

PRACTICAL -03

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>

The screenshot shows a web browser window with the URL [chiark.greenend.org.uk/~sgtatham/putty/latest.html](https://www.chiark.greenend.org.uk/~sgtatham/putty/latest.html). The page title is "Download PuTTY: latest release (0.83)". Below the title, there are links for "Home", "FAQ", "Feedback", "Licence", "Updates", "Mirrors", "Keys", "Links", "Team", "Download: Stable", "Snapshot", "Docs", "Privacy", "Changes", and "Wishlist". A note states: "This page contains download links for the latest released version of PuTTY. Currently this is 0.83, released on 2025-02-08." Another note says: "When new releases come out, this page will update to contain the latest, so this is a good page to bookmark or link to. Alternatively, here is a [permanent link to the 0.83 release](#)." A third note says: "Release versions of PuTTY are versions we think are reasonably likely to work well. However, they are often not the most up-to-date version of the code available. If you have a problem with this release, then it might be worth trying out the [development snapshots](#), to see if the problem has already been fixed in those versions." The main content area is titled "Package files" and lists download links for Windows installers (MSI) and a Unix source archive.

Package files		
You probably want one of these. They include versions of all the PuTTY utilities (except the new and slightly experimental Windows pterm).		
(Not sure whether you want the 32-bit or the 64-bit version? Read the FAQ entry .)		
We also publish the latest PuTTY installers for all Windows architectures as a free-of-charge download at the Microsoft Store ; they usually take a few days to appear there after we release them.		
MSI ('Windows Installer')		
64-bit x86:	putty-64bit-0.83-installer.msi	(signature)
64-bit Arm:	putty-arm64-0.83-installer.msi	(signature)
32-bit x86:	putty-0.83-installer.msi	(signature)
Unix source archive		
.tar.gz:	putty-0.83.tar.gz	(signature)

