IT445 – FINAL TESTBANK

DECISION SUPPORT SYSTEMS

CHAPTER 1

TRUE OR FALSE

Computerized support is only used for organizational decisions that are responses to external pressures, not for taking advantage of opportunities.	**False**
The complexity of today's business environment creates many new challenges for organizations, such as global competition, but creates few new opportunities in return.	**False**
In addition to deploying business intelligence (BI) systems, companies may also perform other actions to counter business pressures, such as improving customer service and entering business alliances.	**True**
The overwhelming majority of competitive actions taken by businesses today feature computerized information system support.	**True**
PCs and, increasingly, mobile devices are the most common means of providing managers with information to directly support decision making, instead of using IT staff intermediaries.	**True**
In today's business environment, creativity, intuition, and interpersonal skills are effective substitutes for analytical decision making.	**False**
In a four-step process for decision making, managers construct a model of the problem before they evaluate potential solutions.	**True**
Due to the fact that business environments are now more complex than ever, trial-and-error is an effective means of arriving at acceptable solutions.	**False**
Group collaboration software has proved generally ineffective at improving decision-making.	**False**
Due to the fact that organizations seek to store greater amounts of data than ever before, the cost per byte of computer-based data storage devices is rapidly rising.	**False**
Computerized information systems help decision makers overcome human cognitive limitations in assembling and processing varied information. However, this is of little use in most analytical applications.	**False**
In the Gorry and Scott-Morton framework of structured, semi-structured, and unstructured decisions, computerized decision support can bring benefits to unstructured decisions.	**True**
The term decision support system is a very specific term that implies the same tool, system, and development approach to most developers.	**False**
The access to data and ability to manipulate data (frequently including real-time data) are key elements of business intelligence (BI) systems.	**True**
One of the four components of BI systems, business performance management, is a collection of source data in the data warehouse.	**False**
Actionable intelligence is the primary goal of modern-day Business Intelligence (BI) systems vs. historical reporting that characterized Management Information Systems (MIS).	**True**
The use of dashboards and data visualizations is seldom effective in finding efficiencies in organizations, as demonstrated by the Seattle Children's Hospital Case Study.	**False**
The use of statistics in baseball by the Oakland Athletics, as described in the Moneyball case study, is an example of the effectiveness of prescriptive analytics.	**True**
Pushing programming out to distributed data is achieved solely by using the Hadoop Distributed File System or HDFS.	**False**
Volume, velocity, and variety of data characterize the Big Data paradigm. (Chapter 13)	**True**

MULTIPLE CHOICE

1. In the Magpie Sensing case study, the automated collection of temperature and humidity data on shipped goods helped with various types of analytics. Which of the following is an example of prescriptive analytics?
   - A) location of the shipment
   - B) real time reports of the shipment's temperature
   - C) warning of an open shipment seal
   - **D) optimal temperature setting**

2. In the Magpie Sensing case study, the automated collection of temperature and humidity data on shipped goods helped with various types of analytics. Which of the following is an example of predictive analytics?
   - **A) warning of an open shipment seal**
   - B) optimal temperature setting
   - C) real time reports of the shipment's temperature
   - D) location of the shipment
3. Which of the following is NOT an example that falls within the four major categories of business environment factors for today's organizations?
   - A) globalization
   - **B) fewer government regulations**
   - C) increased pool of customers
   - D) increased competition
4. Organizations counter the pressures they experience in their business environments in multiple ways. Which of the following is NOT an effective way to counter these pressures?
   - A) reactive actions
   - B) anticipative actions
   - **C) retroactive actions**
   - D) adaptive actions
5. Which of the following activities permeates nearly all managerial activity?
   - A) planning
   - **B) decision-making**
   - C) directing
   - D) controlling
6. Why are analytical decision making skills now viewed as more important than interpersonal skills for an organization's managers?
   - A) because personable and friendly managers are always the least effective
   - B) because interpersonal skills are never important in organizations
   - C) because analytical-oriented managers tend to be flashier and less methodical
   - **D) because analytical-oriented managers produce better results over time**
7. Business environments and government requirements are becoming more complex. All of the following actions to manage this complexity would be appropriate EXCEPT
   - A) deploying more sophisticated tools and technique.
   - B) hiring more sophisticated and computer-savvy managers.
   - **C) seeking new ways to avoid government compliance.**
   - D) avoiding expensive trial and error to find out what works.
8. The deployment of large data warehouses with terabytes or even petabytes of data been crucial to the growth of decision support. All the following explain why EXCEPT
   - A) data warehouses have enabled the affordable collection of data for analytics.
   - B) data warehouses have assisted the collection of data for data mining.
   - **C) data warehouses have enabled the collection of decision makers in one place.**
   - D) data warehouses have assisted the collection of data from multiple sources.
9. Which of the following statements about cognitive limits of organizational decision makers is true?
   - **A) Cognitive limits affect both the recall and use of data by decision makers.**
   - B) Only top managers make decisions where cognitive limits are strained.
   - C) The most talented and effective managers do not have cognitive limitations.
   - D) All organizational decision-making requires data beyond human cognitive limits.
10. For the majority of organizations, evaluating the credit rating of a potential business partner is a(n)
    - A) structured decision.
    - B) unstructured decision.
    - **C) managerial control decision.**

D) strategic decision.

11. For the majority of organizations, a daily accounts receivable transaction is a(n)

A) strategic decision.
B) managerial control decision.
C) structured decision.
D) unstructured decision.

12. All of the following may be viewed as decision support systems EXCEPT

A) an expert system to diagnose a medical condition.
B) a system that helps to manage the organization's supply chain management.
C) a knowledge management system to guide decision makers.
D) a retail sales system that processes customer sales transactions.

13. Business intelligence (BI) can be characterized as a transformation of

A) data to information to decisions to actions.
B) Big Data to data to information to decisions.
C) data to processing to information to actions.
D) actions to decisions to feedback to information.

14. In answering the question "Which customers are most likely to click on my online ads and purchase my goods?", you are most likely to use which of the following analytic applications?

A) customer attrition
B) channel optimization
C) customer profitability
D) propensity to buy

15. In answering the question "Which customers are likely to be using fake credit cards?", you are most likely to use which of the following analytic applications?

A) customer segmentation
B) channel optimization
C) fraud detection
D) customer profitability

16. When Sabre developed their Enterprise Data Warehouse, they chose to use near-real-time updating of their database. The main reason they did so was

A) to be able to assess internal operations.
B) to aggregate performance metrics in an understandable way.
C) to provide a 360 degree view of the organization.
D) to provide up-to-date executive insights.

17. How are descriptive analytics methods different from the other two types?

A) They answer "what to do?" queries, not "what-if?" queries.
B) They answer "what-is?" queries, not "what will be?" queries.
C) They answer "what-if?" queries, not "how many?" queries.
D) They answer "what will be?" queries, not "what to do?" queries.

18. Prescriptive BI capabilities are viewed as more powerful than predictive ones for all the following reasons EXCEPT

A) prescriptive models generally build on (with some overlap) predictive ones.
B) only prescriptive BI capabilities have monetary value to top-level managers.
C) understanding the likelihood of certain events often leaves unclear remedies.
D) prescriptive BI gives actual guidance as to actions.

19. Which of the following statements about Big Data is true?

A) MapReduce is a storage filing system.
B) Data chunks are stored in different locations on one computer.
C) Hadoop is a type of processor used to process Big Data applications.
D) Pure Big Data systems do not involve fault tolerance.

20. Big Data often involves a form of distributed storage and processing using Hadoop and MapReduce. One reason for this is
    A) the processing power needed for the centralized model would overload a single computer.
    B) Big Data systems have to match the geographical spread of social media.
    C) centralized storage creates too many vulnerabilities.
    D) the "Big" in Big Data necessitates over 10,000 processing nodes.

FILL-IN-THE-BLANK

- The desire by a customer to customize a product falls under the **consumer demand** category of business environment factors.
- An older and more diverse workforce falls under the **societal** category of business environment factors.
- Organizations using BI systems are typically seeking to **close** the gap between the organization's current and desired performance.
- Mintzberg defines the **entrepreneur** as a managerial role that involves searching the environment for new opportunities.
- Group communication and **collaboration** involves decision makers who are likely to be in different locations.
- **Wireless** technology enables managers to access and analyze information anytime and from anyplace.
- A(n) **semistructured** problem such as setting budgets for products is one that has some structured elements and some unstructured elements also.
- A(n) **unstructured** problem such as new technology development is one that has very few structured elements.
- **Business intelligence (BI)** is an umbrella term that combines architectures, tools, databases, analytical tools, applications, and methodologies.
- A(n) **data warehouse** is a major component of a Business Intelligence (BI) system that holds source data.
- A(n) **user interface** is a major component of a Business Intelligence (BI) system that is usually browser based and often presents a portal or dashboard.
- **Business** cycle times are now extremely compressed, faster, and more informed across industries.
- The fraud **detection** analytic application helps determine fraudulent events and take action.
- Sabre used executive **dashboards** to present performance metrics in a concise way to its executives.
- **Descriptive** analytics help managers understand current events in the organization including causes, trends, and patterns.
- **Predictive** analytics help managers understand probable future outcomes.
- **Prescriptive** analytics help managers make decisions to achieve the best performance in the future.
- The Google search engine is an example of Big Data in that it has to search and index billions of **web pages** in fractions of a second for each search.
- The filing system developed by Google to handle Big Data storage challenges is known as the **Hadoop** Distributed File System.
- The programming algorithm developed by Google to handle Big Data computational challenges is known as **MapReduce**.

SHORT ANSWER

The environment in which organizations operate today is becoming more and more complex. Business environment factors can be divided into four major categories. What are these categories?

- **Markets.**
- **Consumer demands.**
- **Technology.**
- **Societal.**

List four of Mintzberg's Decisional roles of managers.

- **Entrepreneur:**
  - **Searches the organization and its environment for opportunities and initiates improvement projects to bring about change; supervises design of certain projects**
- **Disturbance handler:**
  - **Is responsible for corrective action when the organization faces important, unexpected disturbances**
- **Resource allocator:**
  - **Is responsible for the allocation of organizational resources of all kinds; in effect, is responsible for the making or approval of all significant organizational decisions**
- **Negotiator:**
  - **Is responsible for representing the organization at major negotiations**

Managers usually make decisions by following a four-step process. What are the steps?

- **Define the problem (i.e., a decision situation that may deal with some difficulty or with an opportunity).**
- **Construct a model that describes the real-world problem.**
- **Identify possible solutions to the modeled problem and evaluate the solutions.**
- **Compare, choose, and recommend a potential solution to the problem.**

List three developments that have contributed to facilitating growth of decision support and analytics.

- **Group communication and collaboration**
- **Improved data management**
- **Managing giant data warehouses and Big Data**
- **Analytical support**
- **Overcoming cognitive limits in processing and storing information**
- **Knowledge management**
- **Anywhere, anytime support**

Describe the types of computer support that can be used for structured, semistructured, and unstructured decisions.

- **Structured Decisions:**
  - **Structured problems, which are encountered repeatedly, have a high level of structure.**
  - **It is therefore possible to abstract, analyze, and classify them into specific categories and use a scientific approach for automating portions of this type of managerial decision making.**
- **Semistructured Decisions:**
  - **Semistructured problems may involve a combination of standard solution procedures and human judgment.**
  - **Management science can provide models for the portion of a decision-making problem that is structured.**

    - **For the unstructured portion, a DSS can improve the quality of the information on which the decision is based by providing, for example, not only a single solution but also a range of alternative solutions, along with their potential impacts.**
- **Unstructured Decisions:**
    - **These can be only partially supported by standard computerized quantitative methods.**
    - **It is usually necessary to develop customized solutions.**
    - **However, such solutions may benefit from data and information generated from corporate or external data sources.**

What are the four major components of a Business Intelligence (BI) system?

- **A data warehouse, with its source data;**
- **Business analytics, a collection of tools for manipulating, mining, and analyzing the data in the data warehouse;**
- **Business performance management (BPM) for monitoring and analyzing performance;**
- **user interface (e.g., a dashboard);**

List and describe three levels or categories of analytics that are most often viewed as sequential and independent, but also occasionally seen as overlapping.

- **Descriptive or reporting analytics refers to knowing what is happening in the organization and understanding some underlying trends and causes of such occurrences.**
- **Predictive analytics aims to determine what is likely to happen in the future.**
    - **This analysis is based on statistical techniques as well as other more recently developed techniques that fall under the general category of data mining.**
- **Prescriptive analytics recognizes what is going on as well as the likely forecast and make decisions to achieve the best performance possible.**

How does Amazon.com use predictive analytics to respond to product searches by the customer?

- **Amazon uses clustering algorithms to segment customers into different clusters to be able to target specific promotions to them.**
- **The company also uses association mining techniques to estimate relationships between different purchasing behaviors.**
- **That is, if a customer buys one product, what else is the customer likely to purchase? That helps Amazon recommend or promote related products.**
- **For example, any product search on Amazon.com results in the retailer also suggesting other similar products that may interest a customer.**

Describe and define Big Data. Why is a search engine a Big Data application?

- **Big Data is data that cannot be stored in a single storage unit.**
    - **Big Data typically refers to data that is arriving in many different forms, be they structured, unstructured, or in a stream.**
    - **Major sources of such data are clickstreams from Web sites, postings on social media sites such as Facebook, or data from traffic, sensors, or weather.**
- **A Web search engine such as Google needs to search and index billions of Web pages in order to give you relevant search results in a fraction of a second.**
    - **Although this is not done in real time, generating an index of all the Web pages on the Internet is not an easy task.**

What storage system and processing algorithm were developed by Google for Big Data?

- **Google developed and released as an Apache project the Hadoop Distributed File System (HDFS) for storing large amounts of data in a distributed way.**
- **Google developed and released as an Apache project the MapReduce algorithm for pushing computation to the data, instead of pushing data to a computing node.**

CHAPTER 2

TRUE OR FALSE

When HP approaches problem-solving, the first step in solving business problems is building a model that enables decision makers to develop a good understanding of the problem.	**False**
In a decision making environment, continuous change always validates the assumptions of the decision makers.	**False**
The most important feature of management support systems is the computational efficiency involved in making a decision.	**False**
Web-based decision support systems can provide support to both individuals and groups that act in a decision-making capacity.	**True**
Single decision makers rarely face decisions with multiple objectives in organizations and so are not the focus of data analytics tools.	**False**
The design phase of decision making is where the decision maker examines reality and identifies and defines the problem.	**False**
Only after the failed implementation of a decision can the decision maker return a prior stage of decision making.	**False**
Web-based collaboration tools (e.g., GSS) can assist in multiple stages of decision making, not just the intelligence phase.	**True**
Uncovering the existence of a problem can be achieved through monitoring and analyzing of the organization's productivity level. The derived measurements of productivity are based on real data.	**True**
Qualitative elements of a problem cannot be incorporated into formal decision models, so one can only seek to minimize their impact.	**False**
Since the business environment involves considerable uncertainty, a manager cannot use modeling to estimate the risks resulting from specific actions.	**False**
A normative model examines all the possible alternatives in order to prove that the one selected is the best.	**True**
Since a descriptive model checks the performance of the system for only a subset of all possible alternatives, there is no guarantee that a selected alternative will be optimal.	**True**
Generating alternatives manually is often necessary in the model-building process. The best option for the decision makers is to generate as many of these alternatives as is conceivable.	**False**
Generally speaking, people intuitively estimate risk quite accurately.	**False**
A data warehouse can support the intelligence phase of decision making by continuously monitoring both internal and external information, looking for early signs of problems and opportunities through a Web-based enterprise information portal or dashboard.	**True**
Business intelligence systems typically support solving a certain problem or evaluate an opportunity, while decision support systems monitor situations and identify problems and/or opportunities, using analytic methods.	**False**
Artificial intelligence-based DSS fall into this category of document-driven DSS.	**False**
The DSS component that includes the financial, statistical, management science, or other quantitative models is called the model management subsystem.	**True**
Knowledge-based management subsystems provide intelligence to augment the decision maker's own intelligence.	**True**

MULTIPLE CHOICE

1. The HP Case illustrates that after analytics are chosen to solve a problem, building a new decision model from scratch or purchasing one may not always be the best approach. Why is that?
   A) Analytic models work better when they are built from scratch or purchased.
   B) A related tool requiring slight modification may already exist.
   C) CIOs are more likely to allocate funds to new development.
   D) Decision models should never be purchased, only developed in house.
2. Groupthink in a decision-making environment occurs when
   A) group members accept the same timeframe for problem solving without complaining.
   B) group members all accept a course of action without thinking for themselves.
   C) group members all use the same analytic tools without having a choice.
   D) group members are all working together for the firm's success.
3. All of the following statements about decision style are true EXCEPT
   A) heuristic styles can also be democratic.
   B) autocratic styles are authority-based.
   C) decision styles may vary among lower-level managers.

D) decision styles are consistent among top managers.

4. A search for alternatives occurs in which phase of the decision making/action model?
   A) the intelligence phase
   B) the implementation phase
   C) the choice phase
   D) the design phase
5. All of the following are benefits of using models for decision support EXCEPT
   A) you can find out probable outcomes of an action before actually taking it.
   B) using well-designed models always guarantees you success in implementation.
   C) it is easier to manipulate a model than a real system.
   D) the cost of a model is usually much lower than manipulating the system in implementation.
6. In the design phase of decision making, selecting a principle of choice or criteria means that
   A) optimality is not the only criterion for acceptable solutions.
   B) if an objective model is used with hard data, all decision makers will make the same choice.
   C) risk acceptability is a subjective concept and plays little part in modeling.
   D) using well-designed models guarantees you success in real life.
7. What form of decision theory assumes that decision makers are rational beings who always seek to strictly maximize economic goals?
   A) satisficing decision theory
   B) human optimal decision theory
   C) normative decision theory
   D) the theory of bounded rationality
8. When an Accounts Payable department improves their information system resulting in faster payments to vendors, without the Accounts Receivable Department doing the same, leading to a cash flow crunch, what can we say happened in decision-theoretic terms?
   A) cash flow problems
   B) profit minimization
   C) suboptimization
   D) optimization
9. All of the following statements about risk in decision making are correct EXCEPT
   A) all business decisions incorporate an element of risk.
   B) decision makers frequently measure risk and uncertainty incorrectly.
   C) most decision makers are pessimistic about decision outcomes.
   D) methodologies are available for handling extreme uncertainty.
10. The Web can play a significant role in making large amounts of information available to decision makers. Decision makers must be careful that this glut of information does not
    A) detract from the quality and speed of decision making.
    B) take on the same credibility of internally-generated data.
    C) increase their enthusiasm for data available on the web.
    D) take on the same role as human intuition.
11. All of the following statements about the decision implementation phases are true EXCEPT
    A) ES and KMS can help in training and support for decision implementation.
    B) implementation is every bit as important as the decision itself.
    C) ERP, CRP, and BPM tools can all help track decision implementation.
    D) employees need only the decisions from the CEO, not the rationale.
12. For DSS, why are semistructured or unstructured decisions the main focus of support?
    A) MIS staff prefer to work on solving unstructured and semistructured decisions.
    B) There are many more unstructured and semistructured decisions than structured in organizations.
    C) Unstructured and semistructured decisions are the easiest to solve.

D) They include human judgment, which is incorporated into DSS.

13. What class of DSS incorporates simulation and optimization?
    - **A) model-driven DSS**
    - B) communications-driven/Group DSS
    - C) knowledge-driven DSS
    - D) data-driven DSS
14. When a DSS is built, used successfully and integrated into the company's business processes, it was most likely built for a(n)
    - A) one-off decision.
    - B) unimportant decision.
    - C) ambiguous decision.
    - **D) recurrent decision.**
15. The fact that many organizations share many similar problems means that in sourcing a DSS, it is often wiser to acquire a(n)
    - A) offshored DSS.
    - B) consultant-developed DSS.
    - C) custom-made DSS.
    - **D) ready-made DSS.**
16. The software that manages the DSS database and enables relevant data to be accessed by DSS application programs is called
    - A) KWS.
    - B) CRM.
    - C) ERP.
    - **D) DBMS.**
17. The model management subsystem provides the system's analytical capabilities and appropriate software management. Which of the following is NOT an element of the model management subsystem?
    - A) MBMS
    - B) model base
    - C) model execution, integration, and command processor
    - **D) DBMS**
18. While Microsoft Excel can be an efficient tool for developing a DSS, compared to using a programming language like C++, a shortcoming of Excel is
    - **A) errors can creep into formulas somewhat easily.**
    - B) Excel is not widely understood compared to a language like C++.
    - C) it cannot be used effectively for small or medium sized problems.
    - D) it is not widely available for purchase.
19. What type of user interface has been recognized as an effective DSS GUI because it is familiar, user friendly, and a gateway to almost all sources of necessary information and data?
    - A) visual basic interfaces
    - **B) Web browsers**
    - C) ASP.net
    - D) mainframe interfaces
20. The user communicates with and commands the DSS through the user interface subsystem. Researchers assert that some of the unique contributions of DSS are derived from
    - A) the user being considered part of the system.
    - **B) the intensive interaction between the computer and the decision maker.**
    - C) some DSS user interfaces utilizing natural-language input (i.e., text in a human language).
    - D) the Web browser.

FILL-IN-THE-BLANK

- At two opposite ends of the spectrum are autocratic and **democratic** decision styles.
- **Intelligence** in decision making involves scanning the environment, either intermittently or continuously.

- The elevators case study shows that correct problem **identification** is important in decision-making.
- **Problem classification** is the conceptualization of a problem in an attempt to place it in a definable category, possibly leading to a standard solution approach.
- In creating a normative model, a decision maker examines all the alternatives to prove that the one selected is indeed the best, and is what the person would normally want. This process is basically known as **optimization**.
- A(n) **descriptive model** is a typically mathematically based model that describes things as they are or as they are believed to be.
- A(n) **cognitive** map can help a decision maker sketch out the important qualitative factors and their causal relationships in a messy decision-making situation.
- The best decision makers accurately estimate the **risk** associated with decision alternatives to aid their selection.
- The **implementation** phase involves putting a recommended solution to work, not necessarily implementing a computer system.
- Early definitions of a(n) **decision support system (DSS)** identified it as a system intended to support managerial decision makers in semistructured and unstructured decision situations.
- DSS applications have been classified in several different ways. **Document**-driven DSS rely on knowledge coding, analysis, search, and retrieval for decision support.
- DSS developed around optimization or simulation models and incorporate model formulation, maintenance, and management in distributed computing environments, are known as **model**-driven DSS.
- A DSS application can employ a data management subsystem, a model management subsystem, a user interface subsystem, and a(n) **knowledge-based management subsystem**.
- The model management subsystem includes financial, statistical, management science, and other quantitative models that provide the system's analytical capabilities plus appropriate software management. This software is often called a(n) **model base management system (MBMS)**.
- In the Station Casinos case, the decision support system brought about benefits from being able to capture, analyze and segment **customers**.
- Because DSS deal with semistructured or unstructured problems, it is often necessary to customize models, using programming tools and languages. For small and medium-sized DSS or for less complex ones, **spreadsheet** software is usually used.
- The user communicates with and commands the DSS through the **user interface** subsystem.
- The knowledge-based management subsystem can be interconnected with the organization's knowledge repository (part of a knowledge management system [KMS]), which is sometimes called the **organizational knowledge base**.
- The Watson Question Answering computing platform uses machine **learning** to acquire vast amounts of new medical knowledge.
- Geographical Information Systems (GIS) can be readily integrated with other, more traditional **decision support system (DSS)** components and tools for improved decision making.

SHORT ANSWER

Olavson and Fry (2008) have worked on many spreadsheet models for assisting decision makers at HP and have identified several lessons from both their successes and their failures when it comes to constructing and applying spreadsheet-based tools. How do they define a tool?

- **They define a tool as "a reusable, analytical solution designed to be handed off to nontechnical end users to assist them in solving a repeated business problem."**

According to Simon (1977), managerial decision making is synonymous with the entire management process. Give a working definition of decision making.

- **Decision making is a process of choosing among two or more alternative courses of action for the purpose of attaining one or more goals.**

Computer support can be provided at a broad level, enabling members of whole departments, divisions, or even entire organizations to collaborate online. Name some of the various systems that have evolved from computer support.

- **Computer support has evolved over the past few years into enterprise information systems (EIS) and includes group support systems (GSS), enterprise resource management (ERM)/enterprise resource planning (ERP), supply chain management (SCM), knowledge management systems (KMS), and customer relationship management (CRM) systems.**

Name Simon's four phases of decision making and mention how they are impacted by the web.

