Which statement describes SQLite?

**Question 1** 0 / 2 pts

Which statement describes SQLite?

○ It is an example of flat file database.

**Correct Answer** ○ It is an embedded SQL database engine.

**You Answered** ● It is a free version of RDBMS suitable for enterprises.

○ It is a fully functional RDBMS for distributed data processing.

Refer to curriculum topic: 1.3.2
SQLite is an embedded SQL database engine in that it does not follow the traditional client/server model like SQL RDBMS (relational database management system). SQLite reads and writes directly to ordinary disk files.

Which two statements describe characteristics of data in motion? (Choose two.)

**Question 2** 0 / 2 pts

Which two statements describe characteristics of data in motion? (Choose two.)

**Correct!** ☑ Its value changes over time.

**You Answered** ☑ It is stored at a central data center.

**Correct Answer** ☐ It requires real-time processing close to the source.

☐ It is the data in RAM during a data analysis process.

☐ It is stored in removable devices for easy transportation.

Refer to curriculum topic: 1.2.4
Data in motion describes the status of data to be distributed among different locations, the need of data to be analyzed close to the source, and how its value changes dynamically over time.

Which characteristic of big data describes different types of datasets that include both structured and unstructured data?

**Question 3**                                                    2 / 2 pts

Which characteristic of big data describes different types of datasets that include both structured and unstructured data?

- ○ velocity

- ○ volume

**Correct!**  ◉ variety

- ○ veracity

Refer to curriculum topic: 1.2.1
The characteristics of big data can be described in four Vs:

- **Volume** - the amount of data being transported and stored
- **Velocity** - the rate at which this data is generated
- **Variety** - the different types of data both structured and unstructured: video, audio, text
- **Veracity** - the process of preventing inaccurate data from spoiling the data sets

What has contributed to the exponential growth in data generation?

**Question 4**                                                    2 / 2 pts

What has contributed to the exponential growth in data generation?

**Correct!**  ◉ the increasing number of mobile devices

- ○ the increasing number of standalone devices

- ○ the increasing number of isolated software applications

- ○ the increasing number of physical installations for protecting environment facilities

Refer to curriculum topic: 1.1.2
An increased number of sensors and other end devices as well as mobile devices are contributing to an exponential growth in data generation.

How do sensors relate to Big Data?

**Question 5**                                                    0 / 2 pts

How do sensors relate to Big Data?

**You Answered**   ◉ They are types of multimedia applications that are sources for Big Data.

**Correct Answer**   ○ They are devices that collectively generate large amounts of data.

○ They are devices that can only be used with static data.

○ They produce structured data.

Refer to curriculum topic: 1.2.1
The use of sensors in IoT systems is growing exponentially. Each sensor has a multiplicative effect on the amount of data generated. Sensors are quickly becoming the greatest contributors toward Big Data.

A multi-campus school wants to perform analytics on classes held during the past 5 years. The school wants to know which classes filled up the quickest across all campuses and which classes filled up the quickest at each campus. The school also wants to know if there is a relationship between the number of passing students and the speed in which a class taught by a particular teacher fills. If the school could only choose one type of database to store the data on one server, which type would be best suited for this task?

## Question 6

A multi-campus school wants to perform analytics on classes held during the past 5 years. The school wants to know which classes filled up the quickest across all campuses and which classes filled up the quickest at each campus. The school also wants to know if there is a relationship between the number of passing students and the speed in which a class taught by a particular teacher fills. If the school could only choose one type of database to store the data on one server, which type would be best suited for this task?

**You Answered**

○ flat

○ local

○ Hadoop

**Correct Answer**

○ relational

Refer to curriculum topic: 1.3.2
A relational database, even though it has multiple, connected tables, can reside on one server and would be best for this type of data. A local database is typically used to collect and store local data, for example, a database of all movies and music for a particular family. A flat database would most likely not be used in a multi-location school to store student data such as this. Hadoop is best to use when distributing processing power across server clusters.

Which type of information is captured and stored as events happen?

## Question 7

**0 / 2 pts**

Which type of information is captured and stored as events happen?

**You Answered**

- ⦿ critical

- ○ analytical

- ○ comparative

**Correct Answer**

- ○ transactional

Refer to curriculum topic: 1.2.1
The two primary types of business information useful to a company are transactional information and analytical information. Transactional information is captured and stored as events happen. Transactional information can be used to analyze daily sales reports and production schedules to determine how much inventory to carry. Analytical information supports managerial analysis tasks like determining whether the organization should build a new manufacturing plant or hire additional sales personnel.

When is data considered to be information?

## Question 8

**0 / 2 pts**

When is data considered to be information?

**You Answered** ⊙ when it is stored

○ when it is recorded

**Correct Answer** ○ when it is processed and analyzed

○ when it is generated

> Refer to curriculum topic: 1.1.1
> Data that has been processed, organized, analyzed, or presented in a meaningful way becomes information.

Which method does openPDS use to protect user privacy of GPS records on a mobile device?

## Question 9

**0 / 2 pts**

Which method does openPDS use to protect user privacy of GPS records on a mobile device?

**You Answered** ⊙ requiring authentication to be completed first

○ encrypting the communication from data to the app

**Correct Answer** ○ providing answers to specific queries instead of raw data

○ removing identifiable personal information before sending data to the app

> Refer to curriculum topic: 1.2.2
> Using the SafeAnswers framework, openPDS provides only answers to specific queries and no raw data is sent. The calculation for the answer is done within the personal data store (PDS) of the user.

What are two examples of unstructured data? (Choose two.)

**Question 10**      1 / 2 pts

What are two examples of unstructured data? (Choose two.)

