Orchestrating Data, ML & Generative AI Pipelines with Apache Airflow



Sneha Rao
Solutions Architect
Amazon Web Services (AWS)

Sushmita Barthakur
Sr. Data Solutions Architect
Amazon Web Services (AWS)

Suba Palanisamy
Sr. Technical Account Manager
Amazon Web Services (AWS)

Two ways to run Apache Airflow, fully managed on AWS

Apache Airflow is the open-source standard for orchestration.
Define pipelines as Python code (DAGs), with built-in dependencies, retries, and scheduling.



Amazon Managed Workflows for Apache Airflow (Amazon MWAA)

Amazon MWAA runs it for you in two flavors:

Amazon MWAA (Provisioned)

- Provisioned Airflow environment. Always-on, sized for your workload
- Easy Airflow deployment and AWS-managed security updates
- Multi-AZ deployment and Native AWS integration
- Auto scaling prevents performance bottlenecks
- Best for complex DAGs, Consistently high-volume workloads

MWAA Serverless

- No environment to manage. Pay only for task execution time
- Declarative YAML workflows (DAG Factory format)
- Per-task isolation. Each task in its own IAM-scoped container
- Best for event-driven, bursty workloads

Data, ML and GenAI Orchestration patterns with Amazon MWAA



1

Orchestrating Cross-Region Multi-Account Data & AI/ML Pipeline with provisioned Amazon MWAA

2

Orchestrating AI Workflows using Amazon MWAA Serverless



1 Orchestrating Cross-Region Multi-Account Data & AI/ML Pipeline with provisioned Amazon MWAA

Building data and ML pipelines individually is not the most difficult part.
Orchestrating them together across accounts and regions is where the real complexity lives.

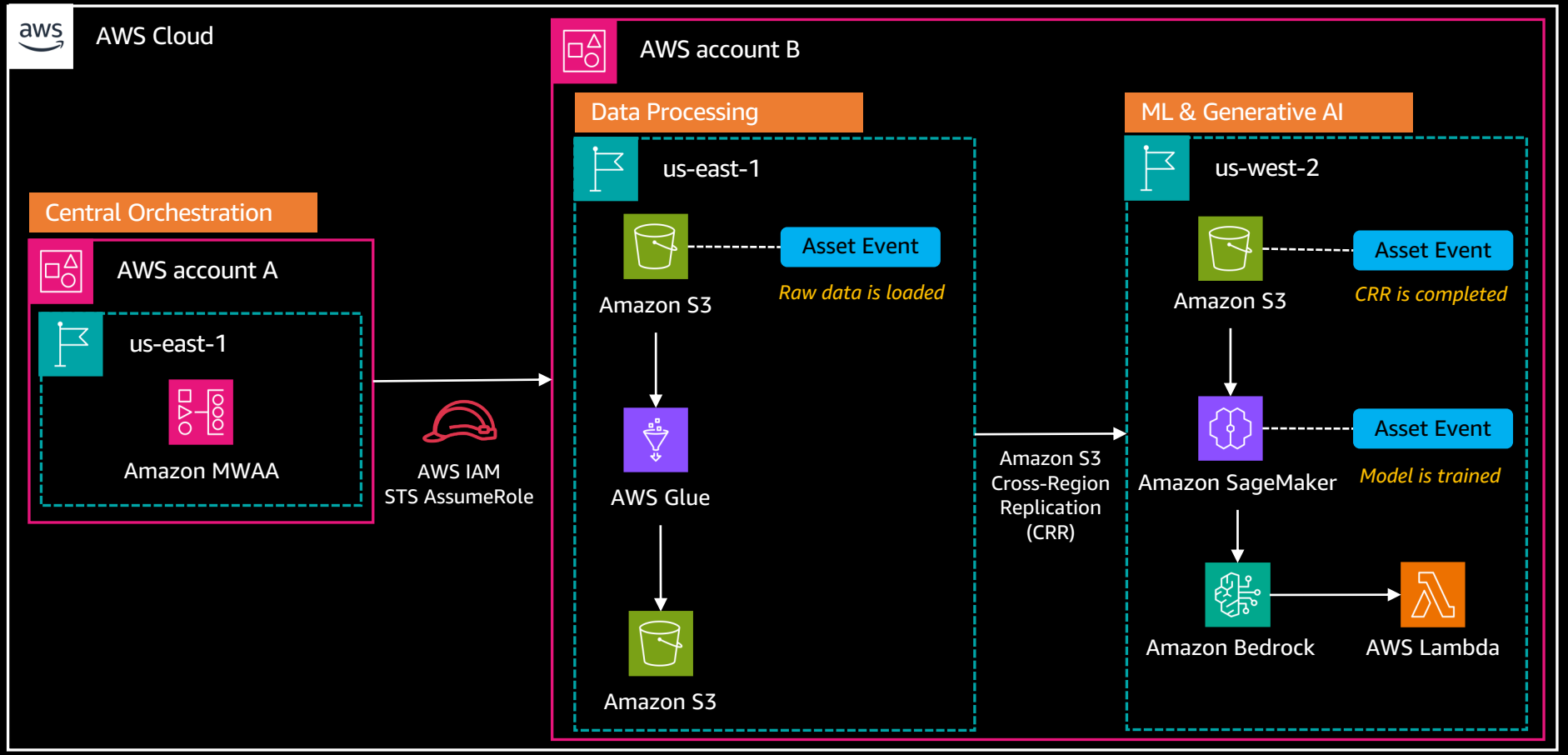
A global enterprise needs to automate an end-to-end ML pipeline — from raw Data processing to ML Model training to Generative AI evaluation for a production deployment decision — across multiple AWS accounts and regions, without manual intervention between stages

