

COMPTIA DATA+ CERTIFICATION (DAO-002) PRACTICE TEST (SAMPLE)

STUDY GUIDE



SAMPLE STUDY GUIDE WITH LIMITED QUESTIONS

VISIT US ONLINE FOR
THE FULL VERSION STUDY GUIDE

<https://dao002.examzify.com>



Copyright © 2026 by Examzify - A Kaluba Technologies Inc. product.

ALL RIGHTS RESERVED.

No part of this book may be reproduced or transferred in any form or by any means, graphic, electronic, or mechanical, including photocopying, recording, web distribution, taping, or by any information storage retrieval system, without the written permission of the author.

Notice: Examzify makes every reasonable effort to obtain accurate, complete, and timely information about this product from reliable sources.

SAMPLE

Table of Contents

Copyright	1
Table of Contents	2
Introduction	3
How to Use This Guide	4
Questions	5
Answers	8
Explanations	10
Next Steps	15

SAMPLE

Introduction

Preparing for a certification exam can feel overwhelming, but with the right tools, it becomes an opportunity to build confidence, sharpen your skills, and move one step closer to your goals. At Examzify, we believe that effective exam preparation isn't just about memorization, it's about understanding the material, identifying knowledge gaps, and building the test-taking strategies that lead to success.

This guide was designed to help you do exactly that.

Whether you're preparing for a licensing exam, professional certification, or entry-level qualification, this book offers structured practice to reinforce key concepts. You'll find a wide range of multiple-choice questions, each followed by clear explanations to help you understand not just the right answer, but why it's correct.

The content in this guide is based on real-world exam objectives and aligned with the types of questions and topics commonly found on official tests. It's ideal for learners who want to:

- Practice answering questions under realistic conditions,
- Improve accuracy and speed,
- Review explanations to strengthen weak areas, and
- Approach the exam with greater confidence.

We recommend using this book not as a stand-alone study tool, but alongside other resources like flashcards, textbooks, or hands-on training. For best results, we recommend working through each question, reflecting on the explanation provided, and revisiting the topics that challenge you most.

Remember: successful test preparation isn't about getting every question right the first time, it's about learning from your mistakes and improving over time. Stay focused, trust the process, and know that every page you turn brings you closer to success.

Let's begin.

How to Use This Guide

This guide is designed to help you study more effectively and approach your exam with confidence. Whether you're reviewing for the first time or doing a final refresh, here's how to get the most out of your Examzify study guide:

1. Start with a Diagnostic Review

Skim through the questions to get a sense of what you know and what you need to focus on. Your goal is to identify knowledge gaps early.

2. Study in Short, Focused Sessions

Break your study time into manageable blocks (e.g. 30 – 45 minutes). Review a handful of questions, reflect on the explanations.

3. Learn from the Explanations

After answering a question, always read the explanation, even if you got it right. It reinforces key points, corrects misunderstandings, and teaches subtle distinctions between similar answers.

4. Track Your Progress

Use bookmarks or notes (if reading digitally) to mark difficult questions. Revisit these regularly and track improvements over time.

5. Simulate the Real Exam

Once you're comfortable, try taking a full set of questions without pausing. Set a timer and simulate test-day conditions to build confidence and time management skills.

6. Repeat and Review

Don't just study once, repetition builds retention. Re-attempt questions after a few days and revisit explanations to reinforce learning. Pair this guide with other Examzify tools like flashcards, and digital practice tests to strengthen your preparation across formats.

There's no single right way to study, but consistent, thoughtful effort always wins. Use this guide flexibly, adapt the tips above to fit your pace and learning style. You've got this!

Questions

SAMPLE

- 1. Which statement about data lineage is true?**
 - A. It only tracks data movement within a single database.
 - B. It is documented using source code comments.
 - C. It traces data from source to destination across transformations and is documented via lineage diagrams, metadata catalogs, and automated lineage tools.
 - D. It is not useful for governance.
- 2. Which activity is part of a data governance process that involves documenting data sources and communicating data limitations?**
 - A. Developing data storage hardware.
 - B. Setting network firewall rules.
 - C. Performing data backups.
 - D. Documenting data sources and communicating data limitations.
- 3. Which deployment approach enables running applications with their dependencies across different environments?**
 - A. Virtualization
 - B. Version control
 - C. Continuous integration
 - D. Containerization
- 4. Which join type returns all records from both tables including unmatched rows, with NULLs where there is no match?**
 - A. INNER JOIN
 - B. FULL OUTER JOIN
 - C. CROSS JOIN
 - D. LEFT OUTER JOIN
- 5. What is ELT and why is advantageous with cloud data warehouses?**
 - A. It loads raw data into the warehouse and performs transformations inside the warehouse.
 - B. It loads transformed data into the warehouse and then stores raw data separately.
 - C. It transforms data before loading to the warehouse.
 - D. It avoids using the warehouse compute.

