Summer 2021 CS213 Project Part 1

> contributors:
>
> Zhu Yueming , Yu Tiancheng, Lu Hongyi, Wang Ziqin, Wang Weiyu, He Yirui, Yang Xiaosu

Overview

It is a one-person group project. Each student should finish the project by himself/herself and submit **one report in PDF**.

You should submit a report before the deadline, the **Topscore will be 100**; for the report submitted after the deadline and before the resubmission date, the **Topscore will be 80** ; for the report submitted after the resubmission date, the **score will be 0** .

Please be honest. DO NOT copy ANY words, figures and others from Internet and others.

DBMS can help us to manage data conveniently, and improve the efficiency of data

retrieval greatly.

Your work of the project Part 1 is mainly divided into two parts below:

1. Designing a relational database(create suitable tables) using postgreSQL according to the given data.
2. Finding ways to import the whole data into your database, try to be accurate, highly efficient and automated please.

Task Requirements:

Task 1: Database design

Design a database by **PostgreSQL** to manage all information mentioned in **course_info.json** and **select_course.csv** . Attention, there are some unreasonable data in these two files, correct them first. The total quantity of tables, the content in each table, all details should be determined by yourself.

Your design needs to meet the follow the **requirements**:

1. The tables created should satisfy the three normal forms;
2. Use primary key and foreign keys to indicate important attributes and relationships about your data. Every row in each table should to be uniquely identified by its primary key;
3. Every table should be involved in a link. No isolate tables included;(每个表要有外键或者有其他表的外键指向，也就是说不能有孤立表)
4. Your design should contain no circular links; (对于表之间的外键方向，不能有环)
5. Each table should always have at least one mandatory (“Not Null”) column (including the primary key but not the system-generated ID column);（每个表格中必须包含有至少一个非空的属性列，主键属于这个范畴，但是自增ID不属于）
6. Tables with no other unique columns than possibly a system-generated ID is not allowed; (除了主键自增的id之外，需要有其他unique约束的列)

7. Use appropriate types for different fields of data;
8. Your design should better be as easy to expand as possible.

Task 2: Import data

Design programs/scripts to import data into your database from those two files (**course_info.json** and

select_course.csv).

Your design needs to meet the follow the **requirements**:

1. Finding ways to improve the efficiency of time consuming during your importing process, and compare different importing methods.
2. Make sure all data are imported accurately , highly effectively and automatically .

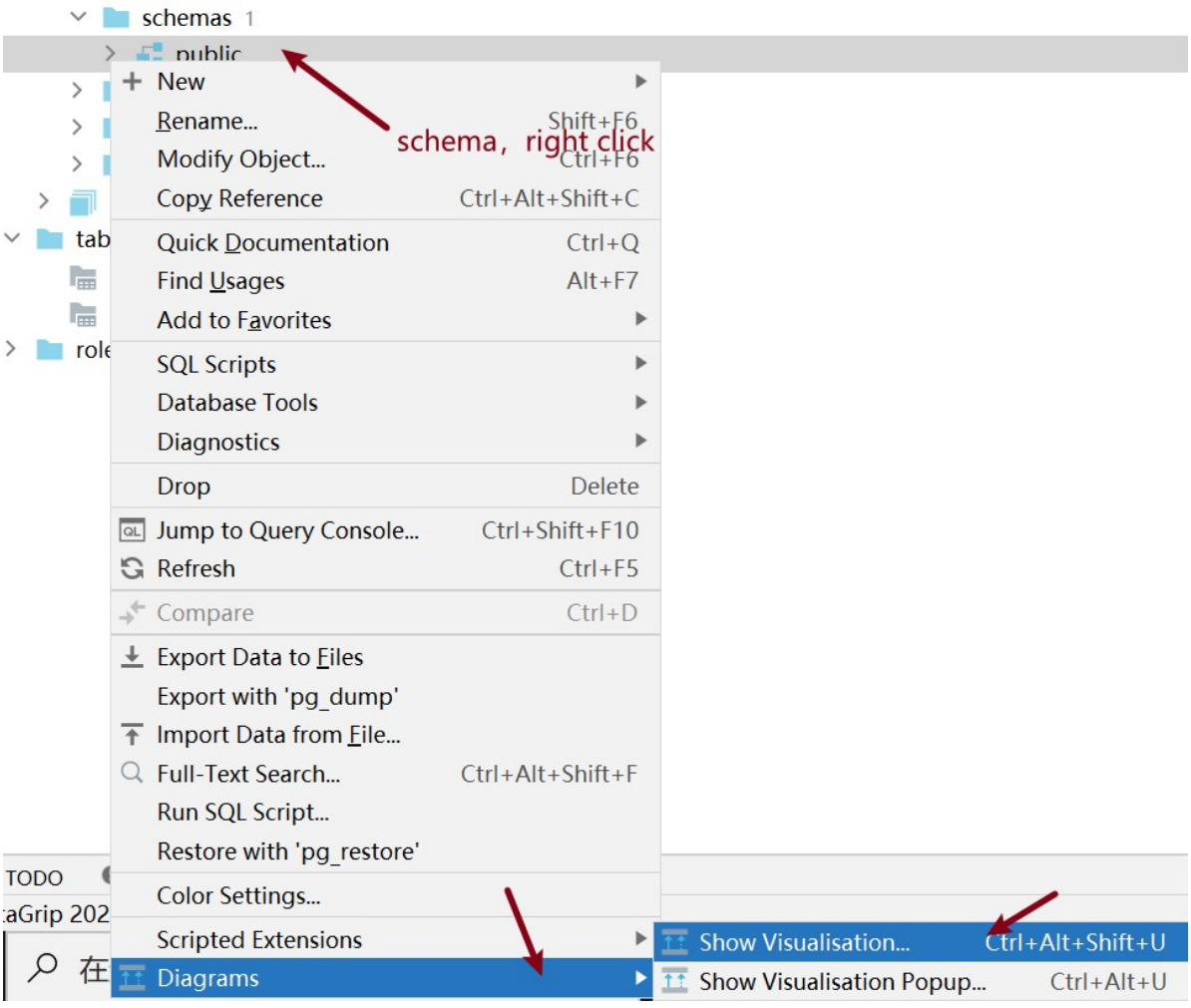
Report Structure:

Part 1. Group Info and Contribution

Need to write down your name and sid.

Part 2. Task 1

Provide a **clearly formatted** diagram of table structure that generated by DataGrip.



Give clearly explanations for the designing of your database, tables and some of columns (if needed).

Part 3. Task 2

Introduce how to design programs/scripts of importing data, and give the core code of your script.

Introduce how to improve the efficiency of importing data and give the core code. You can design some experiments to improve the efficiency in your work. Make sure that the experiments should be reasonable, and the improvement should be obvious.