aws

(What's New With)

Amazon Managed Workflows for Apache Airflow

John Jackson (he/him)

Principal Product Manager

Agenda

About Amazon MWAA

Architecture Review

Private vs Public Web Servers

Local Runner

Best Practices

Open Source

What's Next

About Amazon MWAA

What is Amazon MWAA?

Open Source

Same Apache Airflow, no forks

Secure

AWS Identity and Access Management (IAM)

Integrated

AWS Fargate, Amazon CloudWatch, S3

Easy to Deploy

Console, CloudFormation, API, CLI

Amazon MWAA > Environments

Airflow environments

Environments (3)   〔⟳〕 〔Edit〕 〔Delete〕 〔Actions ▾〕 〔**Create environment**〕

🔍 Find environments					‹ 1 › ⚙
Name ▽	**Status** ▽	**Created date** ▼	**Airflow version** ▽	**Airflow UI** ▽	
○ MWAA-Demo-ARN-3	⊘ Available	Jan 25, 2022 11:22:18 (U...	2.2.2	Open Airflow UI ↗	
○ MWAA-Demo-ARN-2	⊘ Available	Dec 21, 2021 09:02:39 (U...	1.10.12	Open Airflow UI ↗	
○ MWAA-Demo-ARN-1	⊘ Available	May 20, 2021 16:34:14 (U...	2.0.2	Open Airflow UI ↗	

