

Student Assignment Brief

This document is intended for Coventry University Group students for their own use in completing their assessed work for this module. It must not be passed to third parties or posted on any website. If you require this document in an alternative format, please contact your Module Leader.

Contents:

- [Assignment Information](#)
- [Assignment Task](#)
- [Marking and Feedback](#)
- [Assessed Module Learning Outcomes](#)
- [Assignment Support and Academic Integrity](#)
- [Assessment Marking Criteria](#)

The work you submit for this assignment must be your own independent work, or in the case of a group assignment your own groups' work. More information is available in the '[Assignment Task](#)' section of this assignment brief.

Assignment Information

Module Name: Data Management Systems

Module Code: 7086CEM

Assignment Title: Database management

Assignment Due: 27 March 2024, 18:00 UK Time

Assignment Credit: 15

Word Count (or equivalent): 600 words for Part D only.

Assignment Type: **Percentage Grade** (Applied Core Assessment). You will be provided with an overall grade between 0% and 100%. You have one opportunity to pass the assignment at or above 40%.

Assignment Task

7086CEM – Data Management Systems

This assignment is made up of four parts to be tackled by a group of two students:

- Part A deals with database design, using E-R modelling.
- Part B concerns database creation and querying, using SQL.
- Part C covers distributed processing frameworks.
- Part D involves a research report.

Part A. Conceptual modelling

An equipment company wishes to create a database to support the hiring of tools and machinery to clients. The company has four types of equipment: power tools, such as drills and vacuum cleaners; plants such as excavators and floor sanders; scaffolding such as mobile scaffolding and cantilever scaffolding; and work wear such as protective clothing and high visibility clothing. Each piece of equipment is identified by a number. Power tools are described by their brand, their model and the voltage they use, whereas plants are classified by their model and their size in tonnes. Scaffolding type can be traditional, aluminium or fibreglass; in addition, its width can be single or double, and its strength is specified. Work wear is described by its tolerance and its size.

The company has various outlets and each has staff including a manager and several senior technicians who are responsible for supervising the work of allocated groups of technicians. A supervision record is also kept for a specific date.

All employees are identified by their number, name, date of birth (DOB) and address. Two types of worker have been identified: shop floor workers and office workers. Each shop floor worker has a specific skill and performs a specific task. Office workers on the other hand, are identified by their role and the department to which they belong. Furthermore, a record is kept on the qualifications and the employment history of each employee. In order to ensure orderly transfer of skills between employees, a mentoring scheme is supported by the company.

Each outlet has a stock of equipment that may be hired by clients for varying periods of time, from a minimum of four hours to a maximum of six months. Each hire agreement between a client and the company is uniquely identified by using a hire number. Each client is identified by a number and a name. The company insists that a client must take out an insurance cover for each equipment hire period, offered by the company and which is administered by an employee. Each piece of equipment is checked for faults when it is returned by the client, and the faults/defects/damage recorded. The company keeps a record of the hire history of each client.

1. Create an ER diagram for the above scenario and indicate the cardinality of relationships and the nature of the associations (mandatory or optional). You should allocate adequate attributes to the entities/relationships of interest, especially the identifiers.
(20%)
2. Generate, with justification, relational tables from the ER diagram. Indicate clearly the names of the tables, the attributes, the primary keys and the foreign keys.
(5%)

Guidance: i) Please use the notation given in the lecture notes. Create the ER diagram and clearly identify any identifiers, indicate the cardinality of relationships and the nature of the associations (mandatory or optional). ii) Generate tables and include primary and foreign keys. Use the schema notation for the tables; you do not have to produce SQL statements.

Example of table generation in schema form:

Course(courseId, courseName)
Student (studentId, name, courseId*)

Part B: SQL programming

Consider the following Employee database and sample data. You may wish to add more data records.

Employee (empId, name, address, DOB, job, salaryCode, deptId, manager, schemeld)

Department (deptId, name)

SalaryGrade (salaryCode, startSalary, finishSalary)

PensionScheme (schemeld, name, rate)

TABLE: Employee

empId	name	address	DOB	job	salaryCode	deptId	manager	schemeld
E101	Keita, J.	1 high street	06/03/76	Clerk	S1	D10	E110	S116
E301	Wang,F.	22 railway road	11/04/80	Sales person	S2	D30	E310	S124
E310	Flavel,K.	14 crescent road	25/11/69	Manager	S5	D30		S121
E501	Payne, J.	7 heap street	09/02/72	Analyst	S5	D50	E310	S121
E102	Patel,R.	16 glade close	13/07/74	Clerk	S1	D10	E110	S116
E110	Smith,B.	199 London road	22/05/70	Manager	S5	D10		S121

