

EXP NO: 3

MAP REDUCE PROGRAM TO PROCESS A WEATHER DATASET

Scd DA-Lab

\$mkdir exp3

Scd exp3

\$nano dataset.txt

```

File Actions Edit View Help
GNU nano 7.2 dataset.txt
210701101 2.423 -98.00 30.62 2.2 -8.6 0.0 0.9 1.47 C 3.7 1.1 2.5 69.9 83.4 97.2 0.209 0.208 -99.000 -99.000 -99.000 7.8 8.1 -9999.0 -9999.0 -9999.0
210701102 2.423 -98.00 30.62 3.5 1.3 2.4 2.2 10.2 1.43 C 4.9 2.3 3.1 100.0 99.8 99.6 0.391 0.327 -99.000 -99.000 -99.000 7.1 7.9 -9999.0 -9999.0 -9999.0
210701103 2.423 -98.00 30.62 15.9 2.3 9.1 7.5 3.1 11.00 C 16.4 2.9 7.3 100.0 34.8 73.7 0.450 0.397 -99.000 -99.000 -99.000 7.6 7.9 -9999.0 -9999.0 -9999.0
210701104 2.423 -98.00 30.62 9.2 -1.3 3.9 4.2 0.0 13.24 C 12.4 -8.5 4.9 82.0 48.6 61.7 0.413 0.252 -99.000 -99.000 -99.000 7.3 7.9 -9999.0 -9999.0 -9999.0
210701105 2.423 -98.00 30.62 10.9 -2.7 3.6 2.6 0.0 13.37 C 14.7 -3.0 3.8 77.9 33.3 57.4 0.399 0.340 -99.000 -99.000 -99.000 6.3 7.0 -9999.0 -9999.0 -9999.0
210701106 2.423 -98.00 30.62 20.2 2.9 11.6 10.9 0.0 12.90 C 22.0 1.6 9.9 67.7 10.2 49.3 0.395 0.335 -99.000 -99.000 -99.000 8.0 8.0 -9999.0 -9999.0 -9999.0
210701107 2.423 -98.00 30.62 10.9 -3.4 3.8 4.5 0.0 12.68 C 12.4 -2.1 3.5 82.7 16.3 55.7 0.387 0.328 -99.000 -99.000 -99.000 7.6 8.3 -9999.0 -9999.0 -9999.0
210701108 2.423 -98.00 30.62 0.6 -7.9 -3.6 -3.3 0.0 4.98 C 3.9 -4.8 -8.5 57.7 37.6 48.1 0.372 0.316 -99.000 -99.000 -99.000 4.7 6.1 -9999.0 -9999.0 -9999.0
210701109 2.423 -98.00 30.62 2.8 0.1 1.8 0.6 0.0 2.52 C 4.1 1.2 2.5 67.6 68.9 64.4 0.368 0.312 -99.000 -99.000 -99.000 5.4 6.2 -9999.0 -9999.0 -9999.0
210701110 2.423 -98.00 30.62 0.5 -2.8 -0.8 -0.6 1.9 2.11 C 7.5 -0.1 1.4 99.9 67.7 83.8 0.371 0.311 -99.000 -99.000 -99.000 5.1 6.8 -9999.0 -9999.0 -9999.0
210701111 2.423 -98.00 30.62 10.9 0.0 5.4 4.4 2.6 6.38 C 12.7 1.3 5.8 100.0 77.8 97.1 0.420 0.362 -99.000 -99.000 -99.000 6.5 6.7 -9999.0 -9999.0 -9999.0
210701112 2.423 -98.00 30.62 20.0 9.5 12.8 15.9 0.0 14.57 C 27.9 6.5 14.5 88.4 16.1 58.2 0.373 0.328 -99.000 -99.000 -99.000 10.9 10.4 -9999.0 -9999.0 -9999.0
210701113 2.423 -98.00 30.62 3.0 -0.7 1.1 1.2 0.0 3.26 C 5.6 0.7 2.9 99.7 80.7 90.7 0.401 0.337 -99.000 -99.000 -99.000 6.1 6.8 -9999.0 -9999.0 -9999.0
210701114 2.423 -98.00 30.62 2.9 0.9 1.9 1.8 0.7 1.88 C 4.7 2.0 3.1 99.6 90.8 97.9 0.395 0.321 -99.000 -99.000 -99.000 6.1 6.7 -9999.0 -9999.0 -9999.0
210701115 2.423 -98.00 30.62 13.2 1.2 7.2 6.4 0.0 13.37 C 10.4 1.4 6.7 98.9 46.7 73.4 0.395 0.333 -99.000 -99.000 -99.000 6.7 7.8 -9999.0 -9999.0 -9999.0
210701116 2.423 -98.00 30.62 10.7 3.5 10.1 9.9 0.0 13.68 C 19.2 1.3 8.7 80.2 38.1 58.2 0.391 0.330 -99.000 -99.000 -99.000 7.3 7.4 -9999.0 -9999.0 -9999.0
