

Task-1 : HDFS Commands

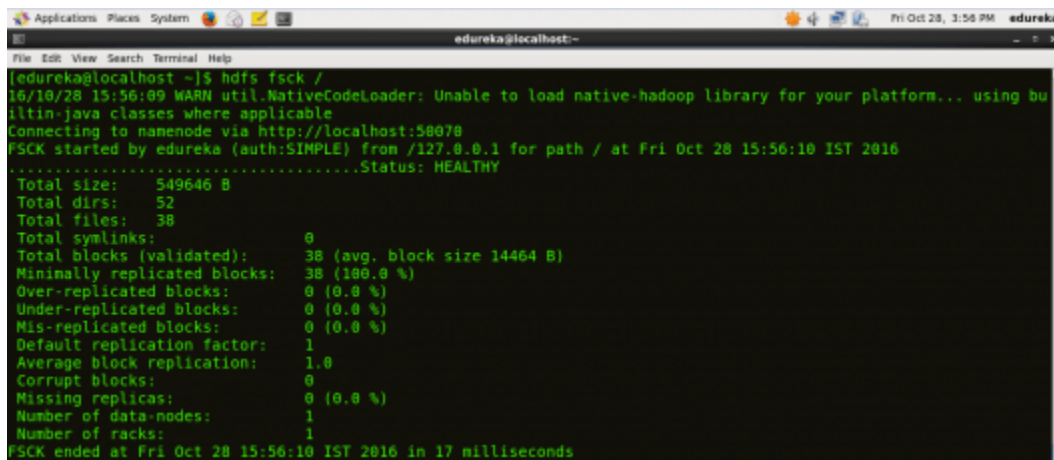
- wget <https://archive.apache.org/dist/hadoop/core/hadoop-2.7.3/hadoop-2.7.3.tar.gz>
- tar -xvf hadoop-2.7.3.tar.gz

HDFS Commands

• fsck

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

Command: `hdfs fsck /`

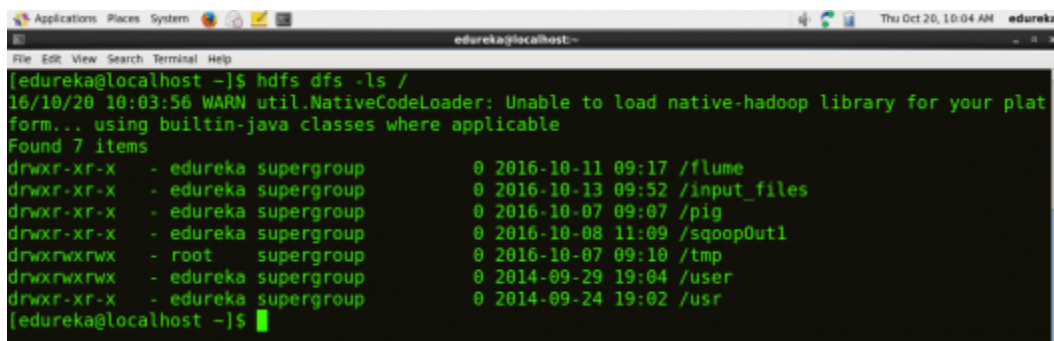


```
[edureka@localhost ~]$ hdfs fsck /
16/10/28 15:56:09 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Connecting to namenode via http://localhost:58070
FSCK started by edureka (auth:SIMPLE) from /127.0.0.1 for path / at Fri Oct 28 15:56:10 IST 2016
.....Status: HEALTHY
Total size: 549646 B
Total dirs: 52
Total files: 38
Total symlinks: 0
Total blocks (validated): 38 (avg. block size 14464 B)
Minimally replicated blocks: 38 (100.0 %)
Over-replicated blocks: 0 (0.0 %)
Under-replicated blocks: 0 (0.0 %)
Mis-replicated blocks: 0 (0.0 %)
Default replication factor: 1
Average block replication: 1.0
Corrupt blocks: 0
Missing replicas: 0 (0.0 %)
Number of data-nodes: 1
Number of racks: 1
FSCK ended at Fri Oct 28 15:56:10 IST 2016 in 17 milliseconds
```

