

Designing a Data Warehouse on the Microsoft SQL Server Platform

UNDERSTANDING THE FUNDAMENTAL CONCEPTS OF
A DATA WAREHOUSE



Ana Voicu

@ana_voicu



Overview



Goals and purpose of a data warehouse

Introducing dimensional modeling

- What are facts and fact tables?
- What are dimensions and dimension tables?
- Star schemas

Putting it all together

- The 4-step dimensional design process



Goals of a Data Warehouse Solution

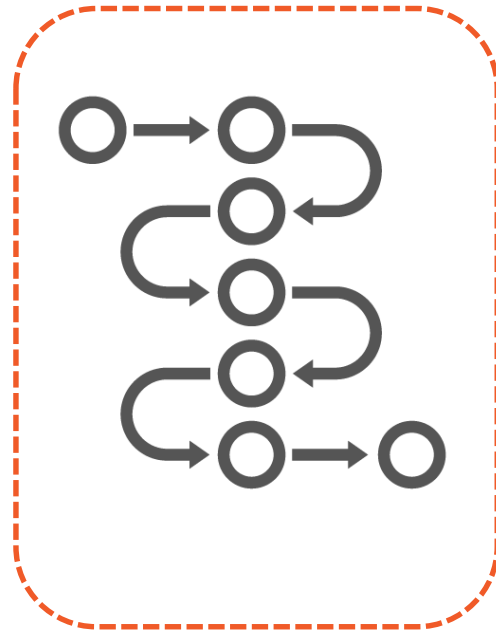


What Is a Data Warehouse?

Data sources



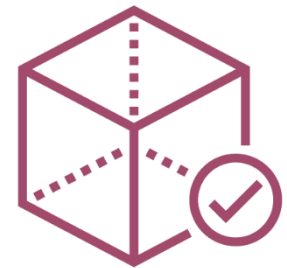
ETL process



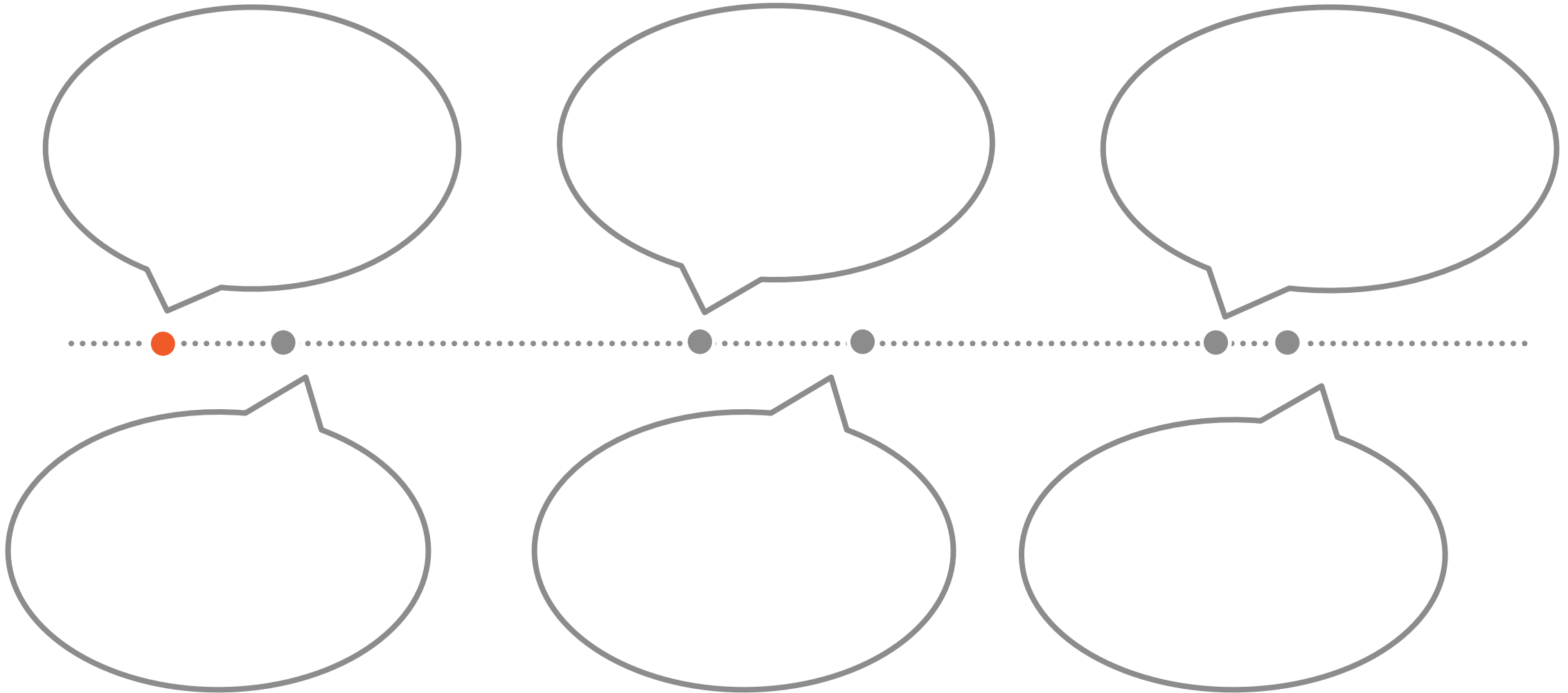
Data warehouse



Presentation



Problems a Data Warehouse Can Solve



Requirements of a Data Warehouse Solution

Easily accessible

Fast

Consistent

Flexible

Secure

**Foundation for
decision-making**



Responsibilities of a Data Warehouse Designer

**Understand the
business users**

**Deliver high-quality,
relevant, and
accessible
information**

**Sustain the DW
environment**



Introduction to Dimensional Modeling



Dimensional Modeling



Database design method optimized for data warehouse solutions

Popular technique because it addresses two important requirements:

- Deliver data in an understandable format
- Deliver fast query performance

Key word is “simplicity”

Elements of a Dimensional Model



Facts (the measurements/metrics or facts from your business process)



Dimensions (for providing the context of a business process event)



Attributes (the various characteristics of a dimension)



Star schema (and/or OLAP cubes)



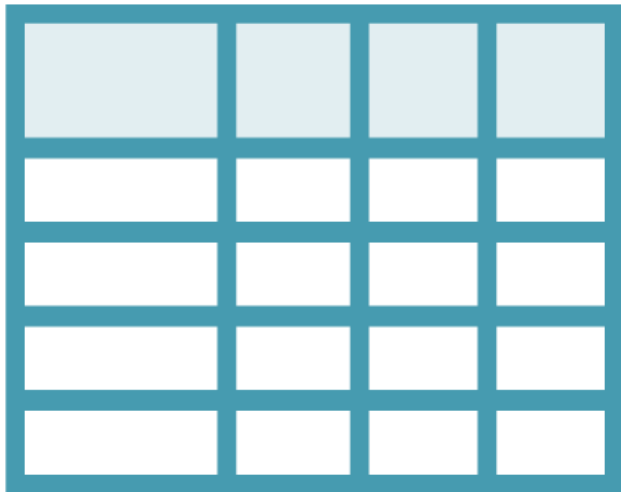
“We sell ice-cream and other products in various locations and measure our achievements over time.”

Jane Poppins

Happy Scoopers CEO



Fact Tables and Facts



Fact table = table that stores the performance measurements resulting from an organization's business process events



Fact = a business measure

- Sales
- Profit
- Volume
- Number of transactions

Fact Tables and Facts



Facts answer questions like:

- “What are we doing?” (sell, buy, count)
- “What do we want to achieve?” (more sales, bigger profit)

1 row in the fact table is 1 measurement in real life

Fact columns in a fact table should be additive

Facts make sense in combination with dimensions

- Linked with foreign keys
- Date/Time dimension is present in most data warehouses

Example of a Fact Table



Sales fact

Date key

Product key

Store key

Employee key

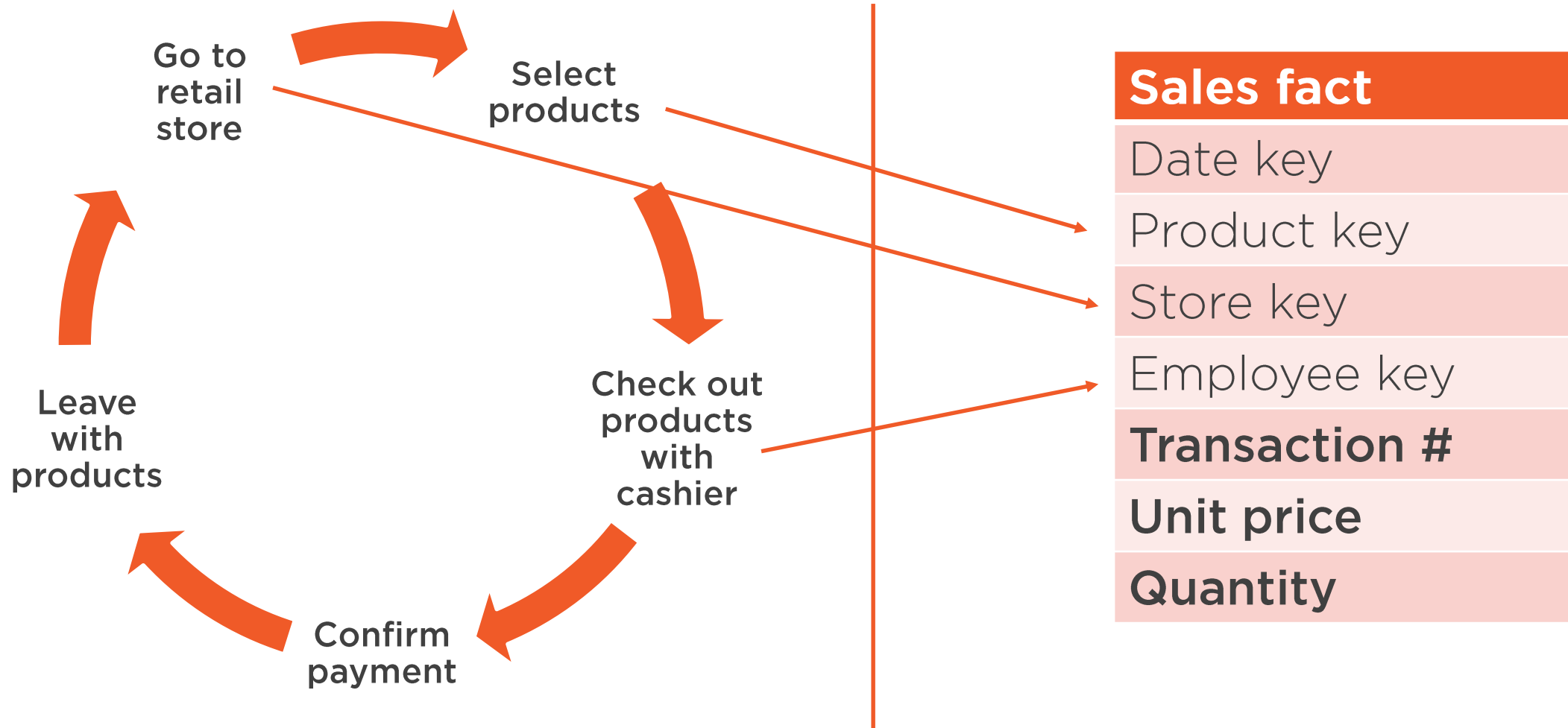
Transaction #

Unit price

Quantity



Example of a Fact Table



Example of a Fact Table



Sales Fact
Date key
Product key
Store key
Employee key
Transaction #
Unit price
Quantity

