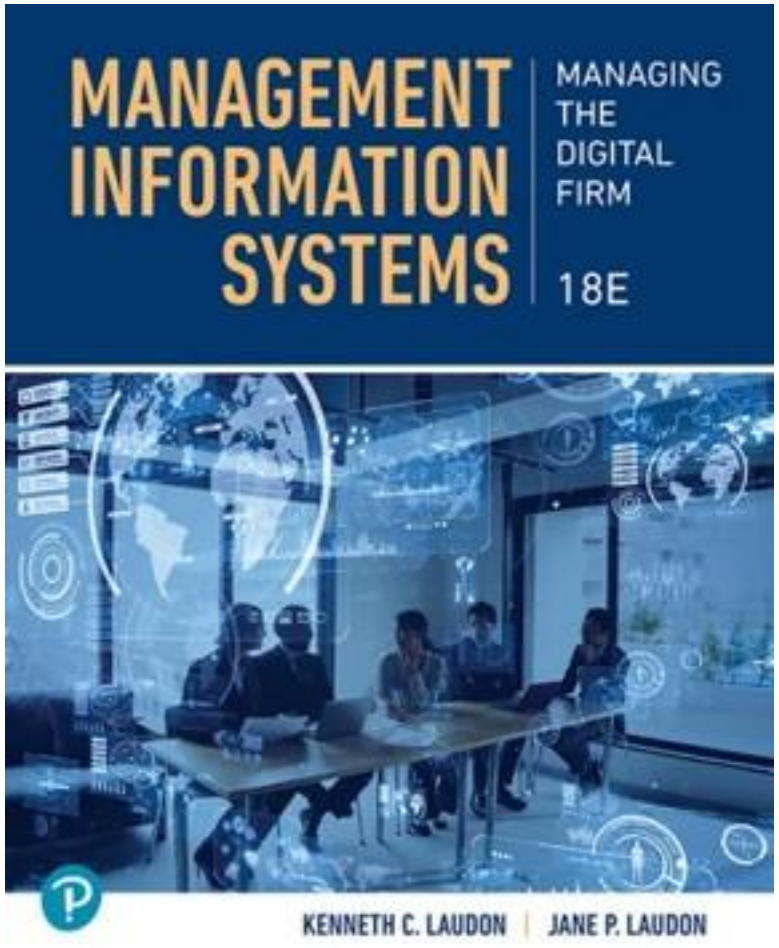
Management Information Systems: Managing the Digital Firm

Eighteenth Edition



Chapter 6

Foundations of Business Intelligence: Databases and Information Management

Learning Objectives (1 of 2)

6.1 Describe traditional file environment data management problems.

6.2 Describe DBMS and relational DBMS capabilities.

6.3 Discuss relational database design principles.

6.4 Describe non-relational, cloud, and blockchain database capabilities.

Learning Objectives (2 of 2)

6.5 Describe tools used to analyze data in databases.

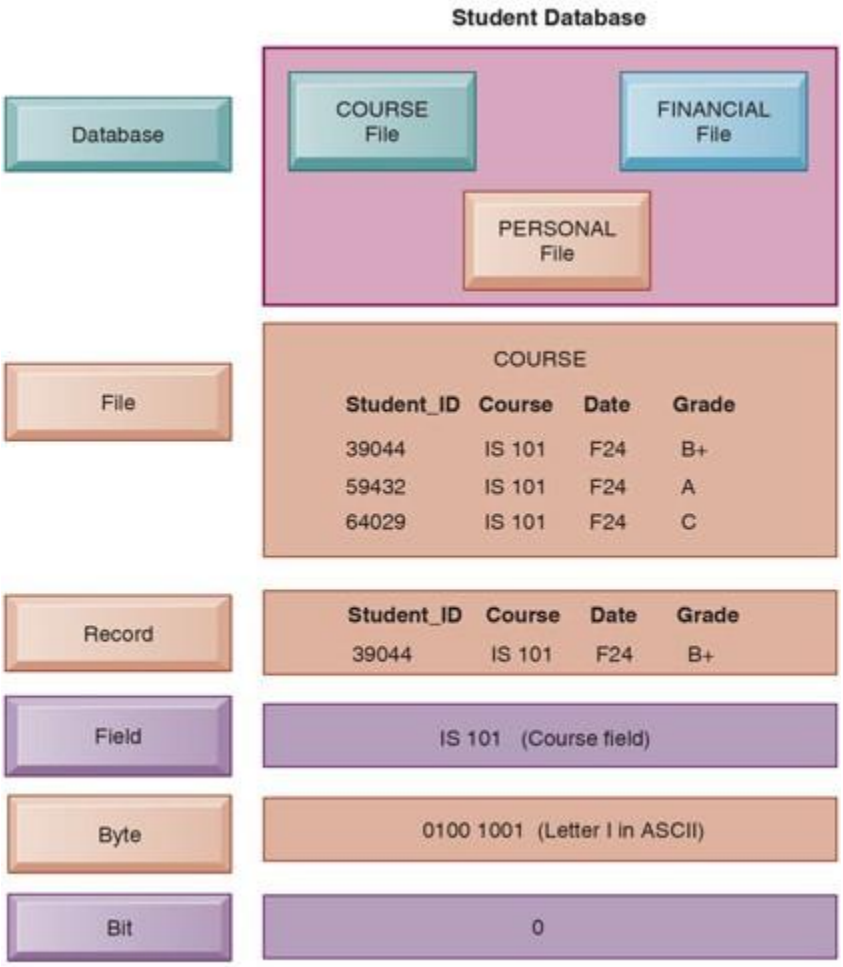
6.6 Discuss data governance, data quality assurance, and data management.

