

EXP NO: 3

MAP REDUCE PROGRAM TO PROCESS A WEATHER DATASET

Scd DA-Lab

\$mkdir exp3

Scd exp3

\$nano dataset.txt

The screenshot shows a terminal window with a large dataset of weather data. The data is organized into columns with headers like 'date', 'temp', 'humidity', 'wind', etc. The rows contain numerical values representing weather metrics over time.

\$nano mapper.py

```

File Actions Edit View Help
GNU nano 7.2
from operator import itemgetter
import sys
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))

```

\$nano reducer.py

```

File Actions Edit View Help
GNU nano 7.2
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
2024-09-21 04:59:16,479 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
14436 NodeManager
16772 Jps
13830 SecondaryNameNode
14311 ResourceManager
13397 DataNode
13471 NameNode

```

\$hdfs dfs -mkdir /exp3

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

```

(hadoop@kali)~$ hdfs dfs -ls /exp3
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 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 00:13:59,903 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2024-09-21 00:13:59,904 INFO client.DefaultHadoopFailoverProxyProvider: Connecting to ResourceManager at /B.0.0.0:8032
2024-09-21 00:13:59,904 INFO client.DefaultHadoopFailoverProxyProvider: Connecting to ResourceManager at /B.0.0.0:8032
2024-09-21 00:13:59,904 INFO mapreduce.JobSubmitter: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1726891437845_0001
2024-09-21 00:13:59,904 INFO mapreduce.JobSubmitter: Total input files to process : 1
2024-09-21 00:13:59,904 INFO mapreduce.JobSubmitter: number of splits:2
2024-09-21 00:13:59,904 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1726891437845_0001
2024-09-21 00:13:59,904 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-09-21 00:13:59,904 INFO conf.Configuration: resource-types.xml not found
2024-09-21 00:13:59,904 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'
2024-09-21 00:13:59,904 INFO impl.YarnClientImpl: Submitted application application_1726891437845_0001
2024-09-21 00:13:59,904 INFO mapreduce.Job: The url to track the job: http://knl13hnaa.proxy/application_1726891437845_0001/
2024-09-21 00:13:59,904 INFO mapreduce.Job: Running job: job_1726891437845_0001
2024-09-21 00:13:59,904 INFO mapreduce.Job: Job job_1726891437845_0001 running in sher mode = false
2024-09-21 00:13:59,904 INFO mapreduce.Job: map 0% reduce 0%
2024-09-21 00:13:59,904 INFO mapreduce.Job: map 100% reduce 0%
2024-09-21 00:14:00,904 INFO mapreduce.Job: map 100% reduce 100%
2024-09-21 00:14:01,904 INFO mapreduce.Job: Job job_1726891437845_0001 completed successfully
2024-09-21 00:14:01,904 INFO mapreduce.Job: Counters: 54
File System Counters
  FILE: Number of bytes read=102894
  FILE: Number of bytes written=128413
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=83461
  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)=6496
  Total time spent by all map tasks (ms)=14691
  Total time spent by all reduce tasks (ms)=6496
  Total vcore-milliseconds taken by all map tasks=14691
  Total vcore-milliseconds taken by all reduce tasks=6496
  Total megabyte-milliseconds taken by all map tasks=15841584
  Total megabyte-milliseconds taken by all reduce tasks=4688784
Map-Reduce Framework
  Map input records=105
  Map output records=10228
  Map output bytes=81668
  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
```