Airflow
Summit 2022

Key Milestones

November 2020

Amazon Managed
Workflows for Apache
Airflow is generally
available

May 2021

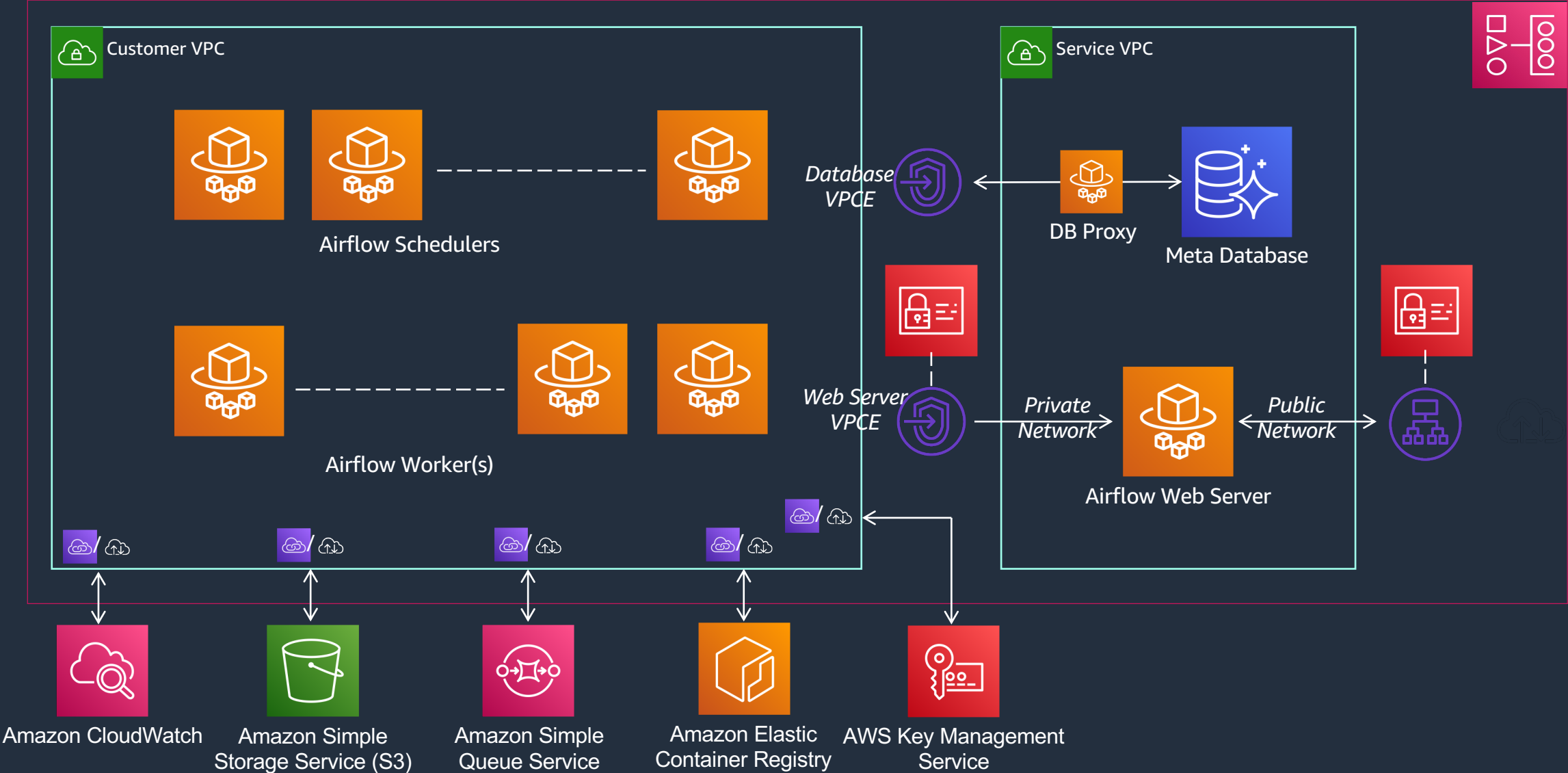Amazon MWAA
adds Airflow 2.0
support

August 2021

Amazon MWAA
expands to a total
of 15 commercial
regions

January 2022

Amazon MWAA
adds Airflow 2.2
support with ability
to install
requirements and
plugins on the web
server

Amazon MWAA Architecture

AMAZON MANAGED WORKFLOWS FOR APACHE AIRFLOW (MWAA)

Airflow
Summit 2022

Customer VPC

Airflow Schedulers

Airflow Worker(s)

Database VPCE

Web Server VPCE

Service VPC

DB Proxy

Meta Database

Private Network

Public Network

Airflow Web Server

Amazon CloudWatch

Amazon Simple Storage Service (S3)

Amazon Simple Queue Service

Amazon Elastic Container Registry

AWS Key Management Service

© 2022, Amazon Web Services, Inc. or its affiliates.

Web Server Options

Airflow ✖
Summit 2022

Web Server Options

Private Network

For customers that require their Airflow web server to be completely disconnected from the Internet

Public Network

Uses the same IAM authentication as AWS console pages to allow access to the Airflow web server

