

Microsoft Azure Data Engineer Certification (DP-203) Practice Test (Sample)

Study Guide



Everything you need from our exam experts!

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

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

- 1. What function is used to parse JSON data in Transact-SQL?**
 - A. JSON_PARSE**
 - B. JSON_READ**
 - C. JSON_VALUE**
 - D. JSON_EXTRACT**

- 2. What is the primary purpose of Azure Functions in data integration scenarios?**
 - A. To provide a user interface for data management**
 - B. To enable event-driven data processing and automate data workflows**
 - C. To store unstructured data efficiently**
 - D. To serve as a cloud database solution**

- 3. Which Azure service is responsible for detecting anomalies in account activity?**
 - A. Azure Defender for Storage**
 - B. Azure Storage Account Security Feature**
 - C. Azure Key Vault**
 - D. Azure Monitor**

- 4. What must you create to query an existing analytical store in Azure Cosmos DB using a Spark pool in Azure Synapse Analytics?**
 - A. Create a linked service for the Azure Cosmos DB database**
 - B. Configure a new security role for the Spark pool**
 - C. Disable automatic pausing for the Spark pool**
 - D. Set up a virtual network for the Spark pool**

- 5. Azure Functions is particularly useful for which type of computing model?**
 - A. Infrastructure as a Service (IaaS)**
 - B. Event-driven serverless computing**
 - C. Platform as a Service (PaaS)**
 - D. Cloud storage management**

- 6. What is the best tool for a data analyst to use when analyzing data with Python code?**
- A. A notebook connected to an Apache Spark pool**
 - B. A SQL script connected to a serverless SQL pool**
 - C. A KQL script connected to a Data Explorer pool**
 - D. A Power BI report**
- 7. What should be considered when selecting a distribution method for optimizing load times in Azure Data services?**
- A. Database size alone**
 - B. Data alignment between compute and storage nodes**
 - C. User demand on resources**
- 8. To enable analytical queries over Azure Synapse Link from Azure Synapse Analytics, what is required for a container in Cosmos DB?**
- A. Delete and recreate the container**
 - B. Create an indexing policy for the container**
 - C. Enable Azure Synapse Link in the container to create an analytical store**
 - D. Increase the storage capacity of the container**
- 9. In what language can the Azure Synapse Apache Spark to Synapse SQL connector be used?**
- A. Python**
 - B. SQL**
 - C. Scala**
 - D. Java**
- 10. What is the primary function of Sink transformation in Mapping Data Flow?**
- A. Extract data from sources**
 - B. Transform data**
 - C. Load data into a data store**
 - D. Validate data**

Answers

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

SAMPLE

Explanations

SAMPLE

1. 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.

2. What is the primary purpose of Azure Functions in data integration scenarios?

- A. To provide a user interface for data management
- B. To enable event-driven data processing and automate data workflows**
- C. To store unstructured data efficiently
- D. To serve as a cloud database solution

The primary purpose of Azure Functions in data integration scenarios is to enable event-driven data processing and automate data workflows. Azure Functions is a serverless compute service that allows you to execute code in response to various events, such as data changes in databases, messages from queues, or HTTP requests. This ability to respond to events in real time makes Azure Functions particularly valuable for building data pipelines and processing data as it flows from one source to another. Using Azure Functions, developers can create lightweight functions that can react to data triggering events, allowing for efficient and automated workflows. For example, when new data is ingested into a data lake, an Azure Function can be triggered to process this data, perform transformations, or move it to another system as needed. This flexibility allows organizations to streamline their data integration processes and facilitate more dynamic and real-time data management. The other choices do not align with the core capabilities of Azure Functions in this context. For instance, Azure Functions do not provide a user interface for data management, store unstructured data, or serve as a cloud database solution. Thus, the focus on event-driven processing and automation highlights why the correct answer is specifically B.