6.7 Understand how the information in this chapter can help your career.

Describe Traditional File Environment Data Management Problems

- An effective information system provides users with accurate, timely, and relevant information
  - Byte: a single character of data
  - Field: Group of characters as word(s) or number(s)
  - Record: Group of related fields
  - File: Group of records of same type
  - Database: Group of related files
  - Entity: Person, place, thing on which we store information
  - Attribute: Each characteristic, or quality, describing entity

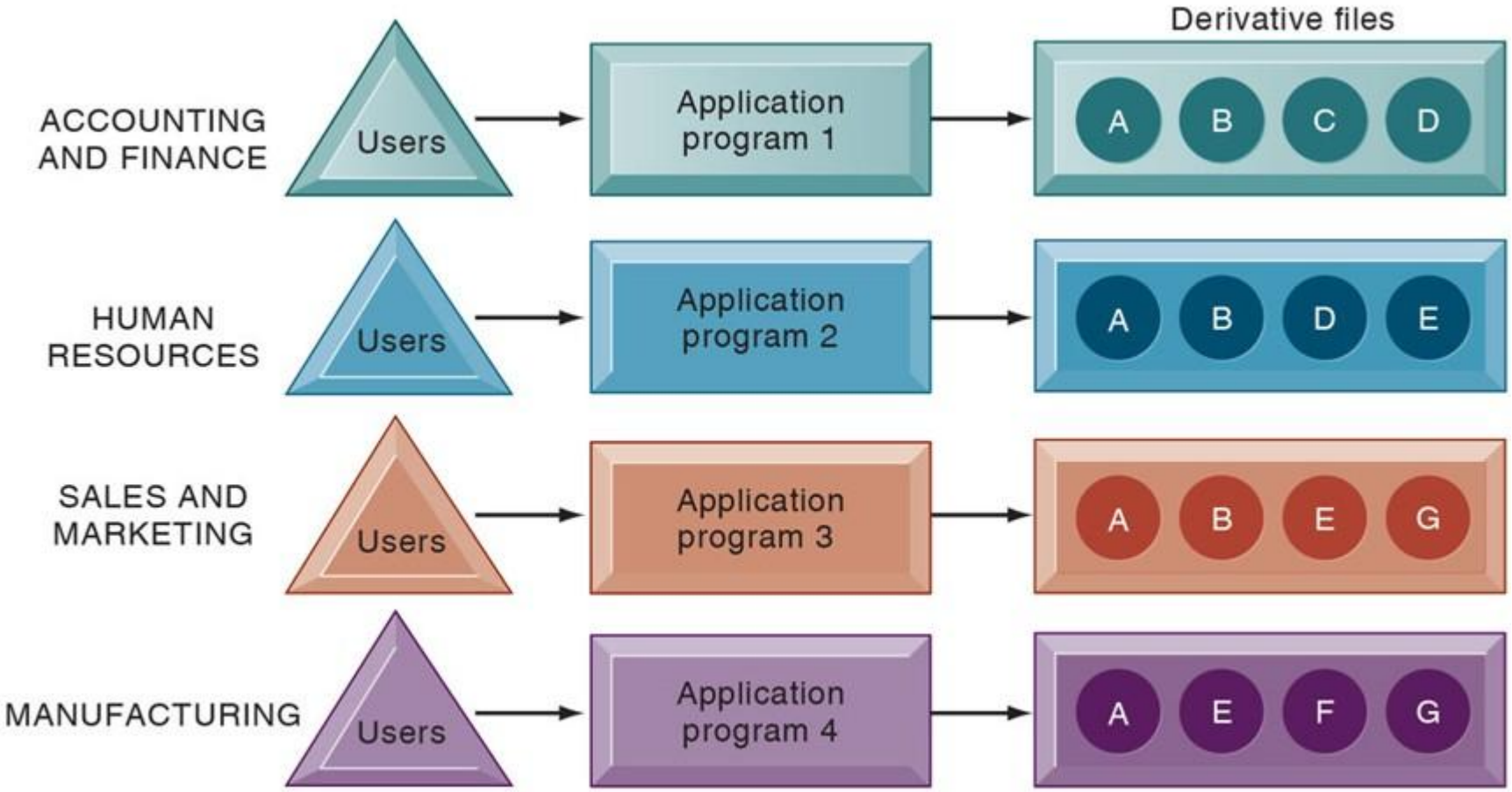
Figure 6.1 The Data Hierarchy



Describe Traditional File Environment Data Management Problems

- Problems with the traditional file environment
  - Files maintained separately by different departments
  - Data redundancy
  - Data inconsistency
  - Program-data dependence
  - Lack of flexibility
  - Poor data security
  - Inability to share data among applications