TABLE: Department

deptId	name
D10	Administration
D20	Finance
D30	Sales
D40	Maintenance
D50	IT Support

TABLE: SalaryGrade

salaryCode	startSalary	finishSalary
S1	17000	19000
S2	19001	24000
S3	24001	26000
S4	26001	30000
S5	30001	39000

TABLE: PensionScheme

schemeld	name	rate
S110	AXA	0.5
S121	Premier	0.6
S124	Stakeholder	0.4
S116	Standard	0.4

1. Use appropriate Oracle data types and write the SQL statements to create the tables defined in the schema above.

(10%)

2. Write SQL statements to return the following data from the Employee database:

a) The name (in ascending order), the starting salary and department id of each employee within a descending order of department ids.

(5%)

b) Give the number of employees for each of the pension schemes offered by the company. Result listing should include the name of each scheme and its corresponding number of employees who join the scheme.

(5%)

c) Give the total number of employees who are not managers but currently receive an annual salary of over £35,000.

(5%)

d) List the id and name of each employee along with his/her manager's name.

(5%)

Guidance: Please use Oracle standard SQL. Indicate clearly the primary keys and the foreign keys. State the SQL statements and give the results. Select and use appropriate data. The presentation of each query should have a text summary which includes i) the query itself, ii) the corresponding SQL statement solution, iii) the result of the execution of the statement and iv) evidence that you have used standard SQL and implemented each statement on a database (use screenshots or spool facility). Links to external datasets or code are not acceptable. The report must be self-contained.

C. Sequential and parallel processing

Consider a flight data store with the following data structure, where all times are in GMT. Each record consists of the 12 attributes; the set of allowable values of the attributes and format are specified in the description (metadata).

	Data Value	Description
1	Year	2000-2020
2	Month	1-12
3	Day of Month	1-31
4	Day of the Week	1 (Monday) – 7(Sunday)
5	Departure Time	Recorded Departure time (hhmm)
6	Actual Departure time	Scheduled Departure time (hhmm)
7	Arrival Time	Recorded Arrival time (hhmm)
8	Airlines carrier	Carrier code (unique)
9	Flight Number	Flight Number
10	Departure Delay	minutes
11	Arrival Delay	minutes
12	Weather Delay	minutes

An example record would have the following values:

(2015, 4, 20, 5, 1430, 1400, 1820, 131, JL729, 30, 15, 0)

A record will contain a list of key-value pairs, as presented in the lectures. A block will consist of a sequence of records. You are expected to create similar additional records for this part, so that you can give a solution to question 2, below.

Flight monitors would like to determine the number of flights which were delayed for each carrier.

1. Assuming that the data is stored in a relational database produce, with justification a) the SQL statement to create the corresponding table b) the SQL statement to determine the number of flights which were delayed for each carrier. (You do not need to implement the SQL statements on Oracle).

(5%)

2. Assuming that the data is too large to be processed in a centralised manner, and that it is stored in an ordinary file, produce a decentralised solution which applies MapReduce to the data processing. Justify your decisions and all the steps of your solution. Use diagrams if required.

(20%)

Guidance: You should study carefully the examples of mapReduce covered in the lecture notes. You should consider the structure of the key in the (key, value) pair in the original record and in the mapping stage. This is not a programming exercise. The solution should follow the structure given in the lecture notes and in the examples/exercises on mapReduce.

Part D. Research report

Consider the following quote from an online article:

“With more than 1.5 million new ads posting every day, Craigslist (a site for classified listings) users have generated over a billion records – some might even consider that ‘big data.’ What’s more, legislation demands that after a 60 day retention period in the live portion of the site, records must be migrated over to an archival space for legislative compliance.

Craigslist faced several challenges due to the nature and volume of data being stored in their relational, MySQL servers. For example, the structure of their data had changed several times over the years. This alone made any change to the database schema a costly, prolonged nightmare, as changes often meant downtime. And if database alterations were a challenge, just imagine how difficult introducing entirely new features became? What’s more, each change to the live database schema required a corresponding change to the entire archive – a process that took months every time“.

In 2011 Craigslist decided to migrate the data and the processing from the relational MySQL servers to NoSQL MongoDB servers.

Refer to the characteristics of relational databases and NoSQL databases and the context, and explain, **in no more than 600 words** why Craigslist moved their operations from MySQL to MongoDB.

(20%)

Guidance: Use your own words for the report. **Copying and pasting is plagiarism.** You should include relevant references. The maximum length of the report is 600 words. Longer reports will be penalised.

