

DATA CHALLENGE STONE

DESAFIO PLATAFORMA

Quem sou eu?



- Afonso Rodrigues
- Tenho 28 anos
- Sou de BH/MG
- DevOps/SRE e DBA



linkedin.com/in/afonsoavr



github.com/afonsoaugusto

MOTIVAÇÃO

A **stone** dispõe de inúmeras
fontes de dados e
possibilidades

Assim ela tem um time de
inteligência de Dados em
crescimento

É necessário ter
ferramentas que
possibilitem o
processamento desses dados
em grande escala



O Spark é o indicado para
processamento de grandes
volumes de dados



E é necessário ter uma
plataforma de execução que
seja escalável



kubernetes

Kubernetes é uma plataforma
portátil e extensiva para o
gerenciamento de cargas de
trabalho e serviços
distribuídos em contêineres



kubernetes

DeSaFIO

Spark

Simplificar a
operação desta
combinação

+



kubernetes

Algo que seja
modular e escalável
(ambiente e pessoas)

Uma solução que provê uma plataforma
de **orquestração** de workflows

PROPOSTA



+



kubernetes

Spark

+



kubernetes



Apache
Airflow

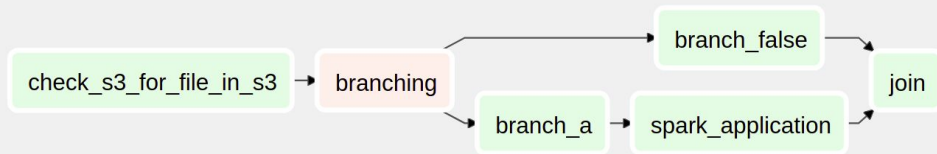


Apache
Airflow

Airflow é uma
plataforma para
criar, programar e
monitorar workflows
de maneira
programática



