Big Data Infrastructures
SQL Review

Review what we saw during the lab. session, install PostgreSQL in your computer if you haven't done it already, and complete the following tasks.

A. Create a table `position(timestamp, lat, lon, alt)` that logs all the GPS coordinates of a person together with the altitude in meters above sea level.
The field "`timestamp`" has to be of type `TIMESTAMP` and you can use usual numeric data types for the other fields.

   i. Populate it with 100,000 reasonable random records; you can do it either with a script/stored procedure, or with your favorite programming language… or by hand, if you are foolish enough ☺

   ii. Choose a rectangular area (for example, all points with latitude in [60, 90] and longitude in [20, 44]) and write a query that selects all the records referring to points in that area.

   iii. Once you are sure that your query is correct, check that PostgreSQL behaves correctly if an index can be exploited to speed-up the execution of your query. In order to do that:
      a. Describe the execution plan of the query and report its execution time by interpreting the result of the statement '`EXPLAIN ANALYZE <query>`';
      b. Create two indices, one on the field `lat` and another on the field `long` by using a `CREATE INDEX` statement;
      c. Re-check the execution plan of the query and highlight the differences between the two plans and the two execution times.
      d. [OPTIONAL: what is a "bitmap index scan"?].

B. With reference to the relational schema introduced last week, write SQL statements to complete the following tasks. Don't use the `CONTAINS` operator, avoid returning duplicate results, try to use the `DISTINCT` clause only when you really need it, and try to use aggregate operators only when they are really needed.

   i. Select the names of all the authors that published at least one solo-paper (that is, a paper with only one author).

   ii. Select the emails of all the a-social authors, that is, all authors that submits only solo-papers.

   iii. Select the affiliation(s) that has *written* the greatest number of papers.

   iv. Select the name of the author(s) who published the greatest number of papers; also show the maximum number of papers published by the selected authors.

   v. Select the IDs of the authors who write papers always with the same co-authors (notice that this does not imply that the co-authors cannot write papers with somebody else).

The solutions of the tasks will be available on ILIAS on Thursday, October the 13th. Feel free to contact me by email if you have any doubt or if you want to send me your answers to have some feedback (alberto@exascale.info).

Good luck ☺