# MICROSOFT AZURE DATA ENGINEER CERTIFICATION (DP-203) PRACTICE TEST (SAMPLE)

## STUDY GUIDE



## SAMPLE STUDY GUIDE WITH LIMITED QUESTIONS

**VISIT US ONLINE FOR THE FULL VERSION STUDY GUIDE**

https://microsoftazuredataengineer.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 ..... 16


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. **Which of the following is a layer in data lake architecture?**

   A. Ingested data layer

   B. Raw data layer

   C. Delta data layer

   D. Historical data layer

2. **What is a key benefit of using materialized views?**

   A. Increased resiliency benefits.

   B. Reducing the execution time for complex queries.

   C. Improved data integrity during transactions.

   D. Enhanced data encryption capabilities.

3. **Which storage access tier would be most suitable for infrequently accessed data?**

   A. Hot

   B. Cool

   C. Archive

   D. Premium

4. **What is the function of Azure Event Hubs?**

   A. A platform for database management

   B. A big data streaming platform for event ingestion

   C. A service for storing historical data

   D. A data visualization tool

5. **What is Azure HDInsight primarily used for?**

   A. A fully-managed cloud service for processing big data using open-source frameworks

   B. A platform for managing virtual machines

   C. A service for implementing relational databases

   D. A tool for basic analytics reporting

SAMPLE

**6. How do column statistics improve query performance?**

A. By keeping track of which columns are being queried

B. By keeping track of how much data exists between ranges in columns

C. By caching column values for queries

D. By optimizing the execution plan

**7. Which T-SQL Statement loads data directly from Azure Storage?**

A. LOAD DATA

B. COPY

C. INSERT FROM FILE

**8. What function is used to parse JSON data in Transact-SQL?**

A. JSON_PARSE

B. JSON_READ

C. JSON_VALUE

D. JSON_EXTRACT

**9. What happens when you drop a non-managed table in Delta Lake?**

A. The table data is retained on disk.

B. Only the table schema is removed.

C. The table and its data are deleted.

D. The metadata is retained for future reference.

**10. The schema of what type of data can be defined at query time?**

A. Structured data

B. Azure Cosmos DB

C. Unstructured data

D. Time-series data

SAMPLE

# Answers

SAMPLE

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

SAMPLE

# Explanations

SAMPLE

## 1. Which of the following is a layer in data lake architecture?

A. Ingested data layer

B. Raw data layer

C. Delta data layer

D. Historical data layer

*The raw data layer is a fundamental component of data lake architecture. This layer is responsible for storing unprocessed or minimally processed data, which is crucial because it preserves the original state of the data as it was ingested into the data lake. This allows for flexibility in data processing and analysis later on, as different applications and users can access the raw data and apply their own transformations and processing as needed. In the context of a data lake, the raw data layer typically contains data in its native format, whether structured, semi-structured, or unstructured. This means that the data can come from various sources like databases, IoT devices, and file systems, enabling a diverse range of analytical capabilities. Future data processing efforts can leverage this raw data to derive insights, build models, or create transformed datasets for more specific use cases. Data lake architecture generally includes various additional layers, such as refined or curated data layers, but the importance of having a raw data layer lies in its role as the foundational source from which all data exploration and transformation originate.*

## 2. What is a key benefit of using materialized views?

A. Increased resiliency benefits.

B. Reducing the execution time for complex queries.

C. Improved data integrity during transactions.

D. Enhanced data encryption capabilities.

*Using materialized views significantly reduces the execution time for complex queries because they store pre-computed results derived from one or more base tables. When a materialized view is created, it captures the results of a query and stores that result set, which can then be accessed directly rather than recalculated with each query. This means that when users or applications need to perform queries against the data, they can retrieve information from the materialized view much faster than if they were querying the underlying tables directly, especially if those queries involve complex computations, aggregations, or joins. Materialized views are particularly useful in scenarios where data does not change frequently, allowing for efficient storage and quick access to frequently requested outputs without the overhead of recalculating them each time a query is executed. As a result, they are widely used in reporting and decision support systems where performance is critical. While the other choices may have their own benefits within broader contexts, they do not directly pertain to the core functionality and advantages of materialized views in terms of query performance.*

