CHAPTER

MODULE 2

Data and Knowledge Management

Business intelligence (BI)

Data and Knowledge Management

1. Database Approach
2. Big Data
3. Data Warehouses and Data Marts
4. Knowledge Management

Managing Data

- All IT applications require data.
- These data should be of high quality, meaning that they should be accurate, complete, timely, consistent, accessible, relevant, and concise.
- Unfortunately, the process of acquiring, keeping, and managing data is becoming increasingly difficult.

- Difficulties of Managing Data
- Data Governance

Difficulties of Managing Data

- The amount of data increases exponentially over time
- Data are scattered throughout organizations
- Data are generated from multiple sources (internal, personal, external)
  - **Internal Data Sources** (e.g., corporate databases and company documents)
  - **Personal Data Sources** (e.g., personal thoughts, opinions, and experiences)
  - **External Data Sources** (e.g., commercial databases, government reports, and corporate Web sites).
- New sources of data (e.g., blogs, podcasts, videocasts, and RFID tags and other wireless sensors)

Difficulties of Managing Data(cntd)

- Data Degradation(e.g., customers move to new addresses, change their names, etc.)
- **Data Rot:** refers primarily to problems with the media on which the data are stored. Over time, temperature, humidity, and exposure to light can cause physical problems with storage media and thus make it difficult to access the data.
- Data security, quality, and integrity are critical
- Legal requirements change frequently and differ among countries & industries

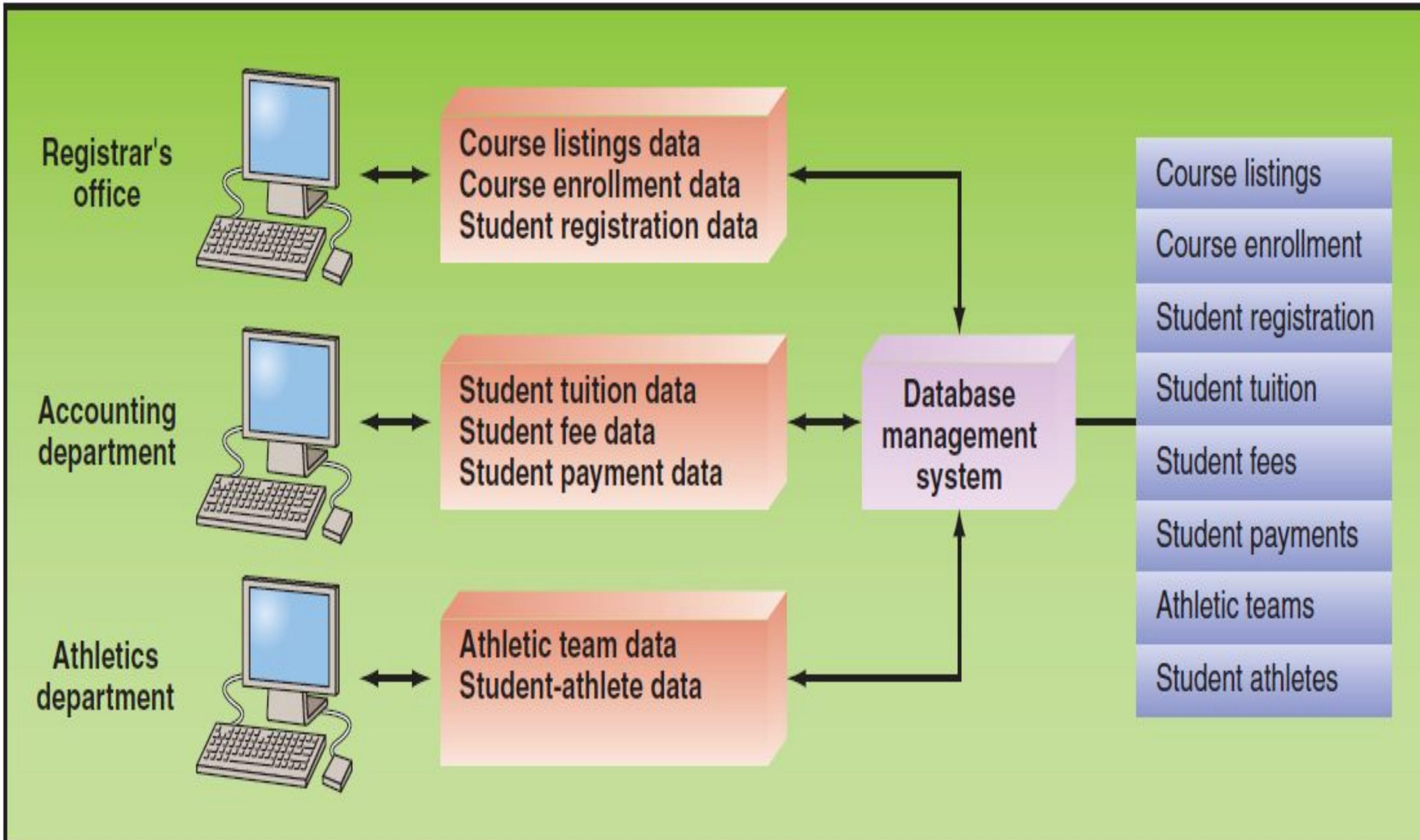
Data Governance

- **Data Governance:** is an approach to managing information across an entire organization involving a formal set of unambiguous rules for creating, collecting, handling, and protecting its information.
- **Master Data Management:** a strategy for data governance involving a process that spans all organizational business processes and applications providing companies with the ability to store, maintain, exchange, and synchronize a consistent, accurate, and timely for the company's master data.
- **Master Data:** a set of core data (e.g., customer, product, employee, vendor

Database Approach

- Data File: a collection of logically related records.
- Data Hierarchy
- Relational Database Model

Database Management System



DBMS Minimizes:

- Data redundancy:
  - Presence of duplicate data in multiple files
- Data inconsistency:
  - Same attribute has different values
- Data Isolation: Applications cannot access data associated with other applications
- Lack of flexibility
- Poor security
- Lack of data sharing and availability

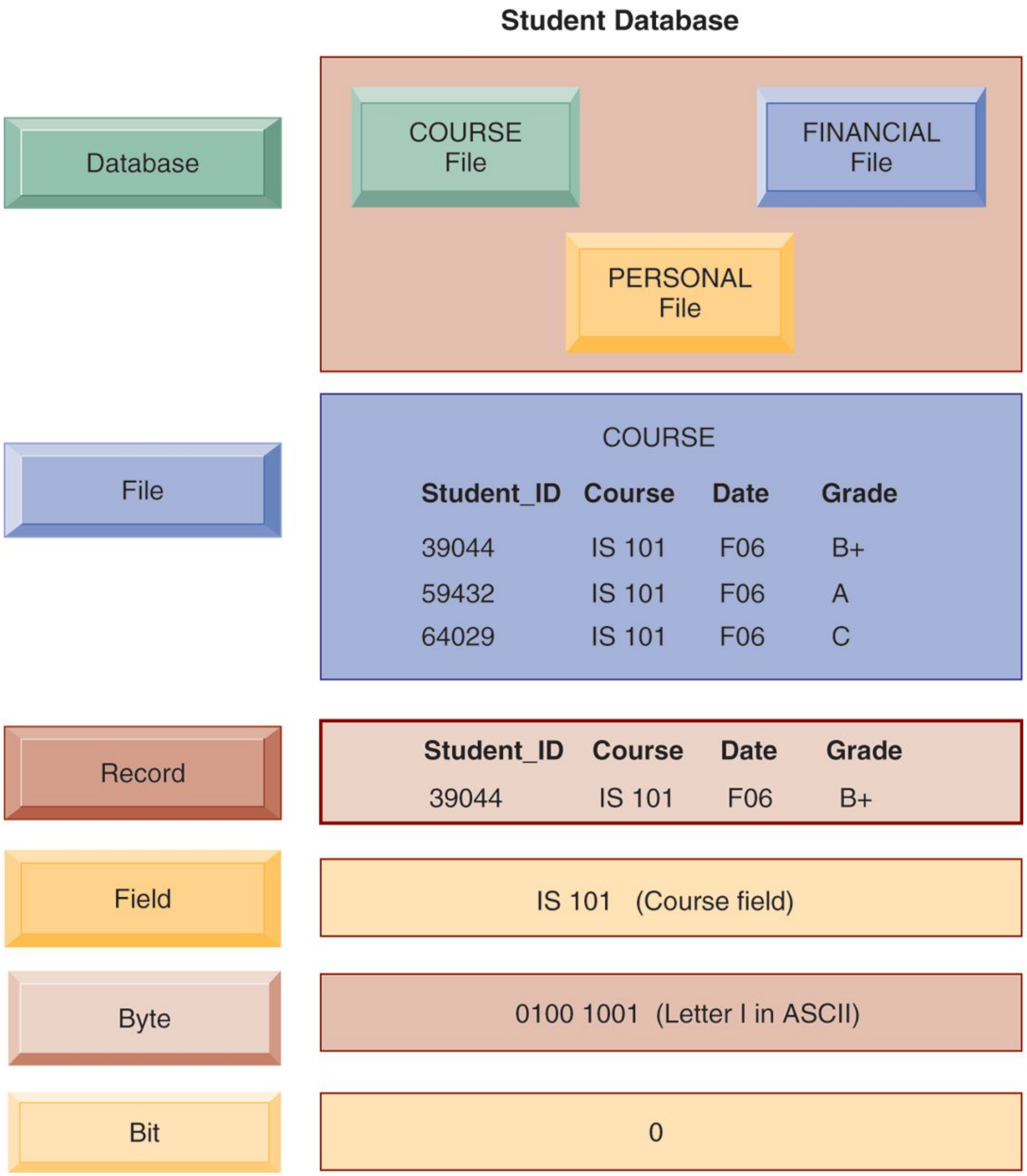
Database Management Systems (DBMS) Maximize:

- Data Security
- Data Integrity
- Data Independence

Data Hierarchy

- **Bit (binary digit):** represents the smallest unit of data a computer can process and it consists only of a 0 or a 1.
- **Byte:** A group of eight bits represents a single character (letter, number, or symbol).
- **Field:** A column of data containing a logical grouping of characters into a word, a small group of words (e.g., last name, social security number, etc.).
- **Record:** A logical grouping of related fields in a row (e.g., student’s name, the courses taken, the date, and the grade).
- **Data File:** logical grouping of related records is called a data file or a table similar in appearance to a spreadsheet in Excel consisting of multiple columns and multiple rows.
- **Database:** logical grouping of related data files (database tables).

