CHAPTER 3

Data and Knowledge Management

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

>>>

1. Discuss ways that common challenges in managing data can be addressed using data governance.
2. Discuss the advantages and disadvantages of relational databases.
3. Define Big Data, and discuss its basic characteristics.

>>>

4. Recognize the necessary environment to successfully implement and maintain data warehouses.
5. Describe the benefits and challenges of implementing knowledge management systems in organizations.

OPENING >

- **Flurry Gathers Data from Smartphone Users**

1. Do you feel that Flurry should be installed on your smartphone by various app makers without your consent? Why or why not? Support your answer.
2. What problems would Flurry encounter if someone other than the smartphone's owner uses the device? (Hint: Note how Flurry gathers data.)
3. Can Flurry survive the privacy concerns that are being raised about its business model?

3.1 Managing Data

- Difficulties of Managing Data
- Data Governance

The 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)
- New sources of data

The Difficulties of Managing Data (continued)

- Data Degradation
- Data Rot
- Data security, quality, and integrity are critical
- Legal requirements change frequently and differ among countries & industries

IT'S ABOUT BUSINESS 3.1

- **New York City Opens Its Data to All**


© mura/iStockphoto

1. What are some other creative applications addressing city problems that could be developed using NYC's open data policy?
2. List some disadvantages of providing all city data in an open, accessible format.

