

A retail company in the United States is looking for a supplier to build a system that provides information for specifically two areas of the company. It is looking for the information to be easily interpreted and to not consume many technological resources. To win the contract, I will present a proposal using the MySQL engine that includes the following:

- Load all the databases available in the company in MYSQL.
- Build a denormalized data warehouse using the ETL pipeline.
- Develop data marts providing information to each area.

1. Operational Layer

1.1. Databases

The company has three databases (Orders, Products and Customers). The first corresponds to the history of purchase orders between 2014 and 2017¹. The second includes the information of all the customers. The third includes information on available products. All files are in CSV format.

1.1.1 Orders

The table has 9.994 rows and 17 variables:

Variable Name	Type	Description
ID	INT	Primary key
Order_ID	Varchar	Unique code of purchase orders
Order_Date	Datetime	Date of order received
Ship_Date	Datetime	Date of Shipping the order
Customer_ID	Varchar	Unique code of a customer, foreign key
Ship_Mode	Varchar	Type of shipping
Country	Varchar	Country of destiny of the order
City	Varchar	City of destiny of the order
State	Varchar	State of destiny of the order
Postal_Code	INT	Postal Code
Region	Varchar	Region of the company that manages the purchase order
Product_ID	Varchar	Unique code of the product, foreign key
Sales	Double	Sales in US Dollars
Quantity	Double	Sales in units
Discount	Double	Percentage of the discount applied
Profit	Double	Profits of each order
Returnss	Varchar	Indicate if the order were returned by the customer

¹ Public resource from tableau plataform:
<https://community.tableau.com/s/question/0D54T00000CWeX8SAL/sample-superstore-sales-excelxls>

1.1.2. Products

The table has 1.862 Rows and 5 variables

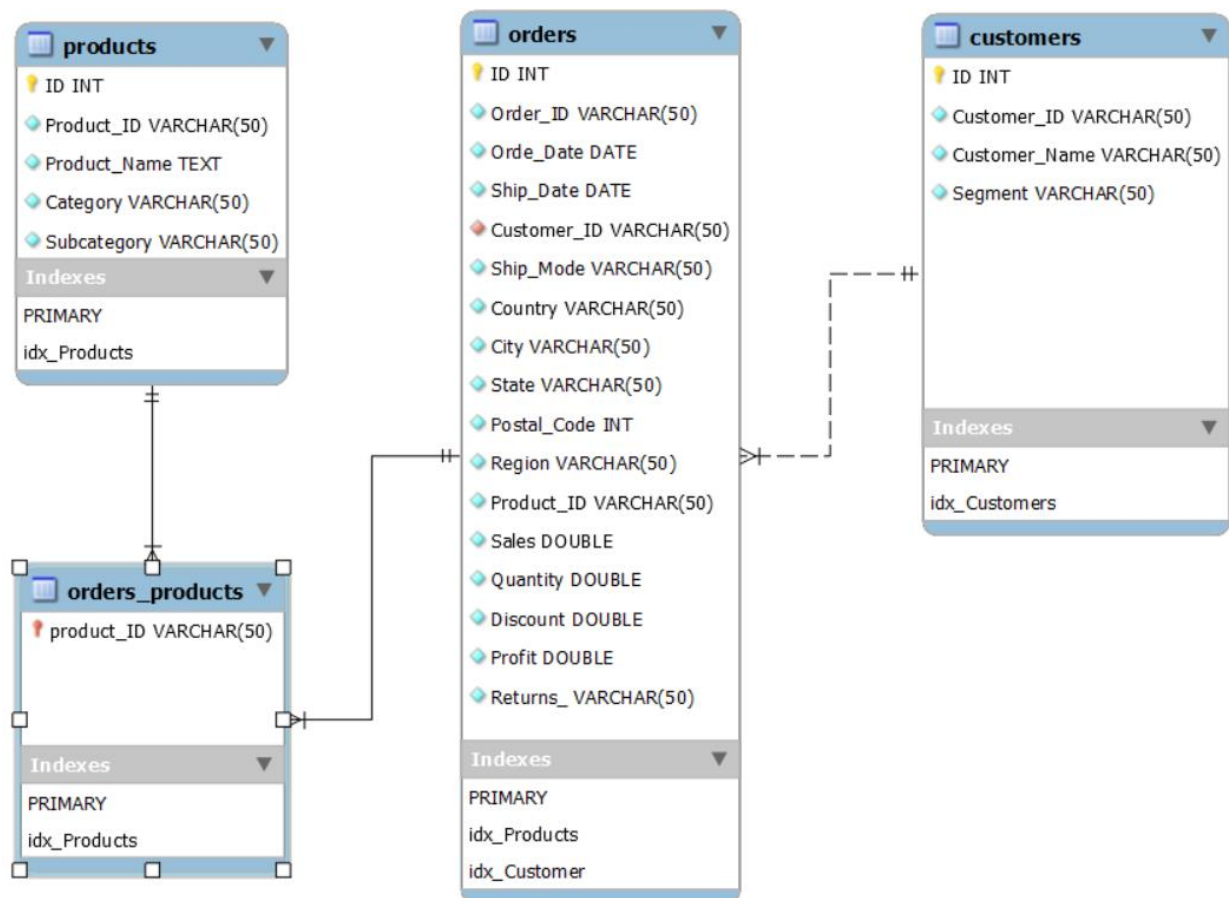
Variable Name	Type	Description
ID	INT	Primary key
Product_ID	Varchar	Unique code of the product, foreign key
Product_Name	TEXT	Product Name
Category	Varchar	The category assigned by the store to each product
Subcategory	Varchar	Subcategory assigned by the store to each product