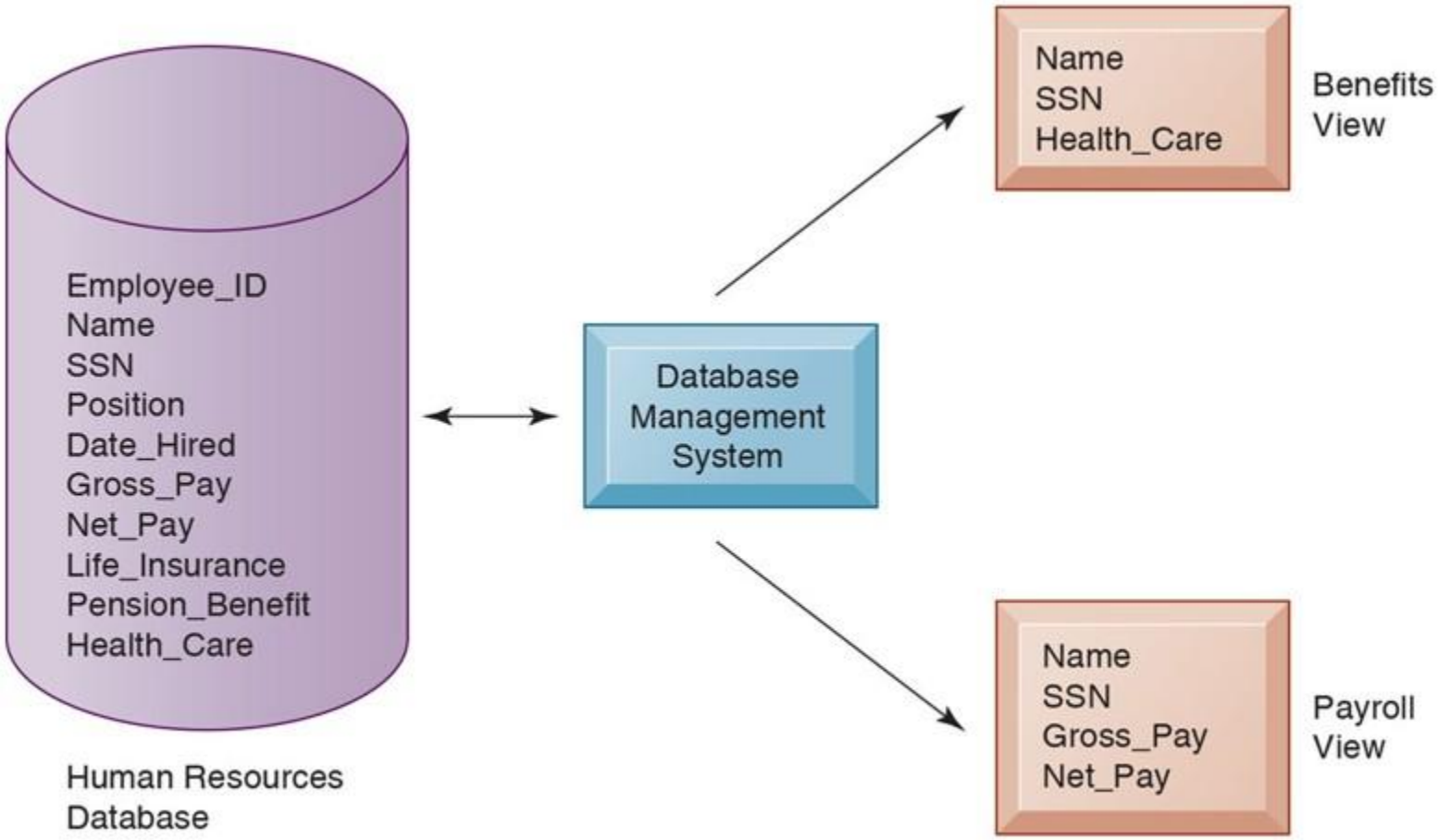
Figure 6.2 Traditional File Processing



Describe DBMS and Relational DBMS Capabilities (1 of 4)

- A database management system (DBMS)
  - Software that enables an organization to
    - Centralize data
    - Manage that data efficiently, and
    - Provide access to the stored data by application programs
  - A DBMS reduces data redundancy and inconsistency by minimizing isolated files in which the same data are repeated

Figure 6.3 Human Resources Database with Multiple Views



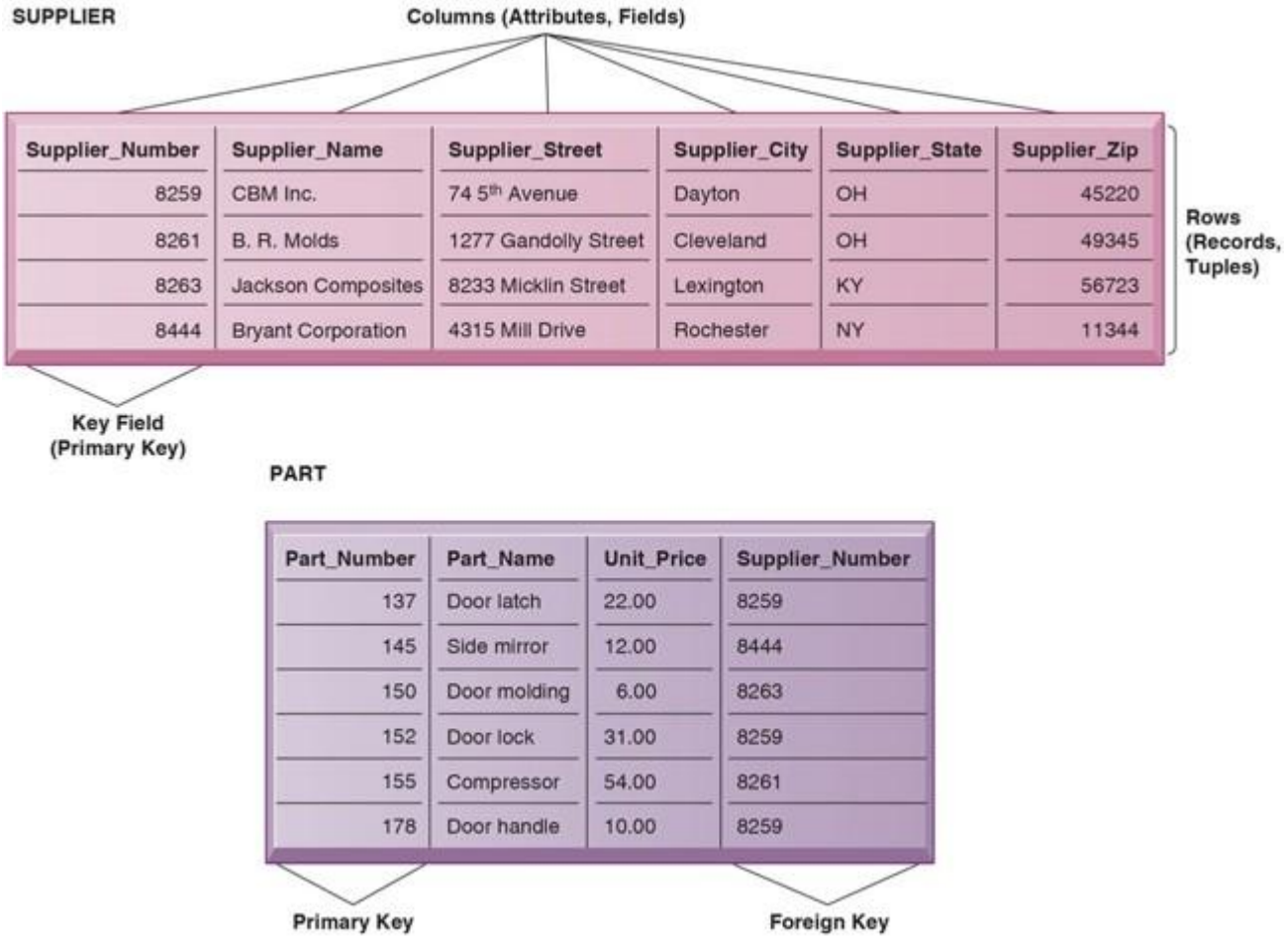
Describe DBMS and Relational DBMS Capabilities (2 of 4)

