

# MSIS 2621 Project Report

Group 7

Yue Ma, Yining Ou, Yidong Mou, Aniketh Satyanarayana

## **Business Analysis of Walmart Retail Store Scenario**

Walmart wants to analyze how to order type and order frequency differ by region and the type of item ordered. The project focuses on developing a data warehouse utilizing the ETL tool Pentaho/Tableau PrepBuilder. ETL process performs extraction of data from available databases and loading into a data warehouse. This data warehouse will help the stakeholders to provide business solutions to improve the business processes involved in Walmart. We utilize Tableau tools to build dashboards and provide actionable insights to the stakeholders.

## **Some basis for the data analysis**

- Store: Sales for the given store
- Date: Sales by day of a week, month, quarter, etc.
- Holiday Flag: Whether the week is a special holiday
- Temperature: Temperature on the day of sale
- Fuel Price: Cost of fuel in the region
- CPI: What's the influence of prevailing consumer price index on sales
- Unemployment: How prevailing unemployment rate influence on sales
- Product category: Sales by different category
- Brand: Sales of same product with different brands

## **Steps involved:**

- Extract data from Walmart OLTP data sources.
- Loading required dimensions and fact tables in Data Warehouse.
- Fetching data from the External Sources database which has some useful datasets for data warehousing.
- Developing a TableauDashboard for mining actionable business insights.

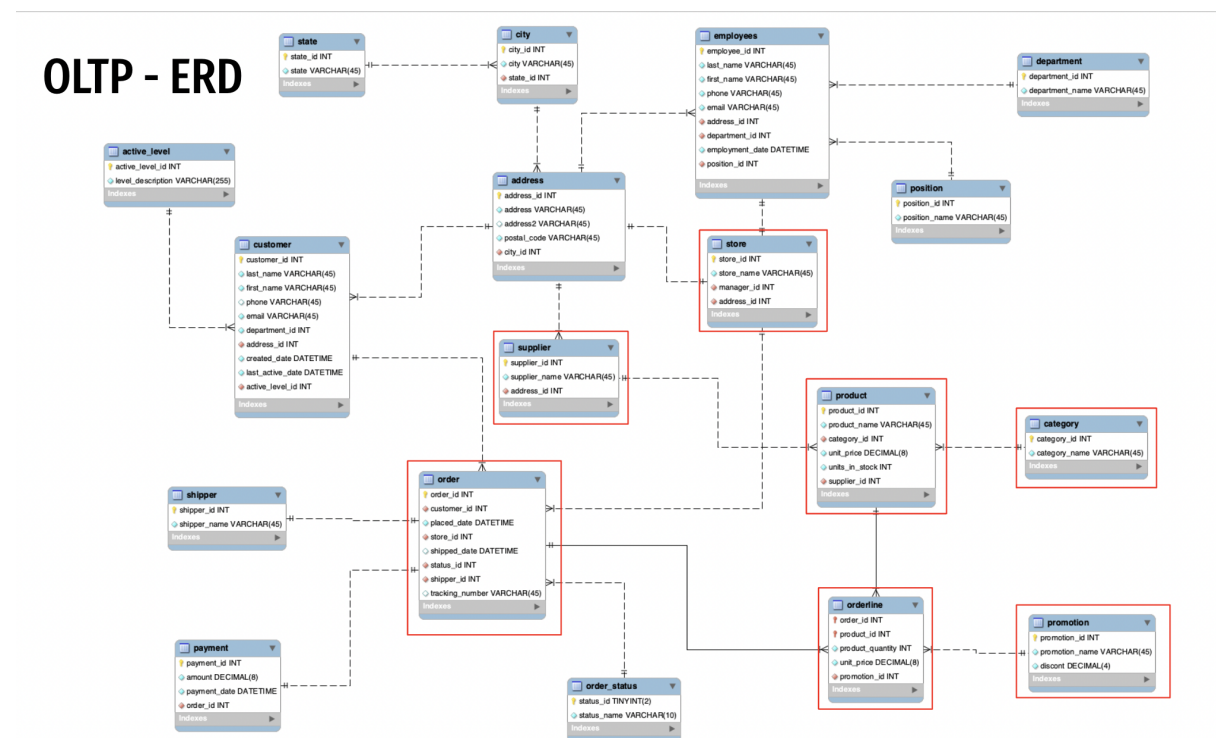
# Data Modeling

## Takeaways:

To figure out a reasonable structure of the schema, we learned about the process of typical online shopping transaction and the 3 Normalizations to ensure the efficiency of OLTP model.

