

Practical 3
Big Data Analytics
2CS702

Mistry Unnat
20BCE515



Department of Computer Science and Engineering
Institute of Technology
Nirma University
Ahmedabad

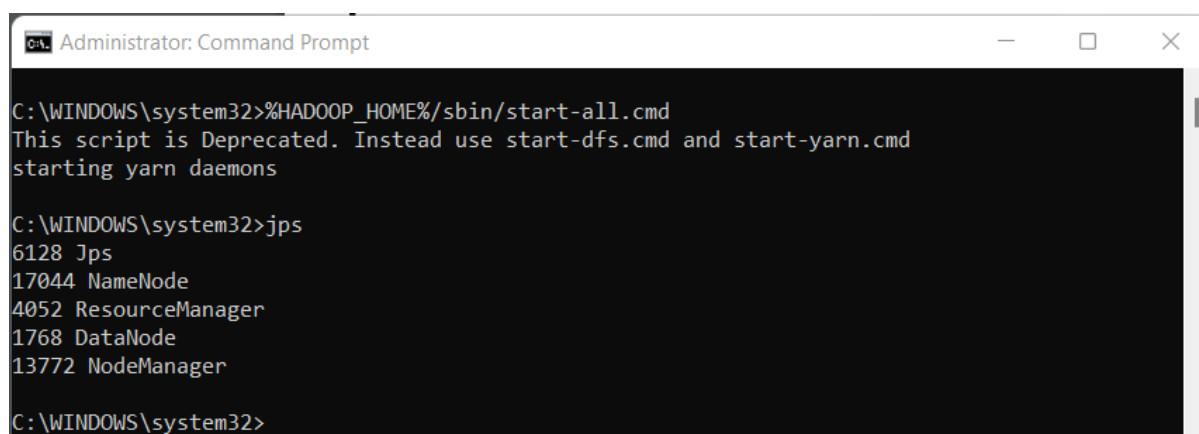
AIM : Setup Single Node Hadoop Cluster and Apply HDFS Commands on single node Hadoop Cluster

Commands :

Starting the daemons... and checking if they are up and working

Commands used :

