NAME:TUSHAR PATIL
PRN;2020BTECS00075 ADS:7

Build the data warehouse for X-Mart

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.

1. Daily, weekly, monthly, quarterly profit analysis for each store
2. Comparison of sales and profit on various time periods
3. Comparison of sales in different time bands of the day
4. Identification of products with high demand in different locations
5. Trend analysis of sales by time period (day/week/month/year)
6. Identification of day(s) with the highest sales
7. Sales and profit analysis on every Sunday of the month
8. Trend analysis of sales on weekdays and weekends
9. Comparison of weekly, monthly, and yearly sales to determine growth and KPIs

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 (TransactionId, 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.

Implementation in mysql
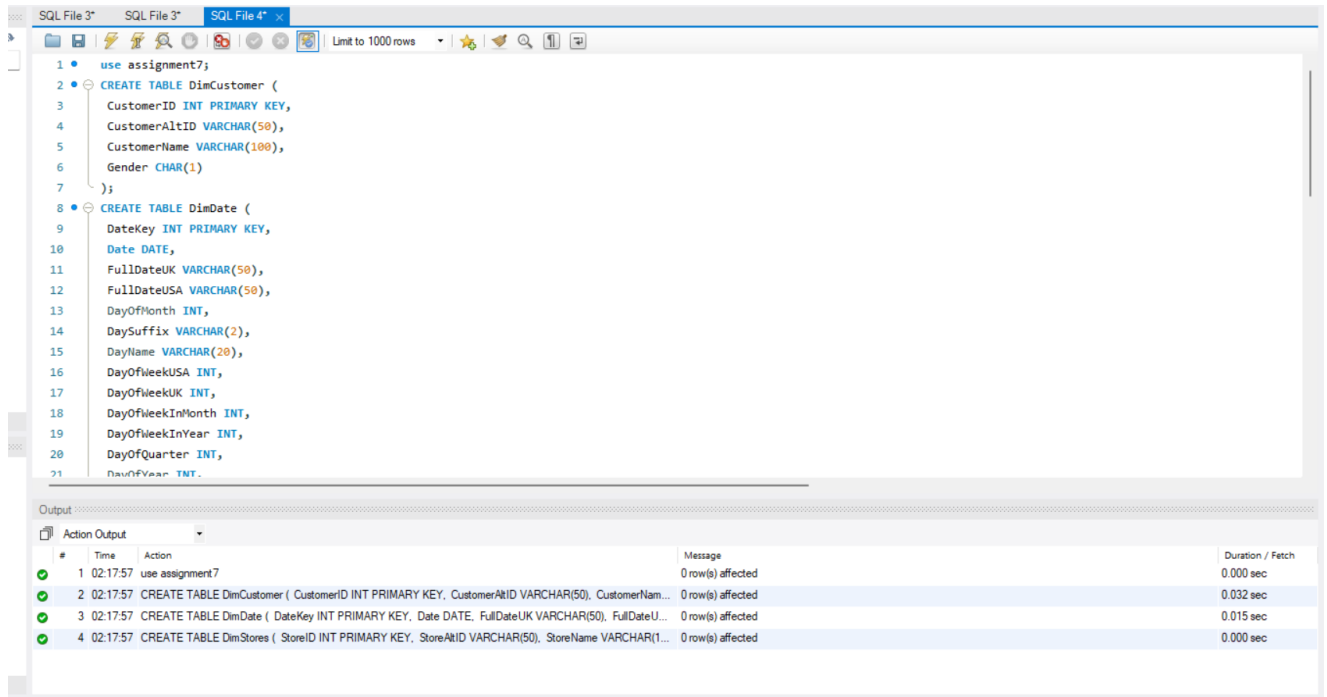
```
use assignment7;
CREATE TABLE DimCustomer (
 CustomerID INT PRIMARY KEY,
 CustomerAltID VARCHAR(50),
 CustomerName VARCHAR(100),
 Gender CHAR(1)
);
CREATE TABLE DimDate (
 DateKey INT PRIMARY KEY,
 Date DATE,
 FullDateUK VARCHAR(50),
 FullDateUSA VARCHAR(50),
 DayOfMonth INT,
 DaySuffix VARCHAR(2),
 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,
```

```sql
    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),
```

```sql
    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)
);
```

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.