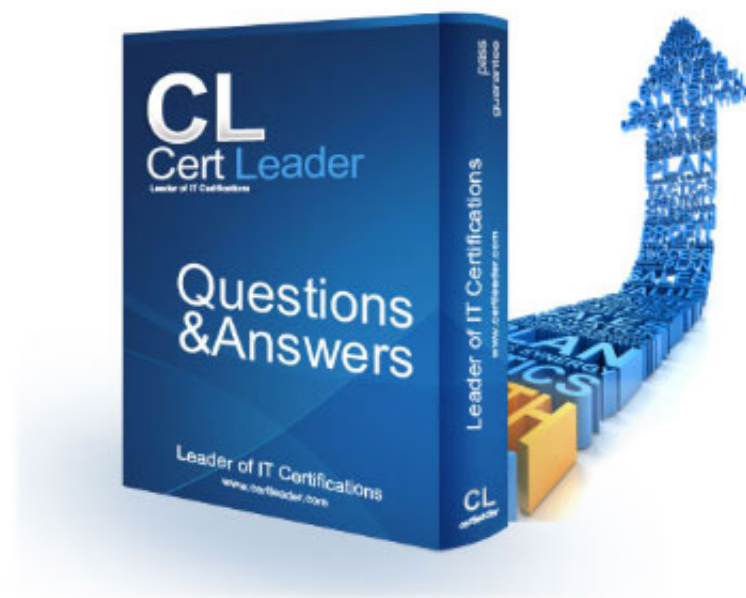


DA0-002 Dumps

CompTIA Data+ Exam (2025)

<https://www.certleader.com/DA0-002-dumps.html>



NEW QUESTION 1

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 2

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 3

A data analyst created a dashboard to illustrate the traffic volume and mean response time for a call center. The traffic data is current, but the mean response time has not updated for more than an hour. Which of the following is the best way to verify the data's freshness?

- A. Refactoring the code base
- B. Testing for network connectivity issues
- C. Checking the last time the calculation script ran
- D. Determining the number of calls with no timestamps

Answer: C

Explanation:

This question pertains to the Data Governance domain, which in DA0-002 includes ensuring data quality and freshness, especially in dashboards. The issue is that the mean response time isn't updating, while traffic data is current, indicating a potential issue with the data refresh process for the response time metric.

? Refactoring the code base (Option A): Refactoring might improve long-term performance but doesn't directly address verifying data freshness.

? Testing for network connectivity issues (Option B): Network issues could cause delays, but since traffic data is updating, connectivity is likely not the issue.

? Checking the last time the calculation script ran (Option C): Mean response time is a calculated metric, likely derived from a script. Checking when the script last ran directly verifies if the data refresh process failed, making this the best approach.

? Determining the number of calls with no timestamps (Option D): Missing timestamps might indicate data quality issues, but it doesn't directly verify why the mean response time isn't updating.

The DA0-002 Data Governance domain focuses on "data quality control concepts," including ensuring data freshness in reporting. Checking the script's last run time aligns with this objective.

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

=====

NEW QUESTION 4

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 5

A data analyst needs to provide a weekly sales report for the Chief Financial Officer. Which of the following delivery methods is the most appropriate?

- A. A granular daily report in a dashboard
- B. A detailed text document
- C. A spreadsheet with raw data
- D. A high-level email

Answer: D

Explanation:

This question pertains to the Visualization and Reporting domain, focusing on report delivery methods for a specific audience. The Chief Financial Officer (CFO) needs a weekly sales report, suggesting a concise, executive-level summary.

? A granular daily report in a dashboard (Option A): Daily granularity is too frequent for a weekly report, and a dashboard might be too interactive for a CFO's needs.

? A detailed text document (Option B): A detailed document is too lengthy for a CFO, who typically needs a summary.

? A spreadsheet with raw data (Option C): Raw data requires further analysis, which isn't appropriate for an executive-level report.

? A high-level email (Option D): A high-level email provides a concise summary of weekly sales, tailored for an executive like a CFO, making it the most appropriate delivery method.

The DA0-002 Visualization and Reporting domain emphasizes "translating business requirements to form the appropriate visualization," and a high-level email is best for delivering a weekly summary to a CFO.

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

NEW QUESTION 6

A data analyst needs to modify a dashboard that was created by another employee. Upon opening the dashboard, the analyst notices that the information is not loading properly. Which of the following should the analyst do to troubleshoot this error?

