Case Study for Information Management
資訊管理個案
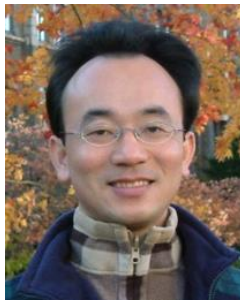
Foundations of Business Intelligence: IBM and Big Data (Chap. 6)

Min-Yuh Day
戴敏育
Assistant Professor
專任助理教授
Dept. of Information Management, Tamkang University
淡江大學 資訊管理學系
http://mail. tku.edu.tw/myday/
2015-10-27

課程大綱 (Syllabus)

週次 (Week)　日期 (Date)　內容 (Subject/Topics)

1　2015/09/15, 17　Introduction to Case Study for Information Management

2　2015/09/22, 24　Information Systems in Global Business: UPS (Chap. 1) (pp.53-54)

3　2015/09/29, 10/01　Global E-Business and Collaboration: P&G (Chap. 2) (pp.84-85)

4　2015/10/06, 08　Information Systems, Organization, and Strategy: Starbucks (Chap. 3) (pp.129-130)

5　2015/10/13, 15　Ethical and Social Issues in Information Systems: Facebook (Chap. 4) (pp.188-190)

課程大綱 (Syllabus)

週次 (Week)　日期 (Date)　內容 (Subject/Topics)

6　2015/10/20, 22　IT Infrastructure and Emerging Technologies:
　　　　　　　Amazon and Cloud Computing
　　　　　　　(Chap. 5) (pp. 234-236)

7　2015/10/27, 29　Foundations of Business Intelligence:
　　　　　　　IBM and Big Data (Chap. 6) (pp.261-262)

8　2015/11/03, 05　Telecommunications, the Internet, and Wireless
　　　　　　　Technology: Google, Apple, and Microsoft
　　　　　　　(Chap. 7) (pp.318-320)

9　2015/11/10, 12　Midterm Report (期中報告)

10　2015/11/17, 19　期中考試週

課程大綱 (Syllabus)

週次　　日期　　　內容（Subject/Topics）

11　2015/11/24, 26　Enterprise Applications: Summit and SAP
　　　　　　　　　　(Chap. 9) (pp.396-398)

12　2015/12/01, 03　E-commerce: Zagat　(Chap. 10) (pp.443-445)

13　2015/12/08, 10　Enhancing Decision Making: Zynga
　　　　　　　　　　(Chap. 12) (pp.512-514)

14　2015/12/15, 17　Building Information Systems: USAA
　　　　　　　　　　(Chap. 13) (pp.547-548)

15　2015/12/22, 24　Managing Projects: NYCAPS and CityTime
　　　　　　　　　　(Chap. 14) (pp.586-588)

16　2015/12/29, 31　Final Report I (期末報告 I)

17　2016/01/05, 07　Final Report II (期末報告 II)

18　2016/01/12, 14　期末考試週

Chap. 6
Foundations of
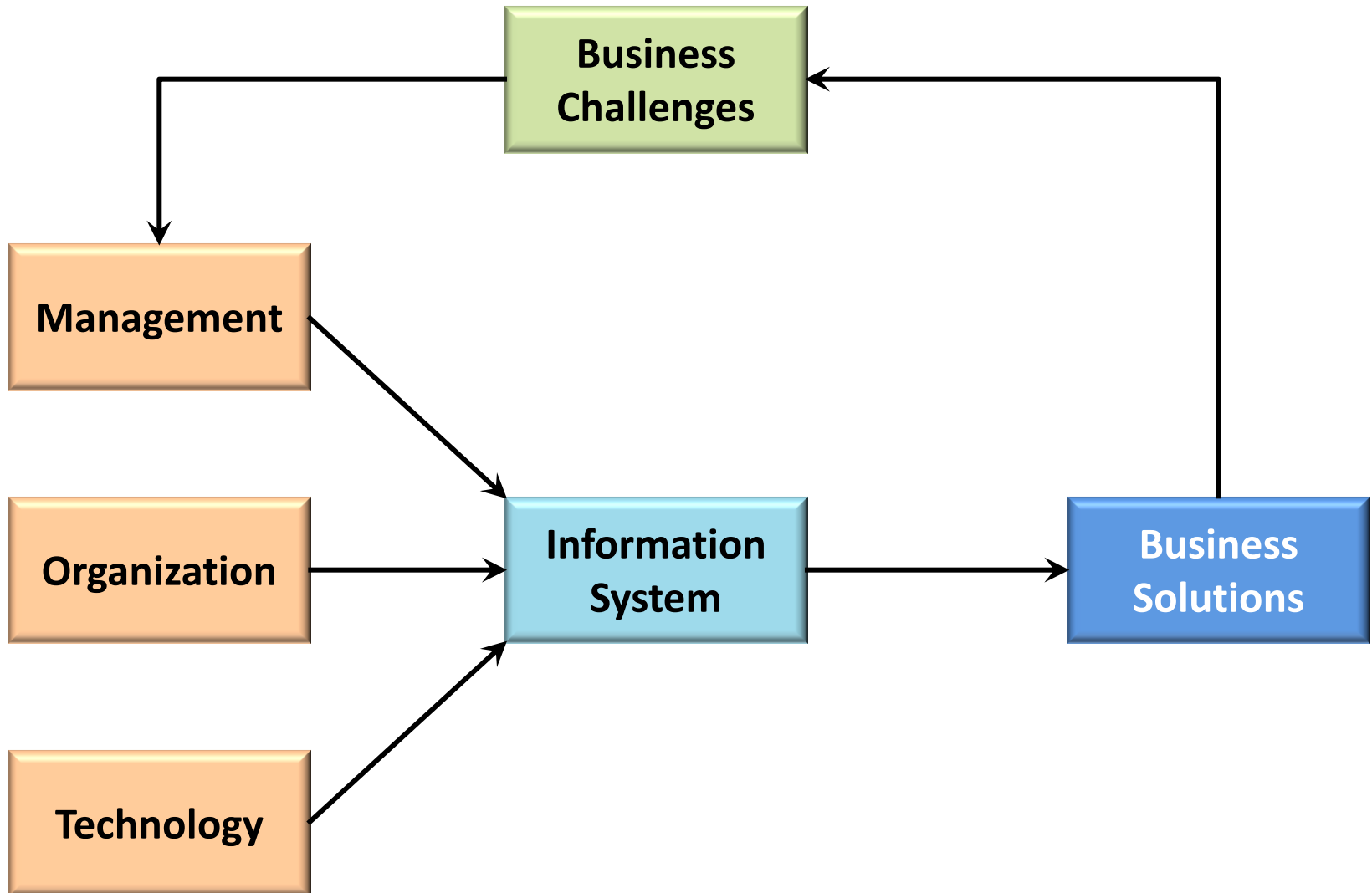# Business Intelligence:
IBM and Big Data

