Data Modeling for the Modern Warehouse — Fundamentals of Dimensional Modeling

Project 1 Instructions

Introduction

Welcome to your first project! In this project, you'll work through a prompt from a real analytics engineering job interview. The goal for the project is for you to apply what you've learned so far about data modeling to a realistic business problem.

There is not a single "correct" way to complete the project; the prompt is intentionally open-ended, allowing you to make choices and assumptions about the structure of the data and reporting objectives. If you already have some data modeling experience, I encourage you to widen the scope of the project by making assumptions that increase complexity! However, if you are new to data modeling, I encourage you to keep the project scope narrow.

Project Detail

Prompt

Imagine that we are helping a grocery store, VeggieMart, model their data for reporting purposes. The leadership team at VeggieMart is particularly interested in gaining insights relating to the following items:

- Stores

- Regions
- Managers
- Employees
- Transactions

Assume the following constraints:

- All managers are employees, but not all employees are managers.
- Each store has only one manager.
- Employees have a unique role at a store of either "manager" or "store worker".
- A manager can lead multiple stores, but store workers can only work at one store.
- Transactions can have multiple items.
- Transactions occur only in USD.

Directions

Using Google Sheets, create an Enterprise Bus Matrix for business processes and dimensions related to the project prompt. Within the same worksheet, below the bus matrix, create table documentation for the fact and dimension tables that relate to the project prompt. As you build out these resources, consider the questions you want may want this data model to answer, for example:

- How many stores exist in each region?
- Which store had the highest revenue last year?
- What store(s) is a manger currently assigned to?
- How many employees work at each store?
- Which employees are working over 40 hours per week?
- Which stores have the highest average transactions amount?
- What percent of transactions included at least one non-food item?

When a store manager goes from managing one store to two stores, how is each store's revenue impacted?

Anything else that seems relevant or interesting.

Try to include at least 4 business processes that a grocery store would engage in and at least 5 dimensions that would be relevant in analysis.

Bonus #1

Write some pseudo SQL queries based off the tables you documented that would give valuable insights to the VeggieMart team.

Bonus #2

Within Google Sheets, create a table that corresponds to each dimension and fact you documented, and populate those tables with 5-10 records of sample data. What are the possible values that could be populated? Can a value be null? How would that impact how the data should be queried?

Bonus #3

Imagine that you are an Analytics Engineer at YouTube who deals with data related to video playback. What are the types of facts and dimensions you would be dealing with on a day-to-day basis? The data generated by YouTube is probably a lot different than the data generated by VeggieMart! Create an Enterprise Bus Matrix for YouTube video playback with a few facts and dimensions.

Submission Instructions

IMPORTANT SUBMISSION INSTRUCTIONS

Before submitting your project, please set the sharing permissions such that anyone with the link can **comment** on the sheet.