- **Intelligence**
    - **Access to information to identify problems and opportunities from internal and external data sources**
    - **Access to analytics methods to identify opportunities**
    - **Collaboration through group support systems (GSS) and knowledge management systems (KMS)**
- **Design**
    - **Access to data, models, and solution methods**
    - **Use of online analytical processing (OLAP), data mining, and data warehouses**
    - **Collaboration through GSS and KMS**
    - **Similar solutions available from KMS**
- **Choice**
    - **Access to methods to evaluate the impacts of proposed solutions**
- **Implementation**
    - **Web-based collaboration tools (e.g., GSS) and KMS, which can assist in implementing decisions**
    - **Tools, which monitor the performance of e-commerce and other sites, including intranets, extranets, and the Internet**

A major characteristic of a DSS and many BI tools (notably those of business analytics) is the inclusion of at least one model. How does the text describe a model?

- **A model is a simplified representation or abstraction of reality.**
    - **It is usually simplified because reality is too complex to describe exactly and because much of the complexity is actually irrelevant in solving a specific problem.**

According to Simon (1977), most human decision making, whether organizational or individual, involves a willingness to settle for a satisfactory solution, "something less than the best." This is called satisficing. How does a decision maker go about satisficing?

- **When satisficing, the decision maker sets up an aspiration, a goal, or a desired level of performance and then searches the alternatives until one is found that achieves this level.**

A scenario is a statement of assumptions about the operating environment of a particular system at a given time; that is, it is a narrative description of the decision-situation setting. What does a scenario describe, and what may it also provide?

- **A scenario describes the decision and uncontrollable variables and parameters for a specific modeling situation. It may also provide the procedures and constraints for the modeling.**

Relate four specific technologies that support all phases of the decision making process, and describe what they provide.

- **Databases, data marts, and especially data warehouses are important technologies in supporting all phases of decision making.**
  - **They provide the data that drive decision making.**

A DSS is typically built to support the solution of a certain problem or to evaluate an opportunity. Describe three key characteristics and capabilities of DSS.

- **Support for decision makers, mainly in semistructured and unstructured situations, by bringing together human judgment and computerized information.**
  - **Such problems cannot be solved (or cannot be solved conveniently) by other computerized systems or through use of standard quantitative methods or tools.**
  - **Generally, these problems gain structure as the DSS is developed. Even some structured problems have been solved by DSS.**
- **Support for all managerial levels, ranging from top executives to line managers.**
- **Support for individuals as well as groups.**
  - **Less-structured problems often require the involvement of individuals from different departments and organizational levels or even from different organizations.**
  - **DSS support virtual teams through collaborative Web tools.**
  - **DSS have been developed to support individual and group work, as well as to support individual decision making and groups of decision makers working somewhat independently.**
- **Support for interdependent and/or sequential decisions.**
  - **The decisions may be made once, several times, or repeatedly.**
- **Support in all phases of the decision-making process: intelligence, design, choice, and implementation.**
- **Support for a variety of decision-making processes and styles.**
- **The decision maker should be reactive, able to confront changing conditions quickly, and able to adapt the DSS to meet these changes.**
  - **DSS are flexible, so users can add, delete, combine, change, or rearrange basic elements.**
  - **They are also flexible in that they can be readily modified to solve other, similar problems.**

Name and give a brief description of each of the components of a DSS application.

- **A data management subsystem, which includes a database that contains relevant data for the situation**
- **A model management subsystem, which is the component that includes financial, statistical, management science, or other quantitative models that provide the system's analytical capabilities and appropriate software management**
- **A user interface subsystem, through which the user communicates with and commands the DSS, and which a user is considered a part of**
- **A knowledge-based management subsystem, which provides intelligence to augment the decision maker's own.**

CHAPTER 3

TRUE OR FALSE

In the Isle of Capri case, the only capability added by the new software was increased processing speed of processing reports.	**False**
The "islands of data" problem in the 1980s describes the phenomenon of unconnected data being stored in numerous locations within an organization.	**True**
Subject oriented databases for data warehousing are organized by detailed subjects such as disk drives, computers, and networks.	**False**
Data warehouses are subsets of data marts.	**False**
One way an operational data store differs from a data warehouse is the recency of their data.	**True**
Organizations seldom devote a lot of effort to creating metadata because it is not important for the effective use of data warehouses.	**False**
Without middleware, different BI programs cannot easily connect to the data warehouse.	**True**
Two-tier data warehouse/BI infrastructures offer organizations more flexibility but cost more than three-tier ones.	**False**
Moving the data into a data warehouse is usually the easiest part of its creation.	**False**
The hub-and-spoke data warehouse model uses a centralized warehouse feeding dependent data marts.	**True**
Because of performance and data quality issues, most experts agree that the federated architecture should supplement data warehouses, not replace them.	**True**
Bill Inmon advocates the data mart bus architecture whereas Ralph Kimball promotes the hub-and-spoke architecture, a data mart bus architecture with conformed dimensions.	**False**
The ETL process in data warehousing usually takes up a small portion of the time in a data-centric project.	**False**
In the Starwood Hotels case, up-to-date data and faster reporting helped hotel managers better manage their occupancy rates.	**True**
Large companies, especially those with revenue upwards of $500 million consistently reap substantial cost savings through the use of hosted data warehouses.	**False**
OLTP systems are designed to handle ad hoc analysis and complex queries that deal with many data items.	**False**
The data warehousing maturity model consists of six stages: prenatal, infant, child, teenager, adult, and sage.	**True**
A well-designed data warehouse means that user requirements do not have to change as business needs change.	**False**
Data warehouse administrators (DWAs) do not need strong business insight since they only handle the technical aspect of the infrastructure.	**False**
Because the recession has raised interest in low-cost open source software, it is now set to replace traditional enterprise software.	**False**

MULTIPLE CHOICE

1. The "single version of the truth" embodied in a data warehouse such as Capri Casinos' means all of the following EXCEPT
   - A) decision makers get to see the same results to queries.
   - B) decision makers have the same data available to support their decisions.
   - **C) decision makers have unfettered access to all data in the warehouse.**
   - D) decision makers get to use more dependable data for their decisions.
2. Operational or transaction databases are product oriented, handling transactions that update the database. In contrast, data warehouses are
   - **A) subject-oriented and nonvolatile.**
   - B) subject-oriented and volatile.
   - C) product-oriented and nonvolatile.
   - D) product-oriented and volatile.
3. Which kind of data warehouse is created separately from the enterprise data warehouse by a department and not reliant on it for updates?
   - A) sectional data mart
   - B) volatile data mart
   - C) public data mart
   - **D) independent data mart**

4. All of the following statements about metadata are true EXCEPT
   - A) metadata gives context to reported data.
   - **B) for most organizations, data warehouse metadata are an unnecessary expense.**
   - C) metadata helps to describe the meaning and structure of data.
   - D) there may be ethical issues involved in the creation of metadata.
5. A Web client that connects to a Web server, which is in turn connected to a BI application server, is reflective of a
   - **A) three tier architecture.**
   - B) one tier architecture.
   - C) two tier architecture.
   - D) four tier architecture.
6. Which of the following BEST enables a data warehouse to handle complex queries and scale up to handle many more requests?
   - A) Microsoft Windows
   - B) a larger IT staff
   - C) use of the web by users as a front-end
   - **D) parallel processing**
7. Which data warehouse architecture uses metadata from existing data warehouses to create a hybrid logical data warehouse comprised of data from the other warehouses?
   - A) centralized data warehouse architecture
   - B) hub-and-spoke data warehouse architecture
   - C) independent data marts architecture
   - **D) federated architecture**
8. Which data warehouse architecture uses a normalized relational warehouse that feeds multiple data marts?
   - A) federated architecture
   - **B) hub-and-spoke data warehouse architecture**
   - C) independent data marts architecture
   - D) centralized data warehouse architecture
9. Which approach to data warehouse integration focuses more on sharing process functionality than data across systems?
   - A) enterprise function integration
   - B) enterprise information integration
   - C) extraction, transformation, and load
   - **D) enterprise application integration**
10. In which stage of extraction, transformation, and load (ETL) into a data warehouse are data aggregated?
    - A) extraction
    - B) load
    - **C) transformation**
    - D) cleanse
11. In which stage of extraction, transformation, and load (ETL) into a data warehouse are anomalies detected and corrected?
    - A) load
    - B) transformation
    - **C) cleanse**
    - D) extraction
12. Data warehouses provide direct and indirect benefits to using organizations. Which of the following is an indirect benefit of data warehouses?
    - A) extensive new analyses performed by users
    - B) simplified access to data
    - C) better and more timely information
    - **D) improved customer service**
13. All of the following are benefits of hosted data warehouses EXCEPT
    - **A) greater control of data.**
    - B) frees up in-house systems.

C) better quality hardware.
D) smaller upfront investment.

14. When representing data in a data warehouse, using several dimension tables that are each connected only to a fact table means you are using which warehouse structure?
    A) relational schema
    B) dimensional schema
    C) star schema
    D) snowflake schema
15. When querying a dimensional database, a user went from summarized data to its underlying details. The function that served this purpose is
    A) slice.
    B) roll-up.
    C) drill down.
    D) dice.
16. Which of the following online analytical processing (OLAP) technologies does NOT require the precomputation and storage of information?
    A) MOLAP
    B) SQL
    C) HOLAP
    D) ROLAP
17. Active data warehousing can be used to support the highest level of decision making sophistication and power. The major feature that enables this in relation to handling the data is
    A) nature of the data.
    B) speed of data transfer.
    C) country of (data) origin.
    D) source of the data.
18. Which of the following statements is more descriptive of active data warehouses in contrast with traditional data warehouses?
    A) large numbers of users, including operational staffs
    B) restrictive reporting with daily and weekly data currency
    C) detailed data available for strategic use only
    D) strategic decisions whose impacts are hard to measure
19. How does the use of cloud computing affect the scalability of a data warehouse?
    A) Hardware resources are dynamically allocated as use increases.
    B) Cloud vendors are mostly based overseas where the cost of labor is low.
    C) Cloud computing has little effect on a data warehouse's scalability.
    D) Cloud computing vendors bring as much hardware as needed to users' offices.
20. All of the following are true about in-database processing technology EXCEPT
    A) it pushes the algorithms to where the data is.
    B) it is the same as in-memory storage technology.
    C) it is often used for apps like credit card fraud detection and investment risk management.
    D) it makes the response to queries much faster than conventional databases.

FILL-IN-THE-BLANK

- With **real-time** data flows, managers can view the current state of their businesses and quickly identify problems.
- In **product** oriented data warehousing, operational databases are tuned to handle transactions that update the database.
- The three main types of data warehouses are data marts, operational **data stores**, and enterprise data warehouses.
- **Metadata** describe the structure and meaning of the data, contributing to their effective use.
- Most data warehouses are built using **relational** database management systems to control and manage the data.

- A(n) **hub-and-spoke** architecture is used to build a scalable and maintainable infrastructure that includes a centralized data warehouse and several dependent data marts.
- The **federated** data warehouse architecture involves integrating disparate systems and analytical resources from multiple sources to meet changing needs or business conditions.
- Data **integration** comprises data access, data federation, and change capture.
- **Enterprise application integration (EAI)** is a mechanism that integrates application functionality and shares functionality (rather than data) across systems, thereby enabling flexibility and reuse.
- **Enterprise information integration (EII)** is a mechanism for pulling data from source systems to satisfy a request for information. It is an evolving tool space that promises real-time data integration from a variety of sources, such as relational databases, Web services, and multidimensional databases.
- Performing extensive **extraction, transformation, and load (ETL)** to move data to the data warehouse may be a sign of poorly managed data and a fundamental lack of a coherent data management strategy.
- The **Inmon** Model, also known as the EDW approach, emphasizes top-down development, employing established database development methodologies and tools, such as entity-relationship diagrams (ERD), and an adjustment of the spiral development approach.
- The **Kimball** Model, also known as the data mart approach, is a "plan big, build small" approach. A data mart is a subject-oriented or department-oriented data warehouse. It is a scaled-down version of a data warehouse that focuses on the requests of a specific department, such as marketing or sales.
- **Dimensional** modeling is a retrieval-based system that supports high-volume query access.
- Online **analytical processing** is arguably the most commonly used data analysis technique in data warehouses.
- Online **transaction processing** is a term used for a transaction system that is primarily responsible for capturing and storing data related to day-to-day business functions such as ERP, CRM, SCM, and point of sale.
- In the Michigan State Agencies case, the approach used was a(n) **enterprise** one, instead of developing separate BI/DW platforms for each business area or state agency.
- The role responsible for successful administration and management of a data warehouse is the **data warehouse administrator (DWA)**, who should be familiar with high-performance software, hardware, and networking technologies, and also possesses solid business insight.
- **SaaS (software as a service)**, or "The Extended ASP Model," is a creative way of deploying information system applications where the provider licenses its applications to customers for use as a service on demand (usually over the Internet)
- **In-database processing** (also called in-database analytics) refers to the integration of the algorithmic extent of data analytics into data warehouse.

SHORT ANSWER

What is the definition of a data warehouse (DW) in simple terms?

- **In simple terms, a data warehouse (DW) is a pool of data produced to support decision making; it is also a repository of current and historical data of potential interest to managers throughout the organization.**

A common way of introducing data warehousing is to refer to its fundamental characteristics. Describe three characteristics of data warehousing.

- **Subject oriented.**
  - **Data are organized by detailed subject, such as sales, products, or customers, containing only information relevant for decision support.**
- **Integrated.**
  - **Integration is closely related to subject orientation.**
  - **Data warehouses must place data from different sources into a consistent format.**
  - **To do so, they must deal with naming conflicts and discrepancies among units of measure.**
  - **A data warehouse is presumed to be totally integrated.**
- **Time variant (time series).**
  - **A warehouse maintains historical data.**
  - **The data do not necessarily provide current status (except in real-time systems).**
  - **They detect trends, deviations, and long-term relationships for forecasting and comparisons, leading to decision making.**
  - **Every data warehouse has a temporal quality.**
  - **Time is the one important dimension that all data warehouses must support.**
  - **Data for analysis from multiple sources contains multiple time points (e.g., daily, weekly, monthly views).**
- **Nonvolatile.**
  - **After data are entered into a data warehouse, users cannot change or update the data.**
  - **Obsolete data are discarded, and changes are recorded as new data.**
- **Web based.**
  - **Data warehouses are typically designed to provide an efficient computing environment for Web-based applications.**
- **Relational/multidimensional.**
  - **A data warehouse uses either a relational structure or a multidimensional structure.**
  - **A recent survey on multidimensional structures can be found in Romero and Abell? (2009).**
- **Client/server.**
  - **A data warehouse uses the client/server architecture to provide easy access for end users.**
- **Real time.**
  - **Newer data warehouses provide real-time, or active, data-access and analysis capabilities (see Basu, 2003; and Bonde and Kuckuk, 2004).**
- **Include metadata.**
  - **A data warehouse contains metadata (data about data) about how the data are organized and how to effectively use them.**

What is the definition of a data mart?

- **A data mart is a subset of a data warehouse, typically consisting of a single subject area (e.g., marketing, operations).**
  - **Whereas a data warehouse combines databases across an entire enterprise, a data mart is usually smaller and focuses on a particular subject or department.**

Mehra (2005) indicated that few organizations really understand metadata, and fewer understand how to design and implement a metadata strategy. How would you describe metadata?

- **Metadata are data about data.**
    - **Metadata describe the structure of and some meaning about data, thereby contributing to their effective or ineffective use.**

According to Kassam (2002), business metadata comprise information that increases our understanding of traditional (i.e., structured) data. What is the primary purpose of metadata?

- **The primary purpose of metadata should be to provide context to the reported data; that is, it provides enriching information that leads to the creation of knowledge.**

In the MultiCare case, how was data warehousing able to reduce septicemia mortality rates in MultiCare hospitals?

- **The Adaptive Data WarehouseTM organized and simplified data from multiple data sources across the continuum of care.**
    - **It became the single source of truth required to see care improvement opportunities and to measure change, integrated teams consisting of clinicians, technologists, analysts, and quality personnel were essential for accelerating MultiCare's efforts to reduce septicemia mortality.**
- **Together the collaborative effort addressed three key bodies of work–standard of care definition, early identification, and efficient delivery of defined-care standard.**

Briefly describe four major components of the data warehousing process.

- **Data sources.**
    - **Data are sourced from multiple independent operational "legacy" systems and possibly from external data providers (such as the U.S. Census).**
    - **Data may also come from an OLTP or ERP system.**
- **Data extraction and transformation.**
    - **Data are extracted and properly transformed using custom-written or commercial ETL software.**
- **Data loading.**
    - **Data are loaded into a staging area, where they are transformed and cleansed.**
    - **The data are then ready to load into the data warehouse and/or data marts.**
- **Comprehensive database.**
    - **Essentially, this is the EDW to support all decision analysis by providing relevant summarized and detailed information originating from many different sources.**
- **Metadata.**
    - **Metadata include software programs about data and rules for organizing data summaries that are easy to index and search, especially with Web tools.**
- **Middleware tools.**
    - **Middleware tools enable access to the data warehouse.**
    - **There are many front-end applications that business users can use to interact with data stored in the data repositories, including data mining, OLAP, reporting tools, and data visualization tools.**

There are several basic information system architectures that can be used for data warehousing. What are they?

- **Generally speaking, these architectures are commonly called client/server or n-tier architectures, of which two-tier and three-tier architectures are the most common, but sometimes there is simply one tier.**

More data, coming in faster and requiring immediate conversion into decisions, means that organizations are confronting the need for real-time data warehousing (RDW). How would you define real-time data warehousing?

- **Real-time data warehousing, also known as active data warehousing (ADW), is the process of loading and providing data via the data warehouse as they become available.**

Mention briefly some of the recently popularized concepts and technologies that will play a significant role in defining the future of data warehousing.

- **Sourcing (mechanisms for acquisition of data from diverse and dispersed sources):**
  - **Web, social media, and Big Data**
  - **Open source software**
  - **SaaS (software as a service)**
  - **Cloud computing**
- **Infrastructure (architectural–hardware and software–enhancements):**
  - **Columnar (a new way to store and access data in the database)**
  - **Real-time data warehousing**
  - **Data warehouse appliances (all-in-one solutions to DW)**
  - **Data management technologies and practices**
  - **In-database processing technology (putting the algorithms where the data is)**
  - **In-memory storage technology (moving the data in the memory for faster processing)**
  - **New database management systems**
  - **Advanced analytics**

CHAPTER 4

TRUE OR FALSE

The WebFOCUS BI platform in the Travel and Transport case study decreased clients' reliance on the IT function when seeking system reports.	**True**
The dashboard for the WebFOCUS BI platform in the Travel and Transport case study required client side software to operate.	**False**
Data is the contextualization of information, that is, information set in context.	**True**
The main difference between service level agreements and key performance indicators is the audience.	**True**
The balanced scorecard is a type of report that is based solely on financial metrics.	**False**
The data storage component of a business reporting system builds the various reports and hosts them for, or disseminates them to users. It also provides notification, annotation, collaboration, and other services.	**False**
In the FEMA case study, the BureauNet software was the primary reason behind the increased speed and relevance of the reports FEMA employees received.	**True**
Google Maps has set new standards for data visualization with its intuitive Web mapping software.	**True**
There are basic chart types and specialized chart types. A Gantt chart is a specialized chart type.	**True**
Visualization differs from traditional charts and graphs in complexity of data sets and use of multiple dimensions and measures.	**True**
When telling a story during a presentation, it is best to avoid describing hurdles that your character must overcome, to avoid souring the mood.	**False**
For best results when deploying visual analytics environments, focus only on power users and management to get the best return on your investment.	**False**
Information density is a key characteristic of performance dashboards.	**True**
In the Dallas Cowboys case study, the focus was on using data analytics to decide which players would play every week.	**False**
One comparison typically made when data is presented in business intelligence systems is a comparison against historical values.	**True**
The best key performance indicators are derived independently from the company's strategic goals to enable developers to "think outside of the box."	**False**
The BPM development cycle is essentially a one-shot process where the requirement is to get it right the first time.	**False**
With key performance indicators, driver KPIs have a significant effect on outcome KPIs, but the reverse is not necessarily true.	**True**
With the balanced scorecard approach, the entire focus is on measuring and managing specific financial goals based on the organization's strategy.	**False**
A Six Sigma deployment can be deemed effective even if the number of defects are not reduced to 3.4 defects per million.	**False**

MULTIPLE CHOICE

1. For those executives who do not have the time to go through lengthy reports, the best alternative is the
   A) last page of the report.
   B) raw data that informed the report.
   C) charts in the report.
   D) executive summary.
2. All of the following are true about external reports between businesses and the government EXCEPT
   A) their primary focus is government.
   B) they can be filed nationally or internationally.
   C) they are standardized for the most part to reduce the regulatory burden.
   D) they can include tax and compliance reporting.
3. Kaplan and Norton developed a report that presents an integrated view of success in the organization called
   A) dashboard-type reports.
   B) balanced scorecard-type reports.
   C) metric management reports.
   D) visual reports.

4. Which component of a reporting system contains steps detailing how recorded transactions are converted into metrics, scorecards, and dashboards?
   - A) assurance
   - B) extract, transform and load
   - C) data supply
   - **D) business logic**
5. Which of the following is LEAST related to data/information visualization?
   - A) statistical graphics
   - B) information graphics
   - **C) graphic artwork**
   - D) scientific visualization
6. The Internet emerged as a new medium for visualization and brought all the following EXCEPT
   - A) immersive environments for consuming data.
   - **B) new forms of computation of business logic.**
   - C) worldwide digital distribution of visualization.
   - D) new graphics displays through PC displays.
7. Which kind of chart is described as an enhanced variant of a scatter plot?
   - A) heat map
   - **B) bubble chart**
   - C) pie chart
   - D) bullet
8. Which type of visualization tool can be very helpful when the intention is to show relative proportions of dollars per department allocated by a university administration?
   - A) heat map
   - B) bubble chart
   - **C) pie chart**
   - D) bullet
9. Which type of visualization tool can be very helpful when a data set contains location data?
   - **A) geographic map**
   - B) tree map
   - C) bar chart
   - D) highlight table
10. Which type of question does visual analytics seeks to answer?
   - A) What is happening today?
   - B) What happened yesterday?
   - C) When did it happen?
   - **D) Why did it happen?**
11. When you tell a story in a presentation, all of the following are true EXCEPT
   - A) stories and their lessons should be easy to remember.
   - **B) a well-told story should have no need for subsequent discussion.**
   - C) a story should make sense and order out of a lot of background noise.
   - D) the outcome and reasons for it should be clear at the end of your story.
12. Benefits of the latest visual analytics tools, such as SAS Visual Analytics, include all of the following EXCEPT
   - A) there is less demand on IT departments for reports.
   - B) mobile platforms such as the iPhone are supported by these products.
   - **C) they explore massive amounts of data in hours, not days.**
   - D) it is easier to spot useful patterns and trends in the data.
13. What is the management feature of a dashboard?
   - **A) operational data that identify what actions to take to resolve a problem**

B) summarized dimensional data to analyze the root cause of problems

C) graphical, abstracted data to monitor key performance metrics

D) summarized dimensional data to monitor key performance metrics

14. What is the fundamental challenge of dashboard design?

A) ensuring that the organization has access to the latest web browsers

B) ensuring that the organization has the appropriate hardware onsite to support it

C) ensuring that the required information is shown clearly on a single screen

D) ensuring that users across the organization have access to it

15. Contextual metadata for a dashboard includes all the following EXCEPT

A) which operating system is running the dashboard server software.

B) whether any high-value transactions that would skew the overall trends were rejected as a part of the loading process.

C) whether the dashboard is presenting "fresh" or "stale" information.

D) when the data warehouse was last refreshed.

16. Dashboards can be presented at all the following levels EXCEPT

A) the static report level.

B) the visual dashboard level.

C) the self-service cube level.

D) the visual cube level.

17. Why is a performance management system superior to a performance measurement system?

A) because measurement alone has little use without action

B) because performance management systems cost more

C) because performance measurement systems are only in their infancy

D) because measurement automatically leads to problem solution

18. Why is the customer perspective important in the balanced scorecard methodology?

A) because customers should always be included in any design methodology

B) because companies need customer input into the design of the balanced scorecard

C) because dissatisfied customers will eventually hurt the bottom line

D) because customers understand best how the firm's internal processes should work

19. All of the following statements about balanced scorecards and dashboards are true EXCEPT

A) scorecards are less preferred at operational and tactical levels.

B) scorecards are best for real-time tracking of a marketing campaign.

C) dashboards would be the preferred choice to monitor production quality.

D) scorecards are preferred for tracking the achievement of strategic goals.

