

Prueba Técnica Globant

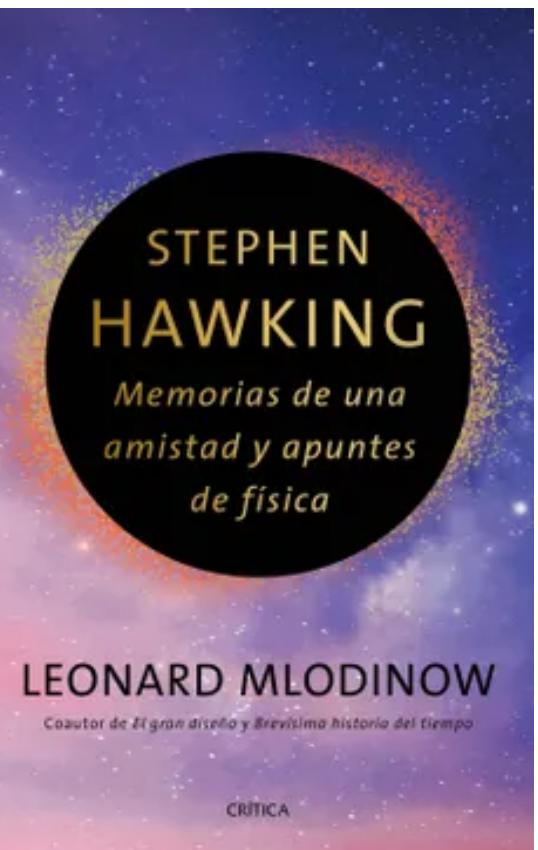
Prueba técnica para Data Engineer

Frederick Adolfo Salazar Sanchez - Senior Data Engineer

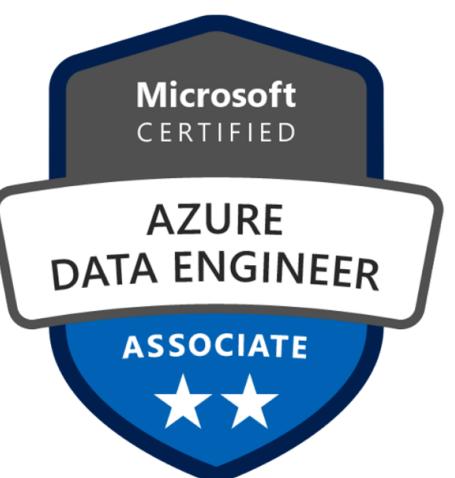


Frederick Salazar Sanchez

Data Engineer



Libro recomendado



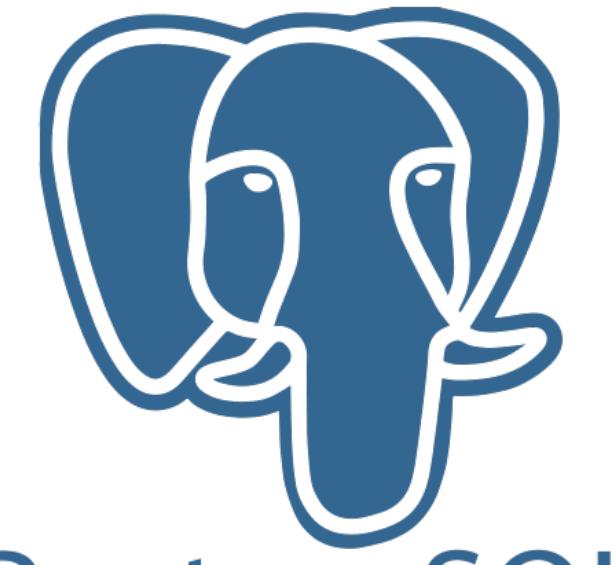
Consideraciones Técnicas

- Se propone dar un tratamiento al procesamiento de datos mas orientado a ingeniería de datos
- El flujo de datos será orquestado usando apache airflow
- El procesamiento de datos se hará en capas estilo datalake (Bronze, Silver, Gold)
- Para la ingesta y procesamiento se usará pandas
- Para el almacenamiento se usará postgresql con esquemas Bronze, Silver, Gold
- Se creará una tabla Horizontal con la integración de las 3 fuentes de datos
- El consumo de datos se hará mediante un API usando FastAPI
- Se propone la creación de un dashboard usando Apache Superset

