



Xi'an Jiaotong-Liverpool University

西交利物浦大学

XJTLU Entrepreneur College (Taicang) Cover Sheet

| | | |
|---|---|------------|
| Module code and Title | DTS106TC: Introduction to Database | |
| School Title | School of AI and Advanced Computing | |
| Assignment Title | Assessment Task 001 (CW): Individual Coursework | |
| Submission Deadline | May 16th, 2025 at 11:59 PM | |
| Final Word Count | NA | |
| If you agree to let the university use your work anonymously for teaching and learning purposes, please type " yes " here. | | Yes |

I certify that I have read and understood the University's Policy for dealing with Plagiarism, Collusion and the Fabrication of Data (available on Learning Mall Online). With reference to this policy I certify that:

- My work does not contain any instances of plagiarism and/or collusion.
My work does not contain any fabricated data.

By uploading my assignment onto Learning Mall Online, I formally declare that all of the above information is true to the best of my knowledge and belief.

| Scoring – For Tutor Use | |
|-------------------------|--|
| Student ID | |

| Stage of Marking | Marker Code | Learning Outcomes Achieved (F/P/M/D) (please modify as appropriate) | | | | Final Score |
|--|----------------|---|-------------------------------------|---|---|-------------|
| | | A | B | C | D | |
| 1 st Marker – red pen | | | | | | |
| Moderation – green pen | IM Initials | The original mark has been accepted by the moderator (please circle as appropriate): | | | | Y / N |
| | | Data entry and score calculation have been checked by another tutor (please circle): | | | | |
| 2 nd Marker if needed – green pen | | | | | | |
| For Academic Office Use | | Possible Academic Infringement (please tick as appropriate) | | | | |
| Date Received | Days late | Late Penalty | <input type="checkbox"/> Category A | Total Academic Infringement Penalty (A,B, C, D, E, Please modify where necessary) _____ | | |
| | | | <input type="checkbox"/> Category B | | | |
| | | | <input type="checkbox"/> Category C | | | |
| | | | <input type="checkbox"/> Category D | | | |
| | | | <input type="checkbox"/> Category E | | | |



Assessment 001: Coursework

Due: May 16th, 2025 @ 23:59

Weight: 60%

Maximum Marks: 100

The coursework will be assessed for the following learning outcomes (LO's):

- A. Demonstrate a basic understanding of the design of databases. (**Bloom's Level:** Understand (Comprehension))
- B. Show a fundamental grounding in the operation and usage of database management systems including "hands-on" experience of a basic database management system. (**Bloom's Level:** Apply)
- C. Demonstrate in-depth knowledge of the database language, SQL. (**Bloom's Level:** Remember, Understand, Apply, Analyze, and Evaluate)
- D. Show understanding of the legal processes and implications of creating and maintaining information systems. (**Bloom's Level:** Understand)

Individual/Group: Individual

Late policy: 5% of the total marks available for the assessment shall be deducted from the assessment mark for each working day after the submission date, up to a maximum of five working days.

Risks:

- Please read the coursework instructions and requirements carefully. Not following these instructions and requirements may result in loss of marks.
- The formal procedure for submitting coursework at XJTLU is strictly followed. Submission link on Learning Mall will be provided in due course. The submission timestamp on Learning Mall will be used to check late submission.
- Ensure that each question's answer is written in one place within the coursework. Do not split your response across different sections.

Overview:

The purpose of this coursework is to design and implement a relational database system to solve a business information need. Working individually you will be given a scenario based on a real-world business example where a database would be needed and work your way through the conceptual, logical, and physical designs of a DBMS solution. This will require substantial research of best practices in design and



the legal and ethical standards to which you must adhere during design. The skills required in this assessment will be valuable in the role of a DBMS professional or an IT manager, as these individuals are often tasked with developing solutions to various organizational data problems while also adhering to legal, ethical, and financial considerations.

There are the following parts of this coursework.

1. Requirements Gathering, Description and Elicitation (Marks 20)
2. Conceptual design using Entity-Relationship Diagram (ERD). (Marks 15)
3. Logical design using the relational model. (Marks 15)
4. Physical design (Create tables, insert sample data into the created tables). (Marks 20)
5. SQL queries that can be run against the database. (Marks 20)
6. Understanding user privacy and legal implication. (Marks 10)

You will be required to write a brief report for each component and note down your process, thoughts, and assumptions made. You will also need to construct a database and produce a set of queries that can be run against that database.