COMPOSITE KEY



Characteristics of Fact Tables



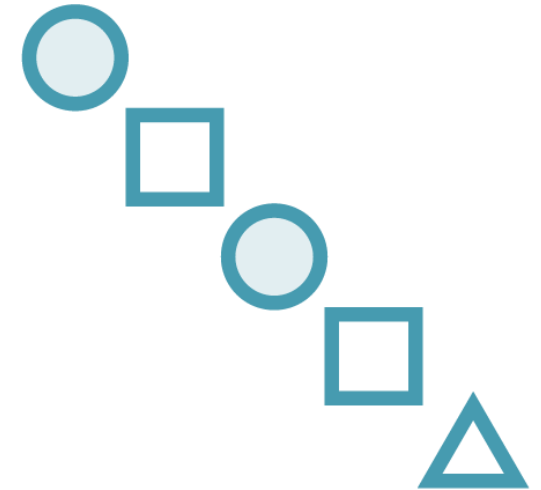
1:1 relationship
between fact table
row and real-world
event



Most facts
should be
additive



Foreign keys to
dimension
tables



Composite key
as PK for fact
table

What Are Dimensions?



Companions to a fact table

Textual context associated with a business process measurement event

Questions Answered by Dimension Tables



Who



What



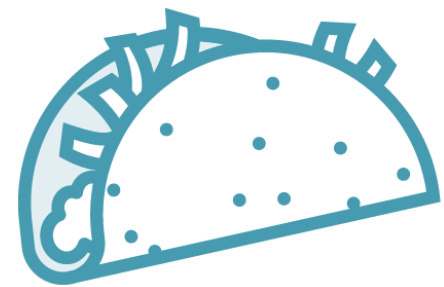
Where



When



How



Why



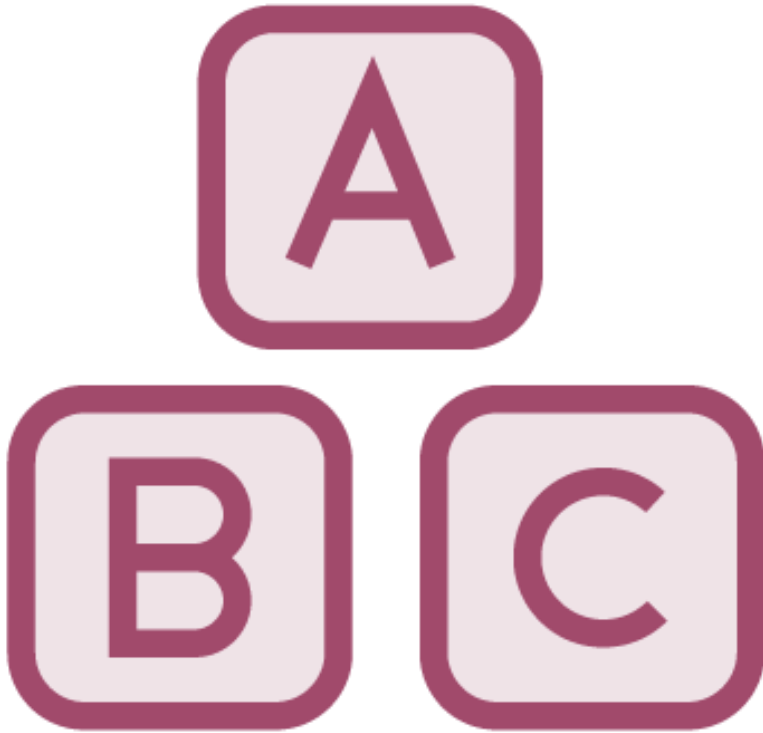
Example of a Dimension Table



Product Dimension
Product key
Product name
Brand name
Category name
Subcategory name
Package type
Package size
Weight
Weight unit of measure



Characteristics of Dimension Tables



No limit for the number of attributes in a dimension table

- Common to have tables with 50 to 100 attributes
- Some dimension tables have only a handful of attributes

Have fewer rows than fact tables

- But can be much wider

Defined by a single primary key

- Basis for the referential integrity with the fact table

Denormalized

- Flattened many-to-one relationships within a single dimension table



Dimension Attributes

The primary source of

- Query constraints
- Groupings
- Report labels

Quality of attributes \propto quality of the system

- Use real words vs. cryptic abbreviations
- Minimize codes in the dimension tables

Product Key	Name	Category name	Subcategory name
1	Beery cotton candy	Candy	French candy
2	Cotton candy	Candy	French candy
3	Peppermint candy (seasonal)	Candy	Fudge
4	Green tea ice cream	Ice-cream	Reduced fat
5	Chocolate chip cookie dough ice cream	Ice-cream	Fat-free frozen dairy
6	Neapolitan ice cream	Ice-cream	Lactose-free
7	Cantuccini	Cookie	Biscotti
8	Chocolate mint cookie	Cookie	Retro snacks
9	Lemon cookie	Cookie	Fruity cookies



Dimensional Model

Characteristics:

