BIG DATA HADOOP AND SPARK DEVELOPMENT

ASSIGNMENT – 2

Table of Contents:

1. Introduction

In this assignment, the given tasks are performed and Output of the tasks performed and Screenshots are attached.

2. Objective

This assignment consolidates the deeper understanding of the Session – 2 Introduction to HDFS, Which Consists of HDFS Commands, HDFS Permissions and HDFS Storage.

3. Problem Statement

- Task 1:

  Check whether /user/acadgild directory exits or not in HDFS.

  If it doesn’t exist, then create this.

  Create a directory /user/acadgild/hadoop

- Task 2:

  Create 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.

- 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.

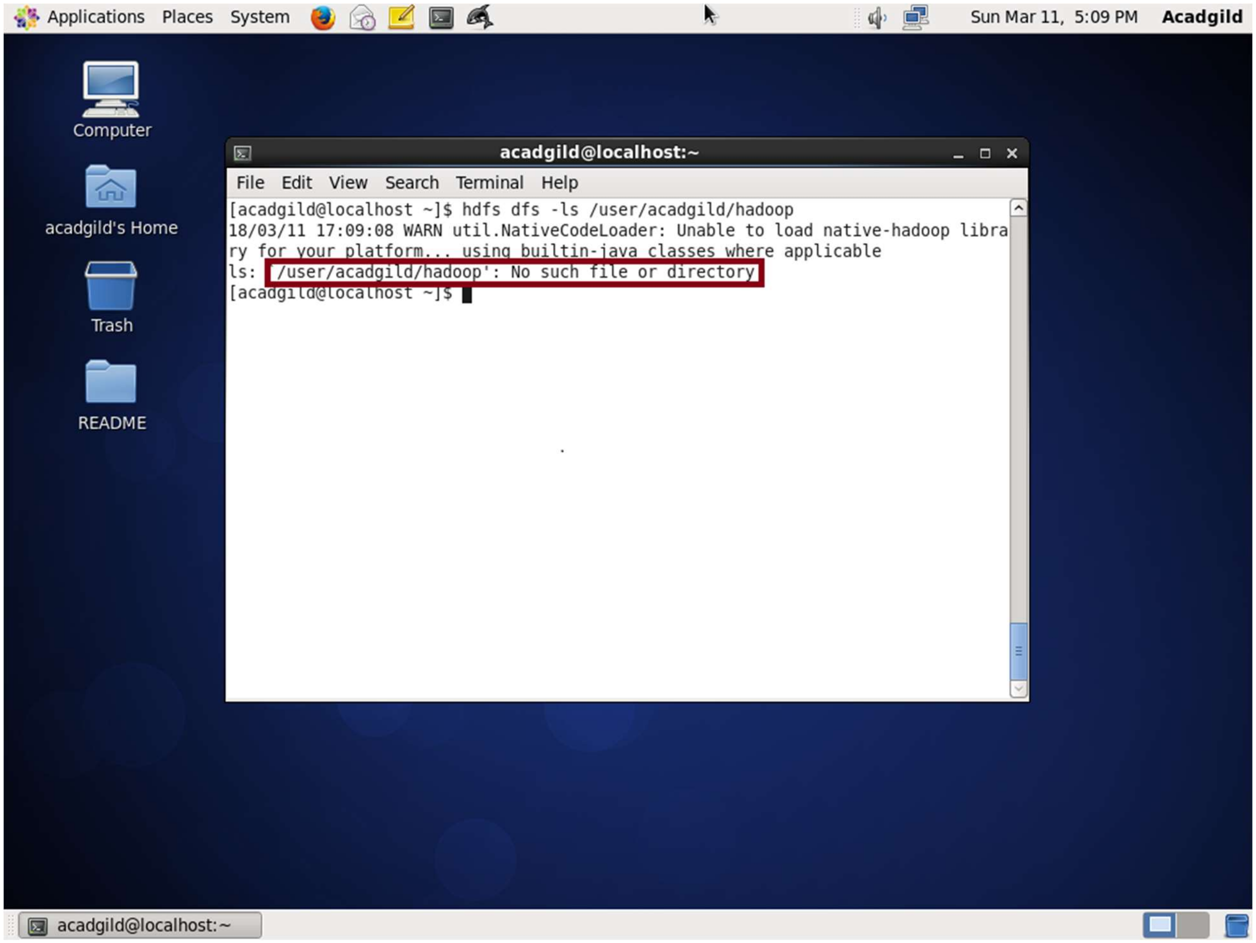
- Task 4:

  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.

4. Expected Output

- Task 1:

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



Applications Places System
Sun Mar 11, 5:09 PM Acadgild
Computer
acadgild's Home
Trash
README
acadgild@localhost:~
File Edit View Search Terminal Help
acadgild@localhost:~

```
[acadgild@localhost ~]$ hdfs dfs -ls /user/acadgild/hadoop
18/03/11 17:09:08 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
ls: `/user/acadgild/hadoop': No such file or directory
[acadgild@localhost ~]$
```

There is no such directory. So Creating parent directory along the path.