20. What is Six Sigma?

A) a methodology aimed at measuring the amount of variability in a business process

B) a methodology aimed at reducing the number of defects in a business process

C) a letter in the Greek alphabet that statisticians use to measure process variability

D) a methodology aimed at reducing the amount of variability in a business process

FILL-IN-THE-BLANK

- A(n) **business report** is a communication artifact, concerning business matters, prepared with the specific intention of relaying information in a presentable form.
- Travel and Transport created an online BI self-service system that allowed **clients** to access information directly.
- There are only a few categories of business report: informal, **formal**, and short.
- In the Delta Lloyd Group case study, the **last mile** is the stage of the reporting process in which consolidated figures are cited, formatted, and described to form the final text of the report.
- **Metric** management reports are used to manage business performance through outcome-oriented metrics in many organizations.

- In the Blastrac case study, Tableau analytics software was used to replace massive **spreadsheets** that were loaded with data from multiple ERP systems.
- **Bar** charts are useful in displaying nominal data or numerical data that splits nicely into different categories so you can quickly see comparative results and trends.
- **PERT** charts or network diagrams show precedence relationships among the project activities/tasks.
- **Maps** are typically used together with other charts and graphs, as opposed to by themselves, and show postal codes, country names, etc.
- Typical charts, graphs, and other visual elements used in visualization-based applications usually involve **two** dimensions.
- Visual analytics is widely regarded as the combination of visualization and **predictive** analytics.
- Dashboards present visual displays of important information that are consolidated and arranged on a single **screen**.
- With dashboards, the layer of information that uses graphical, abstracted data to keep tabs on key performance metrics is the **monitoring** layer.
- In the Saudi Telecom company case study, information **visualization** software allowed managers to see trends and correct issues before they became problems.
- Performance dashboards enable **drill-down/drill-through** operations that allow the users to view underlying data sources and obtain more detail.
- With a dashboard, information on sources of the data being presented, the quality and currency of underlying data provide contextual **metadata** for users.
- Business performance management comprises a **closed-loop** set of processes that link strategy to execution with the goal of optimizing business performance.
- In the Mace case study, the IBM Cognos software enabled the rapid creation of integrated reports across 60 countries, replacing a large and complex **spreadsheet**.
- A strategically aligned metric is also known as a key **performance indicator**.
- The **internal business process** perspective of the organization suggested by the balanced scorecard focuses on business processes and how well they are running.

SHORT ANSWER

List and describe the three major categories of business reports.

- **Metric management reports.**
  - **Many organizations manage business performance through outcome-oriented metrics.**
  - **For external groups, these are service-level agreements (SLAs).**
  - **For internal management, they are key performance indicators (KPIs).**
- **Dashboard-type reports.**
  - **This report presents a range of different performance indicators on one page, like a dashboard in a car.**
  - **Typically, there is a set of predefined reports with static elements and fixed structure, but customization of the dashboard is allowed through widgets, views, and set targets for various metrics.**
- **Balanced scorecard—type reports.**
  - **This is a method developed by Kaplan and Norton that attempts to present an integrated view of success in an organization.**
  - **In addition to financial performance, balanced scorecard—type reports also include customer, business process, and learning and growth perspectives.**

List five types of specialized charts and graphs.

- **Histograms**
- **Gantt charts**
- **PERT charts**
- **Geographic maps**

- **Bullets**
- **Heat maps**
- **Highlight tables**
- **Tree maps**

According to Eckerson (2006), a well-known expert on BI dashboards, what are the three layers of information of a dashboard?

- **Monitoring.**
  - **Graphical, abstracted data to monitor key performance metrics.**
- **Analysis.**
  - **Summarized dimensional data to analyze the root cause of problems.**
- **Management.**
  - **Detailed operational data that identify what actions to take to resolve a problem.**

List five best practices of dashboard design.

- **Benchmark key performance indicators with industry standards**
- **Wrap the dashboard metrics with contextual metadata**
- **Validate the dashboard design by a usability specialist**
- **Prioritize and rank alerts/exceptions streamed to the dashboard**
- **Enrich the dashboard with business users' comments**
- **Present information in three different levels**
- **Pick the right visual construct using dashboard design principles**
- **Provide for guided analytics**

What are the four processes that define a closed-loop BPM cycle?

- **Strategize:**
  - **This is the process of identifying and stating the organization's mission, vision, and objectives, and developing plans (at different levels of granularity–strategic, tactical and operational) to achieve these objectives.**
- **Plan:**
  - **When operational managers know and understand the what (i.e., the organizational objectives and goals), they will be able to come up with the how (i.e., detailed operational and financial plans).**
  - **Operational and financial plans answer two questions:**
    - **What tactics and initiatives will be pursued to meet the performance targets established by the strategic plan?**
    - **What are the expected financial results of executing the tactics?**
- **Monitor/Analyze:**
  - **When the operational and financial plans are underway, it is imperative that the performance of the organization be monitored.**
  - **A comprehensive framework for monitoring performance should address two key issues:**
    - **What to monitor and how to monitor.**
- **Act and Adjust:**
  - **What do we need to do differently? Whether a company is interested in growing its business or simply improving its operations, virtually all strategies depend on new projects–creating new products, entering new markets, acquiring new customers or businesses, or streamlining some processes.**
  - **The final part of this loop is taking action and adjusting current actions based on analysis of problems and opportunities.**

List and describe five distinguishing features of key performance indicators.

- **Strategy.**
    - **KPIs embody a strategic objective.**
- **Targets.**
    - **KPIs measure performance against specific targets.**
    - **Targets are defined in strategy, planning, or budgeting sessions and can take different forms (e.g., achievement targets, reduction targets, absolute targets).**
- **Ranges.**
    - **Targets have performance ranges (e.g., above, on, or below target).**
- **Encodings.**
    - **Ranges are encoded in software, enabling the visual display of performance (e.g., green, yellow, red).**
    - **Encodings can be based on percentages or more complex rules.**
- **Time frames.**
    - **Targets are assigned time frames by which they must be accomplished.**
    - **A time frame is often divided into smaller intervals to provide performance mileposts.**
- **Benchmarks.**
    - **Targets are measured against a baseline or benchmark.**
    - **The previous year's results often serve as a benchmark, but arbitrary numbers or external benchmarks may also be used.**

What are the three nonfinancial objectives of the balanced scorecard?

- **Customer.**
    - **This defines how the organization should appear to its customers if it is to accomplish its vision.**
- **Internal business process.**
    - **This specifies the processes the organization must excel at in order to satisfy its shareholders and customers.**
- **Learning and growth.**
    - **This indicates how an organization can improve its ability to change and improve in order to achieve its vision.**

Six Sigma rests on a simple performance improvement model known as DMAIC. What are the steps involved?

- **Define.**
    - **Define the goals, objectives, and boundaries of the improvement activity.**
    - **At the top level, the goals are the strategic objectives of the company.**
    - **At lower levels–department or project levels–the goals are focused on specific operational processes.**
- **Measure.**
    - **Measure the existing system.**
    - **Establish quantitative measures that will yield statistically valid data.**
    - **The data can be used to monitor progress toward the goals defined in the previous step.**
- **Analyze.**
    - **Analyze the system to identify ways to eliminate the gap between the current performance of the system or process and the desired goal.**
- **Improve.**
    - **Initiate actions to eliminate the gap by finding ways to do things better, cheaper, or faster.**
    - **Use project management and other planning tools to implement the new approach.**
- **Control.**
    - **Institutionalize the improved system by modifying compensation and incentive systems, policies, procedures, manufacturing resource planning, budgets, operation instructions, or other management systems.**

What are the basic ingredients of a good collection of performance measures?

- **Measures should focus on key factors.**
- **Measures should be a mix of past, present, and future.**
- **Measures should balance the needs of shareholders, employees, partners, suppliers, and other stakeholders.**
- **Measures should start at the top and flow down to the bottom.**
- **Measures need to have targets that are based on research and reality rather than be arbitrary.**

In the Expedia case study, what three steps were taken to convert drivers of departmental performance into a scorecard?

- **Deciding how to measure satisfaction.**
    - **This required the group to determine which measures in the 20 databases would be useful for demonstrating a customer's level of satisfaction.**
    - **This became the basis for the scorecards and KPIs.**
- **Setting the right performance targets.**
    - **This required the group to determine whether KPI targets had short-term or long-term payoffs.**
- **Putting data into context.**
    - **The group had to tie the data to ongoing customer satisfaction projects.**

CHAPTER 5

TRUE OR FALSE

In the Cabela's case study, the SAS/Teradata solution enabled the direct marketer to better identify likely customers and market to them based mostly on external data sources.	**False**
The cost of data storage has plummeted recently, making data mining feasible for more firms.	**True**
Data mining can be very useful in detecting patterns such as credit card fraud, but is of little help in improving sales.	**False**
The entire focus of the predictive analytics system in the Infinity P&C case was on detecting and handling fraudulent claims for the company's benefit.	**False**
If using a mining analogy, "knowledge mining" would be a more appropriate term than "data mining".	**True**
Data mining requires specialized data analysts to ask ad hoc questions and obtain answers quickly from the system.	**False**
Ratio data is a type of categorical data.	**False**
Interval data is a type of numerical data.	**True**
In the Memphis Police Department case study, predictive analytics helped to identify the best schedule for officers in order to pay the least overtime.	**False**
In data mining, classification models help in prediction.	**True**
Statistics and data mining both look for data sets that are as large as possible.	**False**
Using data mining on data about imports and exports can help to detect tax avoidance and money laundering.	**True**
In the cancer research case study, data mining algorithms that predict cancer survivability with high predictive power are good replacements for medical professionals.	**False**
During classification in data mining, a false positive is an occurrence classified as true by the algorithm while being false in reality.	**True**
When training a data mining model, the testing dataset is always larger than the training dataset.	**False**
When a problem has many attributes that impact the classification of different patterns, decision trees may be a useful approach.	**True**
In the 2degrees case study, the main effectiveness of the new analytics system was in dissuading potential churners from leaving the company.	**True**
Market basket analysis is a useful and entertaining way to explain data mining to a technologically less savvy audience, but it has little business significance.	**False**
The number of users of free/open source data mining software now exceeds that of users of commercial software versions.	**True**
Data that is collected, stored, and analyzed in data mining is often private and personal. There is no way to maintain individuals' privacy other than being very careful about physical data security.	**False**

MULTIPLE CHOICE

1. In the Cabela's case study, what types of models helped the company understand the value of customers, using a five-point scale?
   A) simulation and geographical models
   B) clustering and association models
   C) simulation and regression models
   D) reporting and association models
2. Understanding customers better has helped Amazon and others become more successful. The understanding comes primarily from
   A) collecting data about customers and transactions.
   B) asking the customers what they want.
   C) analyzing the vast data amounts routinely collected.
   D) developing a philosophy that is data analytics-centric.
3. All of the following statements about data mining are true EXCEPT
   A) the valid aspect means that the discovered patterns should hold true on new data.
   B) the process aspect means that data mining should be a one-step process to results.
   C) the novel aspect means that previously unknown patterns are discovered.
   D) the potentially useful aspect means that results should lead to some business benefit.
4. What is the main reason parallel processing is sometimes used for data mining?

A) because the hardware exists in most organizations and it is available to use
B) because any strategic application requires parallel processing
C) because the most of the algorithms used for data mining require it
D) because of the massive data amounts and search efforts involved

5. The data field "ethnic group" can be best described as
A) ordinal data.
B) ratio data.
C) interval data.
D) nominal data.

6. The data field "salary" can be best described as
A) ratio data.
B) nominal data.
C) ordinal data.
D) interval data.

7. Which broad area of data mining applications analyzes data, forming rules to distinguish between defined classes?
A) associations
B) classification
C) visualization
D) clustering

8. Which broad area of data mining applications partitions a collection of objects into natural groupings with similar features?
A) visualization
B) clustering
C) classification
D) associations

9. The data mining algorithm type used for classification somewhat resembling the biological neural networks in the human brain is
A) decision trees.
B) artificial neural networks.
C) association rule mining.
D) cluster analysis.

10. Identifying and preventing incorrect claim payments and fraudulent activities falls under which type of data mining applications?
A) retailing and logistics
B) insurance
C) computer hardware and software
D) customer relationship management

11. All of the following statements about data mining are true EXCEPT
A) understanding the data, e.g., the relevant variables, is critical to success.
B) understanding the business goal is critical.
C) building the model takes the most time and effort.
D) data is typically preprocessed and/or cleaned before use.

12. Which data mining process/methodology is thought to be the most comprehensive, according to kdnuggets.com rankings?
A) CRISP-DM
B) SEMMA
C) KDD Process
D) proprietary organizational methodologies

13. Prediction problems where the variables have numeric values are most accurately defined as
A) associations.
B) regressions.
C) computations.
D) classifications.

14. What does the robustness of a data mining method refer to?

A) its ability to construct a prediction model efficiently given a large amount of data
B) its ability to overcome noisy data to make somewhat accurate predictions
C) its speed of computation and computational costs in using the mode
D) its ability to predict the outcome of a previously unknown data set accurately

15. What does the scalability of a data mining method refer to?
A) its ability to construct a prediction model efficiently given a large amount of data
B) its ability to overcome noisy data to make somewhat accurate predictions
C) its ability to predict the outcome of a previously unknown data set accurately
D) its speed of computation and computational costs in using the mode

16. In estimating the accuracy of data mining (or other) classification models, the true positive rate is
A) the ratio of correctly classified positives divided by the sum of correctly classified positives and incorrectly classified negatives.
B) the ratio of correctly classified positives divided by the sum of correctly classified positives and incorrectly classified positives.
C) the ratio of correctly classified positives divided by the total positive count.
D) the ratio of correctly classified negatives divided by the total negative count.

17. In data mining, finding an affinity of two products to be commonly together in a shopping cart is known as
A) association rule mining.
B) decision trees.
C) cluster analysis.
D) artificial neural networks.

18. Third party providers of publicly available datasets protect the anonymity of the individuals in the data set primarily by
A) letting individuals in the data know their data is being accessed.
B) asking data users to use the data ethically.
C) removing identifiers such as names and social security numbers.
D) leaving in identifiers (e.g., name), but changing other variables.

19. In the Target case study, why did Target send a teen maternity ads?
A) Target's analytic model suggested she was pregnant based on her buying habits.
B) Target's analytic model confused her with an older woman with a similar name.
C) Target was using a special promotion that targeted all teens in her geographical area.
D) Target was sending ads to all women in a particular neighborhood.

20. Which of the following is a data mining myth?
A) Data mining requires a separate, dedicated database.
B) Newer Web-based tools enable managers of all educational levels to do data mining.
C) The current state-of-the-art is ready to go for almost any business.
D) Data mining is a multistep process that requires deliberate, proactive design and use.

FILL-IN-THE-BLANK

- In the opening vignette, Cabela's uses SAS data mining tools to create **predictive** models to optimize customer selection for all customer contacts.
- There has been an increase in data mining to deal with global competition and customers' more sophisticated **needs** and wants.
- Knowledge extraction, pattern analysis, data archaeology, information harvesting, pattern searching, and data dredging are all alternative names for **data mining**.
- Data are often buried deep within very large **databases**, which sometimes contain data from several years.
- **Categorical data** represent the labels of multiple classes used to divide a variable into specific groups, examples of which include race, sex, age group, and educational level.
- In the Memphis Police Department case study, shortly after all precincts embraced Blue CRUSH, **predictive analytics** became one of the most potent weapons in the Memphis police department's crime-fighting arsenal.

- Patterns have been manually **extracted** from data by humans for centuries, but the increasing volume of data in modern times has created a need for more automatic approaches.
- While prediction is largely experience and opinion based, **forecasting** is data and model based.
- Whereas **statistics** starts with a well-defined proposition and hypothesis, data mining starts with a loosely defined discovery statement.
- Customer **relationship** management extends traditional marketing by creating one-on-one relationships with customers.
- In the terrorist funding case study, an observed price **deviation** may be related to income tax avoidance/evasion, money laundering, or terrorist financing.
- Data preparation, the third step in the CRISP-DM data mining process, is more commonly known as **data preprocessing**.
- The data mining in cancer research case study explains that data mining methods are capable of extracting patterns and **relationships** hidden deep in large and complex medical databases.
- Fayyad et al. (1996) defined **knowledge discovery** in databases as a process of using data mining methods to find useful information and patterns in the data.
- In **k-fold cross-validation**, a classification method, the complete data set is randomly split into mutually exclusive subsets of approximately equal size and tested multiple times on each left-out subset, using the others as a training set.
- The basic idea behind a **decision tree** is that it recursively divides a training set until each division consists entirely or primarily of examples from one class.
- As described in the 2degrees case study, a common problem in the mobile telecommunications industry is defined by the term **customer churn**, which means customers leaving.
- Because of its successful application to retail business problems, association rule mining is commonly called **market-basket analysis**.
- The **Apriori algorithm** is the most commonly used algorithm to discover association rules. Given a set of itemsets, the algorithm attempts to find subsets that are common to at least a minimum number of the itemsets.
- One way to accomplish privacy and protection of individuals' rights when data mining is by **de-identification** of the customer records prior to applying data mining applications, so that the records cannot be traced to an individual.

SHORT ANSWER

List five reasons for the growing popularity of data mining in the business world.

- **More intense competition at the global scale driven by customers' ever-changing needs and wants in an increasingly saturated marketplace**
- **General recognition of the untapped value hidden in large data sources**
- **Consolidation and integration of database records, which enables a single view of customers, vendors, transactions, etc.**
- **Consolidation of databases and other data repositories into a single location in the form of a data warehouse**
- **The exponential increase in data processing and storage technologies**
- **Significant reduction in the cost of hardware and software for data storage and processing**
- **Movement toward the de-massification (conversion of information resources into nonphysical form) of business practices**

What are the differences between nominal, ordinal, interval and ratio data? Give examples.

- **Nominal data contain measurements of simple codes assigned to objects as labels, which are not measurements.**
    - **For example, the variable marital status can be generally categorized as (1) single, (2) married, and (3) divorced.**
- **Ordinal data contain codes assigned to objects or events as labels that also represent the rank order among them.**
    - **For example, the variable credit score can be generally categorized as (1) low, (2) medium, or (3) high.**
    - **Similar ordered relationships can be seen in variables such as age group (i.e., child, young, middle-aged, elderly) and educational level (i.e., high school, college, graduate school).**
- **Interval data are variables that can be measured on interval scales.**
    - **A common example of interval scale measurement is temperature on the Celsius scale.**
    - **In this particular scale, the unit of measurement is 1/100 of the difference between the melting temperature and the boiling temperature of water in atmospheric pressure; that is, there is not an absolute zero value.**
- **Ratio data include measurement variables commonly found in the physical sciences and engineering.**
    - **Mass, length, time, plane angle, energy, and electric charge are examples of physical measures that are ratio scales.**
    - **Informally, the distinguishing feature of a ratio scale is the possession of a nonarbitrary zero value.**
    - **For example, the Kelvin temperature scale has a nonarbitrary zero point of absolute zero.**

List and briefly describe the six steps of the CRISP-DM data mining process.

- **Step 1: Business Understanding - The key element of any data mining study is to know what the study is for.**
    - **Answering such a question begins with a thorough understanding of the managerial need for new knowledge and an explicit specification of the business objective regarding the study to be conducted.**
- **Step 2: Data Understanding - A data mining study is specific to addressing a well-defined business task, and different business tasks require different sets of data.**
    - **Following the business understanding, the main activity of the data mining process is to identify the relevant data from many available databases.**
- **Step 3: Data Preparation - The purpose of data preparation (or more commonly called data preprocessing) is to take the data identified in the previous step and prepare it for analysis by data mining methods.**
    - **Compared to the other steps in CRISP-DM, data preprocessing consumes the most time and effort; most believe that this step accounts for roughly 80 percent of the total time spent on a data mining project.**
- **Step 4: Model Building - Here, various modeling techniques are selected and applied to an already prepared data set in order to address the specific business need.**
    - **The model-building step also encompasses the assessment and comparative analysis of the various models built.**
- **Step 5: Testing and Evaluation - In step 5, the developed models are assessed and evaluated for their accuracy and generality.**

  - **This step assesses the degree to which the selected model (or models) meets the business objectives and, if so, to what extent (i.e., do more models need to be developed and assessed).**
- **Step 6: Deployment - Depending on the requirements, the deployment phase can be as simple as generating a report or as complex as implementing a repeatable data mining process across the enterprise.**
  - **In many cases, it is the customer, not the data analyst, who carries out the deployment steps.**

Describe the role of the simple split in estimating the accuracy of classification models.

- **The simple split (or holdout or test sample estimation) partitions the data into two mutually exclusive subsets called a training set and a test set (or holdout set).**
- **It is common to designate two-thirds of the data as the training set and the remaining one-third as the test set.**
- **The training set is used by the inducer (model builder), and the built classifier is then tested on the test set.**
- **An exception to this rule occurs when the classifier is an artificial neural network.**
- **In this case, the data is partitioned into three mutually exclusive subsets: training, validation, and testing.**

Briefly describe five techniques (or algorithms) that are used for classification modeling.

- **Decision tree analysis.**
  - **Decision tree analysis (a machine-learning technique) is arguably the most popular classification technique in the data mining arena.**
- **Statistical analysis.**
  - **Statistical techniques were the primary classification algorithm for many years until the emergence of machine-learning techniques.**
  - **Statistical classification techniques include logistic regression and discriminant analysis.**
- **Neural networks.**
  - **These are among the most popular machine-learning techniques that can be used for classification-type problems.**
- **Case-based reasoning.**
  - **This approach uses historical cases to recognize commonalities in order to assign a new case into the most probable category.**
- **Bayesian classifiers.**
  - **This approach uses probability theory to build classification models based on the past occurrences that are capable of placing a new instance into a most probable class (or category).**
- **Genetic algorithms.**
  - **This approach uses the analogy of natural evolution to build directed-search-based mechanisms to classify data samples.**
- **Rough sets.**
  - **This method takes into account the partial membership of class labels to predefined categories in building models (collection of rules) for classification problems.**

Describe cluster analysis and some of its applications.

- **Cluster analysis is an exploratory data analysis tool for solving classification problems.**
- **The objective is to sort cases (e.g., people, things, events) into groups, or clusters, so that the degree of association is strong among members of the same cluster and weak among members of different clusters.**
- **Cluster analysis is an essential data mining method for classifying items, events, or concepts into common groupings called clusters.**
- **The method is commonly used in biology, medicine, genetics, social network analysis, anthropology, archaeology, astronomy, character recognition, and even in MIS development.**
- **As data mining has increased in popularity, the underlying techniques have been applied to business, especially to marketing.**

- **Cluster analysis has been used extensively for fraud detection (both credit card and e-commerce fraud) and market segmentation of customers in contemporary CRM systems.**

In the data mining in Hollywood case study, how successful were the models in predicting the success or failure of a Hollywood movie?

- **The researchers claim that these prediction results are better than any reported in the published literature for this problem domain.**
- **Fusion classification methods attained up to 56.07% accuracy in correctly classifying movies and 90.75% accuracy in classifying movies within one category of their actual category.**
- **The SVM classification method attained up to 55.49% accuracy in correctly classifying movies and 85.55% accuracy in classifying movies within one category of their actual category.**

In lessons learned from the Target case, what legal warnings would you give another retailer using data mining for marketing?

- **If you look at this practice from a legal perspective, you would conclude that Target did not use any information that violates customer privacy; rather, they used transactional data that most every other retail chain is collecting and storing (and perhaps analyzing) about their customers.**
- **What was disturbing in this scenario was perhaps the targeted concept: pregnancy.**
- **There are certain events or concepts that should be off limits or treated extremely cautiously, such as terminal disease, divorce, and bankruptcy.**

List four myths associated with data mining.

- **Data mining provides instant, crystal-ball-like predictions.**
- **Data mining is not yet viable for business applications.**
- **Data mining requires a separate, dedicated database.**
- **Only those with advanced degrees can do data mining.**
- **Data mining is only for large firms that have lots of customer data.**

List six common data mining mistakes.

