

ASSIGNMENT 3: To build the data warehouse for X-Mart

DATE : 26-09-2020

SUBMITTED BY : DWDM20G05

Problem Statement :

X-Mart is having different malls in city, where daily sales take place for various products. Higher management is facing an issue while decision making due to non availability of integrated data they can't do study on their data as per their requirement. So objective is to design a system which can help them quickly in decision making and provide Return on Investment (ROI).

IDENTIFICATION AND COLLECTION OF REQUIREMENTS

- ❖ We need to interview the key decision makers to know, what factors define the success in the business? How does management want to analyze their data? What are the most important business questions, which need to be satisfied by this new system?
- ❖ We also need to work with persons in different departments to know the data and their common relations if any, document their entire requirement which need to be satisfied by this system.
- ❖ Let us first identify the requirement from management about their requirements.
- ❖ Need to see daily, weekly, monthly, quarterly profit of each store. Comparison of sales and profit on various time periods.
- ❖ Comparison of sales in various time bands of the day.
- ❖ Need to know which product has more demand on which location?
- ❖ Need to study trend of sales by time period of the day over the week, month, and year?
- ❖ On what day sales is higher?
- ❖ On every Sunday of this month, what is sales and what is profit? What is trend of sales on weekday and weekend?
- ❖ Need to compare weekly, monthly and yearly sales to know growth and KPI

DESIGN THE DIMENSIONAL MODEL

We need to design Dimensional Model to suit requirements of users which must address business needs and contains information which can be easily accessible. Design of model should be easily extensible according to future needs. This model design must supports OLAP cubes to provide "instantaneous" query results for analysts.

Let us take a quick look at a few new terms and then we will identify/derive it for our requirement.

DIMENSION

The dimension is a master table composed of individual, non-overlapping data elements. The primary functions of dimensions are to provide filtering, grouping and labeling on your data. Dimension tables contain textual descriptions about the subjects of the business.

Let me give you a glimpse on different types of dimensions available like confirmed dimension, Role Playing dimension, Degenerated dimension, Junk Dimension.

Slowly changing dimension (SCD) specifies the way using which you are storing values of your dimension which is changing over a time and preserver

the history. Different methods / types are available to store the history of this change **E.g.** SCD1, SCD2, and SCD3 you can use as per your requirement.

Let us identify dimensions related to the above case study. Product, Customer, Store, Date, Time, Sales person Measure

