## Schedule and Deliverables

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Item | Value | Format | Deadline | Late/EC Deadline |
| ITEM 1 (CW1) | 50% | Moodle - 2 Files (PDF + SQL) | **18/11/2022 | 23:00** | 02/12/2022 | 23:00 |
| ITEM 2 (CW2) | 30% | Moodle - 1 File (PDF) | **16/12/2022 | 23:00** | 30/12/2022 | 23:00 |
| ITEM 3 (Labs) | 20% | Moodle - 2 Files (PDF + SQL) | **18/12/2022 | 23:00** | N/A |

***Important: Do not leave your submission to the last minute as your connection or the server might be slow, technical errors etc. Aim to submit a minimum one day before the deadline or in the morning. No submission will be accepted after 23:00. No Excuses.***

## Plagiarism and Poor Scholarship

This piece of work requires you to provide a database solution for a business scenario. You are expected to understand how to reference using the [APA 7 system](https://library.port.ac.uk/ref/page2.html). Therefore, marks will be deducted for poor scholarship and NO Marks will be awarded for the entire assignment if ANY part of it is found to be directly copied from printed / published work or from another student.

Use respected websites for your information gathering. E.g. technical sites such as PostgreSQL, Oracle, IBM DB2, MySQL technical pages. To avoid plagiarism rewrite using your own words, but ensure where you have used the ideas of others you acknowledge the source within the text and provide the full source in the references at the end of the document (using Harvard APA 7 format). If you are not sure how to reference, use the information provided by the [Library](https://library.port.ac.uk/ref/page2.html).

## Presentation

Proof read you work for spelling, grammar and English prior to submission as clarity of expression is an important part of a research coursework. Please include page numbers and your student ID in the header/footer of the document (on every page).

## Anonymity of work

For a fair marking of your coursework, *please use only your UP number on all documents* instead of your name.

## Assessment marks

This assessment has multiple submissions and deadlines and is covering LO: 1-4 as follow:

* **ITEM 1 (CW1 18/11/2022 - 50%)** 
  + Evaluate & design a functional database that will address business needs; Recommend hardware and software solutions for design
* **ITEM 2 (CW2 16/12/2022 - 30%)**
  + Evaluate database architecture as distributed system with regard of scalability and performance
* **ITEM 3 (CW3 18/12/2022 - 20%)**
  + Lab exercises

## Assessment Notes

*As this module is based on Advanced Databases Concepts you can use any relational DBMS (PostgreSQL, MySQL, MariaDB) however, you should write all the codes(s) yourself without using 3rd party tools (such as pgAdmin, PHPMyAdmin etc). Using external applications to automatically generate CREATE statements might lead to 0 marks for coding*.

# Assessment Description

*The design and approach of this assessment MUST be specifically related to the case study and not generalisation.*

You are part of a team of junior database developers working for a Performance Systems PLC and have been in post approximately 9 months. During this time you have been working on different existing database systems that have been developed by the company’s Senior DBA (You can assume that Val would have this role).

One of the Company’s clients has asked for a system to be developed for them. The Senior DBA does not have the time to devote to this project and so has asked you to work out the design, development, implementation and justification for him. He will then check it over before it is delivered to the clients.

Therefore the work you produce is for a Senior DBA and as such you can assume they understand more about databases than you do. Your report should take this into account and you should not spend time describing or explaining topics that the senior DBA already knows (e.g. why you have an PK in a table) but you should analyse their importance and relevance in the given scenario.This coursework follows the lifecycle of database implementation. The first stage is to try and identify the data and the organisational / linkage of the data for the system. They are working with another software development company that will do the front/back end hence ***there is no need to cover any of the web interfaces***.

At this stage the business did not provide too much information about what they want and there is room for many interpretations and assumptions in the proposed solution.

# Case Study

Industrial Works LTD is a UK company that has as its core activity environmental engineering, and is specialised in wastewater treatment solutions. They are involved in the entire process from designing, manufacturing and implementation of the projects for national and international clients. For each project there are various teams allocated and coordinated, projects consisting multiple stages - planning, design, approval, manufacturing implementation, revision, maintenance.

