**Scheduling Hive Job Using Oozie**

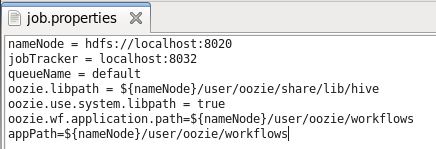
**Note: Cloudera quickstart VM is used to schedule Hive job using oozie.**

**In Cloudera quickstart VM all daemons are started at the time when we start VM, so there is no need to start all the required daemons manually, like as in acadgild VM we start hadoop daemons with “start-all.sh” command and mysql service with “sudo service mysqld start”.**

To schedule Hive job using Oozie, we need to write a Hive-action. Oozie job consists of mainly three things.

* **workflow.xml**
* **job.properties**
* **hive script**

**Step 1: Created job.properties file inside /home/cloudera/mydata/oozie/ directory using gedit job.properties command. This file consists of all the variables definitions that are used in workflow.xml.**

****

**Line 1: In nameNode variable, assigning address of namenode**

**Line 2: In jobTracker variable, assigning address of resource manager**

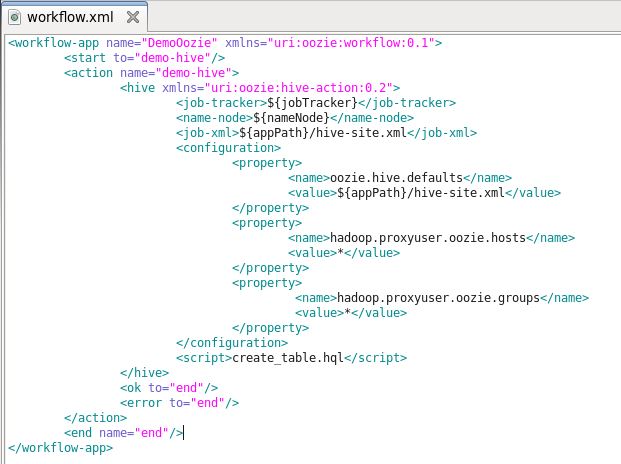
**Line 3: queueName variable stores default value**

**Line 4: oozie.libpath stores that path where all hive .jar files are present**

**Line 5: oozie.use.system.libpath is set to true so that path specified at Line 4 is picked.**

**Line 6 & 7: oozie.wf.application.path and appPath store the path where all dependent files are present like workflow.xml, hive-script, hive-site.xml**

**Step 2: Created workflow.xml file inside /home/cloudera/mydata/oozie/ directory using gedit workflow.xml command.** **This is the place where we write our Oozie action. It contains all the details of files, scripts required to schedule and run Oozie job. As the name suggests, it is an XML file where we need to mention the details in a proper tag.**

****

**Explanation:**

The first line creates a workflow app and we assign a name (according to our convenience) to recognize the job.

**<workflow-app name=”DemoOozie”>**

Indicates, we are creating a workflow app whose name is ‘DemoOozie’. All the other properties will remain inside this main tag.

**<start to=”demo-hive”/>**

**<action name=”demo-hive”>**

First tag <start to/> gives a name to hive action (i.e. ‘demo-hive’) and when <action name> matches with <start to/> then it starts oozie job.

**<hive xmlns=”uri:oozie:hive-action:0.2″>**

The line above is very important as, it says what kind of action we are going to run. It can be a MR action, or a Pig action, or Hive. Here, name is specified as Hive-action.

**<job-tracker>${jobTracker}</job-tracker>**

**<name-node>${nameNode}</name-node>**

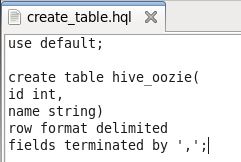
**<job-xml>${appPath}/hive-site.xml</job-xml>**

All the above tags point to the variable where job-tracker, NameNode, and Hive-site.xml are present. The exact declaration of these variables is done in Job.properties file.

**<script>create\_table.hql</script>**

In this tag we need to write the exact name of script file (here, it is a Hive script file i.e. create\_table.hql) which will be looked for and the query will get executed.

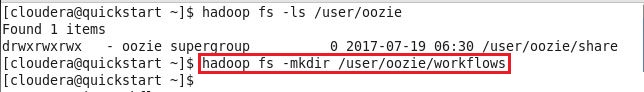
**Step 3: Created hive script i.e. create\_table.hql which we want to schedule in Oozie, inside /home/cloudera/mydata/oozie/ directory using gedit workflow.xml command.**

