Lab Program 2

HDFS COMMANDS

- HDFS is the primary or major component of the Hadoop ecosystem which is responsible for storing large data sets of structured or unstructured data across various nodes and thereby maintaining the metadata in the form of log files.
- Note: In order to use hdfs commands we need to start services of Hadoop

Command to start Hadoop

1. ./start-all.sh
2. jps {To check whether all daemon processes are running)

1. Listing files and Directories using ls command

Note: Every hdfs commands must have a prefix *hdfs dfs* or *hadoop fs*

```
disha@disha:~$ hadoop fs -ls
Found 3 items
-rw-r--r--   1 disha supergroup          0 2021-04-22 12:26 empty.txt
-rw-r--r--   1 disha supergroup          0 2021-04-22 11:05 file.txt
-rw-r--r--   1 disha supergroup         44 2021-04-23 11:52 wordcount.txt
disha@disha:~$
```

2. Creating Directory using mkdir command

```
disha@disha:~$ hadoop fs -mkdir prog2
disha@disha:~$ hadoop fs -ls
Found 4 items
-rw-r--r--   1 disha supergroup          0 2021-04-22 12:26 empty.txt
-rw-r--r--   1 disha supergroup          0 2021-04-22 11:05 file.txt
drwxr-xr-x   - disha supergroup          0 2021-04-25 21:31 prog2
-rw-r--r--   1 disha supergroup         44 2021-04-23 11:52 wordcount.txt
```

3. Create an empty file using touchz command

```
disha@disha:~$ hadoop fs -touchz empty.txt
disha@disha:~$ hadoop fs -ls
Found 2 items
-rw-r--r--   1 disha supergroup          0 2021-04-25 21:34 empty.txt
drwxr-xr-x   - disha supergroup          0 2021-04-25 21:31 prog2
disha@disha:~$
```

4. To copy files/folders from local file system to hdfs store using copyFromLocal or put command

- First create one file and enter few contents

```
disha@disha:~$ hadoop fs -copyFromLocal f1.txt prog2/f1.txt
disha@disha:~$ hadoop fs -ls prog2/
Found 1 items
-rw-r--r--   1 disha supergroup         20 2021-04-25 21:38 prog2/f1.txt
disha@disha:~$
```

```
disha@disha:~$ hadoop fs -put f1.txt prog2/f2.txt
disha@disha:~$ hadoop fs -ls prog2/
Found 2 items
-rw-r--r--   1 disha supergroup         20 2021-04-25 21:38 prog2/f1.txt
-rw-r--r--   1 disha supergroup         20 2021-04-25 21:39 prog2/f2.txt
disha@disha:~$
```

5. To print the content of the file using cat command

```
disha@disha:~$ hadoop fs -cat prog2/f1.txt
hi this is testing

disha@disha:~$
```

6. To copy files/folders from hdfs store to local file system using copyToLocal

```
disha@disha:~$ hadoop fs -copyToLocal prog2/
disha@disha:~$ ls
a.out                     f1.txt                          Public
Desktop                   hadoop-3.2.2.tar.gz             snap
dfsdata                   hdfs.txt                        sorting.c
Documents                 mongodb-compass_1.26.0_amd64.deb Templates
Downloads                 Music                           tmpdata
eclipse                   newdbexport.json                Videos
eclipse-inst-jre-linux64  Pictures                        wordcount_mapper.py
eclipse-workspace         prog2                           wordcount_reducer.py
disha@disha:~$ cd prog2/
disha@disha:~/prog2$ ls
f1.txt  f2.txt
disha@disha:~/prog2$
```

7. **moveFromLocal:** This command will move file from local to hdfs.

```
disha@disha:~$ hadoop fs -moveFromLocal prog2 prog3
disha@disha:~$ hadoop fs -ls
Found 3 items
-rw-r--r--   1 disha supergroup          0 2021-04-25 21:34 empty.txt
drwxr-xr-x   - disha supergroup          0 2021-04-25 21:39 prog2
drwxr-xr-x   - disha supergroup          0 2021-04-25 21:43 prog3
disha@disha:~$ ls
a.out                    f1.txt                          snap
Desktop                  hadoop-3.2.2.tar.gz             sorting.c
dfsdata                  hdfs.txt                        Templates
Documents                mongodb-compass_1.26.0_amd64.deb tmpdata
Downloads                Music                           Videos
eclipse                  newdbexport.json                wordcount_mapper.py
eclipse-inst-jre-linux64 Pictures                        wordcount_reducer.py
eclipse-workspace        Public
disha@disha:~$
```

8. cp: This command is used to copy files within hdfs.

```
disha@disha:~$ hadoop fs -cp prog2/f1.txt prog3/f1_copy.txt
disha@disha:~$ hadoop fs -cat prog3/f1_copy.txt
hi this is testing

disha@disha:~$
```

