DDR DATA SCIENCE INTERVIEW PRACTICE TEST (SAMPLE)

STUDY GUIDE

SAMPLE STUDY GUIDE WITH LIMITED QUESTIONS

VISIT US ONLINE FOR THE FULL VERSION STUDY GUIDE

https://ddrdatascienceinterview.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

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.

SAMPLE

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!

SAMPLE

Questions

SAMPLE

1. **What are best practices for ensuring reproducibility and versioning in a data science project?**

   A. Rely only on notebooks; no versioning.

   B. Version control for code; data versioning; environment management; dependency pinning; logging; and documentation.

   C. Hard code parameters in scripts.

   D. Share code but not environment.

2. **Which statement about a unique key is true?**

   A. A unique key ensures all values in a column are distinct and allows one NULL value.

   B. A unique key cannot have NULL values.

   C. A unique key identifies a row uniquely.

   D. A unique key functions identically to a primary key.

3. **What does TCL manage?**

   A. Defining database schemas.

   B. Querying data with SELECT statements.

   C. Managing user permissions.

   D. Controlling transactions, such as COMMIT and ROLLBACK.

4. **Which statement best describes the difference between UNION and UNION ALL?**

   A. UNION preserves duplicates, making it slower.

   B. UNION removes duplicates and UNION ALL keeps duplicates, which is faster.

   C. UNION sorts the result set, while UNION ALL does not.

   D. UNION can only be used with numeric columns.

5. **Which statement correctly contrasts embedding and referencing in MongoDB?**

   A. Embedding stores related data as a nested document; Referencing stores related data in a separate collection

   B. Embedding stores related data in a separate collection

   C. Referencing stores related data as a nested document

   D. Referencing stores related data in the same document

SAMPLE

6. Which NoSQL approach is commonly used to scale across many machines?

A. Vertical scaling by upgrading a single server

B. Scaling by adding more servers

C. Using only in-memory caches

D. Sharding data across multiple servers

7. In MongoDB terminology, which concept maps to a SQL table?

A. Database

B. Collection

C. Document

D. Index

8. What is the purpose of SQL constraints?

A. They enforce data accuracy and integrity

B. To speed up queries

C. To store computed data

D. To encrypt data

9. In k-fold cross-validation for classification, how many models are trained and evaluated, and why is stratification important?

A. k models trained and evaluated; stratification preserves class distribution in folds, improving estimate for imbalanced data.

B. k models trained but not evaluated.

C. Stratification speeds up computation but harms accuracy.

D. Cross-validation uses random folds without preserving class distribution.

10. What is a View in SQL?

A. A view stores data physically

B. A view is a virtual table defined by a stored SELECT query that does not store data but re-runs the query each time it is accessed

C. A view can always be updated directly

D. A view cannot be queried

SAMPLE

Answers

SAMPLE

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

SAMPLE

Explanations

SAMPLE

1. What are best practices for ensuring reproducibility and versioning in a data science project?

A. Rely only on notebooks; no versioning.

B. Version control for code; data versioning; environment management; dependency pinning; logging; and documentation.

C. Hard code parameters in scripts.

D. Share code but not environment.

Reproducibility in data science hinges on capturing code, data, and the runtime environment, and making them versionable and well documented. Version control for code keeps a traceable history of changes, so you can understand how results evolved and revert to a previous state if needed. Data versioning ensures the exact dataset used (including preprocessing steps and transformations) can be retrieved later, preventing shifts in results when data changes. Environment management guarantees the same runtime setup—Python or R version, OS specifics, and a consistent set of libraries—so runs aren't affected by subtle system differences. Dependency pinning locks specific library versions, protecting against drift that can alter behavior or results across runs. Logging captures the details of each experiment—parameters, seeds, data splits, and metrics—so you can reproduce a particular run precisely and compare results across attempts. Documentation ties everything together, explaining how to reproduce steps, what each parameter means, and how data flows through the pipeline. Together, these practices create a stable, auditable trail from code to results, which is essential for reliable replication, collaboration, and long-term maintenance.

2. Which statement about a unique key is true?

A. A unique key ensures all values in a column are distinct and allows one NULL value.

B. A unique key cannot have NULL values.

C. A unique key identifies a row uniquely.

D. A unique key functions identically to a primary key.

A unique key enforces that the values in its column (or set of columns) are distinct across all rows. This means a given non-NULL value can appear at most once, so you can identify a row by its unique value. It also allows NULL values; in practice, many databases permit multiple NULLs because NULL represents unknown and NULLs are not considered equal to each other. So the core idea is that a unique key guarantees uniqueness of values and permits NULLs, unlike a primary key which cannot be NULL.

3. What does TCL manage?

A. Defining database schemas.

B. Querying data with SELECT statements.

C. Managing user permissions.

D. Controlling transactions, such as COMMIT and ROLLBACK.

Transaction control language focuses on managing groups of database operations as a single unit. It lets you start a transaction, perform multiple changes, and then decide whether to apply all of them or undo them. The main commands COMMIT and ROLLBACK are used to finalize changes or revert them if something goes wrong. This control is what preserves atomicity and consistency across operations, ensuring that either everything in the transaction takes effect or none of it does. Some systems also support SAVEPOINT to mark a point to which you can roll back within a transaction. In contrast, defining database schemas, querying data, or setting permissions belong to other SQL sublanguages, while TCL specifically handles transaction flow.

