4.15 Course Project: Dimensional Data Model ERD- Commentary on Design

Lizzy Schmitz

Department of Information Management, University of Washington

IMT 577: Business Intelligence Systems

Professor Janak Rajani

July 17th, 2022

**Overall Design:**

Based on the key measurements and data points the business needs to analyze to help answer questions and make decisions, I decided to structure the model (database) by creating two fact tables focused on 1) location and 2) customer-centric metrics, with product information withing both tables as both locations and customers are tied to products.

Below outlines the key measurements that will be presented in each fact table:

|  |  |
| --- | --- |
| Fact Table | Key Measurements Answered |
| Location Performance | Location/Product Sales, Performance to Targets, & Profitability (by Store/Reseller, Channel, Date, Product, and Overall) |
| Customer | Customer Purchase Trends(by Store/Reseller, Channel, Date, Product, Reseller, Subsegment, Overall) |

I then leveraged existing CSVs/normalized tables as supporting DIM tables and kept all critical columns from those tables that would help provide metrics needed for existing business questions and future proof for any additional business questions that might be asked. Below outlines what each record (grain) captures within each dimension table:

|  |  |
| --- | --- |
| Dimension Table | What each Record (Grain) Captures |
| Channel | Each grain contains one unique channel type and whether that is considered Direct/Indirect (based on channel category) |
| Channel Category | Each grain contains one unique type of channel classification used for stores/resellers |
| Customer | Each grain contains one unique customer and their identification details (customer ID, name, address, contact information, etc.) |
| Product | Each grain contains one unique product, and the associated description |
| Product Category | Each grain contains one unique category of product used to classify/group the products |
| Product Type | Each grain contains one unique product type and what product category it's classified under |
| Reseller | Each grain contains one unique reseller and identification details (reseller ID, name, address, contact information, etc.) |
| Sales Detail | Each grain contains one unique sale and the quantity and product that was purchased, and the sales amount |
| Sales Header | Each grain contains one unique transaction and the date it was purchased, and in which channel/store/reseller |
| Segment | Each grain contains one unique type of segment |
| SubSegment | Each grain contains one unique type of subsegment |
| Store | Each grain contains one unique store (location) and store details (store number, manager, location, contact information) |
| Target Data Channel Reseller | Each grain contains one unique target sales amount by year for each store/ reseller |
| Target Data Product | Each grain contains one unique sales quantity target by product, by year |

Additionally, this model leverages natural keys (the assumption that all data in CSVs was created in the source system) as many of the columns had unique identifiers that passed from the source system. Only surrogate keys were needed for target tables as those didn't have established unique identifiers from source systems (Natural keys).