

# **BIG DATA ENGINEERING WITH HADOOP & SPARK**

## **ASSIGNMENT 1.1 (BASED ON HDFS)**

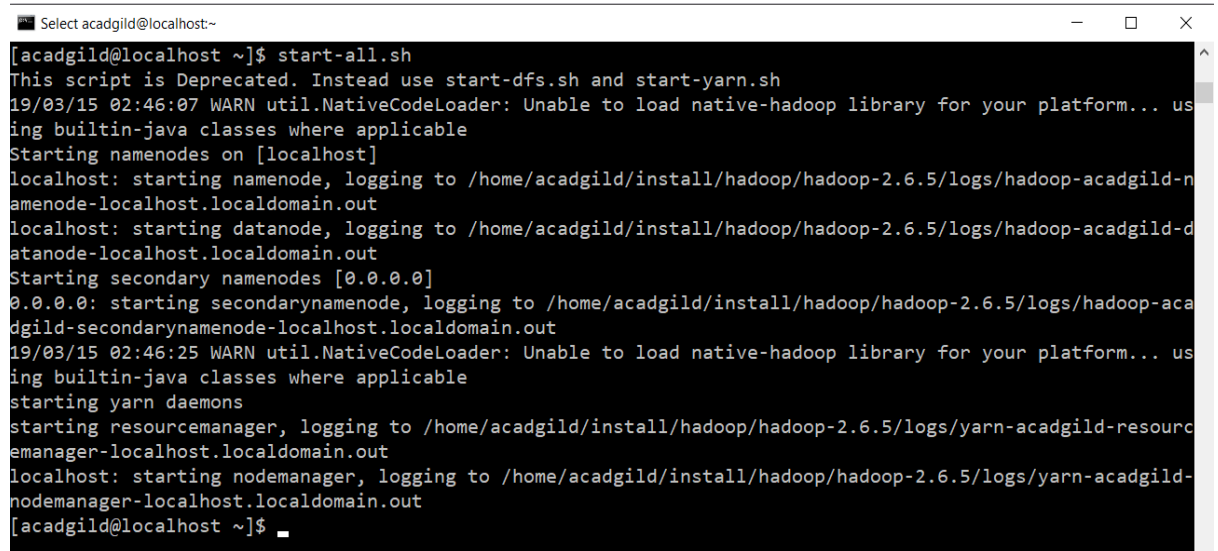
**Submitted By: Anmoldeep Singh**

## TASK 1:

- > Starting Hadoop on Acadgild VM

Command: start-all.sh

Screenshot:

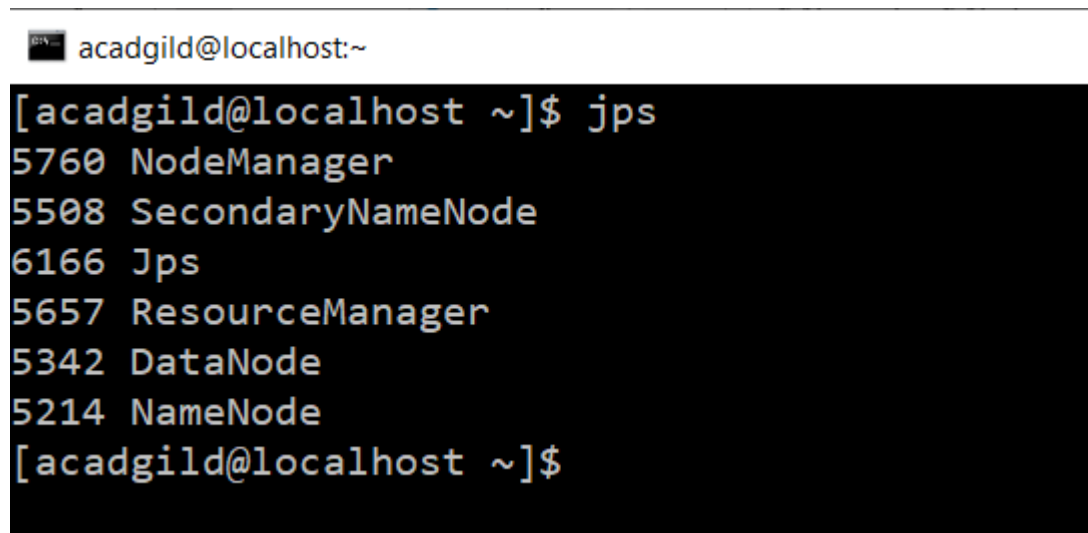
A terminal window titled "Select acadgild@localhost:~" showing the execution of the "start-all.sh" script. The output indicates that the script is deprecated and suggests using "start-dfs.sh" and "start-yarn.sh" instead. It shows the starting of namenodes, datanodes, secondary namenodes, and yarn daemons (resource manager and node manager) on localhost. There are warnings about the native-hadoop library not being loaded, but the script continues using builtin-java classes. The terminal ends with a prompt for the user.

```
[acadgild@localhost ~]$ start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
19/03/15 02:46:07 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... us
ing 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-n
amenode-localhost.localdomain.out
localhost: starting datanode, logging to /home/acadgild/install/hadoop/hadoop-2.6.5/logs/hadoop-acadgild-d
atanode-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-aca
dgild-secondarynamenode-localhost.localdomain.out
19/03/15 02:46:25 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... us
ing builtin-java classes where applicable
starting yarn daemons
starting resourcemanager, logging to /home/acadgild/install/hadoop/hadoop-2.6.5/logs/yarn-acadgild-resourc
emanager-localhost.localdomain.out
localhost: starting nodemanager, logging to /home/acadgild/install/hadoop/hadoop-2.6.5/logs/yarn-acadgild-
nodemanager-localhost.localdomain.out
[acadgild@localhost ~]$
```

