

DA0-001^{Q&As}

CompTIA Data+

Pass CompTIA DA0-001 Exam with 100% Guarantee

Free Download Real Questions & Answers **PDF** and **VCE** file from:

<https://www.lead4pass.com/da0-001.html>

100% Passing Guarantee
100% Money Back Assurance

Following Questions and Answers are all new published by CompTIA
Official Exam Center

-  **Instant Download** After Purchase
-  **100% Money Back** Guarantee
-  **365 Days** Free Update
-  **800,000+** Satisfied Customers



QUESTION 1

Which of the following would be considered non-personally identifiable information?

- A. Cell phone device name
- B. Customer's name
- C. Government ID number
- D. Telephone number

Correct Answer: A

Explanation: Non-personally identifiable information (non-PII) is any data that cannot be used to identify, contact, or locate a specific individual, either alone or combined with other sources. Non-PII can include aggregated statistics, anonymous data, device identifiers, IP addresses, cookies, and other types of information that do not reveal the identity or location of a person. Cell phone device name is an example of non-PII, as it does not reveal any personal information about the owner or user of the device. Therefore, the correct answer is A. References: What is Non-Personally Identifiable Information (Non-PII)? | Definition and Examples, What is Personally Identifiable Information (PII)? | Definition and Examples

QUESTION 2

A data analyst has been asked to create a daily manufacturing report for the floor manager Which of the following metrics should be included in the report?

- A. Tons of steel produced per hour
- B. Annual sales budget
- C. End-of-day stock price
- D. Daily corporate employee count

Correct Answer: A

QUESTION 3

Which of the following is the BEST reason to use database views instead of tables?

- A. Views reduce the need for repetitive, complex data joins.
- B. Views allow for the storage of temporary data. whereas tables do not.
- C. Views allow for the joining of multiple data sources, whereas tables do not.
- D. Views can be used to restrict sensitive information.

Correct Answer: A

Explanation: Views are virtual tables that are created by querying one or more base tables or other views. Views do not store any data, but only show the result of a query. One of the main advantages of using views is that they can reduce the need for repetitive, complex data joins. For example, if a query involves joining multiple tables with many conditions, creating a view can simplify the query and make it easier to reuse. Therefore, the correct answer is A. References: [What is a Database View? | Definition and Examples - Vertabelo], [Database Views - GeeksforGeeks]

QUESTION 4

You have two databases tables that you would like to join together using a foreign key relationship.

What term best describes this action?

- A. Blending.
- B. Appending.
- C. Mixing.
- D. Merging.

Correct Answer: D

Data merging is the process of combining two or more data sets into a single data set. Most often, this process is necessary when you have raw data stored in multiple files, worksheets, or data tables, that you want to analyze all in one go.

QUESTION 5

Which of the following differentiates a flat text file from other data types?

- A. Data is separated by a delimiter.
- B. Data is stored in defined rows.
- C. Data is defined with key-value pairs.
- D. Data is housed in a markup language.

Correct Answer: A

Explanation: A flat text file is a type of data file that contains only plain text without any formatting or markup. Data in a flat text file is usually separated by a delimiter, which is a character that marks the boundary between different fields or values. For example, a comma-separated values (CSV) file is a flat text file that uses commas as delimiters. Other common delimiters are tabs, spaces, semicolons, and pipes. Therefore, the correct answer is A. References: Plain text -Wikipedia, Comparison of document markup languages - Wikipedia

QUESTION 6

Amanda needs to create a dashboard that will draw information from many other data sources and present it to business leaders.

Which one of the following tools is least likely to meet her needs?

- A. QuickSight.
- B. Tableau.
- C. Power BI.
- D. SPSS Modeler.

Correct Answer: D

SPSS Modeler.

QuickSight, Tableau, and Power BI are all powerful analytics and reporting tools that can pull data from a variety of sources. SPSS Modeler is a powerful predictive analytics platform that is designed to bring predictive intelligence to decisions

made by individuals, groups, systems and your enterprise.

