

EXPERIMENT-5

AIM: Write a MapReduce to analyse weather data set and print whether the day is sunny or cool day

```
weather_reducer.py (~/Desktop/20010396/weather) - gedit
File Edit View Search Tools Documents Help
Open Save Undo
weather_reducer.py
#!/usr/bin/env python
import sys

counts={}

for line in sys.stdin:
    data = line.split()
    year = data[0]
    temp = float(data[1])

    if year in counts.keys():
        if temp<=25.0:
            counts[year][0] += 1
        else:
            counts[year][1] += 1
    else:
        counts[year] = [0, 0]
        if temp<=25.0:
            counts[year][0] += 1
        else:
            counts[year][1] += 1

for key in sorted(counts):
    print "In year {0}, there were {1} cold days and {2} sunny days.".format(key, counts[key][0], counts[key][1])
```

```
weather_mapper.py (~/Desktop/20010396/weather) - gedit
File Edit View Search Tools Documents Help
Open Save Undo
weather_mapper.py
#!/usr/bin/env python
import sys

cnt=0

for line in sys.stdin:
    if cnt==0:
        cnt=1
        continue

    data = line.strip().split(',')
    year = data[0].split('-')[-1]
    temp = data[1]

    if len(temp) != 0:
        print "{0} {1}".format(year, temp)
```

Creating Program files

```
File Edit View Search Terminal Help
[cloudera@quickstart 20010396]$ cd
[cloudera@quickstart ~]$ touch weather_mapper.py
[cloudera@quickstart ~]$ touch weather_reducer.py
[cloudera@quickstart ~]$
```

```
-rw-rwxr-x 1 cloudera cloudera 0 Feb 26 06:45 weather_mapper.py
-rw-rwxr-x 1 cloudera cloudera 0 Feb 26 06:45 weather_reducer.py
[cloudera@quickstart ~]$ chmod go+x weather_mapper.py
[cloudera@quickstart ~]$ chmod go+x weather_reducer.py
[cloudera@quickstart ~]$ ls -l
```

Granting execution permission

• Copying file from local to HDFS

```
[cloudera@quickstart ~]$ hdfs dfs -put /home/cloudera/Desktop/20010396/weather/weather_Bhubhreshwar_1990_2022.csv /user/cloudera/20010396/exp5/weather_output
```

```
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
File Input Format Counters
  Bytes Read=415903
File Output Format Counters
  Bytes Written=1960
```

- 23/02/26 06:21:54 INFO streaming.StreamJob: Output directory: /user/cloudera/20010396/exp5/weather output

```
[cloudera@quickstart 20010396]$ hadoop jar /usr/lib/hadoop-mapreduce/hadoop-streaming.jar \
> -input /user/cloudera/20010396/exp5/weather_Bhubhreshwar_1990_2022.csv \
> -output /user/cloudera/20010396/exp5/weather_output \
> -mapper /home/cloudera/Desktop/20010396/weather/weather_mapper.py \
> -reducer /home/cloudera/Desktop/20010396/weather/weather_reducer.py
packageJobJar: [] [/usr/lib/hadoop-mapreduce/hadoop-streaming-2.6.0-cdh5.13.0.jar] /tmp/streamjob6910765683807646897.jar tmpDir=null
23/02/26 06:20:51 INFO client.RMPProxy: Connecting to ResourceManager at /0.0.0.0:8032
23/02/26 06:20:52 INFO client.RMPProxy: Connecting to ResourceManager at /0.0.0.0:8032
23/02/26 06:20:53 INFO mapred.FileInputFormat: Total input paths to process : 1
23/02/26 06:20:53 INFO mapreduce.JobSubmitter: number of splits:2
23/02/26 06:20:54 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1677416662241_0004
23/02/26 06:20:54 INFO impl.YarnClientImpl: Submitted application application_1677416662241_0004
23/02/26 06:20:54 INFO mapreduce.Job: The url to track the job: http://quickstart.cloudera:8088/proxy/application_1677416662241_0004
23/02/26 06:20:54 INFO mapreduce.Job: Running job: job_1677416662241_0004
23/02/26 06:21:01 INFO mapreduce.Job: Job job_1677416662241_0004 running in uber mode : false
23/02/26 06:21:01 INFO mapreduce.Job:  map 0% reduce 0%
23/02/26 06:21:36 INFO mapreduce.Job:  map 33% reduce 0%
23/02/26 06:21:37 INFO mapreduce.Job:  map 50% reduce 0%
23/02/26 06:21:40 INFO mapreduce.Job:  map 100% reduce 0%
23/02/26 06:21:49 INFO mapreduce.Job:  map 100% reduce 100%
23/02/26 06:21:53 INFO mapreduce.Job: Job job_1677416662241_0004 completed successfully
```

Program execution using Hadoop-streaming.jar

- Displaying the output in the terminal

```
[cloudera@quickstart 20010396]$ hdfs dfs -cat /user/cloudera/20010396/exp5/weather_output/part-000000
In year 1990, there were 90 cold days and 262 sunny days.
In year 1991, there were 103 cold days and 260 sunny days.
In year 1992, there were 100 cold days and 263 sunny days.
In year 1993, there were 92 cold days and 271 sunny days.
In year 1994, there were 99 cold days and 261 sunny days.
In year 1995, there were 109 cold days and 254 sunny days.
In year 1996, there were 104 cold days and 259 sunny days.
In year 1997, there were 84 cold days and 278 sunny days.
In year 1998, there were 87 cold days and 276 sunny days.
In year 1999, there were 92 cold days and 259 sunny days.
In year 2000, there were 98 cold days and 263 sunny days.
In year 2001, there were 93 cold days and 269 sunny days.
In year 2002, there were 102 cold days and 261 sunny days.
In year 2003, there were 102 cold days and 255 sunny days.
In year 2004, there were 104 cold days and 261 sunny days.
In year 2005, there were 100 cold days and 265 sunny days.
In year 2006, there were 97 cold days and 267 sunny days.
In year 2007, there were 95 cold days and 267 sunny days.
In year 2008, there were 110 cold days and 256 sunny days.
In year 2009, there were 75 cold days and 288 sunny days.
In year 2010, there were 82 cold days and 283 sunny days.
In year 2011, there were 99 cold days and 261 sunny days.
In year 2012, there were 95 cold days and 271 sunny days.
In year 2013, there were 113 cold days and 252 sunny days.
In year 2014, there were 109 cold days and 256 sunny days.
In year 2015, there were 80 cold days and 285 sunny days.
In year 2016, there were 86 cold days and 280 sunny days.
In year 2017, there were 91 cold days and 274 sunny days.
In year 2018, there were 97 cold days and 268 sunny days.
In year 2019, there were 102 cold days and 263 sunny days.
In year 2020, there were 106 cold days and 260 sunny days.
In year 2021, there were 88 cold days and 277 sunny days.
In year 2022, there were 51 cold days and 196 sunny days.
```

Submitted By:

Name: Sohan Choudhury

Regd. No.: 20010396

Branch: CSE (AI & ML)

Group: 4 (H)