THE DATA HIERARCHY



Hierarchy of Data for a Computer-Based File

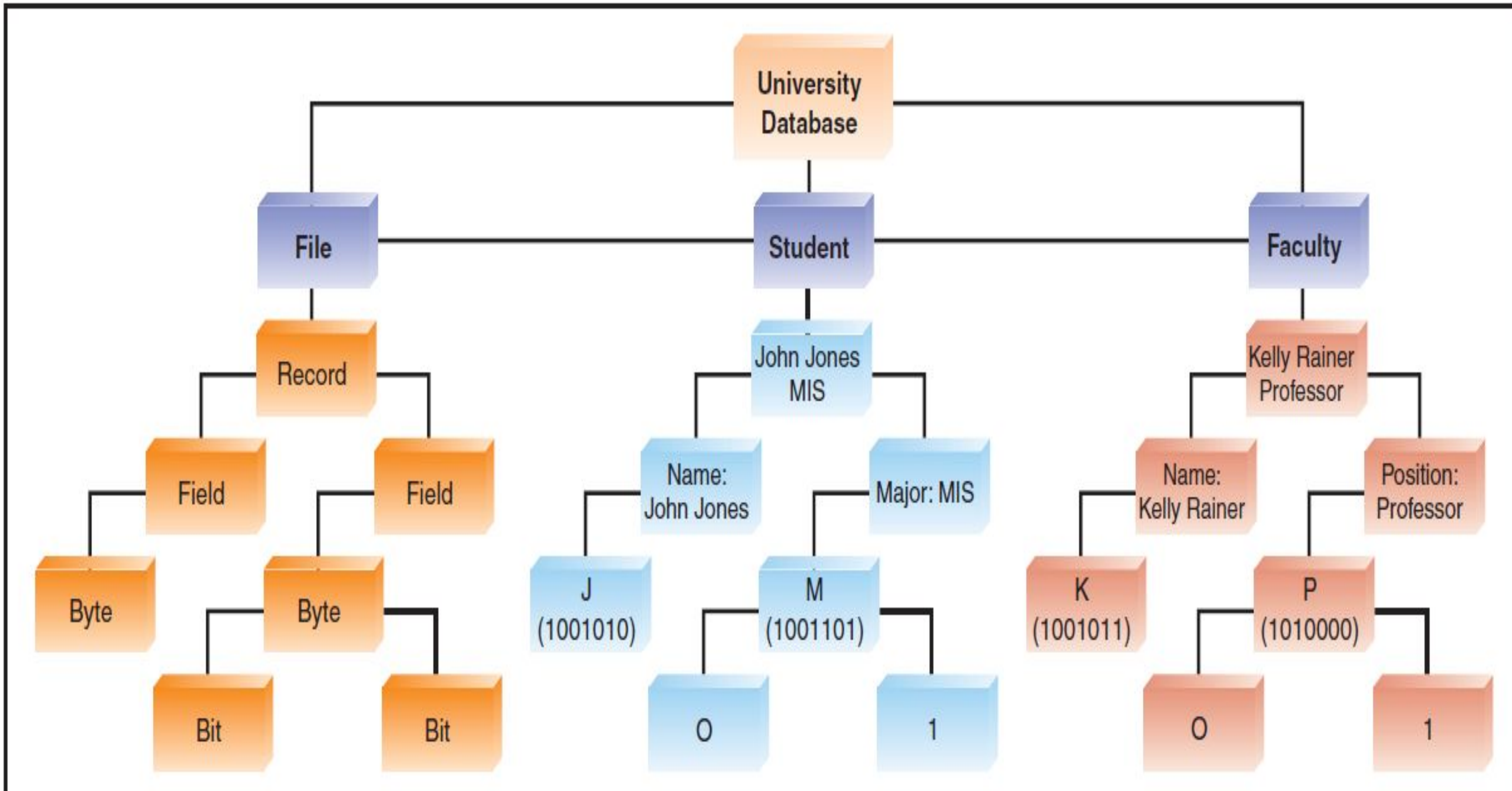


Figure 3.2 Hierarchy of data for a computer-based file.

The Relational Database Model

- Database Management System (DBMS)
- Relational Database Model
- Data Model
- Entity
- Instance
- Attribute

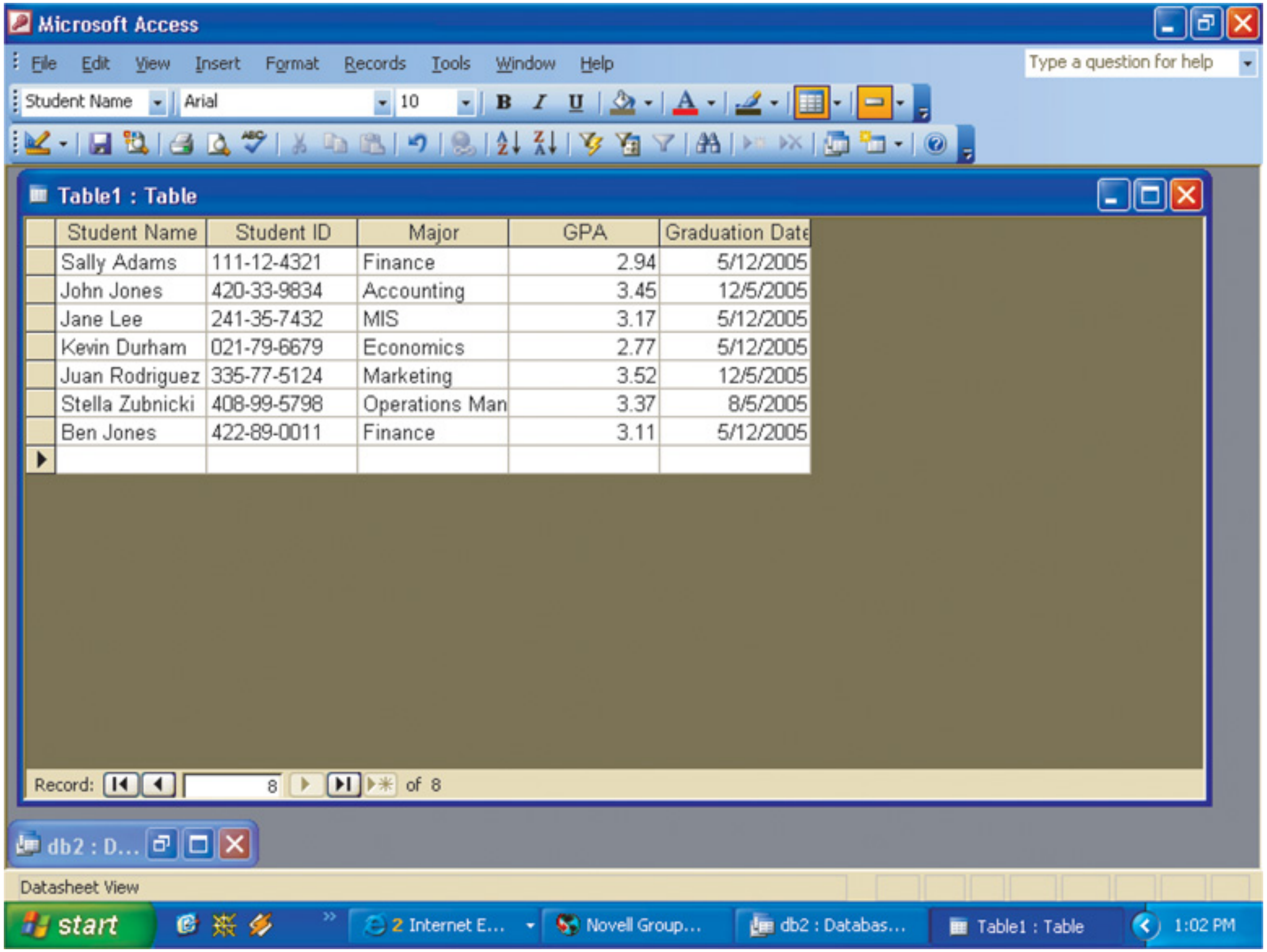
The Relational Database Model

- **Database Management System (DBMS):** a set of programs that provide users with tools to create and manage a database.
- **Relational Database Model:** is based on the concept of two-dimensional tables and is usually designed with a number of related tables with each of these tables contains records (listed in rows) and attributes (listed in columns).
- **Data Model:** a diagram that represents entities in the database and their relationships.
- **Entity:** a person, place, thing, or event (e.g., customer, an employee, or a product).
- **Record:** generally describes an entity and an instance of an entity refers to each row in a relational table.
- **Attribute:** each characteristic or quality of a particular entity.

The Relational Database Model (continued)

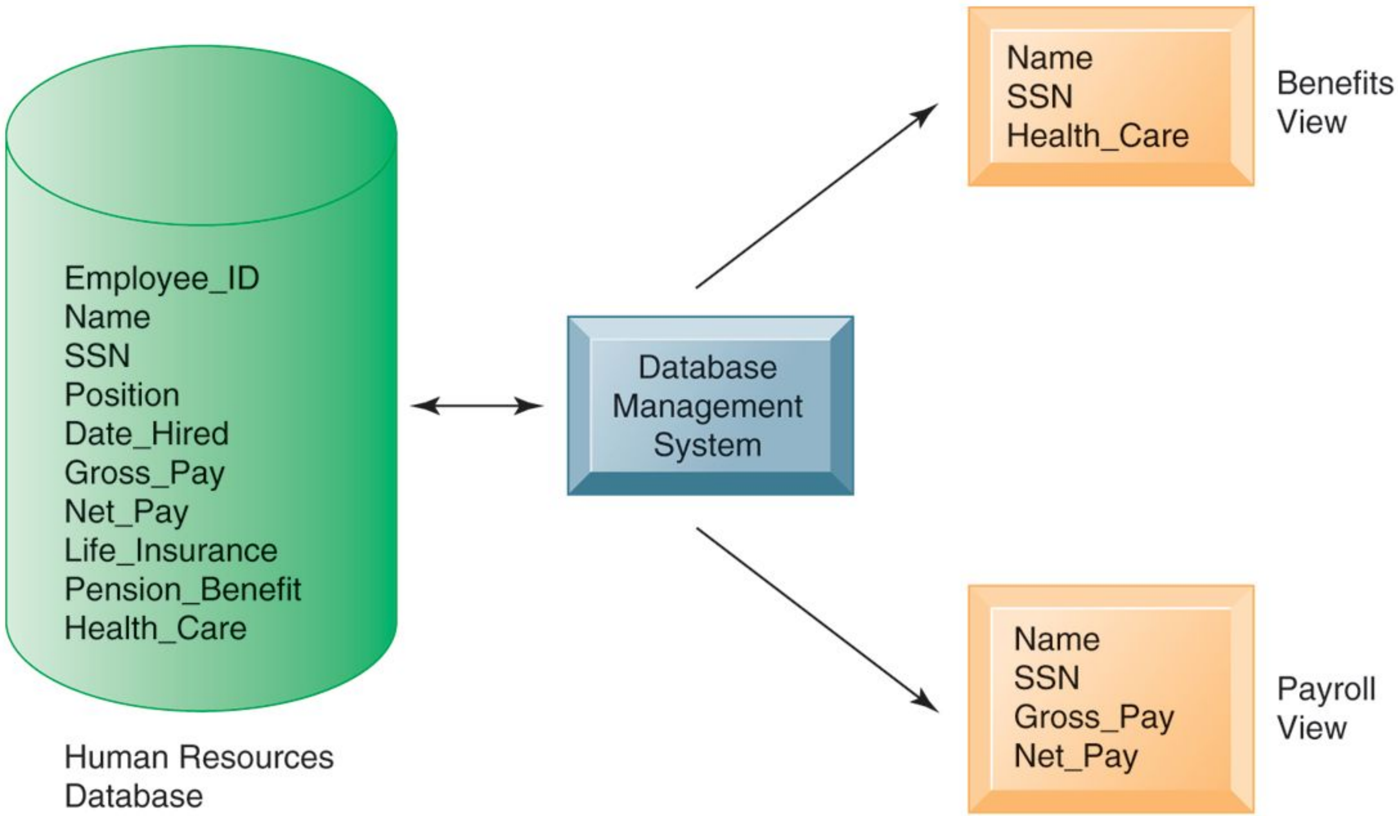
- **Primary Key:** a field in a database that uniquely identify each record so that it can be retrieved, updated, and sorted.
- **Secondary Key:** Secondary Key is the key that has not been selected to be the primary key. However, it is considered a candidate key for the primary key.
- Therefore, a candidate key not selected as a primary key is called secondary key.
- **Foreign Key:** a field (or group of fields) in one table that uniquely identifies a row of another table. It is used to establish and enforce a link between two tables.

Figure 3.3: Student Database Example



Student Name	Student ID	Major	GPA	Graduation Date
Sally Adams	111-12-4321	Finance	2.94	5/12/2005
John Jones	420-33-9834	Accounting	3.45	12/5/2005
Jane Lee	241-35-7432	MIS	3.17	5/12/2005
Kevin Durham	021-79-6679	Economics	2.77	5/12/2005
Juan Rodriguez	335-77-5124	Marketing	3.52	12/5/2005
Stella Zubnicki	408-99-5798	Operations Man	3.37	8/5/2005
Ben Jones	422-89-0011	Finance	3.11	5/12/2005

HUMAN RESOURCES DATABASE WITH MULTIPLE VIEWS



A single human resources database provides many different views of data, depending on the information requirements of the user. Illustrated here are two possible views, one of interest to a benefits specialist and one of interest to a member of the company's payroll department.

The Database Approach to Data Management

- **Relational DBMS**
  - Represent data as two-dimensional tables
  - Each table contains data on entity and attributes
- **Table: grid of columns and rows**
  - Rows (tuples): Records for different entities
  - Fields (columns): Represents attribute for entity
  - Key field: Field used to uniquely identify each record

Relational Database Tables

A relational database organizes data in the form of two-dimensional tables. Illustrated here are tables for the entities SUPPLIER and PART showing how they represent each entity and its attributes. Supplier Number is a primary key for the SUPPLIER table and a foreign key for the PART table.