Case Study:

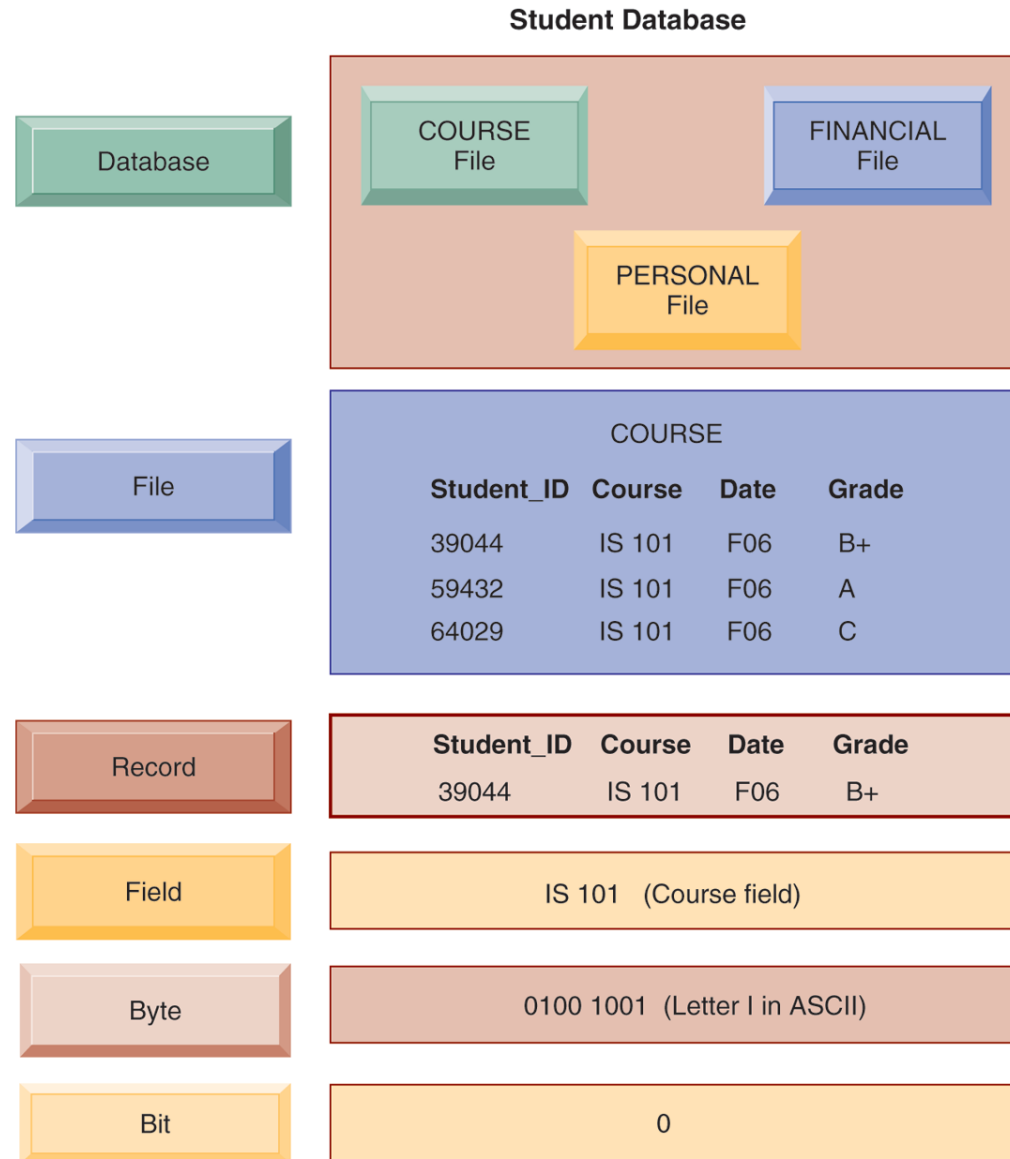## IBM and Big Data (Chap. 6) (pp. 261-262)
Interactive Session: Technology: Big Data, Big Rewards

