



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.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 are designing an Azure Synapse Analytics workspace.

You need to recommend a solution to provide double encryption of all the data at rest.

Which two components should you include in the recommendation? Each correct answer presents part of the solution

NOTE: Each correct selection is worth one point.

- A. an X509 certificate
- B. an RSA key
- C. an Azure key vault that has purge protection enabled
- D. an Azure virtual network that has a network security group (NSG)
- E. an Azure Policy initiative

Correct Answer: BE

Synapse workspaces encryption uses existing keys or new keys generated in Azure Key Vault. A single key is used to encrypt all the data in a workspace. Synapse workspaces support RSA 2048 and 3072 byte-sized keys, and RSA-HSM keys.

The Key Vault itself needs to have purge protection enabled.

Reference: <https://docs.microsoft.com/en-us/azure/synapse-analytics/security/workspaces-encryption>

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 OPENROWSETfunction. The solution must assign a name and override the inferred data type of each column.

What should you include in the OPENROWSETfunction?

- A. the WITH clause
- B. the ROWSET_OPTIONSbulk option



C. the DATAFILETYPEbulk option

D. the DATA_SOURCEparameter

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 workspace named MyWorkspace that contains an Apache Spark database named mytestdb.

You run the following command in an Azure Synapse Analytics Spark pool in MyWorkspace.

```
CREATE TABLE mytestdb.myParquetTable(EmployeeID int,EmployeeName string,EmployeeStartDate date)
```

```
USING Parquet
```

You then use Spark to insert a row into mytestdb.myParquetTable. The row contains the following data.

EmployeeName	EmployeeID	EmployeeStartDate
Alice	24	2020-01-25

One minute later, you execute the following query from a serverless SQL pool in MyWorkspace.

```
SELECT EmployeeIDFROM mytestdb.dbo.myParquetTableWHERE EmployeeName = '\\Alice\\';
```

What will be returned by the query?



- A. 24
- B. an error
- C. a null value

Correct Answer: B

Once a database has been created by a Spark job, you can create tables in it with Spark that use Parquet as the storage format. Table names will be converted to lower case and need to be queried using the lower case name. These tables will immediately become available for querying by any of the Azure Synapse workspace Spark pools. They can also be used from any of the Spark jobs subject to permissions.

Note: For external tables, since they are synchronized to serverless SQL pool asynchronously, there will be a delay until they appear.

Reference: <https://docs.microsoft.com/en-us/azure/synapse-analytics/metadata/table>

QUESTION 4

You have two fact tables named Flight and Weather. Queries targeting the tables will be based on the join between the following columns.

Table	Column
Flight	ArrivalAirportID ArrivalDateTime
Weather	AirportID ReportDateTime

You need to recommend a solution that maximizes query performance. What should you include in the recommendation?

- A. In the tables use a hash distribution of ArrivalDateTime and ReportDateTime.
- B. In the tables use a hash distribution of ArrivalAirportID and AirportID.
- C. In each table, create an IDENTITY column.
- D. In each table, create a column as a composite of the other two columns in the table.

Correct Answer: B

Hash-distribution improves query performance on large fact tables. Incorrect Answers:

A: Do not use a date column for hash distribution. All data for the same date lands in the same distribution. If several users are all filtering on the same date, then only 1 of the 60 distributions do all the processing work.

**QUESTION 5**

You need to trigger an Azure Data Factory pipeline when a file arrives in an Azure Data Lake Storage Gen2 container. Which resource provider should you enable?

- A. Microsoft.Sql
- B. Microsoft.Automation
- C. Microsoft.EventGrid
- D. Microsoft.EventHub

Correct Answer: C

Event-driven architecture (EDA) is a common data integration pattern that involves production, detection, consumption, and reaction to events. Data integration scenarios often require Data Factory customers to trigger pipelines based on events happening in storage account, such as the arrival or deletion of a file in Azure Blob Storage account. Data Factory natively integrates with Azure Event Grid, which lets you trigger pipelines on such events.

Reference: <https://docs.microsoft.com/en-us/azure/data-factory/how-to-create-event-trigger>
<https://docs.microsoft.com/en-us/azure/data-factory/concepts-pipeline-execution-triggers>

[DP-203 PDF Dumps](#)

[DP-203 Practice Test](#)

[DP-203 Study Guide](#)