

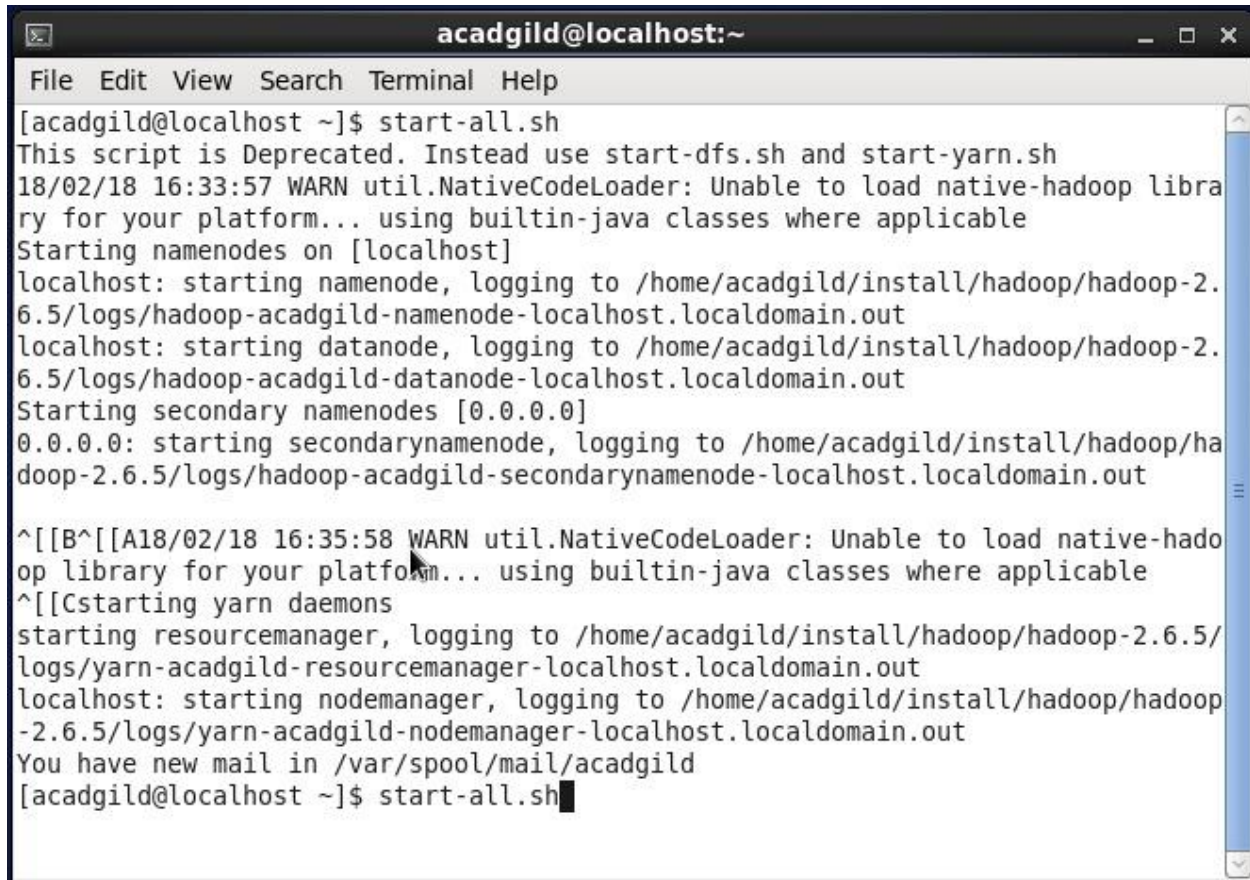


Big Data Hadoop & Spark Training

Assignment-1

Task1-Start Hadoop single node on AcadGild VM.