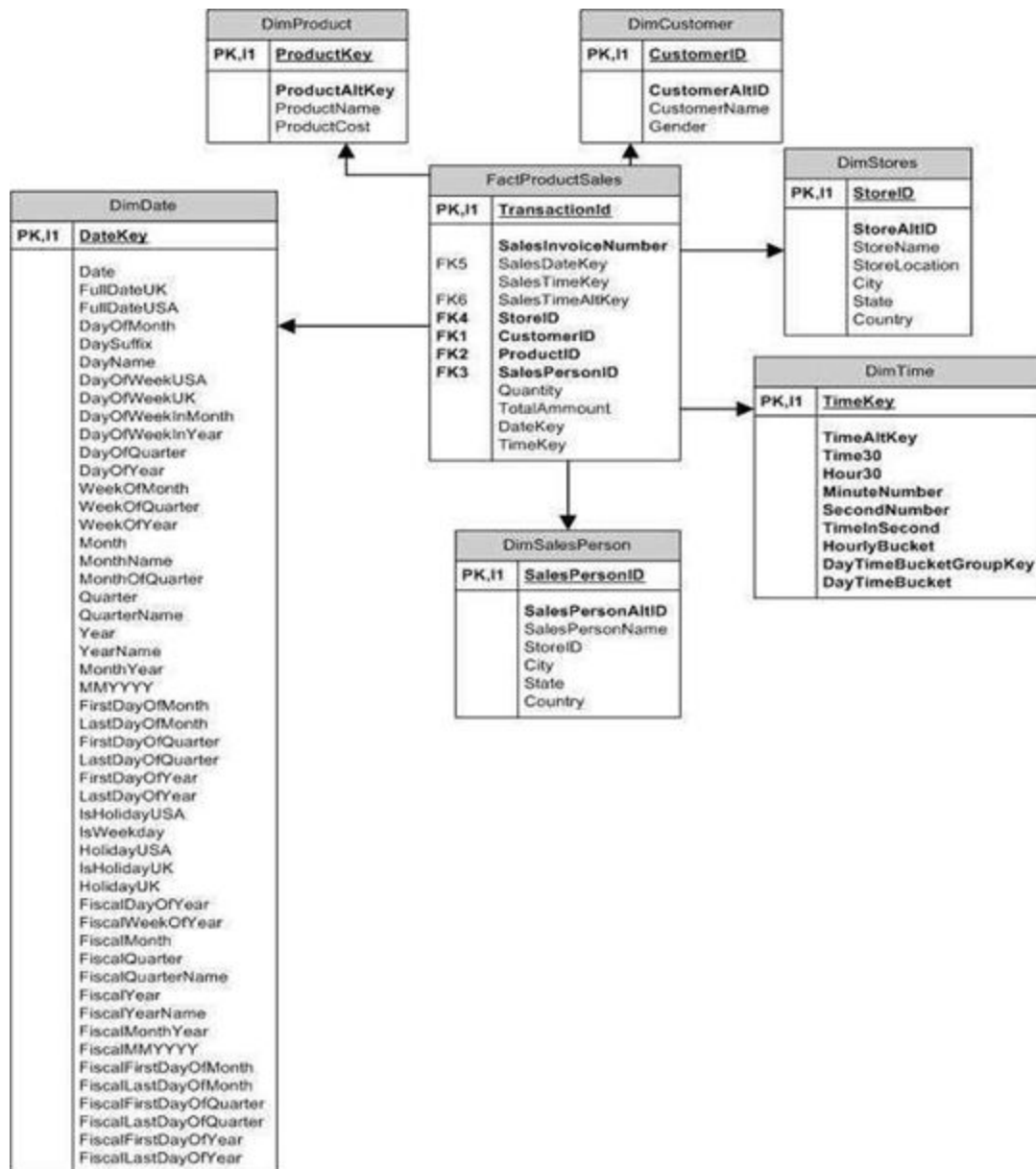
A measure represents a column that contains quantifiable data, usually numeric, that can be aggregated. A measure is generally mapped to a column in a fact table. For your information, various types of measures are there. E.g. Additive, semi additive and Non additive. Let us define what will be the Measures in our case. Actual Cost, Total Sales, Quantity, Fact table record count

FACT TABLE

Data in fact table are called measures (or dependent attributes), Fact table provides statistics for sales broken down by customer, salesperson, product, period and store dimensions. Fact table usually contains historical transactional entries of your live system, it is mainly made up of Foreign key column which references to various dimension and numeric measure values on which aggregation will be performed. Fact tables are of different types, **E.g.** Transactional, Cumulative and Snapshot. Let us identify what attributes should be there in our Fact Sales Table.

FOREIGN KEY COLUMN: Sales Date key, Sales Time key, Invoice Number, Sales Person ID, Store ID, Customer ID

MEASURES: Actual Cost, Total Sales, Quantity, Fact table record count



Create table DimCustomer
(
CustomerID int primary key,
CustomerAltID varchar(10) not null,
CustomerName varchar(50),
Gender varchar(20)
);

Create table DimProduct

```
(  
ProductKey int primary key,  
ProductAltKey varchar(10)not null,  
ProductName varchar(100),  
ProductActualCost real,  
ProductSalesCost real  
);
```

Create table DimStores

```
(  
StoreID int primary key,  
StoreAltID varchar(10)not null,  
StoreName varchar(100),  
StoreLocation varchar(100),  
City varchar(100),  
State varchar(100),  
Country varchar(100)  
);
```