- 6. Which category of analytics is used to summarize data and describe what has happened?**
- A. Descriptive
 - B. Diagnostic
 - C. Predictive
 - D. Prescriptive
- 7. Before creating dashboards for multiple teams, which step should be taken first?**
- A. Determine the number of team members who need access
 - B. Evaluate the user persona type for the dashboards.
 - C. Determine the delivery method of the dashboards
 - D. Evaluate the KPIs for the dashboards.
- 8. What approach supports accessibility for visually impaired stakeholders when presenting dashboards?**
- A. Add alt text to images only
 - B. Increase color contrasts only
 - C. Incorporate audio accessibility functions into graphics for all instances of the interface
 - D. Create more charts with hover interactivity
- 9. Which tool would a data analyst most likely use to modify a simple data connection property in a configuration file?**
- A. RStudio
 - B. SSMS
 - C. Notepad++
 - D. Toad
- 10. In data warehousing, what term refers to a dataset created specifically for drill-down reports?**
- A. Dimension
 - B. Measure
 - C. Cube
 - D. Cluster

Answers

SAMPLE

1. C
2. D
3. D
4. B
5. A
6. A
7. B
8. C
9. C
10. C

SAMPLE

Explanations

SAMPLE

1. Which statement about data lineage is true?

- A. It only tracks data movement within a single database.
- B. It is documented using source code comments.
- C. It traces data from source to destination across transformations and is documented via lineage diagrams, metadata catalogs, and automated lineage tools.**
- D. It is not useful for governance.

Data lineage tracks where data originates, how it moves, and how it's transformed from source to final destination across the entire data pipeline. It isn't limited to a single database; it follows data through multiple systems and processes, showing every transformation and movement along the way. This understanding is documented with lineage diagrams that visualize data flow, metadata catalogs that describe data assets and their relationships, and automated lineage tools that capture and maintain the lineage over time. This combination provides a complete picture of data provenance, which is essential for governance, impact analysis, and auditing. For example, if a revenue figure comes from a source file, undergoes several ETL steps, and ends up in a data warehouse, lineage would show the source, each transformation, and the destination. The idea that lineage is documented only in source code comments or that it isn't useful for governance doesn't fit how lineage is used in practice.

2. Which activity is part of a data governance process that involves documenting data sources and communicating data limitations?

- A. Developing data storage hardware.
- B. Setting network firewall rules.
- C. Performing data backups.
- D. Documenting data sources and communicating data limitations.**

In data governance, documenting data sources and communicating data limitations is all about metadata management and data lineage. By capturing where data comes from, how it was collected, what transformations were applied, and any known constraints (such as timeliness, completeness, or biases), you create context that helps data producers, stewards, and users evaluate trustworthiness and appropriate use. This transparency supports data quality, compliance, and informed decision-making, since people know the provenance and limits of the data they rely on. The other activities focus on infrastructure and operations rather than governance documentation: building storage hardware relates to system capacity and architecture, configuring firewall rules pertains to security, and performing backups deals with data resilience and recovery. While important, they don't embody the governance practice of documenting sources and communicating limitations.

3. Which deployment approach enables running applications with their dependencies across different environments?

- A. Virtualization
- B. Version control
- C. Continuous integration
- D. Containerization**

Containerization packages the application together with its runtime, libraries, and configuration into a portable unit that can run the same way in different environments. These containers are lightweight and share the host's kernel, so the same image behaves consistently whether it's deployed in development, testing, or production, on-premises or in the cloud. This "build once, run anywhere" approach minimizes environment-specific issues and simplifies deployment. Virtualization, while allowing isolation, uses full virtual machines with their own OS, which is heavier and not as streamlined for moving just the application and its dependencies. Version control tracks code changes, and continuous integration automates building and testing, but neither by themselves ensures consistent runtime environments across deployments. Hence, containerization is the best fit.

4. Which join type returns all records from both tables including unmatched rows, with NULLs where there is no match?

- A. INNER JOIN
- B. FULL OUTER JOIN**
- C. CROSS JOIN
- D. LEFT OUTER JOIN

The concept tested is how join types handle rows that don't have a match in the other table. A full outer join returns every row from both tables, and when a row from one table has no matching row in the other, the missing side is filled with NULLs. That's why it's the correct choice here: you see all records from both sides, with NULLs where there is no match. Inner join would only include rows with matches in both tables, so unmatched rows are excluded. Cross join produces every possible combination of rows from the two tables, not about matching at all. Left outer join includes all rows from the left table (with NULLs on the right when there's no match) but does not bring in unmatched rows from the right table.

5. What is ELT and why is advantageous with cloud data warehouses?

- A. It loads raw data into the warehouse and performs transformations inside the warehouse.**
- B. It loads transformed data into the warehouse and then stores raw data separately.
- C. It transforms data before loading to the warehouse.
- D. It avoids using the warehouse compute.

