



CompTIA

Exam Questions DA0-002

CompTIA Data+ Exam (2025)

About ExamBible

[Your Partner of IT Exam](#)

Found in 1998

ExamBible is a company specialized on providing high quality IT exam practice study materials, especially Cisco CCNA, CCDA, CCNP, CCIE, Checkpoint CCSE, CompTIA A+, Network+ certification practice exams and so on. We guarantee that the candidates will not only pass any IT exam at the first attempt but also get profound understanding about the certificates they have got. There are so many alike companies in this industry, however, ExamBible has its unique advantages that other companies could not achieve.

Our Advances

* 99.9% Uptime

All examinations will be up to date.

* 24/7 Quality Support

We will provide service round the clock.

* 100% Pass Rate

Our guarantee that you will pass the exam.

* Unique Gurantee

If you do not pass the exam at the first time, we will not only arrange FULL REFUND for you, but also provide you another exam of your claim, ABSOLUTELY FREE!

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

Which of the following is found in metadata?

- A. Transformations
- B. Data lineage
- C. Syntax
- D. Variable types

Answer: D

Explanation:

This question pertains to the Data Concepts and Environments domain, focusing on the content of metadata. Metadata describes data attributes, and the task is to identify what it typically includes.

? Transformations (Option A): Transformations (e.g., data cleaning steps) are part of data lineage, not metadata.

? Data lineage (Option B): Data lineage tracks data flow and transformations, which is related to metadata but not a direct component.

? Syntax (Option C): Syntax refers to code structure, not a metadata component.

? Variable types (Option D): Metadata includes information about data fields, such as variable types (e.g., integer, string), which is a standard component.

The DA0-002 Data Concepts and Environments domain includes understanding "data schemas and dimensions," and metadata typically contains details like variable types to describe the dataset.

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

NEW QUESTION 3

A data analyst needs to get an accurate idea of how data components are automated. Which of the following types of documentation should the analyst review first?

- A. Data flow diagram
- B. Data explainability report
- C. Data dictionary
- D. Data lineage

Answer: A

Explanation:

This question pertains to the Data Concepts and Environments domain, focusing on documentation for understanding data processes. The analyst needs to understand automation of data components, which involves data movement and processes.

? Data flow diagram (Option A): A data flow diagram (DFD) visualizes how data moves through systems, including automated processes, making it the best starting point.

? Data explainability report (Option B): This is related to AI/ML model transparency, not data automation.

? Data dictionary (Option C): A data dictionary defines data elements, not how they're automated.

? Data lineage (Option D): Data lineage tracks data origin and transformations but doesn't focus on automation processes.

The DA0-002 Data Concepts and Environments domain includes understanding "data schemas and dimensions," and a data flow diagram is key for visualizing automation. Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 1.0 Data Concepts and Environments.

=====

NEW QUESTION 4

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 5

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?

- 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.

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

=====

NEW QUESTION 6

A data analyst is gathering data from multiple tables in a database. The analyst needs certain columns from each table. Which of the following is the best method to accomplish this task?

- A. Aggregate
- B. Union
- C. Nest
- D. Join

Answer: D

Explanation:

This question falls under the Data Acquisition and Preparation domain, focusing on combining data from multiple tables. The analyst needs specific columns from each table, suggesting a method to combine data horizontally based on relationships.

? Aggregate (Option A): Aggregation (e.g., SUM, COUNT) summarizes data, not suitable for combining columns from tables.

? Union (Option B): Union stacks tables vertically, requiring identical structures, but the analyst needs specific columns, likely based on relationships, not a vertical stack.

? Nest (Option C): Nesting is used for hierarchical data (e.g., JSON), not for combining relational tables.

? Join (Option D): A join (e.g., INNER JOIN) combines tables horizontally based on a common key, allowing the analyst to select specific columns from each table, which fits the task.

The DA0-002 Data Acquisition and Preparation domain includes "executing data manipulation," and joining tables is the best method for combining specific columns from multiple tables.

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

=====

NEW QUESTION 7

Which of the following supports capabilities such as automatic versioning, corruption checks, KPIs, and user authentication?

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

=====

NEW QUESTION 8

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 9

A data analyst is working on an initial analysis of the dataset in the following table:

DateTime Count 2024-01-01

12

2024-01-02

245

2024-01-02

13

2024-01-03

13

2024-01-03

12

00:00:00

12

Which of the following issues should the analyst flag in the data report?

A. Completeness

B. Outlier

C. Mismatch

D. Duplication

Answer: B

Explanation:

This question falls under the Data Analysis domain, focusing on identifying data quality issues. The table shows counts over time, and the analyst needs to flag an issue in the data.

? Completeness (Option A): Completeness refers to missing data, but all rows have values for DateTime and Count.

? Outlier (Option B): The count of 245 on 2024-01-02 is significantly higher than other counts (12-13), indicating an outlier that should be investigated for accuracy.

? Mismatch (Option C): Mismatch implies inconsistent data types or formats, but the DateTime and Count columns appear consistent except for the last row (addressed separately).

? Duplication (Option D): Duplication refers to identical rows, but no rows are identical (same DateTime and Count).

The last row ("00:00:00", 12) has a formatting issue, but the most significant issue for analysis is the outlier (245), as it could skew results. The DA0-002 Data Analysis domain includes "applying the appropriate descriptive statistical methods," such as identifying outliers in datasets.

