Creating and Working with Hierarchies

Ana Voicu

@ana_voicu

Overview

In short, a hierarchy is:

- A data structure
- Created with attributes from a dimension
- Used for data aggregation

Topics elaborated in this chapter:

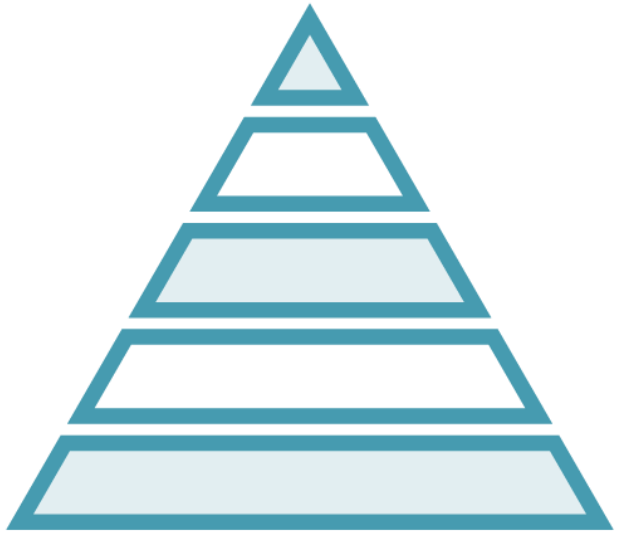
- What is a hierarchy?
- Why is it useful?
- What does drilling-down mean?
- Implementing a hierarchy in a data warehouse

What Is a Hierarchy?

What Is a Hierarchy?

Data structure with multiple levels

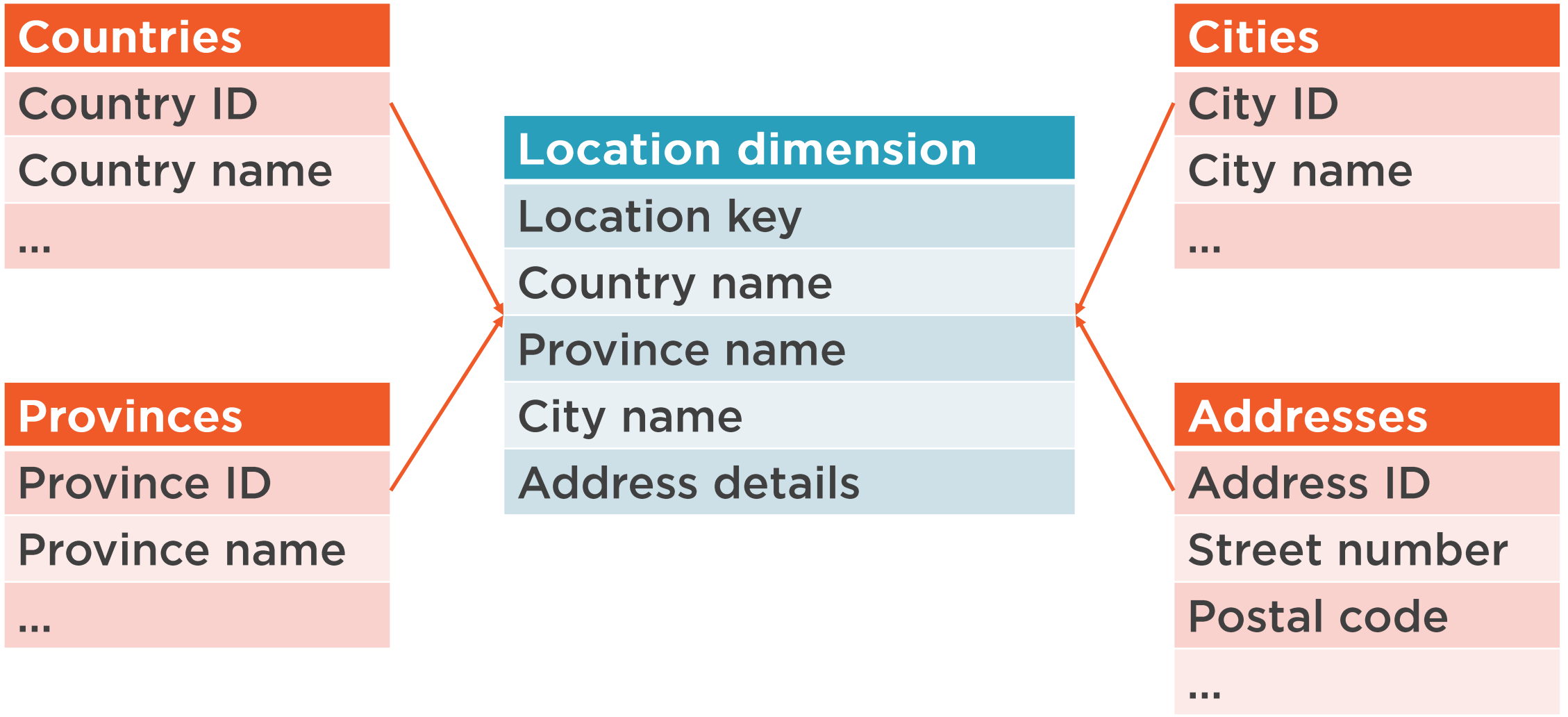
- The levels are dimension attributes
- The elements on each level are called nodes