Configure advanced settings

Networking Info

Virtual private cloud (VPC)
Defines the networking infrastructure setup of your Airflow environment. An environment needs 2 private subnets in different availability zones. To create a new VPC with private subnets, choose Create MWAA VPC. **Learn more** ↗
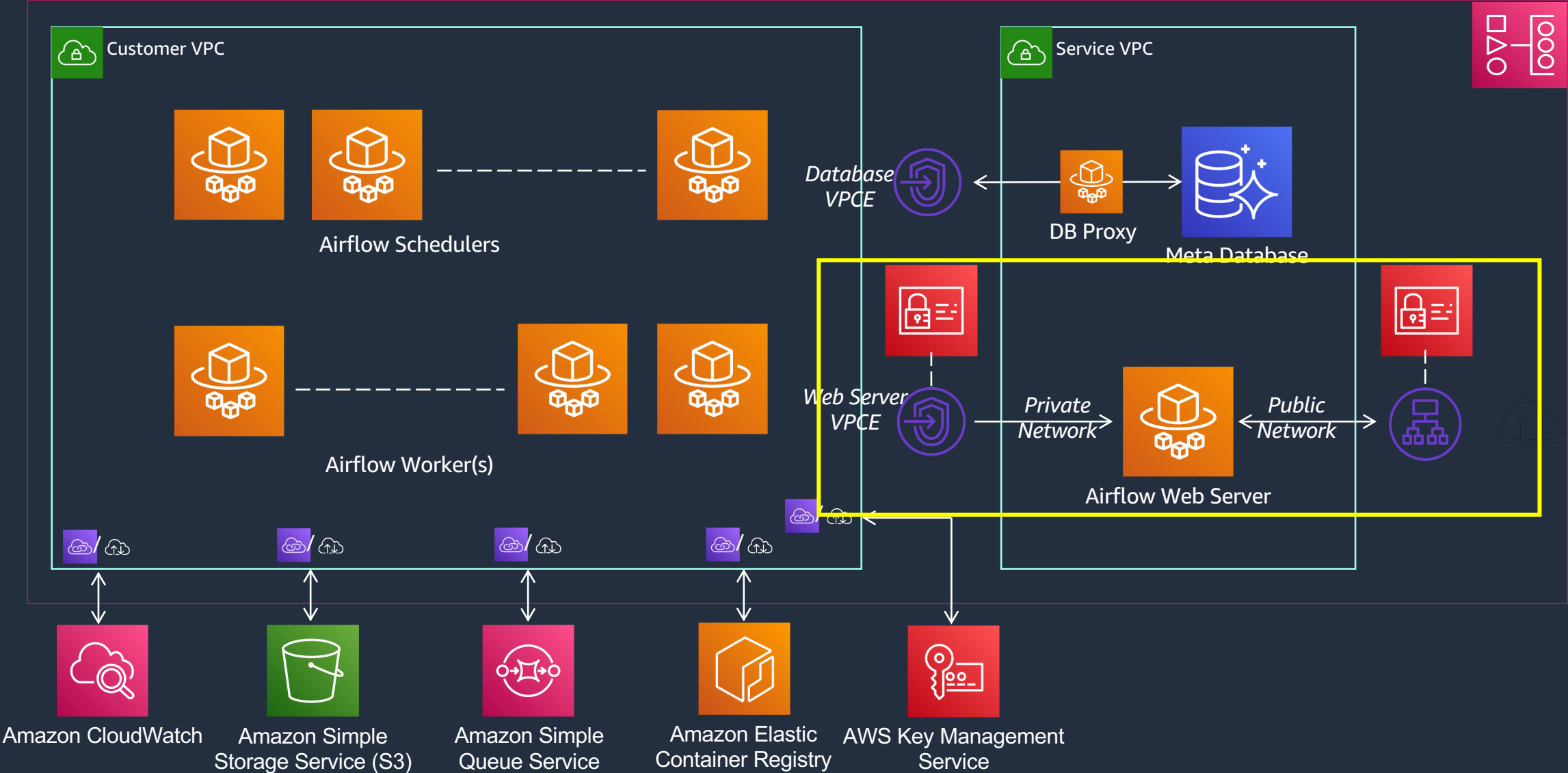
[Choose VPC ▼] [⟳] [**Create MWAA VPC** ↗]

ⓘ VPC and subnet selections can't be changed after an environment is created.

