

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.pass2lead.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

A database consists of one fact table that is composed of multiple dimensions. Each dimension is represented by a denormalized table. This structure is an example of a:

- A. non-relational schema.
- B. galaxy schema.
- C. snowflake schema.
- D. star schema.

Correct Answer: D

Explanation: A star schema is a type of database schema that consists of one fact table and multiple dimension tables. The fact table contains the measures or metrics of the business process, such as sales, orders, or transactions. The dimension tables contain the attributes or characteristics of the business entities, such as products, customers, or locations. The fact table is connected to the dimension tables by foreign keys that reference the primary keys of the dimension tables. The fact table is located at the center of the schema, while the dimension tables are located at the edges, forming a star-like shape¹. A star schema is an example of a denormalized schema, which means that the dimension tables are not normalized and may contain redundant or repeated data. This is done to improve the performance and simplicity of queries, as there are fewer joins and tables involved. A star schema is suitable for data warehouses and business intelligence applications that require fast and efficient data retrieval².

QUESTION 2

An analyst modified a data set that had a number of issues. Given the original and modified versions:

Original data:

| Var001 | Var002 | Var003 | Var004 |
|--------|--------|--------|--------|
| 1 | 0 | 0 | 0 |
| 0 | 1 | 0 | 1 |
| 1 | 1 | 1 | 2 |
| 0 | 0 | 0 | 1 |

Modified data:

| Var001 | Var002 | Var003 | Var004 |
|--------|---------|------------|--------|
| Yes | Absent | No payment | No |
| No | Present | No payment | Yes |
| Yes | Present | Payment | Maybe |
| No | Absent | No payment | Yes |

Which of the following data manipulation techniques did the analyst use?

- A. Imputation
- B. Recoding
- C. Parsing
- D. Deriving

Correct Answer: B

The correct answer is B. Recoding.

Recoding is a data manipulation technique that involves changing the values or categories of a variable to make it more suitable for analysis. Recoding can be used to simplify or group the data, to correct errors or inconsistencies, or to create

new variables from existing ones¹²

In the example, the analyst used recoding to change the values of Var001, Var002, Var003, and Var004 from numerical to textual form. The analyst also used recoding to assign meaningful labels to the values, such as "Absent" for 0,

"Present" for 1, "Low" for 2, "Medium" for 3, and "High" for 4. This makes the data more understandable and easier to analyze.

QUESTION 3

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 4

A data analyst for a media company needs to determine the most popular movie genre. Given the table below:

| MovieID | Name | Genre | Actors | Rating |
|---------|-------------------|----------------------------|---|--------|
| 01 | Ghost Writer | Comedy, Actions | Joshua Wellington, Susana Summons | 6.5 |
| 02 | Life of Suffering | Drama. Foreign, Historical | Shelly May, Rita Moralle, Ethan Warner, Sean Houser | 7.2 |

Which of the following must be done to the Genre column before this task can be completed?

- A. Append
- B. Merge
- C. Concatenate
- D. Delimit

Correct Answer: D

Explanation: The action that must be done to the Genre column before this task can be completed is delimit. Delimit is a process of separating or splitting a string of text into multiple parts based on a delimiter, which is a character or a sequence of characters that marks the boundary between the parts. For example, a comma (,) or a semicolon (;) can be used as a delimiter. In this case, the Genre column contains multiple genres for each movie, separated by commas. To

determine the most popular movie genre, the data analyst needs to delimit the Genre column by commas, so that each genre can be counted and compared separately. The other options are not relevant for this task, as they are related to combining or joining strings or tables, not separating them. Append is a process of adding or attaching one string or table to the end of another string or table. Merge is a process of combining or joining two or more tables into one table based on a common column or key. Concatenate is a process of joining or linking two or more strings together into one string. Reference: [How to Split Text in Excel - Exceljet]

QUESTION 5

Which of the following BEST describes standard deviation?

- A. A measure that is used to establish a relationship between two variables
- B. A measure of how data is distributed
- C. A measure of the amount of dispersion of a set of values
- D. A measure that is used to find the significant difference between variables

Correct Answer: C

Explanation: A measure of the amount of dispersion of a set of values. This is because standard deviation is a type of statistical measure that quantifies how much the values in a data set vary or deviate from the mean or the average of the

data set. Standard deviation can be used to describe the spread or the distribution of the data, as well as to identify any outliers or extreme values in the data. For example, a low standard deviation indicates that the values are close to the

mean, while a high standard deviation indicates that the values are far from the mean. The other options are not correct descriptions of standard deviation.

Here is why:

A measure that is used to establish a relationship between two variables is not a correct description of standard deviation, but rather a description of correlation or regression, which are types of statistical measures that quantify how two

variables are related or associated with each other. Correlation or regression can be used to test or model the dependence or the influence of one variable on another variable, as well as to predict or estimate the value of one variable based

on the value of another variable.

A measure of how data is distributed is not a correct description of standard deviation, but rather a description of frequency or probability, which are types of statistical measures that quantify how often or how likely a value or an event occurs

in a data set. Frequency or probability can be used to describe the occurrence or the chance of the data, as well as to compare or contrast different categories or groups of the data.

A measure that is used to find the significant difference between variables is not a correct description of standard deviation, but rather a description of hypothesis testing or inferential statistics, which are types of statistical methods that use

sample data to make generalizations or conclusions about a population or a parameter. Hypothesis testing or inferential

statistics can be used to test or verify a claim or an assumption about the data, as well as to measure the confidence or the error of the estimation.

[Latest DA0-001 Dumps](#)

[DA0-001 Practice Test](#)

[DA0-001 Braindumps](#)