

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

You are designing database for an Azure Synapse Analytics dedicated SQL pool to support workloads for detecting ecommerce transaction fraud.

Data will be combined from multiple ecommerce sites and can include sensitive financial information such as credit card numbers.

You need to recommend a solution that meets the following requirements:

Users must be able to identify potentially fraudulent transactions. Users must be able to use credit cards as a potential feature in models. Users must NOT be able to access the actual credit card numbers.

What should you include in the recommendation?

- A. Transparent Data Encryption (TDE)
- B. row-level security (RLS)
- C. column-level encryption
- D. Azure Active Directory (Azure AD) pass-through authentication

Correct Answer: C

Use Always Encrypted to secure the required columns. You can configure Always Encrypted for individual database columns containing your sensitive data. Always Encrypted is a feature designed to protect sensitive data, such as credit card numbers or national identification numbers (for example, U.S. social security numbers), stored in Azure SQL Database or SQL Server databases.

Reference: <https://docs.microsoft.com/en-us/sql/relational-databases/security/encryption/always-encrypted-database-engine>

QUESTION 2

HOTSPOT

You are building an Azure Synapse Analytics dedicated SQL pool that will contain a fact table for transactions from the first half of the year 2020.

You need to ensure that the table meets the following requirements:

1.
Minimizes the processing time to delete data that is older than 10 years
2.
Minimizes the I/O for queries that use year-to-date values

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:

Answer Area

```
CREATE TABLE [dbo].[FactTransaction]
(
    [TransactionTypeID] int NOT NULL
,   [TransactionDateID] int NOT NULL
,   [CustomerID] int NOT NULL
,   [RecipientID] int NOT NULL
,   [Amount] money NOT NU::
)
WITH
```

	▼
CLUSTERED COLUMNSTORE INDEX	
DISTRIBUTION	
PARTITION	
TRUNCATE_TARGET	

	▼
[TransactionDateID]	
[TransactionDateID], [TransactionTypeID]	
HASH([TransactionTypeID])	
ROUND ROBIN	

RANGE RIGHT FOR VALUES

```
(20200101,20200201,20200301,20200401,20200501,20200601)
```

Correct Answer:

Answer Area

```
CREATE TABLE [dbo].[FactTransaction]
(
    [TransactionTypeID] int NOT NULL
,   [TransactionDateID] int NOT NULL
,   [CustomerID] int NOT NULL
,   [RecipientID] int NOT NULL
,   [Amount] money NOT NU::
)
WITH
```

	▼
CLUSTERED COLUMNSTORE INDEX	
DISTRIBUTION	
PARTITION	
TRUNCATE_TARGET	

	▼
[TransactionDateID]	
[TransactionDateID], [TransactionTypeID]	
HASH([TransactionTypeID])	
ROUND ROBIN	

RANGE RIGHT FOR VALUES

```
(20200101, 20200201, 20200301, 20200401, 20200501, 20200601)
```

Box 1: PARTITION

RANGE RIGHT FOR VALUES is used with PARTITION.

Part 2: [TransactionDateID]

Partition on the date column.

Example: Creating a RANGE RIGHT partition function on a datetime column

The following partition function partitions a table or index into 12 partitions, one for each month of a year's worth of values in a datetime column.

```
CREATE PARTITION FUNCTION [myDateRangePF1] (datetime)
```

```
AS RANGE RIGHT FOR VALUES ('20030201', '20030301', '20030401', '20030501', '20030601',
'20030701', '20030801', '20030901', '20031001', '20031101', '20031201');
```

Reference: <https://docs.microsoft.com/en-us/sql/t-sql/statements/create-partition-function-transact-sql>

QUESTION 3

You have an enterprise data warehouse in Azure Synapse Analytics.

Using PolyBase, you create an external table named [Ext].[Items] to query Parquet files stored in Azure Data Lake Storage Gen2 without importing the data to the data warehouse.

The external table has three columns.

You discover that the Parquet files have a fourth column named ItemID.

Which command should you run to add the ItemID column to the external table?

- A.

```
ALTER EXTERNAL TABLE [Ext].[Items]
  ADD [ItemID] int;
```
- B.

```
DROP EXTERNAL FILE FORMAT parquetfile1;
CREATE EXTERNAL FILE FORMAT parquetfile1
WITH (
  FORMAT_TYPE = PARQUET,
  DATA_COMPRESSION = 'org.apache.hadoop.io.compress.SnappyCodec'
);
```
- C.

```
DROP EXTERNAL TABLE [Ext].[Items]
CREATE EXTERNAL TABLE [Ext].[Items]
([ItemID] [int] NULL,
 [ItemName] nvarchar(50) NULL,
 [ItemType] nvarchar(20) NULL,
 [ItemDescription] nvarchar(250))
WITH
(
  LOCATION= '/Items/',
  DATA_SOURCE = AzureDataLakeStore,
  FILE_FORMAT = PARQUET,
  REJECT_TYPE = VALUE,
  REJECT_VALUE = 0
);
```
- D.

```
ALTER TABLE [Ext].[Items]
  ADD [ItemID] int;
```

A. Option A

B. Option B

C. Option C

D. Option D

Correct Answer: C

Incorrect Answers:

A, D: Only these Data Definition Language (DDL) statements are allowed on external tables:

1.

