CSCI-5408 DATA MANAGEMENT, WAREHOUSING, & ANALYTICS

ASSIGNMENT - 1

Banner ID: B00977669

GitLab Assignment Link:

https://git.cs.dal.ca/saji/csci5408 w24 b00977669 christin saji

Problem 1: Build a Conceptual Model

Task 1: List of Entities and Attributes

Table 1 20 Entity set with their attributes and reasoning

S.No.	Entity	Attributes	Reason
1.	Student	StudentID, FirstName, LastName, DateOfBirth, Email	To uniquely identify each student, and communicate with them
2.	Faculty	FacultyID, FirstName, LastName, DepartmentID, Email, OfficeNumber	To uniquely identify faculty members, associate them with departments, and provide contact information
3.	Course	CourseID, CourseName, DepartmentID, CreditHours	To catalog courses, link them to departments, and define their weight
4.	Program	ProgramID, ProgramName, DepartmentID, DegreeLevel, TotalCreditHours	To define the various academic programs and their requirements
5.	Department	DepartmentID, DepartmentName, FacultyCount, Budget	To manage the administrative aspects of academic departments
6.	Lecture Hall	HallID, BuildingName,	To manage the physical spaces where courses are taught and their features

		Capacity, EquipmentDetails	
7.	Enrollment	EnrollmentID, StudentID, CourseID, Semester, Grade	To record what courses students are taking and their performance
8.	Academic Advisor	AdvisorID, FirstName, LastName, Email, DepartmentID	To provide information on individuals who guide students academically
9.	Research Group	GroupID, GroupName, FocusArea, LeadFacultyID, DepartmentID	To organize research efforts and link them to academic departments
10.	Research Project	ProjectID, Title, Summary, StartDate, EndDate	To manage individual research initiatives and track their timelines
11.	Library	LibraryID, Name, Location, OpenHours, ResourceCount	To manage the operations of university libraries
12.	Book	ISBN, Title, Author, LibraryID, CheckOutStatus	To catalog library books and track their availability
13.	Club	ClubID, Name, Purpose, FacultyAdvisorID	To manage student clubs and their activities
14.	Event	EventID, Title, Description, StartTime, EndTime	To schedule and promote university events
15.	Sports Team	TeamID, Sport, CaptainStudentID, TeamRanking	To manage sports teams and their competitive standings
16.	Scholarship	ScholarshipID, Name, EligibilityCriteria,	To manage financial aid offerings and their recipients

		Amount, ApplicationDeadline	
17.	Services	ServiceID, Name, Description, Availability, DepartmentID	To detail the services provided by the university to students and staff
18.	Job	JobID, Title, DepartmentID, Description, SalaryRange	To manage employment opportunities available within the university
19.	Alumni	AlumniID, FirstName, LastName, GraduationYear	To maintain a network of former students
20.	Dormitory	DormID, Name, Capacity, Amenities, RoomType	To manage student housing options and their features

Task 2: Initial Conceptual Model (ERD_Initial)

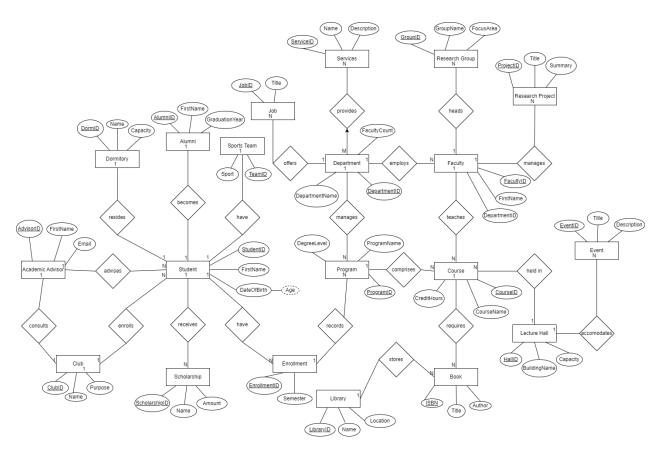


Figure 1 Initial ER Diagram for Halifax City University

Task 3: Design Issues Identification and Justification

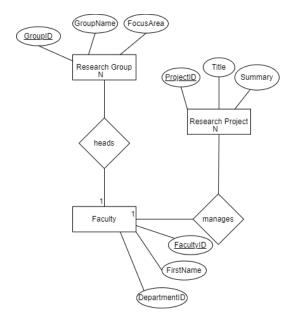


Figure 2 Chasm Trap between Research Group, Faculty and Research Project

The design issue present here is the chasm trap. Because of this, we can't determine which Research Group is working on which Research Project, and Research Groups and Research Projects can exist independently, which doesn't satisfy the requirements.

To solve this, we need to enforce a direct relationship between Research Group and Research Project.