The projects are coordinated from different sources (software) and there is no way to automatically migrate the information from one software to another. The company clients are unable to view/participate in any of the project stages and the only communication is done through phone/email.

They are looking for an all-in-one software solution where they would be able to overview a project through all stages. The proposed solution should be able to track:

* All the stages of a project (e.g. drawing, execution, completed, re-design, maintenance etc)
* The start/end data of a project
* The staff members (teams) allocated to a specific project - a staff member can work on multiple projects at the same time and be part of multiple teams and each project will have a project manager. A project manager (PM) can be manager only for one project at the time but they can have multiple projects during their employment.
* The meeting dates and notes of each team for each project
* The documents (e.g. plans, drawings, permissions, notes) for each project. All documents will be submitted as PDF files. *You do not need to build up any API to handle the PDF files but a way of tracking the documents IDs*

Each project is tracking three major aspects of the project:

* Components (Different components allocated to the project - e.g. metal pipe, 50cm; metal sheet 50x50; plastic tube diameter 50mm etc, the price per component, how many components per project, part number etc.)
* Defects - if any defect was found to a component before installation; Notes and status of the component with the defect (e.g. replaced, fixed )
* Incidents - If during the project built there were any incidents (e.g. pipe crack; gear broken etc). The incidents should record the date/time where happened, for which project and the person who identified.

# Assessment Requirements

**Requirement 1 | ITEM 1 - (CW1 18/11/2022) - 50% (100 Marks)**

1. Using the information given in the [case study](#co83dhwwlv0u), draw an ***Enhanced Entity Relationship Diagram*** (EERD) for the system. You MUST resolve any M:M relationships and include the Primary and any Foreign Keys (PK/FK) in each of the tables (entities) including all suitable attributes, data type and data size. The EERD should be electronically created (e.g. not hand drawn) and ***the design should be as realistic as possible*** (e.g. attribute names, data type/ size). Include any assumptions you have made in design. The assumptions should clarify ambiguous aspects of your design. Do not repeat the brief statements (e.g. “*Customers table will record customers*” or “*a customer can make more than one booking*” ) ***(15 Marks)***
2. A complete Data Dictionary that will show all constraints and additional description if necessary. The Data Dictionary should follow [this template](https://docs.google.com/document/u/0/d/1TuTHzQEtm79MsXopJv2y-5nK9Q4h2g2fkYSdVnp_D20/edit). For obvious attributes a description is not needed (e.g. user\_name; user\_email etc). ***(5 Marks)***
3. Based on your design (EERD), identify the most important transactions (critical transactions - CRUD/IRUD) of your database creating a Transaction Analysis (TA) matrix. Based on the results of the typical TA results, select one of the transactions identified and create a daily Transaction Volume analysis (TV). ***(10 Marks)***
4. Discuss the ***physical design decisions*** you will have to make, specific to your chosen client to ensure successful implementation of the database on ONE Site. (i.e A centralised database that will be accessed by multiple staff/users over an internet connection (web interface probably - you do not design any interfaces). These decisions include, but are not limited to, how to obtain the exact requirements (clarification of assumptions), selection of a RDBMS, chosen file systems for the tables, indexes, security considerations, underlying hardware decisions, CPU, RAM HDD/SSD capacity based on software requirements and estimated database size, cost implications and more. These decisions will be based on the completed Transaction Analysis and you only need discuss your decision for the KEY tables identified. You should try to include all issues that you would need to address to ensure that you are successful when implementing the database system for the client. ***You should enhance your discussion with specific examples relevant to your chosen project.*** ***(10 Marks)***
5. Critically evaluate how you would monitor and optimise the Database. Your answer MUST focus on the case study and NOT just general optimization issues. Add examples to showcase your understanding of optimization processes to include optimization at database structure level, tuning and query optimization - referencing the use of indexing, query design best practices.***(10 Marks)***
6. ***The physical implementation of the database***. ALL SQL code (CREATE) and sample data that you have used for testing the database (INSERT). The code must be organised in the correct CREATE/INSERT sequence as it will be tested on the VM. The code should work as it was provided just using COPY/PASTE into the VM. The physical database should follow the Data Dictionary (1.2).. *Check your code before submission. If the database will not work, there will be no way to assess 1.7 either.* ***(10 Marks)***
7. In order to evaluate the database's usability, based on your design you are required to ***create 6 queries*** that you are considering will be useful for the business. You can create VIEWs and base other queries on those VIEWs. The queries output must be inserted as a screenshot into the main report, along with a query description - what is that query useful for? A copy of all queries should be available at the end of the database creation SQL file. Some useful queries would be identified through your ***Transaction Analysis & Transaction Volume*** (see 1.3). Each query should have their Relational Algebra (RA) and Relational Algebra Tree (RAT) **(30 Marks - 6 x 5)**
8. Innovation and excellence - Awarded for extraordinary work that goes beyond minimum scenario description and requirements, including but not limited to organised layout, grammar, references and anything else that is outstanding **(10 Marks)**

