What is Data Warehouse?

A data warehouse is a centralized repository of integrated data from one or more disparate sources. It is designed to support business analytics and decision making. Key characteristics of a data warehouse include:

- **Subject-Oriented**: Organized around major subjects such as customer, product, sales, etc. rather than the organization's functional areas.
- **Integrated**: Data is gathered into the data warehouse from various sources and merged into a coherent whole.
- **Time-Variant**: Historical data is kept in the data warehouse, providing information from a historical perspective (e.g. daily sales for the past 5 years).
- **Non-Volatile**: Data is never over-written or deleted, new data is simply added.

Database and Database Management System

- **Database**: A structured collection of data that is organized and stored in a way that allows for efficient retrieval, manipulation, and management of information.
- **Database Management System (DBMS)**: Software that provides an interface between the database and the applications or users that access the data. Examples: MySQL, PostgreSQL, Oracle, SQL Server.

Data Model and Relational Model

- **Data Model**: An abstract model that describes how data is represented and organized. It defines the structure of the data, the relationships between data elements, and the rules that govern the data.
- **Relational Model**: A data model where data is stored in tables consisting of rows and columns. The key elements of the relational model are:
  - **Entity**: A real-world object or concept represented in the database.
  - **Attribute**: A property or characteristic of an entity.
  - **Relationship**: An association between two or more entities.
  - **Primary Key**: A unique identifier for each row in a table.
  - **Foreign Key**: An attribute in one table that refers to the primary key of another table, establishing a relationship.

SQL File

```sql
-- Create a sample database schema for a data warehouse

-- Create the Dimension Tables
CREATE TABLE Dim_Customer (
    CustomerID INT PRIMARY KEY,
    CustomerName VARCHAR(50),
    CustomerAddress VARCHAR(100),
    CustomerCity VARCHAR(50),
    CustomerState VARCHAR(50),
    CustomerCountry VARCHAR(50),
    CustomerZipCode VARCHAR(20)
);

CREATE TABLE Dim_Product (
    ProductID INT PRIMARY KEY,
    ProductName VARCHAR(50),
    ProductDescription VARCHAR(200),
    ProductCategory VARCHAR(50),
    ProductSubCategory VARCHAR(50),
    ProductPrice DECIMAL(10,2)
);

CREATE TABLE Dim_Time (
    TimeID INT PRIMARY KEY,
    DateValue DATE,
    Year INT,
    Quarter INT,
    Month INT,
    Day INT,
    DayOfWeek INT
);

-- Create the Fact Table
CREATE TABLE Fact_Sales (
    SalesID INT PRIMARY KEY,
    CustomerID INT FOREIGN KEY REFERENCES Dim_Customer(CustomerID),
    ProductID INT FOREIGN KEY REFERENCES Dim_Product(ProductID),
    TimeID INT FOREIGN KEY REFERENCES Dim_Time(TimeID),
    Quantity INT,
    Revenue DECIMAL(10,2)
);
```

This SQL file creates a simple data warehouse schema with three dimension tables (Customer, Product, and Time) and one fact table (Sales). The dimension tables store the

attributes of the key entities, while the fact table stores the actual sales transactions, linking to the dimension tables through foreign key relationships.