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

=====

NEW QUESTION 10

A data analyst calculated the average score per student without making any changes to the following table:

Student Subject Score

123 Math 100

234 Biology 80

234 Biology 80

345 Math 96

123 Biology 80

345 Biology 88

234 Math 96

345 Math 96

234

Math 96

Math 96

Which of the following exploration techniques should the analyst have considered before calculating the average?

- A. Duplication
- B. Redundancy
- C. Binning
- D. Grouping

Answer: A

Explanation:

This question pertains to the Data Governance domain, focusing on data quality issues that affect analysis. The table contains duplicate rows, which would skew the average score calculation if not addressed.

? Student 123: Math (100), Biology (80), Biology (80)– Duplicate Biology score.

? Student 234: Math (96), Math (96)– Duplicate Math score.

? Student 345: Biology (88)– No duplicates.

? Duplication (Option A): The table has duplicate rows (e.g., Student 123's Biology score of 80 appears twice), which would inflate the average if not removed.

The analyst should have checked for duplicates before calculating the average.

? Redundancy (Option B): Redundancy refers to unnecessary fields (e.g., storing the same data in multiple columns), not duplicate rows.

? Binning (Option C): Binning groups data into categories, not relevant for

addressing duplicates in averaging.

? Grouping (Option D): Grouping (e.g., GROUP BY in SQL) might be part of the solution, but the issue to identify is duplication.

The DA0-002 Data Governance domain includes "data quality control concepts," and checking for duplication is critical to ensure accurate calculations like averages. Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 5.0 Data Governance.

=====

NEW QUESTION 10

Which of the following data sources makes online data consumption easier?

- A. Data mart
- B. Web scraping
- C. Database
- D. Application programming interface

Answer: D

Explanation:

This question pertains to the Data Concepts and Environments domain, focusing on data sources that facilitate online data access. The task is to identify a source that simplifies online data consumption.

? Data mart (Option A): A data mart stores structured data for specific business

areas, typically accessed internally, not designed for online consumption.

? Web scraping (Option B): Web scraping extracts data from websites but requires parsing and cleaning, which isn't necessarily "easier."

? Database (Option C): Databases store data but aren't inherently designed for online consumption without an interface.

? Application programming interface (Option D): An API provides a structured way to access data online, often in formats like JSON, making data consumption

easier for applications and users.

The DA0-002 Data Concepts and Environments domain includes understanding "data sources," and APIs are specifically designed to simplify online data access.

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

=====

NEW QUESTION 14

An administrator needs to design a table that will include foreign words. Which of the following is the best option for storing non-native language characters?

- A. VARCHAR
- B. NVARCHAR
- C. CLOB
- D. CHAR

Answer: B

Explanation:

This question falls under the Data Concepts and Environments domain, focusing on selecting appropriate data types for storing specific kinds of data. The task requires storing foreign words, which often include non-native characters (e.g., accents, Unicode characters).

? VARCHAR (Option A): VARCHAR stores variable-length strings but typically uses ASCII or single-byte encoding, which may not support all foreign characters.

? NVARCHAR (Option B): NVARCHAR (National VARCHAR) stores variable-length strings in Unicode, supporting a wide range of non-native characters, making it the best choice.

? CLOB (Option C): CLOB (Character Large Object) is for large text data, but it's overkill for most foreign words and not specifically designed for Unicode.

? CHAR (Option D): CHAR stores fixed-length strings, but like VARCHAR, it often uses single-byte encoding, which may not support foreign characters.

The DA0-002 Data Concepts and Environments domain includes understanding "data schemas and dimensions," such as selecting data types like NVARCHAR for Unicode support.

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

=====

NEW QUESTION 19

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 24

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 27

A data analyst is generating a custom report for a Chief Executive Officer's executive meeting. Later, the analyst learns that other custom reports will be required for future executive meetings. Which of the following delivery methods should the analyst use?

- A. Ad hoc
- B. Real-time
- C. Recurring
- D. Self-service

Answer: C

Explanation:

This question falls under the Visualization and Reporting domain of DA0-002, which involves selecting appropriate delivery methods for reports. The scenario describes a need for custom reports for future executive meetings, implying a scheduled, repeated delivery.

? Ad hoc (Option A): Ad hoc reports are generated on-demand for one-time use, not suitable for ongoing needs.

? Real-time (Option B): Real-time delivery provides live data updates, which isn't necessary for scheduled executive meetings.

? Recurring (Option C): Recurring delivery involves scheduling reports to be generated and delivered at regular intervals (e.g., weekly or monthly), which fits the need for future executive meetings.

? Self-service (Option D): Self-service allows users to generate reports themselves, but the scenario implies the analyst will create the reports.

The DA0-002 Visualization and Reporting domain includes understanding "the appropriate visualization in the form of a report" with delivery methods, and recurring delivery aligns with scheduled reporting needs.

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

=====

NEW QUESTION 32

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

- 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.

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

=====

NEW QUESTION 33

A product goes viral on social media, creating high demand. Distribution channels are facing supply chain issues because the testing and training models that are used for sales forecasting have not encountered similar demand. Which of the following best describes this situation?

- A. Model bias
- B. Data drift
- C. Incorrect sizing
- D. Skewing

Answer: B

Explanation:

This question pertains to the Data Analysis domain, focusing on issues with forecasting models. The scenario describes a sudden change in demand (viral product) that the model couldn't predict because it hasn't seen similar patterns before.

