6. Problem Statement

# HadoopSession

A simple session call and recall.

Complete the following Hadoop

exercises in a GCP Dataproc cluster. For each step, capture and insert screenshots to show

the commands and results.

Step

Step 1. Create a Dataproc cluster associating with an existing storage bucket. | Step 2. SSH connect to the master node of the cluster and open a terminal.

You should complete the following steps in the terminal.

Step 3. Use a command to show the Hadoop version.

Step 4. List the content of the HDFS in your cluster.

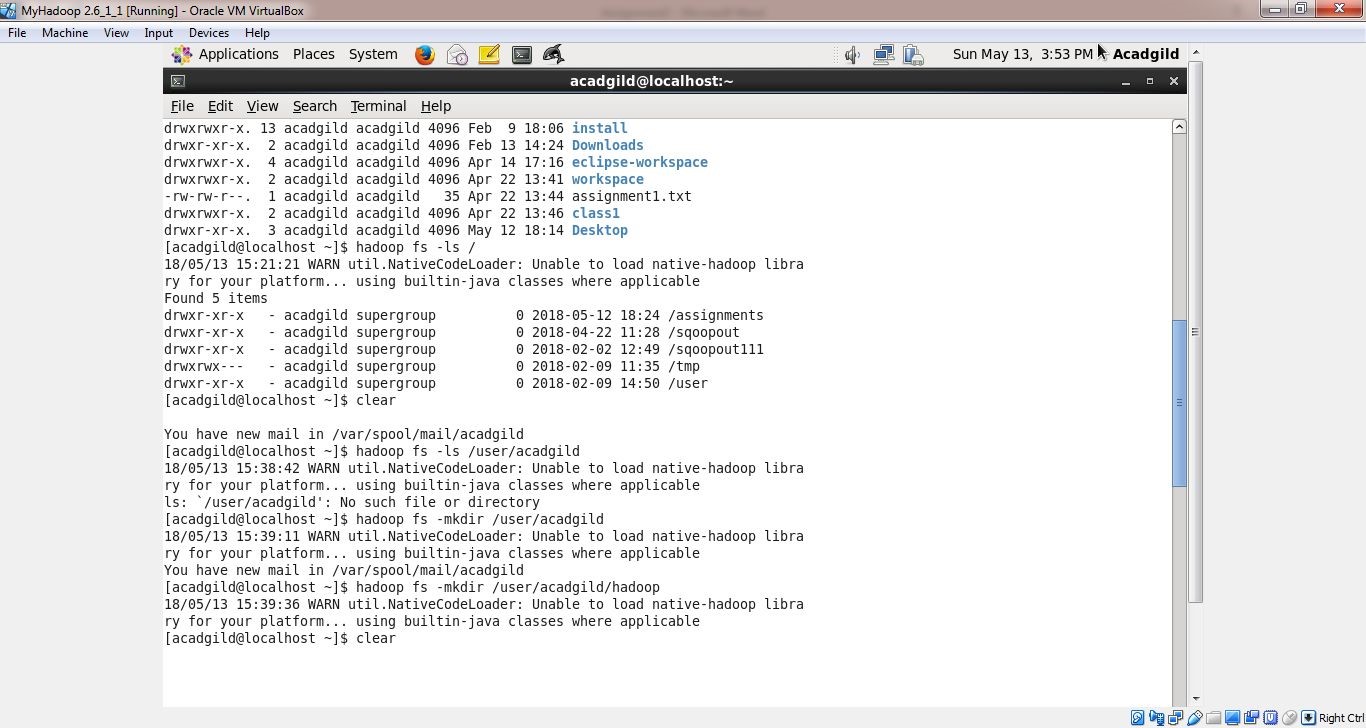
Step 5. Create a folder /user/info323' in the HDFS.

Step 6. Copy a file “test.txt" from the file system on the master node to the HDFS folder ‘info323' you just created.

