

# Airflow

- DAG: é um pipeline do Airflow
  - O que este pipeline vai fazer?
- Operators/Task
- **Arquitetura:**
  - Webserver (UI)
  - Scheduler
  - Executor
  - Metadata DB
  - Triggerer
  - Queue





# DAG (Directed Acyclic Graph)

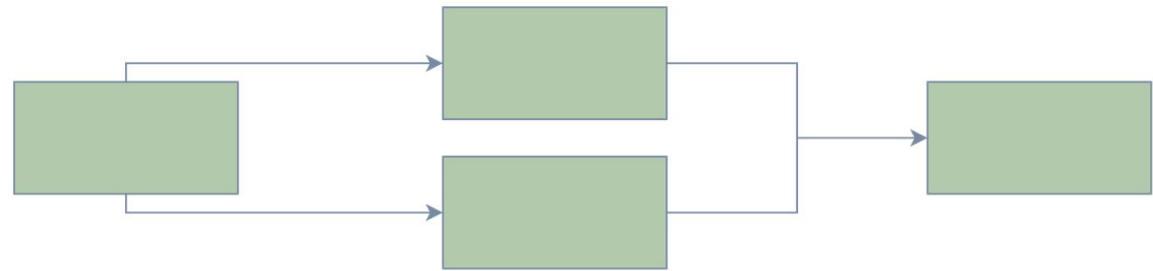
**Directed acyclic graph: Grafo direcionado acíclico**

**Grafo: Conjunto de Vértices e Arestas**

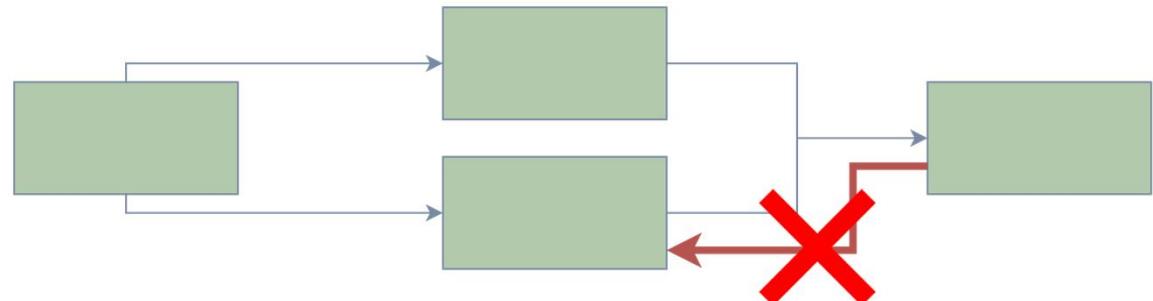
**Direcionado: Possui Direção Específica**

