

Session 1: BIG DATA & HADOOP – INTRODUCTION & CASE STUDIES

Assignment 1

1. Start Hadoop single node on AcadGild VM. The command is start-all.sh.

```
[acadgild@localhost ~]$ echo $HADOOP_HOME
/home/acadgild/install/hadoop/hadoop-2.6.5
[acadgild@localhost ~]$ cd $HADOOP_HOME
[acadgild@localhost hadoop-2.6.5]$ start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
18/08/13 14:08:52 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
Starting namenodes on [localhost]
localhost: starting namenode, logging to /home/acadgild/install/hadoop/hadoop-2.
6.5/logs/hadoop-acadgild-namenode-localhost.localdomain.out
localhost: starting datanode, logging to /home/acadgild/install/hadoop/hadoop-2.
6.5/logs/hadoop-acadgild-datanode-localhost.localdomain.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /home/acadgild/install/hadoop/ha
dooop-2.6.5/logs/hadoop-acadgild-secondarynamenode-localhost.localdomain.out
18/08/13 14:11:36 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
starting yarn daemons
starting resourcemanager, logging to /home/acadgild/install/hadoop/hadoop-2.6.5/
logs/yarn-acadgild-resourcemanager-localhost.localdomain.out
localhost: starting nodemanager, logging to /home/acadgild/install/hadoop/hadoo
p-2.6.5/logs/yarn-acadgild-nodemanager-localhost.localdomain.out
```

2) Run a JPS command to see if all Hadoop daemons are running.

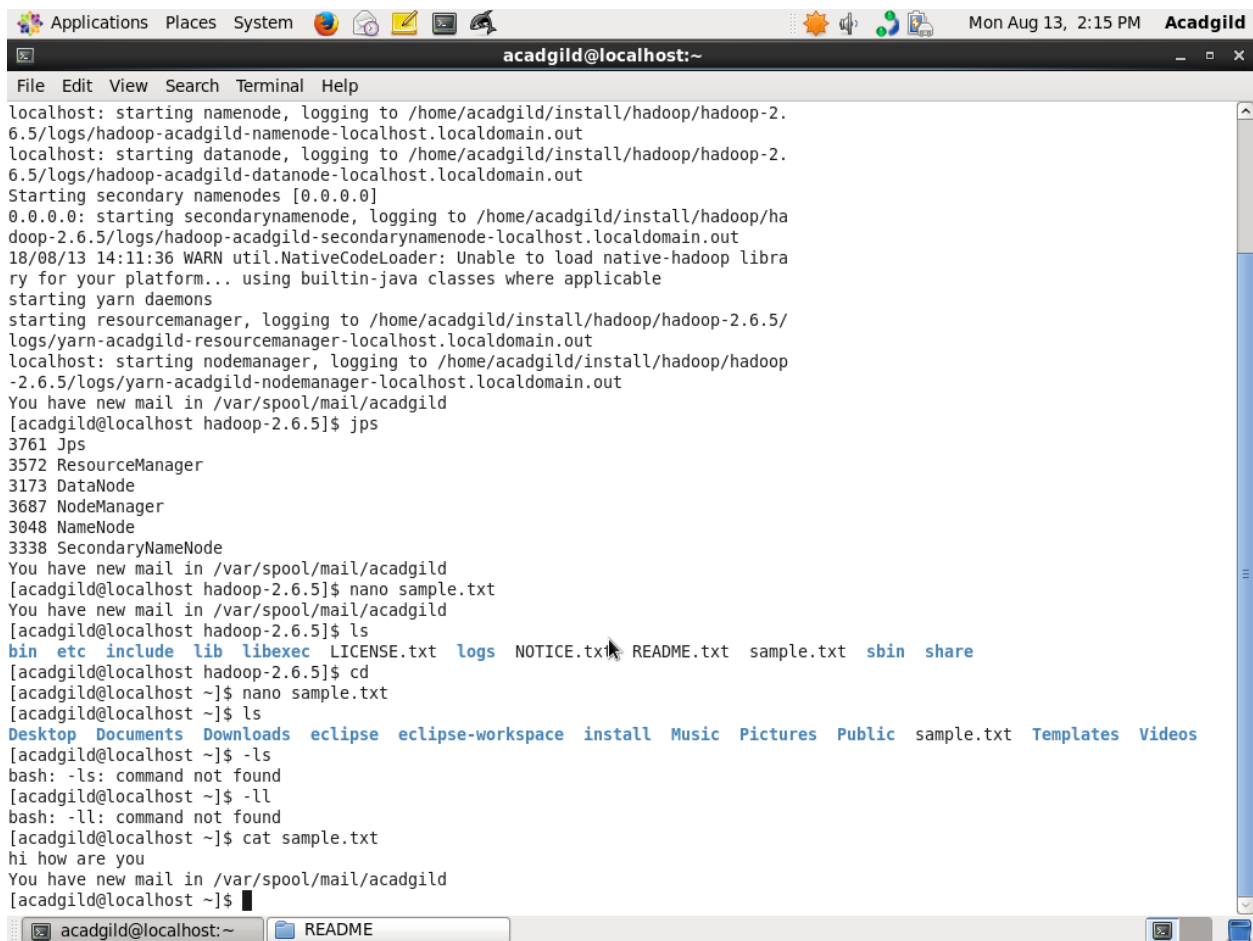
```
[acadgild@localhost hadoop-2.6.5]$ jps
3761 Jps
3572 ResourceManager
3173 DataNode
3687 NodeManager
3048 NameNode
3338 SecondaryNameNode
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost hadoop-2.6.5]$
```

3) Run few Unix commands like pwd, ls -ls, etc.

