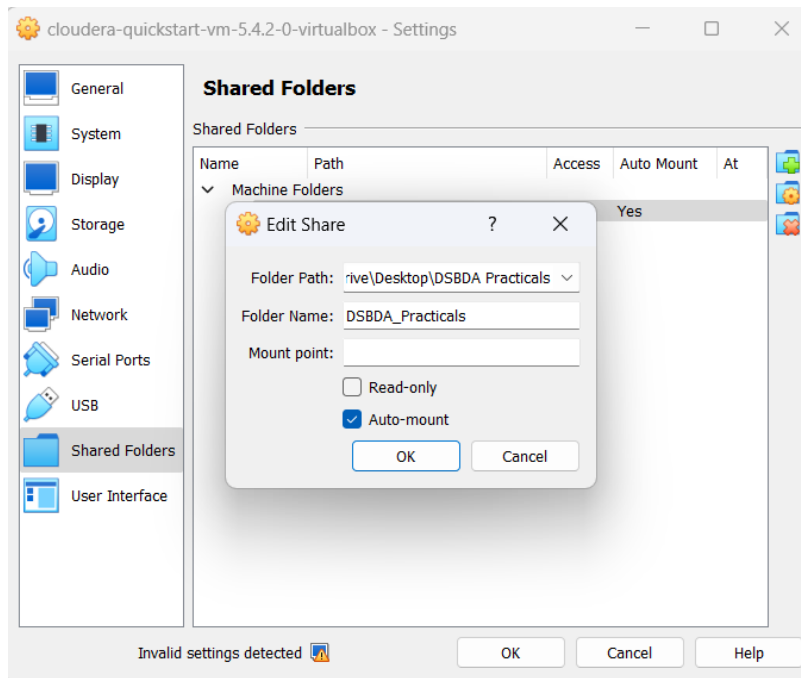
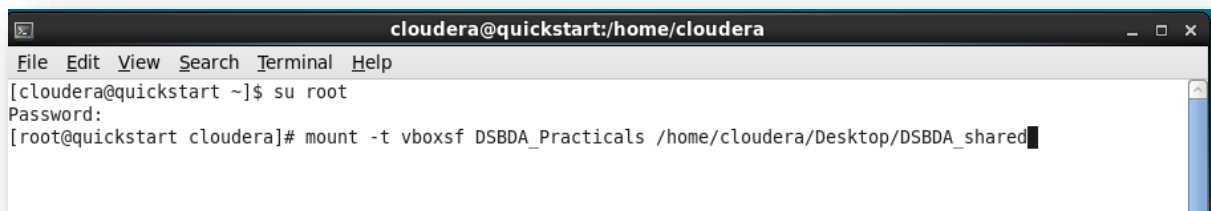


## Shared Folder:

- 1) Create a folder on Windows/Linux OS and copy the path  
Ex. C:\Users\jossa\OneDrive\Desktop\DSBDA Practicals
- 2) Click on VM settings and enter the path and Shared folder name

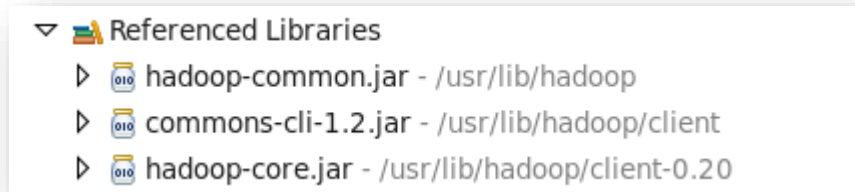


- 3) Create a folder in Cloudera  
Ex. /home/cloudera/Desktop/DSBDA\_shared
- 4) Now open Terminal and type  
su root  
Password: cloudera
- 5) mount -t vboxsf DSBDA\_Practicals /home/cloudera/Desktop/DSBDA\_shared



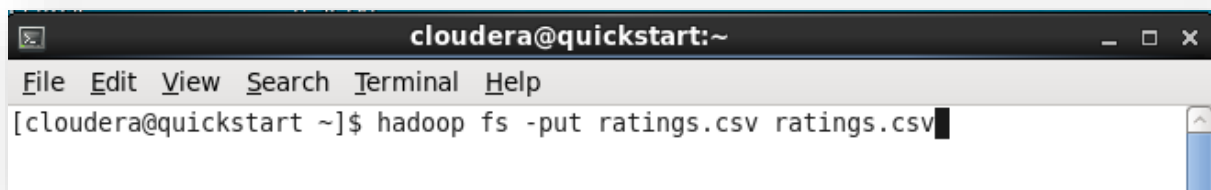
## MapReduce

- 1) Create a new project with class ex. Movie
- 2) Add 3 External Archives



Path:

- File System/usr/lib/hadoop/hadoop-common.jar
  - File System/usr/lib/hadoop/client/commons-cli-1.2.jar
  - File System/usr/lib/Hadoop/client-0.20/hadoop-core.jar
- 3) Export the project in JAR file after everytime any changes are done  
Ex. movie.jar
  - 4) Load the csv file in Cloudera through Shared Folder (Remove the column names after copying it into Cloudera's Home)
  - 5) Put the csv file into HDFS  
hadoop fs -put movie.csv movie.csv (or any different name)



- 6) Now for executing MapReduce job:  
hadoop jar movie.jar Movie movie.csv temp

```
[cloudera@quickstart ~]$ hadoop jar MapReduce.jar Movie ratings.csv temp9
24/05/08 13:23:14 INFO client.RMPProxy: Connecting to ResourceManager at /0.0.0.0:8032
24/05/08 13:23:39 INFO input.FileInputFormat: Total input paths to process : 1
```

7) Output:

hadoop fs -ls temp

hadoop fs -cat temp/part-r-00000

```
[cloudera@quickstart ~]$ hadoop fs -cat temp8/part-r-00000  
Suggested movie with highest rating: /n1      3.97165991902834
```