1. %HADOOP_HOME%/sbin/start-all.cmd
2. jps



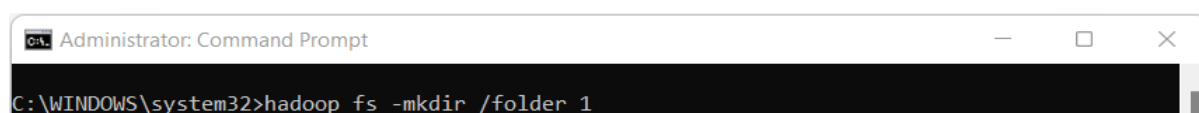
```
Administrator: Command Prompt

C:\WINDOWS\system32>%HADOOP_HOME%/sbin/start-all.cmd
This script is deprecated. Instead use start-dfs.cmd and start-yarn.cmd
starting yarn daemons

C:\WINDOWS\system32>jps
6128 Jps
17044 NameNode
4052 ResourceManager
1768 DataNode
13772 NodeManager

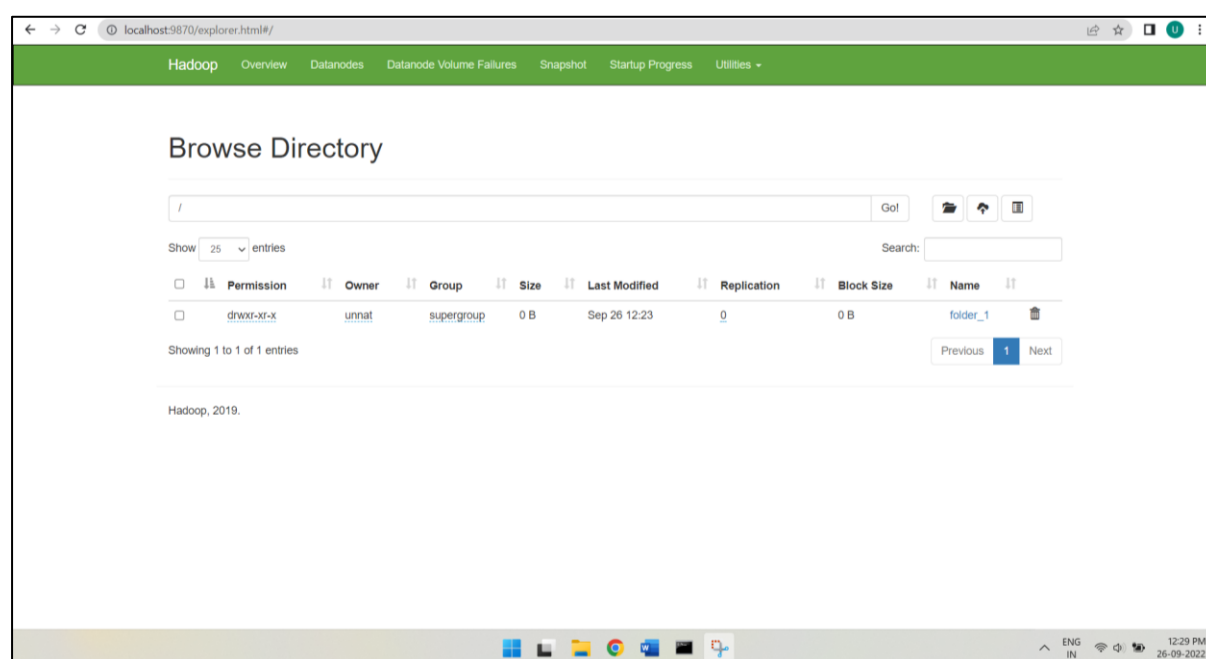
C:\WINDOWS\system32>
```

1. **mkdir** : creates directory



```
Administrator: Command Prompt

C:\WINDOWS\system32>hadoop fs -mkdir /folder_1
```



2. **ls** : lists files and folders

```

Administrator: Command Prompt
C:\WINDOWS\system32>hadoop fs -ls /
Found 1 items
drwxr-xr-x  - unnat supergroup          0 2022-09-26 12:23 /folder_1
C:\WINDOWS\system32>

```

3. **chmod** : To change file and directory permissions, use the command chmod (change mode).

| Permission | Files | Directories |
|------------|-----------------------------|--|
| r | can read the file | can ls the directory |
| w | can write the file | can modify the directory's contents |
| x | can execute the file | can cd to the directory |

```

Administrator: Command Prompt
C:\WINDOWS\system32>hadoop fs -ls /
Found 1 items
drwxr-xr-x  - unnat supergroup          0 2022-09-26 12:23 /folder_1

C:\WINDOWS\system32>hadoop fs -chmod -r /folder_1

C:\WINDOWS\system32>hadoop fs -ls /
Found 1 items
d-wx--x--x  - unnat supergroup          0 2022-09-26 12:23 /folder_1
C:\WINDOWS\system32>

