

Homework 3: SQL

DSCI 551 – Spring 2025

Due: 11:59pm, March 10, 2025, Monday

Points: 100

In this homework, you are provided a banking dataset with four CSV files on customers, accounts, branches, and (bank) employees. Note that your solution should work on similar CSV files with similar formats.

1. [15 points] Reverse engineer the tables into an ER model. Assume that an account belongs to exactly one customer and exactly one branch, a branch has exactly one manager, and an employee works for exactly one branch. Draw the ER model electronically.
2. [15 points] Write an SQL script `banking.sql`. The script should create a database “banking”, four tables, and populate the tables using the data in the CSV files. You can refer to `world.sql`, the in-class example, for an idea on such an SQL script.
3. [45 points] Write SQL queries for each of the following questions. For each query, show the output of the query.
 - 1) [5 points] Find ids of customers who have the largest total balance (which is the sum of balances of all accounts owned by the customer). Note that there may be more than one such customer.
 - 2) [5 points] Find names of customers who have accounts in the Downtown branch. Output the name in the format of “First Last” (e.g., “John Smith”). Output the same name only once.
 - 3) [5 points] Find the name of the manager who manages the Downtown branch. Return the name in the format of “First Name”.
 - 4) [5 points] Find names of branches which have the largest number of customers. Note there may be multiple such branches.
 - 5) [5 points] Find names and job titles of employees who are NOT managers. Output the name in the format of “First Last” (e.g., “John Smith”). Sort the output by name ascending. Note that such employees are ones whose IDs do not appear in the `ManagerID` column of the `branches` table. Do not use Job title to determine if an employee is a manager.

6) [15 points] Find ids of customers who have both checking and saving accounts in Branch B001.

- a. Using set operation
- b. Using join
- c. Using subquery (without using set operation)

7) [5 points] Find names of branches which have at least 3 employees.

4. [25 points] Write a Python program `search_customer.py` that takes the customer first and last name and returns the AccountID, type, and balance of accounts held by the customer.

Execution format:

```
python3 search_customer.py <first> <last>
```

For example,

```
python3 search_customer.py John Doe
```

will return:

```
A001,Savings,1500
A003,Checking,    500
```

Permitted libraries: sqlalchemy, pymysql, sys, and pandas.

● Submission details:

Submit 4 files in total

- **ER.jpg** : ER Diagram drawn electronically
- **banking.sql**: SQL file with create and insert statements required for Question 2
- **queries.sql**: SQL file with the select statements for Question 3. Include results as comments below each SQL statement.
- **search_customer.py**: py script for Question 4