LABORATORY MANUAL

CZ2007: Introduction to Databases

Implementation of a Database Application

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

NANYANG TECHNOLOGICAL UNIVERSITY

1. OBJECTIVES

Upon completion of the assignment, the student should be able to:

a. Construct an entity-relationship model at a conceptual level.
b. Map the model into a schema of a relational DBMS.
c. Implement the given schema on a relational DBMS.
d. Use a database language (SQL) to retrieval data from a relational DBMS.

2. LABORATORY

This is a team-based assignment. Each team consists of **six to seven** members from your laboratory group. (Do note that your lab supervisor may prefer a smaller team size.) Please inform your lab supervisor at the start of Lab 1 if you are selecting your own team members. The lab supervisor may also form team, add, or remove members from your teams. The final members of your team must be submitted to the lab technician during your **first laboratory session**.

There are five scheduled lab sessions for this team assignment. Laboratory sessions will start from the **third week** of the semester. For students whose lab sessions are scheduled on odd weeks, the first lab session is on Week 3. For students whose lab sessions are scheduled on even weeks, the first lab session is on Week 4.

You might need more than the scheduled five sessions for the actual implementation. You are encouraged to **start early** with your lab work (as soon as the topic is covered in the lectures).

Attendance is taken for the **first**, **third** and **fifth** lab sessions only. Attendance for the second and fourth lab sessions is not mandatory. Failing to sign-in for the first, third, or fifth lab session may result in F grade for the respective assessment.

To encourage fair contribution to the lab work, each team needs to indicate contributions from each member. The final mark of a team member may be adjusted based on the team score and individual contributions. **Appendix C is to be submitted with each submission**.

For each lab session, there is a lab supervisor and a lab technician assisting you. The lab supervisor is a professor or a teaching assistant whom you may approach for clarifications on lab work, lab report submission, graded lab reports, etc. The lab technician is a technical staff whom you may approach for lab logistics (lab attendance, SQL Server account matter, lab submission deadline, computer problem, etc.)

3. INTRODUCTION

The assignment covers the portion of the course concerning data modelling, database design and implementation from the user's viewpoint. Thus, the assignment involves modelling as well as implementation aspects of the database course.

The overall aim of the laboratory is to develop an application based on a given data model using a given database management system. This exercise will bring you through a crucial first part of the life cycle of a database application. It is assumed that the data analysis has been performed. Note that this manual provides you with more information than is required for the first laboratory session, e.g., not all constraints can be modelled in the beginning but are included at a later implementation stage. In contrast you might require additional information for an understanding of the application. Proceed by stating your assumptions in written form and / or ask your laboratory supervisor.

4. DESCRIPTION OF THE ASSIGNMENT

The description of the application is given in Appendix A and B. This includes background and general requirements of the application, conceptual information about the system and its users as well as a list of SQL queries that must be fulfilled as a minimum. Note that teamwork is required. Each team will submit one solution. **No individual submission will be accepted.**

4.1 First Laboratory Session: Creating an ER Diagram

Appendix A gives conceptual information about the project obtained after a partial system analysis was performed. Based on the description, construct a suitable ER diagram. Analyze the choice of entity sets, different types of relationships required, the usage of weak entity sets, subclasses, etc. and compare them with alternative solutions from your team members. You need to submit the following, latest **three working days** after the first laboratory session:

- A PDF document of your ER diagram, including a written discussion of your solution (maximum one page) that highlights the reasons for the chosen design. Marks will be given for neat presentation of your ER diagram. Label the PDF document as: Lab1_XXX_TeamY.pdf, where XXX is your lab group number and Y is your team number.

- Assessment for Lab 1 is based on whether the submitted ER diagram reflects correct understanding of ER diagram artefacts (entity sets, relationships, weak entities, subclasses, etc.) and whether they are used correctly and appropriately. Do note that not all information can be represented in an ER diagram.

4.2 Second Laboratory Session: Finalization of the ER Diagram

There is no submission for the second laboratory session. In this lab, each team should finalize their database design based on the feedback received from their lab supervisor. Please note that the second laboratory session is a free access session, i.e., attendance is not mandatory.

4.3 Third Laboratory Session: Generation of Normalized Database Schema

In this lab, you convert the ER diagram into relational schema and ensure that the relations are at least in 3NF. Follow the general guidelines covered during the lectures and tutorials to produce suitably normalized relations. For each relation, the key(s), primary key, and functional dependencies must be specified. If a relation is generated due to the normalization of an original relation, then the normalization steps must be presented. You need to submit the following, latest **three working days** after the third laboratory session:

- A PDF document of the normalized database schema and FDs associated with each relation. Label the PDF document as: Lab3_XXX_TeamY.pdf, where XXX is your lab group number and Y is your team number. If a relation created from the ER diagram violates 3NF, then this should be highlighted along with the decomposed normalized relations. Note that for this lab, no SQL code should be submitted. Hence, the structure of your solution shall be similar to the following example:

  R1(A, B, C, D)
  Keys: AB, AD
  Primary Key: AB
  FDs: AB → CD, A → D
  The relation is in 3NF. (If relation is not in 3NF, perform the steps of the 3NF normalization.)

- Assessment for Lab 3 is based on whether the submitted report reflects correct understanding of keys in relations, identification of appropriate functional dependencies in each relation, how normalized relations are formed, and whether the normalizations are correctly and appropriately performed.

4.4 Fourth Laboratory Session: Implementation of the database schema

There is no submission for the fourth laboratory session. In this lab, the finalized database schema is to be implemented using SQL DDL commands. Your tables should be appropriately populated with sufficient realistic records using SQL INSERT statements so that your query solution for Appendix B results in some output records (3 to 5) for each query. Your implementation should clearly incorporate the primary and foreign keys, data types, and any form of constraints. The lab provides MS SQL Server software for your implementation.

