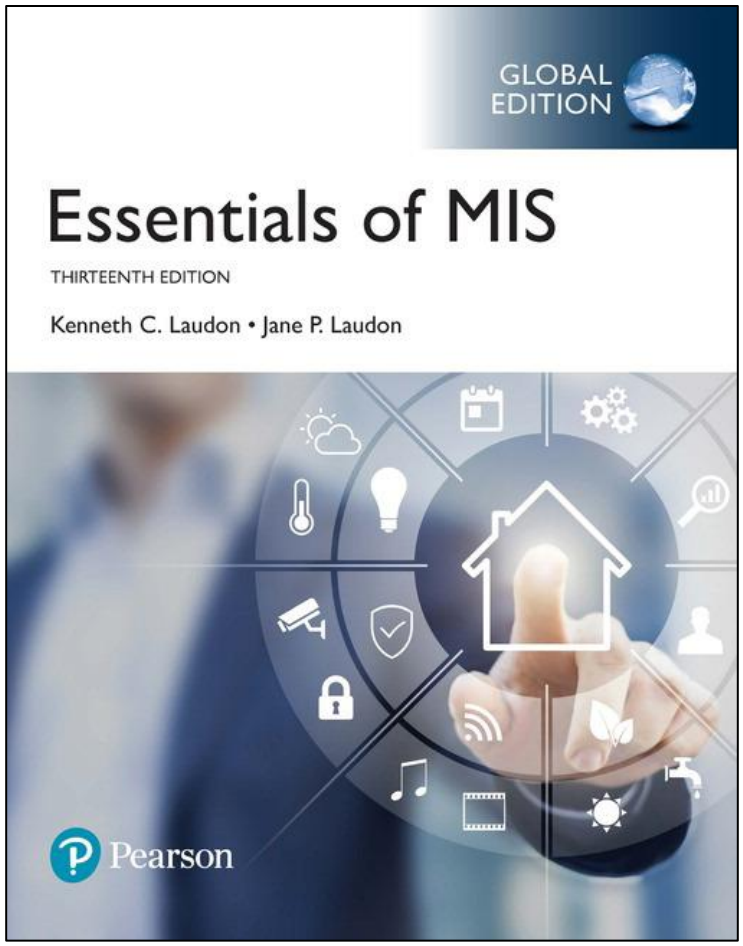
Essentials of Management Information Systems

Thirteenth Edition



Chapter 6

Foundations of Business Intelligence: Databases and Information Management

Learning Objectives

6.1 What is a database, and how does a relational database organize data?

6.2 What are the principles of a database management system?

6.3 What are the principal tools and technologies for accessing information from databases to improve business performance and decision making?

6.4 Why are information policy, data administration, and data quality assurance essential for managing the firm's data resources?

6.5 How will MIS help my career?

Video Cases

- Case 1: Dubuque Uses Cloud Computing and Sensors to Build a Smarter City
- Case 2: Brooks Brothers Closes in on Omnichannel Retail
- Case 3: Maruti Suzuki Business Intelligence and Enterprise Databases

Data Management Helps the Charlotte Hornets Learn More About Their Fans

- Problem
  - Large volumes of data in isolated databases
  - Outdated data management technology
- Solutions
  - SAP HANA
  - Data warehouse
  - FanTracker
- Illustrates the importance of data management for better decision making and customer analysis

What is a Database?

- Database:
  - Collection of related files containing records on people, places, or things
- Entity:
  - Generalized category representing person, place, thing
  - E.g., SUPPLIER, PART
- Attributes:
  - Specific characteristics of each entity:
    - SUPPLIER name, address
    - PART description, unit price, supplier

Relational Databases

- Organize data into two-dimensional tables (relations) with columns and rows
- One table for each entity:
  - E.g., (CUSTOMER, SUPPLIER, PART, SALES)
  - Fields (columns) store data representing an attribute
  - Rows store data for separate records, or tuples
- Key field: uniquely identifies each record
- Primary key