- A. Review the data layer and data source.
- B. Validate that the database is up-to-date.
- C. Check that the program is updated to the latest version.
- D. Ensure the correct filters are displaying on the dashboard.

Answer: A

Explanation:

This question falls under the Data Governance domain of CompTIA Data+ DA0-002, focusing on troubleshooting data quality issues in dashboards. The dashboard isn't loading properly, indicating a potential issue with the data connection or configuration.

? Review the data layer and data source (Option A): The data layer (e.g., queries, connections) and data source (e.g., database) are the foundation of a dashboard. If the information isn't loading, the issue likely lies in the data connection or query configuration, making this the first step.

? Validate that the database is up-to-date (Option B): While this might be a subsequent step, it assumes the connection is working, which should be confirmed first.

? Check that the program is updated to the latest version (Option C): Software updates might fix bugs, but this isn't the most immediate cause of data not loading.

? Ensure the correct filters are displaying on the dashboard (Option D): Filters affect data display, but if the data isn't loading at all, the issue is more fundamental.

The DA0-002 Data Governance domain includes "data quality control concepts," and reviewing the data layer and source is the primary step in troubleshooting dashboard loading issues.

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

=====

NEW QUESTION 7

Which of the following best describes the method used to combine files, software, and libraries for use on various operating systems and environments?

- A. Package manager
- B. Code repository
- C. Virtual machine
- D. Containerization

Answer: D

Explanation:

This question pertains to the Data Concepts and Environments domain, focusing on methods for managing software and data environments. The task is to identify a method that combines files, software, and libraries for use across different systems.

? Package manager (Option A): Package managers (e.g., npm) manage software dependencies but don't combine files and libraries for cross-system use.

? Code repository (Option B): Code repositories (e.g., GitHub) store code but don't package it for deployment across environments.

? Virtual machine (Option C): Virtual machines emulate entire operating systems, which is heavier than needed for combining files and libraries.

? Containerization (Option D): Containerization (e.g., Docker) packages files, software, and libraries into a container that can run consistently across different operating systems and environments, making it the best choice.

The DA0-002 Data Concepts and Environments domain includes understanding "data environments," and containerization is a standard method for ensuring consistency across systems.

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

=====

NEW QUESTION 8

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 9

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 10

A data company needs a visualization that shows the availability zones from the last ten years and any future availability zones that the company will be using in the next five years. Which of the following is the most appropriate visualization to display this information?

- A. Bar chart
- B. Mosaic plot
- C. Map
- D. Pie chart

Answer: C

Explanation:

This question falls under the Visualization and Reporting domain of CompTIA Data+ DA0-002, focusing on selecting the appropriate visualization for a specific dataset. The task is to show availability zones over a 15-year period (past and future), which involves a geographical element since availability zones are typically location-based.

? Bar chart (Option A): Bar charts are good for comparing categorical data but don't

effectively show geographical locations or time-based trends across zones.

? Mosaic plot (Option B): Mosaic plots display relationships between categorical variables, not suitable for geographical or time-series data.

? Map (Option C): A map can display availability zones geographically, with annotations or layers to show changes over time (past 10 years and future 5 years), making it the most appropriate visualization.

? Pie chart (Option D): Pie charts show proportions of a whole, not suitable for geographical or time-based data.

The DA0-002 Visualization and Reporting domain emphasizes "translating business requirements to form the appropriate visualization," and a map is the best choice for displaying geographical availability zones over time.

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

=====

NEW QUESTION 10

A data professional wants to identify all customers who made a purchase in January. Given the following table:

CustomerID Month Sales

0001

January 13000

0002

March 10000

0003

April 23000

0004

May 10000

Which of the following types of functions should the professional use to flag the customers?

- A. Statistical
- B. Logical
- C. Mathematical
- D. Date

Answer: B

Explanation:

This question falls under the Data Analysis domain, focusing on selecting the appropriate function type to filter data in a query. The task is to flag customers who made a purchase in January, which involves a conditional check.

? Statistical (Option A): Statistical functions (e.g., AVG, STDEV) analyze data distributions, not suitable for flagging specific months.

? Logical (Option B): Logical functions (e.g., WHERE Month = 'January' in SQL) are used to apply conditions and flag rows based on criteria, which fits the task.

? Mathematical (Option C): Mathematical functions (e.g., SUM, ROUND) perform calculations, not conditional flagging.

? Date (Option D): Date functions (e.g., MONTH()) manipulate dates, but the Month column is already in text format, so a logical comparison is sufficient.