210701117 2.423 -98.00 30.62 19.5 5.0 12.2 12.3 0.0 10.96 C 20.9 3.3 10.6 87.7 10.4 55.7 0.388 0.327 -99.000 -99.000 -99.000 6.7 8.4 -9999.0 -9999.0 -9999.0
210701118 2.423 -98.00 30.62 20.9 7.6 14.3 13.7 0.0 15.03 C 21.4 3.5 11.9 45.9 14.6 31.4 0.383 0.325 -99.000 -99.000 -99.000 9.5 9.2 -9999.0 -9999.0 -9999.0
210701119 2.423 -98.00 30.62 23.9 6.7 15.3 14.3 0.0 14.10 C 25.6 3.8 12.6 65.3 26.8 45.6 0.376 0.321 -99.000 -99.000 -99.000 9.9 9.5 -9999.0 -9999.0 -9999.0
210701120 2.423 -98.00 30.62 20.0 9.5 12.8 15.9 0.0 14.57 C 27.9 6.5 14.5 88.4 16.1 58.2 0.373 0.328 -99.000 -99.000 -99.000 10.9 10.4 -9999.0 -9999.0 -9999.0
210701121 2.423 -98.00 30.62 11.0 6.9 6.9 8.9 1.7 2.71 C 13.1 6.8 9.7 99.2 68.0 88.1 0.369 0.317 -99.000 -99.000 -99.000 10.7 10.6 -9999.0 -9999.0 -9999.0
210701122 2.423 -98.00 30.62 8.6 3.5 6.1 5.6 40.0 1.28 C 9.1 4.1 6.3 99.6 95.2 98.0 0.546 0.418 -99.000 -99.000 -99.000 9.8 9.3 -9999.0 -9999.0 -9999.0
210701123 2.423 -98.00 30.62 9.4 2.2 3.8 4.2 7.5 6.58 C 11.1 2.0 4.3 98.4 58.0 56.5 0.554 0.409 -99.000 -99.000 -99.000 7.6 8.1 -9999.0 -9999.0 -9999.0
210701124 2.423 -98.00 30.62 16.0 1.4 8.7 8.0 0.0 14.26 C 18.8 0.4 7.7 92.0 33.0 63.0 0.494 0.381 -99.000 -99.000 -99.000 7.7 7.9 -9999.0 -9999.0 -9999.0
210701125 2.423 -98.00 30.62 20.2 6.4 13.3 12.7 0.0 14.99 C 22.0 4.4 11.0 69.2 18.9 43.8 0.456 0.357 -99.000 -99.000 -99.000 9.1 8.9 -9999.0 -9999.0 -9999.0
210701126 2.423 -98.00 30.62 21.5 7.2 14.4 14.1 0.0 12.41 C 22.9 5.5 12.2 56.8 23.7 48.6 0.423 0.349 -99.000 -99.000 -99.000 10.0 9.7 -9999.0 -9999.0 -9999.0
210701127 2.423 -98.00 30.62 26.5 10.7 18.6 17.5 0.0 15.18 C 28.9 8.1 15.5 52.2 21.4 38.8 0.420 0.344 -99.000 -99.000 -99.000 11.4 10.8 -9999.0 -9999.0 -9999.0
210701128 2.423 -98.00 30.62 20.2 13.3 19.8 19.1 0.0 15.11 C 28.1 7.9 16.2 54.9 19.4 35.5 0.410 0.339 -99.000 -99.000 -99.000 12.1 11.5 -9999.0 -9999.0 -9999.0
210701129 2.423 -98.00 30.62 23.1 9.8 10.5 10.4 0.0 13.74 C 27.4 9.7 16.4 87.0 34.2 55.6 0.442 0.334 -99.000 -99.000 -99.000 13.1 12.4 -9999.0 -9999.0 -9999.0
210701130 2.423 -98.00 30.62 13.0 6.9 10.8 9.0 0.2 7.19 C 19.2 8.3 11.0 67.6 48.4 58.8 0.389 0.328 -99.000 -99.000 -99.000 11.6 11.8 -9999.0 -9999.0 -9999.0
210701131 2.423 -98.00 30.62 15.1 7.4 11.3 10.2 0.5 1.18 C 14.5 8.4 10.7 100.0 63.1 98.3 0.401 0.327 -99.000 -99.000 -99.000 11.2 11.3 -9999.0 -9999.0 -9999.0
210701132 2.423 -98.00 30.62 10.3 3.9 11.1 13.3 0.0 8.69 C 22.1 4.1 13.8 98.8 53.6 79.1 0.450 0.386 -99.000 -99.000 -99.000 12.9 12.6 -9999.0 -9999.0 -9999.0
