Database Systems -

Introduction to Databases and Data Warehouses

CHAPTER 8 - Data Warehouse and Data Mart Modeling

Copyright (c) 2014 Pearson Education, Inc.

INTRODUCTION

- ER modeling
  - A predominant technique for visualizing database requirements, used extensively for conceptual modeling of operational databases
- Relational modeling
  - Standard method for logical modeling of operational databases
- Both of these techniques can also be used during the development of data warehouses and data marts
- Dimensional modeling
  - A modeling technique tailored specifically for analytical database design purposes
  - Regularly used in practice for modeling data warehouses and data marts

Copyright (c) 2014 Pearson Education, Inc.

DIMENSIONAL MODELING

- Dimensional modeling
  - A data design methodology used for designing subject-oriented analytical databases, such as data warehouses or data marts
  - Commonly, dimensional modeling is employed as a relational data modeling technique
  - In addition to using the regular relational concepts (primary keys, foreign keys, integrity constraints, etc.) dimensional modeling distinguishes two types of tables:
    - **Dimensions**
    - **Facts**

Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

As a relational modeling technique, dimensional modeling, just like standard relational modeling, designs relational tables that have primary keys and are connected to each other via foreign keys, while conforming to the standard relational integrity constraints

Troy University Resources

DIMENSIONAL MODELING

- Dimension tables (dimensions)
  - Contain descriptions of the business, organization, or enterprise to which the subject of analysis belongs
  - Columns in dimension tables contain descriptive information that is often textual (e.g., product brand, product color, customer gender, customer education level), but can also be numeric (e.g., product weight, customer income level)
  - This information provides a basis for analysis of the subject

Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

For example, if the subject of the business analysis is sales, it can be analyzed by dimension columns such as product brand, customer gender, customer income level, and so on.

DIMENSIONAL MODELING

- Fact tables
  - Contain measures related to the subject of analysis and the foreign keys (associating fact tables with dimension tables)
  - The measures in the fact tables are typically numeric and are intended for mathematical computation and quantitative analysis

Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

For example, if the subject of the business analysis is sales, one of the measures in the fact table sales could be the sale's dollar amount. The sale amounts can be calculated and recalculated using different mathematical functions across various dimension columns. For example, the total and average sale can be calculated per product brand, customer gender, customer income level, and so on.

Troy University Resources

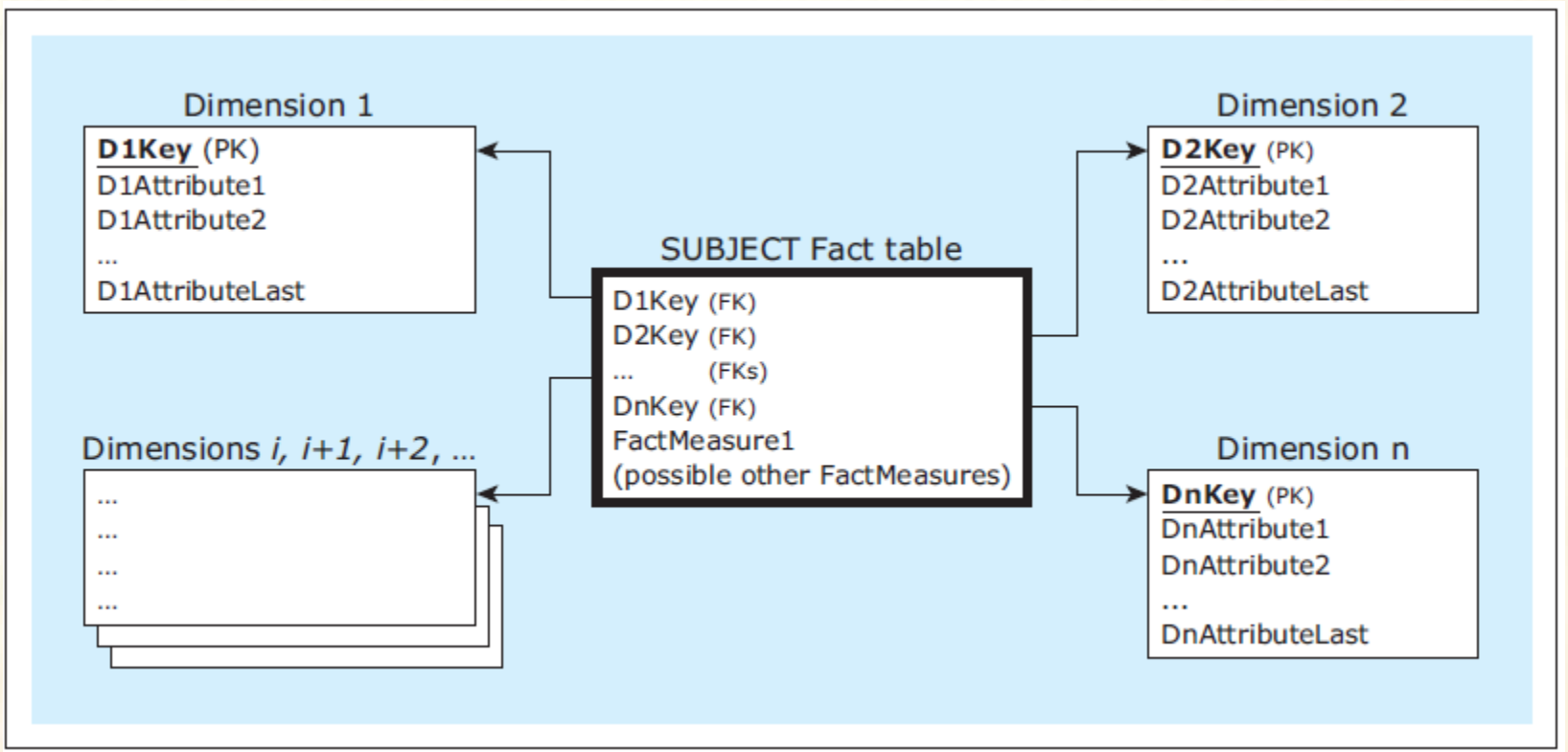
DIMENSIONAL MODELING

- **Star schema**
  - The result of dimensional modeling is a dimensional schema containing facts and dimensions
  - The dimensional schema is often referred to as the *star schema*

Copyright (c) 2014 Pearson Education, Inc.

DIMENSIONAL MODELING

A dimensional model (star schema)

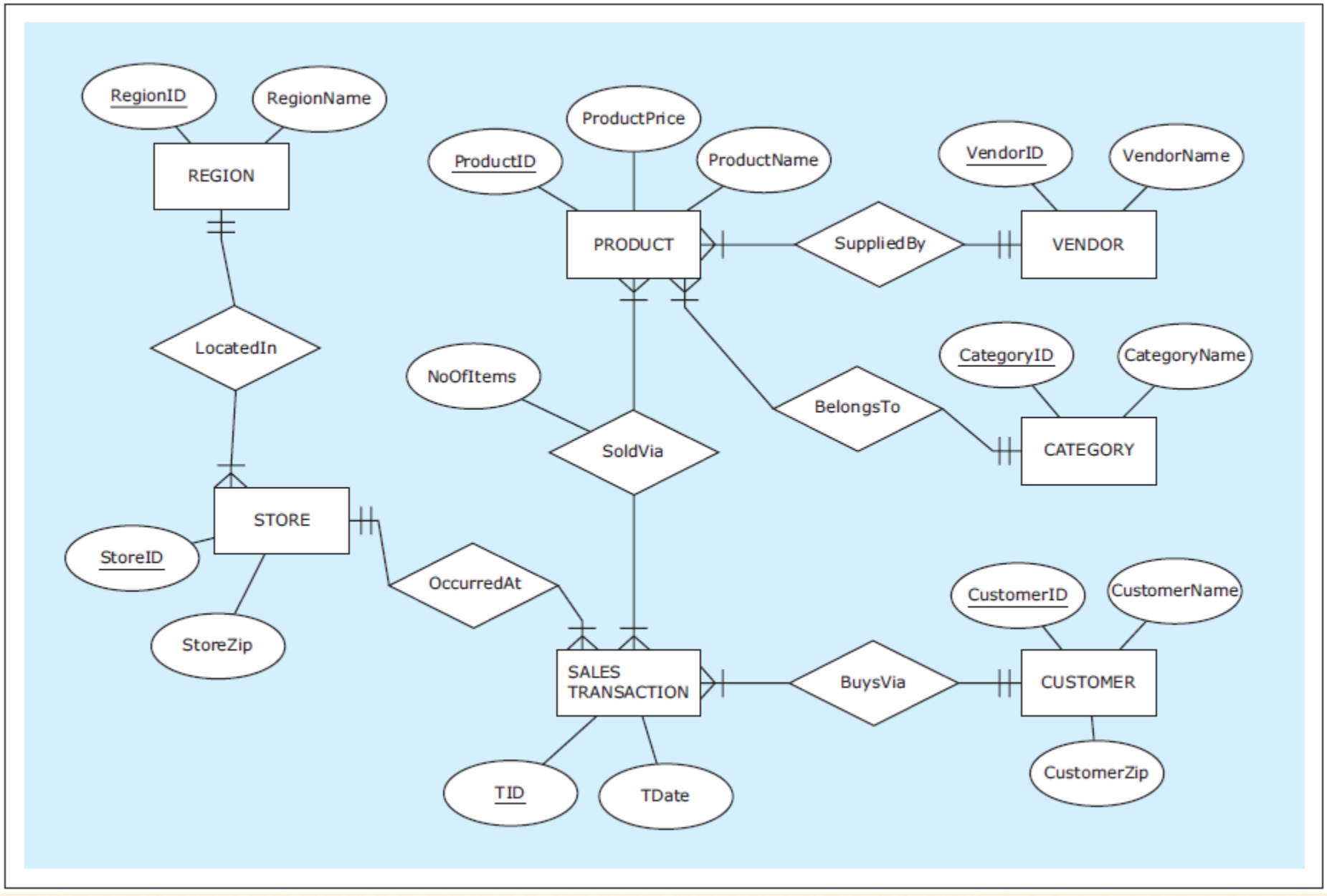


Copyright (c) 2014 Pearson Education, Inc.

Troy University Resources

Initial Example: Dimensional Model Based on A Single Source

ER diagram : ZAGI Retail Company Sales Department Database (Source)



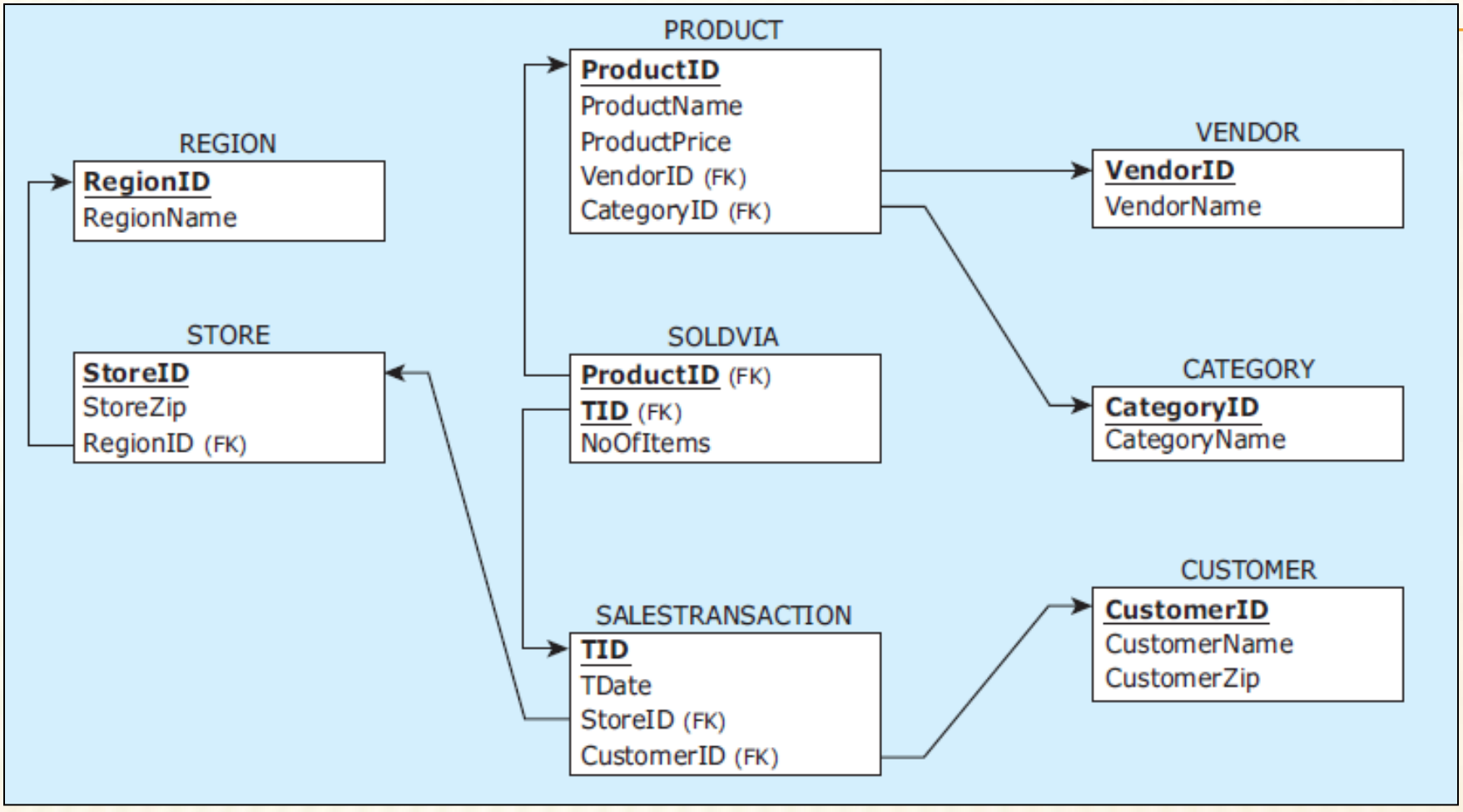
Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Pages 226-229.

Initial Example: Dimensional Model Based on A Single Source

Relational schema : ZAGI Retail Company Sales Department Database (Source)



Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Pages 226-229.

Initial Example: Dimensional Model Based on A Single Source

Data records: ZAGI Retail Company Sales Department Database (Source)

REGION

RegionID	RegionName
C	Chicagoland
T	Tristate

PRODUCT

ProductID	ProductName	ProductPrice	VendorID	CategoryID
1X1	Zzz Bag	$100	PG	CP
2X2	Easy Boot	$70	MK	FW
3X3	Cosy Sock	$15	MK	FW
4X4	Dura Boot	$90	PG	FW
5X5	Tiny Tent	$150	MK	CP
6X6	Biggy Tent	$250	MK	CP

VENDOR

VendorID	VendorName
PG	Pacifica Gear
MK	Mountain King

STORE

StoreID	StoreZip	RegionID
S1	60600	C
S2	60605	C
S3	35400	T

CATEGORY

CategoryID	CategoryName
CP	Camping
FW	Footwear

SALES TRANSACTION

TID	CustomerID	StoreID	TDate
T111	1-2-333	S1	1-Jan-2013
T222	2-3-444	S2	1-Jan-2013
T333	1-2-333	S3	2-Jan-2013
T444	3-4-555	S3	2-Jan-2013
T555	2-3-444	S3	2-Jan-2013

SOLDVIA

ProductID	TID	NoOfItems
1X1	T111	1
2X2	T222	1
3X3	T333	5
1X1	T333	1
4X4	T444	1
2X2	T444	2
4X4	T555	4
5X5	T555	2
6X6	T555	1

CUSTOMER

CustomerID	CustomerName	CustomerZip
1-2-333	Tina	60137
2-3-444	Tony	60611
3-4-555	Pam	35401

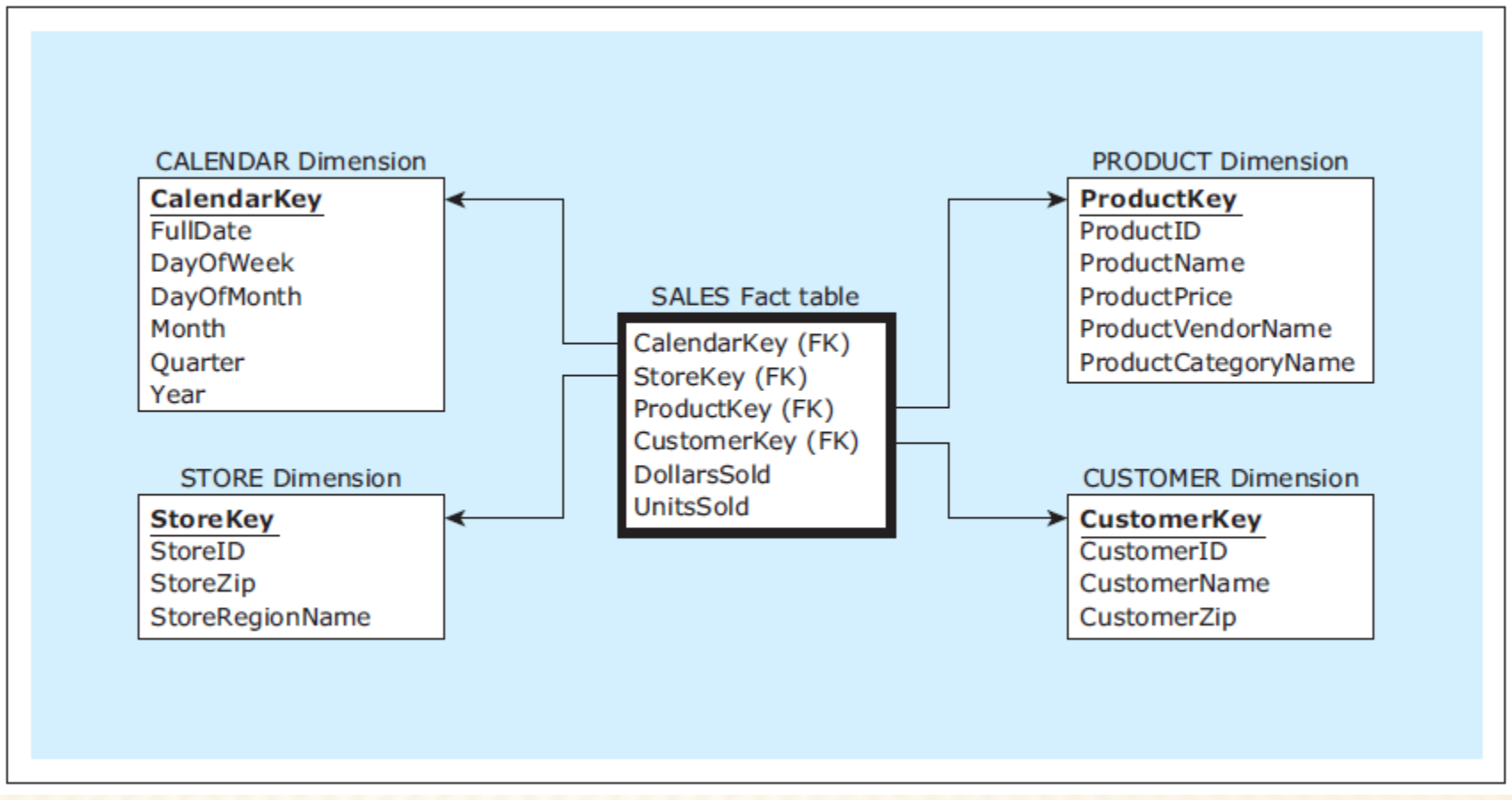
Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Pages 226-229.

Initial Example: Dimensional Model Based on A Single Source

ZAGI Retail Company dimensional model for the subject ***sales***



Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Pages 226-229.

Troy University Resources

DIMENSIONAL MODELING

- **Star schema**
  - In the star schema, the chosen subject of analysis is represented by a fact table
  - Designing the star schema involves considering which dimensions to use with the fact table representing the chosen subject
  - For every dimension under consideration, two questions must be answered:
    - Question 1*: Can the dimension table be useful for the analysis of the chosen subject?*
    - Question 2: *Can the dimension table be created based on the existing data sources?*

Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

In this example the chosen subject of analysis (sales) is represented by the SALES fact table.

For each of the four dimensions in the ZAGI example the answer to Question 1 and Question 2 was yes. Therefore those dimensions were included in the star schema.

