Assignment Solution

Week1: Getting Started with Big Data and Understanding HDFS Concept along with Linux Commands

Week 1 Assignment Solutions

Note:

Total Marks : 50
Each Subpart carries 2 Marks

Question 1) (18 Marks)

1. Create two directories 'dir1' and 'dir2' using a single hdfs command inside
home directory of hdfs in cloudera VM.dir2 should be subdirectory of dir1.
2. Verify that the two folders have been created in the above path.
3. Inside dir 2 Create an empty file, file1.txt
4. Create a file file2.txt in local filesystem with some text inside it
5. Copy file2.txt from local to hdfs inside dir2.
6. List the subdirectories and files inside dir1 recursively
7. List the files inside dir2 ,sorted by size but size should be displayed in KBs/MBs and not bytes
8. Rename the file, file2.txt to file3.txt
9. Remove the directory dir1 using a single command.

Solution:

1) hadoop fs -mkdir -p /user/cloudera/dir1/dir2

2) hadoop fs -ls /user/cloudera
hadoop fs -ls /user/cloudera/dir1

3) hadoop fs -touchz /user/cloudera/dir1/dir2/file1.txt

4) gedit ./Desktop/file2.txt

5) hadoop fs -put /home/cloudera/Desktop/file2.txt /user/cloudera/dir1/dir2/

6) hadoop fs -ls -R /user/cloudera/dir1

7) hadoop fs -ls -S -h /user/cloudera/dir1/dir2/

8) hadoop fs -mv /user/cloudera/dir1/dir2/file2.txt /user/cloudera/dir1/dir2/file3.txt

9) hadoop fs -rm -R /user/cloudera/dir1

Question 2) (2 Marks)

Suppose there is file of size 514 MB stored in HDFS (Hadoop 2.x) using default block size configuration and default replication factor.Then, how many blocks will be created in total and what will be the size of each block?

Solution:

There will be 5 blocks created. 4 blocks of size 128MB each and 1 Block of size 2MB. Default replication factor is 3. Considering RF ,Totally 5*3 = 15 blocks are created.

Question 3) (8 Marks)

1. Create a directory inside home directory of local filesystem named 'test'
2. Create few empty files inside the test directory namely a.pdf, b.html, c.xml
3. List the files in reverse alphabetical order of file name
4. Display only the file which ends with .html extension

Solution:

i) mkdir test
ii) touch test/a.pdf test/b.html test/c.xml
iii) ls -lr test
iv) ls -ltr test | grep html OR ls -l | grep .html

Question 4) **(8 Marks)**

1. Create two new text files, file1 and file2 , with following content using cat command in your linux home directory.
file1 : This is from file1
file2: This is from file2
2. Display the contents of the file 1 and file2 using cat command
3. Concatenate the contents of the two files and put them into a new file file3 and display the results.
4. Count the number of lines and number of words in the file3.

Solution:

i) cat > file1
cat > file2

ii) cat file1
cat file2

Or

cat file1 file2

iii) cat file1 file2 > file3
cat file3

Or

cat file1file2 >> file3

iv) wc -l file3
wc -w file3

Or

wc-l-w file3

Question 5) **(4 Marks)**

Create a text file myfile.txt with 5 lines in home directory of local filesystem

1. Display last 3 lines of that file.
2. Display all lines of that text file except first line.

Solution:

i) cat > myfile.txt
tail -n3 myfile.txt

Or

tail -3 myfile.txt

ii) tail -n+2 myfile.txt

Question 6) **(10 Marks)**

The getmerge command in Hadoop is for merging files existing in the HDFS file system into a single file in the local file system.
1. Use the help for getmerge command to see the arguments it takes
2. Create file1.txt in local with contents " Hello, this is from file1 ",Create file2.txt in local with contents "Hello, this is from file2"
3. Copy the file1.txt and file2.txt into a location in hdfs inside home directory in hdfs
4. Use the getmerge command to merge the contents of two files present in hdfs and put the merged content into a single local destination file named filenew.txt.
5. Display the merged contents of the file filenew.txt

Solution:

1) hadoop fs -help getmerge

2) cd Desktop

gedit file1.txt
gedit file2.txt

3) hadoop fs -put file1.txt /user/cloudera/
hadoop fs -put file2.txt /user/cloudera/

4) hadoop fs -getmerge /user/cloudera/file1.txt /user/cloudera/file2.txt /home/cloudera/Desktop/filenew.txt

5) cat ./Desktop/filenew.txt

****