DATAWAREHOUSE ASSESSMENT2

1A)How many dimensions and Facts are present?

Number of dimensions:6

Number of facts:1 (SALES FACT)

1B) Please identify the cardinality between each table?

Cardinality

- Year – Month -> One to Many
- Month – Time -> One to Many
- Time – Sales Facts -> One to Many
- Customer – Sales Facts -> One to Many
- Store – Sales Facts -> One to Many
- Products – Sales Facts -> One to Many

1C) How to create a Sales_Aggr fact using the following structure (SQL Statement):

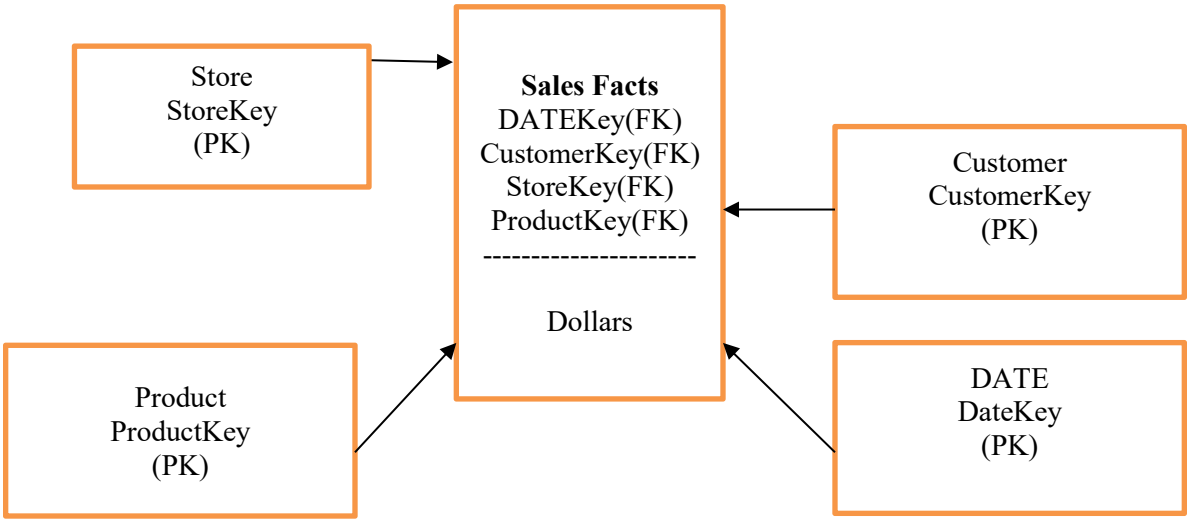
Create table sales_aggregate(select customer_key, storekey, product_key,y.yearkey from sales customer,store,product ,year y);

