Foundations of BI: Databases and Information Management Managing Knowledge

Part II & III

Information Technology Infrastructure

Key System Applications for the Digital Age

Paulo Faroleiro

Information Systems

Part II & III
Information Technology Infrastructure
Key System Applications for the Digital Age

Learning Objectives

1. Describe the role of knowledge management and knowledge management programs in business.
2. Describe the types of systems used for enterprise-wide knowledge management and how they provide value for businesses.
3. What are the problems of managing data resources in a traditional file environment?
4. What are the major capabilities of database management systems (DBMS), and why is a relational DBMS so powerful?
5. What are the principal tools and technologies for accessing information from databases to improve business performance and decision making?
6. Why are information policy, data administration, and data quality assurance essential for managing the firm's data resources?
7. Describe the major types of knowledge work systems and how they provide value for firms.
8. Describe the business benefits of using intelligent techniques for knowledge management.

The Knowledge Management Landscape

- **Knowledge management systems among fastest growing areas of software investment**
- **Information economy**
  - 37% U.S. labor force: knowledge and information workers
  - 45% U.S. GDP from knowledge and information sectors
- **Substantial part of a firm's stock market value is related to intangible assets: knowledge, brands, reputations, and unique business processes**
- **Well-executed knowledge-based projects can produce extraordinary ROI**

The Knowledge Management Landscape

- **Important dimensions of knowledge**
  - **Knowledge is a firm asset.**
    - Intangible
    - Creation of knowledge from data, information, requires organizational resources
    - As it is shared, experiences network effects
  - **Knowledge has different forms.**
    - May be *explicit* (documented) or *tacit* (residing in minds)
    - Know-how, craft, skill
    - How to follow procedure
    - Knowing why things happen (causality)

The Knowledge Management Landscape

- **Important dimensions of knowledge (cont.)**
  - **Knowledge has a location.**
    - Cognitive event
    - 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 circumstances to use certain tool

The Knowledge Management Landscape

- **To transform information into knowledge, firm must expend additional resources to discover patterns, rules, and contexts where knowledge works**
- **Wisdom:**
  - Collective and individual experience of applying knowledge to solve problems
  - Involves where, when, and how to apply knowledge
- **Knowing how to do things effectively and efficiently in ways others cannot duplicate is prime source of profit and competitive advantage**
  - For example, Having a unique build-to-order production system

The Knowledge Management Landscape

- **Organizational learning**
  - **Process in which organizations learn**
    - Gain experience through collection of data, measurement, trial and error, and feedback
    - Adjust behavior to reflect experience
      - Create new business processes
      - Change patterns of management decision making

The Knowledge Management Landscape

- **Knowledge management**
  - Set of business processes developed in an organization to create, store, transfer, and apply knowledge
- **Knowledge management value chain:**
  - Each stage adds value to raw data and information as they are transformed into usable knowledge
  - Knowledge acquisition
  - Knowledge storage
  - Knowledge dissemination
  - Knowledge application

The Knowledge Management Landscape

- **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 TPS and external sources

The Knowledge Management Landscape

- **Knowledge management value chain (cont.)**
  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.

The Knowledge Management Landscape

- **Knowledge management value chain (cont.)**
  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.

The Knowledge Management Landscape

- **Knowledge management value chain (cont.)**

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

The Knowledge Management Value Chain

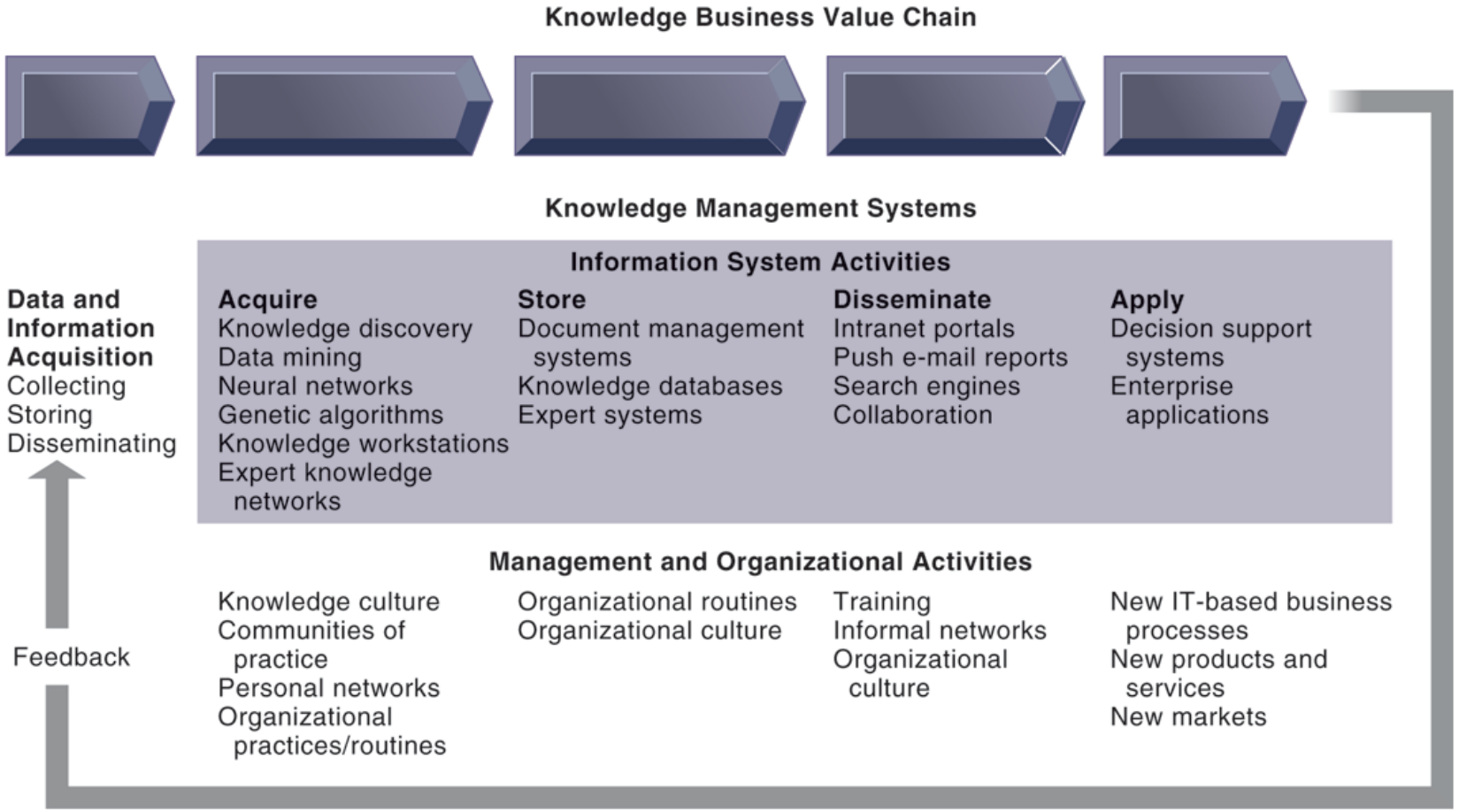


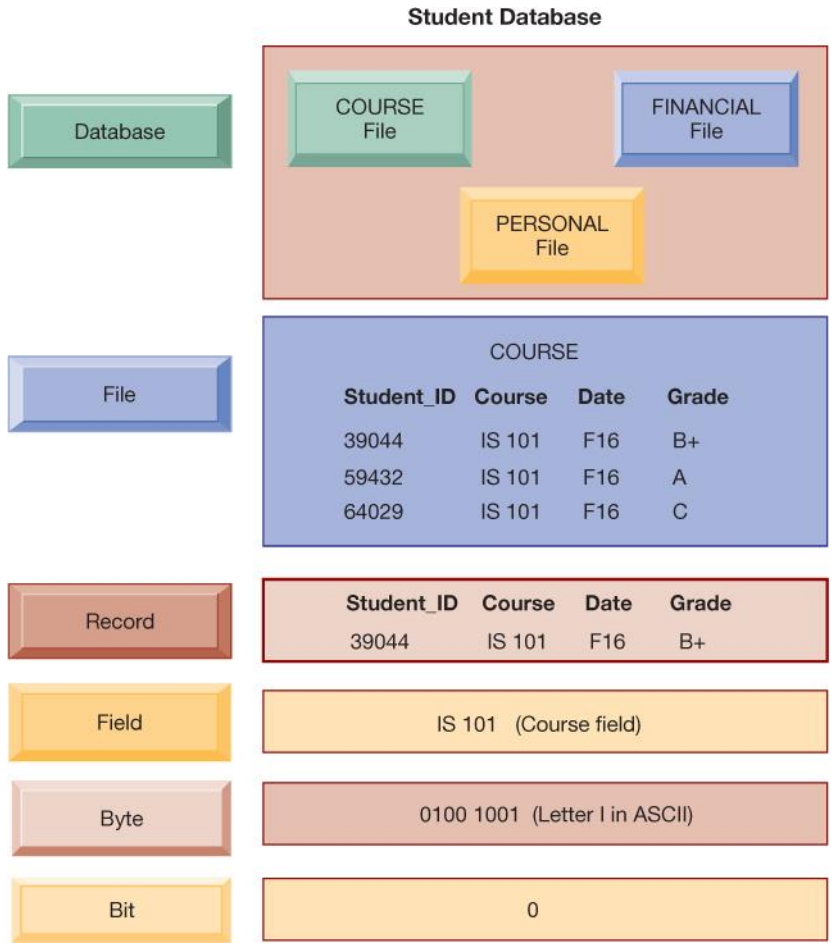
FIGURE 11-1 Knowledge management today involves both information systems activities and a host of enabling management and organizational activities.

