

Question 1

In the modern data ecosystem, what is one of the key challenges when organizing data within a data repository?

Scalability

Security

Replicability

Marketability

Correct

When organizing data, one of the key challenges is working with data repositories that provide high availability, flexibility, accessibility, and security.

Question 2

In the data ecosystem, what are data engineers responsible for?

Converting raw data from disparate sources into usable data

Providing business solutions based on insights and predictions

Using the data to generate insights and identify patterns

Creating predictive models based on the data

Correct

Data engineers collect data from disparate sources. They clean and prepare this data, then store and manage it. Finally, they make the data accessible for business applications and other stakeholders in the data ecosystem.

Question 3

Which of the following data sources contain structured data?

Tweets

Emails

Spreadsheets

Sensors

Correct

Spreadsheets are an example of structured data since data within a spreadsheet has a well-defined structure that can be represented in a tabular manner with rows and columns.

Question 4

Unix/Linux shell is ideal for which of the following tasks?

Automating repetitive tasks that are time-consuming to execute one line at a time

Developing statistical software and performing data analytics

Accessing information from relational databases

Creating mobile and desktop web applications quickly

Correct

Unix/Linux shell is most useful for repetitive tasks that may be time-consuming to execute by typing one line at a time.

Question 5

Document-based databases are ideal for storing which type of data?

User preferences

Social network information

Weather data

Medical records

Correct

Document-based databases are preferable for eCommerce platforms, medical records storage, CRM platforms, and analytics platforms.

Question 6

What is the relationship between data pipelines and data integration?

A data pipeline is a subprocess within data integration.

A data pipeline is a process used to perform data integration.

Data integration is a process that functions separately from the data pipeline.

Data integration is a process used to create a data pipeline.

Correct

A data pipeline is a process used to perform data integration, while ETL is a process within data integration.

Question 7

In a data platform architecture, the data analysis and user interface layer performs which of the following tasks?

Maintains information about the data

Supports querying tools and programming languages

Delivers data to data consumers

Makes data available for processing

Correct

The data analysis and user interface layer deliver processed data to data consumers.

Question 8

When using SQL, which of the following functions can you use to count the number of unique values in a column?

PRIMARY KEY

COUNT

DISTINCT

UNIQUE

Correct

Use the DISTINCT function to count the number of unique, or distinct, values in a column.

Question 9

At which stages of the data lifecycle must you maintain an auditable trail of personal data?

Acquisition, processing, storage, and disposal

Acquisition, processing, storage, sharing, and retention

Storage, sharing, and retention

Acquisition, processing, storage, sharing, retention, and disposal

Question 10

Which of the following is a mainstay for your career in data engineering?

Presentation experience

Academic degree

Technical abilities

Professional certification

Correct

You can start a career in data engineering through various paths. However, your technical abilities will be your mainstay in this profession, along with a broad understanding of the scope and use of data in business.