mkdir

Usage: **hadoop fs -mkdir [-p] <paths>**

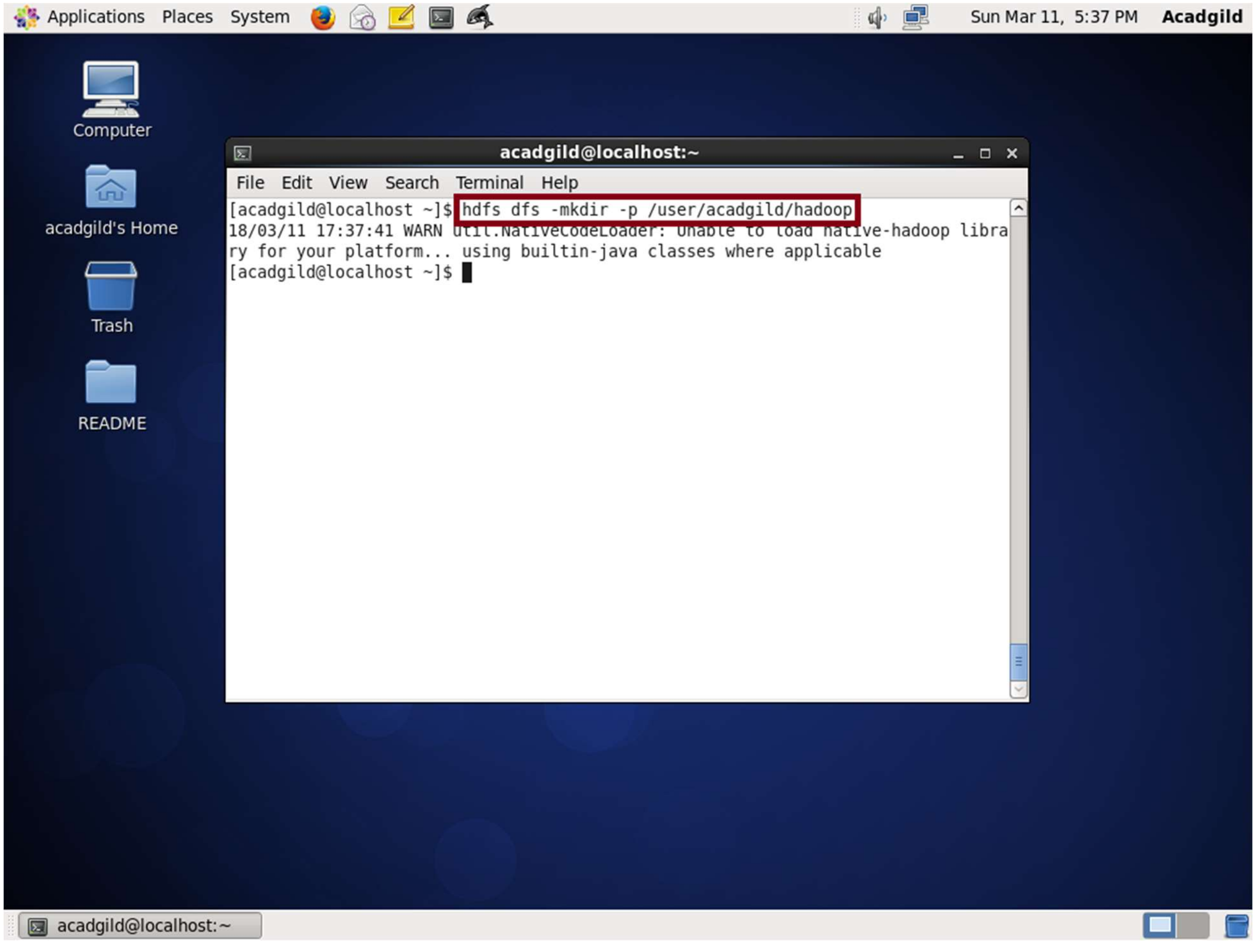
hdfs dfs –mkdir [-p] <paths>

Takes path uri's as argument and creates directories.

Options:

- The -p option behavior is much like Unix mkdir -p, creating parent directories along the path.

Here I have used **hdfs dfs -mkdir -p /user/acadgild/hadoop** to create the parent directory along the path.



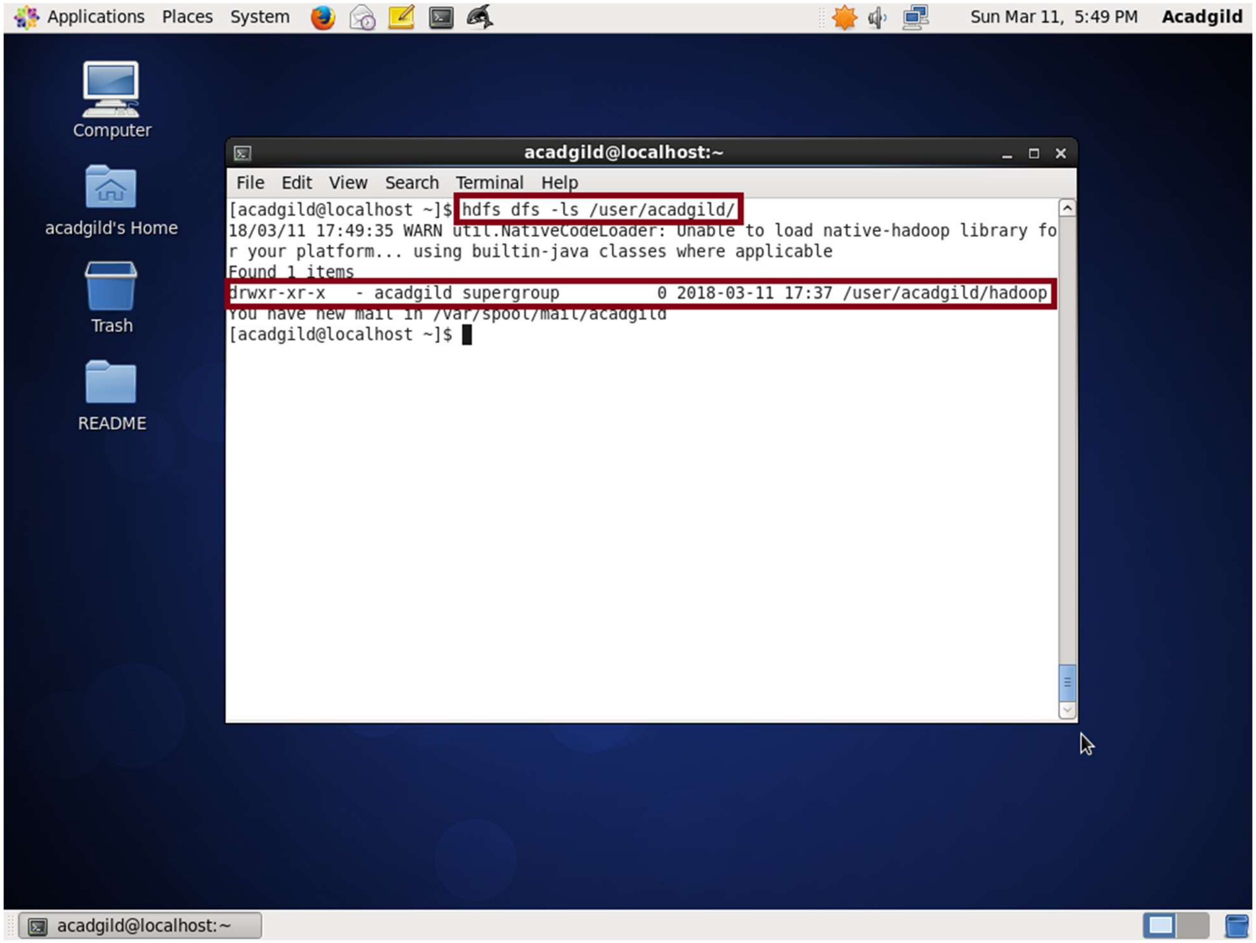
```
[acadgild@localhost ~]$ hdfs dfs -mkdir -p /user/acadgild/hadoop
18/03/11 17:37:41 WARN util.NativeCodeLoader: Unable to load native-hadoop libra
ry for your platform... using builtin-java classes where applicable
[acadgild@localhost ~]$ 
```

Here I have used **hdfs dfs -ls /user/acadgild/**

-ls

List the content of the directory.