Please note that the fourth laboratory session is a free access session, i.e., attendance is not mandatory.

4.5 Fifth Laboratory Session: Final demonstration

In this lab, the implementation obtained from the previous laboratory session must now be extended to provide SQL query solutions for the queries in Appendix B. **At the end of the lab session**, you need to submit a single PDF document containing the followings:

- SQL DDL commands for table creation (Lab 4).
- SQL statements to solve the queries in Appendix B and additional queries. Each query should be immediately followed by the query output. Briefly explain how the output is obtained.
- A printout of all table records.
- Description of any additional effort made.

Label the PDF document as: Lab5_XXX_TeamY.pdf, where XXX is your lab group number and Y is your team number. You should prepare the PDF document in advance before coming to the lab. Some DDL commands may look like this:

```
CREATE TABLE name (
        attr1 datatype NOT NULL,
        attr2 datatype,
        ...
        PRIMARY KEY (attr1),
        FOREIGN KEY (attr3) REFERENCES name(attr1)
                ON DELETE ... ON UPDATE ...,
);
```

In addition to the PDF document, you are to capture screen recording of query execution as a mp4 video file. For each query in Appendix B and additional queries, first show the SQL statement, then execute the query and show the query results, all recorded as a mp4 video file. Each query video should be no more than 30 seconds and labeled as: Lab5_XXX_TeamY_Q#.mp4 where # is the query number. Zip the PDF and all mp4 files into one single ZIP file.

During the lab session, you may be given additional queries to solve. In addition, your lab supervisor may require in-person live demonstration and Q&A. If required, all team members are to actively contribute during the demonstration session and be familiar with all aspects of the project. No slides presentation required.

APPENDIX A: APPLICATION DESCRIPTION

Suppose you are asked to construct a database for Shiokee, an online shopping website. The requirements are as follows:

- The Shiokee website hosts a number of shops, each of which has a unique name. Each shop sells a number of products, each of which has a name, a category, a maker, a price, and the quantity in stock.
- Each shop gives a unique ID to each of the products it sells. The same product, e.g., iPhone 13x, may have different IDs under different shops.
- The same product could be sold at multiple shops at difference prices. In addition, the price of a product in a shop may change over time. We need to record the history of price changes.
- Shiokee allows users to place orders from the shops. Each user has an ID number and a name. Each order has an ID and timestamp. Each order involves one or more products, which could be from different shops. For each product involved in an order, its price and quantity are recorded. Each order has a total shipping cost and a shipping address.
- After an order is made, the user can track the status of the order on Shiokee. Initially, the status of each product in the order is shown to be "being processed". After the shop (that sells the product) ships the product, the status of the product will be changed to "shipped". Once a product is delivered to the user (as reported by the courier), the status of the product is changed to "delivered", and the delivery date is recorded. Within 30 days from the delivery date of a product, the user may return the product for a refund. Once the shop refunds the product, its status will be changed to "returned".
- After a user purchases a product, he/she is allowed to rate and comment on any product once. There are five possible ratings: 1, 2, 3, 4, and 5, with 5 being the highest. The average rating for a product, as well as the number of users that have rated the product, are shown on the web page that displays the product information to the users. In addition, a user can modify his/her ratings and comments anytime.
- Shiokee users are allowed to file complaints on any product and shop. For example, if a user does not receive a product that has been shown to be "delivered" in an order, he/she can file a complaint to Shiokee. If he/she is not happy about a certain shop, he/she can file a complaint. After a complaint is filed, the user can check the status of his/her complaint. Initially, the status of the complaint is set to "pending". After the complaint is picked up by a Shiokee employee, the status is changed to "being handled", and the name of the employee is shown. Once the complaint is addressed, its status is changed to "addressed".
- Shiokee has a number of employees that handles complaints from users. Each employee has an ID, a name, and a monthly salary. Each complaint is handled by one employee.
- The database should support the queries listed in Appendix B.

Note that the information above may not be complete. Some aspects of the database application's details may have been omitted. It is expected that you come up with their own solutions in case of inconsistencies or missing information. However, you have to keep track of these aspects and explain your assumptions in your submitted report. Extensions to the implementation of the basic system are encouraged.

APPENDIX B: QUERIES

1. Find the average price of "iPhone Xs" on Shiokee from 1 August 2021 to 31 August 2021.
2. Find products that received at least 100 ratings of "5" in August 2021, and order them by their average ratings.
3. For all products purchased in June 2021 that have been delivered, find the average time from the ordering date to the delivery date.
4. Let us define the "latency" of an employee by the average that he/she takes to process a complaint. Find the employee with the smallest latency.
5. Produce a list that contains (i) all products made by Samsung, and (ii) for each of them, the number of shops on Shiokee that sell the product.
6. Find shops that made the most revenue in August 2021.
7. For users that made the most amount of complaints, find the most expensive products he/she has ever purchased.
8. Find products that have never been purchased by some users, but are the top 5 most purchased products by other users in August 2021.
9. Find products that are increasingly being purchased over at least 3 months.

APPENDIX C: INDIVIDUAL CONTRIBUTION FORM

Name	Individual Contribution to Submission 1 (Lab 1)	Percentage of Contribution	Signature

Name	Individual Contribution to Submission 2 (Lab 3)	Percentage of Contribution	Signature

Name	Individual Contribution to Submission 3 (Lab 5)	Percentage of Contribution	Signature