**Exp.No: 3**

**Map Reduce program to process a weather dataset**

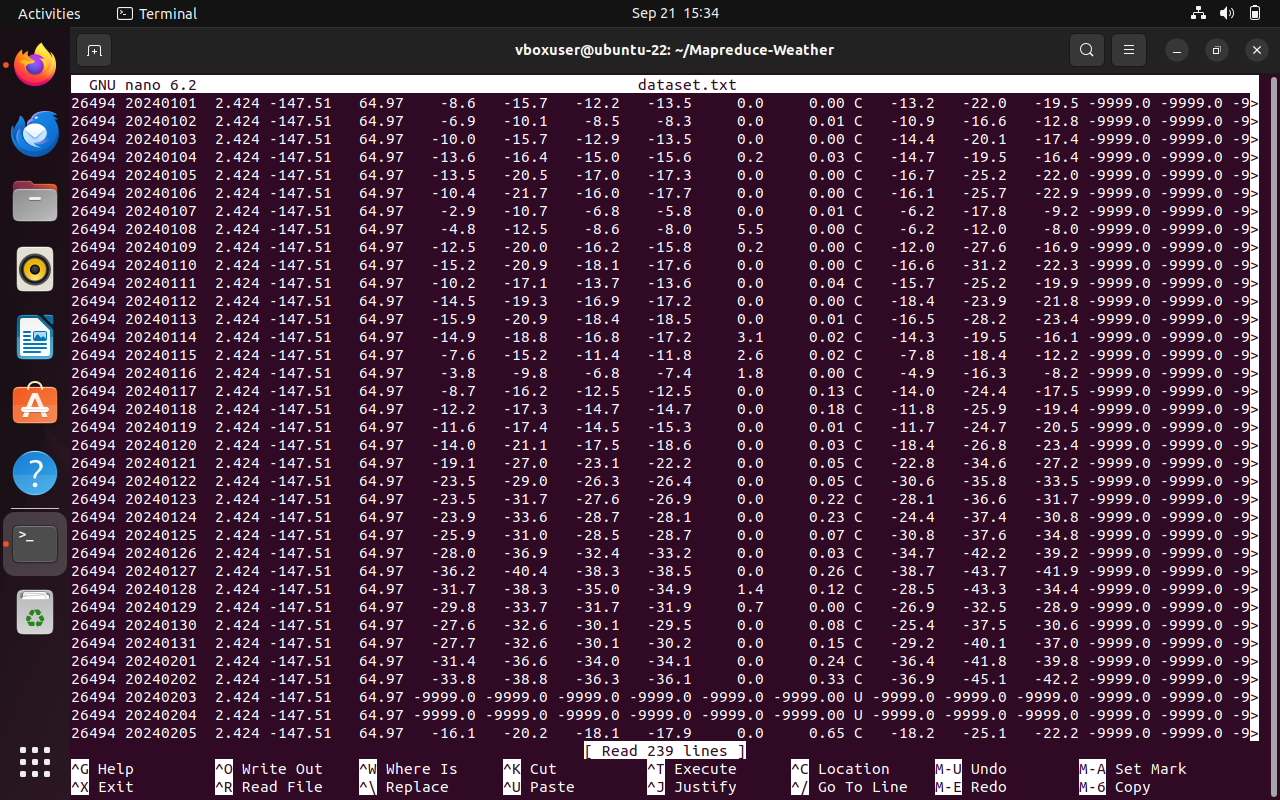
**AIM:**

To implement MapReduce program to process a weather dataset.

