VDM1 — VDM1 TASK 1: AUTOMATING DATA INTEGRATION

ADVANCED DATA MANAGEMENT — D191
PRFA — VDM1

TASK OVERVIEW | SUBMISSIONS | EVALUATION REPORT

COMPETENCIES

4037.4.1 : Writing Complex SQL Statement
The graduate writes complex SQL statements in order to implement functions, stored procedures, and triggers to prepare data sets for data analysis and manipulation.

4037.4.2 : Configuring Extraction, Transformation, and Loading (ETL) Workflows
The graduate configures data extraction, transformation, and loading tasks in order to automate 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 submission must be your original work. No more than a combined total of 30% of the submission and no more than a 10% match to any one individual source can be directly quoted or closely paraphrased from

sources, even if cited correctly. The similarity report that is provided when you submit your task can be used as a guide.

You must use the rubric to direct the creation of your submission because it provides detailed criteria that will be used to evaluate your work. Each requirement below may be evaluated by more than one rubric aspect. The rubric aspect titles may contain hyperlinks to relevant portions of the course.

*Tasks may **not** be submitted as cloud links, such as links to Google Docs, Google Slides, OneDrive, etc., unless specified in the task requirements. All other submissions must be file types that are uploaded and submitted as attachments (e.g., .docx, .pdf, .ppt).*

To work on this task, please follow the link in the Web Links section below to the LoDs PostgreSQL lab environment. In this environment, you will be able to write and test your PostgreSQL code and access the databases to complete this task.

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.

Note: For instructions on how to access and use Panopto, use the "Panopto How-To Videos" web link provided below. To access Panopto's website, navigate to the web link titled "Panopto Access," and then choose to log in using the "WGU" option. If prompted, log in using your WGU student portal credentials, and then it will forward you to Panopto's website.

To submit your recording, upload it to the Panopto drop box titled "XXX." Once the recording has been uploaded and processed in Panopto's system, retrieve the URL of the recording from Panopto and copy and paste it into the Links option. Upload the remaining task requirements using the Attachments option.

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.

File Restrictions

File name may contain only letters, numbers, spaces, and these symbols: ! - _ . * ' ( )
File size limit: 200 MB
File types allowed: doc, docx, rtf, xls, xlsx, ppt, pptx, odt, pdf, txt, qt, mov, mpg, avi, mp3, wav, mp4, wma, flv, asf, mpeg, wmv, m4v, svg, tif, tiff, jpeg, jpg, gif, png, zip, rar, tar, 7z

RUBRIC

A: BUSINESS REPORT:

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
A summary of 1 real-world business report created from a given dataset is not provided.	The submission summarizes the 1 real-world business report created from the chosen dataset, but it is *either* incomplete or inaccurate.	The submission accurately and completely summarizes 1 real-world business report created from the chosen dataset.

A1: DATA DESCRIPTION:

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
A description of the data to be used in the report is not provided.	The submission describing the data to be used in the report is *either* inaccurate or incomplete.	The submission accurately and completely describes the data that will be used in the report.

A2: IDENTIFYING SPECIFIC TABLES:

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
A submission identifying 2 specific tables from the given dataset is not provided.	The submission does not identify 1 or *both* of the tables from the given data set, or 1 or *both* of the identified tables will not provide the data necessary for the detailed and summary sections of the report.	The submission identifies 2 tables from the given dataset that provide the data necessary for the detailed and summary sections of the report.

A3: IDENTIFYING SPECIFIC FIELDS:

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
A submission identifying specific fields that will be included in the detailed and summary sections of the report is not provided.	The submission does not identify 1 or more of the specific fields that will be included in the detailed and summary sections of the report.	The submission identifies *all* of the specific fields that will be included in the detailed and summary sections of the report.

A4: FIELD TRANSFORMATION:

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
A submission identifying 1 field in the detailed section of the report that will require a custom transformation is not provided.	The submission *either* identified 1 field in the summary section of the report or it did not explain why the field should be transformed.	The submission identifies 1 field in the detailed section of the report that will require a custom transformation and explains why it should be transformed.

A5: BUSINESS USES:

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
An explanation of the business uses of the summary and detailed sections of the report is not provided.	The explanation is missing *either* the explanation of the detailed section or the summary section. Or the explanation did not accurately and logically explain how the sections can be used to improve business.	The submission thoroughly explains the business uses of the summary and the detailed sections of your report and how these sections can be used to improve business.

A6: REPORT FRESHNESS:

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
An explanation of how frequently the report should be refreshed to remain relevant to stakeholders is not provided.	The explanation of how frequently the report should be refreshed has an unreasonable frequency, or the explanation is inadequate or illogical.	The explanation of how frequently the report should be refreshed to remain relevant to stakeholders has a reasonable frequency. The explanation is adequate and logical.

