

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

DRAG DROP

You have an Azure Data Lake Storage Gen2 account that contains a JSON file for customers. The file contains two attributes named FirstName and LastName.

You need to copy the data from the JSON file to an Azure Synapse Analytics table by using Azure Databricks. A new column must be created that concatenates the FirstName and LastName values.

You create the following components:

1.

A destination table in Azure Synapse

2.

An Azure Blob storage container

3.

A service principal

Which five actions should you perform in sequence next in is Databricks notebook? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

Select and Place:

Actions

Answer Area

- Mount the Data Lake Storage onto DBFS.
- Write the results to a table in Azure Synapse.
- Perform transformations on the file.
- Specify a temporary folder to stage the data.
- Write the results to Data Lake Storage.
- Read the file into a data frame.
- Drop the data frame.
- Perform transformations on the data frame.

Correct Answer:

Actions

Perform transformations on the file.
Write the results to Data Lake Storage.
Drop the data frame.

Answer Area

Mount the Data Lake Storage onto DBFS.
Read the file into a data frame.
Perform transformations on the data frame.
Specify a temporary folder to stage the data.
Write the results to a table in Azure Synapse.

Step 1: Mount the Data Lake Storage onto DBFS

Begin with creating a file system in the Azure Data Lake Storage Gen2 account.

Step 2: Read the file into a data frame.

You can load the json files as a data frame in Azure Databricks.

Step 3: Perform transformations on the data frame.

Step 4: Specify a temporary folder to stage the data Specify a temporary folder to use while moving data between Azure Databricks and Azure Synapse.

Step 5: Write the results to a table in Azure Synapse.

You upload the transformed data frame into Azure Synapse. You use the Azure Synapse connector for Azure Databricks to directly upload a dataframe as a table in a Azure Synapse. <https://docs.databricks.com/data/data-sources/azure/azure-datalake-gen2.html> <https://docs.microsoft.com/en-us/azure/databricks/scenarios/databricks-extract-load-sql-data-warehouse>

QUESTION 2

You have an Azure Data Factory that contains 10 pipelines.

You need to label each pipeline with its main purpose of either ingest, transform, or load. The labels must be available for grouping and filtering when using the monitoring experience in Data Factory.

What should you add to each pipeline?

- A. a resource tag
- B. a correlation ID
- C. a run group ID
- D. an annotation

Correct Answer: D

Annotations are additional, informative tags that you can add to specific factory resources: pipelines, datasets, linked services, and triggers. By adding annotations, you can easily filter and search for specific factory resources.

Reference: <https://www.cathrinewilhelmsen.net/annotations-user-properties-azure-data-factory/>

QUESTION 3

A company purchases IoT devices to monitor manufacturing machinery. The company uses an Azure IoT Hub to communicate with the IoT devices.

The company must be able to monitor the devices in real-time.

You need to design the solution.

What should you recommend?

- A. Azure Analysis Services using Microsoft Visual Studio
- B. Azure Data Factory instance using Azure PowerShell
- C. Azure Analysis Services using Azure PowerShell
- D. Azure Stream Analytics cloud job using Azure Portal

Correct Answer: D

In a real-world scenario, you could have hundreds of these sensors generating events as a stream. Ideally, a gateway device would run code to push these events to Azure Event Hubs or Azure IoT Hubs. Your Stream Analytics job would ingest these events from Event Hubs and run real-time analytics queries against the streams.

Create a Stream Analytics job:

In the Azure portal, select + Create a resource from the left navigation menu. Then, select Stream Analytics job from Analytics.

Reference:

<https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-get-started-with-azure-stream-analytics-to-process-data-from-iot-devices>

QUESTION 4

HOTSPOT

You are creating dimensions for a data warehouse in an Azure Synapse Analytics dedicated SQL pool.

You create a table by using the Transact-SQL statement shown in the following exhibit.

```
CREATE TABLE [DBO].[DimProduct] (  
    [ProductKey] [int] IDENTITY(1,1) NOT NULL,  
    [ProductSourceID] [int] NOT NULL,  
    [ProductName] [nvarchar](100) NOT NULL,  
    [ProductNumber] [nvarchar](25) NOT NULL,  
    [Color] [nvarchar](15) NULL,  
    [Size] [nvarchar](5) NULL,  
    [Weight] [decimal](8, 2) NULL,  
    [ProductCategory] [nvarchar](100) NULL,  
    [SellStartDate] [date] NOT NULL,  
    [SellEndDate] [date] NULL,  
    [RowInsertedDateTime] [datetime] NOT NULL,  
    [RowUpdatedDateTime] [datetime] NOT NULL,  
    [ETLAuditID] [int] NOT NULL  
)
```

Use the drop-down menus to select the answer choice that completes each statement based on the information presented in the graphic. NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

DimProduct is a **[answer choice]** slowly changing dimension (SCD).

	▼
Type 0	
Type 1	
Type 2	

The ProductKey column is **[answer choice]**.

	▼
a surrogate key	
a business key	
an audit column	

Correct Answer:

Answer Area

DimProduct is a **[answer choice]** slowly changing dimension (SCD).

	▼
Type 0	
Type 1	
Type 2	

The ProductKey column is **[answer choice]**.

	▼
a surrogate key	
a business key	
an audit column	

Box 1: Type 2

A Type 2 SCD supports versioning of dimension members. Often the source system doesn't store versions, so the data warehouse load process detects and manages changes in a dimension table. In this case, the dimension table must use

a surrogate key to provide a unique reference to a version of the dimension member. It also includes columns that define the date range validity of the version (for example, StartDate and EndDate) and possibly a flag column (for example,

IsCurrent) to easily filter by current dimension members.

Incorrect Answers:

A Type 1 SCD always reflects the latest values, and when changes in source data are detected, the dimension table data is overwritten.

Box 2: a business key

A business key or natural key is an index which identifies uniqueness of a row based on columns that exist naturally in a table according to business rules. For example business keys are customer code in a customer table, composite of sales

order header number and sales order item line number within a sales order details table.

Reference:

<https://docs.microsoft.com/en-us/learn/modules/populate-slowly-changing-dimensions-azure-synapse-analytics-pipelines/3-choose-between-dimension-types>

QUESTION 5

You have an Azure subscription that contains an Azure SQL database named DB1 and a storage account named storage1. The storage1 account contains a file named File1.txt. File1.txt contains the names of selected tables in DB1.

You need to use an Azure Synapse pipeline to copy data from the selected tables in DB1 to the files in storage1. The solution must meet the following requirements:

1.

The Copy activity in the pipeline must be parameterized to use the data in File1.txt to identify the source and destination of the copy.

2.

Copy activities must occur in parallel as often as possible.

Which two pipeline activities should you include in the pipeline? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

A. If Condition

B. ForEach

C. Lookup

D. Get Metadata

Correct Answer: BC

Lookup: This is a control activity that retrieves a dataset from any of the supported data sources and makes it available for use by subsequent activities in the pipeline. You can use a Lookup activity to read File1.txt from storage1 and store its content as an array variable1. ForEach: This is a control activity that iterates over a collection and executes specified activities in a loop. You can use a ForEach activity to loop over the array variable from the Lookup activity and pass each table name as a parameter to a Copy activity that copies data from DB1 to storage11.

[DP-203 VCE Dumps](#)

[DP-203 Study Guide](#)

[DP-203 Exam Questions](#)