

**Department of Computing and Mathematics**

**ASSIGNMENT COVER SHEET**

<b>Unit title:</b>	6G5Z1111: Database Systems
<b>Assignment set by:</b>	Dr Anthony Kleerekoper & Dr Andrew Schofield
<b>Assignment ID:</b>	2CWK30
<b>Assignment title:</b>	Data Migration Project
<b>Assignment weighting:</b>	30%
<b>Type: (Group/Individual)</b>	Group Assignment
<b>Hand-in deadline:</b>	See hand-in date on Moodle
<b>Hand-in format and mechanism:</b>	Report to be submitted online via the unit's area on Moodle
<b>Support:</b>	Face to face support is available from tutors during the unit's lab sessions. Lab worksheets and associated videos support is available on the unit's Moodle page. For additional one-to-one support see tutor office hours on the unit's Moodle page.

**Learning outcomes being assessed:**

- Learning Outcome 1: Design relational databases using advanced modelling techniques (e.g., UML class diagrams), and create, maintain and write advanced queries on relational databases using a declarative language (e.g., SQL).
- Learning Outcome 4: Perform core database operations on a DBMS from within a programming language or environment.

**Note:** it is your responsibility to make sure that your work is complete and available for marking by the deadline. Make sure that you have followed the submission instructions carefully, and your work is submitted in the correct format, using the correct hand-in mechanism (e.g. Moodle upload). If submitting via Moodle, you are advised to check your work after upload, to make sure it has uploaded properly. Do not alter your work after the deadline. You should make at least one full backup copy of your work.

**Penalties for late hand-in:** see Regulations for Undergraduate Programmes of Study (<http://www.mmu.ac.uk/academic/casqe/regulations/assessment.php>). The timeliness of submissions is strictly monitored and enforced.

All coursework has a late submission window of **5 working days**, but any work submitted within the late window will be capped at 40%, unless you have an agreed extension. Work submitted after the 5-day window will be capped at zero, unless you have an agreed extension.

Please note that individual tutors are unable to grant extensions to coursework. Extensions can only be granted on the basis of a PLP, or approved Exceptional Factors (see below).

**Exceptional Factors affecting your performance:** see Regulations for Undergraduate Programmes of Study (<https://www.mmu.ac.uk/academic/casqe/regulations/assessment/docs/ug-regs.pdf>). For advice relating to exceptional factors, please see the following website: <https://www2.mmu.ac.uk/student-case-management/guidance-for-students/exceptional-factors/> or visit a Student Hub for more information.

**Plagiarism:** Plagiarism is the unacknowledged representation of another person's work, or use of their ideas, as one's own. Manchester Metropolitan University takes care to detect plagiarism, employs plagiarism detection software, and imposes severe penalties, as outlined in the Student Handbook ([http://www.mmu.ac.uk/academic/casqe/regulations/docs/policies\\_regulations.pdf](http://www.mmu.ac.uk/academic/casqe/regulations/docs/policies_regulations.pdf) and Regulations for Undergraduate Programmes (<http://www.mmu.ac.uk/academic/casqe/regulations/assessment.php>). Bad referencing or submitting the wrong assignment may still be treated as plagiarism. If in doubt, seek advice from your tutor.

**As part of a plagiarism check, you may be asked to attend a meeting with the Unit Leader, or another member of the unit delivery team, where you will be asked to explain your work (e.g. explain the code in a programming assignment). If you are called to one of these meetings, it is very important that you attend.**

<b>Assessment Criteria:</b>	Indicated in the attached assignment specification.
<b>Formative Feedback:</b>	Formative feedback and guidance for the project will be given verbally in the lab sessions. Formative feedback on lab exercises will also be given in this manner.
<b>Summative Feedback Format:</b>	Summative feedback sheets for the project will be provided after submission. See marking scheme for details.

## **ASSIGNMENT BRIEF**

### **AIMS:**

The primary aims of this project are to give you the opportunity to:

- Demonstrate your understanding of UML/ERD modelling by applying it to the development of a database system.
- Develop your understanding of the issues that can arise when moving from design to implementation in database systems development.
- Demonstrate your understanding of database development as used to implement a system.
- Apply skills of critical analysis to real world situations within a defined range of contexts.
- Demonstrate a high degree of professionalism e.g. initiative, creativity, motivation, professional practice and self-management.
- Express ideas effectively and communicate information appropriately and accurately using a range of media including ICT

### **LEARNING OUTCOMES:**

- Learning Outcome 1: Design relational databases using advanced modelling techniques (e.g., UML class diagrams), and create, maintain and write advanced queries on relational databases using a declarative language (e.g., SQL).
- Learning Outcome 4: Perform core database operations on a DBMS from within a programming language or environment.

