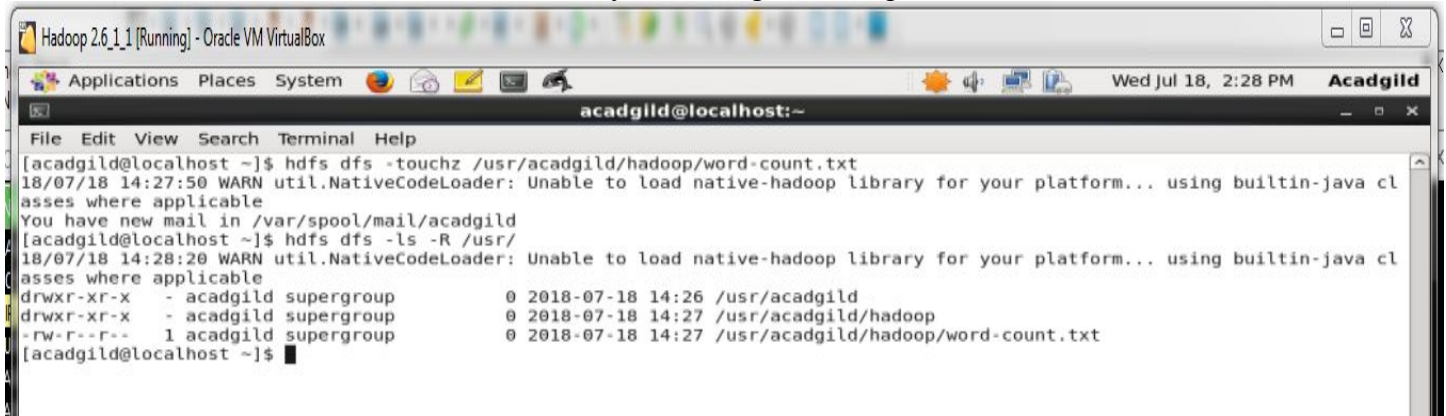


Task 2:

1. Create a file in HDFS under the directory `/usr/acadgild/hadoop`, with a name `word-count.txt`.



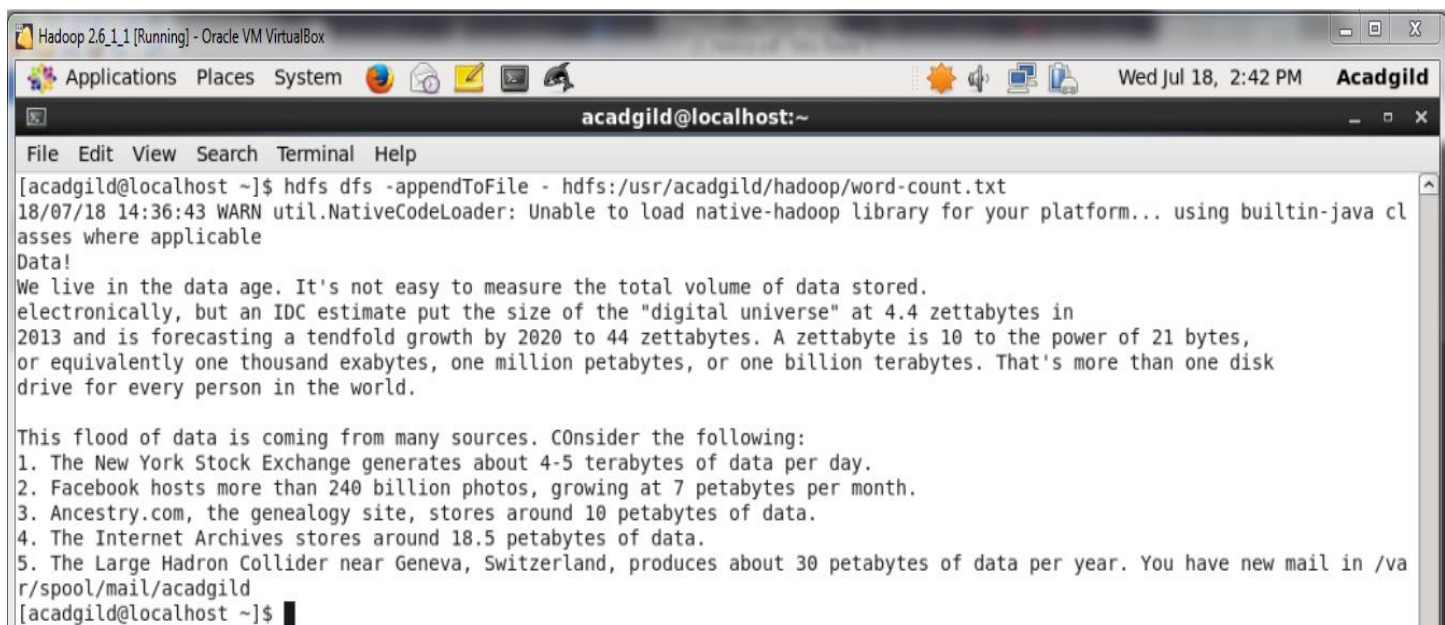
The screenshot shows a terminal window titled 'Hadoop 2.6.1_1 [Running] - Oracle VM VirtualBox'. The user 'Acadgild' is logged in. The terminal shows the command `hdfs dfs -touchz /usr/acadgild/hadoop/word-count.txt` being executed. The output includes a warning message from 'util.NativeCodeLoader' and a confirmation that the file was created. The `ls -R` command is also used to verify the file's existence.

```
[acadgild@localhost ~]$ hdfs dfs -touchz /usr/acadgild/hadoop/word-count.txt
18/07/18 14:27:50 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
You have new mail in /var/spool/mail/acadgild
[acadgild@localhost ~]$ hdfs dfs -ls -R /usr/
18/07/18 14:28:20 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
drwxr-xr-x  - acadgild supergroup          0 2018-07-18 14:26 /usr/acadgild
drwxr-xr-x  - acadgild supergroup          0 2018-07-18 14:27 /usr/acadgild/hadoop
-rw-r--r--  1 acadgild supergroup          0 2018-07-18 14:27 /usr/acadgild/hadoop/word-count.txt
[acadgild@localhost ~]$
```