- **Selecting the wrong problem for data mining**
- **Ignoring what your sponsor thinks data mining is and what it really can and cannot do**
- **Leaving insufficient time for data preparation**
- **Looking only at aggregated results and not at individual records**
- **Being sloppy about keeping track of the data mining procedure and results**
- **Ignoring suspicious findings and quickly moving on**
- **Running mining algorithms repeatedly and blindly**
- **Believing everything you are told about the data**
- **Believing everything you are told about your own data mining analysis**
- **Measuring your results differently from the way your sponsor measures them**

CHAPTER 6

TRUE OR FALSE

In the opening vignette, the high accuracy of the models in predicting the outcomes of complex medical procedures showed that data mining tools are ready to replace experts in the medical field.	**False**
Though useful in business applications, neural networks are a rough, inexact model of how the brain works, not a precise replica.	**True**
The use of hidden layers and new topologies and algorithms renewed waning interest in neural networks.	**True**
Compared to the human brain, artificial neural networks have many more neurons.	**False**
In the mining industry case study, the input to the neural network is a verbal description of a hanging rock on the mine wall.	**False**
The network topology that allows only one-way links between layers, with no feedback linkage permitted, is known as backpropagation.	**True**
With a neural network, outputs are attributes of the problem while inputs are potential solutions to the problem.	**False**
The most complex problems solved by neural networks require one or more hidden layers for increased accuracy.	**True**
The task undertaken by a neural network does not affect the architecture of the neural network; in other words, architectures are problem-independent.	**False**
Prior to starting the development of a neural network, developers must carry out a requirements analysis.	**True**
No matter the topology or architecture of a neural network, they all use the same algorithm to adjust weights during training.	**False**
Neural networks are called "black boxes" due to the lack of ability to explain their reasoning.	**True**
Generally speaking, support vector machines are less accurate a prediction method than other approaches such as decision trees and neural networks.	**False**
Unlike other "black box" predictive models, support vector machines have a solid mathematical foundation in statistics.	**True**
In the student retention case study, support vector machines used in prediction had proportionally more true positives than true negatives.	**True**
Using support vector machines, you must normalize the data before you numericize it.	**False**
The k-nearest neighbor algorithm is overly complex when compared to artificial neural networks and support vector machines.	**False**
The k-nearest neighbor algorithm appears well-suited to solving image recognition and categorization problems.	**True**
In the Coors case study, a neural network was used to more skillfully identify which beer flavors could be predicted.	**True**
In the Coors case study, genetic algorithms were of little use in solving the flavor prediction problem.	**False**

MULTIPLE CHOICE

1. In the opening vignette, predictive modeling is described as

   A) estimating the future using the past.
   B) not yet accepted in the business world.
   C) unable to handle complex predictive problems.
   D) the least practiced branch of data mining.

2. In the opening vignette, which method was the best in both accuracy of predicted outcomes and sensitivity?

   A) CART
   B) SVM
   C) ANN
   D) C5

3. Neural networks have been described as "biologically inspired." What does this mean?

   A) They are faithful to the entire process of computation in the human brain.
   B) They have the power to undertake every task the human brain can.
   C) They crudely model the biological makeup of the human brain.
   D) They were created to look identical to human brains.

4. Which element in an artificial neural network roughly corresponds to a synapse in a human brain?

   A) node
   B) weight

C) output
D) input

5. Which element in an artificial neural network roughly corresponds to a dendrite in a human brain?
   - **A)** output
   - **B)** node
   - **C) input**
   - **D)** weight
6. All the following statements about hidden layers in artificial neural networks are true EXCEPT
   - **A)** more hidden layers increase required computation exponentially.
   - **B)** hidden layers are not direct inputs or outputs.
   - **C) many top commercial ANNs forgo hidden layers completely.**
   - **D)** more hidden layers include many more weights.
7. In developing an artificial neural network, all of the following are important reasons to pre-select the network architecture and learning method EXCEPT
   - **A) most neural networks need special purpose hardware, which may be absent.**
   - **B)** development personnel may be more experienced with certain architectures.
   - **C)** some neural network software may not be available in the organization.
   - **D)** some configurations have better success than others with specific problems.
8. Backpropagation learning algorithms for neural networks are
   - **A)** used without a training set of data.
   - **B)** used without hidden layers for effectiveness.
   - **C)** the least popular algorithm due to their inaccuracy.
   - **D) required to have error tolerance set in advance.**
9. Why is sensitivity analysis frequently used for artificial neural networks?
   - **A)** because it is generally informative, although it cannot help to identify cause-and-effect relationships among variables
   - **B) because some consequences of mistakes by the network might be fatal, so justification may matter**
   - **C)** because it provides a complete description of the inner workings of the artificial neural network
   - **D)** because it is required by all major artificial neural networks
10. Support vector machines are a popular machine learning technique primarily because of
    - **A)** their relative cost and relative ease of use.
    - **B)** their relative cost and superior predictive power.
    - **C) their superior predictive power and their theoretical foundation.**
    - **D)** their high effectiveness in the very few areas where they can be used.
11. In the student retention case study, which of the following variables was MOST important in determining whether a student dropped out of college?
    - **A)** marital status and hours enrolled
    - **B)** college and major
    - **C) completed credit hours and hours enrolled**
    - **D)** high school GPA and SAT high score math
12. In the student retention case study, of the four data mining methods used, which was the most accurate?
    - **A)** ANN
    - **B)** LR
    - **C) SVM**
    - **D)** DT(C5)
13. When using support vector machines, in which stage do you transform the data?
    - **A)** deploying the model
    - **B)** experimentation
    - **C)** developing the model
    - **D) preprocessing the data**

14. When using support vector machines, in which stage do you select the kernel type (e.g., RBF, Sigmoid)?
    - **A)** deploying the model
    - **B)** preprocessing the data
    - **C)** experimentation
    - **D) developing the model**
15. For how long do SVM models continue to be accurate and actionable?
    - **A)** for as long as you choose to use them
    - **B)** for as long as the developers stay with the firm
    - **C)** for as long as management support continues to exist for the project
    - **D) for as long as the behavior of the domain stays the same**
16. All of the following are disadvantages/limitations of the SVM technique EXCEPT
    - **A)** they have high algorithmic complexity and extensive memory requirements for complex tasks.
    - **B) their accuracy is poor in many domains compared to neural networks.**
    - **C)** model building involves complex and time-demanding calculations.
    - **D)** selection of the kernel type and kernel function parameters is difficult.
17. The k-nearest neighbor machine learning algorithm (kNN) is
    - **A)** highly mathematical and computationally intensive.
    - **B) regarded as a "lazy" learning method.**
    - **C)** very complex in its inner workings.
    - **D)** a method that has little in common with regression.
18. Using the k-nearest neighbor machine learning algorithm for classification, larger values of k
    - **A)** do not change the effect of noise on the classification.
    - **B)** increase the effect of noise on the classification.
    - **C) reduce the effect of noise on the classification.**
    - **D)** sharpen the distinction between classes.
19. What is a major drawback to the basic majority voting classification in kNN?
    - **A)** Classes that are more clustered tend to dominate prediction.
    - **B)** Even the naive version of the algorithm is hard to implement.
    - **C)** It requires frequent human subjective input during computation.
    - **D) Classes with more frequent examples tend to dominate prediction.**
20. In the Coors case study, why was a genetic algorithm paired with neural networks in the prediction of beer flavors?
    - **A) to complement the neural network by reducing the error term**
    - **B)** to best model how the flavor of beer evolves as it ages
    - **C)** to enhance the neural network by pre-selecting output classes for the neural network
    - **D)** to replace the neural network in harder cases

FILL-IN-THE-BLANK

- The opening vignette teaches us that **evidence-based** medicine is a relatively new term coined in the healthcare arena, where the main idea is to dig deep into past experiences to discover new and useful knowledge to improve medical and managerial procedures in healthcare.
- Neural computing refers to a **pattern-recognition** methodology for machine learning.
- A thorough analysis of an early neural network model called the **perceptron**, which used no hidden layer, in addition to a negative evaluation of the research potential by Minsky and Papert in 1969, led to a diminished interest in neural networks.
- In a neural network, groups of neurons can be organized in a number of different ways; these various network patterns are referred to as **topologies**.
- In a typical network structure of an ANN consisting of three layers–input, intermediate, and output–the intermediate layer is called the **hidden** layer.
- In an ANN, **connection weights** express the relative strength (or mathematical value) of the input data or the many connections that transfer data from layer to layer.
- Kohonen's **self-organizing** feature maps provide a way to represent multidimensional data in much lower dimensional spaces, usually one or two dimensions.
- In the power generators case study, data mining—driven software tools, including data-driven **predictive modeling** technologies with historical data, helped an energy company reduce emissions of NOx and CO.
- The development process for an ANN application involves **nine** steps.
- **Backpropagation** is the most widely used supervised learning algorithm in neural computing.
- **Sensitivity analysis** has proved the most popular of the techniques proposed for shedding light into the "black-box" characterization of trained neural networks.
- In the formulation of the traffic accident study in the traffic case study, the five-class prediction problem was decomposed into a number of **binary classification** models in order to obtain the granularity of information needed.
- **Support vector machines (SVMs)** are of particular interest to modeling highly nonlinear, complex problems, systems, and processes and use hyperplanes to separate output classes in training data.
- The student retention case study shows that, given sufficient data with the proper variables, data mining techniques are capable of predicting freshman student attrition with approximately **80** percent accuracy.
- In the mathematical formulation of SVM's, the normalization and/or scaling are important steps to guard against variables/attributes with **larger variance** that might otherwise dominate the classification formulae.
- Writing the SVM classification rule in its dual form reveals that classification is only a function of the **support vectors**, i.e., the training data that lie on the margin.
- In machine learning, the **kernel trick** is a method for converting a linear classifier algorithm into a nonlinear one by using a nonlinear function to map the original observations into a higher-dimensional space.
- Due largely to their better classification results, support vector machines (SVMs) have recently become a popular technique for **classification**-type problems.
- Historically, the development of ANNs followed a heuristic path, with applications and extensive experimentation preceding theory. In contrast to ANNs, the development of SVMs involved sound **statistical learning** theory first, then implementation and experiments.
- In the process of image recognition (or categorization), images are first transformed into a multidimensional **feature space** and then, using machine-learning techniques, are categorized into a finite number of classes.

SHORT ANSWER

Predictive modeling is perhaps the most commonly practiced branch in data mining. What are three of the most popular predictive modeling techniques?

- **Artificial neural networks**
- **Support vector machines**
- **k-nearest neighbor**

Why have neural networks shown much promise in many forecasting and business classification applications?

- **Because of their ability to "learn" from the data, their nonparametric nature (i.e., no rigid assumptions), and their ability to generalize**

Each ANN is composed of a collection of neurons that are grouped into layers. One of these layers is the hidden layer. Define the hidden layer.

- **A hidden layer is a layer of neurons that takes input from the previous layer and converts those inputs into outputs for further processing.**

How is a general Hopfield network represented architecturally?

- **Architecturally, a general Hopfield network is represented as a single large layer of neurons with total interconnectivity; that is, each neuron is connected to every other neuron within the network.**

Describe the nine steps in the development process for an ANN application.

- **Collect, organize, and format the data**
- **Separate data into training, validation, and testing sets**
- **Decide on a network architecture and structure**
- **Select a learning algorithm**
- **Set network parameters and initialize their values**
- **Initialize weights and start training (and validation)**
- **Stop training, freeze the network weights**
- **Test the trained network**
- **Deploy the network for use on unknown new cases**

What are the five steps in the backpropagation learning algorithm?

- **Initialize weights with random values and set other parameters.**
- **Read in the input vector and the desired output.**
- **Compute the actual output via the calculations, working forward through the layers.**
- **Compute the error.**
- **Change the weights by working backward from the output layer through the hidden layers.**

Define the term sensitivity analysis as it relates to ANNs.

- **Sensitivity analysis is a method for extracting the cause-and-effect relationships among the inputs and the outputs of a trained neural network model.**

In 1992, Boser, Guyon, and Vapnik suggested a way to create nonlinear classifiers by applying the kernel trick to maximum-margin hyperplanes. How does the resulting algorithm differ from the original optimal hyperplane algorithm proposed by Vladimir Vapnik in 1963?

- **The resulting algorithm is formally similar, except that every dot product is replaced by a nonlinear kernel function.**
- **This allows the algorithm to fit the maximum-margin hyperplane in the transformed feature space.**
- **The transformation may be nonlinear and the transformed space high dimensional; thus, though the classifier is a hyperplane in the high-dimensional feature space it may be nonlinear in the original input space.**

What are the three steps in the process-based approach to the use of support vector machines (SVMs)?

- **Numericizing the data**
- **Normalizing the data**
- **Selecting the kernel type and kernel parameters**

Describe the k-nearest neighbor (kNN) data mining algorithm.

- **k-NN is a prediction method for classification- as well as regression-type prediction problems.**
- **k-NN is a type of instance-based learning (or lazy learning) where the function is only approximated locally and all computations are deferred until the actual prediction.**

CHAPTER 7

TRUE OR FALSE

In the chapter's opening vignette, IBM's computer named Watson outperformed human game champions on the game show Jeopardy!	**True**
Text analytics is the subset of text mining that handles information retrieval and extraction, plus data mining.	**False**
In text mining, inputs to the process include unstructured data such as Word documents, PDF files, text excerpts, e-mail and XML files.	**True**
During information extraction, entity recognition (the recognition of names of people and organizations) takes place after relationship extraction.	**False**
Categorization and clustering of documents during text mining differ only in the preselection of categories.	**True**
Articles and auxiliary verbs are assigned little value in text mining and are usually filtered out.	**True**
In the patent analysis case study, text mining of thousands of patents held by the firm and its competitors helped improve competitive intelligence, but was of little use in identifying complementary products.	**False**
The bag-of-words model is appropriate for spam detection but not for text analytics.	**True**
Chinese, Japanese, and Thai have features that make them more difficult candidates for natural language processing.	**True**
Regional accents present challenges for natural language processing.	**True**
In the Hong Kong government case study, reporting time was the main benefit of using SAS Business Analytics to generate reports.	**True**
Detecting lies from text transcripts of conversations is a future goal of text mining as current systems achieve only 50% accuracy of detection.	**False**
In the financial services firm case study, text analysis for associate-customer interactions were completely automated and could detect whether they met the company's standards.	**True**
In text mining, creating the term-document matrix includes all the terms that are included in all documents, making for huge matrices only manageable on computers.	**False**
In text mining, if an association between two concepts has 7% support, it means that 7% of the documents had both concepts represented in the same document.	**True**
In sentiment analysis, sentiment suggests a transient, temporary opinion reflective of one's feelings.	**False**
Current use of sentiment analysis in voice of the customer applications allows companies to change their products or services in real time in response to customer sentiment.	**True**
In sentiment analysis, it is hard to classify some subjects such as news as good or bad, but easier to classify others, e.g., movie reviews, in the same way.	**True**
The linguistic approach to speech handles processes elements such as intensity, pitch and jitter from speech recorded on audio.	**False**
In the BBVA case study, text analytics was used to help the company defend and enhance its reputation in social media.	**True**

MULTIPLE CHOICE

1. In the opening vignette, the architectural system that supported Watson used all the following elements EXCEPT
   A) a core engine that could operate seamlessly in another domain without changes.
   B) massive parallelism to enable simultaneous consideration of multiple hypotheses.
   C) integration of shallow and deep knowledge.
   D) an underlying confidence subsystem that ranks and integrates answers.
2. According to a study by Merrill Lynch and Gartner, what percentage of all corporate data is captured and stored in some sort of unstructured form?
   A) 15%
   B) 85%
   C) 25%
   D) 75%
3. Which of these applications will derive the LEAST benefit from text mining?
   A) sales transaction files
   B) patent description files
   C) customer comment files

D) patients' medical files

4. In text mining, stemming is the process of
   A) creating new branches or stems of recorded paragraphs.
   B) reducing multiple words to their base or root.
   C) categorizing a block of text in a sentence.
   D) transforming the term-by-document matrix to a manageable size.
5. In text mining, tokenizing is the process of
   A) transforming the term-by-document matrix to a manageable size.
   B) reducing multiple words to their base or root.
   C) categorizing a block of text in a sentence.
   D) creating new branches or stems of recorded paragraphs.
6. All of the following are challenges associated with natural language processing EXCEPT
   A) dividing up a text into individual words in English.
   B) understanding the context in which something is said.
   C) recognizing typographical or grammatical errors in texts.
   D) distinguishing between words that have more than one meaning.
7. What application is MOST dependent on text analysis of transcribed sales call center notes and voice conversations with customers?
   A) CRM
   B) OLAP
   C) finance
   D) ERP
8. In text mining, which of the following methods is NOT used to reduce the size of a sparse matrix?
   A) eliminating rarely occurring terms
   B) using singular value decomposition
   C) using a domain expert
   D) normalizing word frequencies
9. What data discovery process, whereby objects are categorized into predetermined groups, is used in text mining?
   A) classification
   B) trend analysis
   C) association
   D) clustering
10. In the research literature case study, the researchers analyzing academic papers extracted information from which source?
   A) the paper abstract
   B) the paper references
   C) the main body of the paper
   D) the paper keywords
11. Sentiment classification usually covers all the following issues EXCEPT
   A) range of polarity (e.g., star ratings for hotels and for restaurants).
   B) classes of sentiment (e.g., positive versus negative).
   C) range in strength of opinion.
   D) biometric identification of the consumer expressing the sentiment.
12. In sentiment analysis, which of the following is an implicit opinion?
   A) The cruise we went on last summer was a disaster.
   B) Our new mayor is great for the city.
   C) The customer service I got for my TV was laughable.
   D) The hotel we stayed in was terrible.
13. In the Whirlpool case study, the company sought to better understand information coming from which source?
   A) delivery information

B) customer transaction data

C) goods moving through the internal supply chain

D) customer e-mails

14. What do voice of the market (VOM) applications of sentiment analysis do?

A) They examine the "market of ideas" in politics.

B) They examine employee sentiment in the organization.

C) They examine customer sentiment at the aggregate level.

D) They examine the stock market for trends.

15. How is objectivity handled in sentiment analysis?

A) It is clarified with the customer who expressed it.

B) It is incorporated as a type of sentiment.

C) It is identified and removed as facts are not sentiment.

D) It is ignored because it does not appear in customer sentiment.

16. Identifying the target of an expressed sentiment is difficult for all the following reasons EXCEPT

A) the review may not be directly connected to the target through the topic name.

B) sometimes there are multiple targets expressed in a sentiment.

C) strong sentiments may be generated by a computer, not a person.

D) blogs and articles with the sentiment may be general in nature.

17. In text analysis, what is a lexicon?

A) a catalog of words, their synonyms, and their meanings

B) a catalog of customers, their words, and phrase

C) a catalog of customers, products, words, and phrase

D) a catalog of letters, words, phrases and sentences

18. What types of documents are BEST suited to semantic labeling and aggregation to determine sentiment orientation?

A) collections of documents

B) medium- to large-sized documents

C) large-sized documents

D) small- to medium-sized documents

19. Inputs to speech analytics include all of the following EXCEPT

A) written transcripts of calls to service centers.

B) recorded conversations of customer call-ins.

C) videos of customer focus groups.

D) live customer interactions with service representatives.

20. In the Blue Cross Blue Shield case study, speech analytics were used to identify "confusion" calls by customers. What was true about these calls?

A) They were not documented by customer service reps for speech analytics.

B) They took less time than others as frustrated customers hung up.

C) They led customers to rely more on self-serve options.

D) They were difficult to identify using standard phrases like "I don't get it."

FILL-IN-THE-BLANK

- IBM's Watson utilizes a massively parallel, text mining—focused, probabilistic evidence-based computational architecture called **DeepQA**.
- **Named entity extraction** is probably the most often used form of information extraction.
- **Polysemes**, also called homonyms, are syntactically identical words with different meanings.
- When a word has more than one meaning, selecting the meaning that makes the most sense can only be accomplished by taking into account the context within which the word is used. This concept is known as **word sense disambiguation**.
- **Sentiment analysis** is a technique used to detect favorable and unfavorable opinions toward specific products and services using large numbers of textual data sources.

- In the text mining system developed by Ghani et al., treating products as sets of **attribute-value pairs** rather than as atomic entities can potentially boost the effectiveness of many business applications.
- In the Mining for Lies case study, a text based deception-detection method used by Fuller and others in 2008 was based on a process known as **message-feature mining**, which relies on elements of data and text mining techniques.
- At a very high level, the text mining process can be broken down into three consecutive tasks, the first of which is to establish the **Corpus**.
- Because the term-document matrix is often very large and rather sparse, an important optimization step is to reduce the **dimensionality** of the matrix.
- Where **sentiment** appears in text, it comes in two flavors: explicit, where the subjective sentence directly expresses an opinion, and implicit, where the text implies an opinion.
- **Voice of the customer (VOC)** is mostly driven by sentiment analysis and is a key element of customer experience management initiatives, where the goal is to create an intimate relationship with the customer.
- **Brand management** focuses on listening to social media where anyone can post opinions that can damage or boost your reputation.
- In sensitivity analysis, the task of differentiating between a fact and an opinion can also be characterized as calculation of **Objectivity-Subjectivity (OS)** polarity.
- When identifying the polarity of text, the most granular level for polarity identification is at the **word** level.
- When viewed as a binary feature, **polarity** classification is the binary classification task of labeling an opinionated document as expressing either an overall positive or an overall negative opinion.
- When labeling each term in the WordNet lexical database, the group of cognitive synonyms (or synset) to which this term belongs is classified using a set of **ternary classifiers**, each of which is capable of deciding whether the synset is Positive, or Negative, or Objective.
- In automated sentiment analysis, two primary methods have been deployed to predict sentiment within audio: acoustic/phonetic and **linguistic** modeling.
- The time-demanding and laborious process of the **acoustic/phonetic** approach makes it impractical for use with live audio streams.
- **Linguistic** models operate on the premise that, when in a charged state, a speaker has a higher probability of using specific words, exclamations, or phrases in a particular order.
- Among the significant advantages associated with the **phonetic indexing and search** approach to linguistic modeling is the method's ability to maintain a high degree of accuracy no matter what the quality of the audio source, and its incorporation of conversational context through the use of structured queries.

SHORT ANSWER

When IBM Research began looking for a major research challenge to rival the scientific and popular interest of Deep Blue, the computer chess-playing champion, what was the company's goal?

- **The goal was to advance computer science by exploring new ways for computer technology to affect science, business, and society, and which would also have clear relevance to IBM business interests.**

What is the definition of text analytics according to the experts in the field?

- **Text analytics is a broader concept that includes information retrieval as well as information extraction, data mining, and Web mining.**

How would you describe information extraction in text mining?

- **Information extraction is the identification of key phrases and relationships within text by looking for predefined objects and sequences in text by way of pattern matching.**

Natural language processing (NLP), a subfield of artificial intelligence and computational linguistics, is an important component of text mining. What is the definition of NLP?

- **NLP is a discipline that studies the problem of "understanding" the natural human language, with the view of converting depictions of human language into more formal representations in the form of numeric and symbolic data that are easier for computer programs to manipulate.**

In the security domain, one of the largest and most prominent text mining applications is the highly classified ECHELON surveillance system. What is ECHELON assumed to be capable of doing?

- **Identifying the content of telephone calls, faxes, e-mails, and other types of data and intercepting information sent via satellites, public switched telephone networks, and microwave links**

Describe the query-specific clustering method as it relates to clustering.

- **This method employs a hierarchical clustering approach where the most relevant documents to the posed query appear in small tight clusters that are nested in larger clusters containing less similar documents, creating a spectrum of relevance levels among the documents.**

Name and briefly describe four of the most popular commercial software tools used for text mining.

- **ClearForest offers text analysis and visualization tools.**
- **IBM offers SPSS Modeler and data and text analytics toolkits.**
- **Megaputer Text Analyst offers semantic analysis of free-form text, summarization, clustering, navigation, and natural language retrieval with search dynamic refocusing.**
- **SAS Text Miner provides a rich suite of text processing and analysis tools.**
- **KXEN Text Coder (KTC) offers a text analytics solution for automatically preparing and transforming unstructured text attributes into a structured representation for use in KXEN Analytic Framework.**
- **The Statistica Text Mining engine provides easy-to-use text mining functionality with exceptional visualization capabilities.**
- **VantagePoint provides a variety of interactive graphical views and analysis tools with powerful capabilities to discover knowledge from text databases.**
- **The WordStat analysis module from Provalis Research analyzes textual information such as responses to open-ended questions, interviews, etc.**
- **Clarabridge text mining software provides end-to-end solutions for customer experience professionals wishing to transform customer feedback for marketing, service, and product improvements.**

Sentiment analysis has many names. Which other names is it often known by?

- **Sentiment analysis is often referred to as opinion mining, subjectivity analysis, and appraisal extraction.**

Identify, with a brief description, each of the four steps in the sentiment analysis process.

- **Sentiment Detection:**
    - **Here the goal is to differentiate between a fact and an opinion, which may be viewed as classification of text as objective or subjective.**
- **N-P Polarity Classification:**
    - **Given an opinionated piece of text, the goal is to classify the opinion as falling under one of two opposing sentiment polarities, or locate its position on the continuum between these two polarities.**
- **Target Identification:**
    - **The goal of this step is to accurately identify the target of the expressed sentiment.**
- **Collection and Aggregation:**
    - **In this step all text data points in the document are aggregated and converted to a single sentiment measure for the whole document.**

Within the context of speech analytics, what does the linguistic approach focus on?