The Knowledge Management Landscape

- **Organizational roles and responsibilities**
  - **Chief knowledge officer executives**
  - **Dedicated staff / knowledge managers**
  - **Communities of practice (COPs)**
    - Informal social networks of professionals and employees within and outside firm who have similar work-related activities and interests
    - Activities include education, online newsletters, sharing experiences and techniques
    - Facilitate reuse of knowledge, discussion
    - Reduce learning curves of new employees

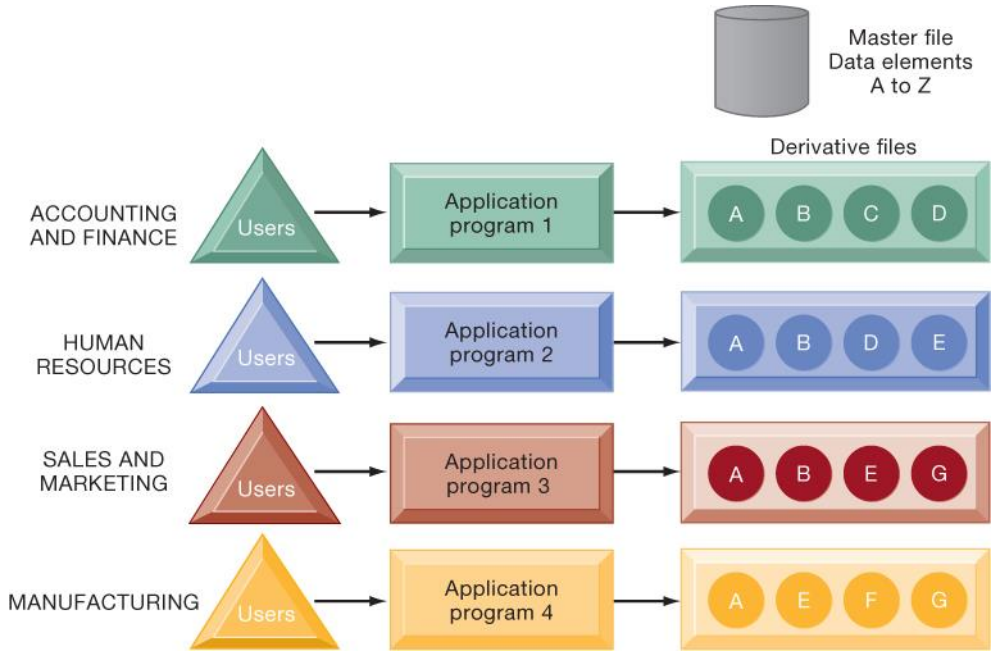
File Organization Terms and Concepts

- **Database: Group of related files**
- **File: Group of records of same type**
- **Record: Group of related fields**
- **Field: Group of characters as word(s) or number(s)**
- **Entity: Person, place, thing on which we store information**
- **Attribute: Each characteristic, or quality, describing entity**



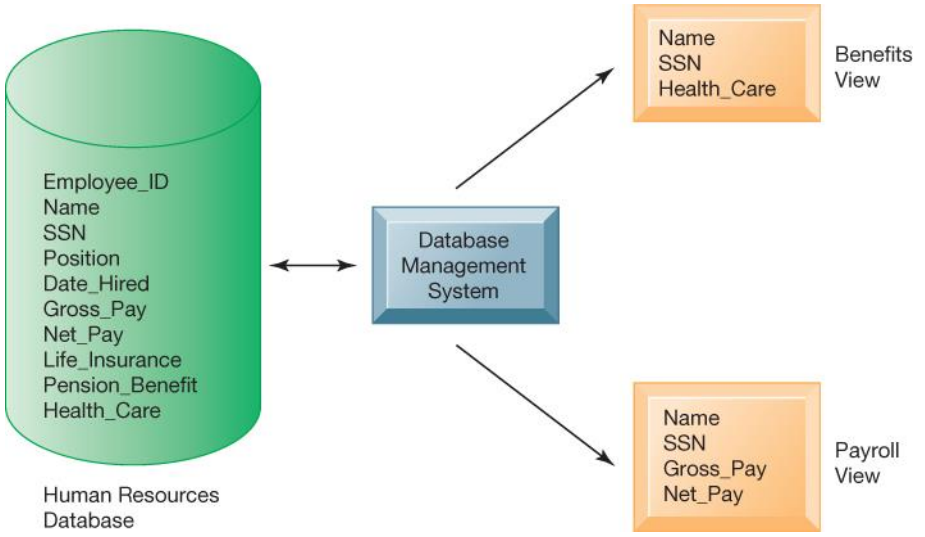
Problems with the Traditional File Environment

- **Files maintained separately by different departments**
- **Data redundancy**
- **Data inconsistency**
- **Program-data dependence**
- **Lack of flexibility**
- **Poor security**
- **Lack of data sharing and availability**



Database Management Systems

- **Database**
  - **Serves many applications by centralizing data and controlling redundant data**
- **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 centrally manage data and data security



Non-relational Databases and Databases in the Cloud

- **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**
- **Databases in the cloud**
  - **Appeal to start-ups, smaller businesses**
  - **Amazon Relational Database Service, Microsoft SQL Azure**
  - **Private clouds**

The Challenge of Big Data

- **Big data**
  - **Massive sets of unstructured/semi-structured data from web traffic, social media, sensors, and so on**
- **Volumes too great for typical DBMS**
  - **Petabytes, exabytes of data**
- **Can reveal more patterns, relationships and anomalies**
- **Requires new tools and technologies to manage and analyze**

Business Intelligence Infrastructure (1 of 3)

- **Array of tools for obtaining information from separate systems and from big data**
- **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**

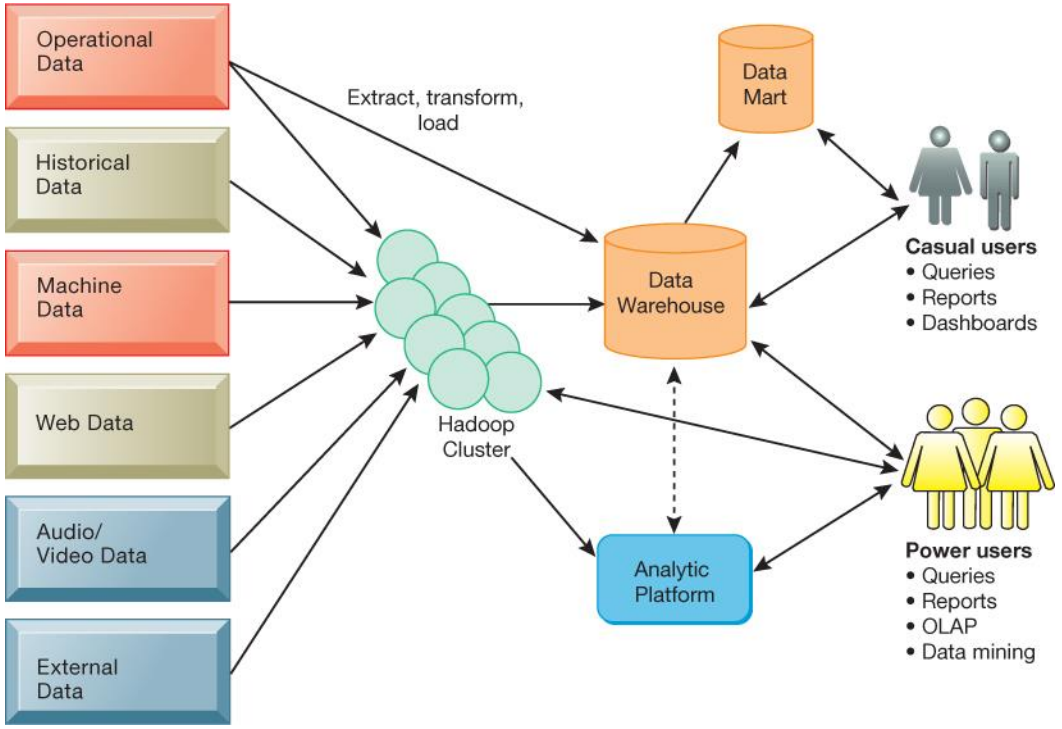
Business Intelligence Infrastructure (2 of 3)

- **Data marts**
  - **Subset of data warehouse**
  - **Typically focus 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 Yahoo, NextBio**

Business Intelligence Infrastructure (3 of 3)

- **In-memory computing**
  - **Used in big data analysis**
  - **Uses 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**

