

Fabric Analytics Engineer Associate 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	9
Explanations	11
Next Steps	17

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. To display the value breakdown between starting balance and ending balance across customer segments, which visual should you use?**
 - A. ribbon chart**
 - B. stacked bar chart**
 - C. treemap**
 - D. waterfall chart**

- 2. For near-real-time reporting with minimal development effort on a lakehouse with heavy historical data and frequent incremental updates, which storage mode is most suitable?**
 - A. Direct Lake**
 - B. DirectQuery**
 - C. Import**
 - D. Import and Direct Lake combined**

- 3. In Dataflows Query Editor connected to an Azure SQL customer table, which option helps identify which column contains non-duplicate values per customer?**
 - A. Column Distribution - distinct values**
 - B. Column distribution - unique values**
 - C. Column profile - values count**
 - D. Column quality - valid values**

- 4. What determines the data model size limits on a dataset when large dataset storage format is enabled?**
 - A. Power BI Premium 10 GB size limit.**
 - B. Power BI Premium capacity size or the maximum size set by the administrator.**
 - C. There are no size limits if large dataset storage format is enabled.**
 - D. The size is determined by the number of users accessing the dataset.**

- 5. If you drop a managed Delta table whose data resides in a folder, what happens to the table and data?**
- A. The table definition is removed from the metastore, but the data files remain intact.**
 - B. The table metadata and data files are deleted.**
 - C. The data files are deleted, leaving the table metadata intact.**
 - D. The table metadata and data files are deleted.**
- 6. What data storage solution located in a Fabric workspace supports access via T-SQL or Python and can store unstructured or semi-structured data?**
- A. Data lake**
 - B. Data warehouse**
 - C. Datamart**
 - D. Lakehouse**
- 7. Which statement best describes Direct Lake mode?**
- A. It loads all data into memory**
 - B. It queries data directly from the lake without ingesting**
 - C. It requires staging tables**
 - D. It caches results in memory**
- 8. If you want a Power BI semantic model to be reusable by multiple reports without an enterprise data warehouse, which component should you create?**
- A. Linked table**
 - B. Shared dataset**
 - C. Staging dataflow**
 - D. Transformation dataflow**
- 9. Which method should be used for a one-time copy of a small local file into the lakehouse when you want to minimize setup?**
- A. Dataflow Gen2 query**
 - B. Spark notebook**
 - C. Local file upload by using Lakehouse explorer**
 - D. Pipeline with the Copy data activity**

10. Which Fabric data store should be used to store data on OneLake and support CREATE TABLE statements?

- A. SQL Endpoint**
- B. Lakehouse**
- C. KQL database**
- D. Warehouse**

SAMPLE

Answers

SAMPLE

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

SAMPLE

Explanations

SAMPLE

1. To display the value breakdown between starting balance and ending balance across customer segments, which visual should you use?

- A. ribbon chart**
- B. stacked bar chart**
- C. treemap**
- D. waterfall chart**

You're aiming to show how the total value in each customer segment breaks down into starting balance and ending balance, and you want to compare both across multiple segments at a glance. A treemap fits this well because it shows parts of a whole as rectangles, with size representing magnitude. By organizing the data so each segment is a block and then subdividing that block into starting balance and ending balance tiles, you can immediately see how much each balance contributes within a segment and compare those contributions across all segments. This makes the relative proportions easy to scan, even with many segments. Ribbon charts focus on ranking and changes over time, not a fixed part-to-whole breakdown. A stacked bar chart can show two values per category but often emphasizes the total of the bar and can be harder to compare proportions across many categories at once. A waterfall chart illustrates how a starting value evolves through sequential changes to reach an ending value, which isn't the intended cross-segment breakdown here.

2. For near-real-time reporting with minimal development effort on a lakehouse with heavy historical data and frequent incremental updates, which storage mode is most suitable?

- A. Direct Lake**
- B. DirectQuery**
- C. Import**
- D. Import and Direct Lake combined**

Direct Lake mode is the best fit here because it lets you run queries directly against the data stored in the data lake, keeping data in place rather than duplicating it in a separate storage layer. This means: - You see the latest updates without waiting for a data import or refresh, which is ideal for near-real-time reporting. - It scales well with large historical datasets since you're not moving a massive amount of data into an analytics engine; the data remains in the lake and is accessed as needed. - Development effort is minimal because there's no need to build and maintain import pipelines or incremental refresh logic; the lake's files and metadata drive the query. In contrast, importing data would involve copying a huge historical dataset into the analytics layer and then maintaining it with incremental refreshes to stay current, adding latency and maintenance. A Direct Query-like approach would query the source live, which can introduce latency and depend on the source's performance and availability. Combining Import with Direct Lake adds complexity without offering a clear advantage for the given scenario.

3. In Dataflows Query Editor connected to an Azure SQL customer table, which option helps identify which column contains non-duplicate values per customer?

- A. Column Distribution - distinct values**
- B. Column distribution - unique values**
- C. Column profile - values count**
- D. Column quality - valid values**

Profiling the data to see how many different values appear in each column quickly reveals duplication patterns. The column distribution feature that shows distinct values tells you how many unique values exist for a column across all rows. If a column has as many distinct values as there are rows, every row has a unique value in that column, meaning there are no duplicates for that column across customers. That direct relationship makes it the best way to identify columns with non-duplicate values. The other options don't provide the same direct signal. Column distribution that highlights unique values (values that occur only once) is related but focuses on values that appear a single time, which is not as straightforward for assessing overall non-duplication per column. Column profile with values count shows how many values exist but not how they're distributed or whether duplicates occur. Column quality focusing on valid values checks data validity, not duplication patterns.

