

---

# Airflow



1.

# Introdução

O que é e o que não é o Airflow?



*Airflow é uma ferramenta de construção,  
orquestração e gerenciamento de  
“workflows”*

# O que o Airflow não é

- ▷ Biblioteca de processamento e manipulação de dados. Ex: Pandas, Spark, Dask
- ▷ Ferramenta de monitoramento

# Ferramentas semelhantes

- ▷ Scripts agendados no crontab
- ▷ Ferramentas de ETL: Pentaho, Knime etc
- ▷ Luigi (Spotify)

# Porquê usar o Airflow?

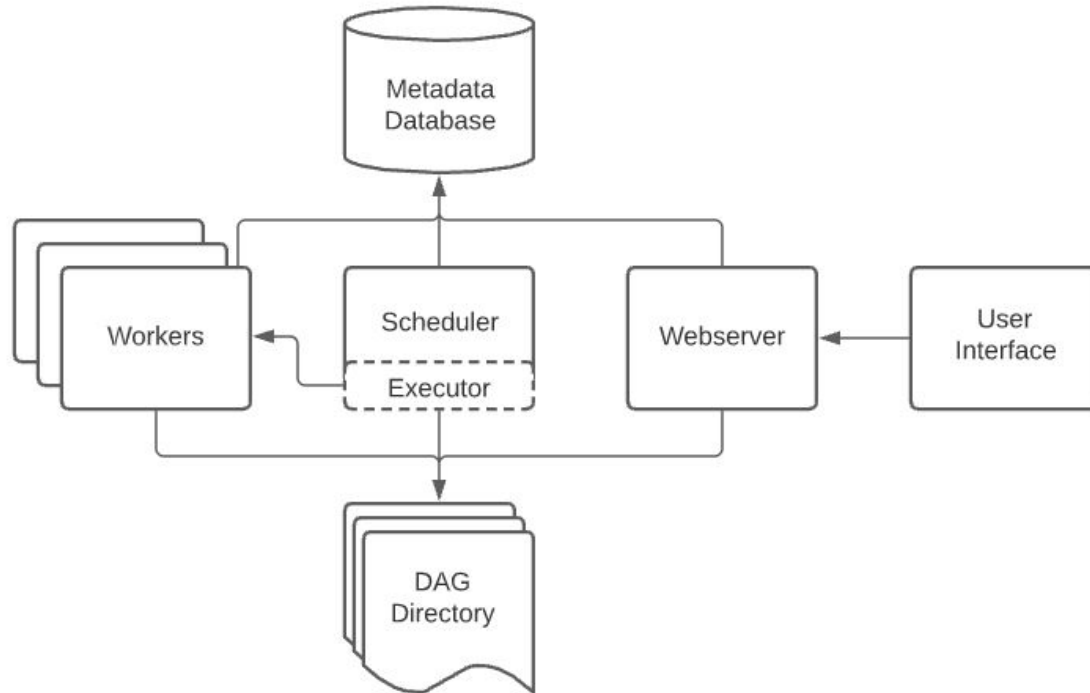
- ▷ Open Source (Apache foundation)
- ▷ Escalável
- ▷ Ferramenta com a maior comunidade
- ▷ Excelente documentação
- ▷ Escrito em Python

2.