SAMPLE

4. Which statement best describes the difference between UNION and UNION ALL?

A. UNION preserves duplicates, making it slower.

B. UNION removes duplicates and UNION ALL keeps duplicates, which is faster.

C. UNION sorts the result set, while UNION ALL does not.

D. UNION can only be used with numeric columns.

When combining results from multiple queries, the key idea is how duplicates are treated. UNION returns only distinct rows, removing any duplicates that appear across the combined results. UNION ALL keeps every row from both sides, including duplicates, so the same row can appear more than once if it comes from both queries. This difference in handling duplicates is what drives the typical performance gap: removing duplicates requires extra work to identify and drop duplicates, so UNION usually has more overhead and can be slower than UNION ALL. Use UNION when you need a unique set of rows; use UNION ALL when you want to preserve duplicates or when you want the faster, simpler operation and duplicates don't matter for your result. For example, if one query returns (1, 'x') and the other also returns (1, 'x'), UNION would yield a single (1, 'x'), while UNION ALL would yield two copies of (1, 'x').

5. Which statement correctly contrasts embedding and referencing in MongoDB?

A. Embedding stores related data as a nested document; Referencing stores related data in a separate collection

B. Embedding stores related data in a separate collection

C. Referencing stores related data as a nested document

D. Referencing stores related data in the same document

Embedding and referencing describe how related data is modeled in MongoDB. The correct statement is that embedding stores related data as a nested document, while referencing stores related data in a separate collection. Embedding puts the related data inside the parent document, which makes reads fast and atomic for the whole object and is great for tightly linked, one-to-few relationships. But it can blow up document size and create duplication if the related data is large or needs to be updated independently. Referencing keeps related data in separate documents across collections and uses identifiers to link them, which avoids duplication and supports more flexible, many-to-many relationships, but reads require extra queries or a lookup stage to assemble the full picture. The other statements misstate where the data lives: embedding is not in a separate collection, and referencing is not a nested document or stored in the same document.

6. Which NoSQL approach is commonly used to scale across many machines?

A. Vertical scaling by upgrading a single server

B. Scaling by adding more servers

C. Using only in-memory caches

D. Sharding data across multiple servers

Scaling across many machines in a NoSQL setup relies on distributing data so no single machine becomes a bottleneck. This is done through sharding, where the dataset is partitioned and the shards live on different servers. A shard key decides where each piece of data belongs, allowing reads and writes to happen in parallel across the cluster. As demand grows, you can add more machines to handle the increased load, achieving horizontal scaling. This approach contrasts with vertical scaling, which only upgrades a single machine, and with relying solely on in-memory caches, which doesn't provide durable storage or true distributed data handling. The idea of simply "adding more servers" is vague without a concrete data-distribution mechanism like sharding that actually spreads the data and the load.

SAMPLE

7. In MongoDB terminology, which concept maps to a SQL table?

A. Database

B. Collection

C. Document

D. Index

In MongoDB, the concept that maps to a SQL table is a collection. A collection is where you store documents, and you query against that set of documents much like you would query a table in a relational database. Each document is a JSON-like object (BSON) with fields, and while SQL tables have fixed schemas, a collection doesn't require a rigid schema—documents can vary in structure, though you'll typically index and query on common fields. A database is the container that holds multiple collections, and an index is a data structure used to speed up lookups, not a container of rows. For example, a table of customers in SQL corresponds to a collection named customers containing documents like { _id: ..., name: ..., email: ... }.

8. What is the purpose of SQL constraints?

A. They enforce data accuracy and integrity

B. To speed up queries

C. To store computed data

D. To encrypt data

SQL constraints enforce data accuracy and integrity by defining rules that data must follow in a table. They ensure things like a value exists when required (NOT NULL), values are unique (UNIQUE or PRIMARY KEY), and relationships between tables stay consistent (FOREIGN KEY). They can also enforce acceptable value ranges (CHECK) and supply sensible defaults (DEFAULT). Because these rules are automatically checked on insert, update, or delete, the database stays reliable and free from inconsistent or invalid data even with concurrent access. Other options describe different goals: speeding up queries relies on indexing and optimization, not constraints; storing computed data uses computed columns or views; encrypting data is about security measures, not data validity rules.

9. In k-fold cross-validation for classification, how many models are trained and evaluated, and why is stratification important?

A. k models trained and evaluated; stratification preserves class distribution in folds, improving estimate for imbalanced data.

B. k models trained but not evaluated.

C. Stratification speeds up computation but harms accuracy.

D. Cross-validation uses random folds without preserving class distribution.

In classification, k-fold cross-validation trains k models and evaluates each on a held-out fold, giving you k performance estimates that you typically average for a final score. The key idea here is not just splitting the data, but ensuring the evaluation reflects how the model would perform on new data. Stratification is important because real-world datasets often have imbalanced classes. If you split the data randomly, some folds might have very few or no instances of the minority class, biasing the evaluation and giving an unreliable picture of the model's performance. By stratifying, each fold preserves the same class distribution as the full dataset, so every evaluation run tests the model on a representative mix of classes. This leads to more stable, realistic performance estimates, especially for metrics that care about the minority class.

SAMPLE

10. What is a View in SQL?

A. A view stores data physically

B. A view is a virtual table defined by a stored SELECT query that does not store data but re-runs the query each time it is accessed

C. A view can always be updated directly

D. A view cannot be queried

A view is a virtual table defined by a stored SELECT statement. It doesn't store data itself; instead, when you query the view, the database runs the underlying SELECT against the base tables and returns the results. This setup lets you present a simplified or restricted view of data, or combine data from multiple tables behind one name, without duplicating data storage. Because the view doesn't hold data, it automatically reflects any changes made to the underlying tables. You can query a view just like a regular table, and in some cases a view can be updatable, but that's not guaranteed for all views or database systems. Since it doesn't store data, it's not correct to say a view stores data physically; and it's also incorrect to say a view cannot be queried.

SAMPLE

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://ddrdatascienceinterview.examzify.com

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

SAMPLE