

Star Schema: 4-Step Process

Gabriela Crețu

June 2025

Step 1: Choose the Fact Table

The first step in building a star schema is selecting the appropriate fact table — the central focus of our analysis. In this case, we have two potential candidates: **sales** and **payment**.

Since **sales** provides a more comprehensive view for analytics, we choose it as our fact table. To ensure completeness, we incorporate the **amount** field (originally from **payment**) into the **sales** table.

The final structure of the fact table **SALES** includes:

- **sale_id** BIGINT (Primary Key)
- **store_id** (Foreign Key)
- **product_id** (Foreign Key)
- **employee_id** (Foreign Key)
- **customer_id** (Foreign Key)
- **sales_date** DATE
- **amount** NUMERIC(10,2)
- **update_dt** DATE

Step 2: Identify the Dimension Tables

Next, we identify the dimension tables. These are easily derived by analyzing the foreign keys in the **sales** table. Each foreign key typically points to a dimension entity.

The dimension tables are:

- **DIM_CUSTOMERS**
- **DIM_EMPLOYEES**
- **DIM_PRODUCTS**
- **DIM_STORES**

Step 3: Denormalize the Dimension Tables

Since we start with a normalized (3NF) schema, we need to denormalize our dimension tables to create a flat, analysis-friendly structure for the star schema.

For each dimension identified in Step 2, we must trace and incorporate data from any related tables (via foreign keys) into a single, flat table. This means including additional descriptive attributes such as:

- For DIM_EMPLOYEES: fields from related tables like **address**, **manager**, and **store**
- For DIM_CUSTOMERS: fields from **addresses**, **cities**, **countries**, and **regions**
- And similarly for other dimensions

This step ensures that each dimension contains all necessary descriptive data, without the need for further joins.

Step 4: Build the Star Schema

Finally, we bring everything together.

The central fact table, **FACT_SALES**, will reference each of the dimension tables through foreign keys. The complete star schema now includes:

- One fact table: **FACT_SALES**
- Four dimension tables: **DIM_CUSTOMERS**, **DIM_EMPLOYEES**, **DIM_PRODUCTS**, and **DIM_STORES**

Each component is fully denormalized and optimized for efficient querying and reporting.