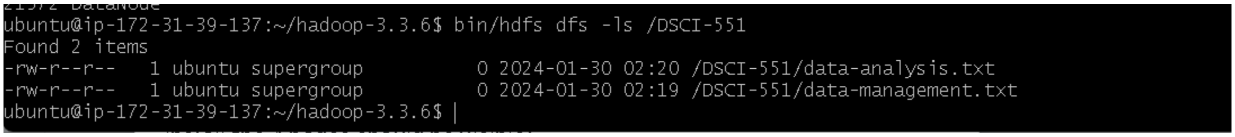
Lab 2: EC2
DSCI 551 – Spring 2024

For each of the following HDFS dfs commands, submit a screenshot showing an example of the command and also explain what the command does.

- ls

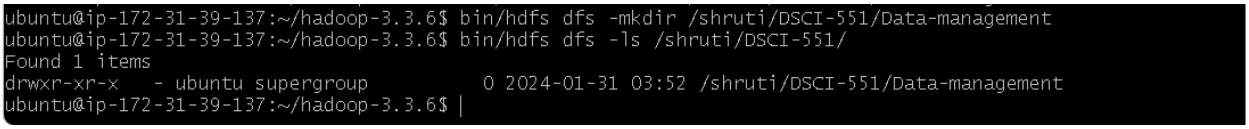
Example: bin/hdfs dfs -ls /DSCI-551
Description: The above command lists the contents of the specified directory in HDFS. It provides details such as file permissions, ownership, size, modification time, and the name of each file or subdirectory in the given path. In this example, it lists the contents of the /DSCI-551 folder.



```
21372 DataNode
ubuntu@ip-172-31-39-137:~/hadoop-3.3.6$ bin/hdfs dfs -ls /DSCI-551
Found 2 items
-rw-r--r--   1 ubuntu supergroup          0 2024-01-30 02:20 /DSCI-551/data-analysis.txt
-rw-r--r--   1 ubuntu supergroup          0 2024-01-30 02:19 /DSCI-551/data-management.txt
ubuntu@ip-172-31-39-137:~/hadoop-3.3.6$ |
```

- mkdir

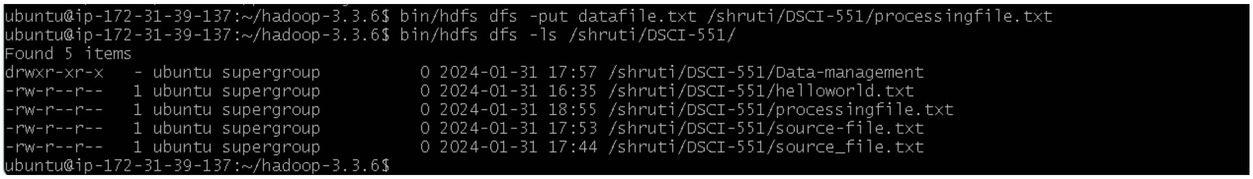
Example: bin/hdfs dfs -mkdir /shruti/DSCI-551/Data-management
Description: The above command creates a new directory named Data-management within the /shruti/DSCI-551 directory in HDFS. It is useful for organizing data by creating a directory structure that fits our needs.



```
ubuntu@ip-172-31-39-137:~/hadoop-3.3.6$ bin/hdfs dfs -mkdir /shruti/DSCI-551/Data-management
ubuntu@ip-172-31-39-137:~/hadoop-3.3.6$ bin/hdfs dfs -ls /shruti/DSCI-551/
Found 1 items
drwxr-xr-x   - ubuntu supergroup          0 2024-01-31 03:52 /shruti/DSCI-551/Data-management
ubuntu@ip-172-31-39-137:~/hadoop-3.3.6$ |
```

- put

Example: bin/hdfs dfs -put datafile.txt /shruti/DSCI-551/processingfile.txt
Description: The above command copies the local file datafile.txt to HDFS, saving it as processingfile.txt in the /shruti/DSCI-551/directory. It helps in transferring data from your local machine to the Hadoop Distributed File System for processing by distributed computing frameworks.