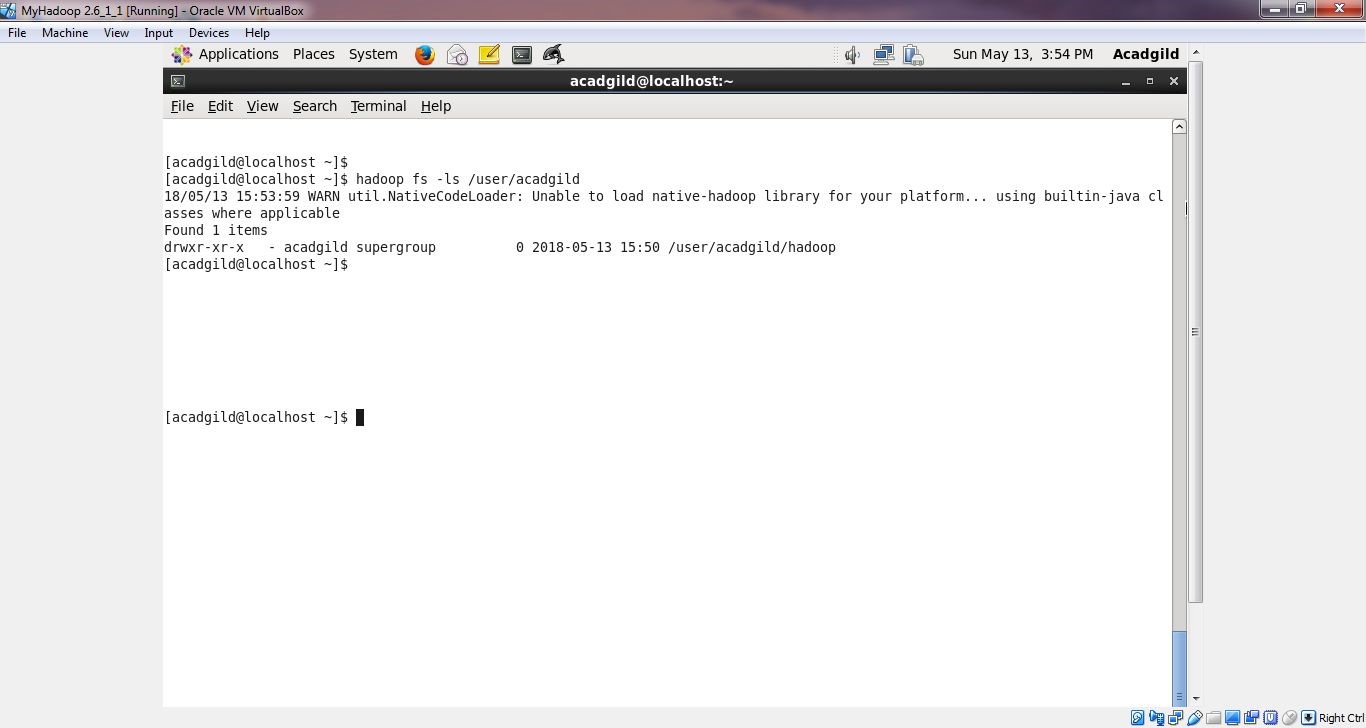
Step 7. Copy the “test.txt file from the HDFS to the home directory on the master node as “test-from-HDFS.txt". Step 8. Delete the file and the folder in the HDFS.

Create a directory for user

Hadoop fs –mkdir creates a directory with the name specified in the arguments.