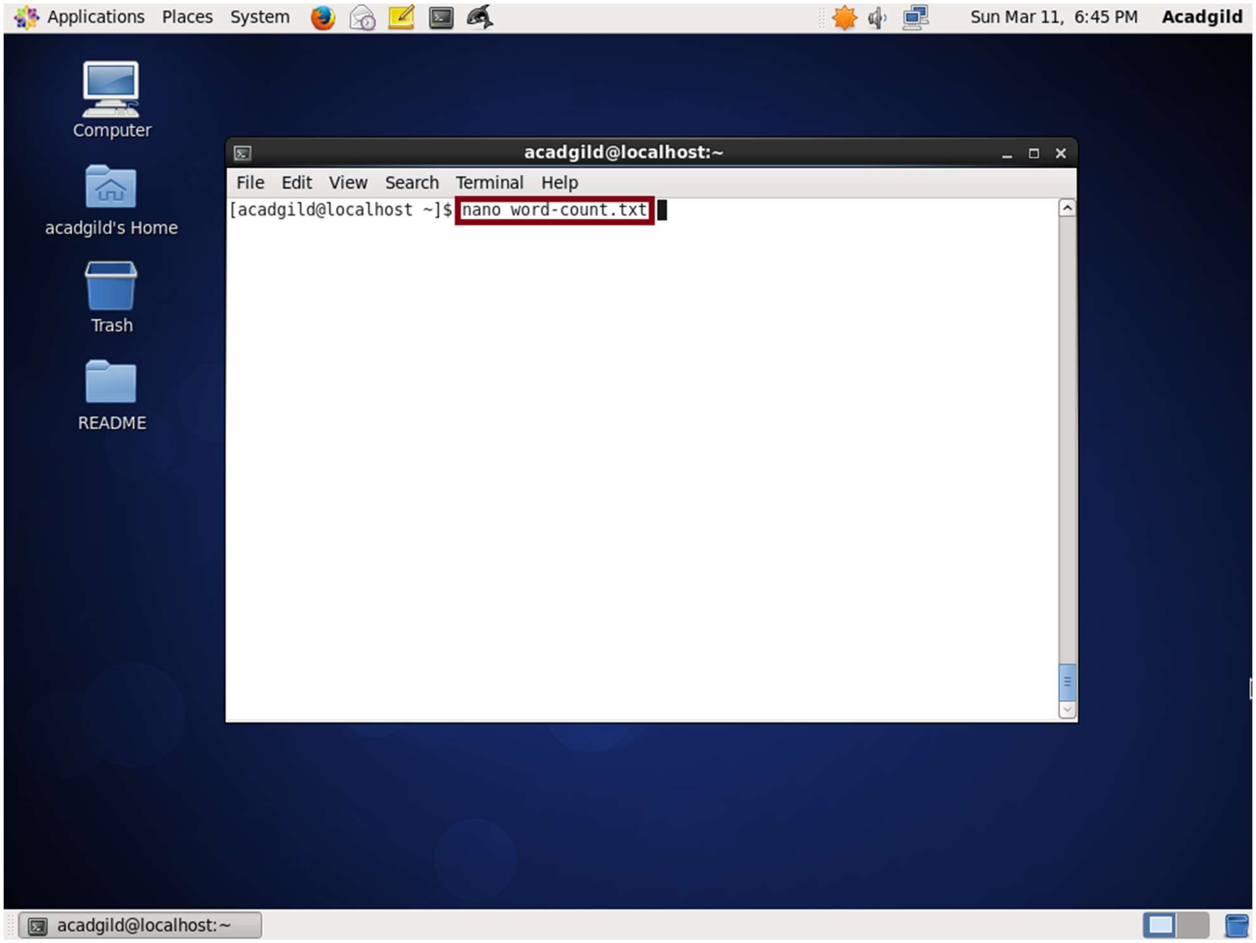
Here **/user/acadgild/** contains the hadoop directory.



```
[acadgild@localhost ~]$ hdfs dfs -ls /user/acadgild/
18/03/11 17:49:35 WARN util.NativeCodeLoader: Unable to load native-hadoop library fo
r your platform... using builtin-java classes where applicable
Found 1 items
drwxr-xr-x   - acadgild supergroup          0 2018-03-11 17:37 /user/acadgild/hadoop
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ 
```

- Task 2:

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

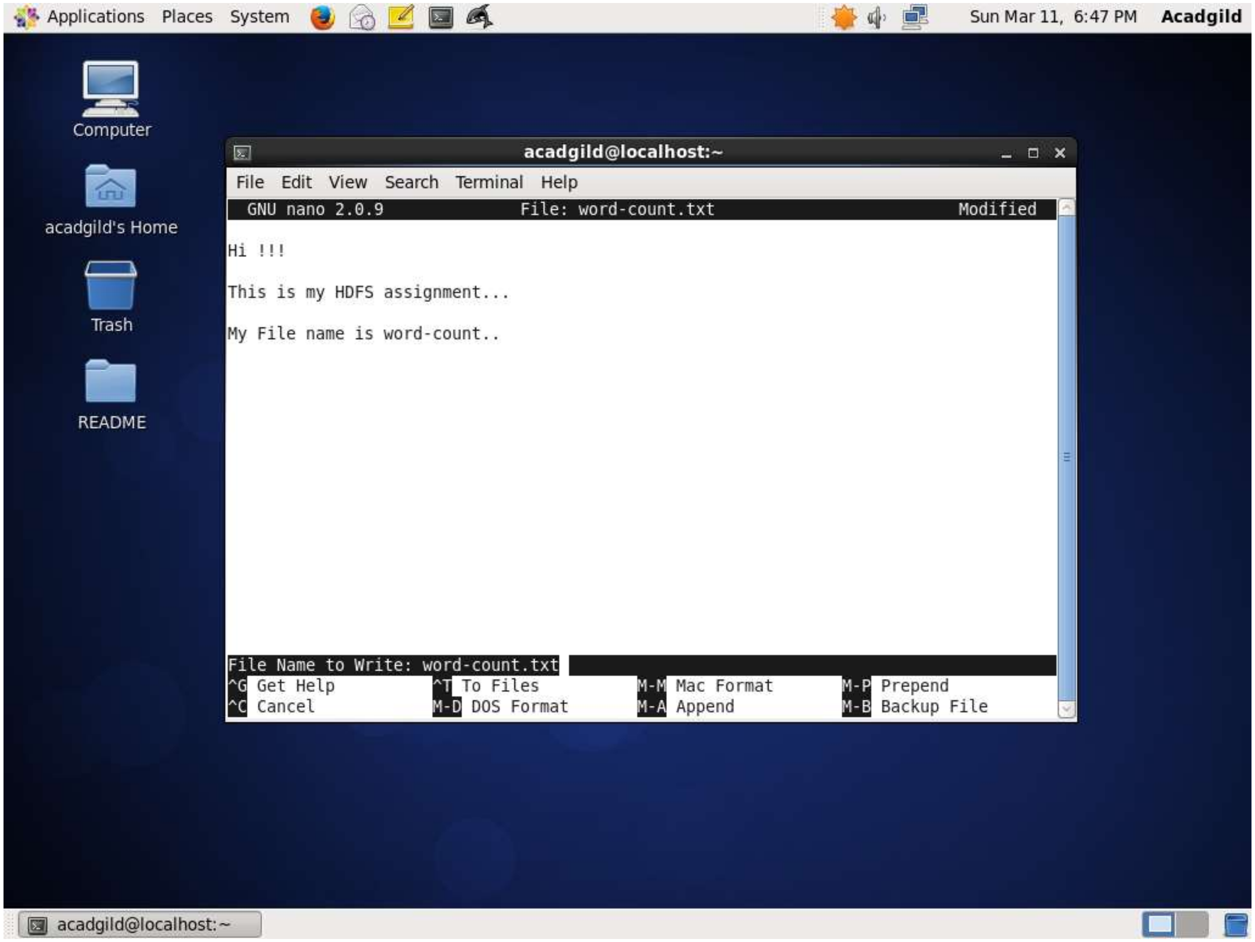


Applications Places System
Sun Mar 11, 6:45 PM Acadgild
Computer
acadgild's Home
Trash
README
acadgild@localhost:~
File Edit View Search Terminal Help
acadgild@localhost:~

```
[acadgild@localhost ~]$ nano word-count.txt
```

Word-count.txt is created.

