



# 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

### HOTSPOT

You need to implement an Azure Databricks cluster that automatically connects to Azure Data Lake Storage Gen2 by using Azure Active Directory (Azure AD) integration.

How should you configure the new cluster? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

### Answer Area

Tier:

	▼
Premium	
Standard	

Advanced option to enable:

	▼
Azure Data Lake Storage Credential Passthrough	
Table Access Control	

Correct Answer:

### Answer Area

Tier:

	▼
Premium	
Standard	

Advanced option to enable:

	▼
Azure Data Lake Storage Credential Passthrough	
Table Access Control	

Box 1: Premium

Credential passthrough requires an Azure Databricks Premium Plan

Box 2: Azure Data Lake Storage credential passthrough

You can access Azure Data Lake Storage using Azure Active Directory credential passthrough.



When you enable your cluster for Azure Data Lake Storage credential passthrough, commands that you run on that cluster can read and write data in Azure Data Lake Storage without requiring you to configure service principal credentials for

access to storage.

Reference:

<https://docs.microsoft.com/en-us/azure/databricks/security/credential-passthrough/adls-passthrough>

## QUESTION 2

### HOTSPOT

You are designing an Azure Synapse Analytics dedicated SQL pool.

Groups will have access to sensitive data in the pool as shown in the following table.

Name	Enhanced access
Executives	No access to sensitive data
Analysts	Access to in-region sensitive data
Engineers	Access to all numeric sensitive data

You have policies for the sensitive data. The policies vary by region as shown in the following table.

Region	Data considered sensitive
RegionA	Financial, Personally Identifiable Information (PII)
RegionB	Financial, Personally Identifiable Information (PII), medical
RegionC	Financial, medical

You have a table of patients for each region. The tables contain the following potentially sensitive columns.

Name	Sensitive data	Description
CardOnFile	Financial	Debit/credit card number for charges
Height	Medical	Patient's height in cm
ContactEmail	PII	Email address for secure communications

You are designing dynamic data masking to maintain compliance.

For each of the following statements, select Yes if the statement is true. Otherwise, select No.

NOTE: Each correct selection is worth one point.

Hot Area:



## Answer Area

Statements	Yes	No
Analysts in RegionA require dynamic data masking rules for [Patients_RegionA].	<input type="radio"/>	<input type="radio"/>
Engineers in RegionC require a dynamic data masking rule for [Patients_RegionA], [Height]	<input type="radio"/>	<input type="radio"/>
Engineers in RegionB require a dynamic data masking rule for [Patients_RegionB], [Height]	<input type="radio"/>	<input type="radio"/>

Correct Answer:

## Answer Area

Statements	Yes	No
Analysts in RegionA require dynamic data masking rules for [Patients_RegionA].	<input checked="" type="radio"/>	<input type="radio"/>
Engineers in RegionC require a dynamic data masking rule for [Patients_RegionA], [Height]	<input type="radio"/>	<input checked="" type="radio"/>
Engineers in RegionB require a dynamic data masking rule for [Patients_RegionB], [Height]	<input checked="" type="radio"/>	<input type="radio"/>

Reference: <https://docs.microsoft.com/en-us/azure/azure-sql/database/dynamic-data-masking-overview>

### QUESTION 3

You are designing an Azure Data Lake Storage solution that will transform raw JSON files for use in an analytical workload.

You need to recommend a format for the transformed files. The solution must meet the following requirements:

1.  
Contain information about the data types of each column in the files.
- 2.



---

Support querying a subset of columns in the files.

3.

Support read-heavy analytical workloads.

4.

Minimize the file size.

What should you recommend?

A. JSON

B. CSV

C. Apache Avro

D. Apache Parquet

Correct Answer: D

Parquet, an open-source file format for Hadoop, stores nested data structures in a flat columnar format.

Compared to a traditional approach where data is stored in a row-oriented approach, Parquet file format is more efficient in terms of storage and performance.

It is especially good for queries that read particular columns from a “wide” (with many columns) table since only needed columns are read, and IO is minimized.

Incorrect:

Not C:

The Avro format is the ideal candidate for storing data in a data lake landing zone because:

1.

Data from the landing zone is usually read as a whole for further processing by downstream systems (the row-based format is more efficient in this case).

2.

Downstream systems can easily retrieve table schemas from Avro files (there is no need to store the schemas separately in an external meta store).

3.

Any source schema change is easily handled (schema evolution). Reference: <https://www.clairvoyant.ai/blog/big-data-file-formats>

---

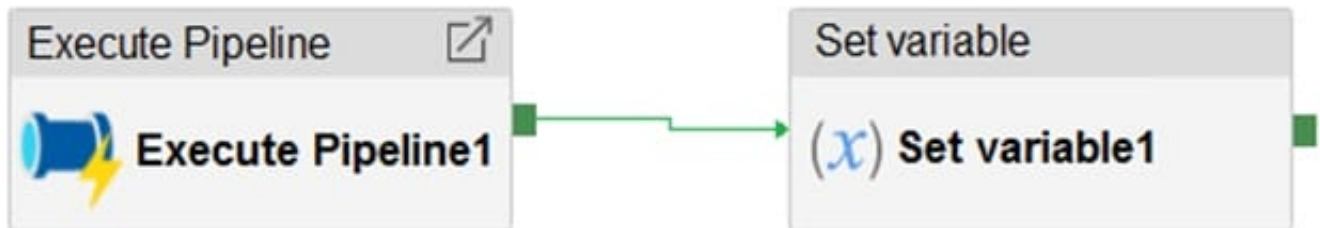
#### QUESTION 4

You have an Azure Data Factory instance that contains two pipelines named Pipeline1 and Pipeline2. Pipeline1 has the activities shown in the following exhibit.





Pipeline2 has the activities shown in the following exhibit.



You execute Pipeline2, and Stored procedure1 in Pipeline1 fails. What is the status of the pipeline runs?

- A. Pipeline1 and Pipeline2 succeeded.
- B. Pipeline1 and Pipeline2 failed.
- C. Pipeline1 succeeded and Pipeline2 failed.
- D. Pipeline1 failed and Pipeline2 succeeded.

Correct Answer: A

Activities are linked together via dependencies. A dependency has a condition of one of the following: Succeeded, Failed, Skipped, or Completed.

Consider Pipeline1:

If we have a pipeline with two activities where Activity2 has a failure dependency on Activity1, the pipeline will not fail just because Activity1 failed. If Activity1 fails and Activity2 succeeds, the pipeline will succeed. This scenario is treated as a

try-catch block by Data Factory.



The failure dependency means this pipeline reports success.

Note:

If we have a pipeline containing Activity1 and Activity2, and Activity2 has a success dependency on Activity1, it will only



---

execute if Activity1 is successful. In this scenario, if Activity1 fails, the pipeline will fail.

Reference:

<https://datasavvy.me/category/azure-data-factory/>

---

## QUESTION 5

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>

[Latest DP-203 Dumps](#)

[DP-203 Study Guide](#)

[DP-203 Braindumps](#)