aws

User Guide

Amazon Managed Workflows for Apache Airflow



Copyright © 2026 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Amazon Managed Workflows for Apache Airflow: User Guide

Copyright © 2026 Amazon Web Services, Inc. and/or its affiliates. All rights reserved.

Amazon's trademarks and trade dress may not be used in connection with any product or service that is not Amazon's, in any manner that is likely to cause confusion among customers, or in any manner that disparages or discredits Amazon. All other trademarks not owned by Amazon are the property of their respective owners, who may or may not be affiliated with, connected to, or sponsored by Amazon.

Table of Contents

What Is Amazon MWAA? 1
Features 1
Architecture 2
Integration 4
Supported versions 4
What's next? 4
Quick start 5
In this tutorial 5
Prerequisites 6
Step one: Save the CloudFormation template locally 7
Step two: Create the stack using the AWS CLI 17
Step three: Upload a DAG to Amazon S3 and run in the Apache Airflow UI 17
Step four: access logs in CloudWatch Logs 18
What's next? 19
Get started 20
Prerequisites 20
About this guide 20
Before you begin 21
Available regions 21
Create a bucket 23
Before you begin 24
Create the bucket 24
What's next? 26
Create the VPC network 26
Prerequisites 27
Before you begin 27
Options to create the Amazon VPC network 27
What's next? 40
Create an environment 40
Before you begin 41
Apache Airflow versions 41
Create an environment 43
What's next? 26
Managing access 50

Accessing an Amazon MWAA environment ... 50
How it works ... 51
Full console access ... 53
Full API access ... 59
Read-only console access ... 64
Apache Airflow UI access ... 64
Apache Airflow Rest API access ... 65
Apache Airflow CLI access ... 66
Creating a JSON policy ... 67
Example use case ... 67
Permission Set policy examples ... 70
What's next? ... 72
Service-linked role ... 73
Service-linked role permissions for Amazon MWAA ... 73
Creating a service-linked role for Amazon MWAA ... 77
Editing a service-linked role for Amazon MWAA ... 77
Deleting a service-linked role for Amazon MWAA ... 78
Supported regions for Amazon MWAA service-linked roles ... 78
Policy updates ... 78
Execution role ... 79
Execution role overview ... 80
Create a new role ... 82
Access and update an execution role policy ... 82
Grant access to Amazon S3 bucket with account-level public access block ... 84
Use Apache Airflow connections ... 85
Sample policies ... 85
What's next? ... 91
Cross-service confused deputy prevention ... 91
Apache Airflow access modes ... 93
Apache Airflow access modes ... 94
Access modes overview ... 96
Setup for access modes ... 98
Accessing the VPC endpoint for your Apache Airflow webserver (private network access) . 100
Accessing Apache Airflow ... 101
Prerequisites ... 101
Access ... 101

AWS CLI ... 102
Open the Apache Airflow UI ... 102
Log in to Apache Airflow ... 102
Log in using IAM Identity Center ... 102
Create a webserver access token ... 103
Prerequisites ... 103
Using the AWS CLI ... 104
Using a bash script ... 104
Using a Python script ... 105
What's next? ... 106
Setting up a custom domain ... 106
Configure the custom domain ... 107
Set up the networking infrastructure ... 107
Apache Airflow CLI token ... 112
Prerequisites ... 113
Using the AWS CLI ... 113
Using a curl script ... 113
Using a bash script ... 116
Using a Python script ... 117
What's next? ... 120
Using the Apache Airflow REST API ... 120
Granting access to the Apache Airflow REST API: `airflow:InvokeRestApi` ... 123
Calling the Apache Airflow REST API ... 124
Creating a webserver session token and calling the Apache Airflow REST API ... 125
Apache Airflow CLI command reference ... 131
Prerequisites ... 132
What changed? ... 133
Supported CLI commands ... 133
Sample code ... 140
Managing connections ... 143
Overview ... 143
Apache Airflow packages ... 144
Constraints file ... 144
Version-specific provider packages ... 145
Connection types ... 156
Example connection URI string ... 156

Example connection template ... 157
Example using an HTTP connection template for a Jdbc connection ... 158
Configuring Secrets Manager ... 160
Step one: Provide Amazon MWAA with permission to access Secrets Manager secret keys ... 161
Step two: Create the Secrets Manager backend as an Apache Airflow configuration option ... 162
Step three: Generate an Apache Airflow AWS connection URI string ... 164
Step four: Add the variables in Secrets Manager ... 166
Step five: Add the connection in Secrets Manager ... 168
Sample code ... 169
Resources ... 169
What's next? ... 169
Managing environments ... 170
Configuring the environment class ... 170
Environment capabilities ... 171
Apache Airflow schedulers ... 173
Cost allocation tags ... 174
Adding and activating cost allocation tags ... 174
Viewing costs by tag ... 175
Limitations ... 176
Related resources ... 176
Configuring worker auto scaling ... 177
How worker scaling works ... 177
Using the Amazon MWAA console ... 178
Example high performance use case ... 178
Troubleshooting tasks stuck in the running state ... 180
What's next? ... 180
Configuring webserver auto scaling ... 180
How webserver scaling works ... 181
Using the Amazon MWAA console ... 181
Using configuration options ... 182
Prerequisites ... 183
How it works ... 183
Using configuration options to load plugins ... 183
Configuration options overview ... 184

Configuration reference ... 185
Unsupported configurations ... 192
Examples and sample code ... 193
What's next? ... 194
Update an environment ... 194
Before you begin ... 195
Worker replacement strategy ... 196
How environment updates work ... 196
Update environment resources ... 198
Update an environment ... 198
Multi-team environments ... 202
Prerequisites ... 203
Key concepts ... 203
Choosing a multi-team environment or separate environments ... 205
Limitations ... 206
Next steps ... 207
Managing teams ... 207
Adding teams (console) ... 207
Adding and updating teams (AWS CLI) ... 209
Removing teams ... 210
Changing the version ... 211
Upgrade or downgrade your workflow resources ... 212
Specify the new version ... 213
Using a startup script ... 214
Configure a startup script ... 214
Install Linux runtimes ... 218
Set environment variables ... 219
Working with DAGs ... 223
Amazon S3 bucket overview ... 223
Adding or updating DAGs ... 224
Prerequisites ... 224
How it works ... 225
What's changed? ... 226
Testing DAGs using the Amazon MWAA CLI utility ... 226
Uploading DAG code to Amazon S3 ... 226
Specifying the path to a DAGs folder ... 227

Accessing changes on your Apache Airflow UI ... 228
DAG folders in multi-team environments ... 228
What's next? ... 228
Installing custom plugins ... 228
Prerequisites ... 229
How it works ... 230
When to use the plugins ... 230
Custom plugins overview ... 231
Examples of custom plugins ... 231
Creating a plugins.zip file ... 241
Uploading `plugins.zip` to Amazon S3 ... 242
Installing custom plugins on your environment ... 244
Example use cases for plugins.zip ... 245
What's next? ... 245
Installing Python dependencies ... 245
Prerequisites ... 246
How it works ... 246
Python dependencies overview ... 247
Creating a requirements.txt file ... 247
Uploading `requirements.txt` to Amazon S3 ... 251
Installing Python dependencies on your environment ... 252
Accessing logs for your `requirements.txt` ... 253
What's next? ... 254
Deleting files on Amazon S3 ... 254
Prerequisites ... 255
Versioning overview ... 255
How it works ... 255
Deleting a DAG on Amazon S3 ... 256
Removing "current" plugins.zip or requirements.txt ... 257
Delete "non-current" plugins.zip or requirements.txt ... 257
Deleting files with lifecycles ... 257
Example lifecycle policy ... 258
What's next? ... 258
Using agent skills ... 259
Overview ... 259
Installing the skills ... 260

Prerequisites ... 260
Installing the aws-data-analytics plugin ... 260
Example prompts ... 261
Related resources ... 262
Networking ... 263
About networking ... 263
Terms ... 264
What's supported ... 264
VPC infrastructure overview ... 264
Example use cases for an Amazon VPC and Apache Airflow access mode ... 268
Security in your VPC ... 270
Terms ... 271
Security overview ... 271
Network access control lists (ACLs) ... 271
VPC security groups ... 272
VPC endpoint policies (private routing only) ... 274
Managing access to VPC endpoints ... 275
Pricing ... 276
VPC endpoint overview ... 276
Permission to use other AWS services ... 277
Accessing VPC endpoints ... 277
Accessing the VPC endpoint for your Apache Airflow webserver (private network access) . 279
VPC service endpoints in private Amazon VPCs ... 281
Pricing ... 282
Private network and private routing ... 282
(Required) VPC endpoints ... 283
Attaching the required VPC endpoints ... 284
(Optional) Enable private IP addresses for your Amazon S3 VPC interface endpoint ... 287
Managing your own Amazon VPC endpoints ... 288
Creating an environment in a shared Amazon VPC ... 289
Tutorials ... 300
Tutorial: AWS Client VPN ... 300
Private network ... 301
Use cases ... 302
Before you begin ... 302
Objectives ... 302

(Optional) Step one: Identify your VPC, CIDR rules, and VPC security ... 303
Step two: Create the server and client certificates ... 304
Step three: Save the CloudFormation template locally ... 305
Step four: Create the Client VPN CloudFormation stack ... 307
Step five: Associate subnets to your Client VPN ... 307
Step six: Add an authorization ingress rule to your Client VPN ... 308
Step seven: Download the Client VPN endpoint configuration file ... 309
Step eight: Connect to the AWS Client VPN ... 310
What's next? ... 311
Tutorial: Linux Bastion Host ... 311
Private network ... 312
Use cases ... 313
Before you begin ... 313
Objectives ... 313
Step 1: Create the bastion instance ... 314
Step 2: Create the SSH tunnel ... 314
Step three: Configure the bastion security group as an inbound rule ... 316
Step four: Copy the Apache Airflow URL ... 316
Step five: Configure proxy settings ... 316
Step six: Open the Apache Airflow UI ... 319
What's next? ... 319
Tutorial: Restricting users to a subset of DAGs ... 320
Prerequisites ... 320
Step one: Provide Amazon MWAA webserver access to your IAM principal with the default `Public` Apache Airflow role. ... 321
Step two: Create a new Apache Airflow custom role ... 322
Step three: Assign the role you created to your Amazon MWAA user ... 323
Next steps ... 324
Related resources ... 324
Tutorial: Automate managing your own environment endpoints ... 324
Prerequisites ... 325
Create the Amazon VPC ... 325
Create the Lambda function ... 326
Create the EventBridge rule ... 326
Create the environment ... 327
Code examples ... 329

Import variables DAG ... 330
Version ... 330
Prerequisites ... 330
Permissions ... 330
Dependencies ... 330
Code sample ... 330
What's next? ... 332
Using the `SSHOperator` ... 332
Version ... 333
Prerequisites ... 333
Permissions ... 333
Requirements ... 333
Copy your secret key to Amazon S3 ... 334
Create a new Apache Airflow connection ... 334
Code sample ... 335
Apache Airflow Snowflake connection in Secrets Manager ... 337
Version ... 337
Prerequisites ... 337
Permissions ... 337
Requirements ... 338
Code sample ... 338
What's next? ... 339
Using a DAG to write custom metrics ... 339
Version ... 340
Prerequisites ... 340
Permissions ... 340
Dependencies ... 340
Code example ... 340
Aurora PostgreSQL database cleanup ... 343
Version ... 344
Prerequisites ... 344
Dependencies ... 344
Code sample ... 344
Exporting environment metadata to Amazon S3 ... 351
Version ... 352
Prerequisites ... 352

Permissions ... 352
Requirements ... 352
Code sample ... 353
Using an Apache Airflow variable in Secrets Manager ... 355
Version ... 355
Prerequisites ... 356
Permissions ... 356
Requirements ... 356
Code sample ... 356
What's next? ... 357
Using an Apache Airflow connection in Secrets Manager ... 357
Version ... 358
Prerequisites ... 358
Permissions ... 358
Requirements ... 356
Code sample ... 358
What's next? ... 360
Custom plugin with Oracle ... 360
Version ... 361
Prerequisites ... 361
Permissions ... 361
Requirements ... 361
Code sample ... 362
Create the custom plugin ... 362
Airflow configuration options ... 366
What's next? ... 366
Changing a DAG's timezone ... 366
Version ... 367
Prerequisites ... 367
Permissions ... 367
Create a plugin to change the timezone in Airflow logs ... 367
Create a `plugins.zip` ... 368
Code sample ... 368
What's next? ... 369
Refreshing an AWS CodeArtifact token at runtime ... 370
Version ... 370

Prerequisites ... 370
Permissions ... 370
Code sample ... 371
What's next? ... 372
Custom plugin with Apache Hive and Hadoop ... 373
Version ... 373
Prerequisites ... 373
Permissions ... 373
Requirements ... 356
Download dependencies ... 374
Custom plugin ... 375
Plugins.zip ... 375
Code sample ... 376
Airflow configuration options ... 376
What's next? ... 376
Invoking DAGs with Lambda ... 377
Version ... 377
Prerequisites ... 377
Permissions ... 378
Dependencies ... 378
Code example ... 378
Invoking DAGs in different environments ... 380
Version ... 380
Prerequisites ... 380
Permissions ... 380
Dependencies ... 381
Code example ... 381
Amazon RDS server ... 383
Version ... 383
Prerequisites ... 384
Dependencies ... 344
Apache Airflow v2 connection ... 384
Code sample ... 385
What's next? ... 387
Amazon EKS (eksctl) ... 387
Version ... 388

Prerequisites ... 388
Create a public key for Amazon EC2 ... 388
Create the cluster ... 389
Create a `mwaa` namespace ... 390
Create a role for the `mwaa` namespace ... 390
Create and attach an IAM role for the Amazon EKS cluster ... 392
Create the requirements.txt file ... 395
Grant the Amazon MWAA execution role access to the cluster ... 395
Create the `kubeconfig` ... 397
Create a DAG ... 397
Add the DAG and `kube_config.yaml` to the Amazon S3 bucket ... 398
Enable and trigger the example ... 398
Using the `ECSOperator` ... 399
Version ... 399
Prerequisites ... 399
Permissions ... 400
Create an Amazon ECS cluster ... 401
Code sample ... 406
Using dbt with Amazon MWAA ... 410
Version ... 411
Prerequisites ... 411
Dependencies ... 411
Upload a dbt project to Amazon S3 ... 413
Use a DAG to verify dbt dependency installation ... 413
Create and upload a dbt profiles.yml ... 414
Use a DAG to run a dbt project ... 415
Using DuckDB with Amazon MWAA ... 416
Version ... 417
Prerequisites ... 417
How DuckDB connects in Amazon MWAA ... 417
Using DuckDB with Apache Airflow v2.11 through v3.2 ... 418
Customize DuckDB in your environment ... 419
Troubleshooting DuckDB tasks ... 419
Next steps ... 421
Related resources ... 421
Using a DAG to run DuckDB queries in memory ... 421

Using a DAG to read and write Amazon S3 data with DuckDB ... 424
AWS blogs and tutorials ... 430
Best practices ... 431
Performance tuning for Apache Airflow ... 431
Adding an Apache Airflow configuration option ... 431
Apache Airflow scheduler ... 432
DAG folders ... 436
DAG files ... 437
Tasks ... 440
Managing Python dependencies ... 444
Testing DAGs using the Amazon MWAA CLI utility ... 445
Installing Python dependencies using PyPi.org Requirements File Format ... 445
Enabling logs on the Amazon MWAA console ... 452
Accessing logs on the CloudWatch Logs console ... 453
Accessing errors in the Apache Airflow UI ... 454
Example `requirements.txt` scenarios ... 454
Monitoring and metrics ... 456
Overview ... 456
Amazon CloudWatch overview ... 457
AWS CloudTrail overview ... 457
Accessing audit logs ... 457
Creating a trail in CloudTrail ... 458
Logging data plane events ... 458
Accessing events with CloudTrail Event History ... 459
Example trail for `CreateEnvironment` ... 459
InvokeRestApi log entry example ... 461
What's next? ... 462
Accessing Airflow logs ... 462
Pricing ... 463
Before you begin ... 463
Log types ... 463
Enabling Apache Airflow logs ... 464
Accessing Apache Airflow logs ... 465
Example scheduler logs ... 465
What's next? ... 466
Monitoring dashboards and alarms ... 466

View the metrics dashboard ... 467
Create and monitor alarms ... 470
Alarm states overview ... 473
Example custom dashboards and alarms ... 473
Deleting metrics and dashboards ... 479
What's next? ... 479
Apache Airflow environment metrics ... 479
Terms ... 480
Dimensions ... 480
Accessing metrics in the CloudWatch console ... 481
Apache Airflow metrics available in CloudWatch ... 482
Choosing which metrics are reported ... 501
What's next? ... 501
Container, queue, and database metrics ... 501
Terms ... 502
Dimensions ... 503
Accessing metrics ... 503
List of metrics ... 504
Security ... 508
Shared responsibility model ... 509
Infrastructure that AWS manages ... 509
Workflow code and content that you manage ... 511
Operational excellence that you manage ... 512
Configuration and access that you manage ... 514
Related resources ... 515
Data Protection ... 515
Encryption ... 516
Using customer-managed keys ... 518
AWS Identity and Access Management ... 522
Audience ... 523
Authenticating With Identities ... 523
Managing Access Using Policies ... 524
Allowing users to access their own permissions ... 526
Troubleshooting Amazon Managed Workflows for Apache Airflow identity and access ... 527
How Amazon MWAA works with IAM ... 528
IAM Identity Center integration ... 533

IAM Identity Center integration ... 533
Prerequisites ... 534
How it works ... 534
Enabling IAM Identity Center integration ... 535
Configuring access with Permission Sets ... 536
Teams and the `airflow:TeamNames` condition key ... 544
Logging in to Apache Airflow with IAM Identity Center ... 544
Known limitations ... 546
Troubleshooting ... 547
Compliance Validation ... 548
Resilience ... 549
Infrastructure Security ... 549
AWS PrivateLink ... 549
Creating a VPC endpoint ... 550
FIPS endpoints ... 550
Configuration and Vulnerability Analysis ... 550
Best practices ... 551
Security best practices in Apache Airflow ... 551
Security best practices for multi-team environments ... 553
Versions ... 554
About Amazon MWAA versions ... 554
Latest version ... 554
Apache Airflow versions ... 554
Apache Airflow components ... 556
Schedulers ... 556
Workers ... 557
Upgrading the Apache Airflow version ... 557
Downgrading the Apache Airflow version ... 557
Apache Airflow deprecated versions ... 557
End-of-support versions ... 558
Endpoints and quotas ... 560
Service endpoints ... 560
Service quotas ... 560
Increasing quotas ... 561
FAQs ... 562
Supported versions ... 563

Apache Airflow support ... 563
Python version ... 563
Use cases ... 565
Can I use Amazon MWAA with Amazon SageMaker Unified Studio? ... 565
When can I use AWS Step Functions vs. Amazon MWAA? ... 565
Environment specifications ... 565
How much task storage is available to each environment? ... 565
Default OS ... 565
Custom images ... 566
HIPAA compliance ... 566
Does Amazon MWAA support Spot Instances? ... 566
Custom domain ... 566
SSH access ... 566
Self-referencing rule ... 567
Custom metrics ... 567
Store data ... 567
Worker quota ... 568
Shared Amazon VPCs ... 568
Shared Amazon VPCs ... 568
Metrics ... 568
Worker metrics ... 568
Custom metrics ... 568
DAGs, Operators, Connections, and other questions ... 569
PythonVirtualenvOperator ... 569
How long does it take Amazon MWAA to recognize a new DAG file? ... 569
Why is my DAG file not picked up by Apache Airflow? ... 569
Remove plugins.zip or requirements.txt ... 569
Remove plugins.zip or requirements.txt ... 570
Can I use AWS Database Migration Service (DMS) Operators? ... 570
When I access the Airflow REST API using the AWS credentials, can I increase the throttling limit to more than 10 transactions per second (TPS)? ... 570
Where does the Airflow Task Execution API server run in Amazon MWAA? ... 570
Troubleshooting ... 571
Apache Airflow v2 and v3 ... 574
Connections ... 574
Webserver ... 577

Tasks ... 578
CLI ... 581
Operators ... 582
Amazon MWAA Create/Update ... 583
Environment errors ... 585
Updating `requirements.txt` ... 590
Plugins ... 591
Create bucket ... 591
Create environment ... 592
Update environment ... 595
Access environment ... 596
Multi-team environments ... 596
CloudWatch Logs and CloudTrail ... 597
Logs ... 598
Amazon MWAA User Guide history ... 604

What Is Amazon Managed Workflows for Apache Airflow?

Use Amazon Managed Workflows for Apache Airflow, a managed service for Apache Airflow, to set up and run data pipelines in the cloud at scale. Apache Airflow is an open-source tool used to create, schedule, and monitor *workflows*.

With Amazon MWAA, you can use Apache Airflow and Python to create workflows without managing infrastructure for scalability, availability, and security. Amazon MWAA automatically scales to meet your workflow needs. It integrates with AWS security services to provide fast, secure access to your data.

Content

- Features
- Architecture
- Integration
- Supported versions
- What's next?

Features

Review the following features to learn how Amazon MWAA can simplify managing your Apache Airflow workflows.

- **Automatic Airflow setup** – Quickly set up Apache Airflow by choosing an Apache Airflow version when you create an Amazon MWAA environment. Amazon MWAA sets up Apache Airflow for you using the same Apache Airflow user interface and open-source code available on the internet.
- **Automatic scaling** – Automatically scale Apache Airflow workers (the compute resources that run your tasks) by setting minimum and maximum limits. Amazon MWAA monitors the workers in your environment and uses its autoscaling component to add workers to meet demand, up to the maximum number you defined.
- **Built-in authentication** – Enable role-based authentication and authorization for your Apache Airflow webserver by defining the access control policies in AWS Identity and Access Management (IAM). The Apache Airflow workers assume these policies for secure access to AWS

services. For environments that require per-user identity and IAM-native authorization, you can also enable IAM Identity Center integration. For more information, see Using IAM Identity Center with Amazon Managed Workflows for Apache Airflow.

- **Built-in security** – The Apache Airflow workers and schedulers run in Amazon MWAA's Amazon VPC. Data is also automatically encrypted using AWS Key Management Service, so your environment is secure by default.
- **Public or private access modes** – Access your Apache Airflow webserver using a private, or public access mode. The **Public network** access mode uses a VPC endpoint for your Apache Airflow webserver that is accessible over the internet. The **Private network** access mode uses a VPC endpoint for your Apache Airflow webserver that is accessible *in your VPC*. In both cases, access for your Apache Airflow users is controlled by the access control policy you define in AWS Identity and Access Management (IAM), and AWS SSO.
- **Streamlined upgrades and patches** – Amazon MWAA provides new versions of Apache Airflow periodically. The Amazon MWAA team will update and patch the images for these versions.
- **Workflow monitoring** – access Apache Airflow logs and Apache Airflow metrics in Amazon CloudWatch to identify Apache Airflow task delays or workflow errors without the need for additional third-party tools. Amazon MWAA automatically sends environment metrics—and if enabled—Apache Airflow logs to CloudWatch.
- **AWS integration** – Amazon MWAA supports open-source integrations with Amazon Athena, AWS Batch, Amazon CloudWatch, Amazon DynamoDB, AWS DataSync, Amazon EMR, AWS Fargate, Amazon EKS, Amazon Data Firehose, AWS Glue, AWS Lambda, Amazon Redshift, Amazon SQS, Amazon SNS, Amazon SageMaker AI, and Amazon S3, as well as hundreds of built-in and community-created operators and sensors.
- **Worker fleets** – Amazon MWAA offers support for using containers to scale the worker fleet on demand and reduce scheduler outages using Amazon ECS on AWS Fargate. Operators that invoke tasks on Amazon ECS containers, and Kubernetes operators that create and run pods on a Kubernetes cluster are supported.

Architecture

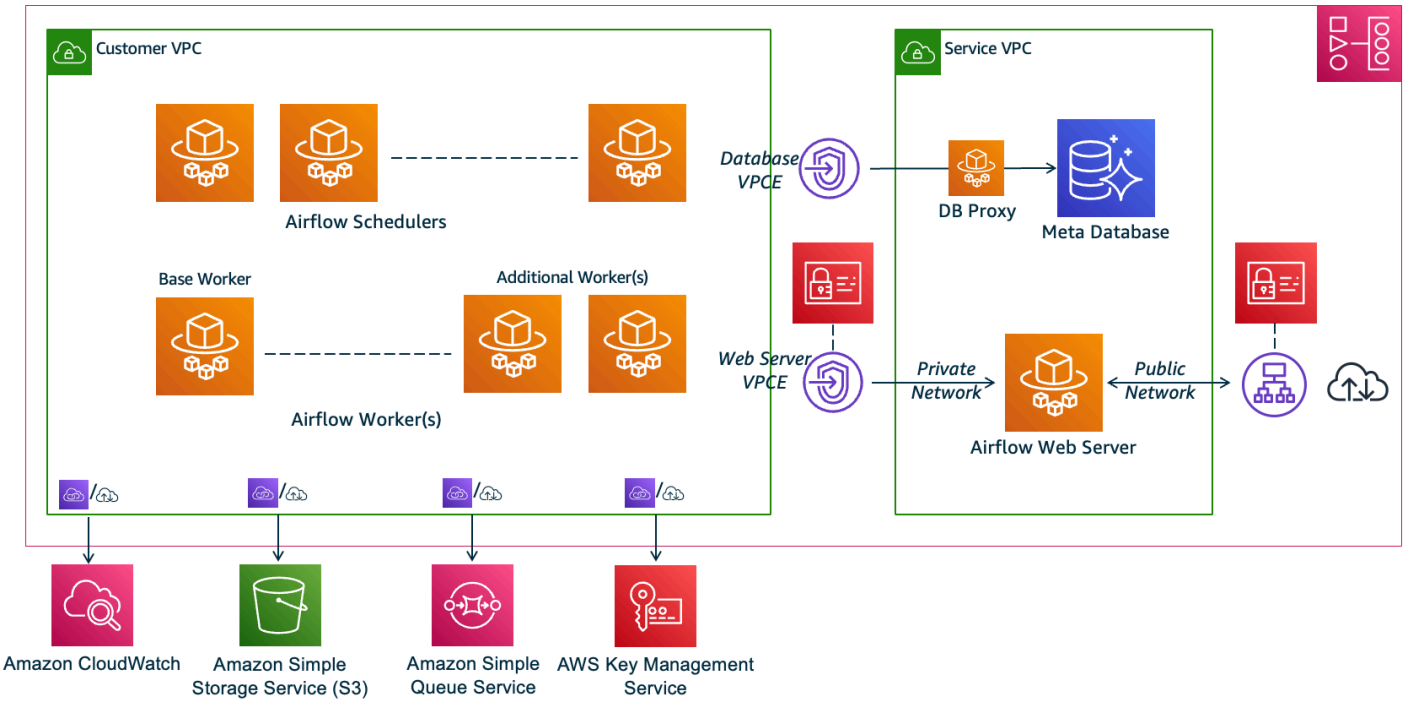
All of the components contained in the outer box (in the following image) are shown as a single Amazon MWAA environment in your account. The Apache Airflow scheduler and workers are AWS Fargate containers that connect to the private subnets in the Amazon VPC for your environment. Each environment has its own Apache Airflow metadatabase managed by AWS that is accessible to the scheduler and workers Fargate containers through a privately-secured VPC endpoint.

Amazon CloudWatch, Amazon S3, Amazon SQS, and AWS KMS are separate from Amazon MWAA and need to be accessible from the Apache Airflow schedulers and workers in the Fargate containers. Multiple Apache Airflow schedulers are only available with Apache Airflow v2 and later. Learn more about the Apache Airflow task lifecycle at Concepts in the *Apache Airflow reference guide*.

The Apache Airflow webserver can be accessed either through the internet by selecting the **Public network** Apache Airflow access mode, or *within your VPC* by selecting the **Private network** Apache Airflow access mode. In both cases, access for your Apache Airflow users is controlled by the access control policy you define in AWS Identity and Access Management (IAM).

> **Note**
>
> Starting with Apache Airflow v3, the Amazon MWAA webserver also hosts Apache Airflow's execution API server.

Amazon MWAA Architecture



Integration

The active and growing Apache Airflow open-source community provides operators (plugins that simplify connections to services) for Apache Airflow to integrate with AWS services. This includes services such as Amazon S3, Amazon Redshift, Amazon EMR, AWS Batch, and Amazon SageMaker AI, as well as services on other cloud platforms.

Using Apache Airflow with Amazon MWAA fully supports integration with AWS services and popular third-party tools such as Apache Hadoop, Presto, Hive, and Spark to perform data processing tasks. Amazon MWAA is committed to maintaining compatibility with the Apache Airflow API, and Amazon MWAA intends to provide reliable integrations to AWS services and make them available to the community, and be involved in community feature development.

For sample code, refer to Code examples for Amazon Managed Workflows for Apache Airflow.

Supported versions

Amazon MWAA supports multiple versions of Apache Airflow. For more information about the Apache Airflow versions we support and the Apache Airflow components included with each version, refer to Apache Airflow versions on Amazon Managed Workflows for Apache Airflow.

What's next?

- Get started with a single CloudFormation template that creates an Amazon S3 bucket for your Airflow DAGs and supporting files, an Amazon VPC with public routing, and an Amazon MWAA environment in Quick start tutorial for Amazon Managed Workflows for Apache Airflow.
- Get started incrementally by creating an Amazon S3 bucket for your Airflow DAGs and supporting files, choosing from one of three Amazon VPC networking options, and creating an Amazon MWAA environment in Get started with Amazon Managed Workflows for Apache Airflow.

Quick start tutorial for Amazon Managed Workflows for Apache Airflow

This quick start tutorial uses an AWS CloudFormation template that creates the Amazon VPC infrastructure, an Amazon S3 bucket with a `dags` folder, and an Amazon Managed Workflows for Apache Airflow environment together.

Topics

- In this tutorial
- Prerequisites
- Step one: Save the CloudFormation template locally
- Step two: Create the stack using the AWS CLI
- Step three: Upload a DAG to Amazon S3 and run in the Apache Airflow UI
- Step four: access logs in CloudWatch Logs
- What's next?

In this tutorial

Use this tutorial to upload a DAG to Amazon S3, run the DAG in Apache Airflow, and access logs in CloudWatch using three AWS Command Line Interface (AWS CLI) commands. Finally, you'll learn to create an IAM policy for an Apache Airflow development team.

> **Note**
>
> The CloudFormation template on this page creates an Amazon Managed Workflows for Apache Airflow environment for the latest version of Apache Airflow available in CloudFormation. The latest available version is Apache Airflow v3.3.1.

The CloudFormation template creates the following:

- **VPC infrastructure**. The template uses Public routing over the internet. It uses the Public network access mode for the Apache Airflow webserver in `WebserverAccessMode: PUBLIC_ONLY`.

- **Amazon S3 bucket**. The template creates an Amazon S3 bucket with a `dags` folder. It's configured to **Block all public access**, with **Bucket Versioning** enabled, as defined in Create an Amazon S3 bucket for Amazon MWAA.
- **Amazon MWAA environment**. The template creates an Amazon MWAA environment that's associated to the `dags` folder on the Amazon S3 bucket, an execution role with permission to AWS services used by Amazon MWAA, and the default for encryption using an AWS-owned key, as defined in Create an Amazon MWAA environment.
- **CloudWatch Logs**. The template turns on Apache Airflow logs in CloudWatch at the **INFO** level and up for the *Airflow scheduler log group*, *Airflow webserver log group*, *Airflow worker log group*, *Airflow DAG processing log group*, and the *Airflow task log group*, as defined in Accessing Airflow logs in Amazon CloudWatch.

In this tutorial, you'll complete the following tasks:

- **Upload and run a DAG**. Upload Apache Airflow's tutorial DAG for the latest Amazon MWAA supported Apache Airflow version to Amazon S3, and then run in the Apache Airflow UI, as defined in Adding or updating DAGs.
- **Access logs**. Access the *Airflow webserver log group* in CloudWatch Logs, as defined in Accessing Airflow logs in Amazon CloudWatch.
- **Create an access control policy**. Create an access control policy in IAM for your Apache Airflow development team, as defined in Accessing an Amazon MWAA environment.

> **Note**
>
> In the VPC that hosts the Amazon MWAA environment, set `assignIpv6AddressOnCreation` to `true` for all attached subnets. This setting ensures automatic Internet Protocol version 6 (IPv6) address assignment to the resources within these subnets.

Prerequisites

The AWS Command Line Interface (AWS CLI) is an open source tool that you can use to interact with AWS services using commands in your command-line shell. To complete the steps on this page, you need the following:

- AWS CLI – Install version 2.
- AWS CLI – Quick configuration with `aws configure`.

Step one: Save the CloudFormation template locally

- Copy the contents of the following template and save locally as `mwaa-public-network.yml`. You can also download the template.

```
AWSTemplateFormatVersion: "2010-09-09"

Parameters:

  EnvironmentName:
    Description: An environment name that is prefixed to resource names
    Type: String
    Default: MWAAEnvironment

  VpcCIDR:
    Description: The IP range (CIDR notation) for this VPC
    Type: String
    Default: 10.192.0.0/16

  PublicSubnet1CIDR:
    Description: The IP range (CIDR notation) for the public subnet in the first
 Availability Zone
    Type: String
    Default: 10.192.10.0/24

  PublicSubnet2CIDR:
    Description: The IP range (CIDR notation) for the public subnet in the second
 Availability Zone
    Type: String
    Default: 10.192.11.0/24

  PrivateSubnet1CIDR:
    Description: The IP range (CIDR notation) for the private subnet in the first
 Availability Zone
    Type: String
    Default: 10.192.20.0/24
  PrivateSubnet2CIDR:
```

```
    Description: The IP range (CIDR notation) for the private subnet in the second
 Availability Zone
    Type: String
    Default: 10.192.21.0/24
  MaxWorkerNodes:
    Description: The maximum number of workers that can run in the environment
    Type: Number
    Default: 2
  DagProcessingLogs:
    Description: Log level for DagProcessing
    Type: String
    Default: INFO
  SchedulerLogsLevel:
    Description: Log level for SchedulerLogs
    Type: String
    Default: INFO
  TaskLogsLevel:
    Description: Log level for TaskLogs
    Type: String
    Default: INFO
  WorkerLogsLevel:
    Description: Log level for WorkerLogs
    Type: String
    Default: INFO
  WebserverLogsLevel:
    Description: Log level for WebserverLogs
    Type: String
    Default: INFO
  AvailabilityZone1:
    Description: >-
      First Availability Zone for the subnets. Any two different AZs work.
      Changing this on an existing stack replaces the subnets.
    Type: AWS::EC2::AvailabilityZone::Name
  AvailabilityZone2:
    Description: >-
      Second Availability Zone. Must differ from the first.
      Changing this on an existing stack replaces the subnets.
    Type: AWS::EC2::AvailabilityZone::Name

Rules:
  DifferentAZs:
    Assertions:
      - Assert: !Not [ !Equals [ !Ref AvailabilityZone1, !Ref AvailabilityZone2 ] ]
        AssertDescription: The two Availability Zones must be different.
```

```
Resources:

 ###############################################################################################
  # CREATE VPC

 ###############################################################################################

  VPC:
    Type: AWS::EC2::VPC
    Properties:
      CidrBlock: !Ref VpcCIDR
      EnableDnsSupport: true
      EnableDnsHostnames: true
      Tags:
        - Key: Name
          Value: MWAAEnvironment

  InternetGateway:
    Type: AWS::EC2::InternetGateway
    Properties:
      Tags:
        - Key: Name
          Value: MWAAEnvironment

  InternetGatewayAttachment:
    Type: AWS::EC2::VPCGatewayAttachment
    Properties:
      InternetGatewayId: !Ref InternetGateway
      VpcId: !Ref VPC

  PublicSubnet1:
    Type: AWS::EC2::Subnet
    Properties:
      VpcId: !Ref VPC
      AvailabilityZone: !Ref AvailabilityZone1
      CidrBlock: !Ref PublicSubnet1CIDR
      MapPublicIpOnLaunch: true
      Tags:
        - Key: Name
          Value: !Sub ${EnvironmentName} Public Subnet (AZ1)

  PublicSubnet2:
    Type: AWS::EC2::Subnet
```

```
    Properties:
      VpcId: !Ref VPC
      AvailabilityZone: !Ref AvailabilityZone2
      CidrBlock: !Ref PublicSubnet2CIDR
      MapPublicIpOnLaunch: true
      Tags:
        - Key: Name
          Value: !Sub ${EnvironmentName} Public Subnet (AZ2)

  PrivateSubnet1:
    Type: AWS::EC2::Subnet
    Properties:
      VpcId: !Ref VPC
      AvailabilityZone: !Ref AvailabilityZone1
      CidrBlock: !Ref PrivateSubnet1CIDR
      MapPublicIpOnLaunch: false
      Tags:
        - Key: Name
          Value: !Sub ${EnvironmentName} Private Subnet (AZ1)

  PrivateSubnet2:
    Type: AWS::EC2::Subnet
    Properties:
      VpcId: !Ref VPC
      AvailabilityZone: !Ref AvailabilityZone2
      CidrBlock: !Ref PrivateSubnet2CIDR
      MapPublicIpOnLaunch: false
      Tags:
        - Key: Name
          Value: !Sub ${EnvironmentName} Private Subnet (AZ2)

  NatGateway1EIP:
    Type: AWS::EC2::EIP
    DependsOn: InternetGatewayAttachment
    Properties:
      Domain: vpc

  NatGateway2EIP:
    Type: AWS::EC2::EIP
    DependsOn: InternetGatewayAttachment
    Properties:
      Domain: vpc

  NatGateway1:
```

```
    Type: AWS::EC2::NatGateway
    Properties:
      AllocationId: !GetAtt NatGateway1EIP.AllocationId
      SubnetId: !Ref PublicSubnet1

  NatGateway2:
    Type: AWS::EC2::NatGateway
    Properties:
      AllocationId: !GetAtt NatGateway2EIP.AllocationId
      SubnetId: !Ref PublicSubnet2

  PublicRouteTable:
    Type: AWS::EC2::RouteTable
    Properties:
      VpcId: !Ref VPC
      Tags:
        - Key: Name
          Value: !Sub ${EnvironmentName} Public Routes

  DefaultPublicRoute:
    Type: AWS::EC2::Route
    DependsOn: InternetGatewayAttachment
    Properties:
      RouteTableId: !Ref PublicRouteTable
      DestinationCidrBlock: 0.0.0.0/0
      GatewayId: !Ref InternetGateway

  PublicSubnet1RouteTableAssociation:
    Type: AWS::EC2::SubnetRouteTableAssociation
    Properties:
      RouteTableId: !Ref PublicRouteTable
      SubnetId: !Ref PublicSubnet1

  PublicSubnet2RouteTableAssociation:
    Type: AWS::EC2::SubnetRouteTableAssociation
    Properties:
      RouteTableId: !Ref PublicRouteTable
      SubnetId: !Ref PublicSubnet2


  PrivateRouteTable1:
    Type: AWS::EC2::RouteTable
    Properties:
      VpcId: !Ref VPC
```

```
      Tags:
        - Key: Name
          Value: !Sub ${EnvironmentName} Private Routes (AZ1)

  DefaultPrivateRoute1:
    Type: AWS::EC2::Route
    Properties:
      RouteTableId: !Ref PrivateRouteTable1
      DestinationCidrBlock: 0.0.0.0/0
      NatGatewayId: !Ref NatGateway1

  PrivateSubnet1RouteTableAssociation:
    Type: AWS::EC2::SubnetRouteTableAssociation
    Properties:
      RouteTableId: !Ref PrivateRouteTable1
      SubnetId: !Ref PrivateSubnet1

  PrivateRouteTable2:
    Type: AWS::EC2::RouteTable
    Properties:
      VpcId: !Ref VPC
      Tags:
        - Key: Name
          Value: !Sub ${EnvironmentName} Private Routes (AZ2)

  DefaultPrivateRoute2:
    Type: AWS::EC2::Route
    Properties:
      RouteTableId: !Ref PrivateRouteTable2
      DestinationCidrBlock: 0.0.0.0/0
      NatGatewayId: !Ref NatGateway2

  PrivateSubnet2RouteTableAssociation:
    Type: AWS::EC2::SubnetRouteTableAssociation
    Properties:
      RouteTableId: !Ref PrivateRouteTable2
      SubnetId: !Ref PrivateSubnet2

  SecurityGroup:
    Type: AWS::EC2::SecurityGroup
    Properties:
      VpcId: !Ref VPC
      GroupDescription: !Sub "Security Group for Amazon MWAA Environment
${AWS::StackName}-MwaaEnvironment"
```

```
      GroupName: !Sub "airflow-security-group-${AWS::StackName}-MwaaEnvironment"

  SecurityGroupIngress:
    Type: AWS::EC2::SecurityGroupIngress
    Properties:
      GroupId: !Ref SecurityGroup
      IpProtocol: "-1"
      SourceSecurityGroupId: !Ref SecurityGroup

  EnvironmentBucket:
    Type: AWS::S3::Bucket
    Properties:
      VersioningConfiguration:
        Status: Enabled
      PublicAccessBlockConfiguration:
        BlockPublicAcls: true
        BlockPublicPolicy: true
        IgnorePublicAcls: true
        RestrictPublicBuckets: true


##################################################################################################
  # CREATE MWAA

##################################################################################################

  MwaaEnvironment:
    Type: AWS::MWAA::Environment
    DependsOn: MwaaExecutionPolicy
    Properties:
      Name: !Sub "${AWS::StackName}-MwaaEnvironment"
      SourceBucketArn: !GetAtt EnvironmentBucket.Arn
      ExecutionRoleArn: !GetAtt MwaaExecutionRole.Arn
      DagS3Path: dags/
      NetworkConfiguration:
        SecurityGroupIds:
          - !GetAtt SecurityGroup.GroupId
        SubnetIds:
          - !Ref PrivateSubnet1
          - !Ref PrivateSubnet2
      WebserverAccessMode: PUBLIC_ONLY
      MaxWorkers: !Ref MaxWorkerNodes
      LoggingConfiguration:
        DagProcessingLogs:
```

```
          LogLevel: !Ref DagProcessingLogs
          Enabled: true
        SchedulerLogs:
          LogLevel: !Ref SchedulerLogsLevel
          Enabled: true
        TaskLogs:
          LogLevel: !Ref TaskLogsLevel
          Enabled: true
        WorkerLogs:
          LogLevel: !Ref WorkerLogsLevel
          Enabled: true
        WebserverLogs:
          LogLevel: !Ref WebserverLogsLevel
          Enabled: true

  MwaaExecutionRole:
    Type: AWS::IAM::Role
    Properties:
      AssumeRolePolicyDocument:
        Version: 2012-10-17
        Statement:
          - Effect: Allow
            Principal:
              Service:
                - airflow-env.amazonaws.com
                - airflow.amazonaws.com
            Action:
             - "sts:AssumeRole"
      Path: "/service-role/"

  MwaaExecutionPolicy:
    DependsOn: EnvironmentBucket
    Type: AWS::IAM::ManagedPolicy
    Properties:
      Roles:
        - !Ref MwaaExecutionRole
      PolicyDocument:
        Version: 2012-10-17
        Statement:
          - Effect: Allow
            Action: airflow:PublishMetrics
            Resource:
              - !Sub "arn:aws:airflow:${AWS::Region}:${AWS::AccountId}:environment/
${EnvironmentName}"
```

```
          - Effect: Deny
            Action: s3:ListAllMyBuckets
            Resource:
              - !Sub "${EnvironmentBucket.Arn}"
              - !Sub "${EnvironmentBucket.Arn}/*"

          - Effect: Allow
            Action:
              - "s3:GetObject*"
              - "s3:GetBucket*"
              - "s3:List*"
            Resource:
              - !Sub "${EnvironmentBucket.Arn}"
              - !Sub "${EnvironmentBucket.Arn}/*"
          - Effect: Allow
            Action:
              - logs:DescribeLogGroups
            Resource: "*"

          - Effect: Allow
            Action:
              - logs:CreateLogStream
              - logs:CreateLogGroup
              - logs:PutLogEvents
              - logs:GetLogEvents
              - logs:GetLogRecord
              - logs:GetLogGroupFields
              - logs:GetQueryResults
              - logs:DescribeLogGroups
            Resource:
              - !Sub "arn:aws:logs:${AWS::Region}:${AWS::AccountId}:log-
group:airflow-${AWS::StackName}*"
          - Effect: Allow
            Action: cloudwatch:PutMetricData
            Resource: "*"
          - Effect: Allow
            Action:
              - sqs:ChangeMessageVisibility
              - sqs:DeleteMessage
              - sqs:GetQueueAttributes
              - sqs:GetQueueUrl
              - sqs:ReceiveMessage
              - sqs:SendMessage
            Resource:
```

```
              - !Sub "arn:aws:sqs:${AWS::Region}:*:airflow-celery-*"
          - Effect: Allow
            Action:
              - kms:Decrypt
              - kms:DescribeKey
              - "kms:GenerateDataKey*"
              - kms:Encrypt
            NotResource: !Sub "arn:aws:kms:*:${AWS::AccountId}:key/*"
            Condition:
              StringLike:
                "kms:ViaService":
                  - !Sub "sqs.${AWS::Region}.amazonaws.com"
Outputs:
  VPC:
    Description: A reference to the created VPC
    Value: !Ref VPC

  PublicSubnets:
    Description: A list of the public subnets
    Value: !Join [ ",", [ !Ref PublicSubnet1, !Ref PublicSubnet2 ]]

  PrivateSubnets:
    Description: A list of the private subnets
    Value: !Join [ ",", [ !Ref PrivateSubnet1, !Ref PrivateSubnet2 ]]

  PublicSubnet1:
    Description: A reference to the public subnet in the 1st Availability Zone
    Value: !Ref PublicSubnet1

  PublicSubnet2:
    Description: A reference to the public subnet in the 2nd Availability Zone
    Value: !Ref PublicSubnet2

  PrivateSubnet1:
    Description: A reference to the private subnet in the 1st Availability Zone
    Value: !Ref PrivateSubnet1

  PrivateSubnet2:
    Description: A reference to the private subnet in the 2nd Availability Zone
    Value: !Ref PrivateSubnet2

  SecurityGroupIngress:
    Description: Security group with self-referencing inbound rule
    Value: !Ref SecurityGroupIngress
```

```
  MwaaApacheAirflowUI:
    Description: MWAA Environment
    Value: !Sub  "https://${MwaaEnvironment.WebserverUrl}"
```

Step two: Create the stack using the AWS CLI

1. In your command prompt, navigate to the directory where `mwaa-public-network.yml` is stored. For example:

```
cd mwaaproject
```

2. Use the create-stack command to create the stack using the AWS CLI. Replace *availability-zone-1* and *availability-zone-2* with two different, supported Availability Zones in your Region.

```
aws cloudformation create-stack --stack-name mwaa-environment-public-
network --template-body file://mwaa-public-network.yml --parameters
 ParameterKey=AvailabilityZone1,ParameterValue=availability-zone-1
 ParameterKey=AvailabilityZone2,ParameterValue=availability-zone-2 --capabilities
 CAPABILITY_IAM
```

> **Note**
>
> It takes more than 30 minutes to create the Amazon VPC infrastructure, Amazon S3 bucket, and Amazon MWAA environment.

Step three: Upload a DAG to Amazon S3 and run in the Apache Airflow UI

1. Copy the contents of the `tutorial.py` file for the latest supported Apache Airflow version and save locally as `tutorial.py`.
2. In your command prompt, navigate to the directory where `tutorial.py` is stored. For example:

```
cd mwaaproject
```

3. Use the following command to list all of your Amazon S3 buckets.

```
aws s3 ls
```

4. Use the following command to list the files and folders in the Amazon S3 bucket for your environment.

```
aws s3 ls s3://YOUR_S3_BUCKET_NAME
```

5. Use the following script to upload the `tutorial.py` file to your `dags` folder. Substitute the sample value in *amzn-s3-demo-bucket*.

```
aws s3 cp tutorial.py s3://amzn-s3-demo-bucket/dags/
```

6. Open the Environments page on the Amazon MWAA console.
7. Choose an environment.
8. Choose **Open Airflow UI**.
9. On the Apache Airflow UI, from the list of available DAGs, choose the **tutorial** DAG.
10. On the DAG details page, choose the **Pause/Unpause DAG** toggle next to your DAG name to unpause the DAG.
11. Choose **Trigger DAG**.

Step four: access logs in CloudWatch Logs

You can access Apache Airflow logs in the CloudWatch console for all of the Apache Airflow logs that were turned on by the CloudFormation stack. The following section explains how to access logs for the *Airflow webserver log group*.

1. Open the Environments page on the Amazon MWAA console.
2. Choose an environment.
3. Choose the **Airflow webserver log group** on the **Monitoring** pane.
4. Choose the `webserver_console_ip` log in **Log streams**.

What's next?

- Learn more about how to upload DAGs, specify Python dependencies in a `requirements.txt` and custom plugins in a `plugins.zip` in Working with DAGs on Amazon MWAA.
- Learn more about the best practices we recommend to tune the performance of your environment in Performance tuning for Apache Airflow on Amazon MWAA.
- Create a monitoring dashboard for your environment in Monitoring dashboards and alarms on Amazon MWAA.
- Run some of the DAG code samples in Code examples for Amazon Managed Workflows for Apache Airflow.

Get started with Amazon Managed Workflows for Apache Airflow

Amazon Managed Workflows for Apache Airflow uses the Amazon VPC, DAG files and supporting files in your Amazon S3 storage bucket to create an environment. This chapter describes the prerequisites and AWS resources needed to get started with Amazon MWAA.

Topics

- Prerequisites
- About this guide
- Before you begin
- Available regions
- Create an Amazon S3 bucket for Amazon MWAA
- Create the VPC network
- Create an Amazon MWAA environment
- What's next?

Prerequisites

To create an Amazon MWAA environment, ensure you have permission to the AWS resources you need to create.

- **AWS account** – An AWS account with permission to use Amazon MWAA and the AWS services and resources used by your environment.

About this guide

This guide covers the AWS infrastructure and resources you'll create.

- **Amazon VPC** – The Amazon VPC networking components required by an Amazon MWAA environment. You can configure an existing VPC that meets these requirements (advanced) as found in About networking on Amazon MWAA, or create the VPC and networking components, as defined in the section called “Create the VPC network”.

- **Amazon S3 bucket** – An Amazon S3 bucket to store your DAGs and associated files, such as `plugins.zip` and `requirements.txt`. Your Amazon S3 bucket must be configured to **Block all public access**, with **Bucket Versioning** enabled, as defined in Create an Amazon S3 bucket for Amazon MWAA.
- **Amazon MWAA environment** – An Amazon MWAA environment configured with the location of your Amazon S3 bucket, the path to your DAG code and any custom plugins or Python dependencies, and your Amazon VPC and its security group, as defined in Create an Amazon MWAA environment.

Before you begin

To create an Amazon MWAA environment, you can take additional steps to create and configure other AWS resources before you create your environment.

To create an environment, you need the following:

- **AWS KMS key** – An AWS KMS key for data encryption on your environment. You can choose the default option on the Amazon MWAA console to create an AWS-owned key when you create an environment, or specify an existing Customer-managed key with permissions to other AWS services used by your environment configured (advanced). To learn more, refer to Using customer-managed keys for encryption.
- **Execution role** – An execution role that allows Amazon MWAA to access AWS resources in your environment. You can choose the default option on the Amazon MWAA console to create an execution role when you create an environment. To learn more, refer to Amazon MWAA execution role.
- **VPC security group** – A VPC security group that allows Amazon MWAA to access other AWS resources in your VPC network. You can choose the default option on the Amazon MWAA console to create a security group when you create an environment, or provide a security group with the appropriate inbound and outbound rules (advanced). To learn more, refer to Security in your VPC on Amazon MWAA.

Available regions

Amazon MWAA is available in the following AWS Regions. To learn more about each region, such as which are enabled or disabled by default, refer to AWS Regions.

Code	Name
us-east-1	US East (N. Virginia)
us-east-2	US East (Ohio)
us-west-1	US West (N. California)
us-west-2	US West (Oregon)
af-south-1	Africa (Cape Town)
ap-east-1	Asia Pacific (Hong Kong)
ap-south-1	Asia Pacific (Mumbai)
ap-south-2	Asia Pacific (Hyderabad)
ap-southeast-1	Asia Pacific (Singapore)
ap-southeast-2	Asia Pacific (Sydney)
ap-southeast-3	Asia Pacific (Jakarta)
ap-southeast-4	Asia Pacific (Melbourne)
ap-southeast-5	Asia Pacific (Malaysia)
ap-southeast-7	Asia Pacific (Thailand)
ap-northeast-1	Asia Pacific (Tokyo)
ap-northeast-2	Asia Pacific (Seoul)
ap-northeast-3	Asia Pacific (Osaka)
ca-central-1	Canada (Central)
ca-west-1	Canada West (Calgary)
eu-central-1	Europe (Frankfurt)

Code	Name
eu-central-2	Europe (Zurich)
eu-west-1	Europe (Ireland)
eu-west-2	Europe (London)
eu-west-3	Europe (Paris)
eu-south-1	Europe (Milan)
eu-south-2	Europe (Spain)
eu-north-1	Europe (Stockholm)
il-central-1	Israel (Tel Aviv)
me-south-1	Middle East (Bahrain)
me-central-1	Middle East (UAE)
mx-central-1	Mexico (Central)
sa-east-1	South America (São Paulo)

Create an Amazon S3 bucket for Amazon MWAA

This guide describes the steps to create an Amazon S3 bucket to store your Apache Airflow Directed Acyclic Graphs (DAGs), custom plugins in a `plugins.zip` file, and Python dependencies in a `requirements.txt` file.

Contents

- Before you begin
- Create the bucket
- What's next?

Before you begin

- The Amazon S3 bucket name can't be changed after you create the bucket. To learn more, refer to Rules for bucket naming in the *Amazon Simple Storage Service User Guide*.
- An Amazon S3 bucket used for an Amazon MWAA environment must be configured to **Block all public access**, with **Bucket Versioning** enabled.
- An Amazon S3 bucket used for an Amazon MWAA environment must be located in the same AWS Region as an Amazon MWAA environment. To access a list of AWS Regions for Amazon MWAA, refer to Amazon MWAA endpoints and quotas in the *AWS General Reference*.

Create the bucket

This section describes the steps to create the Amazon S3 bucket for your environment.

To create a bucket

1. Sign in to the AWS Management Console and open the Amazon S3 console at https://console.aws.amazon.com/s3/.
2. Choose **Create bucket**.
3. In **Bucket name**, enter a DNS-compliant name for your bucket.

   The bucket name must:

   - Be unique across all of Amazon S3.
   - Be between 3 and 63 characters long.
   - Not contain uppercase characters.
   - Start with a lowercase letter or number.

   > **Important**
   >
   > Avoid including sensitive information, such as account numbers, in the bucket name. The bucket name is available in the URLs that point to the objects in the bucket.

4. Choose an AWS Region in **Region**. This must be the same AWS Region as your Amazon MWAA environment.

- We recommend choosing a region close to you to minimize latency and costs and address regulatory requirements.

5. Choose **Block all public access**.
6. Choose **Enable** in **Bucket Versioning**.
7. **Optional** - *Tags*. Add key-value tag pairs to identify your Amazon S3 bucket in **Tags**. For example, `Bucket : Staging`.
8. **Optional** - *Server-side encryption*. You can optionally **Enable** one of the following encryption options on your Amazon S3 bucket.

   a. Choose **Amazon S3 key (SSE-S3)** in **Server-side encryption** to enable server-side encryption for the bucket.

   b. Choose **AWS Key Management Service key (SSE-KMS)** to use an AWS KMS key for encryption on your Amazon S3 bucket:

      i. **AWS managed key (aws/s3)** - If you choose this option, you can either use an AWS-owned key managed by Amazon MWAA, or specify a Customer-managed key for encryption of your Amazon MWAA environment.

      ii. **Choose from your AWS KMS keys** or **Enter AWS KMS key ARN** - If you choose to specify a Customer-managed key in this step, you must specify an AWS KMS key ID or ARN. AWS KMS aliases and multi-region keys are not supported by Amazon MWAA. The AWS KMS key you specify must also be used for encryption on your Amazon MWAA environment.

9. **Optional** - *Advanced settings*. If you want to enable Amazon S3 Object Lock:

   a. Choose **Advanced settings**, **Enable**.

   > **Important**
   >
   > Enabling Object Lock will permanently allow objects in this bucket to be locked. To learn more, refer to Locking Objects Using Amazon S3 Object Lock in the *Amazon Simple Storage Service User Guide*.

   b. Choose the acknowledgement.

10. Choose **Create bucket**.

What's next?

- Learn how to create the required Amazon VPC network for an environment in Create the VPC network.
- Learn how to how to manage access permissions in How do I set ACL bucket permissions?
- Learn how to delete a storage bucket in How do I delete an S3 Bucket?.

Create the VPC network

Amazon Managed Workflows for Apache Airflow requires an Amazon VPC and specific networking components to support an environment. This guide describes the different options to create the Amazon VPC network for an Amazon Managed Workflows for Apache Airflow environment.

> **Note**
>
> Apache Airflow works best in a low-latency network environment. If you are using an existing Amazon VPC which routes traffic to another region or to an on-premise environment, we recommended adding AWS PrivateLink endpoints for Amazon SQS, CloudWatch, Amazon S3, and AWS KMS. For more information about configuring AWS PrivateLink for Amazon MWAA, refer to Creating an Amazon VPC network without internet access.

Contents

- Prerequisites
- Before you begin
- Options to create the Amazon VPC network
  - Option one: Creating the VPC network on the Amazon MWAA console
  - Option two: Creating an Amazon VPC network with internet access
  - Option three: Creating an Amazon VPC network without internet access
- What's next?

Prerequisites

The AWS Command Line Interface (AWS CLI) is an open source tool that you can use to interact with AWS services using commands in your command-line shell. To complete the steps on this page, you need the following:

- AWS CLI – Install version 2.
- AWS CLI – Quick configuration with `aws configure`.

Before you begin

- The VPC network you specify for your environment can't be changed after the environment is created.
- You can use private or public routing for your Amazon VPC and Apache Airflow webserver. To access a list of options, refer to the section called "Example use cases for an Amazon VPC and Apache Airflow access mode".

Options to create the Amazon VPC network

The following section describes the options available to create the Amazon VPC network for an environment.

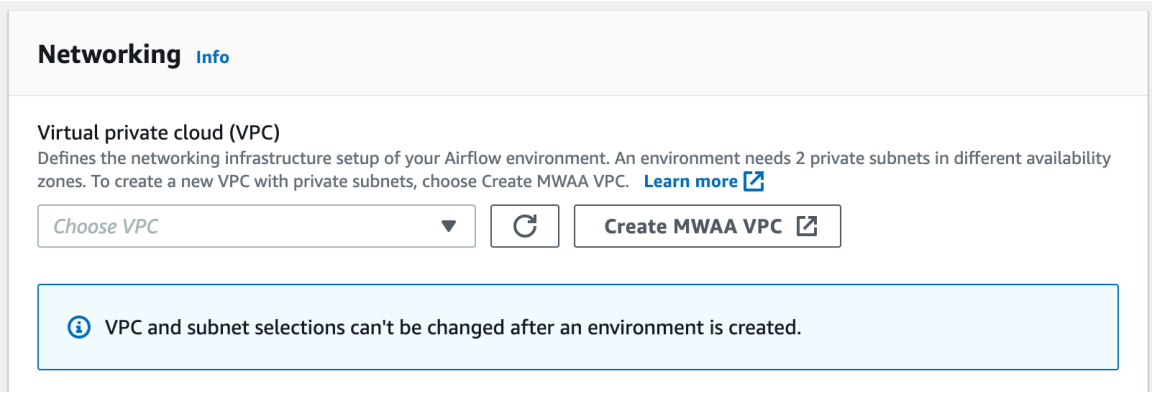
> **Note**
>
> Amazon MWAA does not support the use of `use1-az3` Availability Zone (AZ) in the US East (N. Virginia) Region. When creating the VPC for Amazon MWAA in the US East (N. Virginia) region, you must explicitly assign the `AvailabilityZone` in the CloudFormation (CFN) template. The assigned availability zone name must not be mapped to `use1-az3`. You can retrieve the detailed mapping of AZ names to their corresponding AZ IDs by running the following command:
>
> ```
> aws ec2 describe-availability-zones --region us-east-1
> ```

Choose two supported, different Availability Zones in your Region. If you run several environments, spread them across different zones.

Option one: Creating the VPC network on the Amazon MWAA console

The following section explains how to create an Amazon VPC network on the Amazon MWAA console. This option uses Public routing over the internet. It can be used for an Apache Airflow webserver with the **Private network** or **Public network** access modes.

The following image depicts where you can find the **Create MWAA VPC** button on the Amazon MWAA console.



Option two: Creating an Amazon VPC network *with* internet access

The following CloudFormation template creates an Amazon VPC network with internet access in your default AWS Region. This option uses Public routing over the internet. This template can be used for an Apache Airflow webserver with the **Private network** or **Public network** access modes.

1. Copy the contents of the following template and save locally as `cfn-vpc-public-private.yaml`. You can also download the template.

```
Description:  This template deploys a VPC, with a pair of public and private
 subnets spread
  across two Availability Zones. It deploys an internet gateway, with a default
  route on the public subnets. It deploys a pair of NAT gateways (one in each AZ),
  and default routes for them in the private subnets.

Parameters:
  EnvironmentName:
    Description: An environment name that is prefixed to resource names
    Type: String
    Default: mwaa-

  VpcCIDR:
    Description: Please enter the IP range (CIDR notation) for this VPC
```

```
    Type: String
    Default: 10.192.0.0/16

  PublicSubnet1CIDR:
    Description: Please enter the IP range (CIDR notation) for the public subnet in
 the first Availability Zone
    Type: String
    Default: 10.192.10.0/24

  PublicSubnet2CIDR:
    Description: Please enter the IP range (CIDR notation) for the public subnet in
 the second Availability Zone
    Type: String
    Default: 10.192.11.0/24

  PrivateSubnet1CIDR:
    Description: Please enter the IP range (CIDR notation) for the private subnet
 in the first Availability Zone
    Type: String
    Default: 10.192.20.0/24

  PrivateSubnet2CIDR:
    Description: Please enter the IP range (CIDR notation) for the private subnet
 in the second Availability Zone
    Type: String
    Default: 10.192.21.0/24

  AvailabilityZone1:
    Description: >-
      First Availability Zone for the subnets. Any two different AZs work.
      Changing this on an existing stack replaces the subnets.
    Type: AWS::EC2::AvailabilityZone::Name
  AvailabilityZone2:
    Description: >-
      Second Availability Zone. Must differ from the first.
      Changing this on an existing stack replaces the subnets.
    Type: AWS::EC2::AvailabilityZone::Name

Rules:
  DifferentAZs:
    Assertions:
      - Assert: !Not [ !Equals [ !Ref AvailabilityZone1, !Ref AvailabilityZone2 ] ]
        AssertDescription: The two Availability Zones must be different.
```

```
Resources:
  VPC:
    Type: AWS::EC2::VPC
    Properties:
      CidrBlock: !Ref VpcCIDR
      EnableDnsSupport: true
      EnableDnsHostnames: true
      Tags:
        - Key: Name
          Value: !Ref EnvironmentName

  InternetGateway:
    Type: AWS::EC2::InternetGateway
    Properties:
      Tags:
        - Key: Name
          Value: !Ref EnvironmentName

  InternetGatewayAttachment:
    Type: AWS::EC2::VPCGatewayAttachment
    Properties:
      InternetGatewayId: !Ref InternetGateway
      VpcId: !Ref VPC

  PublicSubnet1:
    Type: AWS::EC2::Subnet
    Properties:
      VpcId: !Ref VPC
      AvailabilityZone: !Ref AvailabilityZone1
      CidrBlock: !Ref PublicSubnet1CIDR
      MapPublicIpOnLaunch: true
      Tags:
        - Key: Name
          Value: !Sub ${EnvironmentName} Public Subnet (AZ1)

  PublicSubnet2:
    Type: AWS::EC2::Subnet
    Properties:
      VpcId: !Ref VPC
      AvailabilityZone: !Ref AvailabilityZone2
      CidrBlock: !Ref PublicSubnet2CIDR
      MapPublicIpOnLaunch: true
      Tags:
        - Key: Name
```

```
        Value: !Sub ${EnvironmentName} Public Subnet (AZ2)

PrivateSubnet1:
  Type: AWS::EC2::Subnet
  Properties:
    VpcId: !Ref VPC
    AvailabilityZone: !Ref AvailabilityZone1
    CidrBlock: !Ref PrivateSubnet1CIDR
    MapPublicIpOnLaunch: false
    Tags:
      - Key: Name
        Value: !Sub ${EnvironmentName} Private Subnet (AZ1)

PrivateSubnet2:
  Type: AWS::EC2::Subnet
  Properties:
    VpcId: !Ref VPC
    AvailabilityZone: !Ref AvailabilityZone2
    CidrBlock: !Ref PrivateSubnet2CIDR
    MapPublicIpOnLaunch: false
    Tags:
      - Key: Name
        Value: !Sub ${EnvironmentName} Private Subnet (AZ2)

NatGateway1EIP:
  Type: AWS::EC2::EIP
  DependsOn: InternetGatewayAttachment
  Properties:
    Domain: vpc

NatGateway2EIP:
  Type: AWS::EC2::EIP
  DependsOn: InternetGatewayAttachment
  Properties:
    Domain: vpc

NatGateway1:
  Type: AWS::EC2::NatGateway
  Properties:
    AllocationId: !GetAtt NatGateway1EIP.AllocationId
    SubnetId: !Ref PublicSubnet1

NatGateway2:
  Type: AWS::EC2::NatGateway
```

```
    Properties:
      AllocationId: !GetAtt NatGateway2EIP.AllocationId
      SubnetId: !Ref PublicSubnet2

  PublicRouteTable:
    Type: AWS::EC2::RouteTable
    Properties:
      VpcId: !Ref VPC
      Tags:
        - Key: Name
          Value: !Sub ${EnvironmentName} Public Routes

  DefaultPublicRoute:
    Type: AWS::EC2::Route
    DependsOn: InternetGatewayAttachment
    Properties:
      RouteTableId: !Ref PublicRouteTable
      DestinationCidrBlock: 0.0.0.0/0
      GatewayId: !Ref InternetGateway

  PublicSubnet1RouteTableAssociation:
    Type: AWS::EC2::SubnetRouteTableAssociation
    Properties:
      RouteTableId: !Ref PublicRouteTable
      SubnetId: !Ref PublicSubnet1

  PublicSubnet2RouteTableAssociation:
    Type: AWS::EC2::SubnetRouteTableAssociation
    Properties:
      RouteTableId: !Ref PublicRouteTable
      SubnetId: !Ref PublicSubnet2


  PrivateRouteTable1:
    Type: AWS::EC2::RouteTable
    Properties:
      VpcId: !Ref VPC
      Tags:
        - Key: Name
          Value: !Sub ${EnvironmentName} Private Routes (AZ1)

  DefaultPrivateRoute1:
    Type: AWS::EC2::Route
    Properties:
```

```
      RouteTableId: !Ref PrivateRouteTable1
      DestinationCidrBlock: 0.0.0.0/0
      NatGatewayId: !Ref NatGateway1

  PrivateSubnet1RouteTableAssociation:
    Type: AWS::EC2::SubnetRouteTableAssociation
    Properties:
      RouteTableId: !Ref PrivateRouteTable1
      SubnetId: !Ref PrivateSubnet1

  PrivateRouteTable2:
    Type: AWS::EC2::RouteTable
    Properties:
      VpcId: !Ref VPC
      Tags:
        - Key: Name
          Value: !Sub ${EnvironmentName} Private Routes (AZ2)

  DefaultPrivateRoute2:
    Type: AWS::EC2::Route
    Properties:
      RouteTableId: !Ref PrivateRouteTable2
      DestinationCidrBlock: 0.0.0.0/0
      NatGatewayId: !Ref NatGateway2

  PrivateSubnet2RouteTableAssociation:
    Type: AWS::EC2::SubnetRouteTableAssociation
    Properties:
      RouteTableId: !Ref PrivateRouteTable2
      SubnetId: !Ref PrivateSubnet2

  SecurityGroup:
    Type: AWS::EC2::SecurityGroup
    Properties:
      GroupName: "mwaa-security-group"
      GroupDescription: "Security group with a self-referencing inbound rule."
      VpcId: !Ref VPC

  SecurityGroupIngress:
    Type: AWS::EC2::SecurityGroupIngress
    Properties:
      GroupId: !Ref SecurityGroup
      IpProtocol: "-1"
      SourceSecurityGroupId: !Ref SecurityGroup
```

```
Outputs:
  VPC:
    Description: A reference to the created VPC
    Value: !Ref VPC

  PublicSubnets:
    Description: A list of the public subnets
    Value: !Join [ ",", [ !Ref PublicSubnet1, !Ref PublicSubnet2 ]]

  PrivateSubnets:
    Description: A list of the private subnets
    Value: !Join [ ",", [ !Ref PrivateSubnet1, !Ref PrivateSubnet2 ]]

  PublicSubnet1:
    Description: A reference to the public subnet in the 1st Availability Zone
    Value: !Ref PublicSubnet1

  PublicSubnet2:
    Description: A reference to the public subnet in the 2nd Availability Zone
    Value: !Ref PublicSubnet2

  PrivateSubnet1:
    Description: A reference to the private subnet in the 1st Availability Zone
    Value: !Ref PrivateSubnet1

  PrivateSubnet2:
    Description: A reference to the private subnet in the 2nd Availability Zone
    Value: !Ref PrivateSubnet2

  SecurityGroupIngress:
    Description: Security group with self-referencing inbound rule
    Value: !Ref SecurityGroupIngress
```

2. In your command prompt, navigate to the directory where `cfn-vpc-public-private.yaml` is stored. For example:

```
cd mwaaproject
```

3. Use the create-stack command to create the stack using the AWS CLI. Replace *availability-zone-1* and *availability-zone-2* with two different, supported Availability Zones in your Region.

```
aws cloudformation create-stack --stack-name mwaa-environment
 --template-body file://cfn-vpc-public-private.yaml --parameters
 ParameterKey=AvailabilityZone1,ParameterValue=availability-zone-1
 ParameterKey=AvailabilityZone2,ParameterValue=availability-zone-2
```

> **Note**
>
> It takes about 30 minutes to create the Amazon VPC infrastructure.

Option three: Creating an Amazon VPC network *without* internet access

The following CloudFormation template creates an Amazon VPC network *without internet access* in your default AWS Region.

This option uses Private routing without internet access. This template can be used for an Apache Airflow webserver with the **Private network** access mode only. It creates the required VPC endpoints for the AWS services used by an environment.

1. Copy the contents of the following template and save locally as `cfn-vpc-private.yaml`. You can also download the template.

```
AWSTemplateFormatVersion: "2010-09-09"

Parameters:
   VpcCIDR:
     Description: The IP range (CIDR notation) for this VPC
     Type: String
     Default: 10.192.0.0/16

   PrivateSubnet1CIDR:
     Description: The IP range (CIDR notation) for the private subnet in the first
 Availability Zone
     Type: String
     Default: 10.192.10.0/24

   PrivateSubnet2CIDR:
     Description: The IP range (CIDR notation) for the private subnet in the second
 Availability Zone
     Type: String
```

```
      Default: 10.192.11.0/24
    AvailabilityZone1:
      Description: >-
        First Availability Zone for the subnets. Any two different AZs work.
        Changing this on an existing stack replaces the subnets.
      Type: AWS::EC2::AvailabilityZone::Name
    AvailabilityZone2:
      Description: >-
        Second Availability Zone. Must differ from the first.
        Changing this on an existing stack replaces the subnets.
      Type: AWS::EC2::AvailabilityZone::Name

Rules:
  DifferentAZs:
    Assertions:
      - Assert: !Not [ !Equals [ !Ref AvailabilityZone1, !Ref AvailabilityZone2 ] ]
        AssertDescription: The two Availability Zones must be different.

Resources:
    VPC:
      Type: AWS::EC2::VPC
      Properties:
        CidrBlock: !Ref VpcCIDR
        EnableDnsSupport: true
        EnableDnsHostnames: true
        Tags:
         - Key: Name
           Value: !Ref AWS::StackName

    RouteTable:
      Type: AWS::EC2::RouteTable
      Properties:
        VpcId: !Ref VPC
        Tags:
         - Key: Name
           Value: !Sub "${AWS::StackName}-route-table"

    PrivateSubnet1:
      Type: AWS::EC2::Subnet
      Properties:
        VpcId: !Ref VPC
        AvailabilityZone: !Ref AvailabilityZone1
        CidrBlock: !Ref PrivateSubnet1CIDR
        MapPublicIpOnLaunch: false
```

```
      Tags:
       - Key: Name
         Value: !Sub "${AWS::StackName} Private Subnet (AZ1)"

  PrivateSubnet2:
    Type: AWS::EC2::Subnet
    Properties:
      VpcId: !Ref VPC
      AvailabilityZone: !Ref AvailabilityZone2
      CidrBlock: !Ref PrivateSubnet2CIDR
      MapPublicIpOnLaunch: false
      Tags:
       - Key: Name
         Value: !Sub "${AWS::StackName} Private Subnet (AZ2)"

  PrivateSubnet1RouteTableAssociation:
    Type: AWS::EC2::SubnetRouteTableAssociation
    Properties:
      RouteTableId: !Ref RouteTable
      SubnetId: !Ref PrivateSubnet1

  PrivateSubnet2RouteTableAssociation:
    Type: AWS::EC2::SubnetRouteTableAssociation
    Properties:
      RouteTableId: !Ref RouteTable
      SubnetId: !Ref PrivateSubnet2

  S3VpcEndoint:
    Type: AWS::EC2::VPCEndpoint
    Properties:
      ServiceName: !Sub "com.amazonaws.${AWS::Region}.s3"
      VpcEndpointType: Gateway
      VpcId: !Ref VPC
      RouteTableIds:
       - !Ref RouteTable

  SecurityGroup:
    Type: AWS::EC2::SecurityGroup
    Properties:
      VpcId: !Ref VPC
      GroupDescription: Security Group for Amazon MWAA Environments to access VPC
endpoints
      GroupName: !Sub "${AWS::StackName}-mwaa-vpc-endpoints"
```

```
  SecurityGroupIngress:
    Type: AWS::EC2::SecurityGroupIngress
    Properties:
      GroupId: !Ref SecurityGroup
      IpProtocol: "-1"
      SourceSecurityGroupId: !Ref SecurityGroup

  SqsVpcEndoint:
    Type: AWS::EC2::VPCEndpoint
    Properties:
      ServiceName: !Sub "com.amazonaws.${AWS::Region}.sqs"
      VpcEndpointType: Interface
      VpcId: !Ref VPC
      PrivateDnsEnabled: true
      SubnetIds:
       - !Ref PrivateSubnet1
       - !Ref PrivateSubnet2
      SecurityGroupIds:
       - !Ref SecurityGroup

  CloudWatchLogsVpcEndoint:
    Type: AWS::EC2::VPCEndpoint
    Properties:
      ServiceName: !Sub "com.amazonaws.${AWS::Region}.logs"
      VpcEndpointType: Interface
      VpcId: !Ref VPC
      PrivateDnsEnabled: true
      SubnetIds:
       - !Ref PrivateSubnet1
       - !Ref PrivateSubnet2
      SecurityGroupIds:
       - !Ref SecurityGroup

  CloudWatchMonitoringVpcEndoint:
    Type: AWS::EC2::VPCEndpoint
    Properties:
      ServiceName: !Sub "com.amazonaws.${AWS::Region}.monitoring"
      VpcEndpointType: Interface
      VpcId: !Ref VPC
      PrivateDnsEnabled: true
      SubnetIds:
       - !Ref PrivateSubnet1
       - !Ref PrivateSubnet2
      SecurityGroupIds:
```

```
        - !Ref SecurityGroup

   KmsVpcEndoint:
     Type: AWS::EC2::VPCEndpoint
     Properties:
       ServiceName: !Sub "com.amazonaws.${AWS::Region}.kms"
       VpcEndpointType: Interface
       VpcId: !Ref VPC
       PrivateDnsEnabled: true
       SubnetIds:
        - !Ref PrivateSubnet1
        - !Ref PrivateSubnet2
       SecurityGroupIds:
        - !Ref SecurityGroup


Outputs:
   VPC:
     Description: A reference to the created VPC
     Value: !Ref VPC

   MwaaSecurityGroupId:
     Description: Associates the Security Group to the environment to allow access
 to the VPC endpoints
     Value: !Ref SecurityGroup

   PrivateSubnets:
     Description: A list of the private subnets
     Value: !Join [ ",", [ !Ref PrivateSubnet1, !Ref PrivateSubnet2 ]]

   PrivateSubnet1:
     Description: A reference to the private subnet in the 1st Availability Zone
     Value: !Ref PrivateSubnet1

   PrivateSubnet2:
     Description: A reference to the private subnet in the 2nd Availability Zone
     Value: !Ref PrivateSubnet2
```

2. In your command prompt, navigate to the directory where `cfn-vpc-private.yaml` is stored. For example:

```
cd mwaaproject
```

3. Use the create-stack command to create the stack using the AWS CLI. Replace *availability-zone-1* and *availability-zone-2* with two different, supported Availability Zones in your Region.

```
aws cloudformation create-stack --stack-name mwaa-private-
environment --template-body file://cfn-vpc-private.yaml --parameters
 ParameterKey=AvailabilityZone1,ParameterValue=availability-zone-1
 ParameterKey=AvailabilityZone2,ParameterValue=availability-zone-2
```

> **Note**
>
> It takes about 30 minutes to create the Amazon VPC infrastructure.

4. You'll need to create a mechanism to access these VPC endpoints from your computer. To learn more, refer to Managing access to service-specific Amazon VPC endpoints on Amazon MWAA.

> **Note**
>
> You can further restrict outbound access in the CIDR of your Amazon MWAA security group. For example, you can restrict to itself by adding a self-referencing outbound rule, the prefix list for Amazon S3, and the CIDR of your Amazon VPC.

What's next?

- Learn how to create an Amazon MWAA environment in Create an Amazon MWAA environment.
- Learn how to create a VPN tunnel from your computer to your Amazon VPC with private routing in Tutorial: Configuring private network access using an AWS Client VPN.

Create an Amazon MWAA environment

Amazon Managed Workflows for Apache Airflow sets up Apache Airflow on an environment in your chosen version using the same open-source Apache Airflow and user interface available from Apache. This guide describes the steps to create an Amazon MWAA environment.

Contents

- Before you begin
- Apache Airflow versions
- Create an environment
  - Step one: Specify details
  - Step two: Configure advanced settings
  - Step three: Review and create

Before you begin

- The VPC network you specify for your environment cannot be modified after the environment is created.
- You need an Amazon S3 bucket configured to **Block all public access**, with **Bucket Versioning** enabled.
- You need an AWS account with permissions to use Amazon MWAA, and permission in AWS Identity and Access Management (IAM) to create IAM roles. If you choose the **Private network** access mode for the Apache Airflow webserver, which limits Apache Airflow access within your Amazon VPC, you'll need permission in IAM to create Amazon VPC endpoints.

> **Note**
>
> Amazon MWAA dynamically determines the network during creation. If you use IPv6 subnets, Amazon MWAA creates IPv6 private link connectivity to the database and webserver. Amazon MWAA does not support transitioning between network types and cannot upgrade existing environments to IPv6.

Apache Airflow versions

The following Apache Airflow versions are supported on Amazon Managed Workflows for Apache Airflow.

Note

- Effective December 30, 2025, Amazon MWAA will end support for Apache Airflow versions v2.4.3, v2.5.1, and v2.6.3. For more information, refer to End-of-support versions.
- Beginning with Apache Airflow v2.2.2, Amazon MWAA supports installing Python requirements, provider packages, and custom plugins directly on the Apache Airflow webserver.
- Beginning with Apache Airflow v2.7.2, your requirements file must include a `--constraint` statement. If you don't provide a constraint, Amazon MWAA will specify one for you to ensure the packages listed in your requirements are compatible with the version of Apache Airflow you're using.

  For more information about setting up constraints in your requirements file, refer to Installing Python dependencies.
- Starting with Apache Airflow 3.3.1, Amazon MWAA supports multi-team environments. For more information, see Multi-team environments.

Apache Airflow version	Apache Airflow release date	Amazon MWAA availability date	Apache Airflow constraints	Python version
v3.3.1	August 12, 2026	September 1, 2026	v3.3.1 constraints file	Python 3.12
v3.2.1	April 22, 2026	May 19, 2026	v3.2.1 constraints file	Python 3.12
v2.11.2	March 14, 2026	July 23, 2026	v2.11.2 constraints file	Python 3.12
v2.11.0	May 20, 2025	January 7, 2026	v2.11.0 constraints file	Python 3.12
v3.0.6	August 29, 2025	October 1, 2025	v3.0.6 constraints file	Python 3.12

Apache Airflow version	Apache Airflow release date	Amazon MWAA availability date	Apache Airflow constraints	Python version
v2.10.3	November 4, 2024	December 18, 2024	v2.10.3 constraints file	Python 3.11
v2.10.1	September 5, 2024	September 26, 2024	v2.10.1 constraints file	Python 3.11
v2.9.2	June 10, 2024	July 9, 2024	v2.9.2 constraints file	Python 3.11
v2.8.1	January 19, 2024	February 23, 2024	v2.8.1 constraints file	Python 3.11
v2.7.2	October 12, 2023	November 6, 2023	v2.7.2 constraints file	Python 3.11

For more information about migrating your self-managed Apache Airflow deployments, or migrating an existing Amazon MWAA environment, including instructions for backing up your metadata database, refer to the Amazon MWAA Migration Guide.

Create an environment

The following section describes the steps to create an Amazon MWAA environment.

Step one: Specify details

To specify details for the environment

1. Open the Amazon MWAA console.
2. Select your AWS Region.
3. Choose **Create environment**.
4. On the **Specify details** page, under **Environment details**:
    a. Enter a unique name for your environment in **Name**.
    b. Choose the Apache Airflow version in **Airflow version**.

> **Note**
>
> If no value is specified, defaults to the latest Apache Airflow version. The latest available version is Apache Airflow v3.3.1.

c. Enter the maintenance period for your environment in the **Weekly maintenance window**.

> **Note**
>
> Amazon MWAA performs two types of maintenance on your environment:
>
> Routine database maintenance
>
> Amazon MWAA passes a maintenance window to the underlying Aurora PostgreSQL metadata database. This maintenance does not interrupt your running DAGs, tasks, or other Apache Airflow operations.
>
> Critical patching
>
> In rare cases, Amazon MWAA applies security or stability patches to your environment. You receive a notification through the AWS Health Dashboard before any patch is applied. Your environment restarts during patching and is temporarily unavailable.
>
> You cannot opt out of maintenance on managed services.

5. Under **DAG code in Amazon S3** specify the following:

a. **S3 Bucket**. Choose **Browse S3** and select your Amazon S3 bucket, or enter the Amazon S3 URI.

b. **DAGs folder**. Choose **Browse S3** and select the `dags` folder in your Amazon S3 bucket, or enter the Amazon S3 URI.

c. **Plugins file - *optional***. Choose **Browse S3** and select the `plugins.zip` file on your Amazon S3 bucket, or enter the Amazon S3 URI.

d. **Requirements file - *optional***. Choose **Browse S3** and select the `requirements.txt` file on your Amazon S3 bucket, or enter the Amazon S3 URI.

e. **Startup script file - *optional***, Choose **Browse S3** and select the script file on your Amazon S3 bucket, or enter the Amazon S3 URI.

6. Choose **Next**.

Step two: Configure advanced settings

To configure advanced settings

1. On the **Configure advanced settings** page, under **Networking**:

   - Choose your Amazon VPC.

     This step populates two of the private subnets in your Amazon VPC.

2. Under **webserver access**, select your preferred Apache Airflow access mode:

   a. **Private network**. This limits access of the Apache Airflow UI to users *within your Amazon VPC* that have been granted access to the IAM policy for your environment. You need permission to create Amazon VPC endpoints for this step.

   > **Note**
   >
   > Choose the **Private network** option if your Apache Airflow UI is only accessed within a corporate network, and you do not require access to public repositories for webserver requirements installation. If you choose this access mode option, you need to create a mechanism to access your Apache Airflow webserver in your Amazon VPC. For more information, refer to Accessing the VPC endpoint for your Apache Airflow webserver (private network access).

   b. **Public network**. This allows the Apache Airflow UI to be accessed over the internet by users granted access to the IAM policy for your environment.

   c. **Both public and private network access**. Available for Apache Airflow version 3.2.1 and later. This allows the Apache Airflow UI to be accessed over the internet, while workers communicate with the webserver through a private VPC endpoint. Choose this option if the Amazon VPC hosting your environment lacks internet access.

3. Under **Security groups**, choose the security group used to secure your Amazon VPC:

a. By default, Amazon MWAA creates a security group in your Amazon VPC with specific inbound and outbound rules in **Create new security group**.

b. **Optional**. Deselect the check box in **Create new security group** to select up to 5 security groups.

> **Note**
>
> An existing Amazon VPC security group must be configured with specific inbound and outbound rules to allow network traffic. To learn more, refer to Security in your VPC on Amazon MWAA.

4. Under **Environment class**, choose an environment class.

   We recommend choosing the smallest size necessary to support your workload. You can change the environment class at any time.

5. For **Maximum worker count**, specify the maximum number of Apache Airflow workers to run in the environment.

   For more information, refer to Example high performance use case.

6. Specify the **Maximum web server count** and **Minimum web server count** to configure how Amazon MWAA scales the Apache Airflow web servers in your environment.

   For more information about web server automatic scaling, refer to the section called "Configuring webserver auto scaling".

7. Under **Encryption**, choose a data encryption option:

   a. By default, Amazon MWAA uses an AWS-owned key to encrypt your data.

   b. **Optional**. Choose **Customize encryption settings (advanced)** to choose a different AWS KMS key. If you choose to specify a Customer-managed key in this step, you must specify an AWS KMS key ID or ARN. AWS KMS aliases and multi-region keys are not supported by Amazon MWAA. If you specified an Amazon S3 key for server-side encryption on your Amazon S3 bucket, you must specify the same key for your Amazon MWAA environment.

> **Note**
>
> You must have permissions to the key to select it on the Amazon MWAA console. You must also grant permissions for Amazon MWAA to use the key by attaching the policy described in Attach key policy.

8. **Recommended**. Under **Monitoring**, choose one or more log categories for **Airflow logging configuration** to send Apache Airflow logs to CloudWatch Logs:

   a. **Airflow task logs**. Choose the type of Apache Airflow task logs to send to CloudWatch Logs in **Log level**.

   b. **Airflow webserver logs**. Choose the type of Apache Airflow webserver logs to send to CloudWatch Logs in **Log level**.

   c. **Airflow scheduler logs**. Choose the type of Apache Airflow scheduler logs to send to CloudWatch Logs in **Log level**.

   d. **Airflow worker logs**. Choose the type of Apache Airflow worker logs to send to CloudWatch Logs in **Log level**.

   e. **Airflow DAG processing logs**. Choose the type of Apache Airflow DAG processing logs to send to CloudWatch Logs in **Log level**.

9. **Optional**. For **Airflow configuration options**, choose **Add custom configuration option**.

   You can choose from the suggested dropdown list of Apache Airflow configuration options for your Apache Airflow version, or specify custom configuration options. For example, `core.default_task_retries : 3`.

10. **Optional**. Under **Tags**, choose **Add new tag** to associate tags to your environment. For example, `Environment: Staging`.

11. Under **Permissions**, choose an execution role:

    a. By default, Amazon MWAA creates an execution role in **Create a new role**. You must have permission to create IAM roles to use this option.

    b. **Optional**. Choose **Enter role ARN** to enter the Amazon Resource Name (ARN) of an existing execution role.

12. **Optional**. Under **Identity Center configuration** (visible only for Apache Airflow 3.3.1 or later), configure IAM Identity Center integration:

a. Select **Enable IAM Identity Center integration**.

b. Enter the **AWS access portal URL**, found in the IAM Identity Center console under **Settings**. Example: `https://d-xxxxxxxxxx.awsapps.com/start`.

c. Enter the **Identity Store ID**, found in the IAM Identity Center console under **Settings**. Example: `d-xxxxxxxxxx`.

> **Optional configuration**
>
> This section is optional. Environments without IAM Identity Center integration continue to use the existing IAM token-based access model. IAM Identity Center integration requires an organization-level IAM Identity Center instance. For more information, see Using IAM Identity Center with Amazon Managed Workflows for Apache Airflow.

> **Irreversible action**
>
> Once IAM Identity Center integration is enabled on an environment, it cannot be disabled. Use the parallel-environment approach in Using IAM Identity Center with Amazon Managed Workflows for Apache Airflow for a safe migration path.

13. Choose **Next**.

Step three: Review and create

To review an environment summary

- Review the environment summary, choose **Create environment**.

> **Note**
>
> It takes about twenty to thirty minutes to create an environment.

What's next?

- Learn how to create an Amazon S3 bucket in Create an Amazon S3 bucket for Amazon MWAA.
- For information about using the AWS CDK to create Amazon MWAA resources, see AWS CDK MWAA module in the AWS CDK API Reference.

Managing access to an Amazon MWAA environment

Amazon Managed Workflows for Apache Airflow needs to be permitted to use other AWS services and resources used by an environment. You also need to be granted permission to access an Amazon MWAA environment and your Apache Airflow UI in AWS Identity and Access Management (IAM). This section describes the execution role used to grant access to the AWS resources for your environment and how to add permissions, and the AWS account permissions you need to access your Amazon MWAA environment and Apache Airflow UI.

Topics

- Accessing an Amazon MWAA environment
- Service-linked role for Amazon MWAA
- Amazon MWAA execution role
- Cross-service confused deputy prevention
- Apache Airflow access modes

Accessing an Amazon MWAA environment

To use Amazon Managed Workflows for Apache Airflow, you must use an account and IAM entities with the necessary permissions. This topic describes the access policies you can attach to your Apache Airflow development team and Apache Airflow users for your Amazon Managed Workflows for Apache Airflow environment.

We recommend using temporary credentials and configuring federated identities with groups and roles to access your Amazon MWAA resources. As a best practice, avoid attaching policies directly to your IAM users. Instead, define groups or roles to provide temporary access to AWS resources.

An IAM role is an IAM identity that you can create in your account that has specific permissions. An IAM role is similar to an IAM user in that it is an AWS identity with permissions policies that determine what the identity can and cannot do in AWS. However, instead of being uniquely associated with one person, a role is intended to be assumable by anyone who needs it. Also, a role does not have standard long-term credentials such as a password or access keys associated with it. Instead, when you assume a role, it provides you with temporary security credentials for your role session.

To assign permissions to a federated identity, you create a role and define permissions for the role. When a federated identity authenticates, the identity is associated with the role and is granted the permissions that are defined by the role. For information about roles for federation, see Create a role for a third-party identity provider (federation) in the *IAM User Guide*. If you use IAM Identity Center, you configure a permission set. To control what your identities can access after they authenticate, IAM Identity Center correlates the permission set to a role in IAM. For information about permissions sets, see Permission sets in the *AWS IAM Identity Center User Guide*.

You can use an IAM role in your account to grant another AWS account permissions to access your account's resources. For an example, see IAM tutorial: Delegate access across AWS accounts using IAM roles in the *IAM User Guide*.

Sections

- How it works
- Full console access policy: AmazonMWAAFullConsoleAccess
- Full API and console access policy: AmazonMWAAFullApiAccess
- Read-only console access policy: AmazonMWAAReadOnlyAccess
- Apache Airflow UI access policy: AmazonMWAAWebServerAccess
- Apache Airflow Rest API access policy: AmazonMWAARestAPIAccess
- Apache Airflow CLI policy: AmazonMWAAAirflowCliAccess
- Creating a JSON policy
- Example use case to attach policies to a developer group
- IAM Identity Center Permission Set policy examples
- What's next?

How it works

The resources and services used in an Amazon MWAA environment are not accessible to all AWS Identity and Access Management (IAM) entities. You must create a policy that grants Apache Airflow users permission to access these resources. For example, you need to grant access to your Apache Airflow development team.

Amazon MWAA uses these policies to validate whether a user has the permissions needed to perform an action on the AWS console or through the APIs used by an environment.

You can use the JSON policies in this topic to create a policy for your Apache Airflow users in IAM, and then attach the policy to a user, group, or role in IAM.

- AmazonMWAAFullConsoleAccess – Use this policy to grant permission to configure an environment on the Amazon MWAA console.
- AmazonMWAAFullApiAccess – Use this policy to grant access to all Amazon MWAA APIs used to manage an environment.
- AmazonMWAAReadOnlyAccess – Use this policy to grant access to the resources used by an environment on the Amazon MWAA console.
- AmazonMWAAWebServerAccess – Use this policy to grant access to the Apache Airflow webserver.
- AmazonMWAAAirflowCliAccess – Use this policy to grant access to run Apache Airflow CLI commands.

To provide access, add permissions to your users, groups, or roles:

- Users and groups in AWS IAM Identity Center:

  Create a permission set. Follow the instructions in Create a permission set in the *AWS IAM Identity Center User Guide*.
- Users managed in IAM through an identity provider:

  Create a role for identity federation. Follow the instructions in Create a role for a third-party identity provider (federation) in the *IAM User Guide*.
- IAM users:
  - Create a role that your user can assume. Follow the instructions in Create a role for an IAM user in the *IAM User Guide*.
  - (Not recommended) Attach a policy directly to a user or add a user to a user group. Follow the instructions in Adding permissions to a user (console) in the *IAM User Guide*.

IAM Identity Center integration

If your organization uses IAM Identity Center, you can configure Amazon MWAA to authenticate users through your corporate identity provider and authorize access using Permission Sets instead of attaching policies directly to IAM users or roles. For more

information, see Using IAM Identity Center with Amazon Managed Workflows for Apache Airflow.

Full console access policy: AmazonMWAAFullConsoleAccess

A user might need access to the `AmazonMWAAFullConsoleAccess` permissions policy if they need to configure an environment on the Amazon MWAA console.

Note

Your full console access policy must include permissions to perform `iam:PassRole`. This allows the user to pass service-linked roles, and execution roles, to Amazon MWAA. Amazon MWAA assumes each role to call other AWS services on your behalf. The following example uses the `iam:PassedToService` condition key to specify the Amazon MWAA service principal (`airflow.amazonaws.com`) as the service to which a role can be passed.
For more information about `iam:PassRole`, refer to Granting a user permissions to pass a role to an AWS service in the *IAM User Guide*.

Use the following policy if you want to create, and manage, your Amazon MWAA environments using an AWS owned key for encryption at-rest.

Using an AWS owned key

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "airflow:*",
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "iam:PassRole"
            ],
```

```
            "Resource": "*",
            "Condition": {
                "StringLike": {
                    "iam:PassedToService": "airflow.amazonaws.com"
                }
            }
        },
        {
            "Effect": "Allow",
            "Action": [
                "iam:ListRoles"
            ],
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "iam:CreatePolicy"
            ],
            "Resource": "arn:aws:iam::111122223333:policy/service-role/MWAA-
Execution-Policy*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "iam:AttachRolePolicy",
                "iam:CreateRole"
            ],
            "Resource": "arn:aws:iam::111122223333:role/service-role/AmazonMWAA*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "iam:CreateServiceLinkedRole"
            ],
            "Resource": "arn:aws:iam::*:role/aws-service-role/
airflow.amazonaws.com/AWSServiceRoleForAmazonMWAA"
        },
        {
            "Effect": "Allow",
            "Action": [
                "s3:GetBucketLocation",
                "s3:ListAllMyBuckets",
                "s3:ListBucket",
```

```
                "s3:ListBucketVersions"
            ],
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "s3:CreateBucket",
                "s3:PutObject",
                "s3:GetEncryptionConfiguration"
            ],
            "Resource": "arn:aws:s3:::*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "ec2:DescribeSecurityGroups",
                "ec2:DescribeSubnets",
                "ec2:DescribeVpcs",
                "ec2:DescribeRouteTables"
            ],
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "ec2:AuthorizeSecurityGroupIngress",
                "ec2:CreateSecurityGroup"
            ],
            "Resource": "arn:aws:ec2:*:*:security-group/airflow-security-group-*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "kms:ListAliases"
            ],
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": "ec2:CreateVpcEndpoint",
            "Resource": [
                "arn:aws:ec2:*:*:vpc-endpoint/*",
                "arn:aws:ec2:*:*:vpc/*",
```

```
                "arn:aws:ec2:*:*:subnet/*",
                "arn:aws:ec2:*:*:security-group/*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "ec2:CreateNetworkInterface"
            ],
            "Resource": [
                "arn:aws:ec2:*:*:subnet/*",
                "arn:aws:ec2:*:*:network-interface/*"
            ]
        }
    ]
}
```

Use the following policy if you want to create, and manage, your Amazon MWAA environments using a customer managed key for encryption at-rest. To use a customer managed key, the IAM principal must have permission to access AWS KMS resources using the key stored in your account.

Using a customer managed key

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "airflow:*",
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "iam:PassRole"
            ],
            "Resource": "*",
            "Condition": {
                "StringLike": {
```

```
                    "iam:PassedToService": "airflow.amazonaws.com"
                }
            }
        },
        {
            "Effect": "Allow",
            "Action": [
                "iam:ListRoles"
            ],
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "iam:CreatePolicy"
            ],
            "Resource": "arn:aws:iam::111122223333:policy/service-role/MWAA-
Execution-Policy*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "iam:AttachRolePolicy",
                "iam:CreateRole"
            ],
            "Resource": "arn:aws:iam::111122223333:role/service-role/AmazonMWAA*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "iam:CreateServiceLinkedRole"
            ],
            "Resource": "arn:aws:iam::*:role/aws-service-role/
airflow.amazonaws.com/AWSServiceRoleForAmazonMWAA"
        },
        {
            "Effect": "Allow",
            "Action": [
                "s3:GetBucketLocation",
                "s3:ListAllMyBuckets",
                "s3:ListBucket",
                "s3:ListBucketVersions"
            ],
            "Resource": "*"
```

```
        },
        {
            "Effect": "Allow",
            "Action": [
                "s3:CreateBucket",
                "s3:PutObject",
                "s3:GetEncryptionConfiguration"
            ],
            "Resource": "arn:aws:s3:::*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "ec2:DescribeSecurityGroups",
                "ec2:DescribeSubnets",
                "ec2:DescribeVpcs",
                "ec2:DescribeRouteTables"
            ],
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "ec2:AuthorizeSecurityGroupIngress",
                "ec2:CreateSecurityGroup"
            ],
            "Resource": "arn:aws:ec2:*:*:security-group/airflow-security-group-*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "kms:ListAliases"
            ],
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "kms:DescribeKey",
                "kms:ListGrants",
                "kms:CreateGrant",
                "kms:RevokeGrant",
                "kms:Decrypt",
                "kms:Encrypt",
```

```
                "kms:GenerateDataKey*",
                "kms:ReEncrypt*"
            ],
            "Resource": "arn:aws:kms:*:111122223333:key/YOUR_KMS_ID"
        },
        {
            "Effect": "Allow",
            "Action": "ec2:CreateVpcEndpoint",
            "Resource": [
                "arn:aws:ec2:*:*:vpc-endpoint/*",
                "arn:aws:ec2:*:*:vpc/*",
                "arn:aws:ec2:*:*:subnet/*",
                "arn:aws:ec2:*:*:security-group/*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "ec2:CreateNetworkInterface"
            ],
            "Resource": [
                "arn:aws:ec2:*:*:subnet/*",
                "arn:aws:ec2:*:*:network-interface/*"
            ]
        }
    ]
}
```

Full API and console access policy: AmazonMWAAFullApiAccess

A user might need access to the `AmazonMWAAFullApiAccess` permissions policy if they need access to all Amazon MWAA APIs used to manage an environment. It does not grant permissions to access the Apache Airflow UI.

> **Note**
>
> A full API access policy must include permissions to perform `iam:PassRole`. This allows the user to pass service-linked roles, and execution roles, to Amazon MWAA. Amazon MWAA assumes each role to call other AWS services on your behalf. The following example uses the `iam:PassedToService` condition key to specify the Amazon MWAA service principal (`airflow.amazonaws.com`) as the service to which a role can be passed.

For more information about `iam:PassRole`, refer to Granting a user permissions to pass a role to an AWS service in the *IAM User Guide*.

Use the following policy if you want to create, and manage, your Amazon MWAA environments using an AWS owned key for encryption at-rest.

Using an AWS owned key

JSON

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Action":"airflow:*",
         "Resource":"*"
      },
      {
         "Effect":"Allow",
         "Action":[
            "iam:PassRole"
         ],
         "Resource":"*",
         "Condition":{
            "StringLike":{
               "iam:PassedToService":"airflow.amazonaws.com"
            }
         }
      },
      {
         "Effect":"Allow",
         "Action":[
            "iam:CreateServiceLinkedRole"
         ],
         "Resource":"arn:aws:iam::*:role/aws-service-role/airflow.amazonaws.com/
AWSServiceRoleForAmazonMWAA"
      },
      {
         "Effect":"Allow",
         "Action":[
```

```
            "ec2:DescribeSecurityGroups",
            "ec2:DescribeSubnets",
            "ec2:DescribeVpcs",
            "ec2:DescribeRouteTables"
         ],
         "Resource":"*"
      },
      {
         "Effect":"Allow",
         "Action":[
            "s3:GetEncryptionConfiguration"
         ],
         "Resource":"arn:aws:s3:::*"
      },
      {
         "Effect":"Allow",
         "Action":"ec2:CreateVpcEndpoint",
         "Resource":[
            "arn:aws:ec2:*:*:vpc-endpoint/*",
            "arn:aws:ec2:*:*:vpc/*",
            "arn:aws:ec2:*:*:subnet/*",
            "arn:aws:ec2:*:*:security-group/*"
         ]
      },
      {
         "Effect":"Allow",
         "Action":[
            "ec2:CreateNetworkInterface"
         ],
         "Resource":[
            "arn:aws:ec2:*:*:subnet/*",
            "arn:aws:ec2:*:*:network-interface/*"
         ]
      }
   ]
}
```

Use the following policy if you want to create, and manage, your Amazon MWAA environments using a customer managed key for encryption at-rest. To use a customer managed key, the IAM principal must have permission to access AWS KMS resources using the key stored in your account.

Using a customer managed key

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "airflow:*",
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "iam:PassRole"
            ],
            "Resource": "*",
            "Condition": {
                "StringLike": {
                    "iam:PassedToService": "airflow.amazonaws.com"
                }
            }
        },
        {
            "Effect": "Allow",
            "Action": [
                "iam:CreateServiceLinkedRole"
            ],
            "Resource": "arn:aws:iam::*:role/aws-service-role/
airflow.amazonaws.com/AWSServiceRoleForAmazonMWAA"
        },
        {
            "Effect": "Allow",
            "Action": [
                "ec2:DescribeSecurityGroups",
                "ec2:DescribeSubnets",
                "ec2:DescribeVpcs",
                "ec2:DescribeRouteTables"
            ],
            "Resource": "*"
        },
        {
```

```
            "Effect": "Allow",
            "Action": [
                "kms:DescribeKey",
                "kms:ListGrants",
                "kms:CreateGrant",
                "kms:RevokeGrant",
                "kms:Decrypt",
                "kms:Encrypt",
                "kms:GenerateDataKey*",
                "kms:ReEncrypt*"
            ],
            "Resource": "arn:aws:kms:*:111122223333:key/YOUR_KMS_ID"
        },
        {
            "Effect": "Allow",
            "Action": [
                "s3:GetEncryptionConfiguration"
            ],
            "Resource": "arn:aws:s3:::*"
        },
        {
            "Effect": "Allow",
            "Action": "ec2:CreateVpcEndpoint",
            "Resource": [
                "arn:aws:ec2:*:*:vpc-endpoint/*",
                "arn:aws:ec2:*:*:vpc/*",
                "arn:aws:ec2:*:*:subnet/*",
                "arn:aws:ec2:*:*:security-group/*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "ec2:CreateNetworkInterface"
            ],
            "Resource": [
                "arn:aws:ec2:*:*:subnet/*",
                "arn:aws:ec2:*:*:network-interface/*"
            ]
        }
    ]
}
```

Read-only console access policy: AmazonMWAAReadOnlyAccess

A user might need access to the `AmazonMWAAReadOnlyAccess` permissions policy if they need to access the resources used by an environment on the Amazon MWAA console environment details page. It doesn't allow a user to create new environments, edit existing environments, or allow a user to access the Apache Airflow UI.

JSON

```
{
        "Version":"2012-10-17",
        "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "airflow:ListEnvironments",
                "airflow:GetEnvironment",
                "airflow:ListTagsForResource"
            ],
            "Resource": "*"
        }
    ]
}
```

Apache Airflow UI access policy: AmazonMWAAWebServerAccess

You might need access to the `AmazonMWAAWebServerAccess` permissions policy to access the Apache Airflow UI. This policy does not allow you to access environments on the Amazon MWAA console or use the Amazon MWAA APIs to perform any actions. Specify the `Admin`, `Op`, `User`, `Viewer` or the `Public` role in `{airflow-role}` to customize your level of access for the web token. For more information, see Access control on the Apache Airflow website.

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
```

```
            "Action": "airflow:CreateWebLoginToken",
            "Resource": [
             "arn:aws:airflow:us-east-1:111122223333:role/{your-environment-
name}/{airflow-role}"
            ]
        }
    ]
}
```

Note

- Amazon MWAA provides IAM integration with the five default Apache Airflow role-based access control (RBAC) roles. For more information about working with custom Apache Airflow roles, see the section called "Tutorial: Restricting users to a subset of DAGs".
- The `Resource` field in this policy can be used to specify the Apache Airflow role-based access control roles for the Amazon MWAA environment. However, it does not support the Amazon MWAA environment ARN (Amazon Resource Name) in the `Resource` field of the policy.

Apache Airflow Rest API access policy: AmazonMWAARestAPIAccess

To access the Apache Airflow REST API, you must grant the `airflow:InvokeRestApi` permission in your IAM policy. In the following policy sample, specify the `Admin`, `Op`, `User`, `Viewer` or the `Public` role in `{airflow-role}` to customize the level of user access. For more information, see Access control on the Apache Airflow website.

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Sid": "AllowMwaaRestApiAccess",
            "Effect": "Allow",
            "Action": "airflow:InvokeRestApi",
```

```
            "Resource": [
            "arn:aws:airflow:us-east-1:111122223333:role/{your-environment-name}/
{airflow-role}"
            ]
        }
    ]
}
```

Note

- For environments with a private webserver, you can invoke the `InvokeRestApi` action over the public internet. This requires the `airflow:InvokeRestApi` and `airflow:CreateWebLoginToken` permissions in your IAM policy.
- To confine `InvokeRestApi` calls to a Virtual Private Cloud (VPC), use the `aws:SourceVpc` or `aws:SourceVpce` key to apply more granular access control for this operation. For more information about this condition key, see aws:SourceVpc
- The `Resource` field in this policy can be used to specify the Apache Airflow role-based access control roles for the Amazon MWAA environment. However, it does not support the Amazon MWAA environment ARN (Amazon Resource Name) in the `Resource` field of the policy.

Apache Airflow CLI policy: AmazonMWAAAirflowCliAccess

A user might need access to the `AmazonMWAAAirflowCliAccess` permissions policy if they need to run Apache Airflow CLI commands (such as `trigger_dag`). It does not allow the user to access environments on the Amazon MWAA console or use the Amazon MWAA APIs to perform any actions.