? Model bias (Option A): Model bias occurs when a model systematically favors certain outcomes due to flawed training data, but this scenario is about a change in data patterns, not bias.

? Data drift (Option B): Data drift occurs when the statistical properties of the data change over time (e.g., sudden high demand due to virality), causing the model to perform poorly because it was trained on different patterns, which fits the scenario.

? Incorrect sizing (Option C): This term is vague and not a standard concept in data analysis for this context.

? Skewing (Option D): Skewing refers to data distribution asymmetry, not a change in data patterns affecting model performance.

The DA0-002 Data Analysis domain includes understanding "applying the appropriate descriptive statistical methods," and data drift is a key concept in forecasting when data patterns change unexpectedly.

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

NEW QUESTION 34

An analyst must use the same data to create reports for police patrol supervisors, the city council, and a public website. Which of the following is the best way to differentiate the reports?

- A. Persona type
- B. Detail levels
- C. Accessibility
- D. Sensitivity

Answer: A

Explanation:

This question falls under the Visualization and Reporting domain, focusing on tailoring reports for different audiences. The same data is used for three distinct groups (supervisors, city council, public), requiring differentiation.

? Persona type (Option A): Persona types define the audience's needs and

preferences (e.g., supervisors need operational details, the city council needs summaries, the public needs simplified data), making this the best way to

differentiate the reports.

? Detail levels (Option B): Detail levels are a result of persona types, not the method of differentiation.

? Accessibility (Option C): Accessibility ensures access for all users (e.g., screen readers), but it's not the primary way to differentiate content.

? Sensitivity (Option D): Sensitivity determines data access (e.g., confidential vs. public), but the scenario implies all reports use the same data, just presented differently.

The DA0-002 Visualization and Reporting domain emphasizes "translating business requirements to form the appropriate visualization," and persona types guide report differentiation for diverse audiences.

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

=====

NEW QUESTION 36

A report triggers an error that prevents information from being displayed. However, the report was functional before a database upgrade. Which of the following should a data analyst do first to troubleshoot the problem?

- A. Ensure the system has permissions for the report service.
- B. Change the report's refresh rate.
- C. Verify the connection to the database.
- D. Check whether the data structures were modified.

Answer: C

Explanation:

This question falls under the Data Concepts and Environments domain, which in DA0-002 involves understanding database environments, connections, and troubleshooting issues related to data access. The scenario describes a report failing after a database upgrade, indicating a potential issue with the database environment or connectivity.

? Ensure the system has permissions for the report service (Option A): While permissions issues can cause report failures, they are less likely to be the first issue after a database upgrade unless explicitly mentioned.

? Change the report's refresh rate (Option B): Refresh rate adjustments might address performance issues but won't resolve a fundamental error preventing data display.

? Verify the connection to the database (Option C): A database upgrade often involves changes to connection strings, drivers, or network configurations. Verifying the connection ensures the report can access the database, making this the most logical first step.

? Check whether the data structures were modified (Option D): While possible, checking data structures (e.g., schema changes) is a deeper troubleshooting step that should follow after confirming basic connectivity.

The DA0-002 Data Concepts and Environments domain includes understanding database connectivity, similar to DA0-001's focus on "data schemas and dimensions" and environments (web ID: 1). Verifying the connection is the first recommended step in troubleshooting post-upgrade issues.

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

=====

NEW QUESTION 39

A table contains several rows of cellular numbers with call timestamps, call durations, called numbers, and carriers of the called number. Which of the following allows a data analyst to sort the cellular numbers based on the carriers of the called numbers and include the total call durations?

- A. `SELECT cellular_number, called_number_carrier, SUM(call_duration) FROM calls GROUP BY cellular_number ORDER BY called_number_carrier;`
- B. `SELECT cellular_number, SUM(call_duration) FROM calls GROUP BY call_duration ORDER BY called_number_carrier;`
- C. `SELECT cellular_number, called_number_carrier, SUM(call_duration) FROM calls GROUP BY cellular_number, called_number_carrier ORDER BY called_number_carrier;`
- D. `SELECT cellular_number, called_number_carrier, SUM(call_duration) FROM calls GROUP BY call_duration ORDER BY called_number_carrier;`

Answer: C

Explanation:

This question falls under the Data Analysis domain of CompTIA Data+ DA0-002, focusing on SQL queries for data analysis. The task requires sorting cellular numbers by the carrier of the called number (called_number_carrier) and calculating the total call durations (SUM(call_duration)).

? Option A: `SELECT cellular_number, called_number_carrier, SUM(call_duration) FROM calls GROUP BY cellular_number ORDER BY called_number_carrier` This query groups by cellular_number only, but called_number_carrier is in the SELECT clause without being in the GROUP BY, which is invalid in SQL (it would raise an error in most databases).

? Option B: `SELECT cellular_number, SUM(call_duration) FROM calls GROUP BY call_duration ORDER BY called_number_carrier` This query doesn't include called_number_carrier in the SELECT clause, so it cannot be used in the ORDER BY clause, making it invalid. Grouping by call_duration also doesn't align with the task.

? Option C: `SELECT cellular_number, called_number_carrier, SUM(call_duration) FROM calls GROUP BY cellular_number, called_number_carrier ORDER BY called_number_carrier` This query correctly groups by both cellular_number and called_number_carrier (since both are in the SELECT clause), calculates the total call duration with SUM(call_duration), and sorts by called_number_carrier as required.