CREATE TABLE and DROP TABLE

2.

CREATE STATISTICS and DROP STATISTICS

3.

CREATE VIEW and DROP VIEW

Reference: <https://docs.microsoft.com/en-us/sql/t-sql/statements/create-external-table-transact-sql>

QUESTION 4

HOTSPOT

You are building an Azure Data Factory solution to process data received from Azure Event Hubs, and then ingested into an Azure Data Lake Storage Gen2 container.

The data will be ingested every five minutes from devices into JSON files. The files have the following naming pattern.

`/{deviceType}/in/{YYYY}/{MM}/{DD}/{HH}/{deviceID}_{YYYY}{MM}{DD}HH}{mm}.json`

You need to prepare the data for batch data processing so that there is one dataset per hour per deviceType. The solution must minimize read times.

How should you configure the sink for the copy activity? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

Parameter:

	▼
@pipeline(),TriggerTime	
@pipeline(),TriggerType	
@trigger().outputs.windowStartTime	
@trigger().startTime	

Naming pattern:

	▼
/{deviceId}/out/{YYYY}/{MM}/{DD}/{HH}.json	
/{YYYY}/{MM}/{DD}/{deviceType}.json	
/{YYYY}/{MM}/{DD}/{HH}.json	
/{YYYY}/{MM}/{DD}/{HH}_{deviceType}.json	

Copy behavior:

	▼
Add dynamic content	
Flatten hierarchy	
Merge files	

Correct Answer:

Answer Area

Parameter:

	▼
@pipeline(),TriggerTime	
@pipeline(),TriggerType	
@trigger().outputs.windowStartTime	
@trigger().startTime	

Naming pattern:

	▼
/{deviceId}/out/{YYYY}/{MM}/{DD}/{HH}.json	
/{YYYY}/{MM}/{DD}/{deviceType}.json	
/{YYYY}/{MM}/{DD}/{HH}.json	
/{YYYY}/{MM}/{DD}/{HH}_{deviceType}.json	

Copy behavior:

	▼
Add dynamic content	
Flatten hierarchy	
Merge files	

Box 1: @trigger().startTime

startTime: A date-time value. For basic schedules, the value of the startTime property applies to the first occurrence. For complex schedules, the trigger starts no sooner than the specified startTime value.

Box 2: /{YYYY}/{MM}/{DD}/{HH}_{deviceType}.json

One dataset per hour per deviceType.

Box 3: Flatten hierarchy

- FlattenHierarchy: All files from the source folder are in the first level of the target folder. The target files have autogenerated names.

Reference: <https://docs.microsoft.com/en-us/azure/data-factory/concepts-pipeline-execution-triggers>
<https://docs.microsoft.com/en-us/azure/data-factory/connector-file-system>

QUESTION 5

You have two Azure Data Factory instances named ADFdev and ADFprod. ADFdev connects to an Azure DevOps Git repository.

You publish changes from the main branch of the Git repository to ADFdev.

You need to deploy the artifacts from ADFdev to ADFprod.

What should you do first?

- A. From ADFdev, modify the Git configuration.
- B. From ADFdev, create a linked service.
- C. From Azure DevOps, create a release pipeline.
- D. From Azure DevOps, update the main branch.

Correct Answer: C

In Azure Data Factory, continuous integration and delivery (CI/CD) means moving Data Factory pipelines from one environment (development, test, production) to another.

Note:

The following is a guide for setting up an Azure Pipelines release that automates the deployment of a data factory to multiple environments.

1.

In Azure DevOps, open the project that's configured with your data factory.

2.

On the left side of the page, select Pipelines, and then select Releases.

3.

Select New pipeline, or, if you have existing pipelines, select New and then New release pipeline.

4.

In the Stage name box, enter the name of your environment.

5.

Select Add artifact, and then select the git repository configured with your development data factory. Select the publish branch of the repository for the Default branch. By default, this publish branch is adf_publish.

6.

Select the Empty job template.

Reference: <https://docs.microsoft.com/en-us/azure/data-factory/continuous-integration-deployment>