1. Describe the kinds of "big data" collected by the organizations described in this case.

2. List and describe the business intelligence technologies described in this case.

3. Why did the companies described in this case need to maintain and analyze big data? What business benefits did they obtain?

4. Identify three decisions that were improved by using "big data."

5. What kinds of organizations are most likely to need "big data" management and analytical tools? Why?
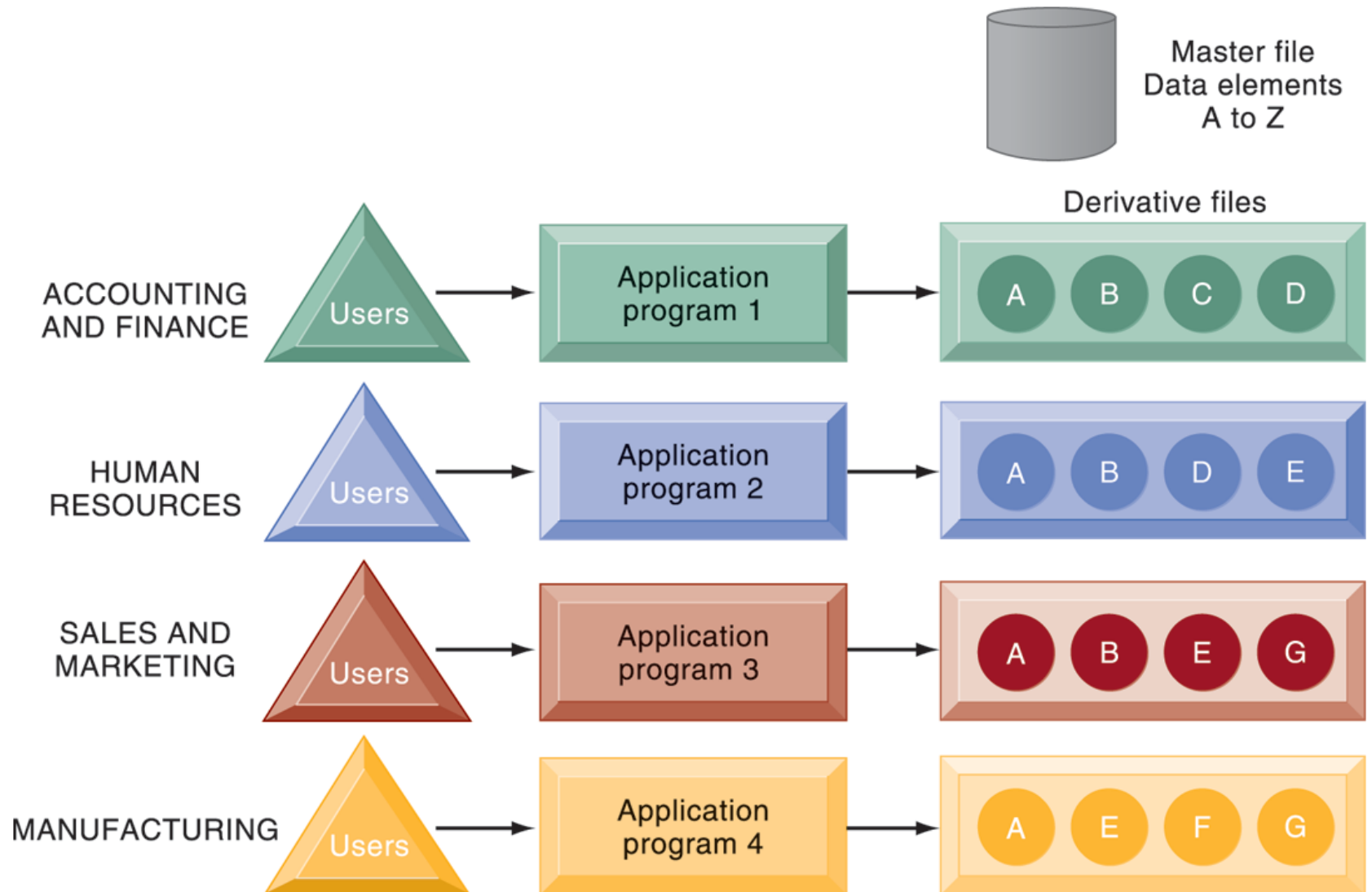
Overview of Fundamental MIS Concepts

Source: Kenneth C. Laudon & Jane P. Laudon (2014), Management Information Systems: Managing the Digital Firm, Thirteenth Edition, Pearson.

