Vendor: CompTIA

Exam Code:DA0-001

Exam Name: CompTIA Data+

Version: Demo

QUESTION 1

A site reliability team wants to monitor the stability of their website. so they can proactively diagnose issues when they occur Which of the following deliverables would best suit their needs?

A. A self-serve dashboard of website performance that updates in real time

B. A weekly log report of site visits and user actions

C. A portal that is refreshed daily and reports errors classified by type

D. A daily summary email indicating website outages for the previous day

Correct Answer: A

Explanation: The best deliverable that would suit the site reliability team\\'s needs is A. A self-serve dashboard of website performance that updates in real time. A self-serve dashboard is a visual display of the most important information needed to achieve one or more objectives, consolidated and arranged on a single screen so the information can be monitored at a glance. A self-serve dashboard of website performance that updates in real time would allow the site reliability team to easily and quickly access the information they need about the stability of their website, such as uptime, response time, error rate, traffic volume, etc. A self-serve dashboard would also enable the team to proactively diagnose issues when they occur, by providing alerts, notifications, or drill-down options. A self-serve dashboard would also be more interactive and engaging than a report or an email. A weekly log report of site visits and user actions would not be a good deliverable for the site reliability team\\'s needs, because it would not provide timely or relevant information about the stability of their website. A weekly log report would be too infrequent and delayed to monitor and diagnose issues when they occur. A weekly log report would also focus on the behavior and actions of the users, rather than the performance and functionality of the website. A portal that is refreshed daily and reports errors classified by type would not be a good deliverable for the site reliability team\\'s needs, because it would not provide real-time or comprehensive information about the stability of their website. A portal that is refreshed daily would be too slow and outdated to monitor and diagnose issues when they occur. A portal that reports errors classified by type would be too narrow and limited to capture the full picture of the website performance. A daily summary email indicating website outages for the previous day would not be a good deliverable for the site reliability team\\'s needs, because it would not provide real-time or actionable information about the stability of their website. A daily summary email would be too late and retrospective to monitor and diagnose issues when they occur. A daily summary email indicating website outages would also be too passive and generic to help the team resolve or prevent issues in the future.

QUESTION 2

Which of the following is a difference between a primary key and a unique key?

A. A unique key cannot take null values, whereas a primary key can take null values.

B. There can be only one primary key in a data set, whereas there can be multiple unique keys.

C. A primary key can take a value more than once, whereas a unique key cannot take a value more than once.

D. A primary key cannot be a date variable, whereas a unique key can be.

Correct Answer: B

The correct answer is B. There can be only one primary key in a data set, whereas there can be multiple unique keys. A primary key is a column or a set of columns that uniquely identifies each row in a table. A table can have only one primary key, which also enforces the NOT NULL constraint on the column(s) involved. A primary key can also be

referenced by a foreign key of another table to establish a relationship between the tables12 A unique key is a column or a set of columns that also uniquely identifies each row in a table, but it is not the primary key. A table can have more than one unique key, which also allows one NULL value for the column(s) involved. A unique key can also be referenced by a foreign key of another table to establish a relationship between the tables12 Some of the differences between a primary key and a unique key are: A primary key creates a clustered index on the column(s), whereas a unique key creates a non-clustered index on the column(s)3 A primary key does not allow any NULL values, whereas a unique key allows one NULL value for the column(s)123 A primary key can be a unique key, but a unique key cannot be a primary key12

QUESTION 3

Which one of the following is a common data warehouse schema?

A. Snowflake.

B. Square.

C. Spiral.

D. Sphere.

Correct Answer: A

Snowflake enables data storage, processing, and analytic solutions that are faster, easier to use, and far more flexible than traditional offerings. The Snowflake data platform is not built on any existing database technology or "big data" software platforms such as Hadoop.

QUESTION 4

A sales director has requested a report for individual team members within the division be developed. The director would like the report to be shared with all team members, but individual team members should not be identifiable within the report Which of the following access requirements would support the director\\'s needs?

A. Create an acceptable use policy for the sales data.

B. Release the report as user-group-based access and include data masking.

C. Get a data use agreement from the individual team members.

D. Provide the report based on role and include data encryption.

Correct Answer: B

QUESTION 5

Which of the following descriptive statistical methods are measures of central tendency? (Choose two.)

A. Mean

B. Minimum

C. Mode

D. Variance

E. Correlation

F. Maximum

Correct Answer: AC

Explanation: Mean and mode are measures of central tendency, which describe the typical or most common value in a distribution of data. Mean is the arithmetic average of all the values in a dataset, calculated by adding up all the values and dividing by the number of values. Mode is the most frequently occurring value in a dataset. Other measures of central tendency include median, which is the middle value when the data is sorted in ascending or descending order.

QUESTION 6

A data analyst has been asked to merge the tables below, first performing an INNER JOIN and then a LEFT JOIN:

Customer_ID	Segment	Region
001	New	BC
002	Existing	ON
003	New	MB
004	New	ON
005	Existing	AT
006	Existing	MB
007	New	QC
008	New	QC
009	Existing	BC

Customer Table In-store Transactions ?

Order_ID	Customer_ID	Date	Amount	Quantity
006A	006	04/01/2020	$200	59
007B	007	03/01/2020	$500	54
008C	008	02/01/2020	$600	15
009D	009	05/01/2020	$800	18
001E	001	07/01/2020	$300	50
003F	003	08/01/2020	$200	55

Which of the following describes the number of rows of data that can be expected after performing both joins in the order stated, considering the customer table as the main table?

A. INNER: 6 rows; LEFT: 9 rows

B. INNER: 9 rows; LEFT: 6 rows

C. INNER: 9 rows; LEFT: 15 rows

D. INNER: 15 rows; LEFT: 9 rows

Correct Answer: C

An INNER JOIN returns only the rows that match the join condition in both tables. A LEFT JOIN returns all the rows from the left table, and the matched rows from the right table, or NULL if there is no match. In this case, the customer table is

the left table and the in-store transactions table is the right table. The join condition is based on the customer_id column, which is common in both tables.

To perform an INNER JOIN, we can use the following SQL query:

SELECT * FROM customer INNER JOIN in_store_transactions ON customer.customer_id = in_store_transactions.customer_id;

This query will return 9 rows of data, as shown below:

customer_id | name | lastname | gender | marital_status | transaction_id | amount | date 1 | MARC | TESCO | M | Y | 1 | 1000 | 2020-01-01 1 | MARC | TESCO | M | Y | 2 | 5000 | 2020-01-02 2 | ANNA | MARTIN | F | N | 3 | 2000 | 2020-01-03 2

| ANNA | MARTIN | F | N | 4 | 3000 | 2020-01-04 3 | EMMA | JOHNSON | F | Y | 5 | 4000 | 2020-01-05 4 | DARIO | PENTAL | M | N | 6 | 5000 | 2020-01-06 5 | ELENA | SIMSON| F| N|7|6000|2020-01-07 6|TIM|ROBITH|M|N|8|7000|2020-01-08

7|MILA|MORRIS|F|N|9|8000|2020-01-09 To perform a LEFT JOIN, we can use the following SQL query:

SELECT * FROM customer LEFT JOIN in_store_transactions ON customer.customer_id = in_store_transactions.customer_id;

This query will return 15 rows of data, as shown below:

customer_id|name|lastname|gender|marital_status|transaction_id|amount|date 1|MARC|TESCO|M|Y|1|1000|2020-01-01
1|MARC|TESCO|M|Y|2|5000|2020-01-02 2|ANNA|MARTIN|F|N|3|2000|2020-01-03
2|ANNA|MARTIN|F|N|4|3000|202001-04 3|EMMA|JOHNSON|F|Y|5|4000|2020-01-05
4|DARIO|PENTAL|M|N|6|5000|2020-01-06 5|ELENA|SIMSON||F||N||7||6000||2020-01-07
6||TIM||ROBITH||M||N||8||7000||2020-01-08 7||MILA||MORRIS||F||N||9||8000||2020-01-09

8||JENNY||DWARTH||F||Y||NULL||NULL||NULL

As you can see, the customers who do not have any transactions (customer_id = 8) are still included in the result, but with NULL values for the transaction_id, amount, and date columns.

