Designing and Implementing a Data Warehouse

Background

Galleria Holdings is a fast-food chain which originated in Italy but has now acquired a number of businesses in the UK. The company believes that the UK menus are too large and inconsistent across the outlets they have acquired. They also believe that the menus should be consolidated to suit the preferences of the customers and therefore reduce unnecessary purchases of food which is not chosen by customers.

In order to achieve this, the parent company wishes to create a data warehouse for Business Intelligence purposes.

They wish to analyse sales to determine the most popular menu items and also those which produce the most revenue. They also wish to find out the most popular product groups and those which provide the most revenue. Finally, they wish to construct a “league table” of the total weekly and monthly sales of the various outlets so that their sales performance can be monitored easily.

**The project brief is as follows:**

* Review the case study and identify the business requirements from a reporting perspective. Your overall objective is to design a data warehouse capable of reporting on the questions the business wish to answer.
* Perform an assessment of the initial dataset to:
  + Understand the data items and how they relate together.
  + Identify any data quality issues which could affect your implementation.
* Design the schema to satisfy the requirements.
* Implement the ETL to create and populate the tables in PowerBI ensuring all counts reconcile.
* Create a set of visualisations to meet the reporting requirements contained in the brief.

**Detailed requirements**

In its simplest terms, your task can be broken down into six parts which should be evidenced as a report that includes screen shots and commentary.

Part 1: Problem Domain Understanding

**Task 1**

Given the nature of the business you have been asked to conduct a comparative analysis of on-premises, cloud-based, and hybrid data engineering solutions e.g., ETL and ELT.

The evaluation criteria should consider the following:

* Security
* Compliance
* Scalability
* Efficiency
* Reliability
* Fidelity
* Flexibility
* Portability

**Task 2**

Review the case study and identify the business requirements from a reporting perspective and identify additional relevant business questions for reporting purposes.

Part 2: Data Understanding

**Task 1**

Perform an assessment of the initial dataset to:

* Understand the data items and how they relate together.
* Identify any data quality issues which could affect your implementation.

Part 3: OLAP Schema Design

**Task 1**

Identify and justify your OLAP schema design.

**Task 2**

Design and implement the schema in PowerBI to satisfy the requirements.

Part 4: Data deliverables

**Task 1**

Produce a document (Source to Target Mapping) that contains the mapping of source system fields to the fields of the target system.

**Task 2**

Create the tables in PowerBI.

**Task 3**

Extract Transform and Load the data using PowerBI.

Part 5: Test approaches

**Task 1**

Perform some tests to check the four different levels of data testing.

**Entity Level:**

Review the reporting requirements that were listed as part of the initial data warehouse project brief. Can your design support the requirements?

**Record Level:**

Run some queries to get the counts for each of your populated tables.

Look back at the original source data and analyse what each entity count should be.

Do your counts align? If not, consider what needs to be fixed.

**Column Level:**

Check your table structures against the original source data. Have you accounted for all the data items?

**Column Value Level:**

Run some queries to get frequency counts of your columns. As a minimum, pick one column per table and do a frequency count of that column.

Do the same in your source data. Do the values match?

Part 6: Measures and Visualisations

**Task 1**

Create visualisations to meet the original reporting requirements. Include any additional visualisations to support the extra business questions you identified in Part 1.

**Task 2**

Identify and create measures in Power BI for non-additive and semi-additive facts, so they can be used in visualisations and reports.

Marking Scheme

The skills evaluated within this project are described within the SFIA 8 framework; please see https://sfia-online.org/en/framework for further information.

The skills which this project will evaluate are the following:

* Data Engineering
  + Evaluating, designing, and implementing on-premises, cloud-based, and hybrid data engineering solutions.
  + Structuring and storing data for uses including — but not limited to — analytics, machine learning, data mining, and sharing with applications and organisations.
  + Integrating, consolidating, and cleansing data.
  + Building in security, compliance, scalability, efficiency, reliability, fidelity, flexibility, and portability.
* Data Visualisation
  + Presenting findings and data insights in creative ways to facilitate the understanding of data across a range of technical and non-technical audiences.

Submissions

Completed project report should be emailed to: [AcademyQAC@qa.com](mailto:AcademyQAC@qa.com).

The email should be titled “DFEDATA2 Final Project”.