

# PRÁCTICA Nº 2.8

## Primeros pasos Oozie



- Nombre y apellidos: Alvaro Lucio-Villegas de Cea



## Índice

---

Carga de archivos Local	3
Cargar archivos HDFS	3
Ejecutar un Workflow.	4
Ejecutamos el job	4
Personalización del job	5
Comprobación	6

## Carga de archivos Local

Cargamos los datos del directorio “/usr/share/doc/oozie-4.1.0+cdh5.13.0+458/” y movemos el fichero comprimido a la carpeta workspace del home del usuario cloudera.

```
[cloudera@quickstart ~]$ ls -l workspace/
total 52
-rw-r--r-- 1 cloudera cloudera 46650 Oct  4 2017 oozie-examples.tar.gz
drwxrwxr-x 6 cloudera cloudera 4096 Oct 23 2017 training
[cloudera@quickstart ~]$
```

3 It

Una vez en la carpeta lo descomprimos.

```
[cloudera@quickstart ~]$ cd workspace/
[cloudera@quickstart workspace]$ tar -xvzf oozie-examples.tar.gz
examples/input-data/text/data.txt
examples/input-data/rawLogs/2010/01/01/00/20/log01.txt
examples/input-data/rawLogs/2010/01/01/00/20/ SUCCESS
examples/input-data/rawLogs/2010/01/01/00/40/log02.txt
examples/input-data/rawLogs/2010/01/01/00/40/ SUCCESS
examples/input-data/rawLogs/2010/01/01/01/00/log03.txt
examples/input-data/rawLogs/2010/01/01/01/00/ SUCCESS
examples/input-data/rawLogs/2010/01/01/01/20/log04.txt
examples/input-data/rawLogs/2010/01/01/01/20/ SUCCESS
examples/input-data/rawLogs/2010/01/01/01/40/log05.txt
examples/input-data/rawLogs/2010/01/01/01/40/ SUCCESS
examples/input-data/rawLogs/2010/01/01/02/00/log06.txt
examples/input-data/rawLogs/2010/01/01/02/00/ SUCCESS
examples/apps/aggregator/lib/oozie-examples-4.1.0-cdh5.13.0.jar
examples/apps/custom-main/lib/oozie-examples-4.1.0-cdh5.13.0.jar
examples/apps/demo/lib/oozie-examples-4.1.0-cdh5.13.0.jar
examples/apps/hadoop-el/lib/oozie-examples-4.1.0-cdh5.13.0.jar
examples/apps/java-main/lib/oozie-examples-4.1.0-cdh5.13.0.jar
examples/apps/datelist-java-main/lib/oozie-examples-4.1.0-cdh5.13.0.jar
examples/apps/map-reduce/lib/oozie-examples-4.1.0-cdh5.13.0.jar
examples/apps/spark/lib/oozie-examples.jar
examples/src/
```

## Cargar archivos HDFS

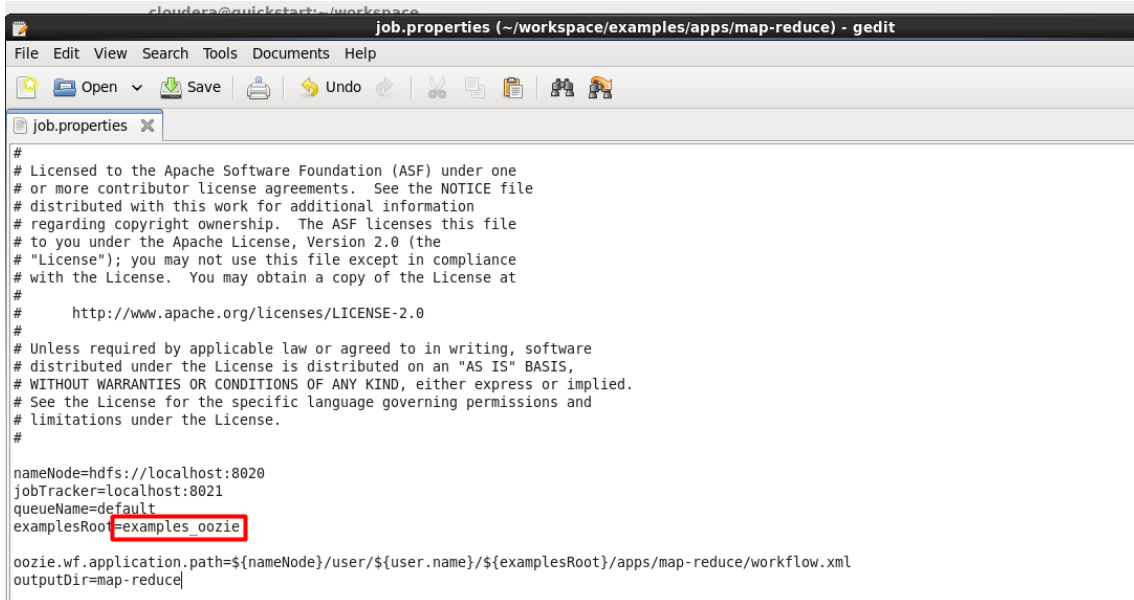
Ahora subimos a HDFS las 3 carpetas que contenían el fichero comprimido.

```
examples/
input-data/ src/
[cloudera@quickstart workspace]$ hdfs dfs -put examples /user/cloudera/examples_oozie
[cloudera@quickstart workspace]$ hdfs dfs -ls /user/cloudera/examples_oozie
Found 3 items
drwxr-xr-x - cloudera cloudera 0 2023-05-02 08:48 /user/cloudera/examples_oozie/apps
drwxr-xr-x - cloudera cloudera 0 2023-05-02 08:48 /user/cloudera/examples_oozie/input-data
drwxr-xr-x - cloudera cloudera 0 2023-05-02 08:48 /user/cloudera/examples_oozie/src
[cloudera@quickstart workspace]$
```