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
```

```
                "airflow:CreateCliToken"
            ],
            "Resource": "arn:aws:airflow:us-east-1:111122223333:environment/
${EnvironmentName}"
        }
    ]
}
```

Creating a JSON policy

You can create the JSON policy, and attach the policy to your user, role, or group on the IAM console. The following steps describe how to create a JSON policy in IAM.

To create the JSON policy

1. Open the Policies page on the IAM console.
2. Choose **Create policy**.
3. Choose the **JSON** tab.
4. Add your JSON policy.
5. Choose **Review policy**.
6. Enter a value in the text field for **Name** and **Description** (optional).

   For example, you can name the policy `AmazonMWAAReadOnlyAccess`.
7. Choose **Create policy**.

Example use case to attach policies to a developer group

Let's say you're using a group in IAM named `AirflowDevelopmentGroup` to apply permissions to all of the developers on your Apache Airflow development team. These users need access to the `AmazonMWAAFullConsoleAccess`, `AmazonMWAAAirflowCliAccess`, and `AmazonMWAAWebServerAccess` permission policies. This section describes how to create a group in IAM, create and attach these policies, and associate the group to an IAM user. The steps assume you're using an AWS-owned key.

To create the AmazonMWAAFullConsoleAccess policy

1. Download the AmazonMWAAFullConsoleAccess access policy.

2. Open the Policies page on the IAM console.
3. Choose **Create policy**.
4. Choose the **JSON** tab.
5. Paste the JSON policy for `AmazonMWAAFullConsoleAccess`.
6. Substitute the following values:
    a. *123456789012* – Your AWS account ID (such as `0123456789`)
    b. *{your-kms-id}* – The unique identifer for a customer managed key, applicable only if you use a customer managed key for encryption at-rest.
7. Choose the **Review policy**.
8. Type `AmazonMWAAFullConsoleAccess` in **Name**.
9. Choose **Create policy**.

To create the AmazonMWAAWebServerAccess policy

1. Download the AmazonMWAAWebServerAccess access policy.
2. Open the Policies page on the IAM console.
3. Choose **Create policy**.
4. Choose the **JSON** tab.
5. Paste the JSON policy for `AmazonMWAAWebServerAccess`.
6. Substitute the following values:
    a. *us-east-1* – the region of your Amazon MWAA environment (such as `us-east-1`)
    b. *123456789012* – your AWS account ID (such as `0123456789`)
    c. *{your-environment-name}* – your Amazon MWAA environment name (such as `MyAirflowEnvironment`)
    d. *{airflow-role}* – the `Admin` Apache Airflow default role. For more information, see Access control on the Apache Airflow website.
7. Choose **Review policy**.
8. Type `AmazonMWAAWebServerAccess` in **Name**.
9. Choose **Create policy**.

To create the AmazonMWAAAirflowCliAccess policy

1. Download the AmazonMWAAAirflowCliAccess access policy.
2. Open the Policies page on the IAM console.
3. Choose **Create policy**.
4. Choose the **JSON** tab.
5. Paste the JSON policy for `AmazonMWAAAirflowCliAccess`.
6. Choose the **Review policy**.
7. Type `AmazonMWAAAirflowCliAccess` in **Name**.
8. Choose **Create policy**.

To create the group

1. Open the Groups page on the IAM console.
2. Enter a name of `AirflowDevelopmentGroup`.
3. Choose **Next Step**.
4. Type `AmazonMWAA` to filter results in **Filter**.
5. Select the three policies you created.
6. Choose **Next Step**.
7. Choose **Create Group**.

To associate to a user

1. Open the Users page on the IAM console.
2. Choose a user.
3. Choose **Groups**.
4. Choose **Add user to groups**.
5. Select the **AirflowDevelopmentGroup**.
6. Choose **Add to Groups**.

IAM Identity Center Permission Set policy examples

When IAM Identity Center integration is enabled on an Amazon MWAA environment, you control Apache Airflow access by attaching IAM policies to IAM Identity Center Permission Sets, not directly to IAM users or roles. The following examples show only the `airflow:AccessAirflowResource` statement. Each Permission Set policy must also include the required `identitystore` lookup statement. For details, see Configuring access with Permission Sets.

Example Grant full Apache Airflow access

```
{
  "Effect": "Allow",
  "Action": "airflow:AccessAirflowResource",
  "Resource": "arn:aws:airflow:us-east-1:123456789012:environment/my-env"
}
```

Example Restrict to DAG read access

Always include both `GET` and `LIST`. The DAG list page uses `ResourceAction=LIST`; home page dashboard widgets use `ResourceAction=GET` with no DAG ID. A policy that grants only `GET` or only `LIST` causes 403 errors on the home page or shows zero DAGs on the list page respectively.

```
{
  "Effect": "Allow",
  "Action": "airflow:AccessAirflowResource",
  "Resource": "arn:aws:airflow:us-east-1:123456789012:environment/my-env",
  "Condition": {
    "StringEquals": {
      "airflow:ResourceType": "DAG",
      "airflow:ResourceAction": ["GET", "LIST"]
    }
  }
}
```

Example Restrict to a specific DAG

Use two separate statements. The first allows `GET` and `LIST` without a `ResourceId` — this is required for the DAG list page (`LIST`) and for home page dashboard widgets (`GET` with no DAG ID). The second allows all actions on the specific DAG. Do not combine `ResourceId` with `LIST` in the same `StringEquals` block — `LIST` calls do not supply a `ResourceId`, so the condition never matches.

```
{
  "Effect": "Allow",
  "Action": "airflow:AccessAirflowResource",
  "Resource": "arn:aws:airflow:us-east-1:123456789012:environment/my-env",
  "Condition": {
    "StringEquals": {
      "airflow:ResourceType": "DAG",
      "airflow:ResourceAction": ["GET", "LIST"]
    }
  }
},
{
  "Effect": "Allow",
  "Action": "airflow:AccessAirflowResource",
  "Resource": "arn:aws:airflow:us-east-1:123456789012:environment/my-env",
  "Condition": {
    "StringEquals": {
      "airflow:ResourceType": "DAG",
      "airflow:ResourceId": "my_etl_dag"
    }
  }
}
```

Example Restrict by team using `airflow:TeamNames`

Use `ForAnyValue:StringEquals` for `airflow:TeamNames`. The single-value `StringEquals` operator does not match when a user belongs to more than one team.

```
{
  "Effect": "Allow",
  "Action": "airflow:AccessAirflowResource",
  "Resource": "arn:aws:airflow:us-east-1:123456789012:environment/my-env",
  "Condition": {
    "ForAnyValue:StringEquals": {
      "airflow:TeamNames": ["platform-team", "data-eng"]
    }
  }
}
```

Note

If the environment has teams, Apache Airflow isolates each team's resources automatically, and you don't need `airflow:TeamNames` conditions. A user can access a team's resources only if they belong to the IAM Identity Center group that matches the team name and their Permission Set policy allows the access. A policy can't give a user access to a team they don't belong to. To give a user access to a team, add the user to that team's IAM Identity Center group.

Example Restrict access to members of an IAM Identity Center group

With this policy, a user can access resources only when they belong to the `analytics` IAM Identity Center group. Use this approach for finer-grained control in an environment without teams.

```
{
  "Effect": "Allow",
  "Action": "airflow:AccessAirflowResource",
  "Resource": "arn:aws:airflow:us-east-1:123456789012:environment/my-env",
  "Condition": {
    "ForAnyValue:StringEquals": {
      "airflow:TeamNames": ["analytics"]
    }
  }
}
```

For more information about multi-team environments, see Multi-team environments.

For the full IAM Identity Center integration guide including prerequisites, setup steps, login procedure, and troubleshooting, see Using IAM Identity Center with Amazon Managed Workflows for Apache Airflow.

What's next?

- Learn how to generate a token to access the Apache Airflow UI in Accessing Apache Airflow.
- Learn more about creating IAM policies in Creating IAM policies.

Service-linked role for Amazon MWAA

Amazon Managed Workflows for Apache Airflow uses AWS Identity and Access Management (IAM) service-linked roles. A service-linked role is a unique type of IAM role that is linked directly to Amazon MWAA. Service-linked roles are predefined by Amazon MWAA and include all the permissions that the service requires to call other AWS services on your behalf.

A service-linked role makes setting up Amazon MWAA easier because you do not need to manually add the necessary permissions. Amazon MWAA defines the permissions of its service-linked roles, and unless defined otherwise, only Amazon MWAA can assume its roles. The defined permissions include the trust policy and the permissions policy, and that permissions policy cannot be attached to any other IAM entity.

You can delete a service-linked role only after first deleting their related resources. This protects your Amazon MWAA resources because you can't inadvertently remove permission to access the resources.

For information about other services that support service-linked roles, refer to AWS Services That Work with IAM and search for the services that have **Yes** in the **Service-linked roles** column. Choose a **Yes** with a link to access the service-linked role documentation for that service.

Service-linked role permissions for Amazon MWAA

Amazon MWAA uses the service-linked role named `AWSServiceRoleForAmazonMWAA` – The service-linked role created in your account grants Amazon MWAA access to the following AWS services:

- Amazon CloudWatch Logs (CloudWatch Logs) – To create log groups for Apache Airflow logs.
- Amazon CloudWatch (CloudWatch) – To publish metrics related to your environment and its underlying components to your account.
- Amazon Elastic Compute Cloud (Amazon EC2) – To create the following resources:
  - An Amazon VPC endpoint in your VPC for an AWS-managed Amazon Aurora PostgreSQL database cluster to be used by the Apache Airflow scheduler and worker.
  - An additional Amazon VPC endpoint to enable network access to the webserver if you choose the private network option for your Apache Airflow webserver.
  - Elastic Network Interfaces (ENIs) in your Amazon VPC to enable network access to AWS resources hosted in your Amazon VPC.

When you create or update an Amazon MWAA environment, Amazon MWAA performs a dry-run validation of the `ec2:CreateVpcEndpoint` permission against your IAM identity. This dry-run does not create any resources. It verifies that your identity has explicit permission to initiate VPC endpoint creation. If the dry-run fails, Amazon MWAA returns an `UnauthorizedOperation` error and the environment creation or update does not proceed.

The following trust policy allows the service principal to assume the service-linked role. The service principal for Amazon MWAA is `airflow.amazonaws.com` as demonstrated by the policy.

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Principal": {
                "Service": "airflow.amazonaws.com"
            },
            "Action": "sts:AssumeRole"
        }
    ]
}
```

The role permissions policy named `AmazonMWAAServiceRolePolicy` allows Amazon MWAA to complete the following actions on the specified resources:

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "logs:CreateLogStream",
                "logs:CreateLogGroup",
                "logs:DescribeLogGroups"
            ],
```

```
            "Resource": "arn:aws:logs:*:*:log-group:airflow-*:*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "ec2:AttachNetworkInterface",
                "ec2:CreateNetworkInterface",
                "ec2:CreateNetworkInterfacePermission",
                "ec2:DeleteNetworkInterface",
                "ec2:DeleteNetworkInterfacePermission",
                "ec2:DescribeDhcpOptions",
                "ec2:DescribeNetworkInterfaces",
                "ec2:DescribeSecurityGroups",
                "ec2:DescribeSubnets",
                "ec2:DescribeVpcEndpoints",
                "ec2:DescribeVpcs",
                "ec2:DetachNetworkInterface"
            ],
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": "ec2:CreateVpcEndpoint",
            "Resource": "arn:aws:ec2:*:*:vpc-endpoint/*",
            "Condition": {
                "ForAnyValue:StringEquals": {
                    "aws:TagKeys": "AmazonMWAAManaged"
                }
            }
        },
        {
            "Effect": "Allow",
            "Action": [
                "ec2:ModifyVpcEndpoint",
                "ec2:DeleteVpcEndpoints"
            ],
            "Resource": "arn:aws:ec2:*:*:vpc-endpoint/*",
            "Condition": {
                "Null": {
                    "aws:ResourceTag/AmazonMWAAManaged": false
                }
            }
        },
        {
```

```
            "Effect": "Allow",
            "Action": [
                "ec2:CreateVpcEndpoint",
                "ec2:ModifyVpcEndpoint"
            ],
            "Resource": [
                "arn:aws:ec2:*:*:vpc/*",
                "arn:aws:ec2:*:*:security-group/*",
                "arn:aws:ec2:*:*:subnet/*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": "ec2:CreateTags",
            "Resource": "arn:aws:ec2:*:*:vpc-endpoint/*",
            "Condition": {
                "StringEquals": {
                    "ec2:CreateAction": "CreateVpcEndpoint"
                },
                "ForAnyValue:StringEquals": {
                    "aws:TagKeys": "AmazonMWAAManaged"
                }
            }
        },
        {
            "Effect": "Allow",
            "Action": "cloudwatch:PutMetricData",
            "Resource": "*",
            "Condition": {
                "StringEquals": {
                    "cloudwatch:namespace": [
                        "AWS/MWAA"
                    ]
                }
            }
        }
    ]
}
```

You must configure permissions to allow an IAM entity (such as a user, group, or role) to create, edit, or delete a service-linked role. For more information, refer to Service-linked role permissions in the *IAM User Guide*.

We recommend scoping the `ec2:CreateVpcEndpoint` permission using the `aws:CalledVia` condition key to restrict the permission to calls made through the Amazon MWAA service:

```
{
  "Effect": "Allow",
  "Action": "ec2:CreateVpcEndpoint",
  "Resource": "arn:aws:ec2:*:*:vpc-endpoint/*",
  "Condition": {
    "ForAnyValue:StringEquals": {
      "aws:CalledVia": ["airflow.amazonaws.com"]
    }
  }
}
```

With this condition, the permission is only effective when the call originates through the Amazon MWAA service.

> **Note**
>
> If your security requirements do not allow granting `ec2:CreateVpcEndpoint` to end users, you can use customer-managed endpoints instead. For more information, see Managing your own Amazon VPC endpoints on Amazon MWAA.

Creating a service-linked role for Amazon MWAA

You don't need to manually create a service-linked role. When you create a new Amazon MWAA environment using the AWS Management Console, the AWS CLI, or the AWS API, Amazon MWAA creates the service-linked role for you.

If you delete this service-linked role, and then need to create it again, you can use the same process to recreate the role in your account. When you create another environment, Amazon MWAA creates the service-linked role for you again.

Editing a service-linked role for Amazon MWAA

Amazon MWAA does not permit editing the AWSServiceRoleForAmazonMWAA service-linked role. After you create a service-linked role, you cannot change the name of the role because various entities might reference the role. However, you can edit the description of the role using IAM. For more information, refer to Editing a service-linked role in the *IAM User Guide*.

Deleting a service-linked role for Amazon MWAA

If you no longer need to use a feature or service that requires a service-linked role, we recommend that you delete that role. That way you don't have an unused entity that is not actively monitored or maintained.

When you delete an Amazon MWAA environment, Amazon MWAA deletes all the associated resources it uses as a part of the service. However, you must wait before Amazon MWAA completes deleting your environment, before attempting to delete the service-linked role. If you delete the service-linked role before Amazon MWAA deletes the environment, Amazon MWAA might be unable to delete all of the environment's associated resources.

To manually delete the service-linked role using IAM

Use the IAM console, the AWS CLI, or the AWS API to delete the AWSServiceRoleForAmazonMWAA service-linked role. For more information, refer to Deleting a service-linked role in the *IAM User Guide*.

Supported regions for Amazon MWAA service-linked roles

Amazon MWAA supports using service-linked roles in all of the regions where the service is available. For more information, refer to Amazon Managed Workflows for Apache Airflow endpoints and quotas.

Policy updates

Change	Description	Date
Amazon MWAA update its service-linked role permission policy	`AmazonMWAAServiceRolePolicy` – Amazon MWAA updates the permission policy for its service-linked role to grant Amazon MWAA permission to publish additional metrics related to the service's underlying resources to customer accounts. These new metrics	November 18, 2022

Change	Description	Date
	are published in the `AWS/MWAA`	
Amazon MWAA started tracking changes	Amazon MWAA started tracking changes for its AWS-managed service-linked role permission policy.	November 18, 2022

Amazon MWAA execution role

An execution role is an AWS Identity and Access Management (IAM) role with a permissions policy that grants Amazon Managed Workflows for Apache Airflow permission to invoke the resources of other AWS services on your behalf. This can include resources such as your Amazon S3 bucket, AWS-owned key, and CloudWatch Logs. Amazon MWAA environments need one execution role per environment. This topic describes how to use and configure the execution role for your environment to allow Amazon MWAA to access other AWS resources used by your environment.

Contents

- Execution role overview
  - Permissions attached by default
  - How to add permission to use other AWS services
  - How to associate a new execution role
- Create a new role
- Access and update an execution role policy
  - Attach a JSON policy to use other AWS services
- Grant access to Amazon S3 bucket with account-level public access block
- Use Apache Airflow connections
- Sample JSON policies for an execution role
  - Sample policy for a customer-managed key
  - Sample policy for an AWS-owned key
- What's next?

Execution role overview

Permission for Amazon MWAA to use other AWS services used by your environment comes from the execution role. An Amazon MWAA execution role needs permission to the following AWS services used by an environment:

- Amazon CloudWatch (CloudWatch) – to send Apache Airflow metrics and logs.
- Amazon Simple Storage Service (Amazon S3) – to parse your environment's DAG code and supporting files (such as a `requirements.txt`).
- Amazon Simple Queue Service (Amazon SQS) – to queue your environment's Apache Airflow tasks in an Amazon SQS queue owned by Amazon MWAA.
- AWS Key Management Service (AWS KMS) – for your environment's data encryption (using either an AWS-owned key or your Customer-managed key).

> **Note**
>
> If you have elected for Amazon MWAA to use an AWS owned KMS key to encrypt your data, then you must define permissions in a policy attached to your Amazon MWAA execution role that grant access to arbitrary KMS keys stored outside of your account through Amazon SQS. The following two conditions are required in order for your environment's execution role to access arbitrary KMS keys:
>
> - A KMS key in a third-party account needs to allow this cross account access through its resource policy.
> - Your DAG code needs to access an Amazon SQS queue that starts with `airflow-celery-` in the third-party account and uses the same KMS key for encryption.
>
> To mitigate the risks associated with cross-account access to resources, we recommend reviewing the code placed in your DAGs to ensure that your workflows are not accessing arbitrary Amazon SQS queues outside your account. Furthermore, you can use a customer-managed KMS key stored in your own account to manage encryption on Amazon MWAA. This limits your environment's execution role to access only the KMS key in your account.
>
> Keep in mind that after you choose an encryption option, you cannot change your selection for an existing environment.

An execution role also needs permission to the following IAM actions:

- `airflow:PublishMetrics` – to allow Amazon MWAA to monitor the health of an environment.

Permissions attached by default

You can use the default options on the Amazon MWAA console to create an execution role and an AWS-owned key, then use the steps on this page to add permission policies to your execution role.

- When you choose the **Create new role** option on the console, Amazon MWAA attaches the minimal permissions needed by an environment to your execution role.
- In some cases, Amazon MWAA attaches the maximum permissions. For example, we recommend choosing the option on the Amazon MWAA console to create an execution role when you create an environment. Amazon MWAA adds the permissions policies for all CloudWatch Logs groups automatically by using the regex pattern in the execution role as "`arn:aws:logs:`*`us-east-1`*`:`*`111122223333`*`:log-group:`*`airflow-your-environment-name-*`*"".

How to add permission to use other AWS services

Amazon MWAA can't add or edit permission policies to an existing execution role after an environment is created. You must update your execution role with additional permission policies needed by your environment. For example, if your DAG requires access to AWS Glue, Amazon MWAA can't automatically detect these permissions are required by your environment, or add the permissions to your execution role.

You can add permissions to an execution role in two ways:

- By modifying the JSON policy for your execution role inline. You can use the sample JSON policy documents on this page to either add to or replace the JSON policy of your execution role on the IAM console.
- By creating a JSON policy for an AWS service and attaching it to your execution role. You can use the steps on this page to associate a new JSON policy document for an AWS service to your execution role on the IAM console.

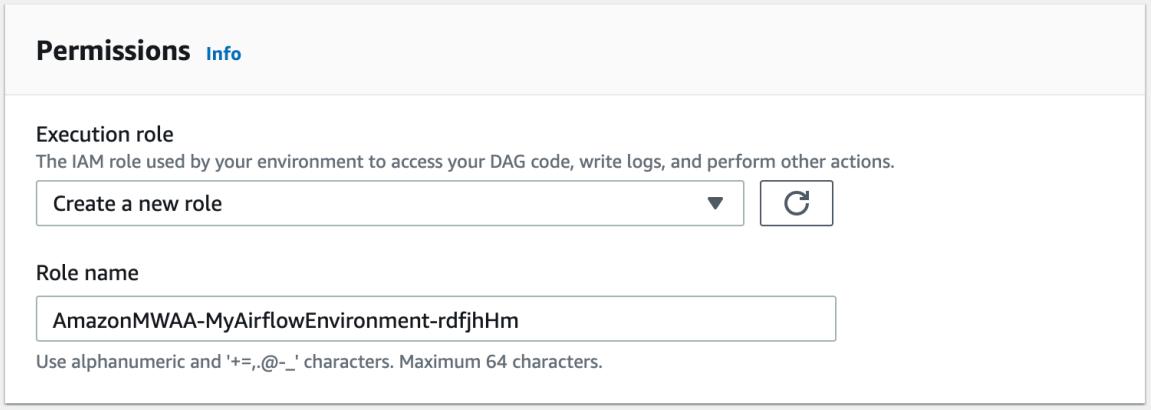
Assuming the execution role is already associated to your environment, Amazon MWAA can start using the added permission policies immediately. This also means if you remove any required permissions from an execution role, your DAGs might fail.

How to associate a new execution role

You can change the execution role for your environment at any time. If a new execution role is not already associated with your environment, use the steps on this page to create a new execution role policy, and associate the role to your environment.

Create a new role

By default, Amazon MWAA creates an AWS-owned key for data encryption and an execution role on your behalf. You can choose the default options on the Amazon MWAA console when you create an environment. The following image displays the default option to create an execution role for an environment.



Important

When you create a new execution role, do not reuse the name of a deleted execution role. Unique names can help prevent conflicts and ensure proper resource management.

Access and update an execution role policy

You can access the execution role for your environment on the Amazon MWAA console, and update the JSON policy for the role on the IAM console.

To update an execution role policy

1. Open the Environments page on the Amazon MWAA console.
2. Choose an environment.
3. Choose the execution role on the **Permissions** pane to open the permissions page in IAM.

4. Choose the execution role name to open the permissions policy.
5. Choose **Edit policy**.
6. Choose the **JSON** tab.
7. Update your JSON policy.
8. Choose **Review policy**.
9. Choose **Save changes**.

Attach a JSON policy to use other AWS services

You can create a JSON policy for an AWS service and attach it to your execution role. For example, you can attach the following JSON policy to grant read-only access to all resources in AWS Secrets Manager.

JSON

```
{
   "Version":"2012-10-17",
   "Statement":[
      {
         "Effect":"Allow",
         "Action":[
            "secretsmanager:GetResourcePolicy",
            "secretsmanager:GetSecretValue",
            "secretsmanager:DescribeSecret",
            "secretsmanager:ListSecretVersionIds"
         ],
         "Resource":[
            "*"
         ]
      }
   ]
}
```

To attach a policy to your execution role

1. Open the Environments page on the Amazon MWAA console.
2. Choose an environment.

3. Choose your execution role on the **Permissions** pane.
4. Choose **Attach policies**.
5. Choose **Create policy**.
6. Choose **JSON**.
7. Paste the JSON policy.
8. Choose **Next: Tags**, **Next: Review**.
9. Enter a descriptive name (such as `SecretsManagerReadPolicy`) and a description for the policy.
10. Choose **Create policy**.

Grant access to Amazon S3 bucket with account-level public access block

You might want to block access to all buckets in your account by using the `PutPublicAccessBlock` Amazon S3 operation. When you block access to all buckets in your account, your environment execution role must include the `s3:GetAccountPublicAccessBlock` action in a permission policy.

The following example demonstrates the policy you must attach to your execution role when blocking access to all Amazon S3 buckets in your account.

JSON

```
{
  "Version":"2012-10-17",
  "Statement": [
     {
       "Effect": "Allow",
       "Action": "s3:GetAccountPublicAccessBlock",
       "Resource": "*"
     }
  ]
}
```

For more information about restricting access to your Amazon S3 buckets, refer to Blocking public access to your Amazon S3 storage in the *Amazon Simple Storage Service User Guide*.

Use Apache Airflow connections

You can also create an Apache Airflow connection and specify your execution role and its ARN in your Apache Airflow connection object. To learn more, refer to Managing connections to Apache Airflow.

Sample JSON policies for an execution role

You can use the two sample permission policies in this section to replace the permissions policy used for your existing execution role, or to create a new execution role and use for your environment. These policies contain Resource ARN placeholders for Apache Airflow log groups, an Amazon S3 bucket, and an Amazon MWAA environment.

We recommend copying the example policy, replacing the sample ARNs or placeholders, then using the JSON policy to create or update an execution role.

Sample policy for a customer-managed key

The following example presents an execution role policy you can use for an Customer-managed key.

JSON

```
{
 "Version":"2012-10-17",
 "Statement": [
  {
   "Effect": "Allow",
   "Action": "airflow:PublishMetrics",
   "Resource": "arn:aws:airflow:us-east-1:111122223333:environment/your-
environment-name"
  },
  {
   "Effect": "Deny",
   "Action": "s3:ListAllMyBuckets",
   "Resource": [
     "arn:aws:s3:::amzn-s3-demo-bucket",
    "arn:aws:s3:::amzn-s3-demo-bucket/*"
   ]
  },
  {
```

```
    "Effect": "Allow",
    "Action": [
     "s3:GetObject*",
     "s3:GetBucket*",
     "s3:List*"
    ],
    "Resource": [
      "arn:aws:s3:::amzn-s3-demo-bucket",
     "arn:aws:s3:::amzn-s3-demo-bucket/*"
    ]
   },
   {
    "Effect": "Allow",
    "Action": [
     "logs:CreateLogStream",
     "logs:CreateLogGroup",
     "logs:PutLogEvents",
     "logs:GetLogEvents",
     "logs:GetLogRecord",
     "logs:GetLogGroupFields",
     "logs:GetQueryResults",
     "logs:DescribeLogStreams"
    ],
    "Resource": [
     "arn:aws:logs:us-east-1:111122223333:log-group:airflow-your-environment-name-
*"
    ]
   },
   {
    "Effect": "Allow",
    "Action": [
     "logs:DescribeLogGroups"
    ],
    "Resource": [
     "*"
    ]
   },
   {
    "Effect": "Allow",
    "Action": [
     "s3:GetAccountPublicAccessBlock"
    ],
    "Resource": [
     "*"
```

```
    ]
   },
   {
    "Effect": "Allow",
    "Action": "cloudwatch:PutMetricData",
    "Resource": "*"
   },
   {
    "Effect": "Allow",
    "Action": [
     "sqs:ChangeMessageVisibility",
     "sqs:DeleteMessage",
     "sqs:GetQueueAttributes",
     "sqs:GetQueueUrl",
     "sqs:ReceiveMessage",
     "sqs:SendMessage"
    ],
    "Resource": "arn:aws:sqs:us-east-1:*:airflow-celery-*"
   },
   {
    "Effect": "Allow",
    "Action": [
     "kms:Decrypt",
     "kms:DescribeKey",
     "kms:GenerateDataKey*",
     "kms:Encrypt"
    ],
    "Resource": "arn:aws:kms:us-east-1:111122223333:key/your-kms-cmk-id",
    "Condition": {
     "StringLike": {
      "kms:ViaService": [
       "sqs.us-east-1.amazonaws.com",
       "s3.us-east-1.amazonaws.com"
      ]
     }
    }
   }
  ]
}
```

Next, you need to allow Amazon MWAA to assume this role to perform actions on your behalf. This can be done by adding `"airflow.amazonaws.com"` and `"airflow-env.amazonaws.com"`

service principals to the list of trusted entities for this execution role using the IAM console, or by placing these service principals in the assume role policy document for this execution role through the IAM create-role command using the AWS CLI. Refer to the following sample assume role policy document:

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
      {
        "Effect": "Allow",
        "Principal": {
            "Service": ["airflow.amazonaws.com","airflow-env.amazonaws.com"]
        },
        "Action": "sts:AssumeRole"
      }
   ]
}
```

Then attach the following JSON policy to your Customer-managed key. This policy uses the `kms:EncryptionContext` condition key prefix to permit access to your Apache Airflow logs group in CloudWatch Logs.

JSON

```
{
  "Sid": "Allow logs access",
  "Effect": "Allow",
  "Principal": {
    "Service": "logs.us-east-1.amazonaws.com"
  },
  "Action": [
    "kms:Encrypt*",
    "kms:Decrypt*",
    "kms:ReEncrypt*",
    "kms:GenerateDataKey*",
    "kms:Describe*"
  ],
  "Resource": "*",
```

```
  "Condition": {
    "ArnLike": {
      "kms:EncryptionContext:aws:logs:arn": "arn:aws:logs:us-east-1:111122223333:*"
    }
  }
}
```

Sample policy for an AWS-owned key

The following example presents an execution role policy you can use for an AWS-owned key.

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "airflow:PublishMetrics",
            "Resource": "arn:aws:airflow:us-east-1:111122223333:environment/your-environment-name"
        },
        {
            "Effect": "Deny",
            "Action": "s3:ListAllMyBuckets",
            "Resource": [
                "arn:aws:s3:::amzn-s3-demo-bucket",
                "arn:aws:s3:::amzn-s3-demo-bucket/*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "s3:GetObject*",
                "s3:GetBucket*",
                "s3:List*"
            ],
            "Resource": [
                "arn:aws:s3:::amzn-s3-demo-bucket",
                "arn:aws:s3:::amzn-s3-demo-bucket/*"
            ]
        },
```

```
        {
            "Effect": "Allow",
            "Action": [
                "logs:CreateLogStream",
                "logs:CreateLogGroup",
                "logs:PutLogEvents",
                "logs:GetLogEvents",
                "logs:GetLogRecord",
                "logs:GetLogGroupFields",
                "logs:GetQueryResults",
                "logs:DescribeLogStreams"
            ],
            "Resource": [
            "arn:aws:logs:us-east-1:111122223333:log-group:airflow-your-
environment-name-*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "logs:DescribeLogGroups"
            ],
            "Resource": [
                "*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "s3:GetAccountPublicAccessBlock"
            ],
            "Resource": [
                "*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": "cloudwatch:PutMetricData",
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "sqs:ChangeMessageVisibility",
```

```
                "sqs:DeleteMessage",
                "sqs:GetQueueAttributes",
                "sqs:GetQueueUrl",
                "sqs:ReceiveMessage",
                "sqs:SendMessage"
            ],
            "Resource": "arn:aws:sqs:us-east-1:*:airflow-celery-*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "kms:Decrypt",
                "kms:DescribeKey",
                "kms:GenerateDataKey*",
                "kms:Encrypt"
            ],
            "NotResource": "arn:aws:kms:*:111122223333:key/*",
            "Condition": {
                "StringLike": {
                    "kms:ViaService": [
                    "sqs.us-east-1.amazonaws.com"
                    ]
                }
            }
        }
    ]
}
```

What's next?

- Learn about the required permissions you and your Apache Airflow users need to access your environment in Accessing an Amazon MWAA environment.
- Learn about Using customer-managed keys for encryption.
- Explore more Customer-managed policy examples.

Cross-service confused deputy prevention

The confused deputy problem is a security issue where an entity that doesn't have permission to perform an action can coerce a more-privileged entity to perform the action. In AWS, cross-service

impersonation can result in the confused deputy problem. Cross-service impersonation can occur when one service (the *calling service*) calls another service (the *called service*). The calling service can be manipulated to use its permissions to act on another customer's resources in a way it does not otherwise have permission to access. To prevent this, AWS provides tools that help you protect your data for all services with service principals that have been given access to resources in your account.

We recommend using the `aws:SourceArn` and `aws:SourceAccount` global condition context keys in your environment's execution role to limit the permissions that Amazon MWAA gives another service to access the resource. Use `aws:SourceArn` if you want only one resource to be associated with the cross-service access. Use `aws:SourceAccount` if you want to allow any resource in that account to be associated with the cross-service use.

The most effective way to protect against the confused deputy problem is to use the `aws:SourceArn` global condition context key with the full ARN of the resource. If you don't know the full ARN of the resource or if you are specifying multiple resources, use the `aws:SourceArn` global context condition key with wildcard characters (*) for the unknown portions of the ARN. For example, `arn:aws:airflow:*:123456789012:environment/*`.

The value of `aws:SourceArn` must be your Amazon MWAA environment ARN, for which you are creating an execution role.

Use the following example to apply the `aws:SourceArn` and `aws:SourceAccount` global condition context keys in your environment's execution role trust policy to prevent the confused deputy problem. You can use the following trust policy when you create a new execution role.

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Principal": {
                "Service": [
                    "airflow.amazonaws.com",
                    "airflow-env.amazonaws.com"
                ]
            },
            "Action": "sts:AssumeRole",
```

```
            "Condition": {
                "ArnLike": {
                    "aws:SourceArn": "arn:aws:airflow:us-
east-1:123456789012:environment/your-environment-name"
                },
                "StringEquals": {
                    "aws:SourceAccount": "123456789012"
                }
            }
        }
    ]
}
```

Apache Airflow access modes

The Amazon Managed Workflows for Apache Airflow console contains built-in options to configure private, public, or both public and private routing to the Apache Airflow webserver on your environment. This guide describes the access modes available for the Apache Airflow webserver on your Amazon Managed Workflows for Apache Airflow environment, and the additional resources you'll need to configure in your Amazon VPC if you choose the private network option.

Contents

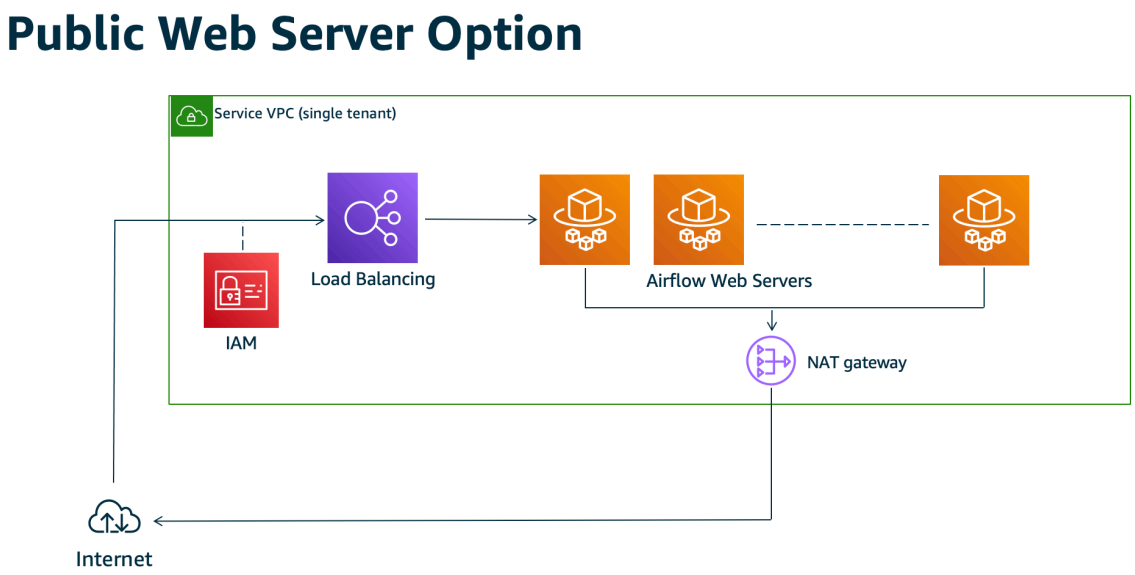
- Apache Airflow access modes
  - Public network
  - Private network
  - Both public and private network access
- Access modes overview
  - Public network access mode
  - Private network access mode
  - Both public and private network access mode
- Setup for access modes
  - Setup for public network
  - Setup for private network
  - Setup for both public and private network access
- Accessing the VPC endpoint for your Apache Airflow webserver (private network access)

Apache Airflow access modes

You can choose private, public, or both public and private routing for your Apache Airflow webserver. To enable private routing, choose **Private network**. This limits user access to an Apache Airflow webserver within an Amazon VPC. To enable public routing, choose **Public network**. This allows users to access the Apache Airflow webserver over the internet. To enable both public and private routing, choose **Both public and private network access**. This allows users to access the Apache Airflow webserver over the internet while workers communicate with the webserver through a private VPC endpoint.

Public network

The following architectural diagram depicts an Amazon MWAA environment with a public webserver.



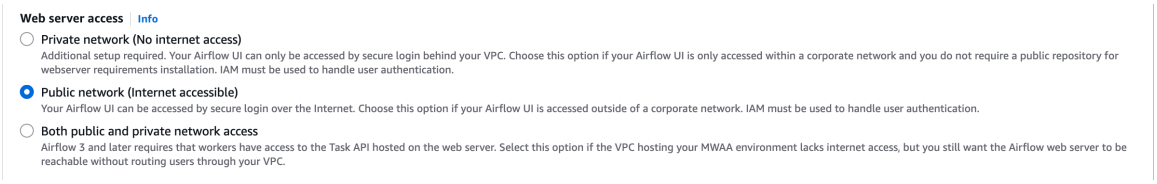
The public network access mode allows the Apache Airflow UI to be accessed over the internet by users granted access to the IAM policy for your environment.

Important

If your environment uses Apache Airflow version 3 or later with the **Public network** access mode, workers must be able to reach the webserver over the internet to communicate task state. If the subnets hosting your workers do not have internet access (for example, private subnets without a NAT gateway), DAG tasks will fail. To resolve this, upgrade to Apache

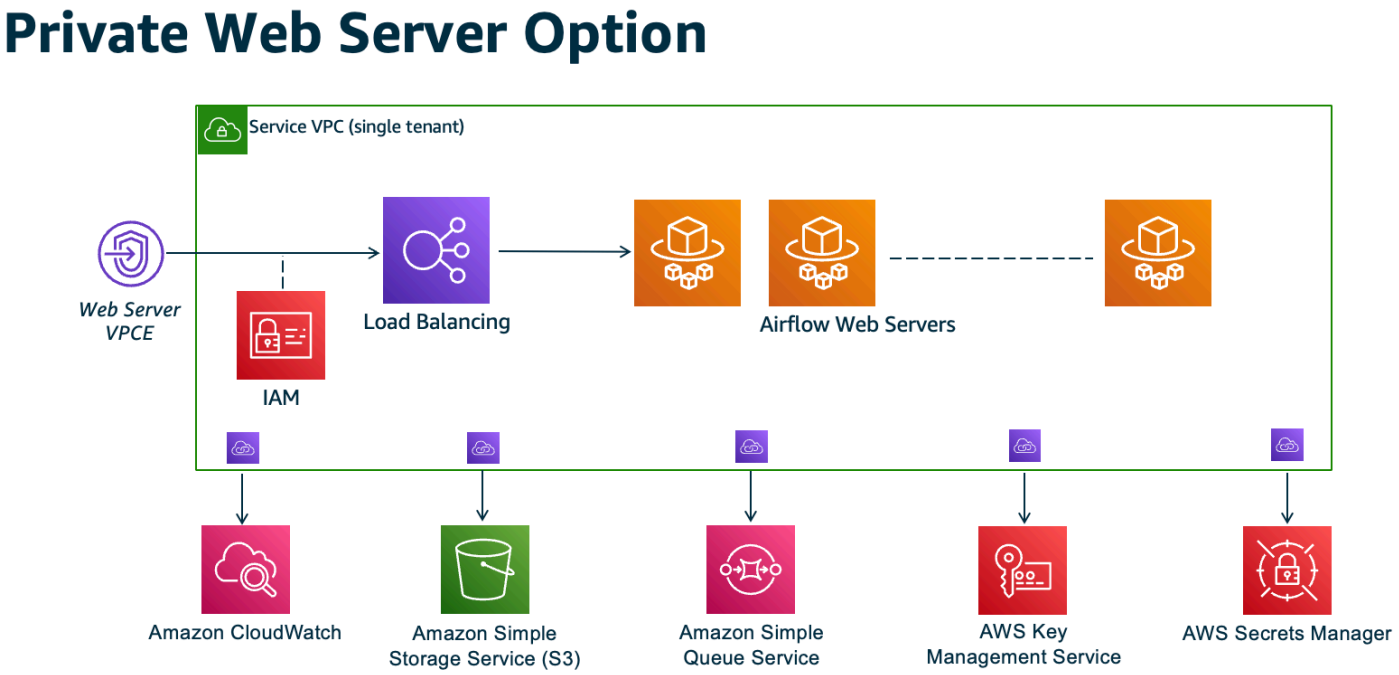
Airflow version 3.2.1 or later and switch to **Both public and private network access** mode, which routes worker communication through a private VPC endpoint.

The following image depicts where to find the **Public network** option on the Amazon MWAA console.



Private network

The following architectural diagram depicts an Amazon MWAA environment with a private webserver.

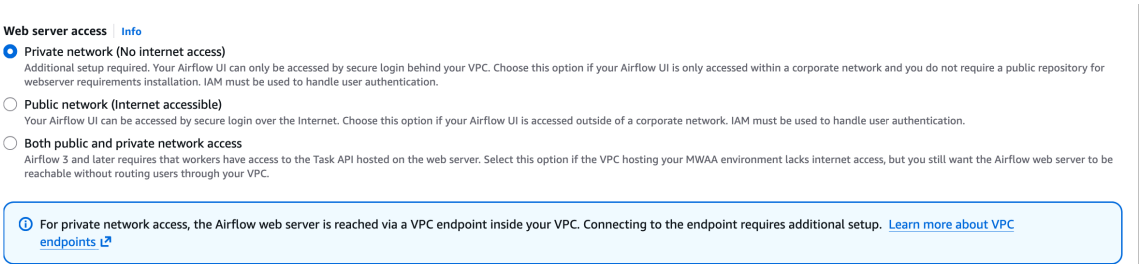


The private network access mode limits access to the Apache Airflow UI to users *within your Amazon VPC* that have been granted access to the IAM policy for your environment.

When you create an environment with private webserver access, you must package all of your dependencies in a Python wheel archive (`.whl`), then reference the `.whl` in your

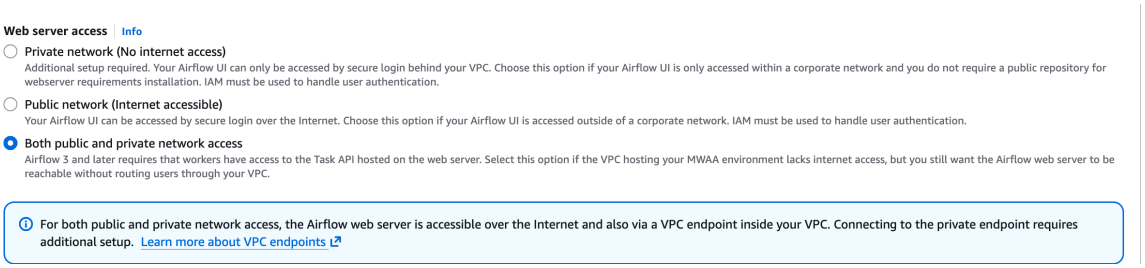
`requirements.txt`. For instructions on packaging and installing your dependencies using wheel, refer to Managing dependencies using Python wheel.

The following image depicts where to find the **Private network** option on the Amazon MWAA console.



Both public and private network access

Available for Apache Airflow version 3.2.1 and later. In Apache Airflow version 3 and later, workers communicate task state to the webserver through the Task API. If your Amazon VPC lacks internet access, workers cannot reach a public webserver, causing DAG tasks to fail. This mode creates both a public network load balancer for browser access to the Apache Airflow UI and a private VPC endpoint for worker-to-webserver communication, allowing workers to reach the webserver without internet access. Refer to the **Public network** and **Private network** architecture diagrams above for each component.



> **Note**
>
> With this mode, browser access to the Apache Airflow UI goes through the public URL. The private VPC endpoint is used by workers for internal communication and is not intended for browser access.

Access modes overview

This section describes the VPC endpoints (AWS PrivateLink) created in your Amazon VPC when you choose the **Public network**, **Private network**, or **Both public and private network access** mode.

Public network access mode

If you chose the **Public network** access mode for your Apache Airflow webserver, network traffic is publicly routed over the internet.

- Amazon MWAA creates a VPC interface endpoint for your Amazon Aurora PostgreSQL metadata database. The endpoint is created in the Availability Zones mapped to your private subnets and is independent from other AWS accounts.
- Amazon MWAA then binds an IP address from your private subnets to the interface endpoints. This is designed to support the best practice of binding a single IP from each Availability Zone of the Amazon VPC.

Private network access mode

If you chose the **Private network** access mode for your Apache Airflow webserver, network traffic is privately routed *within your Amazon VPC*.

- Amazon MWAA creates a VPC interface endpoint for your Apache Airflow webserver, and an interface endpoint for your Amazon Aurora PostgreSQL metadata database. The endpoints are created in the Availability Zones mapped to your private subnets and is independent from other AWS accounts.
- Amazon MWAA then binds an IP address from your private subnets to the interface endpoints. This is designed to support the best practice of binding a single IP from each Availability Zone of the Amazon VPC.

Both public and private network access mode

If you chose the **Both public and private network access** mode for your Apache Airflow webserver, network traffic to the Apache Airflow UI is publicly routed over the internet, while worker-to-webserver communication is privately routed within your Amazon VPC.

- Amazon MWAA creates a VPC interface endpoint for your Apache Airflow webserver (for worker connectivity), and an interface endpoint for your Amazon Aurora PostgreSQL metadata database. The endpoints are created in the Availability Zones mapped to your private subnets and are independent from other AWS accounts.

- Amazon MWAA then binds an IP address from your private subnets to the interface endpoints. This is designed to support the best practice of binding a single IP from each Availability Zone of the Amazon VPC.
- The Apache Airflow UI is accessible over the internet via a public network load balancer. Users access the UI the same way as with the **Public network** access mode.

To learn more, refer to the section called "Example use cases for an Amazon VPC and Apache Airflow access mode".

Setup for access modes

The following section describes the additional setup and configurations you'll need based on the Apache Airflow access mode you've chosen for your environment.

Setup for public network

If you choose the **Public network** option for your Apache Airflow webserver, you can begin using the Apache Airflow UI after you create your environment.

You'll need to take the following steps to configure access for your users, and permission for your environment to use other AWS services.

1. **Add permissions**. Amazon MWAA needs permission to use other AWS services. When you create an environment, Amazon MWAA creates a service-linked role that allows it to use certain IAM actions for Amazon Elastic Container Registry (Amazon ECR), CloudWatch Logs, and Amazon EC2.

   You can add permission to use additional actions for these services, or to use other AWS services by adding permissions to your execution role. To learn more, refer to Amazon MWAA execution role.
2. **Create user policies**. You might need to create multiple IAM policies for your users to configure access to your environment and Apache Airflow UI. To learn more, refer to Accessing an Amazon MWAA environment.

Setup for private network

If you choose the **Private network** option for your Apache Airflow webserver, you'll need to configure access for your users, permission for your environment to use other AWS services, and create a mechanism to access the resources in your Amazon VPC from your computer.

1. **Add permissions**. Amazon MWAA needs permission to use other AWS services. When you create an environment, Amazon MWAA creates a service-linked role that allows it to use certain IAM actions for Amazon Elastic Container Registry (Amazon ECR), CloudWatch Logs, and Amazon EC2.

   You can add permission to use additional actions for these services, or to use other AWS services by adding permissions to your execution role. To learn more, refer to Amazon MWAA execution role.
2. **Create user policies**. You might need to create multiple IAM policies for your users to configure access to your environment and Apache Airflow UI. To learn more, refer to Accessing an Amazon MWAA environment.
3. **Enable network access**. You'll need to create a mechanism in your Amazon VPC to connect to the VPC endpoint (AWS PrivateLink) for your Apache Airflow webserver. For example, by creating a VPN tunnel from your computer using an AWS Client VPN.

Setup for both public and private network access

If you choose the **Both public and private network access** option for your Apache Airflow webserver, you can begin using the Apache Airflow UI after you create your environment. No VPN or VPC endpoint access mechanism is required for browser access. The Apache Airflow UI is accessible over the internet. Workers connect to the webserver via the private VPC endpoint automatically.

You'll need to take the following steps to configure access for your users, and permission for your environment to use other AWS services.

1. **Add permissions**. Amazon MWAA needs permission to use other AWS services. When you create an environment, Amazon MWAA creates a service-linked role that allows it to use certain IAM actions for Amazon Elastic Container Registry (Amazon ECR), CloudWatch Logs, and Amazon EC2.

You can add permission to use additional actions for these services, or to use other AWS services by adding permissions to your execution role. To learn more, refer to Amazon MWAA execution role.

2. **Create user policies**. You might need to create multiple IAM policies for your users to configure access to your environment and Apache Airflow UI. To learn more, refer to Accessing an Amazon MWAA environment.

Accessing the VPC endpoint for your Apache Airflow webserver (private network access)

If you've chosen the **Private network** option, you'll need to create a mechanism in your Amazon VPC to access the VPC endpoint (AWS PrivateLink) for your Apache Airflow webserver. We recommend using the same Amazon VPC, VPC security group, and private subnets as your Amazon MWAA environment for these resources.

If you've chosen **Both public and private network access**, you do not need to create a mechanism to access the Apache Airflow UI. It is accessible over the internet. The private VPC endpoint is used automatically by workers for internal communication.

To learn more, refer to Managing access for VPC endpoints.

Accessing Apache Airflow

Amazon MWAA lets you access your Apache Airflow environment using multiple methods: the Apache Airflow user interface (UI) console, the Apache Airflow CLI, and the Apache Airflow REST API. You can use the Amazon MWAA console to access and invoke a DAG in your Apache Airflow UI, or use Amazon MWAA APIs to get a token and invoke a DAG. This section describes the permissions needed to access the Apache Airflow UI, how to generate a token to make Amazon MWAA API calls directly in your command shell, and the supported commands in the Apache Airflow CLI.

Topics

- Prerequisites
- Open the Apache Airflow UI
- Log in to Apache Airflow
- Log in using IAM Identity Center
- Create a Apache Airflow webserver access token
- Setting up a custom domain for the Apache Airflow webserver
- Creating an Apache Airflow CLI token
- Using the Apache Airflow REST API
- Apache Airflow CLI command reference

Prerequisites

The following section describes the preliminary steps required to use the commands and scripts in this section.

Access

- AWS account access in AWS Identity and Access Management (IAM) to the Amazon MWAA permissions policy in Apache Airflow UI access policy: AmazonMWAAWebServerAccess.
- AWS account access in AWS Identity and Access Management (IAM) to the Amazon MWAA permissions policy Full API and console access policy: AmazonMWAAFullApiAccess.

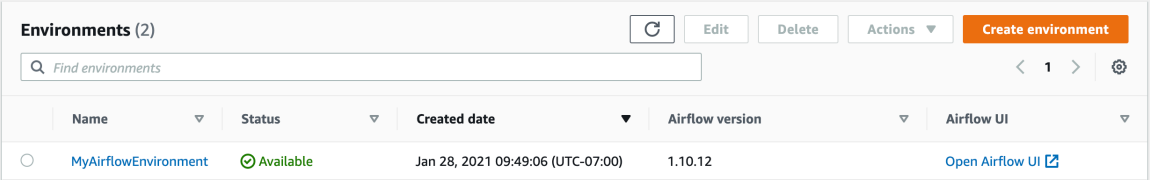
AWS CLI

The AWS Command Line Interface (AWS CLI) is an open source tool that you can use to interact with AWS services using commands in your command-line shell. To complete the steps on this page, you need the following:

- AWS CLI – Install version 2.
- AWS CLI – Quick configuration with `aws configure`.

Open the Apache Airflow UI

The following image displays the link to your Apache Airflow UI on the Amazon MWAA console.



Log in to Apache Airflow

You need Apache Airflow UI access policy: AmazonMWAAWebServerAccess permissions for your AWS account in AWS Identity and Access Management (IAM) to access your Apache Airflow UI.

To access your Apache Airflow UI

1. Open the Environments page on the Amazon MWAA console.
2. Choose an environment.
3. Choose **Open Airflow UI**.

Log in using IAM Identity Center

If your environment has IAM Identity Center integration enabled on Apache Airflow 3.3.1 or later, use the IAM Identity Center device code flow to log in. Amazon MWAA redirects you to a device authorization page where you authenticate with your corporate identity provider in a separate browser tab, then enter a one-time code to complete sign-in.

For instructions on signing in using IAM Identity Center, see Using IAM Identity Center with Amazon Managed Workflows for Apache Airflow.

Create a Apache Airflow webserver access token

You can use the commands on this page to create a web server access token. An access token provides access to your Amazon MWAA environment. For example, you can get a token, then deploy DAGs programmatically using Amazon MWAA APIs. The following section includes the steps to create an Apache Airflow web login token using the AWS CLI, a bash script, a POST API request, or a Python script. The token returned in the response is valid for 60 seconds.

Important

Effective August 19, 2025, Amazon MWAA added support for IPv6 endpoints, and now supports IPv4 and IPv6 endpoints. As of this date, all newly created environments will use `.on.aws` domains for the Airflow user interface (UI). Customers must migrate their Airflow UI from `.amazonaws.com` to `.on.aws` domains for these newly created environments. Virtual Private Cloud (VPC) endpoint services for webserver and database will maintain their current `.amazonaws.com` domains with no required changes.

Contents

- Prerequisites
  - Access
  - AWS CLI
- Using the AWS CLI
- Using a bash script
- Using a Python script
- What's next?

Prerequisites

The following section describes the preliminary steps required to use the commands and scripts on this page.

Access

- AWS account access in AWS Identity and Access Management (IAM) to the Amazon MWAA permissions policy in Apache Airflow UI access policy: AmazonMWAAWebServerAccess.

- AWS account access in AWS Identity and Access Management (IAM) to the Amazon MWAA permissions policy Full API and console access policy: AmazonMWAAFullApiAccess.

AWS CLI

The AWS Command Line Interface (AWS CLI) is an open source tool that you can use to interact with AWS services using commands in your command-line shell. To complete the steps on this page, you need the following:

- AWS CLI – Install version 2.
- AWS CLI – Quick configuration with `aws configure`.

Using the AWS CLI

The following example uses the create-web-login-token command in the AWS CLI to create an Apache Airflow web login token.

```
aws mwaa create-web-login-token --name YOUR_ENVIRONMENT_NAME
```

Using a bash script

The following example uses a bash script to call the create-web-login-token command in the AWS CLI to create an Apache Airflow web login token.

1. Copy the contents of the following code sample and save locally as `get-web-token.sh`.

```
#!/bin/bash
HOST=YOUR_HOST_NAME
YOUR_URL=https://$HOST/aws_mwaa/aws-console-sso?login=true#
WEB_TOKEN=$(aws mwaa create-web-login-token --name YOUR_ENVIRONMENT_NAME --query
 WebToken --output text)
echo $YOUR_URL$WEB_TOKEN
```

2. Substitute the placeholders in *red* for YOUR_HOST_NAME and YOUR_ENVIRONMENT_NAME. For example, a host name for a public network resembles (without the *https://*):

```
123456a0-0101-2020-9e11-1b159eec9000.c2.us-east-1.airflow.amazonaws.com
```

3. (optional) macOS and Linux users might need to run the following command to ensure the script is executable.

```
chmod +x get-web-token.sh
```

4. Run the following script to get a web login token.

```
./get-web-token.sh
```

Your command prompt displays:

```
https://123456a0-0101-2020-9e11-1b159eec9000.c2.us-east-1.airflow.amazonaws.com/
aws_mwaa/aws-console-sso?login=true#{your-web-login-token}
```

Using a Python script

The following example uses the boto3 create_web_login_token method in a Python script to create an Apache Airflow web login token. You can run this script outside of Amazon MWAA. The only thing you need to do is install the boto3 library. You might want to create a virtual environment to install the library. It assumes you have configured AWS authentication credentials for your account.

1. Copy the contents of the following code sample and save locally as `create-web-login-token.py`.

```
import boto3
  mwaa = boto3.client('mwaa')
  response = mwaa.create_web_login_token(
    Name="YOUR_ENVIRONMENT_NAME"
  )
  webServerHostName = response["WebServerHostname"]
  webToken = response["WebToken"]
  airflowUIUrl = 'https://{0}/aws_mwaa/aws-console-sso?
login=true#{1}'.format(webServerHostName, webToken)
  print("Here is your Airflow UI URL: ")
  print(airflowUIUrl)
```

2. Substitute the placeholder in *red* for YOUR_ENVIRONMENT_NAME.
3. Run the following script to get a web login token.

```
python3 create-web-login-token.py
```

What's next?

- Explore the Amazon MWAA API operation used to create a web login token at CreateWebLoginToken.

Setting up a custom domain for the Apache Airflow webserver

Amazon Managed Workflows for Apache Airflow (Amazon MWAA) lets you set up a custom domain for the managed Apache Airflow webserver. Using a custom domain, you can access your environment's Amazon MWAA-managed Apache Airflow webserver using the Apache Airflow UI, the Apache Airflow CLI, or the Apache Airflow webserver.

> **Note**
>
> You can only use custom domain with a private webserver without internet access.

Use cases for a custom domain on Amazon MWAA

1. Share the webserver domain across your cloud application on AWS — Using a custom domain lets you define a user-friendly URL for accessing the webserver, instead of the generated service domain name. You can store this custom domain and share it as an environment variable in your applications.
2. Access a private webserver — If you want to configure access for a webserver in a VPC with no internet access, using a custom domain simplifies the URL redirection work flow.

Topics

- Configure the custom domain
- Set up the networking infrastructure

Configure the custom domain

To configure the custom domain feature, you need to provide the custom domain value through the `webserver.base_url` Apache Airflow configuration when creating or updating your Amazon MWAA environment. The following constraints apply to your custom domain name:

- The value must be a fully qualified domain name (FQDN) without any protocol or path. For example, `your-custom-domain.com`.
- Amazon MWAA does not allow a path in the URL. For example, `your-custom-domain.com/dags/` is not a valid custom domain name.
- The URL length is limited to 255 ASCII characters.
- If you provide an empty string, by default, the environment will be created with a webserver URL generated by Amazon MWAA.

Use the following example to create an environment with a custom webserver domain name using the AWS CLI.

```
aws mwaa create-environment \
--name my-mwaa-env \
--source-bucket-arn arn:aws:s3:::amzn-s3-demo-bucket \
--airflow-configuration-options '{"webserver.base_url":"my-custom-domain.com"}' \
--network-configuration '{"SubnetIds":["subnet-0123456789abcdef","subnet-
fedcba9876543210"]}' \
--execution-role-arn arn:aws:iam::123456789012:role/my-execution-role
```

After the environment is created or updated, you need to set up the networking infrastructure in your AWS account to access the private webserver through the custom domain.

To revert back to the default service-generated URL, update your private environment and remove the `webserver.base_url` configuration option.

Set up the networking infrastructure

Use the following steps to set up the required networking infrastructure to use with your custom domain in your AWS account.

1. Get the IP addresses for the Amazon VPC Endpoint Network Interfaces (ENI). To do this, first, use `get-environment` to find the `WebserverVpcEndpointService` for your environment.

```
aws mwaa get-environment --name your-environment-name
```

If successful, you'll get output similar to the following.

```
{
  "Environment": {
    "AirflowConfigurationOptions": {},
    "AirflowVersion": "latest-version",
    "Arn": "environment-arn",
    "CreatedAt": "2024-06-01T01:00:00-00:00",
    "DagS3Path": "dags",
    .
    .
    .
    "WebserverVpcEndpointService": "web-server-vpc-endpoint-service",
    "WeeklyMaintenanceWindowStart": "TUE:21:30"
  }
}
```

Note the `WebserverVpcEndpointService` value and use it for `web-server-vpc-endpoint-service` in the following Amazon EC2 `describe-vpc-endpoints` command. `--filters Name=service-name,Values=`*`web-server-vpc-endpoint-service-id`* in the following command.

2. Retrieve the Amazon VPC endpoint details. This command fetches details about Amazon VPC endpoints that match a specific service name, returning the endpoint ID and associated network interface IDs in a text format.

```
aws ec2 describe-vpc-endpoints \
 --filters Name=service-name,Values=web-server-vpc-endpoint-service \
 --query 'VpcEndpoints[*].
{EndpointId:VpcEndpointId,NetworkInterfaceIds:NetworkInterfaceIds}' \
 --output text
```

3. Get the network interface details. This command retrieves private IP addresses for each network interface associated with the Amazon VPC endpoints identified in the previous step.

```
for eni_id in $(
  aws ec2 describe-vpc-endpoints \
   --filters Name=service-name,Values=service-id \
```

```
   --query 'VpcEndpoints[*].NetworkInterfaceIds' \
   --output text
 ); do
 aws ec2 describe-network-interfaces \
  --network-interface-ids $eni_id \
  --query 'NetworkInterfaces[*].PrivateIpAddresses[*].PrivateIpAddress' \
  --output text
      done
```

4. Use `create-target-group` to create a new target group. You will use this target group to register the IP addresses for your webserver Amazon VPC endpoints.

```
aws elbv2 create-target-group \
--name new-target-group-namne \
--protocol HTTPS \
--port 443 \
--vpc-id web-server-vpc-id \
--target-type ip \
--health-check-protocol HTTPS \
--health-check-port 443 \
--health-check-path / \
--health-check-enabled \
--matcher 'HttpCode="200,302"'
```

   Register the IP addresses using the `register-targets` command.

```
aws elbv2 register-targets \
--target-group-arn target-group-arn \
--targets Id=ip-address-1 Id=ip-address-2
```

5. Request an ACM certificate. Skip this step if you are using an existing certificate.

```
aws acm request-certificate \
--domain-name my-custom-domain.com \
--validation-method DNS
```

6. Configure an Application Load Balancer. First, create the load balancer, then create a listener for the load balancer. Specify the ACM certificate you created in the previous step.

```
aws elbv2 create-load-balancer \
--name my-mwaa-lb \
--type application \
```

```
--subnets subnet-id-1 subnet-id-2
```

```
aws elbv2 create-listener \
--load-balancer-arn load-balancer-arn \
--protocol HTTPS \
--port 443 \
--ssl-policy ELBSecurityPolicy-2016-08 \
--certificates CertificateArn=acm-certificate-arn \
--default-actions Type=forward,TargetGroupArn=target-group-arn
```

If you use a Network Load Balancer in a private subnet, set up a bastion host or Site-to-Site VPN tunnel to access the webserver.

7. Create a hosted zone using Route 53 for the domain.

```
aws route53 create-hosted-zone --name my-custom-domain.com \
--caller-reference 1
```

Create an A record for the domain. To do this using the AWS CLI, get the hosted zone ID using `list-hosted-zones-by-name` then apply the record with `change-resource-record-sets`.

```
HOSTED_ZONE_ID=$(aws route53 list-hosted-zones-by-name \
--dns-name my-custom-domain.com \
--query 'HostedZones[0].Id' --output text)
```

```
aws route53 change-resource-record-sets \
--hosted-zone-id $HOSTED_ZONE_ID \
--change-batch '{
  "Changes": [
    {
      "Action": "CREATE",
      "ResourceRecordSet": {
        "Name": "my-custom-domain.com",
        "Type": "A",
        "AliasTarget": {
          "HostedZoneId": "load-balancer-hosted-zone-id>",
          "DNSName": "load-balancer-dns-name",
          "EvaluateTargetHealth": true
        }
      }
```

```
    }
  ]
}'
```

8. Update the security group rules for the webserver Amazon VPC endpoint to follow the principle of least privilege by allowing HTTPS traffic only from the public subnets where the Application Load Balancer is located. Save the following JSON locally. For example, as `sg-ingress-ip-permissions.json`.

```
[
  {
    "IpProtocol": "tcp",
    "FromPort": 443,
    "ToPort": 443,
    "UserIdGroupPairs": [
      {
        "GroupId": "load-balancer-security-group-id"
      }
    ],
    "IpRanges": [
      {
        "CidrIp": "public-subnet-1-cidr"
      },
      {
        "CidrIp": "public-subnet-2-cidr"
      }
    ]
  }
]
```

Run the following Amazon EC2 command to update your ingress security group rules. Specify the JSON file for `--ip-permissions`.

```
aws ec2 authorize-security-group-ingress \
--group-id <security-group-id> \
--ip-permissions file://sg-ingress-ip-permissions.json
```

Run the following Amazon EC2 command to update your egress rules.

```
aws ec2 authorize-security-group-egress \
--group-id webserver-vpc-endpoint-security-group-id \
```

```
--protocol tcp \
--port 443 \
--source-group load-balancer-security-group-id
```

Open the Amazon MWAA console and navigate to the Apache Airflow UI. If you are setting up an Network Load Balancer in a private subnet instead of the Application Load Balancer used here, you must access the webserver with one of the following options.

- the section called “Tutorial: Linux Bastion Host”
- the section called “Tutorial: AWS Client VPN”

Creating an Apache Airflow CLI token

> **Tip**
>
> REST API is more modern than the CLI and is designed for programmatic integration with external systems. REST is the preferred way of interacting with Apache Airflow.

You can use the commands on this page to generate a CLI token, and then make Amazon Managed Workflows for Apache Airflow API calls directly in your command shell. For example, you can get a token, then deploy DAGs programmatically using Amazon MWAA APIs. The following section includes the steps to create an Apache Airflow CLI token using the AWS CLI, a curl script, a Python script, or a bash script. The token returned in the response is valid for 60 seconds.

The AWS CLI token is intended as a replacement for synchronous shell actions, not asynchronous API commands. As such, available concurrency is limited. To ensure that the webserver remains responsive for users, we recommend not opening a new AWS CLI request until the previous one completes successfully.

Contents

- Prerequisites
  - Access
  - AWS CLI
- Using the AWS CLI
- Using a curl script

- Using a bash script
- Using a Python script
- What's next?

Prerequisites

The following section describes the preliminary steps required to use the commands and scripts on this page.

Access

- AWS account access in AWS Identity and Access Management (IAM) to the Amazon MWAA permissions policy in Apache Airflow UI access policy: AmazonMWAAWebServerAccess.
- AWS account access in AWS Identity and Access Management (IAM) to the Amazon MWAA permissions policy Full API and console access policy: AmazonMWAAFullApiAccess.

AWS CLI

The AWS Command Line Interface (AWS CLI) is an open source tool that you can use to interact with AWS services using commands in your command-line shell. To complete the steps on this page, you need the following:

- AWS CLI – Install version 2.
- AWS CLI – Quick configuration with `aws configure`.

Using the AWS CLI

The following example uses the create-cli-token command in the AWS CLI to create an Apache Airflow CLI token.

```
aws mwaa create-cli-token --name YOUR_ENVIRONMENT_NAME
```

Using a curl script

Use the following curl script to call the create-cli-token command in the AWS CLI to invoke the Apache Airflow CLI through an endpoint on the Apache Airflow webserver and trigger a DAG.

Apache Airflow v3

1. Copy the curl statement from your text file and paste it in your command shell.

   > **Note**
   >
   > After copying it to your clipboard, you might need to use **Edit > Paste** from your shell menu.

   ```
   CLI_JSON=$(aws mwaa --region us-east-1 create-cli-token --
   name YOUR_ENVIRONMENT_NAME) \
   && CLI_TOKEN=$(echo $CLI_JSON | jq -r '.CliToken') \
   && WEB_SERVER_HOSTNAME=$(echo $CLI_JSON | jq -r '.WebServerHostname') \
   && CLI_RESULTS=$(curl -L --request POST "https://$WEB_SERVER_HOSTNAME/aws_mwaa/
   cli" \
   --header "Authorization: Bearer $CLI_TOKEN" \
   --header "Content-Type: text/plain" \
   --data-raw "dags trigger YOUR_DAG_NAME --logical-date $(date -u +"%Y-%m-%dT%H:
   %M:%SZ")") \
   && echo "Output:" \
   && echo $CLI_RESULTS | jq -r '.stdout' | base64 --decode \
   && echo "Errors:" \
   && echo $CLI_RESULTS | jq -r '.stderr' | base64 --decode
   ```

2. Replace *YOUR_DAG_NAME* with the name of your DAG, *YOUR_ENVIRONMENT_NAME* with your environment name, and the AWS Region with your environment's Region. For example, a host name for a public network resembles (without the *https://*):

   ```
   123456a0-0101-2020-9e11-1b159eec9000.c2.us-east-1.airflow.amazonaws.com
   ```

   Your command prompt displays:

   ```
   {
     "stderr":"<STDERR of the CLI execution (if any), base64 encoded>",
     "stdout":"<STDOUT of the CLI execution, base64 encoded>"
   }
   ```

Apache Airflow v2

1. Copy the curl statement from your text file and paste it in your command shell.

   > **Note**
   >
   > After copying it to your clipboard, you might need to use **Edit > Paste** from your shell menu.

   ```
   CLI_JSON=$(aws mwaa --region us-east-1 create-cli-token --
   name YOUR_ENVIRONMENT_NAME) \
   && CLI_TOKEN=$(echo $CLI_JSON | jq -r '.CliToken') \
   && WEB_SERVER_HOSTNAME=$(echo $CLI_JSON | jq -r '.WebServerHostname') \
   && CLI_RESULTS=$(curl --request POST "https://$WEB_SERVER_HOSTNAME/aws_mwaa/cli"
    \
   --header "Authorization: Bearer $CLI_TOKEN" \
   --header "Content-Type: text/plain" \
   --data-raw "dags trigger YOUR_DAG_NAME") \
   && echo "Output:" \
   && echo $CLI_RESULTS | jq -r '.stdout' | base64 --decode \
   && echo "Errors:" \
   && echo $CLI_RESULTS | jq -r '.stderr' | base64 --decode
   ```

2. Replace *YOUR_DAG_NAME* with the name of your DAG, *YOUR_ENVIRONMENT_NAME* with your environment name, and the AWS Region with your environment's Region. For example, a host name for a public network resembles (without the *https://*):

   ```
   123456a0-0101-2020-9e11-1b159eec9000.c2.us-east-1.airflow.amazonaws.com
   ```

   Your command prompt displays:

   ```
   {
     "stderr":"<STDERR of the CLI execution (if any), base64 encoded>",
     "stdout":"<STDOUT of the CLI execution, base64 encoded>"
   }
   ```

Using a bash script

Use the following bash script to call the create-cli-token command in the AWS CLI, create an Apache Airflow CLI token, and trigger a DAG.

Apache Airflow v3

1. Copy the contents of the following code sample and save locally as `get-cli-token.sh`.

```
# brew install jq
CLI_TOKEN=$(aws mwaa create-cli-token --name YOUR_ENVIRONMENT_NAME | jq -r
 '.CliToken') \
&& curl -L --request POST "https://YOUR_HOST_NAME/aws_mwaa/cli" \
--header "Authorization: Bearer $CLI_TOKEN" \
--header "Content-Type: text/plain" \
--data-raw "dags trigger YOUR_DAG_NAME --logical-date $(date -u +"%Y-%m-%dT%H:
%M:%SZ")"
```

2. Replace *YOUR_ENVIRONMENT_NAME* with your environment name, *YOUR_HOST_NAME* with your webserver hostname, and *YOUR_DAG_NAME* with the name of your DAG. For example, a host name for a public network resembles (without the *https://*):

```
123456a0-0101-2020-9e11-1b159eec9000.c2.us-east-1.airflow.amazonaws.com
```

3. (optional) macOS and Linux users might need to run the following command to ensure the script is executable.

```
chmod +x get-cli-token.sh
```

4. Run the following script to create an Apache Airflow CLI token.

```
./get-cli-token.sh
```

Apache Airflow v2

1. Copy the contents of the following code sample and save locally as `get-cli-token.sh`.

```
# brew install jq
CLI_TOKEN=$(aws mwaa create-cli-token --name YOUR_ENVIRONMENT_NAME | jq -r
 '.CliToken') \
```

```
&& curl --request POST "https://YOUR_HOST_NAME/aws_mwaa/cli" \
--header "Authorization: Bearer $CLI_TOKEN" \
--header "Content-Type: text/plain" \
--data-raw "dags trigger YOUR_DAG_NAME"
```

2. Replace *YOUR_ENVIRONMENT_NAME* with your environment name, *YOUR_HOST_NAME* with your webserver hostname, and *YOUR_DAG_NAME* with the name of your DAG. For example, a host name for a public network resembles (without the *https://*):

```
123456a0-0101-2020-9e11-1b159eec9000.c2.us-east-1.airflow.amazonaws.com
```

3. (optional) macOS and Linux users can run the following command to ensure the script is executable.

```
chmod +x get-cli-token.sh
```

4. Run the following script to create an Apache Airflow CLI token.

```
./get-cli-token.sh
```

Using a Python script

The following example uses the boto3 create_cli_token method in a Python script to create an Apache Airflow CLI token and trigger a DAG. You can run this script outside of Amazon MWAA. The only thing you need to do is install the boto3 library. You might want to create a virtual environment to install the library. It assumes you have configured AWS authentication credentials for your account.

Apache Airflow v3

1. Copy the contents of the following code sample and save locally as `create-cli-token.py`.

```
"""
Copyright Amazon.com, Inc. or its affiliates. All Rights Reserved.

Permission is hereby granted, free of charge, to any person obtaining a copy of
this software and associated documentation files (the "Software"), to deal in
the Software without restriction, including without limitation the rights to
use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of
```

```
the Software, and to permit persons to whom the Software is furnished to do so.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR
IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS
FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR
COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER
IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN
CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.
"""
import boto3
import json
import requests
import base64

mwaa_env_name = 'YOUR_ENVIRONMENT_NAME'
dag_name = 'YOUR_DAG_NAME'
mwaa_cli_command = 'dags trigger'

client = boto3.client('mwaa')

mwaa_cli_token = client.create_cli_token(
    Name=mwaa_env_name
)

mwaa_auth_token = 'Bearer ' + mwaa_cli_token['CliToken']
mwaa_webserver_hostname = 'https://{0}/aws_mwaa/
cli'.format(mwaa_cli_token['WebServerHostname'])
raw_data = '{0} {1}'.format(mwaa_cli_command, dag_name)

mwaa_response = requests.post(
        mwaa_webserver_hostname,
        headers={
            'Authorization': mwaa_auth_token,
            'Content-Type': 'text/plain'
            },
        data=raw_data
        )

mwaa_std_err_message = base64.b64decode(mwaa_response.json()
['stderr']).decode('utf8')
mwaa_std_out_message = base64.b64decode(mwaa_response.json()
['stdout']).decode('utf8')

print(mwaa_response.status_code)
```

```
print(mwaa_std_err_message)
print(mwaa_std_out_message)
```

2. Substitute the placeholders for YOUR_ENVIRONMENT_NAME and YOUR_DAG_NAME.
3. Run the following script to create an Apache Airflow CLI token.

```
python3 create-cli-token.py
```

Apache Airflow v2

1. Copy the contents of the following code sample and save locally as `create-cli-token.py`.

```
"""
Copyright Amazon.com, Inc. or its affiliates. All Rights Reserved.

Permission is hereby granted, free of charge, to any person obtaining a copy of
this software and associated documentation files (the "Software"), to deal in
the Software without restriction, including without limitation the rights to
use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of
the Software, and to permit persons to whom the Software is furnished to do so.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR
IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS
FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR
COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER
IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN
CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.
"""
import boto3
import json
import requests
import base64

mwaa_env_name = 'YOUR_ENVIRONMENT_NAME'
dag_name = 'YOUR_DAG_NAME'
mwaa_cli_command = 'dags trigger'

client = boto3.client('mwaa')

mwaa_cli_token = client.create_cli_token(
```

```
    Name=mwaa_env_name
)

mwaa_auth_token = 'Bearer ' + mwaa_cli_token['CliToken']
mwaa_webserver_hostname = 'https://{0}/aws_mwaa/
cli'.format(mwaa_cli_token['WebServerHostname'])
raw_data = '{0} {1}'.format(mwaa_cli_command, dag_name)

mwaa_response = requests.post(
        mwaa_webserver_hostname,
        headers={
            'Authorization': mwaa_auth_token,
            'Content-Type': 'text/plain'
            },
        data=raw_data
        )

mwaa_std_err_message = base64.b64decode(mwaa_response.json()
['stderr']).decode('utf8')
mwaa_std_out_message = base64.b64decode(mwaa_response.json()
['stdout']).decode('utf8')

print(mwaa_response.status_code)
print(mwaa_std_err_message)
print(mwaa_std_out_message)
```

2. Substitute the placeholders for YOUR_ENVIRONMENT_NAME and YOUR_DAG_NAME.
3. Run the following script to create an Apache Airflow CLI token.

```
python3 create-cli-token.py
```

What's next?

- Explore the Amazon MWAA API operation used to create a CLI token at CreateCliToken.

Using the Apache Airflow REST API

Amazon Managed Workflows for Apache Airflow (Amazon MWAA) supports interacting with your Apache Airflow environments directly using the Apache Airflow REST API for environments running Apache Airflow v2.4.3 and later. This lets you access and manage your Amazon MWAA

environments programmatically, providing a standardized way to invoke data orchestration workflows, manage your DAGs, and monitor the status of various Apache Airflow components such as the metadata database, triggerer, and scheduler.

To support scalability while using the Apache Airflow REST API, Amazon MWAA provides you with the option to horizontally scale webserver capacity to handle increased demand, whether from REST API requests, command line interface (CLI) usage, or more concurrent Apache Airflow user interface (UI) users. For more information about how Amazon MWAA scales webservers, refer to the section called “Configuring webserver auto scaling”.

You can use the Apache Airflow REST API to implement the following use cases for your environments:

- **Programmatic access** – You can now start Apache Airflow DAG runs, manage datasets, and retrieve the status of various components such as the metadata database, triggerers, and schedulers without relying on the Apache Airflow UI or CLI.
- **Integrate with external applications and microservices** – REST API support you can use to build custom solutions that integrate your Amazon MWAA environments with other systems. For example, you can start workflows in response to events from external systems, such as completed database jobs or new user sign-ups.
- **Centralized monitoring** – You can build monitoring dashboards that aggregate the status of your DAGs across multiple Amazon MWAA environments, enabling centralized monitoring and management.

For more information about the Apache Airflow REST API, refer to the Apache Airflow REST API Reference.

By using `InvokeRestApi`, you can access the Apache Airflow REST API using AWS credentials. Alternatively, you can also access it by obtaining a webserver access token and then using the token to call it.

If you encounter an error with the message `Update your environment to use InvokeRestApi` while using the `InvokeRestApi` operation, it indicates that you need to update your Amazon MWAA environment. This error occurs when your Amazon MWAA environment is not compatible with the latest changes related to the `InvokeRestApi` feature. To resolve this issue, update your Amazon MWAA environment to incorporate the necessary changes for the `InvokeRestApi` feature.

The `InvokeRestApi` operation has a default timeout duration of 10 seconds. If the operation does not complete within this 10-second timeframe, it is automatically terminated, and an error is raised. Ensure that your REST API calls are designed to complete within this timeout period to avoid encountering errors.

To support scalability while using the Apache Airflow REST API, Amazon MWAA provides you with the option to horizontally scale web server capacity to handle increased demand, whether from REST API requests, command line interface (CLI) usage, or more concurrent Apache Airflow user interface (UI) users. For more information about how Amazon MWAA scales web servers, refer to the section called “Configuring webserver auto scaling”.

You can use the Apache Airflow REST API to implement the following use cases for your environments:

- **Programmatic access** – You can now start Apache Airflow DAG runs, manage datasets, and retrieve the status of various components such as the metadata database, triggerers, and schedulers without relying on the Apache Airflow UI or CLI.
- **Integrate with external applications and microservices** – REST API support you can use to build custom solutions that integrate your Amazon MWAA environments with other systems. For example, you can start workflows in response to events from external systems, such as completed database jobs or new user sign-ups.
- **Centralized monitoring** – You can build monitoring dashboards that aggregate the status of your DAGs across multiple Amazon MWAA environments, enabling centralized monitoring and management.

For more information about the Apache Airflow REST API, refer to The Apache Airflow REST API Reference.

By using `InvokeRestApi`, you can access the Apache Airflow REST API using AWS credentials. Alternatively, you can also access it by obtaining a web server access token and then using the token to call it.

- If you encounter an error with the message `Update your environment to use InvokeRestApi` while using the `InvokeRestApi` operation, it indicates that you need to update your Amazon MWAA environment. This error occurs when your Amazon MWAA environment is not compatible with the latest changes related to the `InvokeRestApi` feature. To resolve this issue, update your Amazon MWAA environment to incorporate the necessary changes for the `InvokeRestApi` feature.

- The `InvokeRestApi` operation has a default timeout duration of 10 seconds. If the operation does not complete within this 10-second timeframe, it is automatically terminated, and an error is raised. Ensure that your REST API calls are designed to complete within this timeout period to avoid encountering errors.

> **Important**
>
> The response payload size cannot exceed 6 MB. Your `RestApi` fails if this limit is exceeded.

> **`InvokeRestApi` limitations in Amazon MWAA environments with IAM Identity Center**
>
> If your environment has IAM Identity Center integration enabled, refer to Known limitations before you use `InvokeRestApi`.

Use the following examples to make API calls to the Apache Airflow REST API and start a new DAG run:

Topics

- Granting access to the Apache Airflow REST API: airflow:InvokeRestApi
- Calling the Apache Airflow REST API
- Creating a webserver session token and calling the Apache Airflow REST API

Granting access to the Apache Airflow REST API: `airflow:InvokeRestApi`

To access the Apache Airflow REST API using AWS credentials, you must grant the `airflow:InvokeRestApi` permission in your IAM policy. In the following policy sample, specify the `Admin`, `Op`, `User`, `Viewer`, or `Public` role in `{airflow-role}` to customize the level of user access. For more information, refer to Default Roles in the *Apache Airflow reference guide*.

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
```

```
            "Sid": "AllowMwaaRestApiAccess",
            "Effect": "Allow",
            "Action": ["airflow:InvokeRestApi","airflow:CreateWebLoginToken"],
            "Resource": [
            "arn:aws:airflow:{your-region}:{your-account-id}:role/{your-environment-
name}/{airflow-role}"
            ]
        }
    ]
}
```

Note

`InvokeRestApi` is an Amazon MWAA service API call authenticated with AWS credentials (SigV4/IAM). When you call it, Amazon MWAA retrieves data from the Apache Airflow webserver over AWS internal networking. The webserver itself is never exposed to the public internet.

For environments with a private webserver, you can call `InvokeRestApi` over the public internet. This requires the `airflow:InvokeRestApi` and `airflow:CreateWebLoginToken` permissions in your IAM policy. Interactive access to the Apache Airflow web UI is restricted to the Virtual Private Cloud (VPC).

To confine `InvokeRestApi` calls to a VPC, apply an IAM policy condition using the `aws:SourceVpc` or `aws:SourceVpce` key. For more information about this condition key, see aws:SourceVpc.

Calling the Apache Airflow REST API

The following sample Python script shows how to use the Apache Airflow REST API. The script lists the available DAGs in your environment and creates an Apache Airflow variable:

```
import boto3

env_name = "MyAirflowEnvironment"

def list_dags(client):
    request_params = {
        "Name": env_name,
        "Path": "/dags",
        "Method": "GET",
```

```
        "QueryParameters": {
            "paused": False
        }
    }
    response = client.invoke_rest_api(
        **request_params
    )
    print("Airflow REST API response: ", response['RestApiResponse'])

def create_variable(client):
    request_params = {
        "Name": env_name,
        "Path": "/variables",
        "Method": "POST",
        "Body": {
            "key": "test-restapi-key",
            "value": "test-restapi-value",
            "description": "Test variable created by MWAA InvokeRestApi API",
        }
    }
    response = client.invoke_rest_api(
        **request_params
    )
    print("Airflow REST API response: ", response['RestApiResponse'])

if __name__ == "__main__":
    client = boto3.client("mwaa")
    list_dags(client)
    create_variable(client)
```

Creating a webserver session token and calling the Apache Airflow REST API

To create a webserver access token, use the following Python function. This function first calls the Amazon MWAA API to obtain a web login token. The web login token, which expires after 60 seconds, is then exchanged for a web *session* token, which lets you access the webserver and use the Apache Airflow REST API. If you require more than 10 transactions per second (TPS) of throttling capacity, you can use this method to access the Apache Airflow REST API.

The session token expires after 12 hours.

Tip

Key changes in the following code samples from Apache Airflow v2 to v3 are:

- REST API path changed from `/api/v1` to `/api/v2`
- Login path changed from `/aws_mwaa/login` to `/pluginsv2/aws_mwaa/login`
- Response from login `response.cookies["_token"]` contains token information that you must use for subsequent API calls
- For a REST API call, you must pass `jwt_token` information in headers as:

```
headers = {
  "Authorization": f"Bearer {jwt_token}",
  "Content-Type": "application/json"
}
```

Apache Airflow v3

```
def get_token_info(region, env_name):
    logging.basicConfig(level=logging.INFO)

    try:
        # Initialize MWAA client and request a web login token
        mwaa = boto3.client('mwaa', region_name=region)
        response = mwaa.create_web_login_token(Name=env_name)

        # Extract the web server hostname and login token
        web_server_host_name = response["WebServerHostname"]
        web_token = response["WebToken"]

        # Construct the URL needed for authentication
        login_url = f"https://{web_server_host_name}/pluginsv2/aws_mwaa/login"
        login_payload = {"token": web_token}

        # Make a POST request to the MWAA login url using the login payload
        response = requests.post(
            login_url,
            data=login_payload,
            timeout=10
        )
```

```
        # Check if login was successful
        if response.status_code == 200:
            # Return the hostname and the session cookie
            return (
                web_server_host_name,
                response.cookies['_token']
            )
        else:
            # Log an error
            logging.error("Failed to log in: HTTP %d", response.status_code)
            return None
    except requests.RequestException as e:
        # Log any exceptions raised during the request to the MWAA login endpoint
        logging.error("Request failed: %s", str(e))
        return None
    except Exception as e:
        # Log any other unexpected exceptions
        logging.error("An unexpected error occurred: %s", str(e))
        return None
```

Apache Airflow v2

```
def get_session_info(region, env_name):
    logging.basicConfig(level=logging.INFO)

    try:
        # Initialize MWAA client and request a web login token
        mwaa = boto3.client('mwaa', region_name=region)
        response = mwaa.create_web_login_token(Name=env_name)

        # Extract the web server hostname and login token
        web_server_host_name = response["WebServerHostname"]
        web_token = response["WebToken"]

        # Construct the URL needed for authentication
        login_url = f"https://{web_server_host_name}/aws_mwaa/login"
        login_payload = {"token": web_token}

        # Make a POST request to the MWAA login url using the login payload
        response = requests.post(
            login_url,
            data=login_payload,
```

```
            timeout=10
        )

        # Check if login was successful
        if response.status_code == 200:
            # Return the hostname and the session cookie
            return (
                web_server_host_name,
                response.cookies["session"]
            )
        else:
            # Log an error
            logging.error("Failed to log in: HTTP %d", response.status_code)
            return None
    except requests.RequestException as e:
        # Log any exceptions raised during the request to the MWAA login endpoint
        logging.error("Request failed: %s", str(e))
        return None
    except Exception as e:
        # Log any other unexpected exceptions
        logging.error("An unexpected error occurred: %s", str(e))
        return None
```

After authentication is complete, you have the credentials to start sending requests to the API endpoints. In the example in the following section, use the endpoint `dags/{dag_name}/dagRuns`.

Apache Airflow v3

```
def trigger_dag(region, env_name, dag_id):
    """
    Triggers a DAG in a specified MWAA environment using the Airflow REST API.

    Args:
    region (str): AWS region where the MWAA environment is hosted.
    env_name (str): Name of the MWAA environment.
    dag_id (str): ID of the DAG to trigger.
    """

    logging.info(f"Attempting to trigger DAG {dag_id} in environment {env_name} at
 region {region}")
```

```
    # Retrieve the web server hostname and token for authentication
    try:
        web_server_host_name, jwt_token = get_token_info(region, env_name)
        if not jwt_token:
            logging.error("Authentication failed, no jwt token retrieved.")
            return
    except Exception as e:
        logging.error(f"Error retrieving token info: {str(e)}")
        return

    # Prepare headers and payload for the request
    request_headers = {
        "Authorization": f"Bearer {jwt_token}",
        "Content-Type": "application/json"
    }

    # sample request body input
    json_body = {"logical_date": "2025-09-17T14:15:00Z"}

    # Construct the URL for triggering the DAG
    url = f"https://{web_server_host_name}/api/v2/dags/{dag_id}/dagRuns"

    # Send the POST request to trigger the DAG
    try:
        response = requests.post(url, headers=request_headers, json=json_body)
        # Check the response status code to determine if the DAG was triggered
 successfully
        if response.status_code == 200:
            logging.info("DAG triggered successfully.")
        else:
            logging.error(f"Failed to trigger DAG: HTTP {response.status_code} -
 {response.text}")
    except requests.RequestException as e:
        logging.error(f"Request to trigger DAG failed: {str(e)}")

if __name__ == "__main__":
    logging.basicConfig(level=logging.INFO)

    # Check if the correct number of arguments is provided
    if len(sys.argv) != 4:
        logging.error("Incorrect usage. Proper format: python script_name.py
 {region} {env_name} {dag_id}")
        sys.exit(1)
```

```
    region = sys.argv[1]
    env_name = sys.argv[2]
    dag_id = sys.argv[3]

    # Trigger the DAG with the provided arguments
    trigger_dag(region, env_name, dag_id)
```

Apache Airflow v2

```
def trigger_dag(region, env_name, dag_name):
    """
    Triggers a DAG in a specified MWAA environment using the Airflow REST API.

    Args:
    region (str): AWS region where the MWAA environment is hosted.
    env_name (str): Name of the MWAA environment.
    dag_name (str): Name of the DAG to trigger.
    """

    logging.info(f"Attempting to trigger DAG {dag_name} in environment {env_name} at
 region {region}")

    # Retrieve the web server hostname and session cookie for authentication
    try:
        web_server_host_name, session_cookie = get_session_info(region, env_name)
        if not session_cookie:
            logging.error("Authentication failed, no session cookie retrieved.")
            return
    except Exception as e:
        logging.error(f"Error retrieving session info: {str(e)}")
        return

    # Prepare headers and payload for the request
    cookies = {"session": session_cookie}
    json_body = {"conf": {}}

    # Construct the URL for triggering the DAG
    url = f"https://{web_server_host_name}/api/v1/dags/{dag_name}/dagRuns"

    # Send the POST request to trigger the DAG
    try:
        response = requests.post(url, cookies=cookies, json=json_body)
```

```
        # Check the response status code to determine if the DAG was triggered
 successfully
        if response.status_code == 200:
            logging.info("DAG triggered successfully.")
        else:
            logging.error(f"Failed to trigger DAG: HTTP {response.status_code} -
 {response.text}")
    except requests.RequestException as e:
        logging.error(f"Request to trigger DAG failed: {str(e)}")

if __name__ == "__main__":
    logging.basicConfig(level=logging.INFO)

    # Check if the correct number of arguments is provided
    if len(sys.argv) != 4:
        logging.error("Incorrect usage. Proper format: python script_name.py
 {region} {env_name} {dag_name}")
        sys.exit(1)

    region = sys.argv[1]
    env_name = sys.argv[2]
    dag_name = sys.argv[3]

    # Trigger the DAG with the provided arguments
    trigger_dag(region, env_name, dag_name)
```

Apache Airflow CLI command reference

This topic describes the supported and unsupported Apache Airflow CLI commands on Amazon Managed Workflows for Apache Airflow.

> **Tip**
>
> REST API is more modern than the CLI and is designed for programmatic integration with external systems. REST is the preferred way of interacting with Apache Airflow.

Contents

- Prerequisites
  - Access

- AWS CLI
- What changed?
- Supported CLI commands
  - Supported commands
  - Using commands that parse DAGs
- Sample code
  - Set, get or delete an Apache Airflow v2 variable
  - Add a configuration when triggering a DAG
  - Run CLI commands on an SSH tunnel to a bastion host

Prerequisites

The following section describes the preliminary steps required to use the commands and scripts on this page.

Access

- AWS account access in AWS Identity and Access Management (IAM) to the Amazon MWAA permissions policy in Apache Airflow UI access policy: AmazonMWAAWebServerAccess.
- AWS account access in AWS Identity and Access Management (IAM) to the Amazon MWAA permissions policy Full API and console access policy: AmazonMWAAFullApiAccess.

AWS CLI

The AWS Command Line Interface (AWS CLI) is an open source tool that you can use to interact with AWS services using commands in your command-line shell. To complete the steps on this page, you need the following:

- AWS CLI – Install version 2.
- AWS CLI – Quick configuration with `aws configure`.

What changed?

- **v3: Airflow architecture**. Apache Airflow v3 introduces breaking architectural changes to provide improved security and scalability, and to make maintenance easier. To learn more, refer to Upgrading To Airflow 3 on the *Apache Airflow* website.
- **v2: Airflow CLI command structure**. The Apache Airflow v2 CLI is organized so that related commands are grouped together as subcommands, which means you need to update Apache Airflow v1 scripts if you want to upgrade to Apache Airflow v2. For example, `unpause` in Apache Airflow v1 is `dags unpause` in Apache Airflow v2. To learn more, refer to Airflow CLI changes in 2.0.

Supported CLI commands

The following section lists the Apache Airflow CLI commands available on Amazon MWAA.

Supported commands

Apache Airflow v3

Minor versions	Command
v3.0.6, v3.2.1, v3.3.1	assets details
v3.0.6, v3.2.1, v3.3.1	assets list
v3.0.6, v3.2.1, v3.3.1	assets materialize
v3.0.6, v3.2.1, v3.3.1	backfill create
v3.0.6, v3.2.1, v3.3.1	cheat-sheet
v3.0.6, v3.2.1, v3.3.1	connections add
v3.0.6, v3.2.1, v3.3.1	connections delete

Minor versions	Command
v3.0.6, v3.2.1, v3.3.1	dags delete
v3.0.6, v3.2.1, v3.3.1	dags list
v3.0.6, v3.2.1, v3.3.1	dags list-jobs
v3.0.6, v3.2.1, v3.3.1	dags list-import-errors
v3.0.6, v3.2.1, v3.3.1	dags list-runs
v3.0.6, v3.2.1, v3.3.1	dags next-execution
v3.0.6, v3.2.1, v3.3.1	dags pause
v3.0.6, v3.2.1, v3.3.1	dags report
v3.0.6, v3.2.1, v3.3.1	dags reserialize
v3.0.6, v3.2.1, v3.3.1	dags show
v3.0.6, v3.2.1, v3.3.1	dags state
v3.0.6, v3.2.1, v3.3.1	dags test
v3.0.6, v3.2.1, v3.3.1	dags trigger
v3.0.6, v3.2.1, v3.3.1	dags unpause

Minor versions	Command
v3.0.6, v3.2.1, v3.3.1	db clean
v3.0.6, v3.2.1, v3.3.1	providers behaviours
v3.0.6, v3.2.1, v3.3.1	providers get
v3.0.6, v3.2.1, v3.3.1	providers hooks
v3.0.6, v3.2.1, v3.3.1	providers links
v3.0.6, v3.2.1, v3.3.1	providers list
v3.0.6, v3.2.1, v3.3.1	providers notifications
v3.0.6, v3.2.1, v3.3.1	providers secrets
v3.0.6, v3.2.1, v3.3.1	providers triggers
v3.0.6, v3.2.1, v3.3.1	providers widgets
v3.0.6, v3.2.1, v3.3.1	tasks clear
v3.0.6, v3.2.1, v3.3.1	tasks failed-deps
v3.0.6, v3.2.1, v3.3.1	tasks list
v3.0.6, v3.2.1, v3.3.1	tasks render

Minor versions	Command
v3.0.6, v3.2.1, v3.3.1	tasks state
v3.0.6, v3.2.1, v3.3.1	tasks states-for-dag-run
v3.0.6, v3.2.1, v3.3.1	tasks test
v3.0.6, v3.2.1, v3.3.1	variables delete
v3.0.6, v3.2.1, v3.3.1	variables get
v3.0.6, v3.2.1, v3.3.1	variables set
v3.0.6, v3.2.1, v3.3.1	variables list
v3.0.6, v3.2.1, v3.3.1	version

Note

In Apache Airflow 3.x, you can no longer use most of the RBAC role and user management commands in the CLI. Instead, you manage roles and permissions through the REST API or the Apache Airflow UI. This change reflects Apache Airflow 3's shift toward API-first management.
To manage roles in Apache Airflow 3, you can do the following:

- Use the REST API — The `/api/v2/roles` endpoint supports listing, creating, updating, and deleting roles.
- Use the Apache Airflow UI — Navigate to **Security** > **List Roles** in the web interface.

Apache Airflow v2

Minor versions	Command
v2.0+	cheat-sheet
v2.0+	connections add
v2.0+	connections delete
v2.2+ (note)	dags backfill
v2.0+	dags delete
v2.2+ (note)	dags list
v2.0+	dags list-jobs
v2.6+	dags list-import-errors
v2.2+ (note)	dags list-runs
v2.2+ (note)	dags next-execution
v2.0+	dags pause
v2.0+	dags report
v2.4+	dags reserialize
v2.0+	dags show

Minor versions	Command
v2.0+	dags state
v2.0+	dags test
v2.0+	dags trigger
v2.0+	dags unpause
v2.4+	db clean
v2.0+	providers behaviours
v2.0+	providers get
v2.0+	providers hooks
v2.0+	providers links
v2.0+	providers list
v2.8+	providers notifications
v2.6+	providers secrets
v2.7+	providers triggerer
v2.0+	providers widgets
v2.6+	roles add-perms

Minor versions	Command
v2.6+	roles del-perms
v2.6+	roles create
v2.0+	roles list
v2.0+	tasks clear
v2.0+	tasks failed-deps
v2.0+	tasks list
v2.0+	tasks render
v2.0+	tasks run
v2.0+	tasks state
v2.0+	tasks states-for-dag-run
v2.0+	tasks test
v2.0+	variables delete
v2.0+	variables get
v2.0+	variables set

Minor versions	Command
v2.0+	variables list
v2.0+	version

Using commands that parse DAGs

If your environment is running Apache Airflow v2.0.2, CLI commands that parse DAGs will fail if the DAG uses plugins that depend on packages installed through a `requirements.txt`:

Apache Airflow v2.0.2

- `dags backfill`
- `dags list`
- `dags list-runs`
- `dags next-execution`

You can use these CLI commands if your DAGs do not use plugins that depend on packages installed through a `requirements.txt`.

Sample code

The following section contains examples of different ways to use the Apache Airflow CLI.

Set, get or delete an Apache Airflow v2 variable

You can use the following sample code to set, get or delete a variable in the format of `<script> <mwaa env name> get | set | delete <variable> <variable value> </variable> </variable>`.

```
[ $# -eq 0 ] && echo "Usage: $0 MWAA environment name " && exit

if [[ $2 == "" ]]; then
    dag="variables list"

elif  [ $2 == "get" ] ||  [ $2 == "delete" ] ||  [ $2 == "set" ]; then
```

```
    dag="variables $2 $3 $4 $5"

else
    echo "Not a valid command"
    exit 1
fi

CLI_JSON=$(aws mwaa --region $AWS_REGION create-cli-token --name $1) \
    && CLI_TOKEN=$(echo $CLI_JSON | jq -r '.CliToken') \
    && WEB_SERVER_HOSTNAME=$(echo $CLI_JSON | jq -r '.WebServerHostname') \
    && CLI_RESULTS=$(curl --request POST "https://$WEB_SERVER_HOSTNAME/aws_mwaa/cli" \
    --header "Authorization: Bearer $CLI_TOKEN" \
    --header "Content-Type: text/plain" \
    --data-raw "$dag" ) \
    && echo "Output:" \
    && echo $CLI_RESULTS | jq -r '.stdout' | base64 --decode \
    && echo "Errors:" \
    && echo $CLI_RESULTS | jq -r '.stderr' | base64 --decode
```

Add a configuration when triggering a DAG

You can use the following sample code with Apache Airflow v2 to add a configuration when triggering a DAG, such as `airflow dags trigger 'dag_name' --conf '{"key":"value"}'`.

```
import boto3
import json
import requests
import base64

mwaa_env_name = 'YOUR_ENVIRONMENT_NAME'
dag_name = 'YOUR_DAG_NAME'
key = "YOUR_KEY"
value = "YOUR_VALUE"
conf = "{\"" + key + "\":\"" + value + "\"}"

client = boto3.client('mwaa')

mwaa_cli_token = client.create_cli_token(
  Name=mwaa_env_name
)

mwaa_auth_token = 'Bearer ' + mwaa_cli_token['CliToken']
```

```
mwaa_webserver_hostname = 'https://{0}/aws_mwaa/
cli'.format(mwaa_cli_token['WebServerHostname'])
raw_data = "dags trigger {0} -c '{1}'".format(dag_name, conf)

mwaa_response = requests.post(
  mwaa_webserver_hostname,
  headers={
    'Authorization': mwaa_auth_token,
    'Content-Type': 'text/plain'
  },
  data=raw_data
)

mwaa_std_err_message = base64.b64decode(mwaa_response.json()['stderr']).decode('utf8')
mwaa_std_out_message = base64.b64decode(mwaa_response.json()['stdout']).decode('utf8')

print(mwaa_response.status_code)
print(mwaa_std_err_message)
print(mwaa_std_out_message)
```

Run CLI commands on an SSH tunnel to a bastion host

Use the following example to run Airflow CLI commands using an SSH tunnel proxy to a Linux Bastion Host.

Using curl

1. ```
   ssh -D 8080 -f -C -q -N YOUR_USER@YOUR_BASTION_HOST
   ```
2. ```
   curl -x socks5h://0:8080 --request POST https://YOUR_HOST_NAME/aws_mwaa/cli --
   header YOUR_HEADERS --data-raw YOUR_CLI_COMMAND
   ```

Managing connections to Apache Airflow

This chapter describes how to configure an Apache Airflow connection for an Amazon Managed Workflows for Apache Airflow environment.

Topics

- Overview of Apache Airflow variables and connections
- Apache Airflow provider packages installed on Amazon MWAA environments
- Overview of connection types
- Configuring an Apache Airflow connection using a AWS Secrets Manager secret

Overview of Apache Airflow variables and connections

In some cases, you might want to specify additional connections or variables for an environment, such as an AWS profile, or to add your execution role in a connection object in the Apache Airflow metastore, then refer to the connection from within a DAG.

- **Self-managed Apache Airflow**. On a self-managed Apache Airflow installation, you set Apache Airflow configuration options in `airflow.cfg`.

  ```
  [secrets]
  backend = airflow.providers.amazon.aws.secrets.secrets_manager.SecretsManagerBackend
  backend_kwargs = {"connections_prefix" : "airflow/connections", "variables_prefix" :
   "airflow/variables"}
  ```

- **Apache Airflow on Amazon MWAA**. On Amazon MWAA, you need to add these configuration settings as Apache Airflow configuration options on the Amazon MWAA console. Apache Airflow configuration options are written as environment variables to your environment and override all other existing configurations for the same setting.

Apache Airflow provider packages installed on Amazon MWAA environments

This page lists the Apache Airflow provider packages installed by Amazon MWAA for all supported Apache Airflow environments. For more information about these packages, refer to the Apache Airflow reference for package extras.

> **Note**
>
> To ensure that compatibility with CloudWatch logging is not overridden by other Python library installations, Amazon MWAA installs Watchtower version 2.0.1 after performing `pip3 install -r requirements.txt`.

Topics

- Constraints file
- Version-specific provider packages

Constraints file

Beginning with Apache Airflow v2.7.2, your requirements file must include a `--constraint` statement. If you don't provide a constraint, Amazon MWAA will specify one for you to ensure the packages listed in your requirements are compatible with the version of Apache Airflow you're using.

Apache Airflow constraints files specify the provider versions available at the time of an Apache Airflow release. In many cases, however, newer providers are compatible with that version of Apache Airflow. Because you must use constraints, to specify a newer version of a provider package, you can modify the constraints file for a specific provider version:

1. Download the version-specific constraints file from GitHub, for example https://raw.githubusercontent.com/apache/airflow/constraints-2.7.2/constraints-3.11.txt (replace '2.7.2' with the version you want to use).
2. Save the modified constraints file to the Amazon S3 dags folder of your Amazon MWAA environment, for example, as `constraints-3.11-updated.txt`.
3. Specify your requirements as listed in the following.

