Step 1: Fact vs. Dimension Identification

Attribute/Entity	Classification	Reasoning
Order Amount	Fact	A quantitative, numeric value that can be summed or averaged to measure business performance.
Quantity Sold	Fact	A measurable count of items. Used to calculate total volume and inventory turnover.
Customer	Dimension	Descriptive data used to group (segment) and filter (loyal vs. new) the sale facts.
Product	Dimension	Descriptive attributes (category, brand) used to slice sales data.
Date	Dimension	The most common dimension; allows for time-series analysis (YoY, MoM trends).
Region	Dimension	Geographic attributes used to identify high-performing territories.

Step 2: Fact Table Design (`fact_sales`)

The Grain: One row represents one individual line item in an order. *Justification:* This is the lowest possible level of detail. It allows us to roll up to the total order level or drill down into specific product performance.

Column	Type	Purpose
sale_id	PK	Surrogate key for the unique record.
customer_id	FK	Links to dim_customer.
product_id	FK	Links to dim_product.
date_id	FK	Links to dim_date.
region_id	FK	Links to dim_region.
order_amount	Measure	The gross revenue for this line item.
quantity	Measure	Number of units sold.

discount_amount	Measure	The dollar value subtracted from the price.
profit_amount	Measure	Calculated as (Price - Cost) * Quantity.

Step 3: Dimension Tables Design

`dim_customer`

- **Attributes:** `name`, `email`, `segment` (Enterprise/Retail), `city`, `state`.
- **SCD Strategy: SCD Type 2**. If a customer moves from "New York" to "Texas," we create a new row with a new surrogate key to preserve historical sales data for New York while attributing new sales to Texas.

`dim_product`

- **Attributes:** `product_name`, `category`, `subcategory`, `brand`, `unit_price`.
- **SCD Strategy: SCD Type 1** for corrections (fixing a typo in a name) and **Type 2** for price changes if historical margin analysis is required.

`dim_date`

- **Attributes:** `full_date`, `year`, `quarter`, `month_name`, `day_of_week`, `is_holiday`.
- **Purpose:** Pre-calculating these avoids expensive SQL functions (like `EXTRACT`) during runtime.

`dim_region`

- **Attributes:** `region_name` (North, South, East, West), `country`, `state`.

- **Purpose:** Allows the business to compare performance across different geographic hierarchies.

Step 4: Star Schema Diagram

In a Star Schema, the fact table sits in the center, directly connected to each dimension. This "single-hop" relationship is what makes queries so fast.