MIS 636 Data Warehousing and Business Intelligence

Individual Accreditation Assignment

Balaji Katakam
CWID: 10423274
STEVENS INSTITUTE OF TECHNOLOGY

Problem Statement

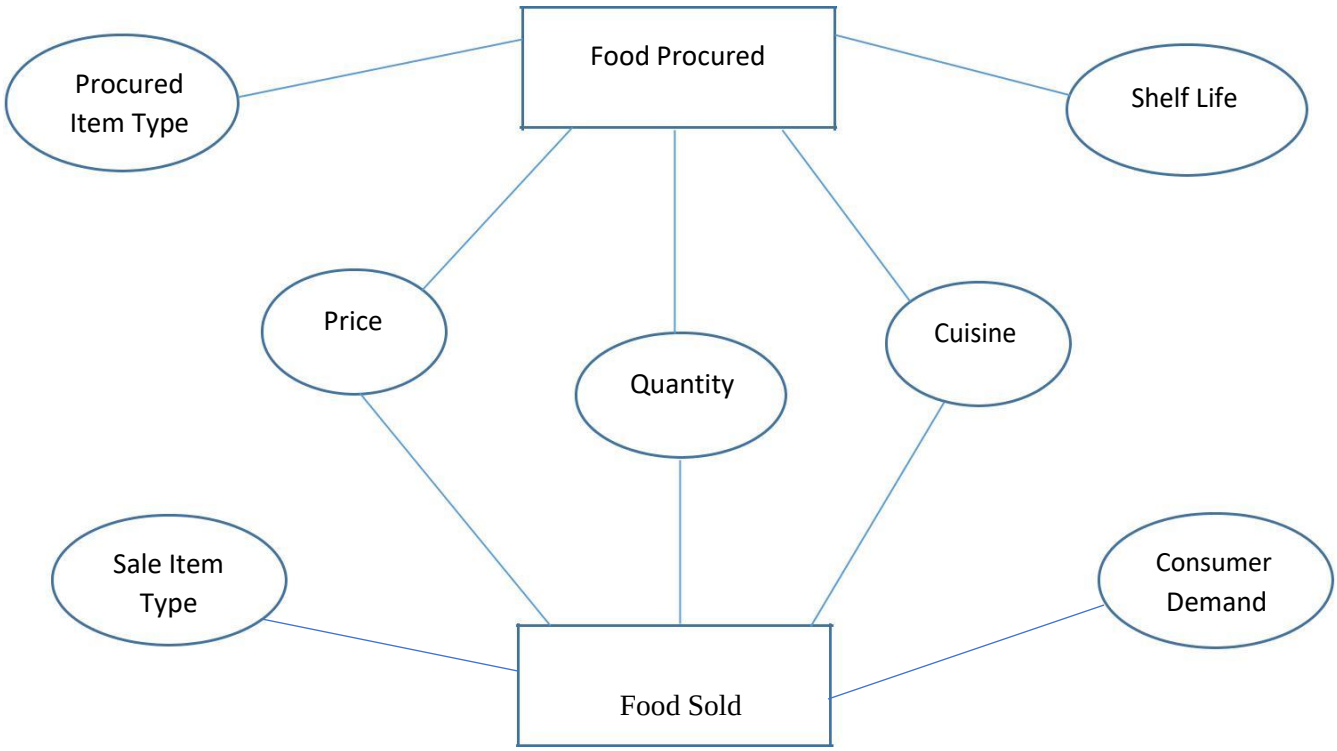
I have considered a Franchise restaurant as a possible case for my business process identification. This can also be called as Food Chain based restaurants. Restaurants like Subway, Chipotle, McDonalds are Food Chain restaurants. The need here is to build a multi-dimensional model to identify the amount of food procured by the restaurant and the amount of food sold, to decide if the restaurant is selling as much as it is buying or if it's not, it's either breaking even or if it's returning a profit.

