From Centrailization to Autonomy
Managing Airflow Pipeline through Multi-Tenancy

3.0

SILVER PANG
Software Developer @ Expedia Group

Technology landscape keep shifting

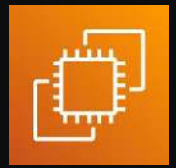
Organizational Challenges

Platform Engineering

Design, Build, Maintain the shared infrastructure

Time Travel back 2018

Own its own tech stack

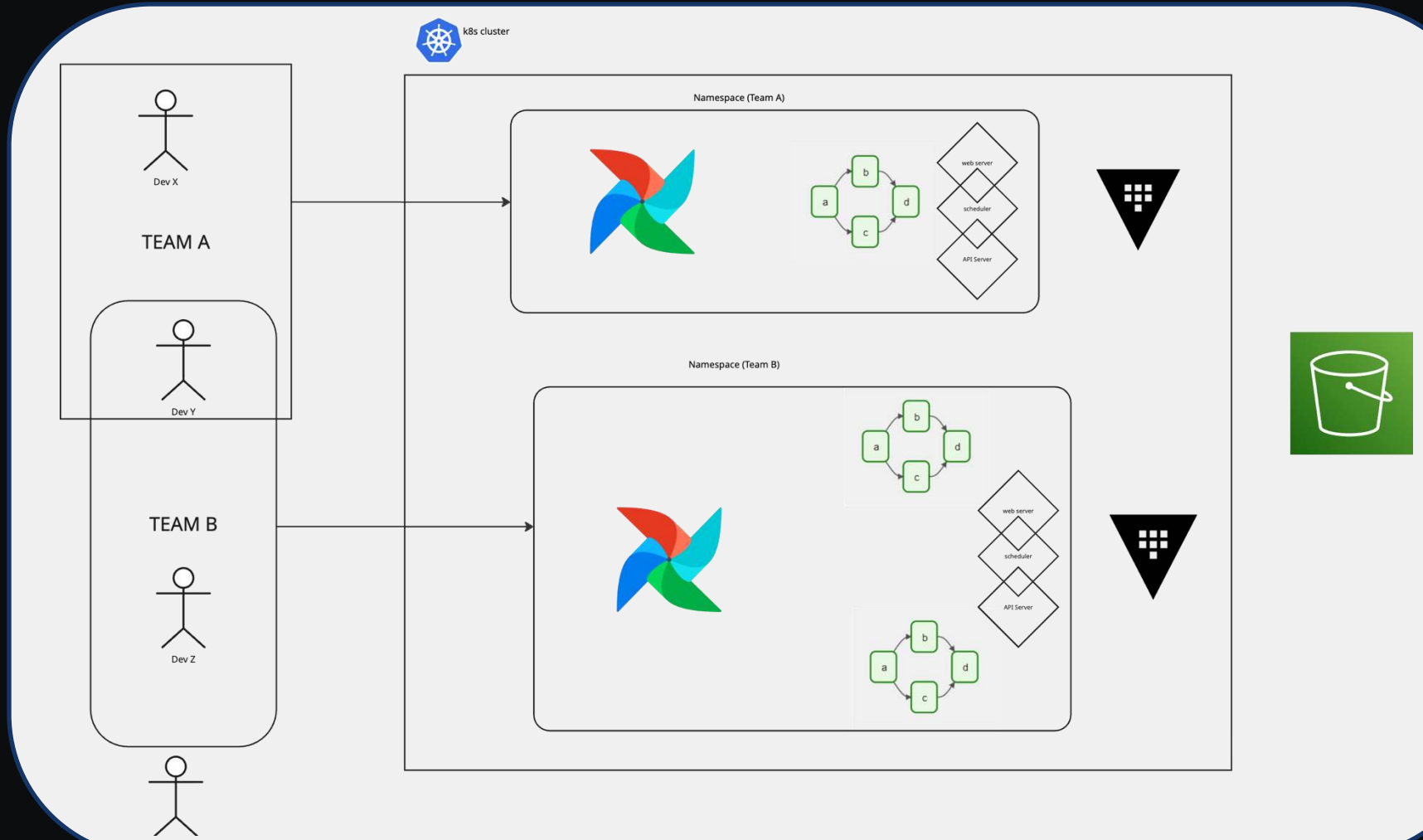


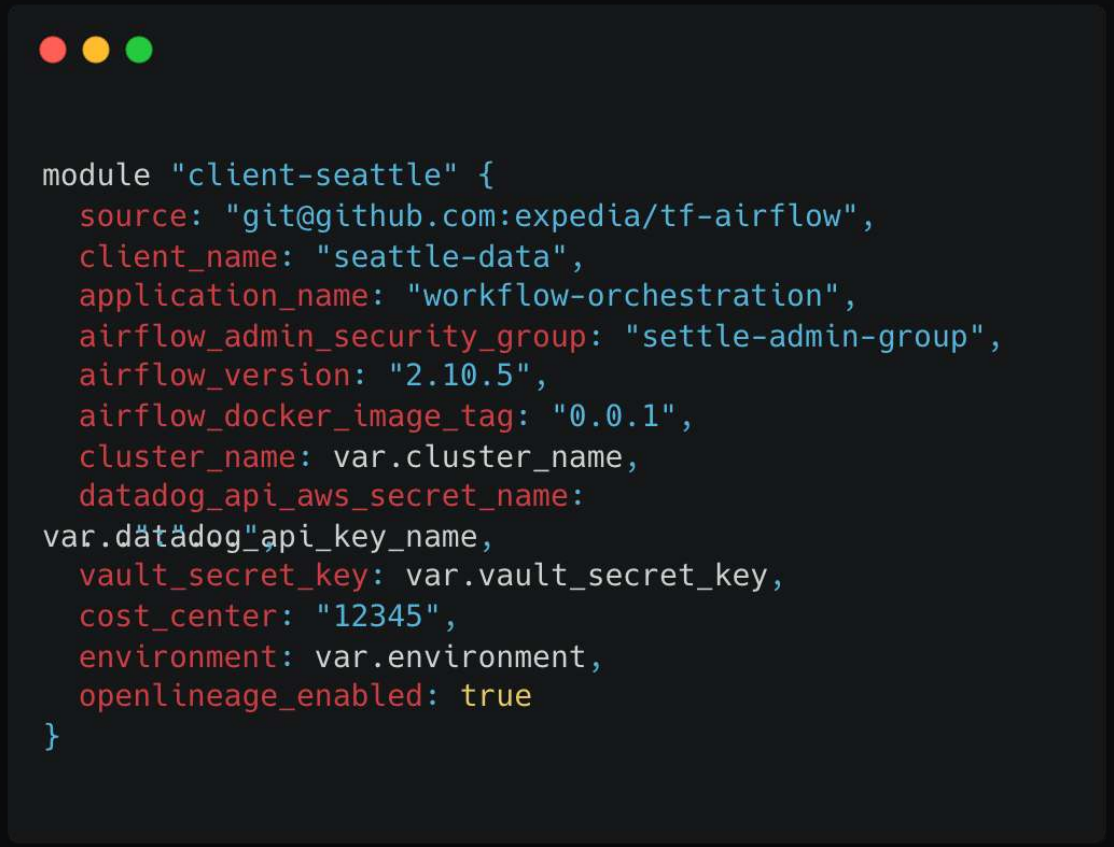
The Enterprise Dilemma: When Autonomy Creates Chaos.

- Bottlenecks & Slowed Development
- Inconsistent Standards
- High Operational Overhead for every team

Platform Engineering stepped in

Providing toolkit
Data Processing Framework
Batch / Stream compute / Workflow Orchestration


k8s cluster
Dev X
TEAM A
Dev Y
TEAM B
Dev Z
Namespace (Team A)
web server
scheduler
API Server
a
b
c
d
Namespace (Team B)
web server
scheduler
API Server
a
b
c
d



```
module "client-seattle" {
  source: "git@github.com:expedia/tf-airflow",
  client_name: "seattle-data",
  application_name: "workflow-orchestration",
  airflow_admin_security_group: "settle-admin-group",
  airflow_version: "2.10.5",
  airflow_docker_image_tag: "0.0.1",
  cluster_name: var.cluster_name,
  datadog_api_aws_secret_name:
var.datadog_api_key_name,
  vault_secret_key: var.vault_secret_key,
  cost_center: "12345",
  environment: var.environment,
  openlineage_enabled: true
}
```

apache-airflow-providers-amazon==8.12.0
apache-airflow-providers-datadog==3.4.0
awscli
boto3
datadog
trino
pandas

The DevEx | Templates for speed
When Autonomy Creates Chaos.

// TODO add vault

From zero to pipeline in “minutes”

1 Provide Application Metadata — 2 Provide additional information — 3 Include TechDocs — 4 Review

This data is used to register your application in AMS

GitHub Instance *

