

# Lab 3: Data Manipulation Language

## University of Toronto Mississauga

Due: Friday March 11<sup>th</sup>, 2022 at 11:59AM ET

Released: February 25<sup>th</sup>, 2022

---

### Overview

For this assignment, we will be working with the *Coruscant Public Transit* (CPT), and perform data analytics over real data, tracking ship routes over the period September 1 - September 7, 2019.

On the CSC343 download site, you will find the following:

- a. `createTables.ddl`, a DDL file for CREATE TABLE statements.
- b. `loadData.ddl`, a DDL file for INSERT TABLE statements.

Please execute scripts `createTables.ddl` and `loadData.ddl` on your PostgreSQL database on the MCS server as you will use this schema for the questions below. Remember, your submission must work on the MCS lab machines!

### Structured Query Language (70 marks)

Write and provide SQL statements for each of the 10 questions, and their respective subsections. Execute each of your SQL queries against your CPT database and give the result of each query. You may consider writing views for some of these questions; if you do, only add it to your file once, not per question. Return attributes in the order the question below states them to be returned.

- (q1) [3 marks] Find the number of seniors (as “totalSeniors”) in the database. A ‘senior’ is defined as being 65-years of age or older (as of 12:00am on November 14th, 2019). (*Hint*: you may use the `date()` function.)
- (q2) [4 marks] Identify the total number of students who took ship route #1 on September 4th, 2019 as “taken”. Do not specify the column name.
- (q3) Profitability and revenue from advertising.
  - (a) [3 mark] Identify all your profitable ships. ‘Profitable’ is defined as a ship which has advertising revenue strictly greater than \$10,000. The only information required from this query is the ships’ unique identifier, age, and manufacturer.
  - (b) [3 marks] In addition to (a), above, identify the most profitable ship’s pilot(s). Return only their first name, last name, years of service, and the ship’s unique identifier.

- (c) [4 marks] Identify the total advertising revenue for each ship's route. The information required to be returned for this query is the ships' route identifier and total advertising revenue (named "Total Revenue"). Results are to be outputted in descending order of total revenue.

(q4) Profitability and revenue from fares.

- (a) [4 marks] For each passenger type, find the total fares' revenue. Return the passenger type and its associated total revenue (the column must be named "revenue").
- (b) [2 marks] Extend your query in part (a); only return passenger types and their revenue when the total revenue is greater than \$500. You will return the same columns as in (a).
- (c) [2 marks] Given what you did to solve part (a); return the most profitable passenger type (in terms of total fares' revenue) on September 1, 2019 and their "revenue".

(q5) Pilots and their violations/infractions.

- (a) [4 marks] Find all pilots who have less than 3 infractions. Return the pilot's unique identification, first name, last name, and age (named "Age").
- (b) [5 marks] For each pilot, return the total demerit points and total fines incurred (names "totalFine"). Pilots with less than 2 demerit points are irrelevant and can be omitted. Results must be sorted in a manner which places the most offending pilot (by demerit points then highest amount of fines) first on the list.

(q6) Continual growth of the CPT's service.

- (a) [3 marks] Identify what ships are different (i.e., are the only ship made by their manufacturer). Return the ship's unique identification and the name of the manufacturer.
- (b) [4 marks] Determine the most popular ship route on September 7, 2019. 'Popularity' is defined by the largest number of passengers. Results must provide the route's unique identification and the number of passenger trips (as "number of times").
- (c) [4 marks] Find what day contained the largest volume of passenger trips. Return the date and the number of trips taken (as "trips taken").

- (q7) (a) [4 marks] Find all the people who visited any "library" on either September 5, 2019 or September 6, 2019. We want to know their occupation, but we do not want to list duplicates. Therefore, return their occupation and "occurrences", where occurrences reflects the count of the number of times that specific occupation has appeared.
- (b) [3 marks] Extend your query in part (a); return the occupation with the most visits to the library separated by date (i.e., on September 5th, 2019 and on September 6th, 2019).

(q8) [5 marks] Find the pilots who have worked with CPT for more than 5 years, with a salary greater than \$75,000, and with less than 9 demerit points on their driving record. Return the pilot's first name, last name, and unique identifier.

(q9) [6 marks] Find all students who attended the "Jedi Knight Basketball" game at the Jedi Temple and arrived via a ship on route #4. Return the student's first name, last name, gender, and Phone Number. Note: CPT may not have a phone number for all people.

(q10) [7 marks] Assuming that the ship schedule has not changed since September 2019. Suppose you would like to attend the “YG 4hunnid Concert” (an event). The concert starts at 6:30pm and you would like to arrive at the site between 4:00pm and 6:00pm, which routes can you take? Use September 6, 2019 as a reference date for the ship schedule information. Return the route’s unique identification, the ship’s stop name (i.e., where should you get off the ship), and the scheduled arrival time.

## Requirements and Submission

This lab is to be completed in partners (in the same pairs of 2 selected in Lab 1 & 2) unless written permission is given by the Course Coordinator. You and your partner are required to work together, equally contribute to, and understand all parts of your submission. Please refer to the syllabus for additional details on groups and the “Minimum Standards for Submitted Work”.

You are **required** to use the `skeleton.dml` file provided on the ([CSC343 download site](#)). You will rename that file to `lab3.dml` and add your SQL queries (i.e., your solutions) to it. Please ensure that your SQL statements are syntactically correct and that they are executable on the MCS PostgreSQL server. Non-executable queries (or those producing no output) will receive a grade of 0. Do not leave your solutions commented out, if you do a penalty will be applied.

You are submitting **one file** called **lab3.dml**. You must include your, and your partner’s, full name and student ID number at the top this DDL file where specified (see “PREAMBLE” in `skeleton.dml`).

All files are to be submitted using the MarkUs platform (<https://markus108.utm.utoronto.ca/csc343s22/>). You or your partner must create the “group” and the other must accept the invitation to join. Once your group is formed on MarkUs, only one person from each group is required to submit the file(s). You may submit as many times as you like, in fact you are encouraged to do so! And do so early!! Groups that were created in Lab 1 will persist (so there may be no need for you to create a group!).