**Correct!**
- ☑ video content

- ☐ user account data

- ☐ SQL queries

**Correct Answer**
- ☐ blog entry

- ☐ customer account spreadsheet

Refer to curriculum topic: 1.2.3
Unstructured data is raw data, data that is not organized in a predefined way.
Examples of unstructured data would be contents of photos, audio, video, web pages, blogs, books, journals, and white papers.

Which term describes the growth rate of data in the IoT?

**Question 11**      0 / 2 pts

Which term describes the growth rate of data in the IoT?

**You Answered**
- ⦿ linear

- ○ cyclical

- ○ uniform

**Correct Answer**
- ○ exponential

Refer to curriculum topic: 1.1.2
Current trends and forecasts all indicate that today data grows exponentially.

What is a characteristic of structured data?

## Question 12

0 / 2 pts

What is a characteristic of structured data?

**You Answered**

⦿ Structured data is subject to intellectual property restrictions.

**Correct Answer**

○ It has a predefined organization.

○ It is raw data.

○ It generates new knowledge.

Refer to curriculum topic: 1.2.3
Structured data is data that is structured and can be entered, classified, and queried by a computer. Data that is found in databases and spreadsheets is an example of structured data.

In the data analysis process, which sequence depicts the work flow suitable for data at rest?

## Question 13

0 / 2 pts

In the data analysis process, which sequence depicts the work flow suitable for data at rest?

**You Answered**

- ⦿ act > analyze > store > notify
- ○ analyze > notify > act > store
- ○ notify > store > act > analyze

**Correct Answer**

- ○ store > analyze > notify > act

Refer to curriculum topic: 1.2.4
Data at rest is static data that is stored in a database first and then analyzed and interpreted. Data at rest follows the traditional analysis flow of store > analyze > notify > act. Once the data is analyzed, decision makers are notified and determine whether action is needed.

What is Hadoop?

## Question 14

**0 / 2 pts**

What is Hadoop?

**You Answered** ⦿ a method of preventing loops when analyzing Big Data

○ a groundbreaking method of moving large amounts of data through micro loops

**Correct Answer** ○ a framework that allows distributed processing of data across clusters of computers

○ a method of sharing data across multiple companies using computing resources housed within each respective company

Refer to curriculum topic: 1.3.2
Data management and analysis today are characterized by the use of flat file databases, relational database management system (RDBMS), and the Hadoop framework that allows distributed processing of data across clusters of computers using simple programming models.

What is true of Big Data in comparison to traditional data?

## Question 15

**0 / 2 pts**

What is true of Big Data in comparison to traditional data?

**You Answered** ○ Traditional data is represented through binary strings, whereas Big Data is represented through hexadecimal strings.

○ Both types of data require the same hardware for processing and storage.

**Correct Answer** ○ Big Data requires a different approach to analysis, computing, and storage mechanisms.

○ Big Data means that the data sets are being sent through the network in larger packets than the sets that contain legacy data.

What is a characteristic of open data?

**Question 16**                                                 2 / 2 pts

What is a characteristic of open data?

**Correct!**    ⦿ data that lacks intellectual property restrictions

○ data that lacks predefined organization

○ data that does not need to be stored

○ data that does not generate new knowledge

Refer to curriculum topic: 1.2.2
Open data is not protected by intellectual property restrictions and can be used
and redistributed without legal, technical, or social restrictions.

What is a purpose of applying data anonymization process to a data sets?

**Question 17**                                                 0 / 2 pts

What is a purpose of applying data anonymization process to a data sets?

**You Answered**    ⦿ to compress the data sets

○ to reduce the size of the data sets

**Correct Answer**    ○ to remove identifiable personal information
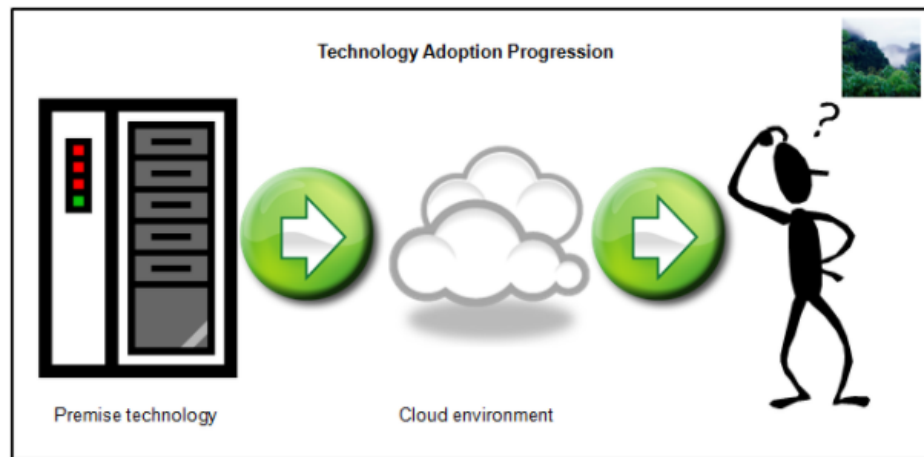
○ to adjust the value length of certain data fields

Refer to curriculum topic: 1.2.2
Data anonymization is a process of either encrypting or removing identifiable
personal information from data sets to achieve privacy protection.

Refer to the exhibit. To remain competitive, a company has progressed from on-premise
technology to the cloud environment. What technology environment would a manager need
to consider to accommodate long-term storage and immediate analysis of data in motion.

Technology Adoption Progression

Premise technology

Cloud environment

Refer to the exhibit. To remain competitive, a company has progressed from on-premise technology to the cloud environment. What technology environment would a manager need to consider to accommodate long-term storage and immediate analysis of data in motion.