SUPPLIER

Columns (Attributes, Fields)

Supplier_Number	Supplier_Name	Supplier_Street	Supplier_City	Supplier_State	Supplier_Zip
8259	CBM Inc.	74 5th Avenue	Dayton	OH	45220
8261	B. R. Molds	1277 Gandolly Street	Cleveland	OH	49345
8263	Jackson Composites	8233 Micklin Street	Lexington	KY	56723
8444	Bryant Corporation	4315 Mill Drive	Rochester	NY	11344

Rows (Records, Tuples)

Key Field (Primary Key)

PART

Part_Number	Part_Name	Unit_Price	Supplier_Number
137	Door latch	22.00	8259
145	Side mirror	12.00	8444
150	Door molding	6.00	8263
152	Door lock	31.00	8259
155	Compressor	54.00	8261
178	Door handle	10.00	8259

Primary Key

Foreign Key

The Database Approach to Data Management

- **Operations of a Relational DBMS**
  - **Three basic operations used to develop useful sets of data**
    - **SELECT:** Creates subset of data of all records that meet stated criteria
    - **JOIN:** Combines relational tables to provide user with more information than available in individual tables
    - **PROJECT:** Creates subset of columns in table, creating tables with only the information specified

THE THREE BASIC OPERATIONS OF A RELATIONAL DBMS

PART

Part_Number	Part_Name	Unit_Price	Supplier_Number
137	Door latch	22.00	8259
145	Side mirror	12.00	8444
150	Door molding	6.00	8263
152	Door lock	31.00	8259
155	Compressor	54.00	8261
178	Door handle	10.00	8259

Select Part_Number = 137 or 150

SUPPLIER

Supplier_Number	Supplier_Name	Supplier_Street	Supplier_City	Supplier_State	Supplier_Zip
8259	CBM Inc.	74 5th Avenue	Dayton	OH	45220
8261	B. R. Molds	1277 Gandolly Street	Cleveland	OH	49345
8263	Jackson Components	8233 Micklin Street	Lexington	KY	56723
8444	Bryant Corporation	4315 Mill Drive	Rochester	NY	11344

Join by Supplier_Number

Part_Number	Part_Name	Supplier_Number	Supplier_Name
137	Door latch	8259	CBM Inc.
150	Door molding	8263	Jackson Components

Project selected columns

The select, join, and project operations enable data from two different tables to be combined and only selected attributes to be displayed.

The Database Approach to Data Management

- **object-oriented DBMS**

- Many applications today and in the future require databases that can store and retrieve not only structured numbers and characters but also drawings, images, photographs, voice, and full-motion video.
- DBMS designed for organizing structured data into rows and columns are not well suited to handling graphics based or multimedia applications. Object-oriented databases are better suited for this purpose.
- An **object-oriented DBMS** stores the data and procedures that act on those data as objects that can be automatically retrieved and shared.

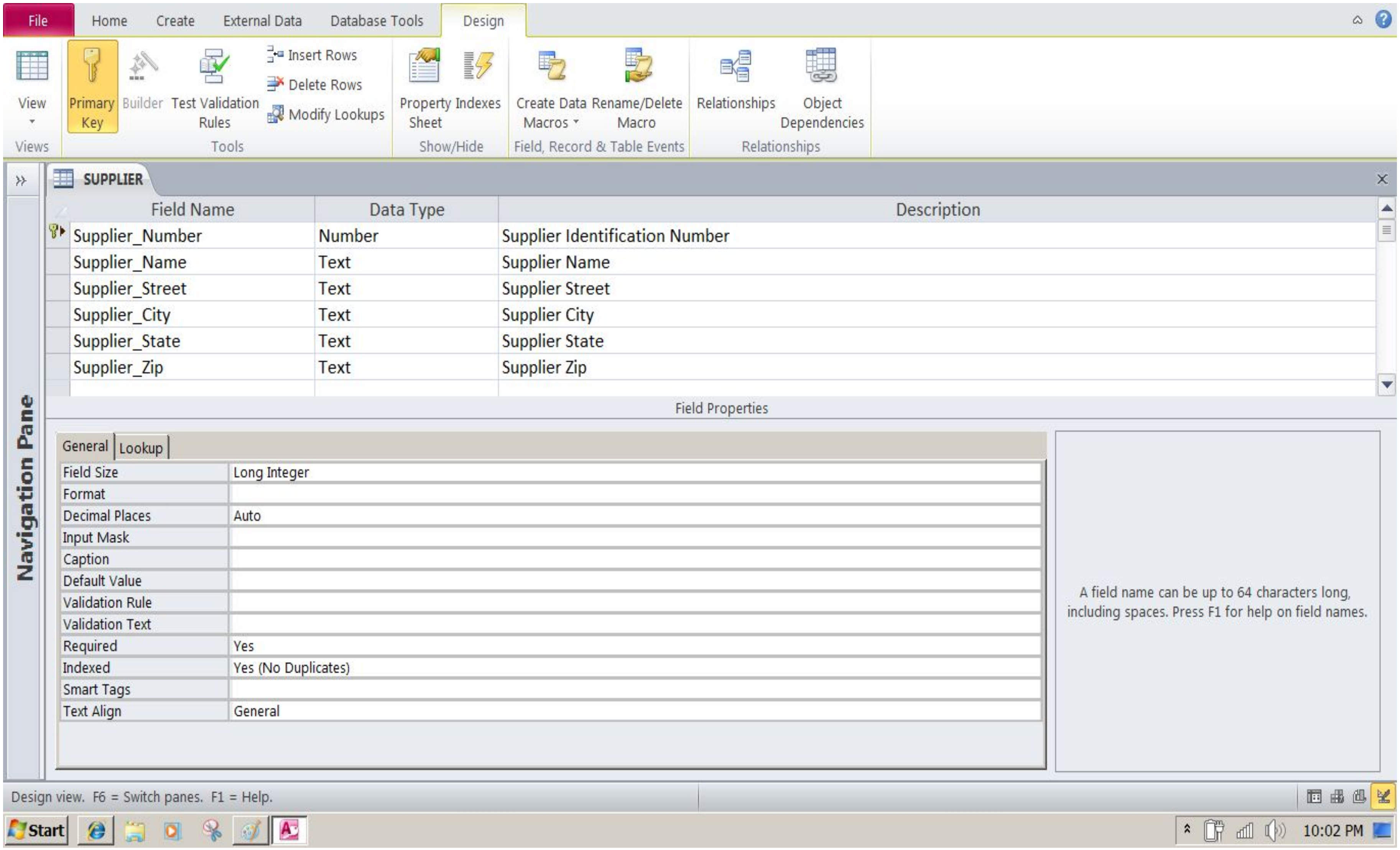
The Database Approach to Data Management

- Non-relational databases: “NoSQL”
  - More flexible data model
  - Data sets stored across distributed machines
  - Easier to scale
  - Handle large volumes of unstructured and structured data (Web, social media, graphics)
- Databases in the cloud
  - Typically, less functionality than on-premises DBs
  - Amazon Relational Database Service, Microsoft SQL Azure
  - Private clouds

The Database Approach to Data Management

- **Capabilities of database management systems**
  - **Data definition capability:**
    - Specifies structure of database content
    - used to create tables and define characteristics of fields
  - **Data dictionary:**
    - Automated or manual file storing definitions of data elements and their characteristics(name, description, size, type, format etc.)
  - **Data manipulation language:**
    - Used to add, change, delete, retrieve data from database
      - Structured Query Language (SQL)
      - Microsoft Access user tools for generating SQL
  - **Many DBMS have report generation capabilities for creating polished reports (Crystal Reports)**

MICROSOFT ACCESS DATA DICTIONARY FEATURES



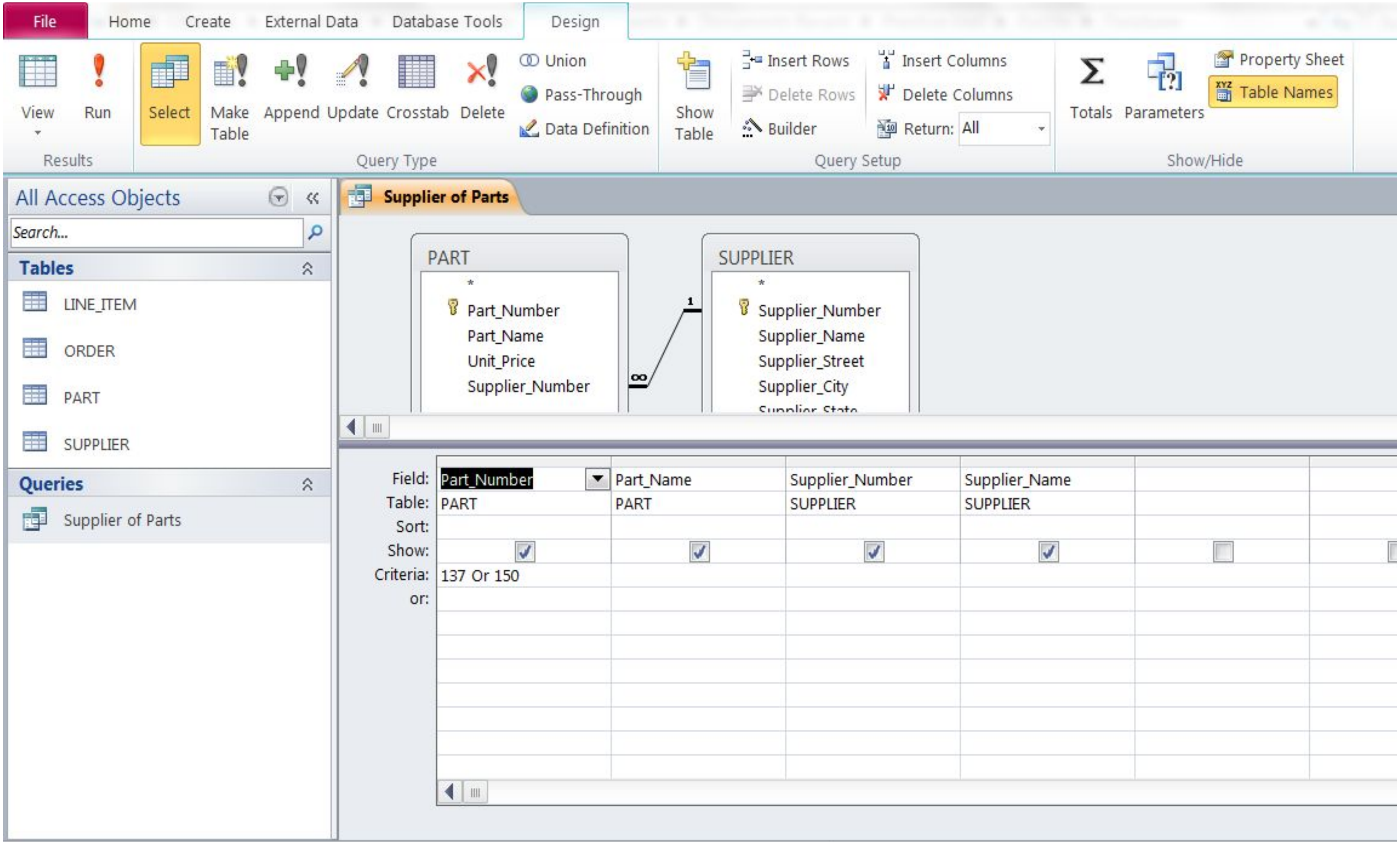
Supplier table: Microsoft Access has a rudimentary data dictionary capability that displays information about the size, format, and other characteristics of each field in a database

EXAMPLE OF AN SQL QUERY

```
SELECT PART.Part_Number, PART.Part_Name, SUPPLIER.Supplier_Number,
SUPPLIER.Supplier_Name
FROM PART, SUPPLIER
WHERE PART.Supplier_Number = SUPPLIER.Supplier_Number AND
Part_Number = 137 OR Part_Number = 150;
```

Illustrated here are the SQL statements for a query to select suppliers for parts 137 or 150. They produce a list with the same results as Figure 6-5.

AN ACCESS QUERY



Illustrated here is how the query in Figure 6-7 would be constructed using Microsoft Access query building tools. It shows the tables, fields, and selection criteria used for the query.

