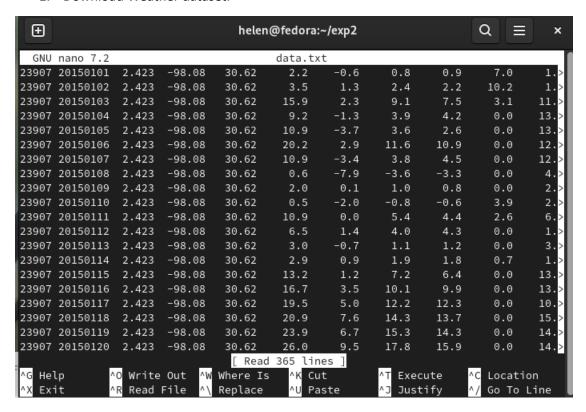
Exp. No: 3

Map Reduce program to process Weather dataset

1. Download Weather dataset.



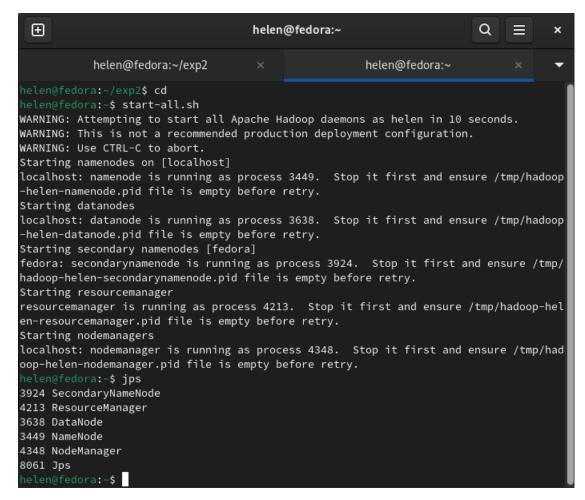
2. Create mapper.py program

```
GNU nano 7.2
                                          mapper.py
mport sys
 input comes from STDIN (standard input)
the mapper will get daily max temperature and group it by month. so output w>
(month,dailymax_temperature)
for line in sys.stdin:
        line = line.strip()
        words = line.split()
        month = line[10:12]
        daily_max = line[38:45]
        daily_max = daily_max.strip()
         for word in words:
                  print ('%s\t%s' % (month ,daily_max))
                 ^O Write Out
                                  ^W Where Is
                                                    ^K Cut
\G Help
                                                                      ^T Execute
                 ^R Read File
                                  ^\ Replace
                                                    ^U Paste
                                                                      ^J Justify
```

