**Lab 3**

**Aim:** "Interfacing to Distributed File System – HDFS commands and Web GUI file system browser usage"

**HDFS commands:**

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 maintaining the metadata in the form of log files. To use the HDFS commands, 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:

Jsp

**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>

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>

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

**Syntax**:bin/hdfs dfs -touchz<file\_path>

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

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

1. **cat:** To print file contents.

**Syntax:** bin/hdfs dfs -cat <path>

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>

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

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

1. **cp:**This command is used to copy files within hdfs.

**Syntax:** bin/hdfs dfs -cp<src(on hdfs)><dest(on hdfs)>

1. **mv:** This command is used to move files within hdfs.

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

1. **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>

1. **du:** It will give the size of each file in directory.

**Syntax**: bin/hdfs dfs -du <dirName>

1. **dus:** This command will give the total size of directory/file.

**Syntax:**bin/hdfs dfs -dus<dirName>

1. **stat:** It will give the last modified time of directory or path. In short it will give stats of the directory or file.

**syntax:**bin/hdfs dfs -stat <hdfs file>

1. **setrep:** This command is used to change the replication factor of a file/directory in HDFS. By default it is 3 for anything which is stored in HDFS (as set in hdfs core-site.xml).

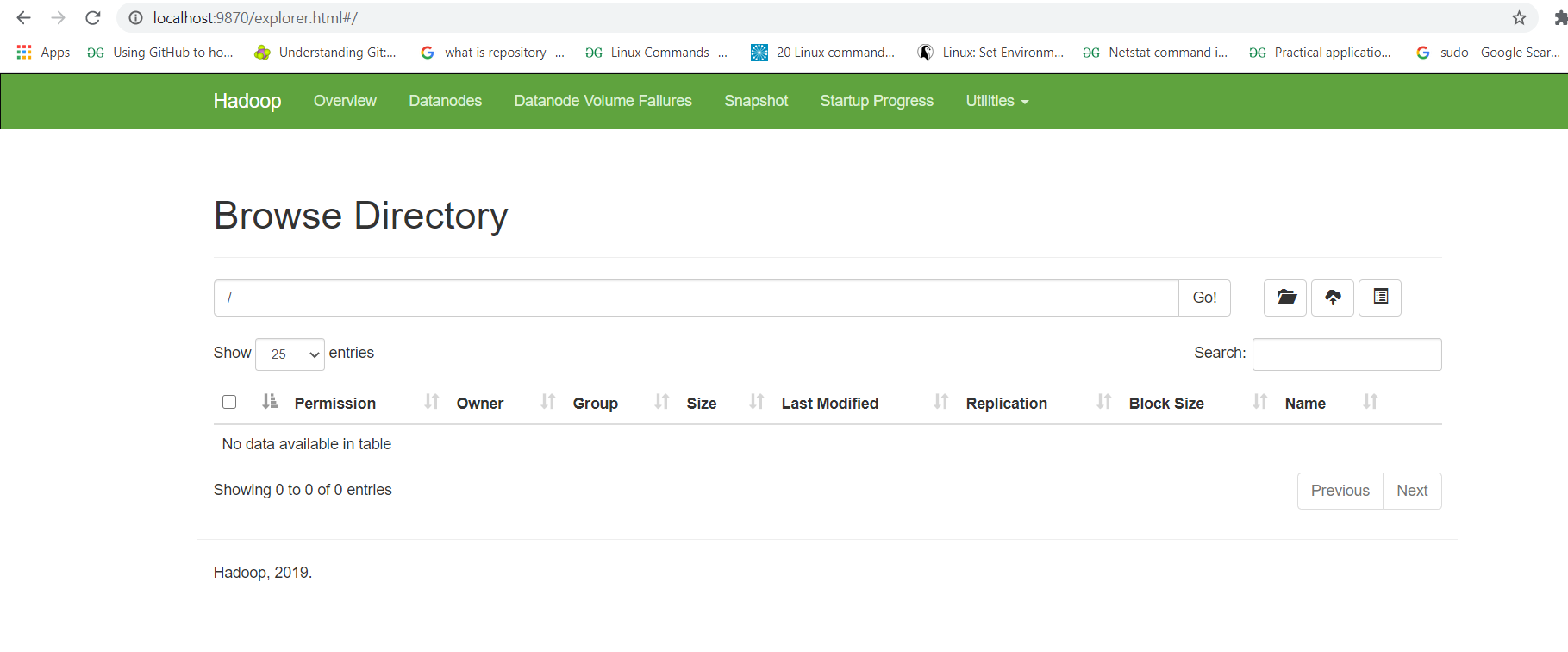
**Syntex:**bin/hdfs dfs -setrep -R -w 6 geeks.txt

