Task-1 : HDFS Commands

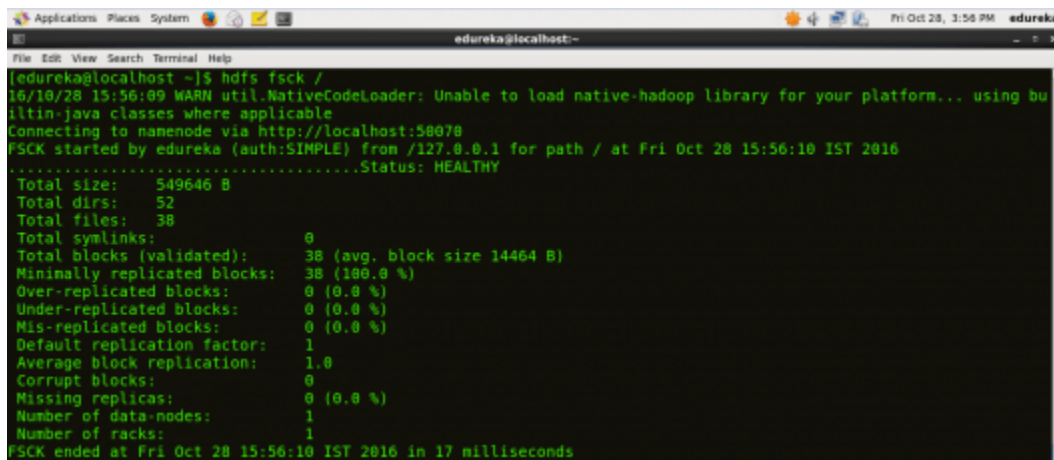- o   wget https://archive.apache.org/dist/hadoop/core/hadoop-2.7.3/hadoop-2.7.3.tar.gz
- o   tar -xvf hadoop-2.7.3.tar.gz

HDFS Commands

• fsck

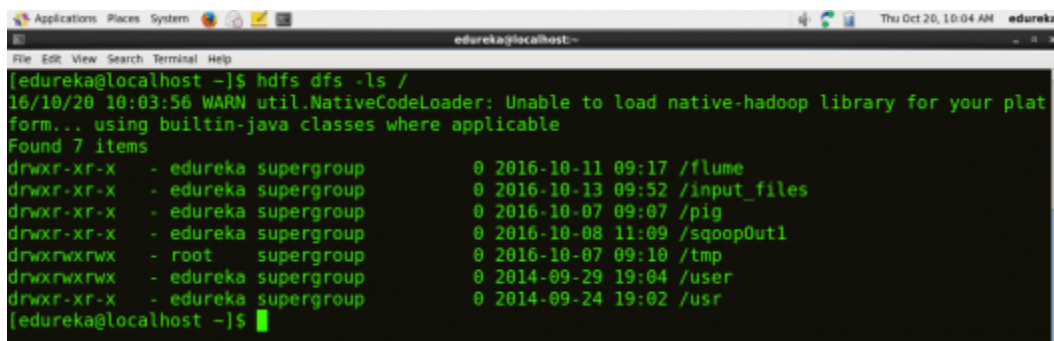HDFS Command to check the health of the Hadoop file system.

Command: **hdfs fsck /**

• ls

HDFS Command to display the list of Files and Directories in HDFS.

