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.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

HOTSPOT

You have a SQL pool in Azure Synapse.

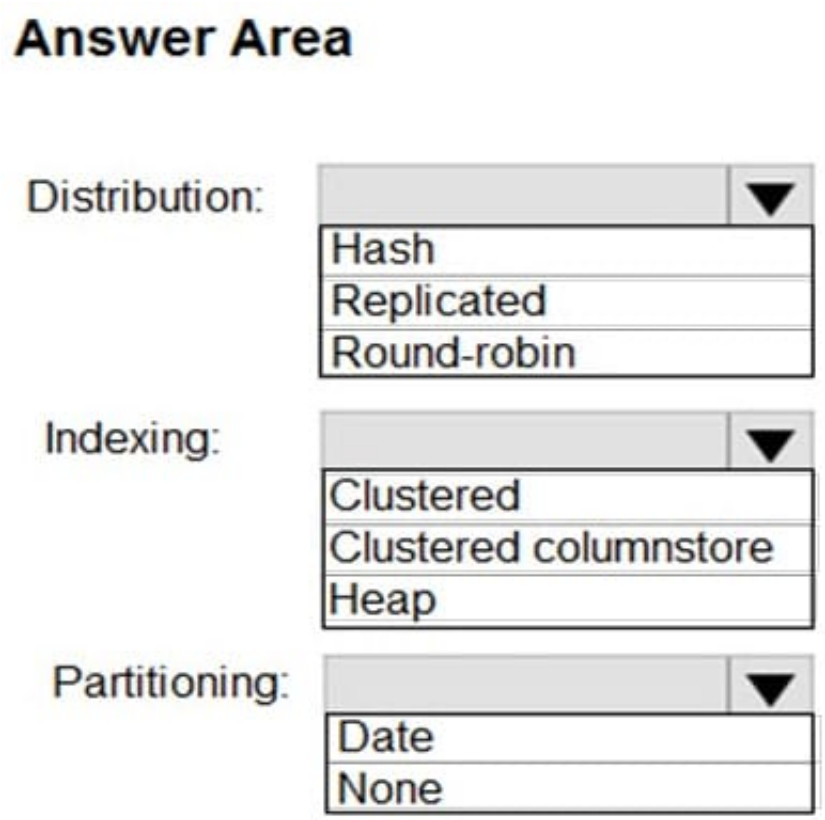
You plan to load data from Azure Blob storage to a staging table. Approximately 1 million rows of data will be loaded daily. The table will be truncated before each daily load.

You need to create the staging table. The solution must minimize how long it takes to load the data to the staging table.

How should you configure the table? To answer, select the appropriate options in the answer area.

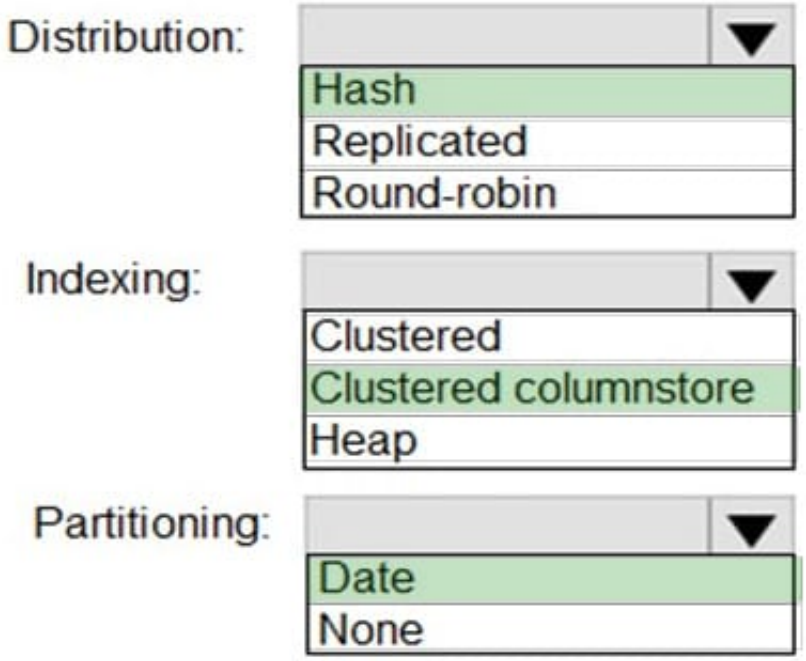
NOTE: Each correct selection is worth one point.

