

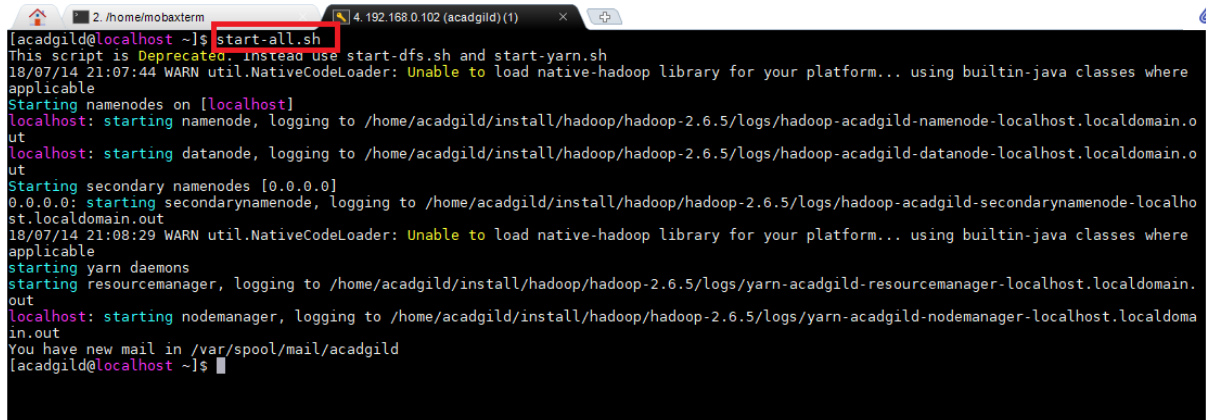
Assignment 1.1  
Session 1 : BIG DATA & HADOOP – INTRODUCTION & CASE STUDIES

Tasks :

**1. start-all.sh**

This command is used to start all Hadoop daemons.

They are Namenode,Datanodes, SecondaryNameNode, ResourceManager and NodeManagers.

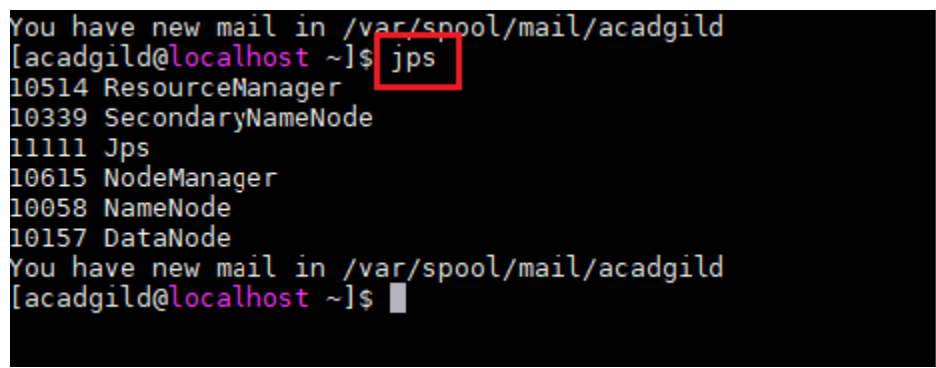
A terminal window showing the execution of the 'start-all.sh' script. The script starts by displaying a warning about deprecated commands and then proceeds to start various Hadoop daemons: namenodes, datanodes, secondarynamenodes, yarn daemons (resource manager and node manager), and finally the namenode again. The terminal output includes log paths and status messages for each daemon.

```
[acadgild@localhost ~]$ start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
18/07/14 21:07:44 WARN util.NativeCodeLoader: Unable to load native-hadoop library 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/hadoop-2.6.5/logs/hadoop-acadgild-secondarynamenode-localhost.localdomain.out
18/07/14 21:08:29 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 ~]$
```

**2. Jps :**

Jps stands for Java Virtual Machine Process Status Tool.

Jps command is used to display all java based processes. It is used to check whether all Hadoop daemons like Namenode,Datanode, etc are running or not.

A terminal window showing the output of the 'jps' command. It lists several Java processes running on the system, including Resource Manager, Secondary Name Node, Jps, Node Manager, Name Node, and Data Node. The terminal also shows a notification about new mail.

