Devon Ankar
3/24/2019

DSE 6300 – Oozie Assignment

Run examples from Oozie textbook/GitHub

Executed following code:

```
$ git clone https://github.com/oozie-book/examples.git
$ cd examples/chapter-01/identity-wf
$ mvn clean assembly:single
$ tree target/example
```

Result: successful

```
[INFO] ------------------------------------------------------------
[INFO] BUILD SUCCESS
[INFO] ------------------------------------------------------------
[INFO] Total time: 12.681s
[INFO] Finished at: Tue Mar 26 11:07:18 PDT 2019
[INFO] Final Memory: 8M/238M
[INFO] ------------------------------------------------------------
[cloudera@quickstart identity-wf]$ tree target/example
target/example
├── ch01-identity
│   ├── app
│   │   └── workflow.xml
│   └── data
│       └── input
│           └── input.txt
└── job.properties

4 directories, 3 files
```

Now we need to transfer to HDFS. Executed following code:

```
$ hdfs dfs -put target/example/ch01-identity ch01-identity
```

```
[cloudera@quickstart identity-wf]$ hdfs dfs -put target/example/ch01-identity ch01-identity
put: Failed on local exception: com.google.protobuf.InvalidProtocolBufferException: Protocol message end-group tag did not
match expected tag.; Host Details : local host is: "quickstart.cloudera/127.0.0.1"; destination host is: "localhost":11000;
```

Received error:

```
put: Failed on local exception:
com.google.protobuf.InvalidProtocolBufferException: Protocol message end-
group tag did not match expected tag.; Host Details : local host is:
"quickstart.cloudera/127.0.0.1"; destination host is: "localhost":11000;
```

Followed directions to edit `core-site.xml` files to reflect correct default value for `fs.defaultFS` to `hdfs://quickstart.cloudera:8020`

DSE 6300 – Oozie Assignment

```xml
<!--Autogenerated by Cloudera Manager-->
<configuration>
  <property>
    <name>fs.defaultFS</name>
    <value>hdfs://quickstart.cloudera:8020</value>
  </property>
  <property>
    <name>fs.trash.interval</name>
    <value>1</value>
  </property>
  <property>
    <name>io.compression.codecs</name>
```

It is unclear why it was set to 11000 by default, but after setting to 8020, the above command worked:

```
^C[cloudera@quickstart identity-wf]$ hdfs dfs -put target/example/ch01-identity ch01-identity
[cloudera@quickstart identity-wf]$ hdfs dfs -ls -R ch01-identity
drwxr-xr-x   - cloudera cloudera          0 2019-03-26 11:29 ch01-identity/app
-rw-r--r--   1 cloudera cloudera       2141 2019-03-26 11:29 ch01-identity/app/workflow.xml
drwxr-xr-x   - cloudera cloudera          0 2019-03-26 11:29 ch01-identity/data
drwxr-xr-x   - cloudera cloudera          0 2019-03-26 11:29 ch01-identity/data/input
-rw-r--r--   1 cloudera cloudera         25 2019-03-26 11:29 ch01-identity/data/input/input.txt
[cloudera@quickstart identity-wf]$ ▮
```

Per textbook, confirm that `job.properties` exists and has the correct configuration:

```
nameNode=hdfs://localhost:8020
jobTracker=localhost:8021
exampleDir=${nameNode}/user/${user.name}/ch01-identity

oozie.wf.application.path=${exampleDir}/app
```

But we want to change to `quickstart.cloudera`:

```
# nameNode=hdfs://localhost:8020
nameNode=hdfs://quickstart.cloudera:8020
jobTracker=localhost:8021
exampleDir=${nameNode}/user/${user.name}/ch01-identity

oozie.wf.application.path=${exampleDir}/app
```

