

# CSE 460 Project 2: SQL Query

Due 23:59 11/10/2018 EST

## 1 Submission

*Failure to follow the submission specifications will incur penalties for EACH violation.*

### 1.1 What to submit

A zip file has to be submitted through the ‘submit\_cse460’ or ‘submit\_cse560’ script by 11/10/2018 11:59PM EST. **ONLY** zip extension will be accepted.

### 1.2 Zip file naming convention

Use *ubit\_proj2* (**NO SPACE!**) for the filename, e.g., *jsmith\_proj2.zip*, where *jsmith* is the ubit. The project is an **INDIVIDUAL** project, copied solutions will be considered violations of academic integrity.

### 1.3 Sub-structure of zip file

- On unzipping the zip file, there should be a folder named as *ubit\_proj2*, where *ubit* is your ubit.
- Under the folder *ubit\_proj2*, there should be a sql file, name the sql file as ***ubit\_proj2.sql***, where *ubit* is your ubit, this file should contain your answers of the questions in this project, use comments to separate your answers for the questions.

## 2 Problem 1 (7 pts)

Read and **understand** the DB schema in `personnel.sql`, `EMPLOYEE` contains the basic information of the employees, `DEPARTMENT` contains the basic information of the departments, `WORKSIN` contains information about which employees work in which departments, `PAYROLL` contains the information of payments, the column `PAYROLL.year` stores the year of payments in the form of ‘YYYY’, e.g., ‘2016’. Write the following SQL queries in your sql file. You’ll need to create your own test data to verify your queries, but do **not** include any of your test data in your submission.

- **1.1 (1 pt)** Find the names of employees who work in **at most** two departments.
- **1.2 (2 pts)** An employee is said to be *loyal* to the company if the employee has been working in the company for at least 3 years, i.e., the employee got paid in at least 3 different years, find the ids and names of loyal employees.
- **1.3 (2 pts)** Find for every employee his/her total income for the year 2016 (if the income is not applicable to an employee, still list the employee name and return *null* as the value for the total income), the query result should have the schema (name, income).
- **1.4 (2 pts)** For every department, list the department id and the total number of employees in the department that had total income in 2016 more than the departmental average income of 2016, i.e., the average income of all the employees in the department in 2016.

### 3 Problem 2 (3 pts)

Consider the following relational DB schema: `TRAIN(Origin, Destination, Price)`, where the domains of `Origin` and `Destination` are strings, the domain of `Price` is integer, and the primary key is as underlined. A tuple  $\langle A, B, C \rangle$  in `TRAIN` can be interpreted as “from station  $A$  there is a train to station  $B$  with the ticket price  $C$ ”. Answer the following questions and write the answers in your sql file.

A *cycle* in the relation `TRAIN` is defined as the following: start growing a path from station  $A$ , if there’s a path to travel back to  $A$ , then we say there’s a *cycle* for the station  $A$ . E.g., if we have tuples  $\langle A, B, 1 \rangle, \langle B, C, 2 \rangle, \langle C, A, 3 \rangle, \langle D, A, 4 \rangle$ , then there’s a cycle for  $A$ , a cycle for  $B$  and a cycle for  $C$  (with the path  $A \rightarrow B \rightarrow C \rightarrow A$ ), but there is no cycle for  $D$ .

- **2.1 (2 pts)** Assume that there’s **no** cycle for any of the station in the instance, write a SQL query to find all the stations that are reachable from the station ‘BUF’, and also compute the lowest costs to travel to those stations from ‘BUF’, an example result schema is (Destination, LeastCost).
- **2.2 (1 pt)** What could happen if you run your query on an instance with cycles? and why? Write your answer for this question as comments in your sql file.