```
--constraint "/usr/local/airflow/dags/constraints-3.11-updated.txt"
apache-airflow-providers-amazon==version-number
```

> **Note**
>
> If you are using a private webserver, we recommend that you package the required libraries as WHL files by using aws-mwaa-docker-images.

Version-specific provider packages

Installing provider packages you can use to access a connection type in the Apache Airflow UI. It also means you don't need to specify these packages as a Python dependency in your `requirements.txt` file. This page lists the Apache Airflow provider packages installed by Amazon MWAA for all supported Apache Airflow environments.

> **Note**
>
> For Apache Airflow v2 and later, Amazon MWAA installs Watchtower version 2.0.1 after performing `pip3 install -r requirements.txt`, to ensure compatibility with CloudWatch logging is not overridden by other Python library installations.

You can specify the latest supported version of `apache-airflow-providers-amazon` to upgrade this provider.

Supported Apache Airflow versions:

v3.3.1

Connection type	Package
AWS Connection	apache-airflow-providers-amazon[aiobotocore]==9.34.0
Postgres Connection	apache-airflow-providers-postgres==7.0.1

Connection type	Package
FTP Connection	apache-airflow-providers-ftp==3.15.2
Fab Connection	apache-airflow-providers-fab==3.8.0
Celery Connection	apache-airflow-providers-celery==3.23.1
HTTP Connection	apache-airflow-providers-http==6.0.5
IMAP Connection	apache-airflow-providers-imap==3.12.1
Common SQL	apache-airflow-providers-common-sql==2.1.0
SQLite Connection	apache-airflow-providers-sqlite==4.3.3

v3.2.1

Connection type	Package
AWS Connection	apache-airflow-providers-amazon[aiobotocore]==9.25.0
Postgres Connection	apache-airflow-providers-postgres==6.6.3
FTP Connection	apache-airflow-providers-ftp==3.14.3
Fab Connection	apache-airflow-providers-fab==3.6.1
Celery Connection	apache-airflow-providers-celery==3.18.0

Connection type	Package
HTTP Connection	apache-airflow-providers-http==6.0.2
IMAP Connection	apache-airflow-providers-imap==3.11.2
Common SQL	apache-airflow-providers-common-sql==1.34.0
SQLite Connection	apache-airflow-providers-sqlite==4.3.2

v3.0.6

Connection type	Package
AWS Connection	apache-airflow-providers-amazon[aiobotocore]==9.9.0
Postgres Connection	apache-airflow-providers-postgres==6.2.1
FTP Connection	apache-airflow-providers-ftp==3.13.1
Fab Connection	apache-airflow-providers-fab==2.3.0
Celery Connection	apache-airflow-providers-celery==3.12.1
HTTP Connection	apache-airflow-providers-http==5.3.2
IMAP Connection	apache-airflow-providers-imap==3.9.1
Common SQL	

Connection type	Package
	apache-airflow-providers-common-sql==1.27.3
SQLite Connection	apache-airflow-providers-sqlite==4.1.1

v2.11.2

Connection type	Package
AWS Connection	apache-airflow-providers-amazon[aiobotocore]==9.22.0
Postgres Connection	apache-airflow-providers-postgres==6.6.0
FTP Connection	apache-airflow-providers-ftp==3.14.1
Celery Connection	apache-airflow-providers-celery==3.15.2
HTTP Connection	apache-airflow-providers-http==6.0.0
IMAP Connection	apache-airflow-providers-imap==3.11.0
Common SQL	apache-airflow-providers-common-sql==1.32.0
SQLite Connection	apache-airflow-providers-sqlite==4.3.0
SMTP Connection	apache-airflow-providers-smtp==2.4.2

v2.11.0

Connection type	Package
AWS Connection	apache-airflow-providers-amazon[aiobotocore]==9.8.0
Postgres Connection	apache-airflow-providers-postgres==6.2.0
FTP Connection	apache-airflow-providers-ftp==3.13.0
Fab Connection	apache-airflow-providers-fab==1.5.3
Celery Connection	apache-airflow-providers-celery==3.11.0
HTTP Connection	apache-airflow-providers-http==5.3.0
IMAP Connection	apache-airflow-providers-imap==3.9.0
Common SQL	apache-airflow-providers-common-sql==1.27.1
SQLite Connection	apache-airflow-providers-sqlite==4.1.0
SMTP Connection	apache-airflow-providers-smtp==2.1.0

v2.10.3

Connection type	Package
AWS Connection	apache-airflow-providers-amazon[aiobotocore]==9.0.0

Connection type	Package
Postgres Connection	apache-airflow-providers-postgres==5.13.1
FTP Connection	apache-airflow-providers-ftp==3.11.1
Fab Connection	apache-airflow-providers-fab==1.5.0
Celery Connection	apache-airflow-providers-celery==3.8.3
HTTP Connection	apache-airflow-providers-http==4.13.2
IMAP Connection	apache-airflow-providers-imap==3.7.0
Common SQL	apache-airflow-providers-common-sql==1.19.0
SQLite Connection	apache-airflow-providers-sqlite==3.9.0
SMTP Connection	apache-airflow-providers-smtp==1.8.0

v2.10.1

Connection type	Package
AWS Connection	apache-airflow-providers-amazon[aiobotocore]==8.28.0
Postgres Connection	apache-airflow-providers-postgres==5.12.0
FTP Connection	apache-airflow-providers-ftp==3.11.0

Connection type	Package
Fab Connection	apache-airflow-providers-fab==1.3.0
Celery Connection	apache-airflow-providers-celery==3.8.1
HTTP Connection	apache-airflow-providers-http==4.13.0
IMAP Connection	apache-airflow-providers-imap==3.7.0
Common SQL	apache-airflow-providers-common-sql==1.16.0
SQLite Connection	apache-airflow-providers-sqlite==3.9.0
SMTP Connection	apache-airflow-providers-smtp==1.8.0

v2.9.2

Connection type	Package
AWS Connection	apache-airflow-providers-amazon[aiobotocore]==8.24.0
Postgres Connection	apache-airflow-providers-postgres==5.11.1
FTP Connection	apache-airflow-providers-ftp==3.9.1
Fab Connection	apache-airflow-providers-fab==1.1.1
Celery Connection	apache-airflow-providers-celery==3.7.2

Connection type	Package
HTTP Connection	apache-airflow-providers-http==4.11.1
IMAP Connection	apache-airflow-providers-imap==3.6.1
Common SQL	apache-airflow-providers-common-sql==1.14.0
SQLite Connection	apache-airflow-providers-sqlite==3.8.1
SMTP Connection	apache-airflow-providers-smtp==1.7.1

v2.8.1

Connection type	Package
AWS Connection	apache-airflow-providers-amazon[aiobotocore]==8.16.0
Postgres Connection	apache-airflow-providers-postgres==5.10.0
FTP Connection	apache-airflow-providers-ftp==3.7.0
Celery Connection	apache-airflow-providers-celery==3.5.1
HTTP Connection	apache-airflow-providers-http==4.8.0
IMAP Connection	apache-airflow-providers-imap==3.5.0
Common SQL	

Connection type	Package
	apache-airflow-providers-common-sql==1.10.0
SQLite Connection	apache-airflow-providers-sqlite==3.7.0

v2.7.2

Connection type	Package
AWS Connection	apache-airflow-providers-amazon[aiobotocore]==8.7.1
Postgres Connection	apache-airflow-providers-postgres==5.6.1
FTP Connection	apache-airflow-providers-ftp==3.5.2
Celery Connection	apache-airflow-providers-celery==3.3.4
HTTP Connection	apache-airflow-providers-http==4.5.2
IMAP Connection	apache-airflow-providers-imap==3.3.2
Common SQL	apache-airflow-providers-common-sql==1.7.2
SQLite Connection	apache-airflow-providers-sqlite==3.4.3

v2.6.3

Connection type	Package
AWS Connection	apache-airflow-providers-amazon[aiobotocore]==8.2.0
Postgres Connection	apache-airflow-providers-postgres==5.5.1
FTP Connection	apache-airflow-providers-ftp==3.4.2
Celery Connection	apache-airflow-providers-celery==3.2.1
HTTP Connection	apache-airflow-providers-http==4.4.2
IMAP Connection	apache-airflow-providers-imap==3.2.2
Common SQL	apache-airflow-providers-common-sql==1.5.2
SQLite Connection	apache-airflow-providers-sqlite==3.4.2

v2.5.1

Connection type	Package
AWS Connection	apache-airflow-providers-amazon==7.1.0
Postgres Connection	apache-airflow-providers-postgres==5.4.0
FTP Connection	apache-airflow-providers-ftp==3.3.0

Connection type	Package
Celery Connection	apache-airflow-providers-celery==3.1.0
HTTP Connection	apache-airflow-providers-http==4.1.1
IMAP Connection	apache-airflow-providers-imap==3.1.1
Common SQL	apache-airflow-providers-common-sql==1.3.3
SQLite Connection	apache-airflow-providers-sqlite==3.3.1

v2.4.3

Connection type	Package
AWS Connection	apache-airflow-providers-amazon==6.0.0
Postgres Connection	apache-airflow-providers-postgres==5.2.2
FTP Connection	apache-airflow-providers-ftp==3.1.0
Celery Connection	apache-airflow-providers-celery==3.0.0
HTTP Connection	apache-airflow-providers-http==4.0.0
IMAP Connection	apache-airflow-providers-imap==3.0.0
Common SQL	apache-airflow-providers-common-sql==1.2.0

Connection type	Package
SQLite Connection	apache-airflow-providers-sqlite==3.2.1

Overview of connection types

Apache Airflow stores connections as a connection URI string. It provides a connections template in the Apache Airflow UI to generate the connection URI string, regardless of the connection type. If a connection template is not available in the Apache Airflow UI, an alternate connection template can be used to generate this connection URI string, such as using the HTTP connection template. The primary difference is the URI prefix, such as `my-conn-type://`, which Apache Airflow providers typically ignore for a connection. This page describes how to use connection templates in the Apache Airflow UI interchangeably for different connection types.

Warning

Don't overwrite the `aws_default` connection in Amazon MWAA. Amazon MWAA uses this connection to perform a variety of critical tasks, such as collecting task logs. Overwriting this connection might result in data loss and disruptions to your environment availability.

Topics

- Example connection URI string
- Example connection template
- Example using an HTTP connection template for a Jdbc connection

Example connection URI string

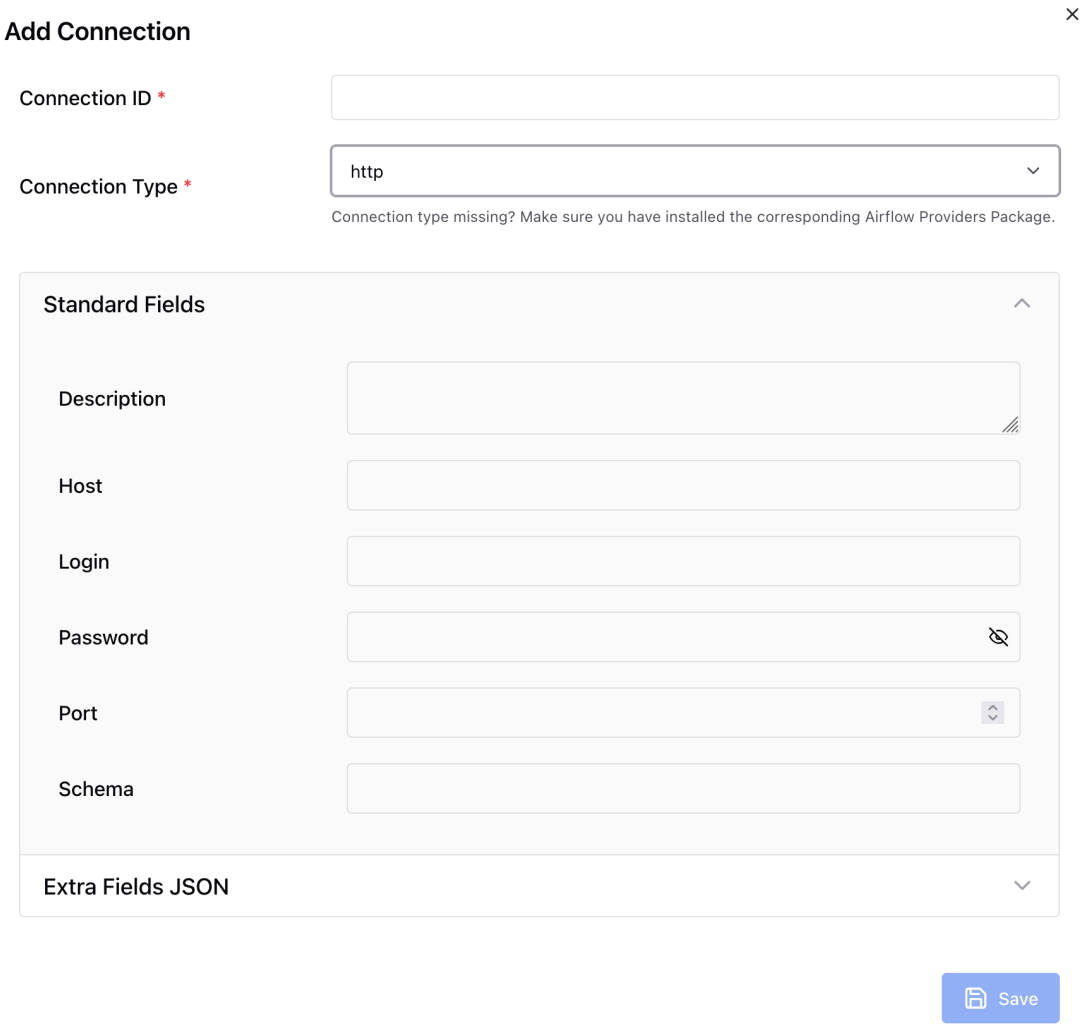
The following example presents a connection URI string for the MySQL connection type.

```
'mysql://288888a0-50a0-888-9a88-1a111aaa0000.a1.us-east-1.airflow.amazonaws.com
%2Fhome?role_arn=arn%3Aaws%3Aiam%3A%3A001122332255%3Arole%2Fservice-role%2FAmazonMWAA-
MyAirflowEnvironment-iAaaaA&region_name=us-east-1'
```

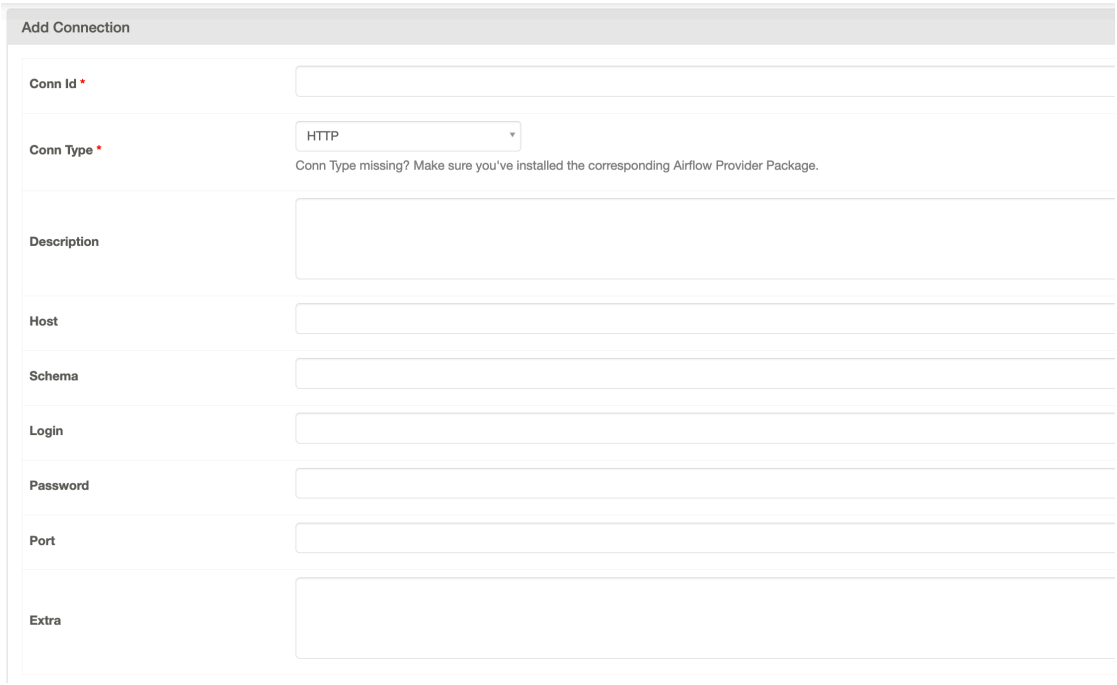
Example connection template

The following examples depict the HTTP connection template in the Apache Airflow UI.

Apache Airflow v3



Apache Airflow v2



Example using an HTTP connection template for a Jdbc connection

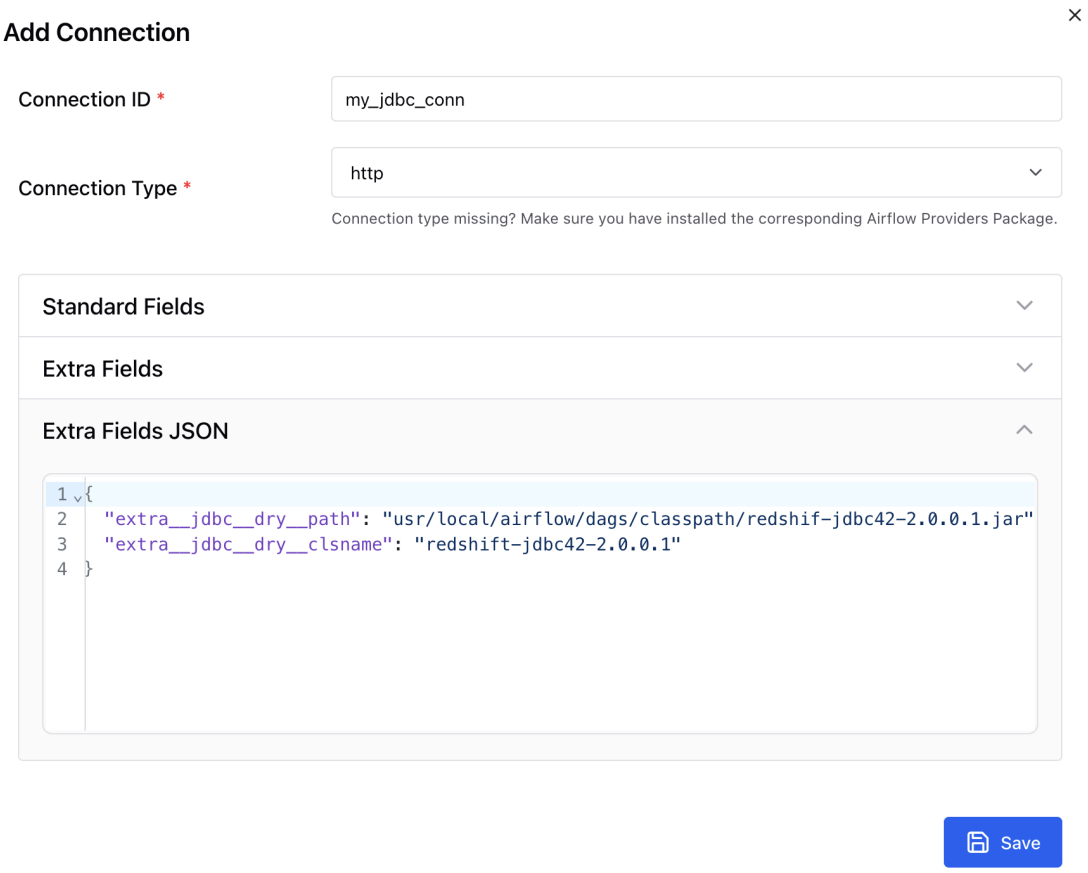
Use the following example to apply the **HTTP** connection template for a *Jdbc* connection type in the Apache Airflow UI.

Apache Airflow v3

The following example displays the connection URI string generated by Apache Airflow for the example in this section.

```
http://myconnectionurl/some/path&login=mylogin&extra__jdbc__dry__path=usr/local/
airflow/dags/classpath/redshif-
jdbc42-2.0.0.1.jar&extra__jdbc__dry__clsname=redshift-jdbc42-2.0.0.1
```

Use the following example to apply the HTTP connection template for a *Jdbc* connection for Apache Airflow v3 in the Apache Airflow UI.

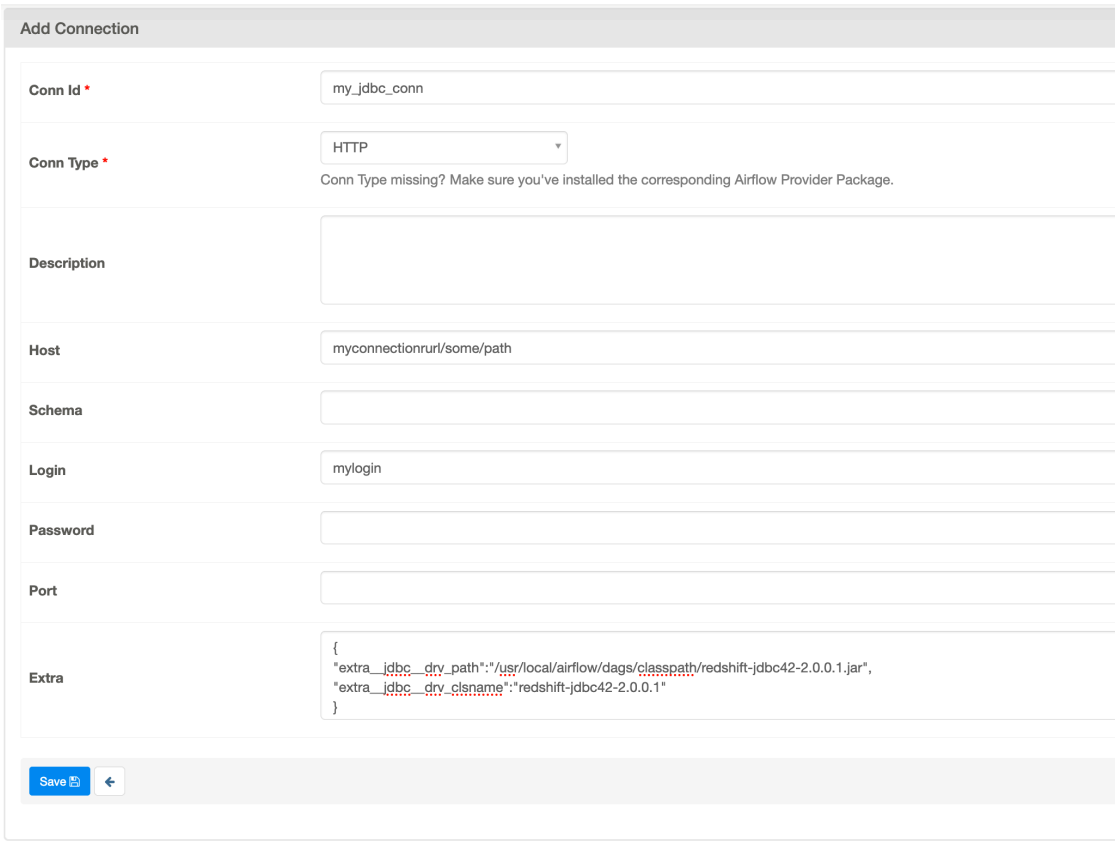


Apache Airflow v2

The following example displays the connection URI string generated by Apache Airflow for the example in this section.

```
http://myconnectionurl/some/path&login=mylogin&extra__jdbc__dry__path=usr/local/
airflow/dags/classpath/redshif-
jdbc42-2.0.0.1.jar&extra__jdbc__dry__clsname=redshift-jdbc42-2.0.0.1
```

Use the following example to apply the HTTP connection template for a *Jdbc* connection for Apache Airflow v2 in the Apache Airflow UI.



Configuring an Apache Airflow connection using a AWS Secrets Manager secret

AWS Secrets Manager is a supported alternative Apache Airflow backend on an Amazon Managed Workflows for Apache Airflow environment. This topic explains how to use AWS Secrets Manager to securely store secrets for Apache Airflow variables and an Apache Airflow connection on Amazon Managed Workflows for Apache Airflow.

Note

- You are charged for the secrets you create. For more information about Secrets Manager pricing, refer to AWS Pricing.
- AWS Systems Manager Parameter Store is also supported as a secrets backend in Amazon MWAA. For more information, refer to Amazon Provider Package documentation.

Contents

- Step one: Provide Amazon MWAA with permission to access Secrets Manager secret keys
- Step two: Create the Secrets Manager backend as an Apache Airflow configuration option
- Step three: Generate an Apache Airflow AWS connection URI string
- Step four: Add the variables in Secrets Manager
- Step five: Add the connection in Secrets Manager
- Sample code
- Resources
- What's next?

Step one: Provide Amazon MWAA with permission to access Secrets Manager secret keys

The execution role for your Amazon MWAA environment needs read access to the secret key in AWS Secrets Manager. The following IAM policy allows read-write access using the AWS-managed SecretsManagerReadWrite policy.

To attach the policy to your execution role

1. Open the Environments page on the Amazon MWAA console.
2. Choose an environment.
3. Choose your execution role on the **Permissions** pane.
4. Choose **Attach policies**.
5. Type `SecretsManagerReadWrite` in the **Filter policies** text field.
6. Choose **Attach policy**.

If you do not want to use an AWS-managed permission policy, you can directly update your environment's execution role to allow any level of access to your Secrets Manager resources. The following JSON policy grants read access to all secrets in a specific AWS Region. Replace *us-east-1* with your Region, and replace *111122223333* with your AWS account ID:

Example – IAM policy for Secrets Manager read access

```
{
    "Version": "2012-10-17",
```

```
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "secretsmanager:GetSecretValue"
            ],
            "Resource": "arn:aws:secretsmanager:us-east-1:111122223333:secret:*"
        },
        {
            "Effect": "Allow",
            "Action": "secretsmanager:ListSecrets",
            "Resource": "*"
        }
    ]
}
```

Wildcard resource for ListSecrets

The preceding policy grants `ListSecrets` access to all secrets (`"Resource": "*"`). `ListSecrets` does not support resource-level permissions, so the wildcard is required. If you use this policy in production, any principal that assumes this role can list every secret in your account. Restrict the role's trust policy and other permissions to limit the scope of access.

Optional: DescribeSecret permission

If your provider version checks for secret existence before retrieval, add `secretsmanager:DescribeSecret`. This permission might also reduce `ResourceNotFoundException` errors in your logs.

Step two: Create the Secrets Manager backend as an Apache Airflow configuration option

The following section describes how to create an Apache Airflow configuration option on the Amazon MWAA console for the AWS Secrets Manager backend. If you're using a configuration setting of the same name in `airflow.cfg`, the configuration you create in the following steps takes precedence and override the configuration settings.

1. Open the Environments page on the Amazon MWAA console.
2. Choose an environment.
3. Choose **Edit**.
4. Choose **Next**.
5. Choose **Add custom configuration** in the **Airflow configuration options** pane. Add the following key-value pairs:

   a. **`secrets.backend`**: **`airflow.providers.amazon.aws.secrets.secrets_manager.SecretsManagerBacken`**

   b. **`secrets.backend_kwargs`**: **`{"connections_prefix":"airflow/connections","variables_prefix":"airflow/variables"}`** This configures Apache Airflow to search for connection strings and variables at `airflow/connections/*` and `airflow/variables/*` paths.

   You can use a lookup pattern on the Apache Airflow website to reduce the number of API calls Amazon MWAA makes to Secrets Manager on your behalf. If you do not specify a lookup pattern, Apache Airflow searches for all connections and variables in the configured backend. By specifying a pattern, you narrow the possible paths that Apache Airflow searches. This lowers your costs when using Secrets Manager with Amazon MWAA.

   To specify a lookup pattern, specify the `connections_lookup_pattern` and `variables_lookup_pattern` parameters. These parameters accept a RegEx string as input. For example, to search for secrets that start with `test`, enter the following for `secrets.backend_kwargs`:

   ```
   {
     "connections_prefix": "airflow/connections",
     "connections_lookup_pattern": "^test",
     "variables_prefix" : "airflow/variables",
     "variables_lookup_pattern": "^test"
   }
   ```

   Note

   To use `connections_lookup_pattern` and `variables_lookup_pattern`, you must install `apache-airflow-providers-amazon` version 7.3.0 or higher.

For more information about updating provder pacakges for to newer versions, refer to Constraints file.

6. Choose **Save**.

Step three: Generate an Apache Airflow AWS connection URI string

To create a connection string, use the "tab" key on your keyboard to indent the key-value pairs in the Connection object. We also recommend creating a variable for the `extra` object in your shell session. The following section walks you through the steps to generate an Apache Airflow connection URI string for an Amazon MWAA environment using Apache Airflow or a Python script.

Apache Airflow CLI

The following shell session uses your local Airflow CLI to generate a connection string. If you don't have the CLI installed, we recommend using the Python script.

1. Open a Python shell session:

```
python3
```

2. Enter the following command:

```
>>> import json
```

3. Enter the following command:

```
>>> from airflow.models.connection import Connection
```

4. Create a variable in your shell session for the `extra` object. Substitute the sample values in *YOUR_EXECUTION_ROLE_ARN* with the execution role ARN, and the region in *us-east-1* (such as `us-east-1`).

```
>>> extra=json.dumps({'role_arn': 'YOUR_EXECUTION_ROLE_ARN', 'region_name': 'us-east-1'})
```

5. Create the connection object. Substitute the sample value in `myconn` with the name of the Apache Airflow connection.

```
>>> myconn = Connection(
```

6. Use the "tab" key on your keyboard to indent each of the following key-value pairs in your connection object. Substitute the sample values in *red*.

   a. Specify the AWS connection type:

   ```
   ... conn_id='aws',
   ```

   b. Specify the Apache Airflow database option:

   ```
   ... conn_type='mysql',
   ```

   c. Specify the Apache Airflow UI URL on Amazon MWAA:

   ```
   ... host='288888a0-50a0-888-9a88-1a111aaa0000.a1.us-
   east-1.airflow.amazonaws.com/home',
   ```

   d. Specify the AWS access key ID (username) to log in to Amazon MWAA:

   ```
   ... login='YOUR_AWS_ACCESS_KEY_ID',
   ```

   e. Specify the AWS secret access key (password) to log in to Amazon MWAA:

   ```
   ... password='YOUR_AWS_SECRET_ACCESS_KEY',
   ```

   f. Specify the `extra` shell session variable:

   ```
   ... extra=extra
   ```

   g. Close the connection object.

   ```
   ... )
   ```

7. Print the connection URI string:

   ```
   >>> myconn.get_uri()
   ```

   Refer to the connection URI string in the response:

```
'mysql://288888a0-50a0-888-9a88-1a111aaa0000.a1.us-east-1.airflow.amazonaws.com
%2Fhome?role_arn=arn%3Aaws%3Aiam%3A%3A001122332255%3Arole%2Fservice-role
%2FAmazonMWAA-MyAirflowEnvironment-iAaaaA&region_name=us-east-1'
```

Python script

The following Python script does not require the Apache Airflow CLI.

1. Copy the contents of the following code sample and save locally as `mwaa_connection.py`.

```
import urllib.parse

        conn_type = 'YOUR_DB_OPTION'
        host = 'YOUR_MWAA_AIRFLOW_UI_URL'
        port = 'YOUR_PORT'
        login = 'YOUR_AWS_ACCESS_KEY_ID'
        password = 'YOUR_AWS_SECRET_ACCESS_KEY'
        role_arn = urllib.parse.quote_plus('YOUR_EXECUTION_ROLE_ARN')
        region_name = 'us-east-1'

        conn_string = '{0}://{1}:{2}@{3}:{4}?
role_arn={5}&region_name={6}'.format(conn_type, login, password, host, port,
 role_arn, region_name)
        print(conn_string)
```

2. Substitute the placeholders in *red*.
3. Run the following script to generate a connection string.

```
python3 mwaa_connection.py
```

Step four: Add the variables in Secrets Manager

The following section describes how to create the secret for a variable in Secrets Manager.

To create the secret

1. Open the AWS Secrets Manager console.

2. Choose **Store a new secret**.
3. Choose **Other type of secret**.
4. On the **Specify the key/value pairs to be stored in this secret** pane, choose **Plaintext**.
5. Add the variable value as **Plaintext** in the following format.

   ```
   "YOUR_VARIABLE_VALUE"
   ```

   For example, to specify an integer:

   ```
   14
   ```

   For example, to specify a string:

   ```
   "mystring"
   ```

6. For **Encryption key**, choose an AWS KMS key option from the dropdown list.
7. Enter a name in the text field for **Secret name** in the following format.

   ```
   airflow/variables/YOUR_VARIABLE_NAME
   ```

   For example:

   ```
   airflow/variables/test-variable
   ```

8. Choose **Next**.
9. On the **Configure secret** page, on the **Secret name and description** pane, do the following.

   a. For **Secret name**, provide a name for your secret.

   b. (Optional) For **Description**, provide a description for your secret.

   Choose **Next**.
10. On the **Configure rotation - optional** leave the default options and choose **Next**.
11. Repeat these steps in Secrets Manager for any additional variables you want to add.
12. On the **Review** page, review your secret, then choose **Store**.

Step five: Add the connection in Secrets Manager

The following section describes how to create the secret for your connection string URI in Secrets Manager.

To create the secret

1. Open the AWS Secrets Manager console.
2. Choose **Store a new secret**.
3. Choose **Other type of secret**.
4. On the **Specify the key/value pairs to be stored in this secret** pane, choose **Plaintext**.
5. Add the connection URI string as **Plaintext** in the following format.

   ```
   YOUR_CONNECTION_URI_STRING
   ```

   For example:

   ```
   mysql://288888a0-50a0-888-9a88-1a111aaa0000.a1.us-east-1.airflow.amazonaws.com
   %2Fhome?role_arn=arn%3Aaws%3Aiam%3A%3A001122332255%3Arole%2Fservice-role
   %2FAmazonMWAA-MyAirflowEnvironment-iAaaaA&region_name=us-east-1
   ```

   Warning

   Apache Airflow parses each of the values in the connection string. You must **not** use single nor double quotes, or it parses the connection as a single string.

6. For **Encryption key**, choose an AWS KMS key option from the dropdown list.
7. Enter a name in the text field for **Secret name** in the following format.

   ```
   airflow/connections/YOUR_CONNECTION_NAME
   ```

   For example:

   ```
   airflow/connections/myconn
   ```

8. Choose **Next**.
9. On the **Configure secret** page, on the **Secret name and description** pane, do the following.

a. For **Secret name**, provide a name for your secret.

b. (Optional) For **Description**, provide a description for your secret.

Choose **Next**.

10. On the **Configure rotation - optional** leave the default options and choose **Next**.
11. Repeat these steps in Secrets Manager for any additional variables you want to add.
12. On the **Review** page, review your secret, then choose **Store**.

Sample code

- Learn how to use the secret key for the Apache Airflow connection (`myconn`) on this page using the sample code at Using a secret key in AWS Secrets Manager for an Apache Airflow connection.
- Learn how to use the secret key for the Apache Airflow variable (`test-variable`) on this page using the sample code at Using a secret key in AWS Secrets Manager for an Apache Airflow variable.

Resources

- For more information about configuring Secrets Manager secrets using the console and the AWS CLI, refer to Create a secret in the *AWS Secrets Manager User Guide*.
- Use a Python script to migrate a large volume of Apache Airflow variables and connections to Secrets Manager in Move your Apache Airflow connections and variables to AWS Secrets Manager.

What's next?

- Learn how to generate a token to access the Apache Airflow UI in Accessing Apache Airflow.

Managing Amazon MWAA environments

The Amazon Managed Workflows for Apache Airflow console contains built-in options to configure private or public access to the Apache Airflow UI. It also contains built-in options to configure the environment size, when to scale workers, and Apache Airflow configuration options that you can use to override Apache Airflow configurations that are normally only accessible in `airflow.cfg`. This chapter describes how to use these configurations on the Amazon MWAA console.

Topics

- Configuring the Amazon MWAA environment class
- Cost allocation tags in Amazon MWAA
- Configuring Amazon MWAA worker automatic scaling
- Configuring Amazon MWAA webserver automatic scaling
- Using Apache Airflow configuration options on Amazon MWAA
- Update an Amazon MWAA environment
- Multi-team environments
- Managing teams
- Changing the Apache Airflow version
- Using a startup script with Amazon MWAA

Configuring the Amazon MWAA environment class

The environment class you choose for your Amazon MWAA environment determines the size of the AWS-managed AWS Fargate containers where the Celery Executor runs, and the AWS-managed Amazon Aurora PostgreSQL metadata database where the Apache Airflow schedulers creates task instances. This topic describes each Amazon MWAA environment class, and how to update the environment class on the Amazon MWAA console.

Sections

- Environment capabilities
- Apache Airflow schedulers

Environment capabilities

The following section contains the default concurrent Apache Airflow tasks, Random Access Memory (RAM), and the virtual centralized processing units (vCPUs) for each environment class. The concurrent tasks listed assume that task concurrency does not exceed the Apache Airflow worker capacity in the environment.

In the following table, DAG capacity refers to DAG definitions, not executions, and assumes that your DAGs are dynamic in a single Python file and written with Apache Airflow best practices.

Task executions depend on how many are scheduled simultaneously, and assumes that the number of DAG runs set to start at the same time doesn't exceed the default `max_dagruns_per_loop_to_schedule`, as well as the size and number of workers as detailed in this topic.

mw1.micro

- Up to 25 DAG capacity
- 3 concurrent tasks (by default)
- Components:
  - Webserver: 1 vCPU, 3GB RAM
  - Worker and scheduler: 1 vCPU, 3GB RAM
  - Database: 2 vCPU, 4GB RAM

> **Note**
>
> mw1.micro does not support auto-scaling.

mw1.small

- Up to 50 DAG capacity
- 5 concurrent tasks (by default)
- Components:
  - Webservers: 1 vCPU, 2GB RAM each
  - Workers: 1 vCPU, 2GB RAM each
  - Schedulers: 1 vCPU, 2GB RAM each

- Database: 2 vCPU, 4GB RAM

mw1.medium

- Up to 250 DAG capacity
- 10 concurrent tasks (by default)
- Components:
  - Webservers: 1 vCPU 2GB RAM each
  - Workers: 2 vCPU 4GB RAM each
  - Schedulers: 2 vCPU 4GB RAM each
  - Database: 2 vCPU 8GB RAM

mw1.large

- Up to 1000 DAG capacity
- 20 concurrent tasks (by default)
- Components:
  - Webservers: 2 vCPU 4GB RAM each
  - Workers: 4 vCPU 8GB RAM each
  - Schedulers: 4 vCPU 8GB RAM each
  - Database: 2 vCPU 8GB RAM

mw1.xlarge

- Up to 2000 DAG capacity
- 40 concurrent tasks (by default)
- Components:
  - Webservers: 4 vCPU 12GB RAM each
  - Workers: 8 vCPU 24GB RAM each
  - Schedulers: 8 vCPU 24GB RAM each
  - Database: 4 vCPU 32GB RAM

mw1.2xlarge

- Up to 4000 DAG capacity
- 80 concurrent tasks (by default)
- Components:
  - Webservers: 8 vCPU 24GB RAM each
  - Workers: 16 vCPU 48GB RAM each
  - Schedulers: 16 vCPU 48GB RAM each
  - Database: 8 vCPU 64GB RAM

You can use `celery.worker_autoscale` to increase tasks per worker. For more information, refer to the the section called “Example high performance use case”.

Apache Airflow schedulers

The following section contains the Apache Airflow scheduler options available on the Amazon MWAA, and how the number of schedulers affects the number of triggerers.

In Apache Airflow, a triggerer manages tasks which it defers until certain conditions specified using a trigger have been met. In Amazon MWAA the triggerer runs alongside the scheduler on the same Fargate task. Increasing the scheduler count correspondingly increases the number of available triggerers, optimizing how the environment manages deferred tasks. This ensures efficient handling of tasks, promptly scheduling them to run when conditions are satisfied.

Apache Airflow v3

- **v3** - For environments larger than mw1.micro, accepts values from 2 to 5. Defaults to 2 for all environment sizes except mw1.micro, which defaults to 1.

Apache Airflow v2

- **v2** - For environments larger than mw1.micro, accepts values from 2 to 5. Defaults to 2 for all environment sizes except mw1.micro, which defaults to 1.

Cost allocation tags in Amazon MWAA

A cost allocation tag is a key-value pair that you attach to an Amazon MWAA environment. After you activate a tag, AWS groups the environment's charges under that tag in your cost reports. You can then attribute the cost of each environment to a team, project, or cost center.

Topics

- Adding and activating cost allocation tags
- Viewing costs by tag
- Limitations
- Related resources

Adding and activating cost allocation tags

You tag your environments with the key that you want to report on, and then activate that key as a cost allocation tag. You activate a key one time for your organization, in the AWS Billing and Cost Management console of your management account.

To track Amazon MWAA costs with tags

1. Tag your environments with the key that you want to report on, such as `CostCenter` or `Team`. You can add tags when you create an environment, or add them to an existing environment, by using the Amazon MWAA console, the AWS CLI, or the TagResource API operation.
2. Open the Cost allocation tags page in the AWS Billing and Cost Management console of your management account, select your tag key, and then choose **Activate**. For more information about activating cost allocation tags, see Activating user-defined cost allocation tags in the *AWS Billing and Cost Management User Guide*.
3. In AWS Cost Explorer, for **Group by**, select your tag key. Your cost reports now group each environment's charges by the tag key.

To stop tracking costs with a tag, deactivate the tag key in the AWS Billing and Cost Management console.

> **Note**
>
> Tag keys and values are case sensitive. Activating or deactivating a tag applies only to usage that you incur afterward, and does not change cost data from earlier billing periods. After you activate a tag, it can take up to 24 hours for it to appear in your cost reports.

Viewing costs by tag

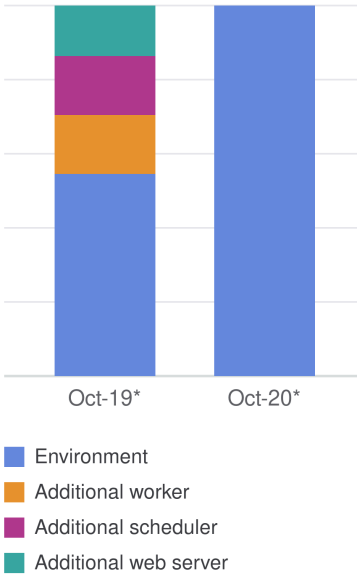
When you group your costs by a tag in AWS Cost Explorer or the AWS Cost and Usage Report, each tagged environment contributes its cost to that tag. After you onboard an environment, a single tag on the environment covers the cost of the environment and all of its components: its workers, its schedulers, and its web servers.

After you onboard an environment, grouping by resource in AWS Cost Explorer also aggregates these costs. Amazon MWAA attributes the usage for the environment, its workers, its schedulers, and its web servers to the single environment Amazon Resource Name (ARN), rather than to a separate resource for each component. Before onboarding, each additional worker and web server appears as its own resource in a per-resource view. After onboarding, they appear as a single line under the environment ARN.

The following chart compares an environment's cost before and after onboarding:

- Before onboarding, Amazon MWAA reports the cost as separate amounts for the environment, its additional workers, its additional schedulers, and its additional web servers.
- After onboarding, Amazon MWAA reports the same total cost as a single amount that it attributes to the environment ARN.

The chart shows these amounts as two stacked columns of equal height, because onboarding changes how Amazon MWAA attributes the cost, not the cost itself. The column for the day before onboarding has four segments: environment, additional worker, additional scheduler, and additional web server. The column for the day after has a single segment for the environment ARN. The dates and amounts are illustrative, not actual costs.



Limitations

Environments that you create after this cost-tracking model is available are onboarded automatically.

To fully onboard an existing environment, update the environment. The update onboards the environment and all of its components. For more information, see the section called “Update an environment”.

If you don't update the environment, Amazon MWAA still migrates the workers and web servers of recently created or recently updated environments to the new model as it replaces them. Environments that haven't been updated in a long time don't migrate their workers and web servers automatically. To migrate them, update the environment.

Until you onboard a component, Amazon MWAA reports the component's usage under its own resource rather than under the environment. As a result, a tag on the environment doesn't include that component's cost.

Related resources

For more information about cost allocation tags, see the following resources.

- For more information about filtering and grouping your costs by tag in AWS Cost Explorer, see Filtering the data that you want to view in the *AWS Cost Management User Guide*.

- For more information about how AWS uses tags to categorize your costs, see Use cost allocation tags in the *AWS Billing and Cost Management User Guide*.

Configuring Amazon MWAA worker automatic scaling

The auto scaling mechanism automatically increases the number of Apache Airflow workers in response to running and queued tasks on your Amazon Managed Workflows for Apache Airflow environment and disposes of extra workers when there are no more tasks queued or executing. This topic describes how you can configure auto scaling by specifying the maximum number of Apache Airflow workers that run on your environment using the Amazon MWAA console.

> **Note**
>
> Amazon MWAA uses Apache Airflow metrics to determine when additional Celery Executor workers are needed, and as required increases the number of Fargate workers up to the value specified by `max-workers`. As the additional workers complete work and work load decreases, Amazon MWAA removes them, thus downscaling back to the value set by `min-workers`.
>
> If workers pick up new tasks while downscaling, Amazon MWAA keeps the Fargate resource and does not remove the worker. For more information, refer to How Amazon MWAA auto scaling works.

Sections

- How worker scaling works
- Using the Amazon MWAA console
- Example high performance use case
- Troubleshooting tasks stuck in the running state
- What's next?

How worker scaling works

Amazon MWAA uses `RunningTasks` and `QueuedTasks` metrics, where *(tasks running + tasks queued) / (tasks per worker) = (required workers)*. If the required number of workers is greater than

the current number of workers, Amazon MWAA will add Fargate worker containers to that value, up to the maximum value specified by `max-workers`.

As the workload decreases and the `RunningTasks` and `QueuedTasks` metric sum reduces, Amazon MWAA requests Fargate to scale down the workers for the environment. Any workers which still completing work remain protected during downscaling until they complete their work. Depending on the workload, tasks might be queued while workers downscale.

Using the Amazon MWAA console

You can choose the maximum number of workers that can run on your environment concurrently on the Amazon MWAA console. By default, you can specify a maximum value up to 25.

To configure the number of workers

1. Open the Environments page on the Amazon MWAA console.
2. Choose an environment.
3. Choose **Edit**.
4. Choose **Next**.
5. On the **Environment class** pane, enter a value in **Maximum worker count**.
6. Choose **Save**.

> **Note**
>
> It can take a few minutes before changes take effect on your environment.

Example high performance use case

The following section describes the type of configurations you can use to enable high performance and parallelism on an environment.

On-premise Apache Airflow

Typically, in an on-premise Apache Airflow platform, you configure task parallelism, auto scaling, and concurrency settings in your `airflow.cfg` file:

- `core.parallelism` – The maximum number of task instances that can run simultaneously per scheduler.
- `core.dag_concurrency` – The maximum concurrency for DAGs (not workers).
- `celery.worker_autoscale` – The maximum and minimum number of tasks that can run concurrently on any worker.

For example, if `core.parallelism` was set to `100` and `core.dag_concurrency` was set to 7, you can only run a total of 14 tasks concurrently if you had 2 DAGs. Given, each DAG is set to run only seven tasks concurrently (in `core.dag_concurrency`), even though overall parallelism is set to `100` (in `core.parallelism`).

> **Note**
>
> `core.dag_concurrency` is not available in Apache Airflow v3.

On an Amazon MWAA environment

On an Amazon MWAA environment, you can configure these settings directly on the Amazon MWAA console using Using Apache Airflow configuration options on Amazon MWAA, Configuring the Amazon MWAA environment class, and the **Maximum worker count** auto scaling mechanism. While `core.dag_concurrency` is not available in the drop down list as an **Apache Airflow configuration option** on the Amazon MWAA console, you can add it as a custom Apache Airflow configuration option.

Let's say, when you created your environment, you chose the following settings:

1. The **mw1.small** environment class which controls the maximum number of concurrent tasks each worker can run by default and the vCPU of containers.
2. The default setting of `10` workers in **Maximum worker count**.
3. An Apache Airflow configuration option for `celery.worker_autoscale` of `5,5` tasks per worker.

This means you can run 50 concurrent tasks in your environment. Any tasks beyond 50 are queued, and wait for the running tasks to complete.

Run more concurrent tasks. You can modify your environment to run more tasks concurrently using the following configurations:

1. Increase the maximum number of concurrent tasks each worker can run by default and the vCPU of containers by choosing the `mw1.medium` (10 concurrent tasks by default) environment class.
2. Add `celery.worker_autoscale` as an Apache Airflow configuration option.
3. Increase the **Maximum worker count**. In this example, increasing maximum workers from `10` to `20` doubles the number of concurrent tasks the environment can run.

Specify Minimum workers. You can also specify the minimum and maximum number of Apache Airflow workers that run in your environment using the AWS Command Line Interface (AWS CLI). For example:

```
aws mwaa update-environment --max-workers 10 --min-workers 10 --
name YOUR_ENVIRONMENT_NAME
```

To learn more, refer to the update-environment command in the AWS CLI.

Troubleshooting tasks stuck in the running state

In rare cases, Apache Airflow might think there are tasks still running. To resolve this issue, you need to clear the stranded task in your Apache Airflow UI. For more information, refer to the Troubleshooting Amazon Managed Workflows for Apache Airflow troubleshooting topic.

What's next?

- Learn more about the best practices we recommend to tune the performance of your environment in Performance tuning for Apache Airflow on Amazon MWAA.

Configuring Amazon MWAA webserver automatic scaling

For environments running Apache Airflow v2.2.2 and later, Amazon MWAA dynamically scales your webservers to handle fluctuating workloads, which in turn prevents performance issues during peak loads. By automatically scaling the number of webservers based on CPU utilization and active connection count, Amazon MWAA ensures that your Apache Airflow environment can seamlessly

accommodate increased demand, whether from REST API requests, CLI usage, or more concurrent Apache Airflow user interface users.

Sections

- How webserver scaling works
- Using the Amazon MWAA console

How webserver scaling works

Amazon MWAA uses the container metric, `CPUUtilization`, and the load balancer metric, `ActiveConnectionCount`, to determine if scaling the webservers is required based on the amount of traffic. If `CPUUtilization` is greater than 70 or `ActiveConnectionCount` is greater than 15, Amazon MWAA will add additional Fargate webserver containers up to the maximum value specified by `MaxWebservers`.

As traffic decreases and the `CPUUtilization` and `ActiveConnectionCount` values reduce, Amazon MWAA requests Fargate to scale down the webserver containers for the environment to the minimum value set by `MinimumWebservers`.

Using the Amazon MWAA console

You can choose the number of webservers that can run on your environment concurrently on the Amazon MWAA console. By default, the minimum number of webservers is two, and the maximum number of webservers is five.

To configure the number of webservers

1. Open the Environments page on the Amazon MWAA console.
2. Choose an environment.
3. Choose **Edit**.
4. Choose **Next**.
5. On the **Environment class** pane, enter a value in **Maximum webserver count**.
6. Next, enter a value in **Minimum webserver count**.
7. Choose **Save**.

Note

It can take a few minutes before changes take effect on your environment.

Using Apache Airflow configuration options on Amazon MWAA

Apache Airflow configuration options can be attached to your Amazon Managed Workflows for Apache Airflow environment as environment variables. You can choose from the suggested dropdown list or specify custom configuration options for your Apache Airflow version on the Amazon MWAA console. This topic describes the Apache Airflow configuration options available and how to use these options to override Apache Airflow configuration settings on your environment.

Contents

- Prerequisites
- How it works
- Using configuration options to load plugins
- Configuration options overview
  - Apache Airflow configuration options
  - Apache Airflow reference
  - Using the Amazon MWAA console
- Configuration reference
  - Email configurations
    - Customizing the email subject and content
  - Task configurations
  - Scheduler configurations
  - Worker configurations
  - Webserver configurations
  - Triggerer configurations
- Unsupported configurations
- Examples and sample code
  - Example DAG

- Example email notification settings
- What's next?

Prerequisites

You'll need the following before you can complete the steps on this page.

- **Permissions** — Your AWS account must have been granted access by your administrator to the AmazonMWAAFullConsoleAccess access control policy for your environment. In addition, your Amazon MWAA environment must be permitted by your execution role to access the AWS resources used by your environment.
- **Access** — If you require access to public repositories to install dependencies directly on the webserver, your environment must be configured with **public network** webserver access. For more information, refer to the section called “Apache Airflow access modes”.
- **Amazon S3 configuration** — The Amazon S3 bucket used to store your DAGs, custom plugins in `plugins.zip`, and Python dependencies in `requirements.txt` must be configured with *Public Access Blocked* and *Versioning Enabled*.

How it works

When you create an environment, Amazon MWAA attaches the configuration settings you specify on the Amazon MWAA console in **Airflow configuration options** as environment variables to the AWS Fargate container for your environment. If you're using a setting of the same name in `airflow.cfg`, the options you specify on the Amazon MWAA console override the values in `airflow.cfg`.

While we don't expose the `airflow.cfg` in the Apache Airflow UI of an Amazon MWAA environment by default, you can change the Apache Airflow configuration options directly on the Amazon MWAA console, including setting `webserver.expose_config` to expose the configurations.

Using configuration options to load plugins

By default in Apache Airflow v2 and later, plugins are configured to be "lazily" loaded using the `core.lazy_load_plugins : True` setting. If you're using custom plugins, you must add `core.lazy_load_plugins : False` as an Apache Airflow configuration option to load plugins at the start of each Airflow process to override the default setting.

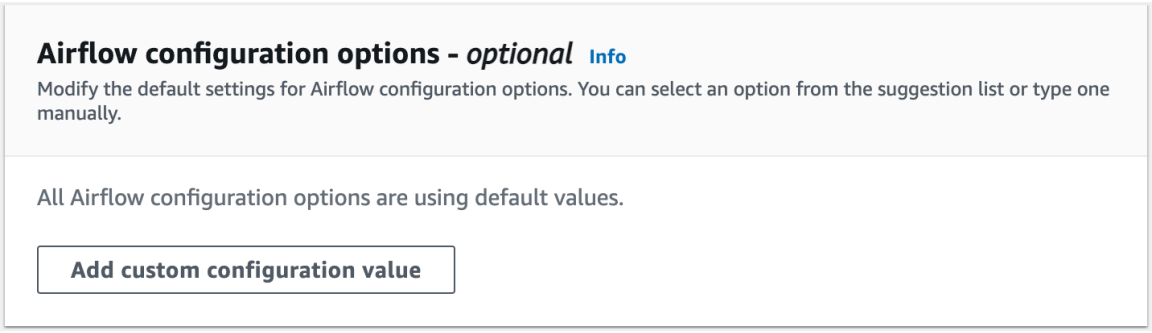
Configuration options overview

When you add a configuration on the Amazon MWAA console, Amazon MWAA writes the configuration as an environment variable.

- **Listed options**. You can choose from one of the configuration settings available for your Apache Airflow version in the dropdown list. For example, `dag_concurrency : 16`. The configuration setting is translated to your environment's Fargate container as `AIRFLOW__CORE__DAG_CONCURRENCY : 16`
- **Custom options**. You can also specify Airflow configuration options that are not listed for your Apache Airflow version in the dropdown list. For example, `foo.user : YOUR_USER_NAME`. The configuration setting is translated to your environment's Fargate container as `AIRFLOW__FOO__USER : YOUR_USER_NAME`

Apache Airflow configuration options

The following image depicts where you can customize the **Apache Airflow configuration options** on the Amazon MWAA console.



Apache Airflow reference

For a list of configuration options supported by Apache Airflow, refer to Configuration Reference in the *Apache Airflow reference guide*. To access the options for the version of Apache Airflow you are running on Amazon MWAA, select the version from the drop down list.

Using the Amazon MWAA console

The following procedure walks you through the steps of adding an Airflow configuration option to your environment.

1. Open the Environments page on the Amazon MWAA console.

2. Choose an environment.
3. Choose **Edit**.
4. Choose **Next**.
5. Choose **Add custom configuration** in the **Airflow configuration options** pane.
6. Choose a configuration from the dropdown list and enter a value, or enter a custom configuration and enter a value.
7. Choose **Add custom configuration** for each configuration you want to add.
8. Choose **Save**.

Configuration reference

The following section contains the list of available Apache Airflow configurations in the dropdown list on the Amazon MWAA console.

Email configurations

The following list displays the Airflow email notification configuration options available on Amazon MWAA for Apache Airflow v2 and v3.

We recommend using port 587 for SMTP traffic. By default, AWS blocks outbound SMTP traffic on port 25 of all Amazon EC2 instances. If you want to send outbound traffic on port 25, you can request for this restriction to be removed.

Airflow configuration option	Description	Example value
email.email_backend	The Apache Airflow utility used for email notifications in email_backend.	airflow.utils.email.send_email_smtp
email.subject_template	The path to a Jinja2 template file used to render the **subject** of Apache Airflow alert emails in subject_template. If this option is not set, Apache Airflow uses the default	/usr/local/airflow/dags/my_subject_template_file

Airflow configuration option	Description	Example value
	subject `Airflow alert: {{ti}}`.	
email.html_content_template	The path to a Jinja2 template file used to render the **body** of Apache Airflow alert emails in html_content_template. If this option is not set, Apache Airflow uses the default built-in body template.	/usr/local/airflow/dags/my_html_content_template_file
smtp.smtp_host	The name of the outbound server used for the email address in smtp_host.	localhost
smtp.smtp_starttls	Transport Layer Security (TLS) is used to encrypt the email over the internet in smtp_starttls.	True
smtp.smtp_ssl	Secure Sockets Layer (SSL) is used to connect the server and email client in smtp_ssl.	False
smtp.smtp_port	The Transmission Control Protocol (TCP) port designated to the server in smtp_port.	587
smtp.smtp_mail_from	The outbound email address in smtp_mail_from.	myemail@domain.com

SMTP port and encryption settings

Use 587 with STARTTLS (`smtp_starttls` : `True`, `smtp_ssl` : `False`), or 465 with the SSL/TLS wrapper (`smtp_ssl` : `True`).

Customizing the email subject and content

By default, Apache Airflow sends alert emails using its built-in subject and body templates. You can override these defaults to send customized notifications. Set the `email.subject_template` and `email.html_content_template` configuration options to point to your own Jinja2 template files. Jinja2 is an open-source templating engine; for more information, see the Jinja website.

Note

Customizing email content requires a working email backend. Configure your SMTP settings as described in the Email configurations table before you customize the subject and body. For example, you can use Amazon Simple Email Service (Amazon SES) as the SMTP host.

If you do not set `subject_template` and `html_content_template`, Apache Airflow uses the following base templates:

Default subject:

```
Airflow alert: {{ti}}
```

Default HTML content:

```
Try {{try_number}} out of {{max_tries + 1}}<br>
Exception:<br>{{exception_html}}<br>
Log: <a href="{{ti.log_url}}">Link</a><br>
Host: {{ti.hostname}}<br>
Mark success: <a href="{{ti.mark_success_url}}">Link</a><br>
```

To use custom templates on Amazon MWAA

1. Create your subject and HTML content template files. The following are examples.

   `my_subject_template_file`

```
Production Airflow alert: {{ti}}
```

`my_html_content_template_file`

```
Exception:<br>{{exception_html}}<br>
State: {{ti.state}}<br>
Mark success: <a href="{{ti.mark_success_url}}">Link</a><br>
```

2. Upload the template files to the DAGs folder in your environment's Amazon S3 bucket. Amazon MWAA syncs this folder to `/usr/local/airflow/dags/` on the environment's containers.
3. On the Amazon MWAA console, add the `email.subject_template` and `email.html_content_template` configuration options shown in the preceding Email configurations table, pointing each to the container path of the corresponding template file:

Airflow configuration option	Value
email.subject_template	/usr/local/airflow/dags/my_subject_template_file
email.html_content_template	/usr/local/airflow/dags/my_html_content_template_file

4. Save your changes. After the environment finishes updating, alert emails use your custom subject and body templates.

For more information about sending email through Amazon SES as the SMTP host, see the AWS re:Post article How to send emails from MWAA DAG tasks using SES as the SMTP host.

Task configurations

The following list displays the configurations available in the dropdown list for Airflow tasks on Amazon MWAA for Apache Airflow v2 and v3.

Airflow configuration option	Description	Example value

Airflow configuration option	Description	Example value
core.default_task_retries	The number of times to retry an Apache Airflow task in default_task_retries.	3
core.parallelism	The maximum number of task instances that can run simultaneously across the entire environment in parallel (parallelism).	40

Scheduler configurations

The following list displays the Apache Airflow scheduler configurations available in the dropdown list on Amazon MWAA for Apache Airflow v2 and v3.

Airflow configuration option	Description	Example value
scheduler.catchup_by_default	Tells the scheduler to create a DAG run to "catch up" to the specific time interval in catchup_by_default.	False
scheduler.scheduler_zombie_task_threshold **Note** Not available in Apache Airflow v3.	Tells the scheduler whether to mark the task instance as failed and reschedule the task in scheduler_zombie_task_threshold.	300

Worker configurations

The following list displays the Airflow worker configurations available in the dropdown list on Amazon MWAA for Apache Airflow v2 and v3.

Airflow configuration option	Description	Example value
celery.worker_autoscale	The maximum and minimum number of tasks that can run concurrently on any worker using the Celery Executor in worker_autoscale. Value must be comma-separated in the following order: `max_concurrency,min_concurrency` .	16,12
mwaa.worker_replacement_threshold_seconds	Sets the number of seconds after which Amazon MWAA terminates and replaces an unhealthy worker. When set to a value greater than 0, Amazon MWAA replaces any worker that has not returned to a healthy state within the configured threshold. If set to `0` (the default), unhealthy workers are only replaced when they reach red health status.	300

Webserver configurations

The following list displays the Apache Airflow webserver configurations available in the dropdown list on Amazon MWAA for Apache Airflow v2 and v3.

Airflow configuration option	Description	Example value
webserver.default_ui_timezone **Note** Not available in Apache Airflow v3.	The default Apache Airflow UI datetime setting in default_ui_timezone. **Note** Setting the `default_ui_timezone` option does not change the time zone in which your DAGs are scheduled to run. To change the time zone for your DAGs, you can use a custom plugin. For more information, refer to the section called "Changing a DAG's timezone".	America/New_York

Triggerer configurations

The following list displays the Apache Airflow triggerer configurations available on Amazon MWAA for Apache Airflow v2 and v3.

Airflow configuration option	Description	Example value
mwaa.triggerer_enabled	Used for activating and deactivating the triggerer on Amazon MWAA. By default,	True

Airflow configuration option	Description	Example value
	this value is set to `True`. If set to `False`, Amazon MWAA will not start any triggerer processes on schedulers.	
triggerer.default_capacity (in v2) triggerer.capacity (in v3)	Defines the number of triggers each triggerer can run in parallel. On Amazon MWAA, this capacity is set for each triggerer and for each scheduler as both component s run alongside each other. The default per scheduler is set to `60`, `125`, `250`, `500`, and `1000` for small, medium, large, xlarge, and 2xlarge instances, respectively.	125

Unsupported configurations

The following Apache Airflow configuration options are not available in Amazon MWAA. You cannot set or override these options using the Amazon MWAA console or API.

Airflow configuration option	Apache Airflow version	Default value	Reason
core.multi_team	3.2	False	Amazon MWAA sets this option when you add teams. You can't set it directly. See Multi-team environments.
triggerer .queues_enabled	3.2	False	Amazon MWAA does not support triggerer queue assignment.

Airflow configuration option	Apache Airflow version	Default value	Reason
			Enabling this option causes deferred tasks to hang indefinitely.

Examples and sample code

Example DAG

You can use the following DAG to print your `email_backend` Apache Airflow configuration options. To run in response to Amazon MWAA events, copy the code to your environment's DAGs folder on your Amazon S3 storage bucket.

```
from airflow.decorators import dag
from airflow.operators.python import PythonOperator
from datetime import datetime

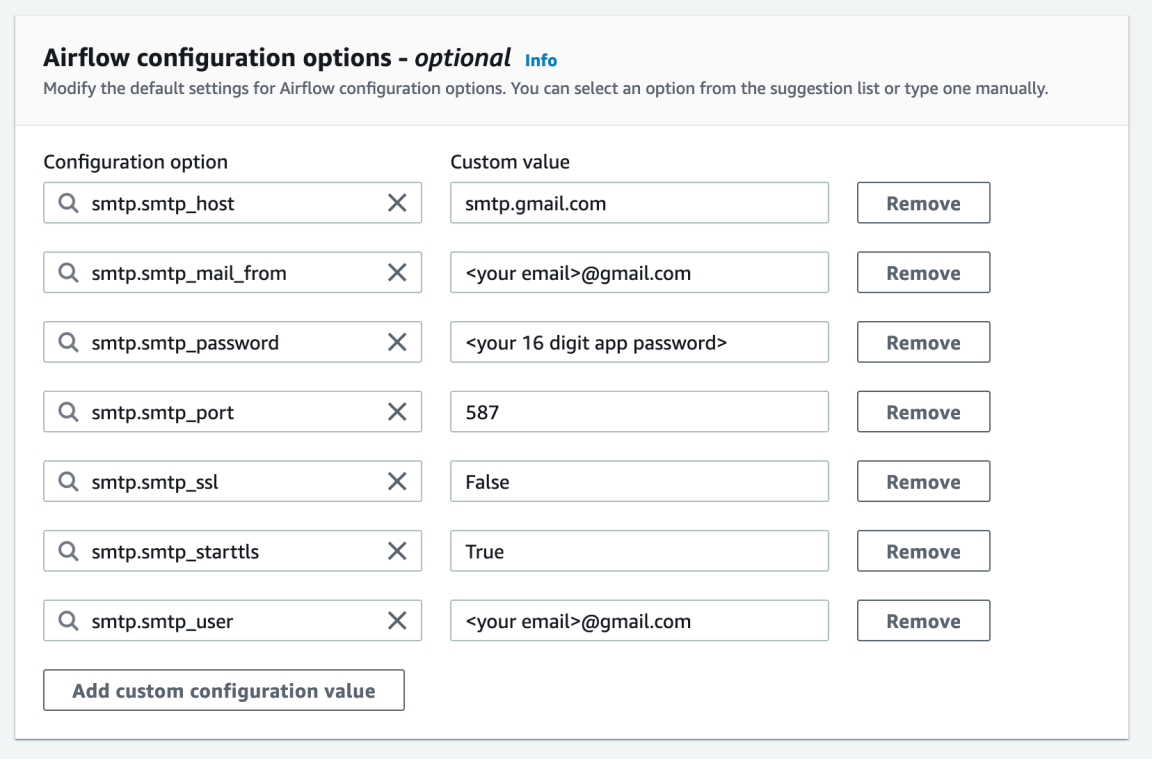

def print_var(**kwargs):
    email_backend = kwargs['conf'].get(section='email', key='email_backend')
    print(email_backend)
    return email_backend


@dag(
    dag_id="print_env_variable_example",
    schedule_interval=None,
    start_date=datetime(yyyy, m, d),
    catchup=False,
)
def print_variable_dag():
    email_backend_test = PythonOperator(
        task_id="email_backend_test",
        python_callable=print_var,
        provide_context=True
    )
```

```
print_variable_test = print_variable_dag()
```

Example email notification settings

The following Apache Airflow configuration options can be used for a Gmail.com email account using an app password. For more information, refer to Sign in using app passwords in the *Gmail Help reference guide*.



What's next?

- Learn how to upload your DAG folder to your Amazon S3 bucket in Adding or updating DAGs.

Update an Amazon MWAA environment

> **Note**
>
> Amazon MWAA graceful updates are not yet supported in the Canada West (Calgary) and Asia Pacific (Malaysia) regions.

Amazon MWAA environment updates apply the latest changes and security patches. You can also edit existing configurations and upgrade the Apache Airflow version. This guide describes the steps to update an Amazon MWAA environment.

Contents

- Before you begin
- Worker replacement strategy
- How environment updates work
  - Forced updates
  - Graceful updates
  - If an update fails
- Update environment resources
- Update an environment
  - Step one: Specify details
  - Step two: Configure advanced settings
  - Step three: Review and save

Before you begin

- The VPC network you specified for your environment cannot be modified after the environment is created.
- You need an Amazon S3 bucket configured to **Block all public access**, with **Bucket Versioning** enabled.
- You need an AWS account with permissions to use Amazon MWAA, and permission in AWS Identity and Access Management (IAM) to create IAM roles. If you choose the **Private network** access mode for the Apache Airflow webserver, which limits Apache Airflow access within your Amazon VPC, you'll need permission in IAM to create Amazon VPC endpoints.
- To enable Graceful environment updates, you need to upgrade to Apache Airflow version 2.4.3 or higher. To upgrade the Airflow version, refer to Changing the Apache Airflow version.

Worker replacement strategy

You can choose a worker replacement strategy to control how Amazon MWAA handles active workers during an environment update. You can select one of the following strategies:

Forced updates

Forced update is the default worker replacement strategy. Forced updates immediately stop all active workers, causing running tasks to fail during the update.

Graceful updates

Graceful updates allow workers to continue running tasks for up to 12 hours before shutting down. It prevents tasks failing due to update interruptions, as long as they finish under 12 hours. New tasks are routed to updated workers.

To enable Graceful updates on an existing environment, you must complete one **Forced update** and ensure the environment is on Apache Airflow version 2.4.3 or higher.

> **Note**
>
> If you perform an update while your environment is in `MAINTENANCE` status, the worker replacement strategy for any ongoing environment update switches from `GRACEFUL` to `FORCED`. Your update is performed after maintenance is complete.

> **Important**
>
> `WorkerReplacementStrategy` defaults to `FORCED` and resets on every environment update—it does not carry over from a previous update.
> To continue using `GRACEFUL`, you must explicitly set `WorkerReplacementStrategy=GRACEFUL` each time you update your environment.

How environment updates work

Your environment stays available while Amazon MWAA applies an update, and its status shows as `UPDATING`. During the update, the Apache Airflow UI might be unavailable for a short time while Amazon MWAA replaces your webservers.

How your workers and running tasks behave depends on the worker replacement strategy that you choose.

Forced updates

When you use the `FORCED` worker replacement strategy, Amazon MWAA immediately stops your active workers and replaces them with workers that use your updated environment configuration.

When Amazon MWAA replaces a worker, it stops every task running on that worker. Those tasks fail and run again only if you've configured retries for them. To reduce the impact on your workflows, schedule forced updates when few or no tasks are running. You can also set task retries, for example with the `core.default_task_retries` Apache Airflow configuration option.

Graceful updates

When you use the `GRACEFUL` worker replacement strategy, Amazon MWAA provisions new workers with your updated configuration and sends new tasks to them. Existing workers stop accepting new tasks but get up to 12 hours to finish tasks that started before the update.

Amazon MWAA replaces each existing worker after its tasks finish or the 12-hour limit passes, whichever comes first. Amazon MWAA stops any tasks that are still running after 12 hours.

If an update fails

If the new workers with the updated configuration don't become healthy, Amazon MWAA stops the update and sets your environment's status to `UPDATE_FAILED`. New workers are provisioned with the last known stable configuration, so your workflows continue without the changes you requested.

If you have not configured versioning for your Amazon S3 bucket, or Amazon MWAA is not able to revert to the last stable configuration, your environment is moved to `UNAVAILABLE`. Your workflows will not continue until you upload new stable configurations to Amazon S3 and update your environment to use them.

To find out why the update failed, check your environment's CloudWatch Logs. For more information about troubleshooting, see Troubleshooting Amazon Managed Workflows for Apache Airflow.

For more information, see the following:

- Worker replacement strategy
- Update environment resources
- Using Apache Airflow configuration options on Amazon MWAA

Update environment resources

Amazon MWAA environment updates use the existing environment configuration by default. To update the environment without changing your current configuration:

1. Open the Environments page on the Amazon MWAA console.
2. From the **Environments** list, choose the environment that you want to update.
3. On the environment page, choose **Edit** to edit the environment.
4. Choose **Next** until you are on the **Review and save** page.
5. On the **Review and save** page, review your changes, then choose **Save**.

Update an environment

The following section describes the steps to update an Amazon MWAA environment.

Step one: Specify details

To specify details for the environment

1. Open the Environments page on the Amazon MWAA console.
2. From the **Environments** list, choose the environment that you want to update.
3. On the environment page, choose **Edit** to edit the environment.
4. In the Environment details section, for Airflow version, choose the new Apache Airflow version number that you want to upgrade the environment to from the dropdown list.

> **Note**
>
> Before you upgrade, make sure that your DAGs and other workflow resources are compatible with the new Apache Airflow version. For more information, refer to Changing the Apache Airflow version.

5. Under **DAG code in Amazon S3** specify the following:

    a. **S3 Bucket**. Choose **Browse S3** and select your Amazon S3 bucket, or enter the Amazon S3 URI.

    b. **DAGs folder**. Choose **Browse S3** and select the `dags` folder in your Amazon S3 bucket, or enter the Amazon S3 URI.

    c. **Plugins file - *optional***. Choose **Browse S3** and select the `plugins.zip` file on your Amazon S3 bucket, or enter the Amazon S3 URI.

    d. **Requirements file - *optional***. Choose **Browse S3** and select the `requirements.txt` file on your Amazon S3 bucket, or enter the Amazon S3 URI.

    e. **Startup script file - *optional***, Choose **Browse S3** and select the script file on your Amazon S3 bucket, or enter the Amazon S3 URI.

6. Choose **Next**.

Step two: Configure advanced settings

To configure advanced settings

1. Under **webserver access**, select your preferred Apache Airflow access mode:

    a. **Private network**. This limits access of the Apache Airflow UI to users *within your Amazon VPC* that have been granted access to the IAM policy for your environment. You need permission to create Amazon VPC endpoints for this step.

    > **Note**
    >
    > Choose the **Private network** option if your Apache Airflow UI is only accessed within a corporate network, and you do not require access to public repositories for webserver requirements installation. If you choose this access mode option, you need to create a mechanism to access your Apache Airflow webserver in your Amazon VPC. For more information, refer to Accessing the VPC endpoint for your Apache Airflow webserver (private network access).

    b. **Public network**. This allows the Apache Airflow UI to be accessed over the internet by users granted access to the IAM policy for your environment.

    c. **Both public and private network access**. Available for Apache Airflow version 3.2.1 and later. This allows the Apache Airflow UI to be accessed over the internet, while workers

communicate with the webserver through a private VPC endpoint. Choose this option if the Amazon VPC hosting your environment lacks internet access.

2. Under **Security groups**, choose the security group used to secure your Amazon VPC:

   a. By default, Amazon MWAA creates a security group in your Amazon VPC with specific inbound and outbound rules in **Create new security group**.

   b. **Optional**. Deselect the check box in **Create new security group** to select up to 5 security groups.

   > **Note**
   >
   > An existing Amazon VPC security group must be configured with specific inbound and outbound rules to allow network traffic. To learn more, refer to Security in your VPC on Amazon MWAA.

3. Under **Environment class**, choose an environment class.

   We recommend choosing the smallest size necessary to support your workload. You can change the environment class at any time.

4. For **Maximum worker count**, specify the maximum number of Apache Airflow workers to run in the environment.

   For more information, refer to Example high performance use case.

5. Specify the **Maximum web server count** and **Minimum webserver count** to configure how Amazon MWAA scales the Apache Airflow web servers in your environment.

   For more information about web server automatic scaling, refer to the section called "Configuring webserver auto scaling".

6. Under **Encryption**, choose a data encryption option:

   a. By default, Amazon MWAA uses an AWS-owned key to encrypt your data.

   b. **Optional**. Choose **Customize encryption settings (advanced)** to choose a different AWS KMS key. If you choose to specify a Customer-managed key in this step, you must specify an AWS KMS key ID or ARN. AWS KMS aliases and multi-region keys are not supported by Amazon MWAA. If you specified an Amazon S3 key for server-side encryption on your Amazon S3 bucket, you must specify the same key for your Amazon MWAA environment.

> **Note**
>
> You must have permissions to the key to select it on the Amazon MWAA console. You must also grant permissions for Amazon MWAA to use the key by attaching the policy described in Attach key policy.

7. **Recommended**. Under **Monitoring**, choose one or more log categories for **Airflow logging configuration** to send Apache Airflow logs to CloudWatch Logs:

   a. **Airflow task logs**. Choose the type of Apache Airflow task logs to send to CloudWatch Logs in **Log level**.

   b. **Airflow webserver logs**. Choose the type of Apache Airflow webserver logs to send to CloudWatch Logs in **Log level**.

   c. **Airflow scheduler logs**. Choose the type of Apache Airflow scheduler logs to send to CloudWatch Logs in **Log level**.

   d. **Airflow worker logs**. Choose the type of Apache Airflow worker logs to send to CloudWatch Logs in **Log level**.

   e. **Airflow DAG processing logs**. Choose the type of Apache Airflow DAG processing logs to send to CloudWatch Logs in **Log level**.

8. **Optional**. For **Airflow configuration options**, choose **Add custom configuration option**.

   You can choose from the suggested dropdown list of Apache Airflow configuration options for your Apache Airflow version, or specify custom configuration options. For example, `core.default_task_retries : 3`.

9. Under **Permissions**, choose an execution role:

   a. By default, Amazon MWAA creates an execution role in **Create a new role**. You must have permission to create IAM roles to use this option.

   b. **Optional**. Choose **Enter role ARN** to enter the Amazon Resource Name (ARN) of an existing execution role.

10. Under **Update specifications**, choose a Worker replacement strategy to control how active workers are handled during an update.

11. Choose **Next**.

Step three: Review and save

To review an environment summary

- Review the environment summary, choose **Save**.

> **Note**
>
> It takes about twenty to thirty minutes to update an environment using forced updates. Graceful environment updates might take up to twelve hours to complete, as it waits for your ongoing tasks to finish.

Multi-team environments

In a multi-team environment, several teams share a single Amazon Managed Workflows for Apache Airflow environment. Each team has its own DAG folder in Amazon S3, and IAM Identity Center groups control which teams a user can access. Amazon MWAA multi-team environments are based on the Apache Airflow multi-team feature. For more information, see Multi-Team on the Apache Airflow website.

> **Important**
>
> Amazon MWAA multi-team support is based on the Apache Airflow multi-team feature. This Apache Airflow feature is experimental and in preview as of Apache Airflow version 3.3.1. The feature is not complete, and its behavior and APIs can change in breaking ways, or be removed, between releases without notice. We recommend that you use multi-team environments with caution and that you don't rely on them for production-critical workloads. For more information, see Experimental features on the Apache Airflow website.

Contents

- Prerequisites
- Key concepts
  - Teams
  - Global and team DAG folders
  - DAG visibility

- Team isolation and the airflow:TeamNames condition key
- Choosing a multi-team environment or separate environments
- Limitations
- Next steps

Prerequisites

- The environment runs Apache Airflow version 3.3.1 or later.
- IAM Identity Center integration is enabled on the environment. For more information, see Using IAM Identity Center with Amazon Managed Workflows for Apache Airflow.
- Each team has an IAM Identity Center group whose display name exactly matches the team name. For example, a team named `data-engineering` needs an IAM Identity Center group with the display name `data-engineering`. Matching is case-sensitive, so a team named `data-engineering` doesn't match a group named `Data-Engineering`.

Key concepts

The following concepts describe how teams, DAG folders, and access control work together in a multi-team environment.

Teams

A *team* is a named group with its own DAG folder in Amazon S3. You can configure up to 20 teams per environment. Team names must meet the following requirements:

- Between 3 and 50 characters long.
- Starts with a letter or number.
- Contains only letters (`a-z`, `A-Z`), numbers (`0-9`), hyphens (`-`), and underscores (`_`).
- Unique within the environment.

Each team maps to the IAM Identity Center group with the same display name. When a user signs in, Amazon MWAA uses the user's IAM Identity Center group memberships to determine which teams they belong to. If no IAM Identity Center group matches a team name, the team has no members, and no users can access it.

Global and team DAG folders

A multi-team environment has two kinds of DAG folders:

- **Global DAG folder** – The DAG folder that you specify when you create the environment. All users can see and run the DAGs in this folder. Put DAGs that several teams need here.
- **Team DAG folders** – Each team has its own folder in Amazon S3. Only members of the team can see and run the DAGs in it. A team folder can be in the environment's bucket or in a different bucket. If you use a different bucket, the environment's execution role must have permission to read from it. For more information, see Amazon MWAA execution role.

You configure plugins and the `requirements.txt` file at the environment level, and they apply to all teams.

DAG folders can't overlap. A team folder can't be inside the global folder or another team's folder, and the global folder can't be inside a team folder. Amazon MWAA compares folder paths as Amazon S3 prefixes, so the trailing slash matters:

- `s3://amzn-s3-demo-bucket/dags/` and `s3://amzn-s3-demo-bucket/dags/team-a/` overlap, because the first path is a prefix of the second.
- `s3://amzn-s3-demo-bucket/dags/` and `s3://amzn-s3-demo-bucket/dags-team-a/` don't overlap.
- `s3://amzn-s3-demo-bucket/teams/team-a/` and `s3://amzn-s3-demo-bucket/teams/team-b/` don't overlap.

If folders overlap, the Amazon MWAA console shows a validation error, and the API rejects the request.

DAG visibility

When you sign in to the Apache Airflow UI in a multi-team environment, you see the DAGs in the global DAG folder and the DAGs of the teams you belong to. You can't see or access the DAGs of other teams.

Team isolation and the `airflow:TeamNames` condition key

How you control access to DAGs and other Apache Airflow resources depends on whether the environment has teams. Don't confuse the `airflow:TeamNames` IAM condition key with teams.

The condition key is available in every environment that uses IAM Identity Center, but you don't need it to isolate teams.

Environments with teams

When you add teams, Amazon MWAA turns on the Apache Airflow multi-team mode. Each DAG, connection, and variable either belongs to a team or is shared by all users. Apache Airflow isolates resources automatically based on this association. A user can access shared resources and the resources of the teams whose IAM Identity Center groups they belong to. You don't need to add `airflow:TeamNames` conditions to your policies for this isolation.

The user's Permission Set policy must still allow access to the Apache Airflow resources. You can use `airflow:TeamNames` to narrow access further, but a policy can't give a user access to a team they don't belong to.

Environments without teams

Apache Airflow multi-team mode is off, and resources don't belong to teams. The `airflow:TeamNames` condition key still contains the display names of the IAM Identity Center groups that the user belongs to. If you want finer-grained access control without adding teams, use `airflow:TeamNames` in your Permission Set policies. For example, combine it with the `airflow:ResourceType` and `airflow:ResourceId` condition keys so that only members of a specific IAM Identity Center group can access specific DAGs.

For how Apache Airflow associates resources with teams, see Multi-Team on the Apache Airflow website. For policy examples, see Using IAM Identity Center with Amazon Managed Workflows for Apache Airflow and Accessing an Amazon MWAA environment.

Choosing a multi-team environment or separate environments

A multi-team environment is a good fit when teams can share the same Apache Airflow version and environment configuration. It provides the following:

- **Lower cost** – Teams share one environment's infrastructure instead of each paying for their own.
- **Simpler management** – You configure environment settings, networking, and the Apache Airflow version once for all teams.
- **Central governance** – The same security policies, monitoring, and compliance settings apply to all teams.

- **DAG isolation** – Each team manages its own DAGs and can't see the DAGs of other teams.

Use separate Amazon MWAA environments when teams need any of the following:

- Different Apache Airflow versions.
- Different network configurations or VPCs.
- Separate compute, such as their own workers and DAG processors.
- Different plugins or Python dependencies.
- Different compliance or security requirements.
- Independent scaling.

Limitations

Multi-team environments have the following limitations:

- **Experimental feature** – Multi-team support is based on an Apache Airflow experimental feature that is in preview and still evolving. Some functionality is incomplete, and behavior can change without notice in future releases. Use multi-team environments with caution.
- **Shared compute** – All teams share the same DAG processor and workers. Teams are isolated only by DAG folder and access control. A team's workload can affect the performance of other teams.
- **Shared metadata database** – All teams share the same SQL database instance containing Airflow metadata. A team's workload disproportionally loading the database can affect the performance of other teams.
- **Shared plugins and Python dependencies** – All teams use the same plugins and the same `requirements.txt` file.
- **IAM Identity Center dependency** – Teams require IAM Identity Center integration. After you enable IAM Identity Center integration on an environment, you can't disable it, even if you remove all teams.
- **No renaming** – You can't rename a team. To change a team's name, remove the team and add a new one.
- **Version downgrade** – You can't downgrade an environment that has teams to an Apache Airflow version earlier than 3.3.1. To downgrade, first remove all teams.
- **Team quota** – An environment can have up to 20 teams. You can't increase this quota.

Next steps

To add, change, or remove teams, see Managing teams.

Managing teams

You can add, update, and remove teams by using the Amazon MWAA console or the AWS CLI. Before you start, review the prerequisites and the team name and DAG folder rules in Multi-team environments.

> **Important**
>
> Amazon MWAA multi-team support is based on the Apache Airflow multi-team feature. This Apache Airflow feature is experimental and in preview as of Apache Airflow version 3.3.1. The feature is not complete, and its behavior and APIs can change in breaking ways, or be removed, between releases without notice. We recommend that you use multi-team environments with caution and that you don't rely on them for production-critical workloads. For more information, see Experimental features on the Apache Airflow website.

Contents

- Adding teams (console)
  - Adding teams to a new environment
  - Adding teams to an existing environment
- Adding and updating teams (AWS CLI)
  - Adding teams to a new environment
  - Adding or updating teams in an existing environment
- Removing teams
  - Removing teams (console)
  - Removing teams (AWS CLI)
  - Returning to single-team mode

Adding teams (console)

You can add teams when you create an environment or when you edit an existing environment.

Adding teams to a new environment

To add teams when you create an environment

1. Open the Environments page on the Amazon MWAA console.
2. Choose **Create environment**.
3. On the **Specify details** page, configure the environment name and the DAG code settings. The DAG folder that you specify becomes the global DAG folder.
4. On the **Configure advanced settings** page, expand **Identity Center configuration**.
5. Turn on **Enable IAM Identity Center integration**, and then enter the **AWS access portal URL** and **Identity Store ID**.

   You can't disable IAM Identity Center integration after you enable it.
6. Expand **Teams**.
7. Choose **Add new team**.
8. For **Team name**, enter the display name of the team's IAM Identity Center group.
9. For **DAGs folder**, keep the default location or enter a different Amazon S3 URI. You can also choose **Browse S3**. The default location is a folder named after the team in the bucket that you chose on the **Specify details** page, such as `s3://amzn-s3-demo-bucket/data-engineering/`.
10. To add more teams, repeat the previous three steps.
11. Finish configuring the environment, and then choose **Create environment**.

Adding teams to an existing environment

To add teams to an existing environment

1. Open the Environments page on the Amazon MWAA console.
2. Choose the environment, and then choose **Edit**.
3. Choose **Next** to go to the **Configure advanced settings** page.
4. Under **Identity Center configuration**, make sure that **Enable IAM Identity Center integration** is turned on, and then expand **Teams**.

   If the environment doesn't use IAM Identity Center yet, enabling it is permanent. Before you enable it, see Alternative: enable IAM Identity Center on an existing environment.

5. Choose **Add new team**, and then enter the team name and DAG folder as described in the previous procedure.
6. Choose **Next** until you are on the **Review and save** page, review your changes, and then choose **Save**.

Adding and updating teams (AWS CLI)

Use the `--teams` parameter of the `create-environment` and `update-environment` commands. The parameter has a different format for each command:

- `create-environment` takes a map of teams. Each key is a team name, and each value specifies the team's Amazon S3 bucket ARN (`SourceBucketArn`) and DAG folder path (`DagS3Path`). Each team can use a different bucket, as long as the environment's execution role can read from it.
- `update-environment` takes an object with two optional fields. `AddOrUpdate` is a map of teams in the same format as `create-environment`. `Remove` is a list of team names to remove.

Adding teams to a new environment

Example Create an environment with two teams

```
aws mwaa create-environment \
  --name my-multi-team-environment \
  --airflow-version 3.3.1 \
  --source-bucket-arn arn:aws:s3:::amzn-s3-demo-bucket \
  --dag-s3-path dags \
  --execution-role-arn arn:aws:iam::123456789012:role/MyMWAAExecutionRole \
  --network-configuration SubnetIds=subnet-12345,subnet-67890,SecurityGroupIds=sg-12345
 \
  --identity-center-configuration StartUrl=https://my-org.awsapps.com/
start,IdentityStoreId=d-1234567890 \
  --teams '{
    "data-engineering": {
      "SourceBucketArn": "arn:aws:s3:::amzn-s3-demo-bucket",
      "DagS3Path": "teams/data-engineering/"
    },
    "analytics": {
      "SourceBucketArn": "arn:aws:s3:::amzn-s3-demo-bucket",
      "DagS3Path": "teams/analytics/"
    }
```

```
  }'
```

Adding or updating teams in an existing environment

Example Add a team to an existing environment

```
aws mwaa update-environment \
  --name my-multi-team-environment \
  --teams '{
    "AddOrUpdate": {
      "ml-ops": {
        "SourceBucketArn": "arn:aws:s3:::amzn-s3-demo-bucket",
        "DagS3Path": "teams/ml-ops/"
      }
    }
  }'
```

Removing teams

When you save the change, Amazon MWAA stops processing the team's DAGs. Existing DAG runs remain in the metadata database, but the team's DAGs, variables, and connections become inaccessible. You can't undo this action. Amazon MWAA doesn't delete the team's DAG files in Amazon S3.

You can't rename a team. To change a team's name, remove the team and add a new one. You can't remove a team and add a team with the same name in the same update.

Removing teams (console)

To remove a team

1. Open the Environments page on the Amazon MWAA console.
2. Choose the environment, and then choose **Edit**.
3. Choose **Next** to go to the **Configure advanced settings** page, and then expand **Teams**.
4. Choose the remove icon (✕) next to the team that you want to remove.
5. In the **Remove team?** dialog box, enter the team name to confirm, and then choose **Remove team**.
6. Choose **Next** until you are on the **Review and save** page, review your changes, and then choose **Save**.

Removing teams (AWS CLI)

Use the `Remove` field of the `--teams` parameter.

Example Remove two teams

```
aws mwaa update-environment \
  --name my-multi-team-environment \
  --teams '{
    "Remove": ["ml-ops", "analytics"]
  }'
```

Returning to single-team mode

When you remove all teams, the environment returns to single-team mode. Users still sign in through IAM Identity Center, and all users see the DAGs in the environment's DAG folder. IAM Identity Center integration stays enabled, because you can't disable it.

Changing the Apache Airflow version

Amazon MWAA supports minor and major version upgrades and downgrades.

A minor version upgrade means you can update your environment from version x**.4.**z to x**.5.**z or from x**.5.**z to x**.4.**z.

To perform a major version upgrade, for example from version **2.**y.z to **3.**y.z, you must first perform a minor upgrade to the latest Airflow `2.11.x` version followed by upgrade to `3.x`. Alternatively, to perform a major upgrade from Airflow `2.10.3` and below, you need to create a new environment and migrate your resources. You can also downgrade from Airflow version **3.**y.z to Airflow version `2.11.x`. For more information about upgrading to a new major version of Apache Airflow, see Migrating to a new Amazon MWAA environment.

During the upgrade or downgrade process, Amazon MWAA captures a snapshot of your environment metadata, upgrades or downgrades the workers, schedulers, the web server to the new Apache Airflow version, and finally restores the metadata database using the snapshot.

Before you upgrade or downgrade, make sure that your DAGs and other workflow resources are compatible with the new Apache Airflow version you're upgrading to. If you use

`requirements.txt` to manage dependencies, you must also ensure that the dependencies you specify in your requirements are compatible with the new version.

Topics

- Upgrade or downgrade your workflow resources
- Specify the new version

Upgrade or downgrade your workflow resources

Whenever you're changing Apache Airflow versions, ensure that you reference the correct `--constraint` URL in your `requirements.txt`.

> **Warning**
>
> Specifying requirements that are incompatible with your target Apache Airflow version during an upgrade or downgrade might result in a lengthy rollback process to the previous version of Apache Airflow with the previous requirements version.

Migrate your workflow resources

1. Create a fork of the aws-mwaa-docker-images repository, and clone a copy of the Amazon MWAA local runner.
2. Checkout to the branch of the aws-mwaa-docker-images repository that matches the version you are upgrading or downgrading to.
3. To update your `requirements.txt`, follow the best practices we recommend in Managing Python dependencies, in the *Amazon MWAA User Guide*.
4. (Optional) To speed up the upgrade or downgrade process, clean up the environment's metadata database. Environments with a large amount of metadata can take significantly longer to upgrade.
5. After you have successfully tested your workflow resources, copy your DAGs, `requirements.txt`, and plugins to your environment's Amazon S3 bucket.

You are now ready to edit the environment, specify a new Apache Airflow version, and start the update procedure.

Specify the new version

After you have completed updating your workflow resources to ensure compatibility with the new Apache Airflow version, do the following to edit environment details and specify the version of Apache Airflow that you want to upgrade to.

> **Note**
>
> When you perform an upgrade or downgrade, all tasks currently running on the environment are terminated during the procedure. The update procedure can take up to two hours, during which time your environment will be unavailable.

Specify a new version using the console

1. Open the Environments page on the Amazon MWAA console.
2. From the **Environments** list, choose the environment that you want to upgrade or downgrade.
3. On the environment page, choose **Edit** to edit the environment.
4. In the **Environment details** section, for **Airflow version**, choose the Apache Airflow version number that you want to upgrade or downgrade the environment to from the dropdown list.
5. Choose **Next** until you are on the **Review and save** page.
6. On the **Review and save** page, review your changes, then choose **Save**.

When you apply changes, your environment begins the upgrade or downgrade procedure. During this period, the status of your environment indicates what actions Amazon MWAA is taking, and whether the procedure is successful.

In a successful upgrade or downgrade scenario, the status will be `UPDATING`, then `CREATING_SNAPSHOT` as Amazon MWAA captures a backup of your metadata. Finally, the status will return first to `UPDATING`, then to `AVAILABLE` when the procedure is done.

If the environment fails to upgrade or downgrade, your environment status will be `ROLLING_BACK`. If the rollback is successful, the status will first present `UPDATE_FAILED`, indicating that the update failed but the environment is available. If the rollback fails, the status will be `UNAVAILABLE`, indicating that you cannot access the environment.

Using a startup script with Amazon MWAA

A startup script is a shell (`.sh`) script you host in your environment's Amazon S3 bucket similar to your DAGs, requirements, and plugins. Amazon MWAA runs this script during startup on every individual Apache Airflow component (worker, scheduler, and web server) before installing requirements and initializing the Apache Airflow process. Use a startup script to do the following:

- **Install runtimes** – Install Linux runtimes required by your workflows and connections.
- **Configure environment variables** – Set environment variables for each Apache Airflow component. Overwrite common variables such as `PATH`, `PYTHONPATH`, and `LD_LIBRARY_PATH`.
- **Manage keys and tokens** – Pass access tokens for custom repositories to `requirements.txt` and configure security keys.

The following topics describe how to configure a startup script to install Linux runtimes, set environment variables, and troubleshoot related issues using CloudWatch Logs.

Topics

- Configure a startup script
- Install Linux runtimes using a startup script
- Set environment variables using a startup script

Configure a startup script

To use a startup script with your existing Amazon MWAA environment, upload a `.sh` file to your environment's Amazon S3 bucket. Then, to associate the script with the environment, specify the following in your environment details:

- **The Amazon S3 URL path to the script** – The relative path to the script hosted in your bucket, for example, `s3://mwaa-environment/startup.sh`
- **The Amazon S3 version ID of the script** – The version of the startup shell script in your Amazon S3 bucket. You must specify the version ID that Amazon S3 assigns to the file every time you update the script. Version IDs are Unicode, UTF-8 encoded, URL-ready, opaque strings that are no more than 1,024 bytes long, for example, `3sL4kqtJlcpXroDTDmJ+rmSpXd3dIbrHY+MTRCxf3vjVBH40Nr8X8gdRQBpUMLUo`.

To complete the steps in this section, use the following sample script. The script outputs the value assigned to `MWAA_AIRFLOW_COMPONENT`. This environment variable identifies each Apache Airflow component that the script runs on.

Copy the code and save it locally as `startup.sh`.

```
#!/bin/sh

echo "Printing Apache Airflow component"
echo $MWAA_AIRFLOW_COMPONENT
```

Next, upload the script to your Amazon S3 bucket.

AWS Management Console

To upload a shell script (console)

1. Sign in to the AWS Management Console and open the Amazon S3 console at https://console.aws.amazon.com/s3/.
2. From the **Buckets** list, choose the name of the bucket associated with your environment.
3. On the **Objects** tab, choose **Upload**.
4. On the **Upload** page, drag and drop the shell script you created.
5. Choose **Upload**.

The script is listed in the list of **Objects**. Amazon S3 creates a new version ID for the file. If you update the script and upload it again using the same file name, a new version ID is assigned to the file.

AWS CLI

To create and upload a shell script (CLI)

1. Open a new command prompt, and run the Amazon S3 `ls` command to list and identify the bucket associated with your environment.

   ```
   aws s3 ls
   ```

2. Navigate to the folder where you saved the shell script. Use `cp` in a new prompt window to upload the script to your bucket. Replace *amzn-s3-demo-bucket* with your information.

```
aws s3 cp startup.sh s3://amzn-s3-demo-bucket/startup.sh
```

If successful, Amazon S3 outputs the URL path to the object:

```
upload: ./startup.sh to s3://amzn-s3-demo-bucket/startup.sh
```

3. Use the following command to retrieve the latest version ID for the script.

```
aws s3api list-object-versions --bucket amzn-s3-demo-bucket --prefix startup --
query 'Versions[?IsLatest].[VersionId]' --output text
```

```
BbdVMmBRjtestta1EsVnbybZp1Wqh1J4
```

You specify this version ID when you associate the script with an environment.

Now, associate the script with your environment.

AWS Management Console

To associate the script with an environment (console)

1. Open the Environments page on the Amazon MWAA console.
2. Select the row for the environment you want to update, then choose **Edit**.
3. On the **Specify details** page, for **Startup script file - *optional***, enter the Amazon S3 URL for the script, for example: `s3://amzn-s3-demo-bucket/startup-sh..`
4. Choose the latest version from the drop down list, or **Browse S3** to find the script.
5. Choose **Next**, then proceed to the **Review and save** page.
6. Review changes, then choose **Save**.

Environment updates can take between 10 to 30 minutes. Amazon MWAA runs the startup script as each component in your environment restarts.

AWS CLI

To associate the script with an environment (CLI)

- Open a command prompt and use `update-environment` to specify the Amazon S3 URL and version ID for the script.

```
aws mwaa update-environment \
--name your-mwaa-environment \
--startup-script-s3-path startup.sh \
--startup-script-s3-object-version BbdVMmBRjtestta1EsVnbybZp1Wqh1J4
```

If successful, Amazon MWAA returns the Amazon Resource Name (ARN) for the environment:

```
arn:aws:airflow:us-west-2:123456789012:environment/your-mwaa-environment
```

Environment update can take between 10 to 30 minutes. Amazon MWAA runs the startup script as each component in your environment restarts.

Finally, retrieve log events to verify that the script is working as expected. When you activate logging for an each Apache Airflow component, Amazon MWAA creates a new log group and log stream. For more information, refer to Apache Airflow log types.

AWS Management Console

To get the Apache Airflow log stream (console)

1. Open the Environments page on the Amazon MWAA console.
2. Choose your environment.
3. In the **Monitoring** pane, choose the log group for which you want to access logs, for example, **Airflow scheduler log group** .
4. In the CloudWatch console, from the **Log streams** list, choose a stream with the following prefix: `startup_script_execution_ip`.
5. On the **Log events** pane, you will refer to the output of the command printing the value for `MWAA_AIRFLOW_COMPONENT`. For example, for scheduler logs, you will the following:

```
Printing Apache Airflow component
        scheduler
        Finished running startup script. Execution time: 0.004s.
        Running verification
        Verification completed
```

You can repeat the previous steps to access worker and webserver logs.

Install Linux runtimes using a startup script

Use a startup script to update the operating system of an Apache Airflow component, and install additional runtime libraries to use with your workflows. For example, the following script runs `yum update` to update the operating system.

When running `yum update` in a startup script, you must exclude Python using `--exclude=python*` as listed in the example. For your environment to run, Amazon MWAA installs a specific version of Python compatible with your environment. Therefore, you can't update the environment's Python version using a startup script.

```
#!/bin/sh

echo "Updating operating system"
sudo yum update -y --exclude=python*
```

To install runtimes on specific Apache Airflow component, use `MWAA_AIRFLOW_COMPONENT` and `if` and `fi` conditional statements. This example runs a single command to install the `libaio` library on the scheduler and worker, but not on the webserver.

Important

- If you have configured a private webserver, you must either use the following condition or provide all installation files locally to avoid installation timeouts.
- Use `sudo` to run operations that require administrative privileges.

```
#!/bin/sh
```

```
if [[ "${MWAA_AIRFLOW_COMPONENT}" != "webserver" ]]
  then
    sudo yum -y install libaio

fi
```

You can use a startup script to get the Python version.

```
#!/bin/sh

export PYTHON_VERSION_CHECK=`python -c 'import sys; version=sys.version_info[:3];
 print("{0}.{1}.{2}".format(*version))'`
echo "Python version is $PYTHON_VERSION_CHECK"
```

Amazon MWAA does not support overriding the default Python version, as this might lead to incompatibilities with the installed Apache Airflow libraries.

Set environment variables using a startup script

Use startup scripts to set environment variables and modify Apache Airflow configurations. The following defines a new variable, `ENVIRONMENT_STAGE`. You can reference this variable in a DAG or in your custom modules.

```
#!/bin/sh

export ENVIRONMENT_STAGE="development"
echo "$ENVIRONMENT_STAGE"
```

Use startup scripts to overwrite common Apache Airflow or system variables. For example, you set `LD_LIBRARY_PATH` to instruct Python to search for binaries in the path you specify. This lets you provide custom binaries for your workflows using plugins:

```
#!/bin/sh

export LD_LIBRARY_PATH=/usr/local/airflow/plugins/your-custom-binary
```

Reserved environment variables

Amazon MWAA reserves a set of critical environment variables. If you overwrite a *reserved* variable, Amazon MWAA restores it to its default. The following lists the reserved variables:

- `MWAA__AIRFLOW__COMPONENT` – Used to identify the Apache Airflow component with one of the following values: `scheduler`, `worker`, or `webserver`.
- `AIRFLOW__WEBSERVER__SECRET_KEY` – The secret key used for securely signing session cookies in the Apache Airflow web server.
- `AIRFLOW__CORE__FERNET_KEY` – The key used for encryption and decryption of sensitive data stored in the metadata database, for example, connection passwords.
- `AIRFLOW_HOME` – The path to the Apache Airflow home directory where configuration files and DAG files are stored locally.
- `AIRFLOW__CELERY__BROKER_URL` – The URL of the message broker used for communication between the Apache Airflow scheduler and the Celery worker nodes.
- `AIRFLOW__CELERY__RESULT_BACKEND` – The URL of the database used to store the results of Celery tasks.
- `AIRFLOW__CORE__EXECUTOR` – The executor class that Apache Airflow must use. In Amazon MWAA this is a `CeleryExecutor`
- `AIRFLOW__CORE__LOAD_EXAMPLES` – Used to activate, or deactivate, the loading of example DAGs.
- `AIRFLOW__METRICS__METRICS_BLOCK_LIST` – Used to manage which Apache Airflow metrics are emitted and captured by Amazon MWAA in CloudWatch.
- `SQL_ALCHEMY_CONN` – The connection string for the RDS for PostgreSQL database used to store Apache Airflow metadata in Amazon MWAA.
- `AIRFLOW__CORE__SQL_ALCHEMY_CONN` – Used for the same purpose as `SQL_ALCHEMY_CONN`, but following the new Apache Airflow naming convention.
- `AIRFLOW__CELERY__DEFAULT_QUEUE` – The default queue for Celery tasks in Apache Airflow.
- `AIRFLOW__OPERATORS__DEFAULT_QUEUE` – The default queue for tasks using specific Apache Airflow operators.
- `AIRFLOW_VERSION` – The Apache Airflow version installed in the Amazon MWAA environment.
- `AIRFLOW_CONN_AWS_DEFAULT` – The default AWS credentials used to integrate with other AWS services in.
- `AWS_DEFAULT_REGION` – Sets the default AWS Region used with default credentials to integrate with other AWS services.
- `AWS_REGION` – If defined, this environment variable overrides the values in the environment variable `AWS_DEFAULT_REGION` and the profile setting region.
- `PYTHONUNBUFFERED` – Used to send `stdout` and `stderr` streams to container logs.

- `AIRFLOW__METRICS__STATSD_ALLOW_LIST` – Used to configure an allow list of comma-separated prefixes to send the metrics that start with the elements of the list.
- `AIRFLOW__METRICS__STATSD_ON` – Activates sending metrics to `StatsD`.
- `AIRFLOW__METRICS__STATSD_HOST` – Used to connect to the `StatSD` daemon.
- `AIRFLOW__METRICS__STATSD_PORT` – Used to connect to the `StatSD` daemon.
- `AIRFLOW__METRICS__STATSD_PREFIX` – Used to connect to the `StatSD` daemon.
- `AIRFLOW__CELERY__WORKER_AUTOSCALE` – Sets the maximum and minimum concurrency.
- `AIRFLOW__CORE__DAG_CONCURRENCY` – Sets the number of task instances that can run concurrently by the scheduler in one DAG.
- `AIRFLOW__CORE__MAX_ACTIVE_TASKS_PER_DAG` – Sets the maximum number of active tasks per DAG.
- `AIRFLOW__CORE__PARALLELISM` – Defines the maximum number of task instances that can simultaneously.
- `AIRFLOW__SCHEDULER__PARSING_PROCESSES` – Sets the maximum number of processes parsed by the scheduler to schedule DAGs.
- `AIRFLOW__CELERY_BROKER_TRANSPORT_OPTIONS__VISIBILITY_TIMEOUT` – Defines the number of seconds a worker waits to acknowledge the task before the message is redelivered to another worker.
- `AIRFLOW__CELERY_BROKER_TRANSPORT_OPTIONS__REGION` – Sets the AWS Region for the underlying Celery transport.
- `AIRFLOW__CELERY_BROKER_TRANSPORT_OPTIONS__PREDEFINED_QUEUES` – Sets the queue for the underlying Celery transport.
- `AIRFLOW_SCHEDULER_ALLOWED_RUN_ID_PATTERN` – Used to verify the validity of your input for the `run_id` parameter when triggering a DAG.
- `AIRFLOW__WEBSERVER__BASE_URL` – The URL of the web server used to host the Apache Airflow UI.
- `PYTHONPATH` (only for Apache Airflow v2.9 and later) – Reserved by Amazon MWAA to ensure that all basic environment functionalities operate correctly.

> **Note**
>
> For Apache Airflow versions earlier than v2.9, `PYTHONPATH` is an *unreserved* environment variable.

Unreserved environment variables

You can use a startup script to overwrite *unreserved* environment variables. The following lists some of these common variables:

- `PATH` – Specifies a list of directories where the operating system searches for executable files and scripts. When a command runs in the command line, the system checks the directories in `PATH` to find and execute the command. When you create custom operators or tasks in Apache Airflow, you might need to rely on external scripts or executables. If the directories containing these files are not in the specified in the `PATH` variable, the tasks fail to run when the system is unable to locate them. By adding the appropriate directories to `PATH`, Apache Airflow tasks can find and run the required executables.
- `PYTHONPATH` – Used by the Python interpreter to determine which directories to search for imported modules and packages. It is a list of directories that you can add to the default search path. This lets the interpreter find and load Python libraries not included in the standard library, or installed in system directories. Use this variable to add your modules and custom Python packages and use them with your DAGs.

  > **Note**
  >
  > For Apache Airflow v2.9 and later, `PYTHONPATH` is a *reserved* environment variable.

- `LD_LIBRARY_PATH` – An environment variable used by the dynamic linker and loader in Linux to find and load shared libraries. It specifies a list of directories containing shared libraries, which are searched before the default system library directories. Use this variable to specify your custom binaries.
- `CLASSPATH` – Used by the Java Runtime Environment (JRE) and Java Development Kit (JDK) to locate and load Java classes, libraries, and resources at runtime. It is a list of directories, JAR files, and ZIP archives that contain compiled Java code.

Working with DAGs on Amazon MWAA

To run Directed Acyclic Graphs (DAGs) on an Amazon Managed Workflows for Apache Airflow environment, you copy your files to the Amazon S3 storage bucket attached to your environment, then let Amazon MWAA know where your DAGs and supporting files are located on the Amazon MWAA console. Amazon MWAA takes care of synchronizing the DAGs among workers, schedulers, and the webserver. This guide describes how to add or update your DAGs and install custom plugins and Python dependencies on an Amazon MWAA environment.

Topics

- Amazon S3 bucket overview
- Adding or updating DAGs
- Installing custom plugins
- Installing Python dependencies
- Deleting files on Amazon S3

Amazon S3 bucket overview

An Amazon S3 bucket for an Amazon MWAA environment must have *Public Access Blocked*. By default, all Amazon S3 resources—buckets, objects, and related sub-resources (for example, lifecycle configuration)—are private.

- Only the resource owner, the AWS account that created the bucket, can access the resource. The resource owner (for example, your administrator) can grant access permissions to others by writing an access control policy.
- The access policy you set up must have permission to add DAGs, custom plugins in `plugins.zip`, and Python dependencies in `requirements.txt` to your Amazon S3 bucket. For an example policy that contains the required permissions, refer to AmazonMWAAFullConsoleAccess.

An Amazon S3 bucket for an Amazon MWAA environment must have *Versioning Enabled*. When Amazon S3 bucket versioning is enabled, anytime a new version is created, a new copy is created.

- Versioning is enabled for the custom plugins in a `plugins.zip`, and Python dependencies in a `requirements.txt` on your Amazon S3 bucket.

- You must specify the version of a `plugins.zip`, and `requirements.txt` on the Amazon MWAA console each time these files are updated on your Amazon S3 bucket.

Adding or updating DAGs

Directed Acyclic Graphs (DAGs) are defined within a Python file that defines the DAG's structure as code. You can use the AWS CLI or the Amazon S3 console to upload DAGs to your environment. This topic describes the steps to add or update Apache Airflow DAGs on your Amazon Managed Workflows for Apache Airflow environment using the `dags` folder in your Amazon S3 bucket.

Sections

- Prerequisites
- How it works
- What's changed?
- Testing DAGs using the Amazon MWAA CLI utility
- Uploading DAG code to Amazon S3
- Specifying the path to your DAGs folder on the Amazon MWAA console (the first time)
- Accessing changes on your Apache Airflow UI
- DAG folders in multi-team environments
- What's next?

Prerequisites

You'll need the following before you can complete the steps on this page.

- **Permissions** — Your AWS account must have been granted access by your administrator to the AmazonMWAAFullConsoleAccess access control policy for your environment. In addition, your Amazon MWAA environment must be permitted by your execution role to access the AWS resources used by your environment.
- **Access** — If you require access to public repositories to install dependencies directly on the webserver, your environment must be configured with **public network** webserver access. For more information, refer to the section called “Apache Airflow access modes”.

- **Amazon S3 configuration** — The Amazon S3 bucket used to store your DAGs, custom plugins in `plugins.zip`, and Python dependencies in `requirements.txt` must be configured with *Public Access Blocked* and *Versioning Enabled*.

How it works

A Directed Acyclic Graph (DAG) is defined within a single Python file that defines the DAG's structure as code. It consists of the following:

- A DAG definition.
- Operators that describe how to run the DAG and the tasks to run.
- Operator relationships that describe the order in which to run the tasks.

To run an Apache Airflow platform on an Amazon MWAA environment, you need to copy your DAG definition to the `dags` folder in your storage bucket. For example, the DAG folder in your storage bucket should resemble:

Example DAG folder

```
dags/
# dag_def.py
```

Amazon MWAA automatically syncs new and changed objects from your Amazon S3 bucket to Amazon MWAA scheduler and worker containers' `/usr/local/airflow/dags` folder every 30 seconds, preserving the Amazon S3 source's file hierarchy, regardless of file type. The time that new DAGs take to be listed in your Apache Airflow UI is controlled by `scheduler.dag_dir_list_interval`. Changes to existing DAGs will be picked up on the next DAG processing loop.

> **Note**
>
> You do not need to include the `airflow.cfg` configuration file in your DAG folder. You can override the default Apache Airflow configurations from the Amazon MWAA console. For more information, refer to Using Apache Airflow configuration options on Amazon MWAA.

What's changed?

To review changes for a specific Apache Airflow release, refer to the Release Notes page.

- Apache Airflow v3 configurations: Configuration Reference
- Apache Airflow v2 public interface information: Public Interface of Airflow

Testing DAGs using the Amazon MWAA CLI utility

- The command line interface (CLI) utility replicates an Amazon Managed Workflows for Apache Airflow environment locally.
- The CLI builds a Docker container image locally that's similar to an Amazon MWAA production image. You can use this to run a local Apache Airflow environment to develop and test DAGs, custom plugins, and dependencies before deploying to Amazon MWAA.
- To run the CLI, refer to aws-mwaa-docker-images on GitHub.

Uploading DAG code to Amazon S3

You can use the Amazon S3 console or the AWS Command Line Interface (AWS CLI) to upload DAG code to your Amazon S3 bucket. The following steps assume you are uploading code (`.py`) to a folder named `dags` in your Amazon S3 bucket.

Using the AWS CLI

The AWS Command Line Interface (AWS CLI) is an open source tool that you can use to interact with AWS services using commands in your command-line shell. To complete the steps on this page, you need the following:

- AWS CLI – Install version 2.
- AWS CLI – Quick configuration with `aws configure`.

To upload using the AWS CLI

1. Use the following command to list all of your Amazon S3 buckets.