```

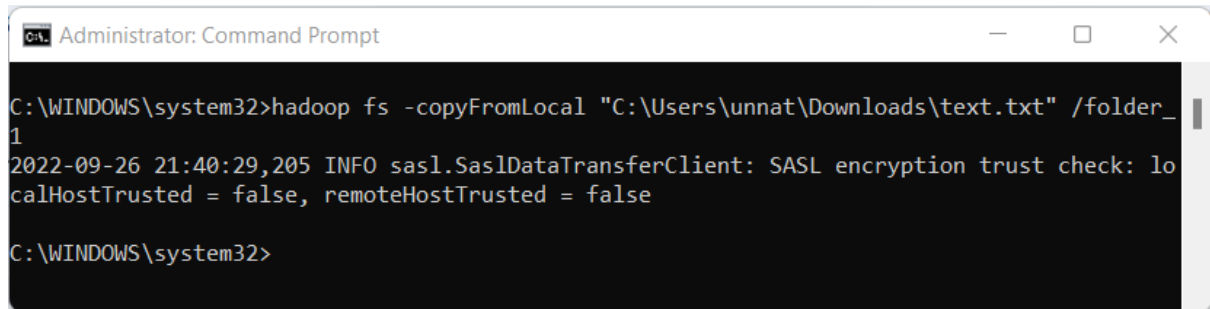
4. **count** : Hadoop HDFS count option is used to count a number of directories, number of files, number of characters in a file and file size.

```

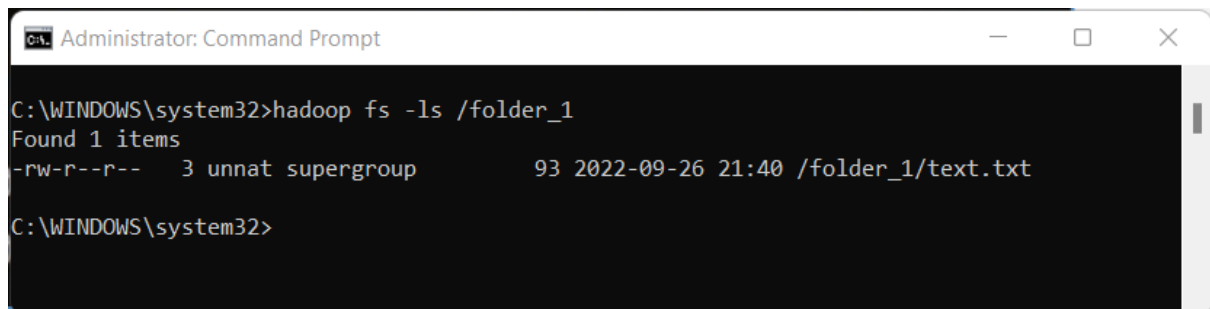
Administrator: Command Prompt
C:\WINDOWS\system32>hadoop fs -count -v /folder_1
DIR_COUNT  FILE_COUNT  CONTENT_SIZE  PATHNAME
          1           0              0  /folder_1
C:\WINDOWS\system32>

```

5. **copyFromLocal** : Hadoop copyFromLocal command is used to copy the file from your local file system to the HDFS(Hadoop Distributed File System).

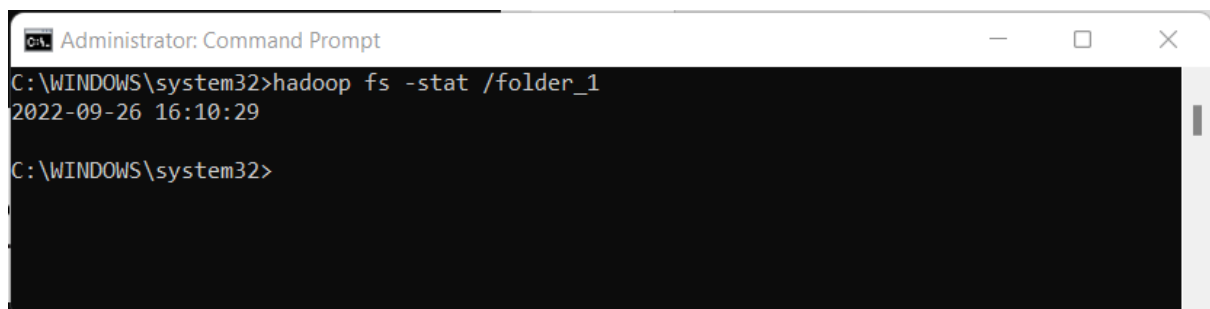


```
C:\WINDOWS\system32>hadoop fs -copyFromLocal "C:\Users\unnat\Downloads\text.txt" /folder_1
2022-09-26 21:40:29,205 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
C:\WINDOWS\system32>
```



