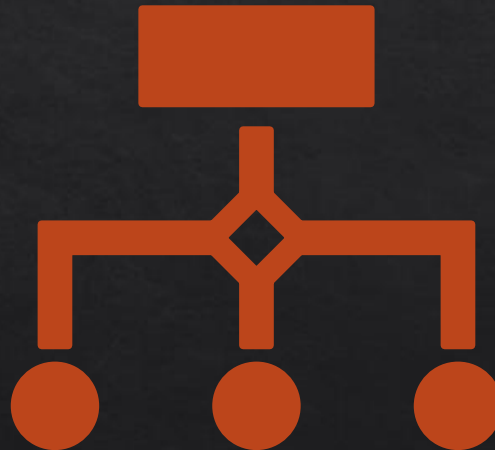


# Formação Engenheiro de Dados

Outros Sistemas Hadoop

# Oozie: Workflow e Agendamento

- ◇ Alguns atividades previsão ser executadas em sequência, de forma coordenada

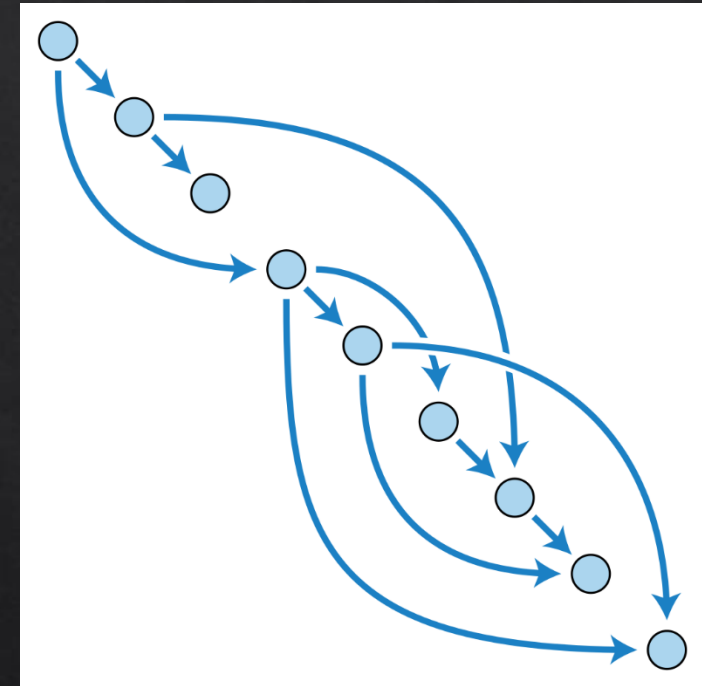


# OOZIE

- ◇ Oozie Server:
  - ◇ Workflow: definição de processo
  - ◇ Coordinator: coordena a execução
  - ◇ Bundles: Conjunto de aplicações coordenadas
  - ◇ SLA: Service Level Agreement

# Workflow

- ◆ DAG: Directed Acyclic Graph
- ◆ Grafo direcionado, em direção única (sem loops)
- ◆ Hadoop Process Definition Language (hPDL), com estrutura XML



# Workflow

- ◊ Workflow: conjunto de ações e transições ordenados
- ◊ Action: Tarefa do Workflow, por exemplo, batch mapreduce
  - ◊ Synchronous
  - ◊ Asynchrhronous
- ◊ Transition: Descreve a transição entre as ações
  - ◊ Structural conditions: definidos estaticamente, como transitions e fork-join
  - ◊ Runtime conditions:: definidos em tempo de execução, como erros



# Workflow: nós de controle

0.0 Start: Início



End: Fim



Decision: Decisão



Fork: Divisão para execução em paralelo



Join: Aguarda ações concorrentes encerrarem para continuar



Sub-Workflow: executa um outro workflow



Kill: encerra o processo

# Ações

- ◊ Java: executa o método main() de uma classe java
- ◊ Pig: executa um script Pig
- ◊ MapReduce: Executa uma tarefa mapreduce
- ◊ Hive: Executa uma tarefa Hive
- ◊ Sqoop: Executa um job do Sqoop
- ◊ Shell: Executa um comando no Shell do SO
- ◊ Email: Envia e-mail

# Recuperação

- ◇ Parâmetros
  - ◇ Máximo de tentativas
  - ◇ Intervalos de tentativas



# Coordinator

- ◊ Executa um Workflow
- ◊ Pode definir dependências entre workflows
  - ◊ Nominal time: tempo em que o Job deveria começar
  - ◊ Actual time: tempo em que o Job de fato iniciou

# Ciclo de vida de um Job

## ◇ Estados:

- ◇ Prep (carregado)
- ◇ Running
- ◇ Killed
- ◇ Suspended
- ◇ Succeeded
- ◇ Failed

# SLA - Service Level Agreement

Permite definir,  
registrar e monitorar  
a SLA dos processos  
do Oozie

# Comandos

## ◆ Enviar Job

```
sudo oozie job -oozie http://localhost:11000/oozie -config  
examples/apps/map-reduce/job.properties -run
```

