

Chapter 7

- Attribute hierarchies
- Snowflake design
- Outriggers
- Refer to pages 147, 157-159, 163-170

Attribute hierarchies

An attribute hierarchy describes parent–child relationships between groups of attributes within a dimension

Multiple attribute hierarchies may exist

Aside: <https://danischnider.wordpress.com/2017/05/31/attribute-dimensions-and-hierarchies-in-oracle-12c/>

Attribute hierarchies

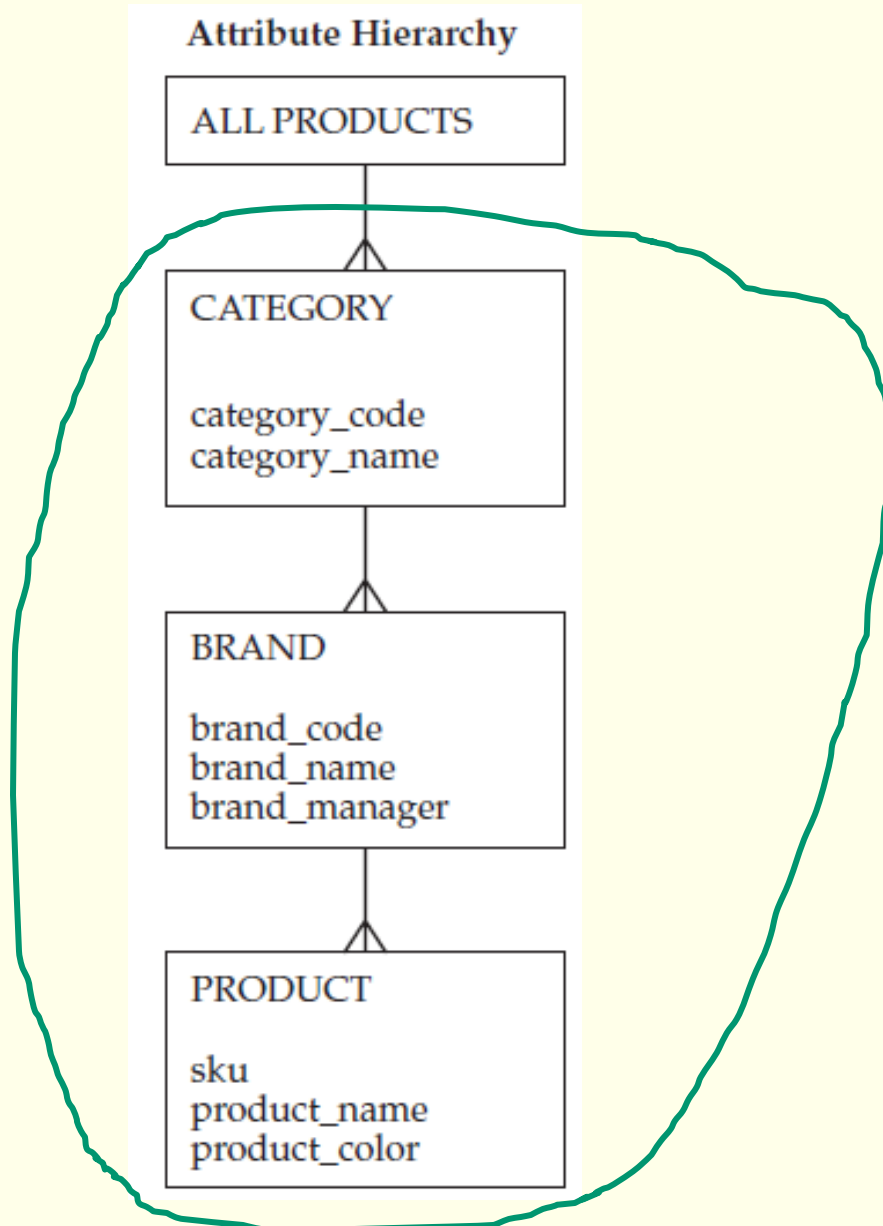
Product Dimension Table

PRODUCT
product_key sku product_name product_color brand_code brand_name brand_manager category_code category_name

