Oozie

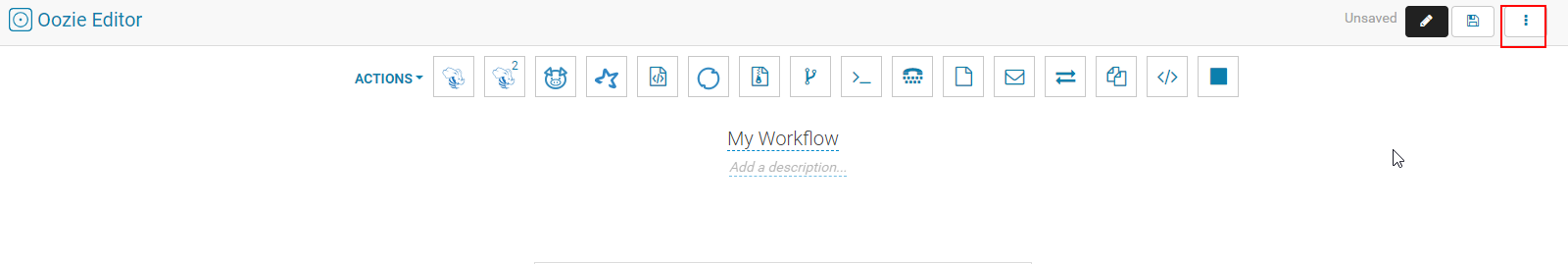
https://www.cloudera.com/documentation/enterprise/latest/topics/cdh\_oozie\_sqoop\_jdbc.html

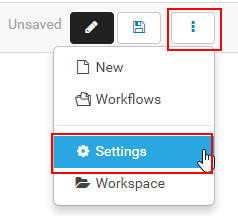
https://oozie.apache.org/docs/4.3.1/DG\_ActionAuthentication.html

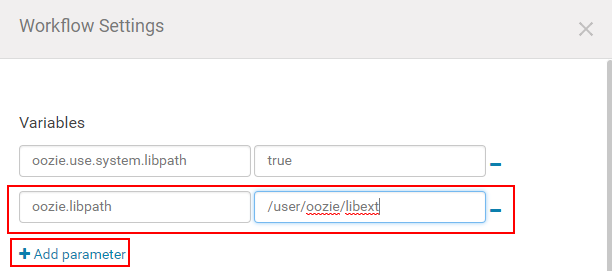
# Workflow com Sqoop

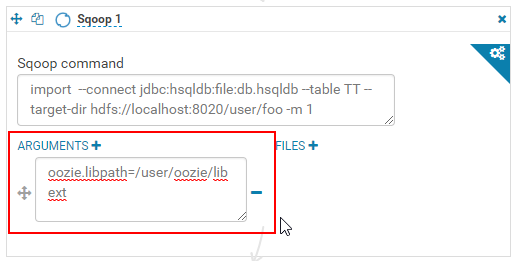
Para rodar sqoop import/export com bibliotecas externas seguir o procedimento a seguir:

1. Copiar o JARs e trocar as permissões
2. Entrar na pasta onde está o JAR do driver
   * cd /var/lib/sqoop
3. Criar a pasta no HDFS
   * hdfs dfs -mkdir /user/oozie/libext
4. Trocar o owner e grupo da pasta
   * hdfs dfs -chown oozie:oozie /user/oozie/libext
5. Copiar os JARS para a pasta no HDFS
   * hdfs dfs -put \*.jar /user/oozie/libext
6. Trocar o owner e o grupo da pasta
   * hdfs dfs -chown oozie:oozie /user/oozie/libext/\*.jar
7. Modificar as permissões dos JARs
   * hdfs dfs -chmod 755 /user/oozie/libext/\*.jar
8. Listar a pasta
   * hdfs dfs -ls /user/oozie/libext
9. No Workflow





