

Assignment 2 : Introduction to HDFS Assignment Problems

Problem Statement

Initial Terminal Execution :

```
[acadgild@localhost ~]$ jps
6973 Jps
[acadgild@localhost ~]$ sudo service sshd start
[sudo] password for acadgild:
[acadgild@localhost ~]$ start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
18/08/02 05:13:01 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.localdo
main.out
localhost: starting datanode, logging to
/home/acadgild/install/hadoop/hadoop-2.6.5/logs/hadoop-acadgild-datanode-localhost.localdo
main.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-localho
st.localdomain.out
18/08/02 05:13:38 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.loc
aldomain.out
localhost: starting nodemanager, logging to
/home/acadgild/install/hadoop/hadoop-2.6.5/logs/yarn-acadgild-nodemanager-localhost.localdo
main.out
[acadgild@localhost ~]$ jps
7121 NameNode
7555 ResourceManager
7380 SecondaryNameNode
7220 DataNode
7657 NodeManager
7980 Jps
[acadgild@localhost ~]$
```

Task 1:

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

If it doesn't exist, then create this.

Create a directory /user/acadgild/hadoop.

Terminal Execution :

```
[acadgild@localhost ~]$ hdfs dfs -ls /
18/08/02 05:23:58 WARN util.NativeCodeLoader: Unable to load native-hadoop library for
your platform... using builtin-java classes where applicable
Found 8 items
drwxr-xr-x - acadgild supergroup      0 2018-07-04 22:39 /SQOOPOUT
drwxr-xr-x - acadgild supergroup      0 2018-07-04 23:13 /SQOOPOUT1
drwxr-xr-x - acadgild supergroup      0 2018-07-26 01:49 /hadoopdata
drwxr-xr-x - acadgild supergroup      0 2018-07-24 21:34 /hbase
drwxr-xr-x - acadgild supergroup      0 2018-07-16 07:56 /home
drwxr-xr-x - acadgild supergroup      0 2018-07-04 09:16 /sqoopout
drwx-wx-wx - acadgild supergroup      0 2018-07-11 00:19 /tmp
drwxr-xr-x - acadgild supergroup      0 2018-07-15 22:56 /user
```

```
[acadgild@localhost ~]$ hdfs dfs -ls /user
18/08/02 05:24:38 WARN util.NativeCodeLoader: Unable to load native-hadoop library for
your platform... using builtin-java classes where applicable
Found 2 items
drwxr-xr-x - acadgild supergroup      0 2018-06-24 01:29 /user/acadgild
drwxr-xr-x - acadgild supergroup      0 2018-07-15 22:56 /user/hive
```

Thus , we can check and verify from the above execution that directory '/user/acadgild' exists in HDFS.

Now we will create a directory '/user/acadgild/hadoop'.

Terminal Execution :

```
[acadgild@localhost ~]$ hdfs dfs -mkdir /user/acadgild/hadoop
18/08/02 05:30:53 WARN util.NativeCodeLoader: Unable to load native-hadoop library for
your platform... using builtin-java classes where applicable
```