Hadoop fs –ls lists all files in a directory



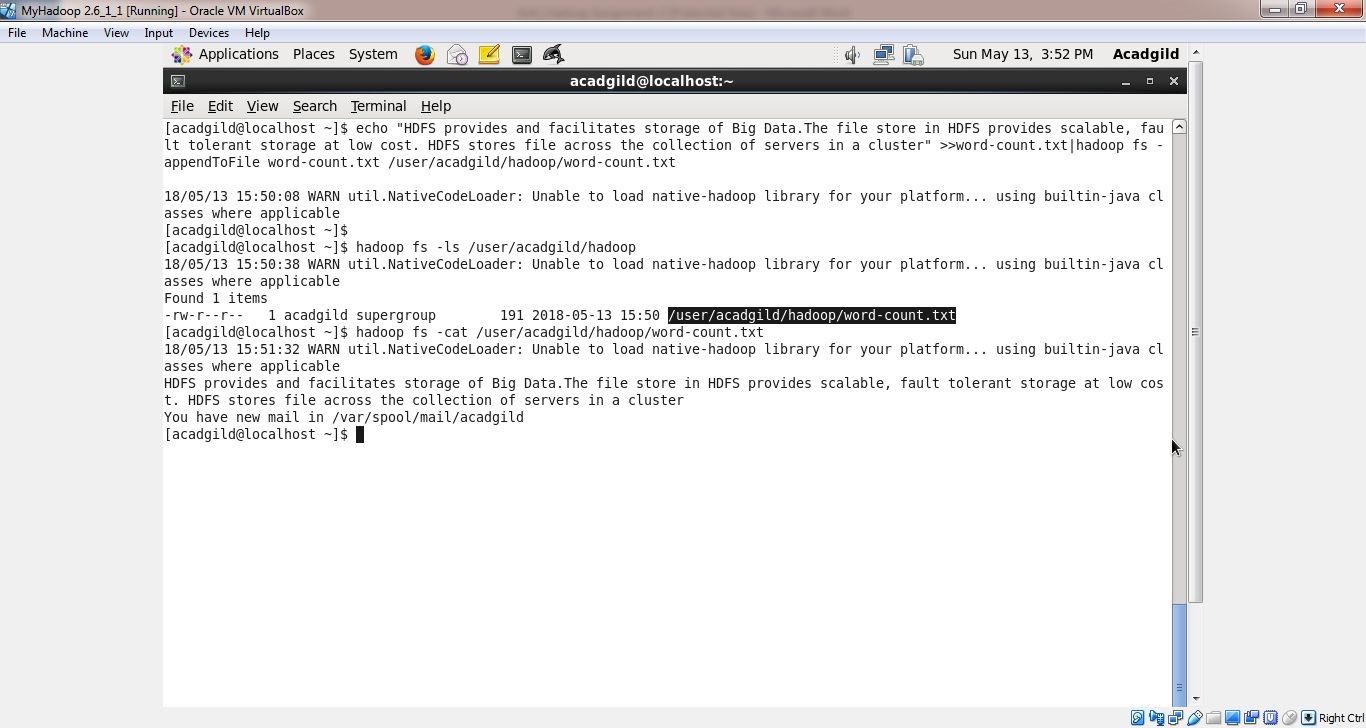
Task 2:

Create a file in HDFS under directory /user/filename/hadoop, with name word-count.txt.

Whatever we type on screen should get appended to the file.

Try to type (on screen) few lines from any online article or textbook.

Output of echo is appended to file word-count.txt and appendToFile command in Hadoop adds the file to Hadoop file system.



Task 3:

Create a file max-temp.txt in local FS.

Put some 10-15 records of date and temperature example: dd-mm-yyyy,temperature

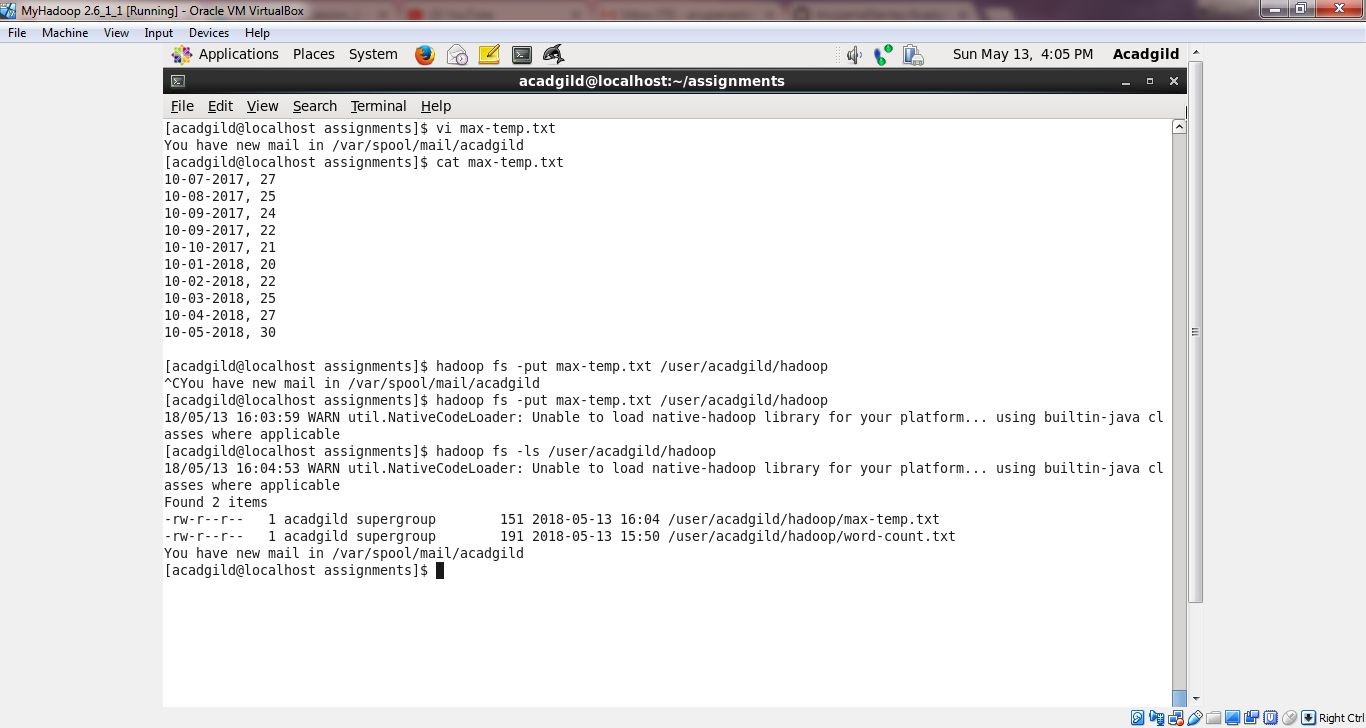
Example:

10-01-1990,10

10-02-1991,20

Move this file to HDFS at /user/acadgild/hadoop.

Created a file max-temp.txt using vi command and using Hadoop fs –put commnd, moved this file to HDFS directory /user/acadgild/Hadoop.

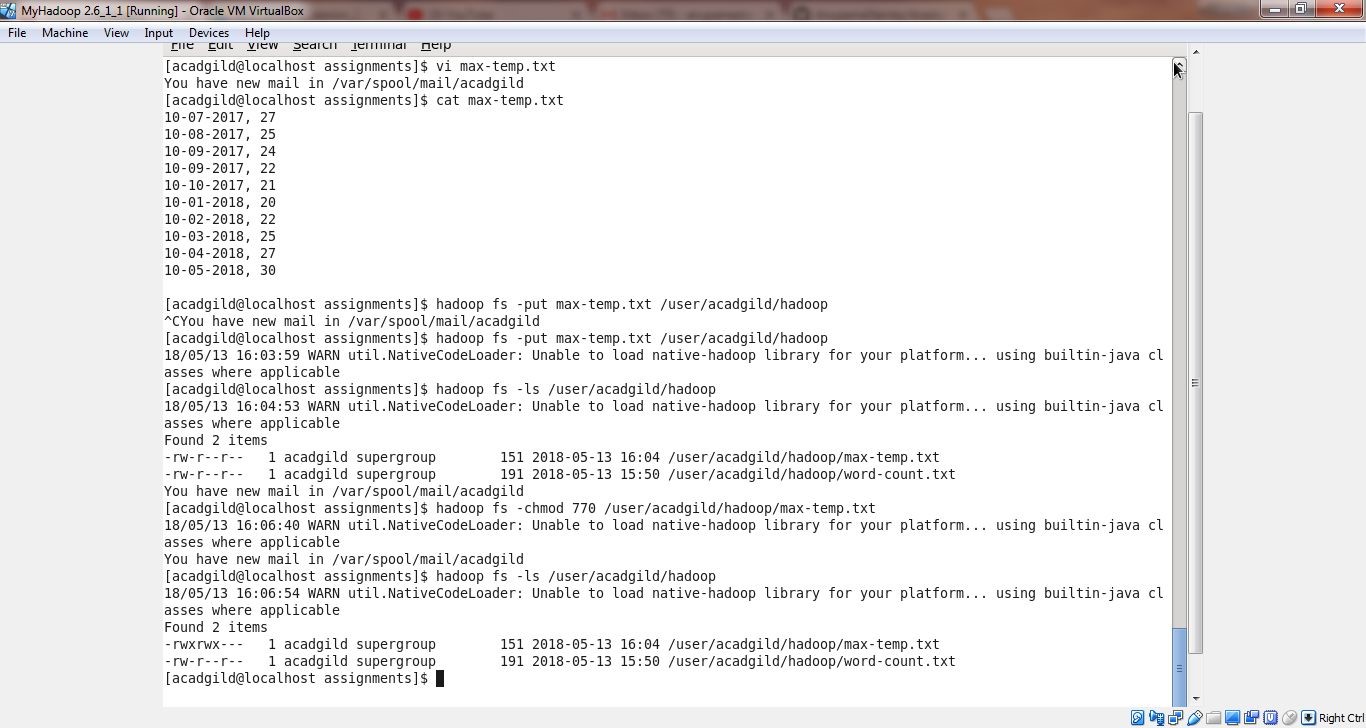


Task 4:

Change the permission of the file /user/acadgild/hadoop/max-temp.txt, such that only the owner and the group members have full control over the file.

Others do not have any control over it.

Hadoop fs –chmod 770 provides full permission to owner and group and no permission to others on the file.



Step 5. Create a folder /user/info323' in the HDFS.

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. To use the HDFS commands, first you need to start the Hadoop services using the following command:

sbin/start-all.sh

To check the Hadoop services are up and running use the following command:

jps