# Arquitetura do Airflow

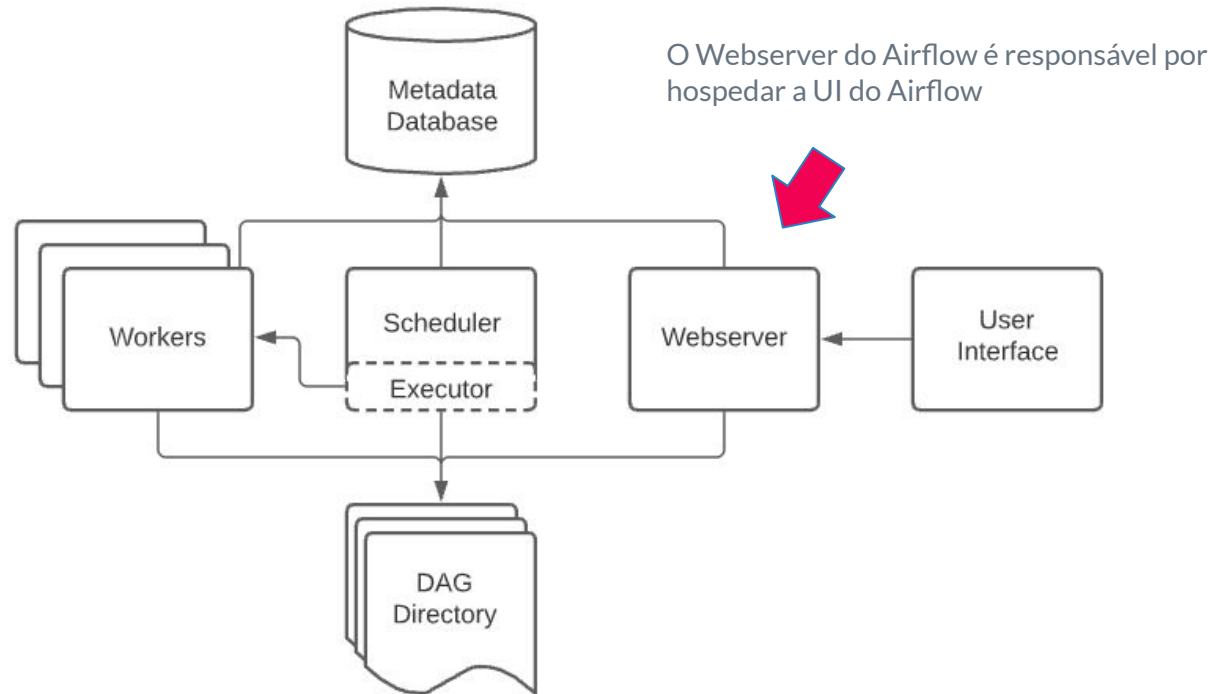
Como o Airflow funciona?

