Implementing Fact Tables

Ana Voicu

@ana_voicu

Overview

Structure of a fact table

The most common types of fact tables

- Transaction
- Periodic snapshot
- Accumulating snapshot

Handling nulls in fact tables

Demo: Creating and working with fact tables

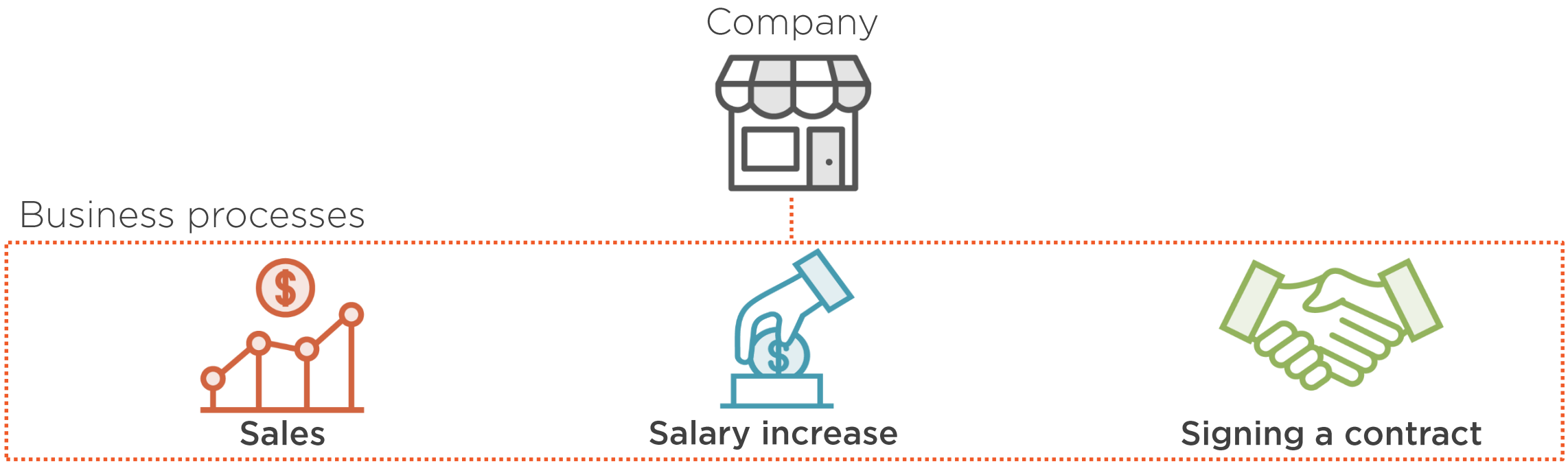
Structure of a Fact Table

Examining a Fact Table's Purpose

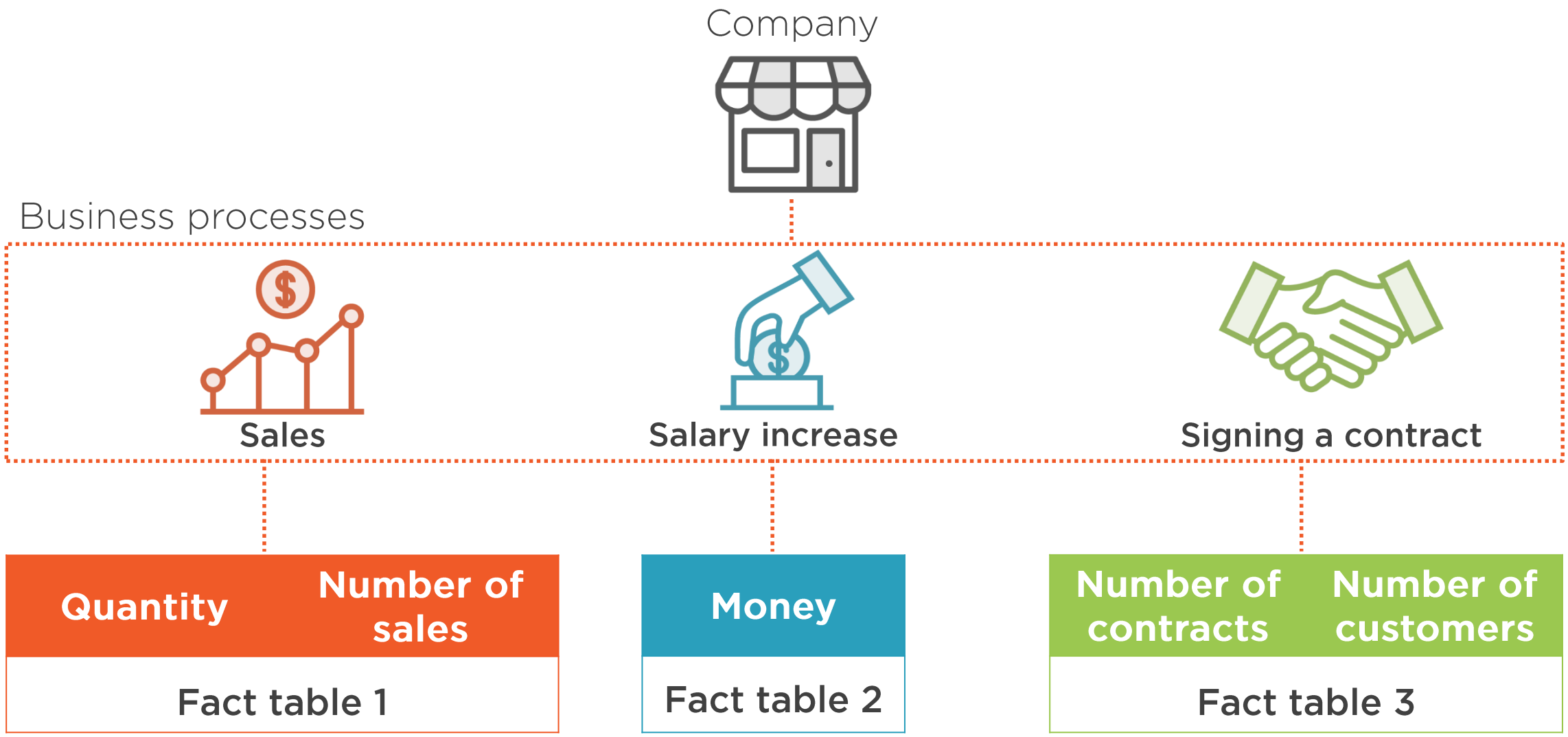
Examining a Fact Table's Purpose

Company

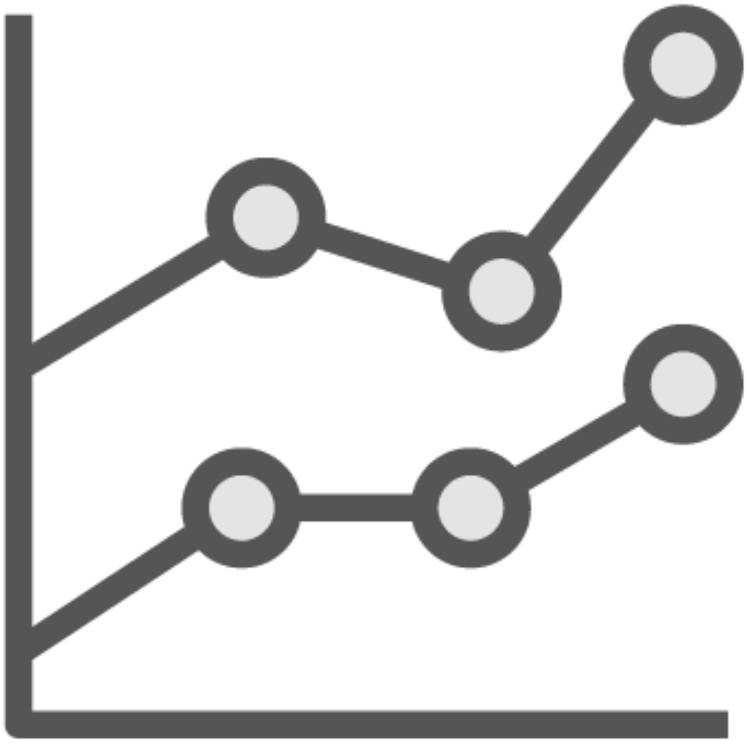
Examining a Fact Table's Purpose



Examining a Fact Table's Purpose



Fact Table Characteristics



Used with dimensions for:

- Reporting purposes
- Measuring performance
- Analyzing trends in time

Contain a lot of measurements

They tend to grow very fast

Performance is an important design factor

Walmart Case Study

As of April 2019

Number of stores	11.300
Annual revenue in previous fiscal year	514 billion dollars
Days in a year	365
Avg. revenue per day	1.4 billion dollars
Avg. revenue per day per store	124.000 dollars
Avg. price of a banana	19 cents