**Commands:**

1. **ls:** This command is used to list all the files. Use *lsr* for recursive approach. It is useful when we want a hierarchy of a folder.

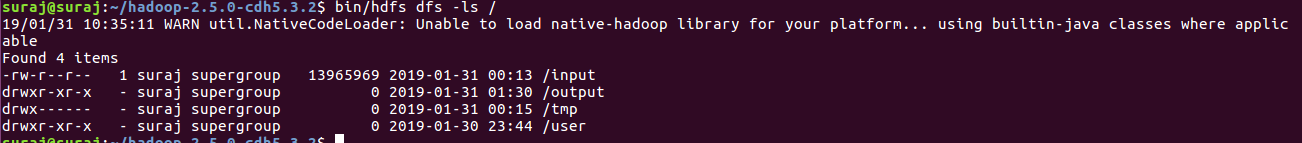
**Syntax:**

bin/hdfs dfs -ls <path>

**Example:**

bin/hdfs dfs -ls /

It will print all the directories present in HDFS. bin directory contains executables so, *bin/hdfs* means we want the executables of hdfs particularly *dfs*(Distributed File System) commands.



1. **mkdir**: To create a directory. In Hadoop *dfs*there is no home directory by default. So let’s first create it.

**Syntax:**

bin/hdfs dfs -mkdir <folder name>

creating home directory:

hdfs/bin -mkdir /user

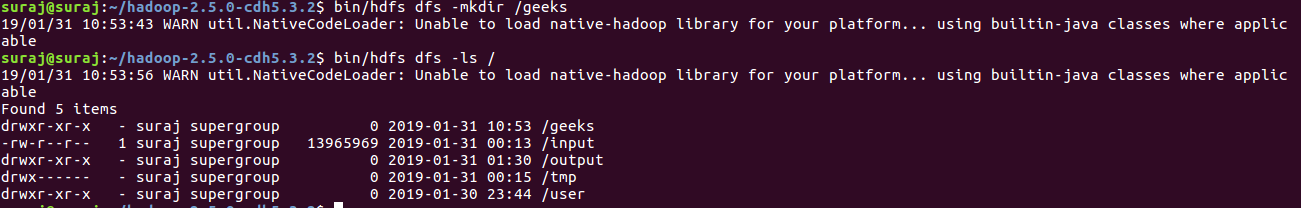
hdfs/bin -mkdir /user/username -> write the username of your computer

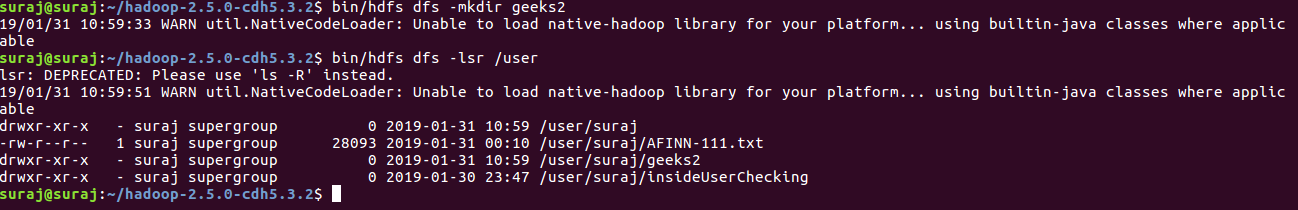
**Example:**

bin/hdfs dfs -mkdir /hadoop => '/' means absolute path

bin/hdfs dfs -mkdir hadoop => Relative path -> the folder will be

created relative to the home directory.





