

D191 – ADVANCED DATA MANAGEMENT

AUTOMATING DATA INTEGRATION

INTRODUCTION

Data analysts frequently perform periodic extraction of data from a larger database. They then use this extracted data for analysis. A single data extract may serve different organizational needs such as populating an executive summary or providing a complete, detailed collection of data. This task will emulate such a process where you will create a repeatable data export, import, and analysis for various stakeholders.

In this task, the data will be extracted and analyzed within the same database to reduce the technical challenges of manipulating multiple databases.

This task defines a report as a collection of data that answers a real-world business question. The report contains two distinct sections that differ in the granularity of the data they present and how directly they support the answering of the business question of interest. The Detailed section will contain all data that informs the answers to the question, and the Summary section will include contains relevant aggregated data that provide the answer to the business question. For example, if the business question of interest in the context of a computer factory is focused on the number of the computers manufactured in the past six months, the detailed portion of a report could contain details on each individual computer made, while the summary portion contains only the total number of computers made in each factory.

Your business question must rely on the aggregation of data. In this task, the summary data will be automatically created from the detailed data.

REQUIREMENTS

Your business question must rely on the aggregation of data. In this assessment, the summary data will be automatically created from the detailed data.

Plan for and compose the sections of a real-world business report that can be created from the DVD Database on Labs on Demand (see link below), and demonstrate the functionality of the supporting SQL code by doing the following:

- A. Summarize **one** real-world business report that can be created from the attached Data Sets and Associated Dictionaries.
 1. Describe the data used for the report.
 2. Identify **two** or more specific tables from the given dataset that will provide the data necessary for the detailed and the summary sections of the report.
 3. Identify the specific fields that will be included in the detailed and the summary sections of the report.

4. Identify **one** field in the detailed section that will require a custom transformation and explain why it should be transformed. For example, you might translate a field with a value of 'N' to 'No' and 'Y' to 'Yes'.
 5. Explain the different business uses of the detailed and the summary sections of the report.
 6. Explain how frequently your report should be refreshed to remain relevant to stakeholders.
- B. Write a SQL code that creates the tables to hold your report sections.
 - C. Write a SQL query that will extract the raw data needed for the Detailed section of your report from the source database and verify the data's accuracy.
 - D. Write code for function(s) that perform the transformation(s) you identified in part A4.
 - E. Write a SQL code that creates a trigger on the detailed table of the report that will continually update the summary table as data is added to the detailed table.
 - F. Create a stored procedure that can be used to refresh the data in *both* your detailed and summary tables. The procedure should clear the contents of the detailed and summary tables and perform the ETL load process from part C and include comments that identify how often the stored procedure should be executed.
 1. Explain how the stored procedure can be run on a schedule to ensure data freshness.
 - G. Provide a Panopto video recording that includes a demonstration of the functionality of the code used for the analysis and a summary of the programming environment.
 - H. Record the web sources you used to acquire data or segments of third-party code to support the application if applicable. Be sure the web sources are reliable.
 - I. Acknowledge sources, using in-text citations and references, for content that is quoted, paraphrased, or summarized.
 - J. Demonstrate professional communication in the content and presentation of your submission.