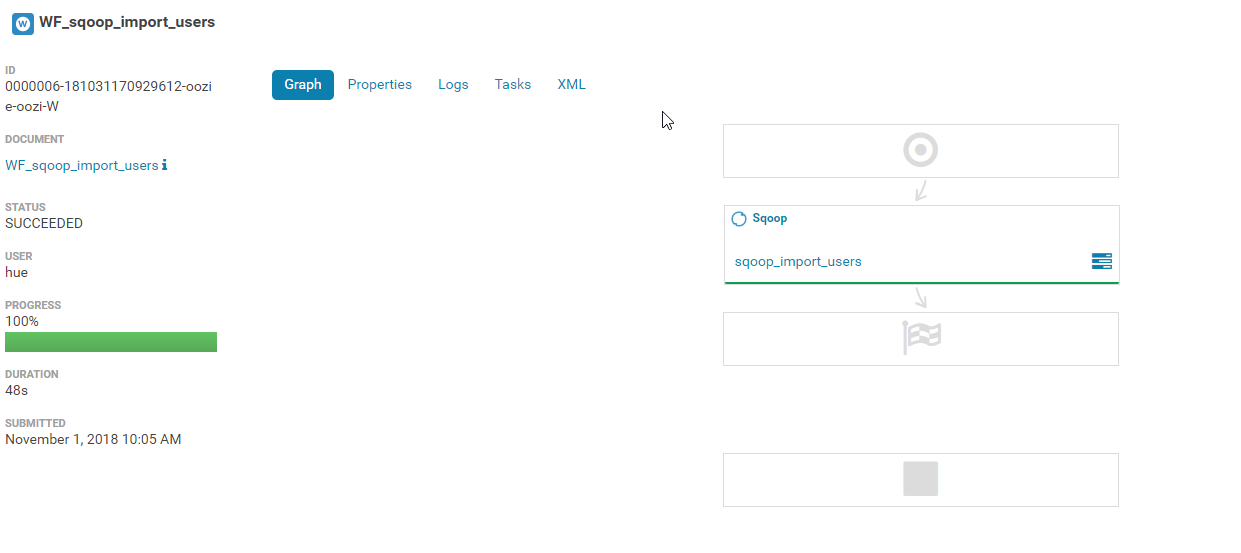




No workflow em Settings adicionar as seguintes variavéis:

- oozie.use.system.libpath=true

- oozie.libpath=/user/oozie/libext



## Outros parâmetros

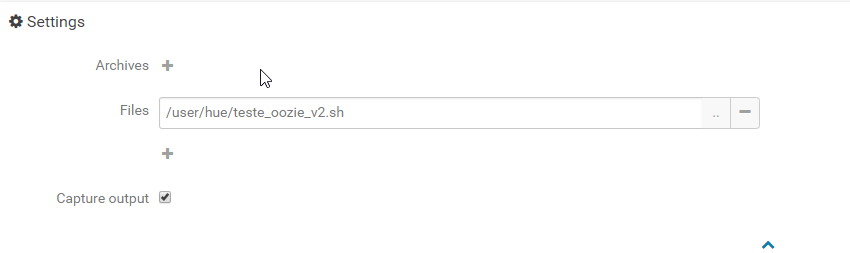
yarn.nodemanager.resource.memory-mb = 2GB

