CompTIA

Exam Questions DA0-002

CompTIA Data+ Exam (2025)

NEW QUESTION 1

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

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

Answer: D

Explanation:

This question pertains to theData Governancedomain, focusing on troubleshooting andmaintaining data quality in software processes. The analyst has already tried basic debugging and confirmed compatibility, so the next step involves seeking external resources.
? Use the old software and preexisting code, since both were functional (Option A):
Reverting to old software avoids solving the issue and may introduce other risks (e.g., security vulnerabilities).
? Contact the IT department and inform them that the software has a bug (Option B):
Assuming a bug without further investigation is premature, especially since compatibility was confirmed.
? Escalate to the department manager and ask for assistance (Option C): Escalation
is a later step after exhausting technical troubleshooting options.
? Research the issue online and see if a solution is available (Option D): Researching online (e.g., forums, documentation) is a logical next step to find solutions or identify known issues, especially after local debugging fails.
The DA0-002 Data Governance domain includes "data quality control concepts," and researching online is a standard troubleshooting step to maintain data process integrity. Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 5.0 Data Governance.
==============

NEW QUESTION 2

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

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

Answer: A

Explanation:

This question pertains to theData Analysisdomain, focusing on date calculations. The task is to calculate the difference between dates in a file and the current date, requiring proper date handling.
? Convert data to date format and use date functions (Option A): Flat files often store
dates as strings (e.g., "2023-01-01"). Converting them to a date format (e.g., using Python??s datetime or SQL??s TO_DATE) allows the use of date functions (e.g., DATEDIFF) to calculate the difference to the current date, which is the best approach.
? Validate the date format with logical functions and use date functions to analyze
(Option B): Validation is unnecessary if conversion handles format issues, making this overly complex.
? Use date functions to analyze the data with no conversion (Option C): Without
converting to a date format, date functions may fail if the data is stored as strings.
? Transform data to a numerical value and use mathematical functions (Option D): This is inefficient and error-prone compared to using date functions.
The DA0-002 Data Analysis domain includes "applying the appropriate descriptive statistical methods," and converting to date format followed by date functions is the standard method for such calculations.
Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 3.0 Data Analysis.
==============

NEW QUESTION 3

Which of the following is business intelligence software?

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

Answer: D

Explanation:

This question falls under theVisualization and Reportingdomain, focusing on identifying tools used for business intelligence (BI), which typically involves data visualization and reporting.
? SAS (Option A): SAS is a statistical analysis software, not primarily a BI tool
focused on visualization.
? Python (Option B): Python is a programming language, not a BI software, though it can be used for data analysis with libraries.
? Notepad++ (Option C): Notepad++ is a text editor, not a BI tool.
? Tableau (Option D): Tableau is a leading BI software designed for data visualization, dashboards, and reporting, making it the correct choice.
The DA0-002 Visualization and Reporting domain includes understanding "the appropriate visualization in the form of a report or dashboard," and Tableau is a recognized BI tool for this purpose.
Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 4.0 Visualization and
Reporting.
==============

NEW QUESTION 4
Which of the following best describes the function of a data type?

A. To provide a generic identifier for files used in analysis
B. To identify the program needed to open a file
C. To differentiate the real value of the field in its context
D. To make the addition of individual records simpler

Answer: C

Explanation:
This question falls under theData Concepts and Environmentsdomain, focusing on the purpose of data types in data management. Data types define how data is stored and interpreted.
? To provide a generic identifier for files used in analysis (Option A): Data types
apply to fields within datasets, not files.
? To identify the program needed to open a file (Option B): File extensions (e.g.,
.csv) identify programs, not data types.
? To differentiate the real value of the field in its context (Option C): Data types (e.g.,
integer, string, date) define how a field??s value is interpreted (e.g., ??123?? as a number vs. text), ensuring its real meaning in context, making this the correct answer.
? To make the addition of individual records simpler (Option D): Data types don??t
directly simplify record addition; they ensure proper data handling.
The DA0-002 Data Concepts and Environments domain includes understanding "data schemas and dimensions," and data types ensure fields are interpreted correctly in their context.
Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 1.0 Data Concepts and Environments.
==============

NEW QUESTION 5
A data analyst receives four files that need to be unified into a single spreadsheet for further analysis. All of the files have the same structure, number of columns, and field names, but each file contains different values. Which of the following methods will help the analyst convert the files into a single spreadsheet?