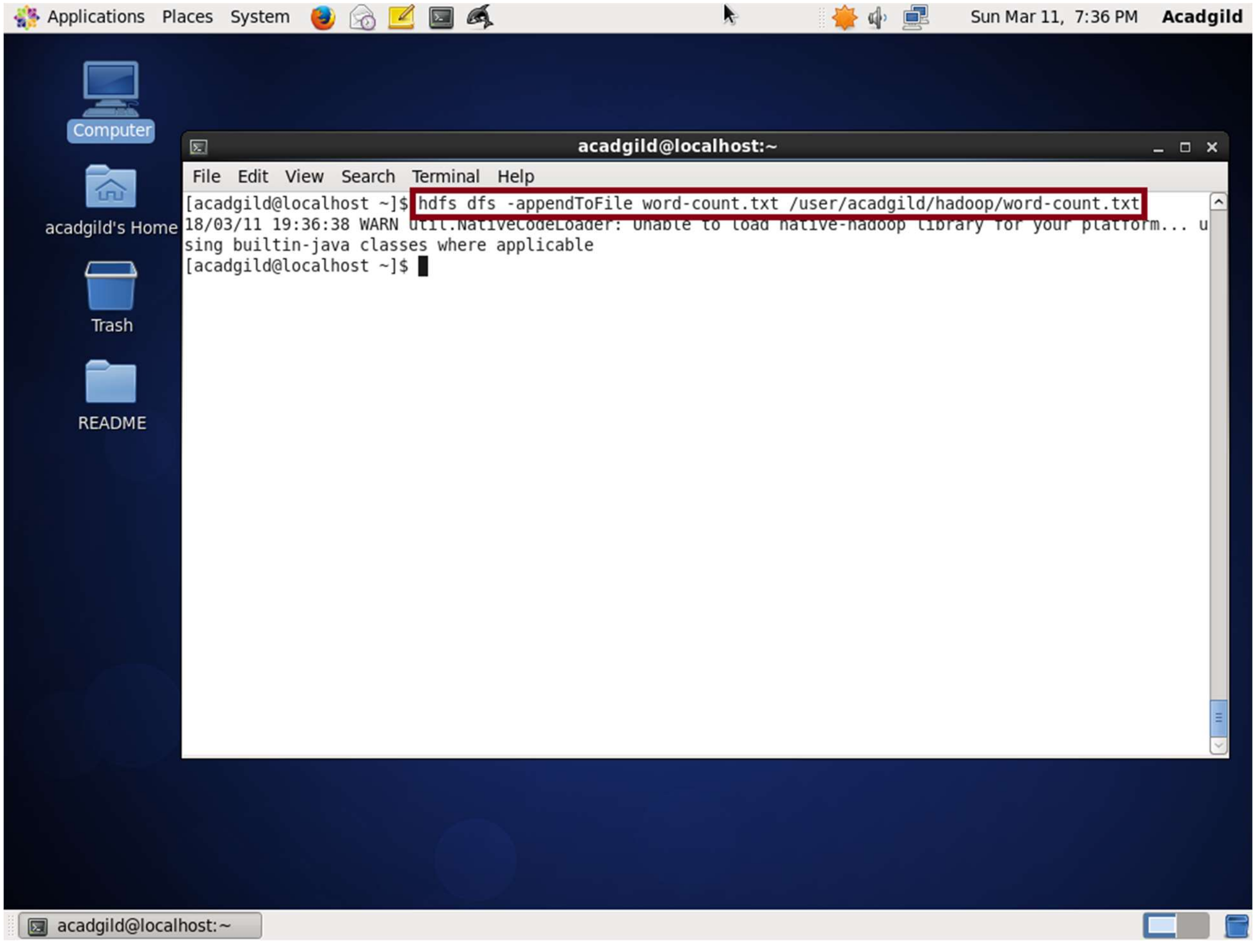
Following screenshot shows the word-count.txt file with contents added.


Applications Places System
Sun Mar 11, 6:47 PM Acadgild
Computer
acadgild's Home
Trash
README
acadgild@localhost:~
File Edit View Search Terminal Help
acadgild@localhost:~

```
  GNU nano 2.0.9              File: word-count.txt                    Modified

Hi !!!

This is my HDFS assignment...

My File name is word-count..









File Name to Write: word-count.txt
^G Get Help           ^T To Files           M-M Mac Format        M-P Prepend
^C Cancel             M-D DOS Format        M-A Append            M-B Backup File
```

Whatever we type on screen is appended into the file word-count.txt using this command

hdfs dfs –appendToFile word-count.txt /user/acadgild/Hadoop/word-count.txt



```
[acadgild@localhost ~]$ hdfs dfs -appendToFile word-count.txt /user/acadgild/hadoop/word-count.txt
18/03/11 19:36:38 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... u
sing builtin-java classes where applicable
[acadgild@localhost ~]$ 
```

-appendToFile

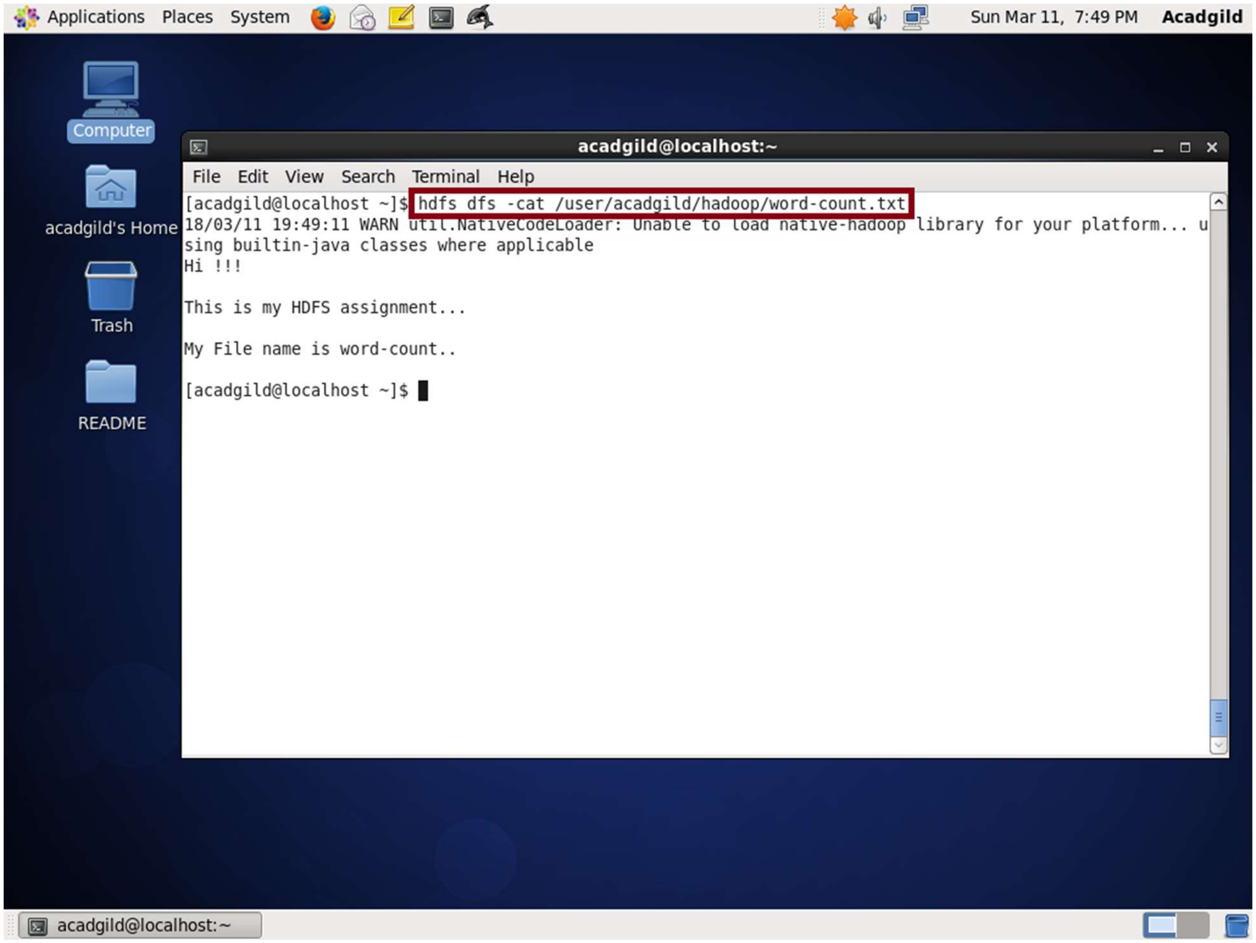
Appends the contents of all the given local files to the given destination file. The destination file will be created if it does not exist.

