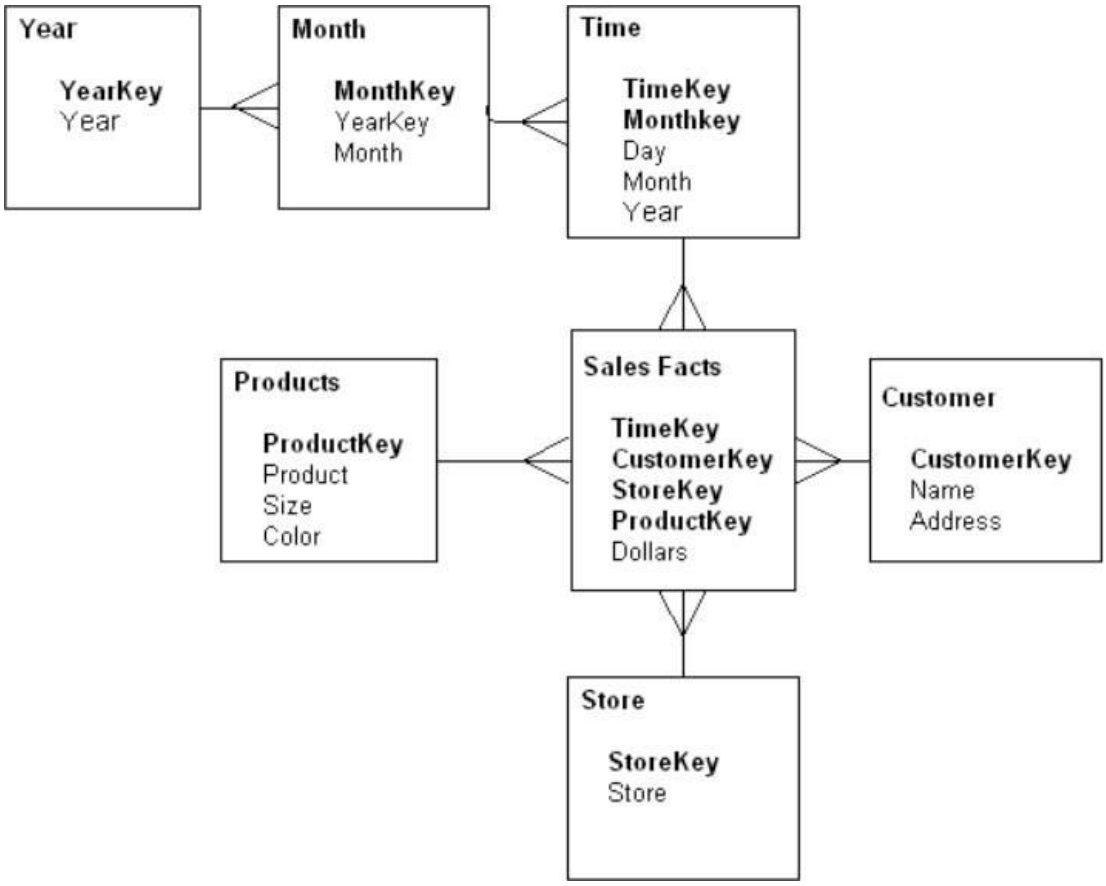
Data Warehouse Assessment 2

1. For the given Dimensional Modelling, please identify the following:



Ans: 1.1
A **fact** table is found at the center.
A **dimension** is a structure that categorizes facts.

So by these defination we can say that we have 1 Fact and 6 Dimensions in this Dimension modelling.

Fact is: **Sales Facts**
Dimensions are :**Year ,Month ,Time ,Products, Customer, Store**

Ans:1.2

Table Name	Table Name	Cardinality
Year (Dimension)	Month (Dimension)	1 : Many (Year : Month)
Month (Dimension)	Time (Dimension)	1: Many (Month : Time)
Time (Dimension)	Sales Facts (Fact)	1: Many (Time : Sales Fact)
Products (Dimension)	SalesFacts (Fact)	1: Many (Products : Sales Facts)
Customer (Dimension)	Sales Facts (Fact)	1: Many (Customer : Sales Facts
Store (Dimension)	Sales Facts (Fact)	1: Many (Store : SalesFacts)

Ans:1.3

I am writing here this query for Mysql and if we want then we can also modify this by change their data type for SqlPlus.

