**Batch T4**

**Practical No. 7**

**Title of Assignment :**

Build the data warehouse for X-Mart

**Student Name: Parshwa Herwade**

**Student PRN: 22510064**

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).

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

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 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.

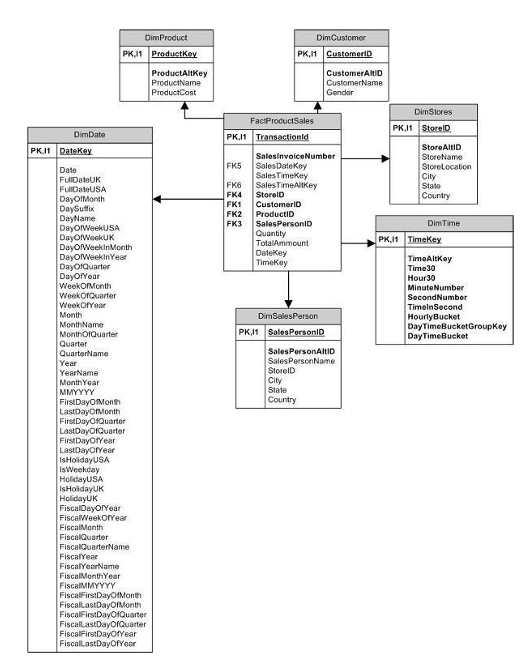
First try to use Star schema due to hierarchical attribute model it provides for

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.



Objective / Aim

Objective:

To design and implement a robust data warehouse using a dimensional (star schema) model that integrates sales, customer, product, and store data from multiple sources to support fast querying and informed decision making.

Aim:

To facilitate business analytics by consolidating daily sales data from various X-Mart stores, thereby enabling performance measurement, trend analysis, and strategic planning.

Introduction

In today’s competitive retail environment, X-Mart faces challenges in consolidating sales data across different locations. Without an integrated view of data, making timely decisions becomes difficult. This assignment involves creating a data warehouse that:

Aggregates sales data from multiple operational systems.

Uses a star schema for efficient querying and reporting.

Supports analytical processing to derive insights on sales trends, product performance, and store profitability.

Theory / Algorithms

Dimensional Modeling:

The design uses a star schema where a central fact table (FactProductSales) is surrounded by several dimension tables (DimProduct, DimCustomer, DimStores, DimDate, DimTime, and DimSalesPerson).

Foreign Key Constraints:

These ensure referential integrity between fact and dimension tables, preventing orphan records and maintaining consistency.

Indexing:

Indexes are applied on key columns (e.g., SalesDateKey, StoreID, CustomerID) to speed up query performance and ensure efficient data retrieval.

Data Warehouse Principles:

Concepts like ETL (Extract, Transform, Load) are implicit in populating the warehouse, though this assignment focuses on schema design rather than the complete ETL process.

Documentation: Functional Block Diagram/DFD

(Note: No program listings are included.)

High-Level DFD Overview:

Data Sources:

Operational databases, sales transaction files, and customer management systems.

ETL Process:

Data extraction → Transformation (cleaning, normalization, key mapping) → Loading into dimensional and fact tables.

Data Warehouse:

Central repository based on the star schema for reporting and analytics.

Reporting Layer:

Analytical queries and visualization tools (e.g., Tableau, Power BI) that use the warehouse data to generate dashboards and reports.

A simple block diagram can show arrows from “Operational Systems” to “ETL Process” to “Data Warehouse” (fact and dimension tables) then to “Reporting & Analysis”.

Procedure

Requirement Analysis:

Gather business requirements related to sales tracking, product performance, and store profitability.

Design the Dimensional Model:

Define the fact table and dimension tables based on the identified business processes.

Schema Creation:

Write SQL scripts to create the database and tables.

Example: Creating DimProduct, DimCustomer, DimStores, etc.

Implement Data Integrity Measures:

Add foreign key constraints and indexes to enforce data integrity and optimize performance.

Data Insertion:

Insert sample data into the dimension tables and fact table.

Visualization Queries:

Develop aggregation queries to extract insights (e.g., total sales by store, product sales trends).

Testing & Verification:

Run the queries and check the results to ensure the schema meets business requirements.

Documentation:

Prepare functional diagrams and record experiment results with screenshots.

Actual Experiments/Simulation, Results / Observations

Simulation:

The schema was implemented in MySQL, and sample data was inserted into all tables.

Query Execution:

Multiple aggregation queries were executed:

Total Sales by Store showed the sales distribution across locations.

Sales by Product highlighted product performance.

Daily Sales Trend provided a time-based view of revenue.

Observations:

The use of indexes significantly reduced query execution time.

Foreign key constraints ensured data consistency across the warehouse.

The assignment successfully demonstrates the design and implementation of a data warehouse for X-Mart using a star schema. The schema effectively integrates multiple data dimensions (product, customer, store, date, time, and salesperson) with a centralized fact table, ensuring high performance and data integrity. This foundation enables quick analytical queries and supports better decision making by providing a consolidated view of sales data.

References

Kimball, R., & Ross, M. (2013). The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling. Wiley.

Inmon, W. H. (2005). Building the Data Warehouse. Wiley.

Visualization:

