Module 3: Hadoop MapReduce Framework

Assignment – Find the Hot and Cold Days

Module 3: Hadoop MapReduce Framework

Assignment – Find the Hot and Cold Days

Table of Contents

Introduction

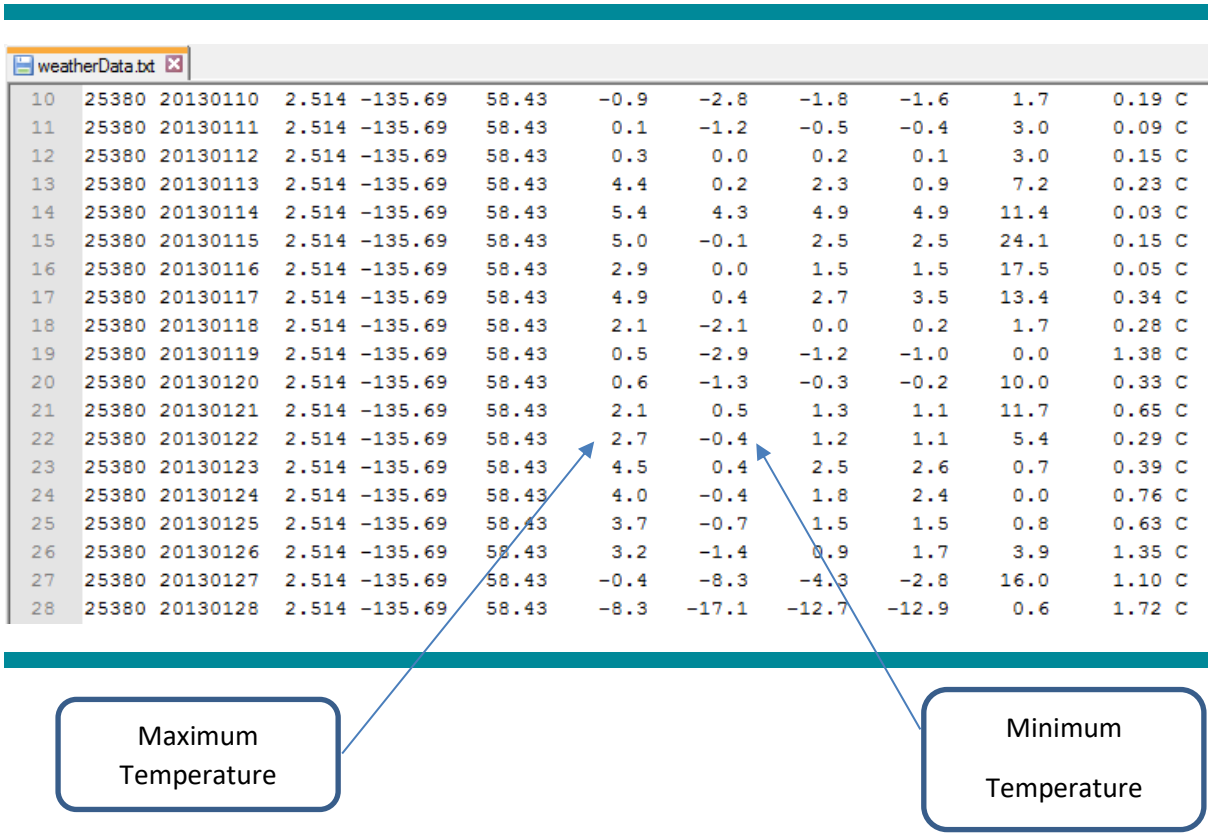
Apply your MapReduce programming knowledge and write a MapReduce program to process a dataset with temperature records. You need to find the Hot and Cold days in a year based on the maximum and minimum temperatures on those days.

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:

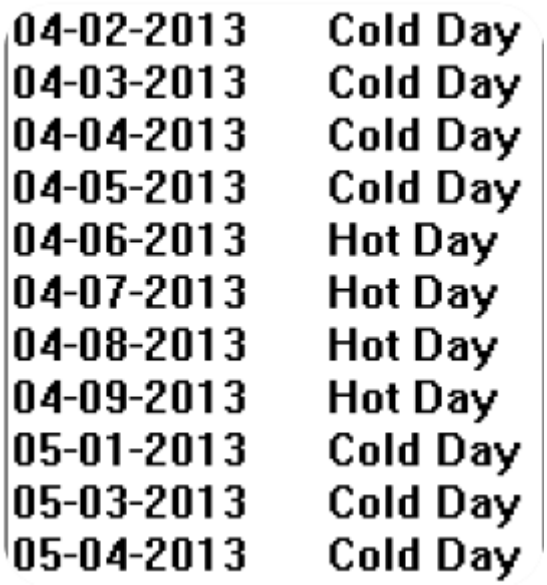
FIGURE 1-1 WEATHER RECORDS



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.