Job ID : 00000000-190727045839257-oozie-oozi-W

## ◆ Verificar o Status

```
oozie job -oozie http://localhost:11000/oozie -info 00000000-  
190727045839257-oozie-oozi-W
```

## ◆ Suspende

```
oozie job -oozie http://localhost:11000/oozie -suspend 00000000-  
190727045839257-oozie-oozi-W
```

# Verificar o Status no Navegador

Oozie Web Console

quickstart.cloudera:11000/oozie/

Cloudera Hue Hadoop HBase Impala Spark Solr Oozie Cloudera Manager Getting Started

Documentation

Oozie Web Console

Workflow Jobs Coordinator Jobs Bundle Jobs SLA System Info Instrumentation Metrics Settings

All Jobs Active Jobs Done Jobs Custom Filter

	Job Id	Name	Status	Run	User	Group	Created	Started	Last Modified	Ended
1	0000012-190727045839257-oozie-oozi-W	streaming-wf	RUNNING	0	root		Sat, 27 Jul 2019 13:34:11 GMT	Sat, 27 Jul 2019 13:34:11 GMT	Sat, 27 Jul 2019 13:34:12 GMT	
2	0000011-190727045839257-oozie-oozi-W	scoop-wf	RUNNING	0	root		Sat, 27 Jul 2019 13:34:02 GMT	Sat, 27 Jul 2019 13:34:02 GMT	Sat, 27 Jul 2019 13:34:02 GMT	
3	0000010-190727045839257-oozie-oozi-W	one-op-wf	RUNNING	0	root		Sat, 27 Jul 2019 13:33:51 GMT	Sat, 27 Jul 2019 13:33:51 GMT	Sat, 27 Jul 2019 13:33:52 GMT	
4	0000008-190727045839257-oozie-oozi-W	no-op-wf	RUNNING	0	root		Sat, 27 Jul 2019 13:33:40 GMT	Sat, 27 Jul 2019 13:33:40 GMT	Sat, 27 Jul 2019 13:33:40 GMT	
5	0000007-190727045839257-oozie-oozi-W	datelist-java-mai...	RUNNING	0	root		Sat, 27 Jul 2019 13:33:25 GMT	Sat, 27 Jul 2019 13:33:25 GMT	Sat, 27 Jul 2019 13:33:25 GMT	
6	0000006-190727045839257-oozie-oozi-W	aggregator-wf	RUNNING	0	root		Sat, 27 Jul 2019 13:32:57 GMT	Sat, 27 Jul 2019 13:32:57 GMT	Sat, 27 Jul 2019 13:32:57 GMT	
7	0000003-190727045839257-oozie-oozi-W	SparkFileCopy	RUNNING	0	root		Sat, 27 Jul 2019 13:32:41 GMT	Sat, 27 Jul 2019 13:32:41 GMT	Sat, 27 Jul 2019 13:32:41 GMT	
8	0000002-190727045839257-oozie-oozi-W	hive2-wf	RUNNING	0	root		Sat, 27 Jul 2019 13:23:16 GMT	Sat, 27 Jul 2019 13:23:16 GMT	Sat, 27 Jul 2019 13:23:16 GMT	
9	0000001-190727045839257-oozie-oozi-W	pig-wf	RUNNING	0	root		Sat, 27 Jul 2019 13:21:49 GMT	Sat, 27 Jul 2019 13:21:49 GMT	Sat, 27 Jul 2019 13:21:49 GMT	

Page 1 of 1

Job Info Job Definition Job Configuration Job Log Job DAG

Job Id: 0000000-190727045839257-oozie-oozi-W

Name: map-reduce-wf

App Path: hdfs://localhost:8020/user/root/examples/apps/map-reduce/workflow.xml

Run: 0

Status: RUNNING

User: root

Group:

Parent Coord:

Create Time: Sat, 27 Jul 2019 12:39:05 GMT

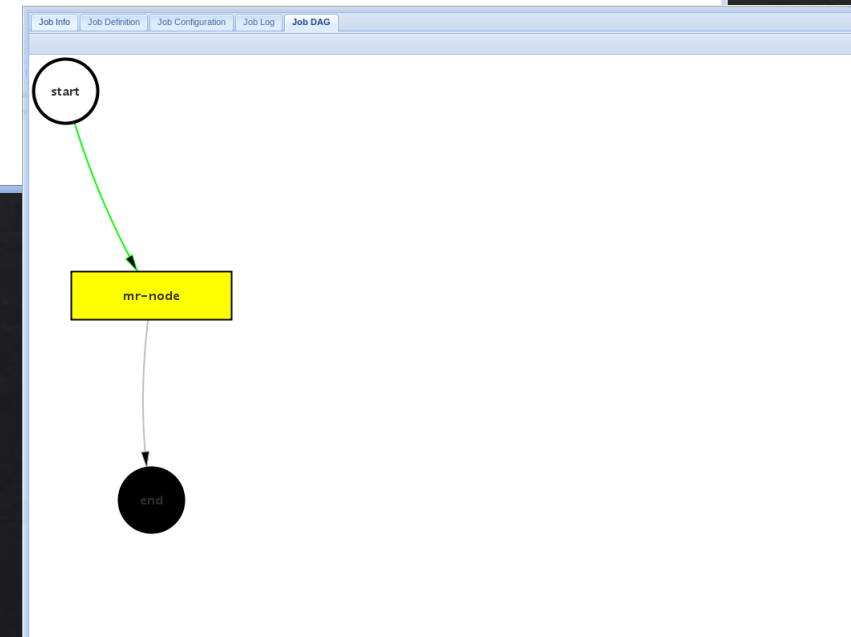
Start Time: Sat, 27 Jul 2019 12:39:06 GMT