Initial Example: Dimensional Model Based on A Single Source

*ZAGI Retail Company dimensional model for the subject **sales**, populated with the data from the operational data source*

CALENDAR Dimension

CalendarKey	FullDate	DayOf Week	DayOf Month	Month	Qtr	Year
1	1/1/2013	Tuesday	1	January	Q1	2013
2	1/2/2013	Wednesday	2	January	Q1	2013

STORE Dimension

StoreKey	StoreID	StoreZip	StoreRegionName
1	S1	60600	Chicagoland
2	S2	60605	Chicagoland
3	S3	35400	Tristate

PRODUCT Dimension

ProductKey	ProductID	Product Name	Product Price	Product Vendor Name	Product Category Name
1	1X1	Zzz Bag	$100	Pacifica Gear	Camping
2	2X2	Easy Boot	$70	Mountain King	Footwear
3	3X3	Cosy Sock	$15	Mountain King	Footwear
4	4X4	Dura Boot	$90	Pacifica Gear	Footwear
5	5X5	Tiny Tent	$150	Mountain King	Camping
6	6X6	Biggy Tent	$250	Mountain King	Camping

CUSTOMER Dimension

CustomerKey	CustomerID	CustomerName	CustomerZip
1	1-2-333	Tina	60137
2	2-3-444	Tony	60611
3	3-4-555	Pam	35401

SALES Fact table

CalendarKey	StoreKey	ProductKey	CustomerKey	DollarsSold	UnitsSold
1	1	1	1	$100	1
1	2	2	2	$70	1
2	3	3	1	$75	5
2	3	1	1	$100	1
2	3	4	3	$90	1
2	3	2	3	$140	2
2	3	4	2	$360	4
2	3	5	2	$300	2
2	3	6	2	$250	1

Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Pages 226-229.

Troy University Resources

DIMENSIONAL MODELING

- Characteristics of dimensions and facts
  - A typical dimension contains relatively static data, while in a typical fact table, records are added continually, and the table rapidly grows in size.
  - In a typical dimensionally modeled analytical database, dimension tables have orders of magnitude fewer records than fact tables

Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

The discussion about the number of records in the example star schema is given on Page 230.

DIMENSIONAL MODELING

- Surrogate key
  - Typically, in a star schema all dimension tables are given a simple, non-composite system-generated key, also called a *surrogate key*
  - Values for the surrogate keys are typically simple auto-increment integer values
  - Surrogate key values have no meaning or purpose except to give each dimension a new column that serves as a primary key within the dimensional model instead of the operational key

Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

For example, instead of using the primary key ProductID as the primary key of the PRODUCT dimension, a new surrogate key column ProductKey is created.

One of the main reasons for creating a surrogate primary key and not using the operational primary key as a primary key of the dimension, is to enable the handling of so called slowly changing dimensions (covered later in this chapter).

Initial Example: Dimensional Model Based on A Single Source

Example query

Query A: Compare the quantities of sold products on Saturdays in the category Camping provided by the vendor Pacifica Gear within the Tristate region between the 1st and 2nd quarter of the year 2013

Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Pages 230-231.

*Example query - **Query A, dimensional version***

```
SELECT    SUM(SA.UnitsSold)
,         P.ProductCategoryName
,         P.ProductVendorName
,         C.DayofWeek
,         C.Qtr
FROM
          Calendar C
,         Store    S
,         Product  P
,         Sales    SA
WHERE
          C.CalendarKey = SA.CalendarKey
AND       S.StoreKey = SA.StoreKey
AND       P.ProductKey = SA.ProductKey
AND       P.ProductVendorName = 'Pacifica Gear'
AND       P.ProductCategoryName = 'Camping'
AND       S.StoreRegionName = 'Tristate'
AND       C.DayofWeek = 'Saturday'
AND       C.Year = 2013
AND       C.Qtr IN ( 'Q1', 'Q2' )
GROUP BY
          P.ProductCategoryName,
          P.ProductVendorName,
          C.DayofWeek,
          C.Qtr;
```


Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Pages 230-231.

Troy University Resources

*Example query - **Query A, nondimensional version***

```
SELECT    SUM( SV.NoOfItems )
,         C.CategoryName
,         V.VendorName
,         EXTRACTWEEKDAY(ST.Date)
,         EXTRACTQUARTER(ST.Date)
FROM
          Region    R
,         Store     S
,         SalesTransaction ST
,         SoldVia   SV
,         Product   P
,         Vendor    V
,         Category  C
WHERE
          R.RegionID = S.RegionID
AND       S.StoreID = ST.StoreID
AND       ST.Tid = SV.Tid
AND       SV.ProductID = P.ProductID
AND       P.VendorID = V.VendorID
AND       P.CateoryID = C.CategoryID
AND       V.VendorName = 'Pacifica Gear'
AND       C.CategoryName = 'Camping'
AND       R.RegionName = 'Tristate'
AND       EXTRACTWEEKDAY(St.Date) = 'Saturday'
AND       EXTRACTYEAR(ST.Date) = 2013
AND       EXTRACTQUARTER(ST.Date) IN ( 'Q1', 'Q2' )
GROUP BY
          C.CategoryName,
          V.VendorName,
          EXTRACTWEEKDAY(ST.Date),
          EXTRACTQUARTER(ST.Date);
```

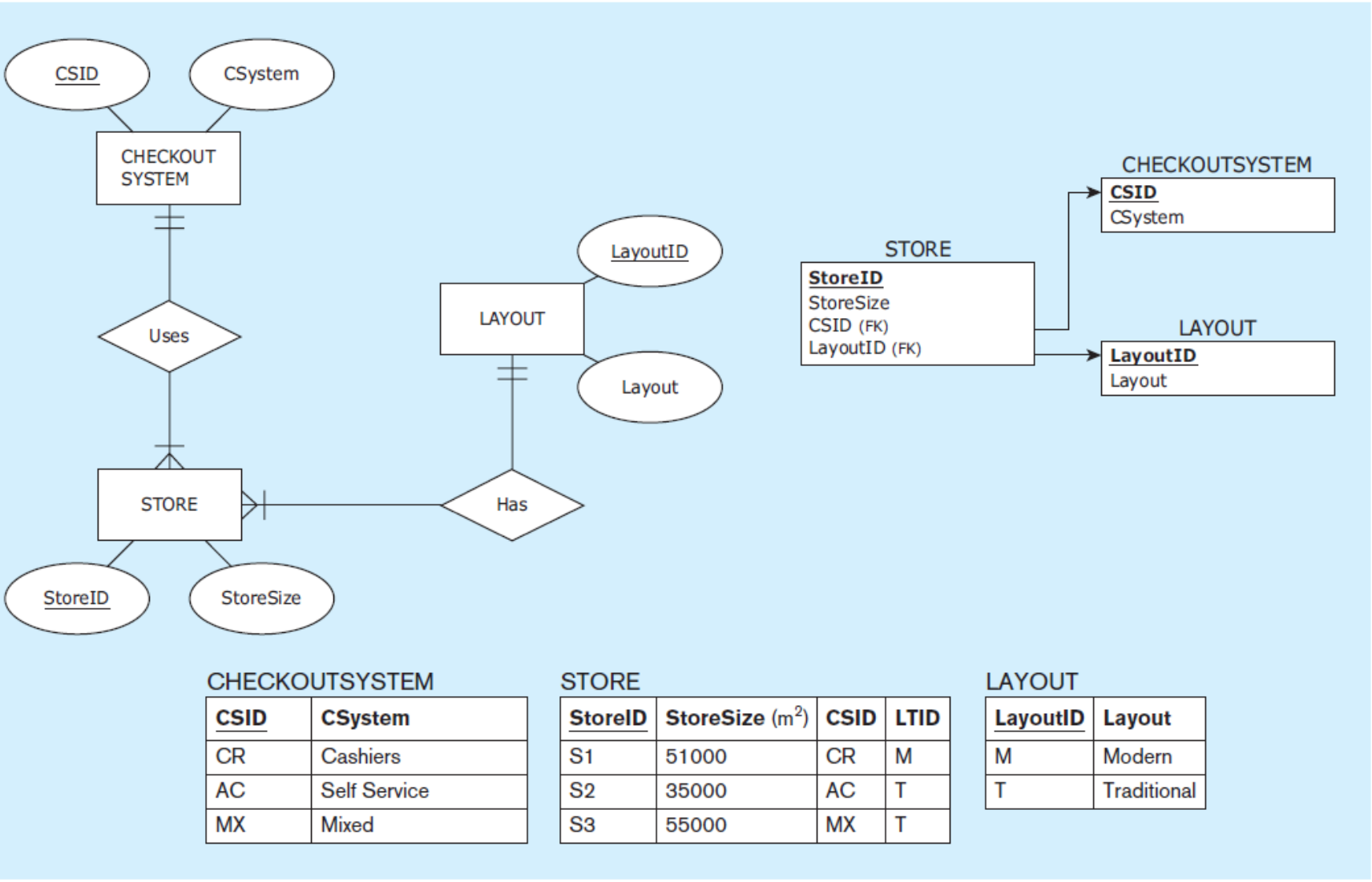

Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Pages 230-231.

Expanded Example: Dimensional Model Based on Multiple Sources

ZAGI Retail Company Facilities Department Database (Source 2)



CHECKOUTSYSTEM

CSID	CSystem
CR	Cashiers
AC	Self Service
MX	Mixed

STORE

StoreID	StoreSize (m^2)	CSID	LTID
S1	51000	CR	M
S2	35000	AC	T
S3	55000	MX	T

LAYOUT

LayoutID	Layout
M	Modern
T	Traditional

Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Pages 231-234.

Troy University Resources

Expanded Example: Dimensional Model Based on Multiple Sources

Customer Demographic Data Table - external source acquired from a market research company (Source 3)

CUSTOMER TABLE

CustomerID	Customer Name	Gender	Marital Status	Education Level	Credit Score
1-2-333	Tina	Female	Single	College	700
2-3-444	Tony	Male	Single	High School	650
3-4-555	Pam	Female	Married	College	623

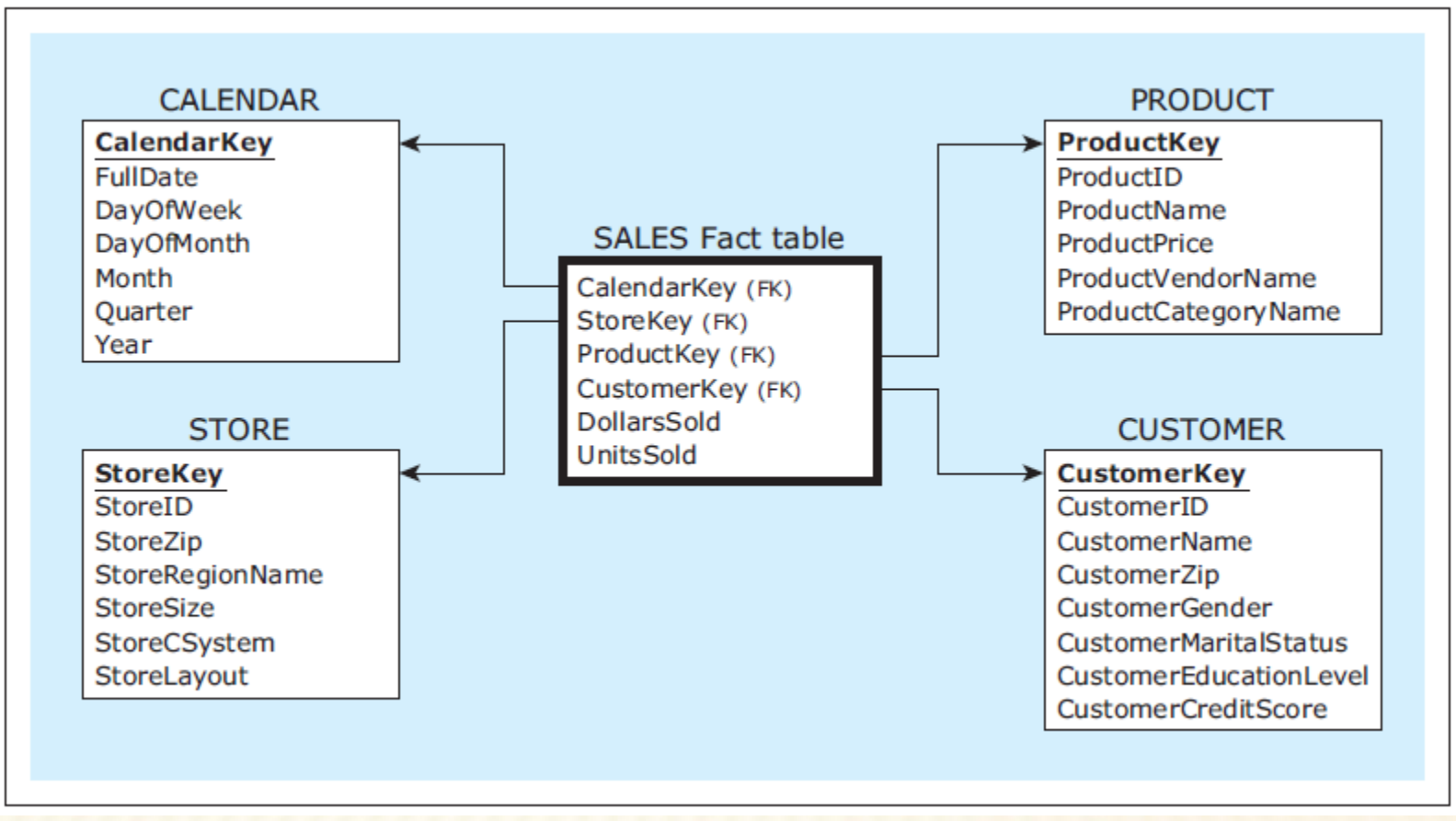
Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Pages 231-234.

Expanded Example: Dimensional Model Based on Multiple Sources

ZAGI Retail Company dimensional model for the subject ***sales***



Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Pages 231-234.

Expanded Example: Dimensional Model Based on Multiple Sources

*ZAGI Retail Company dimensional model for the subject **sales** , populated with the data from the three sources*

CALENDAR Dimension

CalendarKey	FullDate	DayOf Week	DayOf Month	Month	Qtr	Year
1	1/1/2013	Tuesday	1	January	Q1	2013
2	1/2/2013	Wednesday	2	January	Q1	2013

PRODUCT Dimension

ProductKey	ProductID	Product Name	Product Price	Product Vendor Name	Product Category Name
1	1X1	Zzz Bag	$100	Pacifica Gear	Camping
2	2X2	Easy Boot	$70	Mountain King	Footwear
3	3X3	Cosy Sock	$15	Mountain King	Footwear
4	4X4	Dura Boot	$90	Pacifica Gear	Footwear
5	5X5	Tiny Tent	$150	Mountain King	Camping
6	6X6	Biggy Tent	$250	Mountain King	Camping

STORE Dimension

StoreKey	StoreID	StoreZip	StoreRegion Name	Store Size (m^2)	Store CSystem	Store Layout
1	S1	60600	Chicagoland	51000	Cashiers	Modern
2	S2	60605	Chicagoland	35000	Self Service	Traditional
2	S3	35400	Tristate	55000	Mixed	Traditional

CUSTOMER Dimension

CustomerKey	CustomerID	Customer Name	Customer Zip	Customer Gender	Customer MaritalStatus	Customer EducationLevel	Customer CreditScore
1	1-2-333	Tina	60137	Female	Single	College	700
2	2-3-444	Tony	60611	Male	Single	High School	650
3	3-4-555	Pam	35401	Female	Married	College	623

SALES Fact table

CalendarKey	StoreKey	ProductKey	CustomerKey	DollarsSold	UnitsSold
1	1	1	1	$100	1
1	2	2	2	$70	1
2	3	3	1	$75	5
2	3	1	1	$100	1
2	3	4	3	$90	1
2	3	2	3	$140	2
2	3	4	2	$360	4
2	3	5	2	$300	2
2	3	6	2	$250	1

Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Pages 231-234.

Expanded Example: Dimensional Model Based on Multiple Sources

Example query

Query B: Compare the quantities of sold products to male customers in Modern stores on Saturdays in the category Camping provided by the vendor Pacifica Gear within the Tristate region between the 1st and 2nd quarter of the year 2013.

Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Pages 231-234.

Troy University Resources

*Example query - **Query B, dimensional version***

```
SELECT    SUM(SA.UnitsSold)
,         P.ProductCategoryName
,         P.ProductVendorName
,         C.DayofWeek
,         C.Qtr
FROM
          Calendar C
,         Store    S
,         Product  P
,         Customer CU
,         Sales    SA
WHERE
          C.CalendarKey = SA.CalendarKey
AND       S.StoreKey = SA.StoreKey
AND       P.ProductKey = SA.ProductKey
AND       CU.CustomerKey = SA.CustomerKey
AND       P.ProductVendorName = 'Pacifica Gear'
AND       P.ProductCategoryName = 'Camping'
AND       S.StoreRegionName = 'Tristate'
AND       C.DayofWeek = 'Saturday'
AND       C.Year = 2013
AND       C.Qtr IN ( 'Q1', 'Q2' )
AND       S.StoreLayout = 'Modern'
AND       CU.Gender = 'Male'
GROUP BY
          P.ProductCategoryName,
          P.ProductVendorName,
          C.DayofWeek,
          C.Qtr;
```


Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Pages 231-234.

DIMENSIONAL MODELING

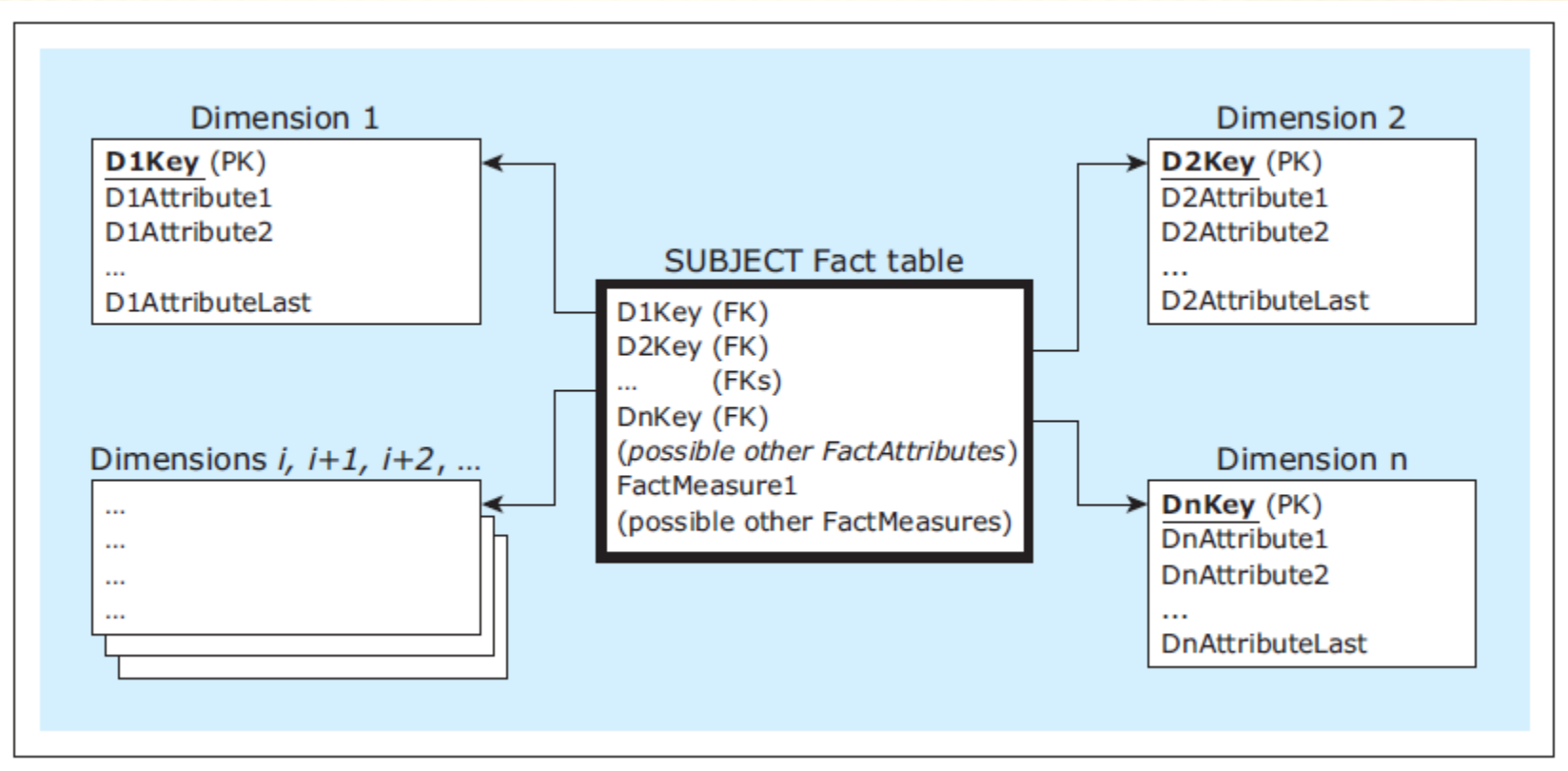
- **Additional possible fact attributes**
  - A fact table contains
    - Foreign keys connecting the fact table to the dimension tables
    - The measures related to the subject of analysis
  - In addition to the measures related to the subject of analysis, in certain cases fact tables can contain other attributes that are not measures
  - Two of the most typical additional attributes that can appear in the fact table are:
    - **Transaction identifier**
    - **Transaction time**

Copyright (c) 2014 Pearson Education, Inc.