A. Merging
B. Appending
C. Parsing
D. Clustering

Answer: B

Explanation:
This question is part of theData Acquisition and Preparationdomain, which involves combining data from multiple sources. The files have the same structure but different
values, meaning theyneed to be stacked vertically into one dataset.
? Merging (Option A): Merging typically involves joining datasets on a common key (e.g., a customer ID), which isn??t indicated here since the files only differ in values, not keys.
? Appending (Option B): Appending stacks datasets vertically, combining rows from files with the same structure into a single dataset, which matches the scenario.
? Parsing (Option C): Parsing involves breaking down data (e.g., splitting text), not combining files.
? Clustering (Option D): Clustering is a machine learning technique for grouping similar data points, not for combining files.
The DA0-002 Data Acquisition and Preparation domain includes "executing data manipulation," such as appending datasets with identical structures.
Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 2.0 Data Acquisition and Preparation.
==============

NEW QUESTION 6
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 theData Governancedomain, 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 7
An analyst needs to produce a final dataset using the following tables:
CourseID SectionNumber StudentID MATH1000
1
10009
MATH1000 2

10007
PSYC1500 1
10009
PSYC1500 1
10015
StudentID FirstName LastName
10009
Jane Smith
10007
John Doe
10015
Robert Roe
The expected output should be formatted as follows:
| CourseID | SectionNumber | StudentID | FirstName | LastName |
Which of the following actions is the best way to produce the requested output?

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

Answer: B

Explanation:
This question falls under theData Acquisition and Preparationdomain, focusing on combining tables to produce a dataset. The task requires combining the Courses and Students tables to include student names with course details, based on the StudentID.
? Aggregate (Option A): Aggregation (e.g., SUM, COUNT) summarizes data, not
suitable for combining tables to include names.
? Join (Option B): A join operation (e.g., INNER JOIN on StudentID) combines the tables, matching records to produce the requested output with CourseID, SectionNumber, StudentID, FirstName, and LastName.
? Group (Option C): Grouping is used for aggregation (e.g., GROUP BY in SQL), not for combining tables.
? Filter (Option D): Filtering selects specific rows, not relevant for combining tables. The DA0-002 Data Acquisition and Preparation domain includes "executing data manipulation," such as joining tables to create a unified dataset.
Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 2.0 Data Acquisition and Preparation.
==============

NEW QUESTION 8
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 theData Analysisdomain, 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 9
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 thefollowing 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 theVisualization and Reportingdomain 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
SIMULATION
The director of operations at a power company needs data to help identify where company resources should be allocated in order to monitor activity for outages and restoration of power in the entire state. Specifically, the director wants to see the following:
* County outages
* Status
* Overall trend of outages
INSTRUCTIONS:
Please, select each visualization to fit the appropriate space on the dashboard and choose an appropriate color scheme. Once you have selected all visualizations, please, select the appropriate titles and labels, if applicable. Titles and labels may be used more than once.
If at any time you would like to bring back the initial state of the simulation, please click the Reset All button.

Dashboard Editor

Dashboard Editor

Select a dashboard title
- Power Outages Enterprise-wide
- Power Outages Over Time
- EnPOWER Me! Dashboard
- Outages in Sheridan County

A. Mastered
B. Not Mastered

Answer: A

Explanation:
This is a simulation question that requires you to create a dashboard with visualizations that meet the director??s needs.Here are the steps to complete the task:
? Drag and drop the visualization that shows the county outages on the top left
space of the dashboard. This visualization is a map of the state with different colors indicating the number of outages in each county. You can choose any color scheme that suits your preference, but make sure that the colors are consistent and clear. For example, you can use a gradient of red to show the counties with more outages and green to show the counties with less outages.
? Drag and drop the visualization that shows the status of the outages on the top
right space of the dashboard. This visualization is a pie chart that shows the percentage of outages that are active, restored, or pending. You can choose any color scheme that suits your preference, but make sure that the colors are distinct and easy to identify. For example, you can use red for active, green for restored, and yellow for pending.
? Drag and drop the visualization that shows the overall trend of outages on the
bottom space of the dashboard. This visualization is a line graph that shows the number of outages over time. You can choose any color scheme that suits your preference, but make sure that the color is visible and contrasted with the background. For example, you can use blue for the line and white for the background.
? Select appropriate titles and labels for each visualization. Titles and labels may be
used more than once. For example, you can use ??County Outages?? as the title for the map, ??Status?? as the title for the pie chart, and ??Trend?? as the title for the line graph. You can also use ??County??, ??Number of Outages??, ??Active??, ??Restored??, ??Pending??, ??Time??, and ??Number of Outages?? as labels for the axes and legends of the visualizations.