B:CREATING TABLES

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
The SQL code that creates the tables to hold the report sections is not provided.	The SQL code that creates the tables to hold the report sections is incorrect or incomplete.	The SQL code that creates the tables to hold the report sections is correct and complete.

C:SQL QUERY

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
The SQL query to extract raw data is not provided.	The SQL query provided does not correctly query the raw data needed for the Detailed section of the report from the source database, or the query does not verify the data's accuracy.	The SQL query provided correctly queries the raw data needed for the Detailed section of the report from the source database and verifies the data's accuracy.

D:FUNCTIONS

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
The code for the function(s) that perform the transformation(s) identified in part A4 is not provided.	The code for the function(s) that perform the transformation(s) identified in part A4 *either* did not result in the expected transformations, or 1 or more of the previously identified functions is not included in the submission.	The code for the function(s) that perform the transformation(s) identified in part A4 produced the expected transformations and *all* previously identified functions are included in the submission.

E:TRIGGERS

NOT EVIDENT

The SQL code that creates a trigger on the detailed table of the report is not provided.

APPROACHING COMPETENCE

The provided SQL code does not create a trigger or the trigger does not continually update the summary table as data is added to the detailed table.

COMPETENT

The SQL code successfully 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:STORED PROCEDURE

NOT EVIDENT

A stored procedure to refresh the data is not provided.

APPROACHING COMPETENCE

The stored procedure provided does not refresh the data in 1 or *both* of the detailed and summary tables, or it does not clear the contents of the detailed and summary tables, or it does not perform the ETL load process from part C. Or the submission does not include comments that identify how often the stored procedure should be executed.

COMPETENT

The stored procedure provided refreshes the data in *both* the detailed and summary tables, and it can clear the contents of the detailed and summary tables and perform the ETL load process from part C. The submission includes comments that identify how often the stored procedure should be executed.

F1: DATA FRESHNESS:

NOT EVIDENT

An explanation of how the stored procedure can be run on a schedule to ensure data freshness is not provided.

APPROACHING COMPETENCE

The explanation of how the stored procedure can be run on a schedule to ensure data freshness is inaccurate or incomplete.

COMPETENT

The explanation of how the stored procedure can be run on a schedule to ensure data freshness is accurate and complete.

G: PANOPTO VIDEO:

NOT EVIDENT

A Panopto video recording is not provided.

APPROACHING COMPETENCE

The Panopto video recording does not include a full demonstration of the functionality of the code used for the analysis, or it does not include an accurate

COMPETENT

The Panopto video recording includes a full demonstration of the functionality of the code used for the analysis and an accurate and complete summary of the programming environment.

and complete summary of the programming environment.

H:WEB SOURCES

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
A record of the web sources used to acquire data or segments of third-party code to support the application is not provided.	The record the web sources used to acquire data or segments of third-party code to support the application is incomplete or inaccurate. Or the web sources cited are not reliable.	The record the web sources used to acquire data or segments of third-party code to support the application is both complete and accurate, and the web sources cited are reliable.Or the candidate did not use *any* web sources to acquire data or segments of third-party code and stated this in their submission.

I:SOURCES

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
The submission does not include both in-text citations and a reference list for sources that are quoted, paraphrased, or summarized.	The submission includes in-text citations for sources that are quoted, paraphrased, or summarized and a reference list; however, the citations or reference list is incomplete or inaccurate.	The submission includes in-text citations for sources that are properly quoted, paraphrased, or summarized and a reference list that accurately identifies the author, date, title, and source location as available.

J: PROFESSIONAL COMMUNICATION:

NOT EVIDENT	APPROACHING COMPETENCE	COMPETENT
Content is unstructured, is disjointed, or contains pervasive errors in mechanics, usage, or grammar. Vocabulary or tone is unprofessional or distracts from the topic.	Content is poorly organized, is difficult to follow, or contains errors in mechanics, usage, or grammar that cause confusion. Terminology is misused or ineffective.	Content reflects attention to detail, is organized, and focuses on the main ideas as prescribed in the task or chosen by the candidate. Terminology is pertinent, is used correctly, and effectively conveys the intended meaning. Mechanics, usage, and grammar promote accurate interpretation and understanding.

WEB LINKS

Labs on Demand Assessment Environment and DVD Database

Panopto Access

Sign in using the "WGU" option. If prompted, log in with your WGU student portal credentials, which should forward you to Panopto's website. If you have any problems accessing Panopto, please contact Assessment Services at assessmentservices@wgu.edu. It may take up to two business days to receive your WGU Panopto recording permissions once you have begun the course.

Panopto How-To Videos