Troy University Resources

DIMENSIONAL MODELING

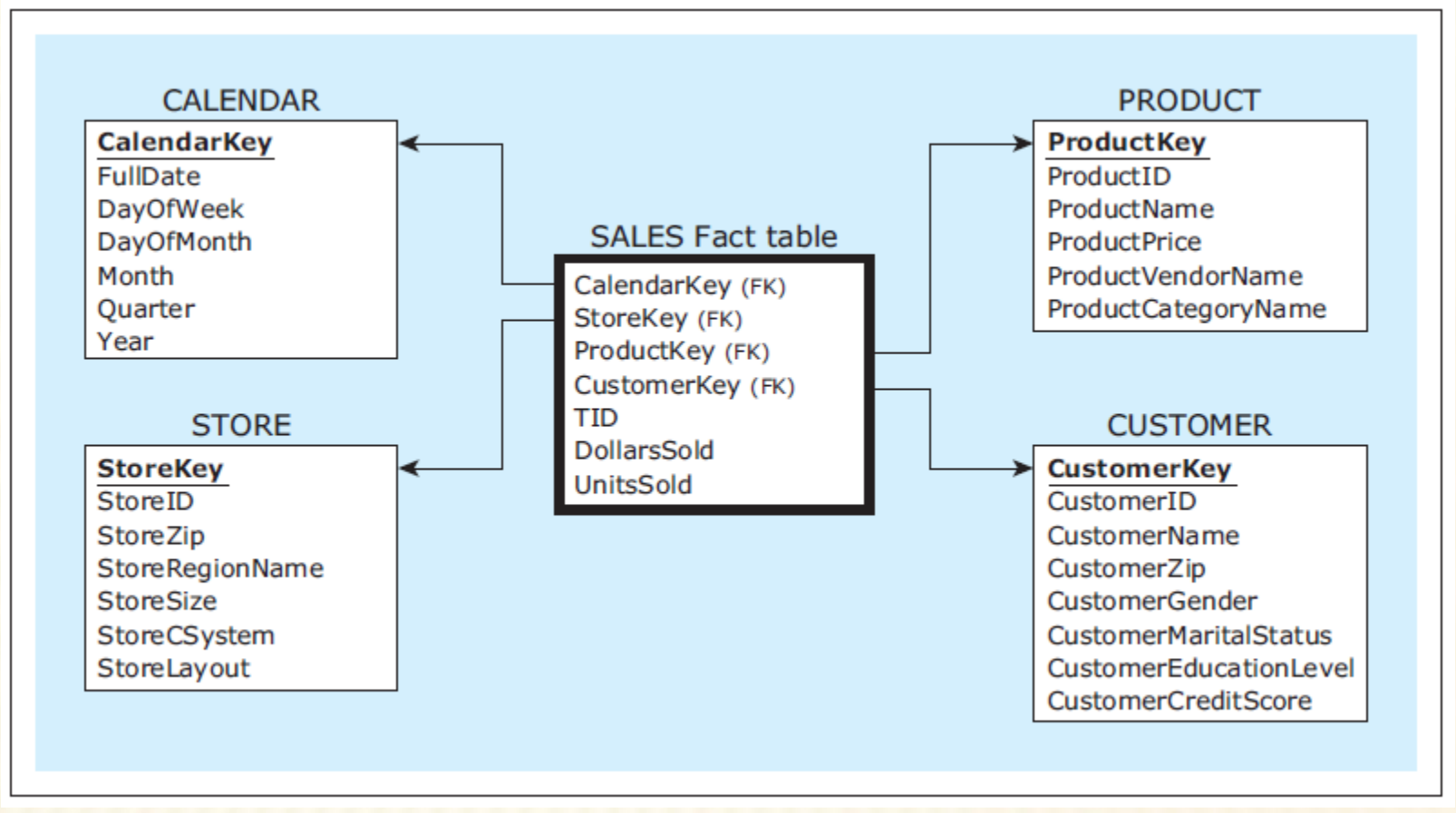
Additional possible fact attributes



Copyright (c) 2014 Pearson Education, Inc.

Expanded Example: Dimensional Model Based on Multiple Sources

*ZAGI Retail Company dimensional model for the subject sales with **transaction identifier** included*



Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Pages 235-237.

Expanded Example: Dimensional Model Based on Multiple Sources

ZAGI Retail Company dimensional model for the subject sales, populated with the data, including the ***transaction identifier*** *values*

CALENDAR Dimension

CalendarKey	FullDate	DayOf Week	DayOf Month	Month	Qtr	Year
1	1/1/2013	Tuesday	1	January	Q1	2013
2	1/2/2013	Wednesday	2	January	Q1	2013

PRODUCT Dimension

ProductKey	ProductID	Product Name	Product Price	Product Vendor Name	Product Category Name
1	1X1	Zzz Bag	$100	Pacifica Gear	Camping
2	2X2	Easy Boot	$70	Mountain King	Footwear
3	3X3	Cosy Sock	$15	Mountain King	Footwear
4	4X4	Dura Boot	$90	Pacifica Gear	Footwear
5	5X5	Tiny Tent	$150	Mountain King	Camping
6	6X6	Biggy Tent	$250	Mountain King	Camping

STORE Dimension

StoreKey	StoreID	StoreZip	StoreRegion Name	Store Size (m^2)	Store CSystem	Store Layout
1	S1	60600	Chicagoland	51000	Cashiers	Modern
2	S2	60605	Chicagoland	35000	Self Service	Traditional
3	S3	35400	Tristate	55000	Mixed	Traditional

CUSTOMER Dimension

CustomerKey	CustomerID	Customer Name	Customer Zip	Customer Gender	Customer MaritalStatus	Customer EducationLevel	Customer CreditScore
1	1-2-333	Tina	60137	Female	Single	College	700
2	2-3-444	Tony	60611	Male	Single	High School	650
3	3-4-555	Pam	35401	Female	Married	College	623

SALES Fact table

CalendarKey	StoreKey	ProductKey	CustomerKey	TID	DollarsSold	UnitsSold
1	1	1	1	T111	$100	1
1	2	2	2	T222	$70	1
2	3	3	1	T333	$75	5
2	3	1	1	T333	$100	1
2	3	4	3	T444	$90	1
2	3	2	3	T444	$140	2
2	3	4	2	T555	$360	4
2	3	5	2	T555	$300	2
2	3	6	2	T555	$250	1

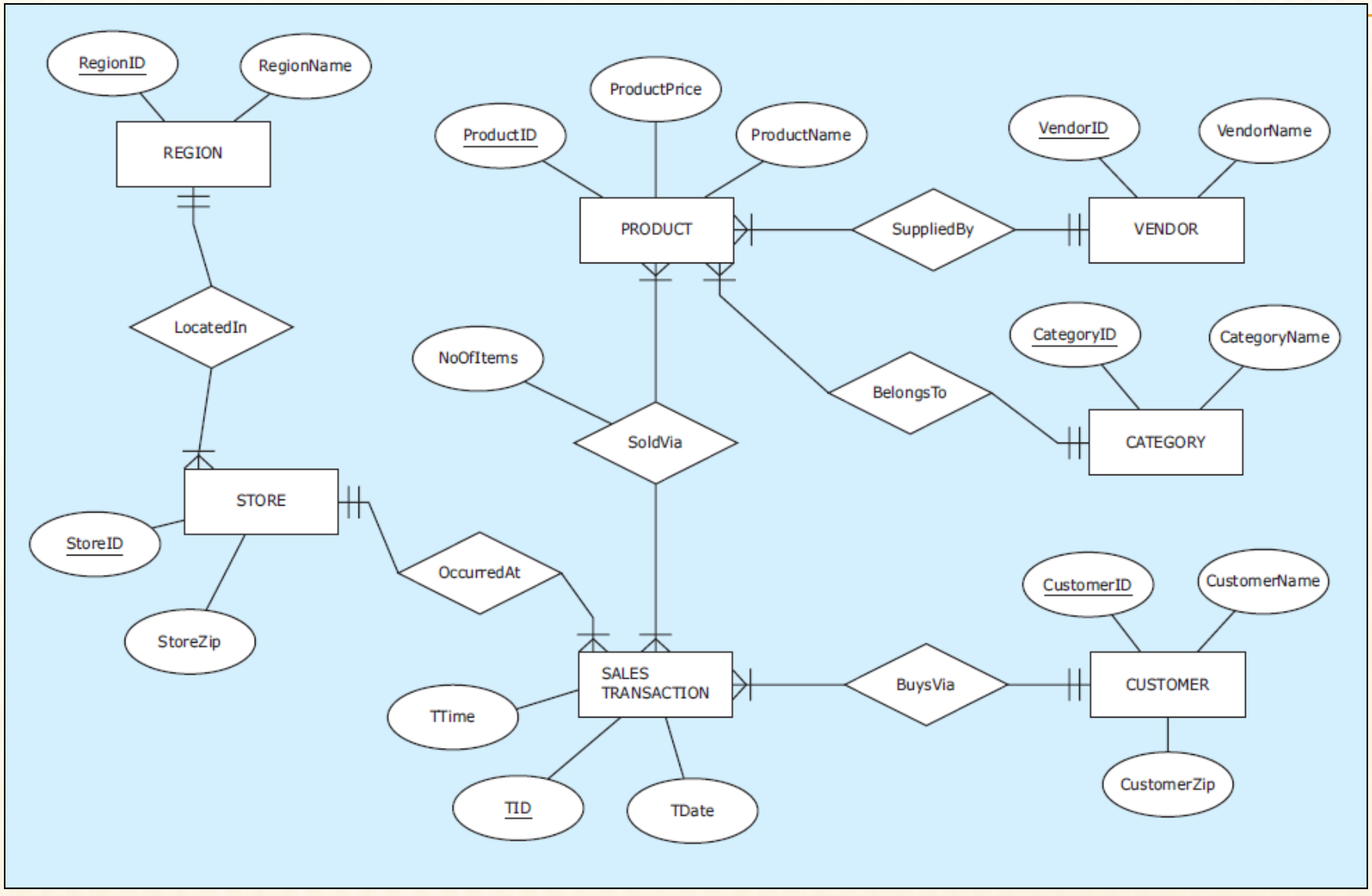
Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Pages 235-237.

Expanded Example: Dimensional Model Based on Multiple Sources

*ER diagram : ZAGI Retail Company Sales Department Database (Source 1) with the **time** attribute included*



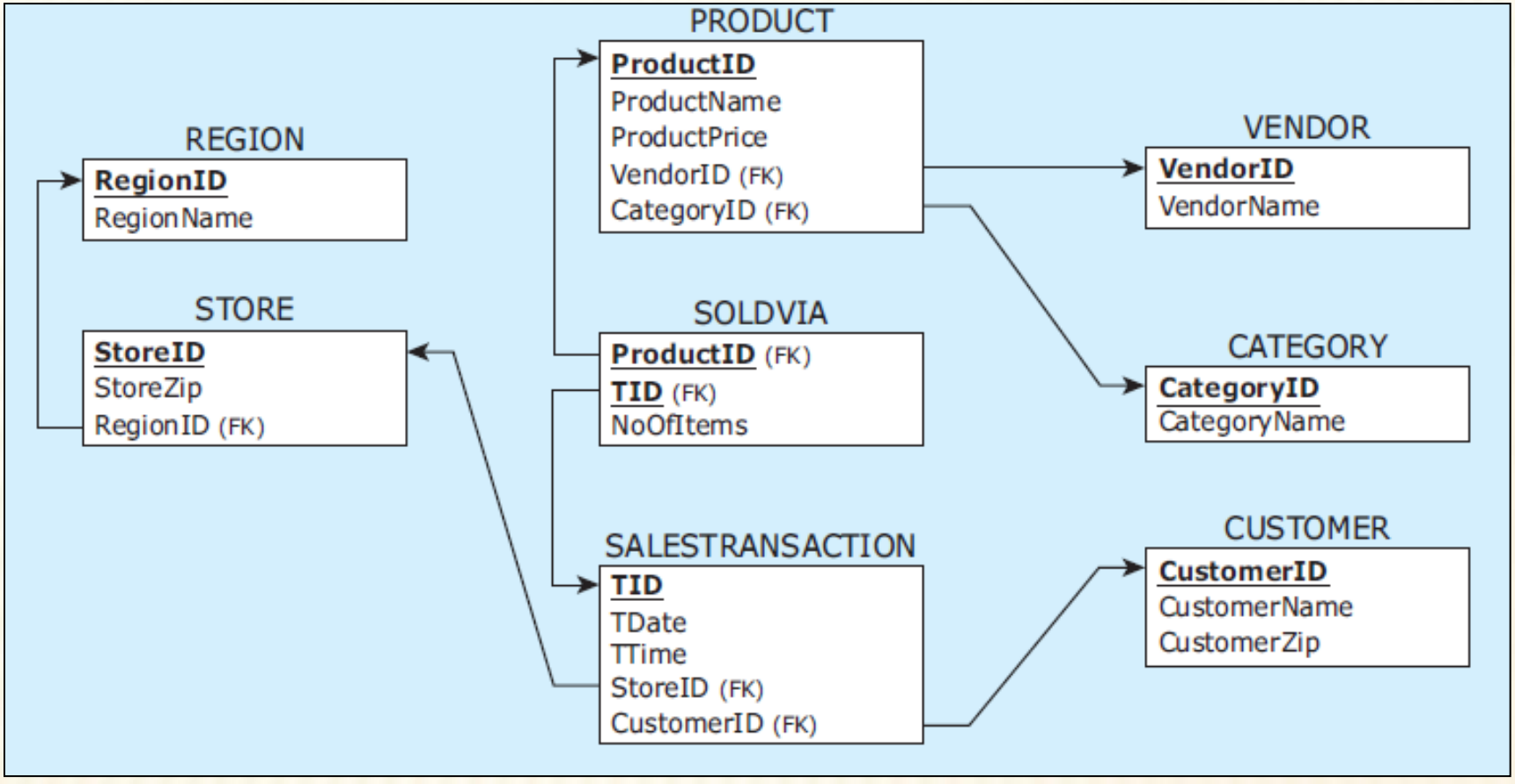
Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Pages 237-239.

Expanded Example: Dimensional Model Based on Multiple Sources

*Relational schema : ZAGI Retail Company Sales Department Database (Source 1) with the **time** column included*



Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Pages 237-239.

Expanded Example: Dimensional Model Based on Multiple Sources

*Data records: ZAGI Retail Company Sales Department Database (Source 1) with **time** data included*

REGION

RegionID	RegionName
C	Chicagoland
T	Tristate

PRODUCT

ProductID	ProductName	ProductPrice	VendorID	CategoryID
1X1	Zzz Bag	$100	PG	CP
2X2	Easy Boot	$70	MK	FW
3X3	Cosy Sock	$15	MK	FW
4X4	Dura Boot	$90	PG	FW
5X5	Tiny Tent	$150	MK	CP
6X6	Biggy Tent	$250	MK	CP

VENDOR

VendorID	VendorName
PG	Pacifica Gear
MK	Mountain King

STORE

StoreID	StoreZip	RegionID
S1	60600	C
S2	60605	C
S3	35400	T

CATEGORY

CategoryID	CategoryName
CP	Camping
FW	Footwear

SALESTRANSACTION

TID	CustomerID	StoreID	TDate	TTime
T111	1-2-333	S1	1-Jan-2013	8:23:59 AM
T222	2-3-444	S1	1-Jan-2013	8:24:30 AM
T333	1-2-333	S3	2-Jan-2013	8:15:08 AM
T444	3-4-555	S3	2-Jan-2013	8:20:33 AM
T555	2-3-444	S3	2-Jan-2013	8:30:00 AM

SOLDVIA

ProductID	TID	NoOfItems
1X1	T111	1
2X2	T222	1
3X3	T333	5
1X1	T333	1
4X4	T444	1
2X2	T444	2
4X4	T555	4
5X5	T555	2
6X6	T555	1

CUSTOMER

CustomerID	CustomerName	CustomerZip
1-2-333	Tina	60137
2-3-444	Tony	60611
3-4-555	Pam	35401

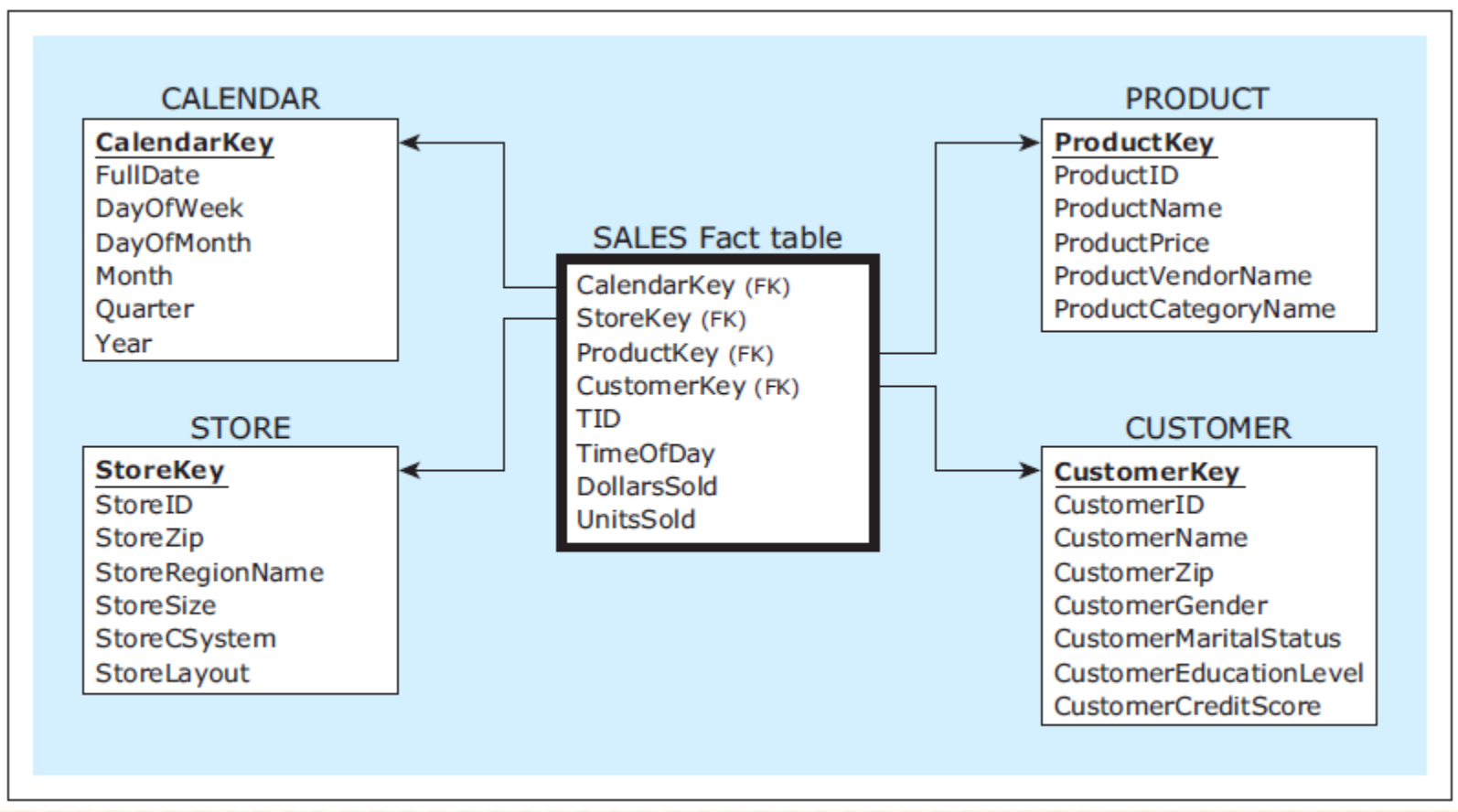
Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Pages 237-239.

Expanded Example: Dimensional Model Based on Multiple Sources

*ZAGI Retail Company dimensional model for the subject sales with **time** included*



Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Pages 237-239.

