WordCount Program with Hadoop 2.7 with Ubuntu 14.04



Data Science Lab, The Department of Computer Science,

KSKV Kachchh University.

Web: http://cs.kutchuni.edu.in

**Steps:**

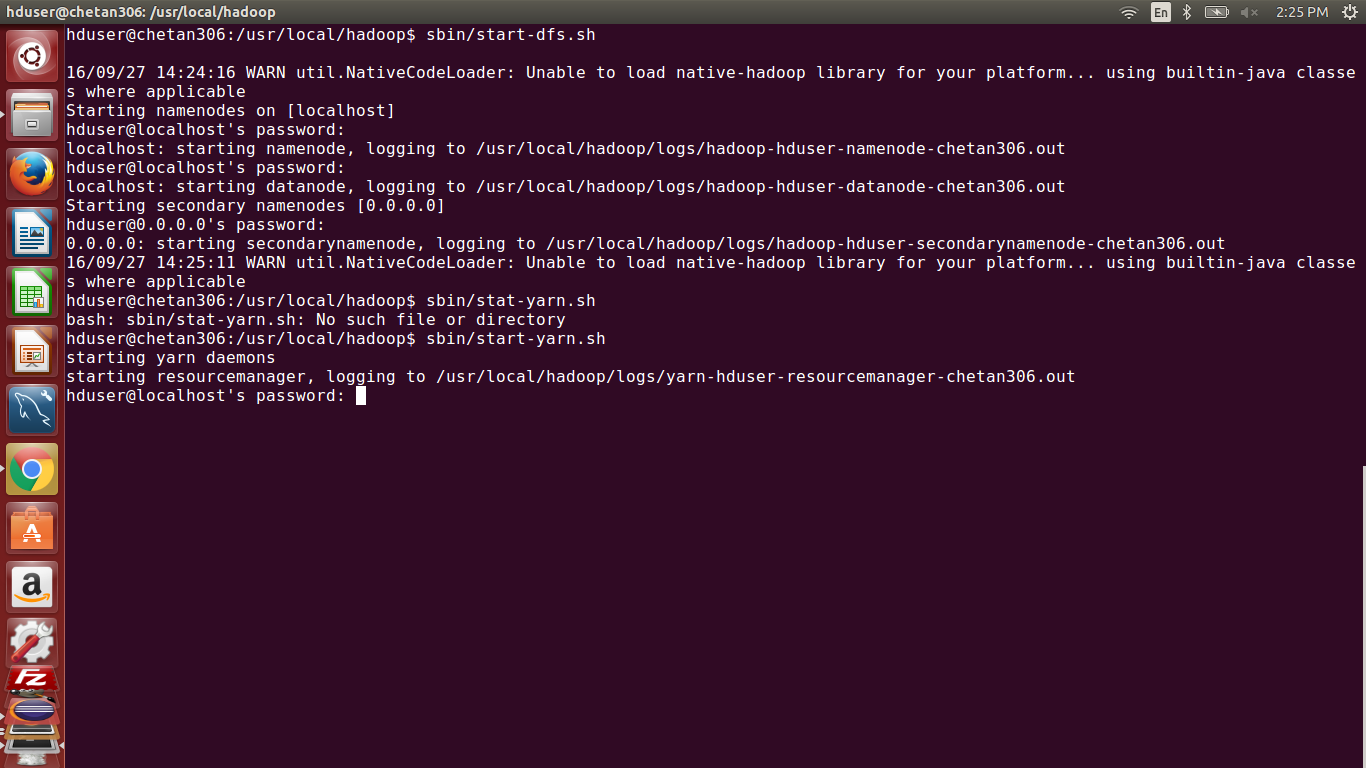
1) Login as a Hadoop User:

su hduser

2) Start hadoop deamon services

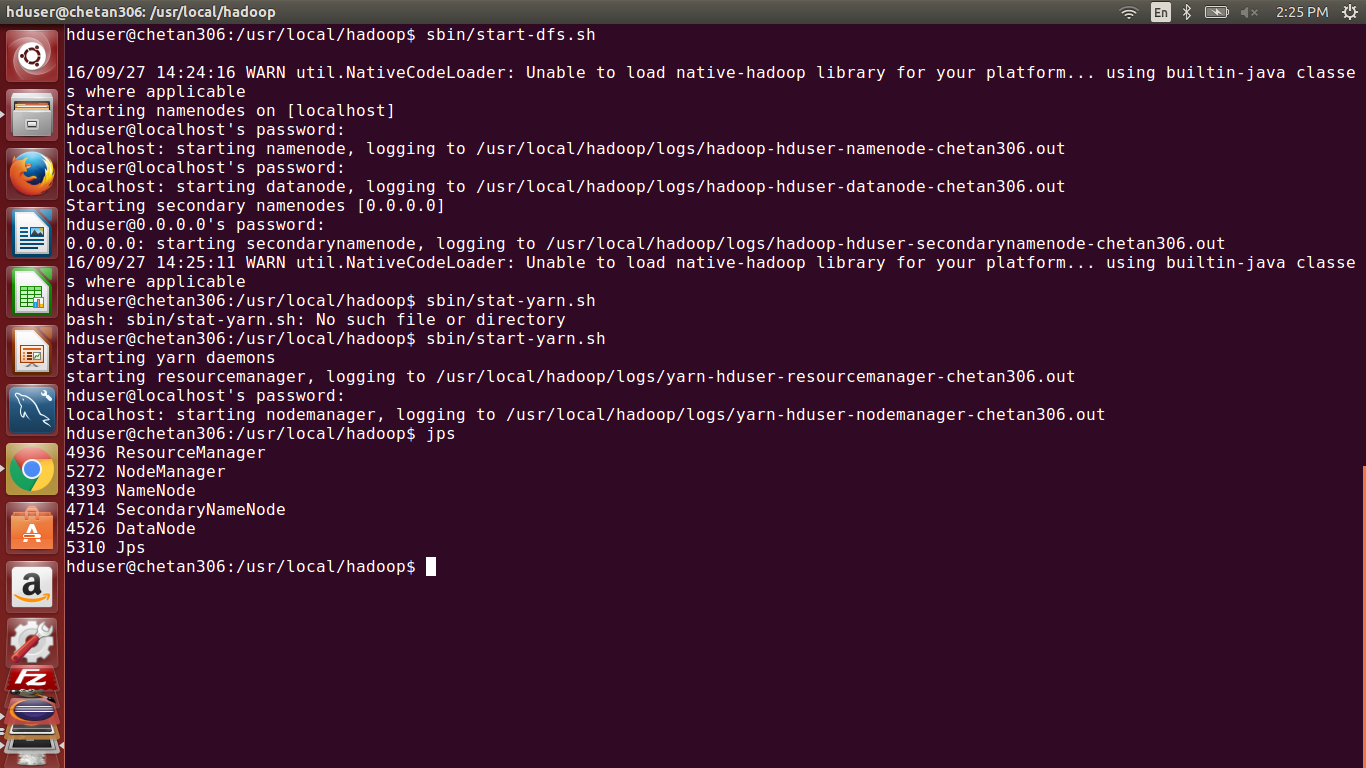
sbin/start-dfs.sh

sbin/start-yarn.sh



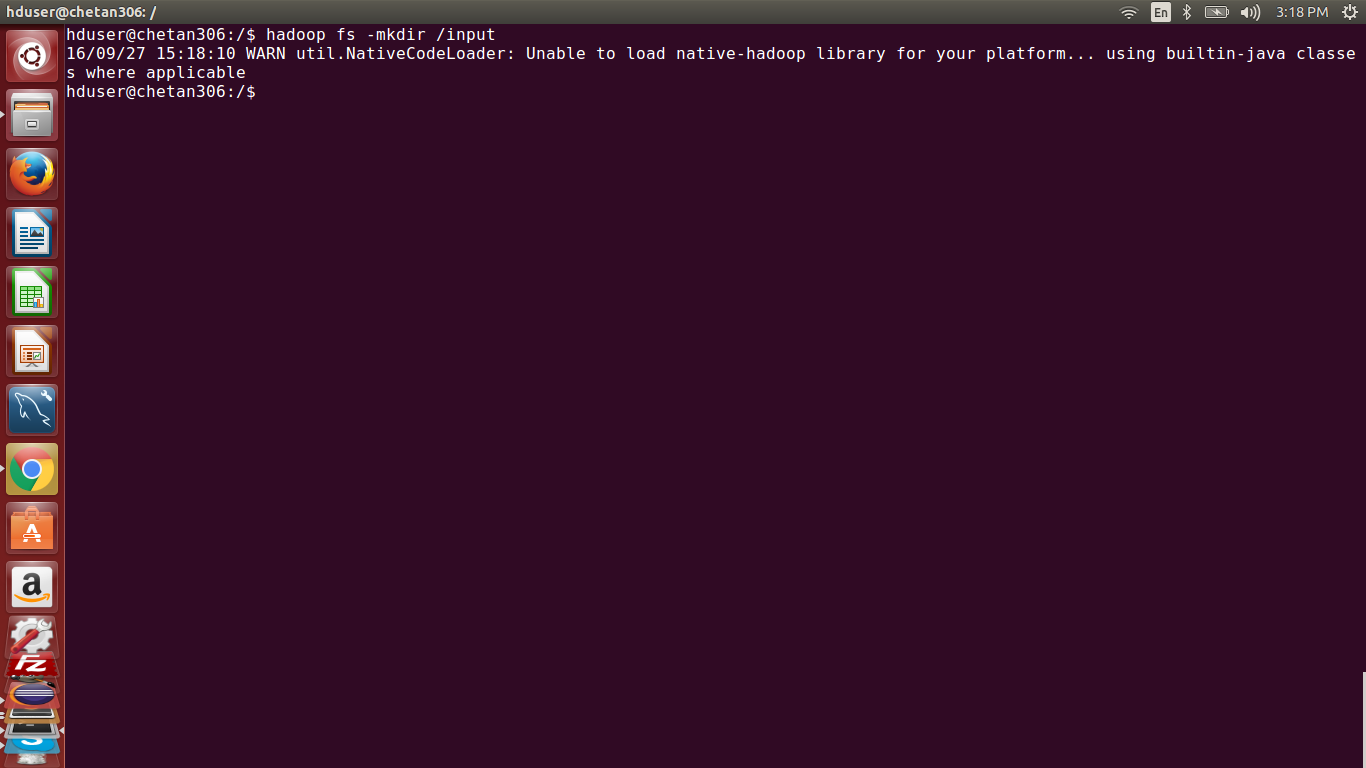
3) Check whether all deamon services are up or not

jps



4) Create directory in HDFS, Note: make sure wherever you are in the console , Hadoop user should have previlegies to access it.

