2. File Management tasks in Hadoop

In Hadoop, a distributed file system called Hadoop Distributed File System (HDFS) is used for storing and managing large volumes of data across multiple nodes in a cluster. File management tasks in Hadoop involve operations related to storing, retrieving, organising, and deleting data within the HDFS.

Connect Local windows with Hadoop Linux machine using Putty

Putty is a popular open-source terminal emulator, serial console, and network file transfer application. It is commonly used on Windows systems to establish secure shell (SSH), telnet, and serial connections to remote devices, especially Linux and Unix servers.

Host Name: maria_dev@localhost Port: 2222

1. Listing files in HDFS:

Use the hadoop fs -ls command to list files and directories in a given HDFS path.

2. Remove a file in HDFS

Hadoop fs -rm path_of_file

3. Remove a directory from HDFS

Hadoop fs -rmdir direc_name

4. Create a directory in HDFS

Hadoop fs -mkdir lab2

5. List the files in local host

Ls

6. Download the files from internet to the local host

Wget https://github.com/DoctorVinay8097/bigdata/blob/main/ml-100k.zip

7. Rename the file in the local host

Mv old_file_name new_file_name

8. Copy local file or (put) to HDFS

hadoop fs -copyFromLocal file_name dir_name

9. Copy file to local system or (get)

hadoop fs -copyToLocal file_in_hdfs

I. Create a text file sample.txt in the local system and write some content to it using touch and cat commands. Put this file in HDFS

Touch sample.txt
Cat > sample.txt
Hello world (ctrl+d)

Hadoop fs -copyFromLocal sample.txt dir_name

II. Explore moveFromLocal and moveToLocal