Then proceed with the textbook example:

```
[cloudera@quickstart identity-wf]$ oozie job -run -config target/example/job.properties
job: 0000000-190326111908439-oozie-oozi-W
[cloudera@quickstart identity-wf]$ ▮
```
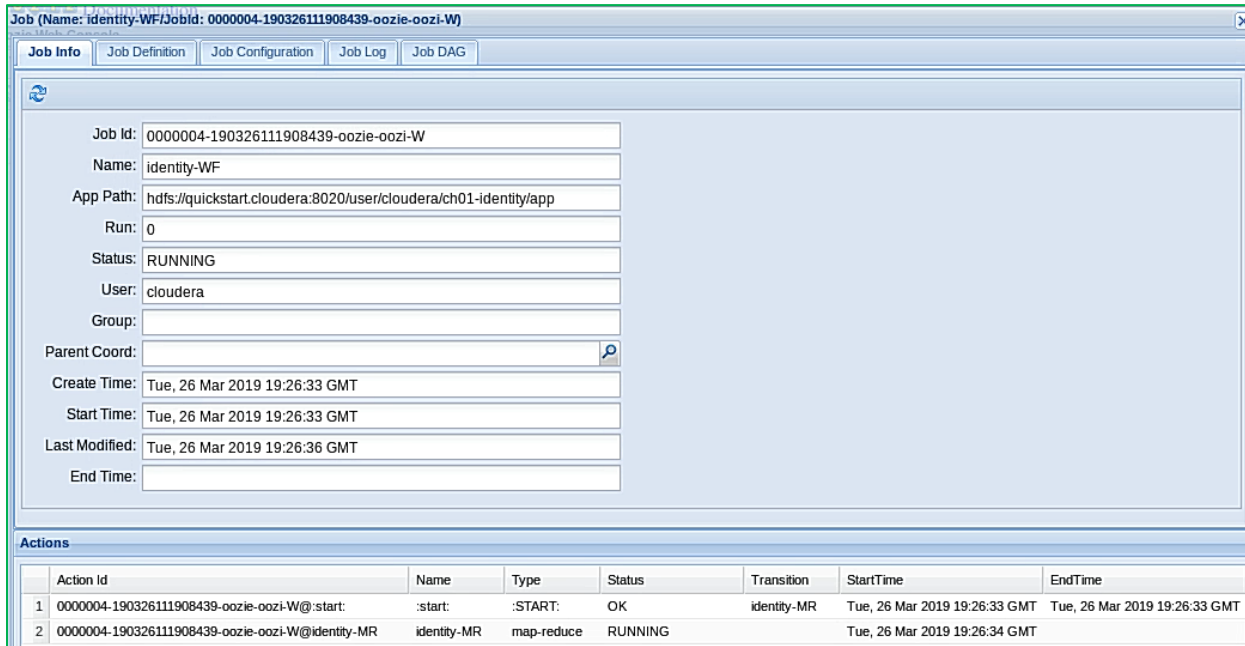
Job got killed. Let's see why:

```
HadoopAccessorException: E0900: JobTracker [localhost:8021] not allowed, not
in Oozie's whitelist. Allowed values are: [quickstart.cloudera:8032]
```

I had forgotten to change `jobTracker`. Further modification needed to `job.properties`:

```
# nameNode=hdfs://localhost:8020
nameNode=hdfs://quickstart.cloudera:8020
# jobTracker=localhost:8021
jobTracker=quickstart.cloudera:8032
exampleDir=${nameNode}/user/${user.name}/ch01-identity
```

DSE 6300 – Oozie Assignment

Now job runs (indicates `RUNNING`):

Can also check through CLI that the job has succeeded:

```
[cloudera@quickstart identity-wf]$ oozie job -info 0000004-190326111908439-oozie-oozi-W
Job ID : 0000004-190326111908439-oozie-oozi-W
------------------------------------------------------------------------------------------------------------------------
Workflow Name : identity-WF
App Path      : hdfs://quickstart.cloudera:8020/user/cloudera/ch01-identity/app
Status        : SUCCEEDED
Run           : 0
User          : cloudera
Group         : -
Created       : 2019-03-26 19:26 GMT
Started       : 2019-03-26 19:26 GMT
Last Modified : 2019-03-26 19:27 GMT
Ended         : 2019-03-26 19:27 GMT
CoordAction ID: -

Actions
------------------------------------------------------------------------------------------------------------------------
ID                                                         Status    Ext ID                 Ext Status Err Code
------------------------------------------------------------------------------------------------------------------------
0000004-190326111908439-oozie-oozi-W@:start:               OK        -                      OK         -
------------------------------------------------------------------------------------------------------------------------
0000004-190326111908439-oozie-oozi-W@identity-MR           OK        job_1553624249577_0001 SUCCEEDED  -
------------------------------------------------------------------------------------------------------------------------
0000004-190326111908439-oozie-oozi-W@success               OK        -                      OK         -
------------------------------------------------------------------------------------------------------------------------
```

```
[cloudera@quickstart identity-wf]$ hdfs dfs -ls -R ch01-identity/data/output
-rw-r--r--   1 cloudera cloudera          0 2019-03-26 12:27 ch01-identity/data/output/_SUCCESS
-rw-r--r--   1 cloudera cloudera         31 2019-03-26 12:27 ch01-identity/data/output/part-00000
```