Similar to a pyramid

- Bottom level is the weakest
- Highest level is at the top

A Hierarchy in the Location Dimension



A Hierarchy in the Location Dimension

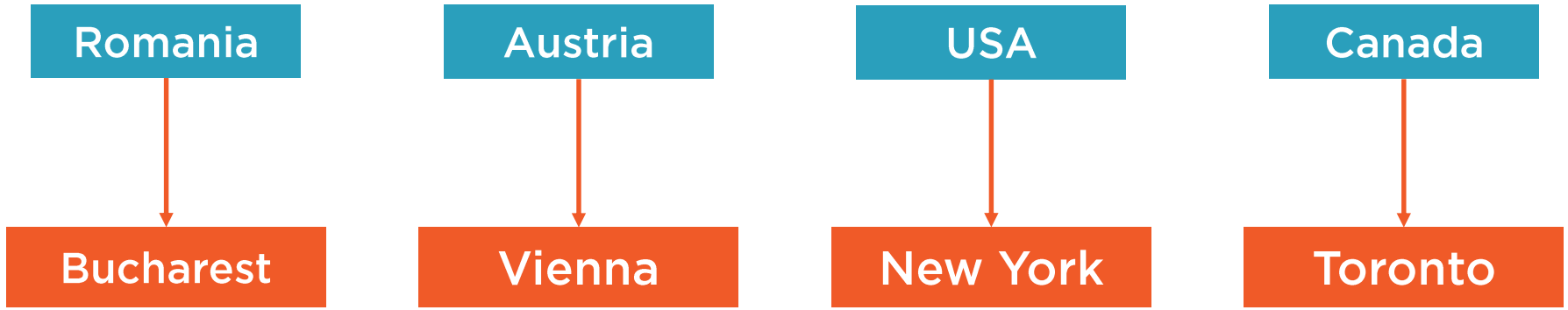
Bucharest

Vienna

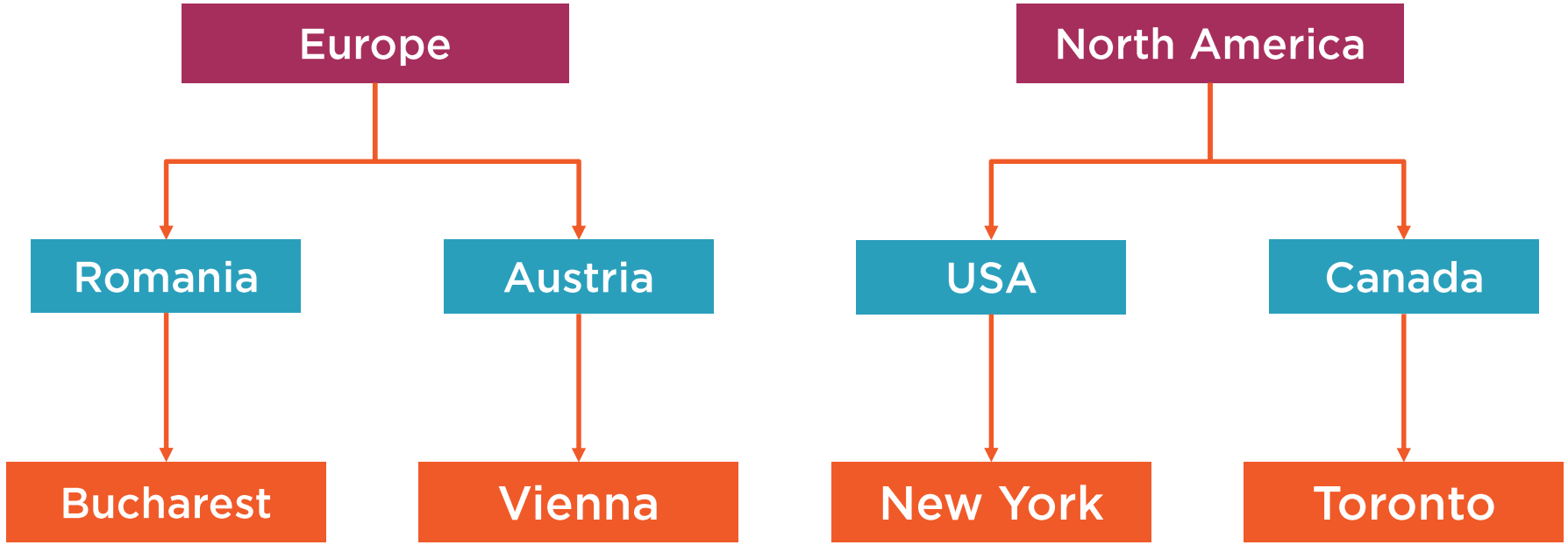
New York

Toronto

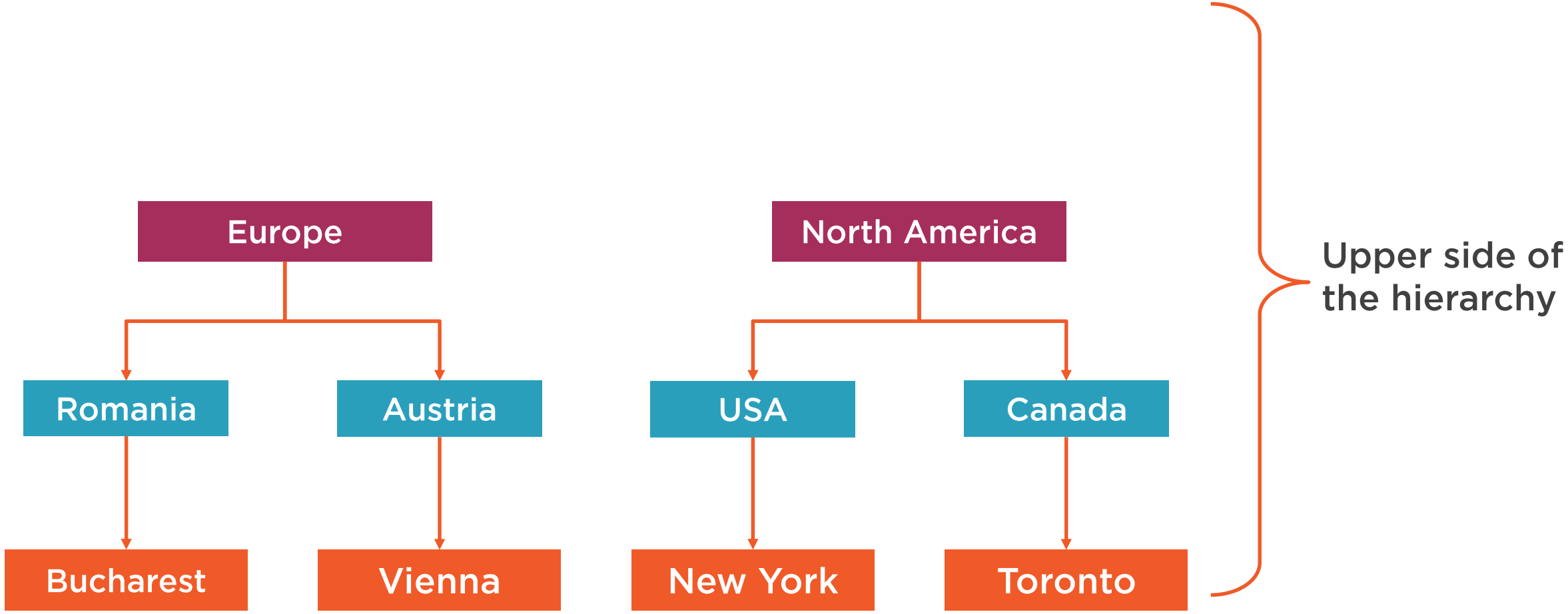
A Hierarchy in the Location Dimension



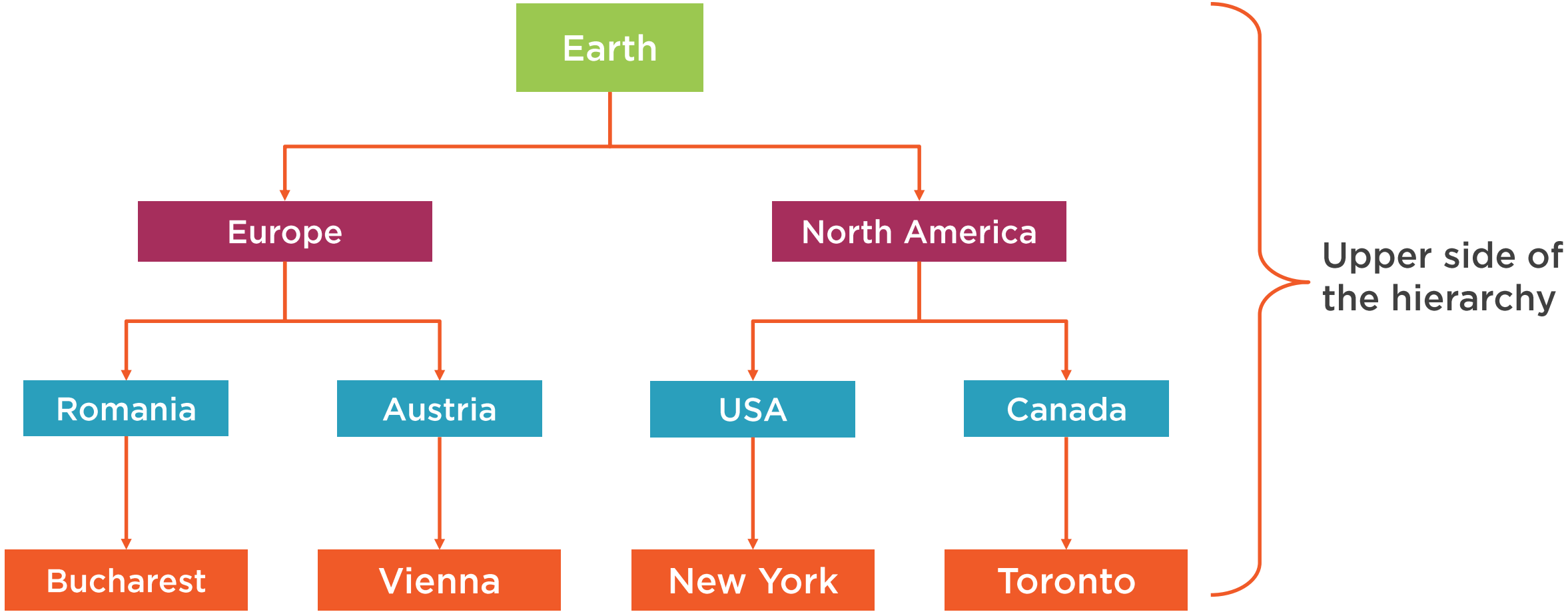
A Hierarchy in the Location Dimension



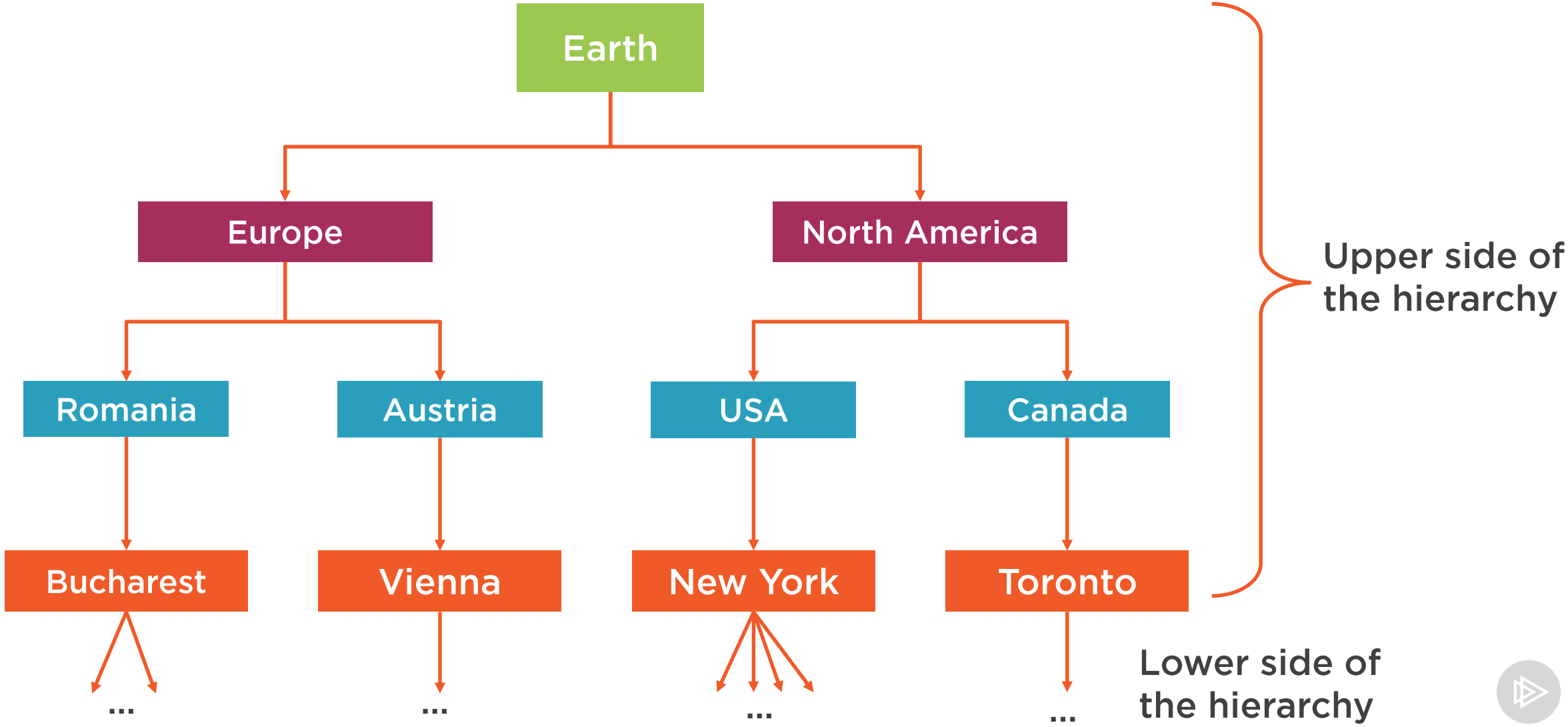
A Hierarchy in the Location Dimension



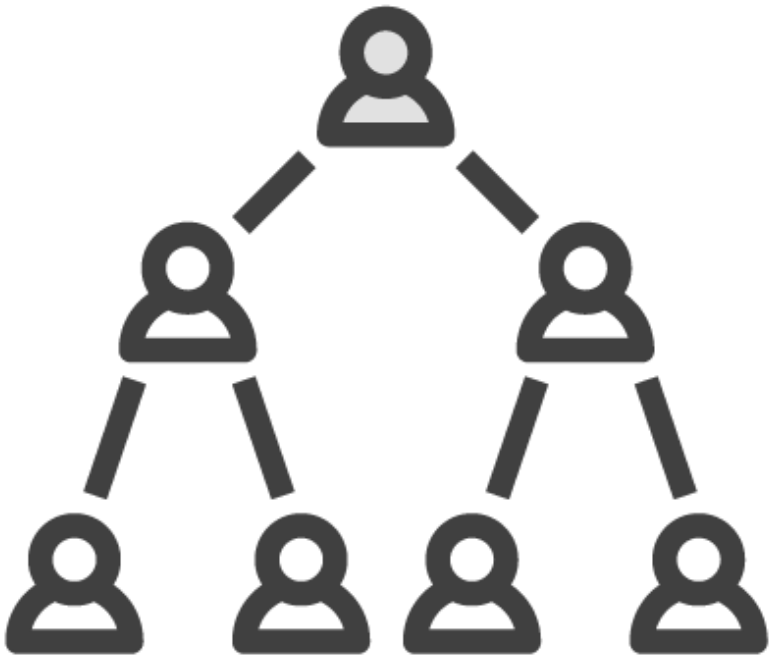
A Hierarchy in the Location Dimension



A Hierarchy in the Location Dimension



What Is a Hierarchy?



Data structure

- Attributes of a dimension are organized together
- Going down the hierarchy nodes -> more details
- Going up the hierarchy -> summarized data

Advantages of Using Hierarchies

Structure

- Important dimensions have plenty attributes
- Analyzing important data can become overwhelming
- Hierarchies provide order in the data

Multiple perspectives

Merchandise hierarchy

Advantages of Using Hierarchies

Structure

- Important dimensions have plenty attributes
- Analyzing important data can become overwhelming
- Hierarchies provide order in the data

Multiple perspectives

Merchandise hierarchy

- Department
  - Category
    - Subcategory
      - Product name

Advantages of Using Hierarchies

Structure

- Important dimensions have plenty attributes
- Analyzing important data can become overwhelming
- Hierarchies provide order in the data

Multiple perspectives

