Designing Dimensional Models for Data Warehouses

<u>Each of the three parts of the assignment is worth **1 point**.</u>

Part 1: The simplified Inventory Data Mart supports analysis and reporting on inventory transactions along multiple dimensions, from dates and products to warehouses and stores, throughout the last several years. To support these functions, the data mart contains data about theses dimensions, as well as over ten thousand inventory reordering transactions in a single fact table.

If you have Microsoft Access on your computer, open `AccDB9_Inventory_Data_Mart.accdb` database in `Assign` subfolder of `09_Dim_Model` folder. Under **Database Tools** click on the **Relationship** button to see the dimensional model's star-schema with the four dimensions and one fact table. Otherwise, open `Assign9_Inventory_Data_Mart.xlsx` file, where the data on each of the data mart dimensions, as well as the inventory fact table, are stored in separate sheets. Also, please review the date, product, store and warehouse dimensions, as well as inventory facts metadata described in different sheets in `Assign9_Inventory_Meta_Data.xlsx` file.

To learn about the product data, examine the data in **product dimension** sheet, counting the number of products and average weight by product family, department, category and brand. Using Excel's pivot tables is the simplest way to go, but feel free to do it any other way that will provide adequate summaries. All of the results need to end up in `Assign9_Inventory_Meta_Data.xlsx` Excel file, **product_meta** sheet. Continue by doing the same for the **store dimension**. Counting the number of stores by store type, average square footage for the whole store and various sections could be a way to start. Also, looking at where the stores are located, how old are they and how long has it been since last remodel should provide some insight into the store data as well. All the results need to end up in the **store_meta** sheet in the same Excel file. Quickly review the **warehouse dimension**, which seems to indicate the warehouses are right next to the stores. Finish by examining the **inventory facts** table, consisting of the keys to each of the 4 dimension tables, as well as the 6 fact measures: units reordered, units shipped, warehouse sales dollars and cost, supply time and store invoice dollar amounts. The **inventory_meta** gives you the ranges for those relatively small amounts. Find the averages and medians for each of the measurement facts. Which way are the distributions skewed, if at all? You must create histograms (frequency charts) for the warehouse sales, reordered quantities and supply time, three charts in all. All of this needs to end up in the **inventory_meta** sheet of the same Excel file.

Now that you have a good grasp on each of the 4 dimension tables (including the **date dimension**) and the inventory facts table, you need to use your **db_x500** database on the **SQL Server**, and use the SQL code in `MySQL9_Inventory_Data_Mart.sql` to create all the tables and relationships forming a star-schema dimensional model (use BIT instead of BOOLEAN data type). Import all 4 dimensions and the inventory facts data using **Import Data** (right-click db_x500 and choose Tasks) directly from `Assign9_Inventory_Data_Mart.xlsx` file. Use **Data Source Microsoft Excel version 2007**, sheet names ending with **$** as **sources**, but make sure to select correct **destination** tables from the dropdowns. You are going to have to stumble your way through the import, which is not going to necessarily be as smooth as we would like. After you finish with importing, proceed with `MySQL9_Explore_Inventory_DM.sql` file, in which you need to design three (3) queries that would be representative of the actual warehouse use. Suppose you want to better understand the difference between units ordered and units shipped by product category during 2028 for warehouses in the state of Washington. Next you want to examine the number of warehouse sales and average sales dollar volume by store country and state for food items broken by product department during 2028. For the third and last query examine total store invoices for beverages by store and month during 2028. Save the results in the `Assign9_Explore_Inventory_DM.xlsx` Excel file.

