BIG DATA ENGINEERING FOR HADOOP & SPARK TRAINING

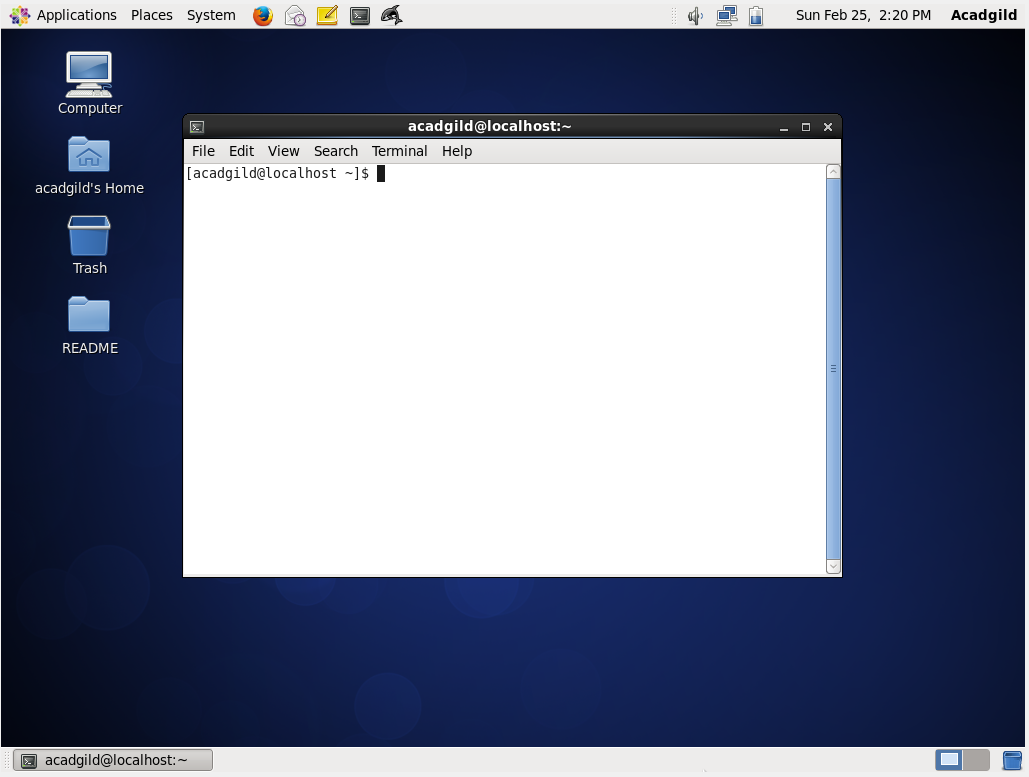
Assignment 1

1. Start Hadoop Single Node VM

i) First open Virtual Box and then Import Acadgild appliance and then start the VM using default settings



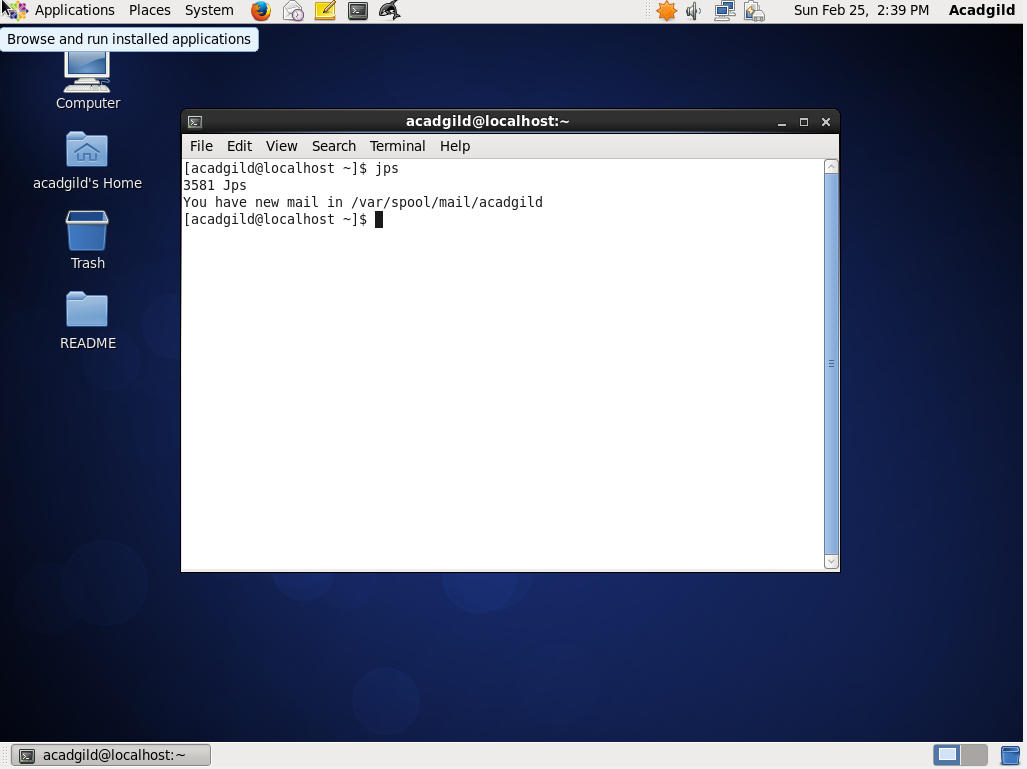
ii)Now Start the VM and provide password for Acadgild user and then open terminal