SAMPLE

## 3. Which storage access tier would be most suitable for infrequently accessed data?

A. Hot

B. Cool

C. Archive

D. Premium

*The most suitable storage access tier for infrequently accessed data is the Cool tier. This tier is designed specifically for storing large volumes of data that do not need to be accessed frequently but still require quick retrieval when needed. The Cool tier offers lower storage costs compared to the Hot tier, which is optimized for data that is accessed frequently. This makes it a cost-effective solution for data that is expected to be accessed less than once a month. In addition to cost efficiency, the Cool tier provides a balance of performance and cost, enabling organizations to keep their data affordable while still being readily accessible when necessary. For example, it is often used for backup and disaster recovery scenarios, where data may only be accessed occasionally. The Archive tier, while providing even lower storage costs, is designed for data that is rarely accessed and incurs higher retrieval costs and latency compared to the Cool tier. Therefore, it would not be as suitable for data that may require more timely access, even if infrequent. The Premium tier is tailored for high-performance scenarios where low latency and high throughput are critical, making it inappropriate for infrequent access. Thus, the Cool tier stands out as the most appropriate choice for infrequently accessed data, as it effectively balances cost and access needs.*

## 4. What is the function of Azure Event Hubs?

A. A platform for database management

B. A big data streaming platform for event ingestion

C. A service for storing historical data

D. A data visualization tool

*Azure Event Hubs serves as a big data streaming platform primarily designed for event ingestion. It allows users to easily collect, process, and analyze large volumes of events in real-time, making it ideal for scenarios that require handling and processing huge streams of data. For instance, Event Hubs can be used to ingest data from multiple sources such as IoT devices, web applications, and other data-producing entities. This service is designed to handle millions of events per second, enabling organizations to gain insights from the data as it flows through their systems. It supports real-time analytics and can work seamlessly with other Azure services like Azure Stream Analytics, which enhances its capability to perform streaming analytics on the ingested data. By using Azure Event Hubs, organizations ensure scalability and reliability in their data ingestion processes, which is critical for making timely decisions based on real-time data. Other options such as database management, historical data storage, and data visualization do not reflect the core capabilities of Azure Event Hubs, which is uniquely focused on event ingestion and real-time data processing.*

SAMPLE

## 5. What is Azure HDInsight primarily used for?

A. A fully-managed cloud service for processing big data using open-source frameworks

B. A platform for managing virtual machines

C. A service for implementing relational databases

D. A tool for basic analytics reporting

*Azure HDInsight is primarily used as a fully-managed cloud service for processing big data using open-source frameworks. This includes frameworks such as Apache Hadoop, Apache Spark, Apache Hive, and others. These technologies facilitate the storage, processing, and analysis of large datasets efficiently and at scale. HDInsight allows organizations to leverage popular big data analytics frameworks without the overhead of managing the underlying infrastructure, providing a cloud-native solution that simplifies deployment and management. The focus on processing big data enables companies to analyze large volumes of data in real-time, extract insights, and solve complex analytics problems effectively. This capability is essential for businesses that require scalable and flexible solutions to handle big data workloads. HDInsight's integration with other Azure services further enhances its functionality for tasks like machine learning, data warehousing, and business intelligence. In contrast, the other options describe services or solutions that do not align with the primary purpose of HDInsight. Managing virtual machines, implementing relational databases, and basic analytics reporting represent different aspects of cloud services that are better handled by other Azure offerings rather than HDInsight.*

## 6. How do column statistics improve query performance?

A. By keeping track of which columns are being queried

B. By keeping track of how much data exists between ranges in columns

C. By caching column values for queries

D. By optimizing the execution plan

*Column statistics significantly enhance query performance by providing essential information about the distribution of data within specific columns. By maintaining an overview of how much data exists within ranges of column values, the query optimizer can make more informed decisions when formulating execution plans. When the optimizer understands the data distribution, it can accurately estimate the number of rows that will be returned for a query. This capability allows it to choose the most efficient execution plan, like selecting appropriate indexes or deciding on join strategies. For instance, if a particular value in a column is highly skewed (where a small number of distinct values occur very frequently), the optimizer can avoid scanning large swaths of data unnecessarily, thus speeding up query performance. The other options do not capture the core function of column statistics effectively. Keeping track of which columns are being queried offers some insight but does not specifically relate to how data distribution impacts performance. Caching column values could theoretically enhance performance but is more about memory management than statistics. Optimizing the execution plan is the end result of insights gained from statistics rather than a function of statistics themselves. Thus, the focus on data ranges and their cardinality is what makes column statistics a vital tool for improving the processing speed of queries.*

