T. Y. B. Tech Computer Engineering

2024-2025

Pursued in Department of Computer Engineering Faculty of Science & Technology

Vishwakarma University, Pune-411048

NAME	**Swanand Mangesh Garge**
YEAR	**THIRD YEAR**
DIV	**D**
BATCH	**D2**
ROLL NO	**39**
SRN NO	**202201589**
PRN NO	**2280030433**
COURSE NAME	**DATAWAREHOUSE & DATA MINING LAB**
COURSE CODE	**BTECCE22509**
COURSE TEACHER NAME	**PROF. RAHUL PAPALKAR**

ASSIGNMENT NO: 2

Problem Statement:

Consider an order management operational database that tracks order numbers, dates, the requested ship dates, customers and their shipping and billing addresses, products and their quantity and gross dollar amount, sales representatives that take and process orders, the deals (promotions) and discounts proposed/offered to customers.

You have to design a data warehouse that will be updated from the above operational database and should support decision making by helping to answer analytical questions about the net order dollar amounts per customer, products, promotions or deals, and the performance of their sales representatives or agents.

Analysis of requested ship dates is important for analysis as well. It is also important to allow for performing order amount analysis in various currencies: dollars, dirhams, euros.

1) Draw the star schema(s) showing the main attributes, including primary keys, foreign keys, and facts.

2) Insert appropriate values in the database. Write one SQL statement that runs on your schema and returns the net order dollar amount per customer, products, promotions, and performance of sales representatives.

3) Make necessary assumptions to compute an approximate size (in MB) of your DW over 5 years.

THEORY:

1. Star Schema:

- A type of database schema that organizes data into fact and dimension tables.
- The fact table is at the center, and it connects to dimension tables, which "radiate" out like the points of a star.
- It is simple, easy to understand, and supports efficient queries.

2. Fact Table:

- Stores quantitative data (measurable metrics) related to business processes, like sales, revenue, or performance metrics.
- Each record in a fact table is a combination of foreign keys to dimension tables and facts (numeric data).
- Typically, it has many rows and fewer columns.

3. Dimension Table:

- Contains descriptive attributes (context) about the facts, like product details, time, location, or customer information.
- Dimension tables are used to filter, group, and label the data in the fact table.
- They usually have fewer rows but more columns compared to fact tables.

CustomerID	CustomerName	ShippingAddress	BillingAddress
1	Alice	123 Elm St	456 Oak St
2	Bob	789 Pine St	101 Maple St
3	Charlie	654 Birch St	789 Cedar St
4	David	321 Spruce St	654 Pine St
5	Eva	987 Oak St	321 Maple St
6	Frank	456 Willow St	987 Elm St
7	Grace	789 Cedar St	321 Spruce St
NULL	NULL	NULL	NULL

Result Grid | Filter Rows: | Edit: | Export/Import:

ProductID	ProductName	ProductCategory
1	Laptop	Electronics
2	Smartphone	Electronics
3	Tablet	Electronics
4	Headphones	Accessories
5	Charger	Accessories
6	Monitor	Electronics
7	Keyboard	Accessories
NULL	NULL	NULL

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Content:

PromotionID	PromotionName	DiscountPercent
1	Summer Sale	10.00
2	Holiday Discount	15.00
3	New Year Offer	20.00
4	Festive Season	25.00
5	Clearance Sale	30.00
6	Black Friday	35.00
7	Cyber Monday	40.00
NULL	NULL	NULL

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Content:

SalesRepID	SalesRepName
1	John Doe
2	Jane Smith
3	Chris Evans
4	Natalie Portman
5	Tom Holland
6	Scarlett Johansson
7	Robert Downey Jr.
NULL	NULL

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Content:

CurrencyCode	CurrencyName	ConversionRateToDollar
AED	Dirham	0.2700
CAD	Canadian Dollar	0.7500
EUR	Euro	1.1000
GBP	Pound	1.2500
INR	Rupee	0.0130
JPY	Yen	0.0091
USD	Dollar	1.0000
NULL	NULL	NULL

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Content:

OrderID	CustomerID	ProductID	PromotionID	SalesRepID	OrderDate	RequestedShipDate	Quantity	GrossAmount	NetAmount	Currency
1	1	1	1	1	2024-08-01	2024-08-05	2	2000.00	1800.00	INR
2	2	2	2	2	2024-08-02	2024-08-06	1	1500.00	1275.00	INR
3	3	3	3	3	2024-08-03	2024-08-07	3	3000.00	2400.00	INR
4	4	4	4	4	2024-08-04	2024-08-08	4	1000.00	750.00	INR
5	5	5	5	5	2024-08-05	2024-08-09	5	500.00	350.00	INR
6	6	6	6	6	2024-08-06	2024-08-10	1	3000.00	1950.00	INR
7	7	7	7	7	2024-08-07	2024-08-11	2	4000.00	2400.00	INR
NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL	NULL

Based on the SQL code you provided, here are the entities (tables) with their attributes:

1. **Customers**
   - **Attributes:**
     - `CustomerID` (Primary Key)
     - `CustomerName`
     - `ShippingAddress`
     - `BillingAddress`

2. **Products**

- **Attributes:**
  - `ProductID` (Primary Key)
  - `ProductName`
  - `ProductCategory`

3. **Promotions**
- **Attributes:**
  - `PromotionID` (Primary Key)
  - `PromotionName`
  - `DiscountPercent`

4. **SalesRepresentatives**
- **Attributes:**
  - `SalesRepID` (Primary Key)
  - `SalesRepName`

5. **Currency**
- **Attributes:**
  - `CurrencyCode` (Primary Key)
  - `CurrencyName`
  - `ConversionRateToDollar`

6. **OrderFacts (Fact Table)**
- **Attributes:**
  - `OrderID` (Primary Key)
  - `CustomerID` (Foreign Key referencing `Customers.CustomerID`)
  - `ProductID` (Foreign Key referencing `Products.ProductID`)
  - `PromotionID` (Foreign Key referencing `Promotions.PromotionID`)
  - `SalesRepID` (Foreign Key referencing `SalesRepresentatives.SalesRepID`)
  - `OrderDate`
  - `RequestedShipDate`
  - `Quantity`
  - `GrossAmount`
  - `NetAmount`
  - `Currency` (Foreign Key referencing `Currency.CurrencyCode`)

Relationships:
- The `OrderFacts` table is a fact table and has many-to-one relationships with the `Customers`, `Products`, `Promotions`, `SalesRepresentatives`, and `Currency` tables.
- The `Customers`, `Products`, `Promotions`, `SalesRepresentatives`, and `Currency` tables are dimension tables that provide details about each dimension of the orders.

These entities together define the structure of your `Parth` database, capturing the details of customers, products, promotions, sales representatives, currencies, and orders.

