Chapter 6

Foundations of Business Intelligence: Databases and Information Management

True-False Questions

1. A grouping of characters into a word, a group of words, or a complete number is called a record.

   Answer: False   **Difficulty: Easy**   **Reference: p. 226**

2. Each characteristic or quality describing a particular entity is called an attribute.

   Answer: True   **Difficulty: Easy**   **Reference: pp. 226–227**

3. A DBMS separates the logical and physical views of the data.

   Answer: True   **Difficulty: Medium**   **Reference: p. 229**

4. Every record in a file should contain at least one key field.

   Answer: True   **Difficulty: Easy**   **Reference: p. 230**

5. Many applications today require databases that can store and retrieve multimedia.

   Answer: True   **Difficulty: Easy**   **Reference: p. 233**

6. One of the drawbacks to OODBMS are that they cannot work with applets.

   Answer: False   **Difficulty: Medium**   **Reference: p. 233**

7. OODBMS are slower than relational DBMS.

   Answer: True   **Difficulty: Medium**   **Reference: p.233**

8. A data dictionary is a language associated with a database management system that end users and programmers use to manipulate data in the database.

   Answer: False   **Difficulty: Easy**   **Reference: pp. 233–234**

9. In a relational database, complex groupings of data must be streamlined to eliminate awkward many-to-many relationships.

   Answer: True   **Difficulty: Easy**   **Reference: p. 236**

10. A data warehouse may be updated by a legacy system.

   Answer: True   **Difficulty: Easy**   **Reference: p. 239**

11. A data warehouse is typically comprised of several smaller data marts.

   Answer: False   **Difficulty: Medium**   **Reference: p. 239**

12. OLAP is a key tool of BI.

   Answer: True   **Difficulty: Easy**   **Reference: p. 240**

13. OLAP enables users to obtain online answers to ad-hoc questions in a fairly rapid amount of time, except when the data are stored in very large databases.

   Answer: False   **Difficulty: Medium**   **Reference: p. 241**

14. Data mining can be a challenge to individual privacy.

   Answer: True   **Difficulty: Easy**   **Reference: p. 243**

15. Predictive analysis is synonymous with data mining.

   Answer: False   **Difficulty: Hard**   **Reference: p. 243**

16. In a Web-based database, middleware is used to transfer information from the organization's internal database back to the Web server for delivery in the form of a Web page to the user.

   Answer: True   **Difficulty: Medium**   **Reference: p. 245**

17. A Web interface requires changes to the internal database.

   Answer: False   **Difficulty: Easy**   **Reference: p. 246**

18. Common Gateway Interface is a specification for processing data on a Web server.

   Answer: True   **Difficulty: Medium**   **Reference: p. 245**

19. Data administration is a special organizational function that manages the policies and procedures through which data can be managed as an organizational resource.

   Answer: True   **Difficulty: Medium**   **Reference: p. 247**

20. Data cleansing is the same process as data scrubbing.

   Answer: True   **Difficulty: Easy**   **Reference: p. 248**

Multiple-Choice Questions

21. *Analysis*

In the chapter-opening case, which of the four generic strategies against competitive forces did NASCAR focus on in implementing its new information systems?

a. Customer and supplier intimacy
b. Product differentiation
c. Low-cost leadership
d. Focus on market niche

Answer: a   **Difficulty: Medium**   **Reference: 223–234**

Analysis in terms of categorize

22. *Analysis*

For which function was NASCAR's new information system geared toward?

a. Supply chain management
b. Sales and marketing
c. Production and manufacturing
d. Finance and accounting

Answer: b   **Difficulty: Easy**   **Reference: 223–234**

Analysis in terms of categorize

23. *Analysis*

Which of the following best illustrates the relationship between entities and attributes?

a. The entity CUSTOMER with the attribute PRODUCT
b. The entity CUSTOMER with the attribute PURCHASE
c. The entity PRODUCT with the attribute PURCHASE
d. The entity PRODUCT with the attribute CUSTOMER

Answer: b   **Difficulty: Hard**   **Reference: pp. 226–227**

Analysis in terms of categorize

24. A characteristic or quality describing an entity is called a(n):

   a. field.
   b. tuple.
   c. key field.
   d. attribute.

