

Airflow

- DAG: é um pipeline do Airflow
 - O que este pipeline vai fazer?
- Operators/Task





DAG

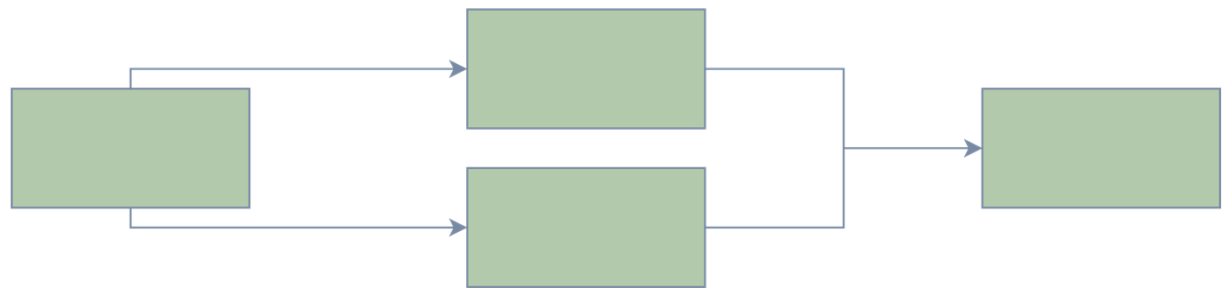
Directed acyclic graph: Grafo direcionado acíclico

Grafo: Conjunto de Vértices e Arestas

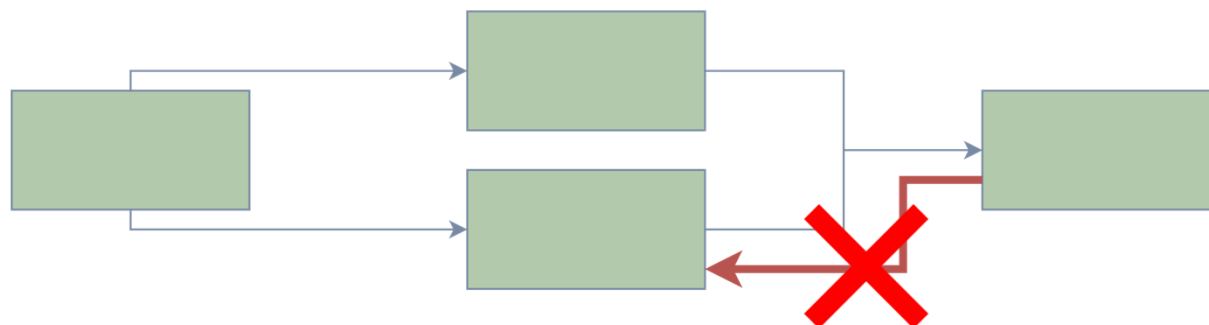
Direcionado: Possui Direção Específica

Acíclico: Não Circular

DAG



Sem Ciclos



Airflow

DAG é um script python

Não precisa ter a lógica (classe externa)

Um mesmo script pode ter múltiplas DAGs

Uma DAG pode ter uma ou mais tasks

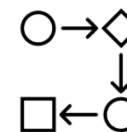
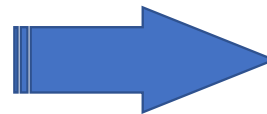
Tasks possuem dependência



