

Microsoft.DP-203.v2025-07-14.q173

Exam Code:	DP-203
Exam Name:	Data Engineering on Microsoft Azure
Certification Provider:	Microsoft
Free Question Number:	173
Version:	v2025-07-14
# of views:	104
# of Questions views:	1730
https://www.freepdfdumps.com/Microsoft.DP-203.v2025-07-14.q173.html	

NEW QUESTION: 1

You need to implement versioned changes to the integration pipelines. The solution must meet the data integration requirements.

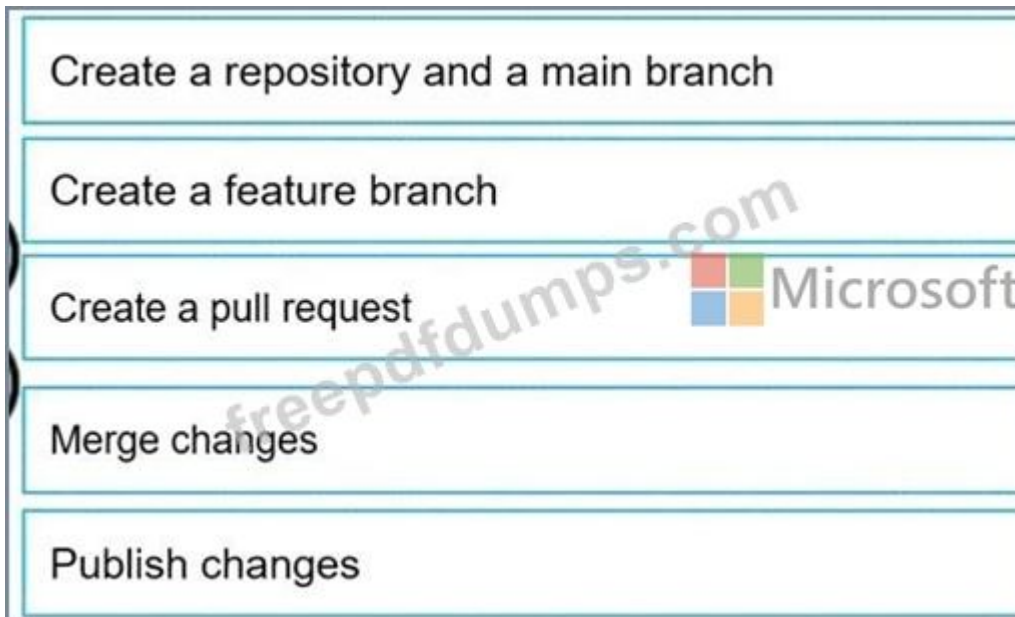
In which order should you perform the actions? To answer, move all actions from the list of actions to the answer area and arrange them in the correct order.

The screenshot shows a drag-and-drop interface. On the left, under the heading "Actions", there is a list of five items: "Publish changes.", "Create a feature branch.", "Merge changes.", "Create a repository and a main branch.", and "Create a pull request.". On the right, under the heading "Answer Area", there is an empty space with a dashed red border. A Microsoft logo is visible in the background, and a watermark "freepdfdumps.com" is overlaid on the image. Two circular arrows indicate the direction of movement: one pointing right from the actions list to the answer area, and one pointing left from the answer area back to the actions list.

Answer:

The screenshot shows the same interface as above, but with the actions moved to the answer area in the correct order. The "Answer Area" now contains five items: "Create a repository and a main branch.", "Create a feature branch.", "Create a pull request.", "Merge changes.", and "Publish changes.". The "Actions" list on the left is empty. A Microsoft logo is visible in the background, and a watermark "freepdfdumps.com" is overlaid on the image. Two circular arrows indicate the direction of movement: one pointing right from the actions list to the answer area, and one pointing left from the answer area back to the actions list.

Explanation:



Scenario: Identify a process to ensure that changes to the ingestion and transformation activities can be version-controlled and developed independently by multiple data engineers.

Step 1: Create a repository and a main branch

You need a Git repository in Azure Pipelines, TFS, or GitHub with your app.

Step 2: Create a feature branch

Step 3: Create a pull request

Step 4: Merge changes

Merge feature branches into the main branch using pull requests.

Step 5: Publish changes

Reference:

<https://docs.microsoft.com/en-us/azure/devops/pipelines/repos/pipeline-options-for-git>

NEW QUESTION: 2

You have an Azure Data Factory pipeline that performs an incremental load of source data to an Azure Data Lake Storage Gen2 account.

Data to be loaded is identified by a column named LastUpdatedDate in the source table.

You plan to execute the pipeline every four hours.

You need to ensure that the pipeline execution meets the following requirements:

- * Automatically retries the execution when the pipeline run fails due to concurrency or throttling limits.
- * Supports backfilling existing data in the table.

Which type of trigger should you use?

- A. Storage event
- B. on-demand
- C. schedule
- D. tumbling window

Answer: ([SHOW ANSWER](#))

In case of pipeline failures, tumbling window trigger can retry the execution of the referenced pipeline automatically, using the same input parameters, without the user intervention. This can be specified using the property "retryPolicy" in the trigger definition.

Reference:

<https://docs.microsoft.com/en-us/azure/data-factory/how-to-create-tumbling-window-trigger>

NEW QUESTION: 3

You are designing an application that will store petabytes of medical imaging data. When the data is first created, the data will be accessed frequently during the first week. After one month, the data must be accessible within 30 seconds, but files will be accessed infrequently. After one year, the data will be accessed infrequently but must be accessible within five minutes.

You need to select a storage strategy for the data. The solution must minimize costs.

Which storage tier should you use for each time frame? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

The screenshot shows a Microsoft exam question interface. On the left, there is the Microsoft logo. The question text is as follows: "First week:", "After one month:", and "After one year:". To the right of each text label is a dropdown menu. Each dropdown menu is currently open, showing three options: "Archive", "Cool", and "Hot". A watermark "freepdfmups.com" is visible diagonally across the interface.

Answer:

First week: ▼

Archive
Cool
Hot

After one month: ▼

Archive
Cool
Hot

After one year: ▼

Archive
Cool
Hot

Explanation:

The screenshot shows a Microsoft OneDrive interface with three dropdown menus for selecting storage tiers. The first dropdown, labeled 'First week:', has 'Hot' selected. The second dropdown, labeled 'After one month:', has 'Cool' selected. The third dropdown, labeled 'After one year:', has 'Cool' selected. The Microsoft logo is visible in the top left corner of the interface.

First week: Hot

Hot - Optimized for storing data that is accessed frequently.

After one month: Cool

Cool - Optimized for storing data that is infrequently accessed and stored for at least 30 days.

After one year: Cool

NEW QUESTION: 4

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 High Concurrency cluster for each data scientist, a High Concurrency cluster for the data engineers, and a Standard cluster for the jobs.

Does this meet the goal?

A. Yes

B. No

Answer: B (LEAVE A REPLY)

Need a High Concurrency cluster for the jobs.

Standard clusters are recommended for a single user. Standard can run workloads developed in any language:

Python, R, Scala, and SQL.

A high concurrency cluster is a managed cloud resource. The key benefits of high concurrency clusters are that they provide Apache Spark-native fine-grained sharing for maximum resource utilization and minimum query latencies.

Reference:

<https://docs.azuredatabricks.net/clusters/configure.html>

NEW QUESTION: 5

You are batch loading a table in an Azure Synapse Analytics dedicated SQL pool.

You need to load data from a staging table to the target table. The solution must ensure that if an error occurs while loading the data to the target table, all the inserts in that batch are undone.

How should you complete the Transact-SQL code? 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.

Values

BEGIN DISTRIBUTED TRANSACTION

BEGIN TRAN

COMMIT TRAN

ROLLBACK TRAN

SET RESULT_SET_CACHING ON

Answer Area

```

BEGIN TRY
    INSERT INTO dbo.Table1 (col1, col2, col3)
    SELECT col1, col2, col3 FROM stage.Table1;
END TRY
BEGIN CATCH
    IF @@TRANCOUNT > 0
    BEGIN
        ;
    END
END CATCH;
IF @@TRANCOUNT > 0
BEGIN
    COMMIT TRAN;
END

```

Microsoft

Answer:

Values

BEGIN DISTRIBUTED TRANSACTION

BEGIN TRAN

COMMIT TRAN

ROLLBACK TRAN

SET RESULT_SET_CACHING ON

Answer Area

```

BEGIN TRY
    INSERT INTO dbo.Table1 (col1, col2, col3)
    SELECT col1, col2, col3 FROM stage.Table1;
END TRY
BEGIN CATCH
    IF @@TRANCOUNT > 0
    BEGIN
        ;
    END
END CATCH;
IF @@TRANCOUNT > 0
BEGIN
    COMMIT TRAN;
END

```

Microsoft

Explanation:

Values

- BEGIN DISTRIBUTED TRANSACTION
- BEGIN TRAN
- COMMIT TRAN
- ROLLBACK TRAN
- SET RESULT_SET_CACHING ON

Answer Area

```
BEGIN TRAN
BEGIN TRY
    INSERT INTO dbo.Table1 (col1, col2, col3)
    SELECT col1, col2, col3 FROM stage.Table1;
END TRY
BEGIN CATCH
    IF @@TRANCOUNT > 0
    BEGIN
        ROLLBACK TRAN ;
    END
END CATCH;
IF @@TRANCOUNT > 0
BEGIN
    COMMIT TRAN;
END
```

freepdfdumps.com Microsoft

NEW QUESTION: 6

You have an Azure Synapse Analytics serverless SQL pool, an Azure Synapse Analytics dedicated SQL pool, an Apache Spark pool, and an Azure Data Lake Storage Gen2 account.

You need to create a table in a lake database. The table must be available to both the serverless SQL pool and the Spark pool.

Where should you create the table, and Which file format should you use for data in the table? TO answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Create the table in: 

- The dedicated SQL pool
- The serverless SQL pool
- The Spark pool

File format:

- Apache Parquet
- Delta
- JSON

freepdfdumps.com

Answer:



Explanation:

The dedicated SQL pool

Apache Parquet

NEW QUESTION: 7

You are designing an inventory updates table in an Azure Synapse Analytics dedicated SQL pool. The table will have a clustered columnstore index and will include the following columns:

Table	Comment
EventDate	One million records are added to the table each day
EventTypeID	The table contains 10 million records for each event type.
WarehouseID	The table contains 100 million records for each warehouse.
ProductCategoryTypeID	The table contains 25 million records for each product category type.

You identify the following usage patterns:

- * Analysts will most commonly analyze transactions for a warehouse.
- * Queries will summarize by product category type, date, and/or inventory event type.

You need to recommend a partition strategy for the table to minimize query times.

On which column should you partition the table?

- A. ProductCategoryTypeID
- B. EventDate
- C. WarehouseID
- D. EventTypeID

Answer: C (LEAVE A REPLY)

The number of records for each warehouse is big enough for a good partitioning.

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

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. Before partitions are created, dedicated SQL pool already divides each table into 60 distributed databases.

NEW QUESTION: 8

You have an Azure Synapse Analytics dedicated SQL pool.

You need to monitor the database for long-running queries and identify which queries are waiting on resources. Which dynamic management view should you use for each requirement? To answer, select the appropriate options in the answer area.

NOTE; Each correct answer is worth one point.

Answer Area

Monitor the database for long-running queries: sys.dm_pdw_exec_requests

Identify which queries are waiting on resources: sys.dm_pdw_lock_waits

Microsoft

Answer:

Answer Area

Monitor the database for long-running queries: sys.dm_pdw_exec_requests

Identify which queries are waiting on resources: sys.dm_pdw_lock_waits

Microsoft

Explanation:

Answer Area

Monitor the database for long-running queries: sys.dm_pdw_exec_requests

Identify which queries are waiting on resources: sys.dm_pdw_lock_waits

Microsoft

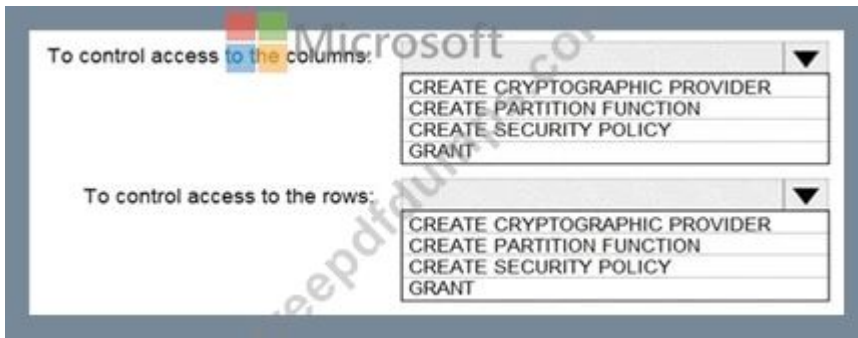
NEW QUESTION: 9

You have an Azure subscription that contains the following resources:

- * An Azure Active Directory (Azure AD) tenant that contains a security group named Group1
- * An Azure Synapse Analytics SQL pool named Pool1

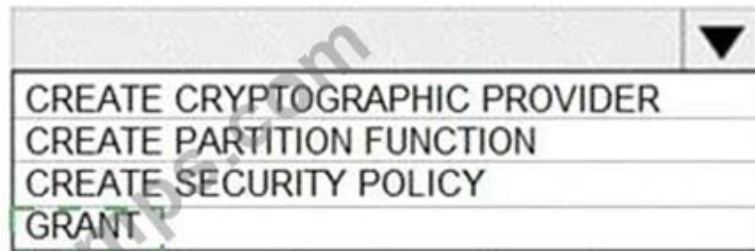
You need to control the access of Group1 to specific columns and rows in a table in Pool1.

Which Transact-SQL commands should you use? To answer, select the appropriate options in the answer area.

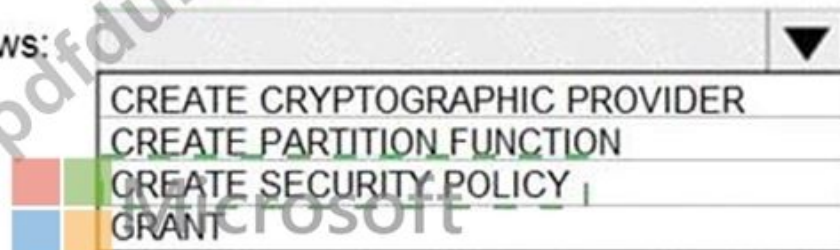


Answer:

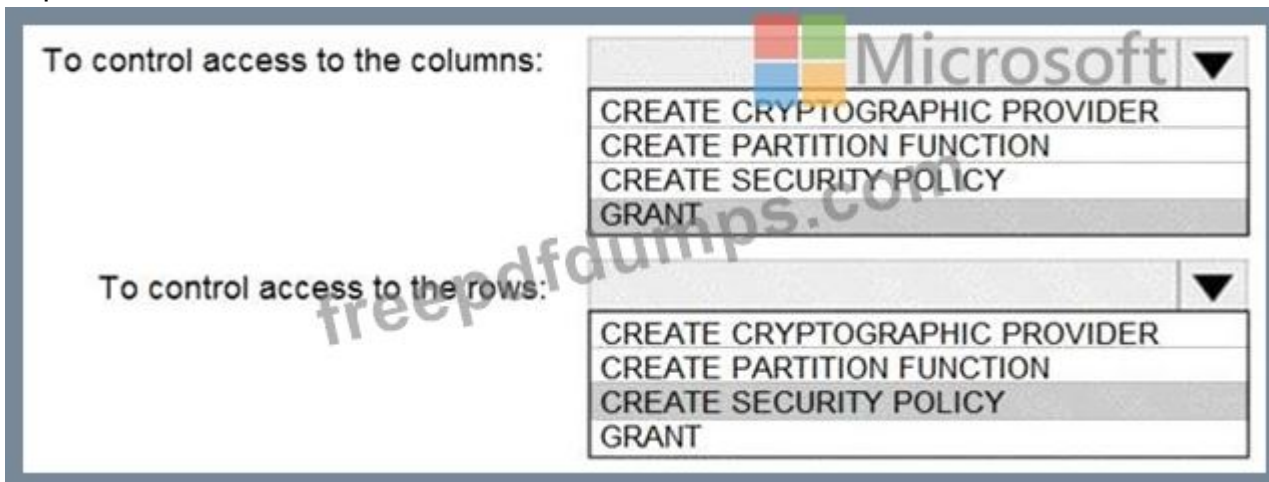
To control access to the columns:



To control access to the rows:



Explanation:



Box 1: GRANT

You can implement column-level security with the GRANT T-SQL statement.

Box 2: CREATE SECURITY POLICY

Implement Row Level Security by using the CREATE SECURITY POLICY Transact-SQL statement

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/column-level-security>

NEW QUESTION: 10

What should you recommend to prevent users outside the Litware on-premises network from accessing the analytical data store?

A. a server-level virtual network rule

- B. a database-level virtual network rule
- C. a database-level firewall IP rule
- D. a server-level firewall IP rule

Answer: A (LEAVE A REPLY)

Virtual network rules are one firewall security feature that controls whether the database server for your single databases and elastic pool in Azure SQL Database or for your databases in SQL Data Warehouse accepts communications that are sent from particular subnets in virtual networks.

Server-level, not database-level: Each virtual network rule applies to your whole Azure SQL Database server, not just to one particular database on the server. In other words, virtual network rule applies at the serverlevel, not at the database-level.

References:

<https://docs.microsoft.com/en-us/azure/sql-database/sql-database-vnet-service-endpoint-rule-overview>

NEW QUESTION: 11

You are designing an Azure Synapse Analytics dedicated SQL pool.

Groups will have access to sensitive data in the pool as shown in the following table.

Name	Enhanced access
Executives	No access to sensitive data
Analysts	Access to in-region sensitive data
Engineers	Access to all numeric sensitive data

You have policies for the sensitive data. The policies vary by region as shown in the following table.

Region	Data considered sensitive
RegionA	Financial, Personally Identifiable Information (PII)
RegionB	Financial, Personally Identifiable Information (PII), medical
RegionC	Financial, medical

You have a table of patients for each region. The tables contain the following potentially sensitive columns.

Name	Sensitive data	Description
CardOnFile	Financial	Debit/credit card number for charges
Height	Medical	Patient's height in cm
ContactEmail	PII	Email address for secure communications

You are designing dynamic data masking to maintain compliance.

For each of the following statements, select Yes if the statement is true. Otherwise, select No.

NOTE: Each correct selection is worth one point.

Statements	Yes	No
Analysts in RegionA require dynamic data masking rules for [Patients_RegionA]	<input type="radio"/>	<input type="radio"/>
Engineers in RegionC require a dynamic data masking rule for [Patients_RegionA], [Height]	<input type="radio"/>	<input type="radio"/>
Engineers in RegionB require a dynamic data masking rule for [Patients_RegionB], [Height]	<input type="radio"/>	<input type="radio"/>

Answer:

Statements	Yes	No
Analysts in RegionA require dynamic data masking rules for [Patients_RegionA]	<input checked="" type="radio"/>	<input type="radio"/>
Engineers in RegionC require a dynamic data masking rule for [Patients_RegionA], [Height]	<input type="radio"/>	<input checked="" type="radio"/>
Engineers in RegionB require a dynamic data masking rule for [Patients_RegionB], [Height]	<input checked="" type="radio"/>	<input type="radio"/>

Explanation:

Answer Area	Statements	Yes	No
	Analysts in RegionA require dynamic data masking rules for [Patients_RegionA],[Height].	<input checked="" type="radio"/>	<input type="radio"/>
	Engineers in RegionC require a dynamic data masking rule for [Patients_RegionA],[Height].	<input checked="" type="radio"/>	<input type="radio"/>
	Executives in RegionB require a dynamic data masking rule for [Patients_RegionB],[Height].	<input type="radio"/>	<input checked="" type="radio"/>

Reference:

<https://docs.microsoft.com/en-us/azure/azure-sql/database/dynamic-data-masking-overview>

NEW QUESTION: 12

You are designing an enterprise data warehouse in Azure Synapse Analytics that will store website traffic analytics in a star schema.

You plan to have a fact table for website visits. The table will be approximately 5 GB.

You need to recommend which distribution type and index type to use for the table. The solution must provide the fastest query performance.

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

NOTE: Each correct selection is worth one point.



Answer:



Explanation:

Box 1: Hash

Consider using a hash-distributed table when:

The table size on disk is more than 2 GB.

The table has frequent insert, update, and delete operations.

Box 2: Clustered columnstore

Clustered columnstore tables offer both the highest level of data compression and the best overall query performance.

Reference:

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

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

NEW QUESTION: 13

You are designing an Azure Synapse Analytics workspace.

You need to recommend a solution to provide double encryption of all the data at rest.

Which two components should you include in the recommendation? Each coned answer presents part of the solution NOTE: Each correct selection is worth one point.

A. an X509 certificate

- B. an RSA key
- C. an Azure key vault that has purge protection enabled
- D. an Azure virtual network that has a network security group (NSG)
- E. an Azure Policy initiative

Answer: (SHOW ANSWER)

Synapse workspaces encryption uses existing keys or new keys generated in Azure Key Vault. A single key is used to encrypt all the data in a workspace. Synapse workspaces support RSA 2048 and 3072 byte-sized keys, and RSA-HSM keys.

The Key Vault itself needs to have purge protection enabled.

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/security/workspaces-encryption>

NEW QUESTION: 14

You have an Azure Databricks workspace that contains a Delta Lake dimension table named Table1. Table1 is a Type 2 slowly changing dimension (SCD) table. You need to apply updates from a source table to Table1.

Which Apache Spark SQL operation should you use?

- A. CREATE
- B. UPDATE
- C. MERGE
- D. ALTER

Answer: (SHOW ANSWER)

The Delta provides the ability to infer the schema for data input which further reduces the effort required in managing the schema changes. The Slowly Changing Data(SCD) Type 2 records all the changes made to each key in the dimensional table. These operations require updating the existing rows to mark the previous values of the keys as old and then inserting new rows as the latest values. Also, Given a source table with the updates and the target table with dimensional data, SCD Type 2 can be expressed with the merge.

Example:

```
// Implementing SCD Type 2 operation using merge function
customersTable
as("customers")
merge(
stagedUpdates.as("staged_updates"),
"customers.customerId = mergeKey")
whenMatched("customers.current = true AND customers.address <> staged_updates.address")
updateExpr(Map(
"current" -> "false",
"endDate" -> "staged_updates.effectiveDate"))
whenNotMatched()
insertExpr(Map(
```

```
"customerid" -> "staged_updates.customerId",
"address" -> "staged_updates.address",
"current" -> "true",
"effectiveDate" -> "staged_updates.effectiveDate",
"endDate" -> "null"))
execute()
}
```

Reference:

<https://www.projectpro.io/recipes/what-is-slowly-changing-data-scd-type-2-operation-delta-table-databricks>

NEW QUESTION: 15

You are designing a real-time dashboard solution that will visualize streaming data from remote sensors that connect to the internet. The streaming data must be aggregated to show the average value of each 10-second interval. The data will be discarded after being displayed in the dashboard.

The solution will use Azure Stream Analytics and must meet the following requirements:

- * Minimize latency from an Azure Event hub to the dashboard.
- * Minimize the required storage.
- * Minimize development effort.

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

NOTE: Each correct selection is worth one point

Azure Stream Analytics input type:

Azure Event Hub
Azure SQL Database
Azure Stream Analytics
Microsoft Power BI

Azure Stream Analytics output type:

Azure Event Hub
Azure SQL Database
Azure Stream Analytics
Microsoft Power BI

Aggregation query location:

Azure Event Hub
Azure SQL Database
Azure Stream Analytics
Microsoft Power BI

Answer:

Azure Stream Analytics input type:

Azure Stream Analytics output type:

Aggregation query location:

Microsoft


freepdfdumps.com

Azure Event Hub
Azure SQL Database
Azure Stream Analytics
Microsoft Power BI

Azure Event Hub
Azure SQL Database
Azure Stream Analytics
Microsoft Power BI

Azure Event Hub
Azure SQL Database
Azure Stream Analytics
Microsoft Power BI

Explanation:

 Azure Stream Analytics Input type: ▼

- Azure Event Hub
- Azure SQL Database
- Azure Stream Analytics
- Microsoft Power BI

Azure Stream Analytics output type: ▼

- Azure Event Hub
- Azure SQL Database
- Azure Stream Analytics
- Microsoft Power BI

Aggregation query location: ▼

- Azure Event Hub
- Azure SQL Database
- Azure Stream Analytics
- Microsoft Power BI

Reference:

<https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-power-bi-dashboard>

NEW QUESTION: 16

You are planning a solution to aggregate streaming data that originates in Apache Kafka and is output to Azure Data Lake Storage Gen2. The developers who will implement the stream processing solution use Java, Which service should you recommend using to process the streaming data?

- A. Azure Data Factory
- B. Azure Stream Analytics
- C. Azure Databricks
- D. Azure Event Hubs

Answer: (SHOW ANSWER)

<https://docs.microsoft.com/en-us/azure/architecture/data-guide/technology-choices/stream-processing>

Valid DP-203 Dumps shared by Actual4test.com for Helping Passing DP-203 Exam! Actual4test.com now offer the **newest DP-203 exam dumps**, the Actual4test.com DP-203 exam **questions have been updated** and **answers have been corrected** get the **newest** Actual4test.com DP-203 dumps with Test

Engine here: https://www.actual4test.com/DP-203_examcollection.html (365 Q&As Dumps, 30%OFF Special Discount: **Freepdfdumps**)

NEW QUESTION: 17

You have an Azure data factory.

You need to examine the pipeline failures from the last 180 flays.

What should you use?

- A. the Activity tog blade for the Data Factory resource
- B. Azure Data Factory activity runs in Azure Monitor
- C. Pipeline runs in the Azure Data Factory user experience
- D. the Resource health blade for the Data Factory resource

Answer: B (LEAVE A REPLY)

Data Factory stores pipeline-run data for only 45 days. Use Azure Monitor if you want to keep that data for a longer time.

Reference:

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

NEW QUESTION: 18

You are designing an Azure Data Lake Storage Gen2 structure for telemetry data from 25 million devices distributed across seven key geographical regions. Each minute, the devices will send a JSON payload of metrics to Azure Event Hubs.

You need to recommend a folder structure for the data. The solution must meet the following requirements:

- * Data engineers from each region must be able to build their own pipelines for the data of their respective region only.
- * The data must be processed at least once every 15 minutes for inclusion in Azure Synapse Analytics serverless SQL pools.

How should you recommend completing the structure? 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.

Values	Answer Area
<input data-bbox="154 1688 456 1727" type="text" value="{deviceID}"/>	/ <input data-bbox="511 1688 820 1727" type="text" value="Value"/> / <input data-bbox="851 1688 1161 1727" type="text" value="Value"/> / <input data-bbox="1192 1688 1501 1727" type="text" value="Value"/> .json
<input data-bbox="154 1733 456 1771" type="text" value="{mm}/{HH}/{DD}/{MM}/{YYYY}"/>	
<input data-bbox="154 1778 456 1816" type="text" value="{regionID}/{deviceID}"/>	
<input data-bbox="154 1823 456 1861" type="text" value="{regionID}/raw"/>	
<input data-bbox="154 1868 456 1906" type="text" value="{YYYY}/{MM}/{DD}/{HH}"/>	
<input data-bbox="154 1912 456 1951" type="text" value="{YYYY}/{MM}/{DD}/{HH}/{mm}"/>	
<input data-bbox="154 1957 456 1995" type="text" value="raw/{deviceID}"/>	
<input data-bbox="154 2002 456 2040" type="text" value="raw/{regionID}"/>	

Answer:

Explanation:

Box 1: {YYYY}/{MM}/{DD}/{HH}

Date Format [optional]: if the date token is used in the prefix path, you can select the date format in which your files are organized. Example: YYYY/MM/DD Time Format [optional]: if the time token is used in the prefix path, specify the time format in which your files are organized. Currently the only supported value is HH.

Box 2: {regionID}/raw

Data engineers from each region must be able to build their own pipelines for the data of their respective region only.

Box 3: {deviceID}

Reference:

<https://github.com/paolosalvatori/StreamAnalyticsAzureDataLakeStore/blob/master/README.md>

NEW QUESTION: 19

A company purchases IoT devices to monitor manufacturing machinery. The company uses an IoT appliance to communicate with the IoT devices.

The company must be able to monitor the devices in real-time.

You need to design the solution.

What should you recommend?

- A. Azure Stream Analytics cloud job using Azure PowerShell
- B. Azure Analysis Services using Azure Portal
- C. Azure Data Factory instance using Azure Portal
- D. Azure Analysis Services using Azure PowerShell

Answer: A (LEAVE A REPLY)

Stream Analytics is a cost-effective event processing engine that helps uncover real-time insights from devices, sensors, infrastructure, applications and data quickly and easily.

Monitor and manage Stream Analytics resources with Azure PowerShell cmdlets and powershell scripting that execute basic Stream Analytics tasks.

Reference:

<https://cloudblogs.microsoft.com/sqlserver/2014/10/29/microsoft-adds-iot-streaming-analytics-data-production-and-workflow-services-to-azure/>

NEW QUESTION: 20

You are designing a statistical analysis solution that will use custom proprietary Python functions on near real-time data from Azure Event Hubs.

You need to recommend which Azure service to use to perform the statistical analysis. The solution must minimize latency.

What should you recommend?

- A. Azure Stream Analytics
- B. Azure SQL Database
- C. Azure Databricks
- D. Azure Synapse Analytics

Answer: A (LEAVE A REPLY)

Reference:

<https://docs.microsoft.com/en-us/azure/event-hubs/process-data-azure-stream-analytics>

NEW QUESTION: 21

A company plans to use Platform-as-a-Service (PaaS) to create the new data pipeline process. The process must meet the following requirements:

Ingest:

- * Access multiple data sources.
- * Provide the ability to orchestrate workflow.
- * Provide the capability to run SQL Server Integration Services packages.

Store:

- * Optimize storage for big data workloads.
- * Provide encryption of data at rest.
- * Operate with no size limits.

Prepare and Train:

- * Provide a fully-managed and interactive workspace for exploration and visualization.
- * Provide the ability to program in R, SQL, Python, Scala, and Java.
- * Provide seamless user authentication with Azure Active Directory.

Model & Serve:

- * Implement native columnar storage.
- * Support for the SQL language
- * Provide support for structured streaming.

You need to build the data integration pipeline.

Which technologies should you use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.



Architecture requirement

Technology

Ingest

- Logic Apps
- Azure Data Factory
- Azure Automation

Store

- Azure Data Lake Storage
- Azure Blob storage
- Azure files

Prepare and Train

- HDInsight Apache Spark cluster
- Azure Databricks
- HDInsight Apache Storm cluster

Model and Serve

- HDInsight Apache Kafka cluster
- Azure Synapse Analytics
- Azure Data Lake Storage


Answer:

Answer Area	
Architecture requirement	Technology
Ingest	<div style="border: 1px solid gray; padding: 2px;"> ▼ </div> <div style="border: 1px solid gray; padding: 2px;"> Logic Apps Azure Data Factory Azure Automation </div>
Store	<div style="border: 1px solid gray; padding: 2px;"> ▼ </div> <div style="border: 1px solid gray; padding: 2px;"> Azure Data Lake Storage Azure Blob storage Azure files </div>
Prepare and Train	<div style="border: 1px solid gray; padding: 2px;"> ▼ </div> <div style="border: 1px solid gray; padding: 2px;"> HDInsight Apache Spark cluster Azure Databricks HDInsight Apache Storm cluster </div>
Model and Serve	<div style="border: 1px solid gray; padding: 2px;"> ▼ </div> <div style="border: 1px solid gray; padding: 2px;"> HDInsight Apache Kafka cluster Azure Synapse Analytics Azure Data Lake Storage </div>

Explanation:

Architecture requirement

Technology



Ingest	<div style="border: 1px solid black; padding: 2px;"><div style="background-color: #f0f0f0; padding: 2px; display: flex; justify-content: space-between;">▼</div><ul style="list-style-type: none">Logic Apps<li style="background-color: #e0e0e0;">Azure Data FactoryAzure Automation</div>
Store	<div style="border: 1px solid black; padding: 2px;"><div style="background-color: #f0f0f0; padding: 2px; display: flex; justify-content: space-between;">▼</div><ul style="list-style-type: none">Azure Data Lake Storage<li style="background-color: #e0e0e0;">Azure Blob storageAzure files</div>
Prepare and Train	<div style="border: 1px solid black; padding: 2px;"><div style="background-color: #f0f0f0; padding: 2px; display: flex; justify-content: space-between;">▼</div><ul style="list-style-type: none">HDInsight Apache Spark cluster<li style="background-color: #e0e0e0;">Azure DatabricksHDInsight Apache Storm cluster</div>
Model and Serve	<div style="border: 1px solid black; padding: 2px;"><div style="background-color: #f0f0f0; padding: 2px; display: flex; justify-content: space-between;">▼</div><ul style="list-style-type: none">HDInsight Apache Kafka cluster<li style="background-color: #e0e0e0;">Azure Synapse AnalyticsAzure Data Lake Storage</div>

NEW QUESTION: 22

You build an Azure Data Factory pipeline to move data from an Azure Data Lake Storage Gen2 container to a database in an Azure Synapse Analytics dedicated SQL pool.