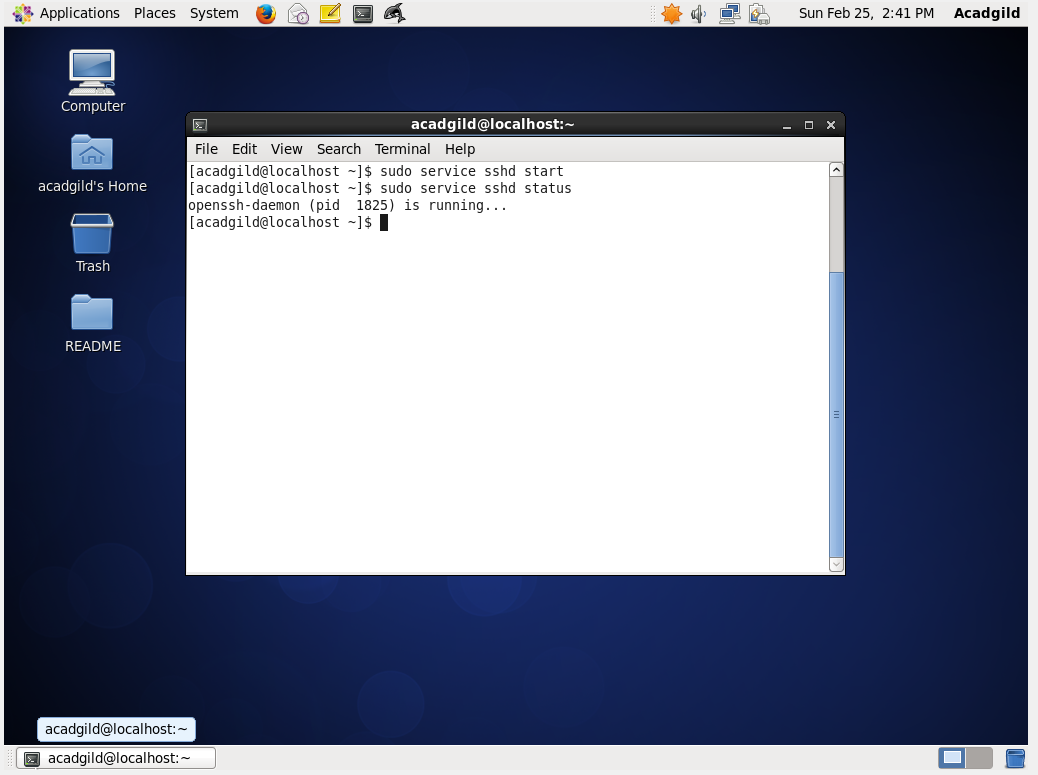


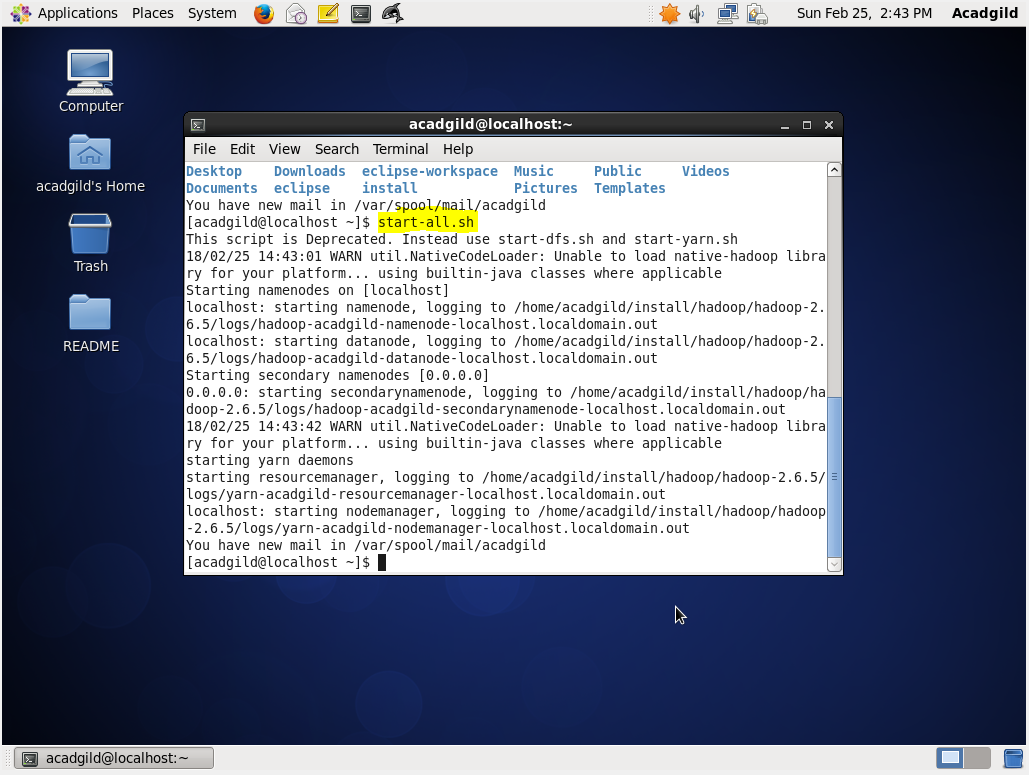
iii) Type “jps” command to check any deamons are running if not use “start-all.sh” to start the deamons

