Big Data Infrastructures
SQL Review

Review what we saw during the lab. session, install PostgreSQL in your computer, and complete the following tasks by using SQL statements. Also, hand in a snapshot of the relations produced as a result

A. Create a new database.

B. Create the following tables: choose the right type for each field and justify the choice of referential actions (cascade, restrict, update, set null, set default) for foreign keys.

Paper	(paperID, title, abstract)
Author	(authorID, name, email, affiliation)
Conference	(confID, name, ranking)
Writes	(authorID, paperID)
Submits	(paperID, confID, isAccepted, date)
Cites	(paperIDfrom, paperIDto)

C. Use your favorite programming language to connect to the database and populate it with random data. Use a search engine to find PostgreSQL drivers or API for the language you chose.
For python, you can use psycopg2 module:
https://www.tutorialspoint.com/postgresql/postgresql_python.htm
For C/C++, you can use **libpqxx** library:
https://www.tutorialspoint.com/postgresql/postgresql_c_cpp.htm
For JAVA: https://www.tutorialspoint.com/postgresql/postgresql_java.htm

D. With reference to the schema just introduced, write SQL code 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 number of authors belonging to each affiliation.

   ii. Choose an `authorID` from Author and select all the abstracts of the papers written (or co-written) by the author with the selected `authorID`.

   iii. Create a view `PublishesIn(authorID, confID)` containing all the pairs `(authorID, confID)` such that there is at least one publication by `authorID` accepted at `confID`.

   iv. Select the IDs of all the authors that cited themselves.

   v. Select the title of all the papers *co-authored* by the author selected in ii. (the papers written only by the selected authors do not have to be part of the result).

The solutions of the tasks will be available on ILIAS on Thursday, October 4th. 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 (akbwaj@exascale.info)

Good luck ☺