

2023 Fall CS504 Project

1. Introduction

Welcome to the final project for *CS504 Principles of Data Management and Mining*. This project aims to test your understanding of the concepts covered throughout the course, such as database design, implementation, and querying. You will be required to develop a functional database system for a given scenario, ensuring that the data is efficiently stored and easily accessible.

2. Project Scenario

2.1. Overview

The objective of this project is to design and implement a database management system for a public library. The system will help the library staff manage their resources, including books, magazines, digital media, and other materials. Additionally, it will provide efficient access to library member's information and facilitate borrowing and tracking of library materials. The library management system should ensure data integrity and minimize redundancy for both library staff and members.

2.2. Features

- **Materials Management:** The system should store and maintain information about all library materials, such as *books*, *magazines*, *e-books*, and *audiobooks*, including their *titles*, *authors*, *publication dates*, and *genres*.
- **Membership Management:** The system should store and manage information about library members, including their *names*, *contact information*, *membership numbers*, and *borrowing history*.
- **Borrowing:** The system should facilitate the borrowing process, allowing members to check out items, and providing library staff with the necessary information to manage the circulation of library materials. Once a material is checked out, a librarian should record its borrow date, anticipated due date. And once the material is returned, its return date should be updated.
- **Reporting and Analytics:** The system should generate reports on library usage, popular materials, and other relevant statistics, enabling the library staff to make data-driven decisions about resource acquisition and management.

2.3. Entities and Relationships

1. Material

Represents individual items available in the library, such as books, magazines, e-books, and audiobooks.

Attributes:

- Material ID: A unique identifier for each material.
- Title: The title of the material.
- Publication Date: The date of publication of the material.
- Catalog ID: A reference to the catalog entry for the material.
- Genre ID: A reference to the genre of the material.

2. Catalog

Represents a record of library materials with information on their availability and location.

Attributes:

- Catalog ID: A unique identifier for each catalog entry.
- Name: The name of the catalog.
- Location: The location of the material within the library.

3. Genre

Represents the various genres or categories of library materials.

Attributes:

- Genre ID: A unique identifier for each genre.
- Name: The name of the genre.
- Description: The brief introduction of the genre.

4. Borrow

Represents the borrowing activity of library materials by members.

Attributes:

- Borrow ID: A unique identifier for each borrowing transaction.
- Material ID: A reference to the borrowed material.
- Member ID: A reference to the member who borrowed the material.
- Staff ID: A reference to the staff who processed the transaction.
- Borrow Date: The date the material was borrowed.
- Due Date: The date the material is due.
- Return Date: The date the material is returned.

5. Author

Represents authors who have created library materials.

Attributes:

- Author ID: A unique identifier for each author.
- Name: The name of the author.
- Birth Date: The birth date of the author.
- Nationality: The nationality of the author.

6. Authorship

Represents the relationship between authors and the materials they have created.

Attributes:

- Authorship_ID: A unique identifier for each authorship record.
- Author_ID: A reference to the author.
- Material_ID: A reference to the material authored.

7. Member

Represents library members who can borrow and reserve materials.

Attributes:

- Member_ID: A unique identifier for each member.
- Name: The name of the member.
- Contact_Info: Email address (or phone number) of the member.
- Join_Date: The date the member joined the library.

8. Staff

Represents library staff who manage library resources and assist members.

Attributes:

- Staff_ID: A unique identifier for each staff member.
- Name: The name of the staff member.
- Contact_Info: Email address (or phone number) of the member.
- Job_Title: The job title of the staff member (e.g., librarian, assistant librarian).
- Hire_Date: The date the staff member was hired by the library.

3. Requirements

3.1. Database Design

1. Define the scope of the project and identify the entities and their relationships.
2. Create an Entity-Relationship (ER) diagram to represent the database schema.
3. (Optional) Normalize the database schema to ensure data integrity and minimize redundancy.

3.2. Database Implementation

1. Choose an appropriate Database Management System (DBMS) for the project. We recommend using Postgre.
2. Implement the database schema using SQL or the chosen DBMS's data definition language (DDL).
3. Populate the database with sample data.

3.3. Querying and Manipulation

Example queries are included in the Tests section below.

1. Develop a set of SQL queries or stored procedures to perform common tasks, such as searching, updating, inserting, and deleting records.
2. Demonstrate the use of advanced querying techniques, such as joins, aggregation, and subqueries.

3.4. Documentation and Presentation

1. Prepare a project report that covers the design and implementation
2. Include the ER diagram, SQL code (DDL and DML), and sample output for each query.

4. Tests and Evaluation

Your final project will be evaluated based on the following criteria:

1. (30pts) Database Design: Correctness and completeness of the ER diagram and constraints.
2. (10pts) Implementation: Functionality of the implemented SQL to define the database schema and insert sample data.
3. (40pts) Querying and Manipulation: Correctness of SQL queries or updates. Reasonable system designs.
4. (20pts) Documentation: Quality of the project report.

4.1. Queries/Updates

1. Which materials are currently available in the library? If a material is borrowed and not returned, it's not considered as available.
2. Which materials are currently overdue? Suppose today is 04/01/2023, and show the borrow date and due date of each material.
3. What are the top 10 **most borrowed** materials in the library? Show the title of each material and order them based on their **available counts**.
4. How many materials has the author Lucas Piki written?
5. How many materials were written by two or more authors?
6. What are the most popular genres in the library ranked by the total number of borrowed times of each genre?
7. How many materials had been borrowed from 09/2020-10/2020?
8. How do you update the "Harry Potter and the Philosopher's Stone" when it is returned on 04/01/2023?
9. How do you delete the member Emily Miller and all her related records from the database?
10. How do you add the following material to the database?

Title: New book

Date: 2020-08-01

Catalog: E-Books

Genre: Mystery & Thriller
Author: Lucas Luke

4.2. Design

You are required to explain how to extend the existing database system to incorporate the following features. While you can provide SQL statements to illustrate your concepts, you don't need to execute the statements.

1. Alert staff about overdue materials on a daily-basis?
2. Automatically deactivate the membership based on the member's overdue occurrence (\geq three times). And reactivate the membership once the member pays the overdue fee.

5. Submission Guidelines

1. **Honor Code:** You are to work on this project alone and not in groups. All projects will be checked for plagiarism against each other and against the Internet using software. Suspicious projects will be sent to the honor committee for further review.
2. Submit your project report in **PDF (or docx) format**, including the description, ER diagram, constraints, answers to Section 4, including sample output.
3. Submit a compressed file in **tar (or zip) format**, containing the report, the project code, sample data, instructions to load/execute the code, and any additional resources used.
4. **Postgre** compatible SQL statements are preferred.

Please feel free to reach out to the instructor or TA if you have any questions or need clarification on any aspect of the project. Good luck!

6. Data

There are eight independent csv files consisting of all data required for this project. You may insert the data into your database using any means as long as the content remain the same. You can remove quotes (") or commas (,) if necessary, other characters are not allowed to change.

- Borrow.csv
- Genre.csv
- Catalog.csv
- Authorship.csv
- Author.csv
- Material.csv
- Staff.csv
- Member.csv