

# CompTIA

## Exam Questions DA0-002

CompTIA Data+ Exam (2025)



### NEW QUESTION 1

A data analyst encounters an issue with new software and a code that they are using. The analyst includes print statements in the code to try to identify the issue, without success. An informal peer review of the code also produces the same result. The analyst confirms that the software is updated to the latest version and compatible with the code. Which of the following troubleshooting steps should the analyst take next?

- A. Use the old software and preexisting code, since both were functional.
- B. Contact the IT department and inform them that the software has a bug.
- C. Escalate to the department manager and ask for assistance.
- D. Research the issue online and see if a solution is available.

**Answer:** D

#### Explanation:

This question pertains to the Data Governance domain, focusing on troubleshooting and maintaining data quality in software processes. The analyst has already tried basic debugging and confirmed compatibility, so the next step involves seeking external resources.

? Use the old software and preexisting code, since both were functional (Option A):

Reverting to old software avoids solving the issue and may introduce other risks (e.g., security vulnerabilities).

? Contact the IT department and inform them that the software has a bug (Option B):

Assuming a bug without further investigation is premature, especially since compatibility was confirmed.

? Escalate to the department manager and ask for assistance (Option C): Escalation

is a later step after exhausting technical troubleshooting options.

? Research the issue online and see if a solution is available (Option D): Researching online (e.g., forums, documentation) is a logical next step to find solutions or identify known issues, especially after local debugging fails.

The DA0-002 Data Governance domain includes "data quality control concepts," and researching online is a standard troubleshooting step to maintain data process integrity. Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 5.0 Data Governance.

=====

### NEW QUESTION 2

A data analyst receives a flat file that includes dates. The analyst needs to calculate the number of days from the dates on the file to the current date. Which of the following is the best way to complete this task?

- A. Convert data to date format and use date functions.
- B. Validate the date format with logical functions and use date functions to analyze.
- C. Use date functions to analyze the data with no conversion.
- D. Transform data to a numerical value and use mathematical functions.

**Answer:** A

#### Explanation:

This question pertains to the Data Analysis domain, focusing on date calculations. The task is to calculate the difference between dates in a file and the current date, requiring proper date handling.

? Convert data to date format and use date functions (Option A): Flat files often store

dates as strings (e.g., "2023-01-01"). Converting them to a date format (e.g., using Python's datetime or SQL's TO\_DATE) allows the use of date functions (e.g., DATEDIFF) to calculate the difference to the current date, which is the best approach.

? Validate the date format with logical functions and use date functions to analyze

(Option B): Validation is unnecessary if conversion handles format issues, making this overly complex.

? Use date functions to analyze the data with no conversion (Option C): Without

converting to a date format, date functions may fail if the data is stored as strings.

? Transform data to a numerical value and use mathematical functions (Option D): This is inefficient and error-prone compared to using date functions.

The DA0-002 Data Analysis domain includes "applying the appropriate descriptive statistical methods," and converting to date format followed by date functions is the standard method for such calculations.

Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 3.0 Data Analysis.

=====

### NEW QUESTION 3

Which of the following is business intelligence software?

- A. SAS
- B. Python
- C. Notepad++
- D. Tableau

**Answer:** D

#### Explanation:

This question falls under the Visualization and Reporting domain, focusing on identifying tools used for business intelligence (BI), which typically involves data visualization and reporting.

? SAS (Option A): SAS is a statistical analysis software, not primarily a BI tool focused on visualization.

? Python (Option B): Python is a programming language, not a BI software, though it can be used for data analysis with libraries.

? Notepad++ (Option C): Notepad++ is a text editor, not a BI tool.

? Tableau (Option D): Tableau is a leading BI software designed for data visualization, dashboards, and reporting, making it the correct choice.

The DA0-002 Visualization and Reporting domain includes understanding "the appropriate visualization in the form of a report or dashboard," and Tableau is a recognized BI tool for this purpose.

Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 4.0 Visualization and Reporting.

=====

**NEW QUESTION 4**

Which of the following is a NoSQL database?

- A. PostgreSQL
- B. MySQL
- C. Oracle
- D. MongoDB

**Answer: D**

**Explanation:**

This question falls under the Data Concepts and Environments domain, focusing on types of databases. The task is to identify a NoSQL database among the options.

? PostgreSQL (Option A): PostgreSQL is a relational (SQL) database, not NoSQL.

? MySQL (Option B): MySQL is a relational (SQL) database, not NoSQL.