QUESTION 7

A data analyst needs to create a dashboard using the company's yearly revenue data sets. Which of the following would be the best way to plot the information to show the top-performing region?

- A. A line chart
- B. A waterfall chart
- C. A heat map
- D. A stacked bar chart

Correct Answer: D

QUESTION 8

Which of the following is an example of a flat file?

- A. CSV file
- B. PDF file
- C. JSON file
- D. JPEG file

Correct Answer: D

QUESTION 9

A Chief Executive Officer (CEO) is requesting more up-to-date sales data for improved visibility prior to month-end. An analyst must determine the frequency of a sales report that was previously distributed on an as-needed basis. Which of the following would be the most appropriate frequency for this report?

- A. Monthly
- B. Quarterly
- C. Weekly
- D. Every other month

Correct Answer: C

Explanation: The most appropriate frequency for the sales report is weekly, as this will provide the CEO with more up-to-date sales data for improved visibility prior to month-end. A weekly sales report can show the sales performance, trends, and issues of the sales team on a regular basis, and help the CEO to monitor and evaluate the progress and results of the sales activities. A weekly sales report can also help the CEO to identify and address any problems or opportunities that may arise during the month, and to make timely and informed decisions.

QUESTION 10

Which of the following is a process that is used during data integration to collect, blend, and load data?

- A. MDM
- B. ETL
- C. OLTP
- D. BI

Correct Answer: B

Explanation: ETL is a process that is used during data integration to collect, blend, and load data. ETL stands for extract, transform, and load, which are the three main steps involved in moving data from different sources to a common destination, such as a data warehouse or a data lake. ETL helps to consolidate and standardize data for analysis and reporting purposes. References: CompTIA Data+ Certification Exam Objectives, page 12

QUESTION 11

After completing web scraping, which of the following file formats needs to be parsed?

- A. .html
- B. .txt
- C. .csv
- D. .tsv

Correct Answer: A

The correct answer is .html. Short explanation: Web scraping is the process of extracting data from websites by parsing the HTML code of the web pages. HTML stands for HyperText Markup Language and it is the standard markup language for creating web pages and web applications. HTML files have the extension .html and they contain tags, elements, attributes, and content that define the structure and appearance of a web page. Web scraping tools need to parse the HTML files to extract the relevant data from the web pages12

QUESTION 12

Consider this dataset showing the retirement age of 11 people, in whole years:

54, 54, 54, 55, 56, 57, 57, 58, 58, 60, 60

This tables show a simple frequency distribution of the retirement age data.

Age	Frequency
54	3
55	1
56	1
57	2
58	2
60	2

- A. 56
- B. 55
- C. 57
- D. 54

Correct Answer: D

A measure of central tendency (also referred to as measures of centre or central location) is a summary measure that attempts to describe a whole set of data with a single value that represents the middle or centre of its distribution. There

are three main measures of central tendency: the mode, the median and the mean. Each of these measures describes a different indication of the typical or central value in the distribution.

What is the mode?

The mode is the most commonly occurring value in a distribution. The most commonly occurring value is 54, therefore the mode of this distribution is 54 years.

QUESTION 13

Given the following:

Candy	Has_nuts	Date_purchased	Cost	Quantity	Ext_cost
Snickers	Y	2021-08-24	\$1.00	2	2.00
Starburst	N	8/24/2021	null	10	null
Snickers	Y	2020-11-13	\$2.00	3	6.00

Which of the following is the most important thing for an analyst to do when transforming the table for a trend analysis?

- A. Fill in the missing cost where it is null.
- B. Separate the table into two tables and create a primary key
- C. Replace the extended cost field with a calculated field.
- D. Correct the dates so they have the same format.