Contemporary Business Intelligence Infrastructure/Architecture



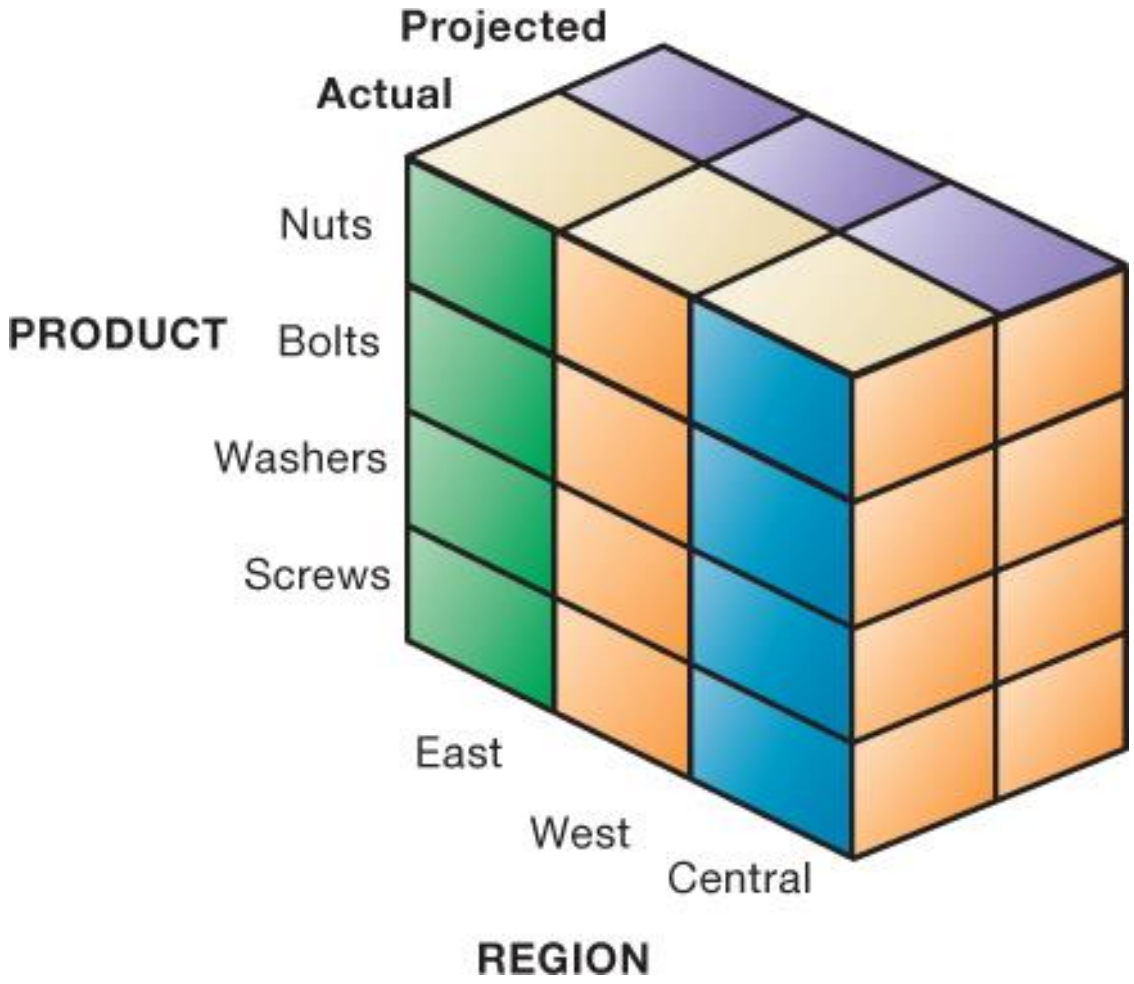
Analytical Tools: Relationships, Patterns, Trends

- **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 the 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**
- **Types of information obtainable from data mining:**
  - **Associations**
  - **Sequences**
  - **Classification**
  - **Clustering**
  - **Forecasting**

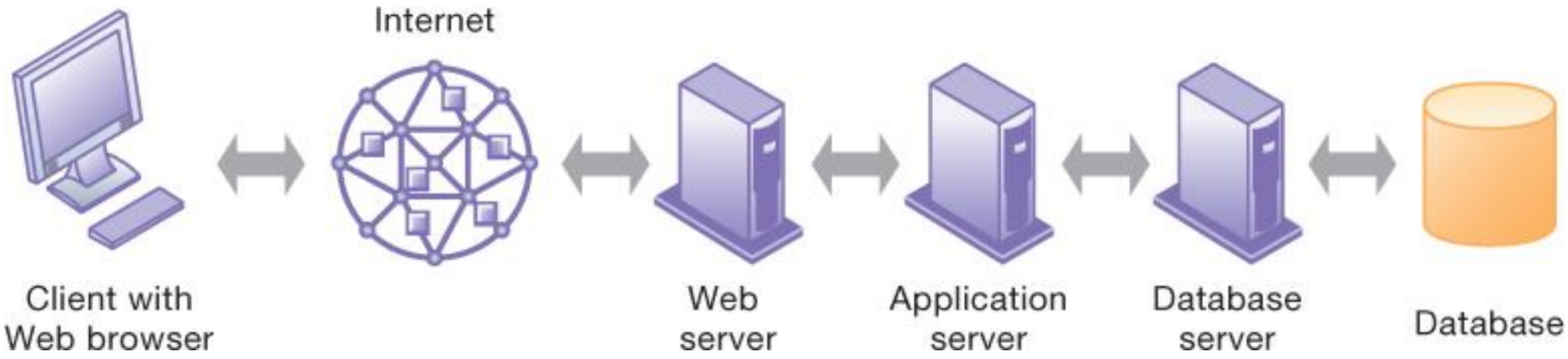
Text Mining and Web Mining

- **Text mining**
  - **Extracts key elements from large unstructured data sets**
  - **Sentiment analysis software**
- **Web mining**
  - **Discovery and analysis of useful patterns and information from web**
  - **Web content mining**
  - **Web structure mining**
  - **Web usage mining**

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/CGI scripts**
  - **Database server (hosting DBMS)**
- **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**

Linking Internal Databases to the Web



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**

Ensuring 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
- Data quality audit
- Data cleansing

The Knowledge Management Landscape

- **Three major types of knowledge management systems:**
  1. **Enterprise-wide knowledge management systems**
     - General-purpose firm-wide efforts to collect, store, distribute, and apply digital content and knowledge
  2. **Knowledge work systems (KWS)**
     - Specialized systems built for engineers, scientists, other knowledge workers charged with discovering and creating new knowledge
  3. **Intelligent techniques**
     - Diverse group of techniques such as data mining used for various goals: discovering knowledge, distilling knowledge, discovering optimal solutions

MAJOR TYPES OF KNOWLEDGE MANAGEMENT SYSTEMS

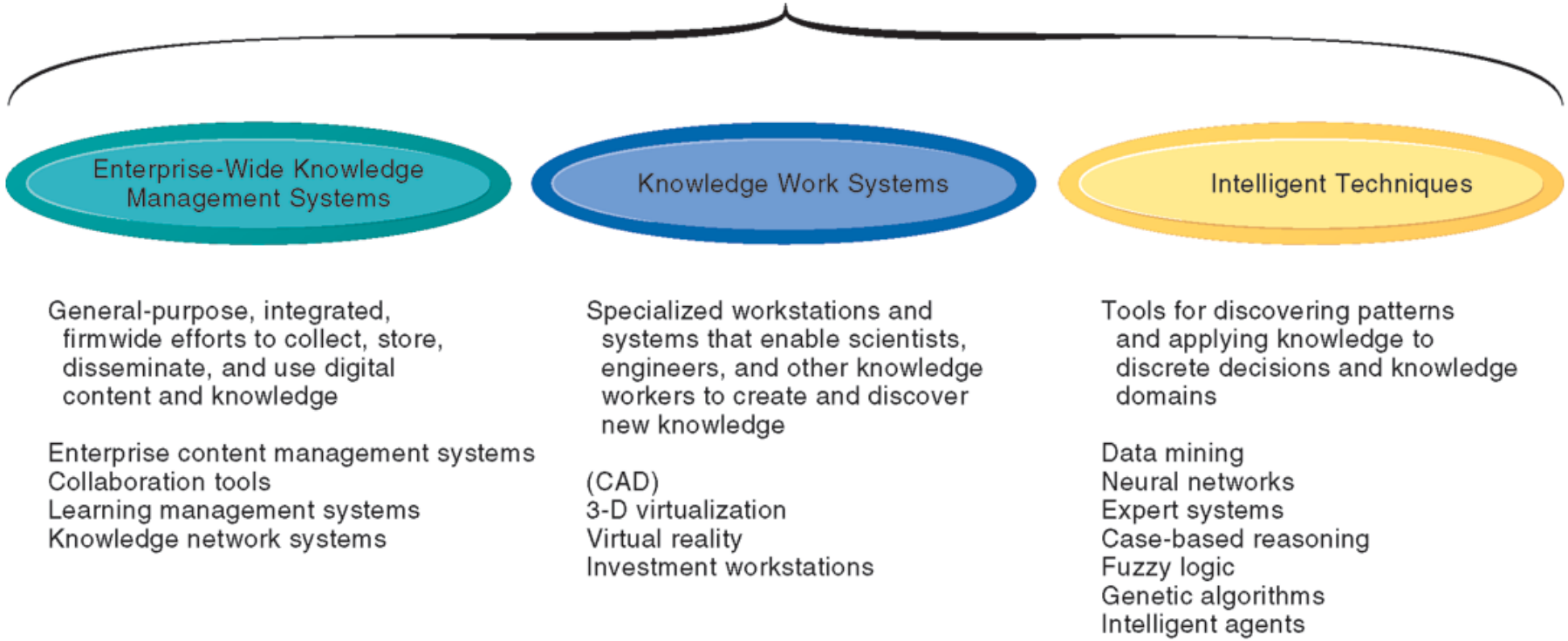


FIGURE 11-2 There are three major categories of knowledge management systems, and each can be broken down further into more specialized types of knowledge management systems.

