



DP-203^{Q&As}

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.pass4itsure.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 have several Azure Data Factory pipelines that contain a mix of the following types of activities.

1.

Wrangling data flow

2.

Notebook

3.

Copy

4.

jar

Which two Azure services should you use to debug the activities? Each correct answer presents part of the solution

NOTE: Each correct selection is worth one point.

A. Azure Machine Learning

B. Azure Data Factory

C. Azure Synapse Analytics

D. Azure HDInsight

E. Azure Databricks

Correct Answer: BE

B: Azure Data Factory: Azure Data Factory itself provides debugging capabilities for its activities. You can monitor and debug the execution of pipeline activities directly within the Azure Data Factory interface. It allows you to view activity run details, input/output data, logs, and diagnose any errors or issues encountered during execution.

Use Azure Data Factory for wrangling and Copy.

E: Azure Databricks: Azure Databricks is a powerful analytics platform that integrates well with Azure Data Factory. You can use it to debug and analyze Notebook activities within the Data Factory pipelines. Azure Databricks provides an interactive environment to run and debug notebooks, allowing you to inspect intermediate data, execute code step-by-step, and troubleshoot any issues.

Azure Databricks for Jar and Notbooks.

Reference: <https://learn.microsoft.com/en-us/azure/data-factory/iterative-development-debugging>

QUESTION 2

You have an Azure subscription that contains the resources shown in the following table.



Name	Type	Description
storage1	Azure Blob storage account	Contains publicly accessible TSV files that do NOT have a header row
WS1	Azure Synapse Analytics workspace	Contains a serverless SQL pool

You need to read the TSV files by using ad-hoc queries and the OPENROWSET function. The solution must assign a name and override the inferred data type of each column.

What should you include in the OPENROWSET function?

- A. the WITH clause
- B. the ROWSET_OPTIONS bulk option
- C. the DATAFILETYPE bulk option
- D. the DATA_SOURCE parameter

Correct Answer: D

Data source usage

Previous example uses full path to the file. As an alternative, you can create an external data source with the location that points to the root folder of the storage:

create external data source covid

with (location = '\\https://pandemicdatalake.blob.core.windows.net/public/curated/covid-19/ecdc_cases\\');

Once you create a data source, you can use that data source and the relative path to the file in OPENROWSET function:

select top 10 *

from openrowset(bulk '\\latest/ecdc_cases.csv\\', data_source = '\\covid\\', format = '\\csv\\', parser_version = '\\2.0\\', firstrow = 2) as rows

Incorrect:

Not B: ROWSET_OPTIONS not need if there is no header.

Reference:

<https://learn.microsoft.com/en-us/azure/synapse-analytics/sql/query-single-csv-file>

QUESTION 3

You have an Azure Synapse Analytics Apache Spark pool named Pool1.

You plan to load JSON files from an Azure Data Lake Storage Gen2 container into the tables in Pool1. The structure



and data types vary by file.

You need to load the files into the tables. The solution must maintain the source data types.

What should you do?

- A. Use a Get Metadata activity in Azure Data Factory.
- B. Use a Conditional Split transformation in an Azure Synapse data flow.
- C. Load the data by using the OPEHRowset Transact-SQL command in an Azure Synapse Analytics serverless SQL pool.
- D. Load the data by using PySpark.

Correct Answer: A

Serverless SQL pool can automatically synchronize metadata from Apache Spark. A serverless SQL pool database will be created for each database existing in serverless Apache Spark pools.

Serverless SQL pool enables you to query data in your data lake. It offers a T-SQL query surface area that accommodates semi-structured and unstructured data queries. To support a smooth experience for in place querying of data that's

located in Azure Storage files, serverless SQL pool uses the OPENROWSET function with additional capabilities.

The easiest way to see to the content of your JSON file is to provide the file URL to the OPENROWSET function, specify csv FORMAT.

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/query-json-files>

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/query-data-storage>

QUESTION 4

You have an Azure subscription that contains an Azure Synapse Analytics workspace named ws1 and an Azure Cosmos DB database account named Cosmos1. Cosmos1 contains a container named container1 and ws1 contains a serverless SQL pool.

You need to ensure that you can query the data in container1 by using the serverless SQL pool.

Which three actions should you perform? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. Enable Azure Synapse Link for Cosmos1.
- B. Disable the analytical store for container1.
- C. In ws1, create a linked service that references Cosmos1.
- D. Enable the analytical store for container1.



E. Disable indexing for container1.

Correct Answer: ACD

Query Azure Cosmos DB data with a serverless SQL pool in Azure Synapse Link

Prerequisites include:

(D, not B) Make sure that you have prepared Analytical store:

Enable analytical store on your Azure Cosmos DB containers.

(A) An Azure Cosmos DB database account that's Azure Synapse Link enabled. Etc.

Note: A serverless SQL pool allows you to analyze data in your Azure Cosmos DB containers that are enabled with Azure Synapse Link in near real time without affecting the performance of your transactional workloads. It offers a familiar TSQL syntax to query data from the analytical store and integrated connectivity to a wide range of business intelligence (BI) and ad-hoc querying tools via the T-SQL interface.

Reference: <https://learn.microsoft.com/en-us/azure/synapse-analytics/sql/query-cosmos-db-analytical-store>

QUESTION 5

You are designing an inventory updates table in an Azure Synapse Analytics dedicated SQL pool. The table will have a clustered columnstore index and will include the following columns:

Table	Comment
EventDate	One million records are added to the table each day
EventTypeID	The table contains 10 million records for each event type.
WarehouseID	The table contains 100 million records for each warehouse.
ProductCategoryTypeID	The table contains 25 million records for each product category type.

You identify the following usage patterns:

1.

Analysts will most commonly analyze transactions for a warehouse.

2.

Queries will summarize by product category type, date, and/or inventory event type.

You need to recommend a partition strategy for the table to minimize query times.

On which column should you partition the table?

A. EventTypeID

B. ProductCategoryTypeID

C. EventDate

D. WarehouseID



Correct Answer: D

The number of records for each warehouse is big enough for a good partitioning.

Note: Table partitions enable you to divide your data into smaller groups of data. In most cases, table partitions are created on a date column.

When creating partitions on clustered columnstore tables, it is important to consider how many rows belong to each partition. For optimal compression and performance of clustered columnstore tables, a minimum of 1 million rows per

distribution and partition is needed. Before partitions are created, dedicated SQL pool already divides each table into 60 distributed databases.

[Latest DP-203 Dumps](#)

[DP-203 Practice Test](#)

[DP-203 Exam Questions](#)