jps(Java Vitual Machine Process Status ) is a command which dispays all the hadoop processes that are running on JVM

start-all.sh script will start all the Hadoop deamons



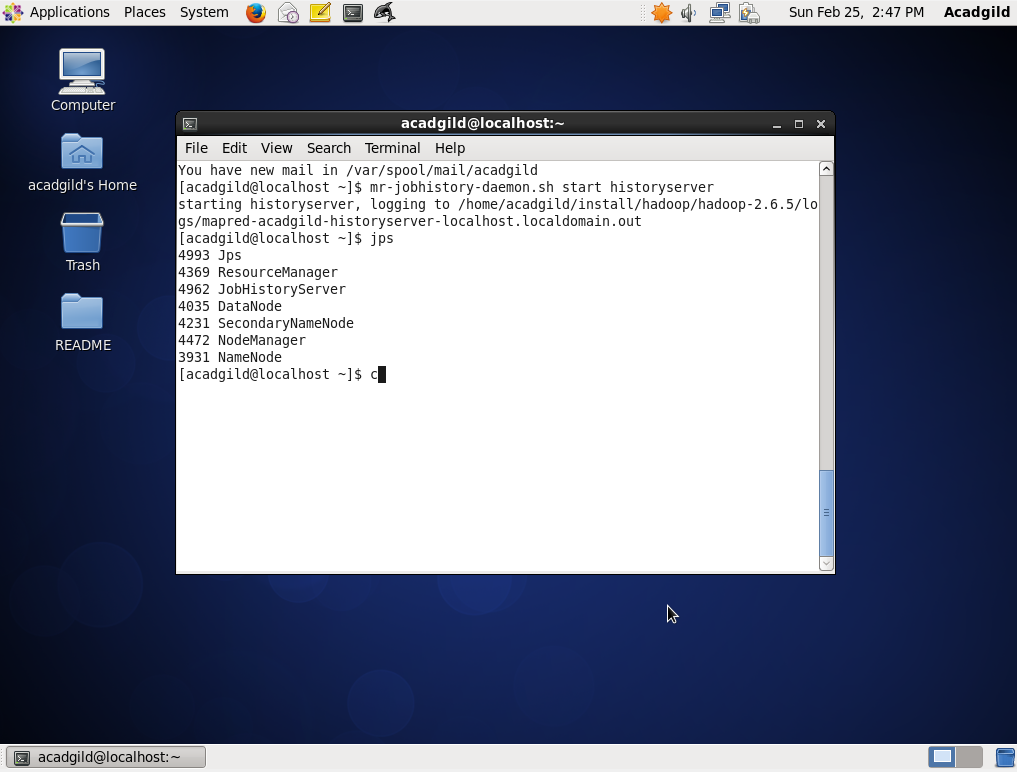




start-all.sh script will start **namenodes,secondary namenodes, resource manager and node manager** hadoop deamons.

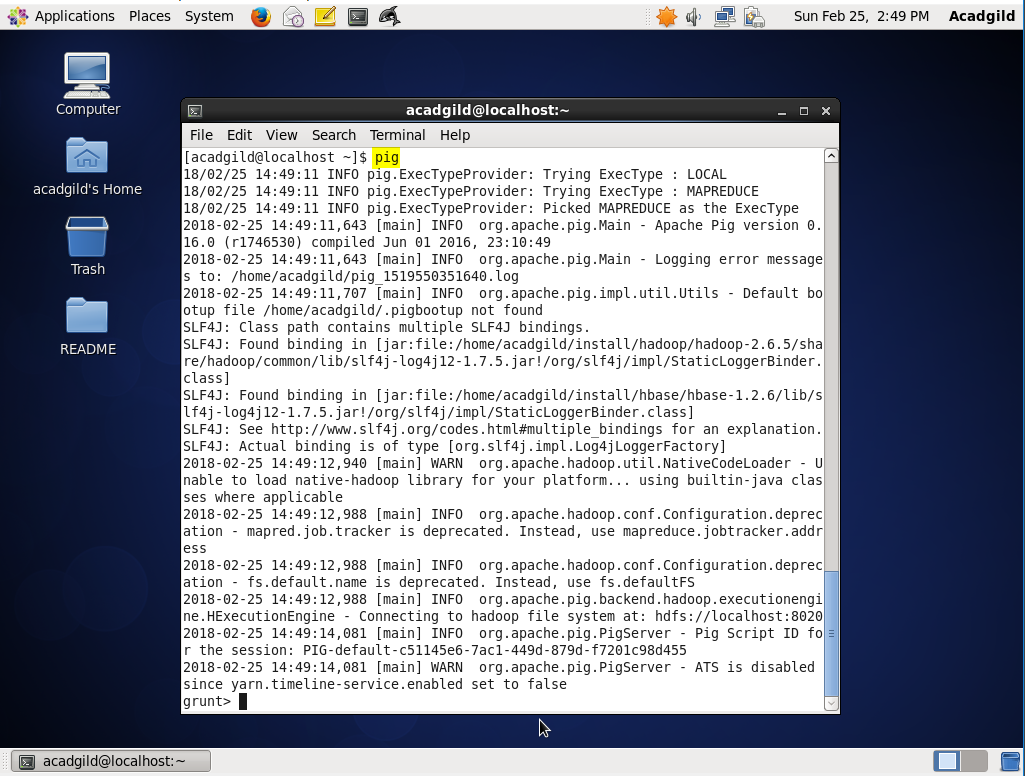
iv )for starting historyserver use the command “mr-jobhistory-deamon.sh start historyserver”