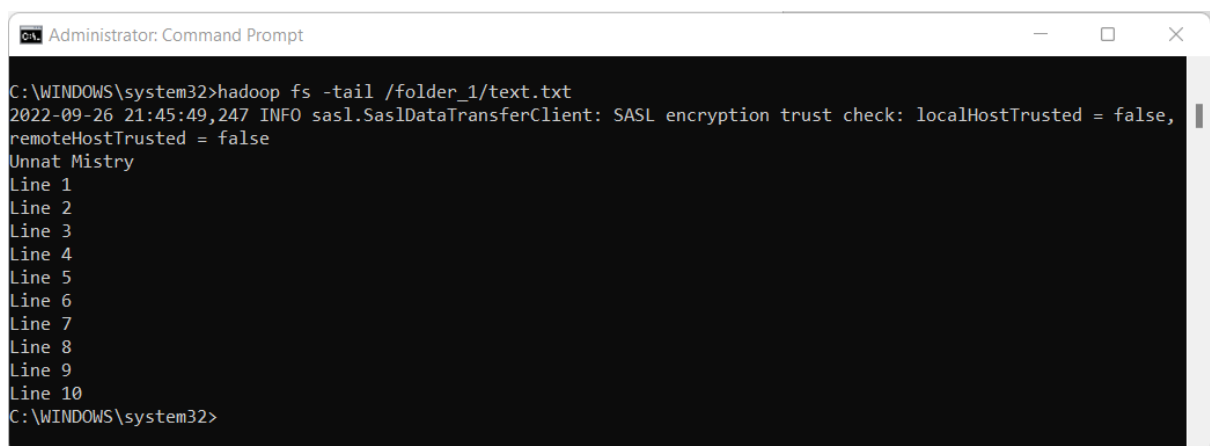
```
C:\WINDOWS\system32>hadoop fs -ls /folder_1
Found 1 items
-rw-r--r-- 3 unnat supergroup 93 2022-09-26 21:40 /folder_1/text.txt
C:\WINDOWS\system32>
```

6. **stat** : Print statistics about the file/directory at <path> in the specified format.



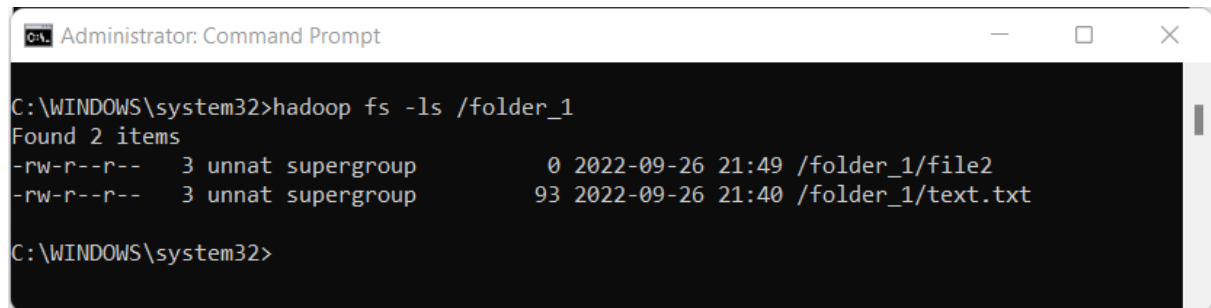
```
C:\WINDOWS\system32>hadoop fs -stat /folder_1
2022-09-26 16:10:29
C:\WINDOWS\system32>
```

7. **tail** : The Hadoop fs shell tail command shows the last 1KB of a file on console or stdout.



```
C:\WINDOWS\system32>hadoop fs -tail /folder_1/text.txt
2022-09-26 21:45:49,247 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
Unnat Mistry
Line 1
Line 2
Line 3
Line 4
Line 5
Line 6
Line 7
Line 8
Line 9
Line 10
C:\WINDOWS\system32>
```

8. **touchz** : touchz command creates a file in HDFS with file size equals to 0 byte. The directory is the name of the directory where we will create the file, and filename is the name of the new file we are going to create.

