

BIG DATA
DEVELOPMENT

ACADGILD

Session 2: HDFS -INTRODUCTION

Assignment 1(Solution)

Big Data Hadoop and Spark Development

Assignment 1 – You have to perform the given tasks.

1. Table of Contents

1. Introduction	3
2. Objective	3
3. Prerequisites	3
4. Associated Data Files	3
5. Problem Statement	3
6. Expected Output	4
7. Approximate Time to Complete Task	4

Big Data Hadoop and Spark Development

2. Introduction

In this assignment, you need to perform the given tasks.

3. Objective

This assignment will help you consolidate the concepts learnt in Session 2.

4. Prerequisites:

None

5. Associated Data Files

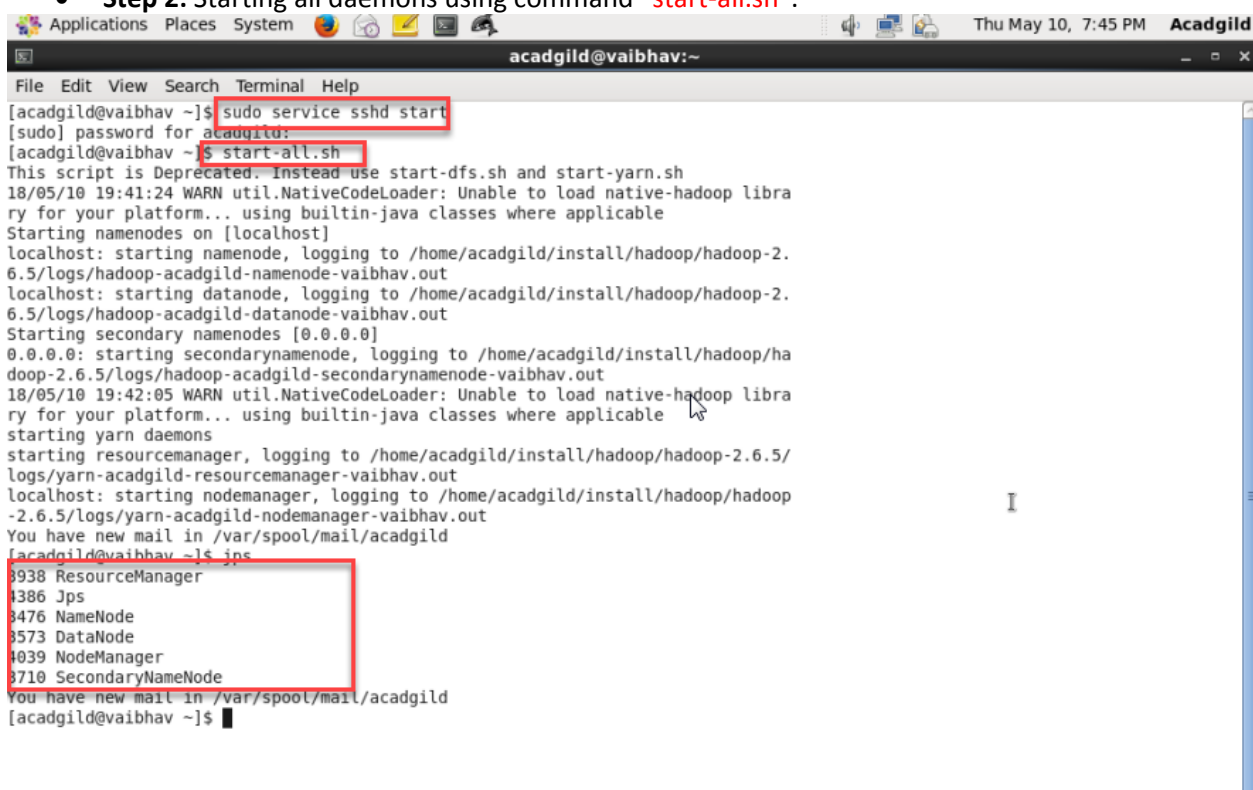
None

6. Problem Statement

Task 1: Check whether /user/acadgild directory exists or not in the HDFS.

