



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.passapply.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 data cleansing issues will be fixed when a DISTINCT function is applied?

- A. Missing data
- B. Duplicate data
- C. Redundant data
- D. Invalid data

Correct Answer: B

Explanation: This is because duplicate data refers to data that is repeated or copied in a data set, which can affect the quality and validity of the analysis. A DISTINCT function is a type of function that removes duplicate values from a column or a table, leaving only unique values. For example, a DISTINCT function in SQL that can achieve this is:

```
SELECT DISTINCT column_name FROM table_name;
```

The other data cleansing issues will not be fixed by applying a DISTINCT function. Here is why:

Missing data refers to data that is absent or incomplete in a data set, which can affect the accuracy and reliability of the analysis. A DISTINCT function does not help with missing data, because it does not fill in or impute the missing values. Redundant data refers to data that is unnecessary or irrelevant for the analysis, which can affect the efficiency and performance of the analysis. A DISTINCT function does not help with redundant data, because it does not remove or filter out the redundant values. Invalid data refers to data that is incorrect or inaccurate in a data set, which can affect the validity and reliability of the analysis. A DISTINCT function does not help with invalid data, because it does not validate or correct the invalid values.

QUESTION 2

A user receives a large custom report to track company sales across various date ranges. The user then completes a series of manual calculations for each date range. Which of the following should an analyst suggest so the user has a dynamic, seamless experience?

- A. Create multiple reports, one for each needed date range.
- B. Build calculations into the report so they are done automatically.
- C. Add macros to the report to speed up the filtering and calculations process.
- D. Create a dashboard with a date range picker and calculations built in.

Correct Answer: D

Explanation: Create a dashboard with a date range picker and calculations built in. This is because a dashboard is a type of visualization that displays multiple charts or graphs on a single page, usually to provide an overview or summary of some data or information. A dashboard can be used to track company sales across various date ranges by showing different metrics and indicators related to sales, such as revenue, volume, or growth. By creating a dashboard with a date range picker and calculations built in, the analyst can suggest a way for the user to have a dynamic, seamless



experience, which means that the user can interact with and customize the dashboard according to their needs or preferences, as well as avoid any manual work or errors. For example, a date range picker is a type of feature or function that allows users to select or adjust the time period for which they want to see the data on the dashboard, such as daily, weekly, monthly, or quarterly. A date range picker can make the dashboard dynamic, as it can automatically update or refresh the dashboard with new data based on the selected time period. Calculations are mathematical operations or expressions that can be performed on the data on the dashboard, such as addition, subtraction, multiplication, division, average, sum, etc. Calculations can make the dashboard seamless, as they can eliminate the need for manual calculations for each date range, as well as ensure accuracy and consistency of the results. The other ways are not the best ways to provide a dynamic, seamless experience for the user. Here is why:

Creating multiple reports, one for each needed date range would not provide a dynamic, seamless experience for the user, but rather create a static, cumbersome experience, which means that the user cannot interact with or customize the reports according to their needs or preferences, as well as have to deal with multiple files or pages. For example, creating multiple reports would make it difficult for the user to compare or contrast the sales across different date ranges, as well as increase the workload and complexity of managing and maintaining the reports. Building calculations into the report so they are done automatically would not provide a dynamic, seamless experience for the user, but rather provide a partial, limited experience, which means that the user can only benefit from one aspect or feature of the report, but not from others. For example, building calculations into the report would help with avoiding manual work or errors, but it would not help with interacting with or customizing the report according to different date ranges. Adding macros to the report to speed up the filtering and calculations process would not provide a dynamic, seamless experience for the user, but rather provide an advanced, complex experience, which means that the user would need to have some technical skills or knowledge to use or apply the macros, as well as face some potential risks or challenges. For example, adding macros to the report would require the user to know how to write or run the macros, which are a type of code or script that automates certain tasks or actions on the report, such as filtering or calculating the data. Adding macros to the report could also expose the user to some security or compatibility issues, such as viruses, malware, or errors.

QUESTION 3

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 4

Which of the following database schemas features normalized dimension tables?

- A. Flat



- B. Snowflake
- C. Hierarchical
- D. Star

Correct Answer: B

Explanation: The correct answer is B. Snowflake.

A snowflake schema is a type of database schema that features normalized dimension tables. A database schema is a way of organizing and structuring the data in a database. A dimension table is a table that contains descriptive attributes or characteristics of the data, such as product name, category, color, etc. A normalized table is a table that follows the rules of normalization, which is a process of reducing data redundancy and improving data integrity by organizing the data into smaller and simpler tables¹² A snowflake schema is a variation of the star schema, which is another type of database schema that features denormalized dimension tables. A denormalized table is a table that does not follow the rules of normalization, and may contain redundant or duplicated data. A star schema consists of a central fact table that contains quantitative measures or facts, such as sales amount, order quantity, etc., and several dimension tables that are directly connected to the fact table. A snowflake schema differs from a star schema in that the dimension tables are further split into sub-dimension tables, creating a snowflake-like shape¹³ A snowflake schema has some advantages and disadvantages over a star schema. Some advantages are: It reduces the storage space required for the dimension tables, as it eliminates the redundant data. It improves the data quality and consistency, as it avoids the update anomalies that may occur in denormalized tables. It allows more detailed analysis and queries, as it provides more levels of dimensions. Some disadvantages are: It increases the complexity and number of joins required to retrieve the data from multiple tables, which may affect the query performance and speed. It reduces the readability and simplicity of the schema, as it has more tables and relationships to understand. It may require more maintenance and administration, as it has more tables to manage and update¹³

QUESTION 5

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 VCE Dumps](#)

[DA0-001 Practice Test](#)