Lecture 3: Data Warehousing

Dr Anesu Nyabadza

INTRODUCTION

A typical organization maintains and utilizes a number of operational data sources.

- The operational data sources include the databases and other data repositories which are used to support the organization's day-to-day operations

A data warehouse is created within an organization as a separate data store whose primary purpose is data analysis.

Two main reasons for the creation of a data warehouse as a separate analytical database

- The performance of operational day-to-day tasks involving data use can be severely diminished if such tasks have to compete for computing resources with analytical queries

- It is often impossible to structure a database which can be used in an efficient manner for both operational and analytical purposes

INTRODUCTION

Operational information (transactional information) - the information collected and used in support of day to day operational needs in businesses and other organizations

Analytical information - the information collected and used in support of analytical tasks

Analytical information is based on operational (transactional) information

OPERATIONAL VS. ANALYTICAL INFORMATION

Operational Data	Analytical Data
Data Makeup Differences	
Typical TimeData Warehousing-Horizon: Days/Months	Typical Time-Horizon: Years
Detailed	Summarized (and/or Detailed)
Current	Values over time (Snapshots)
Technical Differences	
Small Amounts used in a Process	Large Amounts used in a Process
High frequency of Access	Low/Modest frequency of Access
Can be Updated	Read (and Append) Only
Non-Redundant	Redundancy not an Issue
Functional Differences	
Used by all types of employees for tactical purposes	Used by a narrower set of users for decision making
Application Oriented	Subject Oriented

Application Oriented vs. Subject Oriented– Example

Application-oriented databases are designed to support the day-to-day operations of an organization. These databases are optimized for transaction processing and are often referred to as OLTP (Online Transaction Processing) systems.

Data Structure →Typically normalized to reduce redundancy and ensure data integrity.
Usage→Supports routine business operations such as order processing, inventory management, and customer transactions.
Data Granularity→Contains detailed, current data relevant to specific business applications.
Access Pattern→High frequency of read and write operations with small amounts of data processed per transaction.
Examples→ Customer Relationship Management (CRM) systems, Enterprise Resource Planning (ERP) systems, and point-of-sale systems.

Application Oriented vs. Subject Oriented– Example

Subject-oriented databases are designed for data analysis and decision support. These databases are optimized for querying and reporting and are often referred to as OLAP (Online Analytical Processing) systems or data warehouses.

Data Structure→ Often denormalized to simplify queries and improve read performance. Data is organized around key subjects or business areas such as sales, finance, or marketing.
Usage→Supports strategic decision-making by providing historical, summarized, and consolidated data.
Data Granularity→Contains detailed and summarized data over a long time horizon, often spanning years.
Access Pattern→ Low to moderate frequency of read operations with large amounts of data processed per query.
Examples→Data warehouses, data marts **(subset of a data warehouse focused on a particular line of business, department or subject area)**, and business intelligence systems.

Application Oriented vs. Subject Oriented– Example

An **application-oriented** database serving the Vitality Health Club Visits and Payments Application

Application Oriented" and "Subject Oriented" refer to different approaches to organizing and managing data. Understanding these concepts is crucial for designing databases that meet specific business requirements and for optimizing data retrieval and analysis.

HEALTH CLUB MEMBER

MemberID	MemberName	MemberGender	MLevelID	DateMembershipPaid
111	Joe	M	A	1/1/2013
222	Sue	F	B	1/1/2013
333	Pam	F	A	1/2/2013
...	...	...	...	...

MEMBERSHIP LEVEL

MLevelID	MLevelType	MLevelFee	MLevelDescription
A	Gold	$100	Includes the Pool Usage
B	Basic	$50	No Pool Usage

DAILY VISIT FROM NONMEMBERS

DVisitTID	DVisitLevelID	DVisitDate	DVisitorGender
11xx22	YP	1/1/2013	M
11xx23	NP	1/2/2013	M
11xx24	YP	1/2/2013	F
...	...	...	...

VISIT LEVEL

DVisitLevelID	DVisitLevelFee	DVisitLevelType
YP	$15	With Pool Usage
NP	$10	Without Pool Usage

Application Oriented vs. Subject Oriented– Example

A **subject-oriented** database for the analysis of the subject *revenue* in the Vitality Health Club

REVENUE

RevenueRecordID	Date	GeneratedBy	ClientGender	Pool Use Included in Purchase	Amount
1000	1/1/2013	Member	M	Yes	$100
1001	1/1/2013	Member	F	No	$50
1002	1/1/2013	Nonmember	M	Yes	$15
1003	1/2/2013	Member	F	Yes	$100
1004	1/2/2013	Nonmember	M	No	$10
1005	1/2/2013	Nonmember	F	Yes	$15
…				…	…

THE DATA WAREHOUSE DEFINITION

The data warehouse is a **structured repository** of **integrated**, **subject-oriented**, **enterprise-wide**, **historical**, and **time-variant** data.

The purpose of the data warehouse is the **retrieval of analytical information**. A data warehouse can store **detailed and/or summarized data**.

THE DATA WAREHOUSE DEFINITION

Retrieval of analytical information

A data warehouse is developed for the *retrieval of analytical information*, and it is not meant for direct data entry by the users.

The only functionality available to the users of the data warehouse is retrieval

The data in the data warehouse is not subject to changes.

The data in the data warehouse is referred to as non-volatile, static, or read-only

Detailed and/or summarized data

A data warehouse, depending on its purpose, may include the *detailed data* or *summary data* or *both*

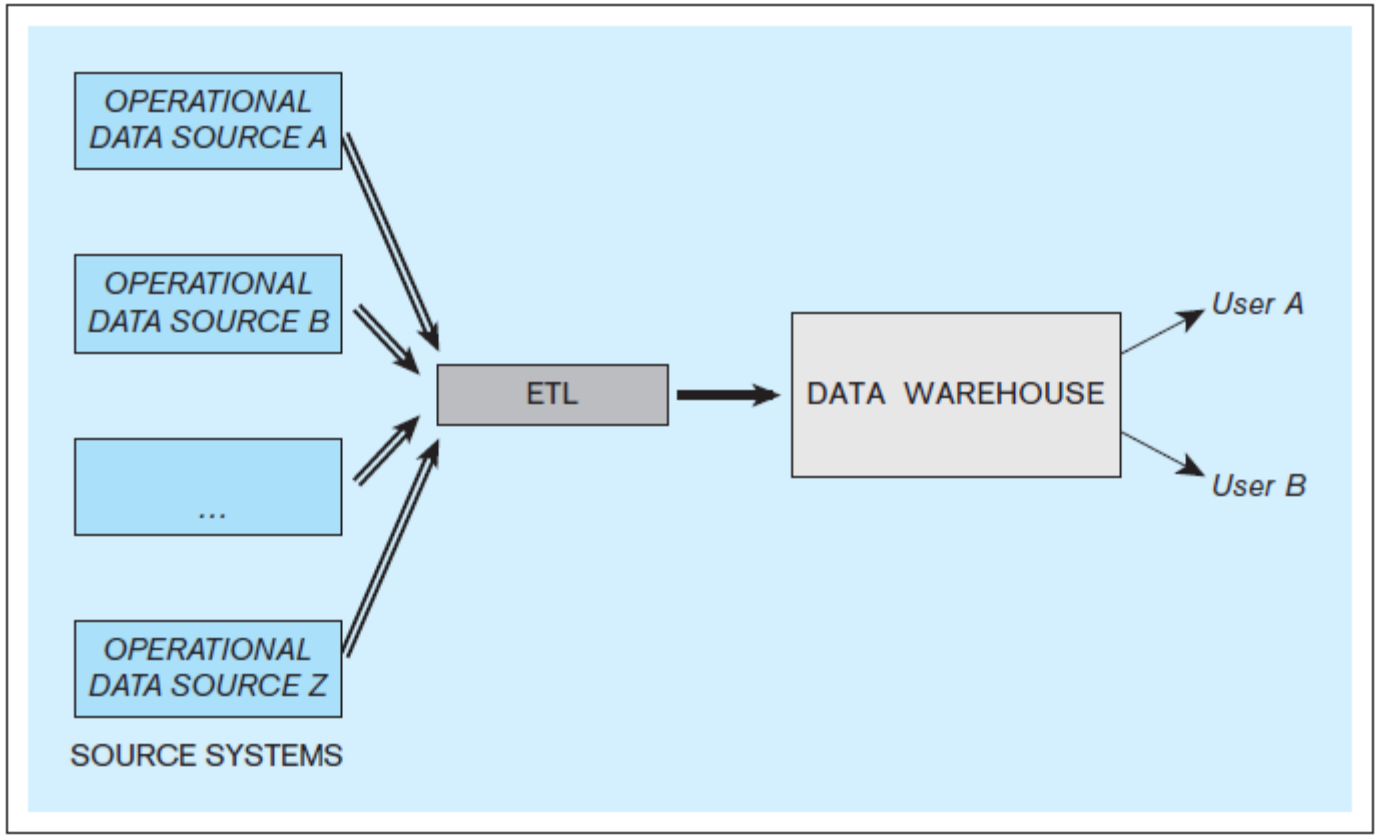
A data warehouse that contains the data at the finest level of detail is the most powerful

DATA WAREHOUSE COMPONENTS

Data warehouse components

- Source systems
- Extraction-transformation-load (ETL) infrastructure
- Data warehouse
- Front-end applications

Example - The core components of a data warehousing system



DATA WAREHOUSE COMPONENTS

Source systems

In the context of data warehousing, *source systems* are operational databases and other operational data repositories (in other words, any sets of data used for operational purposes) that provide analytically useful information for the data warehouse's subjects of analysis.

Every operational data store that is used as a source system for the data warehouse has two purposes:

The original operational purpose

As a source system for the data warehouse

Source systems can include *external data sources*

External data sources examples include market research data, census data, stock market data, weather data.

DATA WAREHOUSE COMPONENTS

Data warehouse

The *data warehouse* is sometimes referred to as the *target system*, to indicate the fact that it is a destination for the data from the source systems

A typical data warehouse periodically retrieves selected analytically useful data from the operational data sources

DATA WAREHOUSE COMPONENTS

ETL infrastructure

The infrastructure that facilitates the retrieval of data from operational databases into the data warehouses

ETL includes the following tasks:

Extracting analytically useful data from the operational data sources

Transforming such data so that it conforms to the structure of the subject-oriented target data warehouse model (while ensuring the quality of the transformed data)

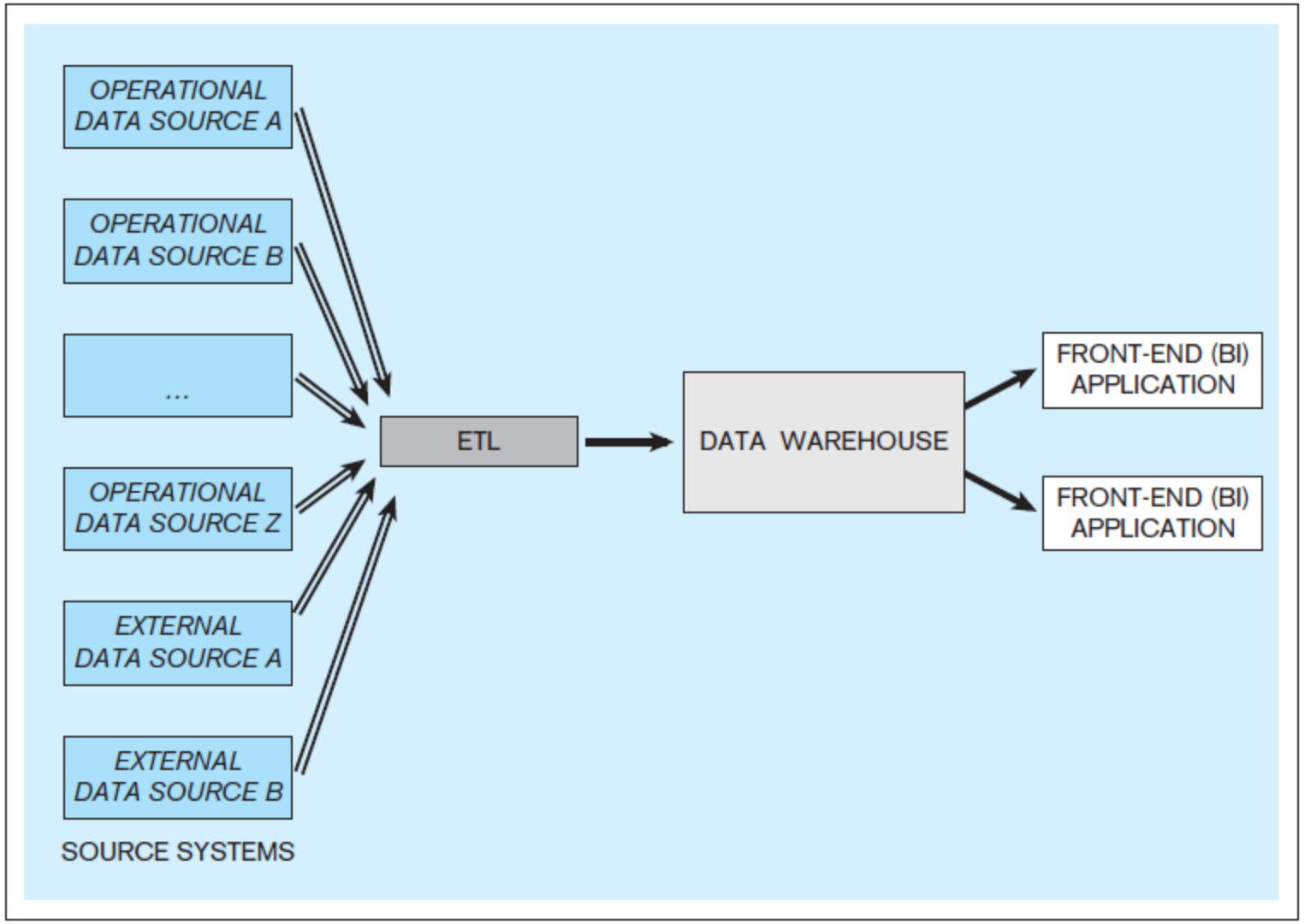
Loading the transformed and quality-assured data into the target data warehouse