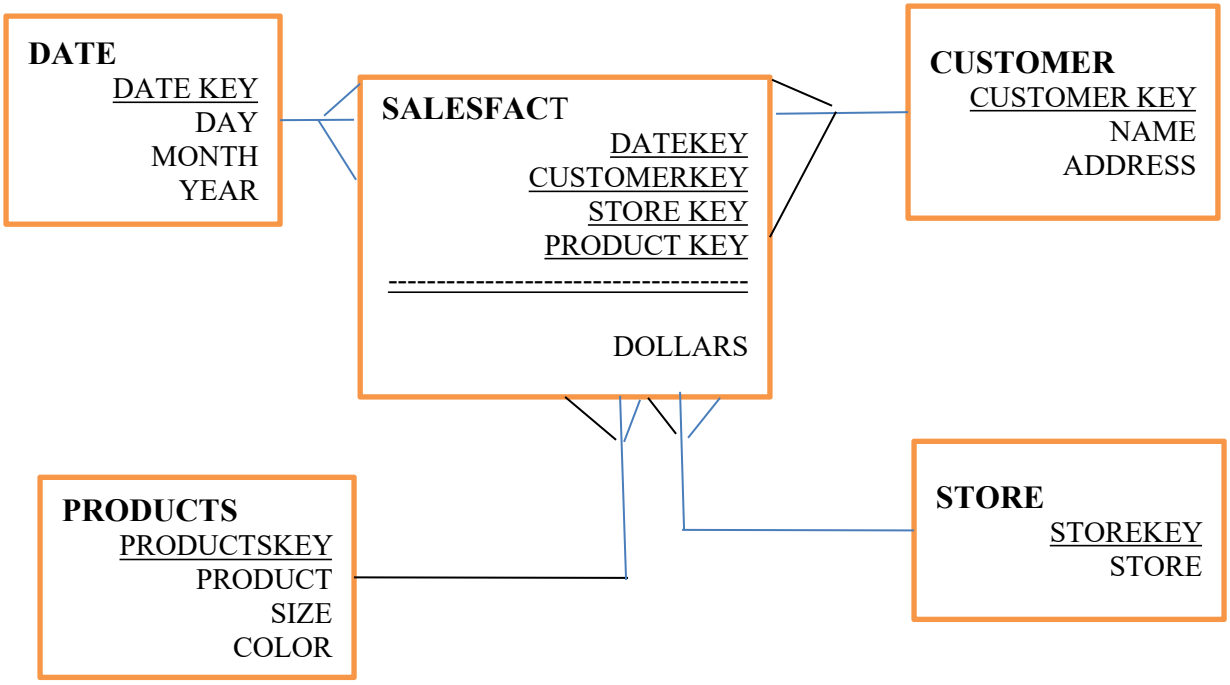
FOR ADDING DOLLARS

Alter table sales_aggegrate add dollars double(40);

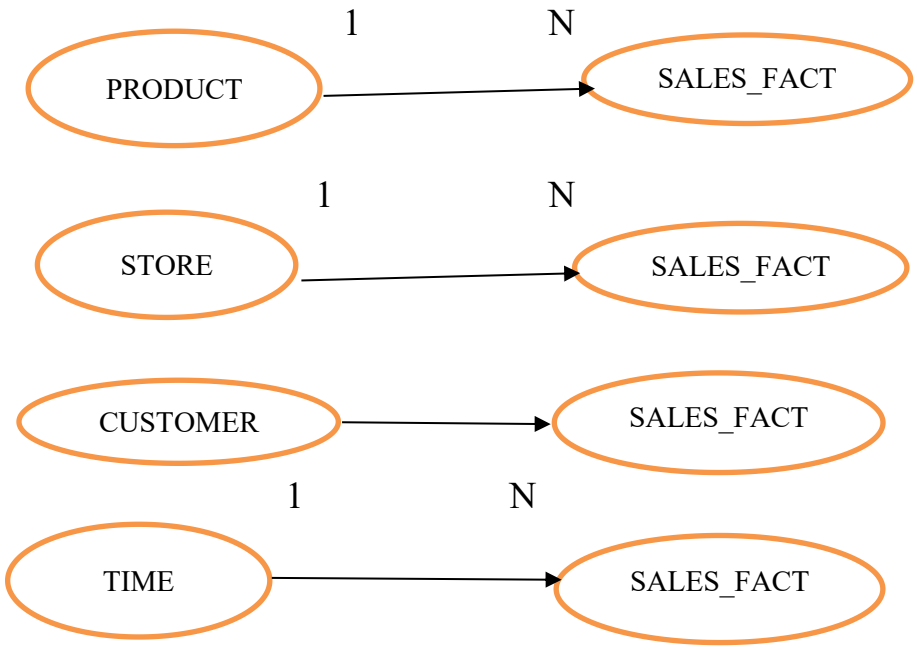
1D) Can you Please Modify the above snowflake schema to Star schema and draw the dimension model, showing all the cardinality?

- PRODUCT- SALES_FACT ->ONE TO MANY
- STORE – SALES_FACT -> ONE TO MANY
- CUSTOMER- SALES_FACT -> ONE TO MANY
- DATE -SALES_FACT ->ONE TO MANY

