

# Apache Airflow

---

- Open Source – Criado pelo Airbnb em 2015. Mantido pela fundação Apache
- Desenvolvido em Python
- Extensível





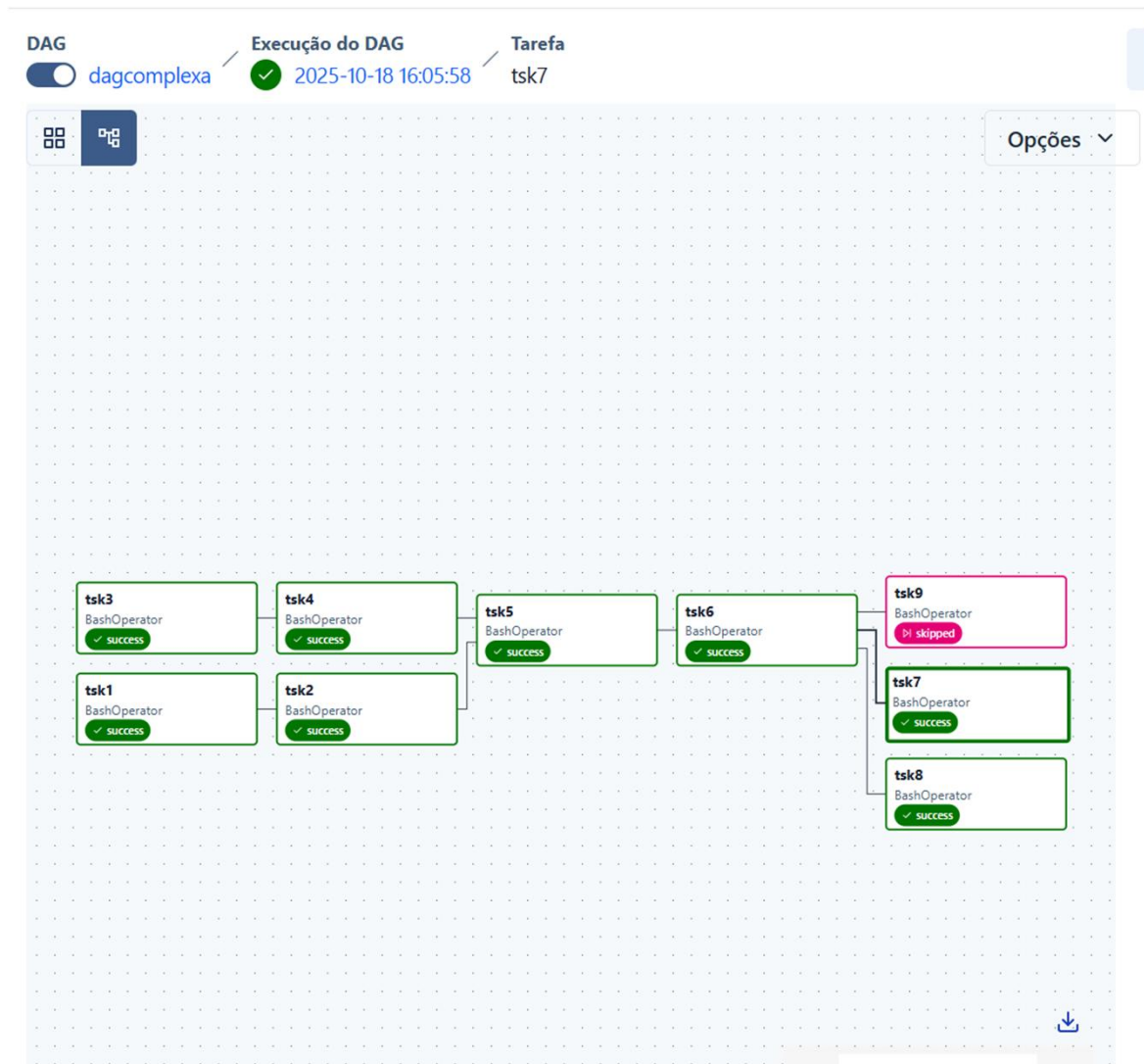
# Orquestrador

- Airflow não processa dados, ele coordena o processamento
- Orquestra workloads (principalmente batch)
  - Sensores e operadores “*deferrable*” → permite *near real time*
- O processamento ocorre em:
  - Sistema Operacional
  - Banco de Dados
  - Spark
  - Elastic Search
  - Etc.



# Interfaces

- Web
- REST API
- CLI



# DAGs

```
import pendulum
from airflow import DAG
from airflow.operators.bash import BashOperator

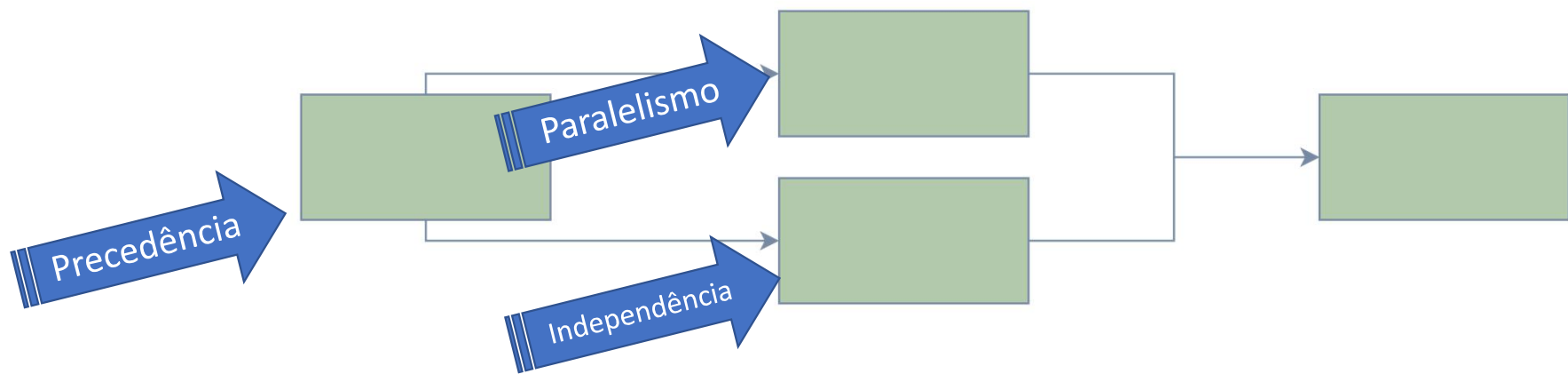
with DAG(
    dag_id="primeira_dag",
    description="Nossa primeira DAG",
    schedule=None,
    start_date=pendulum.datetime(2023, 3, 5, tz="America/Sao_Paulo"),
    catchup=False,
    tags=["curso", "exemplo"],
) as dag:
    task1 = BashOperator(task_id="tsk1", bash_command="sleep 5")
    task2 = BashOperator(task_id="tsk2", bash_command="sleep 5")
    task3 = BashOperator(task_id="tsk3", bash_command="sleep 5")

    task1 >> task2 >> task3
```





## DAG: Pipeline



# Pipeline

Erros

Log/Auditoria

Monitoramento/Alertas

Recuperação a partir de um Ponto

Dados Históricos/ Diferenciação

Alta disponibilidade

Distribuição de Carga