SOURCE CODE:

```
-- Step 1: Create the database
CREATE DATABASE Parth;

-- Step 2: Use the database
USE Parth;

-- Step 3: Create Customers Table
CREATE TABLE Customers (
    CustomerID INT PRIMARY KEY,
    CustomerName VARCHAR(100),
    ShippingAddress VARCHAR(255),
    BillingAddress VARCHAR(255)
);

-- Step 4: Insert data into Customers
INSERT INTO Customers (CustomerID, CustomerName, ShippingAddress, BillingAddress)
VALUES (1, 'Alice', '123 Elm St', '456 Oak St'),
       (2, 'Bob', '789 Pine St', '101 Maple St'),
       (3, 'Charlie', '654 Birch St', '789 Cedar St'),
       (4, 'David', '321 Spruce St', '654 Pine St'),
       (5, 'Eva', '987 Oak St', '321 Maple St'),
       (6, 'Frank', '456 Willow St', '987 Elm St'),
       (7, 'Grace', '789 Cedar St', '321 Spruce St');

-- Step 5: Create Products Table
CREATE TABLE Products (
    ProductID INT PRIMARY KEY,
    ProductName VARCHAR(100),
    ProductCategory VARCHAR(100)
);

-- Step 6: Insert data into Products
INSERT INTO Products (ProductID, ProductName, ProductCategory)
VALUES (1, 'Laptop', 'Electronics'),
       (2, 'Smartphone', 'Electronics'),
       (3, 'Tablet', 'Electronics'),
       (4, 'Headphones', 'Accessories'),
       (5, 'Charger', 'Accessories'),
       (6, 'Monitor', 'Electronics'),
       (7, 'Keyboard', 'Accessories');

-- Step 7: Create Promotions Table
CREATE TABLE Promotions (
```

```
    PromotionID INT PRIMARY KEY,
    PromotionName VARCHAR(100),
    DiscountPercent DECIMAL(5, 2)
);

-- Step 8: Insert data into Promotions
INSERT INTO Promotions (PromotionID, PromotionName, DiscountPercent)
VALUES (1, 'Summer Sale', 10),
       (2, 'Holiday Discount', 15),
       (3, 'New Year Offer', 20),
       (4, 'Festive Season', 25),
       (5, 'Clearance Sale', 30),
       (6, 'Black Friday', 35),
       (7, 'Cyber Monday', 40);

-- Step 9: Create SalesRepresentatives Table
CREATE TABLE SalesRepresentatives (
    SalesRepID INT PRIMARY KEY,
    SalesRepName VARCHAR(100)
);

-- Step 10: Insert data into SalesRepresentatives
INSERT INTO SalesRepresentatives (SalesRepID, SalesRepName)
VALUES (1, 'John Doe'),
       (2, 'Jane Smith'),
       (3, 'Chris Evans'),
       (4, 'Natalie Portman'),
       (5, 'Tom Holland'),
       (6, 'Scarlett Johansson'),
       (7, 'Robert Downey Jr.');

-- Step 11: Create Currency Table
CREATE TABLE Currency (
    CurrencyCode CHAR(3) PRIMARY KEY,
    CurrencyName VARCHAR(50),
    ConversionRateToDollar DECIMAL(10, 4)
);

-- Step 12: Insert data into Currency
INSERT INTO Currency (CurrencyCode, CurrencyName, ConversionRateToDollar)
VALUES ('USD', 'Dollar', 1.0),
       ('AED', 'Dirham', 0.27),
       ('EUR', 'Euro', 1.1),
       ('INR', 'Rupee', 0.013),
       ('GBP', 'Pound', 1.25),
```