ELT means loading data in its raw form into the data warehouse and then performing transformations inside the warehouse using its processing power. In cloud data warehouses, storage and compute can scale independently, so you can ingest large volumes of raw data quickly without pre-processing. Once the data is inside the warehouse, you transform it with SQL or built-in processing, create curated tables or views, and reuse the raw data for future analyses. This approach is advantageous because it reduces data movement and pipeline complexity, preserves the original raw data for flexibility, and lets you leverage scalable, on-demand compute to run intensive transformations as needed. It also aligns well with handling semi-structured data and evolving analytics requirements. The other options describe pre-transforming data before loading, or avoiding warehouse compute, which misses the main benefit of ELT in a cloud environment.

6. Which category of analytics is used to summarize data and describe what has happened?

- A. Descriptive**
- B. Diagnostic
- C. Predictive
- D. Prescriptive

Descriptive analytics focuses on summarizing historical data to describe what happened. It uses simple measures and visuals—totals, averages, counts, distributions, and dashboards—to provide a clear picture of past performance. For example, a monthly sales report showing total revenue, average order value, and transaction counts tells you what occurred during that period. This sets the stage for deeper analysis by offering a clear summary before exploring why it happened (diagnostic), what might happen next (predictive), or what actions to take (prescriptive).

7. Before creating dashboards for multiple teams, which step should be taken first?

- A. Determine the number of team members who need access
- B. Evaluate the user persona type for the dashboards.**
- C. Determine the delivery method of the dashboards
- D. Evaluate the KPIs for the dashboards.

Understanding who will use the dashboards is the first step because it determines what to show, how to present it, and how the data should be interacted with. By evaluating the user personas, you identify each group's role, decisions they need to support, and the level of detail they require. This drives which KPIs are relevant, what visualizations work best, and how to tailor the content and interactivity for different audiences. It also helps set appropriate access and security needs from the outset. Once you know the personas, decisions about delivery method and specific KPIs can be aligned to those users, making the dashboards more useful and adopted across teams.

8. What approach supports accessibility for visually impaired stakeholders when presenting dashboards?

- A. Add alt text to images only
- B. Increase color contrasts only
- C. Incorporate audio accessibility functions into graphics for all instances of the interface**
- D. Create more charts with hover interactivity

The key idea is making dashboards usable for visually impaired users by providing non-visual ways to access the information. Adding audio accessibility functions that describe graphics for every part of the interface ensures charts and data can be understood through sound, captions, or speech, and that this accessibility is available regardless of how the user navigates (keyboard, screen reader, etc.). This approach directly supports perceivable and navigable information for those who can't rely on visuals. Partial measures like alt text help some images but often don't convey the full data story of a chart. Simply increasing color contrast aids readability for some users but doesn't help those who rely on non-visual access. More charts with hover interactivity can create additional barriers for screen readers or keyboard-only users and doesn't guarantee accessible descriptions. The audio-enabled, across-the-interface approach provides a comprehensive solution for visually impaired stakeholders.

9. Which tool would a data analyst most likely use to modify a simple data connection property in a configuration file?

- A. RStudio
- B. SSMS
- C. Notepad++**
- D. Toad

Editing a simple data connection property stored in a configuration file is about quickly and safely changing plain text without the overhead of a full development or database tool. Notepad++ is ideal for this task because it's a lightweight, general-purpose text editor designed for working with plain text files such as configuration files that use simple key-value pairs, JSON, XML, or YAML. With Notepad++, you can open the config file, locate the data connection property, modify its value, and save the file easily. It handles plain text well and provides handy features like search and replace and straightforward line editing, which makes updating a single property quick and low-risk. The other tools are oriented toward specific development or database-management tasks. RStudio is aimed at data analysis and R programming, SSMS (SQL Server Management Studio) is focused on managing SQL Server databases and writing queries, and Toad is a database development and administration tool. While they can work with text, they're not designed for quick, lightweight edits to configuration files, so they're less efficient for this particular need. So, for modifying a simple data connection property in a configuration file, Notepad++ is the best fit.

10. In data warehousing, what term refers to a dataset created specifically for drill-down reports?

- A. Dimension
- B. Measure
- C. Cube**
- D. Cluster

A dataset built for drill-down reports is a data cube. In OLAP, a cube organizes data across multiple dimensions (such as time, product, and region) and holds numeric measures (like sales or units sold) at their intersections. This structure lets you explore data by drilling down into finer levels, for example from yearly totals to quarterly, monthly, or by specific regions and products, with fast, aggregated calculations pre-computed for quick analysis. A dimension is an attribute used to slice data, such as product or region. A measure is the numeric value you analyze, like revenue. A cluster refers to a grouping used in clustering, not a storage structure for drill-down reporting. The cube combines dimensions and measures to support multi-dimensional, drill-down analysis, making it the best fit.

Next Steps

Congratulations on reaching the final section of this guide. You've taken a meaningful step toward passing your certification exam and advancing your career.

As you continue preparing, remember that consistent practice, review, and self-reflection are key to success. Make time to revisit difficult topics, simulate exam conditions, and track your progress along the way.

If you need help, have suggestions, or want to share feedback, we'd love to hear from you. Reach out to our team at hello@examzify.com.

Or visit your dedicated course page for more study tools and resources:

<https://dao002.examzify.com>

We wish you the very best on your exam journey. You've got this!

SAMPLE