

McMaster University  
SFWRENG 3DB3 Fall 2021  
Assignment 2  
Due: November 5, 2021 at 10:00pm

October 11, 2021

For this assignment, you will continue to work with *StuffNOW* (SN), and perform data analytics over product transactions data. On Avenue and Microsoft Teams, under Assignments/Assignment 2, 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).

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 (85 marks)

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

- q1) [5 marks] Find all customers (first name, last name, date of birth) who are 18 years of age or older, and placed an order on July 22, 2020. (Hint: the DATE and LIKE function may be used, if needed). 2 records
- q2) [5 marks] For all products purchased by customers between the ages of 20-35, return the product ID and category name. Do not return duplicates. 1861 records
- q3) [5 marks] Which customers wrote the most reviews (across all products)? Return the customers' last name, first name, date of birth, city, and country. 4 records
- q4) (a) [5 marks] To save delivery costs, compute the number of shipments with more than one order. =546

- (b) [5 marks] Find the shipments from (a) that are delivered to a postal code in Toronto. Note: postal codes in Toronto begin with the letter 'M'. Return the tracking number of the shipment(s). **33 records**
- q5) [5 marks] Find all products (productID) that belong to exactly one category. **344 records**
- q6) (a) [5 marks] Find those products that are unique to their brand. That is, the product is the only one made by that brand. Return the productID, name, and brand. **5 records**
- (b) [6 marks] Find all orders (return the orderID) with the highest sales amount. For example, if the maximum sales is \$1000 (for a single order, among all orders), return all orders with sales greater than or equal to this amount.
- q7) [6 marks] List the stores (StoreID, description, and start date), and their revenue (in ascending order) for the month of July, 2020 (i.e., from July 1, 2020 to July 31, 2020). **10 records**
- q8) (a) [5 marks] Which products (productID, name, and brand) have never been purchased by any customer? **10 records → 991~1000**
- (b) [5 marks] Find all products (productID) in (a) that have been on promotion. **2 records**
- q9) (a) [6 marks] Find the product categories (product category ID, name) such that *all* the contained products have a warranty. For example, if every storage drive in the 'Computer Storage' category has a warranty, then return 'Computer Storage' and its ID. **3 records**
- (b) [7 marks] From the product categories in (a), find the stores (return their storeID) that sell *all* those products. For example, store ID #101 sells all the storage drives (products) from the category in (a). **1 record → 102**
- q10) (a) [7 marks] Which products have an average rating greater than the average rating (among all products) in their product category? Return the product ID, product model number, and product name. **152 records**
- (b) [8 marks] Compute the total sales for each product in part (a). Display your output in descending order of the product revenue. **152 records** **27 6027**

## II. Relational Algebra (60 marks)

For each of the SQL queries in Part I, give the corresponding relational algebra expression. Please submit your **typed** answers, using your choice of software (some suggestions are LaTeX or MS Word). **Handwritten relational algebra solutions will not be accepted.**

## III. Indexes (15 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 four indexes that should be defined on the tables. What four 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 - q10) 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 servers. Non-executable queries 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`.