• ls

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

Command: `hdfs dfs -ls /`



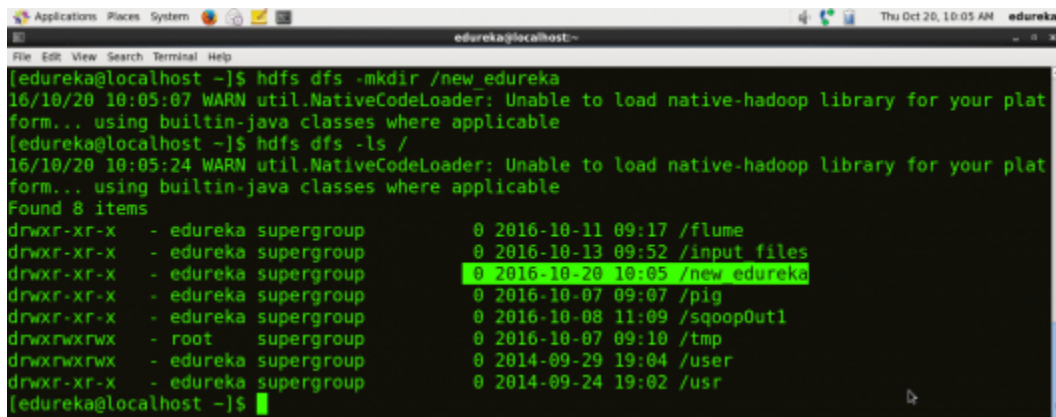
```
[edureka@localhost ~]$ hdfs dfs -ls /
16/10/28 10:03:56 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 7 items
drwxr-xr-x - edureka supergroup 0 2016-10-11 09:17 /flume
drwxr-xr-x - edureka supergroup 0 2016-10-13 09:52 /input_files
drwxr-xr-x - edureka supergroup 0 2016-10-07 09:07 /pig
drwxr-xr-x - edureka supergroup 0 2016-10-08 11:09 /sqoopOut1
drwxrwxrwx - root supergroup 0 2016-10-07 09:10 /tmp
drwxrwxrwx - edureka supergroup 0 2014-09-29 19:04 /user
drwxr-xr-x - edureka supergroup 0 2014-09-24 19:02 /usr
[edureka@localhost ~]$
```

- **mkdir**

HDFS Command to create the directory in HDFS.

Usage: `hdfs dfs -mkdir /directory_name`

Command: `hdfs dfs -mkdir /new_edureka`



```
[edureka@localhost ~]$ hdfs dfs -mkdir /new_edureka
16/10/20 10:05:07 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your plat
form... using builtin-java classes where applicable
[edureka@localhost ~]$ hdfs dfs -ls /
16/10/20 10:05:24 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your plat
form... using builtin-java classes where applicable
Found 8 items
drwxr-xr-x - edureka supergroup      0 2016-10-11 09:17 /flume
drwxr-xr-x - edureka supergroup      0 2016-10-13 09:52 /input_files
drwxr-xr-x - edureka supergroup      0 2016-10-20 10:05 /new_edureka
drwxr-xr-x - edureka supergroup      0 2016-10-07 09:07 /pig
drwxr-xr-x - edureka supergroup      0 2016-10-08 11:09 /sqoopOut1
drwxrwxrwx - root supergroup         0 2016-10-07 09:10 /tmp
drwxrwxrwx - edureka supergroup      0 2014-09-29 19:04 /user
drwxr-xr-x - edureka supergroup      0 2014-09-24 19:02 /usr
[edureka@localhost ~]$
```

Note: Here we are trying to create a directory named “new_edureka” in HDFS.

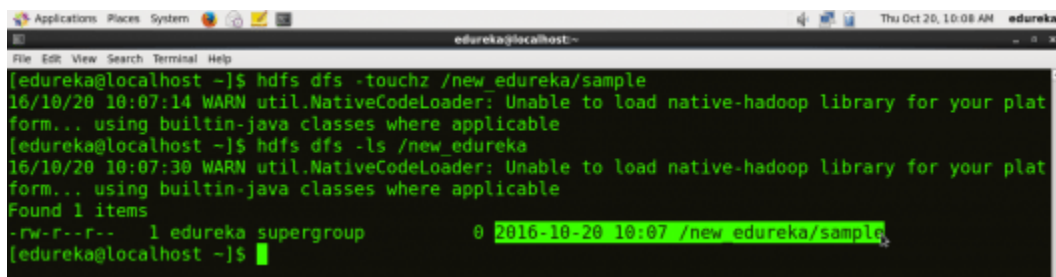
You can get a better understanding with the [Azure Data Engineering certification](#).