SOLUTION:

- **Step 1:** Start the service using command “`sudo service sshd start`” in the command line interface of the linux VDI system.
- **Step 2:** Starting all daemons using command “`start-all.sh`”.



The screenshot shows a terminal window titled 'acadgild@vaibhav:~'. The user has executed the command 'sudo service sshd start' and 'start-all.sh'. The output of 'start-all.sh' shows the following:

```
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
18/05/10 19:41:24 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-vaibhav.out
localhost: starting datanode, logging to /home/acadgild/install/hadoop/hadoop-2.
6.5/logs/hadoop-acadgild-datanode-vaibhav.out
Starting secondary namenodes [0.0.0.0]
0.0.0.0: starting secondarynamenode, logging to /home/acadgild/install/hadoop/ha
doop-2.6.5/logs/hadoop-acadgild-secondarynamenode-vaibhav.out
18/05/10 19:42:05 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-vaibhav.out
localhost: starting nodemanager, logging to /home/acadgild/install/hadoop/hadoop
-2.6.5/logs/yarn-acadgild-nodemanager-vaibhav.out
You have new mail in /var/spool/mail/acadgild
[acadgild@vaibhav ~]$ jps
3938 ResourceManager
1386 Jps
3476 NameNode
3573 DataNode
4039 NodeManager
3710 SecondaryNameNode
You have new mail in /var/spool/mail/acadgild
[acadgild@vaibhav ~]$
```

- **Step 3:** To check whether the hadoop system is having the directory: /user/acadgild. Command “`hadoop hdfs -ls /user/acadgild`” was used.

- It shows the hdfs command returns a result that there is no such directory.

```

Applications Places System Thu May 10, 7:50 PM Acadgild
acadgild@vaibhav:~
File Edit View Search Terminal Help
[acadgild@vaibhav ~]$ hadoop dfs -ls /user/acadgild/
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

18/05/10 19:49:14 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
ls: '/user/acadgild/': No such file or directory
[acadgild@vaibhav ~]$

```

If it doesn't exist, then create this.