```
aws s3 ls
```

2. Use the following command to list the files and folders in the Amazon S3 bucket for your environment.

```
aws s3 ls s3://YOUR_S3_BUCKET_NAME
```

3. The following command uploads a `dag_def.py` file to a `dags` folder.

```
aws s3 cp dag_def.py s3://amzn-s3-demo-bucket/dags/
```

   If a folder named `dags` does not already exist on your Amazon S3 bucket, this command creates the `dags` folder and uploads the file named `dag_def.py` to the new folder.

Using the Amazon S3 console

The Amazon S3 console is a web-based user interface that you can use to create and manage the resources in your Amazon S3 bucket. The following steps assume you have a DAGs folder named `dags`.

To upload using the Amazon S3 console

1. Open the Environments page on the Amazon MWAA console.
2. Choose an environment.
3. Select the **S3 bucket** link in the **DAG code in S3** pane to open your storage bucket in the console.
4. Choose the `dags` folder.
5. Choose **Upload**.
6. Choose **Add file**.
7. Select the local copy of your `dag_def.py`, choose **Upload**.

Specifying the path to your DAGs folder on the Amazon MWAA console (the first time)

The following steps assume you are specifying the path to a folder on your Amazon S3 bucket named `dags`.

1. Open the Environments page on the Amazon MWAA console.

2. Choose the environment where you want to run DAGs.
3. Choose **Edit**.
4. On the **DAG code in Amazon S3** pane, choose **Browse S3** adjacent to the **DAG folder** field.
5. Select your `dags` folder.
6. Choose **Choose**.
7. Choose **Next**, **Update environment**.

Accessing changes on your Apache Airflow UI

You need Apache Airflow UI access policy: AmazonMWAAWebServerAccess permissions for your AWS account in AWS Identity and Access Management (IAM) to access your Apache Airflow UI.

To access your Apache Airflow UI

1. Open the Environments page on the Amazon MWAA console.
2. Choose an environment.
3. Choose **Open Airflow UI**.

DAG folders in multi-team environments

In a multi-team environment, the DAG folder that you configure for the environment is the global DAG folder, and all users can see its DAGs. Each team also has its own DAG folder, and only members of that team can see its DAGs. DAG folders can't overlap. For the folder rules and examples, see Multi-team environments.

What's next?

Test your DAGs, custom plugins, and Python dependencies locally using aws-mwaa-docker-images on GitHub.

Installing custom plugins

Amazon Managed Workflows for Apache Airflow supports Apache Airflow's built-in plugin manager, allowing you to use custom Apache Airflow operators, hooks, sensors, or interfaces. This page describes the steps to install Apache Airflow custom plugins on your Amazon MWAA environment using a `plugins.zip` file.

Contents

- Prerequisites
- How it works
- When to use the plugins
- Custom plugins overview
  - Custom plugins directory and size limits
- Examples of custom plugins
  - Example using a flat directory structure in plugins.zip
  - Example using a nested directory structure in plugins.zip
- Creating a plugins.zip file
  - Step one: Test custom plugins using the Amazon MWAA CLI utility
  - Step two: Create the plugins.zip file
- Uploading plugins.zip to Amazon S3
  - Using the AWS CLI
  - Using the Amazon S3 console
- Installing custom plugins on your environment
  - Specifying the path to plugins.zip on the Amazon MWAA console (the first time)
  - Specifying the plugins.zip version on the Amazon MWAA console
- Example use cases for plugins.zip
- What's next?

Prerequisites

You'll need the following before you can complete the steps on this page.

- **Permissions** — Your AWS account must have been granted access by your administrator to the AmazonMWAAFullConsoleAccess access control policy for your environment. In addition, your Amazon MWAA environment must be permitted by your execution role to access the AWS resources used by your environment.
- **Access** — If you require access to public repositories to install dependencies directly on the webserver, your environment must be configured with **public network** webserver access. For more information, refer to the section called “Apache Airflow access modes”.

- **Amazon S3 configuration** — The Amazon S3 bucket used to store your DAGs, custom plugins in `plugins.zip`, and Python dependencies in `requirements.txt` must be configured with *Public Access Blocked* and *Versioning Enabled*.

How it works

To run custom plugins on your environment, you must do three things:

1. Create a `plugins.zip` file locally.
2. Upload the local `plugins.zip` file to your Amazon S3 bucket.
3. Specify the version of this file in the **Plugins file** field on the Amazon MWAA console.

> **Note**
>
> If this is the first time you're uploading a `plugins.zip` to your Amazon S3 bucket, you also need to specify the path to the file on the Amazon MWAA console. You only need to complete this step once.

When to use the plugins

Plugins are required only for extending the Apache Airflow user interface, as outlined in the Apache Airflow documentation. Custom operators can be placed directly in the `/dags` folder alongside your DAG code.

If you need to create your own integrations with external systems, place them in the `/dags` folder or a subfolder within it, but not in the `plugins.zip` folder. In Apache Airflow 2.x, plugins are primarily used for extending the UI.

Similarly, other dependencies can not be placed in `plugins.zip`. Instead, they can be stored in a location in the Amazon S3 `/dags` folder, where they will be synchronized to each Amazon MWAA container before Apache Airflow starts.

> **Note**
>
> Any file in the `/dags` folder or in `plugins.zip` that does not explicitly define an Apache Airflow DAG object must be listed in an `.airflowignore` file.

Custom plugins overview

Apache Airflow's built-in plugin manager can integrate external features to its core by simply dropping files in an `$AIRFLOW_HOME/plugins` folder. It you can use to use custom Apache Airflow operators, hooks, sensors, or interfaces. The following section provides an example of flat and nested directory structures in a local development environment and the resulting import statements, which determines the directory structure within a plugins.zip.

Custom plugins directory and size limits

The Apache Airflow scheduler and the workers search for custom plugins during startup on the AWS-managed Fargate container for your environment at `/usr/local/airflow/plugins/*`.

- **Directory structure**. The directory structure (at `/*`) is based on the contents of your `plugins.zip` file. For example, if your `plugins.zip` contains the `operators` directory as a main-level directory, then the directory will be extracted to `/usr/local/airflow/plugins/operators` on your environment.
- **Size limit**. We recommend a `plugins.zip` file less than than 1 GB. The larger the size of a `plugins.zip` file, the longer the startup time on an environment. Although Amazon MWAA doesn't limit the size of a `plugins.zip` file explicitly, if dependencies can't be installed within ten minutes, the Fargate service will time-out and attempt to rollback the environment to a stable state.

> **Note**
>
> For environments using Apache Airflow v2.0.2, Amazon MWAA limits outbound traffic on the Apache Airflow webserver, and does not permit installing plugins nor Python dependencies directly on the webserver. Starting with Apache Airflow v2.2.2, Amazon MWAA can install plugins and dependencies directly on the webserver.

Examples of custom plugins

The following section uses sample code in the *Apache Airflow reference guide* to explain how to structure your local development environment.

Example using a flat directory structure in plugins.zip

Apache Airflow v3

The following example presents a `plugins.zip` file with a flat directory structure for Apache Airflow v3.

Example flat directory with `PythonVirtualenvOperator` plugins.zip

The following example displays the main-level tree of a plugins.zip file for the `PythonVirtualenvOperator` custom plugin.

```
### virtual_python_plugin.py
```

Example plugins/virtual_python_plugin.py

The following example displays the `PythonVirtualenvOperator` custom plugin.

```
"""
Copyright Amazon.com, Inc. or its affiliates. All Rights Reserved.

Permission is hereby granted, free of charge, to any person obtaining a copy of
this software and associated documentation files (the "Software"), to deal in
the Software without restriction, including without limitation the rights to
use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of
the Software, and to permit persons to whom the Software is furnished to do so.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR
IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS
FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR
COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER
IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN
CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.
"""
from airflow.plugins_manager import AirflowPlugin
import airflow.utils.python_virtualenv
from typing import List

def _generate_virtualenv_cmd(tmp_dir: str, python_bin: str, system_site_packages:
 bool) -> List[str]:
    cmd = ['python3','/usr/local/airflow/.local/lib/python3.7/site-packages/
virtualenv', tmp_dir]
    if system_site_packages:
```

```
        cmd.append('--system-site-packages')
    if python_bin is not None:
        cmd.append(f'--python={python_bin}')
    return cmd

airflow.utils.python_virtualenv._generate_virtualenv_cmd=_generate_virtualenv_cmd

class VirtualPythonPlugin(AirflowPlugin):
    name = 'virtual_python_plugin'
```

Apache Airflow v2

The following example presents a `plugins.zip` file with a flat directory structure for Apache Airflow v2.

Example flat directory with `PythonVirtualenvOperator` plugins.zip

The following example displays the main-level tree of a plugins.zip file for the `PythonVirtualenvOperator` custom plugin.

```
### virtual_python_plugin.py
```

Example plugins/virtual_python_plugin.py

The following example displays the `PythonVirtualenvOperator` custom plugin.

```
"""
Copyright Amazon.com, Inc. or its affiliates. All Rights Reserved.

Permission is hereby granted, free of charge, to any person obtaining a copy of
this software and associated documentation files (the "Software"), to deal in
the Software without restriction, including without limitation the rights to
use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of
the Software, and to permit persons to whom the Software is furnished to do so.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR
IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS
FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR
COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER
IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN
CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.
"""
from airflow.plugins_manager import AirflowPlugin
```

```
import airflow.utils.python_virtualenv
from typing import List

def _generate_virtualenv_cmd(tmp_dir: str, python_bin: str, system_site_packages:
 bool) -> List[str]:
    cmd = ['python3','/usr/local/airflow/.local/lib/python3.7/site-packages/
virtualenv', tmp_dir]
    if system_site_packages:
        cmd.append('--system-site-packages')
    if python_bin is not None:
        cmd.append(f'--python={python_bin}')
    return cmd

airflow.utils.python_virtualenv._generate_virtualenv_cmd=_generate_virtualenv_cmd

class VirtualPythonPlugin(AirflowPlugin):
    name = 'virtual_python_plugin'
```

Example using a nested directory structure in plugins.zip

Apache Airflow v3

The following example presents a `plugins.zip` file with separate directories for hooks, `operators`, and a `sensors` directory.

Example plugins.zip

```
__init__.py
my_airflow_plugin.py
 hooks/
  |-- __init__.py
  |-- my_airflow_hook.py
 operators/
  |-- __init__.py
  |-- my_airflow_operator.py
  |-- hello_operator.py
 sensors/
  |-- __init__.py
  |-- my_airflow_sensor.py
```

The following example displays the import statements in the DAG (DAGs folder) that uses the custom plugins.

Example dags/your_dag.py

```
from airflow import DAG
from datetime import datetime, timedelta
from operators.my_airflow_operator import MyOperator
from sensors.my_airflow_sensor import MySensor
from operators.hello_operator import HelloOperator

default_args = {
 'owner': 'airflow',
 'depends_on_past': False,
 'start_date': datetime(2018, 1, 1),
 'email_on_failure': False,
 'email_on_retry': False,
 'retries': 1,
 'retry_delay': timedelta(minutes=5),
}


with DAG('customdag',
   max_active_runs=3,
   schedule_interval='@once',
   default_args=default_args) as dag:

 sens = MySensor(
  task_id='taskA'
 )

 op = MyOperator(
  task_id='taskB',
  my_field='some text'
 )

 hello_task = HelloOperator(task_id='sample-task', name='foo_bar')



 sens >> op >> hello_task
```

Example plugins/my_airflow_plugin.py

```
from airflow.plugins_manager import AirflowPlugin
from hooks.my_airflow_hook import *
```

```
from operators.my_airflow_operator import *

class PluginName(AirflowPlugin):

    name = 'my_airflow_plugin'

    hooks = [MyHook]
    operators = [MyOperator]
    sensors = [MySensor]
```

The following examples present each of the import statements needed in the custom plugin files.

Example hooks/my_airflow_hook.py

```
from airflow.hooks.base import BaseHook


class MyHook(BaseHook):

    def my_method(self):
        print("Hello World")
```

Example sensors/my_airflow_sensor.py

```
from airflow.sensors.base import BaseSensorOperator
from airflow.utils.decorators import apply_defaults


class MySensor(BaseSensorOperator):

    @apply_defaults
    def __init__(self,
                 *args,
                 **kwargs):
        super(MySensor, self).__init__(*args, **kwargs)

    def poke(self, context):
        return True
```

Example operators/my_airflow_operator.py

```
from airflow.operators.bash import BaseOperator
from airflow.utils.decorators import apply_defaults
from hooks.my_airflow_hook import MyHook


class MyOperator(BaseOperator):

    @apply_defaults
    def __init__(self,
                 my_field,
                 *args,
                 **kwargs):
        super(MyOperator, self).__init__(*args, **kwargs)
        self.my_field = my_field

    def execute(self, context):
        hook = MyHook('my_conn')
        hook.my_method()
```

Example operators/hello_operator.py

```
from airflow.models.baseoperator import BaseOperator
from airflow.utils.decorators import apply_defaults

class HelloOperator(BaseOperator):

    @apply_defaults
    def __init__(
            self,
            name: str,
            **kwargs) -> None:
        super().__init__(**kwargs)
        self.name = name

    def execute(self, context):
        message = "Hello {}".format(self.name)
        print(message)
        return message
```

Follow the steps in Testing custom plugins using the Amazon MWAA CLI utility, and then Creating a plugins.zip file to zip the contents **within** your `plugins` directory. For example, `cd plugins`.

Apache Airflow v2

The following example presents a `plugins.zip` file with separate directories for `hooks`, `operators`, and a `sensors` directory.

Example plugins.zip

```
__init__.py
 my_airflow_plugin.py
 hooks/
  |-- __init__.py
  |-- my_airflow_hook.py
 operators/
  |-- __init__.py
  |-- my_airflow_operator.py
  |-- hello_operator.py
 sensors/
  |-- __init__.py
  |-- my_airflow_sensor.py
```

The following example displays the import statements in the DAG (DAGs folder) that uses the custom plugins.

Example dags/your_dag.py

```
from airflow import DAG
from datetime import datetime, timedelta
from operators.my_airflow_operator import MyOperator
from sensors.my_airflow_sensor import MySensor
from operators.hello_operator import HelloOperator

default_args = {
 'owner': 'airflow',
 'depends_on_past': False,
 'start_date': datetime(2018, 1, 1),
 'email_on_failure': False,
 'email_on_retry': False,
 'retries': 1,
 'retry_delay': timedelta(minutes=5),
```

```
}


with DAG('customdag',
   max_active_runs=3,
   schedule_interval='@once',
   default_args=default_args) as dag:

 sens = MySensor(
  task_id='taskA'
 )

 op = MyOperator(
  task_id='taskB',
  my_field='some text'
 )

 hello_task = HelloOperator(task_id='sample-task', name='foo_bar')



 sens >> op >> hello_task
```

Example plugins/my_airflow_plugin.py

```
from airflow.plugins_manager import AirflowPlugin
from hooks.my_airflow_hook import *
from operators.my_airflow_operator import *

class PluginName(AirflowPlugin):

    name = 'my_airflow_plugin'

    hooks = [MyHook]
    operators = [MyOperator]
    sensors = [MySensor]
```

The following examples present each of the import statements needed in the custom plugin files.

Example hooks/my_airflow_hook.py

```
from airflow.hooks.base import BaseHook
```

```
class MyHook(BaseHook):

    def my_method(self):
        print("Hello World")
```

Example sensors/my_airflow_sensor.py

```
from airflow.sensors.base import BaseSensorOperator
from airflow.utils.decorators import apply_defaults


class MySensor(BaseSensorOperator):

    @apply_defaults
    def __init__(self,
                 *args,
                 **kwargs):
        super(MySensor, self).__init__(*args, **kwargs)

    def poke(self, context):
        return True
```

Example operators/my_airflow_operator.py

```
from airflow.operators.bash import BaseOperator
from airflow.utils.decorators import apply_defaults
from hooks.my_airflow_hook import MyHook


class MyOperator(BaseOperator):

    @apply_defaults
    def __init__(self,
                 my_field,
                 *args,
                 **kwargs):
        super(MyOperator, self).__init__(*args, **kwargs)
        self.my_field = my_field

    def execute(self, context):
        hook = MyHook('my_conn')
```

```
        hook.my_method()
```

Example operators/hello_operator.py

```
from airflow.models.baseoperator import BaseOperator
from airflow.utils.decorators import apply_defaults

class HelloOperator(BaseOperator):

    @apply_defaults
    def __init__(
            self,
            name: str,
            **kwargs) -> None:
        super().__init__(**kwargs)
        self.name = name

    def execute(self, context):
        message = "Hello {}".format(self.name)
        print(message)
        return message
```

Follow the steps in Testing custom plugins using the Amazon MWAA CLI utility, and then Creating a plugins.zip file to zip the contents **within** your `plugins` directory. For example, `cd plugins`.

Creating a plugins.zip file

The following steps describe the steps we recommend to create a plugins.zip file locally.

Step one: Test custom plugins using the Amazon MWAA CLI utility

- The command line interface (CLI) utility replicates an Amazon Managed Workflows for Apache Airflow environment locally.
- The CLI builds a Docker container image locally that's similar to an Amazon MWAA production image. You can use this to run a local Apache Airflow environment to develop and test DAGs, custom plugins, and dependencies before deploying to Amazon MWAA.
- To run the CLI, refer to aws-mwaa-docker-images on GitHub.

Step two: Create the plugins.zip file

You can use a built-in ZIP archive utility, or any other ZIP utility (such as 7zip) to create a .zip file.

> **Note**
>
> The built-in zip utility for Windows OS might add subfolders when you create a .zip file. We recommend verifying the contents of the plugins.zip file before uploading to your Amazon S3 bucket to ensure no additional directories were added.

1. Change directories to your local Airflow plugins directory. For example:

```
myproject$ cd plugins
```

2. Run the following command to ensure that the contents have executable permissions (macOS and Linux only).

```
plugins$ chmod -R 755 .
```

3. Zip the contents **within** your `plugins` folder.

```
plugins$ zip -r plugins.zip .
```

Uploading `plugins.zip` to Amazon S3

You can use the Amazon S3 console or the AWS Command Line Interface (AWS CLI) to upload a `plugins.zip` file to your Amazon S3 bucket.

Using the AWS CLI

The AWS Command Line Interface (AWS CLI) is an open source tool that you can use to interact with AWS services using commands in your command-line shell. To complete the steps on this page, you need the following:

- AWS CLI – Install version 2.
- AWS CLI – Quick configuration with `aws configure`.

To upload using the AWS CLI

1. In your command prompt, navigate to the directory where your `plugins.zip` file is stored. For example:

```
cd plugins
```

2. Use the following command to list all of your Amazon S3 buckets.

```
aws s3 ls
```

3. Use the following command to list the files and folders in the Amazon S3 bucket for your environment.

```
aws s3 ls s3://YOUR_S3_BUCKET_NAME
```

4. Use the following command to upload the `plugins.zip` file to the Amazon S3 bucket for your environment.

```
aws s3 cp plugins.zip s3://amzn-s3-demo-bucket/plugins.zip
```

Using the Amazon S3 console

The Amazon S3 console is a web-based user interface that you can use to create and manage the resources in your Amazon S3 bucket.

To upload using the Amazon S3 console

1. Open the Environments page on the Amazon MWAA console.
2. Choose an environment.
3. Select the **S3 bucket** link in the **DAG code in S3** pane to open your storage bucket in the console.
4. Choose **Upload**.
5. Choose **Add file**.
6. Select the local copy of your `plugins.zip`, choose **Upload**.

Installing custom plugins on your environment

This section describes how to install the custom plugins you uploaded to your Amazon S3 bucket by specifying the path to the plugins.zip file, and specifying the version of the plugins.zip file each time the zip file is updated.

Specifying the path to `plugins.zip` on the Amazon MWAA console (the first time)

If this is the first time you're uploading a `plugins.zip` to your Amazon S3 bucket, you also need to specify the path to the file on the Amazon MWAA console. You only need to complete this step once.

1. Open the Environments page on the Amazon MWAA console.
2. Choose an environment.
3. Choose **Edit**.
4. On the **DAG code in Amazon S3** pane, choose **Browse S3** adjacent to the **Plugins file - optional** field.
5. Select the `plugins.zip` file on your Amazon S3 bucket.
6. Choose **Choose**.
7. Choose **Next**, **Update environment**.

Specifying the `plugins.zip` version on the Amazon MWAA console

You need to specify the version of your `plugins.zip` file on the Amazon MWAA console each time you upload a new version of your `plugins.zip` in your Amazon S3 bucket.

1. Open the Environments page on the Amazon MWAA console.
2. Choose an environment.
3. Choose **Edit**.
4. On the **DAG code in Amazon S3** pane, choose a `plugins.zip` version in the dropdown list.
5. Choose **Next**.

Example use cases for plugins.zip

- Learn how to create a custom plugin in Custom plugin with Apache Hive and Hadoop.
- Learn how to create a custom plugin in Custom plugin with Oracle.
- Learn how to create a custom plugin in the section called “Changing a DAG's timezone”.

What's next?

Test your DAGs, custom plugins, and Python dependencies locally using aws-mwaa-docker-images on GitHub.

Installing Python dependencies

A Python dependency is any package or distribution not included in the Apache Airflow base install for your Apache Airflow version on your Amazon Managed Workflows for Apache Airflow environment. This topic describes the steps to install Apache Airflow Python dependencies on your Amazon MWAA environment using a `requirements.txt` file in your Amazon S3 bucket.

Contents

- Prerequisites
- How it works
- Python dependencies overview
  - Python dependencies location and size limits
- Creating a requirements.txt file
  - Step one: Test Python dependencies using the Amazon MWAA CLI utility
  - Step two: Create the requirements.txt
- Uploading requirements.txt to Amazon S3
  - Using the AWS CLI
  - Using the Amazon S3 console
- Installing Python dependencies on your environment
  - Specifying the path to requirements.txt on the Amazon MWAA console (the first time)
  - Specifying the requirements.txt version on the Amazon MWAA console
- Accessing logs for your requirements.txt

- What's next?

Prerequisites

You'll need the following before you can complete the steps on this page.

- **Permissions** — Your AWS account must have been granted access by your administrator to the AmazonMWAAFullConsoleAccess access control policy for your environment. In addition, your Amazon MWAA environment must be permitted by your execution role to access the AWS resources used by your environment.
- **Access** — If you require access to public repositories to install dependencies directly on the webserver, your environment must be configured with **public network** webserver access. For more information, refer to the section called "Apache Airflow access modes".
- **Amazon S3 configuration** — The Amazon S3 bucket used to store your DAGs, custom plugins in `plugins.zip`, and Python dependencies in `requirements.txt` must be configured with *Public Access Blocked* and *Versioning Enabled*.

How it works

On Amazon MWAA, you install all Python dependencies by uploading a `requirements.txt` file to your Amazon S3 bucket, then specifying the version of the file on the Amazon MWAA console each time you update the file. Amazon MWAA runs `pip3 install -r requirements.txt` to install the Python dependencies on the Apache Airflow scheduler and each of the workers.

To run Python dependencies on your environment, you must do three things:

1. Create a `requirements.txt` file locally.
2. Upload the local `requirements.txt` to your Amazon S3 bucket.
3. Specify the version of this file in the **Requirements file** field on the Amazon MWAA console.

> **Note**
>
> If this is the first time you're creating and uploading a `requirements.txt` to your Amazon S3 bucket, you also need to specify the path to the file on the Amazon MWAA console. You only need to complete this step once.

Python dependencies overview

You can install Apache Airflow extras and other Python dependencies from the Python Package Index (PyPi.org), Python wheels (`.whl`), or Python dependencies hosted on a private PyPi/PEP-503 Compliant Repo on your environment.

Python dependencies location and size limits

The Apache Airflow scheduler and the workers search for the packages in the `requirements.txt` file and the packages are installed on the environment at `/usr/local/airflow/.local/bin`.

- **Size limit**. We recommend a `requirements.txt` file that references libraries whose combined size is less than 1 GB. The more libraries Amazon MWAA needs to install, the longer the *startup* time on an environment. Although Amazon MWAA doesn't limit the size of installed libraries explicitly, if dependencies can't be installed within ten minutes, the Fargate service will time-out and attempt to rollback the environment to a stable state.

Creating a requirements.txt file

The following steps describe the steps we recommend to create a requirements.txt file locally.

Step one: Test Python dependencies using the Amazon MWAA CLI utility

- The command line interface (CLI) utility replicates an Amazon Managed Workflows for Apache Airflow environment locally.
- The CLI builds a Docker container image locally that's similar to an Amazon MWAA production image. You can use this to run a local Apache Airflow environment to develop and test DAGs, custom plugins, and dependencies before deploying to Amazon MWAA.
- To run the CLI, refer to aws-mwaa-docker-images on GitHub.

Step two: Create the `requirements.txt`

The following section describes how to specify Python dependencies from the Python Package Index in a `requirements.txt` file.

Apache Airflow v3

1. **Test locally**. Add additional libraries iteratively to find the right combination of packages and their versions, before creating a `requirements.txt` file. To run the Amazon MWAA CLI utility, refer to aws-mwaa-docker-images on GitHub.
2. **Review the Apache Airflow package extras**. To access a list of the packages installed for Apache Airflow v3 on Amazon MWAA, refer to aws-mwaa-docker-images `requirements.txt` on the GitHub website.
3. **Add a constraints statement**. Add the constraints file for your Apache Airflow v3 environment at the top of your `requirements.txt` file. Apache Airflow constraints files specify the provider versions available at the time of a Apache Airflow release.

   In the following example, replace *{Airflow-version}* with your environment's Airflow version. Replace *{Python-version}* with the Python version compatible with your environment.

   For information about the version of Python compatible with your Apache Airflow environment, refer to Apache Airflow Versions.

   ```
   --constraint "https://raw.githubusercontent.com/apache/airflow/
   constraints-{Airflow-version}/constraints-{Python-version}.txt"
   ```

   If the constraints file determines that `xyz==1.0` package is not compatible with other packages in your environment, `pip3 install` will fail to prevent incompatible libraries from being installed to your environment. If installation fails for any packages, you can access error logs for each Apache Airflow component (the scheduler, worker, and webserver) in the corresponding log stream on CloudWatch Logs. For more information about log types, refer to the section called “Accessing Airflow logs”.

   > **Updating constraint file versions**
   >
   > If you need to update the versions in the constraint file, refer to Constraints file.

4. **Apache Airflow packages**. Add the package extras and the version (==). This helps to prevent packages of the same name, but different version, from being installed on your environment.

```
apache-airflow[package-extra]==3.2.1
```

5. **Python libraries**. Add the package name and the version (==) in your `requirements.txt` file. This helps to prevent a future breaking update from PyPi.org from being automatically applied.

   ```
   library == version
   ```

   Example Boto3 and psycopg2-binary

   This example is provided for demonstration purposes. The boto and psycopg2-binary libraries are included with the base install for Apache Airflow v3 and don't need to be specified in a `requirements.txt` file.

   ```
   boto3==1.42.84
   botocore==1.42.84
   psycopg2-binary==2.9.11
   ```

   If a package is specified without a version, Amazon MWAA installs the latest version of the package from PyPi.org. This version might conflict with other packages in your `requirements.txt`.

Apache Airflow v2

1. **Test locally**. Add additional libraries iteratively to find the right combination of packages and their versions, before creating a `requirements.txt` file. To run the Amazon MWAA CLI utility, refer to aws-mwaa-docker-images on GitHub.
2. **Review the Apache Airflow package extras**. To access a list of the packages installed for Apache Airflow v2 on Amazon MWAA, access aws-mwaa-docker-images `requirements.txt` on the GitHub website.
3. **Add a constraints statement**. Add the constraints file for your Apache Airflow v2 environment at the top of your `requirements.txt` file. Apache Airflow constraints files specify the provider versions available at the time of a Apache Airflow release.

   Beginning with Apache Airflow v2.7.2, your requirements file must include a `--constraint` statement. If you do not provide a constraint, Amazon MWAA will specify one

for you to ensure the packages listed in your requirements are compatible with the version of Apache Airflow you are using.

In the following example, replace *{Airflow-version}* with your environment's Airflow version. Replace *{Python-version}* with the Python version compatible with your environment.

For information about the version of Python compatible with your Apache Airflow environment, refer to Apache Airflow Versions.

```
--constraint "https://raw.githubusercontent.com/apache/airflow/
constraints-{Airflow-version}/constraints-{Python-version}.txt"
```

If the constraints file determines that `xyz==1.0` package is not compatible with other packages in your environment, `pip3 install` will fail to prevent incompatible libraries from being installed to your environment. If installation fails for any packages, you can access error logs for each Apache Airflow component (the scheduler, worker, and webserver) in the corresponding log stream on CloudWatch Logs. For more information about log types, refer to the section called "Accessing Airflow logs".

> **Updating constraint file versions**
>
> If you need to update the versions in the constraint file, refer to Constraints file.

4. **Apache Airflow packages**. Add the package extras and the version (==). This helps to prevent packages of the same name, but different version, from being installed on your environment.

```
apache-airflow[package-extra]==2.5.1
```

5. **Python libraries**. Add the package name and the version (==) in your `requirements.txt` file. This helps to prevent a future breaking update from PyPi.org from being automatically applied.

```
library == version
```

Example Boto3 and psycopg2-binary

This example is provided for demonstration purposes. The boto and psycopg2-binary libraries are included with the Apache Airflow v2 base install and don't need to be specified in a `requirements.txt` file.

```
boto3==1.17.54
boto==2.49.0
botocore==1.20.54
psycopg2-binary==2.8.6
```

If a package is specified without a version, Amazon MWAA installs the latest version of the package from PyPi.org. This version might conflict with other packages in your `requirements.txt`.

Uploading `requirements.txt` to Amazon S3

You can use the Amazon S3 console or the AWS Command Line Interface (AWS CLI) to upload a `requirements.txt` file to your Amazon S3 bucket.

Using the AWS CLI

The AWS Command Line Interface (AWS CLI) is an open source tool that you can use to interact with AWS services using commands in your command-line shell. To complete the steps on this page, you need the following:

- AWS CLI – Install version 2.
- AWS CLI – Quick configuration with `aws configure`.

To upload using the AWS CLI

1. Use the following command to list all of your Amazon S3 buckets.

   ```
   aws s3 ls
   ```

2. Use the following command to list the files and folders in the Amazon S3 bucket for your environment.

```
aws s3 ls s3://YOUR_S3_BUCKET_NAME
```

3. The following command uploads a `requirements.txt` file to an Amazon S3 bucket.

```
aws s3 cp requirements.txt s3://amzn-s3-demo-bucket/requirements.txt
```

Using the Amazon S3 console

The Amazon S3 console is a web-based user interface that you can use to create and manage the resources in your Amazon S3 bucket.

To upload using the Amazon S3 console

1. Open the Environments page on the Amazon MWAA console.
2. Choose an environment.
3. Select the **S3 bucket** link in the **DAG code in S3** pane to open your storage bucket in the console.
4. Choose **Upload**.
5. Choose **Add file**.
6. Select the local copy of your `requirements.txt`, choose **Upload**.

Installing Python dependencies on your environment

This section describes how to install the dependencies you uploaded to your Amazon S3 bucket by specifying the path to the requirements.txt file, and specifying the version of the requirements.txt file each time it's updated.

Specifying the path to `requirements.txt` on the Amazon MWAA console (the first time)

If this is the first time you're creating and uploading a `requirements.txt` to your Amazon S3 bucket, you also need to specify the path to the file on the Amazon MWAA console. You only need to complete this step once.

1. Open the Environments page on the Amazon MWAA console.
2. Choose an environment.

3. Choose **Edit**.
4. On the **DAG code in Amazon S3** pane, choose **Browse S3** adjacent to the **Requirements file - optional** field.
5. Select the `requirements.txt` file on your Amazon S3 bucket.
6. Choose **Choose**.
7. Choose **Next**, **Update environment**.

You can begin using the new packages immediately after your environment finishes updating.

Specifying the `requirements.txt` version on the Amazon MWAA console

You need to specify the version of your `requirements.txt` file on the Amazon MWAA console each time you upload a new version of your `requirements.txt` in your Amazon S3 bucket.

1. Open the Environments page on the Amazon MWAA console.
2. Choose an environment.
3. Choose **Edit**.
4. On the **DAG code in Amazon S3** pane, choose a `requirements.txt` version in the dropdown list.
5. Choose **Next**, **Update environment**.

You can begin using the new packages immediately after your environment finishes updating.

Accessing logs for your `requirements.txt`

You can view Apache Airflow logs for the scheduler scheduling your workflows and parsing your `dags` folder. The following steps describe how to open the log group for the scheduler on the Amazon MWAA console, and access Apache Airflow logs on the CloudWatch Logs console.

To access logs for a `requirements.txt`

1. Open the Environments page on the Amazon MWAA console.
2. Choose an environment.
3. Choose the **Airflow scheduler log group** on the **Monitoring** pane.
4. Choose the `requirements_install_ip` log in **Log streams**.

5. Refer to the list of packages that were installed on the environment at `/usr/local/airflow/.local/bin`. For example:

```
Collecting appdirs==1.4.4 (from -r /usr/local/airflow/.local/bin (line 1))
Downloading https://files.pythonhosted.org/
packages/3b/00/2344469e2084fb28kjdsfiuyweb47389789vxbmnbjhsdgf5463acd6cf5e3db69324/
appdirs-1.4.4-py2.py3-none-any.whl
Collecting astroid==2.4.2 (from -r /usr/local/airflow/.local/bin (line 2))
```

6. Review the list of packages and whether any of these encountered an error during installation. If something went wrong, you get an error similar to the following:

```
2021-03-05T14:34:42.731-07:00
No matching distribution found for LibraryName==1.0.0 (from -r /usr/local/
airflow/.local/bin (line 4))
No matching distribution found for LibraryName==1.0.0 (from -r /usr/local/
airflow/.local/bin (line 4))
```

What's next?

Test your DAGs, custom plugins, and Python dependencies locally using aws-mwaa-docker-images on GitHub.

Deleting files on Amazon S3

This page describes how versioning works in an Amazon S3 bucket for an Amazon Managed Workflows for Apache Airflow environment, and the steps to delete a DAG, `plugins.zip`, or `requirements.txt` file.

Contents

- Prerequisites
- Versioning overview
- How it works
- Deleting a DAG on Amazon S3
- Removing a "current" requirements.txt or plugins.zip from an environment
- Deleting a "non-current" (previous) requirements.txt or plugins.zip version
- Using lifecycles to delete "non-current" (previous) versions and delete markers automatically

- Example lifecycle policy to delete requirements.txt "non-current" versions and delete markers automatically
- What's next?

Prerequisites

You'll need the following before you can complete the steps on this page.

- **Permissions** — Your AWS account must have been granted access by your administrator to the AmazonMWAAFullConsoleAccess access control policy for your environment. In addition, your Amazon MWAA environment must be permitted by your execution role to access the AWS resources used by your environment.
- **Access** — If you require access to public repositories to install dependencies directly on the webserver, your environment must be configured with **public network** webserver access. For more information, refer to the section called “Apache Airflow access modes”.
- **Amazon S3 configuration** — The Amazon S3 bucket used to store your DAGs, custom plugins in `plugins.zip`, and Python dependencies in `requirements.txt` must be configured with *Public Access Blocked* and *Versioning Enabled*.

Versioning overview

The `requirements.txt` and `plugins.zip` in your Amazon S3 bucket are versioned. When Amazon S3 bucket versioning is enabled for an object, and an artifact (for example, plugins.zip) is deleted from an Amazon S3 bucket, the file doesn't get deleted entirely. Anytime an artifact is deleted on Amazon S3, a new copy of the file is created that is a 404 (Object not found) error/0k file that says `I'm not here`. Amazon S3 calls this a *delete marker*. A delete marker is a "null" version of the file with a key name (or key) and version ID like any other object.

We recommend deleting file versions and delete markers periodically to reduce storage costs for your Amazon S3 bucket. To delete "non-current" (previous) file versions entirely, you must delete the versions of the files, and then the *delete marker* for the version.

How it works

Amazon MWAA runs a sync operation on your Amazon S3 bucket every thirty seconds. This causes any DAG deletions in an Amazon S3 bucket to be synced to the Airflow image of your Fargate container.

For `plugins.zip` and `requirements.txt` files, changes occur only after an environment update when Amazon MWAA builds a new Airflow image of your Fargate container with the custom plugins and Python dependencies. If you delete the *current* version of any of a `requirements.txt` or `plugins.zip` file, and then update your environment without providing a new version for the deleted file, then the update will fail with an error message, such as, `Unable to read version {version number} of file {file name}`.

Deleting a DAG on Amazon S3

A DAG file (`.py`) is not versioned and can be deleted directly on the Amazon S3 console. The following steps describe how to delete a DAG on your Amazon S3 bucket.

To delete a DAG

1. Open the Environments page on the Amazon MWAA console.
2. Choose an environment.
3. Select the **S3 bucket** link in the **DAG code in S3** pane to open your storage bucket in the console.
4. Choose the `dags` folder.
5. Select the DAG, **Delete**.
6. Under **Delete objects?**, type `delete`.
7. Choose **Delete objects**.

> **Note**
>
> Apache Airflow preserves historical DAG runs. After a DAG has been run in Apache Airflow, it remains in the Airflow DAGs list regardless of the file status, until you delete it in Apache Airflow. To delete a DAG in Apache Airflow, choose the red "delete" button in the **Links** column.

Removing a "current" requirements.txt or plugins.zip from an environment

Currently, there isn't a way to remove a plugins.zip or requirements.txt from an environment after they've been added, but we're working on the issue. In the interim, a workaround is to point to an empty text or zip file, respectively.

Deleting a "non-current" (previous) requirements.txt or plugins.zip version

The `requirements.txt` and `plugins.zip` files in your Amazon S3 bucket are versioned on Amazon MWAA. If you want to delete these files on your Amazon S3 bucket entirely, you must retrieve the current version (121212) of the object (for example, plugins.zip), delete the version, and then remove the *delete marker* for the file versions.

You can also delete "non-current" (previous) file versions on the Amazon S3 console; however, you'll still need to delete the *delete marker* using one of the following options.

- To retrieve the object version, refer to Retrieving object versions from a versioning-enabled bucket *in the Amazon S3 guide*.
- To delete the object version, refer to Deleting object versions from a versioning-enabled bucket *in the Amazon S3 guide*.
- To remove a delete marker, refer to Managing delete markers *in the Amazon S3 guide*.

Using lifecycles to delete "non-current" (previous) versions and delete markers automatically

You can configure a lifecycle policy for your Amazon S3 bucket to delete "non-current" (previous) versions of the plugins.zip and requirements.txt files in your Amazon S3 bucket after a certain number of days, or to remove an expired object's delete marker.

1. Open the Environments page on the Amazon MWAA console.
2. Choose an environment.
3. Under **DAG code in Amazon S3**, choose your Amazon S3 bucket.
4. Choose **Create lifecycle rule**.

Example lifecycle policy to delete requirements.txt "non-current" versions and delete markers automatically

Use the following example to create a lifecycle rule that permanently deletes "non-current" versions of a requirements.txt file and their delete markers after thirty days.

1. Open the Environments page on the Amazon MWAA console.
2. Choose an environment.
3. Under **DAG code in Amazon S3**, choose your Amazon S3 bucket.
4. Choose **Create lifecycle rule**.
5. In **Lifecycle rule name**, type `Delete previous requirements.txt versions and delete markers after thirty days`.
6. In **Prefix**, **requirements**.
7. In **Lifecycle rule actions**, choose **Permanently delete previous versions of objects** and **Delete expired delete markers or incomplete multipart uploads**.
8. In **Number of days after objects become previous versions**, type `30`.
9. In **Expired object delete markers**, choose **Delete expired object delete markers, objects are permanently deleted after 30 days**.

What's next?

- Learn more about Amazon S3 delete markers in Managing delete markers.
- Learn more about Amazon S3 lifecycles in Expiring objects.

Using agent skills with Amazon MWAA

Agent skills are instruction packages that guide an AI coding assistant through Amazon Managed Workflows for Apache Airflow operations. The AWS Data Analytics plugin (`aws-data-analytics`) in the open-source Agent Toolkit for AWS provides a set of skills that follow the open Agent Skills standard. After you install the skills, you describe what you want to do in natural language, and your assistant uses the matching skill to do the work. This guide describes the Amazon MWAA skills, how to install them for several assistants, and example prompts that you can use.

Topics

- Overview
- Installing the skills
- Example prompts
- Related resources

Overview

The AWS Data Analytics plugin (`aws-data-analytics`) in the open-source Agent Toolkit for AWS on the GitHub website provides agent skills for Amazon Managed Workflows for Apache Airflow.

A skill is an instruction package that guides an AI coding assistant (such as Kiro, Claude Code, Amazon Q, Cursor, or Codex) through a common Amazon MWAA operation. These skills follow the open Agent Skills standard on the Anthropic website, so you can use them across assistants that support the standard.

The Amazon MWAA skills work with both Amazon MWAA provisioned environments and Amazon MWAA Serverless, except where noted. In provisioned environments, you author workflows as Python directed acyclic graphs (DAGs). In Amazon MWAA Serverless, you author workflows as YAML.

The plugin includes the following four Amazon MWAA skills.

- `authoring-mwaa-workflow` authors and deploys Apache Airflow DAGs for provisioned environments and YAML workflows for Amazon MWAA Serverless.
- `debugging-mwaa-workflow` finds the root cause of failed runs and task instances, DAGs that do not appear, import errors, worker out-of-memory errors, and IAM permission errors.

- `testing-mwaa-workflow` triggers a deployed DAG or workflow, monitors it to completion, and retests after fixes.
- `upgrading-mwaa-environments` upgrades a provisioned environment to a newer Apache Airflow version, including 2.x to 3.x, by using a blue-green or in-place strategy. This skill supports provisioned environments only, not Amazon MWAA Serverless.

The skills run AWS Command Line Interface (AWS CLI) commands to interact with your AWS resources. We recommend that you also install the AWS Model Context Protocol (MCP) Server. The skills use the AWS MCP Server to fetch current AWS documentation at runtime and to run AWS CLI commands with sandboxed execution and audit logging.

Installing the skills

The Amazon MWAA skills are part of the `aws-data-analytics` plugin in the Agent Toolkit for AWS. To install the AWS MCP Server and set up your coding assistant, such as Kiro, Claude Code, Codex, Cursor, or the AWS CLI, follow Getting started in the Agent Toolkit for AWS User Guide. When you install plugins or skills, choose the `aws-data-analytics` plugin, which provides the Amazon MWAA skills.

Prerequisites

Complete the prerequisites in Getting started. In addition, grant your AWS Identity and Access Management (IAM) principal permissions for the Amazon MWAA action namespaces that you use, scoped to the resources that you use. Use the `airflow` namespace for Amazon MWAA environments and the Apache Airflow REST API, and the `airflow-serverless` namespace for Amazon MWAA Serverless workflows and runs.

Installing the aws-data-analytics plugin

Install the `aws-data-analytics` plugin for your assistant. For example, in Claude Code, run the following command:

```
/plugin install aws-data-analytics@claude-plugins-official
```

For Kiro, Codex, Cursor, and other assistants, install the `aws-data-analytics` plugin or its skills as described in Getting started. To install the skills directly from the repository, run the following command:

```
npx skills add aws/agent-toolkit-for-aws/skills
```

Example prompts

After you install the skills, you can give your assistant natural-language prompts to complete Amazon MWAA operations. The following examples are grouped by skill.

Use the following prompts to author and deploy workflows with the `authoring-mwaa-workflow` skill.

- "Create an Apache Airflow DAG that runs a daily data export and deploy it to my environment"
- "Build a YAML workflow for Amazon MWAA Serverless"
- "Write a pipeline that orchestrates three tasks and schedule it to run every hour"

Use the following prompts to test workflows with the `testing-mwaa-workflow` skill.

- "Test my DAG and watch it until it finishes"
- "Trigger a test run of my workflow and report whether it succeeded"
- "Smoke-test the pipeline I just deployed"

Use the following prompts to debug workflows with the `debugging-mwaa-workflow` skill.

- "Find the root cause of my failed DAG"
- "My DAG isn't showing up in the Airflow UI"
- "Why is my worker running out of memory?"

Use the following prompts to upgrade environments with the `upgrading-mwaa-environments` skill.

- "Upgrade my environment from Apache Airflow 2 to 3"
- "Plan a blue-green upgrade that preserves my run history"
- "Migrate my MWAA environment to a newer Apache Airflow version"

You don't need to name a skill in your prompt. Your assistant selects the matching skill automatically based on your request.

Related resources

For more information about the AWS agent skills for Amazon MWAA, see the following resources.

- Getting started in the Agent Toolkit for AWS User Guide.
- The Agent Toolkit for AWS on the GitHub website.
- The AWS Data Analytics plugin on the GitHub website, which provides the Amazon MWAA skills.
- The AWS MCP Server documentation.
- The Agent Skills open standard on the Anthropic website.

Networking

This guide describes the Amazon VPC network setup you'll need for an Amazon MWAA environment.

Sections

- About networking on Amazon MWAA
- Security in your VPC on Amazon MWAA
- Managing access to service-specific Amazon VPC endpoints on Amazon MWAA
- Creating the required VPC service endpoints in an Amazon VPC with private routing
- Managing your own Amazon VPC endpoints on Amazon MWAA

About networking on Amazon MWAA

A Amazon VPC is a virtual network linked to your AWS account. It gives you cloud security and the ability to scale dynamically by providing fine-grained control over your virtual infrastructure and network traffic segmentation. This page describes the Amazon VPC infrastructure with *public routing*, *private routing*, or *both public and private routing* needed to support an Amazon Managed Workflows for Apache Airflow environment.

Contents

- Terms
- What's supported
- VPC infrastructure overview
  - Public routing over the internet
  - Private routing without internet access
- Example use cases for an Amazon VPC and Apache Airflow access mode
  - Internet access is allowed - new Amazon VPC network
  - Internet access is not allowed - new Amazon VPC network
  - Internet access is not allowed - existing Amazon VPC network

Terms

Public routing

An Amazon VPC network that has access to the internet.

Private routing

An Amazon VPC network **without** access to the internet.

What's supported

The following table describes the types of Amazon VPCs Amazon MWAA supports.

Amazon VPC types	Supported	
An Amazon VPC owned by the account that is attemptin g to create the environment.	Yes	
A shared Amazon VPC where multiple AWS accounts create their AWS resources.	Yes	

VPC infrastructure overview

When you create an Amazon MWAA environment, Amazon MWAA creates between one to two VPC endpoints for your environment based on the Apache Airflow access mode you chose for your environment. These endpoints are shown as Elastic Network Interfaces (ENIs) with private IPs in your Amazon VPC. After these endpoints are created, any traffic destined to these IPs is privately or publicly routed to the corresponding AWS services used by your environment.

The following section describes the Amazon VPC infrastructure required to route traffic publicly over the internet, or privately *within your Amazon VPC*.

Public routing over the internet

This section describes the Amazon VPC infrastructure of an environment with public routing. You'll need the following VPC infrastructure:

- **One VPC security group**. A VPC security group acts as a virtual firewall to control ingress (inbound) and egress (outbound) network traffic on an instance.
  - Up to 5 security groups can be specified.
  - The security group must specify a self-referencing inbound rule to itself.
  - The security group must specify an outbound rule for all traffic (`0.0.0.0/0`; for IPv6, use `::/0`).
  - The security group must allow all traffic in the self-referencing rule. For example, (Recommended) Example all access self-referencing security group.
  - The security group can *optionally* restrict traffic further by specifying the port range for HTTPS port range `443` and a TCP port range `5432`. For example, (Optional) Example security group that restricts inbound access to port 5432 and (Optional) Example security group that restricts inbound access to port 443.
- **Two public subnets**. A public subnet is a subnet that's associated with a route table that has a route to an internet gateway.
  - Two public subnets are required. This allows Amazon MWAA to build a new container image for your environment in your other availability zone, if one container fails.
  - The subnets must be in different Availability Zones. For example, `us-east-1a`, `us-east-1b`.
  - The subnets must have a route table that directs internet-bound traffic to an internet gateway.
- **Two private subnets**. A private subnet is a subnet that's **not** associated with a route table that has a route to an internet gateway.
  - Two private subnets are required. This allows Amazon MWAA to build a new container image for your environment in your other availability zone, if one container fails.
  - The subnets must be in different Availability Zones. For example, `us-east-1a`, `us-east-1b`.
  - The subnets *must* have a route table to a NAT device (gateway or instance).
  - The subnets **must not** route to an internet gateway.
  - Set `assignIpV6AddressOnCreation` to `true` for IPv6 subnets.
  - For IPv6 private subnets, you must have a connection to an egress-only internet gateway (EIGW).
- **A network access control list (ACL)**. An NACL manages (by allow or deny rules) inbound and outbound traffic at the subnet level.
  - The NACL must have an inbound rule that allows all traffic (`0.0.0.0/0`; for IPv6, use `::/0`).
  - The NACL must have an outbound rule that allows all traffic (`0.0.0.0/0`; for IPv6, use `::/0`).

  - For example, (Recommended) Example ACLs.
- **A NAT gateway (or NAT instance)**. With a NAT gateway (or NAT instance), you can route traffic from instances in the private subnet to the internet or other AWS services, and receive responses.
  - The NAT device must be attached to a public subnet.
  - The NAT device must have an Elastic IPv4 Address (EIP) attached to each public subnet.

> **High availability**
>
> For high availability, consider provisioning a separate NAT device in each of the public subnets.

- **An internet gateway**. An internet gateway connects an Amazon VPC to the internet and other AWS services.
  - An internet gateway must be attached to the Amazon VPC.

Private routing without internet access

This section describes the Amazon VPC infrastructure of an environment with *private routing*. You'll need the following VPC infrastructure:

- **One VPC security group**. A VPC security group acts as a virtual firewall to control ingress (inbound) and egress (outbound) network traffic on an instance.
  - Up to 5 security groups can be specified.
  - The security group must specify a self-referencing inbound rule to itself.
  - The security group must specify an outbound rule for all traffic (`0.0.0.0/0`; for IPv6, use `::/0`).
  - The security group must allow all traffic in the self-referencing rule. For example, (Recommended) Example all access self-referencing security group.
  - The security group can *optionally* restrict traffic further by specifying the port range for HTTPS port range 443 and a TCP port range 5432. For example, (Optional) Example security group that restricts inbound access to port 5432 and (Optional) Example security group that restricts inbound access to port 443.
- **Two private subnets**. A private subnet is a subnet that's **not** associated with a route table that has a route to an internet gateway.

  - Two private subnets are required. This allows Amazon MWAA to build a new container image for your environment in your other availability zone, if one container fails.
  - The subnets must be in different Availability Zones. For example, `us-east-1a`, `us-east-1b`.
  - The subnets must have a route table to your VPC endpoints.
  - The subnets must have a route table to an EIGW to download from the internet as part of a DAG.
  - The subnets **must not** have a route table to a NAT device (gateway or instance), **nor** an internet gateway.
- **A network access control list (ACL)**. An NACL manages (by allow or deny rules) inbound and outbound traffic at the subnet level.
  - The NACL must have an inbound rule that allows all traffic (`0.0.0.0/0`; for IPv6, use `::/0`).
  - The NACL must have an outbound rule that allows all traffic (`0.0.0.0/0`; for IPv6, use `::/0`).
  - For example, (Recommended) Example ACLs.
- **A local route table**. A local route table is a default route for communication within the VPC.
  - The local route table must be associated to your private subnets.
  - The local route table must enable instances in your VPC to communicate with your own network. For example, if you're using an AWS Client VPN to access the VPC interface endpoint for your Apache Airflow webserver, the route table must route to the VPC endpoint.
- **VPC endpoints** for each AWS service used by your environment, and Apache Airflow VPC endpoints in the same AWS Region and Amazon VPC as your Amazon MWAA environment.
  - A VPC endpoint for each AWS service used by the environment and VPC endpoints for Apache Airflow. For example, (Required) VPC endpoints.
  - The VPC endpoints must have private DNS enabled.
  - The VPC endpoints must be associated to your environment's two private subnets.
  - The VPC endpoints must be associated to your environment's security group.
  - The VPC endpoint policy for each endpoint must be configured to allow access to AWS services used by the environment. For example, (Recommended) Example VPC endpoint policy to allow all access.
  - A VPC endpoint policy for Amazon S3 must be configured to allow bucket access. For example, (Recommended) Example Amazon S3 gateway endpoint policy to allow bucket access.

Example use cases for an Amazon VPC and Apache Airflow access mode

This section descibes the different use cases for network access in your Amazon VPC and the Apache Airflow webserver access mode choose on the Amazon MWAA console.

Internet access is allowed - new Amazon VPC network

If internet access in your VPC is allowed by your organization, *and* you would like users to access your Apache Airflow webserver over the internet:

1. Create an Amazon VPC network with internet access.
2. Create an environment with the **Public network** access mode for your Apache Airflow webserver.
3. **What we recommend**: We recommend using the CloudFormation quick-start template that creates the Amazon VPC infrastructure, an Amazon S3 bucket, and an Amazon MWAA environment at the same time. To learn more, refer to Quick start tutorial for Amazon Managed Workflows for Apache Airflow.

If internet access in your VPC is allowed by your organization, *and* you would like to limit Apache Airflow webserver access to users within your VPC:

1. Create an Amazon VPC network with internet access.
2. Create a mechanism to access the VPC interface endpoint for your Apache Airflow webserver from your computer.
3. Create an environment with the **Private network** access mode for your Apache Airflow webserver.
4. **What we recommend**:
    a. We recommend using the Amazon MWAA console in Option one: Creating the VPC network on the Amazon MWAA console, or the CloudFormation template in Option two: Creating an Amazon VPC network *with* internet access.
    b. We recommend configuring access using an AWS Client VPN to your Apache Airflow webserver in Tutorial: Configuring private network access using an AWS Client VPN.

Internet access is not allowed - new Amazon VPC network

If internet access in your VPC is **not allowed** by your organization, *and* you want the Apache Airflow web server to remain accessible over the internet:

1. Create an Amazon VPC network without internet access.
2. Create an environment with the **Both public and private network access** mode for your Apache Airflow webserver. This option is available for Apache Airflow version 3.2.1 and later.
3. **What we recommend**: We recommend using the CloudFormation template to create an Amazon VPC without internet access and the VPC endpoints for each AWS service used by Amazon MWAA in Option three: Creating an Amazon VPC network *without* internet access.

If internet access in your VPC is **not allowed** by your organization, *and* you want to limit Apache Airflow webserver access to users within your VPC:

1. Create an Amazon VPC network without internet access.
2. Create a mechanism to access the VPC interface endpoint for your Apache Airflow webserver from your computer.
3. Create VPC endpoints for each AWS service used by your environment.
4. Create an environment with the **Private network** access mode for your Apache Airflow webserver.
5. **What we recommend**:
    a. We recommend using the CloudFormation template to create an Amazon VPC without internet access and the VPC endpoints for each AWS service used by Amazon MWAA in Option three: Creating an Amazon VPC network *without* internet access.
    b. We recommend configuring access using an AWS Client VPN to your Apache Airflow webserver in Tutorial: Configuring private network access using an AWS Client VPN.

Internet access is not allowed - existing Amazon VPC network

If internet access in your VPC is **not allowed** by your organization, *and* you already have the required Amazon VPC network without internet access:

1. Create VPC endpoints for each AWS service used by your environment.
2. Create VPC endpoints for Apache Airflow.

3. Create a mechanism to access the VPC interface endpoint for your Apache Airflow webserver from your computer.
4. Create an environment with the **Private network** access mode for your Apache Airflow webserver.
5. **What we recommend**:
    a. We recommend creating and attaching the VPC endpoints needed for each AWS service used by Amazon MWAA, and the VPC endpoints needed for Apache Airflow in Creating the required VPC service endpoints in an Amazon VPC with private routing.
    b. We recommend configuring access using an AWS Client VPN to your Apache Airflow webserver in Tutorial: Configuring private network access using an AWS Client VPN.

Security in your VPC on Amazon MWAA

This page describes the Amazon VPC components used to secure your Amazon Managed Workflows for Apache Airflow environment and the configurations needed for these components.

Contents

- Terms
- Security overview
- Network access control lists (ACLs)
    - (Recommended) Example ACLs
- VPC security groups
    - (Recommended) Example all access self-referencing security group
    - (Optional) Example security group that restricts inbound access to port 5432
    - (Optional) Example security group that restricts inbound access to port 443
- VPC endpoint policies (private routing only)
    - (Recommended) Example VPC endpoint policy to allow all access
    - (Recommended) Example Amazon S3 gateway endpoint policy to allow bucket access

Terms

Public routing

An Amazon VPC network that has access to the internet.

Private routing

An Amazon VPC network without access to the internet.

Security overview

Security groups and access control lists (ACLs) provide ways to control the network traffic across the subnets and instances in your Amazon VPC using rules you specify.

- Network traffic to and from a subnet can be controlled by Access Control Lists (ACLs). You only need one ACL, and the same ACL can be used on multiple environments.
- Network traffic to and from an instance can be controlled by an Amazon VPC security group. You can use between one to five security groups per environment.
- Network traffic to and from an instance can also be controlled by VPC endpoint policies. If internet access within your Amazon VPC is not allowed by your organization and you're using an Amazon VPC network with *private routing*, a VPC endpoint policy is required for the AWS VPC endpoints and Apache Airflow VPC endpoints.

Network access control lists (ACLs)

A network access control list (ACL) can manage (by allow or deny rules) inbound and outbound traffic at the *subnet* level. An ACL is stateless, which means that inbound and outbound rules must be specified separately and explicitly. It is used to specify the types of network traffic that are allowed in or out from the instances in a VPC network.

Every Amazon VPC has a default ACL that allows all inbound and outbound traffic. You can edit the default ACL rules, or create a custom ACL and attach it to your subnets. A subnet can only have one ACL attached to it at any time, but one ACL can be attached to multiple subnets.

(Recommended) Example ACLs

The following example displays the *inbound* and *outbound* ACL rules that can be used for an Amazon VPC with *public routing* or *private routing*.

Rule number	Type	Protocol	Port range	Source	Allow/Deny
100	All IPv4 traffic	All	All	0.0.0.0/0	Allow
*	All IPv4 traffic	All	All	0.0.0.0/0	Deny

VPC security groups

A VPC security group acts as a virtual firewall that controls the network traffic at the *instance* level. A security group is stateful, which means that when an inbound connection is permitted, it is allowed to reply. It is used to specify the types of network traffic that are allowed in from the instances in a VPC network.

Every Amazon VPC has a default security group. By default, it has no inbound rules. It has an outbound rule that allows all outbound traffic. You can edit the default security group rules, or create a custom security group and attach it to your Amazon VPC.

(Recommended) Example all access self-referencing security group

The following example displays the *inbound* security group rules that allows all traffic for an Amazon VPC with *public routing* or *private routing*. The security group in this example is a self-referencing rule to itself.

Type	Protocol	Source Type	Source		
All traffic	All	All	sg-0909e8e81919 / my-mwaa-vpc-security-group		

The following example displays the *outbound* security group rules.

Type	Protocol	Source Type	Source		
All traffic	All	All	0.0.0.0/0		

(Optional) Example security group that restricts inbound access to port 5432

The following example displays the *inbound* security group rules that allow all HTTPS traffic on port 5432 for the Amazon Aurora PostgreSQL metadata database (owned by Amazon MWAA) for your environment.

> **Note**
>
> If you choose to restrict traffic using this rule, you'll need to add another rule to allow TCP traffic on port 443.

Type	Protocol	Port range	Source type	Source	
Custom TCP	TCP	5432	Custom	sg-0909e8e81919 / my-mwaa-vpc-security-group	

(Optional) Example security group that restricts inbound access to port 443

The following example displays the *inbound* security group rules that allow all TCP traffic on port 443 for the Apache Airflow webserver.

Type	Protocol	Port range	Source type	Source	
HTTPS	TCP	443	Custom	sg-0909e8e81919 / my-mwaa-v	

Type	Protocol	Port range	Source type	Source	
				pc-security-group	

VPC endpoint policies (private routing only)

A VPC endpoint (AWS PrivateLink) policy controls access to AWS services from your private subnet. A VPC endpoint policy is an IAM resource policy that you attach to your VPC gateway or interface endpoint. This section describes the permissions needed for the VPC endpoint policies for each VPC endpoint.

We recommend using a VPC interface endpoint policy for each of the VPC endpoints you created that allows full access to all AWS services, and using your execution role exclusively for AWS permissions.

(Recommended) Example VPC endpoint policy to allow all access

The following example presents a VPC interface endpoint policy for an Amazon VPC with *private routing*.

```
{
  "Statement": [
    {
      "Action": "*",
      "Effect": "Allow",
      "Resource": "*",
      "Principal": "*"
    }
  ]
}
```

(Recommended) Example Amazon S3 gateway endpoint policy to allow bucket access

The following example presents a VPC gateway endpoint policy that provides access to the Amazon S3 buckets required for Amazon ECR operations for an Amazon VPC with *private routing*. This is required for your Amazon ECR image to be retrieved, in addition to the bucket where your DAGs and supporting files are stored.

```
{
  "Statement": [
    {
      "Sid": "Access-to-specific-bucket-only",
      "Principal": "*",
      "Action": [
        "s3:GetObject"
      ],
      "Effect": "Allow",
      "Resource": ["arn:aws:s3:::prod-us-east-1-starport-layer-bucket/*"]
    }
  ]
}
```

Managing access to service-specific Amazon VPC endpoints on Amazon MWAA

A VPC endpoint (AWS PrivateLink) lets you privately connect your VPC to services hosted on AWS without requiring an internet gateway, a NAT device, VPN, or firewall proxies. These endpoints are horizontally scalable and highly available virtual devices that allow communication between instances in your VPC and AWS services. This page describes the VPC endpoints created by Amazon MWAA, and how to access the VPC endpoint for your Apache Airflow webserver if you've chosen the **Private network** access mode on Amazon Managed Workflows for Apache Airflow.

> **Note**
>
> If you chose **Both public and private network access**, the VPC endpoint for the webserver is created and managed automatically by Amazon MWAA for worker connectivity. No additional VPC endpoint setup is required, and you do not need to configure access to the VPC endpoint for browser access to the Apache Airflow UI.

Contents

- Pricing
- VPC endpoint overview
  - Public network access mode
  - Private network access mode

- Permission to use other AWS services
- Accessing VPC endpoints
  - Accessing VPC endpoints on the Amazon VPC console
  - Identifying the private IP addresses of your Apache Airflow webserver and its VPC endpoint
- Accessing the VPC endpoint for your Apache Airflow webserver (private network access)
  - Using an AWS Client VPN
  - Using a Linux Bastion Host
  - Using a Load Balancer (advanced)

Pricing

- AWS PrivateLink Pricing

VPC endpoint overview

When you create an Amazon MWAA environment, Amazon MWAA creates between one to two VPC endpoints for your environment. These endpoints are shown as Elastic Network Interfaces (ENIs) with private IPs in your Amazon VPC. After these endpoints are created, any traffic destined to these IPs is privately or publicly routed to the corresponding AWS services used by your environment.

Public network access mode

If you chose the **Public network** access mode for your Apache Airflow webserver, network traffic is publicly routed over the internet.

- Amazon MWAA creates a VPC interface endpoint for your Amazon Aurora PostgreSQL metadata database. The endpoint is created in the Availability Zones mapped to your private subnets and is independent from other AWS accounts.
- Amazon MWAA then binds an IP address from your private subnets to the interface endpoints. This is designed to support the best practice of binding a single IP from each Availability Zone of the Amazon VPC.

Private network access mode

If you chose the **Private network** access mode for your Apache Airflow webserver, network traffic is privately routed *within your Amazon VPC*.

- Amazon MWAA creates a VPC interface endpoint for your Apache Airflow webserver, and an interface endpoint for your Amazon Aurora PostgreSQL metadata database. The endpoints are created in the Availability Zones mapped to your private subnets and are independent from other AWS accounts.
- Amazon MWAA then binds an IP address from your private subnets to the interface endpoints. This is designed to support the best practice of binding a single IP from each Availability Zone of the Amazon VPC.

Permission to use other AWS services

The interface endpoints use the execution role for your environment in AWS Identity and Access Management (IAM) to manage permission to AWS resources used by your environment. As more AWS services are turned on for an environment, each service requires you to configure permission using your environment's execution role. To add permissions, refer to Amazon MWAA execution role.

If you've chosen the **Private network** access mode for your Apache Airflow webserver, you must also allow permission in the VPC endpoint policy for each endpoint. To learn more, refer to the section called “VPC endpoint policies (private routing only)”.

Accessing VPC endpoints

This section describes how to access the VPC endpoints created by Amazon MWAA, and how to identify the private IP addresses for your Apache Airflow VPC endpoint.

Accessing VPC endpoints on the Amazon VPC console

The following section displays the steps to access the VPC endpoints created by Amazon MWAA, and any VPC endpoints you might have created if you're using *private routing* for your Amazon VPC.

To access the VPC endpoints

1. Open the Endpoints page on the Amazon VPC console.

2. Select your AWS Region.
3. Refer to the VPC interface endpoints created by Amazon MWAA, and any VPC endpoints you might have created if you're using *private routing* in your Amazon VPC.

To learn more about the VPC service endpoints that are required for an Amazon VPC with *private routing*, refer to Creating the required VPC service endpoints in an Amazon VPC with private routing.

Identifying the private IP addresses of your Apache Airflow webserver and its VPC endpoint

The following steps describe how to retrieve the host name of your Apache Airflow webserver and its VPC interface endpoint, and their private IP addresses.

1. Use the following AWS Command Line Interface (AWS CLI) command to retrieve the host name for your Apache Airflow webserver.

   ```
   aws mwaa get-environment --name YOUR_ENVIRONMENT_NAME --query
    'Environment.WebserverUrl'
   ```

   You get something similar to the following response:

   ```
   "99aa99aa-55aa-44a1-a91f-f4552cf4e2f5-vpce.c10.us-west-2.airflow.amazonaws.com"
   ```

2. Run a *dig* command on the host name returned in the response of the previous command. For example:

   ```
   dig CNAME +short 99aa99aa-55aa-44a1-a91f-f4552cf4e2f5-vpce.c10.us-
   west-2.airflow.amazonaws.com
   ```

   You get something similar to the following response:

   ```
   vpce-0699aa333a0a0a0-bf90xjtr.vpce-svc-00bb7c2ca2213bc37.us-
   west-2.vpce.amazonaws.com.
   ```

3. Use the following AWS Command Line Interface (AWS CLI) command to retrieve the VPC endpoint DNS name returned in the response of the previous command. For example:

```
aws ec2 describe-vpc-endpoints | grep vpce-0699aa333a0a0a0-bf90xjtr.vpce-
svc-00bb7c2ca2213bc37.us-west-2.vpce.amazonaws.com.
```

You get something similar to the following response:

```
"DnsName": "vpce-066777a0a0a0-bf90xjtr.vpce-svc-00bb7c2ca2213bc37.us-
west-2.vpce.amazonaws.com",
```

4. Run either an *nslookup* or *dig* command on your Apache Airflow host name and its VPC endpoint DNS name to retrieve the IP addresses. For example:

```
dig +short YOUR_AIRFLOW_HOST_NAME YOUR_AIRFLOW_VPC_ENDPOINT_DNS
```

You get something similar to the following response:

```
192.0.5.1
192.0.6.1
```

Accessing the VPC endpoint for your Apache Airflow webserver (private network access)

If you've chosen the **Private network** access mode for your Apache Airflow webserver, you'll need to create a mechanism to access the VPC interface endpoint for your Apache Airflow webserver. You must use the same Amazon VPC, VPC security group, and private subnets as your Amazon MWAA environment for these resources.

Using an AWS Client VPN

AWS Client VPN is a managed client-based VPN service that you can use to securely access your AWS resources and resources in your on-premises network. It provides a secure TLS connection from any location using the OpenVPN client.

We recommend following the Amazon MWAA tutorial to configure a Client VPN: Tutorial: Configuring private network access using an AWS Client VPN.

Using a Linux Bastion Host

A bastion host is a server whose purpose is to provide access to a private network from an external network, such as over the internet from your computer. Linux instances are in a public subnet, and they are set up with a security group that allows SSH access from the security group attached to the underlying Amazon EC2 instance running the bastion host.

We recommend following the Amazon MWAA tutorial to configure a Linux Bastion Host: Tutorial: Configuring private network access using a Linux Bastion Host.

Using a Load Balancer (advanced)

The following section displays the configurations you'll need to apply to an Application Load Balancer.

1. **Target groups**. You'll need to use target groups that point to the private IP addresses for your Apache Airflow webserver, and its VPC interface endpoint. We recommend specifying both private IP addresses as your registered targets, as using only one can reduce availability. For more information about how to identify the private IP addresses, refer to the section called "Identifying the private IP addresses of your Apache Airflow webserver and its VPC endpoint".
2. **Status codes**. We recommend using `200` and `302` status codes in your target group settings. Otherwise, the targets might be flagged as unhealthy if the VPC endpoint for the Apache Airflow webserver responds with a `302 Redirect` error.
3. **HTTPS Listener**. You'll need to specify the target port for the Apache Airflow webserver. For example:

Protocol	Port
HTTPS	443

4. **ACM new domain**. If you want to associate an SSL/TLS certificate in AWS Certificate Manager, you'll need to create a new domain for the HTTPS listener for your load balancer.
5. **ACM certificate region**. If you want to associate an SSL/TLS certificate in AWS Certificate Manager, you'll need to upload to the same AWS Region as your environment. For example:

- **Example region to upload certificate**

```
aws acm import-certificate --certificate fileb://Certificate.pem --certificate-
chain fileb://CertificateChain.pem --private-key fileb://PrivateKey.pem --
region us-west-2
```

Creating the required VPC service endpoints in an Amazon VPC with private routing

An existing Amazon VPC network without internet access needs additional VPC service endpoints (AWS PrivateLink) to use Apache Airflow on Amazon Managed Workflows for Apache Airflow. This page describes the VPC endpoints required for the AWS services used by Amazon MWAA, the VPC endpoints required for Apache Airflow, and how to create and attach the VPC endpoints to an existing Amazon VPC with private routing.

> **Note**
>
> If you chose **Both public and private network access**, the VPC endpoint for the webserver is created and managed automatically by Amazon MWAA. You do not need to create VPC endpoints for Apache Airflow webserver connectivity. However, if your Amazon VPC does not have internet access, you still need VPC endpoints for other AWS services (such as Amazon S3, CloudWatch Logs, SQS, and KMS) as described on this page.

Contents

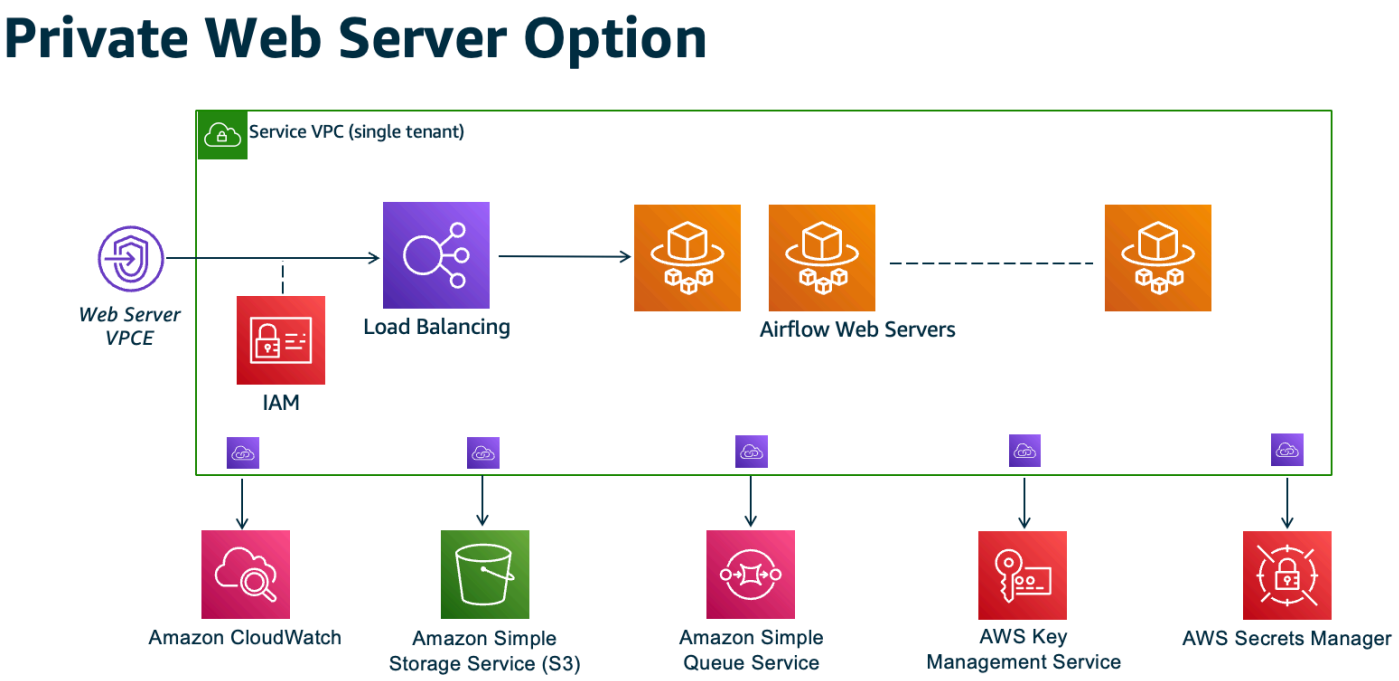
- Pricing
- Private network and private routing
- (Required) VPC endpoints
- Attaching the required VPC endpoints
  - VPC endpoints required for AWS services
  - VPC endpoints required for Apache Airflow
- (Optional) Enable private IP addresses for your Amazon S3 VPC interface endpoint
  - Using Route 53

- VPCs with custom DNS

Pricing

- AWS PrivateLink Pricing

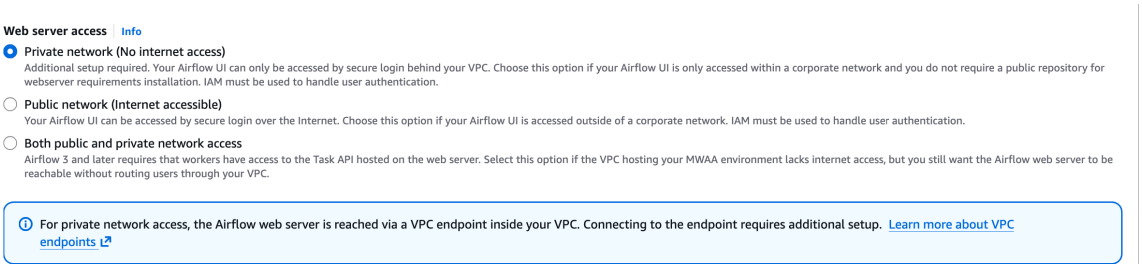
Private network and private routing



The private network access mode limits access to the Apache Airflow UI to users *within your Amazon VPC* who have been granted access to the IAM policy for your environment.

When you create an environment with **Private network** access, you must package all of your dependencies in a Python wheel archive (`.whl`), then reference the `.whl` in your `requirements.txt`. For instructions on packaging and installing your dependencies using wheel, refer to Managing dependencies using Python wheel.

The following image depicts where to find the **Private network** option on the Amazon MWAA console.



- **Private routing**. An Amazon VPC without internet access limits network traffic within the VPC. This page assumes your Amazon VPC does not have internet access and requires VPC endpoints for each AWS service used by your environment, and VPC endpoints for Apache Airflow in the same AWS Region and Amazon VPC as your Amazon MWAA environment.

(Required) VPC endpoints

The following section displays the required VPC endpoints needed for an Amazon VPC without internet access. It lists the VPC endpoints for each AWS service used by Amazon MWAA, including the VPC endpoints needed for Apache Airflow.

```
com.amazonaws.us-east-1.s3
com.amazonaws.us-east-1.monitoring
com.amazonaws.us-east-1.logs
com.amazonaws.us-east-1.sqs
com.amazonaws.us-east-1.kms
```

Note

When using Transit Gateway or any other routing that does not go directly to the AWS API endpoints, we recommend you to add AWS PrivateLink to your Amazon MWAA private subnets for the following services:

- Amazon S3
- Amazon SQS
- CloudWatch Logs
- CloudWatch metrics
- AWS KMS (if applicable)

This ensures that your Amazon MWAA environment can securely and efficiently communicate with these services without routing traffic through the public internet, thereby improving security and performance.

Attaching the required VPC endpoints

This section describes the steps to attach the required VPC endpoints for an Amazon VPC with private routing.

VPC endpoints required for AWS services

The following section displays the steps to attach the VPC endpoints for the AWS services used by an environment to an existing Amazon VPC.

To attach VPC endpoints to your private subnets

1. Open the Endpoints page on the Amazon VPC console.
2. Select your AWS Region.
3. Create the endpoint for Amazon S3:
    a. Choose **Create Endpoint**.
    b. In the *Filter by attributes or search by keyword* text field, type: **`.s3`**, then press *Enter* on your keyboard.
    c. We recommend choosing the service endpoint listed for the **Gateway** type.

    For example, **`com.amazonaws.us-west-2.s3 amazon Gateway`**
    d. Choose your environment's Amazon VPC in **VPC**.
    e. Make sure that you choose the route table for your two private subnets.
    f. Choose **Full Access** in **Policy**.
    g. Choose **Create endpoint**.
4. Create the endpoint for CloudWatch Logs:
    a. Choose **Create Endpoint**.
    b. In the *Filter by attributes or search by keyword* text field, type: **`.logs`**, then press *Enter* on your keyboard.

c. Select the service endpoint.

d. Choose your environment's Amazon VPC in **VPC**.

e. Ensure that your two private subnets in different Availability Zones are selected, and that **Enable DNS name** is enabled.

f. Choose your environment's Amazon VPC security groups.

g. Choose **Full Access** in **Policy**.

h. Choose **Create endpoint**.

5. Create the endpoint for CloudWatch Monitoring:

a. Choose **Create Endpoint**.

b. In the *Filter by attributes or search by keyword* text field, type: **`.monitoring`**, then press *Enter* on your keyboard.

c. Select the service endpoint.

d. Choose your environment's Amazon VPC in **VPC**.

e. Ensure that your two private subnets in different Availability Zones are selected, and that **Enable DNS name** is enabled.

f. Choose your environment's Amazon VPC security groups.

g. Choose **Full Access** in **Policy**.

h. Choose **Create endpoint**.

6. Create the endpoint for Amazon SQS:

a. Choose **Create Endpoint**.

b. In the *Filter by attributes or search by keyword* text field, type: **`.sqs`**, then press *Enter* on your keyboard.

c. Select the service endpoint.

d. Choose your environment's Amazon VPC in **VPC**.

e. Ensure that your two private subnets in different Availability Zones are selected, and that **Enable DNS name** is enabled.

f. Choose your environment's Amazon VPC security groups.

g. Choose **Full Access** in **Policy**.

h. Choose **Create endpoint**.

7. Create the endpoint for AWS KMS:

a. Choose **Create Endpoint**.

b. In the *Filter by attributes or search by keyword* text field, type: **`.kms`**, then press *Enter* on your keyboard.

c. Select the service endpoint.

d. Choose your environment's Amazon VPC in **VPC**.

e. Ensure that your two private subnets in different Availability Zones are selected, and that **Enable DNS name** is enabled.

f. Choose your environment's Amazon VPC security groups.

g. Choose **Full Access** in **Policy**.

h. Choose **Create endpoint**.

VPC endpoints required for Apache Airflow

The following section displays the steps to attach the VPC endpoints for Apache Airflow to an existing Amazon VPC.

To attach VPC endpoints to your private subnets

1. Open the Endpoints page on the Amazon VPC console.
2. Select your AWS Region.
3. Create the endpoint for the Apache Airflow API:

   a. Choose **Create Endpoint**.

   b. In the *Filter by attributes or search by keyword* text field, type: **`.airflow.api`**, then press *Enter* on your keyboard.

   c. Select the service endpoint.

   d. Choose your environment's Amazon VPC in **VPC**.

   e. Ensure that your two private subnets in different Availability Zones are selected, and that **Enable DNS name** is enabled.

   f. Choose your environment's Amazon VPC security groups.

   g. Choose **Full Access** in **Policy**.

   h. Choose **Create endpoint**.

4. Create the first endpoint for the Apache Airflow environment:

a. Choose **Create Endpoint**.

b. In the *Filter by attributes or search by keyword* text field, type: **`.airflow.env`**, then press *Enter* on your keyboard.

c. Select the service endpoint.

d. Choose your environment's Amazon VPC in **VPC**.

e. Ensure that your two private subnets in different Availability Zones are selected, and that **Enable DNS name** is enabled.

f. Choose your environment's Amazon VPC security groups.

g. Choose **Full Access** in **Policy**.

h. Choose **Create endpoint**.

5. Create the second endpoint for Apache Airflow operations:

a. Choose **Create Endpoint**.

b. In the *Filter by attributes or search by keyword* text field, type: **`.airflow.ops`**, then press *Enter* on your keyboard.

c. Select the service endpoint.

d. Choose your environment's Amazon VPC in **VPC**.

e. Ensure that your two private subnets in different Availability Zones are selected, and that **Enable DNS name** is enabled.

f. Choose your environment's Amazon VPC security groups.

g. Choose **Full Access** in **Policy**.

h. Choose **Create endpoint**.

(Optional) Enable private IP addresses for your Amazon S3 VPC interface endpoint

Amazon S3 **Interface** endpoints don't support private DNS. The S3 endpoint requests still resolve to a *public* IP address. To resolve the S3 address to a *private* IP address, you need to add a private hosted zone in Route 53 for the S3 regional endpoint.

Using Route 53

This section describes the steps to enable private IP addresses for an S3 **Interface** endpoint using Route 53.

1. Create a Private Hosted Zone for your Amazon S3 VPC interface endpoint (such as, s3.eu-west-1.amazonaws.com) and associate it with your Amazon VPC.
2. Create an ALIAS A record for your Amazon S3 VPC interface endpoint (such as, s3.eu-west-1.amazonaws.com) that resolves to your VPC Interface Endpoint DNS name.
3. Create an ALIAS A wildcard record for your Amazon S3 interface endpoint (such as, *.s3.eu-west-1.amazonaws.com) that resolves to the VPC Interface Endpoint DNS name.

VPCs with custom DNS

If your Amazon VPC uses custom DNS routing, you need to make the changes in your DNS resolver (not Route 53, typically an EC2 instance running a DNS server) by creating a CNAME record. For example:

```
Name: s3.us-west-2.amazonaws.com
Type: CNAME
Value:  *.vpce-0f67d23e37648915c-e2q2e2j3.s3.us-west-2.vpce.amazonaws.com
```

Managing your own Amazon VPC endpoints on Amazon MWAA

Amazon MWAA uses Amazon VPC endpoints to integrate with various AWS services necessary to set up an Apache Airflow environment. Managing your own endpoints has two primary use cases:

1. It means you can create Apache Airflow environments in a shared Amazon VPC when you use an AWS Organizations to manage multiple AWS accounts and share resources.
2. It lets you use more restrictive access policies by narrowing down your permissions to the specific resources that use your endpoints.

If you choose to manage your own VPC endpoints, you're responsible for creating your own endpoints for the environment RDS for PostgreSQL database and for the environment webserver.

For more information about how Amazon MWAA deploys Apache Airflow in the cloud, refer to the Amazon MWAA architecture diagram.

Important

Amazon MWAA does not validate the IP address type (`AddressType`) selection for customer-managed endpoints, so make sure you correctly specify `AddressType` (valid options are IPv4 or IPv6).

Creating an environment in a shared Amazon VPC

If you use AWS Organizations to manage multiple AWS accounts that share resources, you can use customer-managed VPC endpoints with Amazon MWAA to share environment resources with another account in your organization.

When you configure shared VPC access, the account that owns the main Amazon VPC (*owner*) shares the two private subnets required by Amazon MWAA with other accounts (*participants*) that belong to the same organization. Participant accounts that share those subnets can view, create, modify, and delete environments in the shared Amazon VPC.

Assume you have an account, `Owner`, which acts as the `Root` account in the organization and owns the Amazon VPC resources, and a participant account, `Participant`, a member of the same organization. When `Participant` creates a new Amazon MWAA in Amazon VPC it shares with `Owner`, Amazon MWAA will first create the service VPC resources, then enter a `PENDING` state for up to 72 hours.

After the environment status changes from `CREATING` to `PENDING`, a principal acting on behalf of `Owner` creates the required endpoints. To do this, Amazon MWAA lists the database and webserver endpoint in the Amazon MWAA console. You can also call the `GetEnvironment` API action to get the service endpoints.

Note

If the Amazon VPC you use to share resources is a private Amazon VPC, you must still complete the steps described in the section called "Managing access to VPC endpoints". The topic covers setting up a different set of Amazon VPC endpoints related to other AWS services that AWS integrates with, such as Amazon ECR, Amazon ECS, and Amazon SQS. These services are essential in operating, and managing, your Apache Airflow environment in the cloud.

Prerequisites

Before you create an Amazon MWAA environment in a shared VPC, you need the following resources:

- An AWS account, `Owner` to be used as the account that owns the Amazon VPC.
- An AWS Organizations organization unit, `MyOrganization` created as a *root*.
- A second AWS account, `Participant`, under `MyOrganization` to serve the participant account that creates the new environment.

In addition, we recommend that you familiarize yourself with the responsibilities and permissions for owners and participants when sharing resources in Amazon VPC.

Create the Amazon VPC

First, create a new Amazon VPC that the owner and participant accounts will share:

1. Sign in to the console using `Owner`, then, open the CloudFormation console. Use the following template to create a stack. This stack provisions a number of networking resources including a Amazon VPC, and the subnets that the two accounts will share in this scenario.

```
AWSTemplateFormatVersion: "2010-09-09"
Description: This template deploys a VPC, with a pair of public and private subnets
 spread across two Availability Zones. It deploys an internet gateway, with a
 default route on the public subnets. It deploys a pair of NAT gateways (one in
 each AZ), and default routes for them in the private subnets.
Parameters:
    EnvironmentName:
      Description: An environment name that is prefixed to resource names
      Type: String
      Default: mwaa-
    VpcCIDR:
      Description: Please enter the IP range (CIDR notation) for this VPC
      Type: String
      Default: 10.192.0.0/16
    PublicSubnet1CIDR:
      Description: Please enter the IP range (CIDR notation) for the public subnet
 in the first Availability Zone
      Type: String
      Default: 10.192.10.0/24
    PublicSubnet2CIDR:
```

```
      Description: Please enter the IP range (CIDR notation) for the public subnet
 in the second Availability Zone
      Type: String
      Default: 10.192.11.0/24
    PrivateSubnet1CIDR:
      Description: Please enter the IP range (CIDR notation) for the private subnet
 in the first Availability Zone
      Type: String
      Default: 10.192.20.0/24
    PrivateSubnet2CIDR:
      Description: Please enter the IP range (CIDR notation) for the private subnet
 in the second Availability Zone
      Type: String
      Default: 10.192.21.0/24
    AvailabilityZone1:
      Description: >-
        First Availability Zone for the subnets. Any two different AZs work.
        Changing this on an existing stack replaces the subnets.
      Type: AWS::EC2::AvailabilityZone::Name
    AvailabilityZone2:
      Description: >-
        Second Availability Zone. Must differ from the first.
        Changing this on an existing stack replaces the subnets.
      Type: AWS::EC2::AvailabilityZone::Name
Rules:
    DifferentAZs:
      Assertions:
        - Assert: !Not [ !Equals [ !Ref AvailabilityZone1, !Ref
 AvailabilityZone2 ] ]
          AssertDescription: The two Availability Zones must be different.
Resources:
    VPC:
      Type: 'AWS::EC2::VPC'
      Properties:
        CidrBlock: !Ref VpcCIDR
        EnableDnsSupport: true
        EnableDnsHostnames: true
        Tags:
          - Key: Name
            Value: !Ref EnvironmentName

    InternetGateway:
      Type: 'AWS::EC2::InternetGateway'
      Properties:
```

```
      Tags:
        - Key: Name
          Value: !Ref EnvironmentName

  InternetGatewayAttachment:
    Type: 'AWS::EC2::VPCGatewayAttachment'
    Properties:
      InternetGatewayId: !Ref InternetGateway
      VpcId: !Ref VPC

  PublicSubnet1:
    Type: 'AWS::EC2::Subnet'
    Properties:
      VpcId: !Ref VPC
      AvailabilityZone: !Ref AvailabilityZone1
      CidrBlock: !Ref PublicSubnet1CIDR
      MapPublicIpOnLaunch: true
      Tags:
        - Key: Name
          Value: !Sub '${EnvironmentName} Public Subnet (AZ1)'

  PublicSubnet2:
    Type: 'AWS::EC2::Subnet'
    Properties:
      VpcId: !Ref VPC
      AvailabilityZone: !Ref AvailabilityZone2
      CidrBlock: !Ref PublicSubnet2CIDR
      MapPublicIpOnLaunch: true
      Tags:
        - Key: Name
          Value: !Sub '${EnvironmentName} Public Subnet (AZ2)'

  PrivateSubnet1:
    Type: 'AWS::EC2::Subnet'
    Properties:
      VpcId: !Ref VPC
      AvailabilityZone: !Ref AvailabilityZone1
      CidrBlock: !Ref PrivateSubnet1CIDR
      MapPublicIpOnLaunch: false
      Tags:
        - Key: Name
          Value: !Sub '${EnvironmentName} Private Subnet (AZ1)'

  PrivateSubnet2:
```

```
      Type: 'AWS::EC2::Subnet'
      Properties:
        VpcId: !Ref VPC
        AvailabilityZone: !Ref AvailabilityZone2
        CidrBlock: !Ref PrivateSubnet2CIDR
        MapPublicIpOnLaunch: false
        Tags:
          - Key: Name
            Value: !Sub '${EnvironmentName} Private Subnet (AZ2)'

  NatGateway1EIP:
    Type: 'AWS::EC2::EIP'
    DependsOn: InternetGatewayAttachment
    Properties:
      Domain: vpc

  NatGateway2EIP:
    Type: 'AWS::EC2::EIP'
    DependsOn: InternetGatewayAttachment
    Properties:
      Domain: vpc

  NatGateway1:
    Type: 'AWS::EC2::NatGateway'
    Properties:
      AllocationId: !GetAtt NatGateway1EIP.AllocationId
      SubnetId: !Ref PublicSubnet1

  NatGateway2:
    Type: 'AWS::EC2::NatGateway'
    Properties:
      AllocationId: !GetAtt NatGateway2EIP.AllocationId
      SubnetId: !Ref PublicSubnet2

  PublicRouteTable:
    Type: 'AWS::EC2::RouteTable'
    Properties:
      VpcId: !Ref VPC
      Tags:
        - Key: Name
          Value: !Sub '${EnvironmentName} Public Routes'

  DefaultPublicRoute:
    Type: 'AWS::EC2::Route'
```

```
    DependsOn: InternetGatewayAttachment
    Properties:
      RouteTableId: !Ref PublicRouteTable
      DestinationCidrBlock: 0.0.0.0/0
      GatewayId: !Ref InternetGateway

  PublicSubnet1RouteTableAssociation:
    Type: 'AWS::EC2::SubnetRouteTableAssociation'
    Properties:
      RouteTableId: !Ref PublicRouteTable
      SubnetId: !Ref PublicSubnet1

  PublicSubnet2RouteTableAssociation:
    Type: 'AWS::EC2::SubnetRouteTableAssociation'
    Properties:
      RouteTableId: !Ref PublicRouteTable
      SubnetId: !Ref PublicSubnet2

  PrivateRouteTable1:
    Type: 'AWS::EC2::RouteTable'
    Properties:
      VpcId: !Ref VPC
      Tags:
        - Key: Name
          Value: !Sub '${EnvironmentName} Private Routes (AZ1)'

  DefaultPrivateRoute1:
    Type: 'AWS::EC2::Route'
    Properties:
      RouteTableId: !Ref PrivateRouteTable1
      DestinationCidrBlock: 0.0.0.0/0
      NatGatewayId: !Ref NatGateway1

  PrivateSubnet1RouteTableAssociation:
    Type: 'AWS::EC2::SubnetRouteTableAssociation'
    Properties:
      RouteTableId: !Ref PrivateRouteTable1
      SubnetId: !Ref PrivateSubnet1

  PrivateRouteTable2:
    Type: 'AWS::EC2::RouteTable'
    Properties:
      VpcId: !Ref VPC
      Tags:
```

```
          - Key: Name
            Value: !Sub '${EnvironmentName} Private Routes (AZ2)'

    DefaultPrivateRoute2:
      Type: 'AWS::EC2::Route'
      Properties:
        RouteTableId: !Ref PrivateRouteTable2
        DestinationCidrBlock: 0.0.0.0/0
        NatGatewayId: !Ref NatGateway2

    PrivateSubnet2RouteTableAssociation:
      Type: 'AWS::EC2::SubnetRouteTableAssociation'
      Properties:
        RouteTableId: !Ref PrivateRouteTable2
        SubnetId: !Ref PrivateSubnet2

    SecurityGroup:
      Type: 'AWS::EC2::SecurityGroup'
      Properties:
        GroupName: mwaa-security-group
        GroupDescription: Security group with a self-referencing inbound rule.
        VpcId: !Ref VPC

    SecurityGroupIngress:
      Type: 'AWS::EC2::SecurityGroupIngress'
      Properties:
        GroupId: !Ref SecurityGroup
        IpProtocol: '-1'
        SourceSecurityGroupId: !Ref SecurityGroup
Outputs:
    VPC:
      Description: A reference to the created VPC
      Value: !Ref VPC
    PublicSubnets:
      Description: A list of the public subnets
      Value: !Join
        - ','
        - - !Ref PublicSubnet1
          - !Ref PublicSubnet2
    PrivateSubnets:
      Description: A list of the private subnets
      Value: !Join
        - ','
        - - !Ref PrivateSubnet1
```

```
        - !Ref PrivateSubnet2
    PublicSubnet1:
      Description: A reference to the public subnet in the 1st Availability Zone
      Value: !Ref PublicSubnet1
    PublicSubnet2:
      Description: A reference to the public subnet in the 2nd Availability Zone
      Value: !Ref PublicSubnet2
    PrivateSubnet1:
      Description: A reference to the private subnet in the 1st Availability Zone
      Value: !Ref PrivateSubnet1
    PrivateSubnet2:
      Description: A reference to the private subnet in the 2nd Availability Zone
      Value: !Ref PrivateSubnet2
    SecurityGroupIngress:
      Description: Security group with self-referencing inbound rule
      Value: !Ref SecurityGroupIngress
```

2. After the new Amazon VPC resources have been provisioned, navigate to the AWS Resource Access Manager console, then choose **Create resource share**.
3. Choose the subnets you created in the first step from the list of available subnets you can share with `Participant`.

Create the environment

Complete the following steps to create an Amazon MWAA environment with customer-managed Amazon VPC endpoints.

1. Sign in using `Participant`, and open the Amazon MWAA console. Complete **Step one: Specify details** to specify an Amazon S3 bucket, a DAG folder, and dependencies for your new environment. For more information, refer to getting started.
2. On the **Configure advanced settings** page, under **Networking**, choose the subnets from the shared Amazon VPC.
3. Under **Endpoint management** choose **CUSTOMER** from the dropdown list.
4. Keep the default for the remaining options on the page, then, choose **Create environment** on the **Review and create** page.

The environment begins in a `CREATING` state, then changes to `PENDING`. When the environment is `PENDING`, write down the **Database endpoint service name** and **webserver endpoint service name** (if you set up a private webserver) using the console.

When you create a new environment using the Amazon MWAA console. Amazon MWAA creates a new security group with the required inbound and outbound rules. Write down the security group ID.

In the next section, `Owner` will use the service endpoints and the security group ID to create new Amazon VPC endpoints in the shared Amazon VPC.

Create the Amazon VPC endpoints

Complete the following steps to create the required Amazon VPC endpoints for your environment.

1. Sign in to the AWS Management Console using `Owner`, the open https://console.aws.amazon.com/vpc/.
2. Choose **Security groups** from the left navigation panel, then create a new security group in the shared Amazon VPC using the following inbound, and outbound, rules:

	Type	Protocol	Source type	Source
Inbound	All traffic	All	All	Your environment security group
Outbound	All traffic	All	All	`0.0.0.0/0`

> **Warning**
>
> The `Owner` account must set up a security group in the `Owner` account to allow traffic from the new environment to the shared Amazon VPC. You can do this by creating a new security group in `Owner`, or editing an existing one.

3. Choose **Endpoints**, then choose **Create endpoint**. Under **Endpoint settings**, select **Endpoint services that use NLBs and GWLBs**. In the **Service name** field, enter the database endpoint service name you noted from the Amazon MWAA console, then choose **Verify service**. Choose the shared Amazon VPC, select the subnets you used for the environment, and select the security group you created in the previous step. Choose **Create endpoint**. If your environment uses a private webserver, repeat these steps using the webserver endpoint service name.

> **Note**
>
> If you create VPC endpoints using infrastructure as code, set the endpoint type to `Interface` explicitly. The default endpoint type in both AWS CloudFormation and Terraform is `Gateway`, which is not compatible with Amazon MWAA endpoint services.

If successful, the environment will change from `PENDING` back to `CREATING`, then finally to `AVAILABLE`. When it is `AVAILABLE`, you can sign in to the Apache Airflow console.

Shared Amazon VPC Troubleshooting

Use the following reference to resolve issues you encounter when creating environments in a shared Amazon VPC.

Environment in `CREATE_FAILED` after `PENDING` status

- Verify that `Owner` is sharing the subnets with `Participant` using AWS Resource Access Manager.
- Verify that the Amazon VPC endpoints for the database and webserver are created in the same subnets associated with the environment.
- Verify that the security group used with your endpoints allows traffic from the security groups used for the environment. The `Owner` account creates rules that reference the security group in `Participant` as *`123456789012`*/*`security-group-id`*:.

Type	Protocol	Source type	Source
All traffic	All	All	*`123456789012`* /*`sg-0909e8e81919`*

For more information, refer to Responsibilities and permissions for owners and participants

Environment stuck in `PENDING` status

Verify each VPC endpoint status to ensure it is `Available`. If you configure an environment with a private webserver, you must also create an endpoint for the webserver. If the

environment is stuck in PENDING, this might indicate that the private webserver endpoint is missing.

Received The Vpc Endpoint Service '*vpce-service-name*' does not exist error

If you refer to the following error, verify that the account creating the endpoints in the Owner account that owns the shared VPC:

```
ClientError: An error occurred (InvalidServiceName) when calling the
 CreateVpcEndpoint operation:

The Vpc Endpoint Service 'vpce-service-name' does not exist
```

Tutorials for Amazon Managed Workflows for Apache Airflow

This guide includes step-by-step tutorials to using and configuring an Amazon Managed Workflows for Apache Airflow environment.

Topics

- Tutorial: Configuring private network access using an AWS Client VPN
- Tutorial: Configuring private network access using a Linux Bastion Host
- Tutorial: Restricting an Amazon MWAA user's access to a subset of DAGs
- Tutorial: Automate managing your own environment endpoints on Amazon MWAA

Tutorial: Configuring private network access using an AWS Client VPN

This tutorial walks you through the steps to create a VPN tunnel from your computer to the Apache Airflow webserver for your Amazon Managed Workflows for Apache Airflow environment. To connect to the internet through a VPN tunnel, you'll first need to create a AWS Client VPN endpoint. Once set up, a Client VPN endpoint acts as a VPN server allowing a secure connection from your computer to the resources in your VPC. You'll then connect to the Client VPN from your computer using the AWS Client VPN for Desktop.

> **Note**
>
> This tutorial applies to the **Private network** access mode. If you chose **Both public and private network access**, you do not need a VPN to access the Apache Airflow UI since it is accessible over the internet.

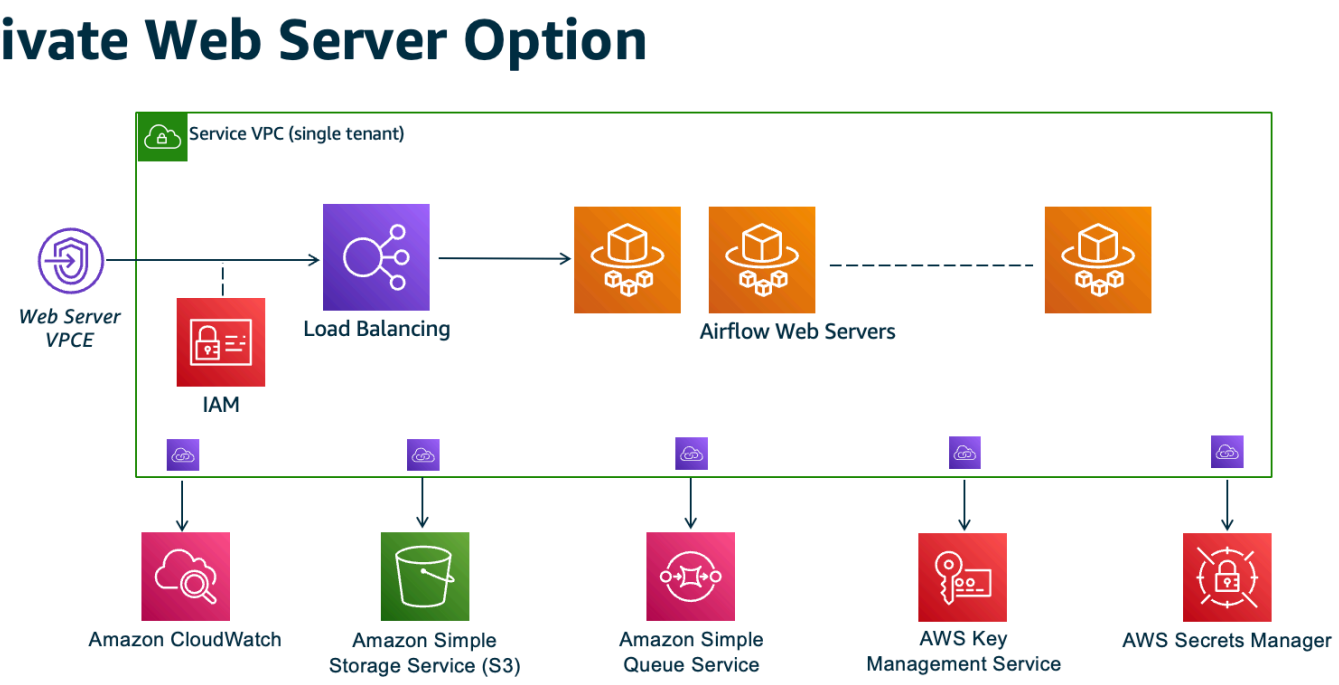
Sections

- Private network
- Use cases
- Before you begin

- Objectives
- (Optional) Step one: Identify your VPC, CIDR rules, and VPC security
- Step two: Create the server and client certificates
- Step three: Save the CloudFormation template locally
- Step four: Create the Client VPN CloudFormation stack
- Step five: Associate subnets to your Client VPN
- Step six: Add an authorization ingress rule to your Client VPN
- Step seven: Download the Client VPN endpoint configuration file
- Step eight: Connect to the AWS Client VPN
- What's next?

Private network

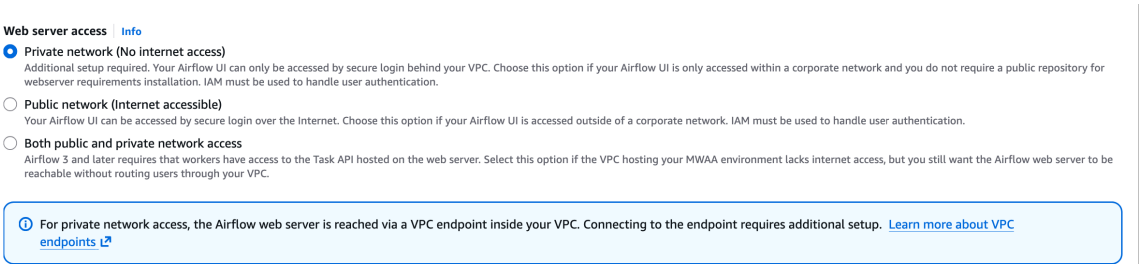
This tutorial assumes you've chosen the **Private network** access mode for your Apache Airflow webserver.



The private network access mode limits access to the Apache Airflow UI to users *within your Amazon VPC* who have been granted access to the IAM policy for your environment.

When you create an environment with **Private network** access, you must package all of your dependencies in a Python wheel archive (`.whl`), then reference the `.whl` in your `requirements.txt`. For instructions on packaging and installing your dependencies using wheel, refer to Managing dependencies using Python wheel.

The following image depicts where to find the **Private network** option on the Amazon MWAA console.



Use cases

You can use this tutorial before or after you've created an Amazon MWAA environment. You must use the same Amazon VPC, VPC security groups, and private subnets as your environment. If you use this tutorial after you've created an Amazon MWAA environment, once you've completed the steps, you can return to the Amazon MWAA console and change your Apache Airflow webserver access mode to **Private network**.

Before you begin

1. Check for user permissions. Be sure that your account in AWS Identity and Access Management (IAM) has sufficient permissions to create and manage VPC resources.
2. Use your Amazon MWAA VPC. This tutorial assumes that you are associating the Client VPN to an existing VPC. The Amazon VPC must be in the same AWS Region as an Amazon MWAA environment and have two private subnets. If you haven't created an Amazon VPC, use the CloudFormation template in Option three: Creating an Amazon VPC network *without* internet access.

Objectives

In this tutorial, you'll do the following:

1. Create a AWS Client VPN endpoint using a CloudFormation template for an existing Amazon VPC.

2. Generate server and client certificates and keys, and then upload the server certificate and key to AWS Certificate Manager in the same AWS Region as an Amazon MWAA environment.
3. Download and modify a Client VPN endpoint configuration file for your Client VPN, and use the file to create a VPN profile to connect using the Client VPN for Desktop.

(Optional) Step one: Identify your VPC, CIDR rules, and VPC security

The following section describes how to find IDs for your Amazon VPC, VPC security group, and a way to identify the CIDR rules you'll need to create your Client VPN in subsequent steps.

Identify your CIDR rules

The following section explains how to identify the CIDR rules, which you'll need to create your Client VPN.

To identify the CIDR for your Client VPN

1. Open the Your Amazon VPCs page on the Amazon VPC console.
2. Use the region selector in the navigation bar to choose the same AWS Region as an Amazon MWAA environment.
3. Choose your Amazon VPC.
4. Assuming the CIDRs for your private subnets are:

   - Private Subnet 1: 10.192.10.0/24
   - Private Subnet 2: 10.192.11.0/24

   If the CIDR for your Amazon VPC is 10.192.0.0/16, then the **Client IPv4 CIDR** you'd specify for your Client VPN would be 172.16.0.0/22.
5. Save this CIDR value, and the value of your VPC ID for subsequent steps.

Identify your VPC and security groups

The following section explains how to find the ID of your Amazon VPC and security groups, which you'll need to create your Client VPN.

> **Note**
>
> You might be using more than one security group. You'll need to specify all of your VPC's security groups in subsequent steps.

To identify the security groups

1. Open the Security Groups page on the Amazon VPC console.
2. Use the region selector in the navigation bar to choose the AWS Region.
3. Search for the Amazon VPC in **VPC ID**, and identify the security groups associated with the VPC.
4. Save the ID of your security groups and VPC for subsequent steps.

Step two: Create the server and client certificates

A Client VPN endpoint supports 1024-bit and 2048-bit RSA key sizes only. The following section explains how to use OpenVPN easy-rsa to generate the server and client certificates and keys, and then upload the certificates to ACM using the AWS Command Line Interface (AWS CLI).

To create the client certificates

1. Follow these quick steps to create and upload the certificates to ACM through the AWS CLI in Client authentication and authorization: Mutual authentication.
2. In these steps, you **must** specify the same AWS Region as an Amazon MWAA environment in the AWS CLI command when uploading your server and client certificates. Here's some examples of how to specify the region in these commands:

   a. **Example region for server certificate**

   ```
   aws acm import-certificate --certificate fileb://server.crt --private-key
    fileb://server.key --certificate-chain fileb://ca.crt --region us-west-2
   ```

   b. **Example region for client certificate**

   ```
   aws acm import-certificate --certificate fileb://client1.domain.tld.crt
    --private-key fileb://client1.domain.tld.key --certificate-chain fileb://
   ca.crt --region us-west-2
   ```

c. After these steps, save the value returned in the AWS CLI response for the server certificate and client certificate ARNs. You'll be specifying these ARNs in your CloudFormation template to create the Client VPN.

3. In these steps, a client certificate and a private key are saved to your computer. Here's an example of where to find these credentials:

   a. **Example on macOS**

   On macOS the contents are saved at `/Users/your-user/custom_folder`. If you list all (`ls -a`) contents of this directory, you get something similar to the following:

   ```
   .
   ..
   ca.crt
   client1.domain.tld.crt
   client1.domain.tld.key
   server.crt
   server.key
   ```

   b. After these steps, save the contents or note the location of the client certificate in `client1.domain.tld.crt`, and the private key in `client1.domain.tld.key`. You'll be adding these values to the configuration file for your Client VPN.

Step three: Save the CloudFormation template locally

The following section contains the CloudFormation template to create the Client VPN. You must specify the same Amazon VPC, VPC security groups, and private subnets as your Amazon MWAA environment.

- Copy the contents of the following template and save locally as `mwaa_vpn_client.yaml`. You can also download the template.

  Substitute the following values:

  - **`YOUR_CLIENT_ROOT_CERTIFICATE_ARN`** – The ARN for your **client1.domain.tld** certificate in `ClientRootCertificateChainArn`.
  - **`YOUR_SERVER_CERTIFICATE_ARN`** – The ARN for your **server** certificate in `ServerCertificateArn`.

- The Client IPv4 CIDR rule in `ClientCidrBlock`. A CIDR rule of `172.16.0.0/22` is provided. This range must not overlap your Amazon VPC CIDR.
- Your Amazon VPC ID in `VpcId`. A VPC of `vpc-010101010101` is provided.
- Your VPC security group IDs in `SecurityGroupIds`. A security group of `sg-0101010101` is provided.