**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

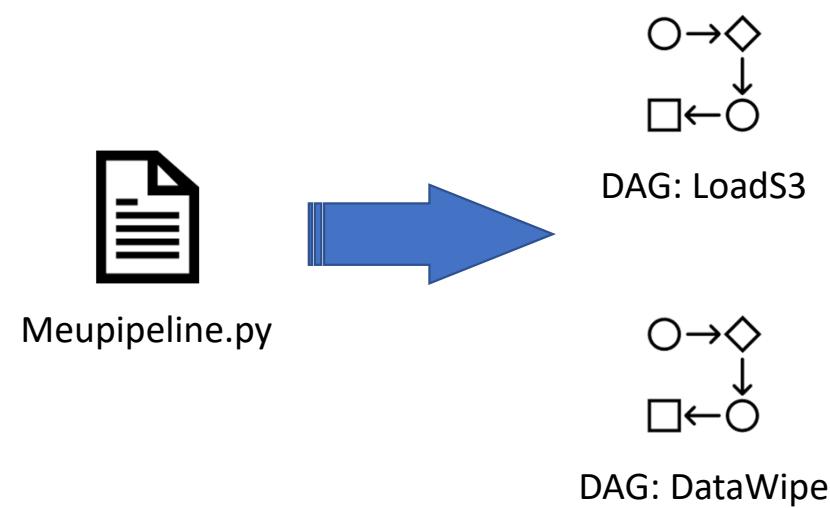
---

Uma DAG pode ter uma ou mais tasks

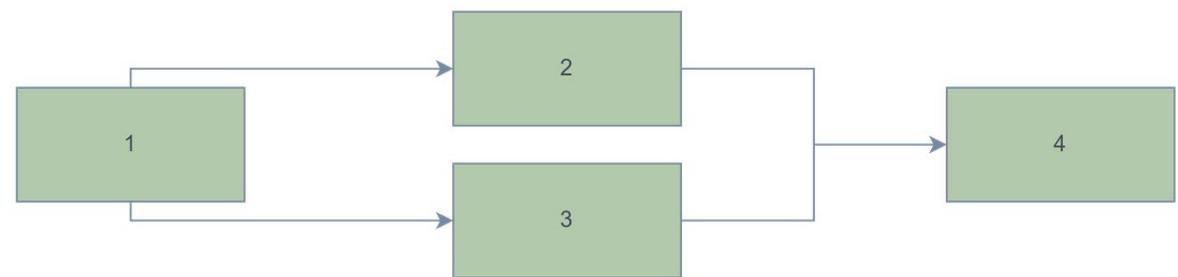
---

Tasks possuem dependência





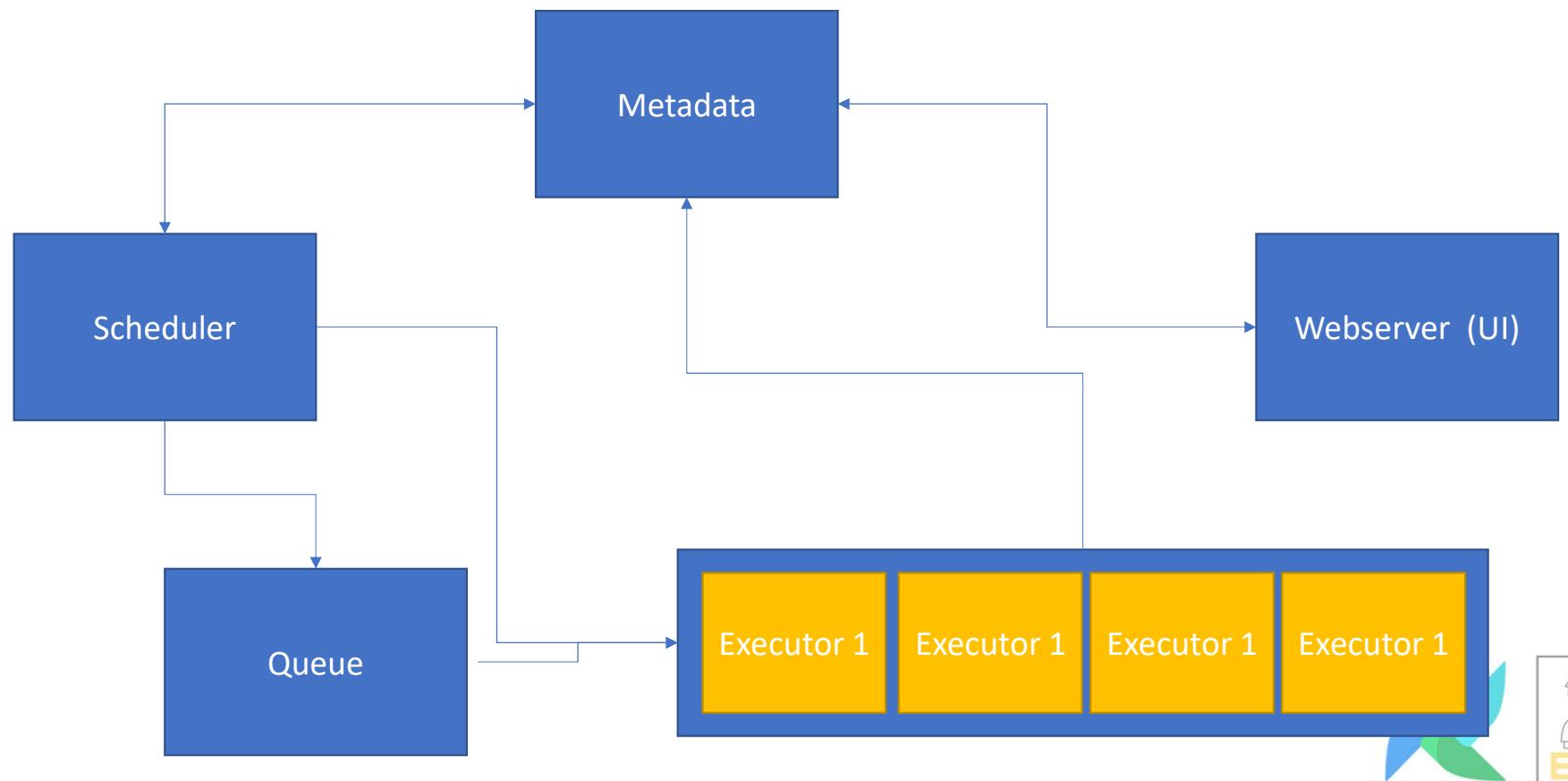
# DAG



- Operators: definição do trabalho
- Tasks: instância de um operator na DAG



# Arquitetura





# Operators

- Definem o ‘o que fazer’ no pipeline
  - Derivam de **BaseOperator**
  - Instanciados na DAG viram **Tasks**
-

# Tipos de Operators



**Action (Python/Bash/... )**



**Transfer**



**Sensors (deferrable)**



**Providers-specific**



# Providers



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