   Answer: d   **Difficulty: Easy**   **Reference: pp. 226–227**

25. *Analysis*

   An example of a pre-digital database is :

   a. a library's card catalog.
   b. a cash register receipt.
   c. a doctor's office invoice.
   d. a list of sales totals on a spreadsheet.

   Answer: a   **Difficulty: Medium**   **Reference: p. 227**

   Analysis in terms of categorize

26. The confusion created by _____ makes it difficult for companies to create customer relationship management, supply chain management, or enterprise systems that integrate data from different sources.

   a. batch processing
   b. data redundancy
   c. data independence
   d. online processing

   Answer: b   **Difficulty: Easy**   **Reference: p. 228**

27. Duplicate data in multiple data files is:

   a. data redundancy.
   b. data multiplication.
   c. data independence.
   d. data backups.

   Answer: a   **Difficulty: Easy**   **Reference: p. 228**

28. A DBMS:

   a. makes the physical database available for different logical views.
   b. makes the logical database available for different analytical views.
   c. makes the physical database available for different analytical views.
   d. makes the relational database available for different analytical views.

   Answer: a   **Difficulty: Medium**   **Reference: p. 229**

29.   The logical view:

        a.   shows how data are organized and structured on the storage media.
        b.   presents an entry screen to the user.
        c.   allows the creation of supplementary reports.
        d.   presents data as perceived by end users.

        Answer: d Difficulty: Medium   Reference: p. 229

30.   The type of logical database model that treats data as if they were stored in two-dimensional tables is the:

        a.   OODBMS.
        b.   pre-digital DBMS.
        c.   relational DBMS.
        d.   hierarchical DBMS.

        Answer: c   Difficulty: Easy   Reference: p. 230

31.   In a table for customers, the information about a single customer would reside in a single:

        a.   field.
        b.   row.
        c.   column.
        d.   table.

        Answer: b   Difficulty: Easy   Reference: p. 230

32.   In a relational database, a record is referred to in technical terms as a(n):

        a.   tuple.
        b.   row.
        c.   entity.
        d.   field.

        Answer: a   Difficulty: Medium   Reference: p. 230

33.   A field identified in a table as holding the unique identifier of the table's records is called the:

        a.   primary key.
        b.   key field.
        c.   primary field.
        d.   unique ID.

        Answer: a   Difficulty: Medium   Reference: p. 230

34. A field identified in a record as holding the unique identifier for that record is called the:

   a. primary key.
   b. key field.
   c. primary field.
   d. unique ID.

   Answer: b   **Difficulty: Medium**   **Reference: p. 230**

35. DBMS for midrange computers include all of the following EXCEPT:

   a. DB2.
   b. Oracle.
   c. Microsoft SQL Server.
   d. Microsoft Access.

   Answer: d   **Difficulty: Medium**   **Reference: p. 230**

36. Oracle Database Lite is a(n):

   a. DBMS for small handheld computing devices.
   b. Internet DBMS.
   c. mainframe relational DBMS.
   d. DBMS for midrange computers.

   Answer: a   **Difficulty: Hard**   **Reference: p. 230**

37. Microsoft SQL Server is a(n):

   a. DBMS for small handheld computing devices.
   b. Internet DBMS.
   c. desktop relational DBMS.
   d. DBMS for midrange computers.

   Answer: d   **Difficulty: Hard**   **Reference: p. 230**

38. In a relational database, the three basic operations used to develop useful sets of data are:

   a. select, project, and where.
   b. select, join, and where.
   c. select, project, and join.
   d. select, from, and join.

   Answer: c   **Difficulty: Medium**   **Reference: p. 231**

39.	The select operation:

    a.   combines relational tables to provide the user with more information than is otherwise available.
    b.   creates a subset consisting of columns in a table.
    c.   identifies the table from which the columns will be selected.
    d.   creates a subset consisting of all records in the file that meet stated criteria.