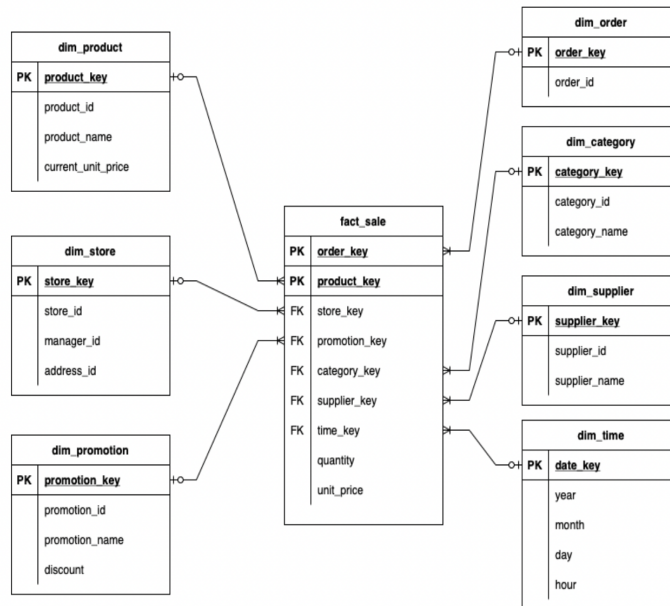
## Process:

Our OLTP schema includes all entities that are related to the transaction. We didn't find proper real data, and since we want to analyze sales grouped by suppliers, stores, products, categories, and promotions, which were highlighted with red boxes, we generated some random data for these tables.



Below is the Star schema that contains the data we use for analysis.

## WH - Star Schema



## ETL Implementation

### Takeaways:

- We used both Pentaho and Tableau Prep to do ETL but we chose Tableau in the end because its visual operation is very convenient. We also learned that though Pentaho has a bad use experience, but it is a professional open source software. It's more powerful compared to Tableau Prep when the datasets are huge. But for smaller datasets, Tableau Prep's convenience can save a lot of time.
- We also learned to create table with SQL scripts in Tableau Prep before we output the table so that we can have incremental key columns without first creating the table in MySQL.
- And there is a tip, we can directly remove useless columns when we load the table, so that we don't need to add a cleaning step to remove it. And it's a good habit to remove useless columns before next step, this can keep everything clear.
- The importance of a proper database transformation step is realised while carrying out the ETL process. Some errors were encountered along the way, where we had to make sure to modify the way the data cleaning transformation pipeline was built.

- We generated some timestamps with python.
- We included different formats and made sure that ultimately we should be having data in the form of SQL table format in the ODS database.
- Tableau Data Prep software helped us seamlessly merge the different sources of data together into a SQL format.

## OLTP to ODS

This is the process we transform OLTP data into ODS data, we get 7 ods tables and those are what the tables look like. Product, Order, Orderline, Promotion, store, supplier, and category.