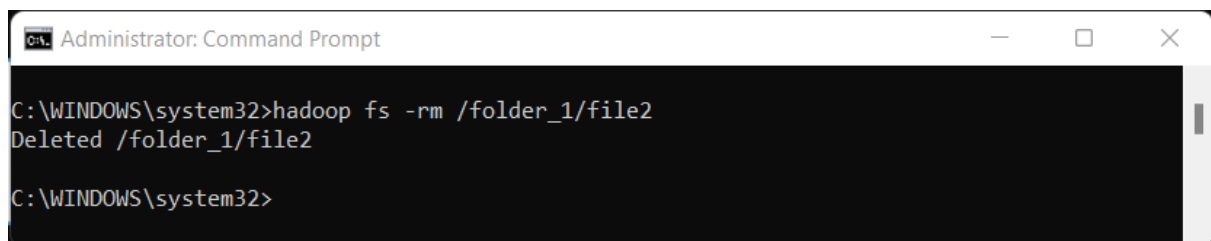


```
Administrator: Command Prompt

C:\WINDOWS\system32>hadoop fs -ls /folder_1
Found 2 items
-rw-r--r--   3 unnat supergroup      0 2022-09-26 21:49 /folder_1/file2
-rw-r--r--   3 unnat supergroup    93 2022-09-26 21:40 /folder_1/text.txt

C:\WINDOWS\system32>
```

9. **rm** : This command is similar to the Linux rm command, and it is used for removing a file from the HDFS file system.

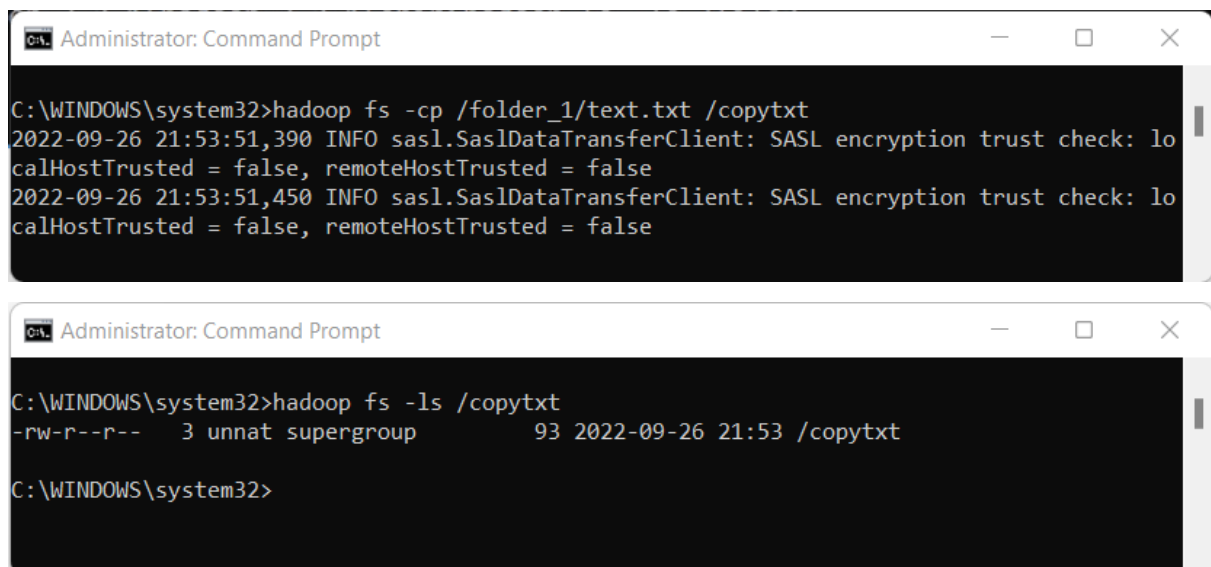


```
Administrator: Command Prompt

C:\WINDOWS\system32>hadoop fs -rm /folder_1/file2
Deleted /folder_1/file2

C:\WINDOWS\system32>
```

