DP-600 Fabric Analytics Engineer Practice Exam (Sample)

Study Guide

Everything you need from our exam experts!

Table of Contents

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. **Which SCD type keeps the latest data, with old data overwritten?**

   A. Type 3 SCD

   B. Type 0 SCD

   C. Type 2 SCD

   D. Type 1 SCD

2. **To enforce access control for embedded reports and prevent public sharing, which tenant setting should you disable?**

   A. Disable External Content

   B. Disable Analyze in Excel

   C. Disable XMLA endpoints

   D. Disable Publish to web

3. **To schedule a pipeline to run every four hours on Mondays and Fridays, which Repeat setting should you use?**

   A. Daily

   B. Hourly

   C. Weekly

   D. Monthly

4. **In a Fabric workspace containing a Power BI report, when access to a data source is restricted to narrow time windows, what is the recommended approach to bring data into a Power BI semantic model using dataflows?**

   A. Create a staging dataflow that will only copy the data from the source as-is

   B. Use a direct query dataflow

   C. Use a dataflow with incremental load

   D. Use a dataflow that pre-aggregates data

5. **If warehouse queries degrade in performance due to throttling, which app should you use to identify throttling?**

   A. Azure Monitor

   B. Power BI Performance Analyzer

   C. SQL Server Profiler

   D. Microsoft Fabric Capacity Metrics app

6. **You have a Fabric lakehouse named Lakehouse1. Forecast data stored in Azure Data Lake Storage Gen2. You plan to ingest the forecast data into Lakehouse1. The data is already formatted, and you do NOT need to apply any further data transformations. The solution must minimize development effort and costs. Which method should you recommend to efficiently ingest the data?**

   A. Use the Copy activity in a pipeline

   B. Use a custom Spark job

   C. Use Dataflow Gen2

   D. Do nothing; data is already in Lakehouse

7. **Which PySpark expression returns the total number of records in the fact table grouped by ProductKey, with results sorted by count in descending order?**

   A. df.groupBy("ProductKey").count().sort("count", descending=True).show()

   B. df.groupBy("ProductKey").count().sort("count", ascending=True).show()

   C. df.groupBy("Key").count().sort("count", descending=True).show()

   D. df.groupBy("ProductKey").count().orderBy("ProductKey").show()

8. **In Microsoft Power BI Desktop, to diagram the model containing only the Sales table and its related tables, which view should you use?**

   A. Report view

   B. Data view

   C. Relationship view

   D. Model view

9. **Which tool would you use to update a measure definition without requiring a service refresh?**

   A. ALM Toolkit

   B. Power BI Desktop

   C. Tabular Editor

   D. DAX Studio

10. Given a model with a Calendar Dimension and a Sales fact with a 1-to-many relationship, which measure expression provides better performance when filtering by a specific year?

A. [Sales]

B. CALCULATE([Sales])

C. CALCULATE([Sales], Calendar[Year] = 2023)

D. SUM(Sales[Amount])

Answers

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

Explanations

1. Which SCD type keeps the latest data, with old data overwritten?

A. Type 3 SCD

B. Type 0 SCD

C. Type 2 SCD

D. Type 1 SCD

When you update a dimension attribute and you want to overwrite the old value, you're using Type 1 SCD. In this approach, the existing attribute value is updated in place, so only the most recent data is stored and no history is kept. If you need to remember changes, you'd use Type 2, which adds a new row for each change to preserve history, or Type 3, which stores a limited history in additional columns. Type 0 means no changes are allowed, which isn't about overwriting with the latest data. So for keeping only the latest data and discarding the past, Type 1 is the correct fit.

2. To enforce access control for embedded reports and prevent public sharing, which tenant setting should you disable?

A. Disable External Content

B. Disable Analyze in Excel

C. Disable XMLA endpoints

D. Disable Publish to web

Disabling the XMLA endpoints blocks the XMLA protocol, which is a common way external tools and apps connect directly to a Fabric data model to query data or metadata. When these endpoints are on, embedded reports or external analytics clients can reach the dataset outside the standard, governed UI, potentially bypassing some access controls and enabling unintended sharing. Turning off the XMLA endpoints enforces a stricter access path: only the service and authenticated users through the official channels can access the data, making it harder for embedded reports to be accessed publicly or teased apart from intended permissions. Other options address different vectors—external content, Excel analytics connections, or public publishing—yet they don't specifically close the direct, programmatic access that XMLA endpoints provide, which is why disabling XMLA endpoints is the most targeted move for tightening access control around embedded reports and preventing unintended exposure.

