

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 plan to create a table in an Azure Synapse Analytics dedicated SQL pool.

Data in the table will be retained for five years. Once a year, data that is older than five years will be deleted.

You need to ensure that the data is distributed evenly across partitions. The solution must minimize the amount of time required to delete old data.

How should you complete the Transact-SQL statement? To answer, drag the appropriate values to the correct targets. Each value may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to

view content.

NOTE: Each correct selection is worth one point.

Select and Place:

Values	Answer Area
CustomerKey	<pre>CREATE TABLE [dbo].[FactSales] ([ProductKey] int NOT NULL , [OrderDateKey] int NOT NULL , [CustomerKey] int NOT NULL , [SalesOrderNumber] nvarchar (20) NOT NULL , [OrderQuantity] smallint NOT NULL , [UnitPrice] money NOT NULL) WITH (CLUSTERED COLUMNSTORE INDEX , DISTRIBUTION = [Value] ([ProductKey]) , PARTITION ([[Value]] RANGE RIGHT FOR VALUES (20170101,20180101,20190101,20200101,20210101)))</pre>
HASH	
ROUND_ROBIN	
REPLICATE	
OrderDateKey	
SalesOrderNumber	

Correct Answer:

Values	Answer Area
CustomerKey	<pre>CREATE TABLE [dbo].[FactSales] ([ProductKey] int NOT NULL , [OrderDateKey] int NOT NULL , [CustomerKey] int NOT NULL , [SalesOrderNumber] nvarchar (20) NOT NULL , [OrderQuantity] smallint NOT NULL , [UnitPrice] money NOT NULL) WITH (CLUSTERED COLUMNSTORE INDEX , DISTRIBUTION = HASH ([ProductKey]) , PARTITION ([OrderDateKey]] RANGE RIGHT FOR VALUES (20170101, 20180101, 20190101, 20200101, 20210101)))</pre>
ROUND_ROBIN	
REPLICATE	
SalesOrderNumber	

Box 1: HASH

Box 2: OrderDateKey

In most cases, table partitions are created on a date column.

A way to eliminate rollbacks is to use Metadata Only operations like partition switching for data management. For example, rather than execute a DELETE statement to delete all rows in a table where the order_date was in October of 2001,

you could partition your data early. Then you can switch out the partition with data for an empty partition from another table.

QUESTION 2

You have an Azure Synapse Analytics dedicated SQL pool.

You plan to create a fact table named Table1 that will contain a clustered columnstore index.

You need to optimize data compression and query performance for Table1.

What is the minimum number of rows that Table1 should contain before you create partitions?

- A. 100,000
- B. 600,000
- C. 1 million

D. 60 million

Correct Answer: A

Target size for rowgroups

For best query performance, the goal is to maximize the number of rows per rowgroup in a columnstore index. A rowgroup can have a maximum of 1,048,576 rows.

It's okay to not have the maximum number of rows per rowgroup. Columnstore indexes achieve good performance when rowgroups have at least 100,000 rows.

Reference:

<https://learn.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-memory-optimizations-for-columnstore-compression>

QUESTION 3

You have an Azure Synapse Analytics dedicated SQL pool named SA1 that contains a table named Table1.

You need to identify tables that have a high percentage of deleted rows.

What should you run?

- A. `sys.pdw_nodes_column_store_segments`
- B. `sys.dm_db_column_store_row_group_operational_stats`
- C. `sys.pdw_nodes_column_store_row_groups`
- D. `sys.dm_db_column_store_row_group_physical_stats`

Correct Answer: C

Use `sys.pdw_nodes_column_store_row_groups` to determine which row groups have a high percentage of deleted rows and should be rebuilt.

Note: `sys.pdw_nodes_column_store_row_groups` provides clustered columnstore index information on a per-segment basis to help the administrator make system management decisions in Azure Synapse Analytics.

`sys.pdw_nodes_column_store_row_groups` has a column for the total number of rows physically stored (including those marked as deleted) and a column for the number of rows marked as deleted.

Incorrect:

Not A: You can join `sys.pdw_nodes_column_store_segments` with other system tables to determine the number of columnstore segments per logical table.

Not B: Use `sys.dm_db_column_store_row_group_operational_stats` to track the length of time a user query must wait to read or write to a compressed rowgroup or partition of a columnstore index, and identify rowgroups that are encountering

significant I/O activity or hot spots.

QUESTION 4

You are designing an Azure Databricks interactive cluster. The cluster will be used infrequently and will be configured for auto-termination. You need to use that the cluster configuration is retained indefinitely after the cluster is terminated. The solution must minimize costs. What should you do?

- A. Pin the cluster.
- B. Create an Azure runbook that starts the cluster every 90 days.
- C. Terminate the cluster manually when processing completes.
- D. Clone the cluster after it is terminated.

Correct Answer: A

Azure Databricks retains cluster configuration information for up to 70 all-purpose clusters terminated in the last 30 days and up to 30 job clusters recently terminated by the job scheduler. To keep an all-purpose cluster configuration even after it has been terminated for more than 30 days, an administrator can pin a cluster to the cluster list.

Reference: <https://docs.microsoft.com/en-us/azure/databricks/clusters/>

QUESTION 5

HOTSPOT

You use Azure Data Lake Storage Gen2 to store data that data scientists and data engineers will query by using Azure Databricks interactive notebooks. Users will have access only to the Data Lake Storage folders that relate to the projects

on which they work.

You need to recommend which authentication methods to use for Databricks and Data Lake Storage to provide the users with the appropriate access. The solution must minimize administrative effort and development effort.

Which authentication method should you recommend for each Azure service? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

Databricks:

	▼
Azure Active Directory credential passthrough	
Azure Key Vault secrets	
Personal access tokens	

Data Lake Storage:

	▼
Azure Active Directory credential passthrough	
Shared access keys	
Shared access signatures	

Correct Answer:

Answer Area

Databricks:

	▼
Azure Active Directory credential passthrough	
Azure Key Vault secrets	
Personal access tokens	

Data Lake Storage:

	▼
Azure Active Directory credential passthrough	
Shared access keys	
Shared access signatures	

Box 1: Personal access tokens

You can use storage shared access signatures (SAS) to access an Azure Data Lake Storage Gen2 storage account directly. With SAS, you can restrict access to a storage account using temporary tokens with fine-grained access control.

You can add multiple storage accounts and configure respective SAS token providers in the same Spark session.

Box 2: Azure Active Directory credential passthrough

You can authenticate automatically to Azure Data Lake Storage Gen1 (ADLS Gen1) and Azure Data Lake Storage Gen2 (ADLS Gen2) from Azure Databricks clusters using the same Azure Active Directory (Azure AD) identity that you use to

log into Azure Databricks. When you enable your cluster for Azure Data Lake Storage credential passthrough, commands that you run on that cluster can read and write data in Azure Data Lake Storage without requiring you to configure

service principal credentials for access to storage.

After configuring Azure Data Lake Storage credential passthrough and creating storage containers, you can access data directly in Azure Data Lake Storage Gen1 using an `adl://` path and Azure Data Lake Storage Gen2 using an `abfss://`

path:

Reference:

<https://docs.microsoft.com/en-us/azure/databricks/data/data-sources/azure/adls-gen2/azure-datalake-gen2-sas-access>

<https://docs.microsoft.com/en-us/azure/databricks/security/credential-passthrough/adls-passthrough>

[Latest DP-203 Dumps](#)

[DP-203 VCE Dumps](#)

[DP-203 Practice Test](#)