This product dimension contains a product hierarchy that could be drawn out as shown →

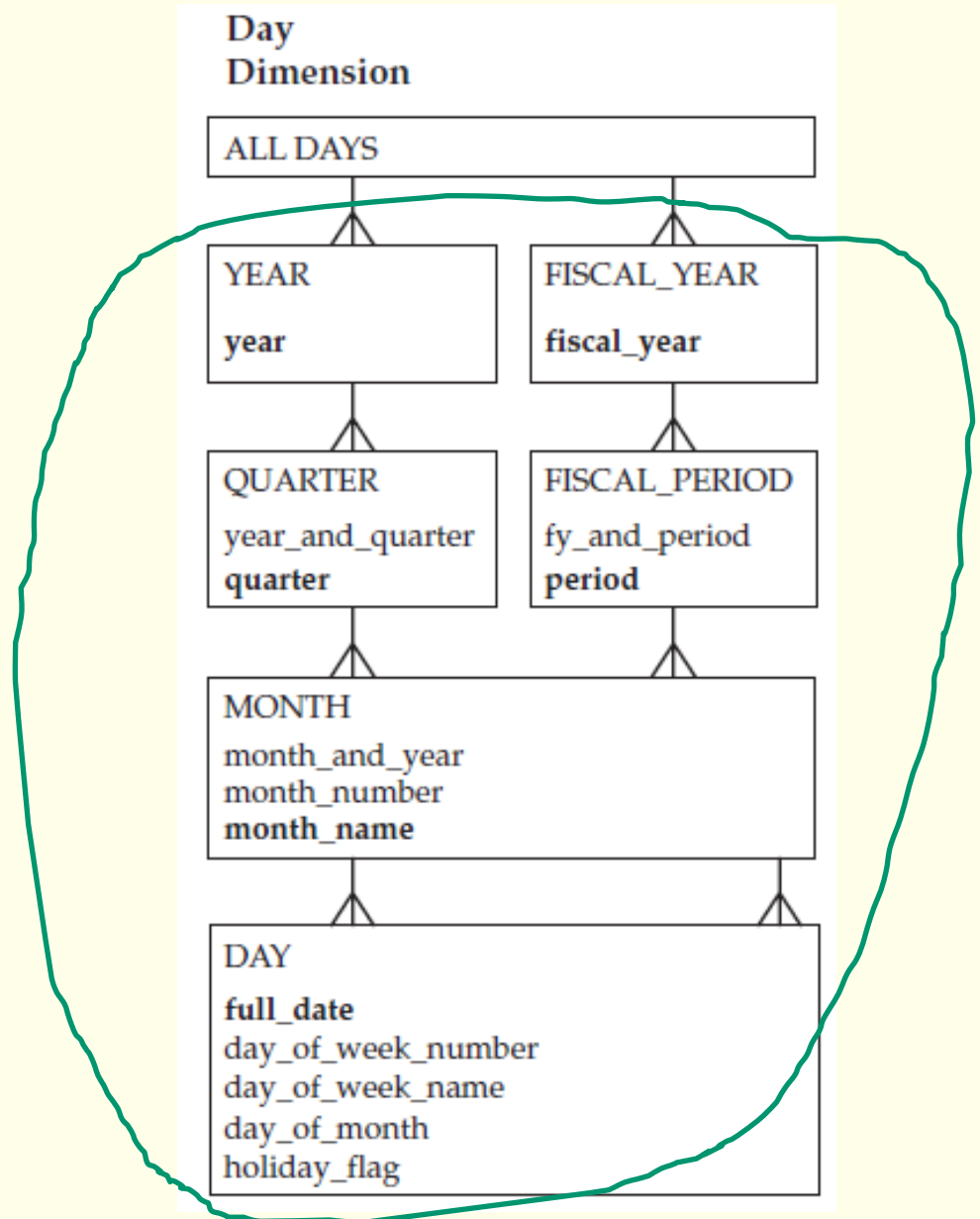
Attribute hierarchies

Products are organized into brands which are organized into categories



Attribute hierarchies

Multiple hierarchies can exist in a single dimension



Attribute hierarchies

Alternatively some describe a hierarchy as:

All Products (1) \rightarrow Categories (25) \rightarrow Brands (650) \rightarrow Products (8000)

The 8000 products are organized into 25 categories. At the next level there are 650 brands, and at the last level we have 8000 products.