-cat

Fetch all files that match the file pattern <src> and display their content on stdout.

By using hdfs dfs –cat /user/acadgild/Hadoop/word-count.txt

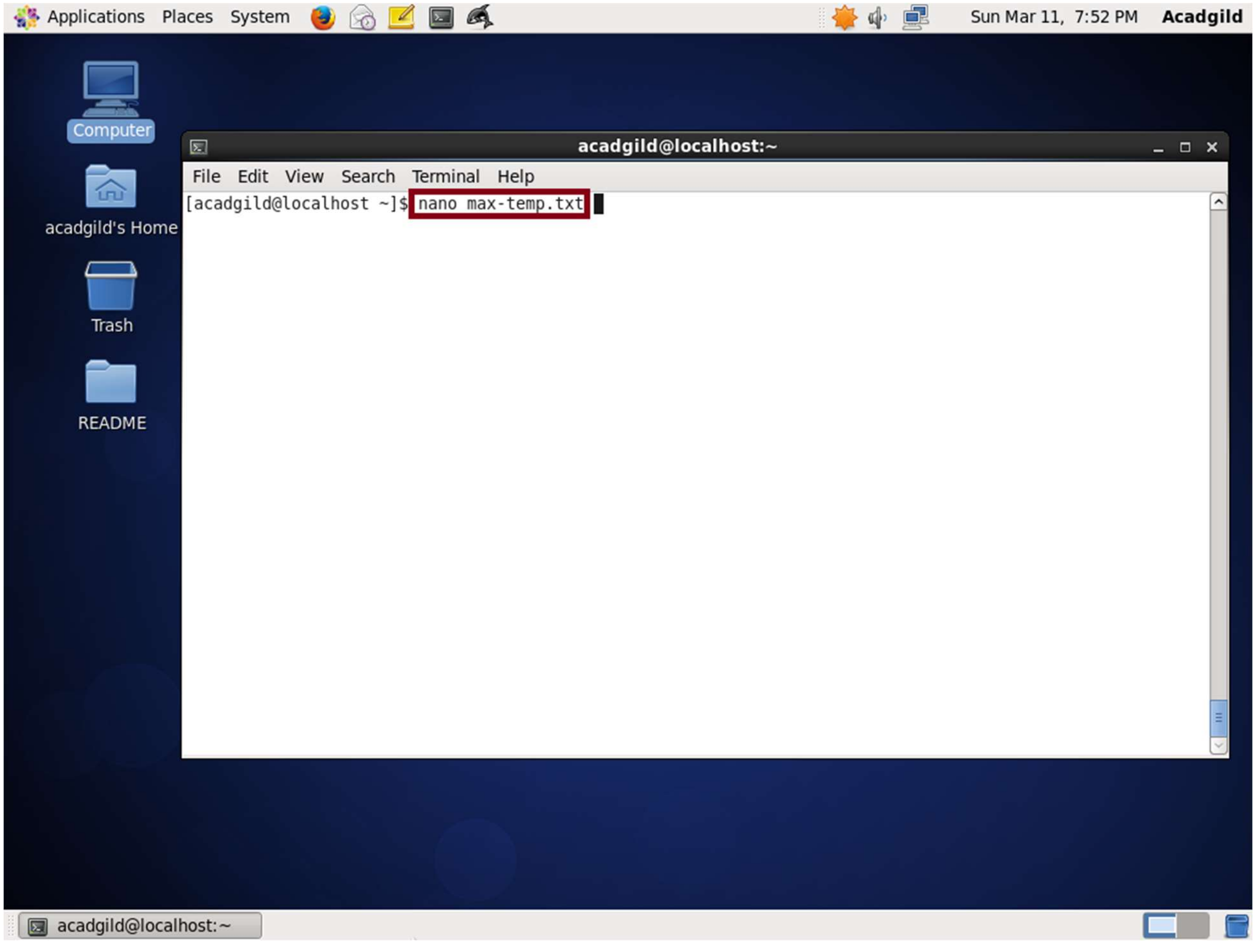
The content is displayed in stdout.


Applications Places System
Sun Mar 11, 7:49 PM Acadgild
Computer
acadgild's Home
Trash
README
acadgild@localhost:~
File Edit View Search Terminal Help
acadgild@localhost:~

```
[acadgild@localhost ~]$ hdfs dfs -cat /user/acadgild/hadoop/word-count.txt
18/03/11 19:49:11 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... u
sing builtin-java classes where applicable
Hi !!!

This is my HDFS assignment...

My File name is word-count..

[acadgild@localhost ~]$ 
```

- Task 3:

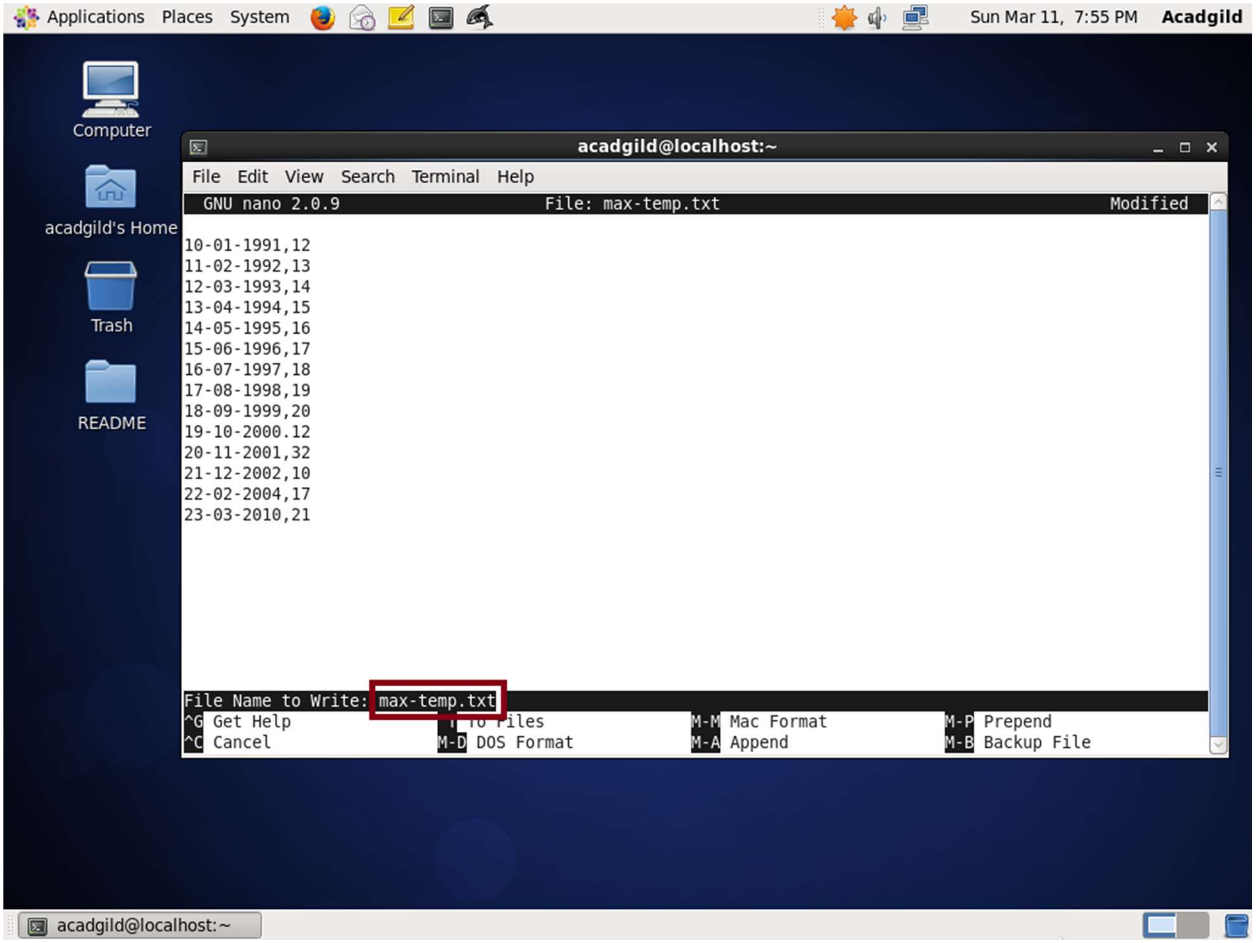
  Create a file max-temp.txt in local FS



Applications Places System
Sun Mar 11, 7:52 PM Acadgild
Computer
acadgild's Home
Trash
README
acadgild@localhost:~
File Edit View Search Terminal Help
acadgild@localhost:~

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

Max-temp.txt file is created and records of date and temperature are added in dd-mm-yyyy,degree format and saved as max-temp.txt


Applications Places System
Sun Mar 11, 7:55 PM Acadgild
Computer
acadgild's Home
Trash
README
acadgild@localhost:~
File Edit View Search Terminal Help
acadgild@localhost:~

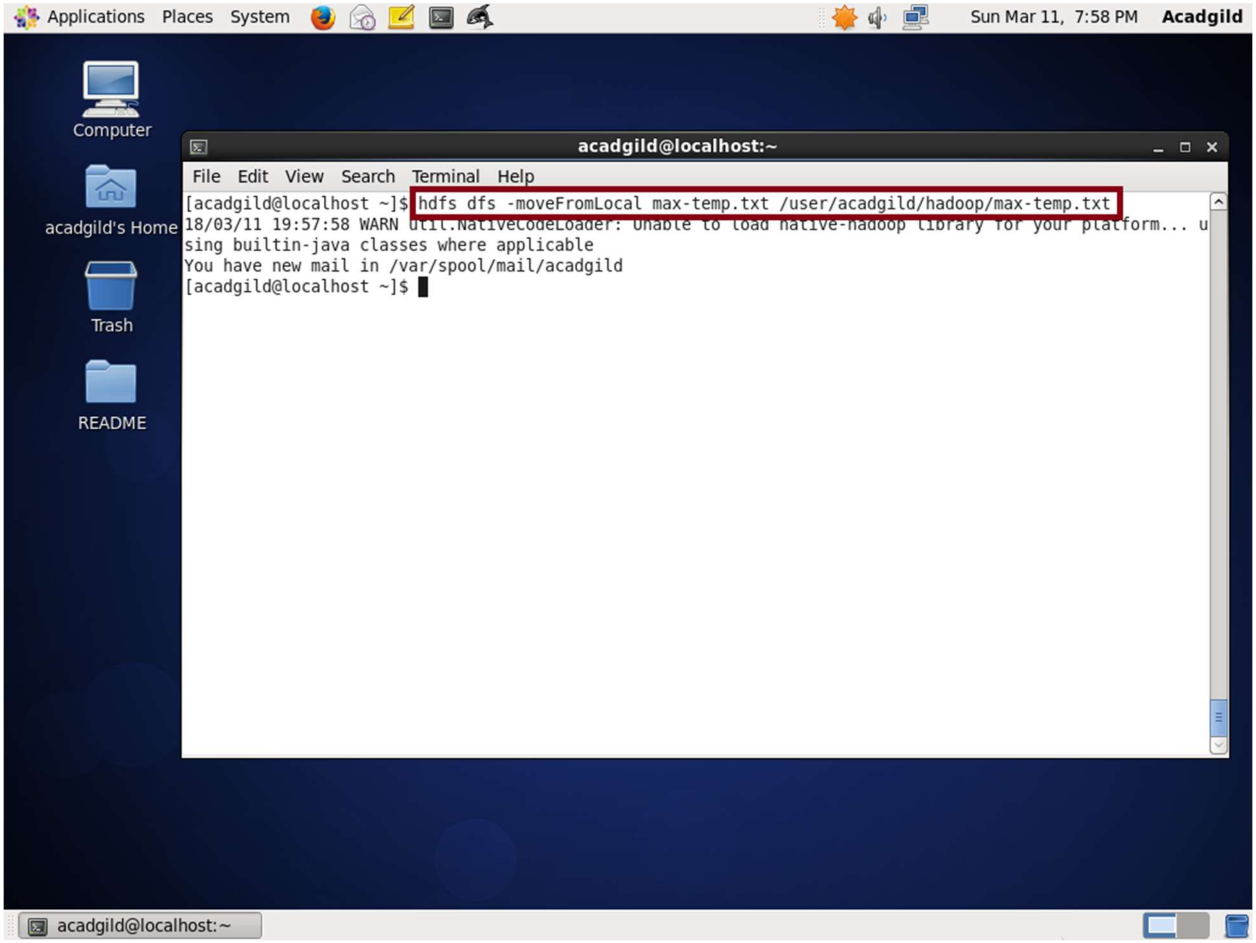
```
  GNU nano 2.0.9                 File: max-temp.txt                      Modified

10-01-1991,12
11-02-1992,13
12-03-1993,14
13-04-1994,15
14-05-1995,16
15-06-1996,17
16-07-1997,18
17-08-1998,19
18-09-1999,20
19-10-2000.12
20-11-2001,32
21-12-2002,10
22-02-2004,17
23-03-2010,21

File Name to Write: max-temp.txt
^G Get Help              ^T To Files              M-M Mac Format          M-P Prepend
^C Cancel                M-D DOS Format           M-A Append              M-B Backup File
```

-moveFromLocal

Copy files from the local file system into fs and deletes the original once it is copied. We can use put command also for same operation but the original file will exist after the copying file.

