**Assignment 10.1**

How to Schedule Hive Job Using Oozie.

In production, where you need to run the same job for multiple times, or, you have multiple jobs that should be executed one after another, you need to schedule your job using some scheduler. There are multiple ways to automate jobs, however, here we will work with Oozie. We will begin with understanding what Oozie is and Oozie job scheduling.

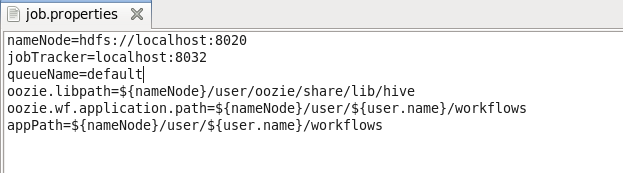
Oozie, an open source Apache project is a job scheduler that manages Hadoop jobs. In short, Oozie schedules long list of works sequentially into one job. For more details, I would suggest you to go through [this link.](https://acadgild.com/blog/running-a-map-reduce-program-using-oozie/)

To schedule Hive job using Oozie, you need to write a Hive-action. Your Oozie job will consist of mainly three things.

1. workflow.xml
2. job.properties
3. Hive script

Let us look at each of them individually.

Job.properties



This file consists of all the variable definition that you use in your workflow.xml. Let’s say, in workflow.xml, you have mentioned a property as below:

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

So, in your Job.properties file, you must declare *$nameNode* and assign the relative path

*oozie.libpath=${nameNode}/user/oozie/share/lib/hive*

Indicates the path (in hdfs) where all the respective jars are present.

*oozie.wf.application.path=${nameNode}/user/${user.name}/workflows*

This is the place where from your application will get the dependent files

Workflow.xml



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”>***

Quite self-explanatory are the above two tags which says, give a name to your action (here ‘demo-hive’) and when <action name> matches, start your oozie job.

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

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

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

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

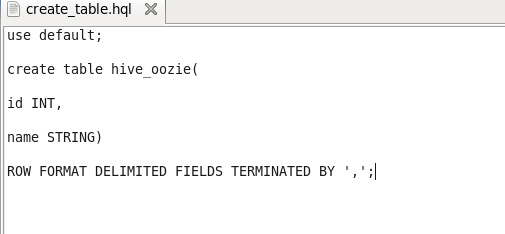
***<job-xml>${appPath}/hive-site.xml</job-xml>***

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

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

You need to fill in the exact name of your script file (here, it is a Hive script file) which will be looked for and the query will get executed.

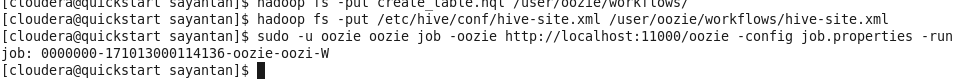
**create\_table.hql**



This is the Hive script which you want to schedule in Oozie. Quite simple and self-explanatory it is.

Now we have to create a hdfs directory and copy the workflow,hive script and hive-site.xml file to run our oozie job





It ran successfully,now we can check the output using oozie UI