- > Run a JPS Command

Command: jps

Screenshot:

A terminal window titled "acadgild@localhost:~" showing the output of the "jps" command. The output lists several Java processes running on the system: NodeManager (PID 5760), SecondaryNameNode (PID 5508), Jps (PID 6166), ResourceManager (PID 5657), DataNode (PID 5342), and NameNode (PID 5214). The terminal ends with a prompt for the user.

```
[acadgild@localhost ~]$ jps
5760 NodeManager
5508 SecondaryNameNode
6166 Jps
5657 ResourceManager
5342 DataNode
5214 NameNode
[acadgild@localhost ~]$
```

- > Run a few Unix commands

Command: pwd (for printing the Present Working Directory)

Screenshot:

```
acadgild@localhost:~  
[acadgild@localhost ~]$ pwd  
/home/acadgild  
[acadgild@localhost ~]$
```

Command: ls (for listing the contents of a directory)

Screenshot:

```
acadgild@localhost:~  
[acadgild@localhost ~]$ ls  
; eclipse MapR.jar.desktop pigudf.txt  
a.txt eclipse-workspace Music Public  
b.txt employee.csv myoutput Templates  
commandline file1.txt Pictures Videos  
Desktop file.txt pig_1551604163230.log wget-log  
Documents google-chrome-stable_current_x86_64.rpm pig_1551611887288.log  
Downloads install pig_udf-0.0.1-SNAPSHOT.jar  
[acadgild@localhost ~]$
```

Command: mkdir (for making a new directory)

Screenshot:

```
acadgild@localhost:~  
[acadgild@localhost ~]$ mkdir helloworld  
[acadgild@localhost ~]$ ls  
; eclipse install pig_udf-0.0.1-SNAPSHOT.jar  
a.txt eclipse-workspace MapR.jar.desktop pigudf.txt  
b.txt employee.csv Music Public  
commandline file1.txt myoutput Templates  
Desktop file.txt Pictures Videos  
Documents google-chrome-stable_current_x86_64.rpm pig_1551604163230.log wget-log  
Downloads helloworld pig_1551611887288.log  
[acadgild@localhost ~]$
```

Command: cp (for copying files from one location to another)

Screenshot:

```
acadgild@localhost:~  
[acadgild@localhost ~]$ ls  
; eclipse install pig_udf-0.0.1-SNAPSHOT.jar  
a.txt eclipse-workspace MapR.jar.desktop pigudf.txt  
b.txt employee.csv Music Public  
commandline file1.txt myoutput Templates  
Desktop file.txt Pictures Videos  
Documents google-chrome-stable_current_x86_64.rpm pig_1551604163230.log wget-log  
Downloads helloworld pig_1551611887288.log  
[acadgild@localhost ~]$ cp a.txt Documents  
[acadgild@localhost ~]$ ls Documents  
a.txt  
[acadgild@localhost ~]$
```

