# TY B.Tech. (CSE) – II [ 2022-23 ]

**5CS372 : Advanced Database System Lab.**

**Assignment No. 7**

**2020BTECS00085, MANJIRI CHANDURE**

**Build the data warehouse for X-Mart**

## Problem Statement :

1. 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).

# Activity :

## Identify and Collect 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](http://en.wikipedia.org/wiki/Data_element) [elements.](http://en.wikipedia.org/wiki/Data_element) 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 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 Design the Relational Database

We have done some basic workout to identify dimensions and measures, now we have to use appropriate schema to relate this dimension and Fact tables.

Few popular schemas used to develop dimensional model are as follows:

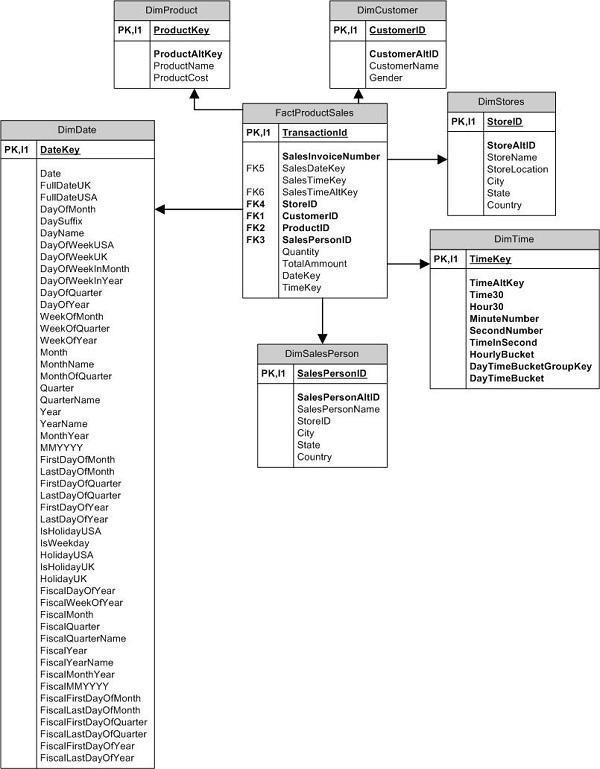
**E.g.** Star Schema, Snow Flake Schema, Star Flake Schema, Distributed Star

Schema, etc. In a different article, we will discuss all these schemas, dimension types, measure types,

etc., in detail.

First try to use Star schema due to hierarchical attribute model it provides for analysis and speedy performance in querying the data.

Star schema the diagram resembles a star, with points radiating from a center. The center of the star consists of fact table and the points of the star are the dimension tables.



**Design the Dimensional Model**

**Dimension tables:**

DimProduct (ProductKey, ProductAltKey, ProductName, ProductCost)

DimCustomer (CustomerID, CustomerAltID, CustomerName, Gender)

DimDate (DateKey, Date, FullDateUK, FullDateUSA, DayOfMonth, DaySuffix, DayName, DayOfWeekUSA, DayOfWeekUK, DayOfWeekInMonth, DayOfWeekInYear, DayOfQuarter, DayOfYear, WeekOfMonth, WeekOfQuarter, WeekOfYear, Month, MonthName, Quarter, QuarterName, Year, YearName, MonthOfQuarter, MonthYear)

DimStores (StorelD, StoreAltID, StoreName, StoreLocation, City, State, Country)

Note: The Time dimension is not included as it is not clear from the information provided whether it is necessary or redundant. It could be added if needed.

Fact table:

FactProductSales (Transactionld, SalesInvoiceNumber, SalesDateKey, SalesTimeKey, SalesTimeAltKey, StorelD, CustomerID, ProductID, SalesPersonID, Quantity, TotalAmount)

**Schema: Star Schema**

In this schema, the FactProductSales table is the center of the star, and the dimension tables are the points radiating from it. The relationship between the fact table and the dimension tables is based on the foreign keys. This schema provides a simple and fast way to query data for analysis.

SQL SCHEMA:

USE database assingmment\_7;

CREATE TABLE DimCustomer (

CustomerID INT PRIMARY KEY,

CustomerAltID VARCHAR(50),

CustomerName VARCHAR(100),

Gender CHAR(1)

);