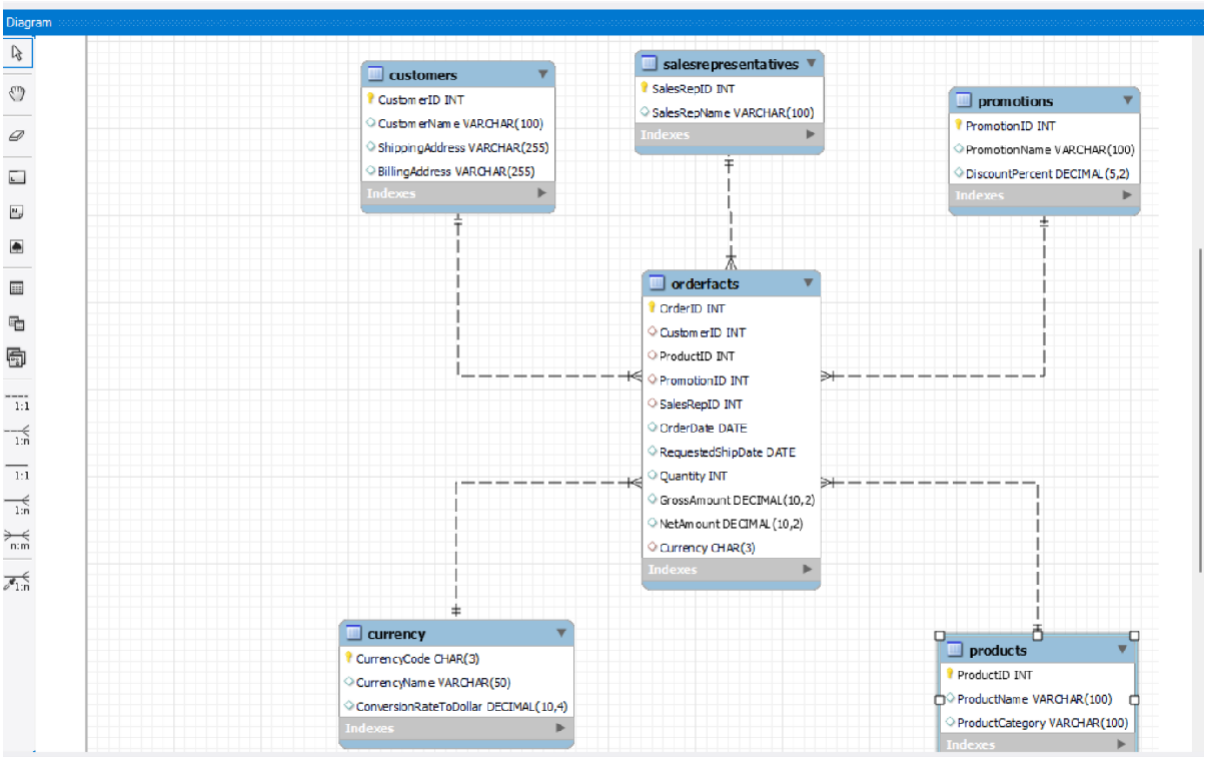
```
    ('CAD', 'Canadian Dollar', 0.75),
    ('JPY', 'Yen', 0.0091);

-- Step 13: Create OrderFacts Table
CREATE TABLE OrderFacts (
    OrderID INT PRIMARY KEY,
    CustomerID INT,
    ProductID INT,
    PromotionID INT,
    SalesRepID INT,
    OrderDate DATE,
    RequestedShipDate DATE,
    Quantity INT,
    GrossAmount DECIMAL(10, 2),
    NetAmount DECIMAL(10, 2),
    Currency CHAR(3),
    FOREIGN KEY (CustomerID) REFERENCES Customers(CustomerID),
    FOREIGN KEY (ProductID) REFERENCES Products(ProductID),
    FOREIGN KEY (PromotionID) REFERENCES Promotions(PromotionID),
    FOREIGN KEY (SalesRepID) REFERENCES SalesRepresentatives(SalesRepID),
    FOREIGN KEY (Currency) REFERENCES Currency(CurrencyCode)
);

-- Step 14: Insert data into OrderFacts
INSERT INTO OrderFacts (OrderID, CustomerID, ProductID, PromotionID, SalesRepID,
OrderDate, RequestedShipDate, Quantity, GrossAmount, NetAmount, Currency)
VALUES (1, 1, 1, 1, 1, '2024-08-01', '2024-08-05', 2, 2000, 1800, 'INR'),
    (2, 2, 2, 2, 2, '2024-08-02', '2024-08-06', 1, 1500, 1275, 'INR'),
    (3, 3, 3, 3, 3, '2024-08-03', '2024-08-07', 3, 3000, 2400, 'INR'),
    (4, 4, 4, 4, 4, '2024-08-04', '2024-08-08', 4, 1000, 750, 'INR'),
    (5, 5, 5, 5, 5, '2024-08-05', '2024-08-09', 5, 500, 350, 'INR'),
    (6, 6, 6, 6, 6, '2024-08-06', '2024-08-10', 1, 3000, 1950, 'INR'),
    (7, 7, 7, 7, 7, '2024-08-07', '2024-08-11', 2, 4000, 2400, 'INR');

-- Step 15: Display data from all tables
SELECT * FROM Customers;
SELECT * FROM Products;
SELECT * FROM Promotions;
SELECT * FROM SalesRepresentatives;
SELECT * FROM Currency;
SELECT * FROM OrderFacts;
```

OUTPUT:



CONCLUSION:

This assignment aimed to design a star schema tailored for a sales environment, with the core focus on the `Fact_Sales` table. This fact table serves as the central repository for essential sales metrics, while the dimension tables—`Dim_Sales_Rep`, `Dim_Time`, `Dim_Customer`, `Dim_Product`, and `Dim_Order`—offer detailed contextual information. The one-to-many relationships among these tables facilitate streamlined querying and comprehensive data analysis. The star schema's structure not only optimizes performance but also supports strategic decision-making, showcasing its efficacy in organizing and interpreting extensive datasets.