Submission Instructions:

Submission process for the group

In order to avoid 100% plagiarism similarity within a group, the CW assignment must be submitted as the work of a group of two students as follows:

- a) you are expected to work as a group on ALL the parts of the CW.
- b) one student will submit individually through Aula, the whole report, with both names and ids on the front page, as a single pdf file, and
- c) the other student must submit individually in his/her name on Aula, half a page to one page only, with both names and ids, as a pdf file, where he/she presents reflections on the collaborative work.

Report structure

Please follow the instructions below. You are expected:

- a) to include the name and the code of the module, the title of the CW, **your names and your student ids**.
- b) to follow the structure of the CW brief, and present your answers accordingly, i.e.:

Part A

- 1. *Answer ...*
- 2. *Answer ...*

.....

Part B

- 1. *Answer ...*
- 2.
 - a) *Answer ...*
 - b) *Answer ...*
 - c) *Answer ...*
 - d) *Answer ...*

..

.....

Marking and Feedback

How will my assignment be marked?

Your assignment will be marked by the module team.

How will I receive my grades and feedback?

Provisional marks will be released once internally moderated.

Feedback will be provided by the module team alongside grades release.

Feedback will be available on Aula.

Your provisional marks and feedback should be available within 2 weeks (10 working days).

What will I be marked against?

Details of the marking criteria for this task can be found at the [bottom of this assignment brief](#).

Assessed Module Learning Outcomes

The Learning Outcomes for this module align to the [marking criteria](#) which can be found at the end of this brief. Ensure you understand the marking criteria to ensure successful achievement of the assessment task. The following module learning outcomes are assessed in this task:

2. Demonstrate a sound understanding of the theoretical and practical issues relevant to data management systems
3. Critically evaluate a range of conceptual and technical tools and to apply them selectively in the design and implementation of an effective data management system
4. Assess and evaluate the theoretical and technological underpinnings of distributed frameworks
5. Review and comment critically on the current trends in distributed processing

Assignment Support and Academic Integrity

If you have any questions about this assignment please see the [Student Guidance on Coursework](#) for more information.

Spelling, Punctuation, and Grammar:

You are expected to use effective, accurate, and appropriate language within this assessment task.

Academic Integrity:

The work you submit must be your own, or in the case of groupwork, that of your group. All sources of information need to be acknowledged and attributed; therefore, you must provide references for all sources of information and acknowledge any tools used in the production of your work, including Artificial Intelligence (AI). We use detection software and make routine checks for evidence of academic misconduct.

Definitions of academic misconduct, including plagiarism, self-plagiarism, and collusion can be found [on the Student Portal](#). All cases of suspected academic misconduct are referred for investigation, the outcomes of which can have profound consequences to your studies. For more information on academic integrity please visit the [Academic and Research Integrity](#) section of the Student Portal.

Support for Students with Disabilities or Additional Needs:

If you have a disability, long-term health condition, specific learning difference, mental health diagnosis or symptoms and have discussed your support needs with health and wellbeing you may be able to access support that will help with your studies.

If you feel you may benefit from additional support, but have not disclosed a disability to the University, or have disclosed but are yet to discuss your support needs it is important to let us know so we can provide the right support for your circumstances. Visit [the Student Portal](#) to find out more.

Unable to Submit on Time?

The University wants you to do your best. However, we know that sometimes events happen which mean that you cannot submit your assessment by the deadline or sit a scheduled exam. If you think this might be the case, guidance on understanding what counts as an extenuating circumstance, and how to apply is [available on the Student Portal](#).

