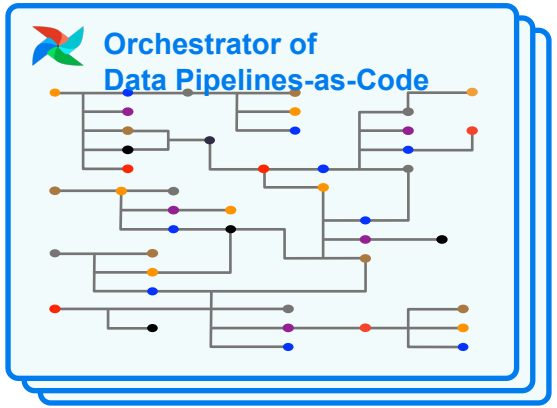
Using Talend and Airflow

Why Combine Talend and Airflow?

- If your team is moving to Astronomer but has existing Talend jobs, using the two together eliminates the need to migrate existing jobs to python code
- Use both tools for what they're good for. Talend for some use cases, python for others, and still have a one-stop-shop for orchestration, monitoring, logs, etc. with Astronomer
- Using Airflow for orchestration allows for easily running multiple jobs with dependencies, parallelizing jobs, monitoring run status and failures, and more



talend

How can Talend + Airflow be used together?

Talend Cloud

Use the Talend API with the Airflow SimpleHTTPOperator
(or other operators)

Talend Studio

Containerize studio jobs, publish to a registry, and execute using the KubernetesPodOperator

Talend Cloud + SimpleHTTPOperator

Pros

- Low barrier to entry
  - Does not require Docker or an image registry
- Does not require Talend Studio
- Very simple DAG code

Cons

- Airflow task completion is based on the API request, not the Talend job itself
  - Could be an issue for downstream dependencies
- Requires a Talend Cloud license to access API

Containerized Talend + KubernetesPodOperator

Pros

- Task completion is based on completion of Talend job within the container (i.e. task does not finish until Talend job is finished)
- Benefits of containerization such as
  - Flexibility
  - Scalability
  - Efficiency

Cons

- Requires Talend Studio. Cloud jobs can’t be containerized
- Requires some basic knowledge of Docker
- Requires a private registry (such as ECR, ACR, etc.) unless the image can be made public

Demo