- **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`



```
[edureka@localhost ~]$ hdfs dfs -touchz /new_edureka/sample
16/10/20 10:07:14 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your plat
form... using builtin-java classes where applicable
[edureka@localhost ~]$ hdfs dfs -ls /new_edureka
16/10/20 10:07:30 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your plat
form... using builtin-java classes where applicable
Found 1 items
-rw-r--r-- 1 edureka supergroup      0 2016-10-20 10:07 /new_edureka/sample
[edureka@localhost ~]$
```

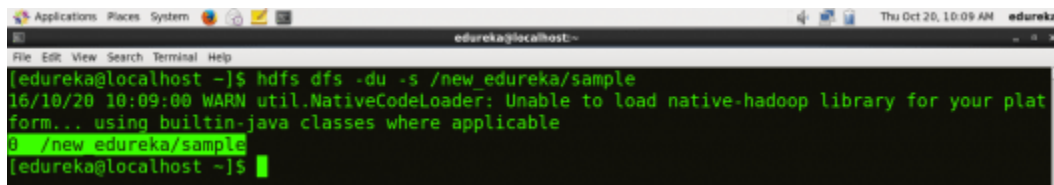
Note: Here we are trying to create a file named “sample” in the directory “new_edureka” of hdfs with file size 0 bytes.

- **du**

HDFS Command to check the file size.

Usage: `hdfs dfs -du -s /directory/filename`

Command: `hdfs dfs -du -s /new_edureka/sample`



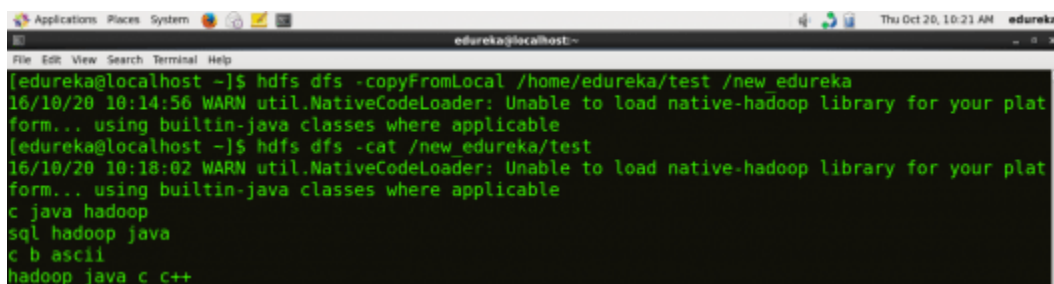
```
Applications Places System Thu Oct 20, 10:09 AM edureka
edureka@localhost:~
File Edit View Search Terminal Help
[edureka@localhost ~]$ hdfs dfs -du -s /new_edureka/sample
16/10/20 10:09:00 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your plat
form... using builtin-java classes where applicable
0 /new_edureka/sample
[edureka@localhost ~]$
```

- **cat**

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`



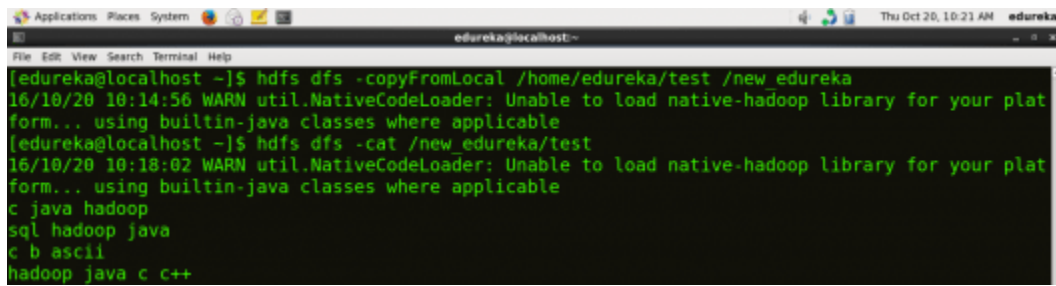
```
Applications Places System Thu Oct 20, 10:21 AM edureka
edureka@localhost:~
File Edit View Search Terminal Help
[edureka@localhost ~]$ hdfs dfs -copyFromLocal /home/edureka/test /new edureka
16/10/20 10:14:56 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your plat
form... using builtin-java classes where applicable
[edureka@localhost ~]$ hdfs dfs -cat /new_edureka/test
16/10/20 10:18:02 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your plat
form... using builtin-java classes where applicable
c java hadoop
sql hadoop java
c b ascii
hadoop java c c++
```