1.1.3. Customers

The table has 793 Rows and 4 variables:

Variable Name	Type	Description
ID	INT	Primary key
Customer_ID	Varchar	Unique code of a customer, foreign key
Customer_Name	Varchar	Customer Name
Segment	Varchar	The segment of each customer

1.2. Database diagram



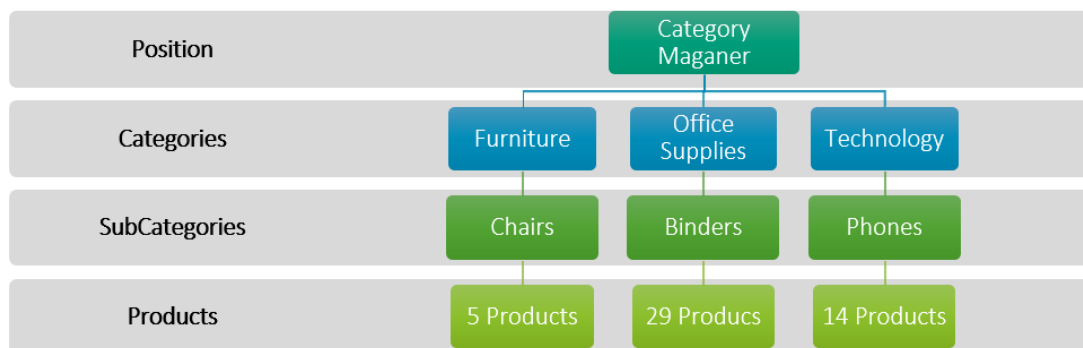
2. Analytical plan

The following scheme is the analytical plan.



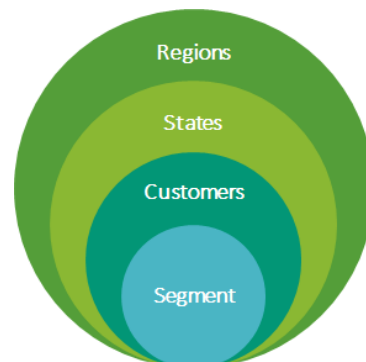
2.1. Category Department

This area is responsible for defining and managing the product mix within the store. The portfolio of the store is composed of 3 categories, 17 subcategories and 1,525 products. For this reason, the area needs information that makes it easier to analyze sales and profitability at each level to make decisions to optimize the store's offer.



2.2. Sales department

The sales area is in charge of managing customer relationships. The department has 4 managers for each region of the country (South, Center, East and West). These positions need information that will allow them to identify the areas of the country with the greatest potential. They also need to monitor customers and plan their visiting schedule based on sales and profitability.



3. Analytical Data Layer

The exercise will be based on the star schema approximation with a transaction fact table associated with sales and four dimensions: products, time, market, and customer. Each of these layers would serve to provide information for Sales and Category departments and would help define the data marts for the analytical requirements.

To ensure that the database do not consume many resources, I decided to sacrifice all the information related to operations (date of shipment and mode of shipment) because the project was not oriented to provide information to this area. We also left out the postal code and returns variables as they did not add much value to the departments involved.

Fact Sales				Dimension 1 Time		Dimension 2 Products				Dimension 3 Customers			Dimension 4 Market	
Order_ID	Sales	Quantity	Profit	Year	Month	Product_ID	Product_Name	Category	Subcategory	Customer_ID	Customer_Name	Segment	Region	State
CA2017140326	825.174	9	-117.88	2017	9	FURBO10000112	Bush Birmingham Collection Bookcase, Dark Cherry	Furniture	Bookcases	HW14935	Helen Wasserman	Corporate	Central	Illinois
CA2015130785	411.332	4	-4.8392	2015	9	FURBO10000330	Sauder Camden County Barrister Bookcase, Planked Cherry Finish	Furniture	Bookcases	AG10900	Arthur Gainer	Consumer	West	California

4. Data Marts

7 data marts have been created: 3 for the category department and 4 for the sales department. All contain transactional information (sales, units, and profit) as well as their variations with respect to the previous year.

Area	Data Mart	Type	Description	User input
Category	Category	Store procedure	The user will be able to observe the transactional information of the 3 categories of the company.	Year, Month
	SubCategory	Store procedure	The user will be able to zoom in on each category and find the transactional information of the subcategories that compose it.	Year, Month, category
	Top Products	Store procedure	The user will find information on the best-selling products in the subcategory. The user defines how many products he wants to include at the top	Year, month, subcategory
Sales	Regions	Store procedure	The user will be able to observe the transactional information of the 4 regions where the company has commercial relationships	Year, Month
	States	Store procedure	User will be able to view the transactional information of the states of the region selected	Year, Month, Region
	Customers	Store procedure	The user will be able to visualize the transactional information of the states of the selected region.	Year, Month, Region
	Segments	Store procedure	The user will be able to visualize the transactional information of the states of each segment of customers (Corporate, consumer, Home office)	Year, Month,