Walmart Case Study

As of April 2019

Number of stores	11.300
Annual revenue in previous fiscal year	514 billion dollars
Days in a year	365
Avg. revenue per day	1.4 billion dollars
Avg. revenue per day per store	124.000 dollars
Avg. price of a banana	19 cents

Walmart Case Study

As of April 2019

Number of stores	11.300
Annual revenue in previous fiscal year	514 billion dollars
Days in a year	365
Avg. revenue per day	1.4 billion dollars
Avg. revenue per day per store	124.000 dollars
Avg. price of a banana	19 cents

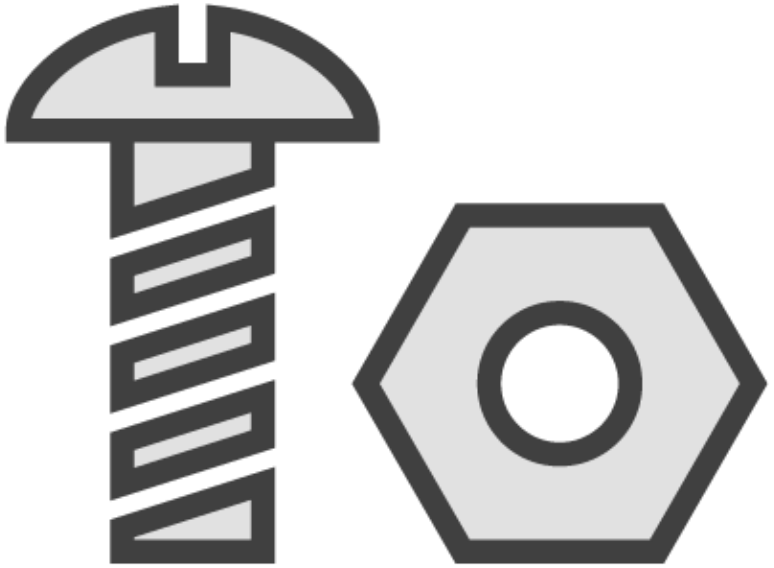
How many transactions are recorded per day?

What are the best sold products from each store?

Data warehouse fact tables

- Sales fact table
- Profit and loss
- Periodic revenue per period

Main Elements of a Fact Table



Facts

- Columns storing the measurements of the process

Primary key

- Uniquely identifies rows

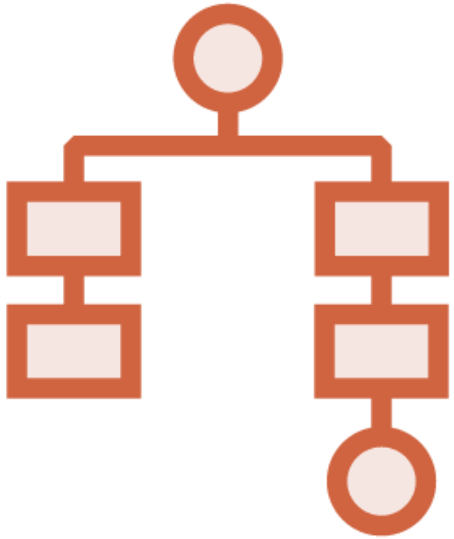
Foreign keys

- Linking the fact and the dimension tables

Steps for Identifying Facts

Analyze what is important in a business process

Identify multiple facts

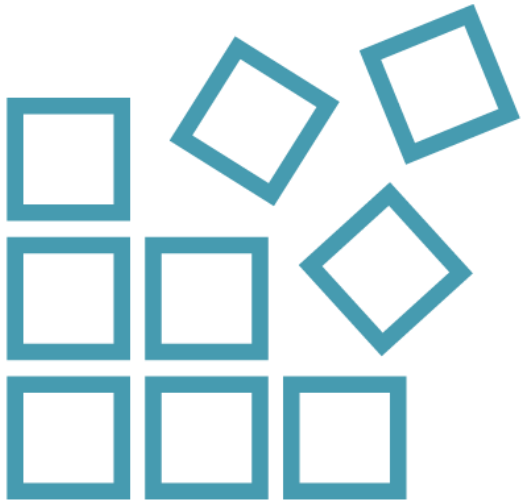


Group facts by granularity

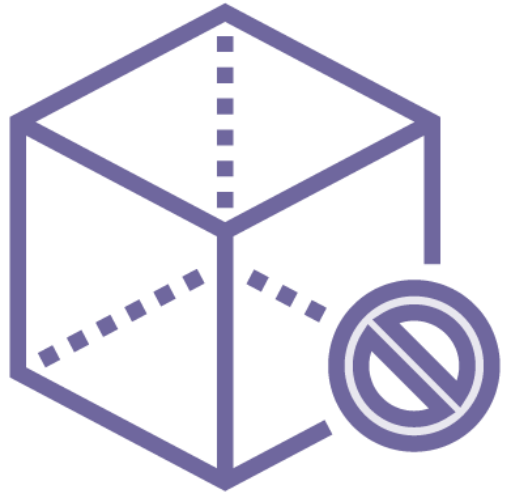
Double-check the facts

Types of Facts

Additive



Semi-additive



Non-additive

Additive Facts

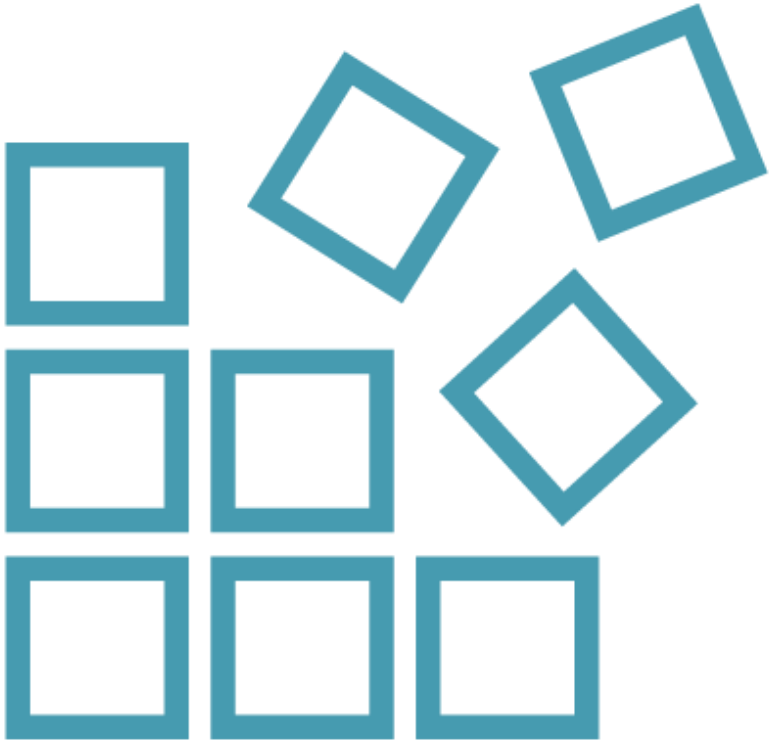
Can be summarized across all dimensions from a fact

The most common type of fact

Examples:

- Sales amount
- Sales quantity
- Purchase amount
- Number of transactions

Semi-additive Measures



Can be summarized across *some* dimensions

Cannot be summarized across time

Example of Semi-additive Measures

Example of Semi-additive Measures

Date key	Product key	Store key	Quantity	Inventory value at price ($)
04-03-2019	777	1	200	400
04-04-2019	777	1	100	200

Example of Semi-additive Measures

Date key	Product key	Store key	Quantity	Inventory value at price ($)
04-03-2019	777	1	200	400
04-04-2019	777	1	100	200
04-03-2019	80	1	50	50
04-04-2019	80	1	25	25

Example of Semi-additive Measures

Date key	Product key	Store key	Quantity	Inventory value at price ($)
04-03-2019	777	1	200	400
04-04-2019	777	1	100	200
04-03-2019	80	1	50	50
04-04-2019	80	1	25	25

How many units were in the store on the 3rd of April?

What was their value?