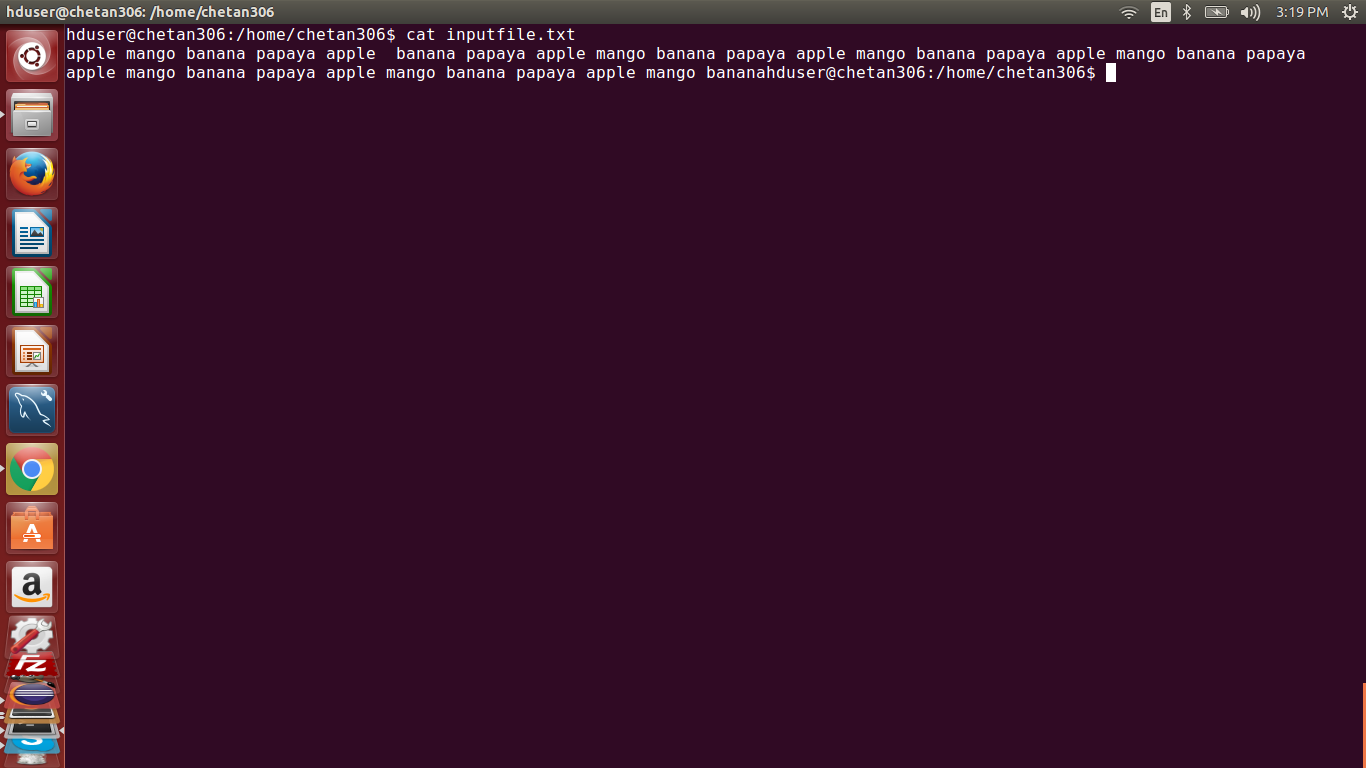
hadoop fs -mkdir /input

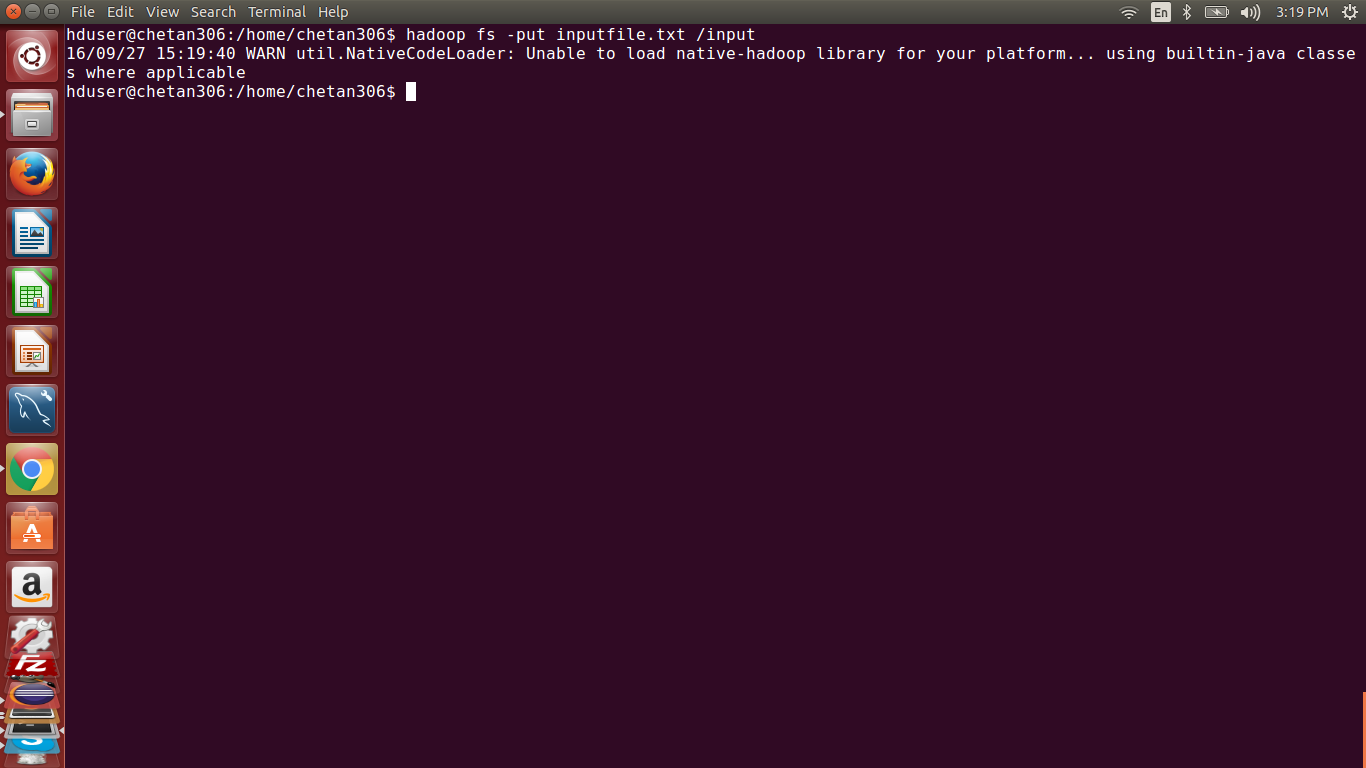


5) Transfer textfile to HDFS

hadoop fs -put inputfile.txt /input