Create Table Sales_Aggr(YearKey as Year_ID INT(78) ,
CustomerKey number as Customer_Key INT(34) ,
StoreKey as Store_Key INT(36) ,
ProductKey number as Product_Key INT(25) ,
Dollars INT(65) ,
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 Products(ProductKey));

Ans1.4:

star schema considered denormalized because the star schema takes the information from the fact table (which is also often denormalized, meaning it is grouped and may contain redundancies) and splits it into denormalized dimension tables. The emphasis for the star schema is on query speed. Only one join is needed to link fact tables to each dimension, so querying each table is easy

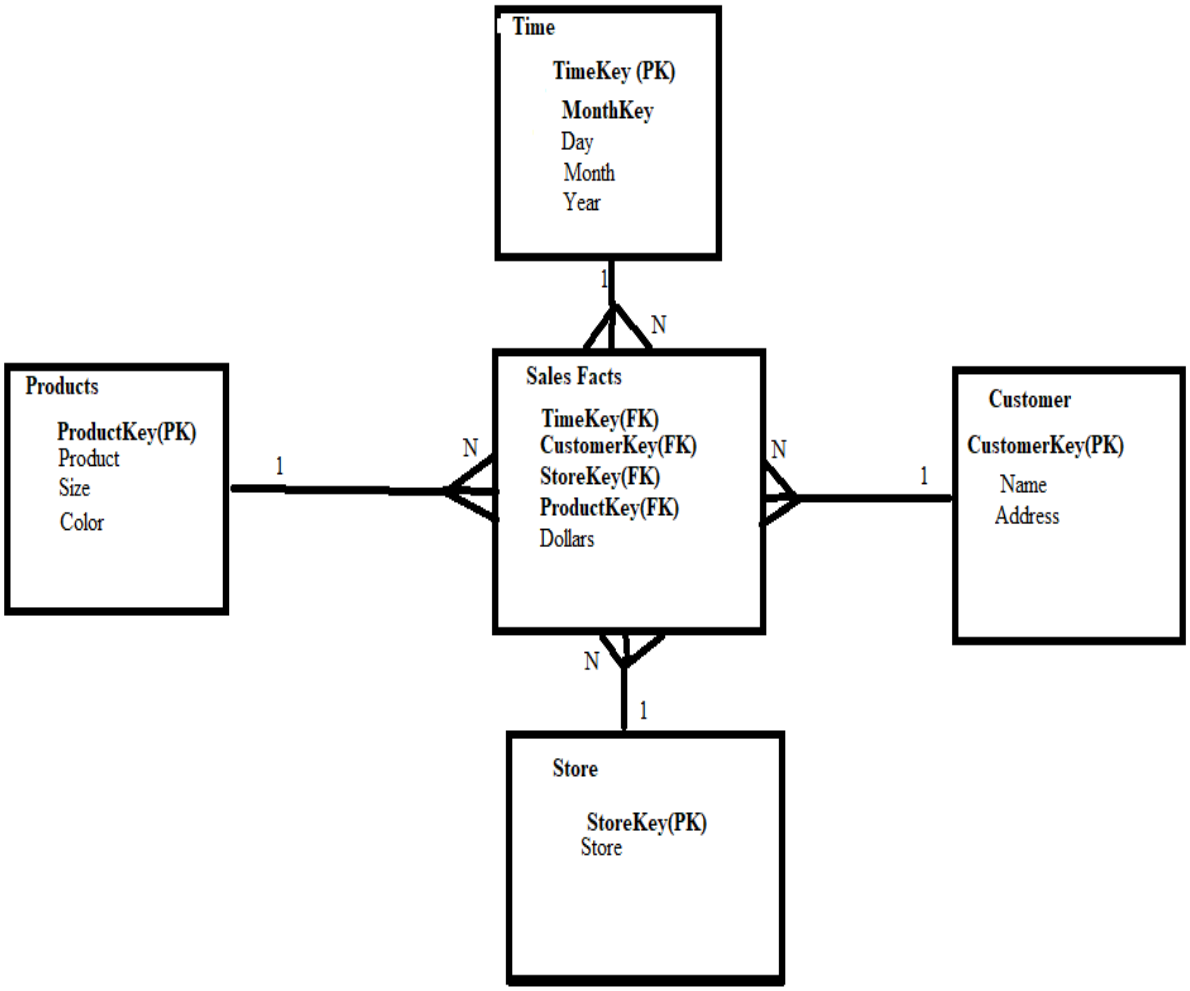
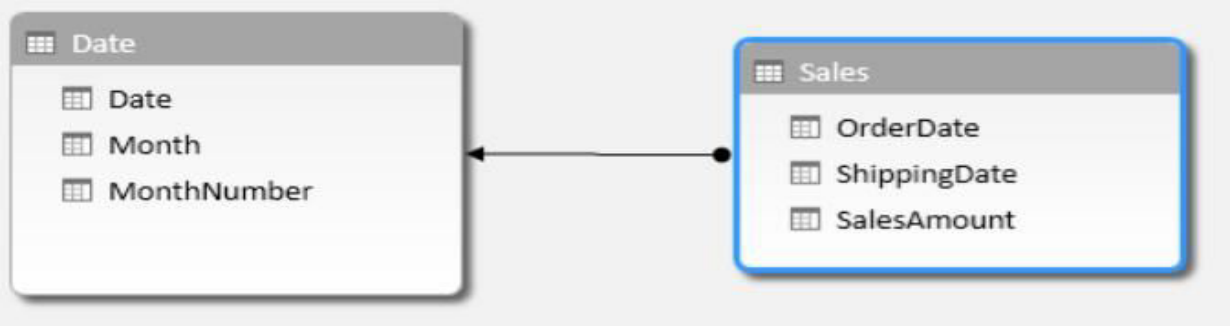