? Option D: `SELECT cellular_number, called_number_carrier, SUM(call_duration) FROM calls GROUP BY call_duration ORDER BY called_number_carrier` Grouping by call_duration is incorrect because cellular_number and called_number_carrier are in the SELECT clause but not in the GROUP BY, making this query invalid.

The DA0-002 Data Analysis domain includes "applying the appropriate descriptive statistical methods using SQL queries," and Option C correctly aggregates and sorts the data as specified.

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

=====

NEW QUESTION 41

Which of the following is the best tool for creating a dynamic dashboard?

- A. Power BI
- B. RStudio
- C. Excel
- D. SAS

Answer: A

Explanation:

The question asks for the best tool to create adynamic dashboard, which falls under the Visualization and Reportingdomain of CompTIA Data+ DA0-002. According to the DA0- 002 draft objectives, this domain includes understanding tools and techniques for creating effective visualizations, such as dashboards, that can be updated dynamically to reflect real-time or changing data. A dynamic dashboard typically allows for interactivity, real-time updates, and user-driven exploration of data, which is a key focus in this domain.

? Power BI (Option A): Power BI is a business intelligence tool by Microsoft

designed specifically for creating interactive and dynamic dashboards. It supports real-time data updates, user interactivity (e.g., filters, slicers), and integration with various data sources, making it ideal for dynamic dashboard creation.

? RStudio (Option B): RStudio is primarily an IDE for the R programming language,

used for statistical computing and data analysis. While it can create visualizations, it??s not optimized for dynamic dashboards without additional packages like Shiny, and even then, it requires more coding effort compared to Power BI.

? Excel (Option C): Excel is a spreadsheet tool that can create static charts and

basic dashboards, but it lacks the interactivity and real-time update capabilities of a true dynamic dashboard tool like Power BI.

? SAS (Option D): SAS is a statistical analysis software suite that excels in data

mining and analytics but is not primarily designed for creating dynamic, interactive dashboards.

The DA0-002 Visualization and Reporting domain emphasizes tools that facilitate "the appropriate visualization in the form of a report or dashboard with the proper design components," as noted in similar DA0-001 objectives (web ID: 1). Power BI aligns best with this requirement due to its focus on dynamic, user-friendly dashboard creation.

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

=====

NEW QUESTION 46

The human resources department wants to understand the relationship between the ages and incomes of all employees. Which of the following graphics is the most appropriate to present the analysis?

A. Scatter plot

B. Area plot

C. Bar chart

D. Pie chart

Answer: A

Explanation:

This question pertains to theVisualization and Reportingdomain, focusing on selecting the appropriate visualization to show a relationship between two continuous variables (ages and incomes).

? Scatter plot (Option A): A scatter plot displays individual data points on two axes

(age vs. income), making it ideal for showing the relationship and potential correlation between two continuous variables.

? Area plot (Option B): Area plots are used for showing trends over time, not

relationships between two variables.

? Bar chart (Option C): Bar charts are better for categorical data comparisons, not continuous variable relationships.

? Pie chart (Option D): Pie charts show proportions of a whole, not suitable for showing relationships between variables.

The DA0-002 Visualization and Reporting domain emphasizes "translating business requirements to form the appropriate visualization," and a scatter plot is best for showing the relationship between age and income.

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

=====

NEW QUESTION 47

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 51

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 56

A data analyst is creating a forecast for a product line introduced early last year. Which of the following should the analyst use to create projected sales and customer satisfaction for next year?

- A. Standard deviation and constraints
- B. Mean and median
- C. Boolean data and an array
- D. Numerical and ordinal attributes

Answer: D

Explanation:

This question pertains to the Data Analysis domain, focusing on data types and methods for forecasting. The task involves projecting sales (numerical) and customer satisfaction (likely ordinal, e.g., ratings), requiring appropriate data attributes.

? Standard deviation and constraints (Option A): Standard deviation measures data spread, and constraints are conditions, neither of which directly supports forecasting.

? Mean and median (Option B): Mean and median are descriptive statistics, not sufficient for forecasting future values.

? Boolean data and an array (Option C): Boolean data (true/false) and arrays (data structures) are not relevant for forecasting sales and satisfaction.

? Numerical and ordinal attributes (Option D): Sales are numerical (e.g., units sold), and customer satisfaction is often ordinal (e.g., 1-5 ratings). These attributes are suitable for forecasting models (e.g., time-series analysis for sales, regression for satisfaction).

The DA0-002 Data Analysis domain includes "applying the appropriate descriptive statistical methods," and numerical and ordinal attributes are key for forecasting sales and satisfaction.

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

=====

NEW QUESTION 61

Which of the following data repositories stores unaltered data?

- A. Data lake
- B. Data warehouse
- C. Data table
- D. Data factory

Answer: A

Explanation:

This question falls under the Data Concepts and Environments domain, focusing on data repositories. The task is to identify a repository that stores data in its original, unaltered form.

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

? Data warehouse (Option B): A data warehouse stores processed, structured data, often transformed for analysis, not unaltered.

? Data table (Option C): A data table is a structure within a database, not a repository, and may contain altered data.

? Data factory (Option D): A data factory (e.g., Azure Data Factory) is a data integration service, not a repository for storing data.