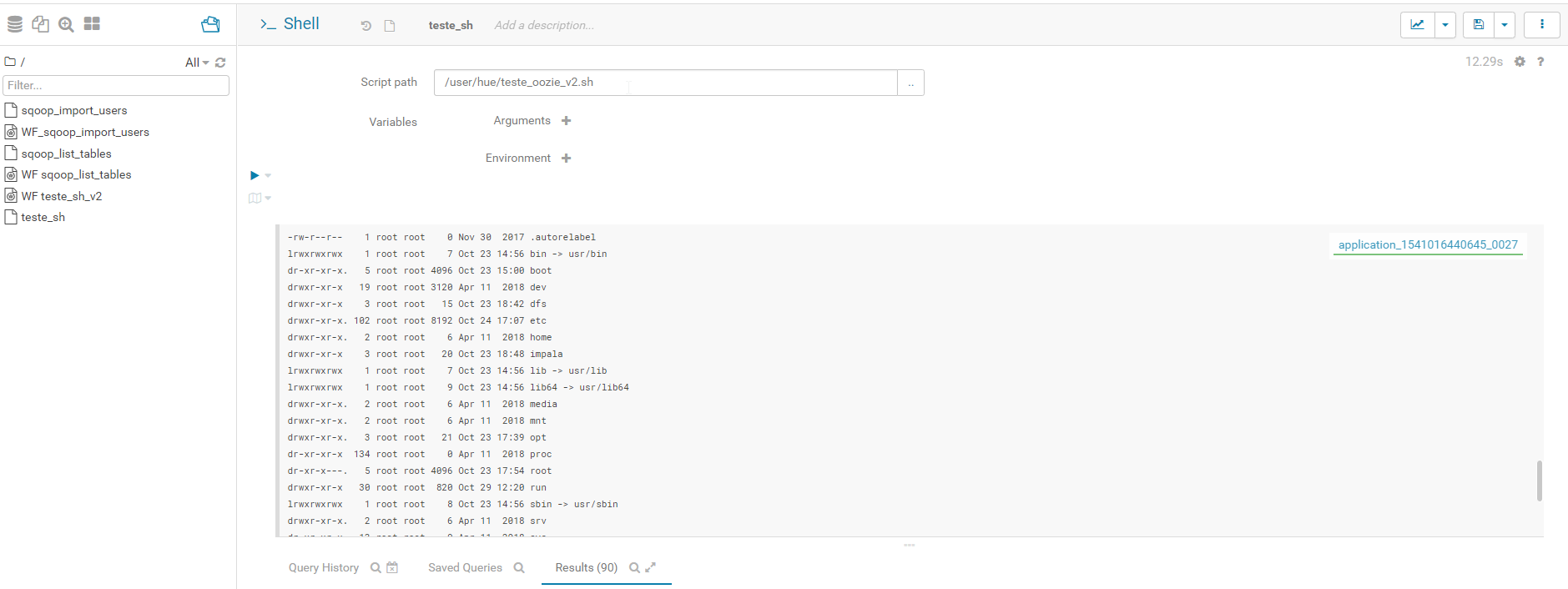
mapreduce.job.user.name = hdfs

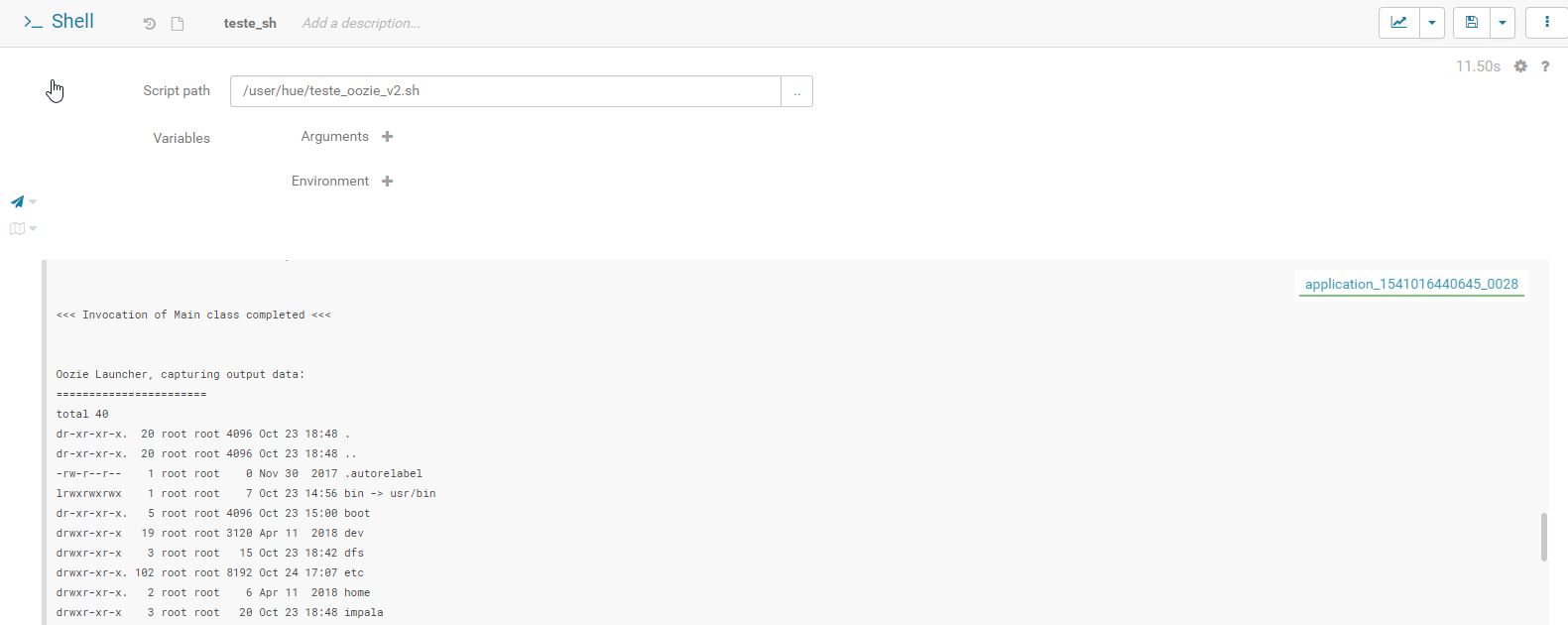
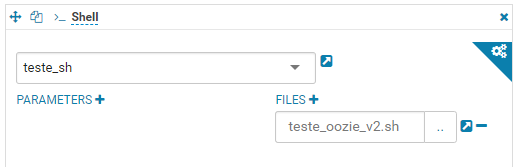
user.name = hdfs

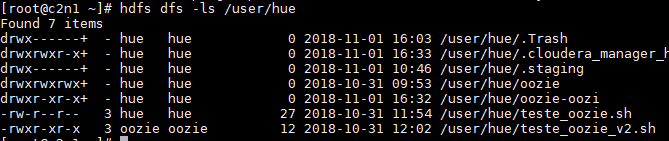
## Workflow para rodar Shell Script



Apontar 





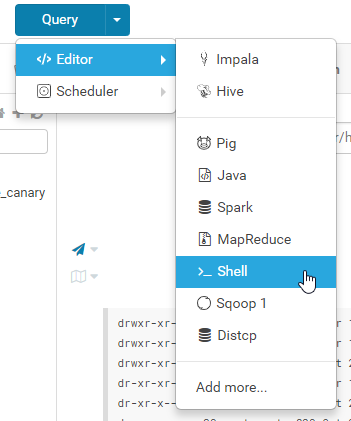


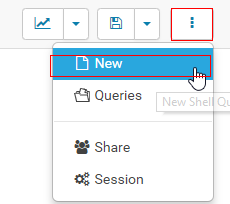
Exemplo:



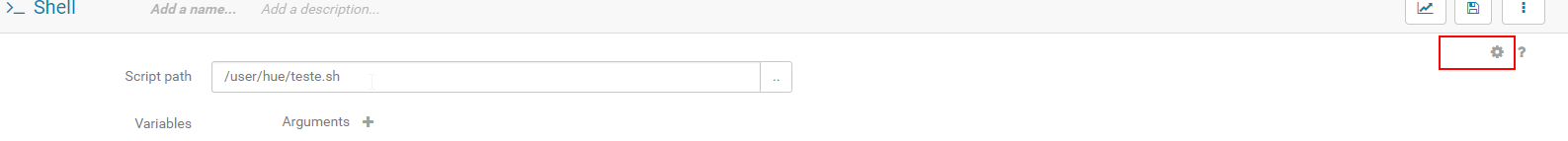
Selecione o arquivo na máquina local

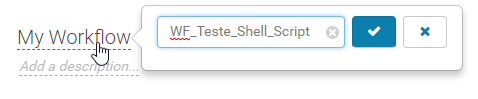
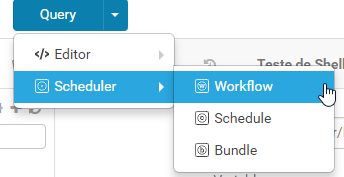
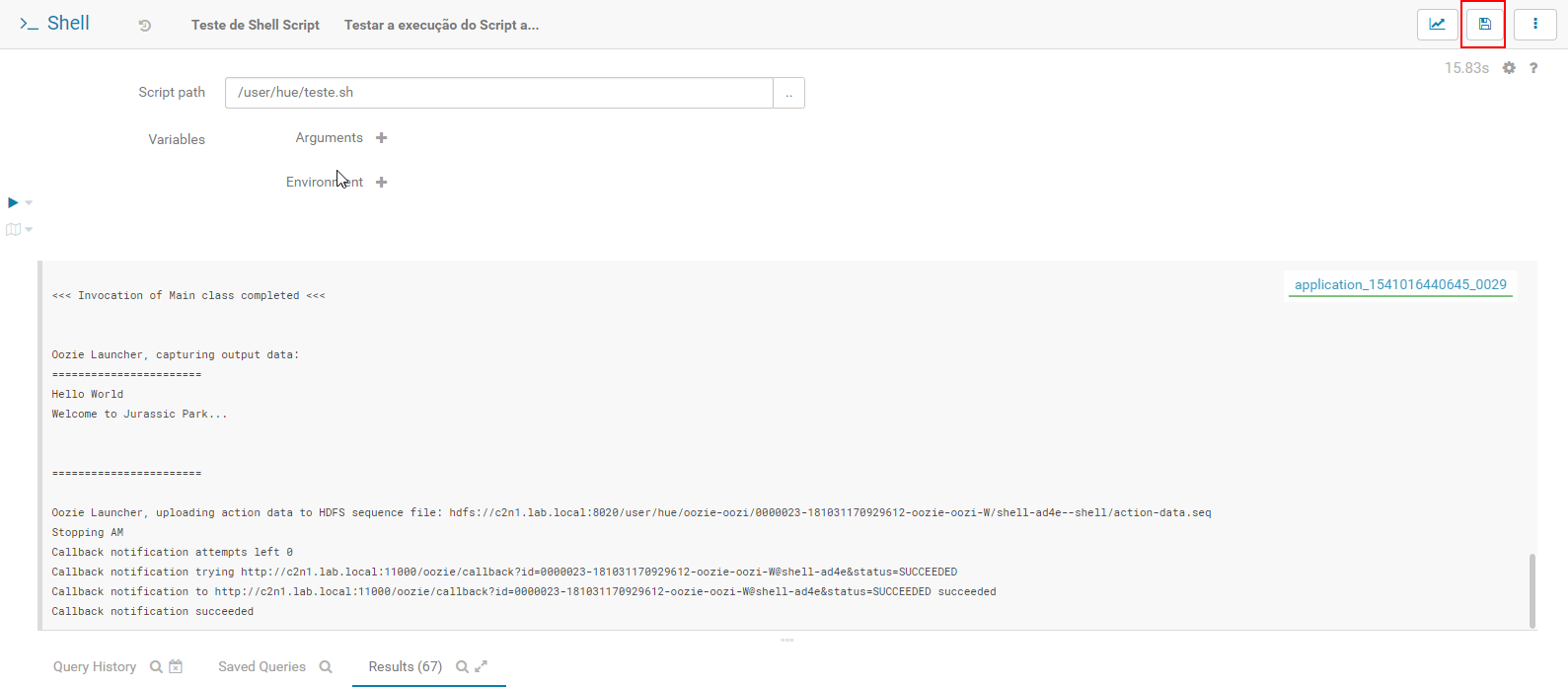
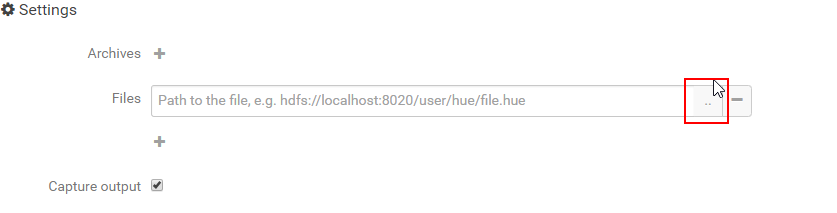




