**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

**Sales Facts**

DATEKey(FK)

CustomerKey(FK)

StoreKey(FK)

ProductKey(FK)

----------------------

Dollars

Product

ProductKey

(PK)

DATE

DateKey

(PK)

Customer

CustomerKey

(PK)

Store

StoreKey

(PK)

Here Month and y

**CUSTOMER**

CUSTOMER KEY

NAME

ADDRESS

**DATE**

DATE KEY

DAY

MONTH

YEAR

**SALESFAC**T

DATEKEY

CUSTOMERKEY

STORE KEY

PRODUCT KEY

-------------------------------------

DOLLARS

**STORE**

STOREKEY

STORE

**PRODUCTS**

PRODUCTSKEY

PRODUCT

SIZE

COLOR

* PRODUCT- SALES\_FACT ->ONE TO MANY
* STORE – SALES\_FACT -> ONE TO MANY
* CUSTOMER- SALES\_FACT -> ONE TO MANY
* DATE -SALES\_FACT ->ONE TO MANY

SALES\_FACT

SALES\_FACT

SALES\_FACT

SALES\_FACT

CUSTOMER

TIME

STORE

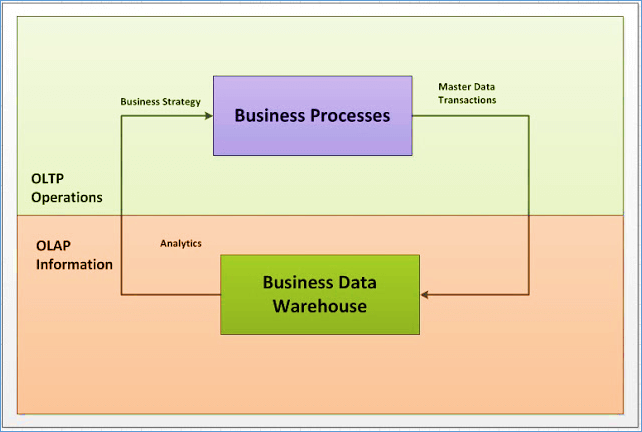
PRODUCT

1 N

1 N

1 N

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

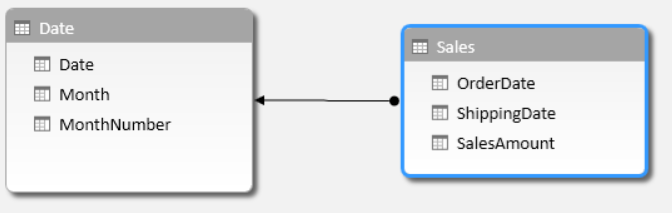


|  |  |  |
| --- | --- | --- |
|  | OLTP | DW |
| size | Less in size in MB | More in size compared to OLTP in TB |
| usage | Designed for real time business operations. | Designed for analysis of business measures by category and attributes. |
| processing | It provides fast result for daily used data.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

1. **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

1. 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

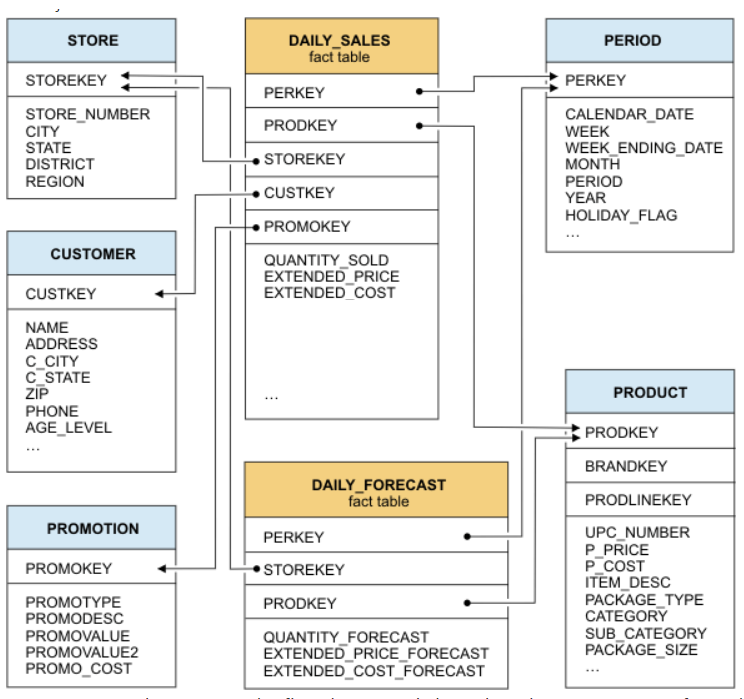
|  |  |  |
| --- | --- | --- |
| DATE | MONTH | MONTHNO |
| D1 | M1 | 1 |
| D2 | M2 | 2 |
| D3 | M3 | 3 |

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;