

CompTIA.Premium.DA0-001.30q - DEMO

Number: DA0-001
Passing Score: 800
Time Limit: 120 min



Exam Code: DA0-001
Exam Name: Data+
Certification Provider: CompTIA
Corresponding Certification: CompTIA Data+
Website: <https://VCEup.com/>
Free Exam: <https://vceup.com/exam-da0-001/>



Exam A**QUESTION 1**

Which one of the following values would not be appropriately stored in integer data type?

- A. -16
- B. 10921
- C. 0
- D. 1.2

Correct Answer: D

Section: (none)

Explanation

Explanation/Reference:

QUESTION 2

What data type is most appropriate for storing postal codes, such as, U.S. ZIP codes?

- A. Location.
- B. Float.
- C. Integer.
- D. String.

Correct Answer: D

Section: (none)

Explanation

Explanation/Reference:

QUESTION 3

Which of the following data elements would not normally be stored in binary format?

- A. Audio recording.
- B. Geo location.
- C. Video recording.
- D. Photograph.

Correct Answer: B

Section: (none)

Explanation

Explanation/Reference:

QUESTION 4

What effort has the goal of maintaining a single source of truth for the most important data elements handled by the organization?

- A. MDM (Master Data Management)
- B. IPS (Intrusion Prevention System)
- C. MAM (Mobile Application Management)
- D. DLP (Data Loss Prevention)

Correct Answer: A

Section: (none)

Explanation

Explanation/Reference:

Master data management (MDM) involves creating a single master record for each person, place, or thing in a business, from across internal and external data sources and applications. This information has been de-duplicated, reconciled and enriched, becoming a consistent, reliable source. Once created, this master data serves as a trusted view of business-critical data that can be managed and shared across the business to promote accurate reporting, reduce data errors, remove redundancy, and help workers make better-informed business decisions.

QUESTION 5

Data definitions should be written in plain language.

- A. True.
- B. False.

Correct Answer: A

Section: (none)

Explanation

Explanation/Reference:

QUESTION 6

What cybersecurity goal protects an organizations data from unauthorized modification?

- A. Confidentiality.
- B. Non-repudiation.
- C. Integrity.
- D. Availability.

Correct Answer: C

Section: (none)

Explanation

Explanation/Reference:

The term data integrity refers to the accuracy and consistency of data. When creating databases, attention needs to be given to data integrity and how to maintain it. A good database will enforce data integrity whenever possible. For example, a user could accidentally try to enter a phone number into a date field.

QUESTION 7

What is an example of data in transit?

- A. Data in memory on a computer.
- B. Data on a smartphone.
- C. Data on a hard disk.
- D. Data on a network.

Correct Answer: D

Section: (none)

Explanation

Explanation/Reference:

A data network is a system designed to transfer data from one network access point to one other or more network access points via data switching, transmission lines, and system controls. Data networks consist of communication systems such as circuit switches, leased lines, and packet switching networks.

QUESTION 8

Q3 2020 has just ended, and now a data analyst needs to create an ad-hoc sales report that demonstrates how well the Q3 2020 promotion went versus last year's Q3 promotion. Which of the following date parameters should the analyst use?

- A. 2019 vs. YTD 2020
- B. Q3 2019 vs. Q3 2020
- C. YTD 2019 vs. YTD 2020
- D. Q4 2019 vs. Q3 2020

Correct Answer: B

Section: (none)

Explanation

Explanation/Reference:

QUESTION 9

A data analyst has been asked to create an ad-hoc sales report for the Chief Executive Officer (CEO). Which of the following should be included in the report?

- A. The sales representatives' home addresses.
- B. Line-item SKU numbers.
- C. YTD total sales.
- D. The customers' first and last names.

Correct Answer: C

Section: (none)

Explanation

Explanation/Reference:

QUESTION 10

Which of the following can be used to translate data into another form so it can only be read by a user who has a key or a password?

- A. Data encryption.
- B. Data transmission.
- C. Data protection.
- D. Data masking.

Correct Answer: A

Section: (none)

Explanation

Explanation/Reference:

Data encryption is a way of translating data from plaintext (unencrypted) to ciphertext (encrypted). Users can access encrypted data with an encryption key and decrypted data with a decryption key.

QUESTION 11

Which of the following is an example of a discrete data type?

- A. 8in (20cm)
- B. 5 kids
- C. 2.5mi (4km)
- D. 10.71bs (4.9kg)

Correct Answer: B

Section: (none)

Explanation

Explanation/Reference:

QUESTION 12

Which of the following contains alphanumeric values?