Enterprise-Wide Knowledge Management Systems

- **Three major types of knowledge in enterprise**
  1. **Structured documents**
     - Reports, presentations
     - Formal rules
  2. **Semistructured documents**
     - E-mails, videos
  3. **Unstructured, tacit knowledge**
- **80% of an organization's business content is semistructured or unstructured**

Enterprise-Wide Knowledge Management Systems

- **Enterprise content management systems**
  - **Help capture, store, retrieve, distribute, preserve**
    - Documents, reports, best practices
    - Semistructured knowledge (e-mails)
  - **Bring in external sources**
    - News feeds, research
  - **Tools for communication and collaboration**
    - Blogs, wikis, and so on

AN ENTERPRISE CONTENT MANAGEMENT SYSTEM

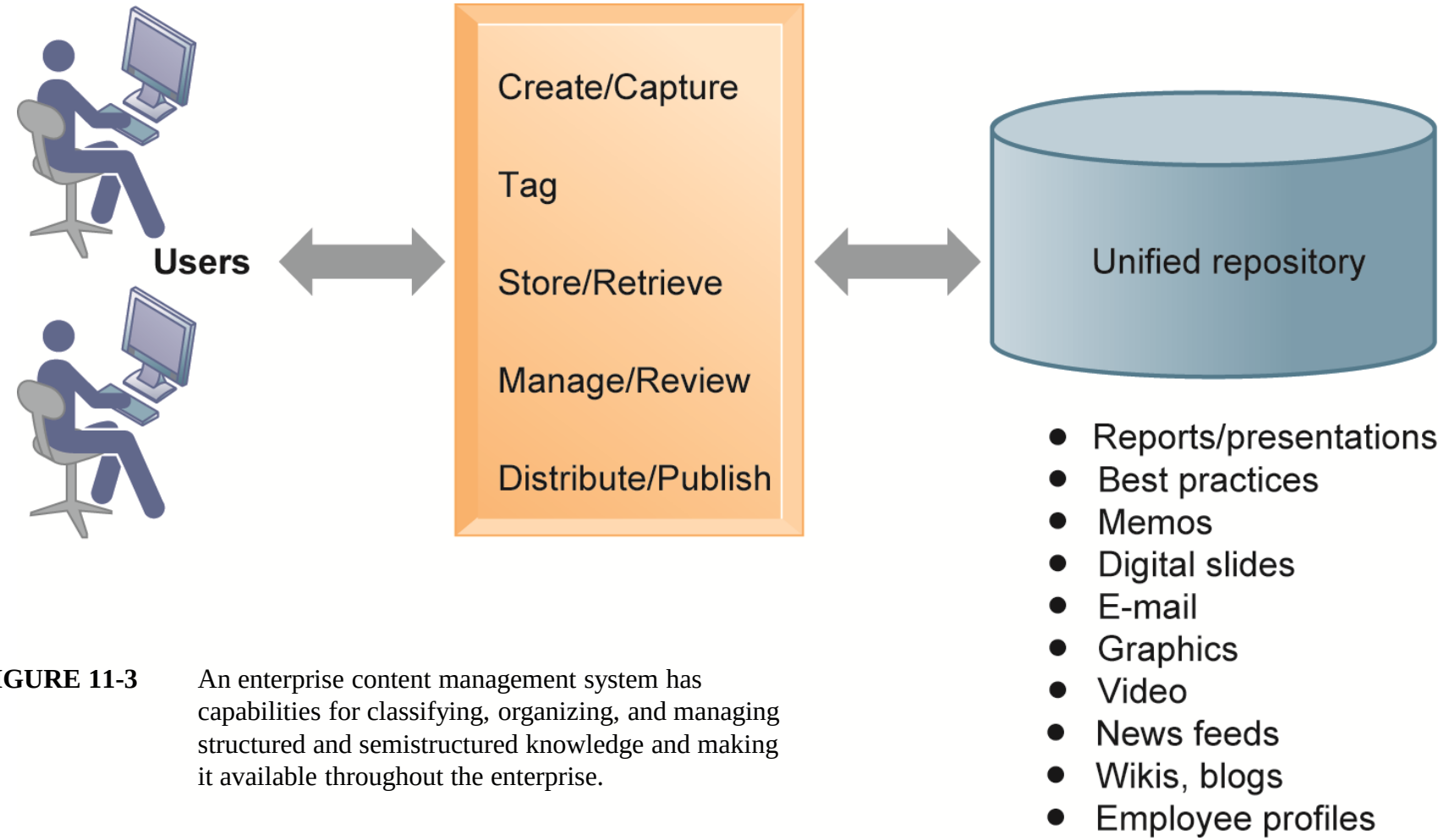


FIGURE 11-3 An enterprise content management system has capabilities for classifying, organizing, and managing structured and semistructured knowledge and making it available throughout the enterprise.

Enterprise-Wide Knowledge Management Systems

- **Enterprise content management systems**
  - **Key problem—Developing taxonomy**
    - Knowledge objects must be tagged with categories for retrieval
  - **Digital asset management systems**
    - Specialized content management systems for classifying, storing, managing unstructured digital data
    - Photographs, graphics, video, audio

Enterprise-Wide Knowledge Management Systems

- **Knowledge network systems**
  - **Provide online directory of corporate experts in well-defined knowledge domains**
  - **Search tools enable employees to find appropriate expert in a company**
  - **Hivemine's AskMe**
    - Includes repositories of expert-generated content
  - **Some knowledge networking capabilities included in leading enterprise content management and collaboration products**

Enterprise-Wide Knowledge Management Systems

- **Collaboration and social tools**
  - **Social bookmarking**
    - Sharing and tagging bookmarks
  - **Folksonomies**
    - User-created taxonomies for tagging
  - **Examples:**
    - Delicious
    - Slashdot
    - Pinterest

Enterprise-Wide Knowledge Management Systems

- **Learning management systems (LMS)**
  - **Provide tools for management, delivery, tracking, and assessment of various types of employee learning and training**
  - **Support multiple modes of learning**
    - CD-ROM, Web-based classes, online forums, live instruction, and so on
  - **Automates selection and administration of courses**
  - **Assembles and delivers learning content**
  - **Measures learning effectiveness**

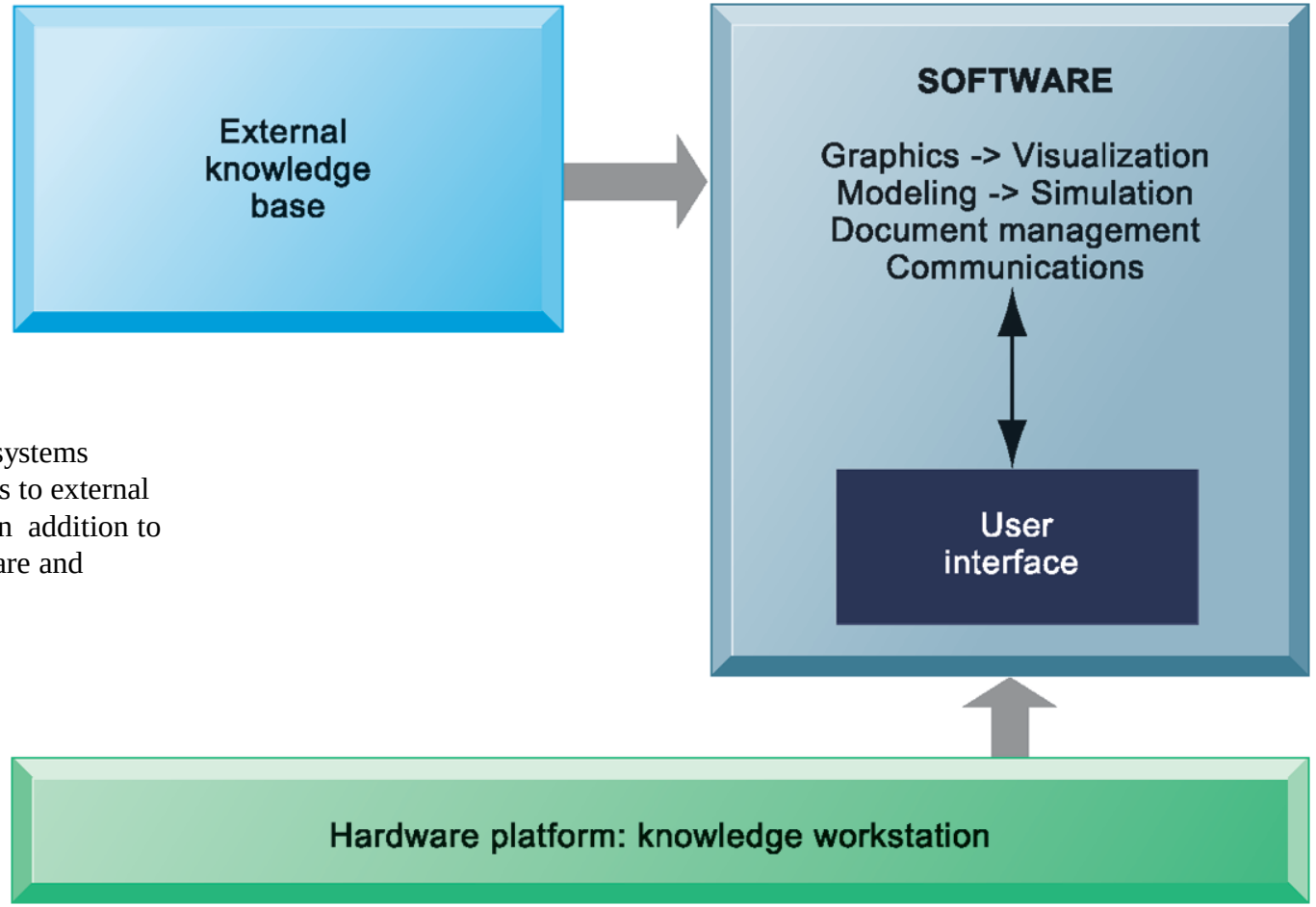
Knowledge Work Systems

- **Knowledge work systems**
  - **Systems for knowledge workers to help create new knowledge and integrate that knowledge into business**
- **Knowledge workers**
  - **Researchers, designers, architects, scientists, engineers who create knowledge for the organization**
  - **Three key roles:**
    1. Keeping organization current in knowledge
    2. Serving as internal consultants regarding their areas of expertise
    3. Acting as change agents, evaluating, initiating, and promoting change projects

Knowledge Work Systems

- **Requirements of knowledge work systems**
  - **Sufficient computing power for graphics, complex calculations**
  - **Powerful graphics and analytical tools**
  - **Communications and document management**
  - **Access to external databases**
  - **User-friendly interfaces**
  - **Optimized for tasks to be performed (design engineering, financial analysis)**

REQUIREMENTS OF KNOWLEDGE WORK SYSTEMS



Knowledge work systems require strong links to external knowledge bases in addition to specialized hardware and software.