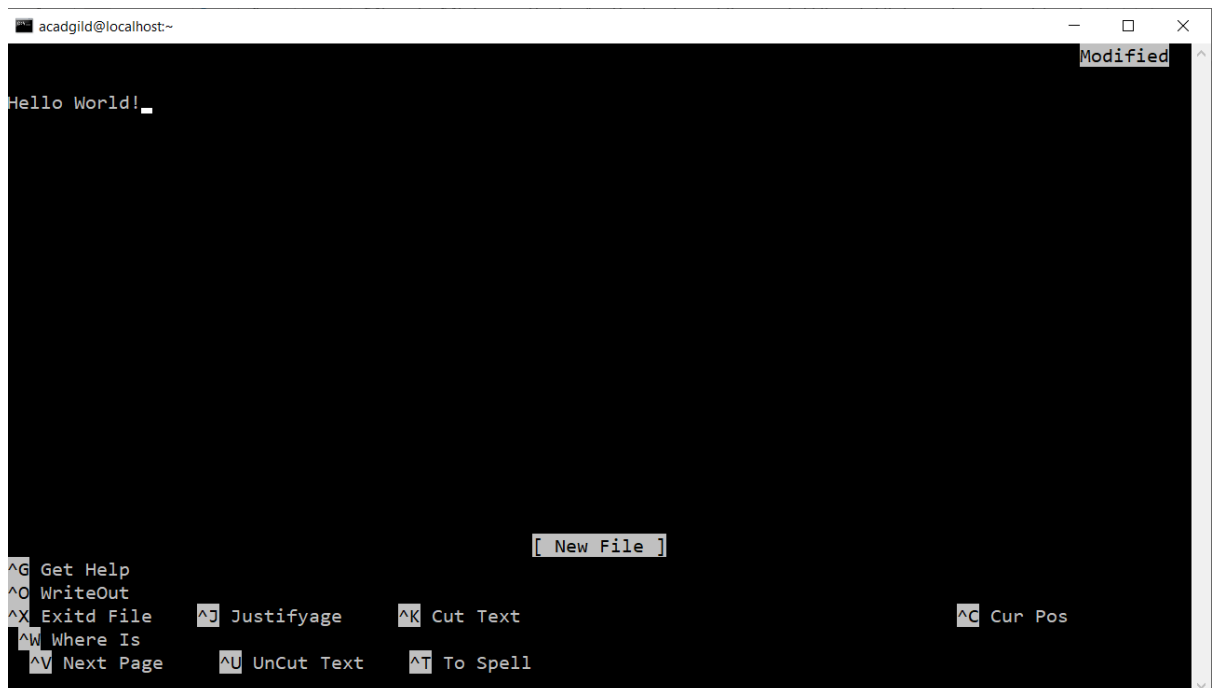
- > Create a file from terminal using Nano editor, add content and Cat it to see the saved contents

Command: nano (for opening nano editor)

Command: cat (for seeing the contents of the saved file)

Screenshots:

```
acadgild@localhost:~  
[acadgild@localhost ~]$ nano test.txt
```

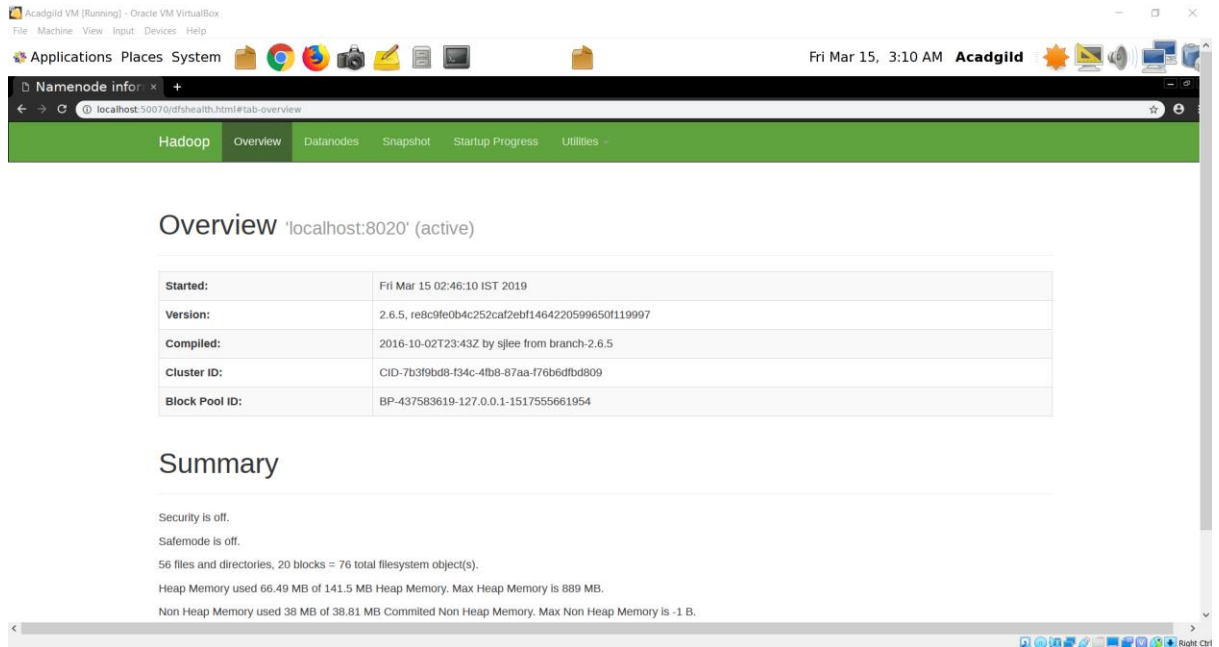


```
acadgild@localhost:~  
[acadgild@localhost ~]$ nano test.txt  
[acadgild@localhost ~]$ cat test.txt  
Hello World!  
[acadgild@localhost ~]$
```