4. What determines the data model size limits on a dataset when large dataset storage format is enabled?

- A. Power BI Premium 10 GB size limit.**
- B. Power BI Premium capacity size or the maximum size set by the administrator.**
- C. There are no size limits if large dataset storage format is enabled.**
- D. The size is determined by the number of users accessing the dataset.**

The key idea is that large dataset storage format expands how big a model can be, but the ceiling is still defined by the resources available in your Premium capacity and any administrator-imposed limits. When you enable this format, the dataset benefits from more efficient storage and can grow larger, but you're not free to exceed what the Premium capacity can support. If the capacity has more headroom, larger datasets are possible; if an admin has set a maximum dataset size for a workspace or policy, that cap applies as well. The number of users accessing the dataset doesn't directly set its maximum size, and there are indeed practical limits beyond zero, so the size isn't unlimited or fixed at an old 10 GB rule.

5. If you drop a managed Delta table whose data resides in a folder, what happens to the table and data?
- A. The table definition is removed from the metastore, but the data files remain intact.
 - B. The table metadata and data files are deleted.
 - C. The data files are deleted, leaving the table metadata intact.
 - D. The table metadata and data files are deleted.**

Dropping a managed Delta table removes both the table's metadata and its data. In a managed table, the system controls the table's storage location and the metastore entry. When you drop the table, the metastore entry is deleted and the entire directory that holds the table data (including the Delta transaction log and all data files) is removed. This is different from external tables, where dropping only removes the metadata and leaves the underlying data intact. So both the metadata and the data files disappear.

6. What data storage solution located in a Fabric workspace supports access via T-SQL or Python and can store unstructured or semi-structured data?
- A. Data lake
 - B. Data warehouse
 - C. Datamart
 - D. Lakehouse**

Access methods like T-SQL and Python alongside the ability to store unstructured or semi-structured data point to a lakehouse. A lakehouse combines the flexible storage of a data lake with the governance, performance, and schema features of a data warehouse, so you can keep raw or semi-structured data in one place and query it with SQL (T-SQL) or analyze it from Python. In Fabric, this unified storage model lets analysts run familiar SQL queries and data scientists work directly in Python on the same data, covering both workloads efficiently. By contrast, a data lake focuses on raw files without centralized SQL querying, a data warehouse centers on highly structured data optimized for SQL, and a datamart is a smaller, subject-area subset of a warehouse—none of these inherently pair unstructured data storage with both SQL and Python access in a single, unified way.

7. Which statement best describes Direct Lake mode?
- A. It loads all data into memory
 - B. It queries data directly from the lake without ingesting**
 - C. It requires staging tables
 - D. It caches results in memory

Direct Lake mode means you query data where it lives—in the data lake—without importing or ingesting it into the analytics model. The engine reads only the necessary files and partitions from the lake (like Parquet or Delta formats), using lake metadata to optimize the query. You don't create staging tables, and you don't load the entire dataset into memory. While some caching of metadata or results may occur for speed, the core idea is accessing data directly from the lake rather than pulling it into the service.

8. If you want a Power BI semantic model to be reusable by multiple reports without an enterprise data warehouse, which component should you create?

A. Linked table

B. Shared dataset

C. Staging dataflow

D. Transformation dataflow

Centralizing the semantic model as a shared dataset in the Power BI service lets multiple reports connect to the same data model. This means the definitions of tables, relationships, measures, and security rules live once and are reused everywhere, ensuring consistency and reducing duplication. You publish the dataset to a workspace and then reports can connect to it (either live or by import) and refresh together, without needing an enterprise data warehouse. The other options don't provide this single, reusable semantic layer across several reports: a linked table stays within an individual report's model, and staging or transformation dataflows are ETL steps that prepare data but don't by themselves serve as the centralized model that multiple reports consume.

9. Which method should be used for a one-time copy of a small local file into the lakehouse when you want to minimize setup?

A. Dataflow Gen2 query

B. Spark notebook

C. Local file upload by using Lakehouse explorer

D. Pipeline with the Copy data activity

For a one-time, small local file, the fastest way to get it into the lakehouse with minimal setup is to upload it directly using Lakehouse explorer. This built-in UI lets you drag in or browse to your local file and place it into the lakehouse storage (often into a specific folder or dataset) without creating pipelines, notebooks, or data flows. Once uploaded, you can query or use the data right away, which is ideal for quick ad hoc loads. The other options require more setup. A Dataflow Gen2 query involves creating a dataflow with connections and processing steps. A Spark notebook needs you to write and run code to read the file and write it to the lakehouse. A pipeline with the Copy data activity requires building a pipeline, defining sources and sinks, and configuring executions. For a simple, one-off upload, Lakehouse explorer keeps things straightforward and fast.

10. Which Fabric data store should be used to store data on OneLake and support CREATE TABLE statements?

- A. SQL Endpoint**
- B. Lakehouse**
- C. KQL database**
- D. Warehouse**

When you need a data store that sits on top of OneLake and supports standard SQL DDL like CREATE TABLE, you use a Warehouse. It provides a relational, schema-defined store whose data is persisted in OneLake, and it accepts SQL commands to define and manage tables, making it suitable for structured, transactional workloads. The SQL Endpoint is just a compute surface for running SQL against existing data, not a dedicated storage layer. The Lakehouse offers lake-like storage with schema, but for explicit relational CREATE TABLE semantics on OneLake, the Warehouse is the intended relational storage. KQL databases use a different query language and do not align with SQL CREATE TABLE in this context.

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://fabricanalyticsengineerassoc.examzify.com>

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

SAMPLE