3 Items in Trash

## Ejecutar un Workflow

Para poder redirigir el archivo LOCAL al archivo en HDFS es necesario configurar el fichero de job.properties y redireccionar el path para que apunte a HDFS.



```

# Licensed to the Apache Software Foundation (ASF) under one
# or more contributor license agreements. See the NOTICE file
# distributed with this work for additional information
# regarding copyright ownership. The ASF licenses this file
# to you under the Apache License, Version 2.0 (the
# "License"); you may not use this file except in compliance
# with the License. You may obtain a copy of the License at
#
# http://www.apache.org/licenses/LICENSE-2.0
#
# Unless required by applicable law or agreed to in writing, software
# distributed under the License is distributed on an "AS IS" BASIS,
# WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
# See the License for the specific language governing permissions and
# limitations under the License.
#

nameNode=hdfs://localhost:8020
jobTracker=localhost:8021
queueName=default
examplesRoot=examples_oozie

oozie.wf.application.path=${nameNode}/user/${user.name}/${examplesRoot}/apps/map-reduce/workflow.xml
outputDir=map-reduce
    
```

## Ejecutamos el job

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

```

[cloudera@quickstart workspace]$ oozie job --oozie http://localhost:11000/oozie -config examples/apps/map-reduce/job.properties -run
job: 0000000-230419025730046-oozie-oozi-W
[cloudera@quickstart workspace]$
    
```

## Personalización del job

Para poder asignar un nombre concreto al los jobs es necesario dirigirse al al archivo LOCAL de propiedades copiarlo y generar uno nuevo apuntando a un nuevo directorio de workflow en HDFS, en este caso “Workflowalv.xml”

```
job.properties x alvarol.properties x workflowalv.xml x
#
# Licensed to the Apache Software Foundation (ASF) under one
# or more contributor license agreements. See the NOTICE file
# distributed with this work for additional information
# regarding copyright ownership. The ASF licenses this file
# to you under the Apache License, Version 2.0 (the
# "License"); you may not use this file except in compliance
# with the License. You may obtain a copy of the License at
#
# http://www.apache.org/licenses/LICENSE-2.0
#
# Unless required by applicable law or agreed to in writing, software
# distributed under the License is distributed on an "AS IS" BASIS,
# WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
# See the License for the specific language governing permissions and
# limitations under the License.
#
nameNode=hdfs://localhost:8020
jobTracker=localhost:8021
queueName=default
examplesRoot=examples_oozie

oozie.wf.application.path=${nameNode}/user/${user.name}/${examplesRoot}/apps/map-reduce workflowalv.xml
outputDir=map-reduce
```

Editamos el fichero “workflowalv.xml” para asignar los nombre que queremos que tengan tanto el job como la acción. Y lo subimos a HDFS

```
job.properties x alvarol.properties x workflowalv.xml x
<!--
Licensed to the Apache Software Foundation (ASF) under one
or more contributor license agreements. See the NOTICE file
distributed with this work for additional information
regarding copyright ownership. The ASF licenses this file
to you under the Apache License, Version 2.0 (the
"License"); you may not use this file except in compliance
with the License. You may obtain a copy of the License at

http://www.apache.org/licenses/LICENSE-2.0

Unless required by applicable law or agreed to in writing, software
distributed under the License is distributed on an "AS IS" BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
See the License for the specific language governing permissions and
limitations under the License.
-->
<workflow-app xmlns="uri:oozie:workflow:0.2" name="Workflow alvarol">
  <start to="action_alvarol"/>
  <action name="action_alvarol">
    <map-reduce>
      <job-tracker>${jobTracker}</job-tracker>
      <name-node>${nameNode}</name-node>
      <prepare>
        <delete path="${nameNode}/user/${wf:user()}/${examplesRoot}/output-data/${outputDir}"/>
      </prepare>
      <configuration>
        <property>
```

Ejecutamos el job de nuevo

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

```
[cloudera@quickstart workspace]$ oozie job --oozie http://localhost:11000/oozie -config examples/apps/map-reduce/alvarol.properties -run  
job: 0000001-230419025730046-oozie-oozi-W  
[cloudera@quickstart workspace]$ █
```

## Comprobación

Nos podemos dirigir desde el navegador a la pestaña de Oozie y vemos que está el trabajo corriendo correctamente si hacemos clic en él y seleccionamos la pestaña “JOB DAG” podemos ver el nombre de la acción para verificar se le asignó correctamente.