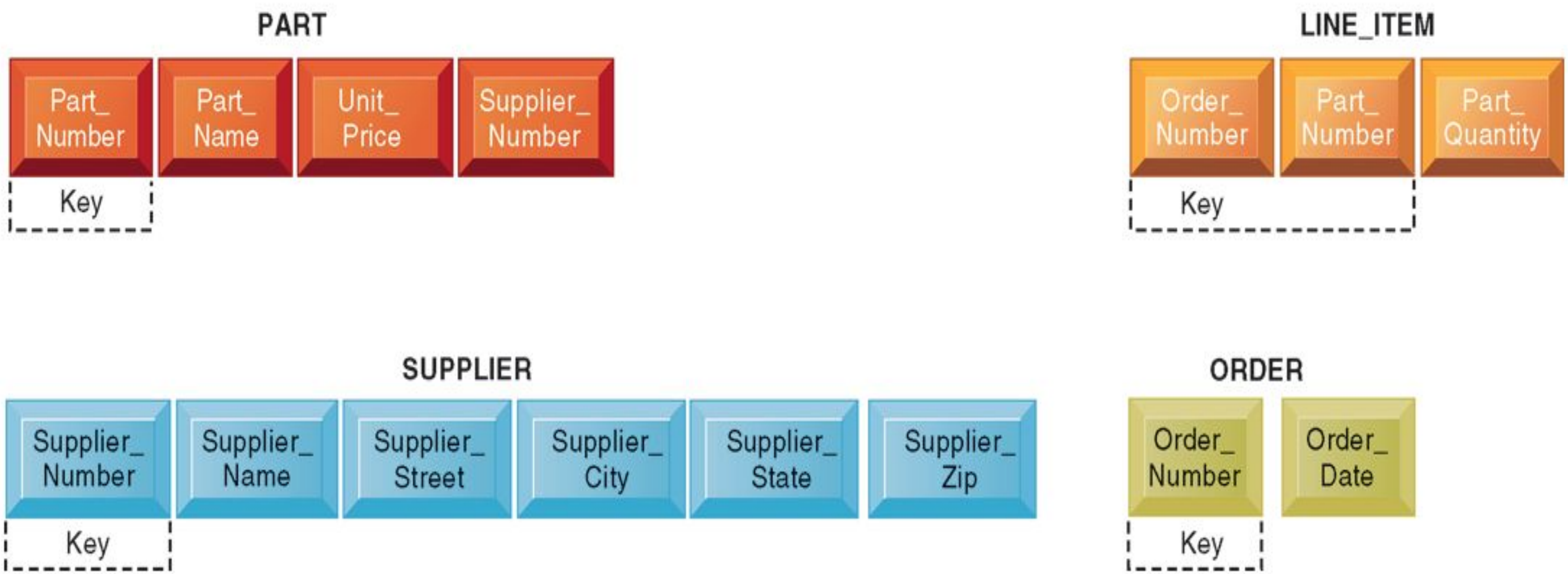
The Database Approach to Data Management

- **Designing Databases**
  - Conceptual (logical) design: abstract model from business perspective
  - Physical design: How database is arranged on direct-access storage devices
- **Design process identifies:**
  - Relationships among data elements, redundant database elements
  - Most efficient way to group data elements to meet business requirements, needs of application programs
- **Normalization**
  - The process of creating small, stable, yet flexible and adaptive data structures from complex groups of data is called **normalization**.
  - minimize redundant data elements.

AN UNNORMALIZED RELATION FOR ORDER

- An unnormalized relation contains repeating groups.
- For example, there can be many parts and suppliers for each order.
- There is only a one-to-one correspondence between Order_Number and Order_Date.

NORMALIZED TABLES CREATED FROM ORDER



- After normalization, the original relation ORDER has been broken down into four smaller relations.
- The relation ORDER is left with only two attributes.
- The relation LINE_ITEM has a combined, or concatenated, key consisting of Order_Number and Part_Number.

Cardinality Symbols

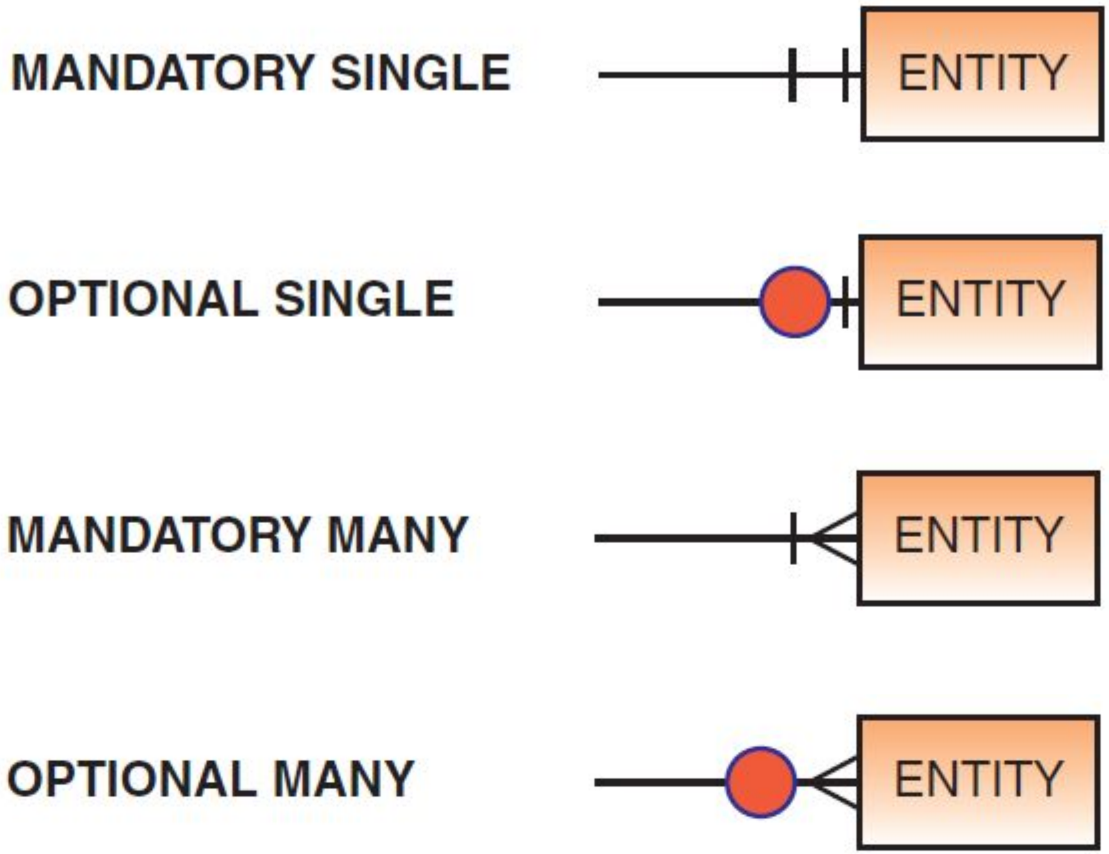


Figure PI3.1 Cardinality symbols.

One-to-One Relationship

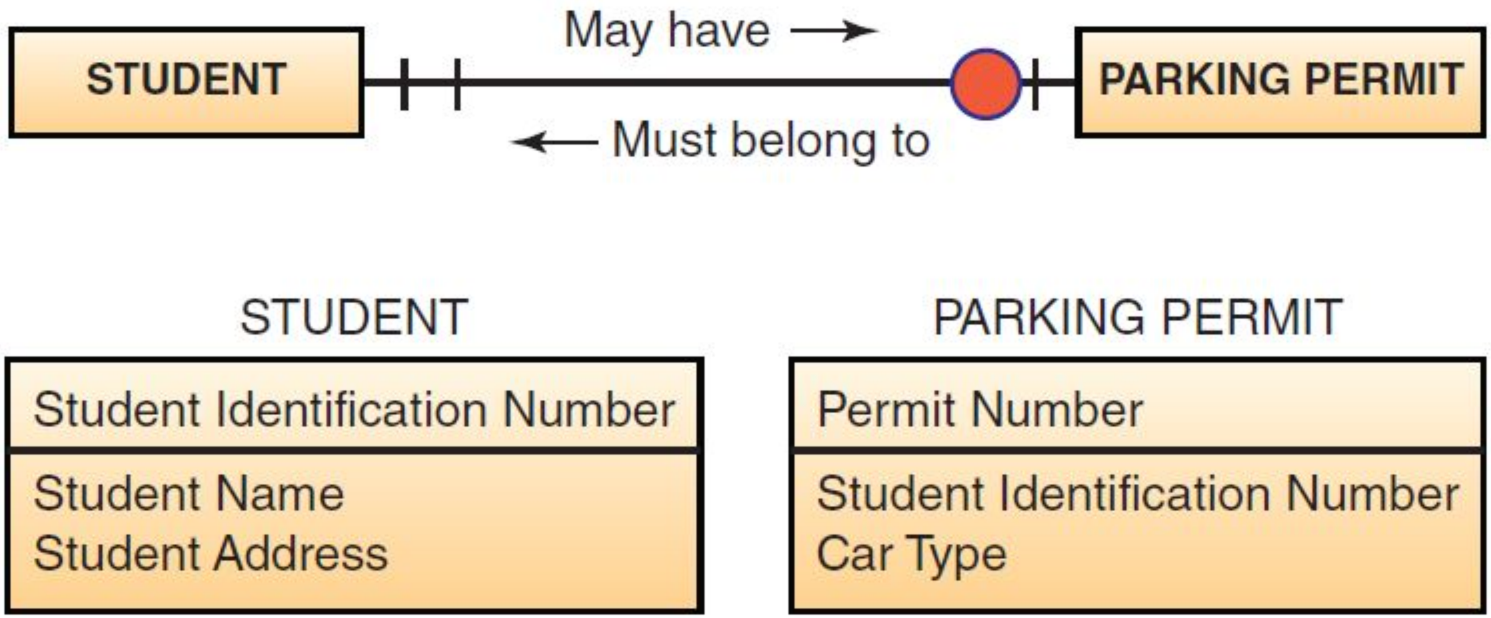


Figure PI3.2 One-to-one relationship.

One-to-Many Relationship

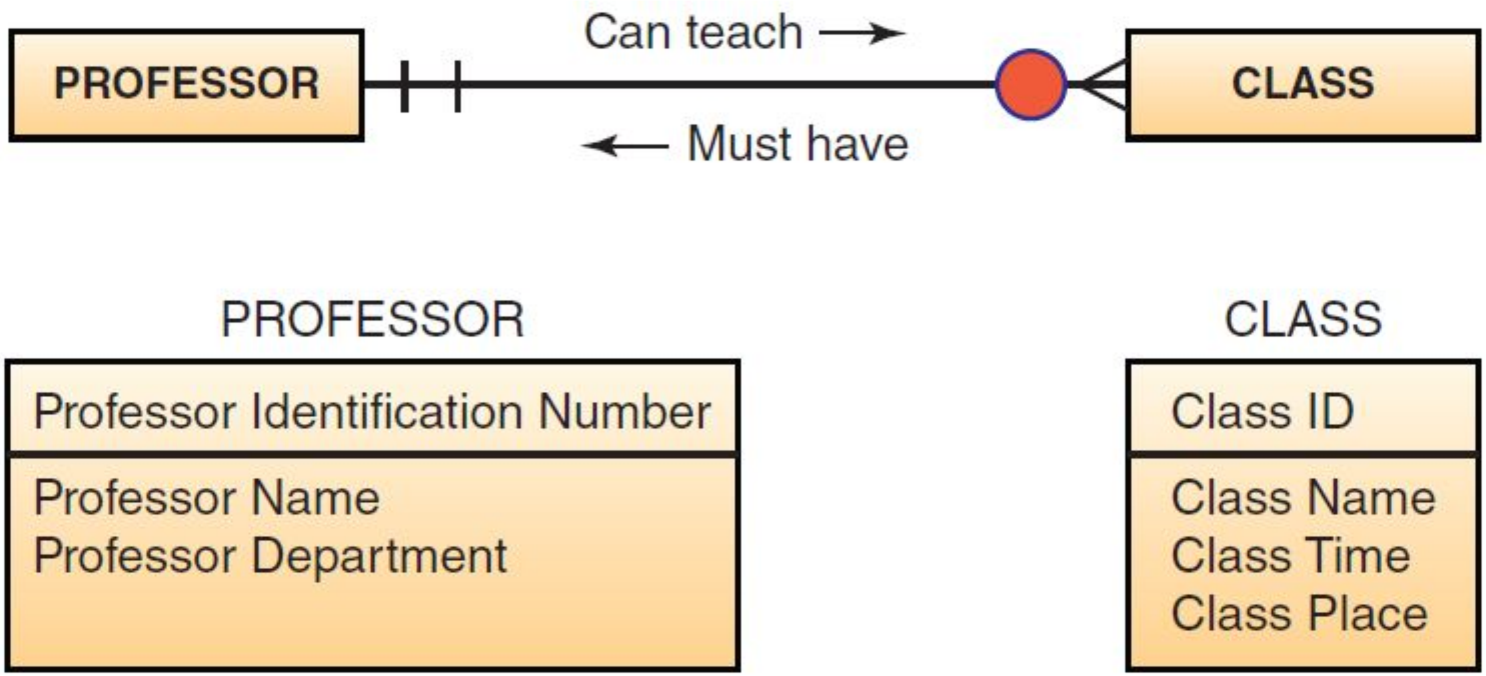


Figure PI3.3 One-to-many relationship.