**Exercise**

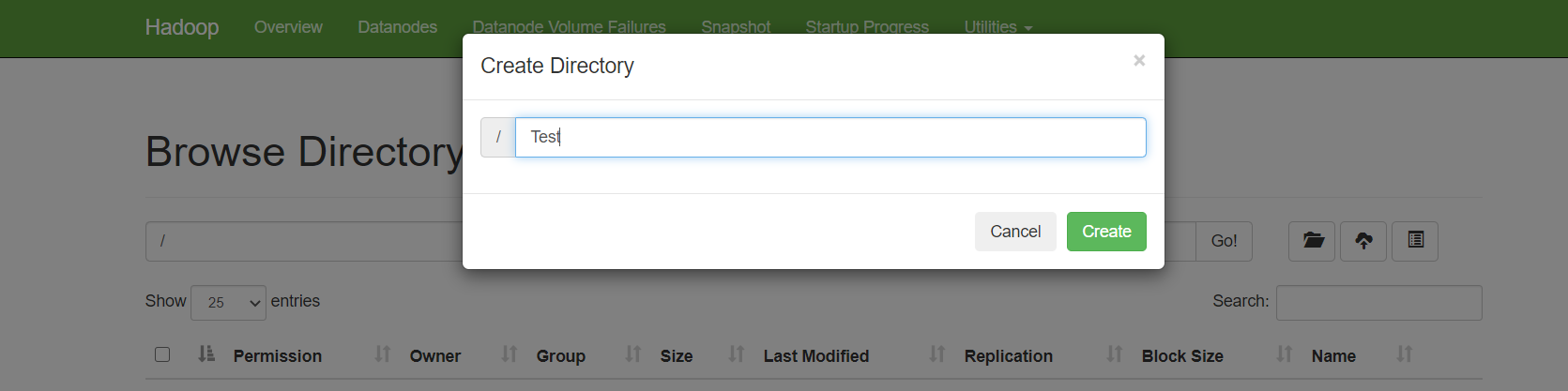
**Exercise 1: Create a directory named 'Test' from HDFS web UI in '/user/hadoop'. It should give error message. Read this message and solve the error.**

**Steps:**

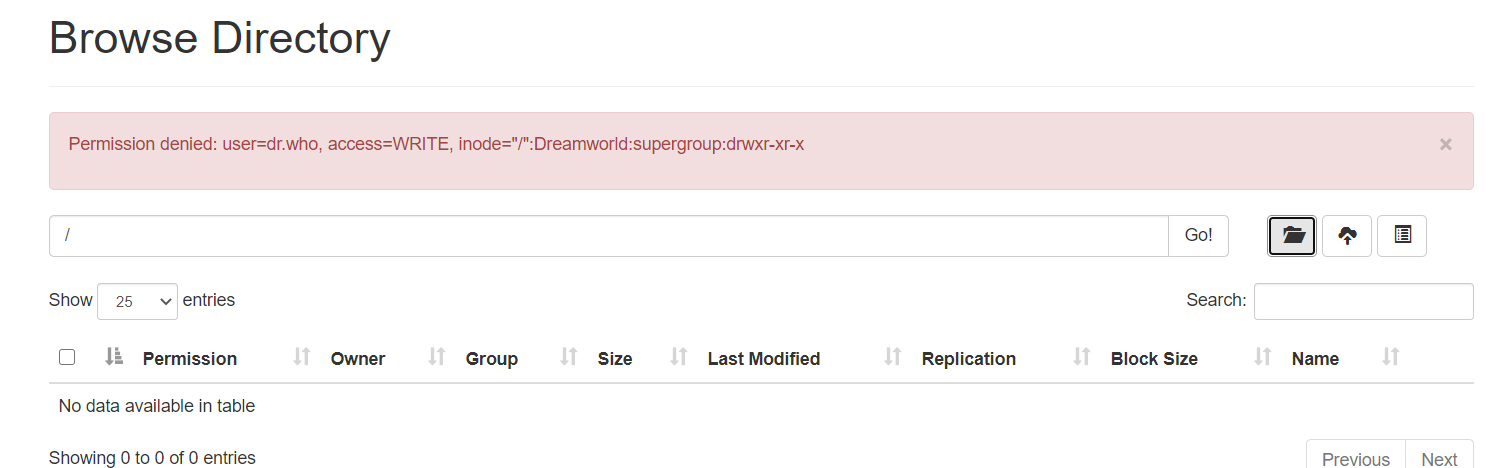
1. Start hadoop daemons.
2. Go to HDFS web GUI by the URL <http://localhost:9870/explorer.html#/>and go to Utilities > Browse the file system.