The DA0-002 Data Analysis domain includes "applying the appropriate descriptive statistical methods using SQL queries," and logical functions are best for conditional flagging.

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

=====

NEW QUESTION 11

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 16

A senior manager needs a report that can be generated and accessed at any time. Which of the following delivery methods should a data analyst use?

- A. Ad hoc

- B. Dynamic
- C. Self-service
- D. Static

Answer: C

Explanation:

This question is part of the Visualization and Reporting domain, focusing on report delivery methods. The requirement for a report that can be generated and accessed at any time suggests user-driven access.

? Ad hoc (Option A): Ad hoc reports are one-time, on-demand reports, not designed for anytime access by the user.

? Dynamic (Option B): Dynamic reports allow interactivity but don't necessarily imply user-generated access at any time.

? Self-service (Option C): Self-service reporting allows users (e.g., the senior manager) to generate and access reports on-demand at any time, fitting the requirement.

? Static (Option D): Static reports are fixed and don't allow on-demand generation by the user.

The DA0-002 Visualization and Reporting domain includes "the appropriate visualization in the form of a report" with delivery methods, and self-service is ideal for anytime access. Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 4.0 Visualization and Reporting.

=====

NEW QUESTION 20

A data analyst wants to use the following tables to find all the customers who have not placed an order:

Customers table ID

Name Address

Products table ID

Name Customer_ID

Which of the following SQL statements is the best way to accomplish this task?

- A. SELECT * FROM CUSTOMERS AS C LEFT JOIN PRODUCTS AS P ON C.ID = P.Customer_ID WHERE P.Customer_ID IS NULL
- B. SELECT * FROM CUSTOMERS AS C INNER JOIN PRODUCTS AS P ON C.ID = C.ID WHERE COUNT(P.) = 0
- C. SELECT * FROM PRODUCTS AS P INNER JOIN CUSTOMERS AS C ON P.Customer_ID = C.ID WHERE (SELECT COUNT(P.) = 0)
- D. SELECT * FROM PRODUCTS AS P LEFT JOIN CUSTOMERS AS C ON P.Customer_ID = C.ID WHERE P.Customer_ID IS NOT NULL

Answer: A

Explanation:

This question pertains to the Data Analysis domain, focusing on SQL queries to analyze data relationships. The task is to find customers who have not placed an order, meaning customers in the Customers table without a matching Customer_ID in the Products table.

? Option A: SELECT * FROM CUSTOMERS AS C LEFT JOIN PRODUCTS AS P

ON C.ID = P.Customer_ID WHERE P.Customer_ID IS NULL A LEFT JOIN includes all customers, even those without orders (where Products columns are NULL). Filtering with WHERE P.Customer_ID IS NULL selects only customers without a match in Products, correctly identifying those who haven't ordered.

? Option B: SELECT * FROM CUSTOMERS AS C INNER JOIN PRODUCTS AS P

ON C.ID = C.ID WHERE COUNT(P.*) = 0 An INNER JOIN only includes matching records, so it won't return customers without orders. The join condition C.ID = C.ID is also incorrect, and COUNT requires a GROUP BY, making this invalid.

? Option C: SELECT * FROM PRODUCTS AS P INNER JOIN CUSTOMERS AS C ON P.Customer_ID = C.ID WHERE (SELECT COUNT(P.*) = 0) An INNER JOIN

excludes customers without orders, and the subquery syntax is incorrect (COUNT needs a GROUP BY or to be part of a HAVING clause).

? Option D: SELECT * FROM PRODUCTS AS P LEFT JOIN CUSTOMERS AS C

ON P.Customer_ID = C.ID WHERE P.Customer_ID IS NOT NULL This starts with Products and joins Customers, returning only records with orders (opposite of the task), and IS NOT NULL further excludes non-ordering customers.

The DA0-002 Data Analysis domain includes "applying the appropriate descriptive statistical methods using SQL queries," and a LEFT JOIN with a NULL check is the standard method for finding non-matching records.

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

=====

NEW QUESTION 24

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 28

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 31

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 36

A company's analytics manager wants all reports to be delivered once every seven days. Which of the following is the best delivery method?

- A. Recurring
- B. Ad hoc
- C. Custom

D. Snapshot

Answer: A

Explanation:

This question pertains to the Visualization and Reporting domain, focusing on report delivery methods. The requirement for delivery every seven days indicates a scheduled, repeating process.

? Recurring (Option A): Recurring delivery schedules reports to be generated and delivered at regular intervals (e.g., weekly), which matches the requirement of every seven days.

