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 **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. It should NOT be the case that a student contributes less to Lab X and make up for it by contributing more to Lab Y. 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) **AI tools**: NTU's policy on the use of AI tools: "The University requires students to (i) identify any generative AI tools used and (ii) declare how the tools are used in submitted work. Please note that

even with acknowledgement, copying of output generated by AI tools (in part or whole) **may still be regarded as plagiarism**." **Appendix D is to be submitted with each submission**.

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, to the NTULearn course site for your lab group:

- 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 that 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, to the NTULearn course site for your lab group:

- 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 that is 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. 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 to the NTULearn course site for your lab group containing the followings:

- SQL DDL commands for table creation (from Lab 4).
- SQL statements to solve the queries in Appendix B and additional queries (if any). 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_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. 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

TotalWealth is a licensed digital investment company that provides a mobile app acting as a digital financial advisor to help people manage their wealth. The mobile app is the client-facing tool for their proprietary investment management system called TotalInvest. You have been asked to build the database for TotalInvest to capture all the data necessary for the operations of the mobile app and TotalInvest.

1. When an investor signs up, the app requests for the investor's phone number, date of birth, name, and gender. The phone number is used as the investor's login identifier. Subsequent investor information such as email, annual income, company, etc. will be requested and recorded in TotalInvest's database system.

2. The app takes the investor through a questionnaire where five questions are asked. The investor's risk tolerance is determined from the answers to the questions. Risk tolerance can be "conservative", "moderate", and "aggressive". People in the 20s to 30s age group tend to be aggressive and are willing to take risk. People in the 40s to 50s are generally moderate in their risk tolerance, and those 60s or higher tend to me more conservative. It is a regulatory requirement that investment platform providers such as TotalWealth explicitly seeks investor's risk tolerance before commencing any investment.

3. The app asks the investor to determine one or more financial goals. Examples of financial goals are: (1) to fund children's education, (2) to fund holiday vacations, (3) to buy a car, (4) to buy a house, (5) to generate sufficient monthly cash after retirement, etc. For each goal, the timeline and the amount of money required to achieve the goal is determined.

4. For each financial goal determined by the investor, TotalInvest generates a portfolio of assets for the investor and asks the investor how much he/she would like to invest in that portfolio. Assets are generally a mix of funds, stocks, bonds, commodities, and cash. There are generally less than 10 assets in a single portfolio, the asset allocation ratio determines how the user's invested money is be distributed among the assets. Clearly, the asset allocation ratios for all assets in a portfolio should sum to one (or 100 percent). Generally, an investor may have one or more active portfolios with TotalWealth.

5. Examples of stocks are Apple (AAPL), Microsoft (MSFT), and Tesla (TSLA). Investors use measures such as price/earnings ratio (P/E ratio), EBITDA, earnings per share (EPS), etc. to evaluate stocks. Examples of bonds are 10-Year Treasury Note (TMUBMUSD10Y), iShares iBoxx $ High Yield Corporate (HYG), SPDR Bloomberg Convertible Securities (CWB), etc. Investors use measures such as interest rate/dividend yields, maturity date, etc. to evaluate bonds. Examples of funds are Vanguard S&P 500 ETF (VOO), iShares Core MSCI Emerging Markets ETF (IEMG), Schwab US Small-Cap ETF (SCHA), etc. Investors use measures such as dividend yields, expense ratio, etc. to evaluate funds.

6. Once the investor puts money into a portfolio, TotalInvest operationalizes the portfolio by buying the required amount of each asset from the market through post-trade companies like Saxo, Clearstream, Interactive Broker, etc. As there are different assets in a portfolio, more than one post-trade companies

may be used in one portfolio. For each asset in the portfolio, buying the asset from the market is considered a transaction and is to be recorded with each asset.

7. Once operationalized, the user's portfolio is now active. It will accrue capital gain or loss regularly over time. The amount of money invested is the invested value of the portfolio. The latest accrued unrealised gain/loss of the portfolio together with its invested value is the market value. The user can use the mobile app to check the performance of the portfolio anytime; whether the portfolio is gaining or losing money. The gains and losses since the inception of the portfolio is used to compute the annualized returns of the portfolio, which is a percentage.

8. The user may top up the amount of money invested in a portfolio if he/she thinks the portfolio will perform in the future. The user may also withdraw any amount of money from the portfolio up to the market value of the user's invested sum of money. If it is a full withdrawal, the portfolio will be closed. Note that the market value may be higher or lower than the amount of money invested. For each asset, any top up or withdrawal is considered a transaction and is to be recorded with each asset.

9. TotalInvest earns a management fee of 0.88 percent of the user's invested sum in a portfolio on an annual basis. TotalInvest will recommend adjustments to the asset allocation ratios on a quarterly basis to the user. If the user accepts the recommended quarterly re-balancing, the user pays a transaction fee of 0.20 percent of the user's invested sum in a portfolio per adjustment. For each asset, any re-balancing is considered a transaction and is to be recorded with each asset. Both management and transaction fee may change over time depending on market dynamics.

10. The app has an optional ChatGPT-like mode where users may interact with the app through typed text or spoken speech, in English, Chinese, Malay, Tamil, and Singlish. Based on users' input prompts, the app learns and personalizes future conversations according to users' preferences. All learned information is currently stored with ChatGPT and TotalWealth is trying to have the learned information stored locally within its database system in future.

11. From the information collected of app users, the system applies artificial intelligence/machine learning (AI/ML) techniques to categorize and predict investor's behaviour in his/her investment journey via TotalWealth. AI/ML techniques may be used to provide insights into correlation between user's age and the financial goals; between user's gender/age and patterns of funds top up. All the learned results will help improve the user interaction with new and existing users.

12. Your 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.

APPENDIX B: QUERIES

1. Find investors who are making on average a loss across all their portfolios in 2024.
2. Find investors who are seeing an annualized return of more than 10% from their portfolios in 2024.
3. Find the monthly average unrealized gain/loss of portfolios for each month in 2024.
4. What is the top three most popular first financial goals for investors in 2024?
5. Find investors who consistently top up their investment at the beginning of every month (dollar-cost averaging) in 2024 for at least one of their portfolios.
6. Find the most popular financial goals for investors working in the same company and whose age is between 30 to 40 years old.
7. Are male investors in their 20s making more money from their investments than their female counterparts in 2024?

Note: For all the above queries, you must populate your tables with sufficient data to generate concrete result outputs. No query should generate a NIL (empty) output.

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

APPENDIX D: USE OF AI TOOL(S) IN LAB WORK

Each team member should indicate either A or B:

A. I affirm that my contribution(s) to the lab work is my own, produced without help from any AI tool(s).
B. I affirm that my contribution(s) to the lab work has been produced with the use of AI tool(s).

Team member (full name)	Signature	Date	A or B

By signing this form, you declare that the above affirmation made is true and that you have read and understood NTU's policy on the use of AI tools.

If any team member answered B, the team member(s) must indicate and replicate the table below for every instance that AI tool(s) is used:

Name of AI tool	< For example, ChatGPT >
Input prompt	< Insert the question that you asked ChatGPT >
Date generated	
Output generated	< Insert the response verbatim from ChatGPT >
Output screenshots	
Impact on submission	< Briefly explain which part of your submitted work was ChatGPT's response applied >