



Amazon MWAA

Our journey and the road ahead

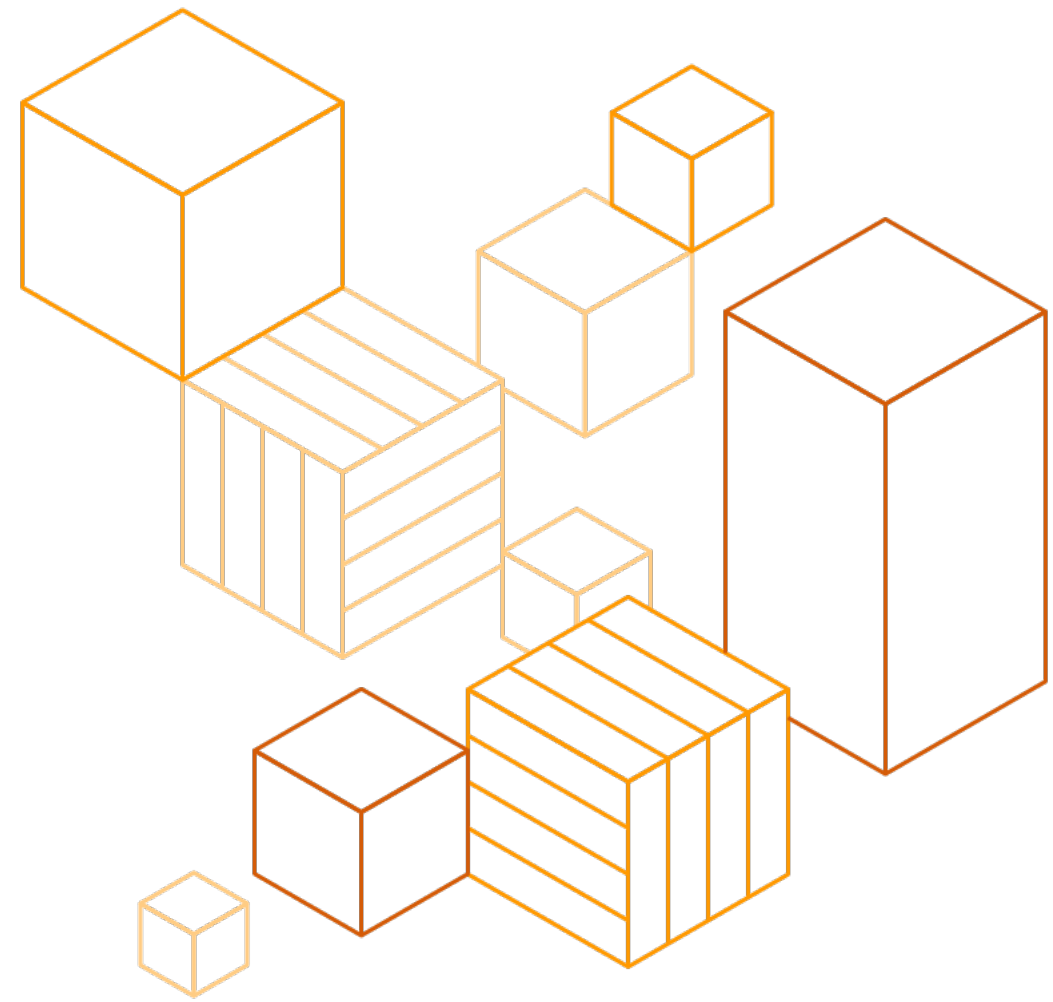
Subash Canapathy & John Jackson

July 16th 2021



Table of contents

- Amazon MWAA Intro
- Our Journey
 - Understanding Airflow
 - Design choices driven by customers
- What are we excited about Airflow's future
- Tenets and Open Source Involvement