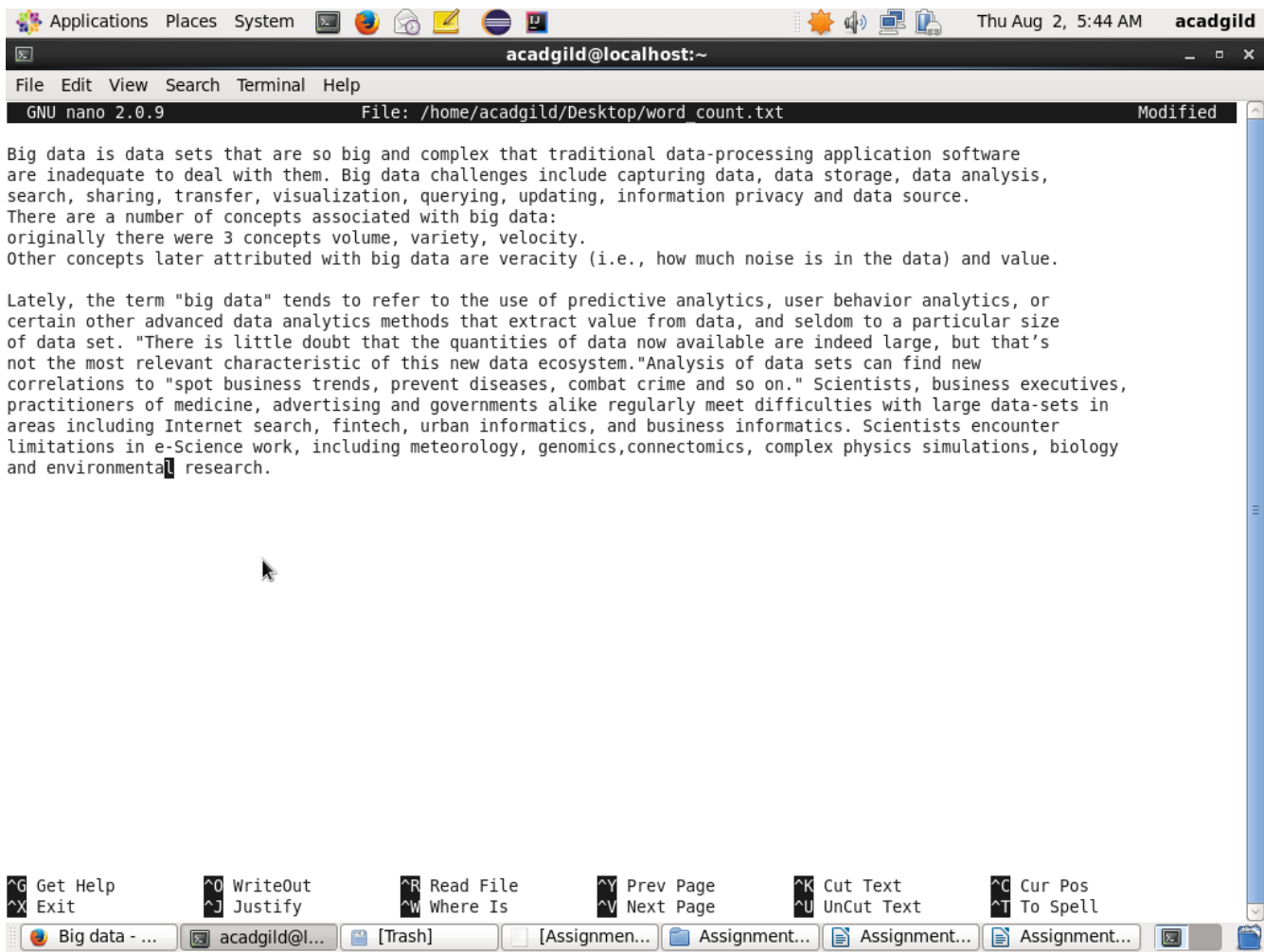
```
[acadgild@localhost ~]$ hdfs dfs -ls /user/acadgild
18/08/02 05:31:21 WARN util.NativeCodeLoader: Unable to load native-hadoop library for
your platform... using builtin-java classes where applicable
Found 2 items
drwxr-xr-x - acadgild supergroup      0 2018-08-02 05:30 /user/acadgild/hadoop
drwxr-xr-x - acadgild supergroup      0 2018-07-06 16:53 /user/acadgild/hadoopdata
```

Task 2:

Create a file in HDFS under directory /user/acadgild/hadoop, with name word-count.txt. Whatever we type on screen should get appended to the file.
Try to type (on screen) few lines from any online article or textbook.

Terminal Execution :

```
[acadgild@localhost ~]$ nano /home/acadgild/Desktop/word_count.txt
```



```
Applications Places System [Icons] [System Tray] Thu Aug 2, 5:44 AM acadgild
acadgild@localhost:~
File Edit View Search Terminal Help
GNU nano 2.0.9 File: /home/acadgild/Desktop/word_count.txt Modified

Big data is data sets that are so big and complex that traditional data-processing application software
are inadequate to deal with them. Big data challenges include capturing data, data storage, data analysis,
search, sharing, transfer, visualization, querying, updating, information privacy and data source.
There are a number of concepts associated with big data:
originally there were 3 concepts volume, variety, velocity.
Other concepts later attributed with big data are veracity (i.e., how much noise is in the data) and value.

Lately, the term "big data" tends to refer to the use of predictive analytics, user behavior analytics, or
certain other advanced data analytics methods that extract value from data, and seldom to a particular size
of data set. "There is little doubt that the quantities of data now available are indeed large, but that's
not the most relevant characteristic of this new data ecosystem." Analysis of data sets can find new
correlations to "spot business trends, prevent diseases, combat crime and so on." Scientists, business executives,
practitioners of medicine, advertising and governments alike regularly meet difficulties with large data-sets in
areas including Internet search, fintech, urban informatics, and business informatics. Scientists encounter
limitations in e-Science work, including meteorology, genomics, connectomics, complex physics simulations, biology
and environmental research.
```

```
[acadgild@localhost ~]$ ls /home/acadgild/Desktop/
```

```
acadgild.conf      Assignment2.png    flume_original_conf~ problem1.pig
query1.pig~ README
apache-flume-1.6.0-bin Assignment Done    hive_local        problem1.pig~
query2.pig~ sample.txt~
apache-flume-1.6.0-bin.tar.gz Assignment_Jars    hive-site.xml     problem2.pig~
query3.pig~ twitter_jars.png
Assignment_12.11199.odt Datasets          logs_hive         problem3.pig~ query4.pig~
word_count.pig~
Assignment_2.11199.odt flume_original_conf PIG               problem4.pig~
query5.pig~ word_count.txt
```

```
[acadgild@localhost ~]$ cat /home/acadgild/Desktop/word_count.txt
```

