

# BDA2 - Spark SQL - Exercises

## Assignments

Redo the exercises BDA1 using Spark SQL whenever possible.

There are two ways to write queries in SparkSQL - using built-in API functions or running SQL-like queries. **To pass this lab, you need to use built-in API functions for all the questions, in addition, SQL-like queries for the second question. The slides of this link (<https://www.ida.liu.se/~732A54/lab/SparkSQLQuickIntro.pdf>) show some examples of Spark SQL.**

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

1.

year, station with the max, max**Value** ORDER BY max**Value** DESC

year, station with the min, min**Value** ORDER BY min**Value** 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 which means there are no stations satisfy the conditions.

5.

year, month, avgMonthlyPrecipitation ORDER BY year DESC, month DESC