

Big Data Analytics Lab Assignment

Name:Bharath.C

Reg No:126015014

Roll No:11

Weather Analysis

Mapper.py:

```
#!/usr/bin/env python
```

```
import sys
```

```
for line in sys.stdin:
```

```
    line = line.strip()
```

```
    if not line:
```

```
        continue
```

```
    try:
```

```
        month, min_temp, max_temp = line.split(',')
```

```
        print("min\t%s\t%s" % (month, min_temp))
```

```
        print("max\t%s\t%s" % (month, max_temp))
```

```
    except:
```

```
        continue
```

Reducer.py:

```
#!/usr/bin/env python
```

```
import sys
```

```
min_temp = None
```

```
min_month = None
```

```
max_temp = None
```

```
max_month = None
```

```
for line in sys.stdin:
```

```
    line = line.strip()
```

```
    parts = line.split('\t')
```

```
    if len(parts) != 3:
```

```
        continue
```

```
    temp_type, month, temp = parts
```

```
    try:
```

```
        temp = float(temp)
```

```
    except:
```

```
        continue
```

```
    if temp_type == "min":
```

```
        if min_temp is None or temp < min_temp:
```

```
            min_temp = temp
```

```
            min_month = month
```

```
    elif temp_type == "max":
```

```
        if max_temp is None or temp > max_temp:
```

```
            max_temp = temp
```

```
            max_month = month
```

```
if min_temp is not None:
```

```
    print("Coldest Month: %s\t%.2f°C" % (min_month, min_temp))
```

```
if max_temp is not None:
```

```
print("Hottest Month: %s\t%.2f°C" % (max_month, max_temp))
```

Weatherdata.txt:

january,10,29

february,14,20

march,22,30

april,33,37

may,40,46

june,37,40

july,30,35

august,30,33

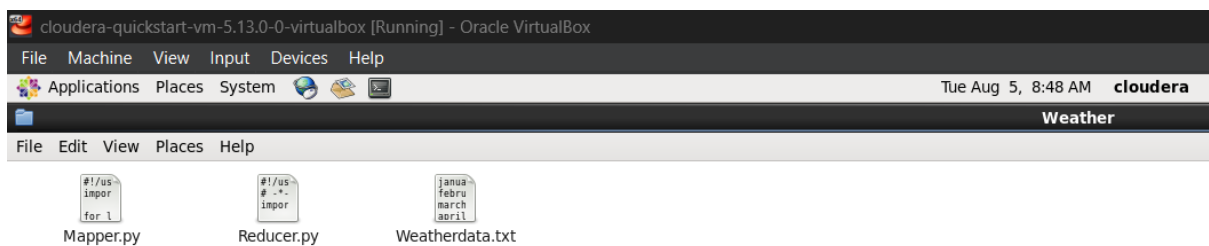
september,32,37

october,29,35

november,25,29

december,20,25

Execution:



```
[cloudera@quickstart Weather]$ cat Weatherdata.txt | ./Mapper.py | sort | ./Reducer.py
: No such file or directory
[cloudera@quickstart Weather]$ ls
Mapper.py Reducer.py Reducer.py~ Weatherdata.txt
[cloudera@quickstart Weather]$ chmod +x Mapper.py
[cloudera@quickstart Weather]$ chmod +x Reducer.py
[cloudera@quickstart Weather]$ cat Weatherdata.txt | ./Mapper.py | sort | ./Reducer.py
: No such file or directory
[cloudera@quickstart Weather]$ dos2unix Mapper.py
dos2unix: converting file Mapper.py to UNIX format ...
[cloudera@quickstart Weather]$ dos2unix Reducer.py
dos2unix: converting file Reducer.py to UNIX format ...
[cloudera@quickstart Weather]$ cat Weatherdata.txt | ./Mapper.py | sort | ./Reducer.py
Coldest Month: january 10.00°C
Hottest Month: may 46.00°C
```

```
[cloudera@quickstart Weather]$ hdfs dfs -mkdir -p /user/cloudera/weather_input
[cloudera@quickstart Weather]$ hdfs dfs -put -f Weatherdata.txt /user/cloudera/weather_input/
[cloudera@quickstart Weather]$ hdfs dfs -ls /user/cloudera/weather_input
Found 1 items
-rw-r--r-- 1 cloudera cloudera 170 2025-08-05 08:45 /user/cloudera/weather_input/Weatherdata.txt
[cloudera@quickstart Weather]$ hdfs dfs -cat /user/cloudera/weather_input/Weatherdata.txt
january,10,29
february,14,20
march,22,30
april,33,37
may,40,46
june,37,40
july,30,35
august,30,33
september,32,37
october,29,35
november,25,29
december,20,25
[cloudera@quickstart Weather]$ hdfs dfs -rm -r /user/cloudera/home/WeatherOutput
Deleted /user/cloudera/home/WeatherOutput
[cloudera@quickstart Weather]$ hdfs dfs -rm -r -skipTrash /user/cloudera/weather_output
Deleted /user/cloudera/weather_output
[cloudera@quickstart Weather]$ hadoop jar /usr/lib/hadoop-mapreduce/hadoop-streaming.jar \
> -input /user/cloudera/weather_input \
> -output /user/cloudera/weather_output \
> -mapper Mapper.py \
> -reducer Reducer.py \
> -file Mapper.py \
> -file Reducer.py
25/08/05 08:46:43 WARN streaming.StreamJob: -file option is deprecated, please use generic option -files
packageJobJar: [Mapper.py, Reducer.py] [/usr/lib/hadoop-mapreduce/hadoop-streaming-2.6.0-cdh5.13.0.jar]
25/08/05 08:46:44 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
25/08/05 08:46:44 INFO client.RMProxy: Connecting to ResourceManager at /0.0.0.0:8032
25/08/05 08:46:44 INFO mapred.FileInputFormat: Total input paths to process : 1
```

```

Total committed heap usage (bytes)=874512384
Shuffle Errors
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
File Input Format Counters
Bytes Read=255
File Output Format Counters
Bytes Written=60
25/08/05 08:46:57 INFO streaming.StreamJob: Output directory: /user/cloudera/weather_output
[cloudera@quickstart Weather]$ hdfs dfs -cat /user/cloudera/weather_output/part-00000
Coldest Month: january 10.00°C
Hottest Month: may 46.00°C
[cloudera@quickstart Weather]$ █
```