INF10002 Database Analysis and Design

Learning summary report

You will submit this report as evidence of the tasks you submitted as part of your Assignment 1: Learning portfolio. This will be checked against the tests you completed and the tasks you submitted and will determine your final grade.

Name: Luong Trac Duc Anh

Student ID: 103488117

Section A: Self-assessment

Self-assessment grade application

The following checklists provide an overview of your self-assessment for this unit. Tick the box with the grade you are applying for.

	Pass	Credit	Distinction	High Distinction
Self-assessment (please tick)				X

Self-assessment statement

Tick the boxes in the following checklist that indicate the tasks you have submitted to support your grade application.

Pass (Minimum Pass checklist)	Included / Completed [please tick]
Learning Summary Report	X
Pass task 1	X

Pass task 2	X
Pass task 3	X
Pass task 4	X
Pass task 5	X
Test 1	X
Test 2	X

Credit (Minimum Credit checklist)	Included / Completed [please tick]
All requirements met in Pass criteria above (including the Learning summary report)	X
Credit task 1	X
Credit task 2	X
Credit task 3	X
Credit task 4	X
Credit task 5	X

Distinction (Minimum Distinction checklist)	Included / Completed [please tick]
All requirements met in Pass and Credit criteria above (including the Learning summary report)	X
Distinction task	X
Test 3	X

High Distinction [Minimum High Distinction Checklist]	Included / Completed [please tick]

All requirements met in Pass, Credit and Distinction criteria above (including the Learning summary report)	X
High distinction task	X

Section B: Reflection

Part 1

For P, C, D & HD

Data has always played an instrumental part in our lives, from keeping track of our media records to managing and developing a business company. However, most of the definitions and attributes of data are not clearly understood by the mass, or simply people who do not work in the field of computer science or data analysis. After completing the course INF10002 – Database, Analysis and Design, I have obtained a better grasp of the concept of data structures, how the data interact with each other, how we can utilise the data to generate information and knowledge that can be a decisive factor in the moderation and development of a business operation.

As the course's topic is new and refreshing to first-year students, there would surely be aspects of the course that is considered interesting or challenging. In this reflection part of the report, I will explain in detail those different viewpoints, and how my learning process has changed after encountering those challenges.

First, I will elaborate on the interesting part. Throughout my 12 years of basic education and even the first semester of university, data to me are simply records that are presented in a table, whether it is using Microsoft Word or for more technical and professional purposes – Excel. There is not much context on how those data should be utilised for the expansion of businesses, and everything was simply to keep track of students' or customers' information, in case of contact or executing a promotion (marketing) plan for a business. My point of view has changed drastically after viewing the course materials, starting in the first week.

In the first week of the course, I get to understand more about the concepts of data, different data categories, RDBMS, servers, SQL interfaces and the introduction to Microsoft Access. Data is important and can be divided into 3 categories: structured, semi-structured, or unstructured. Companies can use data mining techniques to learn the customer buying habits and timing, patterns in purchases, sentiment toward a specific product, etc. That is why data needs to be divided into different sections for management to search and process. For example, structured data with clear patterns and formats like student records are easy to moderate and manage in comparison to unstructured data like social media posts or emails (containing text, images, videos, and other different file types). However, the future opportunity that any data forms

are different and valuable in various aspects. The way of storing data in DBMS and Access was introduced in detail, and for me, the power of query searches in comparison to the humble datasheet filters is truly innovative and intriguing.

Moving to the second week, the presumption of individual data tables was no longer there, as I got to know about relationships, indexes and how queries are used with those concepts. Relationships can retrieve data from multiple tables, saving not only time by increasing productivity, but also by creating knowledge nodes that connect information, assisting business managers in their decision-making process.