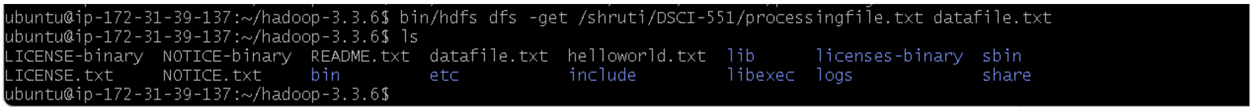


```
ubuntu@ip-172-31-39-137:~/hadoop-3.3.6$ bin/hdfs dfs -put datafile.txt /shruti/DSCI-551/processingfile.txt
ubuntu@ip-172-31-39-137:~/hadoop-3.3.6$ bin/hdfs dfs -ls /shruti/DSCI-551/
Found 5 items
drwxr-xr-x   - ubuntu supergroup          0 2024-01-31 17:57 /shruti/DSCI-551/Data-management
-rw-r--r--   1 ubuntu supergroup          0 2024-01-31 16:35 /shruti/DSCI-551/helloworld.txt
-rw-r--r--   1 ubuntu supergroup          0 2024-01-31 18:55 /shruti/DSCI-551/processingfile.txt
-rw-r--r--   1 ubuntu supergroup          0 2024-01-31 17:53 /shruti/DSCI-551/source-file.txt
-rw-r--r--   1 ubuntu supergroup          0 2024-01-31 17:44 /shruti/DSCI-551/source_file.txt
ubuntu@ip-172-31-39-137:~/hadoop-3.3.6$
```

- get

Example: bin/hdfs dfs -get /shruti/DSCI-551/processingfile.txt datafile.txt

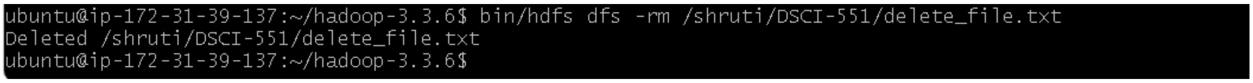
Description: The above command copies the file processingfile.txt from HDFS to the local file system, saving it as datafile.txt in the current directory. The get command is the reverse of put. This is useful when you need to retrieve results or data generated by Hadoop jobs.



```
ubuntu@ip-172-31-39-137:~/hadoop-3.3.6$ bin/hdfs dfs -get /shruti/DSCI-551/processingfile.txt datafile.txt
ubuntu@ip-172-31-39-137:~/hadoop-3.3.6$ ls
LICENSE-binary  NOTICE-binary  README.txt  datafile.txt  helloworld.txt  lib      licenses-binary  sbin
LICENSE.txt     NOTICE.txt     bin         etc           include         libexec  logs             share
ubuntu@ip-172-31-39-137:~/hadoop-3.3.6$
```

- rm

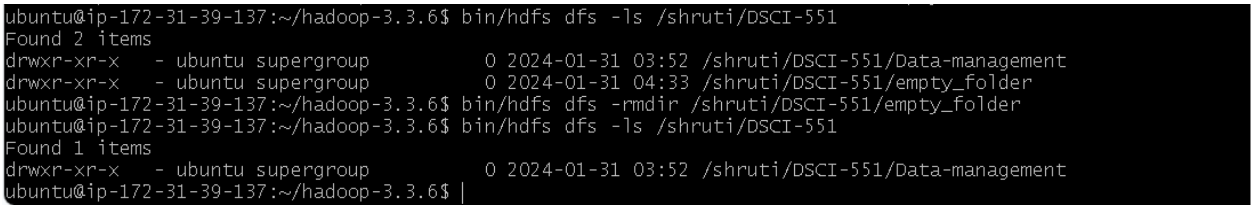
Example: bin/hdfs dfs -rm /shruti/DSCI-551/delete_file.txt
Description: The above command removes the specified file from HDFS. In this example, it deletes the file delete_file.txt. It helps in managing data by allowing users to delete unnecessary files or clean up after processing.



```
ubuntu@ip-172-31-39-137:~/hadoop-3.3.6$ bin/hdfs dfs -rm /shruti/DSCI-551/delete_file.txt
Deleted /shruti/DSCI-551/delete_file.txt
ubuntu@ip-172-31-39-137:~/hadoop-3.3.6$
```

