VE472 — Methods and tools for big data

Homework 4

Manuel — UM-JI (Summer 2021)

Reminders

- Write in a neat and legible handwriting or use LaTeX
- Clearly explain the reasoning process
- Write in a complete style (subject, verb, and object)
- Be critical on your results

To complete this assignment download the `weather.tar` *archive from the course server, using the usual credentials.*

Ex. 1 — *Reminders on database*

1. Explain what is a `Join` operation, and describe its most common types.
2. What is an aggregate operation?
3. Write at least three advanced nested queries on the weather database.

Ex. 2 — *Holidays!*

After a long semester of hard work you want to go on holidays in any place, as long as the weather is nice. Therefore you download the weather data from the year 2017 and decide to analyse it using the knowledge from VE472.

1. Define what is "perfect weather" according to you. Express it in terms of *precipitations*, *average temperature*, and *daily temperature amplitude*.
2. Using Drill, with or without R, determine the perfect location of your next holidays.

Note: the dataset being small Drill can be run in standalone mode, i.e. no need for a Hadoop cluster.

Ex. 3 — *Data visualisation*

Use Drill to retrieve the information, then plot the results in R or using GNUplot.

1. Plot the temperature variation for each continent.
2. Plot the average temperature for each continent.

Note: the file `country_continent.txt` contains a list of the countries with the continent they belong to.