DATA1001 Introduction to Data Science and Decisions

Assignment 2

Date due: 5pm Tuesday October 2

Question 1 (4 Marks)

1. (4 Marks) UNSW is part of the Group of 8 universities (Go8) and consistently ranked in the top 100 universities worldwide. UNSW Library is one of Australia's major research libraries, with extensive scholarly information resources. It serves a large population of academic staff and students on two campuses and affiliated research institutes. To improve the management efficiency, the manager wants to build a Library Management System. Please help the manager to draw an ER diagram based on the following specifications. State your assumptions if any.

- A book is uniquely identified by its book id. For each book, we also record its title, price, and availability.
- A reader is uniquely identified by his/her reader id and we also record his/her name, phone number, expire date and address. The address is composed of street and suburb.
- A publisher is uniquely identified by its publisher id. For each publisher, the name is also recorded.
- An author is uniquely identified by his/her author id. For each author, the name, phone number and birth date are also recorded.
- A reader can borrow zero or more books and a book can be borrowed by zero or more readers. Thus, we need to record the starting date and ending date for the borrowing relationship.
- A publisher can publish zero or more books and a book is published by exactly one publisher. We also need to record the date of publication.
- An author can write zero or more books and a book is written by one or more authors.

Question 2 (7 Marks)

UNSW handles its administrative information using a version of the Peoplesoft product called Campus Solutions. This system is normally accessed via the MyUNSW portal and is maintained by the NSS unit in UNSW IT. The database behind the system is hosted on a large Oracle server. The PeopleSoft system was installed in 2000 and has been modified/extended over the years to encompass:

- human resources (staff/employees, payroll, etc.)
- financials (purchases, income/expenditure, etc.)
- academic (students, courses, classes, enrolment, etc.)

It is in the process of being upgraded to the latest version of Campus Solutions (CS9).

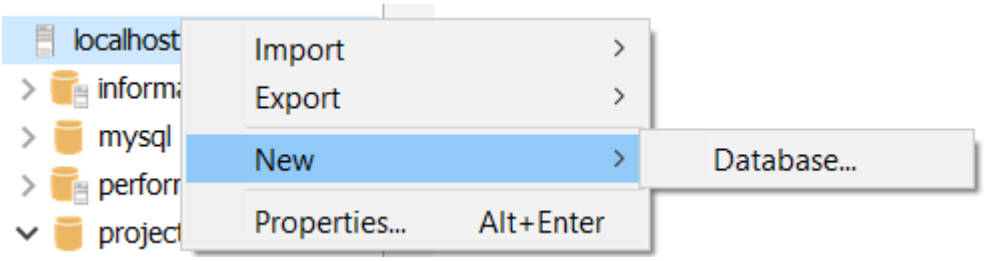
In this exercise, we give a subset of the human resources part which consists of three tables:
BUILDINGS (id, unswid, name, campus)
ROOMS (id, unswid, name, building, capacity)

STAFF (id, office, phone)

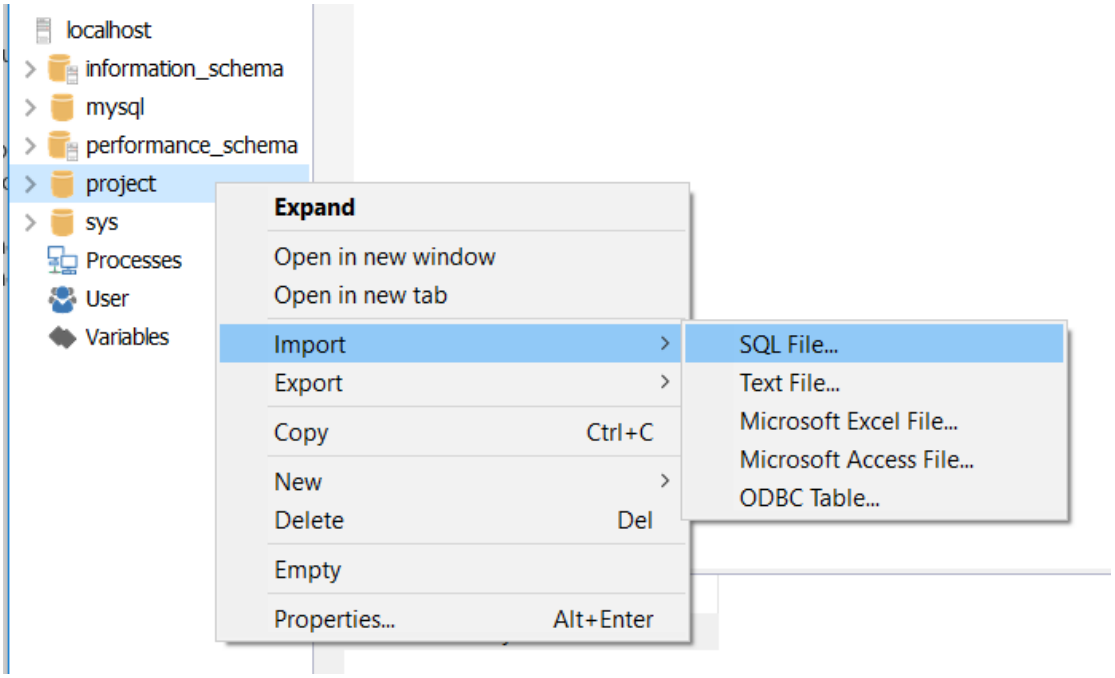
Note:
BUILDINGS.id is the primary key;
ROOMS.id is the primary key. ROOMS.building is the foreign key which references BUILDINGS.id;
STAFF.id is the primary key. STAFF.office is the foreign key which references ROOMS.id.
We omit the STAFF.name for privacy.