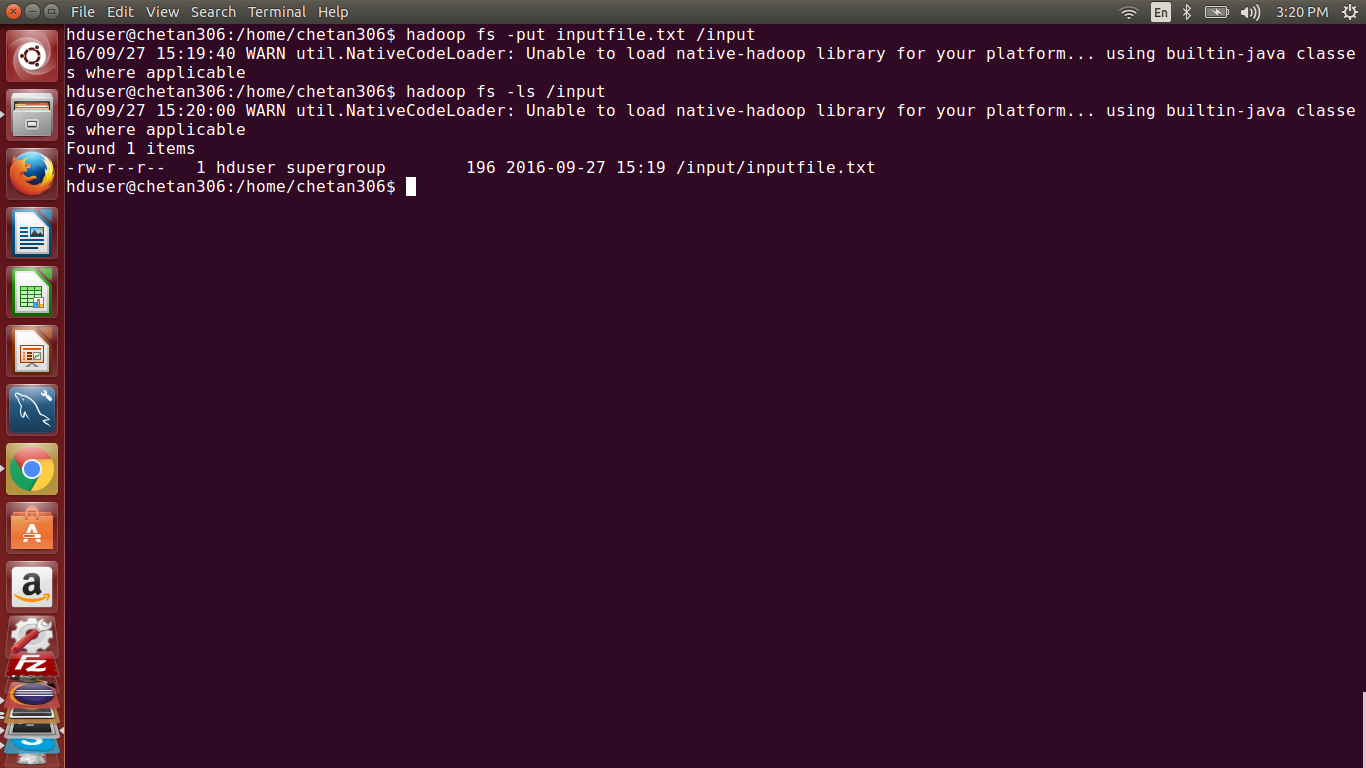
Description: store your inputfile.txt at root / home directory which has contents, you can view it with **cat inputfile.txt**