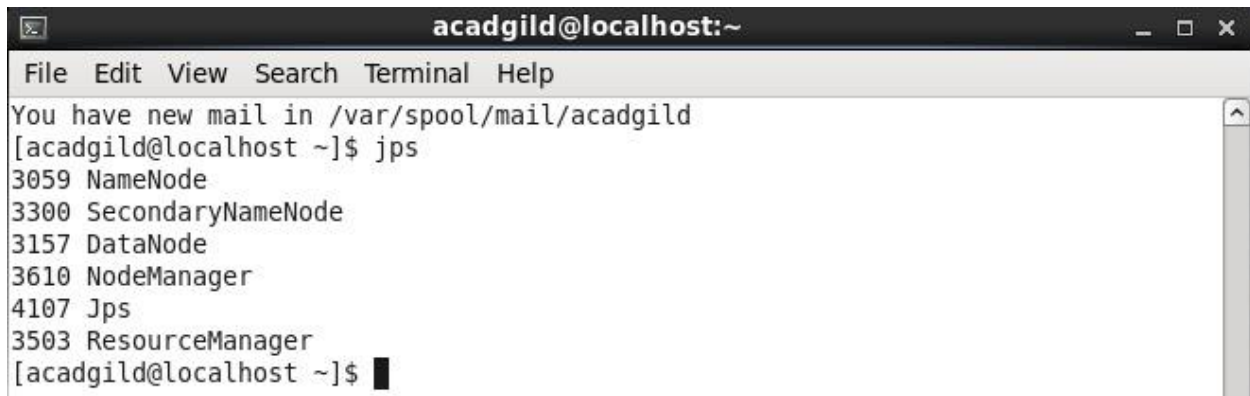
Started Hadoop single node using the command start-all.sh. Below screenshot provides the output of the command.

A terminal window titled 'acadgild@localhost:~' with a menu bar (File, Edit, View, Search, Terminal, Help). The terminal shows the execution of 'start-all.sh'. The output includes a deprecation warning, Hadoop version 2.6.5 logs for namenode, datanode, and secondarynamenode, and Yarn daemon logs for resourcemanager and nodemanager. A mail notification is also shown.

```
acadgild@localhost:~  
File Edit View Search Terminal Help  
[acadgild@localhost ~]$ start-all.sh  
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh  
18/02/18 16:33:57 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  
  
^[[B^[[A18/02/18 16:35:58 WARN util.NativeCodeLoader: Unable to load native-hado  
op library for your platform... using builtin-java classes where applicable  
^[[Cstarting 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 ~]$ start-all.sh
```

Task2- Run a JPS command to see if all Hadoop daemons are running.

Jps command is used to check all the Hadoop daemons like NameNode, DataNode, ResourceManager, NodeManager etc. which are running on the machine



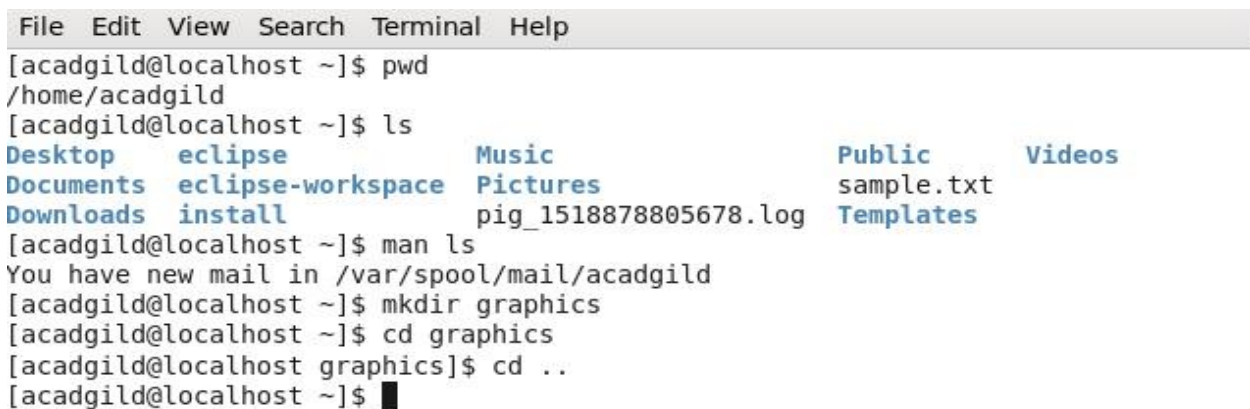
```
acadgild@localhost:~  
File Edit View Search Terminal Help  
You have new mail in /var/spool/mail/acadgild  
[acadgild@localhost ~]$ jps  
3059 NameNode  
3300 SecondaryNameNode  
3157 DataNode  
3610 NodeManager  
4107 Jps  
3503 ResourceManager  
[acadgild@localhost ~]$
```

Task3-Run few Unix commands like pwd, ls -ls, etc.