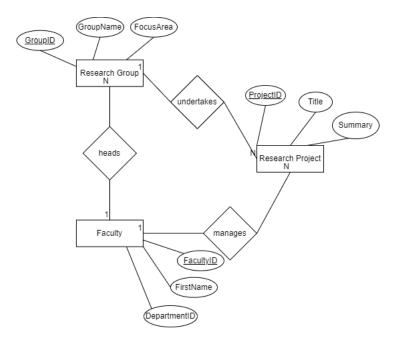


Figure 3 Solution for the Research Group, Faculty and Research Project chasm trap

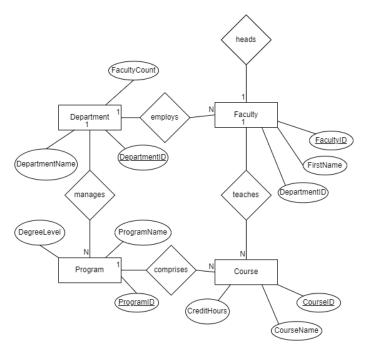


Figure 4 Chasm Trap between Department, Faculty and Course

Here, the chasm trap is present because not all faculty who head departments teach courses, and not all faculty who teach courses are part of the department.

To solve this design issue, we need to form relationships between Department and Course.

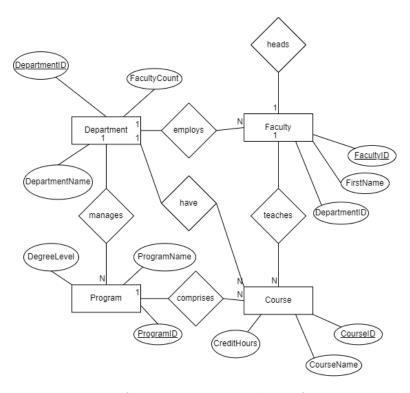


Figure 5 Solution for Department, Faculty and Course for chasm trap

Task 4: Final ERD Model (ERD_Final)

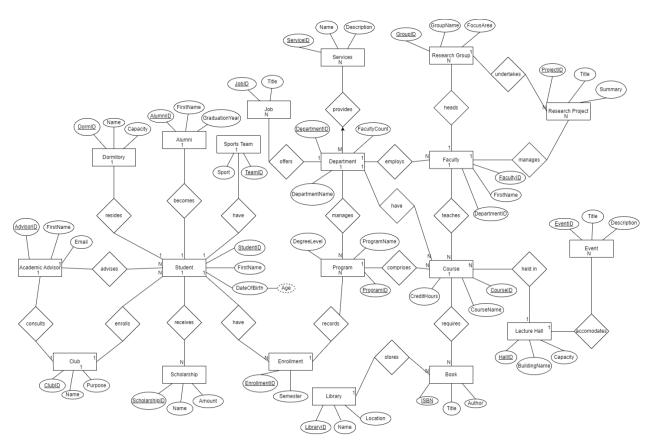


Figure 6 Final ER Diagram for Halifax City University

Task 5: Final ERRD Model (Final_ERRD)

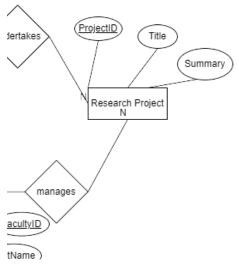


Figure 7 Research Project entity to be extended

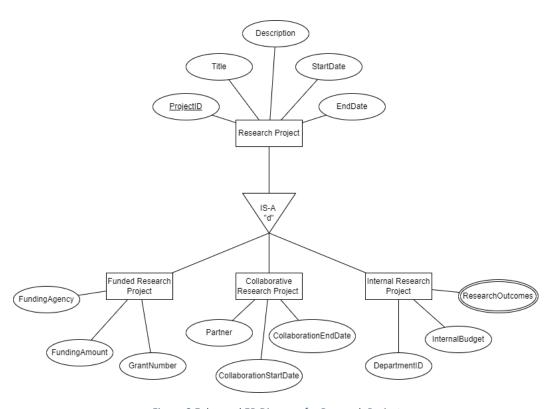


Figure 8 Enhanced ER Diagram for Research Project

Research Project (superclass) can be split into Funded Research Project, Collaborative Research Project, and Internal Research Project (subclasses). Each subclasses inherit the attributes from the Research Project, and they have their own attributes. They represent disjoint constraint because instance of the superclass can be a member of only one of the subclasses at a time.

Problem 2: Custom DBMS Using Java

Task 1: Documentation of Design Principle

Application Architecture

The lightweight database management system is structured using a **layered architecture** style, with this the system is divided into a number of layers, each of which is responsible for a different task and interacts with the layers above and below it [1].

Application is structured as follows:

- **Presentation Layer (application)** This package is the entry point of the application which handles user interaction and provides a command-line interface.
- **Business Logic Layer (services)** This package encapsulates the core business logic and operations of the application.
- Data Access Layer (objects) This package contains various query types which directly manipulates the database files.
- Utility Layer (utils) This package provides supporting functionalities across the application.

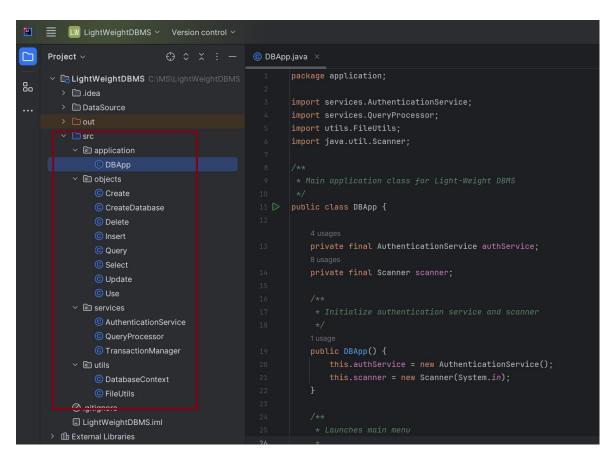


Figure 9 Application Architecture for lightweight database management system.

