## Project Description

To

Mary Sweeney

From

Whitney King

RE

Assignment 4

Course

BUSIT202 – Dimensional Modeling

Difficulties

Originally, I started down the road of a transactional sales table, but realized that was the wrong direction for what the business was asking me to accomplish, so I rethought what was being asked and moved to a periodic snapshot version instead, which I believe to be the correct way to do this project.

At first glance I’d missed the main functional different in Type 2 and Type 3 structures, but once clearing that up was able to move forward with doing this assignment. Overall, I believe this makes sense and that it is appropriate to change the type of change handling for each individual attribute.

In this project, we were requested to update our dimensional model in order to accommodate slowly changing dimensions. In doing so, we were also asked to the business with information about product sales by category, allowing the organization to look at how yearly product re-categorizations affected total product sales. Additionally, we were to provide sample data for each table to show examples of the SCD for each attribute, and attribute types for each field.

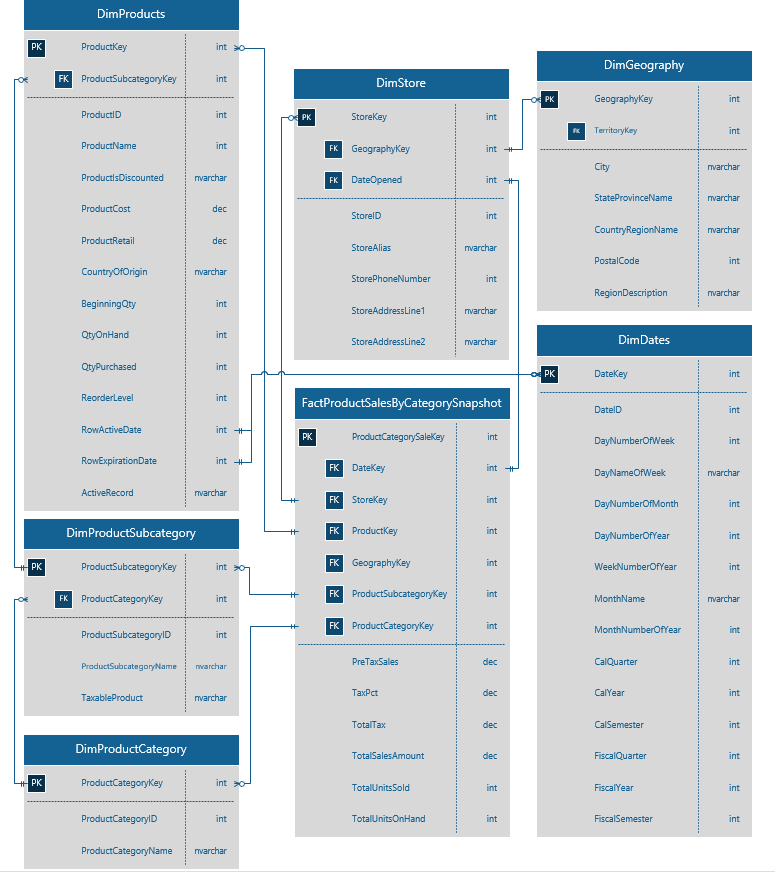
## Overview

For this model, we were asked to provide a slowly changing dimensional structure that would suit the business need of looking at how different products have performed in various categories each year. It is expected that products will shuffle categories once per year at the beginning of each calendar year. The business group has advised us that they do not care to examine current year sales by prior year category, which tells us that **we *are not* considering the Type 3 (Add Attribute) structure** for comparative attribute data. However, if we were to compare year over year sales of a single SKU to the sales it had in the previous year and category, we would still want to retain a historical record of the categories and date ranges that have impacted each SKU. This makes it apparent that we **should be looking into a Type 2 (Add Row) model for category changes** that will be occurring each year. The **primary key in each dimension will be a Type 0 attribute** – this number can never change.

The functionality that the business is currently requesting is pretty standard operationally (and our database is still relatively small), so it doesn’t make sense at this point in time to focus on a dimensional structures with all the bells and whistles. However, **we may find ourselves using several types of attributes (ensuring the categorical data is Type 2, and judging the other attributes on a case by case basis) to take a bit of a hybrid approach**. Any changes made to this dimensional structure were made with previously designed fact tables in mind, my endeavor is to add features, not take those things away or cripple other business processes.

I chose to include certain extra dimensions that seemed like they would bring valuable insights to the business when attempting to analyze product sales by category and year. I’ve included the geography and store dimensions to give users extra flexibility and visibility into if products are also selling better in certain locations at certain times, or if some categories do better in certain locations.