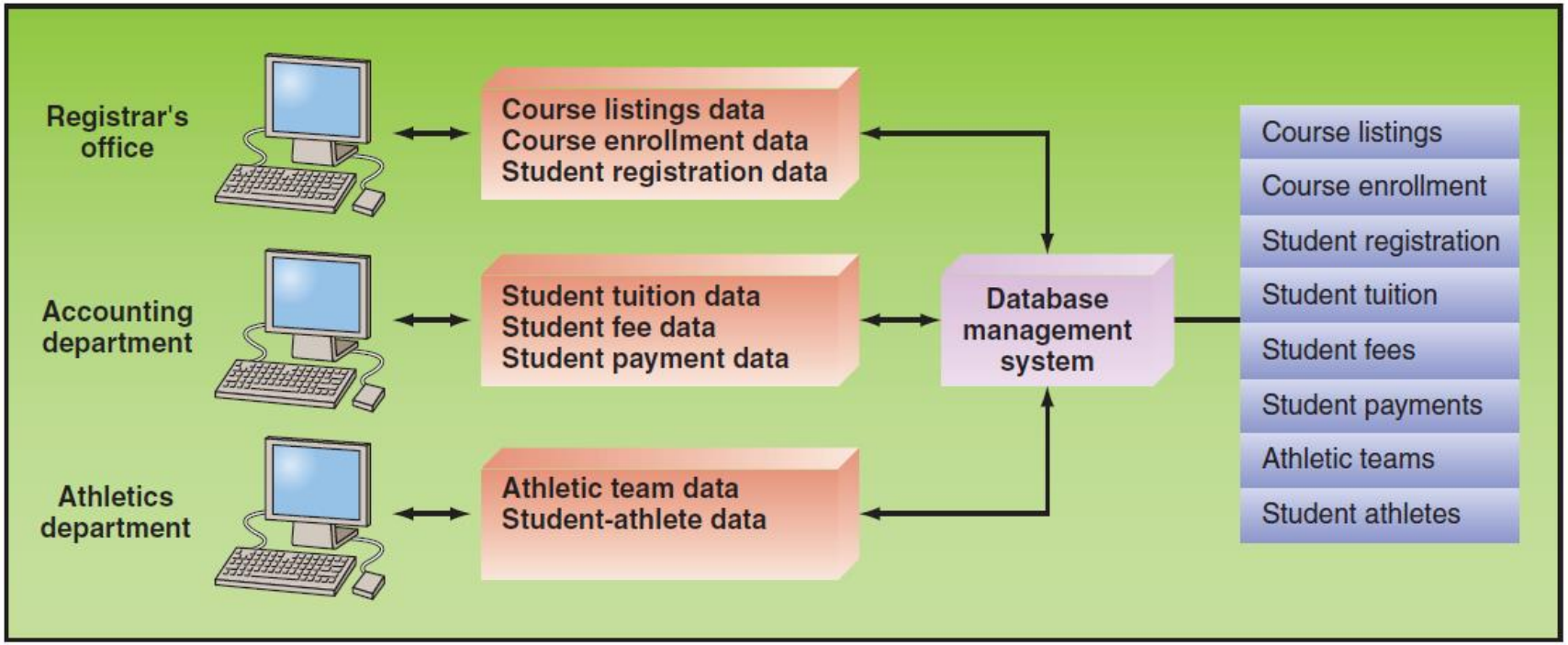
Data Governance

- Master Data Management
- Master Data

3.2 The Database Approach

- Data File
- Database Systems Minimize & Maximize Three Things
- The Data Hierarchy
- The Relational Database Model

Figure 3.1: Database Management System



Database Management Systems (DBMS) Minimize:

- Data Redundancy
- Data Isolation
- Data Inconsistency

Database Management Systems (DBMS) Maximize:

- Data Security
- Data Integrity
- Data Independence

IT'S ABOUT BUSINESS 3.2

- **Google's Knowledge Graph**
  1. Refer to the definition of a relational database. In what way can the Knowledge Graph be considered a database? Provide specific examples to support your answer.
  2. Refer to the definition of an expert system in Plug IT In 5. Could the Knowledge Graph be considered an expert system? If so, provide a specific example to support your answer.
  3. What are the advantages of the Knowledge Graph over traditional Google searches?

