

EXP NO: 3

MAP REDUCE PROGRAM TO PROCESS A WEATHER DATASET

Snano dataset.txt

Date	Location	Temp	Humidity	Wind	Cloud	Pressure	Visibility	UV	PM2.5	PM10	Ozone	CO2	SO2	NO2	CO	CH4	NH3	H2S	H2O	Other						
23/07/20151116	2.423	-96.88	30.62	22.5	17.1	19.8	19.5	0.3	1.56	C	21.2	17.4	19.3	-9999.0	-9999.0	-9999.0	0.432	0.400	-99.000	-99.000	-99.000	18.4	18.0	-9999.0	-9999.0	-9999.0
23/07/20151117	2.423	-96.88	30.62	22.7	10.1	16.4	15.5	7.3	7.53	C	21.2	9.7	16.0	-9999.0	-9999.0	-9999.0	0.467	0.448	-99.000	-99.000	-99.000	17.2	17.7	-9999.0	-9999.0	-9999.0
23/07/20151118	2.423	-96.88	30.62	21.8	7.6	14.4	14.4	0.0	16.83	C	26.4	6.6	14.2	-9999.0	-9999.0	-9999.0	0.443	0.413	-99.000	-99.000	-99.000	15.0	15.6	-9999.0	-9999.0	-9999.0
23/07/20151119	2.423	-96.88	30.62	24.6	11.9	18.2	16.9	0.0	15.70	C	29.1	8.6	16.3	-9999.0	-9999.0	-9999.0	0.426	0.403	-99.000	-99.000	-99.000	15.2	15.6	-9999.0	-9999.0	-9999.0
23/07/20151120	2.423	-96.88	30.62	23.2	12.6	17.9	17.1	0.0	14.74	C	28.4	10.8	17.4	-9999.0	-9999.0	-9999.0	0.416	0.391	-99.000	-99.000	-99.000	16.2	16.2	-9999.0	-9999.0	-9999.0
23/07/20151121	2.423	-96.88	30.62	17.5	2.5	10.0	11.4	0.0	15.07	C	21.2	4.4	14.1	-9999.0	-9999.0	-9999.0	0.402	0.384	-99.000	-99.000	-99.000	15.1	15.9	-9999.0	-9999.0	-9999.0
23/07/20151122	2.423	-96.88	30.62	13.4	0.2	6.8	6.1	0.0	16.16	C	21.9	1.6	8.8	-9999.0	-9999.0	-9999.0	0.383	0.377	-99.000	-99.000	-99.000	11.9	13.2	-9999.0	-9999.0	-9999.0
23/07/20151123	2.423	-96.88	30.62	16.3	3.1	9.7	9.3	0.0	15.50	C	23.2	1.8	10.7	-9999.0	-9999.0	-9999.0	0.376	0.368	-99.000	-99.000	-99.000	11.4	12.3	-9999.0	-9999.0	-9999.0
23/07/20151124	2.423	-96.88	30.62	16.6	7.3	11.9	12.4	0.0	4.91	C	19.2	7.4	13.1	-9999.0	-9999.0	-9999.0	0.375	0.360	-99.000	-99.000	-99.000	13.3	13.4	-9999.0	-9999.0	-9999.0
23/07/20151125	2.423	-96.88	30.62	20.3	14.9	17.6	17.6	0.5	3.65	C	22.0	14.8	17.4	-9999.0	-9999.0	-9999.0	0.377	0.354	-99.000	-99.000	-99.000	16.0	15.6	-9999.0	-9999.0	-9999.0
23/07/20151126	2.423	-96.88	30.62	21.6	18.2	19.9	19.8	14.6	2.21	C	22.6	17.1	19.5	-9999.0	-9999.0	-9999.0	0.402	0.361	-99.000	-99.000	-99.000	17.9	17.3	-9999.0	-9999.0	-9999.0
23/07/20151127	2.423	-96.88	30.62	20.7	2.6	11.6	12.7	21.4	1.10	C	20.8	5.9	14.6	-9999.0	-9999.0	-9999.0	0.543	0.481	-99.000	-99.000	-99.000	17.3	17.4	-9999.0	-9999.0	-9999.0
23/07/20151128	2.423	-96.88	30.62	4.5	2.2	3.3	3.4	23.4	1.55	C	8.0	4.6	6.1	-9999.0	-9999.0	-9999.0	0.605	0.508	-99.000	-99.000	-99.000	10.9	11.8	-9999.0	-9999.0	-9999.0
23/07/20151129	2.423	-96.88	30.62	6.2	3.4	4.8	4.6	3.6	2.00	C	18.1	5.2	7.0	-9999.0	-9999.0	-9999.0	0.546	0.470	-99.000	-99.000	-99.000	18.8	11.7	-9999.0	-9999.0	-9999.0