The DA0-002 Data Concepts and Environments domain includes understanding "different types of databases and data repositories," and a data lake is designed to store unaltered data.

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

=====

NEW QUESTION 64

A company wants to limit an employee's access to a production environment. Which of the following access control practices is the best to implement?

- A. Mandatory
- B. Time-based
- C. Attribute-based
- D. Role-based

Answer: D

Explanation:

This question falls under the Data Governance domain, focusing on access control practices for data security. The task is to limit an employee's access to a production environment, requiring a structured approach.

? Mandatory (Option A): Mandatory access control (MAC) uses strict, system-enforced rules (e.g., military settings), but it's overly rigid for most companies.

? Time-based (Option B): Time-based access limits access to specific times, which doesn't address general production environment access.

? Attribute-based (Option C): Attribute-based access control (ABAC) uses attributes (e.g., department, location), but it's complex and not the simplest solution.
? Role-based (Option D): Role-based access control (RBAC) assigns permissions based on the employee's role, ensuring they only access what's needed for their job, making it the best practice for limiting production access.
The DA0-002 Data Governance domain includes "data privacy concepts," and role-based access control is a widely adopted practice for limiting access in production environments. Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 5.0 Data Governance.
=====

NEW QUESTION 66

A data analyst is joining two tables with different content and one common field. Which of the following should the analyst do to most efficiently meet this requirement?

- A. Match the records of the related columns and merge the tables.
- B. Create a cluster to facilitate data integration between the tables.
- C. Explode both tables to identify unique values and reorder the fields in one table.
- D. Append the values of the matching columns and concatenate the other data fields.

Answer: A

Explanation:

This question falls under the Data Acquisition and Preparation domain, focusing on combining data from multiple tables. The tables have different content but share a common field, indicating a join operation.

? Match the records of the related columns and merge the tables (Option A): This describes a join operation, where records are matched on the common field (e.g., a key like Customer_ID) and the tables are merged, which is the most efficient method.

? Create a cluster to facilitate data integration between the tables (Option B): Clustering is a machine learning technique, not a method for joining tables.

? Explode both tables to identify unique values and reorder the fields in one table (Option C): Exploding is used in nested data (e.g., JSON arrays), and this approach is overly complex and unnecessary.

? Append the values of the matching columns and concatenate the other data fields (Option D): Appending stacks tables vertically, and concatenation applies to text, neither of which is appropriate for joining tables with a common field.

The DA0-002 Data Acquisition and Preparation domain includes "executing data manipulation," such as joining tables using a common field.
Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 2.0 Data Acquisition and Preparation.
=====

NEW QUESTION 69

A data analyst creates a report that identifies the middle 50% of the collected data. Which of the following best describes the analyst's findings?

- A. Interquartile range
- B. The difference between mode and median
- C. Mean variance
- D. Skewness from the slope

Answer: A

Explanation:

This question pertains to the Data Analysis domain, focusing on statistical measures. The middle 50% of a dataset refers to a specific statistical concept related to data distribution.

? Interquartile range (Option A): The interquartile range (IQR) is the range between the first quartile (Q1, 25th percentile) and the third quartile (Q3, 75th percentile), representing the middle 50% of the data, which matches the description.

? The difference between mode and median (Option B): This measures the spread between two central tendency metrics but doesn't represent the middle 50% of the data.

? Mean variance (Option C): Variance measures data dispersion around the mean, not the middle 50%.

? Skewness from the slope (Option D): Skewness measures data asymmetry, and "slope" is irrelevant here.

The DA0-002 Data Analysis domain includes "applying the appropriate descriptive statistical methods," and the IQR is the standard measure for the middle 50% of a dataset. Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 3.0 Data Analysis.
=====

NEW QUESTION 74

A data analyst team needs to segment customers based on customer spending behavior. Given one million rows of data like the information in the following sales order table:

Customer_ID
Region Amount_spent Product_category Quantity_of_items 00123
East 20000
Baby 4
00124
West 30000
Home 6
00125
South 40000
Garden 7
00126
North 50000
Furniture 8
00127
East
60000
Baby 10

Which of the following techniques should the team use for this task?

- A. Standardization
- B. Concatenate
- C. Binning
- D. Appending

Answer: C

Explanation:

This question falls under the Data Analysis domain, focusing on techniques for segmenting data. The task is to segment customers based on spending behavior, which involves grouping numerical data (Amount_spent) into categories.

? Standardization (Option A): Standardization scales numerical data to a common range (e.g., z-scores), but it doesn't segment customers into groups.

? Concatenate (Option B): Concatenation combines text fields, not numerical data for segmentation.

? Binning (Option C): Binning involves grouping numerical data into discrete intervals (e.g., low, medium, high spending), which is ideal for segmenting customers based on spending behavior.

? Appending (Option D): Appending combines datasets vertically, not relevant for segmentation.

The DA0-002 Data Analysis domain includes "applying the appropriate descriptive statistical methods," and binning is a common method for segmenting numerical data like spending amounts.

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

=====

NEW QUESTION 76

A data analyst is analyzing the following dataset:

Transaction Date Quantity

Item

Item Price

12/12/12

11

USB Cords 9.99

11/11/11

3

Charging Block

8.89

10/10/10

5

Headphones

50.15

Which of the following methods should the analyst use to determine the total cost for each transaction?