Github Organization *

The GitHub Organization where the repository is located. Verify the instance above is correct for the organization chosen.

Application Name *

Name of the application. Application name should only contain characters [a - z], [0 - 9] and hyphens. Maximum number of characters allowed is 63.

Description *

Describe this service. Help others understand what this service does. What problem does it solve? What benefits does it provide? When should a person definitely use this template?

Team Name *

Name of the support team for this application service. More details can be found here

Documentation Overview Link

A link to the Overview Documentation for the Application. More details can be found here

Application Type *

Type of Application. More details can be found here

PCI Category *

PCI Category of the application. More details can be found here

Environment Scope *

Environment Scope of the application. More details can be found here

Support Slack Channel

Please include the public support Slack channel of your team for communications. Do not include # in the name, for example: my-support-channel

Funnel

The Funnel attribute is part of the PagerDuty Information for Incidents.This field is mandatory for all the Tier 0 and Tier 1 Applications. More details can be found here

Application Lifetime *

Application Lifetime specifying if the application is temporary or permanent. Choose Temporary if your application belongs to demo/hello world/bootcamp or other limited shelf-life categories. If not, choose permanent. More details can be found here
If you choose Temporary and it meets the criteria for automated decommissioning, the app will be automatically decommissioned. For more details, please refer to FAQs on Automated App Deactivation.

