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.geekcert.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 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 2

Which of the following best describes the law of large numbers?

A. As a sample size decreases, its standard deviation gets closer to the average of the whole population.

B. As a sample size grows, its mean gets closer to the average of the whole population

C. As a sample size decreases, its mean gets closer to the average of the whole population.

D. When a sample size doubles. the sample is indicative of the whole population.

Correct Answer: B

The best answer is B. As a sample size grows, its mean gets closer to the average of the whole population. The law of large numbers, in probability and statistics, states that as a sample size grows, its mean gets closer to the average of the whole population. This is due to the sample being more representative of the population as it increases in size. The law of large numbers guarantees stable long-term results for the averages of some random events1 A. As a sample size decreases, its standard deviation gets closer to the average of the whole population is not correct, because it confuses the concepts of standard deviation and mean. Standard deviation is a measure of how much the values in a data set vary from the mean, not how close the mean is to the population average. Also, as a sample size decreases, its standard deviation tends to increase, not decrease, because the sample becomes less representative of the population.

C. As a sample size decreases, its mean gets closer to the average of the whole population is not correct, because it

contradicts the law of large numbers. As a sample size decreases, its mean tends to deviate from the average of the whole population, because the sample becomes less representative of the population. D. When a sample size doubles, the sample is indicative of the whole population is not correct, because it does not specify how close the sample mean is to the population average. Doubling the sample size does not necessarily make the sample indicative of the whole population, unless the sample size is large enough to begin with. The law of large numbers does not state a specific number or proportion of samples that are indicative of the whole population, but rather describes how the sample mean approaches the population average as the sample size increases indefinitely.

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

Which of the following is used for calculations and pivot tables?

A. IBM SPSS

B. SAS

C. Microsoft Excel

D. Domo

Correct Answer: C

Explanation: This is because Microsoft Excel is a type of software application that allows users to create, edit, and analyze data in spreadsheets, which are composed of rows and columns of cells that can store various types of data, such as

numbers, text, or formulas. Microsoft Excel can be used for calculations and pivot tables, which are two common features or functions in data analysis. Calculations are mathematical operations or expressions that can be performed on the

data in the cells, such as addition, subtraction, multiplication, division, average, sum, etc. Pivot tables are interactive tables that can summarize and display the data in different ways, such as by grouping, filtering, sorting, or aggregating the

data based on various criteria or categories. The other software applications are not used for calculations and pivot tables. Here is why:

IBM SPSS is a type of software application that allows users to perform statistical analysis and modeling on data sets, such as regression, correlation, ANOVA, etc. IBM SPSS does not use spreadsheets or cells to store or manipulate data,

but rather uses data views or variable views to display the data in rows and columns. IBM SPSS does not have pivot tables as a feature or function, but rather has output views or charts to display the results of the analysis.

SAS is a type of software application that allows users to perform data management and analysis using a programming language that consists of statements and commands. SAS does not use spreadsheets or cells to store or manipulate

data, but rather uses data sets or tables that are stored in libraries or folders. SAS does not have pivot tables as a feature or function, but rather has procedures or macros that can produce summary tables or reports based on the data.

Domo is a type of software application that allows users to create and share dashboards and visualizations that display data from various sources and systems, such as databases, cloud services, or web applications. Domo does not use

spreadsheets or cells to store or manipulate data, but rather uses connectors or APIs to access and integrate the data from different sources. Domo does not have pivot tables as a feature or function, but rather has cards or widgets that can

show different aspects or metrics of the data.

QUESTION 5

A development company is constructing a new unit in its apartment complex. The complex has the following floor plans:

Unit name	Sq. Ft.	Price	$/Sq. Ft.
Jasmine	1,000	$345,000	$345
Orchid	1,100	$425,000	$386
Azalea	1,300	$460,000	$354
Tulip	1,640	$525,000	$320
Rose	2,000		

Using the average cost per square foot of the original floor plans, which of the following should be the price of the Rose unit?

A. $640,900

B. $690,000

C. $705,200

D. $702,500

Correct Answer: C

Explanation: This is because the price of the Rose unit can be estimated using the average cost per square foot of the original floor plans, which are Jasmine, Orchid, Azalea, and Tulip. To find the average cost per square foot of the original floor plans, we can use the following formula:

```
Average cost per square foot = Total price / Total square feet
```

Plugging in the values from the original floor plans, we get:

```
Average cost per square foot = ($345,000 + $425,000 + $465,000 + $525,000) / (1,000 + 1,250 + 1,500 + 2,000)
Average cost per square foot = $1,760,000 / 5,750
Average cost per square foot = $306
```

To find the price of the Rose unit, we can use the following formula:

```
Price = Square feet * Average cost per square foot
```

Plugging in the values from the Rose unit, we get:

```
Price = 2,300 * $306

Price = $705,200
```

Therefore, the price of the Rose unit should be $705,200, using the average cost per square foot of the original floor plans.