

1) Online transaction processing (OLTP) systems are the backbone of all cross-functional systems in a company.

- a. True
- b. False

Answer: a

Diff: 1

Type: TF

Page Reference: 258

Topic: Q2

Skill: RECALL

2) Data-mining systems use heuristic techniques to process data to find hidden patterns and relationships among the data.

- a. True
- b. False

Answer: b

Diff: 2

Type: TF

Page Reference: 263

Topic: Q4

Skill: RECALL

3) A reporting system is designed to provide the right information, to the right user, at the right time.

- a. True
- b. False

Answer: a

Diff: 1

Type: TF

Page Reference: 264

Topic: Q4

Skill: RECALL

4) A data-mining system uses statistical analysis to enable fundraisers to predict who will donate to a cause.

- a. True
- b. False

Answer: a

Diff: 2

Type: TF

Page Reference: 262

Topic: Q4

Skill: RECALL

5) Data-mining is about probabilities not certainties. So sometimes a good model may look bad because of a string of bad luck.

a. True

b. False

Answer: a

Diff: 2

Type: TF

Page Reference: 276

Topic: Q7

Skill: RECALL

6) Reporting systems are programs that read data from a variety of sources, process that data, produce and deliver formatted reports to users.

a. True

b. False

Answer: a

Diff: 2

Type: TF

Page Reference: 263

Topic: Q4

Skill: RECALL

7) Fine data are highly summarized.

a. True

b. False

Answer: b

Diff: 1

Type: TF

Page Reference: 257

Topic: Q1

Skill: RECALL

8) Coarse data are highly summarized.

- a. True
- b. False

Answer: a

Diff: 1

Type: TF

Page Reference: 257

Topic: Q1

Skill: RECALL

9) Clickstream data is very fine data.

- a. True
- b. False

Answer: a

Diff: 1

Type: TF

Page Reference: 257

Topic: Q1

Skill: RECALL

10) Generally it is better to have data that is too fine rather than too coarse.

- a. True
- b. False

Answer: a

Diff: 1

Type: TF

Page Reference: 257

Topic: Q1

Skill: RECALL

11) Systems that focus on making data collected in OLTP to support decision making are often referred to as knowledge management systems.

- a. True

b. False

Answer: b

Diff: 2

Type: TF

Page Reference: 258

Topic: Q3

Skill: RECALL

12) Systems that focus on making data collected in OLTP to support decision making are often referred to as expert systems.

a. True

b. False

Answer: b

Diff: 2

Type: TF

Page Reference: 258

Topic: Q3

Skill: RECALL

13) Systems that focus on making data collected in OLTP to support decision making are often referred to as decision support systems.

a. True

b. False

Answer: a

Diff: 2

Type: TF

Page Reference: 258

Topic: Q3

Skill: RECALL

14) OLTP provides the ability to sum, count, average, and perform other simple arithmetic operations on groups of data.

a. True

b. False

Answer: b

Diff: 2

Type: TF

Page Reference: 259

Topic: Q3

Skill: RECALL

15) OLAP provides the ability to sum, count, average, and perform other simple arithmetic operations on groups of data.

a. True

b. False

Answer: a

Diff: 2

Type: TF

Page Reference: 259

Topic: Q3

Skill: RECALL

16) With cluster analysis, analysts create a model or hypothesis before running the analysis.

a. True

b. False

Answer: b

Diff: 2

Type: TF

Page Reference: 268

Topic: Q7

Skill: APPLIED

17) With supervised data mining, analysts create a model or hypothesis before running the analysis.

a. True

b. False

Answer: a

Diff: 2

Type: TF

Page Reference: 268

Topic: Q7

Skill: RECALL

18) Neural networks are a supervised data-mining technique.

- a. True
- b. False

Answer: a

Diff: 2

Type: TF

Page Reference: 267

Topic: Q7

Skill: RECALL

19) Cluster analysis measures the impact of a set of variables on another variable.

- a. True
- b. False

Answer: b

Diff: 2

Type: TF

Page Reference: 268

Topic: Q7

Skill: RECALL

20) Regression analysis measures the impact of a set of variables on another variable.

- a. True
- b. False

Answer: a

Diff: 2

Type: TF

Page Reference: 268

Topic: Q7

Skill: RECALL

21) A data warehouse is a data collection that is created to address the needs of a particular business function, problem, or opportunity.

- a. True
- b. False

Answer: b

Diff: 2

Type: TF

Page Reference: 264

Topic: Q6

Skill: RECALL

22) A data mart is a data collection that is created to address the needs of a particular business function, problem, or opportunity.

a. True

b. False

Answer: a

Diff: 1

Type: TF

Page Reference: 266

Topic: Q6

Skill: RECALL

23) Problematic data are termed dirty data.

a. True

b. False

Answer: a

Diff: 1

Type: TF

Page Reference: 256

Topic: Q1

Skill: RECALL

24) Problematic data are termed missing values.

a. True

b. False

Answer: b

Diff: 1

Type: TF

Page Reference: 256

Topic: Q1

Skill: RECALL

25) The data resource challenge refers to the lack of data to support decision making.

