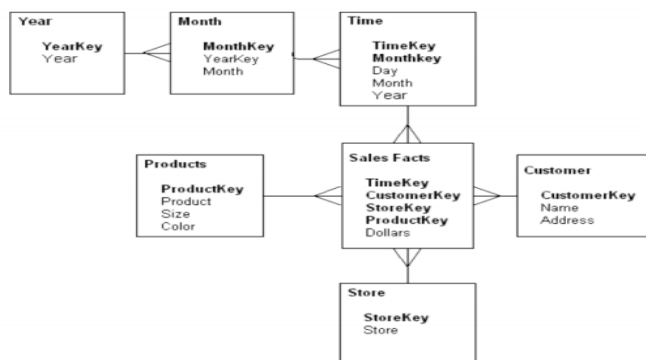# Dataware Housing Project

Q1. For the given Dimension modelling please identify the following:

* How many Dimensions and Facts are present?

➡ In the above Dimension modelling ,there is one Fact present that is "Sales Fact". Rest all Tables are Dimensions.

*Please Identify the cardinality between each table.

➡ Cardinality between:

1. <u>Year and Month: One to Many,</u> Because One Year Contains many months.

2. Month and Time is One to Many, as One month contains many times.

3. Time and Sales Fact is One to Many, as it is possible to have many time in a single sales.

4. Product and Sales Fact is One to Many,Many Products are there in One sale.

5. Customer and Sales Fact is One to Many,Many Customers buys in one Sales.

6. Store and Sales is One to Many,as Many Stores are present in Sales.

7. Customer and Product is Many to Many,Because One customer can buy many products and One Product can be bought by many Customers.

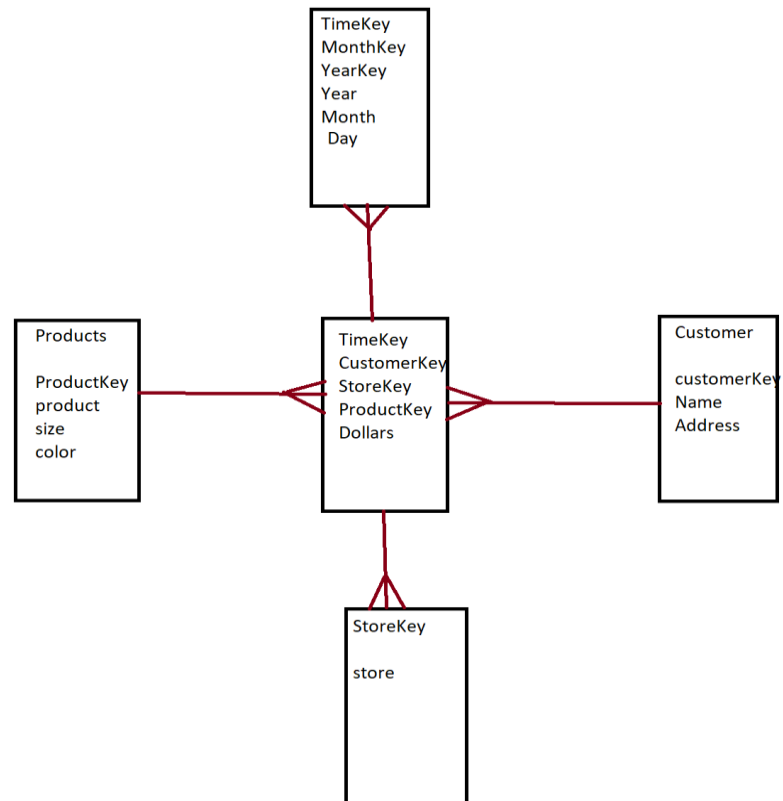8. Customer and Store is Many to Many,as One Customer can buy from many stores and in one store many customer can buy.

9. Product and Store is Many to Many, as 1 product can be present in many stores and 1 Store contains many products

***How to create Sales_Aggr using the Following Structure(SQL Statement)?**

```
Sales_Aggr

Year_ID
Customer_Key
Store_key
Product_key
Dollars
```
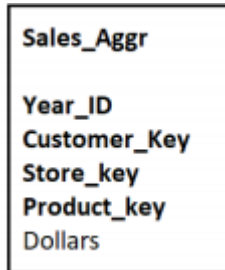
Create table Sales_Aggr(Year_ID INT,Customer_Key INT,StoreKey INT,ProductKey INT,Dollars BIGINT,foreign key(Year_ID) references Year(YearKey),foreign key(Customer_Key) references Customer(CustomerKey),foreign key(store_key) references Store(StoreKey),foreign key(Product_Key) references Product(Product_Key),primary key(Year_Id,Customer_Key,Store_Key,Product_Key));

*Can You Please Modify the above snowflake schema into star schema?

Products		TimeKey		Customer
		MonthKey		
ProductKey		YearKey		customerKey
product		Year		Name
size		Month		Address
color		Day		

TimeKey
CustomerKey
StoreKey
ProductKey
Dollars

StoreKey

store

Q2. For the Following Dimension model can you please give an example of circular join and how to avoid it?