Data Hierarchy

- Bit
- Byte
- Field
- Record
- Data File (Table)
- Database

Figure 3.2: Hierarchy of Data for a Computer-Based File

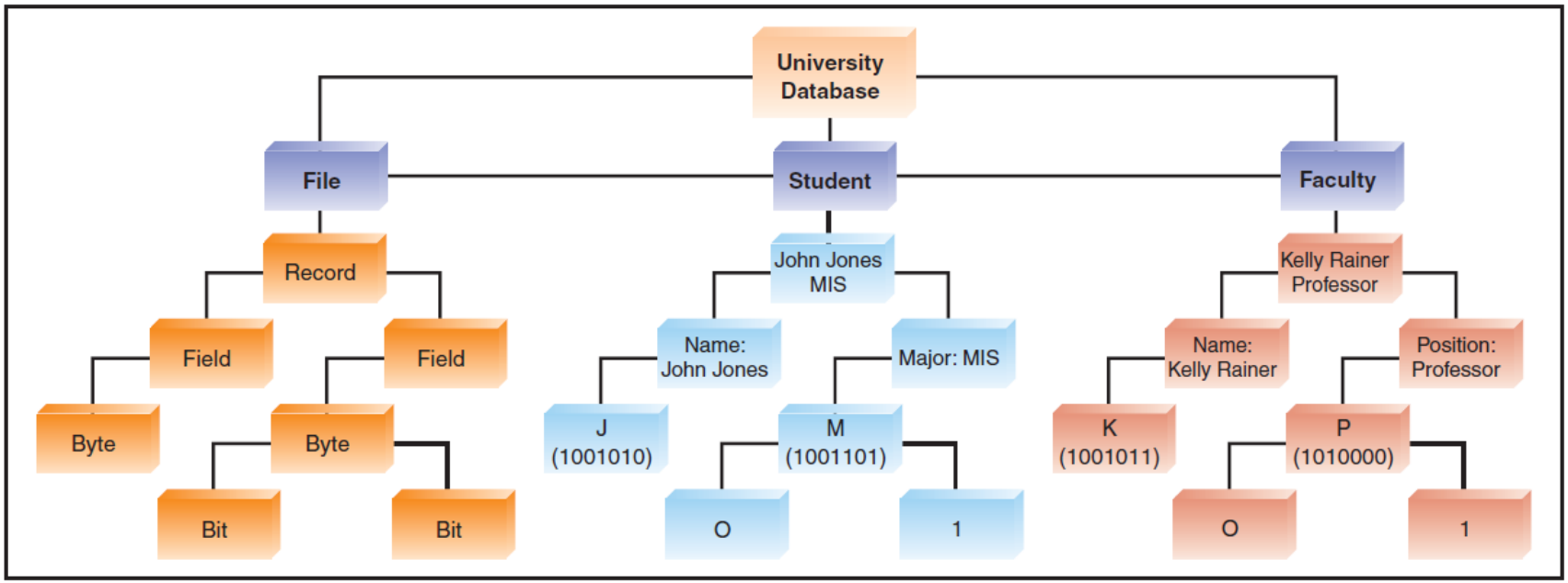


Figure 3.2 Hierarchy of data for a computer-based file.

