Airflow 2.0 on Amazon MWAA
Airflow Summit 2021
Sam Dengler and John Jackson
July 15th, 2021
Table of contents

• How Amazon MWAA does Apache Airflow
• Staying current 2.x and beyond
  • Upgrades and new Apache Airflow version support
• Staying Open and contributing
  • AWS Contributions to Date
  • Support Plan for the Open Source Community
How Amazon MWAA does Apache Airflow
Amazon Managed Workflows for Apache Airflow (MWAA)

- A managed service for Apache Airflow that makes it easy for data engineers and data scientists to execute data processing workflows on AWS
- Released November 24, 2020, added Airflow 2.0 support May 26, 2021

<table>
<thead>
<tr>
<th>Name</th>
<th>Status</th>
<th>Created date</th>
<th>Airflow version</th>
<th>Airflow UI</th>
</tr>
</thead>
<tbody>
<tr>
<td>MWAA-Demo-5</td>
<td>Available</td>
<td>May 25, 2021 17:26:01 (UTC-07:00)</td>
<td>2.0.2</td>
<td>Open Airflow UI</td>
</tr>
<tr>
<td>MWAA-Demo-4</td>
<td>Available</td>
<td>May 25, 2021 14:03:12 (UTC-07:00)</td>
<td>2.0.2</td>
<td>Open Airflow UI</td>
</tr>
<tr>
<td>MWAA-Demo-3</td>
<td>Available</td>
<td>May 24, 2021 18:03:09 (UTC-07:00)</td>
<td>2.0.2</td>
<td>Open Airflow UI</td>
</tr>
<tr>
<td>MWAA-Demo-2</td>
<td>Updating</td>
<td>Dec 14, 2020 08:28:22 (UTC-08:00)</td>
<td>1.10.12</td>
<td>Open Airflow UI</td>
</tr>
<tr>
<td>MWAA-Demo-1</td>
<td>Available</td>
<td>Dec 10, 2020 14:37:40 (UTC-08:00)</td>
<td>1.10.12</td>
<td>Open Airflow UI</td>
</tr>
</tbody>
</table>
Amazon MWAA is Open Source Apache Airflow: No Forks
Amazon MWAA Architecture

- **Customer VPC**
  - Airflow Schedulers
  - Airflow Worker(s)

- **Service VPC**
  - Database VPCE
  - Web Server VPCE
  - DB Proxy
  - Meta Database

- **Private Network**
  - Airflow Web Server

- **Public Network**

- **Support Services**
  - Amazon CloudWatch
  - Amazon Simple Storage Service (S3)
  - Amazon Simple Queue Service
  - Amazon Elastic Container Registry
  - AWS Key Management Service
How Amazon MWAA Helps

- Deployments and Operations
  - Easy to setup up and maintain
- Availability and Sizing
  - Multi AZ/HA with Airflow 2.0 on ECS Fargate
- Scaling
  - Auto scaling with Celery executor
- Security
  - IAM and VPC
Deployments and Operations

- Easy to set up multiple small environments
- Prod/Dev environments
- CloudWatch
  - Logging
  - Metrics/Alarms
  - Dashboards
- CloudFormation/Terraform
  - Configure User Access, Environment (Execution Role) access, and Airflow Configurations without accessing the UI
Scaling/Environment Sizes

- Using Airflow metrics
- \((\text{Tasks running} + \text{tasks queued}) / \text{tasks per container} = \text{number of containers required}\)

- Downscaling occurs when \((\text{Tasks running} + \text{tasks queued}) = 0\)

### Environment class

<table>
<thead>
<tr>
<th>Class</th>
<th>Maximum worker count</th>
<th>Minimum worker count</th>
</tr>
</thead>
<tbody>
<tr>
<td>mw1.small</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>Scheduler count</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>
Security with MWAA

- Auth (IAM/Federation)
- VPC/SG
- Secrets Manager
- Execution Role
CI/CD

- Versioned S3 Bucket deployment target
- Flexible integration with existing CI/CD pipelines
- DAGs automatically synchronized every 30-60s
- Custom Plugin and Python dependency changes require Environment update

Create an MWAA Environment  Copy your DAGs and Plugins to S3  Access the Airflow UI
Amazon MWAA Best Practices

Configuration and Migration
Best Practices – Creating the Environment

• VPC Network
  • Connectivity to AWS Services
  • Private Web Server Option

• IAM Execution Role
  • Access to AWS Services
  • AOK vs CMK

• Python Requirements
  • Compatible Versions
  • Debugging

• Configuration Overrides
  • Performance Tuning
  • Secrets Backend Support
Best Practices – Migration

• Moving tasks off the Worker that require
  • Local Docker
  • Custom runtimes
  • Large CPU/Memory
  • Sudo

• Assume Ephemeral Workers
  • Limited Bash Value
  • Local storage between tasks

• Web Server Restrictions
  • Not currently installing requirements.txt or plugins
Contributions and Staying Current
Staying Current

• Target is to provide minor versions on Amazon MWAA within 30 days of Apache Airflow community release
  • We’re not there yet, but we’re working on it
• Major releases will take a bit longer
• We’ll keep older versions (i.e. 1.10.x) available for as long as we can
  • Depends on security, stability, and customer demand
• Version status will be posted at https://docs.aws.amazon.com/mwaa/latest/userguide/airflow-versions.html
Open Source Contributions

- EKS Native operator
- Improved Apache Airflow AWS integrations
- Performance and security improvements
- Native Serverless executor
Resources

- Amazon MWAA docs [https://docs.aws.amazon.com/mwaa](https://docs.aws.amazon.com/mwaa)
- Amazon MWAA product page [https://aws.amazon.com/mwaa](https://aws.amazon.com/mwaa)
- #airflow-aws Slack Channel: [https://apache-airflow.slack.com](https://apache-airflow.slack.com)
- GitHub samples: [https://github.com/aws-samples/amazon-mwaa-examples](https://github.com/aws-samples/amazon-mwaa-examples)
Q&A

Sam Dengler @samdengler
John Jackson @JohnJacksonPM