

## Review Questions

### Concepts

- Which data compression technique provides the most compression?
  - Lossy
  - Lossless
  - Filtering
  - Classification
- Which of the following best describes abstraction?
  - Adding complexity so the concept can apply to more uses
  - Simplifying complexity to make the concept more general
  - Combining procedures to make a new one
  - A set of instructions to do a task
- Why is cleaning data important?
  - It ensures incomplete data does not hide or skew results.
  - It removes bad data.
  - It repairs incomplete data.
  - All of the above.
- If analyzing data indicates a company should only hire people with a college degree because they stay at the company longer, what is this a potential indication of?
  - Good data management practices leading to good hiring practices
  - Frequency analysis to identify commonalities in the data
  - Bias in collecting the data
  - Data assessment and inquiry of hiring practices
- What is one way in which number systems are abstract?
  - The same amount can be represented by different number representations.
  - A number can only be represented by one number system.
  - Symbols can not be used to add, subtract, multiply, or divide them in their abstract form.
  - They use constants.
- Why is analyzing data with computers important?
  - To identify patterns that humans cannot see
  - To increase the viability of server farms
  - To verify existing solutions to problems
  - To test due diligence
- What is the number system used by computers?
  - Base 10 (decimal)
  - Base 8 (octal)
  - Base 2 (binary)
  - Base 16 (hexadecimal)
- When does an overflow error occur?
  - When the computer runs out of memory to store program instructions
  - When the flow of binary digits reaches a broken pathway and cannot arrive at its destination
  - When an integer requires more bits to represent it than the programming language provides
  - When the numerator in a calculation is larger than the denominator
- What is information about the author of a document considered to be?
  - Metadata
  - Content
  - Context
  - Mididata
- What is the amount of data compression an algorithm can produce reliant upon?
  - No repeating parts of the file being compressed
  - Several patterns in the data
  - A large file size
  - A small file size

11. How can an organization begin the process of analyzing data?
  - (A) By following an iterative development process
  - (B) By establishing measurements the data should show
  - (C) By developing hypotheses and questions to test
  - (D) By checking to see if the data matches previously collected data
12. What is  $214_{10}$  in binary?
  - (A) 11010100
  - (B) 11010110
  - (C) 11010111
  - (D) 01101011
13. What is a reason to perform additional research on correlations found through data analysis?
  - (A) There may not be an actual cause and effect relationship between the correlation variables.
  - (B) A single source may not provide enough data for a conclusion.
  - (C) To understand the relationship between the variables.
  - (D) All of the above.
14. The letter "M" is represented by 01001101 in binary. What is this in decimal?
  - (A) 414
  - (B) 76
  - (C) 77
  - (D) 1101
15. Metadata is used to
  - (A) provide updates to the data
  - (B) help find and organize data
  - (C) brand the data
  - (D) sort the data
16. Which number type is stored imprecisely in memory?
  - (A) Integers
  - (B) Numbers with decimals
  - (C) Both
  - (D) Neither
17. When is *sampling* needed?
  - (A) Sampling is used to store analog data.
  - (B) Sampling is used to approximate real-world data.
  - (C) Sampling is used when converting from digital data to analog data.
  - (D) Sampling is used when converting from analog data to digital data.
18. An example of metadata about sea turtle nests could be
  - (A) number of eggs in the nest
  - (B) location of the nest
  - (C) number of incubation days
  - (D) tracking number assigned to the nest
19. Being able to add or remove resources to store large datasets is called
  - (A) scalability
  - (B) filtering
  - (C) efficiency
  - (D) routing

### Application of Concepts

20. Data compression algorithms are used when the data
  - (A) needs to be shared with a large number of people
  - (B) is used for cryptography to keep data secure
  - (C) is too large to send in a timely manner
  - (D) needs to be sent a large physical distance away
21. How many more bits are available if you go from a 32-bit computer to a 64-bit machine?
  - (A) Twice as many
  - (B) 32 more
  - (C)  $2^{32}$  more
  - (D)  $32^2$  more

22. Your program is comparing temperatures to determine how many patients have a fever. Your selection statement's condition is not working correctly when the variable *patient\_temp* is 98.6. What could be one reason why?

IF (*patient\_temp* = 98.6)

- (A) The format is incorrect for this test.
- (B) A round-off error occurred.
- (C) An overflow error occurred.
- (D) The test condition is invalid.

23. Given a table of lunchroom leftovers, what can be determined from the data?

Date	Total Meals	Total Meals Left Over
1/29/18	800	25
1/30/18	750	5
1/31/18	800	42

- (A) Most popular items
- (B) Days of field trips when classes missed lunch
- (C) Days with high absenteeism
- (D) Amount of wasted budget dollars

24. A company purchases a large block of data from a social media site. If they want to analyze the data to learn more about potential customers, what techniques should they use?

- (A) Simulations to test different hypotheses about what data could be present
- (B) Data analysis to identify patterns and relationships in the data for further analysis
- (C) Maximization to get the highest return on their purchase of the data
- (D) Data processing to use the data with existing company software to see if it will run on their systems or if new ones will need to be developed

25. What order should these be in to go from smallest to largest?

1. Binary—01110111
2. Decimal—111
3. Binary—01111011

- (A) 1, 2, 3
- (B) 2, 1, 3
- (C) 3, 1, 2
- (D) 3, 2, 1

26. A magnet school wants to advertise its students' success taking AP exams to prospective families. What's the best method to share the summarized data?

- (A) Post an image of student results on social media sites.
- (B) Create an interactive pie chart that can drill down to topics and overall scores posted on the school's website.
- (C) Write a report for a marketing pamphlet.
- (D) Send an e-mail to families with middle-school-age children.

27. Which topic needs the use of programs to analyze the data to identify insights?

- (A) The average number of students who drive to school each day
- (B) The record of wins and losses for all sports teams under their current coaches compared to prior years' win/loss records
- (C) The number of library books in a school district that need to be replaced each year
- (D) The standardized test scores for current students compared to test scores across the country for the past decade

28. A school district wants to analyze the data about their high school graduates who attend local community colleges. What information can the school district obtain from the following data that the school has available?

Student name

Number of students in the graduating class by high school.

High school graduation year

Student self-reported plans for after high school

Total numbers from local community colleges of enrolled students and their high school

Total number of local community colleges students enrolled in each degree program

- (A) Average number of in-state students who enrolled in a local community college
- (B) If the number of local graduates who enroll in a local community college is increasing, decreasing, or stable
- (C) Number of local high school students who graduate from a local community college
- (D) Popular degree programs in the community colleges