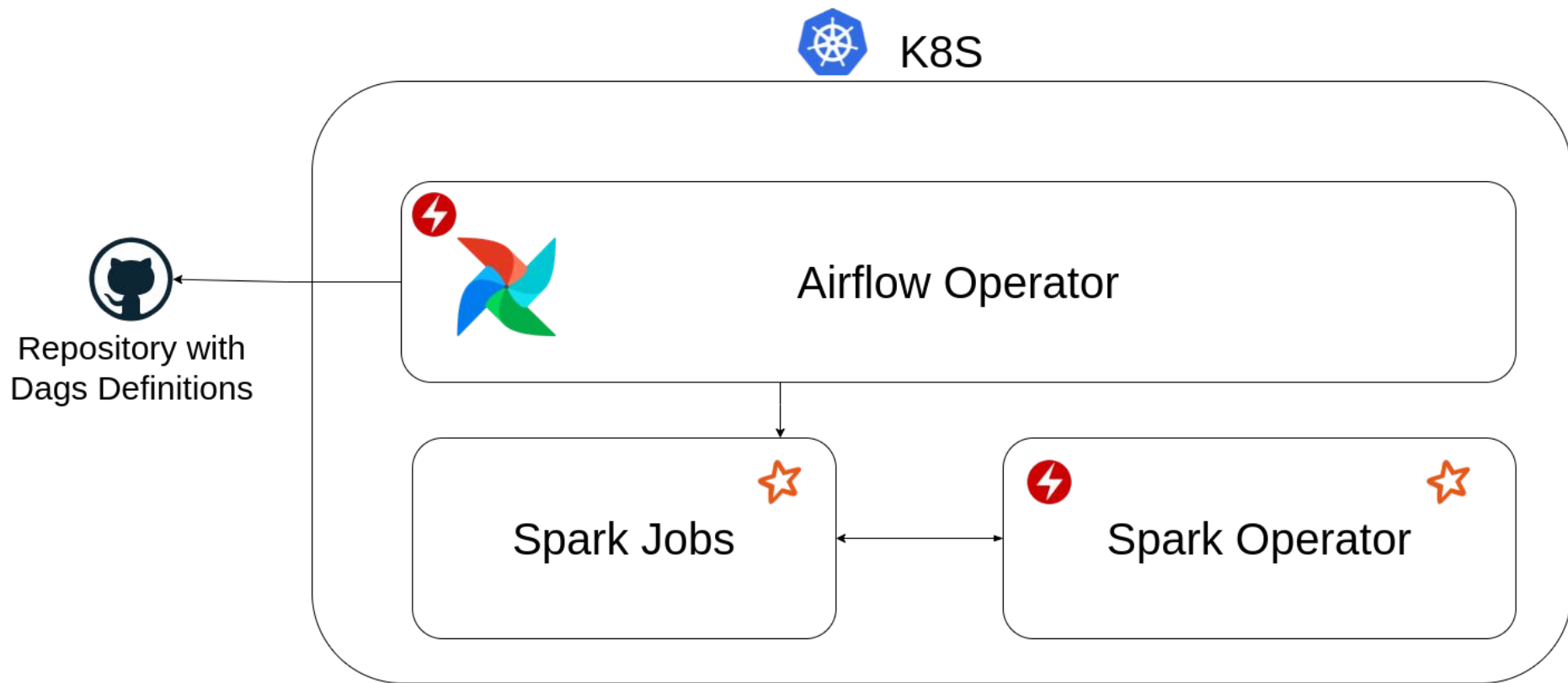
Apache
Airflow





Apache
Airflow

- Escalável
- Dinâmico
- Extensível
- Integrável
- Aberto



ESTRUTURA DA SOLUÇÃO

Contextos de execução



AUTOMAÇÃO
DO CLUSTER

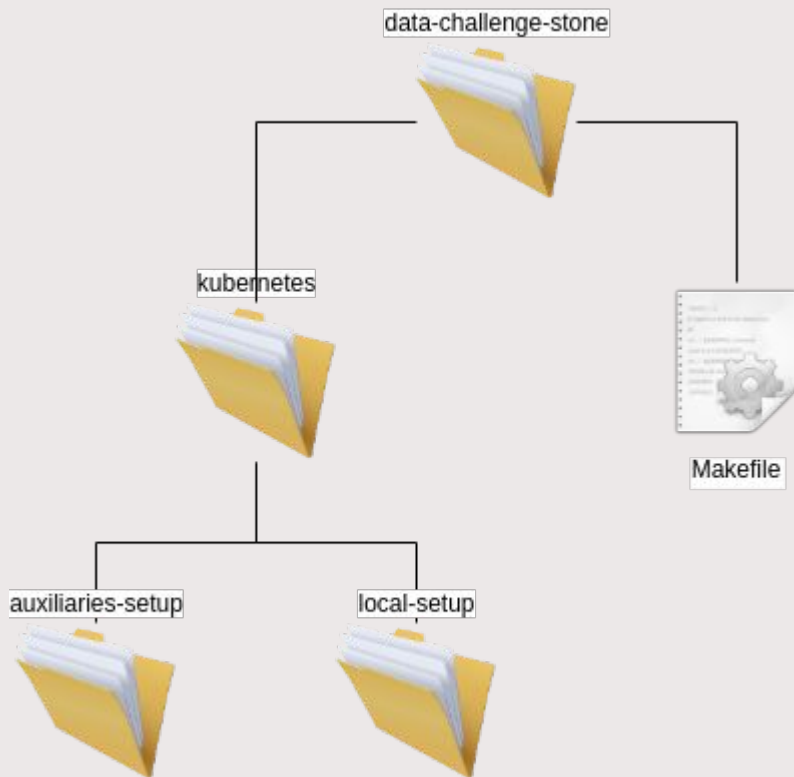