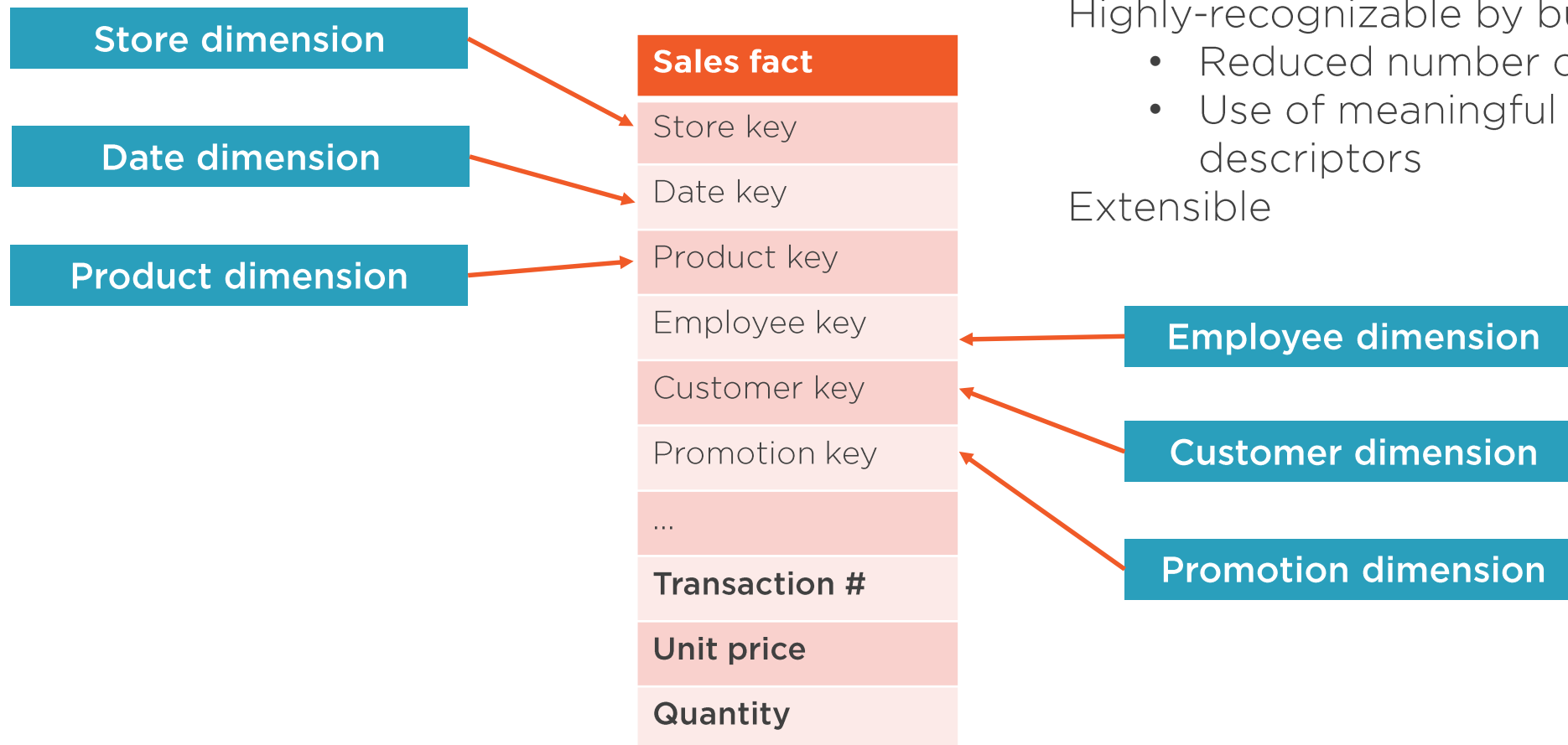
Simple

Symmetric

Highly-recognizable by business users

- Reduced number of tables
- Use of meaningful business descriptors

Extensible



Dimensional Model

Characteristics:

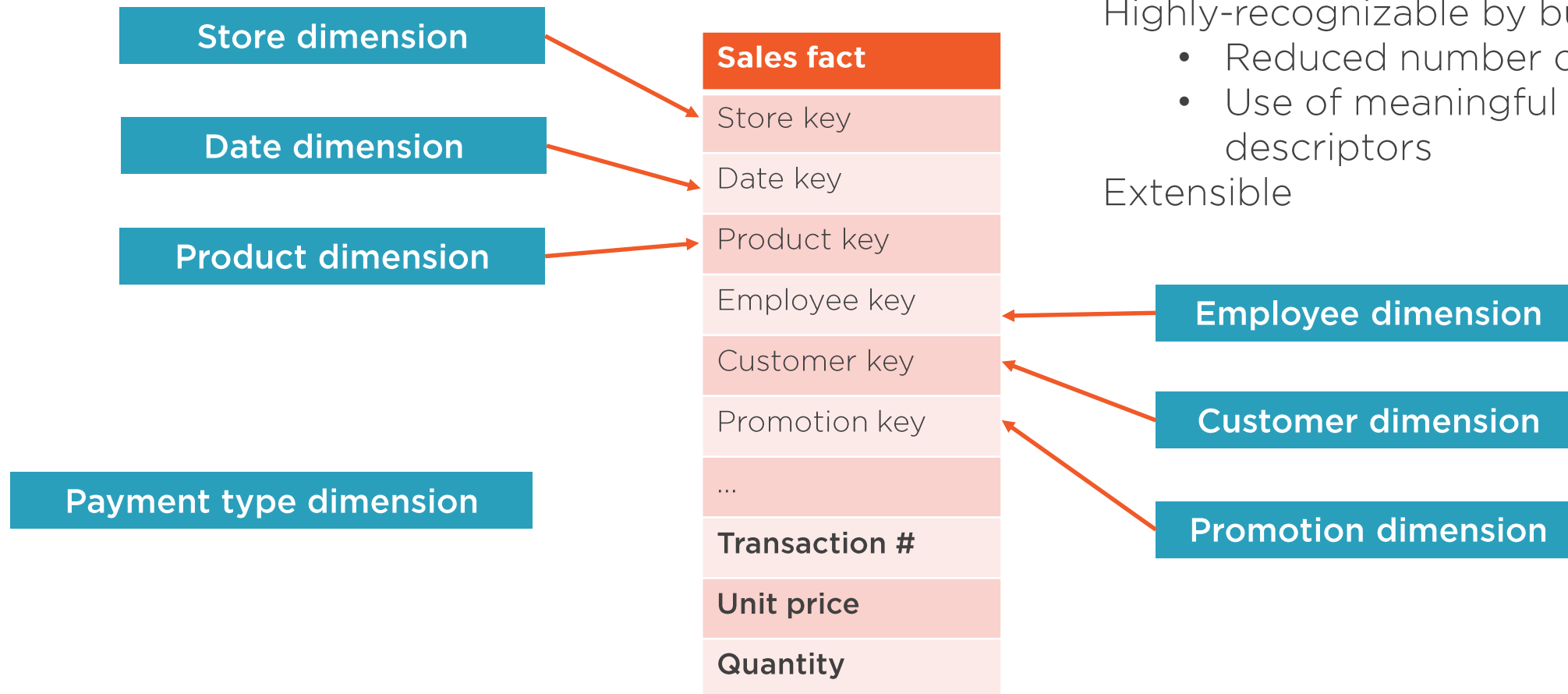
Simple

Symmetric

Highly-recognizable by business users

- Reduced number of tables
- Use of meaningful business descriptors

Extensible



Dimensional Model

Characteristics:

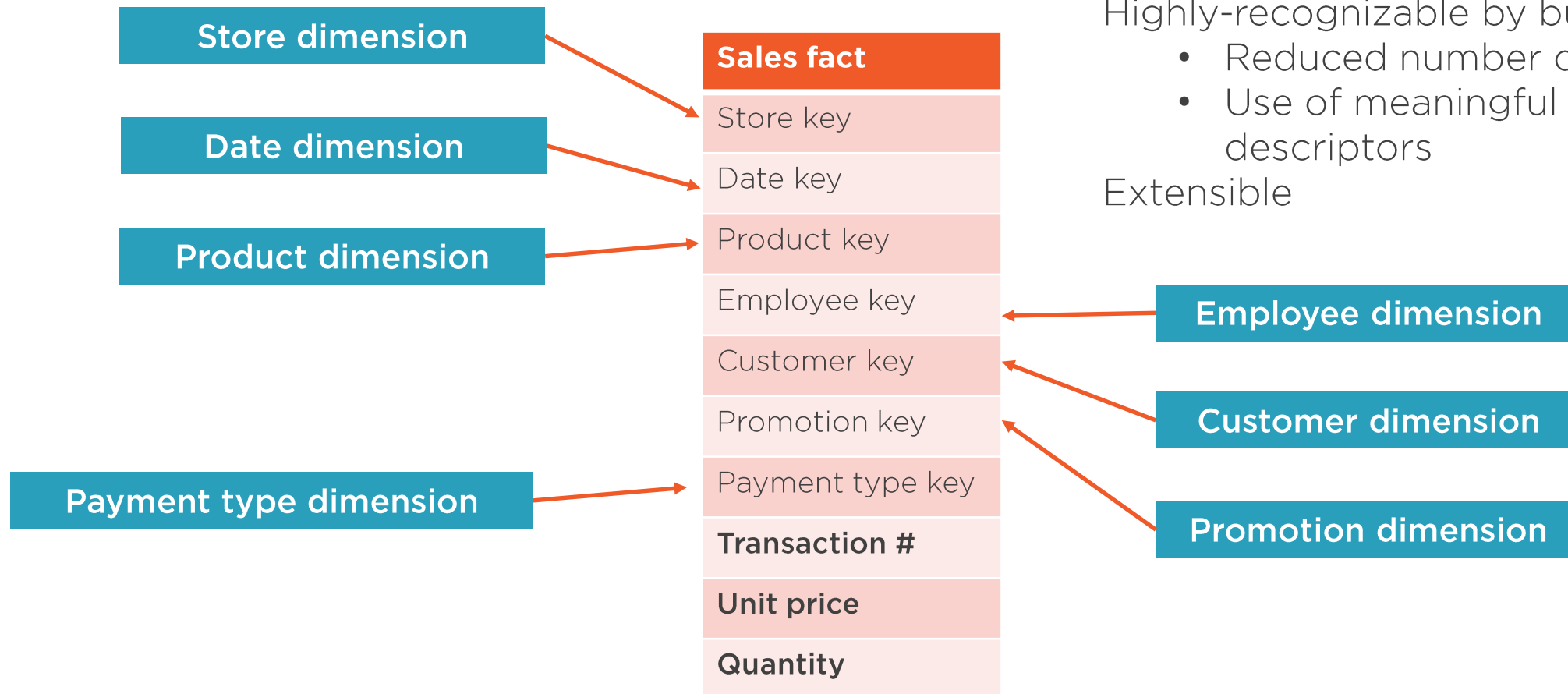
Simple

Symmetric

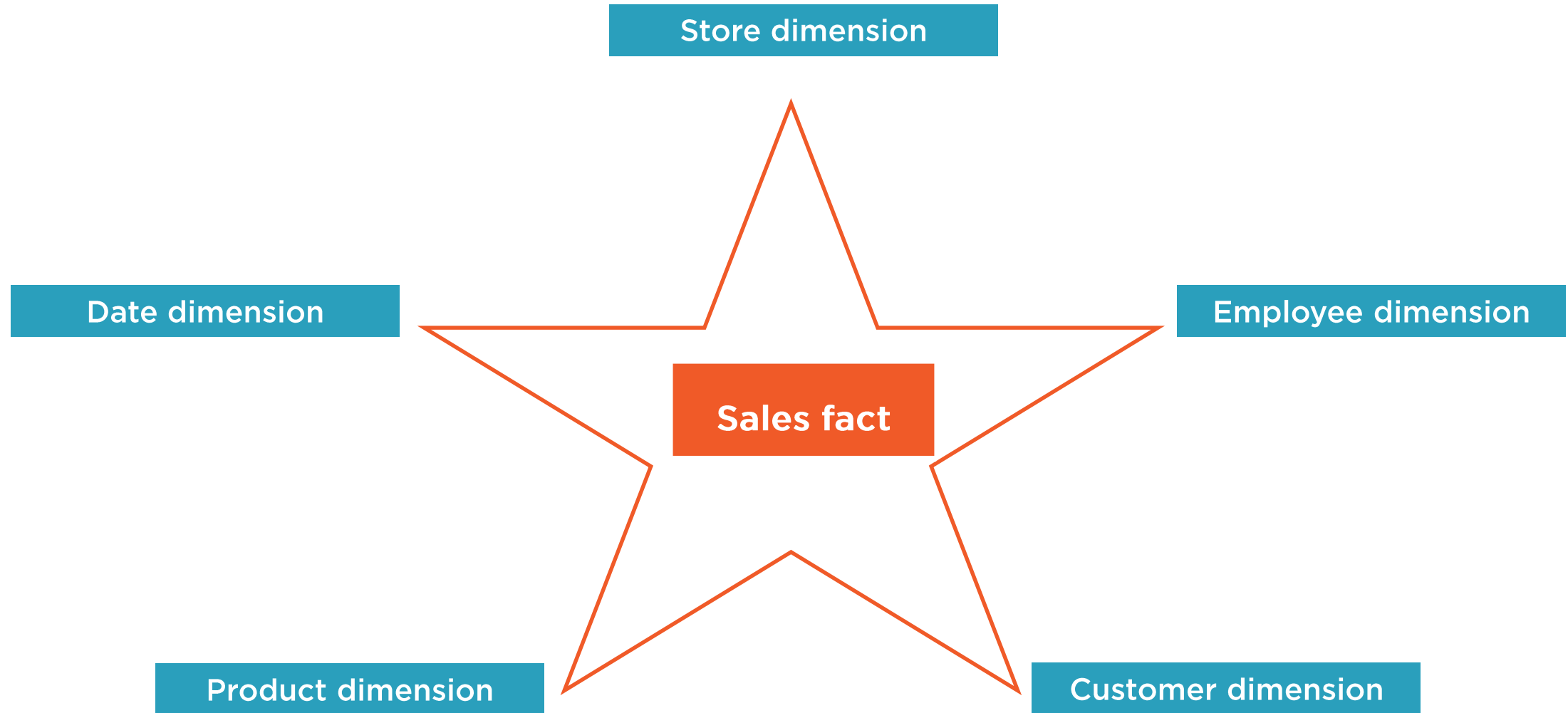
Highly-recognizable by business users

- Reduced number of tables
- Use of meaningful business descriptors

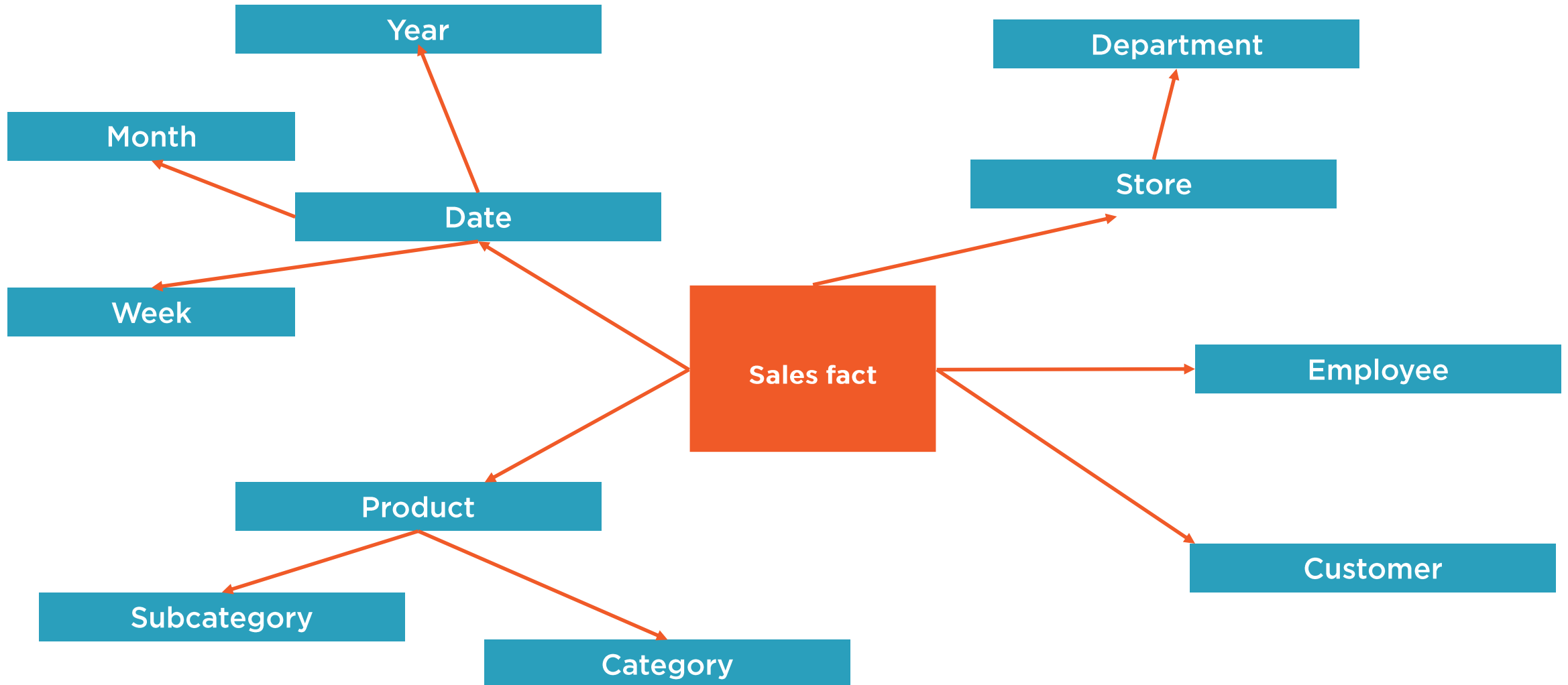
Extensible