THE DATA HIERARCHY

Student Database

Database	COURSE File · FINANCIAL File · PERSONAL File

COURSE

	Student_ID	Course	Date	Grade
File	39044	IS 101	F06	B+
	59432	IS 101	F06	A
	64029	IS 101	F06	C

	Student_ID	Course	Date	Grade
Record	39044	IS 101	F06	B+

Field	IS 101 (Course field)
Byte	0100 1001 (Letter I in ASCII)
Bit	0

Source: Kenneth C. Laudon & Jane P. Laudon (2014), Management Information Systems: Managing the Digital Firm, Thirteenth Edition, Pearson.
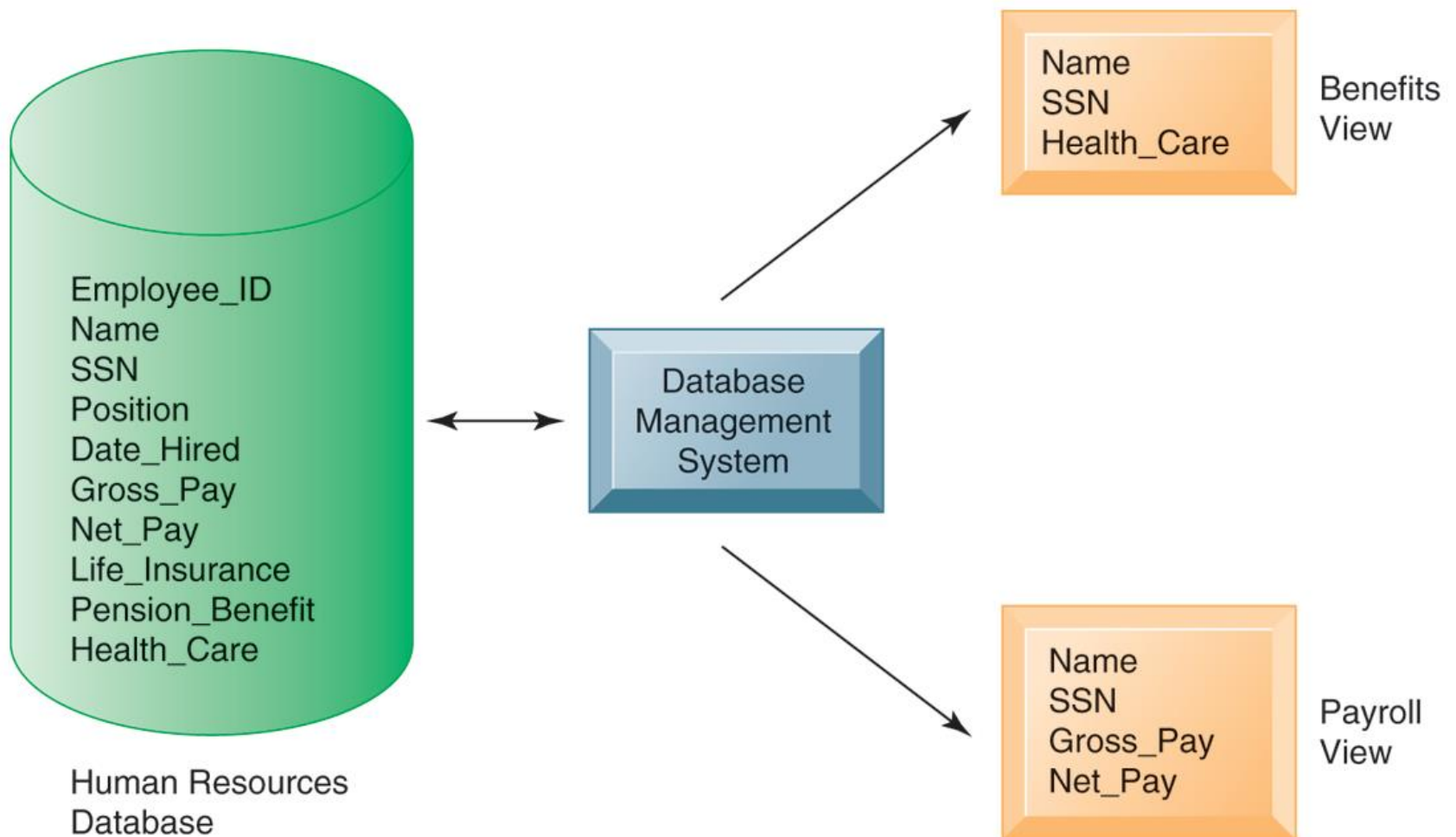
8