Alternative binary files		

After downloading open the file and give following details

Host name- maria_dev@1080

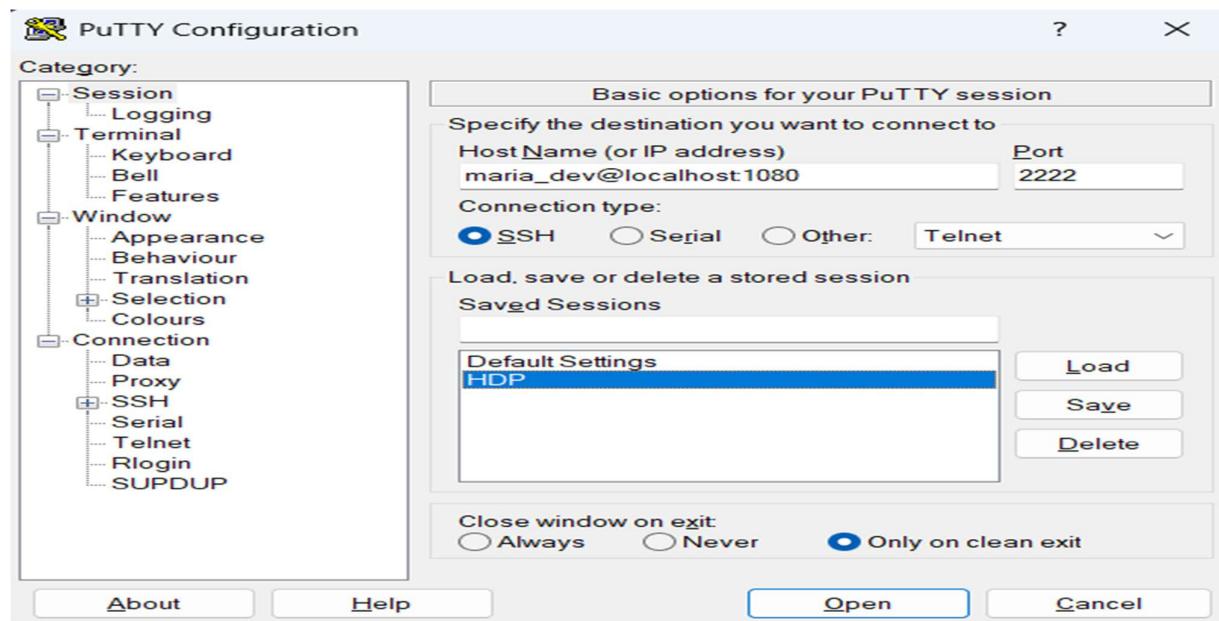
Port- 2222

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`

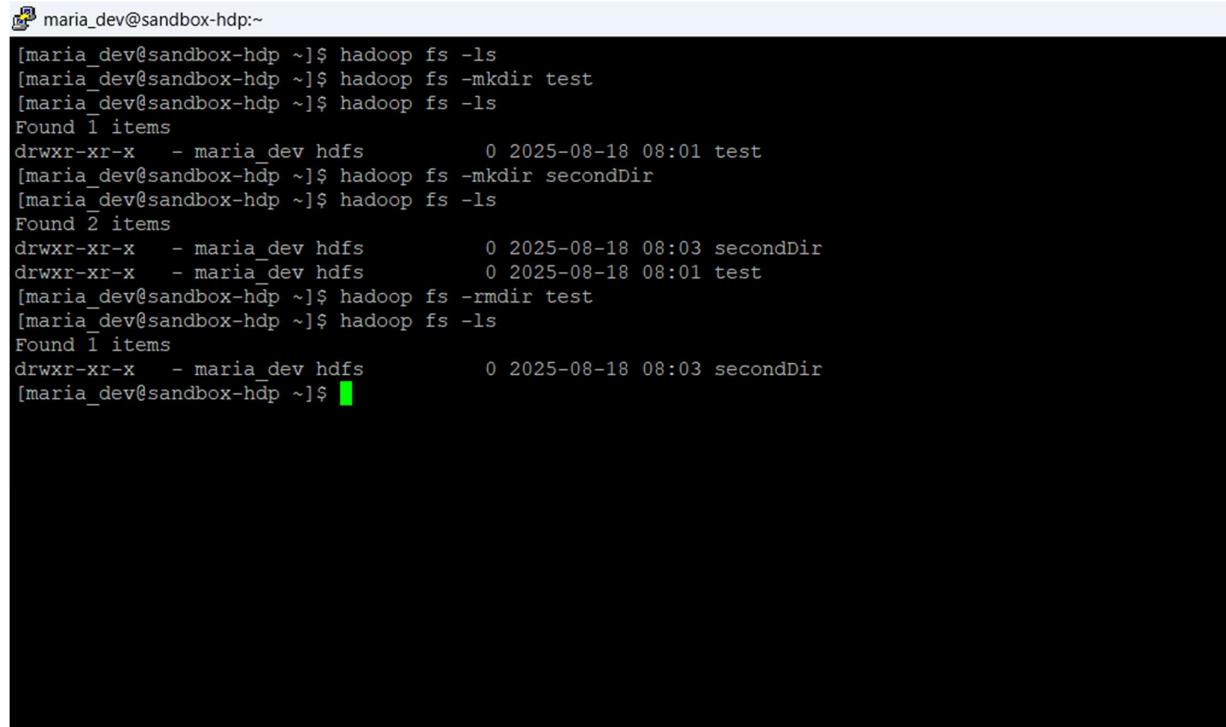
***hadoop fs -ls**

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)



```
[maria_dev@sandbox-hdp:~]$ hadoop fs -ls
[maria_dev@sandbox-hdp:~]$ hadoop fs -mkdir test
[maria_dev@sandbox-hdp:~]$ hadoop fs -ls
Found 1 items
drwxr-xr-x - maria_dev hdfs 0 2025-08-18 08:01 test
[maria_dev@sandbox-hdp:~]$ hadoop fs -mkdir secondDir
[maria_dev@sandbox-hdp:~]$ hadoop fs -ls
Found 2 items
drwxr-xr-x - maria_dev hdfs 0 2025-08-18 08:03 secondDir
drwxr-xr-x - maria_dev hdfs 0 2025-08-18 08:01 test
[maria_dev@sandbox-hdp:~]$ hadoop fs -rmdir test
[maria_dev@sandbox-hdp:~]$ hadoop fs -ls
Found 1 items
drwxr-xr-x - maria_dev hdfs 0 2025-08-18 08:03 secondDir
[maria_dev@sandbox-hdp:~]$ █
```

*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

```
[maria_dev@sandbox-hdp ~]$ wget http://media.sundog-soft.com/hadoop/ml-100k/u.data
--2025-08-18 08:13:46-- http://media.sundog-soft.com/hadoop/ml-100k/u.data
Resolving media.sundog-soft.com (media.sundog-soft.com)... 54.231.172.137, 3.5.3.105, 54.231.
196.241, ...
Connecting to media.sundog-soft.com (media.sundog-soft.com)|54.231.172.137|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 2079229 (2.0M) [application/octet-stream]
Saving to: 'u.data'

100%[=====] 2,079,229 --.-K/s in 0.1s

2025-08-18 08:13:59 (14.0 MB/s) - 'u.data' saved [2079229/2079229]
```

***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).

```
[maria_dev@sandbox-hdp ~]$ ls
u.data
[maria_dev@sandbox-hdp ~]$ ls -la
total 2056
drwx----- 1 maria_dev maria_dev 4096 Aug 18 08:13 .
drwxr-xr-x 1 root      root     4096 Jun 18 2018 ..
-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
```

***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`

```
[maria_dev@sandbox-hdp ~]$ hadoop fs -mkdir ml-100k
[maria_dev@sandbox-hdp ~]$ hadoop fs -copyFromLocal u.data ml-100k/u.data
[maria_dev@sandbox-hdp ~]$ hadoop fs -ls ml-100k
-bash: hadoop: command not found
[maria_dev@sandbox-hdp ~]$ hadoop fs -ls ml-100k
Found 1 items
-rw-r--r-- 1 maria_dev hdfs 2079229 2025-08-18 08:19 ml-100k/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