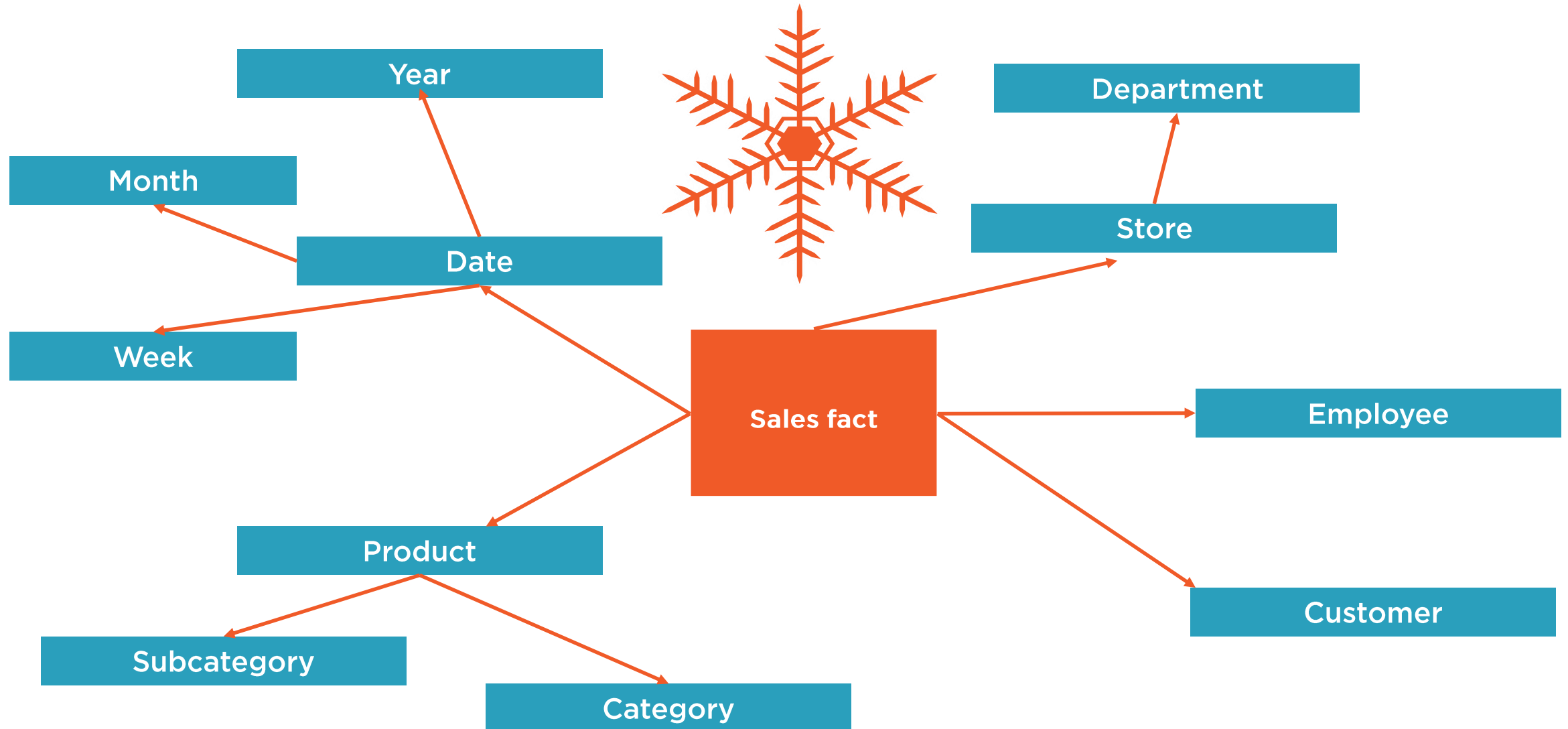
Dimensional Model as a Star Schema



Dimensional Model as a Snowflake



Dimensional Model as a Snowflake



The Four-step Dimensional Design Process



Dimensional Design

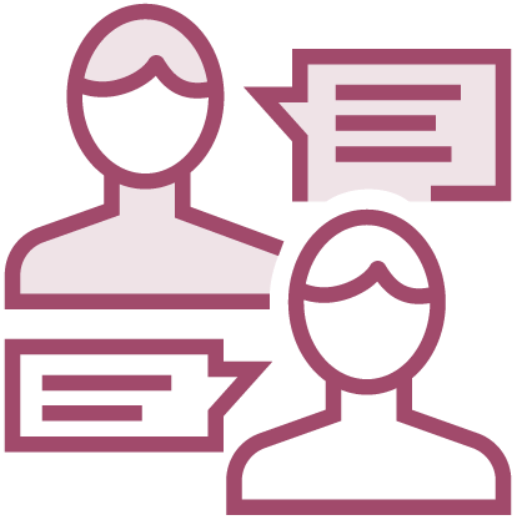


Is done “with pen and paper”