Merchandise hierarchy

- Department
  - Category
    - Subcategory
      - Product name

Packaging hierarchy

Advantages of Using Hierarchies

Structure

- Important dimensions have plenty attributes
- Analyzing important data can become overwhelming
- Hierarchies provide order in the data

Multiple perspectives

Merchandise hierarchy

- Department
  - Category
    - Subcategory
      - Product name

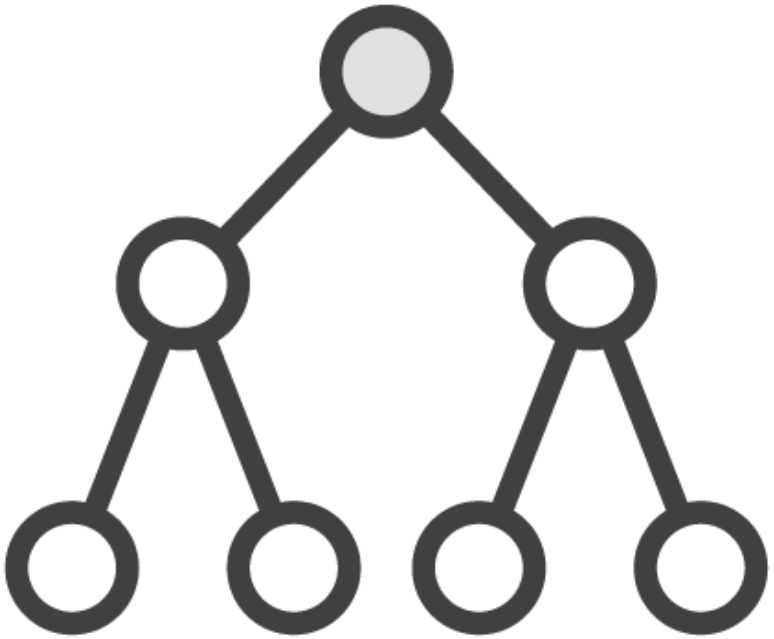
Packaging hierarchy

- Package type
  - Package size
    - Product name

Zoom in/zoom out

- Visualize summarized data and detailed data
- Go as deep as the business requires it

Types of Hierarchies



Fixed-depth (balanced)

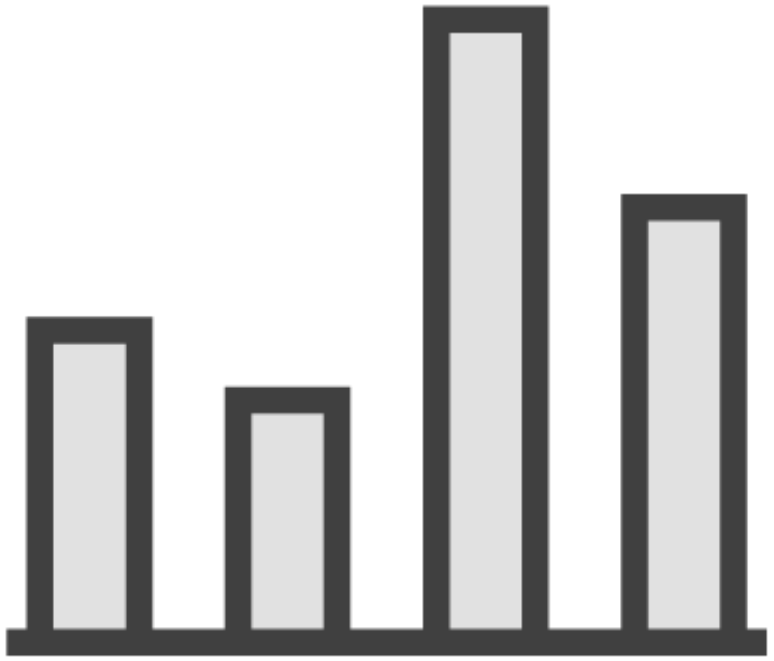
- Fixed number of levels
- Easy to create and work with

Variable-depth (unbalanced or ragged)

- Uneven number of levels
- Creating them is more complex task

Drilling Down a Hierarchy

Data Warehouse Analysis



Minimum requirements for doing data warehouse analysis

- One fact table
- One dimension table

Example: sales per product report

- Sales fact
- Product dimension

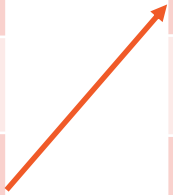
Example of Drilling Down

Sales fact
Store key
Date key
Product key
Employee key
Customer key
...
Transaction #
Unit price
Quantity
Amount

Example of Drilling Down

Sales fact
Store key
Date key
Product key
Employee key
Customer key
...
Transaction #
Unit price
Quantity
Amount

Product dimension
Product key
Product name
Department
Category
Subcategory
Package size
Package type
Description
Unit of measure
...