- **The linguistic approach focuses on the explicit indications of sentiment and context of the spoken content within the audio.**

CHAPTER 8

TRUE OR FALSE

Participating in social media is so new that it is still optional for most companies in the United States.	**False**
Web mining is exactly the same as Web analytics: the analysis of Web site usage data.	**False**
Web crawlers or spiders collect information from Web pages in an automated or semi-automated way. Only the text of Web pages is collected by crawlers.	**False**
Generally, making a search engine more efficient makes it less effective.	**True**
With the PageRank algorithm, a Web page with more incoming links will always rank higher than one with fewer incoming links.	**False**
The main purpose of frequent recrawling of some Web sites is to prevent search users from retrieving stale search results.	**True**
Search engine optimization (SEO) techniques play a minor role in a Web site's search ranking because only well-written content matters.	**False**
Clickstream analysis does not need users to enter their perceptions of the Web site or other feedback directly to be useful in determining their preferences.	**True**
Having more Web traffic coming from organic search than other types of search is the goal of most companies.	**True**
Since little can be done about visitor Web site abandonment rates, organizations have to focus their efforts on increasing the number of new visitors.	**False**
It is possible to use prescriptive tools for Web analytics to describe current Web site use comprehensively.	**False**
Many Web analytics tools are free to download and use, including Google Web Analytics.	**True**
Voice of customer (VOC) applications track and resolve business process and usability obstacles for a Web site.	**False**
Social network analysis can help companies divide their customers into market segments by analyzing their interconnections.	**True**
Decentralization, the need for specialized skills, and immediacy of output are all attributes of Web publishing when compared to industrial publishing.	**False**
Consistent high quality, higher publishing frequency, and longer time lag are all attributes of industrial publishing when compared to Web publishing.	**False**
Web site visitors who critique and create content are more engaged than those who join networks and spectate.	**True**
Descriptive analytics for social media feature such items as your followers as well as the content in online conversations that help you to identify themes and sentiments.	**False**
Companies understand that when their product goes "viral," the content of the online conversations about their product does not matter, only the volume of conversations.	**False**
Social media analytics companies provide integrated support that is helpful to many parts of a business, not only the Sales and Marketing functions.	**True**

MULTIPLE CHOICE

1. What does Web content mining involve?
    - A) analyzing the universal resource locator in Web pages
    - B) analyzing the pattern of visits to a Web site
    - **C) analyzing the unstructured content of Web pages**
    - D) analyzing the PageRank and other metadata of a Web page
2. What does Web structure mining involve?
    - A) analyzing the pattern of visits to a Web site
    - **B) analyzing the universal resource locators in Web pages**
    - C) analyzing the PageRank and other metadata of a Web page
    - D) analyzing the unstructured content of Web pages
3. In the extremist groups case study, what approach is used to discover the ideology and fund raising of extremist groups through their Web sites?
    - **A) content analysis**
    - B) physical visits to addresses on the site
    - C) hyperlink analysis

D) e-mail responses to questions sent to the sites

4. Search engines do not search the entire Web every time a user makes a search request, for all the following reasons EXCEPT

A) it would take longer than the user could wait.

B) it is more efficient to use pre-stored search results.

C) most users are not interested in searching the entire Web.

D) the Web is too complex to be searched each time.

5. Breaking up a Web page into its components to identify worthy words/terms and indexing them using a set of rules is called

A) parsing the documents.

B) document analysis.

C) creating the term-by-document matrix.

D) preprocessing the documents.

6. PageRank for Webpages is useful to Web developers for which of the following reasons?

A) They uniquely identify the Web page developer for greater accountability.

B) It gives developers insight into Web user behavior.

C) Developing many Web pages with low PageRank can help a Web site attract users.

D) It is used in citation analysis for scholarly papers.

7. Search engine optimization (SEO) is a means by which

A) Web site developers can negotiate better deals for paid ads.

B) Web site developers can increase Web site search rankings.

C) Web site developers optimize the artistic features of their Web sites.

D) Web site developers index their Web sites for search engines.

8. In general, what is the best kind of Web traffic to a Web site?

A) European Web traffic

B) organic Web traffic

C) paid Web traffic

D) bot-generated traffic

9. Clickstream analysis is most likely to be used for all the following types of applications EXCEPT

A) predicting user behavior.

B) hiring new functional area managers.

C) designing cross-marketing strategies across products.

D) determining the lifetime value of clients.

10. What are the two main types of Web analytics?

A) off-site and on-site Web analytics

B) data-based and subjective Web analytics

C) old-school and new-school Web analytics

D) Bing and Google Web analytics

11. Web site usability may be rated poor if

A) users fail to click on all pages equally.

B) Web site visitors download few of your offered PDFs and videos.

C) the time spent on your Web site is long.

D) the average number of page views on your Web site is large.

12. Common sources of traffic to your Web site include all of the following EXCEPT

A) direct links.

B) referral Web sites.

C) accidental visitors.

D) paid search from search engines.

13. Understanding which keywords your users enter to reach your Web site through a search engine can help you understand

A) the hardware your Web site is running on.

B) how well visitors understand your products.

C) the type of Web browser being used by your Web site visitors.

D) most of your Web site visitors' wants and needs.

14. Which of the following statements about Web site conversion statistics is FALSE?

A) The conversion rate is the number of people who take action divided by the number of visitors.

B) Analyzing exit rates can tell you why visitors left your Web site.

C) Web site visitors can be classed as either new or returning.

D) Visitors who begin a purchase on most Web sites must complete it.

15. A voice of customer (VOC) strategy involves all of the following EXCEPT

A) capturing both unstructured Web data and enterprise data as a starting point.

B) connecting captured insights to unstructured data in order to take action.

C) taking actions related to your market, customers and services.

D) analyzing unstructured data with minimal effort on the user's part.

16. All of the following statements about social networks are true EXCEPT

A) companies should invest equally to retain all members of a group.

B) it is possible to gain insights into how products go viral.

C) a group with all interconnected individuals is called a clique.

D) members of a group are affected by the behavior of others in the group.

17. What is one major way that Web-based social media is the same as publishing media?

A) They can both reach a global audience.

B) They cost the same to publish.

C) They require the same skill and training to publish.

D) They have the same immediacy of updates.

18. What is one major way in which Web-based social media differs from traditional publishing media?

A) Web-based media have a narrower range of quality.

B) They use different languages of publication.

C) They have different costs to own and operate.

D) Most Web-based media are operated by the government and large firms.

19. What does descriptive analytics for social media do?

A) It examines the content of online conversations.

B) It helps identify your followers.

C) It identifies links between groups.

D) It identifies the biggest sources of influence online.

20. What does advanced analytics for social media do?

A) It helps identify your followers.

B) It identifies the biggest sources of influence online.

C) It identifies links between groups.

D) It examines the content of online conversations.

FILL-IN-THE-BLANK

- The **Web** is perhaps the world's largest data and text repository, and the amount of information on it is growing rapidly.
- Web pages contain both unstructured information and **hyperlinks**, which are connections to other Web pages.
- Web **mining** involves discovering relationships from Web pages.
- Web **crawlers/spiders** are used to automatically read through the contents of Web sites.
- A(n) **hub** is one or more Web pages that provide a collection of links to authoritative Web pages.
- A(n) **search** engine is a software program that searches for Web sites or files based on keywords.
- In the IGN case, IGN Entertainment used search engine optimization to increase their search engine rankings and thereby their **organic** search engine traffic.
- **Google** is far and away the most popular search engine.
- In the Lotte.com retail case, the company deployed SAS for Customer Experience Analytics to better understand the quality of customer traffic on their Web site, classify order rates, and see which **channels** had the most visitors.

- Off-site Web analytics refers to measurement and analysis of data relating to your company that takes place outside your Web site.
- Analyzing server log files is the traditional way to collect Web site information for on-site Web analytics.
- A low number of **page** views may be the result of poor Web site design.
- A **referral** Web site contains links that send traffic directly to your Web site.
- **Conversion** statistics help you understand whether your specific marketing objective for a Web page is being achieved.
- Google Web **Analytics** generates detailed statistics about a Web site's traffic and traffic sources and tracks conversions.
- Social networks consist of nodes, representing individuals or organizations and **ties/connections**, which relate them.
- In the Social Network Analysis (SNA) for Telecommunications case, SNA can be used to detect churners, i.e., those visitors who about to leave the Web site and persuade them to stay with you.
- **Propinquity** is a connections metric for social networks that measures the ties that actors in a network have with others that are geographically close.
- **Cohesion** is a segmentation metric for social networks that measures the strength of the bonds between actors in a social network.
- Advanced **analytics** examine the content in online conversations to identify themes, sentiments, and connections.

SHORT ANSWER

In what ways does the Web pose great challenges for effective and efficient knowledge discovery through data mining?

- **The Web is too big for effective data mining.**
    - **The Web is so large and growing so rapidly that it is difficult to even quantify its size.**
    - **Because of the sheer size of the Web, it is not feasible to set up a data warehouse to replicate, store, and integrate all of the data on the Web, making data collection and integration a challenge.**
- **The Web is too complex.**
    - **The complexity of a Web page is far greater than a page in a traditional text document collection.**
    - **Web pages lack a unified structure.**
    - **They contain far more authoring style and content variation than any set of books, articles, or other traditional text-based document.**
- **The Web is too dynamic.**
    - **The Web is a highly dynamic information source. Not only does the Web grow rapidly, but its content is constantly being updated.**
    - **Blogs, news stories, stock market results, weather reports, sports scores, prices, company advertisements, and numerous other types of information are updated regularly on the Web.**
- **The Web is not specific to a domain.**
    - **The Web serves a broad diversity of communities and connects billions of workstations.**
    - **Web users have very different backgrounds, interests, and usage purposes.**
    - **Most users may not have good knowledge of the structure of the information network and may not be aware of the heavy cost of a particular search that they perform.**
- **The Web has everything.**
    - **Only a small portion of the information on the Web is truly relevant or useful to someone (or some task).**
    - **Finding the portion of the Web that is truly relevant to a person and the task being performed is a prominent issue in Web-related research.**

What is a Web crawler and what function does it serve in a search engine?

- **A Web crawler (also called a spider or a Web spider) is a piece of software that systematically browses (crawls through) the World Wide Web for the purpose of finding and fetching Web pages.**
- **Often Web crawlers copy all the pages they visit for later processing by other functions of a search engine.**

What is search engine optimization (SEO) and why is it important for organizations that own Web sites?

- **Search engine optimization (SEO) is the intentional activity of affecting the visibility of an e-commerce site or a Web site in a search engine's natural (unpaid or organic) search results.**
    - **In general, the higher ranked on the search results page, and more frequently a site appears in the search results list, the more visitors it will receive from the search engine's users.**
- **Being indexed by search engines like Google, Bing, and Yahoo! is not good enough for businesses.**
    - **Getting ranked on the most wide used search engines and getting ranked higher than your competitors are what make the difference.**

What is the difference between white hat and black hat SEO activities?

- **An SEO technique is considered white hat if it conforms to the search engines' guidelines and involves no deception.**
    - **Because search engine guidelines are not written as a series of rules or commandments, this is an important distinction to note.**
    - **White-hat SEO is not just about following guidelines, but about ensuring that the content a search engine indexes and subsequently ranks is the same content a user will see.**
- **Black-hat SEO attempts to improve rankings in ways that are disapproved by the search engines, or involve deception or trying to trick search engine algorithms from their intended purpose.**

How would you define clickstream analysis?

- **Clickstream analysis is the analysis of information collected by Web servers to help companies understand user behavior better.**
    - **By using the data and text mining techniques, companies can frequently discern interesting patterns from the clickstreams.**
- **Data collected from clickstreams include user data, session data, which pages they viewed and when and how often they visited.**
    - **Knowledge extracted from clickstreams includes usage patterns, user profiles, page profiles, visit profiles and customer value.**

Why are the users' page views and time spent on your Web site important metrics?

- **If people come to your Web site and don't view many pages, that is undesirable and your Web site may have issues with its design or structure.**
    - **Another explanation for low page views is a disconnect in the marketing messages that brought them to the site and the content that is actually available.**
- **Generally, the longer a person spends on your Web site, the better it is.**
    - **That could mean they're carefully reviewing your content, utilizing interactive components you have available, and building toward an informed decision to buy, respond, or take the next step you've provided.**
    - **On the contrary, the time on site also needs to be examined against the number of pages viewed to make sure the visitor isn't spending his or her time trying to locate content that should be more readily accessible.**

How is a conversion defined on an organization's Web site? Give examples.

- **Each organization defines a "conversion" according to its specific marketing objectives.**
- **Some Web analytics programs use the term "goal" to benchmark certain Web site objectives, such as a certain number of visitors to a page, a completed registration form, or an online purchase.**

What is the Voice of the customer (VOC) strategy? List and describe its 4 steps.

- **Voice of the customer (VOC) is a term usually used to describe the analytic process of capturing a customer's expectations, preferences, and aversions.**
- **It essentially is a market research technique that produces a detailed set of customer wants and needs, organized into a hierarchical structure, and then prioritized in terms of relative importance and satisfaction with current alternatives.**
    - **Listen encompasses both the capability to listen to the open Web (forums, blogs, tweets, you name it) and the capability to seamlessly access enterprise information (CRM notes, documents, e-mails, etc.).**
    - **Analyze This is taking all of the unstructured data and making sense of it.**
        - **Solutions include keyword, statistical, and natural language approaches that will allow you to essentially tag or barcode every word and the relationships among words, making it data that can be accessed, searched, routed, counted, analyzed, charted, reported on, and even reused.**
    - **Relate After finding insights and analyzing unstructured data, here you connect those insights to your "structured" data about your customers, products, parts, locations and so on.**
    - **Act In this step, you act on the new customer insight you've obtained.**

What are the three categories of social media analytics technologies and what do they do?

- **Descriptive analytics:**
    - **Uses simple statistics to identify activity characteristics and trends, such as how many followers you have, how many reviews were generated on Facebook, and which channels are being used most often.**
- **Social network analysis:**
    - **Follows the links between friends, fans, and followers to identify connections of influence as well as the biggest sources of influence.**
- **Advanced analytics:**
    - **Includes predictive analytics and text analytics that examine the content in online conversations to identify themes, sentiments, and connections that would not be revealed by casual surveillance.**

In social network analysis, who are your most powerful influencers and why are they important?

- **Your most important influencers are the ones who influence the whole realm of conversation about your topic.**
- **You need to understand whether they are saying nice things, expressing support, or simply making observations or critiquing.**
- **What is the nature of their conversations? How is my brand being positioned relative to the competition in that space?**

CHAPTER 9

TRUE OR FALSE

Modeling can be viewed as a science in its entirety.	**False**
In the Midwest ISO opening vignette, the solution provided by the model's output determined the best output level to be produced by each power plant.	**True**
If linear programming can be successfully applied a problem, the output is usually optimal.	**True**
In the ExxonMobil case study, the approach taken was to find individual solutions to routing, transportation, scheduling, and inventory management, and select the best solution for one of the variables.	**False**
In order to be effective, analysts must use models to solve problems with no regard to the organizational culture to find optimal results.	**False**
In the Harrah's Cherokee Casino and Hotel case study, the revenue management system modified room prices based on demand and offered the same price/availability to all customers at any one time.	**False**
AHP can be used effectively for optimization with problems containing a small number of alternatives.	**True**
The trend is towards developing and using Web tools and software to access and run modeling software.	**True**
Using data cubes in OLAP systems opens the data up to analysis by more classes of models.	**False**
Another name for result variables is independent variables.	**False**
Taking a decision under risk is different from taking the decision under uncertainty.	**True**
Spreadsheets are the second most popular tool for modeling.	**False**
Linear programming seeks to optimally allocate resources among competing activities and is likely the best known optimization model.	**True**
When using Excel's Solver, we can have multiple constraints and multiple objective cells.	**False**
Most managerial problems can be properly evaluated and solved using a single goal, such as profit maximization.	**False**
Sensitivity analysis seeks to assess the impact of changes in the input data and parameters on the proposed solution.	**True**
Goal seeking is roughly the opposite of "what-if" analysis.	**True**
Using expected value (EV) with decision trees is totally appropriate for situations where one outcome could lead to an immense loss for the company.	**False**
In the U.S. HUD case study, the use of AHP brought standards and coherence to project selection, resulting in a 10% decrease in project requests from 1999 levels.	**False**
The analytic hierarchy process incorporates both qualitative and quantitative decision making criteria.	**True**

MULTIPLE CHOICE

1. Using modeling for decision support can currently achieve all of the following EXCEPT
   - **A)** enable organizations to see likely results of their decisions.
   - **B) replace strategy formulation at top levels of the organization.**
   - **C)** enhance the decision making process.
   - **D)** reduce the costs of providing services to customers.
2. Environmental scanning is important for all of the following reasons EXCEPT
   - **A)** organizational culture is important and affects the model use.
   - **B) environments have greater impact on a model than the organization does.**
   - **C)** environmental factors may have created the current problem.
   - **D)** it is critical to identify key corporate decision makers.
3. Today, it is critical for companies to consider
   - **A)** how to package products in the right format.
   - **B)** how to sell products at the right price.
   - **C)** how to get products to the right customer.
   - **D) all of the above**
4. Models can be built with the help of human knowledge and expertise. Another source of help in building these models is
   - **A) classification and clustering methods.**
   - **B)** the customer.
   - **C)** business partners.
   - **D)** customer service reps.

5. What is an influence diagram?
   - **A)** a map of the environment around decision makers
   - **B)** a map of the environment around a model
   - **C)** a diagram showing the influence of decision makers
   - **D) a graphical representation of a model**
6. Spreadsheets are particularly useful for all of the following reasons EXCEPT
   - **A)** they can be used to build static and dynamic models.
   - **B) they easily import and manipulate massive databases.**
   - **C)** they are able to import and export to many different file formats.
   - **D)** it is easy to manipulate data and see results instantly.
7. Linear programming belongs to a family of tools called
   - **A)** decision tree models.
   - **B) mathematical programming models.**
   - **C)** qualitative models.
   - **D)** heuristic programming models.
8. Which of the following is NOT a component of a linear programming problem?
   - **A) internal metrics**
   - **B)** decision variables
   - **C)** constraints
   - **D)** objective function
9. In an LP model, what does the fourth hidden component contain?
   - **A)** product mix variables
   - **B)** financial and accounting variables
   - **C)** constraint and limit variables
   - **D) slack and surplus variables**
10. Managers in organizations typically have
    - **A) multiple goals that need to be simultaneously or jointly optimized.**
    - **B)** single goals that cannot be optimized using linear and nonlinear programming.
    - **C)** single goals that can be optimized using linear and nonlinear programming.
    - **D)** a small number of goals that can be independently optimized using linear and nonlinear programming.
11. Sensitivity analysis is important in management support systems for all of the following reasons EXCEPT
    - **A)** it permits the manager to input data to increase his/her confidence in the model.
    - **B)** it allows flexibility and adaptation to changing conditions.
    - **C)** it provides a better understanding of the model and the decision-making situation.
    - **D) it improves the mathematical optimality of the generated solutions.**
12. The question "What will total earnings be if we reduce our inventory stocking costs by 10%?" is a type of
    - **A)** goal-seeking analysis.
    - **B) what-if analysis.**
    - **C)** sensitivity analysis.
    - **D)** utility modeling.
13. The question "What advertising budget is needed to increase market share by 7%?" is a type of
    - **A) goal-seeking analysis.**
    - **B)** what-if analysis.
    - **C)** sensitivity analysis.
    - **D)** utility modeling.
14. The question "How many servers will be needed to reduce the waiting time of restaurant customers to less than 9 minutes?" is a type of
    - **A) goal-seeking analysis.**
    - **B)** what-if analysis.

C) sensitivity analysis.
D) utility modeling.

15. Decision trees are best suited to solve what types of problems?
    A) problems with a single goal
    B) problems with a large number of alternatives
    C) problems where probabilities are unknown
    D) problems with a tabular representation
16. In handling uncertainty in decision modeling, the optimistic approach assumes
    A) the best possible outcome of most alternatives will occur.
    B) the best possible outcome of each alternative will occur.
    C) the best possible outcome of one alternative will occur.
    D) the best possible outcome of some alternatives will occur.
17. In handling uncertainty in decision modeling, what does the pessimistic approach do?
    A) It assumes the worst possible outcome of one alternative will occur and then avoids it.
    B) It assumes the worst possible outcome of some alternatives will occur and then selects the best of them.
    C) It assumes the worst possible outcome of each alternative will occur and then selects the best of them.
    D) It assumes the worst possible outcome of each alternative will occur and then selects the worst of them.
18. Which of the following statements about expected utility is true?
    A) In calculating utility, it assumes the decision will be made thousands of times, making the probabilities more likely on average.
    B) Used in decision making, it can bring huge risk to a small startup with limited resources.
    C) It does not affect decisions made with expected values.
    D) Used in decision making, it is an objective value, not subjective.
19. Which of the following statements about the analytic hierarchy process (AHP) is true?
    A) It can handle multiple criteria and goals.
    B) It is really not a decision model at all.
    C) It is based entirely on quantitative data.
    D) It is an opaque "black box" in the same way as neural networks.
20. Which of the following statements about the end-of-chapter CARE International case study is true?
    A) CARE used a linear programming model for optimization.
    B) CARE ran its own shipping operation with vehicles that needed route optimization.
    C) CARE set out to exclusively use international suppliers with large capacity to better serve people affected by disasters.
    D) CARE's objective was to respond to natural disasters faster.

FILL-IN-THE-BLANK

- In the opening vignette, Midwest ISO used optimization **algorithms/models** in their problem solving.
- Identifying a model's **variables** (e.g., decision and result) and their relationships is very important in creating and using models.
- **Forecasting** models are used to predict the future and are used widely in e-commerce.
- **Heuristic** modeling uses rules to determine solutions that are good enough.
- In non-quantitative models, the relationships are symbolic or **qualitative**.
- In decision-making, fixed factors that affect the result variables but are not manipulated by decision maker are called **uncontrollable** variables.
- Deciding to purchase an FDIC-insured Certificate of Deposit at a U.S. bank can be viewed as decision making under **certainty**.
- In the American Airlines case study, the modeling used for contract bidding could best be described as decision making under **risk**.
- In the Fred Astaire East Side Dance Studio case study, a(n) **scheduling** model was used to organize ballroom showcases and arrange participants and timeslots accordingly.
- In comparison to static models, **dynamic** models represent behavior over time.
- In the Fletcher Allen Health Care case, the **Solver** engine in Excel was used to find a feasible solution to the assignment problem.

- In mathematical programming, of available solutions, the **optimal** solution is the best; i.e., the degree of goal attainment associated with it is the highest.
- Most quantitative models are based on solving for a single **goal/objective**.
- Testing the robustness of decisions under changing conditions is an example of **sensitivity** analysis.
- Utility **theory** is a modeling method for handling multiple goals.
- **Goal** seeking calculates the values of the inputs necessary to achieve a desired level of an output.
- **Decision** tables conveniently organize information and knowledge in a systematic, tabular manner to prepare it for analysis and consideration of alternatives.
- A decision tree can be cumbersome if there are many **alternatives/choices** or states of nature.
- The analytic hierarchy process can be used to great effect to solve multi-**criteria** problems.
- In the end-of-chapter CARE International case study, an optimization model was used to decide on warehouse **location**.

SHORT ANSWER

Can customer relationship management (CRM) systems and revenue management systems (RMS) recommend not selling a particular product to certain customers? If so, why; if not, why not?

- **Yes, CRM and RMS can recommend ignoring certain customers or not selling a bundle of products to a particular set of customers.**
- **Part of this effort involves identifying lifelong customer profitability.**
- **These approaches rely heavily on forecasting techniques, which are typically described as predictive analytics.**
- **These systems attempt to predict who their best (i.e., most profitable) customers (and worst ones as well) are and focus on identifying products and services–or none at all–at appropriate prices to appeal to them.**

List and describe four categories of models. Give examples in each category.

- **Optimization of problems with few alternatives:**
  - **Find the best solution from a small number of alternatives; e.g., decision tables, decision trees, analytic hierarchy process**
- **Optimization via algorithm:**
  - **Find the best solution from a large number of alternatives, using a step-by-step improvement process; e.g., linear and other mathematical programming models, network models**
- **Optimization via an analytic formula:**
  - **Find the best solution in one step, using a formula; e.g., some inventory models**
- **Simulation:**
  - **Find a good enough solution or the best among the alternatives checked, using experimentation; e.g., Monte Carlo simulation**
- **Heuristics:**
  - **Find a good enough solution, using rules; e.g., heuristic programming, expert systems**
- **Predictive models:**
  - **Predict the future for a given scenario; e.g., forecasting models, Markov analysis**

All quantitative models are typically made up of four basic components. List and describe them as well as what links them together.