Following Single Responsibility Principle, each class is responsible for one job [2]. For example, 'QueryProcessor' is responsible for interpreting and directing SQL queries, 'AuthenticationService' handles user authentication, and 'FileUtils' manages file operations.

'Query' class is abstract, allowing for extension to handle various types of queries ('Select', 'Insert', 'Update', 'Delete') without modifying the base class this satisfies Open/Closed Principle. Also, from 'Query' class all classes inherit this ensures than any subclass can replace the 'Query' class without affecting the correctness of the program satisfying Liskov Substitution Principle.

Task 2: User Authentication

Register Process

The menu of the lightweight database management system is displayed when the application is run. When user selects option 2 to register as shown in Figure 10, the user is able to successfully register with the system.

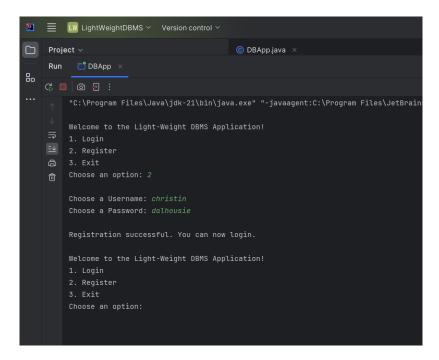


Figure 10 Successful registration process

User credentials are stored safely in a `user_info` text file, and the password is hashed using the MD5 algorithm.

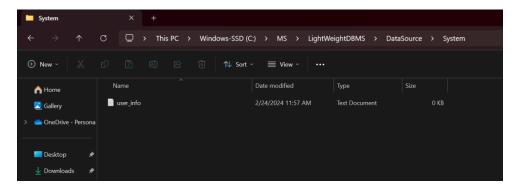


Figure 11 'user_info' contains all the user credentials

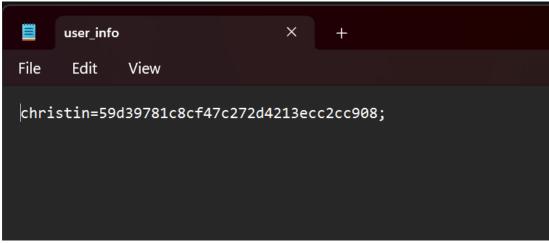


Figure 12 'user_info' contains username with MD5 hashed password

Login Process

When a user enters incorrect credentials, they won't be able to log in to the system.

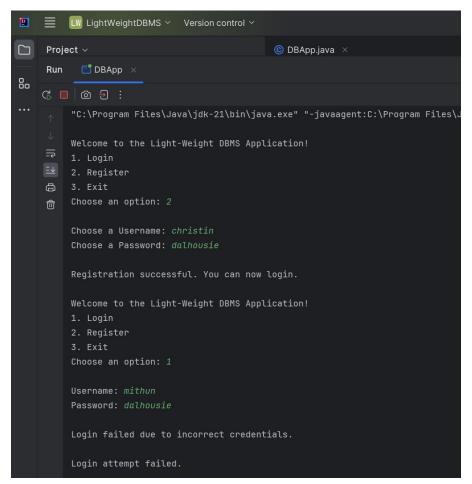


Figure 13 Login process failed due to incorrect credentials

After entering the correct credentials to log in, the user will be prompted to enter a randomly generated 4-digit captcha. If they enter the incorrect captcha, they won't be allowed into the system.

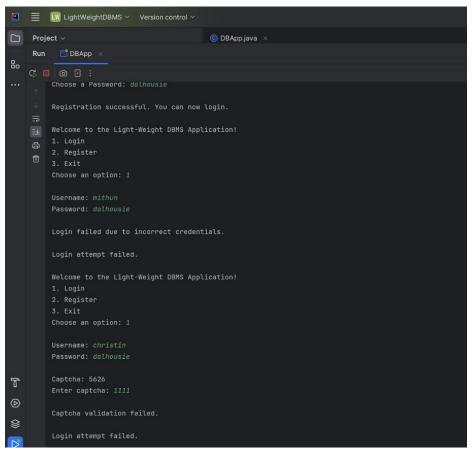


Figure 14 Incorrect captcha

After successfully entering the correct credentials and captcha, the user is allowed to perform SQL operations in the system.

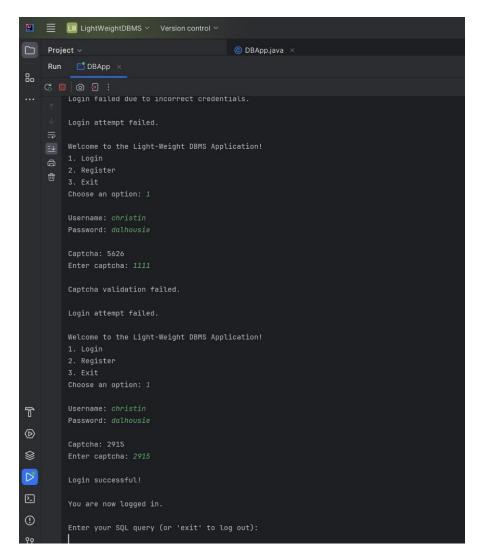


Figure 15 Successful login after entering correct credentials with captcha

Task 3:

Task 3.1: Design of Persistent Storage

Novelty Task: Table data is stored using a custom delimiter separated by \$.