Now type “jps” command to check all the deamons that are started in previous command



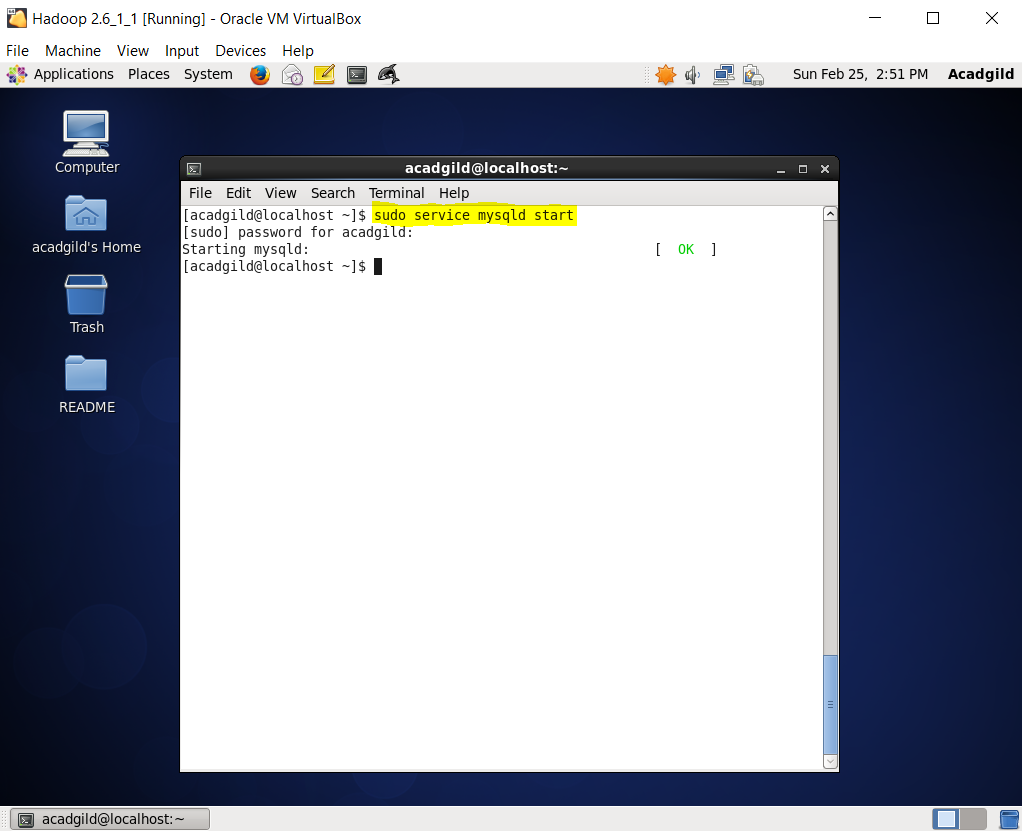
v) Pig is a high level scripting language that is used with Apache Hadoop. Pig enables data workers to write complex data transformations without knowing Java.