- **Step 1** : Creating a directory named hadoop in /USER/acadgild path using command : **"hadoop dfs -mkdir /user/acadgild hadoop"**

```

Applications Places System Thu May 10, 7:50 PM Acadgild
acadgild@vaibhav:~
File Edit View Search Terminal Help
You have new mail in /var/spool/mail/acadgild
[acadgild@vaibhav ~]$ hadoop dfs -mkdir /user/acadgild/ hadoop
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

18/05/10 20:23:27 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using
asses where applicable
[acadgild@vaibhav ~]$ hadoop dfs -ls /
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

18/05/10 20:24:11 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using
asses where applicable
Found 7 items
drwxr-xr-x - acadgild supergroup 0 2018-05-10 20:01 /USER
drwxr-xr-x - acadgild supergroup 0 2018-05-06 11:04 /data1
drwxr-xr-x - acadgild supergroup 0 2018-05-10 20:15 /hadoop
drwxr-xr-x - acadgild supergroup 0 2018-05-06 11:06 /hadoopdata
drwxr-xr-x - acadgild supergroup 0 2018-02-02 12:49 /sqoopout111
drwxrwx--- - acadgild supergroup 0 2018-02-09 11:35 /tmp
drwxr-xr-x - acadgild supergroup 0 2018-05-10 20:23 /user
You have new mail in /var/spool/mail/acadgild
[acadgild@vaibhav ~]$ hadoop dfs -ls /user/acadgild
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

18/05/10 20:24:26 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using
asses where applicable
Found 1 items
drwxr-xr-x - acadgild supergroup 0 2018-05-10 20:23 /user/acadgild/hadoop
[acadgild@vaibhav ~]$

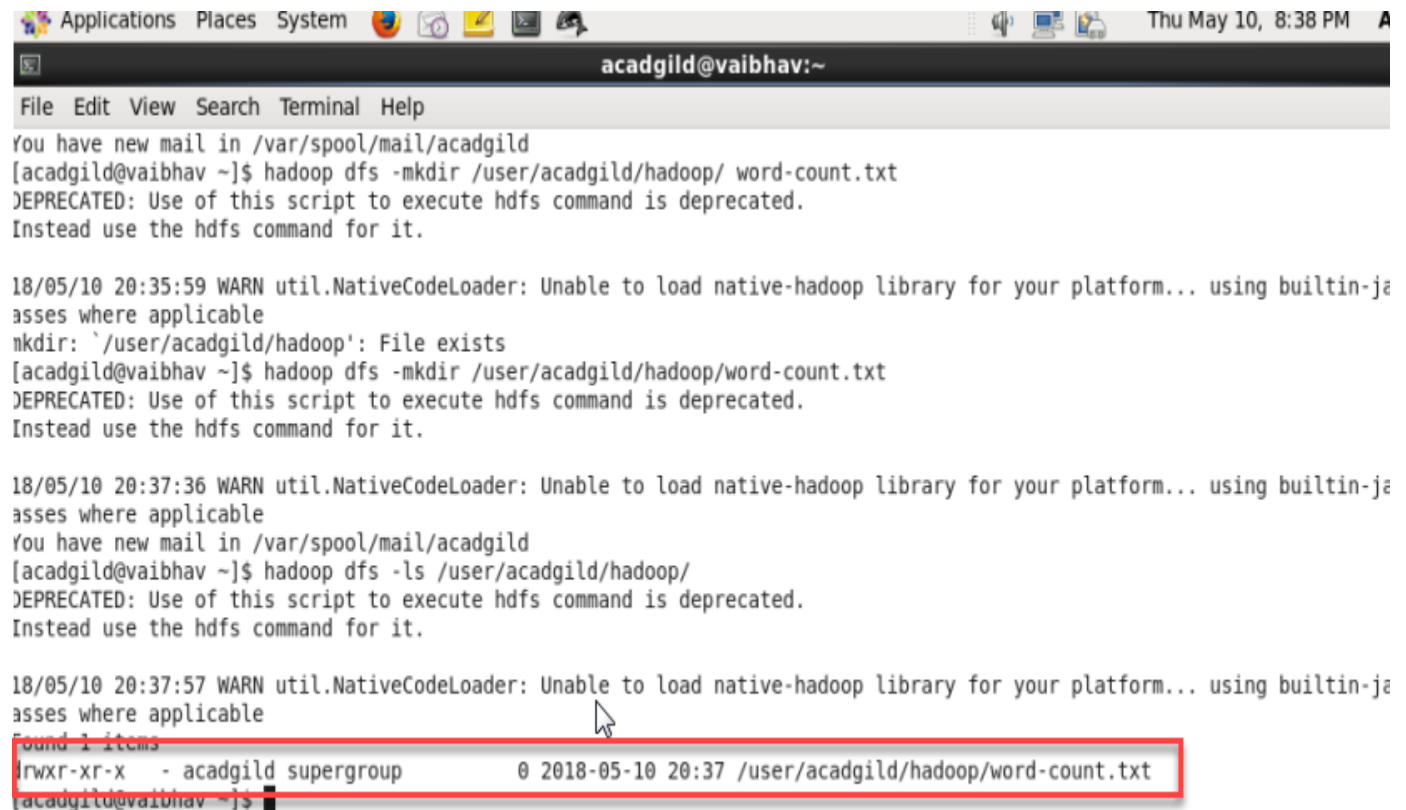
```

Task 2:

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

SOULTION:

Step 1 : The file was created using the command : `"hadoop dfs -mkdir /user/acadgild/hadoop/wordcount.txt"`