Answer: d	**Difficulty: Easy**	**Reference: p. 231**

40.	The join operation:

    a.   combines relational tables to provide the user with more information than is otherwise available.
    b.   identifies the table from which the columns will be selected.
    c.   creates a subset consisting of columns in a table.
    d.   organizes elements into segments.

Answer: a	**Difficulty: Easy**	**Reference: pp. 231–233**

41.	The project operation:

    a.   combines relational tables to provide the user with more information than is otherwise available.
    b.   creates a subset consisting of columns in a table.
    c.   organizes elements into segments.
    d.   identifies the table from which the columns will be selected.

Answer: b	**Difficulty: Easy**	**Reference: p. 233**

42.	The data dictionary serves as an important data management tool by:

    a.   assigning attributes to the data.
    b.   creating an inventory of data contained in the database.
    c.   presenting data as end users or business specialists would perceive them.
    d.   maintaining data in updated form.

Answer: b	**Difficulty: Medium**	**Reference: pp. 233–234**

43.	An automated or manual file that stores information about data elements and data characteristics such as usage, physical representation, ownership, authorization, and security is the:

    a.   data dictionary.
    b.   data definition diagram.
    c.   entity-relationship diagram.
    d.   relationship dictionary.

Answer: a	**Difficulty: Medium**	**Reference: pp. 233–234**

44.   The type of database management approach that can handle multimedia is the:

   a.   hierarchical DBMS.
   b.   relational DBMS.
   c.   network DBMS.
   d.   object-oriented DBMS.

Answer: d   **Difficulty: Easy**   **Reference: p. 233**

45.   Which of the following database types is useful for storing java applets as well as processing large numbers of transactions?

   a.   Relational DBMS
   b.   Hierarchical DBMS
   c.   Object-relational DBMS
   d.   OODBMS

Answer: c   **Difficulty: Medium**   **Reference: p. 233**

46.   The specialized language programmers use to add and change data in the database is called:

   a.   data access language.
   b.   data manipulation language.
   c.   structured Query language.
   d.   data definition language.

Answer: b   **Difficulty: Easy**   **Reference: p. 234**

47.   The most prominent data manipulation language today is:

   a.   Access.
   b.   DB2.
   c.   SQL.
   d.   Crystal Reports.

Answer: c   **Difficulty: Medium**   **Reference: p. 234**

48.   DBMS typically include report-generating tools in order to:

   a.   retrieve and display data.
   b.   display data in an easier-to-read format.
   c.   display data in graphs.
   d.   perform predictive analysis.

Answer: b   **Difficulty: Medium**   **Reference: p. 235**

49. A schematic of the entire database that describes the relationships in a database is called a(n):

   a. data dictionary.
   b. intersection relationship diagram.
   c. entity-relationship diagram.
   d. data definition diagram.

   Answer: c   **Difficulty: Medium**   **Reference: p. 236**

50. A one-to-one relationship between two entities is symbolized in a diagram by a line that:

   a. ends in two short marks.
   b. ends in one short mark.
   c. ends with a crow's foot.
   d. ends with a crow's foot topped by a short mark.

   Answer: a   **Difficulty: Hard**   **Reference: p. 236**

51. A one-to-many relationship between two entities is symbolized in a diagram by a line that:

   a. ends in two short marks.
   b. ends in one short mark.
   c. ends with a crow's foot.
   d. ends with a crow's foot topped by a short mark.

   Answer: d   **Difficulty: Hard**   **Reference: p. 236**

52. The process of streamlining data to minimize redundancy and awkward many-to-many relationships is called:

   a. normalization.
   b. data scrubbing.
   c. data cleansing.
   d. data defining.

   Answer: a   **Difficulty: Easy**   **Reference: p. 236**

53. In what type of a database system is the central database replicated at all remote locations?

   a. Partitioned
   b. Normalized
   c. Distributed
   d. Networked

   Answer: c   **Difficulty: Medium**   **Reference: p. 238**

54. The disadvantage of a distributed database system is:

    a. lack of flexibility.
    b. susceptibility to data inconsistency.
    c. poor responsiveness to local users.
    d. require more expensive computers.