**PROCEDURE:**

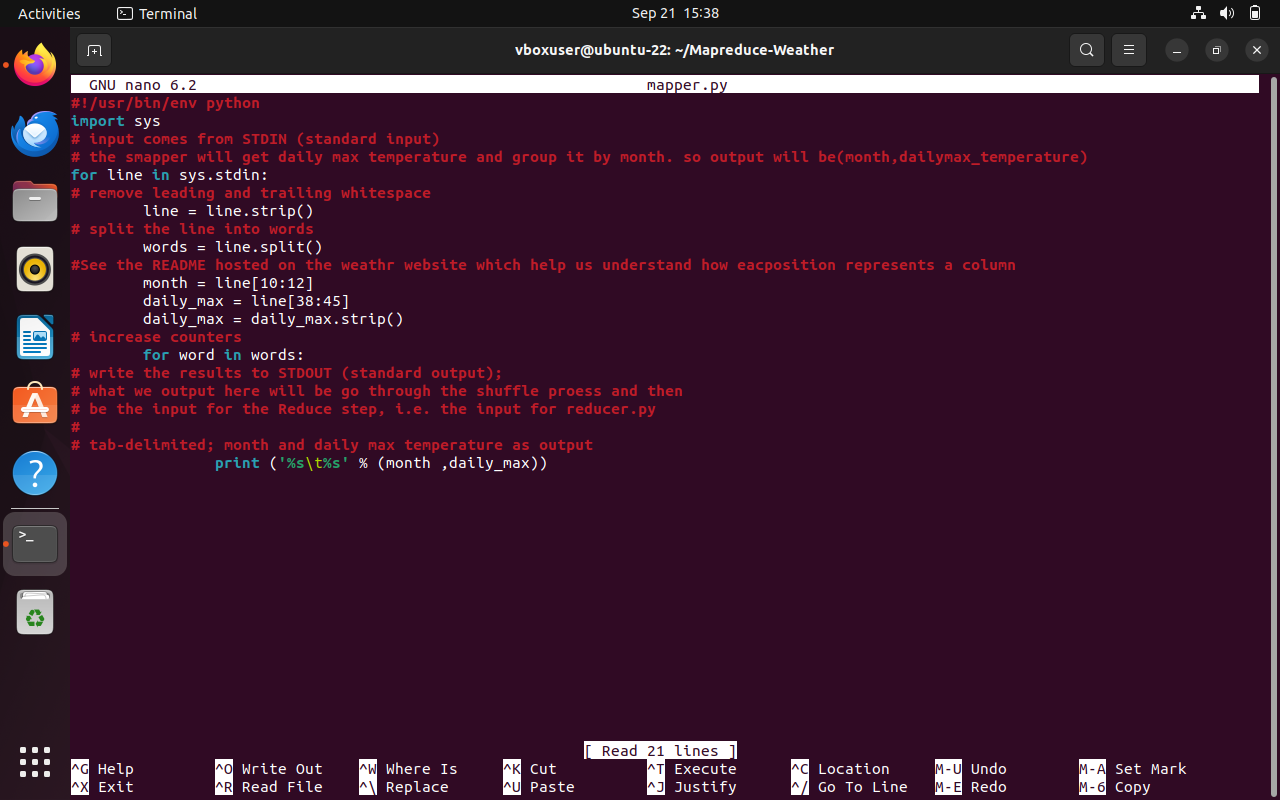
**Step 1: Create Data File:** Create a file named "word\_count\_data.txt" and populate it with text data that you wish to analyse. Login with your hadoop user.