Data in the container is stored in the following folder structure.

`/in/{YYYY}/{MM}/{DD}/{HH}/{mm}`


The earliest folder is `/in/2021/01/01/00/00`. The latest folder is `/in/2021/01/15/01/45`.

You need to configure a pipeline trigger to meet the following requirements:

- * Existing data must be loaded.
- * Data must be loaded every 30 minutes.
- * Late-arriving data of up to two minutes must be included in the load for the time at which the data should have arrived.

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

NOTE: Each correct selection is worth one point.

Type:  ▼

Event
On-demand
Schedule
Tumbling window

Additional properties: ▼

Prefix: /in/, Event: Blob created
Recurrence: 30 minutes, Start time: 2021-01-01T00:00
Recurrence: 30 minutes, Start time: 2021-01-01T00:00, Delay: 2 minutes
Recurrence: 32 minutes, Start time: 2021-01-15T01:45

Answer:

Type: ▼

Event
On-demand
Schedule
Tumbling window

Additional properties: ▼

Prefix: /in/, Event: Blob created
Recurrence: 30 minutes, Start time: 2021-01-01T00:00
Recurrence: 30 minutes, Start time: 2021-01-01T00:00, Delay: 2 minutes
Recurrence: 32 minutes, Start time: 2021-01-15T01:45

Explanation:

Type: ▼

Event
On-demand
Schedule
Tumbling window

Additional properties: ▼

Prefix: /in/, Event: Blob created
Recurrence: 30 minutes, Start time: 2021-01-01T00:00
Recurrence: 30 minutes, Start time: 2021-01-01T00:00, Delay: 2 minutes
Recurrence: 32 minutes, Start time: 2021-01-15T01:45

Box 1: Tumbling window

To be able to use the Delay parameter we select Tumbling window.

Box 2:

Recurrence: 30 minutes, not 32 minutes

Delay: 2 minutes.

The amount of time to delay the start of data processing for the window. The pipeline run is started after the expected execution time plus the amount of delay. The delay defines how long the trigger waits past the due time before triggering a new run. The delay doesn't alter the window start time.

Reference:

<https://docs.microsoft.com/en-us/azure/data-factory/how-to-create-tumbling-window-trigger>

NEW QUESTION: 23

You have an Azure Blob storage account named storage! and an Azure Synapse Analytics serverless SQL pool named Pool!. From Pool!, you plan to run ad-hoc queries that target storage!

You need to ensure that you can use shared access signature (SAS) authorization without defining a data source. What should you create first?

- A. a database scoped credential
- B. a server-level credential
- C. a managed identity
- D. a stored access policy

Answer: D (LEAVE A REPLY)

NEW QUESTION: 24

You have an Azure Active Directory (Azure AD) tenant that contains a security group named Group1. You have an Azure Synapse Analytics dedicated SQL pool named dw1 that contains a schema named schema1.

You need to grant Group1 read-only permissions to all the tables and views in schema1. The solution must use the principle of least privilege.

Which three 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.

Actions

Answer Area

Create a database role named Role1 and grant Role1 SELECT permissions to schema1.

Create a database role named Role1 and grant Role1 SELECT permissions to dw1.


Assign the Azure role-based access control (Azure RBAC) Reader role for dw1 to Group1.

Create a database user in dw1 that represents Group1 and uses the FROM EXTERNAL PROVIDER clause.

Assign Role1 to the Group1 database user.

Answer:

Actions	Answer Area
Create a database role named Role1 and grant Role1 SELECT permissions to schema1.	Create a database role named Role1 and grant Role1 SELECT permissions to schema1.
Create a database role named Role1 and grant Role1 SELECT permissions to dw1.	
Assign the Azure role-based access control (Azure RBAC) Reader role for dw1 to Group1.	Assign Role1 to the Group1 database user.
Create a database user in dw1 that represents Group1 and uses the FROM EXTERNAL PROVIDER clause.	Assign the Azure role-based access control (Azure RBAC) Reader role for dw1 to Group1.
Assign Role1 to the Group1 database user.	



Explanation:

Create a database role named Role1 and grant Role1 SELECT permissions to schema1.
Assign Role1 to the Group1 database user.
Assign the Azure role-based access control (Azure RBAC) Reader role for dw1 to Group1.

Step 1: Create a database role named Role1 and grant Role1 SELECT permissions to schema You need to grant Group1 read-only permissions to all the tables and views in schema1.

Place one or more database users into a database role and then assign permissions to the database role.

Step 2: Assign Rol1 to the Group database user

Step 3: Assign the Azure role-based access control (Azure RBAC) Reader role for dw1 to Group1

Reference:

<https://docs.microsoft.com/en-us/azure/data-share/how-to-share-from-sql>

NEW QUESTION: 25

You have an Azure Synapse Analytics dedicated SQL pool that contains a table named Sales.Orders. Sales.

Orders contains a column named SalesRep.

You plan to implement row-level security (RLS) for Sales.Orders.

You need to create the security policy that will be used to implement RLS. The solution must ensure that sales representatives only see rows for which the value of the SalesRep column matches their username.

How should you complete the code? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

```
CREATE SCHEMA Security;
GO
CREATE FUNCTION Security.tvf_securitypredicate(@SalesRep AS nvarchar(50))
    RETURNS TABLE
WITH SCHEMABINDING
    ENCRYPTION
    RETURNS NULL ON NULL INPUT
AS
    RETURN SELECT 1 AS tvf_securitypredicate_result
    WHERE @SalesRep = USER_NAME();
GO
CREATE SECURITY POLICY SalesFilter
    ADD FILTER PREDICATE Security.tvf_securitypredicate(SalesRep)
    ADD BLOCK PREDICATE Security.tvf_securitypredicate(SalesRep)
    ADD BLOCK PREDICATE tvf_securitypredicate_result
    ADD FILTER PREDICATE Security.tvf_securitypredicate(SalesRep)
```

Answer:
Answer Area

```
CREATE SCHEMA Security;
GO
CREATE FUNCTION Security.tvf_securitypredicate(@SalesRep AS nvarchar(50))
    RETURNS TABLE
WITH SCHEMABINDING
    ENCRYPTION
    RETURNS NULL ON NULL INPUT
AS
    RETURN SELECT 1 AS tvf_securitypredicate_result
    WHERE @SalesRep = USER_NAME();
GO
CREATE SECURITY POLICY SalesFilter
    ADD FILTER PREDICATE Security.tvf_securitypredicate(SalesRep)
    ADD BLOCK PREDICATE Security.tvf_securitypredicate(SalesRep)
    ADD BLOCK PREDICATE tvf_securitypredicate_result
    ADD FILTER PREDICATE Security.tvf_securitypredicate(SalesRep)
```



Explanation:

Answer Area

```
CREATE SCHEMA Security;  
GO  
CREATE FUNCTION Security.tvf_securitypredicate(@SalesRep AS nvarchar(50))  
    RETURNS TABLE  
WITH SCHEMABINDING  
AS  
    RETURN SELECT 1 AS tvf_securitypredicate_result  
    WHERE @SalesRep = USER_NAME();  
GO  
CREATE SECURITY POLICY SalesFilter  
    ADD FILTER PREDICATE Security.tvf_securitypredicate(SalesRep)  
    ON Sales.Orders  
    WITH (STATE = ON);
```

NEW QUESTION: 26

You have an Azure Stream Analytics job that is a Stream Analytics project solution in Microsoft Visual Studio. The job accepts data generated by IoT devices in the JSON format.

You need to modify the job to accept data generated by the IoT devices in the Protobuf format.

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

Actions

Answer Area

- Change the Event Serialization Format to Protobuf in the input.json file of the job and reference the DLL.
- Add an Azure Stream Analytics Custom Deserializer Project (.NET) project to the solution.
- Add .NET deserializer code for Protobuf to the custom deserializer project.
- Add .NET deserializer code for Protobuf to the Stream Analytics project.
- Add an Azure Stream Analytics Application project to the solution.

Answer:



Actions	Answer Area
Change the Event Serialization Format to Protobuf in the input.json file of the job and reference the DLL.	Add an Azure Stream Analytics Custom Deserializer Project (.NET) project to the solution.
Add an Azure Stream Analytics Custom Deserializer Project (.NET) project to the solution.	
Add .NET deserializer code for Protobuf to the custom deserializer project.	Add .NET deserializer code for Protobuf to the custom deserializer project.
Add .NET deserializer code for Protobuf to the Stream Analytics project.	
Add an Azure Stream Analytics Application project to the solution.	Add an Azure Stream Analytics Application project to the solution.

Explanation:

Add an Azure Stream Analytics Custom Deserializer Project (.NET) project to the solution.

Add .NET deserializer code for Protobuf to the custom deserializer project.

Add an Azure Stream Analytics Application project to the solution.

Step 1: Add an Azure Stream Analytics Custom Deserializer Project (.NET) project to the solution.

Create a custom deserializer

1. Open Visual Studio and select File > New > Project. Search for Stream Analytics and select Azure Stream Analytics Custom Deserializer Project (.NET). Give the project a name, like Protobuf Deserializer.

The screenshot shows the 'Create a new project' dialog in Visual Studio. The search bar at the top contains 'stream analytics'. Below the search bar, there are several project templates listed under 'Recent project templates'. The template 'Azure Stream Analytics Custom Deserializer Project (.NET)' is highlighted with a red box. The description for this template is 'A .NET Standard project for Azure Stream Analytics Custom Deserializer.'

2. In Solution Explorer, right-click your Protobuf Deserializer project and select Manage NuGet Packages from the menu. Then install the Microsoft.Azure.StreamAnalytics and Google.Protobuf NuGet packages.

3. Add the MessageBodyProto class and the MessageBodyDeserializer class to your project.

4. Build the Protobuf Deserializer project.

Step 2: Add .NET deserializer code for Protobuf to the custom deserializer project Azure Stream Analytics has built-in support for three data formats: JSON, CSV, and Avro. With custom .NET deserializers, you can read data from other formats such as Protocol Buffer, Bond and other user defined formats for both cloud and edge jobs.

Step 3: Add an Azure Stream Analytics Application project to the solution Add an Azure Stream Analytics project

* In Solution Explorer, right-click the Protobuf Deserializer solution and select Add > New Project.

Under Azure Stream Analytics > Stream Analytics, choose Azure Stream Analytics Application. Name it ProtobufCloudDeserializer and select OK.

* Right-click References under the ProtobufCloudDeserializer Azure Stream Analytics project. Under Projects, add Protobuf Deserializer. It should be automatically populated for you.

Reference:

<https://docs.microsoft.com/en-us/azure/stream-analytics/custom-deserializer>

NEW QUESTION: 27

You have an Azure subscription that contains an Azure Synapse Analytics dedicated SQL pool named Pool1.

Pool1 receives new data once every 24 hours.

You have the following function.

```
create function dbo.udfFtoC(@F decimal)
return decimal
as
begin
return (@F - 32) * 5.0 / 9
end
```

You have the following query.

```
select avg_date, sensorid, avg_f, dbo.udfFtoC(avg_temperature) as avg_c from SensorTemps
where avg_date = @parameter
```

The query is executed once every 15 minutes and the @parameter value is set to the current date. You need to minimize the time it takes for the query to return results.

Which two actions should you perform? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. Create an index on the avg_f column.
- B. Convert the avg_c column into a calculated column.
- C. Create an index on the sensorid column.

- D. Enable result set caching.
- E. Change the table distribution to replicate.

Answer: (SHOW ANSWER)

<https://learn.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/performance-tuning-result-set-caching>

NEW QUESTION: 28

You need to implement the surrogate key for the retail store table. The solution must meet the sales transaction dataset requirements.

What should you create?

- A. a table that has an IDENTITY property
- B. a system-versioned temporal table
- C. a user-defined SEQUENCE object
- D. a table that has a FOREIGN KEY constraint

Answer: A (LEAVE A REPLY)

Scenario: Implement a surrogate key to account for changes to the retail store addresses.

A surrogate key on a table is a column with a unique identifier for each row. The key is not generated from the table data. Data modelers like to create surrogate keys on their tables when they design data warehouse models. You can use the IDENTITY property to achieve this goal simply and effectively without affecting load performance.

Reference:

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

NEW QUESTION: 29

You have an Azure subscription that contains a storage account. The account contains a blob container named blob1 and an Azure Synapse Analytic serve-less SQL pool You need to Query the CSV files stored in blob1. The solution must ensure that all the files in a (older named csv and all its subfolders are queried

How should you complete the query? to answer, select the appropriate options in the answer area NOTE:

Each correct selection is worth one point.

Answer Area

```
SELECT StickerKey, StickerID, Thickness, Width
FROM OPENROWSET
(
    BULK 
    FORMAT = 'CSV',
    PARSER_VERSION = '2.0',
    FIRSTROW = 2) AS Stickers
```




Answer Area

```
SELECT StickKey, StickID, Thickness, Width
FROM OPENROWSET
(
    BULK 'abfs://acct1.blob.core.windows.net/blob1/csv/'
    FORMAT 'https'
    PARSE_V 'tcp'
    FIRSTROW 'Wasd'
)
```




Answer:

Answer Area



```
SELECT StickKey, StickID, Thickness, Width
FROM OPENROWSET
(
    BULK 'abfs://acct1.blob.core.windows.net/blob1/csv/'
    FORMAT 'https'
    PARSE_V 'tcp'
    FIRSTROW 'Wasd'
)
```



Explanation:

Answer Area

```
SELECT StickKey, StickID, Thickness, Width
FROM OPENROWSET
(
    BULK 'abfs://acct1.blob.core.windows.net/blob1/csv/'
    FORMAT 'https'
    PARSE_V 'tcp'
    FIRSTROW 'Wasd'
)
```



NEW QUESTION: 30

You are designing a fact table named FactPurchase in an Azure Synapse Analytics dedicated SQL pool. The table contains purchases from suppliers for a retail store. FactPurchase will contain the following columns.

Name	Data type	Nullable
PurchaseKey	Bigint	No
DateKey	Int	No
SupplierKey	Int	No
StockItemKey	Int	No
PurchaseOrderID	Int	Yes
OrderedQuantity	Int	No
OrderedOuters	Int	No
ReceivedOuters	Int	No
Package	Nvarchar(50)	No
IsOrderFinalized	Bit	No
LineageKey	Int	No

FactPurchase will have 1 million rows of data added daily and will contain three years of data.

Transact-SQL queries similar to the following query will be executed daily.

```
SELECT
```

```
SupplierKey, StockItemKey, COUNT(*)
```

```
FROM FactPurchase
```

```
WHERE DateKey >= 20210101
```

```
AND DateKey <= 20210131
```

```
GROUP BY SupplierKey, StockItemKey
```

Which table distribution will minimize query times?

- A. round-robin
- B. replicated
- C. hash-distributed on DateKey
- D. hash-distributed on PurchaseKey

Answer: D (LEAVE A REPLY)

Hash-distributed tables improve query performance on large fact tables, and are the focus of this article.

Round-robin tables are useful for improving loading speed.

Reference:

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

NEW QUESTION: 31

You are implementing a star schema in an Azure Synapse Analytics dedicated SQL pool.

You plan to create a table named DimProduct.

DimProduct must be a Type 3 slowly changing dimension (SCO) table that meets the following requirements:

- * The values in two columns named ProductKey and ProductSourceID will remain the same.
- * The values in three columns named ProductName, ProductDescription, and Color can change.

You need to add additional columns to complete the following table definition.

```
CREATE TABLE [dbo].[dimproduct]
(
    [ProductKey] INT NOT NULL,
    [ProductSourceID] INT NOT NULL,
    [ProductName] NVARCHAR(100) NOT NULL,
    [ProductDescription] NVARCHAR(2000) NOT NULL,
    [Color] NVARCHAR(50) NOT NULL
)
WITH
(
    DISTRIBUTION = REPLICATE,
    CLUSTERED COLUMNSTORE INDEX
);
```

- A. [EffectiveStartDate] [datetime] NOT NULL
- B. [OriginalColor] NVARCHAR(50) NOT NULL
- C. [IsCurrentRow] [bit] NOT NULL
- D. [OriginalProductName] NVARCHAR(100) NULL
- E. [OriginalProductDescription] NVARCHAR(2000) NOT NULL
- F. [EffectiveEndDate] [datetime] NOT NULL

Answer: A,C,E ([LEAVE A REPLY](#))

Valid DP-203 Dumps shared by Actual4test.com for Helping Passing DP-203 Exam! Actual4test.com now offer the **newest DP-203 exam dumps**, the Actual4test.com DP-203 exam **questions have been updated** and **answers have been corrected** get the **newest** Actual4test.com DP-203 dumps with Test Engine here: https://www.actual4test.com/DP-203_examcollection.html (365 Q&As Dumps, **30%OFF** Special Discount: **Freepdfdumps**)

NEW QUESTION: 32

You have an Azure Data Lake Storage account that contains a staging zone.

You need to design a dairy process to ingest incremental data from the staging zone, transform the data by executing an R script, and then insert the transformed data into a data warehouse in Azure Synapse Analytics.

Solution: You use an Azure Data Factory schedule trigger to execute a pipeline that copies the data to a staging table in the data warehouse, and then uses a stored procedure to execute the R script.

Does this meet the goal?

A. Yes

B. No

Answer: (SHOW ANSWER)

If you need to transform data in a way that is not supported by Data Factory, you can create a custom activity with your own data processing logic and use the activity in the pipeline.

Note: You can use data transformation activities in Azure Data Factory and Synapse pipelines to transform and process your raw data into predictions and insights at scale.

Reference:

<https://docs.microsoft.com/en-us/azure/data-factory/transform-data>

NEW QUESTION: 33

You have an Azure event hub named retailhub that has 16 partitions. Transactions are posted to retailhub. Each transaction includes the transaction ID, the individual line items, and the payment details. The transaction ID is used as the partition key.

You are designing an Azure Stream Analytics job to identify potentially fraudulent transactions at a retail store. The job will use retailhub as the input. The job will output the transaction ID, the individual line items, the payment details, a fraud score, and a fraud indicator.

You plan to send the output to an Azure event hub named fraudhub.

You need to ensure that the fraud detection solution is highly scalable and processes transactions as quickly as possible.

How should you structure the output of the Stream Analytics job? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Number of partitions:

	▼
1	
8	
16	
32	

Partition key:

	▼
Fraud indicator	
Fraud score	
Individual line items	
Payment details	
Transaction ID	

Answer:

Number of partitions:

	▼
1	
8	
16	
32	

Partition key:

	▼
Fraud indicator	
Fraud score	
Individual line items	
Payment details	
Transaction ID	

Explanation:

Number of partitions:

▼

1
8
16
32

Partition key:

▼

Fraud indicator
Fraud score
Individual line items
Payment details
Transaction ID

Box 1: 16

For Event Hubs you need to set the partition key explicitly.

An embarrassingly parallel job is the most scalable scenario in Azure Stream Analytics. It connects one partition of the input to one instance of the query to one partition of the output.

Box 2: Transaction ID

Reference:

<https://docs.microsoft.com/en-us/azure/event-hubs/event-hubs-features#partitions>

NEW QUESTION: 34

You have an Azure subscription that contains the resources shown in the following table.

Name	Type	Description
ws1	Azure Synapse Analytics workspace	Contains a pipeline named pipeline1
storage1	Azure Data Lake Storage account	Used to store Apache Parquet data files that include LIST and STRUCT data types
SQL1	Azure Synapse Analytics dedicated SQL pool	None

You need to ingest the Parquet files from storage1 to SQL1 by using pipeline1. The solution must meet the following requirements:

- * Minimize complexity.
- * Ensure that additional columns in the files are processed as strings.
- * Ensure that files containing additional columns are processed successfully.

How should you configure pipeline1? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area



Ingest the data from storage1 to SQL1 by using:


- A copy activity
- A copy activity
- A custom activity
- A data flow activity
- An Execute SSIS package activity

In Source settings, enable:

- Allow schema drift
- Allow schema drift
- Infer drifted column types
- Sampling
- Validate schema

Answer:

Answer Area



Ingest the data from storage1 to SQL1 by using:

- A copy activity
- A copy activity
- A custom activity
- A data flow activity
- An Execute SSIS package activity

In Source settings, enable:

- Allow schema drift
- Allow schema drift
- Infer drifted column types
- Sampling
- Validate schema

Explanation:

Answer Area



Ingest the data from storage1 to SQL1 by using: A copy activity

In Source settings, enable: Allow schema drift

NEW QUESTION: 35

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

You have files that are ingested and loaded into an Azure Data Lake Storage Gen2 container named container1.

You plan to insert data from the files in container1 into Table1 and transform the data. Each row of data in the files will produce one row in the serving layer of Table1.

You need to ensure that when the source data files are loaded to container1, the DateTime is stored as an additional column in Table1.

Solution: You use an Azure Synapse Analytics serverless SQL pool to create an external table that has an additional DateTime column.

Does this meet the goal?

A. Yes

B. No

Answer: ([SHOW ANSWER](#))

Instead use the derived column transformation to generate new columns in your data flow or to modify existing fields.

Reference:

<https://docs.microsoft.com/en-us/azure/data-factory/data-flow-derived-column>

NEW QUESTION: 36

You have an activity in an Azure Data Factory pipeline. The activity calls a stored procedure in a data warehouse in Azure Synapse Analytics and runs daily.

You need to verify the duration of the activity when it ran last.

What should you use?

- A.** activity runs in Azure Monitor
- B.** Activity log in Azure Synapse Analytics
- C.** the sys.dm_pdw_wait_stats data management view in Azure Synapse Analytics
- D.** an Azure Resource Manager template

Answer: **A** ([LEAVE A REPLY](#))

Reference:

<https://docs.microsoft.com/en-us/azure/data-factory/monitor-visually>

NEW QUESTION: 37

You are designing an Azure Synapse solution that will provide a query interface for the data stored in an Azure Storage account. The storage account is only accessible from a virtual network.

You need to recommend an authentication mechanism to ensure that the solution can access the source data.

What should you recommend?

- A.** a managed identity
- B.** anonymous public read access
- C.** a shared key

Answer: **A** ([LEAVE A REPLY](#))

Managed Identity authentication is required when your storage account is attached to a VNet.

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/quickstart-bulk-load-copy-tsql-examples>

NEW QUESTION: 38

You are building an Azure Stream Analytics job to retrieve game data.

You need to ensure that the job returns the highest scoring record for each five-minute time interval of each game.

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.

SELECT

	▼
Collect(Score)	
CollectTop(1) OVER(ORDER BY Score Desc)	
Game, MAX(Score)	
TopOne() OVER(PARTITION BY Game ORDER BY Score Desc)	

as HighestScore

FROM input TIMESTAMP BY CreatedAt

GROUP BY

	▼
Game	
Hopping(minute,5)	
Tumbling(minute,5)	
Windows(TumblingWindow(minute,5),Hopping(minute,5))	

Answer:

SELECT		▼	as HighestScore
	Collect(Score)		
	CollectTop(1) OVER(ORDER BY Score Desc)		
	Game, MAX(Score)		
	TopOne() OVER(PARTITION BY Game ORDER BY Score Desc)		
FROM input	Timestamp BY	CreatedAt	
GROUP BY		▼	
	Game		
	Hopping(minute,5)		
	Tumbling(minute,5)		
	Windows(TumblingWindow(minute,5),Hopping(minute,5))		

Explanation:

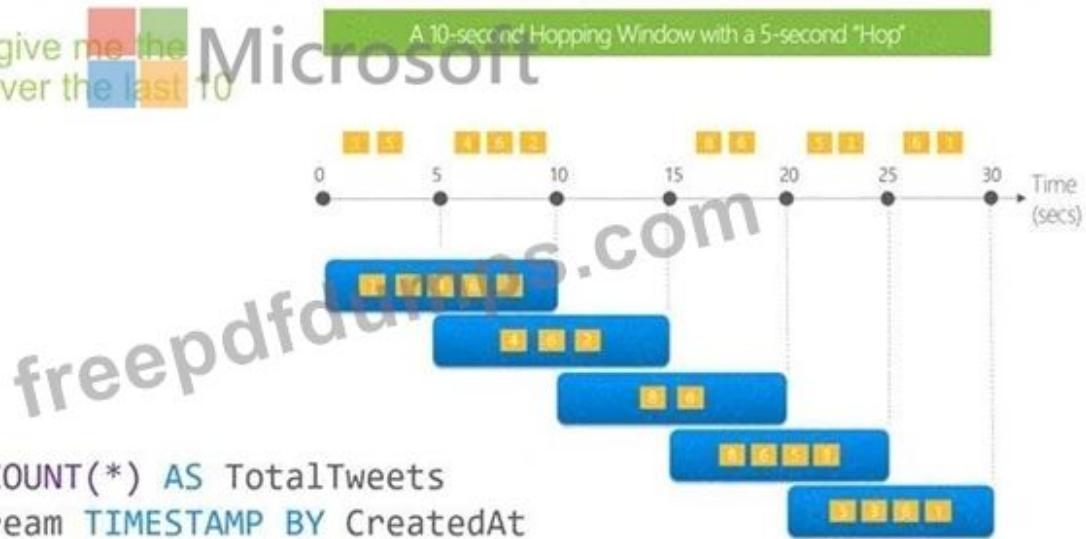
Box 1: TopOne OVER(PARTITION BY Game ORDER BY Score Desc)

TopOne returns the top-rank record, where rank defines the ranking position of the event in the window according to the specified ordering. Ordering/ranking is based on event columns and can be specified in ORDER BY clause.

Box 2: Hopping(minute,5)

Hopping window functions hop forward in time by a fixed period. It may be easy to think of them as Tumbling windows that can overlap and be emitted more often than the window size. Events can belong to more than one Hopping window result set. To make a Hopping window the same as a Tumbling window, specify the hop size to be the same as the window size.

Every 5 seconds give me the count of Tweets over the last 10 seconds



```
SELECT Topic, COUNT(*) AS TotalTweets
FROM TwitterStream TIMESTAMP BY CreatedAt
GROUP BY Topic, HoppingWindow(second, 10 , 5)
```

Reference:

<https://docs.microsoft.com/en-us/stream-analytics-query/topone-azure-stream-analytics>

<https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-window-functions>

NEW QUESTION: 39

What should you do to improve high availability of the real-time data processing solution?

- A. Deploy identical Azure Stream Analytics jobs to paired regions in Azure.
- B. Deploy a High Concurrency Databricks cluster.
- C. Deploy an Azure Stream Analytics job and use an Azure Automation runbook to check the status of the job and to start the job if it stops.
- D. Set Data Lake Storage to use geo-redundant storage (GRS).

Answer: A (LEAVE A REPLY)

Guarantee Stream Analytics job reliability during service updates

Part of being a fully managed service is the capability to introduce new service functionality and improvements at a rapid pace. As a result, Stream Analytics can have a service update deploy on a weekly (or more frequent) basis. No matter how much testing is done there is still a risk that an existing, running job may break due to the introduction of a bug. If you are running mission critical jobs, these risks need to be avoided.

You can reduce this risk by following Azure's paired region model.

Scenario: The application development team will create an Azure event hub to receive real-time sales data, including store number, date, time, product ID, customer loyalty number, price, and discount amount, from the point of sale (POS) system and output the data to data storage in Azure Reference:

<https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-job-reliability>

NEW QUESTION: 40

You have an Azure Synapse Analytics pipeline named pipeline1 that has concurrency set to 1. To run pipeline 1, you create a new trigger as shown in the following exhibit.

Type *
Schedule

Start date * ⓘ
01/03/2023, 9:59:33 pm

Time zone * ⓘ
Coordinated Universal Time (UTC)

Recurrence * ⓘ
Every 1 Day(s)

Advanced recurrence options

Execute at these times ⓘ

Hours
10 × 12 × 17 ×

Minutes
30 × 45 ×

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.

Answer Area

The trigger will run at [answer choice].

If the previous execution of pipeline1 is still running when the trigger fires next, the new triggered execution will [answer choice].

12:30 PM
12:30 AM
12:30 PM
7:30 PM
be queued

be queued
be queued
be skipped
start

Answer:

Answer Area

The trigger will run at [answer choice].

If the previous execution of pipeline1 is still running when the trigger fires next, the new triggered execution will [answer choice].

12:30 PM
12:30 AM
12:30 PM
7:30 PM
be queued

be queued
be queued
be skipped
start

Explanation:

Answer Area

The trigger will run at [answer choice].

If the previous execution of pipeline1 is still running when the trigger fires next, the new triggered execution will [answer choice].

12:30 PM

be queued

NEW QUESTION: 41

You implement an enterprise data warehouse in Azure Synapse Analytics.

You have a large fact table that is 10 terabytes (TB) in size.

Incoming queries use the primary key SaleKey column to retrieve data as displayed in the following table:

SaleKey	CityKey	CustomerKey	StockItemKey	InvoiceDateKey	Quantity	UnitPrice	TotalExcludingTax
49309	90858	70	69	10/22/13	8	16	128
49313	55710	126	69	10/22/13	2	16	32
49343	44710	234	68	10/22/13	10	16	160
49352	66109	163	70	10/22/13	4	16	64
49488	65312	230	70	10/22/13	8	16	128
49646	85877	271	70	10/24/13	1	16	16
49798	41238	288	69	10/24/13	1	16	16

You need to distribute the large fact table across multiple nodes to optimize performance of the table.

Which technology should you use?

- A. hash distributed table with clustered index
- B. hash distributed table with clustered Columnstore index
- C. round robin distributed table with clustered index
- D. round robin distributed table with clustered Columnstore index
- E. heap table with distribution replicate

Answer: B (LEAVE A REPLY)

Hash-distributed tables improve query performance on large fact tables.

Columnstore indexes can achieve up to 100x better performance on analytics and data warehousing workloads and up to 10x better data compression than traditional rowstore indexes.

Reference:

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

<https://docs.microsoft.com/en-us/sql/relational-databases/indexes/columnstore-indexes-query-performance>

NEW QUESTION: 42

You build a data warehouse in an Azure Synapse Analytics dedicated SQL pool.

Analysts write a complex SELECT query that contains multiple JOIN and CASE statements to transform data for use in inventory reports. The inventory reports will use the data and additional WHERE parameters depending on the report. The reports will be produced once daily.

You need to implement a solution to make the dataset available for the reports. The solution must minimize query times.

What should you implement?

- A. a materialized view
- B. a replicated table
- C. in ordered clustered columnstore index
- D. result set chaching

Answer: A (LEAVE A REPLY)

Materialized views for dedicated SQL pools in Azure Synapse provide a low maintenance method for complex analytical queries to get fast performance without any query change.

Note: When result set caching is enabled, dedicated SQL pool automatically caches query results in the user database for repetitive use. This allows subsequent query executions to get results directly from the persisted cache so recomputation is not needed. Result set caching improves query performance and reduces compute resource usage. In addition, queries using cached results set do not use any concurrency slots and thus do not count against existing concurrency limits.

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/performance-tuning-materialized-views>

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/performance-tuning-result-set-caching>

NEW QUESTION: 43

You are designing a date dimension table in an Azure Synapse Analytics dedicated SQL pool. The date dimension table will be used by all the fact tables.

Which distribution type should you recommend to minimize data movement?

- A. HASH
- B. REPLICATE
- C. ROUND ROBIN

Answer: (SHOW ANSWER)

A replicated table has a full copy of the table available on every Compute node. Queries run fast on replicated tables since joins on replicated tables don't require data movement. Replication requires extra storage, though, and isn't practical for large tables.

Reference:

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

NEW QUESTION: 44

You have an Azure Data Factory pipeline named pipeline1 that includes a Copy activity named Copy1. Copy1 has the following configurations:

- * The source of Copy1 is a table in an on-premises Microsoft SQL Server instance that is accessed by using a linked service connected via a self-hosted integration runtime.
- * The sink of Copy1 uses a table in an Azure SQL database that is accessed by using a linked service connected via an Azure integration runtime.

You need to maximize the amount of compute resources available to Copy1. The solution must minimize administrative effort.

What should you do?

- A. Scale up the data flow runtime of the Azure integration runtime and scale out the self-hosted integration runtime.
- B. Scale out the self-hosted integration runtime.
- C. Scale up the data flow runtime of the Azure integration runtime.

Answer: (SHOW ANSWER)

NEW QUESTION: 45

You have an Azure data factory that connects to a Microsoft Purview account. The data factory is registered in Microsoft Purview.

You update a Data Factory pipeline.

