



Xi'an Jiaotong-Liverpool University

西交利物浦大學

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 002 (CW)): Individual Coursework	
Submission Deadline	May 21st ,2025 at 17:00 PM (GMT +8)	
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	E	
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):					Y
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 <input type="checkbox"/> Category B <input type="checkbox"/> Category C <input type="checkbox"/> Category D <input type="checkbox"/> Category E		Total Academic Infringement Penalty (A,B, C, D, E, Please modify where necessary) _____		



Xi'an Jiaotong-Liverpool University
西交利物浦大學

DTS106TC Introduction to Database Coursework- Assessment Task 002

Due: 17:00, June 3rd 2025

Weight: 40%

Maximum Marks: 100

The course work 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)

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:

In detail, Step 1 (Q1: ER Diagram) requires students to analyze the provided dataset, identifying data types and draw the ER diagram. Step 2 (Q2: Normalization) is undertaken to mitigate redundancy and dependency, thereby ensuring data integrity and enhancing database performance. Subsequently, Step 3 (Q3: Database Construction, Q4: Querying) illustrates the process of populating/exporting data of the Database Management System and executing manipulations on the Database Management System using SQL. Step 5 (Q5: Reflection

Question) helps summarize the learning in class and how it helps understanding and solving different parts of coursework. Lastly, Step 6 (Q6: Reflection Question) helps understand the limitations of relational database.

You will be required to write a brief report to answer the questions and note down your process, thoughts, and assumptions made as well as answers.

There are the following 5 parts of this coursework

1. ER Diagram (Marks 15)
2. Normalization (Marks 15)
3. Database Construction and Data Import (Marks 25)
4. SQL Queries and Relational Algebra (Marks 25)
5. Reflection Question on Database Design (Marks 10)
6. Reflection on Understanding the Limitations of Relational Database Design (Marks 10)

You are required to write a concise report addressing the questions, documenting your thought process, assumptions, and solutions. Ensure that your answers are clearly presented in one cohesive section.

Marking Criteria

This coursework will be graded out of 100 marks and contributes **40%** of the overall credit for the module. There are 6 questions different marks available for every question. Marks will be awarded based on the level of correctness of each answer. For example, 100% for fully correct with required explanation/justification, partial marks based on the level of incorrectness or missing required details. For detail rubrics see at the end of the coursework.

DATASET

The dataset to be used for this assessment is random generated film database. The dataset contains four .csv files *films*, *people*, *review*, and *roles*. The dataset / files can be downloaded from LMO under the Assessment 002-Coursework.

SETUP

You can complete this coursework using PostgreSQL.

Q1: ER Diagram (15 Marks)

In this part of the coursework, you are required to analyze the relationships between the given files (through LMO). Based on your analysis, draw the **Entity-Relationship (ER) Diagram** that represents the relationships and constraints between the tables.

1. Use standard symbols to draw the ER diagram. Identify the entities, their attributes, and describe the relationships between them, including the cardinality (e.g., one-to-many, many-to-many). Additionally, provide a clear explanation of the ER diagram to ensure better understanding.

2. Explain any constraints (e.g., Weak entity constraints, Attribute constraints, Key constraints etc).

Note: Make sure your ER diagram is clear, properly labeled, and shows all the relationships between entities. Explain any assumptions you make. The diagram should provide a comprehensive view of how the tables are interconnected and help understand the structure of the database. Also give a brief explanation of the ER diagram. You are also required to write and note down your process, thoughts, and assumptions made (if any). [See the detail rubrics at the end of the coursework.](#)

Q2: Normalization (15 Marks)

Given the tables in the film database, evaluate whether they are in a normalized form. If they are not, convert them into **Third Normal Form (3NF)**.

1. *Evaluate the Current Structure:* Analyze the given table structure and determine if any table violates 1NF, 2NF, or 3NF. If any violations are found, decompose the table into smaller tables that are free from redundancy and dependency issues.
2. *Normalization to 3NF:* If the tables are not in 3NF, perform the necessary decomposition or normalization to bring them to 3NF.
3. *Identify and Suggest Improvements:* Identify potential issues in the current table design and suggest improvements for achieving better normalization and an optimized database design.
4. *Provide Updated Structures:* After normalization, provide the updated table structures, ensuring they accurately reflect the correct relationships and attributes for each table.

Note: Pay special attention to how data is structured and ensure that all non-key attributes are fully functionally dependent on the primary key. Give brief explanation for each point discussed above. Explain any assumptions you make.

