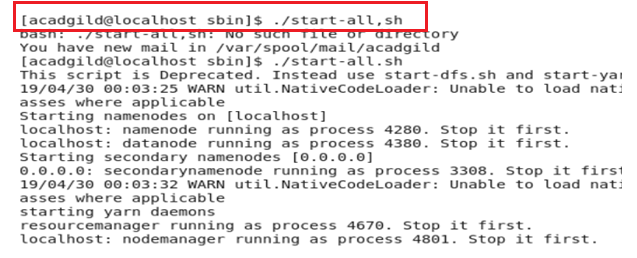
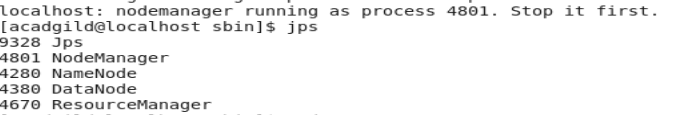
Assignment 1

1. Start-all.sh : used to start all the Hadoop daemons.( A **daemon** is a type of program on Unix-like operating systems that runs unobtrusively in the background.)

**Hadoop** has 5 **daemons**.They are NameNode, DataNode, Secondary NameNode, ResourceManager, NodeManager..



1. JPS command : **JPS** (Java Virtual Machine Process Status Tool) is a command is used to check all the **Hadoop** daemons like NameNode, DataNode, ResourceManager, NodeManager etc. which are running on the machine.



* **Namenode** – It runs on master node for HDFS. It stores Metadata i.e. number of blocks, their location, on which Rack, which Datanode the data is stored and other details.
* **Datanode** – It runs on slave nodes for HDFS. It is responsible for storing actual data in HDFS. Datanode performs read and write operation as per the request of the clients. Replica block of Datanode consists of 2 files on the file system.
* Resource Manager– It runs on YARN master node for MapReduce.
* Node Manager – It runs on YARN slave node for MapReduce.

Map Reduce :



|  |  |  |  |
| --- | --- | --- | --- |
| User id | Movie id | Rating | Timestamp |
| 196 | 242 | 3 | 881250949 |
| 186 | 302 | 3 | 885469756 |
| 196 | 377 | 1 | 965832651 |
| 244 | 51 | 2 | 885695437 |
| 166 | 346 | 1 | 886397596 |
| 186 | 474 | 4 | 698563214 |
| 186 | 265 | 2 | 888885526 |

Mapper



196:242 186:302 196:377 244:51 166:346 186:474 186:265



Shuffle & Sort

166:346 186:302,474,265 196:242,377 244:51



Reducer

166:1 186:3 196:2 244:1

1. Some Unix commands:

ls – list all file and folders

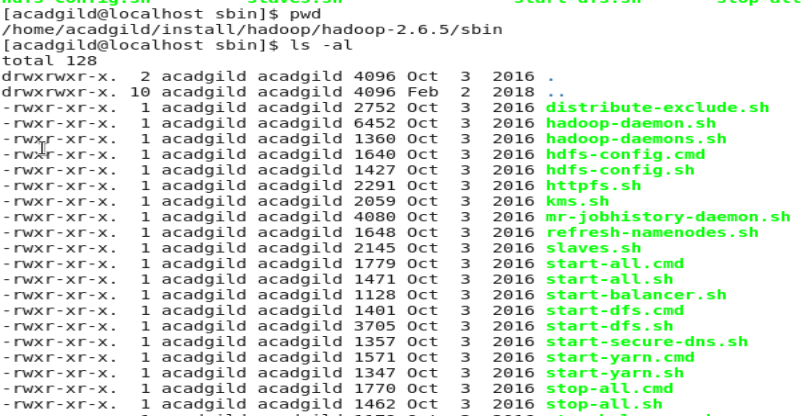
ls – al – ordered listng of files and folders

pwd – print working directory – used to print the current path

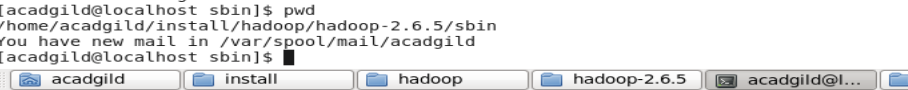
ps – currently active processes.