Command: **hdfs dfs –ls /**
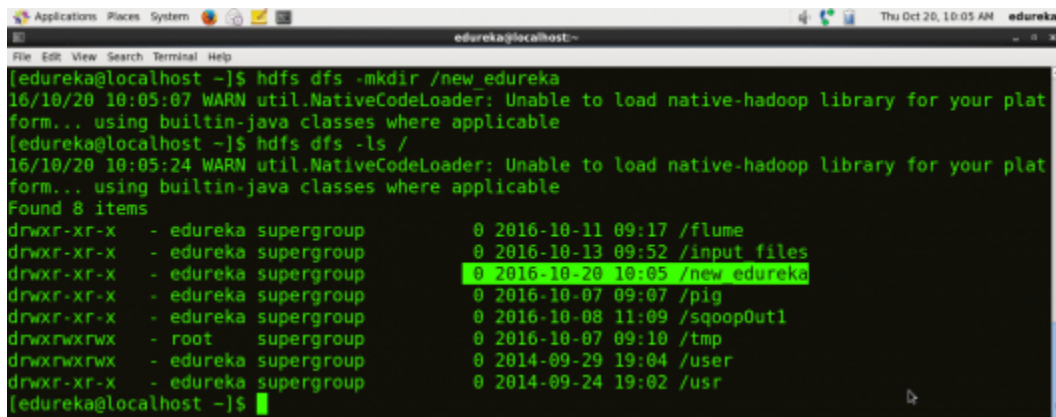
- ## mkdir

HDFS Command to create the directory in HDFS.

Usage: **hdfs dfs –mkdir /directory_name**

Command: **hdfs dfs –mkdir /new_edureka**

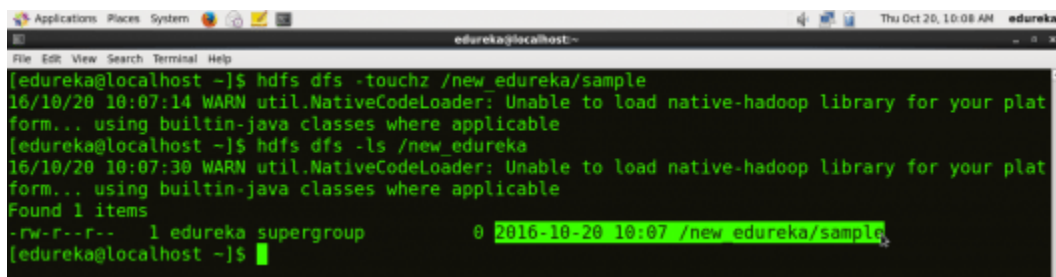*Note:* *Here we are trying to create a directory named "new_edureka" in HDFS.*

You can get a better understanding with the Azure Data Engineering certification.

- ## touchz

HDFS Command to create a file in HDFS with file size 0 bytes.

Usage: **hdfs dfs –touchz /directory/filename**

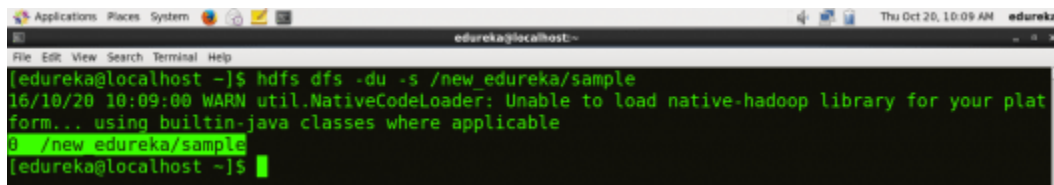*Command:* **hdfs dfs –touchz /new_edureka/sample**

Note: *Here we are trying to create a file named "sample" in the directory "new_edureka" of hdfs with file size 0 bytes.*

- ## <u>du</u>

HDFS Command to check the file size.

Usage: **hdfs dfs –du –s /directory/filename**

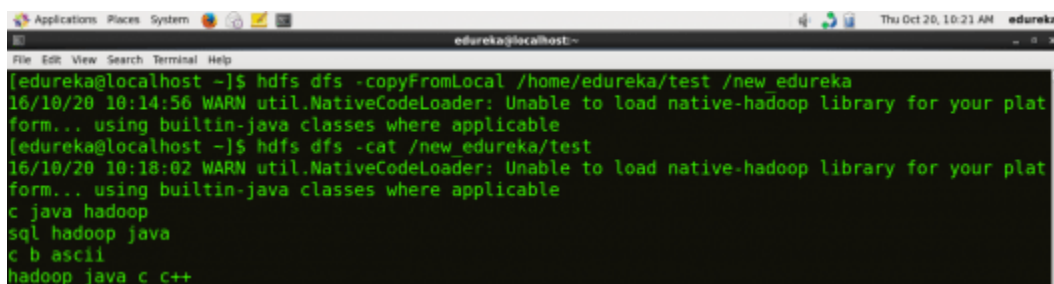*Command:* **hdfs dfs –du –s /new_edureka/sample**

- ## <u>cat</u>

HDFS Command that reads a file on HDFS and prints the content of that file to the standard output.

