



CompTIA Data+ DA0-002 Practice Questions

Shared by Black on 13-08-2026

For More Free Questions and Preparation Resources

Check the Links on Last Page



Question 1

Question Type: MultipleChoice

Which of the following explains the purpose of UAT?

Options:

- A- To begin the software application development process to enhance user experience
- B- To ensure all parts of the software application work together after each sprint
- C- To review software application crashes, create patches, and deploy to users
- D- To validate and verify that a software application meets the needs and requirements of users

Answer:

D

Explanation:

This question is related to the Data Governance domain of DA0-002, which includes understanding processes like User Acceptance Testing (UAT) to ensure data-related applications meet governance and quality standards. UAT is a critical step in ensuring software aligns with user needs and organizational requirements.

To begin the software application development process to enhance user experience (Option A): UAT occurs near the end of development, not at the beginning.

To ensure all parts of the software application work together after each sprint (Option B): This describes integration testing, not UAT, which focuses on user validation.

To review software application crashes, create patches, and deploy to users (Option C): This refers to post-deployment maintenance, not UAT.

To validate and verify that a software application meets the needs and requirements of users (Option D): UAT is specifically designed to ensure the software meets user requirements and functions as intended in a real-world scenario, aligning with governance standards for quality.

The DA0-002 Data Governance domain emphasizes 'data quality control concepts' (similar to DA0-001, web ID: 1), which include ensuring applications meet user needs through processes like UAT.

=====

Question 2

Question Type: MultipleChoice

Which option best supports capabilities such as automatic versioning, corruption checks, KPIs, and user authentication?

Options:

- A- Notebook
- B- REST API
- C- Pipeline
- D- Source control

Answer:

D

Explanation:

This question falls under the Data Governance domain, focusing on tools that support data management and quality control features. The task is to identify a tool with capabilities like versioning, corruption checks, KPIs, and authentication.

Notebook (Option A): Notebooks (e.g., Jupyter) are for data analysis and coding but don't inherently support versioning, corruption checks, or authentication.

REST API (Option B): REST APIs enable data access but don't provide versioning or corruption checks as a primary function.

Pipeline (Option C): Data pipelines automate data workflows but don't typically include versioning or authentication.

Source control (Option D): Source control systems (e.g., Git) support automatic versioning (tracking changes), corruption checks (integrity verification), KPIs (e.g., commit frequency), and user authentication (access control), making this the best fit.

The DA0-002 Data Governance domain includes 'data quality control concepts,' and source control systems provide the listed capabilities to ensure data integrity and security.

=====

Question 3

Question Type: MultipleChoice

Which of the following AI types is the best option for time-series forecasting?

Options:

- A- Generative AI
- B- Foundational models
- C- Natural language processing
- D- Robotic process automation

P2P
exams

Answer:

B

Explanation:

Foundational models are large AI models trained on vast amounts of data, often exhibiting strong generalization capabilities. While not specifically architected for time-series, their ability to learn complex patterns could potentially be leveraged for forecasting tasks through fine-tuning or specialized architectures built upon them.

In reality, the best AI types specifically designed for time-series forecasting include:

Recurrent Neural Networks (RNNs), especially LSTMs and GRUs: These architectures are designed to handle sequential data and capture temporal dependencies.

Transformer Networks: Originally developed for NLP, Transformers have shown remarkable success in time-series forecasting due to their ability to capture long-range dependencies.

Traditional statistical models: ARIMA, Exponential Smoothing, and other statistical methods remain powerful and interpretable options for time-series analysis.

Therefore, while 'foundational models' have some potential, it's important to understand that they aren't the primary or specifically designed AI type for time-series forecasting.

Question 4

Question Type: MultipleChoice

A company reports on seven years of data in a sales dashboard. The dashboard pulls from a sales database that has 30 years of data.

a. The dashboard performance is slow. Which of the following is the best way to improve the dashboard's performance?

Options:

- A- Performing a code review
- B- Checking network connectivity
- C- Filtering to include only relevant data
- D- Adding more RAM and rerunning

Answer:

C

Explanation:

This question falls under the Data Governance domain, focusing on optimizing data quality and performance in dashboards. The dashboard is slow because it pulls from a large database (30 years) but only needs seven years of data.

Performing a code review (Option A): A code review might identify inefficiencies, but it's not the most direct solution for this scenario.

Checking network connectivity (Option B): Network issues might cause delays, but the primary issue is the data volume, not connectivity.

Filtering to include only relevant data (Option C): Filtering the data to include only the last seven years reduces the dataset size, directly improving performance by minimizing the data processed.

Adding more RAM and rerunning (Option D): Adding RAM might help, but it's a hardware solution that doesn't address the root cause of excessive data.

The DA0-002 Data Governance domain includes 'data quality control concepts,' such as optimizing performance by filtering data to improve efficiency.

=====

Question 5

Question Type: MultipleChoice

Which of the following data repositories stores unstructured and structured data?

Options:

- A- Data store
- B- Data silo
- C- Data mart
- D- Data lake

Answer:

D

Explanation:

This question falls under the Data Concepts and Environments domain of CompTIA Data+ DA0-002, which involves understanding different types of data repositories and their characteristics. The task is to identify a repository that can store both unstructured and structured data.