- rmdir

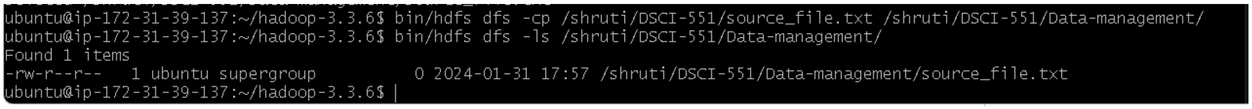
Example: bin/hdfs dfs -rmdir /shruti/DSCI-551/empty_folder
Description: Removes the empty directory named empty_folder from the /shruti/DSCI-551/ directory in HDFS. It's important to note that this command can only remove directories that do not contain any files or subdirectories.



```
ubuntu@ip-172-31-39-137:~/hadoop-3.3.6$ bin/hdfs dfs -ls /shruti/DSCI-551
Found 2 items
drwxr-xr-x   - ubuntu supergroup          0 2024-01-31 03:52 /shruti/DSCI-551/Data-management
drwxr-xr-x   - ubuntu supergroup          0 2024-01-31 04:33 /shruti/DSCI-551/empty_folder
ubuntu@ip-172-31-39-137:~/hadoop-3.3.6$ bin/hdfs dfs -rmdir /shruti/DSCI-551/empty_folder
ubuntu@ip-172-31-39-137:~/hadoop-3.3.6$ bin/hdfs dfs -ls /shruti/DSCI-551
Found 1 items
drwxr-xr-x   - ubuntu supergroup          0 2024-01-31 03:52 /shruti/DSCI-551/Data-management
ubuntu@ip-172-31-39-137:~/hadoop-3.3.6$ |
```

- cp

Example: bin/hdfs dfs -cp /shruti/DSCI-551/source_file.txt /shruti/DSCI-551/Data-management/
Description: The above command copies the file source_file.txt to the destination directory named Data-management within HDFS. It's useful for duplicating data within Hadoop, and it supports copying data across different directories or even different HDFS clusters.

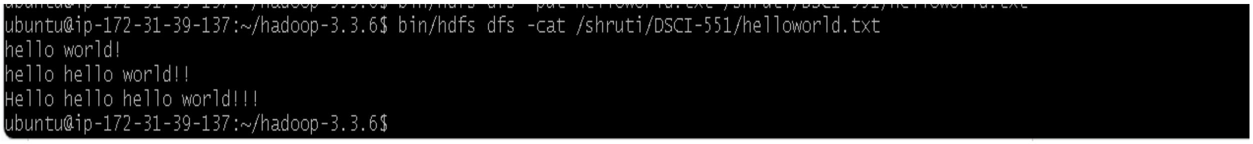


```
ubuntu@ip-172-31-39-137:~/hadoop-3.3.6$ bin/hdfs dfs -cp /shruti/DSCI-551/source_file.txt /shruti/DSCI-551/Data-management/
ubuntu@ip-172-31-39-137:~/hadoop-3.3.6$ bin/hdfs dfs -ls /shruti/DSCI-551/Data-management/
Found 1 items
-rw-r--r--   1 ubuntu supergroup          0 2024-01-31 17:57 /shruti/DSCI-551/Data-management/source_file.txt
ubuntu@ip-172-31-39-137:~/hadoop-3.3.6$ |
```

- cat

Example: bin/hdfs dfs -cat /shruti/DSCI-551/helloworld.txt

Description: The above command displays the content of the specified file in HDFS to the console. In the example above, it displays the content of the helloworld.txt file. It's similar to the Unix cat command and is useful for quickly inspecting the contents of small files.



```
ubuntu@ip-172-31-39-137:~/hadoop-3.3.6$ bin/hdfs dfs -put helloworld.txt /shruti/DSCI-551/helloworld.txt
ubuntu@ip-172-31-39-137:~/hadoop-3.3.6$ bin/hdfs dfs -cat /shruti/DSCI-551/helloworld.txt
hello world!
hello hello world!!
Hello hello hello world!!!
ubuntu@ip-172-31-39-137:~/hadoop-3.3.6$
```