Date
- Date
- Month
- MonthNumber

Sales
- OrderDate
- ShippingDate
- SalesAmount
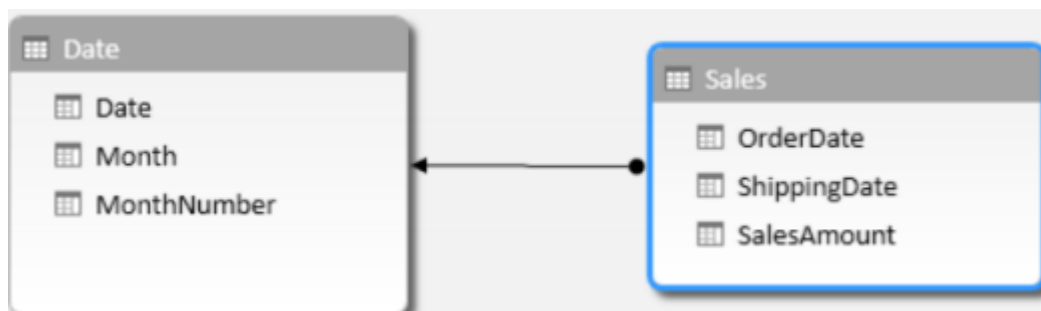
The two tables are present Date and Sales.

DATE

Date	Month	MonthNumber
10/01/2019	January	01
21/03/2019	March	03

Sales

OrderDate	ShippingDate	SalesAmount
10/01/2019	12/01/2019	10000
21/03/2019	24/01/2019	5000

Circular Join Query:

Select s.Shipping_Date, s.Order_Date from Date d.Sales where

s.ShippingDate=d.Date AND d.Date=s.OrderDate;

To Avoid this we will use alias:

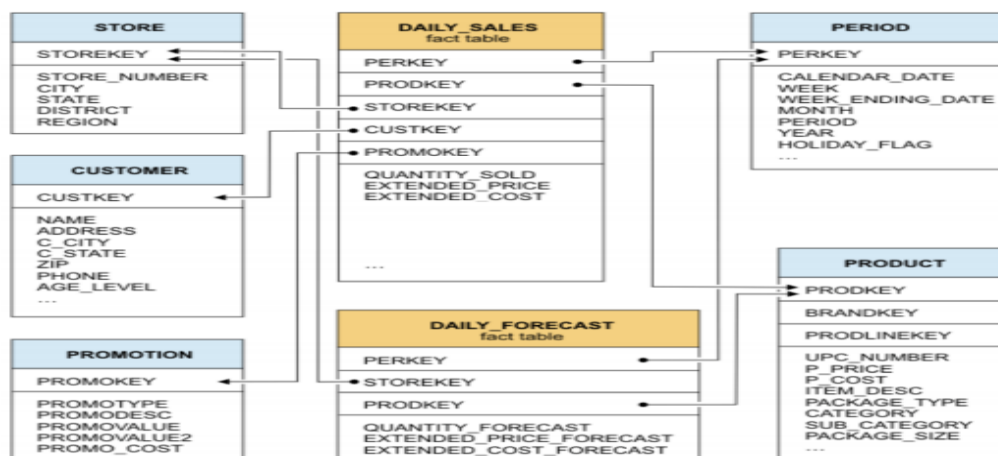Select s.Order_Date, s.SalesAmount, s.Shipping_Date

 from date as "Order_Date", date as "Shipping_Date",

Sales.Date d

where Shipping_Date.date=s.Shipping_Date AND

Order_Date.date=s.Order_Date;

Q3.For the given Dimension model ,can you please generate a SQL to get total Divergence between Quantity sold and Quantity forcast for the current time in each store?

Select sum(Quantity_Sold) - sum(Quantity_Forcast)

As Divergence

from Daily_Sales,Daily_Forcast,Period p

where Month(p.Month)-Month(Current_Date())

 group by store key;

Here we need to find Divergence so need to Calculate sum of the total quantity forcasted and total that really sold and then the difference of total forcasted and total sold is divergence.we need to find it for each store so we grouped it by store key.

Q4. For the above-mentioned dimension model, please identify the conformed and nonconformed dimensions. Additionally, identify the measure types?

Product,Store,Period dimensions are conformed dimensions.

Rest all Dimensions are non conformed dimensions.

Among all the measures in both facts :

Quantity_Sold: Additive

Extended_Price:Semi Additive

Extended_Cost:Additive

Quantity_Forcast:Additive

Extended_price_Forcast:Non Additive

Extended_Cost_Forcast:Non Additive

Q5.Make a list of differences between DW and OLTP based on Size, Usage, Processing and Data Models

OLTP	DataWare (OLAP)
1. It contains operational data and it is original source of data	1.It contains consolidation data and data comes from different OLTP Databases.
2.Basically used to control and run fundamental business task.	2.It basically help with problem solving,planning and decision making
3.Comparatively Faster	3.Depends on size of data present Batch Data refreshes and complex queries may take hours.
4.Relatively smaller in size if historical data is archived.	4. Larger in size,It is due to the presence of aggregation structures and historical data and requirement of more indexes than OLTP.
5.Highly normalized with many tables.	5.It is denormalized with lesser tables with star schema or snowflake schema.
6.It reveals a snapshot of ongoing business processes	6. It gives multidimensional views of various kinds of activities.

Data Model:OLTP (ER Model)

Dimension Model:OLAP