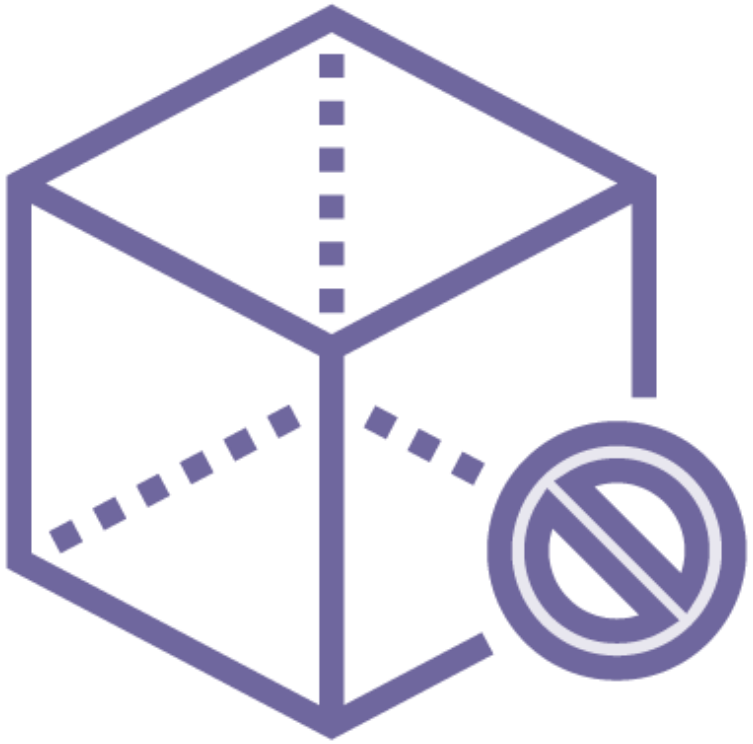
Example of Semi-additive Measures

Date key	Product key	Store key	Quantity	Inventory value at price ($)
04-03-2019	777	1	200	400
04-04-2019	777	1	100	200
04-03-2019	80	1	50	50
04-04-2019	80	1	25	25

Examples of semi-additive measures:

- Inventory information
- Financial account balances
- Water levels on rivers
- Temperature

Non-additive Measures



Ratios or percentages

Cannot be summarized across any dimension

Example of Non-additive Measures

Example of Non-additive Measures

Date key	Product key	Store key	Initial quantity	Removed quantity
04-03-2019	777	1	200	100
04-04-2019	777	1	100	40
04-03-2019	80	1	50	25
04-04-2019	80	1	25	9

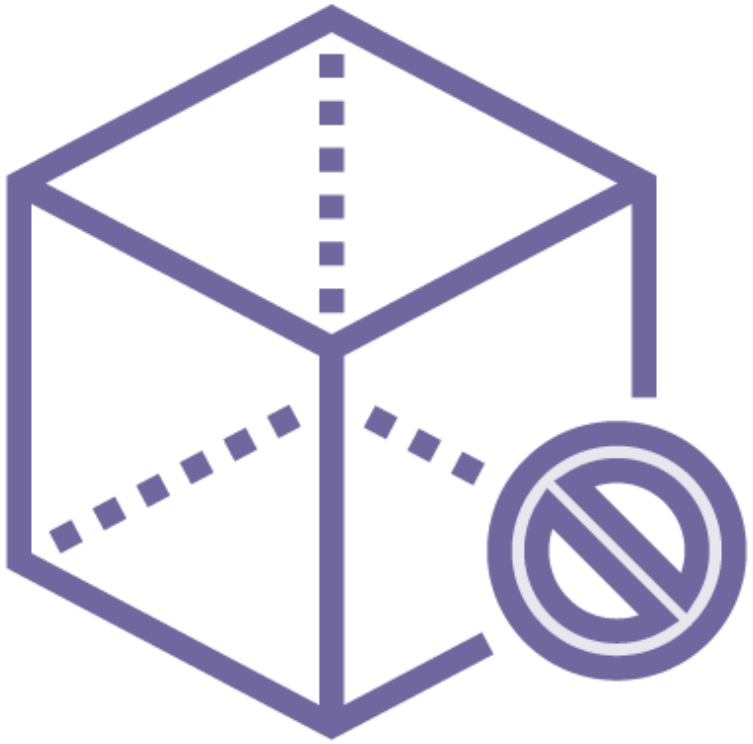
Example of Non-additive Measures

Date key	Product key	Store key	Initial quantity	Removed quantity	Removed quantity (%)
04-03-2019	777	1	200	100	50
04-04-2019	777	1	100	40	40
04-03-2019	80	1	50	25	50
04-04-2019	80	1	25	9	36

Example of Non-additive Measures

Date key	Product key	Store key	Initial quantity	Removed quantity	Removed quantity (%)	Unit price
04-03-2019	777	1	200	100	50	2
04-04-2019	777	1	100	40	40	2
04-03-2019	80	1	50	25	50	1
04-04-2019	80	1	25	9	36	1

Non-additive Measures



Ratios or percentages

Cannot be summarized across any dimension

Why are non-additive measures considered facts?

- Design preferences
  - Storing them in fact tables vs. in dimension tables, as attributes
- They are calculated based on columns from the fact table

They should be used with care by users

Keys in a Fact Table

Identifying the Keys

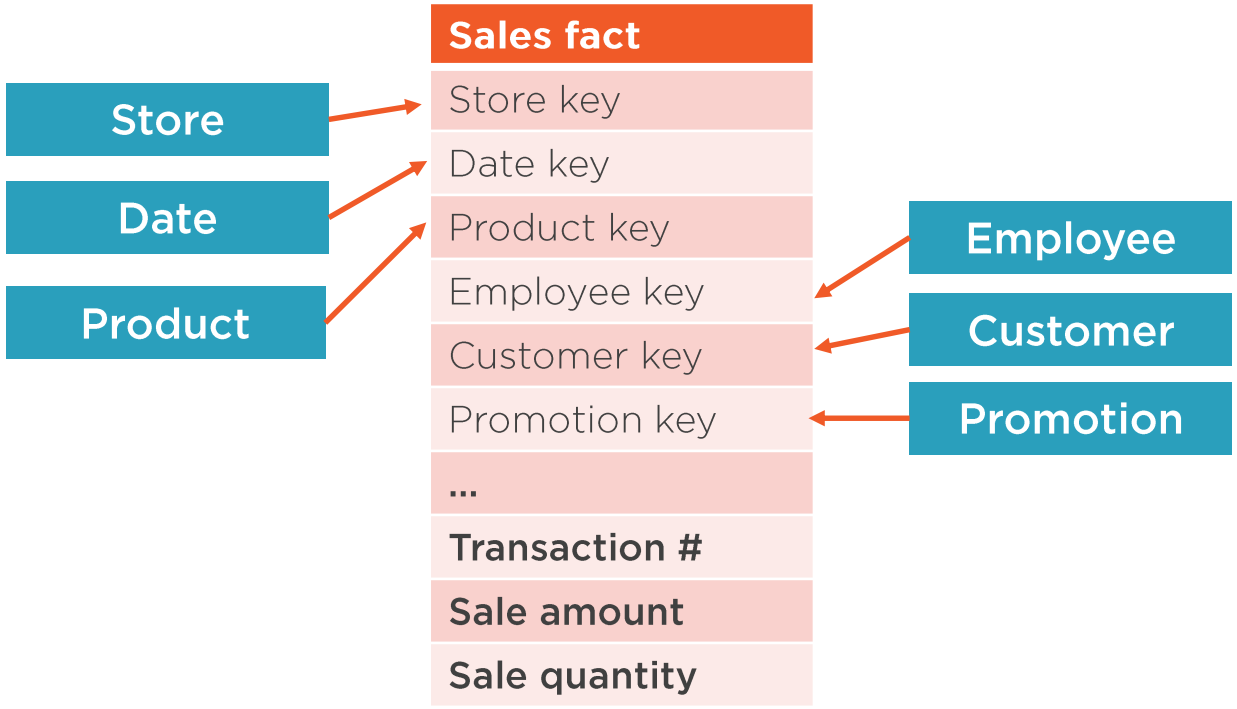
Primary key

- Composite key
- Surrogate key

Foreign keys for dimensions

Identifying the Primary Key

Identifying the Primary Key



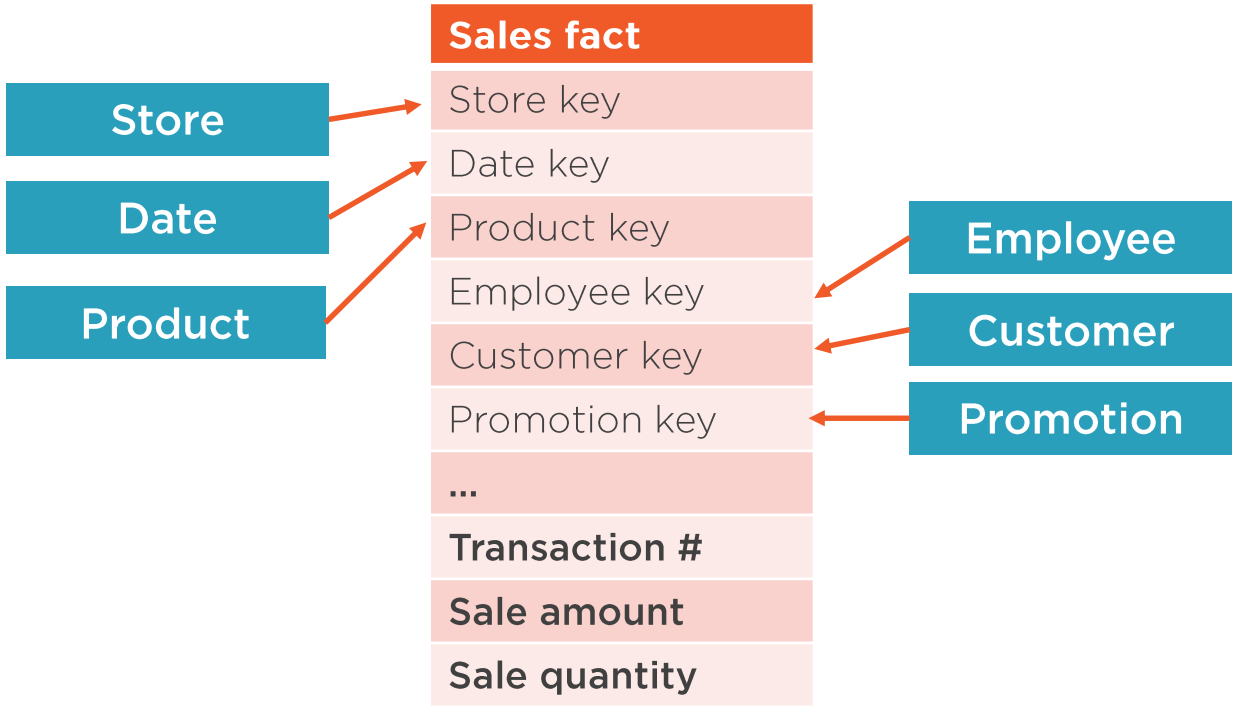
Relationships between fact and dimension are many-to-many