**You Answered**

○ an analytic model

**Correct Answer**

○ a hybrid model

○ the fog model

○ a cloud model will accompish both requirements

○ on-premise clouds

What are two key components in creating data analysis tools from scratch? (Choose two.)

## Question 19

1 / 2 pts

What are two key components in creating data analysis tools from scratch? (Choose two.)

**Correct!**

☑ coding

**Correct Answer**

☐ modeling

☐ data sets

☐ performance

☐ program length

What is an example of data in motion?

## Question 20

**0 / 2 pts**

What is an example of data in motion?

**You Answered**

⊙ recording road traffic volumes and patterns for future highway planning

**Correct Answer**

○ medical information being transmitted from an ambulance to emergency department staff as a critically ill patient is being transported to the hospital

○ hourly weather information being collected in preparation for the next day weather forecast for a specific location

○ collecting sales and transaction records in preparation for a monthly sales report from sales consultants as they travel between customers

Refer to curriculum topic: 1.1.1
Data in motion is dynamic data that requires real-time processing before the data becomes obsolete. It represents the continuous interactions between people, processes, data and things. In this example the real-time medical information enables the emergency staff to be appropriately prepared before the patient arrives at the hospital.

Which statement describes the paradigm that is promoted in the Cisco Fog Computing Model?

## Question 21

**0 / 2 pts**

Which statement describes the paradigm that is promoted in the Cisco Fog Computing Model?

**You Answered** ⊙ All data analysis and decision making should take place near the data source.

**Correct Answer** ○ Some data analysis should take place at the edge of infrastructure rather than at a central location.

○ Data generated by edge devices should be sent to the nearest regional data analysis center for data aggregation.

○ Data collected at the edge of the infrastructure should be stored in a central data center for security and backup operations.

What are three examples of a NoSQL database? (Choose three.)

## Question 22

**0 / 2 pts**

What are three examples of a NoSQL database? (Choose three.)

**You Answered** ☑ Ceph

☐ HDFS

**Correct Answer** ☐ Redis

**Correct Answer** ☐ MongoDB

☐ GlusterFS

**Correct Answer** ☐ Apache Cassandra

Which task is performed within the data analysis step of the Data Analysis Lifecycle?

**Question 1**          2 / 2 pts

Which task is performed within the data analysis step of the Data Analysis Lifecycle?

**Correct!**

◉ testing the model on data to ensure reliability

○ incorporating the findings into corporate strategy

○ selecting a tool for data analysis

○ locating and gathering data

Place the six steps in the data analysis lifecyle in the correct order.

**Analyze the data.**

Step 6 ⌄

Correct Answer  Step 4

**Choose a model.**

Step 1 ⌄

Correct Answer  Step 3

**Gather the data.**

Step 2 ⌄

Correct Answer  Step 1

**Make decisions.**

Step 5 ⌄

Correct Answer  Step 6

**Prepare the data.**

Step 4 ⌄

Correct Answer  Step 2

**Present results.**

Step 4 ⌄

Correct Answer  Step 5

A college has deployed next-generation analytics. Which question would most likely be answered by next generation analytics that could not be answered by traditional analysis tools?

## Question 3

**0 / 2 pts**

A college has deployed next-generation analytics. Which question would most likely be answered by next generation analytics that could not be answered by traditional analysis tools?

**You Answered**

- ◉ Which class has the most sections available?

- ○ What are the 10 most expensive textbooks required in any class?

- ○ Which 100 classes fill the fastest once open enrollment is available for the general population?

- ○ Which teacher had the highest number of students enrolled in classes during the last year?

**Correct Answer**

- ○ Which class sections do students consider the most in the online environment without actually enrolling?

What two aspects of modern data analytics enhance data-driven decision making compared with traditional data analysis methods? (Choose two.)

## Question 4

**1 / 2 pts**

What two aspects of modern data analytics enhance data-driven decision making compared with traditional data analysis methods? (Choose two.)

**Correct!**

- ☑ real-time incoming data

- ☐ well-defined data structures

**Correct Answer**

- ☐ data from multiple sources

- ☐ data stored in a central data warehouse

- ☐ predominant use of descriptive analytical tools

In a data analysis project, which process gathers all grades for students enrolled in a college course?

## Question 5

0 / 2 pts

In a data analysis project, which process gathers all grades for students enrolled in a college course?

**You Answered**

- ⦿ action

**Correct Answer**

- ○ extract

- ○ load

- ○ transform

In the DIKW (data, information, knowledge, and wisdom) model, what is the definition of knowledge?

## Question 6

0 / 2 pts

In the DIKW (data, information, knowledge, and wisdom) model, what is the definition of knowledge?

**You Answered**

- ⦿ Knowledge is unstructured data.

**Correct Answer**

- ○ Knowledge is information that is understood and can be used to make decisions.

- ○ Knowledge is data that is structured in a useful way.

- ○ Knowledge is the understanding of how to apply data.

Refer to curriculum topic: 2.2.1
Knowledge is information that is understood and can be used to make decisions. Information is data that has been structured in a way that it is useful. Wisdom is the understanding of how to apply knowledge. Data is common raw values taken from measurements.

What is an example of data in the DIKW Pyramid model?

## Question 7

**0 / 2 pts**

What is an example of data in the DIKW Pyramid model?

○ data structured in a format that can be analyzed

**You Answered**  ◉ the application of knowledge

○ information that is understood and can be used immediately

**Correct Answer**  ○ facts gathered through observation

Which two tasks may be performed as part of the data transformation process? (Choose two.)

## Question 8

**1 / 2 pts**

Which two tasks may be performed as part of the data transformation process? (Choose two.)