Q3: Database Construction and Importing Data into PostgreSQL (25 Marks)

In this part of the assignment, you are required to create a new database named **film_db** in PostgreSQL. Once the database is created, you will define and create four tables given (See DATASET). For each table, you must define appropriate columns that accurately represent the required information. Ensure that each table has the correct **data types** for each column, and apply **constraints** such as *primary keys*, *candidate keys*, *foreign keys*, and *not null constraints* where necessary to maintain data integrity. Once the tables are defined, you will import data into them from CSV files, ensuring that the data aligns with the table structure and the constraints applied in the previous steps. The successful completion of this task will demonstrate your ability to create a well-structured database and import relevant data into it correctly.

To verify the successful import of data into the tables, execute the SQL query ***SELECT * FROM <table>;*** for each of the tables you have created. This query will display all the rows and columns from the respective tables, allowing you to confirm that the data has been imported correctly. *After running these queries, take screenshots of the results and provide them in your report.* The images should clearly show the output of each query, demonstrating that the tables contain the expected data. Additionally, explain the output of each query in your report, highlighting any important details, such as the correctness of the imported data or any

discrepancies that may require attention. This step ensures that the data import process has been completed successfully and the tables are correctly populated.

Note: Show image of CSV and results generated from the above query side-by-side for comparison. [See the detail rubrics at the end of the coursework.](#)

Q4: Querying and Relational Algebra (25 Marks)

For this question, you need to write SQL queries to perform different operations on a film database that has been imported into a *PostgreSQL* database. The tasks will test your understanding of SQL functions, grouping, and filtering. Some queries might involve multiple steps, while others will be simpler. Additionally, for two of the queries, you are required to write the corresponding relational algebra expressions. The queries to be performed are as follows:

1. Write a query that calculates the average review score for each film and ranks the films from highest to lowest based on this score. Display the top 5 films with the highest average rating.
2. Write a query that returns a list of films that have received more than 5 reviews. (You will need to use grouping and counting functions to achieve this).
3. Write a query that shows the number of films each actor has appeared in. (You will need to join the **Films** and **Roles** tables and count the number of films associated with each actor).
4. For the above Query 1 and Query 2, also write the relational algebra expressions. Explain how relational algebra can be used to represent the operations performed in these queries.

For each query you create, please include the output as an image. Additionally, provide a brief explanation of how the SQL query works. Be sure to also share the SQL query file so that it can be reviewed, executed, and verified. You are also required to write and note down your process, thoughts, and assumptions made (if any). [See the detail rubrics at the end of the coursework.](#)

Q5: Reflection on Database Design (10 Marks)

(This question aligns with Learning Outcome A : "Demonstrate a basic understanding of the design of database". It encourages students to reflect on how the knowledge and skills gained during the learning process have enhanced their understanding and ability to apply key concepts in database design.)

Reflect on any of the one topic covered throughout the Introduction to Database course, mentioning any in-class activities or practical exercises. How have these concepts contributed to your understanding of database design, querying, or management? In your response, please combine the theoretical knowledge gained with any hands-on experience to explain how you applied these learnings in database design? Additionally, identify which concepts were most challenging or intriguing for you, and explain why they stood out. Provide concrete examples from the class / lab to support your reflections.



Provide one example to support your answer. It's important to understand the criteria outlined in the rubric. One key aspect student might be missing is the connection to classroom / lab discussions or class activities / teaching. It is important to ensure that your answer is not simply a random scenario or generic explanation. While online tools and software might help with the answers or provide topic summaries, your response will only be considered correct if it is clearly rooted in the topics and scenarios highlighted / mentioned during our class discussion / sessions or lab activities. Make sure to link your examples directly to the concepts and exercises we've worked on in class / lab for a more comprehensive and relevant response. Highlight important points or claims mentioned in your answer for better readability.

For the explanation of the question, **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 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. [See the detail rubrics at the end of the coursework.](#)

Q6: Reflection on Understanding the Limitations of Relational Database Design (10 Marks)

(This question aligns with Learning Outcomes A, encouraging students to reflect on their experience with relational database design, and related challenges. It helps assess their learning about database design, flexibility, scalability, and handling different data types)

Think about your experience with designing, developing, and querying relational database using SQL in the coursework 1. Reflect on the limitations you encountered while working with simple, structured data (if any). Consider the challenges you might face when dealing with large volumes of data of varying types, particularly in terms of flexibility, scalability, and handling unstructured data. Reflect on how relational databases manage structured versus unstructured data? In your opinion, what types of data, especially those with large volumes or diverse formats, would benefit from a different approach to database design?