23/07/20151130	2.423	-96.88	30.62	11.0	5.1	8.0	7.3	0.0	5.69	C	16.4	6.8	9.5	-9999.0	-9999.0	-9999.0	0.520	0.451	-99.000	-99.000	-99.000	11.4	11.9	-9999.0	-9999.0	-9999.0
23/07/20151201	2.423	-96.88	30.62	13.8	5.1	9.5	8.7	0.0	9.05	C	19.6	6.2	10.3	-9999.0	-9999.0	-9999.0	0.485	0.421	-99.000	-99.000	-99.000	11.6	12.0	-9999.0	-9999.0	-9999.0
23/07/20151202	2.423	-96.88	30.62	17.7	5.1	11.4	10.7	0.0	14.85	C	22.5	4.6	18.9	-9999.0	-9999.0	-9999.0	0.464	0.400	-99.000	-99.000	-99.000	11.4	11.8	-9999.0	-9999.0	-9999.0
23/07/20151203	2.423	-96.88	30.62	16.2	5.4	10.8	10.0	0.0	14.80	C	22.3	4.0	10.5	-9999.0	-9999.0	-9999.0	0.445	0.398	-99.000	-99.000	-99.000	11.0	11.5	-9999.0	-9999.0	-9999.0
23/07/20151204	2.423	-96.88	30.62	15.8	5.5	10.7	9.9	0.0	14.70	C	22.1	3.4	10.1	-9999.0	-9999.0	-9999.0	0.431	0.383	-99.000	-99.000	-99.000	10.5	11.0	-9999.0	-9999.0	-9999.0
23/07/20151205	2.423	-96.88	30.62	16.8	4.3	10.6	10.1	0.0	14.32	C	22.1	3.7	10.6	-9999.0	-9999.0	-9999.0	0.421	0.372	-99.000	-99.000	-99.000	10.6	11.0	-9999.0	-9999.0	-9999.0
23/07/20151206	2.423	-96.88	30.62	18.6	7.4	13.0	12.2	0.0	14.11	C	24.1	6.4	12.3	-9999.0	-9999.0	-9999.0	0.416	0.364	-99.000	-99.000	-99.000	11.5	11.7	-9999.0	-9999.0	-9999.0
23/07/20151207	2.423	-96.88	30.62	19.4	7.3	13.4	12.6	0.0	14.78	C	24.8	5.4	12.3	-9999.0	-9999.0	-9999.0	0.409	0.357	-99.000	-99.000	-99.000	11.7	11.9	-9999.0	-9999.0	-9999.0
23/07/20151208	2.423	-96.88	30.62	23.8	8.2	16.0	14.9	0.0	12.20	C	26.4	6.3	13.5	-9999.0	-9999.0	-9999.0	0.402	0.354	-99.000	-99.000	-99.000	12.4	12.3	-9999.0	-9999.0	-9999.0
23/07/20151209	2.423	-96.88	30.62	22.6	10.0	16.3	16.1	0.0	12.88	C	27.0	7.6	14.9	-9999.0	-9999.0	-9999.0	0.388	0.355	-99.000	-99.000	-99.000	12.3	12.2	-9999.0	-9999.0	-9999.0
23/07/20151210	2.423	-96.88	30.62	25.9	13.9	19.9	18.0	0.0	11.16	C	28.3	11.1	16.8	-9999.0	-9999.0	-9999.0	0.395	0.352	-99.000	-99.000	-99.000	14.6	14.2	-9999.0	-9999.0	-9999.0
23/07/20151211	2.423	-96.88	30.62	22.7	11.4	17.0	17.2	0.0	9.87	C	26.1	9.4	16.4	-9999.0	-9999.0	-9999.0	0.391	0.348	-99.000	-99.000	-99.000	14.8	14.5	-9999.0	-9999.0	-9999.0
23/07/20151212	2.423	-96.88	30.62	22.0	19.4	20.7	20.4	1.2	1.60	C	21.1	17.9	19.3	-9999.0	-9999.0	-9999.0	0.394	0.351	-99.000	-99.000	-99.000	17.4	16.5	-9999.0	-9999.0	-9999.0
23/07/20151213	2.423	-96.88	30.62	22.1	8.7	15.4	11.7	31.7	6.84	C	29.1	8.1	12.4	-9999.0	-9999.0	-9999.0	0.565	0.460	-99.000	-99.000	-99.000	14.2	14.4	-9999.0	-9999.0	-9999.0
23/07/20151214	2.423	-96.88	30.62	24.1	7.4	15.7	15.0	0.0	14.63	C	25.1	5.3	13.0	-9999.0	-9999.0	-9999.0	0.498	0.417	-99.000	-99.000	-99.000	12.9	12.9	-9999.0	-9999.0	-9999.0
