DATA WAREHOUSE ASSESSMENT

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

- How many dimensions and Facts are present?

Ans:-There is one fact table and six dimension table.
Fact table is Sales_Fact and dimensional tables are Products,Store,Costumer,Time,Month,Year.

- Please identify the cardinality between each table?

Ans:-

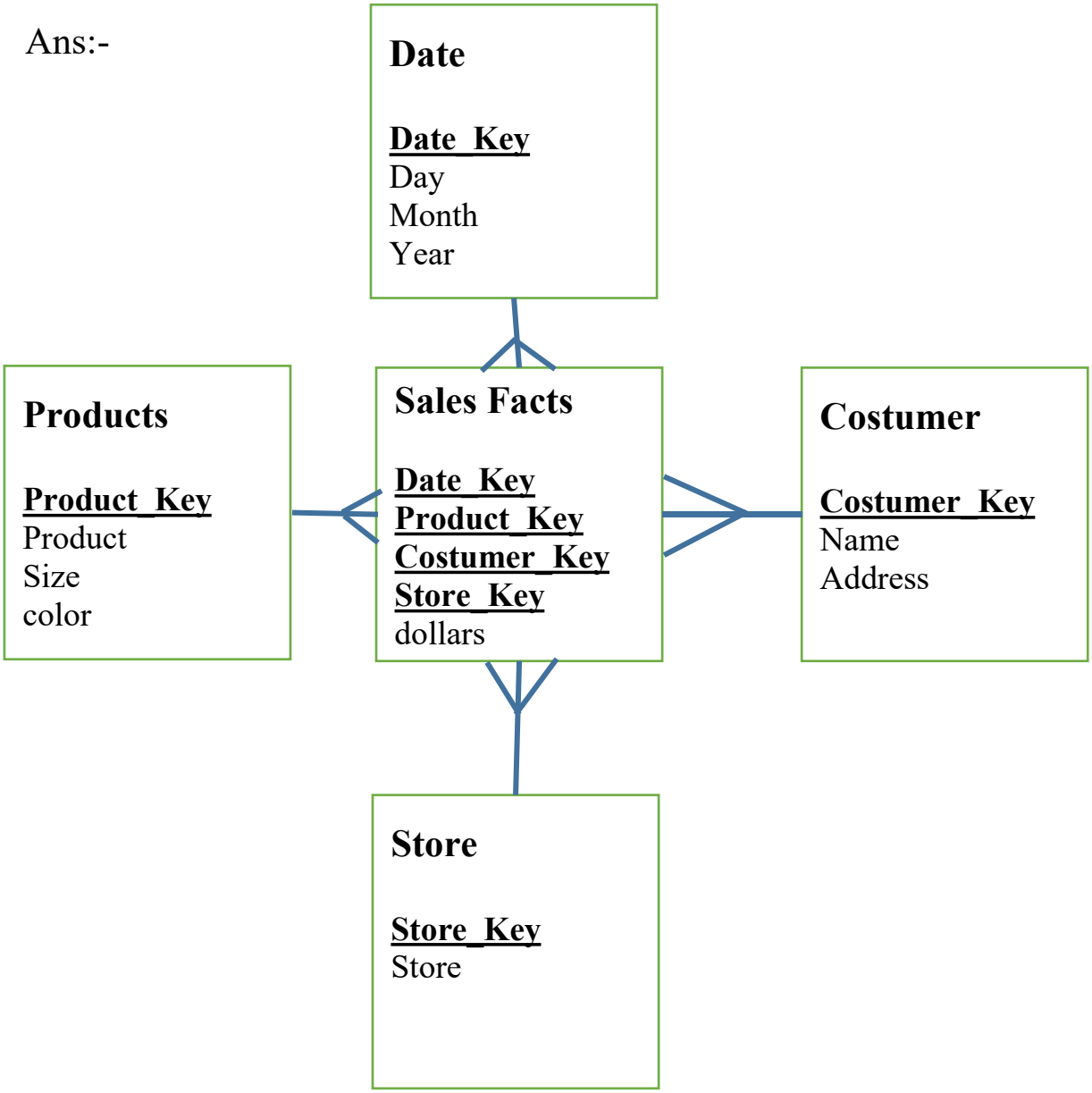
Table	Cardinality	Table
Year	[1:N]	Month
Product	[1:N]	Sales_Fact
Costumer	[1:N]	Sales_Fact
Month	[1:N]	Time
Time	[1:N]	Sales_Fact
Store	[1:N]	Sales_Fact