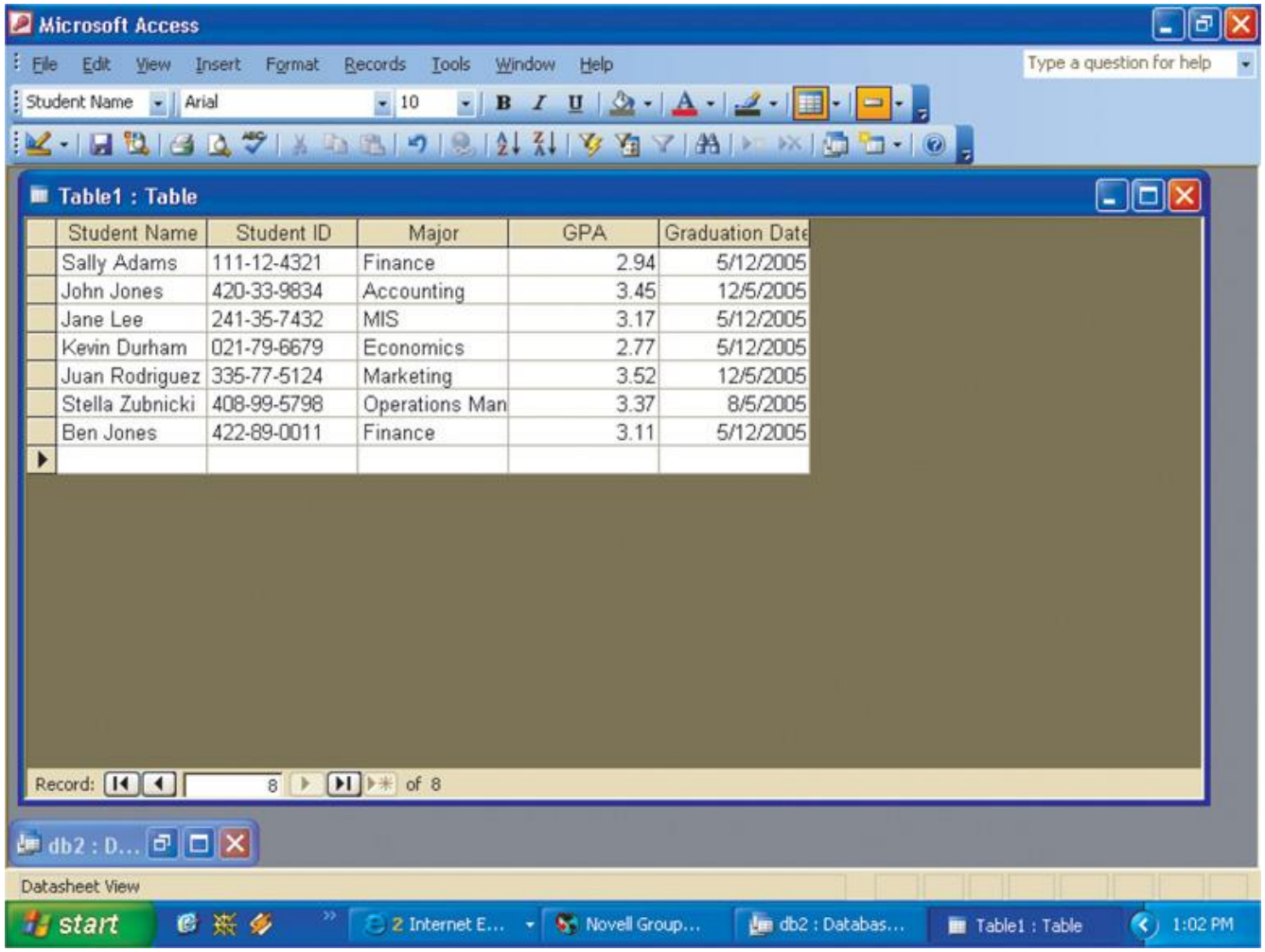
The Relational Database Model

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

The Relational Database Model (continued)

- Primary Key
- Secondary Key
- Foreign Key

Figure 3.3: Student Database Example



3.3 Big Data

- Defining Big Data
- Characteristics of Big Data
- Issues with Big Data
- Managing Big Data
- Putting Big Data to Use

Defining Big Data

- **Gartner (www.gartner.com)**
- **Big Data Institute**

Defining Big Data: Gartner

- Diverse, high volume, high-velocity information assets that require new forms of processing to enable enhanced decision making, insight discovery, and process optimization.

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

- Volume
- Velocity
- Variety

Issues with Big Data

- Untrusted data sources
- Big Data is dirty
- Big Data changes, especially in data streams

Managing Big Data

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

Managing Big Data (continued)

- First Step:
  - Integrate information silos into a database environment and develop data warehouses for decision making.
- Second Step:
  - making sense of their proliferating data.

Managing Big Data (continued)

- Many organizations are turning to NoSQL databases to process Big Data

IT'S ABOUT BUSINESS 3.3

- **The MetLife Wall**
  1. Describe the problems that MetLife was experiencing with customer data before it implemented the MetLife Wall.
  2. Describe how these problems originated.