Fig:Star Schema

2.For the following dimension Model can you please give an example of Circular Join and how to avoid it:



Ans:

Circular Join: A **circular join** is where there are 2 routes to get from one table to another, going from a dimension to a fact. ... A real **circular join** is where usually 2 dimension tables **join** to each other, and each dimension tables also **joins** to a fact.

Example:

We have a circular join:
DimA(1)------->(M)FactA
DimB(1)------->(M)FactA
DimA(1)------->(M)DimB
(In this case DimA is the W_LOV_D table.)

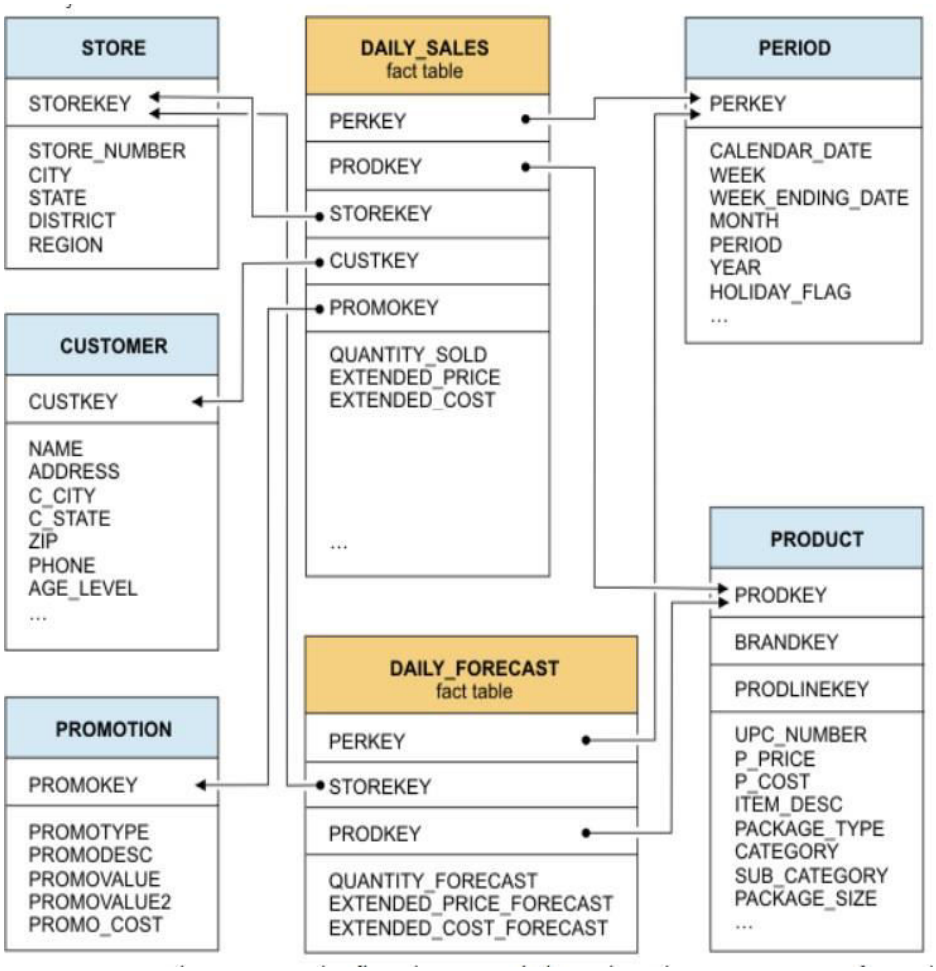
To resolve this we used an alias and came up with:
DimA(1)------->(M)FactA
DimB(1)------->(M)FactA
AliasDimA(1)-->(M)DimB

So the summary is we can resolve or avoid this join by give two alias name.

In our Question we have Date table with Date attribute so it will create confusion so we can avoid it by giving two alias name for its Date Attribute.

Select T.ODate,T.SDate
from Date as SDate ,Sales as T,Date as ODate
Where (ODate.Date=T.OrderDate AND SDate.Date=T.ShippingDate);

3.For the given Dimension Model, can you please generate a sql to get the total divergence between Quantity sold and Quantity Forecast for the current month for all the stores.



Divergence:the process or state of diverging.

Select abs((select sum(QUANTITY_SOLD) as Benefit from DAILY_SALES as DS ,PERIOD as P

Where P.MONTH=(select MONTH(CURDATE())) as CurrentMonth-

(Select sum(QUANTITY_FORECAST) as Prediction

From DAILY_FORECAST as DF,P

Where P.MONTH=(Select CurrentMonth(CURDATE()))as divergence

From DS,DF,P;

4.**For the above-mentioned dimension model, please identify the conformed and non-conformed** dimensions. Additionally, identify the measure types?

Conformed dimensions: In data warehousing, a conformed dimension is a dimension that has the same meaning to every fact with which it relates. Conformed dimensions allow facts and measures to be categorized and described in the same way across multiple facts and/or data marts, ensuring consistent reporting across the enterprise.

So we can say in above question we have conformed dimensions are:

Fig: Conformed dimension

Non Conformed dimensions: A **non-conformed dimension** is where you add an attribute to an analysis which does not logically relate to all the facts. Take this example, here we have an analysis with 2 "facts" and 1 "**conformed dimension**" i.e. the "Store" **dimension** attribute relates to both "# POIs" and "# Orders"

So we can say in above question we have non conformed dimensions are:

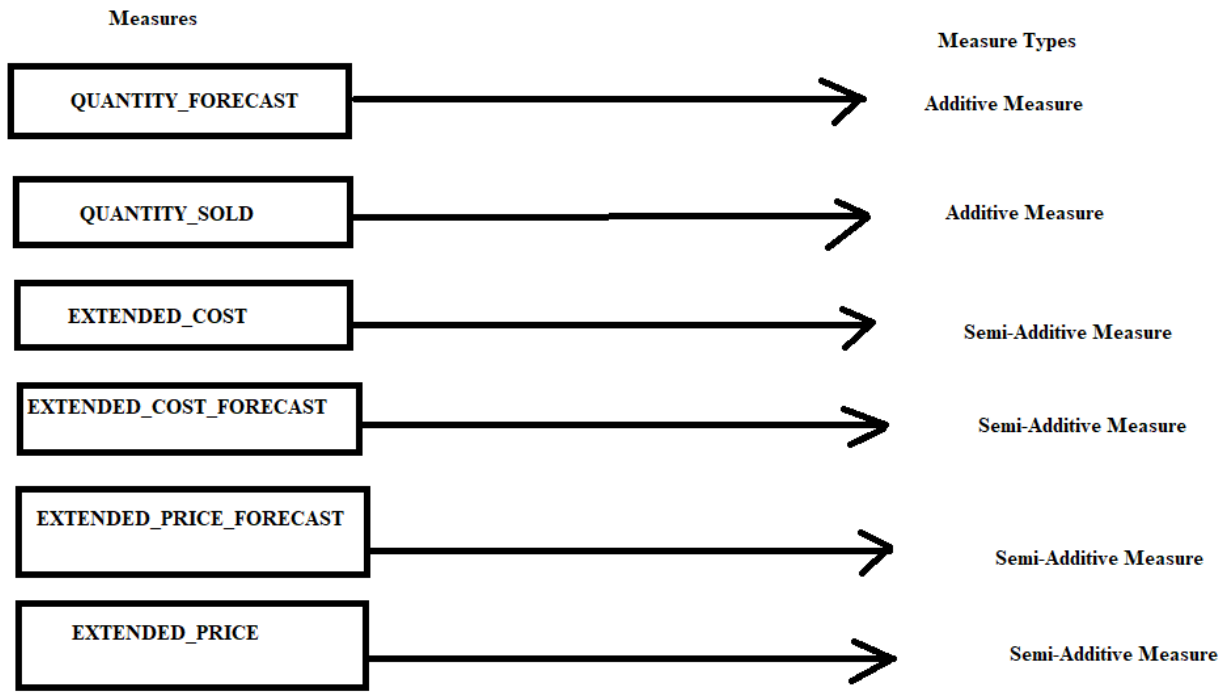
Fig: non conformed dimension

Measure: In a data warehouse, a measure is a property on which calculations (e.g., sum, count, average, minimum, maximum) can be made.

So there are three type of dimension

1. **Additive :** Additive measures can be summed across any of the dimensions associated with the fact table. Example: summation of Sales for a month or year.
2. **Semi-additive :** Measures can be summed across some dimensions, but not all; balance amounts are common semi-additive facts because they are additive across all dimensions except time
3. **Non-additive :** These are those specific class of fact measures which cannot be aggregated across all/any dimension and their hierarchy. Example: Facts which have percentages, ratios calculated.

So by these defination we can conclude that in our above given question we have measures are:



5.MakealistofdifferencesbetweenDWandOLTPbasedonSize,Usage,Processing and DataModels.

Comparisons based on	**DW OR OLAP (Online Analytical Processing)**	**OLTP (Online Transaction processing)**
SIZE	Large amount of data is stored typically in TB, PB	The size of the data is relatively small as the historical data is archived. For ex MB, GB
Usage	It is subject oriented. Used for Data Mining, Analytics, Decision making,etc.	The data is used to perform day to day fundamental operations.
Processing	Relatively slow as the amount of data involved is large. Queries may take hours.	Very Fast as the queries operate on 5% of the data.
Data Models	Dimension model	Entity-Relationship (E-R) model EER MODEL(Enhance Entity Relationship)