**Requirement 2 | ITEM 2 (CW2 16/12/2022) - 30% (100 Marks)**

The initial design was for the single HQ location. However, the company is looking for new solutions as they are planning to expand in the future with 4 more branches in different locations. The current solution as a single relational database will no longer be suitable from a cost and efficiency point of view. You have appointed to provide a rapport than will cover the following aspects:

1. Critically analyse options to the architecture of database systems; they may be Distributed Databases, Web Databases, NoSQL Databases or Mobile Databases that might be suitable for the client. The discussion should not be descriptive (as what is a DD) but an analysis if it is suitable for the business needs (or not) with arguments. (**20 Marks**)
2. Based on the analysis conducted (2.1) recommend one of the suitable architectures. The recommendation should include the additional design requirements, including, but not limited to, data allocation, data fragmentation, data replication, optimization, distributed database security (not general security), distributed backup and recovery strategies, distributed query processing and underlying infrastructure changes from the centralised system discussed in CW1. Chose only one of the following:
   1. Distributed Database System
   2. NoSQL Database
   3. Mobile Database

NB: *This aspect is NOT about mobile databases that reside on phones (mini databases – SQL-Lite), it is about mobile access to a centralised database. Is about what data will be stored/replicated on the mobile device in order to access/sync with the main database.*

*(****50 Marks****)*

1. Following up the architecture decision (2.2) critically discuss and propose some suitable security measures that should be considered for the new proposed solution. What would be the technical implications, flaws, risks, advantages etc. The discussion should be relevant to the proposed architecture and case study and not to be generalised as something that can be applied to any database. (**20 Marks**)
2. Innovation and excellence - Awarded for extraordinary work that goes beyond minimum scenario description and requirements, including but not limited to organised layout, grammar, references and anything else that is outstanding. (**10 Marks**)

**Requirement 3 | ITEM 3 (CW3 18/12/2022) - 20% (100 Marks)**

1. ***Labs***  - All labs are mandatory and they are marked individually. Each lab should follow [this template](https://docs.google.com/document/d/1tpqq7I-UdnQRZS98MIgLllRmOjyPHAy5OqbxdSJBSY4/edit) for the PDF file and [this template](https://drive.google.com/file/d/1SBOp1VFolze6GBhO4TJehNK_j5P4MTcu/view?usp=sharing) for SQL file. The labs will be evaluated as provided solutions but also reflective statements, general understanding, layout and formatting.

# Submission Guidance

***The deadline for each submission is mentioned in this document in several places. Please make note of those deadlines as no extension will be granted. If you are submitting after the initial deadline it will be marked as Late and the marks are capped to 40% unless a valid EC is in place. It is your responsibility to ensure that you are submitting on time. Do not leave it until the last minute as Moodle can be unresponsive due to high traffic volume.***

The PDF file should contain the EERD, Assumptions (if any), Data Dictionary, TA/TV, physical design decisions considerations, the 6 queries (including RA and RAT for each one of them) and the output of the queries as screenshots, any other relevant material or resource.