```
AWSTemplateFormatVersion: 2010-09-09
Description: This template deploys a VPN Client Endpoint.
Resources:
  ClientVpnEndpoint:
    Type: 'AWS::EC2::ClientVpnEndpoint'
    Properties:
      AuthenticationOptions:
        - Type: "certificate-authentication"
          MutualAuthentication:
            ClientRootCertificateChainArn: "YOUR_CLIENT_ROOT_CERTIFICATE_ARN"
      ClientCidrBlock: 172.16.0.0/22
      ClientConnectOptions:
        Enabled: false
      ConnectionLogOptions:
        Enabled: false
      Description: "MWAA Client VPN"
      DnsServers: []
      SecurityGroupIds:
        - sg-0101010101
      SelfServicePortal: ''
      ServerCertificateArn: "YOUR_SERVER_CERTIFICATE_ARN"
      SplitTunnel: true
      TagSpecifications:
        - ResourceType: "client-vpn-endpoint"
          Tags:
          - Key: Name
            Value: MWAA-Client-VPN
      TransportProtocol: udp
      VpcId: vpc-010101010101
      VpnPort: 443
```

Note

If you're using more than one security group for your environment, you can specify multiple security groups in the following format:

```
SecurityGroupIds:
 - sg-0112233445566778b
 - sg-0223344556677889f
```

Step four: Create the Client VPN CloudFormation stack

To create the AWS Client VPN

1. Open the AWS CloudFormation console.
2. Choose **Template is ready**, **Upload a template file**.
3. Choose **Choose file**, and select your `mwaa_vpn_client.yaml` file.
4. Choose **Next**, **Next**.
5. Select the acknowledgement, and then choose **Create stack**.

Step five: Associate subnets to your Client VPN

To associate private subnets to the AWS Client VPN

1. Open the Amazon VPC console.
2. Choose the **Client VPN Endpoints** page.
3. Select your Client VPN, and then choose the **Associations** tab, **Associate**.
4. Choose the following in the dropdown list:
   - Your Amazon VPC in **VPC**.
   - One of your private subnets in **Choose a subnet to associate**.
5. Choose **Associate**.

> **Note**
>
> It takes several minutes for the VPC and subnet to be associated to the Client VPN.

Step six: Add an authorization ingress rule to your Client VPN

You need to add an authorization ingress rule using the CIDR rule for your VPC to your Client VPN. If you want to authorize specific users or groups from your Active Directory Group or SAML-based Identity Provider (IdP), refer to the Authorization rules in the *Client VPN guide*.

To add the CIDR to the AWS Client VPN

1. Open the Amazon VPC console.
2. Choose the **Client VPN Endpoints** page.
3. Select your Client VPN, and then choose the **Authorization** tab, **Authorize Ingress**.
4. Specify the following:
   - Your Amazon VPC's CIDR rule in **Destination network to enable**. For example:

     ```
     10.192.0.0/16
     ```

   - Choose **Allow access to all users** in **Grant access to**.
   - Enter a descriptive name in **Description**.
5. Choose **Add Authorization rule**.

> **Note**
>
> Depending on the networking components for your Amazon VPC, you might also need to add this authorization ingress rule to your network access control list (NACL).

Step seven: Download the Client VPN endpoint configuration file

To download the configuration file

1. Follow these quick steps to download the Client VPN configuration file at Download the Client VPN endpoint configuration file.
2. In these steps, you're asked to prepend a string to your Client VPN endpoint DNS name. Here's an example:

   - **Example endpoint DNS name**

     If your Client VPN endpoint DNS name is:

     ```
     remote cvpn-endpoint-0909091212aaee1.prod.clientvpn.us-west-1.amazonaws.com 443
     ```

     You can add a string to identify your Client VPN endpoint like this:

     ```
     remote mwaavpn.cvpn-endpoint-0909091212aaee1.prod.clientvpn.us-
     west-1.amazonaws.com 443
     ```

3. In these steps, you're asked to add the contents of the client certificate between a new set of `<cert></cert>` tags and the contents of the private key between a new set of `<key></key>` tags. Here's an example:

   a. Open a command prompt and change directories to the location of your client certificate and private key.

   b. **Example macOS client1.domain.tld.crt**

      To show the contents of the `client1.domain.tld.crt` file on macOS, you can use `cat client1.domain.tld.crt`.

      Copy the value from terminal and paste in `downloaded-client-config.ovpn` like this:

      ```
      ZZZ1111dddaBBB
      -----END CERTIFICATE-----
      </ca>
      <cert>
      -----BEGIN CERTIFICATE-----
      YOUR client1.domain.tld.crt
      -----END CERTIFICATE-----
      ```

```
</cert>
```

c. **Example macOS client1.domain.tld.key**

To show the contents of the `client1.domain.tld.key`, you can use `cat client1.domain.tld.key`.

Copy the value from terminal and paste in `downloaded-client-config.ovpn` like this:

```
ZZZ1111dddaBBB
-----END CERTIFICATE-----
</ca>
<cert>
-----BEGIN CERTIFICATE-----
YOUR client1.domain.tld.crt
-----END CERTIFICATE-----
</cert>
<key>
-----BEGIN CERTIFICATE-----
YOUR client1.domain.tld.key
-----END CERTIFICATE-----
</key>
```

Step eight: Connect to the AWS Client VPN

The client for AWS Client VPN is provided free of charge. You can connect your computer directly to AWS Client VPN for an end-to-end VPN experience.

To connect to the Client VPN

1. Download and install the AWS Client VPN for Desktop.
2. Open the AWS Client VPN.
3. Choose **File**, **Managed profiles** in the VPN client menu.
4. Choose **Add profile**, and then choose the `downloaded-client-config.ovpn`.
5. Enter a descriptive name in **Display Name**.
6. Choose **Add profile**, **Done**.
7. Choose **Connect**.

After you connect to the Client VPN, you'll need to disconnect from other VPNs to access any of the resources in your Amazon VPC.

> **Note**
>
> You might need to quit the client, and start again before you're able to get connected.

What's next?

- Learn how to create an Amazon MWAA environment in Get started with Amazon Managed Workflows for Apache Airflow. You must create an environment in the same AWS Region as the Client VPN, and using the same VPC, private subnets, and security group as the Client VPN.

Tutorial: Configuring private network access using a Linux Bastion Host

This tutorial walks you through the steps to create an SSH tunnel from your computer to the Apache Airflow webserver for your Amazon Managed Workflows for Apache Airflow environment. It assumes you've already created an Amazon MWAA environment. Once set up, a Linux Bastion Host acts as a jump server allowing a secure connection from your computer to the resources in your VPC. You then use a SOCKS proxy management add-on to control the proxy settings in your browser to access your Apache Airflow UI.

> **Note**
>
> This tutorial applies to the **Private network** access mode. If you chose **Both public and private network access**, you do not need a bastion host to access the Apache Airflow UI since it is accessible over the internet.

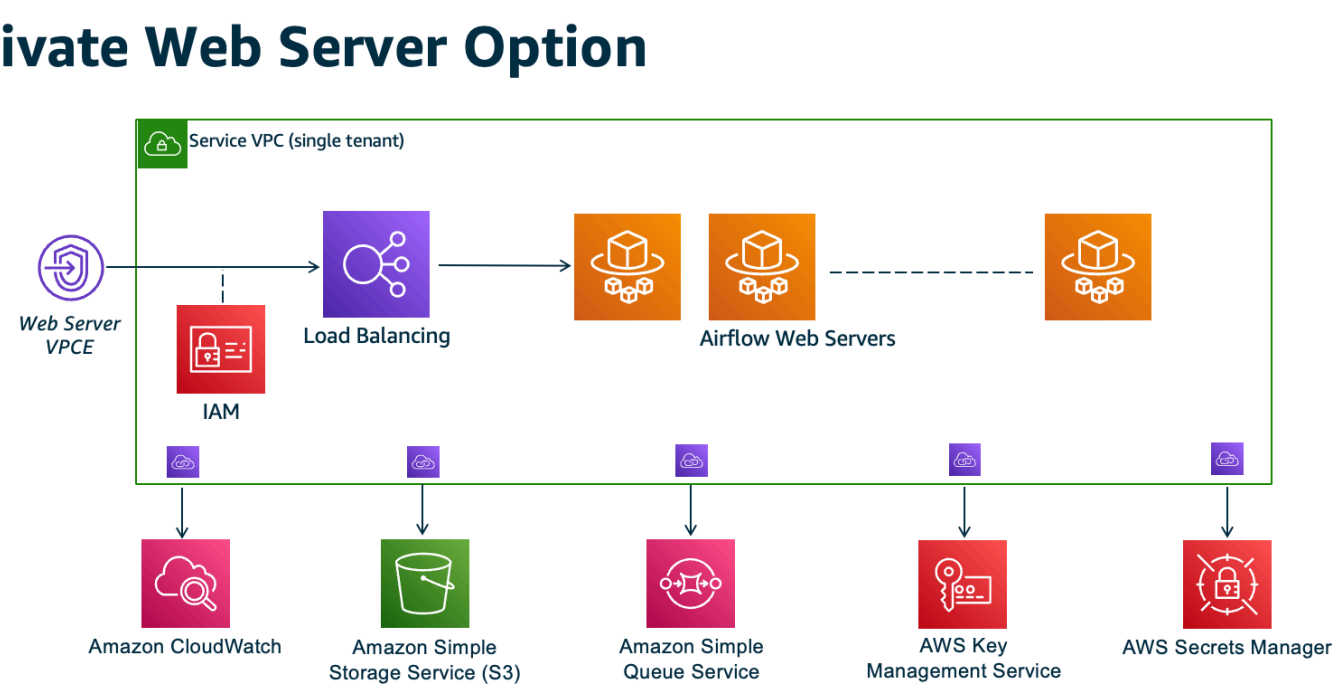
Sections

- Private network
- Use cases
- Before you begin

- Objectives
- Step 1: Create the bastion instance
- Step 2: Create the SSH tunnel
- Step three: Configure the bastion security group as an inbound rule
- Step four: Copy the Apache Airflow URL
- Step five: Configure proxy settings
- Step six: Open the Apache Airflow UI
- What's next?

Private network

This tutorial assumes you've chosen the **Private network** access mode for your Apache Airflow webserver.

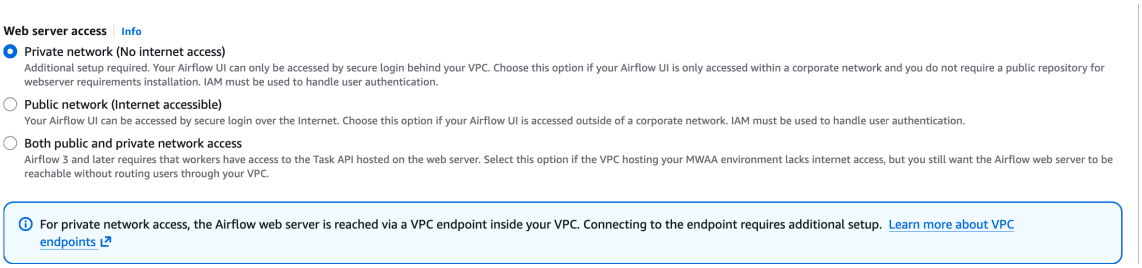


The private network access mode limits access to the Apache Airflow UI to users *within your Amazon VPC* who have been granted access to the IAM policy for your environment.

When you create an environment with **Private network** access, you must package all of your dependencies in a Python wheel archive (`.whl`), then reference the `.whl` in your

`requirements.txt`. For instructions on packaging and installing your dependencies using wheel, refer to Managing dependencies using Python wheel.

The following image depicts where to find the **Private network** option on the Amazon MWAA console.



Use cases

You can use this tutorial after you've created an Amazon MWAA environment. You must use the same Amazon VPC, VPC security groups, and public subnets as your environment.

Before you begin

1. Check for user permissions. Be sure that your account in AWS Identity and Access Management (IAM) has sufficient permissions to create and manage VPC resources.
2. Use your Amazon MWAA VPC. This tutorial assumes that you are associating the bastion host to an existing VPC. The Amazon VPC must be in the same region as your Amazon MWAA environment and have two private subnets, as defined in Create the VPC network.
3. Create an SSH key. You need to create an Amazon EC2 SSH key (**.pem**) in the same Region as your Amazon MWAA environment to connect to the virtual servers. If you don't have an SSH key, refer to Create or import a key pair in the *Amazon EC2 User Guide*.

Objectives

In this tutorial, you'll do the following:

1. Create a Linux Bastion Host instance using a CloudFormation template for an existing VPC.
2. Authorize inbound traffic to the bastion instance's security group using an ingress rule on port 22.
3. Authorize inbound traffic from an Amazon MWAA environment's security group to the bastion instance's security group.

4. Create an SSH tunnel to the bastion instance.
5. Install and configure the FoxyProxy add-on for the Firefox browser to access the Apache Airflow UI.

Step 1: Create the bastion instance

The following section describes the steps to create the Linux bastion instance.

Open the Amazon Elastic Compute Cloud console at https://console.aws.amazon.com/ec2/, and create an Amazon EC2 instance in one of the public subnets in the VPC that is used for your Amazon MWAA environment.

> **Note**
>
> We recommend that you allow TCP forwarding on the Amazon EC2 instance. TCP forwarding means SSH TCP forwarding — allowing `ssh -L/-D` tunnels through the host. The `AllowTcpForwarding` directive in the SSH daemon config (`/etc/ssh/sshd_config`) controls this setting. It defaults to `yes` on most Amazon Linux and Ubuntu AMIs, but is often disabled on hardened bastion AMIs.

Step 2: Create the SSH tunnel

The following steps describe how to create the SSH tunnel to your Linux bastion. An SSH tunnel receives the request from your local IP address to the Linux bastion, which is why TCP forwarding for the Linux bastion was set to `true` in previous steps.

macOS/Linux

To create a tunnel using the command line

1. Open the Instances page on the Amazon EC2 console.
2. Choose an instance.
3. Copy the address in **Public IPv4 DNS**. For example, `ec2-4-82-142-1.compute-1.amazonaws.com`.
4. In your command prompt, navigate to the directory where your SSH key is stored.

5. Run the following command to connect to the bastion instance using ssh. Substitute the sample value with your SSH key name in `mykeypair.pem`.

```
ssh -i mykeypair.pem -N -D 8157 ec2-user@YOUR_PUBLIC_IPV4_DNS
```

Windows (PuTTY)

To create a tunnel using PuTTY

1. Open the Instances page on the Amazon EC2 console.
2. Choose an instance.
3. Copy the address in **Public IPv4 DNS**. For example, `ec2-4-82-142-1.compute-1.amazonaws.com`.
4. Open PuTTY, select **Session**.
5. Enter the host name in **Host Name** as ec2-user@*YOUR_PUBLIC_IPV4_DNS* and the **port** as 22.
6. Expand the **SSH** tab, select **Auth**. In **Private Key file for authentication**, choose your local "ppk" file.
7. Under SSH, choose the **Tunnels** tab, and then select the *Dynamic* and *Auto* options.
8. In **Source Port**, add the 8157 port (or any other unused port), and then leave the **Destination** port blank. Choose **Add**.
9. Choose the **Session** tab and enter a session name. For example `SSH Tunnel`.
10. Choose **Save**, **Open**.

> **Note**
>
> You might need to enter a pass phrase for your public key.

> **Note**
>
> If you receive a `Permission denied (publickey)` error, we recommend using the AWSSupport-TroubleshootSSH tool, and choose **Run this Automation (console)** to troubleshoot your SSH setup.

Step three: Configure the bastion security group as an inbound rule

Access to the servers and regular internet access from the servers is allowed with a special maintenance security group attached to those servers. The following steps describe how to configure the bastion security group as an inbound source of traffic to an environment's VPC security group.

1. Open the Environments page on the Amazon MWAA console.
2. Choose an environment.
3. On the **Networking** pane, choose **VPC security group**.
4. Choose **Edit inbound rules**.
5. Choose **Add rule**.
6. Choose your VPC security group ID in the **Source** dropdown list.
7. Leave the remaining options blank, or set to their default values.
8. Choose **Save rules**.

Step four: Copy the Apache Airflow URL

The following steps describe how to open the Amazon MWAA console and copy the URL to the Apache Airflow UI.

1. Open the Environments page on the Amazon MWAA console.
2. Choose an environment.
3. Copy the URL in **Airflow UI** for subsequent steps.

Step five: Configure proxy settings

If you use an SSH tunnel with dynamic port forwarding, you must use a SOCKS proxy management add-on to control the proxy settings in your browser. For example, you can use the `--proxy-server` feature of Chromium to kick off a browser session, or use the FoxyProxy extension in the Mozilla FireFox browser.

Option one: Setup an SSH Tunnel using local port forwarding

If you do not wish to use a SOCKS proxy, you can set up an SSH tunnel using local port forwarding. The following example command accesses the Apache Airflow webserver by forwarding traffic on local port 8157.

1. Open a new command prompt window.
2. Enter the following command to open an SSH tunnel.

   ```
   ssh -i mykeypair.pem -N -L 8157:YOUR_VPC_ENDPOINT_ID-vpce.us-
   east-1.airflow.amazonaws.com:443 ec2-user@YOUR_PUBLIC_IPV4_DNS
   ```

   `-L` signifies the use of local port forwarding which you can use to specify a local port used to forward data to the identified remote port on the node's local webserver.
3. Enter `http://localhost:8157/` in your browser.

   Note

   You might need to use `https://localhost:8157/`.

Option two: Proxies using the command line

You can use most web browsers to configure proxies using a command line or configuration parameter. For example, with Chromium you can start the browser with the following command:

```
chromium --proxy-server="socks5://localhost:8157"
```

This starts a browser session which uses the ssh tunnel you created in previous steps to proxy its requests. You can open your Private Amazon MWAA environment URL (with *https://*) as follows:

```
https://YOUR_VPC_ENDPOINT_ID-vpce.us-east-1.airflow.amazonaws.com/home.
```

Option three: Proxies using FoxyProxy for Mozilla Firefox

The following example demonstrates a FoxyProxy Standard (version 7.5.1) configuration for Mozilla Firefox. FoxyProxy provides a set of proxy management tools. It lets you use a proxy server for URLs that match patterns corresponding to domains used by the Apache Airflow UI.

1. In Firefox, open the FoxyProxy Standard extension page.
2. Choose **Add to Firefox**.
3. Choose **Add**.
4. Choose the FoxyProxy icon in your browser's toolbar, choose **Options**.
5. Copy the following code and save locally as `mwaa-proxy.json`. Substitute the sample value in *YOUR_HOST_NAME* with your **Apache Airflow URL**.

```
{
  "e0b7kh1606694837384": {
    "type": 3,
    "color": "#66cc66",
    "title": "airflow",
    "active": true,
    "address": "localhost",
    "port": 8157,
    "proxyDNS": false,
    "username": "",
    "password": "",
    "whitePatterns": [
      {
        "title": "airflow-ui",
        "pattern": "YOUR_HOST_NAME",
        "type": 1,
        "protocols": 1,
        "active": true
      }
    ],
    "blackPatterns": [],
    "pacURL": "",
    "index": -1
  },
  "k20d21508277536715": {
    "active": true,
    "title": "Default",
    "notes": "These are the settings that are used when no patterns match a URL.",
    "color": "#0055E5",
    "type": 5,
    "whitePatterns": [
      {
        "title": "all URLs",
        "active": true,
```

```
        "pattern": "*",
        "type": 1,
        "protocols": 1
      }
    ],
    "blackPatterns": [],
      "index": 9007199254740991
  },
  "logging": {
    "active": true,
    "maxSize": 500
  },
  "mode": "patterns",
  "browserVersion": "82.0.3",
  "foxyProxyVersion": "7.5.1",
  "foxyProxyEdition": "standard"
}
```

6. On the **Import Settings from FoxyProxy 6.0+** pane, choose **Import Settings** and select the `mwaa-proxy.json` file.
7. Choose **OK**.

Step six: Open the Apache Airflow UI

The following steps describe how to open your Apache Airflow UI.

1. Open the Environments page on the Amazon MWAA console.
2. Choose **Open Airflow UI**.

What's next?

- Learn how to run Airflow CLI commands on an SSH tunnel to a bastion host in Apache Airflow CLI command reference.
- Learn how to upload DAG code to your Amazon S3 bucket in Adding or updating DAGs.

Tutorial: Restricting an Amazon MWAA user's access to a subset of DAGs

Amazon MWAA manages access to your environment by mapping your IAM principals to one or more of Apache Airflow's default roles. Use the following tutorial to restrict individual Amazon MWAA users to only access and interact with a specific DAG or a set of DAGs.

> **Note**
>
> The steps in this tutorial can be completed using federated access, as long as the IAM roles can be assumed.

Topics

- Prerequisites
- Step one: Provide Amazon MWAA webserver access to your IAM principal with the default Public Apache Airflow role.
- Step two: Create a new Apache Airflow custom role
- Step three: Assign the role you created to your Amazon MWAA user
- Next steps
- Related resources

Prerequisites

To complete the steps in this tutorial, you'll need the following:

- An Amazon MWAA environment with multiple DAGs
- An IAM principal, `Admin` with AdministratorAccess permissions, and an IAM user, `MWAAUser`, as the principal for which you can limit DAG access. For more information about admin roles, refer to Administrator job function in the *IAM User Guide*

> **Note**
>
> Don't attach permission policies directly to your IAM users. We recommend setting up IAM roles that users can assume to gain temporary access to your Amazon MWAA resources.

- AWS Command Line Interface version 2 installed.

Step one: Provide Amazon MWAA webserver access to your IAM principal with the default `Public` Apache Airflow role.

To grant permission using the AWS Management Console

1. Sign in to your AWS account with an `Admin` role and open the IAM console.
2. In the left navigation pane, choose **Users**, then choose your Amazon MWAA IAM user from the users table.
3. On the user details page, under **Summary**, choose the **Permissions** tab, then choose **Permissions policies** to expand the card and choose **Add permissions**.
4. In the **Grant permissions** section, choose **Attach existing policies directly**, then choose **Create policy** to create and attach your own custom permissions policy.
5. On the **Create policy** page, choose **JSON**, then copy and paste the following JSON permissions policy in the policy editor. Tha policy grants webserver access to the user with the default `Public` Apache Airflow role.

   JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "airflow:CreateWebLoginToken",
            "Resource": [
            "arn:aws:airflow:us-
east-1:111122223333:role/YOUR_ENVIRONMENT_NAME/Public"
            ]
```

```
        }
    ]
}
```

Step two: Create a new Apache Airflow custom role

To create a new role using the Apache Airflow UI

1. Using your administrator IAM role, open the Amazon MWAA console and launch your environment's Apache Airflow UI.
2. From the navigation pane at the top, hover on **Security** to open the dropdown list, then choose **List Roles** to access the default Apache Airflow roles.
3. From the roles list, select **User**, then at the beginning of the page choose **Actions** to open the dropdown. Choose **Copy Role**, and confirm **Ok**

   > **Note**
   >
   > Copy the **Ops** or **Viewer** roles to grant more or less access, respectively.

4. Locate the new role you created in the table and choose **Edit record**.
5. On the **Edit Role** page, do the following:

   - For **Name**, enter a new name for the role in the text field. For example, **`Restricted`**.
   - For the list of **Permissions**, remove `can read on DAGs` and `can edit on DAGs`, then add read and write permissions for the set of DAGs you want to provide access to. For example, for a DAG, `example_dag.py`, add **`can read on DAG:`*`example_dag`*** and **`can edit on DAG:`*`example_dag`***.

   Choose **Save**. Now you have a new role that limits access to a subset of DAGs available in your Amazon MWAA environment. You can assign this role to any existing Apache Airflow users.

Step three: Assign the role you created to your Amazon MWAA user

To assign the new role

1. Using access credentials for `MWAAUser`, run the following CLI command to retrieve your environment's webserver URL.

   ```
   aws mwaa get-environment --name YOUR_ENVIRONMENT_NAME | jq
    '.Environment.WebserverUrl'
   ```

   If successful, you'll refer to the following output:

   ```
   "ab1b2345-678a-90a1-a2aa-34a567a8a901.c13.us-west-2.airflow.amazonaws.com"
   ```

2. With `MWAAUser` signed in to the AWS Management Console, open a new browser window and access the following URl. Replace `Webserver-URL` with your information.

   ```
   https://<Webserver-URL>/home
   ```

   If successful, you'll get a `Forbidden` error page because `MWAAUser` has not been granted permission to access the Apache Airflow UI yet.

3. With `Admin` signed in to the AWS Management Console, open the Amazon MWAA console again and launch your environment's Apache Airflow UI.
4. From the UI dashboard, expand the **Security** dropdown, and this time choose **List Users**.
5. In the users table, find the new Apache Airflow user and choose **Edit record**. The user's first name will match your IAM user name in the following pattern: `user/mwaa-user`.
6. On the **Edit User** page, in the **Role** section, add the new custom role you created, then choose **Save**.

   > **Note**
   >
   > The **Last Name** field is required, but a space satisfies the requirement.

   The IAM `Public` principal grants the `MWAAUser` permission to access the Apache Airflow UI, while the new role provides the additional permissions needed to get their DAGs.

Important

Any of the 5 default roles (such as `Admin`) not authorized by IAM which are added using the Apache Airflow UI will be removed on next user log-in.

Next steps

- To learn more about managing access to your Amazon MWAA environment, and to get sample JSON IAM policies you can use for your environment users, refer to the section called "Accessing an Amazon MWAA environment"

Related resources

- Access Control (Apache Airflow Documentation) – Learn more about the default Apache Airflow roles on the Apache Airflow documentation website.

Tutorial: Automate managing your own environment endpoints on Amazon MWAA

If you use AWS Organizations to manage multiple AWS accounts that share resources, Amazon MWAA lets you create and manage your own Amazon VPC endpoints. This means you can use stricter security policies that allow access only to the resources required by your environment.

When you create an environment in a shared Amazon VPC, the account that owns the main Amazon VPC (*owner*) shares the two private subnets required by Amazon MWAA with other accounts (*participants*) that belong to the same organization. Participant accounts that share those subnets can then view, create, modify, and delete environments in the shared VPC.

When you create an environment in a shared, or otherwise policy-restricted, Amazon VPC, Amazon MWAA will first create the service VPC resources, then enter a `PENDING` state for up to 72 hours.

When the environment status changes from `CREATING` to `PENDING`, Amazon MWAA sends an Amazon EventBridge notification of the change in state. This lets the owner account create the required endpoints on behalf of participants based on endpoint service information from the Amazon MWAA console or API, or programmatically In the following, we create new Amazon VPC

endpoints using an Lambda function and an EventBridge rule that listens to Amazon MWAA state change notifications.

Here, we create the new endpoints in the same Amazon VPC as the environment. To set up a shared Amazon VPC, create the EventBridge rule and Lambda function would in the owner account, and the Amazon MWAA environment in the participant account.

Topics

- Prerequisites
- Create the Amazon VPC
- Create the Lambda function
- Create the EventBridge rule
- Create the Amazon MWAA environment

Prerequisites

To complete the steps in this tutorial, you will need the following:

- ...

Create the Amazon VPC

Use the following CloudFormation template and AWS CLI command to create a new Amazon VPC. The template sets up the Amazon VPC resources and modifies the endpoint policy to restrict access to a specific queue.

1. Download the CloudFormation template, then unzip the `.yml` file.
2. In a new command prompt window, navigate to the folder where you saved the template, then use create-stack to create the stack. The `--template-body` flag specifies the path to the template. Replace *`stack-name`*, *`availability-zone-1`*, and *`availability-zone-2`* with your own values. The two Availability Zones must be different and supported in your Region.

```
aws cloudformation create-stack --stack-name stack-name --
template-body file://cfn-vpc-private-network.yml --parameters
 ParameterKey=AvailabilityZone1,ParameterValue=availability-zone-1
 ParameterKey=AvailabilityZone2,ParameterValue=availability-zone-2
```

In the next section, you'll create the Lambda function.

Create the Lambda function

Use the following Python code and IAM JSON policy to create a new Lambda function and execution role. This function creates Amazon VPC endpoints for a private Apache Airflow webserver and an Amazon SQS queue. Amazon MWAA uses Amazon SQS to queue tasks with Celery among multiple workers when scaling your environment.

1. Download the Python function code.
2. Download the IAM permission policy, then unzip the file.
3. Open a command prompt, then navigate to the folder where you saved the JSON permission policy. Use the IAM `create-role` command to create the new role.

   ```
   aws iam create-role --role-name function-role \
         --assume-role-policy-document file://lambda-mwaa-vpce-policy.json
   ```

   Note the role ARN from the AWS CLI response. In the next step, we specify this new role as the function's execution role using its ARN.
4. Navigate to the folder where you saved the function code, then use the`create-function` command to create a new function.

   ```
   aws lambda create-function --function-name mwaa-vpce-lambda \
   --zip-file file://mwaa-lambda-shared-vpc.zip --runtime python3.8 --role
    arn:aws:iam::123456789012:role/function-role --handler lambda_handler
   ```

   Note the function ARN from the AWS CLI response. In the next step we specify the ARN to configure the function as a target for a new EventBridge rule.

In the next section, you create the EventBridge rule that invokes this function when the environment enters a `PENDING` state.

Create the EventBridge rule

Do the following to create a new rule that listens for Amazon MWAA notifications and targets your new Lambda function.

1. Use the EventBridge `put-rule` command to create a new EventBridge rule.

```
aws events put-rule --name "mwaa-lambda-rule" \
--event-pattern "{\"source\":[\"aws.airflow\"],\"detail-type\":[\"MWAA Environment
 Status Change\"]}"
```

The event pattern listens for notifications that Amazon MWAA sends whenever an environment status changes.

```
{
     "source": ["aws.airflow"],
     "detail-type": ["MWAA Environment Status Change"]
     }
```

2. Use the `put-targets` command to add the Lambda function as a target for the new rule.

```
aws events put-targets --rule "mwaa-lambda-rule" \
--targets "Id"="1","Arn"="arn:aws:lambda:us-east-1:123456789012:function:mwaa-vpce-
lambda"
```

You're ready to create a new Amazon MWAA environment with customer-managed Amazon VPC endpoints.

Create the Amazon MWAA environment

Use the Amazon MWAA console to create a new environment with customer-managed Amazon VPC endpoints.

1. Open the Amazon MWAA console, and choose **Create an environment**.
2. For **Name** enter a unique name.
3. For **Airflow version** choose the latest version.
4. Choose an **Amazon S3 bucket** and a **DAGs folder**, such as `dags/` to use with the environment, then choose **Next**.
5. On the **Configure advanced settings** page, do the following:
   a. For **Virtual Private Cloud**, choose the Amazon VPC you created in the previous step.
   b. For **webserver access**, choose **Public network (internet accessible)**.

c. For **Security groups**, choose the security group you created with CloudFormation. Because the security groups for the AWS PrivateLink endpoints from the earlier step are self-referencing, you must choose the same security group for your environment.

d. For **Endpoint management**, choose **Customer managed endpoints**.

6. Keep the remaining default settings, then choose **Next**.

7. Review your selections, then choose **Create environment**.

> **Tip**
>
> For more information about setting up a new environment, refer to Getting started with Amazon MWAA.

When the environment is `PENDING`, Amazon MWAA sends a notification that matches the event pattern you set for your rule. The rule invokes your Lambda function. The function parses the notification event and gets the required endpoint information for the webserver and the Amazon SQS queue. It then creates the endpoints in your Amazon VPC.

When the endpoints are available, Amazon MWAA resumes creating your environment. When ready, the environment status changes to `AVAILABLE` and you can access the Apache Airflow webserver using the Amazon MWAA console.

Code examples for Amazon Managed Workflows for Apache Airflow

This guide contains code samples, including DAGs and custom plugins, that you can use on an Amazon Managed Workflows for Apache Airflow environment. For more examples of using Apache Airflow with AWS services, refer to the `dags` directory in the Apache Airflow GitHub repository.

Samples

- Using a DAG to import variables in the CLI
- Creating an SSH connection using the SSHOperator
- Using a secret key in AWS Secrets Manager for an Apache Airflow Snowflake connection
- Using a DAG to write custom metrics in CloudWatch
- Aurora PostgreSQL database cleanup on an Amazon MWAA environment
- Exporting environment metadata to CSV files on Amazon S3
- Using a secret key in AWS Secrets Manager for an Apache Airflow variable
- Using a secret key in AWS Secrets Manager for an Apache Airflow connection
- Creating a custom plugin with Oracle
- Changing a DAG's timezone on Amazon MWAA
- Refreshing a CodeArtifact token
- Creating a custom plugin with Apache Hive and Hadoop
- Invoking DAGs with a Lambda function
- Invoking DAGs in different Amazon MWAA environments
- Using Amazon MWAA with Amazon RDS for Microsoft SQL Server
- Using Amazon MWAA with Amazon EKS
- Connecting to Amazon ECS using the ECSOperator
- Using dbt with Amazon MWAA
- Using DuckDB with Amazon MWAA
- AWS blogs and tutorials

Using a DAG to import variables in the CLI

The following sample code imports variables using the CLI on Amazon Managed Workflows for Apache Airflow.

Topics

- Version
- Prerequisites
- Permissions
- Dependencies
- Code sample
- What's next?

Version

You can use the code example on this page with **Apache Airflow v2** in Python 3.10 and **Apache Airflow v3** in Python 3.11.

Prerequisites

No additional permissions are required to use the code example on this page.

Permissions

Your AWS account needs access to the `AmazonMWAAAirflowCliAccess` policy. To learn more, refer to Apache Airflow CLI policy: AmazonMWAAAirflowCliAccess.

Dependencies

To use this code example with Apache Airflow v2 and later, no additional dependencies are required. Use aws-mwaa-docker-images to install Apache Airflow.

Code sample

The following sample code takes three inputs: your Amazon MWAA environment name (in `mwaa_env`), the AWS Region of your environment (in `aws_region`), and the local file that contains the variables you want to import (in `var_file`).

```
import boto3
import json
import requests
import base64
import getopt
import sys

argv = sys.argv[1:]
mwaa_env=''
aws_region=''
var_file=''

try:
    opts, args = getopt.getopt(argv, 'e:v:r:', ['environment', 'variable-
file','region'])
    #if len(opts) == 0 and len(opts) > 3:
    if len(opts) != 3:
        print ('Usage: -e MWAA environment -v variable file location and filename -r
 aws region')
    else:
        for opt, arg in opts:
            if opt in ("-e"):
                mwaa_env=arg
            elif opt in ("-r"):
                aws_region=arg
            elif opt in ("-v"):
                var_file=arg

        boto3.setup_default_session(region_name="{}".format(aws_region))
        mwaa_env_name = "{}".format(mwaa_env)

        client = boto3.client('mwaa')
        mwaa_cli_token = client.create_cli_token(
            Name=mwaa_env_name
        )

        with open ("{}".format(var_file), "r") as myfile:
            fileconf = myfile.read().replace('\n', '')

        json_dictionary = json.loads(fileconf)
        for key in json_dictionary:
            print(key, " ", json_dictionary[key])
            val = (key + " " + json_dictionary[key])
```

```
            mwaa_auth_token = 'Bearer ' + mwaa_cli_token['CliToken']
            mwaa_webserver_hostname = 'https://{0}/aws_mwaa/
cli'.format(mwaa_cli_token['WebServerHostname'])
            raw_data = "variables set {0}".format(val)
            mwaa_response = requests.post(
                mwaa_webserver_hostname,
                headers={
                    'Authorization': mwaa_auth_token,
                    'Content-Type': 'text/plain'
                    },
                data=raw_data
                )
            mwaa_std_err_message = base64.b64decode(mwaa_response.json()
['stderr']).decode('utf8')
            mwaa_std_out_message = base64.b64decode(mwaa_response.json()
['stdout']).decode('utf8')
            print(mwaa_response.status_code)
            print(mwaa_std_err_message)
            print(mwaa_std_out_message)

except:
    print('Use this script with the following options: -e MWAA environment -v variable
 file location and filename -r aws region')
    print("Unexpected error:", sys.exc_info()[0])
    sys.exit(2)
```

What's next?

- Learn how to upload the DAG code in this example to the `dags` folder in your Amazon S3 bucket in Adding or updating DAGs.

Creating an SSH connection using the `SSHOperator`

The following example describes how you can use the `SSHOperator` in a directed acyclic graph (DAG) to connect to a remote Amazon EC2 instance from your Amazon Managed Workflows for Apache Airflow environment. You can use a similar approach to connect to any remote instance with SSH access.

In the following example, you upload an SSH secret key (`.pem`) to your environment's `dags` directory on Amazon S3. Then, you install the necessary dependencies using `requirements.txt`

and create a new Apache Airflow connection in the UI. Finally, you write a DAG that creates an SSH connection to the remote instance.

Topics

- Version
- Prerequisites
- Permissions
- Requirements
- Copy your secret key to Amazon S3
- Create a new Apache Airflow connection
- Code sample

Version

You can use the code example on this page with **Apache Airflow v2** in Python 3.10 and **Apache Airflow v3** in Python 3.11.

Prerequisites

To use the sample code on this page, you'll need the following:

- An Amazon MWAA environment.
- An SSH secret key. The code sample assumes you have an Amazon EC2 instance and a `.pem` in the same Region as your Amazon MWAA environment. If you don't have a key, refer to Create or import a key pair in the *Amazon EC2 User Guide*.

Permissions

No additional permissions are required to use the code example on this page.

Requirements

Add the following parameter to `requirements.txt` to install the `apache-airflow-providers-ssh` package on the webserver. Once your environment updates and Amazon MWAA successfully installs the dependency, you will get a new **SSH** connection type in the UI.

```
-c https://raw.githubusercontent.com/apache/airflow/constraints-Airflow-version/
constraints-Python-version.txt
apache-airflow-providers-ssh
```

> **Note**
>
> `-c` defines the constraints URL in `requirements.txt`. This ensures that Amazon MWAA installs the correct package version for your environemnt.

Copy your secret key to Amazon S3

Use the following AWS Command Line Interface command to copy your `.pem` key to your environment's `dags` directory in Amazon S3.

```
aws s3 cp your-secret-key.pem s3://amzn-s3-demo-bucket/dags/
```

Amazon MWAA copies the content in `dags`, including the `.pem` key, to the local `/usr/local/airflow/dags/` directory, By doing this, Apache Airflow can access the key.

Create a new Apache Airflow connection

To create a new SSH connection using the Apache Airflow UI

1. Open the Environments page on the Amazon MWAA console.
2. From the list of environments, choose **Open Airflow UI** for your environment.
3. On the Apache Airflow UI page, choose **Admin** from the main navigation bar to expand the dropdown list, then choose **Connections**.
4. On the **List Connections** page, choose **+**, or **Add a new record** button to add a new connection.
5. On the **Add Connection** page, add the following information:
   a. For **Connection Id**, enter **`ssh_new`**.
   b. For **Connection Type**, choose **SSH** from the dropdown list.

> **Note**
>
> If the **SSH** connection type is not available in the list, Amazon MWAA hasn't installed the required `apache-airflow-providers-ssh` package. Update your `requirements.txt` file to include this package, then try again.

c. For **Host**, enter the IP address for the Amazon EC2 instance that you want to connect to. For example, **`12.345.67.89`**.

d. For **Username**, enter **`ec2-user`** if you are connecting to an Amazon EC2 instance. Your username might be different, depending on the type of remote instance you want Apache Airflow to connect to.

e. For **Extra**, enter the following key-value pair in JSON format:

```
{ "key_file": "/usr/local/airflow/dags/your-secret-key.pem" }
```

This key-value pair instructs Apache Airflow to search for the secret key in the local `/dags` directory.

Code sample

The following DAG uses the `SSHOperator` to connect to your target Amazon EC2 instance, then runs the `hostname` Linux command to print the name of the instance. You can modify the DAG to run any command or script on the remote instance.

1. Open a terminal, and navigate to the directory where your DAG code is stored. For example:

```
cd dags
```

2. Copy the contents of the following code sample and save locally as `ssh.py`.

```
from airflow.decorators import dag
from datetime import datetime
from airflow.providers.ssh.operators.ssh import SSHOperator

@dag(
    dag_id="ssh_operator_example",
    schedule_interval=None,
```

```
    start_date=datetime(2022, 1, 1),
    catchup=False,
    )
def ssh_dag():
    task_1=SSHOperator(
        task_id="ssh_task",
        ssh_conn_id='ssh_new',
        command='hostname',
    )

my_ssh_dag = ssh_dag()
```

3. Run the following AWS CLI command to copy the DAG to your environment's bucket, then trigger the DAG using the Apache Airflow UI.

```
aws s3 cp your-dag.py s3://your-environment-bucket/dags/
```

4. If successful, you'll get output similar to the following in the task logs for `ssh_task` in the `ssh_operator_example` DAG:

```
[2022-01-01, 12:00:00 UTC] {{base.py:79}} INFO - Using connection to: id: ssh_new.
 Host: 12.345.67.89, Port: None,
Schema: , Login: ec2-user, Password: None, extra: {'key_file': '/usr/local/airflow/
dags/your-secret-key.pem'}
[2022-01-01, 12:00:00 UTC] {{ssh.py:264}} WARNING - Remote Identification Change is
 not verified. This won't protect against Man-In-The-Middle attacks [2022-01-01,
 12:00:00 UTC] {{ssh.py:270}} WARNING - No Host Key Verification. This won't
 protect against Man-In-The-Middle attacks
[2022-01-01, 12:00:00 UTC] {{transport.py:1819}} INFO - Connected (version 2.0,
 client OpenSSH_7.4)
[2022-01-01, 12:00:00 UTC] {{transport.py:1819}} INFO - Authentication (publickey)
 successful!
[2022-01-01, 12:00:00 UTC] {{ssh.py:139}} INFO - Running command: hostname
[2022-01-01, 12:00:00 UTC]{{ssh.py:171}} INFO - ip-123-45-67-89.us-
west-2.compute.internal
[2022-01-01, 12:00:00 UTC] {{taskinstance.py:1280}} INFO - Marking task as SUCCESS.
 dag_id=ssh_operator_example, task_id=ssh_task, execution_date=20220712T200914,
 start_date=20220712T200915, end_date=20220712T200916
```

Using a secret key in AWS Secrets Manager for an Apache Airflow Snowflake connection

The following sample calls AWS Secrets Manager to get a secret key for an Apache Airflow Snowflake connection on Amazon Managed Workflows for Apache Airflow. It assumes you've completed the steps in Configuring an Apache Airflow connection using a AWS Secrets Manager secret.

Topics

- Version
- Prerequisites
- Permissions
- Requirements
- Code sample
- What's next?

Version

You can use the code example on this page with **Apache Airflow v2** in Python 3.10 and **Apache Airflow v3** in Python 3.11.

Prerequisites

To use the sample code on this page, you'll need the following:

- The Secrets Manager backend as an Apache Airflow configuration option as listed in Configuring an Apache Airflow connection using a AWS Secrets Manager secret.
- An Apache Airflow connection string in Secrets Manager as listed in Configuring an Apache Airflow connection using a AWS Secrets Manager secret.

Permissions

- Secrets Manager permissions as listed in Configuring an Apache Airflow connection using a AWS Secrets Manager secret.

Requirements

To use the sample code on this page, add the following dependencies to your `requirements.txt`. To learn more, refer to Installing Python dependencies.

```
apache-airflow-providers-snowflake==1.3.0
```

Code sample

The following steps describe how to create the DAG code that calls Secrets Manager to get the secret.

1. In your command prompt, navigate to the directory where your DAG code is stored. For example:

   ```
   cd dags
   ```

2. Copy the contents of the following code sample and save locally as `snowflake_connection.py`.

   ```
   """
   Copyright Amazon.com, Inc. or its affiliates. All Rights Reserved.

   Permission is hereby granted, free of charge, to any person obtaining a copy of
   this software and associated documentation files (the "Software"), to deal in
   the Software without restriction, including without limitation the rights to
   use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of
   the Software, and to permit persons to whom the Software is furnished to do so.

   THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR
   IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS
   FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR
   COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER
   IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN
   CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.
   """
   from airflow import DAG
   from airflow.providers.snowflake.operators.snowflake import SnowflakeOperator
   from airflow.utils.dates import days_ago

   snowflake_query = [
   ```

```
    """use warehouse "MY_WAREHOUSE";""",
    """select * from "SNOWFLAKE_SAMPLE_DATA"."WEATHER"."WEATHER_14_TOTAL" limit
 100;""",
]

with DAG(dag_id='snowflake_test', schedule_interval=None, catchup=False,
 start_date=days_ago(1)) as dag:
    snowflake_select = SnowflakeOperator(
        task_id="snowflake_select",
        sql=snowflake_query,
        snowflake_conn_id="snowflake_conn",
    )
```

What's next?

- Learn how to upload the DAG code in this example to the `dags` folder in your Amazon S3 bucket in Adding or updating DAGs.

Using a DAG to write custom metrics in CloudWatch

You can use the following code example to write a directed acyclic graph (DAG) that runs a `PythonOperator` to retrieve OS-level metrics for an Amazon MWAA environment. The DAG then publishes the data as custom metrics to Amazon CloudWatch.

Custom OS-level metrics provide you with additional visibility about how your environment workers are utilizing resources such as virtual memory and CPU. You can use this information to select the environment class that best suits your workload.

Topics

- Version
- Prerequisites
- Permissions
- Dependencies
- Code example

Version

You can use the code example on this page with **Apache Airflow v2** in Python 3.10 and **Apache Airflow v3** in Python 3.11.

Prerequisites

To use the code example on this page, you need the following:

- An Amazon MWAA environment.

Permissions

No additional permissions are required to use the code example on this page.

Dependencies

- No additional dependencies are required to use the code example on this page.

Code example

1. In your command prompt, navigate to the folder where your DAG code is stored. For example:

```
cd dags
```

2. Copy the contents of the following code example and save it locally as `dag-custom-metrics.py`. Replace `MWAA-ENV-NAME` with your environment name.

```
from airflow import DAG
from airflow.operators.python_operator import PythonOperator
from airflow.utils.dates import days_ago
from datetime import datetime
import os,json,boto3,psutil,socket

def publish_metric(client,name,value,cat,unit='None'):
    environment_name = os.getenv("MWAA_ENV_NAME")
    value_number=float(value)
    hostname = socket.gethostname()
    ip_address = socket.gethostbyname(hostname)
    print('writing value',value_number,'to metric',name)
```

```
    response = client.put_metric_data(
        Namespace='MWAA-Custom',
        MetricData=[
            {
                'MetricName': name,
                'Dimensions': [
                    {
                        'Name': 'Environment',
                        'Value': environment_name
                    },
                    {
                        'Name': 'Category',
                        'Value': cat
                    },
                    {
                        'Name': 'Host',
                        'Value': ip_address
                    },
                ],
                'Timestamp': datetime.now(),
                'Value': value_number,
                'Unit': unit
            },
        ]
    )
    print(response)
    return response

def python_fn(**kwargs):
    client = boto3.client('cloudwatch')

    cpu_stats = psutil.cpu_stats()
    print('cpu_stats', cpu_stats)

    virtual = psutil.virtual_memory()
    cpu_times_percent = psutil.cpu_times_percent(interval=0)

    publish_metric(client=client, name='virtual_memory_total',
 cat='virtual_memory', value=virtual.total, unit='Bytes')
    publish_metric(client=client, name='virtual_memory_available',
 cat='virtual_memory', value=virtual.available, unit='Bytes')
    publish_metric(client=client, name='virtual_memory_used', cat='virtual_memory',
 value=virtual.used, unit='Bytes')
```

```
    publish_metric(client=client, name='virtual_memory_free', cat='virtual_memory',
 value=virtual.free, unit='Bytes')
    publish_metric(client=client, name='virtual_memory_active',
 cat='virtual_memory', value=virtual.active, unit='Bytes')
    publish_metric(client=client, name='virtual_memory_inactive',
 cat='virtual_memory', value=virtual.inactive, unit='Bytes')
    publish_metric(client=client, name='virtual_memory_percent',
 cat='virtual_memory', value=virtual.percent, unit='Percent')

    publish_metric(client=client, name='cpu_times_percent_user',
 cat='cpu_times_percent', value=cpu_times_percent.user, unit='Percent')
    publish_metric(client=client, name='cpu_times_percent_system',
 cat='cpu_times_percent', value=cpu_times_percent.system, unit='Percent')
    publish_metric(client=client, name='cpu_times_percent_idle',
 cat='cpu_times_percent', value=cpu_times_percent.idle, unit='Percent')

    return "OK"


with DAG(dag_id=os.path.basename(__file__).replace(".py", ""),
 schedule_interval='*/5 * * * *', catchup=False, start_date=days_ago(1)) as dag:
    t = PythonOperator(task_id="memory_test", python_callable=python_fn,
 provide_context=True)
```

3. Run the following AWS CLI command to copy the DAG to your environment's bucket, then trigger the DAG using the Apache Airflow UI.

```
aws s3 cp your-dag.py s3://your-environment-bucket/dags/
```

4. If the DAG runs successfully, you get something similar to the following in your Apache Airflow logs:

```
[2022-08-16, 10:54:46 UTC] {{logging_mixin.py:109}} INFO -
 cpu_stats scpustats(ctx_switches=3253992384, interrupts=1964237163,
 soft_interrupts=492328209, syscalls=0)
[2022-08-16, 10:54:46 UTC] {{logging_mixin.py:109}} INFO - writing value
 16024199168.0 to metric virtual_memory_total
[2022-08-16, 10:54:46 UTC] {{logging_mixin.py:109}} INFO - {'ResponseMetadata':
 {'RequestId': 'fad289ac-aa51-46a9-8b18-24e4e4063f4d', 'HTTPStatusCode': 200,
 'HTTPHeaders': {'x-amzn-requestid': 'fad289ac-aa51-46a9-8b18-24e4e4063f4d',
 'content-type': 'text/xml', 'content-length': '212', 'date': 'Tue, 16 Aug 2022
 17:54:45 GMT'}, 'RetryAttempts': 0}}
```

```
[2022-08-16, 10:54:46 UTC] {{logging_mixin.py:109}} INFO - writing value
 14356287488.0 to metric virtual_memory_available
[2022-08-16, 10:54:46 UTC] {{logging_mixin.py:109}} INFO - {'ResponseMetadata':
 {'RequestId': '6ef60085-07ab-4865-8abf-dc94f90cab46', 'HTTPStatusCode': 200,
 'HTTPHeaders': {'x-amzn-requestid': '6ef60085-07ab-4865-8abf-dc94f90cab46',
 'content-type': 'text/xml', 'content-length': '212', 'date': 'Tue, 16 Aug 2022
 17:54:45 GMT'}, 'RetryAttempts': 0}}
[2022-08-16, 10:54:46 UTC] {{logging_mixin.py:109}} INFO - writing value
 1342296064.0 to metric virtual_memory_used
[2022-08-16, 10:54:46 UTC] {{logging_mixin.py:109}} INFO - {'ResponseMetadata':
 {'RequestId': 'd5331438-5d3c-4df2-bc42-52dcf8d60a00', 'HTTPStatusCode': 200,
 'HTTPHeaders': {'x-amzn-requestid': 'd5331438-5d3c-4df2-bc42-52dcf8d60a00',
 'content-type': 'text/xml', 'content-length': '212', 'date': 'Tue, 16 Aug 2022
 17:54:45 GMT'}, 'RetryAttempts': 0}}
...
[2022-08-16, 10:54:46 UTC] {{python.py:152}} INFO - Done. Returned value was: OK
[2022-08-16, 10:54:46 UTC] {{taskinstance.py:1280}} INFO - Marking task as SUCCESS.
 dag_id=dag-custom-metrics, task_id=memory_test, execution_date=20220816T175444,
 start_date=20220816T175445, end_date=20220816T175446
[2022-08-16, 10:54:46 UTC] {{local_task_job.py:154}} INFO - Task exited with return
 code 0
```

Aurora PostgreSQL database cleanup on an Amazon MWAA environment

Amazon Managed Workflows for Apache Airflow uses an Aurora PostgreSQL database as the Apache Airflow metadata database, where DAG runs and task instances are stored. The following sample code periodically clears out entries from the dedicated Aurora PostgreSQL database for your Amazon MWAA environment.

Important

Apache Airflow v3 restricts direct metadata database access from task code. Workers no longer connect to the metadata database, and DAG or task code can't import or use Apache Airflow database sessions or models directly. This change improves security and scalability. However, the DAG-based database cleanup approach that works in Apache Airflow v2 doesn't work in Apache Airflow v3 environments.

Instead, use the `airflow db clean` CLI command through the Amazon MWAA CLI endpoint to perform metadata database cleanup.

Note

Over time, the metadata database accumulates old records, XCom data, and stale task data. This growth uses database connections, slows your environment, and delays tasks. Run regular metadata cleanup to prevent these issues.

Topics

- Version
- Prerequisites
- Dependencies
- Code sample

Version

The code samples on this page are specific to Apache Airflow v2 and v3 supported on Amazon MWAA. Refer to the supported Apache Airflow versions.

Prerequisites

To use the sample code on this page, you'll need the following:

- An Amazon MWAA environment.

Dependencies

To use this code example with Apache Airflow v2, no additional dependencies are required. Use aws-mwaa-docker-images to install Apache Airflow.

Code sample

The following examples show how to clean the metadata database on your Amazon MWAA environment.

Apache Airflow v3.0.6 to 3.3.1

Important considerations

- You must specify `--clean-before-timestamp` to control how far back the cleanup reaches. Use an ISO 8601 formatted timestamp (for example, `2025-01-01T00:00:00+00:00`).
- We recommend that you specify `--tables` to limit the cleanup to specific tables. If omitted, the command cleans all supported tables.
- Start with a small scope — use an older `--clean-before-timestamp` value (closer to your environment creation date) and a single table first. This limits the cleanup to only the oldest records. Because the command deletes everything before the specified timestamp, using a more recent timestamp results in a larger deletion scope. Gradually move the timestamp forward as you gain confidence in the process.
- Large-scale cleanup can impact database performance. Deleting a high volume of records puts pressure on the Aurora PostgreSQL database and might affect the responsiveness of your environment. Use the `--batch-size` parameter to control transaction size, and consider running cleanup during low-traffic periods. Exercise caution when running on production environments.

Dependent table behavior

When you specify a table with `--tables`, the command automatically includes any dependent (child) tables that have foreign key relationships with the specified table. Child table records are deleted first, then the parent table records, to satisfy foreign key constraints. For example, specifying `--tables dag_run` also cleans `task_instance`, `task_instance_history`, `xcom`, `task_state_store`, and `deadline` because these tables reference `dag_run` through foreign keys.

The following table summarizes the dependency chains.

Table specified	Additional tables cleaned (dependents)
`dag_run`	`task_instance` , `task_instance_history` , `xcom`, `task_state_store` , `deadline`

Table specified	Additional tables cleaned (dependents)
dag	dag_version , deadline
task_instance	task_instance_history , xcom
trigger	task_instance , task_instance_history , xcom
dag_version	task_instance , task_instance_history , xcom, dag_run

Tables without dependents (such as log, job, import_error, sla_miss) are cleaned in isolation when specified.

Use --dry-run to see exactly which tables and how many rows would be affected before committing to a cleanup.

Available parameters for `airflow db clean`

The following table describes the available parameters for airflow db clean.

Parameter	Description	Default
--clean-before-timestamp	(Required) The date or timestamp before which data is purged. If no timezone is supplied, the Apache Airflow default timezone is assumed. Example: 2025-01-01T00:00:00+00:00	None
--tables or -t	Table names to perform maintenance on (comma-separated). Options include: dag_run, task_instance , task_inst	None

Parameter	Description	Default
	`ance_history` ,`log`, `job`, `xcom`, `import_er ror` ,`task_reschedule` , `trigger`, `dag`, `dag_versi on` ,`sla_miss`, `callback_ request` ,`celery_ta skmeta` ,`celery_ta sksetmeta` ,`asset_eve nt` ,`deadline`, `revoked_t oken` ,`task_stat e_store` ,`connectio n_test_request` , `_xcom_archive`	
`--batch-size`	Maximum number of rows to delete or archive in a single transaction. Lower values reduce long-running locks but increase the number of batches.	None
`--dry-run`	Perform a dry run without actually deleting data. Recommended for initial testing.	False

Parameter	Description	Default
`--skip-archive`	Don't preserve purged records in an archive table. By default, `db clean` moves purged records into archive tables (named with a `_<table>_archive` convention, for example `_dag_run_archive`) instead of permanently deleting them. This provides a safety net — you can inspect archived data, export it with `airflow db export-archived` , or drop it later with `airflow db drop-archived` . When `--skip-archive` is set, records are permanently deleted without this intermediate step.	False
`--dag-ids`	Only cleanup data related to the given DAG IDs.	None
`--exclude-dag-ids`	Avoid cleaning up data related to the given DAG IDs.	None
`-y`, `--yes`	Skip the confirmation prompt. Required for non-interactive CLI execution through Amazon MWAA.	False
`-v`, `--verbose`	Make logging output more verbose.	False

For more information about the available parameters, see CLI and env variables reference on the Apache Airflow website.

Code samples

The following examples show how to invoke `airflow db clean` through the Amazon MWAA CLI endpoint. For more information about creating CLI tokens, see Creating an Apache Airflow CLI token.

Using a Python script:

```
import boto3
import base64
import requests

# Replace with your environment name and AWS Region
mwaa_env_name = "YOUR_ENVIRONMENT_NAME"
region = "YOUR_REGION"

# Configure cleanup scope
clean_before_timestamp = "2025-06-01T00:00:00+00:00"
tables = "dag_run,task_instance,log,job,xcom"

# Build the Airflow CLI command
# -y flag is required to skip interactive confirmation prompt
airflow_cmd = f"db clean --clean-before-timestamp {clean_before_timestamp} --tables
 {tables} -y"

# Create a CLI token
client = boto3.client("mwaa", region_name=region)
cli_token_response = client.create_cli_token(Name=mwaa_env_name)

cli_token = cli_token_response["CliToken"]
web_server_hostname = cli_token_response["WebServerHostname"]

# Invoke the Airflow CLI through the MWAA endpoint
url = f"https://{web_server_hostname}/aws_mwaa/cli"
response = requests.post(
    url,
    headers={
        "Authorization": f"Bearer {cli_token}",
        "Content-Type": "text/plain",
    },
```

```
    data=airflow_cmd,
)

# Parse and display the results
stdout_message = base64.b64decode(response.json()["stdout"]).decode("utf-8")
stderr_message = base64.b64decode(response.json()["stderr"]).decode("utf-8")

print(f"Status code: {response.status_code}")
print(f"stdout:\n{stdout_message}")
print(f"stderr:\n{stderr_message}")
```

Dry run example (recommended first step)

Before performing an actual cleanup, run with `--dry-run` to see what would be deleted:

```
AIRFLOW_CMD="db clean --clean-before-timestamp 2025-06-01T00:00:00+00:00 --tables
 dag_run,task_instance --dry-run -y"
```

Apache Airflow v2.7.2 to 2.11.2

```
from airflow import DAG
from airflow.models.param import Param
from airflow.operators.bash_operator import BashOperator
from airflow.utils.dates import days_ago

from datetime import datetime, timedelta

# Note: Database commands might time out if running longer than 5 minutes. If this
 occurs, please increase the MAX_AGE_IN_DAYS (or change
# timestamp parameter to an earlier date) for initial runs, then reduce on
 subsequent runs until the desired retention is met.

MAX_AGE_IN_DAYS = 30

# To clean specific tables, please provide a comma-separated list per
# https://airflow.apache.org/docs/apache-airflow/stable/cli-and-env-variables-
ref.html#clean
# A value of None will clean all tables

TABLES_TO_CLEAN = None

with DAG(
    dag_id="clean_db_dag",
```

```
    schedule_interval=None,
    catchup=False,
    start_date=days_ago(1),
    params={
        "timestamp": Param(
            default=(datetime.now()-timedelta(days=MAX_AGE_IN_DAYS)).strftime("%Y-
%m-%d %H:%M:%S"),
            type="string",
            minLength=1,
            maxLength=255,
        ),
    }
) as dag:
    if TABLES_TO_CLEAN:
        bash_command="airflow db clean --clean-before-timestamp
 '{{ params.timestamp }}' --tables '"+TABLES_TO_CLEAN+"' --skip-archive --yes"
    else:
        bash_command="airflow db clean --clean-before-timestamp
 '{{ params.timestamp }}' --skip-archive --yes"

    cli_command = BashOperator(
        task_id="bash_command",
        bash_command=bash_command
    )
```

Exporting environment metadata to CSV files on Amazon S3

Use the following code example to create a directed acyclic graph (DAG) that queries the database for a range of DAG run information, and writes the data to `.csv` files stored on Amazon S3.

You might want to export information from your environment's Aurora PostgreSQL database to inspect the data locally, archive them in object storage, or combine them with tools like the Amazon S3 to Amazon Redshift operator and the database cleanup, to move Amazon MWAA metadata out of the environment, but preserve them for future analysis.

You can query the database for any of the objects listed in Apache Airflow models. This code sample uses three models, `DagRun`, `TaskFail`, and `TaskInstance`, which provide information relevant To DAG runs.

Topics

- Version

- Prerequisites
- Permissions
- Requirements
- Code sample

Version

You can use the code example on this page with **Apache Airflow v2** in Python 3.10 and **Apache Airflow v3** in Python 3.11.

Prerequisites

To use the sample code on this page, you'll need the following:

- An Amazon MWAA environment.
- A new Amazon S3 bucket where you want to export your metadata information.

Permissions

Amazon MWAA needs permission for the Amazon S3 action `s3:PutObject` to write the queried metadata information to Amazon S3. Add the following policy statement to your environment's execution role.

```
{
  "Effect": "Allow",
  "Action": "s3:PutObject*",
  "Resource": "arn:aws:s3:::amzn-s3-demo-bucket"
}
```

This policy limits write access to only *amzn-s3-demo-bucket*.

Requirements

To use this code example with Apache Airflow v2 and later, no additional dependencies are required. Use aws-mwaa-docker-images to install Apache Airflow.

Code sample

The following steps describe how you can create a DAG that queries the Aurora PostgreSQL and writes the result to your new Amazon S3 bucket.

1. In your terminal, navigate to the directory where your DAG code is stored. For example:

```
cd dags
```

2. Copy the contents of the following code example and save it locally as `metadata_to_csv.py`. You can change the value assigned to `MAX_AGE_IN_DAYS` to control the age of the oldest records your DAG queries from the metadata database.

```
from airflow.decorators import dag, task
from airflow import settings
import os
import boto3
from airflow.utils.dates import days_ago
from airflow.models import DagRun, TaskFail, TaskInstance
import csv, re
from io import StringIO

DAG_ID = os.path.basename(__file__).replace(".py", "")

MAX_AGE_IN_DAYS = 30
S3_BUCKET = '<your-export-bucket>'
S3_KEY = 'files/export/{0}.csv'

# You can add other objects to export from the metadatabase,
OBJECTS_TO_EXPORT = [
    [DagRun,DagRun.execution_date],
    [TaskFail,TaskFail.end_date],
    [TaskInstance, TaskInstance.execution_date],
]

@task()
def export_db_task(**kwargs):
    session = settings.Session()
    print("session: ",str(session))

    oldest_date = days_ago(MAX_AGE_IN_DAYS)
    print("oldest_date: ",oldest_date)
```

```
    s3 = boto3.client('s3')

    for x in OBJECTS_TO_EXPORT:
        query = session.query(x[0]).filter(x[1] >= days_ago(MAX_AGE_IN_DAYS))
        print("type",type(query))
        allrows=query.all()
        name=re.sub("[<>']", "", str(x[0]))
        print(name,": ",str(allrows))

        if len(allrows) > 0:
            outfileStr=""
            f = StringIO(outfileStr)
            w = csv.DictWriter(f, vars(allrows[0]).keys())
            w.writeheader()
            for y in allrows:
                w.writerow(vars(y))
            outkey = S3_KEY.format(name[6:])
            s3.put_object(Bucket=S3_BUCKET, Key=outkey, Body=f.getvalue())

@dag(
    dag_id=DAG_ID,
    schedule_interval=None,
    start_date=days_ago(1),
    )
def export_db():
    t = export_db_task()

metadb_to_s3_test = export_db()
```

3. Run the following AWS CLI command to copy the DAG to your environment's bucket, then trigger the DAG using the Apache Airflow UI.

```
aws s3 cp your-dag.py s3://your-environment-bucket/dags/
```

4. If successful, you'll output similar to the following in the task logs for the `export_db` task:

```
[2022-01-01, 12:00:00 PDT] {{logging_mixin.py:109}} INFO - type <class
 'sqlalchemy.orm.query.Query'>
[2022-01-01, 12:00:00 PDT] {{logging_mixin.py:109}} INFO - class
 airflow.models.dagrun.DagRun : [your-tasks]
[2022-01-01, 12:00:00 PDT] {{logging_mixin.py:109}} INFO - type <class
 'sqlalchemy.orm.query.Query'>
```

```
[2022-01-01, 12:00:00 PDT] {{logging_mixin.py:109}} INFO - class
 airflow.models.taskfail.TaskFail :  [your-tasks]
[2022-01-01, 12:00:00 PDT] {{logging_mixin.py:109}} INFO - type <class
 'sqlalchemy.orm.query.Query'>
[2022-01-01, 12:00:00 PDT] {{logging_mixin.py:109}} INFO - class
 airflow.models.taskinstance.TaskInstance :  [your-tasks]
[2022-01-01, 12:00:00 PDT] {{python.py:152}} INFO - Done. Returned value was: OK
[2022-01-01, 12:00:00 PDT] {{taskinstance.py:1280}} INFO - Marking task as
 SUCCESS. dag_id=metadb_to_s3, task_id=export_db, execution_date=20220101T000000,
 start_date=20220101T000000, end_date=20220101T000000
[2022-01-01, 12:00:00 PDT] {{local_task_job.py:154}} INFO - Task exited with return
 code 0
[2022-01-01, 12:00:00 PDT] {{local_task_job.py:264}} INFO - 0 downstream tasks
 scheduled from follow-on schedule check
```

You can now access and download the exported `.csv` files in your new Amazon S3 bucket in `/files/export/`.

Using a secret key in AWS Secrets Manager for an Apache Airflow variable

The following sample calls AWS Secrets Manager to get a secret key for an Apache Airflow variable on Amazon Managed Workflows for Apache Airflow. It assumes you've completed the steps in Configuring an Apache Airflow connection using a AWS Secrets Manager secret.

Topics

- Version
- Prerequisites
- Permissions
- Requirements
- Code sample
- What's next?

Version

You can use the code example on this page with **Apache Airflow v2** in Python 3.10 and **Apache Airflow v3** in Python 3.11.

Prerequisites

To use the sample code on this page, you'll need the following:

- The Secrets Manager backend as an Apache Airflow configuration option as listed in Configuring an Apache Airflow connection using a AWS Secrets Manager secret.
- An Apache Airflow variable string in Secrets Manager as listed in Configuring an Apache Airflow connection using a AWS Secrets Manager secret.

Permissions

- Secrets Manager permissions as listed in Configuring an Apache Airflow connection using a AWS Secrets Manager secret.

Requirements

To use this code example with Apache Airflow v2 and later, no additional dependencies are required. Use aws-mwaa-docker-images to install Apache Airflow.

Code sample

The following steps describe how to create the DAG code that calls Secrets Manager to get the secret.

1. In your command prompt, navigate to the directory where your DAG code is stored. For example:

```
cd dags
```

2. Copy the contents of the following code sample and save locally as `secrets-manager-var.py`.

```
from airflow import DAG
from airflow.operators.python_operator import PythonOperator
from airflow.models import Variable
from airflow.utils.dates import days_ago
from datetime import timedelta
import os
DAG_ID = os.path.basename(__file__).replace(".py", "")
```

```
DEFAULT_ARGS = {
    'owner': 'airflow',
    'depends_on_past': False,
    'email': ['airflow@example.com'],
    'email_on_failure': False,
    'email_on_retry': False,
}
def get_variable_fn(**kwargs):
    my_variable_name = Variable.get("test-variable", default_var="undefined")
    print("my_variable_name: ", my_variable_name)
    return my_variable_name
with DAG(
    dag_id=DAG_ID,
    default_args=DEFAULT_ARGS,
    dagrun_timeout=timedelta(hours=2),
    start_date=days_ago(1),
    schedule_interval='@once',
    tags=['variable']
) as dag:
    get_variable = PythonOperator(
        task_id="get_variable",
        python_callable=get_variable_fn,
        provide_context=True
    )
```

What's next?

- Learn how to upload the DAG code in this example to the `dags` folder in your Amazon S3 bucket in Adding or updating DAGs.

Using a secret key in AWS Secrets Manager for an Apache Airflow connection

The following sample calls AWS Secrets Manager to get a secret key for an Apache Airflow connection on Amazon Managed Workflows for Apache Airflow. It assumes you've completed the steps in Configuring an Apache Airflow connection using a AWS Secrets Manager secret.

Topics

- Version

- Prerequisites
- Permissions
- Requirements
- Code sample
- What's next?

Version

You can use the code example on this page with **Apache Airflow v2** in Python 3.10 and **Apache Airflow v3** in Python 3.11.

Prerequisites

To use the sample code on this page, you'll need the following:

- The Secrets Manager backend as an Apache Airflow configuration option as listed in Configuring an Apache Airflow connection using a AWS Secrets Manager secret.
- An Apache Airflow connection string in Secrets Manager as listed in Configuring an Apache Airflow connection using a AWS Secrets Manager secret.

Permissions

- Secrets Manager permissions as listed in Configuring an Apache Airflow connection using a AWS Secrets Manager secret.

Requirements

To use this code example with Apache Airflow v2 and later, no additional dependencies are required. Use aws-mwaa-docker-images to install Apache Airflow.

Code sample

The following steps describe how to create the DAG code that calls Secrets Manager to get the secret.

1. In your command prompt, navigate to the directory where your DAG code is stored. For example:

```
cd dags
```

2. Copy the contents of the following code sample and save locally as `secrets-manager.py`.

```
"""
Copyright Amazon.com, Inc. or its affiliates. All Rights Reserved.

Permission is hereby granted, free of charge, to any person obtaining a copy of
this software and associated documentation files (the "Software"), to deal in
the Software without restriction, including without limitation the rights to
use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of
the Software, and to permit persons to whom the Software is furnished to do so.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR
IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS
FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR
COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER
IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN
CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.
"""
from airflow import DAG, settings, secrets
from airflow.operators.python import PythonOperator
from airflow.utils.dates import days_ago
from airflow.providers.amazon.aws.hooks.base_aws import AwsBaseHook

from datetime import timedelta
import os

### The steps to create this secret key can be found at: https://
docs.aws.amazon.com/mwaa/latest/userguide/connections-secrets-manager.html
sm_secretId_name = 'airflow/connections/myconn'

default_args = {
    'owner': 'airflow',
    'start_date': days_ago(1),
    'depends_on_past': False
}


### Gets the secret myconn from Secrets Manager
def read_from_aws_sm_fn(**kwargs):
    ### set up Secrets Manager
```

```
    hook = AwsBaseHook(client_type='secretsmanager')
    client = hook.get_client_type(region_name='us-east-1')
    response = client.get_secret_value(SecretId=sm_secretId_name)
    myConnSecretString = response["SecretString"]

    return myConnSecretString

### 'os.path.basename(__file__).replace(".py", "")' uses the file name secrets-
manager.py for a DAG ID of secrets-manager
with DAG(
        dag_id=os.path.basename(__file__).replace(".py", ""),
        default_args=default_args,
        dagrun_timeout=timedelta(hours=2),
        start_date=days_ago(1),
        schedule_interval=None
) as dag:
    write_all_to_aws_sm = PythonOperator(
        task_id="read_from_aws_sm",
        python_callable=read_from_aws_sm_fn,
        provide_context=True
    )
```

What's next?

- Learn how to upload the DAG code in this example to the `dags` folder in your Amazon S3 bucket in Adding or updating DAGs.

Creating a custom plugin with Oracle

The following sample walks you through the steps to create a custom plugin using Oracle for Amazon MWAA and can be combined with other custom plugins and binaries in your plugins.zip file.

Contents

- Version
- Prerequisites
- Permissions
- Requirements

- Code sample
- Create the custom plugin
  - Download dependencies
  - Custom plugin
  - Plugins.zip
- Airflow configuration options
- What's next?

Version

You can use the code example on this page with **Apache Airflow v2** in Python 3.10 and **Apache Airflow v3** in Python 3.11.

Prerequisites

To use the sample code on this page, you'll need the following:

- An Amazon MWAA environment.
- Worker logging enabled at any log level, `CRITICAL` or in the previous section for your environment. For more information about Amazon MWAA log types and how to manage your log groups, refer to the section called “Accessing Airflow logs”

Permissions

No additional permissions are required to use the code example on this page.

Requirements

To use the sample code on this page, add the following dependencies to your `requirements.txt`. To learn more, refer to Installing Python dependencies.

```
-c https://raw.githubusercontent.com/apache/airflow/constraints-2.0.2/
constraints-3.7.txt
cx_Oracle
apache-airflow-providers-oracle
```

Code sample

The following steps describe how to create the DAG code that will test the custom plugin.

1. In your command prompt, navigate to the directory where your DAG code is stored. For example:

```
cd dags
```

2. Copy the contents of the following code sample and save locally as `oracle.py`.

```
from airflow import DAG
from airflow.operators.python_operator import PythonOperator
from airflow.utils.dates import days_ago
import os
import cx_Oracle

DAG_ID = os.path.basename(__file__).replace(".py", "")

def testHook(**kwargs):
    cx_Oracle.init_oracle_client()
    version = cx_Oracle.clientversion()
    print("cx_Oracle.clientversion",version)
    return version

with DAG(dag_id=DAG_ID, schedule_interval=None, catchup=False,
 start_date=days_ago(1)) as dag:
    hook_test = PythonOperator(
        task_id="hook_test",
        python_callable=testHook,
        provide_context=True
    )
```

Create the custom plugin

This section describes how to download the dependencies, create the custom plugin and the plugins.zip.

Download dependencies

Amazon MWAA will extract the contents of plugins.zip into `/usr/local/airflow/plugins` on each Amazon MWAA scheduler and worker container. This is used to add binaries to your environment. The following steps describe how to assemble the files needed for the custom plugin.

Pull the Amazon Linux container image

1. In your command prompt, pull the Amazon Linux container image, and run the container locally. For example:

   ```
   docker pull amazonlinux
         docker run -it amazonlinux:latest /bin/bash
   ```

   Your command prompt can invoke a bash command line. For example:

   ```
   bash-4.2#
   ```

2. Install the Linux-native asynchronous I/O facility (libaio).

   ```
   yum -y install libaio
   ```

3. Keep this window open for subsequent steps. We'll be copying the following files locally: `lib64/libaio.so.1`, `lib64/libaio.so.1.0.0`, `lib64/libaio.so.1.0.1`.

Download client folder

1. Install the unzip package locally. For example:

   ```
   sudo yum install unzip
   ```

2. Create an `oracle_plugin` directory. For example:

   ```
   mkdir oracle_plugin
   cd oracle_plugin
   ```

3. Use the following curl command to download the instantclient-basic-linux.x64-18.5.0.0.0dbru.zip from Oracle Instant Client Downloads for Linux x86-64 (64-bit).

```
curl https://download.oracle.com/otn_software/linux/instantclient/185000/
instantclient-basic-linux.x64-18.5.0.0.0dbru.zip > client.zip
```

4. Unzip the `client.zip` file. For example:

```
unzip *.zip
```

Extract files from Docker

1. In a new command prompt, show and write down your Docker container ID. For example:

```
docker container ls
```

Your command prompt can return all containers and their IDs. For example:

```
debc16fd6970
```

2. In your `oracle_plugin` directory, extract the `lib64/libaio.so.1`, `lib64/libaio.so.1.0.0`, `lib64/libaio.so.1.0.1` files to the local `instantclient_18_5` folder. For example:

```
docker cp debc16fd6970:/lib64/libaio.so.1 instantclient_18_5/
docker cp debc16fd6970:/lib64/libaio.so.1.0.0 instantclient_18_5/
docker cp debc16fd6970:/lib64/libaio.so.1.0.1 instantclient_18_5/
```

Custom plugin

Apache Airflow will execute the contents of Python files in the plugins folder at startup. This is used to set and modify environment variables. The following steps describe the sample code for the custom plugin.

- Copy the contents of the following code sample and save locally as `env_var_plugin_oracle.py`.

```
from airflow.plugins_manager import AirflowPlugin
import os
```

```
os.environ["LD_LIBRARY_PATH"]='/usr/local/airflow/plugins/instantclient_18_5'
os.environ["DPI_DEBUG_LEVEL"]="64"

class EnvVarPlugin(AirflowPlugin):
    name = 'env_var_plugin'
```

Plugins.zip

The following steps explain how to create the `plugins.zip`. The contents of this example can be combined with your other plugins and binaries into a single `plugins.zip` file.

Zip the contents of the plugin directory

1. In your command prompt, navigate to the `oracle_plugin` directory. For example:

   ```
   cd oracle_plugin
   ```

2. Zip the `instantclient_18_5` directory in plugins.zip. For example:

   ```
   zip -r ../plugins.zip ./
   ```

   Your command prompt displays:

   ```
   oracle_plugin$ ls
   client.zip  instantclient_18_5
   ```

3. Remove the `client.zip` file. For example:

   ```
   rm client.zip
   ```

Zip the env_var_plugin_oracle.py file

1. Add the `env_var_plugin_oracle.py` file to the root of the plugins.zip. For example:

   ```
   zip plugins.zip env_var_plugin_oracle.py
   ```

2. Your plugins.zip now includes the following:

   ```
   env_var_plugin_oracle.py
   ```

```
instantclient_18_5/
```

Airflow configuration options

If you're using Apache Airflow v2, add `core.lazy_load_plugins : False` as an Apache Airflow configuration option. To learn more, refer to Using configuration options to load plugins in 2.

What's next?

- Learn how to upload the `requirements.txt` file in this example to your Amazon S3 bucket in Installing Python dependencies.
- Learn how to upload the DAG code in this example to the `dags` folder in your Amazon S3 bucket in Adding or updating DAGs.
- Learn more about how to upload the `plugins.zip` file in this example to your Amazon S3 bucket in Installing custom plugins.

Changing a DAG's timezone on Amazon MWAA

Apache Airflow schedules your directed acyclic graph (DAG) in UTC+0 by default. The following steps present how you can change the timezone in which Amazon MWAA runs your DAGs with Pendulum. Optionally, this topic demonstrates how you can create a custom plugin to change the timezone for your environment's Apache Airflow logs.

Topics

- Version
- Prerequisites
- Permissions
- Create a plugin to change the timezone in Airflow logs
- Create a plugins.zip
- Code sample
- What's next?

Version

You can use the code example on this page with **Apache Airflow v2** in Python 3.10 and **Apache Airflow v3** in Python 3.11.

Prerequisites

To use the sample code on this page, you'll need the following:

- An Amazon MWAA environment.

Permissions

No additional permissions are required to use the code example on this page.

Create a plugin to change the timezone in Airflow logs

Apache Airflow run the Python files in the `plugins` directory at start-up. With the following plugin, you can override the executor's timezone, which modifies the timezone in which Apache Airflow writes logs.

1. Create a directory named `plugins` for your custom plugin, and navigate to the directory. For example:

```
$ mkdir plugins
$ cd plugins
```

2. Copy the contents of the following code sample and save locally as `dag-timezone-plugin.py` in the `plugins` folder.

```
import time
import os

os.environ['TZ'] = 'America/Los_Angeles'
time.tzset()
```

3. In the `plugins` directory, create an empty Python file named `__init__.py`. Your `plugins` directory should be similar to the following:

```
plugins/
```

```
|-- __init__.py
|-- dag-timezone-plugin.py
```

Create a `plugins.zip`

The following steps explain how to create `plugins.zip`. The content of this example can be combined with other plugins and binaries into a single `plugins.zip` file.

1. In your command prompt, navigate to the `plugins` directory from the previous step. For example:

```
cd plugins
```

2. Zip the contents within your `plugins` directory.

```
zip -r ../plugins.zip ./
```

3. Upload `plugins.zip` to your S3 bucket

```
aws s3 cp plugins.zip s3://your-mwaa-bucket/
```

Code sample

To change the default timezone (UTC+0) in which the DAG runs, we'll use a library called Pendulum, a Python library for working with timezone-aware datetime.

1. In your command prompt, navigate to the directory where your DAGs are stored. For example:

```
cd dags
```

2. Copy the content of the following example and save as `tz-aware-dag.py`.

```
from airflow import DAG
from airflow.operators.bash_operator import BashOperator
from datetime import datetime, timedelta
# Import the Pendulum library.
import pendulum

# Instantiate Pendulum and set your timezone.
```

```
local_tz = pendulum.timezone("America/Los_Angeles")

with DAG(
    dag_id = "tz_test",
    schedule_interval="0 12 * * *",
    catchup=False,
    start_date=datetime(2022, 1, 1, tzinfo=local_tz)
) as dag:
    bash_operator_task = BashOperator(
        task_id="tz_aware_task",
        dag=dag,
        bash_command="date"
    )
```

3. Run the following AWS CLI command to copy the DAG to your environment's bucket, then trigger the DAG using the Apache Airflow UI.

```
aws s3 cp your-dag.py s3://your-environment-bucket/dags/
```

4. If successful, you'll output similar to the following in the task logs for the `tz_aware_task` in the `tz_test` DAG:

```
[2022-08-01, 12:00:00 PDT] {{subprocess.py:74}} INFO - Running command: ['bash', '-
c', 'date']
[2022-08-01, 12:00:00 PDT] {{subprocess.py:85}} INFO - Output:
[2022-08-01, 12:00:00 PDT] {{subprocess.py:89}} INFO - Mon Aug  1 12:00:00 PDT 2022
[2022-08-01, 12:00:00 PDT] {{subprocess.py:93}} INFO - Command exited with return
 code 0
[2022-08-01, 12:00:00 PDT] {{taskinstance.py:1280}} INFO - Marking task as
 SUCCESS. dag_id=tz_test, task_id=tz_aware_task, execution_date=20220801T190033,
 start_date=20220801T190035, end_date=20220801T190035
[2022-08-01, 12:00:00 PDT] {{local_task_job.py:154}} INFO - Task exited with return
 code 0
[2022-08-01, 12:00:00 PDT] {{local_task_job.py:264}} INFO - 0 downstream tasks
 scheduled from follow-on schedule check
```

What's next?

- Learn more about how to upload the `plugins.zip` file in this example to your Amazon S3 bucket in Installing custom plugins.

Refreshing a CodeArtifact token

If you're using CodeArtifact to install Python dependencies, Amazon MWAA requires an active token. To allow Amazon MWAA to access an CodeArtifact repository at runtime, you can use a startup script and set the `PIP_EXTRA_INDEX_URL` with the token.

The following topic describes how you can create a startup script that uses the `get_authorization_token` CodeArtifact API operation to retrieve a fresh token every time your environment starts up, or updates.

Topics

- Version
- Prerequisites
- Permissions
- Code sample
- What's next?

Version

You can use the code example on this page with **Apache Airflow v2** in Python 3.10 and **Apache Airflow v3** in Python 3.11.

Prerequisites

To use the sample code on this page, you'll need the following:

- An Amazon MWAA environment.
- A CodeArtifact repository where you store dependencies for your environment.

Permissions

To refresh the CodeArtifact token and write the result to Amazon S3 Amazon MWAA must have the following permissions in the execution role.

- The `codeartifact:GetAuthorizationToken` action allows Amazon MWAA to retrieve a new token from CodeArtifact. The following policy grants permission for every CodeArtifact domain

you create. You can further restrict access to your domains by modifying the resource value in the statement, and specifying only the domains that you want your environment to access.

```
{
  "Effect": "Allow",
  "Action": "codeartifact:GetAuthorizationToken",
  "Resource": "arn:aws:codeartifact:us-west-2:*:domain/*"
}
```

- The `sts:GetServiceBearerToken` action is required to call the CodeArtifact `GetAuthorizationToken` API operation. This operation returns a token that must be used when using a package manager such as `pip` with CodeArtifact. To use a package manager with a CodeArtifact repository, your environment's execution role role must allow `sts:GetServiceBearerToken` as listed in the following policy statement.

```
{
  "Sid": "AllowServiceBearerToken",
  "Effect": "Allow",
  "Action": "sts:GetServiceBearerToken",
  "Resource": "*"
}
```

Code sample

The following steps describe how you can create a start up script that updates the CodeArtifact token.

1. Copy the contents of the following code sample and save locally as `code_artifact_startup_script.sh`.

```
#!/bin/sh

# Startup script for MWAA, refer to https://docs.aws.amazon.com/mwaa/latest/
userguide/using-startup-script.html

set -eu

# setup code artifact endpoint and token
# https://pip.pypa.io/en/stable/cli/pip_install/#cmdoption-0
# https://docs.aws.amazon.com/mwaa/latest/userguide/samples-code-artifact.html
```

```
DOMAIN="amazon"
DOMAIN_OWNER="112233445566"
REGION="us-west-2"
REPO_NAME="MyRepo"
echo "Getting token for CodeArtifact with args: --domain $DOMAIN --region $REGION
 --domain-owner $DOMAIN_OWNER"
TOKEN=$(aws codeartifact get-authorization-token --domain $DOMAIN --region $REGION
 --domain-owner $DOMAIN_OWNER | jq -r '.authorizationToken')
echo "Setting Pip env var for '--index-url' to point to CodeArtifact"
export PIP_EXTRA_INDEX_URL="https://aws:$TOKEN@$DOMAIN-
$DOMAIN_OWNER.d.codeartifact.$REGION.amazonaws.com/pypi/$REPO_NAME/simple/"
echo "CodeArtifact startup setup complete"
```

2. Navigate to the folder where you saved the script. Use `cp` in a new prompt window to upload the script to your bucket. Replace *amzn-s3-demo-bucket* with your information.

```
aws s3 cp code_artifact_startup_script.sh s3://amzn-s3-demo-bucket/
code_artifact_startup_script.sh
```

If successful, Amazon S3 outputs the URL path to the object:

```
upload: ./code_artifact_startup_script.sh to s3://amzn-s3-demo-bucket/
code_artifact_startup_script.sh
```

After you upload the script, your environment updates and runs the script at startup.

What's next?

- Learn how to use startup scripts to customize your environment in the section called “Using a startup script”.
- Learn how to upload the DAG code in this example to the `dags` folder in your Amazon S3 bucket in Adding or updating DAGs.
- Learn more about how to upload the `plugins.zip` file in this example to your Amazon S3 bucket in Installing custom plugins.

Creating a custom plugin with Apache Hive and Hadoop

Amazon MWAA extracts the contents of a `plugins.zip` to `/usr/local/airflow/plugins`. This can be used to add binaries to your containers. In addition, Apache Airflow executes the contents of Python files in the `plugins` folder at *startup*—enabling you to set and modify environment variables. The following sample walks you through the steps to create a custom plugin using Apache Hive and Hadoop on an Amazon Managed Workflows for Apache Airflow environment and can be combined with other custom plugins and binaries.

Topics

- Version
- Prerequisites
- Permissions
- Requirements
- Download dependencies
- Custom plugin
- Plugins.zip
- Code sample
- Airflow configuration options
- What's next?

Version

You can use the code example on this page with **Apache Airflow v2** in Python 3.10 and **Apache Airflow v3** in Python 3.11.

Prerequisites

To use the sample code on this page, you'll need the following:

- An Amazon MWAA environment.

Permissions

No additional permissions are required to use the code example on this page.

Requirements

To use the sample code on this page, add the following dependencies to your `requirements.txt`. To learn more, refer to Installing Python dependencies.

```
-c https://raw.githubusercontent.com/apache/airflow/constraints-2.0.2/
constraints-3.7.txt
apache-airflow-providers-amazon[apache.hive]
```

Download dependencies

Amazon MWAA will extract the contents of plugins.zip into `/usr/local/airflow/plugins` on each Amazon MWAA scheduler and worker container. This is used to add binaries to your environment. The following steps describe how to assemble the files needed for the custom plugin.

1. In your command prompt, navigate to the directory where you would like to create your plugin. For example:

   ```
   cd plugins
   ```

2. Download Hadoop from a mirror, for example:

   ```
   wget https://downloads.apache.org/hadoop/common/hadoop-3.3.0/hadoop-3.3.0.tar.gz
   ```

3. Download Hive from a mirror, for example:

   ```
   wget https://downloads.apache.org/hive/hive-3.1.2/apache-hive-3.1.2-bin.tar.gz
   ```

4. Create a directory. For example:

   ```
   mkdir hive_plugin
   ```

5. Extract Hadoop.

   ```
   tar -xvzf hadoop-3.3.0.tar.gz -C hive_plugin
   ```

6. Extract Hive.

   ```
   tar -xvzf apache-hive-3.1.2-bin.tar.gz -C hive_plugin
   ```

Custom plugin

Apache Airflow will execute the contents of Python files in the plugins folder at startup. This is used to set and modify environment variables. The following steps describe the sample code for the custom plugin.

1. In your command prompt, navigate to the `hive_plugin` directory. For example:

```
cd hive_plugin
```

2. Copy the contents of the following code sample and save locally as `hive_plugin.py` in the `hive_plugin` directory.

```
from airflow.plugins_manager import AirflowPlugin
import os
os.environ["JAVA_HOME"]="/usr/lib/jvm/jre"
os.environ["HADOOP_HOME"]='/usr/local/airflow/plugins/hadoop-3.3.0'
os.environ["HADOOP_CONF_DIR"]='/usr/local/airflow/plugins/hadoop-3.3.0/etc/hadoop'
os.environ["HIVE_HOME"]='/usr/local/airflow/plugins/apache-hive-3.1.2-bin'
os.environ["PATH"] = os.getenv("PATH") + ":/usr/local/airflow/plugins/
hadoop-3.3.0:/usr/local/airflow/plugins/apache-hive-3.1.2-bin/bin:/usr/local/
airflow/plugins/apache-hive-3.1.2-bin/lib"
os.environ["CLASSPATH"] = os.getenv("CLASSPATH") + ":/usr/local/airflow/plugins/
apache-hive-3.1.2-bin/lib"
class EnvVarPlugin(AirflowPlugin):
    name = 'hive_plugin'
```

3. Cope the content of the following text and save locally as `.airflowignore` in the `hive_plugin` directory.

```
hadoop-3.3.0
apache-hive-3.1.2-bin
```

Plugins.zip

The following steps explain how to create `plugins.zip`. The contents of this example can be combined with other plugins and binaries into a single `plugins.zip` file.

1. In your command prompt, navigate to the `hive_plugin` directory from the previous step. For example:

```
cd hive_plugin
```

2. Zip the contents within your `plugins` folder.

```
zip -r ../hive_plugin.zip ./
```

Code sample

The following steps describe how to create the DAG code that will test the custom plugin.

1. In your command prompt, navigate to the directory where your DAG code is stored. For example:

```
cd dags
```

2. Copy the contents of the following code sample and save locally as `hive.py`.

```
from airflow import DAG
from airflow.operators.bash_operator import BashOperator
from airflow.utils.dates import days_ago

with DAG(dag_id="hive_test_dag", schedule_interval=None, catchup=False,
 start_date=days_ago(1)) as dag:
    hive_test = BashOperator(
        task_id="hive_test",
        bash_command='hive --help'
    )
```

Airflow configuration options

If you're using Apache Airflow v2, add `core.lazy_load_plugins : False` as an Apache Airflow configuration option. To learn more, refer to Using configuration options to load plugins in 2.

What's next?

- Learn how to upload the `requirements.txt` file in this example to your Amazon S3 bucket in Installing Python dependencies.

- Learn how to upload the DAG code in this example to the `dags` folder in your Amazon S3 bucket in Adding or updating DAGs.
- Learn more about how to upload the `plugins.zip` file in this example to your Amazon S3 bucket in Installing custom plugins.

Invoking DAGs with a Lambda function

The following code example uses an AWS Lambda function to get an Apache Airflow CLI token and invoke a directed acyclic graph (DAG) in an Amazon MWAA environment.

Topics

- Version
- Prerequisites
- Permissions
- Dependencies
- Code example

Version

You can use the code example on this page with **Apache Airflow v2** in Python 3.10 and **Apache Airflow v3** in Python 3.11.

Prerequisites

To use this code example, you must:

- Use the public network access mode for your Amazon MWAA environment.
- Have a Lambda function using the latest Python runtime.

> **Note**
>
> If the Lambda function and your Amazon MWAA environment are in the same VPC, you can use this code on a private network. For this configuration, the Lambda function's execution role needs permission to call the Amazon Elastic Compute Cloud (Amazon

EC2) **CreateNetworkInterface** API operation. You can provide this permission using the `AWSLambdaVPCAccessExecutionRole` AWS-managed policy.

Permissions

To use the code example on this page, your Amazon MWAA environment's execution role needs access to perform the `airflow:CreateCliToken` action. You can provide this permission using the `AmazonMWAAAirflowCliAccess` AWS-managed policy:

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "airflow:CreateCliToken"
            ],
            "Resource": "*"
        }
    ]
}
```

For more information, refer to Apache Airflow CLI policy: AmazonMWAAAirflowCliAccess.

Dependencies

To use this code example with Apache Airflow v2 and later, no additional dependencies are required. Use aws-mwaa-docker-images to install Apache Airflow.

Code example

1. Open the AWS Lambda console at https://console.aws.amazon.com/lambda/.
2. Choose your Lambda function from the **Functions** list.
3. On the function page, copy the following code and replace the following with the names of your resources:

- YOUR_ENVIRONMENT_NAME – The name of your Amazon MWAA environment.
- YOUR_DAG_NAME – The name of the DAG that you want to invoke.

```
import boto3
import http.client
import base64
import ast
mwaa_env_name = 'YOUR_ENVIRONMENT_NAME'
dag_name = 'YOUR_DAG_NAME'
mwaa_cli_command = 'dags trigger'

client = boto3.client('mwaa')

def lambda_handler(event, context):
    # get web token
    mwaa_cli_token = client.create_cli_token(
        Name=mwaa_env_name
    )

    conn = http.client.HTTPSConnection(mwaa_cli_token['WebServerHostname'])
    payload = mwaa_cli_command + " " + dag_name
    headers = {
      'Authorization': 'Bearer ' + mwaa_cli_token['CliToken'],
      'Content-Type': 'text/plain'
    }
    conn.request("POST", "/aws_mwaa/cli", payload, headers)
    res = conn.getresponse()
    data = res.read()
    dict_str = data.decode("UTF-8")
    mydata = ast.literal_eval(dict_str)
    return base64.b64decode(mydata['stdout'])
```

4. Choose **Deploy**.
5. Choose **Test** to invoke your function using the Lambda console.
6. To verify that your Lambda successfully invoked your DAG, use the Amazon MWAA console to navigate to your environment's Apache Airflow UI, then do the following:

    a. On the **DAGs** page, locate your new target DAG in the list of DAGs.

    b. Under **Last Run**, check the timestamp for the latest DAG run. This timestamp should closely match the latest timestamp for `invoke_dag` in your other environment.

c. Under **Recent Tasks**, check that the last run was successful.

Invoking DAGs in different Amazon MWAA environments

The following code example creates an Apache Airflow CLI token. The code then uses a directed acyclic graph (DAG) in one Amazon MWAA environment to invoke a DAG in a different Amazon MWAA environment.

Topics

- Version
- Prerequisites
- Permissions
- Dependencies
- Code example

Version

You can use the code example on this page with **Apache Airflow v2** in Python 3.10 and **Apache Airflow v3** in Python 3.11.

Prerequisites

To use the code example on this page, you need the following:

- Two Amazon MWAA environments with **public network** webserver access, including your current environment.
- A sample DAG uploaded to your target environment's Amazon Simple Storage Service (Amazon S3) bucket.

Permissions

To use the code example on this page, your environment's execution role must have permission to create an Apache Airflow CLI token. You can attach the AWS-managed policy `AmazonMWAAAirflowCliAccess` to grant this permission.

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": [
                "airflow:CreateCliToken"
            ],
            "Resource": "*"
        }
    ]
}
```

For more information, refer to Apache Airflow CLI policy: AmazonMWAAAirflowCliAccess.

Dependencies

To use this code example with Apache Airflow v2 and later, no additional dependencies are required. Use aws-mwaa-docker-images to install Apache Airflow.

Code example

The following code example assumes that you're using a DAG in your current environment to invoke a DAG in another environment.

1. In your terminal, navigate to the directory where your DAG code is stored. For example:

```
cd dags
```

2. Copy the content of the following code example and save it locally as `invoke_dag.py`. Replace the following values with your information.

   - `your-new-environment-name` – The name of the other environment where you want to invoke the DAG.
   - `your-target-dag-id` – The ID of the DAG in the other environment that you want to invoke.

```
from airflow.decorators import dag, task
import boto3
from datetime import datetime, timedelta
import os, requests

DAG_ID = os.path.basename(__file__).replace(".py", "")

@task()
def invoke_dag_task(**kwargs):
    client = boto3.client('mwaa')
    token = client.create_cli_token(Name='your-new-environment-name')
    url = f"https://{token['WebServerHostname']}/aws_mwaa/cli"
    body = 'dags trigger your-target-dag-id'
    headers = {
        'Authorization' : 'Bearer ' + token['CliToken'],
        'Content-Type': 'text/plain'
        }
    requests.post(url, data=body, headers=headers)

@dag(
    dag_id=DAG_ID,
    schedule_interval=None,
    start_date=datetime(2022, 1, 1),
    dagrun_timeout=timedelta(minutes=60),
    catchup=False
    )
def invoke_dag():
    t = invoke_dag_task()

invoke_dag_test = invoke_dag()
```

3. Run the following AWS CLI command to copy the DAG to your environment's bucket, then trigger the DAG using the Apache Airflow UI.

```
aws s3 cp your-dag.py s3://your-environment-bucket/dags/
```

4. If the DAG runs successfully, you'll get output similar to the following in the task logs for `invoke_dag_task`.

```
[2022-01-01, 12:00:00 PDT] {{python.py:152}} INFO - Done. Returned value was: None
```

```
[2022-01-01, 12:00:00 PDT] {{taskinstance.py:1280}} INFO - Marking task as SUCCESS.
 dag_id=invoke_dag, task_id=invoke_dag_task, execution_date=20220101T120000,
 start_date=20220101T120000, end_date=20220101T120000
[2022-01-01, 12:00:00 PDT] {{local_task_job.py:154}} INFO - Task exited with return
 code 0
[2022-01-01, 12:00:00 PDT] {{local_task_job.py:264}} INFO - 0 downstream tasks
 scheduled from follow-on schedule check
```

To verify that your DAG was successfully invoked, navigate to the Apache Airflow UI for your new environment, then do the following:

a. On the **DAGs** page, locate your new target DAG in the list of DAGs.

b. Under **Last Run**, check the timestamp for the latest DAG run. This timestamp should closely match the latest timestamp for `invoke_dag` in your other environment.

c. Under **Recent Tasks**, check that the last run was successful.

Using Amazon MWAA with Amazon RDS for Microsoft SQL Server

You can use Amazon Managed Workflows for Apache Airflow to connect to an RDS for SQL Server. The following sample code uses DAGs on an Amazon Managed Workflows for Apache Airflow environment to connect to and execute queries on an Amazon RDS for Microsoft SQL Server.

Topics

- Version
- Prerequisites
- Dependencies
- Apache Airflow v2 connection
- Code sample
- What's next?

Version

You can use the code example on this page with **Apache Airflow v2** in Python 3.10 and **Apache Airflow v3** in Python 3.11.

Prerequisites

To use the sample code on this page, you'll need the following:

- An Amazon MWAA environment.
- Amazon MWAA and the RDS for SQL Server are running in the same Amazon VPC/
- VPC security groups of Amazon MWAA and the server are configured with the following connections:
  - An inbound rule for the port 1433 open for Amazon RDS in Amazon MWAA's security group
  - Or an outbound rule for the port of 1433 open from Amazon MWAA To RDS
- Apache Airflow Connection for RDS for SQL Server reflects the hostname, port, username and password from the Amazon RDS SQL server database created in previous process.

Dependencies

To use the sample code in this section, add the following dependency to your `requirements.txt`. To learn more, refer to Installing Python dependencies.

```
apache-airflow-providers-microsoft-mssql==1.0.1
apache-airflow-providers-odbc==1.0.1
pymssql==2.2.1
```

Apache Airflow v2 connection

If you're using a connection in Apache Airflow v2, ensure the Airflow connection object includes the following key-value pairs:

1. **Conn Id:** mssql_default
2. **Conn Type:** Amazon Web Services
3. **Host:** YOUR_DB_HOST
4. **Schema:**
5. **Login:** admin
6. **Password:**
7. **Port:** 1433
8. **Extra:**

Code sample

1. In your command prompt, navigate to the directory where your DAG code is stored. For example:

```
cd dags
```

2. Copy the contents of the following code sample and save locally as `sql-server.py`.

```
"""
Copyright Amazon.com, Inc. or its affiliates. All Rights Reserved.
Permission is hereby granted, free of charge, to any person obtaining a copy of
this software and associated documentation files (the "Software"), to deal in
the Software without restriction, including without limitation the rights to
use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of
the Software, and to permit persons to whom the Software is furnished to do so.
THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR
IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS
FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR
COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER
IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN
CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.
"""
import pymssql
import logging
import sys
from airflow import DAG
from datetime import datetime
from airflow.operators.mssql_operator import MsSqlOperator
from airflow.operators.python_operator import PythonOperator

default_args = {
    'owner': 'aws',
    'depends_on_past': False,
    'start_date': datetime(2019, 2, 20),
    'provide_context': True
}

dag = DAG(
    'mssql_conn_example', default_args=default_args, schedule_interval=None)

drop_db = MsSqlOperator(
   task_id="drop_db",
```

```
   sql="DROP DATABASE IF EXISTS testdb;",
   mssql_conn_id="mssql_default",
   autocommit=True,
   dag=dag
)

create_db = MsSqlOperator(
   task_id="create_db",
   sql="create database testdb;",
   mssql_conn_id="mssql_default",
   autocommit=True,
   dag=dag
)

create_table = MsSqlOperator(
   task_id="create_table",
   sql="CREATE TABLE testdb.dbo.pet (name VARCHAR(20), owner VARCHAR(20));",
   mssql_conn_id="mssql_default",
   autocommit=True,
   dag=dag
)

insert_into_table = MsSqlOperator(
   task_id="insert_into_table",
   sql="INSERT INTO testdb.dbo.pet VALUES ('Olaf', 'Disney');",
   mssql_conn_id="mssql_default",
   autocommit=True,
   dag=dag
)

def select_pet(**kwargs):
   try:
        conn = pymssql.connect(
            server='sampledb.<xxxxxx>.<region>.rds.amazonaws.com',
            user='admin',
            password='<yoursupersecretpassword>',
            database='testdb'
        )

        # Create a cursor from the connection
        cursor = conn.cursor()
        cursor.execute("SELECT * from testdb.dbo.pet")
        row = cursor.fetchone()
```

```
        if row:
            print(row)
   except:
      logging.error("Error when creating pymssql database connection: %s",
 sys.exc_info()[0])

select_query = PythonOperator(
    task_id='select_query',
    python_callable=select_pet,
    dag=dag,
)

drop_db >> create_db >> create_table >> insert_into_table >> select_query
```

What's next?