- A. Parsing
- B. Scaling
- C. Compressing
- D. Deriving

Answer: D

Explanation:

This question falls under the Data Analysis domain, focusing on calculating new values from existing data. The task is to determine the total cost per transaction, which involves multiplying Quantity by Item Price.

? Parsing (Option A): Parsing involves breaking down data (e.g., splitting a string), not calculating totals.

? Scaling (Option B): Scaling adjusts numerical values to a common range (e.g., normalization), not relevant for calculating totals.

? Compressing (Option C): Compressing reduces data size, not applicable to calculating costs.

? Deriving (Option D): Deriving involves creating new data fields by performing calculations on existing ones (e.g., Total Cost = Quantity * Item Price), which fits the task.

The DA0-002 Data Analysis domain includes "applying the appropriate descriptive statistical methods," such as deriving new fields through calculations to analyze data. Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 3.0 Data Analysis.

=====

NEW QUESTION 80

Given the following tables:

Individual table ID

FirstName LastName 1

John Doe Output

ID

FullName

1
JohnDoe
Which of the following is the best option to display output from FirstName and LastName as FullName?

A. Concatenate
B. Filter
C. Join
D. Group

Answer: A

Explanation:

This question falls under theData Acquisition and Preparationdomain of CompTIA Data+ DA0-002, focusing on data manipulation techniques. The task is to combine FirstName and LastName into a single FullName field (e.g., "JohnDoe").
? Concatenate (Option A): Concatenation combines two or more strings into one (e.g., usingCONCAT in SQL or "+" in Python), which is the correct method to create FullName from FirstName and LastName.
? Filter (Option B): Filtering selects specific rows based on conditions, not suitable for combining fields.
? Join (Option C): Joining combines data from multiple tables, but the task involves manipulating data within a single table.
? Group (Option D): Grouping (e.g., GROUP BY in SQL) is for aggregation, not for combining fields into a new column.
The DA0-002 Data Acquisition and Preparation domain includes "executing data manipulation," and concatenation is the standard technique for combining fields like FirstName and LastName into FullName.
Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 2.0 Data Acquisition and Preparation.
=====

NEW QUESTION 85

A manager needs a report to be sent by email every Monday for the next six months. Which of the following is the best way to accomplish this task?

A. Building self-service access
B. Creating a data snapshot
C. Developing a recurring process
D. Waiting for the request each week

Answer: C

Explanation:

This question falls under theVisualization and Reportingdomain, focusing on report delivery methods. The task requires a report to be emailed every Monday for six months, indicating a scheduled, repeating process.
? Building self-service access (Option A): Self-service allows users to generate reports on-demand, but the manager wants automatic delivery.
? Creating a data snapshot (Option B): A snapshot captures data at a specific point, not suitable for recurring delivery over six months.
? Developing a recurring process (Option C): A recurring process schedules the report to be generated and emailed every Monday, meeting the requirement for automated delivery over six months.
? Waiting for the request each week (Option D): This is manual and inefficient, not suitable for a six-month schedule.
The DA0-002 Visualization and Reporting domain includes "the appropriate visualization in the form of a report" with delivery methods, and a recurring process is ideal for scheduled email delivery.
Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 4.0 Visualization and Reporting.
=====

NEW QUESTION 86

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 89

Which of the following tables holds relational keys and numeric values?

- A. Fact
- B. Graph
- C. Dimensional
- D. Transactional

Answer: A

Explanation:

This question falls under the Data Concepts and Environments domain, focusing on understanding table types in data warehousing. The task is to identify a table that holds relational keys and numeric values, typically used in a star schema.

? Fact (Option A): Fact tables in a star schema store quantitative data (numeric values, e.g., sales amounts) and foreign keys (relational keys) linking to dimension tables, making this the correct choice.

? Graph (Option B): Graph tables are used in graph databases for relationships (e.g., nodes, edges), not typically for relational keys and numeric values in a traditional sense.

? Dimensional (Option C): Dimension tables store descriptive attributes (e.g., product names) and primary keys, not typically numeric measures.

? Transactional (Option D): Transactional tables are used in OLTP systems and may contain numeric values, but they're not specifically designed for relational keys in a data warehousing context.

The DA0-002 Data Concepts and Environments domain includes understanding "data schemas and dimensions," and fact tables are designed to hold relational keys and numeric values in a data warehouse.

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

=====

NEW QUESTION 91

A project manager requests an unscheduled report that provides a list of clients. Which of the following frequencies is best for this report?

- A. Annual
- B. Daily
- C. Weekly
- D. Ad hoc

Answer: D

Explanation:

This question pertains to the Visualization and Reporting domain, focusing on report delivery frequencies. The report is described as unscheduled, meaning it's a one-time request.

? Annual (Option A): Annual frequency implies a scheduled report every year, not suitable for an unscheduled request.

? Daily (Option B): Daily frequency implies a scheduled report each day, not suitable.

? Weekly (Option C): Weekly frequency implies a scheduled report each week, not suitable.

? Ad hoc (Option D): Ad hoc reports are generated on-demand for one-time or unscheduled needs, which matches the project manager's request.

The DA0-002 Visualization and Reporting domain includes "the appropriate visualization in the form of a report" with delivery methods, and ad hoc is the best frequency for an unscheduled report.

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

=====

NEW QUESTION 92

A company gives users adequate data access permissions to allow them to fulfill their duties but nothing more. Which of the following concepts best describes this practice?