The SQL file should contain the entire database CREATE TABLE statements, INSERT statements used to test the database and generate the queries, a copy of all 6 queries with query description.

# CW1 - Submission checklist

* EERD (PDF)
* Assumptions (if ay - PDF)
* Data Dictionary ([use template](https://docs.google.com/document/u/0/d/1TuTHzQEtm79MsXopJv2y-5nK9Q4h2g2fkYSdVnp_D20/edit)) (PDF)
* Transaction Analysis and Transaction Volume (PDF)
* Physical design decisions (PDF)
* Optimisation and monitoring (PDF)
* The SQL code with CREATE/INSERT samples (SQL)
* 6 business related queries (PDF & SQL)
* The RA and RAT for the queries.

# CW2 - Submission checklist

* DB Alternative Analysis (PDF)
* DB Alternative Proposed Solution (PDF - Chose only one)
* Security implications (PDF)

# CW3 - Submission checklist

* All labs are present
* The templates were followed ([PDF](https://docs.google.com/document/d/1tpqq7I-UdnQRZS98MIgLllRmOjyPHAy5OqbxdSJBSY4/edit?usp=sharing), [SQL](https://drive.google.com/file/d/1SBOp1VFolze6GBhO4TJehNK_j5P4MTcu/view?usp=sharing))