A Day dimension :

All Days (1) \rightarrow Years (5) \rightarrow Quarters (20) \rightarrow Months (60) \rightarrow Days (1826)

Instance hierarchies

In some cases a hierarchy exists but can't be shown as an attribute hierarchy (**Ch 10**)

e.g. The supervises relationship (Northwind). An employee may supervise other employees.

Snowflaking

When relationships between attributes in a dimension are expressed explicitly (i.e. normalizing a dimension based on hierarchies realized as physical tables), the result is called a snowflake schema.

The text advises against normalization of dimension tables

On rare occasions, a limited form of snowflaking is employed

- help resolve unmanageable row length
- ensure consistent representation of repeating attributes.

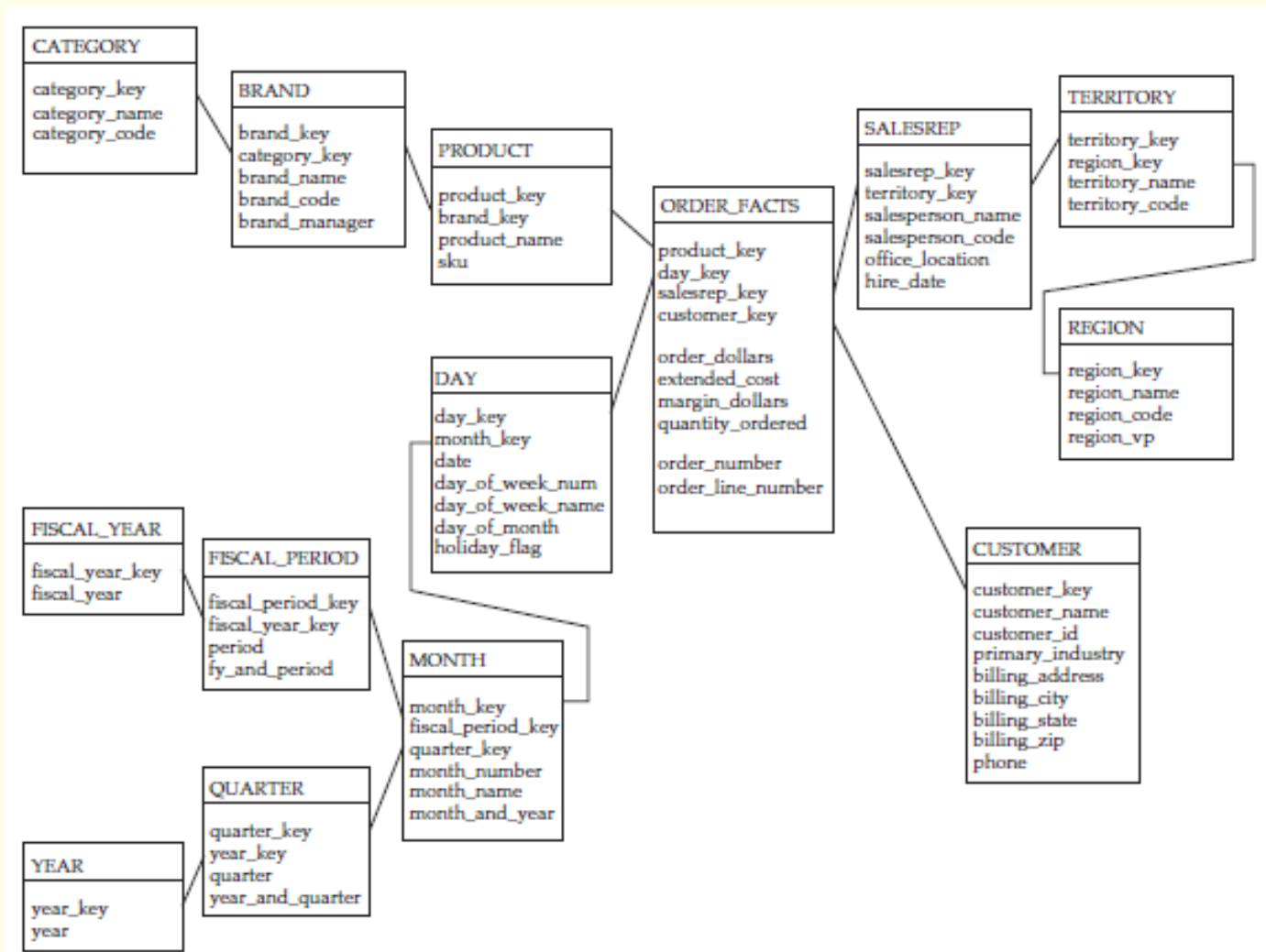
Snowflaking

For those trained in ER modeling, the snowflake reflects some best practices learned in the service of operational systems - it is of little utility for an analytic database, aside from saving some space.

Figure 7-5 illustrates a snowflaked design

Ch 9 will discuss Multi-valued attributes and bridges

Ch 10 will discuss Recursive hierarchies and bridges



Snowflaked design

Outriggers

Repeating attributes or groups of attributes *might* be better handled using Outriggers – to control confusion that may arise with similarly-named attributes