You need to ensure that the updated lineage is available in Microsoft Purview.

What You have an Azure subscription that contains an Azure SQL database named DB1 and a storage account named storage1. The storage1 account contains a file named File1.txt. File1.txt contains the names of selected tables in DB1.

You need to use an Azure Synapse pipeline to copy data from the selected tables in DB1 to the files in storage1. The solution must meet the following requirements:

* The Copy activity in the pipeline must be parameterized to use the data in File1.txt to identify the source and destination of the copy.

* Copy activities must occur in parallel as often as possible.

Which two pipeline activities should you include in the pipeline? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

- A. Lookup
- B. If Condition
- C. Get Metadata
- D. ForEach

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 46

You have an Azure Synapse Analytics dedicated SQL pool named Pool1 that contains a table named Sales.

Sales has row-level security (RLS) applied. RLS uses the following predicate filter.

```
CREATE FUNCTION Security.fn_securitypredicate(@SalesRep AS sysname)
    RETURNS TABLE
WITH SCHEMABINDING
AS
    RETURN SELECT 1 AS fn_securitypredicate_result
    WHERE @SalesRep = USER_NAME() OR USER_NAME() = 'Manager';
```

A user named SalesUser1 is assigned the db_datareader role for Pool1.

A user named SalesUser1 is assigned the db_datareader role for Pool1. Which rows in the Sales table are returned when SalesUser1 queries the table?

- A. all the rows
- B. only the rows for which the value in the User_Name column is SalesUser1
- C. only the rows for which the value in the SalesRep column is Manager
- D. only the rows for which the value in the SalesRep column is SalesUser1

Answer: B ([LEAVE A REPLY](#))

Valid DP-203 Dumps shared by Actual4test.com for Helping Passing DP-203 Exam! Actual4test.com now offer the **newest DP-203 exam dumps**, the Actual4test.com DP-203 exam **questions have been updated** and **answers have been corrected** get the **newest** Actual4test.com DP-203 dumps with Test Engine here: https://www.actual4test.com/DP-203_examcollection.html (**365** Q&As Dumps, **30%OFF** Special Discount: **Freepdfdumps**)

NEW QUESTION: 47

You have an Azure subscription that contains the resources shown in the following table.

Name	Type	Description
storage1	Azure Blob storage account	Contains publicly accessible TSV files that do NOT have a header row
WS1	Azure Synapse Analytics workspace	Contains a serverless SQL pool

You need to read the TSV files by using ad-hoc queries and the openrowset function. The solution must assign a name and override the inferred data type of each column.

What should you include in the openrowset function?

- A. the datafiletype bulk option
- B. the DATA_source parameter
- C. the rowsetoptions bulk option
- D. the with clause

Answer: (SHOW ANSWER)

NEW QUESTION: 48

You have an Azure data factory that has the Git repository settings shown in the following exhibit.

Git repository

Git repository information associated with your data factory. [CI/CD best practices](#)

Edit Overwrite live mode Disconnect Import resources

Repository type Azure DevOps Git

Azure DevOps Account

Project name ADFDeployDemo

Repository name ADFDeployDemo

Collaboration branch main

Publish branch adf_publish

Root folder /

Last published commit 23b144ac4aa7daf16f2fe7c2ab0eb303a8e4ed65

Publish (from ADF Studio) Enabled

Use the drop-down menus to select the answer choose that completes each statement based on the information presented in the graphic.

NOTE: Each correct answer is worth one point.

Answer Area

Changes to pipelines will be saved in Azure DevOps **[answer choice]**

To publish changes by using Azure Data Factory Studio, the changes must first be saved in the **[answer choice]**.

every 20 seconds
every 20 seconds
when the pipeline is published
when the pipeline is saved

root folder
adf_publish branch
main branch
root folder

Answer:

Answer Area

Changes to pipelines will be saved in Azure DevOps **[answer choice]**

To publish changes by using Azure Data Factory Studio, the changes must first be saved in the **[answer choice]**.

every 20 seconds
every 20 seconds
when the pipeline is published
when the pipeline is saved

root folder
adf_publish branch
main branch
root folder

Explanation:



NEW QUESTION: 49

You have an Azure subscription that contains an Azure Data Factory data pipeline named Pipeline1, a Log Analytics workspace named LA1, and a storage account named account1.

You need to retain pipeline-run data for 90 days. The solution must meet the following requirements:

- * The pipeline-run data must be removed automatically after 90 days.
- * Ongoing costs must be minimized.

Which two actions should you perform? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

- A. From the Diagnostic settings (classic) settings of account1, set the retention period to 90 days.
- B. Configure Pipeline1 to send logs to account1.
- C. Configure Pipeline1 to send logs to LA1.
- D. From the Data Retention settings of LA1, set the data retention period to 90 days.

Answer: A,C (LEAVE A REPLY)

NEW QUESTION: 50

You have an Azure Data Factory pipeline named Pipeline1!. Pipeline1 contains a copy activity that sends data to an Azure Data Lake Storage Gen2 account. Pipeline 1 is executed by a schedule trigger.

You change the copy activity sink to a new storage account and merge the changes into the collaboration branch.

After Pipeline1 executes, you discover that data is NOT copied to the new storage account.

You need to ensure that the data is copied to the new storage account.

What should you do?

- A. Publish from the collaboration branch.
- B. Configure the change feed of the new storage account.
- C. Create a pull request.
- D. Modify the schedule trigger.

Answer: A (LEAVE A REPLY)

CI/CD lifecycle

- * A development data factory is created and configured with Azure Repos Git. All developers should have permission to author Data Factory resources like pipelines and datasets.
- * A developer creates a feature branch to make a change. They debug their pipeline runs with their most recent changes
- * After a developer is satisfied with their changes, they create a pull request from their feature branch to the main or collaboration branch to get their changes reviewed by peers.

* After a pull request is approved and changes are merged in the main branch, the changes get published to the development factory.

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

NEW QUESTION: 51

You have a SQL pool in Azure Synapse that contains a table named `dbo.Customers`. The table contains a column named `Email`.

You need to prevent nonadministrative users from seeing the full email addresses in the `Email` column. The users must see values in a format of `aXXX@XXXX.com` instead.

What should you do?

- A. From Microsoft SQL Server Management Studio, set an email mask on the `Email` column.
- B. From the Azure portal, set a mask on the `Email` column.
- C. From Microsoft SQL Server Management studio, grant the `SELECT` permission to the users for all the columns in the `dbo.Customers` table except `Email`.
- D. From the Azure portal, set a sensitivity classification of `Confidential` for the `Email` column.

Answer: (SHOW ANSWER)

From Microsoft SQL Server Management Studio, set an email mask on the `Email` column. This is because "This feature cannot be set using portal for Azure Synapse (use PowerShell or REST API) or SQL Managed Instance." So use Create table statement with Masking e.g. `CREATE TABLE Membership (MemberID int IDENTITY PRIMARY KEY, FirstName varchar(100) MASKED WITH (FUNCTION = 'partial(1, "XXXXXXXX",0)') NULL, . . .` <https://docs.microsoft.com/en-us/azure/azure-sql/database/dynamic-data-masking- overview> upvoted 24 times

NEW QUESTION: 52

You are designing the folder structure for an Azure Data Lake Storage Gen2 container.

Users will query data by using a variety of services including Azure Databricks and Azure Synapse Analytics serverless SQL pools. The data will be secured by subject area. Most queries will include data from the current year or current month.

Which folder structure should you recommend to support fast queries and simplified folder security?

- A. `/{SubjectArea}/{DataSource}/{DD}/{MM}/{YYYY}/{FileData}_{YYYY}_{MM}_{DD}.csv`
- B. `/{DD}/{MM}/{YYYY}/{SubjectArea}/{DataSource}/{FileData}_{YYYY}_{MM}_{DD}.csv`
- C. `/{YYYY}/{MM}/{DD}/{SubjectArea}/{DataSource}/{FileData}_{YYYY}_{MM}_{DD}.csv`
- D. `/{SubjectArea}/{DataSource}/{YYYY}/{MM}/{DD}/{FileData}_{YYYY}_{MM}_{DD}.csv`

Answer: D (LEAVE A REPLY)

There's an important reason to put the date at the end of the directory structure. If you want to lock down certain regions or subject matters to users/groups, then you can easily do so with the POSIX permissions. Otherwise, if there was a need to restrict a certain security group to viewing just the UK data or certain planes, with the date structure in front a separate permission would be required for numerous directories under every hour directory. Additionally, having the date structure in front would exponentially increase the number of directories as time went on.

Note: In IoT workloads, there can be a great deal of data being landed in the data store that spans across numerous products, devices, organizations, and customers. It's important to pre-plan the directory layout for organization, security, and efficient processing of the data for down-stream consumers. A general template to consider might be the following layout:
{Region}/{SubjectMatter(s)}/{yyyy}/{mm}/{dd}/{hh}/

NEW QUESTION: 53

You need to schedule an Azure Data Factory pipeline to execute when a new file arrives in an Azure Data Lake Storage Gen2 container.

Which type of trigger should you use?

- A. on-demand
- B. tumbling window
- C. schedule
- D. storage event

Answer: (SHOW ANSWER)

Event-driven architecture (EDA) is a common data integration pattern that involves production, detection, consumption, and reaction to events. Data integration scenarios often require Data Factory customers to trigger pipelines based on events happening in storage account, such as the arrival or deletion of a file in Azure Blob Storage account.

Reference:

<https://docs.microsoft.com/en-us/azure/data-factory/how-to-create-event-trigger>

NEW QUESTION: 54

You are implementing a batch dataset in the Parquet format.

Data tiles will be produced by using Azure Data Factory and stored in Azure Data Lake Storage Gen2. The files will be consumed by an Azure Synapse Analytics serverless SQL pool.

You need to minimize storage costs for the solution.

What should you do?

- A. Store all the data as strings in the Parquet tiles.
- B. Use OPENROWSET to query the Parquet files.
- C. Create an external table that contains a subset of columns from the Parquet files.
- D. Use Snappy compression for the files.

Answer: (SHOW ANSWER)

An external table points to data located in Hadoop, Azure Storage blob, or Azure Data Lake Storage.

External tables are used to read data from files or write data to files in Azure Storage. With Synapse SQL, you can use external tables to read external data using dedicated SQL pool or serverless SQL pool.

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/develop-tables-external-tables>

NEW QUESTION: 55

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


Answer: C (LEAVE A REPLY)

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

NEW QUESTION: 56

You have an Azure Data Lake Storage Gen2 account that contains two folders named Folder and Folder2.

You use Azure Data Factory to copy multiple files from Folder1 to Folder2.



```
Operation failed: Failure happened on 'Sink' side.
ErrorCode=DelimitedTextMoreColumnsThanDefined,
'Type=Microsoft.DataTransfer.Common.Shared.HybridDeliveryException,
Message=Error found when processing 'Csv/Tsv Format Text' source
'@_2020_11_09_11_43_32.avro' with row number 53: found more columns
than expected column count 27., Source=Microsoft.DataTransfer.Common,'
```

You receive the following error.

What should you do to resolve the error.

- A. Add an explicit mapping.
- B. Enable fault tolerance to skip incompatible rows.
- C. Lower the degree of copy parallelism
- D. Change the Copy activity setting to Binary Copy

Answer: A (LEAVE A REPLY)

Reference: <https://knowledge.informatica.com/s/article/Microsoft-Azure-Data-Lake-Store-Gen2-target-file-names-not-generating-as-expected-using-create-at-runtime-option>

NEW QUESTION: 57

You are designing 2 solution that will use tables in Delta Lake on Azure Databricks.

You need to minimize how long it takes to perform the following:

*Queries against non-partitioned tables

* Joins on non-partitioned columns

Which two options should you include in the solution? Each correct answer presents part of the solution. (Choose Correct Answer and Give Explanation and References to Support the answers based from Data Engineering on Microsoft Azure)

- A. Z-Ordering
- B. Apache Spark caching
- C. dynamic file pruning (DFP)
- D. the clone command

Answer: A,C (LEAVE A REPLY)

According to the information I found on the web, two options that you should include in the solution to minimize how long it takes to perform queries and joins on non-partitioned tables are:

* Z-Ordering: This is a technique to colocate related information in the same set of files. This co-locality is automatically used by Delta Lake in data-skipping algorithms. This behavior dramatically reduces the amount of data that Delta Lake on Azure Databricks needs to read¹²³.

* Apache Spark caching: This is a feature that allows you to cache data in memory or on disk for faster access. Caching can improve the performance of repeated queries and joins on the same data. You can cache Delta tables using the `CACHE TABLE` or `CACHE LAZY` commands.

To minimize the time it takes to perform queries against non-partitioned tables and joins on non-partitioned columns in Delta Lake on Azure Databricks, the following options should be included in the solution:

A: Z-Ordering: Z-Ordering improves query performance by co-locating data that share the same column values in the same physical partitions. This reduces the need for shuffling data across nodes during query execution. By using Z-Ordering, you can avoid full table scans and reduce the amount of data processed.

B: Apache Spark caching: Caching data in memory can improve query performance by reducing the amount of data read from disk. This helps to speed up subsequent queries that need to access the same data. When you cache a table, the data is read from the data source and stored in memory. Subsequent queries can then read the data from memory, which is much faster than reading it from disk.

References:

* Delta Lake on Databricks: <https://docs.databricks.com/delta/index.html>

* Best Practices for Delta Lake on Databricks: <https://databricks.com/blog/2020/05/14/best-practices-for-delta-lake-on-databricks.html>

NEW QUESTION: 58

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.

P1:


 Microsoft	▼
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:


 Microsoft	▼
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	

Answer:

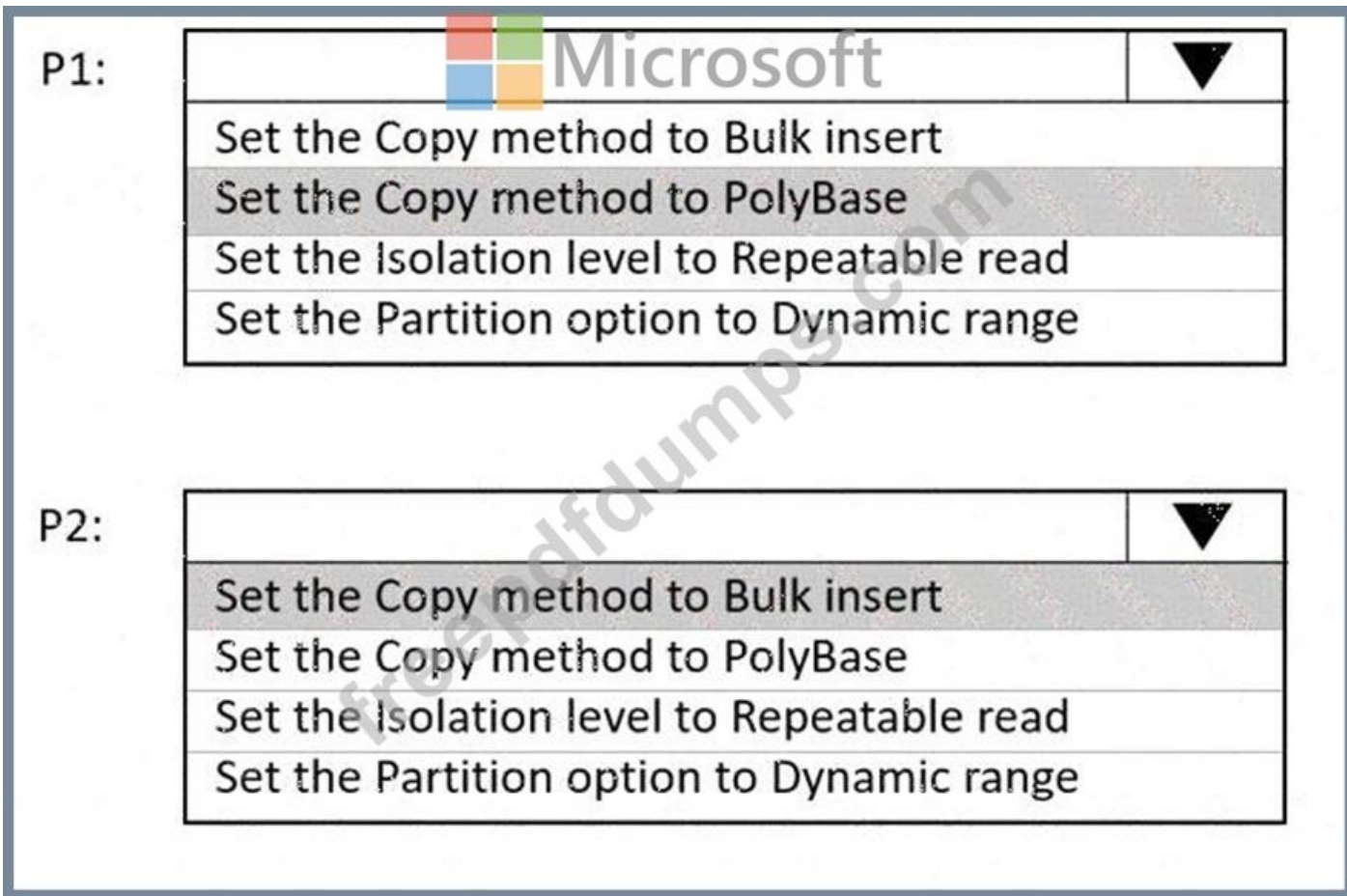
P1:

 Microsoft	▼
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:

 Microsoft	▼
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	

Explanation:



Box 1: Set the Copy method to PolyBase

While SQL pool supports many loading methods including non-Polybase options such as BCP and SQL BulkCopy API, the fastest and most scalable way to load data is through PolyBase. PolyBase is a technology that accesses external data stored in Azure Blob storage or Azure Data Lake Store via the T-SQL language.

Box 2: Set the Copy method to Bulk insert

Polybase not possible for text files. Have to use Bulk insert.

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/load-data-overview>

NEW QUESTION: 59

You have an Azure data factory that connects to a Microsoft Purview account. The data factory is registered in Microsoft Purview.

You update a Data Factory pipeline.

You need to ensure that the updated lineage is available in Microsoft Purview.

What should you do first?

- A. Execute the pipeline.
- B. Execute an Azure DevOps build pipeline.
- C. Locate the related asset in the Microsoft Purview portal.
- D. Disconnect the Microsoft Purview account from the data factory.

Answer: A (LEAVE A REPLY)

NEW QUESTION: 60

You have an Azure Databricks workspace named workspace1 in the Standard pricing tier. Workspace1 contains an all-purpose cluster named cluster1.

You need to reduce the time it takes for cluster 1 to start and scale up. The solution must minimize costs. What should you do first?

- A. Upgrade workspace1 to the Premium pricing tier.
- B. Create a cluster policy in workspace1.
- C. Create a pool in workspace1.
- D. Configure a global init script for workspace1.

Answer: (SHOW ANSWER)

You can use Databricks Pools to Speed up your Data Pipelines and Scale Clusters Quickly.

Databricks Pools, a managed cache of virtual machine instances that enables clusters to start and scale 4 times faster.

Reference:

<https://databricks.com/blog/2019/11/11/databricks-pools-speed-up-data-pipelines.html>

NEW QUESTION: 61

You have an Azure subscription that contains an Azure Blob Storage account named storage1 and an Azure Synapse Analytics dedicated SQL pool named Pool1.

You need to store data in storage1. The data will be read by Pool1. The solution must meet the following requirements:

- * Enable Pool1 to skip columns and rows that are unnecessary in a query.
- * Automatically create column statistics.
- * Minimize the size of files.

Which type of file should you use?

- A. JSON
- B. Parquet
- C. Avro
- D. CSV

Answer: B (LEAVE A REPLY)

Automatic creation of statistics is turned on for Parquet files. For CSV files, you need to create statistics manually until automatic creation of CSV files statistics is supported.

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/develop-tables-statistics>

Valid DP-203 Dumps shared by Actual4test.com for Helping Passing DP-203 Exam! Actual4test.com now offer the **newest DP-203 exam dumps**, the Actual4test.com DP-203 exam **questions have been updated** and **answers have been corrected** get the **newest** Actual4test.com DP-203 dumps with Test

Engine here: https://www.actual4test.com/DP-203_examcollection.html (365 Q&As Dumps, 30%OFF

Special Discount: **Freepdfdumps**)

NEW QUESTION: 62

You manage an enterprise data warehouse in Azure Synapse Analytics.

Users report slow performance when they run commonly used queries. Users do not report performance changes for infrequently used queries.

You need to monitor resource utilization to determine the source of the performance issues.

Which metric should you monitor?

- A. Data IO percentage
- B. Local tempdb percentage
- C. Cache hit percentage
- D. DWU percentage

Answer: C (LEAVE A REPLY)

Monitor and troubleshoot slow query performance by determining whether your workload is optimally leveraging the adaptive cache for dedicated SQL pools.

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-how-to-monitor-cache>

NEW QUESTION: 63

You have an Azure subscription that contains an Azure Data Lake Storage Gen2 account named storage1.

Storage1 contains a container named container1. Container1 contains a directory named directory1.

Directory1 contains a file named file1.

You have an Azure Active Directory (Azure AD) user named User1 that is assigned the Storage Blob Data Reader role for storage1.

You need to ensure that User1 can append data to file1. The solution must use the principle of least privilege.

Which permissions should you grant? To answer, drag the appropriate permissions to the correct resources.

Each permission 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.

Permissions

Read

Write

Execute

Answer Area

container1: Permission

directory1: Permission

file1: Permission

Answer:

The screenshot shows the 'Answer Area' with three permission boxes. The first box, labeled 'container1:', contains the word 'Execute'. The second box, labeled 'directory1:', also contains 'Execute'. The third box, labeled 'file1:', contains 'Write'. The 'Permissions' list on the left has 'Execute' selected, indicated by a red dashed border around the 'Execute' box.

Explanation:

Box 1: Execute

If you are granting permissions by using only ACLs (no Azure RBAC), then to grant a security principal read or write access to a file, you'll need to give the security principal Execute permissions to the root folder of the container, and to each folder in the hierarchy of folders that lead to the file.

Box 2: Execute

On Directory: Execute (X): Required to traverse the child items of a directory Box 3: Write On file: Write (W): Can write or append to a file.

Reference:

<https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-access-control>

NEW QUESTION: 64

you have a project in Azure DevOps that contains a repository named Repo1. Repo1 contains a branch named main.

You create a new Azure Synapse workspace named Workspace1.

You need to create data processing pipelines in Workspace1. The solution must meet the following requirements:

* Pipeline artifacts must be stored in Repo1.

- * Source control must be provided for pipeline artifacts.
- * All development must be performed in a feature branch.

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

The screenshot shows the Synapse Studio interface with two main sections: 'Actions' and 'Answer Area'. The 'Actions' list contains six items: 'Set the main branch as the collaboration branch.', 'Create pipeline artifacts and save them in the main branch.', 'Configure a code repository and select Repo1.', 'Create a new branch.', 'Create pipeline artifacts and save them in the new branch.', and 'Create a pull request to merge the contents of the main branch into the new branch.'. The 'Answer Area' is currently empty. Navigation arrows are visible on the right side of the interface.

Answer:

The screenshot shows the same Synapse Studio interface as above, but with four actions moved from the 'Actions' list to the 'Answer Area'. The actions in the 'Answer Area' are: 'Configure a code repository and select Repo1.', 'Create a new branch.', 'Create pipeline artifacts and save them in the new branch.', and 'Create a pull request to merge the contents of the main branch into the new branch.'. The 'Actions' list now contains three items that were not moved. Navigation arrows are visible on the right side of the interface.

Explanation:

The screenshot shows the Synapse Studio interface with the 'Actions' list containing two items: 'Set the main branch as the collaboration branch.' and 'Create pipeline artifacts and save them in the main branch.'. The 'Answer Area' contains four actions in a numbered sequence: 1. 'Configure a code repository and select Repo1.', 2. 'Create a new branch.', 3. 'Create pipeline artifacts and save them in the new branch.', and 4. 'Create a pull request to merge the contents of the main branch into the new branch.'. Navigation arrows are visible on the right side of the interface.

NEW QUESTION: 65

You have a Microsoft Entra tenant.

The tenant contains an Azure Data Lake Storage Gen2 account named storage! that has two containers named fs1 and fs2. You have a Microsoft Entra group named OdepartmentA. You need to meet the following requirements:

- * OdepartmentA must be able to read, write, and list all the files in fs1.
- * OdepartmentA must be prevented from accessing any files in fs2
- * The solution must use the principle of least privilege.

Which role should you assign to DepartmentA?

- A. Contributor for fs1
- B. Storage Blob Data Owner for fs1
- C. Storage Blob Data Contributor for fs1
- D. Storage Blob Data Contributor for storage1

Answer: C (LEAVE A REPLY)

NEW QUESTION: 66

You plan to create a dimension table in Azure Synapse Analytics that will be less than 1 GB. You need to create the table to meet the following requirements:

- * Provide the fastest Query time.

* Minimize data movement during queries.

Which type of table should you use?

- A. hash distributed
- B. heap
- C. replicated
- D. round-robin

Answer: C (LEAVE A REPLY)

A replicated table has a full copy of the table accessible on each Compute node. Replicating a table removes the need to transfer data among Compute nodes before a join or aggregation. Since the table has multiple copies, replicated tables work best when the table size is less than 2 GB compressed. 2 GB is not a hard limit.

Reference: <https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/design-guidance-for-replicated-tables>

NEW QUESTION: 67

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.

Answer: C (LEAVE A REPLY)

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.

- * In Azure DevOps, open the project that's configured with your data factory.
- * On the left side of the page, select Pipelines, and then select Releases.
- * Select New pipeline, or, if you have existing pipelines, select New and then New release pipeline.
- * In the Stage name box, enter the name of your environment.
- * 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`.
- * Select the Empty job template.

Reference:

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

NEW QUESTION: 68

You are developing an application that uses Azure Data Lake Storage Gen 2.

You need to recommend a solution to grant permissions to a specific application for a limited time period.

What should you include in the recommendation?

- A. Azure Active Directory (Azure AD) identities
- B. shared access signatures (SAS)
- C. account keys
- D. role assignments

Answer: (SHOW ANSWER)

A shared access signature (SAS) provides secure delegated access to resources in your storage account.

With a SAS, you have granular control over how a client can access your data. For example:

What resources the client may access.

What permissions they have to those resources.

How long the SAS is valid.

Reference:

<https://docs.microsoft.com/en-us/azure/storage/common/storage-sas-overview>

NEW QUESTION: 69

You need to create an Azure Data Factory pipeline to process data for the following three departments at your company: Ecommerce, retail, and wholesale. The solution must ensure that data can also be processed for the entire company.

How should you complete the Data Factory data flow script? 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.

The screenshot shows an Azure Data Factory data flow script editor. On the left, there is a 'Values' pane with the following options:

- all, ecommerce, retail, wholesale
- dept=='ecommerce', dept=='retail', dept=='wholesale'
- dept=='ecommerce', dept=='wholesale', dept=='retail'
- disjoint: false
- disjoint: true
- ecommerce, retail, wholesale, all

On the right, there is an 'Answer Area' with a script snippet:

```
cleanData split( [ ] )> SplitByDept@([ ])
```

The script has two empty boxes for the arguments of the split function and the SplitByDept@ function. A watermark 'freepdfdumps.com' is visible across the image.

Answer:

The screenshot shows a configuration for a 'split' transformation in a data pipeline. On the left, under 'Values', there are several rows defining conditions for department splits:

- all, ecommerce, retail, wholesale
- dept=='ecommerce', dept=='retail', dept=='wholesale'
- dept=='ecommerce', dept=='wholesale', dept=='retail'
- disjoint: false
- disjoint: true
- ecommerce, retail, wholesale, all

 On the right, under 'Answer Area', the DFDL code is shown:


```
CleanData
split(
  dept=='ecommerce', dept=='retail',
  dept=='wholesale'
  disjoint: false
) ~> SplitByDept@(ecommerce, retail, wholesale, all)
```

Explanation:

```
CleanData
split(
  dept=='ecommerce', dept=='retail',
  dept=='wholesale'
  disjoint: false
) ~> SplitByDept@(ecommerce, retail, wholesale, all);
```

The conditional split transformation routes data rows to different streams based on matching conditions. The conditional split transformation is similar to a CASE decision structure in a programming language. The transformation evaluates expressions, and based on the results, directs the data row to the specified stream.

Box 1: dept=='ecommerce', dept=='retail', dept=='wholesale'

First we put the condition. The order must match the stream labeling we define in Box 3.

Syntax:

```
<incomingStream>
split(
<conditionalExpression1>
<conditionalExpression2>
disjoint: {true | false}
) ~> <splitTx>@(stream1, stream2, ..., <defaultStream>)
```

Box 2: discount : false

disjoint is false because the data goes to the first matching condition. All remaining rows matching the third condition go to output stream all.

Box 3: ecommerce, retail, wholesale, all

Label the streams

Reference:

<https://docs.microsoft.com/en-us/azure/data-factory/data-flow-conditional-split>

NEW QUESTION: 70

You have an Azure subscription that contains an Azure data factory named ADF1. From Azure Data Factory Studio, you build a complex data pipeline in ADF1.

You discover that the Save button is unavailable and there are validation errors that prevent the pipeline from being published.

You need to ensure that you can save the logic of the pipeline.

Solution: You export ADF1 as an Azure Resource Manager (ARM) template.

A. No

B. Yes

Answer: A ([LEAVE A REPLY](#))

NEW QUESTION: 71

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
604	2776	636	48	2112	1	1
607	2704	576	48	2080	1	2
53	2376	512	16	1848	1	3
98	2376	512	16	1848	1	4
168	2632	528	32	2072	1	5
195	2896	536	32	2128	1	6
5995	3464	1424	32	2008	1	7
0	2232	496	0	1736	1	8
264	2576	544	48	1992	1	9
3008	3016	960	32	2024	1	10
-	-	-	-	-	-	-
1550	2832	752	48	2032	1	50
1258	2832	696	48	2088	1	51
192	2632	528	32	2072	1	52
1127	2768	528	48	2040	1	53
1244	3032	704	64	2264	1	54
609	2632	568	32	2032	1	55
0	2232	496	0	1736	1	56
1417	2832	728	48	2064	1	57
0	2232	496	0	1736	1	58
184	2632	568	32	2040	1	59
225	2768	544	48	2184	1	60

Which statement accurately describes the `dbo,FactInternetSales` table?

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

B. All distributions contain data.

C. The table is skewed.

D. The table uses round-robin distribution.

Answer: B ([LEAVE A REPLY](#))

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>

NEW QUESTION: 72

You plan to implement an Azure Data Lake Gen2 storage account.

You need to ensure that the data lake will remain available if a data center fails in the primary Azure region.

The solution must minimize costs.

Which type of replication should you use for the storage account?

- A. geo-redundant storage (GRS)
- B. zone-redundant storage (ZRS)
- C. geo-zone-redundant storage (GZRS)
- D. locally-redundant storage (LRS)

Answer: [\(SHOW ANSWER\)](#)

NEW QUESTION: 73

You have an Azure Synapse Analytics dedicated SQL pool named Pcol1. Pool1 contains a table named table1. You load 5 TB of data into table1.

You need to ensure that column store compression is maximized for table1.

Which statement should you execute?

- A. ALTER INDEX ALL on table REBUILD
- B. DBCC IIDEXDEFRAG (pool1, table1)
- C. DBCC DBREINDEX (table)
- D. ALTER INDEX ALL on table REORGANIZE

Answer: [C \(LEAVE A REPLY\)](#)

NEW QUESTION: 74

You are designing an Azure Stream Analytics job to process incoming events from sensors in retail environments.

You need to process the events to produce a running average of shopper counts during the previous 15 minutes, calculated at five-minute intervals.

Which type of window should you use?

- A. snapshot
- B. tumbling
- C. hopping
- D. sliding

Answer: [B \(LEAVE A REPLY\)](#)

Tumbling windows are a series of fixed-sized, non-overlapping and contiguous time intervals. The following diagram illustrates a stream with a series of events and how they are mapped into 10-second tumbling windows.

Tell me the count of tweets per time zone every 10 seconds



```
SELECT TimeZone, COUNT(*) AS Count
FROM TwitterStream TIMESTAMP BY CreatedAt
GROUP BY TimeZone, TumblingWindow(second,10)
```

Reference:

<https://docs.microsoft.com/en-us/stream-analytics-query/tumbling-window-azure-stream-analytics>

NEW QUESTION: 75

You have an Azure subscription that contains the resources shown in the following table.