Grep – used for searching text strings in a command, file or anywhere possible

CAT – create single or multiple files, view contain of file, **concatenate** files and **redirect** output in terminal or files.







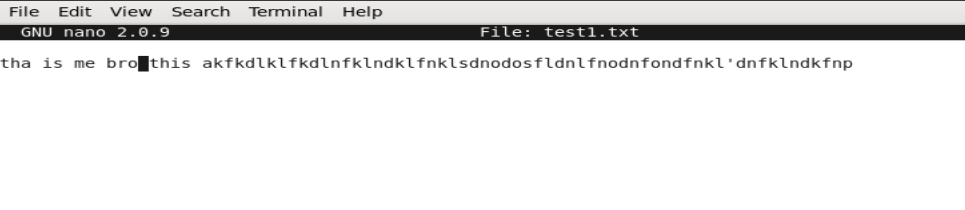
1. Create the file using nano editor and add some content to it.

**Uses of CAT**

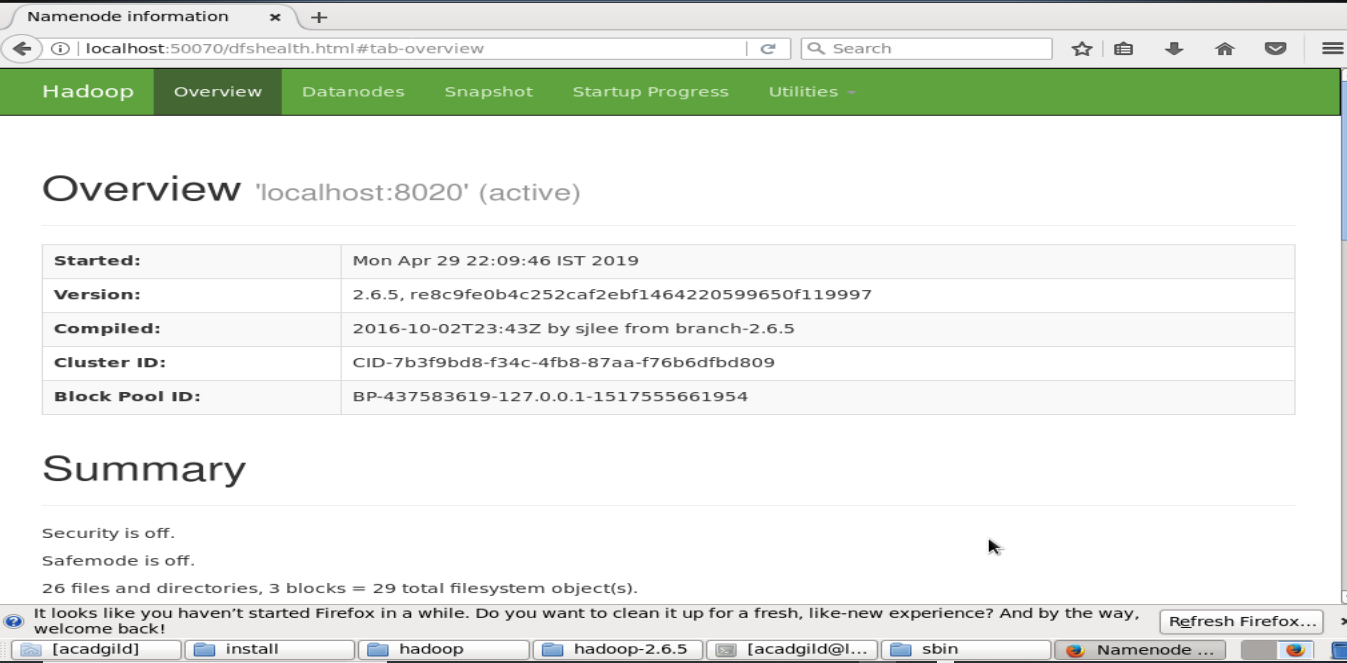
[[1]](#footnote-1)

**Nano - command line text editor in linux.**





1. **Open the HDFS web page by opening the browser http://localhost:50070**



1. [↑](#footnote-ref-1)