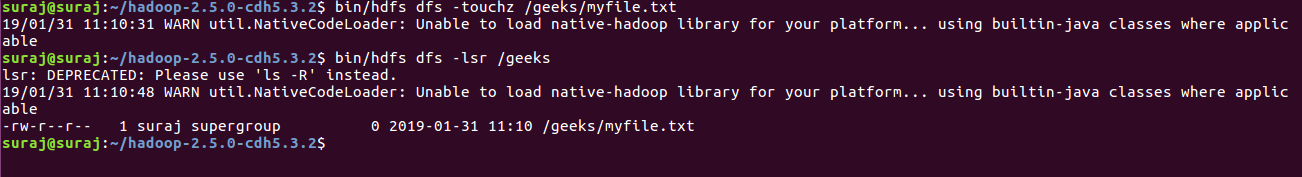
1. **touchz**: It creates an empty file.

**Syntax:**

bin/hdfs dfs -touchz <file\_path>

**Example:**

bin/hdfs dfs -touchz /hadoop/myfile.txt



1. **copyFromLocal (or) put:** To copy files/folders from local file system to hdfs store. This is the most important command. Local filesystem means the files present on the OS.

**Syntax:**

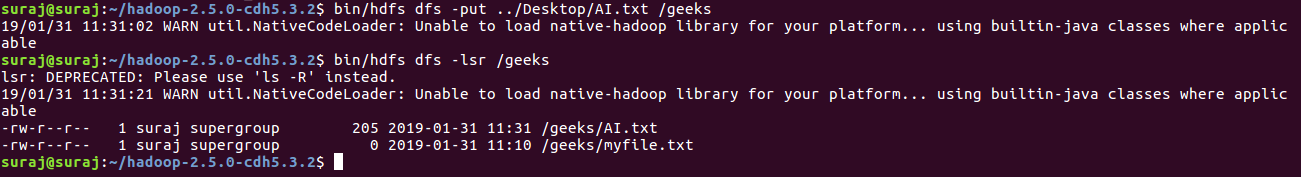
bin/hdfs dfs -copyFromLocal <local file path> <dest(present on hdfs)>

**Example:** Let’s suppose we have a file *AI.txt* on Desktop which we want to copy to folder *hadoop*present on hdfs.

bin/hdfs dfs -copyFromLocal ../Desktop/AI.txt /hadoop

(OR)

bin/hdfs dfs -put ../Desktop/AI.txt /hadoop



1. **cat:** To print file contents.

**Syntax:**

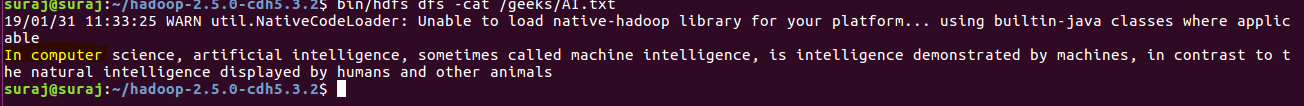
bin/hdfs dfs -cat <path>

**Example:**

// print the content of AI.txt present

// inside hadoop folder.

bin/hdfs dfs -cat /hadoop/AI.txt ->



Step 6. Copy a file “test.txt" from the file system on the master node to the HDFS folder

‘info323' you just created.

1. **copyToLocal (or) get:** To copy files/folders from hdfs store to local file system.

**Syntax:**

bin/hdfs dfs -copyToLocal <<srcfile(on hdfs)> <local file dest>

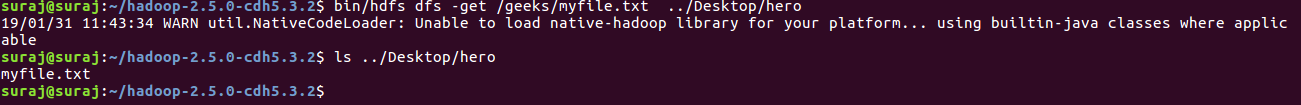
**Example:**

bin/hdfs dfs -copyToLocal /hadoop ../Desktop/hero

(OR)

bin/hdfs dfs -get /hadoop/myfile.txt ../Desktop/hero

*myfile.txt* from *hadoop*folder will be copied to folder *hero*present on *Desktop*.