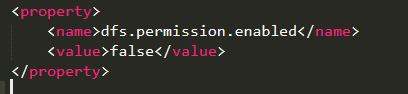
1. Click on create Directory icon.
2. Enter the directory name “Test” as following and click on Create:



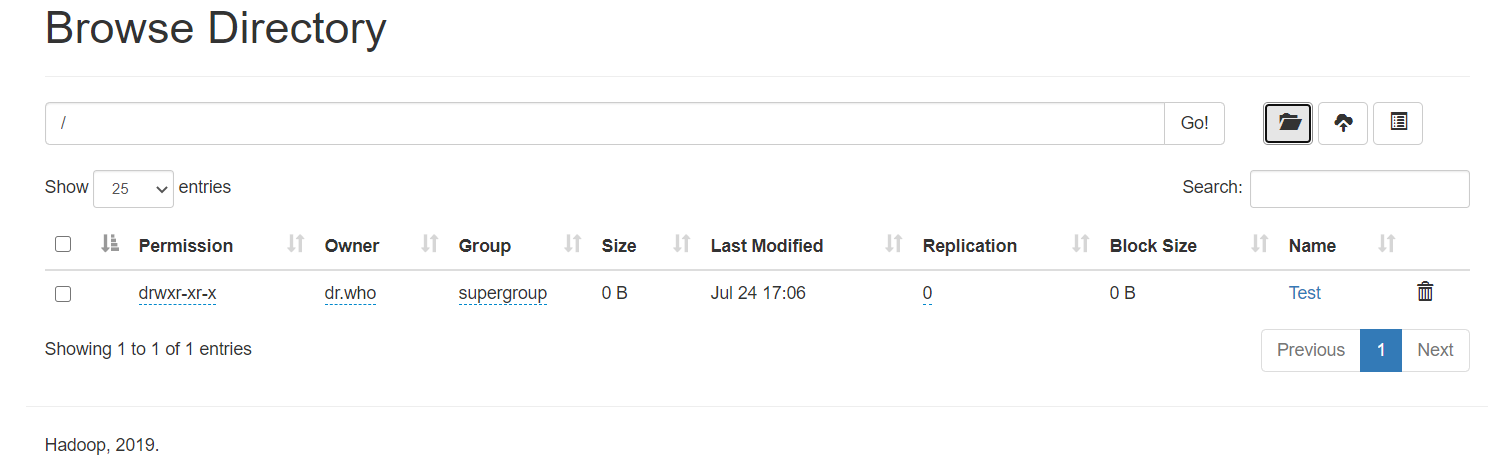
1. Probably you get error like below.



1. To resolve this error, stop HDFS daemons.
2. Make some changes to hdfs-site.xml file as below.



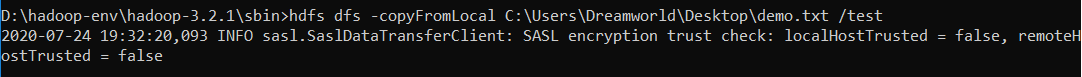
1. Now restart the daemons.
2. Again go to the GUI and repeat above steps.
3. Now you can successfully create the directory.