- Due to the amount of details that have to be considered, creating ETL infrastructure is often the most time- and resource-consuming part of the data warehouse development process

Data warehouse front-end (BI) applications

Used to provide access to the data warehouse for users who are engaging in indirect use

Example - A data warehouse with front-end applications



DATA MARTS

Data mart

A data store based on the same principles as a data warehouse, but with a more limited scope

	DATA WAREHOUSE	DATA MART
Subjects	*Multiple*	*Single*
Data Sources	*Many*	*Fewer*
Typical Size	*Very big (routinely terabytes of data and larger)*	*Not as big*
Implementation Time	*Relatively long (months, years)*	*Not as long*
Focus	*Organization-wide*	*Often narrower than organization-wide*

DATA MARTS

Independent data mart

Stand-alone data mart, created in the same fashion as the data warehouse

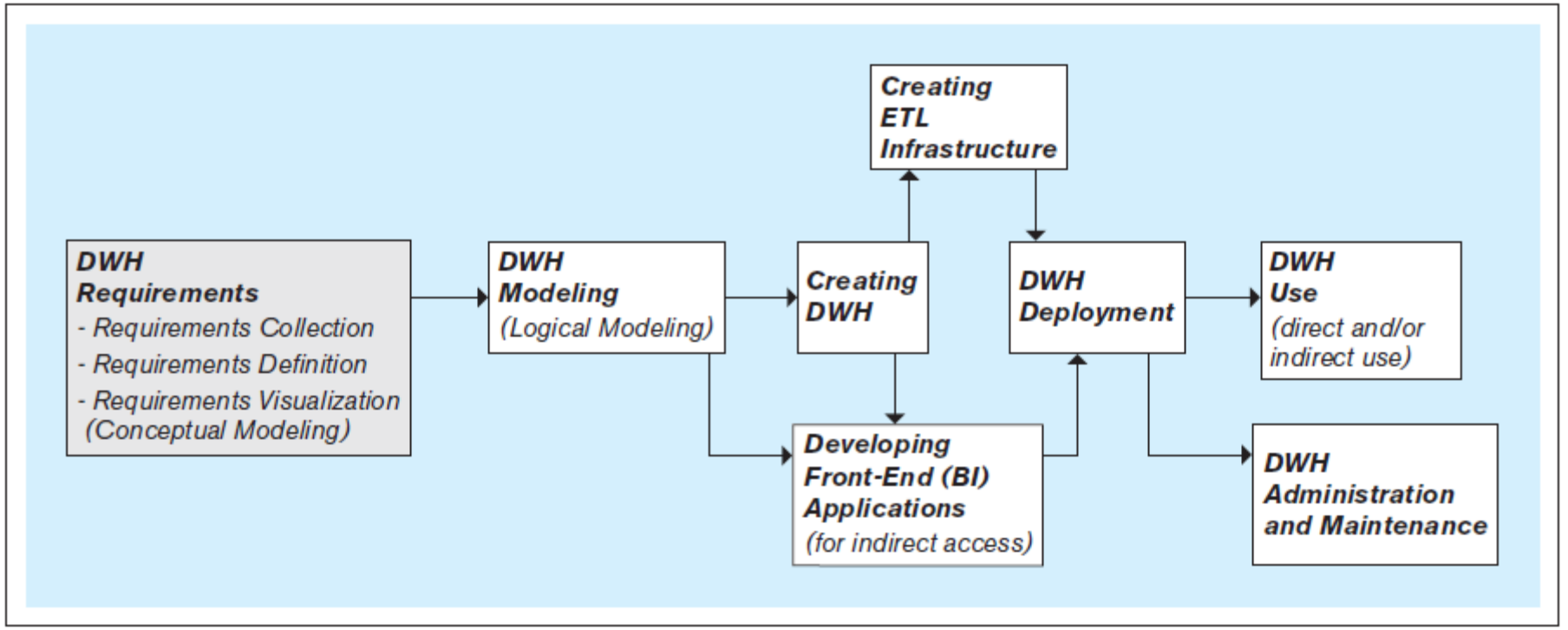
Independent data mart has its own source systems and ETL infrastructure

Dependent data mart

Does not have its own source systems

The data comes from the data warehouse

STEPS IN THE DEVELOPMENT OF DATA WAREHOUSES



STEPS IN THE DEVELOPMENT OF DATA WAREHOUSES

Data Warehouse (DWH) Development Process

DWH Requirements

- **Requirements Collection**: Gather necessary information and specifications from stakeholders.
- **Requirements Definition**: Clearly define what the data warehouse needs to achieve.
- **Requirements Visualization (Conceptual Modeling)**: Create visual representations of the data requirements to aid understanding and communication.

DWH Modeling

- **Logical Modeling**: Develop logical models to represent the structure and relationships within the data warehouse, ensuring alignment with requirements.

Creating ETL Infrastructure

- **ETL (Extract, Transform, Load)**: Build the infrastructure to extract data from source systems, transform it to fit the data warehouse schema, and load it into the data warehouse.

Creating DWH

- **Development**: Construct the data warehouse based on the logical models and ETL processes.

STEPS IN THE DEVELOPMENT OF DATA WAREHOUSES

Developing Front-End (BI) ApplicationsFront-End Development: Create business intelligence applications for users to access and analyze the data indirectly.

DWH DeploymentDeployment: Implement the data warehouse into the operational environment, making it available for use.DWH UseDirect and/or Indirect Use: Enable users to access the data warehouse either directly (for advanced users) or through BI applications (for indirect access).

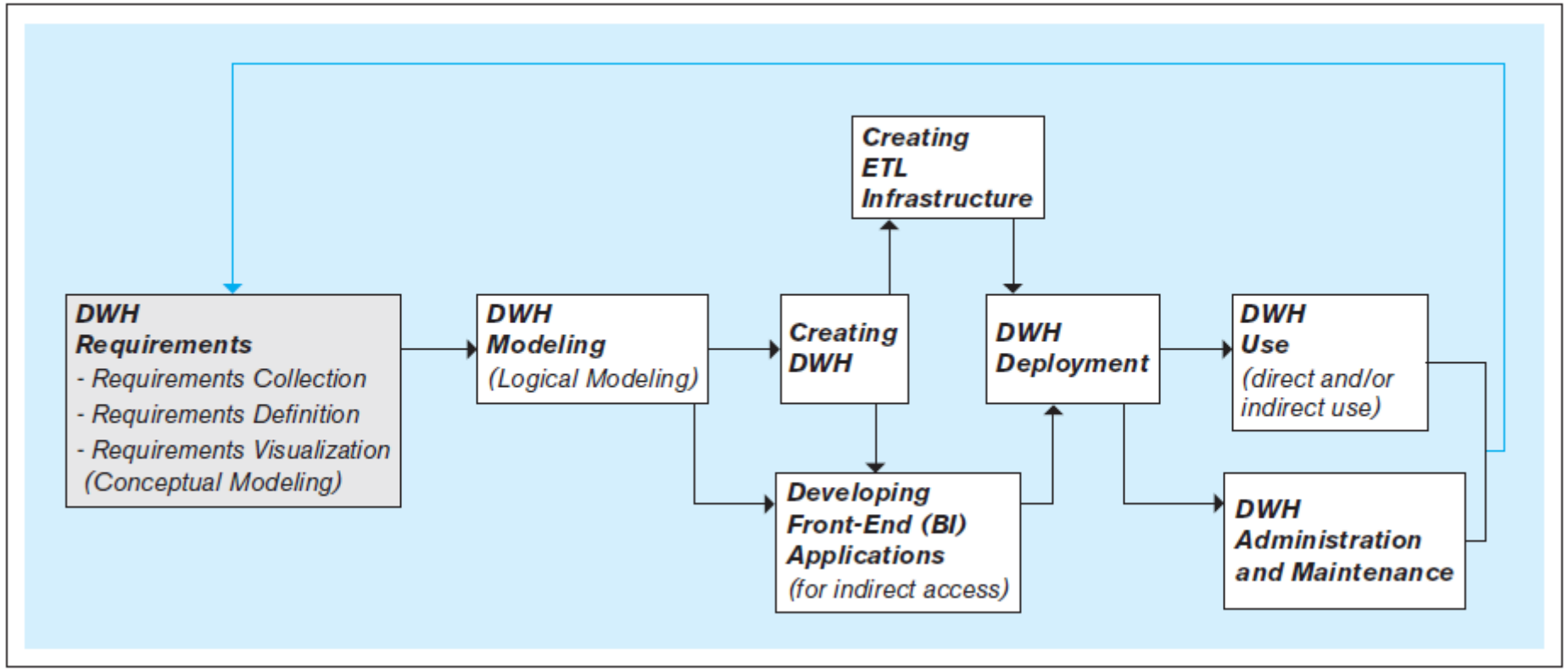
DWH Administration and Maintenance

•**Administration**: Oversee the day-to-day operations, ensuring the data warehouse runs smoothly.

•**Maintenance**: Regularly update and optimize the data warehouse to accommodate new requirements and improve performance.

This structured approach ensures that all aspects of data warehousing are covered, from initial requirements gathering to ongoing maintenance and user access.

THE NEXT VERSION OF THE DATA WAREHOUSE



Data Warehouse Modelling

Data Warehouse Modelling

ER modeling (Entity relationship)

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

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

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

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

Each row represents a single fact, offering fine granularity.

Aggregated Fact Tables summarize multiple facts, providing coarser granularity but faster queries.

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*

A central fact table linked to multiple dimension tables, resembling a star.

Simplifies queries and improves performance for analytical tasks.

DIMENSIONAL MODELING

Consider a retail company tracking sales data:

- **Fact Table**: Stores sales transactions, including measures like quantity sold, total sales amount.
- **Dimension Tables**: Include Product (with attributes like product name, category), Customer (with attributes like customer name, region), and Time (with attributes like day, month, year).

Star Schema Example:

1. **Fact Table**: Sales
    1. **Attributes**: SaleID, ProductID, CustomerID, TimeID, Quantity, SalesAmount
2. **Dimension Tables**:
    1. **Product Dimension**: ProductID, ProductName, Category
    2. **Customer Dimension**: CustomerID, CustomerName, Region
    3. **Time Dimension**: TimeID, Day, Month, Year

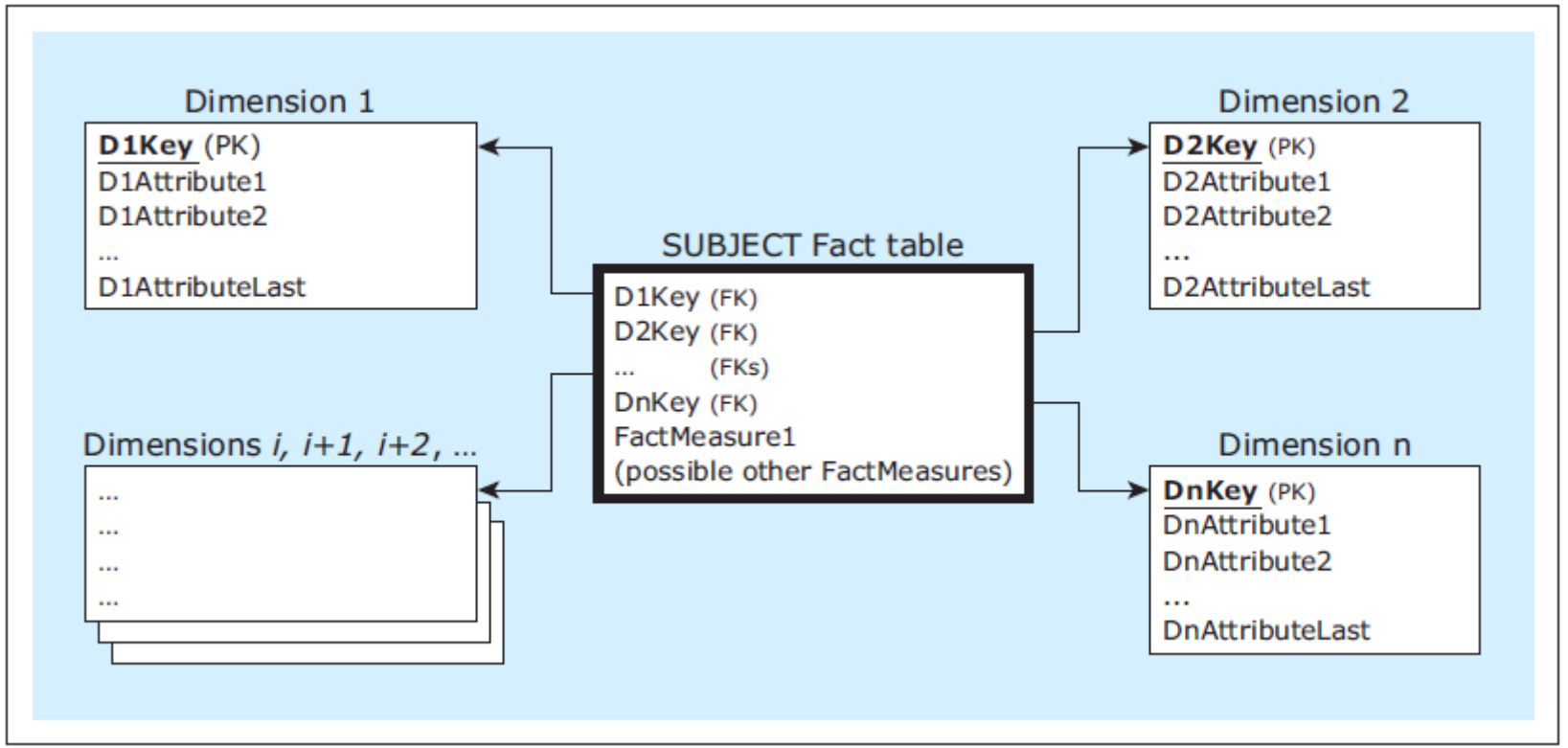
DIMENSIONAL MODELING

To find total sales by product category in SQL for a specific region and time period:

```
SELECT P.Category, SUM(S.SalesAmount)
FROM Sales S JOIN Product P ON S.ProductID = P.ProductID
JOIN Customer C ON S.CustomerID = C.CustomerID
JOIN Time T ON S.TimeID = T.TimeID
WHERE C.Region = 'Tristate' AND T.Year = 2023 AND T.Month = 'January’
GROUP BY P.Category;
```

DIMENSIONAL MODELING

A dimensional model (star schema)



Fact Table→ D1Key, D2Key, …, DnKey (FKs): These foreign keys link the fact table to corresponding dimension tables.
FactMeasure1: Represents the main numerical data collected (e.g., sales amount).
.