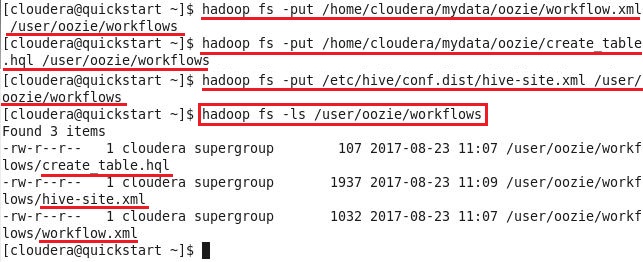
****

**Explanation:**

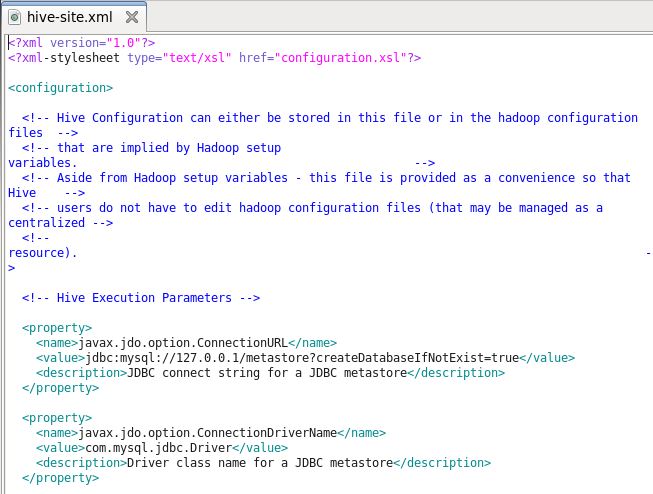
**Line 1: “use default” statement allows to use default database where table will get created, and this default database is present inside /user/hive/warehouse in hdfs.**

**Line 2-5: These statements create table with name hive\_oozie and with two columns id (type int), name (type string) and it’s specified that row format is delimited where all fields are terminated by ‘,’.**

**Step 4: Created workflows directory in hdfs inside /user/oozie, and have put workflow.xml, create\_table.hql and hive-site.xml inside /user/oozie/workflows location using below commands**

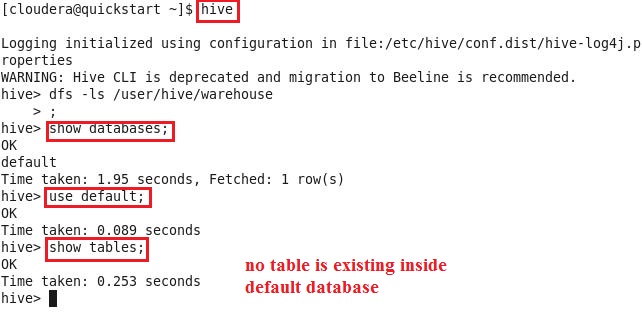
****

**hive-site.xml looks like as below:**

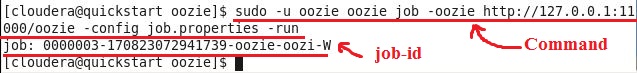
****

****

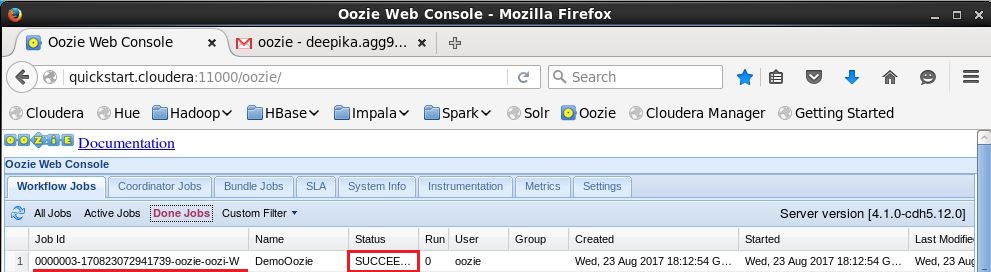
**Step 5: Checked whether hive\_oozie table already exists inside default database or not**

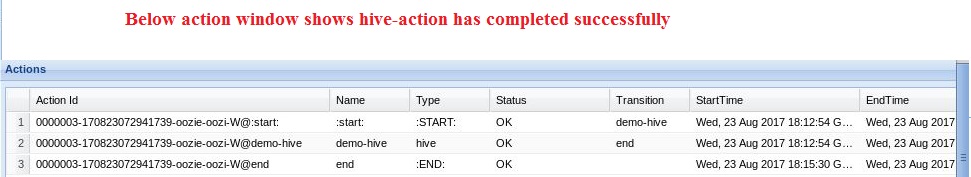
****

**Step 6: After completing above steps, ran Oozie job by using the below command.**

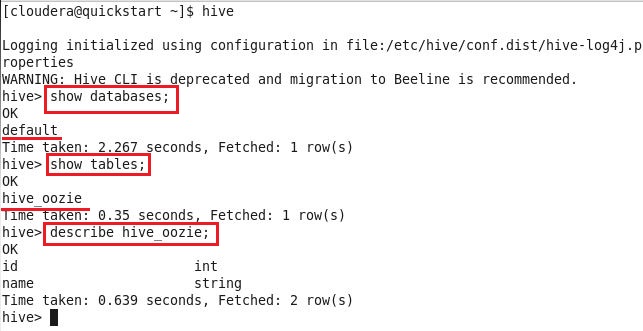
****

**After running the job, checked the status of job in Oozie web console.**





**Step 6: Below screenshot shows that hive\_oozie table has been created successfully.**

****