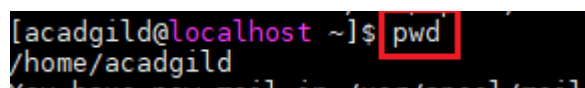
```
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ jps
10514 ResourceManager
10339 SecondaryNameNode
11111 Jps
10615 NodeManager
10058 NameNode
10157 DataNode
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$
```

**3. Unix commands**

**a) Pwd :**

Pwd command stands for Print Working Directory.

It gives full path name of current working directory. Here , '/home/acadgild' is our current path.

A terminal window showing the output of the 'pwd' command, which returns the full path of the current working directory as '/home/acadgild'.

```
[acadgild@localhost ~]$ pwd
/home/acadgild
You have new mail in /var/spool/mail/acadgild
```

## Assignment 1.1

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

#### b) ls -ls :

ls -ls command is used to display detailed information of files or directories like owner, usergroup, size in bytes, File type and Access permissions, Date and Time when it was created. It also gives file size in kilobytes.

Here, for file A.txt, 4 is file size in kB. Owner is Acadgild. Usergroup is Acadgild. Size in Bytes is 12.

Date and time when it was created is 8<sup>th</sup> July 10:37.

For **Documents** directory, File type with Access permission is drwxr-xr-x. Here, 'd' stands for Directory and access permissions like read, write and execute are rwxr-xr-x.

```
[acadgild@localhost ~]$ ls -ls
total 1351916
4 is 4 -rw-rw-r-- 1 acadgild acadgild 12 Jul 8 10:37 A.txt
File 4 -rw-rw-r-- 1 acadgild acadgild 12 Jul 8 10:37 B.txt
Size in 4 drwxr-xr-x 3 acadgild acadgild 4096 Jul 8 10:57 Desktop
kB 4 drwxr-xr-x 2 acadgild acadgild 4096 Feb 2 12:52 Documents
4 drwxr-xr-x 2 acadgild acadgild 4096 Feb 13 14:24 Downloads
4 drwxrwxr-x 3 acadgild acadgild 4096 Dec 29 2017 eclipse
4 drwxrwxr-x 3 acadgild acadgild 4096 Jan 16 14:02 eclipse-workspace
12 -rw-rw-r-- 1 acadgild acadgild 10824 Jul 1 10:01 employee.java
319492 -rw-rw-r-- 1 acadgild acadgild 327155712 Jun 24 16:01 file327.txt
1032140 -rw-rw-r-- 1 acadgild acadgild 1056906185 Jun 24 16:03 file.txt
4 drwxrwxr-x 13 acadgild acadgild 4096 Feb 9 18:06 install
4 drwxr-xr-x 2 acadgild acadgild 4096 Dec 27 2017 Music
8 -rw-rw-r-- 1 acadgild acadgild 4492 Jun 17 11:51 mywordcount.jar
20 -rw-rw-r-- 1 acadgild acadgild 16507 Jul 1 10:45 Person.java
4 drwxr-xr-x 2 acadgild acadgild 4096 Dec 27 2017 Pictures
4 -rw-rw-r-- 1 acadgild acadgild 2615 Jun 2 18:25 pig_1527943593162.log
24 -rw-rw-r-- 1 acadgild acadgild 20968 Jul 7 11:04 pig_1530941132951.log
4 -rw-rw-r-- 1 acadgild acadgild 2247 Jul 8 13:11 pig_1531035041760.log
20 -rw-rw-r-- 1 acadgild acadgild 20230 Jul 8 19:20 pig_1531054879173.log
96 -rw-rw-r-- 1 acadgild acadgild 91246 Jul 8 21:36 pig_1531064088750.log
16 -rw-rw-r-- 1 acadgild acadgild 16160 Jul 14 13:11 pig_1531549232403.log
4 -rw-rw-r-- 1 acadgild acadgild 1288 Jul 14 13:06 pig_1531553768006.log
4 -rw-rw-r-- 1 acadgild acadgild 1288 Jul 14 13:11 pig_1531554102207.log
File type & Access 4 -rw-rw-r-- 1 acadgild acadgild 1288 Jul 14 13:11 pig_1531554109568.log
Permissions 4 -rw-rw-r-- 1 acadgild acadgild 1427 Jul 8 10:37 Pig Commands.txt
4 drwxr-xr-x 2 acadgild acadgild 4096 Dec 27 2017 Public
4 -rw-rw-r-- 1 acadgild acadgild 3677 Jul 7 18:43 Sqoop_Commands_AG.txt
```