You are required to submit the report along with the queries executed on **PostgreSQL**. For each question, your report should include a detailed description of every step taken during the process. Please ensure all details are included, as missing information may result in mark deductions.

For the explanation part of the question (specifically Question 1 and Question 4), **highlight key points** using methods like **bolding**, *italicizing*, underlining, or changing the font color. Combine these with **bullet points** to enhance readability. Ensure the theoretical part for each question remains within **given words or page limit**. Content that surpasses this limit, without a valid and reasonable justification, will not be considered for evaluation.

Marking Criteria

The coursework will be graded out of **100 marks** and contributes **60%** of the overall credit for the module. Please see the assessment rubric at the end of this document outlining the criteria for assessment.

Your final report should be a complete, polished artefact that incorporates all the necessary detail from each of the components. This is an opportunity for you to pull all of your work from the term together into one complete project.

Problem Background

Imagine you're part of a fast-growing online movie platform that's quickly becoming the go-to destination for movie lovers worldwide. People are flocking to your platform not just for the latest blockbusters, but for hidden gems from every genre imaginable. Your platform is rapidly growing, and users are starting to expect more than just a simple movie list – they want detailed information, ratings, and feedback to make decisions about what to watch next.

But there's a challenge. As your database of movies expands, keeping track of key information like movie

titles, ratings, user reviews, and genres becomes more complex. Your current database system just can't keep up – it's disorganized, hard to manage, and doesn't allow for personalized experiences.

You've been brought on board to assist the team in developing a comprehensive movie database. Your role spans from the very beginning—gathering the requirements and understanding the needs of the platform (Requirements Gathering & Eliciting) — to the more technical aspects, such as crafting efficient SQL queries that extract valuable insights. You'll be working on building a robust system that can handle large volumes of data and deliver key information about movies, ratings, comments, and more, ensuring the platform performs at its best. Your contributions will directly impact the platform's ability to manage and display data in a meaningful and efficient way.

Q1: Requirements Description (Marks 20)

You will begin the project by working on the above given topic. You will identify your company's business requirements by doing some search / research. Identify the business requirements that will allow you to understand the business processes. Build a list of business needs, rules and assumptions based on your scenario. Use the following categories to help you with this:

- **Business Scenario:** A business scenario describes a specific situation or context in which a business operates, including its processes, requirements, goals, and challenges. It outlines the need for a solution (e.g., a database) and defines how the solution will address the needs of the business. In the context of database design, the scenario should clarify why the database is necessary, what it aims to achieve, and how it fits into the business workflow. You should clearly state the need for a database and identify its components in paragraphs. Why its important to design a database instead of spreadsheet or file system in the context of problem mentioned? Usually, one paragraph pertains to one or more tables and relationships.
- **Business rules and assumptions:** Business rules and assumptions are foundational elements in database design that help define how a business operates and how its data is structured, managed, and processed. These elements ensure that the database accurately reflects the real-world processes and constraints of the organization. It is used to understand business processes and the nature, role, and scope of the data. For Example, i) A customer cannot place an order without registering in the system; ii) Each product have a unique product ID etc.
- **Problems and possible solutions:** In the context of database design, problems and possible solutions refer to the challenges that arise due to various legal, ethical, financial, or operational considerations that need to be addressed to ensure the database functions effectively within the given business environment. Identifying these challenges early allows the designer to propose practical solutions that minimize risks and optimize performance. These problems can be defined as legal, ethical, and financial considerations that require attention and a possible solution to alleviate the situation.
- **Functional Requirements:** Functional requirements specify the actions and features a system must perform to meet user needs. For the movie database, the focus will be on features or functionalities like movie management (add, update, delete), search and filter options, to name a few for a



seamless user experience. Write the list of functionalities for the database system you design or gather the requirements.