**Note:** Observe that we don’t write *bin/hdfs* while checking the things present on local filesystem.

Step 7. Copy the “test.txt file from the HDFS to the home directory on the master node

as “test-from-HDFS.txt".

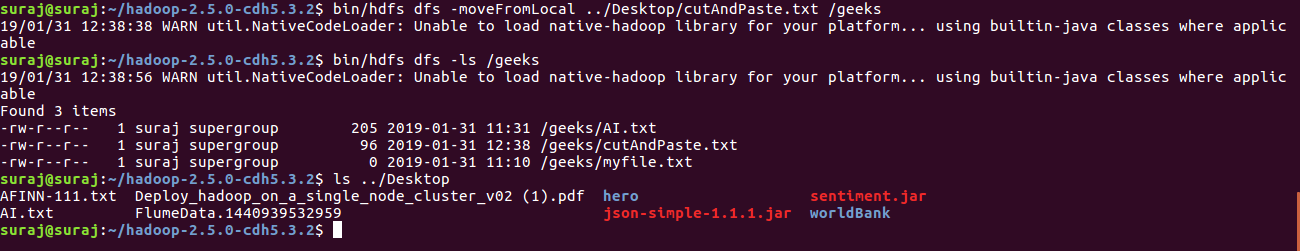
**moveFromLocal:** This command will move file from local to hdfs.

**Syntax:**

bin/hdfs dfs -moveFromLocal <local src> <dest(on hdfs)>

**Example:**

bin/hdfs dfs -moveFromLocal ../Desktop/cutAndPaste.txt /geeks



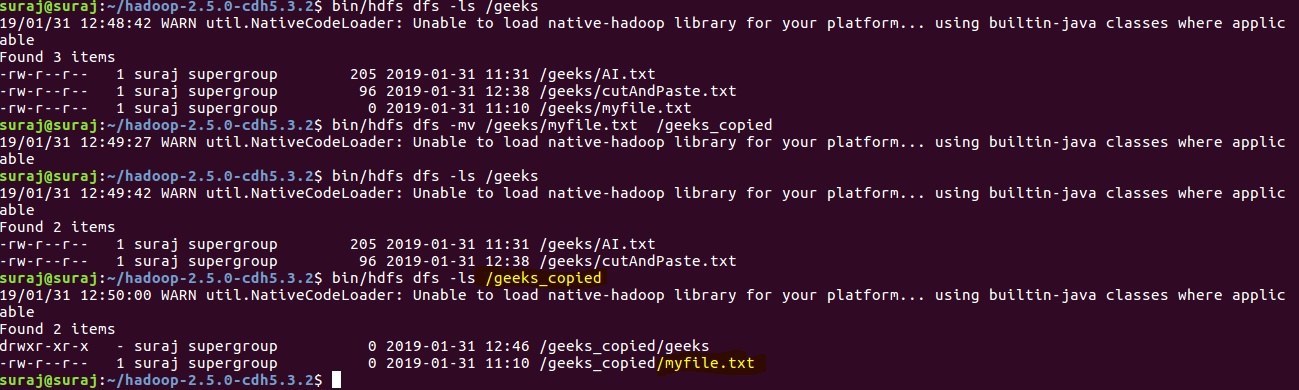
**mv:** This command is used to move files within hdfs. Lets cut-paste a file *myfile.txt* from *hadoop*folder to *hadoop\_copied*.

**Syntax:**

bin/hdfs dfs -mv <src(on hdfs)> <src(on hdfs)>

**Example:**

bin/hdfs -mv /hadoop/myfile.txt /hadoop\_copied



Step 8. Delete the file and the folder in the HDFS.

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

**Syntax:**

bin/hdfs dfs -rmr <filename/directoryName>

**Example:**

bin/hdfs dfs -rmr /hadoop\_copied -> It will delete all the content inside the

directory then the directory itself.