- Learn how to upload the `requirements.txt` file in this example to your Amazon S3 bucket in Installing Python dependencies.
- Learn how to upload the DAG code in this example to the `dags` folder in your Amazon S3 bucket in Adding or updating DAGs.
- Explore example scripts and other pymssql module examples.
- Learn more about executing SQL code in a specific Microsoft SQL database using the mssql_operator in the *Apache Airflow reference guide*.

Using Amazon MWAA with Amazon EKS

The following sample demonstrates how to use Amazon Managed Workflows for Apache Airflow with Amazon EKS.

Topics

- Version
- Prerequisites
- Create a public key for Amazon EC2
- Create the cluster
- Create a mwaa namespace
- Create a role for the mwaa namespace

- Create and attach an IAM role for the Amazon EKS cluster
- Create the requirements.txt file
- Grant the Amazon MWAA execution role access to the cluster
- Create the kubeconfig
- Create a DAG
- Add the DAG and kube_config.yaml to the Amazon S3 bucket
- Enable and trigger the example

Version

You can use the code example on this page with **Apache Airflow v2** in Python 3.10 and **Apache Airflow v3** in Python 3.11.

Prerequisites

To use the example in this topic, you'll need the following:

- An Amazon MWAA environment.
- eksctl. To learn more, refer to Install eksctl.
- kubectl. To learn more, refer to Install and Set Up kubectl. In some case this is installed with eksctl.
- An EC2 key pair in the Region where you create your Amazon MWAA environment. To learn more, refer to Creating or importing a key pair.

> **Note**
>
> When you use an `eksctl` command, you can include a `--profile` to specify a profile other than the default.

Create a public key for Amazon EC2

Use the following command to create a public key from your private key pair.

```
ssh-keygen -y -f myprivatekey.pem > mypublickey.pub
```

To learn more, refer to Retrieving the public key for your key pair.

Create the cluster

Use the following command to create the cluster. If you want a custom name for the cluster or to create it in a different Region, replace the name and Region values. You must create the cluster in the same Region where you create the Amazon MWAA environment. Replace the values for the subnets to match the subnets in your Amazon VPC network that you use for Amazon MWAA. Replace the value for the `ssh-public-key` to match the key you use. You can use an existing key from Amazon EC2 that is in the same Region, or create a new key in the same Region where you create your Amazon MWAA environment.

```
eksctl create cluster \
--name mwaa-eks \
--region us-west-2 \
--version 1.30 \
--nodegroup-name linux-nodes \
--nodes 3 \
--nodes-min 1 \
--nodes-max 4 \
--with-oidc \
--ssh-access \
--ssh-public-key MyPublicKey \
--managed \
--vpc-public-subnets "subnet-11111111111111111, subnet-2222222222222222" \
--vpc-private-subnets "subnet-33333333333333333, subnet-44444444444444444"
```

It takes some time to complete creating the cluster. Once complete, you can verify that the cluster was created successfully and has the IAM OIDC Provider configured by using the following command:

```
eksctl utils associate-iam-oidc-provider \
--region us-west-2 \
--cluster mwaa-eks \
--approve
```

EKS cluster authentication mode

Amazon EKS supports three cluster authentication modes: `API`, `API_AND_CONFIG_MAP`, and `CONFIG_MAP`. The mode determines how you grant the Amazon MWAA execution role access to the cluster:

- `API` (default for clusters created with recent tooling) – grant access by creating an Amazon EKS access entry. Use the steps in the section called "Option A—API mode: create an EKS access entry".
- `CONFIG_MAP` or `API_AND_CONFIG_MAP` – grant access through the `aws-auth` ConfigMap. Use the steps in the section called "Option B—ConfigMap mode: create an identity mapping".

To check your cluster's current mode:

```
aws eks describe-cluster --name mwaa-eks --region us-west-2 \
  --query "cluster.accessConfig.authenticationMode"
```

If you create the cluster with the `eksctl` command shown here, check the mode afterward. If the mode is `API`, follow the access entry steps instead of the ConfigMap steps.

Create a `mwaa` namespace

After confirming that the cluster was successfully created, use the following command to create a namespace for the pods.

```
kubectl create namespace mwaa
```

Create a role for the `mwaa` namespace

After you create the namespace, create a role and role-binding for an Amazon MWAA user on EKS that can run pods in a the MWAA namespace. If you used a different name for the namespace, replace mwaa in `-n` *mwaa* with the name that you used.

```
cat << EOF | kubectl apply -f - -n mwaa
kind: Role
apiVersion: rbac.authorization.k8s.io/v1
```

```
metadata:
  name: mwaa-role
rules:
  - apiGroups:
      - ""
      - "apps"
      - "batch"
      - "extensions"
    resources:
      - "jobs"
      - "pods"
      - "pods/attach"
      - "pods/exec"
      - "pods/log"
      - "pods/portforward"
      - "secrets"
      - "services"
    verbs:
      - "create"
      - "delete"
      - "describe"
      - "get"
      - "list"
      - "patch"
      - "update"
---
kind: RoleBinding
apiVersion: rbac.authorization.k8s.io/v1
metadata:
  name: mwaa-role-binding
subjects:
  - kind: User
    name: mwaa-service
  - kind: Group
    name: mwaa-service
roleRef:
  kind: Role
  name: mwaa-role
  apiGroup: rbac.authorization.k8s.io
EOF
```

Confirm that the new role can access the Amazon EKS cluster by running the following command. Be sure to use the correct name if you did not use *mwaa*:

```
kubectl get pods -n mwaa --as mwaa-service
```

You get a message returned that says:

```
No resources found in mwaa namespace.
```

Create and attach an IAM role for the Amazon EKS cluster

You must create an IAM role and then bind it to the Amazon EKS (k8s) cluster so that it can be used for authentication through IAM. The role is used only to log in to the cluster, and does not have any permissions for the console or API calls.

> **Authentication and permissions**
>
> This role only authenticates to the cluster. In ConfigMap mode, the actual in-cluster permissions come from the Kubernetes Role and RoleBinding. In API mode, permissions come from the access entry's access scope. Make sure you use the same Kubernetes group name in both the RoleBinding and whichever access method you choose.

Create a new role for the Amazon MWAA environment using the steps in Amazon MWAA execution role. However, instead of creating and attaching the policies described in that topic, attach the following policy:

JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Effect": "Allow",
            "Action": "airflow:PublishMetrics",
            "Resource": "arn:aws:airflow:us-east-1:111122223333:environment/
${MWAA_ENV_NAME}"
        },
        {
            "Effect": "Deny",
            "Action": "s3:ListAllMyBuckets",
            "Resource": [
```

```
                "arn:aws:s3:::{MWAA_S3_BUCKET}",
                "arn:aws:s3:::{MWAA_S3_BUCKET}/*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "s3:GetObject*",
                "s3:GetBucket*",
                "s3:List*"
            ],
            "Resource": [
                "arn:aws:s3:::{MWAA_S3_BUCKET}",
                "arn:aws:s3:::{MWAA_S3_BUCKET}/*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": [
                "logs:CreateLogStream",
                "logs:CreateLogGroup",
                "logs:PutLogEvents",
                "logs:GetLogEvents",
                "logs:GetLogRecord",
                "logs:GetLogGroupFields",
                "logs:GetQueryResults",
                "logs:DescribeLogGroups"
            ],
            "Resource": [
            "arn:aws:logs:us-east-1:111122223333:log-group:airflow-
${MWAA_ENV_NAME}-*"
            ]
        },
        {
            "Effect": "Allow",
            "Action": "cloudwatch:PutMetricData",
            "Resource": "*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "sqs:ChangeMessageVisibility",
                "sqs:DeleteMessage",
                "sqs:GetQueueAttributes",
```

```
                "sqs:GetQueueUrl",
                "sqs:ReceiveMessage",
                "sqs:SendMessage"
            ],
            "Resource": "arn:aws:sqs:us-east-1:*:airflow-celery-*"
        },
        {
            "Effect": "Allow",
            "Action": [
                "kms:Decrypt",
                "kms:DescribeKey",
                "kms:GenerateDataKey*",
                "kms:Encrypt"
            ],
            "NotResource": "arn:aws:kms:*:111122223333:key/*",
            "Condition": {
                "StringLike": {
                    "kms:ViaService": [
                    "sqs.us-east-1.amazonaws.com"
                    ]
                }
            }
        },
        {
            "Effect": "Allow",
            "Action": [
                "eks:DescribeCluster"
            ],
            "Resource": "arn:aws:eks:us-east-1:111122223333:cluster/
${EKS_CLUSTER_NAME}"
        }
    ]
}
```

After you create role, edit your Amazon MWAA environment to use the role you created as the execution role for the environment. To change the role, edit the environment to use. You select the execution role under **Permissions**.

Known issues:

- There is a known issue with role ARNs with subpaths not being able to authenticate with Amazon EKS. The workaround for this is to create the service role manually rather than using the one

created by Amazon MWAA itself. To learn more, refer to Roles with paths do not work when the path is included in their ARN in the aws-auth configmap

- If Amazon MWAA service listing is not available in IAM you need to choose an alternate service policy, such as Amazon EC2, and then update the role's trust policy to match the following:

  JSON

```
{
  "Version":"2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Principal": {
        "Service": [
          "airflow-env.amazonaws.com",
          "airflow.amazonaws.com"
        ]
      },
      "Action": "sts:AssumeRole"
    }
  ]
}
```

  To learn more, refer to How to use trust policies with IAM roles.

Create the requirements.txt file

To use the sample code in this section, ensure you've added one of the following database options to your `requirements.txt`. To learn more, refer to Installing Python dependencies.

```
kubernetes
apache-airflow-providers-cncf-kubernetes
```

Grant the Amazon MWAA execution role access to the cluster

The way you grant the Amazon MWAA execution role access to the cluster depends on the cluster's authentication mode (see the note in the section called "Create the cluster"). Follow the path that matches your cluster.

Option A—API mode: create an EKS access entry

If your cluster uses the `API` or `API_AND_CONFIG_MAP` authentication mode, create an access entry for the Amazon MWAA execution role. Replace the AWS Region, cluster name, and role ARN with your values.

```
aws eks create-access-entry \
--region us-west-2 \
--cluster-name mwaa-eks \
--principal-arn arn:aws:iam::123456789012:role/mwaa-execution-role \
--kubernetes-groups mwaa-service \
--username mwaa-service
```

The `mwaa-service` group must match the `subjects` group referenced by the RoleBinding you created in the section called “Create a role for the mwaa namespace”. This binds the execution role to the `mwaa-role` permissions in the `mwaa` namespace.

To confirm that you created the access entry:

```
aws eks list-access-entries --region us-west-2 --cluster-name mwaa-eks
```

Option B—ConfigMap mode: create an identity mapping

If your cluster uses the `CONFIG_MAP` or `API_AND_CONFIG_MAP` authentication mode, create an identity mapping. This writes an entry to the cluster's `aws-auth` ConfigMap. This step has no effect on clusters using the `API` authentication mode.

```
eksctl create iamidentitymapping \
--region us-west-2 \
--cluster mwaa-eks \
--arn arn:aws:iam::123456789012:role/mwaa-execution-role \
--username mwaa-service \
--group mwaa-service
```

The `--group mwaa-service` value must match the Group subject in the RoleBinding you created in the section called “Create a role for the mwaa namespace”. This binds the execution role to the `mwaa-role` permissions in the `mwaa` namespace, consistent with the access entry in Option A.

Create the kubeconfig

Use the following command to create the kubeconfig:

```
aws eks update-kubeconfig \
--region us-west-2 \
--kubeconfig ./kube_config.yaml \
--name mwaa-eks \
--alias aws
```

If you used a specific profile when you ran update-kubeconfig you need to remove the env: section added to the kube_config.yaml file so that it works correctly with Amazon MWAA. To do so, delete the following from the file and then save it:

```
env:
 - name: AWS_PROFILE
 value: profile_name
```

Create a DAG

Use the following code example to create a Python file, such as mwaa_pod_example.py for the DAG.

```
"""
Copyright Amazon.com, Inc. or its affiliates. All Rights Reserved.
Permission is hereby granted, free of charge, to any person obtaining a copy of
this software and associated documentation files (the "Software"), to deal in
the Software without restriction, including without limitation the rights to
use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of
the Software, and to permit persons to whom the Software is furnished to do so.
THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR
IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS
FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR
COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER
IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN
CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.
"""
from airflow import DAG
from datetime import datetime
from airflow.providers.cncf.kubernetes.operators.kubernetes_pod import
 KubernetesPodOperator
```

```
default_args = {
    'owner': 'aws',
    'depends_on_past': False,
    'start_date': datetime(2019, 2, 20),
    'provide_context': True
}

dag = DAG(
    'kubernetes_pod_example', default_args=default_args, schedule_interval=None)

#use a kube_config stored in s3 dags folder for now
kube_config_path = '/usr/local/airflow/dags/kube_config.yaml'

podRun = KubernetesPodOperator(
    namespace="mwaa",
    image="ubuntu:18.04",
    cmds=["bash"],
    arguments=["-c", "ls"],
    labels={"foo": "bar"},
    name="mwaa-pod-test",
    task_id="pod-task",
    get_logs=True,
    dag=dag,
    is_delete_operator_pod=False,
    config_file=kube_config_path,
    in_cluster=False,
    cluster_context='aws'
)
```

Add the DAG and `kube_config.yaml` to the Amazon S3 bucket

Put the DAG you created and the `kube_config.yaml` file into the Amazon S3 bucket for the Amazon MWAA environment. You can put files into your bucket using either the Amazon S3 console or the AWS Command Line Interface.

Enable and trigger the example

In Apache Airflow, enable the example and then trigger it.

After it runs and completes successfully, use the following command to verify the pod:

```
kubectl get pods -n mwaa
```

You get output similar to the following:

```
NAME READY STATUS RESTARTS AGE
mwaa-pod-test-aa11bb22cc3344445555666677778888 0/1 Completed 0 2m23s
```

You can then verify the output of the pod with the following command. Replace the name value with the value returned from the previous command:

```
kubectl logs -n mwaa mwaa-pod-test-aa11bb22cc3344445555666677778888
```

Connecting to Amazon ECS using the ECSOperator

This topic describes how you can use the ECSOperator to connect to an Amazon Elastic Container Service (Amazon ECS) container from Amazon MWAA. In the following steps, you'll add the required permissions to your environment's execution role, use a CloudFormation template to create an Amazon ECS Fargate cluster, and finally create and upload a DAG that connects to your new cluster.

Topics

- Version
- Prerequisites
- Permissions
- Create an Amazon ECS cluster
- Code sample

Version

You can use the code example on this page with **Apache Airflow v2** in Python 3.10 and **Apache Airflow v3** in Python 3.11.

Prerequisites

To use the sample code on this page, you'll need the following:

- An Amazon MWAA environment.

Permissions

- The execution role for your environment needs permission to run tasks in Amazon ECS. You can either attach the AmazonECS_FullAccess AWS-managed policy to your execution role, or create and attach the following policy to your execution role.

  JSON

```
{
    "Version":"2012-10-17",
    "Statement": [
        {
            "Sid": "VisualEditor0",
            "Effect": "Allow",
            "Action": [
                "ecs:RunTask",
                "ecs:DescribeTasks"
            ],
            "Resource": "*"
        },
        {
            "Action": "iam:PassRole",
            "Effect": "Allow",
            "Resource": [
                "*"
            ],
            "Condition": {
                "StringLike": {
                    "iam:PassedToService": "ecs-tasks.amazonaws.com"
                }
            }
        }
    ]
}
```

- In addition to adding the required premissions to run tasks in Amazon ECS, you must also modify the CloudWatch Logs policy statement in your Amazon MWAA execution role to allow access to the Amazon ECS task log group as listed in the following. The Amazon ECS log group is created by the CloudFormation template in the section called “Create an Amazon ECS cluster”.

```
{
  "Effect": "Allow",
  "Action": [
    "logs:CreateLogStream",
    "logs:CreateLogGroup",
    "logs:PutLogEvents",
    "logs:GetLogEvents",
    "logs:GetLogRecord",
    "logs:GetLogGroupFields",
    "logs:GetQueryResults"
  ],
  "Resource": [
    "arn:aws:logs:us-east-1:123456789012:log-group:airflow-environment-name-*",
    "arn:aws:logs:*:*:log-group:ecs-mwaa-group:*"
  ]
}
```

For more information about the Amazon MWAA execution role, and how to attach a policy, refer to Execution role.

Create an Amazon ECS cluster

Using the following CloudFormation template, you will build an Amazon ECS Fargate cluster to use with your Amazon MWAA workflow. For more information, refer to Creating a task definition in the *Amazon Elastic Container Service Developer Guide*.

1. Create a JSON file with the following code and save it as `ecs-mwaa-cfn.json`.

```
{
    "AWSTemplateFormatVersion": "2010-09-09",
    "Description": "This template deploys an ECS Fargate cluster with an Amazon
 Linux image as a test for MWAA.",
    "Parameters": {
        "VpcId": {
            "Type": "AWS::EC2::VPC::Id",
            "Description": "Select a VPC that allows instances access to ECR, as
 used with MWAA."
        },
        "SubnetIds": {
            "Type": "List<AWS::EC2::Subnet::Id>",
```

```
            "Description": "Select at two private subnets in your selected VPC, as
 used with MWAA."
        },
        "SecurityGroups": {
            "Type": "List<AWS::EC2::SecurityGroup::Id>",
            "Description": "Select at least one security group in your selected
 VPC, as used with MWAA."
        }
    },
    "Resources": {
        "Cluster": {
            "Type": "AWS::ECS::Cluster",
            "Properties": {
                "ClusterName": {
                    "Fn::Sub": "${AWS::StackName}-cluster"
                }
            }
        },
        "LogGroup": {
            "Type": "AWS::Logs::LogGroup",
            "Properties": {
                "LogGroupName": {
                    "Ref": "AWS::StackName"
                },
                "RetentionInDays": 30
            }
        },
        "ExecutionRole": {
            "Type": "AWS::IAM::Role",
            "Properties": {
                "AssumeRolePolicyDocument": {
                    "Statement": [
                        {
                            "Effect": "Allow",
                            "Principal": {
                                "Service": "ecs-tasks.amazonaws.com"
                            },
                            "Action": "sts:AssumeRole"
                        }
                    ]
                },
                "ManagedPolicyArns": [
                    "arn:aws:iam::aws:policy/service-role/
AmazonECSTaskExecutionRolePolicy"
```

```
                ]
            }
        },
        "TaskDefinition": {
            "Type": "AWS::ECS::TaskDefinition",
            "Properties": {
                "Family": {
                    "Fn::Sub": "${AWS::StackName}-task"
                },
                "Cpu": 2048,
                "Memory": 4096,
                "NetworkMode": "awsvpc",
                "ExecutionRoleArn": {
                    "Ref": "ExecutionRole"
                },
                "ContainerDefinitions": [
                    {
                        "Name": {
                            "Fn::Sub": "${AWS::StackName}-container"
                        },
                        "Image": "137112412989.dkr.ecr.us-east-1.amazonaws.com/
amazonlinux:latest",
                        "PortMappings": [
                            {
                                "Protocol": "tcp",
                                "ContainerPort": 8080,
                                "HostPort": 8080
                            }
                        ],
                        "LogConfiguration": {
                            "LogDriver": "awslogs",
                            "Options": {
                                "awslogs-region": {
                                    "Ref": "AWS::Region"
                                },
                                "awslogs-group": {
                                    "Ref": "LogGroup"
                                },
                                "awslogs-stream-prefix": "ecs"
                            }
                        }
                    }
                ],
                "RequiresCompatibilities": [
```

```
                    "FARGATE"
                ]
            }
        },
        "Service": {
            "Type": "AWS::ECS::Service",
            "Properties": {
                "ServiceName": {
                    "Fn::Sub": "${AWS::StackName}-service"
                },
                "Cluster": {
                    "Ref": "Cluster"
                },
                "TaskDefinition": {
                    "Ref": "TaskDefinition"
                },
                "DesiredCount": 1,
                "LaunchType": "FARGATE",
                "PlatformVersion": "1.3.0",
                "NetworkConfiguration": {
                    "AwsvpcConfiguration": {
                        "AssignPublicIp": "ENABLED",
                        "Subnets": {
                            "Ref": "SubnetIds"
                        },
                        "SecurityGroups": {
                            "Ref": "SecurityGroups"
                        }
                    }
                }
            }
        }
    }
}
```

2. In your command prompt, use the following AWS CLI command to create a new stack. You must replace the values `SecurityGroups` and `SubnetIds` with values for your Amazon MWAA environment's security groups and subnets.

```
aws cloudformation create-stack \
--stack-name my-ecs-stack --template-body file://ecs-mwaa-cfn.json \
--parameters ParameterKey=SecurityGroups,ParameterValue=your-mwaa-security-group \
ParameterKey=SubnetIds,ParameterValue=your-mwaa-subnet-1\\,your-mwaa-subnet-1 \
```

```
--capabilities CAPABILITY_IAM
```

Alternatively, you can use the following shell script. The script retrieves the required values for your environment's security groups, and subnets using the `get-environment` AWS CLI command, then creates the stack accordingly. To run the script, do the following.

a. Copy, and save the script as `ecs-stack-helper.sh` in the same directory as your CloudFormation template.

```
#!/bin/bash

joinByString() {
  local separator="$1"
  shift
  local first="$1"
  shift
  printf "%s" "$first" "${@/#/$separator}"
}

response=$(aws mwaa get-environment --name $1)

securityGroupId=$(echo "$response" | jq -r
 '.Environment.NetworkConfiguration.SecurityGroupIds[]')
subnetIds=$(joinByString '\,' $(echo "$response" | jq -r
 '.Environment.NetworkConfiguration.SubnetIds[]'))

aws cloudformation create-stack --stack-name $2 --template-body file://ecs-
cfn.json \
--parameters ParameterKey=SecurityGroups,ParameterValue=$securityGroupId \
ParameterKey=SubnetIds,ParameterValue=$subnetIds \
--capabilities CAPABILITY_IAM
```

b. Run the script using the following commands. Replace `environment-name` and `stack-name` with your information.

```
chmod +x ecs-stack-helper.sh
./ecs-stack-helper.bash environment-name stack-name
```

If successful, you'll refer to the following output showing your new CloudFormation stack ID.

```
{
  "StackId": "arn:aws:cloudformation:us-west-2:123456789012:stack/my-ecs-
stack/123456e7-8ab9-01cd-b2fb-36cce63786c9"
}
```

After your CloudFormation stack is completed and AWS has provisioned your Amazon ECS resources, you're ready to create and upload your DAG.

Code sample

1. Open a command prompt, and navigate to the directory where your DAG code is stored. For example:

```
cd dags
```

2. Copy the contents of the following code sample and save locally as `mwaa-ecs-operator.py`, then upload your new DAG to Amazon S3.

   Apache Airflow v3

```
import boto3
import pendulum

from airflow.providers.amazon.aws.operators.ecs import EcsRunTaskOperator
from airflow.sdk import dag, task

CLUSTER_NAME="mwaa-ecs-test-cluster" #Replace value for CLUSTER_NAME with your
 information.
CONTAINER_NAME="mwaa-ecs-test-container" #Replace value for CONTAINER_NAME with
 your information.
LAUNCH_TYPE = "FARGATE"


@dag(
    schedule=None,
    start_date=pendulum.datetime(2024, 1, 1, tz="UTC"),
    catchup=False,
    tags=["ecs", "fargate"],
)
def ecs_fargate_dag():
```

```
    @task()
    def get_ecs_service_config() -> dict:
        """Fetch task definition and network config from the ECS service."""
        client = boto3.client("ecs")
        services = client.list_services(cluster=CLUSTER_NAME,
 launchType=LAUNCH_TYPE)
        service = client.describe_services(cluster=CLUSTER_NAME,
 services=services["serviceArns"])
        return {
            "task_definition": service["services"][0]["taskDefinition"],
            "network_configuration": service["services"][0]
["networkConfiguration"],
        }

    service_config = get_ecs_service_config()

    EcsRunTaskOperator(
        task_id="ecs_operator_task",
        cluster=CLUSTER_NAME,
        task_definition=service_config["task_definition"],
        launch_type=LAUNCH_TYPE,
        overrides={
            "containerOverrides": [
              {
                  "name": CONTAINER_NAME,
                  "command": ["ls", "-l", "/"],
              },
            ],
        },
        network_configuration=service_config["network_configuration"],
        awslogs_group="mwaa-ecs-zero",
        awslogs_stream_prefix=f"ecs/{CONTAINER_NAME}",
    )


ecs_fargate_dag()
```

Apache Airflow v2

```
from http import client
from airflow import DAG
from airflow.providers.amazon.aws.operators.ecs import EcsRunTaskOperator
from airflow.utils.dates import days_ago
```

```
import boto3

CLUSTER_NAME="mwaa-ecs-test-cluster" #Replace value for CLUSTER_NAME with your
 information.
CONTAINER_NAME="mwaa-ecs-test-container" #Replace value for CONTAINER_NAME with
 your information.
LAUNCH_TYPE="FARGATE"

with DAG(
    dag_id = "ecs_fargate_dag",
    schedule_interval=None,
    catchup=False,
    start_date=days_ago(1)
) as dag:
  client=boto3.client('ecs')
  services=client.list_services(cluster=CLUSTER_NAME,launchType=LAUNCH_TYPE)

 service=client.describe_services(cluster=CLUSTER_NAME,services=services['serviceArns'])

  ecs_operator_task = EcsRunTaskOperator(
    task_id = "ecs_operator_task",
    dag=dag,
    cluster=CLUSTER_NAME,
    task_definition=service['services'][0]['taskDefinition'],
    launch_type=LAUNCH_TYPE,
    overrides={
      "containerOverrides":[
        {
          "name":CONTAINER_NAME,
          "command":["ls", "-l", "/"],
        },
      ],
    },
    network_configuration=service['services'][0]['networkConfiguration'],
    awslogs_group="mwaa-ecs-zero",
    awslogs_stream_prefix=f"ecs/{CONTAINER_NAME}",
  )
```

Note

In the example DAG, for `awslogs_group`, you might need to modify the log group with the name for your Amazon ECS task log group. The example assumes a log group

named `mwaa-ecs-zero`. For `awslogs_stream_prefix`, use the Amazon ECS task log stream prefix. The example assumes a log stream prefix, `ecs`.

3. Run the following AWS CLI command to copy the DAG to your environment's bucket, then trigger the DAG using the Apache Airflow UI.

```
aws s3 cp your-dag.py s3://your-environment-bucket/dags/
```

4. If successful, you'll get output similar to the following in the task logs for `ecs_operator_task` in the `ecs_fargate_dag` DAG:

```
[2022-01-01, 12:00:00 UTC] {{ecs.py:300}} INFO - Running ECS Task -
Task definition: arn:aws:ecs:us-west-2:123456789012:task-definition/mwaa-ecs-test-
task:1 - on cluster mwaa-ecs-test-cluster
[2022-01-01, 12:00:00 UTC] {{ecs-operator-test.py:302}} INFO - ECSOperator
 overrides:
{'containerOverrides': [{'name': 'mwaa-ecs-test-container', 'command': ['ls', '-l',
 '/']}]}
.
.
.
[2022-01-01, 12:00:00 UTC] {{ecs.py:379}} INFO - ECS task ID is:
 e012340b5e1b43c6a757cf012c635935
[2022-01-01, 12:00:00 UTC] {{ecs.py:313}} INFO - Starting ECS Task Log Fetcher
[2022-01-01, 12:00:00 UTC] {{ecs.py:119}} INFO - [2022-07-19, 17:54:03 UTC] total
 52
[2022-01-01, 12:00:00 UTC] {{ecs.py:119}} INFO - [2022-07-19, 17:54:03 UTC]
 lrwxrwxrwx   1 root root    7 Jun 13 18:51 bin -> usr/bin
[2022-01-01, 12:00:00 UTC] {{ecs.py:119}} INFO - [2022-07-19, 17:54:03 UTC] dr-xr-
xr-x   2 root root 4096 Apr  9  2019 boot
[2022-01-01, 12:00:00 UTC] {{ecs.py:119}} INFO - [2022-07-19, 17:54:03 UTC] drwxr-
xr-x   5 root root  340 Jul 19 17:54 dev
[2022-01-01, 12:00:00 UTC] {{ecs.py:119}} INFO - [2022-07-19, 17:54:03 UTC] drwxr-
xr-x   1 root root 4096 Jul 19 17:54 etc
[2022-01-01, 12:00:00 UTC] {{ecs.py:119}} INFO - [2022-07-19, 17:54:03 UTC] drwxr-
xr-x   2 root root 4096 Apr  9  2019 home
[2022-01-01, 12:00:00 UTC] {{ecs.py:119}} INFO - [2022-07-19, 17:54:03 UTC]
 lrwxrwxrwx   1 root root    7 Jun 13 18:51 lib -> usr/lib
[2022-01-01, 12:00:00 UTC] {{ecs.py:119}} INFO - [2022-07-19, 17:54:03 UTC]
 lrwxrwxrwx   1 root root    9 Jun 13 18:51 lib64 -> usr/lib64
[2022-01-01, 12:00:00 UTC] {{ecs.py:119}} INFO - [2022-07-19, 17:54:03 UTC] drwxr-
xr-x   2 root root 4096 Jun 13 18:51 local
```

```
[2022-01-01, 12:00:00 UTC] {{ecs.py:119}} INFO - [2022-07-19, 17:54:03 UTC] drwxr-
xr-x   2 root root 4096 Apr  9  2019 media
[2022-01-01, 12:00:00 UTC] {{ecs.py:119}} INFO - [2022-07-19, 17:54:03 UTC] drwxr-
xr-x   2 root root 4096 Apr  9  2019 mnt
[2022-01-01, 12:00:00 UTC] {{ecs.py:119}} INFO - [2022-07-19, 17:54:03 UTC] drwxr-
xr-x   2 root root 4096 Apr  9  2019 opt
[2022-01-01, 12:00:00 UTC] {{ecs.py:119}} INFO - [2022-07-19, 17:54:03 UTC] dr-xr-
xr-x 103 root root    0 Jul 19 17:54 proc
[2022-01-01, 12:00:00 UTC] {{ecs.py:119}} INFO - [2022-07-19, 17:54:03 UTC] dr-xr-
x-\-\-   2 root root 4096 Apr  9  2019 root
[2022-01-01, 12:00:00 UTC] {{ecs.py:119}} INFO - [2022-07-19, 17:54:03 UTC] drwxr-
xr-x   2 root root 4096 Jun 13 18:52 run
[2022-01-01, 12:00:00 UTC] {{ecs.py:119}} INFO - [2022-07-19, 17:54:03 UTC]
 lrwxrwxrwx   1 root root    8 Jun 13 18:51 sbin -> usr/sbin
[2022-01-01, 12:00:00 UTC] {{ecs.py:119}} INFO - [2022-07-19, 17:54:03 UTC] drwxr-
xr-x   2 root root 4096 Apr  9  2019 srv
[2022-01-01, 12:00:00 UTC] {{ecs.py:119}} INFO - [2022-07-19, 17:54:03 UTC] dr-xr-
xr-x  13 root root    0 Jul 19 17:54 sys
[2022-01-01, 12:00:00 UTC] {{ecs.py:119}} INFO - [2022-07-19, 17:54:03 UTC]
 drwxrwxrwt   2 root root 4096 Jun 13 18:51 tmp
[2022-01-01, 12:00:00 UTC] {{ecs.py:119}} INFO - [2022-07-19, 17:54:03 UTC] drwxr-
xr-x  13 root root 4096 Jun 13 18:51 usr
[2022-01-01, 12:00:00 UTC] {{ecs.py:119}} INFO - [2022-07-19, 17:54:03 UTC] drwxr-
xr-x  18 root root 4096 Jun 13 18:52 var
.
.
.
[2022-01-01, 12:00:00 UTC] {{ecs.py:328}} INFO - ECS Task has been successfully
 executed
```

Using dbt with Amazon MWAA

With Amazon MWAA, you can use dbt (data build tool) and PostgreSQL to build and run data transformation workflows. In the following steps, add the required dependencies using a startup script, and upload a sample dbt project to your environment's Amazon S3 bucket. Then, use a sample DAG to verify that Amazon MWAA has installed the dependencies. Finally, use the `BashOperator` to run the dbt project.

Topics

- Version

- Prerequisites
- Dependencies
- Upload a dbt project to Amazon S3
- Use a DAG to verify dbt dependency installation
- Create and upload a dbt profiles.yml
- Use a DAG to run a dbt project

Version

You can use the code example on this page with **Apache Airflow v2** in Python 3.12 on the Python website and **Apache Airflow v3** in Python 3.12 on the Python website.

Prerequisites

Before you can complete the following steps, you need the following:

- An Amazon MWAA environment using Apache Airflow v2.11.2. This sample was written, and tested with v2.11.2. You might need to modify the sample to use with other Apache Airflow versions.
- A sample dbt project. To get started using dbt with Amazon MWAA, you can create a fork and clone the dbt starter project from the dbt-labs GitHub repository.

Dependencies

To use Amazon MWAA with dbt, add the following startup script to your environment. To learn more, refer to Using a startup script with Amazon MWAA.

```
#!/bin/bash

if [[ "${MWAA_AIRFLOW_COMPONENT}" != "worker" ]]
  then
    exit 0
fi

echo "------------------------------"
echo "Installing virtual Python env"
echo "------------------------------"
```

```
pip3 install --upgrade pip

echo "Current Python version:"
python3 --version
echo "..."

sudo pip3 install --user virtualenv
sudo mkdir -p /usr/local/airflow/python3-virtualenv
cd /usr/local/airflow/python3-virtualenv
sudo python3 -m venv dbt-env
sudo chmod -R 777 *

echo "------------------------------"
echo "Activating venv in $DBT_ENV_PATH"
echo "------------------------------"

source dbt-env/bin/activate
pip3 list

echo "------------------------------"
echo "Installing libraries..."
echo "------------------------------"

# do not use sudo, as it will install outside the venv
pip3 install dbt-core==1.9.4 dbt-redshift==1.9.1 dbt-postgres==1.9.0

echo "------------------------------"
echo "Venv libraries..."
echo "------------------------------"

pip3 list
dbt --version

echo "------------------------------"
echo "Deactivating venv..."
echo "------------------------------"

deactivate
```

Setting the DBT_ENV_PATH variable

You can set $DBT_ENV_PATH in the startup script or set it as an Airflow configuration in your Amazon MWAA environment.

In the following sections, upload your dbt project directory to Amazon S3 and run a DAG that validates whether Amazon MWAA has successfully installed the required dbt dependencies.

Upload a dbt project to Amazon S3

To be able to use a dbt project with your Amazon MWAA environment, you can upload the entire project directory to your environment's `dags` folder. When the environment updates, Amazon MWAA downloads the dbt directory to the local `usr/local/airflow/dags/` folder.

To upload a dbt project to Amazon S3

1. Navigate to the directory where you cloned the dbt starter project.
2. Run the following Amazon S3 AWS CLI command to recursively copy the content of the project to your environment's `dags` folder using the `--recursive` parameter. The command creates a sub-directory called `dbt` that you can use for all of your dbt projects. If the sub-directory already exists, the project files are copied into the existing directory, and a new directory is not created. The command also creates a sub-directory within the `dbt` directory for this specific starter project.

   ```
   aws s3 cp dbt-starter-project s3://amzn-s3-demo-bucket/dags/dbt/dbt-starter-project
    --recursive
   ```

   You can use different names for project sub-directories to organize multiple dbt projects within the parent `dbt` directory.

Use a DAG to verify dbt dependency installation

The following DAG uses a `BashOperator` and a bash command to verify whether Amazon MWAA has successfully installed the dbt dependencies specified in the startup script.

```
from airflow import DAG
from airflow.operators.bash_operator import BashOperator
```

```
from airflow.utils.dates import days_ago

with DAG(dag_id="dbt-installation-test", schedule_interval=None, catchup=False,
 start_date=days_ago(1)) as dag:
    cli_command = BashOperator(
        task_id="bash_command",
        bash_command="/usr/local/airflow/python3-virtualenv/dbt-env/bin/dbt --version"
    )
```

Do the following to access task logs and verify that dbt and its dependencies have been installed.

1. Navigate to the Amazon MWAA console, then choose **Open Airflow UI** from the list of available environments.
2. On the Apache Airflow UI, find the `dbt-installation-test` DAG from the list, then choose the date in the `Last Run` column to open the last successful task.
3. Using **Graph View**, choose the `bash_command` task to open the task instance details.
4. Choose **Log** to open the task logs, then verify that the logs successfully list the dbt version specified in the startup script.

Create and upload a dbt profiles.yml

To connect to your target database, dbt requires a `profiles.yml` file. The DAG in the next section passes `--profiles-dir /tmp/dbt`, so dbt looks for `profiles.yml` directly inside the `/tmp/dbt` directory. This is the `dbt` folder you uploaded to Amazon S3.

The profile name in `profiles.yml` must match the `profile:` value defined in the starter project's `dbt_project.yml`. The dbt starter project uses `default`.

To create and upload a profiles.yml

1. In the directory where you cloned the starter project, create a file named `profiles.yml` with your database connection details.

```
default:
  target: dev
  outputs:
    dev:
      type: postgres
      host: your-db-endpoint.region.rds.amazonaws.com
      port: 5432
```

```
      user: your_db_user
      password: your_db_password
      dbname: your_database
      schema: your_schema
      threads: 4
```

2. Upload the file to the dbt sub-directory in your environment's DAGs folder.

```
aws s3 cp profiles.yml s3://amzn-s3-demo-bucket/dags/dbt/profiles.yml
```

Protecting database credentials

To avoid storing plaintext credentials, reference secrets using the dbt `env_var()` function. Supply the values through Amazon MWAA environment variables or AWS Secrets Manager—for example, `password: "{{ env_var('DBT_PASSWORD') }}"`. Also make sure your Amazon MWAA VPC security groups allow your workers to reach your database on the configured port.

Use a DAG to run a dbt project

The following DAG uses a `BashOperator` to copy the dbt projects you uploaded to Amazon S3 from the local `usr/local/airflow/dags/` directory to the write-accessible `/tmp` directory, then runs the dbt project. The bash commands assume a starter dbt project titled `dbt-starter-project`. Modify the directory name according to the name of your project directory.

```
from airflow import DAG
from airflow.operators.bash_operator import BashOperator
from airflow.utils.dates import days_ago

import os

DAG_ID = os.path.basename(__file__).replace(".py", "")

# assumes all files are in a subfolder of DAGs called dbt

with DAG(dag_id=DAG_ID, schedule_interval=None, catchup=False, start_date=days_ago(1))
 as dag:
    cli_command = BashOperator(
        task_id="bash_command",
```

```
        bash_command="source /usr/local/airflow/python3-virtualenv/dbt-env/bin/
activate;\
        cp -R /usr/local/airflow/dags/dbt /tmp;\
        echo 'listing project files:';\
        ls -R /tmp;\
        cd /tmp/dbt/dbt-starter-project;\
        /usr/local/airflow/python3-virtualenv/dbt-env/bin/dbt run --project-dir /tmp/
dbt/dbt-starter-project --profiles-dir /tmp/dbt;\
        cat /tmp/dbt/dbt-starter-project/logs/dbt.log;\
        rm -rf /tmp/dbt/dbt-starter-project"
    )
```

Using DuckDB with Amazon MWAA

Amazon Managed Workflows for Apache Airflow (Amazon MWAA) environments that use Apache Airflow v3.3.1 and later include native DuckDB support. With DuckDB, the tasks in your directed acyclic graphs (DAGs) can run SQL without installing packages, creating connections, or managing credentials. DuckDB is an in-process SQL engine that runs inside an Apache Airflow task on your environment's workers, so you don't need a separate database server or cluster. Use DuckDB for checks, summaries, and transformations on data that your pipelines produce, such as CSV, JSON, and Parquet files in Amazon S3.

This topic describes how DuckDB works in an Amazon MWAA environment. The following code examples show how to write DAGs that run DuckDB queries in memory, and that read and write data in Amazon S3. Each code example includes its own prerequisites, permissions, and cleanup steps.

Topics

- Version
- Prerequisites
- How DuckDB connects in Amazon MWAA
- Using DuckDB with Apache Airflow v2.11 through v3.2
- Customize DuckDB in your environment
- Troubleshooting DuckDB tasks
- Next steps
- Related resources
- Using a DAG to run DuckDB queries in memory

- Using a DAG to read and write Amazon S3 data with DuckDB

Version

Native DuckDB support is available in Amazon MWAA environments that use **Apache Airflow v3.3.1** and later. Earlier Apache Airflow versions don't include the DuckDB provider. To use DuckDB with an earlier version, see Using DuckDB with Apache Airflow v2.11 through v3.2. You can use the code examples on this page with Apache Airflow v3.3.1 in Python 3.12.

If your Apache Airflow v3.3.1 environment was created before DuckDB support was available, update your environment to apply the latest Amazon MWAA image. Don't change any environment settings. The update can take some time to complete.

Prerequisites

To use the code examples for DuckDB, you need an Amazon MWAA environment using Apache Airflow v3.3.1 or later. Each example lists any additional prerequisites, such as Amazon S3 permissions.

How DuckDB connects in Amazon MWAA

Amazon MWAA environments that use Apache Airflow v3.3.1 or later include DuckDB by default, so you don't need to add packages to your `requirements.txt` file. The current Apache Airflow v3.3.1 image includes DuckDB version 2.0.0, version 0.2.0 of the Apache Airflow DuckDB provider, and version 9.38.0 of the Apache Airflow Amazon provider (`apache-airflow-providers-amazon`). The Amazon provider adds the `duckdb_aws` connection type that the `duckdb_default` connection uses. This connection type requires version 9.38.0 or later of the Amazon provider. To check the DuckDB version that your tasks use, run `SELECT version();` in a DuckDB task.

When a DuckDB task doesn't name a connection, the DuckDB query operator, `DuckDBExecuteQueryOperator`, uses the `duckdb_default` connection. Amazon MWAA creates this connection for you when you create or update your environment. With this connection, DuckDB does the following:

- Reads and writes Amazon S3 data as your environment's execution role, the same IAM role that your other tasks use. You don't need to add access keys to Apache Airflow or to your SQL.
- Uses the Amazon S3 support that's included in the Amazon MWAA image, so your environment doesn't need internet access to download it. This applies only to the default DuckDB version. If

you install a different DuckDB version, DuckDB downloads its Amazon S3 support the first time a task uses it, even if your tasks use the `duckdb_default` connection. For more information, see Customize DuckDB in your environment.

- Runs in memory, so each task starts with an empty database. To share data between tasks, write it to Amazon S3.

The `duckdb_default` connection appears in the Apache Airflow UI under **Admin**, **Connections**. You can edit or delete it.

Amazon MWAA doesn't overwrite an existing `duckdb_default` connection. If you delete the connection, Amazon MWAA creates it again the next time you update your environment. For more information about connections, see Managing connections to Apache Airflow.

Using DuckDB with Apache Airflow v2.11 through v3.2

Amazon MWAA environments that use Apache Airflow v3.2 or earlier don't include DuckDB, and Amazon MWAA doesn't create the `duckdb_default` connection for them. You can add DuckDB to environments that use Apache Airflow v2.11 through v3.2. The Apache Airflow DuckDB provider requires Apache Airflow v2.11 or later, so you can't use it with Apache Airflow v2.10 or earlier. To use DuckDB in an environment that uses Apache Airflow v2.11 through v3.2, do the following:

1. Add `apache-airflow-providers-duckdb` and `duckdb` to your `requirements.txt` file. Choose versions that support your Apache Airflow version. For more information about `requirements.txt`, see Installing Python dependencies.
2. To read and write Amazon S3 data, set up Amazon S3 access in your SQL before your queries. The following statements load the DuckDB extensions for Amazon S3 and use your environment's execution role for credentials.

```
INSTALL httpfs;
LOAD httpfs;
INSTALL aws;
LOAD aws;
CREATE SECRET (TYPE s3, PROVIDER credential_chain);
```

DuckDB downloads these extensions from `extensions.duckdb.org` the first time a task uses them. If your environment doesn't have internet access, provide the extensions yourself, as described in Customize DuckDB in your environment.

Customize DuckDB in your environment

You can change how DuckDB works in your environment in the following ways:

- **Use a different AWS identity for some tasks** – Create a connection with the `duckdb_aws` connection type, and pass its ID to the operator as `conn_id`. In the connection's `extra` field, set `aws_conn_id` to an AWS connection, such as one that assumes a role in another account.
- **Change the default connection** – Edit `duckdb_default` in the Apache Airflow UI. For example, set DuckDB options such as `memory_limit` or `threads` in the connection's `extra` field.
- **Use a different DuckDB or provider version** – Add the version that you want to your `requirements.txt` file. If you change the version of the Amazon provider, `apache-airflow-providers-amazon`, use version 9.38.0 or later. Earlier versions don't include the `duckdb_aws` connection type that the `duckdb_default` connection uses. If your environment doesn't have internet access, also provide the matching DuckDB extensions. DuckDB looks for extensions in a folder for each DuckDB version and platform, `~/.duckdb/extensions/`*`version`*`/`*`platform`*`/`. For example, for DuckDB 1.5.6, use a startup script to copy `httpfs.duckdb_extension` and `aws.duckdb_extension` from Amazon S3 to `~/.duckdb/extensions/v1.5.6/linux_amd64/`. For more information about `requirements.txt`, see Installing Python dependencies.

> **Note**
>
> The Amazon MWAA image includes DuckDB extensions that match the default DuckDB version. If you install a different version, DuckDB downloads matching extensions from `extensions.duckdb.org` the first time a task uses them. This download might add time to the first task run.

Troubleshooting DuckDB tasks

If a DuckDB task fails, check the task's log in the Apache Airflow UI for missing permissions, incorrect paths, missing extensions, or memory limits. The following entries describe how to resolve each of these common causes, so that your DAGs can complete successfully.

Resolve HTTP 403 or access denied errors from Amazon S3

Verify that your environment's execution role has the permissions in Permissions for the Amazon S3 code example for the bucket and prefix in your SQL. If the bucket uses AWS KMS keys for encryption, also verify the AWS KMS permissions.

If the role has the required permissions, verify that the task uses a connection with the `duckdb_aws` connection type. A connection with the `duckdb` connection type doesn't create a secret for Amazon S3 access, so Amazon S3 denies the request. Check the connection that you pass as `conn_id`, or the `duckdb_default` connection if the task doesn't specify one.

The `duckdb_aws` connection type requires version 9.38.0 or later of the Amazon provider, `apache-airflow-providers-amazon`. If your `requirements.txt` file installs an earlier version, remove that requirement or change it to version 9.38.0 or later.

Fix wildcard paths that find no files

Verify that the execution role has `s3:ListBucket` on the bucket. DuckDB lists the objects in your bucket to expand paths such as `orders/*/*.csv.gz`. Also verify that the path in your SQL matches where you store your files. If an earlier task writes the files, verify that the earlier task succeeded.

Resolve DuckDB extension installation or loading failures

Check whether you installed a DuckDB version other than the default one, DuckDB 2.0.0, in your `requirements.txt` file. If your environment doesn't have internet access, DuckDB can't download extensions for that version. Remove the version from your `requirements.txt` file to use the default version and the extensions in the Amazon MWAA image. Alternatively, provide the extensions yourself as described in Customize DuckDB in your environment.

Reduce memory use for DuckDB tasks

Reduce the data that each query reads, because DuckDB runs on the worker and shares its memory with other tasks. For example, filter on a partition column such as `order_date`. You can also set `memory_limit` in the connection's `extra` field, or use a larger Configuring the Amazon MWAA environment class. For queries over large datasets, consider using Athena instead.

Next steps

- Run DuckDB tasks when another DAG updates your data, without changing that DAG. To do this, add an `outlets` asset to the task that writes the data. Then use that asset as the `schedule` for your DuckDB DAG. For more information, see Asset-aware scheduling in the Apache Airflow documentation.
- Query JSON and Parquet files in Amazon S3 by using the DuckDB `read_json` and `read_parquet` functions in your DAGs.
- Share results with other teams and BI tools by querying the Parquet files that DuckDB writes to Amazon S3 with Athena.

Related resources

- DuckDB documentation on the DuckDB website
- What is Amazon Athena? in the *Amazon Athena User Guide*
- Amazon MWAA execution role

Using a DAG to run DuckDB queries in memory

You can use the following code example to write a directed acyclic graph (DAG) that runs DuckDB SQL in memory on your Amazon MWAA environment's workers, without reading from or writing to other AWS services.

The DAG has one task, `summarize_sales`, that uses the DuckDB query operator, `DuckDBExecuteQueryOperator`, to run a list of three SQL statements. The statements create a table, add rows to it, and return a per-store summary. The operator runs the statements in order on the same DuckDB connection and returns the result of the last statement.

Topics

- Version
- Prerequisites
- Permissions
- Dependencies
- Code example

Version

You can use the code example on this page with **Apache Airflow v3.3.1** and later in Python 3.12. For information about using DuckDB in an existing Apache Airflow v3.3.1 environment, see Version.

Prerequisites

To use the code example on this page, you need an Amazon MWAA environment using Apache Airflow v3.3.1 or later.

Permissions

You don't need additional permissions to use the code example on this page. The DAG doesn't access other AWS services, so it doesn't incur charges beyond your environment's usage.

Dependencies

You don't need additional dependencies to use the code example on this page. Amazon MWAA environments that use Apache Airflow v3.3.1 or later include the DuckDB provider and the `duckdb` library.

Code example

To run a DuckDB query in memory

1. Copy the following Python code example, and save it locally as `duckdb_local.py`.

```
import pendulum

from airflow.providers.duckdb.operators.duckdb import DuckDBExecuteQueryOperator
from airflow.sdk import DAG

with DAG(
    dag_id="duckdb_local",
    schedule=None,
    start_date=pendulum.datetime(2026, 1, 1, tz="UTC"),
    catchup=False,
):
    summarize_sales = DuckDBExecuteQueryOperator(
        task_id="summarize_sales",
        sql=[
```

```
            """
            CREATE TABLE sales (store_id INTEGER, quantity INTEGER, unit_price
DECIMAL(10, 2))
            """,
            """
            INSERT INTO sales VALUES
                (1, 2, 19.99), (1, 1, 5.50), (2, 4, 12.00), (2, 1, 89.00), (3, 3,
7.25)
            """,
            """
            SELECT store_id,
                   count(*) AS orders,
                   sum(quantity) AS units,
                   sum(quantity * unit_price) AS revenue
            FROM sales
            GROUP BY store_id
            ORDER BY store_id
            """,
        ],
    )
```

If a SQL statement fails, the operator raises an error, and Apache Airflow marks the task as failed. The error appears in the task log. To retry failed tasks automatically, set `retries` on the task or in the DAG's `default_args`.

2. Upload `duckdb_local.py` to the `dags` folder in your environment's Amazon S3 bucket. For instructions on uploading DAGs, see Adding or updating DAGs.
3. Open the Amazon MWAA console, and choose **Open Airflow UI** for your environment.
4. Trigger the `duckdb_local` DAG. The run might take a few minutes to complete. If the run fails, check the task log. For troubleshooting information, see Troubleshooting DuckDB tasks.
5. When the run completes, choose the `summarize_sales` task, and then view the query results on the **XCom** tab.

   The following table shows the expected results, with one row for each store.

store_id	orders	units	revenue
1	2	3	45.48
2	2	5	137.00

store_id	orders	units	revenue
3	1	3	21.75

6. **Clean up resources**

   When you finish, pause the `duckdb_local` DAG, and delete `duckdb_local.py` from the `dags` folder in your environment's Amazon S3 bucket. You don't need to delete the `sales` table. DuckDB deletes the in-memory table when the task ends, so this example doesn't create any other resources.

> **Pass data between tasks**
>
> Downstream tasks often need the results of earlier tasks, such as a file that an earlier task exported. Each task opens its own in-memory DuckDB database, and DuckDB discards the database when the task ends, so tasks can't share data in memory. To pass data between tasks, write it to Amazon S3, as shown in Using a DAG to read and write Amazon S3 data with DuckDB.

Using a DAG to read and write Amazon S3 data with DuckDB

You can use the following code example to write a directed acyclic graph (DAG) that uses DuckDB to generate, check, and summarize sample order data in Amazon S3. The DAG runs on your Amazon MWAA environment's workers and reads and writes files in Amazon S3 as your environment's execution role. The DAG has three tasks:

- `write_orders` generates a day of sample orders and writes them to Amazon S3 as a gzip-compressed CSV file, in a folder for the run's logical date.
- `check_orders` reads the day's orders directly from Amazon S3 and counts duplicate order IDs and invalid quantities. You don't need to define a table first.
- `summarize_orders` writes the day's orders, units, and revenue for each store to Amazon S3 as a Parquet file that you can query with other tools, such as Amazon Athena (Athena).

For more information about the SQL functions that this example uses, such as `read_csv` and `COPY`, see the DuckDB documentation on the DuckDB website.

Topics

- Version
- Prerequisites
- Permissions
- Dependencies
- Code example
- Clean up resources

Version

You can use the code example on this page with **Apache Airflow v3.3.1** and later in Python 3.12. For information about using DuckDB in an existing Apache Airflow v3.3.1 environment, see Version.

Prerequisites

To use the code example on this page, you need the following:

- An Amazon MWAA environment using Apache Airflow v3.3.1 or later.
- An Amazon S3 bucket for your data.
- Permission to update your environment's execution role. For more information about the execution role, see Amazon MWAA execution role.

This example uses AWS services that might incur charges, including Amazon S3 storage and request costs for the sample orders and reports that the DAG creates.

Permissions

DuckDB uses your environment's execution role to access Amazon S3. To use the code example on this page, the execution role needs `s3:GetObject`, `s3:PutObject`, `s3:ListBucket`, and `s3:GetBucketLocation` on your data bucket. If your bucket uses an AWS Key Management Service (AWS KMS) key for encryption, the execution role also needs `kms:Decrypt` and `kms:GenerateDataKey` on that key.

To add DuckDB permissions to the execution role

1. Open the AWS Identity and Access Management (IAM) console.

2. Choose your environment's execution role, and add the following permissions policy to it. Replace *amzn-s3-demo-bucket* with the name of the bucket that stores your data. For more information about the execution role, see Amazon MWAA execution role.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "DuckDBData",
      "Effect": "Allow",
      "Action": [
        "s3:GetObject",
        "s3:PutObject",
        "s3:ListBucket",
        "s3:GetBucketLocation"
      ],
      "Resource": [
        "arn:aws:s3:::amzn-s3-demo-bucket",
        "arn:aws:s3:::amzn-s3-demo-bucket/*"
      ]
    }
  ]
}
```

DuckDB needs `s3:ListBucket` to expand wildcard paths, such as `orders/*/*.csv.gz`. If your DAGs only read data, you can remove `s3:PutObject`.

If the execution role is missing a required permission, the task fails with an HTTP 403 or access denied error. For more information, see Troubleshooting DuckDB tasks.

Dependencies

You don't need additional dependencies to use the code example on this page. Amazon MWAA environments that use Apache Airflow v3.3.1 or later include the DuckDB provider, the `duckdb` library, and the DuckDB extensions for Amazon S3 access.

Code example

To run the DuckDB Amazon S3 code example

> **Important**
>
> This example stores sample data in Amazon S3, so you incur charges for Amazon S3 storage and requests. The sample data is smaller than 1 MB for each DAG run. For more information, see Amazon S3 pricing. To avoid ongoing charges, follow the steps in Clean up resources when you finish.

1. Copy the following Python code example, and save it locally as `duckdb_s3.py`. Replace *`amzn-s3-demo-bucket`* with the name of your bucket.

```
import pendulum

from airflow.providers.duckdb.operators.duckdb import DuckDBExecuteQueryOperator
from airflow.sdk import DAG

DATA = "s3://amzn-s3-demo-bucket/duckdb-sample"

with DAG(
    dag_id="duckdb_s3",
    schedule=None,
    start_date=pendulum.datetime(2026, 1, 1, tz="UTC"),
    catchup=False,
):
    # Generate 10,000 sample orders and write them to S3 as gzip-compressed CSV.
    write_orders = DuckDBExecuteQueryOperator(
        task_id="write_orders",
        sql=f"""
            COPY (
                SELECT i AS order_id,
                       (i % 10) + 1 AS store_id,
                       (i % 5) + 1 AS quantity,
                       round(5 + (i % 50) * 1.5, 2) AS unit_price
                FROM range(1, 10001) AS t(i)
            ) TO '{DATA}/orders/order_date={{{{ ds }}}}/orders.csv.gz'
              (FORMAT csv, HEADER true, COMPRESSION gzip)
        """,
    )
```

```
    # Read the day's orders directly from S3 and check them.
    check_orders = DuckDBExecuteQueryOperator(
        task_id="check_orders",
        sql=f"""
            SELECT count(*) AS orders,
                   count(*) - count(DISTINCT order_id) AS duplicate_order_ids,
                   count(*) FILTER (WHERE quantity <= 0) AS non_positive_quantities
            FROM read_csv('{DATA}/orders/*/*.csv.gz', hive_partitioning = true)
            WHERE order_date = DATE '{{{{ ds }}}}'
        """,
    )

    # Write a per-store summary of the day's orders to S3 as Parquet.
    summarize_orders = DuckDBExecuteQueryOperator(
        task_id="summarize_orders",
        sql=f"""
            COPY (
                SELECT store_id,
                       count(*) AS orders,
                       sum(quantity) AS units,
                       round(sum(quantity * unit_price), 2) AS revenue
                FROM read_csv('{DATA}/orders/*/*.csv.gz', hive_partitioning = true)
                WHERE order_date = DATE '{{{{ ds }}}}'
                GROUP BY store_id
                ORDER BY store_id
            ) TO '{DATA}/reports/daily_store_sales/order_date={{{{ ds }}}}/
sales.parquet'
              (FORMAT parquet)
        """,
    )

    write_orders >> check_orders >> summarize_orders
```

The SQL uses a Python f-string, so the doubled braces (`{{{{ ds }}}}`) render as the Apache Airflow template `{{ ds }}`. Apache Airflow replaces this template with the run's logical date. With `hive_partitioning = true`, DuckDB reads the `order_date` column from the folder names.

If a query fails, for example because of a missing permission or file, the operator raises an error, and Apache Airflow marks the task as failed. The error appears in the task log, and

downstream tasks don't run. To retry failed tasks automatically, set `retries` on the task or in the DAG's `default_args`.

2. Upload `duckdb_s3.py` to the `dags` folder in your environment's Amazon S3 bucket.
3. Open the Amazon MWAA console, and choose **Open Airflow UI** for your environment.
4. Trigger the `duckdb_s3` DAG. The run might take a few minutes to complete. If the run fails, check the task log. For troubleshooting information, see Troubleshooting DuckDB tasks.
5. Verify the results in the Apache Airflow UI after the run completes.

   a. Choose the `check_orders` task, and then on the **XCom** tab, view the results. The results show 10,000 orders, with no duplicate order IDs and no invalid quantities.

   b. In your data bucket, open the `duckdb-sample/reports/daily_store_sales/` folder, and confirm that it contains the Parquet file that the `summarize_orders` task wrote.

6. **Clean up resources**

   When you finish, pause the `duckdb_s3` DAG, delete `duckdb_s3.py` from your `dags` folder, and delete the `duckdb-sample/` folder from your data bucket. If your other DAGs don't need them, remove the `DuckDBData` permissions from the execution role. For detailed steps, see Clean up resources.

Clean up resources

When you finish with this example, delete the resources that you created to avoid unwanted DAG runs and ongoing Amazon S3 charges. To complete these steps, you need access to your environment's Amazon S3 bucket, your data bucket, and the execution role. For more information about the DAGs folder, see Adding or updating DAGs.

To clean up resources

1. Open the Amazon MWAA console, and choose **Open Airflow UI** for your environment.
2. Pause the `duckdb_s3` DAG.
3. Delete `duckdb_s3.py` from the `dags` folder in your environment's Amazon S3 bucket.
4. Delete the `duckdb-sample/` folder, which contains the sample orders and reports, from your data bucket.
5. If your other DAGs don't need them, remove the Amazon S3 permissions that you added to your environment's execution role.

AWS blogs and tutorials

- Working with Amazon EKS and Amazon MWAA for Apache Airflow v2.x

Best practices for Amazon Managed Workflows for Apache Airflow

This guide describes the best practices we recommend when using Amazon Managed Workflows for Apache Airflow.

Topics

- Performance tuning for Apache Airflow on Amazon MWAA
- Managing Python dependencies in requirements.txt

Performance tuning for Apache Airflow on Amazon MWAA

This topic describes how to tune the performance of an Amazon Managed Workflows for Apache Airflow environment using Using Apache Airflow configuration options on Amazon MWAA.

Contents

- Adding an Apache Airflow configuration option
- Apache Airflow scheduler
  - Parameters
  - Limits
- DAG folders
  - Parameters
- DAG files
  - Parameters
- Tasks
  - Parameters

Adding an Apache Airflow configuration option

Use the following procedure to add an Airflow configuration option to your environment.

1. Open the Environments page on the Amazon MWAA console.
2. Choose an environment.

3. Choose **Edit**.
4. Choose **Next**.
5. Choose **Add custom configuration** in the **Airflow configuration options** pane.
6. Choose a configuration from the dropdown list and enter a value, or enter a custom configuration and enter a value.
7. Choose **Add custom configuration** for each configuration you want to add.
8. Choose **Save**.

To learn more, refer to Using Apache Airflow configuration options on Amazon MWAA.

Apache Airflow scheduler

The Apache Airflow scheduler is a core component of Apache Airflow. An issue with the scheduler can prevent DAGs from being parsed and tasks from being scheduled. For more information about Apache Airflow scheduler tuning, refer to Fine-tuning your scheduler performance in the Apache Airflow documentation website.

Parameters

This section describes the configuration options available for the Apache Airflow scheduler (Apache Airflow v2 and later) and their use cases.

Apache Airflow v3

Configuration	Use case
celery.sync_parallelism The number of processes the Celery Executor uses to sync task state. **Default:** 1	Use this option to prevent queue conflicts by limiting the processes the Celery Executor uses. By default, a value is set to `1` to prevent errors in delivering task logs to CloudWatch Logs. Setting the value to `0` means using the maximum number of processes, but can cause errors when delivering task logs.
scheduler.scheduler_idle_sleep_time	Use this option to free up CPU usage on the scheduler by **increasing** the time

Configuration	Use case
The number of seconds to wait between consecutive scheduler "loops" if there was nothing to do in the loop. **Default:** 1	the scheduler sleeps after it finishes a "loop." Increasing this value reduces the scheduler threads available in `dag_processor.parsing_processes` for Apache Airflow v3. This can reduce the capacity of the schedulers to parse DAGs, and increase the time it takes for DAGs to populate in the webserver.
scheduler.max_dagruns_to_create_per_loop The maximum number of DAGs to create *DagRuns* for per scheduler "loop." **Default:** 10	Use this option to free up resources for scheduling tasks by **decreasing** the maximum number of *DagRuns* for the scheduler "loop."
dag_processor.parsing_processes The number of threads the scheduler can run in parallel to schedule DAGs. **Default:** Use `(2 * number of vCPUs) - 1`	Use this option to free up resources by **decreasing** the number of processes the scheduler runs in parallel to parse DAGs. We recommend keeping this number low if DAG parsing is impacting task scheduling. You **must** specify a value that does not exceed the number of threads available for your environment class, up to `(2 * number of vCPUs) - 1` because one thread is reserved for the scheduler. To learn more, refer to Limits.

Apache Airflow v2

Configuration	Use case

Configuration	Use case
celery.sync_parallelism The number of processes the Celery Executor uses to sync task state. **Default:** 1	Use this option to prevent queue conflicts by limiting the processes the Celery Executor uses. By default, a value is set to 1 to prevent errors in delivering task logs to CloudWatch Logs. Setting the value to `0` means using the maximum number of processes, but can cause errors when delivering task logs.
scheduler.scheduler_idle_sleep_time The number of seconds to wait between consecutive scheduler "loops" if there was nothing to do in the loop. **Default:** 1	Use this option to free up CPU usage on the scheduler by **increasing** the time the scheduler sleeps after it finishes a "loop." Increasing this value reduces the scheduler threads available in `scheduler.parsing_processes` for Apache Airflow v2. This can reduce the capacity of the schedulers to parse DAGs, and increase the time it takes for DAGs to populate in the webserver.
scheduler.max_dagruns_to_create_per_loop The maximum number of DAGs to create *DagRuns* for per scheduler "loop." **Default:** 10	Use this option to free up resources for scheduling tasks by **decreasing** the maximum number of *DagRuns* for the scheduler "loop."

Configuration	Use case
scheduler.parsing_processes The number of threads the scheduler can run in parallel to schedule DAGs. **Default:** Use `(2 * number of vCPUs) - 1`	Use this option to free up resources by **decreasing** the number of processes the scheduler runs in parallel to parse DAGs. We recommend keeping this number low if DAG parsing is impacting task scheduling. You **must** specify a value that does not exceed the number of threads available for your environment class, up to `(2 * number of vCPUs) - 1` because one thread is reserved for the scheduler. To learn more, refer to Limits.

Limits

This section describes the limits to consider when adjusting the default parameters for the scheduler.

scheduler.parsing_processes, scheduler.max_threads (v2 only)

Two threads are allowed per vCPU for an environment class. At least one thread must be reserved for the scheduler for an environment class. If you notice a delay in tasks being scheduled, you might need to increase your environment class. For example, a large environment has a 4 vCPU Fargate container instance for its scheduler. This means that a maximum of 7 total threads are available to use for other processes. That is, two threads multiplied four vCPUs, minus one for the scheduler itself. The value you specify in `scheduler.max_threads` (v2 only) and `scheduler.parsing_processes` must not exceed the number of threads available for an environment class, as listed:

- **mw1.small** – Must not exceed 1 thread for other processes. The remaining thread is reserved for the scheduler.
- **mw1.medium** – Must not exceed 3 threads for other processes. The remaining thread is reserved for the scheduler.
- **mw1.large** – Must not exceed 7 threads for other processes. The remaining thread is reserved for the scheduler.

DAG folders

The Apache Airflow scheduler continuously scans the DAGs folder on your environment. Any contained `plugins.zip` files, or Python (`.py`) files containing "airflow" import statements. Any resulting Python DAG objects are then placed into a *DagBag* for that file to be processed by the scheduler to determine what, if any, tasks need to be scheduled. Dag file parsing occurs regardless of whether the files contain any viable DAG objects.

Parameters

This section describes the configuration options available for the DAGs folder (Apache Airflow v2 and later) and their use cases.

Apache Airflow v3

Configuration	Use case
dag_processor.refresh_interval The number of seconds the DAGs folder must be scanned for new files. **Default:** 300 seconds	Use this option to free up resources by **increasing** the number of seconds to parse the DAGs folder. We recommend increasin g this value if you experience long parsing times in `total_parse_time metrics` , which might be due to a large number of files in your DAGs folder.
dag_processor.min_file_process_interval The number of seconds after which the scheduler parses a DAG and updates to the DAG are reflected. **Default:** 30 seconds	Use this option to free up resources by **increasing** the number of seconds that the scheduler waits before parsing a DAG. For example, if you specify a value of `30`, the DAG file is parsed after every 30 seconds. We recommend keeping this number high to decrease the CPU usage on your environme nt.

Apache Airflow v2

Configuration	Use case
scheduler.dag_dir_list_interval The number of seconds the DAGs folder must be scanned for new files. **Default:** 300 seconds	Use this option to free up resources by **increasing** the number of seconds to parse the DAGs folder. We recommend increasin g this value if you experience long parsing times in `total_parse_time metrics` , which might be due to a large number of files in your DAGs folder.
scheduler.min_file_process_interval The number of seconds after which the scheduler parses a DAG and updates to the DAG are reflected. **Default:** 30 seconds	Use this option to free up resources by **increasing** the number of seconds that the scheduler waits before parsing a DAG. For example, if you specify a value of `30`, the DAG file is parsed after every 30 seconds. We recommend keeping this number high to decrease the CPU usage on your environme nt.

DAG files

As part of the Apache Airflow scheduler loop, individual DAG files are parsed to extract DAG Python objects. In Apache Airflow v2 and later, the scheduler parses a maximum of number of parsing processes at the same time. The number of seconds specified in `scheduler.min_file_process_interval` (v2) or `dag_processor.min_file_process_interval` (v3) must pass before the same file is parsed again.

Parameters

This section describes the configuration options available for Apache Airflow DAG files (Apache Airflow v2 and later) and their use cases.

Apache Airflow v3

Configuration	Use case
dag_processor.dag_file_processor_timeout The number of seconds before the *DagFilePr ocessor* times out processing a DAG file. **Default:** 50 seconds	Use this option to increase the time it takes before the *DagFileProcessor* times out. We recommend increasing this value if you experience timeouts in your DAG processin g logs that result in no viable DAGs being loaded.
core.dagbag_import_timeout The number of seconds before importing a Python file times out. **Default:** 30 seconds	Use this option to increase the time it takes before the scheduler times out while importing a Python file to extract the DAG objects. This option is processed as part of the scheduler "loop," and must contain a value less than the value specified in `dag_processor.dag_file_proc essor_timeout` .
core.min_serialized_dag_update_interval The minimum number of seconds after which serialized DAGs in the database are updated. **Default:** 30	Use this option to free up resources by **increasing** the number of seconds after which serialized DAGs in the database are updated. We recommend increasing this value if you have a large number of DAGs, or complex DAGs. Increasing this value reduces the load on the scheduler and the database as DAGs are serialized.
core.min_serialized_dag_fetch_interval The number of seconds a serialized DAG is re-fetched from the database when already loaded in the DagBag. **Default:** 10	Use this option to control how often a serialized DAG is re-fetched from the database when it is already loaded in the DagBag on the webserver. Setting this value higher reduces load on the database, but at

Configuration	Use case
	the expense of displaying a possibly stale cached version of the DAG.

Apache Airflow v2

Configuration	Use case
core.dag_file_processor_timeout The number of seconds before the *DagFilePr ocessor* times out processing a DAG file. **Default:** 50 seconds	Use this option to increase the time it takes before the *DagFileProcessor* times out. We recommend increasing this value if you experience timeouts in your DAG processin g logs that result in no viable DAGs being loaded.
core.dagbag_import_timeout The number of seconds before importing a Python file times out. **Default:** 30 seconds	Use this option to increase the time it takes before the scheduler times out while importing a Python file to extract the DAG objects. This option is processed as part of the scheduler "loop," and must contain a value less than the value specified in `core.dag_file_processor_tim eout` .
core.min_serialized_dag_update_interval The minimum number of seconds after which serialized DAGs in the database are updated. **Default:** 30	Use this option to free up resources by **increasing** the number of seconds after which serialized DAGs in the database are updated. We recommend increasing this value if you have a large number of DAGs, or complex DAGs. Increasing this value reduces the load on the scheduler and the database as DAGs are serialized.

Configuration	Use case
core.min_serialized_dag_fetch_interval The number of seconds a serialized DAG is re-fetched from the database when already loaded in the DagBag. **Default:** 10	Use this option to control how often a serialized DAG is re-fetched from the database when it is already loaded in the DagBag on the webserver. Setting this value higher reduces load on the database, but at the expense of displaying a possibly stale cached version of the DAG.

Tasks

The Apache Airflow scheduler and workers are both involved in queuing and de-queuing tasks. The scheduler takes parsed tasks ready to schedule from a **None** status to a **Scheduled** status. The executor, also running on the scheduler container in Fargate, queues those tasks and sets their status to **Queued**. When the workers have capacity, it takes the task from the queue and sets the status to **Running**, which subsequently changes its status to **Success** or **Failed** based on whether the task succeeds or fails.

Parameters

This section describes the configuration options available for Apache Airflow tasks and their use cases.

The default configuration options that Amazon MWAA overrides are marked in *red*.

Apache Airflow v3

Configuration	Use case
core.parallelism The maximum number of task instances that each scheduler can monitor and execute simultaneously. **Default:** Dynamically set based on `(maxWorkerCount * maxCelery`	Use this option to control the hard global ceiling on total running tasks. For example, you can decrease this value to protect your metadatabase from too many concurrent connections or to cap costs. Increasing this value allows more tasks to run concurren tly. The default value is high so that other

Configuration	Use case
`WorkerSlots) / schedulers * 1.5`.	controls (`worker_autoscale` , pool slots, `max_active_tasks_per_dag`) serve as the effective concurrency limits.
core.max_active_tasks_per_dag The maximum number of task instances allowed to run concurrently in each DAG run. **Default:** 16	Use this option to free up resources by increasing the number of task instances allowed to run concurrently. For example, if you have 100 DAGs with 10 parallel tasks and want all DAGs to run concurrently, calculate the maximum parallelism. Multiply the number of available workers by the task density in `celery.worker_concurrency` , then divide by the number of DAGs.
core.execute_tasks_new_python_interpreter Determines whether Apache Airflow executes tasks by forking the parent process, or by creating a new Python process. **Default:** `True`	When set to `True`, Apache Airflow recognizes changes you make to your plugins as a new Python process so created to execute tasks.
celery.worker_concurrency Amazon MWAA overrides the Airflow base install for this option to scale workers as part of its autoscaling component. **Default:** Not applicable	*Any value specified for this option is ignored.*

Configuration	Use case
celery.worker_autoscale The task concurrency for workers. **Defaults:** • **mw1.micro** - 3,0 • **mw1.small** - 5,0 • **mw1.medium** - 10,0 • **mw1.large** - 20,0 • **mw1.xlarge** - 40,0 • **mw1.2xlarge** - 80,0	Use this option to free up resources by **reducing** the `maximum`, `minimum` task concurrency of workers. Workers accept up to the `maximum` concurrent tasks configure d, regardless of whether there are sufficien t resources to do so. If tasks are scheduled without sufficient resources, the tasks immediately fail. We recommend changing this value for resource-intensive tasks by reducing the values to be less than the defaults to allow more capacity per task.

Apache Airflow v2

Configuration	Use case
core.parallelism The maximum number of task instances that each scheduler can monitor and execute simultaneously. **Default:** Dynamically set based on `(maxWorkerCount * maxCelery WorkerSlots) / schedulers * 1.5`.	Use this option to control the hard global ceiling on total running tasks. For example, you can decrease this value to protect your metadatabase from too many concurrent connections or to cap costs. Increasing this value allows more tasks to run concurren tly. The default value is high so that other controls (`worker_autoscale` , pool slots, `max_active_tasks_per_dag`) serve as the effective concurrency limits.
core.dag_concurrency	Use this option to free up resources by **increasing** the number of task instances allowed to run concurrently. For example,

Configuration	Use case
The number of task instances allowed to run concurrently for each DAG. **Default:** 16 This configuration has been deprecated since Airflow 2.2.0 and was replaced by `core.max_active_tasks_per_dag` .	if you have 100 DAGs with 10 parallel tasks and want all DAGs to run concurrently, calculate the maximum parallelism. Multiply the number of available workers by the task density in `celery.worker_concurrency` , then divide by the number of DAGs.
core.execute_tasks_new_python_interpreter Determines whether Apache Airflow executes tasks by forking the parent process, or by creating a new Python process. **Default:** `True`	When set to `True`, Apache Airflow recognizes changes you make to your plugins as a new Python process so created to execute tasks.
celery.worker_concurrency Amazon MWAA overrides the Airflow base install for this option to scale workers as part of its autoscaling component. **Default:** Not applicable	*Any value specified for this option is ignored.*

Configuration	Use case
celery.worker_autoscale The task concurrency for workers. **Defaults:** • **mw1.micro** - 3,0 • **mw1.small** - 5,0 • **mw1.medium** - 10,0 • **mw1.large** - 20,0 • **mw1.xlarge** - 40,0 • **mw1.2xlarge** - 80,0	Use this option to free up resources by **reducing** the `maximum`, `minimum` task concurrency of workers. Workers accept up to the `maximum` concurrent tasks configure d, regardless of whether there are sufficien t resources to do so. If tasks are scheduled without sufficient resources, the tasks immediately fail. We recommend changing this value for resource-intensive tasks by reducing the values to be less than the defaults to allow more capacity per task.

Managing Python dependencies in requirements.txt

This topic describes how to install and manage Python dependencies in a `requirements.txt` file for an Amazon Managed Workflows for Apache Airflow environment.

Contents

- Testing DAGs using the Amazon MWAA CLI utility
- Installing Python dependencies using PyPi.org Requirements File Format
  - Option one: Python dependencies from the Python Package Index
  - Option two: Python wheels (.whl)
    - Using the plugins.zip file on an Amazon S3 bucket
    - Using a WHL file hosted on a URL
    - Creating a WHL files from a DAG
  - Option three: Python dependencies hosted on a private PyPi/PEP-503 Compliant Repo
- Enabling logs on the Amazon MWAA console
- Accessing logs on the CloudWatch Logs console
- Accessing errors in the Apache Airflow UI

- Log in to Apache Airflow
- Example requirements.txt scenarios

Testing DAGs using the Amazon MWAA CLI utility

- The command line interface (CLI) utility replicates an Amazon Managed Workflows for Apache Airflow environment locally.
- The CLI builds a Docker container image locally that's similar to an Amazon MWAA production image. You can use this to run a local Apache Airflow environment to develop and test DAGs, custom plugins, and dependencies before deploying to Amazon MWAA.
- To run the CLI, refer to aws-mwaa-docker-images on GitHub.

Installing Python dependencies using PyPi.org Requirements File Format

The following section describes the different ways to install Python dependencies according to the PyPi.org Requirements File Format.

Option one: Python dependencies from the Python Package Index

The following section describes how to specify Python dependencies from the Python Package Index in a `requirements.txt` file.

Apache Airflow v3

1. **Test locally**. Add additional libraries iteratively to find the right combination of packages and their versions, before creating a `requirements.txt` file. To run the Amazon MWAA CLI utility, refer to aws-mwaa-docker-images on GitHub.
2. **Review the Apache Airflow package extras**. To access a list of the packages installed for Apache Airflow v3 on Amazon MWAA, refer to aws-mwaa-docker-images `requirements.txt` on the GitHub website.
3. **Add a constraints statement**. Add the constraints file for your Apache Airflow v3 environment at the top of your `requirements.txt` file. Apache Airflow constraints files specify the provider versions available at the time of a Apache Airflow release.

In the following example, replace *{Airflow-version}* with your environment's version number, and *{Python-version}* with the version of Python that's compatible with your environment.

For information about the version of Python compatible with your Apache Airflow environment, refer to Apache Airflow Versions.

```
--constraint "https://raw.githubusercontent.com/apache/airflow/
constraints-{Airflow-version}/constraints-{Python-version}.txt"
```

If the constraints file determines that `xyz==1.0` package is not compatible with other packages in your environment, `pip3 install` fails to prevent incompatible libraries from being installed to your environment. If installation fails for any packages, you can access error logs for each Apache Airflow component (the scheduler, worker, and webserver) in the corresponding log stream on CloudWatch Logs. For more information about log types, refer to the section called “Accessing Airflow logs”.

4. **Apache Airflow packages**. Add the package extras and the version (==). This helps to prevent packages of the same name, but different version, from being installed on your environment.

```
apache-airflow[package-extra]==3.3.1
```

5. **Python libraries**. Add the package name and the version (==) in your `requirements.txt` file. This helps to prevent a future breaking update from PyPi.org from being automatically applied.

```
library == version
```

Example Boto3 and psycopg2-binary

This example is provided for demonstration purposes. The psycopg2-binary library is included with the base install for Apache Airflow v3 and doesn't need to be specified in a `requirements.txt` file.

```
boto3==1.43.56
botocore==1.43.56
psycopg2-binary==2.9.12
```

If a package is specified without a version, Amazon MWAA installs the latest version of the package from PyPi.org. This version can conflict with other packages in your `requirements.txt`.

Apache Airflow v2

1. **Test locally**. Add additional libraries iteratively to find the right combination of packages and their versions, before creating a `requirements.txt` file. To run the Amazon MWAA CLI utility, refer to aws-mwaa-docker-images on GitHub.
2. **Review the Apache Airflow package extras**. To access a list of the packages installed for Apache Airflow v2 on Amazon MWAA, access aws-mwaa-docker-images `requirements.txt` on the GitHub website.
3. **Add a constraints statement**. Add the constraints file for your Apache Airflow v2 environment at the top of your `requirements.txt` file. Apache Airflow constraints files specify the provider versions available at the time of a Apache Airflow release.

   Beginning with Apache Airflow v2.7.2, your requirements file must include a `--constraint` statement. If you do not provide a constraint, Amazon MWAA will specify one for you to ensure the packages listed in your requirements are compatible with the version of Apache Airflow you are using.

   In the following example, replace *`{Airflow-version}`* with your environment's version number, and *`{Python-version}`* with the version of Python that's compatible with your environment.

   For information about the version of Python compatible with your Apache Airflow environment, refer to Apache Airflow Versions.

   ```
   --constraint "https://raw.githubusercontent.com/apache/airflow/
   constraints-{Airflow-version}/constraints-{Python-version}.txt"
   ```

   If the constraints file determines that `xyz==1.0` package is not compatible with other packages in your environment, `pip3 install` fails to prevent incompatible libraries from being installed to your environment. If installation fails for any packages, you can access error logs for each Apache Airflow component (the scheduler, worker, and webserver) in the

corresponding log stream on CloudWatch Logs. For more information about log types, refer to the section called "Accessing Airflow logs".

4. **Apache Airflow packages**. Add the package extras and the version (==). This helps to prevent packages of the same name, but different version, from being installed on your environment.

```
apache-airflow[package-extra]==2.5.1
```

5. **Python libraries**. Add the package name and the version (==) in your `requirements.txt` file. This helps to prevent a future breaking update from PyPi.org from being automatically applied.

```
library == version
```

Example Boto3 and psycopg2-binary

This example is provided for demonstration purposes. The boto and psycopg2-binary libraries are included with the Apache Airflow v2 base install and don't need to be specified in a `requirements.txt` file.

```
boto3==1.17.54
boto==2.49.0
botocore==1.20.54
psycopg2-binary==2.8.6
```

If a package is specified without a version, Amazon MWAA installs the latest version of the package from PyPi.org. This version can conflict with other packages in your `requirements.txt`.

Option two: Python wheels (.whl)

A Python wheel is a package format designed to ship libraries with compiled artifacts. There are several benefits to wheel packages as a method to install dependencies in Amazon MWAA:

- **Faster installation** – the WHL files are copied to the container as a single ZIP, and then installed locally, without having to download each one.
- **Fewer conflicts** – You can determine version compatibility for your packages in advance. As a result, there is no need for `pip` to recursively work out compatible versions.

- **More resilience** – With externally hosted libraries, downstream requirements can change, resulting in version incompatibility between containers on a Amazon MWAA environment. By not depending on an external source for dependencies, every container will have the same libraries regardless of when each container is instantiated.

We recommend the following methods to install Python dependencies from a Python wheel archive (`.whl`) in your `requirements.txt`.

Methods

- Using the plugins.zip file on an Amazon S3 bucket
- Using a WHL file hosted on a URL
- Creating a WHL files from a DAG

Using the `plugins.zip` file on an Amazon S3 bucket

The Apache Airflow scheduler, workers, and webserver (for Apache Airflow v2.2.2 and later) search for custom plugins during startup on the AWS-managed Fargate container for your environment at `/usr/local/airflow/plugins/*`. This process begins prior to Amazon MWAA's `pip3 install -r requirements.txt` for Python dependencies and Apache Airflow service startup. A `plugins.zip` file can be used for any files that you don't want continuously changed during environment execution, or that you do not want to grant access to users that write DAGs. For example, Python library wheel files, certificate PEM files, and configuration YAML files.

The following section describes how to install a wheel that's in the `plugins.zip` file on your Amazon S3 bucket.

1. **Download the necessary WHL files** You can use `pip download` with your existing `requirements.txt` on the Amazon MWAA aws-mwaa-docker-images or another Amazon Linux 2 container to resolve and download the necessary Python wheel files.

   ```
   pip3 download -r "$AIRFLOW_HOME/dags/requirements.txt" -d "$AIRFLOW_HOME/plugins"
   cd "$AIRFLOW_HOME/plugins"
   zip "$AIRFLOW_HOME/plugins.zip" *
   ```

2. **Specify the path in your `requirements.txt`**. Specify the plugins directory at the top of your requirements.txt using `--find-links` and instruct `pip` not to install from other sources using `--no-index`, as listed in the following code:

```
--find-links /usr/local/airflow/plugins
--no-index
```

Example wheel in requirements.txt

The following example assumes you've uploaded the wheel in a `plugins.zip` file at the root of your Amazon S3 bucket. For example:

```
--find-links /usr/local/airflow/plugins
--no-index

numpy
```

Amazon MWAA fetches the `numpy-1.20.1-cp37-cp37m-manylinux1_x86_64.whl` wheel from the `plugins` folder and installs it on your environment.

Using a WHL file hosted on a URL

The following section describes how to install a wheel that's hosted on a URL. The URL must either be publicly accessible, or accessible from within the custom Amazon VPC you specified for your Amazon MWAA environment.

- **Provide a URL**. Provide the URL to a wheel in your `requirements.txt`.

  Example wheel archive on a public URL

  The following example downloads a wheel from a public site.

  ```
  --find-links https://files.pythonhosted.org/packages/
  --no-index
  ```

  Amazon MWAA fetches the wheel from the URL you specified and installs them on your environment.

> **Note**
>
> URLs are not accessible from private webservers installing requirements in Amazon MWAA v2.2.2 and later.

Creating a WHL files from a DAG

If you have a **Private network** environment using Apache Airflow v2.2.2 or later and you're unable to install requirements because your environment does not have access to external repositories, you can use the following DAG to take your existing Amazon MWAA requirements and package them on Amazon S3:

```
from airflow import DAG
 from airflow.operators.bash_operator import BashOperator
 from airflow.utils.dates import days_ago

 S3_BUCKET = 'my-s3-bucket'
 S3_KEY = 'backup/plugins_whl.zip'

 with DAG(dag_id="create_whl_file", schedule_interval=None, catchup=False,
 start_date=days_ago(1)) as dag:
 cli_command = BashOperator(
 task_id="bash_command",
 bash_command=f"mkdir /tmp/whls;pip3 download -r /usr/local/airflow/requirements/
requirements.txt -d /tmp/whls;zip -j /tmp/plugins.zip /tmp/whls/*;aws s3 cp /tmp/
plugins.zip s3://{S3_BUCKET}/{S3_KEY}"
)
```

After running the DAG, use this new file as your Amazon MWAA `plugins.zip`, optionally, packaged with other plugins. Then, update your `requirements.txt` preceded by `--find-links /usr/local/airflow/plugins` and `--no-index` without adding `--constraint`.

This method you can use to use the same libraries offline.

Option three: Python dependencies hosted on a private PyPi/PEP-503 Compliant Repo

The following section describes how to install an Apache Airflow extra that's hosted on a private URL with authentication.

1. Add your user name and password as Apache Airflow configuration options. For example:
   - `foo.user` : *YOUR_USER_NAME*
   - `foo.pass` : *YOUR_PASSWORD*
2. Create your `requirements.txt` file. Substitute the placeholders in the following example with your private URL, and the username and password you've added as Apache Airflow configuration options. For example:

```
--index-url https://${AIRFLOW__FOO__USER}:${AIRFLOW__FOO__PASS}@my.privatepypi.com
```

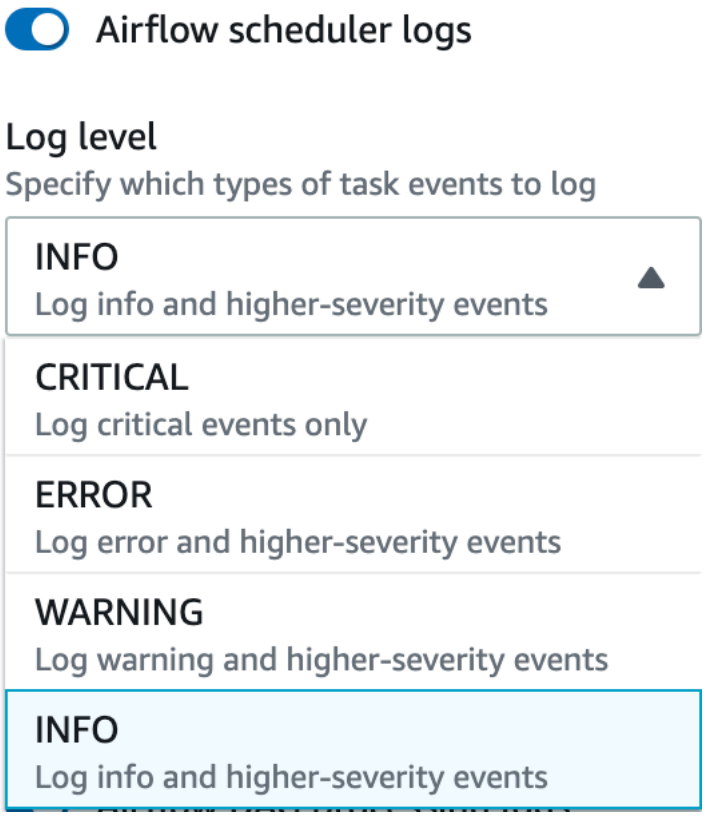
3. Add any additional libraries to your `requirements.txt` file. For example:

```
--index-url https://${AIRFLOW__FOO__USER}:${AIRFLOW__FOO__PASS}@my.privatepypi.com
my-private-package==1.2.3
```

Enabling logs on the Amazon MWAA console

The execution role for your Amazon MWAA environment needs permission to send logs to CloudWatch Logs. To update the permissions of an execution role, refer to Amazon MWAA execution role.

You can enable Apache Airflow logs at the `INFO`, `WARNING`, `ERROR`, or `CRITICAL` level. When you choose a log level, Amazon MWAA sends logs for that level and all higher levels of severity. For example, if you enable logs at the `INFO` level, Amazon MWAA sends `INFO` logs and `WARNING`, `ERROR`, and `CRITICAL` log levels to CloudWatch Logs. We recommend enabling Apache Airflow logs at the `INFO` level for the scheduler to access logs received for the `requirements.txt`.



Accessing logs on the CloudWatch Logs console

You can access Apache Airflow logs for the scheduler scheduling your workflows and parsing your `dags` folder. The following steps describe how to open the log group for the scheduler on the Amazon MWAA console, and access Apache Airflow logs on the CloudWatch Logs console.

To access logs for a `requirements.txt`

1. Open the Environments page on the Amazon MWAA console.
2. Choose an environment.
3. Choose the **Airflow scheduler log group** on the **Monitoring** pane.
4. Choose the `requirements_install_ip` log in **Log streams**.
5. Refer to the list of packages that were installed on the environment at `/usr/local/airflow/.local/bin`. For example:

```
Collecting appdirs==1.4.4 (from -r /usr/local/airflow/.local/bin (line 1))
Downloading https://files.pythonhosted.org/
packages/3b/00/2344469e2084fb28kjdsfiuyweb47389789vxbmnbjhsdgf5463acd6cf5e3db69324/
appdirs-1.4.4-py2.py3-none-any.whl
```

```
Collecting astroid==2.4.2 (from -r /usr/local/airflow/.local/bin (line 2))
```

6. Review the list of packages and whether any of these encountered an error during installation. If something went wrong, you can get an error similar to the following:

```
2021-03-05T14:34:42.731-07:00
No matching distribution found for LibraryName==1.0.0 (from -r /usr/local/
airflow/.local/bin (line 4))
No matching distribution found for LibraryName==1.0.0 (from -r /usr/local/
airflow/.local/bin (line 4))
```

Accessing errors in the Apache Airflow UI

You can also check your Apache Airflow UI to identify whether an error is related to another issue. The most common error you can encounter with Apache Airflow on Amazon MWAA is:

```
Broken DAG: No module named x
```

If you find this error in your Apache Airflow UI, you're likely missing a required dependency in your `requirements.txt` file.

Log in to Apache Airflow

You need Apache Airflow UI access policy: AmazonMWAAWebServerAccess permissions for your AWS account in AWS Identity and Access Management (IAM) to access your Apache Airflow UI.

To access your Apache Airflow UI

1. Open the Environments page on the Amazon MWAA console.
2. Choose an environment.
3. Choose **Open Airflow UI**.

Example `requirements.txt` scenarios

You can mix and match different formats in your `requirements.txt`. The following example uses a combination of the different ways to install extras.

Example Extras on PyPi.org and a public URL

You need to use the `--index-url` option when specifying packages from PyPi.org, in addition to packages on a public URL, such as custom PEP 503 compliant repo URLs.

```
aws-batch == 0.6
    phoenix-letter >= 0.3

    --index-url http://dist.repoze.org/zope2/2.10/simple
    zopelib
```

Monitoring and metrics for Amazon Managed Workflows for Apache Airflow

Monitoring is an important part of maintaining the reliability, availability, and performance of Amazon Managed Workflows for Apache Airflow and your AWS solution. We recommend collecting monitoring data from all parts of your AWS solution so you can more easily debug a multi-point failure if one occurs. This topic describes what resources AWS provides for monitoring your Amazon MWAA environment and responding to potential events.

> **Note**
>
> Apache Airflow metrics and logging are subject to standard Amazon CloudWatch pricing.

For more information about monitoring Apache Airflow, refer to Logging & Monitoring in the Apache Airflow documentation website.

Sections

- Monitoring overview on Amazon MWAA
- Accessing audit logs in AWS CloudTrail
- Accessing Airflow logs in Amazon CloudWatch
- Monitoring dashboards and alarms on Amazon MWAA
- Apache Airflow environment metrics in CloudWatch
- Container, queue, and database metrics for Amazon MWAA

Monitoring overview on Amazon MWAA

This page describes the AWS services used to monitor an Amazon Managed Workflows for Apache Airflow environment.

Contents

- Amazon CloudWatch overview
- AWS CloudTrail overview

Amazon CloudWatch overview

CloudWatch is a metrics repository for AWS services that you can use to retrieve statistics based on the metrics and dimensions published by a service. You can use these metrics to configure alarms, calculate statistics and then present the data in a dashboard that helps you assess the health of your environment in the Amazon CloudWatch console.

Apache Airflow is already set up to send StatsD metrics for an Amazon Managed Workflows for Apache Airflow environment to Amazon CloudWatch.

To learn more, refer to What is Amazon CloudWatch?.

AWS CloudTrail overview

CloudTrail is an auditing service that provides a record of actions taken by a user, role, or an AWS service in Amazon MWAA. Using the information collected by CloudTrail, you can determine the request that was made to Amazon MWAA, the IP address from which the request was made, who made the request, when it was made, and additional details available in audit logs.

To learn more, refer to What is AWS CloudTrail?.

Accessing audit logs in AWS CloudTrail

AWS CloudTrail is enabled on your AWS account when you create it. CloudTrail logs the activity taken by an IAM entity or an AWS service, such as Amazon Managed Workflows for Apache Airflow, which is recorded as a CloudTrail event. You can view, search, and download the past 90 days of event history in the CloudTrail console. CloudTrail captures all events on the Amazon MWAA console and all calls to Amazon MWAA APIs. It doesn't capture read-only actions, such as `GetEnvironment`, or the `PublishMetrics` action. This page describes how to use CloudTrail to monitor events for Amazon MWAA.

Contents

- Creating a trail in CloudTrail
- Logging data plane events
- Accessing events with CloudTrail Event History
- Example trail for CreateEnvironment
- InvokeRestApi log entry example

- What's next?

Creating a trail in CloudTrail

You need to create a trail to access an ongoing record of events in your AWS account, including events for Amazon MWAA. A trail enables CloudTrail to deliver log files to an Amazon S3 bucket. If you don't create a trail, you can still access available event history in the CloudTrail console. For example, using the information collected by CloudTrail, you can determine the request that was made to Amazon MWAA, the IP address from which the request was made, who made the request, when it was made, and additional details. To learn more, refer to the Creating a trail for your AWS account.

Logging data plane events

The following Amazon MWAA API actions are data plane events: `InvokeRestApi`, `CreateWebLoginToken`, and `CreateCliToken`. Unlike management events, AWS CloudTrail does not log data events by default, and they do not appear in CloudTrail Event History. You must configure a trail that explicitly includes data events. Additional charges apply. For more information, see Logging data events in the AWS CloudTrail User Guide.

To log data plane events

1. Open the CloudTrail console.
2. In the navigation pane, choose **Trails**, and then choose **Create trail**.
3. Enter a trail name and configure an Amazon S3 bucket for log files.
4. On the **Choose log events** page, select **Data events**.
5. Under **Data event type**, choose **Managed Apache Airflow**.
6. Review your settings and choose **Create trail**.

After the trail is active, CloudTrail delivers data plane events such as `InvokeRestApi`, `CreateWebLoginToken`, and `CreateCliToken` to your Amazon S3 bucket within approximately 15 minutes. If you configured Amazon CloudWatch Logs, use the following filter pattern to find these events:

```
{ ($.eventName = "InvokeRestApi") || ($.eventName = "CreateWebLoginToken") ||
 ($.eventName = "CreateCliToken") }
```

Response payload is not logged

CloudTrail redacts the `RestApiResponse` field in log entries. CloudTrail records the call and its status code but does not log the response payload.

Accessing events with CloudTrail Event History

You can troubleshoot operational and security incidents over the past 90 days in the CloudTrail console by viewing event history. For example, you can access events related to the creation, modification, or deletion of resources (such as IAM users or other AWS resources) in your AWS account on a per-region basis. To learn more, refer to the Accessing Events with CloudTrail Event History.

1. Open the CloudTrail console.
2. Choose **Event history**.
3. Select the events you want to view, and then choose **Compare event details**.

Example trail for `CreateEnvironment`

A trail is a configuration that enables delivery of events as log files to an Amazon S3 bucket that you specify.

CloudTrail log files contain one or more log entries. An event represents a single request from any source and includes information about the requested action, such as the date and time of the action, or request parameters. CloudTrail log files are not an ordered stack trace of the public API calls, and aren't listed in any specific order. The following example is a log entry for the `CreateEnvironment` action that is denied due to lacking permissions. The values in `AirflowConfigurationOptions` have been redacted for privacy.

```
{
  "eventVersion": "1.05",
  "userIdentity": {
    "type": "AssumedRole",
    "principalId": "00123456ABC7DEF8HIJK",
    "arn": "arn:aws:sts::012345678901:assumed-role/root/myuser",
    "accountId": "012345678901",
    "accessKeyId": "",
```

```
    "sessionContext": {
      "sessionIssuer": {
        "type": "Role",
        "principalId": "00123456ABC7DEF8HIJK",
        "arn": "arn:aws:iam::012345678901:role/user",
        "accountId": "012345678901",
        "userName": "user"
      },
      "webIdFederationData": {},
      "attributes": {
        "mfaAuthenticated": "false",
        "creationDate": "2020-10-07T15:51:52Z"
      }
    }
  },
  "eventTime": "2020-10-07T15:52:58Z",
  "eventSource": "airflow.amazonaws.com",
  "eventName": "CreateEnvironment",
  "awsRegion": "us-west-2",
  "sourceIPAddress": "205.251.233.178",
  "userAgent": "PostmanRuntime/7.26.5",
  "errorCode": "AccessDenied",
  "requestParameters": {
    "SourceBucketArn": "arn:aws:s3:::my-bucket",
    "ExecutionRoleArn": "arn:aws:iam::012345678901:role/AirflowTaskRole",
    "AirflowConfigurationOptions": "***",
    "DagS3Path": "sample_dag.py",
    "NetworkConfiguration": {
      "SecurityGroupIds": [
      "sg-01234567890123456"
      ],
      "SubnetIds": [
        "subnet-01234567890123456",
        "subnet-65432112345665431"
      ]
    },
    "Name": "test-cloudtrail"
  },
  "responseElements": {
    "message": "Access denied."
  },
  "requestID": "RequestID",
  "eventID": "EventID",
  "readOnly": false,
```

```
  "eventType": "AwsApiCall",
  "recipientAccountId": "012345678901"
}
```

InvokeRestApi log entry example

The following is a complete log entry example for the `InvokeRestApi` action. CloudTrail redacts the `RestApiResponse` field.

> **Use temporary credentials in production**
>
> This example shows temporary credentials from an assumed role, which is the recommended approach. For production workloads, use temporary credentials from or AWS Security Token Service (AWS STS) instead of long-term access keys.

```
{
  "eventVersion": "1.09",
  "userIdentity": {
    "type": "AssumedRole",
    "principalId": "00123456ABC7DEF8HIJK:MySession",
    "arn": "arn:aws:sts::012345678901:assumed-role/MyRole/MySession",
    "accountId": "012345678901",
    "accessKeyId": "ASIAIOSFODNN7EXAMPLE",
    "sessionContext": {
      "sessionIssuer": {
        "type": "Role",
        "principalId": "00123456ABC7DEF8HIJK",
        "arn": "arn:aws:iam::012345678901:role/MyRole",
        "accountId": "012345678901",
        "userName": "MyRole"
      },
      "attributes": {
        "creationDate": "2024-01-20T12:00:00Z",
        "mfaAuthenticated": "false"
      }
    }
  },
  "eventTime": "2024-01-20T12:15:00Z",
  "eventSource": "airflow.amazonaws.com",
  "eventName": "InvokeRestApi",
```

```
  "awsRegion": "us-west-2",
  "sourceIPAddress": "203.0.113.10",
  "userAgent": "Boto3/1.28.0 Python/3.11.0",
  "requestParameters": {
    "Name": "MyAirflowEnvironment",
    "Path": "/dags",
    "Method": "GET"
  },
  "responseElements": {
    "RestApiStatusCode": 200,
    "RestApiResponse": "***"
  },
  "requestID": "a1b2c3d4-5678-90ab-cdef-EXAMPLE11111",
  "eventID": "a1b2c3d4-5678-90ab-cdef-EXAMPLE22222",
  "readOnly": false,
  "resources": [
    {
      "accountId": "012345678901",
      "type": "AWS::MWAA::Environment",
      "ARN": "arn:aws:airflow:us-west-2:012345678901:environment/MyAirflowEnvironment"
    }
  ],
  "eventType": "AwsApiCall",
  "managementEvent": false,
  "recipientAccountId": "012345678901",
  "eventCategory": "Data"
}
```


What's next?

- Learn how to configure other AWS services for the event data collected in CloudTrail logs in CloudTrail Supported Services and Integrations.
- Learn how to be notified when CloudTrail publishes new log files to an Amazon S3 bucket in Configuring Amazon SNS Notifications for CloudTrail.

Accessing Airflow logs in Amazon CloudWatch

Amazon MWAA can send Apache Airflow logs to Amazon CloudWatch. You can access logs for multiple environments from a single location to easily identify Apache Airflow task delays or workflow errors without additional third-party tools. Apache Airflow logs need to be enabled

on the Amazon Managed Workflows for Apache Airflow console to access Apache Airflow DAG processing, tasks, webserver, worker logs in CloudWatch.

Contents

- Pricing
- Before you begin
- Log types
- Enabling Apache Airflow logs
- Accessing Apache Airflow logs
- Example scheduler logs
- What's next?

Pricing

- Standard CloudWatch Logs charges apply. For more information, refer to CloudWatch pricing.

Before you begin

- You must have a role that can access logs in CloudWatch. For more information, refer to Accessing an Amazon MWAA environment.

Log types

Amazon MWAA creates a log group for each Airflow logging option you enable, and pushes the logs to the CloudWatch Logs groups associated with an environment. Log groups are named in the following format: `airflow-YourEnvironmentName-`**`LogType`**. For example, if your environment's named `Airflow-v202-Public`, Apache Airflow task logs are sent to `airflow-Airflow-v202-Public-`**`Task`**.

Log type	Description
`airflow-YourEnvironmentName -` **`DAGProcessing`**	The logs of the DAG processor manager (the part of the scheduler that processes DAG files).

Log type	Description
`airflow-YourEnvironmentName - Scheduler`	The logs the Airflow scheduler generates.
`airflow-YourEnvironmentName- Task`	The task logs a DAG generates.
`airflow-YourEnvironmentName - WebServer`	The logs the Airflow web interface generates.
`airflow-YourEnvironmentName - Worker`	The logs generated as part of workflow and DAG execution.

Enabling Apache Airflow logs

You can enable Apache Airflow logs at the `INFO`, `WARNING`, `ERROR`, or `CRITICAL` level. When you choose a log level, Amazon MWAA sends logs for that level and all higher levels of severity. For example, if you enable logs at the `INFO` level, Amazon MWAA sends `INFO` logs and `WARNING`, `ERROR`, and `CRITICAL` log levels to CloudWatch Logs.

1. Open the Environments page on the Amazon MWAA console.
2. Choose an environment.
3. Choose **Edit**.
4. Choose **Next**.
5. Choose one or more of the following logging options:
    a. Choose the **Airflow scheduler log group** on the **Monitoring** pane.
    b. Choose the **Airflow webserver log group** on the **Monitoring** pane.
    c. Choose the **Airflow worker log group** on the **Monitoring** pane.
    d. Choose the **Airflow DAG processing log group** on the **Monitoring** pane.
    e. Choose the **Airflow task log group** on the **Monitoring** pane.
    f. Choose the logging level in **Log level**.
6. Choose **Next**.
7. Choose **Save**.

Accessing Apache Airflow logs

The following section describes how to access Apache Airflow logs in the CloudWatch console.

1. Open the Environments page on the Amazon MWAA console.
2. Choose an environment.
3. Choose a log group in the **Monitoring** pane.
4. Choose a log in **Log stream**.

Example scheduler logs

You can access Apache Airflow logs for the scheduler scheduling your workflows and parsing your `dags` folder. The following steps describe how to open the log group for the scheduler on the Amazon MWAA console, and access Apache Airflow logs on the CloudWatch Logs console.

To access logs for a `requirements.txt`

1. Open the Environments page on the Amazon MWAA console.
2. Choose an environment.
3. Choose the **Airflow scheduler log group** on the **Monitoring** pane.
4. Choose the `requirements_install_ip` log in **Log streams**.
5. Refer to the list of packages that were installed on the environment at `/usr/local/airflow/.local/bin`. For example:

```
Collecting appdirs==1.4.4 (from -r /usr/local/airflow/.local/bin (line 1))
Downloading https://files.pythonhosted.org/
packages/3b/00/2344469e2084fb28kjdsfiuyweb47389789vxbmnbjhsdgf5463acd6cf5e3db69324/
appdirs-1.4.4-py2.py3-none-any.whl
Collecting astroid==2.4.2 (from -r /usr/local/airflow/.local/bin (line 2))
```

6. Review the list of packages and whether any of these encountered an error during installation. If something went wrong, you might get an error similar to the following:

```
2021-03-05T14:34:42.731-07:00
No matching distribution found for LibraryName==1.0.0 (from -r /usr/local/
airflow/.local/bin (line 4))
```

```
No matching distribution found for LibraryName==1.0.0 (from -r /usr/local/
airflow/.local/bin (line 4))
```

What's next?

- Learn how to configure a CloudWatch alarm in Using Amazon CloudWatch alarms.
- Learn how to create a CloudWatch dashboard in Using CloudWatch dashboards.

Monitoring dashboards and alarms on Amazon MWAA

The environment details page for an Amazon Managed Workflows for Apache Airflow environment includes a built-in CloudWatch metrics dashboard and an **Environment alarms** table. You can use these features to monitor the health of your environment directly in the Amazon MWAA console, without any additional setup.