Used few unix commands which are as follows:-

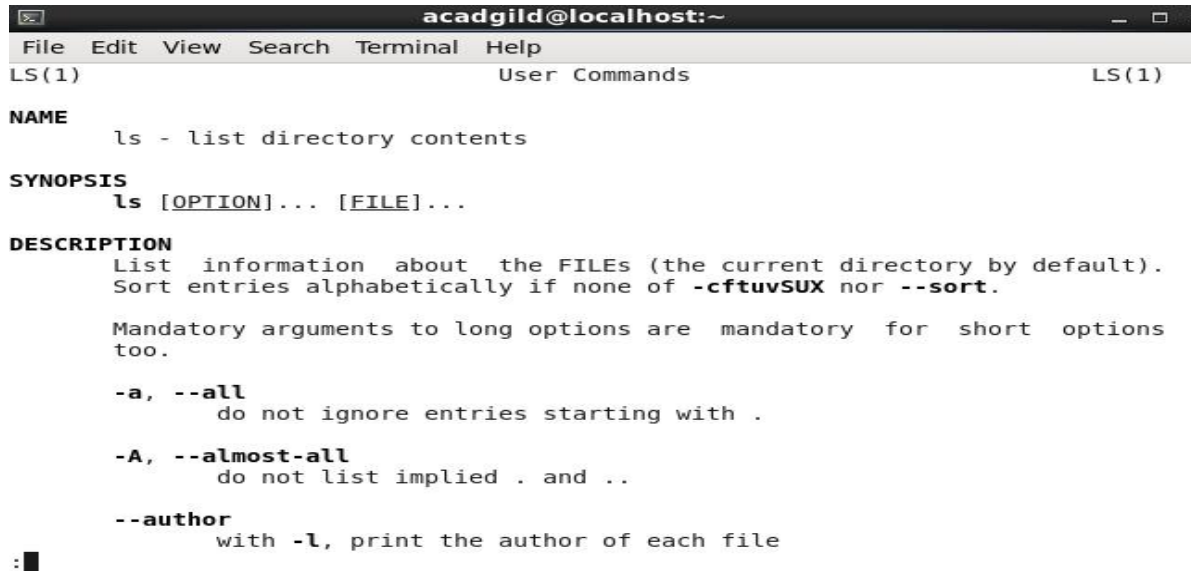
- pwd-This provides the full path name of the current directory.
- ls- This provides the list of all files in the current directory
- man ls-This provides online manual help about command
- mkdir- This helps to make a new directory
- cd- This helps to change the directory
- cd ..-This helps to move back one directory

Screenshots of outputs of the above commands are as below:-



```
File Edit View Search Terminal Help  
[acadgild@localhost ~]$ pwd  
/home/acadgild  
[acadgild@localhost ~]$ ls  
Desktop      eclipse      Music        Public        Videos  
Documents    eclipse-workspace Pictures      sample.txt  
Downloads    install     pig_1518878805678.log Templates  
[acadgild@localhost ~]$ man ls  
You have new mail in /var/spool/mail/acadgild  
[acadgild@localhost ~]$ mkdir graphics  
[acadgild@localhost ~]$ cd graphics  
[acadgild@localhost graphics]$ cd ..  
[acadgild@localhost ~]$
```