Name	Type	Description
Workspace1	Azure Synapse workspace	Contains the Built-in serverless SQL pool
Pool1	Azure Synapse Analytics dedicated SQL pool	Deployed to Workspace1
storage1	Storage account	Hierarchical namespace enabled

The storage1 account contains a container named container1. The container1 container contains the following files.

```

Webdata <root folder>
  Monthly <folder>
    _monthly.csv
    Monthly.csv
  .testdata.csv
  testdata.csv

```

In Pool1, you run the following script.

```

CREATE EXTERNAL DATA SOURCE Ds1
WITH
(
  LOCATION = 'abfss://container1@storage1.dfs.core.windows.net' ,
  CREDENTIAL = credential1,
  TYPE = HADOOP
);

```

In the Built-in serverless SQL pool, you run the following script

```

CREATE EXTERNAL DATA SOURCE Ds2
WITH (
  LOCATION = 'https://storage1.blob.core.windows.net/container1/Webdata/',
  CREDENTIAL = credential2
);

```

For each of the following statements, select Yes if the statement is true. Otherwise, select No. NOTE: Each correct selection is worth one point.

Statements	Yes	No
An external table that uses Ds1 can read the _monthly.csv file.	<input type="radio"/>	<input type="radio"/>
An external table that uses Ds1 can read the Monthly.csv file.	<input type="radio"/>	<input type="radio"/>
An external table that uses Ds2 can read the .testdata.csv file.	<input type="radio"/>	<input type="radio"/>

Answer:

Statements	Yes	No
An external table that uses Ds1 can read the _monthly.csv file.	<input type="radio"/>	<input checked="" type="radio"/>
An external table that uses Ds1 can read the Monthly.csv file.	<input checked="" type="radio"/>	<input type="radio"/>
An external table that uses Ds2 can read the .testdata.csv file.	<input checked="" type="radio"/>	<input type="radio"/>

Explanation:

Answer Area

Statements	Yes	No
An external table that uses Ds1 can read the _monthly.csv file.	<input type="radio"/>	<input checked="" type="radio"/>
An external table that uses Ds1 can read the Monthly.csv file.	<input checked="" type="radio"/>	<input type="radio"/>
An external table that uses Ds2 can read the .testdata.csv file.	<input checked="" type="radio"/>	<input type="radio"/>

NEW QUESTION: 76

You have an Azure Databricks resource.

You need to log actions that relate to changes in compute for the Databricks resource.

Which Databricks services should you log?

- A. clusters
- B. workspace
- C. DBFS
- D. SSH
- E jobs

Answer: B (LEAVE A REPLY)

Cloud Provider Infrastructure Logs.Databricks logging allows security and admin teams to demonstrate conformance to data governance standards within or from a Databricks workspace. Customers, especially in the regulated industries, also need records on activities like:- User access control to cloud data storage- Cloud Identity and Access Management roles- User access to cloud network and compute Azure Databricks offers three distinct workloads on several VM Instances tailored for your data analytics workflow-the Jobs Compute and Jobs Light Compute workloads make it easy for data engineers to build and execute jobs, and the All-Purpose Compute workload makes it easy for data scientists to explore, visualize, manipulate, and share data and insights interactively.

Valid DP-203 Dumps shared by Actual4test.com for Helping Passing DP-203 Exam! Actual4test.com now offer the **newest DP-203 exam dumps**, the Actual4test.com DP-203 exam **questions have been updated** and **answers have been corrected** get the **newest** Actual4test.com DP-203 dumps with Test Engine here: https://www.actual4test.com/DP-203_examcollection.html (**365** Q&As Dumps, **30%OFF** Special Discount: **Freepdfdumps**)

NEW QUESTION: 77

You have an Azure Synapse workspace named MyWorkspace that contains an Apache Spark database named mytestdb.

You run the following command in an Azure Synapse Analytics Spark pool in MyWorkspace.

```
CREATE TABLE mytestdb.myParquetTable(  
EmployeeID int,  
EmployeeName string,
```

EmployeeStartDate date)

USING Parquet

You then use Spark to insert a row into mytestdb.myParquetTable. The row contains the following data.

EmployeeName	EmployeeID	EmployeeStartDate
Alice	24	2020-01-25

One minute later, you execute the following query from a serverless SQL pool in MyWorkspace.

```
SELECT EmployeeID  
FROM mytestdb.dbo.myParquetTable  
WHERE name = 'Alice';
```

What will be returned by the query?

- A. 24
- B. an error
- C. a null value

Answer: B (LEAVE A REPLY)

Once a database has been created by a Spark job, you can create tables in it with Spark that use Parquet as the storage format. Table names will be converted to lower case and need to be queried using the lower case name. These tables will immediately become available for querying by any of the Azure Synapse workspace Spark pools. They can also be used from any of the Spark jobs subject to permissions.

Note: For external tables, since they are synchronized to serverless SQL pool asynchronously, there will be a delay until they appear.

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/metadata/table>

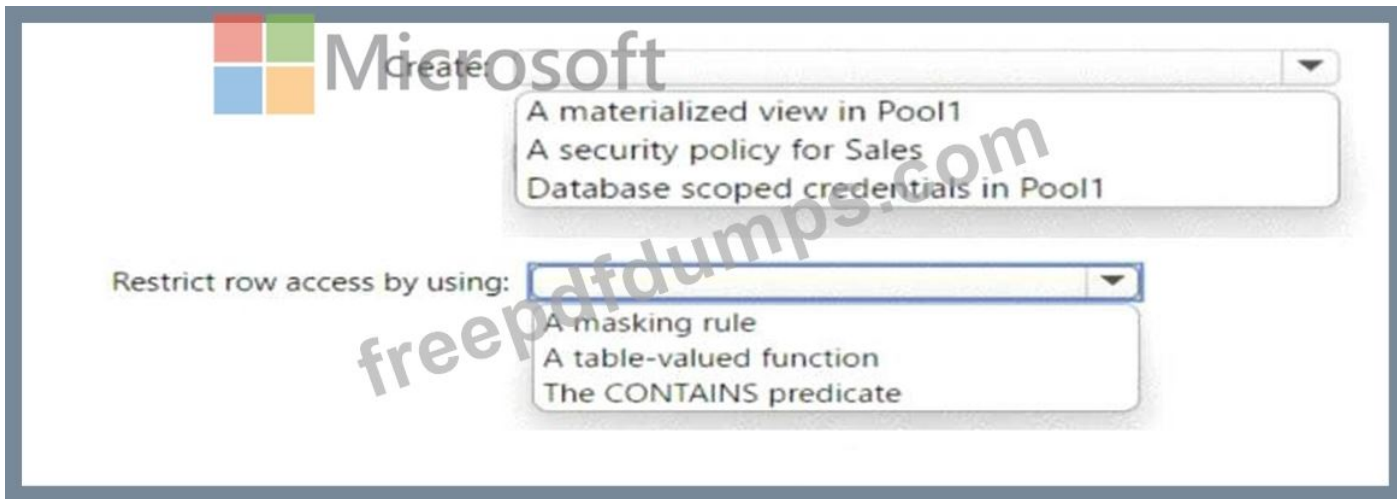
NEW QUESTION: 78

You have an Azure Synapse Analytics dedicated SQL pool named Pool1 that contains an external table named Sales. Sales contains sales data. Each row in Sales contains data on a single sale, including the name of the salesperson.

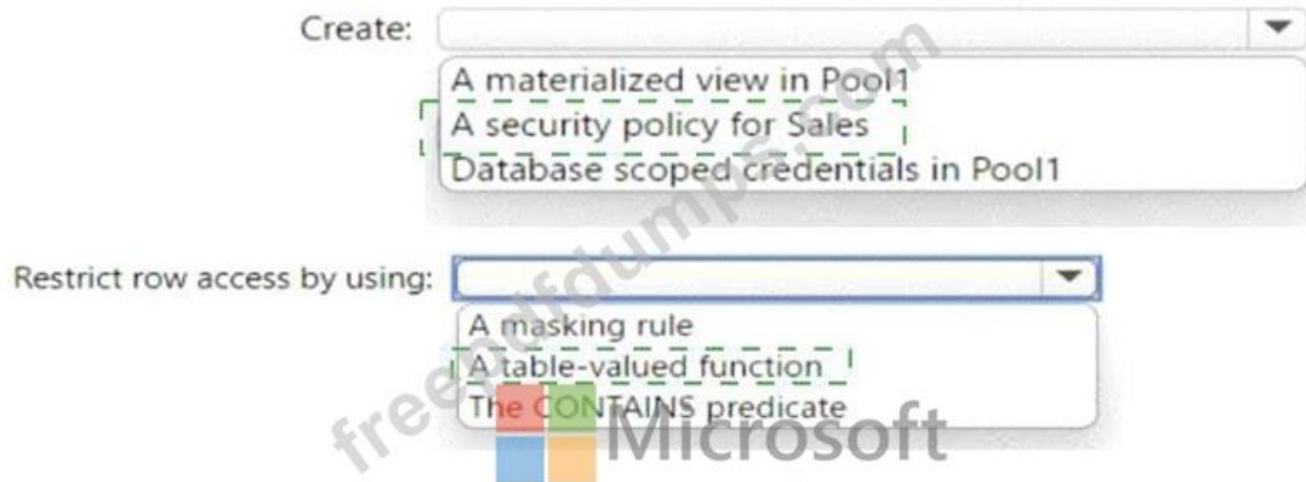
You need to implement row-level security (RLS). The solution must ensure that the salespeople can access only their respective sales.

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

NOTE: Each correct selection is worth one point.



Answer:



Explanation:

Box 1: A security policy for sale

Here are the steps to create a security policy for Sales:

- * Create a user-defined function that returns the name of the current user:
- * CREATE FUNCTION dbo.GetCurrentUser()
- * RETURNS NVARCHAR(128)
- * AS
- * BEGIN
- * RETURN SUSER_SNAME();
- * END;
- * Create a security predicate function that filters the Sales table based on the current user:
- * CREATE FUNCTION dbo.SalesPredicate(@salesperson NVARCHAR(128))
- * RETURNS TABLE
- * WITH SCHEMABINDING
- * AS
- * RETURN SELECT 1 AS access_result
- * WHERE @salesperson = SalespersonName;
- * Create a security policy on the Sales table that uses the SalesPredicate function to filter the data:
- * CREATE SECURITY POLICY SalesFilter

```
* ADD FILTER PREDICATE dbo.SalesPredicate(dbo.GetCurrentUser()) ON dbo.Sales  
* WITH (STATE = ON);
```

By creating a security policy for the Sales table, you ensure that each salesperson can only access their own sales data. The security policy uses a user-defined function to get the name of the current user and a security predicate function to filter the Sales table based on the current user.

Box 2: table-value function

to restrict row access by using row-level security, you need to create a table-valued function that returns a table of values that represent the rows that a user can access. You then use this function in a security policy that applies a predicate on the table.

NEW QUESTION: 79

You have an Azure Synapse Analytics workspace named WS1.

You have an Azure Data Lake Storage Gen2 container that contains JSON-formatted files in the following format.

```

{
  "id": "66532691-ab20-11ea-8b1d-936b3ec64e54",
  "context": {
    "data": {
      "eventTime": "2020-06-10T13:43:34.553Z",
      "samplingRate": "100.0",
      "isSynthetic": "false"
    },
    "session": {
      "isFirst": "false",
      "id": "38619c14-7a23-4687-8268-95862c5326b1"
    },
    "custom": {
      "dimensions": [
        {
          "customerInfo": {
            "ProfileType": "ExpertUser",
            "RoomName": "",
            "CustomerName": "diamond",
            "UserName": "XXXX@yahoo.com"
          }
        },
        {
          "customerInfo": {
            "ProfileType": "Novice",
            "RoomName": "",
            "CustomerName": "topaz",
            "UserName": "XXXX@outlook.com"
          }
        }
      ]
    }
  }
}

```



You need to use the serverless SQL pool in WS1 to read the files.

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.

Values

Answer Area

```

select*
FROM
 (
    BULK 'https://contoso.blob.core.windows.net/contosodw',
    FORMAT= 'CSV',
    fieldterminator = '0x0b',
    fieldquote = '0x0b',
    rowterminator = '0x0b'
)
with (id varchar(50),
contextdateeventTime varchar(50) '$.context.data.eventTime',
contextdatasamplingRate varchar(50) '$.context.data.samplingRate',
contextdataisSynthetic varchar(50) '$.context.data.isSynthetic',
contextsessionisFirst varchar(50) '$.context.session.isFirst',
contextsession varchar(50) '$.context.session.id',
contextcustomdimensions varchar(max) '$.context.custom.dimensions'
) as q
cross apply  (contextcustomdimensions)
with ( ProfileType varchar(50) '$.customerInfo.ProfileType',
RoomName varchar(50) '$.customerInfo.RoomName',
CustomerName varchar(50) '$.customerInfo.CustomerName',
UserName varchar(50) '$.customerInfo.UserName'
)

```

-
-
-
-

Answer:

Values

Answer Area

```

select*
FROM
 (
    BULK 'https://contoso.blob.core.windows.net/contosodw',
    FORMAT= 'CSV',
    fieldterminator = '0x0b',
    fieldquote = '0x0b',
    rowterminator = '0x0b'
)
with (id varchar(50),
contextdateeventTime varchar(50) '$.context.data.eventTime',
contextdatasamplingRate varchar(50) '$.context.data.samplingRate',
contextdataisSynthetic varchar(50) '$.context.data.isSynthetic',
contextsessionisFirst varchar(50) '$.context.session.isFirst',
contextsession varchar(50) '$.context.session.id',
contextcustomdimensions varchar(max) '$.context.custom.dimensions'
) as q
cross apply  (contextcustomdimensions)
with ( ProfileType varchar(50) '$.customerInfo.ProfileType',
RoomName varchar(50) '$.customerInfo.RoomName',
CustomerName varchar(50) '$.customerInfo.CustomerName',
UserName varchar(50) '$.customerInfo.UserName'
)

```

-
-
-
-

Explanation:

```

select
FROM
  openrowset (
    BULK 'https://contoso.blob.core.windows.net/contososodw',
    FORMAT= 'CSV',
    fieldterminator = '0x0b',
    fieldquote = '0x0b',
    rowterminator = '0x0b'
  )
with (id varchar(50),
      contextdateeventTime varchar(50) '$.context.data.eventTime',
      contextdatasamplingRate varchar(50) '$.context.data.samplingRate',
      contextdataisSynthetic varchar(50) '$.context.data.isSynthetic',
      contextsessionisFirst varchar(50) '$.context.session.isFirst',
      contextsession varchar(50) '$.context.session.id',
      contextcustomdimensions varchar(max) '$.context.custom.dimensions'
) as q
cross apply openjson (contextcustomdimensions)

with ( ProfileType varchar(50) '$.customerInfo.ProfileType',
      RoomName varchar(50) '$.customerInfo.RoomName',
      CustomerName varchar(50) '$.customerInfo.CustomerName',
      UserName varchar(50) '$.customerInfo.UserName'
)

```

Box 1: openrowset

The easiest way to see to the content of your CSV file is to provide file URL to OPENROWSET function, specify csv FORMAT.

Example:

```

SELECT *
FROM OPENROWSET(
  BULK 'csv/population/population.csv',
  DATA_SOURCE = 'SqlOnDemandDemo',
  FORMAT = 'CSV', PARSER_VERSION = '2.0',
  FIELDTERMINATOR = ',',
  ROWTERMINATOR = '\n'
)

```

Box 2: openjson

You can access your JSON files from the Azure File Storage share by using the mapped drive, as shown in the following example:

```

SELECT book.* FROM
  OPENROWSET(BULK N't:\books\books.json', SINGLE_CLOB) AS json
CROSS APPLY OPENJSON(BulkColumn)
WITH( id nvarchar(100), name nvarchar(100), price float,
      pages_i int, author nvarchar(100)) AS book

```

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/query-single-csv-file>

NEW QUESTION: 80

You have an Azure Data Factory instance that contains two pipelines named Pipeline1 and Pipeline2. Pipeline1 has the activities shown in the following exhibit.



Pipeline2 has the activities shown in the following exhibit.



You execute Pipeline2, and Stored procedure1 in Pipeline1 fails.

What is the status of the pipeline runs?

- A. Pipeline1 and Pipeline2 succeeded.
- B. Pipeline1 and Pipeline2 failed.
- C. Pipeline1 succeeded and Pipeline2 failed.
- D. Pipeline1 failed and Pipeline2 succeeded.

Answer: A (LEAVE A REPLY)

Activities are linked together via dependencies. A dependency has a condition of one of the following: Succeeded, Failed, Skipped, or Completed.

Consider Pipeline1:

If we have a pipeline with two activities where Activity2 has a failure dependency on Activity1, the pipeline will not fail just because Activity1 failed. If Activity1 fails and Activity2 succeeds, the pipeline will succeed.

This scenario is treated as a try-catch block by Data Factory.



The failure dependency means this pipeline reports success.

Note:

If we have a pipeline containing Activity1 and Activity2, and Activity2 has a success dependency on Activity1, it will only execute if Activity1 is successful. In this scenario, if Activity1 fails, the pipeline will fail.

Reference:

<https://datasavvy.me/category/azure-data-factory/>

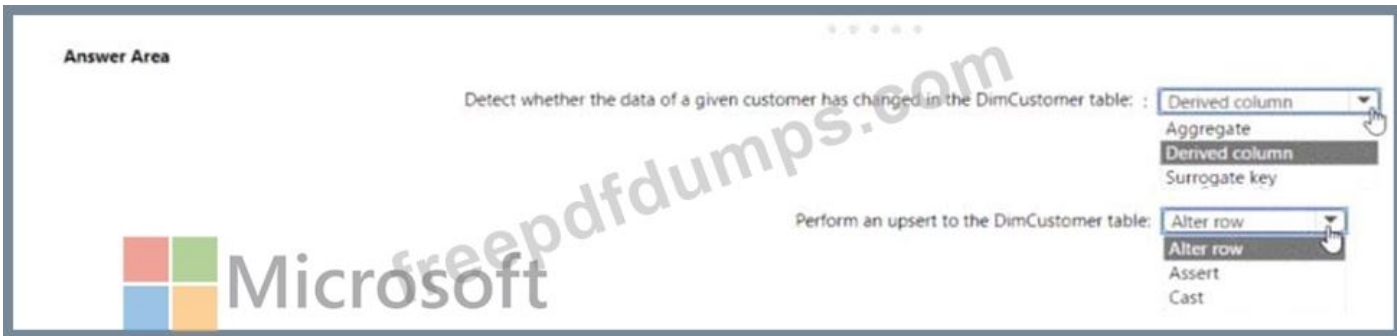
NEW QUESTION: 81

You are developing an Azure Synapse Analytics pipeline that will include a mapping data flow named Dataflow1. Dataflow1 will read customer data from an external source and use a Type 1 slowly changing dimension (SCD) when loading the data into a table named DimCustomer1 in an Azure Synapse Analytics dedicated SQL pool.

You need to ensure that Dataflow1 can perform the following tasks:

- * Detect whether the data of a given customer has changed in the DimCustomer table.
- * Perform an upsert to the DimCustomer table.

Which type of transformation should you use for each task? To answer, select the appropriate options in the answer area NOTE; Each correct selection is worth one point.



Answer:



Explanation:



NEW QUESTION: 82

You have an Azure Synapse Analytics dedicated SQL Pool1. Pool1 contains a partitioned fact table named dbo.Sales and a staging table named stg.Sales that has the matching table and partition definitions.

You need to overwrite the content of the first partition in dbo.Sales with the content of the same partition in stg.Sales. The solution must minimize load times.

What should you do?

- A. Switch the first partition from dbo.Sales to stg.Sales.
- B. Insert the data from stg.Sales into dbo.Sales.
- C. Update dbo.Sales from stg.Sales.
- D. Switch the first partition from stg.Sales to dbo. Sales.

Answer: A (LEAVE A REPLY)

NEW QUESTION: 83

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.

Load methodology:	<div style="display: flex; align-items: center;"><div style="width: 20px; height: 20px; background-color: #f08080; margin-right: 5px;"></div><div style="width: 20px; height: 20px; background-color: #90ee90; margin-right: 5px;"></div><div style="width: 20px; height: 20px; background-color: #ff8c00; margin-right: 5px;"></div><div style="width: 20px; height: 20px; background-color: #6495ed; margin-right: 5px;"></div><div style="font-size: 24px; margin-right: 10px;">Microsoft</div><div style="border: 1px solid black; flex-grow: 1; padding: 2px;">▼</div></div>
	Full Load
	Incremental Load
	Load individual files as they arrive
Trigger:	<div style="border: 1px solid black; padding: 2px;">▼</div>
	Fixed schedule
	New file
	Tumbling window

Answer:

Load methodology:	▼	
	Full Load	
	Incremental Load	
Trigger:	▼	
	Fixed schedule	
	New file	
	Tumbling window	

Explanation:

Load methodology:	▼	
	Full Load	
	Incremental Load	
Trigger:	▼	
	Fixed schedule	
	New file	
	Tumbling window	

Box 1: Incremental load

Box 2: Tumbling window

Tumbling windows are a series of fixed-sized, non-overlapping and contiguous time intervals. The following diagram illustrates a stream with a series of events and how they are mapped into 10-second tumbling windows.

Tell me the count of tweets per time zone every 10 seconds



```
SELECT TimeZone, COUNT(*) AS Count
FROM TwitterStream TIMESTAMP BY CreatedAt
GROUP BY TimeZone, TumblingWindow(second,10)
```

Reference:

<https://docs.microsoft.com/en-us/stream-analytics-query/tumbling-window-azure-stream-analytics>

NEW QUESTION: 84

You have an Azure Synapse Analytics dedicated SQL pool named Pool1 and an Azure Data Lake Storage Gen2 account named Account1.

You plan to access the files in Account1 by using an external table.

You need to create a data source in Pool1 that you can reference when you create the external table.

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.

```
CREATE EXTERNAL DATA SOURCE source1
WITH
( LOCATION = 'https://account1. .core.windows.net',
PUSHDOWN = ON
TYPE = BLOB_STORAGE
TYPE = HADOOP
)
```

Answer:

```
CREATE EXTERNAL DATA SOURCE sourcel
WITH
  ( LOCATION = 'https://account1. .core.windows.net',
    PUSHDOWN = ON
    TYPE = BLOB_STORAGE
    TYPE = HADOOP
  )
```

Explanation:

```
CREATE EXTERNAL DATA SOURCE sourcel
WITH
  ( LOCATION = 'https://account1. .core.windows.net',
    PUSHDOWN = ON
    TYPE = BLOB_STORAGE
    TYPE = HADOOP
  )
```

Box 1: blob

The following example creates an external data source for Azure Data Lake Gen2 CREATE EXTERNAL DATA SOURCE YellowTaxi WITH (LOCATION = 'https://azureopendatastorage.blob.core.windows.net/nyctlc/yellow/', TYPE = HADOOP) Box 2: HADOOP

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/develop-tables-external-tables>

NEW QUESTION: 85

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

You have files that are ingested and loaded into an Azure Data Lake Storage Gen2 container named container1.

You plan to insert data from the files into Table1 and azure Data Lake Storage Gen2 container named container1.

You plan to insert data from the files into Table1 and transform the data. Each row of data in the files will produce one row in the serving layer of Table1.

You need to ensure that when the source data files are loaded to container1, the DateTime is stored as an additional column in Table1.

Solution: In an Azure Synapse Analytics pipeline, you use a Get Metadata activity that retrieves the DateTime of the files.

Does this meet the goal?

- A. Yes
- B. No

Answer: B (LEAVE A REPLY)

Instead use a serverless SQL pool to create an external table with the extra column.

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/create-use-external-tables>

NEW QUESTION: 86

You are creating an Azure Data Factory data flow that will ingest data from a CSV file, cast columns to specified types of data, and insert the data into a table in an Azure Synapse Analytics dedicated SQL pool. The CSV file contains columns named username, comment and date.

The data flow already contains the following:

- * A source transformation
- * A Derived Column transformation to set the appropriate types of data
- * A sink transformation to land the data in the pool

You need to ensure that the data flow meets the following requirements;

- * All valid rows must be written to the destination table.
- * Truncation errors in the comment column must be avoided proactively.
- * Any rows containing comment values that will cause truncation errors upon insert must be written to a file in blob storage.

Which two actions should you perform? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point

- A.** Add a select transformation that selects only the rows which will cause truncation errors.
- B.** Add a sink transformation that writes the rows to a file in blob storage.
- C.** Add a Conditional Split transformation that separates the rows which will cause truncation errors.
- D.** Add a filter transformation that filters out rows which will cause truncation errors.

Answer: B,C (LEAVE A REPLY)

NEW QUESTION: 87

You plan to implement an Azure Data Lake Storage Gen2 container that will contain CSV files. The size of the files will vary based on the number of events that occur per hour.

File sizes range from 4.KB to 5 GB.

You need to ensure that the files stored in the container are optimized for batch processing.

What should you do?

- A.** Compress the files.
- B.** Merge the files.
- C.** Convert the files to JSON
- D.** Convert the files to Avro.

Answer: D (LEAVE A REPLY)

Avro supports batch and is very relevant for streaming.

Note: Avro is framework developed within Apache's Hadoop project. It is a row-based storage format which is widely used as a serialization process. AVRO stores its schema in JSON format making it easy to

read and interpret by any program. The data itself is stored in binary format by doing it compact and efficient.

Reference:

<https://www.adaltas.com/en/2020/07/23/benchmark-study-of-different-file-format/>

NEW QUESTION: 88

You have an Apache Spark DataFrame named temperatures. A sample of the data is shown in the following table.

Date	Temp
...	...
18-01-2021	3
19-01-2021	4
20-01-2021	2
21-01-2021	2
...	...

You need to produce the following table by using a Spark SQL query.

Year	JAN	FEB	MAR	APR	MAY
2019	2.3	4.1	5.2	7.6	9.2
2020	2.4	4.2	4.9	7.8	9.1
2021	2.6	5.3	3.4	7.9	9.5

How should you complete the query? 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.

Values Answer Area

```
SELECT * FROM (  
  SELECT YEAR(Date) Year, MONTH(Date) Month, Temp  
  FROM temperatures  
  WHERE date BETWEEN DATE '2019-01-01' AND DATE '2021-08-31'  
)  
(  
  AVG ( [ ] (Temp AS DECIMAL(4, 1)))  
  FOR Month in (  
    1 JAN, 2 FEB, 3 MAR, 4 APR, 5 MAY, 6 JUN,  
    7 JUL, 8 AUG, 9 SEP, 10 OCT, 11 NOV, 12 DEC  
  )  
)  
ORDER BY Year ASC
```

- CAST
- COLLATE
- CONVERT
- FLATTEN
- PIVOT
- UNPIVOT

Answer:

Values Answer Area

```
SELECT * FROM (  
  SELECT YEAR(Date) Year, MONTH(Date) Month, Temp  
  FROM temperatures  
  WHERE date BETWEEN DATE '2019-01-01' AND DATE '2021-08-31'  
)  
(  
  PIVOT (  
    AVG ( CAST (Temp AS DECIMAL(4, 1)))  
  FOR Month in (  
    1 JAN, 2 FEB, 3 MAR, 4 APR, 5 MAY, 6 JUN,  
    7 JUL, 8 AUG, 9 SEP, 10 OCT, 11 NOV, 12 DEC  
  )  
)  
ORDER BY Year ASC
```

- CAST
- COLLATE
- CONVERT
- FLATTEN
- PIVOT
- UNPIVOT

Explanation:

```

SELECT * FROM (
  SELECT YEAR(Date) Year, MONTH(Date) Month, Temp
  FROM temperatures
  WHERE date BETWEEN DATE '2019-01-01' AND DATE '2021-08-31'
  PIVOT (
    AVG ( CAST (Temp AS DECIMAL(4, 1)))
  FOR Month in (
    1 JAN, 2 FEB, 3 MAR, 4 APR, 5 MAY, 6 JUN,
    7 JUL, 8 AUG, 9 SEP, 10 OCT, 11 NOV, 12 DEC
    )
  )
ORDER BY Year ASC

```

Box 1: PIVOT

PIVOT rotates a table-valued expression by turning the unique values from one column in the expression into multiple columns in the output. And PIVOT runs aggregations where they're required on any remaining column values that are wanted in the final output.

Reference:

<https://learnsql.com/cookbook/how-to-convert-an-integer-to-a-decimal-in-sql-server/>

<https://docs.microsoft.com/en-us/sql/t-sql/queries/from-using-pivot-and-unpivot>

NEW QUESTION: 89

You have an Azure Databricks workspace named workspace1 in the Standard pricing tier.

You need to configure workspace1 to support autoscaling all-purpose clusters. The solution must meet the following requirements:

- * Automatically scale down workers when the cluster is underutilized for three minutes.
- * Minimize the time it takes to scale to the maximum number of workers.
- * Minimize costs.

What should you do first?

- A. Enable container services for workspace1.
- B. Upgrade workspace1 to the Premium pricing tier.
- C. Set Cluster Mode to High Concurrency.
- D. Create a cluster policy in workspace1.

Answer: B (LEAVE A REPLY)

For clusters running Databricks Runtime 6.4 and above, optimized autoscaling is used by all-purpose clusters in the Premium plan Optimized autoscaling:

Scales up from min to max in 2 steps.

Can scale down even if the cluster is not idle by looking at shuffle file state.

Scales down based on a percentage of current nodes.

On job clusters, scales down if the cluster is underutilized over the last 40 seconds.

On all-purpose clusters, scales down if the cluster is underutilized over the last 150 seconds.

The spark.databricks.aggressiveWindowDownS Spark configuration property specifies in seconds how often a cluster makes down-scaling decisions. Increasing the value causes a cluster to scale down more slowly. The maximum value is 600.

Note: Standard autoscaling

Starts with adding 8 nodes. Thereafter, scales up exponentially, but can take many steps to reach the max. You can customize the first step by setting the spark.databricks.autoscaling.standardFirstStepUp Spark configuration property.

Scales down only when the cluster is completely idle and it has been underutilized for the last 10 minutes.

Scales down exponentially, starting with 1 node.

Reference:

<https://docs.databricks.com/clusters/configure.html>

NEW QUESTION: 90

You have an Azure Stream Analytics job named Job1.

The metrics of Job1 from the last hour are shown in the following table.

Metric	Time aggregation	Value
SU (Memory) % Utilization	Average	70
CPU % Utilization	Average	20
Runtime Errors	Total	0
Watermark Delay	Average	25
Input Deserialization Errors	Total	0

The late arrival tolerance for Job1 is set to the five seconds.

You need to optimize Job1.

Which two actions achieve the goal? Each correct answer presents a complete solution.

NOTE: Each correct answer is worth one point.

- A. Resolution errors in inputs processing.
- B. Parallelize the query
- C. Increase the number of SUs.
- D. Resolution errors in output processing

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 91

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 are designing an Azure Stream Analytics solution that will analyze Twitter data.

You need to count the tweets in each 10-second window. The solution must ensure that each tweet is counted only once.

Solution: You use a tumbling window, and you set the window size to 10 seconds.

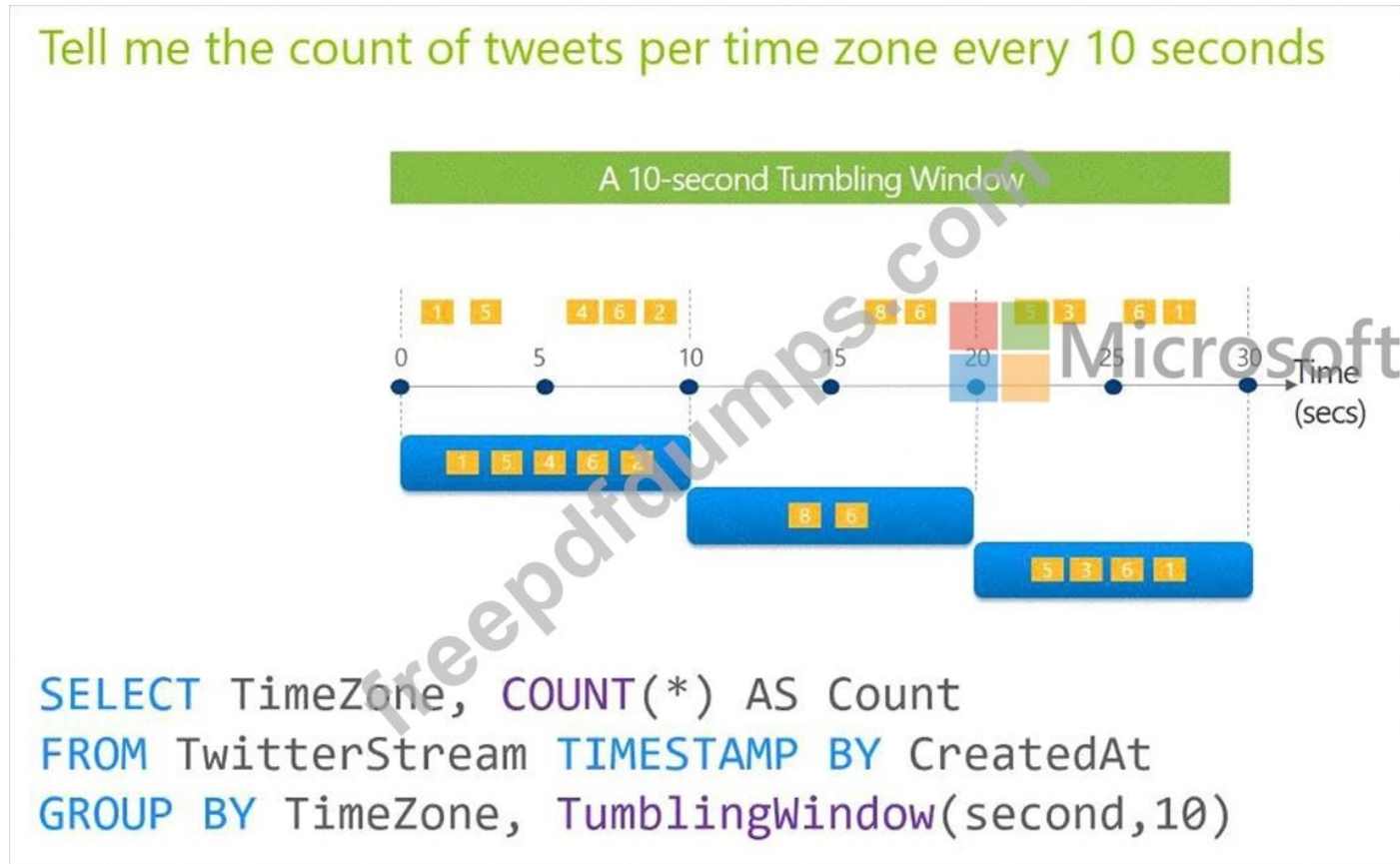
Does this meet the goal?