RECURSOS DO
KUBERNETES



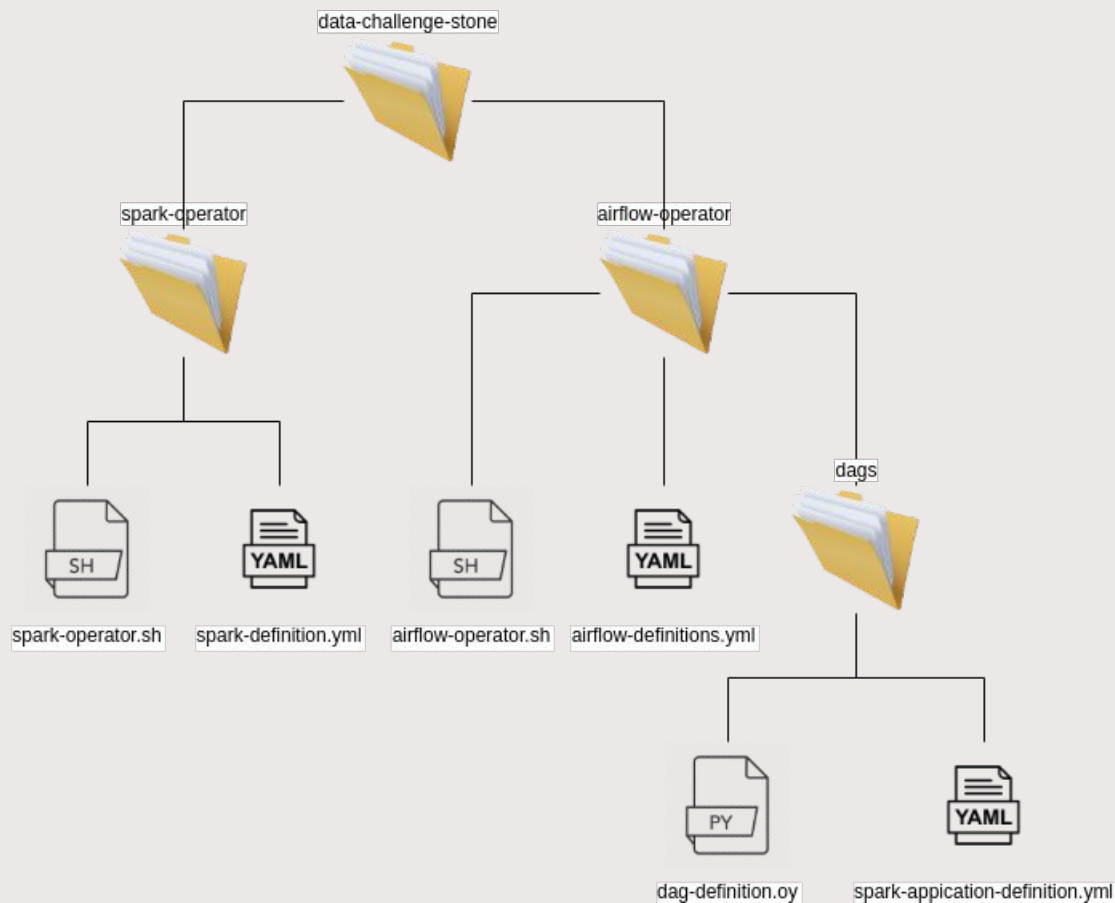
APLICAÇÕES
SPARK E
AIRFLOW

AUTOMAÇÃO DO CLUSTER (Local)



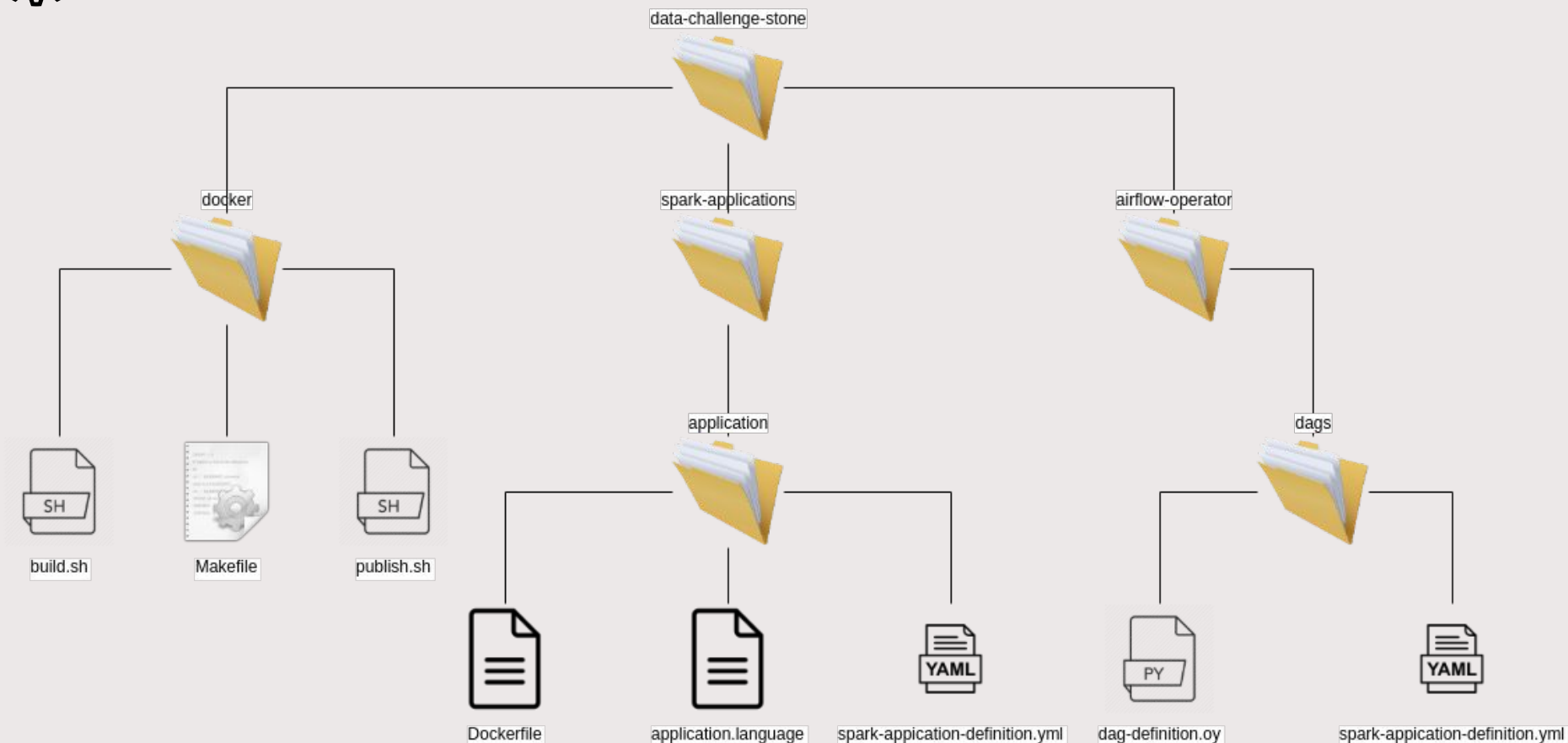


RECURSOS DO KUBERNETES (Operators)