Amazon MWAA

Like most things at AWS, it started with our customers

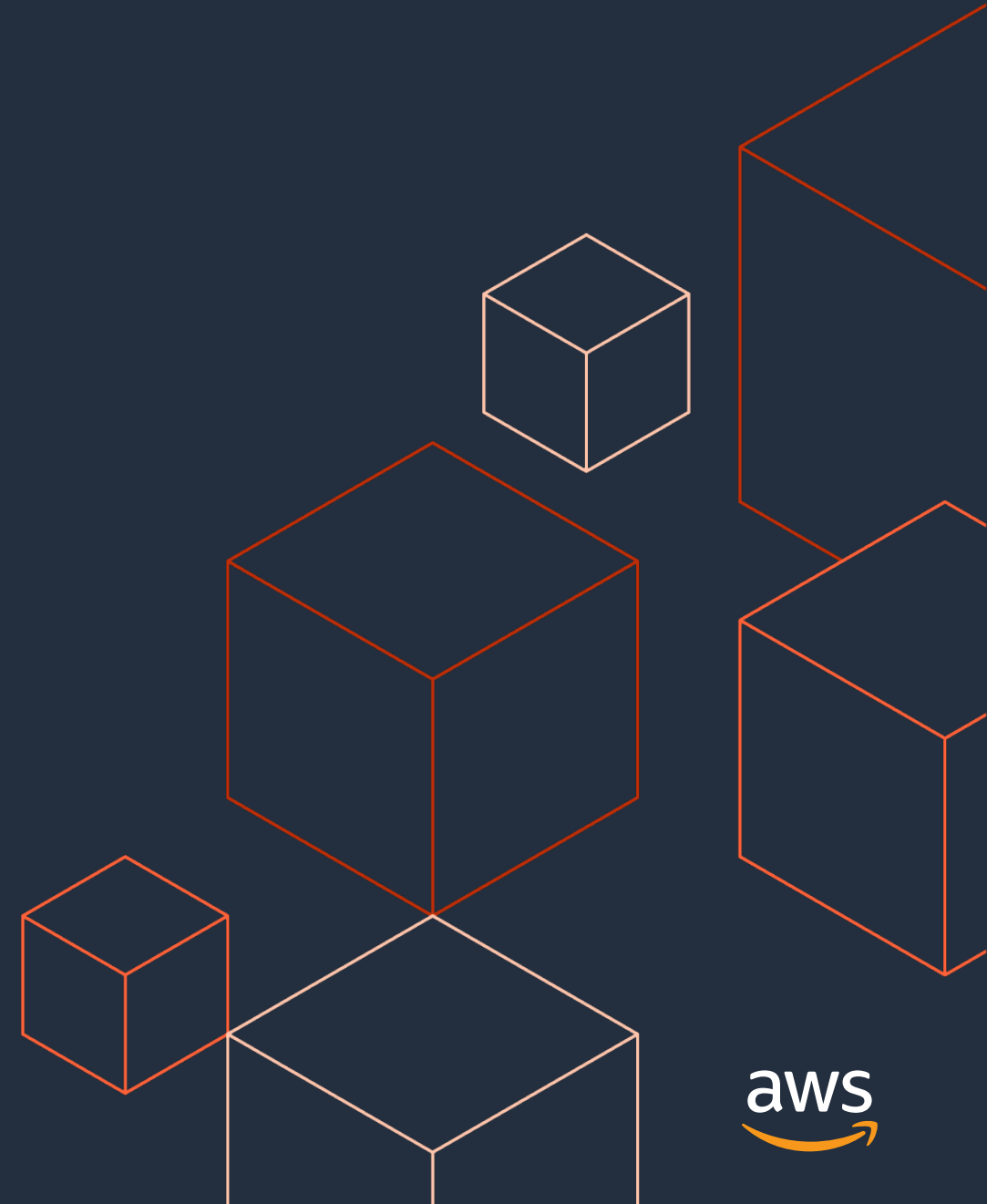
Customers asked if we could

- ... simplify the creation, management and operation of Airflow clusters
- ... increase data engineer efficiency by allowing them to innovate faster
- ... take care of scaling their workers on-demand for cost optimization

Can you do all this,
with the same community version of Airflow we use today?