- **copyFromLocal**

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`



```
[edureka@localhost ~]$ hdfs dfs -copyFromLocal /home/edureka/test /new edureka
16/10/20 10:14:56 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your plat
form... using builtin-java classes where applicable
[edureka@localhost ~]$ hdfs dfs -cat /new_edureka/test
16/10/20 10:18:02 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your plat
form... using builtin-java classes where applicable
c java hadoop
sql hadoop java
c b ascii
hadoop java c c++
```

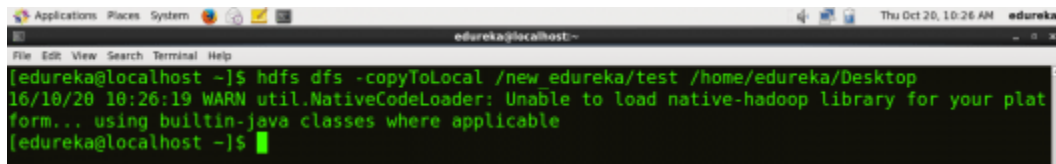
Note: Here the test is the file present in the local directory /home/edureka and after the command gets executed the test file will be copied in /new_edureka directory of HDFS.

- **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`



```
[edureka@localhost ~]$ hdfs dfs -copyToLocal /new edureka/test /home/edureka/Desktop
16/10/20 10:26:19 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your plat
form... using builtin-java classes where applicable
[edureka@localhost ~]$
```

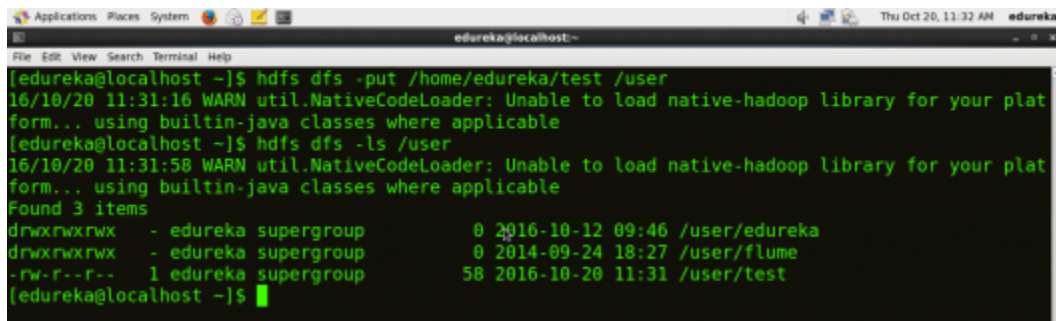
Note: Here test is a file present in the new_edureka directory of HDFS and after the command gets executed the test file will be copied to local directory /home/edureka

- **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`

A terminal window titled 'edureka@localhost:~' showing the execution of HDFS commands. The user runs 'hdfs dfs -put /home/edureka/test /user', which outputs a warning about the native-hadoop library and then lists the contents of '/user'. The output shows three items: 'edureka supergroup' with permissions 'drwxrwxrwx' and size '0' for '/user/edureka', 'edureka supergroup' with permissions 'drwxrwxrwx' and size '0' for '/user/flume', and 'edureka supergroup' with permissions '-rw-r--r--' and size '58' for '/user/test'.

```
[edureka@localhost ~]$ hdfs dfs -put /home/edureka/test /user
16/10/20 11:31:16 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your plat
form... using builtin-java classes where applicable
[edureka@localhost ~]$ hdfs dfs -ls /user
16/10/20 11:31:58 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your plat
form... using builtin-java classes where applicable
Found 3 items
drwxrwxrwx - edureka supergroup          0 2016-10-12 09:46 /user/edureka
drwxrwxrwx - edureka supergroup          0 2014-09-24 18:27 /user/flume
-rw-r--r-- 1 edureka supergroup         58 2016-10-20 11:31 /user/test
[edureka@localhost ~]$
```

Note: The command *copyFromLocal* is similar to *put* command, except that the source is restricted to a local file reference.

