Lab: Using WebHDFS (Cloudera)

------------------

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

curl -i "<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>"

Perform following activities on HDFS

-------------------------------------------------

1. Create a directory in HDFS

2. Create a subdirectory under a directory in step #1

Note - Above directories should be created using a single command

Solution - hdfs dfs -mkdir -p test1/test2

3. Ingest a text file into HDFS in a directory created in step #1

Solution - hdfs dfs -put abc.txt test1/

4. List the contents of the directory created in step #1

Solution - hdfs dfs -ls test1/

5. Print the contents of the file ingested in step #2

Solution - hdfs dfs -cat test1/abc.txt > /tmp/abc.txt

6. Change the replication factor of file ingested in step #2

Solution- hdfs dfs -setrep -w 4 /tmp/abc.txt

7. Display the changed replication factor

Solution- hdfs dfs -ls /tmp/abc.txt

8. Change the permissions of file ingested in step #2 to 644

Solution hdfs dfs -chmod 644 test1/abc.txt

9. Count the number of directories, files and bytes for a directory created in step #1

Solution - hdfs dfs -count test1

10. Delete the entire directory structure you created up to this point

Solution - hdfs dfs -rm -R test1

11. Automate above workflow

Note - Please refer the HDFS commands documentation at

<https://hadoop.apache.org/docs/r2.6.1/hadoop-project-dist/hadoop-common/FileSystemShell.html>

Time Line - 30 Mins

What you mean by building a Java Application?

1) Use an IDE to Create and then build the application (Optional) (Done)

2) Create a Java Project in Eclipse (Done)

3) Study the code that has been shared there with you to understand the package/s need to be used (Done)

4) Based on above study in step #3 crete a respective package in above Eclipse Java project (Done)

5) Import the source code / \*.java files in your project (Done)

6) Resolve the compilation errors by understanding the dependencies (Done)

7) Resolve the dependencies by adding the respective dependencies/libraries in the

classpath of the project (Done)

8) Study the code (Done)

9) Once all dependencies they get resolved create a jar (Done)

10) Run the jar with appropriate command on Hadoop cluster

Lab: Building and running a Java MR application for data ingestion

—---------------------------------------------------------------------------------------

Write down a java program which will do following things

1. Create a java project in eclipse having name as **TestIngestion**
2. Write down a java class **com.cdac.dbda23.IngestFile** which will
   1. Take a file **small\_blocks.txt** to be loaded as a command line parameter to the program
   2. Load the file in your home on hdfs.
   3. File to be loaded must be in a directory **/home/cloudera/hdp/labs/dataingestion**
3. Create a jar having name as **loadfile.jar**
4. Run the jar to load the file at location specified in step #2.
5. Can you automate this entire process?

TimeLine = 40 Mins

<https://stackoverflow.com/questions/48114204/failed-to-restart-hadoop-namenode-using-cloudera-quickstart>

<https://kshitish-bigdata.blogspot.com/2015/02/hadoop-namenode-is-dead-and-pid-file.html>

<https://stackoverflow.com/questions/44161069/clipboard-copy-and-paste-does-not-work-in-vmware-player-12>

Ref - <https://stackoverflow.com/questions/38027877/spark-transformation-why-is-it-lazy-and-what-is-the-advantage>