A. Yes

B. No

Answer: A ([LEAVE A REPLY](#))

Tumbling windows are a series of fixed-sized, non-overlapping and contiguous time intervals. The following diagram illustrates a stream with a series of events and how they are mapped into 10-second tumbling windows.



Reference:

<https://docs.microsoft.com/en-us/stream-analytics-query/tumbling-window-azure-stream-analytics>

Valid DP-203 Dumps shared by Actual4test.com for Helping Passing DP-203 Exam! Actual4test.com now offer the **newest DP-203 exam dumps**, the Actual4test.com DP-203 exam **questions have been updated** and **answers have been corrected** get the **newest** Actual4test.com DP-203 dumps with Test Engine here: https://www.actual4test.com/DP-203_examcollection.html (365 Q&As Dumps, **30%OFF** Special Discount: **Freepdfdumps**)

NEW QUESTION: 92

You are designing an Azure Databricks cluster that runs user-defined local processes. You need to recommend a cluster configuration that meets the following requirements:

* Minimize query latency.

* Maximize the number of users that can run queues on the cluster at the same time Reduce overall costs without compromising other requirements Which cluster type should you recommend?

- A. Standard with Auto termination
- B. Standard with Autoscaling
- C. High Concurrency with Autoscaling
- D. High Concurrency with Auto Termination

Answer: C (LEAVE A REPLY)

A High Concurrency cluster is a managed cloud resource. The key benefits of High Concurrency clusters are that they provide fine-grained sharing for maximum resource utilization and minimum query latencies. Databricks chooses the appropriate number of workers required to run your job. This is referred to as autoscaling. Autoscaling makes it easier to achieve high cluster utilization, because you don't need to provision the cluster to match a workload.

Reference:

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

NEW QUESTION: 93

You have a SQL pool in Azure Synapse.

You discover that some queries fail or take a long time to complete.

You need to monitor for transactions that have rolled back.

Which dynamic management view should you query?

- A. sys.dm_pdw_request_steps
- B. sys.dm_pdw_nodes_tran_database_transactions
- C. sys.dm_pdw_waits
- D. sys.dm_pdw_exec_sessions

Answer: (SHOW ANSWER)

You can use Dynamic Management Views (DMVs) to monitor your workload including investigating query execution in SQL pool.

If your queries are failing or taking a long time to proceed, you can check and monitor if you have any transactions rolling back.

Example:

```
-- Monitor rollback
```

```
SELECT
```

```
SUM(CASE WHEN t.database_transaction_next_undo_lsn IS NOT NULL THEN 1 ELSE 0 END),
```

```
t.pdw_node_id, nod.[type] FROM sys.dm_pdw_nodes_tran_database_transactions t JOIN
```

```
sys.dm_pdw_nodes nod ON t.pdw_node_id = nod.pdw_node_id GROUP BY t.pdw_node_id, nod.[type]
```

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-manage-monitor#monitor-transaction-log-rollback>

NEW QUESTION: 94

You have an Azure Data Lake Storage Gen 2 account named storage1.

You need to recommend a solution for accessing the content in storage1. The solution must meet the following requirements:

- * List and read permissions must be granted at the storage account level.
- * Additional permissions can be applied to individual objects in storage1.
- * Security principals from Microsoft Azure Active Directory (Azure AD), part of Microsoft Entra, must be used for authentication.

What should you use? To answer, drag the appropriate components to the correct requirements. Each component 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.

The screenshot shows a drag-and-drop interface with two main panes. On the left, under the heading "Components", there are four items: "Access control lists (ACLs)", "Role-based access control (RBAC) roles", "Shared access signatures (SAS)", and "Shared account keys". On the right, under the heading "Answer Area", there are two empty text boxes. The first box is labeled "To grant permissions at the storage account level:" and the second box is labeled "To grant permissions at the object level:". A Microsoft logo is visible in the top right corner of the interface.

Answer:

The screenshot shows the same drag-and-drop interface as above, but with the correct answers placed in the boxes. In the "Components" pane, "Role-based access control (RBAC) roles" and "Access control lists (ACLs)" are highlighted with dashed green borders. In the "Answer Area", the box for "To grant permissions at the storage account level:" contains "Role-based access control (RBAC) roles" and the box for "To grant permissions at the object level:" contains "Access control lists (ACLs)". A Microsoft logo is visible in the bottom right corner of the interface.

Explanation:

Box 1: Role-based access control (RBAC) roles

List and read permissions must be granted at the storage account level.

Security principals from Microsoft Azure Active Directory (Azure AD), part of Microsoft Entra, must be used for authentication.

Role-based access control (Azure RBAC)

Azure RBAC uses role assignments to apply sets of permissions to security principals. A security principal is an object that represents a user, group, service principal, or managed identity that is defined in Azure Active Directory (AD). A permission set can give a security principal a "coarse-grain" level of access such as read or write access to all of the data in a storage account or all of the data in a container.

Box 2: Access control lists (ACLs)

Additional permissions can be applied to individual objects in storage1.

Access control lists (ACLs)

ACLs give you the ability to apply "finer grain" level of access to directories and files. An ACL is a permission construct that contains a series of ACL entries. Each ACL entry associates security principal with an access level.

Reference: <https://learn.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-access-control-model>

NEW QUESTION: 95

You have an Azure subscription that contains a Microsoft Purview account named MP1, an Azure data factory named DF1, and a storage account named storage1. MP1 is configured to scan storage1. DF1 is connected to MP1 and contains 3 datasets named DS1. DS1 references 2 files in storage1.

In DF1, you plan to create a pipeline that will process data from DS1.

You need to review the schema and lineage information in MP1 for the data referenced by DS1.

Which two features can you use to locate the information? Each correct answer presents a complete solution.

NOTE: Each correct answer is worth one point.

- A. the Storage browser of storage1 in the Azure portal
- B. the search bar in the Azure portal
- C. the search bar in Azure Data Factory Studio
- D. the search bar in the Microsoft Purview governance portal

Answer: A,D (LEAVE A REPLY)

* The search bar in the Microsoft Purview governance portal: This is a feature that allows you to search for assets in your data estate using keywords, filters, and facets. You can use the search bar to find the files in storage1 that are referenced by DS1, and then view their schema and lineage information in the asset details page¹².

* The search bar in Azure Data Factory Studio: This is a feature that allows you to search for datasets, linked services, pipelines, and other resources in your data factory. You can use the search bar to find DS1 in DF1, and then view its schema and lineage information in the dataset details page. You can also click on the Open in Purview button to open the corresponding asset in MP1¹³.

The two features that can be used to locate the schema and lineage information for the data referenced by DS1 are the search bar in Azure Data Factory Studio and the search bar in the Microsoft Purview governance portal.

The search bar in Azure Data Factory Studio allows you to search for the dataset DS1 and view its properties and lineage. This can help you locate information about the source and destination data stores, as well as the transformations that were applied to the data.

The search bar in the Microsoft Purview governance portal allows you to search for the storage account and view its metadata, including schema and lineage information. This can help you understand the different data assets that are stored in the storage account and how they are related to each other.

The Storage browser of storage1 in the Azure portal may allow you to view the files that are stored in the storage account, but it does not provide lineage or schema information for those files. Similarly, the search bar in the Azure portal may allow you to search for resources in the Azure subscription, but it does not provide detailed information about the data assets themselves.

References:

- * What is Azure Purview?
- * Use Azure Data Factory Studio

NEW QUESTION: 96

You are implementing an Azure Stream Analytics solution to process event data from devices. The devices output events when there is a fault and emit a repeat of the event every five seconds until the fault is resolved. The devices output a heartbeat event every five seconds after a previous event if there are no faults present.

A sample of the events is shown in the following table.

DeviceID	EventType	EventTime
78cc5ht9-w357-684r-w4fr-kr16h6p9874e	HeartBeat	2020-12-01T19:00.000Z
78cc5ht9-w357-684r-w4fr-kr16h6p9874e	HeartBeat	2020-12-01T19:05.000Z
78cc5ht9-w357-684r-w4fr-kr16h6p9874e	TemperatureSensorFault	2020-12-01T19:07.000Z

You need to calculate the uptime between the faults.

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

NOTE: Each correct selection is worth one point.

```

SELECT
DeviceID,
MIN(EventTime) as StartTime,
MAX(EventTime) as EndTime,
DATEDIFF(second, MIN(EventTime), MAX(EventTime)) AS duration_in_seconds
FROM input TIMESTAMP BY EventTime

```

▼

WHERE EventType='HeartBeat'

WHERE LAG(EventType, 1) OVER (LIMIT DURATION(second,5)) <> EventType

WHERE IsFirst(second,5) = 1

```

GROUP BY
DeviceID

```

▼

,SessionWindow(second, 5, 50000) OVER (PARTITION BY DeviceID)

,TumblingWindow(second,5)

HAVING DATEDIFF(second, MIN(EventTime), MAX(EventTime)) > 5

Answer:

```

SELECT
DeviceID,
MIN(EventTime) as StartTime,
MAX(EventTime) as EndTime,
DATEDIFF(second, MIN(EventTime), MAX(EventTime)) AS duration_in_seconds
FROM input TIMESTAMP BY EventTime

```

```

WHERE EventType='HeartBeat'
WHERE LAG(EventType, 1) OVER (LIMIT DURATION(second,5)) <> EventType
WHERE IsFirst(second,5) = 1

```

```

GROUP BY

```

```

DeviceID

```

```

,SessionWindow(second, 5, 50000) OVER (PARTITION BY DeviceID)
,TumblingWindow(second,5)
HAVING DATEDIFF(second, MIN(EventTime), MAX(EventTime)) > 5

```

Explanation:

```

SELECT

```

```

DeviceID,

```

```

MIN(EventTime) as StartTime,

```

```

MAX(EventTime) as EndTime,

```

```

DATEDIFF(second, MIN(EventTime), MAX(EventTime)) AS duration_in_seconds

```

```

FROM input TIMESTAMP BY EventTime

```

```

WHERE EventType='HeartBeat'
WHERE LAG(EventType, 1) OVER (LIMIT DURATION(second,5)) <> EventType
WHERE IsFirst(second,5) = 1

```

```

GROUP BY

```

```

DeviceID

```

```

,SessionWindow(second, 5, 50000) OVER (PARTITION BY DeviceID)
,TumblingWindow(second,5)
HAVING DATEDIFF(second, MIN(EventTime), MAX(EventTime)) > 5

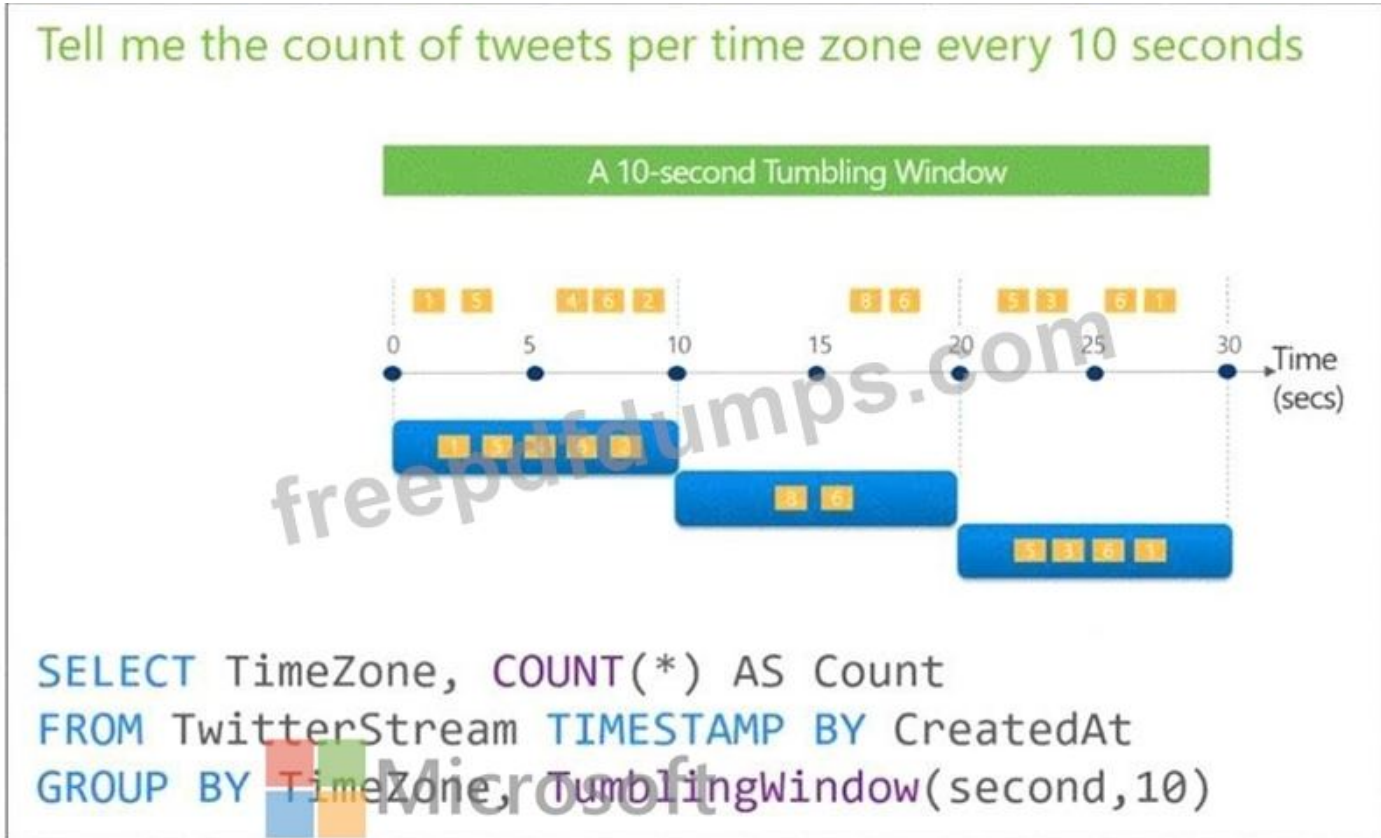
```

Box 1: WHERE EventType='HeartBeat'

Box 2: ,TumblingWindow(Second, 5)

Tumbling windows are a series of fixed-sized, non-overlapping and contiguous time intervals.

The following diagram illustrates a stream with a series of events and how they are mapped into 10-second tumbling windows.



Reference:

<https://docs.microsoft.com/en-us/stream-analytics-query/session-window-azure-stream-analytics>

<https://docs.microsoft.com/en-us/stream-analytics-query/tumbling-window-azure-stream-analytics>

NEW QUESTION: 97

You have an Azure subscription that contains an Azure data factory.

You are editing an Azure Data Factory activity JSON.

The script needs to copy a file from Azure Blob Storage to multiple destinations. The solution must ensure that the source and destination files have consistent folder paths.

How should you complete the script? 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

Values	Answer Area
FlattenHierarchy	<pre> "name": "Pipeline1", "properties": { "activities": [{ "name": "Activity1", "type": <input type="text"/>, "typeProperties": { "isSequential": "true", "items": { "value": "@pipeline () .parameters.mySinkDatasetFolderPath", "type": "Expression"}, "activities" [{ "name": "MyCopyActivity", "type": "Copy", "typeProperties": { "source": { "type": "BlobSource", "recursive": "false" }, "sink": { "type": "BlobSink", "CopyBehavior": <input type="text"/> } } }] } } }] } </pre>
ForEach	
MergeFiles	
PreserveHierarchy	
Switch	
Until	

Answer:

Values	Answer Area
FlattenHierarchy	<pre> "name": "Pipeline1", "properties": { "activities": [{ "name": "Activity1", "type": <input type="text" value="ForEach"/>, "typeProperties": { "isSequential": "true", "items": { "value": "@pipeline () .parameters.mySinkDatasetFolderPath", "type": "Expression"}, "activities" [{ "name": "MyCopyActivity", "type": "Copy", "typeProperties": { "source": { "type": "BlobSource", "recursive": "false" }, "sink": { "type": "BlobSink", "CopyBehavior": <input type="text" value="Switch"/> } } }] } } }] } </pre>
ForEach	
MergeFiles	
PreserveHierarchy	
Switch	
Until	

Explanation:

Values

-
-
-
-
-
-

Answer Area

```

{
  "name": "Pipeline1",
  "properties": {
    "activities": [
      {
        "name": "Activity1",
        "type": ForEach ,
        "typeProperties": {
          "isSequential": "true",
          "items": {
            "value": "@pipeline().parameters.mySinkDatasetFolderPath",
            "type": "Expression" },
          "activities": [
            {
              "name": "MyCopyActivity",
              "type": "Copy",
              "typeProperties": {
                "source": {
                  "type": "BlobSource",
                  "recursive": "false" },
                "sink": {
                  "type": "BlobSink",
                  "CopyBehaviour": Switch
                }
              }
            }
          ]
        }
      }
    ]
  }
}

```

NEW QUESTION: 98

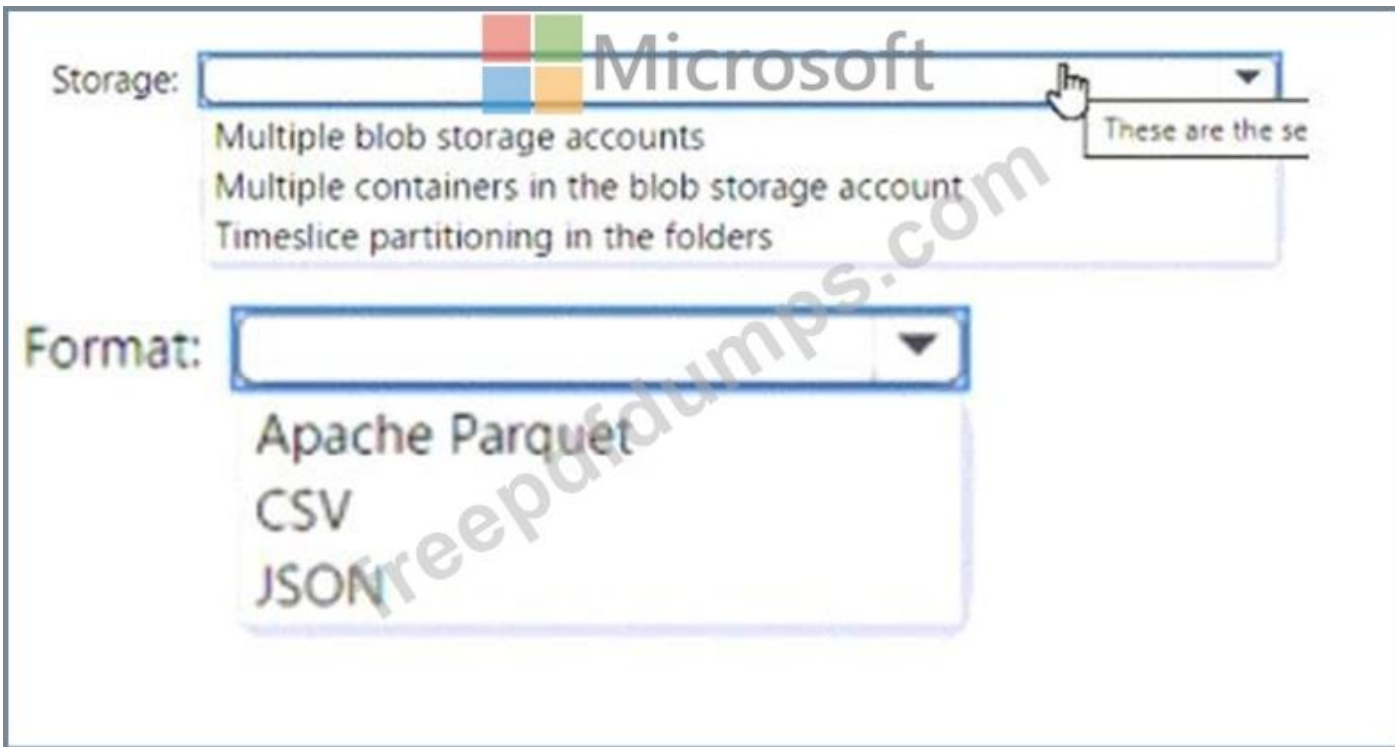
You have an Azure Blob storage account that contains a folder. The folder contains 120,000 files. Each file contains 62 columns.

Each day, 1,500 new files are added to the folder.

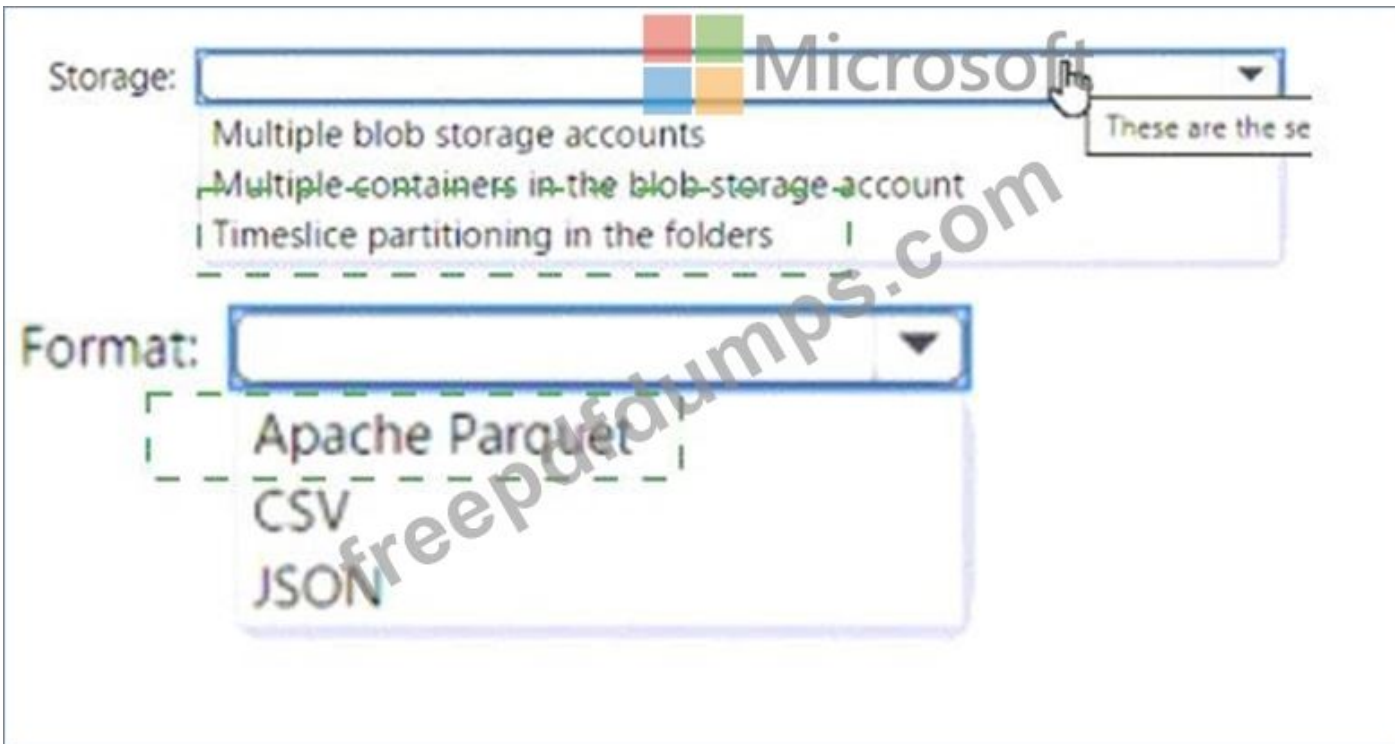
You plan to incrementally load five data columns from each new file into an Azure Synapse Analytics workspace.

You need to minimize how long it takes to perform the incremental loads.

What should you use to store the files and format?



Answer:



Explanation:

Box 1 = timeslice partitioning in the folders This means that you should organize your files into folders based on a time attribute, such as year, month, day, or hour. For example, you can have a folder structure like /yyy

/mm/dd/file.csv. This way, you can easily identify and load only the new files that are added each day by using a time filter in your Azure Synapse pipeline¹². Timeslice partitioning can also improve the performance of data loading and querying by reducing the number of files that need to be scanned Box = 2 Apache Parquet This is because Parquet is a columnar file format that can efficiently store and compress data with many columns. Parquet files can also be partitioned by a time attribute, which can improve the

performance of incremental loading and querying by reducing the number of files that need to be scanned¹²³. Parquet files are supported by both dedicated SQL pool and serverless SQL pool in Azure Synapse Analytics².

NEW QUESTION: 99

You have an Azure subscription that contains an Azure Synapse Analytics dedicated SQL pool named Poo 11 and a storage account. The storage account contains a blob container. The blob container contains multiple CSV files.

You plan to load the files into Pool! by using the following code.

```
COPY INTO [staging] FROM <PATH TO CSV FOLDER> WITH ( FILE_TYPE = 'CSV' FIELDTERMINATOR = ',' ROWTERMINATOR = '0X0A' COMPRESSION = 'GZIP' ) OPTION (LABEL = 'COPY : Staging dataset');
```

For each of the following statements, select Yes if the statement is true. Otherwise, select No.

NOTE: Each correct selection is worth one point.

ANSWER AREA

Statements	Yes	No
Identity values will be imported into a table.	<input type="radio"/>	<input type="radio"/>
The code expects each line in a CSV file to end with a new line character.	<input type="radio"/>	<input type="radio"/>
'COPY : Staging dataset' will be inserted into each row of a column named LABEL.	<input type="radio"/>	<input type="radio"/>

Answer:

Answer Area

Statements	Yes	No
Identity values will be imported into a table.	<input type="radio"/>	<input checked="" type="radio"/>
The code expects each line in a CSV file to end with a new line character.	<input type="radio"/>	<input checked="" type="radio"/>
'COPY : Staging dataset' will be inserted into each row of a column named LABEL.	<input checked="" type="radio"/>	<input type="radio"/>

Explanation:

ANSWER AREA

Statements	Yes	No
Identity values will be imported into a table.	<input type="radio"/>	<input checked="" type="radio"/>
The code expects each line in a CSV file to end with a new line character.	<input type="radio"/>	<input checked="" type="radio"/>
'COPY : Staging dataset' will be inserted into each row of a column named LABEL.	<input checked="" type="radio"/>	<input type="radio"/>

NEW QUESTION: 100

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. 1 million
- C. 60 million
- D. 600,000

Answer: A (LEAVE A REPLY)

NEW QUESTION: 101

You need to design an analytical storage solution for the transactional data. The solution must meet the sales transaction dataset requirements.

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

NOTE: Each correct selection is worth one point.

Table type to store retail store data:

	▼
Hash	
Replicated	
Round-robin	

Table type to store promotional data:

	▼
Hash	
Replicated	
Round-robin	



Answer:

Table type to store retail store data:	<table border="1"><tr><td></td><td>▼</td></tr><tr><td colspan="2">Hash</td></tr><tr><td colspan="2">Replicated</td></tr><tr><td colspan="2">Round-robin</td></tr></table>		▼	Hash		Replicated		Round-robin	
	▼								
Hash									
Replicated									
Round-robin									
Table type to store promotional data:	<table border="1"><tr><td></td><td>▼</td></tr><tr><td colspan="2">Hash</td></tr><tr><td colspan="2">Replicated</td></tr><tr><td colspan="2">Round-robin</td></tr></table>		▼	Hash		Replicated		Round-robin	
	▼								
Hash									
Replicated									
Round-robin									

Explanation:

Table type to store retail store data:

	▼
Hash	
Replicated	
Round-robin	

Table type to store promotional data:

	▼
Hash	
Replicated	
Round-robin	

Box 1: Round-robin

Round-robin tables are useful for improving loading speed.

Scenario: Partition data that contains sales transaction records. Partitions must be designed to provide efficient loads by month.

Box 2: Hash

Hash-distributed tables improve query performance on large fact tables.

Reference:

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

NEW QUESTION: 102

You have the following Azure Data Factory pipelines

- * ingest Data from System 1
- * Ingest Data from System2
- * Populate Dimensions
- * Populate facts

ingest Data from System1 and Ingest Data from System1 have no dependencies. Populate Dimensions must execute after Ingest Data from System1 and Ingest Data from System* Populate Facts must execute after the 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. Create a parent pipeline that contains the four pipelines and use an event trigger.
- C. Create a parent pipeline that contains the four pipelines and use a schedule trigger.
- D. Add a schedule trigger to all four pipelines.

Answer: C (LEAVE A REPLY)

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>

NEW QUESTION: 103

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.

```
Select TimeZone, count (*) AS MessageCount
```

FROM MessageStream

	▼	CreatedAt
LAST		
OVER		
SYSTEM.TIMESTAMP()		
TIMESTAMP BY		

GROUP BY TimeZone,

	▼	(second, 15)
HOPPINGWINDOW		
SESSIONWINDOW		
SLIDINGWINDOW		
TUMBLINGWINDOW		

Answer:

```
Select TimeZone, count (*) AS MessageCount
```

FROM MessageStream

	▼	CreatedAt
LAST		
OVER		
SYSTEM.TIMESTAMP()		
TIMESTAMP BY		

GROUP BY TimeZone,

	▼	(second, 15)
HOPPINGWINDOW		
SESSIONWINDOW		
SLIDINGWINDOW		
TUMBLINGWINDOW		

Explanation:

```
Select TimeZone, count (*) AS MessageCount
```

FROM MessageStream

	▼	CreatedAt
LAST		
OVER		
SYSTEM.TIMESTAMP()		
TIMESTAMP BY		

GROUP BY TimeZone,

	▼	(second, 15)
HOPPINGWINDOW		
SESSIONWINDOW		
SLIDINGWINDOW		
TUMBLINGWINDOW		

Box 1: timestamp by

Box 2: TUMBLINGWINDOW

Tumbling window functions are used to segment a data stream into distinct time segments and perform a function against them, such as the example below. The key differentiators of a Tumbling window are that they repeat, do not overlap, and an event cannot belong to more than one tumbling window.

Tell me the count of Tweets per time zone every 10 seconds

```
SELECT TimeZone, COUNT(*) AS Count
FROM TwitterStream TIMESTAMP BY CreatedAt
GROUP BY TimeZone, TumblingWindow(second, 10)
```

Reference:

<https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-window-functions>

NEW QUESTION: 104

that has the activity shown in the following exhibit.



Use the drop-down menus to select the answer choice that completes each statement based on the information presented in the graphic.

Stored procedure1 will execute if Web1 and Set variable1 [answer choice].

If Web1 fails and Set variable2 succeeds, the pipeline status will be [answer choice].

Answer:

Stored procedure1 will execute if Web1 and Set variable1 [answer choice].

If Web1 fails and Set variable2 succeeds, the pipeline status will be [answer choice].

Explanation:

Answer Area

Stored procedure1 will execute if Web1 and Set variable1 [answer choice] succeed

If Web1 fails and Set variable2 succeeds, the pipeline status will be [answer choice] Failed

NEW QUESTION: 105

You need to design a data retention solution for the Twitter feed data records. The solution must meet the customer sentiment analytics requirements.

Which Azure Storage functionality should you include in the solution?