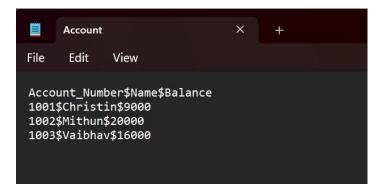


Figure 16 Table data stored using custom delimiter

When each table is created, it creates a `(table_name).data` file extension, and its metadata is stored in the `Schema` folder as `(table_name)_schema.txt`.

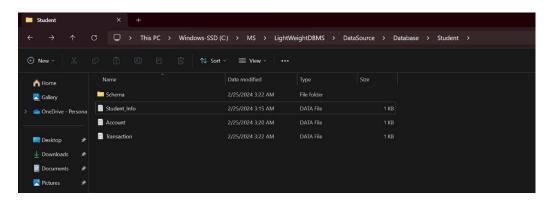


Figure 17 Table data are stored in (table_name).data file extension

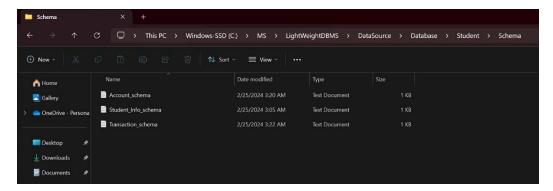


Figure 18 Each table has their own schema file

Task 3.2: Implementation of Queries (DDL & DML)

CREATE DATABASE

Query: CREATE DATABASE Student;

It creates a folder named after the `Database_Name` in the file system.

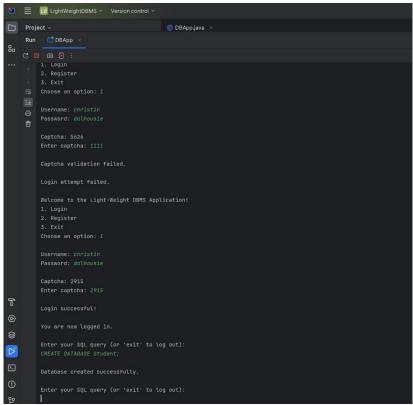


Figure 19 CREATE DATABASE operation

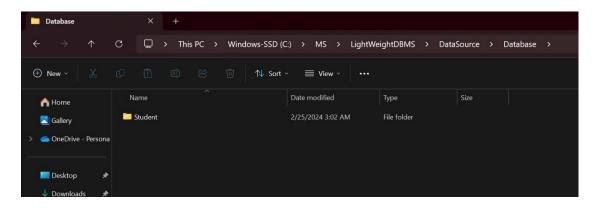


Figure 20 Student folder is generated for the database

USE

If a user tries to access using a DML command, they will be prompted to first select the database.

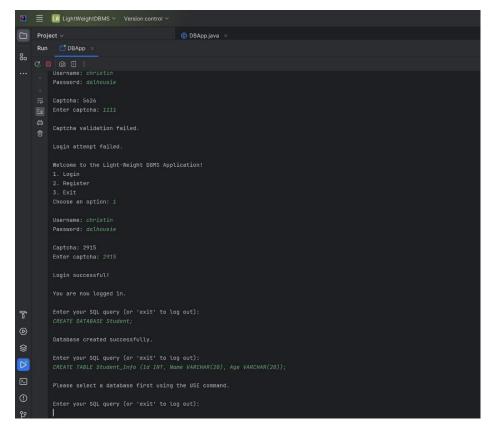


Figure 21 Using DML commands without selecting database

Query: USE DATABASE Student

Selecting the database makes it the active database for the DML operations.

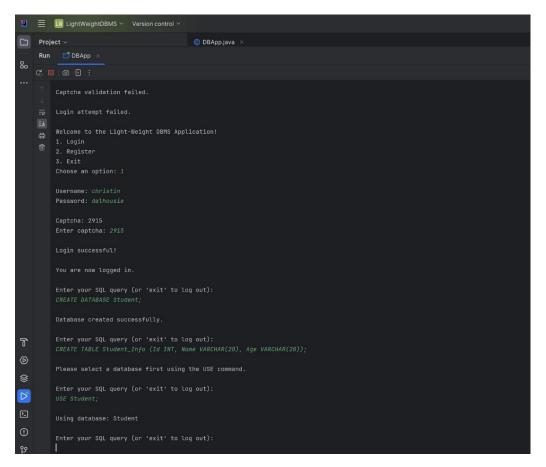


Figure 22 USE Student to select Student database

CREATE TABLE

Query: CREATE TABLE Student_Info (Id INT, Name VARCHAR(20), Age VARCHAR(20));

This creates a `Student_info.data` file to store the table data inside the `Student` folder, and the metadata of the table is stored inside the `Schema` folder.