10. **cp** : The cp command copies a file from one directory to another directory within the HDFS.



```
Administrator: Command Prompt

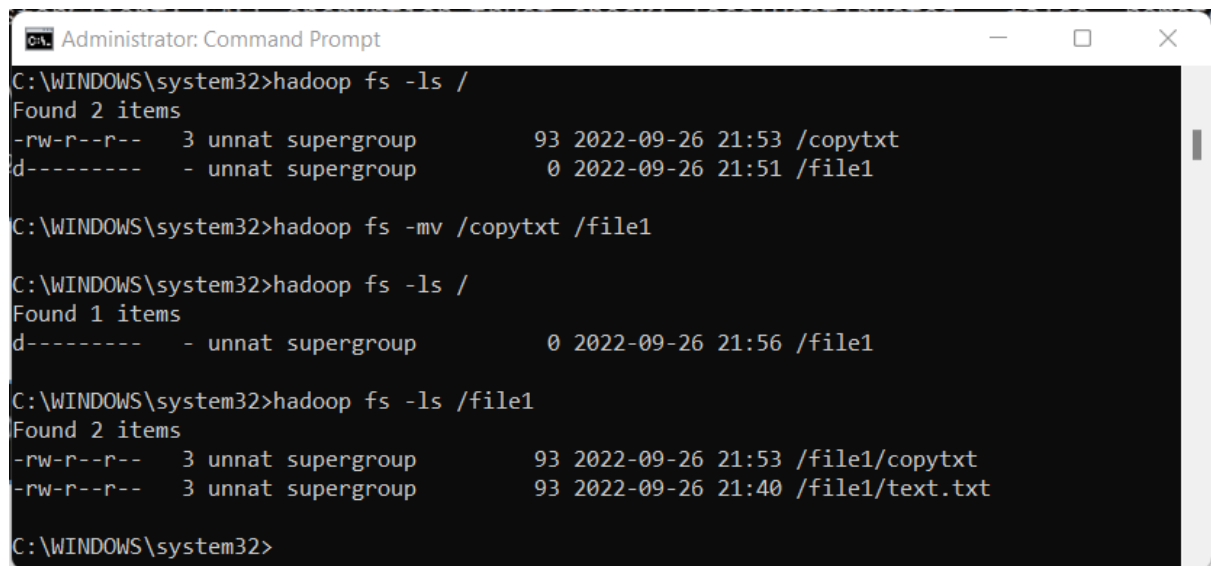
C:\WINDOWS\system32>hadoop fs -cp /folder_1/text.txt /copytxt
2022-09-26 21:53:51,390 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false
2022-09-26 21:53:51,450 INFO sasl.SaslDataTransferClient: SASL encryption trust check: localHostTrusted = false, remoteHostTrusted = false

Administrator: Command Prompt

C:\WINDOWS\system32>hadoop fs -ls /copytxt
-rw-r--r--   3 unnat supergroup    93 2022-09-26 21:53 /copytxt

C:\WINDOWS\system32>
```

11. mv : moves files from source to destination.



```
Administrator: Command Prompt
C:\WINDOWS\system32>hadoop fs -ls /
Found 2 items
-rw-r--r--   3 unnat supergroup          93 2022-09-26 21:53 /copytxt
d-----    - unnat supergroup           0 2022-09-26 21:51 /file1

C:\WINDOWS\system32>hadoop fs -mv /copytxt /file1

C:\WINDOWS\system32>hadoop fs -ls /
Found 1 items
d-----    - unnat supergroup           0 2022-09-26 21:56 /file1

C:\WINDOWS\system32>hadoop fs -ls /file1
Found 2 items
-rw-r--r--   3 unnat supergroup          93 2022-09-26 21:53 /file1/copytxt
-rw-r--r--   3 unnat supergroup          93 2022-09-26 21:40 /file1/text.txt

C:\WINDOWS\system32>
```

Conclusion :

In this practical, we setup a single node Hadoop cluster and applied HDFS commands on single node Hadoop cluster.