

# Prep4sureGuide

## WELCOME USE TEST ENGINE

Prepare your actual test with our sure pass exam guide for successful result

Input your exam code ...

Our sure prep material equipped with the highest experts team and the most authoritative exam items plus the best service, which can ensure you 100% pass. Besides, our exam training guide can support both the fastest delivery speed and the shortest time to get all knowledge.



### Quality and Value

Prep4sureGuide Practice Exams are written to the highest standards of technical accuracy, using only certified subject matter experts and published authors for development - no all dumps.



### Tested and Approved

We are committed to the process of vendor and third party approvals. We believe professionals and executives alike deserve the confidence of quality coverage these authorizations provide.



### Easy to Pass

If you prepare for the exams using our Prep4sureGuide testing engine, It is easy to succeed for all certifications in the first attempt. You don't have to deal with all dumps or any free torrent / rapidshare all stuff.



### Try Before Buy

Prep4sureGuide offers free demo of each product. You can check out the interface, question quality and usability of our practice exams before you decide to buy.



HAPPY CUSTOMERS

32694



DOWNLOADS

62152



TEAM MEMBERS

32694



SHARES

56692

<http://www.prep4sureguide.com>

Prepare your actual test with our sure pass exam guide for successful result

**Exam** : **DA0-002**

**Title** : **CompTIA Data+ Exam**

**Vendor** : **CompTIA**

**Version** : **DEMO**

**NO.1** Which of the following best describes an assessment a data analyst would use to validate that the number of records in a dataset matches the expected results?

- A. Source control
- B. Unit test
- C. Stress test
- D. Health check

**Answer:** B

**NO.2** A data engineer needs to add a process that will run after an employee table is loaded. The engineer does not know the name of the process that loads employee data. Which of the following is the best document for the engineer to review to find this information?

- A. Data flow diagram
- B. Data dictionary
- C. Data explainability report
- D. Data lineage report

**Answer:** D

Explanation:

A data lineage report documents the movement and transformation of data throughout systems and processes. It shows where data originates, how it is processed, and the dependencies between processes. Reviewing the data lineage report would help the engineer identify the process responsible for loading the employee table and determine where to add the subsequent process.

**NO.3** Which of the following is the most efficient to use when programming repeatable tasks?

- A. LLM
- B. Deep learning
- C. NLP
- D. RPA

**Answer:** D

Explanation:

Robotic process automation is designed to automate repetitive, rule-based tasks, making it the most efficient choice for repeatable processes.

**NO.4** A data analyst needs to remove all duplicate values between two tables, "Employees" and "Managers," using SQL SELECT statements. Which of the following should the analyst use for this task?

- A. SELECT \* FROM Employees UNION ALL SELECT \* FROM Managers
- B. SELECT \* FROM Employees UNION SELECT \* FROM Managers
- C. SELECT \* FROM Employees JOIN SELECT \* FROM Managers
- D. SELECT \* FROM Employees CROSS JOIN SELECT \* FROM Managers

**Answer:** B

Explanation:

UNION combines the results of two SELECT statements and automatically removes duplicate rows from the combined result set. This makes it the appropriate choice when the requirement is to eliminate all duplicate values between the Employees and Managers tables.

**NO.5** Which of the following is a flat file data format?

- A. XML
- B. CSV
- C. JSON
- D. HTML

**Answer:** B

Explanation:

CSV (Comma-Separated Values) is a flat file format storing data in plain text, organized in rows and columns without hierarchical structure.

**NO.6** Which of the following is the best example of a slowly changing dimension for a data warehouse?

- A. A customer table that is overwritten every morning without keeping history
- B. A piece of historical customer data that is removed from the table
- C. A new table created each day with updated customer data
- D. An address from the customer table that is overwritten when changed

**Answer:** D

Explanation:

A slowly changing dimension refers to attributes that change over time, such as a customer's address, where updates occur to reflect the latest value in the dimension table.

**NO.7** A data analyst learns that a report detailing employee sales is reflecting sales only for the current month. Which of the following is the most likely cause?

- A. Lack of permissions
- B. An error in SQL code
- C. Report refresh failure
- D. Connectivity issues

**Answer:** B

**NO.8** A data analyst needs to create a combined report that includes information from the following two tables:

**Managers table**

ID	First_name	Last_name	Job_title
1001	John	Doe	Manager
1002	Jane	Roe	Director

---

**Non-managers table**

ID	First_name	Last_name	Job_title
1003	Robert	Roe	Business Analyst
1004	Jane	Doe	Sales Representative
1005	John	Roe	Operations Analyst

Which of the following query methods should the analyst use for this task?

- A. Group
- B. Join
- C. Union
- D. Nested

**Answer:** C

Explanation:

A union combines rows from two tables with the same structure into a single result set, making it appropriate for merging managers and non-managers into one report.

**NO.9** An analyst needs to combine the following two tables. The analyst needs to keep all the records in the Medals table and only the records that match from the Country table.

**Medals**

Person_ID	Medal	Sport
1	Bronze	Cycling
2	Gold	Cricket
3	Silver	Soccer

**Country**

Person_ID	Country
1	USA
2	CAN
3	BRA
4	FRA
5	SWE
6	JPN

Which of the following methods should the analyst use for this task?

