

Advanced Databases TU856/TU857/TU858 (DT228/DT211C/DT282)

Portfolio Task 1

You are required to:

1. Run the script **AdvDB-Task1-applications.sql** (available in Brightspace).
The script creates a non-normalized table with the information about students applications to a college.
2. Normalize (1st, 2nd, 3rd form) the table **Apps_NOT_Normalized** respecting all the assumptions outlined in the description below, documenting your process.
3. Implement the normalized tables in Oracle (you can use Live SQL <https://livesql.oracle.com/>) and move data into the tables.
 - a. Make decisions about datatypes and constraints needed.
 - b. Document your decisions.
4. The normalized tables are usually more storage-efficient.
What is the gain in storage efficiency achieved by your normalization to 3NF?
Storage efficiency=the size in bytes of the normalized tables divided by the size of the starting non normalized table.
In order to estimate the size of the table refer to the maximum values for internal datatypes in Oracle available at https://docs.oracle.com/cd/A97630_01/appdev.920/a96584/oci03typ.htm
5. Create a section in your portfolio where you:
 - a. Explain your normalization process, clearly explaining any assumptions/choices/decisions mad.
 - b. Provide the DDL and SQL to create the tables resulting from your normalisation to 3NF and upload the data into those tables, clearly explaining any assumptions/choices/decisions made.
 - c. Explain the gain in storage efficiency achieved for your normalized tables in comparison to the non-normalised table.
 - d. Explain any assumptions/choices made for both the normalisation and the DDL/SQL.

Table Apps_NOT_Normalized:

- The following data describes information stored about students applying to a graduate school. Initially they are stored in one single table with no data normalization:

Application -No + StudentID+ StudentName + Street + State + Zip-Code + ApplicYear + Reference-Name + RefInstitution + Reference-Statement + Prior-School-Id + Prior-School-Addr + GPA

← data

- The table contains information about the application number (unique for each year), the student ID and name, the student address (street, state, Zip-Code) the year of the application, the name of one or more referees, their institutions and the reference statement written by each referee (so one reference statement for each referee for a specific application), and a list of prior schools the student attended. For each prior school the DB stores the address and the GPA (final score obtained by the student in that school).

- Assume the following:
 - Applications are submitted for each year, applications numbers are reset every year and therefore they are unique only for a particular year.
 - Student ID is unique for all the years (is assigned once for all once the student submit her/his first application).
 - A student might move to another address and the database has to store all the students addresses.
 - An applicant can only apply once during each year.
 - The combinations of Reference-Name and RefInstitution is unique (but Reference-name and RefInstitution are not unique as separate attributes).
 - Prior-School-Id uniquely identifies a university or college.
 - A student might have many prior-schools and if a student sends an application in different years the student may have added a new prior school (therefore prior school lists might change across applications).
 - An applicant has only one GPA from any specific prior school.
 - For each application there could be one or more referees.
 - Referees may write a reference statement for more than one applicants.
 - However, the reference statement for different applicants is different.
 - Reference statements are attached to an application. If a student submits a second application, the reference statements (even from the same referee) could be different.