

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..

```
[acadgild@localhost sbin]$ ./start-all.sh
bash: ./start-all.sh: No such file or directory
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost sbin]$ ./start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
19/04/30 00:03:25 WARN util.NativeCodeLoader: Unable to load native
classes where applicable
Starting namenodes on [localhost]
localhost: namenode running as process 4280. Stop it first.
localhost: datanode running as process 4380. Stop it first.
Starting secondary namenodes [0.0.0.0]
0.0.0.0: secondarynamenode running as process 3308. Stop it first.
19/04/30 00:03:32 WARN util.NativeCodeLoader: Unable to load native
classes where applicable
starting yarn daemons
resourcemanager running as process 4670. Stop it first.
localhost: nodemanager running as process 4801. Stop it first.
```

- 2) 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.

```
localhost: nodemanager running as process 4801. Stop it first.
[acadgild@localhost sbin]$ jps
9328 Jps
4801 NodeManager
4280 NameNode
4380 DataNode
4670 ResourceManager
```

- **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

3) Some Unix commands:

ls – list all file and folders

ls – al – ordered listing 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.

```
[acadgild@localhost sbin]$ pwd
/home/acadgild/install/hadoop-2.6.5/sbin
[acadgild@localhost sbin]$ ls -al
total 128
drwxrwxr-x.  2 acadgild acadgild 4096 Oct  3  2016 .
drwxrwxr-x. 10 acadgild acadgild 4096 Feb  2  2018 ..
-rwxr-xr-x.  1 acadgild acadgild 2752 Oct  3  2016 distribute-exclude.sh
-rwxr-xr-x.  1 acadgild acadgild 6452 Oct  3  2016 hadoop-daemon.sh
-rwxr-xr-x.  1 acadgild acadgild 1360 Oct  3  2016 hadoop-daemons.sh
-rwxr-xr-x.  1 acadgild acadgild 1640 Oct  3  2016 hdfs-config.cmd
-rwxr-xr-x.  1 acadgild acadgild 1427 Oct  3  2016 hdfs-config.sh
-rwxr-xr-x.  1 acadgild acadgild 2291 Oct  3  2016 httpfs.sh
-rwxr-xr-x.  1 acadgild acadgild 2059 Oct  3  2016 kms.sh
-rwxr-xr-x.  1 acadgild acadgild 4080 Oct  3  2016 mr-jobhistory-daemon.sh
-rwxr-xr-x.  1 acadgild acadgild 1648 Oct  3  2016 refresh-namenodes.sh
-rwxr-xr-x.  1 acadgild acadgild 2145 Oct  3  2016 slaves.sh
-rwxr-xr-x.  1 acadgild acadgild 1779 Oct  3  2016 start-all.cmd
-rwxr-xr-x.  1 acadgild acadgild 1471 Oct  3  2016 start-all.sh
-rwxr-xr-x.  1 acadgild acadgild 1128 Oct  3  2016 start-balancer.sh
-rwxr-xr-x.  1 acadgild acadgild 1401 Oct  3  2016 start-dfs.cmd
-rwxr-xr-x.  1 acadgild acadgild 3705 Oct  3  2016 start-dfs.sh
-rwxr-xr-x.  1 acadgild acadgild 1357 Oct  3  2016 start-secure-dns.sh
-rwxr-xr-x.  1 acadgild acadgild 1571 Oct  3  2016 start-yarn.cmd
-rwxr-xr-x.  1 acadgild acadgild 1347 Oct  3  2016 start-yarn.sh
-rwxr-xr-x.  1 acadgild acadgild 1770 Oct  3  2016 stop-all.cmd
-rwxr-xr-x.  1 acadgild acadgild 1462 Oct  3  2016 stop-all.sh
```

```

[acadgild@localhost sbin]$ cat > test.txt
hello can you hear me now then please here me.
^C
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost sbin]$ nano test.txt
[acadgild@localhost sbin]$ grep hear test.txt
hello can you hear me now then please here me.
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost sbin]$ grep hear test
grep: test: No such file or directory
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost sbin]$ grep -r hear test.txt
hello can you hear me now then please here me.
[acadgild@localhost sbin]$ grep -i "hear" test.txt
hello can you hear me now then please here me.
[acadgild@localhost sbin]$ grep -iw "hear" test.txt
hello can you hear me now then please here me.
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost sbin]$ cat > test.txt
can you imagine how much more i can hear please keep imagining you hear me now.
how is that you can hear me and i am not able to hear.
^C
[acadgild@localhost sbin]$ nano test.txt
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost sbin]$ grep -iw "hear" test.txt
can you imagine how much more i can hear please keep imagining you hear me now.
how is that you can hear me and i am not able to hear.
[acadgild@localhost sbin]$ tout test.txt
bash: tout: command not found
[acadgild@localhost sbin]$ touch test.txt
[acadgild@localhost sbin]$ nano test.txt
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost sbin]$ nano test.txt
[acadgild@localhost sbin]$
[acadgild@localhost sbin]$ grep -iw "hear" test.txt
can you imagine how much more i can hear please keep imagining you hear me now.
how is that you can hear me and i am not able to hear.
i can hear i cannot
[acadgild@localhost sbin]$ █

You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mail]$ nano test.txt
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost mail]$ cd ..
[acadgild@localhost spool]$ cd ..
[acadgild@localhost var]$ cd ..
[acadgild@localhost /]$ ls
bin  copyright  etc  lib  lost+found  misc  net  proc  sbin  srv  tmp  var
boot  dev  home  lib64  media  mnt  opt  root  selinux  sys  usr
[acadgild@localhost /]$ cd home
[acadgild@localhost home]$ sudo nano test.txt
[sudo] password for acadgild:
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost home]$ ls
acadgild  test.txt
[acadgild@localhost home]$ cat > test.txt
bash: test.txt: Permission denied
[acadgild@localhost home]$ sudo cat test.txt
this is mario man.
[acadgild@localhost sbin]$ pwd
/home/acadgild/install/hadoop/hadoop-2.6.5/sbin
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost sbin]$ █