## Assignment 1.1

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

#### c) ls -l :

It is similar to ls -ls except it does not display File Size in kilobytes. It displays detailed information of files or directories like owner, user group, size in bytes, File type and Access permissions, Date & Time.

```
[acadgild@localhost ~]$ ls -l
total 1351916
-rw-rw-r--. 1 acadgild acadgild      12 Jul  8 10:37 A.txt
-rw-rw-r--. 1 acadgild acadgild      12 Jul  8 10:37 B.txt
drwxr-xr-x. 3 acadgild acadgild    4096 Jul  8 10:57 Desktop
drwxr-xr-x. 2 acadgild acadgild    4096 Feb  2 12:52 Documents
drwxr-xr-x. 2 acadgild acadgild    4096 Feb 13 14:24 Downloads
drwxrwxr-x. 3 acadgild acadgild    4096 Dec 29  2017 eclipse
drwxrwxr-x. 3 acadgild acadgild    4096 Jan 16 14:02 eclipse-workspace
-rw-rw-r--. 1 acadgild acadgild   10824 Jul  1 10:01 employee.java
-rw-rw-r--. 1 acadgild acadgild  327155712 Jun 24 16:01 file327.txt
-rw-rw-r--. 1 acadgild acadgild 1056906185 Jun 24 16:03 file.txt
drwxrwxr-x. 13 acadgild acadgild    4096 Feb  9 18:06 install
drwxr-xr-x. 2 acadgild acadgild    4096 Dec 27  2017 Music
-rw-rw-r--. 1 acadgild acadgild    4492 Jun 17 11:51 mywordcount.jar
-rw-rw-r--. 1 acadgild acadgild   16507 Jul  1 10:45 Person.java
drwxr-xr-x. 2 acadgild acadgild    4096 Dec 27  2017 Pictures
-rw-rw-r--. 1 acadgild acadgild    2615 Jun  2 18:25 pig_1527943593162.log
-rw-rw-r--. 1 acadgild acadgild   20968 Jul  7 11:04 pig_1530941132951.log
-rw-rw-r--. 1 acadgild acadgild    2247 Jul  8 13:11 pig_1531035041760.log
-rw-rw-r--. 1 acadgild acadgild   20230 Jul  8 19:20 pig_1531054879173.log
-rw-rw-r--. 1 acadgild acadgild   91246 Jul  8 21:36 pig_1531064088750.log
-rw-rw-r--. 1 acadgild acadgild   16160 Jul 14 13:11 pig_1531549232403.log
-rw-rw-r--. 1 acadgild acadgild    1288 Jul 14 13:06 pig_1531553768006.log
-rw-rw-r--. 1 acadgild acadgild    1288 Jul 14 13:11 pig_1531554102207.log
-rw-rw-r--. 1 acadgild acadgild    1288 Jul 14 13:11 pig_1531554109568.log
-rw-rw-r--. 1 acadgild acadgild    1427 Jul  8 10:37 Pig Commands.txt
drwxr-xr-x. 2 acadgild acadgild    4096 Dec 27  2017 Public
-rw-rw-r--. 1 acadgild acadgild    3677 Jul  7 18:43 Sqoop_Commands_AG.txt
drwxr-xr-x. 2 acadgild acadgild    4096 Dec 27  2017 Templates
-rw-rw-r--. 1 acadgild acadgild     36 Jun  9 10:09 test.txt
drwxr-xr-x. 2 acadgild acadgild    4096 Dec 27  2017 Videos
-rw-rw-r--. 1 acadgild acadgild     180 Jul 14 12:32 Wordcount.pig
```

#### 4. Create a file using nano editor and cat command to display content in file :

Nano test.txt creates file 'test.txt' and add content in it and ctrl+X is used to Save and Exit file. Then cat command is used to display content in that file.