- **Result (outcome) variables reflect the level of effectiveness of a system; that is, they indicate how well the system performs or attains its goal(s).**
  - **These variables are outputs.**
- **Decision variables describe alternative courses of action.**
  - **The decision maker controls the decision variables.**

- **Uncontrollable variables or parameters are factors that affect the result variables but are not under the control of the decision maker.**
    - **Either these factors can be fixed, in which case they are called parameters, or they can vary, in which case they are called variables.**
- **Intermediate result variables reflect intermediate outcomes in mathematical models.**

Compare and contrast decision making under uncertainty, risk and certainty.

- **In decision making under certainty, it is assumed that complete knowledge is available so that the decision maker knows exactly what the outcome of each course of action will be (as in a deterministic environment).**
- **In decision making under uncertainty, the decision maker considers situations in which several outcomes are possible for each course of action.**
    - **In contrast to the risk situation, in this case, the decision maker does not know, or cannot estimate, the probability of occurrence of the possible outcomes.**
    - **Decision making under uncertainty is more difficult than decision making under certainty because there is insufficient information.**
- **In decision making under risk (also known as a probabilistic or stochastic decision making situation), the decision maker must consider several possible outcomes for each alternative, each with a given probability of occurrence.**

List four rational economic assumptions the linear programming allocation model is based upon.

- **Returns from different allocations can be compared; that is, they can be measured by a common unit (e.g., dollars, utility).**
- **The return from any allocation is independent of other allocations.**
- **The total return is the sum of the returns yielded by the different activities.**
- **All data are known with certainty.**
- **The resources are to be used in the most economical manner.**

List four difficulties that may arise when analyzing multiple goals.

- **It is usually difficult to obtain an explicit statement of the organization's goals.**
- **The decision maker may change the importance assigned to specific goals over time or for different decision scenarios.**
- **Goals and sub-goals are viewed differently at various levels of the organization and within different departments.**
- **Goals change in response to changes in the organization and its environment.**
- **The relationship between alternatives and their role in determining goals may be difficult to quantify.**
- **Complex problems are solved by groups of decision makers, each of whom has a personal agenda.**
- **Participants assess the importance (priorities) of the various goals differently.**

List four things sensitivity analyses are used for.

- **Revising models to eliminate too-large sensitivities**
- **Adding details about sensitive variables or scenarios**
- **Obtaining better estimates of sensitive external variables**
- **Altering a real-world system to reduce actual sensitivities**
- **Accepting and using the sensitive (and hence vulnerable) real world, leading to the continuous and close monitoring of actual results**

What is the most common method for treating risk in decision trees and tables?

- **The most common method for handling risk in decision trees and tables is to select the alternative with the greatest expected value.**

How are pairwise comparisons used in the analytic hierarchy process (AHP) to select an alternative?

- **To obtain the weights of selection criteria, the decision maker conducts pairwise comparisons of the criteria: first criterion to second, first to third, ..., first to last; then, second to third, ..., second to last; ...; and then the next-to-last criterion to the last one.**
    - **This establishes the importance of each criterion; that is, how much of the goal's weight is distributed to each criterion i.e., how important each criterion is.**
- **Beneath each criterion are the same sets of choices (alternatives) in the simple case described here.**
    - **Like the goal, the criteria decompose their weight into the choices, which capture 100 percent of the weight of each criterion.**
    - **The decision maker performs a pairwise comparison of choices in terms of preferences, as they relate to the specific criterion under consideration.**
    - **Each set of choices must be pairwise compared as they relate to each criterion.**
    - **Finally, the results are synthesized and displayed on a bar graph.**
    - **The choice with the most weight is the correct choice.**

In the CARE International end-of-chapter case study, what were the intended benefits of the model used and the actual

- **The main purpose of the model was to increase the capacity and swiftness to respond to sudden natural disasters like earthquakes, as opposed to other slow-occurring ones like famine.**
    - **Based on up-front cost, the model is able to provide the best optimized configuration of where to locate a warehouse and how much inventory should be kept.**
    - **It is able to provide an optimization result based on estimates of frequency, location, and level of potential demand that is generated by the model.**
- **Based on this model, CARE established three warehouses in the warehouse pre-positioning system in Dubai, Panama, and Cambodia.**
    - **In fact, during the Haiti earthquake crises in 2010, water purification kits were supplied to the victims from the Panama warehouse.**
    - **In the future, the pre-positioning network is expected to be expanded.**

CHAPTER 10

TRUE OR FALSE

In the Fluor case study, redesigning the process of reviewing engineering changes had no discernible impact on the bottom line.	**False**
In the choice phase of problem solving, normative models involve selecting an optimal or best outcome.	**True**
Analytical techniques for problem solving are best for unstructured rather than structured problems.	**False**
Heuristic approaches are typically used to solve more complex problems.	**True**
In the Chilean government case study, the government used complete enumeration to find the optimal solution for deciding meal providers to schools.	**False**
Genetic algorithms are heuristic methods that do not guarantee an optimal solution to a problem.	**True**
A "what-if" model is most typically used for the most structured problems.	**False**
In the Finnish Air Force case, the simulation had to take account of a finite number of possibilities relating to task times, material handling delays, etc.	**True**
The use of simulation models is desirable because they can usually be solved in one pass, without incurring the time and cost of iterations.	**False**
An advantage of simulation is that it allows model builders to solve problems with minimal interaction with users or managers.	**False**
Time compression in a simulation allows managers to test certain strategies with less risk.	**True**
Simulation solutions cannot easily be transferred from one problem domain to another.	**True**
Determining the duration of the simulation occurs before the model is validated and tested.	**False**
Discrete events and agent-based models are usually used for middle or low levels of abstraction.	**True**
In steady-state plant control design, time-independent simulation would be appropriate.	**True**
Simulation does not usually allow decision makers to see how a solution to a complex problem evolves over (compressed) time, nor can decision makers interact with the simulation.	**True**
Visual interactive simulation (VIS) is a simulation method that lets decision makers see what the model is doing and how it interacts with the decisions made, as they are made.	**True**
In the RFID case study, the key variable tested by the simulation model was quicker availability of information about the location of various parts in the supply chain.	**True**
Visual interactive modeling (VIM) systems, especially those developed for the military and the video-game industry, have "thinking" characters who can behave with a relatively high level of intelligence in their interactions with users.	**True**
In the Canadian pandemic case study, the macro-level simulation modeled aggregates of a population that might experience a pandemic.	**False**

MULTIPLE CHOICE

1. How does blind search differ from optimization?

   A) Blind search cannot result in optimal solutions whereas optimization methods do.

   B) Blind search represents a guided approach while optimization is unguided.

   C) Blind search usually does not conclude in one step like some optimization methods.

   D) Blind search is usually a more efficient problem solving approach than optimization.

2. In modeling, an optimal solution is understood to be

   A) a solution that can only be determined by an exhaustive enumeration and testing of alternatives.

   B) a solution found in the least possible time and using the least possible computing resources.

   C) a solution that is the best based on criteria defined in the design phase.

   D) a solution that requires an algorithm for determination.

3. When is a complete enumeration of solutions used?

   A) when a solution that is "good enough" is fine and good heuristics are available

   B) when there is enough time and computational power available

   C) when the modeler requires a guided approach to problem solving

   D) when there are an infinite number of solutions to be searched

4. All of the following are true about heuristics EXCEPT

   A) heuristics are used when the modeler requires a guided approach to problem solving.

B) heuristics are used when a solution that is "good enough" is sought.

C) heuristics are used when there is abundant time and computational power.

D) heuristics are rules of good judgment.

5. Which approach is most suited to structured problems with little uncertainty?

A) simulation

B) human intuition

C) optimization

D) genetic algorithms

6. Genetic algorithms belong to the family of methods in the

A) artificial intelligence area.

B) optimization area.

C) complete enumeration family of methods.

D) non-computer based (human) solutions area.

7. All of the following are suitable problems for genetic algorithms EXCEPT

A) dynamic process control.

B) pattern recognition with complex patterns.

C) simulation of biological models.

D) simple optimization with few variables.

8. Which approach is most suited to complex problems with significant uncertainty, a need for experimentation, and time compression?

A) simulation

B) optimization

C) human intuition

D) genetic algorithms

9. Which of the following is an advantage of simulation?

A) It can incorporate significant real-life complexity.

B) It always results in optimal solutions.

C) Simulation software requires special skills.

D) It solves problems in one pass with no iterations.

10. In which stage of the simulation methodology do you determine the variables and gather data?

A) defining the problem

B) constructing the simulation model

C) testing and validating the model

D) designing the experiment

11. In which stage of the simulation methodology do you determine how long to run the simulation?

A) constructing the simulation model

B) designing the experiment

C) testing and validating the model

D) defining the problem

12. In which stage of the simulation methodology do you determine the system's boundaries and environment?

A) constructing the simulation model

B) defining the problem

C) testing and validating the model

D) designing the experiment

13. What BEST describes a simulation model with a limited number of variables, each with a finite number of values?

A) system dynamics simulation

B) discrete event simulation

C) continuous distribution simulation

D) Monte Carlo simulation

14. What BEST describes a simulation model in which it is not important to know exactly when a modeled event occurred?
    A) continuous distribution simulation
    B) time-independent simulation
    C) system dynamics simulation
    D) discrete event simulation
15. The advantages of visual interactive simulation include all of the following EXCEPT
    A) improvements in training using the simulation.
    B) reduced need for decision maker involvement.
    C) the ability to see how a simulation works.
    D) improved presentation of simulation results.
16. What can system dynamics modeling be used for?
    A) qualitative methods for analyzing a system
    B) simulation models that test each subsystem in isolation
    C) micro-level simulation models that examine individual values
    D) studying system behavior at an instant in time
17. The EHR (electronic health record) system dynamics modeling example showed that
    A) increased electronic note-taking negatively affects compliance.
    B) e-notes negatively affect radiology performance.
    C) increased staff training results in increased electronic prescriptions.
    D) adverse drug events help to decrease patient time.
18. In agent-based modeling, agents are
    A) the human workers or agents who use the system.
    B) communication links between simulations.
    C) autonomous rule-based decision making units.
    D) the hardware platform used to conduct the simulation.
19. Agent-based modeling is best for all the following types of problem features EXCEPT
    A) complex interactions.
    B) low uncertainty.
    C) many interrelated factors.
    D) irregular data.
20. What is the final stage of an agent-based modeling (ABM) methodology?
    A) identifying the agents and determining their behavior
    B) determining agent-related data
    C) validating agent behavior against reality
    D) determining the suitability of ABM

FILL-IN-THE-BLANK

- **Blind** search techniques are arbitrary search approaches that are not guided.
- **Genetic algorithms** are a part of global search techniques used to find approximate solutions to optimization-type problems that are too complex to be solved with traditional optimization methods.
- A genetic algorithm is an iterative procedure that represents its candidate solutions as strings of genes called **chromosomes** and measures their viability with a fitness function.
- Candidate solutions (or chromosomes in genetic algorithms) combine to produce offspring in each algorithmic iteration. Along with the offspring, some of the best solutions are also migrated to the next generation in order to preserve the best solution achieved up until the current iteration. This concept is called **elitism**.
- In MSS, **simulation** is a technique for conducting experiments (e.g., what-if analyses) with a computer on a model of a management system.
- In the Hepatitis B case study, Markov models were used to determine the cost-**effectiveness** of various governmental interventions for Hepatitis B.
- Simulation is not strictly a type of model; models generally represent reality, whereas simulation typically **imitates** it.
- Simulation is the appearance of reality. Simulation is a **descriptive** rather than a normative method.
- One of the advantages of simulation is that a great amount of **time compression** can be attained, quickly giving a manager some feel as to the long-term (1- to 10-year) effects of many policies.
- Simulation involves setting up a **model** of a real system and conducting repetitive experiments on it.
- In **probabilistic** simulation, one or more of the independent variables follows certain probability distributions, which can be either discrete distributions or continuous distributions.
- The most common simulation method for business decision problems is **Monte Carlo** simulation.
- **Discrete event** simulation refers to building a model of a system where the interaction between different entities is studied.
- If simulation results do not match the intuition or judgment of the decision maker, a **confidence** gap in the results can occur.
- In a visual interactive simulation (VIS), **static** models display a visual image of the result of one decision alternative at a time.
- In a visual interactive simulation (VIS), **dynamic** models display systems that evolve over time, and the evolution is represented by animation.
- **System dynamics** models are macro-level simulation models in which aggregate values and trends are considered.
- In a system dynamics model, **causal loop** diagrams show the relationships between variables in a system.
- The performance of the agent-based system should be **validated** against reality.
- An agent-based modeling approach focuses on modeling a(n) "**adaptive learning**" property rather than "optimizing" nature.

SHORT ANSWER

List four well-known search methods used in the choice phase of problem solving.

- **Analytical techniques**
- **Algorithms**
- **Blind searching**
- **Heuristic searching**

Analytical techniques are used in the choice phase of problem solving. How can we define analytical techniques?

- **Analytical techniques use mathematical formulas to derive an optimal solution directly or to predict a certain result.**

Heuristic programming is the process of using heuristics in problem solving. This is done via heuristic search methods. Give a brief definition of the term heuristics.

- **Heuristics are the informal, judgmental knowledge of an application area that constitute the rules of good judgment in the field.**

Give a brief definition of genetic algorithms.

- **Genetic algorithms are sets of computational procedures that conceptually follow the steps of the biological process of evolution.**

List four key terms related to genetic algorithms.

- **Reproduction**
- **Crossover**
- **Mutation**
- **Elitism**

Describe three of the most important limitations of genetic algorithms according to Grupe and Jooste.

- **Not all problems can be framed in the mathematical manner that genetic algorithms demand.**
- **Development of a genetic algorithm and interpretation of the results require an expert who has both the programming and statistical/mathematical skills demanded by the genetic algorithm technology in use.**
- **It is known that in a few situations the "genes" from a few comparatively highly fit (but not optimal) individuals may come to dominate the population, causing it to converge on a local maximum.**
  - **When the population has converged, the ability of the genetic algorithm to continue to search for better solutions is effectively eliminated.**
- **Most genetic algorithms rely on random-number generators that produce different results each time the model runs.**
  - **Although there is likely to be a high degree of consistency among the runs, they may vary.**
- **Locating good variables that work for a particular problem is difficult.**
  - **Obtaining the data to populate the variables is equally demanding.**
- **Selecting methods by which to evolve the system requires thought and evaluation.**
  - **If the range of possible solutions is small, a genetic algorithm will converge too quickly on a solution.**
  - **When evolution proceeds too quickly, thereby altering good solutions too quickly, the results may miss the optimum solution.**

Genetic algorithms provide a set of efficient, domain-independent search heuristics for a broad spectrum of applications. List four possible applications for genetic algorithms.

- **Dynamic process control**
- **Induction of optimization of rules**
- **Discovery of new connectivity topologies (e.g., neural computing connections, neural network design)**
- **Simulation of biological models of behavior and evolution**
- **Complex design of engineering structures**
- **Pattern recognition**
- **Scheduling**
- **Transportation and routing**
- **Layout and circuit design**
- **Telecommunication**
- **Graph-based problems**

Give a simple definition of simulation in MSS.

- **A simulation is the appearance of reality.**

Under what circumstances is simulation normally used?

- **Simulation is normally used only when a problem is too complex to be treated using numerical optimization techniques.**

List five major types of simulation.

- **probabilistic simulation**
- **time-dependent and time-independent simulation**
- **visual simulation**
- **system dynamics modeling**
- **agent-based modeling**

CHAPTER 11

TRUE OR FALSE

Rules used in automated decision systems (ADS) can be derived based on experience.	**True**
Most business decision rules are the same across industries.	**False**
Flight pricing systems are examples of semi-automated decision systems that require managerial input for each decision.	**False**
A revenue management (RM) system for an airline seeks to minimize each customer's ticket price of travel on the airline's flights.	**False**
Rule-based systems have their roots in artificial intelligence.	**True**
Rich and Knight (1991) defined artificial intelligence as "the study of how to make people do things at which, at the moment, computers are better."	**False**
Expert systems (ES) are computer-based information systems that use expert knowledge to attain high-level decision performance in a narrowly defined problem domain.	**True**
A person's decision performance and level of knowledge are typical criteria that determine their level of expertise in a particular subject.	**True**
The basic rationale of artificial intelligence is to use mathematical calculation rather than symbolic reasoning.	**False**
While most first-generation Expert Systems (ES) use if-then rules to represent and store their knowledge, second-generation ES are more flexible in adopting multiple knowledge representation and reasoning methods.	**True**
The case study on chemical, biological, and radiological agents shows that expert systems are widely used in high pressure situations where the human decision makers are confident in taking quick actions.	**False**
A nonexpert uses the development environment of an expert system to obtain advice and to solve problems using the expert knowledge embedded into the system.	**False**
Knowledge acquisition from experts is a complex task that requires specialized expertise to conduct successfully.	**True**
The knowledge base in an expert system must correspond exactly to the format of the knowledge base in the organization where it will be utilized.	**False**
The inference engine, also known as the control structure or the rule interpreter (in rule-based ES), is essentially a computer program that provides a methodology for reasoning about information in the knowledge base and on the blackboard to formulate appropriate conclusions.	**True**
The critical component of a knowledge refinement system is the self-learning mechanism that allows it to adjust its knowledge base and its processing of knowledge based on the evaluation of its recent past performances.	**True**
Validation of knowledge is usually done by a human expert in the knowledge domain.	**True**
Once validated, the knowledge acquired from experts or induced from a set of data must be represented in a format that does not need to be understandable by humans but must be executable on computers.	**False**
Inference rules and knowledge rules are both used to solve problems in a rule-based expert system.	**True**
Unlike human experts, expert systems do not need to explain their views, recommendations, or decisions.	**False**

MULTIPLE CHOICE

1. In the InterContinental Hotel Group case study, the mathematical model used to increase profits was based on
   - **A) an optimization model that used multiple variables.**
   - B) a system that collated the subjective inputs of managers.
   - C) a simulation model that tried out many options.
   - D) a mathematical model that used two variables: price and day of the week.
2. Who are automated decision systems (ADS) primarily designed for?
   - **A) frontline workers who must make decisions rapidly**
   - B) mid-level managers making tactical decision
   - C) operational managers who make shop floor decisions
   - D) strategic level managers making long-term, wide-ranging decisions
3. In the Giant Food Stores case study, the new pricing model deployment system included all the following features EXCEPT
   - A) it could handle large numbers of price changes.
   - **B) it required more staff to make pricing changes.**
   - C) it had a predictive capability.
   - D) it used point of sale and competitive data as inputs.
4. Revenue management systems modify the prices of products and services dynamically based on

A) business rules, demand, and supply.

B) intuition, demand, and supply.

C) intuition, competition, and supply.

D) business rules, supply, and intuition.

5. What would explain why a divorce attorney in New York City may not be considered an expert in Beijing, China?

A) Expertise is frequently domain dependent.

B) The divorce attorney in New York does not speak Mandarin.

C) No criteria to evaluate divorce attorneys exist in Beijing.

D) You need a greater level of experience in Beijing to practice law.

6. What does self-knowledge in an expert system (ES) mean?

A) An ES understands itself in a very human sense.

B) The ES "knows" that it exists.

C) An ES understands the human decision maker.

D) The ES can explain how it reached a conclusion.

7. How does an expert system differ from conventional systems?

A) Expert systems handle qualitative data easily.

B) Execution of expert system programs is algorithmic or step-by-step.

C) The expert system operates only when it is complete.

D) Changes in an expert system are tedious to make.

8. In the sport talents identification case study, the expert system was calibrated with expertise from

A) multiple sports experts.

B) the system developer.

C) subjects in the cases used to create the ES.

D) one overall sports expert.

9. In the chemical, biological, and radiological agents case study, the CBR Advisor program had all the following features EXCEPT

A) it was available online to the general public.

B) it was created by multiple experts.

C) it could provide advice even incomplete information.

D) it was tailored to different types of users.

10. The MYCIN Expert System was used to diagnose bacterial infections using

A) a set of 500 rules on the subject.

B) an expert system whose performance was inferior to human experts.

C) a simulation model that tried out many options.

D) an optimization model.

11. Which module is missing from most expert systems?

A) inference engine

B) user interface subsystem

C) knowledge refinement subsystem

D) knowledge base subsystem

12. All the following statements about how an expert system operates are true EXCEPT

A) incorporated knowledge is drawn exclusively from human experts.

B) a knowledge engineer creates inferencing rules.

C) inference engines contain an explanation subcomponent.

D) knowledge rules are stored in the knowledge base.

13. In the heart disease diagnosis case study, what was a benefit of the SIPMES expert system?

A) No human expert knowledge was needed in development, only textbook knowledge.

B) The SIPMES system agreed with human experts 64% of the time.

C) The SIPMES system could diagnose all types of cardiovascular diseases.

D) Expert systems from other domains were used, saving development time.

14. Which of the following is NOT a stage of knowledge engineering?
    - A) knowledge representation
    - B) knowledge acquisition
    - **C) knowledge consolidation**
    - D) knowledge validation
15. It is difficult to acquire knowledge from experts for all the following reasons EXCEPT
    - A) experts often change their behavior when observed.
    - **B) many business areas have no identifiable experts.**
    - C) testing and refining of knowledge is complex and difficult.
    - D) experts may not be able to put into words how they conduct their work.
16. Using certainty factors, a rule declares that IF competition is strong, CF = 70 AND margins are above 15% CF = 100 THEN sales demand will decline. If both conditions are true, what is the CF of the conclusion?
    - A) 100%
    - B) 21%
    - **C) 70%**
    - D) 30%
17. Using certainty factors, a rule declares that IF competition is strong, CF = 70 OR margins are above 15% CF = 100 THEN sales demand will decline. If both conditions are true, what is the CF of the conclusion?
    - A) 30%
    - B) 21%
    - C) 70%
    - **D) 100%**
18. Which category of expert systems that includes weather forecasting and economic/financial forecasting?
    - A) planning ES
    - B) instruction ES
    - **C) prediction ES**
    - D) diagnostic ES
19. Which tool would be best to use when there is a need to very rapidly and cheaply develop a rule-based expert system?
    - A) ASP.NET
    - B) LISP or Prolog languages
    - **C) an ES shell**
    - D) C++
20. In the Clinical Decision Support System case study, what was the system's output?
    - A) a referral to specialists who could accurately diagnose the tendon injury
    - B) a diagnosis of the type of tendon injury suffered by the patient
    - C) an explanation of the tendon anatomy of the patient
    - **D) a treatment and rehabilitation plan for the patient**

FILL-IN-THE-BLANK

- Rules derived from data **mining** can be used effectively in automated decision systems.
- **Artificial intelligence (AI)** is a collection of concepts and ideas that are related to the development of intelligent systems.
- Expert systems mimic the reasoning process of **human** experts in order to solve problems.
- The accumulation, transfer, and transformation of problem-solving expertise from experts or documented knowledge sources to a computer program for constructing or expanding the knowledge base is known as **knowledge acquisition**.
- The ability of human experts to analyze their own knowledge and its effectiveness, learn from it, and improve on it for future consultations is known as a **knowledge-refining system**.
- The knowledge possessed by human experts is often lacking in **structure** and not explicitly expressed.
- **Knowledge** is a collection of specialized facts, procedures, and judgment usually expressed as rules.
- Knowledge rules, or **declarative** rules, state all the facts and relationships about a problem.

- Inference rules, or **procedural** rules, offer advice on how to solve a problem, given that certain facts are known.
- **Inferencing** (or reasoning) is the process of using the rules in the knowledge base along with the known facts to draw conclusions.
- **Backward** chaining is a goal-driven approach in which you start from an expectation of what is going to happen (i.e., hypothesis) and then seek evidence that supports (or contradicts) your expectation.
- **Forward** chaining is a data-driven approach in which we start from available information as it becomes available or from a basic idea, and then we try to draw conclusions.
- **Certainty factors** express belief in an event (or a fact or a hypothesis) based on the expert's assessment.
- An **interpretation system** infers situation descriptions from observations, and explains observed data by assigning them symbolic meanings that describe the situation.
- An **expert system** shell is a type of development tool that has built-in inference capabilities and a user interface, and is specifically designed for ES development.
- In the popular Corvid ES shell, **variables** define the major factors considered in problem solving.
- In the Corvid ES shell, **logic** blocks are the decision rules acquired from experts.
- **Command** blocks in the Corvid ES determine(s) how the system interacts with the user, including the order of execution and the user interface.
- After an ES system is built, it must be evaluated in a two-step process. The first step, **verification**, ensures that the resulting knowledge base contains knowledge exactly the same as that acquired from the expert.
- After an ES system is built, it must be evaluated in a two-step process. The second step, **validation**, ensures that the system can solve the problem correctly.

SHORT ANSWER

A relatively new approach to supporting decision making is called automated decision systems (ADS), sometimes also known as decision automation systems (DAS). Give a definition of an ADS/DAS in simple terms?