Figure 6.2 A Relational Database Table

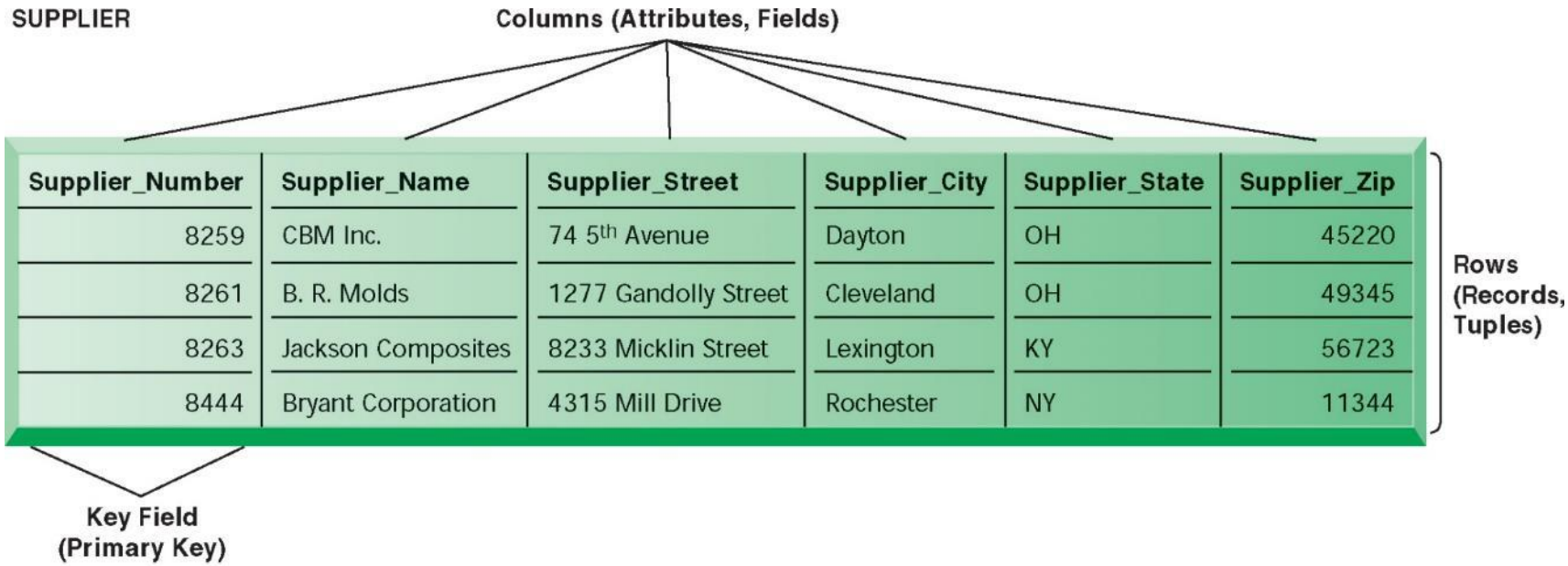


Figure 6.3 The PART Table

PART

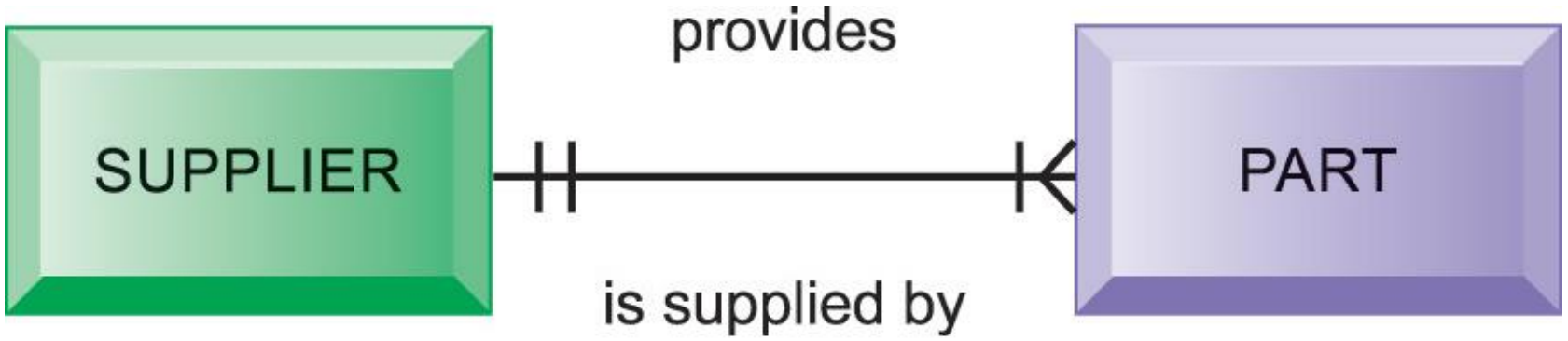
Part_Number	Part_Name	Unit_Price	Supplier_Number
137	Door latch	22.00	8259
145	Side mirror	12.00	8444
150	Door molding	6.00	8263
152	Door lock	31.00	8259
155	Compressor	54.00	8261
178	Door handle	10.00	8259

Primary Key **Foreign Key**

Establishing Relationships (1 of 2)

- Entity-relationship diagram
  - Used to clarify table relationships in a relational database
- Relational database tables may have:
  - One-to-one relationship
  - One-to-many relationship
  - Many-to-many relationship
    - Requires “join table” or intersection relation that links the two tables to join information

Figure 6.4 A Simple Entity-Relationship Diagram



Establishing Relationships (2 of 2)

- Normalization
  - Streamlining complex groups of data
  - Minimizes redundant data elements
  - Minimizes awkward many-to-many relationships
  - Increases stability and flexibility
- Referential integrity rules
  - Ensure that relationships between coupled tables remain consistent

Figure 6.5 Sample Order Report

Order Number: 3502
Order Date: 1/15/2018

Supplier Number: 8259
Supplier Name: CBM Inc.
Supplier Address: 74 5th Avenue, Dayton, OH 45220

Order_Number	Part_Number	Part_Quantity	Part_Name	Unit_Price	Extended Price
3502	137	10	Door latch	22.00	$220.00
3502	152	20	Door lock	31.00	620.00
3502	178	5	Door handle	10.00	50.00
			Order Total:		$890.00