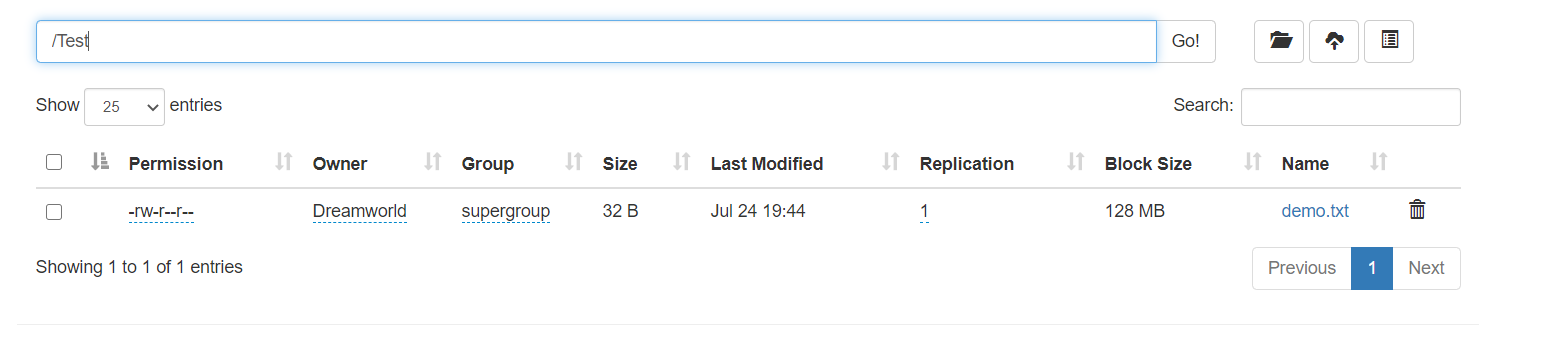
**Exercise 2:Create a text file in local file system and put the same file in HDFS using appropriatecommand.**

