



# DP-203<sup>Q&As</sup>

Data Engineering on Microsoft Azure

**Pass Microsoft DP-203 Exam with 100% Guarantee**

Free Download Real Questions & Answers **PDF** and **VCE** file from:

<https://www.geekcert.com/dp-203.html>

100% Passing Guarantee  
100% Money Back Assurance

Following Questions and Answers are all new published by Microsoft  
Official Exam Center

-  **Instant Download** After Purchase
-  **100% Money Back** Guarantee
-  **365 Days** Free Update
-  **800,000+** Satisfied Customers





## QUESTION 1

You build a data warehouse in an Azure Synapse Analytics dedicated SQL pool.

Analysts write a complex SELECT query that contains multiple JOIN and CASE statements to transform data for use in inventory reports. The inventory reports will use the data and additional WHERE parameters depending on the report. The

reports will be produced once daily.

You need to implement a solution to make the dataset available for the reports. The solution must minimize query times.

What should you implement?

- A. an ordered clustered columnstore index
- B. a materialized view
- C. result set caching
- D. a replicated table

Correct Answer: B

Materialized views for dedicated SQL pools in Azure Synapse provide a low maintenance method for complex analytical queries to get fast performance without any query change. Incorrect Answers:

C: One daily execution does not make use of result cache caching.

Note: When result set caching is enabled, dedicated SQL pool automatically caches query results in the user database for repetitive use. This allows subsequent query executions to get results directly from the persisted cache so recomputation is not needed. Result set caching improves query performance and reduces compute resource usage. In addition, queries using cached results set do not use any concurrency slots and thus do not count against existing concurrency limits.

Reference: <https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/performance-tuning-materialized-views> <https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/performance-tuning-result-set-caching>

---

## QUESTION 2

DRAG DROP

You have an Azure subscription that contains an Azure Synapse Analytics workspace named workspace1. Workspace1 connects to an Azure DevOps repository named repo1. Repo1 contains a collaboration branch named main and a

development branch named branch1. Branch1 contains an Azure Synapse pipeline named pipeline1.

In workspace1, you complete testing of pipeline1.

You need to schedule pipeline1 to run daily at 6 AM.



Which four actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

NOTE: More than one order of answer choices is correct. You will receive credit for any of the correct orders you select.

Select and Place:

**Actions**

**Answer Area**

Create a new branch in Repo1.

Merge the changes from branch1 into main.

Associate the schedule trigger with pipeline1.

Switch to Synapse live mode.

Create a schedule trigger.

Publish the contents of main.

Correct Answer:

**Actions**

**Answer Area**

Create a new branch in Repo1.

Create a schedule trigger.

Associate the schedule trigger with pipeline1.

Merge the changes from branch1 into main.

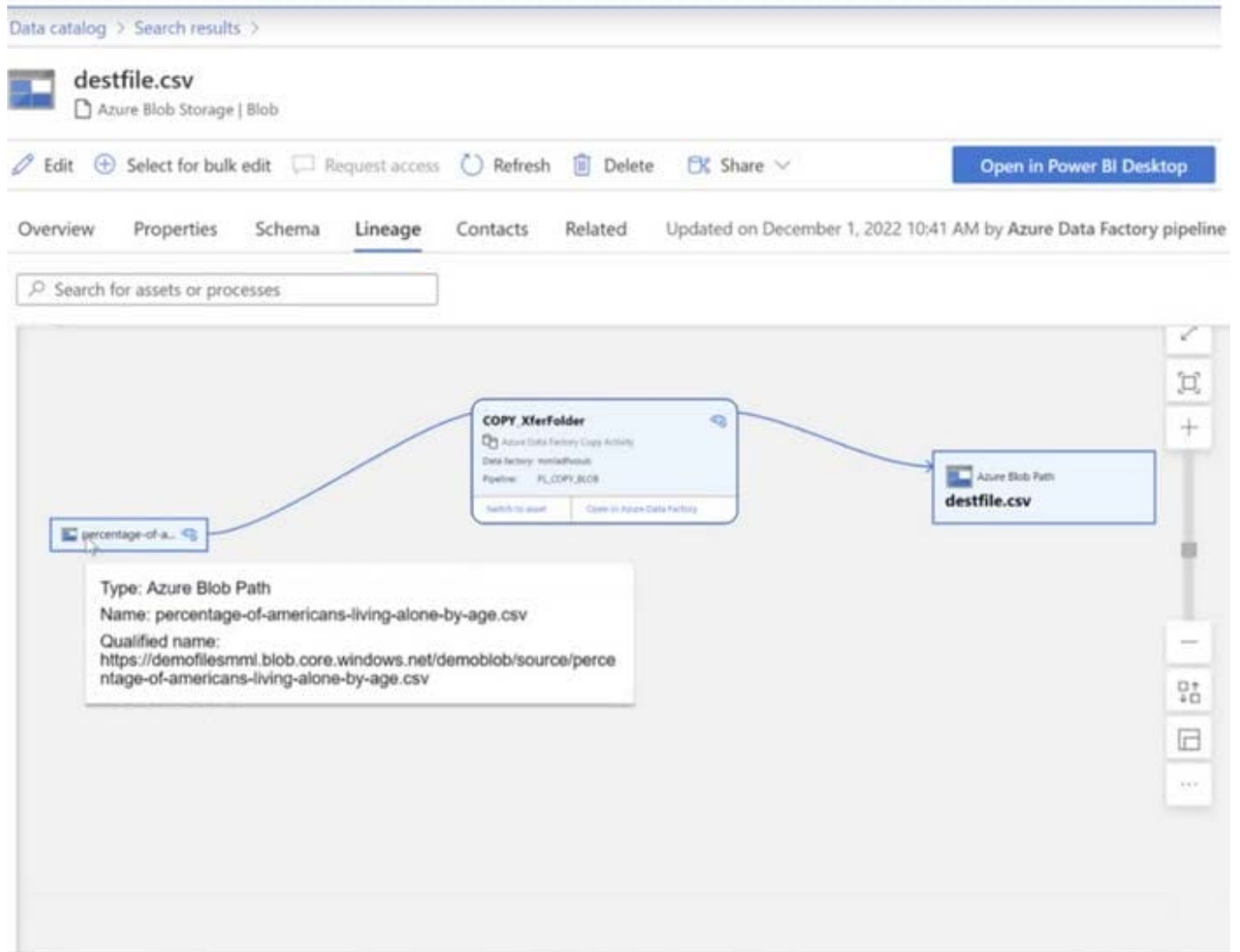
Switch to Synapse live mode.