TRADITIONAL FILE PROCESSING

The Database Approach to Data Management

- Database
  - Serves many applications by centralizing data and controlling redundant data

The Database Approach to Data Management

- Database management system (DBMS)
  - Interfaces between applications and physical data files
  - Separates logical and physical views of data
  - Solves problems of traditional file environment
    - Controls redundancy
    - Eliminates inconsistency
    - Uncouples programs and data
    - Enables organization to central manage data and data security

HUMAN RESOURCES DATABASE WITH MULTIPLE VIEWS

Human Resources Database

Employee_ID
Name
SSN
Position
Date_Hired
Gross_Pay
Net_Pay
Life_Insurance
Pension_Benefit
Health_Care

Database Management System

Name
SSN
Health_Care

Benefits View

Name
SSN
Gross_Pay
Net_Pay

Payroll View

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
- **Primary key**: Field in table used for key fields
- **Foreign key**: Primary key used in second table as look-up field to identify records from original table

RELATIONAL DATABASE TABLES

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

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.

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

Designing Databases

- Conceptual (logical) design:
  - abstract model from business perspective

- Physical design:
  - How database is arranged on direct-access storage devices

Design process identifies and Normalization

- 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
  - Streamlining complex groupings of data to minimize redundant data elements and awkward many-to-many relationships

AN UNNORMALIZED RELATION FOR ORDER

ORDER (Before Normalization)

Order_Number	Order_Date	Part_Number	Part_Name	Unit_Price	Part_Quantity	Supplier_Number	Supplier_Name	Supplier_Street	Supplier_City	Supplier_State	Supplier_Zip

Source: Kenneth C. Laudon & Jane P. Laudon (2014), Management Information Systems: Managing the Digital Firm, Thirteenth Edition, Pearson.
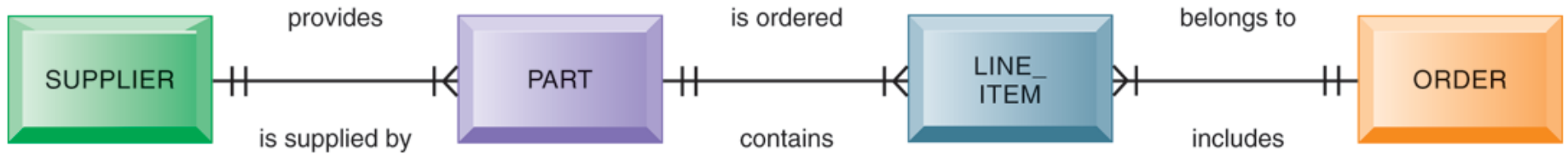
22

NORMALIZED TABLES CREATED FROM ORDER
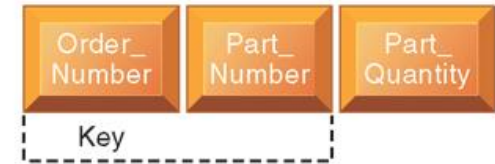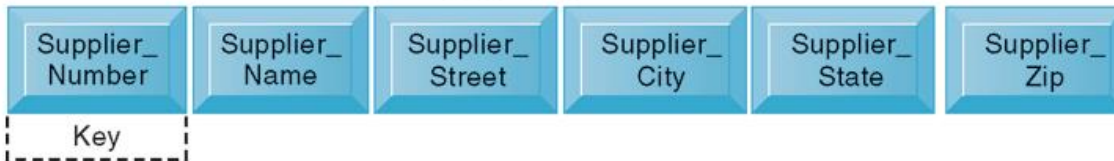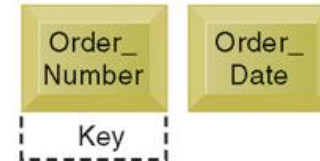
# AN ENTITY-RELATIONSHIP DIAGRAM

Using Databases to Improve Business Performance and Decision Making

- **Big data**
  - **Massive sets of unstructured/semi-structured data from Web traffic, social media, sensors, and so on**
  - **Petabytes, exabytes of data**
    - Volumes too great for typical DBMS
  - **Can reveal more patterns and anomalies**

Using Databases to Improve Business Performance and Decision Making

- **Business intelligence infrastructure**
  - **Today includes an array of tools for separate systems, and big data**

- **Contemporary tools**:
  - **Data warehouses**
  - **Data marts**
  - **Hadoop**
  - **In-memory computing**
  - **Analytical platforms**