Example of Drilling Down

Sales fact
Store key
Date key
Product key
Employee key
Customer key
...
Transaction #
Unit price
Quantity
Amount

Product dimension
Product key
Product name
Department
Category
Subcategory
Package size
Package type
Description
Unit of measure
...

Merchandise hierarchy

- Department
  - Category
    - Subcategory
      - Product name

Example of Drilling Down

Sales

80.000

Department

Sales

Bakery	Sweets	Beverages
25.000	40.000	15.000

Example of Drilling Down

Department
Category
Sales

Bakery			Sweets			Beverages	
Bread	Pie	Croissant	Cookie	Cake	Candy	Juice	Tea
7.000	13.000	5.000	8.700	6.300	5.000	7.000	6.600

Example of Drilling Down

Department

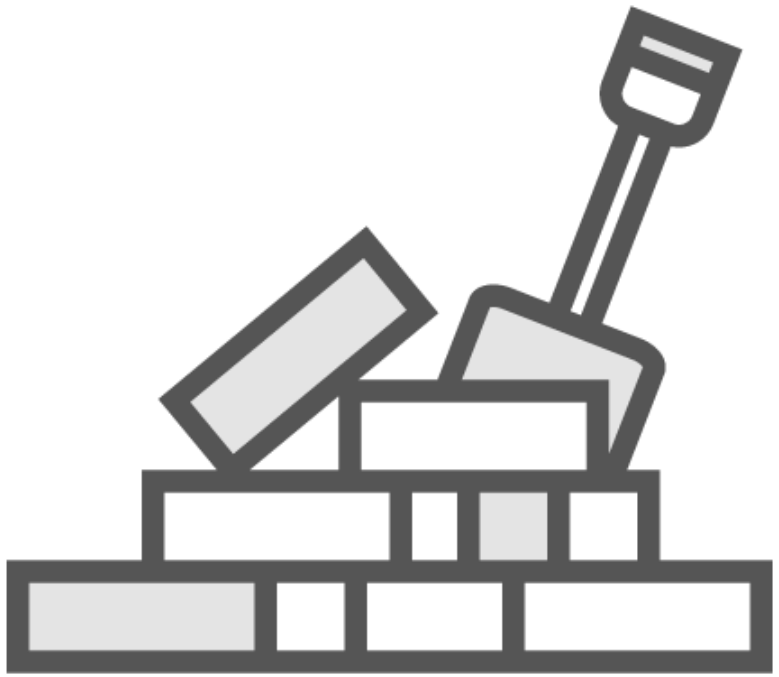
Category

Subcategory

Sales

Bakery							
Bread			Pie			Croissant	
Baguette	Pita	Banana br.	Cream	Fruit	Custard	Sweet	Salty
2.000	3.000	2.000	2.500	2.200	8.300	3.000	2.000

Drilling Down - Summary



Drilling down

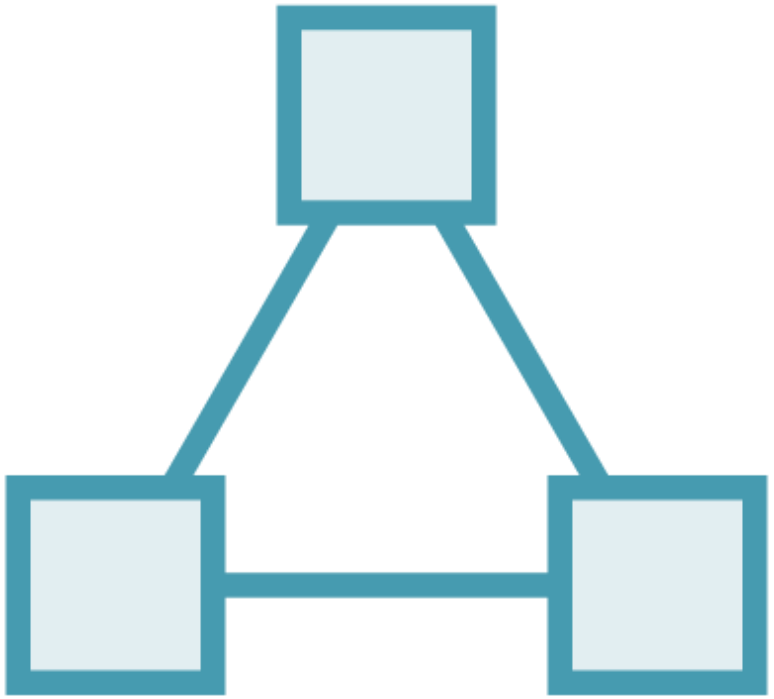
- Adding another member of the dimension to the report
- The member doesn't need to be part of a hierarchy

Drilling up/rolling up

- Taking out an attribute of a dimension from a report

Fixed-depth Positional Hierarchies

Fixed-depth Positional Hierarchies



The number of levels is known upfront

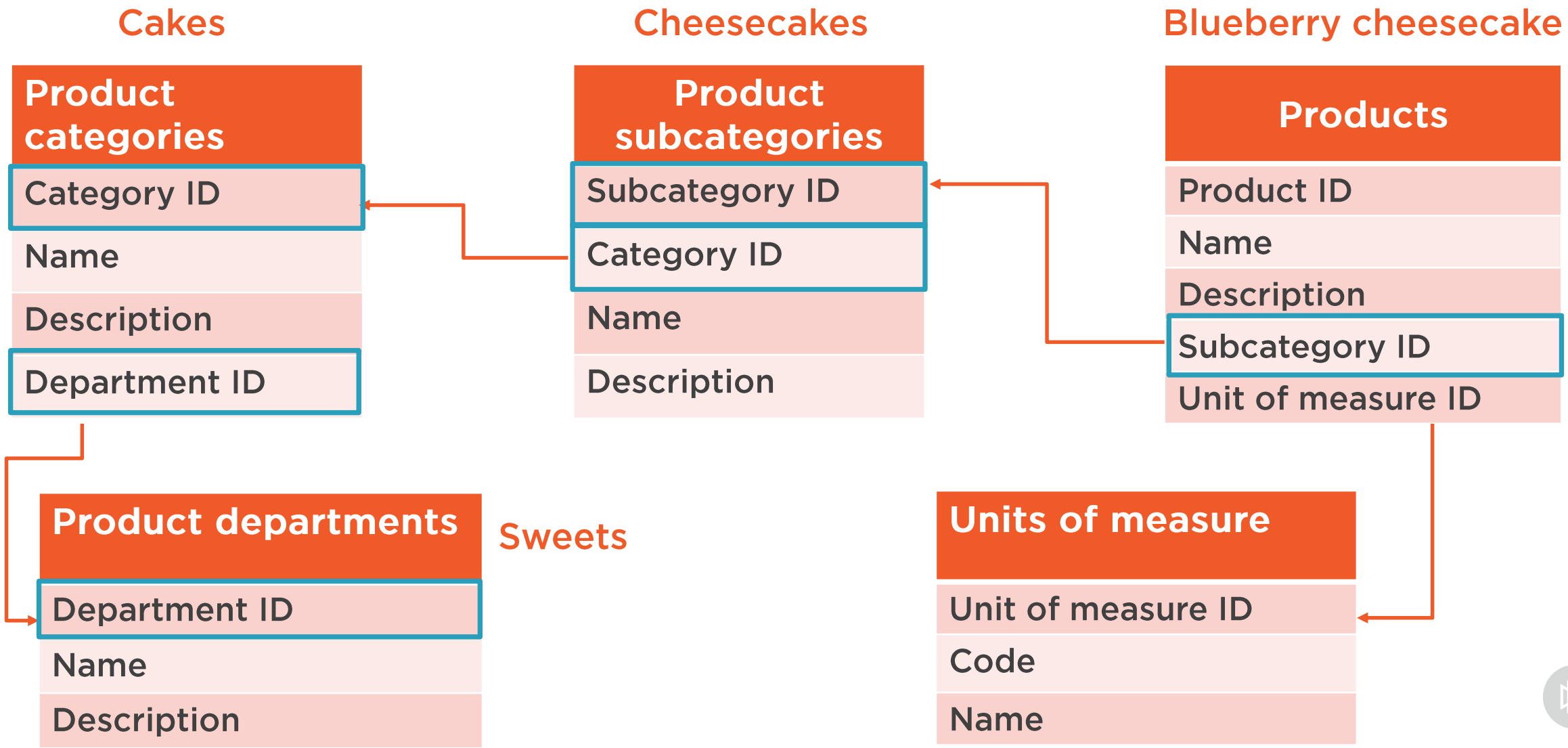
The levels are attributes in the dimension table