## **2CWK30: Group Data Migration Project**

### **SCENARIO:**

A media rental company has been compiling a database of films and TV episodes over many years. Initially, the database was small with limited information about each title. Over time, it has grown to encompass thousands of titles and include information about each title's directors, writers, cast and crew.

Unfortunately, this growth was not planned very well and the database is now a bit of a mess. There are two tables: titles and names. The titles table stores information about each title whilst the names table stores information about named individuals who are related to those titles in some way. The titles table contains 44,733 rows and the names table contains 39,070 rows.

The database is used to provide output to an online front-end where customers and the public can view information about titles. The database is also used to provide title recommendations to users through a combination of SQL queries and analytical algorithms.

## THE PROJECT:

At present, the company manages with its two very large and poorly designed tables. The company now wants to migrate to a better design which will reduce redundancy. The existing data will have to be transferred to the new design. An important consideration is that it must be possible to produce the same output as the existing design.

Your team, of 2 or 3 people, has been contracted to create the new database, transfer the data and design new queries.

## PROJECT REQUIREMENTS

### Section A – Critiquing the Current Database Design (5 marks)

1. Read the scenario and project brief above and examine the provided database dump. You should take some time to understand the columns in the existing tables and notice that they are un-normalised. You must create an ERD describing the original tables as given to you in the dumps.
2. The design is obviously flawed which brings problems with it. You should critique the design, writing a short report outlining the deficiencies with the design and what problems they lead to.

### Section B – New Database Design (10 marks)

3. You must design an Entity Relationship Diagram with an appropriate level of normalisation (i.e. up to and including 5NF if appropriate) that can store the same information as the existing tables. You are free to use either traditional “crow’s foot” or UML Class Diagram notation but should include appropriate relationships/associations (including cardinality/multiplicity) between entities/classes and appropriate attributes, including primary and foreign key definitions. Your final design should resolve any many to many associations.

*Note: Refer to the lecture and lab notes for guidance on drawing these diagrams. You are free to use any software to draw the diagrams but Visual Paradigm is recommended, as this will be covered in the lab material.*

### Section C – Database Implementation and Migration (5 marks)

4. **Implementation:** Implement your database design using Oracle DBMS. You may use the University server (etna), Oracle’s Application Express (APEX) system or a localhost setup (such as Oracle Express) on your own PC or laptop. You should include appropriate tables, attributes and properly set up the table relationships using primary/foreign keys.

*Note: You must only use Oracle DBMS not MySQL.*

5. **Data Migration:** Populate your database with the data from the existing tables by using appropriate and correct SQL statements. This will require creating tables for the dumped data to be inserted into before the data can be inserted from them into the new tables.

## Section D – SQL Query Creation (10 marks)

6. The company has a few complex queries that it has working with the old design. Based on your new design, you should create new queries that produce the same output. To help you, we have included a screenshot of some of the output.
- a. List all actors and actresses and the names of the titles for which they are known. When the title's name is not stored in the database, show the title\_id instead. Order the results by the name\_id of the actor/actress.

SQL | Fetched 50 rows in 0.454 seconds

NAME	Known for Title 1	Known for Title 2	Known for Title 3	Known for Title 4
1 Fred Astaire	tt0043044	tt0072308	tt0045537	tt0050419
2 Lauren Bacall	tt0071877	tt0117057	tt0037382	tt0038355
3 Brigitte Bardot	tt0049189	tt0059956	tt0057345	tt0054452
4 Humphrey Bogart	tt0033870	tt0043265	tt0037382	tt0034583
5 Marlon Brando	tt0078788	tt0070849	tt0068646	tt0047296
6 Richard Burton	tt0057877	tt0059749	tt0061184	tt0087803
7 James Cagney	tt0035575	One, Two, Three	tt0031867	tt0029870
8 Gary Cooper	tt0027996	tt0035211	High Noon	tt0034167
9 Bette Davis	tt0056687	tt0035140	tt0031210	tt0042192
10 Olivia de Havilland	tt0029843	tt0031381	tt0040806	tt0041452
11 Marlene Dietrich	tt0040367	tt0055031	tt0052311	tt0051201
12 Kirk Douglas	tt0049456	tt0052365	tt0056195	tt0054331
13 Henry Fonda	tt0082846	tt0032551	tt0064116	tt0050083
14 Clark Gable	tt0026752	tt0025316	tt0052278	tt0031381
15 John Gielgud	tt0045943	tt0071877	tt0076574	tt0082031
16 Alec Guinness	tt0050212	tt0041546	tt0051739	tt0076759
17 Margaux Hemingway	tt0110138	tt0074802	tt0077800	Inner Sanctum
18 Audrey Hepburn	tt0054698	tt0056923	tt0046250	tt0058385
19 Katharine Hepburn	tt0063227	tt0032904	tt0043265	tt0061735
20 Gene Kelly	tt0041944	tt0041716	tt0043278	tt0045152
21 Grace Kelly	tt0047396	tt0048728	tt0046874	tt0046912