Hot Area:



Correct Answer:

Answer Area



Box 1: Hash

Hash-distributed tables improve query performance on large fact tables. They can have very large numbers of rows and still achieve high performance.

Incorrect Answers:

Round-robin tables are useful for improving loading speed.

Box 2: Clustered columnstore

When creating partitions on clustered columnstore tables, it is important to consider how many rows belong to each partition. For optimal compression and performance of clustered columnstore tables, a minimum of 1 million rows per

distribution and partition is needed.

Box 3: Date

Table partitions enable you to divide your data into smaller groups of data. In most cases, table partitions are created on a date column.

Partition switching can be used to quickly remove or replace a section of a table.

Reference:

https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-partition

https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-distribute

QUESTION 2

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 3

You have a SQL pool in Azure Synapse.

A user reports that queries against the pool take longer than expected to complete.

You need to add monitoring to the underlying storage to help diagnose the issue.

Which two metrics should you monitor? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

A. Cache used percentage

B. DWU Limit

C. Snapshot Storage Size

D. Active queries

E. Cache hit percentage

Correct Answer: AE

A: Cache used is the sum of all bytes in the local SSD cache across all nodes and cache capacity is the sum of the storage capacity of the local SSD cache across all nodes.

E: Cache hits is the sum of all columnstore segments hits in the local SSD cache and cache miss is the columnstore segments misses in the local SSD cache summed across all nodes

Reference: https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-concept-resource-utilization-query-activity

---

QUESTION 4

DRAG DROP

You have an Azure data factory.

You need to ensure that pipeline-run data is retained for 120 days. The solution must ensure that you can query the data by using the Kusto query language.

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

NOTE: More than one order of answer choices is correct. You will receive credit for any of the correct orders you select.

Select and Place:

Actions	Answer Area
Select the PipelineRuns category.	
Create a Log Analytics workspace that has Data Retention set to 120 days.	
Stream to an Azure event hub.	
Create an Azure Storage account that has a lifecycle policy.	
From the Azure portal, add a diagnostic setting.	
Send the data to a Log Analytics workspace.	
Select the TriggerRuns category.	

Correct Answer:

Actions	Answer Area
Select the PipelineRuns category.	Create an Azure Storage account that has a lifecycle policy.
	Create a Log Analytics workspace that has Data Retention set to 120 days.
Stream to an Azure event hub.	From the Azure portal, add a diagnostic setting.
	Send the data to a Log Analytics workspace.
Select the TriggerRuns category.	

Step 1: Create an Azure Storage account that has a lifecycle policy

To automate common data management tasks, Microsoft created a solution based on Azure Data Factory. The service, Data Lifecycle Management, makes frequently accessed data available and archives or purges other data according to

retention policies. Teams across the company use the service to reduce storage costs, improve app performance, and comply with data retention policies.

Step 2: Create a Log Analytics workspace that has Data Retention set to 120 days.

Data Factory stores pipeline-run data for only 45 days. Use Azure Monitor if you want to keep that data for a longer time. With Monitor, you can route diagnostic logs for analysis to multiple different targets, such as a Storage Account: Save

your diagnostic logs to a storage account for auditing or manual inspection. You can use the diagnostic settings to specify the retention time in days.

Step 3: From Azure Portal, add a diagnostic setting.

Step 4: Send the data to a log Analytics workspace,

Event Hub: A pipeline that transfers events from services to Azure Data Explorer.

Keeping Azure Data Factory metrics and pipeline-run data.

Configure diagnostic settings and workspace.

Create or add diagnostic settings for your data factory.

In the portal, go to Monitor. Select Settings > Diagnostic settings.

Select the data factory for which you want to set a diagnostic setting.

If no settings exist on the selected data factory, you\\'re prompted to create a setting. Select Turn on diagnostics.

Give your setting a name, select Send to Log Analytics, and then select a workspace from Log Analytics Workspace.

Select Save.

Reference:

https://docs.microsoft.com/en-us/azure/data-factory/monitor-using-azure-monitor

QUESTION 5

You are designing a sales transactions table in an Azure Synapse Analytics dedicated SQL pool. The table will contains approximately 60 million rows per month and will be partitioned by month. The table will use a clustered column store index and round-robin distribution.

Approximately how many rows will there be for each combination of distribution and partition?

A. 1 million

B. 5 million

C. 20 million

D. 60 million

Correct Answer: D

https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-partition