

210701120

## Map Reduce program to process a weather dataset

Input(dataset.txt):

Open

dataset.txt

dataset.txt

~/Downloads

mapper\_data.py

reducer\_data.py

reducemapexecute.txt

23907	20150101	2.423	-98.08	30.62	2.2	-0.6	0.8	0.9	7.0	1.47	C	3.7	1.1	2.5
99.9	85.4	97.2	0.369	0.308	-99.000	-99.000	-99.000	7.0	8.1	-9999.0	-9999.0	-9999.0		
23907	20150102	2.423	-98.08	30.62	3.5	1.3	2.4	2.2	10.2	1.43	C	4.9	2.3	3.1
100.0	98.8	99.8	0.391	0.327	-99.000	-99.000	-99.000	7.1	7.9	-9999.0	-9999.0	-9999.0		
23907	20150103	2.423	-98.08	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		
23907	20150104	2.423	-98.08	30.62	9.2	-1.3	3.9	4.2	0.0	13.24	C	12.4	-0.5	4.9
82.0	40.6	61.7	0.413	0.352	-99.000	-99.000	-99.000	7.3	7.9	-9999.0	-9999.0	-9999.0		
23907	20150105	2.423	-98.08	30.62	10.9	-3.7	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		
23907	20150106	2.423	-98.08	30.62	20.2	2.9	11.6	10.9	0.0	12.90	C	22.0	1.6	9.9
67.7	30.2	49.3	0.395	0.335	-99.000	-99.000	-99.000	8.0	8.0	-9999.0	-9999.0	-9999.0		
23907	20150107	2.423	-98.08	30.62	10.9	-3.4	3.8	4.5	0.0	12.68	C	12.4	-2.1	5.5
82.7	36.5	55.7	0.387	0.328	-99.000	-99.000	-99.000	7.6	8.3	-9999.0	-9999.0	-9999.0		
23907	20150108	2.423	-98.08	30.62	0.6	-7.9	-3.6	-3.3	0.0	4.98	C	3.9	-4.8	-0.5
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		
23907	20150109	2.423	-98.08	30.62	2.0	0.1	1.0	0.8	0.0	2.52	C	4.1	1.2	2.5
87.8	48.9	64.4	0.368	0.312	-99.000	-99.000	-99.000	5.4	6.2	-9999.0	-9999.0	-9999.0		
23907	20150110	2.423	-98.08	30.62	0.5	-2.0	-0.8	-0.6	3.9	2.11	C	2.5	-0.1	1.4
99.9	47.7	85.8	0.373	0.314	-99.000	-99.000	-99.000	5.1	6.0	-9999.0	-9999.0	-9999.0		
23907	20150111	2.423	-98.08	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		
23907	20150112	2.423	-98.08	30.62	6.5	1.4	4.0	4.3	0.0	1.55	C	6.9	2.7	5.1
100.0	89.4	97.8	0.412	0.350	-99.000	-99.000	-99.000	7.3	7.5	-9999.0	-9999.0	-9999.0		
23907	20150113	2.423	-98.08	30.62	3.0	-0.7	1.1	1.2	0.0	3.26	C	5.6	0.7	2.9