Answer: b   **Difficulty: Medium**   **Reference: p. 238**

55. A data warehouse is composed of:

    a. historical data from legacy systems.
    b. current data.
    c. internal and external data sources.
    d. historic and current internal data.

Answer: d   **Difficulty: Medium**   **Reference: p. 239**

56. A data mart usually can be constructed more rapidly and at lower cost than a data warehouse because:

    a. a data mart typically focuses on a single subject area or line of business.
    b. all the information is historical.
    c. a data mart uses a Web interface.
    d. all of the information belongs to a single company.

Answer: a   **Difficulty: Easy**   **Reference: p. 239**

57. Tools for consolidating, analyzing, and providing access to vast amounts of data to help users make better business decisions are known as:

    a. DSS.
    b. business intelligence.
    c. OLAP.
    d. data mining.

Answer: b   **Difficulty: Medium**   **Reference: p. 240**

58. The tool that enables users to view the same data in different ways using multiple dimensions is:

    a. predictive analysis.
    b. SQL.
    c. OLAP.
    d. data mining.

Answer: c   **Difficulty: Hard**   **Reference: p. 241**

59. OLAP is a tool for enabling:

    a. users to obtain online answers to ad-hoc questions in a rapid amount of time.
    b. users to view both logical and physical views of data.
    c. programmers to quickly diagram data relationships.
    d. programmers to normalize data.

    Answer: a   **Difficulty: Easy**   **Reference: p. 241**

60. Data mining is a tool for allowing users to:

    a. quickly compare transaction data gathered over many years.
    b. find hidden relationships in data.
    c. obtain online answers to ad hoc questions in a rapid amount of time.
    d. summarize massive amounts of data into much smaller, traditional reports.

    Answer: b   **Difficulty: Medium**   **Reference: p. 242**

61. In terms of data relationships, associations refers to:

    a. events linked over time.
    b. patterns that describe a group to which an item belongs.
    c. occurrences linked to a single event.
    d. undiscovered groupings.

    Answer: c   **Difficulty: Hard**   **Reference: p. 242**

62. *Analysis*

    What category best describes the FBI's CODIS system described in the chapter case?

    a. TPS
    b. MIS
    c. KMS
    d. ESS

    Answer: c   **Difficulty: Medium**   **Reference: pp. 244–245**

    Analysis in terms of categorize

63. An alternative to using application server software for interfacing between a Web server and back-end databases is:

    a. CGI.
    b. HTML.
    c. Java.
    d. SQL.

    Answer: a   **Difficulty: Easy**   **Reference: p. 245**

64. The special organizational function whose responsibilities include the technical and operational aspects of managing data, including physical database design and maintenance, is called:

a. data administration.
b. database administration.
c. information policy administration.
d. data auditing.

Answer: b   **Difficulty: Easy**   **Reference: p. 247**

65. *Synthesis*

What is the first step you should take in managing data for a firm?

a. Identify the data needed to run the business.
b. Cleanse the data before importing it to any database.
c. Normalize the data before importing to a database.
d. Audit your data quality.

Answer: a   **Difficulty: Medium**   **Reference: p. 247**

Synthesis in terms of assemble, plan

66. *Analysis*

Which common database challenge is illustrated by the text's discussion of receiving multiple pieces of the same direct mail advertising?

a. Data normalization
b. Data accuracy
c. Data redundancy
d. Data inconsistency

Answer: d   **Difficulty: Hard**   **Reference: p. 248**

Analysis in terms of categorize

67. Detecting and correcting data in a database or file that are incorrect, incomplete, improperly formatted, or redundant is called:

a. data auditing.
b. defragmentation.
c. data scrubbing.
d. data optimization.