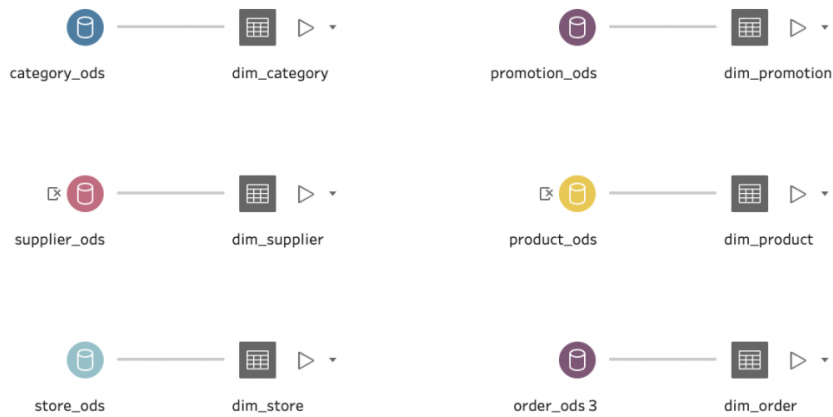
## ETL - OLTP to ODS



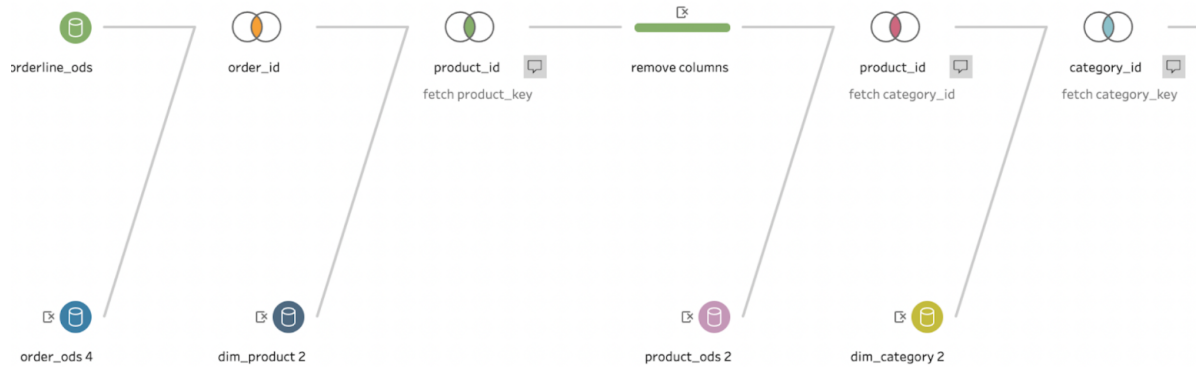
## ODS to WH

And then we transformed the ODS tables into WH tables. We got 7 dimension tables. They are category, supplier, store, promotion, product, and order. The last dimension table is the date dimension. We used the table exported from the pvfc warehouse schema.

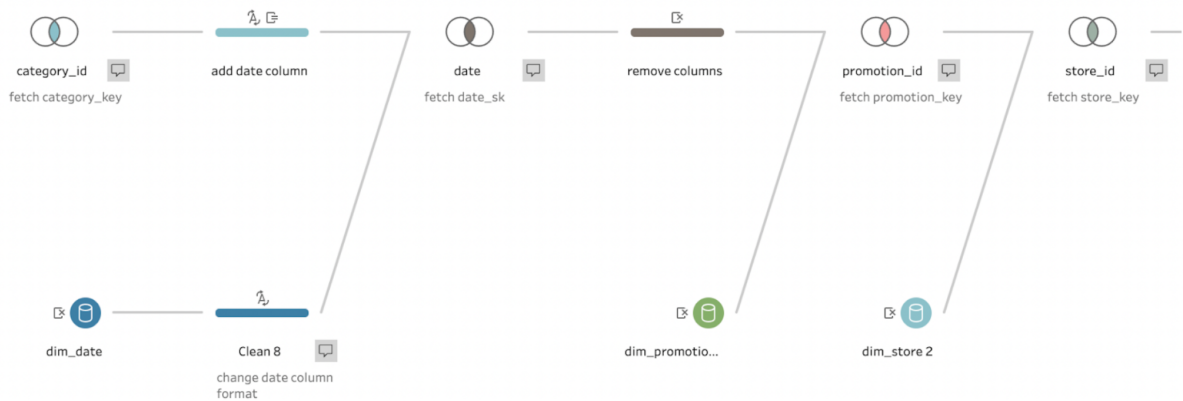
# ETL - ODS to WH



## ETL - ODS to WH



## ETL - ODS to WH



## **Tableau Implementation**

**Takeaways:** Dashboard is a powerful way to present data-based intelligence using data visualisation techniques. And a good dashboard can help the audience quickly catch the information at a glance. By taking this class, we first learned how to implement data visualization using Tableau. Second, We learned how to use the dashboard to integrate relevant, actionable data and track stats and key performance indicators. We also learned how to apply parameters and do forecasting and what-if scenario.