3. Which Azure service is responsible for detecting anomalies in account activity?

- A. Azure Defender for Storage**
- B. Azure Storage Account Security Feature**
- C. Azure Key Vault**
- D. Azure Monitor**

The correct choice is Azure Defender for Storage, which is a comprehensive security solution that incorporates features to protect and monitor the security status of your storage accounts. One of its key capabilities is anomaly detection in account activity. This ensures that any unusual patterns of access, such as unexpected spikes in requests or access from unfamiliar IP addresses, are identified in real time, allowing administrators to respond quickly to potential security threats. In contrast, Azure Storage Account Security Feature primarily focuses on configuring network and security settings for the storage account but does not specifically provide anomaly detection. Azure Key Vault is dedicated to managing secrets and protecting cryptographic keys, whereas Azure Monitor is a broader monitoring service that encompasses various Azure resources but does not specialize in detecting anomalies specifically for storage accounts. Thus, Azure Defender for Storage stands out as the service designed specifically for the purpose of detecting anomalies in account activity within Azure storage services.

4. What must you create to query an existing analytical store in Azure Cosmos DB using a Spark pool in Azure Synapse Analytics?

- A. Create a linked service for the Azure Cosmos DB database**
- B. Configure a new security role for the Spark pool**
- C. Disable automatic pausing for the Spark pool**
- D. Set up a virtual network for the Spark pool**

To query an existing analytical store in Azure Cosmos DB using a Spark pool in Azure Synapse Analytics, creating a linked service for the Azure Cosmos DB database is essential. A linked service acts as a connection string that provides the necessary configuration information for Synapse to access external data sources. In this case, it facilitates communication between the Spark pool and the Azure Cosmos DB analytical store, allowing you to execute queries and process data effectively. Establishing a linked service includes specifying details such as the type of data store, connection strings, authentication methods, and other options necessary for accessing the data within the analytical store. This step is fundamental before any actual querying can take place, as it enables Spark to understand how to connect to the Cosmos DB and retrieve the desired datasets. In contrast, configuring a new security role, disabling automatic pausing, or setting up a virtual network do not address the core requirement of establishing a connection between Azure Synapse Analytics and Azure Cosmos DB for querying purposes. Each of these options may have their own relevance in different contexts or scenarios, but they do not provide the direct linkage necessary for querying an analytical store in this specific situation.

5. Azure Functions is particularly useful for which type of computing model?

- A. Infrastructure as a Service (IaaS)**
- B. Event-driven serverless computing**
- C. Platform as a Service (PaaS)**
- D. Cloud storage management**

Azure Functions is particularly useful for event-driven serverless computing due to its ability to automatically execute code in response to specific events without the need for managing the underlying infrastructure. This serverless architecture enables developers to build applications that can respond to triggers from various sources, such as HTTP requests, changes in data in Azure Blob Storage, messages from queues, and events from other Azure services. The core advantage of using Azure Functions lies in its scalability and flexibility. When an event occurs, Azure Functions can scale up automatically to handle the increased load and then scale down when the demand subsides, ensuring efficient resource utilization and cost management. This makes it ideal for scenarios where workload patterns are unpredictable and where only minimal resources should be consumed without needing a dedicated server. In contrast, while IaaS provides virtual machines and complete control over the operating system, it requires more management and manual configuration. PaaS offers a platform for application development but may involve more overhead compared to serverless functions. Cloud storage management focuses specifically on storing and managing data rather than executing code in response to events. Thus, the event-driven serverless computing model highlights the unique benefits of Azure Functions effectively.

6. What is the best tool for a data analyst to use when analyzing data with Python code?

- A. A notebook connected to an Apache Spark pool**
- B. A SQL script connected to a serverless SQL pool**
- C. A KQL script connected to a Data Explorer pool**
- D. A Power BI report**

The best tool for a data analyst to use when analyzing data with Python code is a notebook connected to an Apache Spark pool. This approach is optimal because notebooks, such as those provided by Azure Databricks, are designed specifically for interactive data analysis and exploration, allowing for a flexible environment where analysts can write, run, and test their Python code in an iterative manner. Using a notebook with Apache Spark facilitates the processing of large datasets efficiently, leveraging Spark's distributed computing capabilities. This setup supports various libraries and frameworks commonly used in data analysis and machine learning, such as Pandas, NumPy, and scikit-learn, enhancing the analytical capabilities of the data analyst. Moreover, notebooks provide an intuitive interface for visualizing results immediately after executing Python code. This allows for a seamless workflow where data can be transformed, analyzed, and visualized within the same environment, ultimately resulting in more efficient analysis and actionable insights. The other tools listed, while useful in certain contexts, are not primarily designed for the direct execution of Python code in an exploratory analysis setting.

