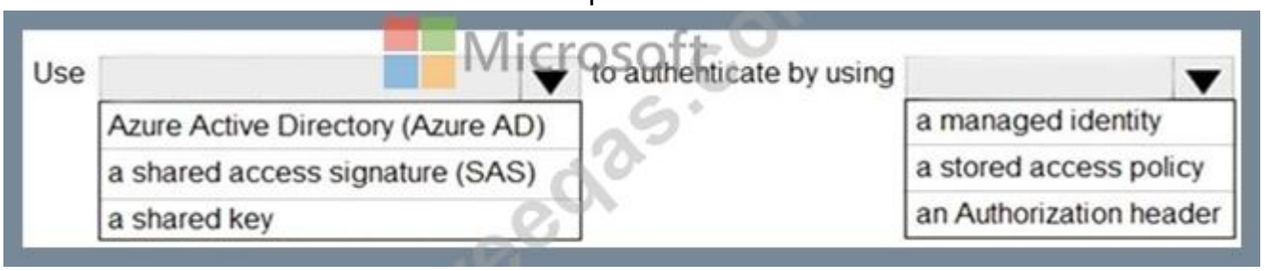
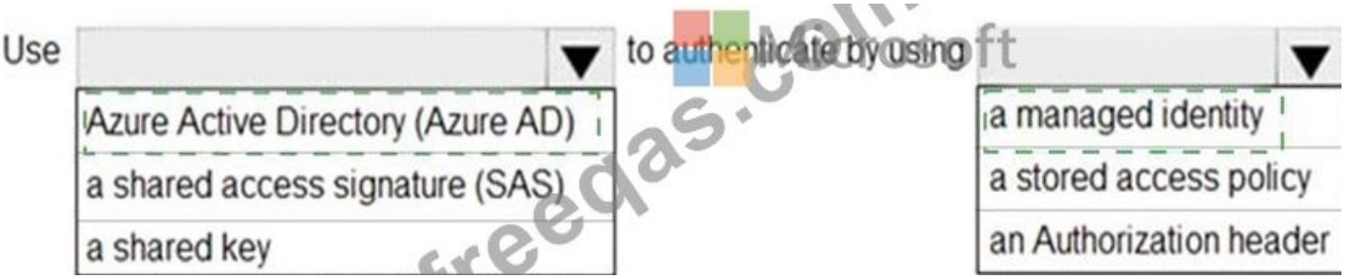
Microsoft.DP-203.v2022-04-01.q227

Exam Code:	DP-203
Exam Name:	Data Engineering on Microsoft Azure
Certification Provider:	Microsoft
Free Question Number:	227
Version:	v2022-04-01
# of views:	4551
# of Questions views:	2270
https://www.freeqas.com/qa/Microsoft/DP-203/Microsoft.DP-203.v2022-04-01.q227.html	

NEW QUESTION: 1
You have an Azure subscription that contains an Azure Data Lake Storage account. The storage account contains a data lake named DataLake1.
You plan to use an Azure data factory to ingest data from a folder in DataLake1, transform the data, and land the data in another folder.
You need to ensure that the data factory can read and write data from any folder in the DataLake1 file system.
The solution must meet the following requirements:
* Minimize the risk of unauthorized user access.
* Use the principle of least privilege.
* Minimize maintenance effort.
How should you configure access to the storage account for the data factory? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

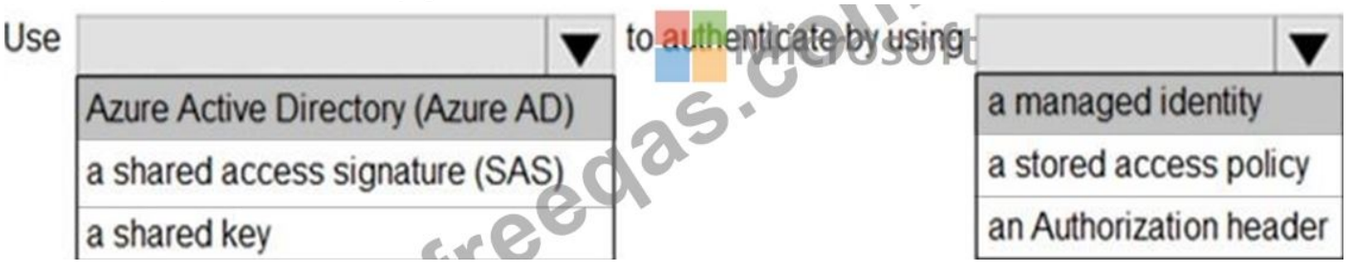


Answer:



Explanation

Text Description automatically generated with low confidence



Box 1: Azure Active Directory (Azure AD)

On Azure, managed identities eliminate the need for developers having to manage credentials by providing an identity for the Azure resource in Azure AD and using it to obtain Azure Active Directory (Azure AD) tokens.

Box 2: a managed identity

A data factory can be associated with a managed identity for Azure resources, which represents this specific data factory. You can directly use this managed identity for Data Lake Storage Gen2 authentication, similar to using your own service principal. It allows this designated factory to access and copy data to or from your Data Lake Storage Gen2.

Note: The Azure Data Lake Storage Gen2 connector supports the following authentication types.

* Account key authentication
* Service principal authentication
* Managed identities for Azure resources authentication

Reference:

https://docs.microsoft.com/en-us/azure/active-directory/managed-identities-azure-resources/overview

https://docs.microsoft.com/en-us/azure/data-factory/connector-azure-data-lake-storage

NEW QUESTION: 2

You have an Azure Synapse Analytics dedicated SQL pool that contains the users shown in the following table.

Name	Role
User1	Server admin
User2	db_datareader

User1 executes a query on the database, and the query returns the results shown in the following exhibit.

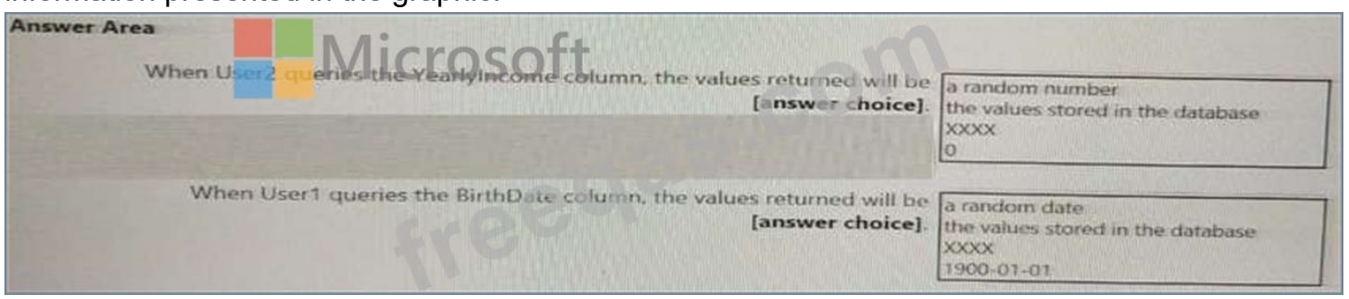
```
SELECT c.name,
    tbl.name as table_name,
    typ.name as datatype,
    c.is_masked,
    c.masking_function
FROM sys.masked_columns AS c
INNER JOIN sys.tables AS tbl ON c.[object_id] = tbl.[object_id]
INNER JOIN sys.types typ ON c.user_type_id = typ.user_type_id
WHERE is_masked = 1;
```

Results | Messages

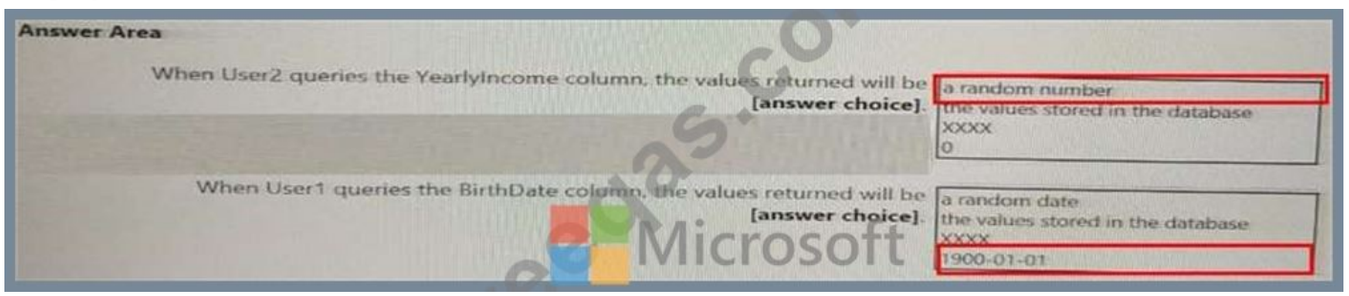
	name	table_name	datatype	is_masked	masking_function
1	BirthDate	DimCustomer	date	1	default()
2	Gender	DimCustomer	nvarchar	1	default()
3	EmailAddress	DimCustomer	nvarchar	1	email()
4	YearlyIncome	DimCustomer	money	1	default()

User1 is the only user who has access to the unmasked data.

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



Answer:



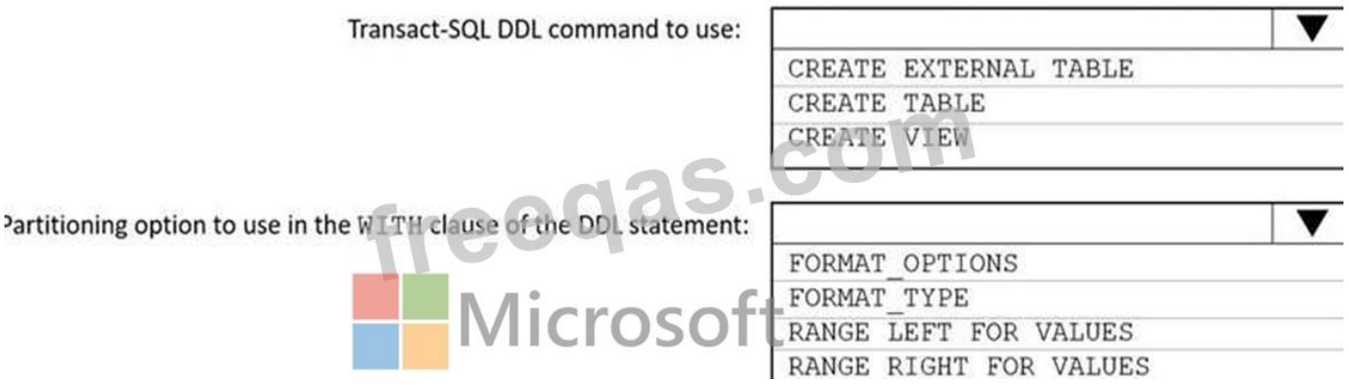
NEW QUESTION: 3

You need to implement an Azure Synapse Analytics database object for storing the sales transactions dat a. The solution must meet the sales transaction dataset requirements.

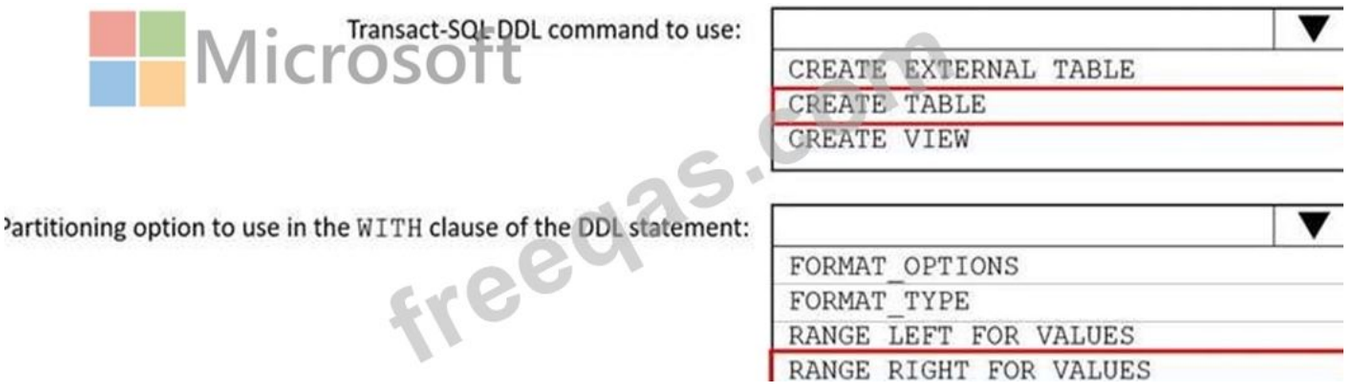
What solution must meet the sales transaction dataset requirements.

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

NOTE: Each correct selection is worth one point.



Answer:



Reference:
https://docs.microsoft.com/en-us/sql/t-sql/statements/create-table-azure-sql-data-warehouse

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 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)
Explanation
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: 5

You use PySpark in Azure Databricks to parse the following JSON input.

You need to output the data in the following tabular format.

owner	age	dog
Keith	30	Fido
Keith	30	Fluffy
Donna	46	Spot

How should you complete the PySpark code? To answer, drag the appropriate values to he 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.