Data store (Option A): A data store is a general term for any data repository, but it's not specific enough to confirm it stores both unstructured and structured data.

Data silo (Option B): A data silo is an isolated data repository, often structured, and not typically designed for unstructured data.

Data mart (Option C): A data mart is a subset of a data warehouse, focused on structured data for specific business areas, not unstructured data.

Data lake (Option D): A data lake is a centralized repository that stores raw data in its native format, including both structured (e.g., tables) and unstructured (e.g., text, images) data, making it the correct choice.

The DA0-002 Data Concepts and Environments domain includes understanding 'different types of databases and data repositories,' and a data lake is specifically designed to handle both unstructured and structured data.

=====

Question 6

Question Type: MultipleChoice

A manager wants to use the information in a recurring report on incomplete timesheets for the prior month to guide employee outreach. Which of the following report types is the best for this task?

Options:

- A- Summary
- B- Infographic
- C- Snapshot
- D- Ad hoc

Answer:

A

Explanation:

This question is part of the Visualization and Reporting domain, focusing on selecting the appropriate report type for a specific purpose. The manager needs a recurring report on incomplete timesheets to guide outreach, which requires a concise, data-driven report.

Summary (Option A): A summary report aggregates data (e.g., total incomplete timesheets per employee) and presents it concisely, making it ideal for recurring use to guide decisions like employee outreach.

Infographic (Option B): Infographics are visual representations for broad audiences, not typically used for recurring, detailed employee outreach tasks.

Snapshot (Option C): A snapshot report captures data at a specific point in time, but it's not ideal for recurring analysis of trends or aggregates.

Ad hoc (Option D): Ad hoc reports are one-time, on-demand reports, not suitable for recurring needs.

The DA0-002 Visualization and Reporting domain includes 'the appropriate visualization in the form of a report' , and a summary report best fits the need for recurring, actionable data.

=====

Question 7

Question Type: MultipleChoice

Which of the following best enables the retrieval and manipulation of data that is stored in a relational database?

Options:

- A- XML
- B- SQL
- C- Excel
- D- JavaScript

Answer:

B

Explanation:

This question pertains to the Data Concepts and Environments domain, focusing on tools for interacting with relational databases. The task is to identify the best method for retrieving and manipulating data.

XML (Option A): XML is a data format, not a language for retrieving or manipulating database data.

SQL (Option B): SQL (Structured Query Language) is specifically designed for querying and manipulating data in relational databases (e.g., SELECT, UPDATE), making it the best choice.

Excel (Option C): Excel can analyze data but isn't designed for direct database manipulation.

JavaScript (Option D): JavaScript is a programming language for web development, not optimized for relational database operations.

The DA0-002 Data Concepts and Environments domain includes understanding 'different types of databases,' and SQL is the standard language for relational database operations.

Question 8

Question Type: MultipleChoice

Which of the following data repositories stores unformatted data in its original, raw form?

Options:

- A- Data warehouse
- B- Data silo
- C- Data mart
- D- Data lake

Answer:

D

Explanation:

This question pertains to the Data Concepts and Environments domain, focusing on data repositories. The task is to identify a repository that stores raw, unformatted data.

Data warehouse (Option A): A data warehouse stores structured, processed data in a predefined schema, not raw data.

Data silo (Option B): A data silo is an isolated repository, often structured, not designed for raw data storage.

Data mart (Option C): A data mart is a subset of a data warehouse, also storing structured data.

Data lake (Option D): A data lake stores raw, unformatted data in its original format (structured, semi-structured, or unstructured), making it the correct choice.

The DA0-002 Data Concepts and Environments domain includes understanding 'different types of databases and data repositories,' and a data lake is designed for raw data storage.

=====

Question 9

Question Type: MultipleChoice

A data analyst pulls a table similar to the following one:

ID	Type	TypeID	Phone
----	------	--------	-------

1	Full Time	Full Time	1 Mobile
---	-----------	-----------	----------

2	Part Time	Part Time	2 Work
---	-----------	-----------	--------

3	Full Time	Full Time	3 Mobile
---	-----------	-----------	----------

Which of the following best explains the data issue with TypeID?

Options:

- A- Redundancy
- B- Outlier
- C- Missing data
- D- Duplication

Answer:

A

Explanation:

This question is part of the Data Concepts and Environments domain, focusing on identifying data quality issues. The table shows Type and TypeID columns, where TypeID seems to repeat information from Type with an additional identifier.

Redundancy (Option A): The TypeID column (e.g., 'Full Time 1') redundantly includes the Type value ('Full Time') with an extra identifier, which is unnecessary and could be simplified by using a numeric ID instead.

Outlier (Option B): Outliers are data points that deviate significantly, which isn't applicable here.

Missing data (Option C): There are no missing values in the table.

Duplication (Option D): Duplication refers to identical rows, but the rows here are unique; the issue is with the column content.

The DA0-002 Data Concepts and Environments domain includes understanding 'data schemas and dimensions,' and redundancy is a common data quality issue in schema design.

=====

To Get Premium Files for DA0-002 Visit

<https://www.p2pexams.com/products/da0-002>

For More Free Questions Visit

<https://www.p2pexams.com/comptia/pdf/da0-002>

20%
DISCOUNT

P2P
exams