Figure 6.6 The Final Database Design with Sample Records

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

LINE_ITEM

Order_Number	Part_Number	Part_Quantity
3502	137	10
3502	152	20
3502	178	5

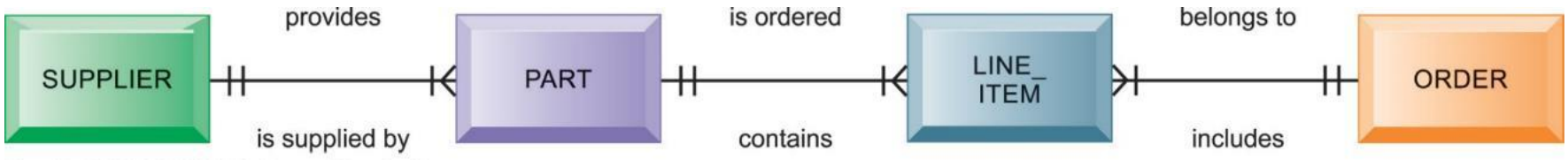
ORDER

Order_Number	Order_Date
3502	1/15/2018
3503	1/16/2018
3504	1/17/2018

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

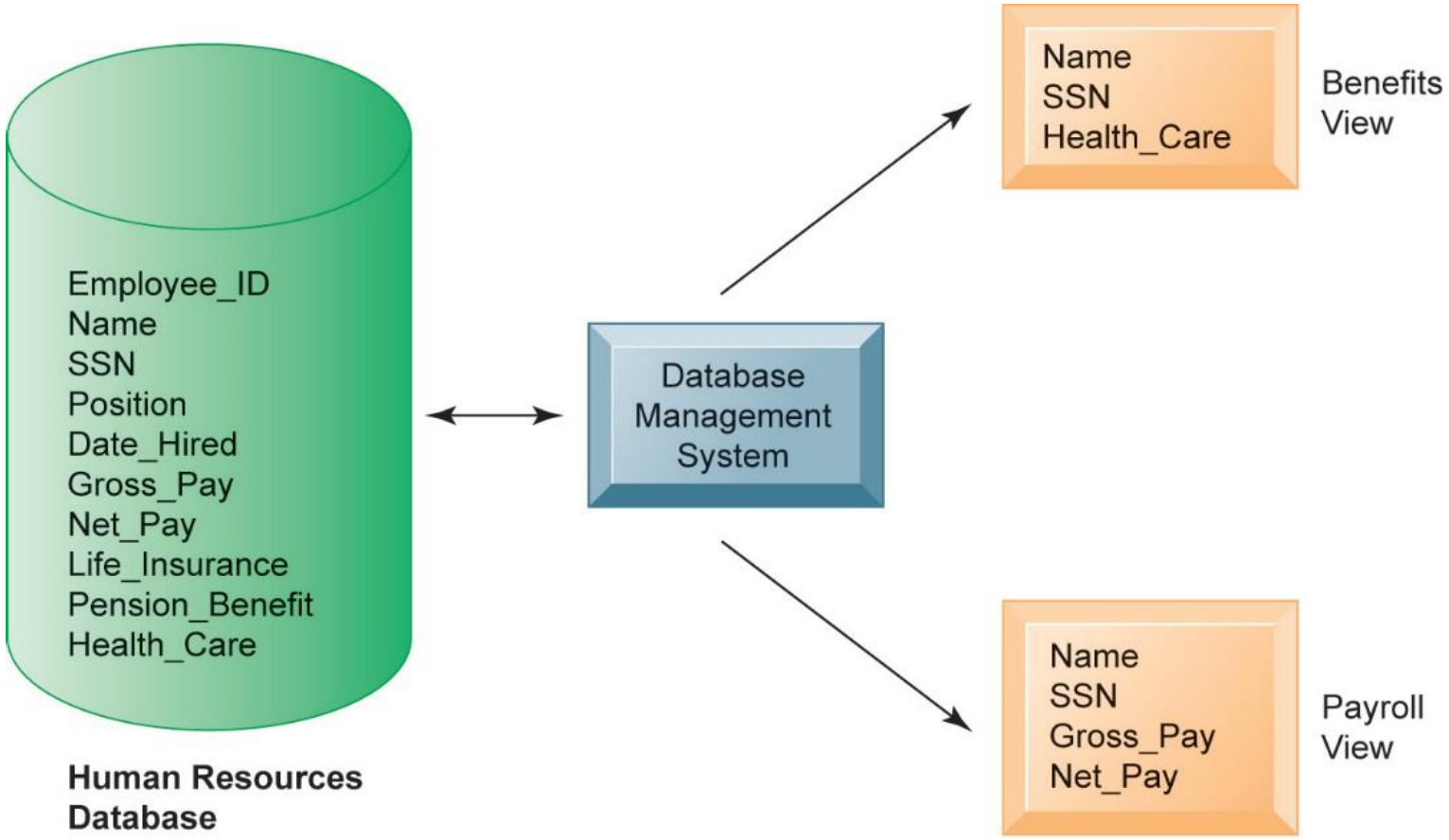
Figure 6.7 Entity-Relationship Diagram for the Database with Four Tables



Database Management Systems (DBMS)

- Software for creating, storing, organizing, and accessing data from a database
- Separates the logical and physical views of the data
  - Logical view: how end users view data
  - Physical view: how data are actually structured and organized
- Examples: Microsoft Access, DB2, Oracle Database, Microsoft SQL Server, MySQL