23/07/20151215	2.423	-96.88	30.62	22.6	12.8	17.7	16.9	0.0	5.28	C	24.2	9.7	15.6	-9999.0	-9999.0	-9999.0	0.473	0.397	-99.000	-99.000	-99.000	14.5	14.2	-9999.0	-9999.0	-9999.0
23/07/20151216	2.423	-96.88	30.62	15.0	5.8	10.4	11.2	0.0	10.91	C	20.9	5.9	12.6	-9999.0	-9999.0	-9999.0	0.452	0.388	-99.000	-99.000	-99.000	13.0	13.6	-9999.0	-9999.0	-9999.0
23/07/20151217	2.423	-96.88	30.62	17.7	3.1	10.4	9.6	0.0	14.54	C	21.5	2.3	9.8	-9999.0	-9999.0	-9999.0	0.431	0.380	-99.000	-99.000	-99.000	11.1	11.8	-9999.0	-9999.0	-9999.0
23/07/20151218	2.423	-96.88	30.62	13.4	0.0	6.7	6.1	0.0	14.49	C	20.6	0.9	8.1	-9999.0	-9999.0	-9999.0	0.416	0.375	-99.000	-99.000	-99.000	10.0	10.8	-9999.0	-9999.0	-9999.0
23/07/20151219	2.423	-96.88	30.62	22.6	10.0	16.3	16.1	0.0	12.88	C	27.0	7.6	14.9	-9999.0	-9999.0	-9999.0	0.388	0.355	-99.000	-99.000	-99.000	12.3	12.2	-9999.0	-9999.0	-9999.0
23/07/20151220	2.423	-96.88	30.62	17.8	10.1	13.9	14.3	0.0	3.77	C	17.5	10.1	14.2	-9999.0	-9999.0	-9999.0	0.409	0.363	-99.000	-99.000	-99.000	12.9	12.6	-9999.0	-9999.0	-9999.0
23/07/20151221	2.423	-96.88	30.62	23.2	14.6	18.9	18.4	0.0	8.88	C	26.6	9.4	17.0	-9999.0	-9999.0	-9999.0	0.411	0.359	-99.000	-99.000	-99.000	15.1	14.6	-9999.0	-9999.0	-9999.0
23/07/20151222	2.423	-96.88	30.62	25.0	11.2	18.1	18.0	0.0	13.64	C	28.4	7.4	16.0	-9999.0	-9999.0	-9999.0	0.402	0.351	-99.000	-99.000	-99.000	14.0	13.8	-9999.0	-9999.0	-9999.0
23/07/20151223	2.423	-96.88	30.62	15.4	4.1	10.2	9.9	0.0	13.54	C	22.0	2.5	10.2	-9999.0	-9999.0	-9999.0	0.409	0.360	-99.000	-99.000	-99.000	10.5	10.8	-9999.0	-9999.0	-9999.0
23/07/20151224	2.423	-96.88	30.62	23.1	14.4	18.8	18.0	0.0	13.53	C	29.4	10.2	17.1	-9999.0	-9999.0	-9999.0	0.388	0.346	-99.000	-99.000	-99.000	14.7	14.5	-9999.0	-9999.0	-9999.0
23/07/20151225	2.423	-96.88	30.62	19.1	9.3	14.2	14.8	0.0	6.05	C	23.4	9.5	16.1	-9999.0	-9999.0	-9999.0	0.383	0.343	-99.000	-99.000	-99.000	14.8	14.7	-9999.0	-9999.0	-9999.0
23/07/20151226	2.423	-96.88	30.62	23.6	18.7	21.2	21.6	0.6	4.03	C	25.4	18.5	20.9	-9999.0	-9999.0	-9999.0	0.387	0.350	-99.000	-99.000	-99.000	18.0	17.0	-9999.0	-9999.0	-9999.0
23/07/20151227	2.423	-96.88	30.62	21.7	9.4	11.0	7.0	18.4	1.56	C	20.2	3.0	9.4	-9999.0	-9999.0	-9999.0	0.456	0.408	-99.000	-99.000	-99.000	14.0	14.9	-9999.0	-9999.0	-9999.0
23/07/20151228	2.423	-96.88	30.62	4.3	-0.1	2.1	1.6	0.0	5.26	C	10.4	1.4	4.2	-9999.0	-9999.0	-9999.0	0.486	0.437	-99.000	-99.000	-99.000	9.1	10.6	-9999.0	-9999.0	-9999.0
23/07/20151229	2.423	-96.88	30.62	9.6	-0.1	4.8	4.1	0.0	8.19	C	16.3	1.2	6.8													