**Approach 1:** Using copyFromLocal command

1. Create one text file on local machine.
2. Now use the following command to put the same file in hdfs file system.



1. You can see the created file using HDFS Web GUI as following:

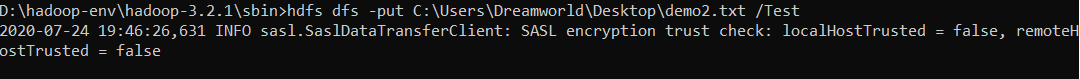


1. You can also verify by –ls command.

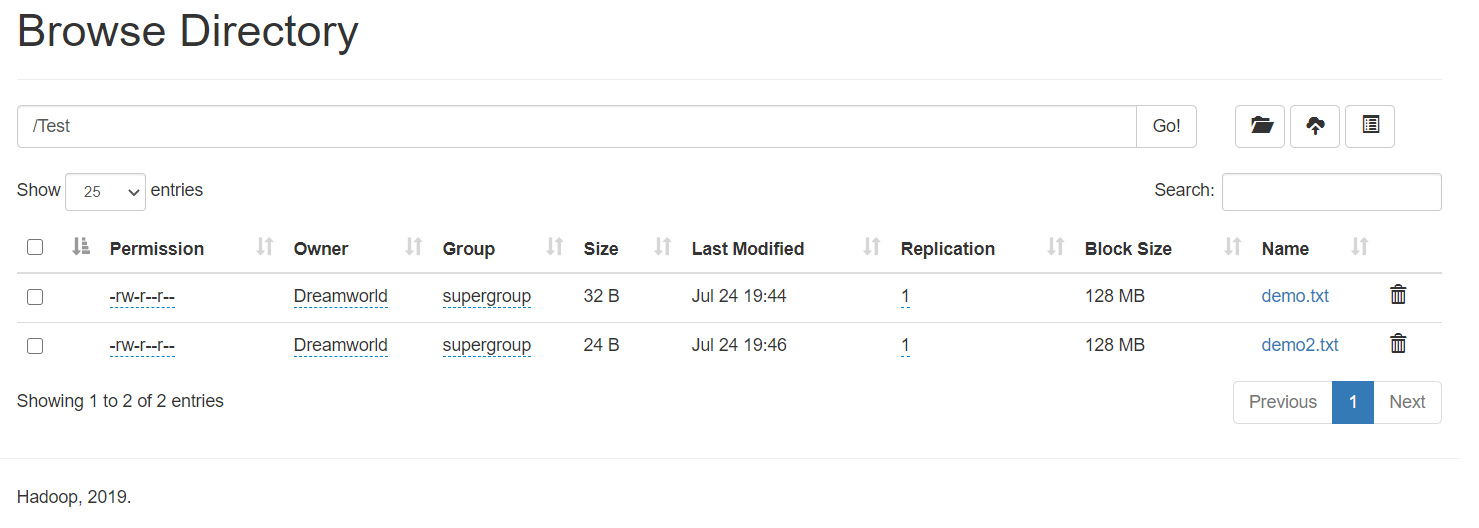


**Approach 2:**Using put command

1. Create text file on local machine.
2. Now use the following command to put the same file on the hdfs file system.

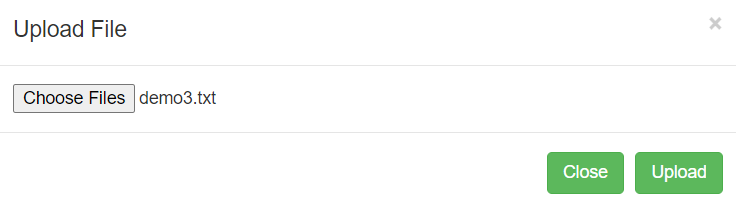


1. You can see the created file using HDFS Web GUI as following:

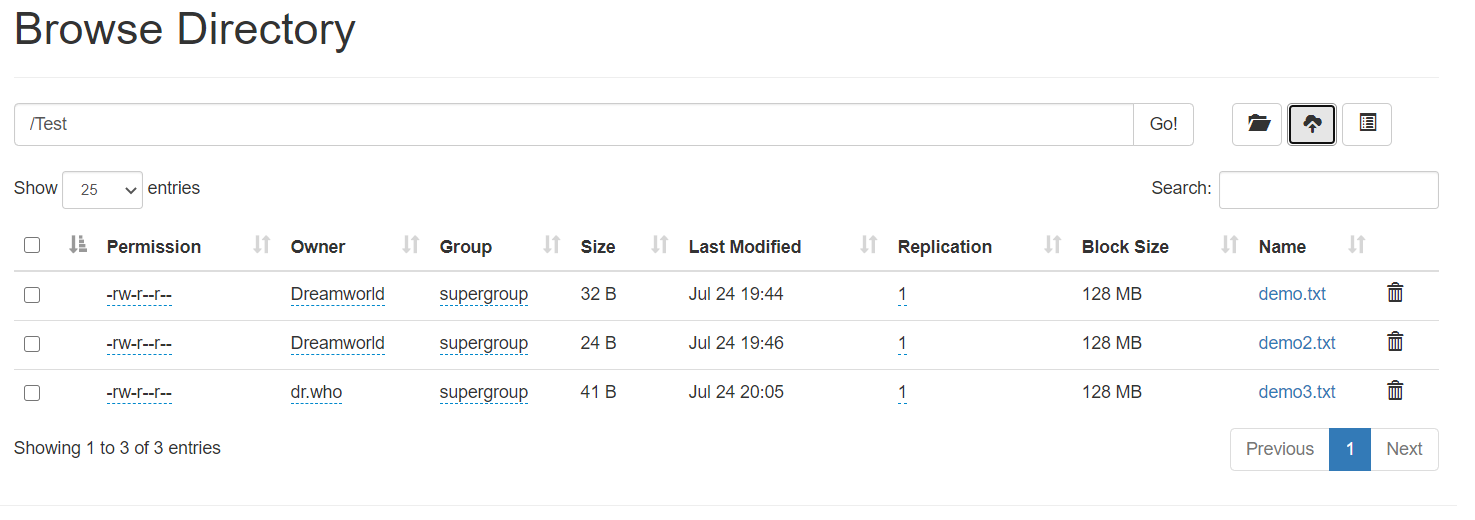


**Approach 3:** Using HDFS Web GUI

1. Create a file on local machine.
2. Go to HDFS Web GUI by the URL<http://localhost:9870/> and go to Utilities > Browse the File System. Click on Upload Files icon.
3. Select the file, which needs to be uploaded form Browse option and click one Upload.