SAMPLE

## 7. Which T-SQL Statement loads data directly from Azure Storage?

A. LOAD DATA

B. COPY

C. INSERT FROM FILE

*The correct answer is the COPY statement. This T-SQL command is specifically designed to facilitate the transfer of data from external storage sources, such as Azure Blob Storage, directly into a database. The COPY statement handles various data formats, allows for efficient bulk loading of large datasets, and can provide options for error handling and data transformation during the load process. In contrast, the other options do not directly refer to T-SQL commands available in Azure SQL Database or Azure Synapse Analytics for loading data from external storage. LOAD DATA is not a recognized T-SQL command in the Azure context, and INSERT FROM FILE is generally not a valid syntax in T-SQL for loading external data. Thus, COPY is clearly the correct choice for loading data from Azure Storage seamlessly and efficiently.*

## 8. What function is used to parse JSON data in Transact-SQL?

A. JSON_PARSE

B. JSON_READ

C. JSON_VALUE

D. JSON_EXTRACT

*The function used to parse JSON data in Transact-SQL is JSON_VALUE. This function extracts a scalar value from a JSON string. It specifically allows users to retrieve a single value from a JSON object by specifying the path to that value. This path can include key names and, if applicable, array indexes. JSON_VALUE is particularly useful for querying JSON data stored in SQL Server and Azure SQL Database, enabling seamless integration of JSON documents with SQL queries. By using this function, data engineers can efficiently handle and manipulate JSON data without the need for complex parsing logic. The other provided options are not valid functions in Transact-SQL for working with JSON data, which differentiates JSON_VALUE as the correct choice. JSON_READ and JSON_PARSE, for example, do not exist in the SQL Server's set of JSON functions, and while JSON_EXTRACT might be familiar from other SQL dialects, it's not applicable in this context. Therefore, JSON_VALUE stands out as the functional choice for parsing JSON data within Transact-SQL.*

## 9. What happens when you drop a non-managed table in Delta Lake?

A. The table data is retained on disk.

B. Only the table schema is removed.

C. The table and its data are deleted.

D. The metadata is retained for future reference.

*When a non-managed table is dropped in Delta Lake, the correct outcome is that the table data is retained on disk. This distinction is crucial; non-managed tables in Delta Lake are created in a way that separates the data storage lifecycle from the table management lifecycle. When you create a non-managed table, you are essentially telling Delta Lake that you only want it to manage the schema of the table, but not the underlying data. Therefore, when you drop the non-managed table, the operation only removes the metadata associated with the table, allowing the data to still exist on disk. This can be particularly useful when you want to retain the data for other purposes, such as creating new tables or performing different operations, without needing to load the data again. This explanation highlights the unique characteristics of non-managed tables within Delta Lake, clarifying why the table data persists even after the metadata has been cleared.*

SAMPLE

## 10. The schema of what type of data can be defined at query time?

A. Structured data

B. Azure Cosmos DB

C. Unstructured data

D. Time-series data

*The schema of unstructured data can be defined at query time because unstructured data lacks a predefined data model, allowing for a flexible approach to organization and analysis. This means that users can specify how they want to interpret or analyze the data when the query is executed, adapting to various needs without the constraints of a fixed schema. In contrast, structured data adheres to a strict schema defined beforehand, making it less flexible in terms of how queries are formulated. Azure Cosmos DB is a multi-model database service that can support both structured and semi-structured data but does not inherently allow for schema definition at query time in the same manner as unstructured data. Time-series data typically has a defined structure over time to accurately represent and analyze temporal changes, making it similar to structured data in its schema rigidity. Thus, the nature of unstructured data allows for this dynamic schema definition at the time of query, making it the correct choice.*

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://microsoftazuredataengineer.examzify.com

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

SAMPLE