DATAWAREHOSE   ASSIGNMENT-1
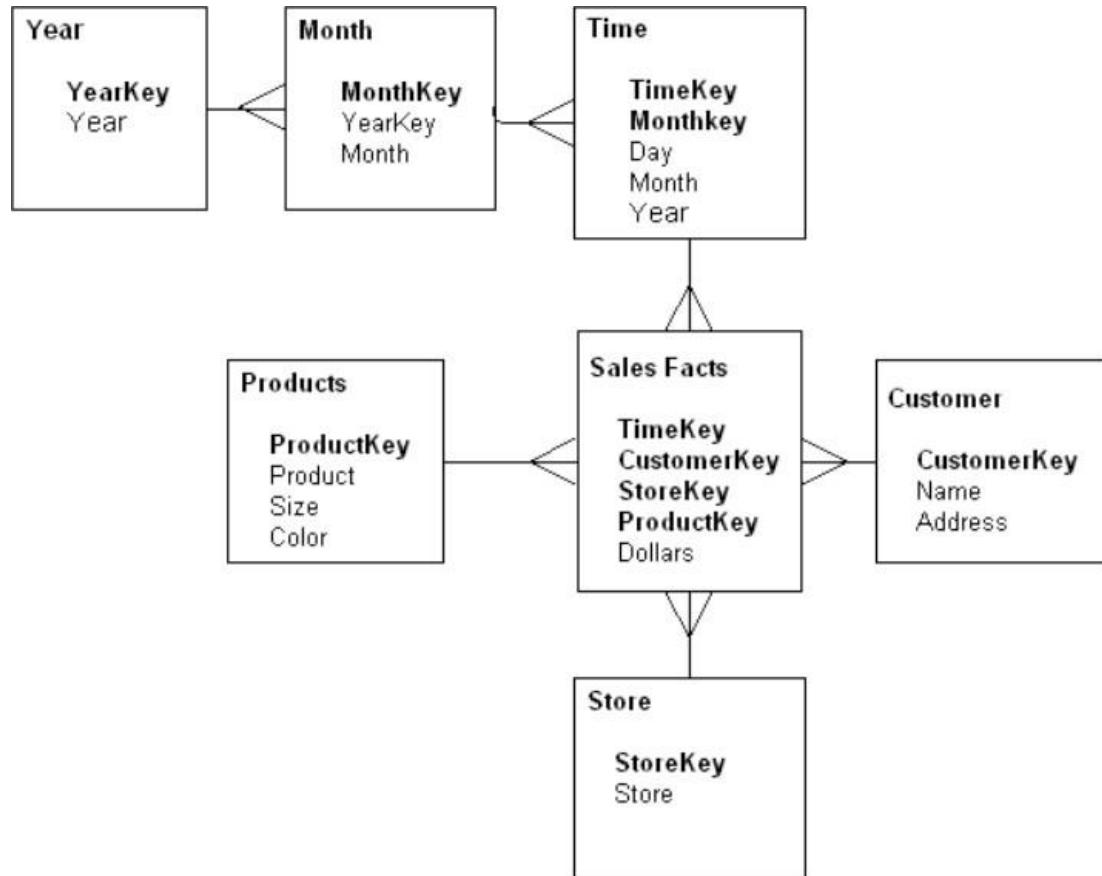
DATAWAREHOUSE ASSIGNMENT-1

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

1.(a) How many dimensions and Facts are present?
ANS-   Dimensions - 6
          Facts - 1

1.(b) Please identify the cardinality between each table?
ANS -

TABLES	CARDINALITY
Year-to-Month	1:N
Month-to-Time	1:N
Time-to-Sales Facts	1:N
Products-to-Sales Facts	1:N
Customer-to-Sales Facts	1:N
Store-to-Sales Facts	1:N