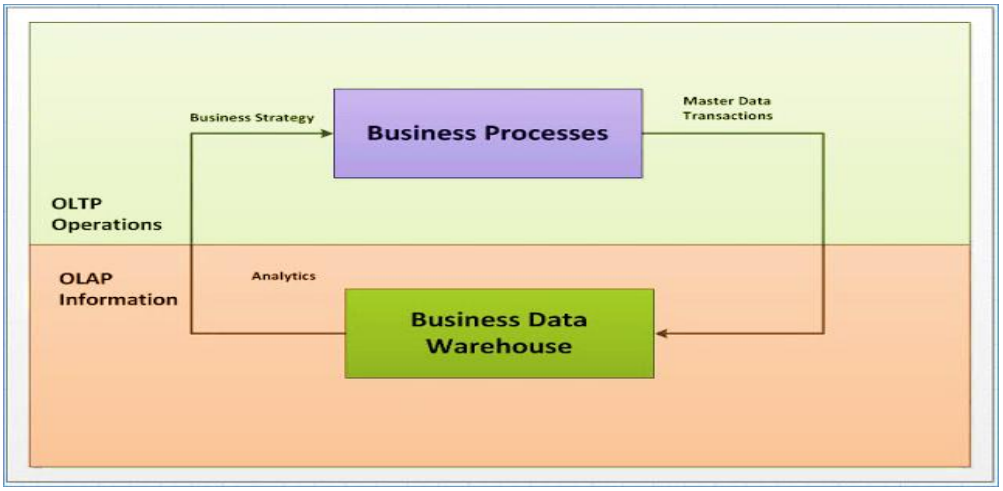

Store
StoreKey
(PK)
Sales Facts
DATEKey(FK)
CustomerKey(FK)
StoreKey(FK)
ProductKey(FK)
Dollars
Customer
CustomerKey
(PK)
Product
ProductKey
(PK)
DATE
DateKey
(PK)


DATE
DATE KEY
DAY
MONTH
YEAR
SALESFACT
DATEKEY
CUSTOMERKEY
STORE KEY
PRODUCT KEY
DOLLARS
CUSTOMER
CUSTOMER KEY
NAME
ADDRESS
PRODUCTS
PRODUCTSKEY
PRODUCT
SIZE
COLOR
STORE
STOREKEY
STORE

- PRODUCT- SALES_FACT ->ONE TO MANY
- STORE – SALES_FACT -> ONE TO MANY
- CUSTOMER- SALES_FACT -> ONE TO MANY
- DATE -SALES_FACT ->ONE TO MANY



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

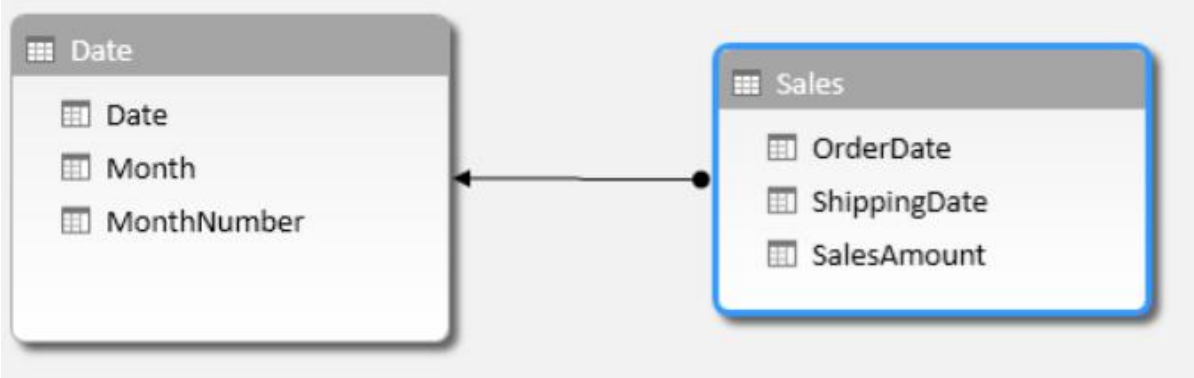


	OLTP	DW
size	Less in size in MB	More in size compared to OLTP in TB
usage	Designed for real time business operations.	Designed for analysis of business measures by category and attributes.
processing	It provides fast result for daily used data.	It ensures that response to the query is quicker consistently.
Data models	ER Modelling	Dimensional modelling

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

Conformed dimensions	Store,period,product
Non conformed dimensions	Customer,promotion
Additive measures	Quantity sold,Quantity_forecast
Semi additive measure	Extended_price,Extended_cost,Extended_price_forecast,Extended_cost_forecast
Non additive measure	None

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



DATE TABLE

DATE	MONTH	MONTHNO
D1	M1	1
D2	M2	2
D3	M3	3

SALES

ORDERDATE	SHIPPINGDATE	SALESAMOUNT
O1	S1	55555
O2	S2	99999
O3	S3	777777

- In circular join,if we order a product ,it can be ordered on the same day , shipped and delivered also sometimes on the same day. So,their might be a

ambiquity in the column and insertion or updating becomes difficult sometimes.

