

Steps to execute map reduce program

1. Open virtual box
2. Create a text file a.txt with the data 1+2+3+4+5+6
3. Copy the text file from local file system to HDFS

Process:

1. Open the terminal
2. Give the command startCDH.sh(to start cloudera distributed hadoop)
3. Type the command

hadoop fs -put file from local file system path file on to HDFS path

example:

```
hadoop fs -put a.txt dc/a.txt
```

dc is the folder on HDFS

How to create the folder???

```
hadoop fs -mkdir dc
```

4. Open eclipse, copy the code by creating 3 java files(runner,mapper,reducer)
5. Right Click on folder goto->build path->configure build path->click on library-> add external jar->hadoop core – 1.2.1 After completion of the program , make a jar file
ow to make a jar file???

Go to the folder where u have written ur code right
click on the folder symbol

Folder→right click-> go to export-→go to java→select
jar→give some name and click on ok

6. Copy the jar file on to the eclipse workspace →project
folder→bin

4. Go to the folder bin copy the path.

5. Open the terminal

Give the command

cd copied path

6. To run the program type the command

hadoop jar created jar name Mainclassname HDFS

filename output directory name

Example:

hadoop jar sum.jar sum_runner dc/a.txt out

The map reduce program starts executing