CREATE TABLE DimDate (

DateKey INT PRIMARY KEY,

Date DATE,

FullDateUK VARCHAR(50),

FullDateUSA VARCHAR(50),

DayOfMonth INT,

DaySuffix VARCHAR(2),

DayName VARCHAR(20),

DayOfWeekUSA INT,

DayOfWeekUK INT,

DayOfWeekInMonth INT,

DayOfWeekInYear INT,

DayOfQuarter INT,

DayOfYear INT,

WeekOfMonth INT,

WeekOfQuarter INT,

WeekOfYear INT,

Month INT,

MonthName VARCHAR(20),

Quarter INT,

QuarterName VARCHAR(20),

Year INT,

YearName VARCHAR(20),

MonthOfQuarter INT,

MonthYear VARCHAR(6),

IsHolidayUSA INT,

IsWeekday INT,

HolidayUSA VARCHAR(50),

IsHolidayUK INT,

HolidayUK VARCHAR(50),

FiscalDayOfYear INT,

FiscalWeekOfYear INT,

FiscalMonth INT,

FiscalQuarter INT,

FiscalQuarterName VARCHAR(20),

FiscalYear INT,

FiscalYearName VARCHAR(20),

FiscalMonthYear VARCHAR(6),

FiscalMMYYYY VARCHAR(6),

FiscalFirstDayOfMonth DATE,

FiscalLastDayOfMonth DATE,

FiscalFirstDayOfQuarter DATE,

FiscalLastDayOfQuarter DATE,

FiscalFirstDayOfYear DATE,

FiscalLastDayOfYear DATE

);

CREATE TABLE DimStores (

StoreID INT PRIMARY KEY,

StoreAltID VARCHAR(50),

StoreName VARCHAR(100),

StoreLocation VARCHAR(100),

City VARCHAR(50),

State VARCHAR(50),

Country VARCHAR(50)

);

CREATE TABLE DimTime (

TimeKey INT PRIMARY KEY,

Time TIME,

TimeAltKey VARCHAR(50),

Hour30 VARCHAR(50),

MinuteNumber INT,

SecondNumber INT,

TimeInSecond INT,

HourlyBucket VARCHAR(50),

DayTimeBucketGroupKey VARCHAR(50),

DayTimeBucket VARCHAR(50),

MMYYYY VARCHAR(6),

FirstDayOfMonth DATE,

LastDayOfMonth DATE,

FirstDayOfQuarter DATE,

LastDayOfQuarter DATE,

FirstDayOfYear DATE,

LastDayOfYear DATE,

IsHolidayUSA INT,

IsWeekday INT,

HolidayUSA VARCHAR(50),

IsHolidayUK INT,

HolidayUK VARCHAR(50),

FiscalDayOfYear INT,

FiscalWeekOfYear INT,

FiscalMonth INT,

FiscalQuarter INT,

FiscalQuarterName VARCHAR(20),

FiscalYear INT,

FiscalYearName VARCHAR(20),

FiscalMonthYear VARCHAR(6),

FiscalMMYYYY VARCHAR(6),

FiscalFirstDayOfMonth DATE,

FiscalLastDayOfMonth DATE,

FiscalFirstDayOfQuarter DATE,

FiscalLastDayOfQuarter DATE,

FiscalFirstDayOfYear DATE,

FiscalLastDayOfYear DATE

);

CREATE TABLE FactProductSales (

TransactionID INT PRIMARY KEY,

SalesInvoiceNumber INT,

SalesDateKey INT,

SalesTimeKey INT,

SalesTimeAltKey VARCHAR(50),

StoreID INT,

CustomerID INT,

ProductID INT,

SalesPersonID INT,

Quantity INT,

TotalAmount DECIMAL(10, 2)

);

CREATE TABLE DimSalesPerson (

SalesPersonKey INT(11) NOT NULL AUTO\_INCREMENT,

SalesPersonAltKey VARCHAR(50),

SalesPersonName VARCHAR(100) NOT NULL,

StoreID INT(11) NOT NULL,

City VARCHAR(50),

PRIMARY KEY (SalesPersonKey),

FOREIGN KEY (StoreID) REFERENCES DimStores(StoreID)

);

**Conclusion:**

In conclusion, the design process for a data warehouse involves several stages, including identifying business requirements, defining a data model, designing data structures, and implementing the system. This process is crucial to ensure that the data warehouse meets the organization's needs and provides useful insights for decision-making.

The data model we have designed includes four dimensions (time, store, salesperson, and product) and one fact table (fact\_product\_sales). These tables provide a framework for storing and analyzing data related to sales transactions. The implementation details involve creating the tables, populating them with data, and defining relationships between them.

Once the data warehouse is implemented, it is essential to perform various types of queries to extract insights from the data. These include daily, weekly, monthly, and quarterly profit analysis for each store, comparison of sales and profit on various time periods, identification of products with high demand in different locations, trend analysis of sales by time period, identification of days with the highest sales, sales and profit analysis on every Sunday of the month, trend analysis of sales on weekdays and weekends, and comparison of weekly, monthly, and yearly sales to determine growth and KPIs.

To execute these queries, we can use SQL or a business intelligence tool that supports OLAP (Online Analytical Processing) operations. By leveraging the data warehouse's capabilities, organizations can gain valuable insights into their sales data and make informed decisions to improve their business operations.