Usage: **hdfs dfs –cat /path/to/file_in_hdfs**
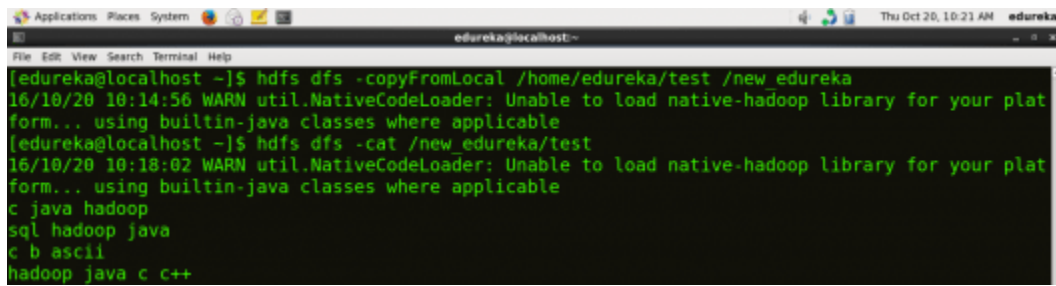
Command: **hdfs dfs –cat /new_edureka/test**

- ## <u>copyFromLocal</u>

HDFS Command to copy the file from a Local file system to HDFS.

Usage: **hdfs dfs -copyFromLocal <localsrc> <hdfs destination>**

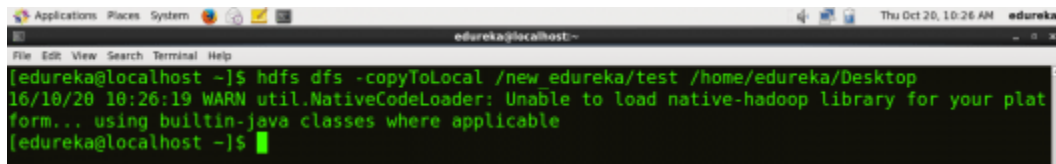*Command:* **hdfs dfs –copyFromLocal /home/edureka/test /new_edureka**

Note: *Here the test is the file present in the local directory /home/edureka and after the command gets executed the test file will be copied in /new_edureka directory of HDFS.*

- ## copyToLocal

HDFS Command to copy the file from HDFS to Local File System.

Usage: **hdfs dfs -copyToLocal <hdfs source> <localdst>**

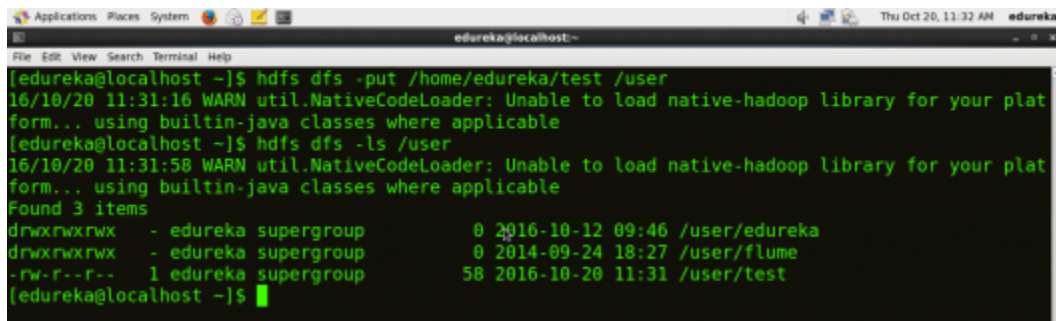*Command:* **hdfs dfs –copyToLocal /new_edureka/test /home/edureka**

Note: *Here test is a file present in the new_edureka directory of HDFS and after the command gets executed the test file will be copied to local directory /home/edureka*

- ## put

HDFS Command to copy single source or multiple sources from local file system to the destination file system.