Our Journey



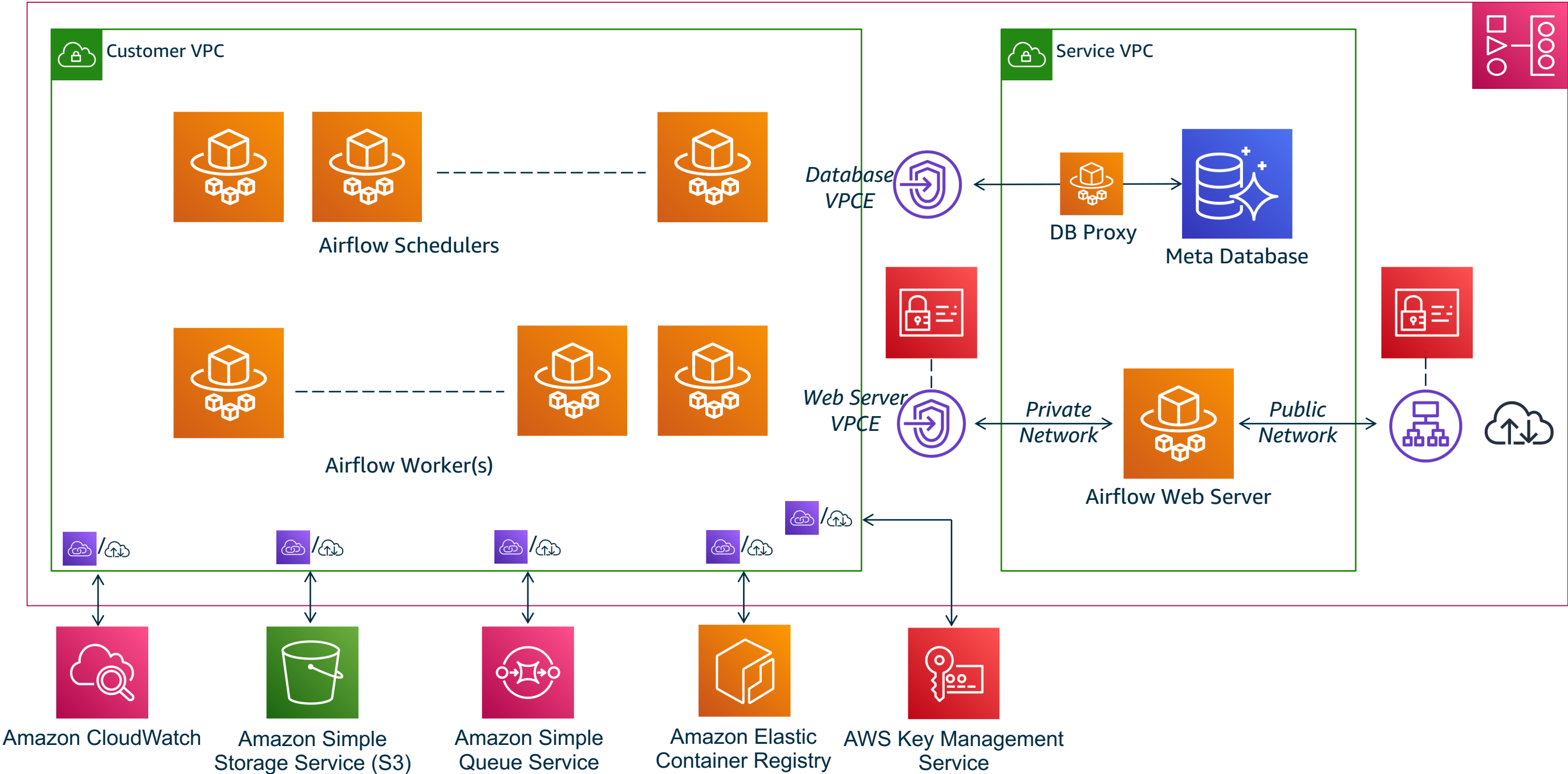
Understanding Airflow

- Learning the domain and how Airflow components interoperate
- Learn the why's behind the evolution of Airflow as a system



- Understanding and appreciating what customers love about Airflow
- Recognizing that every customer sets up Airflow to their unique needs