Answer: c   **Difficulty: Easy**   **Reference: p. 248**

68. As discussed in the Interactive Session: Management, what problem was British Telecom facing in its data file environment?

    a. Program-data dependence
    b. Data redundancy
    c. Lack of data sharing
    d. Data quality

    Answer: d   **Difficulty: Hard**   **Reference: p. 249**

69. *Analysis*

    The systems developed by Panasonic in the chapter case study fall into what category of information system?

    a. Finance and accounting system
    b. Sales and marketing system
    c. Human resources system
    d. Manufacturing and production system

    Answer: b   **Difficulty: Medium**   **Reference: pp. 257–259**

    Analysis in terms of categorize

70. *Evaluation*

    Which was the greatest challenge of the four facing Panasonic in its drive to create a unified information system?

    a. Data definition
    b. Data redundancy
    c. Outdated legacy systems
    d. Poor information policies

    Answer: d   **Difficulty: Hard**   **Reference: pp. 257–259**

    Evaluation in terms of assess

Fill in the Blanks

71. Databases record information about general categories of information referred to as *entities*.

          Difficulty: Easy   **Reference: p. 226**

72. A(n) *attribute* is a piece of information describing a particular entity.

          Difficulty: Easy   **Reference: pp. 226–227**

73. A(n) *physical* view shows data as it is actually organized and structured on the data storage media.

          Difficulty: Easy   **Reference: p. 229**

74. A(n) *field* is a column in a relational database used for storing individual elements of data.

          Difficulty: Medium   **Reference: p. 230**

75. A(n) *record/tuple* is a row of data in a database table.

          Difficulty: Easy   **Reference: p. 230**

76. A(n) *key field/primary key* is a field in a record that uniquely identifies instances of that record so that it can be retrieved, updated, or sorted.

          Difficulty: Easy   **Reference: p. 230**

77. DBMS have a *data definition* capability to specify the structure of the content of the database.

          Difficulty: Medium   **Reference: p. 233**

78. *Structured query language (SQL)* is the most prominent data manipulation language for relational database management systems.

          Difficulty: Easy   **Reference: p. 234**

79. A(n) *application server* is software that handles all application operations between browser-based computers and a company's back-end business applications or databases.

          Difficulty: Easy   **Reference: p. 245**

80. *Information policies* are the formal rules governing the maintenance, distribution, and use of information in an organization.

          Difficulty: Medium   **Reference: p. 247**

Essay Questions

81. *Synthesis*

The small publishing company you work for wants to create a new database for storing information about all of their author contracts. What factors will influence how you design the database?

Student answers will vary, but should include some assessment of data quality; business processes and user needs; and relationship to existing IT systems. Key points to include are:
- Data accuracy when the new data is input,
- Establishing a good data model,
- Determining which data is important and anticipating what the possible uses for the data will be, beyond looking up contract information,
- Technical difficulties linking this system to existing systems,
- New business processes for data input and handling, and contracts management,
- Determining how end users will use the data,
- Making data definitions consistent with other databases,
- Determining what methods to use to cleanse the data.

Difficulty: Medium   **Reference: pp. 228, 247–248**

Synthesis in terms of build, formulate

82. *Analysis*

When you design the new contracts database for the publishing house mentioned above, what fields do you anticipate needing? Which of these fields might be in use in other databases used by the company?

Author first name, author last name, author address, agent name and address, title of book, book ISBN, date of contract, amount of money, payment schedule, date contract ends.

Other databases might be an author database (author names, address, and agent details), a book title database (title and ISBN of book), and financial database (payments made).

Difficulty: Medium   **Reference: pp. 226, 229–233**

Analysis in terms of diagram

83. **List at least three conditions that contribute to data redundancy and inconsistency.**

Data redundancy occurs when different divisions, functional areas, and groups in an organization independently collect the same piece of information. Because it is collected and maintained in so many different places, the same data item may have: (1) different meanings in different parts of the organization, (2) different names may be used for the same item, and (3) different descriptions for the same condition. In addition, the fields into which the data is gathered may have different field names, different attributes, or different constraints.

Difficulty: Medium   **Reference: p. 228**

84. **List and describe three main capabilities or tools of a DBMS.**

A data definition capability to specify the structure of the content of the database. This capability would be used to create database tables and to define the characteristics of the fields in each table.

A data dictionary to store definitions of data elements in the database and their characteristics. In large corporate databases, the data dictionary may capture additional information, such as usage; ownership; authorization; security; and the individuals, business functions, programs, and reports that use each data element.

