6. Problem Statement

# HadoopSession

A simple session call and recall.

Complete the following Hadoop

exercises in a GCP Dataproc cluster. For each step, capture and insert screenshots to show

the commands and results.

Step 1. Create a Dataproc cluster associating with an existing storage bucket.

Step 2. SSH connect to the master node of the cluster and open a terminal

You should complete the following steps in the terminal.

Step 3. Use a command to show the Hadoop version.

Step 4. List the content of the HDFS in your cluster.

Step 5. Create a folder /user/info323' in the HDFS.

Step 6. Copy a file “test.txt" from the file system on the master node to the HDFS folder

‘info323' you just created.

Step 7. Copy the “test.txt file from the HDFS to the home directory on the master node

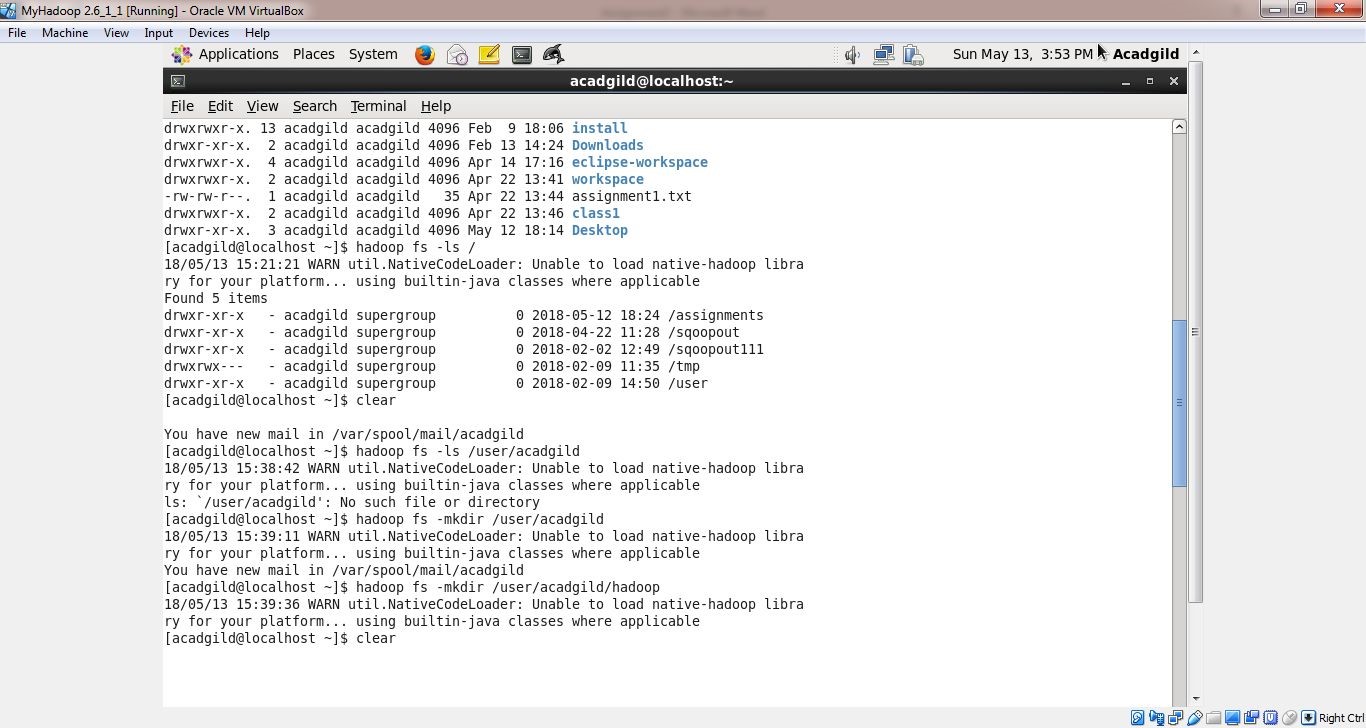
as “test-from-HDFS.txt".

Step 8. Delete the file and the folder in the HDFS.