Therefore, the correct answer is C: INNER: 9 rows; LEFT: 15 rows.

Reference: SQL Joins - W3Schools

QUESTION 7

A data analyst is developing a data dictionary that aligns with a company\\'s data management processes and policies. Which of the following best describes what should be included in the data dictionary?

A. Information containing the links to business data

B. Information explaining the business methodologies

C. Information containing definitions of the business data

D. Information describing the data analysis phases

Correct Answer: C

QUESTION 8

Samantha needs to share a list of her organization\\'s top 50 customers with the VP of sales.

She would like to include the name of the customer, the business they represent, their contact information, and their total sales over the past year. The VP does not have any specialized analytics skills or software but would like to make some personal notes on the dataset.

What would be the best tool for Samantha to use to share this information?

A. Power BI.

B. Microsoft Excel.

C. Minitab.

D. SAS.

Correct Answer: B

Microsoft Excel.

This scenario presents a very simple use case where the business leader needs a dataset in an easy-to-access form and will not be performing any detailed analysis. A simple spreadsheet, such as Microsoft Excel, would be the best tool for

this job. There is no need to use a statistical analysis package, such as SAS or Minitab, as this would likely confuse the VP without adding any value. The same is true of an integrated analytics suite, such as Power BI.

QUESTION 9

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 10

A military commander would like to see the health scorecards of the troops daily and filter them based on gender and rank. Considering this data is PHI, which of the following would be the best way for the commander to view the information?

A. An emailed report

B. A password-protected dashboard

C. A daily printout of a report

D. A cloud-hosted spreadsheet

Correct Answer: B

A password-protected dashboard is a type of web-based application that can display the health scorecards of the troops

in a secure and interactive way. A password-protected dashboard can provide the following benefits for the commander: It can protect the PHI data from unauthorized access or disclosure by requiring a valid username and password to log in. This can ensure that only the commander and other authorized personnel can view the information12 It can allow the commander to filter the data based on gender and rank by using drop-down menus, sliders, checkboxes, or other controls. This can enable the commander to customize the view and focus on the relevant data13 It can update the data daily by connecting to a data source that refreshes automatically or on demand. This can ensure that the commander always sees the latest and most accurate information14 It can present the data in a visual and intuitive way by using charts, graphs, tables, or other elements. This can help the commander to understand and analyze the data more easily and effectively1

QUESTION 11

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 12

A sales analyst needs to report how the sales team is performing to target. Which of the following files will be important in determining 2019 performance attainment?

A. 2018 goal data

B. 2018 actual revenue

C. 2019 goal data

D. 2019 commission plan

Correct Answer: C

Answer: C. 2019 goal data To report how the sales team is performing to target, the sales analyst needs to compare the

actual sales revenue with the expected or planned sales revenue for the same period. The 2019 goal data is the file that contains the expected or planned sales revenue for the year 2019, which is the target that the sales team is aiming to achieve. By comparing the 2019 goal data with the 2019 actual revenue, the sales analyst can calculate the performance attainment, which is the percentage of the goal that was met by the sales team. Option A is incorrect, as 2018 goal data is not relevant for determining 2019 performance attainment. The 2018 goal data contains the expected or planned sales revenue for the year 2018, which is not the target that the sales team is aiming to achieve in 2019. Option B is incorrect, as 2018 actual revenue is not relevant for determining 2019 performance attainment. The 2018 actual revenue contains the actual sales revenue for the year 2018, which is not comparable with the 2019 goal data or the 2019 actual revenue. Option D is incorrect, as 2019 commission plan is not relevant for determining 2019 performance attainment. The 2019 commission plan contains the rules and rates for calculating and paying commissions to the sales team based on their performance attainment, but it does not contain the expected or planned sales revenue for the year 2019.