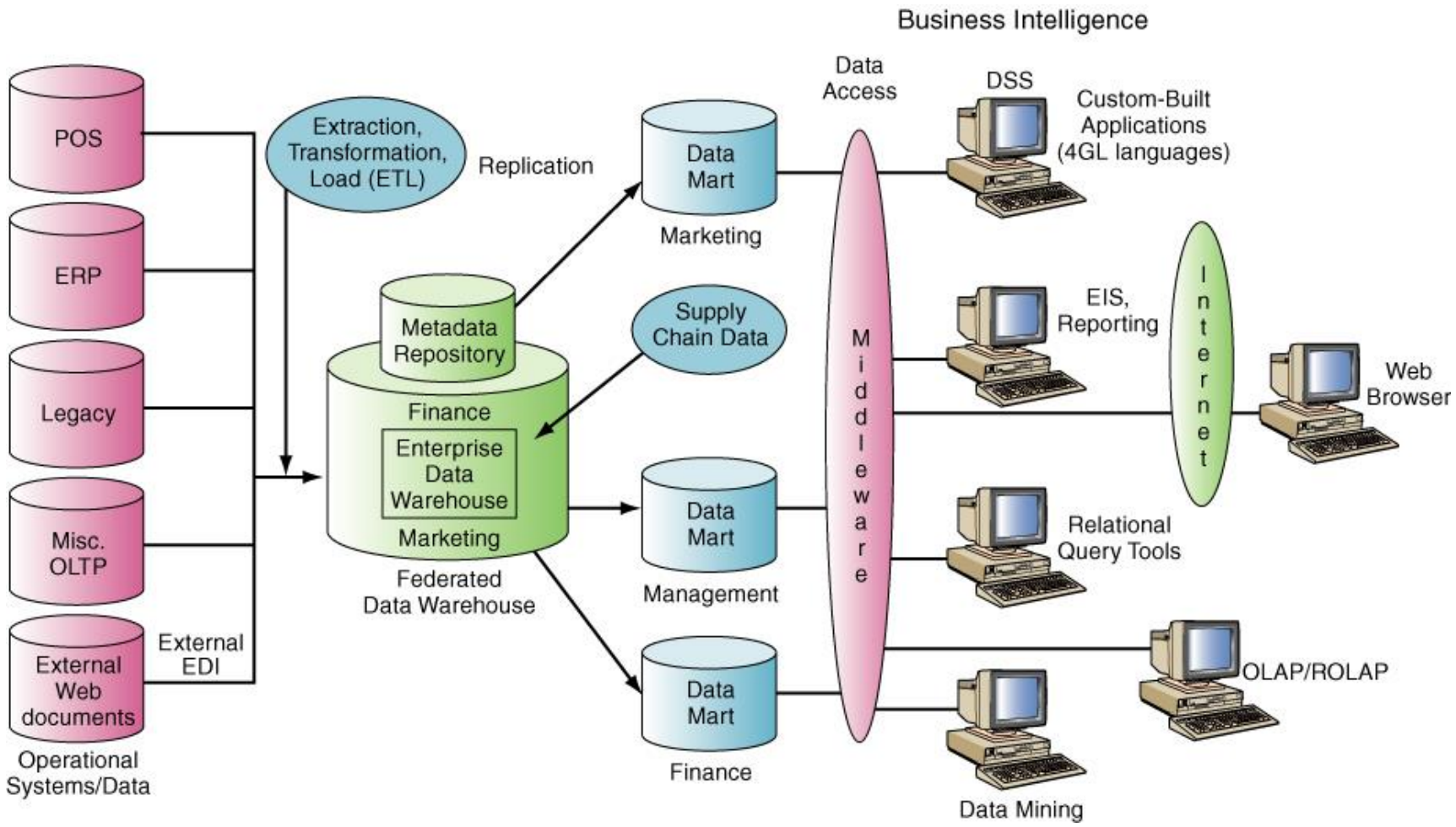
Putting Big Data to Use

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

3.4 Data Warehouses and Data Marts

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

Figure 3.4: Data Warehouse Framework



Describing Data Warehouses and Data Marts

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

A Generic Data Warehouse Environment

- Source Systems
- Data Integration
- Storing the Data
- Metadata
- Data Quality
- Governance
- Users