Create table DimSalesPerson

```
(  
SalesPersonID int primary key,  
SalesPersonAltID varchar(10)not null,  
SalesPersonName varchar(100),  
StoreID int,  
City varchar(100),  
State varchar(100),  
Country varchar(100)  
);
```

```
CREATE TABLE DimTime(  
TimeKey int NOT NULL,
```

TimeAltKey int NOT NULL,
Time30 varchar(8) NOT NULL,
Hour30 int NOT NULL,
MinuteNumber int NOT NULL,
SecondNumber int NOT NULL,
TimeInSecond int NOT NULL,
HourlyBucket varchar(15)not null,
DayTimeBucketGroupKey int not null,
DayTimeBucket varchar(100) not null
);

Create Table FactProductSales

(
TransactionId int primary key,
SalesInvoiceNumber int,
SalesDateKey int,
SalesTimeKey int,
SalesTimeAltKey int,
Quantity int,
SalesTotalCost int,
ProductActualCost int,
Deviation int,
StoreID int,
CustomerID int,
ProductID int,
SalesPersonID int,
FOREIGN KEY (StoreID)REFERENCES DimStores(StoreID),
FOREIGN KEY (CustomerID)REFERENCES Dimcustomer(CustomerID),
FOREIGN KEY (ProductID)REFERENCES Dimproduct(ProductKey),
FOREIGN KEY (SalesPersonID)REFERENCES Dimsalesperson(SalesPersonID)
);

OBSERVATIONS:

DimCustomer relation:

Insert into

```
DimCustomer(CustomerID,CustomerAltID,CustomerName,Gender)values(1,'IMI-001','Aniket Modak','M');
```

Insert into

```
DimCustomer(CustomerID,CustomerAltID,CustomerName,Gender)values(2,'IMI-002','Harshraj Dhote','M');
```

Insert into

```
DimCustomer(CustomerID,CustomerAltID,CustomerName,Gender)values(3,'IMI-003','Neha Jawanjal','F');
```

Insert into

```
DimCustomer(CustomerID,CustomerAltID,CustomerName,Gender)values(4,'IMI-004','Shridhar Patil','M');
```

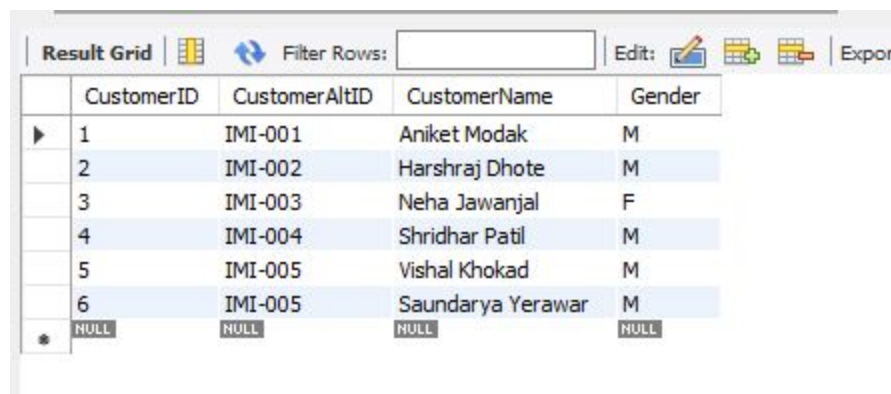
Insert into

```
DimCustomer(CustomerID,CustomerAltID,CustomerName,Gender)values(6,'IMI-005','Saundarya Yerawar','M');
```

Insert into

```
DimCustomer(CustomerID,CustomerAltID,CustomerName,Gender)values(5,'IMI-005','Vishal Khokad','M');
```

```
select * from DimCustomer;
```



The screenshot shows a database result grid with the following data:

	CustomerID	CustomerAltID	CustomerName	Gender
▶	1	IMI-001	Aniket Modak	M
	2	IMI-002	Harshraj Dhote	M
	3	IMI-003	Neha Jawanjal	F
	4	IMI-004	Shridhar Patil	M
	5	IMI-005	Vishal Khokad	M
	6	IMI-005	Saundarya Yerawar	M
*	NULL	NULL	NULL	NULL

DimProduct relation:

```
Insert into DimProduct(ProductKey,ProductAltKey,ProductName, ProductActualCost, ProductSalesCost)values(101,'ITM-001','Idli Instant Mix 1kg',5.50,6.50);
```

```
Insert into DimProduct(ProductKey,ProductAltKey,ProductName, ProductActualCost, ProductSalesCost)values(102,'ITM-002','Gram 1kg',22.50,24);
```

```
Insert into DimProduct(ProductKey,ProductAltKey,Product Name, ProductActualCost, ProductSalesCost)values(103,'ITM-003','Redlabel Tea 1 kg',42,43.5);
```

```
Insert into DimProduct(ProductKey,ProductAltKey,ProductName, ProductActualCost, ProductSalesCost)values(104,'ITM-004','Cinthol Soap',18,20);
```

Insert into DimProduct(ProductKey,ProductAltKey,ProductName, ProductActualCost, ProductSalesCost)values(105,'ITM-005','SurfXL Washing Powder 1kg',135,139);
select * from DimProduct;

ProductKey	ProductAltKey	ProductName	ProductActualCost	ProductSalesCost
101	ITM-001	Idli Instant Mix 1kg	5.5	6.5
102	ITM-002	Gram 1kg	22.5	24
103	ITM-003	Redlabel Tea 1 kg	42	43.5
104	ITM-004	Cinthol Soap	18	20
105	ITM-005	SurfXL Washing Powder 1kg	135	139
NULL	NULL	NULL	NULL	NULL

DimStores relation:

Insert into DimStores(StoreID,StoreAltID,StoreName,StoreLocation,City,State,Country)values(501,'LOC-A1','X-Mart','Hingana','Nagpur','Mah','India');
Insert into DimStores(StoreID,StoreAltID,StoreName,StoreLocation,City,State,Country)values(502,'LOC-A2','X-Mart','Shivaji Nagar','Pune','Mah','India');
Insert into DimStores(StoreID,StoreAltID,StoreName,StoreLocation,City,State,Country)values(503,'LOC-A3','X-Mart','Near Bus Stand','Sangli','Mah','India');
Insert into DimStores(StoreID,StoreAltID,StoreName,StoreLocation,City,State,Country)values(504,'LOC-A4','X-Mart','Nandura','Buldhana','Mah','India');
Insert into DimStores(StoreID,StoreAltID,StoreName,StoreLocation,City,State,Country)values(505,'LOC-A4','X-Mart','Malkapur','Buldhana','Mah','India');
select * from DimStores;

StoreID	StoreAltID	StoreName	StoreLocation	City	State	Country
501	LOC-A1	X-Mart	Hingana	Nagpur	Mah	India
502	LOC-A2	X-Mart	Shivaji Nagar	Pune	Mah	India
503	LOC-A3	X-Mart	Near Bus Stand	Sangli	Mah	India
504	LOC-A4	X-Mart	Nandura	Buldhana	Mah	India
505	LOC-A4	X-Mart	Malkapur	Buldhana	Mah	India
NULL	NULL	NULL	NULL	NULL	NULL	NULL

DimSalesPerson relation:

Insert into
DimSalesPerson(SalesPersonID,SalesPersonAltID,SalesPersonName,StoreID,City,State,Country)values
(11,'SP-DMSPR1','dhananjay',1,'Nagpur','Mah','India');

Insert into

```
DimSalesPerson(SalesPersonID,SalesPersonAltID,SalesPersonName,StoreID,City,State,Country )values
```

```
(12,'SP-DMSPR2','poonam',1,'Nagpur','Mah','India');
```