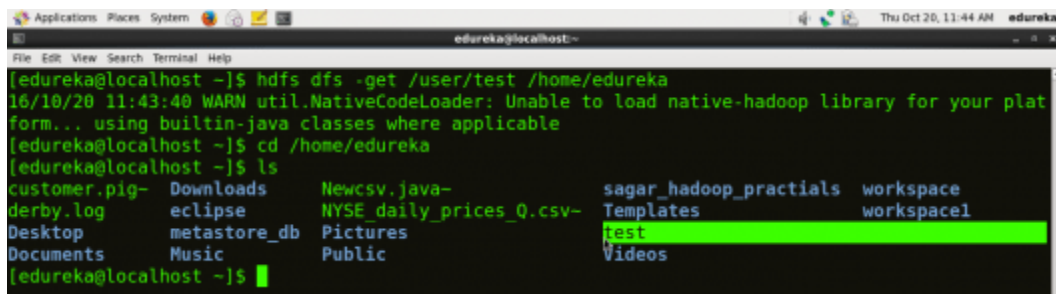
You can even check out the details of Big Data with the [Data Engineering Certification in Canada](#).

- **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`

A terminal window titled 'edureka@localhost:~' showing the execution of the 'hdfs dfs -get' command. The user runs 'hdfs dfs -get /user/test /home/edureka', which outputs a warning about the native-hadoop library. Then, the user runs 'cd /home/edureka' and 'ls', which lists the contents of the local directory. The output shows a list of files and directories, including 'customer.pig', 'derby.log', 'Desktop', 'Documents', 'Downloads', 'eclipse', 'metastore_db', 'Music', 'Newcsv.java-', 'NYSE_daily_prices_Q.csv-', 'Pictures', 'Public', 'sagar_hadoop_practials', 'Templates', 'Videos', 'workspace', and 'workspace1'. The file 'test' is highlighted in green.

```
[edureka@localhost ~]$ hdfs dfs -get /user/test /home/edureka
16/10/20 11:43:40 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your plat
form... using builtin-java classes where applicable
[edureka@localhost ~]$ cd /home/edureka
[edureka@localhost ~]$ ls
customer.pig- Downloads          Newcsv.java-      sagar_hadoop_practials workspace
derby.log      eclipse             NYSE_daily_prices_Q.csv- Templates         workspace1
Desktop        metastore_db       Pictures          test
Documents      Music              Public            Videos
[edureka@localhost ~]$
```

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

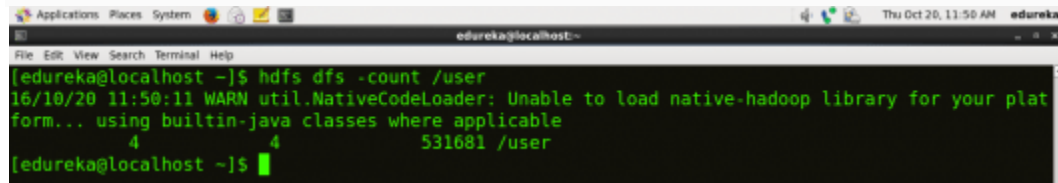
- **count**

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`

Big Data Training



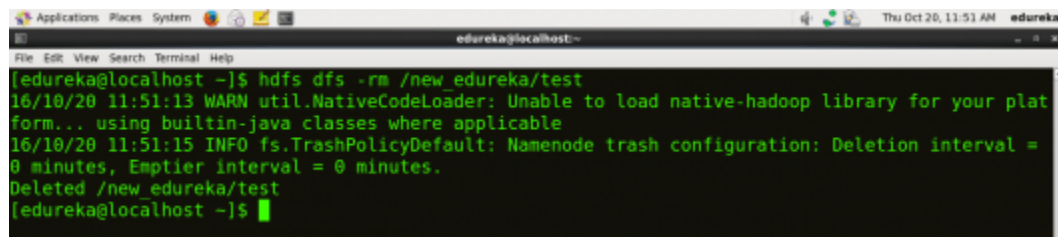
```
Applications Places System edureka@localhost:~
File Edit View Search Terminal Help
[edureka@localhost ~]$ hdfs dfs -count /user
16/10/20 11:50:11 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your plat
form... using builtin-java classes where applicable
      4      4      531681 /user
[edureka@localhost ~]$
```

- **rm**

HDFS Command to remove the file from HDFS.

