

1.

How many dimensions and Facts are present?

Dimensions:Year, Month, Time, Products, Customer, Store

Fact: Sales

Please identify the cardinality between each table?

Year:Month	=>	one to many
Month:Time	=>	one to many,
Time:Salesfact	=>	one to many
Customer:Salesfact	=>	one to many,
Store:Salesfact	=>	one to many

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

Sales_Aggr
Year_ID
Customer_Key
Store_key
Product_key
Dollars

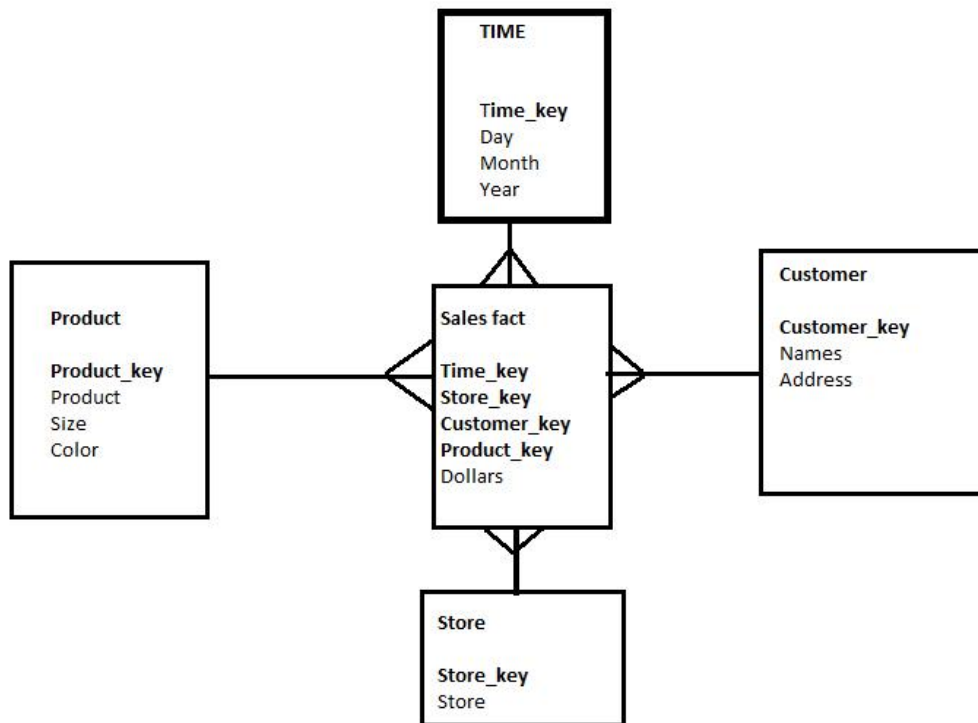
```
CREATE TABLE Sales_Aggr(Year_ID INT(4) , Customer_key INT(10),
Store_key INT(10), Product_key INT(10), Dollars INT(10), FOREIGN
KEY(Year_ID) REFERENCES Year(Year_KEY),FOREIGN KEY(Customer_key)
REFERENCES Customer(Customer_key),FOREIGN KEY(Store_key)
REFERENCES Store(Store_key), PRIMARY KEY(Year_ID, Customer_key,
Store_key, Product_key));
```

```
INSERT INTO Sales_Aggr VALUES(SELECT Y.Year_ID, S.Customer_key,
St. Store_key, P.Product_key, SUM(Dollars)
```

```
FROM Sales S, Year Y, Store St, Customer C, Product P) ;
```

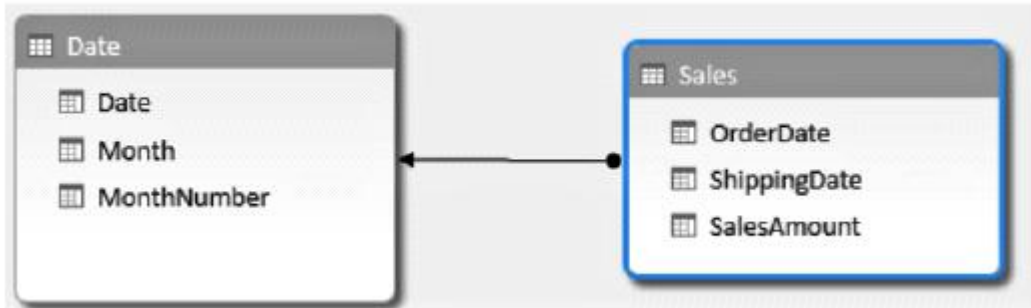
Can you Please Modify the above snowflake schema to Star schema and draw the dimension model, showing all the cardinality?

Star-schema for given snowflake schema is given below:



2.

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



```
Select SalesAmount, S.OrderDate, S.Shippingdate
```

```
FROM DATE D, Sales S,
```

```
WHERE S.OrderDate=D.Date
```

```
AND S.ShippingDate=D.Date;
```

A discrepancy arises in the above query as both OrderDate and ShippingDate should be identical to Date in Date table.

