Module 4: Hadoop MapReduce Framework - II

Assignment

Version 2.0

Module 4: Hadoop MapReduce Framework - II

Assignment – Find the Hot and Cold Days

Table of Contents

Introduction

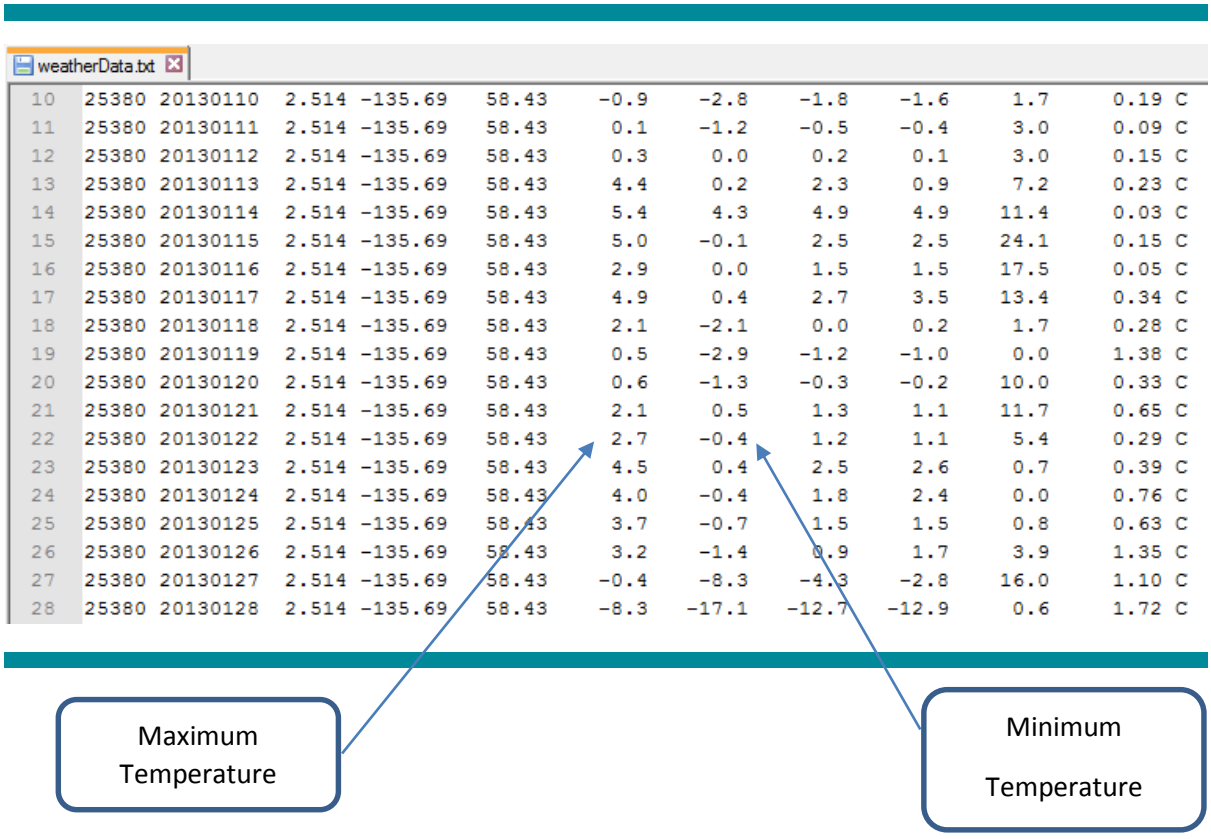
Apply your MapReduce programming knowledge and write a MapReduce program to process a text file. You need to print the count of number of occurrences of each word in the text file.

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:

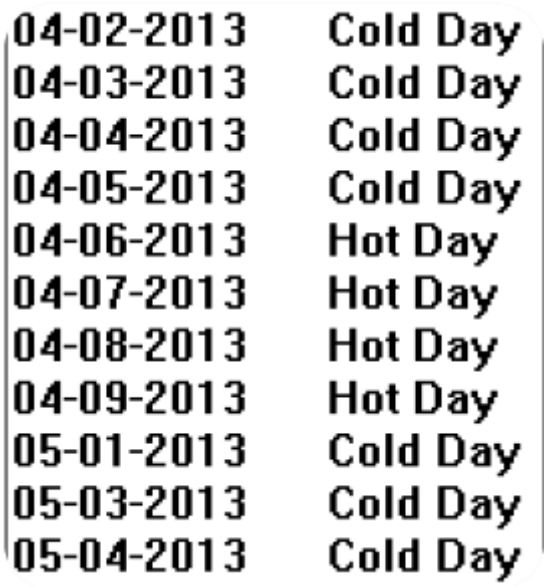
FIGURE 1-1 WEATHER RECORDS



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



You can review the solution in your LMS.