A data manipulation language, such as SQL, that is used to add, change, delete, and retrieve the data in the database. This language contains commands that permit end users and programming specialists to extract data from the database to satisfy information requests and develop applications.

Difficulty: Medium   **Reference: pp. 233–235**

85. *Synthesis*

What types of relationships are possible in a relational database? Describe and give an example of each.

A one-to-one relationship occurs when each record in one table has only one related record in a second table. An example might be a table of salespeople and a separate table of company cars. Each salesperson can only have one car, or be related to the one car in the database.

A one-to-many relationship occurs when a record in one table has many related records in a second table. An example might be a table of salespeople and clients. Each salesperson may have several clients.

A many-to-many relationship occurs when records in one table have many related records in a second table, and the records in the second table have many related records in the first table. An example might be a clients table and a products table. Clients may buy more than one product, and products are sold to more than one client.

Difficulty: Hard   **Reference: pp. 236–237**

Synthesis in terms of build, model

86. **What are the three basic operations used to extract useful sets of data from a relational database?**

- The *select* operation creates a subset consisting of all records (rows) in the table that meets stated criteria.
- The *join* operation combines relational tables to provide the user with more information than is available in individual tables.
- The *project* operation creates a subset consisting of columns in a table, permitting the user to create new tables that contain only the information required.

Difficulty: Medium　　　　**Reference: pp. 231–233**

87. *Evaluation*

Describe the ways in which database technologies could be used by an office stationery supply company to achieve low-cost leadership.

Sales databases could be used to make the supply chain more efficient and minimize warehousing and transportation costs. You can also use sales databases to determine what supplies are in demand by which customers, and whether needs are different in different geographical areas. DSS databases using business intelligence could be used to predict future trends in office supply needs, to help anticipate demand, and to determine the most efficient methods of transportation and delivery.

Difficulty: Hard　　　　**Reference: pp. 238–246**

Evaluation in terms of predict, assess

88. *Evaluation*

Describe the ways in which database technologies could be used by an office stationery supply company to achieve product differentiation.

Product databases could be made available to customers for greater convenience when ordering online. Databases could be used to track customer preferences and to help anticipate customer desires. Sales databases could also help a client anticipate when they would need to re-supply, providing an additional service. Data mining could help anticipate trends in sales or other factors to help determine new services and products to sell to the clients.

Difficulty: Hard　　　　**Reference: pp. 238–246**

Evaluation in terms of predict, assess

89. *Evaluation*

What makes data mining an important business tool? What types of information does data mining produce? In what type of circumstance would you advise a company to use data mining?

Data mining is one of the data analysis tools that helps users make better business decisions and is one of the key tools of business intelligence. Data mining allows users to analyze large amounts of data and find hidden relationships between data that otherwise would not be discovered. For example, data mining might find that a customer that buys product X is ten times more likely to buy product Y than other customers.

Data mining finds information such as:
- Associations or occurrences that are linked to a single event.
- Sequences, events that are linked over time.
- Classification, patterns that describe the group to which an item belongs, found by examining existing items that have been classified and by inferring a set of rules.
- Clusters, unclassified but related groups.

I would advise a company to use data mining when they are looking for new products and services, or when they are looking for new marketing techniques or new markets. Data mining might also be helpful when trying to analyze unanticipated problems with sales whose causes are difficult to identify.

Difficulty: Hard   **Reference: pp. 242–243**

Evaluation in terms of assess

90. *Evaluation*

What are the differences between data mining and OLAP? When would you advise a company to use OLAP?

Data mining uncovers hidden relationships and is used when you are trying to discover data and new relationships. It is used to answer questions such as: Are there any product sales that are related in time to other product sales?

In contrast, OLAP is used to analyze multiple dimensions of data and is used to find answers to complex, but known, questions, such as: What were sales of a product—broken down by month and geographical region, and how did those sales compare to sales forecasts?

Difficulty: Hard   **Reference: pp. 241–243**

Evaluation in terms of assess