# CW1 - Assignment summary

**Please read carefully and submit ALL required elements.**

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Element** | **Description** | **Max Marks** |
| **CW1 - 18/11/2022 @ 23:00 (50%)** | | | |
| **1** | ERD & Assumptions | - Meaningful entities with relevant attributes  - Relevant and correct data type/size  - Correct cardinalities, PKs and FKs  - Good set of justifications and assumptions  - Logical and clear layout (no overlapping lines)  - Crow’s Foot notation (0 Marks for any other notation) | **15**  **(10/5)** |
| **2** | Data Dictionary | * - Correctly layouted data dictionary (use [template](https://docs.google.com/document/u/0/d/1TuTHzQEtm79MsXopJv2y-5nK9Q4h2g2fkYSdVnp_D20/edit)) * - Correctly identified PKs, FKs, AKs and potential INDEXs * - Justified constraints * - Good descriptions (were necessary) * - Tables are in 3NF | **5** |
| **3** | Transaction Analysis & Transactions Volume | - Relevant TA for the business  - Correctly identified all CRUD (IRUD) tables/attributes  - The TV is logical and involves a realistic estimation | **10** |
| **3** | Theoretical Aspects:  -*Physical Implementation*  -*Monitoring & Optimisation* | - Estimation of the database size based on the TA & TV, table data size, indexes, VIEWs, DBMS etc  - Hardware recommendations (CPU, RAM, Storage), costs  - DB and SQL performance tuning  - Security  - Other aspects that would lead to a more efficient database | **20**  **(2x10)** |
| **4** | Physical Design | - The database should be submitted as single SQL file and have all CREATE / INSERT statements included  - You do not need to CREATE the entire database (although it is highly recommended) but ALL the tables (with PKs, FKs, constraints, indexes) and INSERTS necessary to run the 6 queries (1.7).  - The physical implementation follows the ERD and DD as table names, PKs/FKs, attributes, data type/size | **10** |
| **5** | 5 QUERIES | -The queries should be complex enough for a business (not just a SELECT \*) having JOINs, nested queries, multiple WHERE clauses, CASE etc.  -The queries should have the column name output formatted for the end user (e.g. use AS).  -The queries should be included at the end of the SQL file you have created for the database and evidence in the main report with screenshots  -Use of additional features such as constraints etc; Good differentiation among your queries.  -Each query should have the RA and RAT | **30**  **(6x5)** |
| **6** | Innovation & Excellence | - Awarded for extraordinary work that goes beyond minimum scenario description and requirements, including but not limited to organised layout, grammar, references and anything else that is outstanding. | **10** |
| ***Please READ notes.*** | | **TOTAL MARKS 100** | |
| ***Note1:*** *You must submit only 2 files. 1 PDF with the 1.1 to 1.5 and 1.7 and one SQL file with 1.6, 1.7*  ***Note2:*** *The SQL code must be written by you without using 3rd party tools (e.g. pgAdmin, phpMyAdmin). If evidence is found that the code was auto generated, 0 marks will be awarded for the CREATE. You can auto generate INSERTs using* [*Mockaroo*](https://www.mockaroo.com/)*,* [*GenerateData*](https://generatedata.com/) *or any other data generator tool.*  ***Note3:*** *The code should be organised in the correct CREATE & INSERT sequence. All codes will be tested on the Uni’s VMs and you can use PostgreSQL, MySQL or MariaDB.* ***Check the code before submission as if it will not work on a VM it will not be marked.***  ***Note4:*** *Do not submit full code in the PDF (just the SELECT statements for the 6 queries). The CREATE & INSERT statements must be in a separate SQL file.* | | | |
| **Grades Distribution** | | | |
| 0% | | 0 is reserved for non-submissions or plagiarism | |
| 1% - 19% | | Very poor attempt with limited understanding of database principles. Missing elements from requirements, incomplete or grossly incorrect EERD, no code provided or theoretical aspects etc. | |
| 20% - 39% | | Poor understanding of the database principles, incorrect allocation of the PKs/FKs, with minimal coverage of the requirements, errors in code or queries, queries from a single table etc. | |
| 40% - 49% | | Partial theoretical requirements covered, some correct elements in design, some errors in code, basic queries etc. | |
| 50% - 59% | | Mostly a good, logical design with all PKs/FKs and cardinalities correctly represented in EERD, sensible and justified recommendations, correct TA, and security discussion, some errors in code, an average set of queries relevant to the business etc. | |
| 60% - 69% | | Complete CW with very good logical design, including all correct PKs, FKs, cardinalities; Good discussion of theoretical aspects with appropriate examples and references. Good physical implementation with sensible constraints. Good security measures and set of queries etc. . | |
| 70% + | | Excellent and extensive coverage of the study case, good justification of the decisions, exceptional design of the EERD, no errors in SQL code, very complex queries really useful to the business. Very good coverage of all theoretical aspects including solid examples. Almost a “*ready to be used*” database. | |

# CW2 - Assignment summary

**Please read carefully and submit ALL required elements.**

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Element** | **Description** | **Max Marks** |
| **CW2 - 16/12/2022 @ 23:00 (30%)** | | | |
| **1** | DB Alternatives | - A critical analysis and discussion of which architecture would be suitable for migrating from a decentralised database | **20** |
| **2** | DB Distribution Solution | * - Based on your analysis propose (design) the changes for the alternatives * - Chose only one alternative - only some of the bellow will apply dependent of the chosen architecture * - Designs (e.g. data fragmentation, replication, location) * - JSON structure * - Physical partitions, key-values, graph, document etc. * - Anything else that is relevant to the architecture design | **50** |
| **3** | Security Considerations | - Based on the distribution design evaluate some of the security measures that should be put in place  - The measures should be relevant to the case study and proposed solution and not general | **20** |
| **6** | Innovation & Excellence | - Awarded for extraordinary work that goes beyond minimum scenario description and requirements, including but not limited to organised layout, grammar, references and anything else that is outstanding | **10** |
| ***Please READ notes.*** | | **TOTAL MARKS 100** | |
| ***Note1:*** *You must submit only 1 PDF file.*  ***Note2:*** *You don’t need to create any code or to physically implement the proposed design but plan and critically evaluate.*  ***Note3:*** *The code* ***MUST*** *be submitted as PostgreSQL code and be organised in the correct CREATE & INSERT sequence. All codes will be tested on the Uni’s VMs. MySQL, MariaDB, MsSQL code syntaxes will not be accepted. Check the code before submission.*  ***Note4:*** *Do not submit full code in the PDF (just the SELECT statements for the 5 queries). The CREATE & INSERT statements must be in a separate SQL file.* | | | |
| **Grades Distribution** | | | |
| 0% | | 0 is reserved for non-submissions or plagiarism | |
| 1% - 19% | | - No attempt to cover the requirement, very poor scholarship, incomplete CW etc | |
| 20% - 39% | | - Poor attempt at covering CW, no understanding of the database concepts; the proposed solution is ambiguous with different elements from different architectures etc. | |
| 40% - 49% | | - A minimal discussion about DB alternatives; the proposed solution is very general without being referred to the case study etc | |
| 50% - 59% | | A fair attempt of present an workable solution with some discussions about different aspects | |
| 60% - 69% | | - A very good solution that is relevant to the case study having minima errors with good sources of referencing | |
| 70% + | | - Excellent discussion with many relevant examples around the case study; a solution that could be implemented with minimal or no modification; everything is well justified, full and correctly referenced etc | |