It is a series of many-to-one relationships

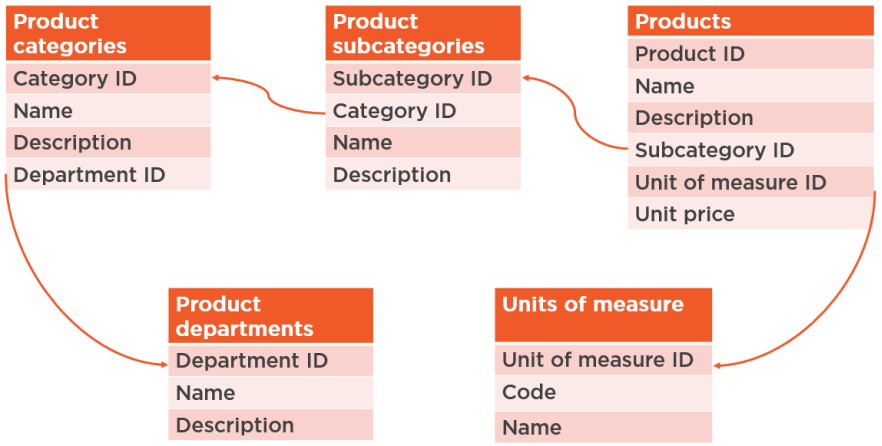
Advantages

- Easy to navigate
- Offers predictable results
- No impact on performance

Product Information from Different Tables



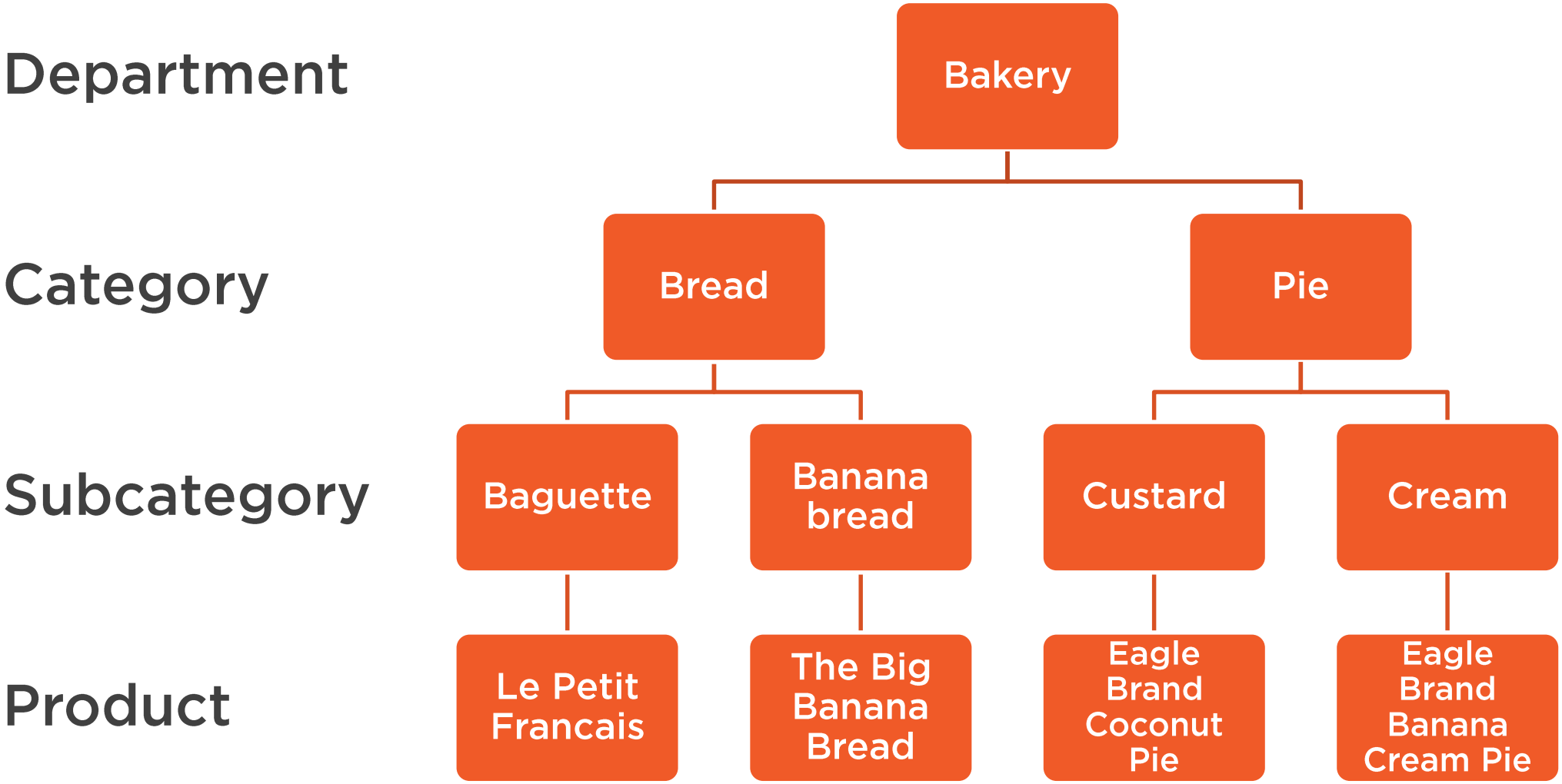
Product Information from the Data Warehouse



Keep in mind:

- Don’t create many snowflake designs
- Most relationships should be from fact tables to the dimensions

Creating the Merchandise Hierarchy

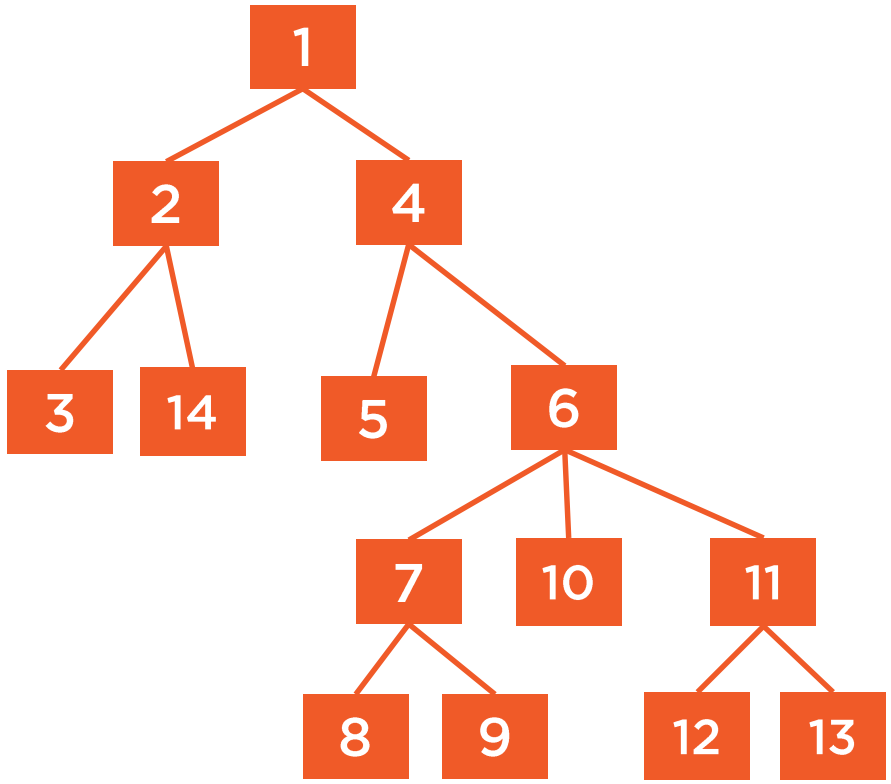


Demo

Creating a fixed-depth positional hierarchy

Variable-depth Positional Hierarchies

Characteristics of Variable-depth Hierarchies



The number of levels is not known at design time

- Example: an organizational chart

Are more complex structures (compared to the fixed-depth hierarchies)

Should be used with moderation

Classification of Variable-depth Hierarchies

Slightly ragged

Ragged, created with a hierarchy bridge

Ragged, created with pathstring attributes

Slightly Ragged Hierarchies

The number of levels is not known beforehand

The range in depth is small

Geographic hierarchies are slightly ragged

Example of a Slightly Ragged Hierarchy

Location hierarchy

- Country
  - Province (or state)
    - City
      - Neighborhood
        - Address

Examples of data with missing levels

Singapore (country and city)

Vatican (independent city-state)

