Final project, Please put the detail of each steps in a project document. And handy in. We will have a presentation on Nov 30, 2016

1.

Create a cluster , using hive to login to the cluster , upload the a text file into S3 location create a hive simple external table point to S3 location

Hint:

**Create a cluster. Use hive to login to the cluster and upload the text file into S3 location; Create a simple hive external table, point to S3 location.**

**CREATE EXTERNAL TABLE actor (**

**actor\_id INT,**

**first\_name VARCHAR(45),**

**last\_name VARCHAR(45),**

**last\_update TIMESTAMP**

**)**

**COMMENT 'Actor Infomation'**

**PARTITIONED BY (date String)**

**LOCATION 's3://<bucket\_name>/input/';**

**create s3://<bucket\_name>/input/ folder in S3**

2).Download data from cluster to a file in your local file system

hint:

do the following in cluster

$hive -e "select \* from actor " > ~/sample\_output.txt

then transfer to local file system

3). Using java or Python to upload the file sample\_output.txt to another S3 location

http://javatutorial.net/java-s3-example

for those who wants to use Java, and don’t’ have eclipse installed

Download Eclipse Standard/SDK •

Version: Kepler Service Release 2 •

URL: <http://www.eclipse.org/>

eclipse-inst-mac64.tar.gz

Then,

* Start Eclipse
* Open Help > Install New Software
* Enter http://aws.amazon.com/eclipse in the text box labeled “Work with” at the top of the dialog.
* Choose the component of the AWS Toolkit for Eclipse that you want to install. Here I, Select “AWS  Toolkit for Eclipse” from the list below.
* Click “Next” Eclipse guides you through the remaining installation steps.
* This will help to get AWS SDK for Java. Once, the update is complete restart the Eclipse.