FIGURE 11-4

Knowledge Work Systems

- **Examples of knowledge work systems**
  - **CAD (computer-aided design):**
    - Creation of engineering or architectural designs
    - 3-D printing
  - **Virtual reality systems:**
    - Simulate real-life environments
    - 3-D medical modeling for surgeons
    - Augmented reality (AR) systems
    - VRML
  - **Investment workstations:**
    - Streamline investment process and consolidate internal, external data for brokers, traders, portfolio managers

Intelligent Techniques

- **Intelligent techniques:** Used to capture individual and collective knowledge and to extend knowledge base
  - **To capture tacit knowledge:** Expert systems, case-based reasoning, fuzzy logic
  - **Knowledge discovery**: Neural networks and data mining
  - **Generating solutions to complex problems:** Genetic algorithms
  - **Automating tasks:** Intelligent agents
- **Artificial intelligence (AI) technology:**
  - **Computer-based systems that emulate human behavior**

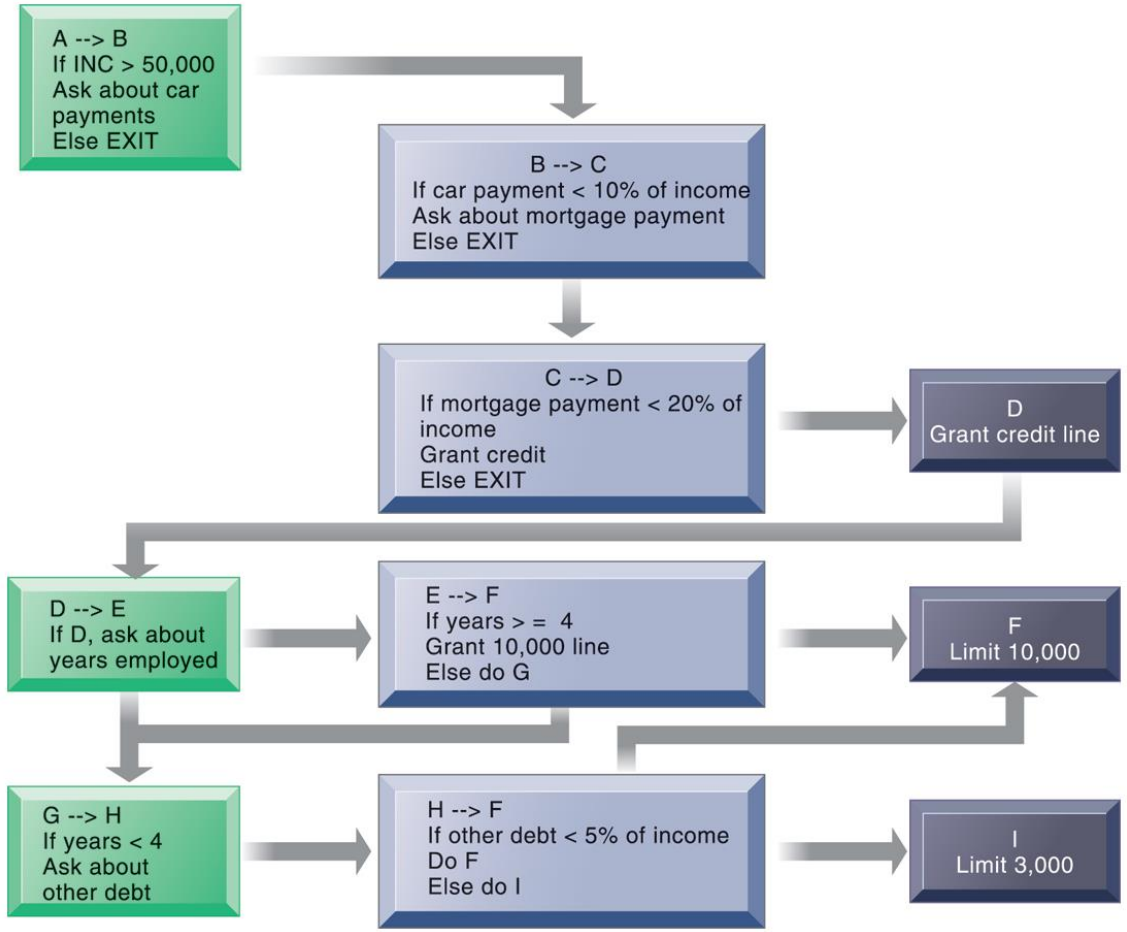
Intelligent Techniques

- **Expert systems:**
  - **Capture tacit knowledge in very specific and limited domain of human expertise**
  - **Capture knowledge of skilled employees as set of rules in software system that can be used by others in organization**
  - **Typically perform limited tasks that may take a few minutes or hours, for example:**
    - Diagnosing malfunctioning machine
    - Determining whether to grant credit for loan
  - **Used for discrete, highly structured decision making**

RULES IN AN EXPERT SYSTEM

An expert system contains a number of rules to be followed. The rules are interconnected; the number of outcomes is known in advance and is limited; there are multiple paths to the same outcome; and the system can consider multiple rules at a single time. The rules illustrated are for simple credit-granting expert systems.

FIGURE 11-5



Intelligent Techniques

- **How expert systems work**
  - **Knowledge base: Set of hundreds or thousands of rules**
  - **Inference engine: Strategy used to search knowledge base**
    - **Forward chaining:** Inference engine begins with information entered by user and searches knowledge base to arrive at conclusion
    - **Backward chaining:** Begins with hypothesis and asks user questions until hypothesis is confirmed or disproved

INFERENCE ENGINES IN EXPERT SYSTEMS

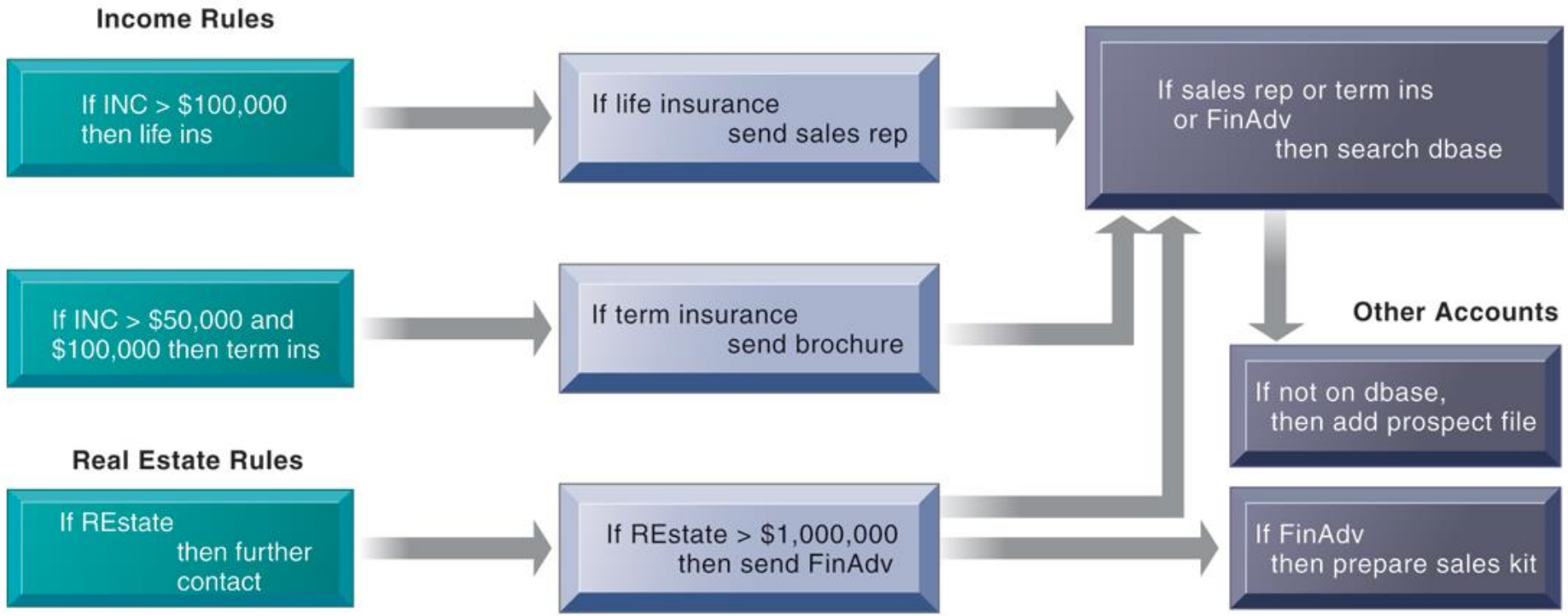


FIGURE 11-6 An inference engine works by searching through the rules and "firing" those rules that are triggered by facts gathered and entered by the user. Basically, a collection of rules is similar to a series of nested IF statements in a traditional software program; however, the magnitude of the statements and degree of nesting are much greater in an expert system.