3. Create reducer.py

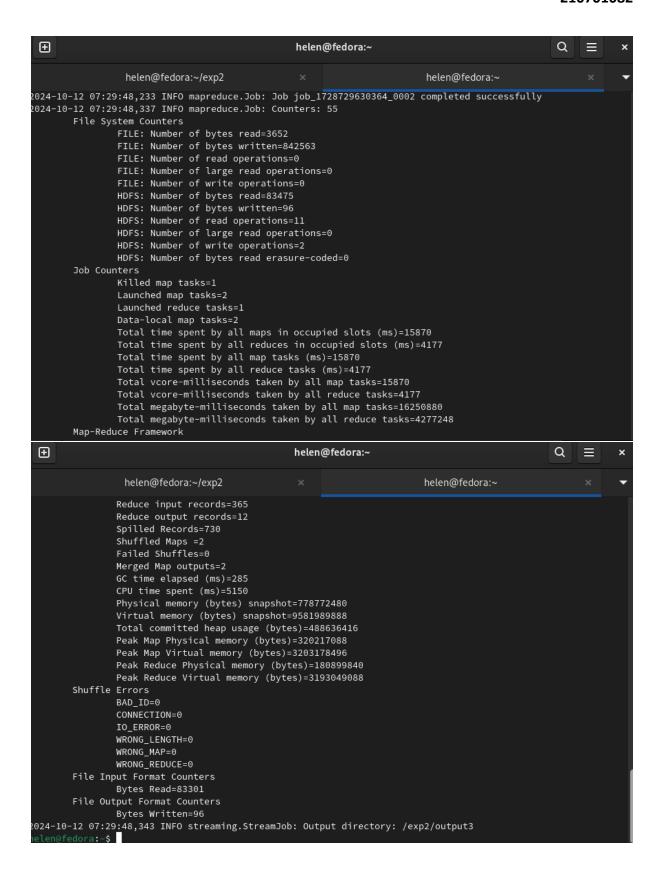
```
GNU nano 7.2
                                                                      Modified
                                    reducer.py
from operator import itemgetter
import sys
current_month = None
current_max = 0
month = None
for line in sys.stdin:
       line = line.strip()
        month, daily_max = line.split('\t', 1)
        try:
                daily_max = float(daily_max)
        except ValueError:
                continue
        if current_month == month:
                if daily_max > current_max:
                        current_max = daily_max
        else:
                if current_month:
                        print ('%s\t%s' % (current_month, current_max))
                current_max = daily_max
                current_month = month
if current_month == month:
        print ('%s\t%s' % (current_month, current_max))
^G Help
                              ^W Where Is
                                              ^K Cut
               ^O Write Out
                                                             ^T Execute
               ^R Read File
                                                                Justify
^X Exit
                                 Replace
                                                Paste
```

4. Start Hadoop services.



5. Run the Map reduce program using Hadoop Streaming.

```
\oplus
                                    helen@fedora:~
                                                                        Q
                                                                              〓
                                                                                    ×
           helen@fedora:~/exp2
                                                      helen@fedora:~
ieten@redora.~/expzș cu
nelen@fedora:~$ hadoop jar $HADOOP_STREAMING -input /exp2/data.txt -output /exp2/outp
ut3 -mapper ~/exp2/mapper.py -reducer ~/exp2/reducer.py
packageJobJar: [/tmp/hadoop-unjar6285464154385550498/] [] /tmp/streamjob1479607642965
193751.jar tmpDir=null
2024-10-12 07:29:18,538 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to
ResourceManager at /0.0.0.0:8032
2024-10-12 07:29:18,750 INFO client.DefaultNoHARMFailoverProxyProvider: Connecting to
ResourceManager at /0.0.0.0:8032
2024-10-12 07:29:19,201 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding
for path: /tmp/hadoop-yarn/staging/helen/.staging/job_1728729630364_0002
2024-10-12 07:29:19,717 INFO mapred.FileInputFormat: Total input files to process : oldsymbol{1}
2024-10-12 07:29:19,994 INFO mapreduce.JobSubmitter: number of splits:2
2024-10-12 07:29:20,746 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1
728729630364_0002
2024-10-12 07:29:20,746 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-10-12 07:29:21,006 INFO conf.Configuration: resource-types.xml not found
2024-10-12 07:29:21,006 INFO resource.ResourceUtils: Unable to find 'resource-types.x
ml'.
2024-10-12 07:29:21,187 INFO impl.YarnClientImpl: Submitted application application_1
728729630364_0002
2024-10-12 07:29:21,216 INFO mapreduce.Job: The url to track the job: http://fedora:8
088/proxy/application_1728729630364_0002/
2024-10-12 07:29:21,218 INFO mapreduce.Job: Running job: job_1728729630364_0002
2024-10-12 07:29:29,626 INFO mapreduce.Job: Job job_1728729630364_0002 running in ube
r mode : false
2024-10-12 07:29:29,627 INFO mapreduce.Job: map 0% reduce 0%
2024-10-12 07:29:39,796 INFO mapreduce.Job: map 100% reduce 0%
```



Output:

```
helen@fedora:~$ hdfs dfs -cat /exp2/output3/part-00000
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
10
        36.5
        36.9
11
        27.6
12
        25.9
```