1. You can see the created file using HDFS Web GUI as following:

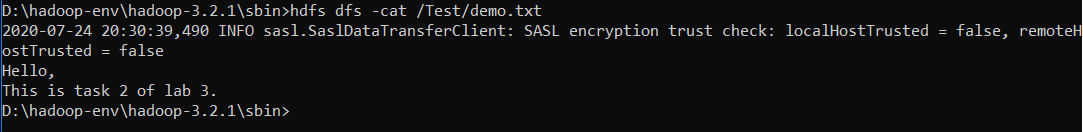


**Exercise 3: Display the contents of this file of HDFS on terminal.**

**Steps:**

1. Use the following command to display contents of HDFS on terminal.

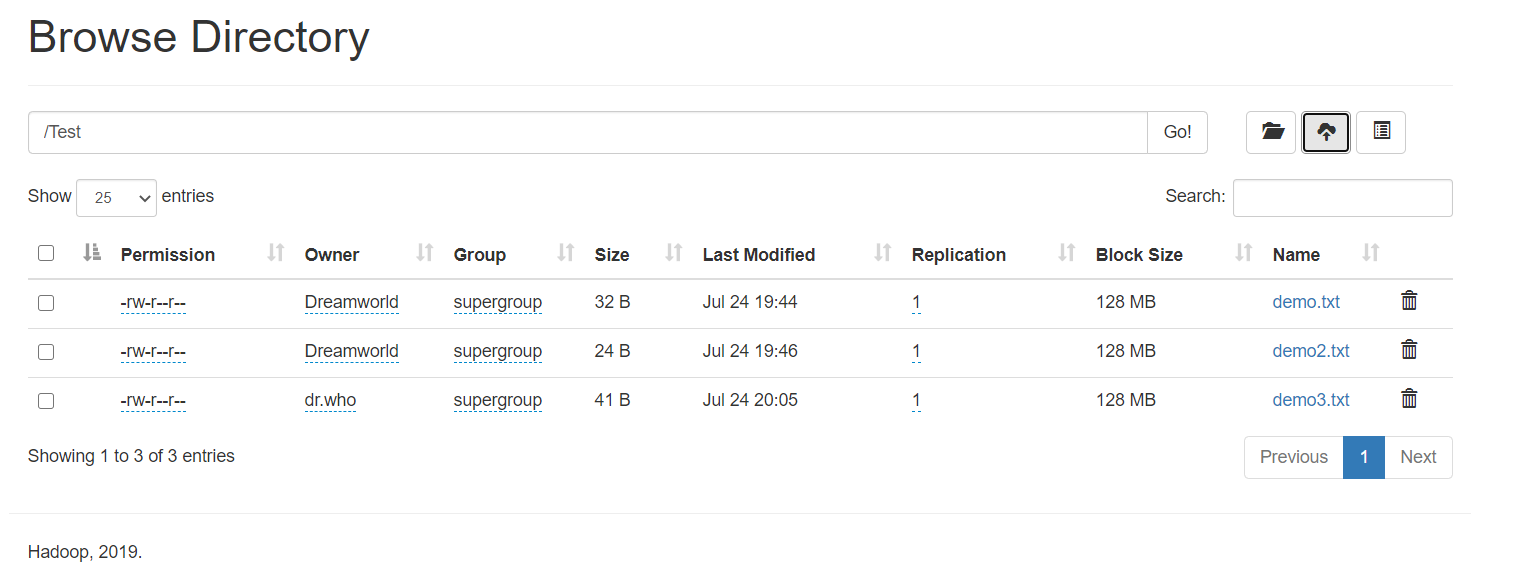
**hdfs dfs -cat <HDFSFilePath>**



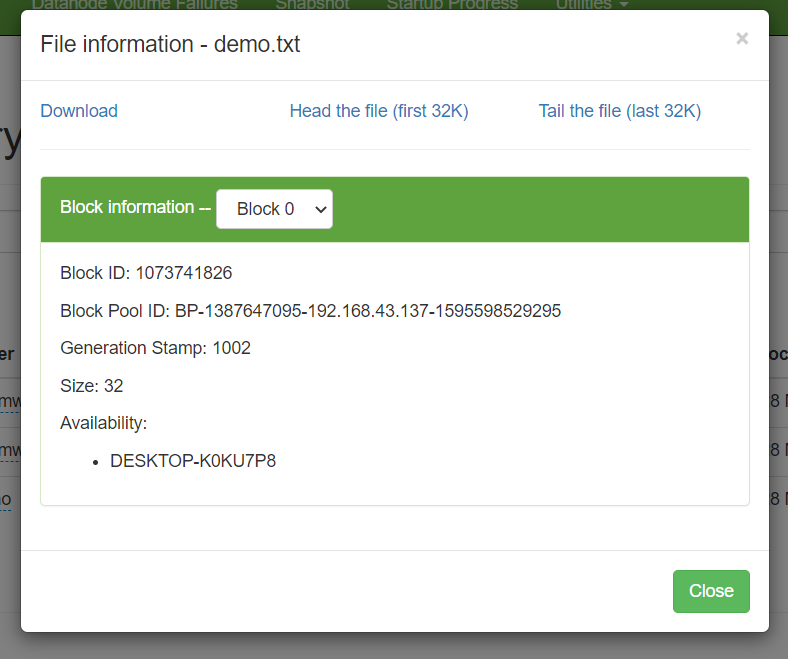
**Exercise 4: Download this file of HDFS on local machine and change its contents. Upload thismodified file again on HDFS and verify its contents. Also make sure that this file should be deleted from local machine automatically**.

**Steps:**

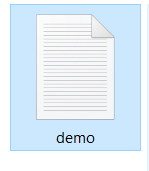
1. Go to file path on HDFS Web GUI. And Click on the file name.



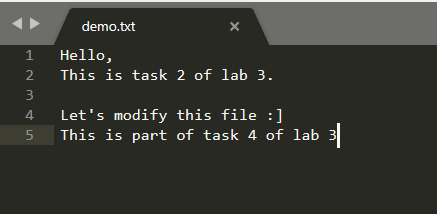
1. It will open one dialog box. Click on Download File option.