**Correct Answer**  ☐ sorting the data

**Correct!**  ☑ applying validation rules to the data

☐ transferring the data into a database

☐ collecting and recording the data

☐ communicating results to decision-makers

Fill in the blank with one word.
Decision-makers today rely on data

## Question 9

**0 / 2 pts**

Fill in the blank with one word.

Decision-makers today rely on data ds to extract the required information from Big Data at the right time to make the right decision.

Answer 1:

**You Answered**

ds

**Correct Answer**    analytics

**Correct Answer**    analysis

Which statement is true concerning data analytics?

## Question 10

**0 / 2 pts**

Which statement is true concerning data analytics?

**You Answered**

○ Prescriptive analytics uses historical data to provide regular reports on events that have already happened.

**Correct Answer**

○ Predictive analytics uses simulation models and forecasting to suggest what could happen.

○ Descriptive analytics uses a feedback system to track the outcome of actions taken.

○ Descriptive analytics attempts to predict what will happen next.

What is true about Big Data and personal privacy?

**Question 11**                                                    0 / 2 pts

What is true about Big Data and personal privacy?

**You Answered**
◉ Privacy issues are not a major concern of Big Data.

○ Ethical issues about the use of Big Data are treated the same way around the world.

○ Most data scientists think that consumers will not have to worry about privacy issues because privacy is the responsibility of government.

**Correct Answer**
○ According to data scientists, the construction of an ethical framework would help the collection and use of Big Data.

What is a characteristic of the six-step Data Analysis Lifecycle methodology?

**Question 12**                                                    0 / 2 pts

What is a characteristic of the six-step Data Analysis Lifecycle methodology?

**Correct Answer**
○ Each step includes many tasks that must be completed before moving on to the next step.

**You Answered**
◉ It is the only methodology available for data analysis.

○ It was designed exclusively for use in a scientific environment.

○ The analysis process is very straightforward, so there are no iterations with previous steps.

What was the main purpose of the "Discover Your Internet Footprint" activity?

## Question 13

**0 / 2 pts**

What was the main purpose of the "Discover Your Internet Footprint" activity?

○ to show that only public information about a person can be found

**Correct Answer** → ○ to show that different types of information can be found about a person

**You Answered** →
◉

to show that only authorized government workers can find information about a person

○

to show that any information about a person can be found, but only if a person gives an authorization to do so

A large online ordering company collects the sales history of its customers. Using this historical data, the online company sends customized sales advertisements to a specific group of customers. Which data analytics method is being used?

## Question 14

**0 / 2 pts**

A large online ordering company collects the sales history of its customers. Using this historical data, the online company sends customized sales advertisements to a specific group of customers. Which data analytics method is being used?

○ indicative

○ predictive

**You Answered** → ◉ descriptive

**Correct Answer** → ○ prescriptive

An analyst is asked to review gathered data and modify any incompatible metric used by various sources of data. Which step in ETL does this represent?

## Question 15

**0 / 2 pts**

An analyst is asked to review gathered data and modify any incompatible metric used by various sources of data. Which step in ETL does this represent?

○ action

○ extract

**You Answered** ◉ load

**Correct Answer** ○ transform

In what way is the Data Analysis Lifecycle similar to using the scientific method?

## Question 16

**0 / 2 pts**

In what way is the Data Analysis Lifecycle similar to using the scientific method?

**Correct Answer** ○ They both begin with a question.

**You Answered** ◉ They both use the same process steps.

○ They both lead to the same results.

○ They both require a scientist to solve a problem.

A college uses a learning management system (LMS) to host lecture materials and provide access to assignments and assessments. A teacher uses the LMS to check activities completed by students in a particular course. Which data analytics method is being used?

## Question 17

**0 / 2 pts**

A college uses a learning management system (LMS) to host lecture materials and provide access to assignments and assessments. A teacher uses the LMS to check activities completed by students in a particular course. Which data analytics method is being used?

○ indicative

**You Answered** ◉ predictive

**Correct Answer** ○ descriptive

○ prescriptive

○ proactive

In what step of the Data Analysis Lifecycle are tools selected for performing data analysis?

## Question 18

**2 / 2 pts**

In what step of the Data Analysis Lifecycle are tools selected for performing data analysis?

**Correct!** ◉ choosing a model

○ gathering data

○ presenting the data

○ preparing the data

Fill in the blank.
Similar to web crawling, a data analyst can use web

## Question 19

0 / 2 pts

Fill in the blank.

Similar to web crawling, a data analyst can use web [sd] to automatically extract data from a specific web database and easily organize it.

---

**Answer 1:**

**You Answered**    sd

**Correct Answer**    scraping

**Correct Answer**    crawling

**Correct Answer**    crawler

**Correct Answer**    scrapping

Which data trend has seen decision making rely increasingly on data analytics?

## Question 20

2 / 2 pts

Which data trend has seen decision making rely increasingly on data analytics?

---

**Correct!**    ⦿ the increase in the amount of data

○ the security constraints of data

○ the lack of information about data

○ the accuracy of decision making

Match the variable with the description.

## Question 1                                      1 / 2 pts

Match the variable with the description.

**You Answered**

nominal                    qualitative values in order of ⌄

> **Correct Answer**    qualitative values based
>                       on identity of the object

**Correct!**

ordinal                    qualitative values in order of r ⌄

**Correct!**

continuous                 quantitative values with infini ⌄

**You Answered**

discrete                   qualitative values based on id ⌄

> **Correct Answer**    quantitative and finite set
>                       of values

In a graphical reppresentation of the distribution for a discrete variable, what does the X and Y axis reppresent? (Choose two.)