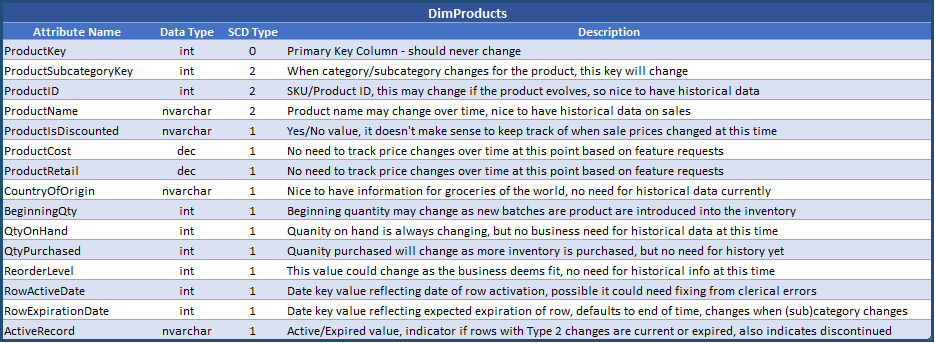
## Dimensional Model



### DimProducts

The products dimension was the most complicated one to put together properly, because I really needed to think about the types of slowly changing dimensions to ensure the structure made sense and was useful to the business. I decided to treat this dimension sort of like a Type 4 (but not quite), and splinter off a couple of smaller dimensions in order to help manage the data and growth of the business. Following some of the information learned about this type of structure, the **DimProductSubcategory** and **DimProductCategory** dimensions were born to help wrangle some of the information that would otherwise clutter up the **DimProducts** table.

Additionally, columns that are important to answering questions the business is asking **in relation to product category year by year are being treated at Type 2 attributes** (such as the **ProductSubcategoryKey**, which will change, and the **ProductID** and **ProductName** which may change) due to the necessity to retain information about past categories for each product (if we want accuracy). To facilitate this, three new (Type 1) columns were added to the **DimProducts** table: **RowActiveDate**, **RowExpirationDate** and **ActiveRecord**. These columns will provide the business information about the status of each row, and enable some of the requested reporting features with dates. **Everything else is being treated as a Type 1 attribute** since the business need is not currently there to treat them otherwise. **CountryOfOrigin** was also added, and could easily be taken from the geography table, but this is nice to have and was overly complicated without being requested as a feature.



#### Sample Rows

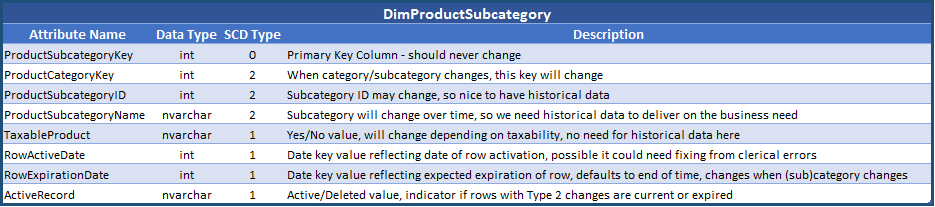
With a row expiration date like 12/31/9999, the date table would possibly get quite large unless we were to restructure it some with some Type 4 mini-dimensions. However, that’s something to think about for the future and I am leaving it out of this current model.



### DimProductSubcategory

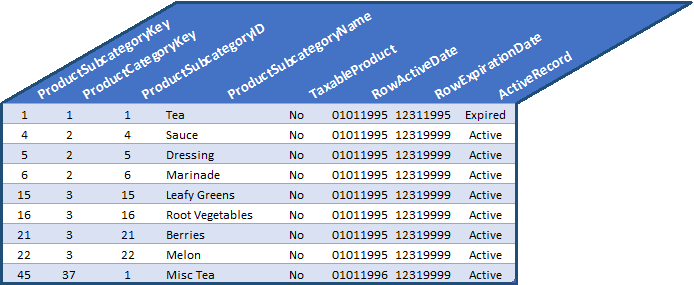
With a business that shuffles product categories as frequently as every year, it didn’t seem to do the dimensional model justice to not break categories into subcategories. When it comes down to it, a beverage is a beverage. You can exercise some thesaurical word play to come up with alternative names for an attribute (such as calling them drink instead), but when it comes to taxonomy, sometimes something simply fits best in its home category (and most of the time it comes down to what the business wants). It is however, generally much easier to diversify subcategorization as you get to a more and more granular view of information.

It’s the above factors that led me to create **not only one new dimension, but two: DimProductSubcategory and DimProductCategory.** Really, it just **makes sense to have subcategories since there are multiple types of products in each category**, but I do *not* believe these new dimensions meet the requirements for Type 4 mini-dimensions due to the lack of bands.



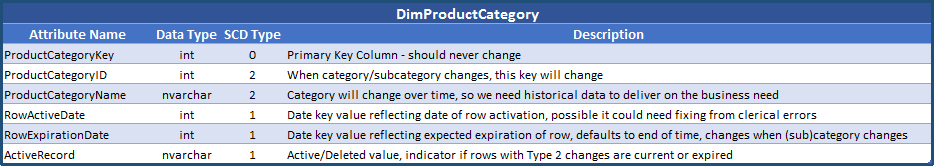
#### Sample Rows

In the event a **DimProductCategory** row expired, its associated **DimProductSubcategory** row will also expire and produce a new active row containing the new **ProductCategoryKey** for the product.



