PRACTICAL -03

Aim : Implement the following file management tasks in Hadoop:-

> Adding files and directories

> Retrieving files from HDFS to local file system

> Deleting files from HDFS

1)To give commands in HDFS download the platform putty it gets directly connected with the HDFS dashboard and from where you can give commands to add & delete the files
Download Links-https://www.chiark.greenend.org.uk/~sgtatham/putty/latest.html
After downloading open the file and give following details
Host name- maria_dev@1080
Port- 2222
Connection type- SSH
Load server- HDP & Save
After saving you will get to see the command prompt where you have to enter the password which you have been set for your browser dashboard
Password- maria_dev

2)To go in the Hadoop system give the command-

*hadoop fs -ls

The command **hadoop fs -ls** is used to **list files and directories stored in Hadoop Distributed File System (HDFS)** or other supported file systems (like local FS, S3, etc., depending on configuration).

Shows the **files and directories** at the given path.

Displays **metadata**:

- File permissions
- Replication factor
- Owner & group
- File size (in bytes)
- Last modification date & time
- Path

*Hadoop fs -mkdir

The **hadoop fs -mkdir** command is used to **create new directories in Hadoop Distributed File System (HDFS)** (or any other file system supported by Hadoop, like S3, local FS, etc., depending on your configuration)

📌 **Purpose**

- To create a **new directory** in HDFS.

Suppose we will give the command for creating a directory for a movielens dataset
Command- hadoop fs -mkdir ml-100k

*hadoop fs -ls

The **hadoop fs -ls** command is used to **list files and directories in Hadoop Distributed File System (HDFS)** or

in any other file system supported by Hadoop (like local FS, S3, etc., depending on configuration)

📌 **Purpose**

- To **view the contents** of a directory in HDFS.
- To **see metadata** of files/directories such as:
    - **Permissions** (read, write, execute)
    - **Replication factor** (for files in HDFS)
    - **Owner** and **Group**
    - **File size** (in bytes)
    - **Modification date & time**
    - **File/Directory name (path)**

***ls**

In **Hadoop**, the ls command is used to **list files and directories** in the Hadoop Distributed File System (**HDFS**)—similar to the ls command in Linux, but it operates on HDFS paths instead of local file system paths.

Purpose:

- To display the list of files/directories in a given HDFS directory.
- To view metadata like **permissions, owner, group, file size, replication factor, modification date, and path**.

***pwd**

📌 **Purpose of pwd in Hadoop**

- pwd stands for **Print Working Directory**.
- It shows the **current working directory in HDFS** where you are operating.
- Useful to confirm your present location before running file operations like ls, put, or get.

***ls**

Command to display the directory

***wget** http://media.sundog-soft.com/hadoop/ml-100k/u.data

The above command is used to copy the data from web server to the Hadoop file system

***ls**

Give the command ls to see whether the data is imported in hdfs

Once it is imported you will see the name as u.data

***ls -la**

📌 **Purpose of ls -la (Linux vs Hadoop)**

- In **Linux**, ls -la lists **all files including hidden ones** (those starting with .), with detailed information (long format).

***hadoop fs -copyFromLocal u.data ml-100k/u.data**

The file will get copied from local file system to the Hadoop named as u.data

***hadoop fs -ls**

The **hadoop fs -ls** command is used to **list files and directories in Hadoop Distributed File System (HDFS)** or in any other file system supported by Hadoop (like local FS, S3, etc., depending on configuration)

***hadoop fs -rm ml-100k/u.data**

📌 **Purpose**

- To **remove (delete) files** from HDFS.
- Works similar to Linux rm, but operates on HDFS.

***hadoop fs -rmdir ml-100k**

The **hadoop fs -rmdir** command is used to **remove (delete) empty directories from HDFS**.