## Question 2                                      0 / 2 pts

In a graphical reppresentation of the distribution for a discrete variable, what does the X and Y axis reppresent? (Choose two.)

**You Answered**

☑ skew

**Correct Answer**

☐ variable

**Correct Answer**

☐ probability

☐ range

☐ standard deviation

What is a purpose of using the **head** Linux command?

## Question 3

**0 / 2 pts**

What is a purpose of using the **head** Linux command?

**You Answered**
- ◉ to determine the size of the file

- ○ to display the permissions on the file

**Correct Answer**
- ○ to verify that the file is available and preview the first few lines

- ○ to determine whether the file contains enough code to be executable

Match the variable with the type of value.

**Unanswered**

## Question 4

**0 / 2 pts**

Match the variable with the type of value.

**You Answered**

**ordinal** [ ▾ ]

**Correct Answer** student class rank

**You Answered**

**nominal** [ ▾ ]

**Correct Answer** eye color

**You Answered**

**continuous** [ ▾ ]

**Correct Answer** sales volume

**You Answered**

**discrete** [ ▾ ]

**Correct Answer** number of users

A data analyst studies two variables over a period. Data set A is (100, 90, 81, 73) and data set B is (3500, 3150, 2835, 2552). The Pearson r value between these two variables is 0.99. Which statement describes the correlation between them?

Which term is used to describe the difference between the highest and the lowest values for a variable?

What are two data structures in pandas? (Choose two.)

## Question 7

**0 / 2 pts**

What are two data structures in pandas? (Choose two.)

☐ text

☐ series

☐ image

☐ number

☐ dataframe

Fill in the blank.
A data analyst uses

## Question 8

**0 / 2 pts**

Fill in the blank.

A data analyst uses [        ] statistics to make generalizations and

predictions about a population.

**Answer 1:**

(You left this blank)

inferential

inferential statistics

inferential statistic

inferental

What relationship is an example of causation?

**Question 9**

0 / 2 pts

What relationship is an example of causation?

○ a relationship between sensors and actuators in IoT

○ a relationship in which two things change at a similar rate

○

a relationship between a computer operating system and one or more programming languages

○ a relationship when a change in one thing directly results in something else changing

Which term is used to describe a graph of distribution where the peak is left of center or right of center?

**Question 10**

0 / 2 pts

Which term is used to describe a graph of distribution where the peak is left of center or right of center?

○ skewed

○ normal

○ bimodal

○ symmetric

What is a measurement of central tendency that is also known as the average?

**Question 11** 0 / 2 pts

What is a measurement of central tendency that is also known as the average?

○ mode

○ mean

○ median

○ Person r

Which term is used to describe the right and left ends of a distribution graph?

**Question 12** 0 / 2 pts

Which term is used to describe the right and left ends of a distribution graph?

○ tail

○ end

○ peak

○ axis

What term describes the key characteristics that are observed or measured as part of an experiment or analysis?

**Question 13** 0 / 2 pts

What term describes the key characteristics that are observed or measured as part of an experiment or analysis?

○ variable

○ data set

○ statistics

○ population

Fill in the blank.
A representative sample is chosen from a

**Question 14**                                           0 / 2 pts

Fill in the blank.

A representative sample is chosen from a ⬚ that is a group of similar

entities that share common characteristics or traits.

**Answer 1:**

(You left this blank)

population

a population

the population

A data analyst performs a correlation analysis between two quantities. The result of the analysis is an r value of 0. What does this mean?

**Question 15**                                           0 / 2 pts

A data analyst performs a correlation analysis between two quantities. The result of the analysis is an r value of 0. What does this mean?

○ The analysis failed.

○ The two variables have no relationship.

○ Neither variables changed during the period.

○ When one variable increases its value, the other variable reduces its value.

Which file type is most similar to a dataframe in pandas?

**Question 16** 0 / 2 pts

Which file type is most similar to a dataframe in pandas?

○ a PDF file

○ a spreadsheet

○ a word processing file

○ a multilayer image file

Which statement is an accurate description of discrete variables?

**Question 17** 0 / 2 pts

Which statement is an accurate description of discrete variables?

○ Discrete variables are qualitative and consist of two or more categories in which order matters.

○ They are quantitative with a continuous range of values.

○ They are quantitative with a finite set of values.

○ They are qualitative and consist of two or more categories of values in which order does not matter.

What category of variables includes nominal or ordinal variables?

**Question 18**                                                    0 / 2 pts

What category of variables includes nominal or ordinal variables?

○ numerical

**Correct Answer**    ○ categorical

○ ratio

○ interval

Fill in the blank.
Continuous variables can take on an infinite number of values whereas

**Question 19**                                                    0 / 2 pts

Fill in the blank.
Continuous variables can take on an infinite number of values whereas

[                    ] variables can only take a value from a finite number of values.

**Answer 1:**

**You Answered**    (You left this blank)

**Correct Answer**    discrete

Fill in the blank.
The  is a measure of central tendency, which is the middle value in a data set.

## Question 20

**0 / 2 pts**

Fill in the blank.

The [                ] is a measure of central tendency, which is the middle value in a data set.

**Answer 1:**

(You left this blank)

median

the median

medain

What is the most commonly used statistical method for analyzing data?

## Question 1

**0 / 2 pts**

What is the most commonly used statistical method for analyzing data?

○ mean estimation

○ regression analysis

○ sample proportion

○ mean analysis

What type of error has occurred when a data scientist records a measurement incorrectly after viewing the correct value on the measuring device?

**Question 2**

0 / 2 pts

What type of error has occurred when a data scientist records a measurement incorrectly after viewing the correct value on the measuring device?