The file can be moved using the following command

hdfs dfs –moveFromLocal max-temp.txt /user/acadgild/hadoop/max-temp.txt



```
[acadgild@localhost ~]$ hdfs dfs -moveFromLocal max-temp.txt /user/acadgild/hadoop/max-temp.txt
18/03/11 19:57:58 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... u
sing builtin-java classes where applicable
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ 
```

- Task 4

The default permission settings for the /user/acadgild/Hadoop/max-temp.txt is shon using getfacl command.

-getfacl

-getfactl [-R] <path> :

Displays the Access Control Lists (ACL) of files and directories. If a directory has a default ACL, then getfacl also displays the default ACL.

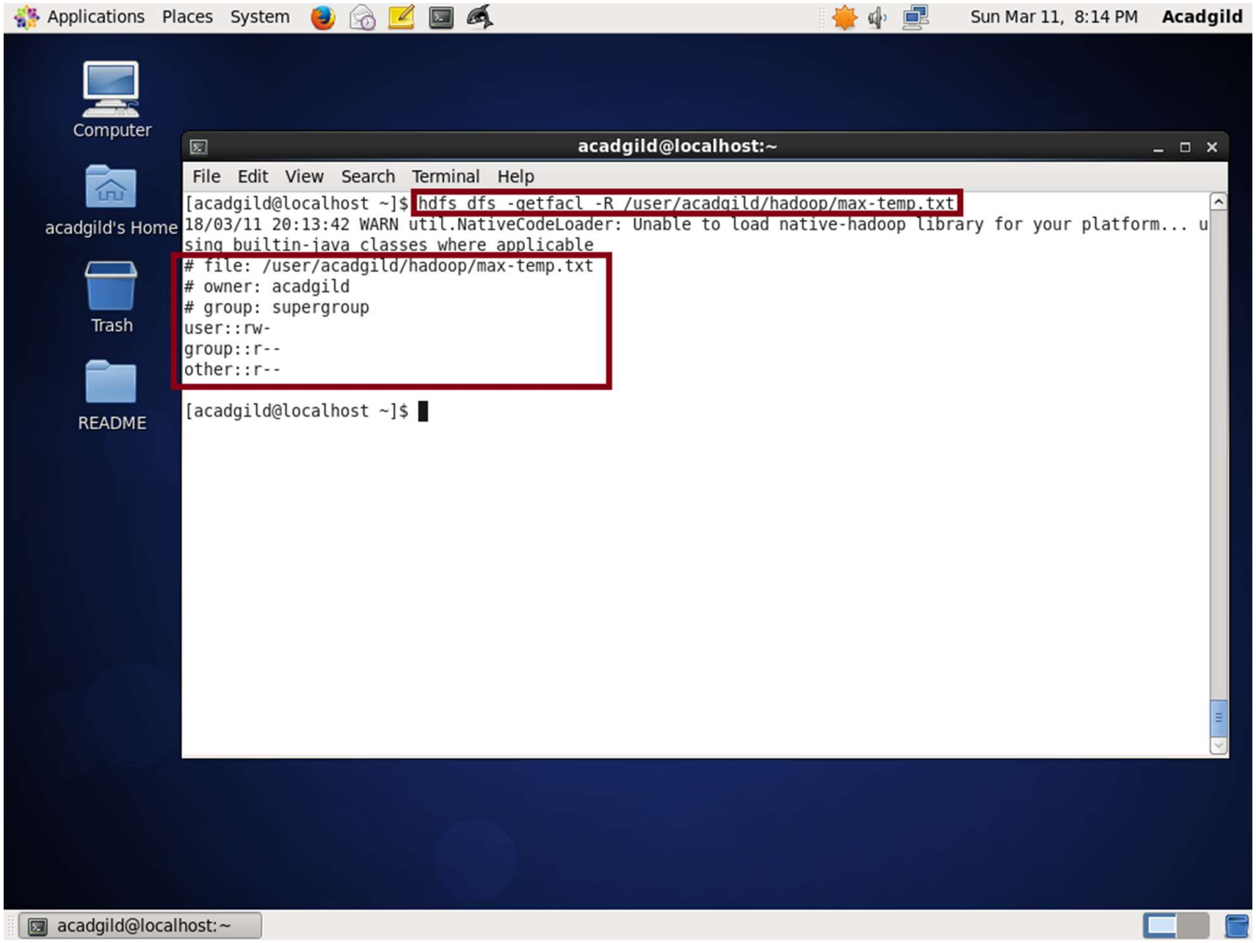
-R

Lists the ACLs of all files and directories recursively.

<path>

File or directory to list.

hdfs dfs –getfacl –R /user/acadgild/Hadoop/max.temp.txt is used to show the current ACL setup for the **max-temp.txt**



Applications Places System
Sun Mar 11, 8:14 PM Acadgild
Computer
acadgild's Home
Trash
README
acadgild@localhost:~
File Edit View Search Terminal Help
acadgild@localhost:~

```
[acadgild@localhost ~]$ hdfs dfs -getfacl -R /user/acadgild/hadoop/max-temp.txt
18/03/11 20:13:42 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... u
sing builtin-java classes where applicable
# file: /user/acadgild/hadoop/max-temp.txt
# owner: acadgild
# group: supergroup
user::rw-
group::r--
other::r--

[acadgild@localhost ~]$ 
```

File is /user/acadgild/Hadoop/max-temp.txt

By default **User** is having **reader and writer access**, **group** is Supergroup and group have **reader** access.

Other have **reader** access.

Now we have to change the permission of user and group to have full access control and we have to remove access to others.

-chmod

-chmod [-R] <MODE[, MODE]… | OCTALMODE> PATH… :

Changes permissions of a file. This works similar to the shell's chmod command with a few exceptions.

-R

Modifies the files recursively. This is the only option currently supported.

<MODE>

Mode is the same as mode used for the shell's command. The only letters recognized are 'rwxXt',

e.g. +t,a+r,g-w,+rwx,o=r.