- A. Active Directory
- B. Hierarchical access
- C. Zero Trust
- D. Least privilege

Answer: D

Explanation:

This question pertains to the Data Governance domain, focusing on data security and access control principles. The company restricts access to the minimum needed for duties, which aligns with a specific security concept.

? Active Directory (Option A): Active Directory is a tool for managing users and permissions, not a concept.

? Hierarchical access (Option B): Hierarchical access implies access based on roles in a hierarchy, but it doesn't specifically focus on minimal access.

? Zero Trust (Option C): Zero Trust requires continuous verification for all access, which is broader than just minimal permissions.

? Least privilege (Option D): Least privilege ensures users have only the permissions necessary for their duties, which matches the scenario. The DA0-002 Data Governance domain includes "data privacy concepts," and least privilege is a fundamental principle for secure access control. Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 5.0 Data Governance.

=====

NEW QUESTION 94

A developer builds an online survey that requires all questions to have an answer. Which of the following inconsistencies does this setting prevent?

- A. Missing values
- B. Duplication
- C. Data corruption
- D. Completeness

Answer: A

Explanation:

This question pertains to the Data Governance domain, focusing on data quality and consistency in survey design. Requiring all questions to have an answer ensures a specific type of data quality.

? Missing values (Option A): Requiring answers prevents missing values (NULLs or blanks) in the survey responses, which is the primary inconsistency this setting addresses.

? Duplication (Option B): Duplication refers to repeated records, not prevented by requiring answers.

? Data corruption (Option C): Data corruption involves damaged or altered data, not related to missing answers.

? Completeness (Option D): Completeness is the concept of having all necessary data, but "missing values" is the specific inconsistency prevented here.

The DA0-002 Data Governance domain includes "data quality control concepts," and preventing missing values ensures data integrity in survey responses.

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

=====

NEW QUESTION 96

A recent server migration applied an update to dataset naming conventions. Multiple users are now reporting stale information in an existing dashboard. The date in the dataset confirms a successful data refresh. Which of the following should a data analyst do first?

- A. Confirm the dashboard is pointed to the newest dataset.
- B. Filter the data in the dashboard.
- C. Escalate user permissions on the server.
- D. Verify that the dashboard subscription is not expired.

Answer: A

Explanation:

This question falls under the Data Governance domain, focusing on troubleshooting data freshness issues in dashboards. The dashboard shows stale data despite a successful refresh, and the server migration updated naming conventions, suggesting a potential mismatch.

? Confirm the dashboard is pointed to the newest dataset (Option A): The server migration updated dataset naming conventions, so the dashboard might still be pointing to an old dataset name, causing stale data. Confirming the dataset connection is the first step.

? Filter the data in the dashboard (Option B): Filtering might adjust the view but doesn't address the root cause of stale data.

? Escalate user permissions on the server (Option C): Permissions issues would likely prevent access, not cause stale data, especially since the dataset refreshed successfully.

? Verify that the dashboard subscription is not expired (Option D): An expired subscription might prevent access, but the dashboard is accessible, just showing stale data.

The DA0-002 Data Governance domain includes "data quality control concepts," such as ensuring dashboards connect to the correct, updated datasets after changes like server migrations.

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

=====

NEW QUESTION 100

Which of the following best describes the semi-structured data that is gathered when web scraping?

- A. JSON
- B. CSV
- C. CSS
- D. HTML

Answer: A

Explanation:

This question pertains to the Data Acquisition and Preparation domain, which in DA0-002 includes understanding data acquisition concepts and the types of data gathered from various sources, such as web scraping. Web scraping involves extracting data from websites, and the data gathered is often semi-structured, meaning it has some organizational structure but isn't fully relational like a database table.

? JSON (Option A): JSON (JavaScript Object Notation) is a semi-structured data format commonly used in web applications. Web scraping often retrieves data in JSON format via APIs or embedded scripts, as it's lightweight and structured with key-value pairs, making it ideal for semi-structured data.

? CSV (Option B): CSV (Comma-Separated Values) is a structured format typically used for tabular data. It's not commonly the direct output of web scraping, though scraped data might be converted to CSV later.

? CSS (Option C): CSS (Cascading Style Sheets) is used for styling web pages and isn't a data format, making it irrelevant for describing scraped data.

? HTML (Option D): HTML (HyperText Markup Language) is the structure of web pages and is often the raw format scraped during web scraping. While HTML is semi-structured due to its tag-based hierarchy, it's primarily a markup language, not a data format, and the actual data extracted is often parsed into formats like JSON.

The DA0-002 Data Acquisition and Preparation domain aligns with the DA0-001 focus on "data acquisition concepts" (web ID: 14), which includes identifying

formats like JSON as semi-structured data commonly acquired through web scraping. JSON is the best fit here due to its prevalence in web data exchange.
Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 2.0 Data Acquisition and Preparation
=====

NEW QUESTION 101

Due to new reporting requirements, a data analyst must add new classification codes to historical data. Which of the following is the best technique for this task?

- A. Append
- B. Binning
- C. Parsing
- D. Union

Answer: A

Explanation:

This question falls under the Data Acquisition and Preparation domain, focusing on modifying historical data. The task is to add new classification codes to existing data, which involves adding new rows or columns.

? Append (Option A): Appending adds new rows to a dataset, which is suitable if the classification codes are new records (e.g., a new table of codes to combine with historical data). If the codes are a new column, a join or update might be used, but append fits the context of adding new data.

