

Lecture 5: Fact tables

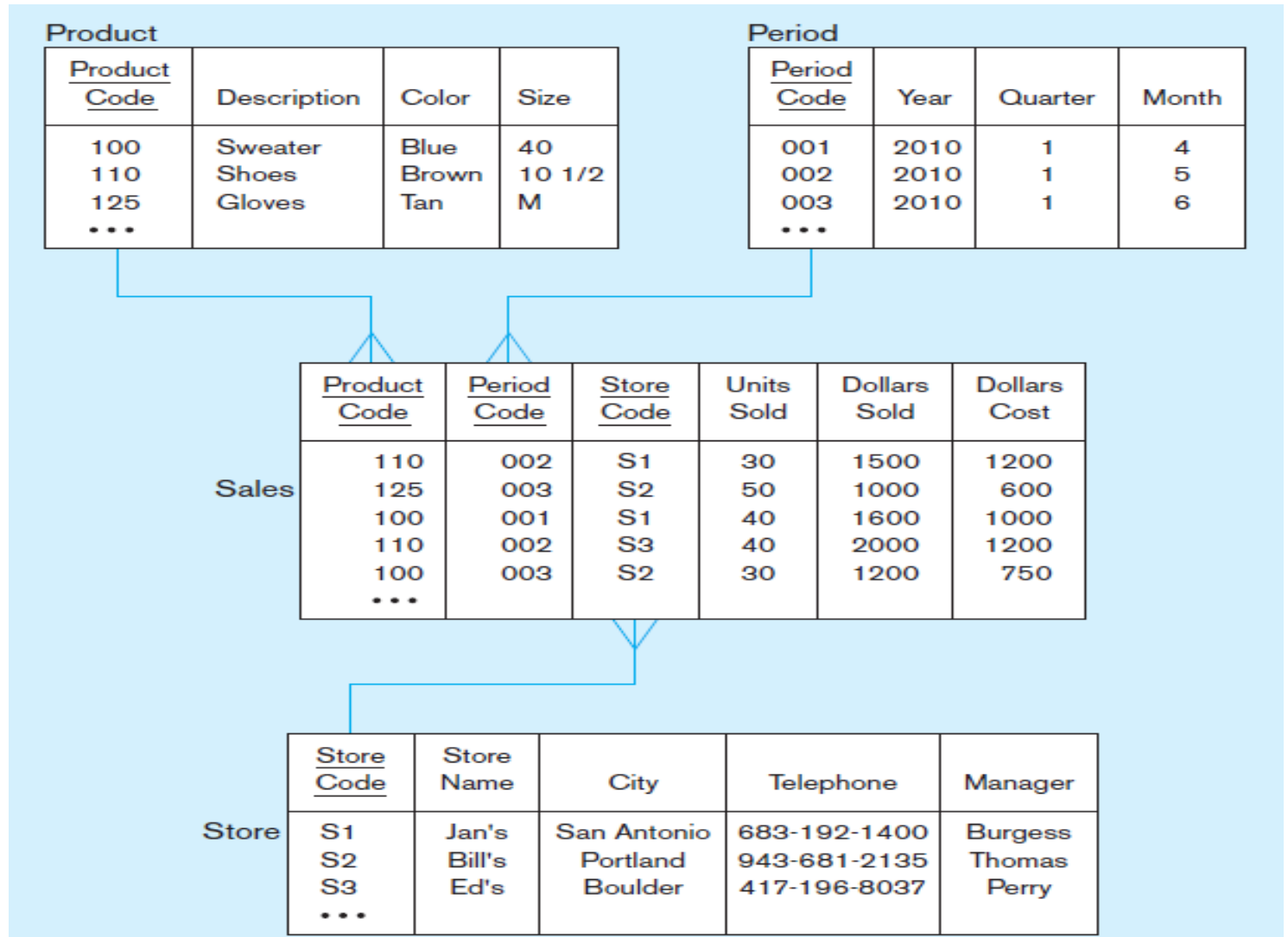
Dr Anesu Nyabadza

Size of a Fact Table

Depends on the number of dimensions and the grain of the fact table

Number of rows = product of number of possible values for each dimension associated with the fact table

Example: Assume the following for the given Figure:



Size of a Fact Table

Total rows calculated as follows (assuming only half the products record sales for a given month):

Total number of stores = 1,000

Total number of products = 10,000

Total number of periods = 24 (2 years' worth of monthly data)

Total rows = 1,000 stores × 5,000 active products × 24 months
= 120,000,000 rows (!)