Challenges

Security	Cross-account access without long-lived credentials
Data Movement	Cross-Region Replication for data movement
Intelligent Handoffs	Chain three distinct pipeline phases together automatically

2 AWS Accounts	2 AWS Regions
6 AWS Services	1 Orchestrator

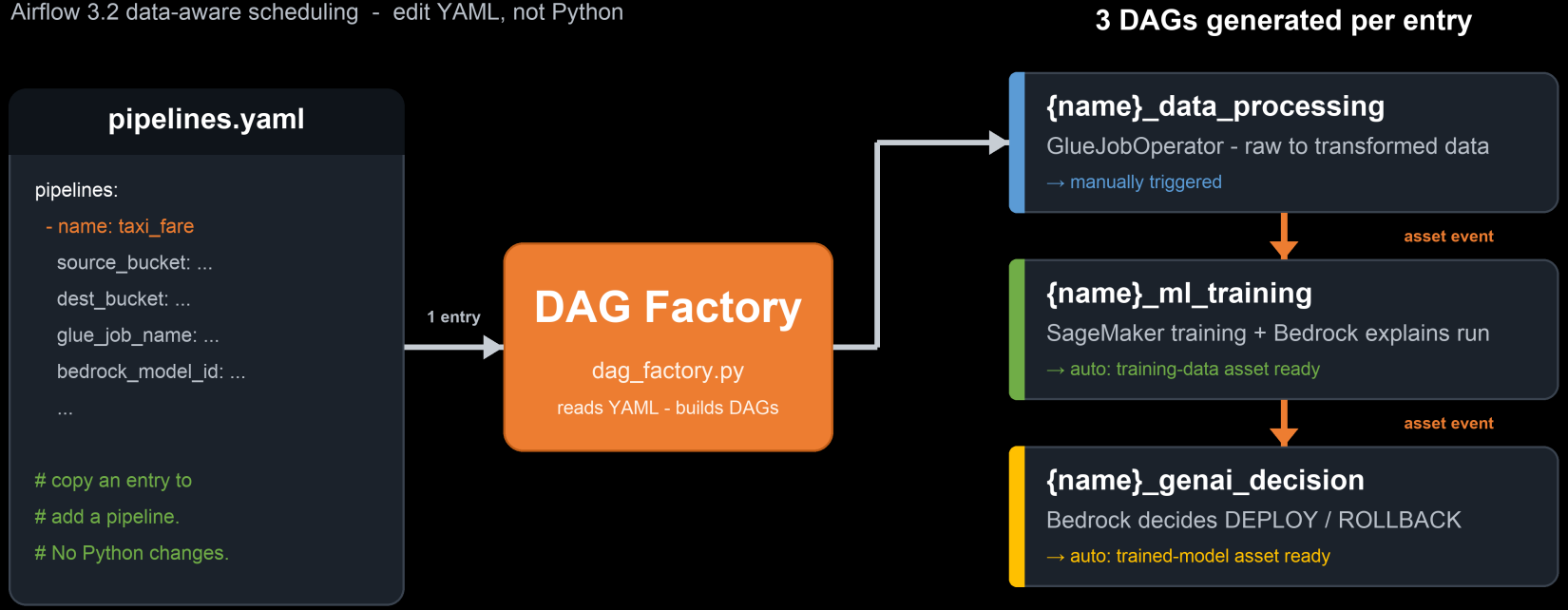
Solution Architecture



DAG Factory Flow

Config-Driven DAG Factory

Airflow 3.2 data-aware scheduling - edit YAML, not Python

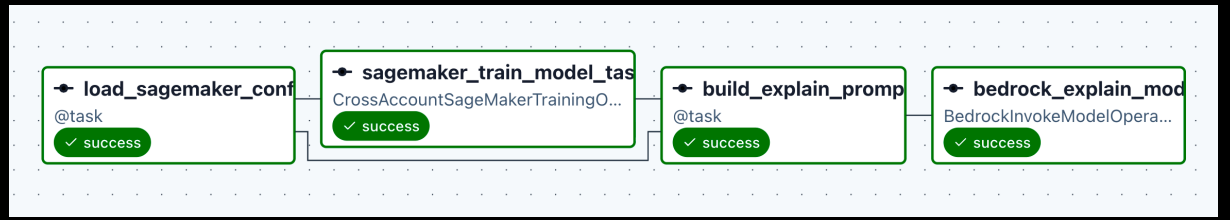


DAGs

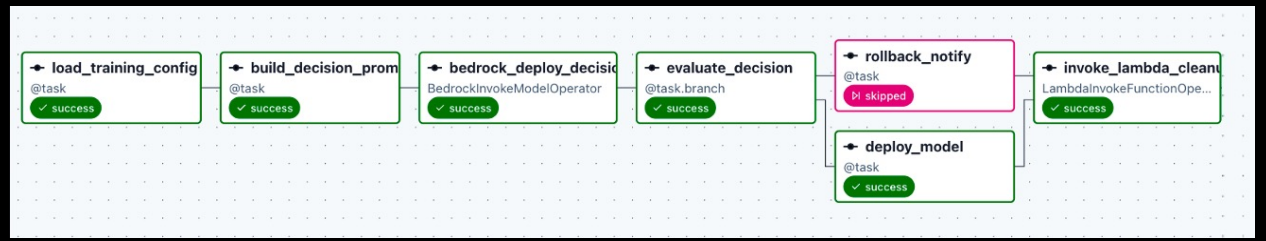
Data Processing DAG