Figure 3.5: Relational Databases

(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

Figure 3.6: Data Cube

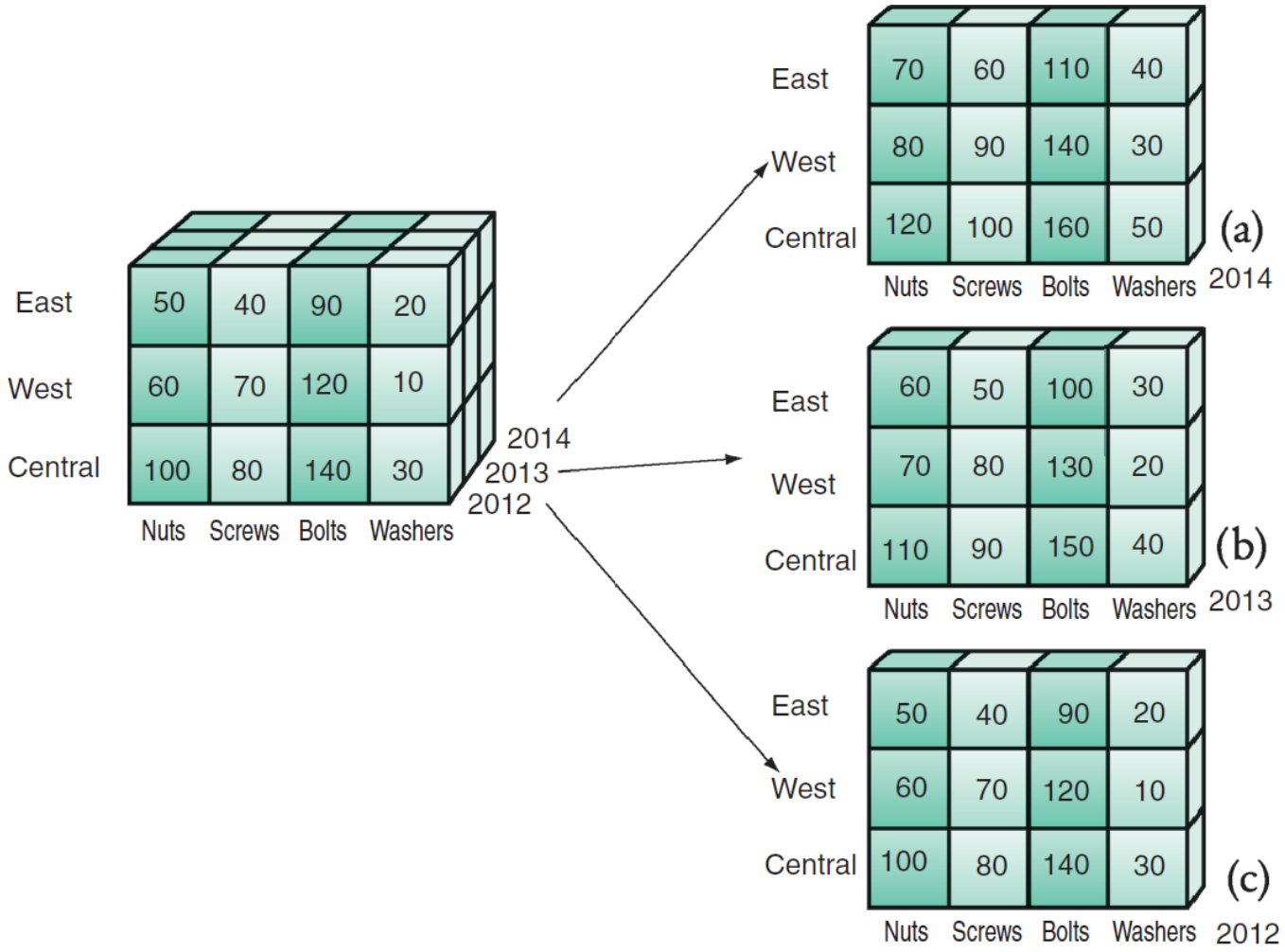
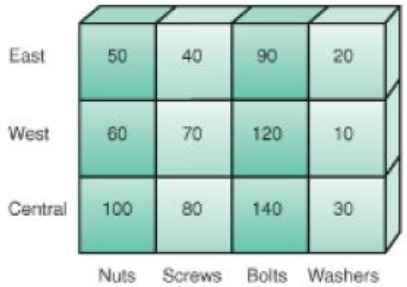


Figure 3.7: Equivalence Between Relational and Multidimensional Databases

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

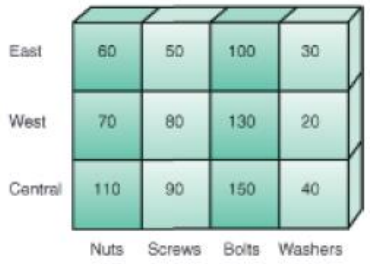
2012



(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

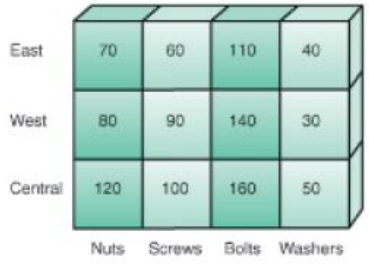
2013



(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.

3.5 Knowledge Management

- Concepts and Definitions
- Knowledge Management Systems
- The KMS Cycle

Concepts and Definitions

- Knowledge Management
- Knowledge
- Explicit and Tacit Knowledge
- Knowledge Management Systems
- The KMS Cycle

Figure 3.8: The Knowledge Management System Cycle