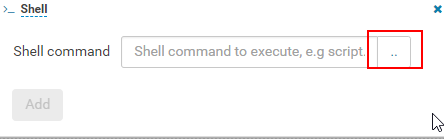




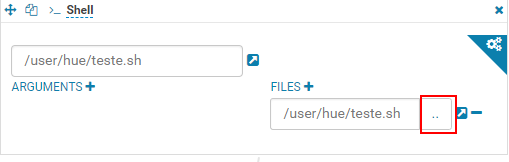


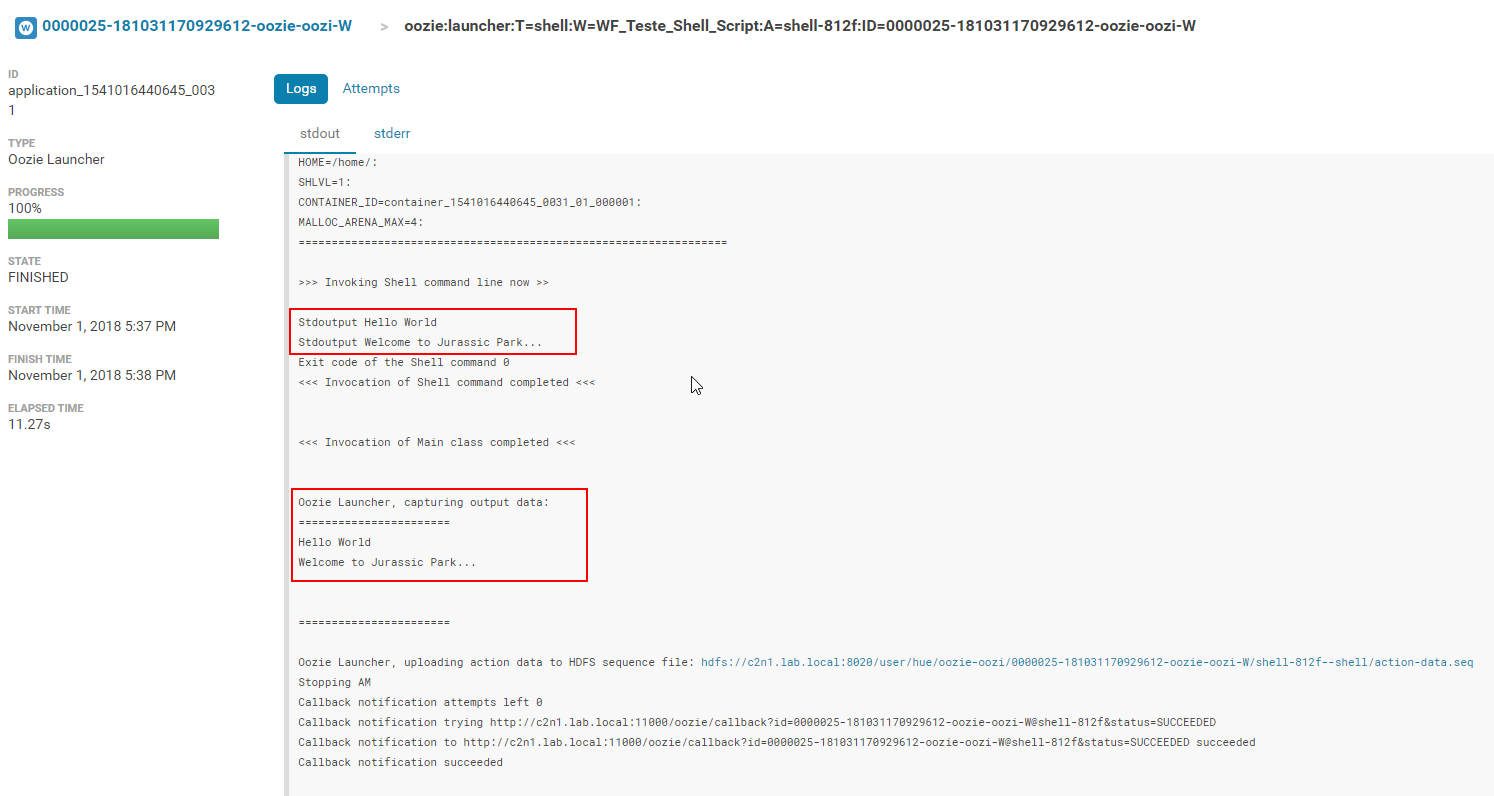












**Throubleshooting:**



# WorkFlow para Impala SEM interface

1. Criar o script shell

#/bin/bash

export PYTHON\_EGG\_CACHE=./myeggs

/usr/bin/kinit -kt impala.keytab impala/\_HOST@REALM

impala-shell -i ${impalahost} -k --ssl --ca\_cert ${impalacert} -username ${username} -f ./commands.sql