Intelligent Techniques

- **Successful expert systems:**
  - Con-Way Transportation built expert system to automate and optimize planning of overnight shipment routes for nationwide freight-trucking business
- **Most expert systems deal with problems of classification.**
  - Have relatively few alternative outcomes
  - Possible outcomes are known in advance
- **Many expert systems require large, lengthy, and expensive development and maintenance efforts.**
  - Hiring or training more experts may be less expensive

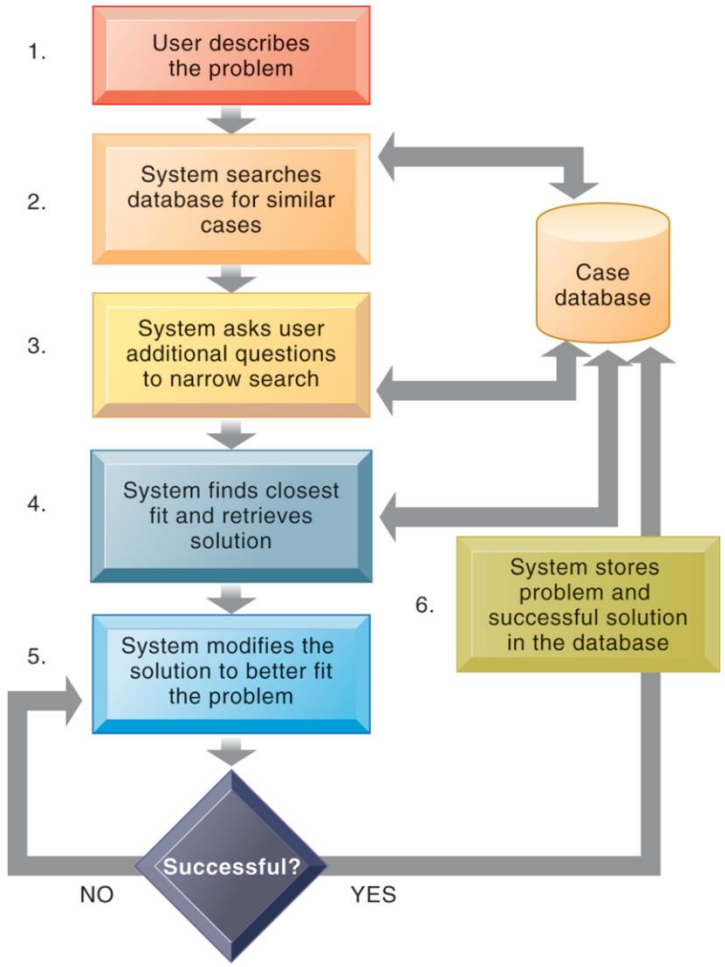
Intelligent Techniques

- **Case-based reasoning (CBR)**
  - **Descriptions of past experiences of human specialists (cases), stored in knowledge base**
  - **System searches for cases with characteristics similar to new one and applies solutions of old case to new case**
  - **Successful and unsuccessful applications are grouped with case**
  - **Stores organizational intelligence: Knowledge base is continuously expanded and refined by users**
  - **CBR found in**
    - Medical diagnostic systems
    - Customer support

HOW CASE-BASED REASONING WORKS

Case-based reasoning represents knowledge as a database of past cases and their solutions. The system uses a six-step process to generate solutions to new problems encountered by the user.

FIGURE 11-7



Intelligent Techniques

- **Fuzzy logic systems**
  - **Rule-based technology that represents imprecision used in linguistic categories (e.g., "cold," "cool") that represent range of values**
  - **Describe a particular phenomenon or process linguistically and then represent that description in a small number of flexible rules**
  - **Provides solutions to problems requiring expertise that is difficult to represent with IF-THEN rules**
    - Autofocus in cameras
    - Detecting possible medical fraud
    - Sendai's subway system acceleration controls

FUZZY LOGIC FOR TEMPERATURE CONTROL

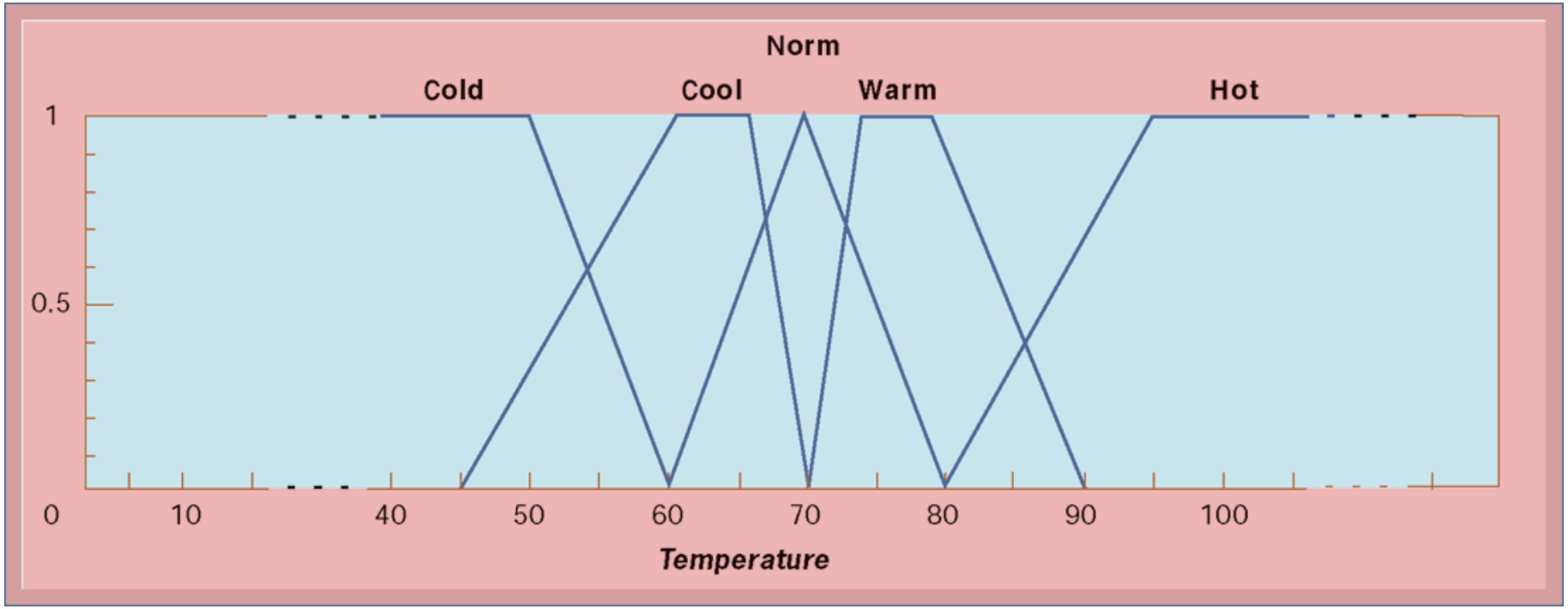


FIGURE 11-8 The membership functions for the input called temperature are in the logic of the thermostat to control the room temperature. Membership functions help translate linguistic expressions such as warm into numbers that the computer can manipulate.

Intelligent Techniques

- **Machine learning**
  - **How computer programs improve performance without explicit programming**
    - Recognizing patterns
    - Experience
    - Prior learnings (database)
  - **Contemporary examples**
    - Google searches
    - Recommender systems on Amazon, Netflix

Intelligent Techniques

- **Neural networks**
  - **Find patterns and relationships in massive amounts of data too complicated for humans to analyze**
  - **"Learn" patterns by searching for relationships, building models, and correcting over and over again**
  - **Humans "train" network by feeding it data inputs for which outputs are known, to help neural network learn solution by example**
  - **Used in medicine, science, and business for problems in pattern classification, prediction, financial analysis, and control and optimization**

HOW A NEURAL NETWORK WORKS

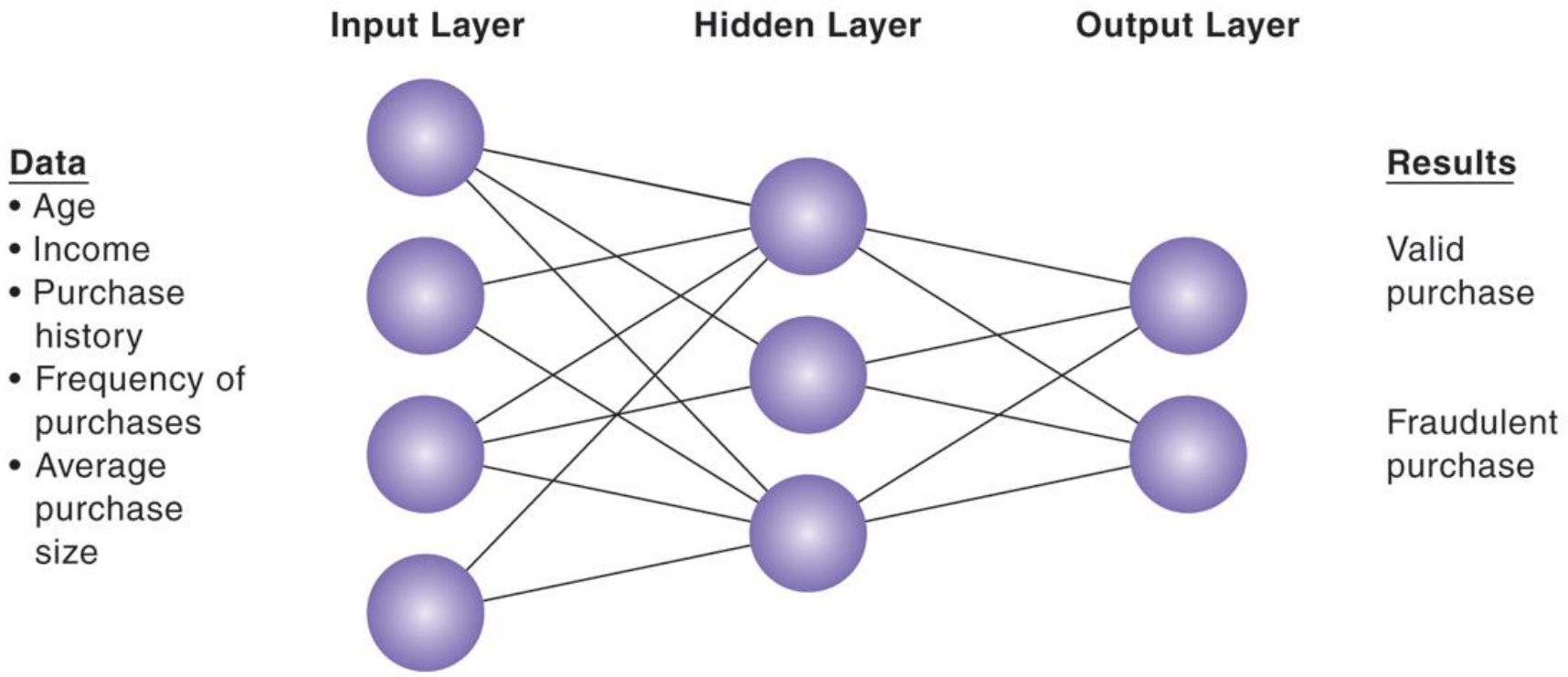


FIGURE 11-9 A neural network uses rules it "learns" from patterns in data to construct a hidden layer of logic. The hidden layer then processes inputs, classifying them based on the experience of the model. In this example, the neural network has been trained to distinguish between valid and fraudulent credit card purchases