Publish the contents of main.



### QUESTION 3

You have a Microsoft Purview account. The Lineage view of a CSV file is shown in the following exhibit.



How is the data for the lineage populated?

- A. manually
- B. by scanning data stores
- C. by executing a Data Factory pipeline

Correct Answer: C

From the exhibit we see Copy\_XferFolder (and even: From Data Factory).

The following example is a typical use case of data moving across multiple systems, where the Data Catalog would connect to each of the systems for lineage.

1.



Data Factory copies data from on-prem/raw zone to a landing zone in the cloud.

2.

Etc.

Reference: <https://learn.microsoft.com/en-us/azure/purview/concept-data-lineage>

---

#### QUESTION 4

You have the following Azure Data Factory pipelines

ingest Data from System 1 Ingest Data from System2 Populate Dimensions Populate facts

ingest Data from System1 and Ingest Data from System1 have no dependencies. Populate Dimensions must execute after Ingest Data from System1 and Ingest Data from System\* Populate Facts must execute after the Populate Dimensions pipeline. All the pipelines must execute every eight hours.

What should you do to schedule the pipelines for execution?

- A. Add an event trigger to all four pipelines.
- B. Create a parent pipeline that contains the four pipelines and use an event trigger.
- C. Create a parent pipeline that contains the four pipelines and use a schedule trigger.
- D. Add a schedule trigger to all four pipelines.

Correct Answer: C

Schedule trigger: A trigger that invokes a pipeline on a wall-clock schedule.

Reference:

<https://docs.microsoft.com/en-us/azure/data-factory/concepts-pipeline-execution-triggers>

---

#### QUESTION 5

##### HOTSPOT

You have an Azure Blob storage account that contains a folder. The folder contains 120,000 files. Each file contains 62 columns.

Each day, 1,500 new files are added to the folder.

You plan to incrementally load five data columns from each new file into an Azure Synapse Analytics workspace.

You need to minimize how long it takes to perform the incremental loads.

What should you use to store the files and format?

Hot Area:



Storage:  These are the se

- Multiple blob storage accounts
- Multiple containers in the blob storage account
- Timeslice partitioning in the folders

Format:

- Apache Parquet
- CSV
- JSON

Correct Answer:

Storage:  These are the se

- Multiple blob storage accounts
- Multiple containers in the blob storage account
- Timeslice partitioning in the folders

Format:

- Apache Parquet
- CSV
- JSON

Box 1 = timeslice partitioning in the folders This means that you should organize your files into folders based on a time attribute, such as year, month, day, or hour. For example, you can have a folder structure like /yyyy/mm/dd/file.csv. This way, you can easily identify and load only the new files that are added each day by using a time filter in your Azure Synapse pipeline<sup>12</sup>. Timeslice partitioning can also improve the performance of data loading and querying by reducing the number of files that need to be scanned

Box = 2 Apache Parquet This is because Parquet is a columnar file format that can efficiently store and compress data with many columns. Parquet files can also be partitioned by a time attribute, which can improve the performance of incremental loading and querying by reducing the number of files that need to be scanned<sup>123</sup>. Parquet files are supported by both dedicated SQL pool and serverless SQL pool in Azure Synapse Analytics<sup>2</sup>.



VCE & PDF

GeekCert.com

<https://www.geekcert.com/dp-203.html>

2024 Latest geekcert DP-203 PDF and VCE dumps Download

---

[Latest DP-203 Dumps](#)

[DP-203 VCE Dumps](#)

[DP-203 Study Guide](#)