McMaster University
SFWRENG 3DB3 Fall 2024
Assignment 2
Due: Nov 4, 2024 at 10:00pm

October 11, 2024

For this assignment, we will continue to work with *MacAir Aviation*, and perform some simple data analytics. On the Assignments webpage, you will find:

a) the DDL for CREATE TABLE statements, `createTables.ddl` (to create the necessary tables), and

b) the INSERT statements `loadData.ddl` (to load data into the tables).
Running this script can take a minimum of 10 - 20 mins, and possibly longer, depending on the server load. It is recommended that you do not wait until the last minute to populate your database!

These files correspond to a simplified schema shown in the E-R diagram `asg2ER.pdf`. Please run these two scripts on your database (i.e., remember to update the CONNECT TO statement with your database name). You will use this schema for the questions below.

I. SQL (77 marks)

Write SQL statements for the following queries. Execute each of your SQL queries against your MacAir database, and give the result of each query.

Q1) [5 marks] Find all passengers, between the ages of 20 and 30 (inclusive), who have a 'Vegan' or 'Vegetarian' dietary preference. Return their ID, name, and age.

Q2) (a) [4 marks] Find the number of airplanes that exist for each model. Return the model and the count for each model.
(b) [4 marks] Extend your query from (a) to find the number of airplanes in each model for any of the following airlines: 'Air Canada', 'Etihad Airways', or 'United Airlines'. Return the name of the airline, the model, and the number of airplanes.

Q3) (a) [5 marks] For each 'Air Canada' ticket, find the average of the total weight, for all baggage associated to the ticket. Return the ticket number, and the average total (baggage) weight.

(b) [5 marks] Find all tickets with 'Oversized', non-fragile baggage with a total weight (strictly) greater than 90 lbs, during the holiday season from Dec. 10, 2023 to Jan. 3, 2024 (inclusive). Return all qualifying ticket numbers, and the total (Oversized) baggage weight.

Q4) [6 marks] Where and when are the cheapest tickets for flights from Toronto 'YYZ' to Orlando 'MCO'? Return the ticket number, the date of departure, the minimum price (rename to minPrice), and the website where the ticket(s) were purchased.

Q5) (a) [5 marks] Which routes are served by at least three airlines? Return the routeID, and display your results in descending order by the number of airlines.
(b) [4 marks] Which routes are not served by any airline? Return the routeID, the source and destination airports.

Q6) (a) [5 marks] Find the number of distinct passengers who also work as either a pilot, cabin crew, or ground staff. Rename this result as NumStaffPassengers.
(b) [6 marks] For each airline, how many pilots or cabin crew are also passengers? Return the airline (alias), and the corresponding count.

Q7) (a) [7 marks] Find all the one-way routes operated by airline 'ACA', i.e., airline alias = 'ACA'. In this context, a one-way route is where the airline serves from a source airport to a destination airport, but not in the reverse direction. Return the route ID, and the corresponding source and destination airports, respectively.
(b) [7 marks] Find the most popular route where the departure date lies between '2023-12-01' to '2023-12-31' (inclusive). Popularity is defined as the maximum number of tickets purchased during this time duration. Return the route ID, the corresponding source and destination airports, and number of tickets sold along this route.

Q8) (a) [7 marks] Which Air Canada (alias 'ACA') flights from source airport 'YYZ' to destination airport 'MCO' have 'First' class tickets? Return all satisfying flight numbers.
(b) [7 marks] Find all airlines that are unique to their country (i.e., they are the only airline for their country). Return the airline alias, airline name, and the country name.

II. Relational Algebra (48 marks)

For the queries Q1 - Q6 in Part I, give the corresponding relational algebra expression. Please submit your **typed** answers, using your choice of software (e.g., LaTeX, MS Word). **Handwritten relational algebra solutions will not be accepted.**

III. Indexes (10 marks)

For the workload given in Part I, several queries have been showing poor performance (i.e., increasing response times). Your task is to improve the performance of this workload as much as possible by recommending two indexes that should be defined on any of the tables. What two indexes would you recommend? For each index, state:

- The attribute(s) the index is defined on.

- Properties of the index (e.g., type of index, clustered/unclustered, composite, etc.)

- Which queries (Q1 - Q8) you think this index will help, and why.

Grading

This assignment is worth 14% towards your final grade.

Submission

All files are to be submitted using the Avenue system. Please ensure you submit all files with the correct names, as described below. In each file, include your name and student ID number. Upload four files with the indicated file extensions (no compression based `.tar, .zip, .rar` files will be accepted).

- For Part I:

  – Submit your SQL statements and the result for each query in two files. Submit your SQL statements in a script file called `queries.sql`, and the corresponding query results in a file called `queries.results`.

  – Ensure your SQL statements are syntactically correct and that they are executable on the DB2 server. **Non-executable queries (that do not run on the DB2 server) will not be marked, and will receive a grade of zero.**

  – Clearly label, with comments, which query the result tuples correspond to in `queries.results`.

- For Part II: Submit your relational algebra expressions in a file named `ra.pdf`.

- For Part III: Submit your index recommendations in a file named `index.pdf`.