? Ad hoc (Option B): Ad hoc reports are one-time, on-demand reports, not suitable for scheduled delivery.

? Custom (Option C): Custom isn't a standard delivery method; it might refer to tailored reports but doesn't imply scheduling.

? Snapshot (Option D): A snapshot captures data at a specific point, not suitable for recurring delivery.

The DA0-002 Visualization and Reporting domain includes "the appropriate visualization in the form of a report" with delivery methods, and recurring delivery is ideal for weekly reports.

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

=====

NEW QUESTION 37

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 the Data Acquisition and Preparation domain 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., using CONCAT 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 40

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 43

A data breach occurs at a company. Which of the following actions should be taken?

A. Make an announcement on social media so customers are aware as soon as possible.

B. Tell the company management team and then tell regulatory agencies.

C. Keep the incident a secret until the issue is resolved.

D. Inform the entire IT sector, but ask for discretion.

Answer: B

Explanation:

This question falls under the Data Governancedomain, focusing on data breach response protocols. A data breach requires a structured response to comply with legal and regulatory requirements.

? Make an announcement on social media so customers are aware as soon as

possible (Option A): Public announcement without internal coordination or

regulatory notification can lead to legal issues and loss of trust.

? Tell the company management team and then tell regulatory agencies (Option B): This follows best practices: inform internal leadership to coordinate a response, then notify regulatory agencies as required by laws (e.g., GDPR mandates notification within 72 hours).

? Keep the incident a secret until the issue is resolved (Option C): This violates regulations requiring timely breach notification.

? Inform the entire IT sector, but ask for discretion (Option D): Sharing with the IT sector is vague and risks leaks; regulatory agencies should be prioritized.

The DA0-002 Data Governance domain includes "data privacy concepts," such as proper breach response procedures, emphasizing internal and regulatory notification. Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 5.0 Data Governance.

=====

NEW QUESTION 47

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 Governancedomain, 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 51

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 Preparationdomain, 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 56

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 60

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 62

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 65

A sales manager wants to understand how sales are trending year over year. Which of the following chart types is the most appropriate to display the information?

- A. Line
- B. Donut
- C. Bubble
- D. Hierarchy

Answer: A

Explanation:

This question falls under the Visualization and Reporting domain, focusing on selecting the appropriate visualization for a specific data trend. The task is to show sales trends over time (year over year).

? Line (Option A): Line charts are ideal for displaying trends over time, such as year-over-year sales, as they clearly show changes and patterns across a continuous time axis.

? Donut (Option B): Donut charts show proportions or percentages of a whole, not suitable for time-based trends.

? Bubble (Option C): Bubble charts display three dimensions of data (e.g., size, x- axis, y-axis), not ideal for simple time trends.

? Hierarchy (Option D): Hierarchy charts (e.g., treemaps) show nested relationships, not time-based trends.

The DA0-002 Visualization and Reporting domain emphasizes "translating business requirements to form the appropriate visualization," and a line chart is best for time-series trends.

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

=====

NEW QUESTION 69

Which of the following is the most efficient to use when programming repeatable tasks?

- A. LLM
- B. Deep learning
- C. NLP
- D. RPA

Answer: D

Explanation:

This question pertains to the Data Acquisition and Preparation domain, focusing on tools for automating repeatable tasks in data processes. The task is to identify the most efficient programming method.

? LLM (Option A): Large Language Models (e.g., GPT) are for text generation, not efficient for repeatable data tasks.

? Deep learning (Option B): Deep learning is for complex pattern recognition (e.g., image classification), not efficient for simple repeatable tasks.

? NLP (Option C): Natural Language Processing is for text analysis, not general repeatable tasks.

? RPA (Option D): Robotic Process Automation (RPA) automates repetitive, rule-based tasks (e.g., data entry, file processing), making it the most efficient for programming repeatable tasks.

The DA0-002 Data Acquisition and Preparation domain includes "executing data manipulation," and RPA is a widely used method for automating repeatable data tasks. Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 2.0 Data Acquisition and Preparation.

=====

NEW QUESTION 71

.....

Thank You for Trying Our Product

* 100% Pass or Money Back

All our products come with a 90-day Money Back Guarantee.

* One year free update

You can enjoy free update one year. 24x7 online support.

* Trusted by Millions

We currently serve more than 30,000,000 customers.

* Shop Securely

All transactions are protected by VeriSign!

100% Pass Your DA0-002 Exam with Our Prep Materials Via below:

<https://www.certleader.com/DA0-002-dumps.html>