Big data is data sets that are so big and complex that traditional data-processing application software are inadequate to deal with them. Big data challenges include capturing data, data storage, data analysis, search, sharing, transfer, visualization, querying, updating, information privacy and data source. There are a number of concepts associated with big data: originally there were 3 concepts volume, variety, velocity. Other concepts later attributed with big data are veracity (i.e., how much noise is in the data) and value.

Lately, the term "big data" tends to refer to the use of predictive analytics, user behavior analytics, or certain other advanced data analytics methods that extract value from data, and seldom to a particular size of data set. "There is little doubt that the quantities of data now available are indeed large, but that's not the most relevant characteristic of this new data ecosystem." Analysis of data sets can find new correlations to "spot business trends, prevent diseases, combat crime and so on." Scientists, business executives, practitioners of medicine, advertising and governments alike regularly meet difficulties with large data-sets in areas including Internet search, fintech, urban informatics, and business informatics. Scientists encounter limitations in e-Science work, including meteorology, genomics, connectomics, complex physics simulations, biology and environmental research.

```
[acadgild@localhost ~]$ hdfs dfs -ls /user/acadgild/hadoop/
```

```
18/08/02 05:48:16 WARN util.NativeCodeLoader: Unable to load native-hadoop library for
```

your platform... using builtin-java classes where applicable

```
[acadgild@localhost ~]$ hdfs dfs -put /home/acadgild/Desktop/word_count.txt  
/user/acadgild/hadoop/
```

18/08/02 05:49:55 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/
```

18/08/02 05:51:36 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

Found 1 items

```
-rw-r--r--  1 acadgild supergroup    1446 2018-08-02 05:49  
/user/acadgild/hadoop/word_count.txt
```