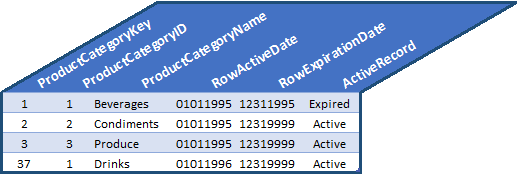
### DimProductCategory

Category is similar to subcategory, but it is one level higher in granularity than subcategory. For this reason, **DimProductCategory** uses **DimProductSubcategory** as a bridge to connect to **DimProducts**. To put it in biology terms, category would be the family, subcategory would be the genus, and product name would be the species of any given product. These dimensions could potentially cause some snowflaking here depending on how fact tables are designed, but I believe the tradeoff is worth this design decision.



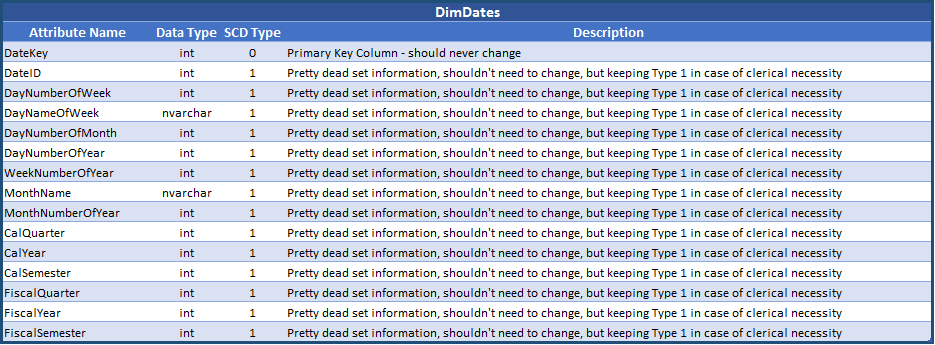
#### Sample Rows

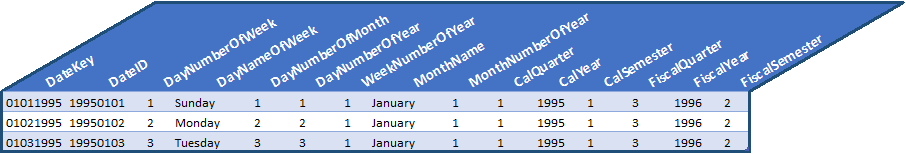
These sample rows follow much of the same format as **DimProductSubcategory**, just with less columns due to the higher level overview.



### DimDates

This dimension **appears to be a Type 1 dimension, even possibly a Type 0 dimension in certain data warehouses**. As long as the date structure were setup properly, there wouldn’t be much of any need that I can think of to change this dimension outside of clerical errors or formatting changes. For that reason, for now I am going **to treat every attribute except the primary key as Type 1**. As time progresses, more and more rows are added to compensate for new dates, and old ones shouldn’t need to change.





### DimStores

This dimension is unchanged from when it was last provided so no attribute table or sample rows are provided. However, I would consider this a Type 1 for now, until further business need determines otherwise.

### DimGeography

This dimension is unchanged from when it was last provided so no attribute table or sample rows are provided. However, I would consider this a Type 1 for now, until further business need determines otherwise.

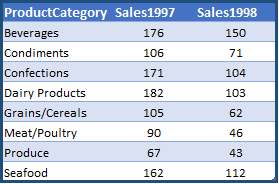
### FactProductSalesByCategorySnapshot

The fact table I provided in the database diagram has a lot more information on it than is needed to generate the simple three column sample report requested below. I tried to design this with the future in mind, and also give it the ability to answer questions that management didn’t think to ask yet (such as category sales by location/store). Since it’s not required in this project, I have not included full sample rows from this fact table, just the requested sample data.

#### Sample Report

Based on the requests of the business for this particular project a sample report that might be produced from the data is included below. This is a nice high level overview of sales by year from current data in the database, and would be subject to change if categories change, or if you wanted to see it broken out by subcategory. Additionally, it would be worthwhile in the long run to look at some of this information on a geographic scale in order to get a better understanding of what items sell best in which regions.

I’ve provided a sample query below, and while it isn’t perfect (it has a lot of NULLs I’ve yet to learn how to clear out, so I relied on Excel to help me organize the information), it is a good starting point for the beginnings of the reports the organization will need.

Use DWNorthwindOrders

Select ProductCategory,

(CASE WHEN Year = '1997' THEN COUNT(ProductCategory) END)

AS Sales1997,

(CASE WHEN Year = '1998' THEN COUNT(ProductCategory) END)

AS Sales1998

FROM DimProducts AS p

JOIN FactOrders AS f

ON f.ProductKey = p.ProductKey

JOIN DimDates AS d

ON d.DateKey = f.OrderDateKey

GROUP BY ProductCategory, Year