Web server access
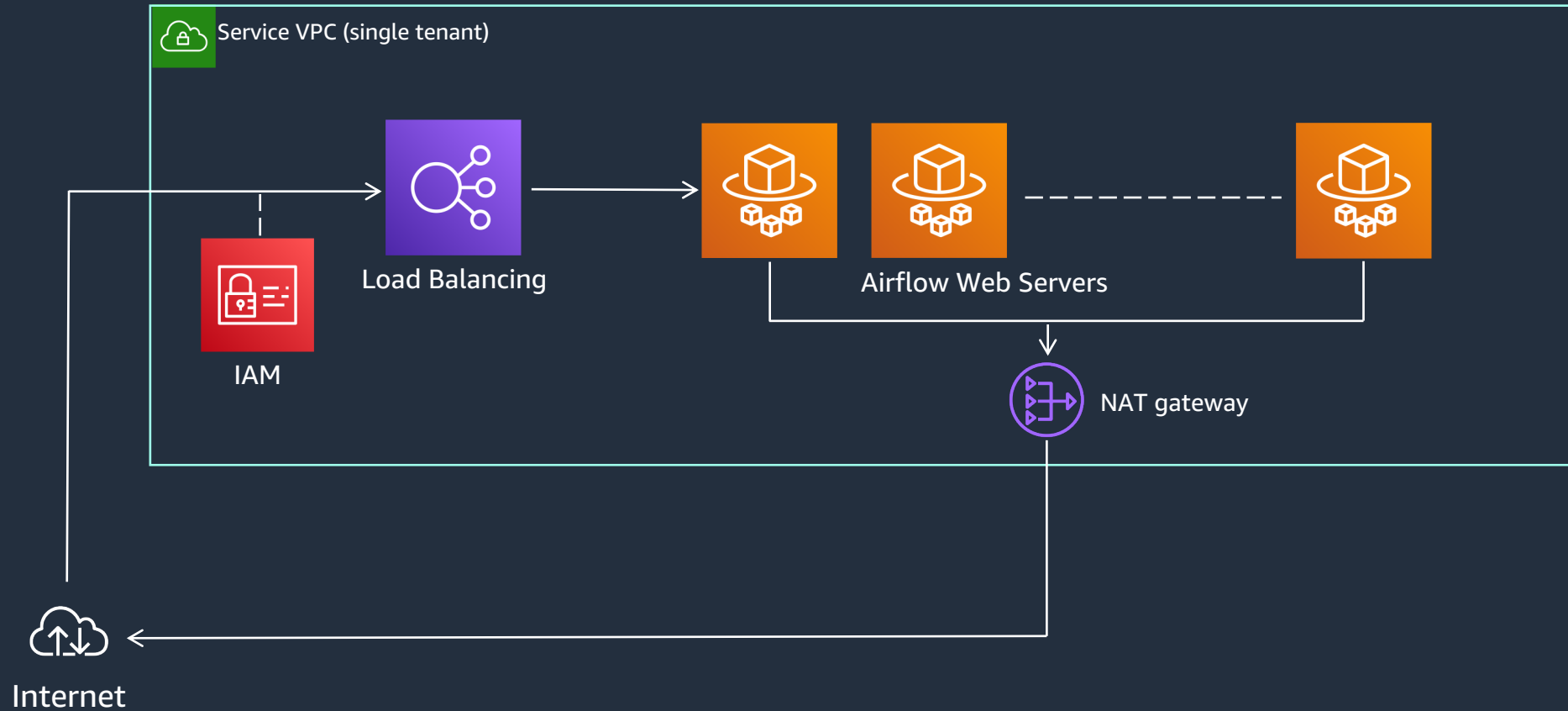
🔘 Private network (No internet access)
Additional setup required. Your Airflow UI can only be accessed by secure login behind your VPC. Choose this option if your Airflow UI is only accessed within a corporate network and you do not require a public repository for webserver requirements installation. IAM must be used to handle user authentication.