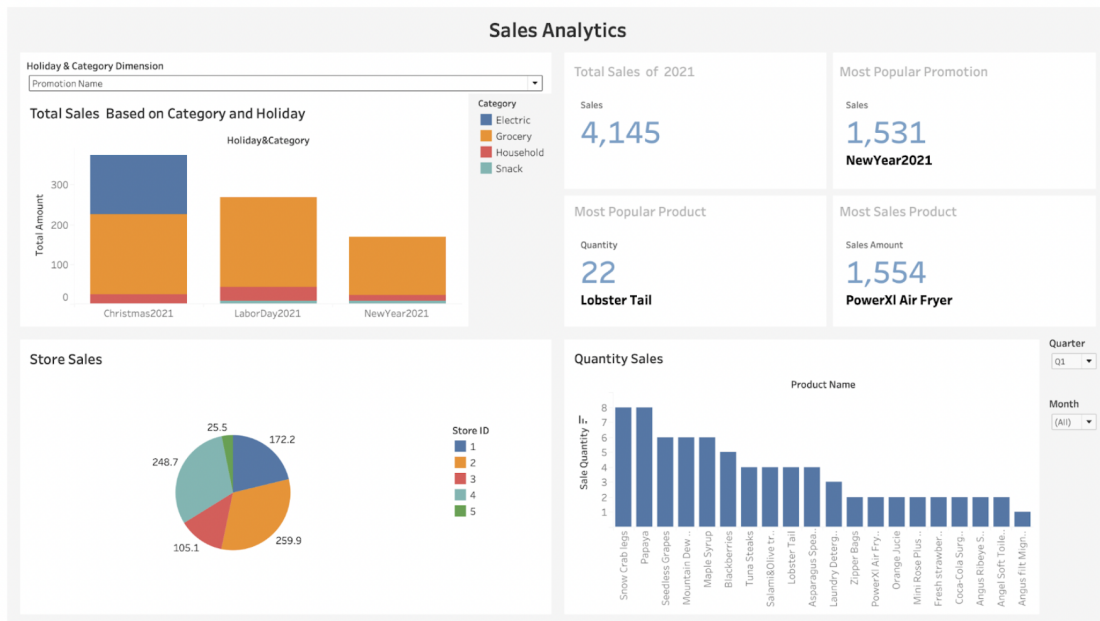
### **Descriptive Analytics**

**Purpose:** Descriptive analytics reveals unique insights, trends, and patterns of our sales data.

#### **Analytic Problem:**

- Which categories of goods provide the most sales
- Which categories of goods provide the least sales
- which product is the best-selling product
- Which promotion attracts more customers
- Products sales based on month/quarter
- Store sales based on month/quarter
- What type of product do people like to shopping during holiday seasons

# Descriptive Analytics Dashboard



## Forecast Analytics

### Takeaways:

After importing the data and getting a blank table, first, Creating a line chart of this data is easy. First, double-click on sales Trips in the left panel. Then double-click on month in the left panel. Tableau will create this line chart. Now we have a line chart showing the forecast of the total number of trips per month. The dataset starts in January 2011 and continues through 2022. We have daily data. Tableau recognizes that the Start Date attribute isn't just an ordinary qualitative or quantitative variable, it's a date. Tableau can aggregate attributes in all sorts of different ways using dates. Then we can make some forecasts for what the number of trips will look like in the future. Our dataset goes through January 2019. we use the chart we created with monthly data to have Tableau forecast some future days. First, click on the Analytics tab in the left panel. The one we want is "Forecast". Click and drag Forecast onto the line chart. We get this forecast result dashboard.

**Purpose:** Predict next-stage sales based on past data to make better business decisions.

### Analytic Method:

- Use Discount and Quantity as parameters.
- Aggregate parameters with their average, sum, min, and max.
- Multiply the parameters with price and quantity to create the new dashboard
- Analysis sales with months

## Forecast Dashboard