- A. 10.1E2
- B. 13.6
- C. 1347

D. A317

Correct Answer: D

Section: (none)

Explanation

Explanation/Reference:

QUESTION 13

A junior web developer is developing a new application where users can upload short videos. The first task is to create a homepage that shows the headline 'Upload Your Short Videos' and a clickable button that says 'upload now'. Which of the following HTML commands would help the developer to complete the task successfully?

- A. < span >Upload Your Short Videos< /span >< button >upload now< /button >
- B. < p >Upload Your Short Videos< /p > < p >upload now< /p >
- C. < h1>Upload Your Short Videos< /h1 > < button >upload now< /button >
- D. < h1 >Upload Your Short Videos< /h1 > < h1 >upload now< /h1 >

Correct Answer: C

Section: (none)

Explanation

Explanation/Reference:

The correct answer is: Upload Your Short Videos upload now The two tags are used to define HTML headings. defines the most important heading. defines the least important heading.

Note: Only use one per page - this should represent the main heading/subject for the whole page. The tag defines a clickable button.

QUESTION 14

A web developer wants to ensure that malicious users can't type SQL statements when they asked for input, like their username/userid. Which of the following query optimization techniques would effectively prevent SQL Injection attacks?

- A. Indexing.
- B. Subset of records.
- C. Temporary table in the query set.
- D. Parameterization.

Correct Answer: D

Section: (none)

Explanation

Explanation/Reference:

The correct answer is D: Parameterization. Parameterized SQL queries allow you to place parameters in an SQL query instead of a constant value. A parameter takes a value only when the query is executed, allowing the query to be reused with different values and purposes. Parameterized SQL statements are available in some analysis clients, and are also available through the Historian SDK.

For example, you could create the following conditional SQL query, which contains a parameter for the collector's name: SELECT* FROM ExamsDigest WHERE coursename=? ORDER BY tagname SQL Injection is best prevented through the use of parameterized queries.

QUESTION 15

Consider the following dataset which contains information about houses that are for sale:

```
sonery=# select * from melb limit 5;
```

houseid	address	regionname	type	rooms	date	distance	price
1	85 Turner St	Northern Metropolitan	h	2	2016-03-12	2.5	1.48e+06
2	25 Bloomburg St	Northern Metropolitan	h	2	2016-04-02	2.5	1.035e+06
3	5 Charles St	Northern Metropolitan	h	3	2017-04-03	2.5	1.465e+06
4	40 Federation La	Northern Metropolitan	h	3	2017-04-03	2.5	850000
5	55a Park St	Northern Metropolitan	h	4	2016-04-06	2.5	1.6e+06

(5 rows)

Which of the following string manipulation commands will combine the address and region name columns to create a full address?

full_address 85 Turner St, Northern Metropolitan 25 Bloomburg St, Northern Metropolitan 5 Charles St, Northern Metropolitan 40 Federation La, Northern Metropolitan 55a Park St, Northern Metropolitan

- A. SELECT CONCAT(address, ' ', regionname) AS full_address FROM melb LIMIT 5;
- B. SELECT CONCAT(address, '-', regionname) AS full_address FROM melb LIMIT 5;

- C. SELECT CONCAT(regionname, ' , ' , address) AS full_address FROM melb LIMIT 5
 D. SELECT CONCAT(regionname, '-' , address) AS full_address FROM melb LIMIT 5;

Correct Answer: A

Section: (none)

Explanation

Explanation/Reference:

The correct answer is A: SELECT CONCAT(address, ' , ' , regionname) AS full_address FROM melb LIMIT 5; String manipulation (or string handling) is the process of changing, parsing, splicing, pasting, or analyzing strings. SQL is used for managing data in a relational database. The CONCATQ function adds two or more strings together. Syntax CONCAT(string1, string2,... string_n) Parameter Values Parameter Description string1, string2, string_n Required. The strings to add together.

QUESTION 16

The ACME Corporation hired an analyst to detect data quality issues in their excel documents. Which of the following are the most common issues? (Select TWO)

- A. Apostrophe.
- B. Commas.
- C. Symbols.
- D. Duplicates.
- E. Misspellings.