NEW QUESTION 10
Which of the following allows a data analyst to send out a spreadsheet containing sensitive information without revealing personal details?

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

Answer: B

Explanation:
This question pertains to theData Governancedomain, focusing on data privacy and security. The task is to share a spreadsheet with sensitive information while protecting personal details.
? Using a UUID in the data file (Option A): A UUID (Universally Unique Identifier)
can anonymize records, but if other PII (e.g., names) remains, personal details are still exposed.
? Redacting all PII (Option B): Redacting personally identifiable information (PII)
removes sensitive details (e.g., names, addresses), ensuring personal information isn??t revealed while sharing the spreadsheet.
? Adding access controls to the ID column (Option C): Access controls limit who can
view the data, but the question focuses on the spreadsheet content itself, not access.
? Encrypting the spreadsheet (Option D): Encryption protects the file during
transmission, but once opened, personal details are still visible unless redacted. The DA0-002 Data Governance domain includes "data privacy concepts," and redacting PII is the most direct method to protect personal details in a shared spreadsheet.
Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 5.0 Data Governance.
==============

NEW QUESTION 14
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 theData Governancedomain, 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 16
Which of the following data repositories should a company use when structured data about the whole company needs to be stored in a predefined data structure?

A. Data mart
B. Data warehouse
C. Data silo

D. Data lake

Answer: B

Explanation:
This question pertains to theData Concepts and Environmentsdomain, focusing on selecting the appropriate repository for structured data across an entire company. The requirement for a predefined structure narrows the options.
? Data mart (Option A): A data mart stores structured data for a specific business
area (e.g., sales), not the whole company.
? Data warehouse (Option B): A data warehouse is designed to store structured data from across the entire company in a predefined schema, optimized for analytics and reporting.
? Data silo (Option C): A data silo is an isolated repository, often structured, but not designed for company-wide integration.
? Data lake (Option D): A data lake stores raw data (structured and unstructured) without a predefined structure, not suitable for this requirement.
The DA0-002 Data Concepts and Environments domain includes understanding "different types of databases and data repositories," and a data warehouse is ideal for company-wide structured data.
Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 1.0 Data Concepts and Environments.
==============

NEW QUESTION 18
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 theData Analysisdomain, 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 23
A data analyst must combine service calls into low-, medium-, and high-priority levels in order to analyze organizational responses. Which of the following techniques should the analyst use for this task?

A. Augmentation
B. Imputation
C. Scaling
D. Binning

Answer: D

Explanation:
This question pertains to theData Analysisdomain, focusing on techniques for categorizing data. The task involves grouping service calls into priority levels (low, medium, high), which requires segmenting numerical or ordinal data into discrete categories.
? Augmentation (Option A): Augmentation involves adding data (e.g., in machine
learning), not categorizing existing data.
? Imputation (Option B): Imputation fills in missing values, not relevant for categorizing priority levels.
? Scaling (Option C): Scaling adjusts numerical data to a common range (e.g., normalization), not suitable for creating priority categories.
? Binning (Option D): Binning groups continuous or ordinal data into discrete categories (e.g., assigning calls to low, medium, or high priority based on a metric like response time), which fits the task.
The DA0-002 Data Analysis domain includes "applying the appropriate descriptive statistical methods," and binning is a standard technique for categorizing data for analysis. Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 3.0 Data Analysis.
==============

NEW QUESTION 26
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 theVisualization and Reportingdomain, 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 30
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 IDDate of birth Address CityState Frank 235253/191234 Harding ChicagoILMartha 112906/13935 Terrace Hills ChinoCAEllen 1214111/42 Speedway MiamiFL
B. Name IDDate of birth Address CityState Frank 235253/191234 Harding ChicagoILMartha112906/13935 Terrace Hills ChinoCAEllen 1214111/42 Speedway MiamiFL
C. Name IDDate of birth Address CityState Frank 235253/191234 Harding ChicagoILMartha 112906/132 Speedway MiamiFLEllen 1214111/4935 Terrace Hills ChinoCA
D. Name IDDate of birth Address CityState Frank 235253/191234 Harding ChicagoILMartha 112906/132 Speedway MiamiFLEllen 1214111/411040935 Terrace Hills ChinoCA

