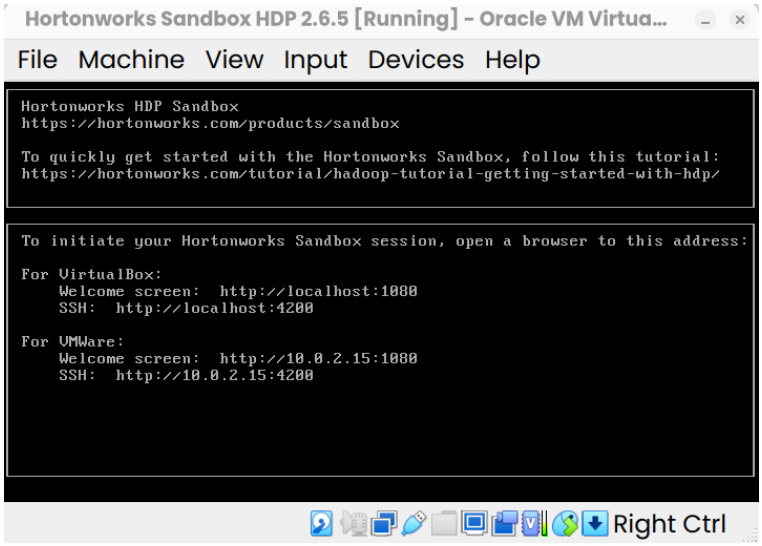
LAB 2

Aim: Installation of HortonWorks Sandbox on Virtual Machine

Procedure:

1. Start the HortonWorks it shows the IP address of machine.



9. Navigate to http://localhost:4200
10. Use the username and password: maria_dev to login.

11. Execute the following commands one-by-one:

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). It also shows the metadata:

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

hadoop fs -mkdir ml-100k

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)
We will give the command for creating a directory for a movielens dataset

```
[maria_dev@sandbox-hdp ~]$ hadoop fs -ls
[maria_dev@sandbox-hdp ~]$ hadoop fs -mkdir ml-100k
[maria_dev@sandbox-hdp ~]$ hadoop fs -ls
Found 1 items
drwxr-xr-x   - maria_dev hdfs          0 2025-08-26 04:51 ml-100k
```

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

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

```
[maria_dev@sandbox-hdp ~]$ wget http://media.sundog-soft.com/hadoop/ml-100k/u.data
--2025-08-26 04:53:21--  http://media.sundog-soft.com/hadoop/ml-100k/u.data
Resolving media.sundog-soft.com (media.sundog-soft.com)... 16.182.70.145, 52.217.194.49, 16.15.219.171, ...
Connecting to media.sundog-soft.com (media.sundog-soft.com)|16.182.70.145|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 2079229 (2.0M) [application/octet-stream]
Saving to: 'u.data'

100%[=======================================================================================================>] 2,079,229    409KB/s   in 5.0s

2025-08-26 04:53:27 (409 KB/s) - 'u.data' saved [2079229/2079229]

[maria_dev@sandbox-hdp ~]$ ls
u.data
[maria_dev@sandbox-hdp ~]$ ls -la
total 2056
drwx------ 1 maria_dev maria_dev    4096 Aug 26 04:53 .
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

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

```
[maria_dev@sandbox-hdp ~]$ ls -la
total 2056
drwx------ 1 maria_dev maria_dev    4096 Aug 26 04:53 .
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
[maria_dev@sandbox-hdp ~]$ hadoop fs  -copyFromLocal u.data ml-100k/u.data
[maria_dev@sandbox-hdp ~]$ hadoop fs -ls
Found 1 items
drwxr-xr-x   - maria_dev hdfs          0 2025-08-26 04:54 ml-100k
[maria_dev@sandbox-hdp ~]$ hadoop fs -ls ml-100k
Found 1 items
-rw-r--r--   1 maria_dev hdfs    2079229 2025-08-26 04:54 ml-100k/u.data
[maria_dev@sandbox-hdp ~]$ hadoop fs -rm ml-100k/u.data
25/08/26 04:55:22 INFO fs.TrashPolicyDefault: Moved: 'hdfs://sandbox-hdp.hortonworks.com:8020/user/maria_dev/ml-100k/u.data' to trash at: hdfs:/
ndbox-hdp.hortonworks.com:8020/user/maria_dev/.Trash/Current/user/maria_dev/ml-100k/u.data
[maria_dev@sandbox-hdp ~]$ hadoop fs -rmdir ml-100k
[maria_dev@sandbox-hdp ~]$ hadoop fs -ls
Found 1 items
drwx------   - maria_dev hdfs          0 2025-08-26 04:55 .Trash
```

-