**Output:**

****

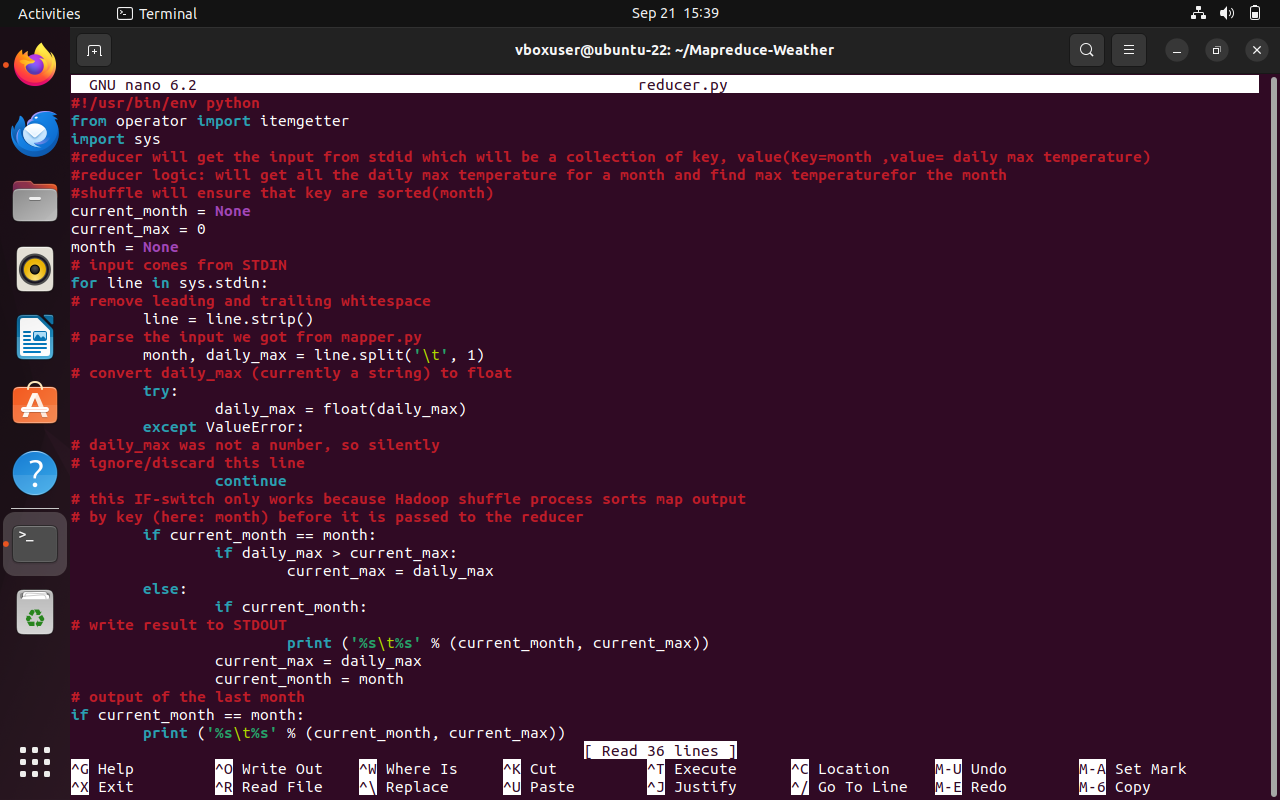
**Step 2: Mapper Logic - mapper.py:** Create a file named "mapper.py" to implement the logic for the mapper. The mapper will read input data from STDIN, split lines into words, and output each word with its count.

**nano mapper.py**

****

**Step 3: Reducer Logic - reducer.py:** Create a file named "reducer.py" to implement the logic for the reducer. The reducer will aggregate the occurrences of each word and generate the final output.

**nano reducer.py**

****

**Step 4: Prepare Hadoop Environment:** Start the Hadoop daemons and create a directory in HDFS to store your data.

**start-all.sh**

**Step 6: Make Python Files Executable:** Give executable permissions to your mapper.py and reducer.py files.

**chmod 777 mapper.py reducer.py**

**Step 7: Run the program using Hadoop Streaming:** Download the latest hadoop-streaming jar file and place it in a location you can easily access.