? Oracle (Option C): Oracle Database is a relational (SQL) database, not NoSQL.

? MongoDB (Option D): MongoDB is a NoSQL database that uses a document-based model, storing data in JSON-like structures, making it the correct choice.

The DA0-002 Data Concepts and Environments domain includes understanding "different types of databases and data repositories," and MongoDB is a well-known NoSQL database.

Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 1.0 Data Concepts and Environments.

=====

**NEW QUESTION 5**

Which of the following pieces of information, if made public, results in a data privacy violation?

- A. Gender
- B. Driver's license
- C. Age
- D. Employment status

**Answer: B**

**Explanation:**

This question falls under the Data Governance domain, which in DA0-002 includes understanding data privacy and compliance with regulations like GDPR. The question asks which piece of information, if made public, constitutes a privacy violation, meaning it must be personally identifiable information (PII).

? Gender (Option A): Gender is not typically considered PII on its own, as it's not uniquely identifiable.

? Driver's license (Option B): A driver's license number is PII because it uniquely identifies an individual and can be linked to other personal information, such as name and address. Making it public violates privacy regulations.

? Age (Option C): Age alone isn't PII, as it's not uniquely identifiable.

? Employment status (Option D): Employment status (e.g., employed, unemployed) isn't PII, as it doesn't uniquely identify an individual.

The DA0-002 Data Governance domain includes "identifying PII and data privacy concepts," and a driver's license is a clear example of PII that, if exposed, results in a privacy violation.

Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 5.0 Data Governance.

=====

**NEW QUESTION 6**

The human resources department wants to know the number of employees who earn \$125,000 or more. However, the department is concerned about duplicates in the dataset. Given the following table:

Employee\_ID Level

Salary

001

1

10000

002

2

20000

003

2

256000

004

2

125000

001

1

10000

002

2

20000

Which of the following SQL statements resolves this issue?

- A. SELECT DISTINCT Employee\_ID FROM Employee WHERE Salary >= 125000
- B. SELECT COUNT(DISTINCT Employee\_ID) FROM Employee WHERE Salary >= 125000
- C. SELECT DISTINCT Employee\_ID FROM Employee WHERE Salary > 125000
- D. SELECT COUNT(Employee\_ID) FROM Employee WHERE Salary >= 125000

**Answer: B**

**Explanation:**

This question falls under the Data Analysis domain, focusing on SQL queries to handle duplicates while counting employees. The task is to count unique employees with a salary of \$125,000 or more, addressing duplicates in the dataset.

? Option A: SELECT DISTINCT Employee\_ID FROM Employee WHERE Salary >= 125000 This lists unique Employee\_IDs but doesn't provide a count, which the department needs.

? Option B: SELECT COUNT(DISTINCT Employee\_ID) FROM Employee WHERE Salary >= 125000 This counts unique Employee\_IDs (using DISTINCT) with a salary of \$125,000 or more, correctly addressing duplicates and providing the required count (2 employees: 003 and 004).

? Option C: SELECT DISTINCT Employee\_ID FROM Employee WHERE Salary > 125000 This lists unique Employee\_IDs with a salary strictly greater than \$125,000 (missing 004), and doesn't provide a count.

? Option D: SELECT COUNT(Employee\_ID) FROM Employee WHERE Salary >= 125000 This counts all rows without addressing duplicates, resulting in an incorrect count (2 rows, but only 2 unique employees).

The DA0-002 Data Analysis domain includes "applying the appropriate descriptive statistical methods using SQL queries," and COUNT(DISTINCT) is the correct method to count unique employees while handling duplicates.

Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 3.0 Data Analysis.

=====

**NEW QUESTION 7**

A data analyst is designing a report for the business review team. The team lists the following requirements for the report:

- Specific data points
- Color branding
- Labels and terminology
- Suggested charts and tables

Which of the following components is missing from the requirements?

- A. Source validation
- B. Design elements
- C. Delivery method
- D. Report type

**Answer: C**

**Explanation:**

This question falls under the Visualization and Reporting domain of CompTIA Data+ DA0-002, which involves understanding the components necessary for designing a report. The given requirements cover data, visuals, and design, but a key aspect of report planning is missing.

? Source validation (Option A): Source validation ensures data accuracy, but it's typically part of the data preparation phase, not a report design requirement.

? Design elements (Option B): Color branding, labels, and terminology are design elements, so this is already included.