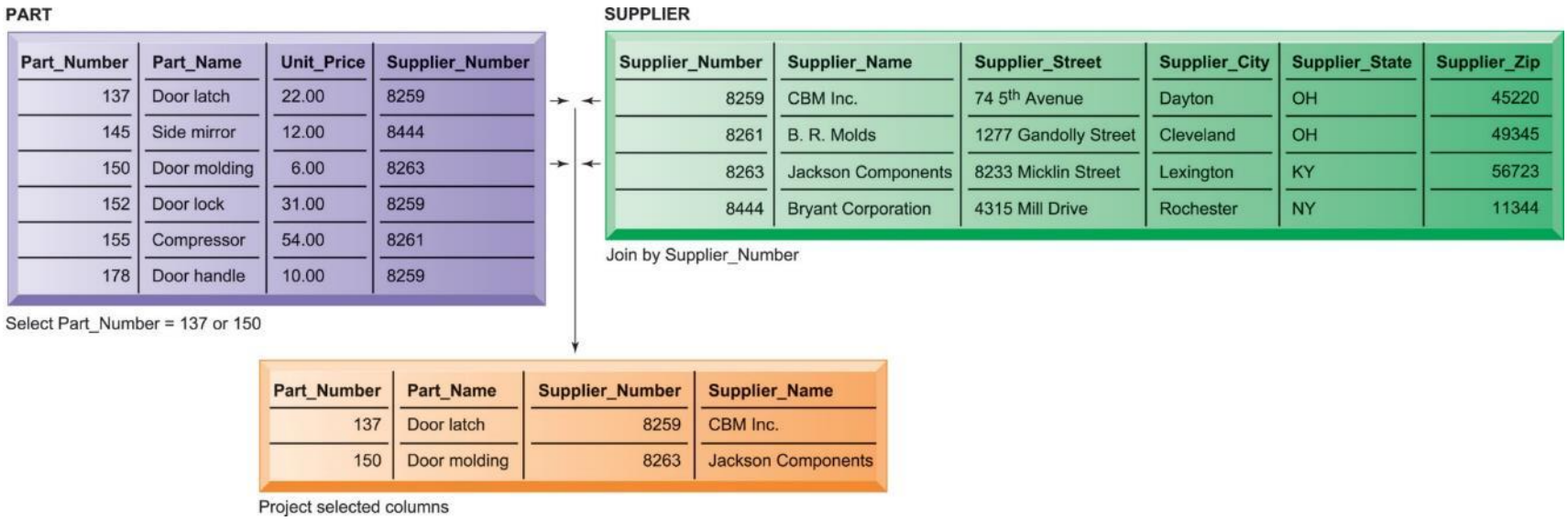
Figure 6.8 Human Resources Database with Multiple Views



Operations of a Relational DBMS

- Select:
  - Creates a subset of all records meeting stated criteria
- Join:
  - Combines relational tables to present the server with more information than is available from individual tables
- Project:
  - Creates a subset consisting of columns in a table
  - Permits user to create new tables containing only desired information

Figure 6.9 The Three Basic Operations of a Relational DBMS



Why do we normalize?

- The goal of data **normalization** is to reduce and even eliminate data redundancy, an important consideration for application developers because it is incredibly difficult to stores objects in a relational database that maintains the same information in several places.

Benefits of Normalization

- Greater overall database organization.
- Reduction of redundant data.
- Data consistency within the database.
- A much more flexible database design.
- A better handle on database security.

Capabilities of Database Management Systems

- Data definition capabilities:
  - Specify structure of content of database
- Data dictionary:
  - Automated or manual file storing definitions of data elements and their characteristics
- Querying and reporting:
  - Data manipulation language
    - Structured query language (SQL)
    - Microsoft Access query-building tools
  - Report generation