Starting Pig in HDFS mode

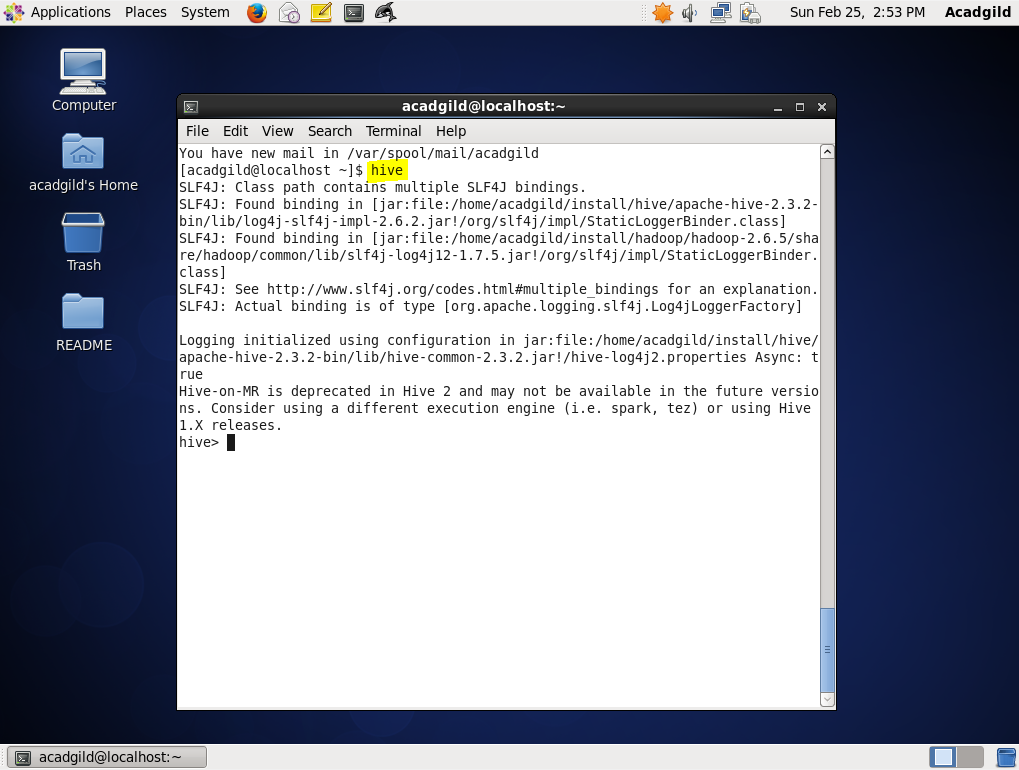


vi) Starting hive service

first start mysql service and then we have to start hive

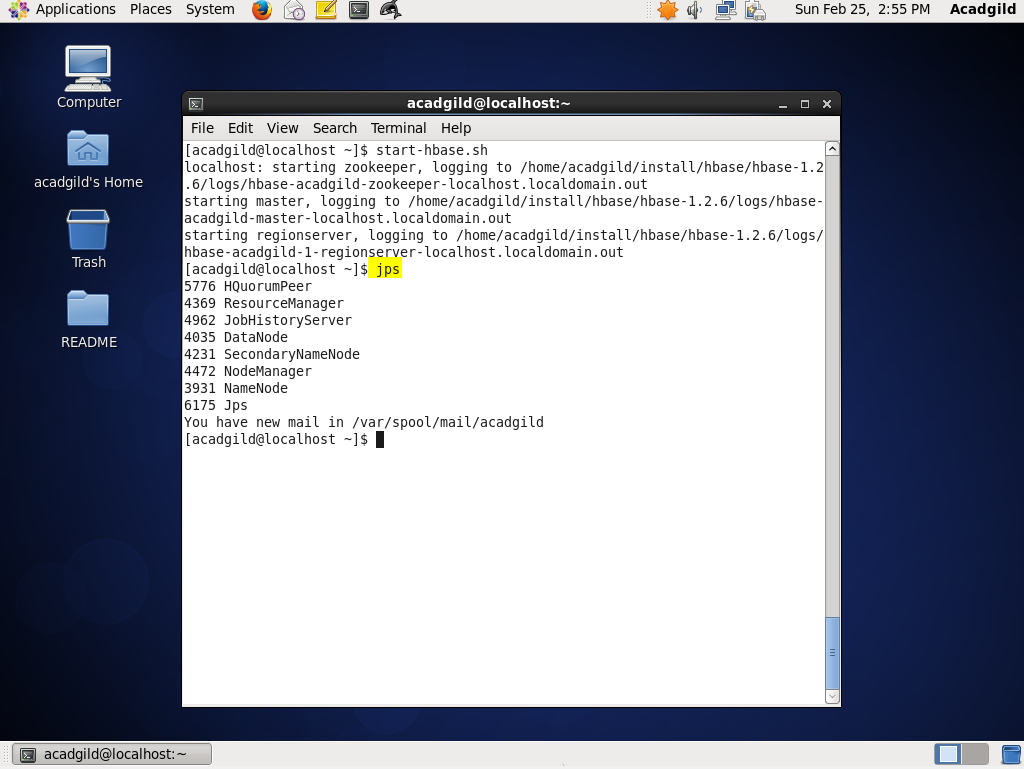


The Apache Hive distributed storage. By using Hive, we can access files stored in Hadoop Distributed File System (HDFS is used to querying and managing large datasets residing in) or in other data storage systems such as Apache HBase



vii) Start hbase using “start-hase.sh” script

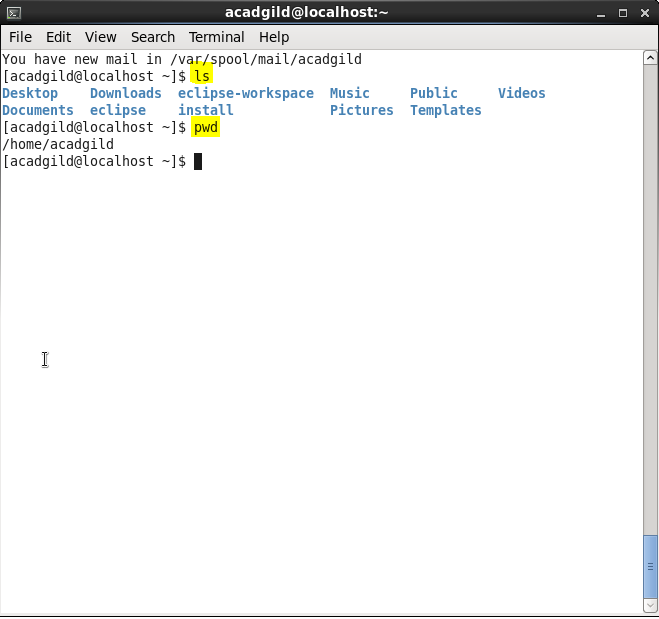
now type jps command to check all the hadoop deamons that we started are running.