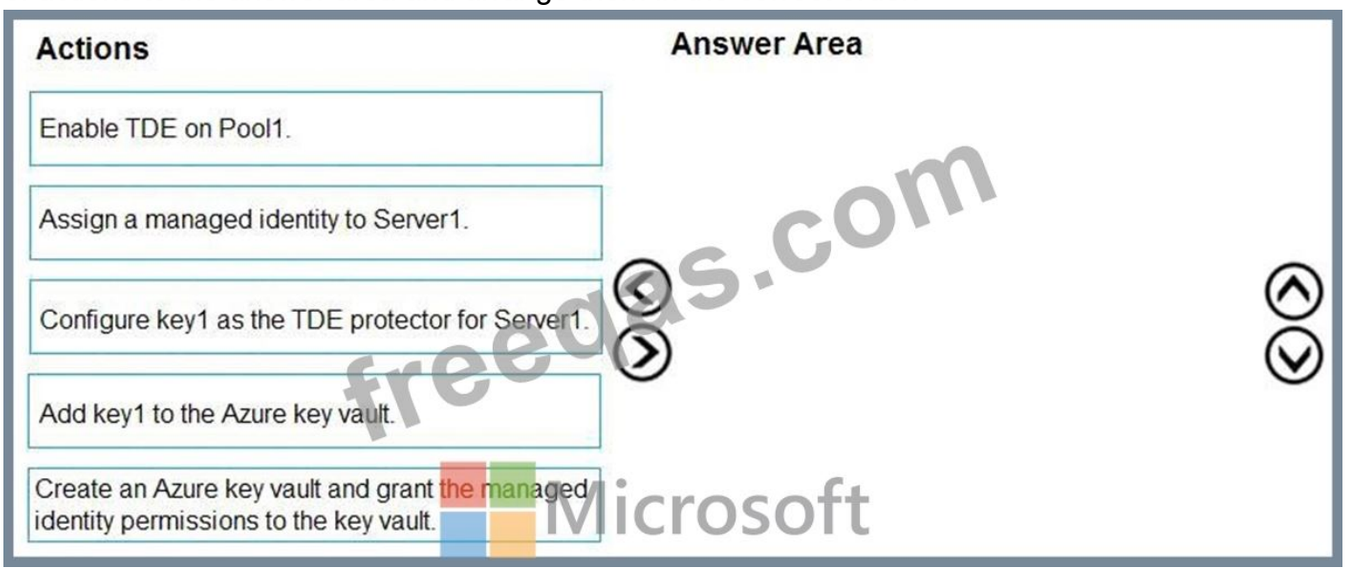
Answer:

NEW QUESTION: 6

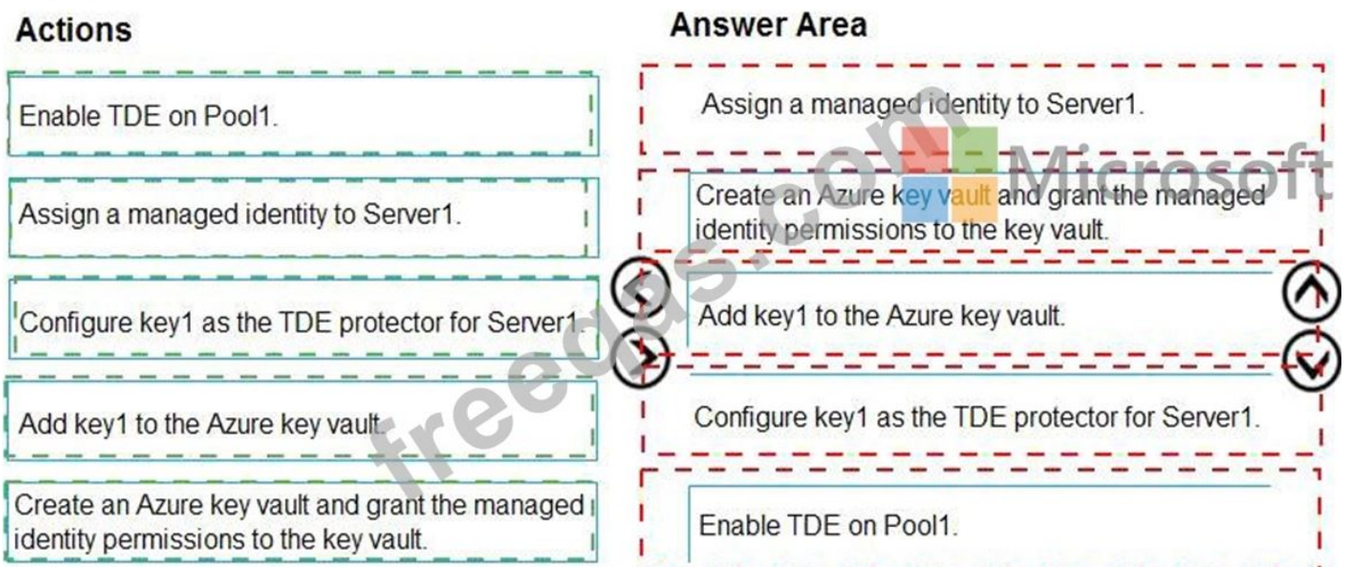
You have an Azure Synapse Analytics SQL pool named Pool1 on a logical Microsoft SQL server named Server1.

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.

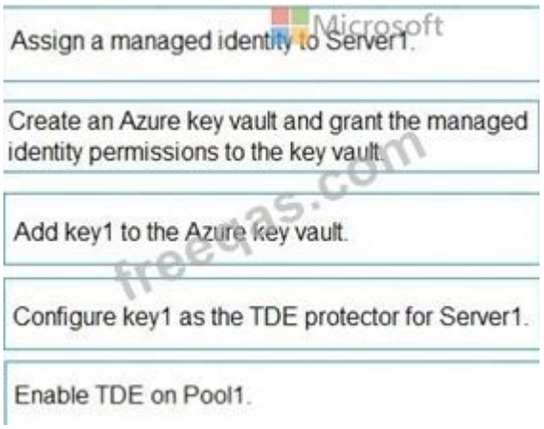


Answer:



Explanation

Graphical user interface, text, application Description automatically generated



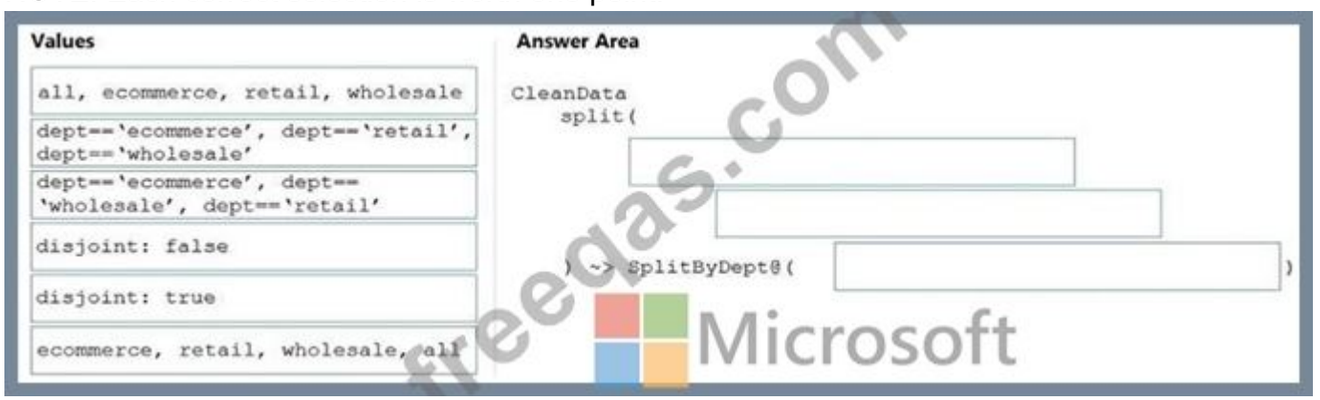
Step 1: Assign a managed identity to Server1
You will need an existing Managed Instance as a prerequisite.
Step 2: Create an Azure key vault and grant the managed identity permissions to the vault Create Resource and setup Azure Key Vault.
Step 3: Add key1 to the Azure key vault
The recommended way is 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-pow

NEW QUESTION: 7
You have a SQL pool in Azure Synapse.
A user reports that queries against the pool take longer than expected to complete.
You need to add monitoring to the underlying storage to help diagnose the issue.
Which two metrics should you monitor? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.
A. Cache used percentage
B. DWU Limit
C. Snapshot Storage Size
D. Active queries
E. Cache hit percentage
Answer: A,E (LEAVE A REPLY)
Explanation
A: Cache used is the sum of all bytes in the local SSD cache across all nodes and cache capacity is the sum of the storage capacity of the local SSD cache across all nodes.
E: Cache hits is the sum of all columnstore segments hits in the local SSD cache and cache miss is the columnstore segments misses in the local SSD cache summed across all nodes Reference:

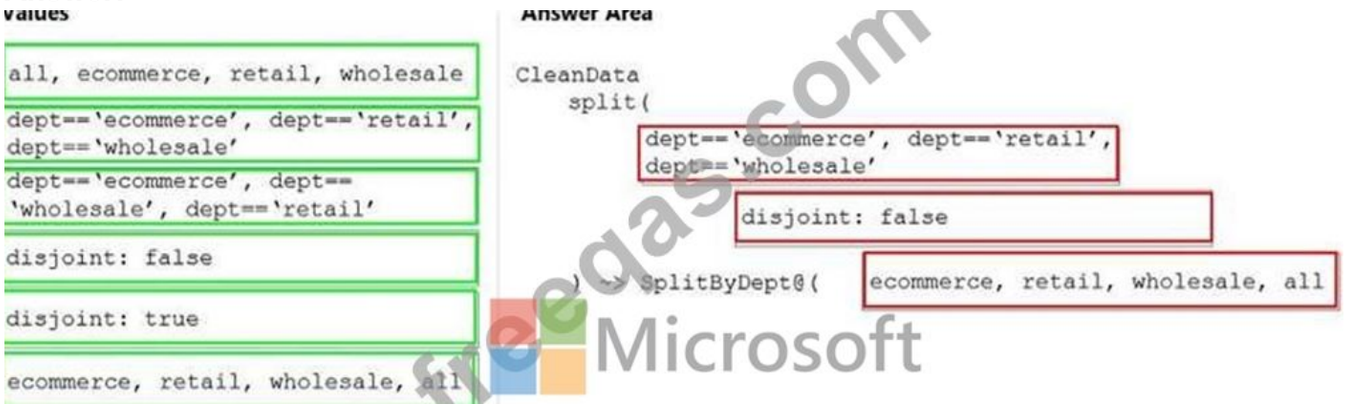
https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-concept-resour

NEW QUESTION: 8
You need to design a data retention solution for the Twitter teed 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. soft delete
C. change feed
D. time-based retention
Answer: B (LEAVE A REPLY)

NEW QUESTION: 9
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.



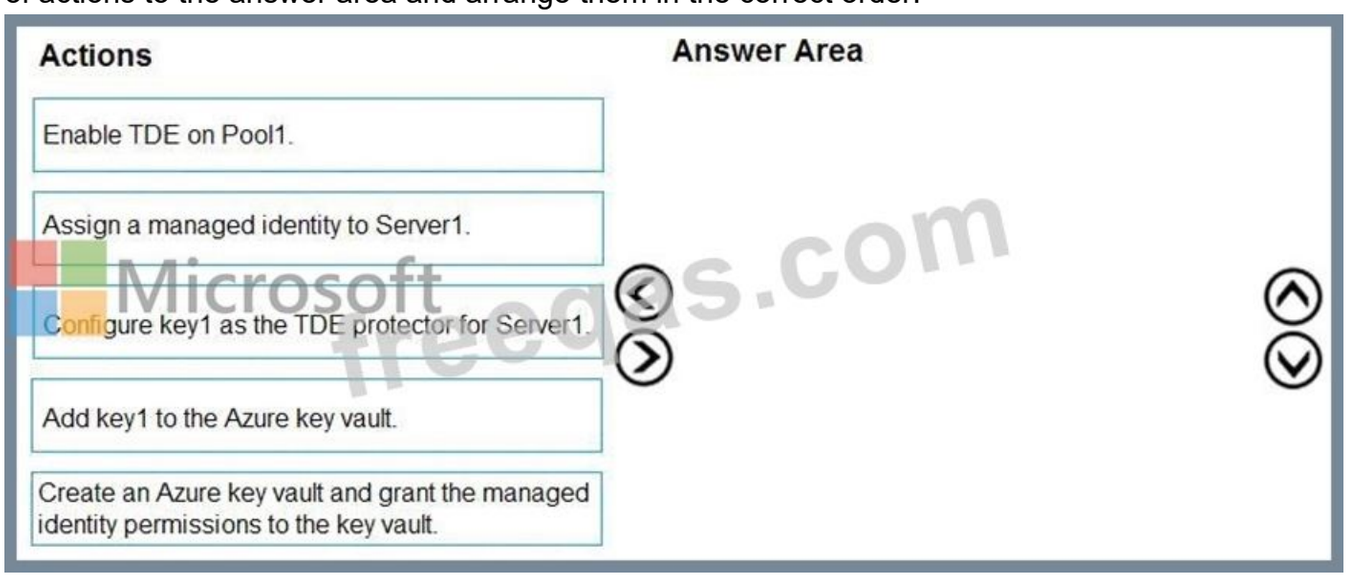
Answer:



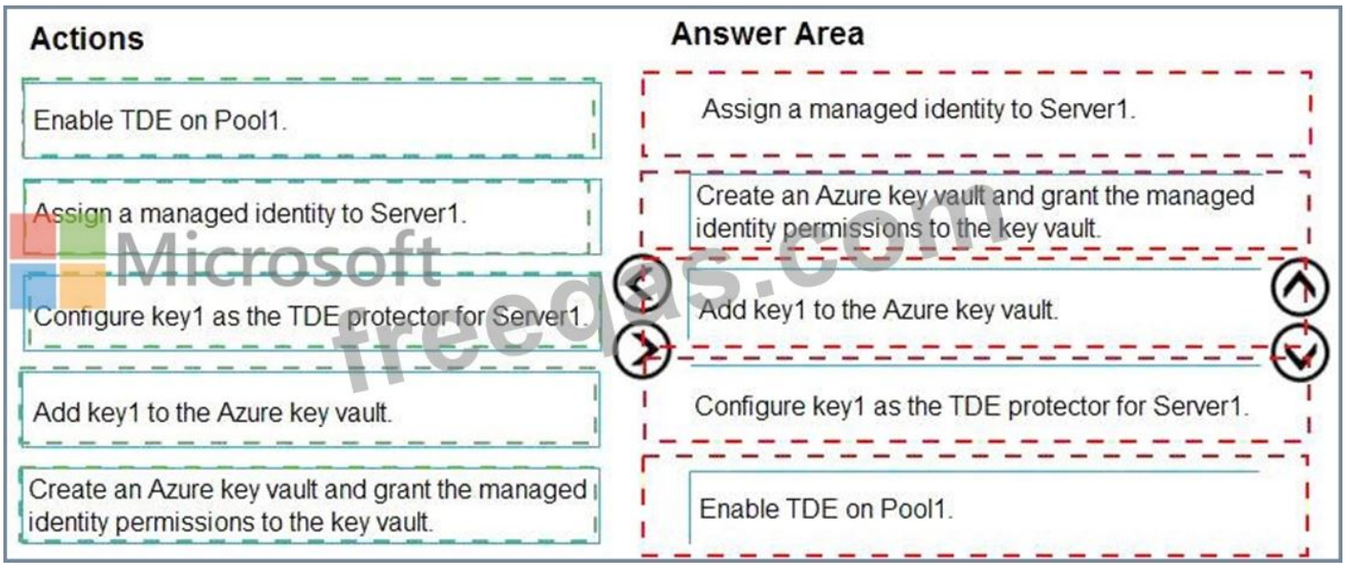
Reference:
https://docs.microsoft.com/en-us/azure/data-factory/data-flow-conditional-split

NEW QUESTION: 10
You are designing a highly available Azure Data Lake Storage solution that will induce geo-zone-redundant storage (GZRS).
You need to monitor for replication delays that can affect the recovery point objective (RPO).
What should you include m the monitoring solution?
A. Last Sync Time
B. Average Success Latency
C. Error errors
D. availability
Answer: (SHOW ANSWER)
Explanation
Because geo-replication is asynchronous, it is possible that data written to the primary region has not yet been written to the secondary region at the time an outage occurs. The Last Sync Time property indicates the last time that data from the primary region was written successfully to the secondary region. All writes made to the primary region before the last sync time are available to be read from the secondary location. Writes made to the primary region after the last sync time property may or may not be available for reads yet.
Reference:
https://docs.microsoft.com/en-us/azure/storage/common/last-sync-time-get

NEW QUESTION: 11
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.

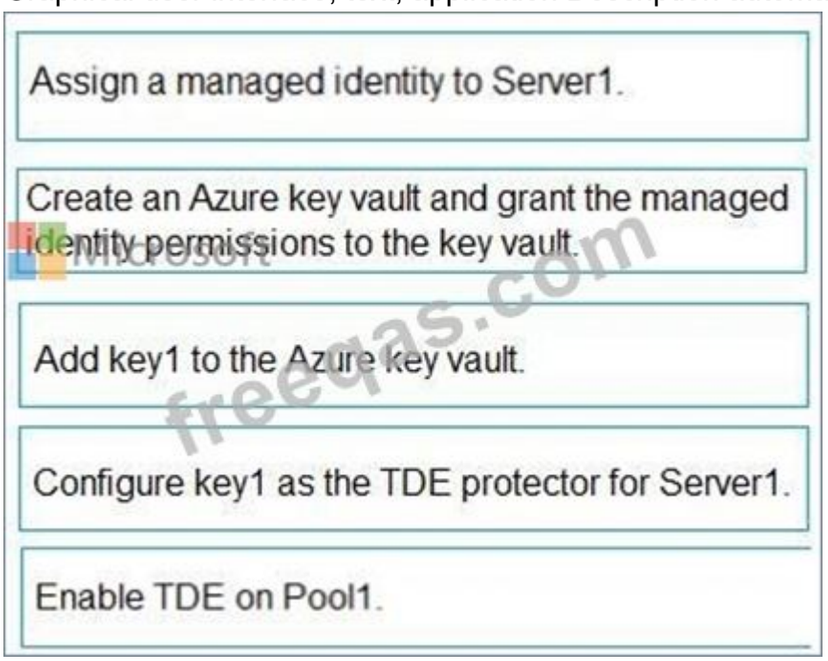


Answer:



Explanation

Graphical user interface, text, application Description automatically generated



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-pow

NEW QUESTION: 12
You have a SQL pool in Azure Synapse that contains a table named dbo.Customers. The table contains a column name 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: A (LEAVE A REPLY)
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,"XXXXXXX",0)') NULL, . . https://docs.microsoft.com/en-us/azure/azure-sql/database/dynamic-data-masking-overview upvoted 24 times

NEW QUESTION: 13
You have an Azure Synapse Analytics job that uses Scala.
You need to view the status of the job.
What should you do?
A. From Synapse Studio, select the workspace. From Monitor, select SQL requests.
B. From Synapse Studio, select the workspace. From Monitor, select Apache Sparks applications.
C. From Azure Monitor, run a Kusto query against the AzureDiagnostics table.
D. From Azure Monitor, run a Kusto query against the SparkLogying1 Event.CL table.
Answer: B (LEAVE A REPLY)

NEW QUESTION: 14
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: (SHOW ANSWER)

Explanation
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: 15
You have an enterprise-wide Azure Data Lake Storage Gen2 account. The data lake is accessible only through an Azure virtual network named VNET1.
You are building a SQL pool in Azure Synapse that will use data from the data lake.
Your company has a sales team. All the members of the sales team are in an Azure Active Directory group named Sales. POSIX controls are used to assign the Sales group access to the files in the data lake.
You plan to load data to the SQL pool every hour.
You need to ensure that the SQL pool can load the sales data from the data lake.
Which three actions should you perform? Each correct answer presents part of the solution.
NOTE: Each area selection is worth one point.
A. Add the managed identity to the Sales group.
B. Use the managed identity as the credentials for the data load process.
C. Create a shared access signature (SAS).
D. Add your Azure Active Directory (Azure AD) account to the Sales group.
E. Use the snared access signature (SAS) as the credentials for the data load process.
F. Create a managed identity.
Answer: A,D,F (LEAVE A REPLY)
The managed identity grants permissions to the dedicated SQL pools in the workspace.
Note: Managed identity for Azure resources is a feature of Azure Active Directory. The feature provides Azure services with an automatically managed identity in Azure AD Reference:
https://docs.microsoft.com/en-us/azure/synapse-analytics/security/synapse-workspace-managed-identity
Design and develop data processing Question Set 1

NEW QUESTION: 16

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.

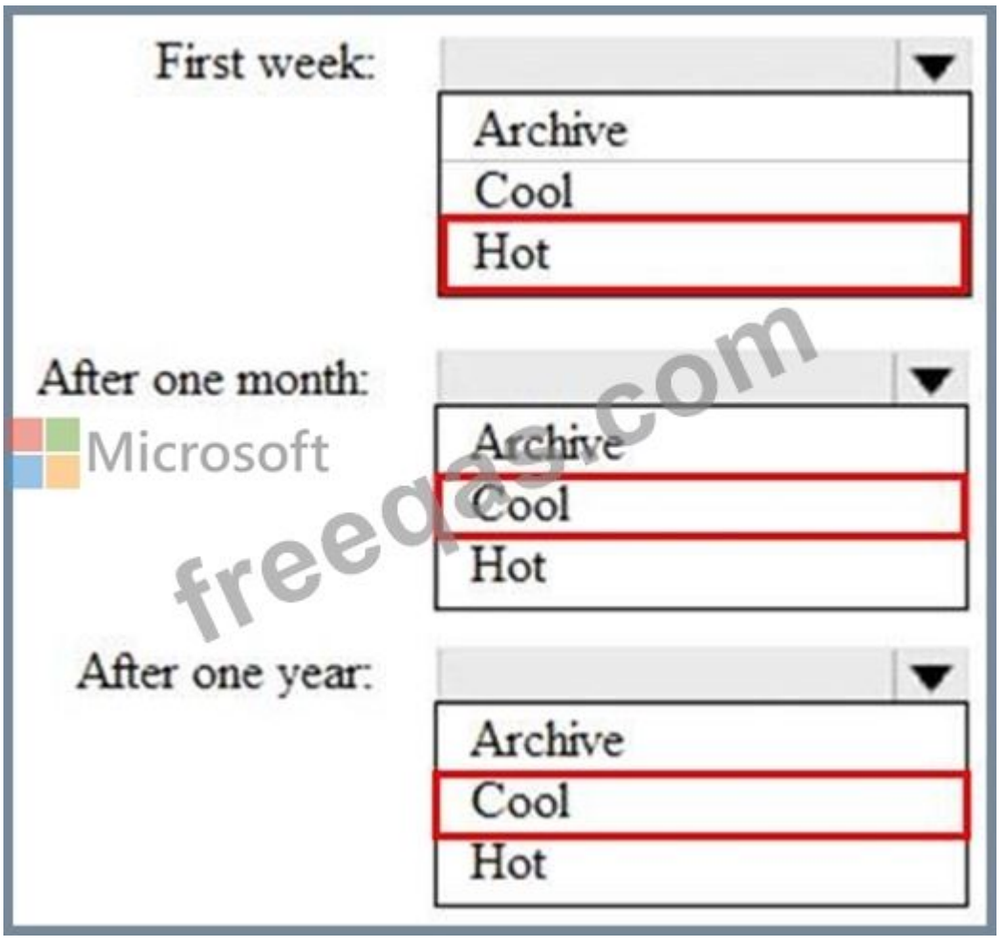
You need to select a storage strategy for the dat

a. 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.

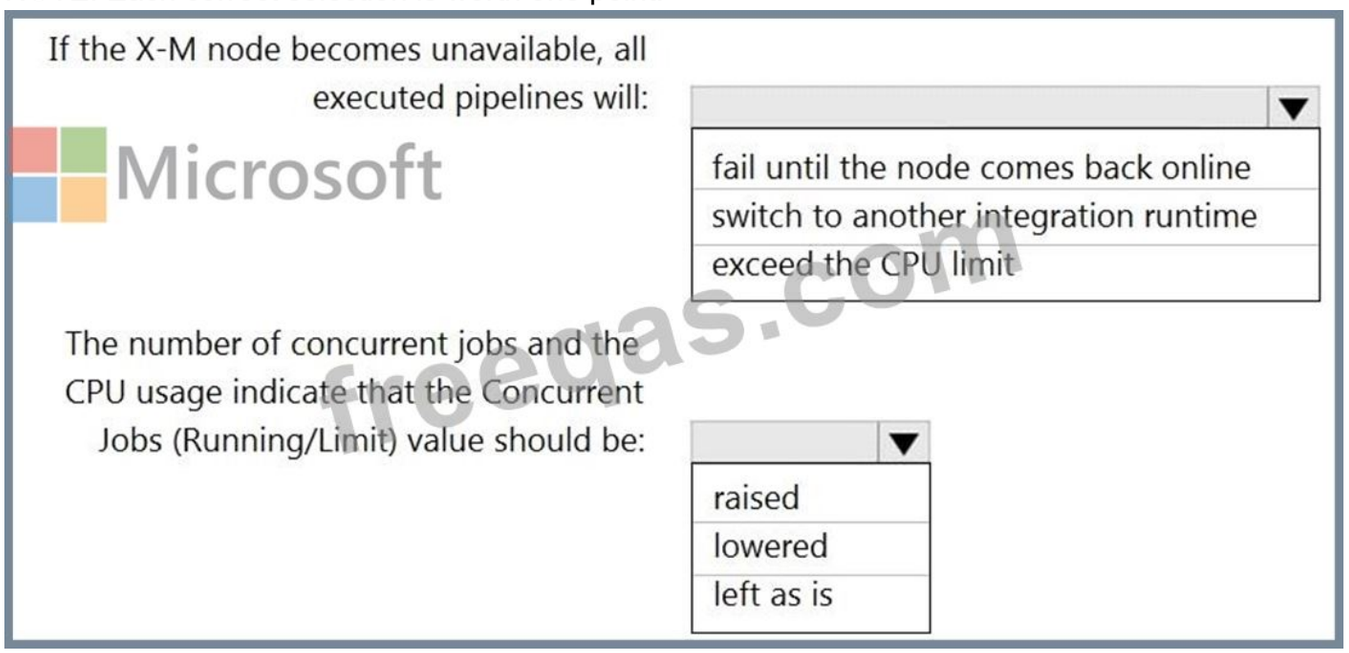
Answer:



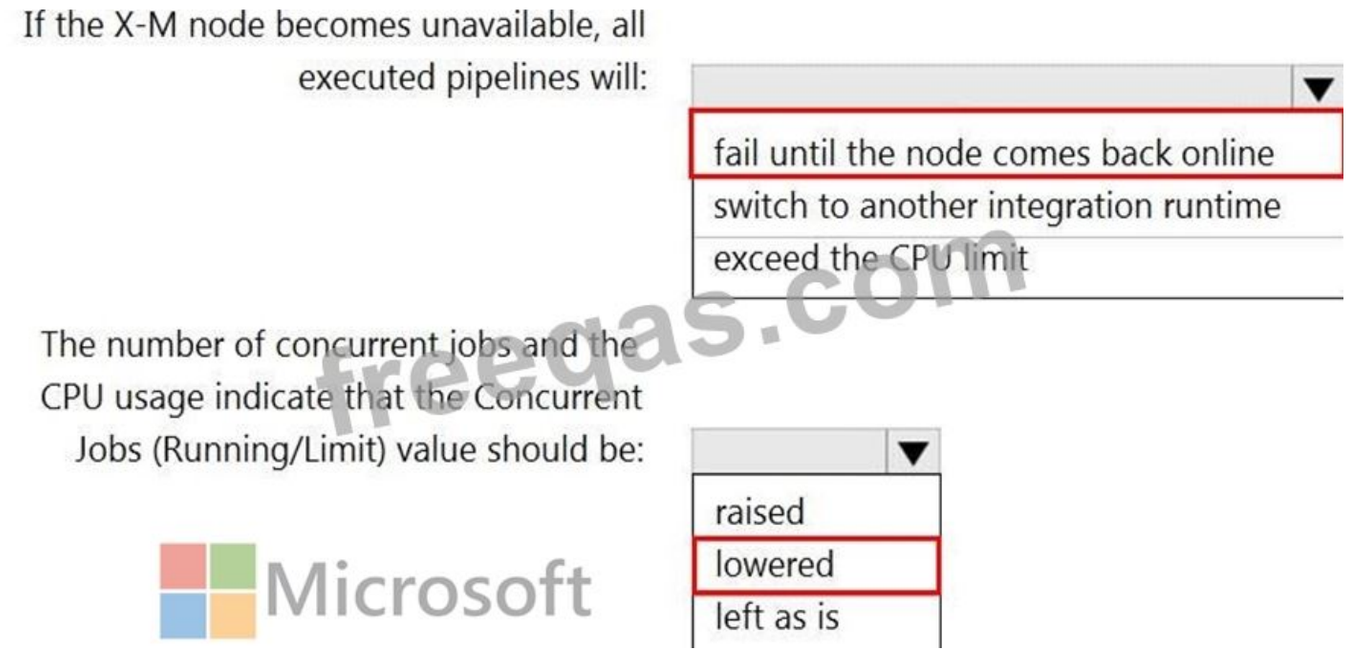
Reference:
https://docs.microsoft.com/en-us/azure/storage/blobs/storage-blob-storage-tiers

NEW QUESTION: 17
You have a self-hosted integration runtime in Azure Data Factory.
The current status of the integration runtime has the following configurations:
Status: Running
Type: Self-Hosted
Running / Registered Node(s): 1/1
High Availability Enabled: False
Linked Count: 0
Queue Length: 0
Average Queue Duration. 0.00s

The integration runtime has the following node details:
Name: X-M
Status: Running
Available Memory: 7697MB
CPU Utilization: 6%
Network (In/Out): 1.21KBps/0.83KBps
Concurrent Jobs (Running/Limit): 2/14
Role: Dispatcher/Worker
Credential Status: In Sync
Use the drop-down menus to select the answer choice that completes each statement based on the information presented.
NOTE: Each correct selection is worth one point.

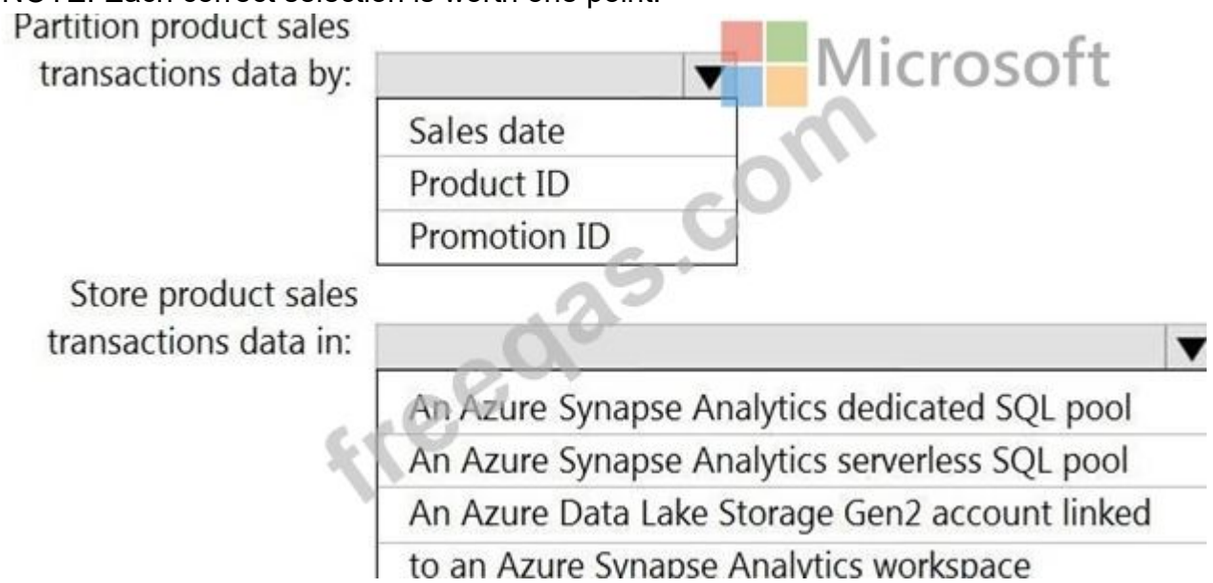


Answer:

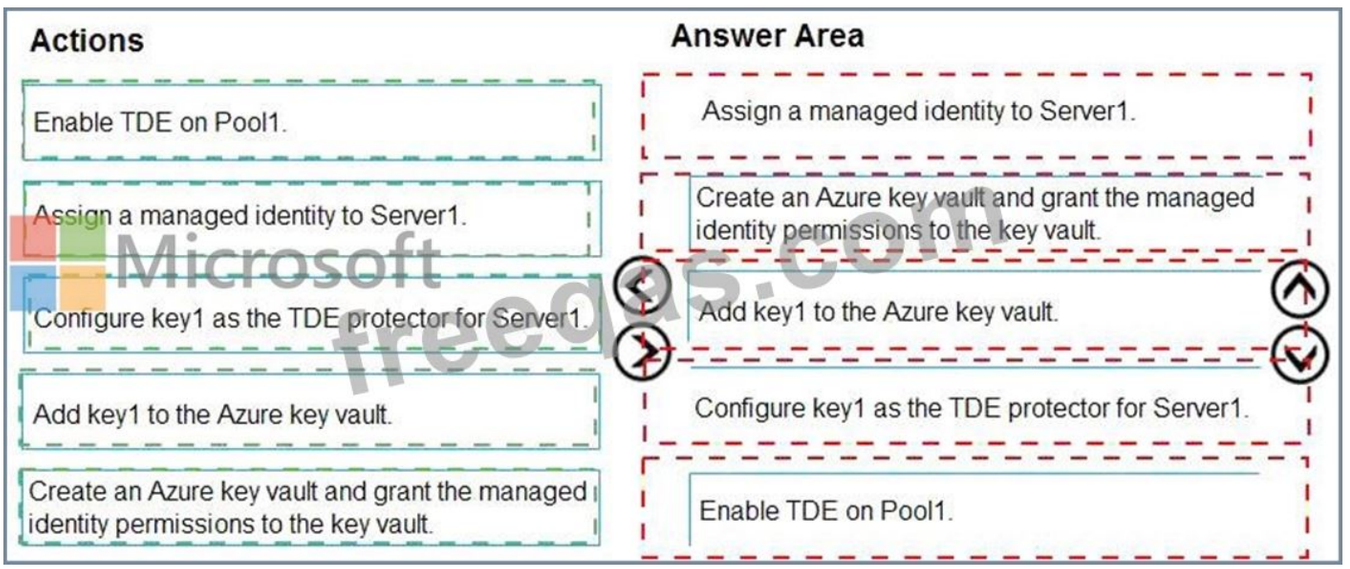


Reference:
https://docs.microsoft.com/en-us/azure/data-factory/create-self-hosted-integration-runtime

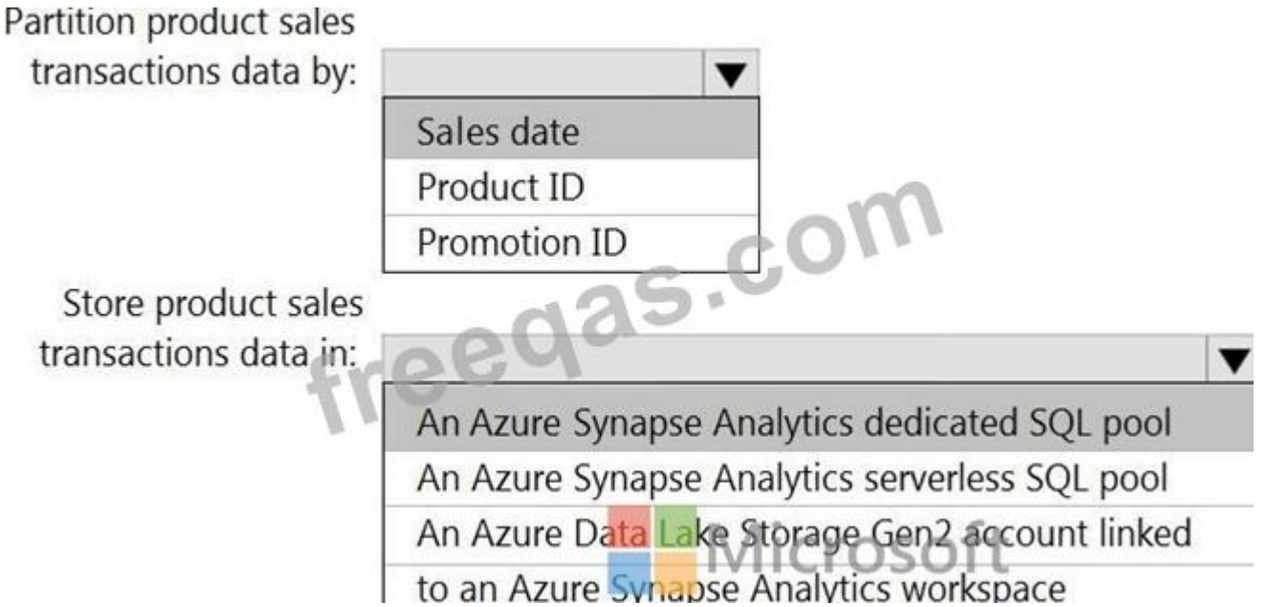
NEW QUESTION: 18
You need to design the partitions for the product sales transactions. The solution must meet the sales transaction dataset requirements.
What should you include in the solution? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.



Answer:



Explanation



Box 1: Sales date

Scenario: Contoso requirements for data integration include:

* Partition data that contains sales transaction records. Partitions must be designed to provide efficient loads by month. Boundary values must belong to the partition on the right.

Box 2: An Azure Synapse Analytics Dedicated SQL pool

Scenario: Contoso requirements for data integration include:

* Ensure that data storage costs and performance are predictable.

The size of a dedicated SQL pool (formerly SQL DW) is determined by Data Warehousing Units (DWU).

Dedicated SQL pool (formerly SQL DW) stores data in relational tables with columnar storage. This format significantly reduces the data storage costs, and improves query performance.

Synapse analytics dedicated sql pool

Reference:

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

NEW QUESTION: 19

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
Design and implement data security
Testlet 1
Case study
This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.
To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.
At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.
To start the case study
To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information

displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

Overview

Litware, Inc. owns and operates 300 convenience stores across the US. The company sells a variety of packaged foods and drinks, as well as a variety of prepared foods, such as sandwiches and pizzas.

Litware has a loyalty club whereby members can get daily discounts on specific items by providing their membership number at checkout.

Litware employs business analysts who prefer to analyze data by using Microsoft Power BI, and data scientists who prefer analyzing data in Azure Databricks notebooks.

Requirements

Business Goals

Litware wants to create a new analytics environment in Azure to meet the following requirements:

* See inventory levels across the stores. Data must be updated as close to real time as possible.
* Execute ad hoc analytical queries on historical data to identify whether the loyalty club discounts increase sales of the discounted products.
* Every four hours, notify store employees about how many prepared food items to produce based on historical demand from the sales data.

Technical Requirements

Litware identifies the following technical requirements:

* Minimize the number of different Azure services needed to achieve the business goals.
* Use platform as a service (PaaS) offerings whenever possible and avoid having to provision virtual machines that must be managed by Litware.
* Ensure that the analytical data store is accessible only to the company's on-premises network and Azure services.
* Use Azure Active Directory (Azure AD) authentication whenever possible.
* Use the principle of least privilege when designing security.
* Stage Inventory data in Azure Data Lake Storage Gen2 before loading the data into the analytical data store. Litware wants to remove transient data from Data Lake Storage once the data is no longer in use. Files that have a modified date that is older than 14 days must be removed.
* Limit the business analysts' access to customer contact information, such as phone numbers, because this type of data is not analytically relevant.
* Ensure that you can quickly restore a copy of the analytical data store within one hour in the event of corruption or accidental deletion.

Planned Environment

Litware plans to implement the following environment:

* 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.
* Customer data, including name, contact information, and loyalty number, comes from Salesforce, a SaaS application, and can be imported into Azure once every eight hours. Row modified dates are not trusted in the source table.

* Product data, including product ID, name, and category, comes from Salesforce and can be imported into Azure once every eight hours. Row modified dates are not trusted in the source table.
* Daily inventory data comes from a Microsoft SQL server located on a private network.
* Litware currently has 5 TB of historical sales data and 100 GB of customer data. The company expects approximately 100 GB of new data per month for the next year.
* Litware will build a custom application named FoodPrep to provide store employees with the calculation results of how many prepared food items to produce every four hours.
* Litware does not plan to implement Azure ExpressRoute or a VPN between the on-premises network and Azure.