2. A file can be created in *hdfs* using the command called `-touchz`:
Ex: `$hdfs dfs -touchz /usr/acadgild/hadoop/word-count.txt`
3. Once the file is get created, you can verify that with `-ls -R` (Recursive) command:
Ex: `$hdfs dfs -ls -R /usr/`

In the above figure you can observe that the file created under the directory `/usr/acadgild/hadoop/word-count.txt`.

4. Now we can write continuous stream of bytes to the file `/usr/acadgild/hadoop/word-count.txt`.



The screenshot shows a terminal window titled 'Hadoop 2.6.1_1 [Running] - Oracle VM VirtualBox'. The user 'Acadgild' is logged in. The terminal shows the command `hdfs dfs -appendToFile - hdfs:/usr/acadgild/hadoop/word-count.txt` being executed. The output includes a warning message from 'util.NativeCodeLoader' and a confirmation that the data was appended to the file. The data being appended is a paragraph about the digital universe and a list of sources.

```
[acadgild@localhost ~]$ hdfs dfs -appendToFile - hdfs:/usr/acadgild/hadoop/word-count.txt
18/07/18 14:36:43 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Data!
We live in the data age. It's not easy to measure the total volume of data stored.
electronically, but an IDC estimate put the size of the "digital universe" at 4.4 zettabytes in
2013 and is forecasting a tendfold growth by 2020 to 44 zettabytes. A zettabyte is 10 to the power of 21 bytes,
or equivalently one thousand exabytes, one million petabytes, or one billion terabytes. That's more than one disk
drive for every person in the world.

This flood of data is coming from many sources. COnsider the following:
1. The New York Stock Exchange generates about 4-5 terabytes of data per day.
2. Facebook hosts more than 240 billion photos, growing at 7 petabytes per month.
3. Ancestry.com, the genealogy site, stores around 10 petabytes of data.
4. The Internet Archives stores around 18.5 petabytes of data.
5. The Large Hadron Collider near Geneva, Switzerland, produces about 30 petabytes of data per year. You have new mail in /va
r/spool/mail/acadgild
[acadgild@localhost ~]$
```

A stream of bytes can be written to the file using commands `-appendToFile`:

Ex: `$hdfs dfs -appendToFile - hdfs:/usr/acadgild/hadoop/word-count.txt`

Once the writing to the file is over press `ctrl+D` (EOF), this will save the contents to the file `word-count.txt`.

5. The contents of the file are verified by using `-cat` command.
Ex: `$hdfs dfs -cat /usr/acadgild/hadoop/word-count.txt`

```
Hadoop 2.6.1_1 [Running] - Oracle VM VirtualBox
Applications Places System
acadmild@localhost:~
File Edit View Search Terminal Help
[acadgild@localhost ~]$ hdfs dfs -cat /usr/acadgild/hadoop/word-count.txt
18/07/18 14:47:47 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java cl
asses where applicable
Data!
We live in the data age. It's not easy to measure the total volume of data stored.
electronically, but an IDC estimate put the size of the "digital universe" at 4.4 zettabytes in
2013 and is forecasting a tendfold growth by 2020 to 44 zettabytes. A zettabyte is 10 to the power of 21 bytes,
or equivalently one thousand exabytes, one million petabytes, or one billion terabytes. That's more than one disk
drive for every person in the world.

This flood of data is coming from many sources. Consider the following:
1. The New York Stock Exchange generates about 4-5 terabytes of data per day.
2. Facebook hosts more than 240 billion photos, growing at 7 petabytes per month.
3. Ancestry.com, the genealogy site, stores around 10 petabytes of data.
4. The Internet Archives stores around 18.5 petabytes of data.
5. The Large Hadron Collider near Geneva, Switzerland, produces about 30 petabytes of data per year. [acadgild@localhost ~]$
[acadgild@localhost ~]$
```

Here you can observe that the contents of the file what we have entered, are displayed.