Business Intelligence Infrastructure
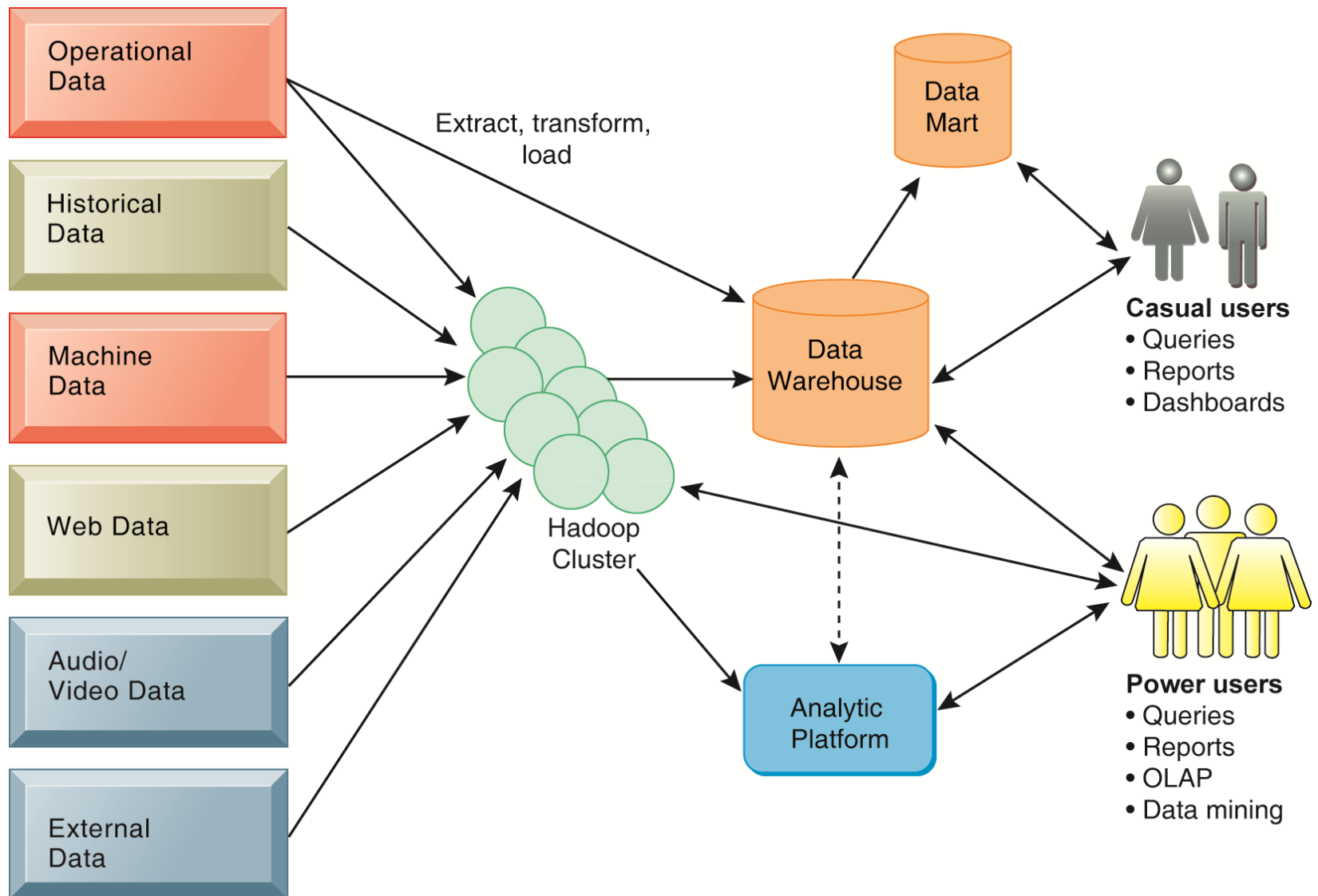
Data Warehouse vs. Data Marts

- **Data warehouse:**
  - Stores current and historical data from many core operational transaction systems
  - Consolidates and standardizes information for use across enterprise, but data cannot be altered
  - Provides analysis and reporting tools

- **Data marts:**
  - Subset of data warehouse
  - Summarized or focused portion of data for use by specific population of users
  - Typically focuses on single subject or line of business

Hadoop

- Enables distributed parallel processing of big data across inexpensive computers
- Key services
  - Hadoop Distributed File System (HDFS): data storage
  - MapReduce: breaks data into clusters for work
  - Hbase: NoSQL database
- Used by Facebook, Yahoo, NextBio

In-memory computing

- Used in big data analysis

- Use computers main memory (RAM) for data storage to avoid delays in retrieving data from disk storage

- Can reduce hours/days of processing to seconds

- Requires optimized hardware

Analytic platforms

- High-speed platforms using both relational and non-relational tools optimized for large datasets

- Examples:
 - IBM Netezza
  - Oracle Exadata

Analytical tools: Relationships, patterns, trends

- Business Intelligence Analytics and Applications
- Tools for consolidating, analyzing, and providing access to vast amounts of data to help users make better business decisions
  - Multidimensional data analysis (OLAP)
  - Data mining
  - Text mining
  - Web mining

Online analytical processing (OLAP)

- Supports multidimensional data analysis
  - Viewing data using multiple dimensions
  - Each aspect of information (product, pricing, cost, region, time period) is different dimension
  - Example: How many washers sold in East in June compared with other regions?