Figure 6.10 Access Data Dictionary Features

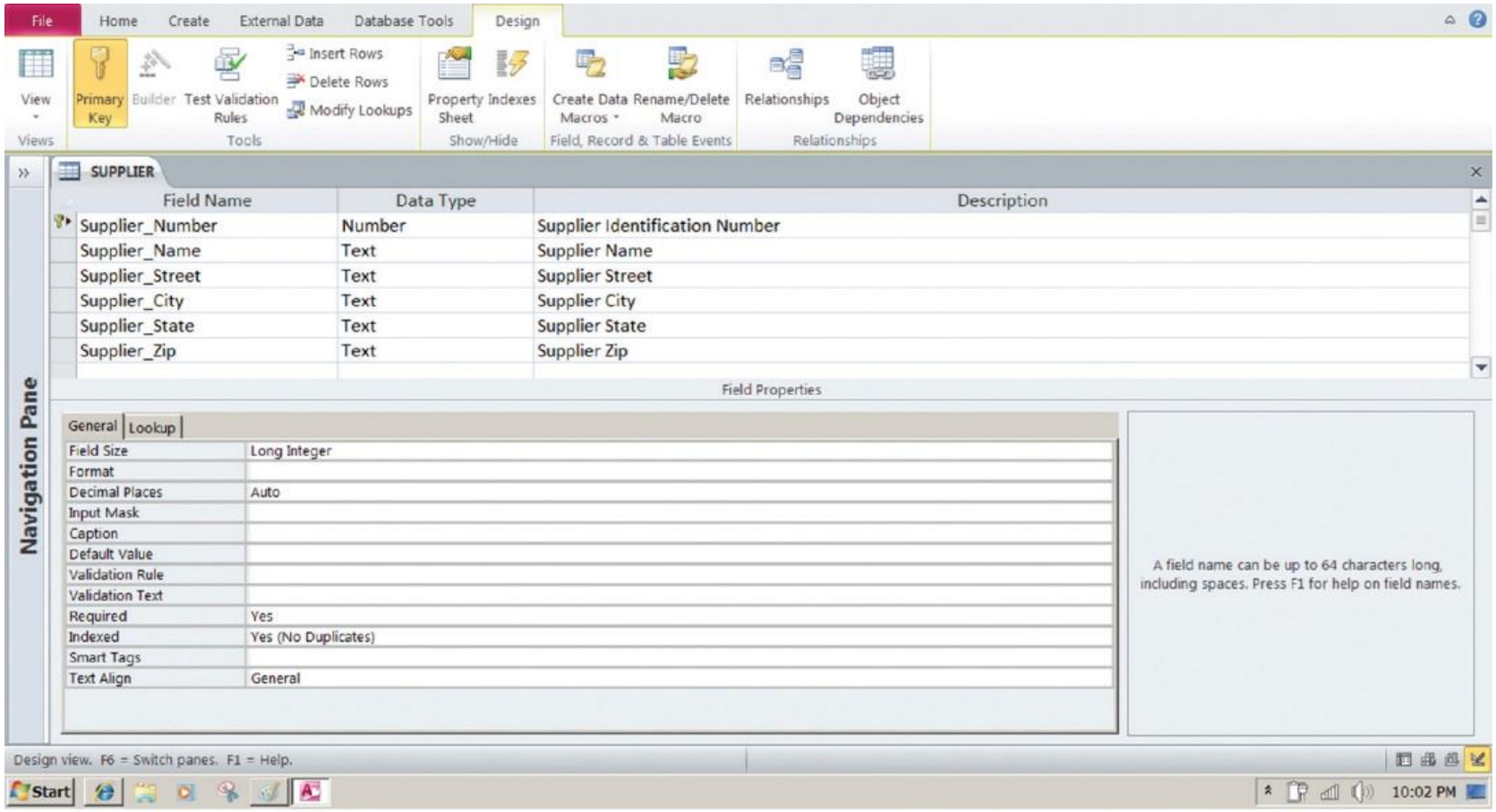
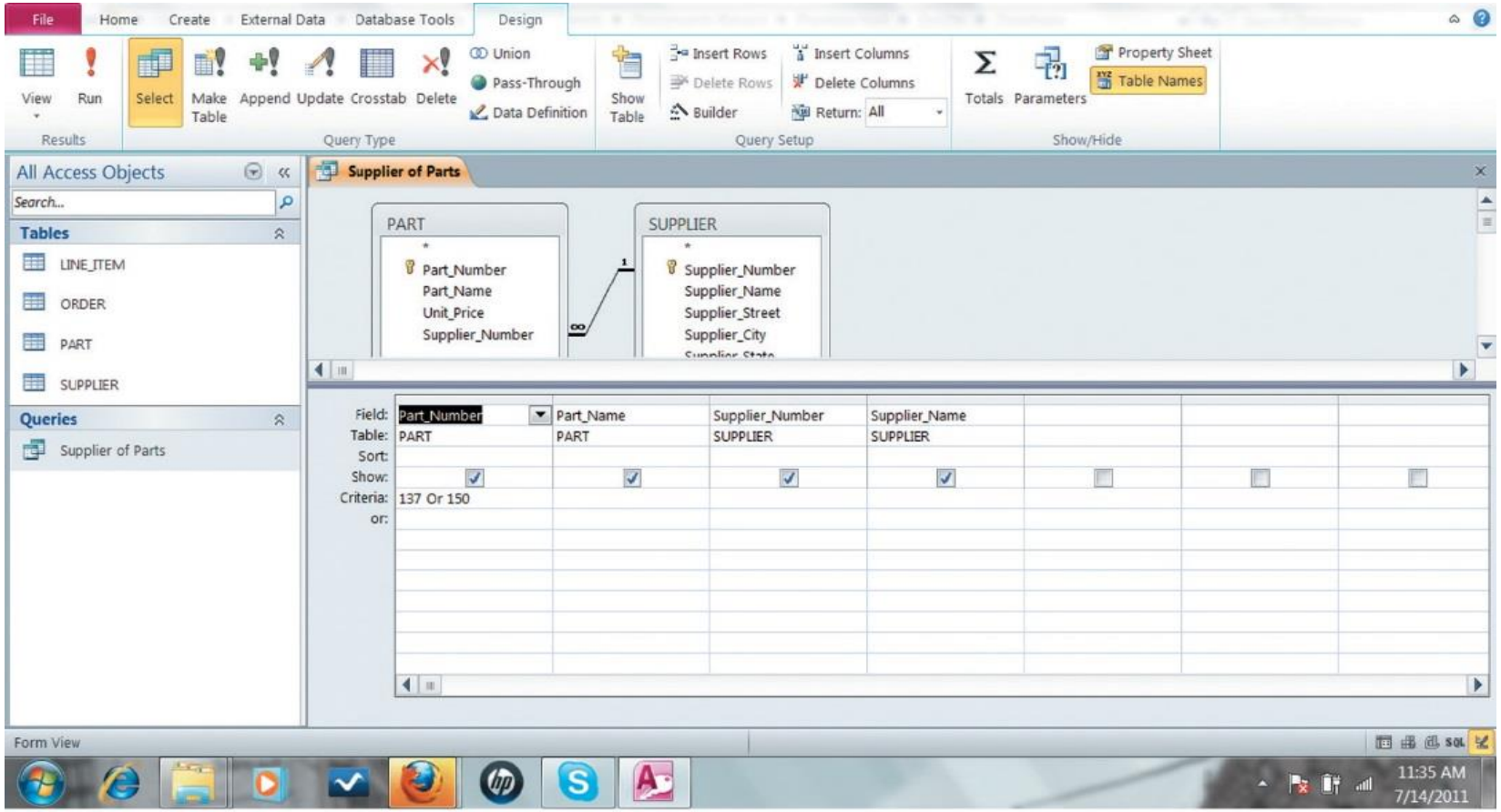