Outriggers

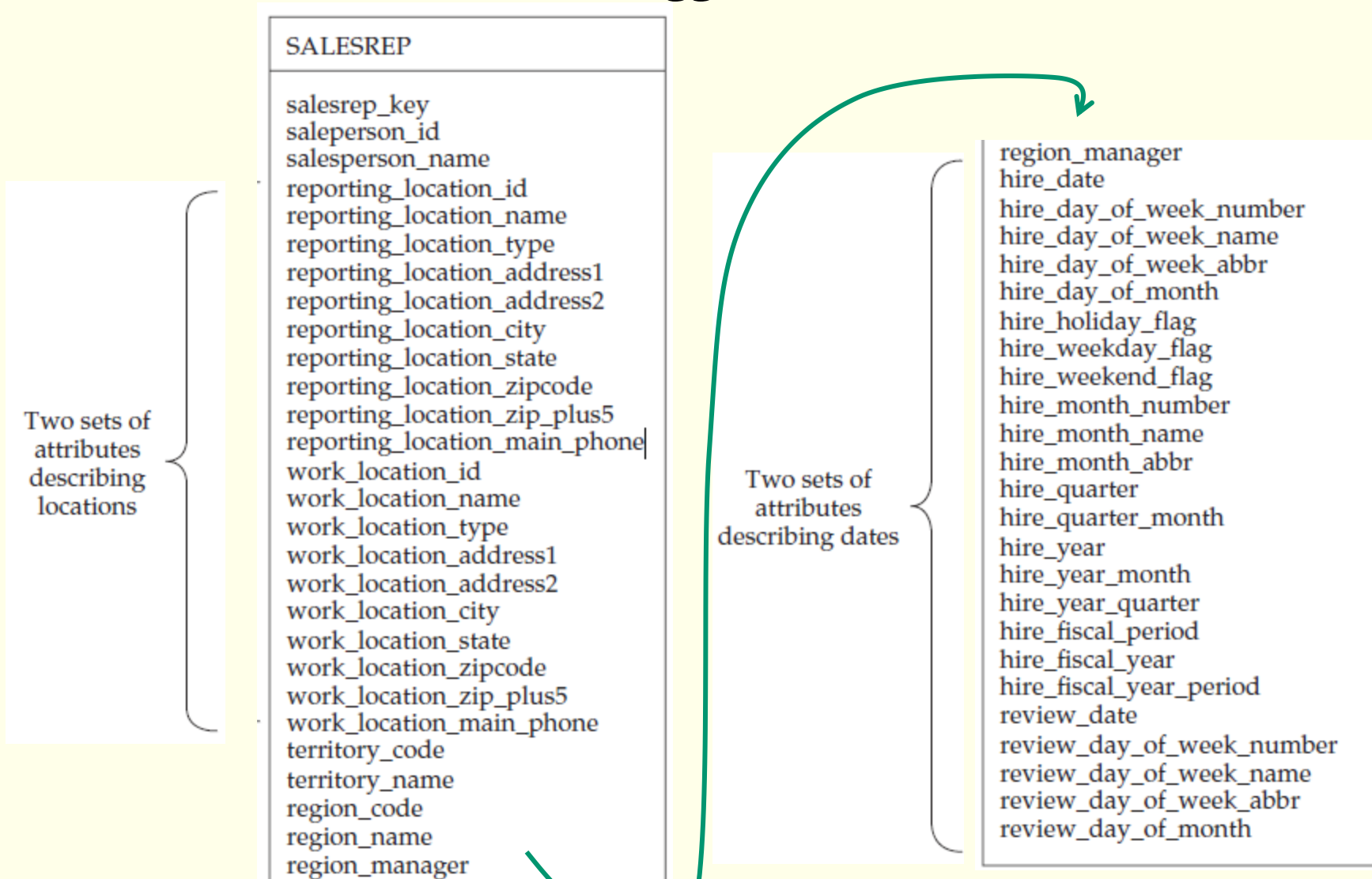


Figure 7-6 Repeating location and date attributes in a dimension table

Outriggers

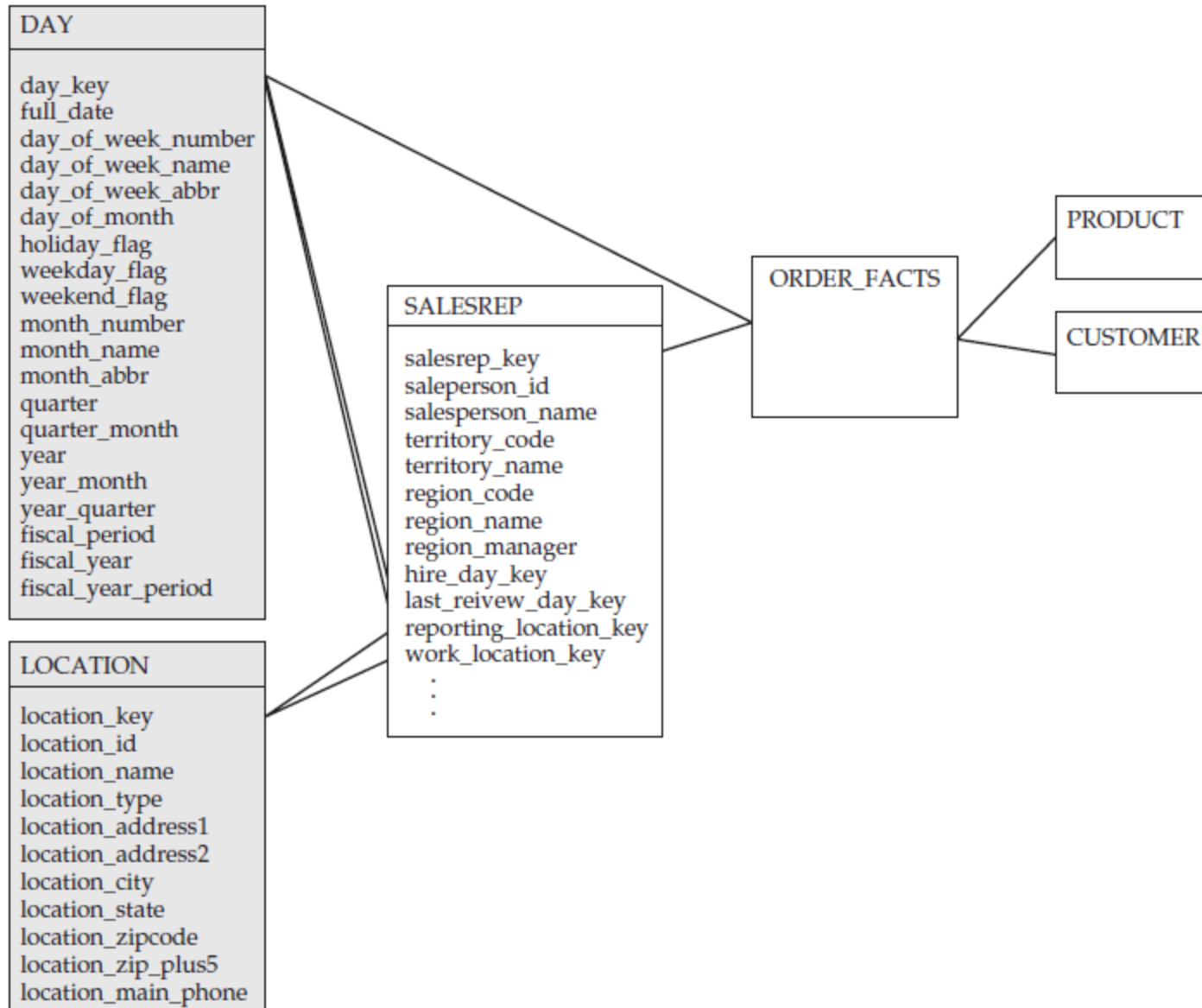


Figure 7-7 Two outriggers for the salesrep dimension

Outriggers

Page 166:

As Figure 7-8 illustrates, the presence of **outriggers** increases the number of joins that may be involved in querying the star.

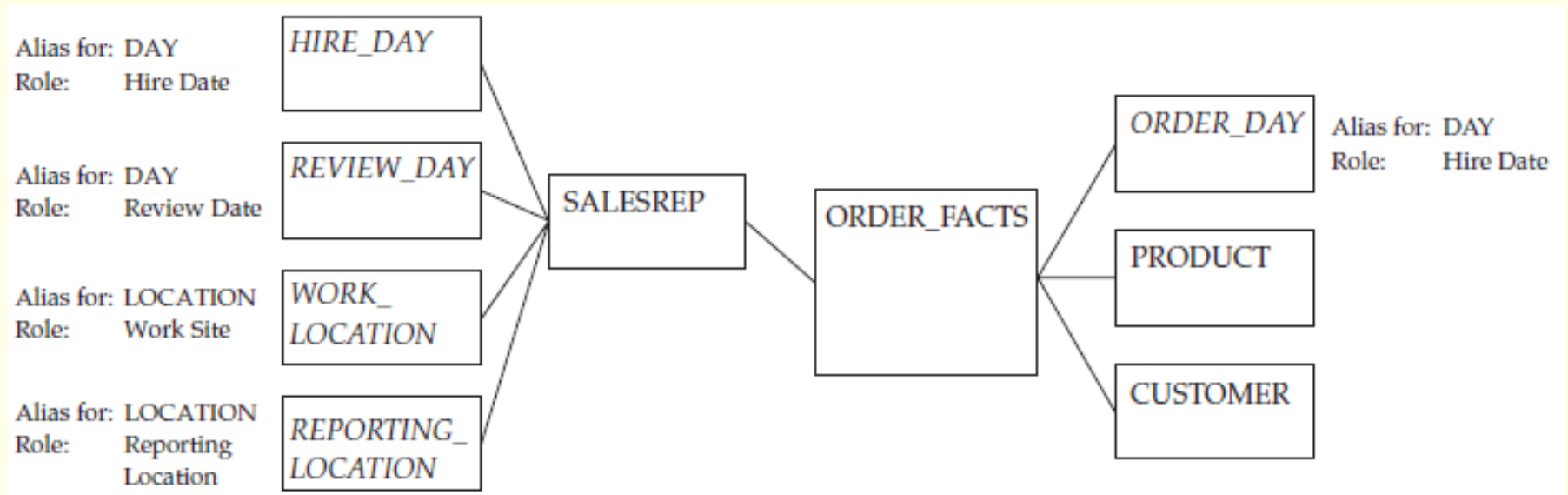
This translates into an increase in **complexity**, and potentially a **decrease in performance**.

Additional complexity is unnecessary from a user's point of view, but it is **possible to hide the outrigger** by replacing the dimension and outriggers with a **view**.

A negative impact on query performance may be more serious. Even if outriggers do not confuse the database's ability to perform a star join, they still increase the number of joins that must be processed.