Many-to-Many Relationship

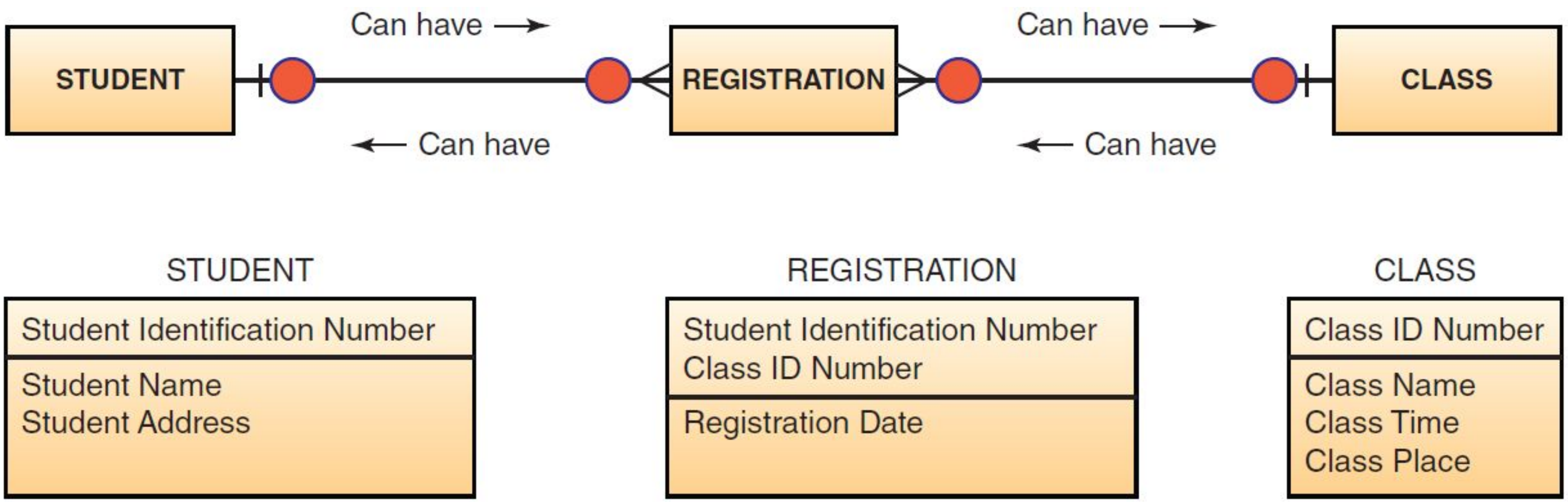


Figure PI3.4 Many-to-many relationship.

Raw Data Gathered from Pizza Shop Orders

Order Number	Order Date	Customer ID	Customer F Name	Customer L Name	Customer Address	Zip Code	Pizza Code	Pizza Name	Quantity	Price	Total Price
1116	9/1/14	16421	Rob	Penny	123 Main St.	37411	P	Pepperoni	1	$11.00	$41.00
							MF	Meat Feast	1	$12.00	
							V	Vegetarian	2	$9.00	
1117	9/2/14	17221	Beth	Jones	41 Oak St.	29416	HM	Ham and Mushroom	3	$10.00	$56.00
							MF	Meat Feast	1	$12.00	
							TH	The Hawaiian	1	$14.00	

Figure PI3.5 Raw data gathered from orders at the pizza shop.

Functional Dependency from Pizza Shop

Order Number	→	Order Date
Order Number	→	Quantity
Order Number	→	Total Price
Customer ID	→	Customer F Name
Customer ID	→	Customer L Name
Customer ID	→	Customer Address
Customer ID	→	Zip Code
Customer ID	→	Total Price
Pizza Code	→	Pizza Name
Pizza Code	→	Price

Figure PI3.6 Functional dependencies in pizza shop example.

1st Normal Form for Pizza Shop Database

Order Number	Order Date	Customer ID	Customer F Name	Customer L Name	Customer Address	Zip Code	Pizza Code	Pizza Name	Quantity	Price	Total Price
1116	9/1/14	16421	Rob	Penny	123 Main St.	37411	P	Pepperoni	1	$11.00	$41.00
1116	9/1/14	16421	Rob	Penny	123 Main St.	37411	MF	Meat Feast	1	$12.00	$41.00
1116	9/1/14	16421	Rob	Penny	123 Main St.	37411	V	Vegetarian	2	$9.00	$41.00
1117	9/2/14	17221	Beth	Jones	41 Oak St.	29416	HM	Ham and Mushroom	3	$10.00	$56.00
1117	9/2/14	17221	Beth	Jones	41 Oak St.	29416	MF	Meat Feast	1	$12.00	$56.00
1117	9/2/14	17221	Beth	Jones	41 Oak St.	29416	TH	The Hawaiian	1	$14.00	$56.00

Figure PI3.7 First normal form for data from pizza shop.

2nd Normal Form for Pizza Shop Database

Order Number	Order Date	Customer ID	Customer F Name	Customer L Name	Customer Address	Zip Code	Total Price
1116	9/1/14	16421	Rob	Penny	123 Main St.	37411	$41.00
1116	9/1/14	16421	Rob	Penny	123 Main St.	37411	$41.00
1116	9/1/14	16421	Rob	Penny	123 Main St.	37411	$41.00
1117	9/2/14	17221	Beth	Jones	41 Oak St.	29416	$56.00
1117	9/2/14	17221	Beth	Jones	41 Oak St.	29416	$56.00
1117	9/2/14	17221	Beth	Jones	41 Oak St.	29416	$56.00

Order Number	Pizza Code	Quantity
1116	P	1
1116	MF	1
1116	V	2
1117	HM	3
1117	MF	1
1117	TH	1

Pizza Code	Pizza Name	Price
P	Pepperoni	$11.00
MF	Meat Feast	$12.00
V	Vegetarian	$9.00
HM	Ham and Mushroom	$10.00
TH	The Hawaiian	$14.00

Figure PI3.8 Second normal form for data from pizza shop.

3rd Normal Form for Pizza Shop Database

ORDER

Order Number	Order Date	Customer ID	Total Price
1116	9/1/14	16421	$41.00
1117	9/2/14	17221	$56.00

CUSTOMER

Customer ID	Customer F Name	Customer L Name	Customer Address	Zip Code
16421	Rob	Penny	123 Main St.	37411
17221	Beth	Jones	41 Oak St.	29416

ORDER-PIZZA

Order Number	Pizza Code	Quantity
1116	P	1
1116	MF	1
1116	V	2
1117	HM	3
1117	MF	1
1117	TH	1

PIZZA

Pizza Code	Pizza Name	Price
P	Pepperoni	$11.00
MF	Meat Feast	$12.00
V	Vegetarian	$9.00
HM	Ham and Mushroom	$10.00
TH	The Hawaiian	$14.00

Figure PI3.9 Third normal form for data from pizza shop.

Join Process with Tables of 3rd Normal Form for Pizza Orders

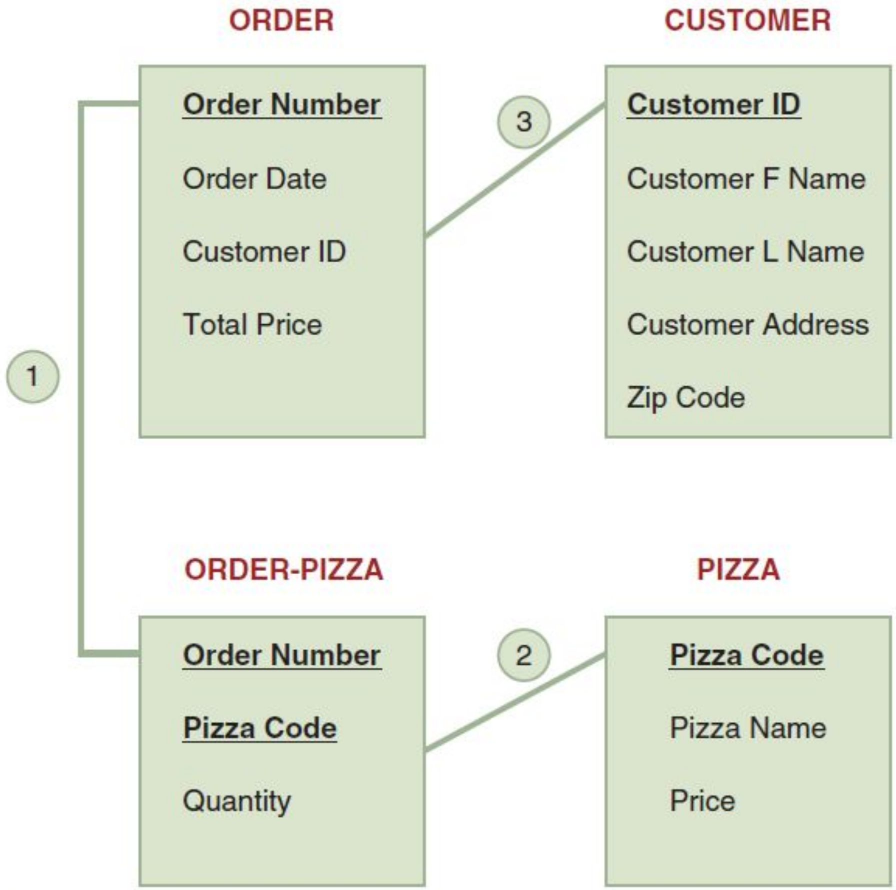
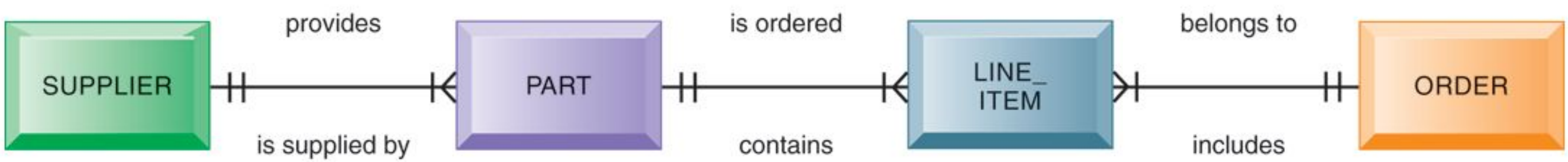


Figure PI3.10 The join process with the tables of third normal form to produce an order.

AN ENTITY-RELATIONSHIP DIAGRAM



This diagram shows the relationships between the entities SUPPLIER, PART, LINE_ITEM, and ORDER that might be used to model the database in Figure 6-10.

The Database Approach to Data Management

- Referential integrity rules
  - Try to ensure relationships between coupled tables remain consistent.
- Entity-relationship diagram
  - Used by database designers to document the data model
  - Illustrates relationships between entities
- Caution: If a business does‘nt get data model right, system won’t be able to serve business well

The Database Approach to Data Management

- Referential integrity rules

- We may not add a new record to the PART table for a part with Supplier_Number 8266 unless there is a corresponding record in the SUPPLIER table for Supplier_Number 8266.