○ instrumental

**Correct Answer**   ○ gross

○ random

○ systematic

Fill in the blank.
Random errors commonly have a normal distribution around the

**Question 3**

0 / 2 pts

Fill in the blank.

Random errors commonly have a normal distribution around the [        ]

of an observation.

**Answer 1:**

**You Answered**   (You left this blank)

**Correct Answer**   mean

**Correct Answer**   mean value

**Correct Answer**   average

When a number of items are grouped together, which type of machine learning algorithm can determine which items in the group predict the presence of other items?

**Question 4**                                                 0 / 2 pts

When a number of items are grouped together, which type of machine learning
algorithm can determine which items in the group predict the presence of other items?

○ association

○ clustering

○ classification

○ regression

Which type of machine learning algorithm uses data sets verified by experts as its learning
basis?

**Question 5**                                                 0 / 2 pts

Which type of machine learning algorithm uses data sets verified by experts as its
learning basis?

○ clustering

○ association

○ routing

○ supervised

In a linear regression, which variable is also known as the target or response variable?

**Question 6**                                                 0 / 2 pts

In a linear regression, which variable is also known as the target or response variable?

○ predictor

○ first

○ dependent

○ independent

Fill in the blank.
A research scientist would perform an

**Question 7**

0 / 2 pts

Fill in the blank.

A research scientist would perform an [          ] validity test to establish if the original experiment included all the steps of the scientific method.

**Answer 1:**

(You left this blank)

internal

Which type of reliability would a scientist measure if the scientist wants to examine the variation between exam scores for a person taking a single test multiple times?

**Question 8**

0 / 2 pts

Which type of reliability would a scientist measure if the scientist wants to examine the variation between exam scores for a person taking a single test multiple times?

○ inter-rater

○ parallel-forms

○ test-retest

○ internal consistency

Which type of regression analysis is often used to model variables that have an exponential relationship?

**Question 9**                                                           0 / 2 pts

Which type of regression analysis is often used to model variables that have an exponential relationship?

○ nonlinear

○ mean

○ median

○ polynomial

What is the goal of linear regression?

**Question 10**                                                          0 / 2 pts

What is the goal of linear regression?

○ to provide a summary of the data

○ to provide a formula that does not require validation

○ to construct a flow chart

○ to compute a line the interpolates the data, and which can be expressed as a weighted average of the predictor variables and any other function

If the results of a study do not align with previous studies, what question should an evaluator ask?

**Question 11**                                                    0 / 2 pts

If the results of a study do not align with previous studies, what question should an evaluator ask?

○ Are there any experts that disagree with the findings?

○ Who paid for the research study?

○ Did the study have an appropriate sample size?

**Correct Answer**   ○ Can the study be replicated to verify the findings?

Refer to the exhibit. What is the purpose of the blue sphere?

Refer to the exhibit. What is the purpose of the blue sphere?

○ to categorize historical data

○ to measure true error

**Correct Answer**   ○ to indicate data clusters

○ to display the mean

When is an experiment considered reliable?

**Question 13**                                                    0 / 2 pts

When is an experiment considered reliable?

○ if someone else can modify the experiment and achieve the same conclusions

○ if someone else can modify the experiment and achieve similar conclusions

**Correct Answer**   ○ if someone else can repeat the experiment and find the same conclusion

○ if someone else can repeat the experiment and find different conclusions

What are two types of supervised machine learning algorithms? (Choose two.)

**Question 14**                                    0 / 2 pts

What are two types of supervised machine learning algorithms? (Choose two.)

☐ mode

☐ mean

☐ clustering

☐ association

☐ regression

☐ classification

Which type of information can distort the results of an analysis and careful consideration should be given to their removal from a data set?

**Question 15**                                    0 / 2 pts

Which type of information can distort the results of an analysis and careful consideration should be given to their removal from a data set?

○ outliers

○ z-axis

○ azimuth

○ units of measurement

When you follow the scientific method, which step would occur after testing the hypotheses through experimentation?

**Question 16**

0 / 2 pts

When you follow the scientific method, which step would occur after testing the hypotheses through experimentation?

Correct Answer

○ Analyze data from an experiment to draw a conclusion.

○ Ask a question about an observation.

○ Communicate the results of the process.

○ Perform research.

learning is often used by Big Data analytics to model existing data in order to predict future outcomes.

Unanswered

**Question 17**

0 / 2 pts

Fill in the blank.

[          ] learning is often used by Big Data analytics to model existing data in order to predict future outcomes.

Answer 1:

You Answered

(You left this blank)

Correct Answer

machine

Correct Answer

machine learning

A researcher has measured the reliability of a test using the parallel-forms method. What is the expected result of this measurement?

## Question 18

A researcher has measured the reliability of a test using the parallel-forms method. What is the expected result of this measurement?

○ How similar are the scores of two different tests that are created from the same content domain?

○ How similarly do different people score on the same test?

○ How much variation exists between scores for the same person taking a test multiple times?

○ What is the variation of scores for different items in the same test?

Match the term to the definition. (Not all options are used.)

## Question 1

Match the term to the definition. (Not all options are used.)

**presented to support a conclusion in explantory analysis** [ dropdown ]

evidence

**a conclusion not justified by the premise of the argument** [ dropdown ]

fallacy

**used to make valid statements about a proposition** [ dropdown ]

logic

What is the purpose of Pyplot?

**Question 2**                                    0 / 2 pts

What is the purpose of Pyplot?

○ to add plotting capabilities to Microsoft Word

○ to analyze variables separately as scatter plots and as pie charts

○ to create and customize a plot using a collection of style functions

○ to analyze a scatter plot using a specific statistical algorithm that shows trends

Which type of chart displays comparisons of values of specific data points in a horizontal position?

