

Lab 3) Implement the following file management tasks in Hadoop:

- **Adding files and directories**
- **Retrieving Files**
- **Deleting Files**

Hint: A typical Hadoop workflow creates data files (such as log files) elsewhere and copies them into HDFS using one of the above command line utilities.

HDFS is a scalable distributed filesystem designed to scale to petabytes of data while running on top of the underlying filesystem of the operating system. HDFS keeps track of where the data resides in a network by associating the name of its rack (or network switch) with the dataset. This allows Hadoop to efficiently schedule tasks to those nodes that contain data, or which are nearest to it, optimizing bandwidth utilization. Hadoop provides a set of command line utilities that work similarly to the Linux file commands, and serve as your primary interface with HDFS. We're going to have a look into HDFS by interacting with it from the command line. We will take a look at the most common file management tasks in Hadoop, which include:

- Adding files and directories to HDFS
- Retrieving files from HDFS to local filesystem
- Deleting files from HDFS

ALGORITHM:-

SYNTAX AND COMMANDS TO ADD, RETRIEVE AND DELETE DATA FROM HDFS

Step-1

Adding Files and Directories to HDFS

Before you can run Hadoop programs on data stored in HDFS, you'll need to put the data into HDFS first. Let's create a directory and put a file in it. HDFS has a default working directory of `/user/$USER`, where `$USER` is your login user name. This directory isn't automatically created for you, though, so let's create it with the `mkdir` command. For the purpose of illustration, we use `chuck`. You should substitute your user name in the example commands.

```
hadoop fs -mkdir /user/chuck
```

```
hadoop fs -put example.txt
```

```
hadoop fs -put example.txt /user/chuck
```

Step-2

Retrieving Files from HDFS

The Hadoop command `get` copies files from HDFS back to the local filesystem. To retrieve `example.txt`, we can run the following command:

```
hadoop fs -cat example.tx
```

Step-3

Deleting Files from HDFS

hadoop fs -rm example.txt

- Command for creating a directory in hdfs is “**hdfs dfs –mkdir /lendicse**”.
- Adding directory is done through the command “**hdfs dfs –put lendi_english /**”.

Step-4

Copying Data from NFS to HDFS

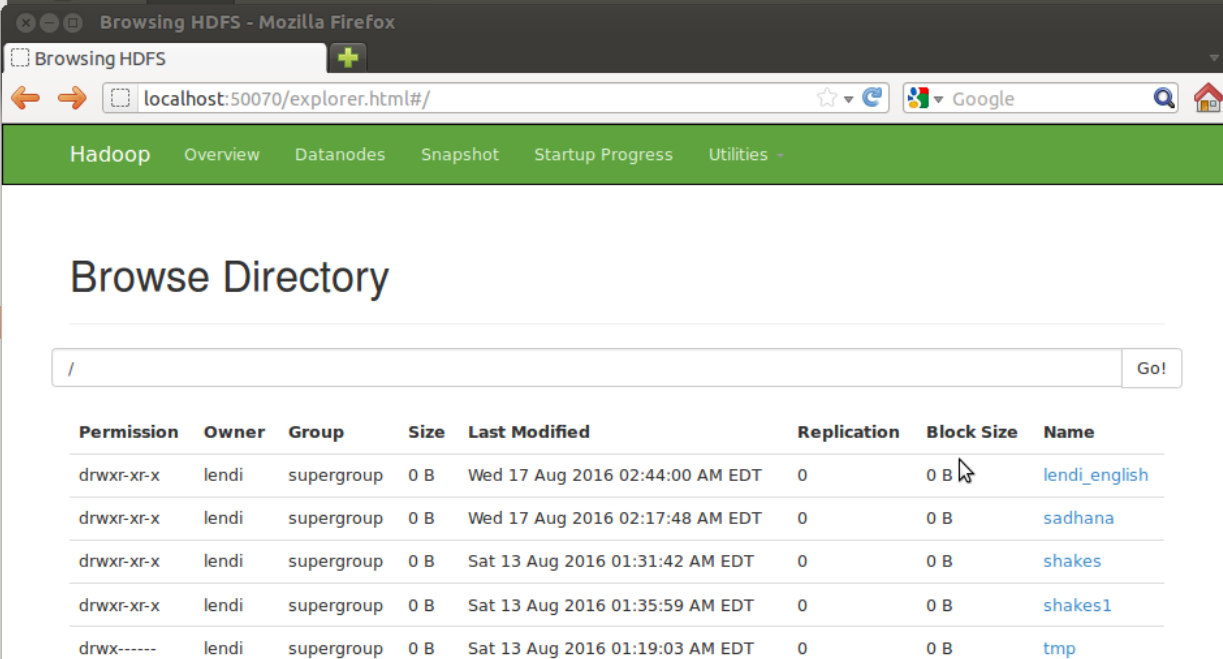
Copying from directory command is “**hdfs dfs –copyFromLocal /home/lendi/Desktop/shakes/glossary /lendicse/**”

- View the file by using the command “**hdfs dfs –cat /lendi_english/glossary**”
- Command for listing of items in Hadoop is “**hdfs dfs –ls hdfs://localhost:9000/**”.
- Command for Deleting files is “**hdfs dfs –rm r /kartheek**”.

SAMPLE INPUT:

Input as any data format of type structured, Unstructured or Semi Structured

EXPECTED OUTPUT:



Browse Directory

/

Permission	Owner	Group	Size	Last Modified	Replication	Block Size	Name
drwxr-xr-x	lendi	supergroup	0 B	Wed 17 Aug 2016 02:44:00 AM EDT	0	0 B	lendi_english
drwxr-xr-x	lendi	supergroup	0 B	Wed 17 Aug 2016 02:17:48 AM EDT	0	0 B	sadhana
drwxr-xr-x	lendi	supergroup	0 B	Sat 13 Aug 2016 01:31:42 AM EDT	0	0 B	shakes
drwxr-xr-x	lendi	supergroup	0 B	Sat 13 Aug 2016 01:35:59 AM EDT	0	0 B	shakes1
drwx-----	lendi	supergroup	0 B	Sat 13 Aug 2016 01:19:03 AM EDT	0	0 B	tmp