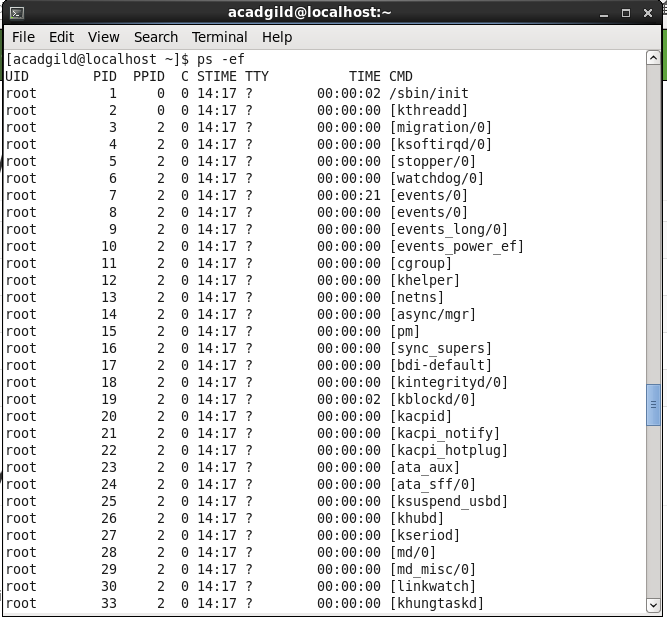
1. Linux Command

ls : lists all the files in the present working directory

pwd : displays the present working directory

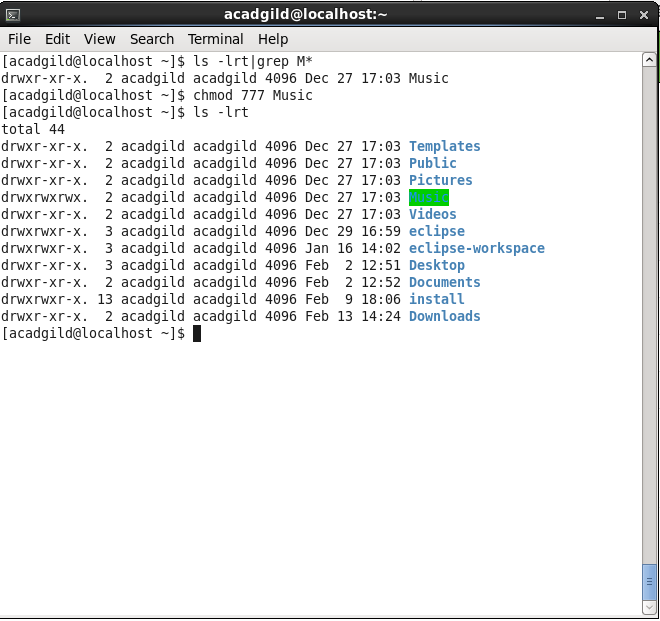


Ps –ef : displays all the processes running

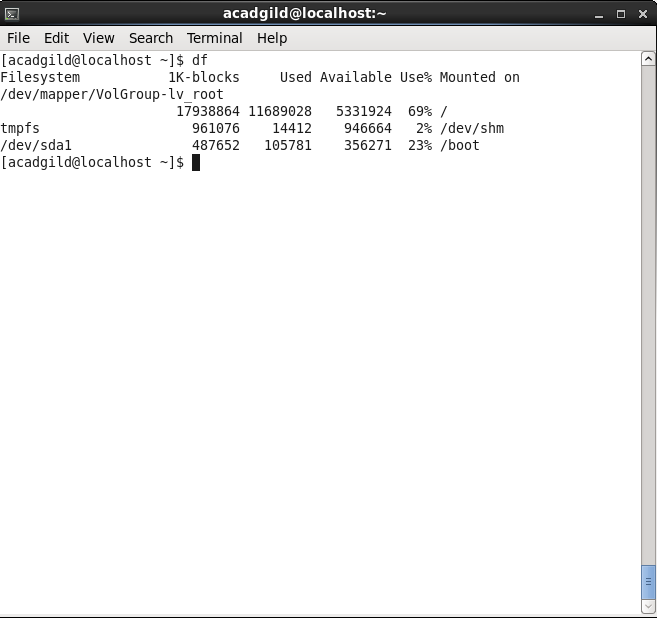


ls –lrt : displays files and their permission set for user,group,others