Access to Data

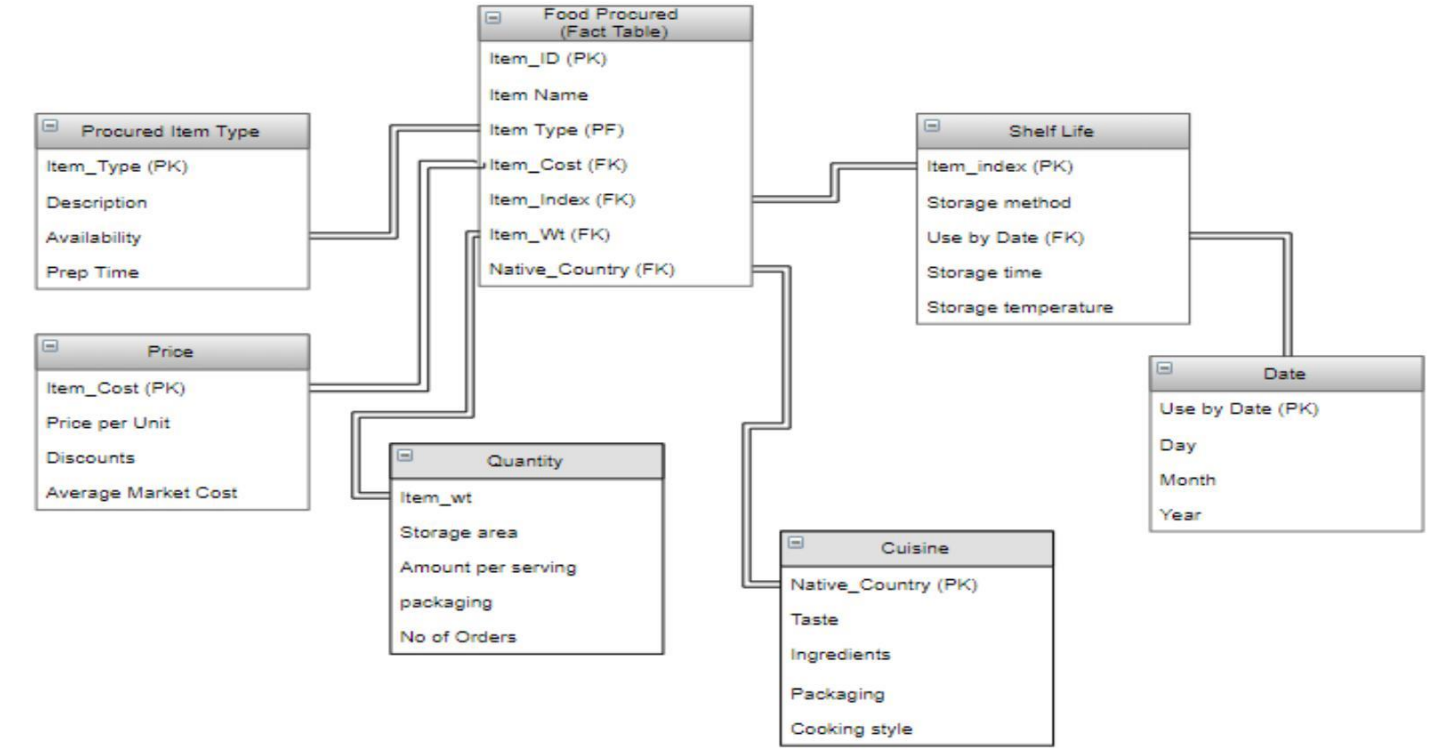
The data can be accessed from multiple available data sets online of Food Chain' retail businesses and other case studies made in the same field. Obtaining physical data like real life user cases would be difficult so one would have to stick with the virtual data sets.