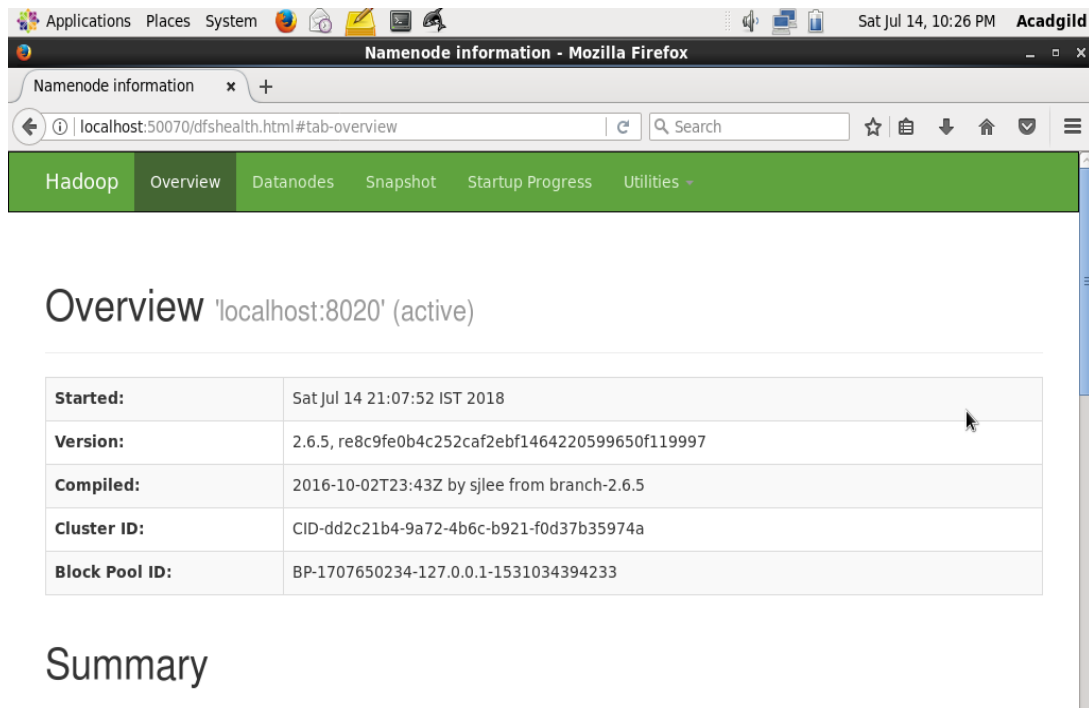
```
[acadgild@localhost ~]$ nano test.txt
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ cat test.txt
Hello Everyone. Welcome to Acadgild.
I am learning Hadoop online course now.
```

## Assignment 1.1

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

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

This page provides health status of Hadoop daemons like Namenode and Datanodes like Percentage of Block Pools and DFS used, Number of under replicated blocks, etc.



The screenshot shows a Mozilla Firefox browser window displaying the 'Namenode information' page. The address bar shows 'localhost:50070/dfshealth.html#tab-overview'. The page has a green navigation bar with tabs: Hadoop, Overview, Datanodes, Snapshot, Startup Progress, and Utilities. The 'Overview' tab is selected, showing the title 'Overview 'localhost:8020' (active)'. Below the title is a table with the following information:

Started:	Sat Jul 14 21:07:52 IST 2018
Version:	2.6.5, re8c9fe0b4c252caf2ebf1464220599650f119997
Compiled:	2016-10-02T23:43Z by sjlee from branch-2.6.5
Cluster ID:	CID-dd2c21b4-9a72-4b6c-b921-f0d37b35974a
Block Pool ID:	BP-1707650234-127.0.0.1-1531034394233

Below the table is a section titled 'Summary'.

## Assignment 1.1

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

<b>Configured Capacity:</b>	17.11 GB
<b>DFS Used:</b>	48.57 MB
<b>Non DFS Used:</b>	14.77 GB
<b>DFS Remaining:</b>	2.29 GB
<b>DFS Used%:</b>	0.28%
<b>DFS Remaining%:</b>	13.4%
<b>Block Pool Used:</b>	48.57 MB
<b>Block Pool Used%:</b>	0.28%
<b>DataNodes usages% (Min/Median/Max/stdDev):</b>	0.28% / 0.28% / 0.28% / 0.00%
<b>Live Nodes</b>	1 (Decommissioned: 0)
<b>Dead Nodes</b>	0 (Decommissioned: 0)
<b>Decommissioning Nodes</b>	0
<b>Number of Under-Replicated Blocks</b>	0
<b>Number of Blocks Pending Deletion</b>	0
<b>Block Deletion Start Time</b>	7/14/2018, 9:07:52 PM

Using Utilities option, we can browse files and directories within directory.