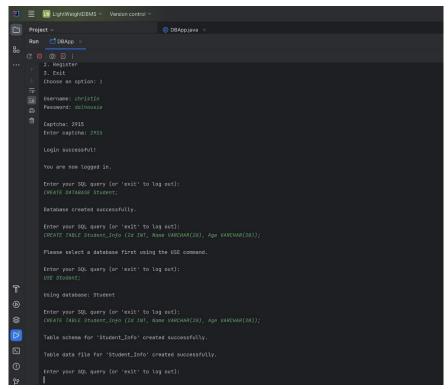


Figure 23 CREATE TABLE operation

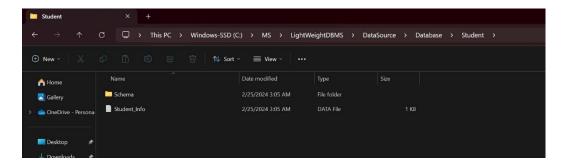


Figure 24 Creates Student_Info.data to store table data

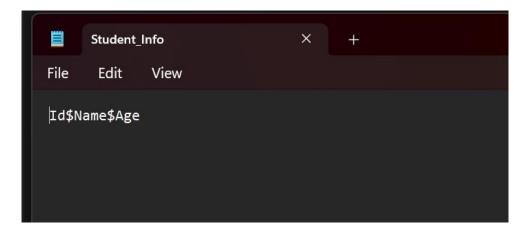


Figure 25 Student_info with header

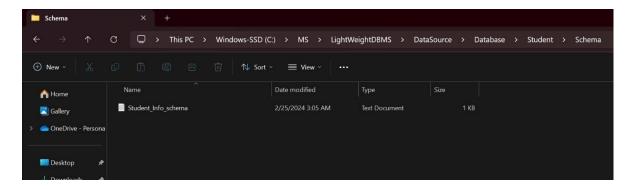


Figure 26 Student_Info_schema to store meta-data

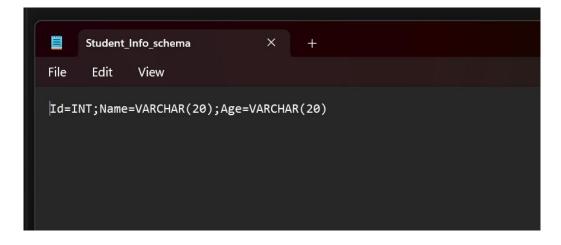


Figure 27 Meta-data of Student_Info table

INSERT

Query: INSERT INTO Student_Info (Id, Name, Age) VALUES (101, 'Christin', 22);

The insert operation puts the data inside the `Student_Info.data`, with each value separated by the custom delimiter \$.

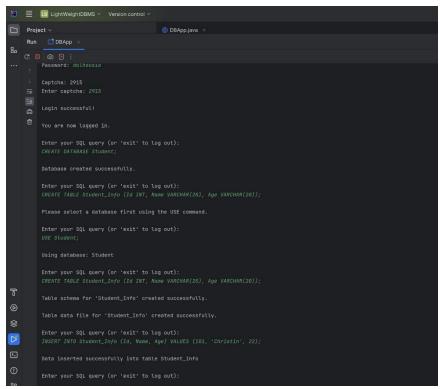


Figure 28 INSERT operation to the Student_Info table

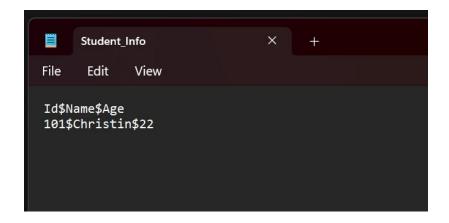


Figure 29 Data stored in Student_Info using custom delimiter

The insert operation also supports multiple insert statements to be entered together, and the values are stored in the data file.

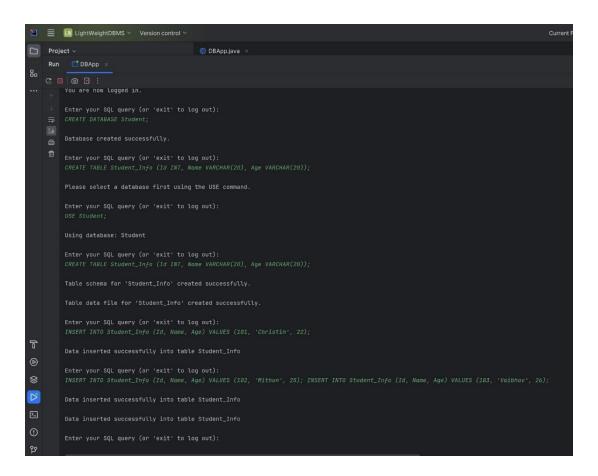


Figure 30 Inserting multiple data to Student_Info table

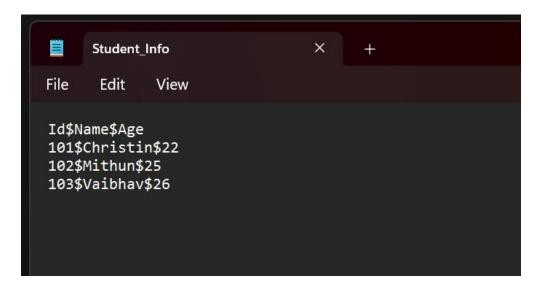


Figure 31 Multiple rows of data in Student_Info table seperated by custom delimiter

SELECT

Query: SELECT * FROM Student_Info;

It selects all the columns and displays them nicely spaced.

```
| Color | Colo
```

Figure 32 SELECT * to select all the columns and display it

We can select particular columns, and it will display that information.