NEW QUESTION: 20
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. change feed
B. soft delete
C. time-based retention
D. lifecycle management
Answer: (SHOW ANSWER)
Explanation
Scenario: Purge Twitter feed data records that are older than two years.
Data sets have unique lifecycles. Early in the lifecycle, people access some data often. But the need for access often drops drastically as the data ages. Some data remains idle in the cloud and is rarely accessed once stored.
Some data sets expire days or months after creation, while other data sets are actively read and modified throughout their lifetimes. Azure Storage lifecycle management offers a rule-based policy that you can use to transition blob data to the appropriate access tiers or to expire data at the end of the data lifecycle.
Reference:
https://docs.microsoft.com/en-us/azure/storage/blobs/lifecycle-management-overview

NEW QUESTION: 21
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.

Answer:

See the answer in explanation.

Explanation

Answer is below

NEW QUESTION: 22

You are building an Azure Stream Analytics job to identify how much time a user spends interacting with a feature on a webpage.

The job receives events based on user actions on the webpage. Each row of data represents an event. Each event has a type of either 'start' or 'end'.

You need to calculate the duration between start and end events.

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:

Explanation

Box 1: DATEDIFF

DATEDIFF function returns the count (as a signed integer value) of the specified datepart boundaries crossed between the specified startdate and enddate.

Syntax: DATEDIFF (datepart , startdate, enddate)

Box 2: LAST

The LAST function can be used to retrieve the last event within a specific condition. In this example, the condition is an event of type Start, partitioning the search by PARTITION BY user and feature. This way, every user and feature is treated independently when searching for the Start event. LIMIT DURATION limits the search back in time to 1 hour between the End and Start events.

Example:

SELECT

[user],

feature,

DATEDIFF(

second,

LAST(Time) OVER (PARTITION BY [user], feature LIMIT DURATION(hour,

1) WHEN Event = 'start'),

Time) as duration

FROM input TIMESTAMP BY Time

WHERE

Event = 'end'

Reference:

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-stream-analytics-query-patterns

NEW QUESTION: 23

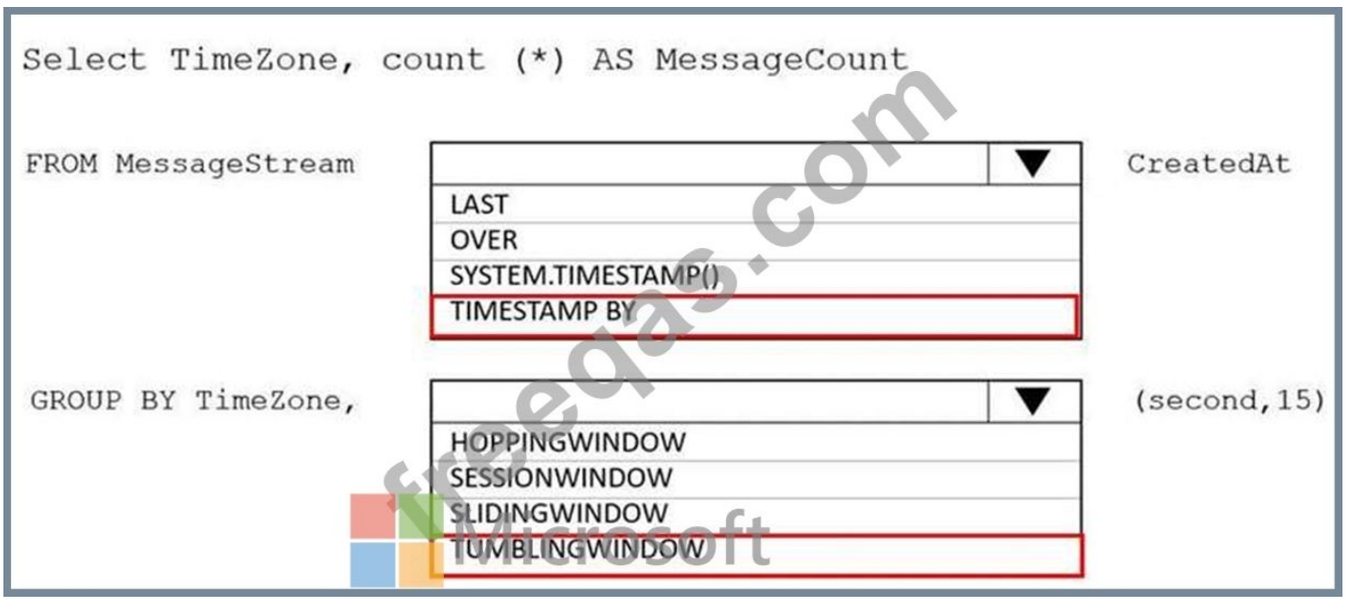
You are designing an Azure Stream Analytics solution that receives instant messaging data from an Azure Event Hub.

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 are a.

NOTE: Each correct selection is worth one point.

Answer:



Reference:

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

NEW QUESTION: 24

You are designing a dimension table for a data warehouse. The table will track the value of the dimension attributes over time and preserve the history of the data by adding new rows as the data changes.
Which type of slowly changing dimension (SCD) should use?

A. Type 0

B. Type 1

C. Type 2

D. Type 3

Answer: C (LEAVE A REPLY)

Type 2 - Creating a new additional record. In this methodology all history of dimension changes is kept in the database. You capture attribute change by adding a new row with a new surrogate key to the dimension table. Both the prior and new rows contain as attributes the natural key(or other durable identifier). Also 'effective date' and 'current indicator' columns are used in this method. There could be only one record with current indicator set to 'Y'. For 'effective date' columns, i.e. start_date and end_date, the end_date for current record usually is set to value 9999-12-31. Introducing changes to the dimensional model in type 2 could be very expensive database operation so it is not recommended to use it in dimensions where a new attribute could be added in the future.
https://www.datawarehouse4u.info/SCD-Slowly-Changing-Dimensions.html

NEW QUESTION: 25

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 hopping window that uses a hop size of 5 seconds and a window size 10 seconds.
Does this meet the goal?

A. Yes

B. No

Answer: (SHOW ANSWER)

Explanation
Instead use a tumbling window. Tumbling windows are a series of fixed-sized, non-overlapping and contiguous time intervals.
Reference:
https://docs.microsoft.com/en-us/stream-analytics-query/tumbling-window-azure-stream-analytics

NEW QUESTION: 26

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:

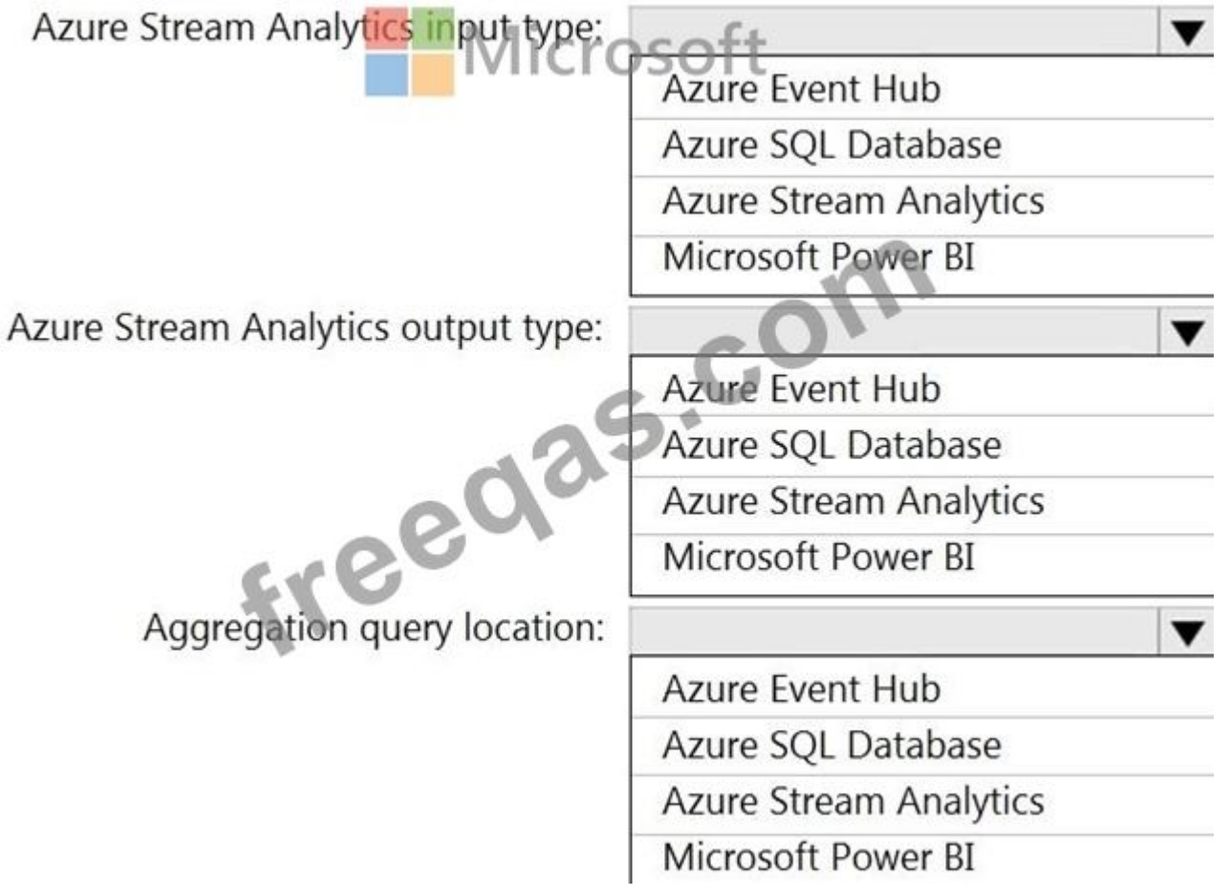
Minimize latency from an Azure Event hub to the dashboard.

Minimize the required storage.

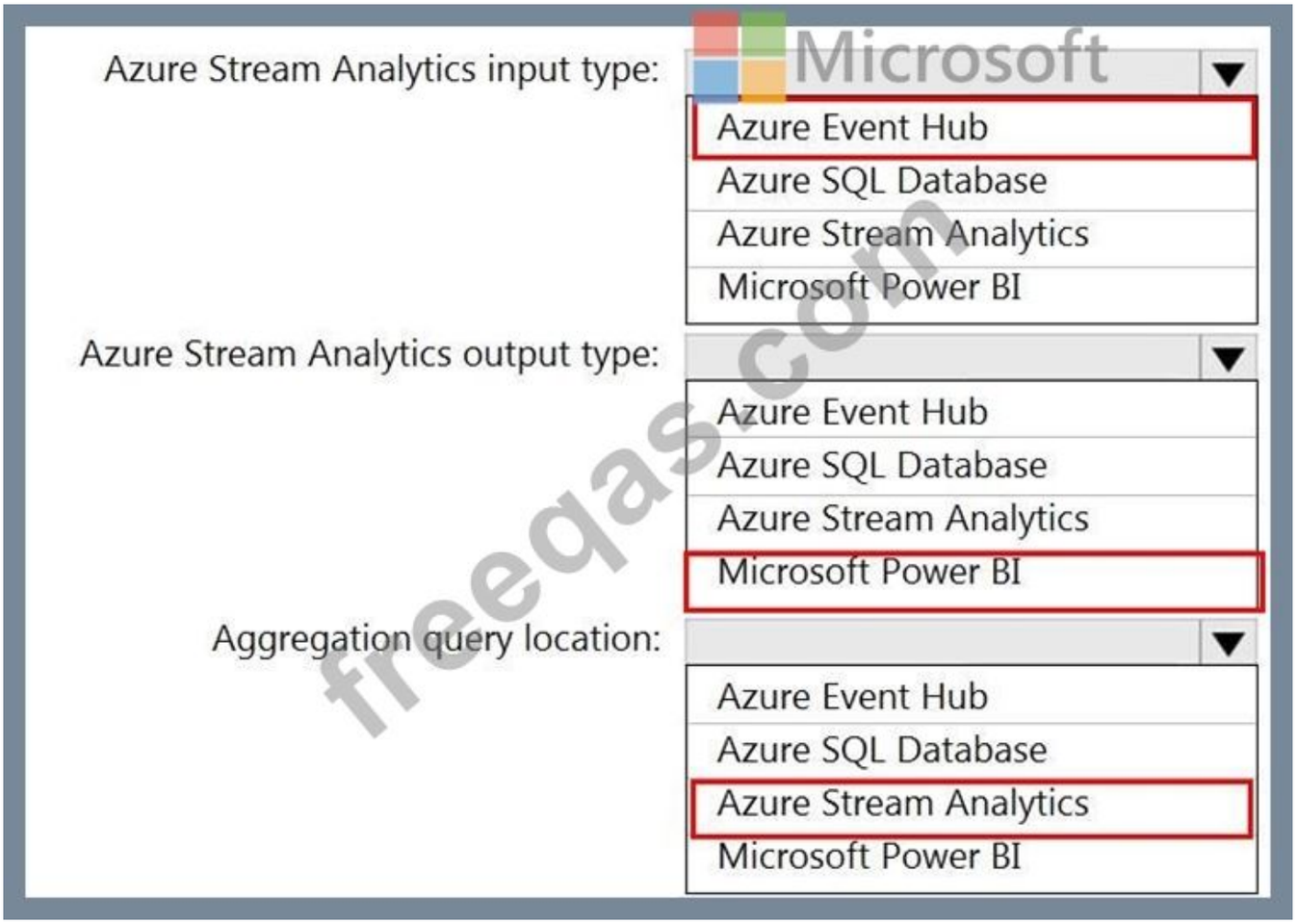
Minimize development effort.