3. To schedule a pipeline to run every four hours on Mondays and Fridays, which Repeat setting should you use?

A. Daily

B. Hourly

C. Weekly

D. Monthly

Start with the pattern that lets you limit runs to specific days of the week. A weekly repeat lets you choose which days are active—in this case Monday and Friday—and you can still configure the trigger to fire every four hours on those days. That combination achieves "every four hours, but only on Mondays and Fridays." Choosing daily would run on every day, which isn't correct. Hourly would apply more broadly or with too frequent runs, again including days you don't want. Monthly doesn't align with a weekly day-of-week restriction. So the weekly setting is the best fit because it restricts execution to the chosen days while allowing the four-hour interval.

4. In a Fabric workspace containing a Power BI report, when access to a data source is restricted to narrow time windows, what is the recommended approach to bring data into a Power BI semantic model using dataflows?

A. Create a staging dataflow that will only copy the data from the source as-is

B. Use a direct query dataflow

C. Use a dataflow with incremental load

D. Use a dataflow that pre-aggregates data

When access to the data source is only available in narrow time windows, the goal is to decouple the report from real-time source availability and create a reliable, reusable copy for analysis. A staging dataflow that copies the data from the source as-is during those open windows delivers a complete snapshot that the Power BI semantic model can consume later without needing to reach the source again. This ensures the report has a stable dataset to query, even when the source is closed outside the allowed window, and supports repeatable, scheduled refreshes. Direct query would require live access every time the report runs, which isn't feasible if the source is restricted. A dataflow with incremental load still depends on the source during each load to detect new data, which can fail or be impractical under restricted access. A dataflow that pre-aggregates data might improve performance but sacrifices data granularity and flexibility, and it still relies on ingestion events during the allowed window.

5. If warehouse queries degrade in performance due to throttling, which app should you use to identify throttling?

A. Azure Monitor

B. Power BI Performance Analyzer

C. SQL Server Profiler

D. Microsoft Fabric Capacity Metrics app

Throttling happens when resource demand exceeds what the system can safely provide, so queries slow down or get limited to protect overall performance. In a Fabric data warehouse, you'll detect throttling by looking at capacity usage and any throttle signals across your deployed capacities. The Microsoft Fabric Capacity Metrics app is designed for this exact purpose: it surfaces capacity utilization, active queries, queue lengths, and throttle events, letting you quickly see if degraded warehouse performance is due to throttling rather than a poorly optimized query or a separate issue. With this insight, you can decide to scale capacity, reallocate resources, or adjust workloads. Other tools exist for general monitoring or for tracing SQL activity, but they don't target Fabric throttling telemetry as directly or comprehensively.

6. You have a Fabric lakehouse named Lakehouse1. Forecast data stored in Azure Data Lake Storage Gen2. You plan to ingest the forecast data into Lakehouse1. The data is already formatted, and you do NOT need to apply any further data transformations. The solution must minimize development effort and costs. Which method should you recommend to efficiently ingest the data?

A. Use the Copy activity in a pipeline

B. Use a custom Spark job

C. Use Dataflow Gen2

D. Do nothing; data is already in Lakehouse

The situation tests choosing a lightweight, managed data movement method when no transformations are needed. Since the forecast data is already formatted and you want to minimize development effort and cost, a simple pipeline with a Copy activity is the best fit. The Copy activity is designed to move data from a source like Azure Data Lake Storage Gen2 into a Fabric lakehouse with little to no coding, handling file-based ingestion efficiently and reliably. It can preserve the existing schema and structure, supports scheduling or automated runs, and avoids the overhead of building and maintaining custom compute jobs. Using a custom Spark job would add coding, cluster management, and compute costs, which isn't necessary here since no transformations are required. Dataflow Gen2 offers rich ETL capabilities, but that extra functionality isn't needed for straightforward data movement and would introduce more setup and cost. Doing nothing would miss the ingestion step entirely.

