Step 3: Schema Design

1. Justification for Schema Choice

Given your data, a **Star Schema** is most appropriate. This is because:

- **Central Fact Table:** Your data is transactional (orders), so the fact table will store quantitative data (sales, profit, etc.).
- **Dimension Tables:** Entities like Customer, Product, Location, and Time can be separated into dimension tables, supporting efficient queries and analytics.
- **Benefits:** Star schemas are simple, denormalized for fast querying, and ideal for
- **Customer**
    1. CustomerID (PK)
    2. CustomerName
    3. Segment
- **Product**
    1. ProductID (PK)
    2. ProductName
    3. Category
    4. Sub_Category
- **Location**
    1. StateID (PK)
    2. State
    3. City
    4. Country
    5. PostalCode
    6. Region
    7. Latitude
    8. Longitude
- **Time**
    1. Date (PK)
    2. Day
    3. Month
    4. Year
    5. Quarter
- **Orders (Fact Table)**

    OrderDate DATE,

    ShipDate DATE,

    CustomerID VARCHAR REFERENCES Customer(CustomerID),

    ProductID VARCHAR REFERENCES Product(ProductID),

    StateID VARCHAR REFERENCES Location(StateID),

    Sales FLOAT,

    Profit FLOAT,

Quantity INT,

Discount FLOAT,

OrderID VARCHAR PRIMARY KEY,

ShipMode VARCHAR,

ShipStatus VARCHAR,

DaystoShipActual INT,

DaystoShipScheduled INT,

ShippingDelay INT,

SalesForecast FLOAT,

OrderProfitable FLOAT,

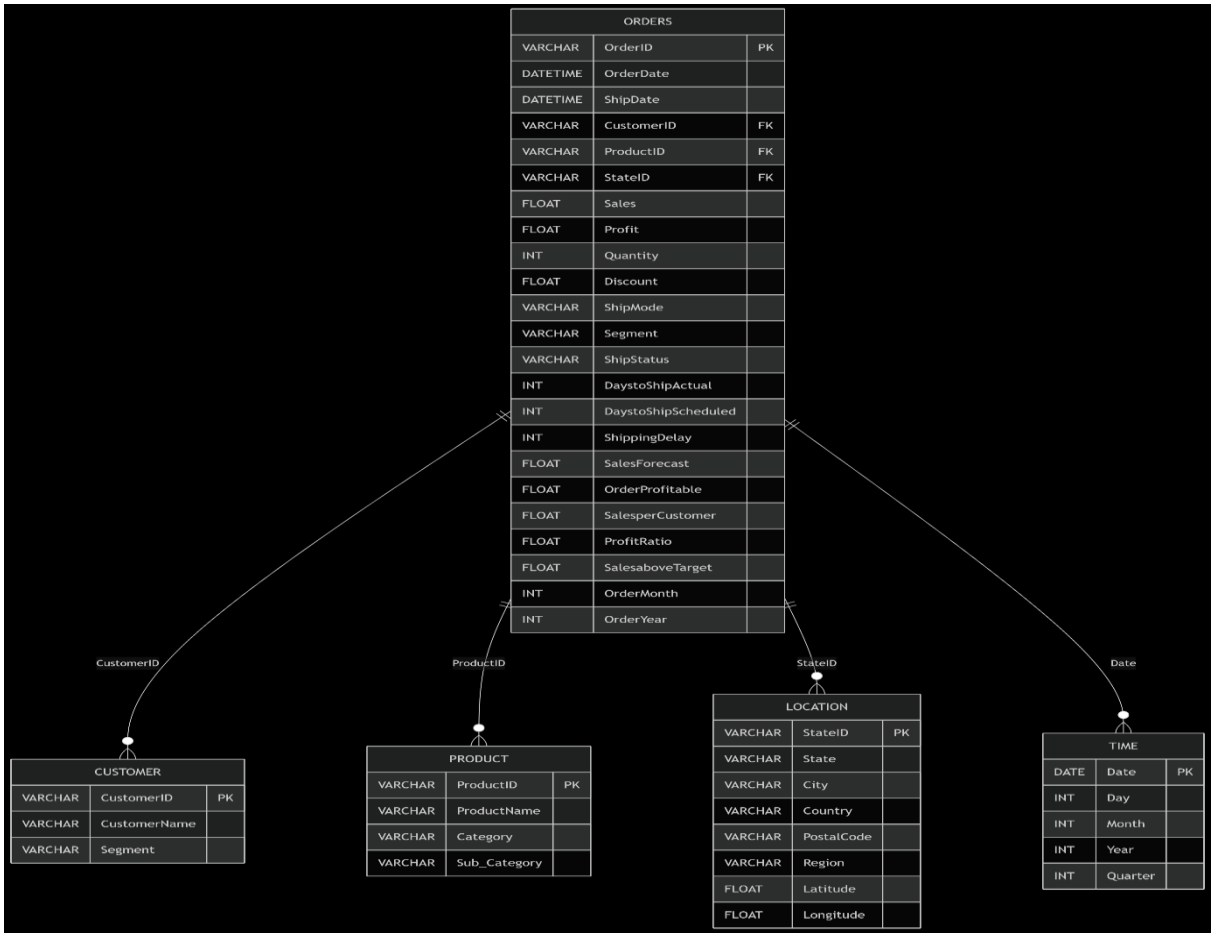
SalesperCustomer FLOAT,

ProfitRatio FLOAT,

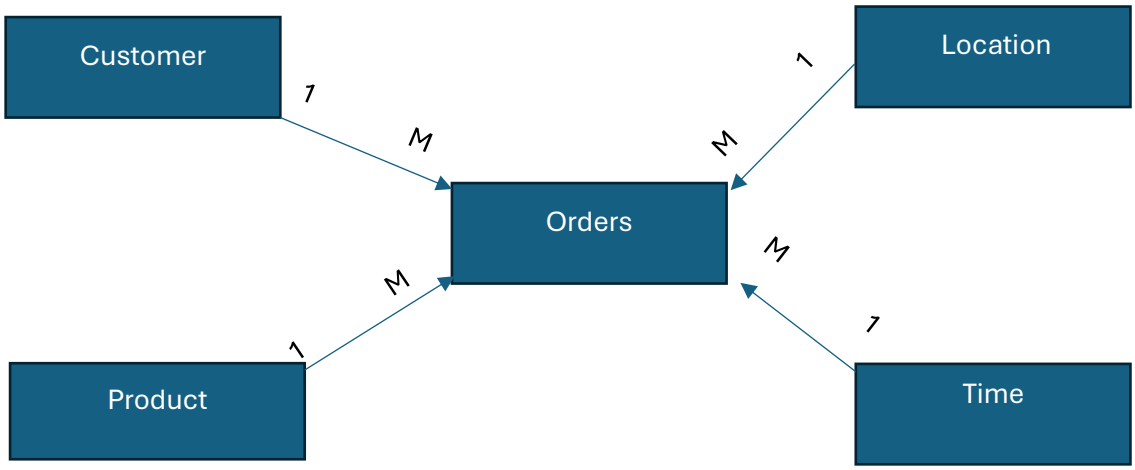
SalesaboveTarget FLOAT,

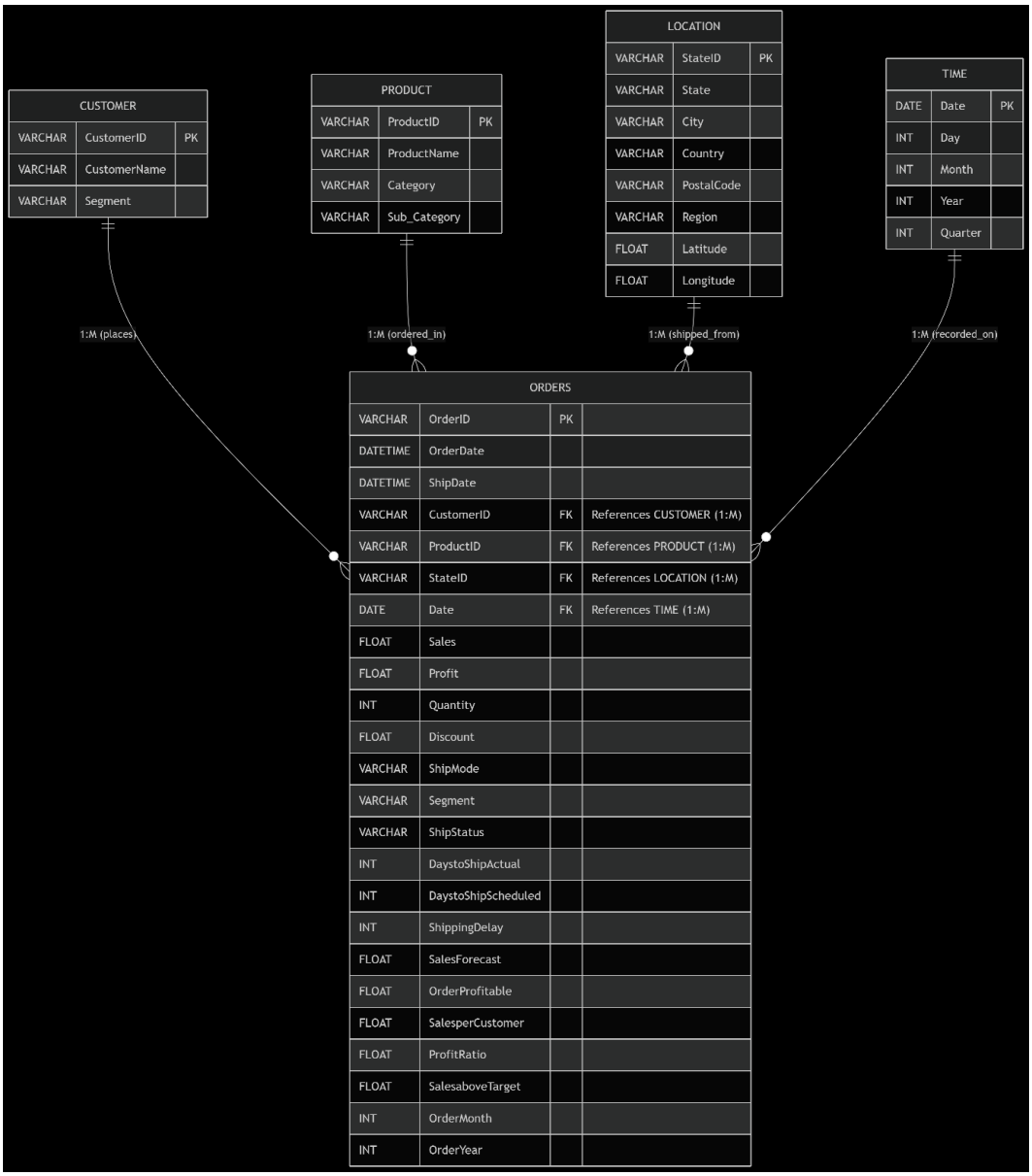
OrderMonth INT,

OrderYear INT



ER Diagram


Customer
Location
Orders
Product
Time
1
M
1
M
M
1
M
1


CUSTOMER
VARCHAR CustomerID PK
VARCHAR CustomerName
VARCHAR Segment
PRODUCT