# Arquitetura do Airflow

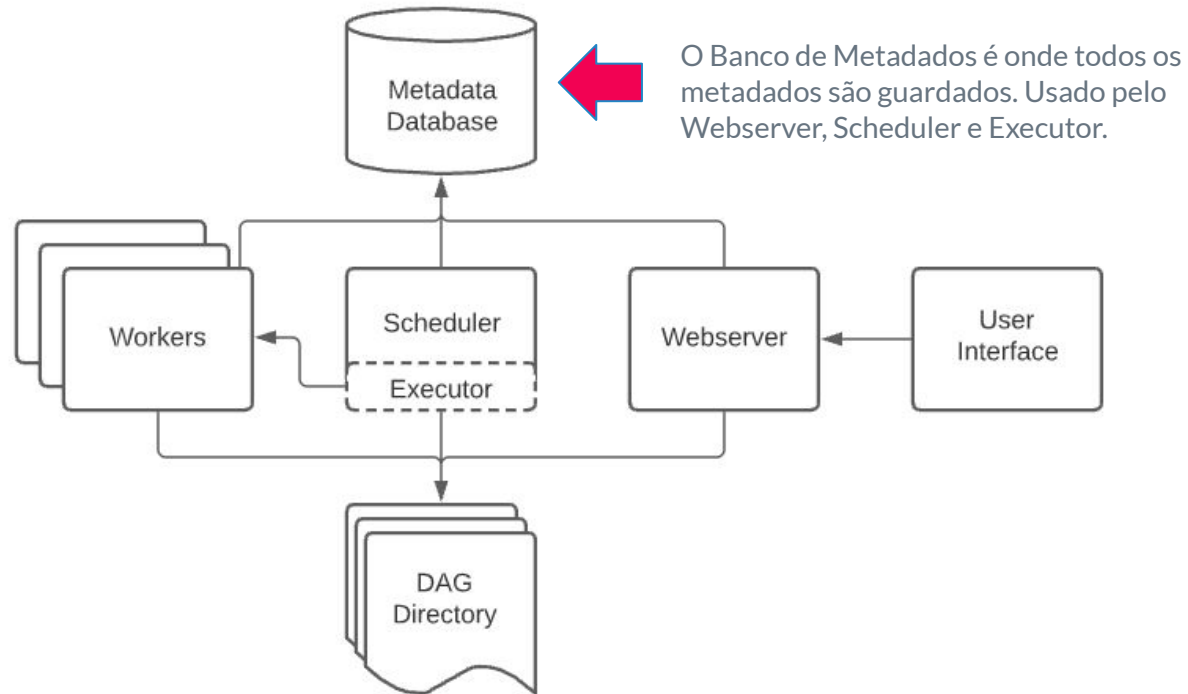




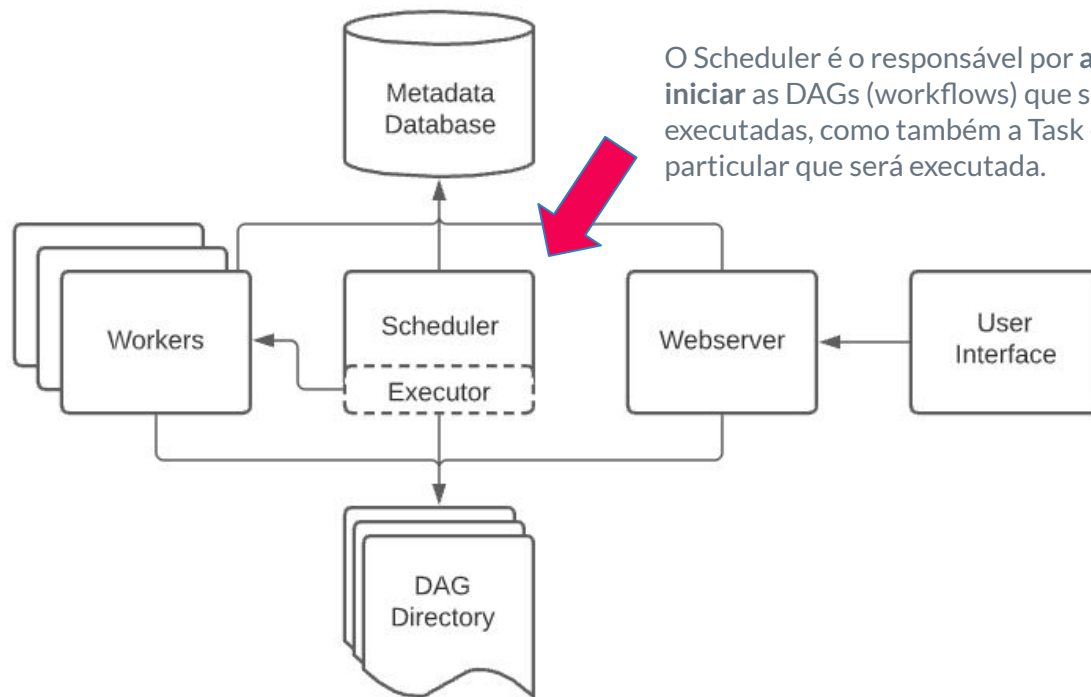
# Arquitetura do Airflow



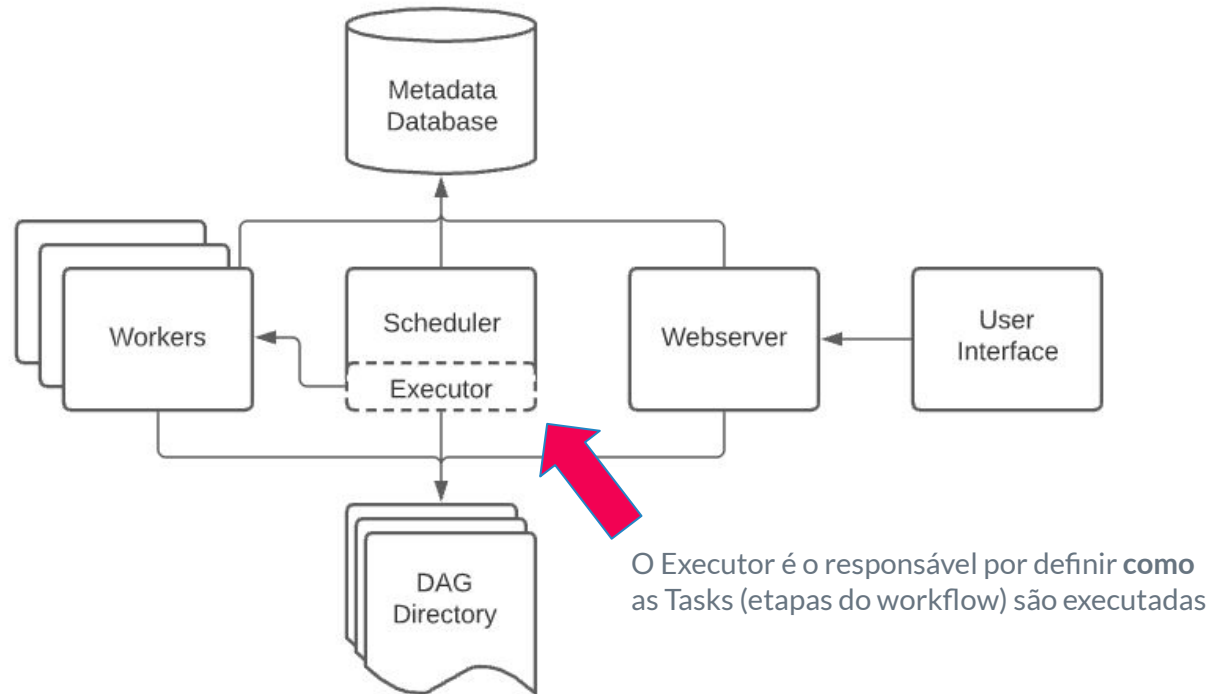
# Arquitetura do Airflow



# Arquitetura