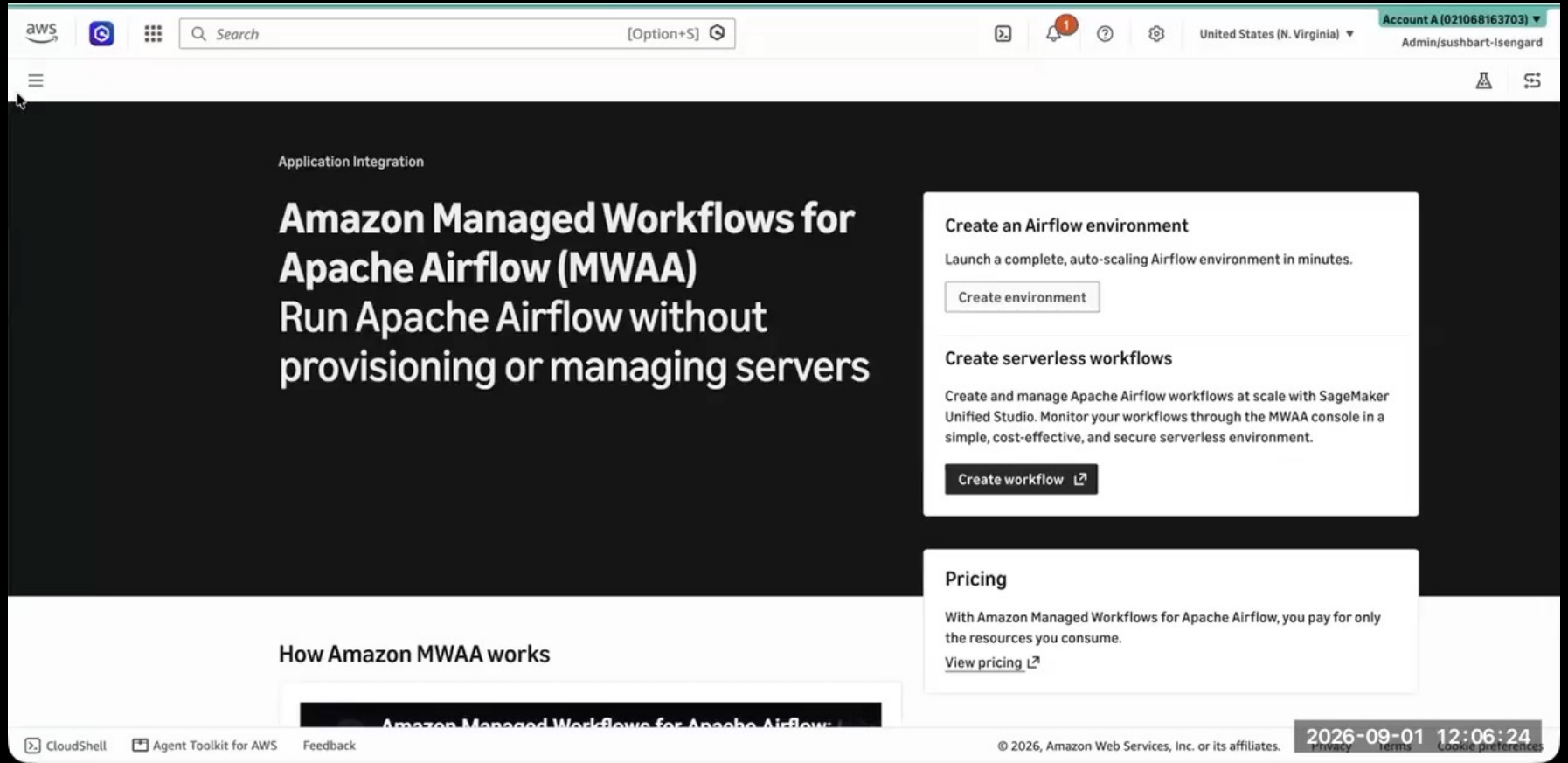
ML Training DAG



GenAI Evaluation DAG



Demo


Search
[Option+S]
United States (N. Virginia)
Account A (021068163703)
Admin/sushbart-Isengard
Application Integration
Amazon Managed Workflows for Apache Airflow (MWAA)
Run Apache Airflow without provisioning or managing servers
Create an Airflow environment
Launch a complete, auto-scaling Airflow environment in minutes.
Create environment
Create serverless workflows
Create and manage Apache Airflow workflows at scale with SageMaker Unified Studio. Monitor your workflows through the MWAA console in a simple, cost-effective, and secure serverless environment.
Create workflow
Pricing
With Amazon Managed Workflows for Apache Airflow, you pay for only the resources you consume.
View pricing
How Amazon MWAA works
CloudShell
Agent Toolkit for AWS
Feedback
© 2026, Amazon Web Services, Inc. or its affiliates.
2026-09-01 12:06:24

2 Orchestrating AI Workflows using Amazon MWAA Serverless

As we move from experimentation to production, organizations need to orchestrate AI-driven decisions as repeatable, auditable, governed workflows, not standalone API calls.