Last Modified: Sat, 27 Jul 2019 12:59:17 GMT

End Time:

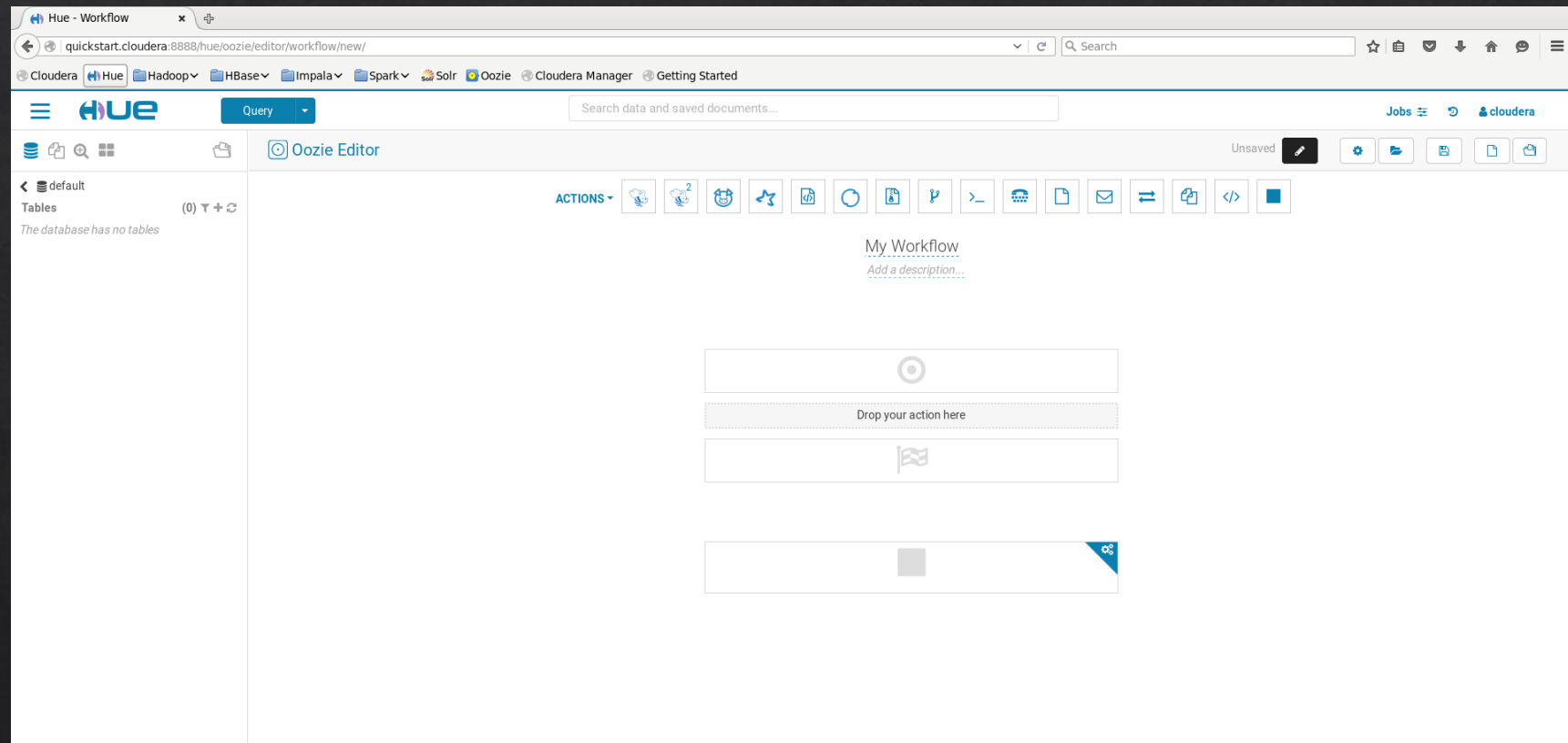
Actions

Action Id	Name	Type	Status	Transition	StartTime	EndTime
1 0000000-190727045839257-oozie-oozi-W@start	start	:START:	OK	mr-node	Sat, 27 Jul 2019 12:39:06 GMT	Sat, 27 Jul 2019 12:39:06 GMT
2 0000000-190727045839257-oozie-oozi-W@mr-node	mr-node	map-reduce	START_RETRY		Sat, 27 Jul 2019 12:39:06 GMT	





# Ferramentas Visuais - Hue



# Plug-in para Eclipse