Administration of Assessment

Module Leader Name: Rachid Anane

Module Leader Email: csx220@coventry.ac.uk

Assignment Category: Written

Attempt Type: Resit

Component Code: Cw

Assessment Marking Criteria

	Part A	Part B	Part C	Part D
80 to 100%	<ul style="list-style-type: none"> Excellent and very clear interpretation of scenario and excellent solution formulation of the two parts Correct and coherent identification of well annotated entities and types of relationships Logical and consistent generation of relational tables Excellent justification of design decisions 	<ul style="list-style-type: none"> Excellent interpretation of requirements and very good formulation of solution Excellent use of DDL and DML statements in solution formulation Very thorough treatment of queries and SQL formulation Excellent rationale 	<ul style="list-style-type: none"> Excellent solution to the initial query in terms of SQL statements Excellent presentation on the need for an overall parallel solution Very logical deployment and annotation of the sequential steps of MapReduce Excellent rationale 	<ul style="list-style-type: none"> Excellent understanding and interpretation of requirements Excellent identification and formulation of key issues Very relevant and specific presentation of structural and processing components Deeply reflective writing supported by an excellent structure
70 to 79%	<ul style="list-style-type: none"> Very good and clear interpretation of scenario and excellent solution formulation of the two parts Correct and coherent identification of well annotated entities and types of relationships Logical and consistent generation of relational tables Very good justification of design decisions 	<ul style="list-style-type: none"> Very good interpretation of requirements and very good formulation of solution Very good use of DDL and DML statements in solution formulation Complete treatment of queries and SQL formulation Very good rationale 	<ul style="list-style-type: none"> Very good solution to the initial query in terms of SQL statements Very good presentation on the need for an overall parallel solution Clear deployment and annotation of the sequential steps of MapReduce Excellent rationale 	<ul style="list-style-type: none"> Very good understanding and interpretation of requirements Very good identification and formulation of key issues Relevant and specific presentation of structural and processing components Reflective writing supported by a very good structure
60 to 69%	<ul style="list-style-type: none"> Good and clear interpretation of scenario and good solution formulation of the two parts Correct and complete identification of entities and relationships Consistent generation of relational tables Well expressed rationale 	<ul style="list-style-type: none"> Good interpretation of requirements and good formulation of solution Correct use of DDL and DML statements in solution formulation Complete and relevant treatment of queries Good justification of decisions 	<ul style="list-style-type: none"> Good solution to the initial query in terms of SQL statements Focused presentation of context of application of parallel processing Clearly stated and correct sequential steps of MapReduce Well expressed rationale 	<ul style="list-style-type: none"> Good understanding and statement of requirements Well focused presentation of main issues Good description of structural and processing components Well presented essay with some reflection
50 to 59%	<ul style="list-style-type: none"> Adequate and consistent 	<ul style="list-style-type: none"> Adequate understanding of 	<ul style="list-style-type: none"> Adequate understanding of 	<ul style="list-style-type: none"> Adequate interpretation of

	<ul style="list-style-type: none"> interpretation of scenario and satisfactory conceptual modelling Mostly correct generation of ER diagram with relevant entities and relationships Relatively competent generation of the relational tables Adequately justified design decisions 	<ul style="list-style-type: none"> requirements and mostly correct interpretation Adequate use of DDL and DML SQL statements in solution formulation Mostly complete SQL statements Adequate justification of solution 	<ul style="list-style-type: none"> requirements and correct SQL formulation Adequate presentation of context of application of parallel processing Mostly correct application of the different steps of MapReduce Adequately justified solution 	<ul style="list-style-type: none"> requirements Key issues well identified and partially addressed Adequate presentation of key structural and processing components Adequately written essay
40 to 49%	<ul style="list-style-type: none"> Partial interpretation of scenario and formulation of solution Partially correct ER diagram with relevant entities and relationships Partial consistency in the generation of the relational tables Partial justification of design decisions 	<ul style="list-style-type: none"> Basic understanding of requirements and partially correct interpretation Relevant use of DDL and DML statements in solution formulation Partially complete SQL statements Limited justification of solution 	<ul style="list-style-type: none"> Partial understanding of requirements and partially correct SQL formulation Partial understanding of context and relevance of parallel processing Incomplete steps in the application of MapReduce Partial justification of design decisions 	<ul style="list-style-type: none"> Partial interpretation of requirements Limited presentation of key issues Relevant description of structural and processing components Mostly descriptive essay
Fail 30, 35%	<ul style="list-style-type: none"> Incorrect interpretation of scenario and Incomplete formulation of solution Limited identification of entities and poor annotation of relationships Incorrect generation of relational tables Limited or absent rationale 	<ul style="list-style-type: none"> Poor interpretation of requirements and of queries DDL and DML SQL statements limited in scope Incomplete and incorrect SQL statements Absence of rationale 	<ul style="list-style-type: none"> Partial understanding of requirements and partially correct SQL formulation Partial understanding of context and relevance of parallel processing Incomplete steps in the application of MapReduce Partial justification of design decisions 	<ul style="list-style-type: none"> Lack of understanding of requirements Inadequate identification of issues Incompetent understanding of structural and processing components Poorly written essay
Fail 0 to 29%	<ul style="list-style-type: none"> Limited interpretation of scenario and Incomplete formulation of solution Limited identification of entities and poor annotation of relationships Limited generation of relational tables Absence of rationale 	<ul style="list-style-type: none"> Limited and incomplete interpretation of requirements and of queries Lack DDL and DML SQL statements or limited in scope Incomplete and limited SQL statements Absence of rationale 	<ul style="list-style-type: none"> Limited understanding of requirements and partially correct SQL formulation Limited understanding of context and relevance of parallel processing Incomplete and limited steps in the application of MapReduce No justification of design decisions 	<ul style="list-style-type: none"> Limited understanding of requirements Limited identification of issues Limited or no understanding of structural and processing components Badly written essay