Dimension Tables→
•Dimension N:

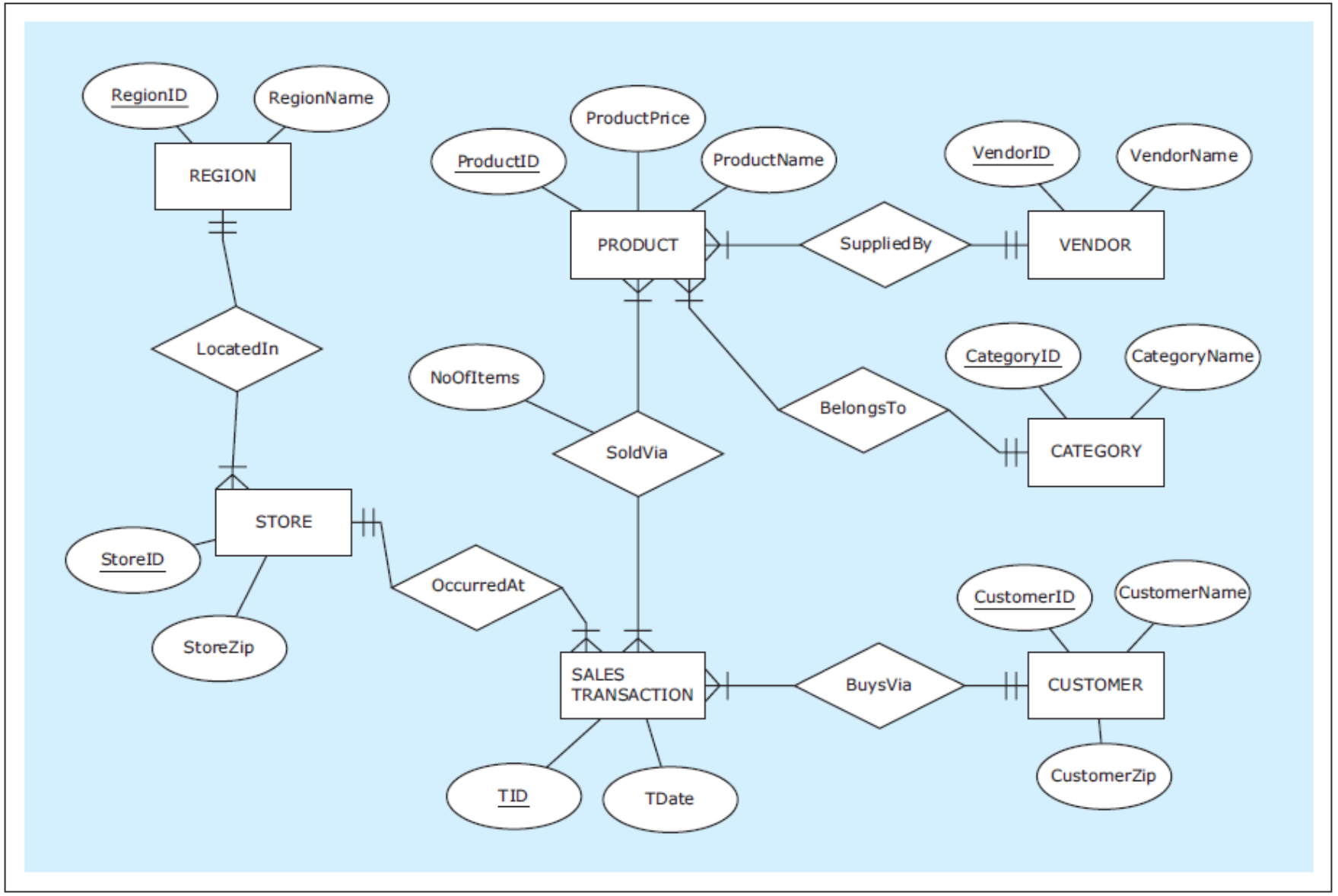
- D1Key (PK): Primary key uniquely identifying each record.
- D1Attribute1, D1Attribute2, …, D1AttributeLast: Descriptive attributes related to Dimension 1.

Consider a retail sales data warehouse example again

- **Fact Table**: Stores sales transactions.
  - **D1Key**: ProductID (FK referencing Product dimension).
  - **D2Key**: CustomerID (FK referencing Customer dimension).
  - **D3Key**: TimeID (FK referencing Time dimension).
  - **FactMeasure1**: QuantitySold.
  - **Other FactMeasures**: TotalSalesAmount.
- **Dimension Tables**:
  - **Product Dimension (Dimension 1)**:
    - **ProductID (D1Key)**: Unique identifier for each product.
    - **ProductName, Category, Brand**: Attributes describing the product.
  - **Customer Dimension (Dimension 2)**:
    - **CustomerID (D2Key)**: Unique identifier for each customer.
    - **CustomerName, Region, AgeGroup**: Attributes describing the customer.
  - **Time Dimension (Dimension n)**:
    - **TimeID (DnKey)**: Unique identifier for each time period.
    - **Date, Month, Year, Quarter**: Attributes describing the time period.

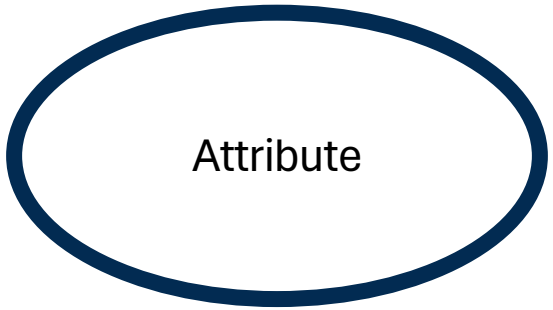
Initial Example: Dimensional Model Based on A Single Source

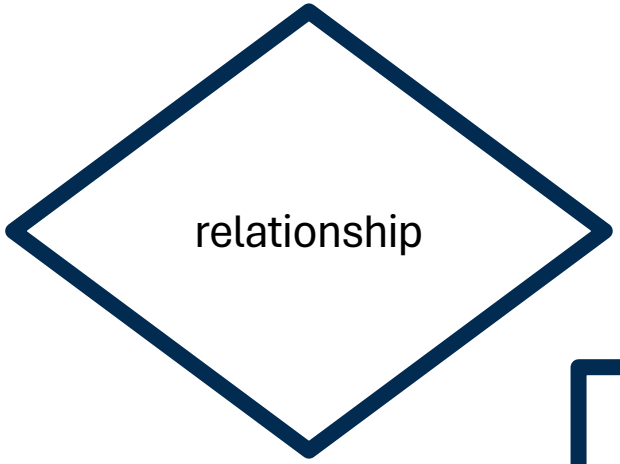
ER diagram : ZAGI Retail Company Sales Department Database (Source)

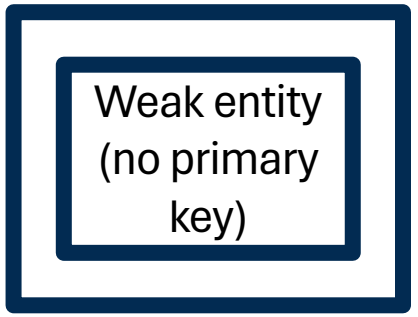


Entity-relationship (ER) diagrams- Components

Entity diagrams, also known as Entity-Relationship (ER) diagrams, are visual representations used in software development to show the structure of a database. They represent entities (things or concepts about which data is stored) and their relationships.