**Question 3**                                    0 / 2 pts

Which type of chart displays comparisons of values of specific data points in a horizontal position?

○ line

○ bar

○ pie

○ scatter

What is the business value of data?

**Question 4**                                    0 / 2 pts

What is the business value of data?

○ It has no value to the departments within an organization.

○ It has equal value to all departments within the organization.

○ It has different value to different departments within the organization.

○ It has equal value to all organizations.

What is a characteristic of the explanatory analysis process?

**Question 5** 0 / 2 pts

What is a characteristic of the explanatory analysis process?

○ produces summaries of data

**Correct Answer** ○ tells a story about the data

○ makes valid statements about a proposition

○ uses statements of truth to arrive at a conclusion

Which term is used to describe the process of making inferences about a population based on attributes of a sample?

**Question 6** 0 / 2 pts

Which term is used to describe the process of making inferences about a population based on attributes of a sample?

○ deductive reasoning

**Correct Answer** ○ inductive reasoning

○ explanatory analysis

○ exploratory analysis

Which type of chart is useful for identifying clustering and outliers in data?

**Question 7** 0 / 2 pts

Which type of chart is useful for identifying clustering and outliers in data?

**Correct Answer** ○ scatter plot

○ bar

○ line

○ pie

Fill in the blank.
Telling a story meant to persuade an audience to adopt a point of view is known as

**You Answered**    (You left this blank)

**Correct Answer**    explanatory analysis

**Correct Answer**    explanatory

**Correct Answer**    explainatory

**Correct Answer**    explanitory

**Correct Answer**    explainatory analysis

**Correct Answer**    explanitory analysis

Which application should be used to create a Pyplot customized style sheet (CSS)?

**Unanswered** | **Question 9** | **0 / 2 pts**

○ Microsoft Excel

○ Microsoft Visio

○ Plotly

**Correct Answer**    ○ text editor

True or False?
A user created a custom style sheet to be used with Pyplot and saved the file in the user home directory. When the style sheet is being referenced in code, the path information for the location of the file must be specified.

**Question 10**

True or False?
A user created a custom style sheet to be used with Pyplot and saved the file in the user home directory. When the style sheet is being referenced in code, the path information for the location of the file must be specified.

○ true

○ false

Fill in the blank.
Using facts or premises to arrive at a conclusion is the process of

**Question 11**

Fill in the blank.

Using facts or premises to arrive at a conclusion is the process of [          ]

reasoning.

**Answer 1:**

(You left this blank)

deductive

deductive reasoning

True or False?
In the story telling step of data analysis, plenty of time should be spent on explaining the entire data-analysis process.

**Question 12**

True or False?
In the story telling step of data analysis, plenty of time should be spent on explaining the entire data-analysis process.

○ true

○ false

Which term is used to describe the process of discovery that may lead to the creation of a hypothesis?

In the story telling step of data analysis, how much time should be spent on explaining the data?

While preparing a presentation, which question would be important in understanding "who" is the audience?

**Question 15** 0 / 2 pts

While preparing a presentation, which question would be important in understanding "who" is the audience?

○ Will the presentation be recorded?

○ What is the possible reaction of the audience?

○ Is the presentation delivered virtually?

○ Are there any security considerations?

True or False?
A drawback to Plotly is that files cannot be exported.

**Question 16** 0 / 2 pts

True or False?
A drawback to Plotly is that files cannot be exported.

○ true

○ false

What is the purpose of Plotly?

**Question 17** 0 / 2 pts

What is the purpose of Plotly?

○ to allow the online creation of data visualizations

○ to add plotting capabilities to Microsoft Excel

○ to analyze variables stored on a Jupyter notebook

○ to analyze a scatter plot using a specific statistical algorithm that shows trends

When you are preparing to give a presentation, which question would be important to ask yourself in order to understand where the audience is located?

**Question 18**                                                      0 / 2 pts

When you are preparing to give a presentation, which question would be important to ask yourself in order to understand where the audience is located?

○ What is the technical knowledge of the participants?

○ Will the presentation be shared?

○ Are there any security considerations?

**Correct Answer**   ○ Will the presentation be delivered virtually?

For preparing a presentation to a targeted audience, why is the "when" of the audience important to consider?

**Question 19**                                                      0 / 2 pts

For preparing a presentation to a targeted audience, why is the "when" of the audience important to consider?

**Correct Answer**   ○ The information may be time sensitive.

○ The presentation may be shared with others.

○ The motivation of the participants may need to be determined.

○ Some participants may be online and some attending in person.

Match the type of question to the example.

## Question 20

0 / 2 pts

Match the type of question to the example.

**Who is the audience?**

[ dropdown ]

What is the motivation of the listener?

**Where is the audience?**

[ dropdown ]

Is the audience online or in-person?

**When is the audience?**

[ dropdown ]

Is recording of the presentation allowed?

Which type of hypervisor would most likely be used in a data center?

## Question 1

0 / 2 pts

Which type of hypervisor would most likely be used in a data center?

○ Type 1

○ Type 2

○ Nexus

○ Hadoop

Fill in the blank.
The use of  logs makes Kafka different from traditional message brokers.

## Question 2

0 / 2 pts

Fill in the blank.

The use of [                    ] logs makes Kafka different from traditional message brokers.

**Answer 1:**

(You left this blank)

transaction

transaction log

transaction logs

Match the big data tool to the description.

**Question 3**

Match the big data tool to the description.