- Contemporary DBMS use different database models to keep track of entities, attributes, and relationships
- Relational database
  - Most common type of database
  - Represents data as two-dimensional tables (called relations)
  - Each table contains data on entity and attributes
- Relational DBMS (RDBMS)
  - Most popular type of DBMS today for desktops as well as for larger computers and mainframes

Describe DBMS and Relational DBMS Capabilities (3 of 4)

- A table is a grid of columns and rows
- The structure of a database includes
  - Rows: 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

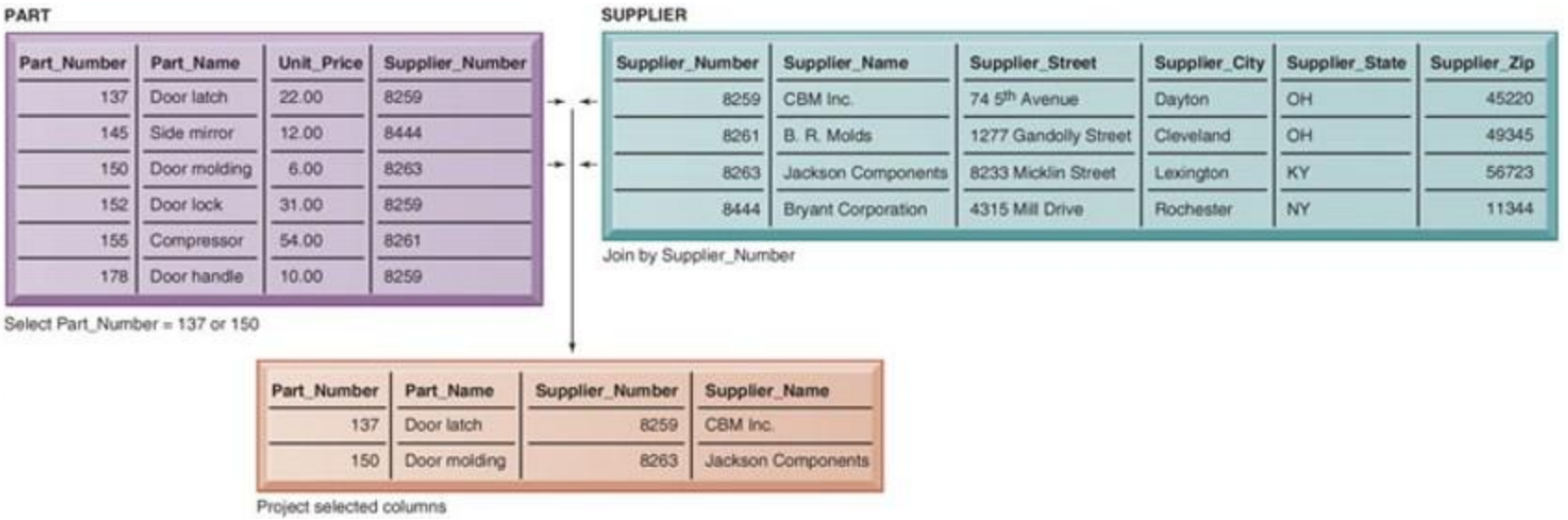
Figure 6.4 Relational Database Tables



Describe DBMS and Relational DBMS Capabilities (4 of 4)

- Three basic operations of a relational DBMS
  - 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

Figure 6.5 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

Capabilities of Database Management Systems

- Data definition
- Data dictionary
- Querying and reporting
  - Data manipulation language
  - Structured Query Language (S Q L)
- Many D B M S have report generation capabilities for creating polished reports (Microsoft Access)

