Snowflake Schema: Normalization Steps

Gabriela Creţu

June 2025

Step 1: Analyze Existing Star Schema

The first step is to analyze the current star schema to determine which dimension tables need further normalization. Upon review, we find that all dimensions—including customers, employees, products, and stores—require normalization.

We then evaluate the repeating attributes, which are primarily address-related attributes found in the employees, customers, and stores tables.

In terms of hierarchies, the most prominent one is the location hierarchy, which spans from a specific address to a region or economic region. We also consider a new table, `Manager`, to represent hierarchical relationships among employees, as multiple employees can share the same manager. Additionally, multiple employees can work in the same store, and multiple entities share address attributes.

Step 2: Identify Hierarchies and Repeating Groups

We continue by identifying specific hierarchies and repeated data groups.

- **Location Hierarchy:** Address → City → Region → Country.
- **Employee Hierarchy:** Employee → Manager.
- **Store Relationship:** Store → Address.

These hierarchies are consistent across multiple dimensions (e.g., customers, employees, and stores), making them suitable candidates for normalization.

Step 3: Normalize Dimension Tables (Typically to 3NF)

Starting from the denormalized star schema, we normalize dimension tables by separating out hierarchical and repeating attribute groups.

We create an `Address` table used by customers, employees, managers, and stores. This address structure includes fields such as street, city, region, and country.

For the `Employee` table, we identify three separate sub-hierarchies:

- **Manager:** A manager can oversee multiple employees.

- **Store:** Multiple employees can work at the same store.
- **Address:** Shared address details are separated into the address hierarchy.

For the `Store` dimension, similar normalization applies: store address, associated managers, and related employees are separated into their respective normalized tables.

For `Customers` and `Products`:

- Customers: Normalized using the address hierarchy.
- Products: Normalized by extracting product category into a separate table.

Step 4: Replace Attributes with Foreign Keys

After normalizing the dimension tables, we replace the original attributes with foreign keys that reference the newly created tables. This includes:

- `Employee` now references `Manager`, `Store`, and `Address` via foreign keys.
- `Store` references `Address` and `Manager`.
- `Customer` references `Address`.
- `Product` references `ProductCategory`.

This completes the transformation into a snowflake schema. The new schema structure improves data consistency, reduces redundancy, and accurately reflects the hierarchical relationships within the data model.