7. Which PySpark expression returns the total number of records in the fact table grouped by ProductKey, with results sorted by count in descending order?

A. df.groupBy("ProductKey").count().sort("count", descending=True).show()

B. df.groupBy("ProductKey").count().sort("count", ascending=True).show()

C. df.groupBy("Key").count().sort("count", descending=True).show()

D. df.groupBy("ProductKey").count().orderBy("ProductKey").show()

Grouping by ProductKey and counting gives the total number of records for each product key. After performing the groupBy and count, you have a column named count that holds the number of records per ProductKey. Sorting by that count in descending order lists the products from the highest to the lowest total. The expression df.groupBy("ProductKey").count().sort("count", descending=True).show() does exactly this: it groups by ProductKey, counts the records in each group, and sorts the results by the count in descending order so the most frequent product keys appear first. Sorting by count ascending would be the opposite and not meet the requirement. Grouping by a different column would produce counts per the wrong key. Sorting by ProductKey would order the results by the key values themselves, not by how many records each key has.

8. In Microsoft Power BI Desktop, to diagram the model containing only the Sales table and its related tables, which view should you use?

A. Report view

B. Data view

C. Relationship view

D. Model view

Diagramming the data model in Power BI Desktop is done in Model view. This view presents the tables and their relationships as a diagram, letting you see at a glance how the Sales table connects to its related tables and to arrange or modify those connections as needed. The report view is for building visuals, the data view shows actual data rows, and while the relationship view can display connections, Model view is the dedicated, comprehensive canvas for modeling and diagramming the entire schema.

9. Which tool would you use to update a measure definition without requiring a service refresh?

A. ALM Toolkit

B. Power BI Desktop

C. Tabular Editor

D. DAX Studio

Updating a measure definition without a service refresh relies on changing only the model's metadata, not the stored data. Measures are part of the tabular model's metadata, so you can push those metadata changes to the running server without having to reprocess or refresh the data. The ALM Toolkit is built for this kind of workflow: it supports comparing a source model to a target and applying deployment updates that are schema-only, meaning you can update measures, calculated columns, and other metadata on the live server without triggering a data refresh. This makes it the right fit when you want to propagate a measure change efficiently and safely. Other tools have different primary roles. Power BI Desktop is focused on authoring and then publishing to the service, which often involves data refresh behavior tied to the published dataset. DAX Studio is a query and debugging tool for running DAX, not a deployment mechanism. Tabular Editor is strong for editing the model directly, but ALM Toolkit is specifically designed for controlled, schema-only deployments across environments, which is why it's the best choice for updating a measure without a service refresh.

10. Given a model with a Calendar Dimension and a Sales fact with a 1-to-many relationship, which measure expression provides better performance when filtering by a specific year?

A. [Sales]

B. CALCULATE([Sales])

C. CALCULATE([Sales], Calendar[Year] = 2023)

D. SUM(Sales[Amount])

When you want to restrict a measure to a specific year in a model where a Calendar dimension relates one-to-many to the Sales fact, applying the filter inside CALCULATE is the most efficient approach. CALCULATE changes the filter context, and by including a precise condition like Calendar[Year] = 2023, you push that year filter directly through the relationship to the fact table. This prunes the Sales rows early to only those from 2023, which the engine can execute faster, especially with a large fact table. Relying on a plain measure without an explicit year filter or using a simple SUM lacks that explicit, targeted filter. Without the year constraint, the result could span all years, and depending on the measure's logic, the engine can't optimize as effectively. Using a bare CALCULATE with no filter doesn't introduce the year constraint at all, so it won't give you a fixed-year result. And summing the column bypasses the measure's logic and its filtering, making it less predictable and potentially less efficient for this targeted scenario. So, applying a filter for the specific year within CALCULATE aligns the evaluation with the related calendar data, yielding the best performance when querying for that year.

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://dp600fabricanalyticsengineer.examzify.com

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