Figure 6.6 Microsoft Access Data Dictionary Features

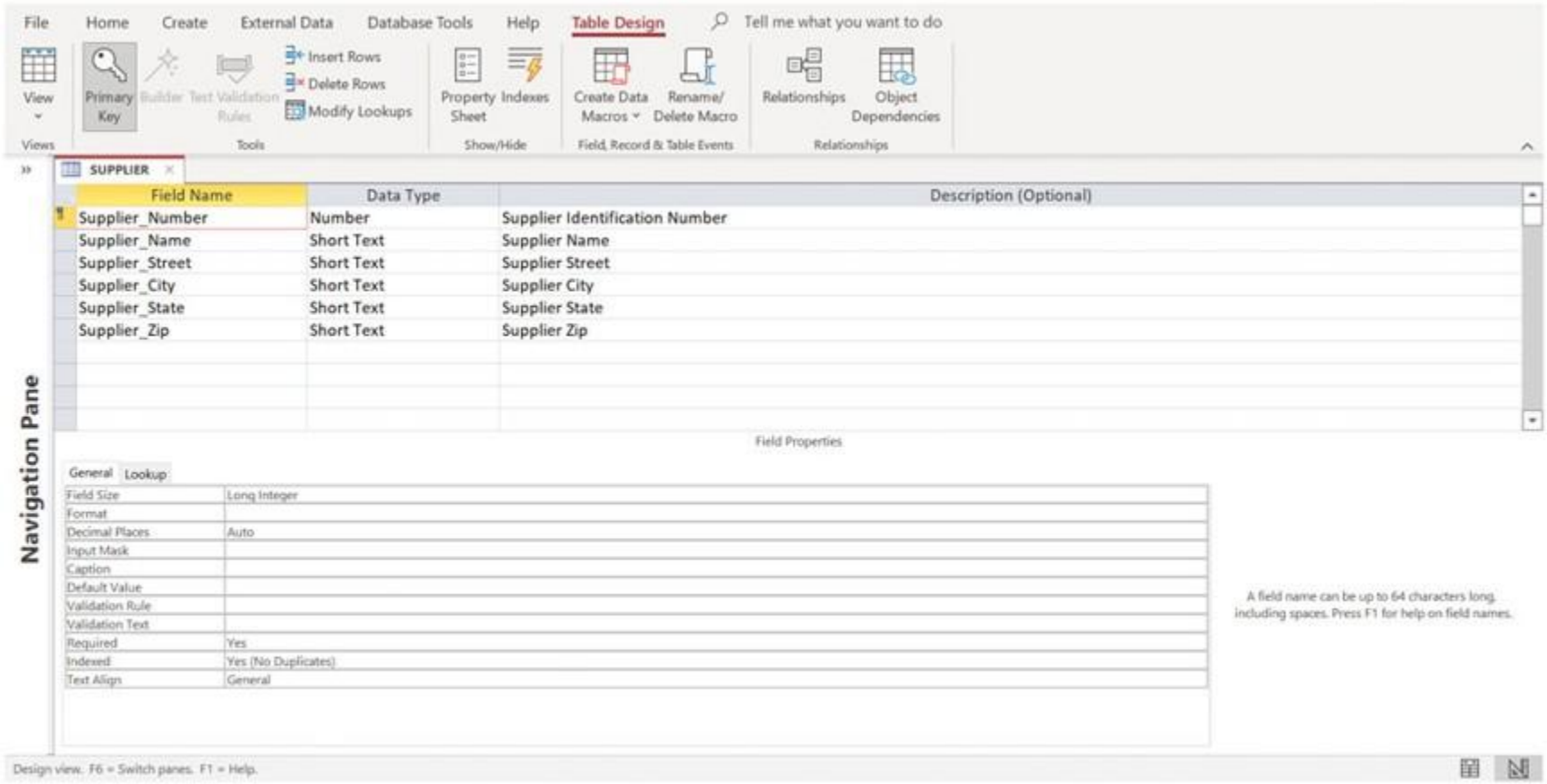
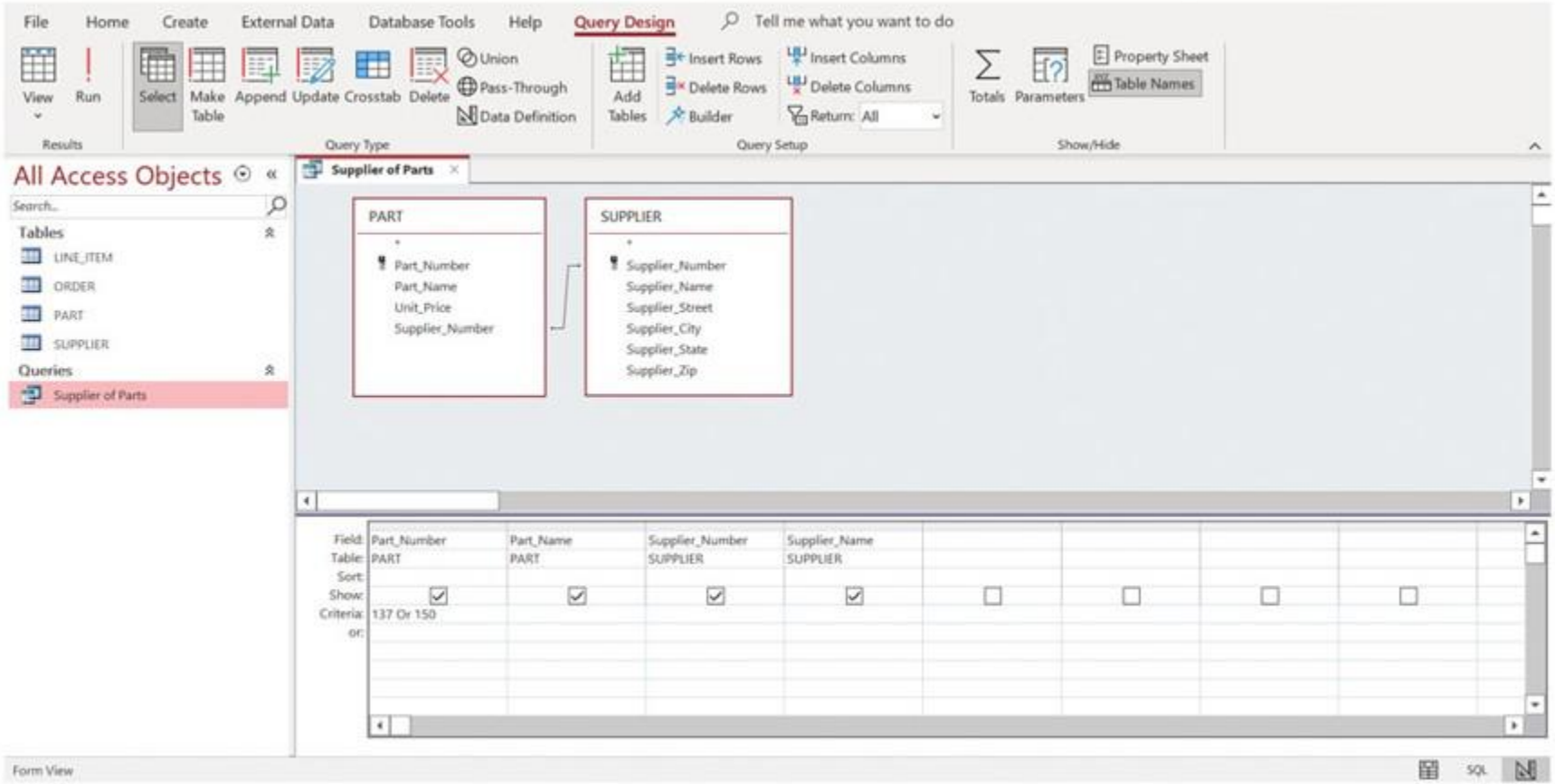


Figure 6.7 Example of an SQL Query

```
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;
```

