

Airflow no Kubernetes Por onde começar?

Neylson Crepalde

Agosto, 2021



THE
DEVELOPER'S
CONFERENCE



A3Data

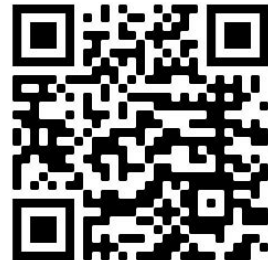
Prazer! Eu sou o Ney.

CDO na A3Data

Prof. no IGTI e na PUC Minas

Doutor em Sociologia Econômica

Viciado em cubo mágico!!!



Agenda

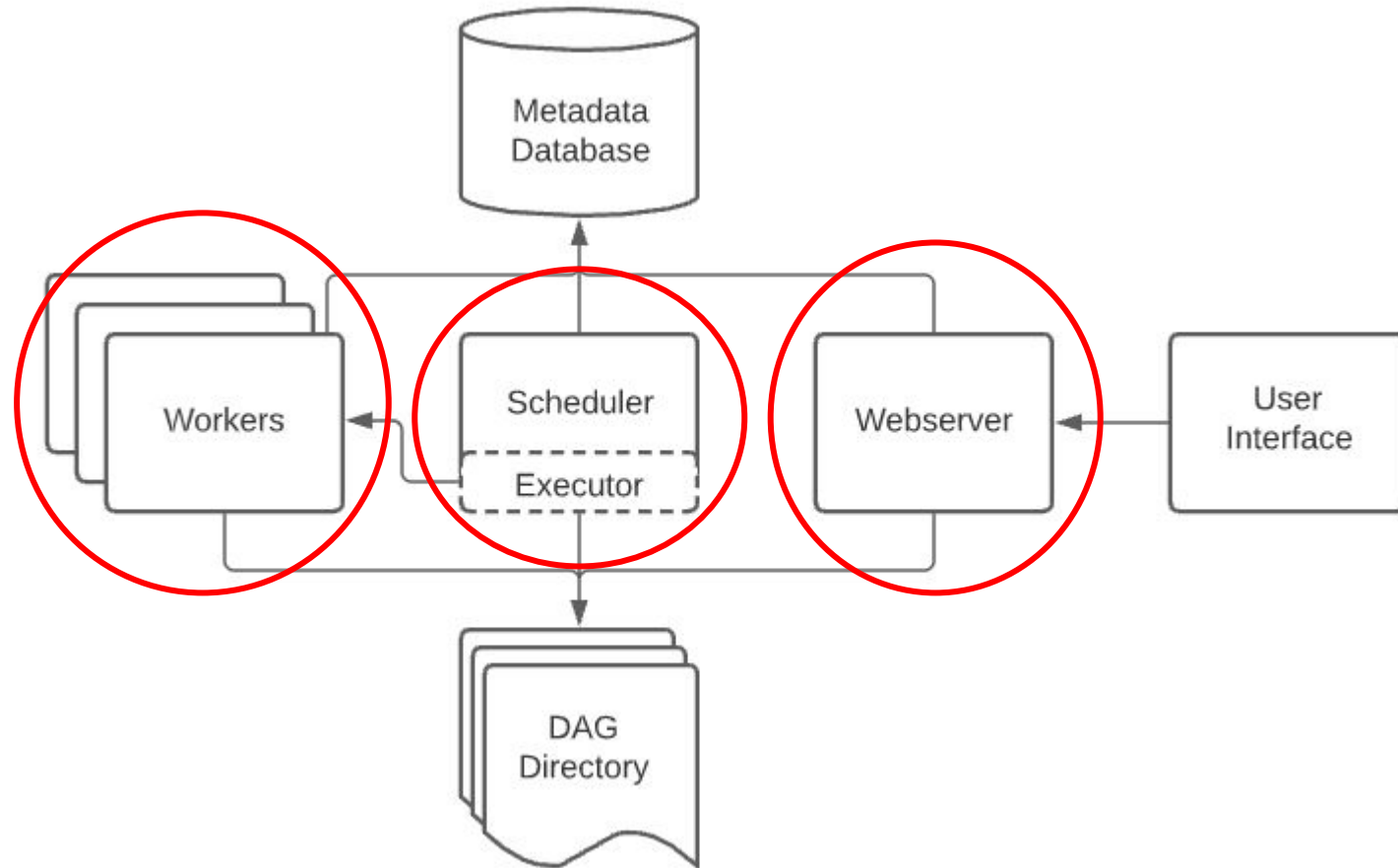
- 1** Airflow
 - 2** Kubernetes
 - 3** Uma estratégia de deploy
 - 4** Algumas escolhas
 - 5** Considerações finais
-

- Ferramenta favorita dos times de dados para orquestração de pipelines!
- Escalável
- Dinâmico
- Extensível
- Elegante



Apache
Airflow

Arquitetura





kubernetes

- Plataforma open source portátil e extensível para gerenciar workloads em containers.
- Automação
- Configuração declarativa
- Nos últimos anos, a tecnologia tem sido utilizada com muito sucesso em projetos de dados

E o Deploy?

- Existem algumas estratégias de deploy em k8s. Para o Airflow, escolhemos usar helm.
- Package manager que simplifica o deploy de aplicações complexas em k8s.
- Há vários helm Charts disponíveis para o Airflow (Bitnami, community, Astronomer...)
- Vamos trabalhar com o helm Chart Oficial



myvalues.yml

```
# Airflow home directory
# Used for mount paths
airflowHome: /opt/airflow

# Default airflow repository -- overrides all t
defaultAirflowRepository: apache/airflow

# Default airflow tag to deploy
defaultAirflowTag: "2.1.2"

# Airflow version (Used to make some decisions
airflowVersion: "2.1.2"

# Images
```

helm install

```
kubectl create namespace airflow

helm repo add apache-airflow \
  https://airflow.apache.org

helm install airflow \
  apache-airflow/airflow \
  -n airflow -f myvalues.yml
```


Exemplo prático

Considerações finais

- Deploy com *Helm* e dags no Git podem facilitar bastante a vida
- É importante escolher o Executor de acordo com o tipo de workloads
- Em produção, será importante trabalhar com alguma ferramenta para gerenciar o deploy e versionamento do Airflow como o Argo CD



empowering
people



A3Data