BDA2 - Spark SQL - Exercises

Assignments

Redo the exercises BDA1 using Spark SQL. An introduction of Spark SQL can be found here (https://www.ida.liu.se/~732A54/lab/SparkSQLQuickIntro.pdf.)

There are two ways to write queries in Spark SQL - using built-in API functions or running SQL-like queries. **To pass this lab, you need to:**

 (1) use built-in API functions for all the questions
 (2) write a regular SQL query for the second question

For each question include the following data in the report and sort it as shown:

1.
year, station with the max, maxValue ORDER BY maxValue DESC
year, station with the min, minValue ORDER BY minValue DESC

2.
year, month, value ORDER BY value DESC
year, month, value ORDER BY value DESC

3.
year, month, station, avgMonthlyTemperature ORDER BY avgMonthlyTemperature DESC

4.
station, maxTemp, maxDailyPrecipitation ORDER BY station DESC
Note: The correct result for this question should be empty.

5.
year, month, avgMonthlyPrecipitation ORDER BY year DESC, month DESC

6.
year, month, difference ORDER BY year DESC, month DESC