- A. lifecycle management
- B. time-based retention
- C. change feed

D. soft delete

Answer: D (LEAVE A REPLY)

NEW QUESTION: 106

You have an Azure Synapse Analytics workspace.

You plan to deploy a lake database by using a database template in Azure Synapse.

Which two elements are included in the template? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point

- A. data formats
- B. table permissions
- C. relationships
- D. table definitions
- E. linked services

Answer: (SHOW ANSWER)

Valid DP-203 Dumps shared by Actual4test.com for Helping Passing DP-203 Exam! Actual4test.com now offer the **newest DP-203 exam dumps**, the Actual4test.com DP-203 exam **questions have been updated** and **answers have been corrected** get the **newest** Actual4test.com DP-203 dumps with Test Engine here: https://www.actual4test.com/DP-203_examcollection.html (**365** Q&As Dumps, **30%OFF** Special Discount: **Freepdfdumps**)

NEW QUESTION: 107

You need to design an Azure Synapse Analytics dedicated SQL pool that meets the following requirements:

- * Can return an employee record from a given point in time.
- * Maintains the latest employee information.
- * Minimizes query complexity.

How should you model the employee data?

- A. as a temporal table
- B. as a SQL graph table
- C. as a degenerate dimension table
- D. as a Type 2 slowly changing dimension (SCD) table

Answer: (SHOW ANSWER)

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.

Reference:

<https://docs.microsoft.com/en-us/learn/modules/populate-slowly-changing-dimensions-azure-synapse-analytics-pipelines/3-choose-between-dimension-types>

NEW QUESTION: 108

You have an Azure subscription that contains an Azure Synapse Analytics workspace named workspace1. Workspace1 connects to an Azure DevOps repository named repo1. Repo1 contains a collaboration branch named main and a development branch named branch1. Branch1 contains an Azure Synapse pipeline named pipeline1.

In workspace1, you complete testing of pipeline1.

You need to schedule pipeline1 to run daily at 6 AM.

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.

Actions

Answer Area

Create a new branch in Repo1.

Merge the changes from branch1 into main.

Associate the schedule trigger with pipeline1.

Switch to Synapse live mode.

Create a schedule trigger.

Publish the contents of main.



Answer:

Actions	Answer Area
Create a new branch in Repo1.	Create a schedule trigger.
Merge the changes from branch1 into main.	Associate the schedule trigger with pipeline1.
Associate the schedule trigger with pipeline1.	Merge the changes from branch1 into main.
Switch to Synapse live mode.	Publish the contents of main.
Create a schedule trigger.	
Publish the contents of main.	

Explanation:

Create a schedule trigger.
Associate the schedule trigger with pipeline1.
Merge the changes from branch1 into main.
Publish the contents of main.

NEW QUESTION: 109

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

Contacts contains a column named Phone.

You need to ensure that users in a specific role only see the last four digits of a phone number when querying the Phone column.

What should you include in the solution?

- A. a default value
- B. dynamic data masking
- C. row-level security (RLS)
- D. column encryption
- E. table partitions

Answer: B (LEAVE A REPLY)

Dynamic data masking helps prevent unauthorized access to sensitive data by enabling customers to designate how much of the sensitive data to reveal with minimal impact on the application layer. It's a policy-based security feature that hides the sensitive data in the result set of a query over designated database fields, while the data in the database is not changed.

Reference:

<https://docs.microsoft.com/en-us/azure/azure-sql/database/dynamic-data-masking-overview>

NEW QUESTION: 110

You have an Azure subscription that contains the resources shown in the following table.

Name	Type	Description
ws1	Azure Synapse Analytics workspace	None
kv1	Azure Key Vault	None
UAMI1	User-assigned managed identity	Associated with ws1
sp1	Apache Spark pool in Azure Synapse Analytics	Associated with ws1

You need to ensure that you can Spark notebooks in ws1. The solution must ensure secrets from kv1 by using UAMI1. What should you do? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

In the Azure portal: Add a role-based access control (RBAC) role to kv1.
 Add a role-based access control (RBAC) role to kv1.
 Add a role-based access control (RBAC) role to ws1.
 Create a linked service to kv1.

In Synapse Studio: Create a linked service to kv1.
 Add a role-based access control (RBAC) role to kv1.
 Add a role-based access control (RBAC) role to ws1.
 Create a linked service to kv1.

Answer:

Answer Area

In the Azure portal: Add a role-based access control (RBAC) role to kv1.
 Add a role-based access control (RBAC) role to kv1.
 Add a role-based access control (RBAC) role to ws1.
 Create a linked service to kv1.

In Synapse Studio: Create a linked service to kv1.
 Add a role-based access control (RBAC) role to kv1.
 Add a role-based access control (RBAC) role to ws1.
 Create a linked service to kv1.

Explanation:

Answer Area

In the Azure portal: Add a role-based access control (RBAC) role to kv1.
 Add a role-based access control (RBAC) role to kv1.
 Add a role-based access control (RBAC) role to ws1.
 Create a linked service to kv1.

In Synapse Studio: Create a linked service to kv1.
 Add a role-based access control (RBAC) role to kv1.
 Add a role-based access control (RBAC) role to ws1.
 Create a linked service to kv1.

NEW QUESTION: 111

You are monitoring an Azure Stream Analytics job.

The Backlogged Input Events count has been 20 for the last hour.

You need to reduce the Backlogged Input Events count.

What should you do?

- A. Drop late arriving events from the job.
- B. Add an Azure Storage account to the job.
- C. Increase the streaming units for the job.
- D. Stop the job.

Answer: (SHOW ANSWER)

General symptoms of the job hitting system resource limits include:

* If the backlog event metric keeps increasing, it's an indicator that the system resource is constrained (either because of output sink throttling, or high CPU).

Note: Backlogged Input Events: Number of input events that are backlogged. A non-zero value for this metric implies that your job isn't able to keep up with the number of incoming events. If this value is slowly increasing or consistently non-zero, you should scale out your job: adjust Streaming Units.

Reference:

<https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-scale-jobs>

<https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-monitoring>

NEW QUESTION: 112

You have an Azure Storage account and a data warehouse in Azure Synapse Analytics in the UK South region.

You need to copy blob data from the storage account to the data warehouse by using Azure Data Factory.

The solution must meet the following requirements:

- * Ensure that the data remains in the UK South region at all times.
- * Minimize administrative effort.

Which type of integration runtime should you use?

- A. Azure integration runtime
- B. Azure-SSIS integration runtime
- C. Self-hosted integration runtime

Answer: A (LEAVE A REPLY)

IR type	Public network	Private network
 Azure	Data Flow Data movement Activity dispatch	
 Self-hosted	Data movement Activity dispatch	Data movement Activity dispatch
Azure-SSIS	SSIS package execution	SSIS package execution

Reference:

<https://docs.microsoft.com/en-us/azure/data-factory/concepts-integration-runtime>

NEW QUESTION: 113

What should you recommend using to secure sensitive customer contact information?

- A. data labels
- B. column-level security
- C. row-level security
- D. Transparent Data Encryption (TDE)

Answer: B (LEAVE A REPLY)

Scenario: All cloud data must be encrypted at rest and in transit.

Always Encrypted is a feature designed to protect sensitive data stored in specific database columns from access (for example, credit card numbers, national identification numbers, or data on a need to know basis).

This includes database administrators or other privileged users who are authorized to access the database to perform management tasks, but have no business need to access the particular data in the encrypted columns.

The data is always encrypted, which means the encrypted data is decrypted only for processing by client applications with access to the encryption key.

References:

<https://docs.microsoft.com/en-us/azure/sql-database/sql-database-security-overview>

NEW QUESTION: 114

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 scenario, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You have an Azure Storage account that contains 100 GB of files. The files contain text and numerical values.

75% of the rows contain description data that has an average length of 1.1 MB.

You plan to copy the data from the storage account to an Azure SQL data warehouse.

You need to prepare the files to ensure that the data copies quickly.

Solution: You modify the files to ensure that each row is less than 1 MB.

Does this meet the goal?

- A. Yes
- B. No

Answer: A (LEAVE A REPLY)

When exporting data into an ORC File Format, you might get Java out-of-memory errors when there are large text columns. To work around this limitation, export only a subset of the columns.

References:

<https://docs.microsoft.com/en-us/azure/sql-data-warehouse/guidance-for-loading-data>

NEW QUESTION: 115

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 scenario, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You have an Azure Storage account that contains 100 GB of files. The files contain text and numerical values.

75% of the rows contain description data that has an average length of 1.1 MB.

You plan to copy the data from the storage account to an enterprise data warehouse in Azure Synapse Analytics.

You need to prepare the files to ensure that the data copies quickly.

Solution: You convert the files to compressed delimited text files.

Does this meet the goal?

A. Yes

B. No

Answer: (SHOW ANSWER)

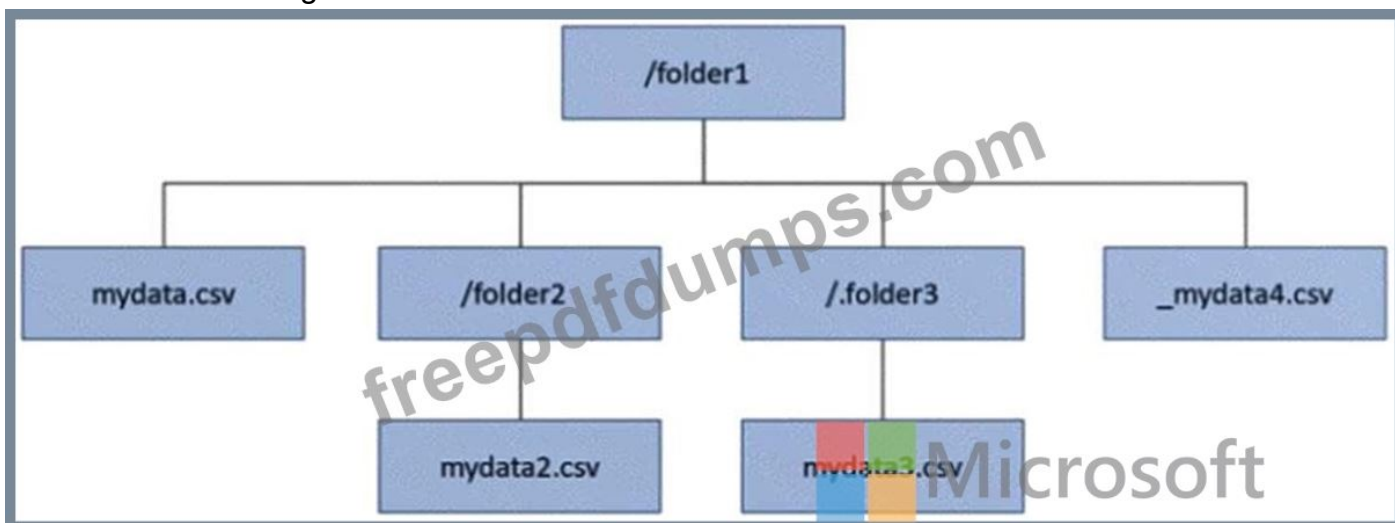
All file formats have different performance characteristics. For the fastest load, use compressed delimited text files.

Reference:

<https://docs.microsoft.com/en-us/azure/sql-data-warehouse/guidance-for-loading-data>

NEW QUESTION: 116

You have an Azure Data Lake Storage Gen2 account that contains a container named container1. You have an Azure Synapse Analytics serverless SQL pool that contains a native external table named dbo.Table1. The source data for dbo.Table1 is stored in container1. The folder structure of container1 is shown in the following exhibit.



The external data source is defined by using the following statement.

```
CREATE EXTERNAL DATA SOURCE DataLake
WITH
(
    LOCATION = 'https://mydatalake.dfs.core.windows.net/container1/folder1/**'
    , CREDENTIAL = DataLakeCred
);
```

For each of the following statements, select Yes if the statement is true. Otherwise, select No.
 NOTE: Each correct selection is worth one point.

Statements	Yes	No
When selecting all the rows in dbo.Table1, data from the mydata2.csv file will be returned.	<input type="radio"/>	<input type="radio"/>
When selecting all the rows in dbo.Table1, data from the mydata3.csv file will be returned.	<input type="radio"/>	<input type="radio"/>
When selecting all the rows in dbo.Table1, data from the _mydata4.csv file will be returned.	<input type="radio"/>	<input type="radio"/>

Answer:

Statements	Yes	No
When selecting all the rows in dbo.Table1, data from the mydata2.csv file will be returned.	<input checked="" type="radio"/>	<input type="radio"/>
When selecting all the rows in dbo.Table1, data from the mydata3.csv file will be returned.	<input checked="" type="radio"/>	<input type="radio"/>
When selecting all the rows in dbo.Table1, data from the _mydata4.csv file will be returned.	<input type="radio"/>	<input checked="" type="radio"/>

Explanation:

Box 1: Yes

In the serverless SQL pool you can also use recursive wildcards /logs/** to reference Parquet or CSV files in any sub-folder beneath the referenced folder.

Box 2: Yes

Box 3: No

Reference: <https://learn.microsoft.com/en-us/azure/synapse-analytics/sql/develop-tables-external-tables>

NEW QUESTION: 117

A company has a real-time data analysis solution that is hosted on Microsoft Azure. The solution uses Azure Event Hub to ingest data and an Azure Stream Analytics cloud job to analyze the data. The cloud job is configured to use 120 Streaming Units (SU).

You need to optimize performance for the Azure Stream Analytics job.

Which two actions should you perform? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. Implement event ordering.
- B. Implement Azure Stream Analytics user-defined functions (UDF).
- C. Implement query parallelization by partitioning the data output.
- D. Scale the SU count for the job up.
- E. Scale the SU count for the job down.
- F. Implement query parallelization by partitioning the data input.

Answer: D,F (LEAVE A REPLY)

Reference:

<https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-parallelization>

NEW QUESTION: 118

You have an Azure SQL database named DB1 and an Azure Data Factory data pipeline named pipeline. From Data Factory, you configure a linked service to DB1.

In DB1, you create a stored procedure named SP1. SP1 returns a single row of data that has four columns.

You need to add an activity to pipeline to execute SP1. The solution must ensure that the values in the columns are stored as pipeline variables.

Which two types of activities can you use to execute SP1? (Refer to Data Engineering on Microsoft Azure documents or guide for Answers/Explanation available at Microsoft.com)

- A. Stored Procedure
- B. Lookup
- C. Script
- D. Copy

Answer: A,B (LEAVE A REPLY)

the two types of activities that you can use to execute SP1 are Stored Procedure and Lookup.

A Stored Procedure activity executes a stored procedure on an Azure SQL Database or Azure Synapse Analytics or SQL Server1. You can specify the stored procedure name and parameters in the activity settings1.

A Lookup activity retrieves a dataset from any data source that returns a single row of data with four columns2

. You can use a query to execute a stored procedure as the source of the Lookup activity2. You can then store the values in the columns as pipeline variables by using expressions2.

<https://learn.microsoft.com/en-us/azure/data-factory/transform-data-using-stored-procedure>

NEW QUESTION: 119

You are developing a solution that will stream to Azure Stream Analytics. The solution will have both streaming data and reference data.

Which input type should you use for the reference data?

- A. Azure Cosmos DB
- B. Azure Blob storage
- C. Azure IoT Hub
- D. Azure Event Hubs

Answer: B (LEAVE A REPLY)

Stream Analytics supports Azure Blob storage and Azure SQL Database as the storage layer for Reference Data.

Reference:

<https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-use-reference-data>

NEW QUESTION: 120

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.

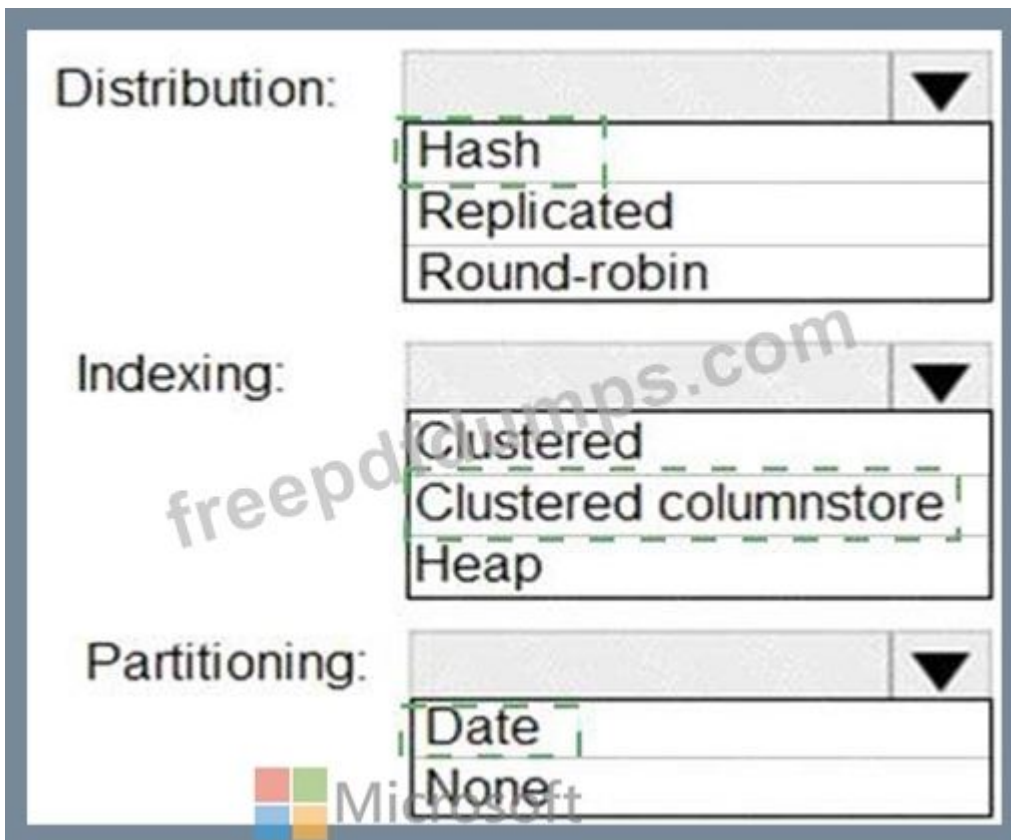
NOTE: Each correct selection is worth one point.

The image shows a configuration interface for a staging table in Azure Synapse. It consists of three dropdown menus:

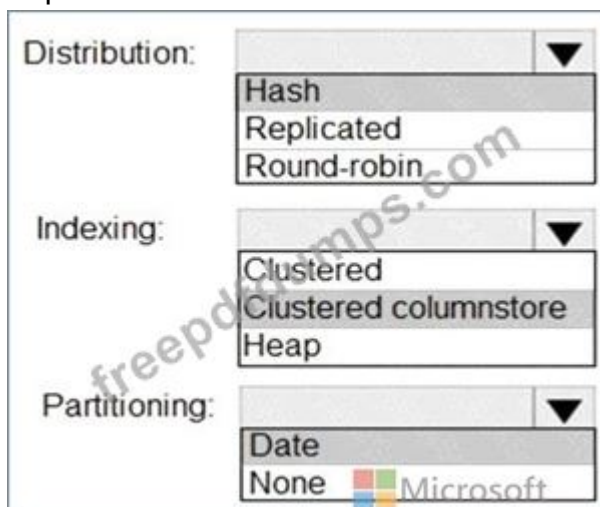
- Distribution:** Hash, Replicated, Round-robin
- Indexing:** Clustered, Clustered columnstore, Heap
- Partitioning:** Date, None

A watermark 'freepdfmumps.com' and the Microsoft logo are visible in the background.

Answer:



Explanation:



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.

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>

NEW QUESTION: 121

You are planning the deployment of Azure Data Lake Storage Gen2.

You have the following two reports that will access the data lake:

* Report1: Reads three columns from a file that contains 50 columns.

* Report2: Queries a single record based on a timestamp.

You need to recommend in which format to store the data in the data lake to support the reports. The solution must minimize read times.

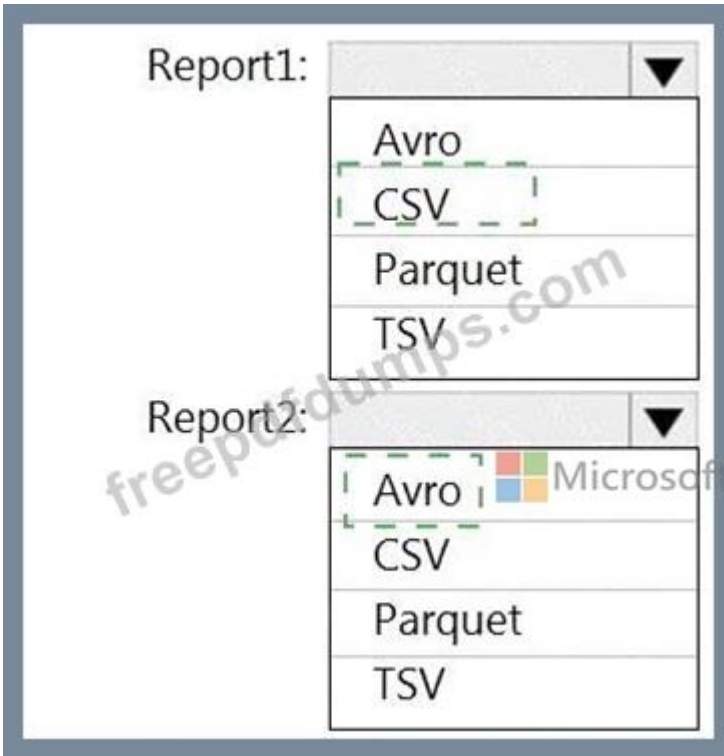
What should you recommend for each report? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

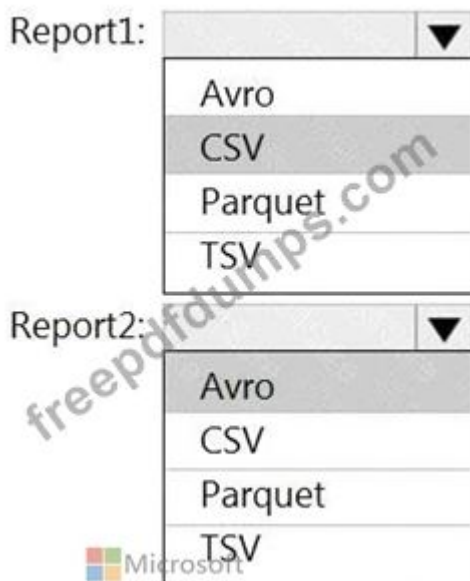
Report1: 

Report2: 

Answer:



Explanation:



Report1: CSV

CSV: The destination writes records as delimited data.

Report2: AVRO

AVRO supports timestamps.

Not Parquet, TSV: Not options for Azure Data Lake Storage Gen2.

Reference:

<https://streamsets.com/documentation/datacollector/latest/help/datacollector/UserGuide/Destinations/ADLS-G2-D.html>

Valid DP-203 Dumps shared by Actual4test.com for Helping Passing DP-203 Exam! Actual4test.com now offer the **newest DP-203 exam dumps**, the Actual4test.com DP-203 exam **questions have been updated** and **answers have been corrected** get the **newest** Actual4test.com DP-203 dumps with Test Engine here: https://www.actual4test.com/DP-203_examcollection.html (**365 Q&As Dumps, 30%OFF** **Special Discount: Freepdfdumps**)

NEW QUESTION: 122

You are designing a partition strategy for a fact table in an Azure Synapse Analytics dedicated SQL pool. The table has the following specifications:

- * Contain sales data for 20,000 products.
- * Use hash distribution on a column named ProductID,
- * Contain 2.4 billion records for the years 2019 and 2020.

Which number of partition ranges provides optimal compression and performance of the clustered columnstore index?

- A. 40
- B. 240
- C. 400
- D. 2,400

Answer: (SHOW ANSWER)

Each partition should have around 1 millions records. Dedicated SQL pools already have 60 partitions.

We have the formula: $\text{Records}/(\text{Partitions} * 60) = 1 \text{ million}$

$\text{Partitions} = \text{Records}/(1 \text{ million} * 60)$

$\text{Partitions} = 2.4 \times 1,000,000,000 / (1,000,000 * 60) = 40$

Note: Having too many partitions can reduce the effectiveness of clustered columnstore indexes if each partition has fewer than 1 million rows. Dedicated SQL pools automatically partition your data into 60 databases. So, if you create a table with 100 partitions, the result will be 6000 partitions.

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/best-practices-dedicated-sql-pool>

NEW QUESTION: 123

You have two fact tables named Flight and Weather. Queries targeting the tables will be based on the join between the following columns.

Table	Column
Flight	ArrivalAirportID ArrivalDateTime
Weather	AirportID ReportDateTime

You need to recommend a solution that maximum query performance.

What should you include in the recommendation?

- A. In the tables, use a hash distribution of ArriveAirPortID and AirportID.
- B. In each table, create an IDENTITY column.

- C. In the tables, use a hash distribution of ArriveDateTime and ReportDateTime.
- D. In each table, create a column as a composite of the other two columns in the table.

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 124

You are building a data flow in Azure Data Factory that upserts data into a table in an Azure Synapse Analytics dedicated SQL pool.

You need to add a transformation to the data flow. The transformation must specify logic indicating when a row from the input data must be upserted into the sink.

Which type of transformation should you add to the data flow?

- A. join
- B. select
- C. surrogate key
- D. alter row

Answer: D ([LEAVE A REPLY](#))

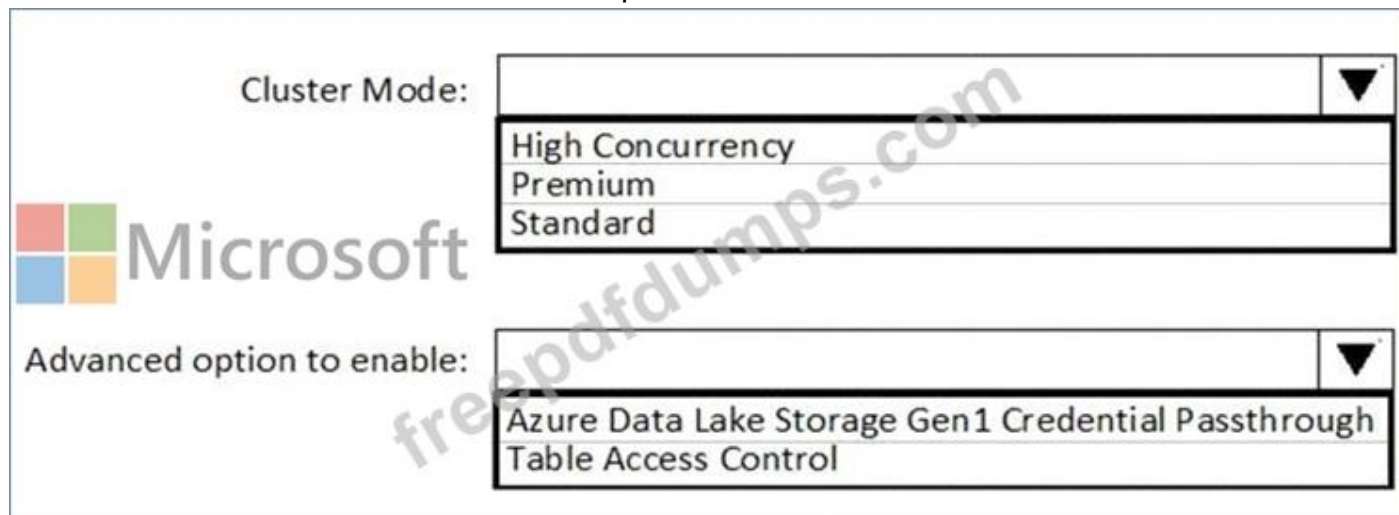
The alter row transformation allows you to specify insert, update, delete, and upsert policies on rows based on expressions. You can use the alter row transformation to perform upserts on a sink table by matching on a key column and setting the appropriate row policy

NEW QUESTION: 125

You need to implement an Azure Databricks cluster that automatically connects to Azure Data Lake Storage Gen2 by using Azure Active Directory (Azure AD) integration.

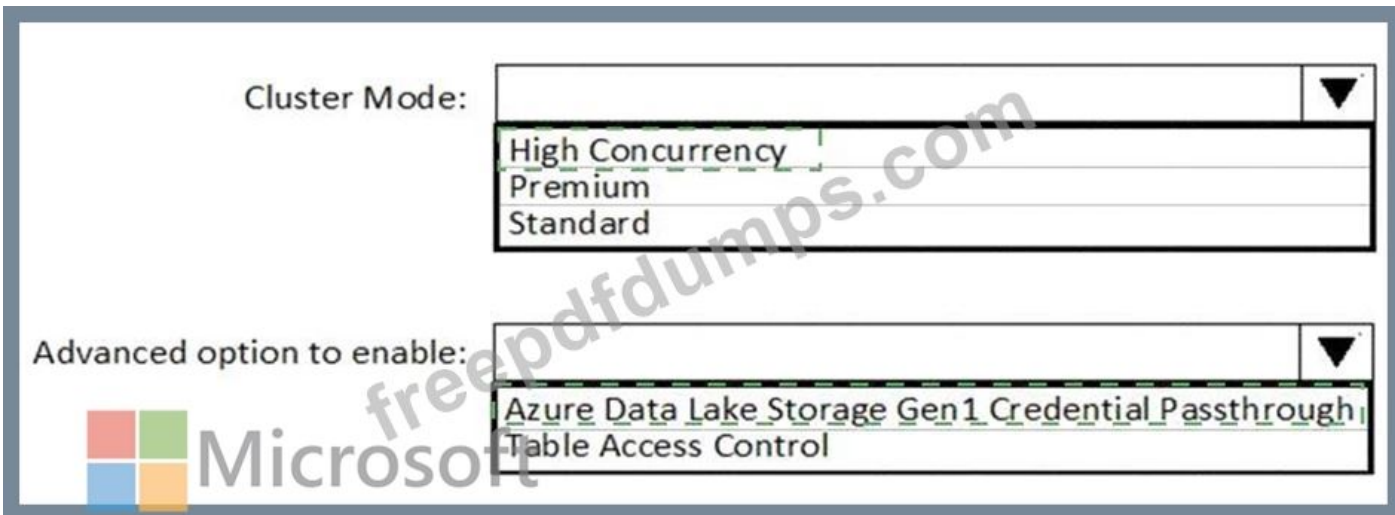
How should you configure the new cluster? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

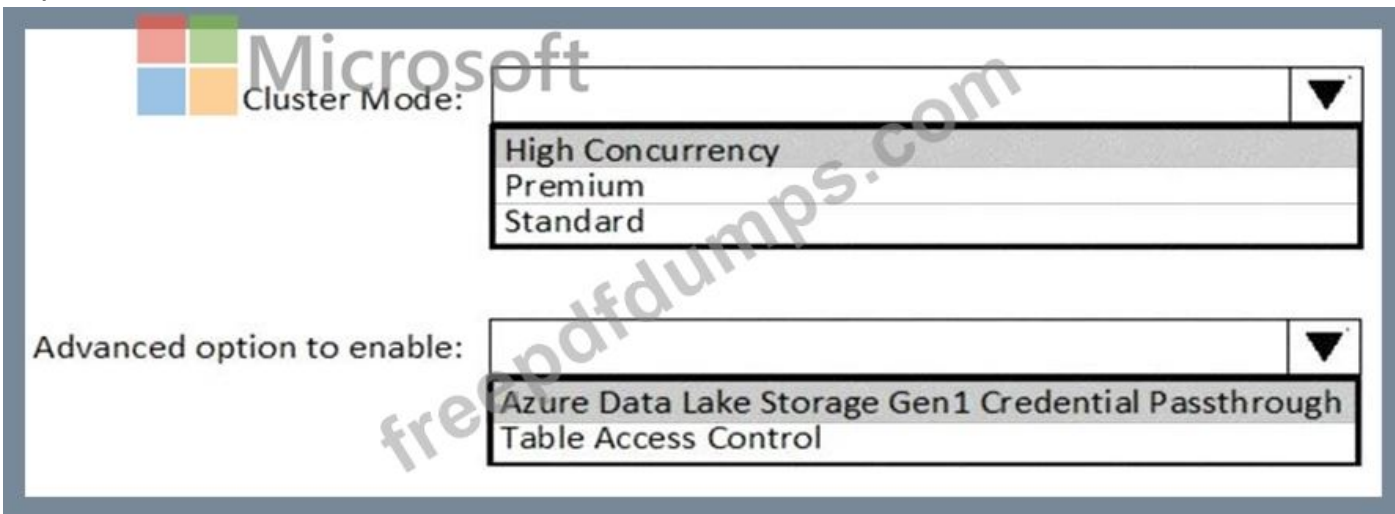


The screenshot shows two dropdown menus for configuring an Azure Databricks cluster. The first dropdown, labeled 'Cluster Mode:', has three options: 'High Concurrency', 'Premium', and 'Standard'. The second dropdown, labeled 'Advanced option to enable:', has two options: 'Azure Data Lake Storage Gen1 Credential Passthrough' and 'Table Access Control'. A Microsoft logo is visible on the left side of the form area.