Figure 6.11 Example of an SQL Query

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

Figure 6.12 An Access Query



Non-Relational Databases

- “NoSQL”
- Handle large data sets of data that are not easily organized into tables, columns, and rows
- Use more flexible data model
  - Don’t require extensive structuring
- Can manage unstructured data, such as social media and graphics
- E.g. Amazon’s SimpleDB

Cloud Databases and Distributed Databases

- Relational database engines provided by cloud computing services
  - Pricing based on usage
  - Appeal to small or medium-sized businesses
- Amazon Relational Database Service
  - Offers MySQL, Microsoft SQL Server, Oracle Database engines
- Distributed databases
  - Stored in multiple physical locations
  - Google's Spanner cloud service

The Challenge of Big Data

- Massive quantities of unstructured and semi-structured data from Internet and more
  - 3Vs: Volume, variety, velocity (speed) or rapidity
- Big datasets offer more patterns and insights than smaller datasets, e.g.
  - Customer behavior
  - Weather patterns
- Requires new technologies and tools

Business Intelligence Infrastructure

- Array of tools for obtaining useful information from internal and external systems and big data
  - Data warehouses
  - Data marts
  - Hadoop
  - In-memory computing
  - Analytical platforms

Data Warehouses

- Data warehouse:
  - Database that stores current and historical data that may be of interest to decision makers
  - Consolidates and standardizes data from many systems, operational and transactional databases
  - Data can be accessed but not altered
- Data mart:
  - Subset of data warehouses that is highly focused and isolated for a specific population of users

Hadoop

- **An open source framework that is used to efficiently store and process large datasets** ranging in size from gigabytes to petabytes of data. Instead of using one large computer to store and process the data, Hadoop allows clustering multiple computers to analyze massive datasets in parallel more quickly.
- Breaks data task into sub-problems and distributes the processing to many inexpensive computer processing nodes
- Combines result into smaller data set that is easier to analyze
- Key services
  - Hadoop Distributed File System (HDFS)
  - MapReduce

