⚙ Custom View Settings

Question #40   *Topic 2*

HOTSPOT -

You configure version control for an Azure Data Factory instance as shown in the following exhibit.

Connections	Git repository
🔗 Linked services	Git repository information associated with your data factory. CI/CD best practices ↗
⊞ Integration runtimes	⚙ Setting 🔗 Disconnect
Source control	Repository type Azure DevOps Git
◆ Git configuration	
⊡ ARM template	Azure DevOps Account CONTOSO
@ Parameterization template	Project name Data
Author	Repository name dwh_batchetl
⚡ Triggers	
[@] Global parameters	Collaboration branch main
Security	Publish branch adf_publish
🔏 Customer managed key	
☁ Managed private endpoints	Root folder /

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

Azure Resource Manager (ARM) templates for the pipeline assets are stored in **[answer choice]**

▼
/
adf_publish
main
Parameterization template

A Data Factory Azure Resource Manager (ARM) template named `contososales` can be found in **[answer choice]**

▼
/
/contososales
/dwh_batchetl/adf_publish/contososales
/main

HOTSPOT -

You are designing an Azure Stream Analytics solution that receives instant messaging data from an Azure Event Hub.

You need to ensure that the output from the Stream Analytics job counts the number of messages per time zone every 15 seconds.

How should you complete the Stream Analytics query? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

Select TimeZone, count (*) AS MessageCount

FROM MessageStream [▼] CreatedAt

| LAST |
| OVER |
| SYSTEM.TIMESTAMP() |
| TIMESTAMP BY |

GROUP BY TimeZone, [▼] (second,15)

| HOPPINGWINDOW |
| SESSIONWINDOW |
| SLIDINGWINDOW |
| TUMBLINGWINDOW |

HOTSPOT -

You have an Azure Data Factory instance named ADF1 and two Azure Synapse Analytics workspaces named WS1 and WS2.

ADF1 contains the following pipelines:

☞ P1: Uses a copy activity to copy data from a nonpartitioned table in a dedicated SQL pool of WS1 to an Azure Data Lake Storage Gen2 account

☞ P2: Uses a copy activity to copy data from text-delimited files in an Azure Data Lake Storage Gen2 account to a nonpartitioned table in a dedicated SQL pool of WS2
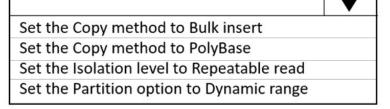
You need to configure P1 and P2 to maximize parallelism and performance.

Which dataset settings should you configure for the copy activity if each pipeline? To answer, select the appropriate options in the answer area.

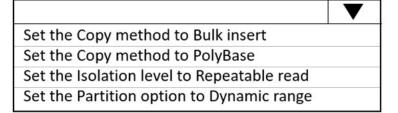NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

P1:

Set the Copy method to Bulk insert
Set the Copy method to PolyBase
Set the Isolation level to Repeatable read
Set the Partition option to Dynamic range

P2:

Set the Copy method to Bulk insert
Set the Copy method to PolyBase
Set the Isolation level to Repeatable read
Set the Partition option to Dynamic range

HOTSPOT -

You have an Azure Storage account that generates 200,000 new files daily. The file names have a format of {YYYY}/{MM}/{DD}/{HH}/{CustomerID}.csv.

You need to design an Azure Data Factory solution that will load new data from the storage account to an Azure Data Lake once hourly. The solution must minimize load times and costs.

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

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

Load methodology:

| Full Load |
| Incremental Load |
| Load individual files as they arrive |

Trigger:

| Fixed schedule |
| New file |
| Tumbling window |

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 plan to create an Azure Databricks workspace that has a tiered structure. The workspace will contain the following three workloads:

☞ A workload for data engineers who will use Python and SQL.

☞ A workload for jobs that will run notebooks that use Python, Scala, and SQL.

☞ A workload that data scientists will use to perform ad hoc analysis in Scala and R.

The enterprise architecture team at your company identifies the following standards for Databricks environments:

☞ The data engineers must share a cluster.

☞ The job cluster will be managed by using a request process whereby data scientists and data engineers provide packaged notebooks for deployment to the cluster.

☞ All the data scientists must be assigned their own cluster that terminates automatically after 120 minutes of inactivity. Currently, there are three data scientists.

You need to create the Databricks clusters for the workloads.

Solution: You create a Standard cluster for each data scientist, a Standard cluster for the data engineers, and a High Concurrency cluster for the jobs.

Does this meet the goal?

A. Yes

B. No