



# 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.pass4itsure.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 have an Azure Synapse Analytics workspace.

You plan to deploy a lake database by using a database template in Azure Synapse.

Which two elements are included in the template? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. relationships
- B. data formats
- C. linked services
- D. table permissions
- E. table definitions

Correct Answer: AE

Reference: <https://learn.microsoft.com/en-us/azure/synapse-analytics/database-designer/overview-database-templates>  
<https://learn.microsoft.com/en-us/azure/synapse-analytics/database-designer/concepts-database-templates>

---

### QUESTION 2

HOTSPOT

You need to output files from Azure Data Factory.

Which file format should you use for each type of output? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:



## Answer Area

Columnar format:

	▼
Avro	
GZip	
Parquet	
TXT	

JSON with a timestamp:

	▼
Avro	
GZip	
Parquet	
TXT	

Correct Answer:



## Answer Area

Columnar format:

	▼
Avro	
GZip	
Parquet	
TXT	

JSON with a timestamp:

	▼
Avro	
GZip	
Parquet	
TXT	

Box 1: Parquet

Parquet stores data in columns, while Avro stores data in a row-based format. By their very nature, column-oriented data stores are optimized for read-heavy analytical workloads, while row-based databases are best for write-heavy transactional workloads.

Box 2: Avro

An Avro schema is created using JSON format.

AVRO supports timestamps.

Note: Azure Data Factory supports the following file formats (not GZip or TXT).

Avro format

Binary format

Delimited text format

Excel format

JSON format



ORC format

Parquet format

XML format

Reference:

<https://www.datanami.com/2018/05/16/big-data-file-formats-demystified>

---

### QUESTION 3

You have an Azure Synapse Analytics dedicated SQL pool named pool1.

You plan to implement a star schema in pool1 and create a new table named DimCustomer by using the following code.

```
CREATE TABLE dbo.[DimCustomer](
    [CustomerKey] int NOT NULL,
    [CustomerSourceID] [int] NOT NULL,
    [Title] [nvarchar](8) NULL,
    [FirstName] [nvarchar](50) NOT NULL,
    [MiddleName] [nvarchar](50) NULL,
    [LastName] [nvarchar](50) NOT NULL,
    [Suffix] [nvarchar](10) NULL,
    [CompanyName] [nvarchar](128) NULL,
    [SalesPerson] [nvarchar](256) NULL,
    [EmailAddress] [nvarchar](50) NULL,
    [Phone] [nvarchar](25) NULL,
    [InsertedDate] [datetime] NOT NULL,
    [ModifiedDate] [datetime] NOT NULL,
    [HashKey] [varchar](100) NOT NULL,
    [IsCurrentRow] [bit] NOT NULL
)
WITH
(
    DISTRIBUTION = REPLICATE,
    CLUSTERED COLUMNSTORE INDEX
);
GO
```

You need to ensure that DimCustomer has the necessary columns to support a Type 2 slowly changing dimension (SCD).

Which two columns should you add? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

A. [HistoricalSalesPerson] [nvarchar] (256) NOT NULL



- B. [EffectiveEndDate] [datetime] NOT NULL
- C. [PreviousModifiedDate] [datetime] NOT NULL
- D. [RowID] [bigint] NOT NULL
- E. [EffectiveStartDate] [datetime] NOT NULL

Correct Answer: BE

"For the SCD Type 2, we need to include three more attributes such as StartDate, EndDate and IsCurrent"

IsCurrentRow is already present! ... ;-)

CustomerKey (in reality is the RowID that many guys wants to add here),

effectiveEndDate will probably set to: 31.12.9999, (to justify the not null).

<https://www.sqlshack.com/implementing-slowly-changing-dimensions-scdds-in-data-warehouses/>

---

#### QUESTION 4

##### HOTSPOT

You have two Azure SQL databases named DB1 and DB2.

DB1 contains a table named Table 1. Table1 contains a timestamp column named LastModifiedOn. LastModifiedOn contains the timestamp of the most recent update for each individual row.

DB2 contains a table named Watermark. Watermark contains a single timestamp column named WatermarkValue.

You plan to create an Azure Data Factory pipeline that will incrementally upload into Azure Blob Storage all the rows in Table1 for which the LastModifiedOn column contains a timestamp newer than the most recent value of the

WatermarkValue column in Watermark.

You need to identify which activities to include in the pipeline. The solution must meet the following requirements:

Minimize the effort to author the pipeline.

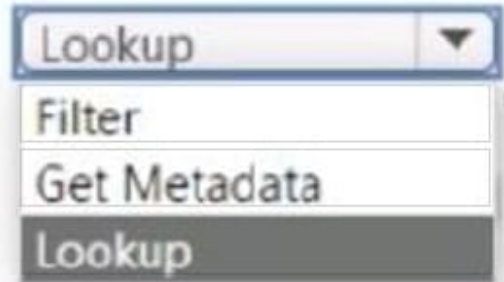
Ensure that the number of data integration units allocated to the upload operation can be controlled.

What should you identify? To answer, select the appropriate options in the answer area.

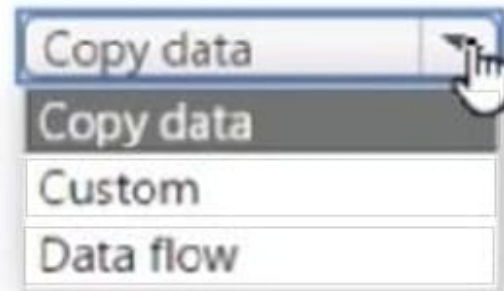
Hot Area:



To retrieve the watermark value, use:

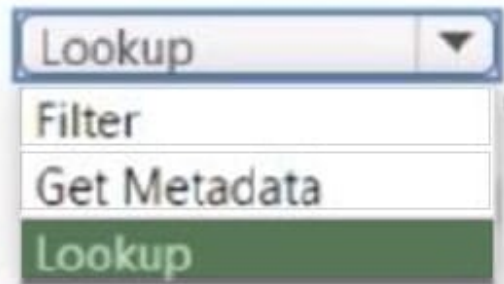


To perform the upload, use:

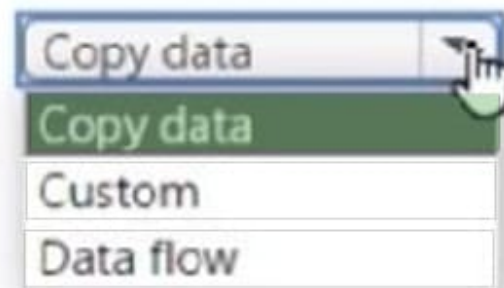


Correct Answer:

To retrieve the watermark value, use:



To perform the upload, use:



**QUESTION 5**

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 have an Azure Synapse Analytics dedicated SQL pool that contains a table named Table1.

You have files that are ingested and loaded into an Azure Data Lake Storage Gen2 container named container1.

You plan to insert data from the files in container1 into Table1 and transform the data. Each row of data in the files will produce one row in the serving layer of Table1.

You need to ensure that when the source data files are loaded to container1, the DateTime is stored as an additional column in Table1.

Solution: You use an Azure Synapse Analytics serverless SQL pool to create an external table that has an additional DateTime column.

Does this meet the goal?

A. Yes

B. No

Correct Answer: B

Instead use the derived column transformation to generate new columns in your data flow or to modify existing fields.

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

[DP-203 PDF Dumps](#)

[DP-203 VCE Dumps](#)

[DP-203 Brindumps](#)