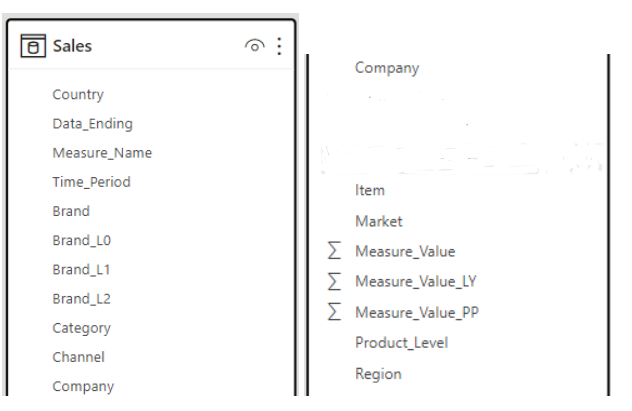
# Guidelines for integrated data model/schema

* ~~Author: Faizan Khurshid~~

## Introduction:

Integrated Schema is an approach towards designing a database structure in a way that efficiently and seamlessly consolidates data from multiple sources with varying granularity and hierarchies into one integrated database. The approach also focuses on reducing the load on downstream data processing layers by pre-calculating necessary measures at required granularities.

integrated approach as OBT(One big Table):



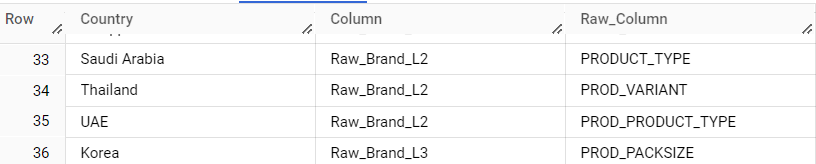
## Some Key aspects of Integrated Schema:

### **Generic Names for Product Hierarchy:**

In Integrated Schema, we utilize standardized naming conventions such as "Brand\_L0," "Brand\_L1," and "Brand\_L2" for Subbrand, Form, and Type. This simplifies the integration and interpretation of product-related data.

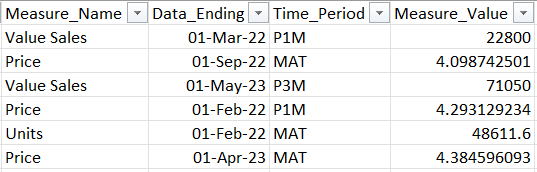


Moreover, an auxiliary metadata table can be maintained which can significantly enhance data comprehension. This table, specifically designed to encompass the detailed definitions and relationships between Brand\_L0 and Brand\_L1 etc columns across different country-categories, prevents the need for hardcoding such specifics into downstream applications.



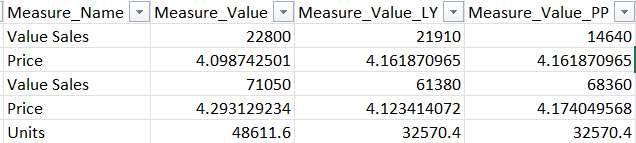
### **Precomputed Time Aggregations (PxMs):**

We calculate and store time-aggregated data, like P3M and P6M, within the fact table. This simplifies downstream calculations, ensuring readily available data for reporting and analysis without overloading the reporting layer.



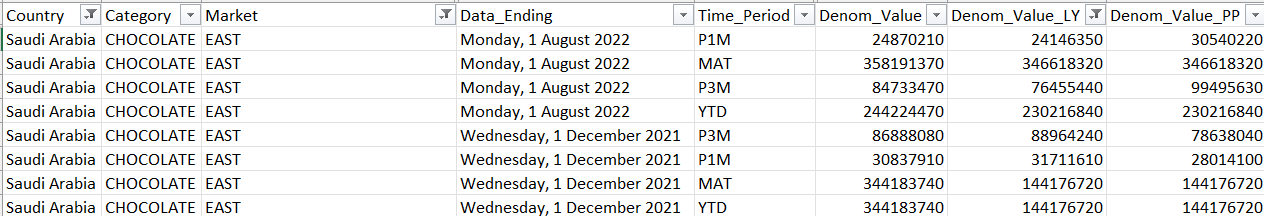
### **Measures in Rows Instead of Columns:**

We store measures and values in rows, not columns. This simplifies integration of data from different sources consisting of numerous measures by using "Measure\_Name" and "Measure\_Value" columns, ensuring a consistent format for all measures. Also if a comparison with an older time period (e.g. LY and PP) analysis is required, Measure\_Value\_LY and Measure\_Value\_PP columns are also added and used to store precomputed values against these time periods. This significantly reduces the delta calculation load in downstream applications.



### **Calculating Denominators:**

We pre-calculate denominator values for share calculation to reduce downstream computation. This eliminates the need for run-time calculations on the reporting layer.



There are two approaches to keep the denominator values in our schema:

1. Keeping the denominators inside the main fact table:  
   In this approach the denominator values are kept inside the fact table, the denominator gets repeated for each market combination in the data
2. Separate Denom Table:  
   The other approach is to create a separate denominator fact table which will act as an independent fact table.

## **Conclusion:**

Integrated Schema approach ensures data integration is seamless and efficient, and it optimizes the availability of essential measures and calculations, ultimately enhancing the overall data processing capabilities within our organization.