- a. True
- b. False

Answer: b

Diff: 2

Type: TF

Page Reference: 256

Topic: Q3

Skill: RECALL

26) Market-basket analysis is a data-mining system, which computes correlations based on past orders to determine items that are frequently purchased together.

- a. True
- b. False

Answer: a

Diff: 2

Type: TF

Page Reference: 264

Topic: Q6

Skill: RECALL

27) Experts systems encapsulate knowledge and use if/then rules to support decisions.

- a. True
- b. False

Answer: a

Diff: 2

Type: TF

Page Reference: 264

Topic: Q4

Skill: RECALL

28) With supervised data mining, data miners develop a model prior to the analysis and apply statistical techniques to data to estimate parameters of the model.



- a. True
- b. False

Answer: a

Diff: 2

Type: TF

Page Reference: 268

Topic: Q7

Skill: RECALL

29) Discovering patterns and relationships in data is the purpose of data mining.

- a. True
- b. False

Answer: a

Diff: 2

Type: TF

Page Reference: 267

Topic: Q7

Skill: RECALL

30) Managers suffer more from an overabundance of irrelevant data.

- a. True
- b. False

Answer: a

Diff: 1

Type: TF

Page Reference: 255

Topic: Q1

Skill: RECALL

31) The total amount of data generated by 2010 is estimated to be

- a. 330 petabytes.
- b. 3.3 zettabytes.
- c. 33 petabytes.
- d. 3.3 exabytes.
- e. 0.33 yottabytes.

Answer: d

Diff: 2

Type: MC

Page Reference: 256

Topic: Q1

Skill: RECALL

32) The hyper abundance of information is sometimes called

- a. data abundance.
- b. Moore's Law.
- c. the Ackoff principle.
- d. information overload.
- e. dirty data.

Answer: d

Diff: 1

Type: MC

Page Reference: 255

Topic: Q1

Skill: RECALL

33) Data \_\_\_\_\_ refers to the degree of summarization or detail.

- a. relevance
- b. granularity
- c. integrity
- d. clarity
- e. quality

Answer: b

Diff: 2

Type: MC

Page Reference: 257

Topic: Q2

Skill: RECALL

34) If you are collecting data electronically and processing the transaction online, then you are using (a) \_\_\_\_\_.

- a. online analytical processing (OLAP)
- b. online transaction processing (OLTP)
- c. data warehouse
- d. data-mining
- e. database management system (DBMS)

Answer: b

Diff: 1

Type: MC

Page Reference: 257

Topic: Q2

Skill: RECALL

35) A person's clicking behavior on the web is termed as their

- a. mouse trails.
- b. webstream.
- c. clickstream.
- d. web tracks.
- e. link journey.

Answer: c

Diff: 1

Type: MC

Page Reference: 257

Topic: Q1

Skill: RECALL

36) Problematic data are termed

- a. problem data.
- b. irrelevant data.
- c. incorrect data.
- d. dirty data.
- e. untrustworthy data.

Answer: d

Diff: 1

Type: MC

Page Reference: 256

Topic: Q1

Skill: RECALL

37) While data may be collected in OLTP, the fact that data may not be used to improve decision making refers to (the)

- a. data overload challenge.
- b. information overload.
- c. data resource challenge.
- d. data integrity challenge.
- e. knowledge challenge.

Answer: c

Diff: 2

Type: MC

Page Reference: 258

Topic: Q3

Skill: RECALL

38) A \_\_\_\_\_ is a data collection that is created to address the needs of a particular business function, problem, or opportunity.

- a. database
- b. data mart
- c. data-mine
- d. data depot
- e. data warehouse

Answer: b

Diff: 2

Type: MC

Page Reference: 266

Topic: Q6

Skill: RECALL

39) The purpose of a business intelligence system is to provide information that

- a. improves decision making.
- b. can be summarized.
- c. is less granular.
- d. follows privacy legislation.
- e. has only one level of complexity.

Answer: a

Diff: 2

Type: MC

Page Reference: 262

Topic: Q4

Skill: RECALL

40) \_\_\_\_\_ is the application of statistical techniques to find patterns and relationships in data and to predict outcomes.

- a. Data mining
- b. Data warehousing
- c. Data depositing
- d. Data marting
- e. Databasing

Answer: a

Diff: 3

Type: MC

Page Reference: 263

Topic: Q4

Skill: RECALL

41) With \_\_\_\_\_, analysts do not create a model or hypothesis before running the analysis. Instead, they apply the data-mining technique to the data and observe the results.

- a. regression analysis
- b. supervised data mining

- c. neural networks
- d. cluster analysis
- e. unsupervised data mining

Answer: e

Diff: 2

Type: MC

Page Reference: 268

Topic: Q7

Skill: RECALL

42) \_\_\_\_\_ measure(s ) the impact of a set of variables on another variable.