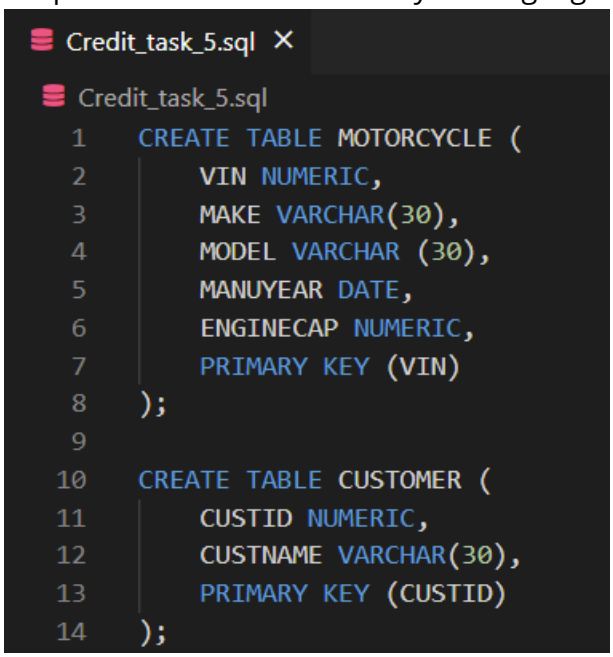
In week 3, the course focused on data manipulation. In addition to high-speed queries, the search of data can be even faster with parameters, a piece of information that the end-user supplies when a query is executed, showing only the necessary fields that can satisfy the user's input. There was a discussion on how to import data efficiently, with Access functionality of importing from various file types (XML, Excel, text, database to database, etc.) triumphed the traditional way of re-typing, which was widely considered as costly, prone to errors, and slow.

For week 4, students got introduced to another powerful tool for data visualisation, which is Power BI. BI stands for Business Intelligence, which analyses the given datasets to make decisions for a business easier with fewer risks involved. There are various forms of visualisations to accommodate the user's needs. For example, fluctuations and changes over time will require a line graph, and workload allocation can be presented in a pie (or doughnut) chart, using a percentage as a form of measurement.

In week 5, we got a more detailed view of the types of keys in addition to the key's input fields that had already been covered in Week 1. ERDs (Entity Relationship Diagram) were introduced as a logical representation of data that strictly follows business rules. After that, we examine the methods of translating the ERD to a Relational Schema, which can assist us in the next progress. The most challenging part of the course was SQL (Structured Query Language). As a student working on the ICT bachelor's degree, I still initially find the language challenging, with syntax not like any programming languages like Python or Ruby. However, after a week of detailed examining and practising using both the course's content and online free materials (W3Schools is a great website for learning SQL that was suggested in the additional reading on Canvas). After week 5, the modules were also structured in a mix of terminologies and sample SQL statements. There are a few tools that also assisted me in the process of learning SQL, two of which that I will mention are Visual Studio Code and Notion.

- With Visual Studio Code and its built-in syntax highlighting function, writing SQL statements for our Pass and Credit Task from 3 to 5 was much easier and

more efficient, with few errors presented in the code itself. The screenshot below is an example of Visual Studio Code syntax highlighting.



- With Notion, the story is again, syntax highlighting. Notion allows us to create code snippets in addition to its various note-taking functionality. It is easy to search materials for revising with the Find Command (Ctrl+P) and create various sub-pages and headings for me to learn. Here is an example of Notion note-taking:

After getting over those difficulties, I have learned the fact that in addition to studying diligently, having the right tools and methods is also as important.

In Week 6, we learned about foreign key constraints and participation constraints. This was a more insightful look into the characteristics of ERD, and how the table can enforce its data fields with necessary rules (like parent and child rules) to avoid errors or Cartesian products. Some additional SQL statements that the week covered were the SELECT clause, INNER JOINS, and operators like AND, OR, NOT.

Moving to Week 7, the relationship between entities got more complex, as composite keys and weak entities were introduced. This is a holistic solution to expand and remove M:M relationships, therefore removing redundant data generated. SQL for aggregate functions is similar to the Total function in Access (Sum, Count, Avg, Min, Max), which can help us either retrieve needed information or create complex queries.

In Week 8, we could then design our ERDs through the detailed steps presented in the lecture time. There were aids from Glenferrie Toy Library Narrative, going through the 9 crucial steps in a case study from the university. In addition to that, many complex SQL functions and statements like OUTER JOINS, INTERSECTION & MINUS were shared for students who wanted to achieve higher grades from the compulsory Pass.

In Week 9, we learn how to create a View, a simple but powerful block of SQL code that even though does not store any data, can be used to select millions of rows. A view will be executed once it is referenced. Different from Week 8, in which we mostly used pen and paper to draw out ERD, an application called Visio from Microsoft was introduced for us to create a visualisation of the ERD. Technical problems arrived as Visio can only be accessed through the school virtual workspace (Citrix), and the web version of Visio had many limitations, we came to the holistic approach of using draw.io, a free to use website that can be used to create any diagrams, from Swimlane to ERDs and so much more.

Once the students had already had a competent view of data, RDBMS and visualisation tools, fixing and modifying tables with redundant data will be the next step, Therefore, in Week 10, we got to learn the ways of normalisation, how the dataset can be converted to 1NF, 2NF, and 3NF. NoSQL database and JSON were introduced as well, another interpretation of data, as the traditional columns and rows in a table cannot satisfy the high volume of queries and searches that big corporations need.