Total size = 120,000,000 rows × 6 fields × 4 bytes/field = 2,880,000,000 bytes (or 2.88 gigabytes)

Variations of the Star Schema

Multiple Facts Tables

- Can improve performance

- Often used to store facts for different combinations of dimensions

- Conformed dimensions

Factless Facts Tables

- No nonkey data, but foreign keys for associated dimensions

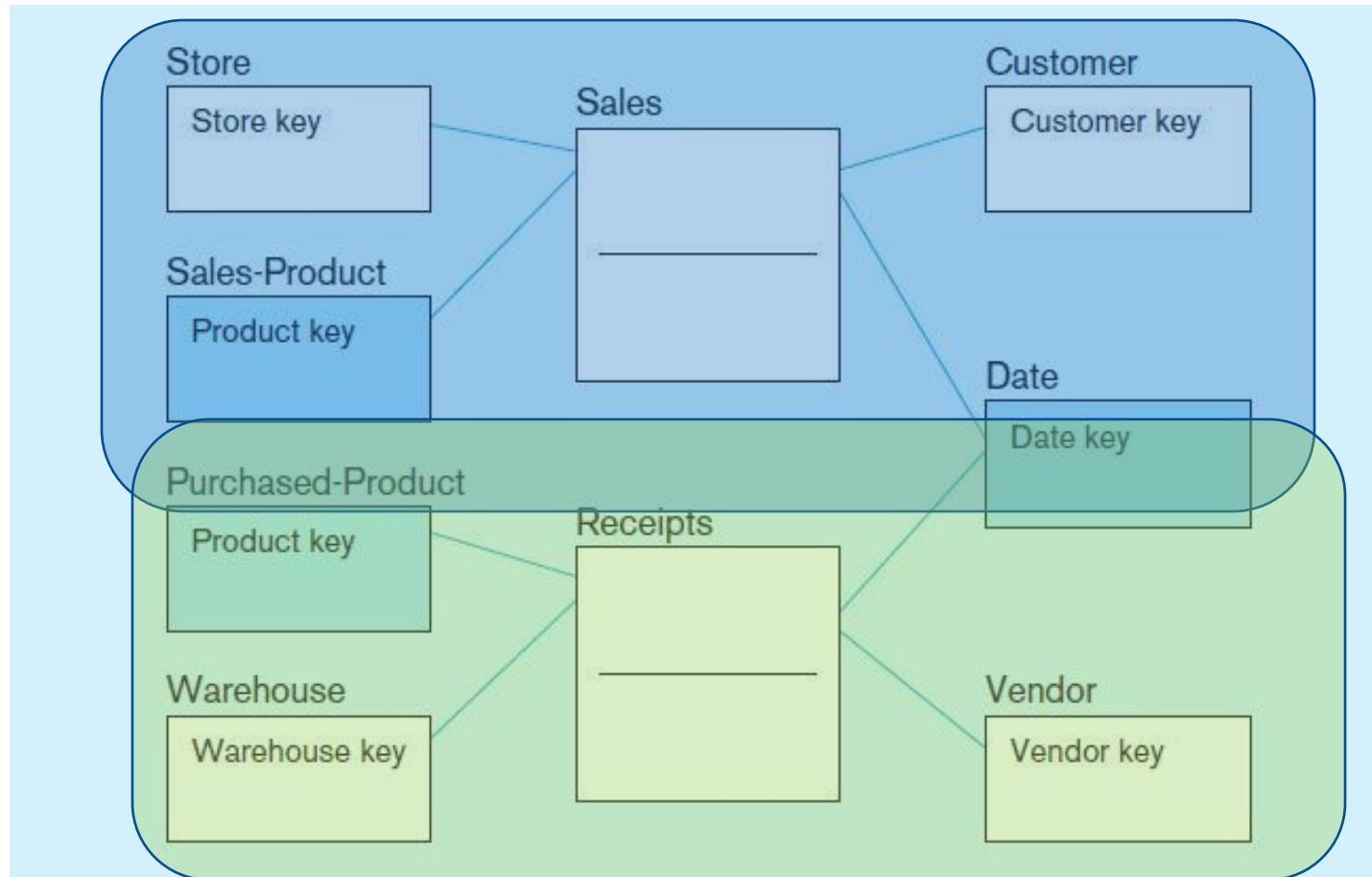
- Used for:

 - Tracking events*

 - Inventory coverage*

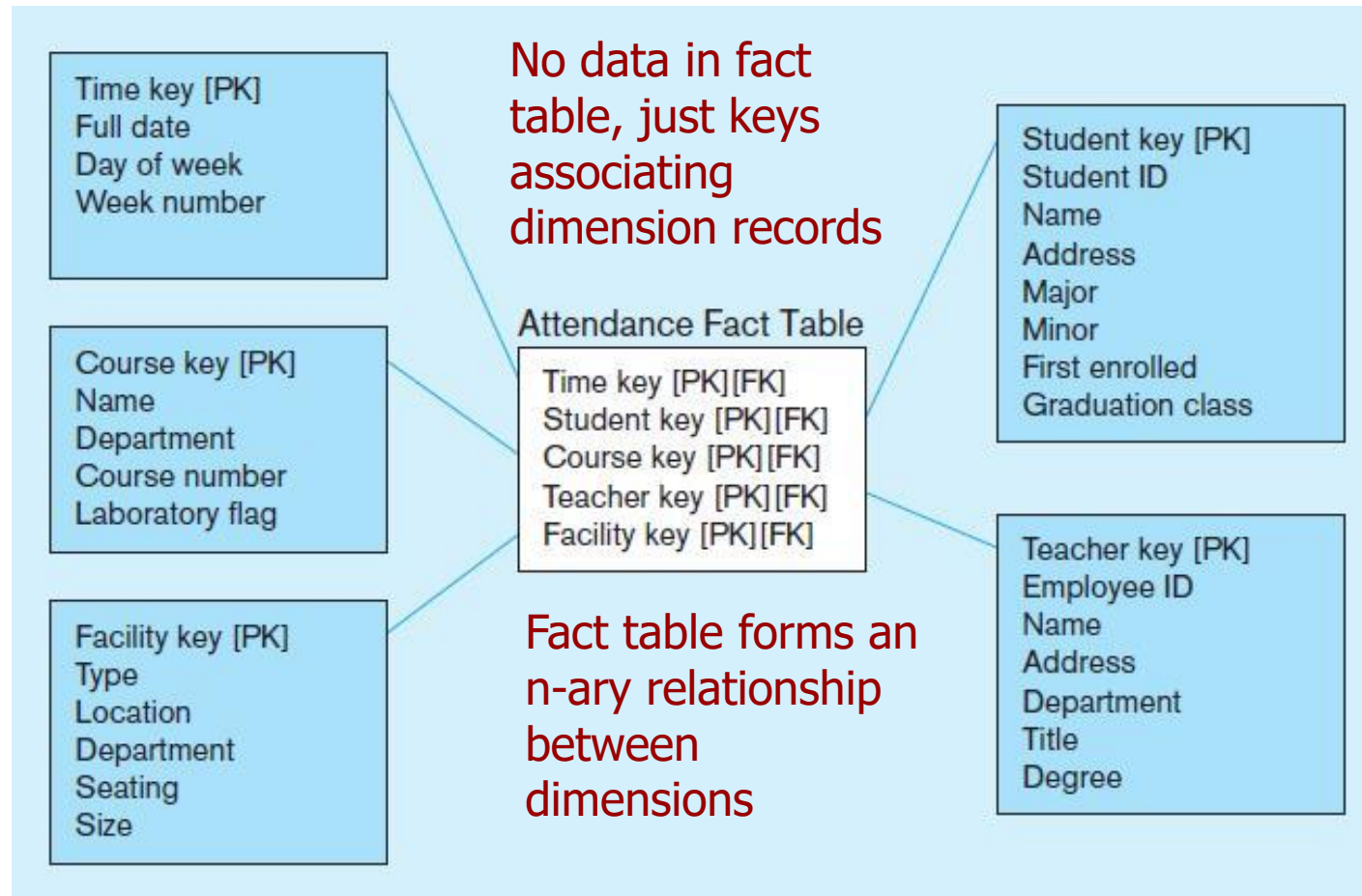
Conformed dimensions

Two fact tables → two (connected) star schemas.