210701133 2.423 -98.00 30.62 8.0 -1.9 3.1 3.3 0.0 12.48 C 15.2 -0.6 5.8 69.4 34.8 54.2 0.420 0.354 -99.000 -99.000 -99.000 9.3 10.1 -9999.0 -9999.0 -9999.0
210701134 2.423 -98.00 30.62 9.3 2.3 3.8 3.8 0.0 2.69 C 8.3 3.9 5.7 100.0 65.1 82.6 0.400 0.363 -99.000 -99.000 -99.000 8.7 9.3 -9999.0 -9999.0 -9999.0
210701135 2.423 -98.00 30.62 11.8 4.3 8.1 7.9 0.3 4.41 C 13.8 5.5 8.8 100.0 80.6 94.1 0.410 0.344 -99.000 -99.000 -99.000 9.6 9.8 -9999.0 -9999.0 -9999.0
210701136 2.423 -98.00 30.62 9.4 0.7 5.0 3.1 0.0 4.98 C 9.3 2.8 5.6 97.3 68.8 81.6 0.402 0.336 -99.000 -99.000 -99.000 8.6 9.3 -9999.0 -9999.0 -9999.0
210701137 2.423 -98.00 30.62 15.3 8.8 8.0 7.4 0.0 14.67 C 20.9 1.8 9.3 95.9 50.7 75.9 0.596 0.331 -99.000 -99.000 -99.000 9.1 9.2 -9999.0 -9999.0 -9999.0
210701138 2.423 -98.00 30.62 19.8 5.9 12.8 12.0 0.0 16.75 C 23.8 4.8 12.0 96.8 48.6 76.2 0.393 0.329 -99.000 -99.000 -99.000 18.5 10.2 -9999.0 -9999.0 -9999.0
210701139 2.423 -98.00 30.62 20.3 12.3 19.3 17.7 0.0 16.94 C 29.8 10.8 22.5 96.6 25.4 65.2 0.393 0.332 -99.000 -99.000 -99.000 13.0 12.2 -9999.0 -9999.0 -9999.0
210701140 2.423 -98.00 30.62 20.6 13.6 20.1 19.3 0.0 16.90 C 31.1 10.3 17.8 64.6 19.7 48.0 0.388 0.329 -99.000 -99.000 -99.000 13.5 12.9 -9999.0 -9999.0 -9999.0
210701141 2.423 -98.00 30.62 26.0 12.7 19.4 18.7 0.0 17.89 C 30.3 9.5 17.6 61.3 26.5 48.1 0.381 0.325 -99.000 -99.000 -99.000 13.6 13.0 -9999.0 -9999.0 -9999.0
210701142 2.423 -98.00 30.62 22.0 10.6 10.6 10.3 0.0 16.58 C 29.2 8.1 10.6 77.6 29.9 55.8 0.375 0.323 -99.000 -99.000 -99.000 14.0 13.5 -9999.0 -9999.0 -9999.0
210701143 2.423 -98.00 30.62 14.8 5.6 10.2 9.4 0.0 12.68 C 22.8 6.8 11.9 71.2 32.3 52.2 0.363 0.317 -99.000 -99.000 -99.000 12.3 12.7 -9999.0 -9999.0 -9999.0
210701144 2.423 -98.00 30.62 18.9 4.7 11.8 11.5 0.0 12.25 C 26.3 4.5 12.3 65.6 37.7 50.7 0.355 0.310 -99.000 -99.000 -99.000 11.4 11.7 -9999.0 -9999.0 -9999.0
210701145 2.423 -98.00 30.62 26.3 10.8 17.5 16.7 0.0 16.94 C 31.3 8.8 17.1 77.8 29.1 56.9 0.354 0.310 -99.000 -99.000 -99.000 13.3 12.8 -9999.0 -9999.0 -9999.0
210701146 2.423 -98.00 30.62 20.2 13.3 18.8 15.9 0.0 5.79 C 23.9 12.0 16.2 96.0 65.3 83.9 0.353 0.312 -99.000 -99.000 -99.000 14.5 13.9 -9999.0 -9999.0 -9999.0
210701147 2.423 -98.00 30.62 17.6 8.0 8.8 5.0 1.0 4.87 C 12.1 2.5 7.4 90.3 66.6 83.8 0.345 0.306 -99.000 -99.000 -99.000 11.9 12.8 -9999.0 -9999.0 -9999.0
210701148 2.423 -98.00 30.62 11.2 -0.4 5.4 4.3 0.0 13.36 C 18.6 1.3 6.8 80.6 39.6 57.9 0.337 0.298 -99.000 -99.000 -99.000 9.3 10.3 -9999.0 -9999.0 -9999.0
210701149 2.423 -98.00 30.62 17.2 -0.4 8.4 9.1 0.0 18.48 C 25.1 0.5 10.1 80.9 21.1 46.1 0.334 0.294 -99.000 -99.000 -99.000 9.5 10.0 -9999.0 -9999.0 -9999.0

