Project 2 Milestone 2: Extract, Transform, and Load

<div>Submit Assignment</div>

Due Apr 20 by 11:59pm   **Points** 25   **Submitting** a file upload

Available Mar 21 at 6am - May 10 at 11:59pm about 2 months

<div align="center">

ACIS 5504 – Project 2

Data Warehouse

Milestone 2 Extract Transform and Load

25 Points

Due: See Assignments Page in Canvas

</div>

Overview

In this milestone you use the DW you designed to address management's top three questions in milestone 1. You first are asked to review and understand the two operational databases from the resort chains. Then you are to evaluate the operational databases to determine that they contain (or perhaps do not contain) the data needed for your DW. Next you write two SQL procedures to simulate the extract, transform, and load process for the DW by modifying example procedures. Finally, you populate your DW with sufficient rows to demonstrate your understanding of data in your DW and provide results for your queries in Milestone 3.

A detailed project report should be submitted on the due date. This project is to be done individually. This may require additional reading and research.

This project will require more than 24 hours of work time to complete.

To complete assignment, you should complete the following activities:

1) Review the ER models and data dictionaries for the operational DBs carefully.

2) Think about the data needed by your DW and the data available from the operational DBs.

3) Complete the activities listed below under submission requirements.

Assignment Resources

1) Your DW design and DDL from Milestone 1;

2) Recorded lecture on Chapter 13;

3) Recorded lecture slides on Chapters 13;

4) The business situation description provided in Milestone 1;

5) ER diagrams for two operational databases for the business situation to review: **Corp1ERD-Revised-Ayyaz.jpg** , **Corp2ERD-Revised-Ayyaz.jpg**

6) Data dictionaries for two operational databases for the business situation to review: **Corp1Data Dictionary.xls** , **Corp2Data Dictionary Revised.xls**

Business Situation Description:

You work for a large corporation that has just purchased 2 hotel and resort corporations each consisting of over 100 hotels. Each Corporation operates a custom database. You are provided the data dictionary and ER diagrams for the two operational databases.

(Note: The databases you will evaluate come from student groups in another class responding to the Hokie Resort problem you reviewed in Milestone 1.)

Submission Requirements (start with #5 to pick up where M1 ended with #4)

When completed, please upload your documents under the "Assignments" tab in Canvas. Please use a document naming convention that allows us to determine your name. EXAMPLE: "LastName_FirstName_Project2Milestone2.docx".

5 Analyze the ER diagram and data dictionary from both of the operational databases to determine if the two operational hotel databases have the data needed for your data warehouse design.

For each DB, create a mapping that shows the tables from that DB that are used to create rows in your data warehouse tables. For each data warehouse table, describe how the operational data is aggregated to create a row in the table. Submit your mapping and aggregation summary in the following format.

DatawarehouseTable	Operational DB Table	Aggregation/Sum
PatientDim	Corp1: Patient	No aggregation, each row is an instance in the DW.
PatientDim	Corp2: Customer	No aggregation, each row is an instance in the DW.
MediationsFact created.	Corp1: Prescriptions	Count and average amount of drug given are
MediationsFact are created.	Corp2: Drugs	Provided Count and average amount of drug given

(Note: If an operational database does not contain the data needed for your data warehouse design, then propose revisions to the existing tables in the DB or define additional tables to be populated in the DB so that it will contain the data needed for your data warehouse). (5 points)

6 Write the SQL necessary to extract and process the data from the two operational databases so that it will be suitable for your elements of your data warehouse.

The mapping you made in question 5 should help in this process. This requires writing SQL procedures that include SELECT statements from the operational DBs and INSERT INTO the data warehouse tables (Note: you do not have to make the procedures work, as you have only the designs of the operational DBs).

You should write the SQL procedure code to extract and load data for two of your dimension tables and one of your fact tables from each operational DB. (Note: you will not be able to test your procedure, yet it should be as correct as possible.)

In addition, you should include the correct time dimension data on your fact table rows. Pay attention to the correct grouping and aggregation necessary to transform the operational data into the form needed for your data warehouse.

Two example procedures are provided using the Hokie Hospital problem in these files: .

Dimension table: **Q5_SampleProcedure_Hokie_Hospital_DimensionV2.sql**

Fact table: **Q5_SampleProcedure_Hokie_Hospital_FactV4.sql**

Your task is to make similar procedures that will extract the data from the operational databases into your data warehouse design.

7 Populate MySQL or other DB with sufficient rows to demonstrate your Data Warehouse

You are required to create rows for all dimension (including time) and fact tables in your Data Warehouse. You must insert sufficient rows in the DW to be able demonstrate that the data warehouse can answer the management questions and provide data for the visualizations in milestone 3. This means that the data you insert must provide consistent keys and foreign keys to allow implementation of the ROLLUP queries needed for milestone 3. (10 points)

This script, consisting of insert statements, is from the textbook and loads a sample data warehouse: **DW-DBINIT.SQL**

Assessment

Your submission document will be checked for correctness and completeness.

Connecting Assignments

This milestone uses the DW design from Milestone 1 to create a database populated with data to needed to complete the requirements for Milestone 3.