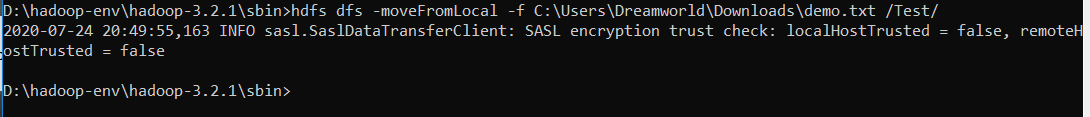
1. To see the downloaded file, go to Browser Download Location. And you can see the downloaded file.



1. Now open the downloaded file and change its content as following:

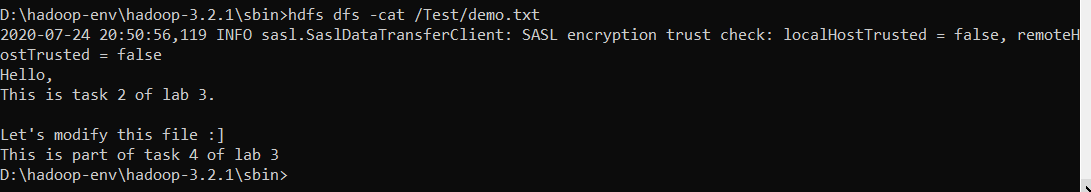


1. Now use following command to upload modified file again on HDFS.



-f option ensures that this file is deleted from local machine automatically.

1. To verify the contents, use the cat commands by following the same steps mentioned in Exercise-3 as following:



1. Go to the Browser Download Location. And you can see that source file is deleted automatically.

**Exercise 5: Find out a CLI way to display all file system commands supported by hdfs. Record the list.**

**Steps:**

1. Use the following command to display all file system commands supported by hdfs.

**hdfs dfs –help**