During the requirements gathering and elicitation phase, it's important to address legal, ethical, and financial factors that may require attention. Identifying potential solutions to these issues is crucial for mitigating any challenges that arise. *Ensure the theoretical part remains within 600-700 words or a maximum of one page for this question. Content that surpasses this limit, without a valid and reasonable justification, will not be considered for evaluation. Also make sure you highlight key points using methods like bolding, italicizing, underlining, or changing the font color. Combine these with bullet points to enhance readability. See the detail rubrics at the end of the coursework.*

Q2: Conceptual Model (Marks 15)

Step 1: Using text analysis on the given scenario identify entities and attributes that will be used to store information about identified entities. Entities are usually the nouns in the scenario, and Attributes are normally found by identifying nouns that describe other nouns.

Step 2: Identify relationships using business rules , write the optionality of the relationship between both entities (remember the relationship exists in both directions) with justification. Determine the cardinality by analyzing the description given for each relationship. If the description uses “a” or “an” then it should be a 1 relation (single toe) however if it says “many” or “one or more” then it should be a M relation. Add the cardinality to the following entities by adding required notation.

Step 3 : Identify the primary key and foreign key in the ER diagram, and explain their roles in establishing relationships between entities.

Step 4: Devise a conceptual model using an **Entity Relationship Diagram (ERD)** that will best address the scenario you are working for the project. Your model should include all necessary entities, relationships, attributes, and business rules with justification. Create a list of assumptions if applicable. The model should be well structured and organized for easy interpretation.

You are also required to write and note down your process, thoughts, and assumptions made (if any). Keep your explanation/reflection part concise, not exceeding five lines. [See the detail rubrics at the end of the coursework.](#)

Q3: Logical Model (Marks 15)

Based on the conceptual model, illustrate a normalized logical model for your DBMS that accurately represents all necessary aspects of the DBMS to address the solution. Use a table instance chart to map ERD into a relational model. The table diagram helps you map out a table before creating it in the database. You should describe the design of each table, by completing the table instance chart for each table mapped.



Your tables' design should correspond to your ERD and must be in the third normal form (**3NF**). State candidate keys and functional dependencies of each table. Explain any assumptions you make applying what you know of the domain to the data and consider future data and the impact it may have as well. You will need to think and determine whether values are 'blank' (a known value of blank) or null (an as yet unknown value) as this may have an impact on your dependencies. Explain any assumptions and decisions you make in the report.

[Here is the sample table instance chart for your reference.](#)

Table name: **CUSTOMERS**

| Column Name | CUSTOMER_ID | LAST_NAME | FIRST_NAME | HOME_PHONE | ADDRESS | CITY | STATE | EMAIL | CELL_PHONE |
|---------------|-------------|-----------|------------|------------|----------|----------|----------|----------|------------|
| Key Type | PK | | | | | | | | |
| Not Null = NN | NN, U | NN | NN | NN | NN | NN | NN | | |
| Unique = U | | | | | | | | | |
| Data Type | NUMBER | VARCHAR2 | VARCHAR2 | VARCHAR2 | VARCHAR2 | VARCHAR2 | VARCHAR2 | VARCHAR2 | VARCHAR2 |
| Length | 30 | 25 | 25 | 12 | 100 | 30 | 2 | 25 | 12 |

The following sections provide a review of the concept of Normal Form, helping to reinforce key principles and ensure a clearer understanding.

- **First Normal Form (1NF):** Ensure that each table has atomic values (no repeating groups or arrays). Break down any multi-valued attributes into separate rows or tables, creating unique rows for each piece of data. For example, each patient can only have one primary doctor, each doctor can only have one speciality etc. **Second Normal Form (2NF):** Ensure that the data is in 1NF and that every non-prime attribute (an attribute not part of a candidate key) is fully functionally dependent on the entire primary key. If partial dependencies exist, move the dependent attributes to separate tables. For example, we need to know each drug's name, purpose and side effects but if we include this in the Prescription entity it will be dependent only on what drug is prescribed not who it's for or what doctor prescribed it - so it does not belong in the same entity as the prescription information itself. **Third Normal Form (3NF):** Ensure that the data is in 2NF and that there are no transitive dependencies. This means non-key attributes should depend only on the primary key and not on other non-key attributes. For example, A patient's insurance ID number will determine what insurance company they are insured with. The ID number determines the insurance company's name.
- In the normalization process, you might create some new entities while resolving functional, transitive and multivalued dependencies. Write down the final list of entities in relational notation.

