

Empowering Teams in your Organization to Self-Service their Airflow Needs

Spencer Tollefson
Data Engineering Team Lead
Rakuten Kobo



About Me

Spencer Tollefson, Data Engineering Team Lead



Agenda

- Use case: multiple teams sharing an Airflow environment
- Share how we address the needs of these teams at Kobo:
 - Delineate responsibilities when it comes to different teams
 - Build guard rails for new Airflow developers
 - Alerts and monitoring
 - Codify access control based on team assignment
 - Operations versus Data Engineering responsibilities



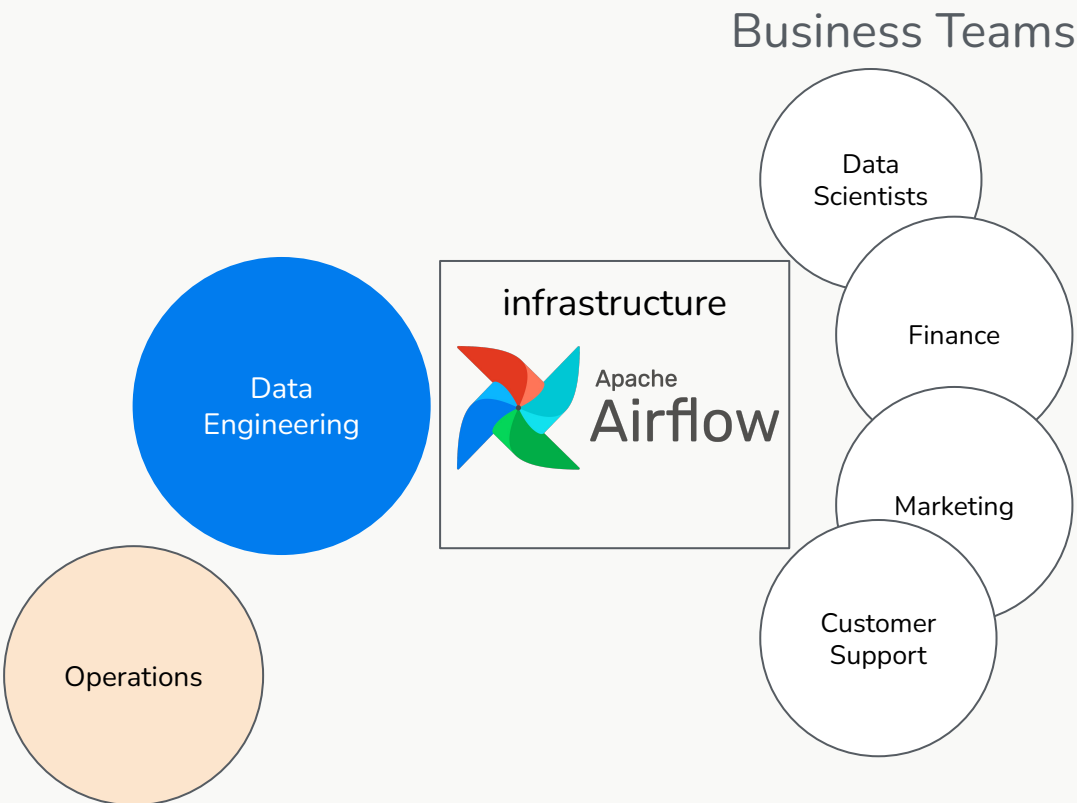
What this talk is:

- A sharing of how teams at Kobo use Airflow successfully

What this talk will not cover:

- Multi-tenancy

Use Case: Multiple Teams Sharing an Airflow Environment



All business teams have in common:

- Business teams have developers
- Devs are not Airflow experts
- The devs are the experts of their domains. The data engineers and operations folks are not



Delineate responsibilities between Data Engineering and Developers

- Developers know the business logic
- Data engineers know Airflow
- Developers author their own Airflow DAGs
 - Pair program between data engineers and developers
- Code review responsibilities:

Data Engineers	Developers	Shared
<ul style="list-style-type: none">● Airflow best practices● Optimization	<ul style="list-style-type: none">● Purpose of DAG● Business logic	<ul style="list-style-type: none">● Knowledge of various systems affected● Clean code● Some parameters (ex: scheduling)● Dependencies

The top right corner of the slide features several circular stickers celebrating Airflow's 10th anniversary. One sticker is red with a red pinwheel and the text 'AIRFLOW 10 YEARS OF AIRFLOW SF-USA 2024'. Another is green with a green pinwheel and 'AIRFLOW 10 YEARS OF AIRFLOW SF-USA 2024'. A third is blue with a blue pinwheel and 'AIRFLOW SUMMIT 2024 10 YEARS OF AIRFLOW SF-USA'. A fourth is red with a red pinwheel and 'AIRFLOW 10 YEARS OF AIRFLOW SF-USA'.

Build guard rails for new Airflow developers

- Have a really good README document
- Example DAGs
- Present and enforce code linters and formatters
- Setting up local Docker environment for Airflow
- Setting up a prod-similar staging environment (running on same infrastructure)
 - Code here unable to touch prod systems
- Leave secret management to Operations
- Access Control (see later slide)

Alerts and Monitoring

- Require DAGs to set “on_*_callbacks”
- Integrate an alert service to notify team-specific Slack channels
- Data Engineers and developers communicate about alert on Slack



Codify DAG Access Control Based on Team Assignment

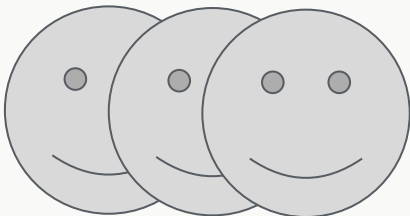
OPERATIONS

New business
team developer



Assign →

Team



DATA ENGINEERS



Apache
Airflow

Create “custom
role” that maps to
the business team

All DAGs require
“access_control”
param

Operations Versus Data Engineering Responsibilities

Operations	Data Engineering
<ul style="list-style-type: none">● Own configuration of our self-hosted infrastructure● Secrets management● Do not know Airflow● Granting users access to Airflow environments	<ul style="list-style-type: none">● Some infrastructure knowledge and troubleshooting● Airflow deployments (upgrades, package changes)● General knowledge of what all Airflow DAGs are doing● But not experts in the business logic● Experts in best practices in terms of resource usage, scheduling, configuration, executors, etc.

Summary

- Have a team that knows Airflow inside-and-out
- This “Airflow team” supports business teams, but does not:
 - Write the DAGs
 - Resolve all the DAG failures
- Invest in ways to make it simpler for non-power users to jump into Airflow
- Automate team assignment within organization to map to DAG permissions in Airflow
- Remove complexities from developers. The data engineers should be their experts

Taking it Farther

- Explore using YAML files to abstract even more Airflow details away from developers
- Multi-tenancy

Questions?

Spencer Tollefson

[linkedin.com/in/stollefson1](https://www.linkedin.com/in/stollefson1)

tollefson.spencer.j@gmail.com