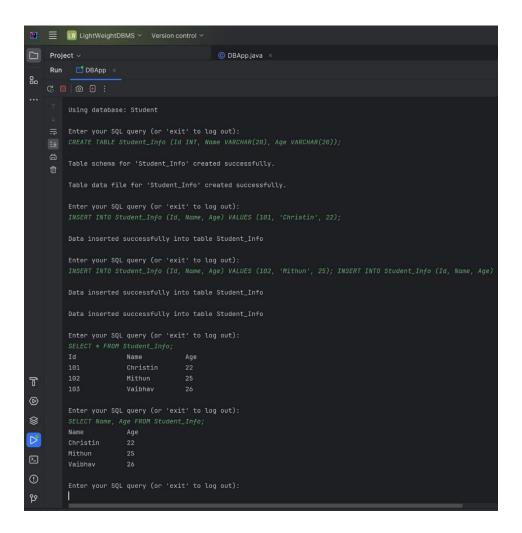


Figure 33 Selecting particular columns

Using the WHERE clause, we can display the particular information we need.

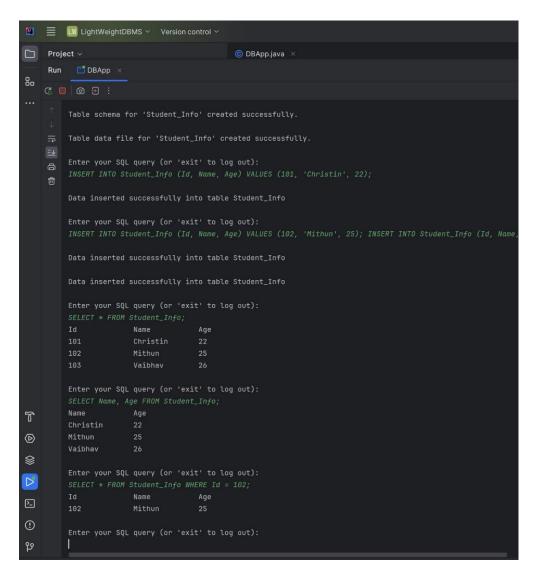


Figure 34 Using WHERE clause to display particular information

UPDATE

Query: UPDATE Student_Info SET Age = 24 WHERE Id = 102;

After a successful operation, it updates the data file.

```
Project V © DBApp.java ×

Run DBApp ×

C DBApp ×

C DBApp.java ×

Run DBApp.java ×
```

Figure 35 UPDATE operation into Student_Info table

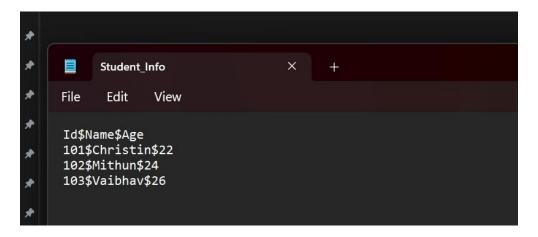


Figure 36 Updated data in Student_Info

DELETE

Query: DELETE FROM Student_Info WHERE Id = 103;

After successful execution, the particular information that is selected is removed from the data file.

Figure 37 Delete operation on Student_Info table

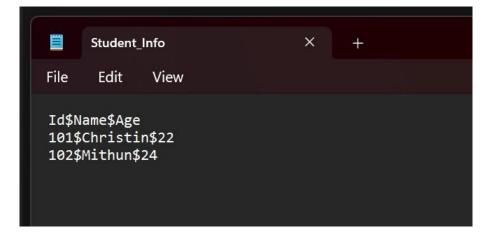


Figure 38 Update data in Student Info table after delete operation

Task 4: Implementation of Transaction

Setup

I created two tables, Account and Transaction, which I will use as base tables to run transaction operations on them. This will create `Account.data` and `Transaction.data` files, along with their respective metadata files in the Schema folder.

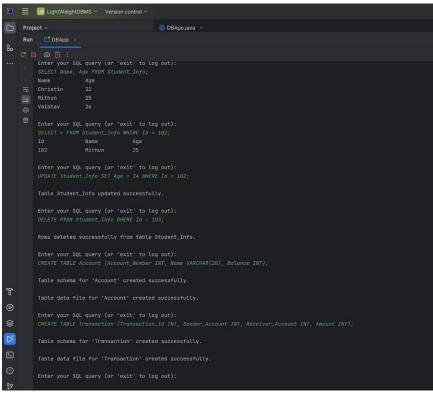


Figure 39 Created Account and Transaction table

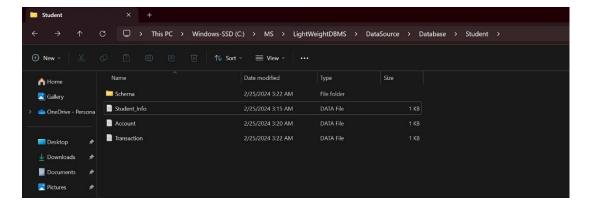


Figure 40 Account.data and Transaction.data generated

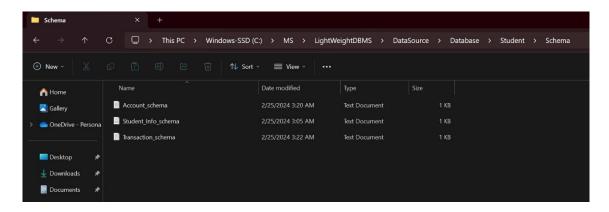


Figure 41 Account_schema and Transaction_schema generated

I added some sample data to the Account table to run the operation on.

