BSc (Hons) in Information Technology
Specializing in Data Science

IT3021 – Data Warehousing and Business Intelligence
Year 3
Assignment 1 - 04 Sections (Answer All 04) **2024**

Duration: 3 Hours

Complete following tasks and demonstrate the same with SQL Server (any version & edition). Additionally, document the steps followed in completing the tasks. Include the screen shots of the steps you followed for each step in the report and upload into shared location.

Step 1: Download the SQL script "MyMusic Sample Database.sql" from the shared location and execute SQL script in the SQL Server Management studio environment. **(5 Marks)**
Draw an ER diagram for the DB. **(10 Marks)**

- Ensure, there should be a database called **'MyMusic'** under the database menu in the SQL Management studio environment.
- There should be 11 tables with or without data.

Documentation: provide a screenshot of the Database menu and each table's data by executing T-SQL query. (First screen first line should show that this was done on your machine and the Table structure inside the newly created DB. Second screen(s) can be record counts of each table or screenshots of each table by showing data using a select query).

Step 2: Owners of the "MyMusic" portal wants to analyze and see how their sales are using past data. This can be timeline based, track based, or they should be able to analyze based on any other sector with data. (there are 11 tables with data). Most of these can be converted to Dimensions individually or after combining 2,3 tables together to analyze against order data. There can be some which can't be brought in as they might create duplicates.

Design the high-level Data Warehouse solution for the above dataset. This should be of star schema. (Do not use snow flake). (Answer should be a star schema shown diagrammatically including column names) (**25 Marks)**

Answer the Below questions based on the given tables and your understanding of the solutions needed. **(5 marks)**

I) What are the table(s) that cannot be used in the main model is there is any?
II) If there are any tables, why they can't be taken into the model?
III) If this needed to be taken in what is the data element needed in which table it should be?

Documentation: provide an architectural diagram to describe the components of your DW & BI solution. (This can be generated after doing step 3 using Power BI which should be accepted as well)

Step 3: Develop the Data warehouse architecture in your SQL server environment. **(25 Marks)**

- Design a data warehouse schema (Table Structure) for the above in step 2.

    - This should be of star schema. (Do not use snow flake)
    - Ensure you have at least two (2) dimensions apart from common dimensions such as Date
    - Ensure you have only one (1) fact table
    - Ensure you have a slowly changing dimension for the customer table in the OLTP.
    - Implement the data warehouse schema in SQL Server

You can use a script from web to generate time/date dimension with generic and suitable columns.

Documentation: You must attach scripts used to create the data warehouse dimensional model and screen shots using the SSMS environment after creating your tables.

Step 4: Develop the ETL package to load OLTP data into Data Warehouse. **(30 Marks)**

- Develop the ETL using SSIS for data extraction, transformation, and loading.
- Ensure you have sufficient SSIS tasks to demonstrate your capabilities around ETL process and implementation.
- Customer dimension should be in SCD Type 2.
- After loading data into dimension and fact table run the “CustomerUpdate.sql” query provided. Then load the Customer dimension again which should give you Customer records loading with updates in SCD Type 2.

Documentation: describe the steps in the ETL process by including screenshots of each step. This should include SSIS screen shots as well.
Show data loaded into dimensions and facts with screenshots.
Show SCD Type 2 loading using screens of data from Customer dimension.