1. No Workflow criar uma action (step)

<workflow-app name="shell-impala-invalidate-wf" xmlns="uri:oozie:workflow:0.5">  
 <start to="shell-impala-invalidate"/>  
    <action name="shell-impala-invalidate">  
       <shell xmlns="uri:oozie:shell-action:0.1">  
         <job-tracker>${jobTracker}</job-tracker>  
         <name-node>${nameNode}</name-node>  
         <configuration>  
           <property>  
             <name>mapred.job.queue.name</name>  
             <value>${queueName}</value>  
           </property>  
         </configuration>  
         <exec>/tmp/teste/shell-impala.sh</exec>  
         <file>/tmp/teste/shell-impala.sh#shell-impala.sh</file>  
         <file>/tmp/teste/impala.keytab#impala.keytab</file> <file>/teste/commands.sql </file>  
       </shell>  
       <ok to=”end”/>  
       <error to=”kill”/>  
     </action>  
     <kill name=”kill”>  
       <message>Action failed, error message[${wf:errorMessage(wf:lastErrorNode())}]</message>  
     </kill>  
   <end name=”end”/>  
 </workflow-app>

1. No job.proprieties adicionar

impalahost=hostname\_impala\_daemon

impaladb=teste\_db

impalacert=/teste/query.sql

username=impala

tablename=teste

1. Script do commands.sql

INVALIDATE METADATA ${impaladb}.${tablename};

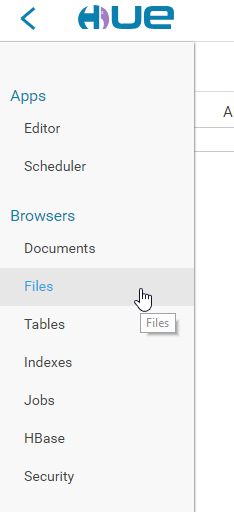
COMPUTE STATS ${impaladb}.${tablename};

CREATE TABLE IF NOT EXISTS ${impaladb}.${tablename} (id int);

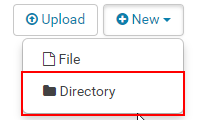
SELECT COUNT(\*) FROM ${impaladb}.${tablename};

1. Carregar os arquivos
   1. workflow.xml
   2. job.proprieties
   3. impala.keytab
   4. shell-impala.sh
2. Dar permissão de execução no script shell
   1. Hdfs dfs -chmod 755 /teste/shell-impala.sh

Ou



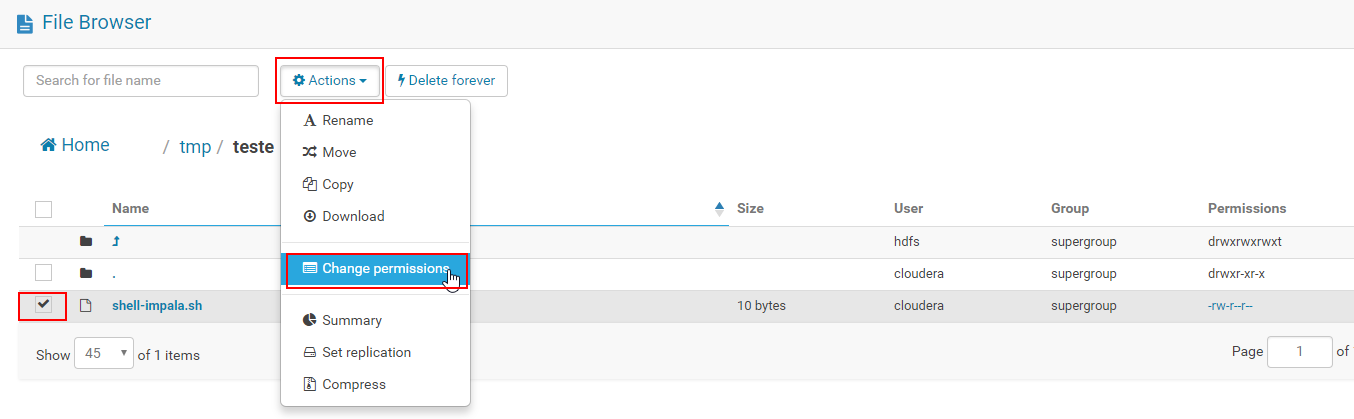
Para criar o diretório teste clique em New -> Directory