SUPPLIER

Columns (Attributes, Fields)

Supplier_Number	Supplier_Name	Supplier_Street	Supplier_City	Supplier_State	Supplier_Zip
8259	CBM Inc.	74 5th Avenue	Dayton	OH	45220
8261	B. R. Molds	1277 Gandolly Street	Cleveland	OH	49345
8263	Jackson Composites	8233 Micklin Street	Lexington	KY	56723
8444	Bryant Corporation	4315 Mill Drive	Rochester	NY	11344

Key Field (Primary Key)

PART

Part_Number	Part_Name	Unit_Price	Supplier_Number
137	Door latch	22.00	8259
145	Side mirror	12.00	8444
150	Door molding	6.00	8263
152	Door lock	31.00	8259
155	Compressor	54.00	8261
178	Door handle	10.00	8259

Primary Key

Foreign Key

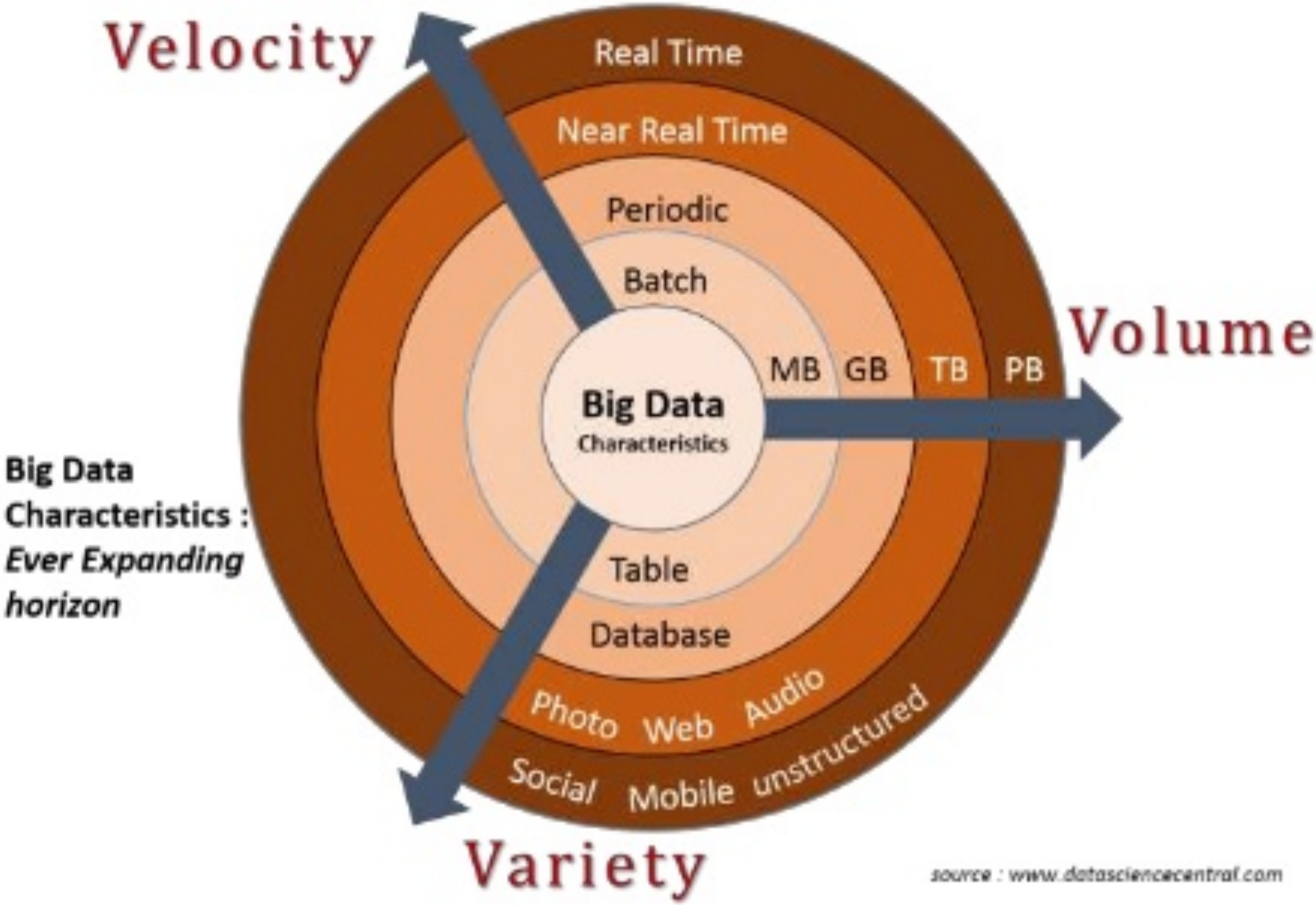
Big Data:

- Big data
  - Massive sets of unstructured/semi-structured data from Web traffic, social media, sensors, and so on
  - Can reveal more patterns and anomalies

Defining Big Data: The Big Data Institute (TBDI)

- Vast Datasets that:
  - Exhibit variety
  - Include structured, unstructured, and semi-structured data
  - Generated at high velocity with an uncertain pattern
  - Do not fit neatly into traditional, structured, relational databases
  - Can be captured, processed, transformed, and analyzed in a reasonable amount of time only by sophisticated information systems.

Characteristics of Big Data



Characteristics of Big Data

- **Volume:** incredible volume of data.
- **Velocity:** The rate at which data flow into an organization is rapidly increasing.
- It is critical because it increases the speed of the feedback loop between a company and its customers.
- **Variety:** Big Data formats change rapidly and can include satellite imagery, broadcast audio streams, digital music files, Web page content.

Issues with Big Data

- Big Data can come from **untrusted sources**.
- **Big Data is dirty:** Dirty data refers to inaccurate, incomplete, incorrect, duplicate, or erroneous data.
- **Big Data changes**:
- Data quality in an analysis can change
- Data itself can change, because the conditions under which the data are captured can change.

Managing Big Data

- Big Data can reveal valuable patterns, trends, and information:
  - tracking the spread of disease
  - tracking crime
  - detecting fraud

Managing Big Data (continued)

- First Step:
  - Integrate information stores into a database environment and develop data warehouses for decision making.
- Second Step:
  - making sense of their growing data.
- Many organizations are turning to NoSQL databases to process Big Data

Putting Big Data to Use

- Making Big Data Available
- Enabling Organizations to Conduct Experiments
- Micro-Segmentation of Customers
- Creating New Business Models
- Organizations Can Analyze Far More Data

Putting Big Data to Use

- **Making Big Data Available**: available for relevant stakeholders can help organizations gain value.
- **Enabling Organizations to Conduct Experiments:**
- For example, Amazon (and many other companies such as Google and LinkedIn) constantly experiments by offering slight different “looks” on its Web site.
- **Micro-Segmentation of Customers:** Segmentation of a company’s customers means dividing them up into groups that share one or more characteristics.

Putting Big Data to Use

- **Creating New Business Models:**

- Companies are able to use Big Data to create new business models.
- For example, a commercial transportation company operated a large fleet of large, long-haul trucks.
- The company recently placed sensors on all its trucks.
- The sensors collect data on vehicle usage (including acceleration, braking, cornering, etc.), driver performance, and vehicle maintenance.
- By analyzing this Big Data, the transportation company was able to improve the condition of its trucks through near-real-time analysis that proactively suggested preventive maintenance.

Putting Big Data to Use

- **Organizations Can Analyze Far More Data:** In some cases, organizations can even process all the data in a population relating to a particular phenomenon, meaning that they do not have to rely as much on sampling.

Data Warehouses and Data Marts

- Describing Data Warehouses and Data Marts
- A Generic Data Warehouse Environment

Data Warehouses and Data Marts

- Most successful companies are those that can respond quickly and flexibly to market changes and opportunities.
- A key to this response is the effective and efficient use of data and information by analysts and managers.

Data Warehouses and Data Marts

- If the manager of a local bookstore wanted to know the profit margin on used books at her store, she could obtain that information from her database, using SQL or QBE.
- However, if she needed to know the trend in the profit margins on used books over the past 10 years, she would have to construct a very complicated SQL or QBE query.

Data Warehouses and Data Marts

- Why organizations are building data warehouses and/or data marts?

1. Bookstore's databases contain the necessary information to answer the manager's query, but this information is not organized in a way that makes it easy for her to find what she needs.
2. Organization's databases are designed to process millions of transactions every day. Therefore, complicated queries might take a long time to answer, and degrade the performance of the databases.
3. Transactional databases are designed to be updated. This update process requires extra processing.
   Data warehouses and data marts are read-only, and the extra processing is eliminated because data already contained in the data warehouse are not updated.
4. Transactional databases are designed to access a single/limited record at a time.
   Data warehouses are designed to access large groups of related records.

Describing Data Warehouses

joined

- A data warehouse possesses consolidated historical data, which helps the organization to analyze its business.
- It helps executives to organize, understand, and use their data to take strategic decisions.
- No frequent updating done in a data warehouse.
- A data warehouse is a database, which is kept separate from the organization's operational database.

Describing Data Marts

- Because data warehouses are so expensive, they are used primarily by large companies.
- A data mart is a low-cost, scaled-down version of a data warehouse that is designed for the end-user needs in a strategic business unit (SBU) or an individual department.
- Data marts can be implemented more quickly than data warehouses,

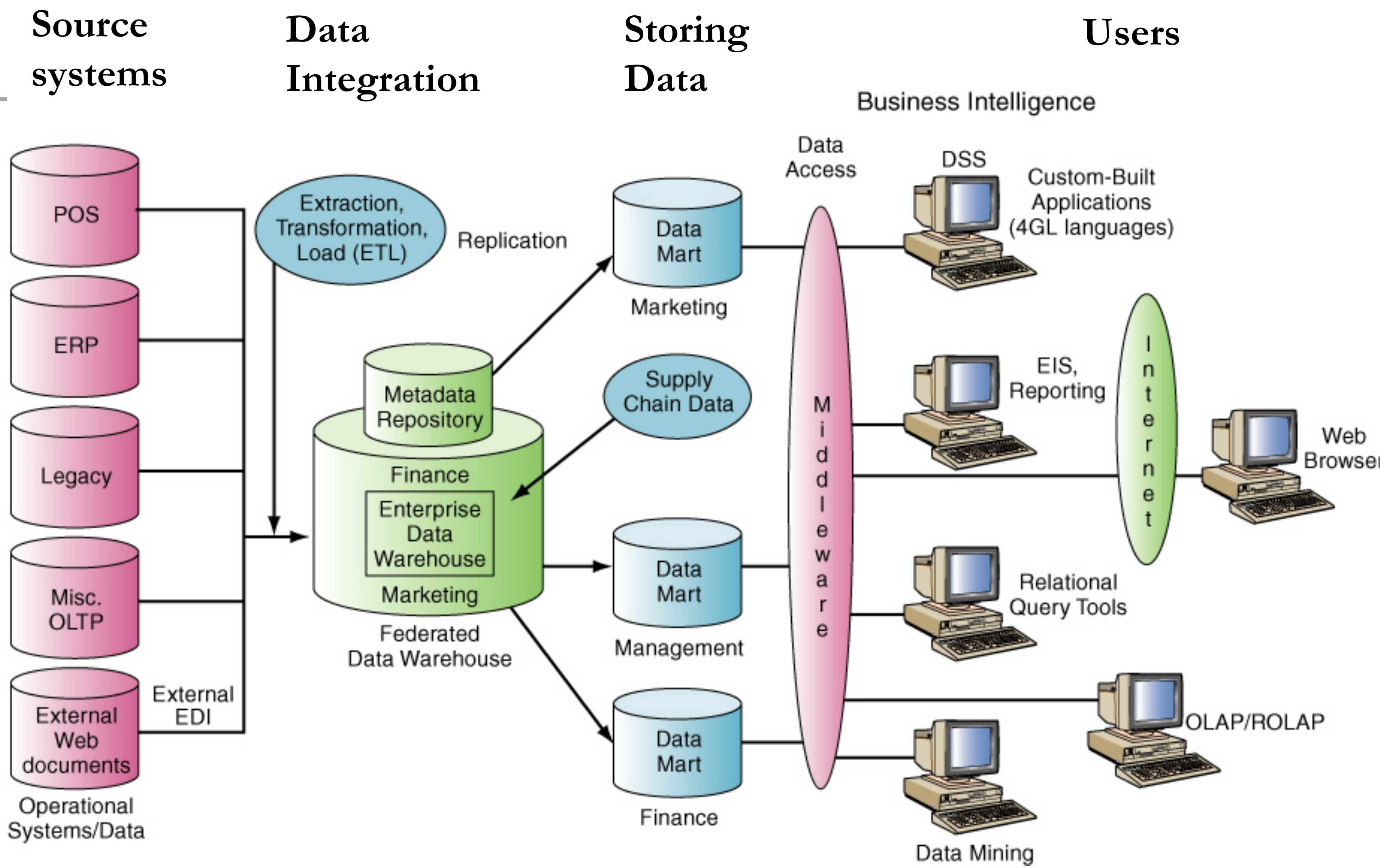
Basic Characteristics of Data Warehouses and Data Marts:

- Organized by business dimension
- Use online analytical processing (OLAP)
- Integrated
- Time variant
- Nonvolatile
- Multidimensional

Basic Characteristics of Data Warehouses and Data Marts:

- **Organized by business dimension or subject** - Data are organized by subject. For example, by customer, vendor, product, price level, and region whereas in transactional systems, data are organized by business process, such as order entry, inventory control, and accounts receivable.
- **Use online analytical processing (OLAP):** used for decision making
- **Integrated** - Data are collected from multiple systems and then integrated around subjects.
- **Time variant** - Data warehouses and data marts maintain historical data (i.e., data that include time as a variable).
- **Nonvolatile**: users cannot change or update the data.
- **Multidimensional** - uses a multidimensional data structure. relational databases store data in two-dimensional tables.