# CW3 - Assignment summary

**Please read carefully and submit ALL required elements.**

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Element** | **Description** | **Max Marks** |
| **CW3 - 16/12/2022 @ 23:00 (20%)** | | | |
| **1** | Labs | - All labs are mandatory  - Submit only 1 PDF and 1 SQL file  - The labs should have screenshots (where required) of the query and output in the VM  - Use the provided template for all your labs the [PDF template](https://docs.google.com/document/d/1tpqq7I-UdnQRZS98MIgLllRmOjyPHAy5OqbxdSJBSY4/edit?usp=sharing) and the [SQL template](https://drive.google.com/file/d/1SBOp1VFolze6GBhO4TJehNK_j5P4MTcu/view?usp=sharing)  - Format your code and create a neat report | **100** |
| ***Please READ notes.*** | | **TOTAL MARKS 100** | |
| ***Note1:*** *You must submit only 2 files. 1 PDF and one SQL file.*  ***Note2:*** *The SQL code must be written by you without using 3rd party tools (e.g. pgAdmin). If evidence is found that the code was auto generated, 0 marks will be awarded for the CREATE. You can auto generate INSERT using* [*Mockaroo*](https://www.mockaroo.com/)*,* [*GenerateData*](https://generatedata.com/) *or any other data generator tool.* | | | |
| **Grades Distribution** | | | |
| 0% | | 0 is reserved for non-submissions or plagiarism | |
| 1% - 19% | | - Minimal or no coverage of the labs; Missing many of lab requirements; | |
| 20% - 39% | | - Some labs present where most of them are incomplete; | |
| 40% - 49% | | - Poor attempt at addressing the lab questions; very cluttered and hard to follow/read the PDF document; missing SQL code etc. | |
| 50% - 59% | | - A fair attempt to answer most of the questions from the labs; Nicely layouted report; Some useful reflections. | |
| 60% - 69% | | - A very good attempt to address all the labs; Logical and easy to follow report; Personal inputs, conclusions and reflections were present | |
| 70% + | | Excellent report; All labs are present and correctly solved; Extremely well presented document; Innovative ways to address some challenges. | |

The CWs will be assessed also, in general for the following criterias:

* Rational and Logical aspects in relation with Database Principles and case study (CW1 & CW2)
* Suitable, sensible and complete set of attributes with appropriate data type and size (CW1)
* Organisation and layout of designs and report document (CW1, 2, 3)
* Discussion and justification of theoretical elements (CW 1,2)
* Logical and suitable architecture (CW 2)
* Database optimisation (through appropriate data type/size; indexes) and query optimisation (CW1, CW3)
* SQL formatting and errors (CW 1,3)
* Overall complexity (logical & physical) (CW 1,2,3)
* References, grammar & punctuation (CW 1,2,3)

***Marks are released after 20 working days***

***(excluding weekends and bank holidays)***