```

 acadgild
  install
  hadoop
  hadoop-2.6.5
  acadgild@l...

- 4) Create the file using nano editor and add some content to it.

Uses of CAT

```
[acadgild@localhost sbin]$ cat < test.txt
can you imagine how much more i can hear please keep imagining you hear me now.
how is that you can hear me and i am not able to hear.
what do you think man
i can hear i cannot
whats the reason you think like that.
you have new mail in /var/spool/mail/acadgild
[acadgild@localhost sbin]$ cat test
cat: test: No such file or directory
you have new mail in /var/spool/mail/acadgild
[acadgild@localhost sbin]$ cat test.txt
can you imagine how much more i can hear please keep imagining you hear me now.
how is that you can hear me and i am not able to hear.
what do you think man
i can hear i cannot
whats the reason you think like that.
[acadgild@localhost sbin]$ cat test.txt more
can you imagine how much more i can hear please keep imagining you hear me now.
how is that you can hear me and i am not able to hear.
what do you think man
i can hear i cannot
whats the reason you think like that.
cat: more: No such file or directory
you have new mail in /var/spool/mail/acadgild
[acadgild@localhost sbin]$ cat test.txt | more
can you imagine how much more i can hear please keep imagining you hear me now.
how is that you can hear me and i am not able to hear.
what do you think man
i can hear i cannot
whats the reason you think like that.
[acadgild@localhost sbin]$ cat -n test.txt
1 can you imagine how much more i can hear please keep imagining you hear me now.
2 how is that you can hear me and i am not able to hear.
3 what do you think man
4 i can hear i cannot
5 whats the reason you think like that.
[acadgild@localhost sbin]$
```

1

Nano - command line text editor in linux.

```
[acadgild@localhost sbin]$ nano test1.txt
File Edit View Search Terminal Help
GNU nano 2.0.9 File: test1.txt
tha is me bro this akfkdlklfkldlnfkldnklfnklsdnodosfldnlnfnodnfondfnkl'dnfnkldnkfnp
```

- 5) Open the HDFS web page by opening the browser <http://localhost:50070>

The screenshot shows a web browser window displaying the Hadoop NameNode information page. The browser's address bar shows the URL `localhost:50070/dfshealth.html#tab-overview`. The page has a green header with navigation tabs: Hadoop, Overview (selected), Datanodes, Snapshot, Startup Progress, and Utilities. The main content area is titled "Overview 'localhost:8020' (active)". Below the title is a table with the following information:

Started:	Mon Apr 29 22:09:46 IST 2019
Version:	2.6.5, re8c9fe0b4c252caf2ebf1464220599650f119997
Compiled:	2016-10-02T23:43Z by sjlee from branch-2.6.5
Cluster ID:	CID-7b3f9bd8-f34c-4fb8-87aa-f76b6dfbd809
Block Pool ID:	BP-437583619-127.0.0.1-1517555661954

Below the table is a "Summary" section. It states: "Security is off.", "Safemode is off.", and "26 files and directories, 3 blocks = 29 total filesystem object(s).". At the bottom of the page, there is a message from Firefox: "It looks like you haven't started Firefox in a while. Do you want to clean it up for a fresh, like-new experience? And by the way, welcome back!" with a "Refresh Firefox..." button. The browser's taskbar at the bottom shows several open windows, including "hadoop", "hadoop-2.6.5", and "NameNode ...".