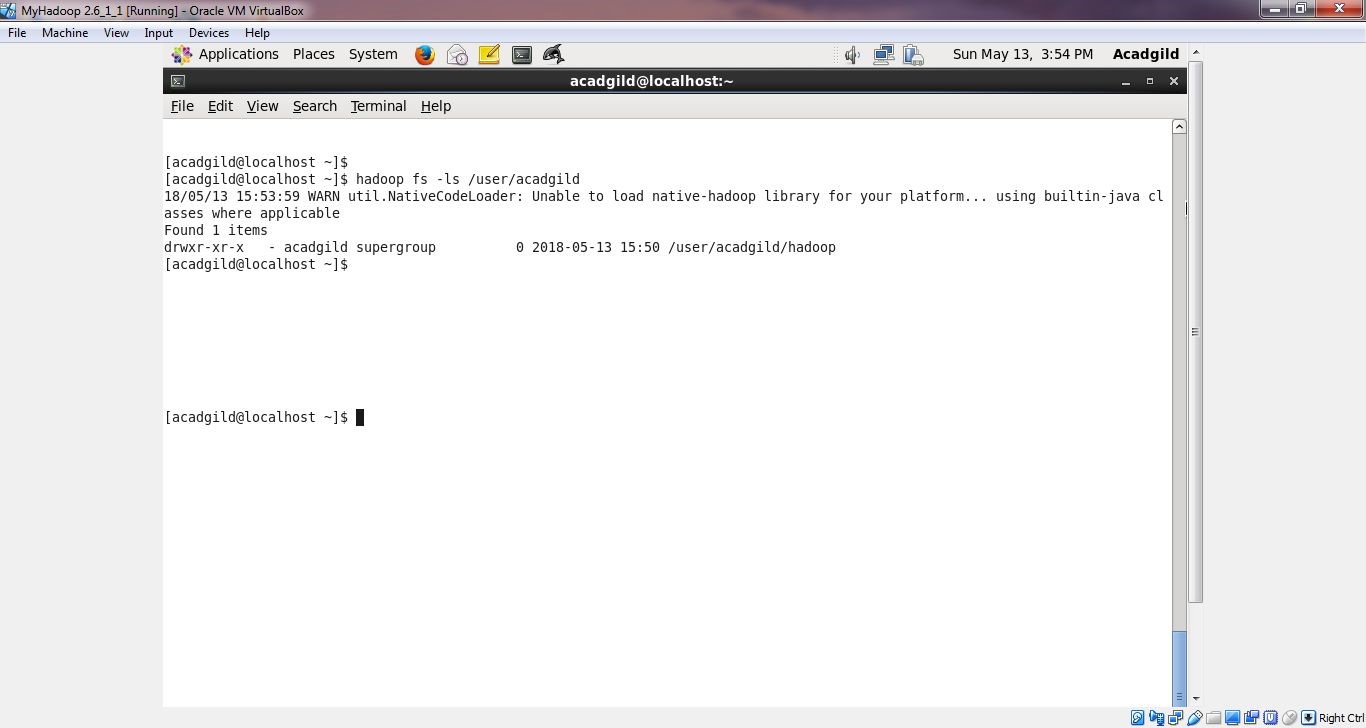
Task 1:

Create a directory for user

Hadoop fs –mkdir creates a directory with the name specified in the arguments.



Hadoop fs –ls lists all files in a directory



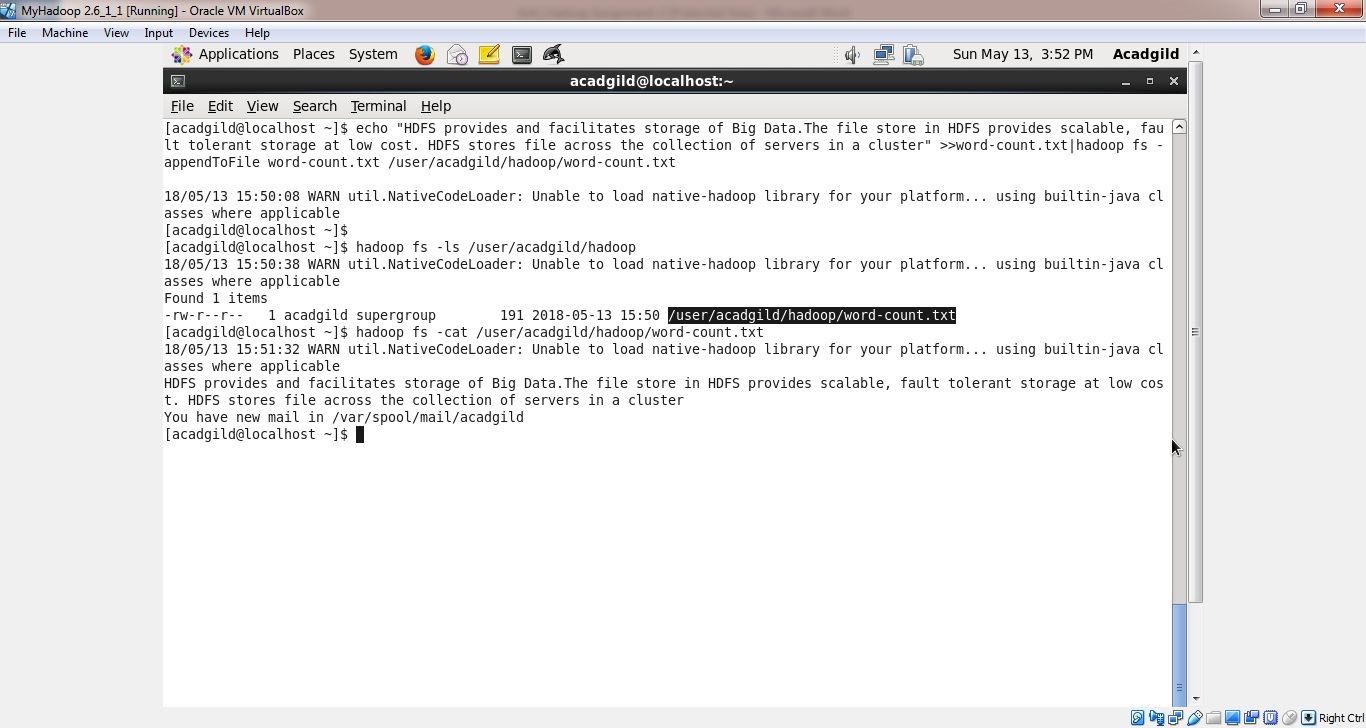
Task 2:

Create a file in HDFS under directory /user/filename/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.

Output of echo is appended to file word-count.txt and appendToFile command in Hadoop adds the file to Hadoop file system.



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

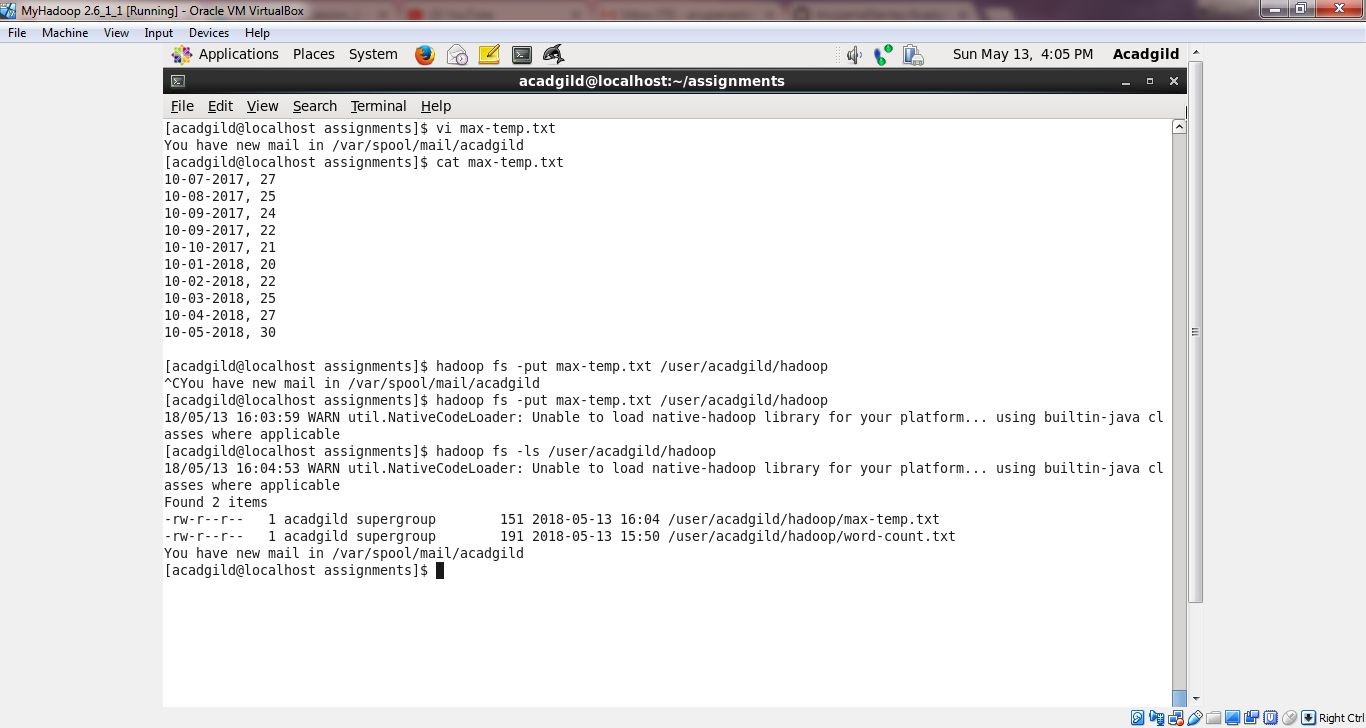
Example:

10-01-1990,10

10-02-1991,20

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

Created a file max-temp.txt using vi command and using Hadoop fs –put commnd, moved this file to HDFS directory /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.

Hadoop fs –chmod 770 provides full permission to owner and group and no permission to others on the file.