The screenshot shows a terminal window titled 'acadgild@vaibhav:~'. The user runs the command `hadoop dfs -mkdir /user/acadgild/hadoop/word-count.txt`. The output shows a warning about the deprecated script and a message that the directory already exists. The user then runs `hadoop dfs -mkdir /user/acadgild/hadoop/word-count.txt` again, which also shows a warning. Finally, the user runs `hadoop dfs -ls /user/acadgild/hadoop/`, and the output shows a file named 'word-count.txt' with permissions 'rw-r--r--' and owner 'acadgild supergroup'. The file is highlighted with a red box.

```
File Edit View Search Terminal Help
You have new mail in /var/spool/mail/acadgild
[acadgild@vaibhav ~]$ hadoop dfs -mkdir /user/acadgild/hadoop/ word-count.txt
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

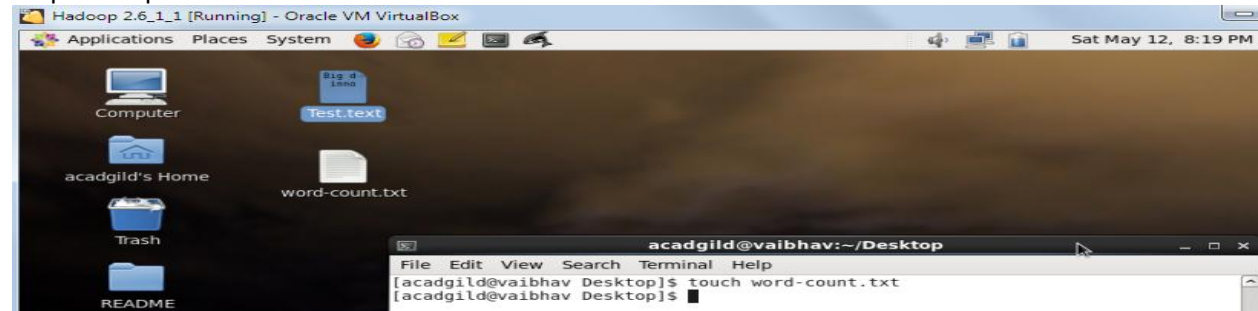
18/05/10 20:35:59 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-jas
asses where applicable
mkdir: '/user/acadgild/hadoop': File exists
[acadgild@vaibhav ~]$ hadoop dfs -mkdir /user/acadgild/hadoop/word-count.txt
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

18/05/10 20:37:36 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-jas
asses where applicable
You have new mail in /var/spool/mail/acadgild
[acadgild@vaibhav ~]$ hadoop dfs -ls /user/acadgild/hadoop/
DEPRECATED: Use of this script to execute hdfs command is deprecated.
Instead use the hdfs command for it.

18/05/10 20:37:57 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-jas
asses where applicable
Found 1 items
-rw-r--r-- - acadgild supergroup          0 2018-05-10 20:37 /user/acadgild/hadoop/word-count.txt
[acadgild@vaibhav ~]$
```

2. Whatever we type on screen should get appended to the file.
Try to type (on screen) few lines from any online article or textbook.

Step 1 : To perform this task one file with the same name was created in the local FS :



Step 2 : Some content was added into the file using echo command.

```
acadgild@vaibhav:~/Desktop
File Edit View Search Terminal Help
[acadgild@vaibhav Desktop]$ touch word-count.txt
[acadgild@vaibhav Desktop]$ echo Hadoop is opensource java framework used for bi
g data processing and storage. > word-count.txt
You have new mail in /var/spool/mail/acadgild
[acadgild@vaibhav Desktop]$ cat word-count.txt
Hadoop is opensource java framework used for big data processing and storage.
[acadgild@vaibhav Desktop]$
```

Step3: To Append this content into the file present in the HDFS path The -appendToFile command was used.

```
File Edit View Search Terminal Help
[acadgild@vaibhav Desktop]$ hdfs dfs -appendToFile word-count.txt /user/acadgild/hadoop/word-count.txt
18/05/12 23:06:53 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-jav
asses where applicable
[acadgild@vaibhav Desktop]$ hdfs -dfs -cat /user/acadgild/hadoop/word-count.txt
Unrecognized option: -dfs
Error: Could not create the Java Virtual Machine.
Error: A fatal exception has occurred. Program will exit.
You have new mail in /var/spool/mail/acadgild
[acadgild@vaibhav Desktop]$ hdfs dfs -cat /user/acadgild/hadoop/word-count.txt
18/05/12 23:08:11 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-jav
asses where applicable
Hadoop is opensource java framework used for big data processing and storage.
Hadoop is opensource java framework used for big data processing and storage.
There are following components in the hadoop ecosystem: 1.HDFS 2. MapReduce 3.YARN
[acadgild@vaibhav Desktop]$
```

