**BIG DATA ANALYTICS**

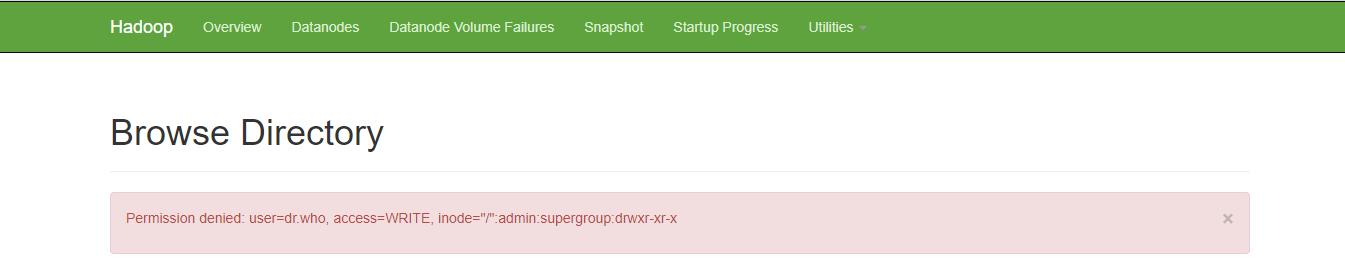
**LAB 3**

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

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

**Solution:**

* When I try to make a new directory, it will give error ‘**permission denied**’.

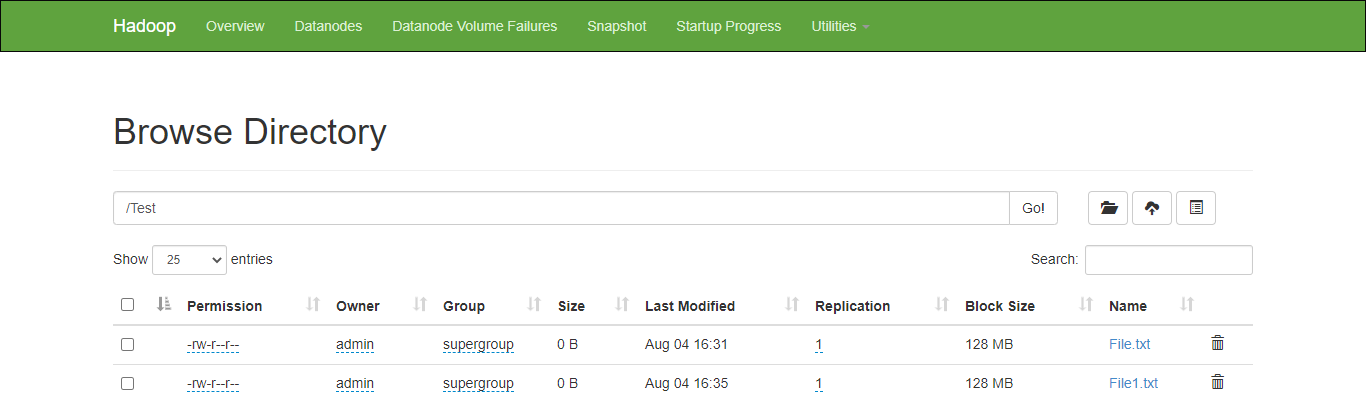


* Error can be solved by **modifying the hdfs-site.xml file**. We need to add the property’s name and value as **dfs.permissions** and **false** respectively.



* Now the ‘Test’ directory will be created successfully.

**Output:**



**2. Create a text file in the local file system and put the same file in HDFS using appropriate command.**

**Solution:**

* This task can be done in **3 ways**

1. Absolute
2. Relative
3. Scheme.

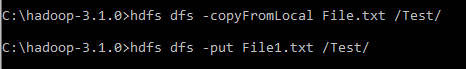
* Both **put** and **copyFromLocal** command will give the desired output.
* **put** command can copy single source, or multiple sources from the local file system to the destination file system. Command for put will be as below:

**hadoop dfs -put <source> <dest>**

* **copyFromLocal**, similar to the fs -put command, except that the source is restricted to a local file reference. Command for put will be as below:

**hadoop dfs -copyFromLocal <source> <dest>**

1. **Relative way**



1. **Absolute way**

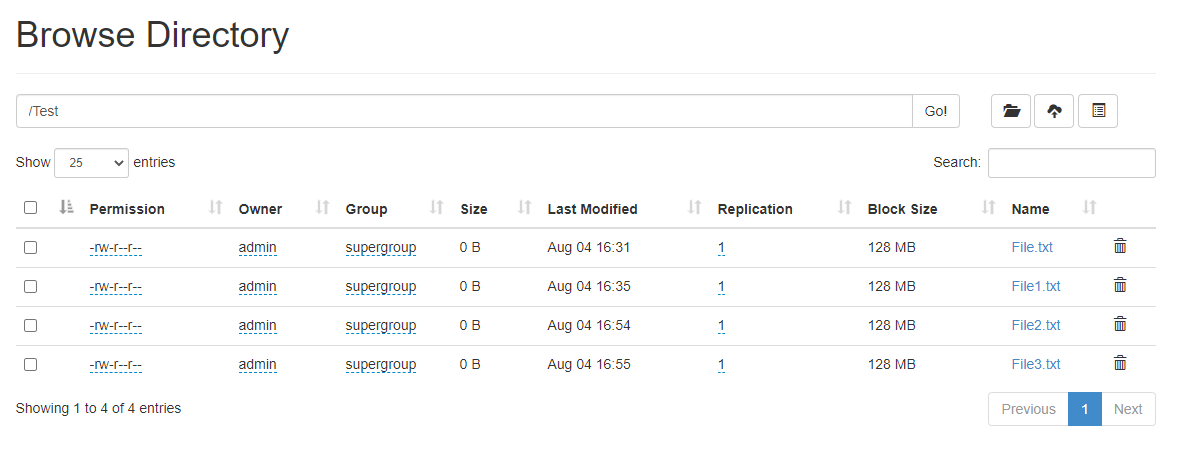
****

1. **Scheme**

****

**Output:**

* All the created text files will be shown in HDFS web UI.



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

**Solution:**

* File contents can be displayed by **2 ways**

1. Scheme
2. Relative way.

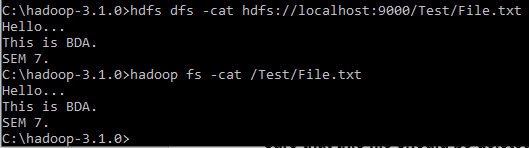
For Scheme, command is will be as below:

**hdfs dfs -cat hdfs://localhost:9000/Test/File.txt**

For relative, command is will be as below:

**hadoop fs -cat /Test/File.txt**

**Output:**



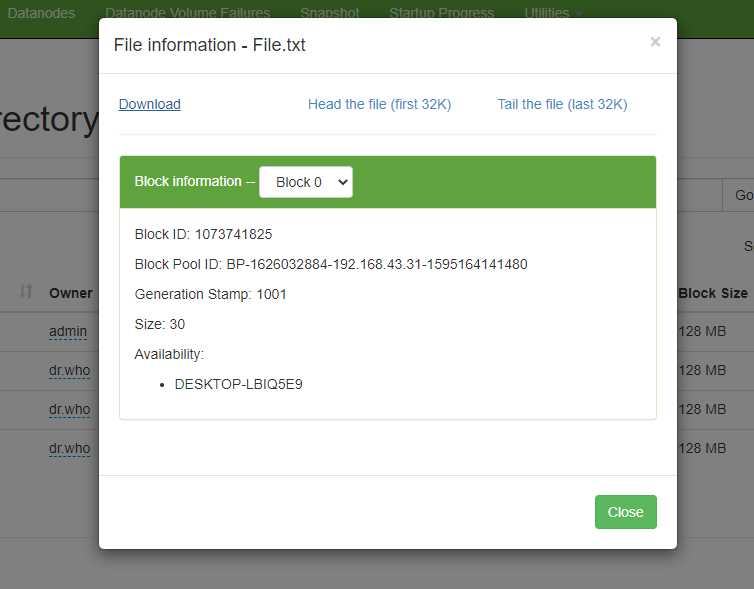
**4. Download this file of HDFS on a local machine and change its contents. Upload this modified file again on HDFS and verify its contents. Also make sure that this file should be deleted from the local machine automatically.**

**Solution:**

* We can perform this task using **both – GUI and CLI** approach.
* Solution using both approaches are discussed below.

1. **Using GUI:**

* Files can be downloaded using the HDFS web UI and saved to the desired location.



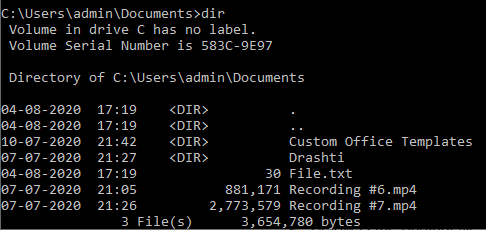
1. **Using CLI:**

* File can be downloaded using **copyToLocal** command and using dir we can see that it is stored in the ‘Documents’ folder.