As you go through these steps, identify if any of the entities are already in 1NF, 2NF, or 3NF. If so, briefly explain why no changes are needed for those entities. Document the modifications you make and provide the final set of normalized tables. Also for each part you are required to write a brief explanation for each component and note down your process, thoughts, and assumptions, and results. [See the detail rubrics at the end of the coursework.](#)



Q4: Physical Model (Marks 20)

Create a physical database design that builds on the conceptual and logical models you crafted.

For the Movie Database, the physical model involves converting the logical structure into actual database tables. This includes defining data types (e.g., VARCHAR for movie titles, INT for movie IDs), specifying primary and foreign keys, and establishing relationships between entities such as Movies, Users, and Rentals. You are required to:

1. Define Appropriate Data Types for Each Column

- Identify the data types for each column in your table(s). Justify the choice of data types for each attribute based on the nature of the data. For example, use VARCHAR for text-based data, INT for numerical identifiers, and DATE for dates.

2. Create Tables with Correct Constraints and Keys

- Write the **DDL** (Data Definition Language) statements to create the necessary tables based on the logical design. Ensure that primary keys, foreign keys, and other constraints (e.g., NOT NULL, UNIQUE) are correctly applied to the tables. Justify the structure and relationships between the tables.

3. Insert Realistic Sample Data

- Write **DML** (Data Manipulation Language) statements to insert sample data into the tables. Ensure that the sample data covers a variety of realistic entries and maintains meaningful relationships between the tables.

4. Ensure Referential Integrity and Constraints

- Apply and explain the use of referential integrity (e.g., foreign key relationships) and other relevant constraints (e.g., NOT NULL, UNIQUE, CHECK) in your table(s). Ensure that these constraints help maintain data integrity when inserting sample data. Justify their use and impact on the system.

5. Provide Justifications for All Choices

- Provide clear and concise explanations for your choices made in data types, table creation, data insertion, and constraints. Explain why these decisions were appropriate for the business requirements and logical design.

Make sure your implementation clearly demonstrates the relationships between the entities.

Write INSERT statements to populate the new tables. Insert enough records (at least 15 rows in each table) so that you can run interesting and nontrivial queries on your database.

- You have to consider the order of the tables when populating them. A table that has a foreign key field cannot be populated before the related table with the primary key
- For each table created in the database take the image of the table resulting query using `SELECT * FROM <table_name>` and show in the report.
- Share all the images of the tables created. The images clearly show the variable name, datatype and other required information.



For each part you are required to write a brief explanation for each component and note down your process, thoughts, and assumptions, results and queries used (if any). [See the detail rubrics at the end of the coursework.](#)

Q5: SQL Queries (Marks 20)

In this task, you are required to write **five** SQL queries, each demonstrating different skills and techniques. For each query, you must explain in simple terms what it is intended to do and provide the corresponding SQL SELECT statement. Additionally, you need to include a screenshot of the query result to confirm that it works.

In this report, please provide a detailed explanation of each SQL query you write. For each query, describe the logic behind your approach and the specific SQL operations you used (such as SELECT, JOIN, GROUP BY, etc.). Explain why you chose those operations and how they contribute to solving the problem or meeting the business requirements. If applicable, include any assumptions you made about the data or the problem. *Additionally, provide a screenshot of the query result to demonstrate that the query works as expected.*

If you submit more than five queries, only the first five will be evaluated. The complexity / design of the queries will influence the marks you receive. **For example**, a query that involves multiple conditions in the SELECT clause will have a lower complexity, while a query that includes multiple JOINs, GROUP BY clauses, and conditions will be rated as more complex. To help you understand how complexity is measured, a SELECT condition counts as 1, a JOIN counts as 1, and a GROUP BY clause counts as 1. For instance, a query that involves two SELECT conditions, one JOIN, and a GROUP BY clause would have a complexity score of 3. Each SQL query mentioned in this part of the coursework must have 4 complexity / design points as discussed.