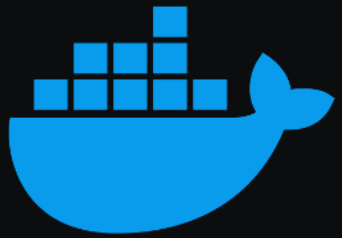
Email *

Fully Automated CI/CD pipeline

Fully Automated CI/CD pipeline

Fully Automated CI/CD pipeline

Fully Automated CI/CD pipeline



Fully Automated CI/CD pipeline

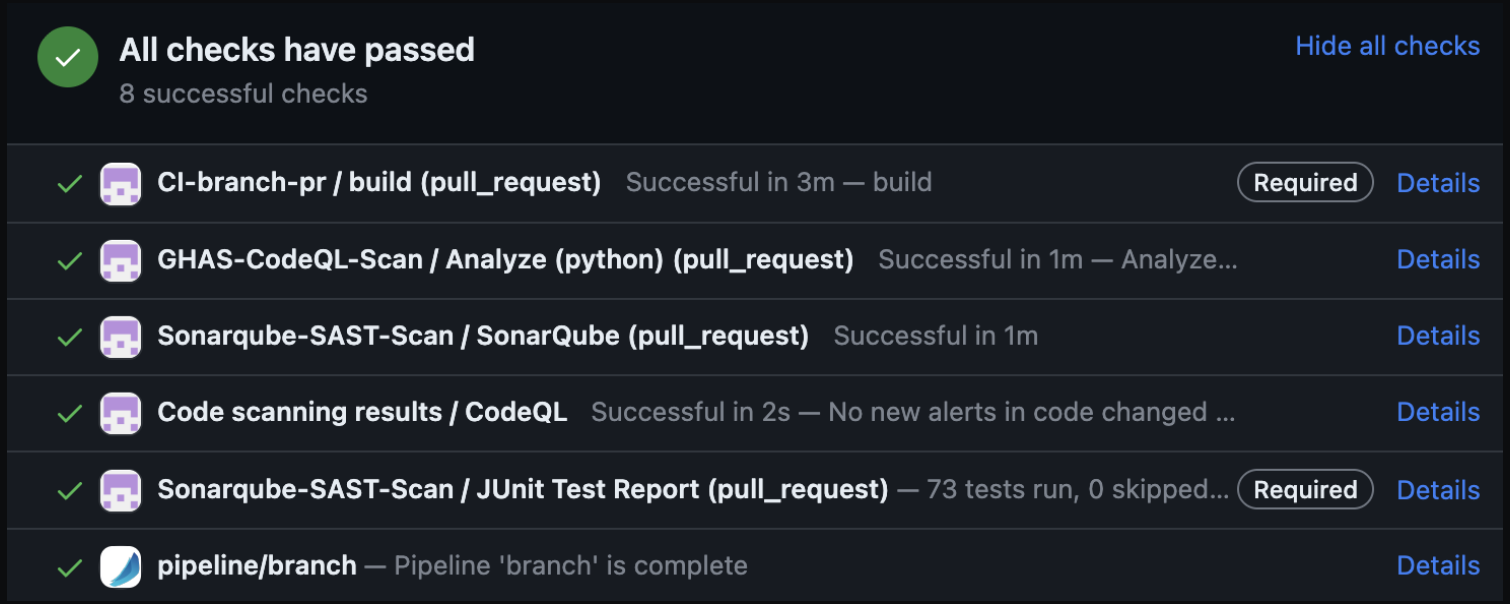
JFrog Artifactory

Fully Automated CI/CD pipeline

Fully Automated CI/CD pipeline

Fully Automated CI/CD pipeline

Fully Automated CI/CD pipeline



20+ New Pipelines per Month

200+ Airflow Clusters

180+ Engineering Teams

14k+ Active DAGs

1.5M+ Task Executions per Month

Reduce Airflow cluster deployment time

Reduce time for DAG deployment

Easier Migration

Hard Lesson Learned

Code Consistency
Compliance & Governance
The Learning Curve

Automated Developer Assistant APP

Job ID [redacted]05e6 in dw


```
{
  "instance_key": "[redacted]",
  "uuid": "[redacted]05e6",
  "name": "[redacted]-spark-job",
  "application_name": "[redacted]-migrator-spark",
```


Show more

To access `marketing-[redacted]` Airflow instance, Ensure you are a member of the security group `[redacted]-marketing-[redacted]` using [redacted] Group Management

Health:

- metadatabase: healthy
- scheduler: healthy
- webui: healthy

Config:

- Airflow version: 2.10.5
- Data[redacted] Operator version: 2.0.9

See less

Key Takeaway

Platform is the riverbed
Autonomy is flow
Adaptability is the current

From Centrailization to Autonomy
Managing Airflow Pipeline through Multi-Tenancy

3.0

SILVER PANG
Software Developer @ Expedia Group