> Open the HDFS webpage

URL: <http://localhost:50070>

Screenshot:

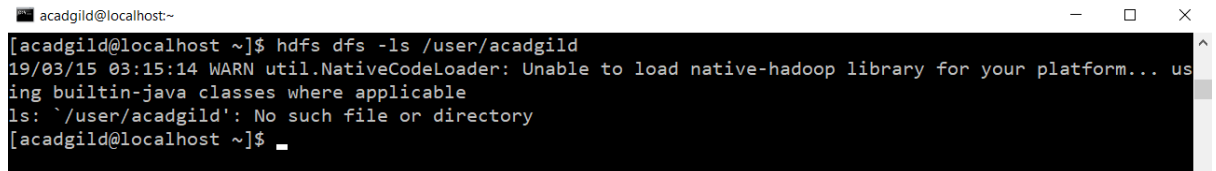


## TASK 2:

- > Check whether /user/acadgild directory exists or not in the HDFS

Command: `hdfs dfs -ls`

Screenshot:

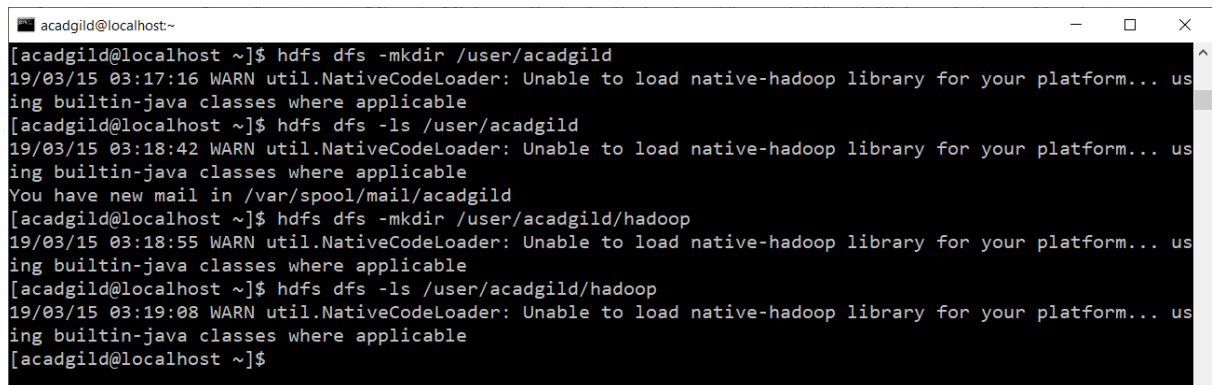


```
acadgild@localhost:~  
[acadgild@localhost ~]$ hdfs dfs -ls /user/acadgild  
19/03/15 03:15:14 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable  
ls: '/user/acadgild': No such file or directory  
[acadgild@localhost ~]$
```

- > Create directory /user/acadgild/hadoop

Command: `hdfs dfs -mkdir`

Screenshot:



```
acadgild@localhost:~  
[acadgild@localhost ~]$ hdfs dfs -mkdir /user/acadgild  
19/03/15 03:17:16 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable  
[acadgild@localhost ~]$ hdfs dfs -ls /user/acadgild  
19/03/15 03:18:42 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable  
You have new mail in /var/spool/mail/acadgild  
[acadgild@localhost ~]$ hdfs dfs -mkdir /user/acadgild/hadoop  
19/03/15 03:18:55 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable  
[acadgild@localhost ~]$ hdfs dfs -ls /user/acadgild/hadoop  
19/03/15 03:19:08 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable  
[acadgild@localhost ~]$
```

### TASK 3:

- > Create a file in HDFS under directory /user/acadgild/hadoop, with name word-count.txt