Examples

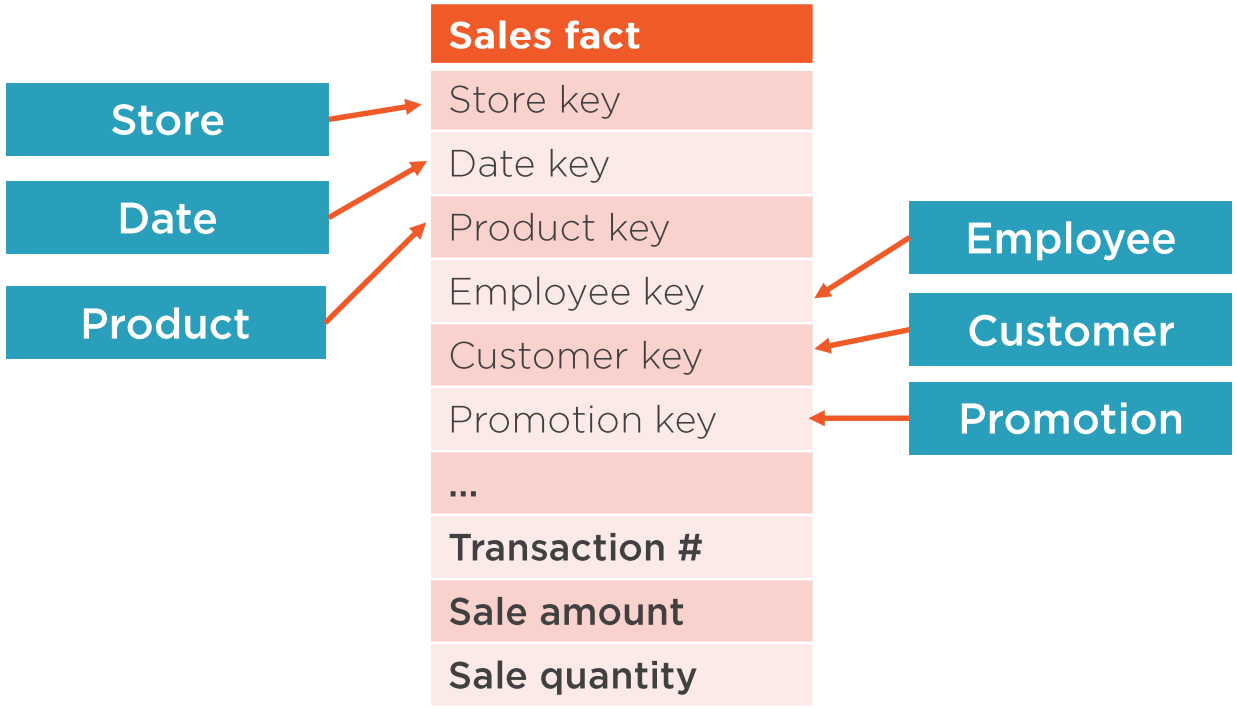
1.a. One sale can have multiple products
1.b. One product can be part of multiple sales

2.a. One sale can have multiple promotions applied
2.b. One promotion can be applied to many sales