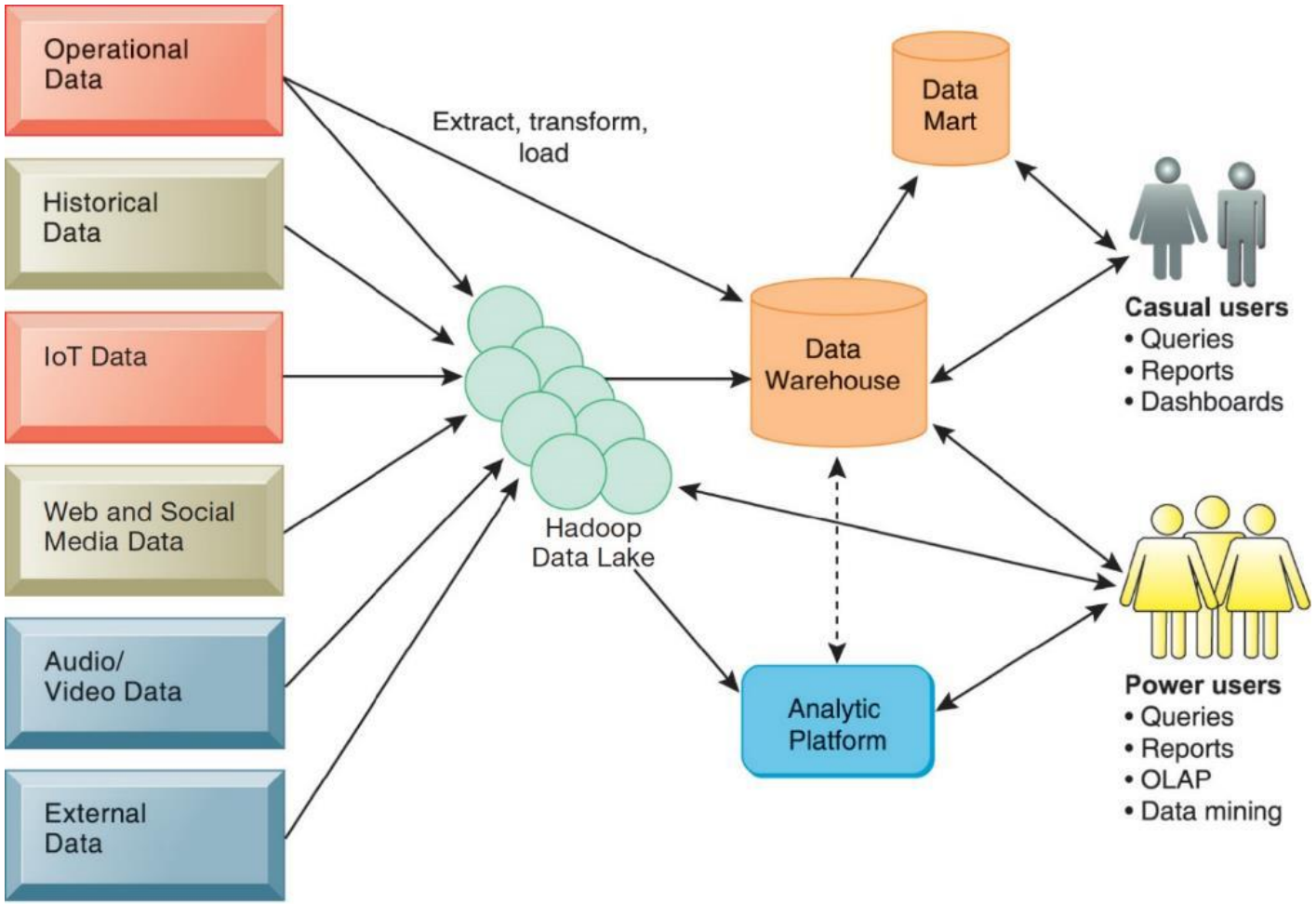
In-Memory Computing

- Relies on computer's main memory (RAM) for data storage
- Eliminates bottlenecks in retrieving and reading data
- Dramatically shortens query response times
- Enabled by high-speed processors, multicore processing
- Lowers processing costs

Analytic Platforms

- Preconfigured hardware-software systems
- Designed for query processing and analytics
- Use both relational and non-relational technology to analyze large data sets
- Include in-memory systems, NoSQL DBMS
- E.g. IBM PureData System for Analytics
  - Integrated database, server, storage components
- Data lakes

Figure 6.13 Business Intelligence Technology Infrastructure



Interactive Session – Society: Societe Generale Builds an Intelligent System to Manage Information Flow

- Class discussion
  - Why did Societe Generale's managers decide to develop an automated transaction processing system?
  - Why did managers decide they needed an "intelligent system?" In what way was the new system "intelligent?"
  - What is the role of human decision makers in the new system?
  - Why did managers select the Infogix platform?

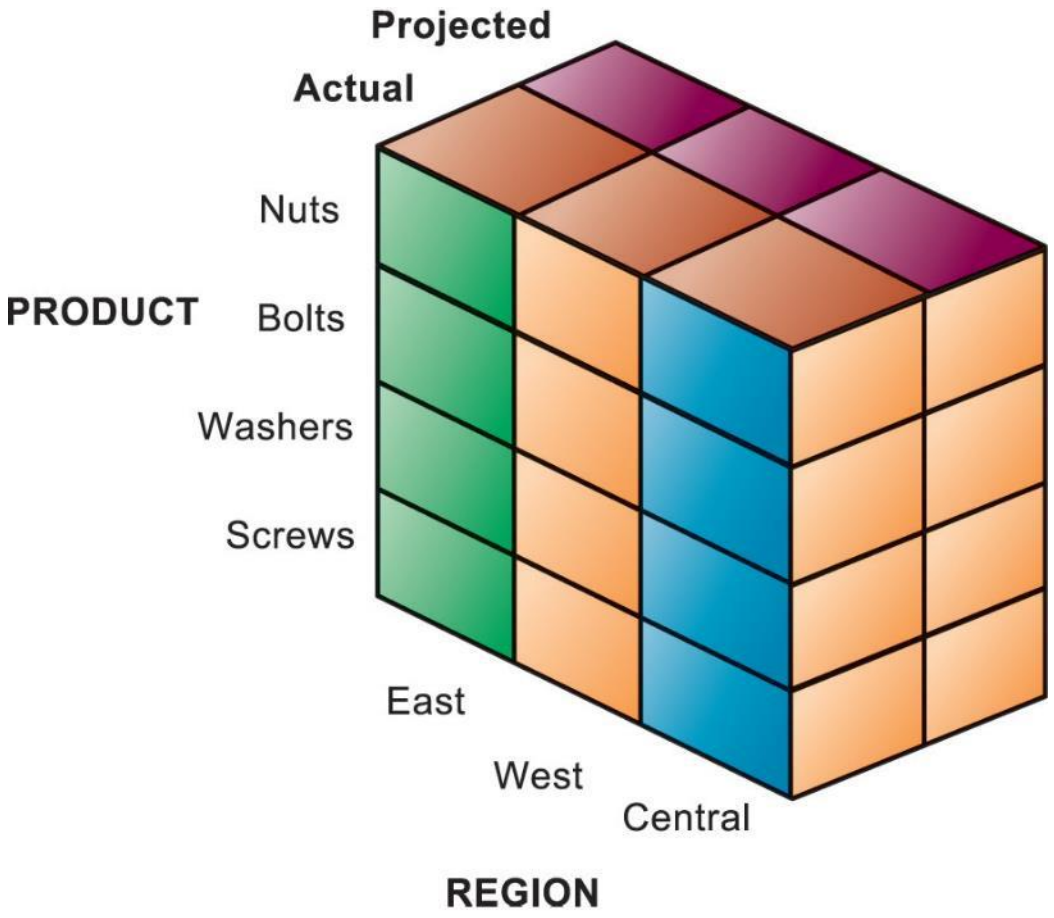
Analytical Tools: Relationships, Patterns, Trends