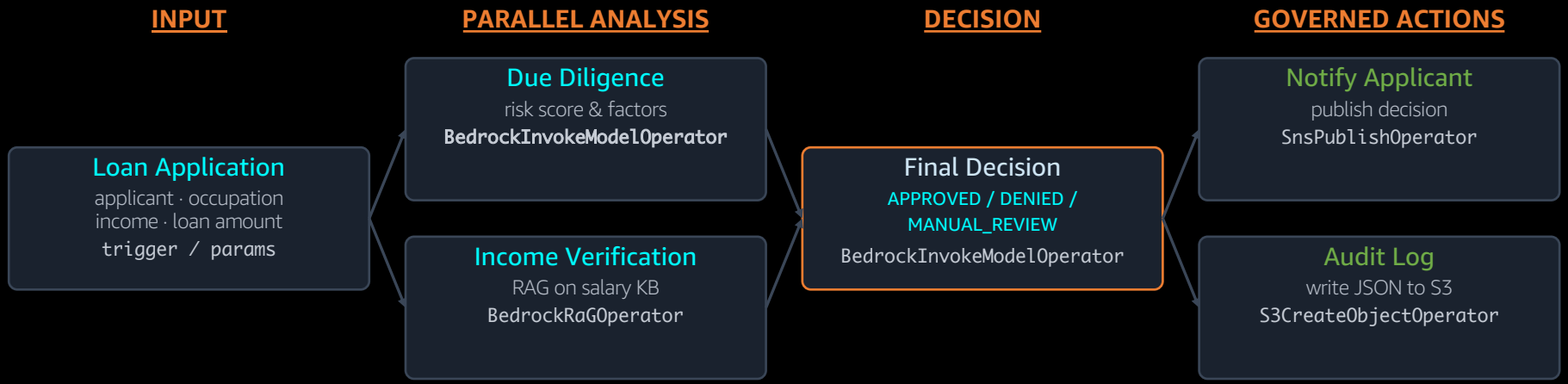
A financial services organization needs to automate loan approval decisions using GenAI — evaluating applicant financial profiles and making governed approval or rejection decisions at scale, without managing Airflow infrastructure or paying for idle compute.