Figure 6.8 Microsoft Access Query

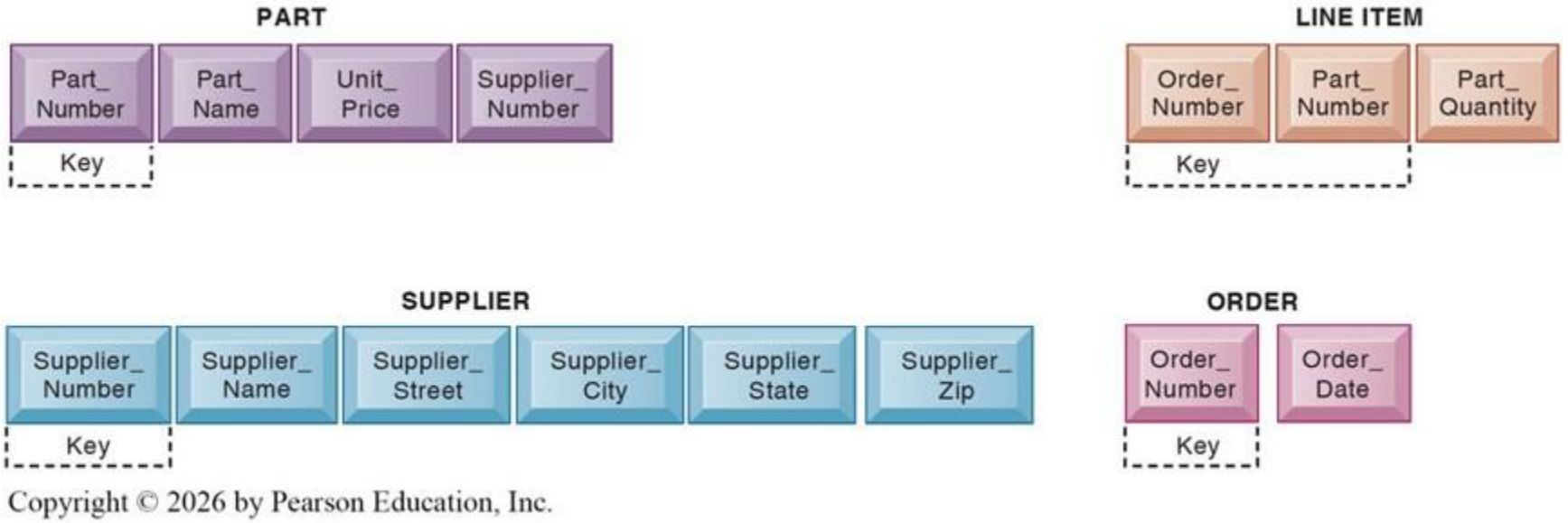


Discuss Relational Database Design Principles (1 of 2)

- Conceptual design vs. physical design
- Normalization
  - Streamlining complex groupings of data to minimize redundant data elements and awkward many-to-many relationships

Figure 6.9 Unnormalized Relation for ORDER

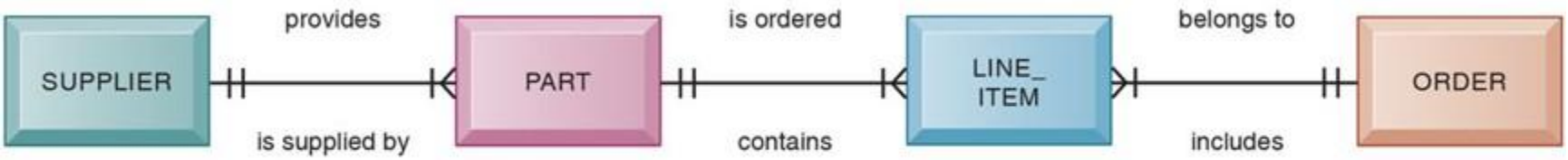
Figure 6.10 Normalized Tables Created from ORDER



Discuss Relational Database Design Principles (2 of 2)

- Referential integrity
  - Rules used by R D B M S to ensure relationships between tables remain consistent
- Entity-relationship diagram
  - Illustrates the relationship among various entities in the database

Figure 6.11 An Entity-Relationship Diagram



Describe Non-Relational Database, Cloud Database, and Blockchain Capabilities (1 of 3)

- Non-relational database management system: “No S Q L”
  - More flexible data model
  - Data sets stored across distributed machines
  - Easier to scale
  - Handle large volumes of unstructured and structured data

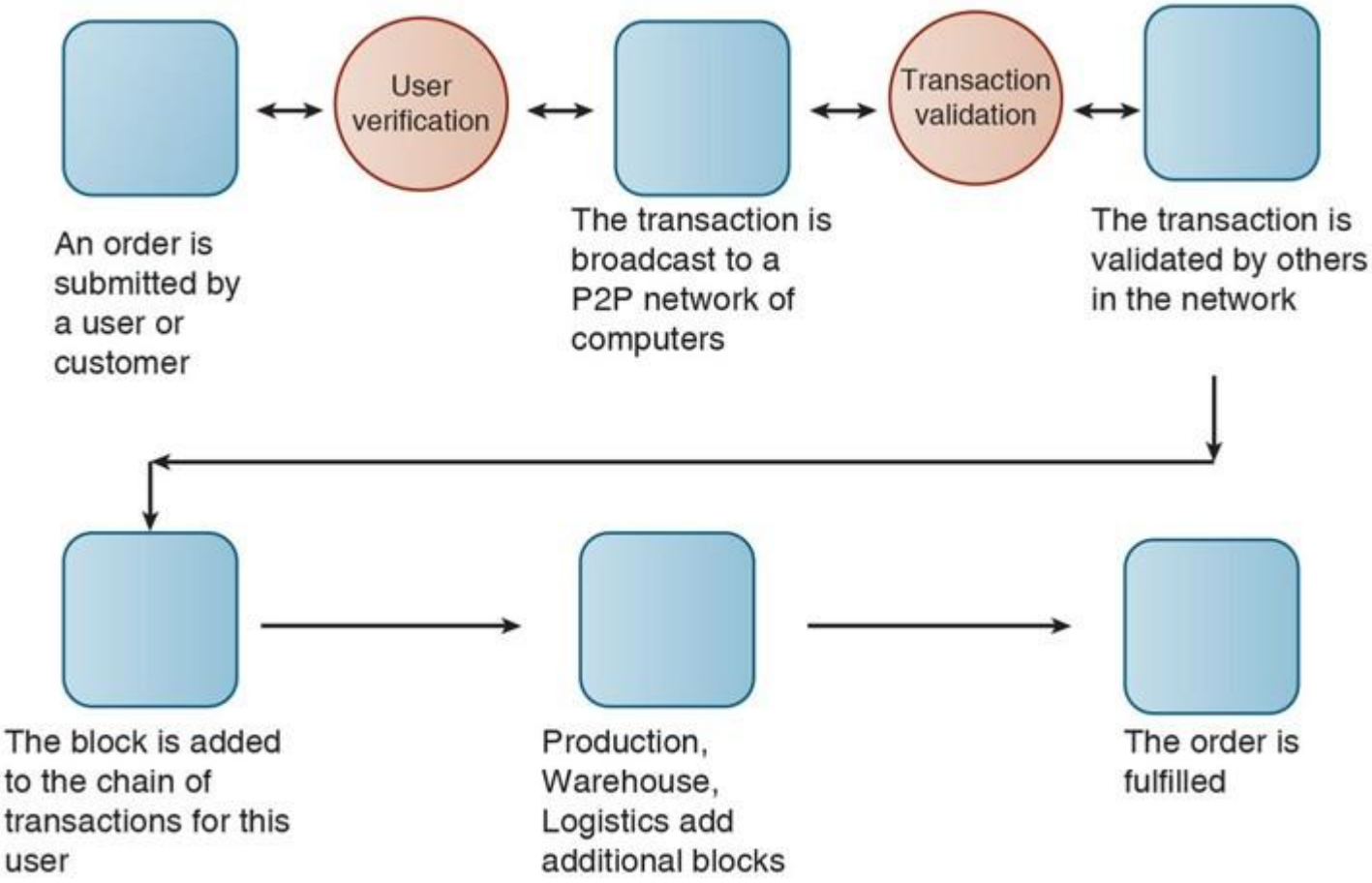
Describe Non-Relational Database, Cloud Database, and Blockchain Capabilities (2 of 3)