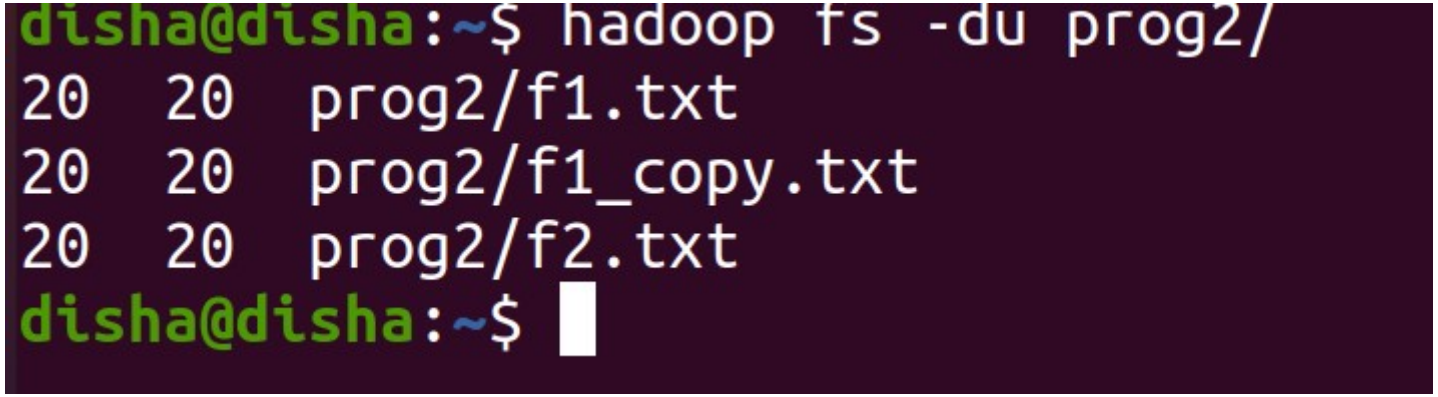
9. mv: This command is used to move files within hdfs.

```
disha@disha:~$ hadoop fs -ls prog3/
Found 3 items
-rw-r--r--   1 disha supergroup         20 2021-04-25 21:43 prog3/f1.txt
-rw-r--r--   1 disha supergroup         20 2021-04-25 21:48 prog3/f1_copy.txt
-rw-r--r--   1 disha supergroup         20 2021-04-25 21:43 prog3/f2.txt
disha@disha:~$ hadoop fs -ls prog2/
Found 2 items
-rw-r--r--   1 disha supergroup         20 2021-04-25 21:38 prog2/f1.txt
-rw-r--r--   1 disha supergroup         20 2021-04-25 21:39 prog2/f2.txt
disha@disha:~$ hadoop fs -mv prog3/f1_copy.txt prog2/
disha@disha:~$ hadoop fs -ls prog2/
Found 3 items
-rw-r--r--   1 disha supergroup         20 2021-04-25 21:38 prog2/f1.txt
-rw-r--r--   1 disha supergroup         20 2021-04-25 21:48 prog2/f1_copy.txt
-rw-r--r--   1 disha supergroup         20 2021-04-25 21:39 prog2/f2.txt
disha@disha:~$ hadoop fs -ls prog3/
Found 2 items
-rw-r--r--   1 disha supergroup         20 2021-04-25 21:43 prog3/f1.txt
-rw-r--r--   1 disha supergroup         20 2021-04-25 21:43 prog3/f2.txt
disha@disha:~$
```

10. rmr: This command deletes a file from HDFS recursively. It is very useful command when you want to delete a non-empty directory.

```
disha@disha:~$ hadoop fs -ls
Found 3 items
-rw-r--r--   1 disha supergroup          0 2021-04-25 21:34 empty.txt
drwxr-xr-x   - disha supergroup          0 2021-04-25 21:50 prog2
drwxr-xr-x   - disha supergroup          0 2021-04-25 21:50 prog3
disha@disha:~$ hadoop fs -rmr prog3/
rmr: DEPRECATED: Please use '-rm -r' instead.
Deleted prog3
disha@disha:~$ hadoop fs -ls
Found 2 items
-rw-r--r--   1 disha supergroup          0 2021-04-25 21:34 empty.txt
drwxr-xr-x   - disha supergroup          0 2021-04-25 21:50 prog2
disha@disha:~$
```

11. du: It will give the size of each file in directory.



12. dus: This command will give the total size of directory/file.

```
disha@disha:~$ hadoop fs -dus prog2/
dus: DEPRECATED: Please use 'du -s' instead.
60  60  prog2
disha@disha:~$
```

13. stat: It will give the last modified time of directory or path. In short it will give stats of the directory or file.

```
disha@disha:~$ hadoop fs -stat prog2
2021-04-25 16:20:55
disha@disha:~$
```

14. AppendToFile

```
disha@disha:~$ hadoop fs -appendToFile f1.txt f2.txt new.txt
disha@disha:~$ hadoop fs -cat new.txt
hi this is testing

content of file 2
disha@disha:~$
```

Tasks To complete

1. Create a directory named by your **name**
2. Add two empty files into it
3. List all the files in the directory that you have created
4. Create one more directory by your **usn** and put that inside the directory of your **name**
5. Create a file in your local system by the name_usn.txt and add the contents as

   “Syntax to list ls command is hadoop fs –ls “

Tasks To complete contd...

- Put the file usn_name.txt in your hdfs directory called usn
- Put empty file of hdfs into the local file system
- Use mv command to move files from local file system and hdfs file system
- Experiment of cp command
- Display the total size of all the files in name directory
- Display each file size in usn directory
- Use stat command on name and usn directories
- Illustrate the use of append command