- a. Supervised data mining
- b. Regression analysis
- c. Unsupervised data mining
- d. Neural networks
- e. Cluster analysis

Answer: b

Diff: 2

Type: MC

Page Reference: 268

Topic: Q7

Skill: RECALL

43) \_\_\_\_\_ is a statistical technique that can identify groups of entities that have similar characteristics.

- a. Regression analysis
- b. Data mining
- c. Cluster analysis
- d. Character analysis
- e. Supervised data mining

Answer: c

Diff: 2

Type: MC

Page Reference: 268

Topic: Q7

Skill: RECALL

44) If transactions are entered and processed immediately upon entry, then the system is operating in

- a. transaction time.
- b. batch mode.
- c. unprotected mode.
- d. online mode.
- e. real time.

Answer: e

Diff: 1

Type: MC

Page Reference: 257

Topic: Q2

Skill: RECALL

45) A phone number of 123-456-7890 and an age of 999 are examples of \_\_\_\_\_ in a database.

- a. granularity
- b. OLTP
- c. incorrect data labels
- d. dirty data
- e. too many attributes

Answer: d

Diff: 3

Type: MC

Page Reference: 256

Topic: Q1

Skill: RECALL

46) \_\_\_\_\_ integrate data from multiple sources, and process those data by sorting, grouping, summing, averaging, and comparing.

- a. Reporting systems
- b. Supplier Relationship Management systems
- c. Business intelligence systems
- d. Expert Systems
- e. Customer Relationship Management systems

Answer: a

Diff: 1

Type: MC

Page Reference: 264

Topic: Q4

Skill: RECALL

47) \_\_\_\_\_ process data using sophisticated statistical techniques like regression analysis and decision tree analysis.

- a. CRM systems
- b. Data-mining systems
- c. Business intelligence systems
- d. Reporting systems
- e. Expert Systems

Answer: b

Diff: 2

Type: MC

Page Reference: 264

Topic: Q4

Skill: RECALL

48) \_\_\_\_\_ create value from intellectual capital by collecting and sharing human knowledge of products, product uses, best practices, and other critical knowledge with employees, managers, customers, suppliers, and other who need it.

- a. CRM applications
- b. Expert Systems
- c. Business intelligence systems
- d. Data mining systems
- e. Knowledge management systems



Answer: e

Diff: 2

Type: MC

Page Reference: 264

Topic: Q4

Skill: RECALL

49) \_\_\_\_\_ encapsulate the knowledge of human experts in the form of if-then rules.

- a. Data mining systems
- b. Expert Systems
- c. Business intelligence applications
- d. Reporting systems
- e. CRM systems

Answer: b

Diff: 1

Type: MC

Page Reference: 264

Topic: Q4

Skill: RECALL

50) The purpose of (a) \_\_\_\_\_ is to extract and clean data from operational systems and other sources.

- a. cluster analysis
- b. data mart
- c. data warehouse
- d. regression analysis
- e. data mining

Answer: c

Diff: 1

Type: MC

Page Reference: 265

Topic: Q5

Skill: RECALL

51) In 2002, a total of \_\_\_\_\_ petabytes of new data were created.

- a. 304
- b. 330
- c. 3.3
- d. 403
- e. 430

Answer: d

Diff: 2

Type: MC

Page Reference: 255

Topic: Q1

Skill: RECALL

52) \_\_\_\_\_ provides the ability to sum, count, average, and perform other simple arithmetic operation on groups of data.

- a. Online database processing (OLDP)
- b. Online transaction processing (OLTP)
- c. Online data processing (OLDP)
- d. Online record processing (OLRP)
- e. Online analytic processing (OLAP)

Answer: e

Diff: 2

Type: MC

Page Reference: 259

Topic: Q3

Skill: RECALL

53) An OLAP report has

- a. measures and dimensions.
- b. averages.
- c. dimensions.
- d. measures and averages.
- e. measures.

Answer: a

Diff: 2

Type: MC

Page Reference: 259

Topic: Q3

Skill: RECALL

54) Systems that focus on making data collected in OLTP useful for allowing multiple decision makers to collaborate are often referred to as

- a. Executive Information Systems.
- b. Data Mining Systems.
- c. Expert Systems.
- d. Knowledge Management Systems.
- e. Decision Support Systems.

Answer: e

Diff: 2

Type: MC

Page Reference: 261

Topic: Q3

Skill: RECALL

55) \_\_\_\_\_ analysis is a data-mining system which computes correlation of items on past orders to determine items that are frequently purchased together.

- a. Market-basket
- b. Marketing
- c. Value chain
- d. Data-mining
- e. Shopping cart

Answer: a

Diff: 2

Type: MC

Page Reference: 269

Topic: Q7

Skill: RECALL

56) If you need to identify groups of entities that have similar characteristics, you need to perform (a)

- a. cluster analysis.
- b. regression analysis.
- c. neural network analysis.
- d. online analytical processing.
- e. grouping analysis.