Correct Answer: D

Correcting the dates so they have the same format is the most important thing for an analyst to do when transforming the table for a trend analysis. Trend analysis is a method of analyzing data over time to identify patterns, changes, or relationships. To perform a trend analysis, the data needs to have a consistent and comparable format, especially for the date or time variables.

In the example, the date purchased column has two different formats: YYYY-MM-DD and MM/DD/YYYY. This could cause errors or confusion when sorting, filtering, or plotting the data over time. Therefore, the analyst should correct the dates so they have the same format, such as YYYY-MM-DD, which is a standard and unambiguous format.

QUESTION 14

Which of the following are reasons to create and maintain a data dictionary? (Choose two.)

- A. To improve data acquisition
- B. To remember specifics about data fields
- C. To specify user groups for databases
- D. To provide continuity through personnel turnover
- E. To confine breaches of PHI data

F. To reduce processing power requirements

Correct Answer: AB

Explanation: The reasons to create and maintain a data dictionary are to improve data acquisition and to remember specifics about data fields. A data dictionary is a document or a database that describes the structure, meaning, and usage of the data elements in a data source or a database. A data dictionary can help to improve data acquisition by providing clear and consistent definitions, rules, and standards for the data collection process. A data dictionary can also help to remember specifics about data fields by providing information such as data type, format, length, range, default value, constraints, relationships, etc. The other options are not reasons to create and maintain a data dictionary, as they are related to other aspects of data management or security. A data dictionary does not specify user groups for databases, as this is a function of access control or authorization. A data dictionary does not provide continuity through personnel turnover, as this is a function of documentation or knowledge transfer. A data dictionary does not confine breaches of PHI data, as this is a function of encryption or anonymization. A data dictionary does not reduce processing power requirements, as this is a function of optimization or compression. Reference: [What is a Data Dictionary? - DataCamp]

QUESTION 15

Given the following data: Which of the following BEST describes the data set?

Name	Gender	Age	Annual income
Ralph	M	27	\$75,000
Jessie	F	3	\$75,000
Monica	F	31	\$125,000
Carlos	M	53	\$75
Sara	F	43	\$0

- A. There is data bias.
- B. The data is incomplete.
- C. The data is inconsistent.
- D. The data is outliers.

Correct Answer: C

Explanation: This is because inconsistency is a type of data quality issue that occurs when the data does not follow a common format, structure, or rule across different sources or systems, which can affect the efficiency and performance of the analysis or process. Inconsistency can be caused by having different spellings, punctuations, capitalizations, or abbreviations for the same or similar values in a data set, such as "M", "m", "Male", or "male" for gender in this case. Inconsistency can be eliminated or reduced by using data cleansing techniques, such as standardizing or normalizing the data values. The other options are not correct descriptions of the data set. Here is why:

Data bias is a type of data quality issue that occurs when the data is not representative or proportional of the population or the parameter, which can affect the validity and reliability of the analysis or process. Data bias can be caused by

having a sample that is too small, too large, or too skewed for the population or the parameter, such as having only male customers for a product that targets both genders in this case. Data bias can be eliminated or reduced by using sampling techniques, such as stratified or cluster sampling. The data is incomplete is a type of data quality issue that occurs when the data is absent or missing in a data set, which can affect the accuracy and reliability of the analysis or process. The data is incomplete can be caused by various factors, such as human error, system error, or non-response. The data is incomplete can be addressed by using various methods, such as replacing or imputing the missing values with some reasonable estimates, such as mean, median, mode, or regression. The data is outliers is a type of data quality issue that occurs when the data has values that are unusually high or low compared to the rest of the data set, which can affect the quality and validity of the analysis or process. The data is outliers can be caused by various factors, such as measurement error, natural variation, or extreme events. The data is outliers can be addressed by using various methods, such as removing or filtering out the outliers, or using robust statistics that are less sensitive to outliers, such as median, interquartile range, or box plot.

[Latest DA0-001 Dumps](#)

[DA0-001 PDF Dumps](#)

[DA0-001 VCE Dumps](#)