A solution to this is aliasing :

```
Select SalesAmount, S.OrderDate, S.Shippingdate
```

```
FROM DATE D, Sales S,DATE D1
```

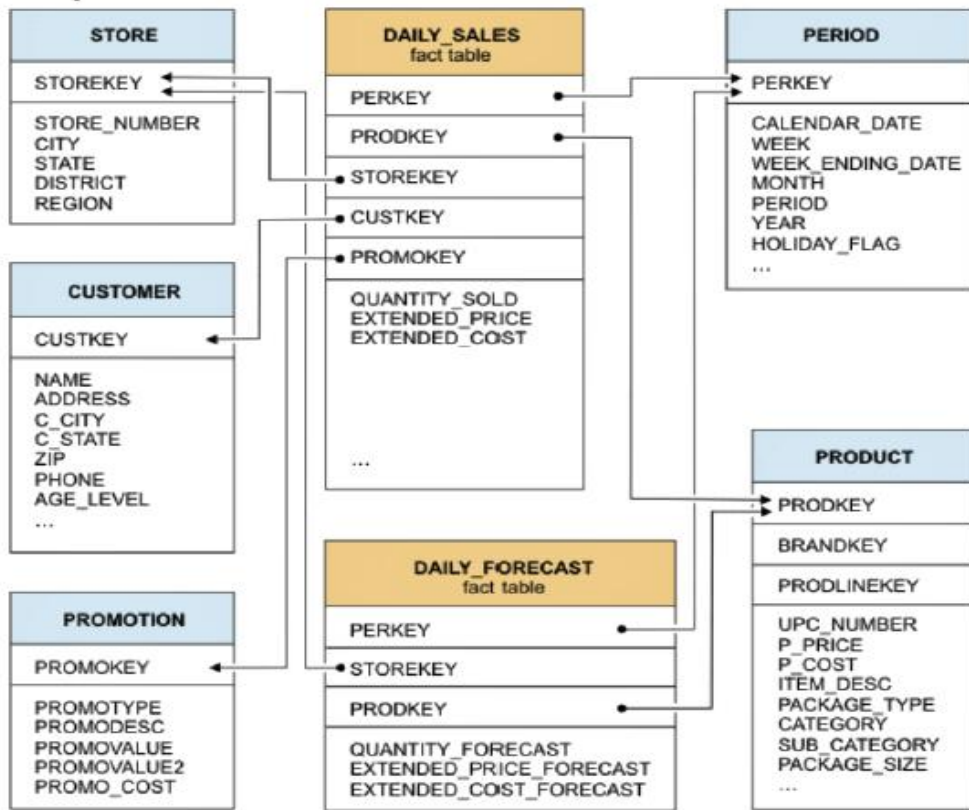
```
WHERE S.OrderDate=D.Date
```

```
AND S.ShippingDate=D1.Date;
```

=====

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:



```

SELECT ABS((
SELECT SUM(QUANTITY_SOLD)
FROM DAILY_SALES DS,PERIOD
WHERE DS.PERKEY IN(SELECT PERKEY FROM PERIOD WHERE
PERIOD.MONTH=SELECT MONTH(CURDATE())) -
(SELECT SUM(QUANTITY_FORECAST)
FROM DAILY_FORECAST DF,PERIOD

```

WHERE DF.PERKEY IN(SELECT PERKEY FROM PERIOD WHERE
 PERIOD.MONTH=SELECT MONTH(CURDATE())) AS DIVERGENCE
 FROM DAILY_SALES, DAILY_FORECAST, PERIOD;

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: Promotion, Customer.

<u>Measure</u>	<u>Measure-type</u>
QUANTITY_SOLD	Additive
EXTENDED_PRICE	Semi-Additive
EXTENDED_COST	Additive
QUANTITY_FORECAST	Additive
EXTENDED_PRICE_FORECAST	Semi-Additive
EXTENDED_COST_FORECAST	Additive

5.

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

Data-warehouse

- i. 100 GB to 2TB data size.

OLTP

- i. 10MB to 100GB data size.

li. Designed for running the business.

ii. Designed for analyzing the business.

iii. Query processing.

iii. Transaction processing.

=====

1.

a) Category of a product may change over a period of time. Historical category information (current category as well as all old categories) has to be stored. Which SCD type will be suitable to implement this requirement? What kind of structure changes are required in a dimension table to implement SCD type 2 and type 3.

SCD type-2 will be the most appropriate as history is maintained for all changes an attribute is undergoing.

The following structural changes in dimension table are needed to implement SCD type-2:

Add a new dimension table row with the new value of the changed attribute.

An effective date field may be included in the dimension table.

There are no changes to the original row in the dimension table.

The key of the original row is not affected.

SCD type-3 is needed when limited or immediate changes in attributes information in maintained. No information on "when" of changes is available.