Tecnologías usadas



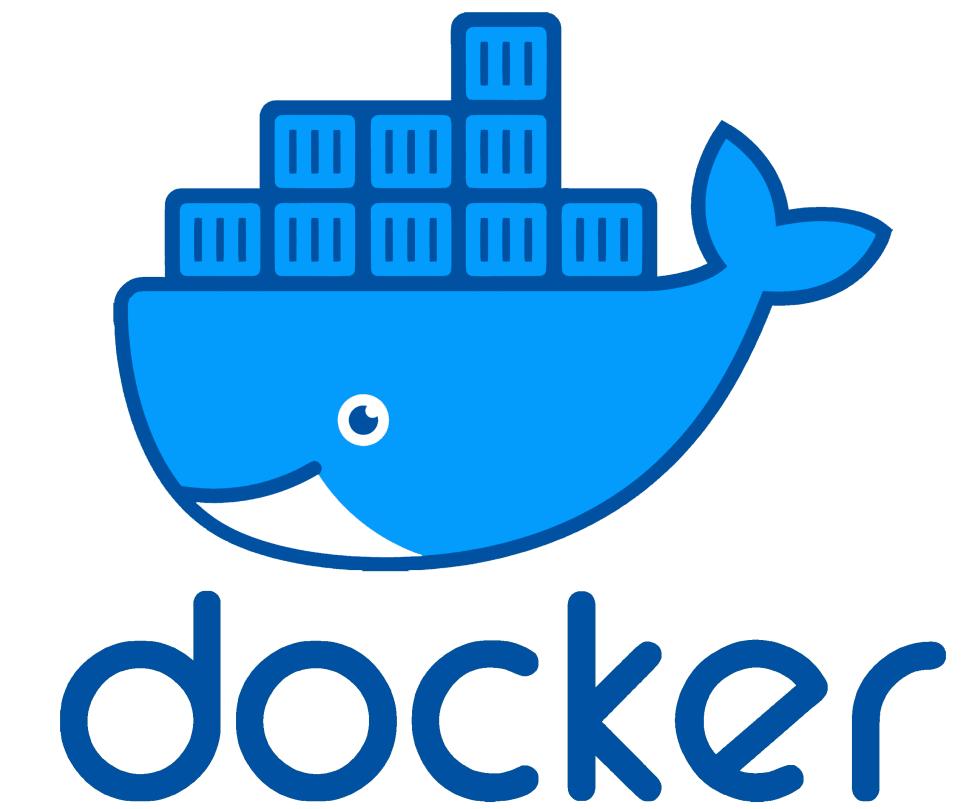
Apache
Airflow



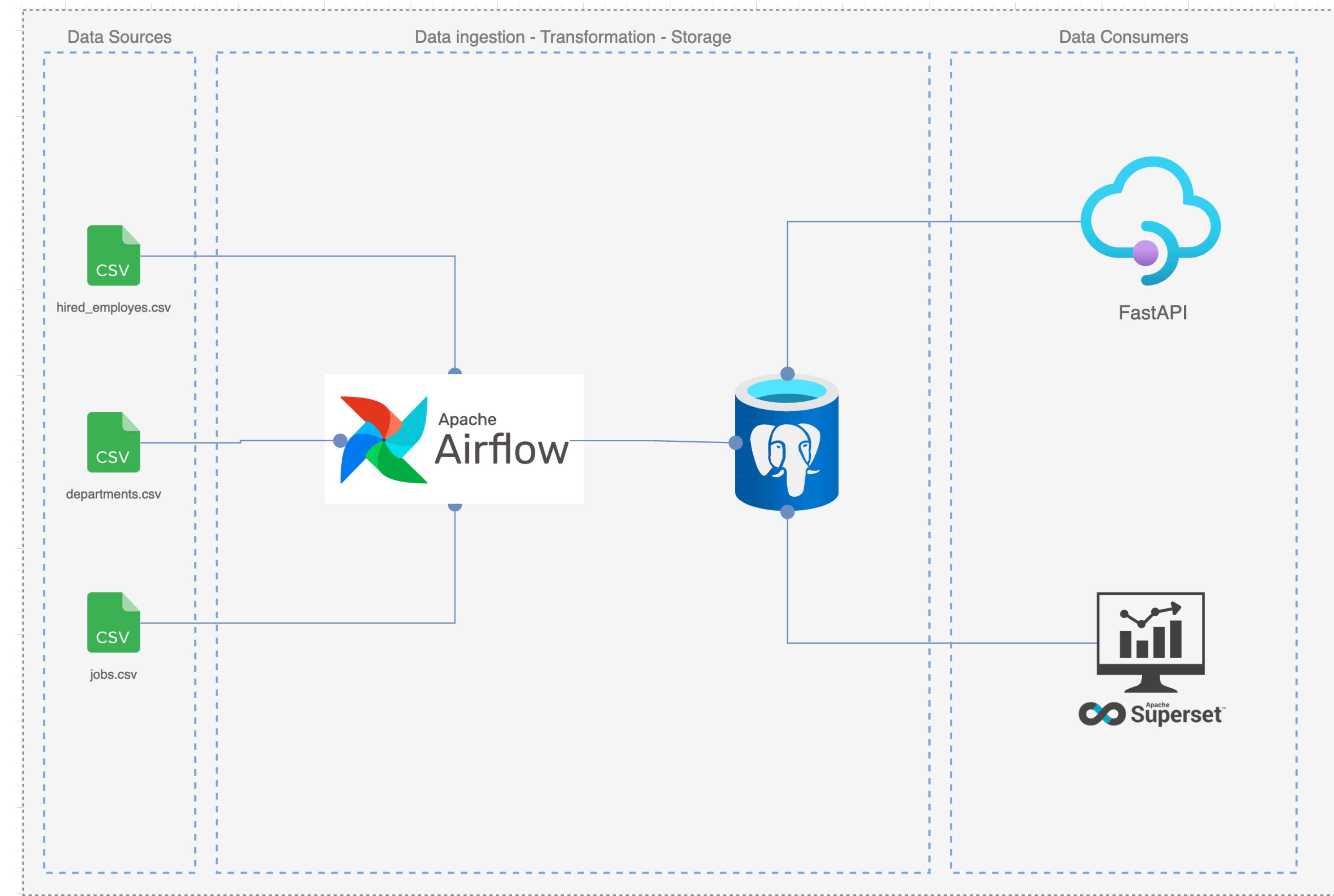
PostgreSQL



python™



Arquitectura de Referencia



Pipeline Globant

DAG: globant_pipeline

Schedule: None | Next Run ID: None |

23/10/2024 08:41:02 p.m. All Run Types All Run States Clear Filters Auto-refresh 25

Press shift + / for Shortcuts

deferred failed queued removed restarting running scheduled shutdown skipped success up_for_reschedule up_for_retry upstream_failed no_status