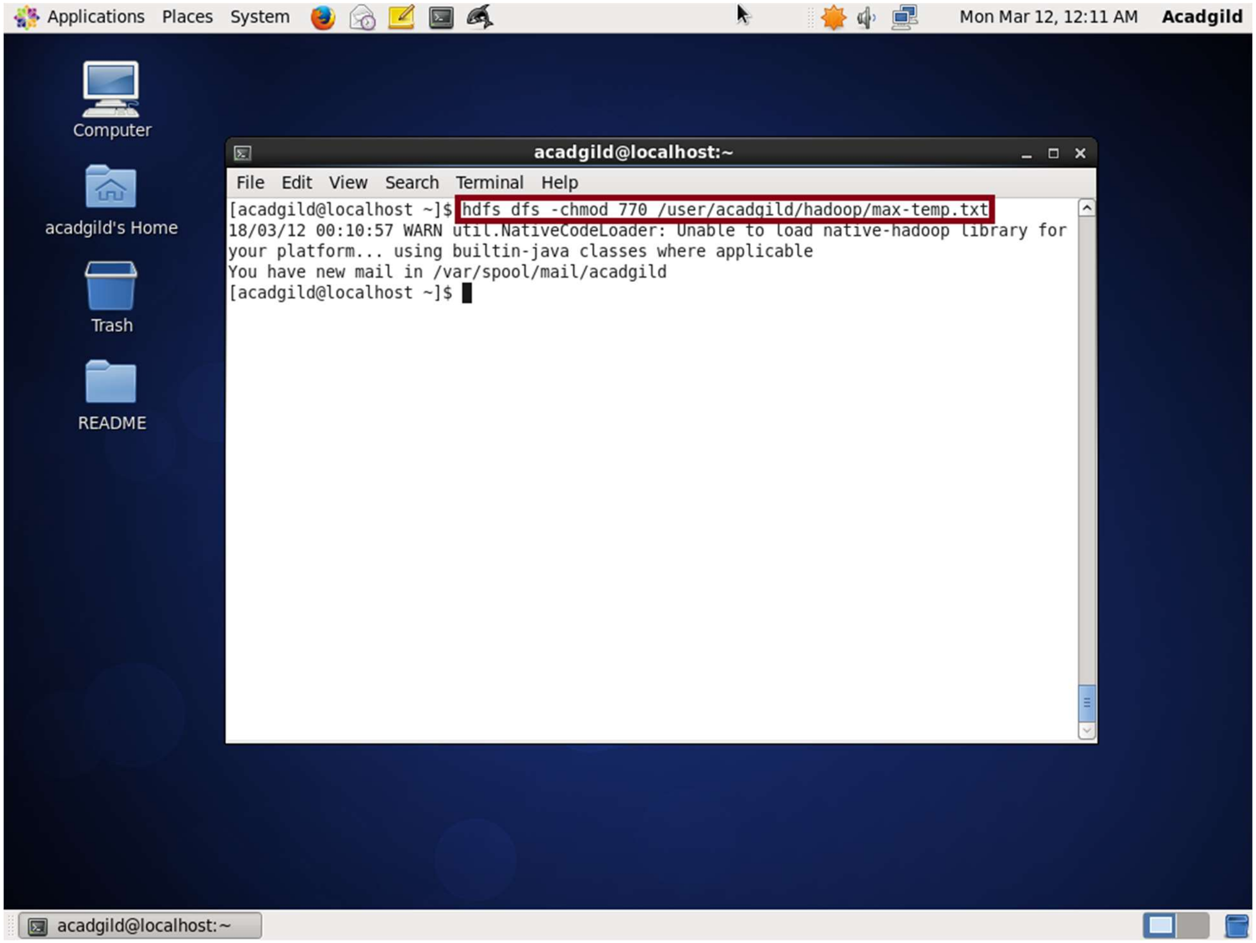
<OCTALMODE>

Mode specified in 3 or 4 digits. If 4 digits, the first may be 1 or 0 to turn the sticky bit on or off, respectively. Unlike the mode, e.g 754 is same as u=rwx,g=rx,o=r.

Here in our case **chmod 554<filename>**

We have to give full access to user and group. Others should not have access.

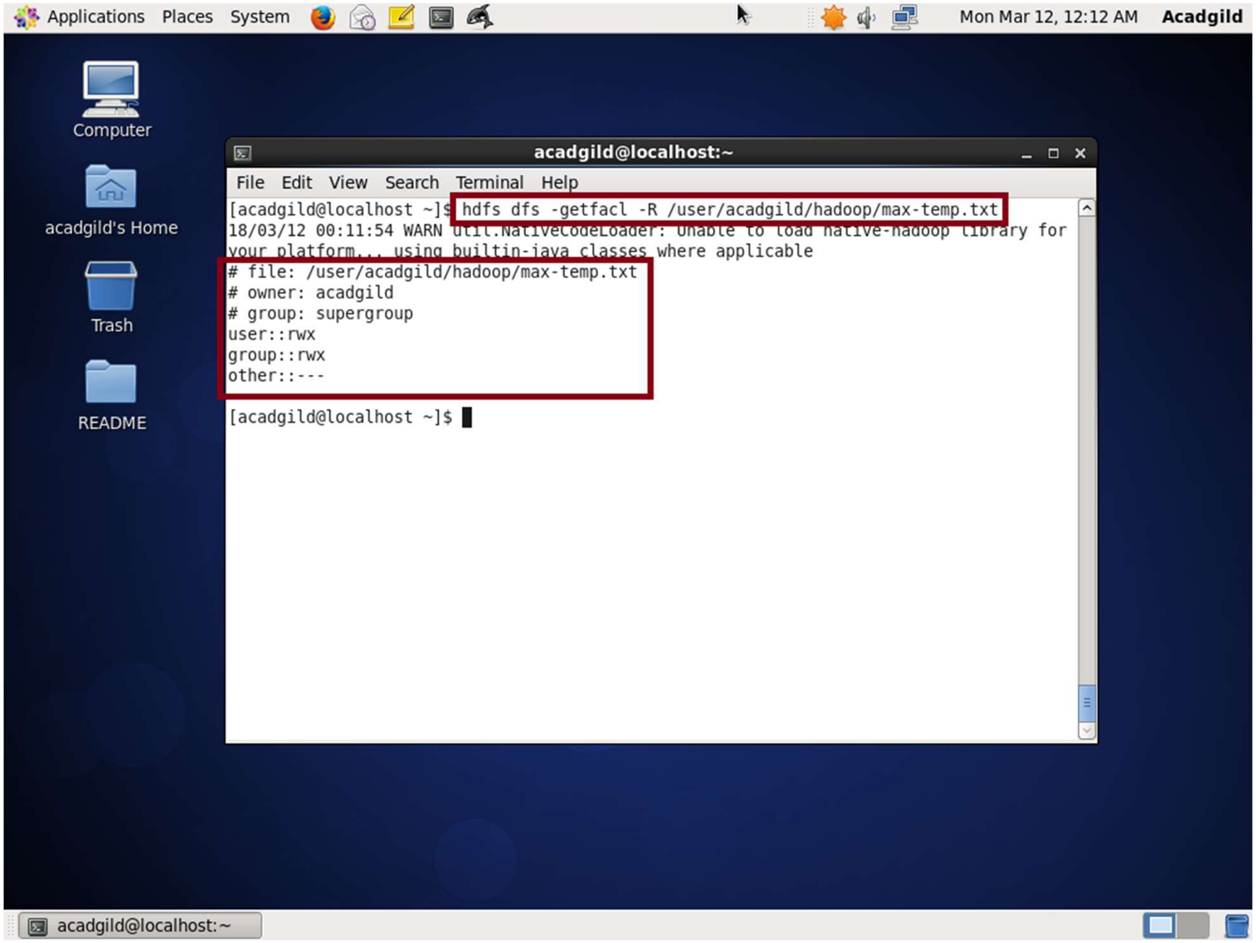
hdfs dfs –Chmod 770 /user/acadgild/Hadoop/max-temp.txt



Applications Places System
Mon Mar 12, 12:11 AM Acadgild
Computer
acadgild's Home
Trash
README
acadgild@localhost:~
File Edit View Search Terminal Help
acadgild@localhost:~

```
[acadgild@localhost ~]$ hdfs dfs -chmod 770 /user/acadgild/hadoop/max-temp.txt
18/03/12 00:10:57 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 –getfacl –R /user/acadgild/Hadoop/max.temp.txt is used to show the current ACL setup for the **max-temp.txt**



```
[acadgild@localhost ~]$ hdfs dfs -getfacl -R /user/acadgild/hadoop/max-temp.txt
18/03/12 00:11:54 WARN util.NativeCodeLoader: Unable to load native-hadoop library for
your platform... using builtin-java classes where applicable
# file: /user/acadgild/hadoop/max-temp.txt
# owner: acadgild
# group: supergroup
user::rwx
group::rwx
other::---

[acadgild@localhost ~]$ 
```

Now user and group is having full access rwx .

Others don’t have access.