The presence of outriggers in Figure 7-8 *doubles the number of potential joins to process when querying order facts*. Without outriggers, this star would require a maximum of four joins; with the outriggers, it includes eight joins.

Outriggers



Querying a star with outriggers
involves more joins than otherwise

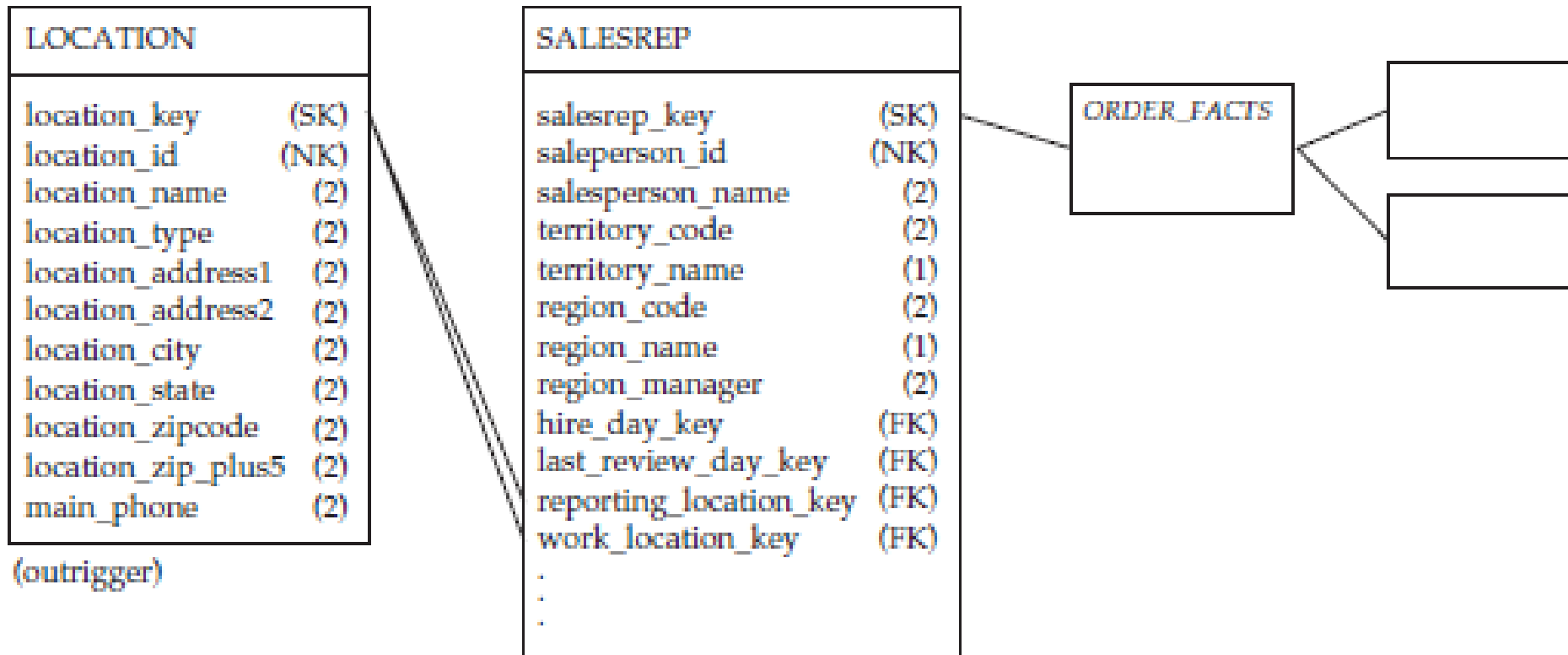
Outriggers

Effect on type 2 dimensions:

With an outrigger in place, it may be necessary to apply a type 2 change to a dimension row, even if none of its attributes have changed. The change is precipitated by a type 2 change in the outrigger.