Expanded Example: Dimensional Model Based on Multiple Sources

*ZAGI Retail Company dimensional model for the subject sales, populated with the data, including the **time** values*

CALENDAR Dimension

CalendarKey	FullDate	DayOf Week	DayOf Month	Month	Qtr	Year
1	1/1/2013	Tuesday	1	January	Q1	2013
2	1/2/2013	Wednesday	2	January	Q1	2013

PRODUCT Dimension

ProductKey	ProductID	Product Name	Product Price	Product Vendor Name	Product Category Name
1	1X1	Zzz Bag	$100	Pacifica Gear	Camping
2	2X2	Easy Boot	$70	Mountain King	Footwear
3	3X3	Cosy Sock	$15	Mountain King	Footwear
4	4X4	Dura Boot	$90	Pacifica Gear	Footwear
5	5X5	Tiny Tent	$150	Mountain King	Camping
6	6X6	Biggy Tent	$250	Mountain King	Camping

STORE Dimension

StoreKey	StoreID	StoreZip	StoreRegion Name	Store Size (m^2)	Store CSystem	Store Layout
1	S1	60600	Chicagoland	51000	Cashiers	Modern
2	S2	60605	Chicagoland	35000	Self Service	Traditional
3	S3	35400	Tristate	55000	Mixed	Traditional

CUSTOMER Dimension

CustomerKey	CustomerID	Customer Name	Customer Zip	Customer Gender	Customer MaritalStatus	Customer EducationLevel	Customer CreditScore
1	1-2-333	Tina	60137	Female	Single	College	700
2	2-3-444	Tony	60611	Male	Single	High School	650
3	3-4-555	Pam	35401	Female	Married	College	623

SALES Fact table

CalendarKey	StoreKey	ProductKey	CustomerKey	TID	TimeOfDay	DollarsSold	UnitsSold
1	1	1	1	T111	8:23:59 AM	$100	1
1	2	2	2	T222	8:24:30 AM	$70	1
2	3	3	1	T333	8:15:08 AM	$75	5
2	3	1	1	T333	8:15:08 AM	$100	1
2	3	4	3	T444	8:20:33 AM	$90	1
2	3	2	3	T444	8:20:33 AM	$140	2
2	3	4	2	T555	8:30:00 AM	$360	4
2	3	5	2	T555	8:30:00 AM	$300	2
2	3	6	2	T555	8:30:00 AM	$250	1

Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Pages 237-239.

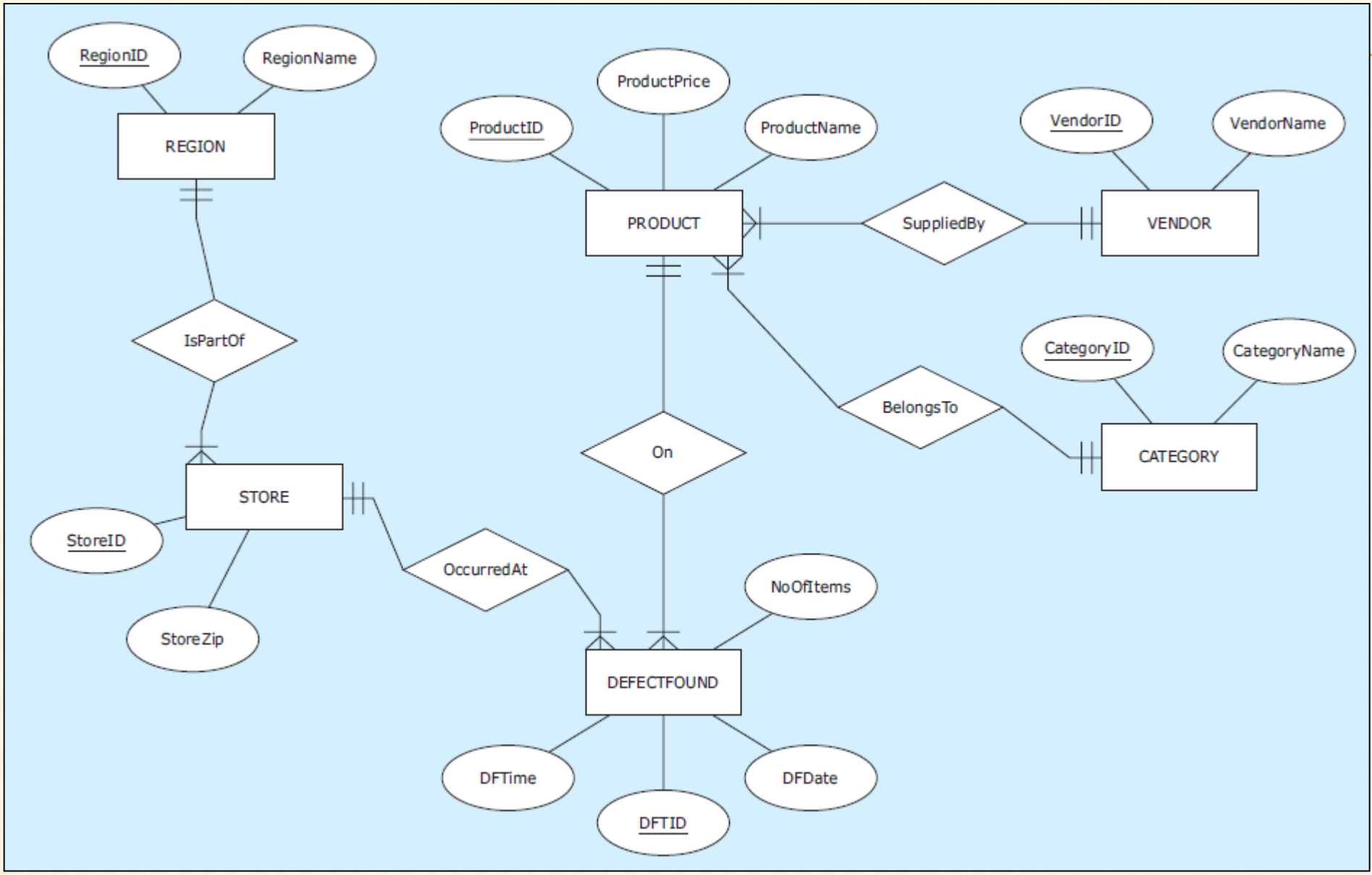
DIMENSIONAL MODELING

- **Multiple facts in a dimensional model**
  - When multiple subjects of analysis can share the same dimensions, a dimensional model contains more than one fact table
  - A dimensional model with multiple fact tables is referred to as a **constellation** or **galaxy of stars**
  - This approach enables:
    - Quicker development of analytical databases for multiple subjects of analysis, because dimensions are re-used instead of duplicated
    - Straightforward cross-fact analysis

Copyright (c) 2014 Pearson Education, Inc.

Expanded Example: Dimensional Model Based on Multiple Sources

ER diagram : ZAGI Retail Company Quality Control Database (Source 4)



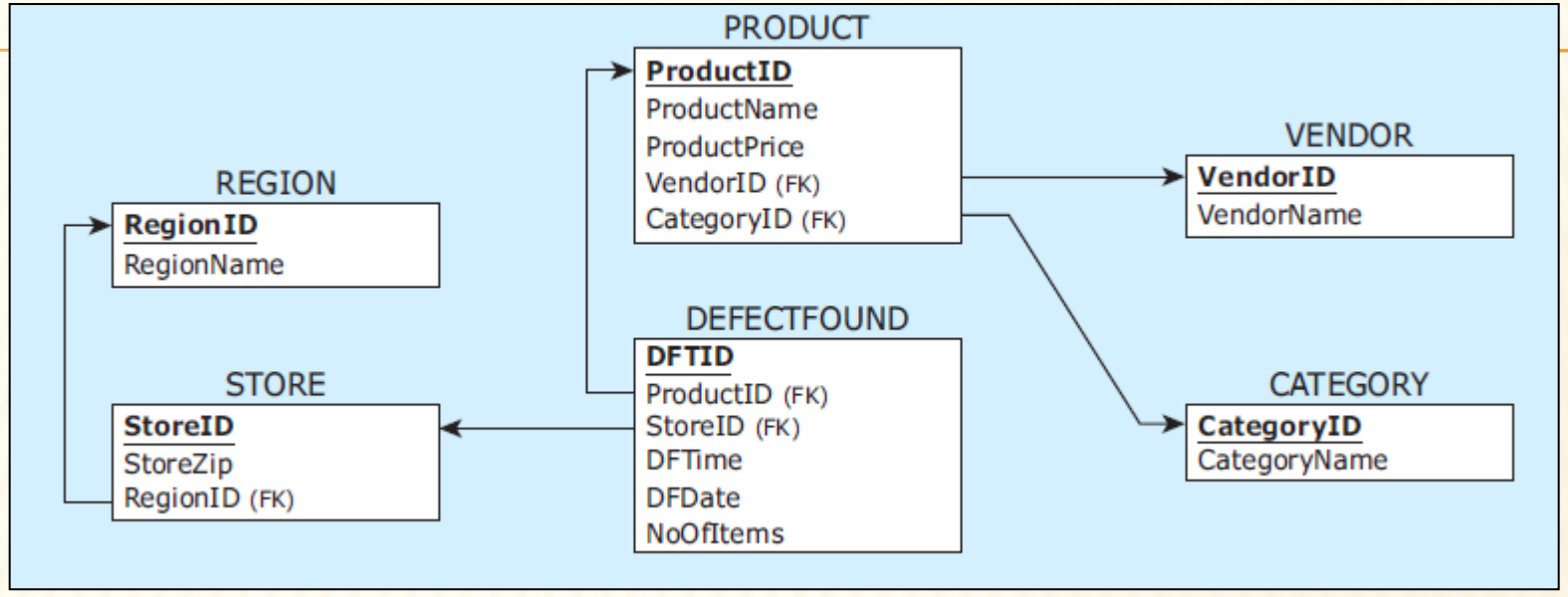
Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Pages 240-242.

Expanded Example: Dimensional Model Based on Multiple Sources

Relational schema and data records: ZAGI Retail Company Quality Control Database (Source 4)



REGION

RegionID	RegionName
C	Chicagoland
T	Tristate

PRODUCT

ProductID	ProductName	ProductPrice	VendorID	CategoryID
1X1	Zzz Bag	$100	PG	CP
2X2	Easy Boot	$70	MK	FW
3X3	Cosy Sock	$15	MK	FW
4X4	Dura Boot	$90	PG	FW
5X5	Tiny Tent	$150	MK	CP
6X6	Biggy Tent	$250	MK	CP

VENDOR

VendorID	VendorName
PG	Pacifica Gear
MK	Mountain King

STORE

StoreID	StoreZip	RegionID
S1	60600	C
S2	60605	C
S3	35400	T

CATEGORY

CategoryID	CategoryName
CP	Camping
FW	Footwear

DEFECTFOUND

DFTID	ProductID	StoreID	DFTDate	DFTTime	NoOfItems
DFT101	1X1	S1	1-Jan-2013	8:00:00 AM	1
DFT202	2X2	S2	1-Jan-2013	8:30:00 AM	2
DFT303	3X3	S3	2-Jan-2013	8:45:00 AM	6

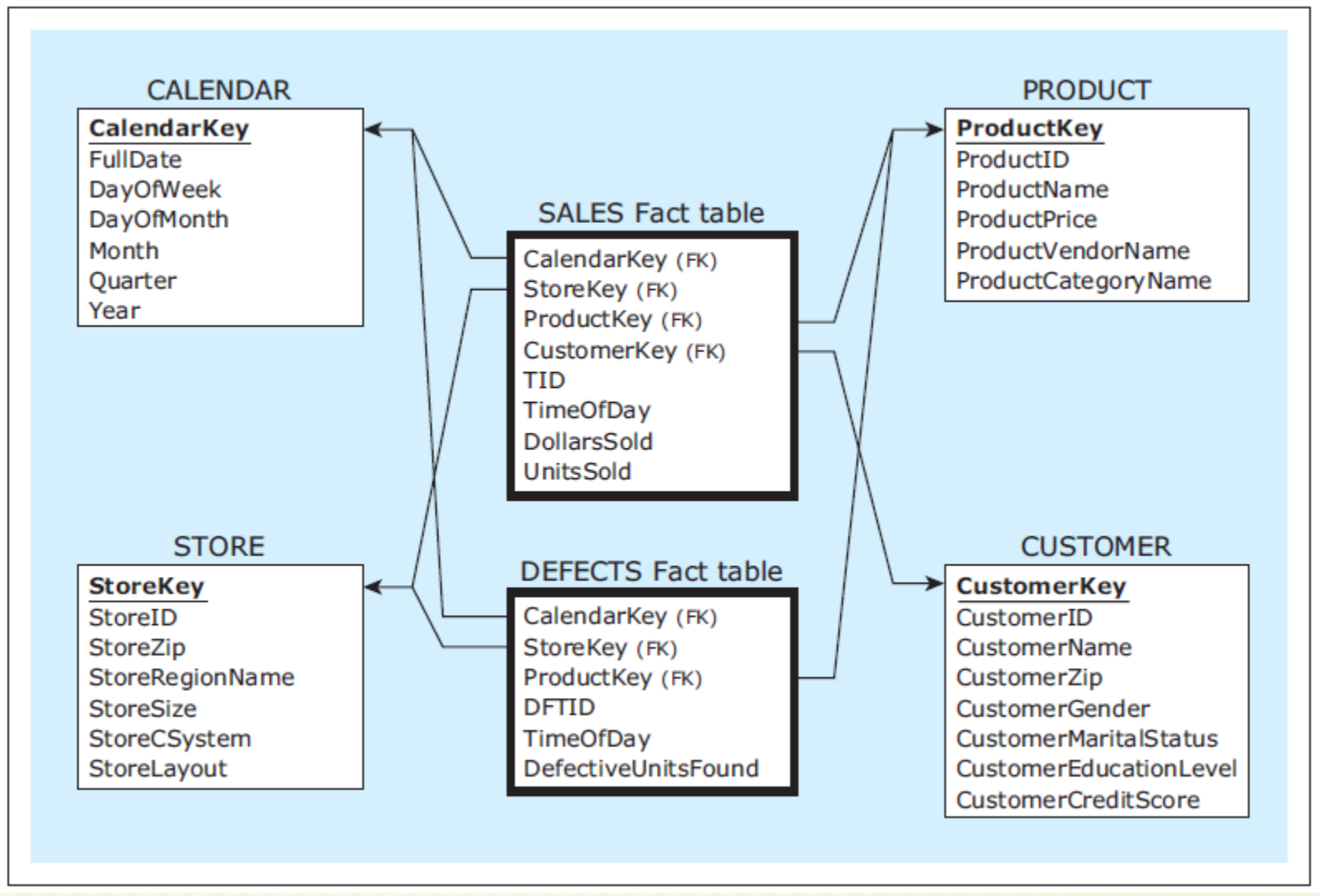
Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Pages 240-242.

Expanded Example: Dimensional Model Based on Multiple Sources

*ZAGI Retail Company dimensional model for the subjects **sales** and **defects***



Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Pages 240-242.

Expanded Example: Dimensional Model Based on Multiple Sources

ZAGI Retail Company dimensional model for the subjects ***sales*** *and* ***defects*** *, populated with the data from the four sources*

CALENDAR Dimension

CalendarKey	FullDate	DayOf Week	DayOf Month	Month	Qtr	Year
1	1/1/2013	Tuesday	1	January	Q1	2013
2	1/2/2013	Wednesday	2	January	Q1	2013

PRODUCT Dimension

ProductKey	ProductID	Product Name	Product Price	Product Vendor Name	Product Category Name
1	1X1	Zzz Bag	$100	Pacifica Gear	Camping
2	2X2	Easy Boot	$70	Mountain King	Footwear
3	3X3	Cosy Sock	$15	Mountain King	Footwear
4	4X4	Dura Boot	$90	Pacifica Gear	Footwear
5	5X5	Tiny Tent	$150	Mountain King	Camping
6	6X6	Biggy Tent	$250	Mountain King	Camping

STORE Dimension

StoreKey	StoreID	StoreZip	StoreRegion Name	Store Size (m^2)	Store CSystem	Store Layout
1	S1	60600	Chicagoland	51000	Cashiers	Modern
2	S2	60605	Chicagoland	35000	Self Service	Traditional
3	S3	35400	Tristate	55000	Mixed	Traditional

CUSTOMER Dimension

CustomerKey	CustomerID	Customer Name	Customer Zip	Customer Gender	Customer MaritalStatus	Customer EducationLevel	Customer CreditScore
1	1-2-333	Tina	60137	Female	Single	College	700
2	2-3-444	Tony	60611	Male	Single	High School	650
3	3-4-555	Pam	35401	Female	Married	College	623

SALES Fact table

CalendarKey	StoreKey	ProductKey	CustomerKey	TID	TimeOfDay	DollarsSold	UnitsSold
1	1	1	1	T111	8:23:59 AM	$100	1
1	2	2	2	T222	8:24:30 AM	$70	1
2	3	3	1	T333	8:15:08 AM	$75	5
2	3	1	1	T333	8:15:08 AM	$100	1
2	3	4	3	T444	8:20:33 AM	$90	1
2	3	2	3	T444	8:20:33 AM	$140	2
2	3	4	2	T555	8:30:00 AM	$360	4
2	3	5	2	T555	8:30:00 AM	$300	2
2	3	6	2	T555	8:30:00 AM	$250	1

DEFECTS Fact table

CalendarKey	StoreKey	ProductKey	DFTID	TimeOfDay	DefectiveUnitsFound
1	1	1	DFT101	8:00:00 AM	1
1	2	2	DFT202	8:30:00 AM	2
2	3	3	DFT303	8:45:00 AM	6

Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Pages 240-242.

DIMENSIONAL MODELING

- **Detailed versus aggregated fact tables**
  - Fact tables in a dimensional model can contain either detailed data or aggregated data
  - In **detailed fact tables** each record refers to a single fact
  - In **aggregated fact tables** each record summarizes multiple facts

Copyright (c) 2014 Pearson Education, Inc.

Troy University Resources

Detailed and Aggregated Fact Table Examples

*ZAGI Retail Company Sales Department Database (Source 1) with **additional** data records included in SALESTRANSACTION and SOLDVIA tables*

REGION

RegionID	RegionName
C	Chicagoland
T	Tristate

PRODUCT

ProductID	ProductName	ProductPrice	VendorID	CategoryID
1X1	Zzz Bag	$100	PG	CP
2X2	Easy Boot	$70	MK	FW
3X3	Cosy Sock	$15	MK	FW
4X4	Dura Boot	$90	PG	FW
5X5	Tiny Tent	$150	MK	CP
6X6	Biggy Tent	$250	MK	CP

VENDOR

VendorID	VendorName
PG	Pacifica Gear
MK	Mountain King

STORE

StoreID	StoreZip	RegionID
S1	60600	C
S2	60605	C
S3	35400	T

CATEGORY

CategoryID	CategoryName
CP	Camping
FW	Footwear

SALESTRANSACTION

TID	CustomerID	StoreID	TDate	TTime
T111	1-2-333	S1	1-Jan-2013	8:23:59 AM
T222	2-3-444	S1	1-Jan-2013	8:24:30 AM
T333	1-2-333	S3	2-Jan-2013	8:15:08 AM
T444	3-4-555	S3	2-Jan-2013	8:20:33 AM
T555	2-3-444	S3	2-Jan-2013	8:30:00 AM
T666	2-3-444	S3	2-Jan-2013	9:30:00 AM

SOLDVIA

ProductID	TID	NoOfItems
1X1	T111	1
2X2	T222	1
3X3	T333	5
1X1	T333	1
4X4	T444	1
2X2	T444	2
4X4	T555	4
5X5	T555	2
6X6	T555	1
5X5	T666	2
6X6	T666	1

CUSTOMER

CustomerID	CustomerName	CustomerZip
1-2-333	Tina	60137
2-3-444	Tony	60611
3-4-555	Pam	35401

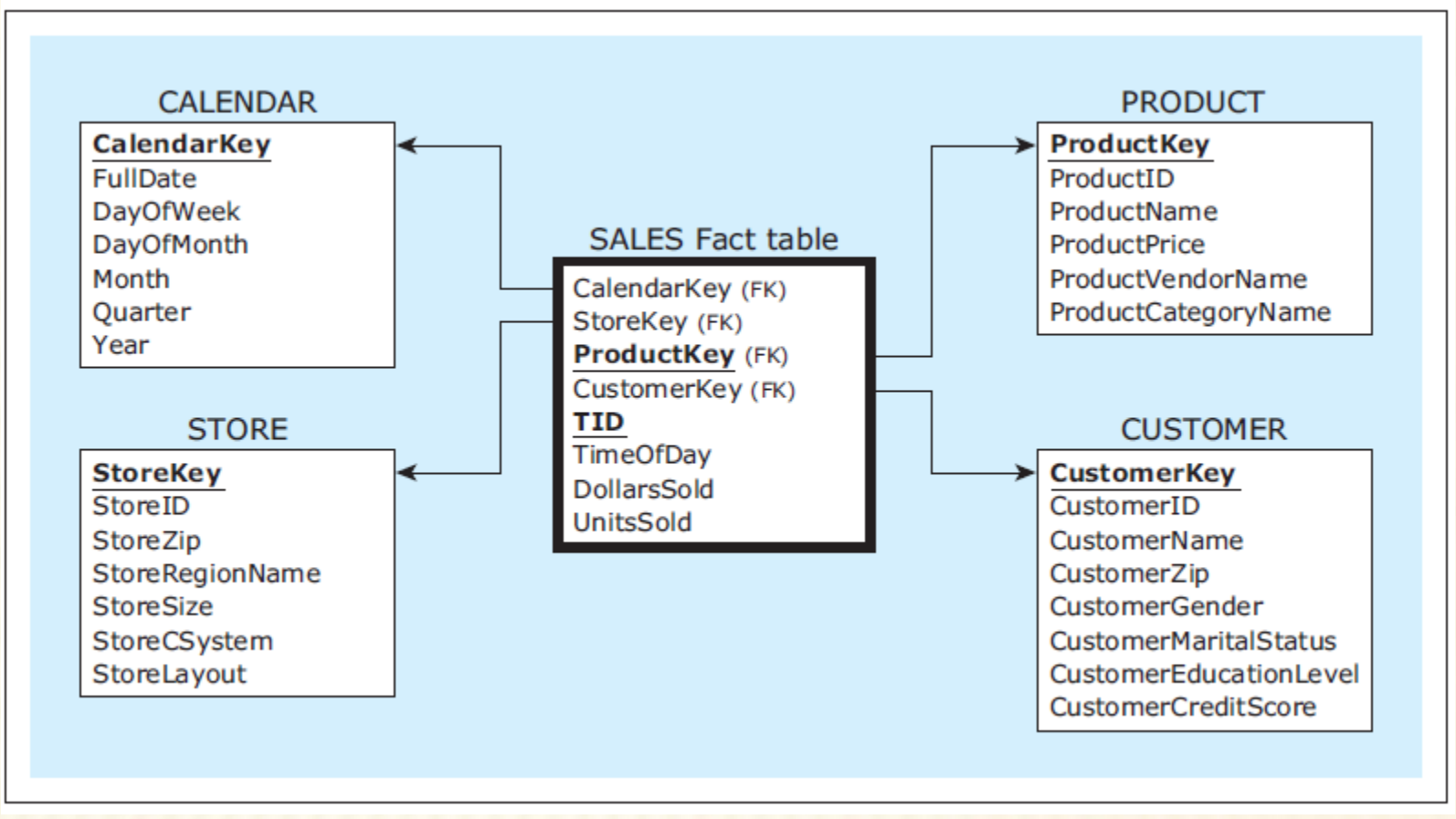
Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Pages 243-248.