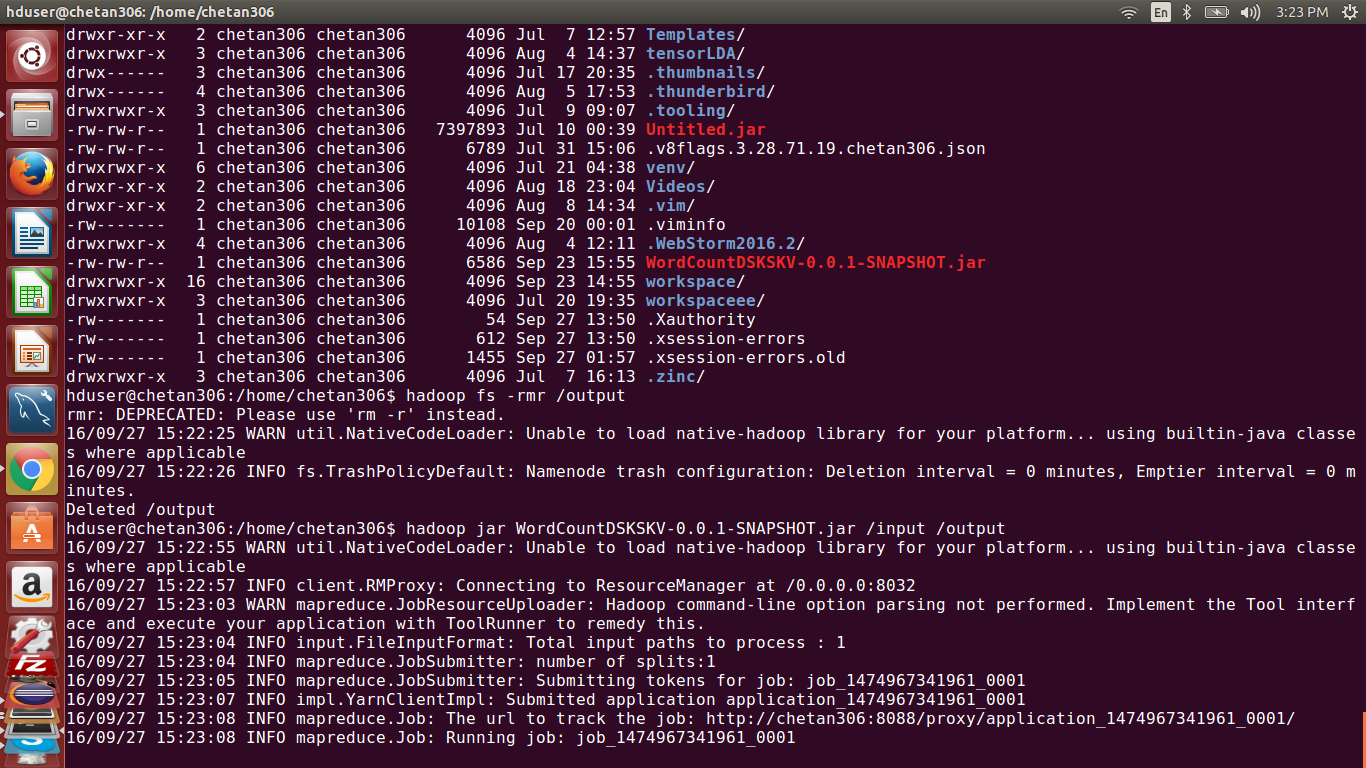
6) Check whether file is transferred successfully