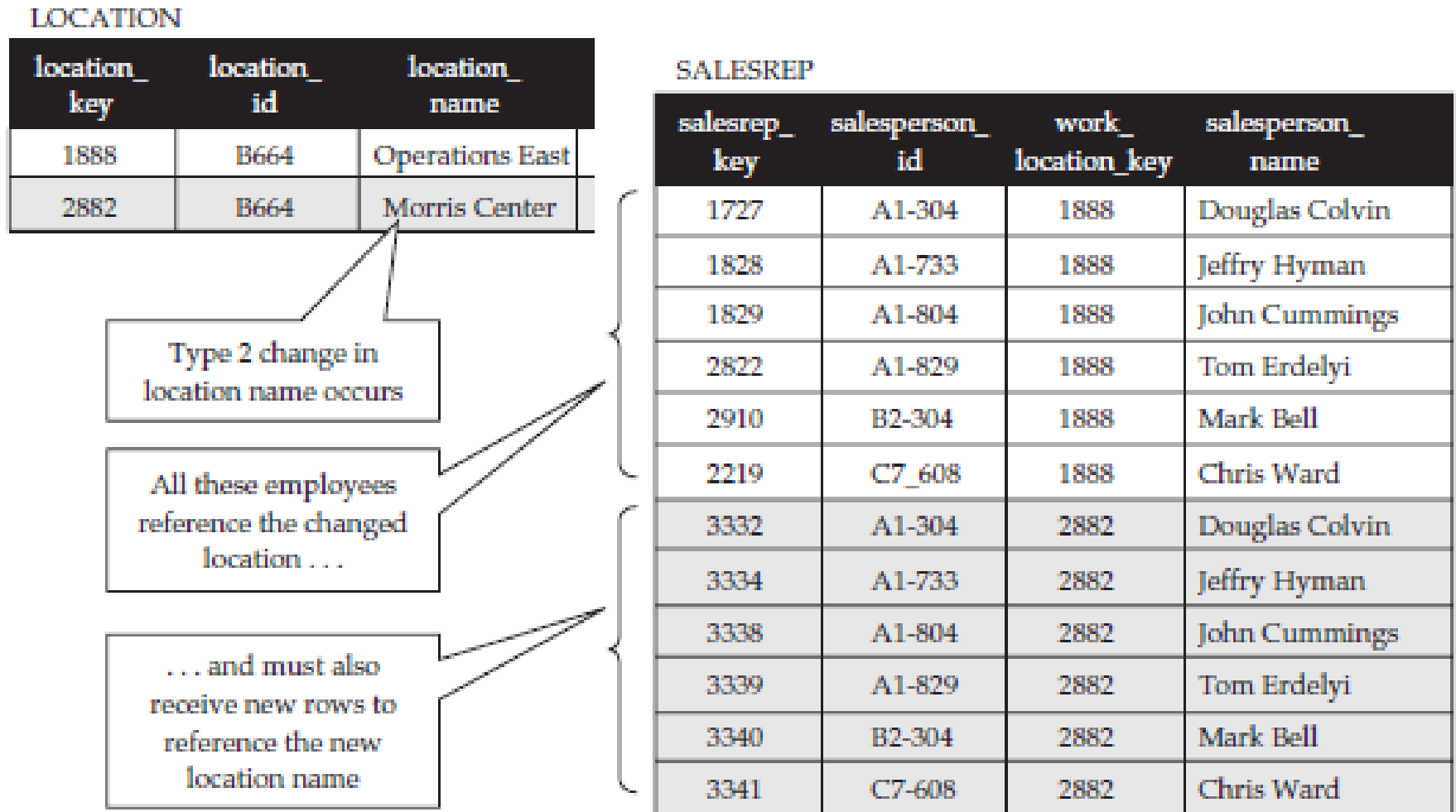
See Figure 7-9 →

Outriggers



Top half of Figure 7-9

Outriggers



Bottom half of Figure 7-9

Advantages of the (denormalized) Star Schema

- Easy to grasp the model
- Redundancy in the data set also simplifies the ETL
- Performance benefit - In a star, all facts and dimensions are separated by a maximum of one join. DBMSs that are 'star'-aware, have special join procedures (star-join ... not discussed in ACS-4902)
- Aside
https://docs.oracle.com/cd/E29633_01/CDMOG/GUID-A21256D6-7A93-4FBA-9688-053F6C5CF432.htm