Usage: **hdfs dfs -put <localsrc> <destination>**

Command: **hdfs dfs –put /home/edureka/test /user**

Note: *The command copyFromLocal is similar to put command, except that the source is restricted to a local file reference.*
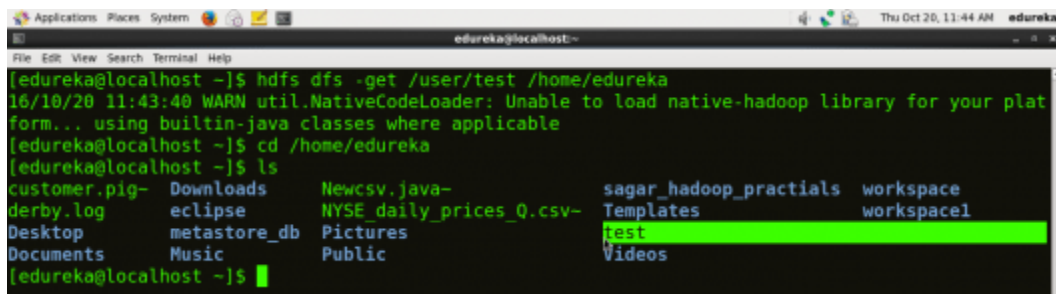
You can even check out the details of Big Data with the Data Engineering Certification in Canada.

- ## **get**

HDFS Command to copy files from hdfs to the local file system.

Usage: **hdfs dfs -get <src> <localdst>**

Command: **hdfs dfs –get /user/test /home/edureka**

Note: *The command copyToLocal is similar to get command, except that the destination is restricted to a local file reference.*
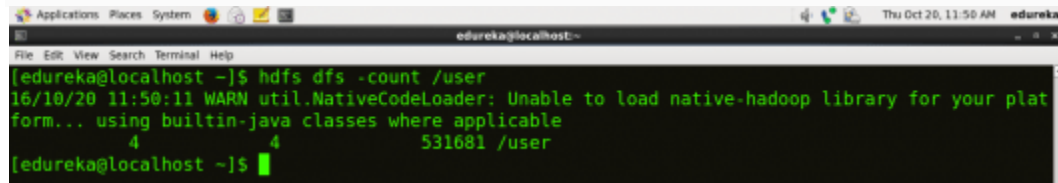
- ## **count**

HDFS Command to count the number of directories, files, and bytes under the paths that match the specified file pattern.

Usage: **hdfs dfs -count <path>**
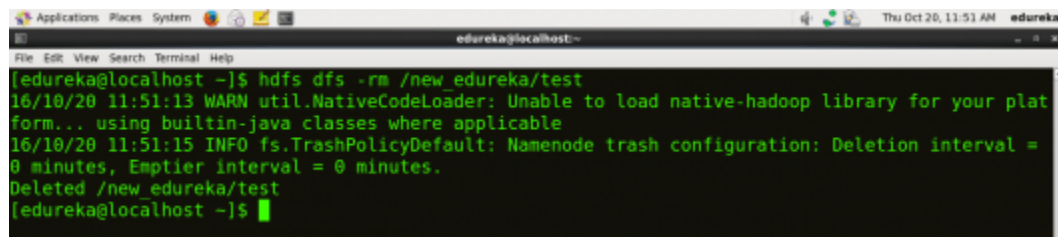
Command: **hdfs dfs –count /user**

- ## rm

HDFS Command to remove the file from HDFS.

Usage: **hdfs dfs –rm <path>**

Command: **hdfs dfs –rm /new_edureka/test**

- ## rm -r

HDFS Command to remove the entire directory and all of its content from HDFS.