ssages - Log X

- b. List all writer and director pairs who have worked together more than once on a drama (each pair should only be listed once for the same title)

TITLE_ID	TITLE	WRITER	DIRECTOR
1 tt0000886	Hamlet, Prince of Denmark	William Shakespeare	Gerard Bourgeois
2 tt0001240	Hamlet	William Shakespeare	August Blom
3 tt0003159	Les Miserables, Part 2: Fantine	Paul Capellani	Albert Capellani
4 tt0003159	Les Miserables, Part 2: Fantine	Victor Hugo	Albert Capellani
5 tt0003442	Tess of the D'Urbervilles	Thomas Hardy	J. Searle Dawley
6 tt0005061	Carmen	Prosper Merimee	Raoul Walsh
7 tt0005061	Carmen	Raoul Walsh	Raoul Walsh
8 tt0006101	The Stolen Voice	Frank Hall Crane	Frank Hall Crane
9 tt0006101	The Stolen Voice	Paul McAllister	Frank Hall Crane
10 tt0006234	Vordertreppe - Hintertreppe	Hermann Sudermann	Urban Gad
11 tt0006234	Vordertreppe - Hintertreppe	Urban Gad	Urban Gad
12 tt0007061	The Microscope Mystery	George Randolph Chester	Paul Powell
13 tt0007061	The Microscope Mystery	William E. Wing	Paul Powell
14 tt0007769	Camille	Adrian Johnson	J. Gordon Edwards
15 tt0007769	Camille	Alexandre Dumas fils	J. Gordon Edwards
16 tt0008652	A Tale of Two Cities	Charles Dickens	Frank Lloyd
17 tt0008652	A Tale of Two Cities	Frank Lloyd	Frank Lloyd
18 tt0009014	The Tenth Symphony	Abel Gance	Abel Gance
19 tt0009309	Little Women	Anne Maxwell	Harley Knoles
20 tt0009309	Little Women	Louisa May Alcott	Harley Knoles
21 tt0009852	Az elet kiralya	Jozsef Pakots	Alfred Deesy
22 tt0009852	Az elet kiralya	Oscar Wilde	Alfred Deesy
23 tt0010061	Deliverance	Francis Trevelyan Miller	George Foster Platt

- c. List every group of “actor/actress, writer, director” who worked on the same title, ordered by title\_ID. That is, each row should contain an actor/actress, a writer and a director who all worked on the same title. Make sure that you exclude groups where the same person appears in more than one column, e.g. where one person was both writer and director.

