

Airflow

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





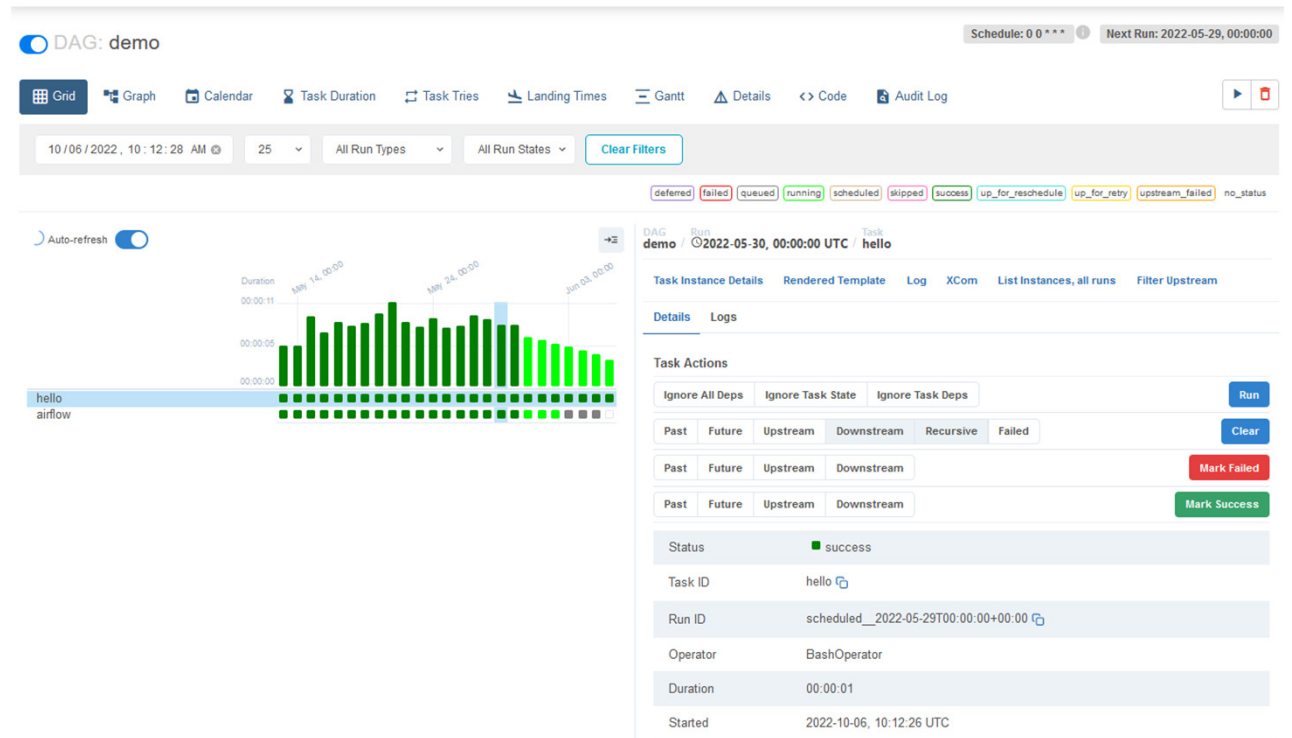
Orquestrador

- Airflow não processa dados, ele coordena o processamento
- Em batch
- O processamento ocorre em:
 - Sistema Operacional
 - Banco de Dados
 - Spark
 - Elastic Search
 - Etc.



Interfaces

- Web
- CLI



DAGs

```
from datetime import datetime

from airflow import DAG
from airflow.decorators import task
from airflow.operators.bash import BashOperator

# A DAG represents a workflow, a collection of tasks
with DAG(dag_id="demo", start_date=datetime(2022, 1, 1), schedule="0 0 * * *") as dag:

    # Tasks are represented as operators
    hello = BashOperator(task_id="hello", bash_command="echo hello")

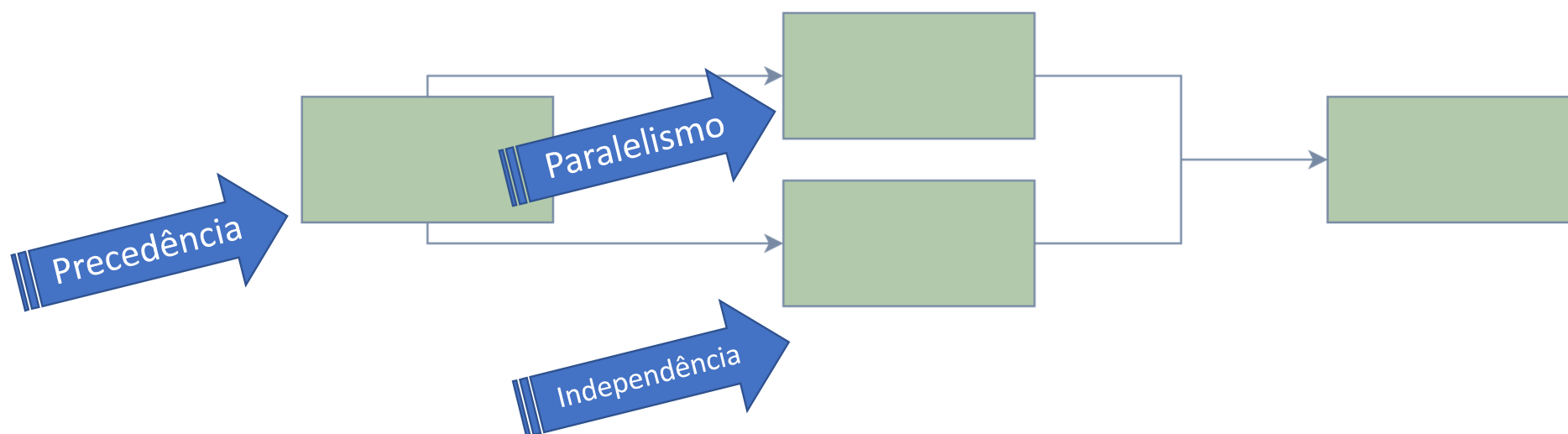
    @task()
    def airflow():
        print("airflow")

    # Set dependencies between tasks
    hello >> airflow()
```





DAG: Pipeline



Pipeline

Erros

Log/Auditoria

Monitoramento/Alertas

Recuperação a partir de um Ponto

Dados Históricos/ Diferenciação

Alta disponibilidade

Distribuição de Carga