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

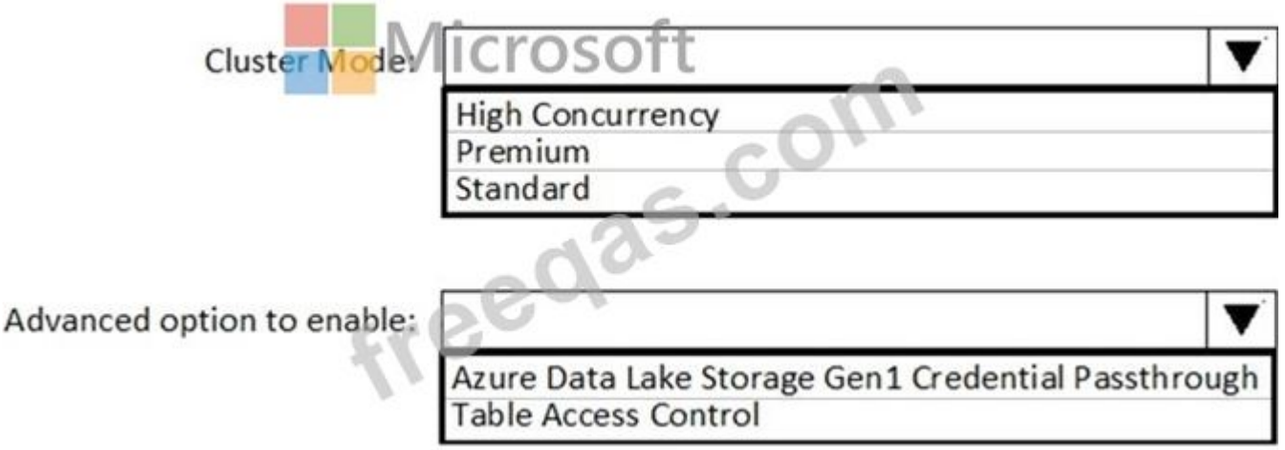


Answer:

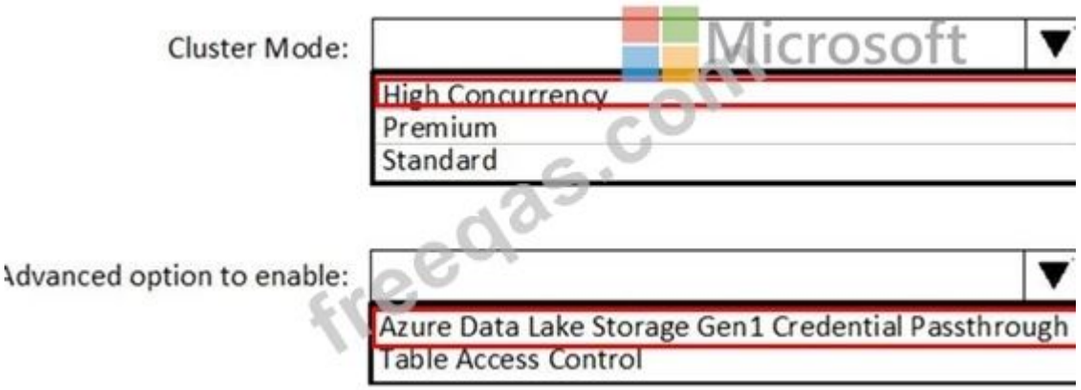


Reference:
https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-power-bi-dashboard

NEW QUESTION: 27
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.



Answer:



References:
https://docs.azuredatabricks.net/spark/latest/data-sources/azure/adls-passthrough.html

NEW QUESTION: 28
The following code segment is used to create an Azure Databricks cluster.

```
{
     "num_workers": null,
     "autoscale": {
          "min_workers": 2,
          "max_workers": 8
     },
     "cluster_name": "MyCluster",
     "spark_version": "latest-stable-scala2.11",
     "spark_conf": {
          "spark.databricks.cluster.profile": "serverless",
          "spark.databricks.repl.allowedLanguages": "sql,python,r"
     },
     "node_type_id": "Standard_DS13_v2",
     "ssh_public_keys": [],
     "custom_tags": {
          "ResourceClass": "Serverless"
     },
     "spark_env_vars": {
          "PYSPARK_PYTHON": "/databricks/python3/bin/python3"
     },
     "autotermination_minutes": 90,
     "enable_elastic_disk": true,
     "init_scripts": []
}
```

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
The Databricks cluster supports multiple concurrent users.	○	○
The Databricks cluster minimizes costs when running scheduled jobs that execute notebooks.	○	○
The Databricks cluster supports the creation of a Delta Lake table	○	○

Answer:

Statements	Yes	No
The Databricks cluster supports multiple concurrent users.	○	○
The Databricks cluster minimizes costs when running scheduled jobs that execute notebooks.	○	○
The Databricks cluster supports the creation of a Delta Lake table	○	○

Reference:
https://adatis.co.uk/databricks-cluster-sizing/
https://docs.microsoft.com/en-us/azure/databricks/jobs
https://docs.databricks.com/administration-guide/capacity-planning/cmbp.html
https://docs.databricks.com/delta/index.html

NEW QUESTION: 29
You have an Azure Storage account that generates 200.000 new files daily. The file names have a format of (YYY)/(MM)/(DD)/|HH])/(CustornerID).csv.
You need to design an Azure Data Factory solution that will toad 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.
Answer:
See the answer below in explanation.
Explanation
Answer as below

NEW QUESTION: 30
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 Hopping window that has one-minute hop
C. a five-minute Sliding window
D. a five-minute Tumbling window
Answer: (SHOW ANSWER)

NEW QUESTION: 31
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)
Explanation
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-v
https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/performance-tuning-result-set-cach

NEW QUESTION: 32
You are designing an enterprise data warehouse in Azure Synapse Analytics that will contain a table named Customers. Customers will contain credit card information.
You need to recommend a solution to provide salespeople with the ability to view all the entries in Customers.
The solution must prevent all the salespeople from viewing or inferring the credit card information.
What should you include in the recommendation?
A. data masking
B. Always Encrypted
C. column-level security
D. row-level security
Answer: A (LEAVE A REPLY)
SQL Database dynamic data masking limits sensitive data exposure by masking it to non-privileged users. The Credit card masking method exposes the last four digits of the designated fields and adds a constant string as a prefix in the form of a credit card.
Example: XXXX-XXXX-XXXX-1234
Reference:
https://docs.microsoft.com/en-us/azure/sql-database/sql-database-dynamic-data-masking-get-started
Monitor and optimize data storage and data processing Testlet 1 Case study This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.
To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.
At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.
To start the case study
To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information

displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

Overview

Litware, Inc. owns and operates 300 convenience stores across the US. The company sells a variety of packaged foods and drinks, as well as a variety of prepared foods, such as sandwiches and pizzas.

Litware has a loyalty club whereby members can get daily discounts on specific items by providing their membership number at checkout.

Litware employs business analysts who prefer to analyze data by using Microsoft Power BI, and data scientists who prefer analyzing data in Azure Databricks notebooks.

Requirements

Business Goals

Litware wants to create a new analytics environment in Azure to meet the following requirements:

* See inventory levels across the stores. Data must be updated as close to real time as possible.
* Execute ad hoc analytical queries on historical data to identify whether the loyalty club discounts increase sales of the discounted products.
* Every four hours, notify store employees about how many prepared food items to produce based on historical demand from the sales data.

Technical Requirements

Litware identifies the following technical requirements:

* Minimize the number of different Azure services needed to achieve the business goals.
* Use platform as a service (PaaS) offerings whenever possible and avoid having to provision virtual machines that must be managed by Litware.
* Ensure that the analytical data store is accessible only to the company's on-premises network and Azure services.
* Use Azure Active Directory (Azure AD) authentication whenever possible.
* Use the principle of least privilege when designing security.
* Stage Inventory data in Azure Data Lake Storage Gen2 before loading the data into the analytical data store. Litware wants to remove transient data from Data Lake Storage once the data is no longer in use. Files that have a modified date that is older than 14 days must be removed.
* Limit the business analysts' access to customer contact information, such as phone numbers, because this type of data is not analytically relevant.
* Ensure that you can quickly restore a copy of the analytical data store within one hour in the event of corruption or accidental deletion.

Planned Environment

Litware plans to implement the following environment:

* 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.
* Customer data, including name, contact information, and loyalty number, comes from Salesforce, a SaaS application, and can be imported into Azure once every eight hours. Row modified dates are not trusted in the source table.

* Product data, including product ID, name, and category, comes from Salesforce and can be imported into Azure once every eight hours. Row modified dates are not trusted in the source table.
* Daily inventory data comes from a Microsoft SQL server located on a private network.
* Litware currently has 5 TB of historical sales data and 100 GB of customer data. The company expects approximately 100 GB of new data per month for the next year.
* Litware will build a custom application named FoodPrep to provide store employees with the calculation results of how many prepared food items to produce every four hours.
* Litware does not plan to implement Azure ExpressRoute or a VPN between the on-premises network and Azure.

NEW QUESTION: 33
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 Standard cluster for the data engineers, and a High Concurrency cluster for the jobs.
Does this meet the goal?
A. Yes
B. No
Answer: B (LEAVE A REPLY)
We need a High Concurrency cluster for the data engineers and 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: 34
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

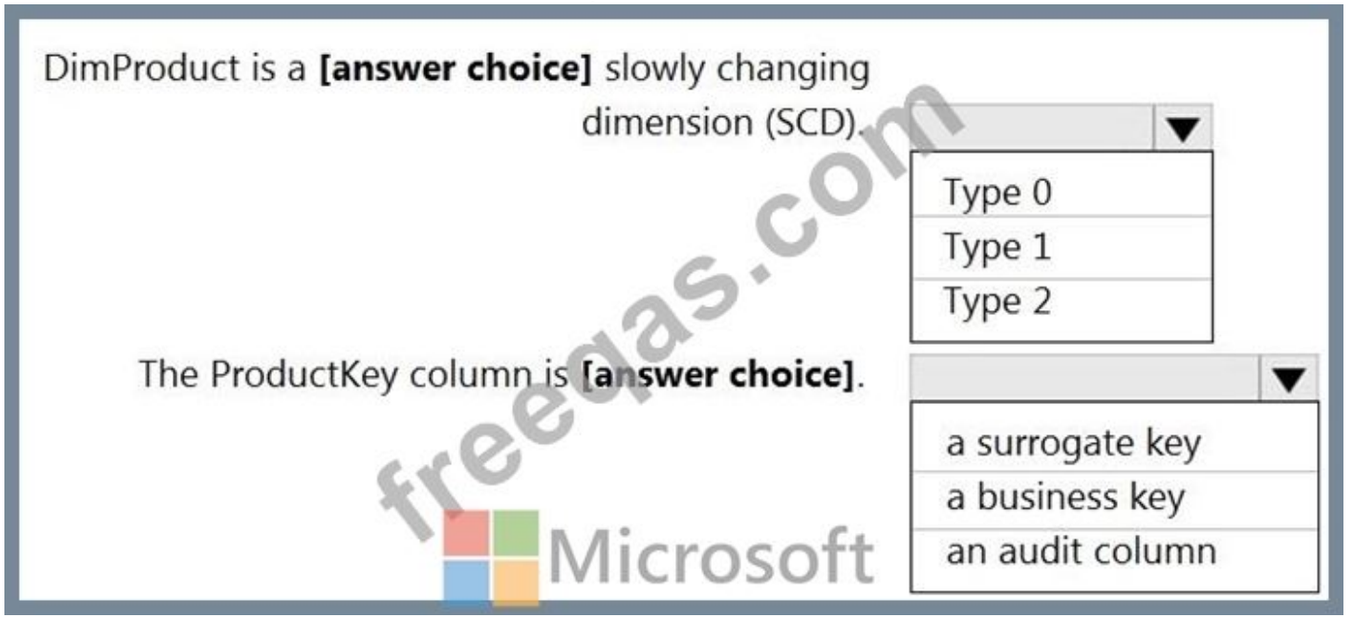
You need to recommend a solution that maximizes query performance.
What should you include in the recommendation?
A. In the tables use a hash distribution of ArrivalDateTime and ReportDateTime.
B. In the tables use a hash distribution of ArrivalAirportID and AirportID.
C. In each table, create an identity column.
D. In each table, create a column as a composite of the other two columns in the table.
Answer: (SHOW ANSWER)
Explanation
Hash-distribution improves query performance on large fact tables.

NEW QUESTION: 35
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.