Detailed Fact Table Example

ZAGI Retail Company dimensional model for the subject ***sales***



Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Pages 243-248.

Detailed Fact Table Example

ZAGI Retail Company dimensional model for the subject sales, populated with the additional data records from Source 1

CALENDAR Dimension

CalendarKey	FullDate	DayOf Week	DayOf Month	Month	Qtr	Year
1	1/1/2013	Tuesday	1	January	Q1	2013
2	1/2/2013	Wednesday	2	January	Q1	2013

PRODUCT Dimension

ProductKey	ProductID	Product Name	Product Price	Product Vendor Name	Product Category Name
1	1X1	Zzz Bag	$100	Pacifica Gear	Camping
2	2X2	Easy Boot	$70	Mountain King	Footwear
3	3X3	Cosy Sock	$15	Mountain King	Footwear
4	4X4	Dura Boot	$90	Pacifica Gear	Footwear
5	5X5	Tiny Tent	$150	Mountain King	Camping
6	6X6	Biggy Tent	$250	Mountain King	Camping

STORE Dimension

StoreKey	StoreID	StoreZip	StoreRegion Name	Store Size (m^2)	Store CSystem	Store Layout
1	S1	60600	Chicagoland	51000	Cashiers	Modern
2	S2	60605	Chicagoland	35000	Self Service	Traditional
3	S3	35400	Tristate	55000	Mixed	Traditional

CUSTOMER Dimension

CustomerKey	CustomerID	Customer Name	Customer Zip	Customer Gender	Customer MaritalStatus	Customer EducationLevel	Customer CreditScore
1	1-2-333	Tina	60137	Female	Single	College	700
2	2-3-444	Tony	60611	Male	Single	High School	650
3	3-4-555	Pam	35401	Female	Married	College	623

SALES Fact table

CalendarKey	StoreKey	ProductKey	CustomerKey	TID	TimeOfDay	DollarsSold	UnitsSold
1	1	1	1	T111	8:23:59 AM	$100	1
1	2	2	2	T222	8:24:30 AM	$70	1
2	3	3	1	T333	8:15:08 AM	$75	5
2	3	1	1	T333	8:15:08 AM	$100	1
2	3	4	3	T444	8:20:33 AM	$90	1
2	3	2	3	T444	8:20:33 AM	$140	2
2	3	4	2	T555	8:30:00 AM	$360	4
2	3	5	2	T555	8:30:00 AM	$300	2
2	3	6	2	T555	8:30:00 AM	$250	1
2	3	5	2	T666	9:30:00 AM	$300	2
2	3	6	2	T666	9:30:00 AM	$250	1

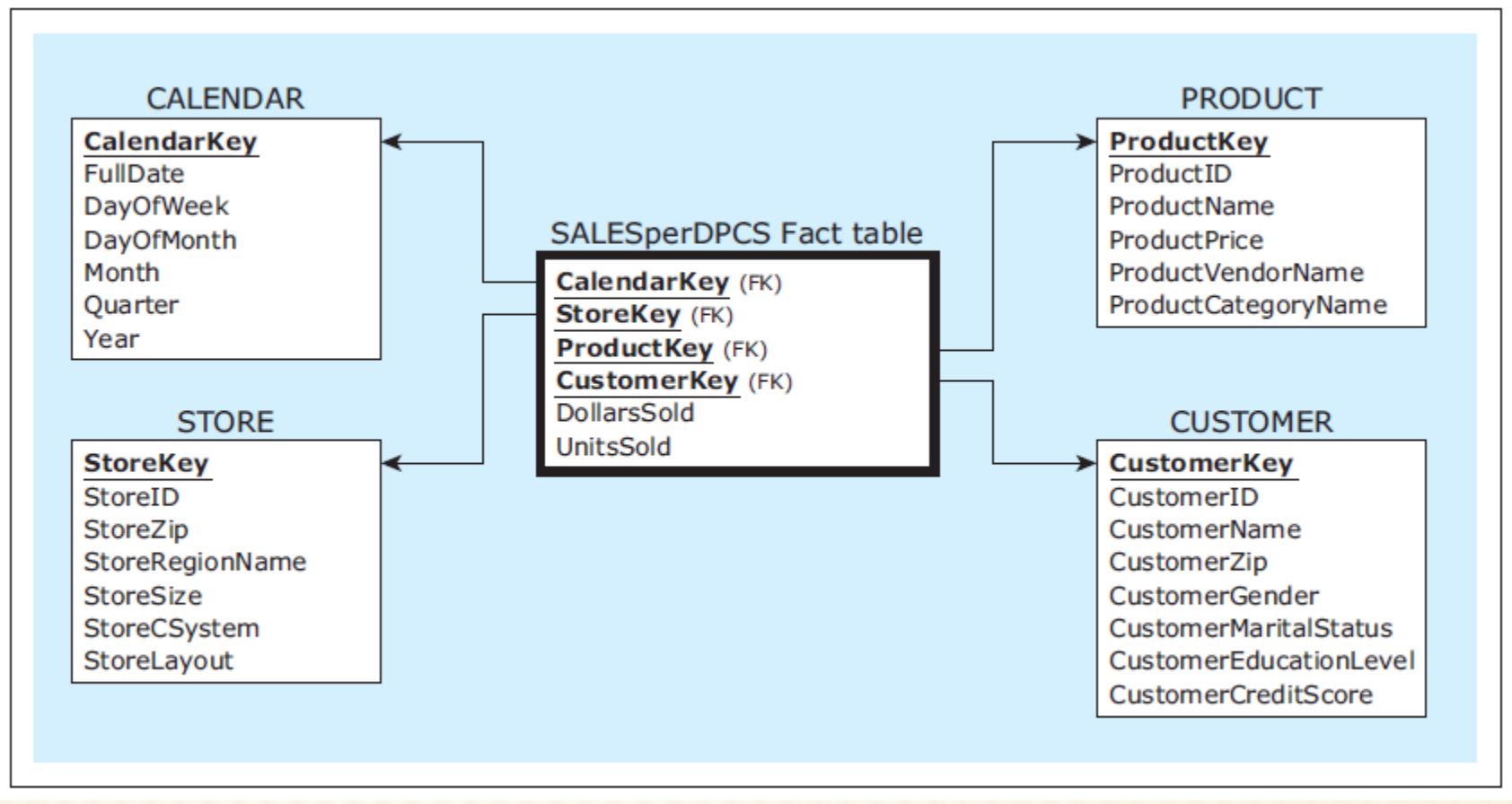
Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Pages 243-248.

Aggregated Fact Table Example 1

ZAGI Retail Company dimensional model with an aggregated fact table ***Sales per day, product, customer, and store***



Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Pages 243-248.

Aggregated Fact Table Example 1

ZAGI Retail Company dimensional model for the subject sales with an aggregated fact table ***Sales per day, product, customer, store,*** *populated with the data*

CALENDAR Dimension

CalendarKey	FullDate	DayOf Week	DayOf Month	Month	Qtr	Year
1	1/1/2013	Tuesday	1	January	Q1	2013
2	1/2/2013	Wednesday	2	January	Q1	2013

PRODUCT Dimension

ProductKey	ProductID	Product Name	Product Price	Product Vendor Name	Product Category Name
1	1X1	Zzz Bag	$100	Pacifica Gear	Camping
2	2X2	Easy Boot	$70	Mountain King	Footwear
3	3X3	Cosy Sock	$15	Mountain King	Footwear
4	4X4	Dura Boot	$90	Pacifica Gear	Footwear
5	5X5	Tiny Tent	$150	Mountain King	Camping
6	6X6	Biggy Tent	$250	Mountain King	Camping

STORE Dimension

StoreKey	StoreID	StoreZip	StoreRegion Name	Store Size (m^2)	Store CSystem	Store Layout
1	S1	60600	Chicagoland	51000	Cashiers	Modern
2	S2	60605	Chicagoland	35000	Self Service	Traditional
3	S3	35400	Tristate	55000	Mixed	Traditional

CUSTOMER Dimension

CustomerKey	CustomerID	Customer Name	Customer Zip	Customer Gender	Customer MaritalStatus	Customer EducationLevel	Customer CreditScore
1	1-2-333	Tina	60137	Female	Single	College	700
2	2-3-444	Tony	60611	Male	Single	High School	650
3	3-4-555	Pam	35401	Female	Married	College	623

SALESPerDPCS Fact table

CalendarKey	StoreKey	ProductKey	CustomerKey	DollarsSold	UnitsSold
1	1	1	1	$100	1
1	2	2	2	$70	1
2	3	3	1	$75	5
2	3	1	1	$100	1
2	3	4	3	$90	1
2	3	2	3	$140	2
2	3	4	2	$360	4
2	3	5	2	$600	4
2	3	6	2	$500	2

← *Amounts from 8th and 10th records in SALES fact table in Figure 8.23 combined (added)*

← *Amounts from 9th and 11th records in SALES fact table in Figure 8.23 combined (added)*

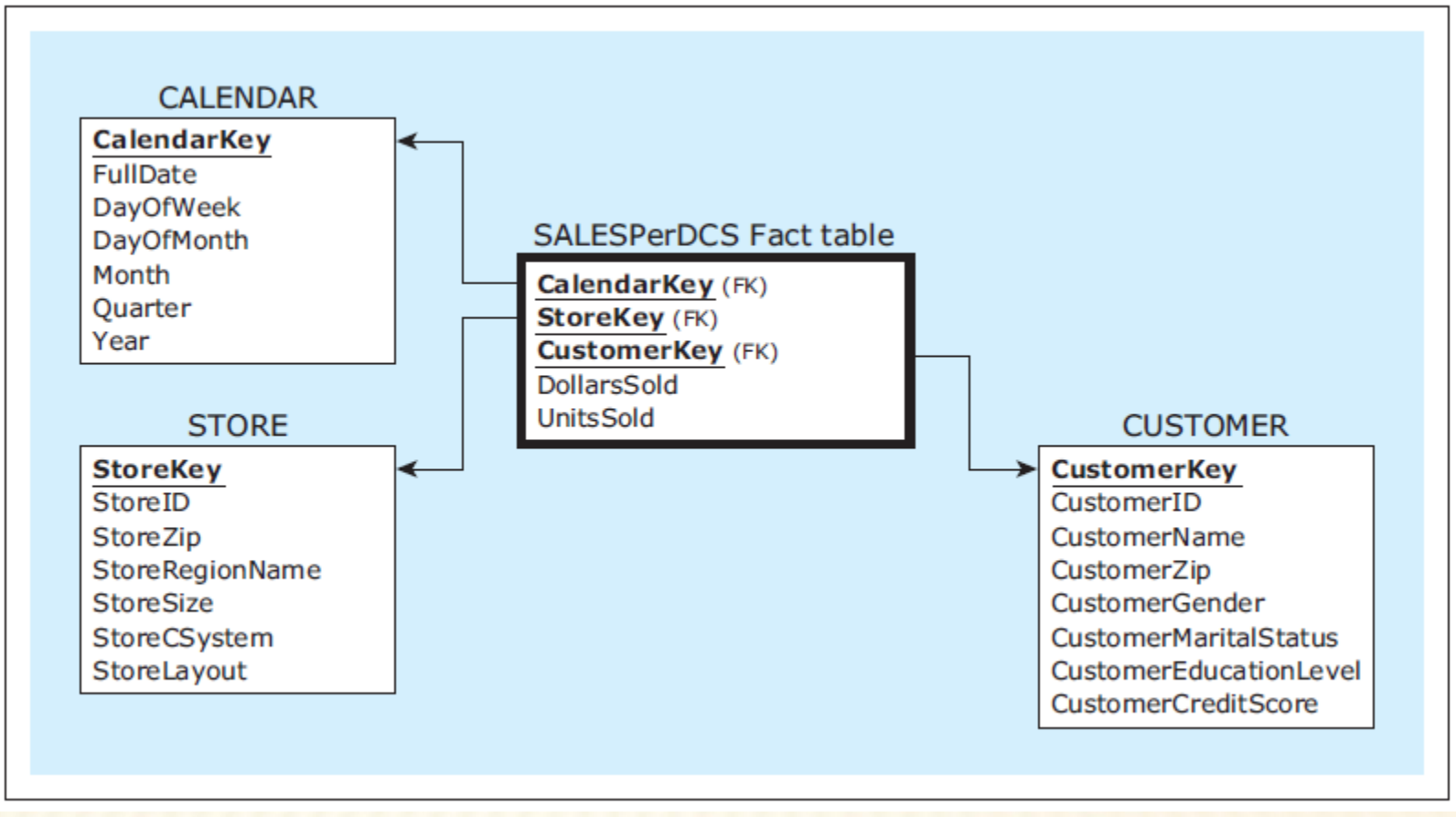
Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Pages 243-248.

Aggregated Fact Table Example 2

ZAGI Retail Company star schema with an aggregated fact table ***Sales per day, customer, and store***



Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Pages 243-248.

Aggregated Fact Table Example 2

ZAGI Retail Company dimensional model for the subject sales with an aggregated fact table ***Sales per day, customer, store,*** *populated with the data*

CALENDAR Dimension

CalendarKey	FullDate	DayOf Week	DayOf Month	Month	Qtr	Year
1	1/1/2013	Tuesday	1	January	Q1	2013
2	1/2/2013	Wednesday	2	January	Q1	2013

STORE Dimension

StoreKey	StoreID	StoreZip	StoreRegion Name	Store Size (m^2)	Store CSystem	Store Layout
1	S1	60600	Chicagoland	51000	Cashiers	Modern
2	S2	60605	Chicagoland	35000	Self Service	Traditional
3	S3	35400	Tristate	55000	Mixed	Traditional

CUSTOMER Dimension

CustomerKey	CustomerID	Customer Name	Customer Zip	Customer Gender	Customer MaritalStatus	Customer EducationLevel	Customer CreditScore
1	1-2-333	Tina	60137	Female	Single	College	700
2	2-3-444	Tony	60611	Male	Single	High School	650
3	3-4-555	Pam	35401	Female	Married	College	623

SALESPerDCS Fact table

CalendarKey	StoreKey	CustomerKey	DollarsSold	UnitsSold	
1	1	1	$100	1	
1	2	2	$70	1	
2	3	1	$175	6	*Amounts from 3rd and 4th records in SALES fact table in Figure 8.23 combined (added)*
2	3	3	$230	3	*Amounts from 5th and 6th records in SALES fact table in Figure 8.23 combined (added)*
2	3	2	$1,460	10	*Amounts from 7th through 11th records in SALES fact table in Figure 8.23 combined (added)*

Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Pages 243-248.

DIMENSIONAL MODELING

- Granularity of the fact tables
  - Granularity describes what is depicted by one row in the fact table
  - Detailed fact tables have fine level of granularity because each record represents a single fact
  - Aggregated fact tables have a coarser level of granularity than detailed fact tables as records in aggregated fact tables always represent summarizations of multiple facts

Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

SALES Per DPCS fact table has a coarser level of granularity than the SALES fact tables because records in the SALES Per DPCS fact table summarize records from the SALES fact table. SALES Per DCS fact table has an even coarser level of granularity, because its records summarize records from the SALES Per DPCS fact tables.

Troy University Resources

DIMENSIONAL MODELING

- Granularity of the fact tables
  - Due to their compactness, coarser granularity aggregated fact tables are quicker to query than detailed fact tables
  - Coarser granularity tables are limited in terms of what information can be retrieved from them
  - One way to take advantage of the query performance improvement provided by aggregated fact tables, while retaining the power of analysis of detailed fact tables, is to have both types of tables coexisting within the same dimensional model, i.e. in the same constellation

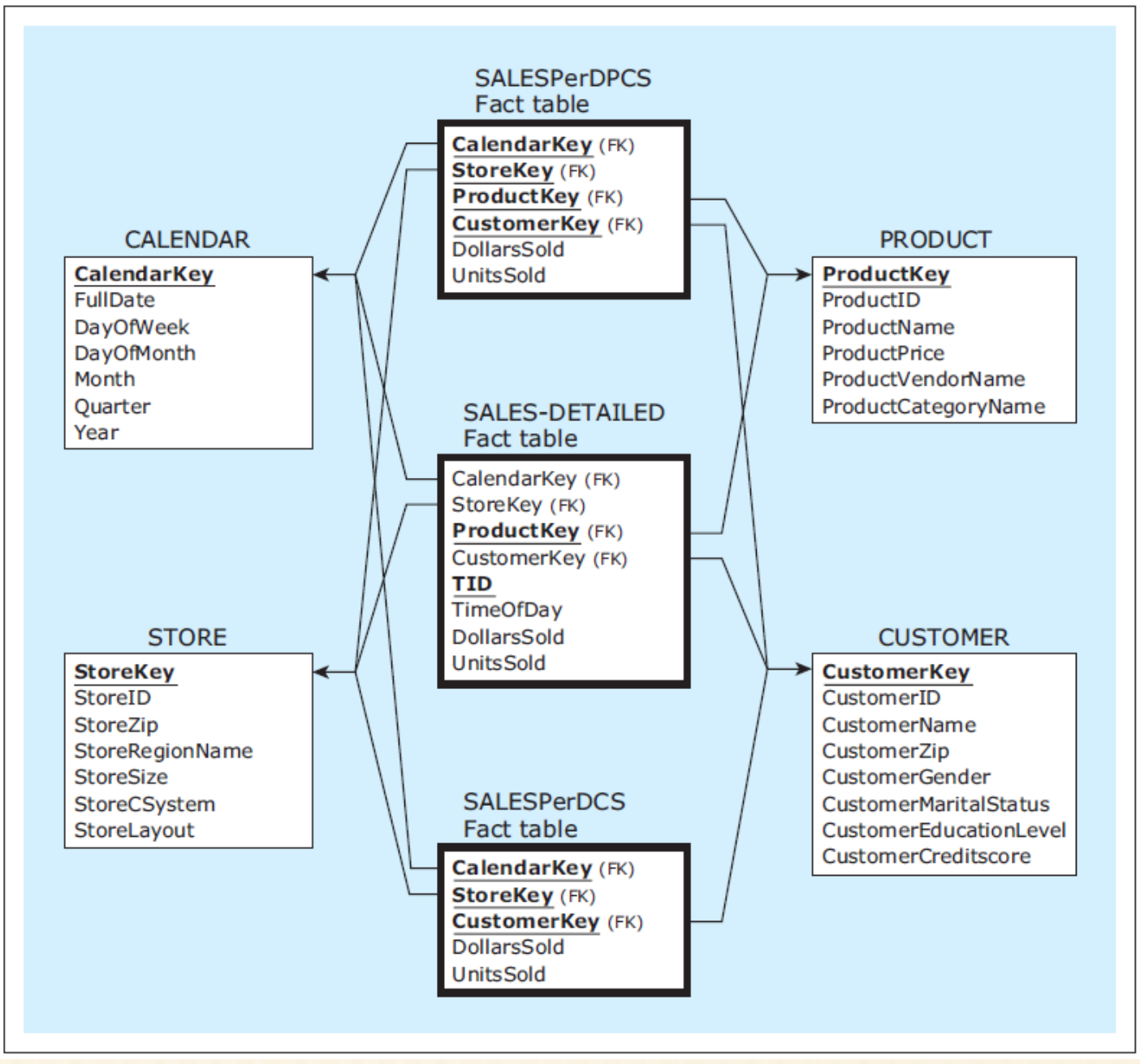
Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

Aggregation is requirement-specific while a detailed granularity provides unlimited possibility for analysis.