- **In simple terms, An ADS is a rule-based system that provides a solution, usually in one functional area, to a specific repetitive managerial problem, usually in one industry.**

What are the various components of an airline revenue management system? Describe the function of each one.

- **The pricing and accounting system:**
    - **This handles ticket data, published fares, and pricing rules.**
- **The aircraft scheduling system:**
    - **This handles flight schedules based on customer demand.**
- **The inventory management system:**
    - **This handles bookings, cancellations, and changes in departure data.**

Describe, with examples, the two basic ideas most experts agree that artificial intelligence (AI) is concerned with.

- **The study of human thought processes (to understand what intelligence is)**
- **The representation and duplication of those thought processes in machines (e.g., computers, robots)**

List five disciplines of artificial intelligence.

- **Philosophy**
- **Human Behavior**
- **Neurology**
- **Logic**
- **Sociology**
- **Psychology**

- **Human Cognition**
- **Linguistics**
- **Biology**
- **Pattern Recognition**
- **Statistics**
- **Information Systems**
- **Robotics**
- **Management Science**
- **Engineering**
- **Computer Science**
- **Mathematics**

List five applications of artificial intelligence.

- **Expert Systems**
- **Game Playing**
- **Computer Vision**
- **Automatic Programming**
- **Speech Understanding**
- **Autonomous Robots**
- **Intelligent Tutoring**
- **Intelligent Agents**
- **Natural Language Processing**
- **Voice Recognition**
- **Neural Networks**
- **Genetic Algorithms**
- **Fuzzy Logic**
- **Machine Learning**

Describe the Turing test for determining whether a computer exhibits intelligent behavior.

- **According to this test, a computer can be considered smart only when a human interviewer cannot identify the computer while conversing with both an unseen human being and an unseen computer.**

What are three components that may be included in an expert system in addition to the three major components found in virtually all expert systems?

- **Knowledge acquisition subsystem**
- **Blackboard (workplace)**
- **Explanation subsystem (justifier)**
- **Knowledge-refining system**

What is knowledge engineering?

- **Knowledge engineering is the collection of intensive activities encompassing the acquisition of knowledge from human experts (and other information sources) and conversion of this knowledge into a repository (commonly called a knowledge base).**

Name and describe three problem areas suitable for expert systems.

- **Interpretation:**

  - **Inferring situation descriptions from observations.**
- **Prediction:**
  - **Inferring likely consequences of given situations.**
- **Diagnosis:**
  - **Inferring system malfunctions from observations.**
- **Design:**
  - **Configuring objects under constraints.**
- **Planning:**
  - **Developing plans to achieve goals.**
- **Monitoring:**
  - **Comparing observations to plans and flagging exceptions.**
- **Debugging:**
  - **Prescribing remedies for malfunctions.**
- **Repair:**
  - **Executing a plan to administer a prescribed remedy.**
- **Instruction:**
  - **Diagnosing, debugging, and correcting student performance.**
- **Control:**
  - **Interpreting, predicting, repairing, and monitoring system behaviors.**

The development of expert systems is often described as a tedious process. What activities does it typically include?

- **Identifying proper experts**
- **Acquiring knowledge**
- **Selecting the building tools**
- **Coding the system**
- **Evaluating the system**

CHAPTER 12

TRUE OR FALSE

In the Army expertise transfer system case, knowledge nuggets from interviewees needed no further vetting before use in the ETS.	**False**
In the Army expertise transfer system case, text mining was used to extract transcribed textual knowledge extracted from interviewed experts.	**True**
Once you have a lot of compiled user generated information collected through the web, it is immediately available for use in a knowledge management system.	**False**
There has not been widespread success in the deployment of knowledge management systems, with many failures reported.	**True**
Tacit knowledge is described as formal, structured knowledge that is well documented.	**False**
The key benefit of knowledge management systems is not having to reinvent the wheel for repetitive problems.	**True**
The key to effective knowledge management is extracting and encoding expert knowledge that can be accessed and reused by others.	**True**
An example of declarative knowledge is understanding why a medical therapy works.	**False**
Tacit knowledge could be structured, just never codified or documented.	**True**
Knowledge management systems are little help to companies when experts leave and take their knowledge with them.	**False**
The process approach to knowledge management may limit innovation and force participants into fixed patterns of thinking.	**True**
The practice approach to knowledge management focuses on formal controls, procedures, and standard operating procedures.	**False**
Once knowledge is captured and entered into a knowledge repository, it must be reevaluated in the future.	**True**
Hybrid knowledge management approaches that combine the process and practice approaches are unworkable for most organizations that must choose between one and the other.	**False**
In the knowledge management system development process, storing the knowledge in the knowledge repository precedes refining the knowledge.	**False**
In modern organizations, the information technology used to support knowledge management is less important than the application of a certain methodology to business practices.	**True**
In group decision making, a virtual team is one that meets in one geographical location using information technology.	**False**
Wikis allow user creation of shared web content in an organizational format and structure determined by the wiki owner.	**False**
Internet- and intranet-based group decision support systems (GDSS) are less popular than special-purpose decision rooms.	**False**
In the digital forensic knowledge repository case study, the knowledge used to populate the National Repository of Digital Forensics Information (NRDFI) system came mostly from the law enforcement user community.	**False**

MULTIPLE CHOICE

1. In the Army expertise transfer system case, what did 'knowledge harvesting' involve?
   A) interviewing experts and transcribing and mining their knowledge
   B) observing experts at work and documenting their expertise in different tasks
   C) interviewing experts, posting their knowledge verbatim and seeking continual feedback
   D) allowing experts to enter their knowledge directly into the knowledge system

2. Knowledge can be best described as
   A) facts, measurements, and statistics.
   B) facts, measurements, and statistics that are validated.
   C) organized facts set in context and actionable.
   D) an organized collection or set of facts.

3. Which of the following is NOT an attribute of knowledge?
   A) Knowledge fragments as it grows.
   B) The value of knowledge is easily quantified.
   C) Knowledge is not subject to diminishing returns.
   D) There is a need to refresh knowledge periodically for competitive advantage.

4. Which is an example of conditional knowledge?
   A) what new part is appropriate to fix a faulty car

B) when to install a new part in a faulty car
C) why a new part fixes a faulty car
D) how to install a new part in a faulty car

5. Which is an example of declarative knowledge?
A) what new part is appropriate to fix a faulty car
B) when to install a new part in a faulty car
C) how to install a new part in a faulty car
D) why a new part fixes a faulty car

6. The term "leaky knowledge" MOST accurately refers to
A) tacit knowledge.
B) individual knowledge.
C) explicit knowledge.
D) social knowledge.

7. Which approach to knowledge management capitalizes on tacit knowledge and requires heavy IT investment?
A) the practice approach
B) the IT approach
C) the process approach
D) the systems approach

8. During which stage of the KMS cycle are human insights captured together with explicit facts?
A) the storing knowledge stage
B) the refining knowledge stage
C) the disseminating knowledge stage
D) the managing knowledge stage

9. During which stage of the KMS cycle is knowledge made available to intended users?
A) the managing knowledge stage
B) the disseminating knowledge stage
C) the storing knowledge stage
D) the refining knowledge stage

10. Which technology element of knowledge management systems enables document sharing among decision makers?
A) communication
B) collaboration
C) storage
D) retrieval

11. Which set of technologies that support knowledge management is characterized by mashups, blogs, social networks, and wikis?
A) Web 2.0
B) artificial intelligence
C) XML
D) ERP

12. Distractions, groupthink, and digressions that occur in groupwork are known as
A) group failure.
B) process losses.
C) malfunctions.
D) misunderstandings.

13. All the following are negative aspects of working in groups EXCEPT
A) participants may be afraid to contribute.
B) it is time-consuming.
C) it is more difficult to catch errors.
D) there can be a lack of coordination.

14. Videoconferencing is an example of what kind of groupware?
    A) different time, same place
    B) different time, different place
    C) same time, different place
    D) same time, same place
15. Which of the following is the best example of asynchronous communication?
    A) multimedia presentation system
    B) e-mail
    C) videoconference
    D) teleconference
16. Which groupware tools are associated with synchronous use?
    A) VoIP and instant messaging
    B) e-mail and web conferencing
    C) VoIP and wikilogs
    D) Web conference and online workspaces
17. All of the following statements about wikis are true EXCEPT
    A) the majority of Wiki content is user-created.
    B) Wiki pages are usually highly interconnected though links.
    C) Wikis run on a server.
    D) the wiki owner is responsible for the wiki organization.
18. What is the major difference between group support systems (GSS) and group decision support systems (GDSS)?
    A) GSS use modern technology; GDSS do not.
    B) GDSS do not support virtual teams, but GSS do.
    C) GSS use Web based technologies; GDSS do not.
    D) GDSS have a narrower focus than GSS.
19. What software or facility is best for a small firm spread out across the Northeast U.S. looking to inexpensively conduct regular videoconferencing?
    A) WebEx
    B) SharePoint
    C) Excel
    D) decision rooms
20. In the digital forensics case study, what was the main obstacle to knowledge sharing?
    A) Many departments and agencies were effectively functional silos.
    B) There was a need to keep each case experience "clean" and untainted by others.
    C) Laws existed to prevent agencies sharing techniques and experience.
    D) There was an extreme lack of Internet security.

FILL-IN-THE-BLANK

- **Expertise transfer system (ETS)** is a knowledge transfer system developed by the Spears School of Business at Oklahoma State University designed to capture the knowledge of experienced ammunition personnel leaving the Army.
- Nonaka (1991) used the term **tacit** knowledge for the knowledge that exists in the head but not on paper.
- Knowledge is dynamic; thus, an organization must continually refresh its knowledge base to maintain it as a source of competitive **advantage**.
- **Procedural** knowledge is defined as "know-how." An example may be a step-by-step description of how to administer a particular drug.
- Historically, management information systems (MIS) departments have focused on capturing, storing, managing, and reporting **explicit** knowledge.
- One main criticism of the **process** approach is that it fails to capture much of the tacit knowledge embedded in firms.

- A **knowledge** repository, also referred to as an organizational knowledge base, is neither a database nor a knowledge base in the strictest sense of the terms, but rather stores knowledge that is often text based and has very different characteristics.
- Knowledge **management** systems are developed using three sets of technologies: communication, collaboration, and storage and retrieval.
- Groupwork may have both potential benefits and potential drawbacks. The benefits of working in groups are called **process gains**.
- Intra-organizational networked decision support can be effectively supported by an **intranet**, which enables people within an organization to work with Internet tools and procedures through enterprise information portals.
- Many computerized tools have been developed to provide group support. These tools are called **groupware** because their primary objective is to support groupwork.
- If a groupware tool has **synchronous** features, it means that communication and collaboration using that tool are done in real time.
- **Collaborative workflow** refers to software products that address project-oriented and collaborative types of processes.
- The term **Web 2.0** refers to what is perceived to be the second generation of Web development and Web design, being characterized as facilitating communication, information sharing, interoperability, user-centered design, and collaboration on the World Wide Web.
- Server software available at a Web site that allows users to freely create and edit Web page content through a Web browser is known as a **wiki**.
- For Web 2.0 collaborative Web sites, RSS technology is used to rapidly notify users of content **changes**.
- A group support system (GSS) is any combination of hardware and **software** that enhances groupwork.
- In a GSS, **parallelism** is the ability of participants in a group to anonymously work simultaneously on a task, such as brainstorming or voting.
- The earliest GDSS were installed in expensive, customized, special- purpose facilities called **decision/electronic meeting** rooms, with PCs and large public screens at the front of each room.
- **Web**-based groupware is the norm for anytime/anyplace collaboration.

SHORT ANSWER

In the Army expertise transfer system case, we learn that the objective of lessons-learned systems is to support the capture, codification, presentation, and application of expertise in organizations. What are the two main reasons why lesson-learned systems have been a failure?

- **Inadequate representation**
- **Lack of integration into an organization's decision-making process**

Define knowledge management (KM), and briefly explain the process through which it is implemented within an organization.

- **Knowledge management is the systematic and active management of ideas, information, and knowledge residing in an organization's employees.**
- **Knowledge management is a process that helps organizations identify, select, organize, disseminate, and transfer important information and expertise that are part of the organization's memory and that typically reside within the organization in an unstructured manner.**

What are the four characteristics of knowledge that make it unlike other organizational assets?

- **Extraordinary leverage and increasing returns.**
  - **Knowledge is not subject to diminishing returns.**
  - **When it is used, it is not decreased (or depleted); rather, it is increased (or improved).**
  - **Its consumers can add to it, thus increasing its value.**
- **Fragmentation, leakage, and the need to refresh.**
  - **As knowledge grows, it branches and fragments.**
  - **Knowledge is dynamic; it is information in action.**
  - **Thus, an organization must continually refresh its knowledge base to maintain it as a source of competitive advantage.**
- **Uncertain value.**

  - **It is difficult to estimate the impact of an investment in knowledge.**
  - **There are too many intangible aspects that cannot be easily quantified.**
- **Value of sharing.**
  - **It is difficult to estimate the value of sharing one's knowledge or even who will benefit most from it.**

One critical goal of knowledge management is to retain the valuable know-how that can so easily and quickly leave an organization. In this context, explain with examples what knowledge management systems refer to within an organization.

- **Knowledge management systems (KMS) refer to the use of modern IT (e.g., the Internet, intranets, extranets, Lotus Notes, software filters, agents, data warehouses, Web 2.0) to systematize, enhance, and expedite intra- and inter-firm KM.**

The two fundamental approaches to knowledge management are the process approach and the practice approach. Explain the differences between them.

- **The process approach to knowledge management attempts to codify organizational knowledge through formalized controls, processes, and technologies, while the practice approach focuses on building the social environments or communities of practice necessary to facilitate the sharing of tacit understanding.**

A functioning knowledge management system (KMS) follows six steps in a cycle. The reason for the cycle is that knowledge is dynamically refined over time. What are the six steps in the KMS cycle?

- **Create knowledge.**
  - **Knowledge is created as people determine new ways of doing things or develop know-how.**
  - **Sometimes external knowledge is brought in.**
  - **Some of these new ways may become best practices.**
- **Capture knowledge.**
  - **New knowledge must be identified as valuable and be represented in a reasonable way.**
- **Refine knowledge.**
  - **New knowledge must be placed in context so that it is actionable.**
  - **This is where human insights (i.e., tacit qualities) must be captured along with explicit facts.**
- **Store knowledge.**
  - **Useful knowledge must be stored in a reasonable format in a knowledge repository so that others in the organization can access it.**
- **Manage knowledge.**
  - **Like a library, a repository must be kept current.**
  - **It must be reviewed to verify that it is relevant and accurate.**
- **Disseminate knowledge.**
  - **Knowledge must be made available in a useful format to anyone in the organization who needs it, anywhere and anytime.**

Briefly describe three benefits (process gains) derived from working in groups.

- **It provides learning.**
  - **Groups are better than individuals at understanding problems.**
- **People readily take ownership of problems and their solutions.**
  - **They take responsibility.**
- **Group members have their egos embedded in the decision, so they are committed to the solution.**
- **Groups are better than individuals at catching errors.**
- **A group has more information (i.e., knowledge) than any one member.**
  - **Group members can combine their knowledge to create new knowledge.**
  - **More and more creative alternatives for problem solving can be generated, and better solutions can be derived (e.g., through stimulation).**

- **A group may produce synergy during problem solving.**
    - **The effectiveness and/or quality of group work can be greater than the sum of what is produced by independent individuals.**
- **Working in a group may stimulate the creativity of the participants and the process.**
- **A group may have better and more precise communication working together.**
- **Risk propensity is balanced.**
    - **Groups moderate high-risk takers and encourage conservatives.**

Groupware products provide a way for groups to share resources and opinions. Groupware implies the use of networks to connect people, even if they are in the same room. What are three general features of groupware products that support commutation, collaboration, and coordination?

- **Built-in e-mail, messaging system**
- **Browser interface**
- **Joint Web-page creation**
- **Sharing of active hyperlinks**
- **File sharing (graphics, video, audio, or other)**
- **Built-in search functions (by topic or keyword)**
- **Workflow tools**
- **Use of corporate portals for communication, collaboration, and search**
- **Shared screens**
- **Electronic decision rooms**
- **Peer-to-peer networks**

A group decision support system (GDSS) is an interactive computer-based system that facilitates the solution of semi structured or unstructured problems by a group of decision makers. What are the major characteristics of a GDSS?

- **Its goal is to support the process of group decision makers by providing automation of sub processes, using information technology tools.**
- **It is a specially designed information system, not merely a configuration of already existing system components.**
    - **It can be designed to address one type of problem or a variety of group-level organizational decisions.**
- **It encourages generation of ideas, resolution of conflicts, and freedom of expression.**
    - **It contains built-in mechanisms that discourage development of negative group behaviors, such as destructive conflict, miscommunication, and groupthink.**

A group support system (GSS) is any combination of hardware and software that enhances group work either in direct or indirect support of decision making. When and how did GSS evolve?

- **GSS evolved after information technology researchers recognized that technology could be developed to support the many activities normally occurring at face-to-face meetings (e.g., idea generation, consensus building, and anonymous ranking).**

CHAPTER 13

TRUE OR FALSE

In the opening vignette, the CERN Data Aggregation System (DAS), built on MongoDB (a Big Data management infrastructure), used relational database technology.	**False**
The term "Big Data" is relative as it depends on the size of the using organization.	**True**
In the Luxottica case study, outsourcing enhanced the ability of the company to gain insights into their data.	**False**
Many analytics tools are too complex for the average user, and this is one justification for Big Data.	**True**
In the investment bank case study, the major benefit brought about by the supplanting of multiple databases by the new trade operational store was providing real-time access to trading data.	**True**
Big Data uses commodity hardware, which is expensive, specialized hardware that is custom built for a client or application.	**False**
MapReduce can be easily understood by skilled programmers due to its procedural nature.	**True**
Hadoop was designed to handle petabytes and extabytes of data distributed over multiple nodes in parallel.	**True**
Hadoop and MapReduce require each other to work.	**False**
In most cases, Hadoop is used to replace data warehouses.	**False**
Despite their potential, many current NoSQL tools lack mature management and monitoring tools.	**True**
The data scientist is a profession for a field that is still largely being defined.	**True**
There is a current undersupply of data scientists for the Big Data market.	**True**
The Big Data and Analysis in Politics case study makes it clear that the unpredictability of elections makes politics an unsuitable arena for Big Data.	**False**
For low latency, interactive reports, a data warehouse is preferable to Hadoop.	**True**
If you have many flexible programming languages running in parallel, Hadoop is preferable to a data warehouse.	**True**
In the Dublin City Council case study, GPS data from the city's buses and CCTV were the only data sources for the Big Data GIS-based application.	**False**
It is important for Big Data and self-service business intelligence go hand in hand to get maximum value from analytics.	**True**
Big Data simplifies data governance issues, especially for global firms.	**False**
Current total storage capacity lags behind the digital information being generated in the world.	**True**

MULTIPLE CHOICE

1. Using data to understand customers/clients and business operations to sustain and foster growth and profitability is
   - **A)** now completely automated with no human intervention required.
   - **B)** essentially the same now as it has always been.
   - **C) an increasingly challenging task for today's enterprises.**
   - **D)** easier with the advent of BI and Big Data.
2. A newly popular unit of data in the Big Data era is the petabyte (PB), which is
   - **A)** 1012 bytes.
   - **B)** 1018 bytes.
   - **C) 1015 bytes.**
   - **D)** 109 bytes.
3. Which of the following sources is likely to produce Big Data the fastest?
   - **A)** order entry clerks
   - **B) RFID tags**
   - **C)** online customers
   - **D)** cashiers
4. Data flows can be highly inconsistent, with periodic peaks, making data loads hard to manage. What is this feature of Big Data called?
   - **A) variability**
   - **B)** periodicity
   - **C)** inconsistency
   - **D)** volatility
5. In the Luxottica case study, what technique did the company use to gain visibility into its customers?

A) focus on growth
B) customer focus
C) visibility analytics
D) data integration

6. Allowing Big Data to be processed in memory and distributed across a dedicated set of nodes can solve complex problems in near—real time with highly accurate insights. What is this process called?

A) in-database analytics
B) appliances
C) in-memory analytics
D) grid computing

7. Which Big Data approach promotes efficiency, lower cost, and better performance by processing jobs in a shared, centrally managed pool of IT resources?

A) in-memory analytics
B) in-database analytics
C) appliances
D) grid computing

8. How does Hadoop work?

A) It integrates Big Data into a whole so large data elements can be processed as a whole on one computer.
B) It integrates Big Data into a whole so large data elements can be processed as a whole on multiple computers.
C) It breaks up Big Data into multiple parts so each part can be processed and analyzed at the same time on multiple computers.
D) It breaks up Big Data into multiple parts so each part can be processed and analyzed at the same time on one computer.

9. What is the Hadoop Distributed File System (HDFS) designed to handle?

A) unstructured and semistructured relational data
B) structured and semistructured relational data
C) structured and semistructured non-relational data
D) unstructured and semistructured non-relational data

10. In a Hadoop "stack," what is a slave node?

A) a node where data is stored and processed
B) a node where bits of programs are stored
C) a node responsible for holding all the source programs
D) a node where metadata is stored and used to organize data processing

11. In a Hadoop "stack," what node periodically replicates and stores data from the Name Node should it fail?

A) backup node
B) slave node
C) secondary node
D) substitute node

12. All of the following statements about MapReduce are true EXCEPT

A) MapReduce is a general-purpose execution engine.
B) MapReduce handles parallel programming.
C) MapReduce handles the complexities of network communication.
D) MapReduce runs without fault tolerance.

13. In the Big Data and Analytics in Politics case study, which of the following was an input to the analytic system?

A) group clustering
B) assessment of sentiment
C) census data
D) voter mobilization

14. In the Big Data and Analytics in Politics case study, what was the analytic system output or goal?

A) group clustering

- **B)** assessment of sentiment
- **C)** census data
- **D) voter mobilization**

15. Traditional data warehouses have not been able to keep up with
    - **A)** expert systems that run on them.
    - **B)** the evolution of the SQL language.
    - **C)** OLAP.
    - **D) the variety and complexity of data.**
16. Under which of the following requirements would it be more appropriate to use Hadoop over a data warehouse?
    - **A)** ANSI 2003 SQL compliance is required
    - **B) unrestricted, ungoverned sandbox explorations**
    - **C)** analysis of provisional data
    - **D)** online archives alternative to tape
17. What is Big Data's relationship to the cloud?
    - **A)** Hadoop cannot be deployed effectively in the cloud just yet.
    - **B)** IBM's homegrown Hadoop platform is the only option.
    - **C) Amazon and Google have working Hadoop cloud offerings.**
    - **D)** Only MapReduce works in the cloud; Hadoop does not.
18. Companies with the largest revenues from Big Data tend to be
    - **A) the largest computer and IT services firms.**
    - **B)** non-U.S. Big Data firms.
    - **C)** pure open source Big Data firms.
    - **D)** small computer and IT services firms.
19. In the health sciences, the largest potential source of Big Data comes from
    - **A)** research administration.
    - **B)** human resources.
    - **C)** accounting systems.
    - **D) patient monitoring.**
20. In the Discovery Health insurance case study, the analytics application used available data to help the company do all of the following EXCEPT
    - **A)** detect fraud.
    - **B)** predict customer health.
    - **C) open its own pharmacy.**
    - **D)** lower costs for members.

FILL-IN-THE-BLANK

- Most Big Data is generated automatically by **machines**.
- **Veracity** refers to the conformity to facts: accuracy, quality, truthfulness, or trustworthiness of the data.
- In-motion **analytics** is often overlooked today in the world of BI and Big Data.
- The **value proposition** of Big Data is its potential to contain more useful patterns and interesting anomalies than "small" data.
- As the size and the complexity of analytical systems increase, the need for more **efficient** analytical systems is also increasing to obtain the best performance.
- **In-database analytics** speeds time to insights and enables better data governance by performing data integration and analytic functions inside the database.
- **Appliances** bring together hardware and software in a physical unit that is not only fast but also scalable on an as-needed basis.
- Big Data employs **parallel** processing techniques and nonrelational data storage capabilities in order to process unstructured and semistructured data.
- In the world of Big Data, **MapReduce** aids organizations in processing and analyzing large volumes of multi-structured data. Examples include indexing and search, graph analysis, etc.
- The **Name** Node in a Hadoop cluster provides client information on where in the cluster particular data is stored and if any nodes fail.
- A job **tracker** is a node in a Hadoop cluster that initiates and coordinates MapReduce jobs, or the processing of the data.
- HBase is a nonrelational **database** that allows for low-latency, quick lookups in Hadoop.
- Hadoop is primarily a(n) **distributed** file system and lacks capabilities we'd associate with a DBMS, such as indexing, random access to data, and support for SQL.
- HBase, Cassandra, MongoDB, and Accumulo are examples of **NoSQL** databases.
- In the eBay use case study, load **balancing** helped the company meet its Big Data needs with the extremely fast data handling and application availability requirements.
- As volumes of Big Data arrive from multiple sources such as sensors, machines, social media, and clickstream interactions, the first step is to **capture** all the data reliably and cost effectively.
- In open-source databases, the most important performance enhancement to date is the cost-based **optimizer**.
- Data **integration** or pulling of data from multiple subject areas and numerous applications into one repository is the raison d'être for data warehouses.
- In the energy industry, **smart** grids are one of the most impactful applications of stream analytics.
- In the U.S. telecommunications company case study, the use of analytics via dashboards has helped to improve the effectiveness of the company's **threat** assessments and to make their systems more secure.