As an example, consider the following query. Suppose you need to find the average rating of movies rented by each customer, but only for movies released in and after 2018. The query might look like as shown below and the below query is equivalent to 4 complexity points (SELECT 1 + JOIN 1 + JOIN 1 + GROUP BY 1 = 4). (Don't use the below query as part of the answer):

```
1 ✓ SELECT c.customer_name, m.movie_title, AVG(r.rating) AS avg_rating
2   FROM customers c
3   JOIN rentals r ON c.customer_id = r.customer_id
4   JOIN movies m ON r.movie_id = m.movie_id
5   WHERE m.release_year >= 2018
6   GROUP BY c.customer_name, m.movie_title;
```



For each part you are required to write a brief explanation for each component and note down your process, thoughts, and assumptions, results and queries used. [See the detail rubrics at the end of the coursework.](#)

Q6: Privacy and Legal Implications (10 Marks)

In today's digital world, privacy and awareness about data protection have become critical concerns. As you design a movie database system, consider the legal implications of creating and maintaining such an information system.

Key Points to Address:

1. What types of user data will your system collect (e.g., data collected during user registration, browsing history, or card details)?
2. Will your system store this data, and if so, how will it be managed securely?
3. Considering that many countries have laws to protect user privacy, what strategies will you implement to address these legal challenges?

Hint: Explore privacy laws such as the GDPR (General Data Protection Regulation), CCPA (California Consumer Privacy Act), or similar regulations applicable to your target region for your database. Suggest practical steps to ensure compliance and safeguard user privacy. *Ensure the theoretical part remains within 500-600 words or a maximum of one page. Content that surpasses this limit, without a valid and reasonable justification, will not be considered for evaluation. Also make sure you highlight key points using methods like bolding, italicizing, underlining, or changing the font color. Combine these with bullet points to enhance readability.*

For each part you are required to write a brief explanation for each component and note down your process, thoughts, and assumptions, results and queries used (if any). [See the detail rubrics at the end of the coursework.](#)

Report Submission Guidelines

All students must download their file and check that it is viewable after submission. Documents may become corrupted during the uploading process (e.g. due to slow internet connections). However, students themselves are responsible for submitting a functional and correct file for assessments. Only electronic submission is accepted and no hard copy submission.

You should submit final report in PDF, .doc and .zip file at LMO.

1. Submit your final report in both PDF and DOC formats. Name the document as *[Your Student Name_ID].pdf* and *[Your Student Name_ID].doc*. The report must be typed in MS Word and uploaded to the Learning Mall in both formats.



2. A single zip file (named as *[Your Student Name_ID].zip*) will contain:
 - a) ddl.sql: The DDL statements of the normalized database
 - b) data.sql: INSERT statements of the normalized dataset
 - c) sqlScript.sql: SQL statements to run queries against normalized tables

Generic Marking Criteria and Assessment Rubric

The below are generic marking criteria and assessment rubric.

Support

Any questions and answers will be added to the FAQ. Please use the coursework discussion channel at LMO as the first point of call for any questions, problems, clarifications, or anything you would like us to go over. If you prefer to ask privately, please send an email to the relevant instructor.

GOOD LUCK !



