**Data Warehousing**

**Data Warehouse Final Project**

**Task 2B – Project**

**Date Due (See Canvas)**

This task will require you to complete the design for your data warehouse and to load your data warehouse with the source data you have identified. Once your star schema has been loaded, you will need to create basic reports that answer the business questions you have previously identified (but you may change these if you wish). The following sections describe the requirements for completing your course project.

**Please note that all project design requirements in the previous assignment must still be met. You may want to assign one member of the team to a QA (Quality Assurance) role. This person can proactively verify that all requirements have been met.**

# Final Design

You will complete the design of your star schema data model. The model you submit should be *identical* to the model that you build in your data warehouse database. You should make updates as necessary for unforeseen problems or to address issues that were identified in your initial design.

This model does not need to match the model from previous project tasks but must be identical to the implemented data warehouse.

You should develop your data model using database project type in MS Visio, or you can submit an image of the database diagram created directly in SQL Server. Models generated as pictures in PowerPoint or another drawing tool will not get full credit. My intent is to force you to define the relationships correctly between the fact and dimension tables.

# Star-Schema Database Development

This step requires you to create your database you presented (or will present) in class in SQL Server. Your star-schema should mimic the tables in your design and contain data loaded that answers your business questions. Be sure to adhere to all rules we’ve discussed in class including the use of surrogate keys and values for “Not Applicable” or “Does Not Apply” so that it is easy to create reports from your model. (A row count on the fact table should be equal to a row count joining all dimension tables. If it does not, a problem is indicated.)

# Type II Dimension Change

You must simulate a Type II dimension change in at least one field of a dimension table. It does not need to be an *actual* data change – it is okay to manually change the data. You must have rows from both the new and old values for the Type II change in the fact table for appropriate time periods that make sense. The goal of this step is to make sure you are familiar with the concept of what it means to have a Type II dimension.

# ETL Procedure Summary

You must submit an MS Word document that describes how you loaded the data into the database. This should be a summary at the table level. You likely will only need one or two sentences that describe how you loaded the data for each table. Just describe the steps that were performed in plain language. Here is one example.

“I used Microsoft SSIS to import the data into a staging table called StageProduct. Then I wrote a merge query to move the data into the dimension table. Finally, I changed the source data in the StageProduct table manually and reloaded the DimProduct dimension to simulate a Type II dimension change.”

However, I do need to see all the steps that you did take to get from the original data source to your final data warehouse tables, so please do not omit anything.

# Database Submission

1. You will need to upload your exported SQL Server database file to a storage location (this location will be given to you soon). To export your database, you will need to use the SQL Server “Generate Scripts” command to do this. Directions for generating a database script can be found at the end of this assignment. Please note that these instructions are slightly different to those given in Task 1 because the database I get from you must contain data as well as your table structures.

## Data Warehouse Reports

This step requires you to create reports to answer four different business questions. You need to have four types of reports:

1. Report of any type to introduce your data and business

1. Report containing a normal aggregation (AVG, SUM, MAX, etc.) employing a windowing function technique. (Note: you define a window by using the OVER() clause). Ensure you use windowing with the aggregation.

1. Report containing an analytical windowing function. (ROW\_NUMBER, RANK, etc.)

1. Reporting utilising hierarchical results. (ROLLUP, CUBE)

You do not have to fully answer each business question, but you need to create reports that at least partially (to a satisfactory level) answer the questions. Each report should target a different business question and you cannot use the same report for more than one of the following categories. Each report should make sense by quickly looking at it; do not simply submit reports that are a series of numbers. If you don’t have the descriptive data to make a report meaningful, feel free to create the descriptive data necessary to do so.

Discuss at least 2 additional business questions that would be derived from the results of any of the above queries. This is the next round of analysis driven by these results. You do not have to create the reports. (Can be generated for extra credit.)

Discuss one additional data source that if available would allow you to further analyse your business. (Can be included in above business questions.)

**Report Submission:** You need to submit your reports in the following manner:

1. Create an Excel workbook with the following worksheets:

* 1. The data results for each report that shows how it answers the business question. You should copy the results from whatever option you choose into Excel. This is supposed to be easy for me to understand, so I suggest applying some level of formatting to make the report presentable.
  2. Submit the SQL statements that you ran. I will test them, so make sure they run against your database.
  3. The additional questions / data source should be added as an additional tab in the workbook.

## Submitting Your Project

Note that there are several deliverables as part of this project, and correct submission of these deliverables is part of the project score. Please be sure to submit all of them correctly.

*It is suggested that you assign one group member to be the “Compliance Officer” to make sure all deliverable requirements have been met.*

1. ***Project Files***

Due by date specified on assignment

Submit via Canvas a zip archive with the name Group\_X\_Project.zip where X is your group number. The archive should contain the following files:

* 1. Group\_X\_Design\_Final.vsd – Data Model design in Visio format. If you are not using Visio submit a PDF file and an explanation of the tool you used along with the original file format. This may be identical to the previous file you submitted, but please submit it again with your final project.
  2. Group\_X\_ETL.doc – The ETL procedures you used to load your database. Remember, that all steps need to be documented.
  3. Group\_X\_Reports.xlsx – An excel spreadsheet containing four worksheets with your report information as described above. Please submit the SQL in the Excel file with the report results as noted in the Reports section above.

1. ***Database Deployment***

Due by date specified on assignment

Deploy your database to the class server as specified above. Your final database should contain ONLY the tables that comprise your star-schema model.

*Note: You cannot deploy your database by restoring a backup file. You must follow the instructions below to generate a database script.*

Please be sure to submit your project by the deadline. Refer to the syllabus for penalties and procedures regarding late homework. I strictly enforce these policies. If you run into any issues submitting files to Canvas, please contact me as soon as the issue arises.

## Database Submission Directions

Generating an SQL Server Database Script

Here is a general outline of the steps required to generate a script to copy your SQL Server database

* 1. In SQL Server Management Studio, right-click on your database and select Tasks

Generate Scripts to open the Generate and Publish Scripts wizard

* 1. On the first screen choose “Select specific database objects”
  2. Mark the checkbox next to the tables you want to script and click “Next”.
  3. Click the “Advanced” button
  4. Under “General -> Script USE DATABASE” choose FALSE.
  5. Under “General -> Types of Data to Script” choose “Schema and Data”
  6. Choose an appropriate location to save the file and click “Next”.
  7. Complete the wizard to generate the script file

Once you have generated your script, you will need to upload this file to a location that I will provide for you at a later date.

Note: You can optionally generate separate scripts for the structure of the database (schema) and the data. This can be useful if you want to load the data more than once, or if you want to use SQL Server import tools to move the data more than once.