⚪ Public network (Internet accessible)
Your Airflow UI can be accessed by secure login over the Internet. Choose this option if your Airflow UI is accessed outside of a corporate network. IAM must be used to handle user authentication.
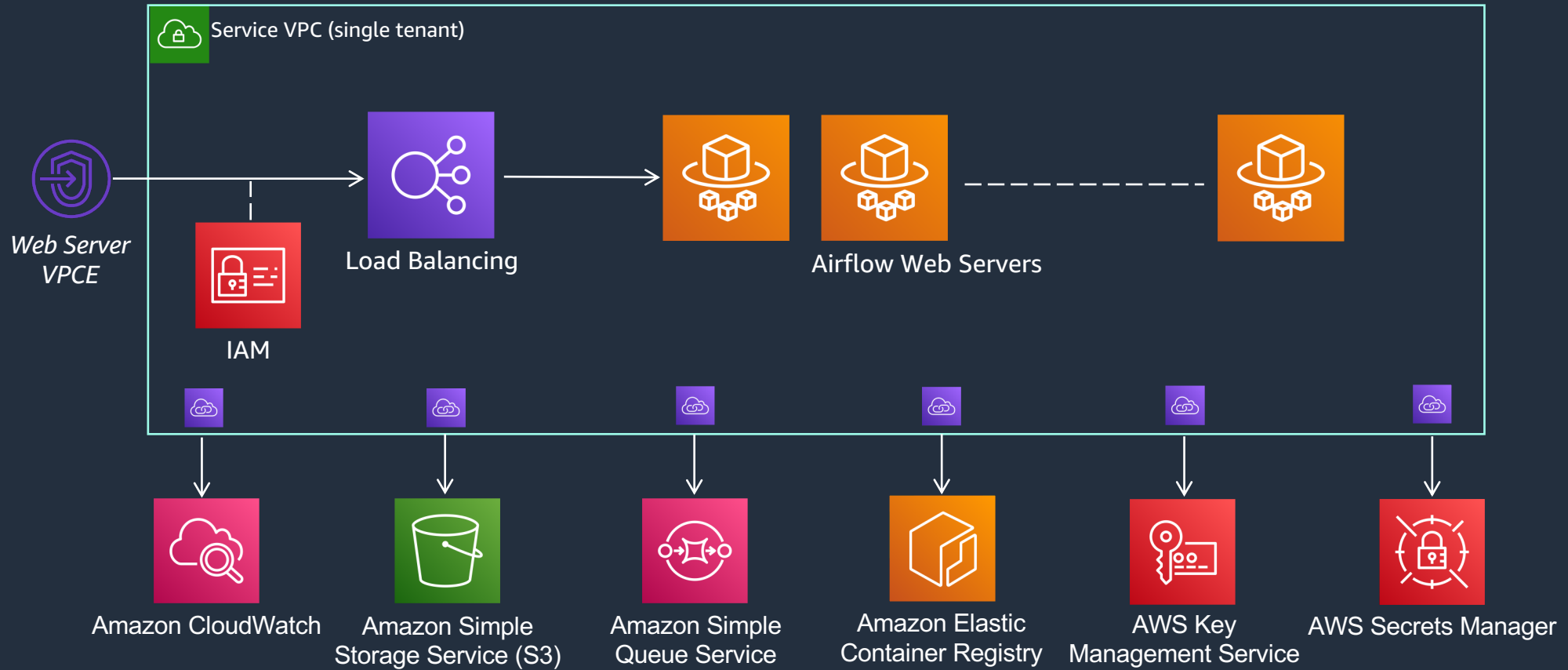
ⓘ For private network access, the Airflow web server is reached via a VPC endpoint inside your VPC. Connecting to the endpoint requires additional setup. **Learn more about VPC endpoints** ↗

AMAZON MANAGED WORKFLOWS FOR APACHE AIRFLOW (MWAA)

Airflow Summit 2022

Customer VPC

Service VPC

Airflow Schedulers

Database VPCE

DB Proxy

Meta Database

Airflow Worker(s)

Web Server VPCE

Private Network

Public Network

Airflow Web Server

Amazon CloudWatch

Amazon Simple Storage Service (S3)

Amazon Simple Queue Service

Amazon Elastic Container Registry

AWS Key Management Service

aws

© 2022, Amazon Web Services, Inc. or its affiliates.

MWAA Local Runner

What is MWAA Local Runner?

- Docker Container

- All supported MWAA versions

- Same libraries and options as MWAA

- Requirements management utilities

```
=====================================
   MWAA Local Runner CLI
=====================================

Syntax: mwaa-local-runner [command]

---commands---
help                        Print CLI help
build-image                 Build Image Locally
reset-db                    Reset local PostgresDB container.
start                       Start Airflow local environment. (LocalExecutor, Using postgres DB)
test-requirements           Install requirements on an ephemeral instance of the container.
package-requirements        Download requirements WHL files into plugins folder.
validate-prereqs            Validate pre-reqs installed (docker, docker-compose, python3, pip3)
```

https://github.com/aws/aws-mwaa-local-runner

> OPEN EDITORS

∨ WORKSPACE (WORKSPACE)

∨ Airflow   ●

> aws-mwaa-local-runner-1.10.12

> aws-mwaa-local-runner-2.0.2

∨ aws-mwaa-local-runner-2.2.2

∨ dags

> __pycache__

≡ requirements.txt

🐍 tutorial.py

> db-data

∨ docker

> config

> script

🐳 docker-compose-local-no-network...