Marking Criteria

| Q1-Q6 | 100 | Components | Description | Maximum Credit |
|-------|-----|---|--|----------------|
| Q 1 | 20 | Clarity and Understanding of Business Scenario | 1. Why the database is necessary, what it aims to achieve? 2. How it fits into the business workflow? 3. You should clearly state the need for a database and identify its components in paragraphs. | 0 - 3 |
| | | Motivation for Database Design | Why its important to design a database instead of spreadsheet or file system in the context of problem mentioned? | 0 - 1 |
| | | Business Rules and assumptions Understanding | Analyzing the explanation on how well the student understand the process, roles and scope of data. Also whether he/she understands the processes and constraints of the organization. | 0 - 3 |
| | | Problems and Possible Solutions | Legal, Ethical and financial considerations, explanation and their possible solutions. | 0 - 3 |
| | | Identifying Requirements / Requirements Gathering | At least 6 complete requirements related to the problem. (Requirements 6 * 0.5 Mark) = 4 Marks Proper Explanation - 3 marks Reflection on the process of gathering the Requirements - 3 marks. | 0 - 10 |
| Q 2 | 15 | Identifying Entities Identifying Attributes | Correct Identifying Entities for the ER Diagram (Marks 2) Correct Identifying Attributes for the ER Diagram (Marks 2) | 0 - 4 |
| | | Identification of Cardinality | Correct Identification of Cardinality for the ER Diagram | 0 - 3 |
| | | Identification of FK, PK etc | Correct Identification of FK, PK ,etc | 0 - 2 |
| | | Making and Explanation of ER Diagram | Proper explanation of ER diagram explaining the over all assumptions, constraints etc | 0 - 3 |
| | | Clarity and Reflection | Evaluation of requirements gathered, attentiveness to details, and the effectiveness of requirement mapping. Does the ER diagram clearly present information and make it easy to read and assess? | 0 - 3 |
| Q 3 | 15 | Correctness of Relational Tables | Evaluates the tables (Marks 3), their attributes, and the relationships between them (Marks 3). | 0 - 6 |
| | | Normalization and Integrity | Assesses the level of normalization applied to the design. The tables should be normalized to at least the third normal form (3NF), ensuring the removal of redundancy and ensuring integrity constraints. Also add the detail of the new entities added as the result of normalization. | 0 - 3 |
| | | Appropriateness of Relationships | Focuses on how well the relationships between tables are established. The relationship types (one-to-one, one-to-many, many-to-many) should be correctly implemented using foreign keys, and any complex relationships should be properly handled. | 0 - 3 |
| | | Consistency and Clarity | Evaluates the overall consistency and clarity of the relational model. Such as, Tables should be clearly named with naming convention, with appropriate attributes and constraints. The design should be easy to understand and logically structured. | 0 - 3 |
| Q 4 | 20 | Correctness of Table Creation | Evaluates the accuracy and correctness of the table creation. This includes proper use of data types, table names, primary | 0 - 4 |



| | | | | |
|----|----|---|---|---|
| | | | keys, and foreign keys (Marks 2). The tables should be created according to the specifications of the logical design (Marks 2). | |
| | | Appropriateness of Data Types | Assesses the appropriateness of the data types used for each attribute in the table (Marks 2). Data types should match the nature of the data (e.g., VARCHAR for names, INT for IDs, DATE for dates, etc.). | 0 - 2 |
| | | Inserting Sample Data | Evaluates the correctness and variety of the sample data inserted into the tables (Marks 2). The data should be realistic, covering a range of possible entries and ensuring the relationships between tables are meaningful (Marks 2). | 0 - 4 |
| | | Referential Integrity and Constraints | Checks whether the referential integrity (foreign key relationships) and other constraints (e.g., NOT NULL, UNIQUE, CHECK) are correctly applied to the tables. This ensures that data integrity is maintained when inserting sample data (Marks 2). | 0 - 3 |
| | | Explanation | Proper Explanation / justification of all the above sections. | 0 - 3 |
| | | Reflection of Challenges | Describe the challenges encountered during the table creation process and the strategies used to overcome them. Highlight the key factors or resources that were most helpful in resolving those challenges. Proper explanation of each point. Two reflection points 2 * 2 = 4 Marks Don't exceed in length. Maximum of 500 words or 1 page for both reflection points. | 0 - 4 |
| Q5 | 20 | SQL Query | Each query [0-4 marks] x [0-4] Queries = 16 Each query [0- 4 marks] depending on the complexity. The complexity of a query will be measured by the number of joins, select conditions and Group By clause, where a select condition counts as a '1' and a join count as a '1' and Group By condition count as '1' (so, a SQL query that probably involved two select conditions, one join condition, and Group By clause will be counted as the complexity of '4'). | Each query [0-4 marks] x [0-4] Queries = 0 - 16 |
| | | Explanation, reflection on the query | Explain how the query processes (Marks 2) the results for the above, including reflection / learning (Marks 2) and its usage w.r.t to the movie database problem given. | 0 - 4 |
| Q6 | 10 | Privacy and Legal Implications | Identification of User Data (storage) and proper explanation. | 0 - 3 |
| | | | Data storage (Privacy and legal implications) and proper explanation | 0 - 3 |
| | | | Privacy strategies (Marks 2) and explanation (Marks 2) w.r.t to problem design | 0 - 4 |
| | | Late Submission? Generally, a 5-point penalty will be applied for each day after the deadline. | | |
| | | <input type="checkbox"/> Yes <input type="checkbox"/> No | | |
| | | Final Marks | | |