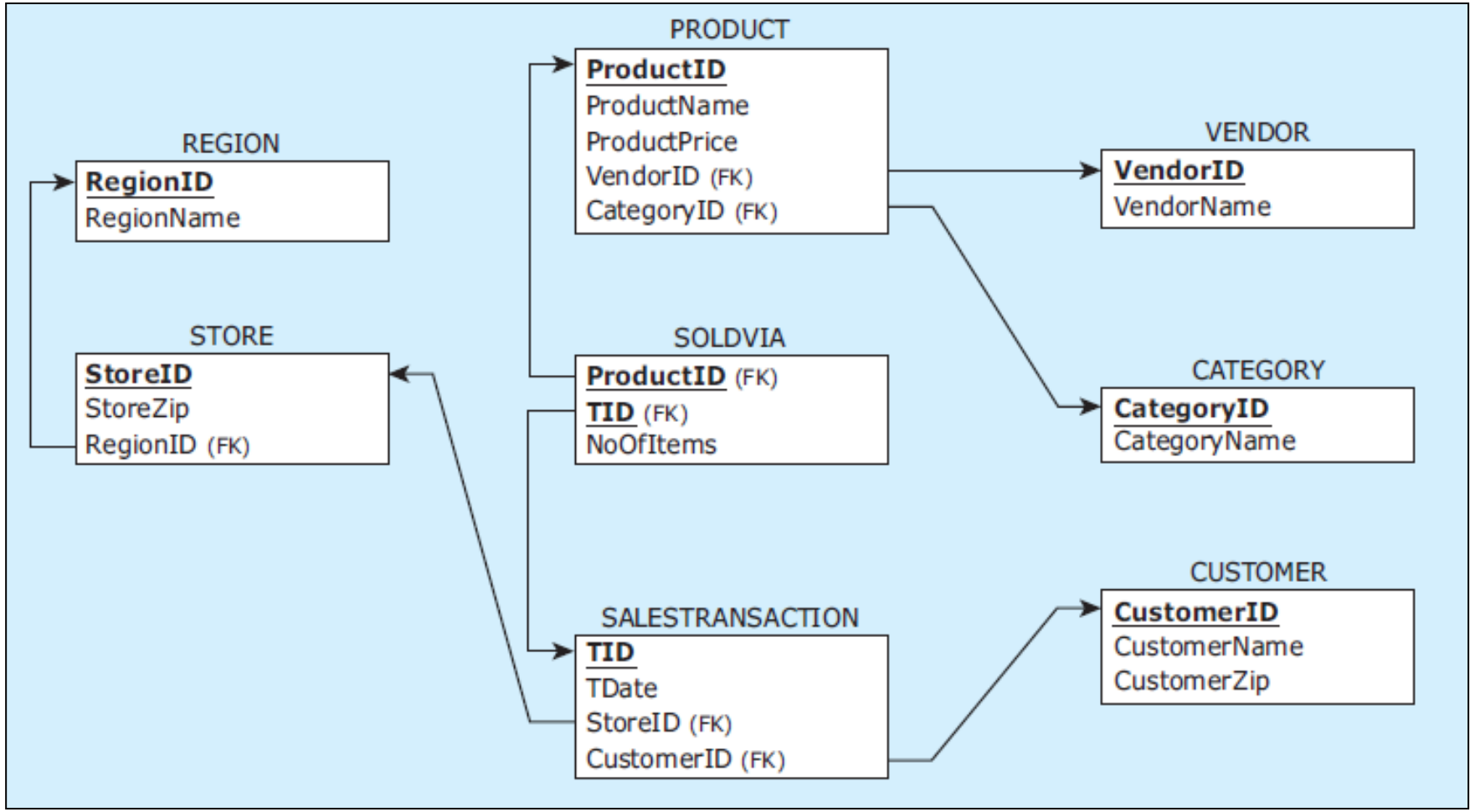


Links

Total participation

Initial Example: Dimensional Model Based on A Single Source

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



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

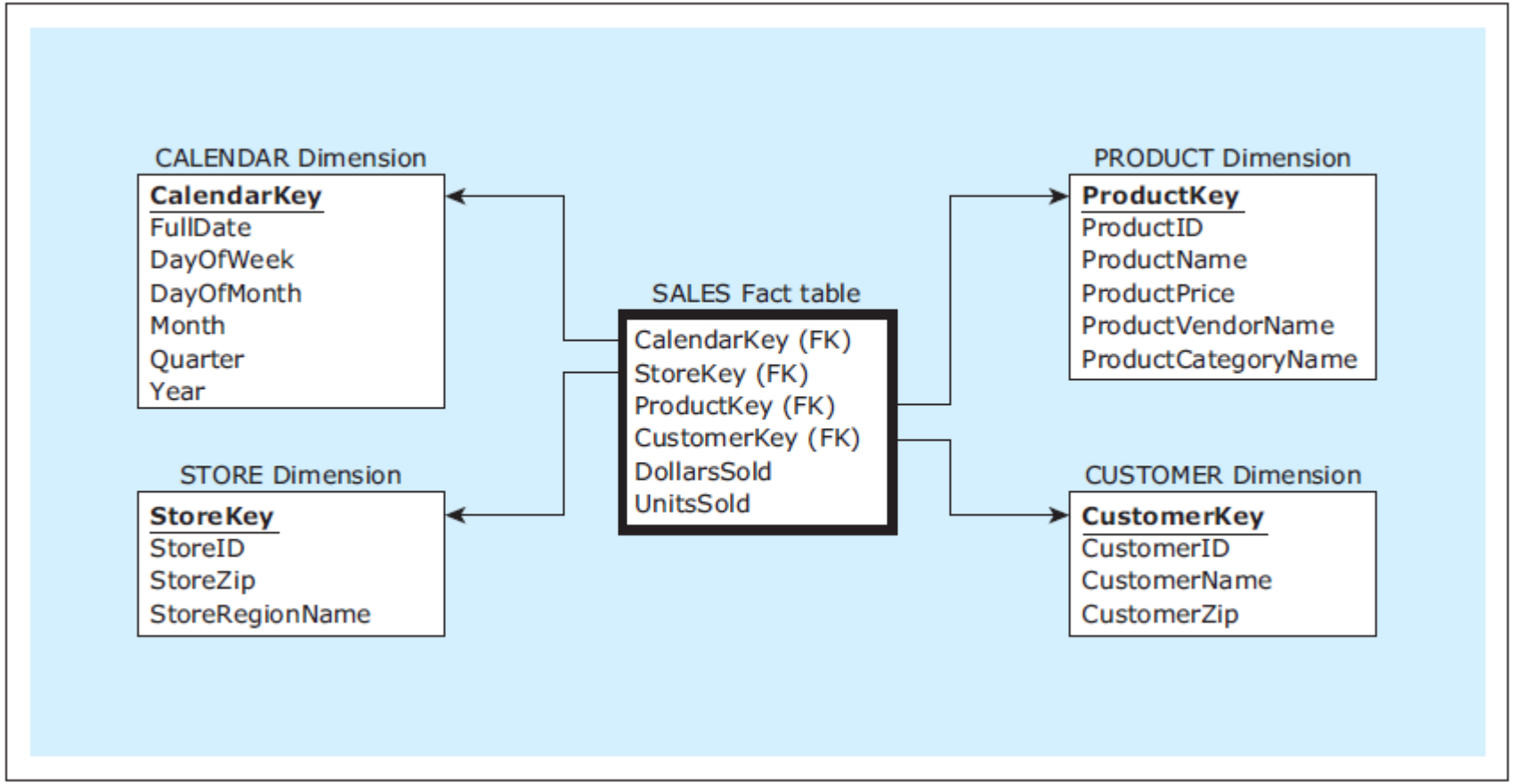
ProductID	TID	NoOfItems
1X1	T111	1
2X2	T222	1
3X3	T333	5
1X1	T333	1
4X4	T444	1
2X2	T444	2
4X4	T555	4
5X5	T555	2
6X6	T555	1

CUSTOMER

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

Initial Example: Dimensional Model Based on A Single Source

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



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?

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

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

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

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

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

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

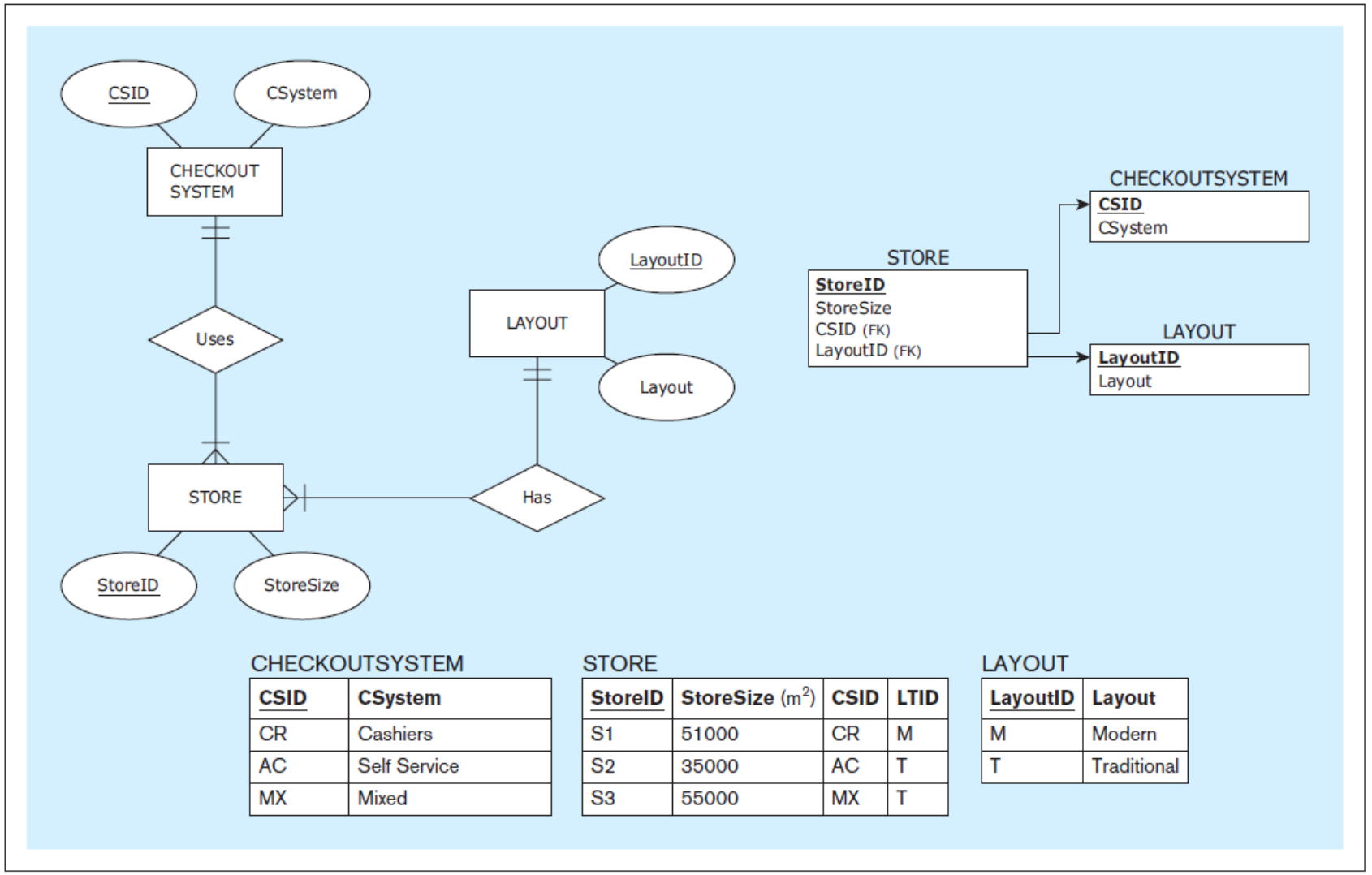
```
SELECT    SUM(SA.UnitsSold)
,         P.ProductCategoryName
,         P.ProductVendorName
,         C.DayofWeek
,         C.Qtr
FROM
          Calendar C
,         Store    S
,         Product  P
,         Sales    SA
WHERE

          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;
```

*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);
```

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

Expanded Example: Dimensional Model Based on Multiple Sources

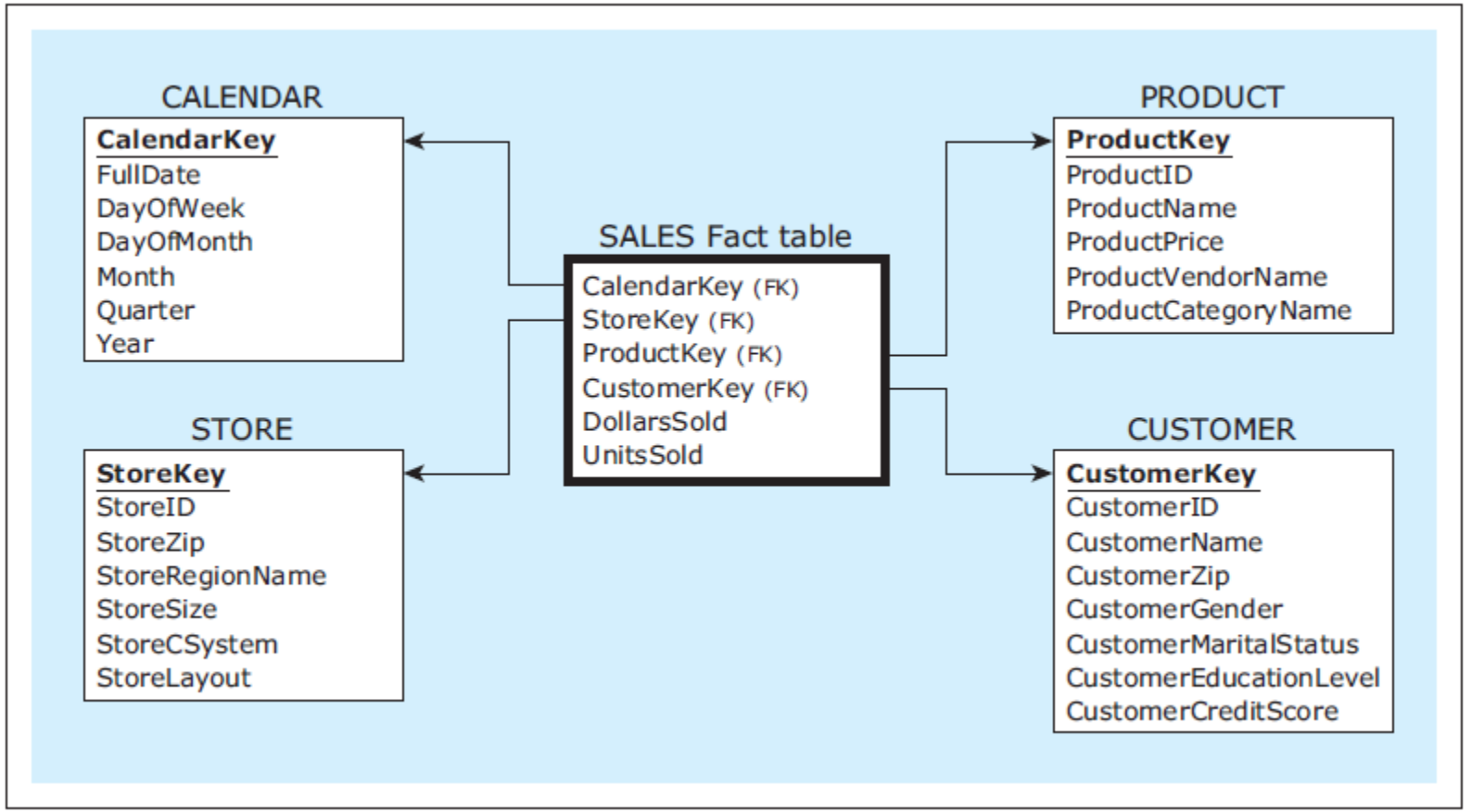
Customer Demographic Data Table - external source acquired from a market research company (Source 3)

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

Expanded Example: Dimensional Model Based on Multiple Sources

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



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

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.