- Cloud databases
  - Appeal to start-ups, smaller businesses
  - Amazon Relational Database Service, Microsoft S Q L Azure
  - Private clouds
- Distributed databases
  - Stored in multiple physical locations
  - Example: Google Spanner

Describe Non-Relational Database, Cloud Database, and Blockchain Capabilities (3 of 3)

- Blockchain
  - Distributed ledgers in a peer-to-peer distributed database
  - Maintains a growing list of records and transactions shared by all
  - Encryption used to identify participants and transactions
  - Used for financial transactions, supply chain, and medical records
  - Foundation of Bitcoin, and other crypto currencies

Figure 6.12 How Blockchain Works



Describe Tools Used to Analyze Data in Databases

- Big data
  - Massive sets of unstructured/semi-structured data from web traffic, social media, sensors, and so on
  - Volumes too great for typical D B M S
    - Petabytes, exabytes of data
- Big data and AI have a synergistic relationship
  - AI requires a massive amount of data for training
  - Machine learning is able to use the information provided by big data to generate more valuable business insights

Business Intelligence Infrastructure (1 of 4)

- Array of tools for obtaining information from separate systems and from big data
  - Data warehouse
  - Data mart
  - Data lake
  - Hadoop
  - In-memory computing
  - Data analytics platform

Business Intelligence Infrastructure
(2 of 4)

- Data warehouse
  - Database that 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 mart
  - Subset of data warehouse
  - Typically focus on single subject or line of business

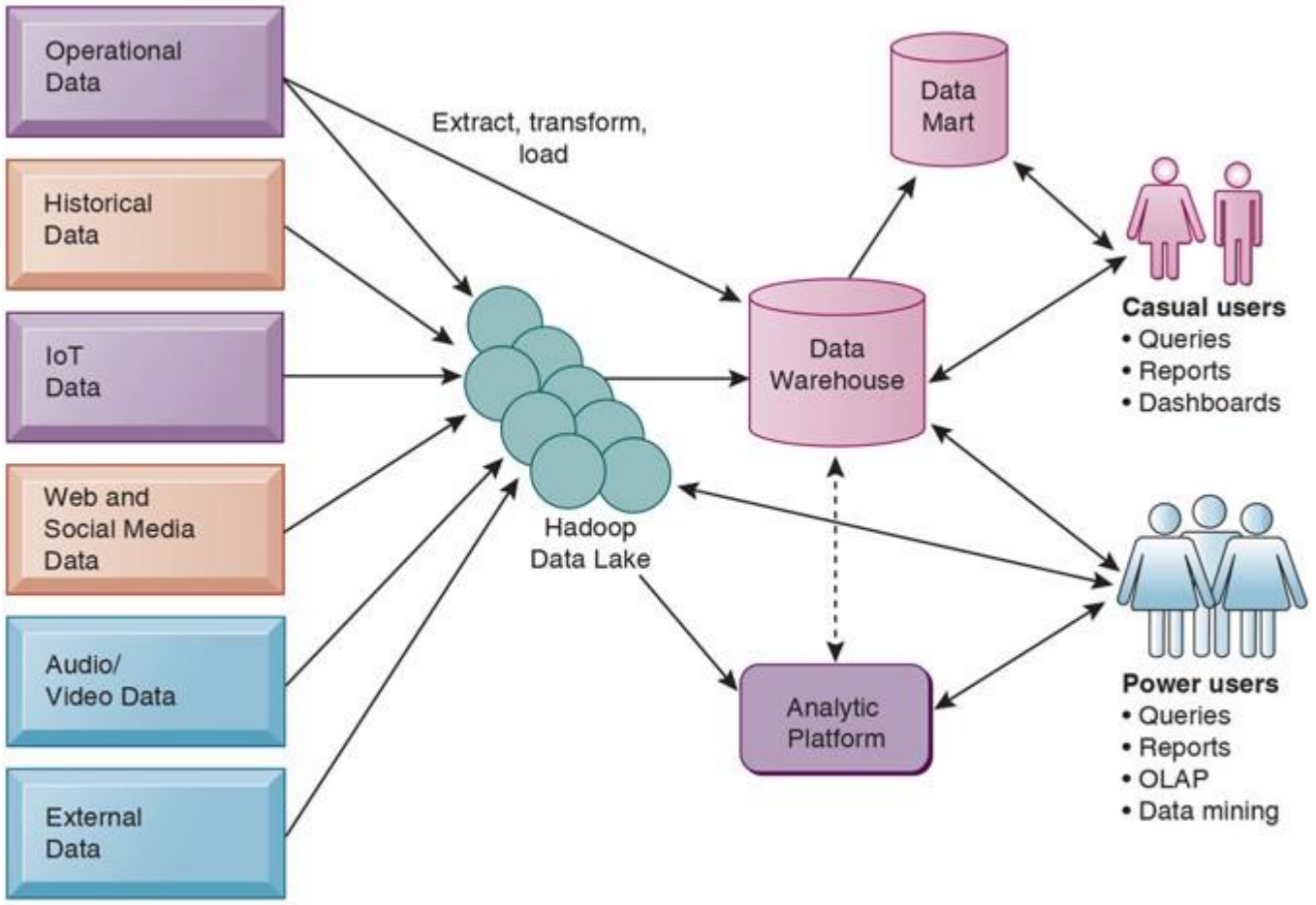
Business Intelligence Infrastructure
(3 of 4)