You can always obtain an aggregation from the finest grain, but the reverse is not true.

A constellation of detailed and aggregated facts - Example



Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

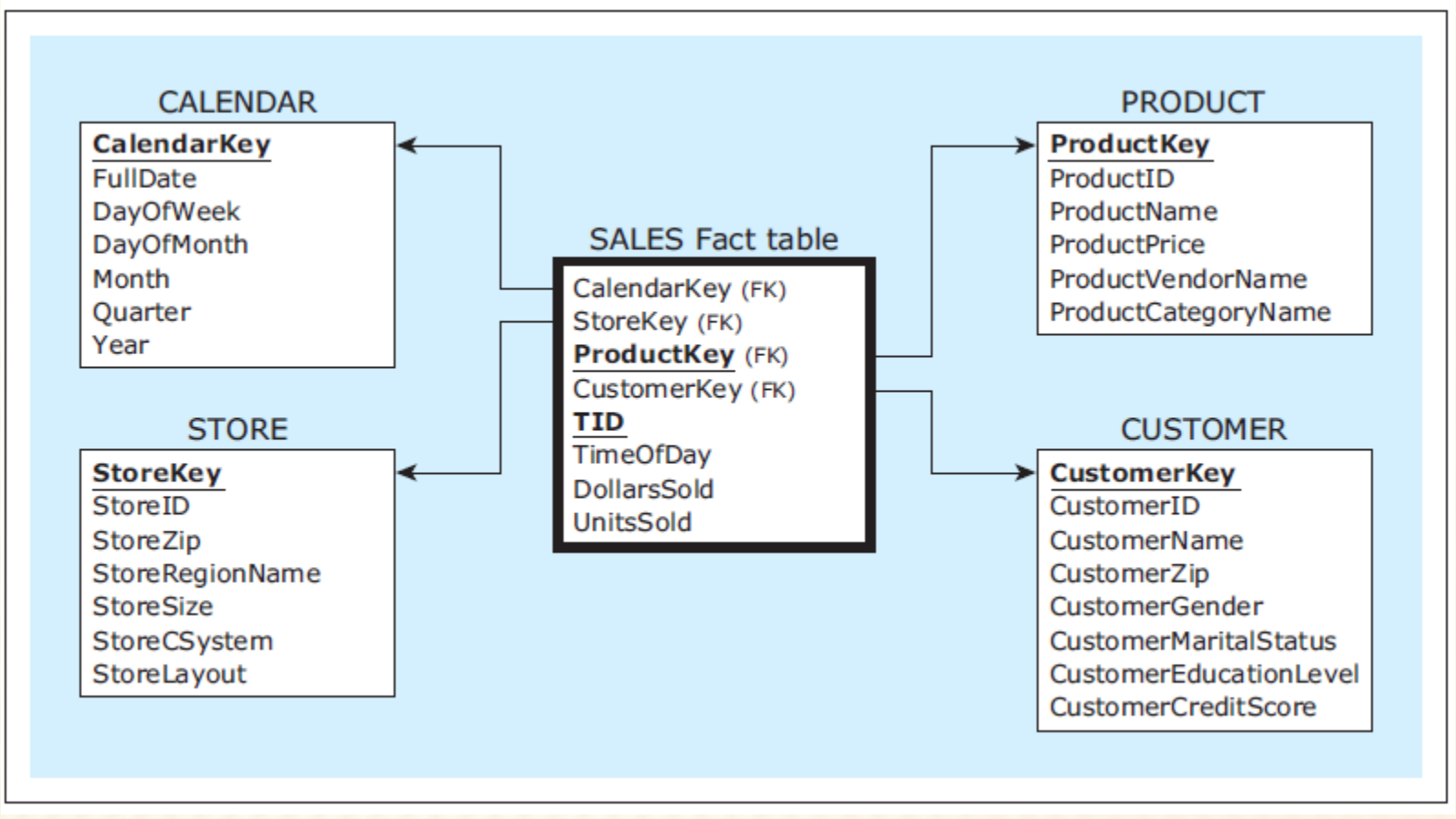
This example is discussed on Pages 243-248.

DIMENSIONAL MODELING

- Line-item versus transaction-level detailed fact table
  - Line-item detailed fact table
    - Each row represents a line item of a particular transaction
  - Transaction-level detailed fact table
    - Each row represents a particular transaction

Copyright (c) 2014 Pearson Education, Inc.

Line-Item Detailed Fact Table Example

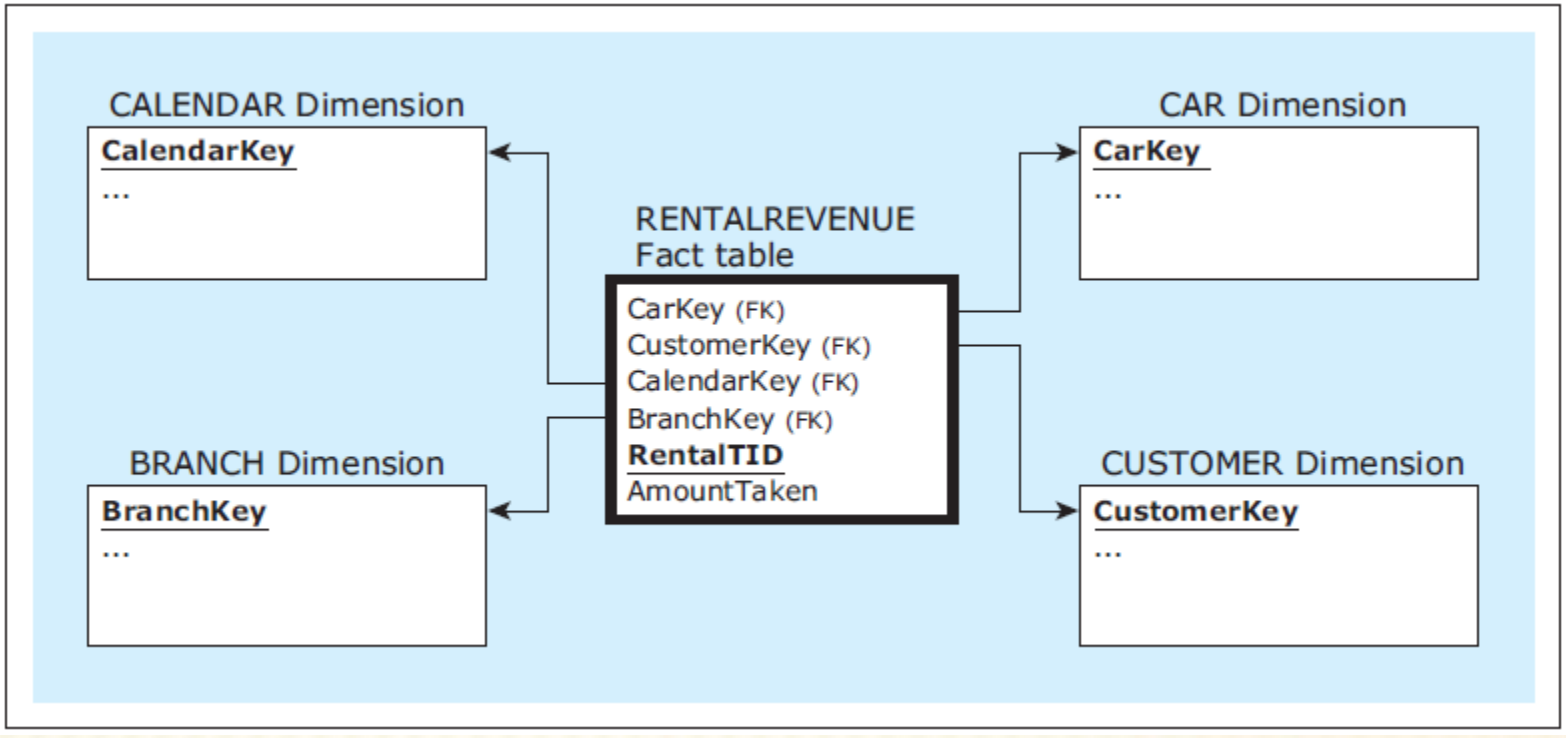


Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Page 249.

Transaction-Level Detailed Fact Table Example



Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Pages 249-250.

DIMENSIONAL MODELING

- **Slowly Changing Dimension**
  - Typical dimension in a star schema contains:
    - Attributes whose values do not change (or change extremely rarely) such as store size and customer gender
    - Attributes whose values change occasionally and sporadically over time, such as customer zip and employee salary.
  - Dimension that contains attributes whose values can change referred to as a **slowly changing dimension**
  - Most common approaches to dealing with slowly changing dimensions
    - **Type 1**
    - **Type 2**
    - **Type 3**

Copyright (c) 2014 Pearson Education, Inc.

Troy University Resources

DIMENSIONAL MODELING

- Type 1
  - Changes the value in the dimension's record
    - The new value replaces the old value.
  - No history is preserved
  - The simplest approach, used most often when a change in a dimension is the result of an error

Copyright (c) 2014 Pearson Education, Inc.

Type 1 Example

Susan's Tax Bracket attribute value changes from *Medium* to *High*

CUSTOMER

CustomerKey	CustomerID	CustomerName	TaxBracket
1	111	Linda	Low
2	222	Susan	Medium
3	333	William	High

CUSTOMER

CustomerKey	CustomerID	CustomerName	TaxBracket
1	111	Linda	Low
2	222	Susan	High
3	333	William	High

Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Pages 250-251.

Troy University Resources

DIMENSIONAL MODELING

- Type 2
  - Creates a new additional dimension record using a new value for the surrogate key every time a value in a dimension record changes
  - Used in cases where history should be preserved
  - Can be combined with the use of *timestamps* and *row indicators*
    - **Timestamps** - columns that indicates the time interval for which the values in the records are applicable
    - **Row indicator** - column that provides a quick indicator of whether the record is currently valid

Copyright (c) 2014 Pearson Education, Inc.

Type 2 Example

Susan's Tax Bracket attribute value changes from *Medium* to *High*

CUSTOMER

CustomerKey	CustomerID	CustomerName	TaxBracket
1	111	Linda	Low
2	222	Susan	Medium
3	333	William	High

CUSTOMER

CustomerKey	CustomerID	CustomerName	TaxBracket
1	111	Linda	Low
2	222	Susan	Medium
3	333	William	High
4	222	Susan	High

Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Pages 251-252.

Troy University Resources

Type 2 Example (with timestamps and row indicator)

Susan's Tax Bracket attribute value changes from *Medium* to *High*

CUSTOMER

CustomerKey	CustomerID	CustomerName	TaxBracket
1	111	Linda	Low
2	222	Susan	Medium
3	333	William	High

CUSTOMER

CustomerKey	CustomerID	CustomerName	TaxBracket	Effective StartDate	Effective EndDate	Row Indicator
1	111	Linda	Low	1.1.2000	n/a	Current
2	222	Susan	Medium	1.1.2000	12.31.2007	Not Current
3	333	William	High	1.1.2000	n/a	Current
4	222	Susan	High	1.1.2008	n/a	Current

Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Page 252.

DIMENSIONAL MODELING

- Type 3
  - Involves creating a “previous” and “current” column in the dimension table for each column where changes are anticipated
  - Applicable in cases in which there is a fixed number of changes possible per column of a dimension, or in cases when only a limited history is recorded.
  - Can be combined with the use of *timestamps*

Copyright (c) 2014 Pearson Education, Inc.

Troy University Resources

Type 3 Example

Susan's Tax Bracket attribute value changes from *Medium* to *High*

CUSTOMER

CustomerKey	CustomerID	CustomerName	TaxBracket
1	111	Linda	Low
2	222	Susan	Medium
3	333	William	High

CUSTOMER

CustomerKey	CustomerID	CustomerName	Previous TaxBracket	Current TaxBracket
1	111	Linda	n/a	Low
2	222	Susan	Medium	High
3	333	William	n/a	High

Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Page 253.

Type 3 Example (with timestamps)

Susan's Tax Bracket attribute value changes from *Medium* to *High*

CUSTOMER

CustomerKey	CustomerID	CustomerName	TaxBracket
1	111	Linda	Low
2	222	Susan	Medium
3	333	William	High

CUSTOMER

CustomerKey	CustomerID	CustomerName	Previous TaxBracket	Previous TaxBracket EffectiveDate	Current TaxBracket	Current TaxBracket EffectiveDate
1	111	Linda	n/a	n/a	Low	1.1.2000
2	222	Susan	Medium	1.1.2000	High	1.1.2008
3	333	William	n/a	n/a	High	1.1.2000

Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Page 253.

Troy University Resources

DIMENSIONAL MODELING

- Snowflake model
  - A star schema that contains the dimensions that are normalized
  - Snowflaking is usually not used in dimensional modeling
    - Not-normalized (not snowflaked) dimensions provide for simpler analysis
    - Normalization is usually not necessary for analytical databases

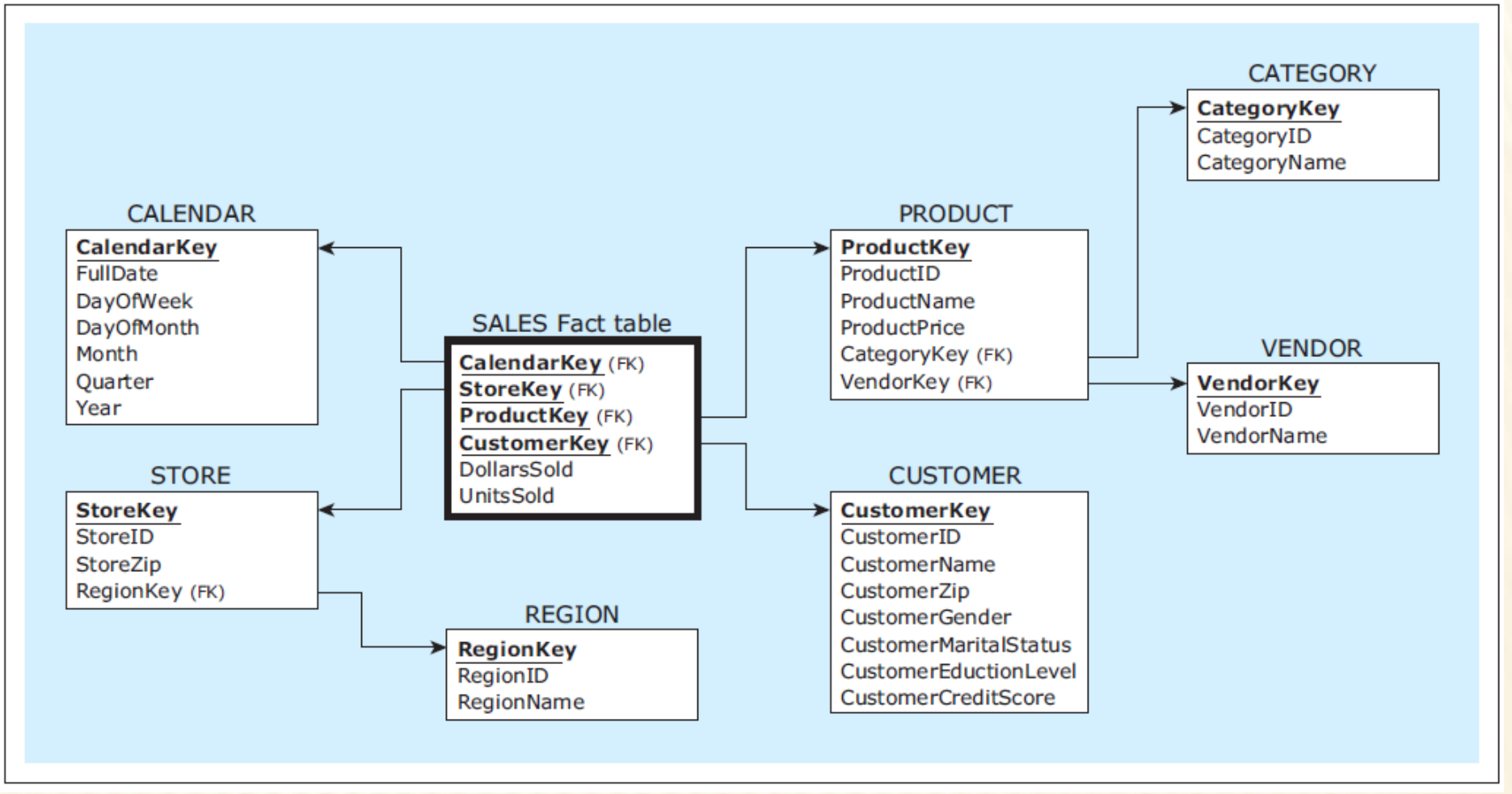
Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

Analytical databases are typically read only. Hence, no danger of update anomalies.

Snowflake Model - Example

A snowflaked version of the *ZAGI Retail Company star schema for the subject sales*



Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Page 254.

Troy University Resources

DATA WAREHOUSE (DATA MART) MODELING APPROACHES

- Data warehouse/data mart modeling approaches
  - Three of the most common data warehouse and data mart modeling approaches:
    - Normalized data warehouse
    - Dimensionally modeled data warehouse
    - Independent data marts

Copyright (c) 2014 Pearson Education, Inc.

DATA WAREHOUSE (DATA MART) MODELING APPROACHES

- **Normalized data warehouse**
  - Envisions a data warehouse as an integrated analytical database modeled by using the traditional database modeling techniques of ER modeling and relational modeling, resulting in a normalized relational database schema
  - Populated with the analytically useful data from the operational data sources via the ETL process
  - Serves as a source of data for dimensionally modeled data marts and for any other non-dimensional analytically useful data sets

Copyright (c) 2014 Pearson Education, Inc.