The following changes in dimension table are needed to implement SCD type-3:

Add an "previous" field in the dimension table for the affected attribute.

Push down the existing value of the attribute from the "current" field to the "previous" field.

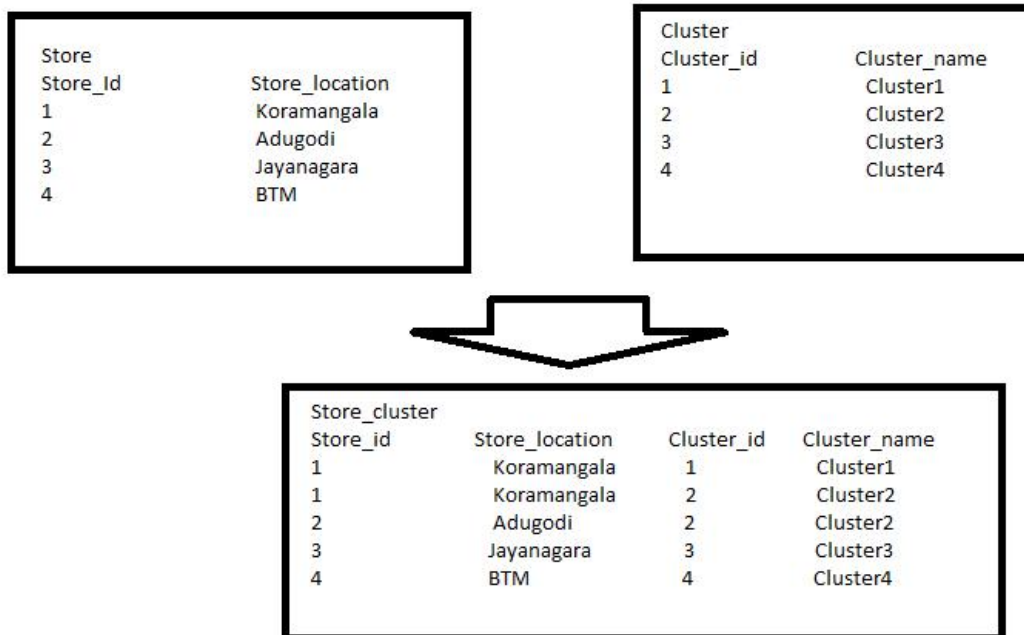
Keep the new value of the attribute in the “current” field.
You may add a “current” effective date field for the attribute.

b)What is surrogate key? Why it is required?

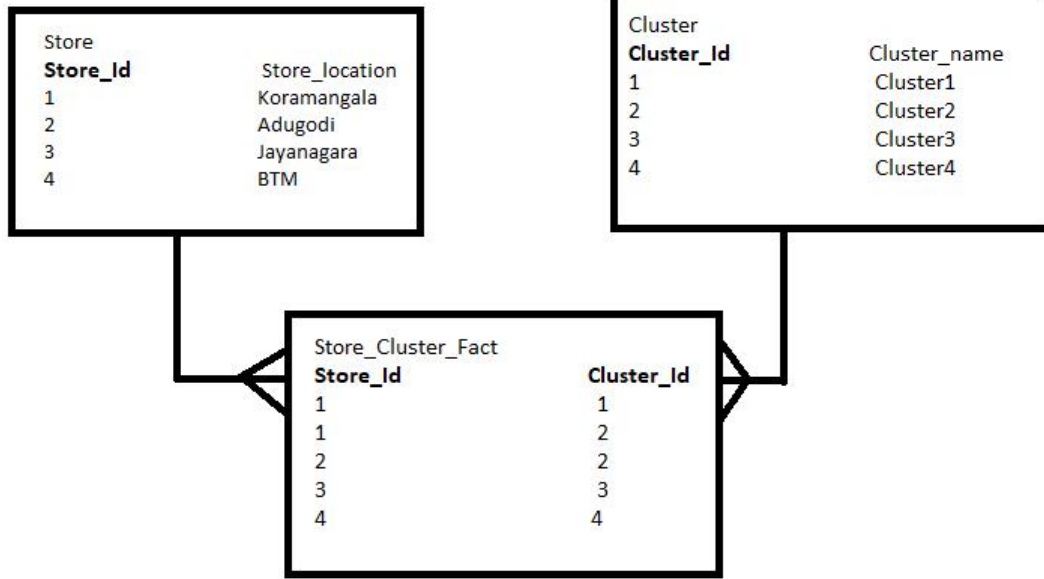
A surrogate key in a data warehouse is more than just a substitute for a natural key. In a data warehouse, a surrogate key is a necessary generalization of the natural production key and is one of the basic elements of data warehouse design.

c) Stores are grouped in to multiple clusters. A store can be part of one or more clusters. Design tables to store this store-cluster mapping information.

The given scenario can be described as follows:



Star-schema for store and cluster mapping:



d) What is a semi-additive measure? Give an example.

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.