? Binning (Option B): Binning groups data into categories, not suitable for adding classification codes.

? Parsing (Option C): Parsing breaks down data (e.g., splitting strings), not relevant for adding codes.

? Union (Option D): Union stacks tables with identical structures, but the task involves adding new data (codes) to historical data, not combining identical tables.

The DA0-002 Data Acquisition and Preparation domain includes "executing data manipulation," and appending is a common technique for adding new data to historical datasets.

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

NEW QUESTION 106

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 108

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 113

A data analyst receives a new data source that contains employee IDs, job titles, dates of birth, addresses, years of service, and employees' birth months. Which of the following inconsistencies should the analyst identify?

- A. Redundancy
- B. Equivalence
- C. Parallel
- D. Duplication

Answer: A

Explanation:

This question falls under the Data Governance domain, focusing on identifying data quality issues. The dataset includes dates of birth and birth months, which suggests a potential inconsistency.

? Redundancy (Option A): The dataset includes both dates of birth (e.g., 1990-05-

15) and birth months (e.g., May), which is redundant because the birth month can be derived from the date of birth, indicating a data quality issue.

? Equivalence (Option B): Equivalence isn't a standard data quality term in this context; it might refer to data matching, which isn't the issue here.

? Parallel (Option C): Parallel isn't a recognized data quality term; it might relate to processing, not data inconsistencies.

? Duplication (Option D): Duplication refers to identical records, but the issue here is redundant fields, not duplicate rows.

The DA0-002 Data Governance domain includes "data quality control concepts," and redundancy is a key inconsistency when the same information is stored in multiple forms unnecessarily.

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

=====

NEW QUESTION 114

Which of the following best explains the purpose of data lineage?

- A. To see the steps and path of data flow through different systems
- B. To better understand the granularity of data variable relationships
- C. To track data transformations from acquisition through reporting
- D. To look up data definitions, ensuring consistent use across business units

Answer: C

Explanation:

This question pertains to the Data Concepts and Environments domain, focusing on the purpose of data lineage. Data lineage involves tracking the lifecycle of data.

? To see the steps and path of data flow through different systems (Option A): This describes a data flow diagram, not data lineage, which focuses on transformations rather than just flow.

? To better understand the granularity of data variable relationships (Option B): This relates to data modeling, not the purpose of data lineage.

? To track data transformations from acquisition through reporting (Option C): Data lineage tracks the journey of data, including transformations (e.g., cleaning, aggregation) from its source to its final use in reporting, which is its primary purpose.

? To look up data definitions, ensuring consistent use across business units (Option D): This describes a data dictionary, not data lineage.

The DA0-002 Data Concepts and Environments domain includes understanding "data schemas and dimensions," and data lineage specifically tracks transformations across the data lifecycle.

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

NEW QUESTION 117

A grocery store wants to view the revenue from the previous year, highlighting individual departments. Which of the following is the most appropriate chart to communicate this information?

- A. Gantt
- B. Pie
- C. Area
- D. Line

Answer: B

Explanation:

This question is part of the Visualization and Reporting domain, focusing on selecting the appropriate visualization for a given dataset. The grocery store wants to view revenue by department, which requires a chart that shows proportions or comparisons across categories.

? Gantt (Option A): Gantt charts are used for project scheduling, not for comparing revenue across categories.

? Pie (Option B): Pie charts are ideal for showing proportions or percentages of a whole, such as revenue distribution across departments, making this the best choice.

? Area (Option C): Area charts are better for showing trends over time, not static categorical comparisons.

? Line (Option D): Line charts are used for trends over time, not for comparing discrete categories like departments.

The DA0-002 Visualization and Reporting domain emphasizes "translating business requirements to form the appropriate visualization", and a pie chart is the most appropriate for showing departmental revenue proportions.

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

=====

NEW QUESTION 118

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

ID	Type	ID	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?

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

Answer: A

Explanation:
This question is part of theData Concepts and Environmentsdomain, 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.
Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 1.0 Data Concepts and Environments.
=====

NEW QUESTION 120

A data analyst deployed a report for public access. A user states that the report is not showing the latest information, even though the user updated the source an hour ago. Which of the following should the data analyst check first?

- A. Event log
- B. User privileges
- C. Database connection
- D. Report corruption

Answer: C

Explanation:
This question pertains to theData Governancedomain, focusing on troubleshooting data freshness issues in reports. The report isn't showing the latest data despite a recent source update, indicating a potential refresh or connectivity issue.
? Event log (Option A): Event logs might provide insight into errors, but they're not the first step for checking data freshness.
? User privileges (Option B): Privileges might affect access, but the user can see the report, so this isn't the issue.
? Database connection (Option C): If the database connection failed or isn't refreshing properly, the report won't reflect the latest data, making this the first thing to check.
? Report corruption (Option D): Corruption might cause errors, but it's less likely than a connectivity issue for this scenario.
The DA0-002 Data Governance domain includes "data quality control concepts," such as ensuring data freshness by verifying database connections.
Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 5.0 Data Governance.
=====

NEW QUESTION 123

.....

Relate Links

100% Pass Your DA0-002 Exam with ExamBible Prep Materials

<https://www.exambible.com/DA0-002-exam/>

Contact us

We are proud of our high-quality customer service, which serves you around the clock 24/7.

Viste - <https://www.exambible.com/>