Identifying the Primary Key



Identifying the Primary Key

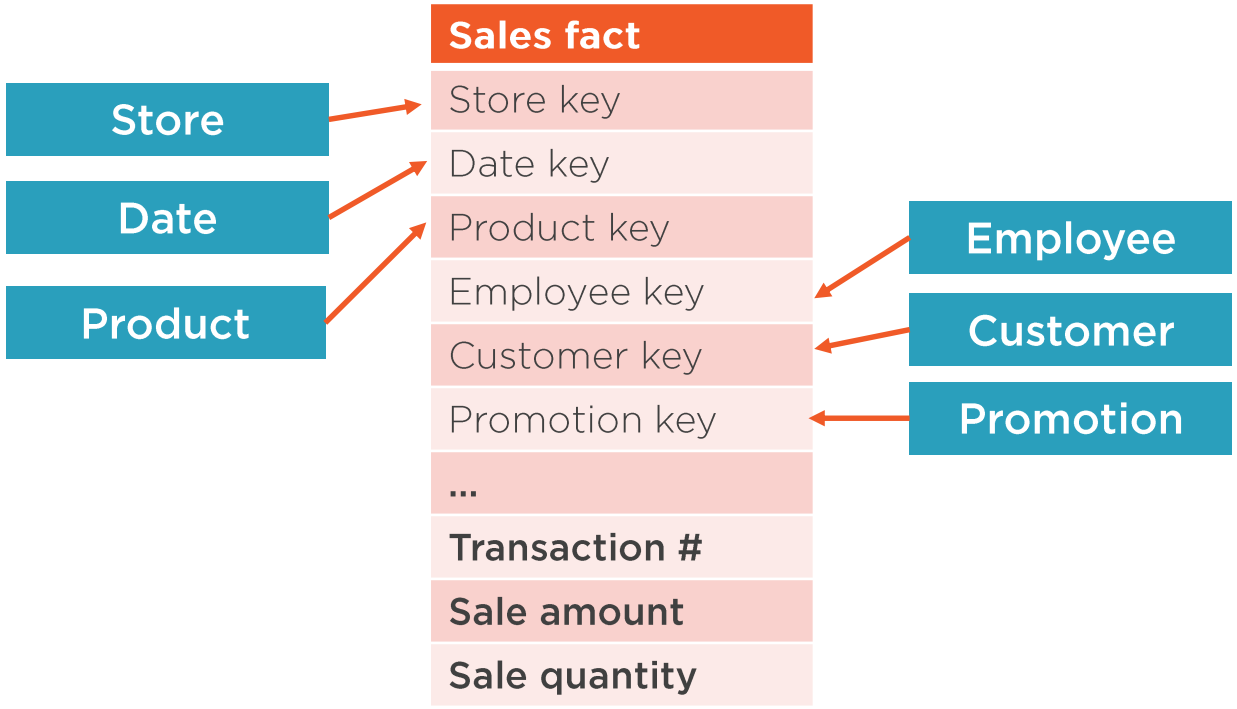


1. Combine a subset of the foreign keys
Also called

- Composite key
- Compound key
- Concatenated key

The combination must ensure row uniqueness

Identifying the Primary Key

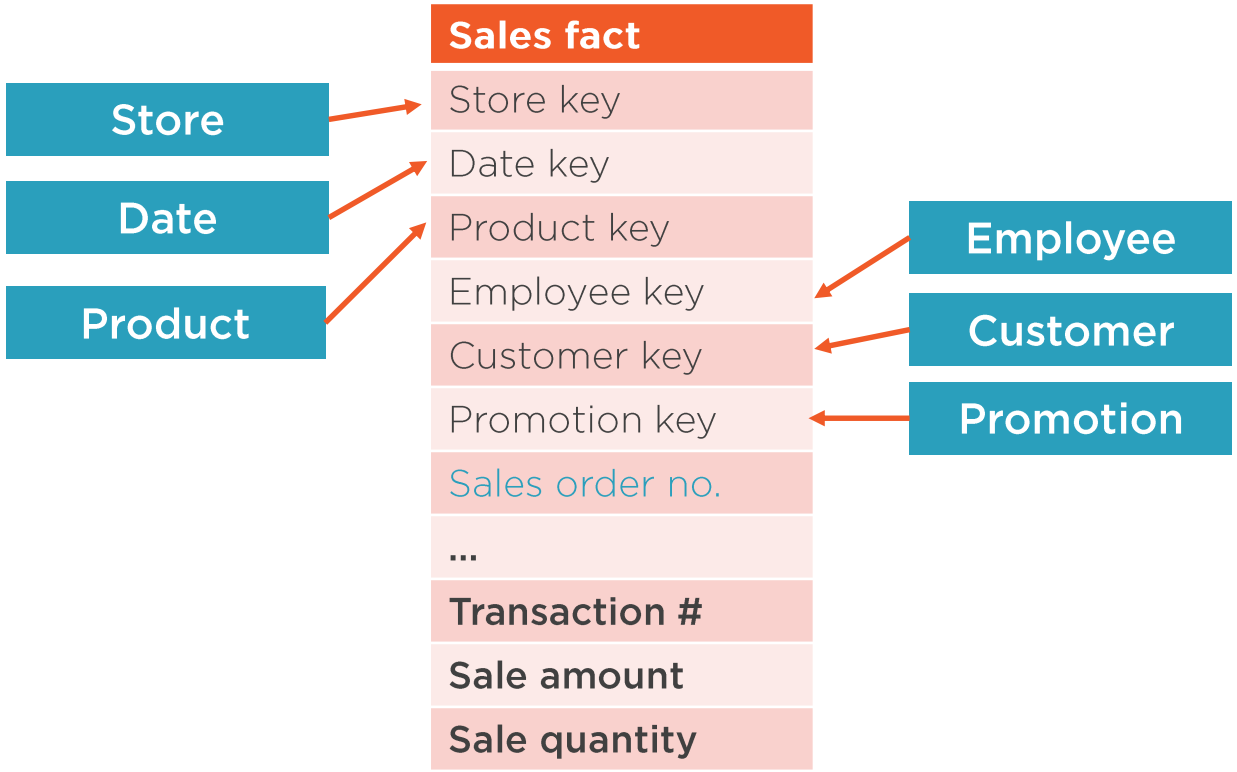


1. Combine a subset of the foreign keys
Also called
- Composite key
- Compound key
- Concatenated key

The combination must ensure row uniqueness

2. Use another column from the source system: the primary key for the transactions
Merge the new column to the subset of FKs

Identifying the Primary Key



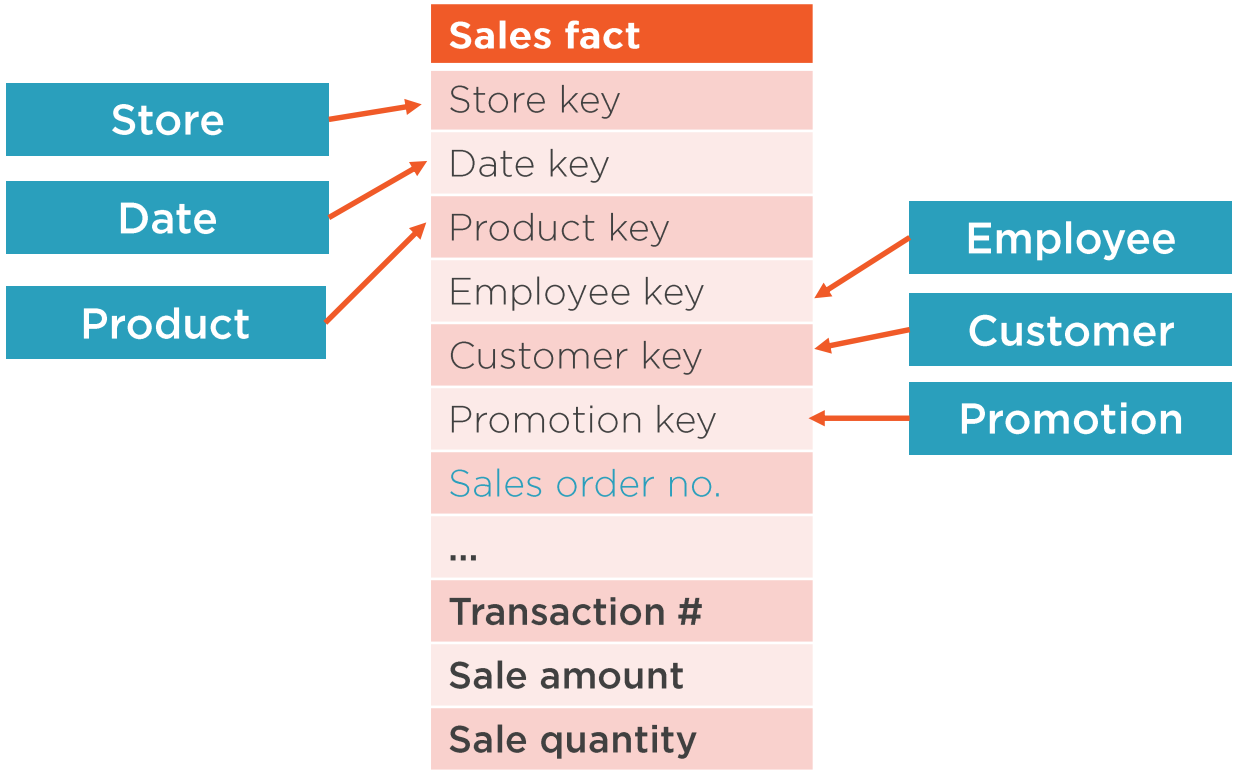
1. Combine a subset of the foreign keys
Also called

- Composite key
- Compound key
- Concatenated key

The combination must ensure row uniqueness

2. Use another column from the source system: the primary key for the transactions
Merge the new column to the subset of FKs

Identifying the Primary Key



1. Combine a subset of the foreign keys
Also called
- Composite key
- Compound key
- Concatenated key

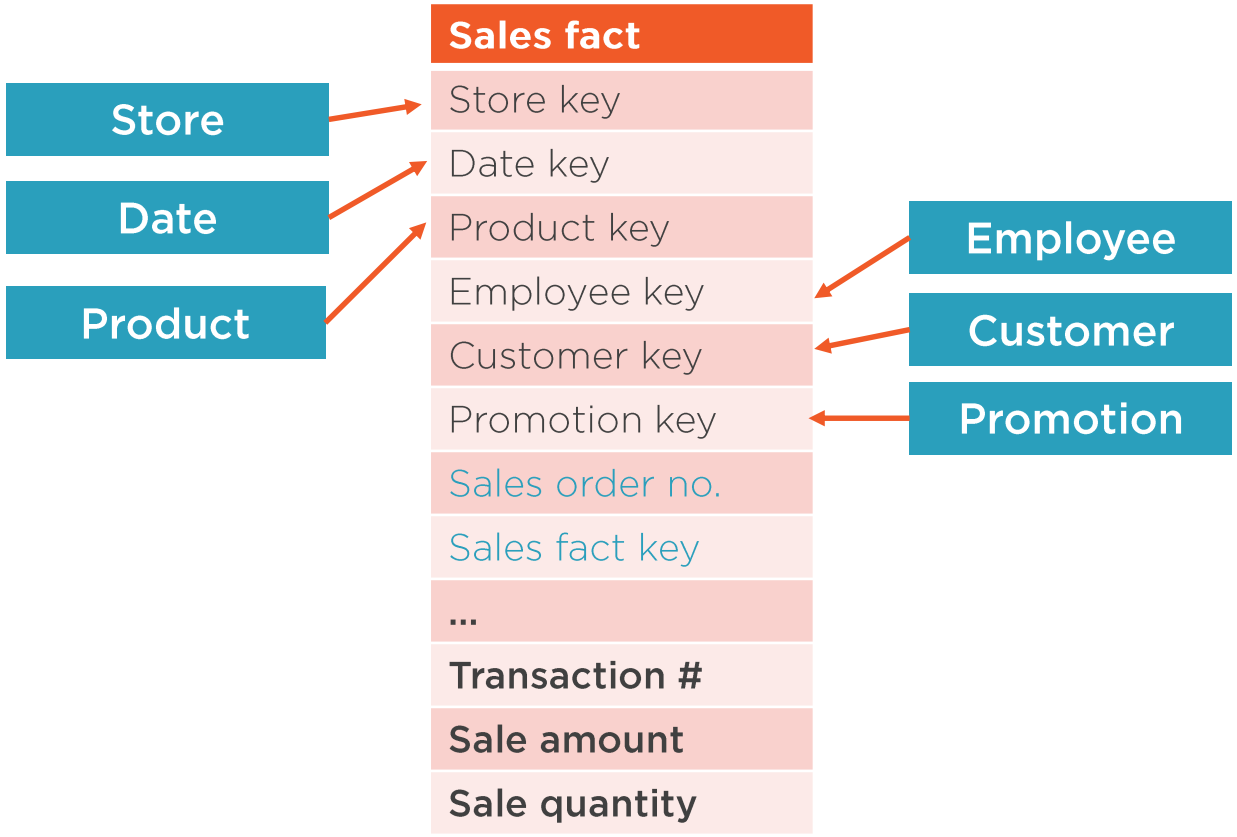
The combination must ensure row uniqueness