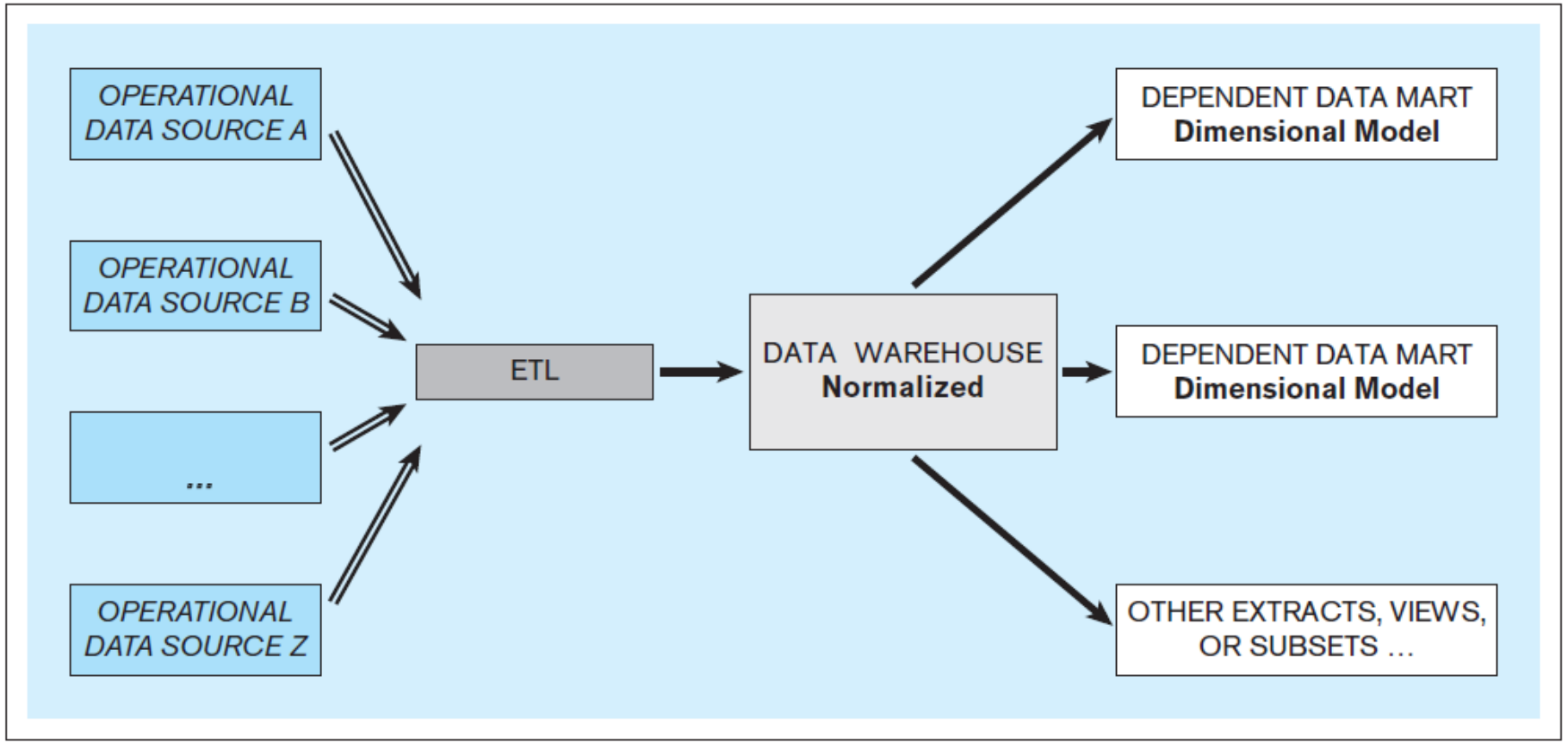
SPEAKER NOTES

Data warehouse as a normalized integrated analytical database was first proposed by Bill Inmon, and hence, the normalized data warehouse approach is often referred to as the Inmon approach.

Troy University Resources

DATA WAREHOUSE (DATA MART) MODELING APPROACHES

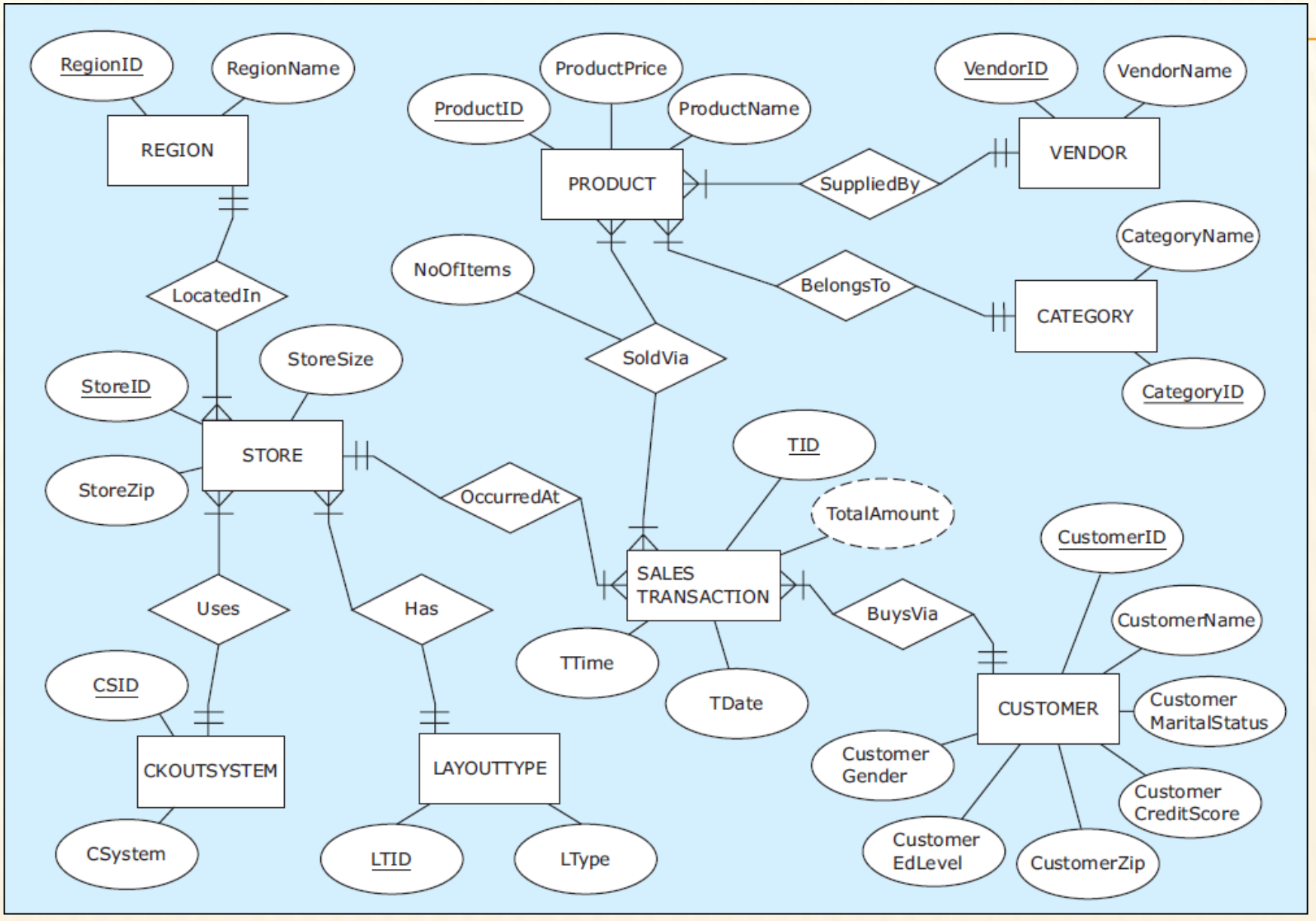
- Normalized data warehouse



Copyright (c) 2014 Pearson Education, Inc.

Normalized Data Warehouse - Example

ER-diagram: *ZAGI Retail Company sales-analysis data warehouse*



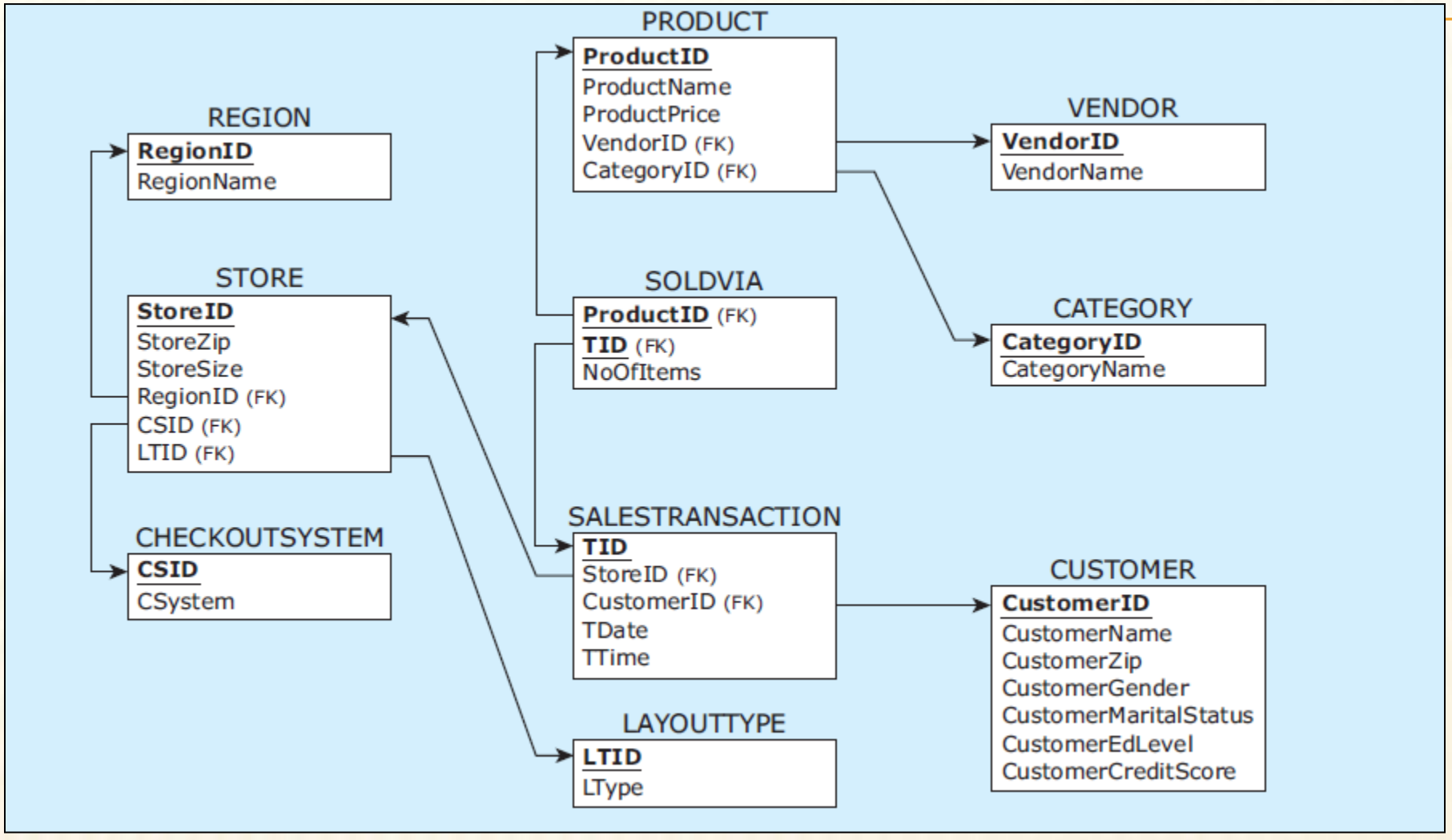
Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Pages 256-259.

Normalized Data Warehouse - Example

Relational schema: *ZAGI Retail Company sales-analysis data warehouse*



Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Pages 256-259.

Normalized Data Warehouse - Example

Data records: *ZAGI Retail Company sales-analysis data warehouse*

PRODUCT

ProductID	ProductName	ProductPrice	VendorID	CategoryID
1X1	Zzz Bag	$100	PG	CP
2X2	Easy Boot	$70	MK	FW
3X3	Cosy Sock	$15	MK	FW
4X4	Dura Boot	$90	PG	FW
5X5	Tiny Tent	$150	MK	CP
6X6	Biggy Tent	$250	MK	CP

CATEGORY

CategoryID	CategoryName
CP	Camping
FW	Footwear

VENDOR

VendorID	VendorName
PG	Pacifica Gear
MK	Mountain King

STORE

StoreID	StoreZip	StoreSize (m^2)	RegionID	CSID	LTID
S1	60600	51000	C	CR	M
S2	60605	35000	C	AC	T
S3	35400	55000	T	MX	T

REGION

RegionID	RegionName
C	Chicagoland
T	Tristate

LAYOUTTYPE

LTID	Layout
M	Modern
T	Traditional

CHECKOUTSYSTEM

CSID	CSystem
CR	Cashiers
AC	Self Service
MX	Mixed

SALESTRANSACTION

TID	CustomerID	StoreID	TDate	TTime
T111	1-2-333	S1	1-Jan-2013	8:23:59 AM
T222	2-3-444	S2	1-Jan-2013	8:24:30 AM
T333	1-2-333	S3	2-Jan-2013	8:15:08 AM
T444	3-4-555	S3	2-Jan-2013	8:20:33 AM
T555	2-3-444	S3	2-Jan-2013	8:30:00 AM

SOLDVIA

ProductID	TID	NoOfItems
1X1	T111	1
2X2	T222	1
3X3	T333	5
1X1	T333	1
4X4	T444	1
2X2	T444	2
4X4	T555	4
5X5	T555	2
6X6	T555	1

CUSTOMER

CustomerID	Customer Name	Customer Zip	Customer Gender	Customer Marital Status	Customer Education Level	Customer Credit Score
1-2-333	Tina	60137	Female	Single	College	700
1-2-333	Tony	60611	Male	Single	High School	650
3-4-555	Pam	35401	Female	Married	College	623

Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Pages 256-259.

Troy University Resources

DATA WAREHOUSE (DATA MART) MODELING APPROACHES

- **Dimensionally modeled data warehouse**
  - Data warehouse as a collection of dimensionally modeled intertwined data marts (i.e. constellation of dimensional models) that integrates analytically useful information from the operational data sources
  - Same as the normalized data warehouse approach when it comes to the utilization of operational data sources and the ETL process

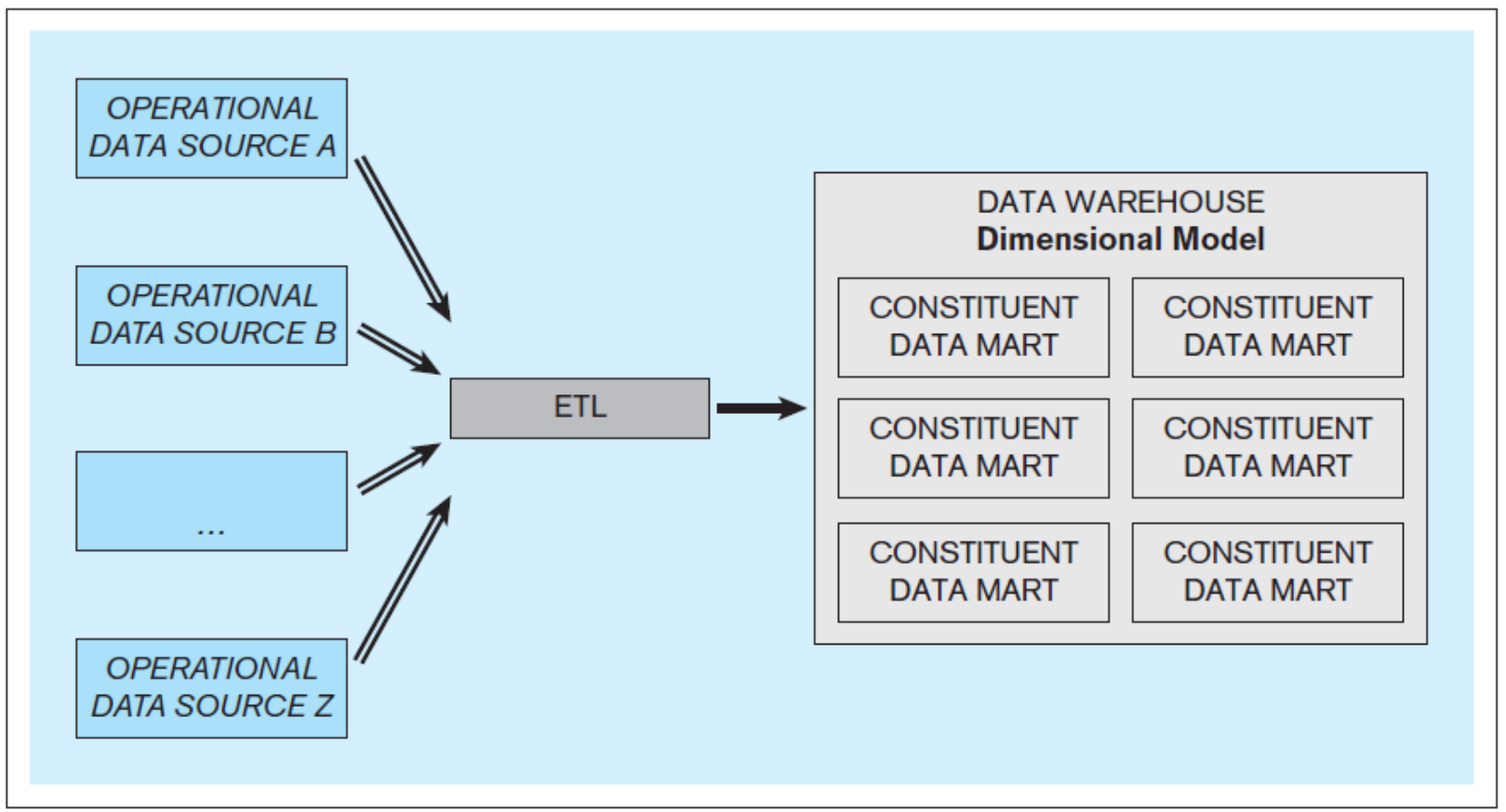
Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

Dimensionally modeled data warehouse approach was championed by Ralph Kimball, and hence, it is often referred to as the Kimball approach.

DATA WAREHOUSE (DATA MART) MODELING APPROACHES

- Dimensionally modeled data warehouse



Copyright (c) 2014 Pearson Education, Inc.

Troy University Resources

DATA WAREHOUSE (DATA MART) MODELING APPROACHES

- **Dimensionally modeled data warehouse**
  - A set of commonly used dimensions known as **conformed dimensions** is designed first
  - Fact tables corresponding to the subjects of analysis are then subsequently added
  - A set of dimensional models is created where each fact table is connected to multiple dimensions, and some of the dimensions are shared by more than one fact table
  - In addition to the originally created set of conformed dimensions, additional dimensions are included as needed
  - The result is a data warehouse that is a collection of intertwined dimensionally modeled data marts, i.e. a constellation of stars

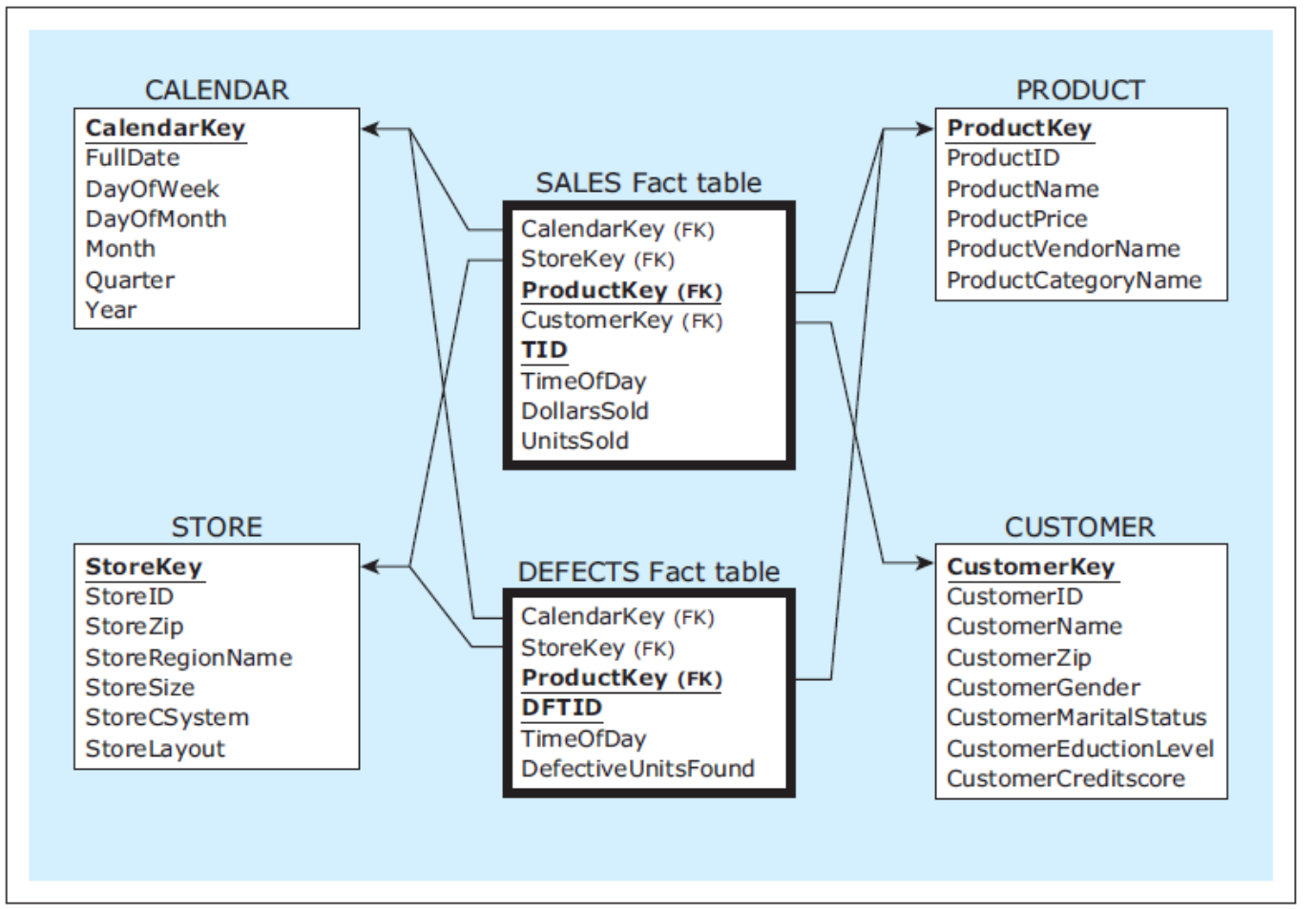
Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

For example, in a retail company, conformed dimensions CALENDAR, PRODUCT, STORE can be designed first, as they will be commonly used by subjects of analysis.