Answer: a

Diff: 3

Type: MC

Page Reference: 268

Topic: Q7

Skill: APPLIED

57) Which of the following systems fosters innovation, improves customer service, and increases organizational responsiveness by sharing knowledge of products, product uses, and best practices?

- a. Market-basket Analysis
- b. Regression Analysis Systems
- c. Data-mining Systems
- d. Knowledge Management Systems
- e. Expert Systems

Answer: d

Diff: 3

Type: MC

Page Reference: 264

Topic: Q4

Skill: RECALL

58) Which of the following is an assumption about information systems that Ackoff disputed?

- a. Computer speed doubles every 18 months.

- b. There is too much uncertainty in making any decision.
- c. Managers have no problem making decisions if they get the data they need.
- d. Managers that make decisions do not get irrelevant data.
- e. There is too much information.

Answer: c

Diff: 3

Type: MC

Page Reference: 255

Topic: Q1

Skill: APPLIED

59) Which of the following is an assumption about information systems that Ackoff disputed?

- a. There is too much uncertainty in making any decision.
- b. Decisions are poor because managers lack relevant information.
- c. Computer speed doubles every 18 months.
- d. There is too much information.
- e. Managers suffer from a lack of perfect data.

Answer: b

Diff: 3

Type: MC

Page Reference: 255

Topic: Q1

Skill: APPLIED

60) Which of the following is an assumption about information systems that Ackoff disputed?

- a. Managers are aware of the data they need.
- b. There is too much information.
- c. There is too much uncertainty in making any decisions.
- d. Computer speed doubles every 18 months.
- e. Managers suffer from a lack of training.

Answer: a

Diff: 3

Type: MC

Page Reference: 255

Topic: Q1

Skill: APPLIED

61) Why does the exponential growth in data matter?

- a. There was too little data before.
- b. Managers know what data they need.
- c. Information overload is a blessing.
- d. Getting the right information at the right time can improve decisions.
- e. You can make money with data.

Answer: d

Diff: 3

Type: MC

Page Reference: 254

Topic: Q1

Skill: APPLIED

62) It is usually better to have data granularity that is \_\_\_\_\_ rather than too coarse.

- a. dirty data
- b. missing values
- c. inconsistent
- d. more recent
- e. too fine

Answer: e

Diff: 1

Type: MC

Page Reference: 257

Topic: Q1

Skill: RECALL

63) The data processing strategy that waits for many transactions to pile up before they are processed is called

- a. LDAP.
- b. delayed processing.
- c. batch processing.
- d. OLTP.
- e. OLAP.

Answer: c

Diff: 1

Type: MC

Page Reference: 255

Topic: Q2

Skill: RECALL

64) \_\_\_\_\_ systems are the backbone of all functional, cross-functional, and inter-organizational systems in an organization.

- a. OLAP
- b. Batch processing
- c. OLTP
- d. LDAP
- e. Real time

Answer: c

Diff: 2

Type: MC

Page Reference: 256

Topic: Q2

Skill: RECALL

65) Thomas Davenport noted that a major grocery chain used less than \_\_\_\_\_ percent of its scanner data

- a. 1
- b. 2
- c. 3
- d. 4

e. 5

Answer: b

Diff: 3

Type: MC

Page Reference: 256

Topic: Q3

Skill: RECALL

66) Which of the following is a synonym for data-mining?

- a. Data discovery
- b. Cluster analysis
- c. Data mart
- d. Data parsing
- e. Knowledge discovery in databases

Answer: e

Diff: 3

Type: MC

Page Reference: 265

Topic: Q4

Skill: APPLIED

67) Experts in a communications company believe that cell phone usage on weekends is determined by the age of the customer and the number of months the customer has the cell phone account. Which of the following technique could be used to analyze the data?

- a. Regression analysis
- b. Unsupervised data-mining
- c. Neural networks
- d. Cluster analysis
- e. Market-based analysis

Answer: a

Diff: 3

Type: MC

Page Reference: 266



Topic: Q7

Skill: APPLIED

68) When an area code changes in a database, the phone number for a given customer before the change will not match the customer's number after the change. This is an example of

- a. too fine data.
- b. too coarse data.
- c. missing values.
- d. incomplete or inaccurate data.
- e. inconsistent data.

Answer: e

Diff: 3

Type: MC

Page Reference: 256

Topic: Q1

Skill: RECALL

69) Clickstream data is often \_\_\_\_\_ and much of it has to be discarded.

- a. incomplete or inaccurate data.
- b. too coarse.
- c. inconsistent data.
- d. missing value.
- e. too fine.

Answer: e

Diff: 3

Type: MC

Page Reference: 257

Topic: Q1

Skill: RECALL

70) A nonprofit organization can process a donation without knowing the donor's gender or age, but a data mining application that looks at differences in donation behaviours based on gender would be impaired. This data problem is known as

- a. too fine data.