Usage: **hdfs dfs -rm -r <path>**
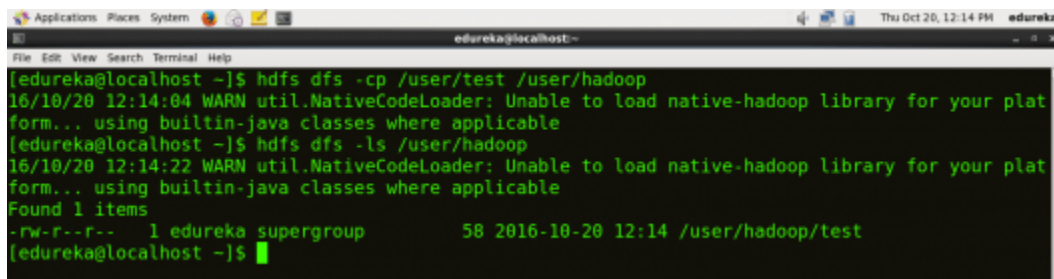
Command: **hdfs dfs -rm -r /new_edureka**

- ## cp

HDFS Command to copy files from source to destination. This command allows multiple sources as well, in which case the destination must be a directory.

Usage: **hdfs dfs -cp <src> <dest>**

Command: **hdfs dfs -cp /user/hadoop/file1 /user/hadoop/file2**

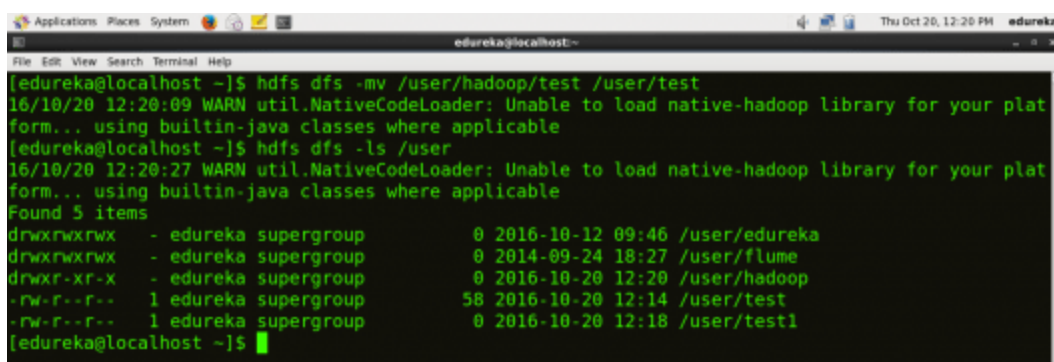*Command:* **hdfs dfs -cp /user/hadoop/file1 /user/hadoop/file2 /user/hadoop/dir**

- ## mv

HDFS Command to move files from source to destination. This command allows multiple sources as well, in which case the destination needs to be a directory.
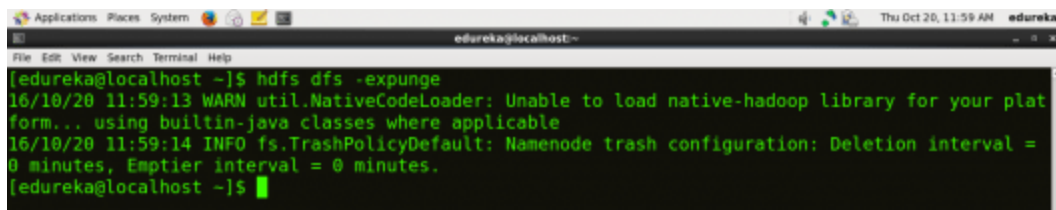
Usage: **hdfs dfs -mv <src> <dest>**

Command: **hdfs dfs -mv /user/hadoop/file1 /user/hadoop/file2**

- ## expunge

HDFS Command that makes the trash empty.