```
[acadgild@localhost ~]$ hdfs dfs -cat /user/acadgild/hadoop/word_count.txt
```

18/08/02 05:52:00 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable

Big data is data sets that are so big and complex that traditional data-processing application software

are inadequate to deal with them. Big data challenges include capturing data, data storage, data analysis,

search, sharing, transfer, visualization, querying, updating, information privacy and data source.

There are a number of concepts associated with big data:

originally there were 3 concepts volume, variety, velocity.

Other concepts later attributed with big data are veracity (i.e., how much noise is in the data) and value.

Lately, the term "big data" tends to refer to the use of predictive analytics, user behavior analytics, or

certain other advanced data analytics methods that extract value from data, and seldom to a particular size

of data set. "There is little doubt that the quantities of data now available are indeed large, but that's

not the most relevant characteristic of this new data ecosystem." Analysis of data sets can find new

correlations to "spot business trends, prevent diseases, combat crime and so on." Scientists, business executives,

practitioners of medicine, advertising and governments alike regularly meet difficulties with large data-sets in

areas including Internet search, fintech, urban informatics, and business informatics. Scientists encounter

limitations in e-Science work, including meteorology, genomics, connectomics, complex physics simulations, biology and environmental research.

Task 3:

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

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

dd-mm-yyyy,temperature

Example:

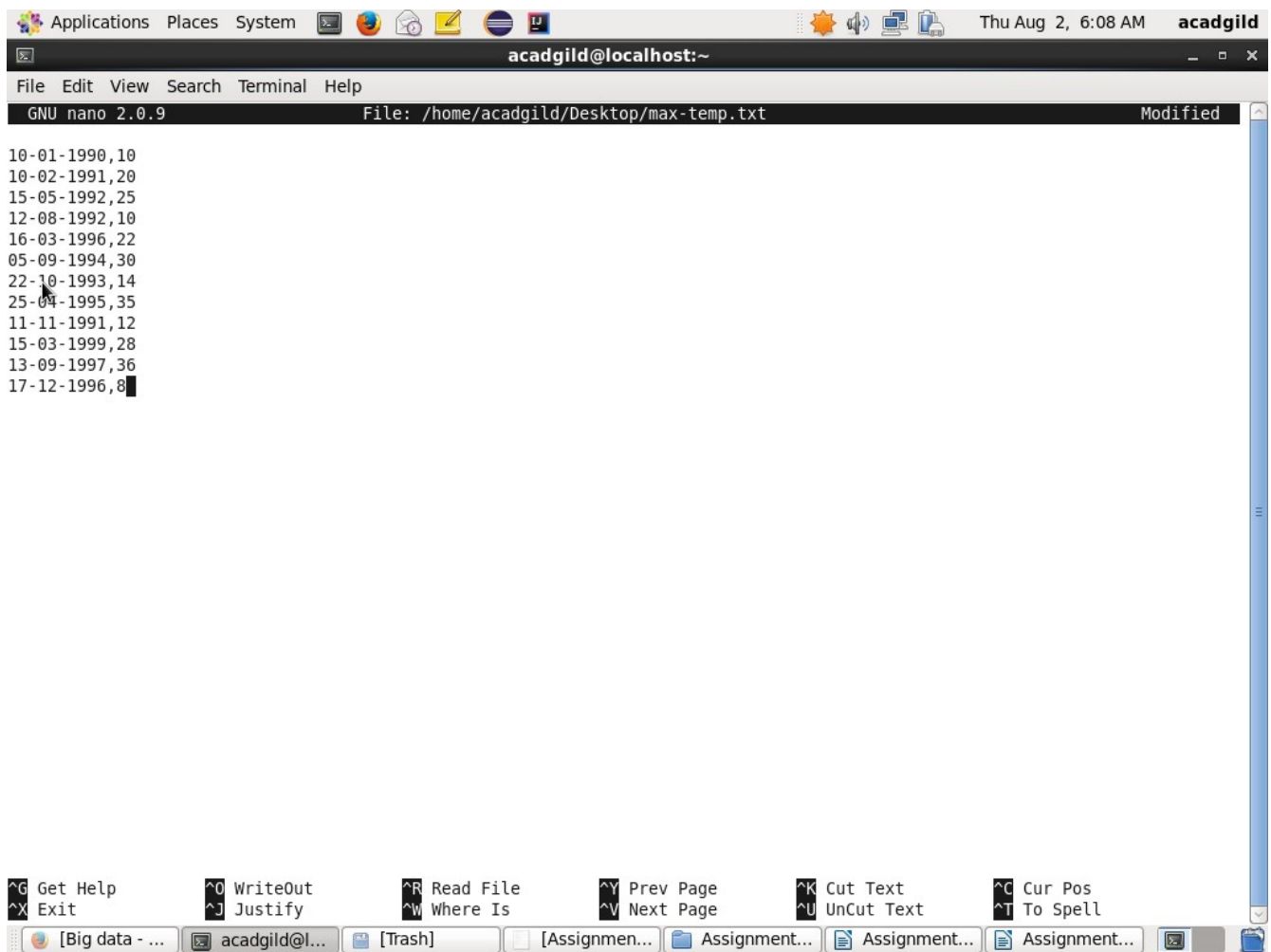
10-01-1990,10

10-02-1991,20

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

Terminal Execution :

```
[acadgild@localhost ~]$ ls /home/acadgild/Desktop/
acadgild.conf      Assignment2.png    flume_original_conf~  problem1.pig
query1.pig~  README
apache-flume-1.6.0-bin  Assignment Done    hive_local            problem1.pig~
query2.pig~  sample.txt~
apache-flume-1.6.0-bin.tar.gz  Assignment_Jars    hive-site.xml         problem2.pig~
query3.pig~  twitter_jars.png
Assignment_12.11199.odt  Datasets          logs_hive             problem3.pig~  query4.pig~
word_count.pig~
Assignment_2.11199.odt    flume_original_conf  PIG                  problem4.pig~
query5.pig~  word_count.txt
[acadgild@localhost ~]$ nano /home/acadgild/Desktop/max-temp.txt
```