- Circular Joins can be avoided by making use of Aliases for the repetitive dimension table.

REMOVAL OF CIRCULAR JOIN BY ALIASING BY EXECUTING QUERY

Select sal.shipping_date,sal.order_date,sal.sales_amt from

Date_table as order_date ,date_table as shipping_date from sales s,date d where

S. orderdate=order.orderdate;

Explanation: Here in the date_table the date column is referring to both orderdate and shippingdate of the sales table.

- So,this might lead to a confusion that which date refers to what.
- So, we divide the date column of the date_table into Order date and Shipping date seperately by aliasing.

BEFORE

DATE_TABLE

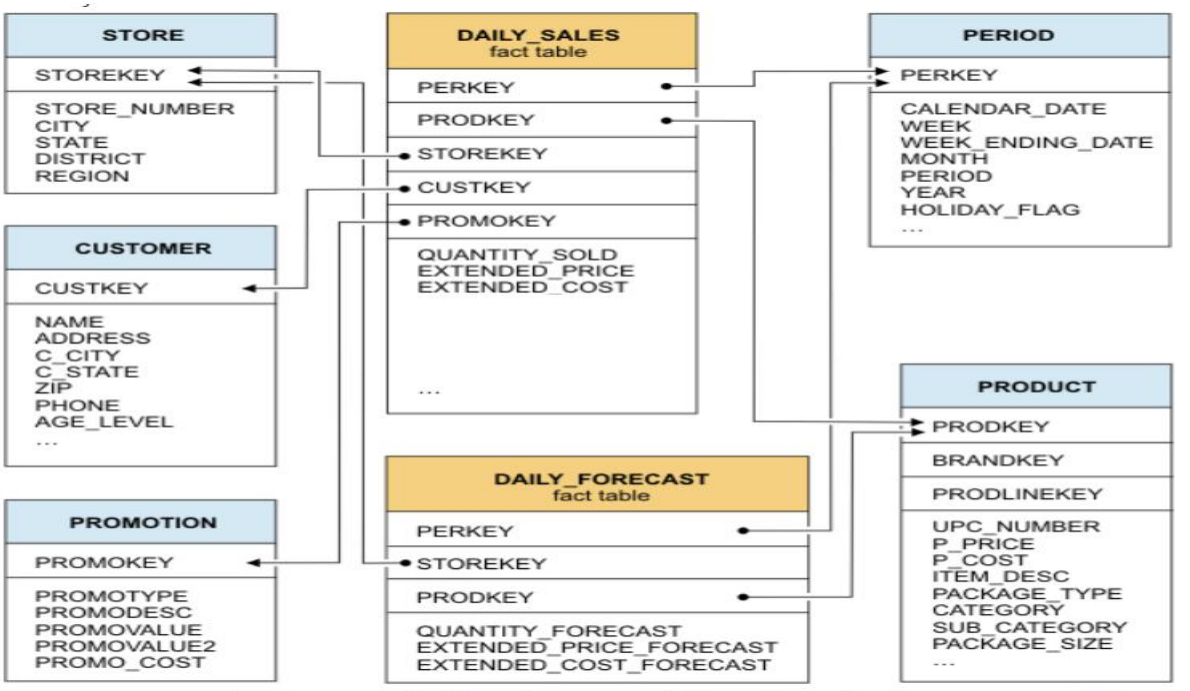
DATE	MONTH	MONTHNO
D1	M1	1
D2	M2	2
D3	M3	3

AFTER

DATE_TABLE

ORDERDATE	SHIPPINGDATE	MONTH	MONTHNO
O1	S1	M1	1
O2	S2	M2	2
O3	S3	M3	3

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 is difference between a table
- Example the difference between prediction and exact value

Select sum(quantity_sold) -sum(quantity_forecast) as d

From daily_forecast,daily_sales,period where

Month(period.month)-month(current_date)

Group by perkey;