Usage: `hdfs dfs -rm <path>`

Command: `hdfs dfs -rm /new_edureka/test`



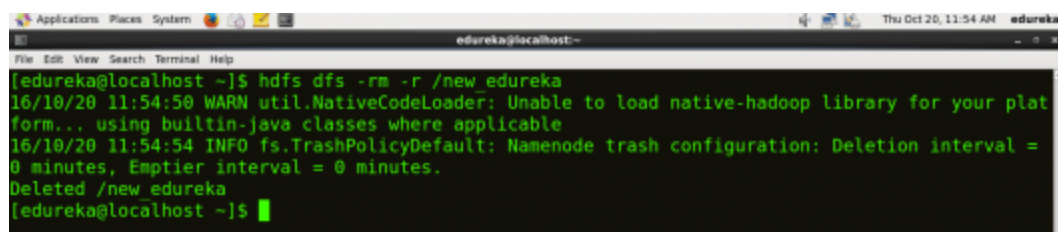
```
Applications Places System edureka@localhost:~
File Edit View Search Terminal Help
[edureka@localhost ~]$ hdfs dfs -rm /new_edureka/test
16/10/20 11:51:13 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your plat
form... using builtin-java classes where applicable
16/10/20 11:51:15 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion interval =
0 minutes, Emptyier interval = 0 minutes.
Deleted /new_edureka/test
[edureka@localhost ~]$
```

- **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`



```
Applications Places System edureka@localhost:~
File Edit View Search Terminal Help
[edureka@localhost ~]$ hdfs dfs -rm -r /new_edureka
16/10/20 11:54:50 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your plat
form... using builtin-java classes where applicable
16/10/20 11:54:54 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion interval =
0 minutes, Emptyier interval = 0 minutes.
Deleted /new_edureka
[edureka@localhost ~]$
```

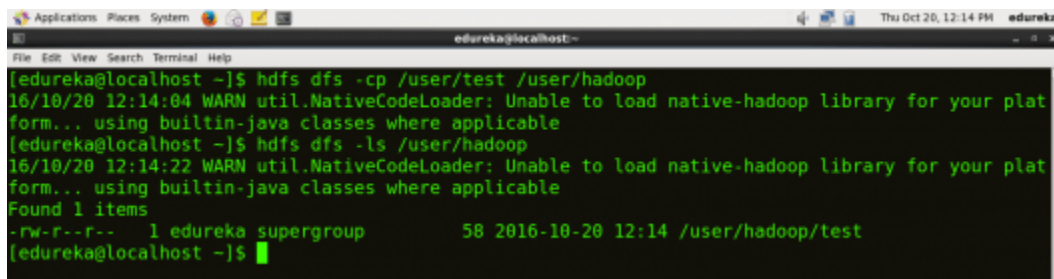
- **cp**

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`

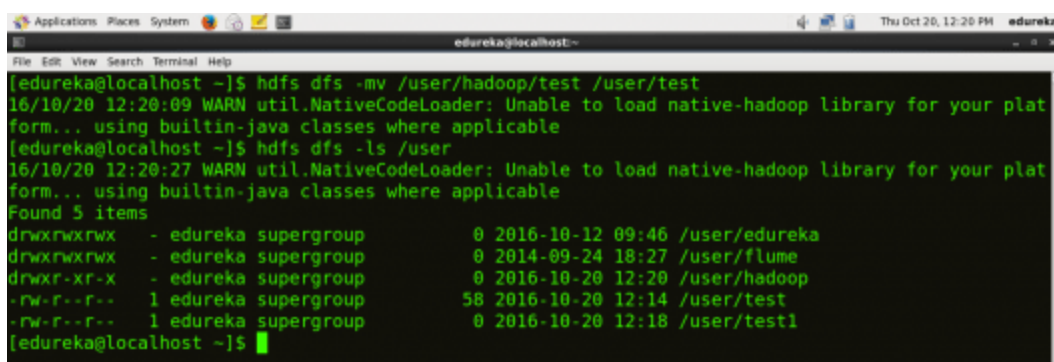
A terminal window titled 'edureka@localhost' showing the execution of HDFS commands. The first command is 'hdfs dfs -cp /user/test /user/hadoop', which results in a warning message from util.NativeCodeLoader. The second command is 'hdfs dfs -ls /user/hadoop', which lists the contents of the directory. The output shows one item: a file named 'test' with permissions '-rw-r--r--', owner 'edureka', group 'supergroup', size '58', and timestamp '2016-10-20 12:14'.