SHORT ANSWER

In the opening vignette, what is the source of the Big Data collected at the European Organization for Nuclear Research or CERN?

- **Forty million times per second, particles collide within the LHC, each collision generating particles that often decay in complex ways into even more particles.**
- **Precise electronic circuits all around LHC record the passage of each particle via a detector as a series of electronic signals, and send the data to the CERN Data Centre (DC) for recording and digital reconstruction.**
- **The digitized summary of data is recorded as a “collision event”.**
- **15 petabytes or so of digitized summary data produced annually and this is processed by physicists to determine if the collisions have thrown up any interesting physics.**

List and describe the three main "V"s that characterize Big Data.

- **Volume:**
  - **This is obviously the most common trait of Big Data.**
  - **Many factors contributed to the exponential increase in data volume, such as transaction-based data stored through the years, text data constantly streaming in from social media, increasing amounts of sensor data being collected, automatically generated RFID and GPS data, and so forth.**
- **Variety:**
  - **Data today comes in all types of formats–ranging from traditional databases to hierarchical data stores created by the end users and OLAP systems, to text documents, e-mail, XML, meter-collected, sensor-captured data, to video, audio, and stock ticker data.**
  - **By some estimates, 80 to 85 percent of all organizations' data is in some sort of unstructured or semistructured format**
- **Velocity:**
  - **This refers to both how fast data is being produced and how fast the data must be processed (i.e., captured, stored, and analyzed) to meet the need or demand.**
  - **RFID tags, automated sensors, GPS devices, and smart meters are driving an increasing need to deal with torrents of data in near—real time.**

List and describe four of the most critical success factors for Big Data analytics.

- **A clear business need (alignment with the vision and the strategy).**
  - **Business investments ought to be made for the good of the business, not for the sake of mere technology advancements.**
  - **Therefore the main driver for Big Data analytics should be the needs of the business at any level–strategic, tactical, and operations.**
- **Strong, committed sponsorship (executive champion).**
  - **It is a well-known fact that if you don't have strong, committed executive sponsorship, it is difficult (if not impossible) to succeed.**
  - **If the scope is a single or a few analytical applications, the sponsorship can be at the departmental level.**
  - **However, if the target is enterprise-wide organizational transformation, which is often the case for Big Data initiatives, sponsorship needs to be at the highest levels and organization-wide.**
- **Alignment between the business and IT strategy.**
  - **It is essential to make sure that the analytics work is always supporting the business strategy, and not other way around.**
  - **Analytics should play the enabling role in successful execution of the business strategy.**
- **A fact-based decision making culture.**
  - **In a fact-based decision-making culture, the numbers rather than intuition, gut feeling, or supposition drive decision making.**

  - There is also a culture of experimentation to see what works and doesn't.
  - To create a fact-based decision-making culture, senior management needs to do the following: recognize that some people can't or won't adjust; be a vocal supporter; stress that outdated methods must be discontinued; ask to see what analytics went into decisions; link incentives and compensation to desired behaviors.
- **A strong data infrastructure.**
  - Data warehouses have provided the data infrastructure for analytics.
  - This infrastructure is changing and being enhanced in the Big Data era with new technologies.
  - Success requires marrying the old with the new for a holistic infrastructure that works synergistically.

When considering Big Data projects and architecture, list and describe five challenges designers should be mindful of in order to make the journey to analytics competency less stressful.

- **Data volume:**
  - The ability to capture, store, and process the huge volume of data at an acceptable speed so that the latest information is available to decision makers when they need it.
- **Data integration:**
  - The ability to combine data that is not similar in structure or source and to do so quickly and at reasonable cost.
- **Processing capabilities:**
  - The ability to process the data quickly, as it is captured.
  - The traditional way of collecting and then processing the data may not work.
  - In many situations data needs to be analyzed as soon as it is captured to leverage the most value.
- **Data governance:**
  - The ability to keep up with the security, privacy, ownership, and quality issues of Big Data.
  - As the volume, variety (format and source), and velocity of data change, so should the capabilities of governance practices.
- **Skills availability:**
  - Big Data is being harnessed with new tools and is being looked at in different ways.
  - There is a shortage of data scientists with the skills to do the job.
- **Solution cost:**
  - Since Big Data has opened up a world of possible business improvements, there is a great deal of experimentation and discovery taking place to determine the patterns that matter and the insights that turn to value.
  - To ensure a positive ROI on a Big Data project, therefore, it is crucial to reduce the cost of the solutions used to find that value.

Define MapReduce.

- **As described by Dean and Ghemawat (2004), "MapReduce is a programming model and an associated implementation for processing and generating large data sets. Programs written in this functional style are automatically parallelized and executed on a large cluster of commodity machines. This allows programmers without any experience with parallel and distributed systems to easily utilize the resources of a large distributed system."**

What is NoSQL as used for Big Data? Describe its major downsides.

- **NoSQL is a new style of database that has emerged to, like Hadoop, process large volumes of multi-structured data.**
  - However, whereas Hadoop is adept at supporting large-scale, batch-style historical analysis, NoSQL databases are aimed, for the most part (though there are some important exceptions), at serving up discrete data stored among large volumes of multi-structured data to end-user and automated Big Data applications.

  - **This capability is sorely lacking from relational database technology, which simply can't maintain needed application performance levels at Big Data scale.**
- **The downside of most NoSQL databases today is that they trade ACID (atomicity, consistency, isolation, durability) compliance for performance and scalability.**
  - **Many also lack mature management and monitoring tools.**

What is a data scientist and what does the job involve?

- **A data scientist is a role or a job frequently associated with Big Data or data science.**
  - **In a very short time it has become one of the most sought-out roles in the marketplace.**
  - **Currently, data scientists' most basic, current skill is the ability to write code (in the latest Big Data languages and platforms).**
  - **A more enduring skill will be the need for data scientists to communicate in a language that all their stakeholders understand–and to demonstrate the special skills involved in storytelling with data, whether verbally, visually, or–ideally–both.**
  - **Data scientists use a combination of their business and technical skills to investigate Big Data looking for ways to improve current business analytics practices (from descriptive to predictive and prescriptive) and hence to improve decisions for new business opportunities.**

Why are some portions of tape backup workloads being redirected to Hadoop clusters today?

- **First, while it may appear inexpensive to store data on tape, the true cost comes with the difficulty of retrieval.**
  - **Not only is the data stored offline, requiring hours if not days to restore, but tape cartridges themselves are also prone to degradation over time, making data loss a reality and forcing companies to factor in those costs.**
  - **To make matters worse, tape formats change every couple of years, requiring organizations to either perform massive data migrations to the newest tape format or risk the inability to restore data from obsolete tapes.**
- **Second, it has been shown that there is value in keeping historical data online and accessible.**
  - **As in the clickstream example, keeping raw data on a spinning disk for a longer duration makes it easy for companies to revisit data when the context changes and new constraints need to be applied.**
  - **Searching thousands of disks with Hadoop is dramatically faster and easier than spinning through hundreds of magnetic tapes.**
  - **Additionally, as disk densities continue to double every 18 months, it becomes economically feasible for organizations to hold many years' worth of raw or refined data in HDFS.**

What are the differences between stream analytics and perpetual analytics? When would you use one or the other?

- **In many cases they are used synonymously.**
  - **However, in the context of intelligent systems, there is a difference.**
  - **Streaming analytics involves applying transaction-level logic to real-time observations.**
  - **The rules applied to these observations take into account previous observations as long as they occurred in the prescribed window; these windows have some arbitrary size (e.g., last 5 seconds, last 10,000 observations, etc.).**
  - **Perpetual analytics, on the other hand, evaluates every incoming observation against all prior observations, where there is no window size.**
  - **Recognizing how the new observation relates to all prior observations enables the discovery of real-time insight.**
- **When transactional volumes are high and the time-to-decision is too short, favoring nonpersistence and small window sizes, this translates into using streaming analytics.**

  - **However, when the mission is critical and transaction volumes can be managed in real time, then perpetual analytics is a better answer.**

Describe data stream mining and how it is used.

- **Data stream mining, as an enabling technology for stream analytics, is the process of extracting novel patterns and knowledge structures from continuous, rapid data records.**
  - **A data stream is a continuous flow of ordered sequence of instances that in many applications of data stream mining can be read/processed only once or a small number of times using limited computing and storage capabilities.**
  - **Examples of data streams include sensor data, computer network traffic, phone conversations, ATM transactions, web searches, and financial data.**
  - **Data stream mining can be considered a subfield of data mining, machine learning, and knowledge discovery.**
  - **In many data stream mining applications, the goal is to predict the class or value of new instances in the data stream given some knowledge about the class membership or values of previous instances in the data stream.**

CHAPTER 14

TRUE OR FALSE

Oklahoma Gas & Electric employs a two-layer information architecture involving data warehouse and improved and expanded integration.	**False**
In the classification of location-based analytic applications, examining geographic site locations falls in the consumer-oriented category.	**False**
In the Great Clips case study, the company uses geospatial data to analyze, among other things, the types of haircuts most popular in different geographic locations.	**False**
From massive amounts of high-dimensional location data, algorithms that reduce the dimensionality of the data can be used to uncover trends, meaning, and relationships to eventually produce human-understandable representations.	**True**
In the life coach case study, Kaggle recently hosted a competition aimed at identifying muscle motions that may be used to predict the progression of Alzheimer's disease.	**True**
Content-based filtering approaches are widely used in recommending textual content such as news items and related Web pages.	**True**
The basic premise behind social networking is that it gives people the power to share, making the world more open and connected.	**True**
Cloud computing originates from a reference to the Internet as a "cloud" and is a combination of several information technology components as services.	**True**
Web-based e-mail such as Google's Gmail are not examples of cloud computing.	**False**
Service-oriented DSS solutions generally offer individual or bundled services to the user as a service.	**True**
Data-as-a-service began with the notion that data quality could happen in a centralized place, cleansing and enriching data and offering it to different systems, applications, or users, irrespective of where they were in the organization, computers, or on the network.	**True**
In service-oriented DSS, an application programming interface (API) serves to populate source systems with raw data and to pull operational reports.	**True**
IaaS helps provide faster information, but provides information only to managers in an organization.	**False**
The trend in the consumption of data analytics is away from in-memory solution and towards mobile devices.	**False**
While cloud services are useful for small and midsize analytic applications, they are still limited in their ability to handle Big Data applications.	**False**
Analytics integration with other organizational systems makes it harder to identify its impact on the organization.	**True**
Use of automated decision systems (ADSs) is likely to result in a reduction of middle management.	**True**
The industry impact of an automated decision system's use is limited to the company's supply chain.	**False**
ES/DSS were found to improve the performance of new managers but not existing managers.	**False**
In designing analytic systems, it must be kept in mind that the right to an individual's privacy is not absolute.	**True**

MULTIPLE CHOICE

1. What kind of location based analytics is real-time marketing promotion?
   - A) consumer-oriented geospatial static approach
   - **B) organization-oriented location-based dynamic approach**
   - C) consumer-oriented location-based dynamic approach
   - D) organization-oriented geospatial static approach
2. GPS Navigation is an example of which kind of location based analytics?
   - **A) consumer-oriented geospatial static approach**
   - B) organization-oriented geospatial static approach
   - C) consumer-oriented location-based dynamic approach
   - D) organization-oriented location-based dynamic approach
3. What new geometric data type in Teradata's data warehouse captures geospatial features?
   - A) NAVTEQ
   - **B) ST_GEOMETRY**
   - C) SQL/MM
   - D) GIS

4. A British company called Path Intelligence has developed a system that ascertains how people move within a city or even within a store. What is this system called?
   - A) Pathdata
   - **B) Footpath**
   - C) Pathfinder
   - D) PathMiner
5. Today, most smartphones are equipped with various instruments to measure jerk, orientation, and sense motion. One of these instruments is an accelerometer, and the other is a(n)
   - A) potentiometer.
   - **B) gyroscope.**
   - C) oscilloscope.
   - D) microscope.
6. Content-based filtering obtains detailed information about item characteristics and restricts this process to a single user using information tags or
   - A) key-pairs.
   - B) passphrases.
   - C) reality mining.
   - **D) keywords.**
7. Service-oriented thinking is one of the fastest growing paradigms in today's economy. Which of the following is NOT a characteristic of service-oriented DSS?
   - A) reusability
   - B) extensibility
   - C) substitutability
   - **D) originality**
8. All of the following are components in a service-oriented DSS environment EXCEPT
   - A) information technology as enabler.
   - B) process as beneficiary.
   - **C) data as infrastructure.**
   - D) people as user.
9. Which of the following is true of data-as-a-Service (DaaS) platforms?
   - A) Knowing where the data resides is critical to the functioning of the platform.
   - B) Business processes can access local data only.
   - C) Data quality happens on each individual platform.
   - **D) There are standardized processes for accessing data wherever it is located.**
10. Which component of service-oriented DSS can be described as a subset of a data warehouse that supports specific decision and analytical needs and provides business units more flexibility, control, and responsibility?
    - A) information services with library and administrator
    - B) extract, transform, load
    - C) information delivery portals
    - **D) data marts**
11. Which component of service-oriented DSS can be described as optimizing the DSS environment use by organizing its capabilities and knowledge, and assimilating them into the business processes?
    - **A) information services with library and administrator**
    - B) extract, transform, load
    - C) data marts
    - D) information delivery portals
12. Which component of service-oriented DSS can be defined as data that describes the meaning and structure of business data, as well as how it is created, accessed, and used?
    - **A) metadata management**

B) operations and administration
C) application programming interface
D) analytics

13. Which component of service-oriented DSS includes such examples as optimization, data mining, text mining, simulation, automated decision systems?
A) analytics
B) operations and administration
C) metadata management
D) application programming interface

14. Which of the following offers a flexible data integration platform based on a newer generation of service-oriented standards that enables ubiquitous access to any type of data?
A) EII
B) IaaS
C) ETL
D) EAI

15. When new analytics applications are introduced and affect multiple related processes and departments, the organization is best served by utilizing
A) multi-department analysis.
B) business flow management.
C) process flow analysis.
D) business process reengineering.

16. Research into managerial use of DSS and expert systems found all the following EXCEPT
A) managers spent more of their time planning.
B) managers spent more time in the office and less in the field.
C) managers were able to devote less of their time fighting fires.
D) managers saw their decision making quality enhanced.

17. Why do analytics applications have the effect of redistributing power among managers?
A) New analytics applications change managers' job expectations.
B) Sponsoring an analytics system automatically confers power to a manager.
C) New analytics systems lead to new budget allocations, resulting in increased power.
D) The more information and analysis managers have, the more power they possess.

18. Services that let consumers permanently enter a profile of information along with a password and use this information repeatedly to access services at multiple sites are called
A) single-sign-on facilities.
B) information collection portals.
C) consumer information sign on facilities.
D) consumer access applications.

19. Which of the following is true about the furtherance of homeland security?
A) The impetus was the need to harvest information related to financial fraud after 2001.
B) There is a greater need for oversight.
C) Most people regard analytic tools as mostly ineffective in increasing security.
D) There is a lessening of privacy issues.

20. Which of the following is considered the economic engine of the whole analytics industry?
A) analytics user organizations
B) analytics industry analysts and influencers
C) application developers and system integrators
D) academic providers and certification industries

FILL-IN-THE-BLANK

- In the opening vignette, the combination of filed infrastructure, geospatial data, enterprise data warehouse, and analytics has enabled OG&E to manage its customer demand in such a way that it can optimize its **long-term** investments.
- A critical emerging trend in analytics is the incorporation of location data. **Geospatial** data is the static location data used by these location-based analytic applications.
- The surge in location-enabled services has resulted in **reality** mining, the analytics of massive databases of historical and real-time streaming location information.
- The Radii mobile app collects information about the user's habits, interests, spending patterns, and favorite locations to understand the user's **personality**.
- Predictive analytics is beginning to enable development of software that is directly used by a consumer. One key concern in employing these technologies is the loss of **privacy**.
- Collaborative filtering is usually done by building a user-item ratings matrix where each row represents a unique user and each column gives the individual item rating made by the user. The resultant matrix is a dynamic, sparse matrix with a huge **dimensionality**.
- **Ajax**, which stands for Asynchronous JavaScript and XML, is an effective and efficient Web development technique for creating interactive Web applications.
- **Information-as-a-service** (IaaS) promises to eliminate independent silos of data that exist in systems and infrastructure and enable sharing real-time information for emerging apps, to hide complexity, and to increase availability with virtualization.
- IaaS, AaaS and other **cloud**-based offerings allow the rapid diffusion of advanced analysis tools among users, without significant investment in technology acquisition.
- A major structural change that can occur when analytics are introduced into an organization is the creation of new organizational **units**.
- When an organization-wide, major restructuring is needed, the process is referred to as **reengineering**.
- ADSs can lead in many cases to improved customer **service** (e.g., responding faster to queries).
- A research study found that employees using ADS systems were more **satisfied** with their jobs.
- Analytics can change the way in which many **decisions** are made by managers and can consequently change their jobs.
- As face-to-face communication is often replaced by e-mail, wikis, and computerized conferencing, leadership qualities attributed to physical **appearance** could become less important.
- Location information from **mobile/cell** phones can be used to create profiles of user behavior and movement.
- For individual decision makers, **personal** values constitute a major factor in the issue of ethical decision making.
- Firms such as Nielsen provide **specialized** data collection, aggregation, and distribution mechanisms and typically focus on one industry sector.
- Possibly the biggest recent growth in analytics has been in **predictive** analytics, as many statistical software companies such as SAS and SPSS embraced it early on.
- Southern States Cooperative used analytics to prepare the customized catalogs to suit the targeted **customer** needs, resulting in better revenue generation.

SHORT ANSWER

How does Oklahoma Gas and Electric use the Teradata platform to manage the electric grid?

- **Oklahoma Gas and Electric uses the Teradata platform to organize the large amounts of data that it gathers from installation of smart meters and other devices on the electronic grid at the consumer end? With Teradata's platform, OG&E has combined its smart meter data, outage data, call center data, rate data, asset data, price signals, billing, and collections into one integrated data platform. The platform also incorporates geospatial mapping of the integrated data using the in-database geospatial analytics that add onto the OG&E's dynamic segmentation capabilities.**

How do the traditional location-based analytic techniques using geocoding of organizational locations and consumers hamper the organizations in understanding "true location-based" impacts?

- **Locations based on postal codes offer an aggregate view of a large geographic area.**
- **This poor granularity may not be able to pinpoint the growth opportunities within a region.**
- **The location of the target customers can change rapidly. An organization's promotional campaigns might not target the right customers.**

In what ways can communications companies use geospatial analysis to harness their data effectively?

- **Communication companies often generate massive amounts of data every day.**
- **The ability to analyze the data quickly with a high level of location-specific granularity can better identify the customer churn and help in formulating strategies specific to locations for increasing operational efficiency, quality of service, and revenue.**

Describe the CabSense application used by the New York City Taxi and Limousine Commission.

- **Sense Networks has built a mobile application called CabSense that analyzes large amounts of data from the New York City Taxi and Limousine Commission.**
- **CabSense helps New Yorkers and visitors in finding the best corners for hailing a taxi based on the person's location, day of the week, and time.**
- **CabSense rates the street corners on a 5-point scale by making use of machine-learning algorithms applied to the vast amounts of historical location points obtained from the pickups and drop-offs of all New York City cabs.**
- **Although the app does not give the exact location of cabs in real time, its data-crunching predictions enable people to get to a street corner that has the highest probability of finding a cab.**

What are recommender systems, how are they developed, and how is the data used to build a recommendation system obtained?

- **The term recommender systems refers to a Web-based information filtering system that takes the inputs from users and then aggregates the inputs to provide recommendations for other users in their product or service selection choices.**
- **Two basic approaches that are employed in the development of recommendation systems are collaborative filtering and content filtering.**
  - **In collaborative filtering, the recommendation system is built based on the individual user's past behavior by keeping track of the previous history of all purchased items.**
    - **This includes products, items that are viewed most often, and ratings that are given by the users to the items they purchased.**
  - **In the content-based filtering approach, the characteristics of an item are profiled first and then content-based individual user profiles are built to store the information about the characteristics of specific items that the user has rated in the past.**

  - **In the recommendation process, a comparison is made by filtering the item information from the user profile for which the user has rated positively and compares these characteristics with any new products that the user has not rated yet.**
  - **Recommendations are made if there are similarities found in the item characteristics.**
- **The data necessary to build a recommendation system are collected by Web-based systems where each user is specifically asked to rate an item on a rating scale, rank the items from most favorite to least favorite, and/or ask the user to list the attributes of the items that the user likes.**

Web 2.0 is the popular term for describing advanced Web technologies and applications. Describe four main representative characteristics of the Web 2.0 environment.

- **Web 2.0 has the ability to tap into the collective intelligence of users. The more users contribute, the more popular and valuable a Web 2.0 site becomes.**
- **Data is made available in new or never-intended ways. Web 2.0 data can be remixed or "mashed up," often through Web service interfaces, much the way a dance-club DJ mixes music.**
- **Web 2.0 relies on user-generated and user-controlled content and data.**
- **Lightweight programming techniques and tools let nearly anyone act as a Web site developer.**
- **The virtual elimination of software-upgrade cycles makes everything a perpetual beta or work-in-progress and allows rapid prototyping, using the Web as an application development platform.**
- **Users can access applications entirely through a browser.**
- **An architecture of participation and digital democracy encourages users to add value to the application as they use it.**
- **A major emphasis is on social networks and computing.**
- **There is strong support for information sharing and collaboration.**
- **Web 2.0 fosters rapid and continuous creation of new business models.**

What is mobile social network and how does it extend the reach of popular social networks?

- **Mobile social networking refers to social networking where members converse and connect with one another using cell phones or other mobile devices.**
- **Virtually all major social networking sites offer mobile services or apps on smartphones to access their services.**
- **The explosion of mobile Web 2.0 services and companies means that many social networks can be based from cell phones and other portable devices, extending the reach of such networks to the millions of people who lack regular or easy access to computers.**

What is cloud computing? What is Amazon's general approach to the cloud computing services it provides?

- **Wikipedia defines cloud computing as "a style of computing in which dynamically scalable and often virtualized resources are provided over the Internet. Users need not have knowledge of, experience in, or control over the technology infrastructures in the cloud that supports them."**
- **Amazon.com has developed an impressive technology infrastructure for e- commerce as well as for business intelligence, customer relationship management, and supply chain management. It has built major data centers to manage its own operations. However, through Amazon.com's cloud services, many other companies can employ these very same facilities to gain advantages of these technologies without having to make a similar investment. Like other cloud-computing services, a user can subscribe to any of the facilities on a pay-as-you-go basis. This model of letting someone else own the hardware and software but making use of the facilities on a pay-per-use basis is the cornerstone of cloud computing.**

Data and text mining is a promising application of AaaS. What additional capabilities can AaaS bring to the analytic world?

- **It can also be used for large-scale optimization, highly-complex multi-criteria decision problems, and distributed simulation models.**

  - **These prescriptive analytics require highly capable systems that can only be realized using service-based collaborative systems that can utilize large-scale computational resources.**

Describe your understanding of the emerging term people analytics. Are there any privacy issues associated with the application?

- **Applications such as using sensor-embedded badges that employees wear to track their movement and predict behavior has resulted in the term people analytics.**
  - **This application area combines organizational IT impact, Big Data, sensors, and has privacy concerns.**
  - **One company, Sociometric Solutions, has reported several such applications of their sensor-embedded badges.**
- **People analytics creates major privacy issues.**
  - **Should the companies be able to monitor their employees this intrusively? Sociometric has reported that its analytics are only reported on an aggregate basis to their clients.**
  - **No individual user data is shared.**
  - **They have noted that some employers want to get individual employee data, but their contract explicitly prohibits this type of sharing.**
  - **In any case, sensors are leading to another level of surveillance and analytics, which poses interesting privacy, legal, and ethical questions.**