- OLAP enables rapid, online answers to ad hoc queries

MULTIDIMENSIONAL DATA MODEL
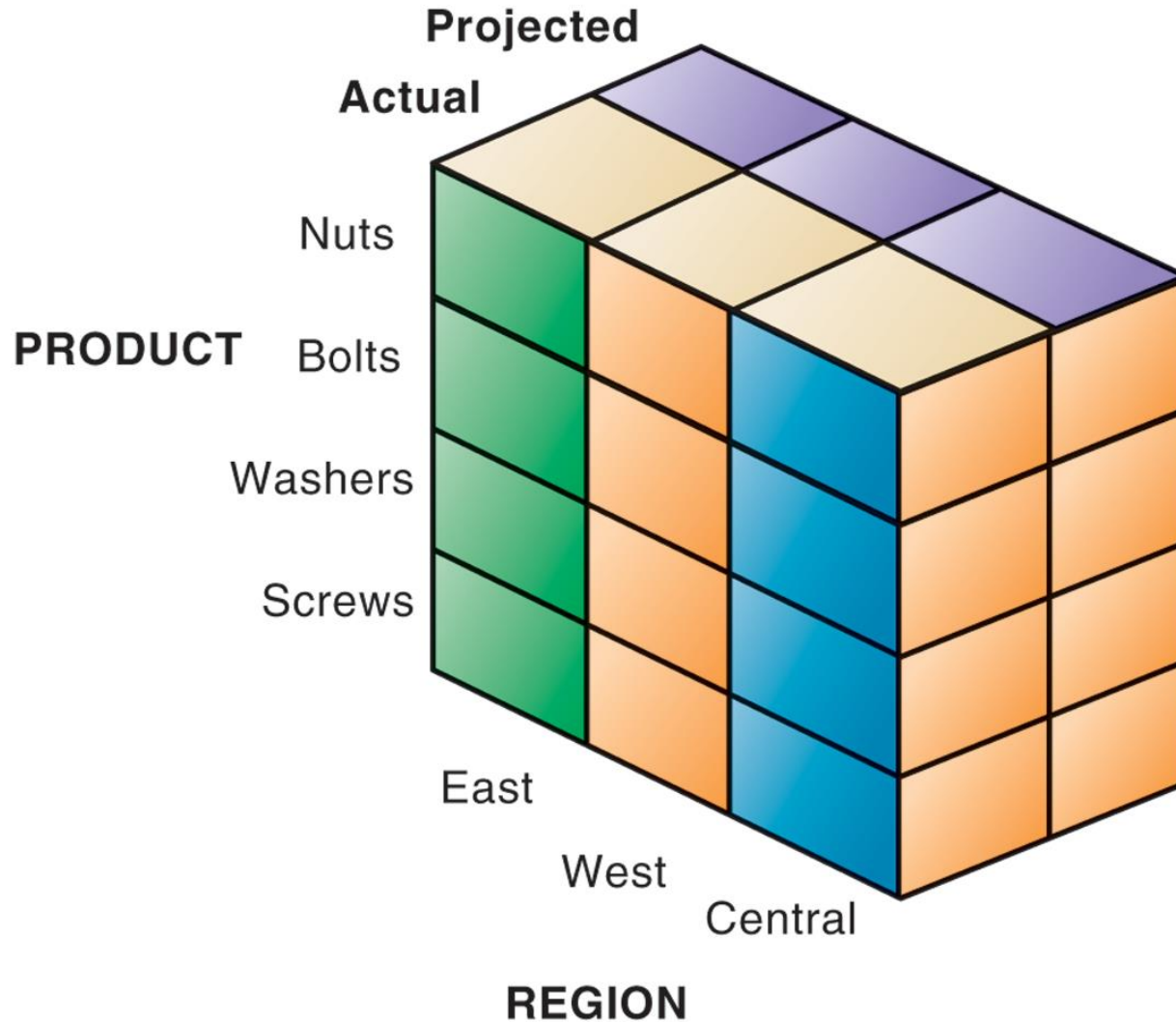
Data mining

- Finds hidden patterns, relationships in datasets
  - Example: customer buying patterns
- Infers rules to predict future behavior
  - Data mining provides insights into data that cannot be discovered through OLAP, by inferring rules from patterns in data.

