

Module 4: Hadoop MapReduce Framework - II

Assignment

edureka!

edureka!

© 2014 Brain4ce Education Solutions Pvt. Ltd.

Module 4: Hadoop MapReduce Framework - II

Assignment – Find the Hot and Cold Days

Table of Contents

Introduction	2
Problem statement	2

edureka!

Introduction

Apply your MapReduce programming knowledge and write a MapReduce program to process a dataset with temperature records. You need to find the Hot and Cold days in a year based on the maximum and minimum temperatures on those days.

The dataset for this problem is the ‘**WeatherData**’ records file available in your LMS. This dataset has been taken from [National Climatic Data Center \(NCDC\)](#) public datasets. You can download more datasets from this FTP site and review the [README](#) file to understand the available datasets.

Problem statement

Let’s understand the problem through a subset of records in the dataset as shown in the following figure:

FIGURE 1-1 WEATHER RECORDS

Line	Station ID	Year	Month	Day	Max Temp	Min Temp	Other 1	Other 2	Other 3	Other 4	Other 5	Unit
10	25380	2013	01	10	2.514	-135.69	58.43	-0.9	-2.8	-1.8	-1.6	1.7
11	25380	2013	01	11	2.514	-135.69	58.43	0.1	-1.2	-0.5	-0.4	3.0
12	25380	2013	01	12	2.514	-135.69	58.43	0.3	0.0	0.2	0.1	3.0
13	25380	2013	01	13	2.514	-135.69	58.43	4.4	0.2	2.3	0.9	7.2
14	25380	2013	01	14	2.514	-135.69	58.43	5.4	4.3	4.9	4.9	11.4
15	25380	2013	01	15	2.514	-135.69	58.43	5.0	-0.1	2.5	2.5	24.1
16	25380	2013	01	16	2.514	-135.69	58.43	2.9	0.0	1.5	1.5	17.5
17	25380	2013	01	17	2.514	-135.69	58.43	4.9	0.4	2.7	3.5	13.4
18	25380	2013	01	18	2.514	-135.69	58.43	2.1	-2.1	0.0	0.2	1.7
19	25380	2013	01	19	2.514	-135.69	58.43	0.5	-2.9	-1.2	-1.0	0.0
20	25380	2013	01	20	2.514	-135.69	58.43	0.6	-1.3	-0.3	-0.2	10.0
21	25380	2013	01	21	2.514	-135.69	58.43	2.1	0.5	1.3	1.1	11.7
22	25380	2013	01	22	2.514	-135.69	58.43	2.7	-0.4	1.2	1.1	5.4
23	25380	2013	01	23	2.514	-135.69	58.43	4.5	0.4	2.5	2.6	0.7
24	25380	2013	01	24	2.514	-135.69	58.43	4.0	-0.4	1.8	2.4	0.0
25	25380	2013	01	25	2.514	-135.69	58.43	3.7	-0.7	1.5	1.5	0.8
26	25380	2013	01	26	2.514	-135.69	58.43	3.2	-1.4	0.9	1.7	3.9
27	25380	2013	01	27	2.514	-135.69	58.43	-0.4	-8.3	-4.3	-2.8	16.0
28	25380	2013	01	28	2.514	-135.69	58.43	-8.3	-17.1	-12.7	-12.9	0.6

Your task is to find out the dates with maximum temperature greater than 40 (**A Hot Day**) and minimum temperature lower than 10 (**A Cold Day**).

Here is the sample output:

FIGURE 1- 2 SAMPLE OUTPUT

04-02-2013	Cold Day
04-03-2013	Cold Day
04-04-2013	Cold Day
04-05-2013	Cold Day
04-06-2013	Hot Day
04-07-2013	Hot Day
04-08-2013	Hot Day
04-09-2013	Hot Day
05-01-2013	Cold Day
05-03-2013	Cold Day
05-04-2013	Cold Day

You can review the solution in your LMS.

edureka!