Answer:



Explanation:



Box 1: High Concurrency

Enable Azure Data Lake Storage credential passthrough for a high-concurrency cluster.

Incorrect:

Support for Azure Data Lake Storage credential passthrough on standard clusters is in Public Preview. Standard clusters with credential passthrough are supported on Databricks Runtime 5.5 and above and are limited to a single user.

Box 2: Azure Data Lake Storage Gen1 Credential Passthrough

You can authenticate automatically to Azure Data Lake Storage Gen1 and Azure Data Lake Storage 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.

References:

<https://docs.azuredatabricks.net/spark/latest/data-sources/azure/adls-passthrough.html>

NEW QUESTION: 126

You use Azure Data Factory to create data pipelines.

You are evaluating whether to integrate Data Factory and GitHub for source and version control. What are two advantages of the integration? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

- A. additional triggers
- B. lower pipeline execution times
- C. the ability to save without publishing
- D. the ability to save pipelines that have validation issues

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 127

You have an Azure subscription that contains an Azure Synapse Analytics dedicated SQL pool. You plan to deploy a solution that will analyze sales data and include the following:

- * A table named Country that will contain 195 rows
- * A table named Sales that will contain 100 million rows
- * A query to identify total sales by country and customer from the past 30 days. You need to create the tables. The solution must maximize query performance.

How should you complete the script? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area



```
CREATE TABLE [dbo].[Sales]
(
    [OrderDate] date NOT NULL
,   [CustomerId] int NOT NULL
,   [CountryId] int NOT NULL
,   [Total] money NOT NULL
)
WITH
(
    DISTRIBUTION = HASH([CustomerId])
    CLUSTERED COLUMNSTORE INDEX
)
CREATE TABLE [dbo].[Country]
```

Answer:

Answer Area



```
CREATE TABLE [dbo].[Sales]
(
    [OrderDate] date NOT NULL
    , [CustomerId] int NOT NULL
    , [CountryId] int NOT NULL
    , [Total] money NOT NULL
)
WITH
(
    DISTRIBUTION = HASH([CustomerId])
    CLUSTERED COLUMNSTORE INDEX
)
CREATE TABLE [dbo].[Country]
```

NEW QUESTION: 128

You have an Azure data solution that contains an enterprise data warehouse in Azure Synapse Analytics named DW1.

Several users execute ad hoc queries to DW1 concurrently.

You regularly perform automated data loads to DW1.

You need to ensure that the automated data loads have enough memory available to complete quickly and successfully when the adhoc queries run.

What should you do?

- A. Hash distribute the large fact tables in DW1 before performing the automated data loads.
- B. Assign a smaller resource class to the automated data load queries.
- C. Assign a larger resource class to the automated data load queries.
- D. Create sampled statistics for every column in each table of DW1.

Answer: C (LEAVE A REPLY)

The performance capacity of a query is determined by the user's resource class. Resource classes are pre-determined resource limits in Synapse SQL pool that govern compute resources and concurrency for query execution.

Resource classes can help you configure resources for your queries by setting limits on the number of queries that run concurrently and on the compute-resources assigned to each query. There's a trade-off between memory and concurrency.

Smaller resource classes reduce the maximum memory per query, but increase concurrency.

Larger resource classes increase the maximum memory per query, but reduce concurrency.

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/resource-classes-for-workload-management>

NEW QUESTION: 129

You are monitoring an Azure Stream Analytics job by using metrics in Azure.

You discover that during the last 12 hours, the average watermark delay is consistently greater than the configured late arrival tolerance.

What is a possible cause of this behavior?

- A. Events whose application timestamp is earlier than their arrival time by more than five minutes arrive as inputs.
- B. There are errors in the input data.
- C. The late arrival policy causes events to be dropped.
- D. The job lacks the resources to process the volume of incoming data.

Answer: D (LEAVE A REPLY)

Watermark Delay indicates the delay of the streaming data processing job.

There are a number of resource constraints that can cause the streaming pipeline to slow down. The watermark delay metric can rise due to:

- * Not enough processing resources in Stream Analytics to handle the volume of input events. To scale up resources, see Understand and adjust Streaming Units.
- * Not enough throughput within the input event brokers, so they are throttled. For possible solutions, see Automatically scale up Azure Event Hubs throughput units.
- * Output sinks are not provisioned with enough capacity, so they are throttled. The possible solutions vary widely based on the flavor of output service being used.

Reference:

<https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-time-handling>

NEW QUESTION: 130

You have an Azure Synapse Analytics dedicated SQL pool named Pool1. Pool1 contains a fact table named Table1. Table1 contains sales data. Sixty-five million rows of data are added to Table1 monthly. At the end of each month, you need to remove data that is older than 36 months. The solution must minimize how long it takes to remove the data.

How should you partition Table1, and how should you remove the old data? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

Partition the data:

Remove the data:

Answer:

Answer Area

Partition the data: Partition by date with one partition per day.

Partition by date with one partition per day.

Partition by date with one partition per month.

Partition by product.

Remove the data: Delete the old data from Table1 by using a WHERE clause.

Delete the old data from Table1 by using a WHERE clause.

Delete the old data from Table1 by using a JOIN.

Switch the oldest partition to another table named Table2 and drop Table2.

Truncate the oldest partition.

Explanation:

Answer Area

Partition the data: Partition by date with one partition per day.

Remove the data: Delete the old data from Table1 by using a WHERE clause.

NEW QUESTION: 131

You are responsible for providing access to an Azure Data Lake Storage Gen2 account.

Your user account has contributor access to the storage account, and you have the application ID and access key.

You plan to use PolyBase to load data into an enterprise data warehouse in Azure Synapse Analytics.

You need to configure PolyBase to connect the data warehouse to storage account.

Which three components should you create in sequence? To answer, move the appropriate components from the list of components to the answer area and arrange them in the correct order.

Components

- a database scoped credential
- an asymmetric key
- an external data source
- a database encryption key
- an external file format

Answer Area

Answer:

Components

- a database scoped credential
- an asymmetric key
- an external data source
- a database encryption key
- an external file format

Answer Area

- a database scoped credential
- an external data source
- an external file format

Explanation:

Answer Area

- 1 a database scoped credential
- 2 an external data source
- 3 an external file format

NEW QUESTION: 132

You have an Azure Data lake Storage account that contains a staging zone.

You need to design a daily process to ingest incremental data from the staging zone, transform the data by executing an R script, and then insert the transformed data into a data warehouse in Azure Synapse Analytics.

Solution: You use an Azure Data Factory schedule trigger to execute a pipeline that executes an Azure Databricks notebook, and then inserts the data into the data warehouse.

Does this meet the goal?

- A. Yes
- B. No

Answer: B (LEAVE A REPLY)

If you need to transform data in a way that is not supported by Data Factory, you can create a custom activity, not an Azure Databricks notebook, with your own data processing logic and use the activity in the pipeline.

You can create a custom activity to run R scripts on your HDInsight cluster with R installed.

Reference:

<https://docs.microsoft.com/en-US/azure/data-factory/transform-data>

NEW QUESTION: 133

You have an Azure Data Lake Storage account that contains a staging zone.

You need to design a daily process to ingest incremental data from the staging zone, transform the data by executing an R script, and then insert the transformed data into a data warehouse in Azure Synapse Analytics.

Solution: You use an Azure Data Factory schedule trigger to execute a pipeline that executes mapping data Flow, and then inserts the data into the data warehouse.

Does this meet the goal?

A. Yes

B. No

Answer: B (LEAVE A REPLY)

If you need to transform data in a way that is not supported by Data Factory, you can create a custom activity, not a mapping flow, with your own data processing logic and use the activity in the pipeline. You can create a custom activity to run R scripts on your HDInsight cluster with R installed.

Reference:

<https://docs.microsoft.com/en-US/azure/data-factory/transform-data>

NEW QUESTION: 134

You have an Azure Synapse Analytics SQL pool named Pool1 on a logical Microsoft SQL server named Server1.

You need to implement Transparent Data Encryption (TDE) on Pool1 by using a custom key named key1. Which five 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.

Actions

Answer Area

- Enable TDE on Pool1.
- Assign a managed identity to Server1.
- Configure key1 as the TDE protector for Server1.
- Add key1 to the Azure key vault.
- Create an Azure key vault and grant the managed identity permissions to the key vault.

freedumps.com



Answer:

Actions

Enable TDE on Pool1.

Assign a managed identity to Server1.

Configure key1 as the TDE protector for Server1.

Add key1 to the Azure key vault.

Create an Azure key vault and grant the managed identity permissions to the key vault.

Answer Area

Assign a managed identity to Server1.

Create an Azure key vault and grant the managed identity permissions to the key vault.

Add key1 to the Azure key vault.

Configure key1 as the TDE protector for Server1.

Enable TDE on Pool1.

Explanation:

Assign a managed identity to Server1.

Create an Azure key vault and grant the managed identity permissions to the key vault.

Add key1 to the Azure key vault.

Configure key1 as the TDE protector for Server1.

Enable TDE on Pool1.

Step 1: Assign a managed identity to Server1

You will need an existing Managed Instance as a prerequisite.

Step 2: Create an Azure key vault and grant the managed identity permissions to the vault Create Resource and setup Azure Key Vault.

Step 3: Add key1 to the Azure key vault

The recommended way is to import an existing key from a .pfx file or get an existing key from the vault.

Alternatively, generate a new key directly in Azure Key Vault.

Step 4: Configure key1 as the TDE protector for Server1

Provide TDE Protector key

Step 5: Enable TDE on Pool1

Reference:

<https://docs.microsoft.com/en-us/azure/azure-sql/managed-instance/scripts/transparent-data-encryption-byok-powershell>

NEW QUESTION: 135

You have an Azure Data Lake Storage account that contains one CSV file per hour for January 1, 2020, through January 31, 2023. The files are partitioned by using the following folder structure.

csv/system1/{year}/{month}/{filename}.csv

You need to query the files by using an Azure Synapse Analytics serverless SQL pool. The solution must return the row count of each file created during the last three months of 2022.

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

NOTE: Each correct selection is worth one point.

```
SELECT
    r.filepath() AS filepath
    ,COUNT_BIG(*) AS [rows]
FROM OPENROWSET (
    BULK 
    DATA_SOURCE = 'MyDataLake',
    FORMAT = 'CSV',
    PARSER_VERSION = '2.0',
    FIRSTROW = 2 )
WITH (vendor_id INT) AS [r]
WHERE  IN ('10', '11', '12')
```

freedumps.com



GROUP BY
Answer:

```

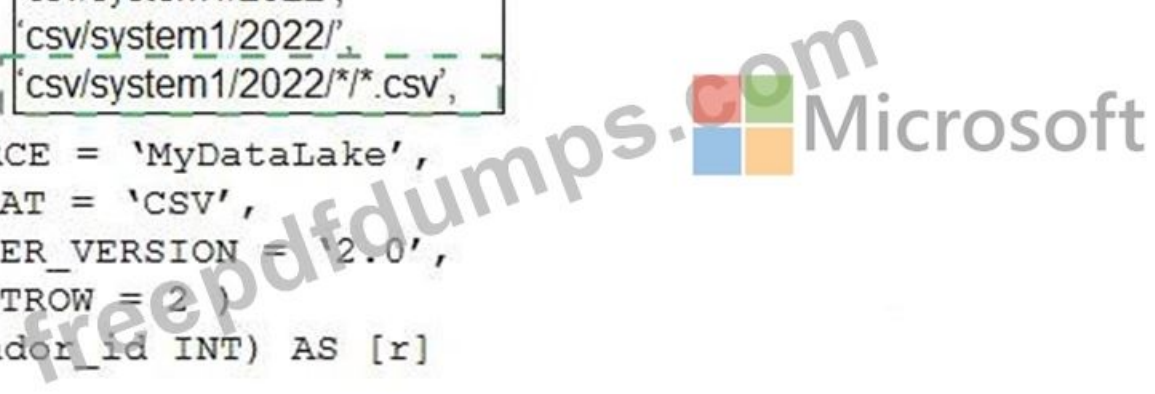
SELECT
    r.filepath() AS filepath
    ,COUNT_BIG(*) AS [rows]
FROM OPENROWSET (
    BULK 
    DATA_SOURCE = 'MyDataLake' ,
    FORMAT = 'CSV' ,
    PARSER_VERSION = '2.0' ,
    FIRSTROW = 2 )
WITH (vendor_id INT) AS [r]
WHERE 




```

GROUP BY

Explanation:



Answer Area

```
SELECT
    r.filepath() AS filepath
    ,COUNT_BIG(*) AS [rows]
FROM OPENROWSET(
    BULK 'csv/system1/2022/*/*.csv'
    DATA_SOURCE = 'MyDataLake'
    FORMAT = 'CSV'
    PARSER_VERSION = '2.0',
    FIRSTROW = 2 )
WITH (vendor_id INT) AS [r]
WHERE
    r.filepath(1) IN ('10', '11', '12')
GROUP BY
    r.filepath()
ORDER BY
    filepath;
```

NEW QUESTION: 136

You have an Azure subscription that contains an Azure Synapse Analytics dedicated SQL pool. You need to identify whether a single distribution of a parallel query takes longer than other distributions.

- A. sys.dm_pdw_dns_workers
- B. sys.dm_pdw Mec_sessions
- C. sys.dm_pdw_request_steps
- D. sys.dm_pdw_sql_requests

Answer: D (LEAVE A REPLY)

Valid DP-203 Dumps shared by Actual4test.com for Helping Passing DP-203 Exam! Actual4test.com now offer the **newest DP-203 exam dumps**, the Actual4test.com DP-203 exam **questions have been updated** and **answers have been corrected** get the **newest** Actual4test.com DP-203 dumps with Test Engine here: https://www.actual4test.com/DP-203_examcollection.html (365 Q&As Dumps, **30%OFF** Special Discount: **Freepdfdumps**)

NEW QUESTION: 137

You have an Azure Synapse Analytics Apache Spark pool named Pool1.

You plan to load JSON files from an Azure Data Lake Storage Gen2 container into the tables in Pool1. The structure and data types vary by file.

You need to load the files into the tables. The solution must maintain the source data types.

What should you do?

A. Use a Get Metadata activity in Azure Data Factory.

B. Use a Conditional Split transformation in an Azure Synapse data flow.

C. Load the data by using the OPEHROWset Transact-SQL command in an Azure Synapse Analytics serverless SQL pool.

D. Load the data by using PySpark.

Answer: A (LEAVE A REPLY)

Serverless SQL pool can automatically synchronize metadata from Apache Spark. A serverless SQL pool database will be created for each database existing in serverless Apache Spark pools.

Serverless SQL pool enables you to query data in your data lake. It offers a T-SQL query surface area that accommodates semi-structured and unstructured data queries.

To support a smooth experience for in place querying of data that's located in Azure Storage files, serverless SQL pool uses the OPENROWSET function with additional capabilities.

The easiest way to see to the content of your JSON file is to provide the file URL to the OPENROWSET function, specify csv FORMAT.

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/query-json-files>

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/query-data-storage>

NEW QUESTION: 138

You have an Azure Synapse Analytics dedicated SQL pool named Pool1. Pool1 contains a table named table1.

You load 5 TB of data into table1.

You need to ensure that columnstore compression is maximized for table1.

Which statement should you execute?

A. ALTER INDEX ALL on table1 REORGANIZE

B. ALTER INDEX ALL on table1 REBUILD

C. DBCC DBREINDEX (table1)

D. DBCC INDEXDEFRAG (pool1, table1)

Answer: C (LEAVE A REPLY)

Columnstore and columnstore archive compression

Columnstore tables and indexes are always stored with columnstore compression. You can further reduce the size of columnstore data by configuring an additional compression called archival compression. To perform archival compression, SQL Server runs the Microsoft XPRESS compression algorithm on the data. Add or remove archival compression by using the following data compression types:

Use COLUMNSTORE_ARCHIVE data compression to compress columnstore data with archival compression.

Use COLUMNSTORE data compression to decompress archival compression. The resulting data continue to be compressed with columnstore compression.

To add archival compression, use ALTER TABLE (Transact-SQL) or ALTER INDEX (Transact-SQL) with the REBUILD option and DATA COMPRESSION = COLUMNSTORE_ARCHIVE.

Reference: <https://learn.microsoft.com/en-us/sql/relational-databases/data-compression/data-compression>

NEW QUESTION: 139

You need to create a partitioned table in an Azure Synapse Analytics dedicated SQL pool.

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.

Values	Answer Area
CLUSTERED INDEX	<pre>CREATE TABLE table1 (ID INTEGER, col1 VARCHAR(10), col2 VARCHAR(10)) WITH ([] = HASH(ID), [] (ID RANGE LEFT FOR VALUES (1, 1000000, 2000000)));</pre>
COLLATE	
DISTRIBUTION	
PARTITION	
PARTITION FUNCTION	
PARTITION SCHEME	

Answer:

Values	Answer Area
CLUSTERED INDEX	<pre>CREATE TABLE table1 (ID INTEGER, col1 VARCHAR(10), col2 VARCHAR(10)) WITH (DISTRIBUTION = HASH(ID), PARTITION (ID RANGE LEFT FOR VALUES (1, 1000000, 2000000)));</pre>
COLLATE	
DISTRIBUTION	
PARTITION	
PARTITION FUNCTION	
PARTITION SCHEME	

Explanation:

```
CREATE TABLE table1
(
  ID INTEGER,
  col1 VARCHAR(10),
  col2 VARCHAR(10)
) WITH
(
  DISTRIBUTION = HASH(ID),
  PARTITION (ID RANGE LEFT FOR VALUES (1, 1000000, 2000000))
);
```

Box 1: DISTRIBUTION

Table distribution options include DISTRIBUTION = HASH (distribution_column_name), assigns each row to one distribution by hashing the value stored in distribution_column_name.

Box 2: PARTITION

Table partition options. Syntax:

PARTITION (partition_column_name RANGE [LEFT | RIGHT] FOR VALUES ([boundary_value [...n]])) Reference:

<https://docs.microsoft.com/en-us/sql/t-sql/statements/create-table-azure-sql-data-warehouse?>

NEW QUESTION: 140

You have a table in an Azure Synapse Analytics dedicated SQL pool. The table was created by using the following Transact-SQL statement.

```
CREATE TABLE [dbo].[DimEmployee]
  ([EmployeeKey] [int] IDENTITY(1,1) NOT NULL,
  [EmployeeID] [int] NOT NULL,
  [FirstName] [varchar](100) NOT NULL,
  [LastName] [varchar](100) NOT NULL,
  [JobTitle] [varchar](100) NULL,
  [LastHireDate] [date] NULL,
  [StreetAddress] [varchar](500) NOT NULL,
  [City] [varchar](200) NOT NULL,
  [StateProvince] [varchar](50) NOT NULL,
  [Portalcode] [varchar](10) NOT NULL
)
```

You need to alter the table to meet the following requirements:

- * Ensure that users can identify the current manager of employees.
- * Support creating an employee reporting hierarchy for your entire company.
- * Provide fast lookup of the managers' attributes such as name and job title.

Which column should you add to the table?

- A. [ManagerEmployeeID] [int] NULL
- B. [ManagerEmployeeID] [smallint] NULL
- C. [ManagerEmployeeKey] [int] NULL
- D. [ManagerName] [varchar](200) NULL

Answer: A (LEAVE A REPLY)

Use the same definition as the EmployeeID column.

Reference:

<https://docs.microsoft.com/en-us/analysis-services/tabular-models/hierarchies-ssas-tabular>

NEW QUESTION: 141

You have a C# application that process data from an Azure IoT hub and performs complex transformations.

You need to replace the application with a real-time solution. The solution must reuse as much code as possible from the existing application.

- A. Azure Databricks
- B. Azure Event Grid
- C. Azure Stream Analytics
- D. Azure Data Factory

Answer: C (LEAVE A REPLY)

Azure Stream Analytics on IoT Edge empowers developers to deploy near-real-time analytical intelligence closer to IoT devices so that they can unlock the full value of device-generated data. UDF are available in C# for IoT Edge jobs Azure Stream Analytics on IoT Edge runs within the Azure IoT Edge framework. Once the job is created in Stream Analytics, you can deploy and manage it using IoT Hub.

References:

<https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-edge>

NEW QUESTION: 142


You have an Azure Data Lake Storage Gen2 account that contains a JSON file for customers. The file contains two attributes named FirstName and LastName.

You need to copy the data from the JSON file to an Azure Synapse Analytics table by using Azure Databricks. A new column must be created that concatenates the FirstName and LastName values.

You create the following components:

- * A destination table in Azure Synapse
- * An Azure Blob storage container
- * A service principal

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

Actions	Answer Area
Mount the Data Lake Storage onto DBFS.	
Write the results to a table in Azure Synapse.	
Perform transformations on the file.	
Specify a temporary folder to stage the data.	
Write the results to Data Lake Storage.	
Read the file into a data frame.	
Drop the data frame.	
Perform transformations on the data frame.	

Answer:

Actions	Answer Area
Mount the Data Lake Storage onto DBFS.	Mount the Data Lake Storage onto DBFS.
Write the results to a table in Azure Synapse.	
Perform transformations on the file.	Read the file into a data frame.
Specify a temporary folder to stage the data.	
Write the results to Data Lake Storage.	Perform transformations on the data frame.
Read the file into a data frame.	Specify a temporary folder to stage the data.
Drop the data frame.	
Perform transformations on the data frame.	Write the results to a table in Azure Synapse.

Explanation:

- 1) mount onto DBFS
- 2) read into data frame
- 3) transform data frame
- 4) specify temporary folder
- 5) write the results to table in in Azure Synapse

<https://docs.databricks.com/data/data-sources/azure/azure-datalake-gen2.html>

<https://docs.microsoft.com/en-us>

</azure/databricks/scenarios/databricks-extract-load-sql-data-warehouse>

NEW QUESTION: 143

You are designing an application that will use an Azure Data Lake Storage Gen 2 account to store petabytes of license plate photos from toll booths. The account will use zone-redundant storage (ZRS).

You identify the following usage patterns:


- * The data will be accessed several times a day during the first 30 days after the data is created. The data must meet an availability SU of 99.9%.
- * After 90 days, the data will be accessed infrequently but must be available within 30 seconds.
- * After 365 days, the data will be accessed infrequently but must be available within five minutes.

First 30 days:

- Archive
- Cool
- Hot

After 90 days:

- Archive
- Cool
- Hot

 Microsoft


After 365 days:

- Archive
- Cool
- Hot

Answer:

First 30 days:

Archive
Cool
Hot



Microsoft

After 90 days:

Archive
Cool
Hot

After 365 days:

Archive
Cool
Hot

Explanation:

Box 1: Hot

The data will be accessed several times a day during the first 30 days after the data is created. The data must meet an availability SLA of 99.9%.

Box 2: Cool

After 90 days, the data will be accessed infrequently but must be available within 30 seconds.

Data in the Cool tier should be stored for a minimum of 30 days.

When your data is stored in an online access tier (either Hot or Cool), users can access it immediately.

The Hot tier is the best choice for data that is in active use, while the Cool tier is ideal for data that is accessed less frequently, but that still must be available for reading and writing.

Box 3: Cool

After 365 days, the data will be accessed infrequently but must be available within five minutes.

Reference: <https://docs.microsoft.com/en-us/azure/storage/blobs/access-tiers-overview>

<https://docs.microsoft.com/en-us/azure/storage/blobs/archive-rehydrate-overview>

NEW QUESTION: 144

You have an Azure Factory instance named DF1 that contains a pipeline named PL1. PL1 includes a tumbling window trigger.

You create five clones of PL1. You configure each clone pipeline to use a different data source.

You need to ensure that the execution schedules of the clone pipeline match the execution schedule of PL1.

What should you do?

- A. Modify the Concurrency setting of each pipeline.
- B. Add a new trigger to each cloned pipeline
- C. Create a tumbling window trigger dependency for the trigger of PL1.
- D. Associate each cloned pipeline to an existing trigger.

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 145

You have a Microsoft SQL Server database that uses a third normal form schema.

You plan to migrate the data in the database to a star schema in an Azure Synapse Analytics dedicated SQL pool.

You need to design the dimension tables. The solution must optimize read operations.

What should you include in the solution? to answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Transform data for the dimension tables by:	<input type="checkbox"/> Maintaining to a third normal form <input type="checkbox"/> Normalizing to a fourth normal form <input type="checkbox"/> Denormalizing to a second normal form
For the primary key columns in the dimension tables, use:	<input type="checkbox"/> New IDENTITY columns <input type="checkbox"/> A new computed column <input type="checkbox"/> The business key column from the source sys

Answer:

Transform data for the dimension tables by:

	▼
Maintaining to a third normal form	
Normalizing to a fourth normal form	
Denormalizing to a second normal form	

For the primary key columns in the dimension tables, use:

	▼
New IDENTITY columns	
A new computed column	
The business key column from the source sys	

Explanation:

Transform data for the dimension tables by:

	▼
Maintaining to a third normal form	
Normalizing to a fourth normal form	
Denormalizing to a second normal form	

For the primary key columns in the dimension tables, use:

	▼
New IDENTITY columns	
A new computed column	
The business key column from the source sys	

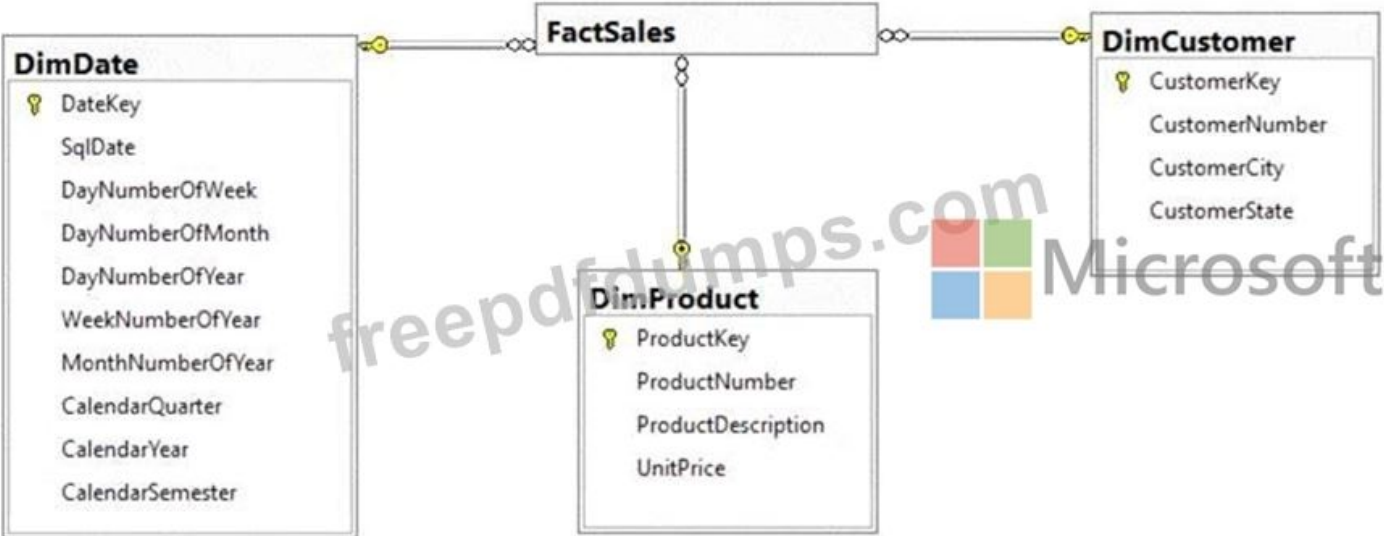
Box 1: Denormalize to a second normal form

Denormalization is the process of transforming higher normal forms to lower normal forms via storing the join of higher normal form relations as a base relation. Denormalization increases the performance in data retrieval at cost of bringing update anomalies to a database.

Box 2: New identity columns

The collapsing relations strategy can be used in this step to collapse classification entities into component entities to obtain #at dimension tables with single-part keys that connect directly to the fact table. The single- part key is a surrogate key generated to ensure it remains unique over time.

Example:



Note: A surrogate key on a table is a column with a unique identifier for each row. The key is not generated from the table data. Data modelers like to create surrogate keys on their tables when they

design data warehouse models. You can use the IDENTITY property to achieve this goal simply and effectively without affecting load performance.

Reference:

<https://www.mssqltips.com/sqlservertip/5614/explore-the-role-of-normal-forms-in-dimensional-modeling/>
<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-identity>

NEW QUESTION: 146

You have an Azure data factory.

You execute a pipeline that contains an activity named Activity1. Activity1 produces the following output.

```
{
  ...
  "dataRead": 1208,
  "dataWritten": 1208,
  "filesRead": 1,
  "filesWritten": 1,
  "sourcePeakConnections": 3,
  "sinkPeakConnections": 2,
  "copyDuration": 13,
  "throughput": 0.147,
  "effectiveIntegrationRuntime": "AutoResolveIntegrationRuntime (West Central US)",
  "usedDataIntegrationUnit": 1,
  "reportLineageToPurview": true,
  "status": "Succeeded",
  "durationInSeconds": "4"
}
```

For each of the following statements select Yes if the statement is true. Otherwise, select No.

NOTE: Each correct selection is worth one point.

Answer Area

Statements	Yes	No
Activity1 is a Copy activity.	<input type="radio"/>	<input type="radio"/>
Activity1 is executed by using a self-hosted integration runtime.	<input type="radio"/>	<input type="radio"/>
The data factory that executed the pipeline is connected to Microsoft Purview.	<input type="radio"/>	<input type="radio"/>

Answer:

Answer Area

Statements	Yes	No
Activity1 is a Copy activity.	<input checked="" type="radio"/>	<input type="radio"/>
Activity1 is executed by using a self-hosted integration runtime.	<input checked="" type="radio"/>	<input type="radio"/>
The data factory that executed the pipeline is connected to Microsoft Purview.	<input type="radio"/>	<input checked="" type="radio"/>

Explanation:

Statements	Yes	No
Activity1 is a Copy activity.	<input checked="" type="radio"/>	<input type="radio"/>
Activity1 is executed by using a self-hosted integration runtime.	<input checked="" type="radio"/>	<input type="radio"/>
The data factory that executed the pipeline is connected to Microsoft Purview.	<input type="radio"/>	<input checked="" type="radio"/>

NEW QUESTION: 147

You have several Azure Data Factory pipelines that contain a mix of the following types of activities.

- * Wrangling data flow
- * Notebook
- * Copy
- * jar

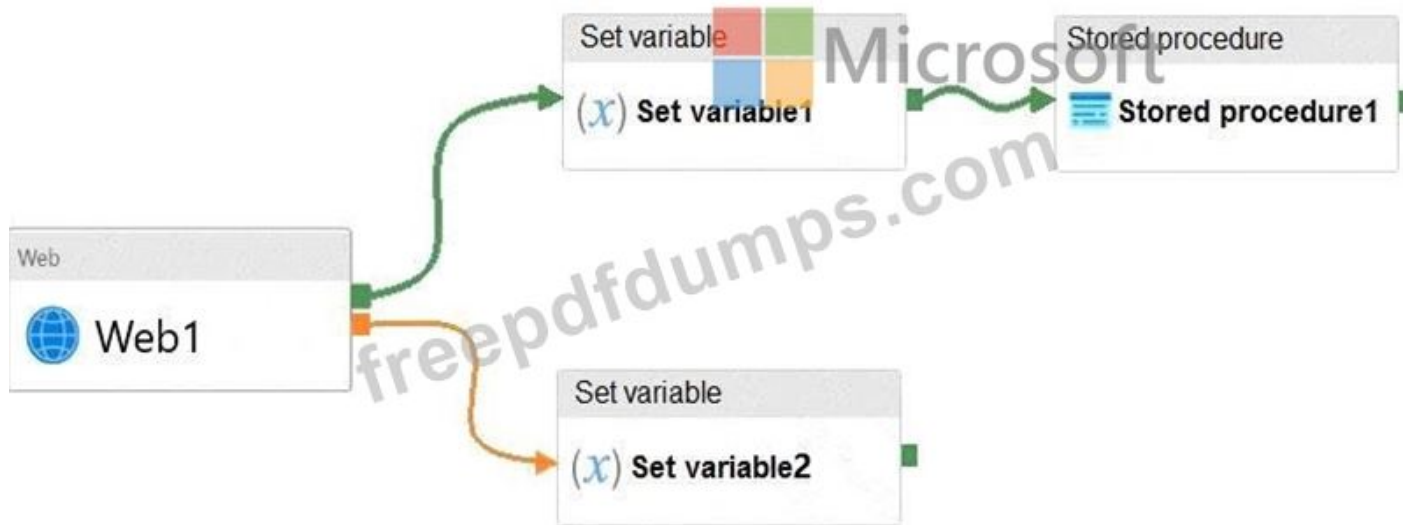
Which two Azure services should you use to debug the activities? Each correct answer presents part of the solution NOTE: Each correct selection is worth one point.