- Once data is gathered, tools are required for consolidating, analyzing, to use insights to improve decision making
  - Software for database querying and reporting
  - Multidimensional data analysis (OLAP)
  - Data mining

Online Analytical Processing (OLAP)

- Supports multidimensional data analysis, enabling users to view the same data in different ways using multiple dimensions
  - Each aspect of information—product, pricing, cost, region, or time period—represents a different dimension
  - E.g., comparing sales in East in June versus May and July
- Enables users to obtain online answers to ad hoc questions such as these in a fairly rapid amount of time

Figure 6.14 Multidimensional Data Model



Data Mining

- Finds hidden patterns and relationships in large databases and infers rules from them to predict future behavior
- Types of information obtainable from data mining
  - Associations: occurrences linked to single event
  - Sequences: events linked over time
  - Classifications: patterns describing a group an item belongs to
  - Clustering: discovering as yet unclassified groupings
  - Forecasting: uses series of values to forecast future values

Text Mining

- Unstructured data (mostly text files) accounts for 80 percent of an organization's useful information.
- Text mining allows businesses to extract key elements from, discover patterns in, and summarize large unstructured data sets.
- Sentiment analysis
  - Mines online text comments online or in email to measure customer sentiment

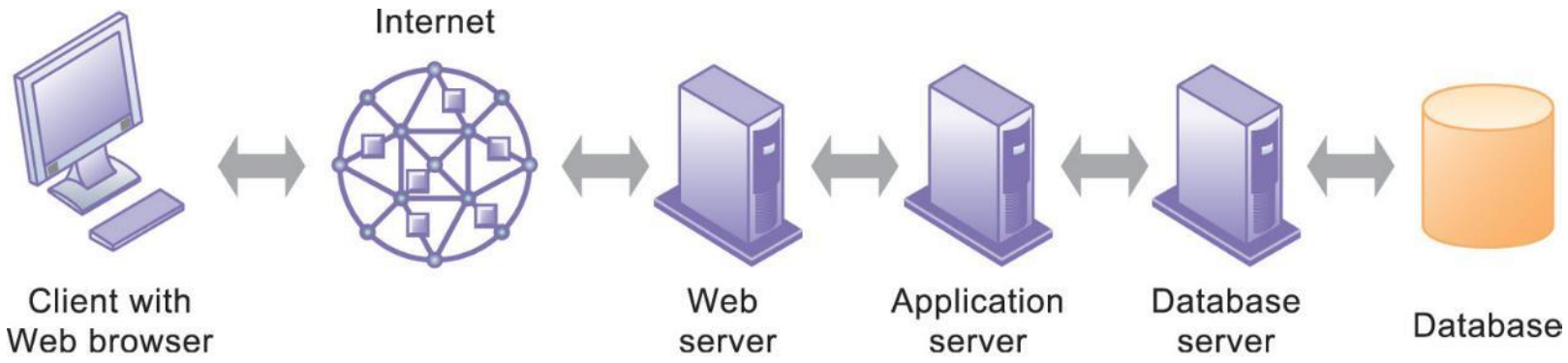
Web Mining

- Discovery and analysis of useful patterns and information from the web
  - E.g. to understand customer behavior, evaluate website, quantify success of marketing
- Content mining – mines content of websites
- Structure mining – mines website structural elements, such as links
- Usage mining – mines user interaction data gathered by web servers

Databases and the Web

- Firms use the web to make information from their internal databases available to customers and partners.
- Middleware and other software make this possible
  - Web server
  - Application servers or CGI
  - Database server
- Web interfaces provide familiarity to users and savings over redesigning legacy systems.

Figure 6.15 Linking Internal Databases to the Web



Establishing an Information Policy

- Information policy
  - States organization's rules for organizing, managing, storing, sharing information
- Data administration
  - Responsible for specific policies and procedures through which data can be managed as a resource
- Database administration
  - Database design and management group responsible for defining and organizing the structure and content of the database, and maintaining the database.

Ensuring Data Quality

- Poor data quality: major obstacle to successful customer relationship management
- Data quality problems caused by:
  - Redundant and inconsistent data produced by multiple systems
  - Data input errors
- Data quality audit
- Data cleansing