For e.g. we can see all files and directories under Root (/) directory like B.txt, test.txt.

Hadoop Overview Datanodes Snapshot Startup Progress Utilities

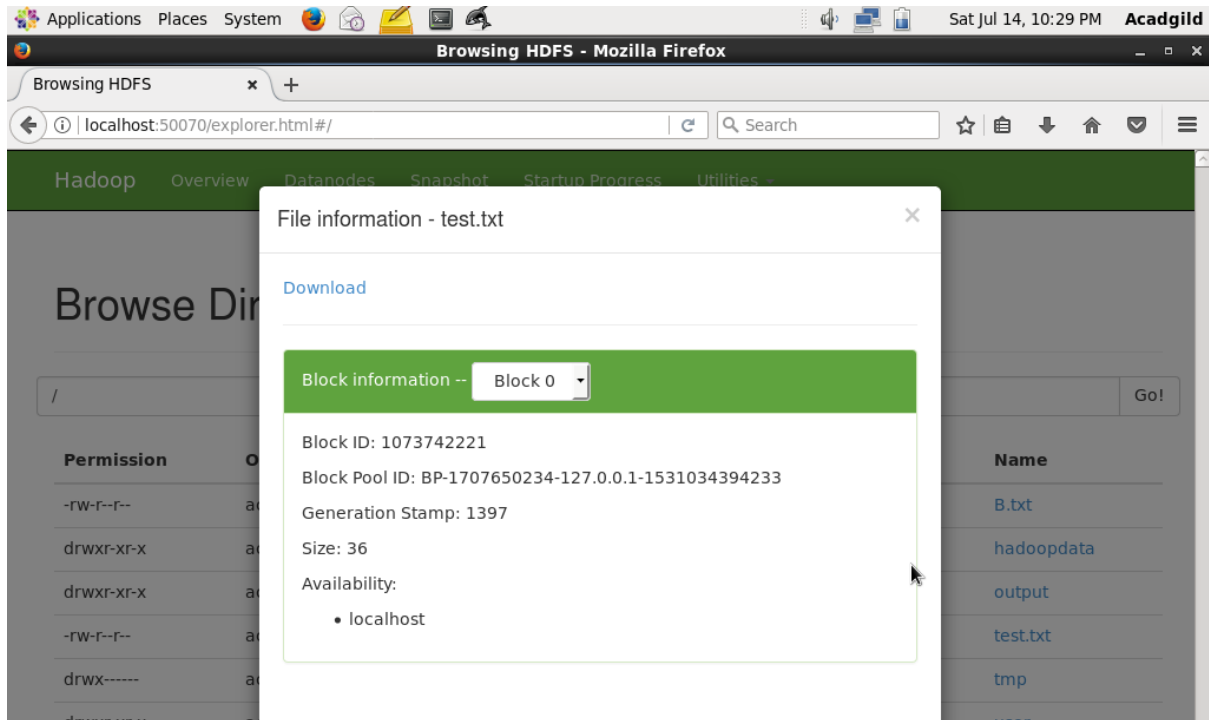
## Browse Directory

/							Go!
Permission	Owner	Group	Size	Replication	Block Size	Name	
-rw-r--r--	acadgild	supergroup	12 B	1	128 MB	B.txt	
drwxr-xr-x	acadgild	supergroup	0 B	0	0 B	hadoopdata	
drwxr-xr-x	acadgild	supergroup	0 B	0	0 B	output	
-rw-r--r--	acadgild	supergroup	36 B	1	128 MB	test.txt	
drwx-----	acadgild	supergroup	0 B	0	0 B	tmp	
drwxr-xr-x	acadgild	supergroup	0 B	0	0 B	user	

## Assignment 1.1

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

Here, test.txt file is 128 MB in size. Hence it shows only 1 block i.e. Block 0 in Block Information.



The screenshot shows the Hadoop Distributed File System (HDFS) browser interface in Mozilla Firefox. The browser address bar shows the URL `localhost:50070/explorer.html#/`. The main content area displays the file 'test.txt' and its block information. The block information panel is open, showing details for Block 0.

**File information - test.txt**

[Download](#)

Block information -- **Block 0**

Block ID: 1073742221  
Block Pool ID: BP-1707650234-127.0.0.1-1531034394233  
Generation Stamp: 1397  
Size: 36  
Availability:

- localhost