```

\$nano mapper.py

```

File Actions Edit View Help
GNU nano 7.2 mapper.py
#!/usr/bin/python3
import sys
for line in sys.stdin: # remove leading and trailing whitespace
    line = line.strip() # split the line into words
    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))

```

\$nano reducer.py

```

File Actions Edit View Help
GNU nano 7.2 reducer.py
#!/usr/bin/python3
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))

```

\$start-all.sh

```

(hadoop@kali)~$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as hadoop in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [kali]
Picked up _JAVA_OPTIONS: -Dawt.useSystemAAFontSettings=on -Dswing.aatext=true
2024-09-11 04:59:16,429 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
Starting resourcemanager
Starting nodemanagers

```

\$ jps

```

(hadoop@kali)~$ jps
Picked up _JAVA_OPTIONS: -Dawt.useSystemAAFontSettings=on -Dswing.aatext=true
14436 NodeManager
16772 Jps
13830 SecondaryNameNode
14311 ResourceManager
13597 DataNode
13471 NameNode

```

\$hdfs dfs -mkdir /exp3

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

```

(hadoop@kali)~/hadoop/bin$ ./hdfs dfs -ls /exp3
Picked up _JAVA_OPTIONS: -Dawt.useSystemAAFontSettings=on -Dswing.aatext=true
2024-09-21 00:11:13,818 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform...
Found 3 items
-rw-r--r-- 1 hadoop supergroup 79205 2024-08-29 10:50 /exp3/dataset.txt
drwxr-xr-x - hadoop supergroup 0 2024-08-29 10:52 /exp3/new_output
drwxr-xr-x - hadoop supergroup 0 2024-09-13 01:00 /exp3/output