DATAWAREHOUSE ASSIGNMENT-1

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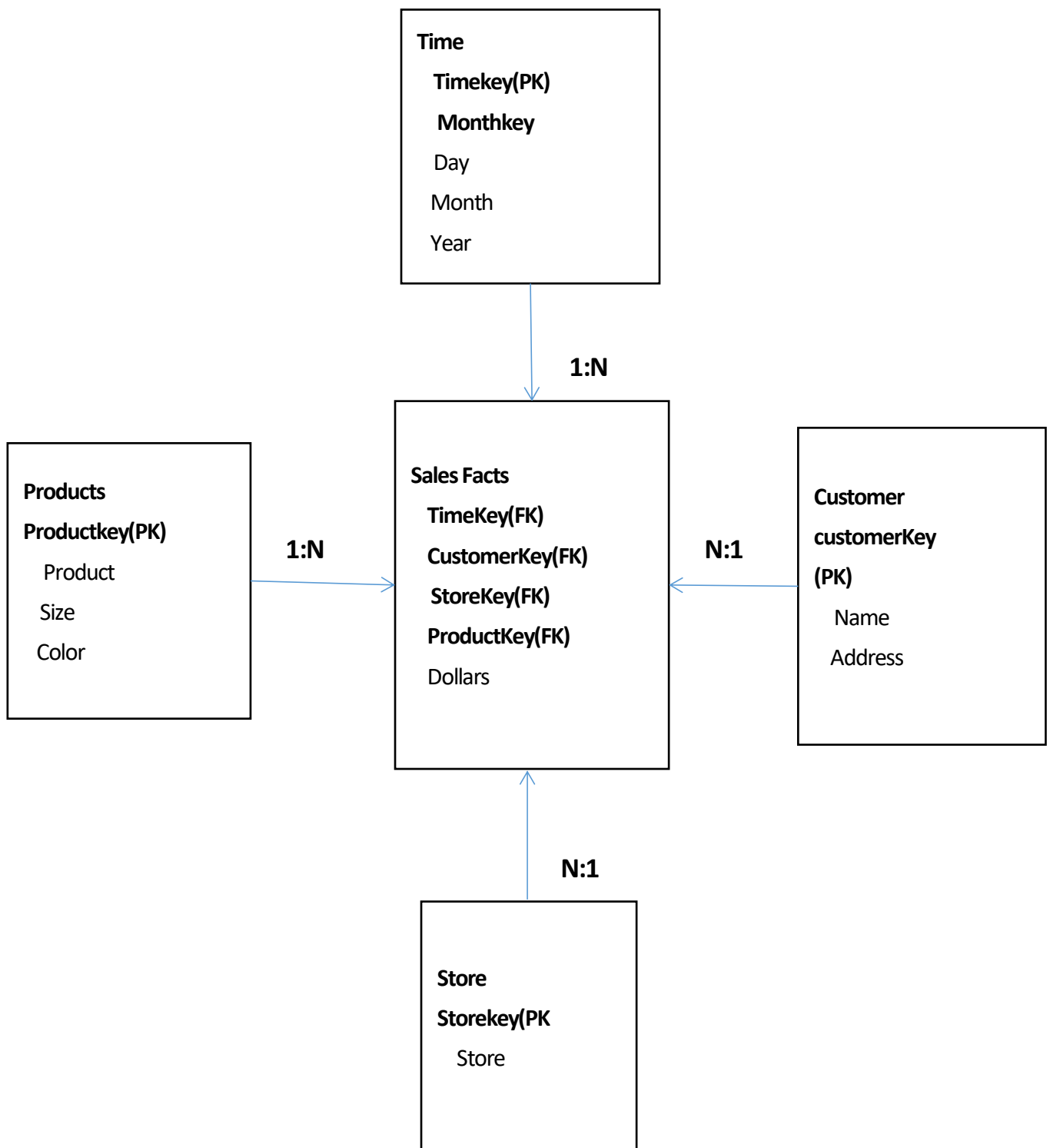
```
Sales_Aggr

Year_ID
Customer_Key
Store_key
Product_key
Dollars
```

ANS- Create table Sales_Aggr(Year_ID INT(10) primary key, Customer_key INT(20) primary key,Store_key int(10) primary key, Product_key int(20) primary key Dollars double 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);

1.(d) Can you Please Modify the above snowflake schema to Star schema and draw the dimension model, showing all the Cardinality?

DATAWAREHOUSE ASSIGNMENT-1

ANS-

Time
  Timekey(PK)
   Monthkey
  Day
  Month
  Year

1:N

Products
Productkey(PK)
  Product
  Size
  Color

1:N

Sales Facts
  TimeKey(FK)
  CustomerKey(FK)
   StoreKey(FK)
  ProductKey(FK)
  Dollars

N:1

Customer
customerKey
(PK)
   Name
   Address

N:1

Store
Storekey(PK
   Store

DATAWAREHOUSE ASSIGNMENT-1

Tables	Cardinality
Time-to-Sales facts	1:N
Products-to-Sales facts	1:N
Customer-to-Sales facts	1:N
Store-to-Sales facts	1:N

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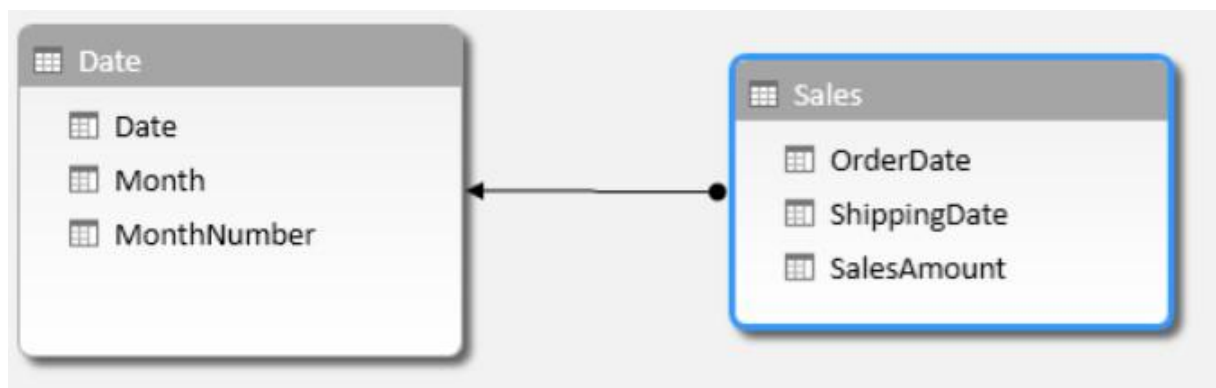
ANS-