Answer: D

Explanation:
This question falls under theData Concepts and Environmentsdomain, 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 34
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 theData Acquisition and Preparationdomain, 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 37
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 theData Concepts and Environmentsdomain, 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 39
A data analyst troubleshoots a dashboard every day for a week. Which of the following techniques best addresses how to validate the data moving forward?

A. Inquiring about structure changes
B. Setting up monitoring alerts
C. Reaching out to users daily
D. Rebuilding the dashboard

Answer: B

Explanation:
This question pertains to theData Governancedomain, focusing on ensuring data quality and reliability in dashboards over time. Daily troubleshooting indicates a recurring issue, and the task is to validate data moving forward.
? Inquiring about structure changes (Option A): This might identify past issues but
doesn??t provide ongoing validation.
? Setting up monitoring alerts (Option B): Monitoring alerts can automatically notify the analyst of data issues (e.g., missing updates, errors), providing a proactive way to validate data continuously.
? Reaching out to users daily (Option C): This is inefficient and reactive, not a sustainable validation method.
? Rebuilding the dashboard (Option D): Rebuilding might fix current issues but doesn??t ensure future validation.
The DA0-002 Data Governance domain includes "data quality control concepts," such as implementing monitoring to ensure data reliability in dashboards.
Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 5.0 Data Governance.

NEW QUESTION 42
A user needs a report that shows the main causes of customer churn rate in a three-year period. Which of the following methods provides this information?

A. Inferential
B. Descriptive
C. Prescriptive
D. Predictive

Answer: B

Explanation:
This question falls under theData Analysisdomain, focusing on analytical methods for reporting. The task is to identify the causes of customer churn over three

years, which involves analyzing historical data.
? Inferential (Option A): Inferential statistics make predictions or generalizations
about a population, not focused on identifying causes in historical data.
? Descriptive (Option B): Descriptive analytics summarizes historical data to identify patterns and causes (e.g., reasons for churn), which fits the task.
? Prescriptive (Option C): Prescriptive analytics provides recommendations, which goes beyond identifying causes.
? Predictive (Option D): Predictive analytics forecasts future outcomes, not focused on historical causes.
The DA0-002 Data Analysis domain includes "applying the appropriate descriptive statistical methods," and descriptive analytics is best for identifying causes in historical data.
Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 3.0 Data Analysis.
==============

NEW QUESTION 45
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 theData Acquisition and Preparationdomain, which in DA0-002
includes understanding data acquisition concepts and the types of data gathered from varioussources, 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 50
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 theData Governancedomain, 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 52
Which of the following data repositories stores unstructured and structured data?

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

Answer: D

Explanation:
This question falls under theData Concepts and Environmentsdomain of CompTIA Data+ DA0-002, which involves understanding different types of data repositories and their characteristics. The task is to identify a repository that can store both unstructured and structured data.
? Data store (Option A): A data store is a general term for any data repository, but
it??s not specific enough to confirm it stores both unstructured and structured data.

? Data silo (Option B): A data silo is an isolated data repository, often structured, and not typically designed for unstructured data.
? Data mart (Option C): A data mart is a subset of a data warehouse, focused on structured data for specific business areas, not unstructured data.
? Data lake (Option D): A data lake is a centralized repository that stores raw data in its native format, including both structured (e.g., tables) and unstructured (e.g., text, images) data, making it the correct choice.
The DA0-002 Data Concepts and Environments domain includes understanding "different types of databases and data repositories," and a data lake is specifically designed to handle both unstructured and structured data.
Reference: CompTIA Data+ DA0-002 Draft Exam Objectives, Domain 1.0 Data Concepts and Environments.
==============

NEW QUESTION 56
A data analyst pulls a table similar to the following one:
ID
Type TypeID Phone 1
Full Time Full Time 1 Mobile
2
Part Time Part Time 2 Work
3
Full Time Full Time 3 Mobile
Which of the following best explains the data issue with TypeID?

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 59
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 63
......