- b. missing values.
- c. inconsistent data.
- d. incomplete or inaccurate data.
- e. too coarse data.

Answer: b

Diff: 2

Type: MC

Page Reference: 254

Topic: Q1

Skill: RECALL

71) Highly summarized data is known as being

- a. coarse.
- b. fine.
- c. inconsistent.
- d. too summarized.
- e. too comprehensive.

Answer: a

Diff: 2

Type: MC

Page Reference: 257

Topic: Q1

Skill: RECALL

72) Clickstream data is

- a. a record of customer's or visitor's behavior on a website.
- b. not used for market basket predictions.
- c. inconsistent.
- d. too summarized.
- e. the end result of data mining.

Answer: a

Diff: 2

Type: MC

Page Reference: 257

Topic: Q1

Skill: RECALL

73) It is better to have data that is too fine than too coarse because fine data can be converted to coarse data by

- a. separating the data into its constituent parts.
- b. using regression analysis.
- c. placing the data into a data mart.
- d. summing and combining the data.
- e. placing the data into a data warehouse.

Answer: d

Diff: 2

Type: MC

Page Reference: 257

Topic: Q1

Skill: RECALL

74) In an OLAP system a measure is (are)

- a. a data item of interest.
- b. a report.
- c. data items that are summed.
- d. processed data.
- e. a characteristic of a dimension.

Answer: a

Diff: 2

Type: MC

Page Reference: 259

Topic: Q3

Skill: RECALL

75) To realize a competitive advantage from data an organization must

- a. collect the data.
- b. view data as a liability.

- c. use the data to make better decisions.
- d. view data as an asset.
- e. analyze the data.

Answer: c

Diff: 2

Type: MC

Page Reference: 264

Topic: Q4

Skill: RECALL

76) The decision to use batch or real time processing should depend upon

- a. the nature of the transactions and the needs of the organization.
- b. the nature of the transactions.
- c. the needs to the organization.
- d. the existing infrastructure of the organization.
- e. the needs to the organization and the existing infrastructure of the organization.

Answer: a

Diff: 2

Type: MC

Page Reference: 258

Topic: Q2

Skill: RECALL

77) Billy Beane is generally credited with introducing

- a. new methods of scouting players for the National Hockey League.
- b. scientific and statistical techniques that capture more data for major league baseball.
- c. market-basket analysis in major league baseball.
- d. developing OLTP software.
- e. real-time processing to business processes

Answer: b

Diff: 2

Type: MC

Page Reference: 260

Topic: Q4

Skill: RECALL

78) By sharing knowledge companies can

- a. improve the decision making of non-experts.
- b. mine data better.
- c. foster innovation.
- d. have more complete information systems.
- e. automate decision making.

Answer: c

Diff: 2

Type: MC

Page Reference: 264

Topic: Q4

Skill: RECALL

79) Your local burger joint wants to reap the benefits of improved decision making by using its sales data. In order to do this, it must

- a. use reporting systems as part of its complete information systems.
- b. incorporate data-mining products as part of its complete information systems.
- c. follow best practices and adopt an ERP.
- d. adopt an expert system into its existing information system.
- e. adopt knowledge management systems into its existing information system.

Answer: b

Diff: 3

Type: MC

Page Reference: 264

Topic: Q4

Skill: RECALL

80) An RFM analysis is a

- a. data-mining system.
- b. way of recording an individual customer's spending habits.
- c. way of analyzing and ranking customers according to their purchasing patterns.

- d. way of recording a customer's most recent purchases.
- e. way of analyzing a customer's most recent purchases.

Answer: c

Diff: 2

Type: MC

Page Reference: 262

Topic: Q4

Skill: RECALL

81) Data-mining techniques include

- a. supplier relationship management.
- b. data warehousing.
- c. data marting.
- d. market basket analysis.
- e. customer relationship management.

Answer: d

Diff: 2

Type: MC

Page Reference: 270

Topic: Q7

Skill: RECALL

82) In most cases, data-mining systems are used to make

- a. assessments.
- b. predictions.
- c. solutions.
- d. decisions.
- e. relationships.

Answer: b

Diff: 2

Type: MC

Page Reference: 267

Topic: Q7

Skill: RECALL

83) \_\_\_\_\_ systems integrate data from a variety of sources, process that data, and produce and deliver formatted reports to users.

- a. Data-mining
- b. Reporting
- c. Web service
- d. XML
- e. RFP

Answer: b

Diff: 2

Type: MC

Page Reference: 264

Topic: Q4

Skill: RECALL

84) Before it can be loaded into the data warehouse, operational data must be extracted and

- a. sorted.
- b. filtered.
- c. dimensioned.
- d. processed.
- e. cleaned.

Answer: e

Diff: 2

Type: MC

Page Reference: 265

Topic: Q5

Skill: RECALL

85) A data warehouse contains a special database that stores the \_\_\_\_\_, which records its source, history, and format of the data.