Data Warehouse Framework



A Generic Data Warehouse Environment

- **Source Systems:** Systems that provide a source of organizational data.

source systems that provide data to the data warehouse or datamart.

- **Data Integration:** reflects the growing number of ways that source system data can be handled. Typically organizations need to Extract, Transform, and Load data from source system into a data warehouse or data mart.
- **Storing the Data:** A variety of architectures can be used to store decision-support data and the most common architecture is one central enterprise data warehouse, without data marts.

Some organizations use one big storage place for all their decision-making data (central data warehouse) and don't use smaller specialized storage areas (Data Marts). This can make things simpler, ensure everyone uses the same data, and save money. However, some organizations do use both, depending on their specific needs.

A Generic Data Warehouse Environment

- **Metadata:** data about the data within the data warehouse. (e.g., database, table, and column names; refresh schedules, source system)
- **Data Quality:** quality of the data in the warehouse must meet users' needs. If it does not, users will not trust the data and ultimately will not use it. Some of the data can be improved with data-cleansing soft ware, but the better, long-term solution is to improve the quality at the source system level.
- **Governance:** To ensure that BI is meeting their needs, organizations must implement governance to plan and control their BI activities. Governance requires that people, committees, and processes be in place.
- **Users:** There are many potential BI users, including IT developers; frontline workers; analysts; information workers; managers and executives; and suppliers, customers, and regulators.

Relational Databases

Company manufactures four products—nuts, screws, bolts, and washers
Sold them in three territories—East, West, and Central
For previous three years— 2011, 2012, and 2013.

(a) 2012

Product	Region	Sales
Nuts	East	50
Nuts	West	60
Nuts	Central	100
Screws	East	40
Screws	West	70
Screws	Central	80
Bolts	East	90
Bolts	West	120
Bolts	Central	140
Washers	East	20
Washers	West	10
Washers	Central	30

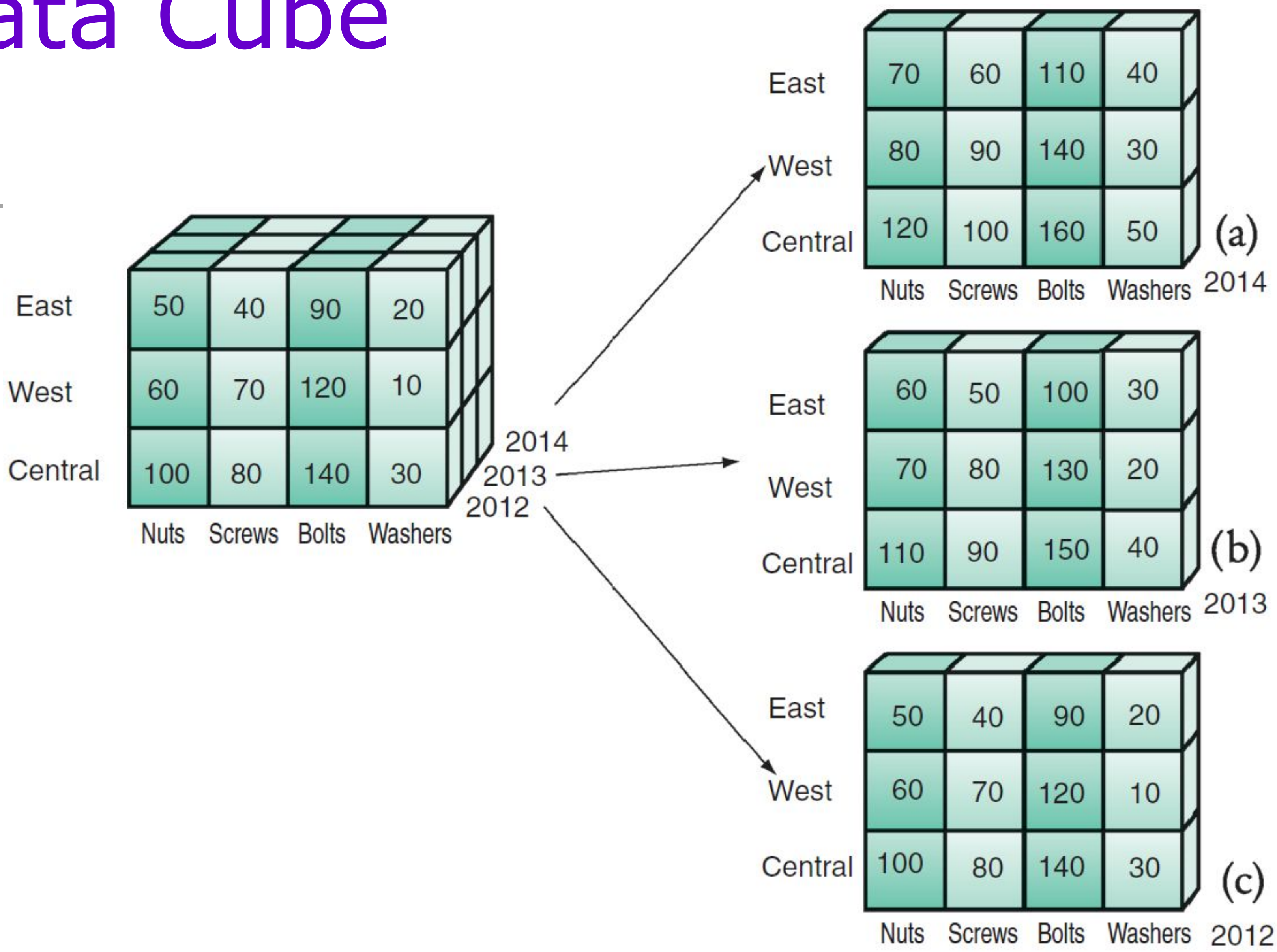
(b) 2013

Product	Region	Sales
Nuts	East	60
Nuts	West	70
Nuts	Central	110
Screws	East	50
Screws	West	80
Screws	Central	90
Bolts	East	100
Bolts	West	130
Bolts	Central	150
Washers	East	30
Washers	West	20
Washers	Central	40

(c) 2014

Product	Region	Sales
Nuts	East	70
Nuts	West	80
Nuts	Central	120
Screws	East	60
Screws	West	90
Screws	Central	100
Bolts	East	110
Bolts	West	140
Bolts	Central	160
Washers	East	40
Washers	West	30
Washers	Central	50

Data Cube

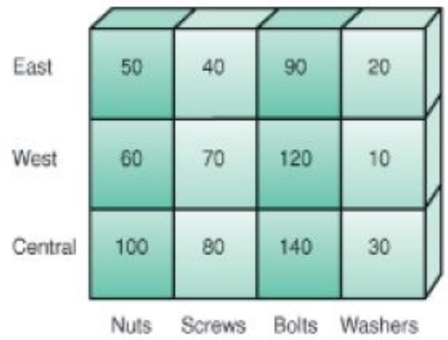


East
West
Central
50
40
90
20
60
70
120
10
100
80
140
30
Nuts
Screws
Bolts
Washers
2014
2013
2012
East
West
Central
70
60
110
40
80
90
140
30
120
100
160
50
Nuts
Screws
Bolts
Washers
(a)
2014
East
West
Central
60
50
100
30
70
80
130
20
110
90
150
40
Nuts
Screws
Bolts
Washers
(b)
2013
East
West
Central
50
40
90
20
60
70
120
10
100
80
140
30
Nuts
Screws
Bolts
Washers
(c)
2012

Equivalence Between Relational and Multidimensional Databases

Product	Region	Sales
Nuts	East	50
Nuts	West	60
Nuts	Central	100
Screws	East	40
Screws	West	70
Screws	Central	80
Bolts	East	90
Bolts	West	120
Bolts	Central	140
Washers	East	20
Washers	West	10
Washers	Central	30

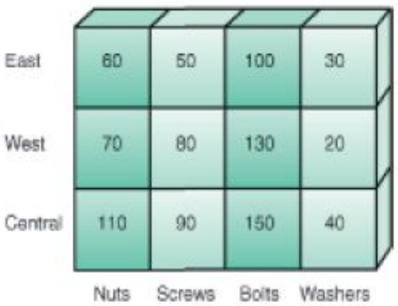
2012



(a)
2012

Product	Region	Sales
Nuts	East	60
Nuts	West	70
Nuts	Central	110
Screws	East	50
Screws	West	80
Screws	Central	90
Bolts	East	100
Bolts	West	130
Bolts	Central	150
Washers	East	30
Washers	West	20
Washers	Central	40

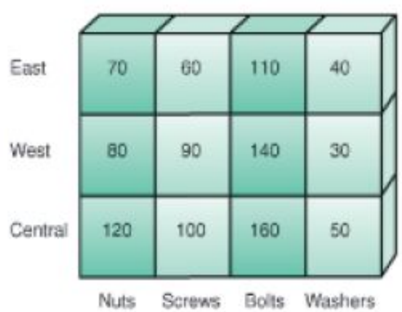
2013



(b)
2013

Product	Region	Sales
Nuts	East	70
Nuts	West	80
Nuts	Central	120
Screws	East	60
Screws	West	90
Screws	Central	100
Bolts	East	110
Bolts	West	140
Bolts	Central	160
Washers	East	40
Washers	West	30
Washers	Central	50

2014



(c)
2014

IT'S ABOUT BUSINESS 3.4

Data Warehouse Gives Nordea Bank a Single Version of the Truth

© halbergman/iStockphoto

1. What are other advantages (not mentioned in the case) that Nordea Bank might realize from its data warehouse?
2. What recommendations would you give to Nordea Bank about incorporating Big Data into their bank's data management? Provide specific examples of what types of Big Data you think Nordea should consider.

Knowledge Management

Knowledge

- Knowledge is very distinct from data and information and provides a higher level of meaning about that data and information.
- The **ability to act** is an integral part of being knowledgeable.

- Data are a collection of:
  - Facts
  - Measurements
  - Statistics
- Information is organized or processed data that are:
  - Timely
  - Accurate
- Knowledge is information that is:
  - Contextual
  - Relevant
  - Actionable
- Having knowledge implies that it can be exercised to solve a problem, whereas having information does not.

Knowledge Positioning

Dimensions of knowledge:

Every organization has four dimensions of knowledge:

- Knowledge is a firm asset.
- Knowledge has different forms.
- Knowledge has a location.
- Knowledge is situational.

Dimensions of knowledge:

- Knowledge is a firm asset.
  - Intangible asset   it doesn't have a physical form
  - Creation of knowledge from data, information, requires organizational resources such as time, expertise, and technology.
  - Knowledge does not diminish as physical assets but increases as more people share it.
- Knowledge has different forms.
  - Can be *explicit* (documented) or *tacit* (residing in minds)
  - Knowledge involves Know-how, craft, skill
  - Knowledge means knowing How to follow procedure
  - Knowing why things happen (causality)

Dimensions of knowledge:

knowledge is not a one-size-fits-all concept. It has a specific location within an organization, is the result of cognitive processes, and can be both an individual and a collective endeavor. Moreover, knowledge is closely tied to the culture and context of the organization, making it "sticky" and situation-dependent.

- Knowledge has a location.
  - Knowledge Cognitive event
  - It ban be both social and individual
  - "Sticky" (hard to move), situated (enmeshed in firm's culture), contextual (works only in certain situations)
- Knowledge is situational.
  - Conditional: Knowing when to apply procedure
  - Contextual: Knowing how to use certain tool under what circumstances.

Explicit knowledge

- **Explicit knowledge** deals with more objective, rational, and technical knowledge.
- In an organization, explicit knowledge consists of the policies, procedural guides, reports, products, strategies, goals, core competencies, and IT infrastructure of the enterprise.
- In other words, explicit knowledge is the knowledge that has been codified (documented) in a form that can be distributed to others or transformed into a process or a strategy.
- A description of how to process a job application that is documented in a firm's human resources policy manual is an example of explicit knowledge.