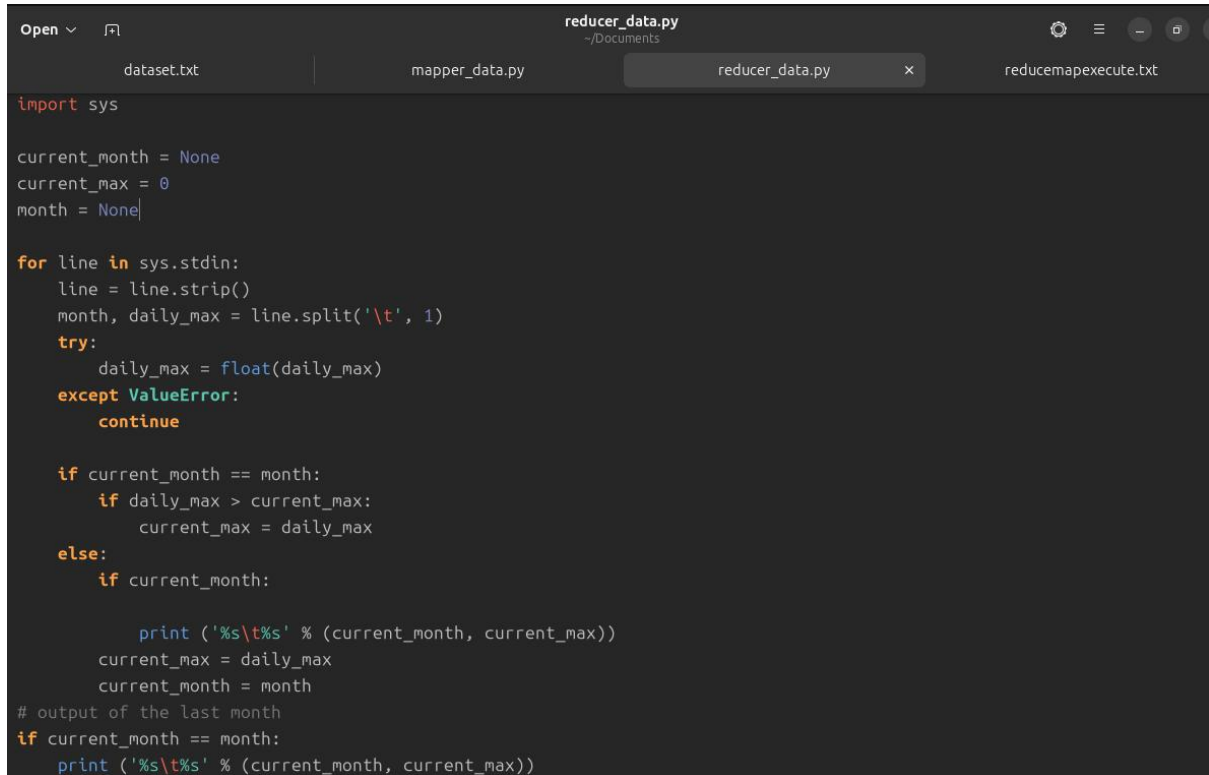
mapper\_data.py:

```
import sys

for line in sys.stdin:
    line = line.strip()
    #words = line.split()
    month = line[10:12]
    daily_max = line[38:45].strip()
    #daily_max = daily_max.strip()
    try:
        print('%s\t%s' % (month, daily_max))
    except ValueError:
        continue
```

reducer\_data.py:

210701120

A screenshot of a code editor window with a dark theme. The title bar shows 'reducer\_data.py' and the path '~/Documents'. There are four tabs open: 'dataset.txt', 'mapper\_data.py', 'reducer\_data.py' (active), and 'reducemapexecute.txt'. The code in the active tab is a Python script for a reducer. It imports sys, initializes current\_month to None, current\_max to 0, and month to None. It then enters a loop reading from sys.stdin. Each line is stripped and split by a tab character. A try-except block attempts to convert the second part of the split line to a float. If successful, it checks if the current month matches the input month. If it does, it checks if the input daily\_max is greater than the current\_max and updates it. If the month changes, it prints the current month and max, resets current\_max to the new daily\_max, and updates current\_month. After the loop, it prints the final month and max. The code is as follows:

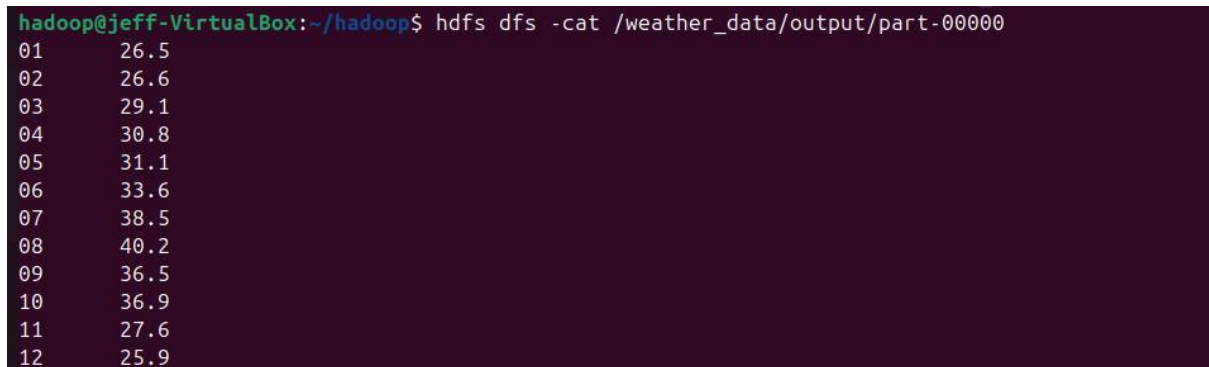
```
import sys

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
# output of the last month
if current_month == month:
    print ('%s\t%s' % (current_month, current_max))
```

## Output:

A screenshot of a terminal window with a dark background. The prompt is 'hadoop@jeff-VirtualBox:~/hadoop\$'. The command entered is 'hdfs dfs -cat /weather\_data/output/part-00000'. The output shows 12 lines of data, each with a line number and a temperature value.

```
hadoop@jeff-VirtualBox:~/hadoop$ hdfs dfs -cat /weather_data/output/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      36.5
10      36.9
11      27.6
12      25.9
```