Small cities, that don't have neighborhoods

Fitting Data into a Slightly Ragged Hierarchy

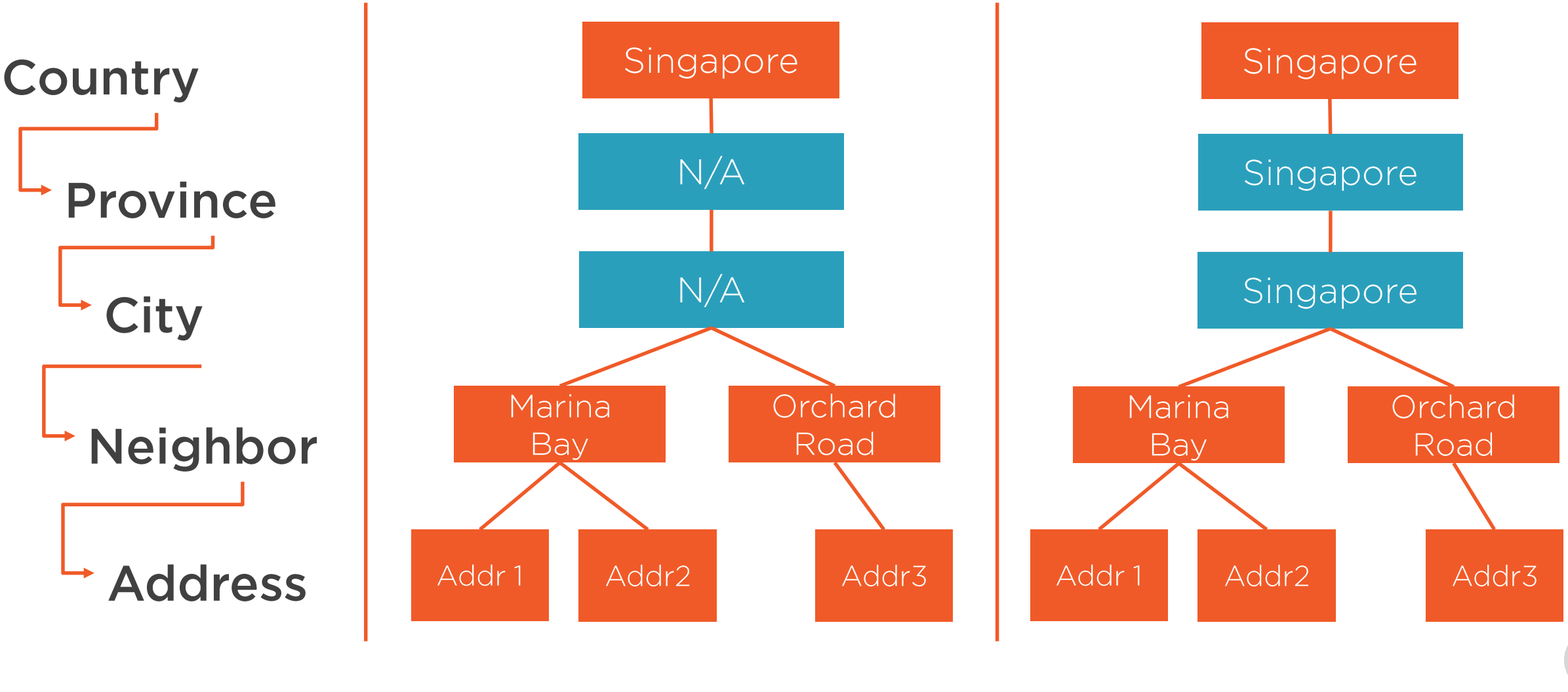
Step 1: Create all possible levels of the hierarchy

Step 2: Fill in the missing values per each level with:

- An expression, similar to “not applicable”
- The value of the next parent member

Step 3: Handle the ragged hierarchy as a fixed-depth one

Populating a Slightly Ragged Hierarchy



Demo

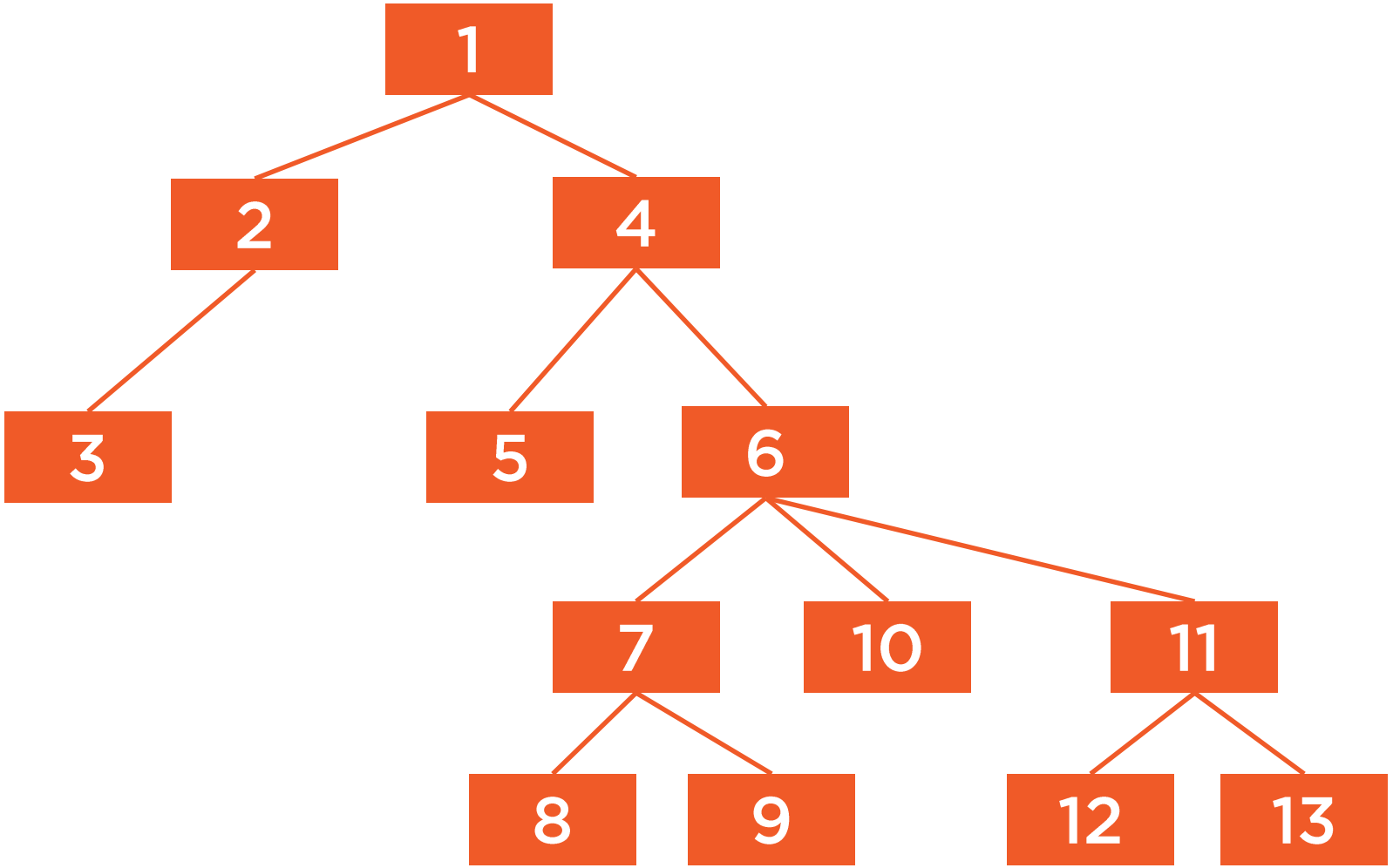
Creating and working with a slightly ragged hierarchy

- Based on the Location dimension

Ragged Hierarchies

Example of a Ragged Hierarchy

The organizational chart



Implementing a Ragged Hierarchy Using a Bridge

A row for each possible path in the hierarchy is stored in a table

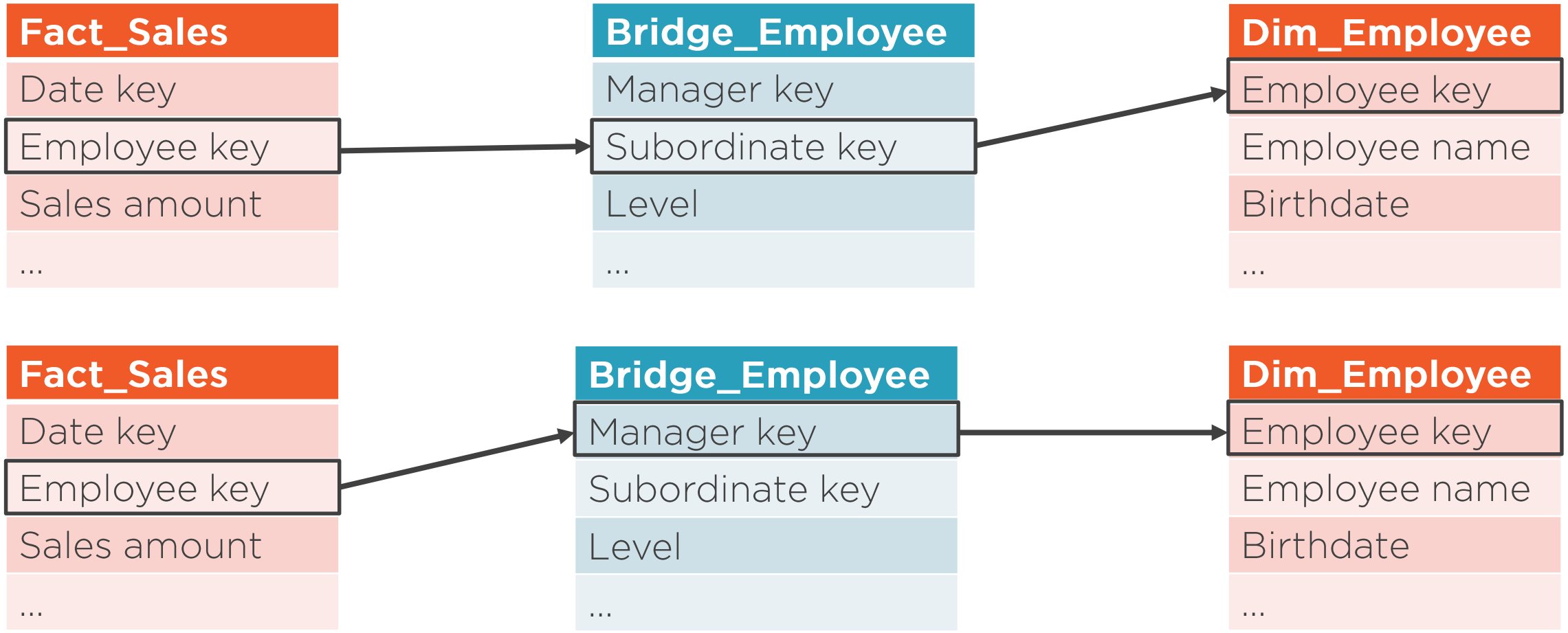
Columns in the bridge table

- ID of the parent
- ID of the child
- Number of levels between them
- Whether the node it a top node or bottom node
- Other information relevant for analysis