Correct Answer: DE

Section: (none)

Explanation

Explanation/Reference:

1. Duplicates
2. Misspellings

The most common data quality issues are difficult to resolve in Excel because of their rigidity. It forces analysts to do a ton of manual work which results in a high probability of an error being introduced to the data set. Those common issues include:

- Blanks
- Nulls
- Outliers
- Duplicates
- Extra spaces
- Misspellings
- Abbreviations and domain-specific variations
- Formula error codes

When introduced, these errors can skew or even invalidate the resulting analysis. A smart tool would minimize the possibility of error by automating the manual work. In Excel, you might look for data quality issues in one of two ways. First, you might use auto filters on specific columns to scan for anomalies and blanks or you might use a pivot table to find gaps and discrepancies.

In either case, you're scanning for the anomalies yourself. Suffice it to say that's not a very efficient process. It also means accuracy is only as good as the analyst's eye, so the probability of error varies throughout the day.

QUESTION 17

Consider this dataset showing the retirement age of 11 people, in whole years:

54, 54, 54, 55, 56, 57, 57, 58, 58, 60, 60

This tables show a simple frequency distribution of the retirement age data.

Age	Frequency
54	3
55	1
56	1
57	2
58	2
60	2

- A. 56
- B. 55
- C. 57
- D. 54

Correct Answer: D

Section: (none)

Explanation

Explanation/Reference:

A measure of central tendency (also referred to as measures of centre or central location) is a summary measure that attempts to describe a whole set of data with a single value that represents the middle or centre of its distribution.

There are three main measures of central tendency: the mode, the median and the mean. Each of these measures describes a different indication of the typical or central value in the distribution.

What is the mode?

The mode is the most commonly occurring value in a distribution.

The most commonly occurring value is 54, therefore the mode of this distribution is 54 years.

QUESTION 18

Which of the following value is the measure of dispersion "range between the scores of ten students in a test.

The scores of ten students in a test are 17, 23, 30, 36, 45, 51, 58, 66, 72, 77.

- A. 90
- B. 60
- C. 70
- D. 80

Correct Answer: B

Section: (none)

Explanation

Explanation/Reference:

The correct answer is: 60

Range is the interval between the highest and the lowest score.

Range is a measure of variability or scatteredness of the varieties or observations among themselves and does not give an idea about the spread of the observations around some central value.

Symbolically $R = H_s - L_s$.

Where R = Range; H_s is the 'Highest score' and L_s is the Lowest Score.

The scores of ten students in a test are: 17, 23, 30, 36, 45, 51, 58, 66, 72, 77.

The highest score is 77 and the lowest score is 17.

So the range is the difference between these two scores $\text{Range} = 77 - 17 = 60$

QUESTION 19

A data scientist wants to see which products make the most money and which products attract the most customer purchasing interest in their company. Which of the following data manipulation techniques would he use to obtain this information?

- A. Data append
- B. Data blending
- C. Normalize data
- D. Data merge

Correct Answer: B

Section: (none)

Explanation

Explanation/Reference:

The correct answer is B: Data blending.

Data blending is combining multiple data sources to create a single, new dataset, which can be presented visually in a dashboard or other visualization and can then be processed or analyzed. Enterprises get their data from a variety of sources, and users may want to temporarily bring together different datasets to compare data relationships or answer a specific question. Data append is incorrect.

Data append is a process that involves adding new data elements to an existing database. An example of a common data append would be the enhancement of a company's customer files. A data append takes the information they have, matches it against a larger database of business data, allowing the desired missing data fields to be added. Normalize data is incorrect.

Data normalization is the process of structuring your relational customer database, following a series of normal forms. This improves the accuracy and integrity of your data while ensuring that your database is easier to navigate. Data merge is incorrect. Data merging is the process of combining two or more data sets into a single data set.

QUESTION 20

A data analyst wants to create "Income Categories" that would be calculated based on the existing variable "Income". The "Income Categories" would be as follows:

Income category 1: less than \$1.

Income category 2: more than \$1 and less than \$20,000.

Income category 3: more than \$20,001 and less than \$40,000.

Income category 4: more than \$40,001.

Which of the following data manipulation techniques should the data analyst use to create "Income Categories"?

- A. Data merge
- B. Derived variables
- C. Data blending
- D. Data append

Correct Answer: B

Section: (none)

Explanation

Explanation/Reference:

The correct answer is B: Derived variables. Derived variables are variables that you create by calculating or categorizing variables that already exist in your data set.

Data merge is incorrect. Data merging is the process of combining two or more data sets into a single data set. Data blending is incorrect.

Data blending involves pulling data from different sources and creating a single, unique, dataset for visualization and analysis.

Data append is incorrect. A data append is a process that involves adding new data elements to an existing database.

QUESTION 21

Angela is aggregating data from CRM system with data from an employee system.

