

Hadoop Steps

- 1) Cloudera → Eclipse → file → new → JavaProject → current version
- 2) ProjectName → Next → Libraries → Add External Jars → file system → user → lib → hadoop → all jar files → ok
- 3) Add External jar → file system → user → lib → hadoop → client → all jar files finish.
- 4) src → right click → new → class → class name → finish
OR
src → right click → new → package (name package same as project)
package → right click → new → class → class Name → finish.
- 5) code copy Paste change and save
- 6) Export Jar file
Right Click on Java Project → export → Java → Jar file → next → home/cloudera/desktop /ProjectName.jar finish.

7) check Tax file on desktop.

8) Mounting folder.

- Ubuntu Desktop

create new folder

- Cloudera desktop

create new folder

On cloudera

Devices → shared folder setting →

machine folder right click → Add

shared folder → other → Select folder

Select

- Auto mount

- Make Permanent

OK

Terminal :-

- su

- password - cloudera

- cd Desktop

- ls

- mount -t vboxsf folderName
(Ubuntu)

folder Name
(cloudera)

copy file on ubuntu so that
they will be seen on cloudera

Commands to Run:-

- ① hdfs dfs -mkdir /inputfolder
- ② hdfs dfs -put /home/cloudera/Desktop/data/music-dataset.csv /inputfolder
- ③ hdfs dfs -cat /inputfolder/music-dataset.csv

hadoop jar /home/cloudera/Desktop/Name.jar
Project Name

Jarfile name
→ .jar

④ hadopp → jar → /home/cloudera/Desktop/jarfile.jar,
→ Package Name. Driver class → /inputfolders/musicdataset
or
direct driver class

↓
1 out 1

- ⑤ hdfs dfs -ls /out1
- ⑥ hdfs dfs -cat /out1/part-r-00000

5 zeros
→