APLICAÇÕES DO SPARK E AIRFLOW



ROTEIRO DE IMPLANTAÇÃO

Execução

simples

make **setup**

Adicione ou remova

ações

```
make setup-k8s setup-auxiliaries
```


Execute o mesmo comando no
contexto local ou no CI

```
version: 2.1

jobs:
  build:
    docker:
      - image: afonsoaugusto/base-ci:latest
    steps:
      - checkout
      - setup_remote_docker
      - run:
          name: build
          command: |
            make build
      - run:
          name: publish
          command: |
            make publish

workflows:
  workflow:
    jobs:
      - build:
          context: CI
```

Makefile

+

Helm

+

Operators

Helm

+

Operators

+

Makefile

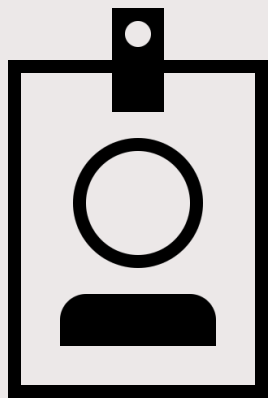
```
dashboard: setup-k8s-dashboard k8s-proxy-dashboard ## setup-k8s-dashboard + k8s-proxy-dashboard
setup-local: setup-local-install-binaries setup-local-init-kind ## Setup environment to running k8s locally - Install binaries and initi k8s kind
setup-k8s: setup-k8s-spark-operator setup-k8s-airflow-operator ## Setup k8s - spark-operator -> airflow-operator
setup-auxiliaries: setup-auxiliaries-minio-operator ## Setup auxiliaries - Minio Operator
setup: setup-local setup-k8s setup-auxiliaries ## Setup environment - setup-local -> setup-k8s -> setup-auxiliaries
clean: setup-local-cleanup ## alias for setup-local-cleanup

build:
    $(MAKE) -C docker/ docker-build
publish:
    $(MAKE) -C docker/ docker-publish
```

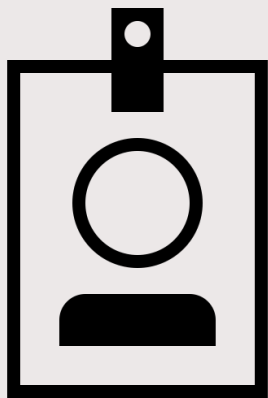
segurança

O **Airflow** provê o controle de autenticação e usuários necessários

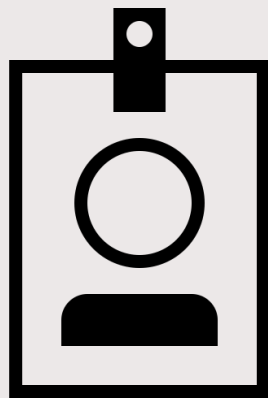
AIRFLOW ROLES



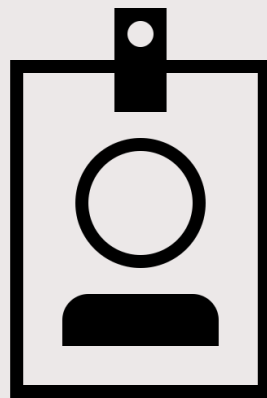
Admin



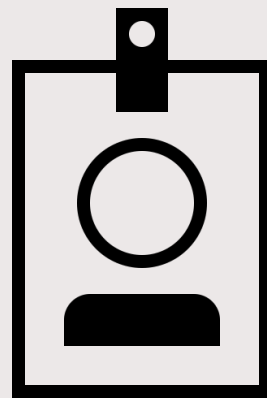
Public



Viewer



User



Op

Demo

Referências

- * [kubectl/cheatsheet] (<https://kubernetes.io/pt/docs/reference/kubectl/cheatsheet/>)
- * [Kind] (<https://kind.sigs.k8s.io/>)
- * [Helm] (<https://helm.sh/docs/>)
- * [spark-on-k8s-operator] (<https://github.com/GoogleCloudPlatform/spark-on-k8s-operator/>)
- * [Airflow-operator] (<https://github.com/bitnami/charts/tree/master/bitnami/airflow>)

OBRIGADO