🐳 docker-compose-local.yml

🐳 docker-compose-resetdb.yml

🐳 docker-compose-sequential.yml

🐳 Dockerfile

∨ plugins

⬦ .gitignore

≡ AWSCLIV2.pkg

📕 CODE_OF_CONDUCT.md

🎗 CONTRIBUTING.md

≡ download_requirements.txt

🔑 LICENSE

$ mwaa-local-env

ⓘ README.md

◔ VERSION

> cloudformation

> dags

> dev

> docker

> easy-rsa   ●

> OUTLINE

> TIMELINE

PROBLEMS   OUTPUT   **TERMINAL**   DEBUG CONSOLE

```
jacnjoh@147dda783c33 aws-mwaa-local-runner-2.2.2 %
```

> zsh - aws-mwaa-local-runner-2.2.2

Best Practices

Things to Watch Out For

Logging

MWAA updated to Watchtower 2.0.1 to improve performance.

Can still run into account-wide PutLogEvent throttle

Access

Connectivity to AWS Services (NAT/ PrivateLink/ Security Groups)

IAM Execution Role access to AWS Services, AOK vs CMK

Migration

Everything on the Worker, Custom runtimes, Large CPU/Memory

Ephemeral Workers / Limited Bash Value
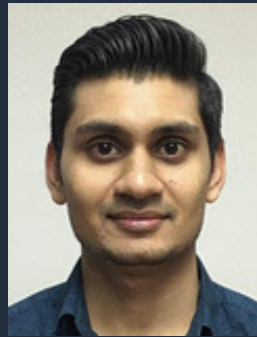
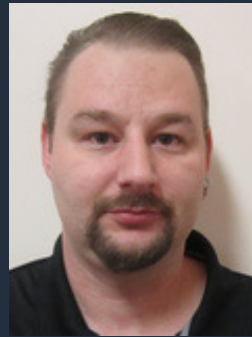## Running Airflow at Scale
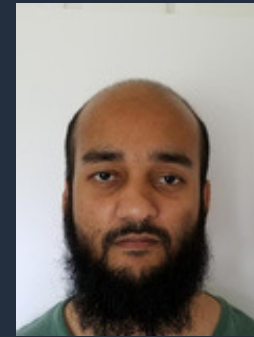
Shameless self-promotion:

https://airflowsummit.org/sessions/2022/managing-apache-airflow-at-scale/

Open Source

AWS Airflow Open Source Team

Niko Oliveira　　**Shubham Mehta**　　**Dennis Ferruzzi**　　**Syed Hussain**　　**Vincent Beck**

What's Next?

Airflow
Summit 2022

What the MWAA Team is Working On

Airflow REST API Access

AWS has a very high bar for security and scalability of APIs

Working with our MWAA service team and Open Source Team

New Airflow Versions Faster

Looking at ways to speed up security reviews and deployments

Improved Performance and Monitoring

Incrementally improving performance and stability, and adding tools that will monitor and when possible recover on your behalf

Resources

Stay "In the Know"…

MWAA docs **https://docs.aws.amazon.com/mwaa**

Amazon MWAA landing page **https://aws.amazon.com/mwaa**

Amazon MWAA resources page **https://aws.amazon.com/managed-workflows-for-apache-airflow/resources/**

#airflow-aws Slack Channel: **https://apache-airflow.slack.com**

Thank you!

John Jackson

jacnjoh@amazon.com

@JohnJacksonPM