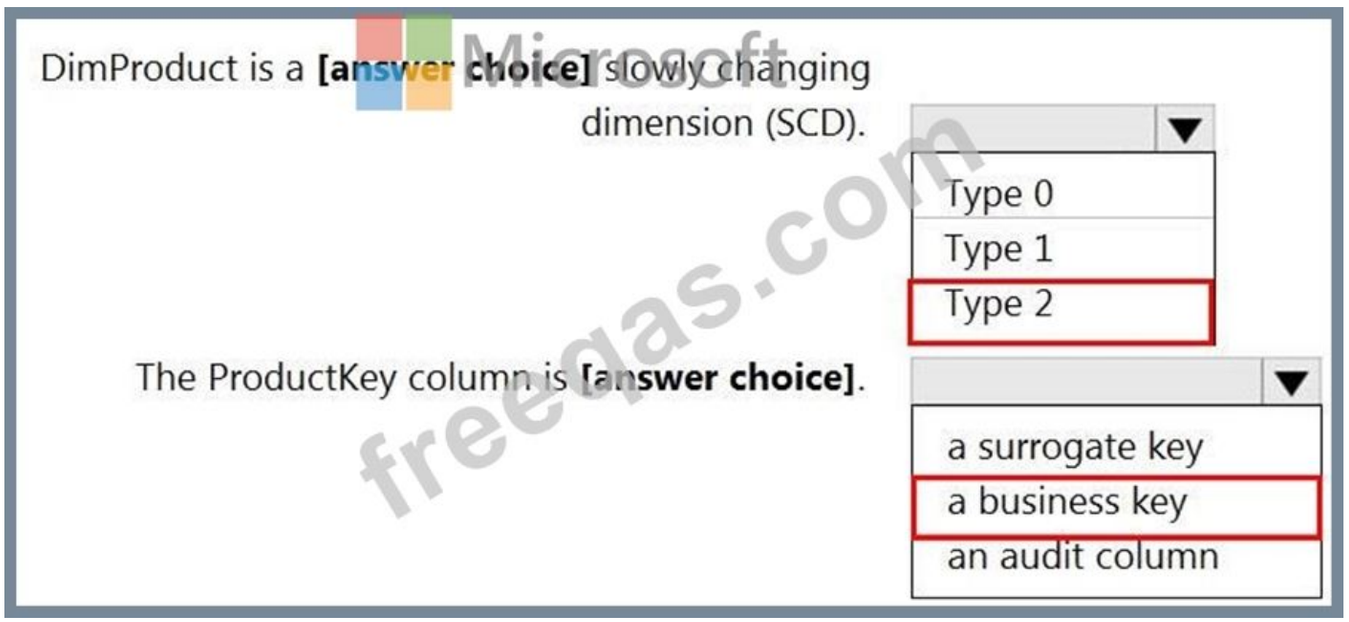
```
CREATE TABLE [DBO].[DimProduct](
       [ProductKey] [int] IDENTITY(1,1) NOT NULL,
       [ProductSourceID] [int] NOT NULL,
       [ProductName] [nvarchar](100) NOT NULL,
       [ProductNumber] [nvarchar](25) NOT NULL,
       [Color] [nvarchar](15) NULL,
       [Size] [nvarchar](5) NULL,
       [Weight] [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.



Answer:



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

NEW QUESTION: 36

You are building an Azure Synapse Analytics dedicated SQL pool that will contain a fact table for transactions from the first half of the year 2020.

You need to ensure that the table meets the following requirements:

* Minimizes the processing time to delete data that is older than 10 years
* Minimizes the I/O for queries that use year-to-date values

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.

Answer Area

```
CREATE TABLE [dbo].[FactTransaction]
(
     [TransactionTypeID]    int      NOT NULL
,    [TransactionDateID]    int      NOT NULL
,    [CustomerID]           int      NOT NULL
,    [RecipientID]          int      NOT NULL
,    [Amount]               money    NOT NU::
)
WITH
(
```

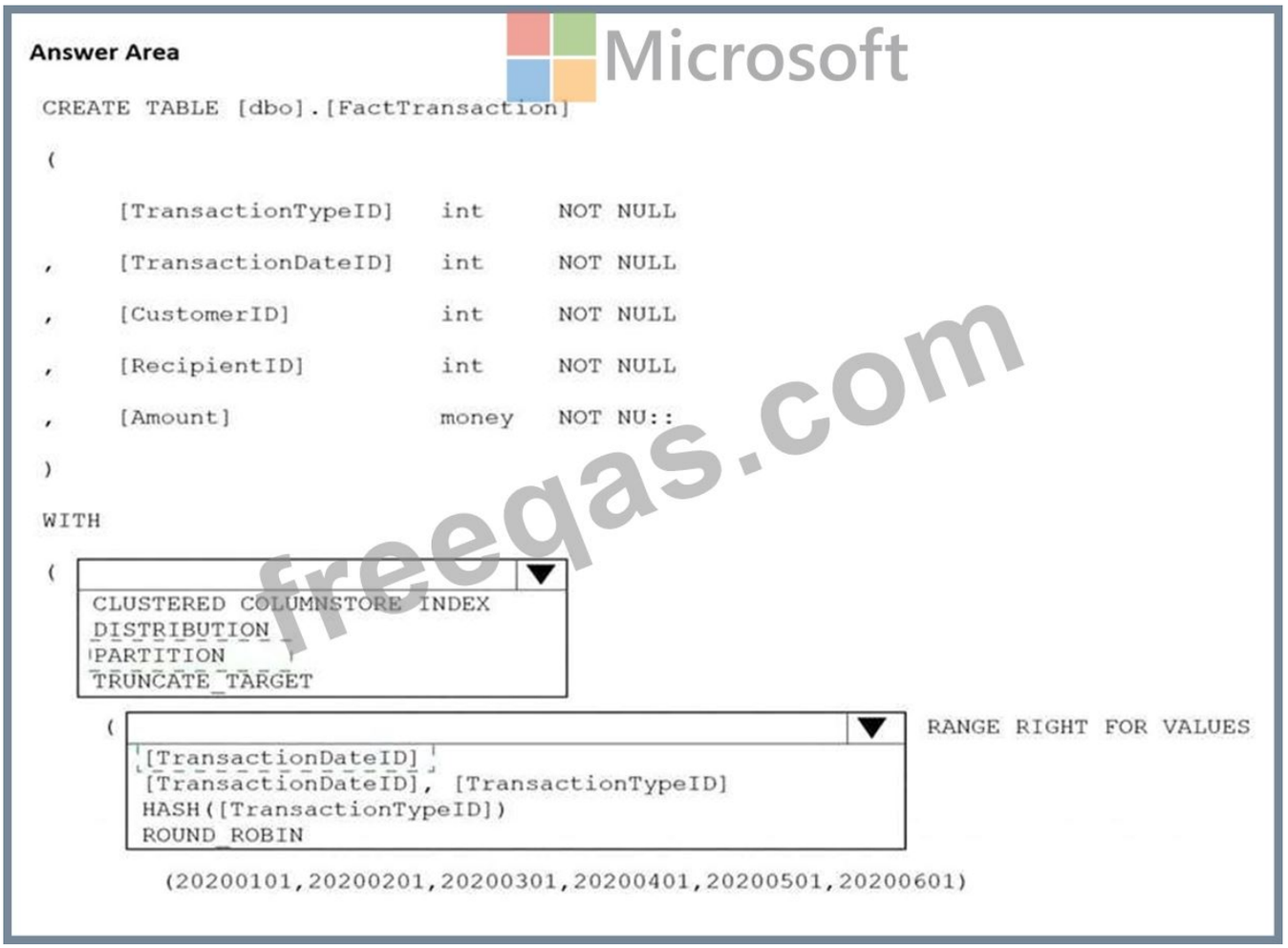
[CLUSTERED COLUMNSTORE INDEX | DISTRIBUTION | PARTITION | TRUNCATE_TARGET]