? Delivery method (Option C): The delivery method (e.g., recurring, ad hoc, self-service) specifies how the report will be distributed or accessed, which is a critical requirement missing from the list.

? Report type (Option D): Suggested charts and tables imply the report type (e.g., summary, dashboard), so this is indirectly covered.

The DA0-002 Visualization and Reporting domain emphasizes "translating business requirements to form the appropriate visualization," and the delivery method is a key component of report planning that's missing here.

Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 4.0 Visualization and Reporting.

=====

**NEW QUESTION 8**

Which of the following explains the purpose of UAT?

- 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.

Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 5.0 Data Governance

=====

**NEW QUESTION 9**

A data analyst creates a report, and some of the fields are empty. Which of the following conditions should the analyst add to a query to provide a list of all the records with empty fields?

- A. WHERE [ColumnName] = NULL
- B. WHERE [ColumnName] IS NULL
- C. WHERE [ColumnName] IS NOT NULL
- D. WHERE [ColumnName] = 'NULL'

**Answer: B**

**Explanation:**

This question falls under the Data Analysis domain, focusing on SQL queries to identify data issues. The task is to find records with empty fields, which in SQL means NULL values.

? WHERE [ColumnName] = NULL (Option A): In SQL, NULL cannot be compared using "="; this syntax is incorrect.

? WHERE [ColumnName] IS NULL (Option B): This is the correct SQL syntax to identify NULL values, which represent empty fields.

? WHERE [ColumnName] IS NOT NULL (Option C): This finds non-empty fields, the opposite of the requirement.

? WHERE [ColumnName] = 'NULL' (Option D): This checks for the string "NULL," not a true NULL value, which is incorrect.

The DA0-002 Data Analysis domain includes "applying the appropriate descriptive statistical methods using SQL queries," such as identifying NULL values with IS NULL. Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 3.0 Data Analysis.

=====

**NEW QUESTION 10**

An analyst needs to create a collection of dashboards for multiple teams within their organization. Which of the following should the analyst do first before starting the project?

- A. Evaluate the user persona type for the dashboards.
- B. Determine the number of team members who need to access the dashboards.
- C. Determine the delivery method of the dashboards.
- D. Evaluate the KPIs for the dashboards.

**Answer: A**

**Explanation:**

This question pertains to the Visualization and Reporting domain, focusing on the initial steps in dashboard creation. The analyst is starting a project for multiple teams, so understanding the audience is critical.

? Evaluate the user persona type for the dashboards (Option A): User personas define the needs, roles, and preferences of the audience (e.g., executives vs. analysts), which is the first step to ensure the dashboards meet user requirements.

? Determine the number of team members who need to access the dashboards (Option B): This is important for access control but not the first step in design.

? Determine the delivery method of the dashboards (Option C): Delivery method (e.g., self-service) is decided later, after understanding user needs.

? Evaluate the KPIs for the dashboards (Option D): KPIs are defined after understanding the user personas and their goals.

The DA0-002 Visualization and Reporting domain emphasizes "translating business requirements to form the appropriate visualization," and evaluating user personas is the first step in this process.

Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 4.0 Visualization and Reporting.

=====

**NEW QUESTION 10**

A data analyst learns that a report detailing employee sales is reflecting sales only for the current month. Which of the following is the most likely cause?

- A. Lack of permissions
- B. An error in SQL code
- C. Report refresh failure
- D. Connectivity issues

**Answer: B**

**Explanation:**

This question falls under the Data Analysis domain, focusing on troubleshooting issues in data reports. The report should show all employee sales but is limited to the current month, suggesting a data retrieval issue.

? Lack of permissions (Option A): Permissions issues would likely prevent access entirely, not limit data to the current month.

? An error in SQL code (Option B): The report likely uses an SQL query to retrieve data, and an error (e.g., a WHERE clause filtering for the current month) could restrict the data to the current month, making this the most likely cause.

? Report refresh failure (Option C): A refresh failure would result in outdated data, not specifically current-month data.

? Connectivity issues (Option D): Connectivity issues would likely prevent the report from running, not limit it to a specific time frame.

The DA0-002 Data Analysis domain includes "applying the appropriate descriptive statistical methods using SQL queries," and errors in SQL code are a common cause of incorrect data retrieval in reports.

Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 3.0 Data Analysis.

=====

**NEW QUESTION 12**

A data analyst receives an email from the IT department about renewing the company password, and the analyst follows the password reset link as required. Later in the week, the analyst receives the following notification when running a recurring analysis that connects to the database:

Log-in failed for user ??<username>??