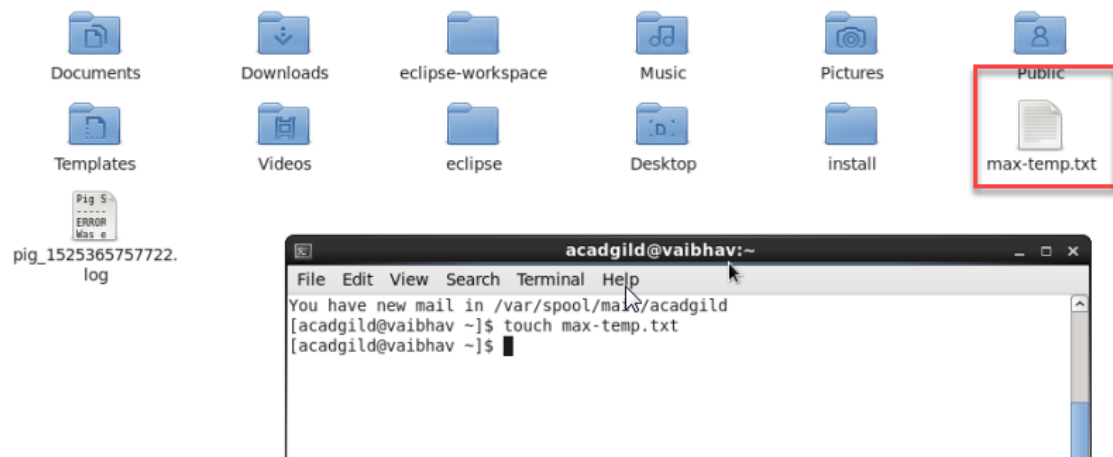
And content was appended with existing content in the file.

Task 3:

SOLUTION:

Create a file max-temp.txt in local FS.

Step 1 : Creating a file in local system :



```
[acadgild@vaibhav ~]$ touch max-temp.txt
[acadgild@vaibhav ~]$ echo 10-01-1990,10
10-01-1990,10
You have new mail in /var/spool/mail/acadgild
[acadgild@vaibhav ~]$ echo 10-01-1990 > max-tmp.txt
You have new mail in /var/spool/mail/acadgild
[acadgild@vaibhav ~]$ echo 10-01-1990 >> max-temp.txt
[acadgild@vaibhav ~]$ echo 10-01-1990,10 >> max-temp.txt
[acadgild@vaibhav ~]$ echo 10-01-1990,20 >> max-temp.txt
[acadgild@vaibhav ~]$ echo 10-01-1990,30 >> max-temp.txt
[acadgild@vaibhav ~]$ echo 10-01-1990,40 >> max-temp.txt
[acadgild@vaibhav ~]$ echo 10-01-1990,50 >> max-temp.txt
You have new mail in /var/spool/mail/acadgild
[acadgild@vaibhav ~]$ echo 10-01-1990,60 >> max-temp.txt
[acadgild@vaibhav ~]$ cat max-temp.txt
10-01-1990
10-01-1990,10
10-01-1990,20
10-01-1990,30
10-01-1990,40
10-01-1990,50
10-01-1990,60
```

Put some 10-15 records of date and temperature example:

dd-mm-yyyy,temperature

Example:

10-01-1990,10

10-02-1991,20

Big Data Hadoop and Spark Development

Move this file to HDFS at /user/acadgild/hadoop.

Task 4:

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.

7. Expected Output

Solution report with commands, explanation to commands and screenshot for output.

Report should be in PDF format and submitted in GitHub.

8. Approximate Time to Complete Task

30 mins