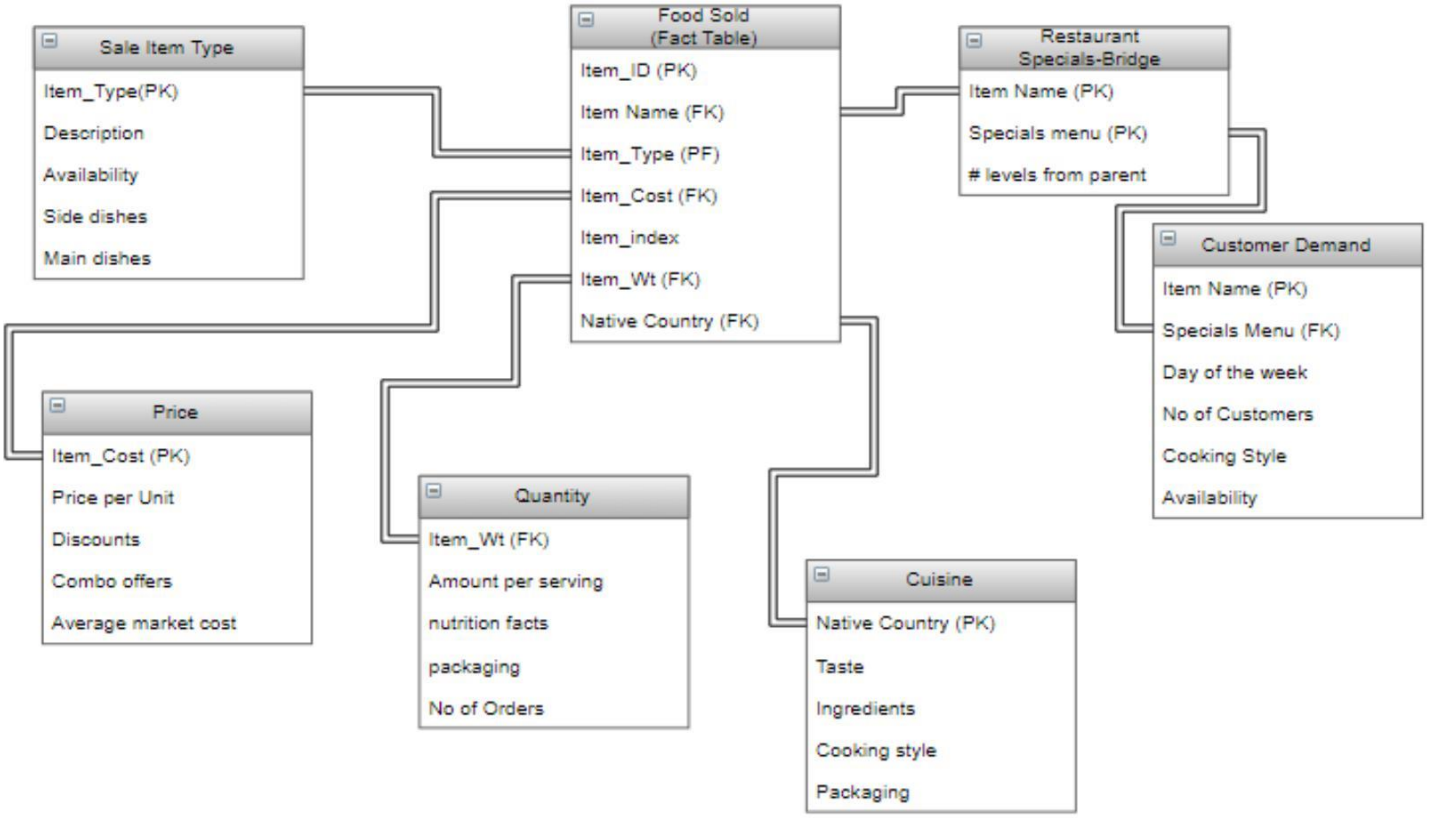
Dimensional Model Design

High Level Diagram



Detailed Fact Table Diagram



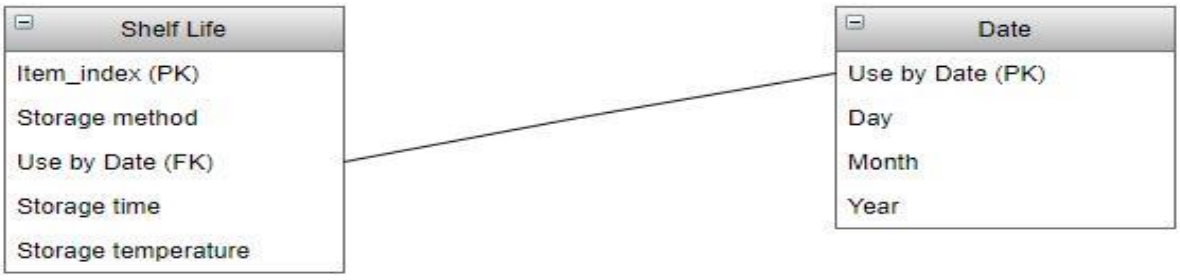


Facts:

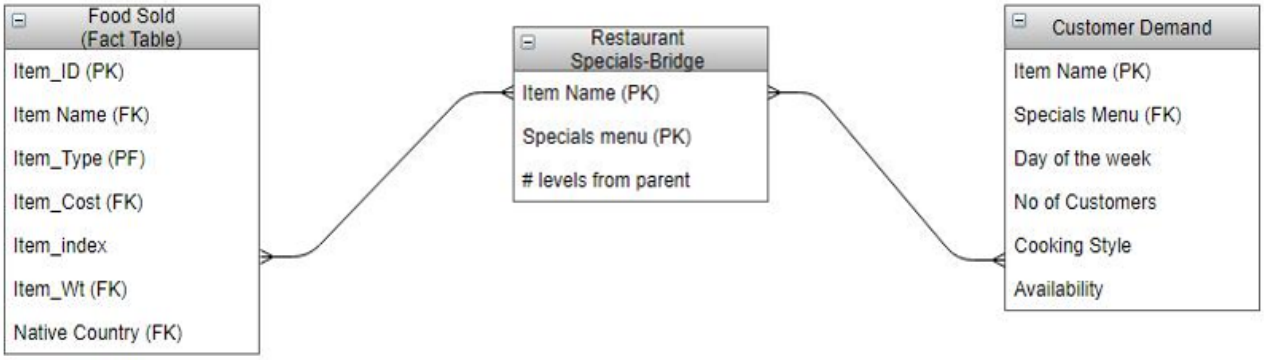
- There are two fact tables in this model, one for the *Food Procured*, and one for the *Food Sold*.
- In the fact tables the foreign keys are taken as composite primary keys so that there is data available on both sides. Ex: Item Name, Item_Cost, Item_Wt, Native Country.
- The tables have both numerical attributes, like price and quantity, and text attributes like item type and cuisine.

Dimensions:

- The dimension tables for the Food Procured table are Procured Item type, Price, Quantity, Cuisine and Shelf Life.
- The dimension tables for the Food Sold table are Sale Item type, Price, Quantity, Cuisine and Customer Demand.
- The tables Price, Quantity and Cuisine can be seen as conformed dimensions as they are used commonly for both tables.
- All the item related attributes are the role changing dimension here. We can see different columns like item_cost, item_wt, item name, item_index, etc., which we can pool in to one common table to update the related items in these tables. This becomes useful when we expand the restaurant table or bring in a larger data set as there is change in only one table required.
- The Date table added to the Food Procured fact table dimension can be seen as an *Outrigger*. This is because we do not have an active dimension related to date in our Fact table model so we would reference the date table externally for the shelf life dimension when we use it.



- When adding the Customer Demands dimension I realized that I would have to include a bridge table in order to correlate the Food sold Fact table and the Customer dimension. The bridge table here is the Specials table which would have a many to many relationship in terms of Food and Customer demands.



Grain:

The fundamental grain of this table can be taken as the *item ID* and the *item type*. Based on a food item's ID and type we can extrapolate the information to fit in to any of the other tables. The basic unit based upon which the rows will be built in the fact table are the item attributes item ID and item Type. Any one transaction will require one of these two columns as a necessary input. For example, based on item type we can determine the price of the item, the quantity of the food item and also its shelf life.

Based on the item name, we can pull out the type of cuisine it represents, and also where in the specials menu it might be present. The item ID and Item type provide the *atomicity* of the Dimensional model.

References:

https://archive.ics.uci.edu/ml/datasets/Restaurant+%26+consumer+data http://connect.factual.com/Restaurants.html
https://dreamtolearn.com/ryan/1001_datasets
https://www.kaggle.com/datasets