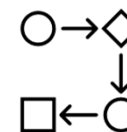
DAGs



Meupipeline.py



DAG: LoadS3

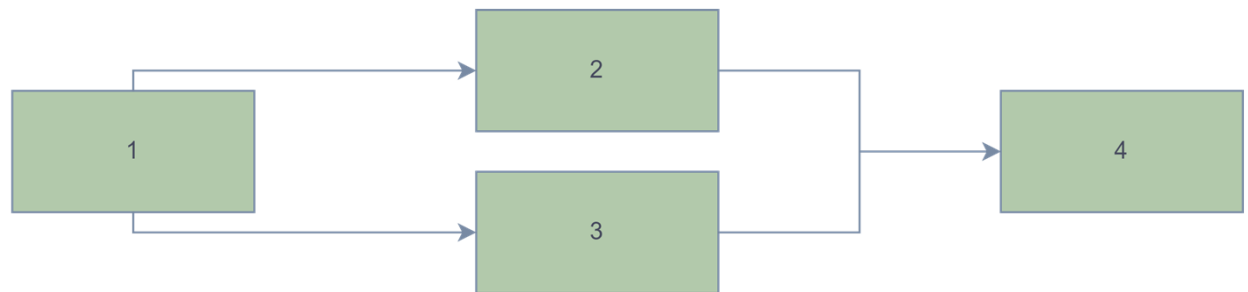


DAG: DataWipe

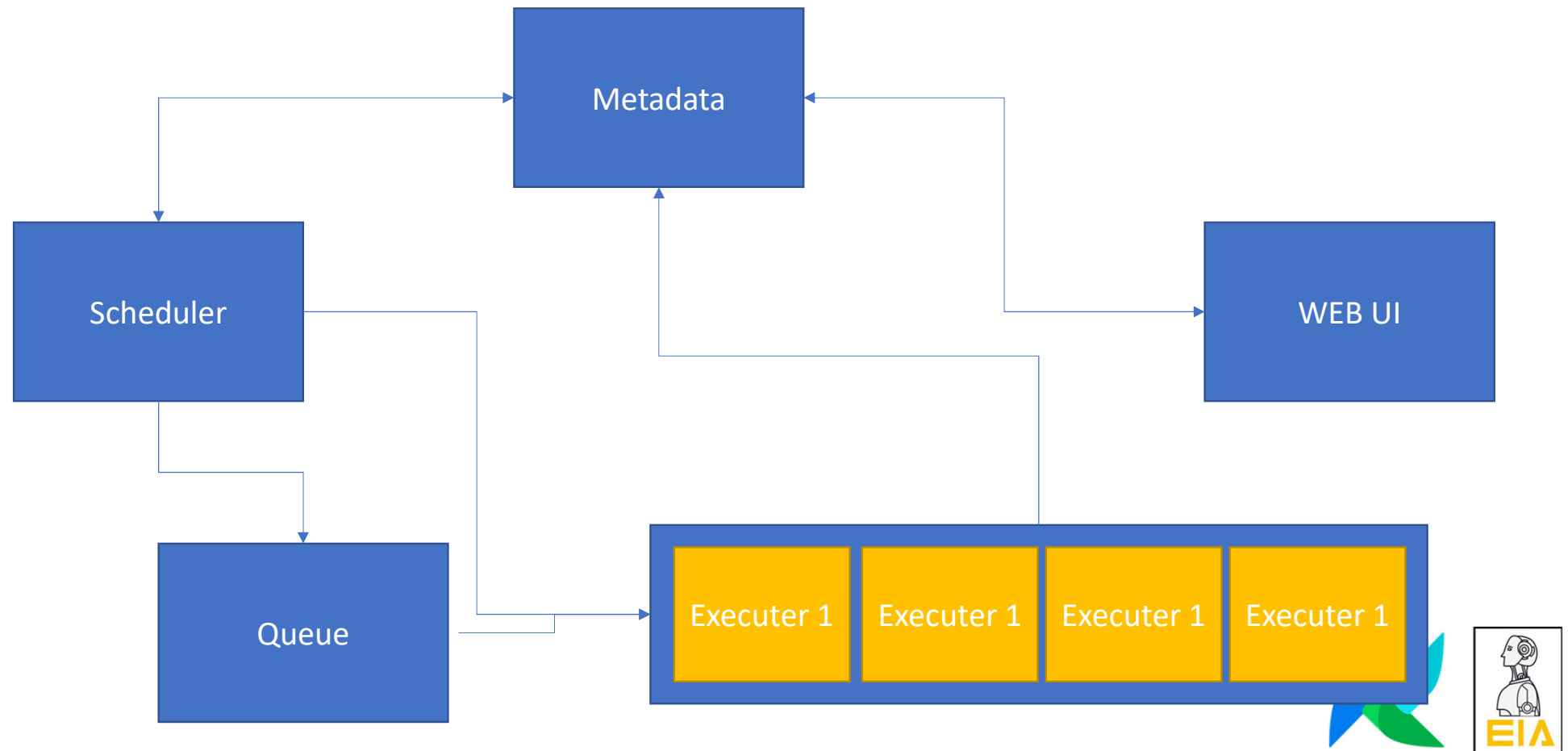


DAG

- Operators: O que é de fato executado.
- Tasks: instância de um operator



Arquitetura



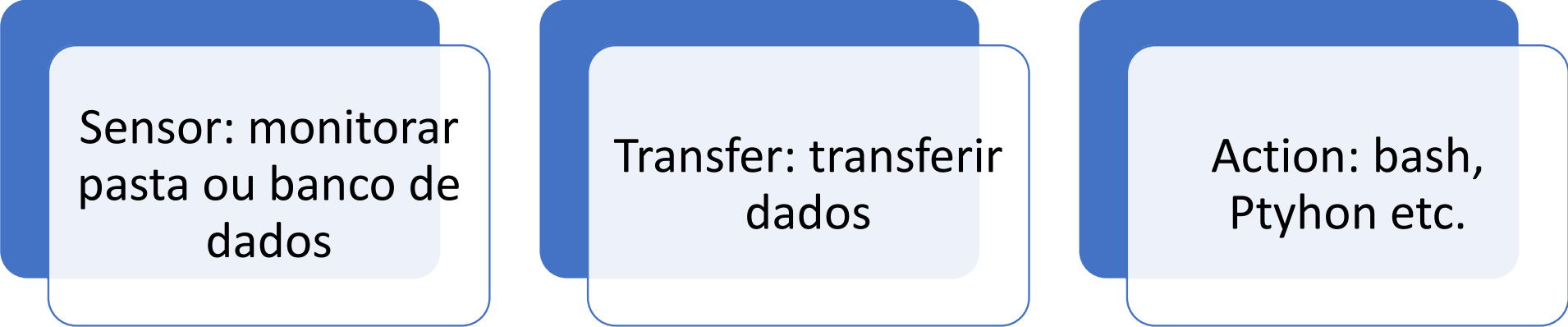


Operators

- Algo a ser feito no pipeline
 - Uma vez instanciado é uma tarefa
 - Existem muitos tipos, derivam de BaseOperator
-



Tipos de Operators



Sensor: monitorar
pasta ou banco de
dados

Transfer: transferir
dados

Action: bash,
Python etc.

Providers

- Interação com outras ferramentas:
 - Aws
 - Databricks
 - MySQL
 - MongoDB