Commands: vi (used for creating files using built-in terminal editor)

hdfs dfs -put (used for copying file from local filesystem to HDFS)

Screenshot:

```
acadgild@localhost:~  
[acadgild@localhost ~]$ vi word-count.txt_
```

```
acadgild@localhost:~  
Now in her village there was a medicine man, a doctor, who lived deep in the mountains. When anyone in the  
village was sick or hurting, a visit to him would do the trick. Most of the time, Yun Ok felt proud that  
she could fix her own problems.  
~  
~  
~  
-- INSERT -- 1,244 All
```

```
acadgild@localhost:~  
[acadgild@localhost ~]$ hdfs dfs -put word-count.txt /user/acadgild/hadoop  
19/03/15 03:35:19 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... us  
ing builtin-java classes where applicable  
[acadgild@localhost ~]$ hdfs dfs -ls /user/acadgild/hadoop  
19/03/15 03:35:37 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... us  
ing builtin-java classes where applicable  
Found 1 items  
-rw-r--r-- 1 acadgild supergroup 244 2019-03-15 03:35 /user/acadgild/hadoop/word-count.txt  
You have new mail in /var/spool/mail/acadgild  
[acadgild@localhost ~]$
```

## TASK 4:

- > Create a file max-temp.txt in local FS with 10-15 records in the following format –

dd-mm-yyyy,temperature

Also, move the file to /user/acadgild/Hadoop in HDFS

## Commands: vi

## hdfs dfs -put

Screenshot:

```
acadgild@localhost:~  
[acadgild@localhost ~]$ vi max-temp.txt
```

```

acadgild@localhost:~
10-01-1990,10
10-01-1991,20
10-01-1992,22
10-01-1993,10
10-01-1994,09
10-01-1995,21
10-01-1996,18
10-01-1997,14
10-01-1998,19
10-01-1999,17
10-01-2000,14
10-01-2001,12_
-- INSERT --
12,14
All

```

```
[acadgild@localhost ~]$ hdfs dfs -put max-temp.txt /user/acadgild/hadoop
19/03/15 03:52:21 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hdfs dfs -ls /user/acadgild/hadoop
19/03/15 03:52:39 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Found 2 items
-rw-r--r--  1 acadgild supergroup      168 2019-03-15 03:52 /user/acadgild/hadoop/max-temp.txt
-rw-r--r--  1 acadgild supergroup      244 2019-03-15 03:35 /user/acadgild/hadoop/word-count.txt
[acadgild@localhost ~]$
```

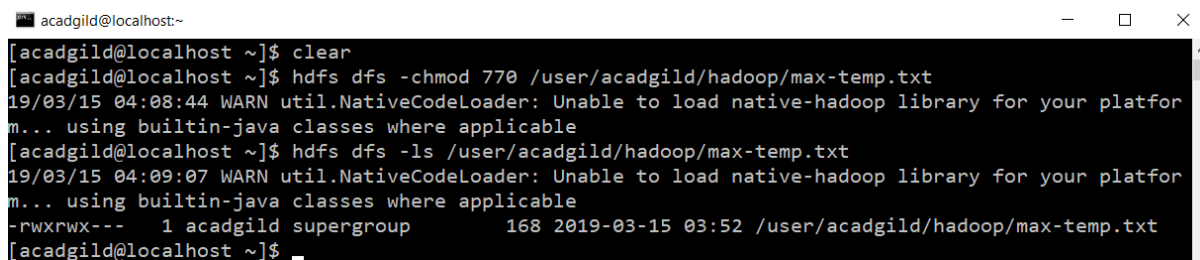


## TASK 5:

- > Change the permission of the file /user/acadgild/hadoop/max-temp.txt, such that only the owner and the group members have full control over the file. Others do not have any control over it.

Command: chmod (used for changing the mode of a file/folder)

Screenshot:



```
acadgild@localhost:~  
[acadgild@localhost ~]$ clear  
[acadgild@localhost ~]$ hdfs dfs -chmod 770 /user/acadgild/hadoop/max-temp.txt  
19/03/15 04:08:44 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platfor  
m... using builtin-java classes where applicable  
[acadgild@localhost ~]$ hdfs dfs -ls /user/acadgild/hadoop/max-temp.txt  
19/03/15 04:09:07 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platfor  
m... using builtin-java classes where applicable  
-rwxrwx--- 1 acadgild supergroup      168 2019-03-15 03:52 /user/acadgild/hadoop/max-temp.txt  
[acadgild@localhost ~]$
```

-----O-----