**Schedule Hive Job Using Oozie**

**To Schedule a workflow we need two files**

workflow.xml : This will be stored on HDFS

Job.properties : This will be stored locally

To schedule hive job we need set of extra files : hive HQL file containing hive scripts and hive-site.xml file which contains hive configurations

**Workflow.xml:**

<workflow-app name="DemoOozie" xmlns="uri:oozie:workflow:0.1">

<start to="demo-hive"/> **<!—This is starting flow which goes to action mentioned , here action is “demo-hive-- >**

<action name="demo-hive"> **<!—Providing the action name -- >**

<hive xmlns="uri:oozie:hive-action:0.2**"> <!—specify which action , here it is hive action , it could be either map-reduce, pig,sqoop .. actions -- >**

<job-tracker>${jobTracker}</job-tracker> **<!—specify the job tracker , here it is getting the value from job.properties file-- >**

<name-node>${nameNode}</name-node>> **<!—specify the namenode , here it is getting the value from job.properties file-- >**

<job-xml>${appPath}/hive-site.xml</job-xml>> **<!—specify the job , here it is getting the value from job.properties file , this hive-site.xml is stored in hdfs location-- >**

<configuration> > **<!—specify the job configurations -- >**

<property>

<name>oozie.hive.defaults</name>

<value>${appPath}/hive-site.xml</value>

</property>

<property>

<name>hadoop.proxyuser.oozie.hosts</name>

<value>\*</value>

</property>

<property>

<name>hadoop.proxyuser.oozie.groups</name>

<value>\*</value>

</property>

</configuration>

<script>create\_table.hql</script> > **<!—specify the script that need to be exexcuted, here specifying create\_table.hql which would create a hive table-- >**

</hive>

<ok to="end"/> **<!—If the action is success , go to end flow -- >**

<error to="end"/> **<!—If the action fails, go to end flow -- >**

</action>

<end name="end"/> **<!—If the action is success , go to end flow -- >**

</workflow-app>

**job.properties file:**

nameNode=hdfs://localhost:9000 --location of namenode

jobTracker=localhost:8032 --location of jobtracker

queueName=default --specify how the jobs should be executed, default is FIFO

oozie.libpath=${nameNode}/user/acadgild/share/lib -- contains the lib path of oozie

oozie.use.system.libpath=false --specify true if the libraries has to taken from the system

oozie.wf.application.path=hdfs://localhost:9000/sample-mr –workflow location

appPath=hdfs://localhost:9000/sample-mr --apppath where our hive-site.xml is located

**create\_table.hql:**

use default;

create table hive\_oozie(

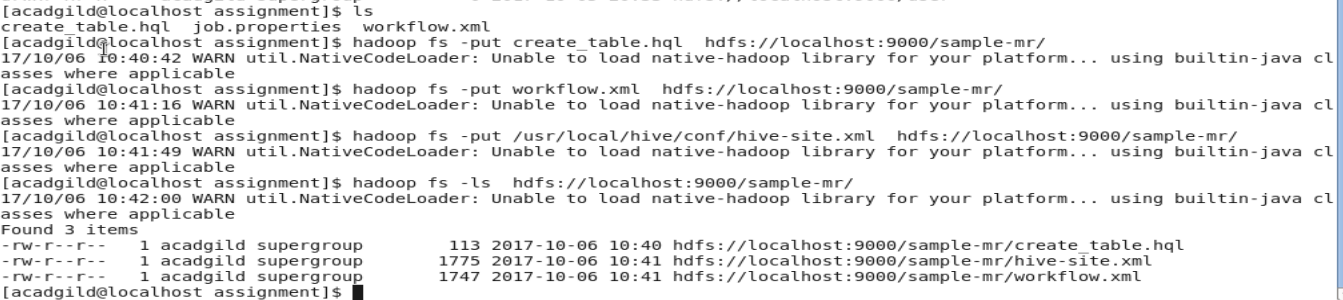
id INT,

name STRING)

ROW FORMAT DELIMITED FIELDS TERMINATED BY ',';

**Steps 1**: Please workflow.xml, hive-site.xml, create\_table.hql script in location mentioned in job.properties , is in hdfs location

Step2: start oozie service using , oozied.sh start



Step 3: run the job using

oozie job -oozie http://127.0.0.1:11000/oozie -config job.properties –run

To get status:

oozie job -oozie http://127.0.0.1:11000/oozie -info 0000005-160904225720176-oozie-oozi-W

In our acadgild VM, the oozie jobs goes to running state, but does not gets finished due to env setting,I had explicitly followed from <https://acadgild.com/blog/beginners-guide-for-oozie-installation/> .We faced this issue during the mentors session as well.