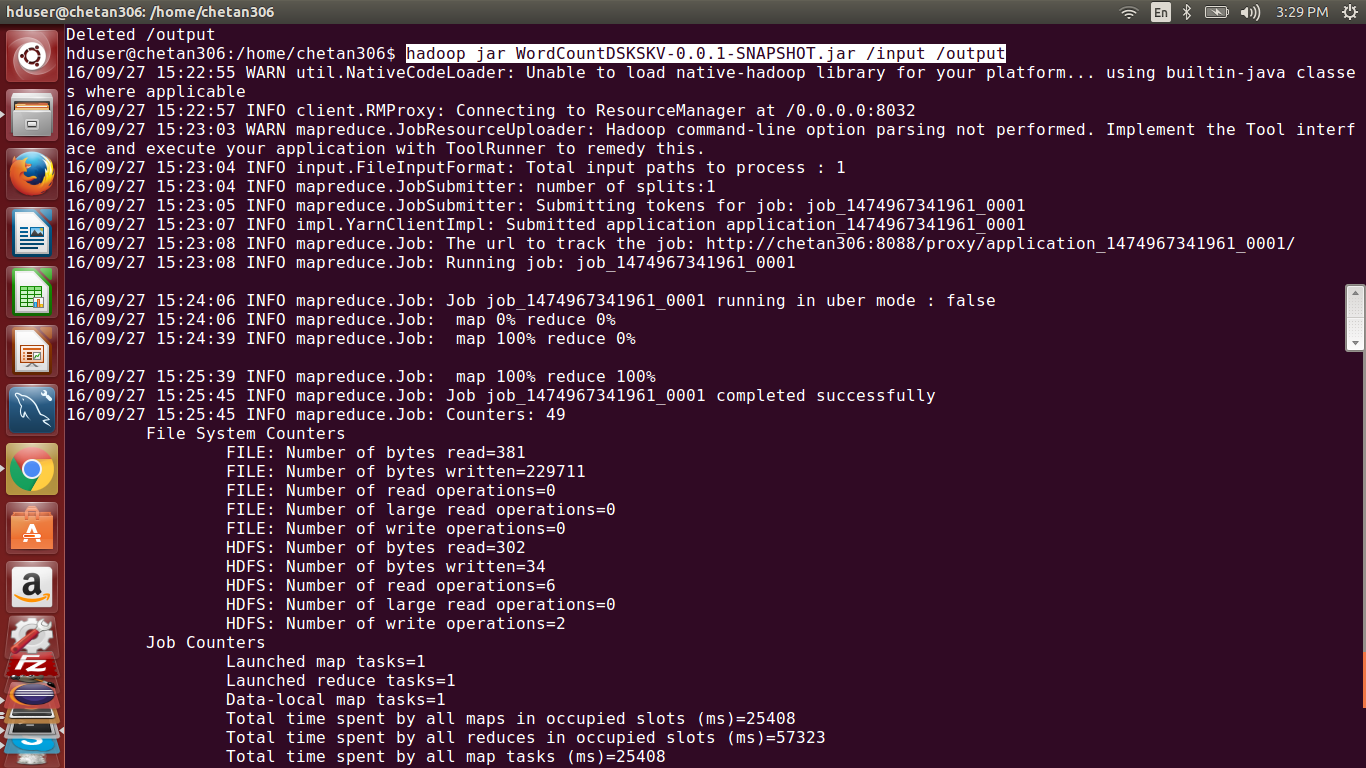
hadoop fs -ls /input



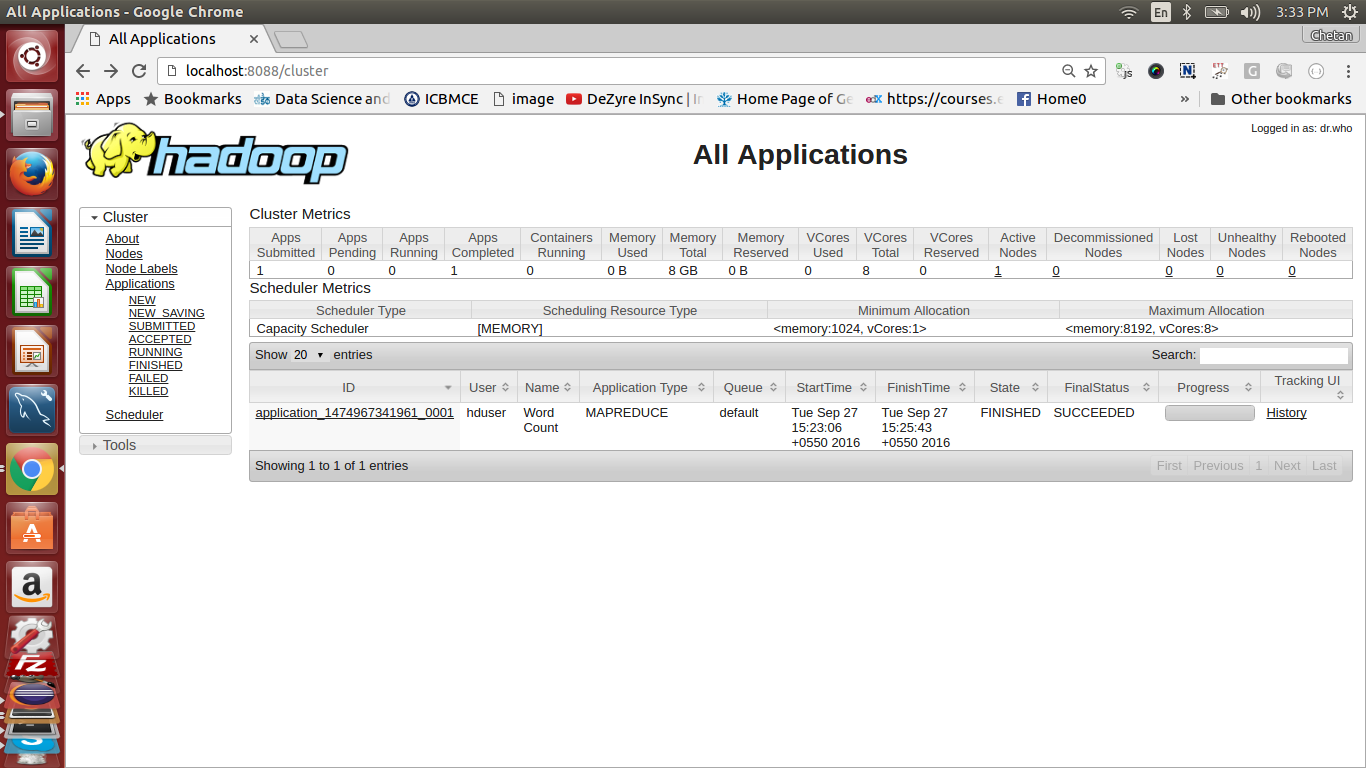
7) execute hadoop job by providing Hadoop Program executable Jar file and input directory path where text file is there and output directory path where you are looking to store process data.

hadoop jar WordCountDSKSKV-0.0.1-SNAPSHOT.jar /input /output





8) Check Job Details and Status at Hadoop Cluster UI: <http://localhost:8088/cluster>



9) Check Processed Directory has processed files ?

hadoop fs -ls /output

10) Read your desired output from Hadoop Job.

hadoop fs -cat /output/part-r-00000

