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

QUESTION 1

A data analyst is asked on the morning of April 9, 2020, to create a sales report that identifies sales year to date. The daily sales data is current through the end of the day. Which of the following date ranges should be on the report?

A. January 1, 2020 to April 1, 2020

B. January 1, 2020 to April 7, 2020

C. January 1, 2020 to April 8, 2020

D. January 1, 2020 to April 9, 2020

Correct Answer: D

This is because sales year to date refers to the sales that have occurred from the beginning of the current year until the current date. By creating a sales report that identifies sales year to date, the analyst can measure and compare the sales performance and progress of the current year. Since the analyst is asked to create the sales report on the morning of April 9, 2020, and the daily sales data is current through the end of the day, the date range that should be on the report is January 1, 2020 to April 9, 2020. The other date ranges are not correct for identifying sales year to date. Here is why:

January 1, 2020 to April 1, 2020 would not include the sales that occurred in the first eight days of April, which would underestimate the sales year to date. January 1, 2020 to April 7, 2020 would not include the sales that occurred in the last two days of April, which would also underestimate the sales year to date. January 1, 2020 to April 8, 2020 would not include the sales that occurred on April 9, which would also underestimate the sales year to date.

QUESTION 2

Which of the following will MOST likely be streamed live?

A. Machine data

B. Key-value pairs

C. Delimited rows

D. Flat files

Correct Answer: A

Explanation: Machine data is the most likely type of data to be streamed live, as it refers to data generated by machines or devices, such as sensors, web servers, network devices, etc. Machine data is often produced continuously and in large volumes, requiring real-time processing and analysis. Other types of data, such as key-value pairs, delimited rows, and flat files, are more likely to be stored in databases or files and processed in batches.

QUESTION 3

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 4

Under which of the following circumstances should the null hypothesis be accepted when a = 0.05?

A. When p is 0.00003

B. When p is 0.001

C. When p is 0.04

D. When p is 0.06

Correct Answer: D

The null hypothesis should be accepted when the p-value is greater than the alpha level, which is the significance level of the test. The p-value is the probability of obtaining a test statistic at least as extreme as the one observed in the sample, assuming that the null hypothesis is true. The alpha level is the probability of rejecting the null hypothesis when it is true, which is also known as a type I error12. In this case, the alpha level is 0.05, which means that there is a 5% chance of rejecting the null hypothesis when it is true. Therefore, to reject the null hypothesis, the p-value must be less than or equal to 0.05, which indicates that the test statistic is very unlikely to occur by chance under the null hypothesis. Conversely, to accept the null hypothesis, the p-value must be greater than 0.05, which indicates that the test statistic is not very unlikely to occur by chance under the null hypothesis. Among the four options, only option D has a p-value that is greater than 0.05 (p = 0.06). Therefore, option D is the correct answer. When p = 0.06, it means that there is a 6% chance of obtaining a test statistic at least as extreme as the one observed in the sample, assuming that the null hypothesis is true. This probability is not very low, and therefore does not provide enough evidence to reject the null hypothesis.

QUESTION 5

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.