

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 text file. You need to print the count of number of occurrences of each word in the text file.

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 patent records as shown in the following figure:

FIGURE 1-1 WEATHER RECORDS

Line	Station ID	Date	Pressure	Humidity	Wind Speed	Wind Dir	Wind Gust	Temp	Dew Point	Clouds	Visibility	Unit
10	25380	20130110	2.514	-135.69	58.43	-0.9	-2.8	-1.8	-1.6	1.7	0.19	C
11	25380	20130111	2.514	-135.69	58.43	0.1	-1.2	-0.5	-0.4	3.0	0.09	C
12	25380	20130112	2.514	-135.69	58.43	0.3	0.0	0.2	0.1	3.0	0.15	C
13	25380	20130113	2.514	-135.69	58.43	4.4	0.2	2.3	0.9	7.2	0.23	C
14	25380	20130114	2.514	-135.69	58.43	5.4	4.3	4.9	4.9	11.4	0.03	C
15	25380	20130115	2.514	-135.69	58.43	5.0	-0.1	2.5	2.5	24.1	0.15	C
16	25380	20130116	2.514	-135.69	58.43	2.9	0.0	1.5	1.5	17.5	0.05	C
17	25380	20130117	2.514	-135.69	58.43	4.9	0.4	2.7	3.5	13.4	0.34	C
18	25380	20130118	2.514	-135.69	58.43	2.1	-2.1	0.0	0.2	1.7	0.28	C
19	25380	20130119	2.514	-135.69	58.43	0.5	-2.9	-1.2	-1.0	0.0	1.38	C
20	25380	20130120	2.514	-135.69	58.43	0.6	-1.3	-0.3	-0.2	10.0	0.33	C
21	25380	20130121	2.514	-135.69	58.43	2.1	0.5	1.3	1.1	11.7	0.65	C
22	25380	20130122	2.514	-135.69	58.43	2.7	-0.4	1.2	1.1	5.4	0.29	C
23	25380	20130123	2.514	-135.69	58.43	4.5	0.4	2.5	2.6	0.7	0.39	C
24	25380	20130124	2.514	-135.69	58.43	4.0	-0.4	1.8	2.4	0.0	0.76	C
25	25380	20130125	2.514	-135.69	58.43	3.7	-0.7	1.5	1.5	0.8	0.63	C
26	25380	20130126	2.514	-135.69	58.43	3.2	-1.4	0.9	1.7	3.9	1.35	C
27	25380	20130127	2.514	-135.69	58.43	-0.4	-8.3	-4.3	-2.8	16.0	1.10	C
28	25380	20130128	2.514	-135.69	58.43	-8.3	-17.1	-12.7	-12.9	0.6	1.72	C

Maximum
Temperature

Minimum
Temperature

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!