```
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;
```

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;
```

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

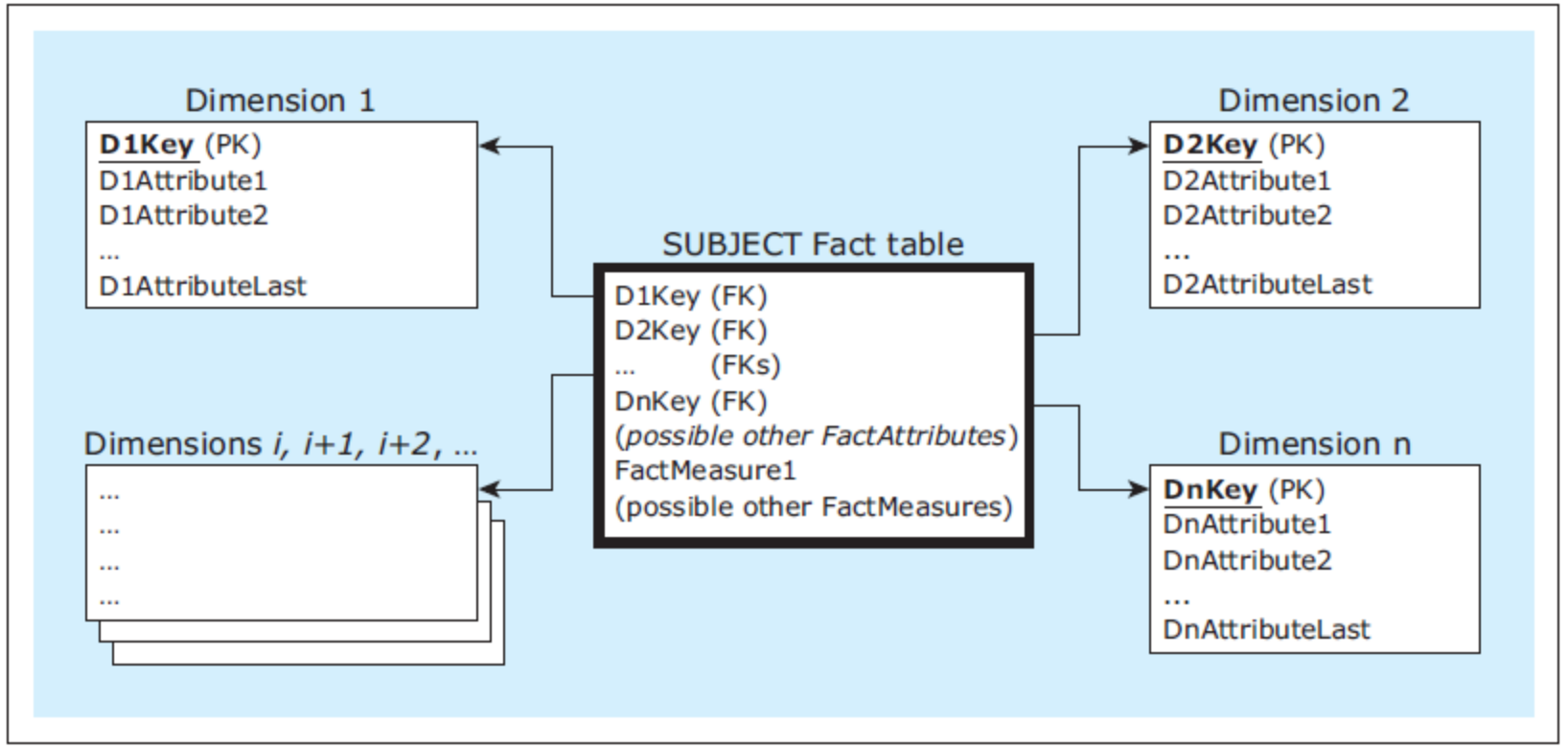
Two of the most typical additional attributes that can appear in the fact table are:

Transaction identifier

Transaction time

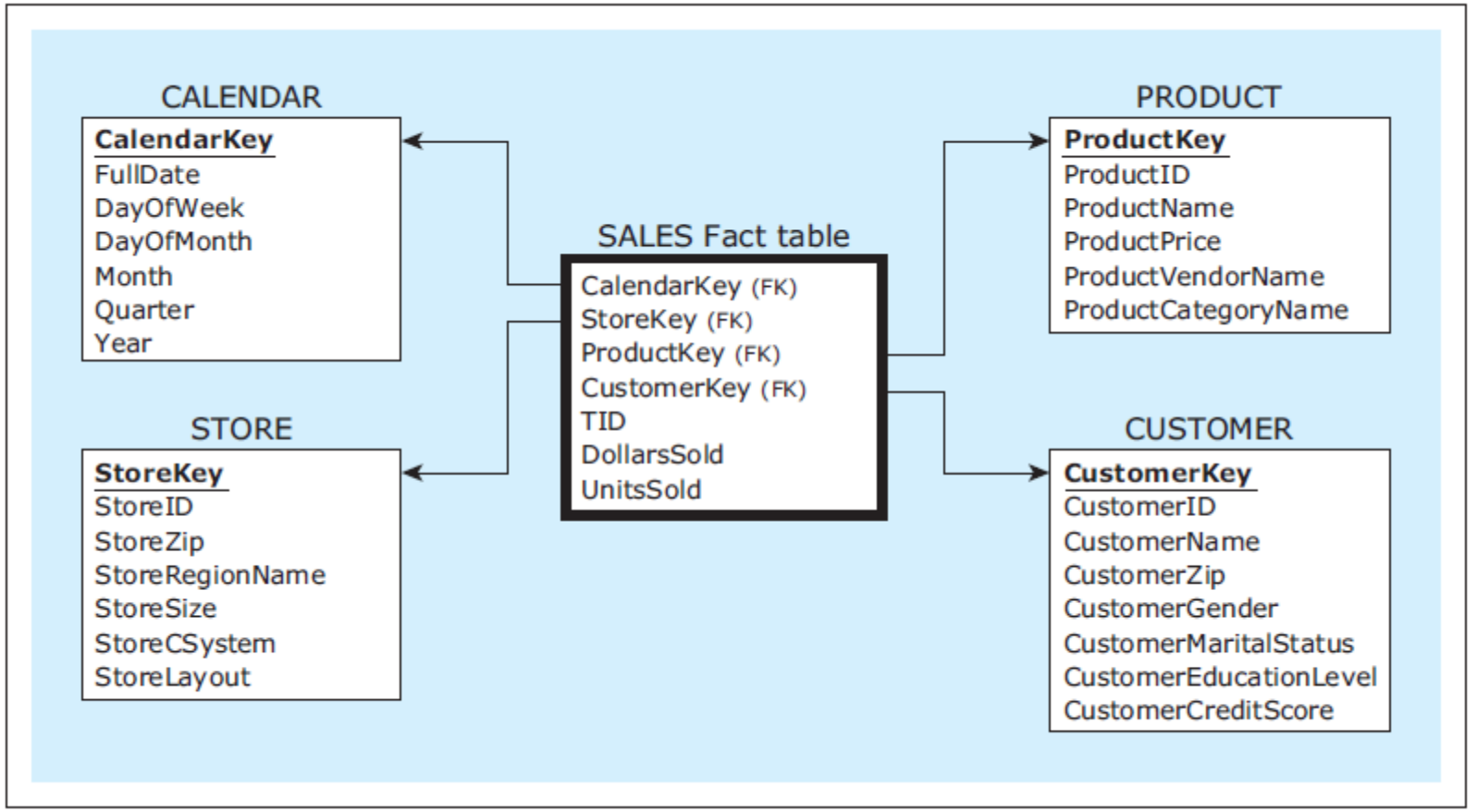
DIMENSIONAL MODELING

Additional possible fact attributes



Expanded Example: Dimensional Model Based on Multiple Sources

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



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

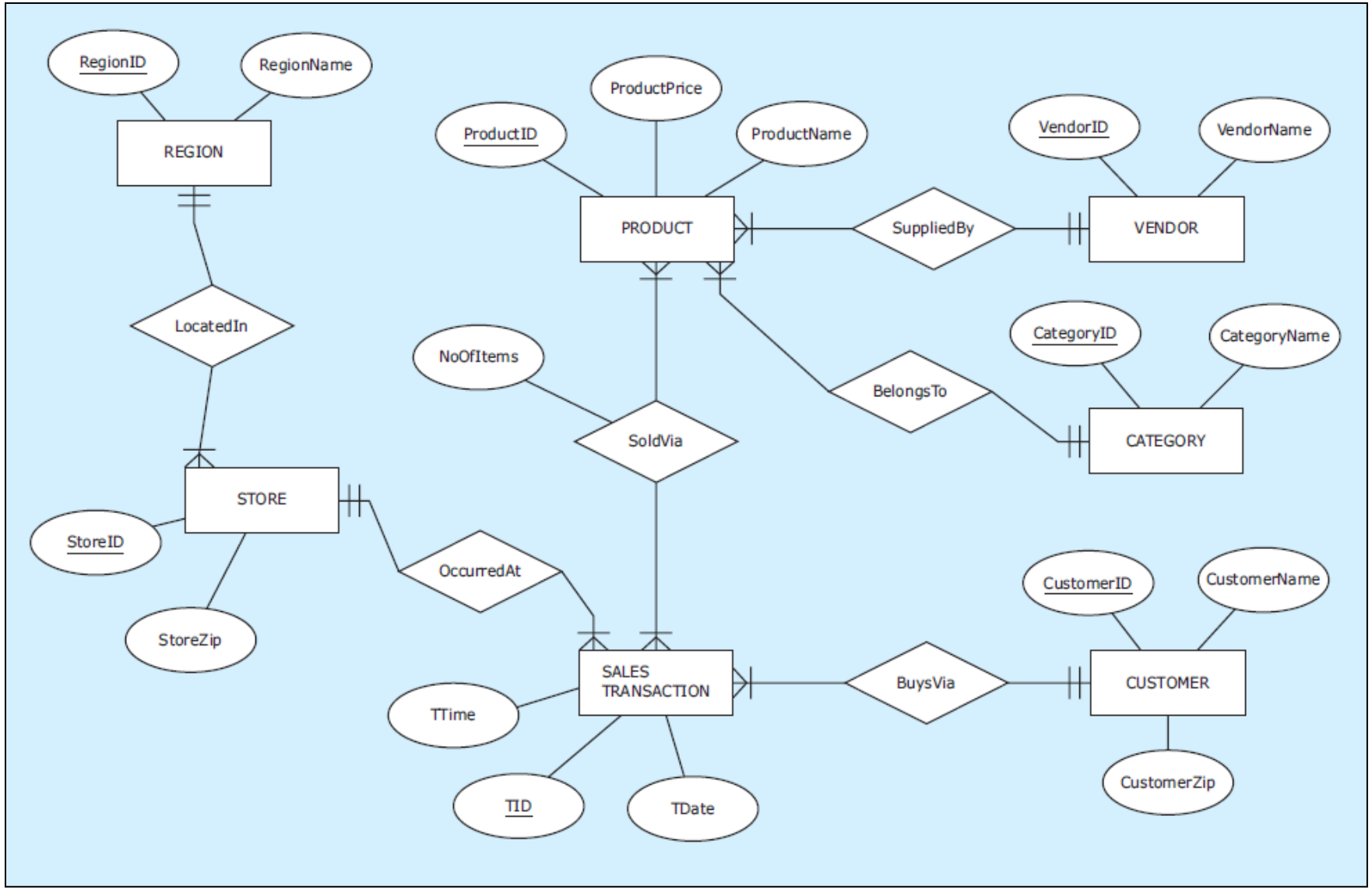
CustomerKey	CustomerID	Customer Name	Customer Zip	Customer Gender	Customer MaritalStatus	Customer EducationLevel	Customer CreditScore
1	1-2-333	Tina	60137	Female	Single	College	700
2	2-3-444	Tony	60611	Male	Single	High School	650
3	3-4-555	Pam	35401	Female	Married	College	623

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

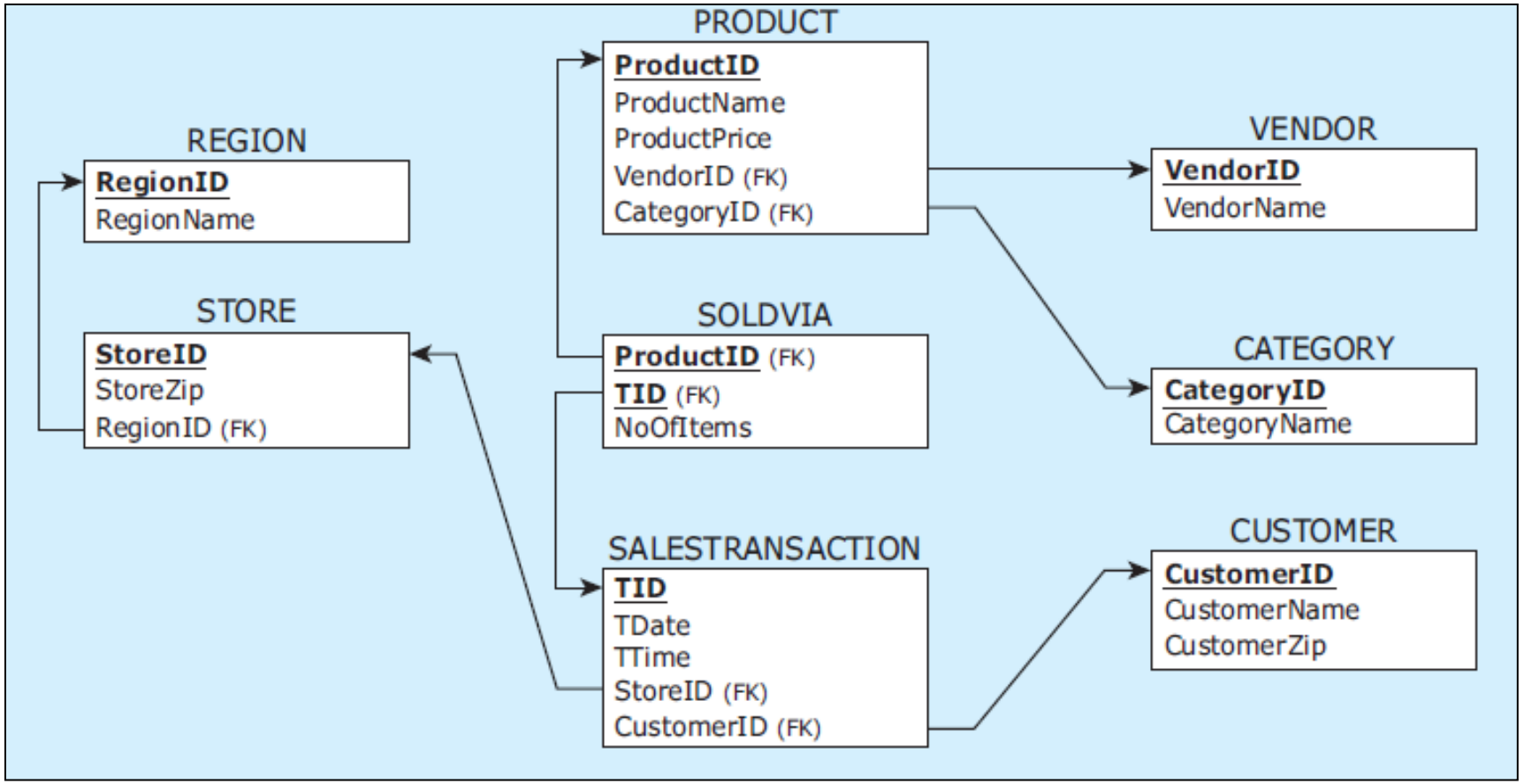
Expanded Example: Dimensional Model Based on Multiple Sources