VARCHAR ProductID PK
VARCHAR ProductName
VARCHAR Category
VARCHAR Sub_Category
LOCATION
VARCHAR StateID PK
VARCHAR State
VARCHAR City
VARCHAR Country
VARCHAR PostalCode
VARCHAR Region
FLOAT Latitude
FLOAT Longitude
TIME
DATE Date PK
INT Day
INT Month
INT Year
INT Quarter
1:M (places)
1:M (ordered_in)
1:M (shipped_from)
1:M (recorded_on)
ORDERS
VARCHAR OrderID PK
DATETIME OrderDate
DATETIME ShipDate
VARCHAR CustomerID FK References CUSTOMER (1:M)
VARCHAR ProductID FK References PRODUCT (1:M)
VARCHAR StateID FK References LOCATION (1:M)
DATE Date FK References TIME (1:M)
FLOAT Sales
FLOAT Profit
INT Quantity
FLOAT Discount
VARCHAR ShipMode
VARCHAR Segment
VARCHAR ShipStatus
INT DaystoShipActual
INT DaystoShipScheduled
INT ShippingDelay
FLOAT SalesForecast
FLOAT OrderProfitable
FLOAT SalesperCustomer
FLOAT ProfitRatio
FLOAT SalesaboveTarget
INT OrderMonth
INT OrderYear