- **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>`

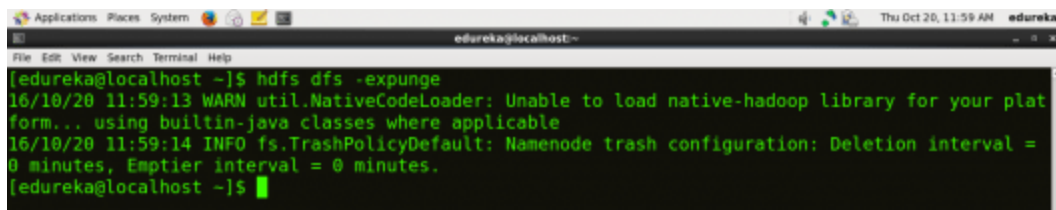
Command: `hdfs dfs -mv /user/hadoop/test /user/test`

A terminal window titled 'edureka@localhost' showing the execution of HDFS commands. The first command is 'hdfs dfs -mv /user/hadoop/test /user/test', which results in a warning message from util.NativeCodeLoader. The second command is 'hdfs dfs -ls /user', which lists the contents of the directory. The output shows five items: 'edureka', 'flume', 'hadoop', 'test', and 'test1'. The 'test' item has permissions '-rw-r--r--', owner 'edureka', group 'supergroup', size '58', and timestamp '2016-10-20 12:14'.

- **expunge**

HDFS Command that makes the trash empty.

Command: `hdfs dfs -expunge`



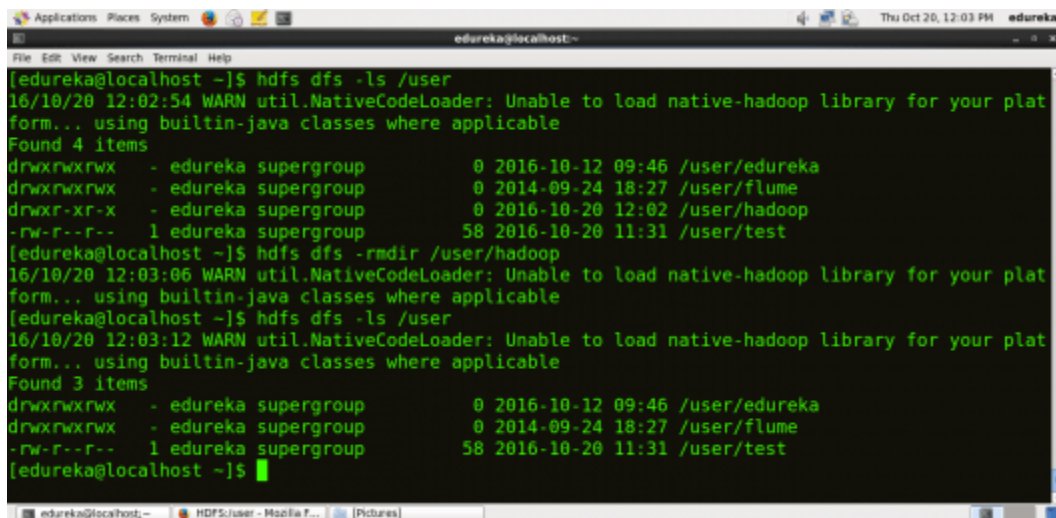
```
[edureka@localhost ~]$ hdfs dfs -expunge
16/10/20 11:59:13 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your plat
form... using builtin-java classes where applicable
16/10/20 11:59:14 INFO fs.TrashPolicyDefault: Namenode trash configuration: Deletion interval =
0 minutes, Empty interval = 0 minutes.
[edureka@localhost ~]$
```

- **rmdir**

HDFS Command to remove the directory.

