

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
21070150101	2,423	-98,00	30,62	2,2
21070150102	2,423	-98,00	30,62	3,5
21070150103	2,423	-98,00	30,62	15,9
21070150104	2,423	-98,00	30,62	9,2
21070150105	2,423	-98,00	30,62	10,9
21070150106	2,423	-98,00	30,62	20,2
21070150107	2,423	-98,00	30,62	10,9
21070150108	2,423	-98,00	30,62	0,6
21070150109	2,423	-98,00	30,62	2,8
21070150110	2,423	-98,00	30,62	0,5
21070150111	2,423	-98,00	30,62	10,9
21070150112	2,423	-98,00	30,62	0,5
21070150113	2,423	-98,00	30,62	0,0
21070150114	2,423	-98,00	30,62	2,9
21070150115	2,423	-98,00	30,62	13,2
21070150116	2,423	-98,00	30,62	10,7
21070150117	2,423	-98,00	30,62	19,5
21070150118	2,423	-98,00	30,62	20,9
21070150119	2,423	-98,00	30,62	23,9
21070150120	2,423	-98,00	30,62	20,0
21070150121	2,423	-98,00	30,62	11,0
21070150122	2,423	-98,00	30,62	8,6
21070150123	2,423	-98,00	30,62	9,4
21070150124	2,423	-98,00	30,62	16,0
21070150125	2,423	-98,00	30,62	20,2
21070150126	2,423	-98,00	30,62	21,5
21070150127	2,423	-98,00	30,62	26,5
21070150128	2,423	-98,00	30,62	20,2
21070150129	2,423	-98,00	30,62	23,1
21070150130	2,423	-98,00	30,62	13,0
21070150131	2,423	-98,00	30,62	15,1
21070150201	2,423	-98,00	30,62	10,3
21070150202	2,423	-98,00	30,62	8,0
21070150203	2,423	-98,00	30,62	9,3
21070150204	2,423	-98,00	30,62	11,8
21070150205	2,423	-98,00	30,62	9,4
21070150206	2,423	-98,00	30,62	15,3
21070150207	2,423	-98,00	30,62	19,8
21070150208	2,423	-98,00	30,62	20,3
21070150209	2,423	-98,00	30,62	20,6
21070150210	2,423	-98,00	30,62	26,0
21070150211	2,423	-98,00	30,62	22,0
21070150212	2,423	-98,00	30,62	14,8
21070150213	2,423	-98,00	30,62	18,9
21070150214	2,423	-98,00	30,62	26,3
21070150215	2,423	-98,00	30,62	20,2
21070150216	2,423	-98,00	30,62	17,6
21070150217	2,423	-98,00	30,62	11,2
21070150218	2,423	-98,00	30,62	17,2

\$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-unjar383059404787382099/] [] /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-milliseconds taken by all map tasks=14691
  Total vcore-milliseconds taken by all reduce tasks=4696
  Total megabyte-milliseconds taken by all map tasks=15043584
  Total megabyte-milliseconds 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
```