Command: hdfs dfs -expunge
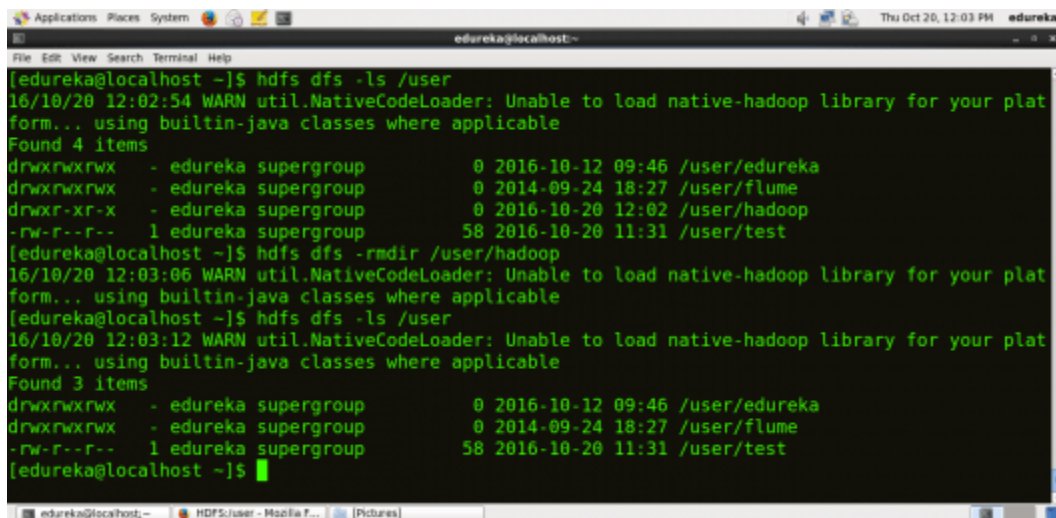
• rmdir

HDFS Command to remove the directory.

Usage: **hdfs dfs -rmdir <path>**

Command: **hdfs dfs –rmdir /user/hadoop**

• usage

HDFS Command that returns the help for an individual command.

Usage: **hdfs dfs -usage <command>**

Command: **hdfs dfs -usage mkdir**
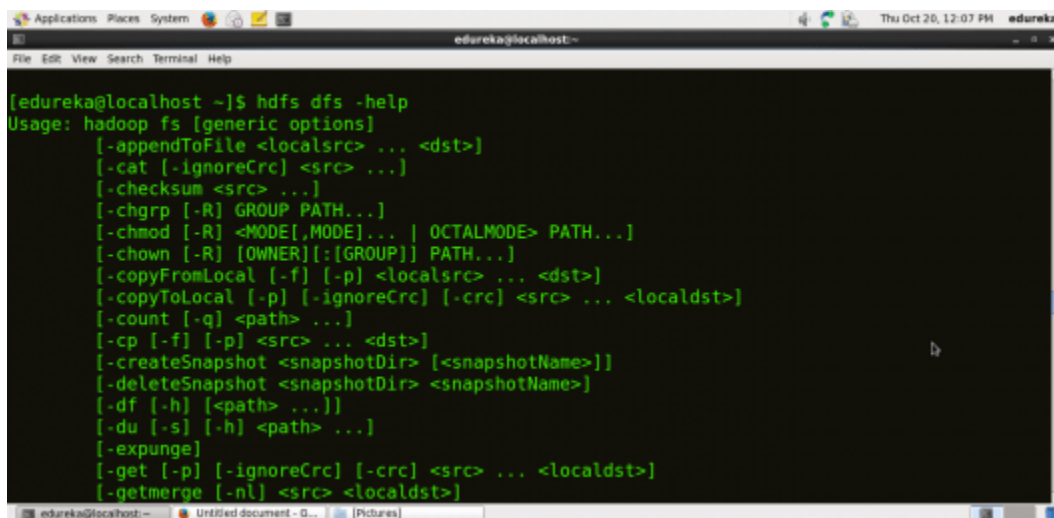
Note: *By using usage command you can get information about any command.*

- # <u>help</u>

HDFS Command that displays help for given command or all commands if none is specified.

Command: **hdfs dfs -help**