***1] Demo: Understanding Block Storage***

Instructions

1) Download the attached file on windows sharable folder (One you configured during VM creation) - **STAGING\_AREA**

2) Pull it on Linux Machine

3) Create a directory structure /home/cloudera/hdp/pigandhive/labs - **LABS\_HOME**

4) Pull the file into LABS\_HOME/demos

5) Open the Demo in Lab Manual

6) Wait for the instructions of an Instructor

**Dataset -** data/stocks.csv

***2] Lab: Using HDFS Commands***

1) Download the attached files in **STAGING\_AREA**

2) Pull it to **LABS\_HOME**

3) See the **File locations entry** from lab manual and take the appropriate action

4) Refer the **Lab Manua**l and follow the instructions in lab manual

**Dataset -**

* data/small\_blocks.txt
* data/data.txt

Timeline = **45 Minutes**

**3] Introduction to WebHDFS**

**1) Following HTTP GET request List a Directory /user/cloudera**

**curl -i "**[**http://quickstart.cloudera:50070/webhdfs/v1/user/cloudera?op=LISTSTATUS**](http://quickstart.cloudera:50070/webhdfs/v1/user/cloudera?op=LISTSTATUS)**"**

**2) Following HTTP GET request Open and Read a File /user/cloudera/stocks.csv**

**curl -i -L "http://quickstart.cloudera:50070/webhdfs/v1/user/cloudera/stocks.csv?op=OPEN"**

**3) The following PUT request makes a new directory in HDFS named /user/cloudera/data:**

**curl -i -X PUT "http://quickstart.cloudera:50070/webhdfs/v1/user/cloudera/data?user.name=cloudera&op=MKDIRS"**

**4) //Below is a command to write the file on hdfs using single curl command instead of 2 commands**

**cd /home/cloudera/labs/demos //Assuming that there is small\_blocks.txt**

**curl -i -X PUT -T small\_blocks.txt "**[**http://quickstart.cloudera:50075/webhdfs/v1/user/cloudera/small\_blocks.txt?op=CREATE&user.name=cloudera&namenoderpcaddress=quickstart.cloudera:8020&overwrite=false**](http://quickstart.cloudera:50075/webhdfs/v1/user/cloudera/small_blocks.txt?op=CREATE&user.name=cloudera&namenoderpcaddress=quickstart.cloudera:8020&overwrite=false)**"**

**4] Ingesting data in hadoop using Java Program**

1) Download the attached file in STAGING\_AREA

2) Extract it in that location

3) Create a folder Lab1.2 in LABS\_HOME

4) Put the extracted folder in step 2 in above location in step 3

5) Wait for the instructions from an Instructor

**Code and Dataset -**

* data/HDFS\_API.rar

5] Importing data using sqoop

1) Download the attached files in **STAGING\_AREA**

2) Create a folder Lab3.1 in **LABS\_HOME**

3) Copy the files in step 1 to step 2 location

4) Wait for the instructions from the trainer.

**Dataset -**

* data/salaries.txt