DATA WAREHOUSE (DATA MART) MODELING APPROACHES

- A dimensionally modeled data warehouse with two constituent data marts using conformed dimensions



Copyright (c) 2014 Pearson Education, Inc.

Troy University Resources

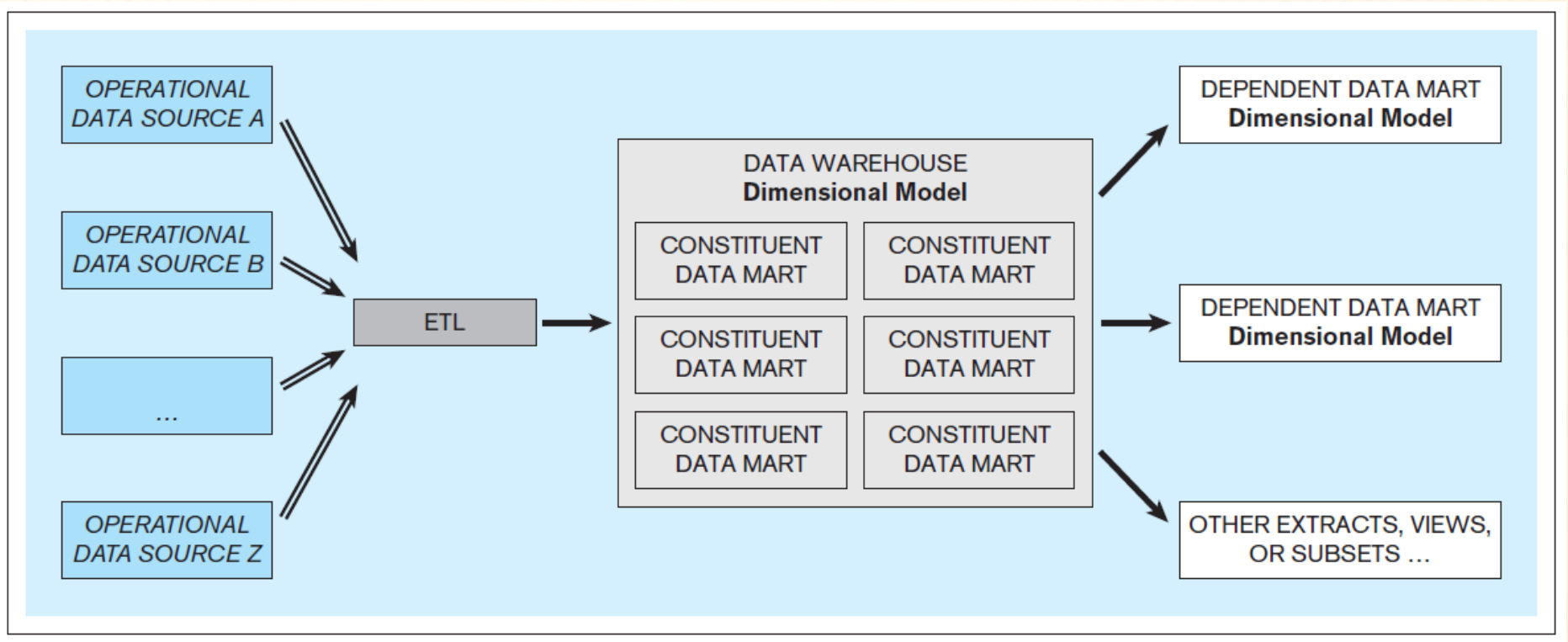
DATA WAREHOUSE (DATA MART) MODELING APPROACHES

- Dimensionally modeled data warehouse
  - Can be used as a source for dependent data marts and other views, subsets, and/or extracts

Copyright (c) 2014 Pearson Education, Inc.

DATA WAREHOUSE (DATA MART) MODELING APPROACHES

- Dimensionally modeled data warehouse

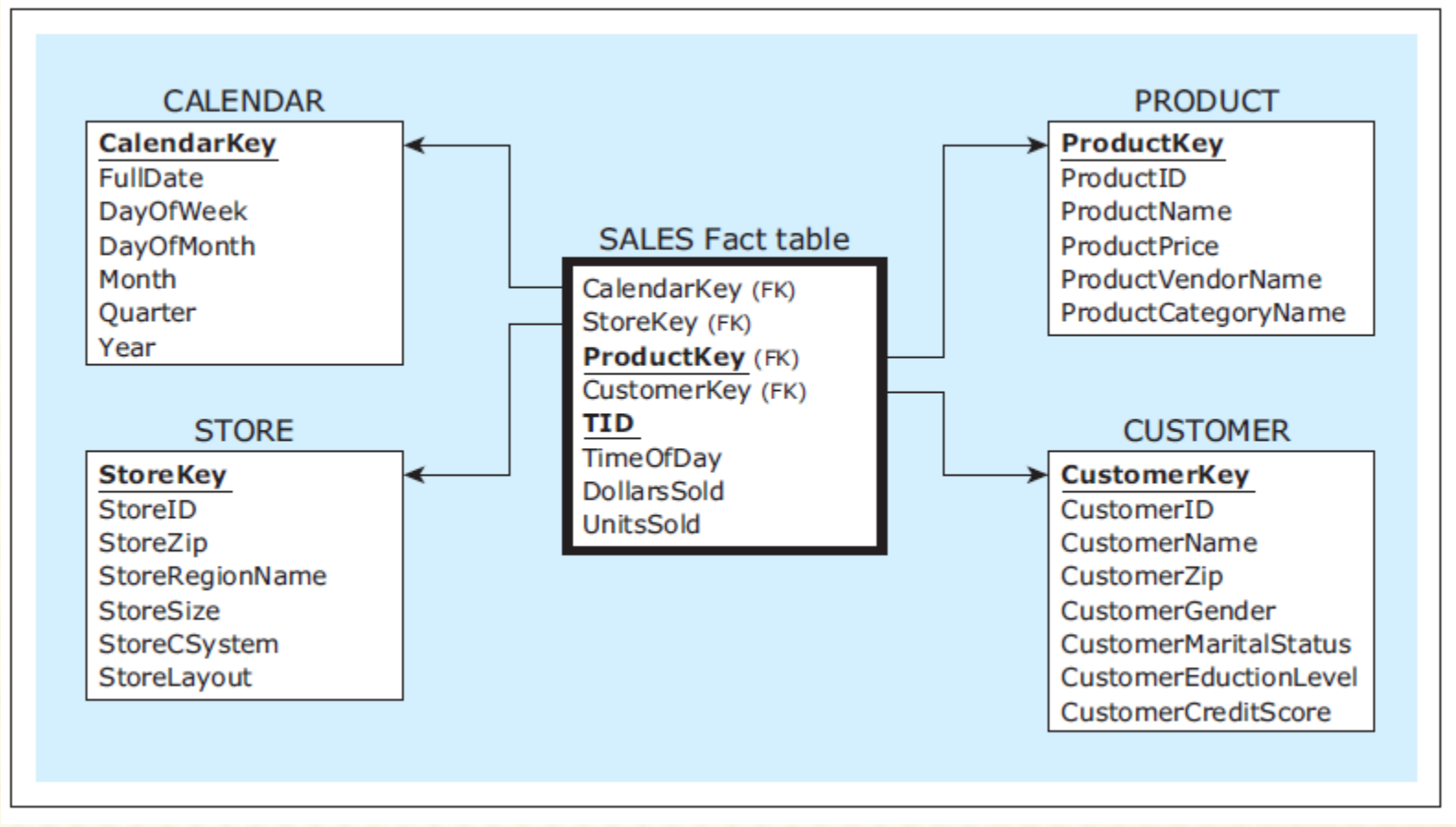


Copyright (c) 2014 Pearson Education, Inc.

Troy University Resources

Dimensionally Modeled Data Warehouse - Example

Star schema: *ZAGI Retail Company sales-analysis data warehouse*



Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Pages 260-262.

Dimensionally Modeled Data Warehouse - Example

Data records:
ZAGI Retail Company sales-analysis data warehouse

CALENDAR Dimension

CalendarKey	Full Date	DayOf Week	DayOf Month	Month	Qtr	Year
1	1/1/2013	Tuesday	1	January	Q1	2013
2	1/2/2013	Wednesday	2	January	Q1	2013

PRODUCT Dimension

ProductKey	ProductID	Product Name	Product Price	Product Vendor Name	Product Category Name
1	1X1	Zzz Bag	$100	Pacifica Gear	Camping
2	2X2	Easy Boot	$70	Mountain King	Footwear
3	3X3	Cosy Sock	$15	Mountain King	Footwear
4	4X4	Dura Boot	$90	Pacifica Gear	Footwear
5	5X5	Tiny Tent	$150	Mountain King	Camping
6	6X6	Biggy Tent	$250	Mountain King	Camping

STORE Dimension

StoreKey	StoreID	StoreZip	StoreRegion Name	Store Size (m^2)	Store Csystem	Store Layout
1	S1	60600	Chicagoland	51000	Cashiers	Modern
2	S2	60605	Chicagoland	35000	Self Service	Traditional
3	S3	35400	Tristate	55000	Mixed	Traditional

CUSTOMER Dimension

CustomerKey	CustomerID	Customer Name	Customer Zip	Customer Gender	Customer MaritalStatus	Customer EducationLevel	Customer CreditScore
1	1-2-333	Tina	60137	Female	Single	College	700
2	2-3-444	Tony	60611	Male	Single	High School	650
3	3-4-555	Pam	35401	Female	Married	College	623

SALES Fact table

CalendarKey	StoreKey	ProductKey	CustomerKey	TID	TimeOfDay	DollarsSold	UnitsSold
1	1	1	1	T111	8:23:59 AM	$100	1
1	2	2	2	T222	8:24:30 AM	$70	1
2	3	3	1	T333	8:15:08 AM	$75	5
2	3	1	1	T333	8:15:08 AM	$100	1
2	3	4	3	T444	8:20:33 AM	$90	1
2	3	2	3	T444	8:20:33 AM	$140	2
2	3	4	2	T555	8:30:00 AM	$360	4
2	3	5	2	T555	8:30:00 AM	$300	2
2	3	6	2	T555	8:30:00 AM	$250	1

Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Pages 260-262.

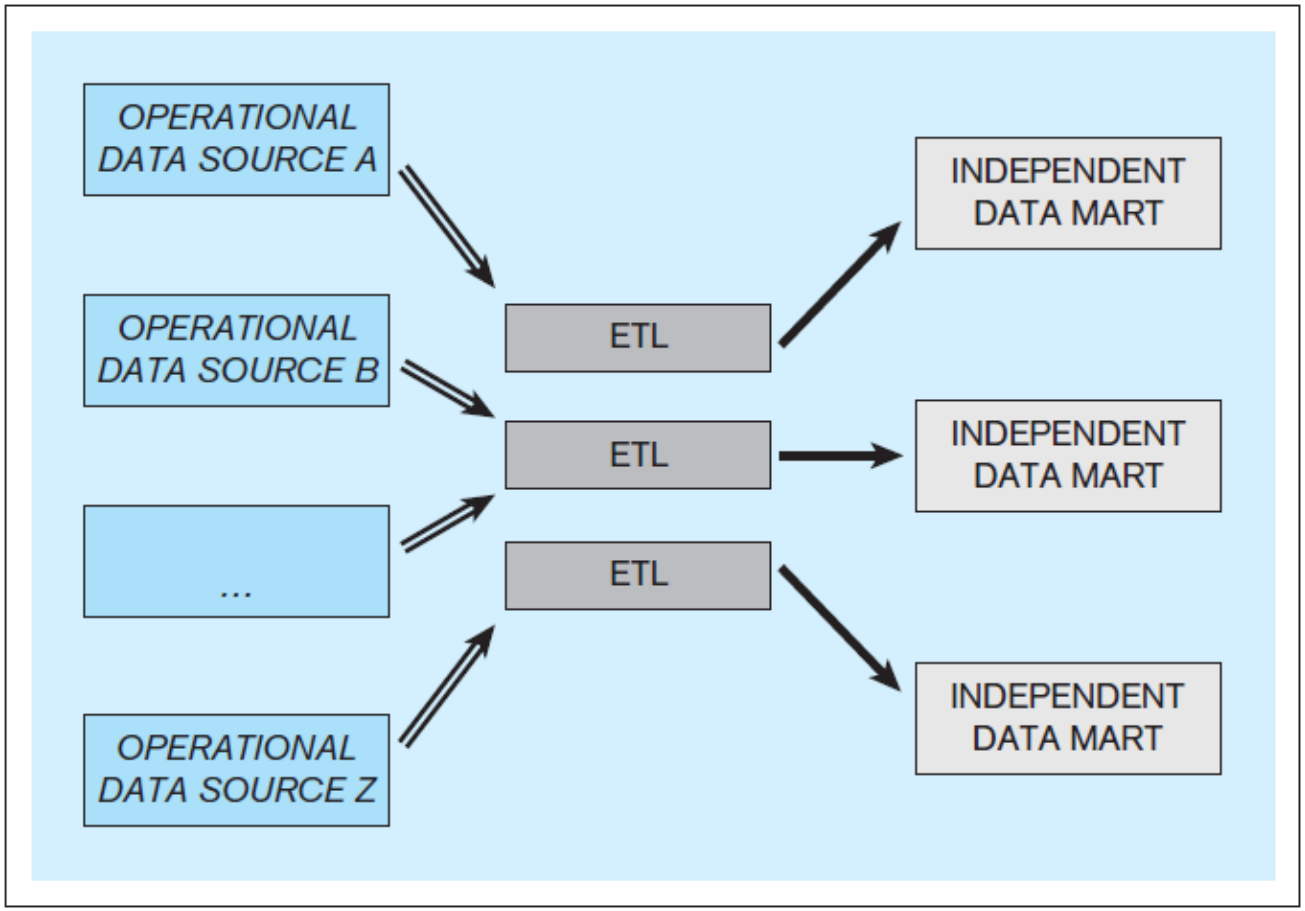
DATA WAREHOUSE (DATA MART) MODELING APPROACHES

- Independent data marts
  - Stand-alone data marts are created by various groups within the organization, independent of other stand-alone data marts in the organization
  - Consequently, multiple ETL systems are created and maintained

Copyright (c) 2014 Pearson Education, Inc.

DATA WAREHOUSE (DATA MART) MODELING APPROACHES

- Independent data marts



Copyright (c) 2014 Pearson Education, Inc.

Troy University Resources

DATA WAREHOUSE (DATA MART) MODELING APPROACHES

- Independent data marts
  - Independent data marts are considered an inferior strategy
    - Inability for straightforward analysis across the enterprise
    - The existence of multiple unrelated ETL infrastructures
  - In spite of obvious disadvantages, a significant number of corporate analytical data stores are developed as a collection of independent data marts

Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This strategy does not result in a data warehouse, but in a collection of unrelated independent data marts. While the independent data marts within one organization may end up as a whole, containing all the necessary analytical information, such information is scattered and difficult or even impossible to analyze as one unit. A discussion on why a significant number of corporate analytical data stores are developed as a collection of independent data marts is given on Pages 263-264.

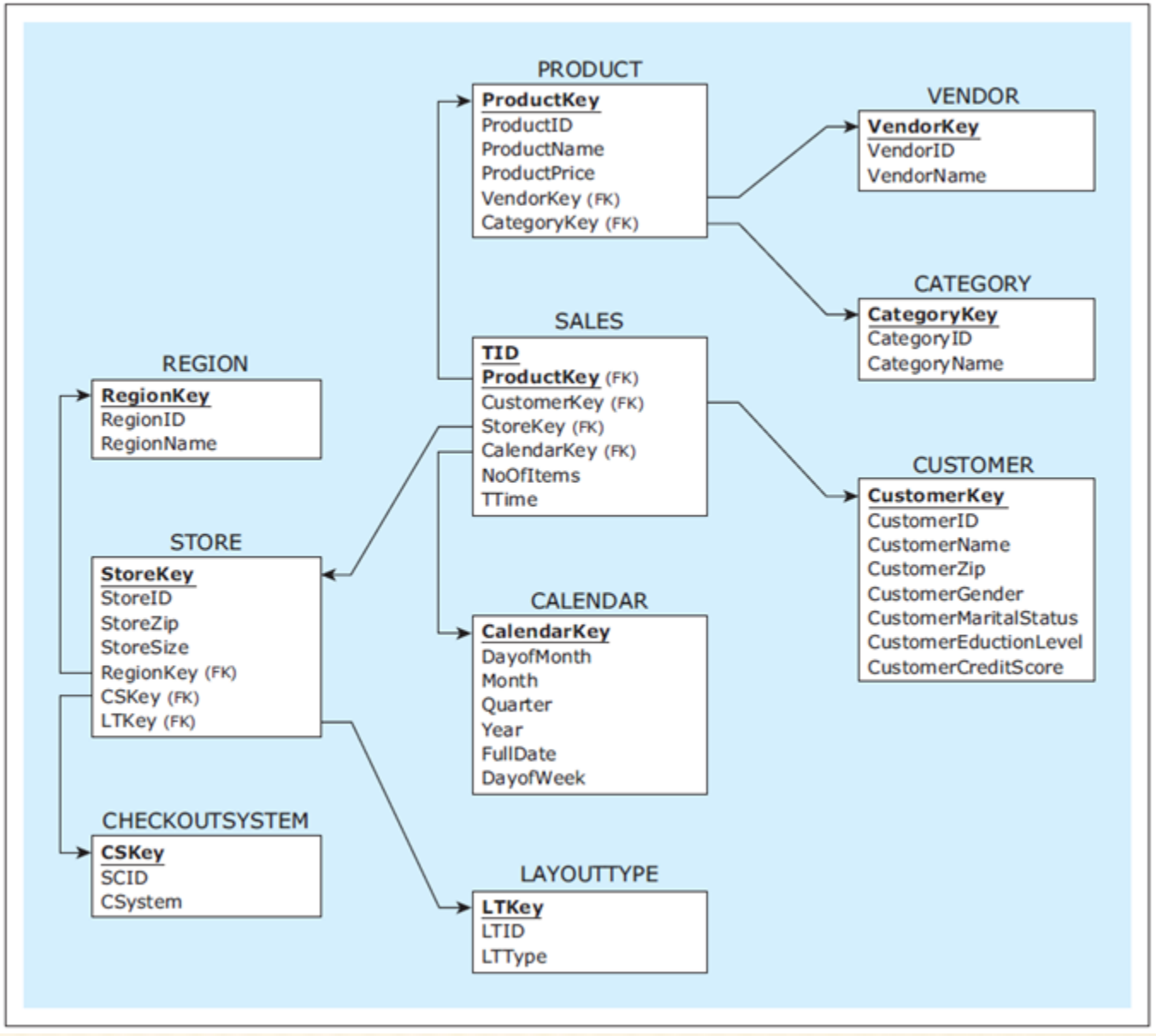
DATA WAREHOUSE (DATA MART) MODELING APPROACHES

- **Comparing dimensional modeling and ER modeling as data warehouse/data mart design techniques**
  - ER modeling can be used as a *conceptual* data warehouse/data mart design technique, followed by relational modeling as *logical* data warehouse/data mart design technique
  - Dimensional modeling can be used both of *conceptual* data warehouse/data mart design and *logical* data warehouse/data mart design

Copyright (c) 2014 Pearson Education, Inc.

Troy University Resources

Example - A modified normalized data warehouse schema



Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

This example is discussed on Pages 264-265.

DATA WAREHOUSE (DATA MART) MODELING APPROACHES

- Comparing dimensional modeling and ER modeling as data warehouse/data mart design techniques
  - Both ER modeling and dimensional modeling are viable alternatives for modeling data warehouses/data marts, and can be used within the same project

Copyright (c) 2014 Pearson Education, Inc.

SPEAKER NOTES

For example, dimensional modeling can be used during the requirements collection process for collecting, refining, and visualizing initial requirements. Based on the resulting requirements visualized as a collection of facts and dimensions, an ER model for a normalized physical data warehouse can be created if there is a preference for a normalized data warehouse. Once a normalized data warehouse is created, a series of dependent data marts can be created using dimensional modeling.