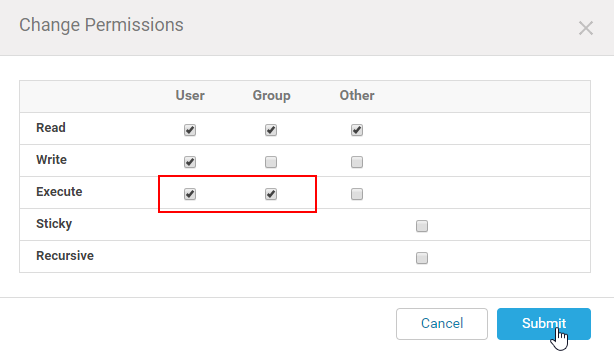


Entre no diretório e clique no botão Upload

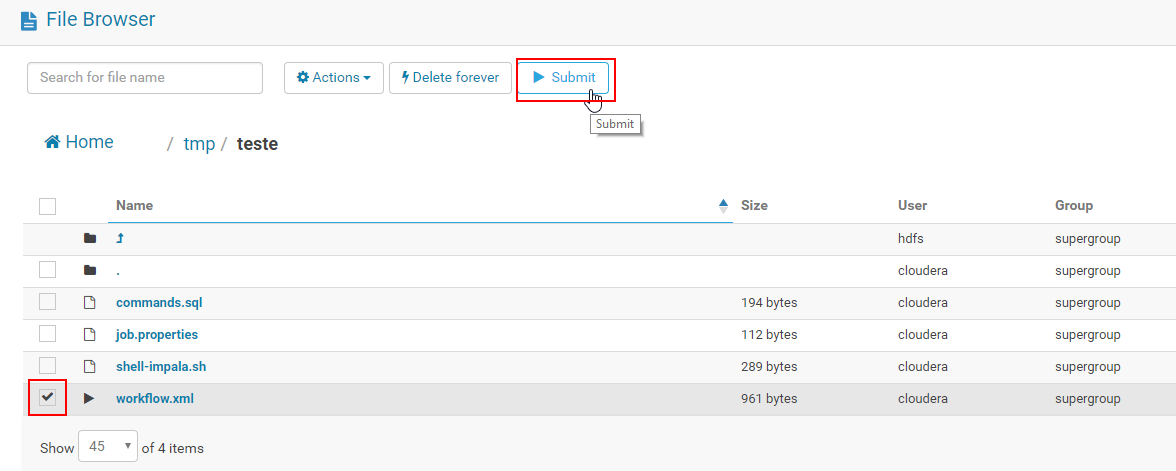


Pela interface do Hue no menu Files e depois Actions -> Change permissions





1. Rodando o workflow

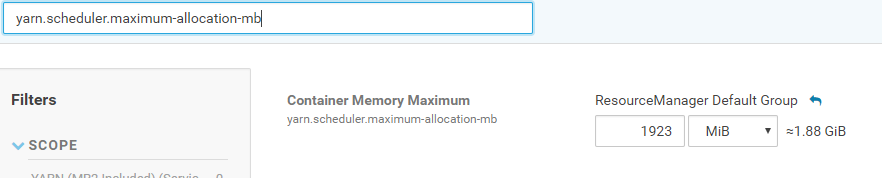


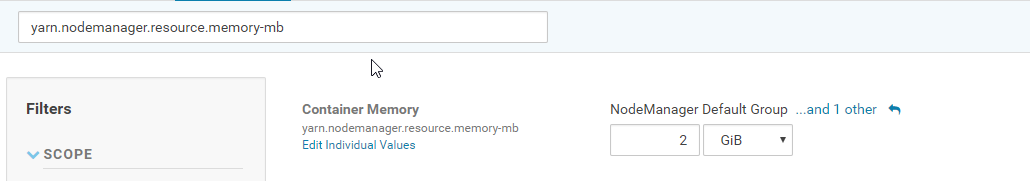
# Workflow com Hive

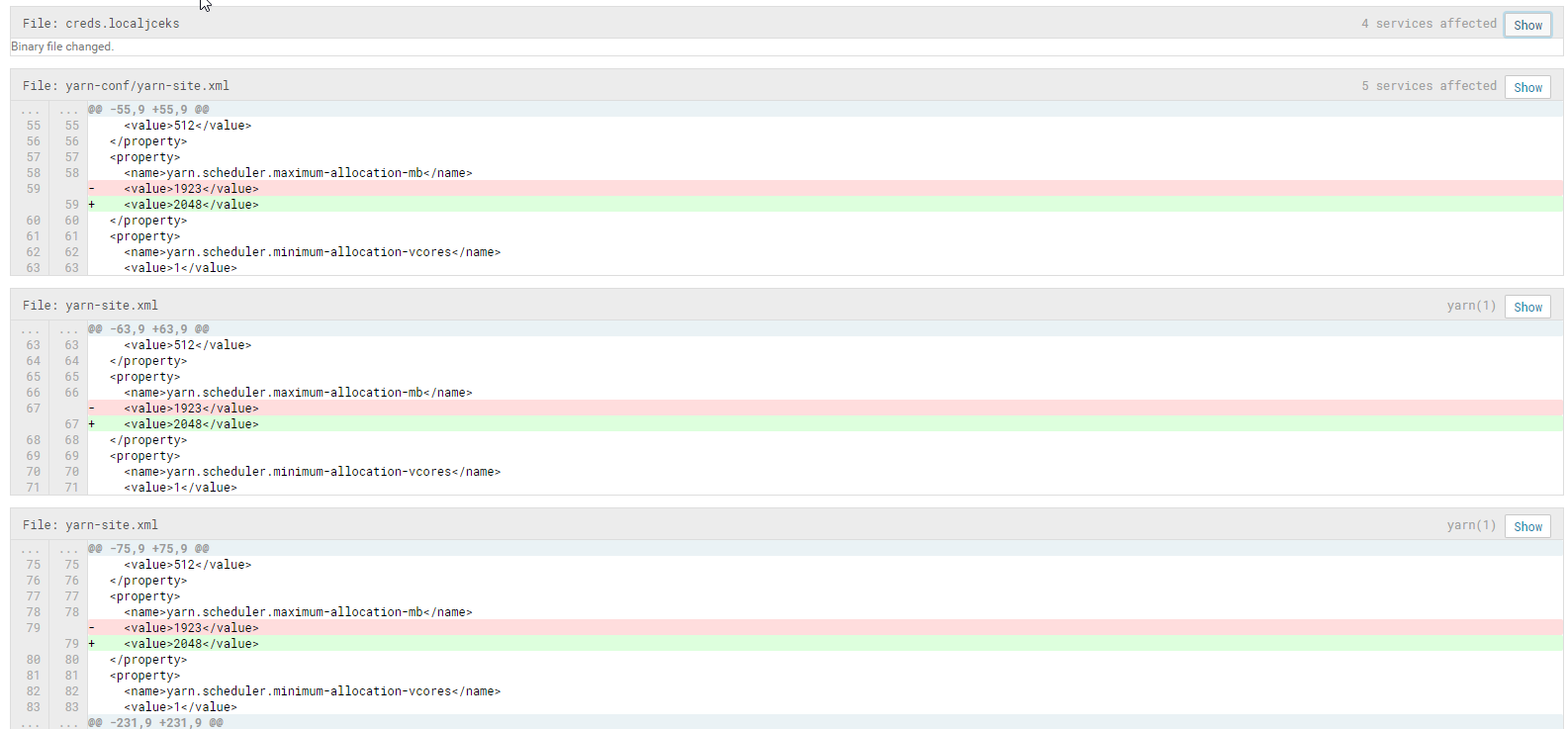
Quando o Metastore do Hive estiver em HA passar na linha de comando do sqoop o seguinte parâmetro

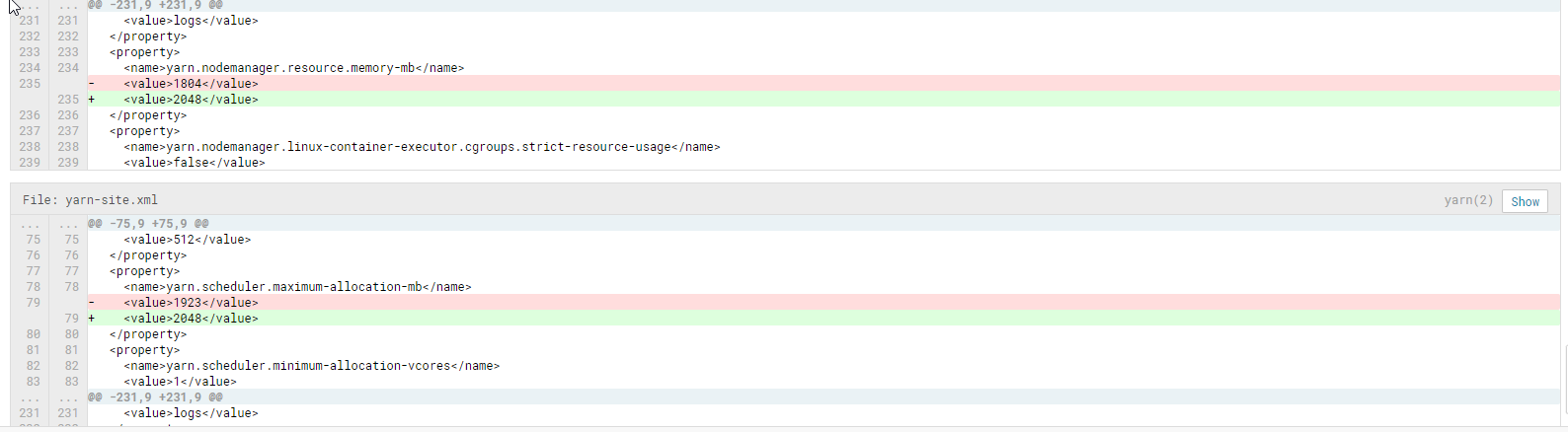
**-Dhive.metastore.uris=thrift://HOST:9083**

Outros Configurações









{u'hcat': {'xml\_name': u'hcat', 'properties': [('hcat.metastore.uri', u'thrift://c2n1.lab.local:9083'), ('hcat.metastore.principal', u'hive/c2n1.lab.local@KRB.LAB')]}, u'hive2': {'xml\_name': u'hive2', 'properties': [('hive2.jdbc.url', u'jdbc:hive2://c2n1.lab.local:10000/default;ssl=true;sslTrustStore=/var/run/cloudera-scm-agent/process/362-hue-HUE\_SERVER/cm-auto-global\_truststore.jks'), ('hive2.server.principal', 'hive/c2n1.lab.local@KRB.LAB')]}, u'hbase': {'xml\_name': u'hbase', 'properties': []}}