2. Use another column from the source system: the primary key for the transactions
Merge the new column to the subset of FKs

3. Create a surrogate key
- Small integer
- Auto-incremented
- No business value

The FKs don't play a special role anymore

Identifying the Primary Key



1. Combine a subset of the foreign keys
Also called
- Composite key
- Compound key
- Concatenated key

The combination must ensure row uniqueness

2. Use another column from the source system: the primary key for the transactions
Merge the new column to the subset of FKs

3. Create a surrogate key
- Small integer
- Auto-incremented
- No business value

The FKs don't play a special role anymore

Advantages of Using a Surrogate Key

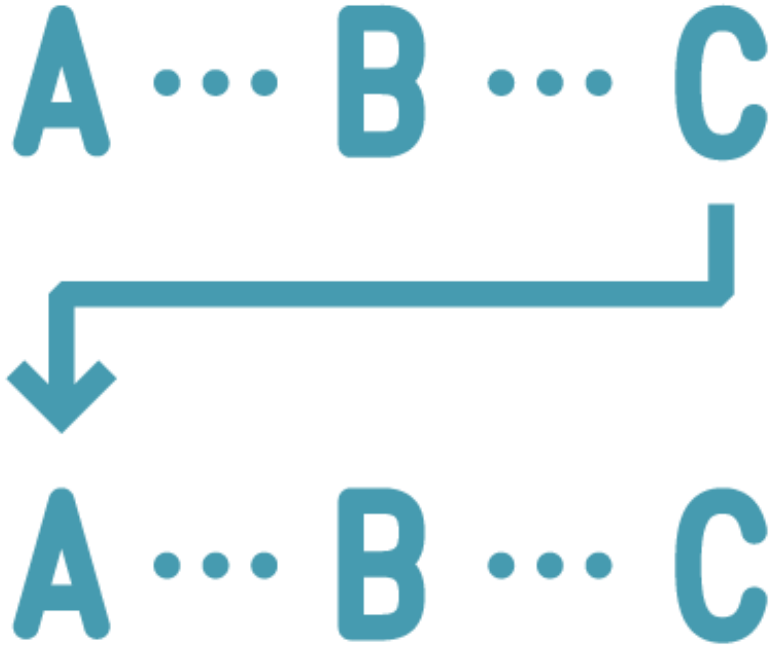
Easier to identify a row in the fact table

Debugging a load that stopped midway

Transforming an update into a combination of

- Inserts
- Deletes

Creating the Foreign Keys



Each dimension has

- A surrogate key (PK)
- A business key (PK of the source system)

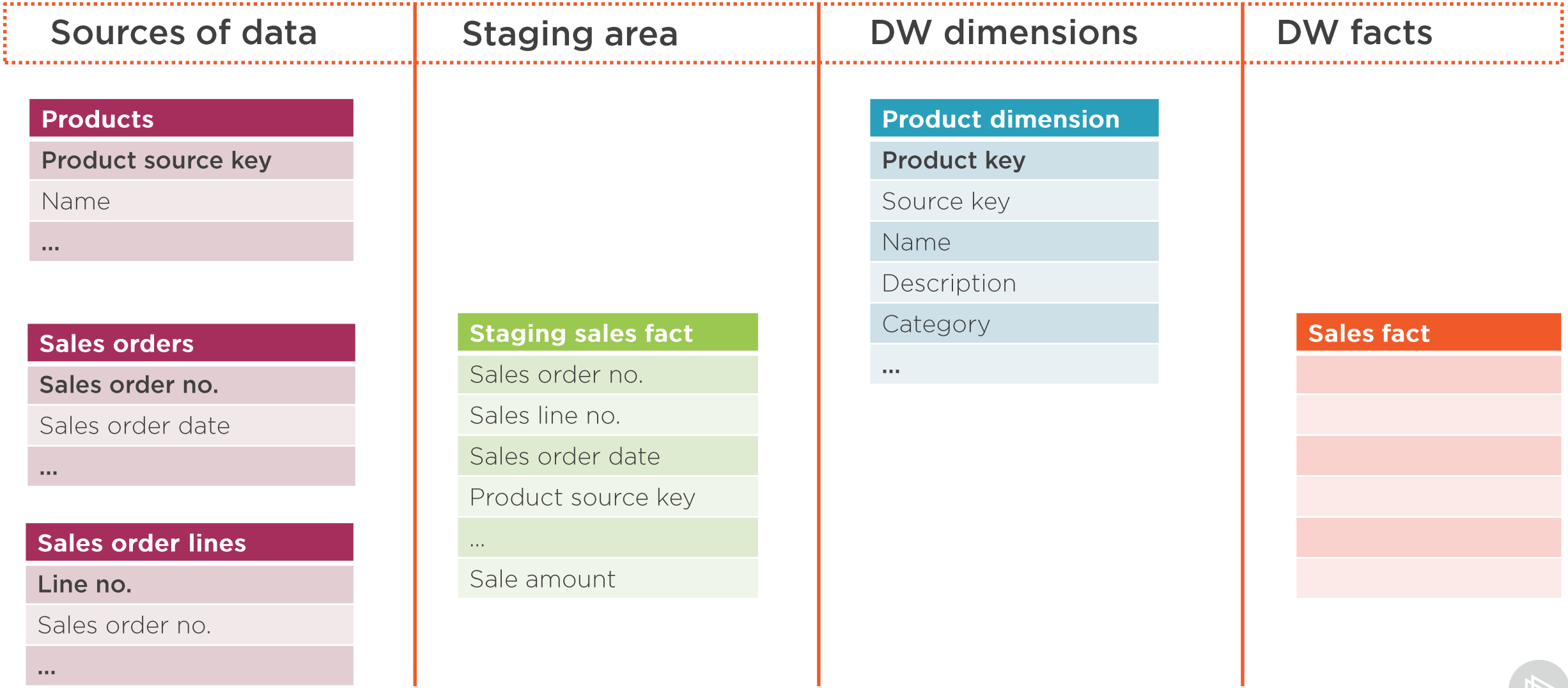
A fact table has

- A primary key
- Foreign keys to its dimensions, linked to the dimensions' surrogate key

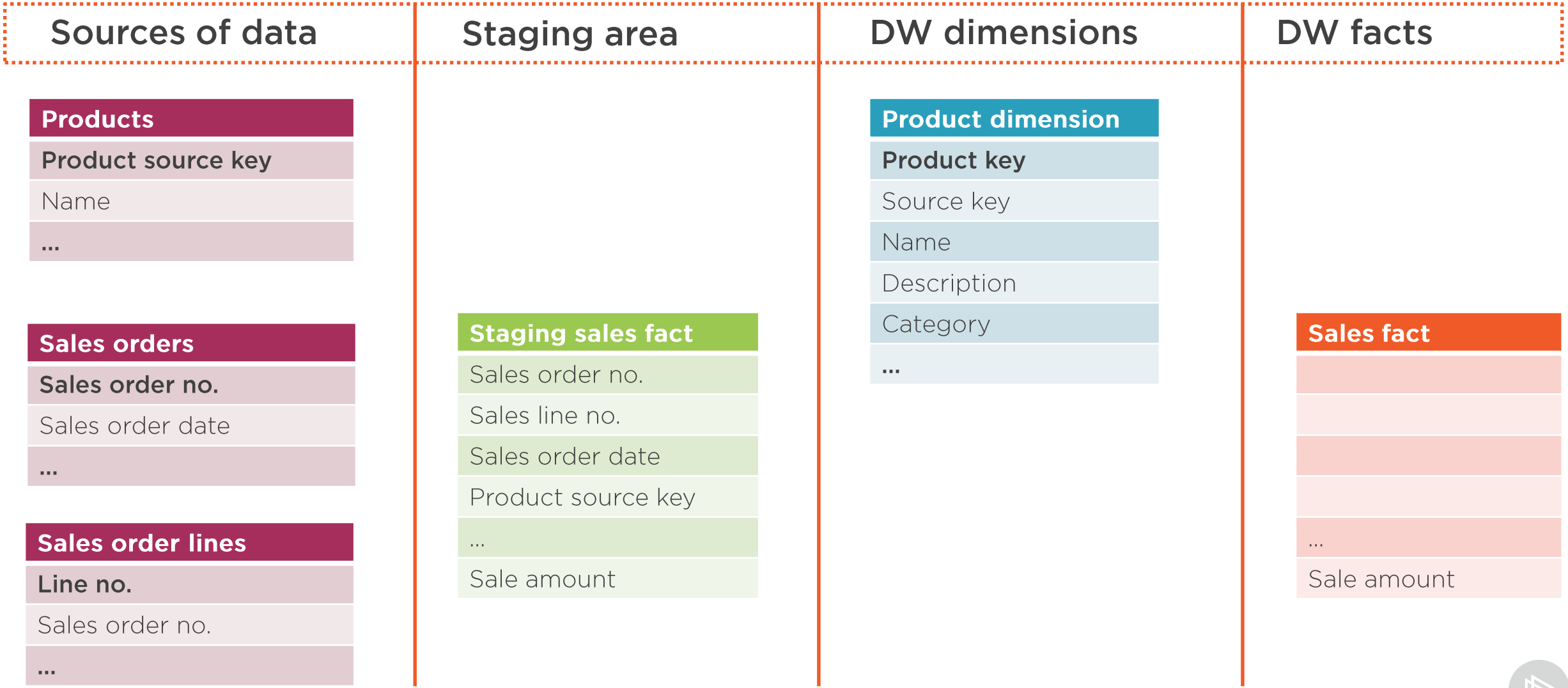
Relationship between the fact and the dimension