- a. dimensional data
- b. clickstream data

- c. metadata
- d. media format
- e. dirty data

Answer: c

Diff: 3

Type: MC

Page Reference: 266

Topic: Q5

Skill: RECALL

86) A data warehouse will usually consist of

- a. large air-conditioned buildings.
- b. thousands of boxes of paper files.
- c. A few fast servers and very large storage devices.
- d. the same data as the company's operational data.
- e. only government related information.

Answer: c

Diff: 2

Type: MC

Page Reference: 265

Topic: Q5

Skill: RECALL

87) Sometimes, large organizations break out \_\_\_\_\_ from their data warehouses so they can be used for a specific functional area of the organization like marketing or finance.

- a. special groupings
- b. Web services
- c. database servers
- d. OLAP applications
- e. data marts

Answer: e

Diff: 2



Type: MC

Page Reference: 266

Topic: Q6

Skill: RECALL

88) \_\_\_\_\_ is a problem that occurs when data-mining is used to create overly complex models that are not suited to making accurate predictions.

- a. Inconsistent data
- b. Overfitting
- c. Model degradation
- d. Dirty data
- e. Clean data

Answer: b

Diff: 3

Type: MC

Page Reference: 273

Topic: Q7

Skill: APPLIED

89) It is almost certain that in reality, when you take on a data-mining project, the data will be

- a. dirty.
- b. full of viruses.
- c. formatted properly.
- d. buggy.
- e. inconsistent.

Answer: a

Diff: 2

Type: MC

Page Reference: 256

Topic: Q1

Skill: APPLIED

90) Data granularity refers to the degree of summarization or detail.

Diff: 2  
Type: FIB  
Page Reference: 257  
Topic: Q1  
Skill: RECALL

91) Customer's web clicking behavior is called clickstream data.

Diff: 2  
Type: FIB  
Page Reference: 257  
Topic: Q1  
Skill: RECALL

92) While data may be collected in OLTP, the data may not be used to improve decision making. We refer to this as the data resource challenge.

Diff: 2  
Type: FIB  
Page Reference: 258  
Topic: Q3  
Skill: RECALL

93) Systems that focus on making data collected in OLTP useful for decision making are often referred to as decision support systems.

Diff: 2  
Type: FIB  
Page Reference: 258  
Topic: Q3  
Skill: RECALL

94) An OLAP report has measures and dimensions.

Diff: 2  
Type: FIB  
Page Reference: 258  
Topic: Q3  
Skill: RECALL

95) Reporting Systems integrate data from multiple sources, and they process those data by sorting, grouping, summing, averaging, and comparing.

Diff: 2  
Type: FIB  
Page Reference: 264  
Topic: Q4  
Skill: RECALL

96) Data-mining systems process data using sophisticated statistical techniques like regression analysis and decision tree analysis.

Diff: 2  
Type: FIB  
Page Reference: 264  
Topic: Q4  
Skill: RECALL

97) Knowledge management systems create value from intellectual capital by collecting and sharing human knowledge of products, product uses, best practices, and other critical knowledge with employees, managers, customers, suppliers, and others who need it.

Diff: 2  
Type: FIB  
Page Reference: 264  
Topic: Q5  
Skill: RECALL

98) Expert systems encapsulate the knowledge of human experts in the form of if/then rules.

Diff: 2  
Type: FIB  
Page Reference: 264  
Topic: Q4  
Skill: RECALL

99) The purpose of a(n) data warehouse is to extract and clean data from operational systems and other sources.

Diff: 2  
Type: FIB  
Page Reference: 265  
Topic: Q5  
Skill: RECALL

100) Metadata is data about data.

Diff: 1  
Type: FIB  
Page Reference: 266  
Topic: Q5  
Skill: RECALL

101) A(n) data mart is a data collection that is made to address the needs of a particular business function, problem, or opportunity.

Diff: 2  
Type: FIB  
Page Reference: 266  
Topic: Q6  
Skill: RECALL

102) Data mining is the application of statistical techniques to find patterns and relationships among data and to classify and predict.

Diff: 2

Type: FIB

Page Reference: 264

Topic: Q4

Skill: RECALL

103) Cluster analysis is a statistical technique that identify groups of entities that have similar characteristics.

Diff: 2

Type: FIB

Page Reference: 268

Topic: Q7

Skill: RECALL

104) Regression analysis measures the impact of a set of variables on another variable.

Diff: 2

Type: FIB

Page Reference: 268

Topic: Q7

Skill: RECALL

105) With supervised data mining, data miners develop a model prior to the analysis and apply statistical techniques to data to estimate parameters of the model.

Diff: 2

Type: FIB

Page Reference: 268

Topic: Q7

Skill: RECALL

106) Many organizations create a special facility called a(n) data warehouse by extracting operational data and cleaning it up for more sophisticated data analysis.

Diff: M

Type: FIB

Page Reference: 265

Topic: Q5

Skill: RECALL

107) market basket is a data-mining system which computes correlations of items on past orders to determine items that are frequently purchased together.