⚡ TITLE_ID	⚡ TITLE	⚡ ACTOR	⚡ WRITER	⚡ DIRECTOR
1	tt0000886 Hamlet, Prince of Denmark	Jean Mounet-Sully	William Shakespeare	Gerard Bourgeois
2	tt0001240 Hamlet	Einar Zangenberg	William Shakespeare	August Blom
3	tt0001240 Hamlet	Oscar Langkilde	William Shakespeare	August Blom
4	tt0001240 Hamlet	Carl Rosenbaum	William Shakespeare	August Blom
5	tt0001240 Hamlet	Alwin Neuss	William Shakespeare	August Blom
6	tt0001240 Hamlet	Ella La Cour	William Shakespeare	August Blom
7	tt0001240 Hamlet	Axel Mattsson	William Shakespeare	August Blom
8	tt0001240 Hamlet	Emilie Sannom	William Shakespeare	August Blom
9	tt0001240 Hamlet	Aage Hertel	William Shakespeare	August Blom
10	tt0002555 The Last Bohemian	Antal Nyaray	Zsolt Harsanyi	Michael Curtiz
11	tt0002555 The Last Bohemian	Bela Bodonyi	Zsolt Harsanyi	Michael Curtiz
12	tt0002555 The Last Bohemian	Zoltan Sipos	Zsolt Harsanyi	Michael Curtiz
13	tt0002555 The Last Bohemian	Elemer Thury	Zsolt Harsanyi	Michael Curtiz
14	tt0003159 Les Miserables, Part 2: Fantine	Henry Krauss	Victor Hugo	Albert Capellani
15	tt0003159 Les Miserables, Part 2: Fantine	Mistinguett	Victor Hugo	Albert Capellani
16	tt0003159 Les Miserables, Part 2: Fantine	Maria Fromet	Victor Hugo	Albert Capellani
17	tt0003159 Les Miserables, Part 2: Fantine	Henri Etievant	Victor Hugo	Albert Capellani
18	tt0003159 Les Miserables, Part 2: Fantine	Maria Ventura	Victor Hugo	Albert Capellani
19	tt0003159 Les Miserables, Part 2: Fantine	Henry Krauss	Paul Capellani	Albert Capellani
20	tt0003159 Les Miserables, Part 2: Fantine	Mistinguett	Paul Capellani	Albert Capellani
21	tt0003159 Les Miserables, Part 2: Fantine	Maria Fromet	Paul Capellani	Albert Capellani
22	tt0003159 Les Miserables, Part 2: Fantine	Maria Ventura	Paul Capellani	Albert Capellani
23	tt0003159 Les Miserables, Part 2: Fantine	Henri Etievant	Paul Capellani	Albert Capellani
24	tt0003442 Tess of the D'Urbervilles	Mary Barker	Thomas Hardy	J. Searle Dawley
25	tt0003442 Tess of the D'Urbervilles	John Steppling	Thomas Hardy	J. Searle Dawley
26	tt0003442 Tess of the D'Urbervilles	David Torrence	Thomas Hardy	J. Searle Dawley

- d. Count the number of people involved (in any capacity) with each genre as well as the total number of people regardless of genre

⚡ GENRE	⚡ Number
1 Action	4451
2 Adult	10
3 Adventure	3165
4 Animation	788
5 Biography	1234
6 Comedy	11651
7 Crime	4068
8 Documentary	4241
9 Drama	18747
10 Family	1837
11 Fantasy	1326
12 Film-Noir	170
13 History	888
14 Horror	3103
15 Music	757
16 Musical	1074
17 Mystery	1803
18 News	79
19 Reality-TV	24
20 Romance	5216
21 Sci-Fi	1136
22 Sport	460
23 Thriller	3566
24 War	889
25 Western	813
26 (null)	37011

- e. List all the living cinematographers who are known for titles with average ratings of 4.5 or less and the name of the title with their lowest rating. Order the results by the lowest average rating from highest to lowest

NAME	MIN(AVERAGERATING)	TITLE
1 Alain Betrancourt	4.5	Top of the World
2 Brian Pratt	4.5	Pauly Shore Is Dead
3 Dharmadjie	4.5	Hantu Jeruk Purut
4 Evan Marlowe	4.5	Horror House
5 Hitoshi Kato	4.5	Tomie: Revenge
6 Jean-Max Bernard	4.5	Va mourire
7 Reza Serkanian	4.5	Hip Moves
8 Richard King	4.5	The First Turn-On!!
9 Yuva	4.5	Jackson Durai
10 Adam Sherer	4.4	Vampire
11 Bryan Koss	4.4	In Stereo
12 Cuneys Denizer	4.4	Super incir
13 Goof de Koning	4.4	The Human Centipede (First Sequence)
14 Ivan Zuccoon	4.4	Wrath of the Crows
15 Lachlan Milne	4.4	Uninhabited
16 Bala Bharani	4.3	Goripalayam
17 Don Stern	4.3	Creature
18 Harry Mathias	4.3	Creature
19 Hector Ortega	4.3	The Night Buffalo
20 Mark Faulkner	4.3	Complete Guide to Guys
21 Peter Roos	4.3	Suzanne og Leonard
22 Tom Adair	4.3	Complete Guide to Guys
23 Aseem Mishra	4.2	Tubelight
24 Bing Rao	4.2	Diamond Dogs
25 Bobby Bukowski	4.2	Boogeyman
26 Erdal Kahraman	4.2	Love Under Siege

## DELIVERABLES

You should submit a single group report (only one report per group is required) containing:

- Your group's ERD for the original design (section A)
- A critique of the original design (word limit about 500 words)
- Your group's ERD for the new design (section B).
- A discussion of any assumptions that you have made about the existing database and any design decisions you have made that have affected your implementation (word limit about 400 words).
- The SQL code used to implement your new design (section C.4).
- The SQL code used to migrate the data including the code used to create the tables to store the dumped data (if appropriate) (section C.5)
- The SQL code for the queries (section D)

Please ensure you provide a list of group members and their Student ID numbers in the report.

