

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.pass2lead.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 have an Azure Synapse Analytics dedicated SQL pool named Pool1 and an Azure Data Lake Storage Gen2 account named Account 1.

You plan to access the files in Account1 by using an external table.

You need to create a data source in Pool1 that you can reference when you create the external table.

How should you complete the Transact-SQL statement? To answer, select the appropriate options in the answer area.

NOTE Each correct selection is worth one point.

Hot Area:

```
CREATE EXTERNAL DATA SOURCE sourcel
WITH
  ( LOCATION = 'https://account1.  .core.windows.net',
    
    
    
  )
```

Correct Answer:

```
CREATE EXTERNAL DATA SOURCE sourcel
WITH
  ( LOCATION = 'https://account1.  .core.windows.net',
    
    
    
  )
```

Box 1: blob The following example creates an external data source for Azure Data Lake Gen2 CREATE EXTERNAL DATA SOURCE YellowTaxi WITH (LOCATION = '\\https://azureopendatastorage.blob.core.windows.net/nyctlc/yellow/' , TYPE = HADOOP)

Box 2: HADOOP

QUESTION 2

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You have an Azure Storage account that contains 100 GB of files. The files contain rows of text and numerical values. 75% of the rows contain description data that has an average length of 1.1 MB.

You plan to copy the data from the storage account to an enterprise data warehouse in Azure Synapse Analytics.

You need to prepare the files to ensure that the data copies quickly.

Solution: You modify the files to ensure that each row is more than 1 MB.

Does this meet the goal?

A. Yes

B. No

Correct Answer: B

Instead convert the files to compressed delimited text files.

Reference: <https://docs.microsoft.com/en-us/azure/sql-data-warehouse/guidance-for-loading-data>

QUESTION 3

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while

others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You plan to create an Azure Databricks workspace that has a tiered structure. The workspace will contain the following three workloads:

1.

A workload for data engineers who will use Python and SQL.

2.

A workload for jobs that will run notebooks that use Python, Scala, and SOL.

3.

A workload that data scientists will use to perform ad hoc analysis in Scala and R.

The enterprise architecture team at your company identifies the following standards for Databricks environments:

1.

The data engineers must share a cluster.

2.

The job cluster will be managed by using a request process whereby data scientists and data engineers provide packaged notebooks for deployment to the cluster.

3.

All the data scientists must be assigned their own cluster that terminates automatically after 120 minutes of inactivity. Currently, there are three data scientists.

You need to create the Databricks clusters for the workloads.

Solution: You create a Standard cluster for each data scientist, a High Concurrency cluster for the data engineers, and a High Concurrency cluster for the jobs.

Does this meet the goal?

A. Yes

B. No

Correct Answer: A

We need a High Concurrency cluster for the data engineers and the jobs.

Note:

Standard clusters are recommended for a single user. Standard can run workloads developed in any language:

Python, R, Scala, and SQL.

A high concurrency cluster is a managed cloud resource. The key benefits of high concurrency clusters are that they provide Apache Spark-native fine-grained sharing for maximum resource utilization and minimum query latencies.

Reference:

<https://docs.azuredatabricks.net/clusters/configure.html>

QUESTION 4

You have an Azure data factory named ADF1.

You currently publish all pipeline authoring changes directly to ADF1.

You need to implement version control for the changes made to pipeline artifacts. The solution must ensure that you can apply version control to the resources currently defined in the Azure Data Factory Studio for ADF1.

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

NOTE: Each correct selection is worth one point.

A. From the Azure Data Factory Studio, run Publish All.

- B. Create an Azure Data Factory trigger.
- C. Create a Git repository.
- D. Create a GitHub action.
- E. From the Azure Data Factory Studio, select Set up code repository.
- F. From the Azure Data Factory Studio, select Publish.

Correct Answer: CE

Source control in Azure Data Factory

C: By default, the Azure Data Factory user interface experience (UX) authors directly against the data factory service. This experience has the following limitations:

The Data Factory service doesn't include a repository for storing the JSON entities for your changes. The only way to save changes is via the Publish All button and all changes are published directly to the data factory service.

The Data Factory service isn't optimized for collaboration and version control.

The Azure Resource Manager template required to deploy Data Factory itself is not included.

E: To provide a better authoring experience, Azure Data Factory allows you to configure a Git repository with either Azure Repos or GitHub. Git is a version control system that allows for easier change tracking and collaboration. Connect to a Git repository There are four different ways to connect a Git repository to your data factory for both Azure Repos and GitHub. After you connect to a Git repository, you can view and manage your configuration in the management hub under Git configuration

in the Source control section Configuration method 1: Home page

1.

In the Azure Data Factory home page, select Set up code repository at the top.

2.

Etc.

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

QUESTION 5

HOTSPOT

You have the following Azure Stream Analytics query.

WITH

```
step1 AS (SELECT *  
  FROM input1  
  PARTITION BY StateID  
  INTO 10),  
step2 AS (SELECT *  
  FROM input2  
  PARTITION BY StateID  
  INTO 10)
```

```
SELECT *  
INTO output  
FROM step1  
PARTITION BY StateID  
UNION step2  
  BY StateID
```

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:

Statements	Yes	No
The query joins two streams of partitioned data.	<input type="radio"/>	<input type="radio"/>
The stream scheme key and count must match the output scheme.	<input type="radio"/>	<input type="radio"/>
Providing 60 streaming units will optimize the performance of the query.	<input type="radio"/>	<input type="radio"/>

Correct Answer:

Statements	Yes	No
The query joins two streams of partitioned data.	<input checked="" type="radio"/>	<input type="radio"/>
The stream scheme key and count must match the output scheme.	<input checked="" type="radio"/>	<input type="radio"/>
Providing 60 streaming units will optimize the performance of the query.	<input checked="" type="radio"/>	<input type="radio"/>

Box 1: Yes

You can now use a new extension of Azure Stream Analytics SQL to specify the number of partitions of a stream when reshuffling the data.

The outcome is a stream that has the same partition scheme. Please see below for an example:

```
WITH step1 AS (SELECT * FROM [input1] PARTITION BY DeviceID INTO 10), step2 AS (SELECT * FROM [input2] PARTITION BY DeviceID INTO 10)
```

```
SELECT * INTO [output] FROM step1 PARTITION BY DeviceID UNION step2 PARTITION BY DeviceID
```

Note: The new extension of Azure Stream Analytics SQL includes a keyword INTO that allows you to specify the number of partitions for a stream when performing reshuffling using a PARTITION BY statement.

Box 2: Yes

When joining two streams of data explicitly repartitioned, these streams must have the same partition key and partition count.

Box 3: Yes

10 partitions x six SUs = 60 SUs is fine.

Note: Remember, Streaming Unit (SU) count, which is the unit of scale for Azure Stream Analytics, must be adjusted so the number of physical resources available to the job can fit the partitioned flow. In general, six SUs is a good number to

assign to each partition. In case there are insufficient resources assigned to the job, the system will only apply the repartition if it benefits the job.

[Latest DP-203 Dumps](#)

[DP-203 PDF Dumps](#)

[DP-203 VCE Dumps](#)