***Conformed
dimension***
Associated with
multiple fact
tables

Factless fact table showing occurrence of an event



Normalizing Dimension Tables

Multivalued Dimensions

Facts qualified by a set of values for the same business subject

Normalization involves creating a table for an associative entity between dimensions

Hierarchies

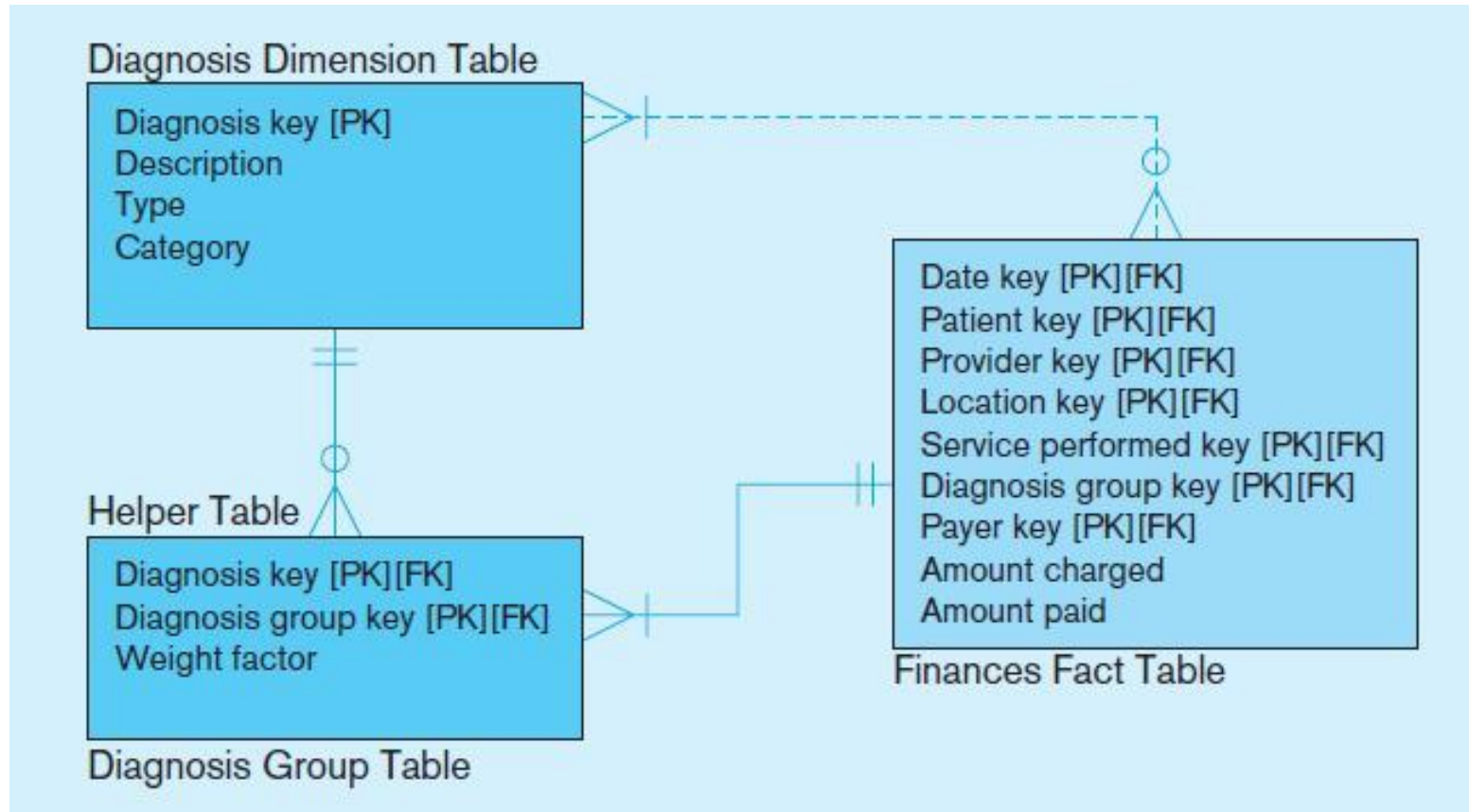
Sometimes a dimension forms a natural, fixed depth hierarchy

