

The Four-Step Dimensional Design Process

Building robust analytics systems that actually
work

Based on Ralph Kimball's Data Warehouse Toolkit

Milan Gabriel
Senior Data Engineer



Data Modelling 101 (1)

Operational Schemas (OLTP) Are Not Built for Analytics

Normalized structures cause join complexity and performance bottlenecks in reporting.

OLTP Focus

Optimized for fast writes, data integrity, and highly normalized 3rd Normal Form (3NF) structures.

The Analytical Challenge

Simple business questions require multiple table joins, making SQL unmaintainable for non-experts.

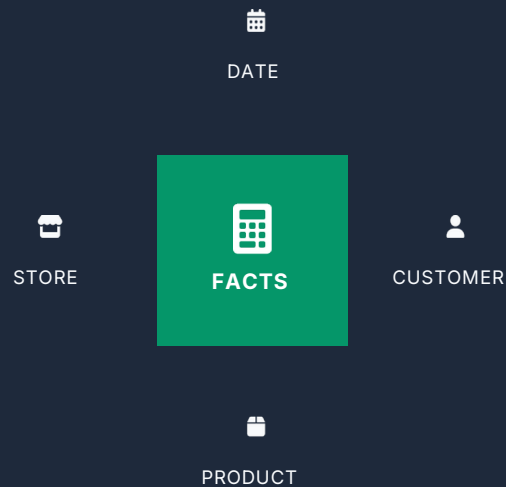
OPERATIONAL (OLTP)

- ✓ Efficient Transaction Processing
- ✓ Minimal Redundancy
- ✗ Complex Query Logic

THE RESULT

Metric Drift: Different departments reporting different numbers for the same KPI due to query complexity.





The Star Schema Optimizes for Read Performance and Clarity

Denormalizing data into facts and dimensions simplifies business logic.

Fact Tables

Central tables containing numeric measurements (e.g., Sales Amount) and foreign keys to dimensions.

Dimension Tables

Surrounding tables providing descriptive context (e.g., Product, Store) for filtering and grouping.

Key Benefit

Queries read like English and execute significantly faster than normalized equivalents.

Kimball's Four-Step Process Ensures Design Consistency

Following a sequential methodology prevents 90% of modeling failures.

ORDER MATTERS

The sequence is the framework. Each decision builds upon the previous one to ensure a robust, scalable model.

01

Select the Business Process

Identify the operational activity to be measured.

02

Declare the Grain

Define exactly what one row in the fact table represents.

03

Identify the Dimensions

Determine the context (Who, What, Where, When).

04

Identify the Facts

Select the numeric measurements that match the grain.

Focus on Measurable Activities, Not Departments

A business process is a verb-based event that generates metrics.

Definition

An operational activity that captures performance metrics, such as "Selling," "Shipping," or "Registering."

Why it Matters

Fact tables should mirror real-world events, not organizational silos, to ensure data remains relevant as structures change.

Key Characteristic

It is always a **VERB**. If you can't describe it as an action, it's likely not a business process.

Process vs. Department

Business Process

- ✓ Retail Sales
- ✓ Inventory Flow
- ✓ Order Fulfillment
- ✓ Claims Processing

Department

- ✗ Marketing
- ✗ Logistics
- ✗ Operations
- ✗ Insurance



Declaring the Grain is the Most Critical Design Decision

The grain determines the maximum level of detail and flexibility of the model.

THE GOLDEN RULE

You must be able to state the grain in one clear sentence:

"One row represents _____"

Example: *"One row per product sold in one transaction at one store."*

ATOMIC GRAIN

The lowest level of detail captured by the business process. Provides maximum flexibility to aggregate up to any level.

AGGREGATE GRAIN

Pre-summarized data (e.g., daily totals). Faster for specific queries but loses the ability to drill down into individual events.

⚠ The Risk of Mixed Grains

Mixing levels of detail in a single fact table leads to double-counting and fundamentally broken metrics.

Dimensions Provide the Descriptive Context for Analysis

Dimensions are the "Who, What, Where, and When" of the process.

Attributes

Wide, denormalized tables containing text-based descriptions (e.g., Brand, Category, Region) used for filtering and grouping.

Surrogate Keys

Warehouse-generated integers that serve as primary keys, protecting the model against changes in source system business keys.

Core Dimension Examples

DATE

Day, Month, Quarter, Year, Fiscal Period

PRODUCT

SKU, Name, Brand, Category, Size

STORE

Name, City, State, Region, Manager

PROMOTION

Type, Discount %, Start/End Dates



Facts are the Numeric Measurements of the Process

Every fact in a table must strictly adhere to the declared grain.

ADDITIVE FACTS

Can be summed across all dimensions (e.g., Sales Amount, Quantity).

SEMI-ADDITIVE FACTS

Can be summed across some dimensions, but not time (e.g., Inventory Levels).

NON-ADDITIVE FACTS

Cannot be summed; usually ratios or percentages (e.g., Unit Price, Margin %).

FACT TABLE STRUCTURE

Fact tables are "**Narrow and Deep**". They consist primarily of foreign keys to dimensions and numeric measurements.