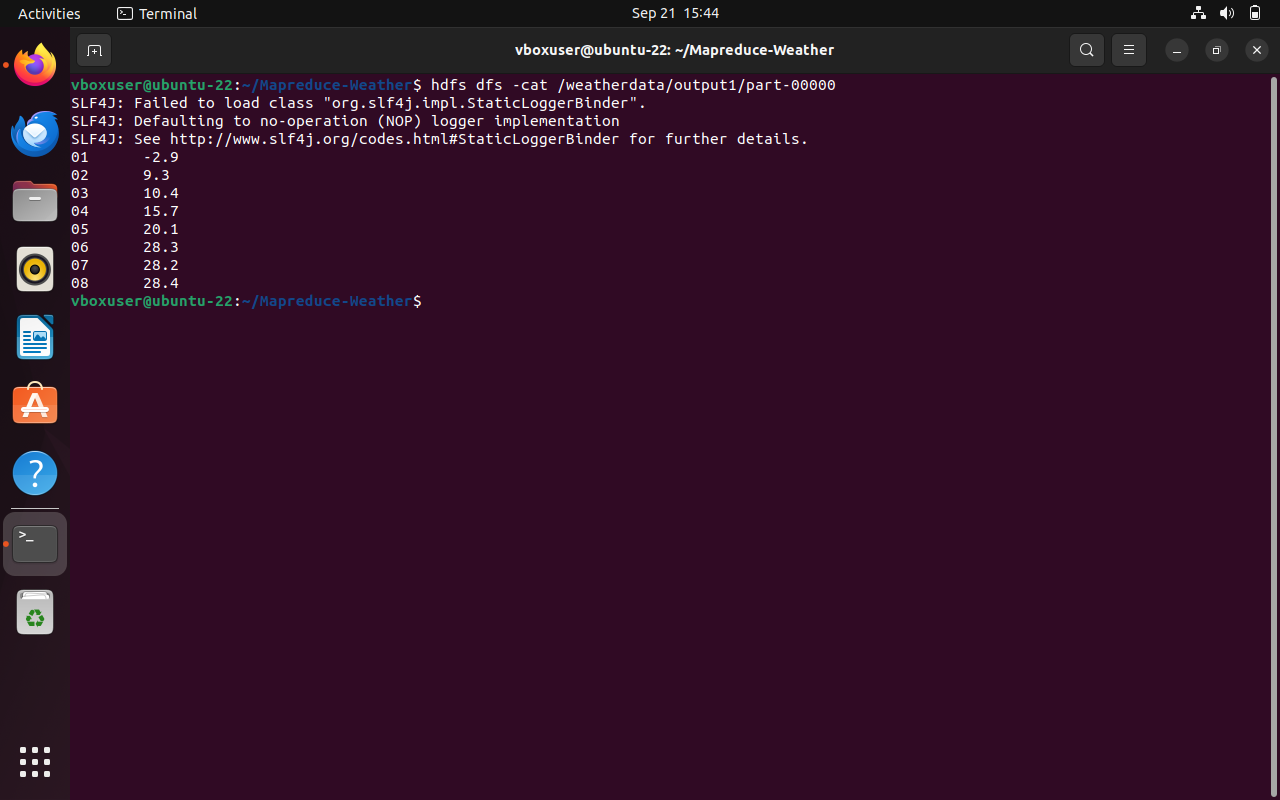
Then run the program using Hadoop Streaming.

**hadoop fs -mkdir -p /weatherdata hadoop fs -copyFromLocal /home/vboxuser/Downloads/dataset.txt /weatherdata**

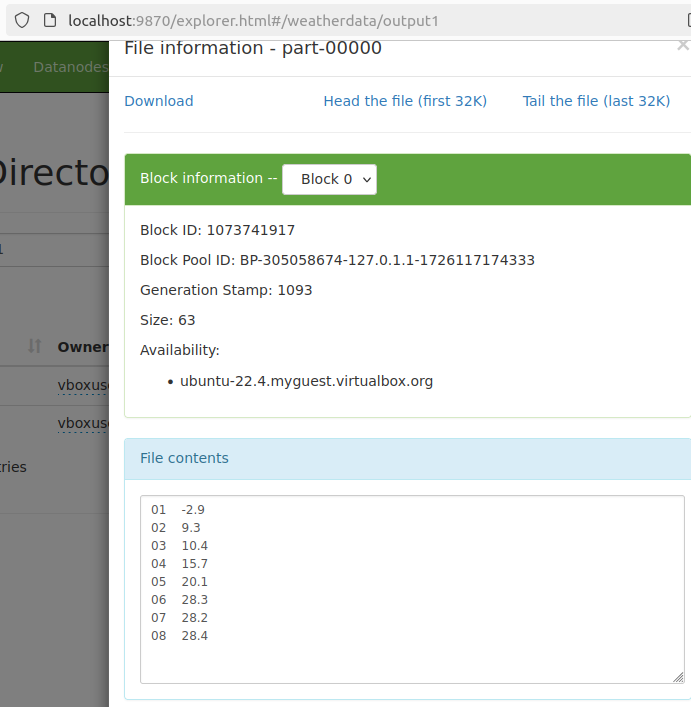
**hdfs dfs -ls /weatherdata hadoop jar /home/vboxuser/hadoop-3.2.3/share/hadoop/tools/lib/hadoop-streaming-3.2.3.jar \ -input /weatherdata/dataset.txt \ -output /weatherdata/output \ -file "/home/vboxuser/Downloads/mapper.py" \ -mapper "python3 mapper.py" \ -file "/home/vboxuser/Downloads/reducer.py" \ -reducer "python3 reducer.py"**

**hdfs dfs -text /weatherdata/output/\* > /home/vboxuser/Downloads/outputfile.txt**

**Step 8: Check Output:** Check the output of the program in the specified HDFS output directory.



**Step 9:** The result in the browser is as follows:



**RESULT:**

Thus, the program for weather dataset using Map Reduce has been executed successfully.