**BSc in Data Science - BSHC/E2 HCCOMP/E BSHDS HCDS**

**Advanced Databases: Project (40% of Overall)**

**Section A - Design and Implement a Data Mart - 70%**

**Section B – Open Book Moodle Quiz- 30%**

**Due before 15th November 2023 at 23:55**

**This is a group project, 2-3 members per group.**

There are two parts to the Assessment, Section A must be answered by each group. Section A accounts for 70% of this assessment. Section B must be completed individually on Moodle in an open book exam format, 30%. The open-book exam will be hosted during class time for 1-hour on the week ending November 17th.

SECTION A - 70%

**Summary:** Students will construct a Data Mart. They will identify an important business process and associated business requirements. Their design will include a dimensional model that will support the data mart requirements. This should be constructed using MySQL scripts to create a corresponding database and set of tables to represent the dimension and fact tables. These dimension and fact tables should be populated.

The business process will be chosen from the list of industries and business areas found in the table below. Both group members student numbers should be used to select the industry and business area.

Student 1 (the student with the lowest student number): using the last digit of your student id, please select a topic from the industry column.

Student 2 (the student with the highest student number): using the last digit of your student id, please select a Business Area.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Industry** |  | **Business Area (Entity relationship)** |
| **0** | Agriculture | **0** | Customer Support (Customers – Cases) |
| **1** | Sports | **1** | Invoicing (Invoice – Items) |
| **2** | Health | **2** | Purchase Orders (PO – Items) |
| **3** | Advertising | **3** | Logistics (Parcel – Contents) |
| **4** | Leisure | **4** | Inventory (Item – Categories) |
| **5** | Retail | **5** | Payroll (Employee – Entries) |
| **6** | Finance | **6** | Recruiting (Candidate – Skills) |
| **7** | Airlines | **7** | Training (Module – Students) |
| **8** | Food | **8** | Sales (Orders – Items) |
| **9** | Education | **9** | Productivity (Projects – Tasks) |

These choices must be used to frame the business process for your data mart. Your group may implement the data mart using either a Star or Snowflake schema.

In order to implement the Data Mart students are required to produce:

1. At least one important business requirements for each student (**each student is responsible for the marks of their business requirement individually**), based on their selected industry and business area, for the Data Mart. Below is a single example of a requirement:

Education payroll: the payroll department for a school or college may want to analyse multiple dimensions to identify when the school was most in need of additional teaching staff.

To do this they could examine a time dimension alongside the number of staff they employed at that given time.

Another dimension table that logged tickets with the teaching support centre could be used to highlight the need for additional teaching staff for a particular module.

Based on these dimensions’ payroll could highlight what stages in the year additional teaching staff were required for specific modules.

Food Inventory: The inventory for a restaurant may want to observe multiple dimensions to examine why ingredients are profitable or whether they ought to raise price or cut costs and obtain the ingredient in a different manner.

To do this they could examine an ingredient dimension to see what produce is bought and sold at.

They could also see which country the produce has arrived from to see if the distance the product travels has proportionality with the profit margin.

Then they would examine the characteristics of the ingredients to see whether the status of gluten within the ingredient has a great impact on the profit margin.

1. A dimensional model using either a Star or Snowflake schema for the Data Mart to support the business requirements. Dimensional model to clearly show the tables, column names, primary keys and foreign keys **(shared marks)**.
2. An SQL Script to create the Fact table and the Dimension tables **(shared marks)**.
3. An SQL script to load data into the Data Mart from (for example the Sakila database) and evidence of the use of manual or automatic data loading into the Data Mart. As there may be quite a lot of data that needs to be loaded into some of the tables it will be sufficient to just present a subset of the data in the document **(shared marks)**.
4. At least one query per student and their outputs to demonstrate that the requirements are satisfied **(each student only requires a query to satisfy the business requirement they specified in part 1)**.
5. The students should highlight changes they made to the structure of the data mart to optimize read access for their specific business process and the associated requirements they listed in part 1. **Each student should highlight how this optimization has an effect on their specific business requirement (whether it is positive or negative).**

Any assumptions that are made must be stated clearly.

All this should go into **one** document and be submitted on or before**, 15th November**. Please include your student names in the filename. One report is to be submitted, with the 2-3 team members names to appear on the cover page.

Breakdown of marks is as follows:

Requirements 20%

Dimensional Model 15%

DDL Script 10%

DML Script 10%

Queries 20%

Optimizations 25%

Background OLTP Database

Students may use the Sakila sample OLTP database during the investigation process for development of their data mart. The Sakila sample database was initially developed by Mike Hillyer and is intended to provide a standard schema. An ER model for the database and a description of each table can be found at:

[**https://dev.mysql.com/doc/sakila/en/sakila-structure.html**](https://dev.mysql.com/doc/sakila/en/sakila-structure.html)

The database itself can be downloaded from:

[**https://dev.mysql.com/doc/index-other.html**](https://dev.mysql.com/doc/index-other.html)

|  |  |
| --- | --- |
|  |  |