Diff: 2

Type: FIB

Page Reference: 269

Topic: Q7

Skill: RECALL

108) When creating a data warehouse that will be used for BI, programs first must extract, clean and prepare the operational data.

Diff: 2

Type: FIB

Page Reference: 265

Topic: Q5

Skill: RECALL

109) When data sets have missing values or invalid entries it is called dirty data.

Diff: 2

Type: FIB

Page Reference: 256

Topic: Q1

Skill: RECALL

110) If a part number is changed, but the data in the data warehouse does not reflect this change the data is said to be inconsistent.

Diff: 2

Type: FIB

Page Reference: 256

Topic: Q1

Skill: RECALL

111) Granularity refers to the relative level of detail that data represents.

Diff: 2

Type: FIB

Page Reference: 257

Topic: Q1

Skill: RECALL

112) If the granularity of the data is too fine, it is easy to make it coarser by summing and combining.

Diff: 2

Type: FIB

Page Reference: 257

Topic: Q1

Skill: RECALL

113) With unsupervised data mining, data miners do not create a model or hypothesis before running the analysis.

Diff: 2

Type: FIB

Page Reference: 268

Topic: Q7

Skill: RECALL

114) Overfitting is the problem of having many data points and generating the same number of equations to predict those points.

Diff: 3

Type: FIB

Page Reference: 276

Topic: Q7

Skill: RECALL

115) For developing BI applications, it is generally better to have too fine a granularity rather than too coarse.

Diff: 3

Type: FIB

Page Reference: 257

Topic: Q1

Skill: RECALL

116) A data mart is a subset of a data warehouse that addresses a particular component or functional area of the business.

Diff: 2

Type: FIB

Page Reference: 266

Topic: Q6

Skill: RECALL

118) There are two general categories of data-mining techniques; supervised and unsupervised.

Diff: 2

Type: FIB

Page Reference: 268

Topic: Q7

Skill: RECALL

119) Data-mining can be described as "knowledge discovery in databases".

Diff: 3

Type: FIB

Page Reference: 268

Topic: Q7

Skill: APPLIED

120) Supervised data mining happens when analysts create a data model or hypothesis before running the analysis.

Diff: 2

Type: FIB

Page Reference: 268

Topic: Q7

Skill: RECALL

121) Cluster analysis that identifies groups of entities with similar characteristics is a common unsupervised data-mining technique.

Diff: 3

Type: FIB

Page Reference: 268

Topic: Q7

Skill: RECALL

122) Regression analysis is the most common kind of supervised data-mining technique.

Diff: 3

Type: FIB

Page Reference: 268

Topic: Q7

Skill: APPLIED

123) Neural networks are another popular data-mining technique, which were originally based on the physiological analogy of how the neurons in the brain process huge amounts of data.

Diff: 3

Type: FIB

Page Reference: 269

Topic: Q7

Skill: RECALL

124) Data mining projects are risky because you never know how they will turn out. Even with a good model, bad luck may prevail.

Diff: 3

Type: FIB

Page Reference: 276

Topic: Q7

Skill: APPLIED

125) What is unsupervised data mining?

Answer:

In unsupervised data mining, analysts do not create a model or hypothesis before running the analysis. Instead, they apply the data mining technique to the data and observe the results. With this method, analysts create hypotheses after the analysis to explain the patterns found.

Diff: 2

Type: ES

Page Reference: 268

Topic: Q7

Skill: RECALL

126) What is supervised data mining?

Answer:

In supervised data mining, data miners develop a model prior to the analysis and apply statistical techniques to data to estimate the parameters of the model. For example, suppose marketing experts in a communications company believe cell phone usage on weekends is determined by the age of the customer and the number of months the customer has had the cell phone account. A data-mining analyst could then do an analysis that estimates the impact of customer and account age.

Diff: 2

Type: ES

Page Reference: 268

Topic: Q7

Skill: RECALL

127) What is online transaction processing?

Answer:

If you are collecting data electronically and processing the transactions online, then you are using an online transaction processing (OLTP) system.

Diff: 2

Type: ES

Page Reference: 257

Topic: Q2

Skill: RECALL

128) What is online analytic processing?

Answer:

OLAP provides the ability to sum, count, average, and perform other simple arithmetic operations on groups of data. The remarkable characteristic of OLAP reports is that their format is dynamic. The viewer of the report can change the report's structure—hence the term *online*.

Diff: 2



Type: ES

Page Reference: 258

Topic: Q3

Skill: RECALL

129) What is clickstream data?

Answer:

It is possible to capture the customers' clicking behaviour on the Internet. This is clickstream data.

Diff: 2

Type: ES

Page Reference: 257

Topic: Q1

Skill: RECALL

130) Using specific examples, describe data granularity.

Answer:

Data granularity refers to the degree of detail in the data. Coarse data are highly summarized; fine data express precise details.

Diff: 3

Type: ES

Page Reference: 257

Topic: Q1

Skill: APPLIED

131) What are the three assumptions that Ackoff disagreed with?