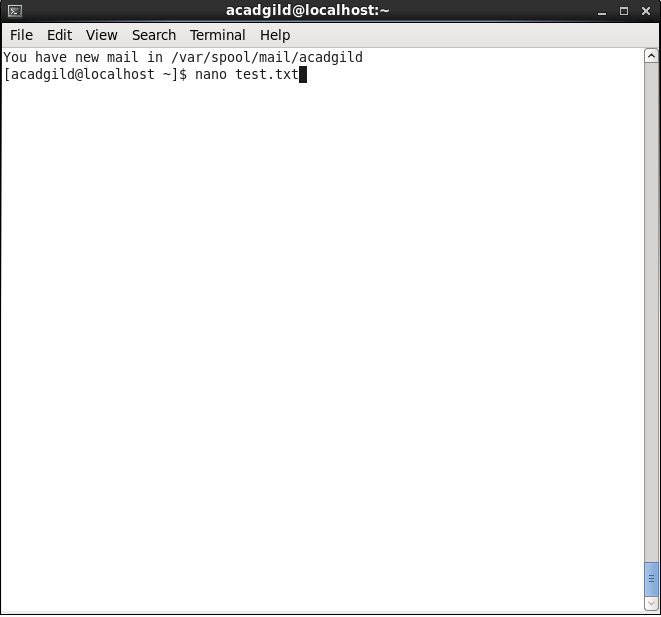
chmod 777 <file name> changes the permissions set for user,group,others



df : displays the free space on the disk

****

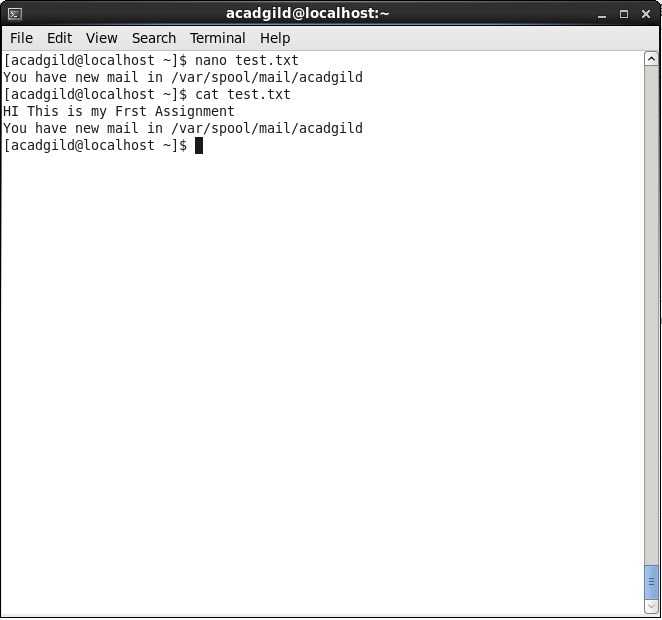
Creating a file using command nano

****

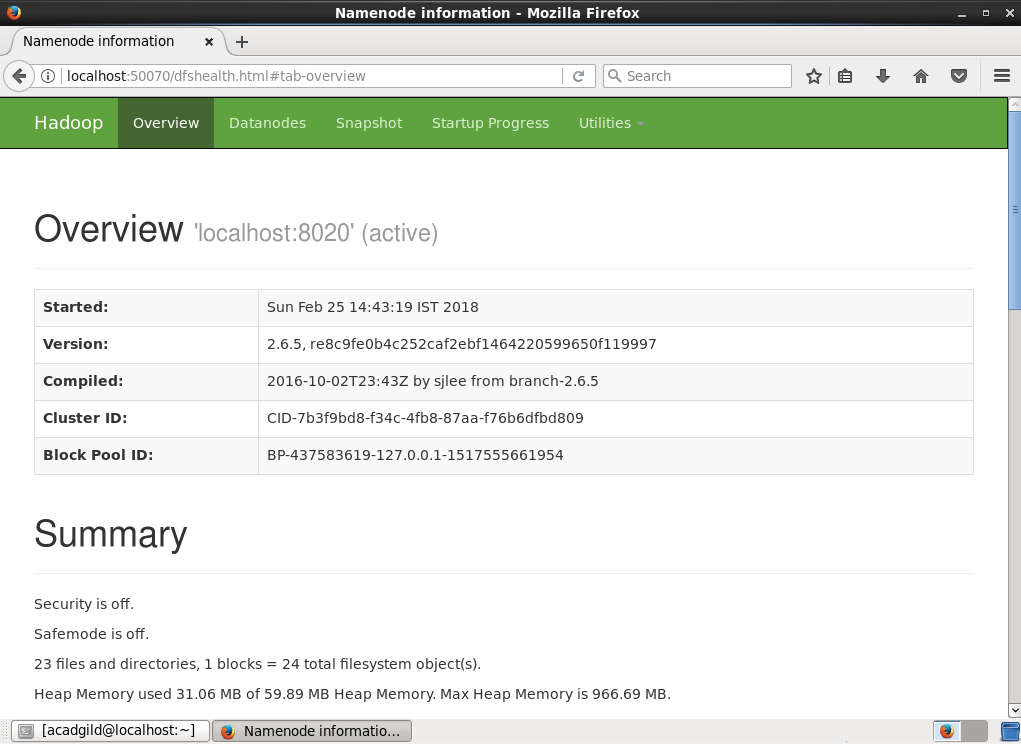
Add some text in the file the save the file using ctrl+x and then type Y and then provide a name.