Design options

Include all information for each level in a single denormalized table

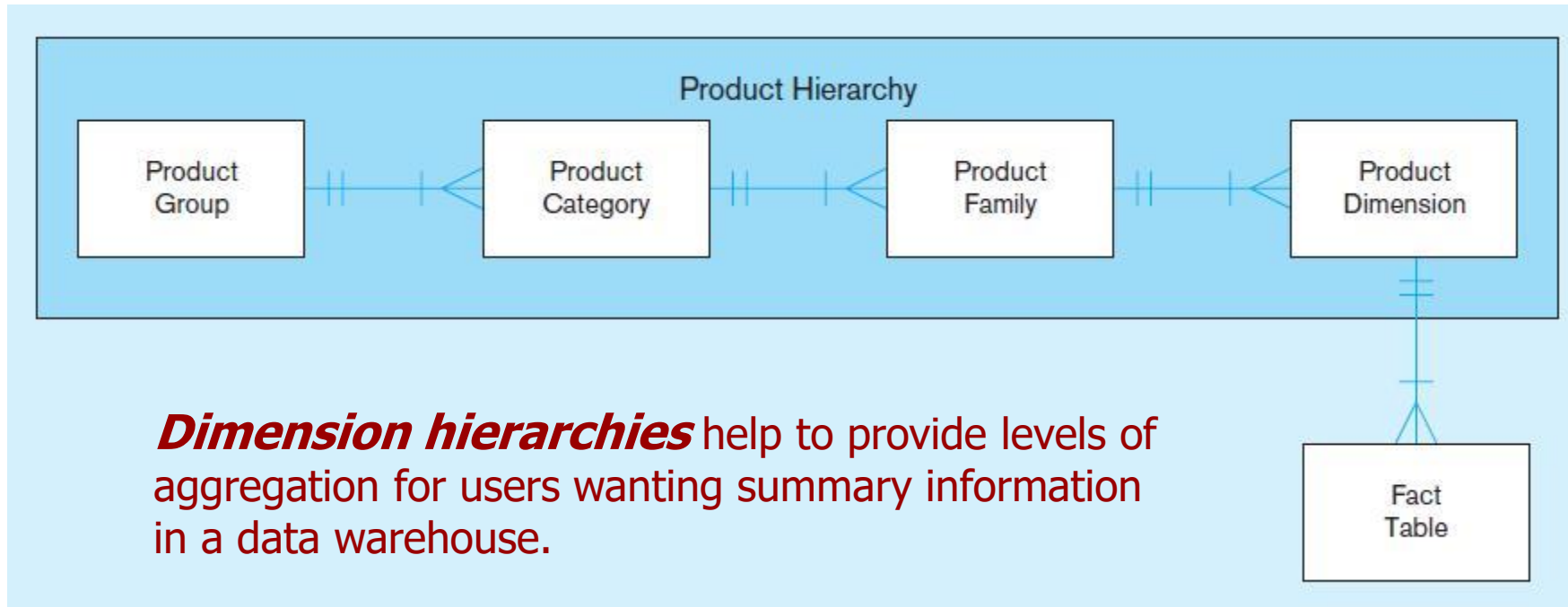
Normalize the dimension into a nested set of 1:M table relationships

Multivalued dimension



Helper table is an associative entity that implements a M:N relationship between dimension and fact.

Fixed product hierarchy



Types of Measures in Fact Tables

Additive

Facts that can be summed up through all of the dimensions in the fact table.

Sales_Amount is an additive fact, because you can sum up this fact along any of the three dimensions present in the fact table -- date, store, and product.

Semi Additive

Facts that can be summed up for some of the dimensions in the fact table, but not the others. Current_Balance is a semi-additive fact, as it can be added for all accounts but you cannot add all current balances for a given account for each day of the month (or any period of time).

Non-Additive

Facts that cannot be summed up for any of the dimensions present in the fact table for example, ratios or percentages, etc.