Intelligent Techniques

- **Genetic algorithms**
  - **Useful for finding optimal solution for specific problem by examining very large number of possible solutions for that problem**
  - **Conceptually based on process of evolution**
    - Search among solution variables by changing and reorganizing component parts using processes such as inheritance, mutation, and selection
  - **Used in optimization problems (minimization of costs, efficient scheduling, optimal jet engine design) in which hundreds or thousands of variables exist**
  - **Able to evaluate many solution alternatives quickly**

THE COMPONENTS OF A GENETIC ALGORITHM

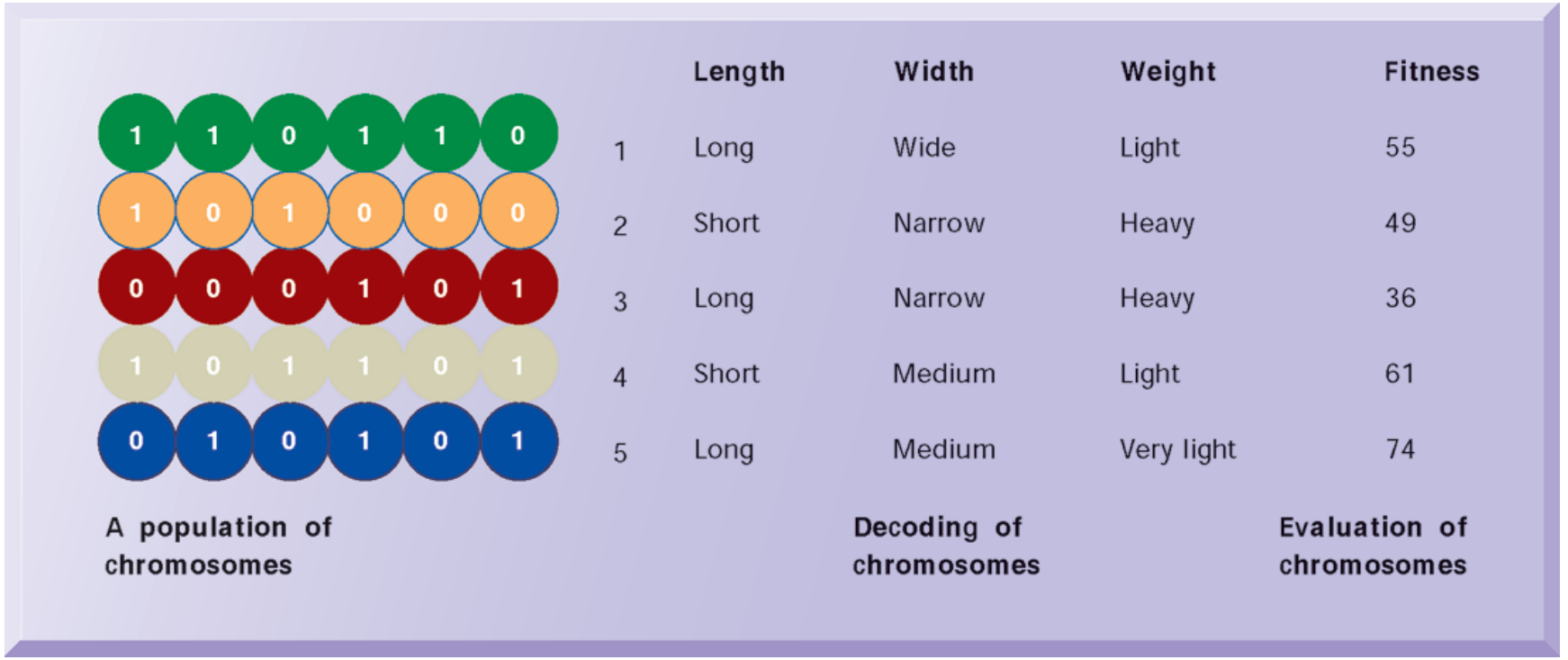


FIGURE 11-11 This example illustrates an initial population of "chromosomes," each representing a different solution. The genetic algorithm uses an iterative process to refine the initial solutions so that the better ones, those with the higher fitness, are more likely to emerge as the best solution.

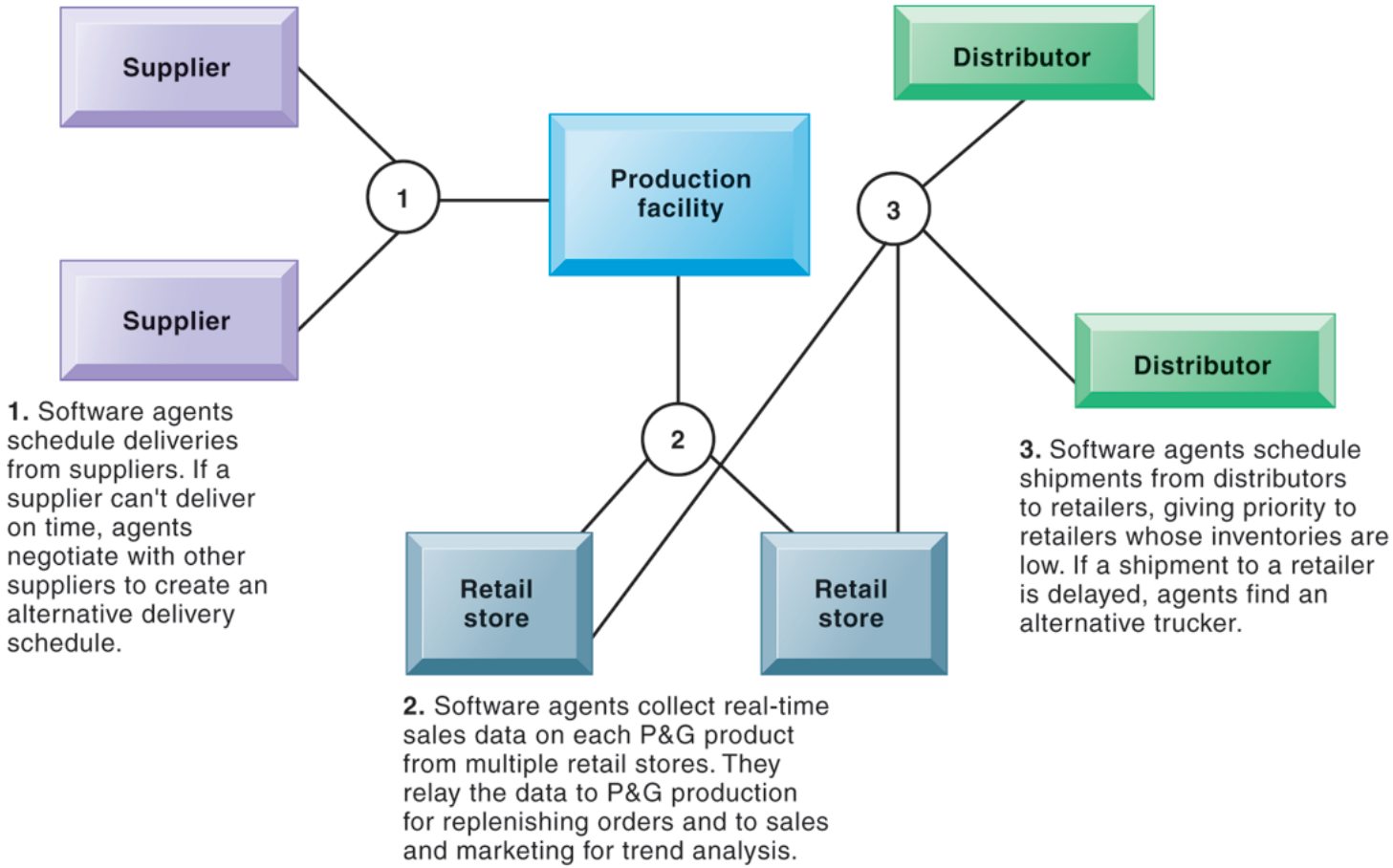
Intelligent Techniques

- **Intelligent agents**
  - **Work without direct human intervention to carry out specific, repetitive, and predictable tasks for user, process, or application**
    - Deleting junk e-mail
      - Finding cheapest airfare
  - **Use limited built-in or learned knowledge base**
    - Some are capable of self-adjustment, for example: Siri
  - **Agent-based modeling applications:**
    - Systems of autonomous agents
    - Model behavior of consumers, stock markets, and supply chains; used to predict spread of epidemics

INTELLIGENT AGENTS IN P&G'S SUPPLY CHAIN NETWORK

Intelligent agents are helping P&G shorten the replenishment cycles for products such as a box of Tide.