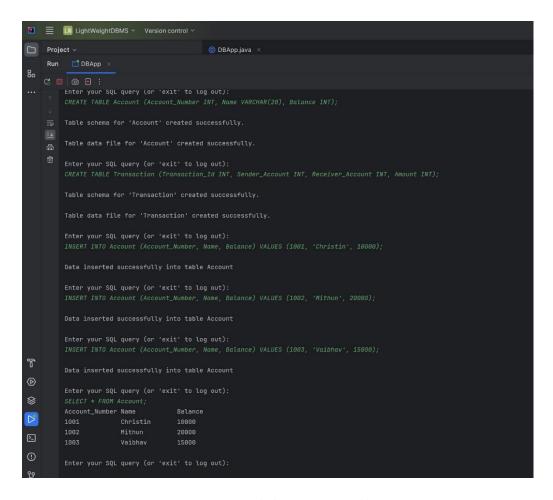


Figure 42 Sample data in Account Table

COMMIT

Query:

BEGIN TRANSACTION;

UPDATE Account SET Balance = Balance - 1000 WHERE Account_Number = 1001;

UPDATE Account SET Balance = Balance + 1000 WHERE Account_Number = 1002;

INSERT INTO Transaction (Transaction_Id, Sender_Account, Receiver_Account, Amount) VALUES (10000, 1001, 1003, 1000);

COMMIT;

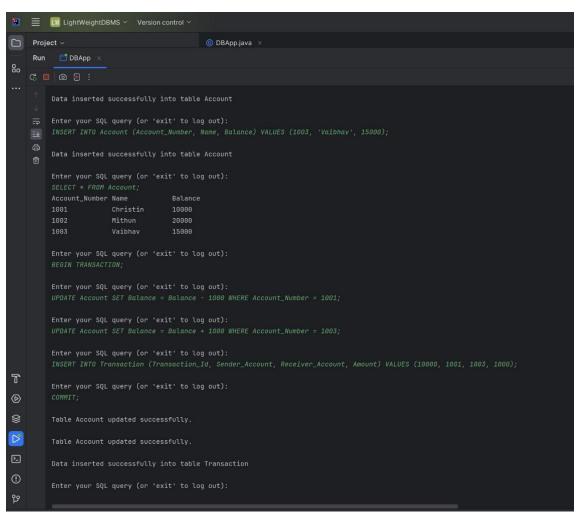


Figure 43 COMMIT operation for the transaction

After the COMMIT command, all the transactions are updated and reflected in the database file.

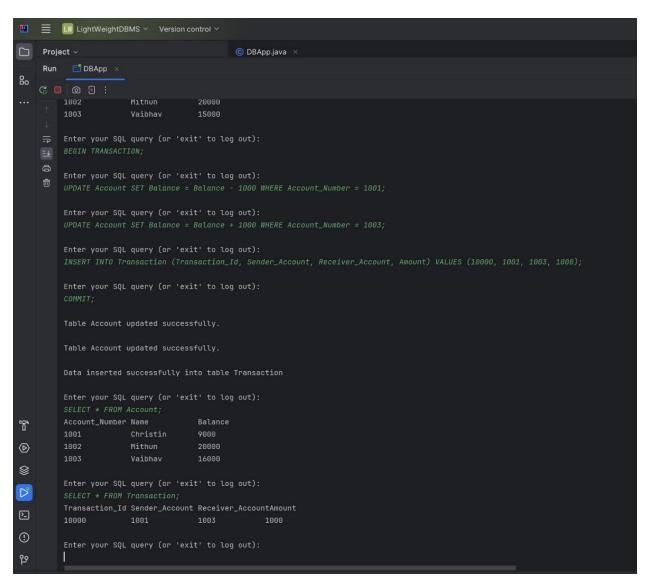


Figure 44 Updated information in the table after successful commit

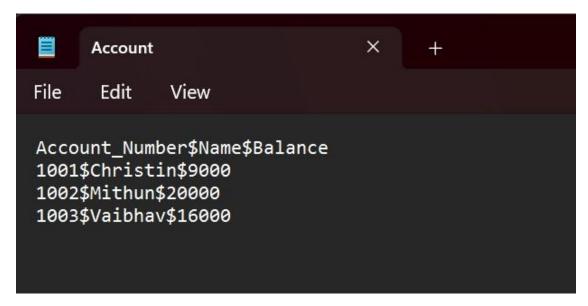


Figure 45 Account table after commit operation

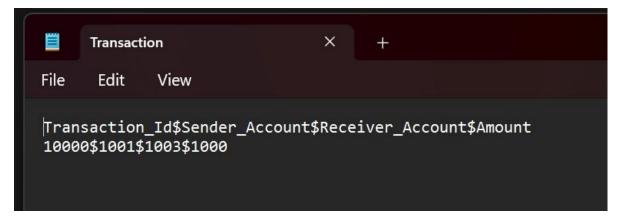


Figure 46 Transaction table after commit operation

ROLLBACK

Query:

BEGIN TRANSACTION;

UPDATE Account SET Balance = Balance - 5000 WHERE Account_Number = 1002;

UPDATE Account SET Balance = Balance + 5000 WHERE Account_Number = 1001;

INSERT INTO Transaction (Transaction_Id, Sender_Account, Receiver_Account, Amount) VALUES (10001, 1002, 1001, 5000);

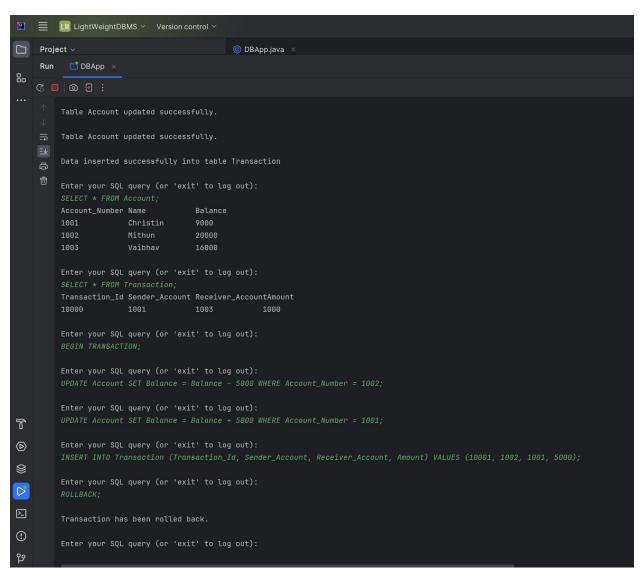


Figure 47 Rollback operation revert back the transaction

As soon as I enter ROLLBACK, all the operations from the point I started the transaction revert without updating the table.

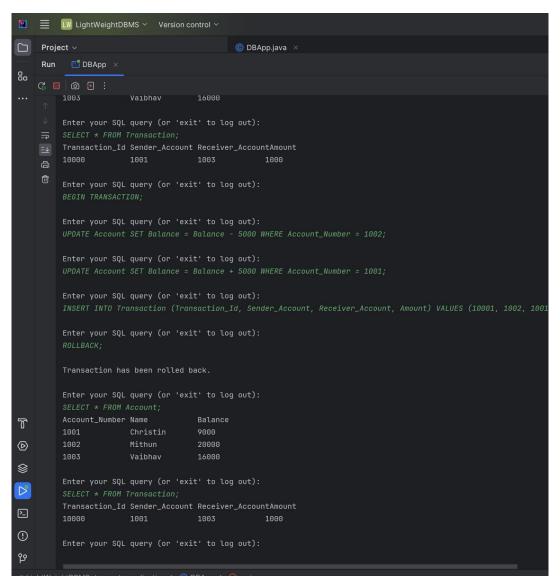


Figure 48 No updates on the Account and Transaction table after rollback

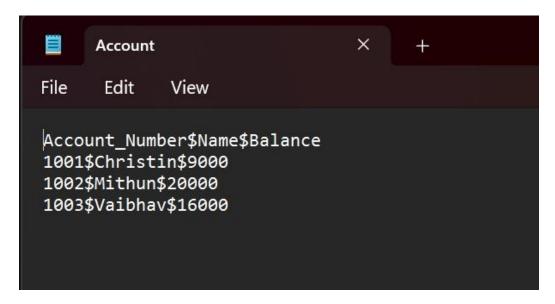


Figure 49 No update on Account table after rollback

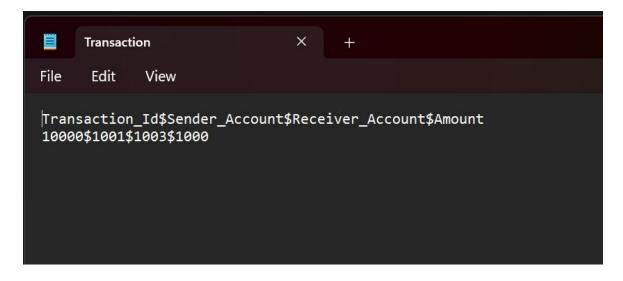


Figure 50 No update on Transaction table after rollback

References

- [1] Sardar Mudassar Ali Khan, "Layered Architecture Used in Software Development," *DEV Community*, [Online], June 14, 2023. Available: https://dev.to/sardarmudassaralikhan/layered-architecture-used-in-software-development-8jd [Accessed: February 24, 2024].
- [2] Samuel Oloruntoba, "Solid: The First 5 Principles of Object Oriented Design," *DigitalOcean*, November 30, 2021. Available: https://www.digitalocean.com/community/conceptual-articles/s-o-l-i-d-the-first-five-principles-of-object-oriented-design [Accessed: February 24, 2024].