Date table

Date	Month	MonthNumber
6/11/2019	November	11
27/2/2019	February	2
19/8/2019	August	8

Sales table

| OrderDate	ShippingDate	SalesAmount
7/4/2019	25/4//2019	8900
10/2/2019	17/2/2019	9800
1/8/2019	9/8/2019	7600
12/5/2019	18/5/2019	1200

DATAWAREHOUSE ASSIGNMENT-1

The following query will create a circular join

Select s.OrderDate,s.ShippingDate
From Date d Sales s
Where
d.Date=s.OrderDate AND
d.Date = s.ShippingDate;

Circular join can be removed by using aliases

Select s.OrderDate,s.ShippingDate
 Sales s,Date d
Where
OrderDate.d=s.OrderDate AND ShippingDate.d=s.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:
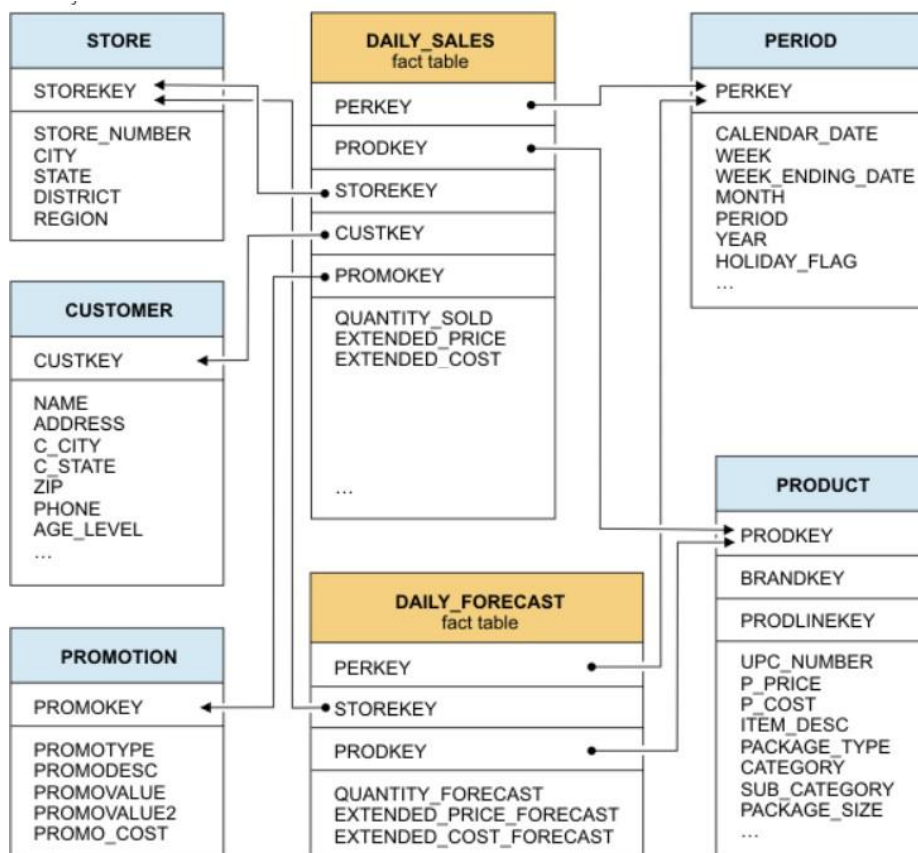
ANS- Select sum(ds.QUANTITY_SOLD) - sum(df.QUANTITY_FORECAST) AS divergence from ds DAILY_SALES , df DAILY_FORECAST p PERIOD where ds.PERKEY=p.PERKEY AND df.PERKEY=p.PERKEY AND p.PERKEY=(select PERKEY from PERIOD where MONTH=MONTHCURDATE());

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

ANS - Conformed dimensions - PERIOD, PRODUCT , STORE

NON - Conformed dimensions - CUSTOMER,PROMO

Measure types--

Additive measures - QUANTITY_SOLD,QUANTITY_FORECAST

Semi-Additive measures - EXTENDED_PRICE,EXTENDED_COST
EXTENDED_PRICE_FORECAST,EXTENDED_COST_FORECAST

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

ANS-

CRITEA	OLTP	DW(OLAP)
SIZE	Can be relatively small if historical date is archived	Large due to existence of aggregation structures and history data.
USAGE	To control and run Fundamental business tasks	To help with planning,problem solving,and decision support

PROCESSING	Typically very fast	Depends on the amount of data involved,batch data refreshes and complex queries may take many hours
DATA MODELS	Highly normalized with many tables	Typically de-normalized with fewer tables,use of star or snowflake schemas.