Below screenshot is the output for command " **man ls**", this provides all details of the commands used in Unix. This is just a screen shot of 1st page.



```
acadgild@localhost:~
File Edit View Search Terminal Help
LS(1) User Commands LS(1)

NAME
  ls - list directory contents

SYNOPSIS
  ls [OPTION]... [FILE]...

DESCRIPTION
  List information about the FILES (the current directory by default).
  Sort entries alphabetically if none of -cftuvSUX nor --sort.

  Mandatory arguments to long options are mandatory for short options
  too.

  -a, --all
        do not ignore entries starting with .

  -A, --almost-all
        do not list implied . and ..

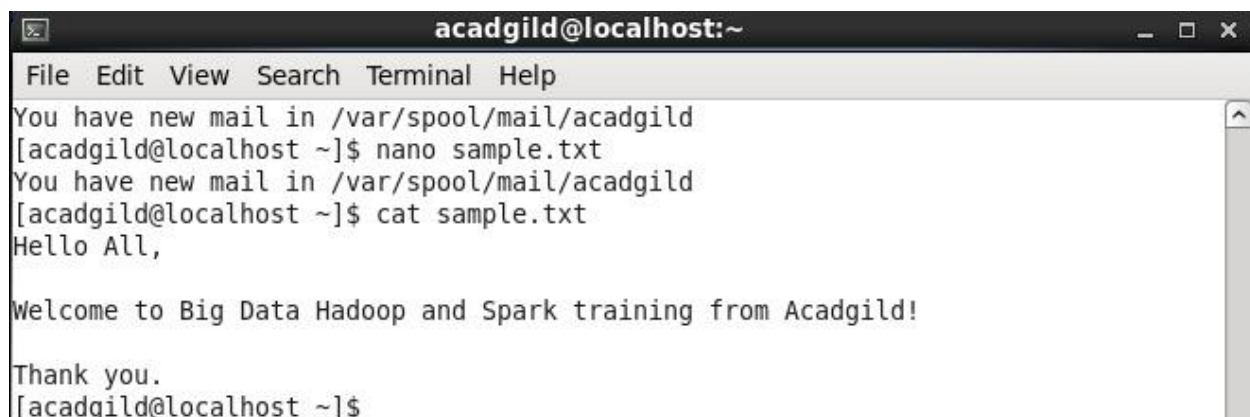
  --author
        with -l, print the author of each file

:
```

Task 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.

Created a text file "**Sample.txt**" in **nano editor** and added the content " *Hello all, welcome to Big Data Hadoop and spark training from acadgild!thank you.*"

Read the content of Sample.txt file using the command "**cat**". Outputs are as shown below.



```
acadgild@localhost:~
File Edit View Search Terminal Help
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ nano sample.txt
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ cat sample.txt
Hello All,

Welcome to Big Data Hadoop and Spark training from Acadgild!

Thank you.
[acadgild@localhost ~]$
```

Task5-Open the hdfs web page by typing localhost:50070 in the browser. Check all the details of the HDFS

Opened hdfs web page- "**localhost:50070**" to check all details of HDFS.

This page provides all details of HDFS like version, started date and time, name node status and so on as shown in the below screen shot.



Overview 'localhost:8020' (active)

Started:	Sun Feb 18 16:34:46 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.

23 files and directories, 1 blocks = 24 total filesystem object(s).

Heap Memory used 33.15 MB of 50 MB Heap Memory. Max Heap Memory is 966.69 MB.

Configured Capacity:	17.11 GB
DFS Used:	48 KB
Non DFS Used:	11.71 GB
DFS Remaining:	5.4 GB
DFS Used%:	0%
DFS Remaining%:	31.58%
Block Pool Used:	48 KB
Block Pool Used%:	0%
DataNodes usages% (Min/Median/Max/stdDev):	0.00% / 0.00% / 0.00% / 0.00%
Live Nodes	1 (Decommissioned: 0)
Dead Nodes	0 (Decommissioned: 0)
Decommissioning Nodes	0
Number of Under-Replicated Blocks	0
Number of Blocks Pending Deletion	0
Block Deletion Start Time	2/18/2018, 4:34:46 PM

NameNode Journal Status

Current transaction ID: 329

Journal Manager	State
FileJournalManager(root=/home/acadgild/install/data/dfs/name)	EditLogFileOutputStream(/home/acadgild/install/data/dfs/name/current/edits_inprogress_0000000000000000329)

NameNode Storage

Storage Directory	Type	State
/home/acadgild/install/data/dfs/name	IMAGE_AND_EDITS	Active

Hadoop, 2016.

Legacy UI