- A. Join
- B. Insert
- C. Union
- D. Append

**Answer:** A

Explanation:

A join is used to combine rows from two tables based on a common field, in this case Person\_ID. Since all records from the Medals table must be retained and only matching records from the Country table included, the appropriate operation is a left join, which is a type of join.

**NO.10** An administrator needs to design a table that will include foreign words. Which of the following is the best option for storing non-native language characters?

- A. VARCHAR
- B. NVARCHAR
- C. CLOB
- D. CHAR

**Answer:** B

**NO.11** A data analyst must combine service calls into low-, medium-, and high-priority levels in order to analyze organizational responses. Which of the following techniques should the analyst use for this task?

- A. Augmentation
- B. Imputation

C. Scaling

D. Binning

**Answer:** D

Explanation:

Binning groups values into defined categories, such as low, medium, and high priority levels, enabling structured analysis of service call responses.

**NO.12** A database administrator is investigating the reason a reporting job failed. The target-layer schema remained unchanged from the last run, and the database server is scalable. All of the users have default execute privileges on the report. Which of the following attribute changes most likely caused this incident?

A. Data lineage

B. Data type

C. Data volume

D. Data access

**Answer:** C

Explanation:

A reporting job that previously ran successfully can fail when the volume of data increases beyond expected limits, causing timeouts, memory constraints, storage issues, or processing bottlenecks. Since the schema was unchanged, the server is scalable, and user permissions remained available, a change in data volume is the most likely cause of the failure.

**NO.13** Which of the following is the best reason for a company to use a CSV file to share data instead of an Excel file?

A. CSV files can store different types of encoding.

B. CSV files are not vendor-specific.

C. CSV files are smaller in size.

D. CSV files are easier to change in text editors.

**Answer:** B

Explanation:

CSV files are not tied to a specific software vendor and can be opened, imported, and processed by a wide variety of applications and platforms. This makes CSV an ideal format for sharing data between different systems and organizations.

**NO.14** Which of the following is a data warehouse term for a data set specifically created for drill-down reports?

A. Dimension

B. Measure

C. Cube

D. Cluster

**Answer:** C

Explanation:

A cube organizes data in multiple dimensions and supports drill-down analysis, allowing users to explore data at different levels of detail.

**NO.15** A company experiences a data breach that exposes some customers' information. A data analyst must categorize the customer data into fully affected, partially affected, and not affected. Which of the following is the best technique to use to quantify the extent of the breach?

- A. Standardizing
- B. Appending
- C. Binning
- D. Parsing

**Answer:** C

Explanation:

Binning groups data into predefined categories or ranges. Classifying customers as fully affected, partially affected, and not affected is a form of binning that enables the analyst to quantify and measure the extent of the breach by placing records into meaningful categories.

**NO.16** A data analyst is creating a document that covers the guidelines and transformation of the suite of sales dashboards that they are responsible for delivering. Which of the following is the analyst most likely creating?

- A. Data flow diagram
- B. Data lineage
- C. Data dictionary
- D. Data versioning

**Answer:** B

Explanation:

Data lineage documents how data is transformed, processed, and flows through systems, including the steps and rules applied to create outputs such as dashboards.

**NO.17** While interacting with a data set, a data analyst learns that copies of the data are distributed across several data repositories around the globe. Which of the following concepts best describes this practice?

- A. Redundancy
- B. High availability
- C. Replication
- D. Duplication

**Answer:** C

Explanation:

Replication involves distributing copies of data across multiple locations or systems to ensure accessibility and consistency across environments.

**NO.18** Which of the following data repositories should a company use when structured data about the whole company needs to be stored in a predefined data structure?

- A. Data mart
- B. Data warehouse
- C. Data silo
- D. Data lake

**Answer:** B

Explanation:

A data warehouse is designed to store structured data from across an entire organization in a predefined schema. It integrates data from multiple sources, supports enterprise-wide reporting and analytics, and provides a consistent view of business information.

**NO.19** A product goes viral on social media, creating high demand. Distribution channels are facing supply chain issues because the testing and training models that are used for sales forecasting have not encountered similar demand. Which of the following best describes this situation?

- A. Model bias
- B. Data drift
- C. Incorrect sizing
- D. Skewing

**Answer:** B

Explanation:

Data drift occurs when real-world data patterns change from what the model was trained on, such as an unexpected surge in demand that the model has not previously encountered, leading to inaccurate predictions.

**NO.20** Which of the following is a reason for sampling?

- A. Data needs to be collected without interference.
- B. The data extraction process needs to be automated.
- C. Many individuals need to be asked a series of questions.
- D. Conclusions need to be drawn from a large population.

**Answer:** D

Explanation:

Sampling is used to draw conclusions about a large population by analyzing a representative subset of data.

**NO.21** A data analyst is cleaning up a data set that contains images. The analyst wants to keep everything that was produced within the last 12 months. Which of the following best describes where the analyst should check for the date of creation on these products?

- A. The data dictionary of the data set
- B. The metadata of each item in the data set
- C. The data set creation date on the database
- D. The folder creation date on the filesystem

**Answer:** B

Explanation:

Metadata contains descriptive information about each file or image, including attributes such as creation date, modification date, file size, and format. To determine which images were produced within the last 12 months, the analyst should examine the metadata of each item rather than the creation date of the dataset or folder containing the files.