Database setup

Please download schema.sql and data.sql from the following address:
schema.sql: https://www.dropbox.com/s/u9qxoa5rckbeaog/schema.sql?dl=0
data.sql: https://www.dropbox.com/s/mlqdq305hdnu4gh/data.sql?dl=0
Step 1:
Open your MySQL front and create a new database named project under localhost.



Step 2:
Right click the database you have created and select “import SQL File”.



Step 3:
Select the schema.sql and data.sql you have just downloaded. Please note that you should import schema.sql first. After importing data.sql you may encounter some warnings, ignore it and click OK.
Please answer the following questions:

We provide answers to these questions based on **current** dataset to help you check your own answers. We will use another data set to test your answers when marking. It is your own responsibility to make sure your answers are logically correct. Note that, the titles of columns and the order of tuples do not matter.

1. (1 Marks) Find the number of different rooms with **capacity** >= 20 in the building named "**Computer Science Building**".
You answer should be like:

count(distinct r...
8

2. (2 Marks) Find all staff's **id** and **phone number** whose office is in the building named "**Biological Sciences Building**".
You answer should be like (note that the phone number can be empty and here just shows the first few lines due to limit of space):

id	phone
5033223	93852506
5033454	93852604
5033462	93858714
5033552	93853328
5033678	93854087
5033745	93852357
5033802	93821862
5033880	93823711
5033981	62688188

3. (2 Marks) Find the **id**, **name** and **the number of different rooms** for each building. Ignore the buildings with no room.
You answer should be like (just the first few lines due to limit of space):

building	name	count(distinct r...
100	Computer Scienc	78
101	Mechanical Engir	39
102	Chemical Science	35
103	Biomedical Theat	6
104	Civil Engineering	10
105	Australian School	35
106	Central Lecture Bl	14
107	Electrical Enginee	51
108	John Goodsell Bu	18
110	Library Stage 2	20
111	Mathews Building	56
112	Morven Brown Bu	41
114	Newton Building	2
115	Old Main Building	47

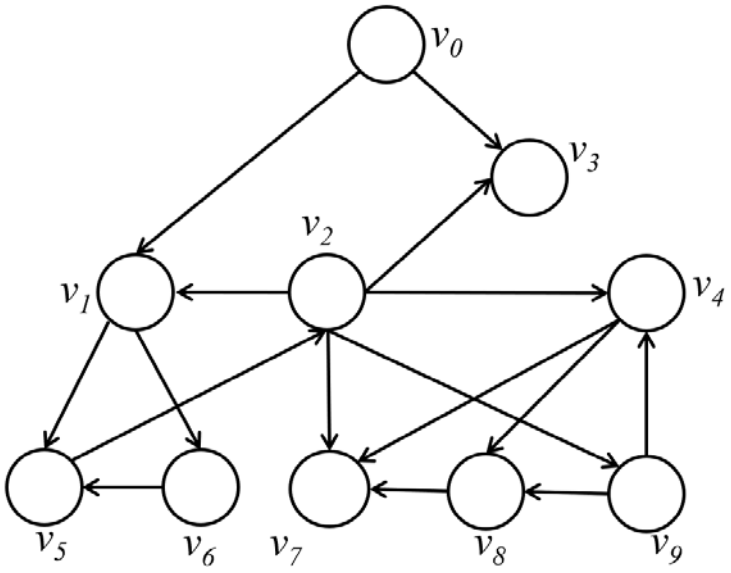
4. (2 Marks) Find the **name** of rooms which have a capacity larger than the average capacity of all the rooms.

You answer should be like (just the first few lines due to limit of space):

name
Seminar Room
Matthews Theatre A
Matthews Theatre B
Matthews Theatre C
Matthews Theatre D
New South Global Theatre
OMB-27
OMB-31
QUAD-1042
E8-219
Electrical Engineering G24
Electrical Engineering G25

Question 3 (4 Marks)

Given a directed graph *G* as follows:



Answer the following questions:

1. (2 Marks) Find the DFS (stack based) traversal order starting from v_0 and **show all the steps**.

2. (2 Marks) Find the BFS (queue based) traversal order starting from v_0 and **show all the steps**.

Assignment Submission

We accept electronic submissions only. Please submit your assignments as follows:

- For question 1 and question 3, save the answer as "**Ass2q1q3.pdf**".
- For questions 2, save the answer as "**Ass2q2.sql**".

- Compress these two files into a file named as your zid (e.g., "**5100000.zip**") and email to data1001cse@gmail.com. Note that only if you receive the **auto-reply** saying that your submission is accepted, your submission is successful.

Note:

1. If you have any problems in submissions, please email to kai.wang@unsw.edu.au
2. All submissions will be checked for plagiarism.

The University regards plagiarism as a form of academic misconduct, and has very strict rules regarding plagiarism. For UNSW policies, penalties, and information to help you avoid plagiarism see: https://student.unsw.edu.au/plagiarism as well as the guidelines in the online ELISE tutorials for all new UNSW students: http://subjectguides.library.unsw.edu.au/elise.

Late Submission Penalty

20% of the value of the submission will be deducted for each day (24 hours). Work submitted more than five days late will not be marked.