Apache Airflow exposes metrics for several processes, including the number of DAG processes, DAG bag size, currently running tasks, task failures, and successes. When you create an environment, Apache Airflow automatically sends metrics for your Amazon MWAA environment to CloudWatch. In addition to the built-in dashboard, you can build custom dashboards in Amazon CloudWatch and add alarms for particular metrics. When an alarm is on a dashboard, it turns red when it enters the ALARM state. This makes it easier to monitor the health of your Amazon MWAA environment proactively.

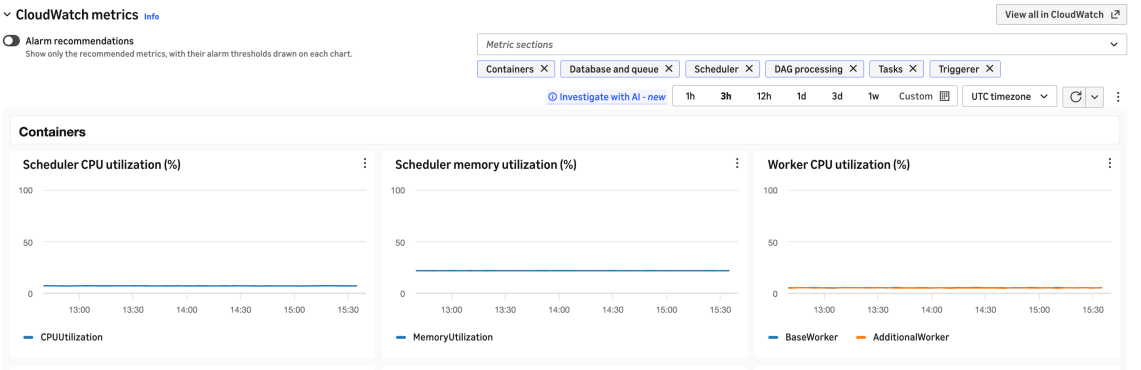
Contents

- View the metrics dashboard
  - Metrics shown on the dashboard
- Create and monitor alarms
  - Recommended alarms
- Alarm states overview
- Example custom dashboards and alarms
  - About these metrics
  - About the dashboard
  - Using AWS tutorials
  - Using CloudFormation

- Deleting metrics and dashboards
- What's next?

View the metrics dashboard

The environment details page includes a built-in CloudWatch metrics dashboard, shown in the console under **CloudWatch metrics**. This embedded CloudWatch dashboard displays charts for your environment without any additional setup.



To open the full set of metrics in the CloudWatch console, choose **View all in CloudWatch**.

The dashboard displays metrics from the following two CloudWatch namespaces:

- `AmazonMWAA` – Apache Airflow metrics.
- `AWS/MWAA` – Container, queue, and database metrics.

You can use the following controls to change what the dashboard displays:

- Turn on the **Alarm recommendations** toggle to show only the recommended metrics. When you turn on this toggle, each chart draws its alarm threshold as an annotation. The annotation reads *Alarms if `metric` above* or *Alarms if `metric` below*, depending on the metric.
- Use the **Metric sections** filter to show only the charts in a specific group. The available groups are **Containers**, **Database and queue**, **Scheduler**, **DAG processing**, **Tasks**, and **Triggerer**.

> **Note**
>
> If you configure an environment to filter which Apache Airflow metrics it publishes by using the metrics allow list or block list options, some charts might not have data.

> **Note**
>
> To view the metrics dashboard, your IAM identity must have permission to read CloudWatch metrics, including `cloudwatch:GetMetricData`, `cloudwatch:GetMetricStatistics`, and `cloudwatch:ListMetrics`. Without these permissions, the charts render with no data.

For the full list of metrics that Amazon MWAA publishes, see Apache Airflow environment metrics in CloudWatch.

Metrics shown on the dashboard

The dashboard organizes charts into six sections that you can filter with **Metric sections**. It shows a curated set of environment-wide metrics; charts come from the `AmazonMWAA` namespace (Apache Airflow metrics) and the `AWS/MWAA` namespace (container, queue, and database metrics). The following table lists the charts in each section and the CloudWatch metrics each one displays.

Dashboard charts and their metrics

Section	Chart	CloudWatch metrics
Containers	Scheduler CPU utilization	`CPUUtilization`
Containers	Scheduler memory utilization	`MemoryUtilization`
Containers	Worker CPU utilization	`CPUUtilization` (BaseWorker and AdditionalWorker)
Containers	Worker memory utilization	`MemoryUtilization` (BaseWorker and AdditionalWorker)
Containers	Web server CPU utilization	`CPUUtilization`
Containers	Web server memory utilization	`MemoryUtilization`

Section	Chart	CloudWatch metrics
Containers	Worker containers running	Derived from `CPUUtilization` sample count
Containers	Worker heartbeat	`CeleryWorkerHeartbeat`
Database and queue	Database freeable memory	`FreeableMemory`
Database and queue	Database CPU utilization	`CPUUtilization`
Database and queue	Database connections	`DatabaseConnections`
Database and queue	Database write latency	`WriteLatency`
Database and queue	Oldest queued task age	`ApproximateAgeOfOldestTask`
Scheduler	Scheduler heartbeat	`SchedulerHeartbeat`
Scheduler	Scheduler critical section duration	`CriticalSectionDuration`
Scheduler	Scheduler task states	`TasksExecutable` , `TasksStarving` , `TasksWithoutDagRun`
Scheduler	Tasks killed externally	`TasksKilledExternally`
Scheduler	Executor open slots	`OpenSlots`
DAG processing	DAG parse time	`TotalParseTime`
DAG processing	DAG bag size	`DagBagSize`
DAG processing	DAG import errors	`ImportErrors` , `DAGFileRefreshError`

Section	Chart	CloudWatch metrics
DAG processing	DAG processor health	`ProcessorTimeouts` , `ManagerStalls`
Tasks	Task queue	`QueuedTasks` , `RunningTasks`
Tasks	Task successes and failures	`TaskInstanceSuccesses` , `TaskInstanceFailures`
Tasks	Zombie tasks killed	`ZombiesKilled`
Triggerer	Triggerer heartbeat	`TriggererHeartbeat`
Triggerer	Trigger outcomes	`TriggersSucceeded` , `TriggersFailed`

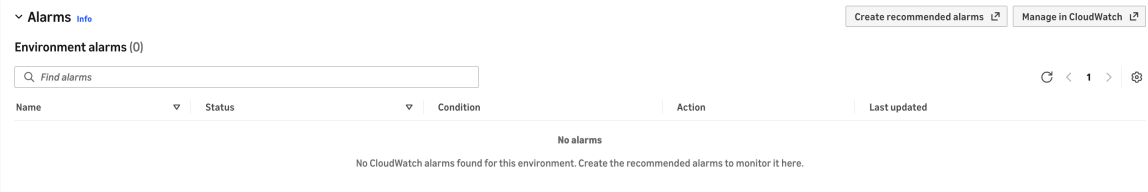
Note

The Triggerer charts appear only on newer Apache Airflow versions. **Trigger outcomes** requires Apache Airflow v2.7.2 or later, and **Triggerer heartbeat** requires v2.8.1 or later.

The dashboard shows aggregate, environment-wide metrics. It does not chart high-cardinality metrics that Apache Airflow publishes per DAG, task, pool, operator, or DAG file, and it charts container CPU and memory but not every container metric. For the complete list of published metrics, see Apache Airflow environment metrics in CloudWatch and Container, queue, and database metrics for Amazon MWAA.

Create and monitor alarms

The environment details page includes an **Alarms** pane where you create and manage your alarms in CloudWatch, and then monitor them in context. The pane shows an **Environment alarms** table that lists the CloudWatch alarms attributed to the environment. Amazon MWAA matches these alarms by CloudWatch namespace and the `Environment` dimension, so the table includes alarms that you author yourself, not only the recommended alarms.



The pane header shows an **In alarm** count of how many of the environment's alarms are currently in the ALARM state. The **Environment alarms** table has the following columns.

Name	Status	Condition	Action	Last updated
The alarm name, which links to the alarm in the CloudWatch console.	The alarm status: In alarm, OK, or Insufficient data.	The alarm condition, shown as *for n datapoints within duration*.	The alarm's action, such as *Notify SNS topic topic*, or - if the alarm has no action.	The date and time when the alarm was last updated.

> **Note**
>
> To view the alarms in this table, you must have the `cloudwatch:DescribeAlarms` permission.

If the environment has no alarms, the **Environment alarms** table is empty and provides guidance to create the recommended alarms to start monitoring your environment.

The **Alarms** pane provides the following two buttons:

- **Create recommended alarms** – Opens a CloudFormation quick-create stack in a new tab. The stack is named `MWAA-alarms-`*`environment-name`* and creates a standard set of recommended alarms for the environment.
- **Manage in CloudWatch** – Opens the alarms page in the CloudWatch console.

To create the recommended alarms

1. Open the Amazon MWAA console, and then choose your environment to open its details page.

2. In the **Monitoring** view, locate the **Alarms** pane.
3. Choose **Create recommended alarms**. The CloudFormation quick-create stack opens in a new browser tab.
4. Review the stack in the CloudFormation console. If prompted, select the acknowledgment that CloudFormation might create IAM resources.
5. Choose **Create stack**.

After CloudFormation creates the stack, the alarms appear in the **Environment alarms** table and in the CloudWatch console. For more information about what each alarm state means, see Alarm states overview.

Recommended alarms

When you choose **Create recommended alarms**, the CloudFormation template creates the following alarms. CloudWatch evaluates all of these alarms over 5-minute (300-second) periods.

Recommended alarms created by the CloudFormation template

Alarm	Metric	Threshold condition
Scheduler heartbeat	`SchedulerHeartbeat`	Sum < 1 for 2 of 2 datapoints; missing data is treated as in alarm.
Oldest queued task age	`ApproximateAgeOfOldestTask`	Maximum > 1800 seconds for 3 of 3 datapoints.
Scheduler CPU utilization	`CPUUtilization` (Scheduler)	Average > 95% for 3 of 3 datapoints.
Scheduler memory utilization	`MemoryUtilization` (Scheduler)	Average > 95% for 3 of 3 datapoints.
Worker CPU utilization	`CPUUtilization` (BaseWorker)	Average > 95% for 6 of 6 datapoints.
Worker memory utilization	`MemoryUtilization` (BaseWorker)	Average > 95% for 6 of 6 datapoints.

Alarm	Metric	Threshold condition
Web server CPU utilization	`CPUUtilization` (WebServer)	Average > 95% for 3 of 3 datapoints.
Web server memory utilizati on	`MemoryUtilization` (WebServer)	Average > 95% for 3 of 3 datapoints.
Database CPU utilization	`CPUUtilization` (WRITER)	Average > 95% for 3 of 3 datapoints.
Database freeable memory	`FreeableMemory` (WRITER)	Average < 512 MiB for 3 of 3 datapoints.

Alarm states overview

A metric alarm has the following possible states:

- `OK` – The metric or expression is within the defined threshold.
- `ALARM` – The metric or expression is outside of the defined threshold.
- `INSUFFICIENT_DATA` – The alarm has just started, the metric is not available, or not enough data is available for the metric to determine the alarm state.

Example custom dashboards and alarms

You can define your own dashboards and alarms as infrastructure as code to build custom dashboards programmatically or to standardize monitoring across multiple Amazon MWAA environments. As an alternative to the built-in dashboard on the environment details page, the following tutorial and CloudFormation template create a custom monitoring dashboard. This dashboard displays charts of selected metrics for your Amazon MWAA environment.

About these metrics

The following list describes each of the metrics created in the custom dashboard by the tutorial and template definitions in this section.

- *QueuedTasks* - The number of tasks with queued state. Corresponds to the `executor.queued_tasks` Apache Airflow metric.

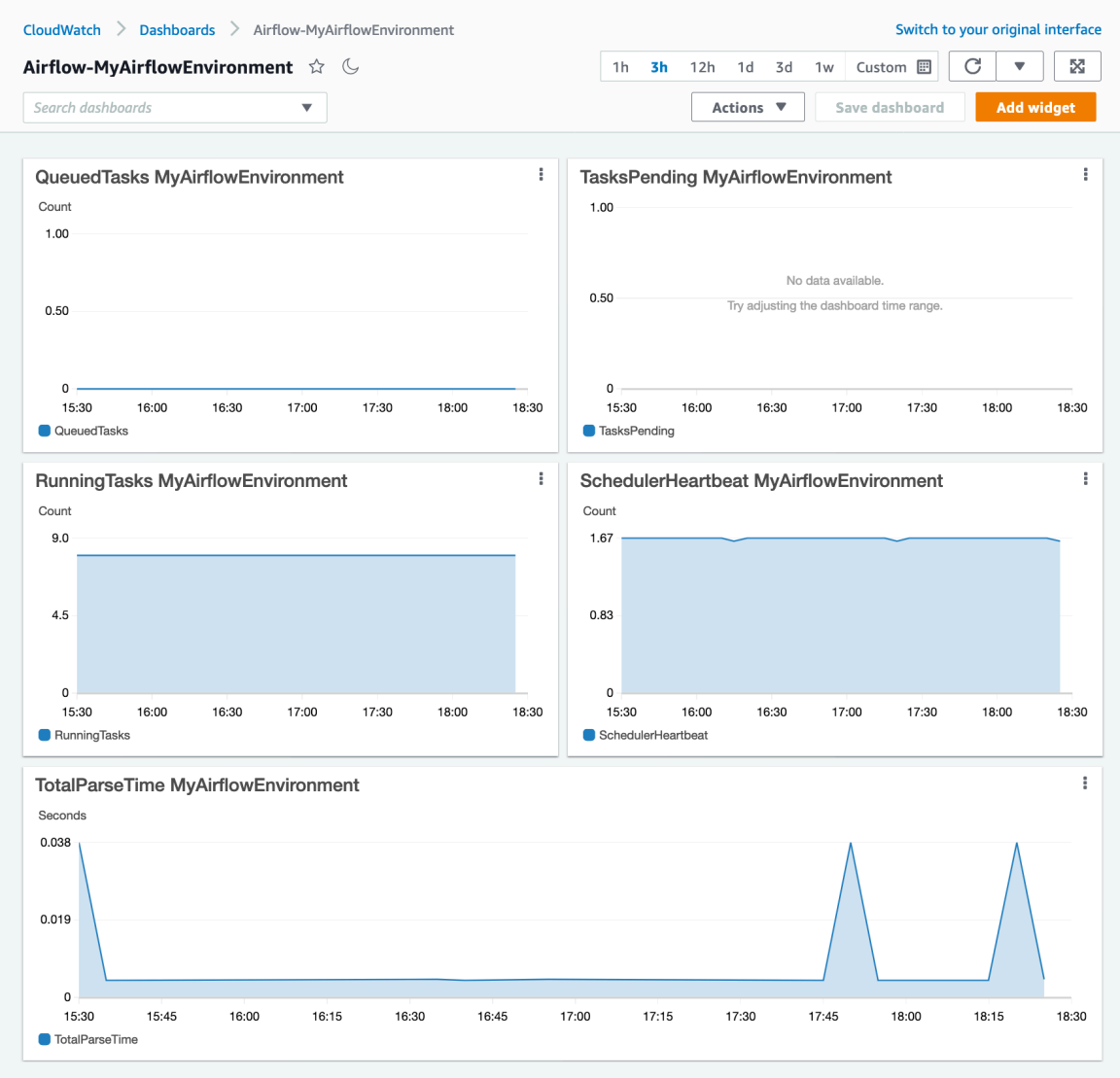
- *TasksPending* - The number of tasks pending in executor. Corresponds to the `scheduler.tasks.pending` Apache Airflow metric.

  > **Note**
  >
  > Does not apply to Apache Airflow v2.2 and later.

- *RunningTasks* - The number of tasks running in executor. Corresponds to the `executor.running_tasks` Apache Airflow metric.
- *SchedulerHeartbeat* - The number of check-ins Apache Airflow performs on the scheduler job. Corresponds to the `scheduler_heartbeat` Apache Airflow metrics.
- *TotalParseTime* - The number of seconds taken to scan and import all DAG files once. Corresponds to the `dag_processing.total_parse_time` Apache Airflow metric.

About the dashboard

The following image displays the monitoring dashboard created by the tutorial and template definition in this section.



Using AWS tutorials

You can use the following AWS tutorial to automatically create a health status dashboard for any Amazon MWAA environments that are currently deployed. It also creates CloudWatch alarms for unhealthy workers and scheduler heartbeat failures across all Amazon MWAA environments.

- CloudWatch Dashboard Automation for Amazon MWAA

Using CloudFormation

You can use the CloudFormation template definition in this section to create a monitoring dashboard in CloudWatch, then add alarms on the CloudWatch console to receive notifications when a metric surpasses a particular threshold. To create the stack using this template definition,

refer to Creating a stack on the CloudFormation console. To add an alarm to the dashboard, refer to Using alarms.

```
AWSTemplateFormatVersion: "2010-09-09"
Description: Creates MWAA Cloudwatch Dashboard
Parameters:
  DashboardName:
    Description: Enter the name of the CloudWatch Dashboard
    Type: String
  EnvironmentName:
    Description: Enter the name of the MWAA Environment
    Type: String
Resources:
  BasicDashboard:
    Type: AWS::CloudWatch::Dashboard
    Properties:
      DashboardName: !Ref DashboardName
      DashboardBody:
        Fn::Sub: '{
              "widgets": [
                  {
                      "type": "metric",
                      "x": 0,
                      "y": 0,
                      "width": 12,
                      "height": 6,
                      "properties": {
                          "view": "timeSeries",
                          "stacked": true,
                          "metrics": [
                              [
                                  "AmazonMWAA",
                                  "QueuedTasks",
                                  "Function",
                                  "Executor",
                                  "Environment",
                                  "${EnvironmentName}"
                              ]
                          ],
                          "region": "${AWS::Region}",
                          "title": "QueuedTasks ${EnvironmentName}",
                          "period": 300
                      }
```

```
                },
                {
                    "type": "metric",
                    "x": 0,
                    "y": 6,
                    "width": 12,
                    "height": 6,
                    "properties": {
                        "view": "timeSeries",
                        "stacked": true,
                        "metrics": [
                            [
                                "AmazonMWAA",
                                "RunningTasks",
                                "Function",
                                "Executor",
                                "Environment",
                                "${EnvironmentName}"
                            ]
                        ],
                        "region": "${AWS::Region}",
                        "title": "RunningTasks ${EnvironmentName}",
                        "period": 300
                    }
                },
                {
                    "type": "metric",
                    "x": 12,
                    "y": 6,
                    "width": 12,
                    "height": 6,
                    "properties": {
                        "view": "timeSeries",
                        "stacked": true,
                        "metrics": [
                            [
                                "AmazonMWAA",
                                "SchedulerHeartbeat",
                                "Function",
                                "Scheduler",
                                "Environment",
                                "${EnvironmentName}"
                            ]
                        ],
```

```
                "region": "${AWS::Region}",
                "title": "SchedulerHeartbeat ${EnvironmentName}",
                "period": 300
            }
        },
        {
            "type": "metric",
            "x": 12,
            "y": 0,
            "width": 12,
            "height": 6,
            "properties": {
                "view": "timeSeries",
                "stacked": true,
                "metrics": [
                    [
                        "AmazonMWAA",
                        "TasksPending",
                        "Function",
                        "Scheduler",
                        "Environment",
                        "${EnvironmentName}"
                    ]
                ],
                "region": "${AWS::Region}",
                "title": "TasksPending ${EnvironmentName}",
                "period": 300
            }
        },
        {
            "type": "metric",
            "x": 0,
            "y": 12,
            "width": 24,
            "height": 6,
            "properties": {
                "view": "timeSeries",
                "stacked": true,
                "region": "${AWS::Region}",
                "metrics": [
                    [
                        "AmazonMWAA",
                        "TotalParseTime",
                        "Function",
```

```
                            "DAG Processing",
                            "Environment",
                            "${EnvironmentName}"
                        ]
                    ],
                    "title": "TotalParseTime  ${EnvironmentName}",
                    "period": 300
                }
            }
        ]
    }'
```

Deleting metrics and dashboards

If you delete an Amazon MWAA environment, the corresponding dashboard is also deleted. CloudWatch metrics are stored for fifteen (15) months and can not be deleted. The CloudWatch console limits the search of metrics to two (2) weeks after a metric is last ingested to ensure that the most up to date instances are shown for your Amazon MWAA environment. To learn more, refer to Amazon CloudWatch FAQs.

What's next?

- Learn how to create a DAG that queries the Amazon Aurora PostgreSQL metadata database for your environment and publishes custom metrics to CloudWatch in Using a DAG to write custom metrics in CloudWatch.

Apache Airflow environment metrics in CloudWatch

Apache Airflow v2 and v3 are already set-up to collect and send StatsD metrics for an Amazon Managed Workflows for Apache Airflow environment to Amazon CloudWatch. The complete list of metrics Apache Airflow sends is available on the Metrics page in the *Apache Airflow reference guide*. This page describes the Apache Airflow metrics available in CloudWatch and how to access metrics in the CloudWatch console.

Contents

- Terms
- Dimensions
- Accessing metrics in the CloudWatch console

- Apache Airflow metrics available in CloudWatch
  - Apache Airflow Counters
  - Apache Airflow Gauges
  - Apache Airflow Timers
- Choosing which metrics are reported
- What's next?

Terms

Namespace

A namespace is a container for the CloudWatch metrics of an AWS service. For Amazon MWAA, the namespace is *AmazonMWAA*.

CloudWatch metrics

A CloudWatch metric represents a time-ordered set of data points that are specific to CloudWatch.

Apache Airflow metrics

The Metrics specific to Apache Airflow.

Dimension

A dimension is a name/value pair that is part of the identity of a metric.

Unit

A statistic has a unit of measure. For Amazon MWAA, units include *Count*, *Seconds*, and *Milliseconds*. For Amazon MWAA, units are set based on the units in the original Airflow metrics.

Dimensions

This section describes the CloudWatch *Dimensions* grouping for Apache Airflow metrics in CloudWatch.

Dimension	Description
DAG	Indicates a specific Apache Airflow DAG name.
DAG Filename	Indicates a specific Apache Airflow DAG file name.
Function	This dimension is used to improve the grouping of metrics in CloudWatch.
Job	Indicates an Apache Airflow job run by the scheduler. Always has a value of `Job`.
Operator	Indicates a specific Apache Airflow operator.
Pool	Indicates a specific Apache Airflow worker pool.
Task	Indicates a specific Apache Airflow task.
HostName	Indicates the hostname for a specific running Apache Airflow process.

Accessing metrics in the CloudWatch console

This section describes how to access performance metrics in CloudWatch for a specific DAG.

To access performance metrics for a dimension

1. Open the Metrics page on the CloudWatch console.
2. Select your AWS Region.
3. Choose the **AmazonMWAA** namespace.
4. In the **All metrics** tab, select a dimension. For example, *DAG, Environment*.
5. Choose a CloudWatch metric for a dimension. For example, *TaskInstanceSuccesses* or *TaskInstanceDuration*. Choose **Graph all search results**.
6. Choose the **Graphed metrics** tab to access performance statistics for Apache Airflow metrics, such as *DAG, Environment, Task*.

Apache Airflow metrics available in CloudWatch

This section describes the Apache Airflow metrics and dimensions sent to CloudWatch.

Apache Airflow Counters

The Apache Airflow metrics in this section contain data about Apache Airflow *Counters*.

CloudWatch metric	Apache Airflow metric	Unit	Dimension	
SLAMissed **Note** Only available for Apache Airflow v2.4.3 to v2.10.3.	sla_missed	Count	Function, Scheduler	
FailedSLACallback	sla_callback_notification_failure	Count	Function, Scheduler	

CloudWatch metric	Apache Airflow metric	Unit	Dimension
Note Only available for Apache Airflow v2.4.3 to v2.10.3.			
Updates **Note** Available for Apache Airflow v2.6.3 and later.	dataset.updates	Count	Function, Scheduler
Orphaned **Note** Available for Apache Airflow v2.6.3 and later.	dataset.orphaned	Count	Function, Scheduler
FailedCeleryTaskExecution **Note** Available for Apache Airflow v2.4.3 and later.	celery.execute_command.failure	Count	Function, Celery

CloudWatch metric	Apache Airflow metric	Unit	Dimension	
FilePathQueueUpdateCount **Note** Available for Apache Airflow v2.6.3 and later.	dag_processing.file_path_queue_update_count	Count	Function, Scheduler	
CriticalSectionBusy	scheduler.critical_section_busy	Count	Function, Scheduler	
DagBagSize	dagbag_size	Count	Function, DAG Processing	
DagCallbackExceptions	dag.callback_exceptions	Count	DAG, All	
FailedSLAEmailAttempts **Note** Not available for Apache Airflow v3.0.6 and later.	sla_email_notification_failure	Count	Function, Scheduler	

CloudWatch metric	Apache Airflow metric	Unit	Dimension	
TaskInstanceFinished	ti.finish.{dag_id}.{task_id}.{state}	Count	DAG, {dag_id} Task, {task_id} State, {state}	
JobEnd	{job_name}_end	Count	Job, {job_name}	
JobHeartbeatFailure	{job_name}_heartbeat_failure	Count	Job, {job_name}	
JobStart	{job_name}_start	Count	Job, {job_name}	
ManagerStalls	dag_processing.manager_stalls	Count	Function, DAG Processing	
OperatorFailures	operator_failures_{operator_name}	Count	Operator, {operator_name}	

CloudWatch metric	Apache Airflow metric	Unit	Dimension
OperatorSuccesses	operator_successes_{operator_name}	Count	Operator, {operator_name}
OtherCallbackCount **Note** Available in Apache Airflow v2.6.3 and later.	dag_processing.other_callback_count	Count	Function, Scheduler
Processes	dag_processing.processes	Count	Function, DAG Processing
SchedulerHeartbeat	scheduler_heartbeat	Count	Function, Scheduler
StartedTaskInstances	ti.start.{dag_id}.{task_id}	Count	DAG, All Task, All

CloudWatch metric	Apache Airflow metric	Unit	Dimension	
SlaCallbackCount	dag_processing.sla_callback_count **Note** Available for Apache Airflow v2.6.3 and later.	Count	Function, Scheduler	
TasksKilledExternally	scheduler.tasks.killed_externally	Count	Function, Scheduler	
TaskTimeoutError	celery.task_timeout_error	Count	Function, Celery	
TaskInstanceCreatedUsingOperator	task_instance_created-{operator_name}	Count	Operator, {operator_name}	

CloudWatch metric	Apache Airflow metric	Unit	Dimension
TaskInstancePreviouslySucceeded	previously_succeeded	Count	DAG, All Task, All
TaskInstanceFailures	ti_failures	Count	DAG, All Task, All
TaskInstanceSuccesses	ti_successes	Count	DAG, All Task, All
TaskRemovedFromDAG	task_removed_from_dag.{dag_id}	Count	DAG, {dag_id}
TaskRestoredToDAG	task_restored_to_dag.{dag_id}	Count	DAG, {dag_id}
TriggersSucceeded **Note** Available for Apache Airflow v2.7.2 and later.	triggers.succeeded	Count	Function, Trigger

CloudWatch metric	Apache Airflow metric	Unit	Dimension
TriggersFailed **Note** Available for Apache Airflow v2.7.2 and later.	triggers.failed	Count	Function, Trigger
TriggersBlockedMainThread **Note** Available for Apache Airflow v2.7.2 and later.	triggers.blocked_main_thread	Count	Function, Trigger
TriggerHeartbeat **Note** Available for Apache Airflow v2.8.1 and later.	triggerer_heartbeat	Count	Function, Triggerer

CloudWatch metric	Apache Airflow metric	Unit	Dimension
TaskInstanceCreatedUsingOperator	airflow.task_instance_created_{operator_name} **Note** Available for Apache Airflow v2.7.2 and later.	Count	Operator, {operator_name}
ZombiesKilled	zombies_killed	Count	DAG, All Task, All

Apache Airflow Gauges

The Apache Airflow metrics in this section contain data about Apache Airflow *Gauges*.

CloudWatch metric	Apache Airflow metric	Unit	Dimension
DAGFileRefreshError	dag_file_refresh_error	Count	Function, DAG Processing

CloudWatch metric	Apache Airflow metric	Unit	Dimension
ImportErrors	dag_processing.import_errors	Count	Function, DAG Processing
Exception Failures	smart_sensor_operator.exception_failures	Count	Function, Smart Sensor Operator
ExecutedTasks	smart_sensor_operator.executed_tasks	Count	Function, Smart Sensor Operator
InfraFailures	smart_sensor_operator.infra_failures	Count	Function, Smart Sensor Operator
LoadedTasks	smart_sensor_operator.loaded_tasks	Count	Function, Smart Sensor Operator
TotalParseTime	dag_processing.total_parse_time	Seconds	Function, DAG Processing

CloudWatch metric	Apache Airflow metric	Unit	Dimension
Triggered DagRuns **Note** Available in Apache Airflow v2.6.3 and later.	dataset.triggered_dagruns	Count	Function, Scheduler
TriggersRunning **Note** Available in Apache Airflow v2.7.2 and later.	triggers.running.*{hostname}*	Count	Function, Trigger HostName, *{hostname}*

CloudWatch metric	Apache Airflow metric	Unit	Dimension
PoolDeferredSlots **Note** Available in Apache Airflow v2.7.2 and later.	pool.deferred_slots.{pool_name}	Count	Pool, {pool_name}
DAGFileProcessingLastRunSecondsAgo	dag_processing.last_run.seconds_ago.{dag_filename}	Seconds	DAG Filename, {dag_filename}
OpenSlots	executor.open_slots	Count	Function, Executor
OrphanedTasksAdopted	scheduler.orphaned_tasks.adopted	Count	Function, Scheduler
OrphanedTasksCleared	scheduler.orphaned_tasks.cleared	Count	Function, Scheduler
PokedExceptions	smart_sensor_operator.poked_exception	Count	Function, Smart Sensor Operator

CloudWatch metric	Apache Airflow metric	Unit	Dimension
PokedSuccess	smart_sensor_operator.poked_success	Count	Function, Smart Sensor Operator
PokedTasks	smart_sensor_operator.poked_tasks	Count	Function, Smart Sensor Operator
PoolFailures	pool.open_slots.{pool_name}	Count	Pool, {pool_name}
PoolStarvingTasks	pool.starving_tasks.{pool_name}	Count	Pool, {pool_name}
PoolOpenSlots	pool.open_slots.{pool_name}	Count	Pool, {pool_name}
PoolQueuedSlots	pool.queued_slots.{pool_name}	Count	Pool, {pool_name}
PoolRunningSlots	pool.running_slots.{pool_name}	Count	Pool, {pool_name}
Processor Timeouts	dag_processing.processor_timeouts	Count	Function, DAG Processing
QueuedTasks	executor.queued_tasks	Count	Function, Executor

CloudWatch metric	Apache Airflow metric	Unit	Dimension
RunningTasks	executor.running_tasks	Count	Function, Executor
TasksExecutable	scheduler.tasks.executable	Count	Function, Scheduler
TasksPending **Note** Does not apply to Apache Airflow v2.2 and later.	scheduler.tasks.pending	Count	Function, Scheduler
TasksRunning	scheduler.tasks.running	Count	Function, Scheduler
TasksStarving	scheduler.tasks.starving	Count	Function, Scheduler
TasksWithoutDagRun	scheduler.tasks.without_dagrun	Count	Function, Scheduler

CloudWatch metric	Apache Airflow metric	Unit	Dimension
DAGFilePr ocessingL astNumOfD bQueries **Note** Available in Apache Airflow v2.10.1 and later.	dag_processing.las t_num_of_db_queries. {dag_filename}	Count	DAG Filename, {dag_file name}
PoolSched uledSlots **Note** Available in Apache Airflow v2.10.1 and later.	pool.scheduled_slots. {pool_name}	Count	Pool, {pool_name}

CloudWatch metric	Apache Airflow metric	Unit	Dimension
TaskCpuUsage **Note** Available in Apache Airflow v2.10.1 and later.	cpu.usage.{dag_id}.{task_id}	Percent	DAG, {dag_id} Task, {task_id}
TaskMemoryUsage **Note** Available in Apache Airflow v2.10.1 and later.	mem.usage.{dag_id}.{task_id}	Percent	DAG, {dag_id} Task, {task_id}

Apache Airflow Timers

The Apache Airflow metrics in this section contain data about Apache Airflow *Timers*.

CloudWatch metric	Apache Airflow metric	Unit	Dimension

CloudWatch metric	Apache Airflow metric	Unit	Dimension
CollectDBDags	collect_db_dags	Milliseconds	Function, DAG Processing
CriticalSectionDuration	scheduler.critical_section_duration	Milliseconds	Function, Scheduler
CriticalSectionQueryDuration **Note** Available for Apache Airflow v2.5.1 and later.	scheduler.critical_section_query_duration	Milliseconds	Function, Scheduler
DAGDependencyCheck	dagrun.dependency-check.{dag_id}	Milliseconds	DAG, {dag_id}
DAGDurationFailed	dagrun.duration.failed.{dag_id}	Milliseconds	DAG, {dag_id}

CloudWatch metric	Apache Airflow metric	Unit	Dimension
DAGDurationSuccess	dagrun.duration.success.{dag_id}	Milliseconds	DAG, {dag_id}
DAGFileProcessingLastDuration	dag_processing.last_duration.{dag_filename}	Seconds	DAG Filename, {dag_filename}
DAGScheduleDelay	dagrun.schedule_delay.{dag_id}	Milliseconds	DAG, {dag_id}
FirstTaskSchedulingDelay	dagrun.{dag_id}.first_task_scheduling_delay	Milliseconds	DAG, {dag_id}
SchedulerLoopDuration **Note** Available for Apache Airflow v2.5.1 and later.	scheduler.scheduler_loop_duration	Milliseconds	Function, Scheduler

CloudWatch metric	Apache Airflow metric	Unit	Dimension
TaskInstanceDuration	dag.{dag_id}.{task_id}.duration	Milliseconds	DAG, {dag_id} Task, {task_id}
TaskInstanceQueuedDuration	dag.{dag_id}.{ta .queued_duration **Note** Available for Apache Airflow v2.7.2 and later.	Milliseconds	DAG, {dag_id} Task, {task_id}
TaskInstanceScheduledDuration **Note** Available for Apache Airflow v2.7.2 and later.	dag.{dag_id}.{ta .scheduled_duration	Milliseconds	DAG, {dag_id} Task, {task_id}

Choosing which metrics are reported

You can choose which Apache Airflow metrics are emitted to CloudWatch, or blocked by Apache Airflow, using the following Amazon MWAA configuration options:

- **`metrics.metrics_allow_list`** — A list of comma-separated prefixes you can use to select which metrics are emitted to CloudWatch by your environment. Use this option if you want Apache Airflow to not send all available metrics and instead select a subset of elements. For example, `scheduler,executor,dagrun`.
- **`metrics.metrics_block_list`** — A list of comma-separated prefixes to filter out metrics that start with the elements of the list. For example, `scheduler,executor,dagrun`.

If you configure both `metrics.metrics_allow_list` and `metrics.metrics_block_list`, Apache Airflow ignores `metrics.metrics_block_list`. If you configure `metrics.metrics_block_list` but not `metrics.metrics_allow_list`, Apache Airflow filters out the elements you specify in `metrics.metrics_block_list`.

> **Note**
>
> The `metrics.metrics_allow_list` and `metrics.metrics_block_list` configuration options only apply to Apache Airflow v2.6.3 and later. For previous version of Apache Airflow use `metrics.statsd_allow_list` and `metrics.statsd_block_list` instead.

What's next?

- Explore the Amazon MWAA API operation used to publish environment health metrics at PublishMetrics.

Container, queue, and database metrics for Amazon MWAA

In addition to Apache Airflow metrics, you can monitor the underlying components of your Amazon Managed Workflows for Apache Airflow environments using CloudWatch, which collects raw data and processes data into readable, near real-time metrics. With these environment metrics, you'll have greater visibility into key performance indicators to help you appropriately size your

environments and debug issues with your workflows. These metrics apply to all supported Apache Airflow versions on Amazon MWAA.

Amazon MWAA will provide CPU and memory utilization for each Amazon Elastic Container Service (Amazon ECS) container and Amazon Aurora PostgreSQL instance, and Amazon Simple Queue Service (Amazon SQS) metrics for the number of messages and the age of the oldest message, Amazon Relational Database Service (Amazon RDS) metrics for database connections, disk queue depth, free local storage, write operations, latency, and throughput, and Amazon RDS Proxy metrics. These metrics also include the number of base workers, additional workers, schedulers, and webservers.

These statistics are kept for 15 months, so that you can access historical information and gain a better perspective on why a schedule is failing, and troubleshoot underlying issues. You can also set alarms that monitor for certain thresholds, and send notifications or take actions when those thresholds are met. For more information, refer to the Amazon CloudWatch User Guide.

Topics

- Terms
- Dimensions
- Accessing metrics in the CloudWatch console
- List of metrics

Terms

Namespace

A namespace is a container for the CloudWatch metrics of an AWS service. For Amazon MWAA, the namespace is `AWS/MWAA`.

CloudWatch metrics

A CloudWatch metric represents a time-ordered set of data points that are specific to CloudWatch.

Dimension

A dimension is a name/value pair that is part of the identity of a metric.

Unit

A statistic has a unit of measure. For Amazon MWAA, units include *Count*.

Dimensions

This section describes the CloudWatch dimensions grouping for Amazon MWAA metrics in CloudWatch.

Dimension	Description
Cluster	Metrics for the minimum three Amazon ECS container that an Amazon MWAA environment uses to run Apache Airflow components: scheduler, worker, and web server.
Queue	Metrics for the Amazon SQS queues that decouple the scheduler from workers. When workers read the messages, they are considered in-flight and not available for other workers. Messages become available for other workers to read if they are not deleted before the 12-hour visibility timeout.
Database	Metrics the Aurora clusters used by Amazon MWAA. This includes metrics for the primary database instance and a read replica to support the read operations. Amazon MWAA publishes database metrics for both READER and WRITER instances.

Accessing metrics in the CloudWatch console

This section describes how to access your Amazon MWAA metrics in CloudWatch.

To access performance metrics for a dimension

1. Open the Metrics page on the CloudWatch console.
2. Select your AWS Region.
3. Choose the **AWS/MWAA** namespace.
4. In the **All metrics** tab, choose a dimension. For example, **Cluster**.
5. Choose a CloudWatch metric for a dimension. For example, *NumSchedulers* or *CPUUtilization*. Then, choose **Graph all search results**.
6. Choose the **Graphed metrics** tab to access performance metrics.

List of metrics

The following tables list the cluster, queue, and database service metrics for Amazon MWAA. To access descriptions for metrics directly emitted from Amazon ECS, Amazon SQS, or Amazon RDS, choose the respective documentation link.

Topics

- Cluster metrics
- Database metrics
- Queue metrics
- Application Load Balancer metrics

Cluster metrics

The following metrics apply to each scheduler, base worker, additional worker, and web server. For more information about these metrics, see Amazon ECS monitoring and Amazon ECS Container Insights.

Namespace	Metric	Unit
AWS/MWAA	CPUUtilization	Percent
AWS/MWAA	MemoryUtilization	Percent

Namespace	Metric	Unit
AWS/MWAA	EphemeralStorageUtilized	Megabytes
AWS/MWAA	EphemeralStorageReserved	Megabytes
AWS/MWAA	NetworkTxBytes	Bytes/Second
AWS/MWAA	NetworkRxBytes	Bytes/Second

Evaluating the number of additional worker and webserver containers

You can use the component metrics provided in the **Cluster** dimension, as described in the following procedure, to assess how many additional workers, or webservers, an environment is using at a given point in time. You can do this by graphing either the **CPUUtilization** or the **MemoryUtilization** metric and setting the statistic type to **Sample Count**. The resulting value is the total number of `RUNNING` tasks for the `AdditionalWorker` component. Understanding the number of additional worker instances utilized by your environment can help you gauge how your environment scales and you can use to optimize the number of additional workers.

Workers

To evaluate the number of additional workers using the AWS Management Console

1. Choose the **AWS/MWAA** namespace.
2. In the **All metrics** tab, choose the **Cluster** dimension.
3. In the **Cluster** dimension, for the **AdditionalWorker**, choose either the **CPUUtilization** or the **MemoryUtilization** metric.
4. On the **Graphed metrics** tab, set **Period** to **1 Minute** and **Statistic** to **Sample Count**.

webservers

To evaluate the number of additional webservers using the AWS Management Console

1. Choose the **AWS/MWAA** namespace.
2. In the **All metrics** tab, choose the **Cluster** dimension.

3. In the **Cluster** dimension, for the **AdditionalWebservers**, choose either the **CPUUtilization** or the **MemoryUtilization** metric.
4. On the **Graphed metrics** tab, set **Period** to **1 Minute** and **Statistic** to **Sample Count**.

For more information, refer to Service RUNNING task count in the *Amazon Elastic Container Service Developer Guide*.

Database metrics

The following metrics apply to each database instance associated with the Amazon MWAA environment.

Namespace	Metric	Unit
AWS/MWAA	CPUUtilization	Percent
AWS/MWAA	DatabaseConnections	Count
AWS/MWAA	DiskQueueDepth	Count
AWS/MWAA	FreeableMemory	Bytes
AWS/MWAA	FreeLocalStorage	Bytes
AWS/MWAA	VolumeWriteIOPS	Count per five minutes
AWS/MWAA	WriteIOPS	Count per second
AWS/MWAA	WriteLatency	Seconds
AWS/MWAA	WriteThroughput	Bytes per second

Queue metrics

For more information about units and descriptions for the following queue metrics, refer to Available CloudWatch metrics for Amazon SQS in the *Amazon Simple Queue Service Developer Guide*.

Namespace	Metric	Unit
AWS/MWAA	ApproximateAgeOfOldestTask	Seconds
AWS/MWAA	RunningTasks	Count
AWS/MWAA	QueuedTasks	Count

Application Load Balancer metrics

Application Load Balancer metrics apply to the web servers running in your environment. Amazon MWAA uses these metrics to for scaling your web servers based on the amount of traffic. For more information about units and descriptions for the following load balancer metrics, refer to CloudWatch metrics for your Application Load Balancer in the *Application Load Balancers User Guide*.

Namespace	Metric	Unit
AWS/MWAA	ActiveConnectionCount	Count

Security in Amazon Managed Workflows for Apache Airflow

Cloud security at AWS is the highest priority. As an AWS customer, you benefit from a data center and network architecture that is built to meet the requirements of the most security-sensitive organizations.

Security is a shared responsibility between AWS and you (the customer). The shared responsibility model describes this as security *of* the cloud and security *in* the cloud:

- **Security of the cloud** – AWS is responsible for protecting the infrastructure that runs AWS services in the AWS Cloud. AWS also provides you with services that you can use securely. Third-party auditors regularly test and verify the effectiveness of our security as part of the AWS Compliance Programs. To learn about the compliance programs that apply to Amazon MWAA, refer to AWS Services in Scope by Compliance Program.
- **Security in the cloud** – Your responsibility is determined by the AWS service that you use. You are also responsible for other factors including the sensitivity of your data, your company's requirements, and applicable laws and regulations.

This documentation helps you understand how to apply the shared responsibility model when using Amazon Managed Workflows for Apache Airflow. Use it to configure Amazon MWAA to meet your security and compliance objectives. You also learn how to use other AWS services that help you monitor and secure your Amazon MWAA resources.

In this section:

- AWS shared responsibility model for Amazon MWAA
- Data Protection in Amazon Managed Workflows for Apache Airflow
- AWS Identity and Access Management
- Using IAM Identity Center with Amazon Managed Workflows for Apache Airflow
- Compliance Validation for Amazon Managed Workflows for Apache Airflow
- Resilience in Amazon Managed Workflows for Apache Airflow
- Infrastructure Security in Amazon MWAA
- AWS PrivateLink for Amazon MWAA

- Configuration and Vulnerability Analysis in Amazon MWAA
- Security best practices on Amazon MWAA

AWS shared responsibility model for Amazon MWAA

This guidance applies to Amazon MWAA Provisioned environments. In a Provisioned environment, you select the environment class and configure the worker and web server scaling limits.

Security and compliance is a shared responsibility between AWS and you. The shared responsibility model describes this as security *of* the cloud and security *in* the cloud. For Amazon MWAA, this model extends beyond security into operational responsibility. We operate, manage, and control the underlying components. These range from the host operating system and virtualization layer down to the physical security of the facilities in which the service operates.

You are responsible for managing the content that you deploy to your environment. On Amazon MWAA, that content includes your DAG code and the Python dependencies that you declare in `requirements.txt`. It also includes your custom plugins and any startup script that you supply. Your responsibility covers the configuration, network, and permissions that you choose. You are also responsible for the day-to-day operation of the workflows that run on the environment. Amazon MWAA runs the content that you supply as you supply it. Amazon MWAA does not inspect, validate, or modify it.

> **Note**
>
> Although AWS operates the infrastructure for your environment, the data that your workflows write to the metadata database directly affects how the service performs. Unbounded metadata growth can degrade scheduler and webserver response times. You are responsible for managing the volume of data in your environment.

Infrastructure that AWS manages

We are responsible for protecting the infrastructure that runs AWS services in the AWS Cloud. For Amazon MWAA, this includes the following:

- **Apache Airflow setup** – Amazon MWAA sets up Apache Airflow for you. It uses the same Apache Airflow user interface and open-source code available on the internet. For more information, see What Is Amazon Managed Workflows for Apache Airflow?.

- **Compute for the managed components** – Amazon MWAA provisions and operates the AWS Fargate compute for your environment. This compute runs the Apache Airflow scheduler, workers, and webserver. Amazon MWAA also provisions and operates the Apache Airflow metadata database. For each environment, Amazon MWAA creates an AWS owned Amazon VPC that hosts the webserver and the metadata database. The scheduler and workers connect to the private subnets in the Amazon VPC for your environment, which you own and configure. For more information, see Explore Amazon MWAA network architecture in the *Amazon MWAA Migration Guide*.
- **The Apache Airflow metadata database** – Amazon MWAA provisions and operates the Apache Airflow metadata database for each environment. You remain responsible for the volume of data that your workflows write to it. For more information, see Operational excellence that you manage.
- **Version images and patching** – Amazon MWAA builds container images that bundle Apache Airflow releases with other common binaries and Python libraries. Amazon MWAA patches those images. An environment keeps using its specified image version until you issue an update environment action to move it to a later version. For more information, see About Amazon MWAA versions. The image definitions are available in the amazon-mwaa-docker-images repository on the GitHub website.
- **Apache Airflow open-source issues** – Amazon MWAA runs open-source Apache Airflow without changing it. The Apache Airflow community fixes bugs in Apache Airflow core and provider packages. Report bugs on the Apache Airflow issues page on the GitHub website. To get those fixes, upgrade your environment to a later supported version. For more information about Apache Airflow versions and environment upgrades, see Apache Airflow versions on Amazon Managed Workflows for Apache Airflow and Changing the Apache Airflow version.
- **Version support lifecycle** – Amazon MWAA publishes the availability and end of support dates for each Apache Airflow version. Amazon MWAA notifies you through the Health Dashboard when an environment in your account runs a version that is nearing the end of support. For the current support commitment and dates, see End-of-support versions.
- **Automatic scaling of the managed components** – Amazon MWAA scales workers and the webserver in response to demand. Scaling stays within the limits that you configure. You choose the environment class, the worker and webserver scaling settings, and the scheduler count. For more information, see Configuring Amazon MWAA worker automatic scaling, Configuring Amazon MWAA webserver automatic scaling, and Configuring the Amazon MWAA environment class.

- **Data encryption** – Amazon MWAA encrypts your data at rest and in transit. Amazon MWAA also attaches the required grants to a customer-managed KMS key on your behalf. For more information, see Data Protection in Amazon Managed Workflows for Apache Airflow.
- **Metrics and logs** – Amazon MWAA publishes environment metrics to Amazon CloudWatch. Amazon MWAA delivers Apache Airflow logs to CloudWatch Logs for the log types that you enable. For more information, see Monitoring overview on Amazon MWAA.
- **Compliance programs** – Third-party auditors regularly test and verify the effectiveness of AWS security. For more information, see Compliance Validation for Amazon Managed Workflows for Apache Airflow.

Workflow code and content that you manage

You are responsible for maintaining control over the content that you host on this infrastructure. This content includes the following:

- **DAG code** – The DAG definitions that you copy to the `dags` folder in your Amazon S3 bucket. You are responsible for what those DAGs access and the operations that they perform. For more information, see Adding or updating DAGs.
- **Python dependencies** – The libraries that you install by uploading a `requirements.txt` file to your Amazon S3 bucket. This includes the security and compatibility of every package that your workflows require. Incompatible or resource-intensive packages can degrade scheduler and worker performance or prevent containers from starting. For more information, see Installing Python dependencies and Managing Python dependencies in requirements.txt.
- **Custom plugins** – The contents of the `plugins.zip` file that you upload to your Amazon S3 bucket. We recommend verifying the contents of the file before you upload it. For more information, see Installing custom plugins.
- **Startup scripts** – The shell script that Amazon MWAA runs on each component at startup. You also specify the Amazon S3 version ID for the script. For more information, see Using a startup script with Amazon MWAA.
- **Testing** – Validating your DAGs, custom plugins, and Python dependencies before you deploy them to an environment. We recommend testing Apache Airflow version upgrades in a development environment before you apply them to production. You can use the amazon-mwaa-docker-images container image on the GitHub website to develop and test locally.
- **Your Amazon S3 bucket and its contents** – The bucket policy and object ACLs for the bucket associated with your environment. Amazon MWAA does not back up the bucket contents on your

behalf. We recommend enabling versioning and configuring cross-region replication if your DAGs, plugins, and requirements files require a disaster recovery posture. We also recommend that you don't store other objects in the bucket or use the bucket with another service.

- **Data that you enter** – We strongly recommend that you never put confidential or sensitive information into tags or free-form fields such as a **Name** field.

Operational excellence that you manage

Amazon MWAA operates the managed infrastructure. You retain responsibility for the workflows that run on that infrastructure. The code, dependencies, and configuration that you deploy determine how your environment performs. Amazon MWAA publishes the metrics and logs that describe environment health. Interpreting that data, setting alarms on it, and acting on what it reports **are your responsibility**.

- **Monitoring environment health** – Reviewing the CloudWatch metrics that Amazon MWAA publishes for your environment. These include container, queue, and database metrics that help you detect saturation and degradation. For more information, see Monitoring and metrics for Amazon Managed Workflows for Apache Airflow and Container, queue, and database metrics for Amazon MWAA.
- **Dashboards and alarms** – Creating the CloudWatch dashboards and alarms that notify you when your environment approaches its limits. For more information, see Monitoring dashboards and alarms on Amazon MWAA.
- **Log analysis** – Enabling the Apache Airflow log types that you need. You then analyze scheduler, worker, webserver, and task logs to diagnose workflow failures. For more information, see Accessing Airflow logs in Amazon CloudWatch and Troubleshooting: CloudWatch Logs and CloudTrail errors.
- **CloudWatch service quotas** – Ensuring that the CloudWatch quotas in your account can handle the log volume that your environment produces. Exceeding CloudWatch Logs quotas such as `CreateLogStream` can cause worker degradation. For more information, see CloudWatch Logs quotas.
- **Metadata database maintenance** – Keeping the Apache Airflow metadata database light. Regularly remove metadata that your workflows no longer need. An unbounded metadata database can degrade scheduler and webserver performance. For more information, see Aurora PostgreSQL database cleanup on an Amazon MWAA environment.

- **DAG design and scheduling frequency** – Your DAG count, tasks per DAG, and schedule intervals directly determine load on the managed scheduler. Amazon MWAA does not throttle or reject workloads that exceed the environment's capacity. You must design your DAGs to fit within the resources that you configure.
- **Task resource governance** – Setting task-level timeouts, Apache Airflow pool sizes, and per-DAG concurrency limits. Amazon MWAA does not enforce per-task guardrails on your behalf. Without these settings, a single long-running or high-concurrency DAG can consume all available workers.
- **Capacity and performance tuning** – Choosing the environment class, the worker and webserver scaling limits, and the Apache Airflow configuration options that suit your workload. Automatic scaling operates within the bounds that you configure and does not compensate for workloads that exceed the environment's design capacity. You must right-size the environment class for your workload profile. For more information, see Performance tuning for Apache Airflow on Amazon MWAA and Configuring the Amazon MWAA environment class.
- **Changes that you make through a startup script** – A startup script runs on every component in your environment. It can change the configuration of those components. You are responsible for the changes that you make this way, including any effect they have on environment stability. We recommend testing a startup script before you apply it to a production environment. For more information, see Using a startup script with Amazon MWAA.
- **Environment updates** – Initiating an environment update is your action. Updates trigger container restarts and can affect running workflows. You are responsible for choosing when to update and for assessing the effect on in-progress tasks. We recommend updating during a time window that you choose, when no critical DAGs are running. To update without pausing and unpausing your DAGs, consider graceful updates, which let workers finish in-progress tasks before they shut down. Conditions and limits apply; for more information, see Update an Amazon MWAA environment.
- **First-level triage** – Using available metrics and logs to diagnose issues before you engage AWS Support. The level of detail that AWS Support can use to help you depends on the logging level that you enabled for your environment.
- **Cross-environment orchestration** – Coordinating workflows across multiple Amazon MWAA environments. If you use mechanisms such as the Apache Airflow REST API or external sensors to link environments, you are responsible for that coordination logic.
- **Following the Amazon MWAA best practices** – Applying the documented guidance for performance, dependency management, and environment configuration. For more information, see Best practices for Amazon Managed Workflows for Apache Airflow.

Configuration and access that you manage

You choose how your environment is configured and who can reach it. You are responsible for the following:

- **Keeping your version current** – You are responsible for keeping your Amazon MWAA versions current. Upgrading an environment to a newer Apache Airflow version is an action that you initiate. For more information, see Changing the Apache Airflow version and End-of-support versions.
- **The execution role** – Amazon MWAA can't add or edit permission policies on an existing execution role after an environment is created. You must update the execution role with the additional permissions that your environment needs. For more information, see Amazon MWAA execution role.
- **IAM access control** – The policies that determine who can access your environment and the Apache Airflow UI. Grant permissions to only the resources and actions that users need. For more information, see AWS Identity and Access Management and Accessing an Amazon MWAA environment.
- **Apache Airflow user privileges** – Apache Airflow is not multi-tenant. DAG authors can write DAGs that change Apache Airflow user privileges and interact with the underlying metadata database. We recommend using separate environments for separate teams that have DAG writing access. For more information, see Security best practices in Apache Airflow.
- **Webserver access mode** – Choosing `PUBLIC_ONLY` or `PRIVATE_ONLY` for access to the Apache Airflow UI. This is a security decision that determines whether the webserver endpoint is reachable from the internet. For more information, see Security in your VPC on Amazon MWAA.
- **Network configuration** – Your Amazon VPC, subnets, security groups, and network ACLs. This includes the inbound and outbound rules that direct traffic on your NAT gateways. If you choose to manage your own Amazon VPC endpoints, you're responsible for creating those endpoints. For more information, see Security in your VPC on Amazon MWAA and Managing your own Amazon VPC endpoints on Amazon MWAA.
- **Encryption key choice** – Whether to use an AWS owned KMS key or a customer-managed KMS key. If you use a customer-managed KMS key, you must attach the required policy statements to the key. For more information, see Data Protection in Amazon Managed Workflows for Apache Airflow.
- **Secrets and connections** – The credentials that your workflows use. We recommend storing secrets in Secrets Manager. For more information, see Configuring an Apache Airflow connection using a AWS Secrets Manager secret.

- **Apache Airflow configuration options** – The configuration options that you override on your environment. For more information, see Using Apache Airflow configuration options on Amazon MWAA.

Related resources

- AWS Shared Responsibility Model
- Shared responsibility in the *AWS Well-Architected Framework*
- Best practices for Amazon Managed Workflows for Apache Airflow
- Performance tuning for Apache Airflow on Amazon MWAA
- Monitoring dashboards and alarms on Amazon MWAA
- Container, queue, and database metrics for Amazon MWAA
- Security best practices on Amazon MWAA

Data Protection in Amazon Managed Workflows for Apache Airflow

The AWS shared responsibility model applies to data protection in Amazon Managed Workflows for Apache Airflow. As described in this model, AWS is responsible for protecting the global infrastructure that runs all of the AWS Cloud. You're responsible for maintaining control over your content hosted on this infrastructure. This content includes the security configuration and management tasks for the AWS services you use. For more information about data privacy, refer to the Data Privacy FAQ. For information about data protection in Europe, refer to the AWS Shared Responsibility Model and GDPR blog post on the *AWS Security Blog*.

For data protection purposes, we recommend that you protect AWS account credentials and set up individual user accounts with AWS Identity and Access Management (IAM). That way each user is given only the permissions necessary to fulfill their job duties. We also recommend that you secure your data in the following ways:

- Use multi-factor authentication (MFA) with each account.
- Use SSL/TLS to communicate with AWS resources. We recommend TLS 1.2 or later.
- Set up API and user activity logging with AWS CloudTrail.
- Use AWS encryption solutions, along with all default security controls within AWS services.

- Use advanced managed security services such as Amazon Macie, which assists in discovering and securing personal data that is stored in Amazon S3.

We strongly recommend that you never put confidential or sensitive information, such as your customers' email addresses, into tags or free-form fields such as a **Name** field. This includes when you work with Amazon MWAA or other AWS services using the console, API, AWS CLI, or AWS SDKs. Any data that you enter into tags or free-form fields used for names can be used for billing or diagnostic logs. If you provide a URL to an external server, we strongly recommend that you do not include credentials information in the URL to validate your request to that server.

Encryption on Amazon MWAA

The following topics describe how Amazon MWAA protects your data at rest, and in transit. Use this information to learn how Amazon MWAA integrates with AWS KMS to encrypt data at rest, and how data is encrypted using Transport Layer Security (TLS) protocol in transit.

Topics

- Encryption at rest
- Encryption in transit

Encryption at rest

On Amazon MWAA, data *at rest* is data that the service saves to persistent media.

You can use an AWS-owned key for data at rest encryption, or optionally provide a Customer-managed key for additional encryption when you create an environment. If you choose to use a customer-managed KMS key, it must be in the same account as the other AWS resources and services you are using with your environment.

To use a customer-managed KMS key, you must attach the required policy statement for CloudWatch access to your key policy. When you use a customer-managed KMS key for your environment, Amazon MWAA attaches four grants on your behalf. For more information about the grants Amazon MWAA attaches to a customer-managed KMS key, refer to Customer-managed keys for data encryption.

If you do not specify a customer-managed KMS key, by default, Amazon MWAA uses an AWS owned KMS key for to encrypt and decrypt your data. We recommend using an AWS owned KMS key to manage data encryption on Amazon MWAA.

> **Note**
>
> You pay for the storage and use of AWS owned, or customer-managed KMS keys on Amazon MWAA. For more information, refer to AWS KMS Pricing.

Encryption artifacts

You specify the encryption artifacts used for at rest encryption by specifying an AWS-owned key or Customer-managed key when you create your Amazon MWAA environment. Amazon MWAA adds the grants needed to your specified key.

Amazon S3 – Amazon S3 data is encrypted at the object-level using Server-Side Encryption (SSE). Amazon S3 encryption and decryption takes place on the Amazon S3 bucket where your DAG code and supporting files are stored. Objects are encrypted when they are uploaded to Amazon S3 and decrypted when they are downloaded to your Amazon MWAA environment. By default, if you are using a customer-managed KMS key, Amazon MWAA uses it to read and decrypt the data on your Amazon S3 bucket.

CloudWatch Logs – If you are using an AWS owned KMS key, Apache Airflow logs sent to CloudWatch Logs are encrypted using SSE with CloudWatch Logs's AWS owned KMS key. If you are using a customer-managed KMS key, you must add a key policy to your KMS key to allow CloudWatch Logs to use your key.

Amazon SQS – Amazon MWAA creates one Amazon SQS queue for your environment. Amazon MWAA handles encrypting data passed to and from the queue using SSE with either an AWS owned KMS key, or a customer-managed KMS key that you specify. You must add Amazon SQS permissions to your execution role regardless of whether you are using an AWS owned or customer-managed KMS key.

Aurora PostgreSQL – Amazon MWAA creates one PostgreSQL cluster for your environment. Aurora PostgreSQL encrypts the content with either an AWS owned or customer-managed KMS key using SSE. If you are using a customer-managed KMS key, Amazon RDS adds at least two grants to the key: one for the cluster and one for the database instance. Amazon RDS can create additional grants if you choose to use your customer-managed KMS key on multiple environments. For more information, refer to Data protection in Amazon RDS.

Encryption in transit

Data in transit is referred to as data that can be intercepted as it travels the network.

Transport Layer Security (TLS) encrypts the Amazon MWAA objects in transit between your environment's Apache Airflow components and other AWS services that integrate with Amazon MWAA, including traffic between Apache Airflow components. The following paths are encrypted:

Metadata database

Schedulers, workers, and the DAG processor connect to the Aurora PostgreSQL database through with `sslmode=require`.

Celery broker

Schedulers and workers communicate with the Amazon SQS Celery broker queue over HTTPS.

AWS services

All components connect to Amazon S3, CloudWatch Logs, AWS KMS, Secrets Manager, and AWS STS using HTTPS endpoints through the AWS SDK for Python (Boto3).

Web server

Client connections terminate TLS at the managed load balancer. The load balancer accepts only HTTPS connections.

For more information about Amazon S3 encryption, see Protecting data using encryption.

Using customer-managed keys for encryption

You can optionally provide a Customer-managed key for data encryption on your environment. You must create the customer-managed KMS key in the same Region as your Amazon MWAA environment instance and your Amazon S3 bucket where you store resources for your workflows. If the customer-managed KMS key that you specify is in a different account from the one you use to configure an environment, you must specify the key using its ARN for cross-account access. For more information about creating keys, refer to Creating Keys in the *AWS Key Management Service Developer Guide*.

What's supported

AWS KMS feature	Supported	
An AWS KMS key ID or ARN.	Yes	

AWS KMS feature	Supported
An AWS KMS key alias.	No
An AWS KMS multi-region key.	No

Using Grants for Encryption

This topic describes the grants Amazon MWAA attaches to a customer-managed KMS key on your behalf to encrypt and decrypt your data.

How it works

There are two resource-based access control mechanisms supported by AWS KMS for customer-managed KMS key: a key policy and grant.

A key policy is used when the permission is mostly static and used in synchronous service mode. A grant is used when more dynamic and granular permissions are required, such as when a service needs to define different access permissions for itself or other accounts.

Amazon MWAA uses and attaches four grant policies to your customer-managed KMS key. This is due to the granular permissions required for an environment to encrypt data at rest from CloudWatch Logs, Amazon SQS queue, Aurora PostgreSQL database database, Secrets Manager secrets, Amazon S3 bucket and DynamoDB tables.

When you create an Amazon MWAA environment and specify a customer-managed KMS key, Amazon MWAA attaches the grant policies to your customer-managed KMS key. These policies allow Amazon MWAA in `airflow.us-east-1.amazonaws.com` to use your customer-managed KMS key to encrypt resources on your behalf that are owned by Amazon MWAA.

Amazon MWAA creates, and attaches, additional grants to a specified KMS key on your behalf. This includes policies to retire a grant if you delete your environment, to use your customer-managed KMS key for Client-Side Encryption (CSE), and for the AWS Fargate execution role that needs to access secrets protected by your customer-managed key in Secrets Manager.

Grant policies

Amazon MWAA adds the following resource based policy grants on your behalf to a customer-managed KMS key. These policies allow the grantee and the principal (Amazon MWAA) to perform actions defined in the policy.

Grant 1: used to create data plane resources

```
{
  "Name": "mwaa-grant-for-env-mgmt-role-environment name",
  "GranteePrincipal": "airflow.us-east-1.amazonaws.com",
  "RetiringPrincipal": "airflow.us-east-1.amazonaws.com",
  "Operations": [
    "kms:Encrypt",
    "kms:Decrypt",
    "kms:ReEncrypt*",
    "kms:GenerateDataKey*",
    "kms:CreateGrant",
    "kms:DescribeKey",
    "kms:RetireGrant"
  ]
}
```

Grant 2: used for `ControllerLambdaExecutionRole` access

```
{
  "Name": "mwaa-grant-for-lambda-exec-environment name",
  "GranteePrincipal": "airflow.us-east-1.amazonaws.com",
  "RetiringPrincipal": "airflow.us-east-1.amazonaws.com",
  "Operations": [
    "kms:Encrypt",
    "kms:Decrypt",
    "kms:ReEncrypt*",
    "kms:GenerateDataKey*",
    "kms:DescribeKey",
    "kms:RetireGrant"
  ]
}
```

Grant 3: used for `CfnManagementLambdaExecutionRole` access

```
{
```

```
  "Name": " mwaa-grant-for-cfn-mgmt-environment name",
  "GranteePrincipal": "airflow.us-east-1.amazonaws.com",
  "RetiringPrincipal": "airflow.us-east-1.amazonaws.com",
  "Operations": [
    "kms:Encrypt",
    "kms:Decrypt",
    "kms:ReEncrypt*",
    "kms:GenerateDataKey*",
    "kms:DescribeKey"
  ]
}
```

Grant 4: used for Fargate execution role to access backend secrets

```
{
  "Name": "mwaa-fargate-access-for-environment name",
  "GranteePrincipal": "airflow.us-east-1.amazonaws.com",
  "RetiringPrincipal": "airflow.us-east-1.amazonaws.com",
  "Operations": [
    "kms:Encrypt",
    "kms:Decrypt",
    "kms:ReEncrypt*",
    "kms:GenerateDataKey*",
    "kms:DescribeKey",
    "kms:RetireGrant"
  ]
}
```

Attaching key policies to a customer-managed key

If you choose to use your own customer-managed KMS key with Amazon MWAA, you must attach the following policy to the key to allow Amazon MWAA to use it to encrypt your data.

If the customer-managed KMS key you used for your Amazon MWAA environment is not already configured to work with CloudWatch, you must update the key policy to allow for encrypted CloudWatch Logs. For more information, refer to the Encrypt log data in CloudWatch using AWS Key Management Service service.

The following example represents a key policy for CloudWatch Logs. Substitute the sample values provided for the region.

```
{
```

```
  "Effect": "Allow",
  "Principal": {
    "Service": "logs.us-east-1.amazonaws.com"
  },
  "Action": [
    "kms:Encrypt*",
    "kms:Decrypt*",
    "kms:ReEncrypt*",
    "kms:GenerateDataKey*",
    "kms:Describe*"
  ],
  "Resource": "*",
  "Condition": {
    "ArnLike": {
      "kms:EncryptionContext:aws:logs:arn": "arn:aws:logs:us-east-1:*:*"
    }
  }
}
```

AWS Identity and Access Management

AWS Identity and Access Management (IAM) is an AWS service that helps an administrator securely control access to AWS resources. IAM administrators control who can be authenticated (signed in) and authorized (have permissions) to use Amazon Managed Workflows for Apache Airflow resources. IAM is an AWS service you can use with no additional charge.

This topic provides a basic overview of how Amazon MWAA uses AWS Identity and Access Management (IAM). To learn about managing access to Amazon MWAA, refer to Managing access to an Amazon MWAA environment.

Contents

- Audience
- Authenticating With Identities
- Managing Access Using Policies
- Allowing users to access their own permissions
- Troubleshooting Amazon Managed Workflows for Apache Airflow identity and access
- How Amazon MWAA works with IAM
- IAM Identity Center integration

Audience

How you use AWS Identity and Access Management (IAM) differs based on your role:

- **Service user** - request permissions from your administrator if you cannot access features (see Troubleshooting Amazon Managed Workflows for Apache Airflow identity and access)
- **Service administrator** - determine user access and submit permission requests (see How Amazon MWAA works with IAM)
- **IAM administrator** - write policies to manage access (see Amazon MWAA identity-based policy examples)

Authenticating With Identities

Authentication is how you sign in to AWS using your identity credentials. You must be authenticated as the AWS account root user, an IAM user, or by assuming an IAM role.

You can sign in as a federated identity using credentials from an identity source like AWS IAM Identity Center (IAM Identity Center), single sign-on authentication, or Google/Facebook credentials. For more information about signing in, see How to sign in to your AWS account in the *AWS Sign-In User Guide*.

For programmatic access, AWS provides an SDK and CLI to cryptographically sign requests. For more information, see AWS Signature Version 4 for API requests in the *IAM User Guide*.

AWS account root user

When you create an AWS account, you begin with one sign-in identity called the AWS account *root user* that has complete access to all AWS services and resources. We strongly recommend that you don't use the root user for everyday tasks. For tasks that require root user credentials, see Tasks that require root user credentials in the *IAM User Guide*.

IAM Users and Groups

An *IAM user* is an identity with specific permissions for a single person or application. We recommend using temporary credentials instead of IAM users with long-term credentials. For more information, see Require human users to use federation with an identity provider to access AWS using temporary credentials in the *IAM User Guide*.

An *IAM group* specifies a collection of IAM users and makes permissions easier to manage for large sets of users. For more information, see Use cases for IAM users in the *IAM User Guide*.

IAM Roles

An *IAM role* is an identity with specific permissions that provides temporary credentials. You can assume a role by switching from a user to an IAM role (console) or by calling an AWS CLI or AWS API operation. For more information, see Methods to assume a role in the *IAM User Guide*.

IAM roles are useful for federated user access, temporary IAM user permissions, cross-account access, cross-service access, and applications running on Amazon EC2. For more information, see Cross account resource access in IAM in the *IAM User Guide*.

Managing Access Using Policies

You control access in AWS by creating policies and attaching them to AWS identities or resources. A policy defines permissions when associated with an identity or resource. AWS evaluates these policies when a principal makes a request. Most policies are stored in AWS as JSON documents. For more information about JSON policy documents, see Overview of JSON policies in the *IAM User Guide*.

Using policies, administrators specify who has access to what by defining which **principal** can perform **actions** on what **resources**, and under what **conditions**.

By default, users and roles have no permissions. An IAM administrator creates IAM policies and adds them to roles, which users can then assume. IAM policies define permissions regardless of the method used to perform the operation.

Identity-Based Policies

Identity-based policies are JSON permissions policy documents that you attach to an identity (user, group, or role). These policies control what actions identities can perform, on which resources, and under what conditions. To learn how to create an identity-based policy, see Define custom IAM permissions with customer managed policies in the *IAM User Guide*.

Identity-based policies can be *inline policies* (embedded directly into a single identity) or *managed policies* (standalone policies attached to multiple identities). To learn how to choose between managed and inline policies, see Choose between managed policies and inline policies in the *IAM User Guide*.

Resource-Based Policies

Resource-based policies are JSON policy documents that you attach to a resource. Examples include IAM *role trust policies* and Amazon S3 *bucket policies*. In services that support resource-based policies, service administrators can use them to control access to a specific resource. You must specify a principal in a resource-based policy.

Resource-based policies are inline policies that are located in that service. You can't use AWS managed policies from IAM in a resource-based policy.

Access Control Lists (ACLs)

Access control lists (ACLs) control which principals (account members, users, or roles) have permissions to access a resource. ACLs are similar to resource-based policies, although they do not use the JSON policy document format.

Amazon S3, AWS WAF, and Amazon VPC are examples of services that support ACLs. To learn more about ACLs, see Access control list (ACL) overview in the *Amazon Simple Storage Service Developer Guide*.

Other Policy Types

AWS supports additional policy types that can set the maximum permissions granted by more common policy types:

- **Permissions boundaries** – Set the maximum permissions that an identity-based policy can grant to an IAM entity. For more information, see Permissions boundaries for IAM entities in the *IAM User Guide*.
- **Service control policies (SCPs)** – Specify the maximum permissions for an organization or organizational unit in AWS Organizations. For more information, see Service control policies in the *AWS Organizations User Guide*.
- **Resource control policies (RCPs)** – Set the maximum available permissions for resources in your accounts. For more information, see Resource control policies (RCPs) in the *AWS Organizations User Guide*.
- **Session policies** – Advanced policies passed as a parameter when creating a temporary session for a role or federated user. For more information, see Session policies in the *IAM User Guide*.

Multiple Policy Types

When multiple types of policies apply to a request, the resulting permissions are more complicated to understand. To learn how AWS determines whether to allow a request when multiple policy types are involved, see Policy evaluation logic in the *IAM User Guide*.

Allowing users to access their own permissions

This example shows how you might create a policy that allows IAM users to view the inline and managed policies that are attached to their user identity. This policy includes permissions to complete this action on the console or programmatically using the AWS CLI or AWS API.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "ViewOwnUserInfo",
            "Effect": "Allow",
            "Action": [
                "iam:GetUserPolicy",
                "iam:ListGroupsForUser",
                "iam:ListAttachedUserPolicies",
                "iam:ListUserPolicies",
                "iam:GetUser"
            ],
            "Resource": ["arn:aws:iam::*:user/${aws:username}"]
        },
        {
            "Sid": "NavigateInConsole",
            "Effect": "Allow",
            "Action": [
                "iam:GetGroupPolicy",
                "iam:GetPolicyVersion",
                "iam:GetPolicy",
                "iam:ListAttachedGroupPolicies",
                "iam:ListGroupPolicies",
                "iam:ListPolicyVersions",
                "iam:ListPolicies",
                "iam:ListUsers"
            ],
            "Resource": "*"
        }
```

```
    ]
}
```

Troubleshooting Amazon Managed Workflows for Apache Airflow identity and access

Use the following information to help you diagnose and fix common issues that you can encounter when working with Amazon MWAA and IAM.

I am not authorized to perform an action in Amazon MWAA

If the AWS Management Console tells you that you're not authorized to perform an action, then you must contact your administrator for assistance. Your administrator is the person that provided you with your user name and password.

I am not authorized to perform iam:PassRole

If you receive an error that you're not authorized to perform the `iam:PassRole` action, your policies must be updated to allow you to pass a role to Amazon MWAA.

Some AWS services allow you to pass an existing role to that service instead of creating a new service role or service-linked role. To do this, you must have permissions to pass the role to the service.

The following example error occurs when an IAM user named `marymajor` tries to use the console to perform an action in Amazon MWAA. However, the action requires the service to have permissions that are granted by a service role. Mary does not have permissions to pass the role to the service.

```
User: arn:aws:iam::123456789012:user/marymajor is not authorized to perform:
 iam:PassRole
```

In this case, Mary's policies must be updated to allow her to perform the `iam:PassRole` action.

If you need help, contact your AWS administrator. Your administrator is the person who provided you with your sign-in credentials.

I want to allow people outside of my AWS account to access my Amazon MWAA resources

You can create a role that users in other accounts or people outside of your organization can use to access your resources. You can specify who is trusted to assume the role. For services that support resource-based policies or access control lists (ACLs), you can use those policies to grant people access to your resources.

To learn more, consult the following:

- To learn whether Amazon MWAA supports these features, see How Amazon MWAA works with IAM.
- To learn how to provide access to your resources across AWS accounts that you own, see Providing access to an IAM user in another AWS account that you own in the *IAM User Guide*.
- To learn how to provide access to your resources to third-party AWS accounts, see Providing access to AWS accounts owned by third parties in the *IAM User Guide*.
- To learn how to provide access through identity federation, see Providing access to externally authenticated users (identity federation) in the *IAM User Guide*.
- To learn the difference between using roles and resource-based policies for cross-account access, see Cross account resource access in IAM in the *IAM User Guide*.

How Amazon MWAA works with IAM

Amazon MWAA uses IAM identity-based policies to grant permissions to Amazon MWAA actions and resources. For recommended examples of custom IAM policies you can use to control access to your Amazon MWAA resources, refer to the section called “Accessing an Amazon MWAA environment”.

To get a high-level access of how Amazon MWAA and other AWS services work with IAM, refer to AWS Services That Work with IAM in the *IAM User Guide*.

Amazon MWAA identity-based policies

With IAM identity-based policies, you can specify allowed or denied actions and resources, as well as the conditions under which actions are allowed or denied. Amazon MWAA supports specific actions, resources, and condition keys.

The following steps present how you can create a new JSON policy using the IAM console. This policy provides read-only access to your Amazon MWAA resources.

To use the JSON policy editor to create a policy

1. Sign in to the AWS Management Console and open the IAM console at https://console.aws.amazon.com/iam/.
2. In the navigation pane on the left, choose **Policies**.

   If this is your first time choosing **Policies**, the **Welcome to Managed Policies** page appears. Choose **Get Started**.
3. At the top of the page, choose **Create policy**.
4. In the **Policy editor** section, choose the **JSON** option.
5. Enter the following JSON policy document:

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Effect": "Allow",
      "Action": [
        "airflow:ListEnvironments",
        "airflow:GetEnvironment",
        "airflow:ListTagsForResource"
      ],
      "Resource": "*"
    }
  ]
}
```

6. Choose **Next**.

> **Note**
>
> You can switch between the **Visual** and **JSON** editor options anytime. However, if you make changes or choose **Next** in the **Visual** editor, IAM might restructure your policy to optimize it for the visual editor. For more information, see Policy restructuring in the *IAM User Guide*.

7. On the **Review and create** page, enter a **Policy name** and a **Description** (optional) for the policy that you are creating. Review **Permissions defined in this policy** to see the permissions that are granted by your policy.
8. Choose **Create policy** to save your new policy.

To learn about all of the elements that you use in a JSON policy, refer to IAM JSON Policy Elements Reference in the *IAM User Guide*.

Actions

Administrators can use AWS JSON policies to specify who has access to what. That is, which **principal** can perform **actions** on what **resources**, and under what **conditions**.

The `Action` element of a JSON policy describes the actions that you can use to allow or deny access in a policy. Include actions in a policy to grant permissions to perform the associated operation.

Policy statements must include either an `Action` element or a `NotAction` element. The `Action` element lists the actions allowed by the policy. The `NotAction` element lists the actions that are not allowed.

The actions defined for Amazon MWAA reflect tasks that you can perform using Amazon MWAA. Policy actions in Detective have the following prefix: `airflow:`.

You can also use wildcards (*) to specify multiple actions. Instead of listing these actions separately, you can grant access to all actions that end with the word, for example, `environment`.

To get a list of Amazon MWAA actions, refer to Actions Defined by Amazon Managed Workflows for Apache Airflow in the *IAM User Guide*.

Amazon MWAA identity-based policy examples

To access the Amazon MWAA policies, refer to Managing access to an Amazon MWAA environment.

By default, IAM users and roles don't have permission to create or modify Amazon MWAA resources. They also can't perform tasks using the AWS Management Console, AWS CLI, or AWS API.

An IAM administrator must create IAM policies that grant users and roles permission to perform specific API operations on the specified resources they need. The administrator then attaches those policies to the IAM users or groups that require those permissions.

> **Important**
>
> We recommend using IAM roles and temporary credentials to provide access to your Amazon MWAA resources. Avoiding attaching permission poicies directly to your IAM users.

To learn how to create an IAM identity-based policy using these example JSON policy documents, refer to Creating Policies on the JSON Tab in the *IAM User Guide*.

Topics

- Policy best practices
- Using the Amazon MWAA console
- Allowing users to access their own permissions

Policy best practices

Identity-based policies determine whether someone can create, access, or delete Amazon MWAA resources in your account. These actions can incur costs for your AWS account. When you create or edit identity-based policies, follow these guidelines and recommendations:

- **Get started with AWS managed policies and move toward least-privilege permissions** – To get started granting permissions to your users and workloads, use the *AWS managed policies* that grant permissions for many common use cases. They are available in your AWS account. We recommend that you reduce permissions further by defining AWS customer managed policies that are specific to your use cases. For more information, see AWS managed policies or AWS managed policies for job functions in the *IAM User Guide*.
- **Apply least-privilege permissions** – When you set permissions with IAM policies, grant only the permissions required to perform a task. You do this by defining the actions that can be taken on specific resources under specific conditions, also known as *least-privilege permissions*. For more information about using IAM to apply permissions, see Policies and permissions in IAM in the *IAM User Guide*.
- **Use conditions in IAM policies to further restrict access** – You can add a condition to your policies to limit access to actions and resources. For example, you can write a policy condition to specify that all requests must be sent using SSL. You can also use conditions to grant access to service actions if they are used through a specific AWS service, such as CloudFormation. For more information, see IAM JSON policy elements: Condition in the *IAM User Guide*.

- **Use IAM Access Analyzer to validate your IAM policies to ensure secure and functional permissions** – IAM Access Analyzer validates new and existing policies so that the policies adhere to the IAM policy language (JSON) and IAM best practices. IAM Access Analyzer provides more than 100 policy checks and actionable recommendations to help you author secure and functional policies. For more information, see Validate policies with IAM Access Analyzer in the *IAM User Guide*.
- **Require multi-factor authentication (MFA)** – If you have a scenario that requires IAM users or a root user in your AWS account, turn on MFA for additional security. To require MFA when API operations are called, add MFA conditions to your policies. For more information, see Secure API access with MFA in the *IAM User Guide*.

For more information about best practices in IAM, see Security best practices in IAM in the *IAM User Guide*.

Using the Amazon MWAA console

To use the Amazon MWAA console, the user or role must have access to the relevant actions, which match corresponding actions in the API.

To access the Amazon MWAA policies, refer to Managing access to an Amazon MWAA environment.

Allowing users to access their own permissions

This example shows how you might create a policy that allows IAM users to view the inline and managed policies that are attached to their user identity. This policy includes permissions to complete this action on the console or programmatically using the AWS CLI or AWS API.

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "ViewOwnUserInfo",
            "Effect": "Allow",
            "Action": [
                "iam:GetUserPolicy",
                "iam:ListGroupsForUser",
                "iam:ListAttachedUserPolicies",
                "iam:ListUserPolicies",
                "iam:GetUser"
            ],
```

```
            "Resource": ["arn:aws:iam::*:user/${aws:username}"]
        },
        {
            "Sid": "NavigateInConsole",
            "Effect": "Allow",
            "Action": [
                "iam:GetGroupPolicy",
                "iam:GetPolicyVersion",
                "iam:GetPolicy",
                "iam:ListAttachedGroupPolicies",
                "iam:ListGroupPolicies",
                "iam:ListPolicyVersions",
                "iam:ListPolicies",
                "iam:ListUsers"
            ],
            "Resource": "*"
        }
    ]
}
```

IAM Identity Center integration

As an alternative to the IAM-based access model described in this section, you can use IAM Identity Center to authenticate users through your corporate identity provider and manage authorization centrally through Permission Sets. For more information, see Using IAM Identity Center with Amazon Managed Workflows for Apache Airflow.

Using IAM Identity Center with Amazon Managed Workflows for Apache Airflow

When you enable IAM Identity Center integration, Amazon Managed Workflows for Apache Airflow uses IAM Identity Center to authenticate and authorize you. Sign in to the Apache Airflow UI with your corporate identity provider through the IAM Identity Center device code flow. Amazon MWAA evaluates authorization per request using IAM policies attached to IAM Identity Center Permission Sets.

Version requirement

IAM Identity Center integration requires Apache Airflow version 3.3.1 or later. The Amazon MWAA console hides the IAM Identity Center configuration section for environments running earlier versions.

In this section:

- Prerequisites
- How it works
- Enabling IAM Identity Center integration
- Configuring access with Permission Sets
- Teams and the airflow:TeamNames condition key
- Logging in to Apache Airflow with IAM Identity Center
- Known limitations
- Troubleshooting

Prerequisites

- Apache Airflow version 3.3.1 or later.
- An *organization-level* IAM Identity Center instance. Account-level IAM Identity Center does not support Permission Sets and is not compatible with this feature.
- Users and groups synced into IAM Identity Center from your corporate identity provider (for example, Okta, Azure AD, or Ping) through SAML or federation, or managed directly in the IAM Identity Center directory.
- The Amazon MWAA environment's AWS account must be accessible in your IAM Identity Center organization.

How it works

When you open the Apache Airflow UI, Amazon MWAA redirects you to an IAM Identity Center device authorization page. You complete authentication in a separate browser tab using your corporate credentials. After successful authentication, Amazon MWAA obtains temporary IAM

Identity Center credentials scoped to your assigned Permission Set and grants you access to the Apache Airflow UI.

Amazon MWAA evaluates authorization on each Apache Airflow UI action using the `airflow:AccessAirflowResource` IAM action and the condition keys described in Configuring access with Permission Sets. You attach the policies to IAM Identity Center Permission Sets, not to individual IAM users or roles, which centralizes access governance in IAM Identity Center.

Apache Airflow Security menu

When IAM Identity Center integration is enabled, the Apache Airflow Security menu, including the List Roles, Users, and Permissions pages, is not available. Access control is managed entirely through IAM Identity Center Permission Sets.

Enabling IAM Identity Center integration

Recommended: create a new environment

We recommend creating a new environment with IAM Identity Center integration enabled from the start and migrating users to it. This approach lets existing users continue working on the original environment while you validate the IAM Identity Center setup, without disrupting access to your production environment.

To migrate to a new environment with IAM Identity Center enabled

1. Create a new environment with IAM Identity Center enabled, using the same Amazon S3 bucket, DAGs, and Amazon VPC as the existing environment. For steps to configure IAM Identity Center during environment creation, refer to Step two: Configure advanced settings.
2. Configure Permission Sets and validate access for a subset of users. For more information, refer to Configuring access with Permission Sets.
3. Cut traffic over by updating bookmarks or the Amazon MWAA console URL to point to the new environment.
4. Decommission the original environment once all users have migrated.

Alternative: enable IAM Identity Center on an existing environment

> **Irreversible action**
>
> Once IAM Identity Center integration is enabled on an environment, it cannot be disabled. If you need a rollback path, use the recommended parallel-environment approach described in Recommended: create a new environment.

To enable IAM Identity Center on an existing environment using the Amazon MWAA console, open the **Update environment** wizard. Expand the **Identity Center configuration** section (visible only for environments on Apache Airflow 3.3.1 or later) and provide the access portal URL and Identity Store ID fields. For field descriptions, refer to Step two: Configure advanced settings.

To enable IAM Identity Center using the AWS CLI:

```
aws mwaa update-environment \
  --name my-env \
  --identity-center-configuration '{
    "startUrl": "https://d-xxxxxxxxxx.awsapps.com/start",
    "identityStoreId": "d-xxxxxxxxxx"
  }'
```

The update takes 20 to 30 minutes for a forced update, which temporarily takes the environment offline. During a graceful update, your DAGs and tasks continue to run.

`startUrl`

The IAM Identity Center access portal URL. Found in the IAM Identity Center console under **Settings**. Example: `https://d-xxxxxxxxxx.awsapps.com/start`.

`identityStoreId`

The IAM Identity Center Identity Store ID. Found in the IAM Identity Center console under **Settings**. Example: `d-xxxxxxxxxx`.

Configuring access with Permission Sets

Authorization in IAM Identity Center-enabled Amazon MWAA environments is controlled by IAM policies attached to IAM Identity Center Permission Sets. You use the

`airflow:AccessAirflowResource` action together with the condition keys described in the following table.

Required identitystore permissions

Every Permission Set policy must include the following `identitystore` statement. Amazon MWAA needs it to look up your display name and group memberships; without it, sign-in fails.

```
{
  "Sid": "AllowIdentityStoreLookup",
  "Effect": "Allow",
  "Action": [
    "identitystore:GetUserId",
    "identitystore:DescribeUser",
    "identitystore:ListGroupMembershipsForMember",
    "identitystore:DescribeGroup"
  ],
  "Resource": "*"
}
```

Additional requirements when your IAM Identity Center identity store uses a customer managed AWS KMS key

If your IAM Identity Center identity store is encrypted with a customer managed AWS KMS key, complete two additional steps. Without them, the `identitystore` API calls that look up your display name and group memberships fail during sign-in with the error: "Could not look up your user information from IAM Identity Center."

To check whether your identity store uses a customer managed AWS KMS key, open the IAM Identity Center console, choose **Settings**, then **Management**. If the console shows a AWS KMS key ARN on **Key for encrypting IAM Identity Center data at rest**, complete both of the following steps. If the default AWS managed key is shown, you do not need to make any additional changes.

Step 1: Add AWS KMS permissions to every Permission Set inline policy. The Permission Set role makes direct `identitystore` API calls using your federated credentials. The following IAM inline policy statement grants the Permission Set role permission to decrypt

data using your AWS KMS key. Add it to the inline policy, replacing the placeholder values with your AWS KMS key ARN:

```
{
  "Sid": "AllowKMSDecryptForIDC",
  "Effect": "Allow",
  "Action": [
    "kms:Decrypt",
    "kms:DescribeKey"
  ],
  "Resource": "arn:aws:kms:key-region:idc-account-id:key/key-id"
}
```

Where: *key-region* is the AWS Region of the AWS KMS key, *idc-account-id* is the account ID of the IAM Identity Center organization management account, and *key-id* is the ID of your customer managed AWS KMS key. The AWS KMS key ARN is shown in the IAM Identity Center console under **Settings**, then **Management**. The key lives in the IAM Identity Center organization management account. This may be a different account from your Amazon MWAA environment account.

Step 2: Add the Permission Set role to the AWS KMS key policy. Because the key lives in the IAM Identity Center organization management account, add an explicit cross-account entry to the AWS KMS key policy. The following AWS KMS key policy statement grants the Permission Set role in your Amazon MWAA environment account permission to use the key. Replace *account-id* with your Amazon MWAA environment account ID, *sso-region* with the AWS Region of the Permission Set role, *permission-set-name-suffix* with the suffix of the Permission Set role name, *key-region* with the AWS Region of the AWS KMS key, *idc-account-id* with the IAM Identity Center organization management account ID, and *key-id* with the AWS KMS key ID. To find the role name suffix, open the IAM console in your Amazon MWAA environment account after the Permission Set is provisioned and search for `AWSReservedSSO_`*PermissionSetName*:

```
{
  "Sid": "AllowMWAAPermissionSetRole",
  "Effect": "Allow",
  "Principal": {
    "AWS": "arn:aws:iam::account-id:role/aws-reserved/sso.amazonaws.com/sso-
region/AWSReservedSSO_permission-set-name-suffix"
  },
  "Action": [
```

```
    "kms:Decrypt",
    "kms:DescribeKey"
  ],
  "Resource": "arn:aws:kms:key-region:idc-account-id:key/key-id"
}
```

Add one entry for each Permission Set for each Amazon MWAA environment account. After updating the key policy, re-provision the Permission Set to the environment account for the changes to take effect.

Supported condition keys for `airflow:AccessAirflowResource`

Condition key	Type	Description
`airflow:ResourceType`	String	The type of Apache Airflow resource. Supported values: `DAG`, `Connection`, `Variable`, `Pool`, `Asset`, `AssetAlias`, `Configuration`, `View`, `CustomView`.
`airflow:ResourceAction`	String	The action being performed. Supported values: `GET`, `POST`, `PUT`, `DELETE`, `LIST`.
`airflow:ResourceId`	String	The ID of a specific resource. For example, a DAG ID.
`airflow:DagAccessEntity`	String	The sub-resource within a DAG being accessed. Only applies when `airflow:ResourceType` is `DAG`. Supported values for Apache Airflow 3.3.1: `AUDIT_LOG`, `CODE`, `DEPENDENCIES`, `HITL_DETAIL`, `RUN`, `TASK`, `TASK_INSTANCE`, `TASK_LOGS`, `VERSION`, `WARNING`, `XCOM`. These values are defined by Apache Airflow core and may change in future Apache Airflow versions.

Condition key	Type	Description
`airflow:TeamNames`	ArrayOfString	The list of IAM Identity Center groups the user belongs to, populated automatically from IAM Identity Center when the user assumes a Permission Set role. Use `ForAnyValue:StringEquals` when writing conditions on this key.

Required ResourceActions for DAG access

The Apache Airflow UI makes two distinct types of DAG authorization calls:

- `ResourceAction=LIST` (no `ResourceId`): Used by the DAG list page. A policy that grants only `GET` shows zero DAGs on the list page.
- `ResourceAction=GET` (no `ResourceId`): Used by home page dashboard widgets (recent run counts, metrics, and deadline data) that query across all DAGs without specifying a DAG ID. A policy that grants only `LIST` causes the home page to return an access denied error.

Always include both `GET` and `LIST` in any Permission Set that needs DAG access.

Policy changes take effect within a few minutes without any environment update.

The following examples show common Permission Set policy patterns. For additional examples, refer to IAM Identity Center Permission Set policy examples.

Example Grant full Apache Airflow access

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "AllowAllAirflowResources",
      "Effect": "Allow",
      "Action": "airflow:AccessAirflowResource",
      "Resource": "arn:aws:airflow:us-east-1:123456789012:environment/my-env"
```

```
    },
    {
      "Sid": "AllowIdentityStoreLookup",
      "Effect": "Allow",
      "Action": [
        "identitystore:GetUserId",
        "identitystore:DescribeUser",
        "identitystore:ListGroupMembershipsForMember",
        "identitystore:DescribeGroup"
      ],
      "Resource": "*"
    }
  ]
}
```

Example Restrict to DAG read access (list and view)

Always include `LIST` alongside `GET`. The DAG list page uses `ResourceAction=LIST`; a policy that grants only `GET` shows zero DAGs.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "AllowDagRead",
      "Effect": "Allow",
      "Action": "airflow:AccessAirflowResource",
      "Resource": "arn:aws:airflow:us-east-1:123456789012:environment/my-env",
      "Condition": {
        "StringEquals": {
          "airflow:ResourceType": "DAG",
          "airflow:ResourceAction": ["GET", "LIST"]
        }
      }
    },
    {
      "Sid": "AllowIdentityStoreLookup",
      "Effect": "Allow",
      "Action": [
        "identitystore:GetUserId",
        "identitystore:DescribeUser",
        "identitystore:ListGroupMembershipsForMember",
        "identitystore:DescribeGroup"
```

```
      ],
      "Resource": "*"
    }
  ]
}
```

Example Restrict to a specific DAG

Use two separate statements. The first allows `GET` and `LIST` without a `ResourceId`. This is required for the DAG list page (`LIST`) and for home page dashboard widgets that query run counts and metrics across all DAGs (`GET` with no DAG ID). The second allows all actions on the specific DAG identified by `ResourceId`. Do not combine `ResourceId` with `LIST` in the same `StringEquals` block, because `LIST` calls do not supply a `ResourceId` and the condition never matches.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "AllowDagGetAndList",
      "Effect": "Allow",
      "Action": "airflow:AccessAirflowResource",
      "Resource": "arn:aws:airflow:us-east-1:123456789012:environment/my-env",
      "Condition": {
        "StringEquals": {
          "airflow:ResourceType": "DAG",
          "airflow:ResourceAction": ["GET", "LIST"]
        }
      }
    },
    {
      "Sid": "AllowSpecificDagAllActions",
      "Effect": "Allow",
      "Action": "airflow:AccessAirflowResource",
      "Resource": "arn:aws:airflow:us-east-1:123456789012:environment/my-env",
      "Condition": {
        "StringEquals": {
          "airflow:ResourceType": "DAG",
          "airflow:ResourceId": "my_etl_dag"
        }
      }
    },
    {
      "Sid": "AllowIdentityStoreLookup",
```

```
      "Effect": "Allow",
      "Action": [
        "identitystore:GetUserId",
        "identitystore:DescribeUser",
        "identitystore:ListGroupMembershipsForMember",
        "identitystore:DescribeGroup"
      ],
      "Resource": "*"
    }
  ]
}
```

Example Restrict by team using `airflow:TeamNames`

Use `ForAnyValue:StringEquals` (the multi-value operator) for `airflow:TeamNames`. This condition matches if the user belongs to *any* of the listed groups (OR logic). Using the single-value `StringEquals` operator causes the condition to never match when a user belongs to more than one team.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "AllowTeamAccess",
      "Effect": "Allow",
      "Action": "airflow:AccessAirflowResource",
      "Resource": "arn:aws:airflow:us-east-1:123456789012:environment/my-env",
      "Condition": {
        "ForAnyValue:StringEquals": {
          "airflow:TeamNames": ["platform-team", "data-eng"]
        }
      }
    },
    {
      "Sid": "AllowIdentityStoreLookup",
      "Effect": "Allow",
      "Action": [
        "identitystore:GetUserId",
        "identitystore:DescribeUser",
        "identitystore:ListGroupMembershipsForMember",
        "identitystore:DescribeGroup"
      ],
      "Resource": "*"
```

```
    }
  ]
}
```

Teams and the `airflow:TeamNames` condition key

The `airflow:TeamNames` condition key contains the display names of the IAM Identity Center groups that the user belongs to. It's available whether or not the environment has teams, but it works differently in each case:

- **Environments with teams** – Apache Airflow isolates DAGs, connections, and variables by team automatically. Each team maps to the IAM Identity Center group whose display name matches the team name. You don't need `airflow:TeamNames` conditions for this isolation. A policy can narrow access further, but it can't give a user access to a team they don't belong to.
- **Environments without teams** – Use `airflow:TeamNames` to restrict access by IAM Identity Center group membership, as shown in the *Restrict by team* example in Configuring access with Permission Sets.

For more information, see Multi-team environments.

Logging in to Apache Airflow with IAM Identity Center

Accessing the Apache Airflow UI

There are two ways to access the Apache Airflow UI when IAM Identity Center integration is enabled:

Amazon MWAA console link

Choose **Open Airflow UI** on the environment page in the Amazon MWAA console. This path requires your IAM Identity Center Permission Set to include the following IAM statement, which grants access to the Amazon MWAA console:

```
{
  "Sid": "AllowMwaaConsoleRead",
  "Effect": "Allow",
  "Action": [
    "airflow:GetEnvironment",
    "airflow:ListEnvironments"
```

```
  ],
  "Resource": "*"
}
```

> **Console visibility**
>
> The Amazon MWAA console lists all environments in the account, regardless of whether you have Apache Airflow UI access to each one. Authorization for Apache Airflow resources is enforced only after you open the environment and log in.

Direct Apache Airflow URL

Administrators can share the direct Apache Airflow URL with users. Users navigate to the URL and the Apache Airflow webserver automatically initiates the IAM Identity Center device code login flow. No Amazon MWAA console access is required. The environment URL has the format: `https://<env-id>.<region>.airflow.amazonaws.com/`

To log in to Apache Airflow using IAM Identity Center

1. Open the Environments page on the Amazon MWAA console and choose your environment.
2. Choose **Open Airflow UI**.
3. On the device authorization page, note the verification URL and the one-time code (for example, `SMCL-BTVJ`).
4. Open the verification URL in a new browser tab. If you already have an active SSO session, the identity provider authentication step is skipped and you land directly on the code confirmation screen. Otherwise, authenticate with your corporate identity provider.
5. Enter the one-time code and confirm.
6. If you have multiple Permission Sets assigned for this AWS account, select the role for the Permission Set you want to use during this session. If you have only one Permission Set, it is selected automatically and this step is not shown.
7. You are now logged in to the Apache Airflow UI as yourself.

You can also navigate to the Apache Airflow URL directly. Amazon MWAA redirects you to the AWS Console for authentication and then continues with the steps above.

Session duration

Session duration is controlled by the Apache Airflow configuration option `aws_mwaa.session_duration_minutes`, the same as non-IAM Identity Center environments. IAM Identity Center credentials are refreshed in the background, so you are not interrupted during an active session.

Logout

Choosing **Sign out** in Apache Airflow ends your Apache Airflow session only. Your IAM Identity Center and corporate identity provider sessions remain active. Re-logging in immediately might complete without re-entering credentials, depending on the identity provider session duration.

To fully sign out of IAM Identity Center, log out from the IAM Identity Center access portal (`https://<alias>.awsapps.com/start`) or your corporate identity provider portal. Signing out of IAM Identity Center also invalidates your Apache Airflow session.

Known limitations

- **One-way enablement.** Once IAM Identity Center integration is enabled on an environment, it cannot be disabled. Use the parallel-environment approach described in Recommended: create a new environment for a safe migration path.
- **Apache Airflow 3.3.1 or later required.** IAM Identity Center integration is not available on earlier Apache Airflow versions. The console version gate enforces this.
- **Role selection for multi-Permission-Set users.** Users with more than one Permission Set assigned in IAM Identity Center must select a role during login. To eliminate this step, assign a single dedicated Permission Set per user for Amazon MWAA access.
- **`InvokeRestApi` is not supported.** On environments with IAM Identity Center integration enabled, calls to the Amazon MWAA `InvokeRestApi` operation return an error. This affects programmatic access to the Apache Airflow REST API through the AWS CLI or AWS SDKs, such as scripts and CI/CD pipelines that start DAG runs, check DAG run status, or retrieve task logs.

  You can still perform these actions through the Apache Airflow UI. If your workflows depend on programmatic access to the Apache Airflow REST API, keep those workflows on an environment without IAM Identity Center integration enabled.

  IAM Identity Center integration cannot be disabled after you enable it. Evaluate this limitation before you enable IAM Identity Center on an existing environment.

Troubleshooting

Apache Airflow shows "Unable to start sign-in"

Root cause: The IAM Identity Center configuration on the environment is incorrect. The access portal URL or Identity Store ID is invalid, the IAM Identity Center instance is not reachable, or the OIDC client registration has not completed.

Resolution:

1. In the Amazon MWAA console, open the environment and verify the **AWS access portal URL** and **Identity Store ID** under **Identity Center configuration** match the values shown in the IAM Identity Center console under **Settings**.
2. If either value is incorrect, use `UpdateEnvironment` to provide the correct values. The update takes 20 to 30 minutes for a forced update.
3. If the values are correct, verify that the IAM Identity Center instance is organization-level and that the Amazon MWAA environment's AWS account is accessible in the IAM Identity Center organization.

User has no roles after IAM Identity Center authentication

Root cause: No Permission Set is assigned to the user for this AWS account in IAM Identity Center.

Resolution: In the IAM Identity Center console, assign the appropriate Permission Set to the user or their group. The change takes effect on the user's next Apache Airflow login, after IAM Identity Center provisions the role.

All Apache Airflow actions denied ("Access Denied")

Root cause: The IAM policies of the selected Permission Set do not include `airflow:AccessAirflowResource`, or condition key values do not match.

Resolution: In the IAM console, find the rendered role of the Permission Set (`AWSReservedSSO_<name>_<id>`) and use IAM Policy Simulator to test `airflow:AccessAirflowResource` on the environment ARN. Common mistakes:

- Condition value case mismatch (for example, `dag` instead of `DAG`).
- `StringEquals` used on `airflow:TeamNames` instead of `ForAnyValue:StringEquals`.
- Policy applied to a different Permission Set than the one the user selected.

- Only `LIST` granted without `GET` (or vice versa). The home page dashboard widgets use `ResourceAction=GET` with no `ResourceId`; without it, the home page returns 403 even if the DAG list page works. Always grant `"airflow:ResourceAction": ["GET", "LIST"]` together.

DAG list shows zero DAGs, or home page shows "Access Denied"

Root cause: The Permission Set policy is missing `GET`, `LIST`, or both for `airflow:ResourceType=DAG`. The DAG list page uses `ResourceAction=LIST`; home page dashboard widgets use `ResourceAction=GET` with no `ResourceId`. A policy granting only `GET` shows zero DAGs. A policy granting only `LIST` causes a 403 on the home page.

Resolution: Include both `GET` and `LIST` in the `airflow:ResourceAction` condition for any Permission Set that needs DAG access: `"airflow:ResourceAction": ["GET", "LIST"]`.

`airflow:TeamNames` condition not matching

Root cause: Policy uses `StringEquals` (single-value operator) instead of `ForAnyValue:StringEquals` (multi-value operator).

Resolution: Update the policy to use `ForAnyValue:StringEquals` with `airflow:TeamNames` (plural).

Compliance Validation for Amazon Managed Workflows for Apache Airflow

Our new AWS sign-up experience is not designed for regulated workloads. If you're using our new AWS sign-up experience, but you want to use AWS for regulated workloads, you can sign up for AWS (advanced) or activate advanced features for your AWS environment.

To learn whether an AWS service is within the scope of specific compliance programs, see AWS services in Scope by Compliance Program and choose the compliance program that you are interested in. For general information, see AWS Compliance Programs.

You can download third-party audit reports using AWS Artifact. For more information, see Downloading Reports in AWS Artifact.

Your compliance responsibility when using AWS services is determined by the sensitivity of your data, your company's compliance objectives, and applicable laws and regulations. For more

information about your compliance responsibility when using AWS services, see AWS Security Documentation.

Resilience in Amazon Managed Workflows for Apache Airflow

The AWS global infrastructure is built around AWS Regions and Availability Zones. Regions provide multiple physically separated and isolated Availability Zones, which are connected through low-latency, high-throughput, and highly redundant networking. With Availability Zones, you can design and operate applications and databases that automatically fail over between zones without interruption. Availability Zones are more highly available, fault tolerant, and scalable than traditional single or multiple data center infrastructures.

For more information about AWS Regions and Availability Zones, refer to AWS Global Infrastructure.

Infrastructure Security in Amazon MWAA

As a managed service, Amazon Managed Workflows for Apache Airflow is protected by AWS global network security. For information about AWS security services and how AWS protects infrastructure, see AWS Cloud Security. To design your AWS environment using the best practices for infrastructure security, see Infrastructure Protection in *Security Pillar AWS Well-Architected Framework*.

You use AWS published API calls to access Amazon MWAA through the network. Clients must support the following:

- Transport Layer Security (TLS). We require TLS 1.2 and recommend TLS 1.3.
- Cipher suites with perfect forward secrecy (PFS) such as DHE (Ephemeral Diffie-Hellman) or ECDHE (Elliptic Curve Ephemeral Diffie-Hellman). Most modern systems such as Java 7 and later support these modes.

AWS PrivateLink for Amazon MWAA

With AWS PrivateLink, you can access Amazon MWAA from within your VPC without crossing the public internet. Interface VPC endpoints are elastic network interfaces with private IP addresses in your subnets. For more information, see Access an AWS service using an interface VPC endpoint in the *Amazon VPC User Guide*.

Creating a VPC endpoint

Amazon MWAA uses prefixes for service endpoints, where each prefix corresponds to a set of API operations:

Environment management

`api.airflow.region.amazonaws.com` (use `api.airflow.region.api.aws` for IPv6)

- `CreateEnvironment`
- `DeleteEnvironment`
- `GetEnvironment`
- `ListEnvironments`
- `ListTagsForResource`
- `TagResource`
- `UntagResource`
- `UpdateEnvironment`

Airflow environment

`env.airflow.region.amazonaws.com` (use `env.airflow.region.api.aws` for IPv6)

- `CreateCliToken`
- `CreateWebLoginToken`

FIPS endpoints

FIPS-compliant endpoints are available in US and Canada Regions:

- `com.amazonaws.region.airflow-fips.api`
- `com.amazonaws.region.airflow-fips.env`

For more information about supported Regions, see Amazon MWAA endpoints and quotas in the *AWS General Reference*.

Configuration and Vulnerability Analysis in Amazon MWAA

Configuration and IT controls are a shared responsibility between AWS and you, our customer.

Amazon Managed Workflows for Apache Airflow periodically patches and upgrades Apache Airflow on your environments. Ensure that the appropriate access policies are used for your VPCs.

For more details, refer to the following resources:

- AWS shared responsibility model for Amazon MWAA
- Compliance Validation for Amazon Managed Workflows for Apache Airflow
- Shared Responsibility Model
- Amazon Web Services: Overview of Security Processes
- Infrastructure Security in Amazon MWAA
- Security best practices on Amazon MWAA

Security best practices on Amazon MWAA

Amazon MWAA provides a number of security features to consider as you develop and implement your own security policies. The following best practices are general guidelines and don't represent a complete security solution. Because these best practices might not be appropriate or sufficient for your environment, treat them as helpful considerations rather than prescriptions.

- Use least-permissive permission policies. Grant permissions to only the resources or actions that users need to perform tasks.
- Use AWS CloudTrail to monitor user activity in your account.
- Ensure that the Amazon S3 bucket policy and object ACLs grant permissions to the users from the associated Amazon MWAA environment to put objects into the bucket. This ensures that users with permissions to add workflows to the bucket also have permissions to run the workflows in Airflow.
- Use the Amazon S3 buckets associated with Amazon MWAA environments. Your Amazon S3 bucket can be any name. Don't store other objects in the bucket, or use the bucket with another service.

Security best practices in Apache Airflow

Apache Airflow is not multi-tenant. While there are access control measures to limit some features to specific users, which Amazon MWAA implements, DAG creators do have the ability to write DAGs that can change Apache Airflow user privileges and interact with the underlying metadatabase.

We recommend the following steps when working with Apache Airflow on Amazon MWAA to ensure your environment's metadatabase and DAGs are secure.

- Use separate environments for separate teams with DAG writing access, or the ability to add files to your Amazon S3 `/dags` folder, assuming anything accessible by the Amazon MWAA Execution Role or Apache Airflow connections will also be accessible to users who can write to the environment.
- Do not provide direct Amazon S3 DAGs folder access. Instead, use CI/CD tools to write DAGs to Amazon S3, with a validation step ensuring that the DAG code meets your team's security guidelines.
- Prevent user access to your environment's Amazon S3 bucket. Instead, use a DAG factory that generates DAGs based on a YAML, JSON, or other definition file stored in a separate location from your Amazon MWAA Amazon S3 bucket where you store DAGs.
- Store secrets in Secrets Manager. While this will not prevent users who can write DAGs from reading secrets, it will prevent them from modifying the secrets that your environment uses.

Detecting changes to Apache Airflow user privileges

You can use CloudWatch Logs Insights to detect occurences of DAGs changing Apache Airflow user privileges. To do so, you can use an EventBridge scheduled rule, a Lambda function, and CloudWatch Logs Insights to deliver notifications to CloudWatch metrics whenever one of your DAGs changes Apache Airflow user privileges.

Prerequisites

To complete the following steps, you will need the following:

- An Amazon MWAA environment with all Apache Airflow log types enabled at the `INFO` log level. For more information, refer to the section called “Accessing Airflow logs”.

To configure notifications for changes to Apache Airflow user privileges

1. Create a Lambda function that runs the following CloudWatch Logs Insights query string against the five Amazon MWAA environment log groups (`DAGProcessing`, `Scheduler`, `Task`, `WebServer`, and `Worker`).

