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.pass4itsure.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

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 2

Jenny wants to study the academic performance of undergraduate sophomores and wants to determine the average grade point average at different points during an academic year.

What best describes the data set she needs?

A. Sample.

B. Observation.

C. Variable.

D. Population.

Correct Answer: A

Correct answer A. Sample.

Jenny does not have data for the entire population of all undergraduate sophomores. While a specific grade point average is an observation of variable, jenny needs sample data.

QUESTION 3

A recurring event is being stored in two databases that are housed in different geographical locations. A data analyst notices the event is being logged three hours earlier in one database than in the other database. Which of the following is the MOST likely cause of the issue?

A. The data analyst is not querying the databases correctly.

B. The databases are recording different events.

C. The databases are recording the event in different time zones.

D. The second database is logging incorrectly.

Correct Answer: C

Explanation: The most likely cause of the issue is that the databases are recording the event in different time zones. A time zone is a region that observes a uniform standard time for legal, commercial, and social purposes. Different time zones have different offsets from Coordinated Universal Time (UTC), which is the primary time standard by which the world regulates clocks and time. For example, UTC-5 is five hours behind UTC, while UTC+3 is three hours ahead of UTC. If an event is being stored in two databases that are housed in different geographical locations with different time zones, it may appear that the event is being logged at different times, depending on how the databases handle the time zone conversion. For example, if one database records the event in UTC-5 and another database records the event in UTC+3, then an event that occurs at 12:00 PM in UTC-5 will appear as 9:00 AM in UTC+3. The other options are not likely causes of the issue, as they are either unrelated or implausible. The data analyst is not querying the databases incorrectly, as this would not affect the time stamps of the events. The databases are not recording different events, as they are supposed to record the same recurring event. The second database is not logging incorrectly, as there is no evidence or reason to assume that. Reference: [Time zone - Wikipedia]

QUESTION 4

An analyst is required to run a text analysis of data that is found in articles from a digital news outlet. Which of the following would be the BEST technique for the analyst to apply to acquire the data?

A. Web scraping

B. Sampling

C. Data wrangling

D. ETL

Correct Answer: A

Explanation: This is because web scraping is a technique that allows the analyst to extract data from web pages, such as articles from a digital news outlet. Web scraping can be done using various tools and methods, such as Python

libraries, browser extensions, or online services. The other techniques are not suitable for acquiring data from web pages.

Here is why:

Sampling is a technique that involves selecting a subset of data from a larger population, usually for statistical analysis or testing purposes. Sampling does not help the analyst to acquire data from web pages, but rather to reduce the amount

of data to be analyzed. Data wrangling is a technique that involves transforming and cleaning data to make it suitable for analysis or visualization. Data wrangling does not help the analyst to acquire data from web pages, but rather to improve

the quality and usability of the data. ETL stands for Extract, Transform, and Load, which is a process that involves moving data from one or more sources to a destination, such as a data warehouse or a database. ETL does not help the

analyst to acquire data from web pages, but rather to store and organize the data.

QUESTION 5

Which of the following data sampling methods involves dividing a population into subgroups by similar characteristics?

A. Systematic

B. Simple random

C. Convenience

D. Stratified

Correct Answer: D

Explanation: Stratified sampling is a data sampling method that involves dividing a population into subgroups by similar characteristics, such as age, gender, income, etc. Then, a simple random sample is drawn from each subgroup. This method ensures that each subgroup is adequately represented in the sample and reduces the sampling error. References: CompTIA Data+ Certification Exam Objectives, page 11.