Part 2: The simplified Order Entry database was designed to support transactions related to customer purchases of computer and electronics products either over the phone or online. The following assumes you are familiar with this operational database from the first several assignments. Review the ERD from MySQL implementation shown below, as well as the data in `AccDB9_OrdEntry_DB.accdb` or `AccDB9_OrdEntry_Data.xlsx`. You want to build a data warehouse from the source data to answer questions such as: What are the top selling products by manufacturer? Who is the top salesperson by state? How do the sales vary by month? This case is similar to the Sporting Goods example from the presentation and videos on dimensional modeling. Kimball's chapters 3 and 6 should be useful as well.
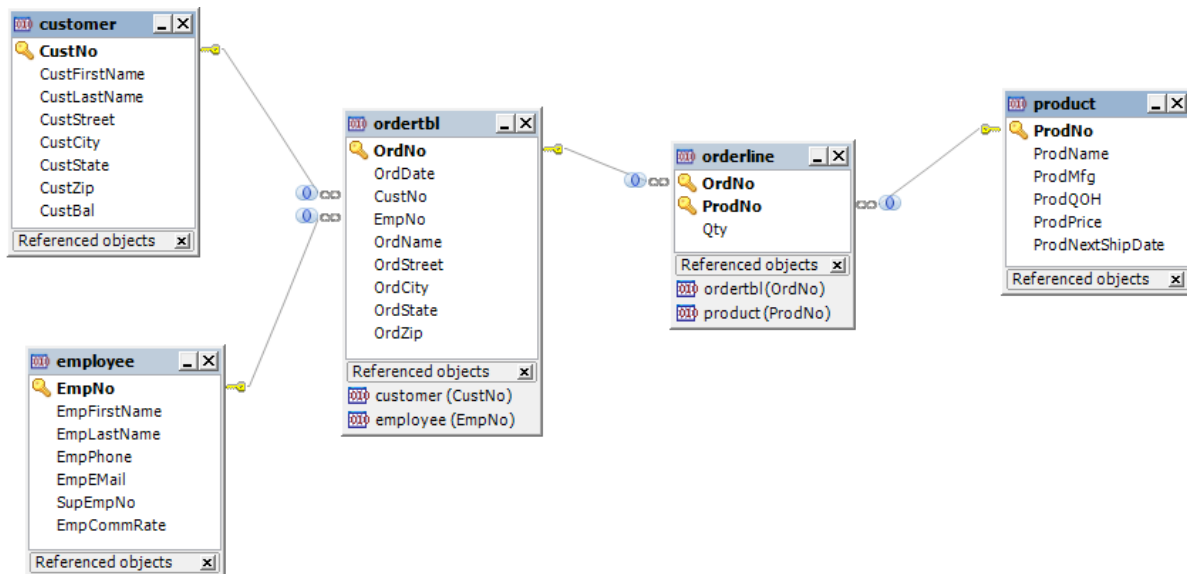
Identify the dimensions and their attributes, as well as the fact measurements for the order management dimensional model. Design it using ER assistant, and save it in `ERD9_OrdEntry_Data_Mart.erd` file. Make sure to identify all the primary keys for the dimension and fact tables.

Part 3: The last part of the assignment is more of an open ended attempt to define a data warehouse that could be useful in the real estate industry. The following assumes you are familiar with the operational ERD for the real estate database from the previous assignment, but please do not feel restricted by that particular operational system. In addition to the information available there, you can assume the MLS can provide you with the listing (asking) prices, listing, pending and sold dates, and any additional transactional info you may need. Also, individual buyer and seller info, such as SSN, names and street addresses should not end up in the warehouse. The warehouse should be able to answer questions such as: What is the number of houses that sold within 10% of the asking price in a given area? What was the average number of days until sold by a particular agent or office? What was the average asking price for 4 bedroom and 2 bathroom houses with 2000 to 2500 square feet by month?

Identify the dimensions and their attributes, as well as the fact measurements for the real estate dimensional model. Design it using ER assistant, and save it in `ERD9_Real_Estate_Data_Mart.erd` file. Make sure to identify all the keys for the dimension and fact tables.

Submission: You must submit `Assign9_Inventory_Meta_Data.xlsx` Excel file, `MySQL9_Explore_Inventory_DM.sql` SQL and `Assign9_Explore_Inventory_DM.xlsx` Excel files, and two ERD files: `ERD9_OrdEntry_Data_Mart.erd` and `ERD9_Real_Estate_Data_Mart.erd` on Canvas by the designated due date.