



# 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.passapply.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 collect application metrics, streaming query events, and application log messages for an Azure Databrick cluster.

Which type of library and workspace should you implement? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

## Answer Area

Library:

	▼
Azure Databricks Monitoring Library	
Microsoft Azure Management Monitoring Library	
PyTorch	
TensorFlow	

Workspace:

	▼
Azure Databricks	
Azure Log Analytics	
Azure Machine Learning	

Correct Answer:



## Answer Area

Library:

	▼
Azure Databricks Monitoring Library	
Microsoft Azure Management Monitoring Library	
PyTorch	
TensorFlow	

Workspace:

	▼
Azure Databricks	
Azure Log Analytics	
Azure Machine Learning	

You can send application logs and metrics from Azure Databricks to a Log Analytics workspace. It uses the Azure Databricks Monitoring Library, which is available on GitHub.

Reference: <https://docs.microsoft.com/en-us/azure/architecture/databricks-monitoring/application-logs>

---

### QUESTION 2

You are performing exploratory analysis of the bus fare data in an Azure Data Lake Storage Gen2 account by using an Azure Synapse Analytics serverless SQL pool. You execute the Transact-SQL query shown in the following exhibit.



```
SELECT
    payment_type,
    SUM(fare_amount) AS fare_total
FROM OPENROWSET (
    BULK 'csv/busfare/tripdata_2020*.csv',
    DATA_SOURCE = 'BusData',
    FORMAT = 'CSV', PARSER_VERSION = '2.0',
    FIRSTROW = 2
)
WITH (
    payment_type INT 10,
    fare_amount FLOAT 11
) AS nyc
GROUP BY payment_type
ORDER BY payment_type;
```

What do the query results include?

- A. Only CSV files in the tripdata\_2020 subfolder.
- B. All files that have file names that beginning with "tripdata\_2020".
- C. All CSV files that have file names that contain "tripdata\_2020".
- D. Only CSV that have file names that beginning with "tripdata\_2020".

Correct Answer: D

---

### QUESTION 3

#### HOTSPOT

You have an Azure Synapse Analytics serverless SQL pool, an Azure Synapse Analytics dedicated SQL pool, an Apache Spark pool, and an Azure Data Lake Storage Gen2 account.

You need to create a table in a lake database. The table must be available to both the serverless SQL pool and the Spark pool.

Where should you create the table, and Which file format should you use for data in the table? To answer, select the appropriate options in the answer area.



NOTE: Each correct selection is worth one point.

Hot Area:

Create the table in:

- The dedicated SQL pool
- The serverless SQL pool
- The Spark pool

File format:

- Apache Parquet
- Delta
- JSON

Correct Answer:

Create the table in:

- The dedicated SQL pool
- The serverless SQL pool
- The Spark pool

File format:

- Apache Parquet
- Delta
- JSON

#### QUESTION 4

##### HOTSPOT

You are designing an application that will store petabytes of medical imaging data

When the data is first created, the data will be accessed frequently during the first week. After one month, the data must be accessible within 30 seconds, but files will be accessed infrequently. After one year, the data will be accessed

infrequently but must be accessible within five minutes.

You need to select a storage strategy for the data. The solution must minimize costs.

Which storage tier should you use for each time frame? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.



Hot Area:

First week:

	▼
Archive	
Cool	
Hot	

After one month:

	▼
Archive	
Cool	
Hot	

After one year:

	▼
Archive	
Cool	
Hot	

Correct Answer:

First week:

	▼
Archive	
Cool	
Hot	

After one month:

	▼
Archive	
Cool	
Hot	

After one year:

	▼
Archive	
Cool	
Hot	

Explanation:

Box 1: Hot



Hot tier - An online tier optimized for storing data that is accessed or modified frequently. The Hot tier has the highest storage costs, but the lowest access costs.

Box 2: Cool

Cool tier - An online tier optimized for storing data that is infrequently accessed or modified. Data in the Cool tier should be stored for a minimum of 30 days. The Cool tier has lower storage costs and higher access costs compared to the Hot

tier.

Box 3: Cool

Not Archive tier - An offline tier optimized for storing data that is rarely accessed, and that has flexible latency requirements, on the order of hours. Data in the Archive tier should be stored for a minimum of 180 days.

	Premium performance	Hot tier	Cool tier	Archive tier
Availability	99.9%	99.9%	99%	Offline
Availability (RA-GRS reads)	N/A	99.99%	99.9%	Offline
Usage charges	Higher storage costs, lower access, and transaction cost	Higher storage costs, lower access, and transaction costs	Lower storage costs, higher access, and transaction costs	Lowest storage costs, highest access, and transaction costs
Minimum object size	N/A	N/A	N/A	N/A
Minimum storage duration	N/A	N/A	30 days <sup>1</sup>	180 days
Latency (Time to first byte)	Single-digit milliseconds	milliseconds	milliseconds	hours <sup>2</sup>

Reference: <https://docs.microsoft.com/en-us/azure/storage/blobs/access-tiers-overview> <https://www.altaro.com/hyper-v/azure-archive-storage/>

## QUESTION 5

You plan to create an Azure Data Factory pipeline that will include a mapping data flow. You have JSON data containing objects that have nested arrays.

You need to transform the JSON-formatted data into a tabular dataset. The dataset must have one row for each item in the arrays. Which transformation method should you use in the mapping data flow?

A. unpivot





B. flatten

C. new branch

D. alter row

Correct Answer: B

Use the flatten transformation to take array values inside hierarchical structures such as JSON and unroll them into individual rows. This process is known as denormalization.

Reference: <https://docs.microsoft.com/en-us/azure/data-factory/data-flow-flatten>

[Latest DP-203 Dumps](#)

[DP-203 VCE Dumps](#)

[DP-203 Braindumps](#)