The screenshot shows a Linux terminal window titled 'acadgild@localhost:~'. The window has a menu bar with 'File', 'Edit', 'View', 'Search', 'Terminal', and 'Help'. The terminal output shows the user running several commands: 'ls', 'pwd', and 'ls -ls'. The 'ls' command lists files in the current directory, including 'Desktop', 'Documents', 'Downloads', 'eclipse', 'eclipse-workspace', 'install', 'Music', 'Pictures', 'Public', 'sample.txt', 'Templates', and 'Videos'. The 'pwd' command shows the current directory is '/home/acadgild'. The 'ls -ls' command provides a detailed listing of the files, including permissions, owner, group, size, date, and filename. The window also shows a taskbar at the bottom with icons for 'Applications', 'Places', 'System', and a clock showing 'Mon Aug 13, 2:22 PM'.

```
acadmild@localhost:~  
File Edit View Search Terminal Help  
You have new mail in /var/spool/mail/acadmild  
[acadmild@localhost ~]$ ls  
Desktop Documents Downloads eclipse eclipse-workspace install Music Pictures Public sample.txt Templates Videos  
[acadmild@localhost ~]$ pwd  
/home/acadmild  
[acadmild@localhost ~]$ /home/acadmild/ -ls  
bash: /home/acadmild/: is a directory  
[acadmild@localhost ~]$ ls -ls  
total 48  
4 drwxr-xr-x. 3 acadmild acadmild 4096 Aug 13 14:15 Desktop  
4 drwxr-xr-x. 2 acadmild acadmild 4096 Feb 2 2018 Documents  
4 drwxr-xr-x. 2 acadmild acadmild 4096 Feb 13 14:24 Downloads  
4 drwxrwxr-x. 3 acadmild acadmild 4096 Dec 29 2017 eclipse  
4 drwxrwxr-x. 3 acadmild acadmild 4096 Jan 16 2018 eclipse-workspace  
4 drwxrwxr-x. 13 acadmild acadmild 4096 Feb 9 2018 install  
4 drwxr-xr-x. 2 acadmild acadmild 4096 Dec 27 2017 Music  
4 drwxr-xr-x. 2 acadmild acadmild 4096 Dec 27 2017 Pictures  
4 drwxr-xr-x. 2 acadmild acadmild 4096 Dec 27 2017 Public  
4 -rw-rw-r--. 1 acadmild acadmild 15 Aug 13 14:14 sample.txt  
4 drwxr-xr-x. 2 acadmild acadmild 4096 Dec 27 2017 Templates  
4 drwxr-xr-x. 2 acadmild acadmild 4096 Dec 27 2017 Videos  
You have new mail in /var/spool/mail/acadmild  
[acadmild@localhost ~]$
```

4) Create a file from the terminal using nano editor (example: nano test.txt), and add some content in it. Cat it to see if the content is saved.



The screenshot shows a terminal window titled 'acadgild@localhost:~'. The window contains the following text:

```
localhost: starting namenode, logging to /home/acadgild/install/hadoop/hadoop-2.6.5/logs/hadoop-acadgild-namenode-localhost.localdomain.out
localhost: starting datanode, logging to /home/acadgild/install/hadoop/hadoop-2.6.5/logs/hadoop-acadgild-datanode-localhost.localdomain.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /home/acadgild/install/hadoop/hadoop-2.6.5/logs/hadoop-acadgild-secondarynamenode-localhost.localdomain.out
18/08/13 14:11:36 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
starting yarn daemons
starting resourcemanager, logging to /home/acadgild/install/hadoop/hadoop-2.6.5/logs/yarn-acadgild-resourcemanager-localhost.localdomain.out
localhost: starting nodemanager, logging to /home/acadgild/install/hadoop/hadoop-2.6.5/logs/yarn-acadgild-nodemanager-localhost.localdomain.out
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost hadoop-2.6.5]$ jps
3761 Jps
3572 ResourceManager
3173 DataNode
3687 NodeManager
3048 NameNode
3338 SecondaryNameNode
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost hadoop-2.6.5]$ nano sample.txt
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost hadoop-2.6.5]$ ls
bin  etc  include  lib  libexec  LICENSE.txt  logs  NOTICE.txt  README.txt  sample.txt  sbin  share
[acadgild@localhost hadoop-2.6.5]$ cd
[acadgild@localhost ~]$ nano sample.txt
[acadgild@localhost ~]$ ls
Desktop  Documents  Downloads  eclipse  eclipse-workspace  install  Music  Pictures  Public  sample.txt  Templates  Videos
[acadgild@localhost ~]$ -ls
bash: -ls: command not found
[acadgild@localhost ~]$ -ll
bash: -ll: command not found
[acadgild@localhost ~]$ cat sample.txt
hi how are you
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```

The terminal window has a menu bar with 'File', 'Edit', 'View', 'Search', 'Terminal', and 'Help'. The status bar at the bottom shows 'acadgild@localhost:~' and a 'README' file icon.

5) Open the hdfs web page by typing localhost:50070 in the browser. Check all the details of the HDFS.

ApplicationsPlacesSystem

Mon Aug 13, 2:24 PMAcadgild

Namenode information - Mozilla Firefox

Namenode information

localhost:50070/dfshealth.html#tab-overview

Search

Hadoop

Overview

Datanodes

Snapshot

Startup Progress

Utilities

Overview

'localhost:8020' (active)

Started:	Mon Aug 13 14:10:01 IST 2018
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

Summary

Security is off.

Safemode is off.

26 files and directories, 3 blocks = 29 total filesystem object(s).

Heap Memory used 31.62 MB of 49.21 MB Heap Memory. Max Heap Memory is 966.69 MB.

acadgild@localhost: ~

[README]

Namenode informatio...