7. What should be considered when selecting a distribution method for optimizing load times in Azure Data services?

A. Database size alone

B. Data alignment between compute and storage nodes

C. User demand on resources

When optimizing load times in Azure Data services, considering the data alignment between compute and storage nodes is crucial. This approach guarantees that the data processing occurs close to where the data is stored, drastically reducing latency. If the compute resources are not in the same location or are misaligned with the storage, this can lead to increased loading times due to the need to transfer data over the network more frequently and potentially across regions. This alignment is particularly important in distributed environments where data may be sharded or partitioned across different nodes. By ensuring that the computing resources can access the relevant data directly and efficiently, you streamline the operations and improve performance, which is a central goal in data engineering tasks. In contrast, looking at database size alone does not take into account how the data is distributed or accessed, potentially leading to inefficient resource utilization. Similarly, while user demand on resources can influence performance, it does not inherently address the optimization of load times in the context of data distribution and architecture alignment. Optimizing for load times requires a holistic approach that prioritizes the physical architecture of data storage and processing resources.

8. To enable analytical queries over Azure Synapse Link from Azure Synapse Analytics, what is required for a container in Cosmos DB?

A. Delete and recreate the container

B. Create an indexing policy for the container

C. Enable Azure Synapse Link in the container to create an analytical store

D. Increase the storage capacity of the container

To enable analytical queries over Azure Synapse Link from Azure Synapse Analytics, it is essential to enable Azure Synapse Link in the container to create an analytical store. This capability allows you to efficiently query your Cosmos DB data using Synapse Analytics without impacting the transactional performance of your workload. When Azure Synapse Link is enabled, it creates a separate analytical store that is specifically optimized for analytical queries. This store is eventually consistent and can be queried using a SQL-based model, making it ideal for large-scale analytics and reporting tasks. The analytical store supports running complex queries, aggregations, and joins, which are typically required in data analysis and business intelligence scenarios. The other options do not directly relate to the requirement for enabling analytical capabilities through Azure Synapse Link. For instance, simply deleting and recreating the container or changing the indexing policy does not fulfill the necessary requirement of enabling the link for analytical querying. Similarly, increasing the storage capacity of the container does not impact the analytical store feature provided by Azure Synapse Link.

9. In what language can the Azure Synapse Apache Spark to Synapse SQL connector be used?

- A. Python**
- B. SQL**
- C. Scala**
- D. Java**

The Azure Synapse Apache Spark to Synapse SQL connector can be used with Scala. This integration allows for seamless interaction between Spark applications and Synapse SQL data, enabling complex data transformations and analytics. Scala is a highly functional language often used with Apache Spark due to its ability to handle distributed data processing efficiently and its close integration with the underlying Spark framework. When using Scala, you have access to Spark's rich API, which streamlines the process of reading from and writing to Synapse SQL. This makes it an ideal choice for data engineers looking to leverage Spark for big data processing while interacting with structured data in Synapse. Python and Java are also commonly used in Spark applications, providing alternative ways to build applications and interact with Synapse SQL. However, the specific focus on Scala reflects its tight integration and performance optimizations within the Spark ecosystem. SQL is not directly applicable as it is primarily a query language and not a programming language used for implementing connectors or APIs in this context.

10. What is the primary function of Sink transformation in Mapping Data Flow?

- A. Extract data from sources**
- B. Transform data**
- C. Load data into a data store**
- D. Validate data**

The primary function of Sink transformation in Mapping Data Flow is to load data into a data store. In data processing pipelines, the sink acts as the final destination where the transformed data is written after being processed. The sink transformation is integral to ensuring that the data flows from various sources, through transformation processes, and ultimately reaches a specified data storage solution—such as Azure Blob Storage, Azure SQL Database, or any other supported data store. By utilizing the sink transformation, users can define how they want to output the data, including options for batching, schema mapping, and writing modes (like upsert or overwrite). This step is crucial in a data pipeline as it facilitates the persistence of processed data so it can be accessed for further analysis or reporting. Options that refer to extracting data from sources or transforming data focus on earlier stages of a data pipeline, while validating data suggests a process of checking the quality or integrity of data, which can happen prior to loading. However, the sink transformation specifically deals with the final loading aspect.

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!