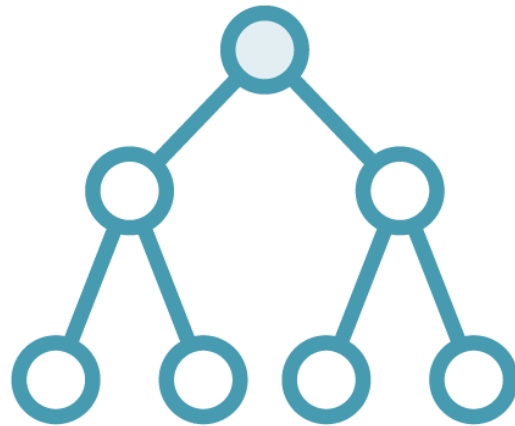
Focuses on understanding the deliverables of the project

Consists of four steps

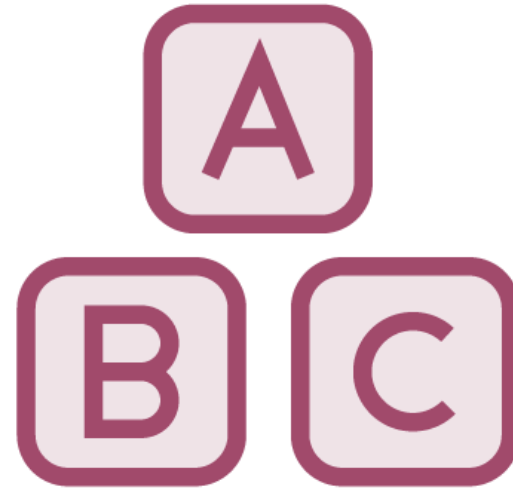
Steps of Dimensional Design



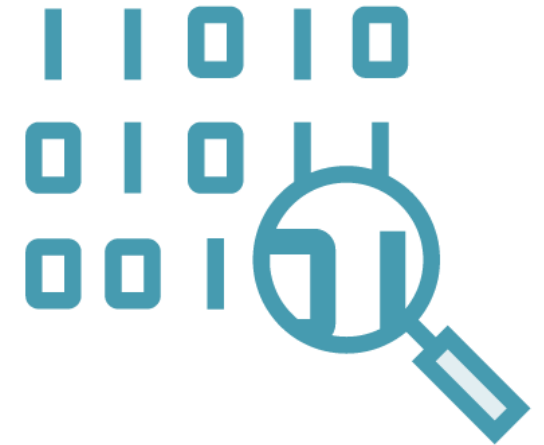
Select the
business
process



Declare the
grain



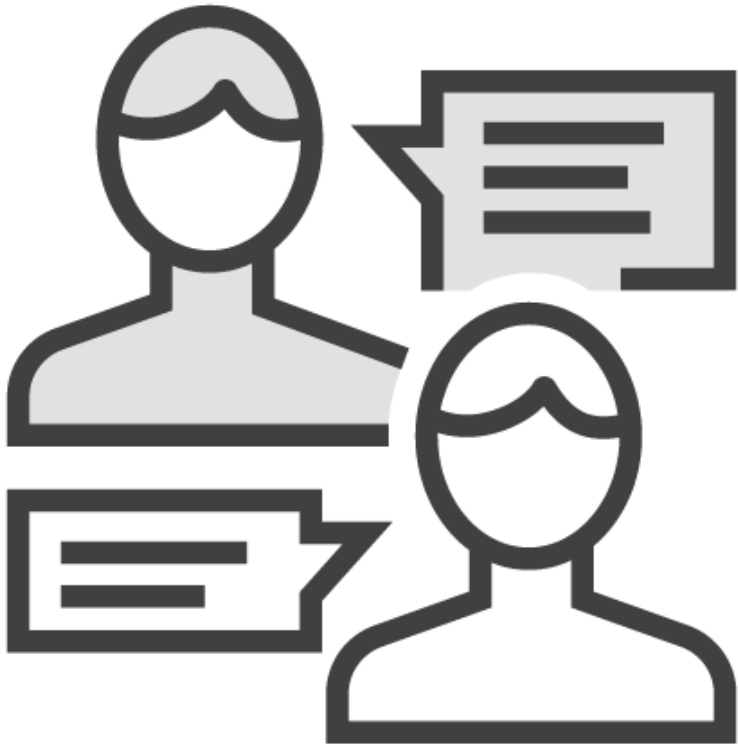
Identify the
dimensions



Identify the
facts



Step 1: Select the Business Process



Low-level activities performed by an organization

Are identified by listening carefully to the business users

Characteristics:

- Expressed as verbs
- Are supported by an operational system
- Generate KPIs



Step 2: Identify the Grain

Specify the detail level of a business process we want to measure

Example	Grain	Questions to ask
Sales/day	One row at the end of the day	How much was paid by each customer? Is this important information?
Sales/day/order	One row for every finished order	What was the most sold product? Is this important information?
Sales/day/order/product	One row for every product sold within an order	Is this enough information?





Step 2: Identify the Grain

Specify the detail level of a business process we want to measure

“How do you describe a single row in the fact table?”

Grain declarations are expressed in business terms



Steps 3 and 4

Identify the dimensions

“How do business users describe the data resulting from the process?”

“who, what, where, when, why, how”

Examples:

- Date
- Product
- Customer
- Employee

Identify the facts

“What is the process measuring?”

All candidate facts must be true to the grain for that fact table

Facts with different grains are split in separate tables

Examples:

- Sales price
- Sales quantity (or Units sold)



Both business requirements and the realities of the source data should be considered when designing the dimensional model