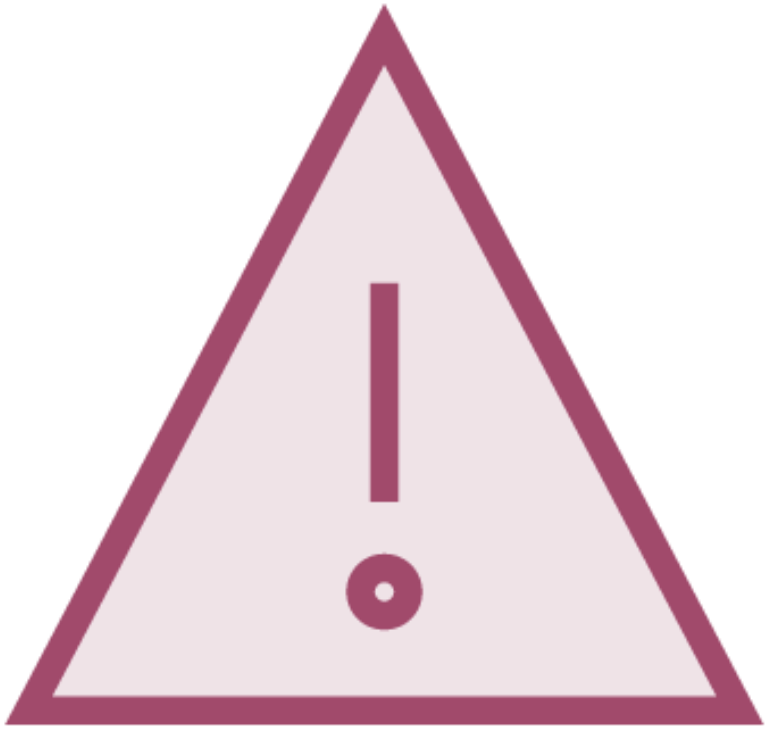
Example of a Hierarchy Bridge Table

Parent key	Child Key	Level	Top flag	Bottom flag
Julia	Julia	0	Y	N
Julia	Marc	1	N	N
Julia	Theodora	2	N	Y
Julia	Greg	2	N	Y
Marc	Marc	0	Y	N
Marc	Theodora	1	N	Y
Marc	Greg	1	N	Y
Theodora	Theodora	0	Y	Y
Greg	Greg	0	Y	Y

Linking the Fact and the Dimension Table



Limitations of the Hierarchy Bridge

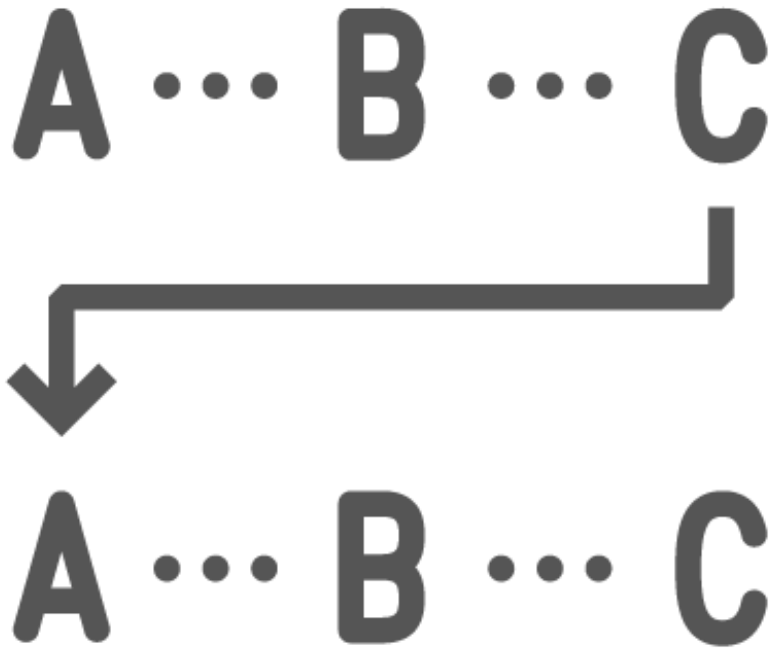


Can grow a lot in size

Performance impact

Difficult to work with

Ragged Hierarchies Created with Pathstring Attributes

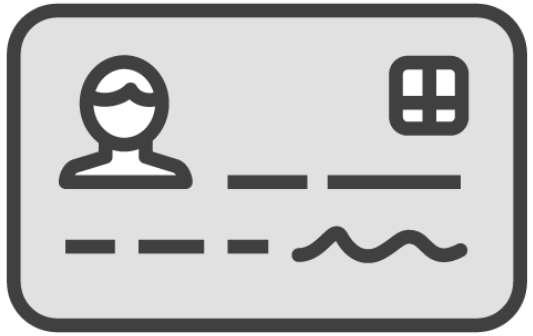


Alternative to the bridge table

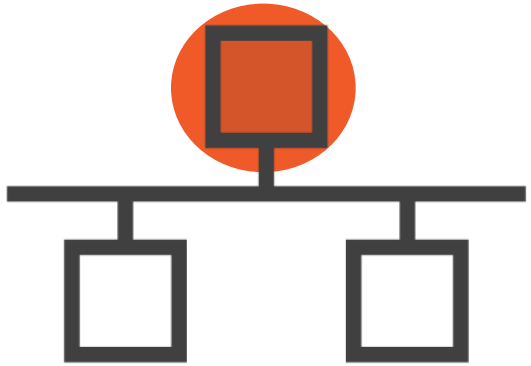
The pathstring attribute:

- A special attribute created in the dimension
- A string of characters
- Consists of all the parents of a member from the top of the hierarchy

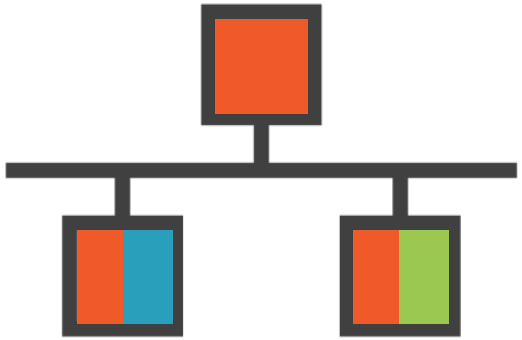
Creating a Hierarchy with Based on Pathstring Attribute



