



LABORATORY MANUAL

SC2207/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. INSTRUCTIONS

- a) **Team Formation**: This is a team-based assignment. Each team consists of **five to seven** members from your laboratory group, to be approved by your lab supervisor, to be formed during the Lab 1 session. The lab supervisor may add or remove members from your teams to ensure an even spread and mix of students in each team. The final members of your team must be submitted to the lab executive during the **Lab 1 session**.
- b) **Lab Submission**: There are five scheduled lab sessions for this team assignment. Laboratory sessions will start from the **Week 3** for those scheduled on odd weeks, or **Week 4** for those scheduled on even weeks. For Lab 1, 3 or 5 submissions, do include a **cover page** indicating the team number and team members. Names of team members must appear as they do in student cards; do not shorten or use nicknames or aliases.
- c) **Lab Attendance**: Attendance is taken for the **first, third and fifth** lab sessions only. Attendance for the second and fourth lab sessions is not mandatory.
- d) **Fair Participation**: Each student is expected to make fair and equal contribution to EACH lab, and thoroughly understand the expectations of EACH lab. E.g., it is NOT acceptable that a student contributes less to Lab 1 and make up for it by contributing more to Lab 5. Each submission needs to indicate contributions from each member. The final marks of a team member may be adjusted based on the team score and individual contribution. **Appendix C is to be submitted with each submission.**
- e) **Lab Supervisors**: For each lab session, there is a lab supervisor and a lab executive 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 executive is a technical staff whom you may approach for lab logistics (lab attendance, SQL Server account matter, lab submission deadline, computer problem, etc.).
- f) **To Note**: You are NOT allowed to use AI tools such as ChatGPT or similar tools to do your lab work. If your submission is detected to use AI tools, you may receive zero mark for your lab work.

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. A good ER diagram is one that is self-explanatory. If you believe certain parts of your ER diagram need explanation, you can include a written description (maximum one page). Combine both the ER diagram and the written explanation (if any) as a single PDF document, labeled as follows: Lab1_XXX_TeamY.pdf, where XXX is your lab group and Y is your team number. Marks are given for neat presentation of your ER diagram.
- 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 given in Appendix A can be represented in an ER diagram and more than one ER diagrams

are possible. It is part of the project work for your team to submit what your team deems to be the best ER diagram among all possible ones.

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 and prepare for Lab 3. 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 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: $AB \rightarrow CD$, $A \rightarrow D$

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. Do note that in your final set of relations, the keys and functional dependencies in each relation may not be explicitly given in the description in Appendix A.

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 sufficiently realistic records using SQL INSERT statements so that your query solution for Appendix B results in some meaningful 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. You should start to work on the queries in Appendix B.

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 (from 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 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**_Team**Y**_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. All team members are to actively contribute during the demonstration session and be familiar with **all aspects** of the project. No slide presentation is required.

APPENDIX A: APPLICATION DESCRIPTION

Suppose you are asked to construct a database for Ahamazon, a hardcopy book and magazine supplier hosting an e-commerce website for online customers to purchase books and magazines. The requirements are as follows:

1. Ahamazon supplies publications (books and magazines) to many bookstores. These bookstores are hosted on Ahamazon's website. Each bookstore has a unique company ID. Each bookstore sells a selection of publications, each of which has a publication ID, year of publication, publisher, selling price, and quantity in stock. For a book publication, there is a book title. A magazine publication has an issue number in addition to a magazine title.
2. Each bookstore gives a unique ID to each publication it sells. The same publication may have different IDs under different bookstores. The same publication can be sold at multiple bookstores at different prices. In addition, the price of a publication in a bookstore may change over time. We need to record the history of price changes.
3. Ahamazon's website allows customers to place purchase orders. Each customer has an ID and a name. Each order has an ID and timestamp. Each order has one or more publications, which could be from different bookstores. For each publication in a purchase order, its price and quantity are recorded. Each order has a total shipping cost and a shipping address.
4. After a purchase order has been submitted, the customer may track the status of the order on Ahamazon's website. Initially, the status of each publication in the order is shown to be "being processed". After the respective bookstore ships the publication, the status is changed to "shipped". Once a publication is delivered to the customer (as reported by the courier), the status is changed to "delivered", and the delivery date is recorded. Within 30 days from the delivery date of a publication, the customer may return the publication for a refund. Once the bookstore refunds the publication, its status is changed to "returned".
5. After a customer purchases a publication, he/she is allowed to rate and comment on the publication once. There are five possible ratings: 1, 2, 3, 4, and 5, with 5 being the highest. The average rating for a publication, as well as the number of users that have rated the publication, are shown on the web page that displays the publication information to customers. In addition, a customer can modify his/her ratings and comments anytime.
6. Ahamazon customers are allowed to file complaints on any publication and bookstore. For example, if a customer did not receive a publication whose status is "delivered", he/she may file a complaint to Ahamazon. If he/she is not happy about a certain bookstore, he/she may file a complaint. After a complaint is filed, the customer 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 Ahamazon 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".

7. Ahamazon has a number of employees that handles complaints from customers. Each employee has an ID, a name, and a monthly salary. Each complaint is handled by one employee.
8. 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 solution(s) 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 "Harry Porter Finale" on Ahamazon from 1 August 2022 to 31 August 2022.
2. Find publications that received at least 10 ratings of "5" in August 2022, and rank them by their average ratings.
3. For all publications purchased in June 2022 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 publications published by Nanyang Publisher Company, and (ii) for each of them, the number of bookstores on Ahamazon that sell them.
6. Find bookstores that made the most revenue in August 2022.
7. For customers that made the most number of complaints, find the most expensive publication he/she has ever purchased.
8. Find publications that have never been purchased by any customer in July 2022, but are the top 3 most purchased publications in August 2022.
9. Find publications that are increasingly being purchased over at least 3 months.

APPENDIX C: INDIVIDUAL CONTRIBUTION FORM

Full Name	Individual Contribution to Lab 1 Submission	Percentage of Contribution	Signature

Full Name	Individual Contribution to Lab 3 Submission	Percentage of Contribution	Signature

Full Name	Individual Contribution to Lab 5 Submission	Percentage of Contribution	Signature