do Airflow

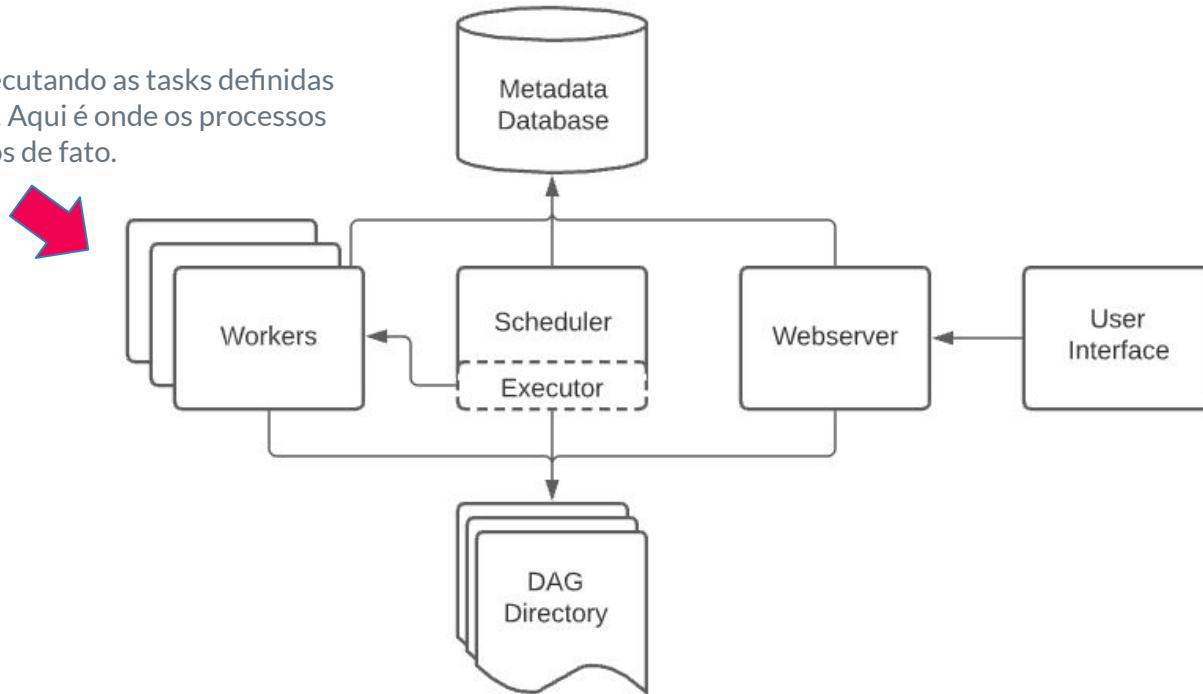


# Arquitetura do Airflow

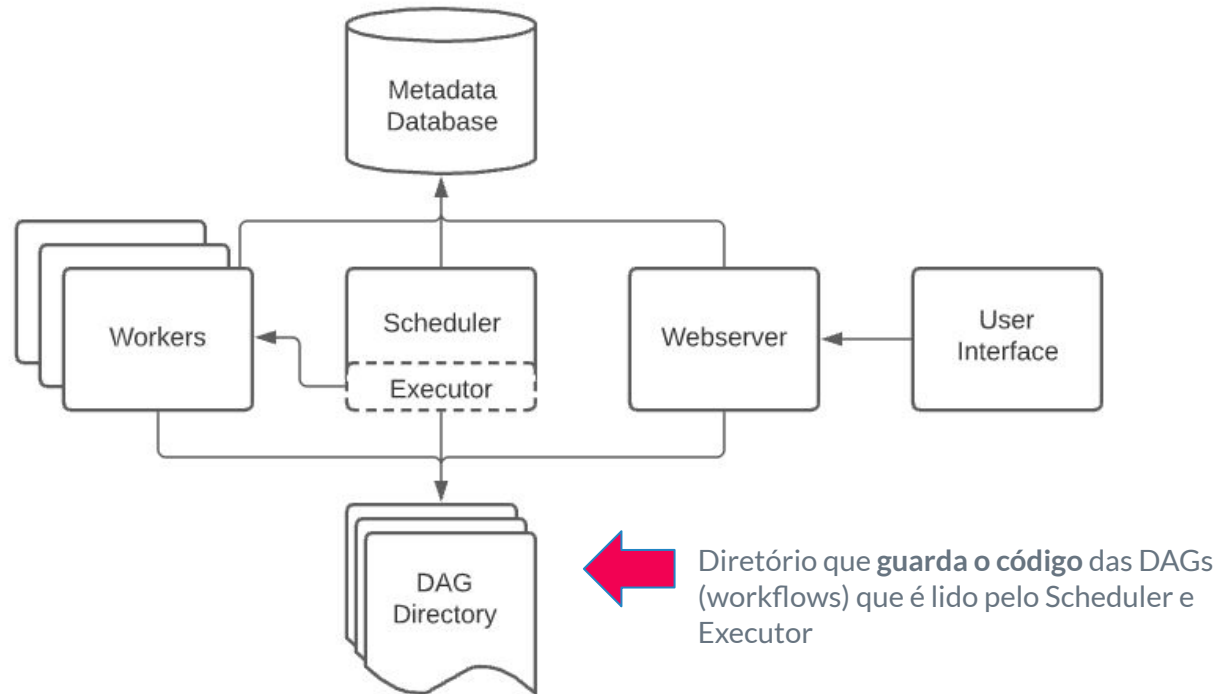


# Arquitetura do Airflow

**Processos** executando as tasks definidas pelo Executor. Aqui é