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.geekcert.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 dedicated SQL pool.

You run PDW_SHOWSPACEUSED(dbo,FactInternetSales\\'); and get the results shown in the following table.

ROWS	RESERVED_SPACE	DATA_SPACE	INDEX_SPACE	UNUSED_SPACE	PDW_NODE_ID	DISTRIBUTION_ID
694	2776	616	48	2112	1	1
407	2704	576	48	2080	1	2
53	2376	512	16	1848	1	3
58	2376	512	16	1848	1	4
168	2632	528	32	2072	1	5
195	2696	536	32	2128	1	6
5995	3464	1424	32	2008	1	7
0	2232	496	0	1736	1	8
264	2576	544	40	1992	1	9
3008	3016	960	32	2024	1	10
...	...	...	...	...	...	...
1550	2832	752	48	2032	1	50
1238	2832	696	40	2096	1	51
192	2632	528	32	2072	1	52
1127	2768	680	48	2040	1	53
1244	3032	704	64	2264	1	54
409	2632	568	32	2032	1	55
0	2232	496	0	1736	1	56
1437	2832	728	40	2064	1	57
0	2232	496	0	1736	1	58
384	2632	560	32	2040	1	59
225	2768	544	40	2194	1	60

Which statement accurately describes the dbo,FactInternetSales table?

A. The table contains less than 1,000 rows.

B. All distribution contain data.

C. The table is skewed.

D. The table uses round-robin distribution.

Correct Answer: C

Data skew means the data is not distributed evenly across the distributions.

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

QUESTION 2

You have the following Azure Data Factory pipelines:

1.

Ingest Data from System1

2.

Ingest Data from System2

3.

Populate Dimensions

4.

Populate Facts

Ingest Data from System1 and Ingest Data from System2 have no dependencies. Populate Dimensions must execute after Ingest Data from System1 and Ingest Data from System2. Populate Facts must execute after Populate Dimensions pipeline. All the pipelines must execute every eight hours.

What should you do to schedule the pipelines for execution?

A. Add an event trigger to all four pipelines.

B. Add a schedule trigger to all four pipelines.

C. Create a patient pipeline that contains the four pipelines and use a schedule trigger.

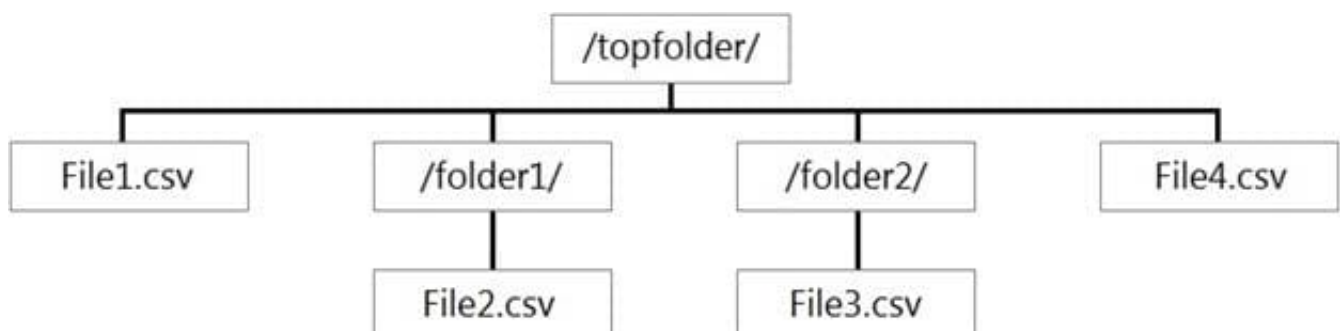D. Create a patient pipeline that contains the four pipelines and use an event trigger.

Correct Answer: C

Schedule trigger: A trigger that invokes a pipeline on a wall-clock schedule.

Reference: https://docs.microsoft.com/en-us/azure/data-factory/concepts-pipeline-execution-triggers

QUESTION 3

You have files and folders in Azure Data Lake Storage Gen2 for an Azure Synapse workspace as shown in the following exhibit.

You create an external table named ExtTable that has LOCATION=\\'/topfolder/\\'.

When you query ExtTable by using an Azure Synapse Analytics serverless SQL pool, which files are returned?

A. File2.csv and File3.csv only

B. File1.csv and File4.csv only

C. File1.csv, File2.csv, File3.csv, and File4.csv

D. File1.csv only

Correct Answer: B

To run a T-SQL query over a set of files within a folder or set of folders while treating them as a single entity or rowset, provide a path to a folder or a pattern (using wildcards) over a set of files or folders.

Reference: https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/query-data-storage#query-multiple-files-or-folders

QUESTION 4

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 5

HOTSPOT

You have an Azure subscription that contains an Azure Databricks workspace named databricks1 and an Azure Synapse Analytics workspace named synapse1. The synapse1 workspace contains an Apache Spark pool named pool1.

You need to share an Apache Hive catalog of pool1 with databricks1.

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

NOTE: Each correct selection is worth one point.

Hot Area:

From synapse1, create a linked service to:	
	Azure Cosmos DB
	Azure Data Lake Storage Gen2
	Azure SQL Database

Configure pool1 to use the linked service as:	
	An Azure Purview account
	A Hive metastore
	A managed Hive metastore service

Correct Answer:

From synapse1, create a linked service to:	
	Azure Cosmos DB
	Azure Data Lake Storage Gen2
	Azure SQL Database

Configure pool1 to use the linked service as:	
	An Azure Purview account
	A Hive metastore
	A managed Hive metastore service

Box 1: Azure SQL Database

Use external Hive Metastore for Synapse Spark Pool Azure Synapse Analytics allows Apache Spark pools in the same workspace to share a managed HMS (Hive Metastore) compatible metastore as their catalog.

Set up linked service to Hive Metastore

Follow below steps to set up a linked service to the external Hive Metastore in Synapse workspace.

Open Synapse Studio, go to Manage > Linked services at left, click New to create a new linked service.

Set up Hive Metastore linked service

Choose Azure SQL Database or Azure Database for MySQL based on your database type, click Continue.

Provide Name of the linked service. Record the name of the linked service, this info will be used to configure Spark shortly.

You can either select Azure SQL Database/Azure Database for MySQL for the external Hive Metastore from Azure subscription list, or enter the info manually.

Provide User name and Password to set up the connection.

Test connection to verify the username and password.

Click Create to create the linked service.

Box 2: A Hive Metastore