Challenges

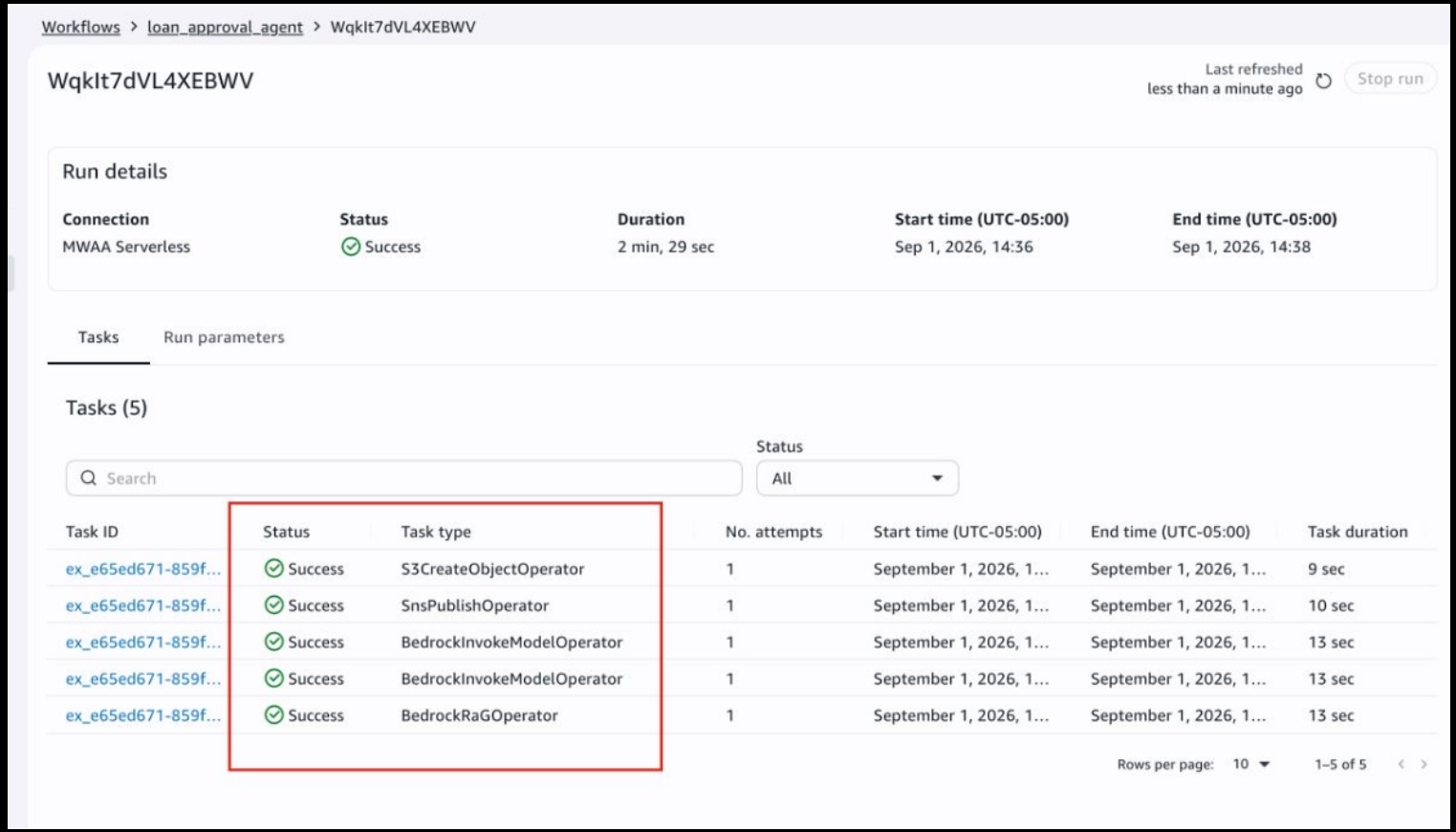
Governance & auditability	Every AI-driven loan decision must be explainable, logged, and traceable for regulatory compliance
Cost at scale	Processing thousands of applications requires GenAI inference on demand without paying for always-on Airflow workers sitting idle between submissions
Security isolation	Each loan evaluation handles sensitive PII (income, credit data) and needs task-level permission scoping no shared execution context across applicants

AI Loan Approval Agent using MWAA Serverless

No custom code, No infrastructure to manage.



Workflow Run



Key takeaways

- **One orchestrator, any pattern**
  - Apache Airflow on AWS is the single control plane for data, ML, and GenAI pipelines.
  - Amazon MWAA runs it fully managed; MWAA Serverless removes the environment entirely - pay per task.
- **Two patterns**
  - Enterprise scale - cross-region, multi-account Data + ML on provisioned MWAA.
  - Build complex AI workflows fast - governed loan approval on Serverless MWAA, no custom code, no idle compute.

Start from these reference samples - adapt them to your own use case



AIRFLOW SUMMIT

Thank you



Sneha Rao
Solutions Architect
Amazon Web Services (AWS)

Sushmita Barthakur
Sr. Data Solutions Architect
Amazon Web Services (AWS)

Suba Palanisamy
Sr. Technical Account Manager
Amazon Web Services (AWS)