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

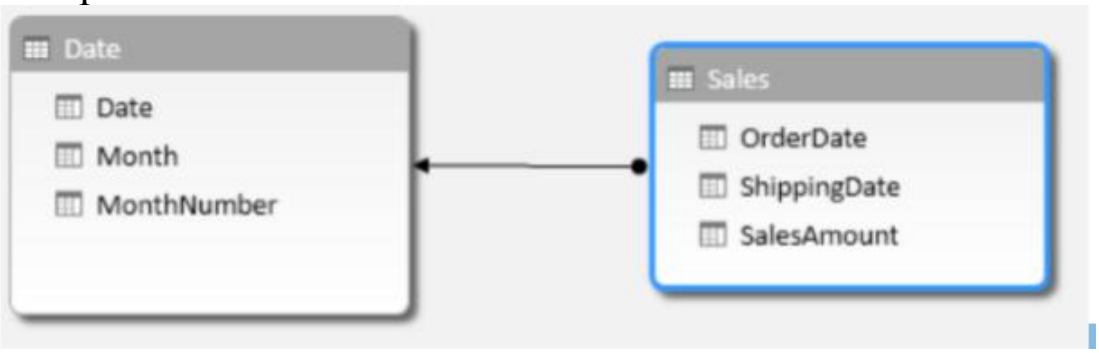
Ans:-Create table Sales_Aggr (Year_Id int,Product_Key int,Costumer_Key int,Store_Key int,dollars double,
Foreign key(Product_Key) references Products(ProductKey),
Foreign key(Costumer_Key) references Costumer(CostumerKey),
Foreign key(Store_Key) references Store(StoreKey),
Foreign key(Year_Id) references year(yearkey),
Primary key(Year_Id,Costumer_Key,Store_Key,Product_Key));

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

Ans:-



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



Ans:-Query for circular join:-

Select S.SalesAmount,S.OrderDate,S.ShippingDate
From Sales S,Date D
Where S.OrderDate=D.Date and
S. ShippingDate=D.Date

If we apply this query there we may get wrong result.

Query to avoid circular join:-

Select S.SalesAmount,S.OrderDate,S.ShippingDate
From Sales S,Date D,Date D1
Where S.OrderDate=D.Date and
S.ShippingDate=D1.Date

By executing this query we can get correct result.

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:

Ans:-
Select Sum(D1.Quantity_Forecast)-Sum(D2.Quantity_Sold) as Divergent_Quantity
From Daily_Sales D1,Daily_Forecast D2,Period P
where D1.Perkey=P.Perkey And D2.Perkey=P.Perkey
And P.month=Extract(month from curdate()
And P.month=Extract(year from curdate());

This query will fetch the diff of quantity by matching the period key and matching the current date and current year.

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

Ans:-In the above dimensional model the conformed dimensions are:-Store,Product,Period
Non-confirmed dimensions are:-Promotion,Store.
Measure types:
Additive
Measures:-Quantity_Sold,Quantity_Forecast,,Extended_Cost,Extended_Cost_Forecast
Semi-Additive
Measures:-Extended_Price,Extended_Price_Forecast.

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

Ans:-

Criteria	OLTP	DW
Size	Less in Size (mostly in Mbs or few Gbs).	Size is hige.(Tbs or Pbs).
Usage	It is used to store	It is used for

	the transactional data or current data.	bussiness analysis.It extracts,tranform and load all the required data.
Processing	IUD operations are faster but to extract data we need to do multiple joins.That reduces the speed of select operations.	Select operations becomes faster because of lesser joins.IUD operations are slower than OLTP.
Data Models	ER Model Normalized.	Dimensional Model. De-Normalized.