\$nano reducer.py

```

Oct 20 10:44 PM
hadoop@fedora:~/hadoop/experiments
reducer.py

GNU nano 7.2
#!/usr/bin/python

from operator import itemgetter
import sys

#reducer will get the input from stdout which will be a collection of key, value(Key=month ,
#value= daily max temperature)
#reducer logic: will get all the daily max temperature for a month and find max temperature
#for the month
#shuffle will ensure that key are sorted(month)
current_month = None
current_max = 0
month = None

# input comes from STDIN
for line in sys.stdin:
    # remove leading and trailing whitespace
    line = line.strip()
    # parse the input we got from mapper.py
    month, daily_max = line.split('\t', 1)

    # convert daily_max (currently a string) to float
    try:
        daily_max = float(daily_max)
    except ValueError:
        # daily_max was not a number, so silently
        # ignore/discard this line
        continue

    # this IF-switch only works because Hadoop shuffle process sorts map output
    # by key (here: month) before it is passed to the reducer
    if current_month == month:
        if daily_max > current_max:
            current_max = daily_max
    else:
        if current_month:
            # write result to STDOUT
            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))

```

```

(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 /dataset.txt -output /output -mapper

mapper.py -reducer reducer.py

```

hadoop@fedora:~/hadoop/experiments$ hadoop jar ../share/hadoop/tools/lib/hadoop-streaming-3.4.0.jar -input /dataset.txt -output /out -mapper $HOME/hadoop/experiments/mapper.py -reducer $HOME/hadoop/experiments/reducer.py
packageJobJar: [/tmp/hadoop-unjar1800474811278489987/] [] /tmp/streamjob17676603820645521190.jar tmpDir=null
2024-08-30 00:52:09,047 INFO client.DefaultHARMAFollowerProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2024-08-30 00:52:09,240 INFO client.DefaultHARMAFollowerProxyProvider: Connecting to ResourceManager at /0.0.0.0:8032
2024-08-30 00:52:09,500 INFO mapreduce.JobResourceUploader: Disabling Erasure Coding for path: /tmp/hadoop-yarn/staging/hadoop/.staging/job_1724955478554_0006
2024-08-30 00:52:09,809 INFO mapred.FileInputFormat: Total input files to process : 1
2024-08-30 00:52:10,710 INFO mapreduce.JobSubmitter: number of splits:2
2024-08-30 00:52:11,295 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1724955478554_0006
2024-08-30 00:52:11,295 INFO mapreduce.JobSubmitter: Executing with tokens: []
2024-08-30 00:52:11,476 INFO conf.Configuration: resource-types.xml not found
2024-08-30 00:52:11,476 INFO resource.ResourceUtils: Unable to find 'resource-types.xml'.
2024-08-30 00:52:11,552 INFO impl.YarnClientImpl: Submitted application application_1724955478554_0006
2024-08-30 00:52:11,583 INFO mapreduce.Job: The url to track the job: http://fedora:8088/proxy/application_1724955478554_0006/
2024-08-30 00:52:11,585 INFO mapreduce.Job: Running job: job_1724955478554_0006
2024-08-30 00:52:17,711 INFO mapreduce.Job: Job job_1724955478554_0006 running in uber mode : false
2024-08-30 00:52:17,712 INFO mapreduce.Job: map 0% reduce 0%
2024-08-30 00:52:23,827 INFO mapreduce.Job: map 100% reduce 0%
2024-08-30 00:52:29,891 INFO mapreduce.Job: map 100% reduce 100%
2024-08-30 00:52:29,898 INFO mapreduce.Job: Job job_1724955478554_0006 completed successfully
2024-08-30 00:52:30,033 INFO mapreduce.Job: Counters: 54

```

\$hdfs dfs -cat /out/*

```
hadoop@fedora:~/hadoop/experiments$ hdfs dfs -cat /out/*
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
```