**Lab 4: Python Hadoop Streaming**

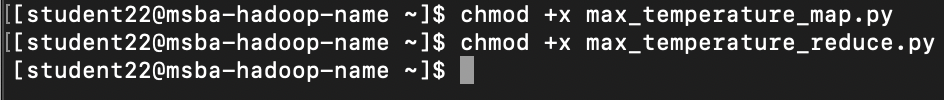
**Write down the commands and paste the corresponding screenshots in the following:**

**Task 1: Execute the map and reduce python scripts at University Hadoop System.**

1). Change the execution permission of the python files.

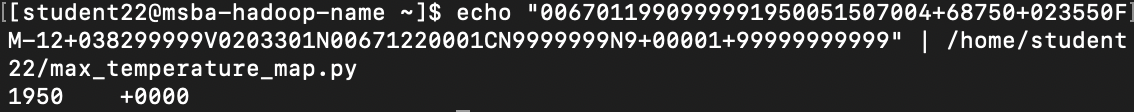
chmod +x max\_temperature\_map.py

chmod +x max\_temperature\_reduce.py

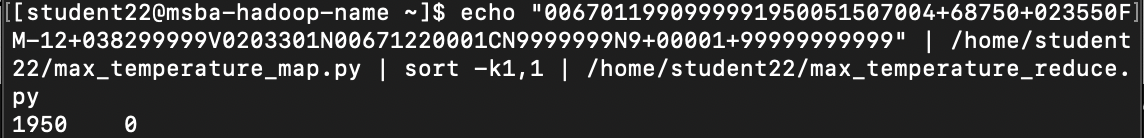


2). Test the two python files locally before running them using Hadoop.

echo "0067011990999991950051507004+68750+023550FM-12+038299999V0203301N00671220001CN9999999N9+00001+99999999999" | /home/student22/max\_temperature\_map.py

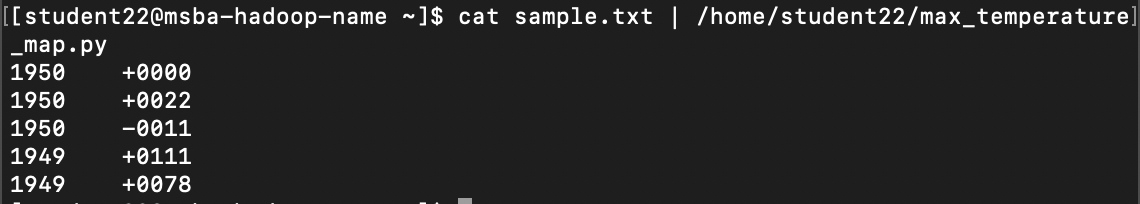


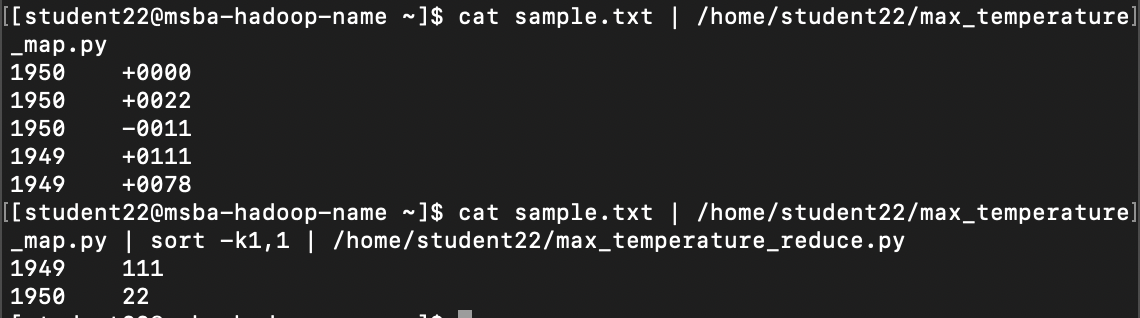
echo "0067011990999991950051507004+68750+023550FM-12+038299999V0203301N00671220001CN9999999N9+00001+99999999999" | /home/student22/max\_temperature\_map.py | sort -k1,1 | /home/student22/max\_temperature\_reduce.py



3). Copy the sample data to HDFS and Execute the mapper and reducer using Hadoop streaming.

cat sample.txt | /home/student22/max\_temperature\_map.py



cat sample.txt | /home/student22/max\_temperature\_map.py | sort -k1,1 | /home/student22/max\_temperature\_reduce.py

4). Find the print out the final output.

Graphical user interface, text, application

Description automatically generated

**Task 2: Execute the map and reduce python scripts at AWS EMR.**

1). Paste a screenshot of streaming before adding the step;

Graphical user interface, text, application, email

Description automatically generated

2). Paste the output directory screenshot;

Graphical user interface, text, application, email

Description automatically generated

3). Open the output file and paste the screenshot of content from the output file.

Graphical user interface, text, application

Description automatically generated