Answer:

One of the erroneous assumptions was that managers have no problem making decisions if they get the data they need. Another of the erroneous assumptions was that decisions are poor because managers lack relevant information. Another erroneous assumption Ackoff suggested was that managers are aware of the data they need.

Diff: 2

Type: ES

Page Reference: 255

Topic: Q1

Skill: RECALL

132) What is a data mart?

Answer:

A data mart is a data collection that is created to address the needs of a particular business function, problem, or opportunity.

Diff: M RECALL

Type: ES

Page Reference: 266

Topic: Q6

133) Describe the data resource challenge.

Answer:

While data may be collected in OLTP, the data may not be used to improve decision making. We refer to this as the data resource challenge.

Diff: 3

Type: ES

Page Reference: 258

Topic: Q3

Skill: RECALL

134) Is it better to have data that is too fine or too coarse? Explain.

Answer:

Generally, it is better to have too fine a granularity than too coarse. If the granularity is too fine, the data can be made coarser by summing and combining. Only analysts' labour and computer processing are required. If the granularity is too coarse, however, there is no way to separate the data into constituent parts.

Diff: 2

Type: ES

Page Reference: 257

Topic: Q1

Skill: APPLIED

135) What is the purpose of business intelligence? What kind of BI tools are there?

Answer:

Business intelligence (BI) is a category of applications and technologies for gathering, storing, analyzing, and providing access to data to help enterprise users make better business decisions. BI tools include reporting and data mining. Reporting tools are programs that read data from a variety of sources, process that data, produce reports, and deliver those reports to the users who need them. Data-mining tools use statistical techniques to process data in order to find hidden patterns and relationships.

Diff: 3

Type: ES

Page Reference: 262

Topic: Q4

Skill: APPLIED

136) Operational data may not be suitable for more sophisticated analyses. What are the problems with operational data? What is a data warehouse and how is it designed to go beyond these issues?

Answer:

Basic reports and simple OLAP analyses can be made directly from operational data. However, operational data may be unsuited for more sophisticated analyses, particularly data mining because there are missing values or inconsistencies with the data. This is why most large companies extract and clean the data and put it into a separate data warehouse. Some dirty data may have illogical entries for gender, phone number, age, etc. Purchased data often have missing values or values that are out of an allowed range. More difficult to correct is data that has become inconsistent over time, because it does not reflect changes in the real world, such as new postal codes, area codes, addresses, income levels, etc. For a data-mining application, a few erroneous data points are worse than no data at all, because errors will introduce bias into the analysis. The level of granularity can also be an issue. The analysis can only go to the level of detail at which the data was originally collected. If a company wants to track sales of candy, they must decide what the granularity of the data will be for each item; i.e. will it be by the case, the box, or the individual candy bar. A data warehouse is a separate facility where the data is extracted, cleaned, and integrated into a single format so that it can support these more complex data-mining applications.

Diff: 3

Type: ES

Page Reference: 265

Topic: Q5

Skill: APPLIED

137) What is meant by supervised and unsupervised data mining? Describe one example of each technique.

Answer:

Data mining is the application of statistical techniques to find patterns and relationships among data and to classify and predict. It represents a convergence of research in fields such as computer science, statistics, and database design. With unsupervised data mining, analysts do not create a model or hypothesis before running the analysis. The data-mining techniques are simply run on the data and the results observed. Cluster analysis is the most common unsupervised technique, since it statistically organizes the data in clusters or groups that share similar characteristics. These may or may not be obvious to the users. Pharmaceutical companies will do this to examine which type of physicians have different prescribing patterns to help them market better to the different groups. With supervised data mining, a model of what factors the users want to test for is developed prior to the analysis. Regression analysis is the most common example of this. It tests a chosen set of variables to see how well they can act as predictors of future behaviour.

Diff: 3

Type: ES

Page Reference: 268

Topic: Q7

Skill: APPLIED

138) What are some of the obstacles to developing data-mining applications in the real world? Does this mean that data-mining projects should be avoided?

Answer:

In the real world there are many obstacles to doing a successful data-mining project that one might not expect. In the first place, it must be expected that all data sets will have dirty data in them. A lot of time will have to be spent just cleaning up data sets with missing values, values that are impossible or just unusable data. Data will also have to be integrated from multiple sources and this will take a substantial amount of time. A more severe problem is that the variables considered important at the beginning of the project may not be the ones that are considered crucial at the end of the project. At that point it may be too late to change variables and collect the data to test them. In the same vein, models can be created to fit any data set. This is called overfitting, and while accurate, these models will not be of much use in predicting new cases that come up. Then there is the problem with deploying the resulting analysis. Managers may have too high expectations about the results and if the model has a few bad results in a row purely due to random luck, the users may lose faith in it and abandon it too soon. All this just means that appropriate caution should be taken when developing and implementing them.

Diff: 3

Type: ES

Page Reference: 276

Topic: Q7

Skill: APPLIED