FIGURE 11-12



Intelligent Techniques

- **Hybrid AI systems**
  - **Genetic algorithms, fuzzy logic, neural networks, and expert systems integrated into single application to take advantage of best features of each**
  - **For example: Matsushita "neurofuzzy" washing machine that combines fuzzy logic with neural networks**

Part II

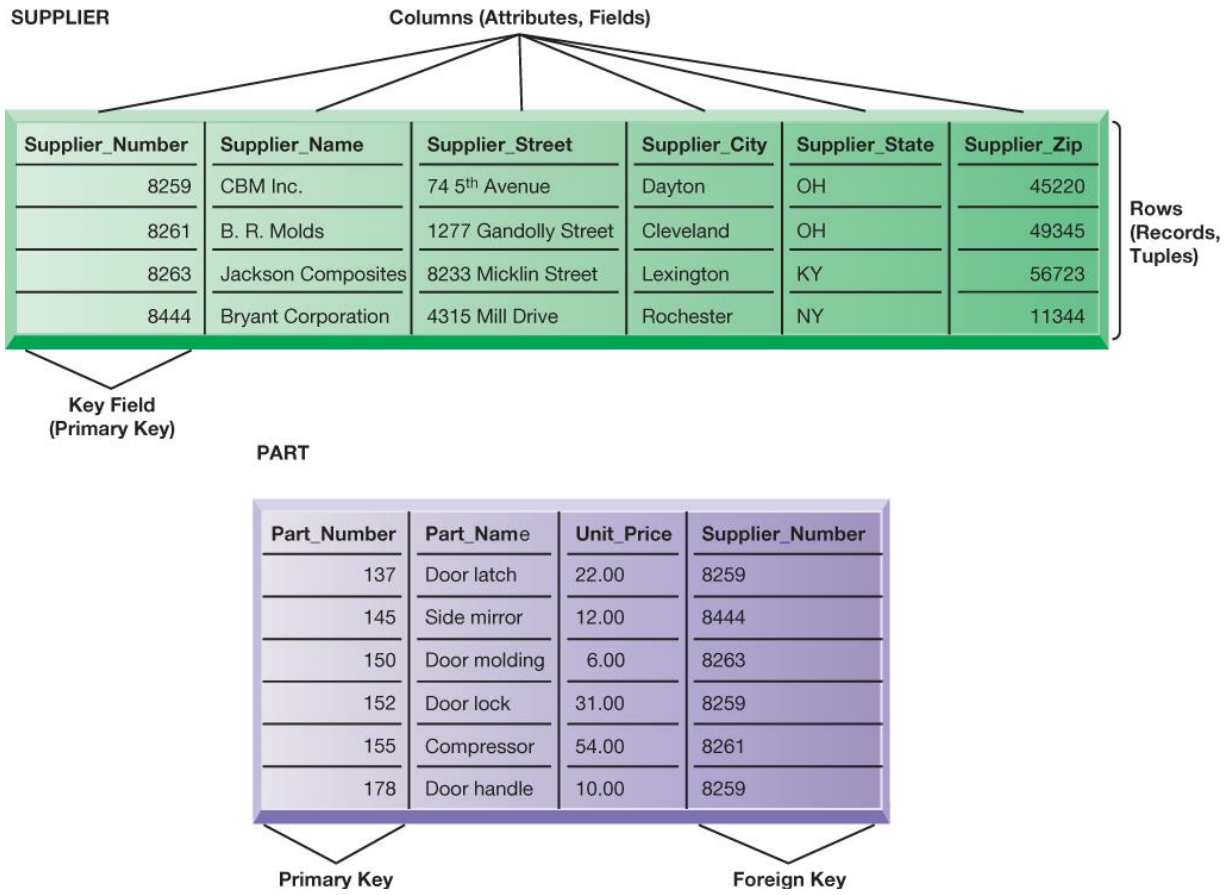
Information Technology Infrastructure

Next Steps

- Answer the Moodle quiz
- Prepare for last class

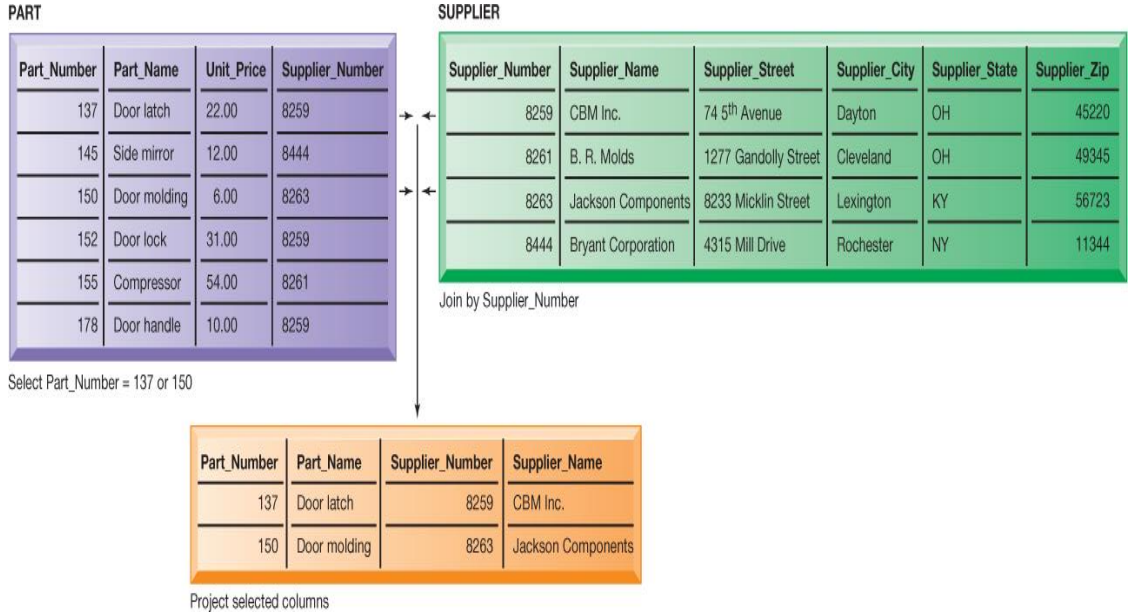
Relational DBMS

- **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**



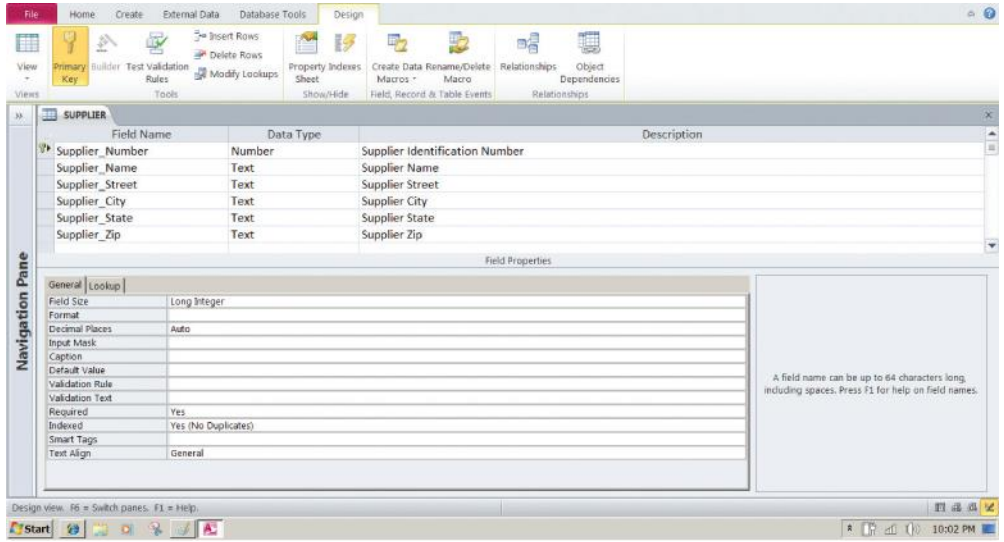
Operations of a Relational DBMS

- **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



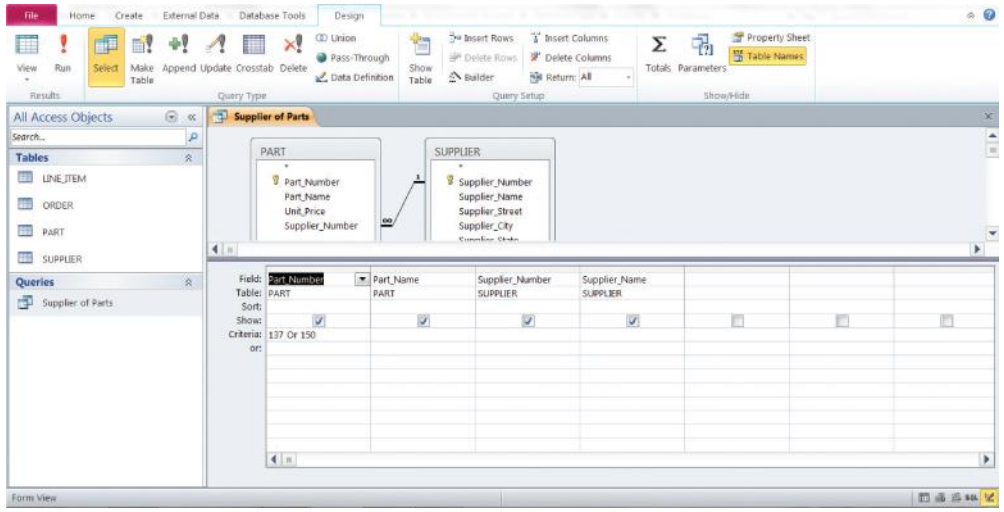
Capabilities of Database Management Systems

- **Data definition capability**
- **Data dictionary**
- **Querying and reporting**
  - **Data manipulation language**
    - Structured Query Language (SQL)
- **Many DBMS have report generation capabilities for creating polished reports (Microsoft Access)**



Example of Queries

```
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;
```



Designing Databases

- **Conceptual design vs. physical design**
- **Normalization**
  - **Streamlining complex groupings of data to minimize redundant data elements and awkward many-to-many relationships**
- **Referential integrity**
  - **Rules used by RDBMS to ensure relationships between tables remain consistent**
- **Entity-relationship diagram**
- **A correct data model is essential for a system serving the business well**

The Normalization Process: The Order Relation and the associated ER Diagram