```

\$chmod 777 mapper.py reducer.py

\$hadoop jar \$HADOOP_STREAMING -input /exp3/dataset.txt -output /exp3/output -mapper ~/DA-Lab/exp3/mapper.py -reducer ~/DA-Lab/exp3/reducer.py

```

(hadoop@kali)~/hadoop/bin$ hadoop jar $HADOOP_STREAMING -input /exp3/dataset.txt -output /exp3/output -mapper ~/DA-Lab/exp3/mapper.py -reducer ~/DA-Lab/exp3/reducer.py
Picked up _JAVA_OPTIONS: -Dawt.useSystemAAFontSettings=on -Dswing.aatext=true
2024-09-21 00:13:19,993 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
packageJobJar: [/tmp/hadoop-unjar383659404787382099/] [] /tmp/streamjob2158010624070613243.jar tmpDir=null
2024-09-21 00:13:20,918 INFO client.DefaultHohARMFailoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2024-09-21 00:13:21,223 INFO client.DefaultHohARMFailoverProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2024-09-21 00:13:27,216 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1726891437845_0001
2024-09-21 00:13:28,262 INFO mapred.FileInputFormat: Total input files to process : 1
2024-09-21 00:13:28,365 INFO mapreduce.JobSubmitter: number of splits:2
2024-09-21 00:13:28,613 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1726891437845_0001
2024-09-21 00:13:28,613 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-09-21 00:13:29,230 INFO conf.Configuration: resource-types.xml not found
2024-09-21 00:13:29,230 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2024-09-21 00:13:29,895 INFO impl.YarnClientImpl: Submitted application application_1726891437845_0001
2024-09-21 00:13:29,993 INFO mapreduce.Job: The url to track the job: http://kali:8088/proxy/application_1726891437845_0001/
2024-09-21 00:13:29,998 INFO mapreduce.Job: Running job: job_1726891437845_0001
2024-09-21 00:13:43,554 INFO mapreduce.Job: Job job_1726891437845_0001 running in uber mode : false
2024-09-21 00:13:43,560 INFO mapreduce.Job: map 0% reduce 0%
2024-09-21 00:13:52,918 INFO mapreduce.Job: map 100% reduce 0%
2024-09-21 00:14:00,992 INFO mapreduce.Job: map 100% reduce 100%
2024-09-21 00:14:01,012 INFO mapreduce.Job: Job job_1726891437845_0001 completed successfully
2024-09-21 00:14:01,189 INFO mapreduce.Job: Counters: 54
File System Counters
  FILE: Number of bytes read=102094
  FILE: Number of bytes written=1138411
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=83481
  HDFS: Number of bytes written=96
  HDFS: Number of read operations=11
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=2
  HDFS: Number of bytes read erasure-coded=0
Job Counters
  Launched map tasks=2
  Launched reduce tasks=1
  Data-local map tasks=2
  Total time spent by all maps in occupied slots (ms)=14691
  Total time spent by all reduces in occupied slots (ms)=4696
  Total time spent by all map tasks (ms)=14691
  Total time spent by all reduce tasks (ms)=4696
  Total vcore-millisecods taken by all map tasks=14691
  Total megabyte-millisecods taken by all map tasks=4606
  Total megabyte-millisecods taken by all map tasks=15043584
  Total megabyte-millisecods taken by all reduce tasks=4808704
Map-Reduce Framework
  Map input records=365
  Map output records=10220
  Map output bytes=81648
  Map output materialized bytes=102100
  Input split bytes=180

```

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

```
(hadoop@kali)-[~/hadoop/bin]
$ ./hdfs dfs -cat /exp3/output/*
Picked up _JAVA_OPTIONS: -Dawt.useSystemAAFontSettings=on -Dswing.aatext=true
2024-09-21 00:15:38,966 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform...
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
12      25.9
```