For the explanation of the question, **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 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. [See the detail rubrics at the end of the coursework.](#)

In your report, explain any assumptions or decisions you made while performing the assignment. Provide insights on how you approached each example.

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 (PDF) and .doc version at LMO.

- 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.

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	Marks
Q1	15	Identification of Entities, Attributes and Cardinality	Entities (3 marks), Attributes (3 marks), Cardinality (3 marks) in the ER diagram. Brief Explanation (1 mark).	0 - 10
		Mentioning of keys, constraints, and explanation.	Mentioning of keys, and constraints (3 marks), explanation and clear diagram structure (2 marks).	0 - 5
Q2	15	Evaluation of Current Table Structure	Evaluate whether the given tables are in 1NF, 2NF, or 3NF. Identify any normalization issues. Total Tables $4 * 2 = 8$	0 - 8
		Decomposition of Tables (if needed)	If any tables violate normalization forms, decompose them into smaller tables to remove redundancy. Proper explain your point with clarity (either decomposition needed or not). Don't exceed from 200 words.	0 - 2
		Identification of Issues and Suggestions	Identify and discuss potential issues in the current design and suggest improvements for better normalization. Adding more tables as a result of normalization process.	0 - 2
		Updated Table Structures, Normalization Process and	Provide the updated table structures after normalization, showing relationships and attributes.	0 - 3



		Explanation		
Q3	25	Database Creation	Successfully creates the film_db database in PostgreSQL.	0 - 3
		Table Definition and Constraints	Defines 4 tables with appropriate columns, data types, and applies constraints (primary keys, foreign keys, not null). $4 * 2 = 8$	0 - 8
		Data Import and Verification	Imports data correctly into tables from 4 CSV files or other data sources, ensuring it aligns with table structure. $4 * 2 = 8$	0 - 8
		Explanation of Results	Data Verification (2 marks), Image sharing (2 marks) and Explanation / reflection of Results (2 marks).	0 - 6
Q4	25	Query 1	Correctly calculates the average review score for each film and ranks them from highest to lowest. Proper explanation (2 marks) and sharing of Images (2 marks). Also share the SQL query for verification (2 marks).	0 - 6
		Query 2	Correctly uses grouping and counting functions to return films with more than 5 reviews. Proper explanation (2 marks) and sharing of Images (2 marks). Also share the SQL query for verification (2 marks).	0 - 6
		Query 3	Correctly joins tables and counts the number of films each actor has appeared in. Proper explanation (3 marks) and sharing of Images (2 marks). Also share the SQL query for verification (2 marks).	0 - 7
		SQL Algebra	SQL Algebra (2 Marks) * 2 = 4	0 - 4
		Explanation, Clarity and Reflection of learning.	Reflect on your approach to solving the queries, the key learnings from developing and executing them, and the clarity of both the explanation and the query for readability and evaluation.	0 - 2
Q5	10	Reflection on Topics	Reflect on one topic covered in the course. Provide explanations of their relevance and impact on database design, querying, or management.	0 - 3
		Connection to Class Activities /Teaching / Learning	Provide examples from class or lab activities /teaching / learning that contributed to your understanding of the concepts. Explain how they applied in practical scenarios.	0 - 3
		Examples and Evidence	Provide concrete examples from the class or lab work to support your reflections. Demonstrate how these examples helped solidify your learning.	0 - 2



		Connection to Course Work	Explain how the concepts from the course helped with your Course Works and how they are relevant to your learning process.	0 - 2
Q6	10	Identification of Limitations & Challenges	The student identifies limitations or challenges faced in terms of flexibility, scalability, and handling complex data structures while working with relational databases. The response is open as reflection.	0 - 2
		Alternative Approaches for non structured Data	The student discusses when a different approach to database design would be needed, particularly when dealing with large volumes or diverse formats of data ?	0 - 3
		Suggestions for Improved Database Design	The student discusses / suggest when a different database design or model or approach might be needed	0 - 2
		Logical Flow and Coherence.	The response is logically organized, with a clear progression of ideas.	0 - 3
Late Submission?				<input type="checkbox"/> Yes <input type="checkbox"/> No
Final Marks				