onde os processos são executados de fato.



# Arquitetura do Airflow

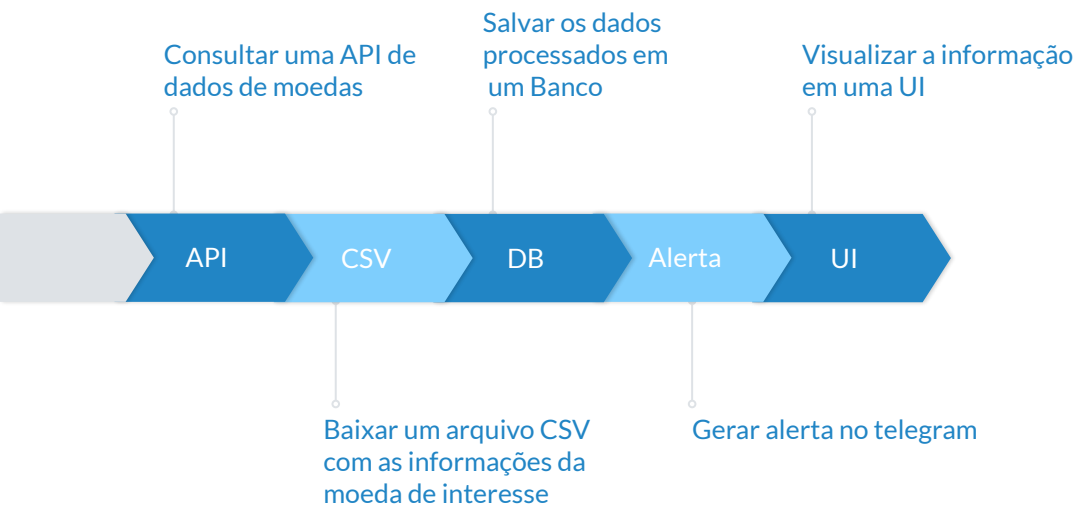


3.

# Live demo

O Airflow na prática

# O fluxo que faremos no Airflow





Obrigado!

**Alguma pergunta?**