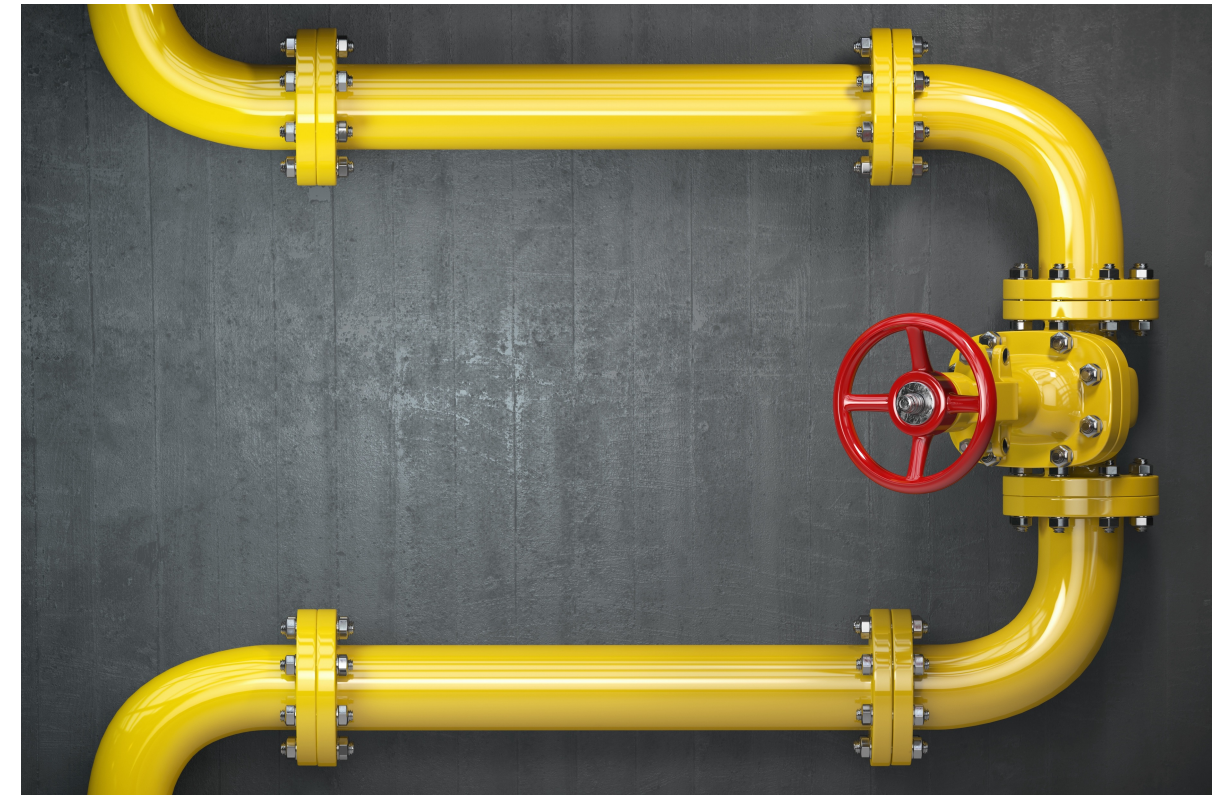
Amazon MWAA Architecture



Customer driven design choices

Deployment Safety

- Prod vs Pre-prod environments
- Effect of plugin and requirement updates on production
- When it comes to deployment, customers preferred safety over speed



Customer driven design choices

Simple and Seamless Auth

- IAM and Console familiarity
- Single Sign-On into Airflow UI
- RBAC Plugin and 1:1 pairing of Airflow user to IAM



Customer driven design choices

Auto Scaling - Celery

- Baseline – single worker
- Scale out based on `tasksQueued`, `tasksRunning` and `task-density/worker`
- Provisioned worker capacity – for customers that need it



Customer driven design choices

Storage

- One source of truth and consistent reads
- Flexible to setup CI/CD pipelines
- S3's durability, familiarity and easy accessibility across regions



Airflow 2.0 and beyond

What are we excited about

What are we excited about

Community's direction on

- Airflow UI modernization
- Manifest and remote DAG fetching
- Granular access and permission controls for multi-tenancy



Our contributions and involvement

- S3 Bucket Tagging Operator, EKS Operator and hook (In progress)
- Code quality and integration test improvements
- More than 4500 lines of code across Airflow, Moto and Watchtower

Amazon MWAA – Operating Tenet

Our customers are all current and future users of Airflow. We will help our customers to be successful on or outside of AWS by always maintaining compatibility and committing to build AWS integrations for Airflow.

Q&A

Talk to us on Apache Airflow Community Slack

@Subash Canapathy

@John Jackson

#airflow-aws