```
fields @log, @timestamp, @message | filter @message like "add-role" | stats count()
 by @log
```

2. Create an EventBridge rule that runs on a schedule, with the Lambda function you created in the previous step as the rule's target. Configure your schedule using a cron or rate expression to run at regular intervals.

Security best practices for multi-team environments

Follow these practices when you use a multi-team environment:

- **Use separate environments for strict isolation** – Teams in a multi-team environment share compute, plugins, and Python dependencies. If teams need isolation beyond separate DAG folders and access control, use separate environments.
- **Review team membership regularly** – Team membership comes from IAM Identity Center groups. Review group memberships regularly, and remove users who change roles or leave your organization.
- **Use separate Amazon S3 buckets for sensitive data** – If teams work with data of different sensitivity, consider using separate Amazon S3 buckets with their own access controls instead of relying only on team DAG folders.

For more information, see Multi-team environments.

Apache Airflow versions on Amazon Managed Workflows for Apache Airflow

This topic describes the Apache Airflow versions Amazon Managed Workflows for Apache Airflow supports, and best-practices for upgrading to the latest version.

Topics

- About Amazon MWAA versions
- Latest version
- Apache Airflow versions
- Apache Airflow components
- Upgrading the Apache Airflow version
- Downgrading the Apache Airflow version
- Apache Airflow deprecated versions

About Amazon MWAA versions

Amazon MWAA builds container images that bundle Apache Airflow releases with other common binaries and Python libraries. The image uses the Apache Airflow base install for the version you specify. When you create an environment, you specify an image version to use. Once an environment is created, it keeps using the specified image version until you upgrade it to a later version.

Latest version

Amazon MWAA supports more than one Apache Airflow version. If you don't specify an image version when you create an environment, Amazon MWAA creates an environment using the latest supported version of Apache Airflow.

Apache Airflow versions

The following Apache Airflow versions are supported on Amazon Managed Workflows for Apache Airflow.

Note

- Effective December 30, 2025, Amazon MWAA will end support for Apache Airflow versions v2.4.3, v2.5.1, and v2.6.3. For more information, refer to End-of-support versions.
- Beginning with Apache Airflow v2.2.2, Amazon MWAA supports installing Python requirements, provider packages, and custom plugins directly on the Apache Airflow webserver.
- Beginning with Apache Airflow v2.7.2, your requirements file must include a `--constraint` statement. If you don't provide a constraint, Amazon MWAA will specify one for you to ensure the packages listed in your requirements are compatible with the version of Apache Airflow you're using.

  For more information about setting up constraints in your requirements file, refer to Installing Python dependencies.
- Starting with Apache Airflow 3.3.1, Amazon MWAA supports multi-team environments. For more information, see Multi-team environments.

Apache Airflow version	Apache Airflow release date	Amazon MWAA availability date	Apache Airflow constraints	Python version
v3.3.1	August 12, 2026	September 1, 2026	v3.3.1 constraints file	Python 3.12
v3.2.1	April 22, 2026	May 19, 2026	v3.2.1 constraints file	Python 3.12
v2.11.2	March 14, 2026	July 23, 2026	v2.11.2 constraints file	Python 3.12
v2.11.0	May 20, 2025	January 7, 2026	v2.11.0 constraints file	Python 3.12
v3.0.6	August 29, 2025	October 1, 2025	v3.0.6 constraints file	Python 3.12

Apache Airflow version	Apache Airflow release date	Amazon MWAA availability date	Apache Airflow constraints	Python version
v2.10.3	November 4, 2024	December 18, 2024	v2.10.3 constraints file	Python 3.11
v2.10.1	September 5, 2024	September 26, 2024	v2.10.1 constraints file	Python 3.11
v2.9.2	June 10, 2024	July 9, 2024	v2.9.2 constraints file	Python 3.11
v2.8.1	January 19, 2024	February 23, 2024	v2.8.1 constraints file	Python 3.11
v2.7.2	October 12, 2023	November 6, 2023	v2.7.2 constraints file	Python 3.11

For more information about migrating your self-managed Apache Airflow deployments, or migrating an existing Amazon MWAA environment, including instructions for backing up your metadata database, refer to the Amazon MWAA Migration Guide.

Apache Airflow components

This section describes the number of Apache Airflow schedulers and workers available for each Apache Airflow version on Amazon MWAA, and provides a list of key Apache Airflow features, indicating the version that supports each feature.

Schedulers

Schedulers for Apache Airflow v2 and later:

Scheduler (default)	Scheduler (min)	Scheduler (max)
2	2	5

Workers

Workers for Apache Airflow v2 and later:

Workers (default)	Workers (min)	Workers (max)
10	1	25

Upgrading the Apache Airflow version

Amazon MWAA supports minor version upgrades. This means you can upgrade your environment from version x**.1.**z to x**.2.**z, but not to a new major version, for example, from **1.**y**.**z to **2.**y**.**z.

For more information, and detailed instructions on updating your workflow resources, and upgrading the environment to a new version, refer to the section called “Changing the version”.

Downgrading the Apache Airflow version

Amazon MWAA supports minor version downgrades to an earlier version that is still supported at the time of downgrade. This means you can downgrade your environment from version x**.2.**z to x**.1.**z, but not to a previous major version, for example, from **2.**y**.**z to **1.**y**.**z.

For more information, and detailed instructions on updating your workflow resources, and upgrading the environment to a new version, refer to the section called “Changing the version”.

Apache Airflow deprecated versions

The following table lists the deprecated versions of Apache Airflow in Amazon MWAA, along with initial release and end-of-support dates for each version. For more information about migrating to a newer version, refer to the Amazon MWAA Migration Guide.

Apache Airflow version	Apache Airflow release date	Amazon MWAA availability date	Amazon MWAA end-of-support date
v1.10.12	August 25, 2020	November 24, 2020	February 21, 2024

Apache Airflow version	Apache Airflow release date	Amazon MWAA availability date	Amazon MWAA end-of-support date
v2.0.2	April 19, 2021	May 25, 2021	April 29, 2024
v2.2.2	November 15, 2021	January 27,2022	June 27, 2024
v2.4.3	November 14, 2022	January 05, 2023	December 30, 2025
v2.5.1	January 20, 2023	April 11, 2023	December 30, 2025
v2.6.3	July 10, 2023	August 09, 2023	December 30, 2025

End-of-support versions

We follow the Apache Airflow community release process and version policy on the Apache Airflow website. We are committed to supporting at least three minor versions of Apache Airflow at any given time. We will announce the end-of-support date at least 180 days in advance. This applies to each Apache Airflow minor version.

We are committed to supporting an Apache Airflow version for at least 12 months after it first becomes available.

If any Amazon MWAA environments in your account run the version nearing the end of support, we send a notice through the Health Dashboard with the end of support date.

On the end of support date:

- You can no longer use a deprecated version to create new Amazon MWAA environments.
- You can no longer upgrade or downgrade existing environments to deprecated versions.
- You can still use your existing environments on deprecated versions and perform in-place updates.
- You can no longer receive technical support for environments running on deprecated versions.

You can continue to access your existing Amazon MWAA environments that run the associated, deprecated version of Apache Airflow at your own risk. For instructions on upgrading to a newer version of Apache Airflow, see the Amazon MWAA Migration Guide.

Important

You are responsible for keeping your Amazon MWAA versions current. We urge you to upgrade to the latest version for current security, privacy, and availability safeguards. Operating past the deprecation date (on a "legacy version") increases security, privacy, and operational risks, including downtime. By continuing on a legacy version, you acknowledge these risks and agree to upgrade as soon as possible.

Legacy versions are not generally available and we no longer support them. We might restrict access to any legacy version at any time if it poses a security or liability risk. Continuing on a legacy version might result in your content becoming unavailable, corrupted, or unrecoverable. SLA exceptions apply.

Legacy environments and related software might contain bugs, errors, defects, and harmful components. We provide the legacy version as is, notwithstanding any contrary terms in your agreement.

For more information about shared responsibility, see Shared responsibility in the *AWS Well-Architected Framework*.

Amazon Managed Workflows for Apache Airflow service endpoints and quotas

Amazon Managed Workflows for Apache Airflow has the following service quotas and endpoints. Service quotas, also referred to as limits, are the maximum number of service resources or operations for your AWS account.

Contents

- Service endpoints
- Service quotas
- Increasing quotas

Service endpoints

To access a list of endpoints for Amazon MWAA, refer to Amazon Managed Workflows for Apache Airflow endpoints and quotas.

Service quotas

Quota name	Description	Default quota	Adjustable
Environments	The maximum number of Amazon MWAA environments per account per Region.	10	Yes
Workers per environment	The maximum number of workers per Amazon MWAA environment.	25	Yes
Webservers per environment	The maximum number of webserver	5	Yes

Quota name	Description	Default quota	Adjustable
	s per Amazon MWAA environment.		
Teams per environment	The maximum number of teams per Amazon MWAA environment.	20	No

Increasing quotas

You can request an increase to an adjustable quota by submitting a quota increase request.

Amazon MWAA frequently asked questions

This page describes common questions you can encounter when using Amazon Managed Workflows for Apache Airflow.

Contents

- Supported versions
  - What does Amazon MWAA support for Apache Airflow v2?
  - What Python version can I use?
- Use cases
  - Can I use Amazon MWAA with Amazon SageMaker Unified Studio?
  - When can I use AWS Step Functions vs. Amazon MWAA?
- Environment specifications
  - How much task storage is available to each environment?
  - What is the default operating system used for Amazon MWAA environments?
  - Can I use a custom image for my Amazon MWAA environment?
  - Is Amazon MWAA HIPAA compliant?
  - Does Amazon MWAA support Spot Instances?
  - Does Amazon MWAA support a custom domain?
  - Can I SSH into my environment?
  - Why is a self-referencing rule required on the VPC security group?
  - Can I hide environments from different groups in IAM?
  - Can I store temporary data on the Apache Airflow worker?
  - Can I specify more than 25 Apache Airflow workers?
  - Does Amazon MWAA support shared Amazon VPCs or shared subnets?
  - Can I create or integrate custom Amazon SQS queues to manage task execution and workflow orchestration in Apache Airflow?
- Metrics
  - What metrics are used to determine whether to scale workers?
  - Can I create custom metrics in CloudWatch?
- DAGs, Operators, Connections, and other questions

- Can I use the PythonVirtualenvOperator?
- How long does it take Amazon MWAA to recognize a new DAG file?
- Why is my DAG file not picked up by Apache Airflow?
- Can I remove a plugins.zip or requirements.txt from an environment?
- Why don't my plugins appear in the Apache Airflow v2.0.2 Admin Plugins menu?
- Can I use AWS Database Migration Service (DMS) Operators?
- When I access the Airflow REST API using the AWS credentials, can I increase the throttling limit to more than 10 transactions per second (TPS)?
- Where does the Airflow Task Execution API server run in Amazon MWAA?

Supported versions

What does Amazon MWAA support for Apache Airflow v2?

To learn what Amazon MWAA supports, refer to Apache Airflow versions on Amazon Managed Workflows for Apache Airflow.

What Python version can I use?

The following Apache Airflow versions are supported on Amazon Managed Workflows for Apache Airflow.

Note

- Effective December 30, 2025, Amazon MWAA will end support for Apache Airflow versions v2.4.3, v2.5.1, and v2.6.3. For more information, refer to End-of-support versions.
- Beginning with Apache Airflow v2.2.2, Amazon MWAA supports installing Python requirements, provider packages, and custom plugins directly on the Apache Airflow webserver.
- Beginning with Apache Airflow v2.7.2, your requirements file must include a `--constraint` statement. If you don't provide a constraint, Amazon MWAA will specify one for you to ensure the packages listed in your requirements are compatible with the version of Apache Airflow you're using.

For more information about setting up constraints in your requirements file, refer to Installing Python dependencies.

- Starting with Apache Airflow 3.3.1, Amazon MWAA supports multi-team environments. For more information, see Multi-team environments.

Apache Airflow version	Apache Airflow release date	Amazon MWAA availability date	Apache Airflow constraints	Python version
v3.3.1	August 12, 2026	September 1, 2026	v3.3.1 constraints file	Python 3.12
v3.2.1	April 22, 2026	May 19, 2026	v3.2.1 constraints file	Python 3.12
v2.11.2	March 14, 2026	July 23, 2026	v2.11.2 constraints file	Python 3.12
v2.11.0	May 20, 2025	January 7, 2026	v2.11.0 constraints file	Python 3.12
v3.0.6	August 29, 2025	October 1, 2025	v3.0.6 constraints file	Python 3.12
v2.10.3	November 4, 2024	December 18, 2024	v2.10.3 constraints file	Python 3.11
v2.10.1	September 5, 2024	September 26, 2024	v2.10.1 constraints file	Python 3.11
v2.9.2	June 10, 2024	July 9, 2024	v2.9.2 constraints file	Python 3.11
v2.8.1	January 19, 2024	February 23, 2024	v2.8.1 constraints file	Python 3.11
v2.7.2	October 12, 2023	November 6, 2023	v2.7.2 constraints file	Python 3.11

For more information about migrating your self-managed Apache Airflow deployments, or migrating an existing Amazon MWAA environment, including instructions for backing up your metadata database, refer to the Amazon MWAA Migration Guide.

Use cases

Can I use Amazon MWAA with Amazon SageMaker Unified Studio?

Yes. With an Amazon SageMaker Unified Studio workflow, you can set up and run a series of tasks in Amazon SageMaker Unified Studio. Amazon SageMaker Unified Studio workflows use Apache Airflow to model data processing procedures and orchestrate your Amazon SageMaker Unified Studio code artifacts. For more information, refer to the Workflows section. To learn more about Amazon SageMaker, refer to What is Amazon SageMaker?

When can I use AWS Step Functions vs. Amazon MWAA?

1. You can use Step Functions to process individual customer orders, since Step Functions can scale to meet demand for one order or one million orders.
2. If you're running an overnight workflow that processes the previous day's orders, you can use Step Functions or Amazon MWAA. Amazon MWAA gives you an open source option to abstract the workflow from the AWS resources you're using.

Environment specifications

How much task storage is available to each environment?

The task storage is limited to 20 GB, and is specified by Amazon ECS Fargate 1.4. The amount of RAM is determined by the environment class you specify. For more information about environment classes, refer to Configuring the Amazon MWAA environment class.

What is the default operating system used for Amazon MWAA environments?

Amazon MWAA environments are created on instances running Amazon Linux 2 for versions 2.6 and older, and on instances running Amazon Linux 2023 for versions 2.7 and later.

Can I use a custom image for my Amazon MWAA environment?

Custom images are not supported. Amazon MWAA uses images that are built on Amazon Linux AMI. Amazon MWAA installs the additional requirements by running `pip3 -r install` for the requirements specified in the requirements.txt file you add to the Amazon S3 bucket for the environment.

Is Amazon MWAA HIPAA compliant?

Amazon MWAA is Health Insurance Portability and Accountability Act (HIPAA) eligible. If you have a HIPAA Business Associate Addendum (BAA) in place with AWS, you can use Amazon MWAA for workflows handling Protected Health Information (PHI) on environments created on, or after, November 14th, 2022.

Does Amazon MWAA support Spot Instances?

Amazon MWAA does not currently support on-demand Amazon EC2 Spot Instance types for Apache Airflow. However, an Amazon MWAA environment can trigger Spot Instances on, for example, Amazon EMR and Amazon EC2.

Does Amazon MWAA support a custom domain?

To be able to use a custom domain for your Amazon MWAA hostname, do one of the following:

- For Amazon MWAA deployments with public web server access, you can use Amazon CloudFront with Lambda@Edge to direct traffic to your environment, and map a custom domain name to CloudFront. For more information and an example of setting up a custom domain for a public environment, refer to the Amazon MWAA custom domain for public web server sample in the Amazon MWAA examples GitHub repository.
- For Amazon MWAA deployments with private web server access, refer to the section called "Setting up a custom domain".

Can I SSH into my environment?

While SSH is not supported on a Amazon MWAA environment, it's possible to use a DAG to run bash commands using the `BashOperator`. For example:

```
from airflow import DAG
```

```
    from airflow.operators.bash_operator import BashOperator
    from airflow.utils.dates import days_ago
    with DAG(dag_id="any_bash_command_dag", schedule_interval=None, catchup=False,
start_date=days_ago(1)) as dag:
    cli_command = BashOperator(
    task_id="bash_command",
    bash_command="{{ dag_run.conf['command'] }}"
    )
```

To trigger the DAG in the Apache Airflow UI, use:

```
{ "command" : "your bash command"}
```

Why is a self-referencing rule required on the VPC security group?

By creating a self-referencing rule, you're restricting the source to the same security group in the VPC, and it's not open to all networks. To learn more, refer to the section called “Security in your VPC”.

Can I hide environments from different groups in IAM?

You can limit access by specifying an environment name in AWS Identity and Access Management, however, access filtering isn't available in the AWS console—if a user can access one environment, they can access all environments.

Can I store temporary data on the Apache Airflow worker?

Your Apache Airflow Operators can store temporary data on the workers. Apache Airflow workers can access temporary files in the `/tmp` on the Fargate containers for your environment.

> **Note**
>
> Total task storage is limited to 20 GB, according to Amazon ECS Fargate 1.4. There's no guarantee that subsequent tasks run on the same Fargate container instance, which can use a different `/tmp` folder.

Can I specify more than 25 Apache Airflow workers?

Yes. Although you can specify up to 25 Apache Airflow workers on the Amazon MWAA console, you can configure up to 50 on an environment by requesting a quota increase. For more information, refer to Requesting a quota increase.

Does Amazon MWAA support shared Amazon VPCs or shared subnets?

Amazon MWAA does not support shared Amazon VPCs or shared subnets. The Amazon VPC you select when you create an environment must be owned by the account that is attempting to create the environment. However, you can route traffic from an Amazon VPC in the Amazon MWAA account to a shared VPC. For more information and an example of routing traffic to a shared Amazon VPC, refer to Centralized outbound routing to the internet in the *Amazon VPC Transit Gateways Guide*.

Can I create or integrate custom Amazon SQS queues to manage task execution and workflow orchestration in Apache Airflow?

No, you cannot create, modify, or use custom Amazon SQS queues within Amazon MWAA. This is because Amazon MWAA automatically provisions and manages its own Amazon SQS queue for each Amazon MWAA environment.

Metrics

What metrics are used to determine whether to scale workers?

Amazon MWAA monitors the **QueuedTasks** and **RunningTasks** in CloudWatch to determine whether to scale Apache Airflow workers on your environment. To learn more, refer to *Monitoring and metrics*.

Can I create custom metrics in CloudWatch?

Not on the CloudWatch console. However, you can create a DAG that writes custom metrics in CloudWatch. For more information, refer to the section called “Using a DAG to write custom metrics”.

DAGs, Operators, Connections, and other questions

Can I use the `PythonVirtualenvOperator`?

With Amazon MWAA, you can use the `PythonVirtualenvOperator`, which comes preinstalled on every environment. Use it to run Python callables in a virtual environment with isolated dependencies. For more information about preinstalled packages, see Apache Airflow provider packages installed on Amazon MWAA environments.

How long does it take Amazon MWAA to recognize a new DAG file?

DAGs are periodically synchronized from the Amazon S3 bucket to your environment. If you add a new DAG file, it takes about 300 seconds for Amazon MWAA to start *using* the new file. If you update an existing DAG, it takes Amazon MWAA about 30 seconds to recognize your updates.

These values, 300 seconds for new DAGs, and 30 seconds for updates to existing DAGs, correspond to Apache Airflow configuration options `dag_dir_list_interval`, and `min_file_process_interval` respectively.

Why is my DAG file not picked up by Apache Airflow?

The following are possible solutions for this issue:

1. Check that your execution role has sufficient permissions to your Amazon S3 bucket. To learn more, refer to Amazon MWAA execution role.
2. Check that the Amazon S3 bucket has *Block Public Access* configured, and *Versioning* enabled. To learn more, refer to Create an Amazon S3 bucket for Amazon MWAA.
3. Verify the DAG file itself. For example, be sure that each DAG has a unique DAG ID.

Can I remove a `plugins.zip` or `requirements.txt` from an environment?

Currently, there is no way to remove a plugins.zip or requirements.txt from an environment once they've been added, but we're working on the issue. In the interim, a workaround is to point to an empty text or zip file, respectively. To learn more, refer to Deleting files on Amazon S3.

Why don't my plugins appear in the Apache Airflow v2.0.2 Admin Plugins menu?

For security reasons, the Apache Airflow webserver on Amazon MWAA has limited network egress, and does not install plugins nor Python dependencies directly on the Apache Airflow webserver for version 2.0.2 environments. The plugin that's listed allows Amazon MWAA to authenticate your Apache Airflow users in AWS Identity and Access Management (IAM).

To be able to install plugins and Python dependencies directly on the webserver, we recommend creating a new environemnt with Apache Airflow v2.2 and later. Amazon MWAA installs Python dependencies and and custom plugins directly on the webserver for Apache Airflow v2.2 and later.

Can I use AWS Database Migration Service (DMS) Operators?

Amazon MWAA supports DMS Operators. However, this operator cannot be used to perform actions on the Amazon Aurora PostgreSQL metadata database associated with an Amazon MWAA environment.

When I access the Airflow REST API using the AWS credentials, can I increase the throttling limit to more than 10 transactions per second (TPS)?

Yes, you can. To increase the throttling limit, please contact AWS Customer Support.

Where does the Airflow Task Execution API server run in Amazon MWAA?

Amazon MWAA runs the Airflow Task Execution API Server in the Webserver component. Task Execution APIs are available only in Apache Airflow v3 and later. For more information about Amazon MWAA architecture, refer to Architecture.

Troubleshooting Amazon Managed Workflows for Apache Airflow

This chapter describes common issues and errors you can encounter when using Apache Airflow on Amazon Managed Workflows for Apache Airflow and recommended steps to resolve these errors.

Contents

- Troubleshooting: DAGs, Operators, Connections, and other issues
  - Connections
    - I can't connect to Secrets Manager
    - How do I configure secretsmanager:ResourceTag/<tag-key> secrets manager conditions or a resource restriction in my execution role policy?
    - I can't connect to Snowflake
    - I can't find my connection in the Airflow UI
  - Webserver
    - I get a 5xx error accessing the webserver
    - I get a The scheduler does not seem to be running error
  - Tasks
    - I get my tasks stuck or not completing
    - I get task failures without logs in Airflow v3
  - CLI
    - I get a '503' error when triggering a DAG in the CLI
    - Why does the dags backfill Apache Airflow CLI command fail? Is there a workaround?
  - Operators
    - I received a PermissionError: [Errno 13] Permission denied error using the S3Transform operator
- Troubleshooting: Creating and updating an Amazon MWAA environment
  - Environment errors
    - AIRFLOW_COMPONENT_FAILURE
      - Before you begin
      - Stabilization and resources

        - My update fails or rolls back with a component at maximum utilization
        - My environment is stuck in UPDATING and old worker tasks will not terminate
      - Dependencies and versions
        - My scheduler or worker cannot install Python packages during the update
  - VALID_KMS_KEY_REQUIRED
    - Before you begin
    - Key policy
      - My environment fails to create with a KMS validation error, or stops logging to CloudWatch after an update
      - I get an AccessDenied error on kms:Decrypt during DAG execution
      - My key policy does not grant the other services that MWAA integrates with
    - Networking
      - My KMS failure occurs alongside VPC or security group problems
  - VPC_ENDPOINT_LIMIT_EXCEEDED
    - Infrastructure prerequisites
      - My update fails when increasing MaxWorkers
  - INCORRECT_CONFIGURATION
- Updating requirements.txt
  - I specified a new version of my requirements.txt and it's taking more than 20 minutes to update my environment
  - I get a Could not install packages due to an OSError [Error 28] No space left on device error
- Plugins
  - Does Amazon MWAA support implementing custom UI?
- Create bucket
  - I can't select the option for S3 Block Public Access settings
- Create environment
  - I tried to create an environment and it's stuck in the Creating state
  - I tried to create an environment but it displays the status as Create failed
  - I tried to select a VPC and received a Network Failure error

    - I tried to create an environment and received a service, partition, or resource "must be passed" error
    - I tried to create an environment and it displays the status as Available but when I try to access the Airflow UI an Empty Reply from Server or 502 Bad Gateway error is shown
    - I tried to create an environment and my user name is a bunch of random character names
  - Update environment
    - I tried changing the environment class but the update failed
  - Access environment
    - I can't access the Apache Airflow UI
  - Multi-team environments
    - I see "Teams require Identity Center to be configured" when I add teams
    - I get an error that a team name is invalid or already in use
    - I get an error that DAG folders overlap
    - I can't disable IAM Identity Center integration
    - Users can't see a team's DAGs
- Troubleshooting: CloudWatch Logs and CloudTrail errors
  - Logs
    - I can't find my task logs, or I received a Reading remote log from Cloudwatch log_group error
    - Tasks are failing without any logs
    - I get a ResourceAlreadyExistsException error in CloudTrail
    - I get an Invalid request error in CloudTrail
    - I get Cannot locate a 64-bit Oracle Client library: "libclntsh.so: cannot open shared object file: No such file or directory in Apache Airflow logs
    - I get psycopg2 'server closed the connection unexpectedly' in my scheduler logs
    - I get Executor reports task instance %s finished (%s) although the task says its %s in my DAG processing logs
    - Triggerer logs do not appear in the Airflow UI
    - I get Could not read remote logs from log_group: airflow-*{*environmentName}-Task log_stream:* {*DAG_ID}/*{*TASK_ID}/*{*time}/*{*n}.log. in my task logs

Troubleshooting: DAGs, Operators, Connections, and other issues

The topics on this page describe resolutions to Apache Airflow v2 and v3 Python dependencies, custom plugins, DAGs, Operators, Connections, tasks, and webserver issues you can encounter on an Amazon Managed Workflows for Apache Airflow environment.

Contents

- Connections
  - I can't connect to Secrets Manager
  - How do I configure secretsmanager:ResourceTag/<tag-key> secrets manager conditions or a resource restriction in my execution role policy?
  - I can't connect to Snowflake
  - I can't find my connection in the Airflow UI
- Webserver
  - I get a 5xx error accessing the webserver
  - I get a The scheduler does not seem to be running error
- Tasks
  - I get my tasks stuck or not completing
  - I get task failures without logs in Airflow v3
- CLI
  - I get a '503' error when triggering a DAG in the CLI
  - Why does the dags backfill Apache Airflow CLI command fail? Is there a workaround?
- Operators
  - I received a PermissionError: [Errno 13] Permission denied error using the S3Transform operator

Connections

The following topic describes the errors you might receive when using an Apache Airflow connection, or using another AWS database.

I can't connect to Secrets Manager

We recommend the following steps:

1. Learn how to create secret keys for your Apache Airflow connection and variables in the section called "Configuring Secrets Manager".
2. Learn how to use the secret key for an Apache Airflow variable (`test-variable`) in Using a secret key in AWS Secrets Manager for an Apache Airflow variable.
3. Learn how to use the secret key for an Apache Airflow connection (`myconn`) in Using a secret key in AWS Secrets Manager for an Apache Airflow connection.

How do I configure `secretsmanager:ResourceTag/<tag-key>` secrets manager conditions or a resource restriction in my execution role policy?

> **Note**
>
> Applies to Apache Airflow version 2.0 and earlier.

Currently, you cannot limit access to Secrets Manager secrets by using condition keys or other resource restrictions in your environment's execution role, due to a known issue in Apache Airflow.

I can't connect to Snowflake

We recommend the following steps:

1. Test your DAGs, custom plugins, and Python dependencies locally using aws-mwaa-docker-images on GitHub.
2. Add the following entries to the requirements.txt for your environment.

```
apache-airflow-providers-snowflake==1.3.0
```

3. Add the following imports to your DAG:

```
from airflow.providers.snowflake.operators.snowflake import SnowflakeOperator
```

Ensure the Apache Airflow connection object includes the following key-value pairs:

1. **Conn Id:** snowflake_conn
2. **Conn Type:** Snowflake
3. **Host:** <my account>.<my region if not us-west-2>.snowflakecomputing.com
4. **Schema:** <my schema>
5. **Login:** <my user name>
6. **Password:** ********
7. **Port:** <port, if any>
8. **Extra:**

```
{
    "account": "<my account>",
    "warehouse": "<my warehouse>",
    "database": "<my database>",
    "region": "<my region if not using us-west-2 otherwise omit this line>"
}
```

For example:

```
>>> import json
>>> from airflow.models.connection import Connection
>>> myconn = Connection(
...    conn_id='snowflake_conn',
...    conn_type='Snowflake',
...    host='123456789012.us-east-1.snowflakecomputing.com',
...    schema='YOUR_SCHEMA',
...    login='YOUR_USERNAME',
...    password='YOUR_PASSWORD',
...    port='YOUR_PORT',
...    extra=json.dumps(dict(account='123456789012', warehouse='YOUR_WAREHOUSE',
 database='YOUR_DB_OPTION', region='us-east-1')),
... )
```

I can't find my connection in the Airflow UI

Apache Airflow provides connection templates in the Apache Airflow UI. It uses this to generate the connection URI string, regardless of the connection type. If a connection template is not available in the Apache Airflow UI, an alternate connection template can be used to generate a connection URI string, such as using the HTTP connection template.

We recommend the following steps:

1. Access the connection types Amazon MWAA's providing in the Apache Airflow UI at Apache Airflow provider packages installed on Amazon MWAA environments.
2. Access the commands to create an Apache Airflow connection in the CLI at Apache Airflow CLI command reference.
3. Learn how to use connection templates in the Apache Airflow UI interchangeably for connection types that aren't available in the Apache Airflow UI on Amazon MWAA at Overview of connection types.

Webserver

The following topic describes the errors you might receive for your Apache Airflow webserver on Amazon MWAA.

I get a 5xx error accessing the webserver

We recommend the following steps:

1. Check Apache Airflow configuration options. Verify that the key-value pairs you specified as an Apache Airflow configuration option, such as AWS Secrets Manager, were configured correctly. To learn more, refer to the section called “I can't connect to Secrets Manager”.
2. Check the `requirements.txt`. Verify the Airflow "extras" package and other libraries listed in your `requirements.txt` are compatible with your Apache Airflow version.
3. Explore ways to specify Python dependencies in a `requirements.txt` file, refer to Managing Python dependencies in requirements.txt.

I get a `The scheduler does not seem to be running` error

If the scheduler doesn't seem to be running, or the last "heart beat" was received several hours ago, your DAGs might not be listed in Apache Airflow, and new tasks will not be scheduled.

We recommend the following steps:

1. Confirm that your VPC security group allows inbound access to port 5432. This port is needed to connect to the Amazon Aurora PostgreSQL metadata database for your environment. After this rule is added, give Amazon MWAA a few minutes, and the error can disappear. To learn more, refer to the section called “Security in your VPC”.

> **Note**
>
> - The Aurora PostgreSQL metadatabase is part of the Amazon MWAA service architecture and is not available in your AWS account.
> - Database-related errors are usually a symptom of scheduler failure and not the root cause.

2. If the scheduler is not running, it might be due to a number of factors such as dependency installation failures, or an overloaded scheduler. Confirm that your DAGs, plugins, and requirements are working correctly by accessing the corresponding log groups in CloudWatch Logs. To learn more, refer to *Monitoring and metrics*.

Tasks

The following topic describes the errors you might receive for Apache Airflow tasks in an environment.

I get my tasks stuck or not completing

If your Apache Airflow tasks are "stuck" or not completing, we recommend the following steps:

1. There might be a large number of DAGs defined. Reduce the number of DAGs and perform an update of the environment (such as changing a log level) to force a reset.

   a. Airflow parses DAGs whether they are enabled or not. If you're using greater than 50% of your environment's capacity you might start overwhelming the Apache Airflow scheduler. This leads to large *Total Parse Time* in CloudWatch Metrics or long DAG processing times in CloudWatch Logs. There are other ways to optimize Apache Airflow configurations which are outside the scope of this guide.

   b. To learn more about the best practices we recommend to tune the performance of your environment, refer to the section called “Performance tuning for Apache Airflow”.

2. There might be a large number of tasks in the queue. This is often shown as a large—and growing—number of tasks in the `None` state, or as a large number in `Queued Tasks` and/or `Tasks Pending` in CloudWatch. This can occur for the following reasons:

a. If there are more tasks to run than the environment has the capacity to run, and/or a large number of tasks that were queued before autoscaling has time to detect the tasks and deploy additional workers.

b. If there are more tasks to run than an environment has the capacity to run, we recommend **reducing** the number of tasks that your DAGs run concurrently, and/or increasing the minimum Apache Airflow workers.

c. If there are a large number of tasks that were queued before autoscaling has had time to detect and deploy additional workers, we recommend **staggering** task deployment and/or increasing the minimum Apache Airflow workers.

d. You can use the update-environment command in the AWS Command Line Interface (AWS CLI) to change the minimum or maximum number of workers that run on your environment.

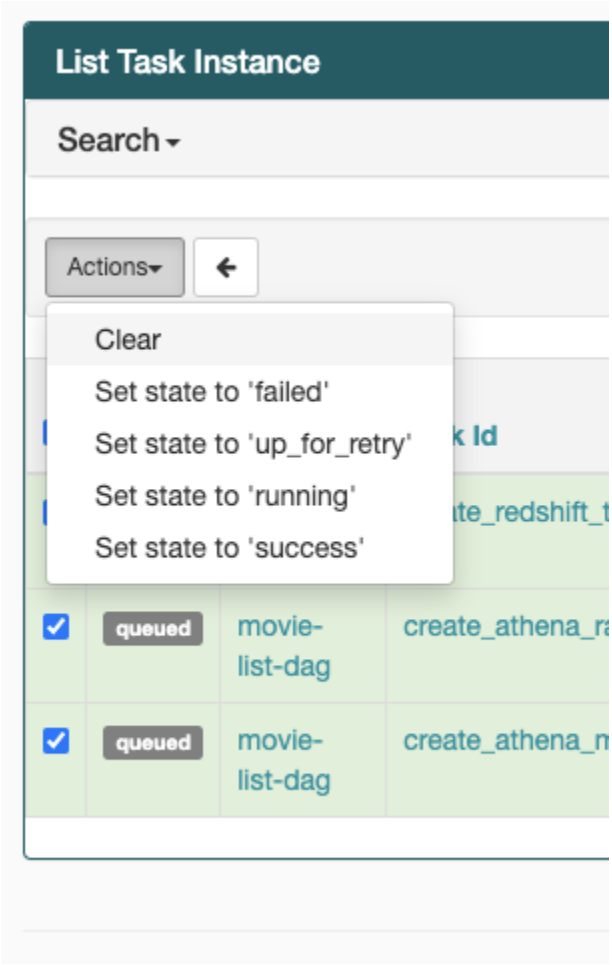
```
aws mwaa update-environment --name MyEnvironmentName --min-workers 2 --max-
workers 10
```

e. To learn more about the best practices we recommend to tune the performance of your environment, refer to the section called “Performance tuning for Apache Airflow”.

3. If your tasks are stuck in the "running" state, you can also clear the tasks or mark them as succeeded or failed. This allows the autoscaling component for your environment to scale down the number of workers running on your environment. The following image depicts an example of a stranded task.



- Choose the circle for the stranded task, and then select **Clear** (as shown). This allows Amazon MWAA to scale down workers; otherwise, Amazon MWAA can't determine which DAGs are enabled or disabled, and can't scale down, if there are still queued tasks.



4. Learn more about the Apache Airflow task lifecycle at Concepts in the *Apache Airflow reference guide*.

I get task failures without logs in Airflow v3

If your Apache Airflow 3 tasks are failing without logs, follow these steps:

- If the worker logs present an error such as `Task handler raised error: WorkerLostError('Worker exited prematurely: exitcode 15 Job: 12.')` around the time the task failed, this indicates that the forked worker process assigned to the task was likely terminated unexpectedly.

  To address this, consider configuring celery.worker_autoscale with the same minimum and maximum values. For example:

  ```
  celery.worker_autoscale=5,5  # for mw1.small
  celery.worker_autoscale=10,10 # for mw1.medium
  ```

```
celery.worker_autoscale=20,20 # for mw1.large
```

This ensures that the worker pool size remains fixed, preventing unexpected worker terminations.

CLI

The following topic describes the errors you might receive when running Airflow CLI commands in the AWS Command Line Interface.

I get a '503' error when triggering a DAG in the CLI

The Airflow CLI runs on the Apache Airflow webserver, which has limited concurrency. Typically a maximum of 4 CLI commands can run simultaneously.

Why does the `dags backfill` Apache Airflow CLI command fail? Is there a workaround?

> **Note**
>
> The following applies only to Apache Airflow v2.0.2 environments.

The `backfill` command, like other Apache Airflow CLI commands, parses all DAGs locally before any DAGs are processed, regardless of which DAG the CLI operation applies to. In Amazon MWAA environments using Apache Airflow v2.0.2, because plugins and requirements are not yet installed on the webserver by the time the CLI command runs, the parsing operation fails, and the `backfill` operation is not invoked. If you did not have any requirements nor plugins in your environment, the `backfill` operation would succeed.

To be able to run the `backfill` CLI command, we recommend invoking it in a bash operator. In a bash operator, `backfill` is initiated from the worker, allowing the DAGs to parse successfully as all necessary requirements and plugins are available and installed. Use the following example to create a DAG with a `BashOperator` to run `backfill`.

```
from airflow import DAG
from airflow.operators.bash_operator import BashOperator
from airflow.utils.dates import days_ago
```

```
with DAG(dag_id="backfill_dag", schedule_interval=None, catchup=False,
 start_date=days_ago(1)) as dag:
    cli_command = BashOperator(
        task_id="bash_command",
        bash_command="airflow dags backfill my_dag_id"
    )
```

Operators

The following topic describes the errors you might receive when using Operators.

I received a `PermissionError: [Errno 13] Permission denied` error using the S3Transform operator

We recommend the following steps if you're trying to run a shell script with the *S3Transform* operator and you're receiving a `PermissionError: [Errno 13] Permission denied` error. The following steps assume you have an existing plugins.zip file. If you're creating a *new* plugins.zip, refer to Installing custom plugins.

1. Test your DAGs, custom plugins, and Python dependencies locally using aws-mwaa-docker-images on GitHub.
2. Create your "transform" script.

   ```
   #!/bin/bash
   cp $1 $2
   ```

3. (optional) macOS and Linux users might need to run the following command to ensure the script is executable.

   ```
   chmod 777 transform_test.sh
   ```

4. Add the script to your plugins.zip.

   ```
   zip plugins.zip transform_test.sh
   ```

5. Follow the steps in Upload the plugins.zip to Amazon S3.
6. Follow the steps in Specifying the plugins.zip version on the Amazon MWAA console.
7. Create the following DAG.

```
from airflow import DAG
from airflow.providers.amazon.aws.operators.s3_file_transform import
 S3FileTransformOperator
from airflow.utils.dates import days_ago
import os

DAG_ID = os.path.basename(__file__).replace(".py", "")

with DAG (dag_id=DAG_ID, schedule_interval=None, catchup=False,
 start_date=days_ago(1)) as dag:
    file_transform = S3FileTransformOperator(
        task_id='file_transform',
        transform_script='/usr/local/airflow/plugins/transform_test.sh',
        source_s3_key='s3://amzn-s3-demo-bucket/files/input.txt',
        dest_s3_key='s3://amzn-s3-demo-bucket/files/output.txt'
    )
```

8. Follow the steps in Uploading DAG code to Amazon S3.

Troubleshooting: Creating and updating an Amazon MWAA environment

The topics on this page contain errors you can encounter when creating and updating an Amazon Managed Workflows for Apache Airflow environment and how to resolve these errors.

Contents

- Environment errors
  - AIRFLOW_COMPONENT_FAILURE
    - Before you begin
    - Stabilization and resources
      - My update fails or rolls back with a component at maximum utilization
      - My environment is stuck in UPDATING and old worker tasks will not terminate
    - Dependencies and versions
      - My scheduler or worker cannot install Python packages during the update
  - VALID_KMS_KEY_REQUIRED
    - Before you begin

- Key policy
  - My environment fails to create with a KMS validation error, or stops logging to CloudWatch after an update
  - I get an AccessDenied error on kms:Decrypt during DAG execution
  - My key policy does not grant the other services that MWAA integrates with
- Networking
  - My KMS failure occurs alongside VPC or security group problems
- VPC_ENDPOINT_LIMIT_EXCEEDED
  - Infrastructure prerequisites
    - My update fails when increasing MaxWorkers
- INCORRECT_CONFIGURATION
- Updating requirements.txt
  - I specified a new version of my requirements.txt and it's taking more than 20 minutes to update my environment
  - I get a Could not install packages due to an OSError [Error 28] No space left on device error
- Plugins
  - Does Amazon MWAA support implementing custom UI?
- Create bucket
  - I can't select the option for S3 Block Public Access settings
- Create environment
  - I tried to create an environment and it's stuck in the Creating state
  - I tried to create an environment but it displays the status as Create failed
  - I tried to select a VPC and received a Network Failure error
  - I tried to create an environment and received a service, partition, or resource "must be passed" error
  - I tried to create an environment and it displays the status as Available but when I try to access the Airflow UI an Empty Reply from Server or 502 Bad Gateway error is shown
  - I tried to create an environment and my user name is a bunch of random character names
- Update environment
  - I tried changing the environment class but the update failed
- Access environment

- I can't access the Apache Airflow UI
- Multi-team environments
  - I see "Teams require Identity Center to be configured" when I add teams
  - I get an error that a team name is invalid or already in use
  - I get an error that DAG folders overlap
  - I can't disable IAM Identity Center integration
  - Users can't see a team's DAGs

Environment errors

The following topics describe the errors you might receive for your Amazon MWAA environment.

`AIRFLOW_COMPONENT_FAILURE`

The worker, scheduler, or webserver component fails to stabilize during environment creation or update.

Before you begin

Before you apply a resolution, determine which update phase the environment is in and which component is failing the stabilization check. The recurring signal is a component failing its health check during stabilization, not a steady-state crash.

We recommend the following steps:

1. Determine whether the environment is in the `UPDATING` or `ROLLING_BACK` state, or is stuck.
2. Identify whether the scheduler, worker, or webserver is the component that fails the stabilization check.
3. Pull the affected component's CloudWatch Logs for the update window.
4. Review the following metrics for the update window:
   - CPU and memory utilization for the affected component
   - Metadata database freeable memory

Stabilization and resources

My update fails or rolls back with a component at maximum utilization

Resource saturation during stabilization is the most common cause of `AIRFLOW_COMPONENT_FAILURE`. Symptoms include the update failing or rolling back, the scheduler or webserver at maximum utilization, and slow worker stabilization.

We recommend the following steps:

- Right-size the environment class.
- Reduce load before you retry the update:
  - Fix DAG import errors and high parse time.
  - Pause and restore DAGs gradually rather than all at once.
  - Align your Apache Airflow concurrency settings with your worker capacity.
- Increase the metadata database capacity to address worker instability, because low freeable memory can block stabilization.

For more information, see Performance tuning for Apache Airflow on Amazon MWAA.

My environment is stuck in `UPDATING` and old worker tasks will not terminate

If the environment is stuck in the `UPDATING` state, tasks on old workers do not terminate. Graceful updates then fail because worker replacement cannot drain running tasks.

We recommend the following steps:

- Set the worker replacement strategy to `FORCED`. Forced updates stop running tasks so that the update does not wait for them to drain, which prevents the environment from getting stuck.
- If the environment is already stuck, release the workers so the update can complete. To do this, mark running tasks as failed.
- To prevent recurrence, pause your DAGs, drain running tasks, and run the update when no extra workers are active.

Dependencies and versions

My scheduler or worker cannot install Python packages during the update

If the scheduler or worker cannot install Python packages mid-update, the install is likely failing because of a transient `pip` or network stall. This stall typically occurs while pulling from public PyPI.

We recommend the following steps:

- Deliver dependencies locally by using `plugins.zip` instead of public PyPI.
- Pre-download and pin wheels so the install can survive transient `pip` or network stalls during the update.

`VALID_KMS_KEY_REQUIRED`

Check that your AWS KMS customer managed key is enabled, and that the permissions policy attached to the key grants access to your environment.

Before you begin

Before you edit the key policy, confirm where the failure occurs and that the key itself is valid. The principals you add to the key policy differ depending on whether the failure happens at creation, at update, or at runtime during DAG execution.

We recommend the following steps:

1. Determine where the failure occurs:
   - Most occurrences are at environment creation, and some at update.
   - A few are runtime `kms:Decrypt` `AccessDenied` errors during DAG execution.
2. Confirm that the key is a valid, enabled customer managed key in the same AWS Region and account as the environment. Verify that you specified the key ARN correctly, and that the configuration does not reference an empty or invalid AWS KMS resource.

Key policy

Several key policy problems can cause this error. Review the following scenarios, and apply the fix that matches your symptoms.

My environment fails to create with a KMS validation error, or stops logging to CloudWatch after an update

A missing CloudWatch Logs grant in the key policy is the most common root cause of `VALID_KMS_KEY_REQUIRED`. Symptoms include creation failing with the AWS KMS validation error, or components that stop logging to CloudWatch Logs after an update.

We recommend the following steps:

- Add the CloudWatch Logs service principal, `logs.region.amazonaws.com`, to the AWS KMS key policy with permission to use the key.
- Scope the grant by using the `kms:ViaService` condition for CloudWatch Logs.

I get an AccessDenied error on kms:Decrypt during DAG execution

Several symptoms point to this problem: a runtime `AccessDenied` error on `kms:Decrypt`, missing dependencies or import errors after a container is replaced, or a AWS KMS access-denied error during DAG execution. These symptoms indicate that the Amazon MWAA execution role lacks the required permissions on the key, or that the key policy scope is too restrictive.

We recommend the following steps:

- Grant the Amazon MWAA execution role the required AWS KMS actions on the key, notably `kms:Decrypt`, as well as `kms:Encrypt` and `kms:GenerateDataKey` as applicable. These permissions are required to access encrypted Amazon S3 objects (your DAGs, plugins, and `requirements.txt`) and other encrypted dependencies.
- Correct a policy scope that is too restrictive.

My key policy does not grant the other services that MWAA integrates with

Creation validation checks that the key is usable by the full set of dependent principals, not CloudWatch Logs alone. If the key policy grants only CloudWatch Logs, validation can still fail.

We recommend the following steps:

- Make sure that the key policy also allows the other services that Amazon MWAA integrates with and that need the customer managed key, including Amazon S3, Amazon SQS, and the execution role.

Networking

Networking problems can occur alongside this error. Review the following scenario.

My KMS failure occurs alongside VPC or security group problems

Several AWS KMS failures occur in combination with Amazon VPC or security group misconfiguration. Networking problems can block the environment from validating or using the key. When you fix the key policy at creation time, verify your networking in the same pass.

We recommend the following steps:

- Check your security group inbound and outbound rules.
- Verify that your Amazon VPC endpoints are reachable, including the AWS KMS endpoint, so that networking does not block key validation and use.

For more information, see Using customer managed keys for encryption.

`VPC_ENDPOINT_LIMIT_EXCEEDED`

The Amazon VPC associated with your environment has reached the maximum number of Amazon VPC endpoints. Delete unused Amazon VPC endpoints or request a service quota increase before creating the environment.

Infrastructure prerequisites

Environment infrastructure limits can cause an update to fail.

My update fails when increasing `MaxWorkers`

An update that increases `MinWorkers` or `MaxWorkers` can fail if the environment's subnets do not have enough free IP addresses for the additional workers.

We recommend the following steps:

- Reclaim orphaned elastic network interfaces (ENIs) to free IP addresses.
- Alternatively, move the environment to larger subnets with greater IP capacity.

INCORRECT_CONFIGURATION

Check the execution role permissions policy for your environment. Verify that each Amazon VPC networking component required by the environment allows traffic.

Updating `requirements.txt`

The following topic describes the errors you might receive when updating your `requirements.txt`.

I specified a new version of my `requirements.txt` and it's taking more than 20 minutes to update my environment

If it takes more than twenty minutes for your environment to install a new version of a `requirements.txt` file, the environment update failed and Amazon MWAA is rolling back to the last stable version of the container image.

1. Check package versions. We recommend always specifying either a specific version (==) or a maximum version (<=) for the Python dependencies in your `requirements.txt`.
2. Check Apache Airflow logs. If you enabled Apache Airflow logs, verify your log groups were created successfully on the Logs groups page on the CloudWatch console. If you get blank logs, the most common reason is due to missing permissions in your execution role for CloudWatch or Amazon S3 where logs are written. To learn more, refer to Execution role.
3. Check Apache Airflow configuration options. If you're using Secrets Manager, verify that the key-value pairs you specified as an Apache Airflow configuration option were configured correctly. To learn more, refer to the section called “Configuring Secrets Manager”.
4. Check VPC network configuration. To learn more, refer to the section called “Environment stuck”.
5. Check execution role permissions. An execution role is an AWS Identity and Access Management (IAM) role with a permissions policy that grants Amazon MWAA permission to invoke the resources of other AWS services (such as Amazon S3, CloudWatch, Amazon SQS, Amazon ECR) on your behalf. Your Customer-managed key or AWS-owned key also needs to be permitted access. To learn more, refer to Execution role.
6. To run a troubleshooting script that checks the Amazon VPC network setup and configuration for your Amazon MWAA environment, refer to the Verify Environment script in AWS Support Tools on GitHub.

I get a `Could not install packages due to an OSError [Error 28] No space left on device` error

Amazon MWAA containers have a finite amount of storage for the OS, DAGs, runtimes, and libraries. For more information, see the section called "How much task storage is available to each environment?".

To reduce the storage requirement, you might be able to use `--no-cache-dir` by installing the libraries with a startup script. For example:

```
#!/bin/sh

if [[ "${MWAA_AIRFLOW_COMPONENT}" != "webserver" ]]
then
     # place requirements.txt in DAGs folder, and replace existing requirements.txt
 with an empty text file
     pip3 install --no-cache-dir -r /usr/local/airflow/dags/requirements.txt
fi
```

For more information, see the section called "Configure a startup script".

You can also install your libraries from Python wheel files. For more information, see the section called "Option two: Python wheels (.whl)".

Plugins

The following topic describes issues you can encounter when configuring or updating Apache Airflow plugins.

Does Amazon MWAA support implementing custom UI?

Starting with Apache Airflow v2.2.2, Amazon MWAA supports installing plugins on the Apache Airflow webserver, and implementing custom UI. If your Amazon MWAA environment is running Apache Airflow v2.0.2 or older, you will not be able to implement custom UI.

For more information about version management, and upgrading your existing environments, refer to *Versions*.

Create bucket

The following topic describes the errors you might receive when creating an Amazon S3 bucket.

I can't select the option for S3 Block Public Access settings

The execution role for your Amazon MWAA environment needs permission to the `GetBucketPublicAccessBlock` action on the Amazon S3 bucket to verify the bucket blocked public access. We recommend the following steps:

1. Follow the steps to Attach a JSON policy to your execution role.
2. Attach the following JSON policy:

```
{
  "Effect":"Allow",
  "Action":[
    "s3:GetObject*",
    "s3:GetBucket*",
    "s3:List*"
  ],
  "Resource":[
    "arn:aws:s3:::amzn-s3-demo-bucket",
    "arn:aws:s3:::amzn-s3-demo-bucket/*"
  ]
}
```

   Substitute the sample placeholders in *amzn-s3-demo-bucket* with your Amazon S3 bucket name.
3. To run a troubleshooting script that checks the Amazon VPC network setup and configuration for your Amazon MWAA environment, refer to the Verify Environment script in AWS Support Tools on GitHub.

Create environment

The following topic describes the errors you might receive when creating an environment.

I tried to create an environment and it's stuck in the `Creating` state

We recommend the following steps:

1. Check VPC network with *public routing*. If you're using an Amazon VPC with internet access, verify the following:

- That your Amazon VPC is configured to allow network traffic between the different AWS resources used by your Amazon MWAA environment, as defined in the section called "About networking". For example, your VPC security group must either allow all traffic in a self-referencing rule, or optionally specify the port range for HTTPS port range 443 and a TCP port range 5432.

2. Check VPC network with *private routing*. If you're using an Amazon VPC without internet access, verify the following:

   - That your Amazon VPC is configured to allow network traffic between the different AWS resources for your Amazon MWAA environment, as defined in the section called "About networking". For example, your two private subnets must **not** have a route table to a NAT gateway (or NAT instance), **nor** an internet gateway.

3. To run a troubleshooting script that checks the Amazon VPC network setup and configuration for your Amazon MWAA environment, refer to the Verify Environment script in AWS Support Tools on GitHub.

I tried to create an environment but it displays the status as `Create failed`

We recommend the following steps:

1. Check VPC network configuration. To learn more, refer to the section called "Environment stuck".
2. Check user permissions. Amazon MWAA performs a dry run against a user's credentials before creating an environment. Your AWS account might not have permission in AWS Identity and Access Management (IAM) to create some of the resources for an environment. For example, if you chose the **Private network** Apache Airflow access mode, your AWS account must have been granted access by your administrator to the AmazonMWAAFullConsoleAccess access control policy for your environment, which allows your account to create VPC endpoints.
3. Check execution role permissions. An execution role is an AWS Identity and Access Management (IAM) role with a permissions policy that grants Amazon MWAA permission to invoke the resources of other AWS services (such as Amazon S3, CloudWatch, Amazon SQS, Amazon ECR) on your behalf. Your Customer-managed key or AWS-owned key also needs to be permitted access. To learn more, refer to Execution role.
4. Check Apache Airflow logs. If you enabled Apache Airflow logs, verify your log groups were created successfully on the Logs groups page on the CloudWatch console. If you get blank

logs, the most common reason is due to missing permissions in your execution role for CloudWatch or Amazon S3 where logs are written. To learn more, refer to Execution role.

5. To run a troubleshooting script that checks the Amazon VPC network setup and configuration for your Amazon MWAA environment, refer to the Verify Environment script in AWS Support Tools on GitHub.
6. If you are using an Amazon VPC *without* internet access, ensure that you've created an Amazon S3 gateway endpoint, and granted the minimum required permissions to Amazon ECR to access Amazon S3. To learn more about creating an Amazon S3 gateway endpoint, refer to the following:

   - Creating an Amazon VPC network without internet access
   - Create the Amazon S3 gateway endpoint in the *Amazon Elastic Container Registry User Guide*

I tried to select a VPC and received a `Network Failure` error

We recommend the following steps:

- If you get a `Network Failure` error when you try to select an Amazon VPC when creating your environment, turn off any in-browser proxies that are running, and then try again.

I tried to create an environment and received a service, partition, or resource "must be passed" error

We recommend the following steps:

- You might be receiving this error because the URI you specified for your Amazon S3 bucket includes a '/' at the end of the URI. We recommend removing the '/' in the path. The value must be in the following format:

  ```
  s3://amzn-s3-demo-bucket
  ```

I tried to create an environment and it displays the status as `Available` but when I try to access the Airflow UI an `Empty Reply from Server` or `502 Bad Gateway` error is shown

We recommend the following steps:

1. Check VPC security group configuration. To learn more, refer to the section called "Environment stuck".
2. Confirm that any Apache Airflow packages you listed in the `requirements.txt` correspond to the Apache Airflow version you're running on Amazon MWAA. To learn more, refer to Installing Python dependencies.
3. To run a troubleshooting script that checks the Amazon VPC network setup and configuration for your Amazon MWAA environment, refer to the Verify Environment script in AWS Support Tools on GitHub.

I tried to create an environment and my user name is a bunch of random character names

- Apache Airflow has a maximum of 64 characters for user names. If your AWS Identity and Access Management (IAM) role exceeds this length, a hash algorithm is used to reduce it, while remaining unique.

Update environment

The following topic describes the errors you might receive when updating an environment.

I tried changing the environment class but the update failed

If you update your environment to a different environment class (such as changing an `mw1.medium` to an `mw1.small`), and the request to update your environment failed, the environment status goes into an `UPDATE_FAILED` state and the environment is rolled back to, and is billed according to, the previous stable version of an environment.

We recommend the following steps:

1. Test your DAGs, custom plugins, and Python dependencies locally using aws-mwaa-docker-images on GitHub.
2. To run a troubleshooting script that checks the Amazon VPC network setup and configuration for your Amazon MWAA environment, refer to the Verify Environment script in AWS Support Tools on GitHub.

Access environment

The following topic describes the errors you might receive when accessing an environment.

I can't access the Apache Airflow UI

We recommend the following steps:

1. Check user permissions. You might not have been granted access to a permissions policy that you can use to access the Apache Airflow UI. To learn more, refer to the section called "Accessing an Amazon MWAA environment".
2. Check network access. This might be because you selected the **Private network** access mode. If the URL of your Apache Airflow UI is in the following format `387fbcn-8dh4-9hfj-0dnd-834jhdfb-vpce.c10.us-west-2.airflow.amazonaws.com`, it means that you're using *private routing* for your Apache Airflow webserver. You can either update the Apache Airflow access mode to the **Public network** access mode, or create a mechanism to access the VPC endpoint for your Apache Airflow *Web server*. To learn more, refer to the section called "Managing access to VPC endpoints".

Multi-team environments

The following topics describe the errors you might receive when you configure teams in a multi-team environment. For team name and DAG folder rules, see Multi-team environments.

I see "Teams require Identity Center to be configured" when I add teams

The Amazon MWAA console shows the following message when IAM Identity Center integration isn't enabled: `Teams require Identity Center to be configured. Enable Identity Center above before adding teams.`

Enable IAM Identity Center integration on the same page, and then add the teams. For more information, see Using IAM Identity Center with Amazon Managed Workflows for Apache Airflow.

I get an error that a team name is invalid or already in use

Team names must be unique within the environment, be 3 to 50 characters long, start with a letter or number, and contain only letters, numbers, hyphens (-), and underscores (_). Choose a name that meets these requirements and matches the display name of an IAM Identity Center group.

I get an error that DAG folders overlap

A team folder can't be inside the global DAG folder or another team's folder, and the global folder can't be inside a team folder. For example, if the global DAG folder is `s3://amzn-s3-demo-bucket/dags/`, use `s3://amzn-s3-demo-bucket/teams/team-a/` for a team instead of `s3://amzn-s3-demo-bucket/dags/team-a/`.

I can't disable IAM Identity Center integration

After you enable IAM Identity Center integration on an environment, you can't disable it, even if you remove all teams. To stop using IAM Identity Center, create a new environment without it. For more information, see Known limitations.

Users can't see a team's DAGs

A user can see a team's DAGs only if they belong to the team's IAM Identity Center group and their Permission Set policy allows the access. We recommend the following steps:

1. Check that an IAM Identity Center group exists whose display name exactly matches the team name. Matching is case-sensitive, so a team named `data-engineering` doesn't match a group named `Data-Engineering`. If no group matches, the team has no members.
2. Check that the user is a member of that IAM Identity Center group.
3. Check the user's Permission Set policy. If it uses the `airflow:TeamNames` condition key, use the `ForAnyValue:StringEquals` operator and include the team name. For examples, see Using IAM Identity Center with Amazon Managed Workflows for Apache Airflow.

Troubleshooting: CloudWatch Logs and CloudTrail errors

The topics on this page contain resolutions to Amazon CloudWatch Logs and AWS CloudTrail errors you can encounter on an Amazon Managed Workflows for Apache Airflow environment.

Contents

- Logs
  - I can't find my task logs, or I received a Reading remote log from Cloudwatch log_group error
  - Tasks are failing without any logs
  - I get a ResourceAlreadyExistsException error in CloudTrail

- I get an Invalid request error in CloudTrail
- I get Cannot locate a 64-bit Oracle Client library: "libclntsh.so: cannot open shared object file: No such file or directory in Apache Airflow logs
- I get psycopg2 'server closed the connection unexpectedly' in my scheduler logs
- I get Executor reports task instance %s finished (%s) although the task says its %s in my DAG processing logs
- Triggerer logs do not appear in the Airflow UI
- I get Could not read remote logs from log_group: airflow-*{*environmentName}-Task log_stream:* {*DAG_ID}/*{*TASK_ID}/*{*time}/*{*n}.log. in my task logs

Logs

The following topic describes the errors you might receive when accessing Apache Airflow logs.

I can't find my task logs, or I received a `Reading remote log from Cloudwatch log_group` error

Amazon MWAA has configured Apache Airflow to read and write logs directly from and to Amazon CloudWatch Logs. If a worker fails to start a task, or fails to write any logs, you will refer to the error:

```
*** Reading remote log from Cloudwatch log_group: airflow-environmentName-Task
 log_stream: DAG_ID/TASK_ID/timestamp/n.log.Could not read remote logs from log_group:
 airflow-environmentName-Task log_stream: DAG_ID/TASK_ID/time/n.log.
```

- We recommend the following steps:

  a. Verify that you have enabled task logs at the `INFO` level for your environment. For more information, refer to Accessing Airflow logs in Amazon CloudWatch.

  b. Verify that the environment execution role has the correct permission policies.

  c. Verify that your operator or task is working correctly, has sufficient resources to parse the DAG, and has the appropriate Python libraries to load. To verify whether you have the correct dependencies, try eliminating imports until you find the one that is causing the issue. We recommend testing your Python dependencies using aws-mwaa-docker-images.

Tasks are failing without any logs

If tasks are failing in a workflow and you can't locate any logs for the failed tasks, check if you are setting the queue parameter in your default arguments, as listed in the following.

```
from airflow import DAG
from airflow.operators.bash_operator import BashOperator
from airflow.utils.dates import days_ago

# Setting queue argument to default.
default_args = {
 "start_date": days_ago(1),
 "queue": "default"
}

with DAG(dag_id="any_command_dag", schedule_interval=None, catchup=False,
 default_args=default_args) as dag:
    cli_command = BashOperator(
        task_id="bash_command",
        bash_command="{{ dag_run.conf['command'] }}"
    )
```

To resolve the issue, remove queue from your code, and invoke the DAG again.

I get a `ResourceAlreadyExistsException` error in CloudTrail

```
"errorCode": "ResourceAlreadyExistsException",
    "errorMessage": "The specified log stream already exists",
    "requestParameters": {
        "logGroupName": "airflow-MyAirflowEnvironment-DAGProcessing",
        "logStreamName": "scheduler_cross-account-eks.py.log"
    }
```

Certain Python requirements such as `apache-airflow-backport-providers-amazon` roll back the `watchtower` library that Amazon MWAA uses to communicate with CloudWatch to an older version. We recommend the following steps:

- Add the following library to your `requirements.txt`

  ```
  watchtower==1.0.6
  ```

I get an `Invalid request` error in CloudTrail

```
Invalid request provided: Provided role does not have sufficient permissions for s3
 location airflow-xxx-xxx/dags
```

If you're creating an Amazon MWAA environment and an Amazon S3 bucket using the same CloudFormation template, you need to add a `DependsOn` section within your CloudFormation template. The two resources (*MWAA Environment* and *MWAA Execution Policy*) have a dependency in CloudFormation. We recommend the following steps:

- Add the following **`DependsOn`** statement to your CloudFormation template.

```
...
      MaxWorkers: 5
      NetworkConfiguration:
        SecurityGroupIds:
          - !GetAtt SecurityGroup.GroupId
        SubnetIds: !Ref subnetIds
      WebserverAccessMode: PUBLIC_ONLY
    DependsOn: MwaaExecutionPolicy

  MwaaExecutionPolicy:
    Type: AWS::IAM::ManagedPolicy
    Properties:
      Roles:
        - !Ref MwaaExecutionRole
      PolicyDocument:
        Version: 2012-10-17
        Statement:
          - Effect: Allow
            Action: airflow:PublishMetrics
            Resource:
...
```

  For an example, refer to Quick start tutorial for Amazon Managed Workflows for Apache Airflow.

I get `Cannot locate a 64-bit Oracle Client library: "libclntsh.so: cannot open shared object file: No such file or directory` in Apache Airflow logs

- We recommend the following steps:
    - If you're using Apache Airflow v2, add `core.lazy_load_plugins : False` as an Apache Airflow configuration option. To learn more, refer to Using configuration options to load plugins in 2.

I get psycopg2 'server closed the connection unexpectedly' in my scheduler logs

If you get an error similar to the following, your Apache Airflow scheduler might have run out of resources.

```
2021-06-14T10:20:24.581-05:00    sqlalchemy.exc.OperationalError:
 (psycopg2.OperationalError) server closed the connection unexpectedly
2021-06-14T10:20:24.633-05:00    This probably means the server terminated abnormally
2021-06-14T10:20:24.686-05:00    before or while processing the request.
```

We recommend the following steps:

- Consider upgrading to Apache Airflow v2.0.2, which you can use to specify up to 5 schedulers.

I get `Executor reports task instance %s finished (%s) although the task says its %s` in my DAG processing logs

If you get an error similar to the following, your long-running tasks might have reached the task time limit on Amazon MWAA. Amazon MWAA has a limit of 12 hours for any one Airflow task, to prevent tasks from getting stuck in the queue and blocking activities like autoscaling.

```
Executor reports task instance %s finished (%s) although the task says its %s. (Info:
 %s) Was the task killed externally
```

We recommend the following steps:

- Consider breaking up the task into multiple, shorter running tasks. Airflow typically has a model whereby operators are asynchronous. It invokes activities on external systems, and

Apache Airflow Sensors poll to check when it's complete. If a Sensor fails, it can be safely retried without impacting the Operator's functionality.

Triggerer logs do not appear in the Airflow UI

If triggerer logs do not appear in the Apache Airflow UI, your execution role might be missing the `logs:DescribeLogStreams` permission. Without this permission, Amazon MWAA cannot discover triggerer log streams. No error appears in the logs, which can make this issue difficult to identify.

Add `logs:DescribeLogStreams` to your execution role. Scope the permission to your environment's log group ARN. For more information, see Amazon MWAA execution role.

> **Note**
>
> This issue affects your environment if you use Apache Airflow version 3.x with deferrable operators. If you created your execution role in the console before the default role included this permission, add it manually.

I get `Could not read remote logs from log_group: airflow-*{*environmentName}-Task log_stream:* {*DAG_ID}/*{*TASK_ID}/*{*time}/*{*n}.log.` in my task logs

If you get an error similar to the following, the execution role for your environment might not contain a permissions policy to create log streams for task logs.

```
Could not read remote logs from log_group: airflow-*{*environmentName}-Task
 log_stream:* {*DAG_ID}/*{*TASK_ID}/*{*time}/*{*n}.log.
```

We recommend the following steps:

- Modify the execution role for your environment using one of the sample policies at the section called “Execution role”.

You might have also specified a provider package in your `requirements.txt` file that is incompatible with your Apache Airflow version. For example, if you're using Apache Airflow v2.0.2,

you might have specified a package, such as the apache-airflow-providers-databricks package, which is only compatible with Airflow 2.1+.

We recommend the following steps:

1. If you're using Apache Airflow v2.0.2, modify the `requirements.txt` file and add `apache-airflow[databricks]`. This installs the correct version of the Databricks package that is compatible with Apache Airflow v2.0.2.
2. Test your DAGs, custom plugins, and Python dependencies locally using aws-mwaa-docker-images on GitHub.

Amazon MWAA User Guide history

The following table describes important additions to the Amazon MWAA service documentation, beginning in November 2020. To receive notifications about updates to this documentation, subscribe to the RSS feed.

Change	Description	Date
IAM Identity Center integration	Added documentation for integrating Amazon Managed Workflows for Apache Airflow with IAM Identity Center to authenticate users through a corporate identity provider and control access using Permission Sets. This feature requires Apache Airflow version 3.3.1 or later. • the section called “IAM Identity Center integration” • IAM Identity Center Permission Set policy examples • the section called “Log in using IAM Identity Center” • Step two: Configure advanced settings	September 24, 2026
New Apache Airflow version	Amazon MWAA now supports Apache Airflow v3.3.1. This update includes information about updated provider packages, and details about using Apache Airflow v3.3.1 on Amazon MWAA.	September 1, 2026

	• *Versions* • the section called “Version-specific provider packages”	
Shared responsibility model	Added a topic that explains the infrastructure that we manage versus the workflow code, dependencies, and configuration that you supply. • the section called “Shared responsibility model”	August 11, 2026
New Apache Airflow version	Amazon MWAA now supports Apache Airflow v2.11.2. This update includes informati on about updated provider packages, and details about using Apache Airflow v2.11.2 on Amazon MWAA. • *Versions* • the section called “Version-specific provider packages”	July 21, 2026

New webserver access mode	Amazon MWAA now supports a third webserver access mode, **Both public and private network access**, for environments running Apache Airflow v3.2.1 and later. This mode creates both a public network load balancer for browser access to the Apache Airflow UI and a private VPC endpoint for worker-to-webserver communication, allowing workers to reach the webserver in VPCs without internet access. • the section called “Apache Airflow access modes” • the section called “About networking”	May 26, 2026

New Apache Airflow version	Amazon MWAA now supports Apache Airflow v3.2.1. This update includes informati on about updated provider packages, details about using Apache Airflow v3.2.1 on Amazon MWAA, a new section on unsupported Apache Airflow configuration options, and a new database cleanup code sample for Apache Airflow v3. • *Versions* • the section called "Version-specific provider packages" • the section called "Unsupporta configurations" • the section called "Aurora PostgreSQL database cleanup"	May 19, 2026
New Apache Airflow version	Amazon MWAA now supports Apache Airflow v2.11.0. This update includes informati on about updated provider packages, and details about using Apache Airflow v2.11.0 on Amazon MWAA. • *Versions* • the section called "Version-specific provider packages"	January 7, 2026

Added Apache Airflow v3 support	Updated documentation to include support for Apache Airflow v3 • *Code examples* • *Versions*	October 1, 2025
IPv6 updates	Added information about IPv6 support. • the section called "About networking"	August 26, 2025
Environment updates	Added a note about the `workerReplacementStrategy` changing from `GRACEFUL` to `FORCED` if you perform an update while your environment is in `MAINTENANCE` status. • the section called "Update an environment"	August 6, 2025
Version deprecation information	Updated topic on version deprecation to include deprecation notices and timelines for Apache Airflow v2.4.3, Apache Airflow v2.5.1, and Apache Airflow v2.6.3. • the section called "Apache Airflow deprecated versions"	June 24, 2025

Added a new environment class: mw1.micro	Amazon MWAA now provides a new environment class: mw1.micro. • the section called "Configuring the environment class" • the section called "Performance tuning for Apache Airflow"	November 19, 2024
Support for simpler method to access Apache Airflow REST API	Amazon MWAA now provides a simplified approach for interacting with the Apache Airflow REST API using AWS credentials. • the section called "Using the Apache Airflow REST API" • the section called "Apache Airflow Rest API access"	October 23, 2024
New Apache Airflow version	Amazon MWAA now supports Apache Airflow v2.10.1. This update includes informati on about updated provider packages, and details about using Apache Airflow v2.10.1 on Amazon MWAA. • *Versions* • the section called "Version-specific provider packages"	September 26, 2024

New Apache Airflow version	Amazon MWAA now supports Apache Airflow v2.9.2. This update includes informati on about updated provider packages, and details about using Apache Airflow v2.9.2 on Amazon MWAA. • *Versions* • the section called "Version-specific provider packages"	July 9, 2024
Amazon MWAA supports configuring a custom webserver domain names	Amazon MWAA supports configuring a custom webserver domain names for private environments with no internet access. This update includes the following new topic that describes setting up a new custom domain. • the section called "Setting up a custom domain"	June 18, 2024
Amazon MWAA supports webserver automatic scaling and the Apache Airflow REST API	Amazon MWAA now supports automatic scaling of webservers as well as the ability to access and use the Apache Airflow REST API. • the section called "Configur ing webserver auto scaling" • the section called "Using the Apache Airflow REST API"	May 16, 2024

Improved description of automatic scaling behavior	Updated the following topic to reflect the new Amazon MWAA automatic scaling behavior when workers pick up new tasks as Fargate workers downscale. • the section called “Configuring worker auto scaling”	May 10, 2024
Support for larger instance sizes	Amazon MWAA now supports two larger instance size options for larger workloads : `mw1.xlarge` , and `mw1.2xlarge` • the section called “Environment capabilities”	April 16, 2024
New Apache Airflow version	Amazon MWAA now supports Apache Airflow v2.8.1. This update includes informati on about updated provider packages, and details about using Apache Airflow v2.8.1 on Amazon MWAA. • *Versions* • the section called “Version-specific provider packages”	February 22, 2024

Support for shared Amazon VPC	Amazon MWAA supports cross-account environment creation for organizations using Amazon OpenSearch Service to manage Amazon MWAA resources using a central shared Amazon VPC in an *owner* account. As part of this launch, Amazon MWAA lets you choose to create, and manage, your own Amazon VPC endpoints. • the section called "Managing your own Amazon VPC endpoints"	November 15, 2023
New Apache Airflow version	Amazon MWAA now supports Apache Airflow v2.7.2. This update includes informati on about updated provider packages, and details about using Apache Airflow v2.7.2 on Amazon MWAA. • *Versions* • the section called "Version-specific provider packages"	November 6, 2023

New Apache Airflow version	Amazon MWAA now supports Apache Airflow v2.6.3. This update includes informati on about updated provider packages, and details about using Apache Airflow v2.6.3 on Amazon MWAA, • *Versions*	August 9, 2023
Version deprecation informati on	Updated topic on version deprecation to include deprecation notices and timelines for Apache Airflow v2.0.2 and Apache Airflow v2.2.2. • the section called "Apache Airflow deprecated versions"	July 31, 2023
New topics and use cases	Amazon MWAA supports minor version upgrades. This updates includes the following new topic that describes how to upgrade the environment and make sure your workflow resources are compatible with the version of Apache Airflow you are upgrading to: • the section called "Changing the version"	June 5, 2023

Updated topic	Updated customer-managed IAM policies that grant a user full console and API access to Amazon MWAA. The update describes why you must provide permission for `iam:PassRole` to allow a user to pass roles to Amazon MWAA. Amazon MWAA uses these permissions to perform actions on a user's behalf. • the section called “Accessing an Amazon MWAA environment”	April 12, 2023
New guidance	Updated topic on configuring AWS Secrets Manager as a backend for Amazon MWAA to provide guidance on using lookup patterns. Using lookup patterns narrow the secrets that Apache Airflow searches for and reduce the number of API calls Amazon MWAA makes to Secrets Manager to retrieve connections and variables. This reduces the costs associated with using Secrets Manager as a backend. • Create the Secrets Manager backend as an Apache Airflow configuration option	April 12, 2023

New Apache Airflow version	Amazon MWAA now supports Apache Airflow v2.5.1. This update includes informati on about updated provider packages, and details about using Apache Airflow v2.5.1 on Amazon MWAA, • *Versions*	April 11, 2023
New topics and use cases	Added a new topic on using a startup script with an Amazon MWAA environment. This topic descibes configuring a startup script for an existing environment, using it to install Linux runtimes, and setting environment variables . • the section called “Using a startup script”	April 3, 2023
Updated section on private webserver access	Updated the following topic on private webserver access. The update clarifies that, in environments with private webserver access, you must use a Python wheel archive (`.whl`) to package, and install, dependencies. • Private webserver access mode	February 24, 2023

Added information about deprecated Apache Airflow versions	Updated the Versions topic with new information about how Amazon MWAA managed deprecating Apache Airflow versions. Removed a section about upgrading to newer version of Apache Airflow, and a section that described changes between Apach e Airflow v1 and Apache Airflow v2. For more informati on about migrating to a newversion of Apache Airflow, refer to the Amazon MWAA Migration Guide. • the section called “Apache Airflow deprecated versions” • the section called “End-of-s upport versions”	February 17, 2023

Fixes in Amazon MWAA container metrics	Updated the container metrics topic, and removed a set of erroneous metrics that did not exist in the `Cluster` dimension. Added an additional section that describes how you can evaluate the number of additional workers that an environment is utilizing at a given time by graphing the `CPUUtilization` or the `MemoryUtilization` metric for the `AdditionalWorker` component, and setting the statistics type to `Sample Count`. • the section called "Evaluating the number of additional worker and webserver containers"	January 20, 2023
New Apache Airflow version	Amazon MWAA now supports Apache Airflow v2.4.3. This update includes information about updated provider packages, details about using Apache Airflow v2.4.3 on Amazon MWAA, and consolidated information about which features are supported in each Apache Airflow version on Amazon MWAA. • *Versions*	January 5, 2023

Updated topic on service-linked role	Updated information about the service-linked role that Amazon MWAA uses to create and manage AWS resources on your behalf, including information about how you can delete the service-linked role when you no longer need it. This includes an updated service-linked role permission policy that allows Amazon MWAA to publishe additional CloudWatch metrics in the `AWS/MWAA` namespace. • the section called "Service-linked role"	November 18, 2022
New topic on service metrics	Added new topic that describes service metrics emitted by Amazon MWAA in the `AWS/MWAA` namespace. These include Amazon ECS cluster metrics schedulers, workers, and webservers, Amazon SQS metrics for the queues that allow Amazon MWAA to decouple schedulers and workers, as well as Amazon RDS metrics for the metadata database. • the section called "Container, queue, and database metrics"	November 18, 2022

New topic	Added new guidance on modifying a constraints file to specify new versions of provider packages to use with your Amazon MWAA environment. • the section called “Constraints file”	November 18, 2022
Updated FAQ entry	Updated information related to Amazon MWAA's HIPAA eligibility. • the section called “HIPAA compliance”	November 15, 2022
New topic	Added new topic on using `aws:SourceArn` and `aws:SourceAccount` global condition context keys in an Amazon MWAA execution role trust policy, to prevent cross-service confused deputy. • the section called “Cross-service confused deputy prevention”	October 21, 2022
New sample code	Added updated instructions and DAG code example that writes custom OS-level metrics to CloudWatch. • the section called “Using a DAG to write custom metrics”	September 13, 2022

New sample code	Added updated instructions and a new AWS Lambda Python code example that retrieves an Apache Airflow CLI token, then invokes a DAG in a specified Amazon MWAA environment. • the section called "Invoking DAGs with Lambda"	September 12, 2022
New architectural diagrams	Added new architectural diagrams that demonstrate an Amazon MWAA environment with a public and private webserver. • the section called "Apache Airflow access modes"	September 12, 2022
New sample code	Added updated instructions and a new DAG code example that retrieves an Apache Airflow CLI token, then invokes another DAG in a different Amazon MWAA environment. • the section called "Invoking DAGs in different environments"	August 16, 2022

New sample code	Added updated instructions and new DAG that queries an environment's Aurora PostgreSQL for metadata information, writes the result To CSV files and stores the files in Amazon S3. • the section called "Exporting environment metadata to Amazon S3"	August 12, 2022
New sample code	Added updated instructions and new DAG that refreshes an AWS CodeArtifact token at runtime and stores the result in Amazon S3. • the section called "Refreshing an AWS CodeArtifact token at runtime"	August 3, 2022
New sample code	Added updated instructions and DAG code sample for using the `ECSOperator` in Amazon MWAA. • the section called "Using the `ECSOperator`"	July 26, 2022
New sample code	Added updated instructions and DAG code sample for using the `SSHOperator` in Amazon MWAA. • the section called "Using the `SSHOperator`"	July 15, 2022

New sample code	Added new instructions and DAG code sample for using dbt Postgres with Amazon MWAA. • the section called “Using dbt with Amazon MWAA”	June 17, 2022
New topics and use cases	Added new instructions and DAG code sample for installing dependencies using Python wheel files for Amazon MWAA environments with public and private access. • Managing dependencies using Python wheels	May 13, 2022
New topics and use cases	Added new guidance on choosing which Apache Airflow metrics Amazon MWAA sends to CloudWatch. • Choosing which Apache Airflow metrics are reported	April 19, 2022
New guides	Amazon MWAA offers a migration guide for migrating Apache Airflow workflows from self-managed deployments, as well as existing Amazon MWAA environments. • Amazon MWAA Migration Guide	March 7, 2022

New topics and use cases	Added new security best practice for working with Apache Airflow, including a solution for detecting changes to the Apache Airflow user privileges. • the section called “Security best practices in Apache Airflow”	February 18, 2022
New sample code	Added new code sample for creating timezone-aware DAGs using Pendulum, and clarified how to use a custom plugin to change the timezone in which Apache Airflow logs are created. • the section called “Changing a DAG's timezone”	February 11, 2022

Apache Airflow v2.2.2 launch	Amazon Managed Workflows for Apache Airflow now supports Apache Airflow v2.2.2. Beginning with v2.2, Amazon MWAA will install Python packages and custom plugins directly on the Apache Airflow web server allowing you greater flexibility to manage your environments. For more information, refer to the following. • Apache Airflow versions on Amazon Managed Workflows for Apache Airflow. • Apache Airflow v2.2.2 changelog on the Apache Airflow documentation website.	January 27, 2022
New tutorials	Added a new tutorial that demonstrates creating a new custom Apache Airflow role, and assigning the role to an Apache Airflow user mapped from IAM to limit the user's access to a subset of specified DAGs. • Tutorial: Restricting an Amazon MWAA user's access to a subset of DAGs	December 8, 2021

Fixes	Fixed a best practices recommendation for setting the value of `scheduler.min_file_process_interval` to optimize CPU usage. Added an IAM policy example granting access to Secrets Manager resources in the execution role. Added troubleshooting topic on using Secrets Manager condition keys. • Performance tuning how the scheduler parses DAGs • Provide Amazon MWAA with permission to access Secrets Manager secret keys • Configuring condition keys in the Amazon MWAA execution role for Secrets Manager	November 22, 2021

New sample code	Added the following new code sample for modifying the time zone in which DAGs are processed using a custom plugin, and new troublesh ooting topic for invoking the `dags backfill` Apache Airflow CLI command from within a bash operator. • the section called "Changing a DAG's timezone" • Backfill CLI command using a bash operator	November 1, 2021
Fixes	Fixed issues in the Amazon ECS operator code sample, and clarified the additional permissions required in the Amazon MWAA execution role to allow the environment to access Amazon ECS task log group in CloudWatch Logs. • Amazon ECS operator permissions.	October 26, 2021

New sample code	Added new code sample that queries the Aurora PostgreSQL database for information relevant to DAG runs and writes the results to CSV file stored on Amazon S3. • the section called "Exporting environment metadata to Amazon S3".	October 1, 2021
Fixes	Corrected information about how Amazon MWAA automatically syncs new and changed objects from your target Amazon S3 bucket to your schedulers and workers. • How the DAG folder works.	October 1, 2021
Now supported	Amazon MWAA now supports additional provider packages for Apache Airflow 2.0+. To learn more about supported packages, refer to the following:	September 24, 2021
New commands and procedures	Added additional guidance and AWS CLI command examples for creating an Amazon S3 gateway endpoint when using an Amazon VPC without internet access: • Creating an Amazon VPC network without internet access.	September 24, 2021

New topics and use cases	Added the following changes: • Added a new code sample that uses an Amazon Elastic Container Service operator in the section called “Using the `ECSOperator`”. • Added new troublesh ooting topics for issues in configuring Apache Airflow plugins in the section called “Plugins”.	September 19, 2021

New supported region	Amazon MWAA is now available in the following regions: • Asia Pacific (Mumbai) - ap-south-1 • Asia Pacific (Seoul) - ap-northeast-2 • Europe (London) - eu-west-2 • Europe (Paris) - eu-west-3 • Canada (Central) - ca-central-1 • South America (São Paulo) - sa-east-1 For more information about region availability and service endpoints, refer to the following: • Amazon MWAA endpoints and quotas in the *AWS General Reference*.	August 31, 2021
New topics and use cases	Added the following changes: • Updated the sample policies to allow Amazon MWAA to fetch account-level Amazon S3 settings (`s3:GetAccountPublicAccessBlock`) in Amazon MWAA execution role.	August 27, 2021

Fixes	Added the following changes: • Fixed the CloudFormation template to use a self-refe rencing inbound rule for the security group in Create the VPC network. • Fixed the CloudFormation template to use a self-refe rencing inbound rule for the security group in Quick start tutorial for Amazon Managed Workflows for Apache Airflow.	August 27, 2021
New topics and use cases	Added the following changes: • Added DAG decorator to the list of what's supported for Apache Airflow v2.0.2 Apache Airflow versions on Amazon Managed Workflows for Apache Airflow.	August 20, 2021

New topics and use cases	Added the following changes: • Added `celery.sync_parallelism` use case to Performance tuning for Apache Airflow on Amazon MWAA. • Added service endpoints to quotas page and changed name to Amazon Managed Workflows for Apache Airflow service endpoints and quotas. • Clarified networking prerequisites based on user feedback at Get started with Amazon Managed Workflows for Apache Airflow. • Moved `dags list-runs` and `dags next-execution` to unsupported Airflow CLI commands in Apache Airflow CLI command reference.	August 13, 2021

New sample code	Added the following changes: • Added bash example to set, get or delete an Apache Airflow v2.0.2 variable in Apache Airflow CLI command reference. • Added Apache Airflow v2.0.2 dependencies and Airflow connection example to Using Amazon MWAA with Amazon RDS for Microsoft SQL Server.	August 13, 2021
Fixes	Added the following changes: • Fixed the Python code sample based on user feedback at Creating an SSH connection using the `SSHOperator` .	August 13, 2021

New topics and use cases	Added the following changes: • Moved `variables set` to supported Airflow CLI commands in Apache Airflow CLI command reference. • Added the summary of **What's changed** in v2.0.2 from the Airflow versions page to Installing Python dependencies based on user feedback. • Added the summary of **What's changed** in v2.0.2 from the Airflow versions page to Apache Airflow CLI command reference based on user feedback. • Added the summary of **What's changed** in v2.0.2 from the Airflow versions page to Overview of connection types based on user feedback. • Added the summary of **What's changed** in v2.0.2 from the Airflow versions page to Installing custom plugins based on user feedback. • Added the summary of **What's changed** in v2.0.2 from the Airflow versions page to Adding or	August 6, 2021

	updating DAGs based on user feedback.	
New sample code	Added the following changes: • Added Apache Airflow v2.0.2 sample code to Using a DAG to import variables in the CLI. • Added Apache Airflow v2.0.2 sample code to Invoking DAGs with a Lambda function.	August 6, 2021
New topics and use cases	Added the following changes: • Added troubleshooting topic for 'I can't find my connection in the Airflow UI' at Troubleshooting Amazon Managed Workflows for Apache Airflow. • Added a list of Amazon VPCs Amazon MWAA supports to About networking on Amazon MWAA.	July 29, 2021

Fixes	Added the following changes: • Fixed the Python code sample based on user feedback to print the web login token at Create a Apache Airflow webserver access token. • Fixed the Snowflake connection topic based on user feedback to use a single quote for the warehouse parameter at Troubleshooting Amazon Managed Workflows for Apache Airflow.	July 29, 2021
Removed or moved topics	Added the following changes: • Restructed the existing page to include all monitoring and metrics documentation pages in Monitoring and metrics for Amazon Managed Workflows for Apache Airflow. • Moved Apache Airflow environment metrics in CloudWatch to the monitoring and metrics navigation menu.	July 23, 2021

New guides	Added the following changes: • Created Apache Airflow provider packages installed on Amazon MWAA environments. • Created Monitoring overview on Amazon MWAA. • Created Accessing audit logs in AWS CloudTrail. • Created Accessing Airflow logs in Amazon CloudWatch.	July 23, 2021
Fixes	Added the following changes: • Fixed the Python code sample based on user feedback to generate an Airflow connection string in the correct sequence and added the port parameter in Configuring an Apache Airflow connection using a AWS Secrets Manager secret. • Added a step to install an unzip package locally based on user feedback in Creating a custom plugin with Oracle.	July 23, 2021

New topics and use cases	Added the following changes: • Added topic for AWS DMS Operators at Amazon MWAA frequently asked questions. • Added troubleshooting topic for a remote logs error to Troubleshooting Amazon Managed Workflows for Apache Airflow. • Moved `variables set` to unsupported Airflow CLI commands in Apache Airflow CLI command reference.	July 16, 2021

New topics and use cases	Added the following changes: • Added sequential steps to create a requirements.txt file based on user feedback at Installing Python dependencies. • Added sequential steps to create a plugins.zip file based on user feedback at Installing custom plugins. • Added cross-reference links throughout the user guide to the API reference guide at Amazon Managed Workflows for Apache Airflow API Reference guide. • Added topic for why plugins aren't listed in the Airflow 2.0 Admin > Plugins menu at Amazon MWAA frequently asked questions.	July 9, 2021
New guides	Added the following changes: • Created Deleting files on Amazon S3.	July 9, 2021

New topics and use cases	Added the following changes: • Added a list of supported values at Using customer-managed keys for encryption. • Updated and clarified the example for a private repo URL based on user feedback in Managing Python dependencies in requirements.txt.	July 2, 2021
New sample code	Added the following changes: • Added Apache Airflow v1.10.12 sample code to use a private key in AWS Secrets Manager for an SSH connection at Creating an SSH connection using the `SSHOperator` .	July 2, 2021
New topics and use cases	Added the following changes: • Added StartedTaskInstances and FinishedTaskInstances metrics to Apache Airflow environment metrics in CloudWatch.	June 25, 2021
New sample code	Added the following changes: • Added Apache Airflow v2.0.2 sample code at Using Amazon MWAA with Amazon EKS.	June 25, 2021

New guides	Added the following changes: • Created Performance tuning for Apache Airflow on Amazon MWAA.	June 25, 2021

New topics and use cases	Added the following changes: • Added `connections add` and `connections delete` to the **supported** Apache Airflow v2.0.2 CLI commands at Apache Airflow CLI command reference. • Added that the latest version available in CloudFormation is Apache Airflow v2.0.2 at Quick start tutorial for Amazon Managed Workflows for Apache Airflow. • Added question for storing temporary data on Apache Airflow workers to Amazon MWAA frequently asked questions. • Added topic for the 'Executor reports task instance %s finished' error to Troubleshooting Amazon Managed Workflows for Apache Airflow. • Added topic for the 'server closed the connectio n unexpectedly' log to Troubleshooting Amazon Managed Workflows for Apache Airflow. • Added example to run CLI commands on an SSH	June 18, 2021

tunnel to a bastion host to Creating an Apache Airflow CLI token.

- Added topic for randomly-generated user names to Troubleshooting Amazon Managed Workflows for Apache Airflow.
- Added topic for a 503 error when running a DAG in the CLI to Troubleshooting Amazon Managed Workflows for Apache Airflow.
- Added topic for custom plugins in Apache Airflow v2.0.2 which need an Airflow configuration option of `core.lazy_load_plugins : False` to load plugins at the start of each Airflow process to override the version's default setting to Using Apache Airflow configuration options on Amazon MWAA.
- Added Airflow configuration options step for Apache Airflow v2.0.2 plugins sample code at Creating a custom plugin with Apache Hive and Hadoop.

	• Added Airflow configuration options step for Apache Airflow v2.0.2 plugins sample code. • Added Airflow configuration options step for Apache Airflow v2.0.2 `PythonVirtualenvOperator` plugins sample code. • Added Airflow configuration options step for Apache Airflow v2.0.2 plugins sample code at Creating a custom plugin with Oracle.	
New sample code	Added the following changes: • Added sample code for an Apache Airflow Snowflake connection at Using a secret key in AWS Secrets Manager for an Apache Airflow Snowflake connection.	June 18, 2021

New topics and use cases	Added the following changes: • Added server-side encryption guidance to Create an Amazon S3 bucket for Amazon MWAA. • Added the secrets backend for Apache Airflow v2.0.2 to Configuring an Apache Airflow connection using a AWS Secrets Manager secret. • Added question for Apache Airflow workers quota increase requests to Amazon MWAA frequently asked questions. • Added question for which metrics are used to determine whether to scale Apache Airflow workers to Amazon MWAA frequently asked questions. • Added question for creating custom metrics in CloudWatch to Amazon MWAA frequently asked questions. • Added steps to enable private IP addresses for an Amazon S3 VPC interface endpoint for a VPC with private routing in Creating the required VPC service	June 2, 2021

	endpoints in an Amazon VPC with private routing. • Added an option to setup an SSH Tunnel using local port forwarding in Tutorial: Configuring private network access using a Linux Bastion Host.	
New sample code	Added the following changes: • Added sample code for a DAG that queries the Amazon Aurora PostgreSQL metadata database and publishes custom metrics to Amazon CloudWatch at Using a DAG to write custom metrics in CloudWatch.	June 2, 2021
New guides	Added the following changes: • Created a guide on how to use connection templates interchangeably in the Apache Airflow UI in Overview of connection types.	June 2, 2021

Fixes	Added the following changes: • Added Apache Airflow VPC endpoints to the CloudFormation template in *Option three: Creating a VPC network without internet access* to Create the VPC network.	June 2, 2021

Apache Airflow v2.0.2 launch	General availability launch of Apache Airflow v2.0.2. • Created Apache Airflow versions on Amazon Managed Workflows for Apache Airflow. • Created Apache Airflow environment metrics in CloudWatch. • Added version-specific links for Apache Airflow v2.0.2 to Using Apache Airflow configuration options on Amazon MWAA. • Added Apache Airflow v2.0.2 version-specific guidance to Installing Python dependencies. • Added Apache Airflow v2.0.2 version-specific guidance to Managing Python dependencies in requirements.txt. • Added Apache Airflow v2.0.2 sample plugins to Installing custom plugins. • Added Apache Airflow v2.0.2 sample code to Aurora PostgreSQL database cleanup on an Amazon MWAA environment. • Added Apache Airflow v2.0.2 sample code to Using	May 26, 2021

	a secret key in AWS Secrets Manager for an Apache Airflow connection. • Added Apache Airflow v2.0.2 sample code for `PythonVirtualenvOperator` custom plugin. • Added Apache Airflow v2.0.2 commands to Apache Airflow CLI command reference. • Added Apache Airflow v2.0.2 scripts to Creating an Apache Airflow CLI token. • Added a note that Amazon MWAA uses the latest Apache Airflow version by default to Create an Amazon MWAA environment.	
New topics and use cases	Added the following changes: • Added guidance to troubleshooting Airflow tasks that are stuck or not running to Troubleshooting Amazon Managed Workflows for Apache Airflow.	May 14, 2021

Fixes	Added the following changes: • We've updated the sample plugins code to use the latest Java version in Creating a custom plugin with Apache Hive and Hadoop. Previously, it was `os.environ["JAVA_HOME"]="/usr/lib/jvm/jre-1.8.0-openjdk-1.8.0.272.b10-1.amzn2.0.1.x86_64"` .	May 12, 2021
Removed or moved topics	Added the following changes: • Moved topics in Troubleshooting Amazon Managed Workflows for Apache Airflow to new pages by category.	May 10, 2021
New topics and use cases	Added the following changes: • Added Amazon S3 bucket overview to Working with DAGs on Amazon MWAA.	May 10, 2021

Removed or moved topics	Added the following changes: • Moved Accessing Apache Airflow to the main-level navigation, and added pages for Create a Apache Airflow webserver access token, Creating an Apache Airflow CLI token, and Apache Airflow CLI command reference.	May 7, 2021
New topics and use cases	Added the following changes: • Added version-specific links to the *Apache Airflow reference guide* for all supported and unsupported Airflow CLI commands in Apache Airflow CLI command reference. • Added version-specific links to the *Apache Airflow reference guide* for all configuration options in Using Apache Airflow configuration options on Amazon MWAA. • Added the Amazon MWAA CLI utility to Managing Python dependencies in requirements.txt.	May 7, 2021

New topics and use cases	Added the following changes: • Added flat and nested examples for how to structure a plugins.zip in Installing custom plugins. • Added the Amazon MWAA CLI utility to the Adding or updating DAGs, Installing custom plugins, and Installing Python dependencies pages. • Restructured content into an overview, upload to Amazon S3, and installing on Amazon MWAA sections based on user feedback in Installing custom plugins, and Installing Python dependencies pages. • Added an example use case to create and attach required VPC endpoints to an existing Amazon VPC without internet access in About networking on Amazon MWAA.	April 30, 2021

New sample code	Added the following changes: • Added sample code that uses a secret key in Secrets Manager for an Apache Airflow variable in Using a secret key in AWS Secrets Manager for an Apache Airflow variable.	April 30, 2021
New guides	Added the following changes: • Created Creating the required VPC service endpoints in an Amazon VPC with private routing.	April 30, 2021
Fixes	Added the following changes: • We've updated `core.default_ui_timezone` to `webserver.default_ui_timezone` in Using Apache Airflow configuration options on Amazon MWAA.	April 30, 2021

New topics and use cases	Added the following changes: • Added Windows (PuTTY) steps for SSH tunnel to Tutorial: Configuring private network access using a Linux Bastion Host. • Added topic for `apache-airflow-providers-amazon` , which is only compatible with Apache Airflow 2.0 to Troublesh ooting Amazon Managed Workflows for Apache Airflow.	April 23, 2021
New sample code	Added the following changes: • Added sample code that uses a secret key in Secrets Manager for an Apache Airflow connection in Using a secret key in AWS Secrets Manager for an Apache Airflow connection.	April 23, 2021
New guides	Added the following changes: • Created About networking on Amazon MWAA. • Created Security in your VPC on Amazon MWAA. • Created Managing access to service-specific Amazon VPC endpoints on Amazon MWAA.	April 23, 2021

New topics and use cases	Added the following changes: • Added a new CloudFormation template to create an Amazon VPC network without internet access in Create the VPC network. • Added a new tutorial to create an AWS Client VPN in Tutorial: Configuring private network access using an AWS Client VPN. • Changed the name of the **Networking access** page to **Apache Airflow access modes** based on user feedback, and streamlined docs in Apache Airflow access modes. • Streamlined docs to include only Amazon VPC getting started information and templates based on user feedback in Create the VPC network. • Added BigQuery operator workaround to Troubleshooting Amazon Managed Workflows for Apache Airflow. • Added an Apache Airflow v1.10.12 constraints file best practice to Installing Python dependencies.	April 16, 2021

New sample code	Added the following changes: • Added sample code to create a custom plugin using Oracle in Creating a custom plugin with Oracle. • Added sample code to create a custom plugin that generates runtime environment variables. •	April 16, 2021

New topics and use cases	Added the following changes: • Added topic for the self-referencing rule requirement on a VPC security group to Amazon MWAA frequently asked questions. • Added custom plugins directory and size limits to Installing custom plugins. • Added requirements directory and size limits to Installing Python dependencies. • Clarified the Apache Airflow configuration options for `foo.user` and `foo.pass` in Managing Python dependencies in requirements.txt. • Added configuration options overview to Using Apache Airflow configuration options on Amazon MWAA.	April 9, 2021

New sample code	Added the following changes: • Added sample code to create a custom plugin using `PythonVirtualenvOperator` . • Added sample code to create a custom plugin with Apache Hive and Hadoop in Creating a custom plugin with Apache Hive and Hadoop.	April 9, 2021
Fixes	Added the following changes: • We've updated the format for a requirements.txt, and added an example that's compatible with Apache Airflow v1.10.12 in Installing Python dependencies.	March 31, 2021

New topics and use cases	Added the following changes: • Added workaround to removing a requirements.txt or plugins.zip to Amazon MWAA frequently asked questions. • Added a bash workaround for SSH on an environment to Amazon MWAA frequently asked questions. • Added topic for CloudTrail ResourceAlreadyExistsException error to Troubleshooting Amazon Managed Workflows for Apache Airflow.	March 26, 2021

New topics and use cases	Added the following changes: • Added list of AWS services used to Amazon MWAA execution role. • Added list of AWS services used to Service-linked role for Amazon MWAA. • Added question for Python 3.7 version for Amazon MWAA to Amazon MWAA frequently asked questions. • Added question for `PythonVirtualenvOperator` to Amazon MWAA frequently asked questions. • Added the troubleshooting script as next steps for all topics related To VPC and environment configuration at Troubleshooting Amazon Managed Workflows for Apache Airflow. • Clarified the docs that a linux bastion must be in the same Region as an environment at Tutorial: Configuring private network access using a Linux Bastion Host.	March 19, 2021

New guides	Added the following changes: • Created Apache Airflow connections guide for AWS Secrets Manager at Configuring an Apache Airflow connection using a AWS Secrets Manager secret. • Created quick start tutorial using a CloudFormation template to create the Amazon VPC infrastructure, Amazon S3 bucket, and Amazon MWAA environme nt at Quick start tutorial for Amazon Managed Workflows for Apache Airflow.	March 19, 2021
New topics and use cases	Added the following changes: • Added the create Amazon S3 bucket troublesh ooting topic Troublesh ooting Amazon Managed Workflows for Apache Airflow. • Added steps to create and attach a JSON policy to Amazon MWAA execution role.	March 12, 2021

New sample code	Added the following changes: • Added sample code to add a configuration when triggering a DAG to Accessing Apache Airflow.	March 12, 2021
New guides	Added the following changes: • Created best practices guide at Managing Python dependencies in requirements.txt.	March 12, 2021
New topics and use cases	Added the following changes: • Added Google/GCP/BigQuery troubleshooting topic to Troubleshooting Amazon Managed Workflows for Apache Airflow. • Added Cython troubleshooting topic to Troubleshooting Amazon Managed Workflows for Apache Airflow. • Added MySQL troubleshooting topic to Troubleshooting Amazon Managed Workflows for Apache Airflow. • Added 5xx webserver error troubleshooting topic to Troubleshooting Amazon Managed Workflows for Apache Airflow.	March 5, 2021

Now supported	Added the following changes: • Previously, `backend_kwargs` was not supported for AWS Secrets Manager and you needed a workaround to override the Secrets Manager function call. Now, `backend_kwargs` is supported. Refer to the AWS Secrets Manager troubleshooting topic in Troubleshooting Amazon Managed Workflows for Apache Airflow.	March 4, 2021
Fixes	Added the following changes: • We've updated the size of each environment class to reflect the actual GB in Configuring the Amazon MWAA environment class.	March 4, 2021

New topics and use cases	Added the following changes: • Added private network access using a VPC endpoint policy to Apache Airflow access modes. • Added additional checks for the creating an environme nt troubleshooting topic to Troubleshooting Amazon Managed Workflows for Apache Airflow. • Added steps to access logs for `requirements.txt` to Installing Python dependencies.	February 26, 2021
New topics and use cases	Added the following changes: • Added Apache Hive use case to Installing Python dependencies. • Clarified the docs that the required dependencies for an Apache Airflow package needs to be included in the `requirements.txt` file at Installing Python dependencies. • Added *Updating requireme nts.txt* troubleshooting topic to Troublesh ooting Amazon Managed Workflows for Apache Airflow.	February 25, 2021

New tutorials	Added the following changes: • Added private network tutorial to Tutorial: Configuring private network access using a Linux Bastion Host.	February 22, 2021
New topics and use cases	Added the following changes: • Added private and public network configurations to Apache Airflow access modes. • Added development group use case and user scenarios to Amazon MWAA execution role.	February 22, 2021
New sample code	Added the following changes: • Added sample Python scripts for web login token and CLI token to Accessing Apache Airflow. • Added sample code to trigger DAG in another environment to Code examples for Amazon Managed Workflows for Apache Airflow. • Added sample code to trigger DAG using a Lambda function to Invoking DAGs with a Lambda function.	February 22, 2021

New commands and procedures	Added the following changes: • Added step by step procedures to all scripts at Accessing Apache Airflow.	February 22, 2021
New sample code	Added the following changes: • Updated curl example for web login token at Accessing Apache Airflow. • Added sample code to connect to an Amazon RDS Microsoft SQL Server to Using Amazon MWAA with Amazon RDS for Microsoft SQL Server.	February 17, 2021

New commands and procedures	Added the following changes: • Added AWS CLI commands to Working with DAGs on Amazon MWAA pages. • Apache Airflow doesn't support serialized DAGs in CLI commands. Since the CLI runs on the webserver , which doesn't have plugins or requireme nts for security reasons, any MWAA environme nts with a plugins.z ip or requirements.txt will not support these commands. Moved Apache Airflow `list_dags` and `backfill` commands to unsupported commands at Accessing Apache Airflow.	February 17, 2021
GitHub launch	User guide docs are now open source on GitHub. Choose **Edit this page on GitHub** on any page.	February 17, 2021

New topics and use cases	Added the following changes: • Added question for Step Functions v. Amazon MWAA use case to Troubleshooting Amazon Managed Workflows for Apache Airflow. • Added CLI access policy to Accessing an Amazon MWAA environment. • Clarified the docs that any supported Apache Airflow configuration option can be specified at Using Apache Airflow configuration options on Amazon MWAA. • Clarified the docs that if a Fargate container in one availability zone fails, MWAA switches to the other container in a different availability zone at Create the VPC network.	February 12, 2021
New topics and use cases	Added the following changes: • Added Configuring the Amazon MWAA environment class.	February 5, 2021

Removed or moved topics	Added the following changes: • Removed requirement for Amazon S3 bucket name to start with `airflow-` at Get started with Amazon Managed Workflows for Apache Airflow. • Moved Accessing an Amazon MWAA environment and Amazon MWAA execution role to Managing access to an Amazon MWAA environment.	February 4, 2021
Amazon MWAA CloudFormation	Update the parameters to create an environment at Amazon MWAA CloudFormation. • Remove SubnetList. • Remove TagList. • Add NetworkConfiguration. • Add TagMap. • Add create environment request examples.	February 4, 2021

New topics and use cases	Added the following changes: • Added example email configuration to Using Apache Airflow configuration options on Amazon MWAA. • Added PostgresHook troubleshooting topic to Troubleshooting Amazon Managed Workflows for Apache Airflow. • Added AWS Secrets Manager troublesh ooting topic to Troublesh ooting Amazon Managed Workflows for Apache Airflow. • Added high performan ce use case to Configuri ng Amazon MWAA worker automatic scaling.	January 29, 2021
Amazon MWAA launch	General availability launch of Amazon Managed Workflows for Apache Airflow. • User guide documentation • CloudFormation documenta tion	November 24, 2020