- Data lake
  - Repository for raw unstructured data or structured data that, for the most part, has not yet been analyzed
- Hadoop
  - Open-source software framework managed by the Apache Software Foundation
  - Enables distributed parallel processing of big data across inexpensive computers

Business Intelligence Infrastructure
(4 of 4)

- In-memory computing
  - Uses computers main memory (R A M) for data storage to avoid delays in retrieving data from disk storage
  - Can reduce hours/days of processing to seconds
  - Requires optimized hardware
- Data-analytic platform
  - High-speed platform using both relational and non-relational tools optimized for large datasets

Figure 6.13 Contemporary Business Intelligence Technology Infrastructure



Analytical Tools: Relationships, Patterns, Trends (1 of 2)

- Tools for consolidating, analyzing, and providing access to vast amounts of data to help users make better business decisions
  - Online analytical processing (O L A P)
    - Supports multidimensional data analysis, enabling users to view the same data in different ways using multiple dimensions
  - Data mining
    - Provides insights into corporate data that cannot be obtained with OLAP

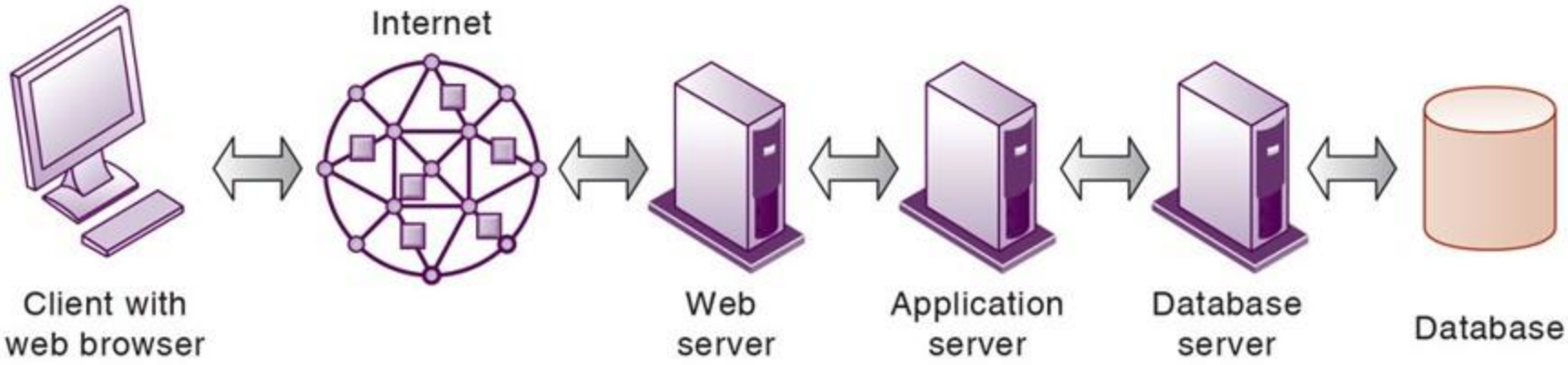
Analytical Tools: Relationships, Patterns, Trends (2 of 2)

- Tools for consolidating, analyzing, and providing access to vast amounts of data to help users make better business decisions
  - Text mining
    - Tools to extract key elements from unstructured natural language text, discover patterns and relationships, and summarize the information
    - Sentiment analysis
      - Involves mining comments in an email, etc. to detect favorable and unfavorable opinions about specific subjects
  - Web mining
    - Discovery and analysis of useful patterns and information from the web

Databases and the Web

- Many companies use the web to make some internal databases available to customers or partners
- Typical configuration includes
  - Web server
  - Application server/middleware/scripts
  - Database server (hosting D B M S)
- Advantages of using the 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

Figure 6.15 Linking Internal Databases to the Web



Discuss Data Governance, Data Quality Assurance, and Data Management (1 of 3)

- Data governance
  - Encompasses policies and procedures through which data can be managed as an organizational resource.
  - Establishes rules for sharing, disseminating, acquiring, standardizing, classifying and inventorying information
    - Example: Firm information policy that specifies that only selected members of a particular department can view certain information

Discuss Data Governance, Data Quality Assurance, and Data Management (2 of 3)

- Data quality
  - More than 25 percent of critical data in Fortune 1000 company databases are inaccurate or incomplete
- Before new database is in place, a firm must:
  - Identify and correct faulty data
  - Establish better routines for editing data once database in operation

Discuss Data Governance, Data Quality Assurance, and Data Management (3 of 3)

- Data quality audit
  - Analysis of data quality often begins with a data quality audit
    - A structured survey of the accuracy and level of completeness of the data in an information system
- Data cleansing
  - Also known as data scrubbing
  - Consists of processes for detecting and correcting data in a database or file that are incorrect, incomplete, improperly formatted, or redundant

How AI Improves Data Management

- Data management is crucial for creating an environment where data can be shared and utilized across the entire organization
  - However, managing all of the organization's data is extremely labor-intensive
  - Artificial intelligence is now being used to improve data management in the following areas: classification, cataloguing, improving data quality, security, data integration, and data analysis

Copyright