Insert into

```
DimSalesPerson(SalesPersonID,SalesPersonAltID,SalesPersonName,StoreID,City,State,Country )values
```

```
(13,'SP-DMNGR1','Shreya',2,'Pune','Mah','India');
```

Insert into

```
DimSalesPerson(SalesPersonID,SalesPersonAltID,SalesPersonName,StoreID,City,State,Country )values
```

```
(14,'SP-DMNGR2','bharambe',2,'Sangli','Mah','India');
```

Insert into

```
DimSalesPerson(SalesPersonID,SalesPersonAltID,SalesPersonName,StoreID,City,State,Country )values
```

```
(15,'SP-DMSVR1','somnath',3,'Buldhana','Mah','India');
```

Insert into

```
DimSalesPerson(SalesPersonID,SalesPersonAltID,SalesPersonName,StoreID,City,State,Country )values
```

```
(16,'SP-DMSVR2','pranay',3,'Buldhana','Mah','India');
```

```
select * from DimSalesPerson;
```

Result Grid

Filter Rows:

Edit:

Export/Import:

Wrap Cell Content:

	SalesPersonID	SalesPersonAltID	SalesPersonName	StoreID	City	State	Country
▶	11	SP-DMSPR1	dhananjay	1	Nagpur	Mah	India
	12	SP-DMSPR2	poonam	1	Nagpur	Mah	India
	13	SP-DMNGR1	Shreya	2	Pune	Mah	India
	14	SP-DMNGR2	bharambe	2	Sangli	Mah	India
	15	SP-DMSVR1	somnath	3	Buldhana	Mah	India
	16	SP-DMSVR2	pranay	3	Buldhana	Mah	India
✱	NULL	NULL	NULL	NULL	NULL	NULL	NULL

DimTime relation:

```
INSERT INTO DimTime (TimeKey,TimeAltKey, Time30,Hour30,MinuteNumber,  
SecondNumber, TimeInSecond,HourlyBucket,  
DayTimeBucketGroupKey,DayTimeBucket)
```

```
values(1,30000,'3:00:00',3,00,00,10800,'3:00-3:59',1,'Early Morning(03:00 AM To 6:59  
AM)');
```

```
INSERT INTO DimTime (TimeKey,TimeAltKey, Time30,Hour30,MinuteNumber,  
SecondNumber, TimeInSecond,HourlyBucket,  
DayTimeBucketGroupKey,DayTimeBucket)
```


values(2,121000,'12:10:00',12,10,00,43800,'12:00-12:59',4,'Lunch (12:00 PM To 13:59 PM)');

Select * from DimTime;

Result Grid									
Filter Rows:									
Export: Wrap Cell Content:									
TimeKey	TimeAltKey	Time30	Hour30	MinuteNumber	SecondNumber	TimeInSeconds	HourlyBucket	DayTimeBucketGroupKey	DayTimeBucket
1	30000	3:00:00	3	0	0	10800	3:00-3:59	1	Early Morning(03:00 AM To 6:59 AM)
2	121000	12:10:00	12	10	0	43800	12:00-12:59	4	Lunch (12:00 PM To 13:59 PM)

FactProductSales relation:

INSERT INTO FactProductSales values(1,2,4,1,10000,1,1,1,1,501,1,101,11);

Select * from FactProductSales;

Result Grid												
Filter Rows:												
Edit: Export/Import: Wrap Cell Content:												
TransactionId	SalesInvoiceNumber	SalesDateKey	SalesTimeKey	SalesTimeAltKey	Quantity	SalesTotalCost	ProductActualCost	Deviation	StoreID	CustomerID	ProductID	SalesPersonID
1	2	4	1	10000	1	1	1	1	501	1	101	11
NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL

CONCLUSION:

Hence we have successfully implemented a data warehouse for Xmart using star schema.