Knowledge – Explicit knowledge

- Explicit knowledge has been codified (documented) in a form that can be distributed to others or transformed into a process or strategy without requiring interpersonal interaction.
- **Explicit knowledge:**
  - Data
  - Policies
  - Procedures
  - Software
  - Documents
  - Products
  - Strategies
  - Goals

Tacit knowledge

- **tacit knowledge** is the cumulative store of subjective or experiential learning.
- In an organization, tacit knowledge consists of an organization's experiences, insights, expertise,know-how, trade secrets, skill sets, understanding, and learning.
- It also includes the organizational culture, which refl ects the past and present experiences of the organization's people and processes, as well as the organization's prevailing values.
- Tacit knowledge is generally imprecise and costly to transfer. It is also highly personal.
- Finally, because it is unstructured, it is difficult to formalize or codify, in contrast to explicit knowledge.

Knowledge – Tacit knowledge

- Tacit knowledge is usually in the domain of subjective, cognitive, and experiential learning; it is highly personal and difficult to formalize.
- It is also referred to as embedded knowledge.
  - Mental maps
  - Insights
  - Expertise
  - Know-how
  - Trade secrets
  - Skill sets
  - The organizational culture

Knowledge Management (KM) is a process used by organizations to handle and make use of important knowledge that is like the organization's memory. This knowledge often exists in a form that's not well-organized or structured. KM aims to organize and utilize this knowledge effectively.

Knowledge Management Systems

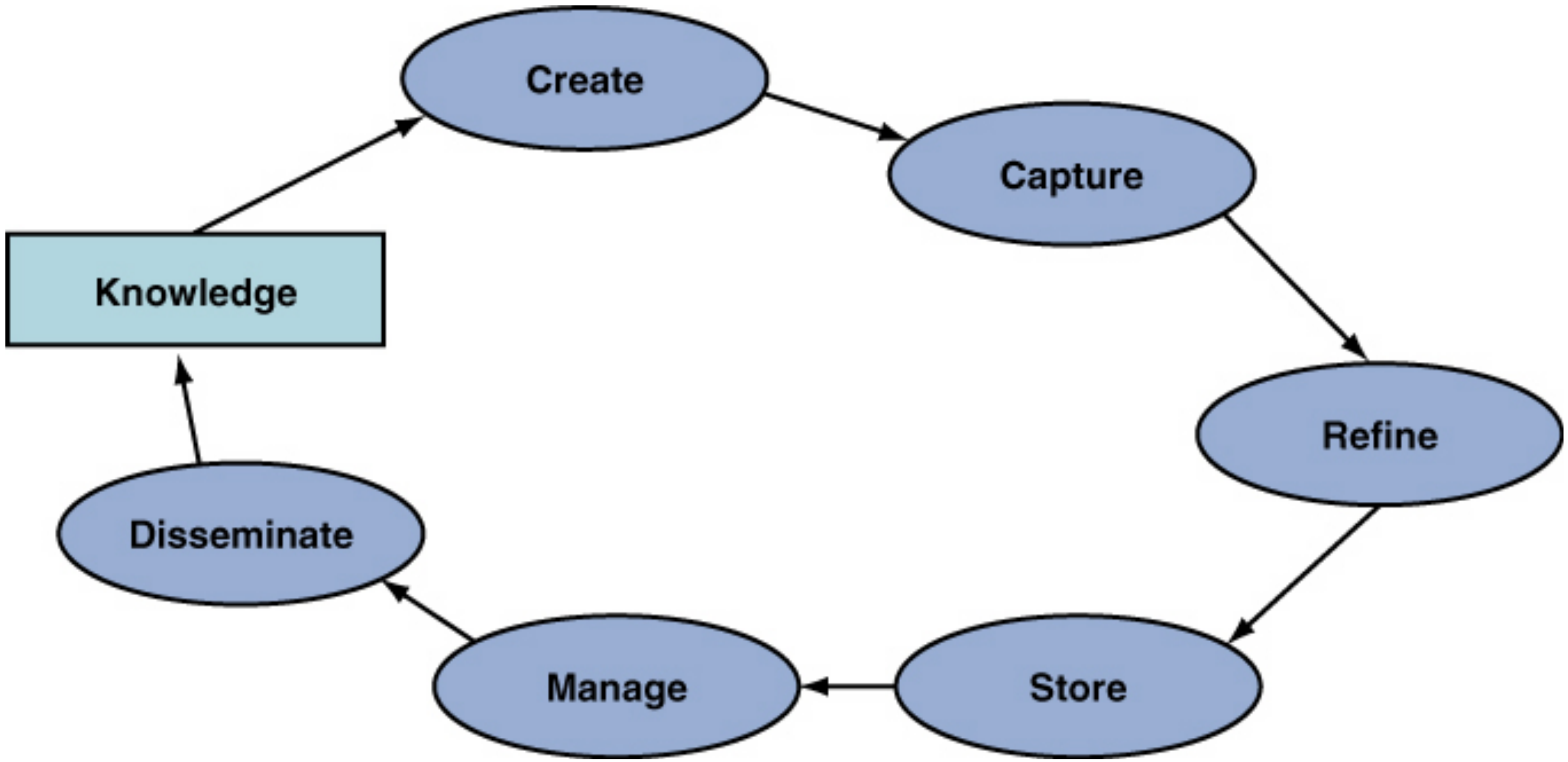
- Organizations need to integrate both explicit and tacit knowledge into a formal information systems - Knowledge Management System (KMS)
- KMS means Set of business processes developed in an organization to create, store, transfer, and apply knowledge
- KMSs are intended to help an organization cope with turnover, rapid change, and downsizing by making the expertise of the organization's human capital widely accessible.

Knowledge Management Systems aim to help an organization make the most optimum use of the knowledge that has been aggregated.

Benefits with KMSs

- best-practice knowledge improves overall organizational performance.
- For example, account managers can now make available their tacit knowledge
  about how best to manage large accounts. The organization can then utilize this knowledge when it trains new account managers.
- Other benefits include improved customer service, more efficient product development, and improved employee morale and retention.
- organizations must create a knowledge management culture that rewards employees who add their expertise to the knowledge base.
- organization must continually maintain and upgrade its knowledge base. Specifically, it must incorporate new knowledge and
  delete old, outdated knowledge.
- Finally, companies must be willing to invest in the resources needed to carry out these operations.

Knowledge Management System Cycle



Knowledge Management System Cycle

- **The KMS Cycle Consists of Six Steps:**

1. **Create knowledge:** Knowledge is created as people determine new ways of doing things or develop know-how. Sometimes external knowledge is brought in.
2. **Capture knowledge:** New knowledge must be identified as valuable and be represented in a reasonable way. standard way
3. **Refine knowledge:** New knowledge must be placed in context so that it is actionable. This is where tacit qualities (human insights) must be captured along with explicit facts.

To make new knowledge useful, it must be put into context so that people can use it effectively. Combining these elements is essential for turning information into action.

The Knowledge Management System Cycle

4. **Store knowledge:** Useful knowledge must then be stored in a reasonable format in a knowledge repository so that other people in the organization can access it.

5. **Manage knowledge:** Like a library, the knowledge must be kept current. It must be reviewed regularly to verify that it is relevant and accurate.

The knowledge in the repository must be continuously managed and updated. It must always have current knowledge

6. **Disseminate knowledge:** Knowledge must be made available in a useful format to anyone in the organization who needs it, anywhere and anytime.

Some examples of knowledge management systems are:
The company employees might be working on some project. So, all the files and information related to the project is shared amonget the team members. This will help every member to be updated what the other one is doing.

Knowledge management value chain

Each stage adds value to raw data and information as they are transformed into usable knowledge

1. Knowledge acquisition
2. Knowledge storage
3. Knowledge dissemination
4. Knowledge application

Knowledge management value chain

1. Knowledge acquisition
    - Documenting tacit and explicit knowledge
        - Storing documents, reports, presentations, best practices
        - Unstructured documents (e.g., e-mails)
        - Developing online expert networks
    - Creating knowledge
    - Tracking data from internal and external sources

Knowledge management value chain

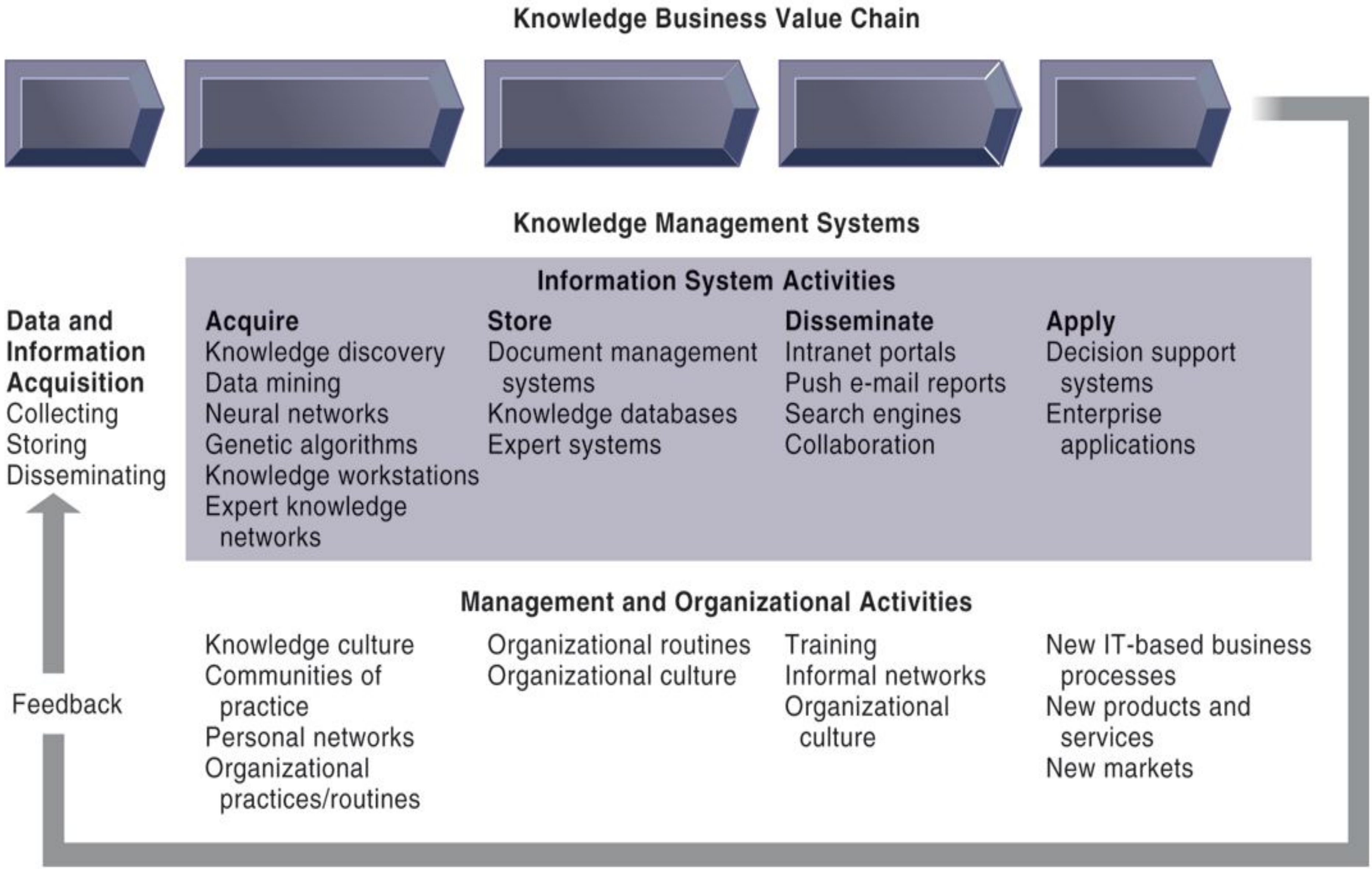
2. Knowledge storage
   - Databases
   - Document management systems
   - Role of management:
     - Support development of planned knowledge storage systems.
     - Encourage development of corporate-wide schemas for indexing documents.
     - Reward employees for taking time to update and store documents properly.

Knowledge management value chain

3. Knowledge dissemination
   - Portals, wikis
   - E-mail, instant messaging
   - Search engines
   - Collaboration tools
   - A deluge of information?
     - Training programs, informal networks, and shared management experience help managers focus attention on important information.

Knowledge management value chain

4. Knowledge application
   - To provide return on investment, organizational knowledge must become systematic part of management decision making and become situated in decision-support systems.
     - New business practices
     - New products and services
     - New markets



Knowledge management today involves both information systems activities and a host of enabling management and organizational activities.