```
[acadgild@localhost ~]$ ls /home/acadgild/Desktop/
acadgild.conf      Assignment2.png    hive_local  problem1.pig~  query3.pig~
word_count.pig~
apache-flume-1.6.0-bin  Assignment Done    hive-site.xml  problem2.pig~  query4.pig~
word_count.txt
apache-flume-1.6.0-bin.tar.gz  Assignment_Jars    logs_hive    problem3.pig~  query5.pig~
Assignment_12.11199.odt      Datasets          max-temp.txt  problem4.pig~  README
Assignment_2.11199.odt      flume_original_conf  PIG          query1.pig~  sample.txt~
Assignment2_1.png          flume_original_conf~  problem1.pig  query2.pig~
twitter_jars.png
```

```
[acadgild@localhost ~]$ cat /home/acadgild/Desktop/max-temp.txt
10-01-1990,10
10-02-1991,20
15-05-1992,25
12-08-1992,10
16-03-1996,22
```

05-09-1994,30
22-10-1993,14
25-04-1995,35
11-11-1991,12
15-03-1999,28
13-09-1997,36
17-12-1996,8

```
[acadgild@localhost ~]$ hdfs dfs -ls /user/acadgild/hadoop
18/08/02 06:13:34 WARN util.NativeCodeLoader: Unable to load native-hadoop library for
your platform... using builtin-java classes where applicable
Found 1 items
-rw-r--r--  1 acadgild supergroup    1446 2018-08-02 05:49
/user/acadgild/hadoop/word_count.txt
```

```
[acadgild@localhost ~]$ hdfs dfs -put /home/acadgild/Desktop/max-temp.txt
/user/acadgild/hadoop
18/08/02 06:17:59 WARN util.NativeCodeLoader: Unable to load native-hadoop library for
your platform... using builtin-java classes where applicable
[acadgild@localhost ~]$ hdfs dfs -ls /home/acadgild/Desktop/max-temp.txt
/user/acadgild/hadoop
18/08/02 06:18:13 WARN util.NativeCodeLoader: Unable to load native-hadoop library for
your platform... using builtin-java classes where applicable
ls: `/home/acadgild/Desktop/max-temp.txt': No such file or directory
Found 2 items
-rw-r--r--  1 acadgild supergroup    170 2018-08-02 06:18
/user/acadgild/hadoop/max-temp.txt
-rw-r--r--  1 acadgild supergroup    1446 2018-08-02 05:49
/user/acadgild/hadoop/word_count.txt
```

```
[acadgild@localhost ~]$ hdfs dfs -cat /user/acadgild/hadoop/max-temp.txt
18/08/02 06:22:02 WARN util.NativeCodeLoader: Unable to load native-hadoop library for
your platform... using builtin-java classes where applicable
10-01-1990,10
10-02-1991,20
15-05-1992,25
12-08-1992,10
16-03-1996,22
05-09-1994,30
22-10-1993,14
25-04-1995,35
11-11-1991,12
15-03-1999,28
13-09-1997,36
```


17-12-1996,8

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.

Terminal Execution :

```
[acadgild@localhost ~]$ hdfs dfs -ls /user/acadgild/hadoop/
18/08/02 06:27:56 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      170 2018-08-02 06:18
/user/acadgild/hadoop/max-temp.txt
-rw-r--r--  1 acadgild supergroup    1446 2018-08-02 05:49
/user/acadgild/hadoop/word_count.txt
```

Here, file '**max-temp.txt**' has got **-rw-r--r--** permissions in HDFS that shows that the given file has read and write for the owner and only read permission for the owner's group and All other users.

Now, to change the permission to Read, Write and Execute for owner and its group and to restrict all other from Read, Write and Execute we will following command on the Terminal :

```
[acadgild@localhost ~]$ hdfs dfs -chmod 770 /user/acadgild/hadoop/max-temp.txt
18/08/02 06:49:03 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/
18/08/02 06:49:15 WARN util.NativeCodeLoader: Unable to load native-hadoop library for
your platform... using builtin-java classes where applicable
Found 2 items
-rwxrwx---  1 acadgild supergroup      170 2018-08-02 06:18
/user/acadgild/hadoop/max-temp.txt
-rw-r--r--  1 acadgild supergroup    1446 2018-08-02 05:49
/user/acadgild/hadoop/word_count.txt
[acadgild@localhost ~]$
```