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



Expanded Example: Dimensional Model Based on Multiple Sources

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



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

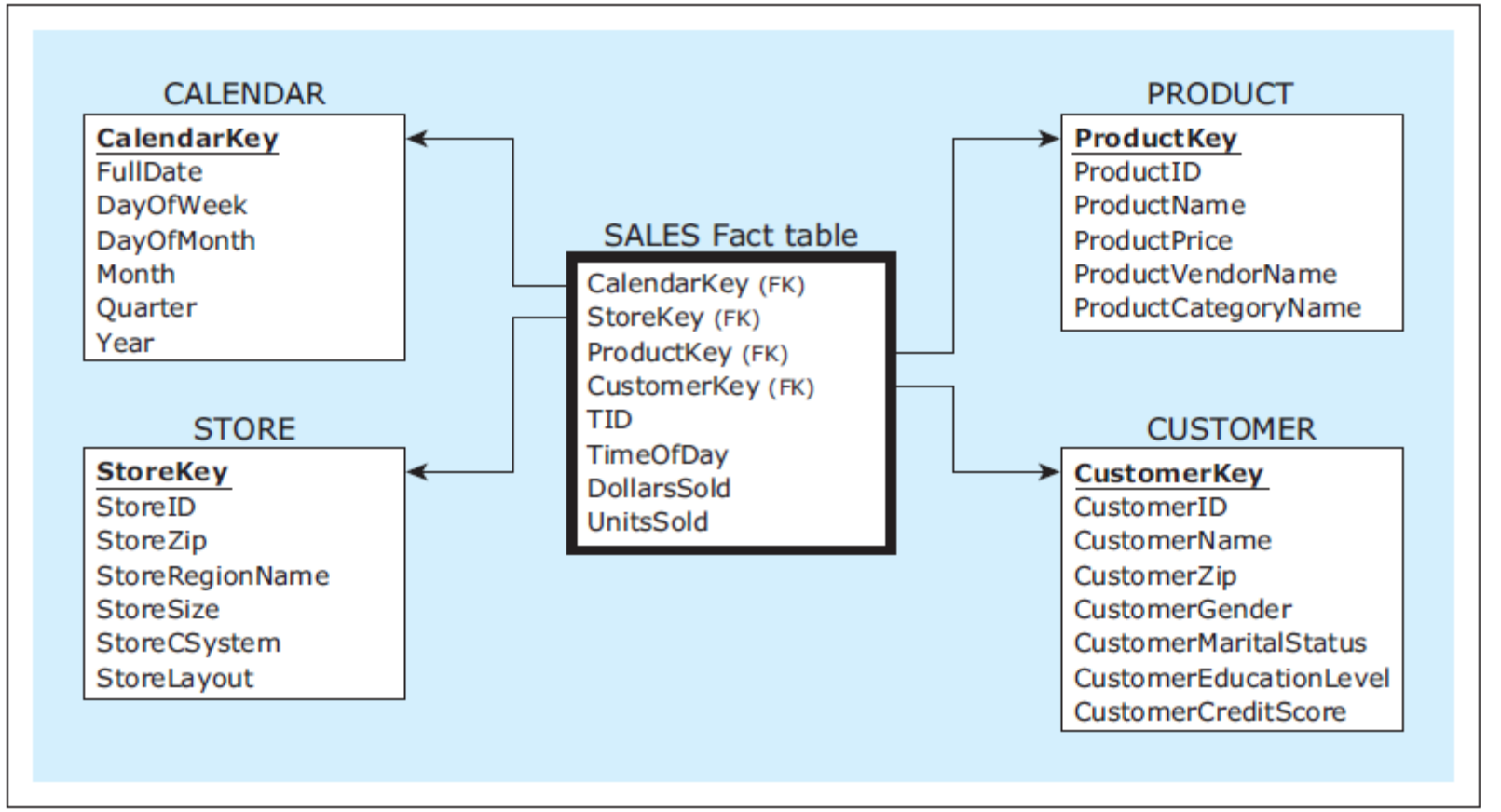
ProductID	TID	NoOfItems
1X1	T111	1
2X2	T222	1
3X3	T333	5
1X1	T333	1
4X4	T444	1
2X2	T444	2
4X4	T555	4
5X5	T555	2
6X6	T555	1

CUSTOMER

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

Expanded Example: Dimensional Model Based on Multiple Sources

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



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

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**

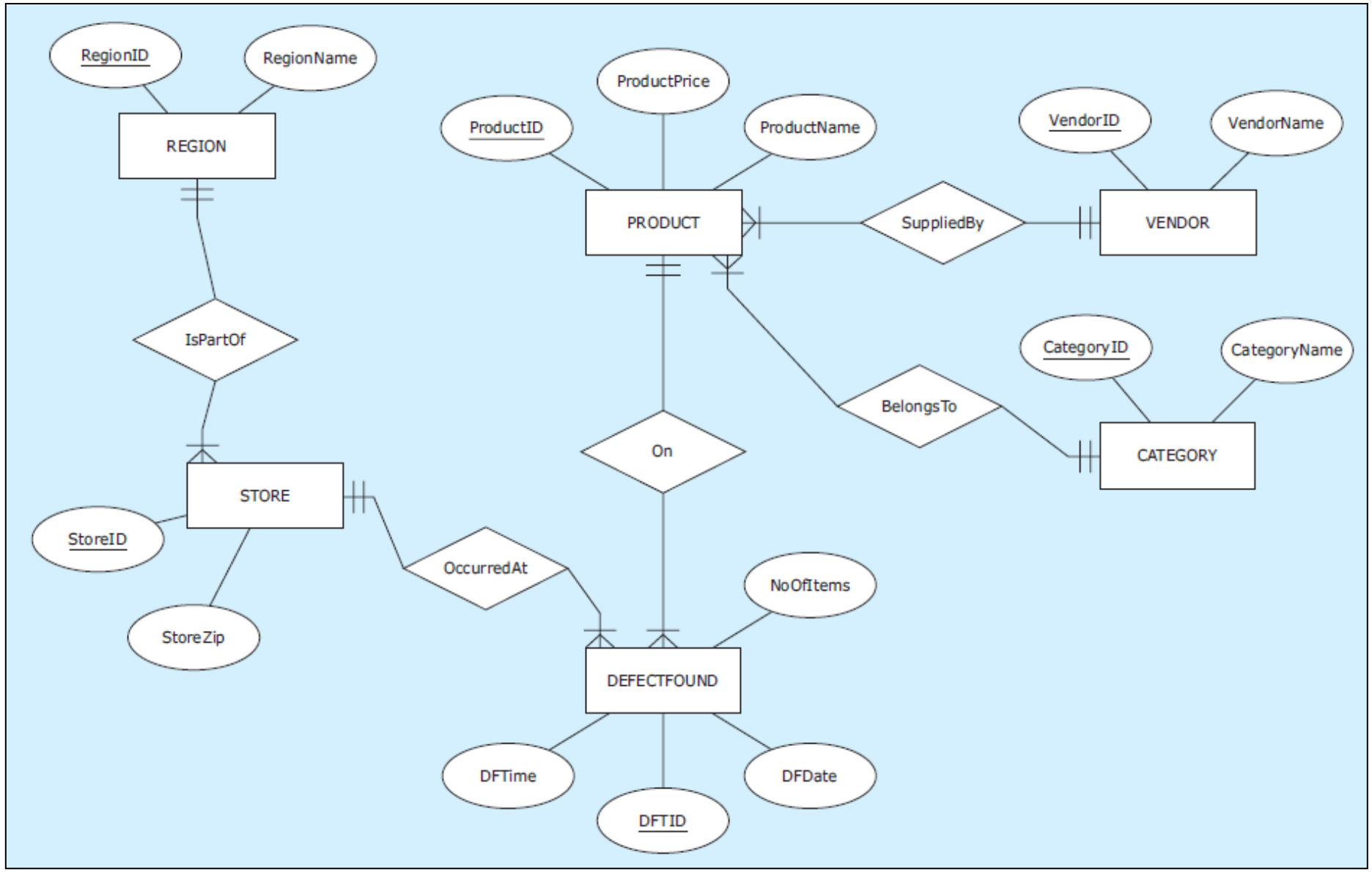
This approach enables:

Quicker development of analytical databases for multiple subjects of analysis, because dimensions are re-used instead of duplicated

Straightforward cross-fact analysis

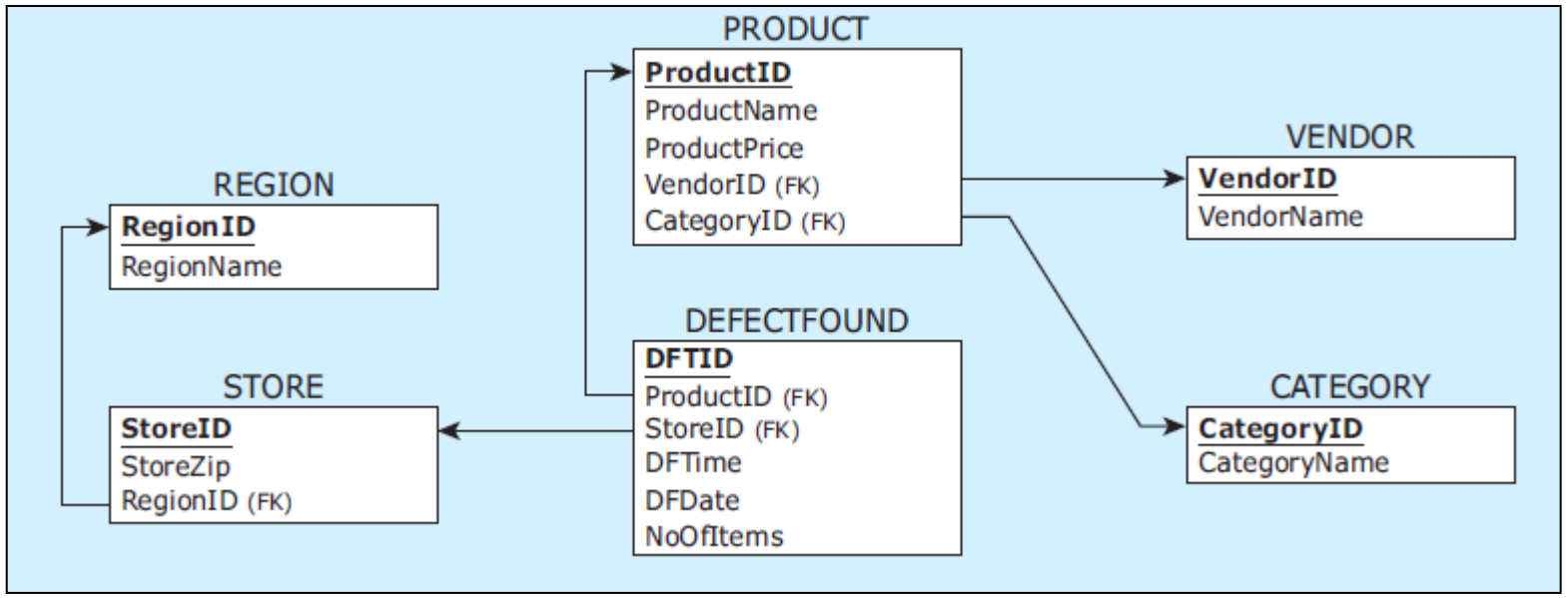
Expanded Example: Dimensional Model Based on Multiple Sources