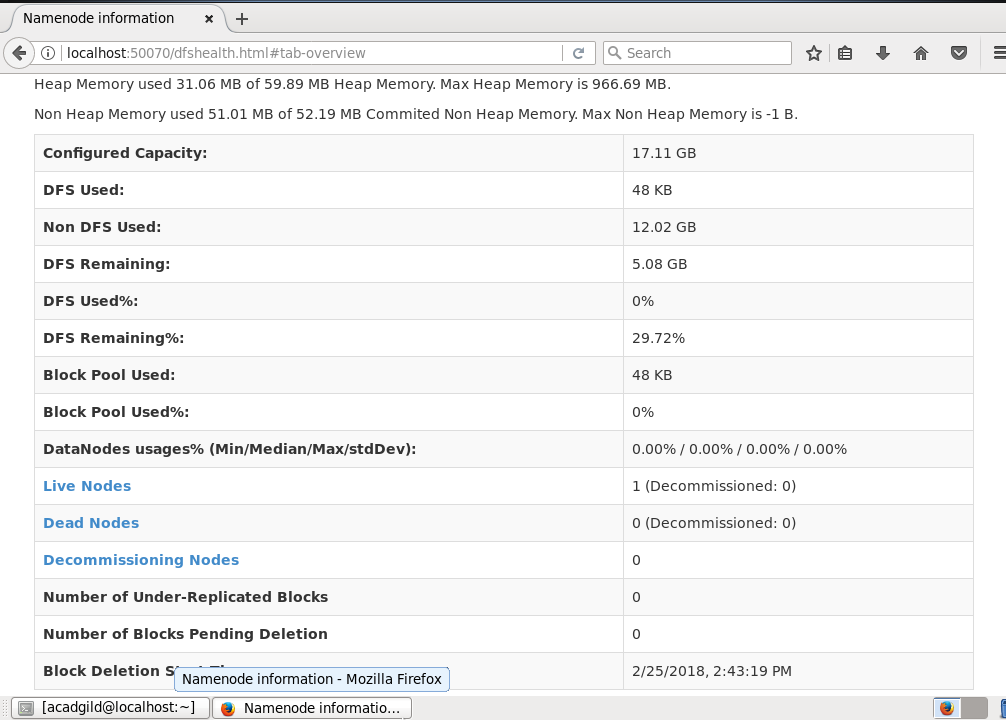
****

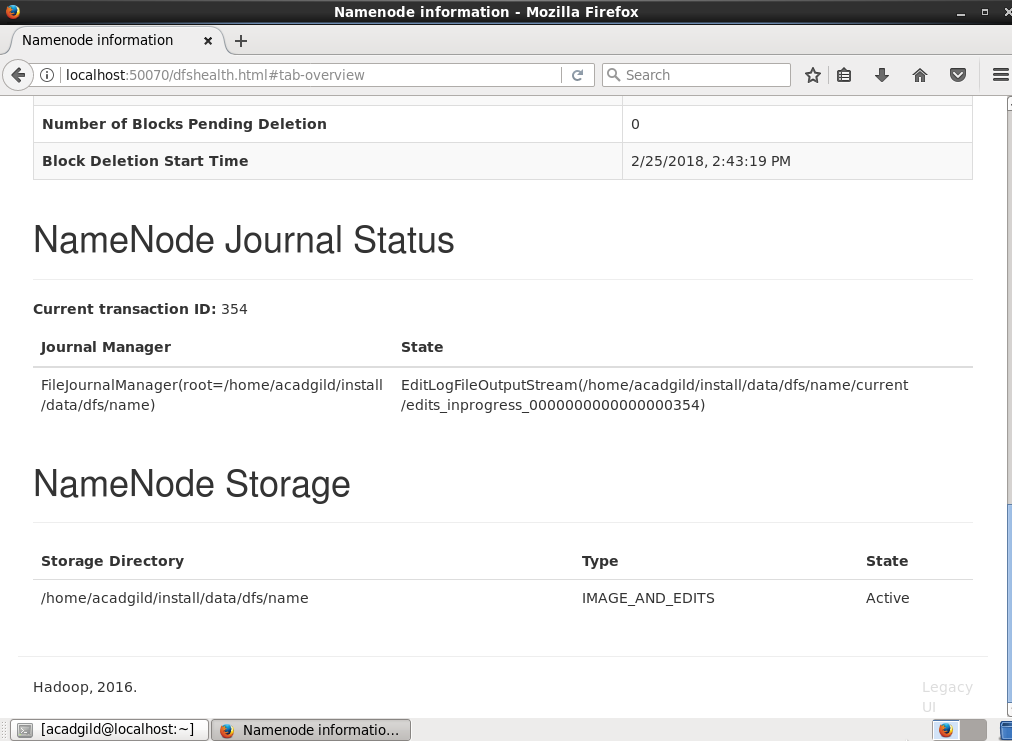
Use cat command to display the contents of the file “cat <filename>”



1. Opening HDFC Homepage using the url “http://localhost:50070”



****

****