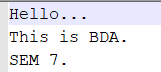
**hadoop fs -copyToLocal <source> <dest>**



* We can locate the copied file ‘File.txt’ in the ‘Documents’ folder.



* Changing the content of ‘File.txt’. Below is the original file.

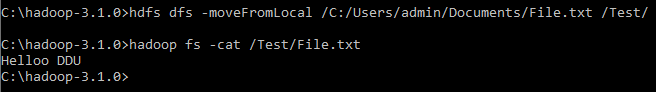


* To change the contents of the file, we need to give permission so **run the notepad as administrator.**

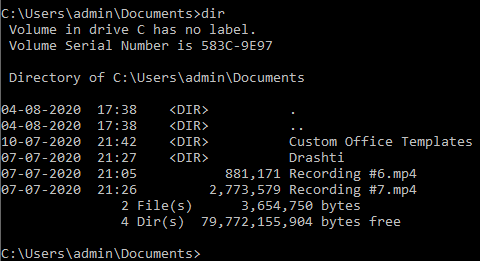


* We have to use **moveFromLocal -f** option so that modified file will be stored in the test directory with same filename.

**hadoop dfs -moveFromLocal <source> <dest>**



* Now we can see that the uploaded file is **deleted automatically**.



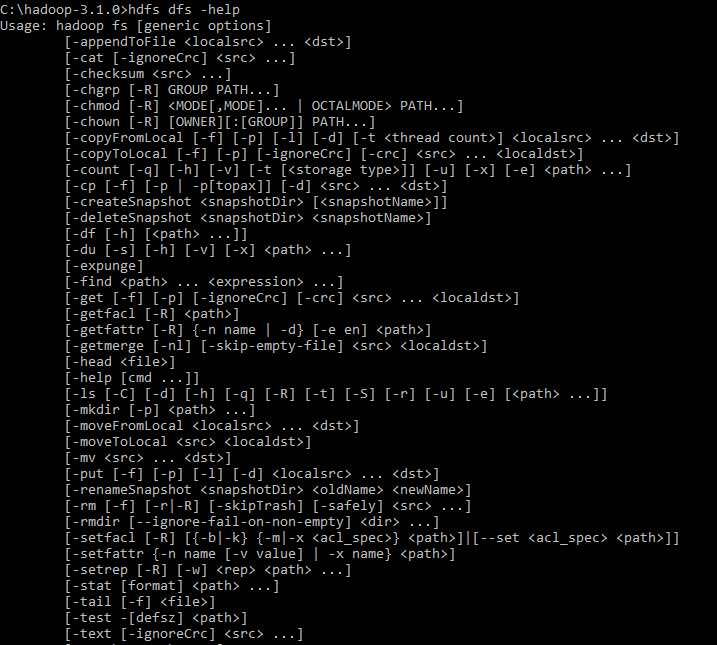
**5. Find out a CLI way to display all file system commands supported by hdfs. Record the list.**

**Solution:**

* Using **hdfs dfs -help** we can display all the file system commands supported by hdfs.

**hdfs** **dfs -help**

**Output:**



* **mkdir**: Similar to Unix mkdir command, it is used for **creating directories** in HDFS.
* **ls**: Similar to Unix ls command, it is used for **listing directories** in HDFS. The **-lsr** command can be used for **recursive listing**.
* **put**: It **copies files** from local file system to HDFS. This is similar to **-copyFromLocal** command.
* **get**: It **copies files** from HDFS to local file system. This is similar to **-copyToLocal command**.
* **cat**: Similar to Unix cat command, it is used for **displaying contents of a file**.
* **cp**: Similar to Unix cp command, it is used **for copying files** from one directory to another **within HDFS.**
* **mv**: Similar to Unix mv command, it is used **for moving a file** from one directory to another **within HDFS**.
* **rm**: Similar to Unix rm command, it is used **for removing a file** from HDFS. The command **-rmr** can be used for recursive delete.
* **getmerge**: **It is one of the important and useful command** when trying to read the contents of map reduce job or pig job’s output files. It is used for **merging a list of files in one directory on HDFS into a single file on a local file system.**
* **setrep:** This command is used to change the replication factor of a file to a specific instead of the default of replication factor for the remaining in HDFS.
* **touchz**: This command can be used to **create a file of zero length** in HDFS.
* **test**: This command can be used to **test a hdfs file’s existence or zero length or is it a directory**.