Types of Information Obtained from Data Mining

- **Associations**: Occurrences linked to single event

- **Sequences**: Events linked over time

- **Classification**: Recognizes patterns that describe group to which item belongs

- **Clustering**: Similar to classification when no groups have been defined; finds groupings within data

- **Forecasting**: Uses series of existing values to forecast what other values will be

Text mining

- Extracts key elements from large unstructured data sets
  - Stored e-mails
  - Call center transcripts
  - Legal cases
  - Patent descriptions
  - Service reports, and so on
- Sentiment analysis software
 - Mines e-mails, blogs, social media to detect opinions

Web mining

- Discovery and analysis of useful patterns and information from Web
  - Understand customer behavior
  - Evaluate effectiveness of Web site, and so on
- 3 Tasks of Web Mining
  - Web content mining
    - Mines content of Web pages
  - Web structure mining
    - Analyzes links to and from Web page
  - Web usage mining
    - Mines user interaction data recorded by Web server

Databases and the Web

- Many companies use Web to make some internal databases available to customers or partners

- Typical configuration includes:
  - Web server
  - Application server/middleware/CGI scripts
  - Database server (hosting DBMS)

- Advantages of using Web for database access:
  - Ease of use of browser software
  - Web interface requires few or no changes to database
  - Inexpensive to add Web interface to system

LINKING INTERNAL DATABASES TO THE WEB
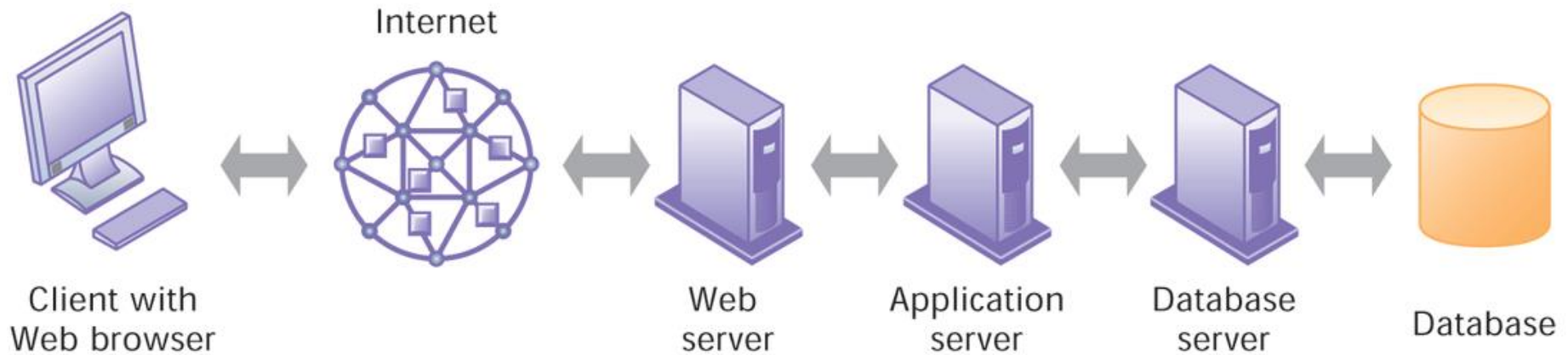
Managing Data Resources

- **Establishing an information policy**
  - Firm's rules, procedures, roles for sharing, managing, standardizing data
  - Data administration
    - Establishes policies and procedures to manage data
  - Data governance
    - Deals with policies and processes for managing availability, usability, integrity, and security of data, especially regarding government regulations
  - Database administration
    - Creating and maintaining database

Managing Data Resources

- **Ensuring data quality**
  - More than 25% of critical data in Fortune 1000 company databases are inaccurate or incomplete
    - Redundant data
    - Inconsistent data
    - Faulty input
  - Before new database in place, need to:
    - Identify and correct faulty data
    - Establish better routines for editing data once database in operation

Managing Data Resources

- **Data quality audit**
  - Structured survey of the accuracy and level of completeness of the data in an information system
    - Survey samples from data files, or
    - Survey end users for perceptions of quality
- **Data cleansing**
  - Software to detect and correct data that are incorrect, incomplete, improperly formatted, or redundant
  - Enforces consistency among different sets of data from separate information systems

Case Study:

Google, Apple, and Microsoft (Chap. 7) (pp. 318-320)

Apple, Google, and Microsoft Battle for Your Internet Experience

1. Define and compare the business models and areas of strength of Apple, Google, and Microsoft.

2. Why is mobile computing so important to these three firms? Evaluate the mobile platform offerings of each firm.

3. What is the significance of applications and app stores, and closed vs. open app standards to the success or failure of mobile computing?

4. Which company and business model do you believe will prevail in this epic struggle? Explain your answer.

5. What difference would it make to a business or to an individual consumer if Apple, Google, or Microsoft dominated the Internet experience? Explain your answer.

資訊管理個案
(Case Study for Information Management)

1. 請同學於資訊管理個案討論前
   應詳細研讀個案，並思考個案研究問題。

2. 請同學於上課前複習相關資訊管理相關理論
   ，以作為個案分析及擬定管理對策的依據。

3. 請同學於上課前
   先繳交個案研究問題書面報告。

References

- Kenneth C. Laudon & Jane P. Laudon (2014), Management Information Systems: Managing the Digital Firm, Thirteenth Edition, Pearson.

- Kenneth C. Laudon & Jane P. Laudon原著，游張松 主編，陳文生 翻譯 (2014)，資訊管理系統，第13版，滄海