**Cassandra**

| ▾ |

Correct Answer ⟩ open-source NoSQL distributed database management system

**Kafka**

| ▾ |

Correct Answer ⟩ messaging system that uses transaction logs

**Spark**

| ▾ |

Correct Answer ⟩ open source distributed data processing engine

**HDFS**

| ▾ |

Correct Answer ⟩ a system that uses a master-slave architecture

Why can it be more beneficial to use Spark than MapReduce when creating a big data solution?

**Question 4**                                    0 / 2 pts

Why can it be more beneficial to use Spark than MapReduce when creating a big data solution?

○ batch processing

○ built-in machine learning library

○ data stored on disk

○ parallelizing algorithms

Fill in the blank.
The big data pipeline includes three components: data processing, data storage, and data

**Question 5**                                    0 / 2 pts

Fill in the blank.
The big data pipeline includes three components: data processing, data storage, and data [    ] .

**Answer 1:**

(You left this blank)

ingestion

What are two data storage problems with big data? (Choose two.)

**Question 6**                                                      0 / 2 pts

What are two data storage problems with big data? (Choose two.)

☐ data management

☐ increased maintenance cost

☐ unstructured data

☐ application performance

☐ loss of predictive analysis

Which transferring methodology do traditional message brokers use?

**Question 7**                                                      0 / 2 pts

Which transferring methodology do traditional message brokers use?

○ publish-subscribe

○ transaction logging

○ client-server

○ real-time

Match the cloud computing service to the description. (Not all options are used.)

**Question 8** 0 / 2 pts

Match the cloud computing service to the description. (Not all options are used.)

**virtual desktop environment**

**Correct Answer** SaaS

**development tools**

**Correct Answer** PaaS

**network equipment**

**Correct Answer** IaaS

Which architecture is used by HDFS?

**Question 9** 0 / 2 pts

Which architecture is used by HDFS?

○ peer-to-peer

**Correct Answer** ○ master-slave

○ client/server

○ stand alone

Match the job in a data center to the description.

## Question 10

0 / 2 pts

Match the job in a data center to the description.

**data analyst**

| ⌄ |

**Correct Answer** — Query and process data, provide reports, and summarize and visualize data.

**data scientist**

| ⌄ |

**Correct Answer** — Take raw data and turn it into meaningful information. Apply statistics, machine learning, and analytic approaches to answer critical business questions.

**data engineer**

| ⌄ |

**data engineer**

Create the infrastructure that supports big data. Design and build the platform on which all of this data is stored and processed.

**business analyst**

Study a business or an industry and then formulate a specific question. Work with company stakeholders to determine the issue of concern.

Which solution improves web response time by deploying multiple web servers and DNS servers?

## Question 11

0 / 2 pts

Which solution improves web response time by deploying multiple web servers and DNS servers?

○ sharding

○ memcaching

○ load balancing

○ distributed databases

What is the purpose of device virtualization?

**Question 12**　　　　　　　　　　　　　　　　　　　　　　　0 / 2 pts

What is the purpose of device virtualization?

○ It allows the physical layer to interface with another device through multiple virtual connections.

○ It allows multiple physical PCs to be linked as one virtual PC.

**Correct Answer**　○ It allows multiple OSs to run on one physical device.

○ It allows a networking device to manage multiple virtual local area networks.

Which function is provided by MapReduce to Hadoop?

**Question 13**　　　　　　　　　　　　　　　　　　　　　　　0 / 2 pts

Which function is provided by MapReduce to Hadoop?

○ storage expansion

○ hypervisor security

**Correct Answer**　○ distributed processing

○ database management

What is hyperjacking?

**Question 14**　　　　　　　　　　　　　　　　　　　　　　　0 / 2 pts

What is hyperjacking?

**Correct Answer**　○ taking over a virtual machine hypervisor as part of a data center attack

○ overclocking the mesh network which connects the data center servers

○ using multiple virtual machines as one large Hadoop server running HDFS

○ using processors from multiple computers to increase data processing power

Fill in the blank.

In the Lambda architecture, the layer is the data at rest layer and in this layer, accuracy is more important than speed.

A cloud service provider is considering building a data center on the coast of southern California. They have found a pre-existing building with large windows facing the ocean. What are two factors the provider should consider? (Choose two.)

Which statement describes the term containers in virtualization technology?

## Question 17

**0 / 2 pts**

Which statement describes the term containers in virtualization technology?

○ a group of VMs with identical OS and applications

○ a subsection of a virtualization environment that contains one or more VMs

○

**a virtual area with multiple independent applications sharing the host OS and hardware**

○

isolated areas of a virtualization environment, where each area is administered by a customer

Which scenario is suitable for fog computing?

## Question 18

**0 / 2 pts**

Which scenario is suitable for fog computing?

○ Multiple sensors are used to collect national climate information.

○ Sales data of a national department store are collected and analyzed.

○ Multiple motion sensors are implemented at major street intersections in a city.

○

Orders on a major online store are processed and fulfilled from multiple warehouse locations.

Which file system does Hadoop use?

**Question 19**                                      0 / 2 pts

Which file system does Hadoop use?

○ FAT32

**Correct Answer**    ○ HDFS

○ NTFS

○ SAS

What are two benefits to an organization if it rented data center services from either a co-location facility or a cloud service provider? (Choose two.)

**Question 20**                                      0 / 2 pts

What are two benefits to an organization if it rented data center services from either a co-location facility or a cloud service provider? (Choose two.)

☐ The organization will reduce the WAN connection cost.

☐ The internal IT staff will get training on data center operations.

☐ Rented data centers are usually close to the organization for easy access.

**Correct Answer**    ☐ The organization pays a smaller up-front cost and requires fewer internal IT staff.

**Correct Answer**    ☐
Rented data centers and equipment are managed by highly experienced professionals 24/7.