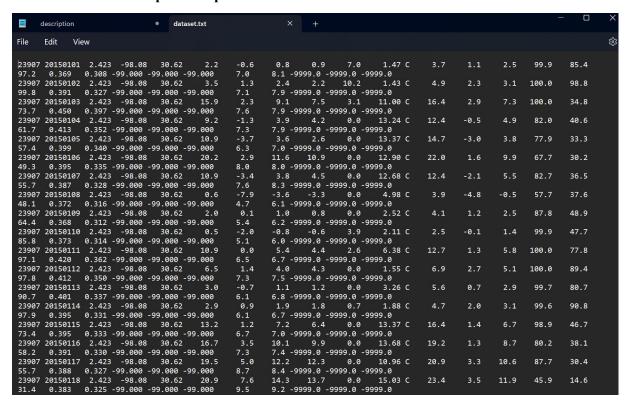
ROLL NUMBER: 210701107

EXP NO: 3 MAP REDUCE PROGRAM TO PROCESS A WEATHER DATASET

\$cd DA-Lab \$mkdir exp3 \$cd exp3 \$nano dataset.txt



\$nano mapper.py

```
Open > Fil

import sys

for line in sys.stdin:
    line = line.strip()
    #words = line.split()
    month = line[10:12]
    daily_max = line[38:45].strip()
    #daily_max = daily_max.strip()
    try:
        print ('%s\t%s' % (month ,daily_max))
    except ValueError:
        continue
```

ROLL NUMBER: 210701107

\$nano reducer.py

```
reducer_data.py
mapper_data.py
#from operator import itemgetter
import sys
current_month = None
current_max = 0
month = None
for line in sys.stdin:
    line = line.strip()
    month, daily_max = line.split('\t', 1)
       daily_max = float(daily_max)
    except ValueError:
        continue
    if current_month == month:
       if daily max > current max:
           current_max = daily_max
        if current_month:
           print ('%s\t%s' % (current_month, current_max))
       current_max = daily_max
       current month = month
# output of the last month
if current_month == month:
    print ('%s\t%s' % (current_month, current_max))
```

Sstart-all.sh

\$ jps

\$hdfs dfs -mkdir /exp3

\$hdfs dfs -copyFromLocal ~/DA-Lab/exp3/dataset.txt /exp2

\$chmod 777 mapper.py reducer.py

 $\label{lem:continuity} $$ \adoop jar $HADOOP_STREAMING -input/exp3/dataset.txt -output/exp3/output -mapper $$ \sim DA-Lab/exp3/mapper.py -reducer $$ \sim DA-Lab/exp3/reducer.py $$$

\$hdfs dfs -cat /exp3/output/*

```
hadoop@ubuntu:~$ nano mapper_data.py
hadoop@ubuntu:-$ cat dataset.txt | python3 mapper_data.py | sort -k1,1 | python3 reducer_da
ta.py
01
        26.5
02
        26.6
03
        29.1
04
        30.8
05
        31.1
06
        33.6
07
        38.5
08
        40.2
09
        36.5
10
        36.9
11
        27.6
hadoop@ubuntu:-$ hadoop jar /home/hadoop/hadoop/share/hadoop/tools/lib/hadoop-streaming-3.3
.6.jar -input /database/dataset.txt -output /database/new_outpu<mark>t</mark> -mapper /home/hadoop/mappe
r_data.py -reducer /home/hadoop/reducer_data.py
```