While performing an initial quality check, she realizes that her employee ID is not associated with her identifier in the CRM system.

What kind of issues is Angela facing?

Choose the best answer.

- A. ETL process.
- B. Record linkage.
- C. ELT process.
- D. System integration.

Correct Answer: B

Section: (none)

Explanation

Explanation/Reference:

While this scenario describes a system integration challenge that can be solved with ETL or ELT. Angela is facing a Record linkage issue.

QUESTION 22

Andy is a pricing analyst for a retailer. Using a hypothesis test, he wants to assess whether people who receive electronic coupons spend more on average. What should Andy's null hypothesis be?

- A. People who receive electronic coupons spend more on average.
- B. People who receive electronic coupons spend less on average.
- C. People who receive electronic coupons do not spend more on average.
- D. People who do not receive electronic coupons spend more on average.

Correct Answer: C

Section: (none)

Explanation

Explanation/Reference:

The null hypothesis presumes the status quo. Andy is testing whether or not people who receive an electronic coupon spend more on average, so, the null hypothesis states that people who receive the coupon do spend more on average.

QUESTION 23

Amanda needs to create a dashboard that will draw information from many other data sources and present it to business leaders. Which one of the following tools is least likely to meet her needs?

- A. QuickSight.
- B. Tableau.
- C. Power BI.
- D. SPSS Modeler.

Correct Answer: D

Section: (none)

Explanation

Explanation/Reference:

SPSS Modeler.

QuickSight, Tableau, and Power BI are all powerful analytics and reporting tools that can pull data from a variety of sources. SPSS Modeler is a powerful predictive analytics platform that is designed to bring predictive intelligence to decisions made by individuals, groups, systems and your enterprise.

QUESTION 24

Daniel is using the structured Query language to work with data stored in relational database. He would like to add several new rows to a database table. What command should he use?

- A. SELECT.
- B. ALTER.
- C. INSERT.
- D. UPDATE.

Correct Answer: C

Section: (none)

Explanation

Explanation/Reference:

INSERT

The INSERT command is used to add new records to a database table.

The SELECT command is used to retrieve information from a database. It's the most commonly used command in SQL because it is used to pose queries to the database and retrieve the data that you're interested in working with.

The UPDATE command is used to modify rows in the database.

The CREATE command is used to create a new table within your database or a new database on your server.

QUESTION 25

Jhon is working on an ELT process that sources data from six different source systems. Looking at the source data, he finds that data about the sample people exists in two of six systems. What does he have to make sure he checks for in his ELT process? Choose the best answer.

- A. Duplicate Data.
- B. Redundant Data.
- C. Invalid Data.
- D. Missing Data.

Correct Answer: C
Section: (none)
Explanation

Explanation/Reference:

Duplicate Data.

While invalid, redundant, or missing data are all valid concerns, data about people exists in two of the six systems. As such, Jhon needs to account for duplicate data issues.

QUESTION 26

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
Section: (none)
Explanation

Explanation/Reference:

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 27

Alex wants to use data from his corporate sale, CRM, and shipping systems to try and predict future sales. Which of the following systems is the most appropriate? Choose the best answer.

- A. Data mart.
- B. OLAP.
- C. Data Warehouse.
- D. OLTP.

Correct Answer: C
Section: (none)
Explanation

Explanation/Reference:

Correct answer. C. Data Warehouse.

Data warehouse bring together data from multiple systems used by an organization.

A data mart is too narrow, as Alex needs data from across multiple divisions.

OLAP is a broad term of analytical processing, and OLTP systems are transactional and not ideal for this task.

QUESTION 28

Analytics reports should follow corporate style guidelines.

- A. True.
- B. False.

Correct Answer: A

Section: (none)

Explanation

Explanation/Reference:

QUESTION 29

Which one of the following is a measure of dispersion?

- A. Variance.
- B. Mode.
- C. Median.
- D. Mean.

Correct Answer: A

Section: (none)

Explanation

Explanation/Reference:

QUESTION 30

Which one of the following is NOT a common data integration tool?

- A. XSS
- B. ELT
- C. ETL
- D. APIs

Correct Answer: A

Section: (none)

Explanation

Explanation/Reference:

Cross-site Scripting (XSS) is a security vulnerability usually found in websites and/or web applications that accept user input.

XSS is a client-side vulnerability that targets other application users, while SQL injection is a server-side vulnerability that targets the applications database. How do I prevent XSS in PHP? Filter your inputs with a whitelist of allowed characters and use type hints or type casting.