- The business key from the dimension
- An intermediary (staging) fact table
  - Fact table data is joined with dimension table data on the real join conditions
  - Staging table stores the PKs from the source tables for dimensions

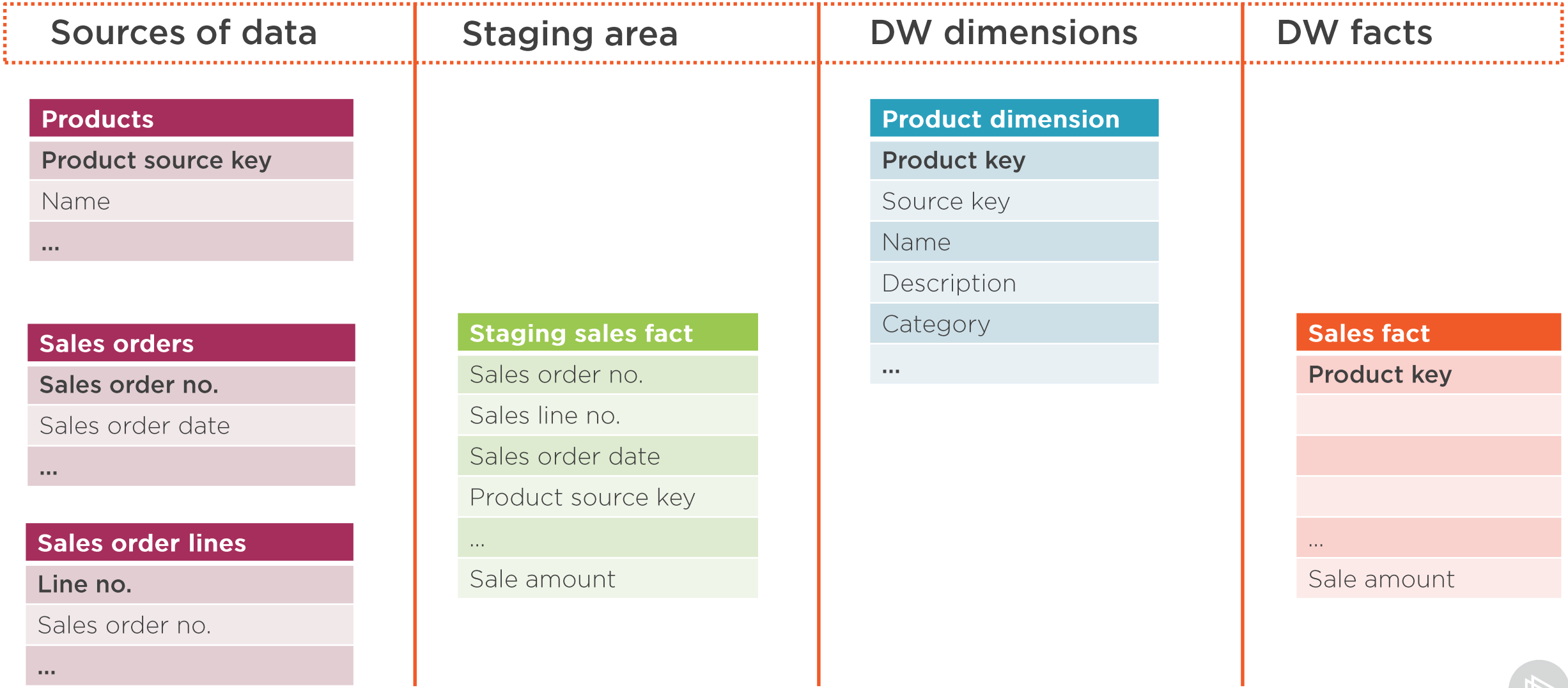
Creating the Foreign Keys



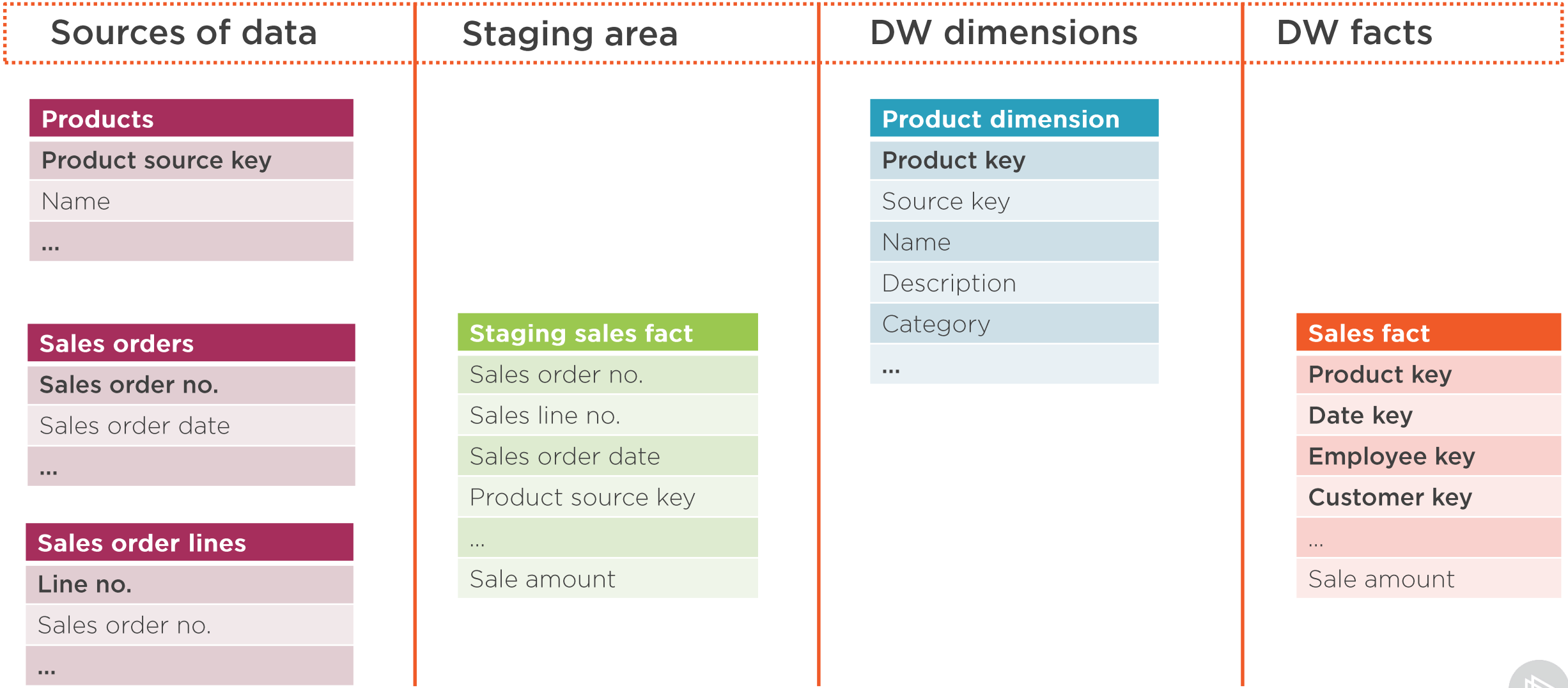
Creating the Foreign Keys



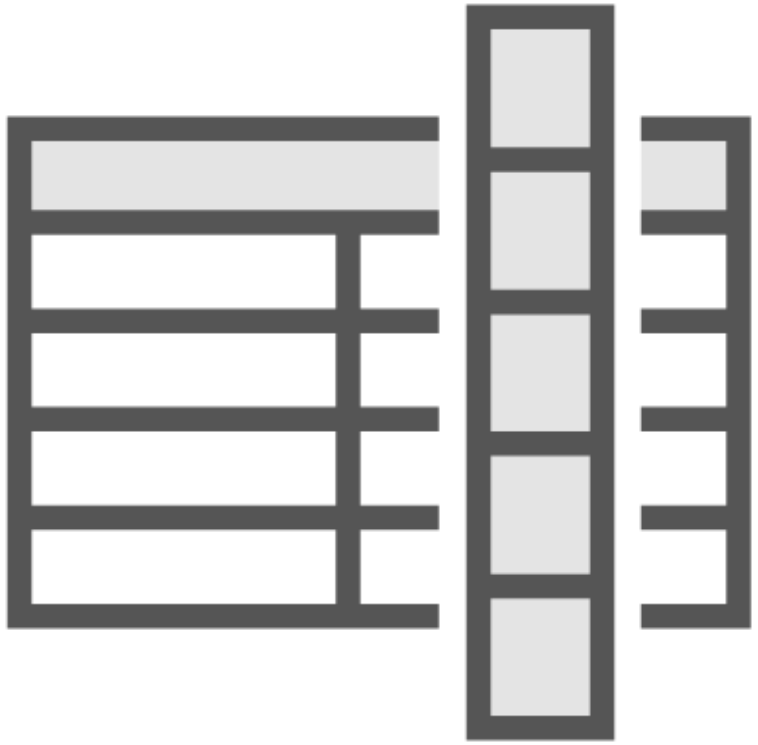
Creating the Foreign Keys



Creating the Foreign Keys



Degenerate Dimensions



Columns added to fact tables

- Different than facts
- Different than keys

Similar to foreign keys, but don't point to any dimension

Examples:

- Sales order number
- Invoice number
- Other transaction numbers

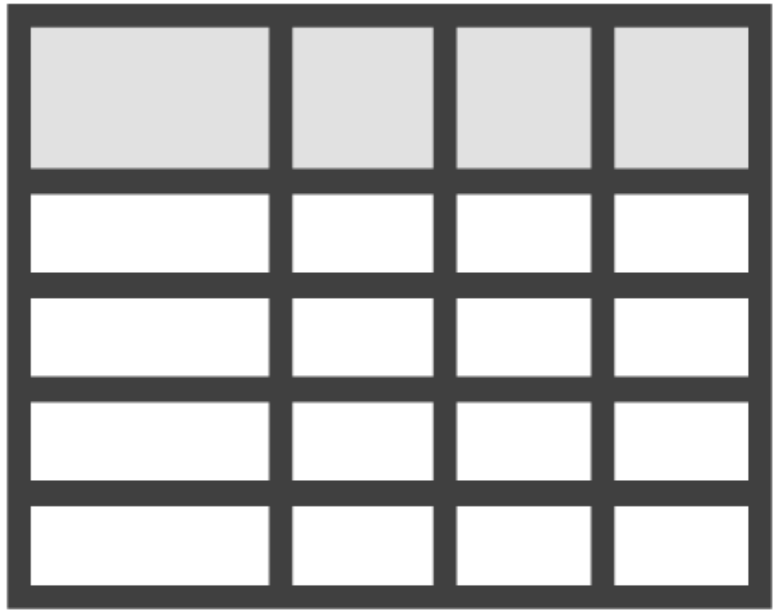
Advantages of degenerate dimensions

- Usually part of the fact's PK
- Group all rows that were part of the same transaction
- Track data back to the operational system

Storing too many may impact the performance of the fact table

Types of Fact Tables

Most Common Types of Fact Tables



Transaction fact tables

Periodic snapshot

Accumulating snapshot

Transaction Fact Tables

The most common type of fact table in a data warehouse

Data inside it represents a process that happened at a certain point in time

A row exists for a dimension member only if the member was involved in a transaction

Properties of Transaction Fact Tables

The granularity is the transaction or transaction line

Tendency to become very large

Relationships with many dimensions

Sparsely populated

The facts are additive

Periodic Snapshot Fact Tables

Show information as it was at the end of a time interval

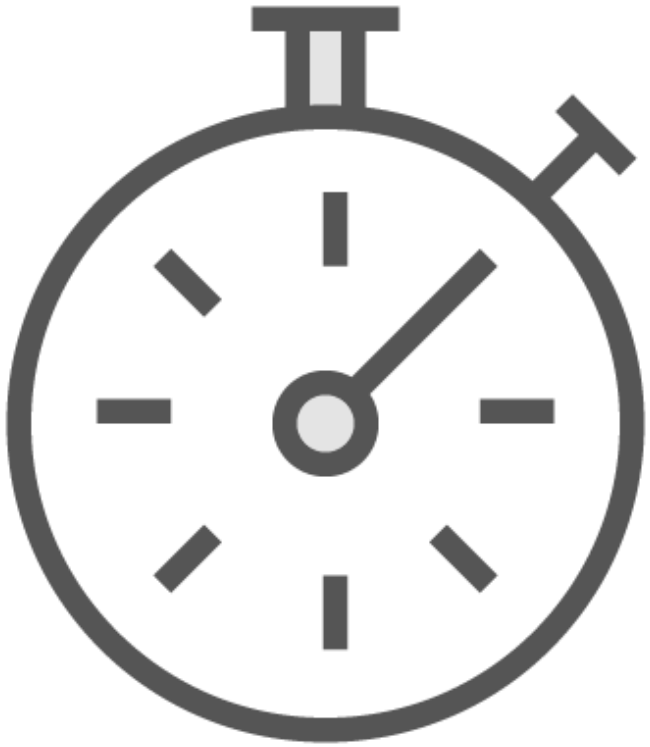
- End of each day
- End of each month, etc.

Are not periodically refreshed

- The new snapshots are added on top of the existing ones
- Are useful for analyzing trends in time

Transactional fact tables can be a starting point for the periodic snapshots

Accumulating Snapshot Fact Tables



Store information about a process that has

- A clear beginning
- A clear end
- A number of intermediary steps

Example: pipeline or workflow processes

- Fulfilling an order
- Progress of a mortgage application
- Status of a support ticket

Accumulating Snapshot Fact Tables

Handling support calls

Accumulating Snapshot Fact Tables

Handling support calls

Accumulating Snapshot Fact Tables

Handling support calls

Created date key	Assigned date key	In progress date key	In testing date key	Done date key
20190404	20190405	20190406	20190412	20190419

Accumulating Snapshot Fact Tables

Handling support calls

Created date key	Assigned date key	In progress date key	In testing date key	Done date key
20190404	20190405	20190406	20190412	20190419
20190407	17530101	17530101	17530101	17530101

Accumulating Snapshot Fact Tables

Handling support calls

Created date key	Assigned date key	In progress date key	In testing date key	Done date key
20190404	20190405	20190406	20190412	20190419
20190407	20190407	17530101	17530101	17530101

Accumulating Snapshot Fact Tables

Handling support calls

Created date key	Assigned date key	In progress date key	In testing date key	Done date key
20190404	20190405	20190406	20190412	20190419
20190407	20190407	20190407	17530101	17530101

Accumulating Snapshot Fact Tables

Handling support calls

Created date key	Assigned date key	In progress date key	In testing date key	Done date key	Duration development	Duration testing
20190404	20190405	20190406	20190412	20190419	6	3
20190407	20190407	20190407	17530101	17530101	NULL	NULL

Accumulating Snapshot Fact Tables

Handling support calls

Created date key	Assigned date key	In progress date key	In testing date key	Done date key	Duration development	Duration testing	Is done
20190404	20190405	20190406	20190412	20190419	6	3	1
20190407	20190407	20190407	17530101	17530101	NULL	NULL	0

Accumulating Snapshot Fact Tables

Handling support calls

Created date key	Assigned date key	In progress date key	In testing date key	Done date key	Duration development	Duration testing	Is done	Current status key
20190404	20190405	20190406	20190412	20190419	6	3	1	6
20190407	20190407	20190407	17530101	17530101	NULL	NULL	0	4

Handling Null Values in Fact Tables

Nulls in the facts

It is alright to have nulls in the facts

The aggregate functions can handle null values

Replacing null with zeros would influence the calculations performed

Value
2
2
null

Average: 2

Value
2
2
0

Average: 1

Nulls in the foreign keys

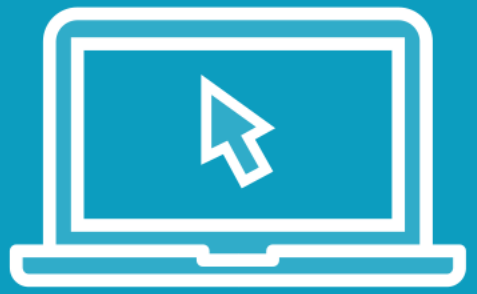
Will produce a referential integrity violation error

This situation must be avoided

Use the "empty row" technique:

- Every dimension should have an empty row, with its own primary key
- When there is no link between fact and dimension, the foreign key column will store the PK of the empty row

Demo



Creating and working with a transaction fact table

- Populate staging table with source data
- Populate fact table with data from staging
- Use the fact table in reports

Summary

Components of a fact table:

- Facts (or numeric measurements)
  - Additive
  - Semi-additive
  - Non-additive
- The primary key
  - Surrogate key
  - Composite key
- Foreign keys for dimensions
- Degenerate dimensions

The most common types of fact tables:

- Transaction fact table
- Periodic snapshot
- Accumulating snapshot

Be aware of the null values in fact tables

- Measures can have null values
- Foreign key columns must not have nulls