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



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

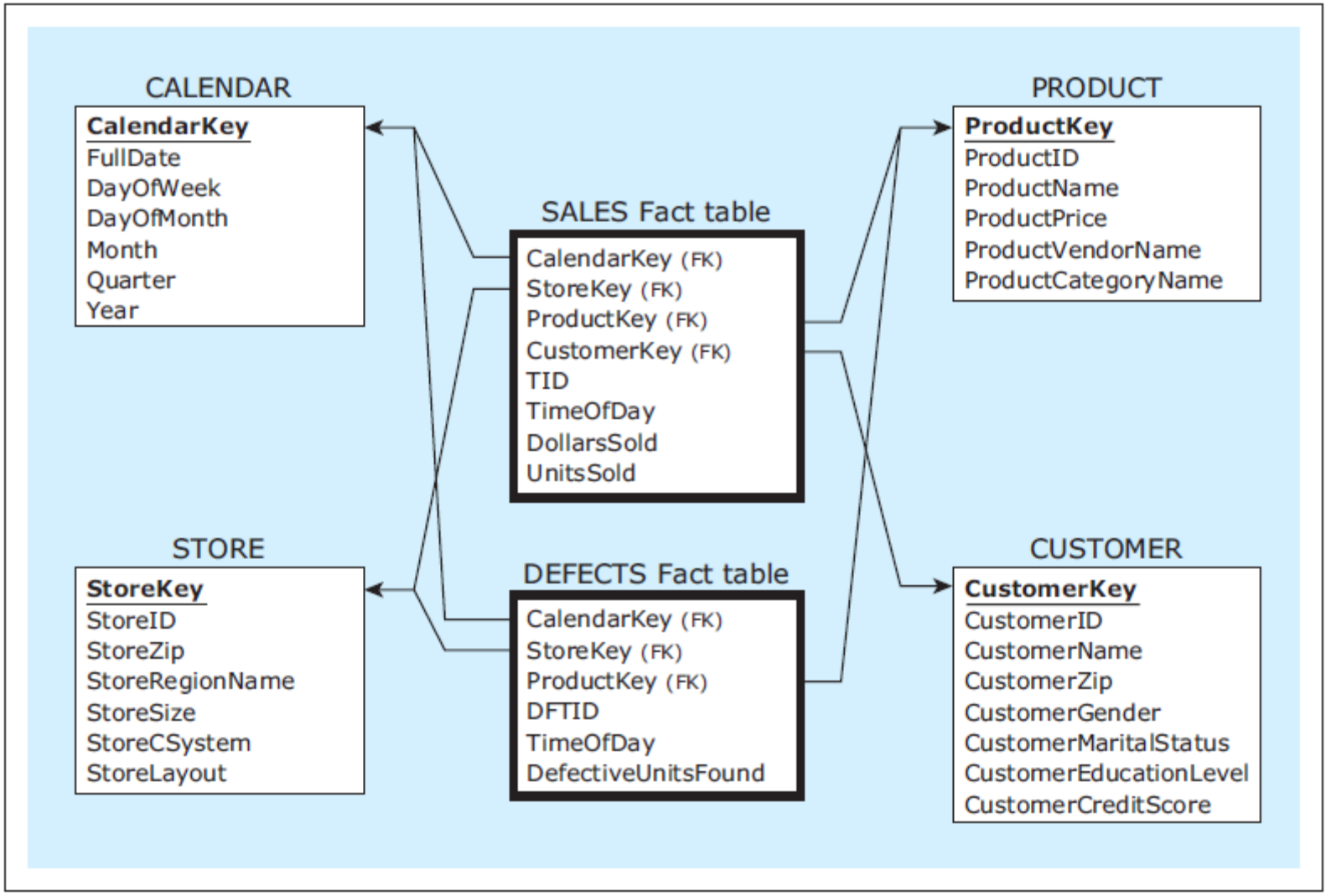
CategoryID	CategoryName
CP	Camping
FW	Footwear

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

Expanded Example: Dimensional Model Based on Multiple Sources

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



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

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

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

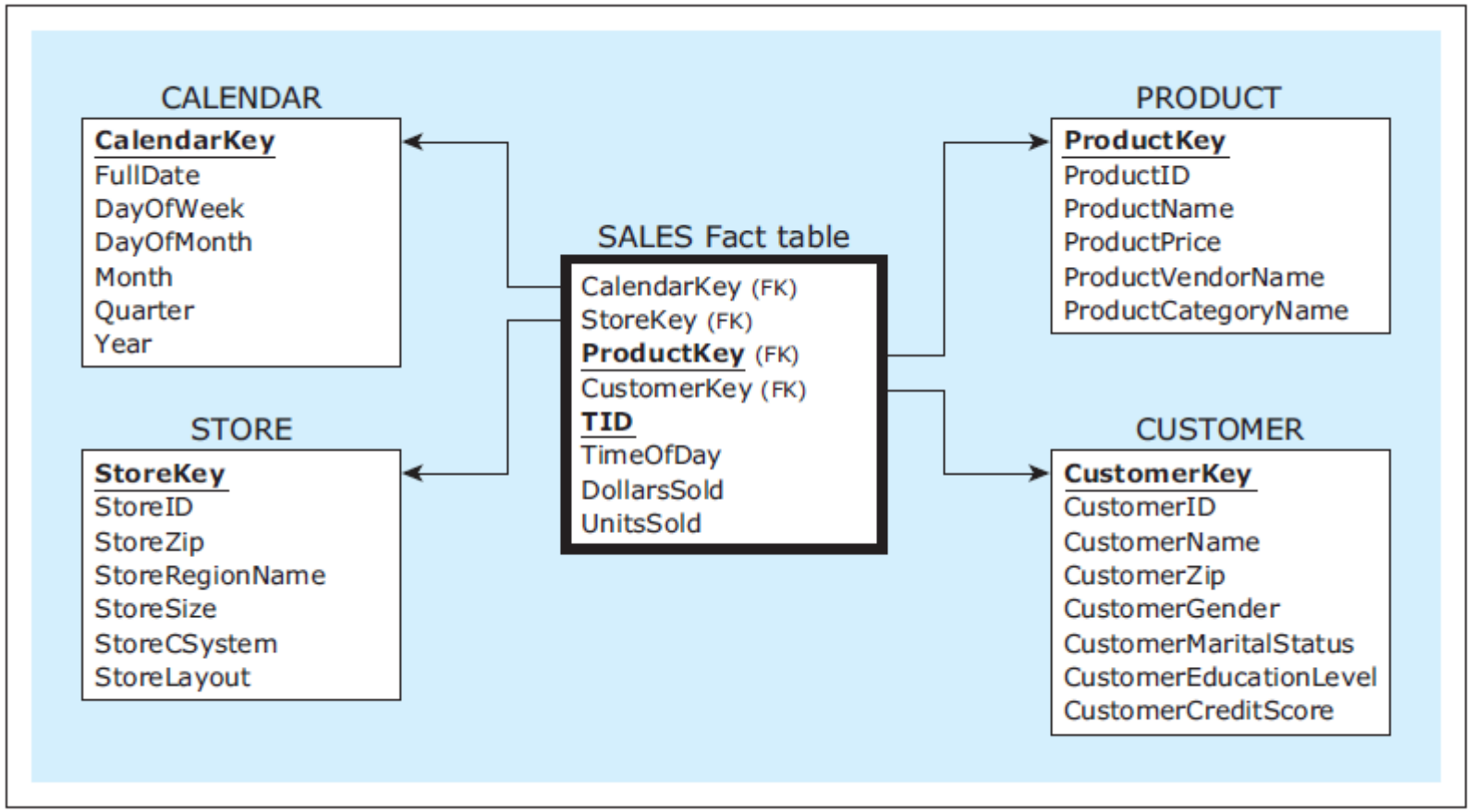
ProductID	TID	NoOfItems
1X1	T111	1
2X2	T222	1
3X3	T333	5
1X1	T333	1
4X4	T444	1
2X2	T444	2
4X4	T555	4
5X5	T555	2
6X6	T555	1
5X5	T666	2
6X6	T666	1

CUSTOMER

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

Detailed Fact Table Example

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



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

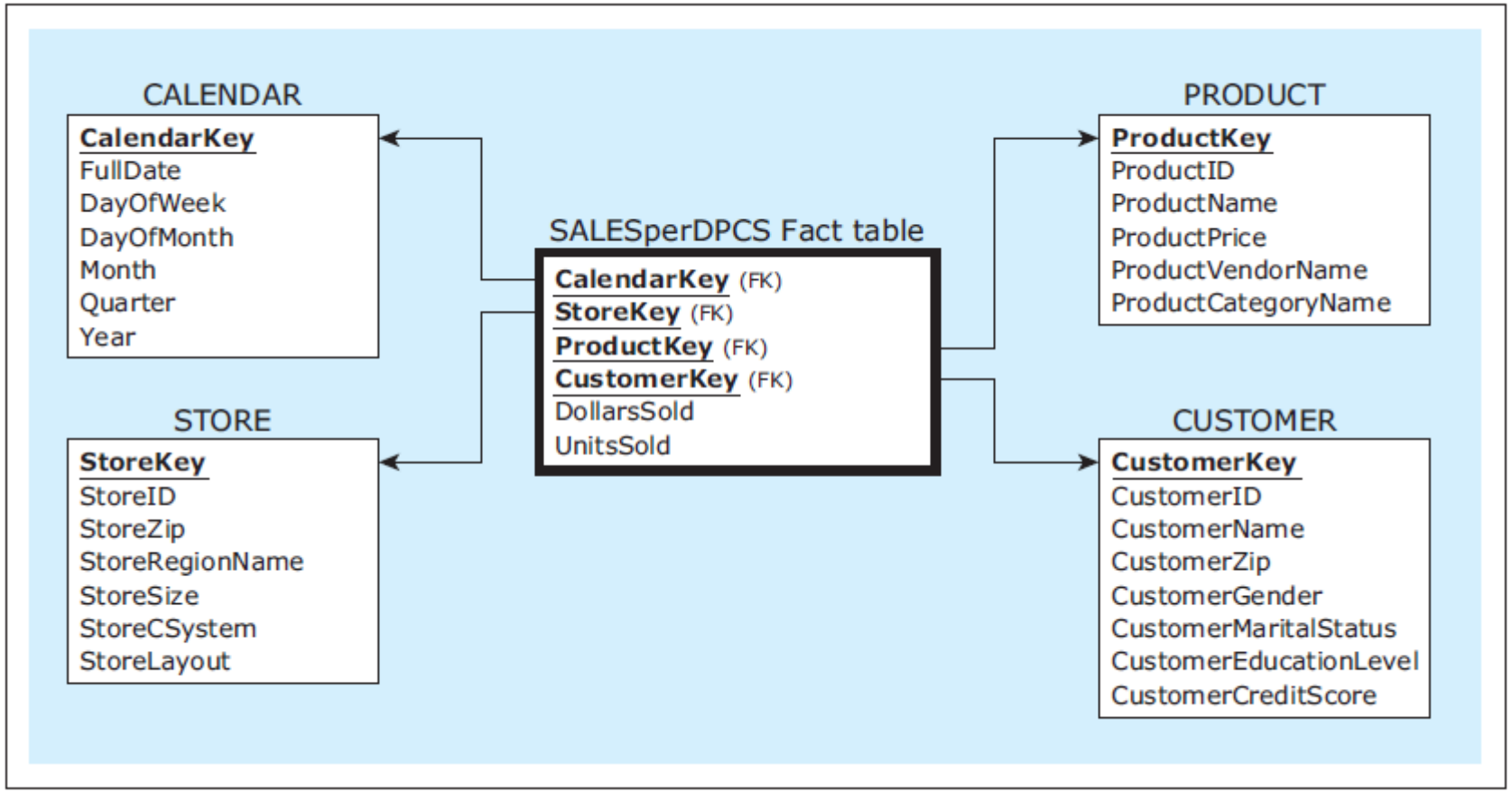
CustomerKey	CustomerID	Customer Name	Customer Zip	Customer Gender	Customer MaritalStatus	Customer EducationLevel	Customer CreditScore
1	1-2-333	Tina	60137	Female	Single	College	700
2	2-3-444	Tony	60611	Male	Single	High School	650
3	3-4-555	Pam	35401	Female	Married	College	623

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

Aggregated Fact Table Example 1

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



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

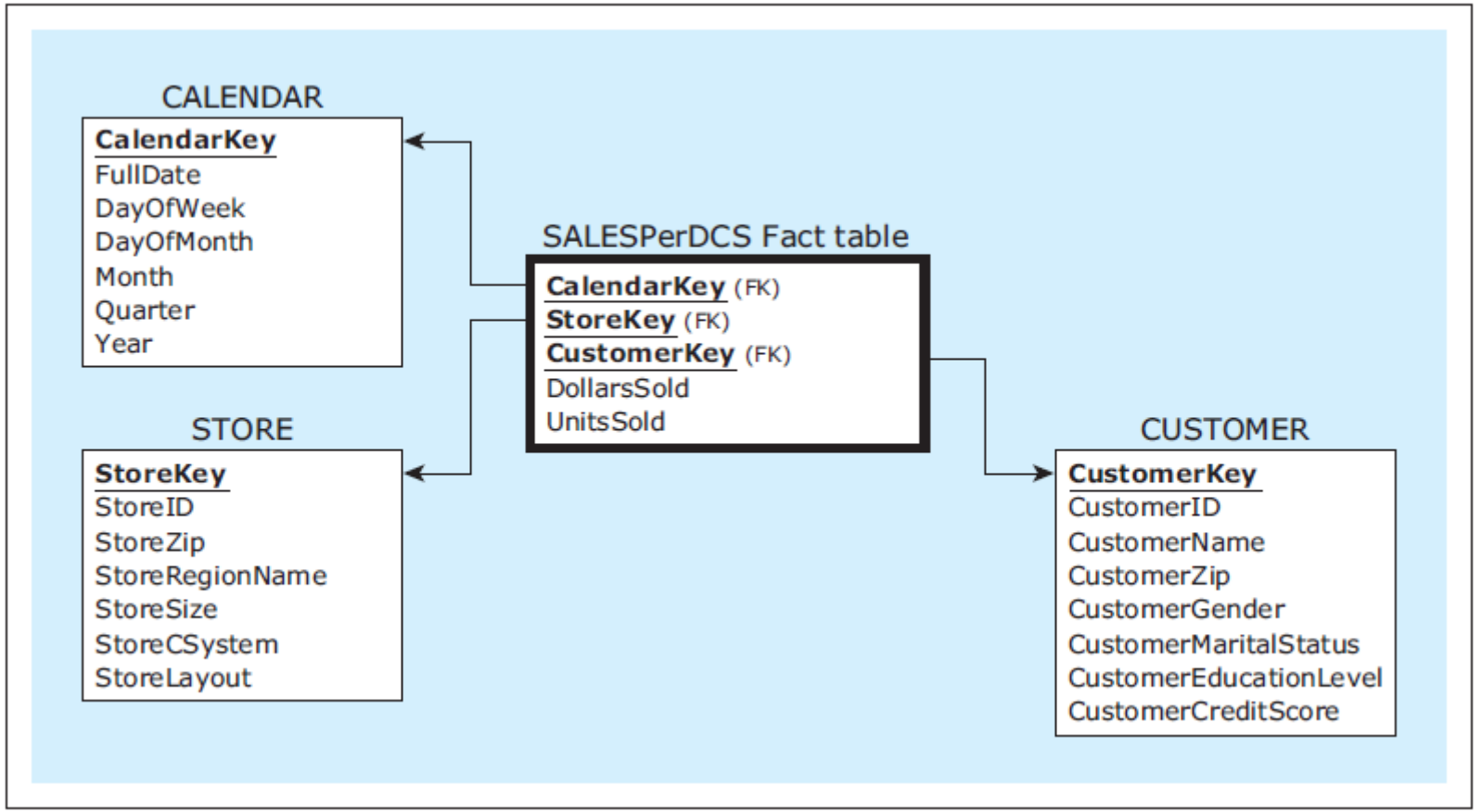
CustomerKey	CustomerID	Customer Name	Customer Zip	Customer Gender	Customer MaritalStatus	Customer EducationLevel	Customer CreditScore
1	1-2-333	Tina	60137	Female	Single	College	700
2	2-3-444	Tony	60611	Male	Single	High School	650
3	3-4-555	Pam	35401	Female	Married	College	623