- A. Azure Synapse Analytics
- B. Azure Machine Learning
- C. Azure HDInsight
- D. Azure Data Factory
- E. Azure Databricks

Answer: ([SHOW ANSWER](#))

NEW QUESTION: 148

You have an Azure Data Factory pipeline that has the activities shown in the following exhibit.



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.

Stored procedure1 will execute Web1 and Set variable1 [answer choice]

complete
fail
succeed

If Web1 fails and Set variable2 succeeds, the pipeline status will be [answer choice]

Canceled
Failed
Succeeded

Answer:

Stored procedure1 will execute Web1 and Set variable1 [answer choice]

complete	▼
fail	
succeed	

If Web1 fails and Set variable2 succeeds, the pipeline status will be [answer choice]

Canceled	▼
Failed	
Succeeded	

Explanation:

Box 1: succeed

Box 2: failed

Example:

Now let's say we have a pipeline with 3 activities, where Activity1 has a success path to Activity2 and a failure path to Activity3. If Activity1 fails and Activity3 succeeds, the pipeline will fail. The presence of the success path alongside the failure path changes the outcome reported by the pipeline, even though the activity executions from the pipeline are the same as the previous scenario.



Activity1 fails, Activity2 is skipped, and Activity3 succeeds. The pipeline reports failure.

Reference:

<https://datasavvy.me/2021/02/18/azure-data-factory-activity-failures-and-pipeline-outcomes/>

NEW QUESTION: 149

You use Azure Stream Analytics to receive Twitter data from Azure Event Hubs and to output the data to an Azure Blob storage account.

You need to output the count of tweets during the last five minutes every five minutes. Each tweet must only be counted once.

Which windowing function should you use?

- A. a five-minute Session window
- B. a five-minute Sliding window
- C. a five-minute Tumbling window
- D. a five-minute Hopping window that has one-minute hop

Answer: (SHOW ANSWER)

Tumbling window functions are used to segment a data stream into distinct time segments and perform a function against them, such as the example below. The key differentiators of a Tumbling window are that they repeat, do not overlap, and an event cannot belong to more than one tumbling window.

References:

<https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-window-functions>

NEW QUESTION: 150

You have an Azure Synapse Analytics workspace named WS1 that contains an Apache Spark pool named Pool1.

You plan to create a database named D61 in Pool1.

You need to ensure that when tables are created in DB1, the tables are available automatically as external tables to the built-in serverless SQL pod.

Which format should you use for the tables in DB1?

- A. Parquet
- B. CSV
- C. ORC
- D. JSON

Answer: (SHOW ANSWER)

Serverless SQL pool can automatically synchronize metadata from Apache Spark. A serverless SQL pool database will be created for each database existing in serverless Apache Spark pools.

For each Spark external table based on Parquet or CSV and located in Azure Storage, an external table is created in a serverless SQL pool database.

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/develop-storage-files-spark-tables>

NEW QUESTION: 151

You plan to create an Azure Data Lake Storage Gen2 account

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

- * Provides the highest degree of data resiliency
 - * Ensures that content remains available for writes if a primary data center fails
- What should you include in the recommendation? To answer, select the appropriate options in the answer area.

Replication mechanism:

Change feed
Zone-redundant storage (ZRS)
Read-access geo-redundant storage (RA-GRS)
Read-access geo-zone-redundant storage (RA-GRS)

Failover process:

Failover initiated by Microsoft
Failover manually initiated by the customer
Failover automatically initiated by an Azure Automation job

Answer:

Answer Area

Replication mechanism:

Change feed
Zone-redundant storage (ZRS)
Read-access geo-redundant storage (RA-GRS)
Read-access geo-zone-redundant storage (RA-GRS)

Failover process:

Failover initiated by Microsoft
Failover manually initiated by the customer
Failover automatically initiated by an Azure Automation job

Explanation:

Answer Area

Replication mechanism:

Change feed
Zone-redundant storage (ZRS)
Read-access geo-redundant storage (RA-GRS)
Read-access geo-zone-redundant storage (RA-GRS)

Failover process:

Failover initiated by Microsoft
Failover manually initiated by the customer
Failover automatically initiated by an Azure Automation job

Availability : "Microsoft recommends RA-GZRS for maximum availability and durability for your applications." Failover: "The customer initiates the account failover to the secondary endpoint. "
<https://docs.microsoft.com/en-us/azure/storage/common/storage-disaster-recovery-guidance?toc=/azure/storage/blobs/toc.json>
<https://docs.microsoft.com/en-us/answers/questions/32583/azure-data-lake-gen2-disaster-recovery-storage-acco.html>

Valid DP-203 Dumps shared by Actual4test.com for Helping Passing DP-203 Exam! Actual4test.com now offer the **newest DP-203 exam dumps**, the Actual4test.com DP-203 exam **questions have been updated** and **answers have been corrected** get the **newest** Actual4test.com DP-203 dumps with Test Engine here: https://www.actual4test.com/DP-203_examcollection.html (365 Q&As Dumps, **30%OFF** Special Discount: **Freepdfdumps**)

NEW QUESTION: 152

You are designing a slowly changing dimension (SCD) for supplier data in an Azure Synapse Analytics dedicated SQL pool.

You plan to keep a record of changes to the available fields.

The supplier data contains the following columns.

Name	Description
SupplierSystemID	Unique supplier ID in an enterprise resource planning (ERP) system
SupplierName	Name of the supplier company
SupplierAddress1	Address of the supplier company
SupplierAddress2	Second address line of the supplier company
SupplierCity	City of the supplier company
SupplierStateProvince	State or province of the supplier company
SupplierCountry	Country of the supplier company
SupplierPostalCode	Postal code of the supplier company
SupplierDescription	Free-text description of the supplier company
SupplierCategory	Category of goods provided by the supplier company

Which three additional columns should you add to the data to create a Type 2 SCD? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. surrogate primary key
- B. foreign key
- C. effective start date
- D. effective end date
- E. last modified date
- F. business key

Answer: (SHOW ANSWER)

Reference:

<https://docs.microsoft.com/en-us/sql/integration-services/data-flow/transformations/slowly-changing-dimension-transformation>

NEW QUESTION: 153

You are designing a dimension table in an Azure Synapse Analytics dedicated SQL pool.

You need to create a surrogate key for the table. The solution must provide the fastest query performance.

What should you use for the surrogate key?

- A. a GUID column
- B. a sequence object
- C. an IDENTITY column

Answer: (SHOW ANSWER)

Use IDENTITY to create surrogate keys using dedicated SQL pool in AzureSynapse Analytics.

Note: A surrogate key on a table is a column with a unique identifier for each row. The key is not generated from the table data. Data modelers like to create surrogate keys on their tables when they design data warehouse models. You can use the IDENTITY property to achieve this goal simply and effectively without affecting load performance.

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

NEW QUESTION: 154

You have an Azure subscription that contains an Azure Data Lake Storage account named myaccount1. The myaccount1 account contains two containers named container1 and contained. The subscription is linked to an Azure Active Directory (Azure AD) tenant that contains a security group named Group1. You need to grant Group1 read access to container1. The solution must use the principle of least privilege. Which role should you assign to Group1?

- A. Storage Blob Data Reader for myaccount1
- B. Storage Table Data Reader for container1
- C. Storage Table Data Reader for myaccount1
- D. Storage Blob Data Reader for container1

Answer: D (LEAVE A REPLY)

NEW QUESTION: 155

You have an Azure Data Lake Storage account that has a virtual network service endpoint configured. You plan to use Azure Data Factory to extract data from the Data Lake Storage account. The data will then be loaded to a data warehouse in Azure Synapse Analytics by using PolyBase. Which authentication method should you use to access Data Lake Storage?

- A. shared access key authentication
- B. managed identity authentication
- C. account key authentication
- D. service principal authentication

Answer: B (LEAVE A REPLY)

Reference:

<https://docs.microsoft.com/en-us/azure/data-factory/connector-azure-sql-data-warehouse#use-polybase-to-load-data-into-azure-sql-data-warehouse>

NEW QUESTION: 156

You have an enterprise data warehouse in Azure Synapse Analytics named DW1 on a server named Server1.

You need to determine the size of the transaction log file for each distribution of DW1.

What should you do?

- A. On DW1, execute a query against the sys.database_files dynamic management view.
- B. From Azure Monitor in the Azure portal, execute a query against the logs of DW1.
- C. Execute a query against the logs of DW1 by using the

Get-AzOperationalInsightsSearchResult PowerShell cmdlet.

D. On the master database, execute a query against the sys.dm_pdw_nodes_os_performance_counters dynamic management view.

Answer: A (LEAVE A REPLY)

For information about the current log file size, its maximum size, and the autogrow option for the file, you can also use the size, max_size, and growth columns for that log file in sys.database_files.

Reference:

<https://docs.microsoft.com/en-us/sql/relational-databases/logs/manage-the-size-of-the-transaction-log-file>

NEW QUESTION: 157

You plan to build a structured streaming solution in Azure Databricks. The solution will count new events in five-minute intervals and report only events that arrive during the interval. The output will be sent to a Delta Lake table.

Which output mode should you use?

A. complete

B. update

C. append

Answer: C (LEAVE A REPLY)

Append Mode: Only new rows appended in the result table since the last trigger are written to external storage. This is applicable only for the queries where existing rows in the Result Table are not expected to change.

<https://docs.databricks.com/getting-started/spark/streaming.html>

NEW QUESTION: 158

You have an Azure Databricks workspace and an Azure Data Lake Storage Gen2 account named storage1. New files are uploaded daily to storage1.

* Incrementally process new files as they are upkorage1 as a structured streaming source. The solution must meet the following requirements:

* Minimize implementation and maintenance effort.

* Minimize the cost of processing millions of files.

* Support schema inference and schema drift.

Which should you include in the recommendation?

A. COPY INTO

B. Auto Loader

C. Azure Data Factory

D. Apache Spark FileStreamSource

Answer: C (LEAVE A REPLY)

NEW QUESTION: 159

You have an Azure Data Lake Storage Gen2 account that contains a JSON file for customers. The file contains two attributes named FirstName and LastName.

You need to copy the data from the JSON file to an Azure Synapse Analytics table by using Azure Databricks. A new column must be created that concatenates the FirstName and LastName values.

You create the following components:

- * A destination table in Azure Synapse
- * An Azure Blob storage container
- * A service principal

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

Actions	Answer Area
Mount the Data Lake Storage onto DBFS.	
Write the results to a table in Azure Synapse.	
Specify a temporary folder to stage the data.	
Read the file into a data frame.	
Perform transformations on the data frame.	

Answer:

Actions	Answer Area
Mount the Data Lake Storage onto DBFS.	Mount the Data Lake Storage onto DBFS.
Write the results to a table in Azure Synapse.	Read the file into a data frame.
Specify a temporary folder to stage the data.	Perform transformations on the data frame.
Read the file into a data frame.	Specify a temporary folder to stage the data.
Perform transformations on the data frame.	Write the results to a table in Azure Synapse.

Explanation:

Mount the Data Lake Storage onto DBFS.
Read the file into a data frame.
Perform transformations on the data frame.
Specify a temporary folder to stage the data.
Write the results to a table in Azure Synapse.

Step 1: Mount the Data Lake Storage onto DBFS

Begin with creating a file system in the Azure Data Lake Storage Gen2 account.

Step 2: Read the file into a data frame.

You can load the json files as a data frame in Azure Databricks.

Step 3: Perform transformations on the data frame.

Step 4: Specify a temporary folder to stage the data

Specify a temporary folder to use while moving data between Azure Databricks and Azure Synapse.

Step 5: Write the results to a table in Azure Synapse.

You upload the transformed data frame into Azure Synapse. You use the Azure Synapse connector for Azure Databricks to directly upload a dataframe as a table in a Azure Synapse.

Reference:

<https://docs.microsoft.com/en-us/azure/azure-databricks/databricks-extract-load-sql-data-warehouse>

NEW QUESTION: 160

You need to implement an Azure Databricks cluster that automatically connects to Azure Data lake Storage Gen2 by using Azure Active Directory (Azure AD) integration. How should you configure the new cluster?

To answer, select the appropriate options in the answers area. NOTE: Each correct selection is worth one point.

Answer Area



Tier: Premium Standard

Advanced option to enable: Azure Data Lake Storage Credential Passthrough Table Access Control

Answer:

Answer Area

Tier: Premium Standard

Advanced option to enable: Azure Data Lake Storage Credential Passthrough Table Access Control

Explanation:

Answer Area



Tier: Premium

Advanced option to enable: Azure Data Lake Storage Credential Passthrough

<https://docs.azuredatabricks.net/spark/latest/data-sources/azure/adls-passthrough.html>

NEW QUESTION: 161

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.

DimProduct is a **[answer choice]** slowly changing dimension (SCD).

The ProductKey column is **[answer choice]**.

Type 0
Type 1
Type 2

a surrogate key
a business key
an audit column

Answer:

DimProduct is a **[answer choice]** slowly changing dimension (SCD).

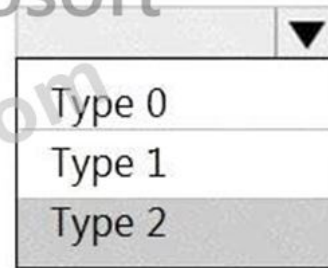
The ProductKey column is **[answer choice]**.

Type 0
Type 1
Type 2

a surrogate key
a business key
an audit column

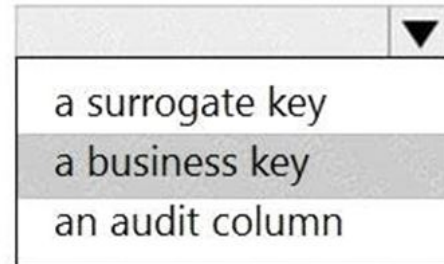
Explanation:

DimProduct is a [answer choice] slowly changing dimension (SCD).



A dropdown menu with a downward arrow on the right. The menu is open, showing three options: 'Type 0', 'Type 1', and 'Type 2'. 'Type 2' is highlighted with a grey background.

The ProductKey column is [answer choice].



A dropdown menu with a downward arrow on the right. The menu is open, showing three options: 'a surrogate key', 'a business key', and 'an audit column'. 'a surrogate key' is highlighted with a grey background.

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.

Reference:

<https://docs.microsoft.com/en-us/learn/modules/populate-slowly-changing-dimensions-azure-synapse-analytics-pipelines/3-choose-between-dimension-types>

NEW QUESTION: 162

You need to design a solution that will process streaming data from an Azure Event Hub and output the data to Azure Data Lake Storage. The solution must ensure that analysts can interactively query the streaming data.

What should you use?

- A. event triggers in Azure Data Factory
- B. Azure Stream Analytics and Azure Synapse notebooks
- C. Structured Streaming in Azure Databricks
- D. Azure Queue storage and read-access geo-redundant storage (RA-GRS)

Answer: (SHOW ANSWER)

Apache Spark Structured Streaming is a fast, scalable, and fault-tolerant stream processing API. You can use it to perform analytics on your streaming data in near real-time.

With Structured Streaming, you can use SQL queries to process streaming data in the same way that you would process static data.

Azure Event Hubs is a scalable real-time data ingestion service that processes millions of data in a matter of seconds. It can receive large amounts of data from multiple sources and stream the prepared data to Azure Data Lake or Azure Blob storage.

Azure Event Hubs can be integrated with Spark Structured Streaming to perform the processing of messages in near real-time. You can query and analyze the processed data as it comes by using a Structured Streaming query and Spark SQL.

Reference: <https://k21academy.com/microsoft-azure/data-engineer/structured-streaming-with-azure-event-hubs/>

NEW QUESTION: 163

You have an Azure Stream Analytics job that receives clickstream data from an Azure event hub.

You need to define a query in the Stream Analytics job. The query must meet the following requirements:

* Count the number of clicks within each 10-second window based on the country of a visitor.

* Ensure that each click is NOT counted more than once.

How should you define the Query?

A. SELECT Country, Avg(*) AS Average

FROM ClickStream TIMESTAMP BY CreatedAt

GROUP BY Country, SlidingWindow(second, 10)

B. SELECT Country, Count(*) AS Count

FROM ClickStream TIMESTAMP BY CreatedAt

GROUP BY Country, TumblingWindow(second, 10)

C. SELECT Country, Avg(*) AS Average

FROM ClickStream TIMESTAMP BY CreatedAt

GROUP BY Country, HoppingWindow(second, 10, 2)

D. SELECT Country, Count(*) AS Count

FROM ClickStream TIMESTAMP BY CreatedAt

GROUP BY Country, SessionWindow(second, 5, 10)

Answer: B (LEAVE A REPLY)

Tumbling window functions are used to segment a data stream into distinct time segments and perform a function against them, such as the example below. The key differentiators of a Tumbling window are that they repeat, do not overlap, and an event cannot belong to more than one tumbling window.

Example:

Reference:

<https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-window-functions>

NEW QUESTION: 164

You have an Azure subscription linked to an Azure Active Directory (Azure AD) tenant that contains a service principal named ServicePrincipal1. The subscription contains an Azure Data Lake Storage account named adls1. Adls1 contains a folder named Folder2 that has a URI of <https://adls1.dfs.core.windows.net/container1/Folder1/Folder2/>.

ServicePrincipal1 has the access control list (ACL) permissions shown in the following table.

Resource	Permission
container1	Access – Execute
Folder1	Access – Execute
Folder2	Access – Read

You need to ensure that ServicePrincipal1 can perform the following actions:

- * Traverse child items that are created in Folder2.
- * Read files that are created in Folder2.

The solution must use the principle of least privilege.

Which two permissions should you grant to ServicePrincipal1 for Folder2? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. Access - Read
- B. Access - Write
- C. Access - Execute
- D. Default-Read
- E. Default - Write
- F. Default - Execute

Answer: (SHOW ANSWER)

Execute (X) permission is required to traverse the child items of a folder.

There are two kinds of access control lists (ACLs), Access ACLs and Default ACLs.

Access ACLs: These control access to an object. Files and folders both have Access ACLs.

Default ACLs: A "template" of ACLs associated with a folder that determine the Access ACLs for any child items that are created under that folder. Files do not have Default ACLs.

Reference:

<https://docs.microsoft.com/en-us/azure/data-lake-store/data-lake-store-access-control>

NEW QUESTION: 165

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 SOL.
- * 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 High Concurrency cluster for the data engineers, and a Standard cluster for the jobs.

Does this meet the goal?

A. Yes

B. No

Answer: B (LEAVE A REPLY)

We would need a High Concurrency cluster for the jobs.

Note:

Standard clusters are recommended for a single user. Standard can run workloads developed in any language:

Python, R, Scala, and SQL.

A high concurrency cluster is a managed cloud resource. The key benefits of high concurrency clusters are that they provide Apache Spark-native fine-grained sharing for maximum resource utilization and minimum query latencies.

Reference:

<https://docs.azuredatabricks.net/clusters/configure.html>

NEW QUESTION: 166

You are designing the folder structure for an Azure Data Lake Storage Gen2 account.

You identify the following usage patterns:

- * Users will query data by using Azure Synapse Analytics serverless SQL pools and Azure Synapse Analytics serverless Apache Spark pods.
- * Most queries will include a filter on the current year or week.
- * Data will be secured by data source.

You need to recommend a folder structure that meets the following requirements:

- * Supports the usage patterns
- * Simplifies folder security
- * Minimizes query times

Which folder structure should you recommend?

- A. `\\YYYY\MM\DataSource\SubjectArea\FileData_YYYY_MM_DD.parquet`
- B. `DataSource\SubjectArea\MM\YYYY\FileData_YYYY_MM_DD.parquet`
- C. `\DataSource\SubjectArea\YYYY\MM\FileData_YYYY_MM_DD.parquet`
- D. `\DataSource\SubjectArea\YYYY-MM\FileData_YYYY_MM_DD.parquet`

E. `WW\YYYY\SubjectArea\DataSource\FileData_YYYY_MM_DD.parquet`

Answer: C (LEAVE A REPLY)

Data will be secured by data source. -> Use DataSource as top folder.

Most queries will include a filter on the current year or week -> Use \YYYY\WW\ as subfolders.

Common Use Cases

A common use case is to filter data stored in a date (and possibly time) folder structure such as /YYYY/MM

/DD/ or /YYYY/MM/YYYY-MM-DD/. As new data is generated/sent/copied/moved to the storage account, a new folder is created for each specific time period. This strategy organises data into a maintainable folder structure.

Reference: <https://www.serverlesssql.com/optimisation/azurestoragefilteringusingfilepath/>

Valid DP-203 Dumps shared by Actual4test.com for Helping Passing DP-203 Exam! Actual4test.com now offer the **newest DP-203 exam dumps**, the Actual4test.com DP-203 exam **questions have been updated** and **answers have been corrected** get the **newest** Actual4test.com DP-203 dumps with Test Engine here: https://www.actual4test.com/DP-203_examcollection.html (**365** Q&As Dumps, **30%OFF**

Special Discount: Freepdfdumps)

NEW QUESTION: 167

You have an Azure Synapse Analytics dedicated SQL pool named pool1.

You plan to implement a star schema in pool1 and create a new table named DimCustomer by using the following code.

```

CREATE TABLE dbo.[DimCustomer](
    [CustomerKey] int NOT NULL,
    [CustomerSourceID] [int] NOT NULL,
    [Title] [nvarchar](8) NULL,
    [FirstName] [nvarchar](50) NOT NULL,
    [MiddleName] [nvarchar](50) NULL,
    [LastName] [nvarchar](50) NOT NULL,
    [Suffix] [nvarchar](10) NULL,
    [CompanyName] [nvarchar](128) NULL,
    [SalesPerson] [nvarchar](256) NULL,
    [EmailAddress] [nvarchar](50) NULL,
    [Phone] [nvarchar](25) NULL,
    [InsertedDate] [datetime] NOT NULL,
    [ModifiedDate] [datetime] NOT NULL,
    [HashKey] [varchar](100) NOT NULL,
    [IsCurrentRow] [bit] NOT NULL
)
WITH
(
    DISTRIBUTION = REPLICATE,
    CLUSTERED COLUMNSTORE INDEX
);
GO

```

You need to ensure that DimCustomer has the necessary columns to support a Type 2 slowly changing dimension (SCD). Which two columns should you add? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

- A. [EffectiveEndDate] [datetime] NOT NULL
- B. [EffectiveStartDate] [datetime] NOT NULL
- C. [RowID] [bigint] NOT NULL
- D. [HistoricalSalesPerson] [nvarchar] (256) NOT NULL
- E. [PreviousModifiedDate] [datetime] NOT NULL

Answer: A,D (LEAVE A REPLY)

NEW QUESTION: 168

You are creating a new notebook in Azure Databricks that will support R as the primary language but will also support Scale and SQL Which switch should you use to switch between languages?

- A. @<Language>
- B. %<Language>
- C. \(<Language>)
- D. \(<Language>)

Answer: (SHOW ANSWER)

To change the language in Databricks' cells to either Scala, SQL, Python or R, prefix the cell with '%', followed by the language.

%python //or r, scala, sql

Reference:

<https://www.theta.co.nz/news-blogs/tech-blog/enhancing-digital-twins-part-3-predictive-maintenance-with-azure-databricks>

NEW QUESTION: 169

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

- * One billion rows
- * A clustered columnstore index
- * A hash-distributed column named Product Key
- * A column named Sales Date that is of the date data type and cannot be null Thirty million rows will be added to Table1 each month.

You need to partition Table1 based on the Sales Date column. The solution must optimize query performance and data loading.

How often should you create a partition?

- A. once per month
- B. once per year
- C. once per day
- D. once per week

Answer: B (LEAVE A REPLY)

Need a minimum 1 million rows per distribution. Each table is 60 distributions. 30 millions rows is added each month. Need 2 months to get a minimum of 1 million rows per distribution in a new partition.

Note: 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. Before partitions are created, dedicated SQL pool already divides each table into 60 distributions.

Any partitioning added to a table is in addition to the distributions created behind the scenes. Using this example, if the sales fact table contained 36 monthly partitions, and given that a dedicated SQL pool has 60 distributions, then the sales fact table should contain 60 million rows per month, or 2.1 billion rows when all months are populated. If a table contains fewer than the recommended minimum number of rows per partition, consider using fewer partitions in order to increase the number of rows per partition.

Reference:

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

NEW QUESTION: 170

You need to build a solution to ensure that users can query specific files in an Azure Data Lake Storage Gen2 account from an Azure Synapse Analytics serverless SQL pool. Which three 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.

Actions	Answer Area
Create an external file format object	
Create an external data source	
Create a query that uses Create Table as Select	
Create a table	
Create an external table	

Answer:

Actions	Answer Area
Create an external file format object	Create an external data source
Create an external data source	Create an external file format object
Create a query that uses Create Table as Select	Create an external table
Create a table	
Create an external table	

Explanation:

Create an external data source
Create an external file format object
Create an external table

Step 1: Create an external data source

You can create external tables in Synapse SQL pools via the following steps:

* CREATE EXTERNAL DATA SOURCE to reference an external Azure storage and specify the credential that should be used to access the storage.

* CREATE EXTERNAL FILE FORMAT to describe format of CSV or Parquet files.

* CREATE EXTERNAL TABLE on top of the files placed on the data source with the same file format.

Step 2: Create an external file format object

Creating an external file format is a prerequisite for creating an external table.

Step 3: Create an external table

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/develop-tables-external-tables>

NEW QUESTION: 171

You are developing a solution using a Lambda architecture on Microsoft Azure.

The data at test layer must meet the following requirements:

Data storage:

*Serve as a repository (or high volumes of large files in various formats).

*Implement optimized storage for big data analytics workloads.

*Ensure that data can be organized using a hierarchical structure.

Batch processing:

*Use a managed solution for in-memory computation processing.

*Natively support Scala, Python, and R programming languages.

*Provide the ability to resize and terminate the cluster automatically.

Analytical data store:

*Support parallel processing.

*Use columnar storage.

*Support SQL-based languages.

You need to identify the correct technologies to build the Lambda architecture.

Which technologies should you use? To answer, select the appropriate options in the answer area NOTE:

Each correct selection is worth one point.

Architecture requirement



Data storage

	▼
Azure SQL Database	
Azure Blob Storage	
Azure Cosmos DB	
Azure Data Lake Store	

Batch processing

	▼
HDInsight Spark	
HDInsight Hadoop	
Azure Databricks	
HDInsight Interactive Query	

Analytical data store

	▼
HDInsight HBase	
Azure SQL Data Warehouse	
Azure Analysis Services	
Azure Cosmos DB	

Answer:

Architecture requirement

Technology

Data storage

	▼
Azure SQL Database	
Azure Blob Storage	
Azure Cosmos DB	
Azure Data Lake Store	

Batch processing

	▼
HDInsight Spark	
HDInsight Hadoop	
Azure Databricks	
HDInsight Interactive Query	

Analytical data store

	▼
HDInsight HBase	
Azure SQL Data Warehouse	
Azure Analysis Services	
Azure Cosmos DB	

Explanation:

Architecture requirement	Technology
Data storage	<div style="border: 1px solid black; padding: 5px;"> <div style="border-bottom: 1px solid black; padding: 2px 5px; display: flex; justify-content: space-between;"> ▼ </div> <div style="padding: 2px 5px;"> <ul style="list-style-type: none"> Azure SQL Database Azure Blob Storage Azure Cosmos DB <li style="background-color: #e0e0e0;">Azure Data Lake Store </div> </div>
Batch processing	<div style="border: 1px solid black; padding: 5px;"> <div style="border-bottom: 1px solid black; padding: 2px 5px; display: flex; justify-content: space-between;"> ▼ </div> <div style="padding: 2px 5px;"> <ul style="list-style-type: none"> <li style="background-color: #e0e0e0;">HDInsight Spark HDInsight Hadoop Azure Databricks HDInsight Interactive Query </div> </div>
Analytical data store	<div style="border: 1px solid black; padding: 5px;"> <div style="border-bottom: 1px solid black; padding: 2px 5px; display: flex; justify-content: space-between;"> ▼ </div> <div style="padding: 2px 5px;"> <ul style="list-style-type: none"> HDInsight HBase <li style="background-color: #e0e0e0;">Azure SQL Data Warehouse Azure Analysis Services Azure Cosmos DB </div> </div>

Data storage: Azure Data Lake Store

A key mechanism that allows Azure Data Lake Storage Gen2 to provide file system performance at object storage scale and prices is the addition of a hierarchical namespace. This allows the collection of objects/files within an account to be organized into a hierarchy of directories and nested subdirectories in the same way that the file system on your computer is organized. With the hierarchical namespace enabled, a storage account becomes capable of providing the scalability and cost-effectiveness of object storage, with file system semantics that are familiar to analytics engines and frameworks.

Batch processing: HD Insight Spark

Aparch Spark is an open-source, parallel-processing framework that supports in-memory processing to boost the performance of big-data analysis applications.

HDInsight is a managed Hadoop service. Use it deploy and manage Hadoop clusters in Azure. For batch processing, you can use Spark, Hive, Hive LLAP, MapReduce.

Languages: R, Python, Java, Scala, SQL

Analytic data store: SQL Data Warehouse

SQL Data Warehouse is a cloud-based Enterprise Data Warehouse (EDW) that uses Massively Parallel Processing (MPP).

SQL Data Warehouse stores data into relational tables with columnar storage.

References:

<https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-namespaces>

<https://docs.microsoft.com/en-us/azure/architecture/data-guide/technology-choices/batch-processing>

<https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-overview-what-is>

NEW QUESTION: 172

You are creating an Apache Spark job in Azure Databricks that will ingest JSON-formatted data.

You need to convert a nested JSON string into a DataFrame that will contain multiple rows.

Which Spark SQL function should you use?

- A. explode
- B. filter
- C. coalesce
- D. extract

Answer: (SHOW ANSWER)

Convert nested JSON to a flattened DataFrame

You can to flatten nested JSON, using only "\$column.*" and explode methods.

Note: Extract and flatten

Use "\$column.*" and explode methods to flatten the struct and array types before displaying the flattened DataFrame.

Scala

```
display(DF.select($"id" as "main_id", $"name", $"batters", $"ppu", explode($"topping")) // Exploding the  
topping column using explode as it is an array type withColumn("topping_id", $"col.id") // Extracting  
topping_id from col using DOT form withColumn("topping_type", $"col.type") // Extracting topping_type  
from col using DOT form drop($"col") select($"*", $"batters.*") // Flattened the struct type batters tto array  
type which is batter drop($"batters") select($"*", explode($"batter")) drop($"batter") withColumn("batter_id",  
$"col.id") // Extracting batter_id from col using DOT form withColumn("battter_type", $"col.type") //  
Extracting battter_type from col using DOT form drop($"col") ) Reference: https://learn.microsoft.com/en-us/azure/databricks/kb/scala/flatten-nested-columns-dynamically
```

NEW QUESTION: 173

You have an Azure Synapse Analytics serverless SQ1 pool.

You have an Azure Data Lake Storage account named aols1 that contains a public container named container1 The container 1 container contains a folder named folder 1.

You need to query the top 100 rows of all the CSV files in folder 1.

How should you complete the query? 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.

The screenshot shows a drag-and-drop interface. On the left, under 'Values', there are four items: BULK, DATA_SOURCE, LOCATION, and OPENROWSET. On the right, under 'Answer Area', there is a SQL query: `SELECT TOP 100 * FROM ('https://ad1s1.dfs.core.windows.net/container1/folder1/*.csv', FORMAT = 'CSV') AS rows`. The FROM clause is currently empty, and the BULK clause is also empty.

Answer:

The screenshot shows the same interface as above, but now the query is completed. The FROM clause contains 'OPENROWSET' and the BULK clause contains 'BULK'. The correct answer is shown with dashed boxes around the selected items in the 'Values' pane.

Explanation:

The screenshot shows the same interface as above, but now the query is completed. The FROM clause contains 'OPENROWSET' and the BULK clause contains 'BULK'. The correct answer is shown with dashed boxes around the selected items in the 'Values' pane.

Valid DP-203 Dumps shared by Actual4test.com for Helping Passing DP-203 Exam! Actual4test.com now offer the **newest DP-203 exam dumps**, the Actual4test.com DP-203 exam **questions have been updated** and **answers have been corrected** get the **newest** Actual4test.com DP-203 dumps with Test Engine here: https://www.actual4test.com/DP-203_examcollection.html (365 Q&As Dumps, **30%OFF** Special Discount: **Freepdfdumps**)