📌 **Purpose**

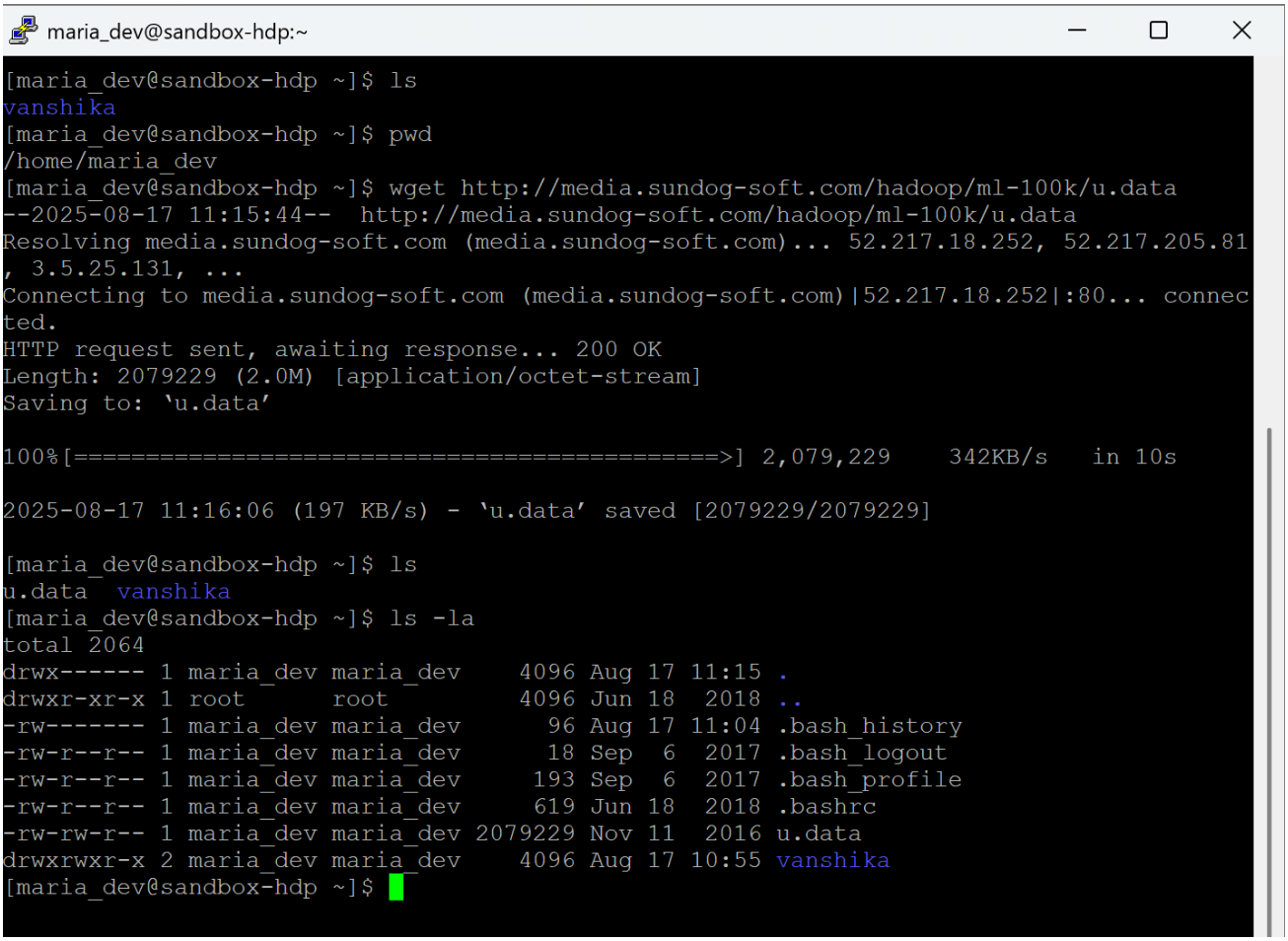
- To delete **empty directories** in Hadoop Distributed File System (HDFS).
- It is similar to the Linux rmdir command.
- ⚠️ Unlike -rm -r, it **cannot delete directories that contain files or subdirectories**.

***hadoop fs -ls**

The commands checks where the directory is removed from the hadoop

***Hadoop fs**

By using this command we may see the activities that we have performed in our Hadoop file system


maria_dev@sandbox-hdp:~
[maria_dev@sandbox-hdp ~]$ ls
vanshika
[maria_dev@sandbox-hdp ~]$ pwd
/home/maria_dev
[maria_dev@sandbox-hdp ~]$ wget http://media.sundog-soft.com/hadoop/ml-100k/u.data
--2025-08-17 11:15:44-- http://media.sundog-soft.com/hadoop/ml-100k/u.data
Resolving media.sundog-soft.com (media.sundog-soft.com)... 52.217.18.252, 52.217.205.81
, 3.5.25.131, ...
Connecting to media.sundog-soft.com (media.sundog-soft.com)|52.217.18.252|:80... connec
ted.
HTTP request sent, awaiting response... 200 OK
Length: 2079229 (2.0M) [application/octet-stream]
Saving to: 'u.data'
100%[=====>] 2,079,229 342KB/s in 10s
2025-08-17 11:16:06 (197 KB/s) - 'u.data' saved [2079229/2079229]
[maria_dev@sandbox-hdp ~]$ ls
u.data vanshika
[maria_dev@sandbox-hdp ~]$ ls -la
total 2064
drwx------ 1 maria_dev maria_dev   4096 Aug 17 11:15 .
drwxr-xr-x 1 root   root   4096 Jun 18 2018 ..
-rw------- 1 maria_dev maria_dev   96 Aug 17 11:04 .bash_history
-rw-r--r-- 1 maria_dev maria_dev   18 Sep 6 2017 .bash_logout
-rw-r--r-- 1 maria_dev maria_dev   193 Sep 6 2017 .bash_profile
-rw-r--r-- 1 maria_dev maria_dev   619 Jun 18 2018 .bashrc
-rw-rw-r-- 1 maria_dev maria_dev 2079229 Nov 11 2016 u.data
drwxrwxr-x 2 maria_dev maria_dev   4096 Aug 17 10:55 vanshika
[maria_dev@sandbox-hdp ~]$