Usage: `hdfs dfs -rmdir <path>`

Command: `hdfs dfs -rmdir /user/hadoop`



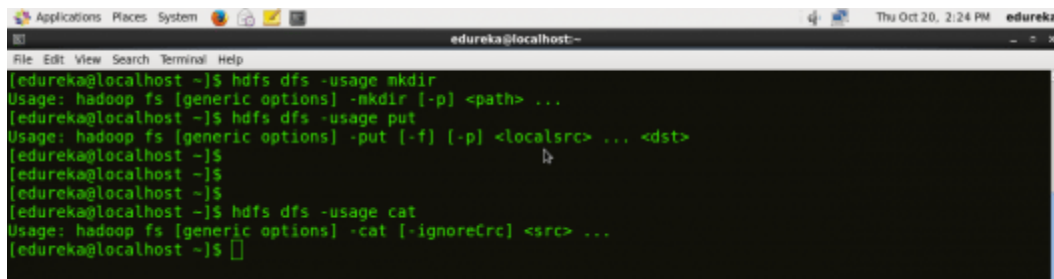
```
[edureka@localhost ~]$ hdfs dfs -ls /user
16/10/20 12:02:54 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your plat
form... using builtin-java classes where applicable
Found 4 items
drwxrwxrwx - edureka supergroup      0 2016-10-12 09:46 /user/edureka
drwxrwxrwx - edureka supergroup      0 2014-09-24 18:27 /user/flume
drwxr-xr-x - edureka supergroup      0 2016-10-20 12:02 /user/hadoop
-rw-r--r-- 1 edureka supergroup     58 2016-10-20 11:31 /user/test
[edureka@localhost ~]$ hdfs dfs -rmdir /user/hadoop
16/10/20 12:03:06 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your plat
form... using builtin-java classes where applicable
[edureka@localhost ~]$ hdfs dfs -ls /user
16/10/20 12:03:12 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your plat
form... using builtin-java classes where applicable
Found 3 items
drwxrwxrwx - edureka supergroup      0 2016-10-12 09:46 /user/edureka
drwxrwxrwx - edureka supergroup      0 2014-09-24 18:27 /user/flume
-rw-r--r-- 1 edureka supergroup     58 2016-10-20 11:31 /user/test
[edureka@localhost ~]$
```

- **usage**

HDFS Command that returns the help for an individual command.

Usage: `hdfs dfs -usage <command>`

Command: `hdfs dfs -usage mkdir`

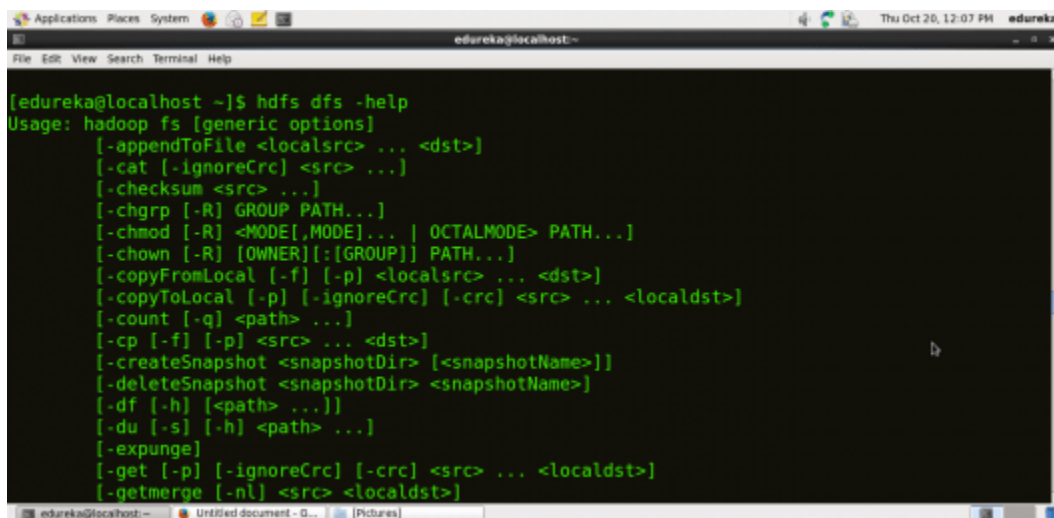
A terminal window titled 'edureka@localhost:~' showing a series of HDFS commands and their usage information. The commands are: 'hdfs dfs -usage mkdir', 'hdfs dfs -usage put', and 'hdfs dfs -usage cat'. Each command is followed by its usage syntax in green text. The window has a menu bar with 'File', 'Edit', 'View', 'Search', 'Terminal', and 'Help'. The system clock shows 'Thu Oct 20, 2:24 PM' and the user is 'edureka'.

Note: By using usage command you can get information about any command.

- **help**

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

Command: hdfs dfs -help

A terminal window titled 'edureka@localhost:~' showing the output of the 'hdfs dfs -help' command. The output lists various HDFS commands and their syntax in green text, including: 'appendToFile', 'cat', 'checksum', 'chgrp', 'chmod', 'chown', 'copyFromLocal', 'copyToLocal', 'count', 'cp', 'createSnapshot', 'deleteSnapshot', 'df', 'du', 'expunge', 'get', and 'getmerge'. The window has a menu bar with 'File', 'Edit', 'View', 'Search', 'Terminal', and 'Help'. The system clock shows 'Thu Oct 20, 12:07 PM' and the user is 'edureka'.