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	← *Amounts from 8th and 10th records in SALES fact table in Figure 8.23 combined (added)*
2	3	6	2	$500	2	← *Amounts from 9th and 11th records in SALES fact table in Figure 8.23 combined (added)*

Aggregated Fact Table Example 2

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



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
2	3	3	$230	3
2	3	2	$1,460	10

Amounts from 3rd and 4th records in SALES fact table in Figure 8.23 combined (added)

Amounts from 5th and 6th records in SALES fact table in Figure 8.23 combined (added)

Amounts from 7th through 11th records in SALES fact table in Figure 8.23 combined (added)

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

DIMENSIONAL MODELING

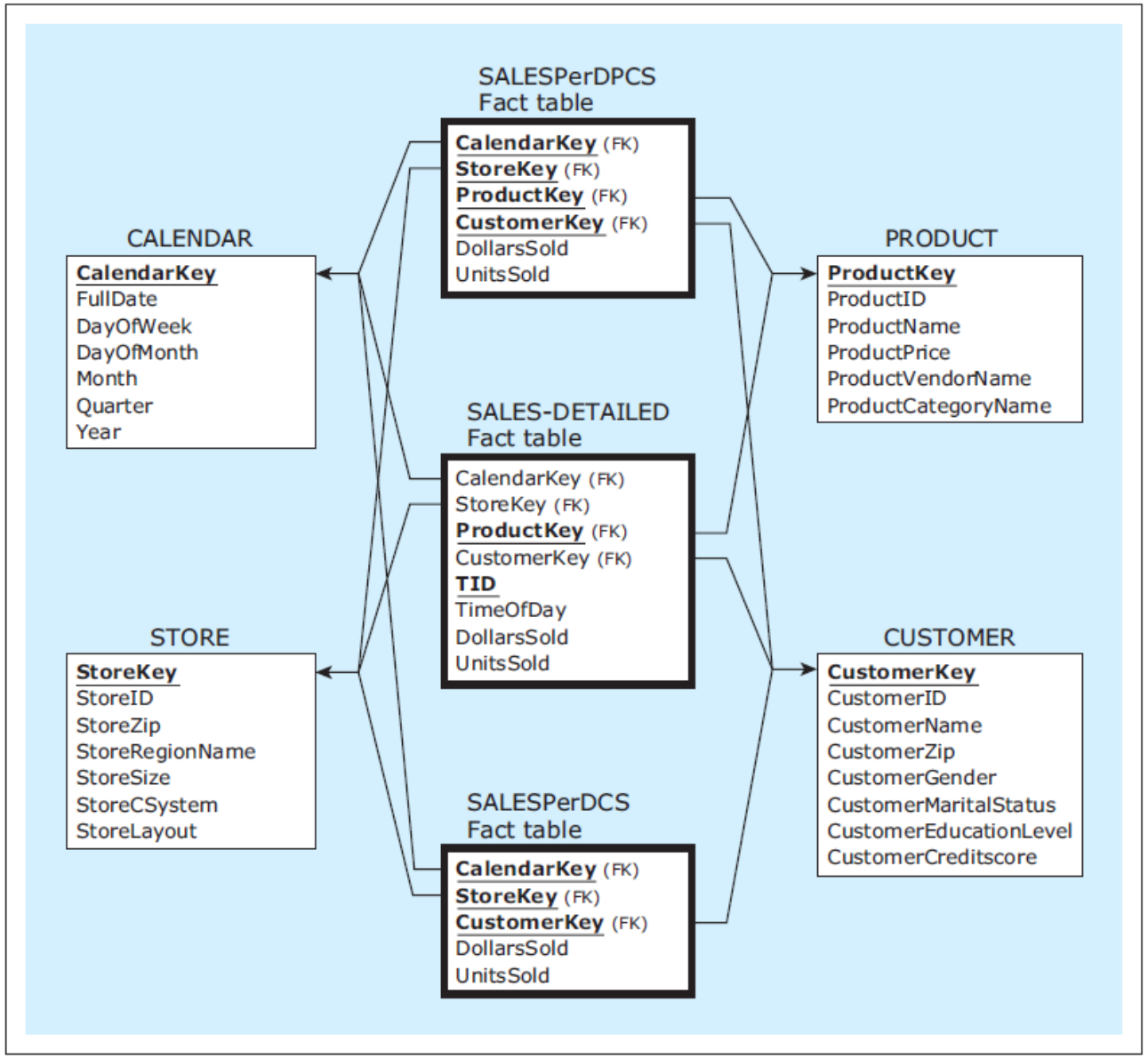
Granularity of the fact tables

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

A constellation of detailed and aggregated facts - Example



DIMENSIONAL MODELING

Line-item versus transaction-level detailed fact table

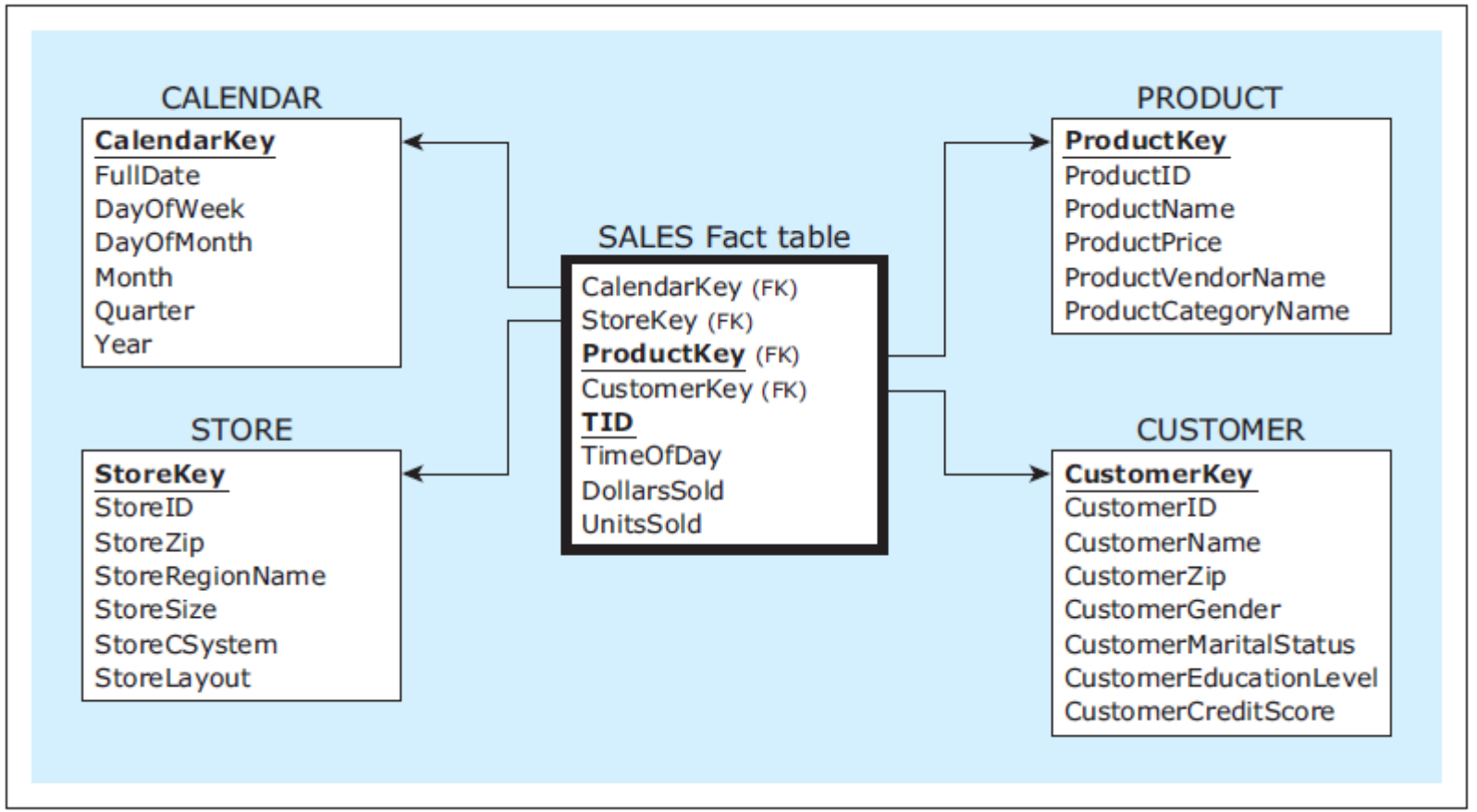
Line-item detailed fact table

Each row represents a line item of a particular transaction

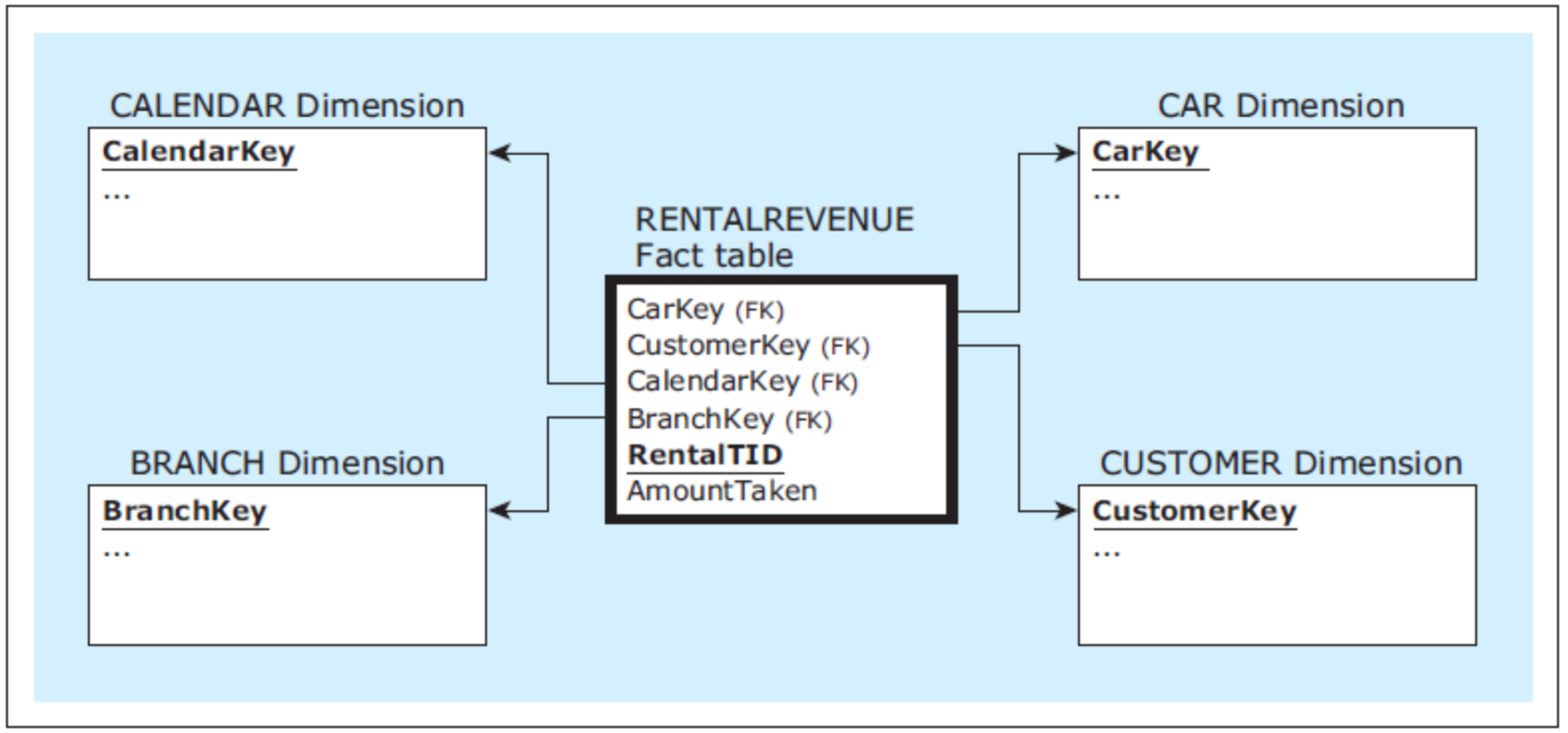
Transaction-level detailed fact table

Each row represents a particular transaction

Line-Item Detailed Fact Table Example



Transaction-Level Detailed Fact Table Example



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

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

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

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*

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

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

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*

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

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

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*

Snowflake Model - Example

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