Which of the following is most likely the reason for this issue?

- A. The company changed its database authentication method.
- B. The password expiration process locked the account.
- C. The analyst did not change the password used to launch the report.
- D. The company is experiencing issues with password replication.

**Answer: C**

**Explanation:**

This question falls under the Data Governance domain, focusing on data access and security troubleshooting. The analyst reset their password, but the recurring analysis failed to log in, indicating a mismatch.

? The company changed its database authentication method (Option A): This would affect all users, not just the analyst, and there's no indication of a broader change.

? The password expiration process locked the account (Option B): The analyst reset the password as required, so the account isn't likely locked due to

expiration.

? The analyst did not change the password used to launch the report (Option C): Recurring analyses often use stored credentials. If the analyst updated their password but didn't update the stored credentials for the analysis, the login would fail, making this the most likely reason.

? The company is experiencing issues with password replication (Option D): This is possible but less likely without evidence of broader system issues.

The DA0-002 Data Governance domain includes "data privacy concepts," and ensuring stored credentials match updated passwords is a common issue in recurring analyses. Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 5.0 Data Governance.

=====

#### NEW QUESTION 14

A data analyst is evaluating all conditions in a query. Which of the following is the best logical function to accomplish this task?

- A. OR
- B. NOT
- C. AND
- D. IF

**Answer: C**

#### Explanation:

This question falls under the Data Analysis domain, focusing on SQL logical functions for query evaluation. The task is to evaluate "all conditions," implying multiple conditions must be true together.

? OR (Option A): OR returns true if any condition is true, not ensuring all conditions are met.

? NOT (Option B): NOT negates a condition, not suitable for combining multiple conditions.

? AND (Option C): AND requires all conditions to be true, which aligns with evaluating "all conditions" in a query.

? IF (Option D): IF is a conditional function for decision-making, not for evaluating multiple conditions together.

The DA0-002 Data Analysis domain includes "applying the appropriate descriptive statistical methods using SQL queries," and AND is the best logical function for ensuring all conditions are met.

Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 3.0 Data Analysis.

=====

#### NEW QUESTION 18

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

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

**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.

#### NEW QUESTION 22

A manager wants a report that contains the days off for each direct report. The manager needs this report to always be up-to-date with the latest data. Which of the following describes the refresh frequency that the manager is requesting?

- A. Real-time
- B. Ad hoc
- C. Snapshot
- D. Dynamic

**Answer: A**

#### Explanation:

This question pertains to the Visualization and Reporting domain, focusing on report refresh frequencies. The manager needs the report to always be up-to-date, implying continuous data updates.

? Real-time (Option A): Real-time refresh frequency ensures the report reflects the latest data as soon as it changes, which matches the requirement to "always be up-to-date."

? Ad hoc (Option B): Ad hoc reports are generated on-demand, not continuously updated.

? Snapshot (Option C): A snapshot captures data at a specific point in time, not suitable for always being up-to-date.

? Dynamic (Option D): Dynamic reports allow interactivity, but the term doesn't specifically imply real-time updates.

The DA0-002 Visualization and Reporting domain includes "the appropriate visualization in the form of a report" with delivery methods, and real-time refresh frequency ensures the report is always current.

Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 4.0 Visualization and Reporting.

=====

#### NEW QUESTION 24

Which of the following allows a data analyst to send out a spreadsheet containing sensitive information without revealing personal details?

- A. Using a UUID in the data file
- B. Redacting all PII
- C. Adding access controls to the ID column
- D. Encrypting the spreadsheet

**Answer: B**

#### Explanation:

This question pertains to the Data Governance domain, focusing on data privacy and security. The task is to share a spreadsheet with sensitive information while protecting personal details.

? Using a UUID in the data file (Option A): A UUID (Universally Unique Identifier) can anonymize records, but if other PII (e.g., names) remains, personal details are still exposed.

? Redacting all PII (Option B): Redacting personally identifiable information (PII) removes sensitive details (e.g., names, addresses), ensuring personal information isn't revealed while sharing the spreadsheet.

? Adding access controls to the ID column (Option C): Access controls limit who can view the data, but the question focuses on the spreadsheet content itself, not access.

? Encrypting the spreadsheet (Option D): Encryption protects the file during transmission, but once opened, personal details are still visible unless redacted. The DA0-002 Data Governance domain includes "data privacy concepts," and redacting PII is the most direct method to protect personal details in a shared spreadsheet.

Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 5.0 Data Governance.

=====

#### NEW QUESTION 25

A data analyst is creating a new dataset that involves bringing together the following datasets:

Name ID

Date of birth

Frank 23525

3/19

Martha 11290

6/13

Ellen 12141

11/4

ID

Address City State 23525

1234 Harding Chicago

IL 11040

935 Terrace Hills Chino

CA 11290

2 Speedway Miami

FL

Which of the following would be the output if the data analyst does a FULL JOIN?

A. Name ID Date of birth Address City State Frank 23525 3/19 1234 Harding Chicago IL Martha 11290 6/13 935 Terrace Hills Chino CA Ellen 12141 11/4 2 Speedway Miami FL

B. Name ID Date of birth Address City State Frank 23525 3/19 1234 Harding Chicago IL Martha 11290 6/13 935 Terrace Hills Chino CA Ellen 12141 11/4 2 Speedway Miami FL

C. Name ID Date of birth Address City State Frank 23525 3/19 1234 Harding Chicago IL Martha 11290 6/13 2 Speedway Miami FL Ellen 12141 11/4 935 Terrace Hills Chino CA

D. Name ID Date of birth Address City State Frank 23525 3/19 1234 Harding Chicago IL Martha 11290 6/13 2 Speedway Miami FL Ellen 12141 11/4 11040 935 Terrace Hills Chino CA

**Answer: D**

#### Explanation:

This question falls under the Data Concepts and Environments domain, focusing on database operations like joins. A FULL JOIN combines all rows from both tables, including matches and non-matches, filling in NULLs where there's no corresponding data.

? The first table has IDs: 23525 (Frank), 11290 (Martha), 12141 (Ellen).

? The second table has IDs: 23525, 11040, 11290.

? A FULL JOIN includes all IDs: 23525, 11290, 12141, 11040.

? Option A: Incorrect; it includes a row for Ellen with "2 Speedway," but Ellen's ID (12141) doesn't match any address, and 11040 is missing.

? Option B: Identical to Option A, so incorrect for the same reasons.

? Option C: Incorrect; it mismatches addresses (e.g., Ellen with 935 Terrace Hills, which belongs to 11040).

? Option D: Correct; it includes all IDs, with NULLs for non-matching rows (Ellen has no address, and 11040 has no name).

The DA0-002 Data Concepts and Environments domain includes understanding "data schemas and dimensions," such as performing joins in relational databases.

Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 1.0 Data Concepts and Environments.

=====

#### NEW QUESTION 30

A data analyst receives the following sales data for a convenience store:

Item Quantity Price

Chocolate Bars 7

\$1.99

Vanilla Ice Bars

2

\$4.99

Chocolate Wafers 6

\$0.99

Peanut Butter 2

\$2.99  
 Cups 3  
 \$4.99  
 Strawberry Jam 3  
 \$4.99  
 Chocolate Cake 9  
 \$6.99  
 Milk Chocolate 2  
 \$2.99  
 Almonds 5  
 \$2.99

The analyst needs to provide information on the products that contain chocolate. Which of the following RegEx should the analyst use to filter the chocolate products?

- A. Chocolate!
- B. Chocolate\$
- C. %Chocolate&
- D. #Chocolate#&

**Answer: B**

**Explanation:**

This question falls under the Data Acquisition and Preparation domain, which includes techniques for manipulating and filtering data, such as using regular expressions (RegEx) to identify specific patterns in text data. The task is to filter items containing the word "Chocolate."

? Chocolate! (Option A): In RegEx, "!" is not a valid pattern for matching a word like "Chocolate." It typically denotes negation in some contexts, but here it's incorrect.

? Chocolate\$ (Option B): The "\$" in RegEx anchors the pattern to the end of the string, meaning it matches "Chocolate" at the end of an item name (e.g., "Milk Chocolate"). This is the most appropriate pattern for identifying items ending with "Chocolate," which applies to the relevant items in the list.

? %Chocolate& (Option C): "%" and "&" are not standard RegEx anchors; they're often used in SQL LIKE patterns, not RegEx, making this incorrect.

? #Chocolate#\$ (Option D): "#" is not a standard RegEx anchor, and this pattern would look for "Chocolate" surrounded by "#", which doesn't match the data.

The DA0-002 Data Acquisition and Preparation domain includes "executing data manipulation", and RegEx is a common technique for filtering text data. The pattern "Chocolate\$" correctly identifies items like "Chocolate Bars," "Chocolate Wafers," "Chocolate Cake," and "Milk Chocolate."

Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 2.0 Data Acquisition and Preparation

=====

**NEW QUESTION 35**

Given the following dataset:

Day  
 Number of Guests  
 Monday 455  
 Tuesday 346  
 Wednesday 382  
 Thursday 563  
 Friday 887  
 Saturday 934  
 Sunday 346

Which of the following is the mode?

- A. 346
- B. 446
- C. 455
- D. 559

**Answer: A**

**Explanation:**

This question falls under the Data Analysis domain, focusing on statistical measures. The mode is the value that appears most frequently in a dataset.

? Monday: 455

? Tuesday: 346

? Wednesday: 382

? Thursday: 563

? Friday: 887

? Saturday: 934

? Sunday: 346

The value 346 appears twice (Tuesday and Sunday), while all other values (455, 382, 563, 887, 934) appear once. Thus, the mode is 346.

? Option A: 346– Correct, as it's the most frequent value.

? Option B: 446– Incorrect, as 446 doesn't appear in the dataset.

? Option C: 455– Incorrect, as 455 appears only once.

? Option D: 559– Incorrect, as 559 doesn't appear in the dataset.

The DA0-002 Data Analysis domain includes "applying the appropriate descriptive statistical methods," and the mode is a fundamental measure of central tendency. Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 3.0 Data Analysis.

=====

**NEW QUESTION 40**

A data analyst is preparing a survey for Paralympic Games athletes. Which of the following should the analyst consider when creating this survey?

- A. Idioms
- B. Color contrast
- C. Refresh speed
- D. Granularity

**Answer:** B

**Explanation:**

This question pertains to the Visualization and Reporting domain, focusing on survey design considerations, particularly for accessibility. The survey is for Paralympic athletes, who may have visual impairments, requiring specific design considerations.

? Idioms (Option A): Idioms (e.g., colloquial phrases) might confuse non-native speakers, but they're not a primary survey design concern for Paralympic athletes.

? Color contrast (Option B): High color contrast ensures readability for athletes with visual impairments (e.g., color blindness), a critical accessibility consideration for Paralympic surveys.

? Refresh speed (Option C): Refresh speed is relevant for dashboards, not static surveys.

? Granularity (Option D): Granularity refers to data detail levels, not a survey design consideration.

The DA0-002 Visualization and Reporting domain includes "translating business requirements to form the appropriate visualization," and color contrast is a key accessibility factor in survey design for diverse audiences.

Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 4.0 Visualization and Reporting.

=====

**NEW QUESTION 44**

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

- 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.

Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 5.0 Data Governance.

=====

**NEW QUESTION 45**

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

- 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.

Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 1.0 Data Concepts and Environments.

=====

**NEW QUESTION 47**

A data analyst wants to understand several datasets at the variable level. Which of the following should the analyst consult to find this information?

- A. Data lineage
- B. Data dictionary
- C. Data flow diagram
- D. Data versioning

**Answer:** B

**Explanation:**

This question pertains to the Data Concepts and Environments domain, focusing on documentation for understanding datasets. The analyst needs variable-level details (e.g., field definitions), which points to a specific type of documentation.

? Data lineage (Option A): Data lineage tracks data flow and transformations, not variable definitions.

? Data dictionary (Option B): A data dictionary provides detailed information about variables (e.g., name, type, description), which is exactly what the analyst

needs.  
? Data flow diagram (Option C): A data flow diagram shows data movement, not variable-level details.  
? Data versioning (Option D): Data versioning tracks changes to datasets, not variable definitions.  
The DA0-002 Data Concepts and Environments domain includes understanding "data schemas and dimensions," and a data dictionary is the primary tool for variable-level information.  
Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 1.0 Data Concepts and Environments.  
=====

**NEW QUESTION 48**

.....

## **Thank You for Trying Our Product**

### **We offer two products:**

1st - We have Practice Tests Software with Actual Exam Questions

2nd - Questions and Answers in PDF Format

### **DA0-002 Practice Exam Features:**

- \* DA0-002 Questions and Answers Updated Frequently
- \* DA0-002 Practice Questions Verified by Expert Senior Certified Staff
- \* DA0-002 Most Realistic Questions that Guarantee you a Pass on Your FirstTry
- \* DA0-002 Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year

**100% Actual & Verified — Instant Download, Please Click**  
**[Order The DA0-002 Practice Test Here](#)**