Your report will be submitted via the unit's Moodle page. Instructions on submission are provided on the page.

Finally, each team member may individually submit a completed peer review form (template available on Moodle) indicating the contributions made by each team member. There will be a second submission area for this on the Moodle page.

## STEPPED MARKING

Please note that stepped marking will be applied to sections A, B and C of this assessment. For section D, the marking scheme will simply be 2 marks per correct query.

Stepped marking means that the marking criteria will be applied to determine an overall mark band. Then, within this band, a decision will be made as to whether the assignment was in the top, middle or bottom of the band. Based on that decision, the assignment will be allocated a mark ending in 2, 5 or 8 as appropriate.

For example, an assignment might be judged, based on the marking criteria, to be in the 2:1 band (60% - 70%). It will therefore be marked as either 62%, 65% or 68% depending on whether it is in the top, middle or bottom of that band. No other marks are possible.

For further details, please visit:

[https://www.celt.mmu.ac.uk/assessment/lifecycle/5\\_step\\_marking.php](https://www.celt.mmu.ac.uk/assessment/lifecycle/5_step_marking.php)

### **ASSESSMENT HELP**

Help will be available during the lab sessions and some sessions may be devoted entirely to the assessment.

For any questions or help outside of the lab sessions, please contact:

Dr Andrew Schofield, E141, JD. Email: [a.schofield@mmu.ac.uk](mailto:a.schofield@mmu.ac.uk)

Office Hours: See Moodle

Dr Anthony Kleerekoper, E152, JD. Email: [a.kleerekoper@mmu.ac.uk](mailto:a.kleerekoper@mmu.ac.uk)

Office Hours: See Moodle



## MARKING SCHEME

The deliverables for the project are a group report and relevant SQL scripts.

The criteria that will be applied to each part of the submission are given in the tables below. Percentages are provided for each section to enable you to work out an overall percentage grade for the work:

<b>Original Database Design (5%)</b>	
Pass (3 <sup>rd</sup> )	ERD provided with little or no critique
Lower 2 <sup>nd</sup>	Mostly correct ERD provided with simple critique covering some of the issues with the design.
Upper 2 <sup>nd</sup>	Correct ERD provided with good critique covering most of the issues with the design.
1 <sup>st</sup>	Correct ERD provided with very good critique covering all of the issues with the design.

<b>New Database Design (10%)</b>	
Pass (3 <sup>rd</sup> )	Basic ERD produced with some normalisation but little or no supporting commentary.
Lower 2 <sup>nd</sup>	ERD provided with acceptable amount of normalisation and some commentary. Appropriate use of diagram features and syntax. Attributes for diagram are mostly correct.
Upper 2 <sup>nd</sup>	ERD and appropriate normalisation provided along with supporting information. Correct and accurate diagram with matching attributes datatype/length definitions and keys. Some commentary and evaluation of issues surrounding development.
1 <sup>st</sup>	High quality ERD specification produced with clear, comprehensive justification and evaluation of design and alternatives, and consideration of business scenario needs. Accurate and detailed specification of attributes datatype/length definitions, keys and constraints. Design is well thought out and suitable for implementation in RDBMS.

<b>Database Implementation and Migration (5%)</b>	
Pass (3 <sup>rd</sup> )	Database tables created successfully. Implementation shows some relation to the ERD and attribute definition in terms of relationships and primary/foreign keys. Data has been transferred with some errors or workarounds.
Lower 2 <sup>nd</sup>	Database implementation has been performed correctly and shows a clear relationship to the ERD in terms of relationships and primary/foreign keys. Data has been transferred with few errors or workarounds.
Upper 2 <sup>nd</sup>	Properly implemented database which clearly reflects the ERD and attribute list. Primary/Foreign keys created properly and matching ERD. Data has been transferred without errors. Good, clear discussion/commentary of implementation.
1 <sup>st</sup>	Well implemented database which clearly reflects the ERD and attribute list. Primary/Foreign keys created properly and matching ERD. Additional and appropriate constraints implemented. Data has been transferred

	using efficient SQL commands. Good, clear discussion/commentary of implementation showing consideration of business scenario requirements.
--	--

<b>SQL Query Creation (10%)</b>
2% for each correct query