In the last week, the course material mainly focuses on database transactions and storing data in warehouses. In a realistic context, any modern DBMS can commit

errors, sometimes because of external factors like network or power source. Therefore, knowing how every transaction is either committed or rolled back is important, as we can debug or fix the action in the shortest time possible. In the case of banking transactions, ensuring how each command is executed successfully is important to the user's assets and trust in the bank system.

In conclusion, the course poses many challenges with new technical terms, concepts, different applications and tools to manipulate, utilise and manage the data, as well as query languages like SQL or JSON. However, the process of learning, and enhancing our skills through those challenges to obtain knowledge and solve problems is fundamental for our future working careers, whether it will be in the field of Data Science or Computer Science in general.

Part 2

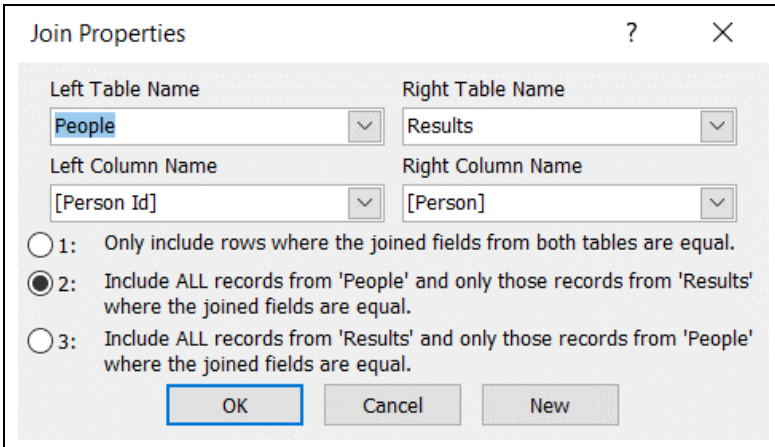
For D & HD

In addition to the normal Pass and Credit tasks that we had every 2 weeks, there is also a Distinction task for us at the end of the semester. The assignment went more in dept in the requirements of creating ERDs, writing queries (using SQL or MS Access) and creating insightful visualisations with Power BI.

Normally, we will have datasets already created and only need to do the specific tasks in the requirement. However, this time, the Distinction task involves generating different tables in the form of a .csv file using an Excel macro button. The data will be then needed to be imported into Access, declare the appropriate columns for primary keys, and create relationships. An additional step of fixing some minor errors with the dataset is also required as columns that have duplicated values cannot have their relationships reinforced by referential integrity. Each student will input their unique student ID, therefore curbing the chance of people copying each other's work.

Different from the previous tasks, the 2 narratives of the Distinction tasks were more complex, with different entities, attributes, and relationships. Upon completing the necessary business requirement in terms of query search and different aggregate functions, students could understand in detail different case studies and therefore can enhance their critical thinking skills in real working scenarios in the future.

In terms of the queries and visualisations, most are manageable, except for the Q6 in Part 4, which requires us to find the null values in the Results table. It took me quite some time to solve joining the 2 tables (People and Results) based on their properties where we need to include all records from 'People' and only those records from 'Results' where the joined fields are equal. (Screenshot below)



The visualisation was another challenging part, as we have to be flexible when creating our visualisations, with the proper graphs, fields displayed and the specific requirements for them.

Part 3

For HD

Different from the Distinction task, in the HD task, the challenge was posed right from the start, as students are required to create their business narrative and get confirmation from the Unit Coordinator. After the scenario has been approved, we need to create our own data tables, inputting every field and setting the necessary relationships.

A scenario can be disapproved because of the practical value that it brings, how feasible will the process of creating the data be, or the formatting and word choice of the narrative can also be rejected. A good place to start would be revising some of the old given narratives to give us insights into the necessary business requirements, and what type of data, columns, and relationships will be involved in the process of creating the data. Combining the 2 factors of given materials and your ideas will be more efficient and effective in comparison to entirely making up a scenario.

Moving to create the data tables, the appropriate number of entities and attributes for each has been briefly mentioned in the assignment instructions (7 entities, 8 queries and visualisations minimum), also with your validation in the creation of an ERD. If these aforementioned steps are executed properly, then making the datasets for the next steps will be more manageable, as you only need to consider the number of records that each table has, preferably less in the strong entities and more in the weak entities.

The reports and visualisations you created had to be complex, with aggregate functions and not just general statistics retrieved from a single table. Throughout this

step, you can be creative with your solutions and overall get a comprehensive view of how data works based on your proposed narrative.