DAG Run Task
globant_pipeline / ► 2024-10-23, 20:41:01 UTC / run_bronze_notebook

Clear task Mark state as... Filter DAG by task

Duration 21:53 Oct 23, 20:40

00:00:12 00:00:06 00:00:00

run_bronze_notebook run_silver_notebook run_gold_notebook

Graph Gantt Code Event Log Logs XCom Task Duration

Layout: Left -> Right

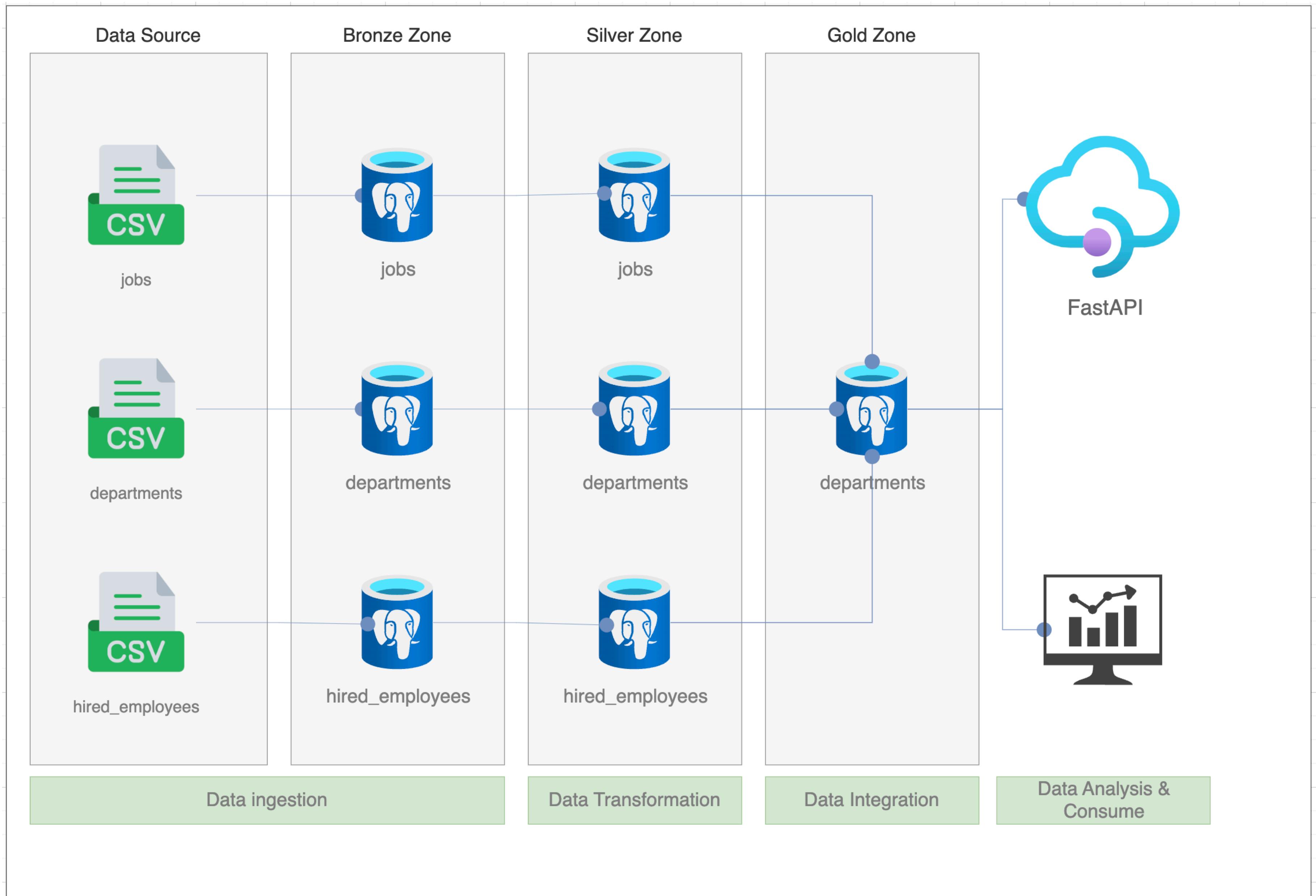
run_bronze_notebook success PapermillOperator

run_silver_notebook success PapermillOperator

run_gold_notebook success PapermillOperator

React Flow

Life Cicle Data



Gracias