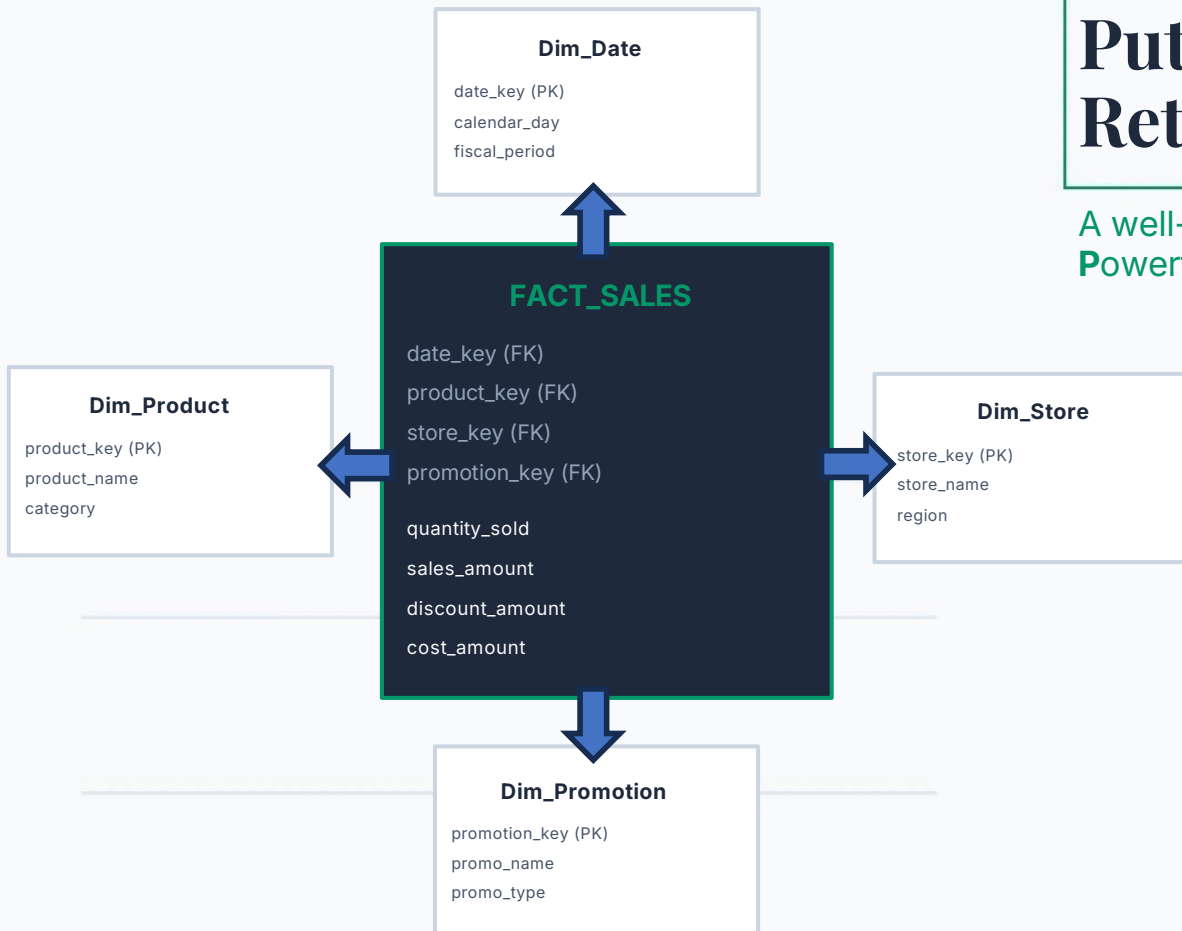
Retail Sales Facts

-  Sales Revenue (Gross & Net)
-  Quantity Sold (Units)
-  Discount Amount
-  Cost of Goods Sold (COGS)



Putting It Together: Retail Sales Star Schema

A well-designed schema enables
Powerful, Simple, and Fast Analytics.



FK = Foreign Key | PK = Primary Key (Surrogate)
Center: Measurements at Declared Grain | Satellites: Descriptive Context

Avoid These Four Common Dimensional Modeling Mistakes

Most failures stem from ignoring the four-step process or mixing grains.



Unclear Grain

01

Failing to define exactly what a row represents leads to ambiguity and inconsistent query results across developers.

Mixed Grain

02

Putting different levels of detail in one fact table (e.g., daily and transaction) causes massive double-counting errors.

Facts in Dimensions

03

Placing performance metrics in descriptive tables causes dimension bloat and breaks historical tracking (SCD) logic.

Premature Aggregation

04

Summarizing data too early to "save space" or "speed up" results in losing the ability to drill down into critical business details.

Master the Process to Build Scalable Data Models

Consistency and atomic detail are the foundations of successful BI.

These five rules prevent 90% of modeling failures. Adhering to Kimball's framework ensures your data warehouse remains a single source of truth.

- 01 Process First**
Always follow the four steps in order: Process → Grain → Dimensions → Facts.
- 02 Atomic is Best**
Start at the lowest level of detail to ensure maximum future flexibility.
- 03 One Grain per Table**
Never mix levels of detail in a single fact table to avoid broken metrics.
- 04 Dimensions Describe**
Keep descriptive attributes in wide, denormalized dimension tables.
- 05 Facts Measure**
Store numeric measurements in narrow, deep fact tables using surrogate keys.

Advance Your Skills with Kimball's Methodology

Real-world modeling requires continuous
practice and learning.



Thank you for your attention.

Go forth and declare grain.

ESSENTIAL READING

-  **The Data Warehouse Toolkit**
by Ralph Kimball (Chapters 1-3)

IMMEDIATE PRACTICE

-  Reverse-engineer an existing report: what is the grain?
-  Redesign a legacy schema using the four-step process.

NEXT TOPICS

-  Slowly Changing Dimensions (SCD Types 1, 2, 3)
-  Conformed Dimensions and Bus Architecture

GET IN TOUCH

-  <https://www.linkedin.com/in/milan-gabriel/>
-  <https://bit.ly/the1onwrongway>