Each node is labeled with a unique value

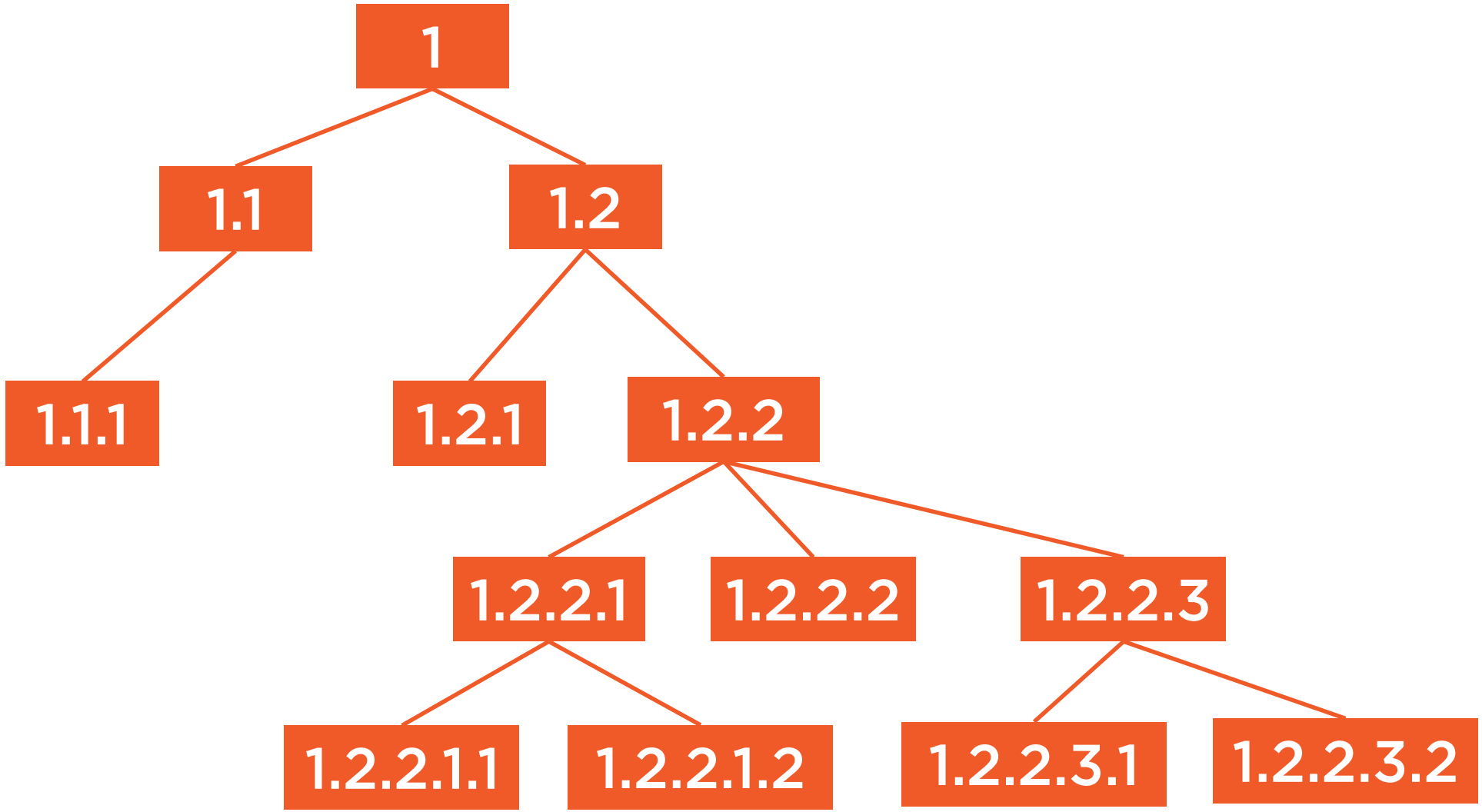


The pathstring of the root node is its unique label

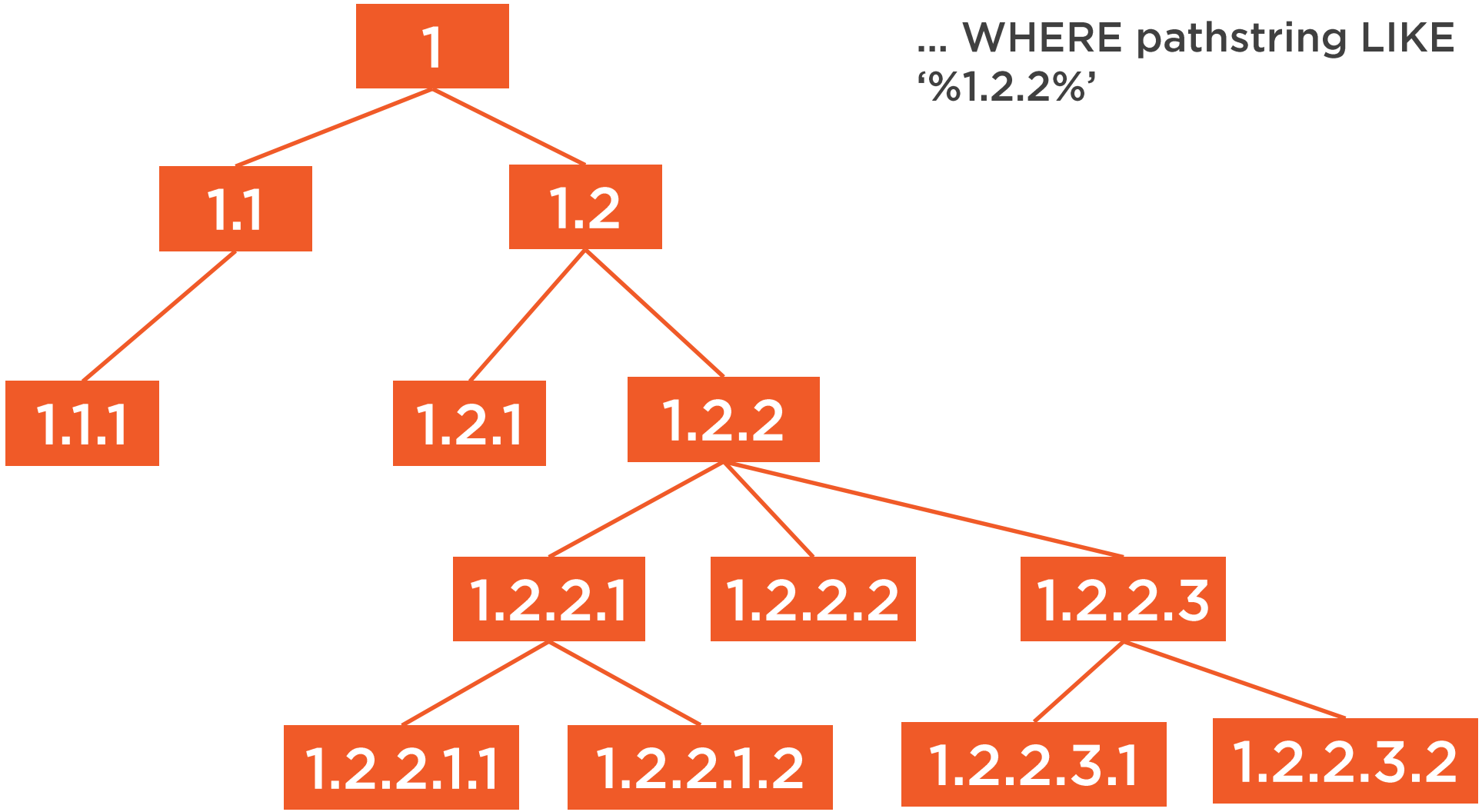


The pathstring on one level includes the pathstring of the parent

Hierarchy Based on Pathstring Attribute

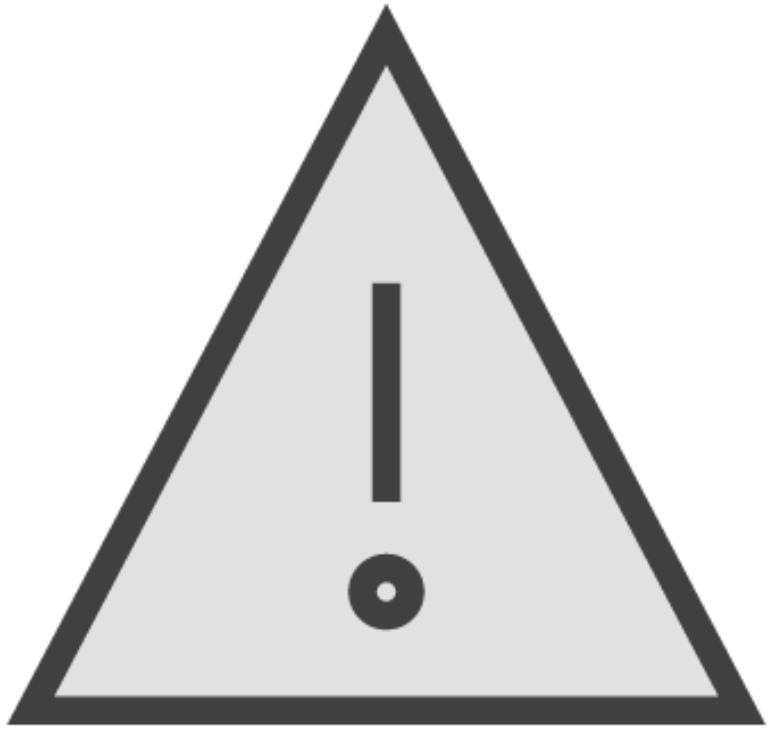


Hierarchy Based on Pathstring Attribute



... WHERE pathstring LIKE '%1.2.2%'

Limitations of the Pathstring Attribute



Vulnerable to structure changes

- If a member is moved within the organization (or added or deleted)
- Entire hierarchy must be relabeled

Complicated to use by business users

- Accessing the database directly
- Using the ‘LIKE’ operator in SQL queries

Summary

There is no “best solution” to this problem

A good enough solution is generated by having a clear understanding of

- The data available
- The business requirements

Summary

Hierarchy definition

- Data structures with multiple levels
- Levels are formed with attributes of a dimension

Advantages of hierarchies

- Structure the data
- Multiple perspectives of the same data
- Easily visualize summarized or detailed data

Types of hierarchies

- Fixed-depth
- Variable-depth
  - Slightly ragged
  - Ragged