```
(
```

[[TransactionDateID] | [TransactionDateID], [TransactionTypeID] | HASH([TransactionTypeID]) | ROUND_ROBIN]

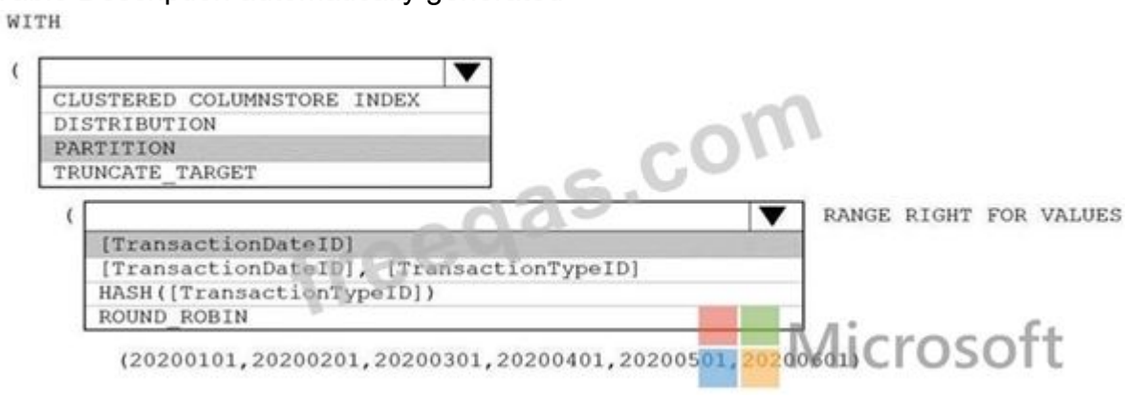
```
RANGE RIGHT FOR VALUES
(20200101,20200201,20200301,20200401,20200501,20200601)
```


Answer:



Explanation

Table Description automatically generated



Box 1: PARTITION

RANGE RIGHT FOR VALUES is used with PARTITION.

Part 2: [TransactionDateID]

Partition on the date column.

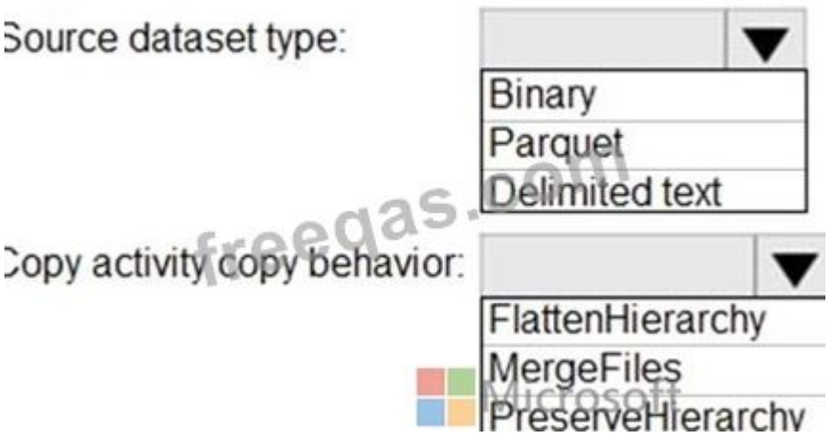
Example: Creating a RANGE RIGHT partition function on a datetime column The following partition function partitions a table or index into 12 partitions, one for each month of a year's worth of values in a datetime column.

CREATE PARTITION FUNCTION [myDateRangePF1] (datetime)

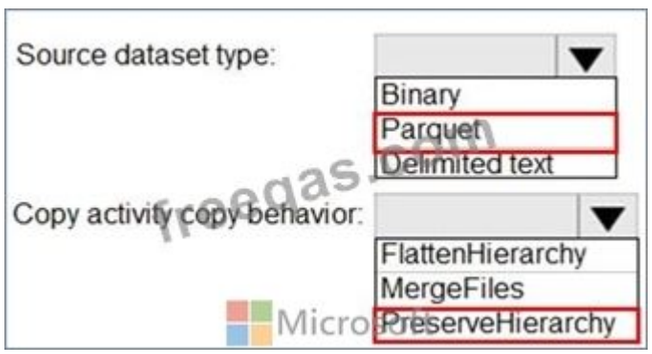
AS RANGE RIGHT FOR VALUES ('20030201', '20030301', '20030401',

'20030501', '20030601', '20030701', '20030801',
'20030901', '20031001', '20031101', '20031201');
Reference:
https://docs.microsoft.com/en-us/sql/t-sql/statements/create-partition-function-transact-sql

NEW QUESTION: 37
You have two Azure Storage accounts named Storage1 and Storage2. Each account holds one container and has the hierarchical namespace enabled. The system has files that contain data stored in the Apache Parquet format.
You need to copy folders and files from Storage1 to Storage2 by using a Data Factory copy activity. The solution must meet the following requirements:
No transformations must be performed.
The original folder structure must be retained.
Minimize time required to perform the copy activity.
How should you configure the copy activity? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.



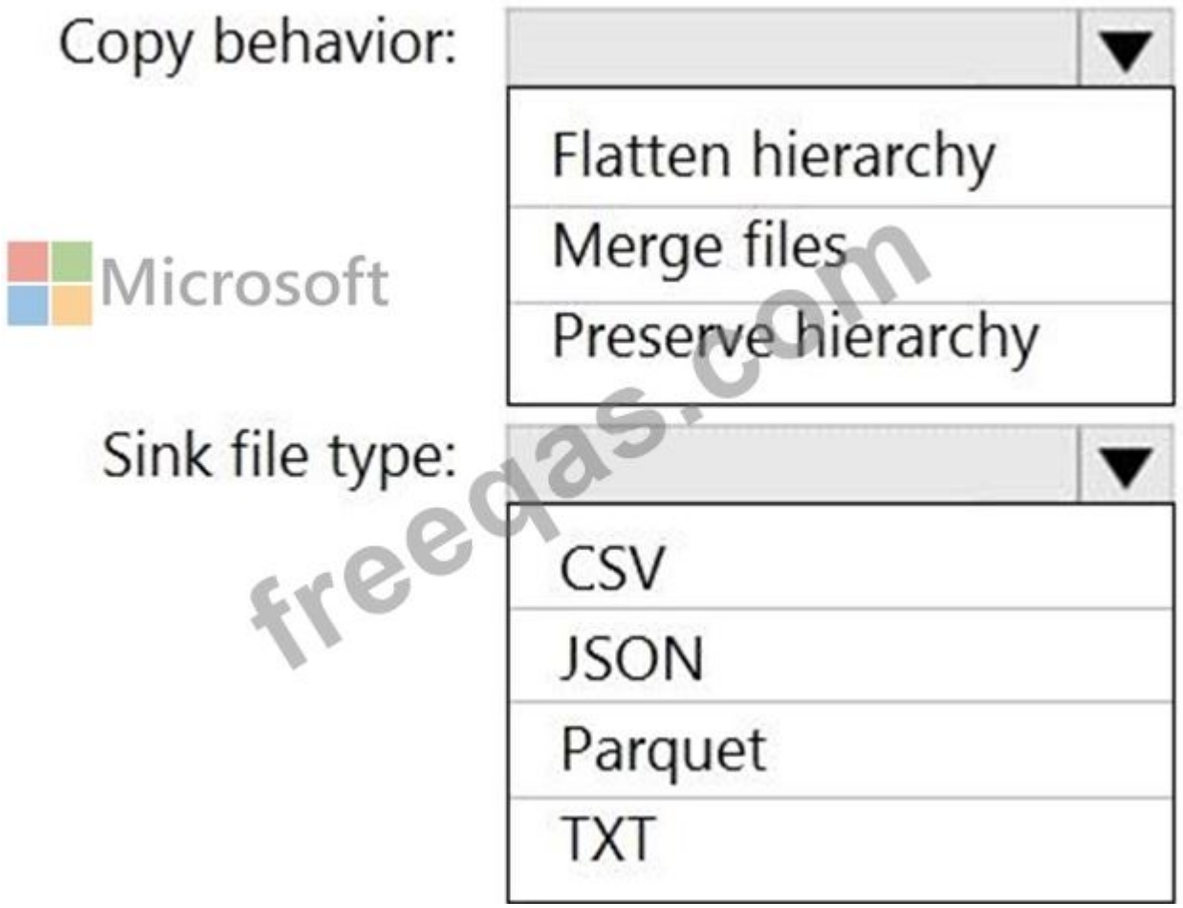
Answer:



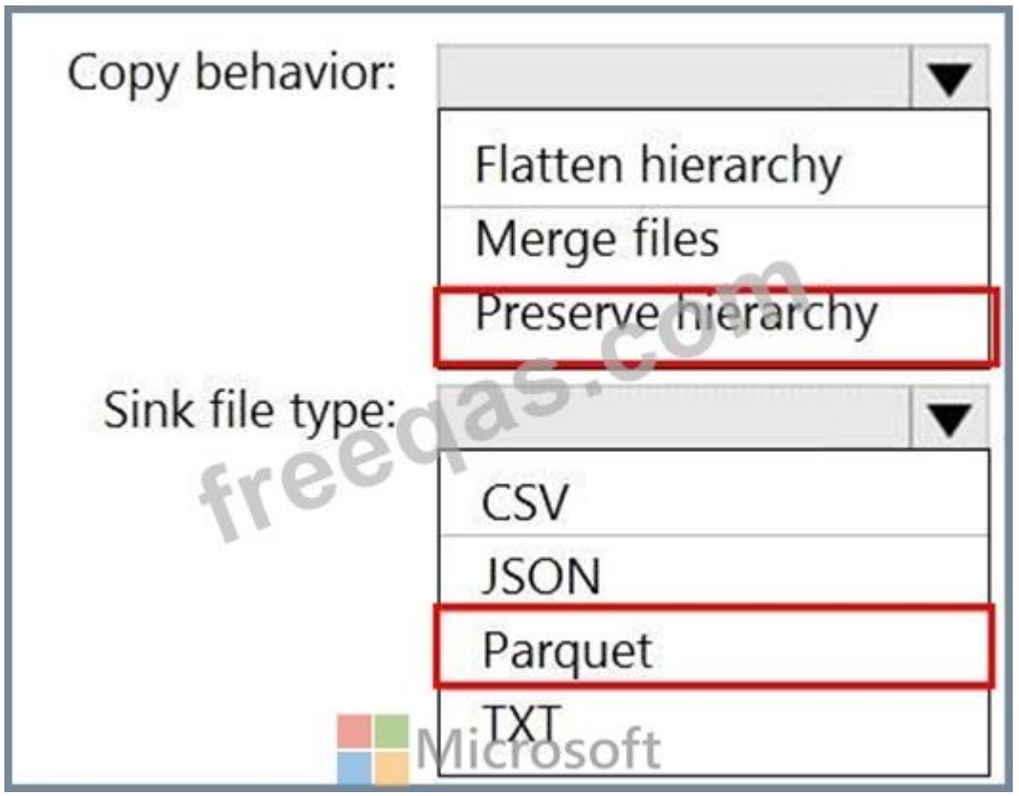
Reference:
https://docs.microsoft.com/en-us/azure/data-factory/format-parquet
https://docs.microsoft.com/en-us/azure/data-factory/connector-azure-data-lake-storage

NEW QUESTION: 38
You use Azure Data Factory to prepare data to be queried by Azure Synapse Analytics serverless SQL pools.

Files are initially ingested into an Azure Data Lake Storage Gen2 account as 10 small JSON files. Each file contains the same data attributes and data from a subsidiary of your company.
You need to move the files to a different folder and transform the data to meet the following requirements:
Provide the fastest possible query times.
Automatically infer the schema from the underlying files.
How should you configure the Data Factory copy activity? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

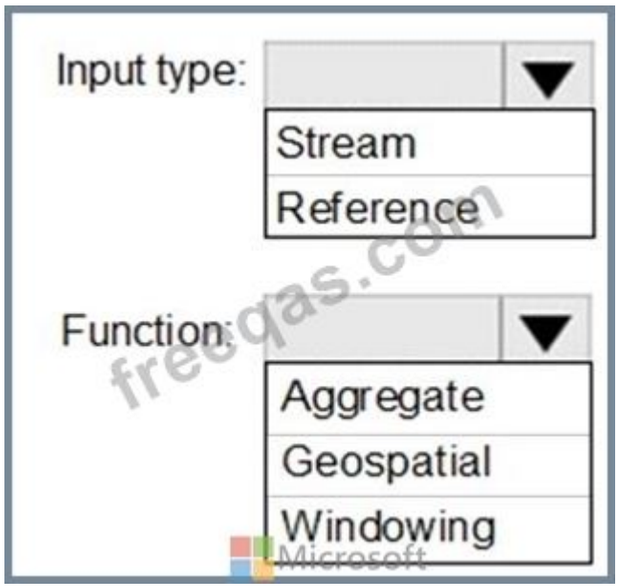


Answer:

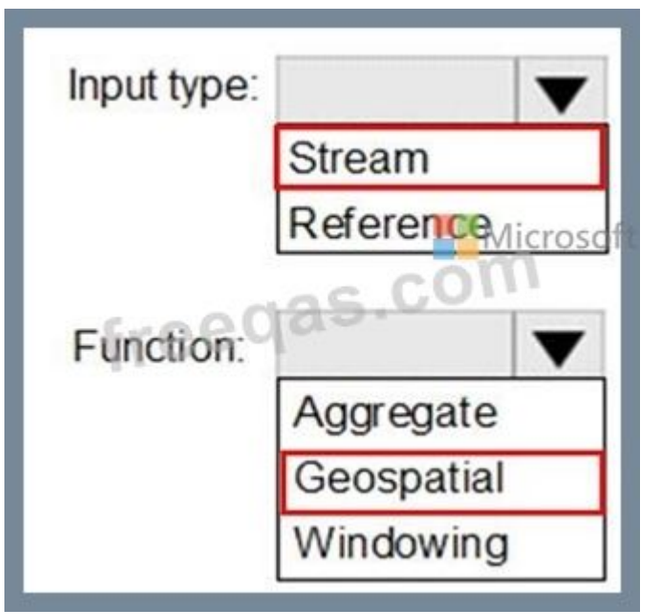


Reference:
https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-introduction
https://docs.microsoft.com/en-us/azure/data-factory/format-parquet

NEW QUESTION: 39
You plan to create a real-time monitoring app that alerts users when a device travels more than 200 meters away from a designated location.
You need to design an Azure Stream Analytics job to process the data for the planned app. The solution must minimize the amount of code developed and the number of technologies used.
What should you include in the Stream Analytics job? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.



Answer:



Reference:
https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-get-started-with-azure-stream-analytics-to-process-data-from-iot-devices
https://docs.microsoft.com/en-us/azure/stream-analytics/geospatial-scenarios

NEW QUESTION: 40
You plan to create an Azure Synapse Analytics dedicated SQL pool.
You need to minimize the time it takes to identify queries that return confidential information as defined by the company's data privacy regulations and the users who executed the queues.
Which two components should you include in the solution? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.
A. sensitivity-classification labels applied to columns that contain confidential information
B. resource tags for databases that contain confidential information

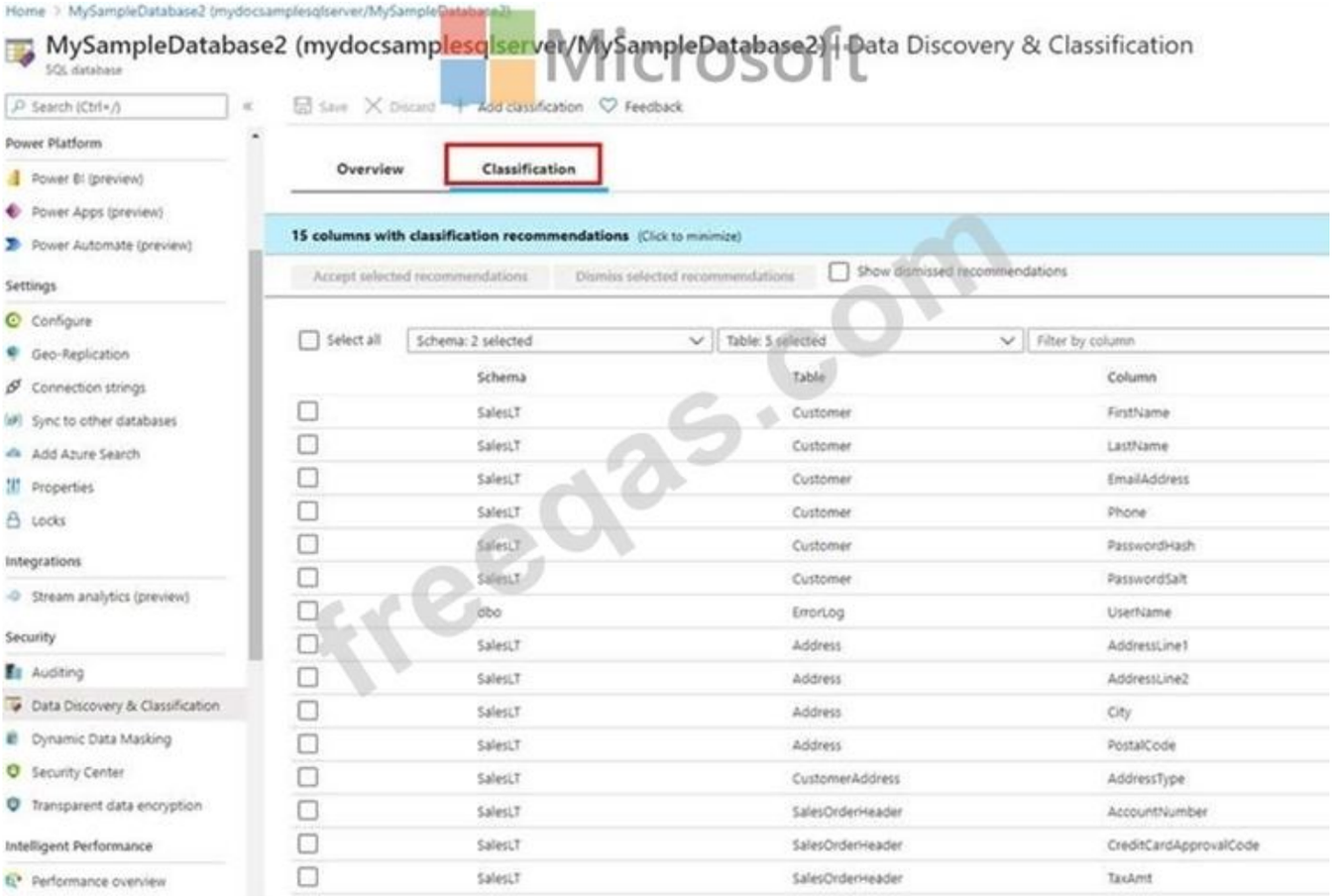
C. audit logs sent to a Log Analytics workspace

D. dynamic data masking for columns that contain confidential information

Answer: A,C (LEAVE A REPLY)

Explanation

A: You can classify columns manually, as an alternative or in addition to the recommendation-based classification:



* Select Add classification in the top menu of the pane.

* In the context window that opens, select the schema, table, and column that you want to classify, and the information type and sensitivity label.

* Select Add classification at the bottom of the context window.

C: An important aspect of the information-protection paradigm is the ability to monitor access to sensitive data. Azure SQL Auditing has been enhanced to include a new field in the audit log called data_sensitivity_information. This field logs the sensitivity classifications (labels) of the data that was returned by a query. Here's an example:

d	client_ip	application_name	duration_milliseconds	response_rows	affected_rows	connection_id	data_sensitivity_information
	[redacted]7.125	Microsoft SQL Server Management Studio - Query	1	847	847	C244A066-2271-...	Confidential - GDPR
	[redacted]7.125	Microsoft SQL Server Management Studio - Query	2	32	32	C244A066-2271-...	Confidential
	[redacted]7.125	Microsoft SQL Server Management Studio - Query	41	32	32	A7088FD4-759E-...	Confidential, Confidential - GDPR

Reference:

https://docs.microsoft.com/en-us/azure/azure-sql/database/data-discovery-and-classification-overview

NEW QUESTION: 41

You have an Azure Synapse Analytics serverless SQL pool named Pool1 and an Azure Data Lake Storage Gen2 account named storage1. The AllowedBlobpublicAccess porperty is disabled for storage1.
You need to create an external data source that can be used by Azure Active Directory (Azure AD) users to access storage1 from Pool1.
What should you create first?
A. an external resource pool
B. a remote service binding
C. database scoped credentials
D. an external library
Answer: (SHOW ANSWER)
Topic 1, Contoso Case StudyTransactional Date
Contoso has three years of customer, transactional, operation, sourcing, and supplier data comprised of 10 billion records stored across multiple on-premises Microsoft SQL Server servers. The SQL server instances contain data from various operational systems. The data is loaded into the instances by using SQL server integration Services (SSIS) packages.
You estimate that combining all product sales transactions into a company-wide sales transactions dataset will result in a single table that contains 5 billion rows, with one row per transaction.
Most queries targeting the sales transactions data will be used to identify which products were sold in retail stores and which products were sold online during different time period. Sales transaction data that is older than three years will be removed monthly.
You plan to create a retail store table that will contain the address of each retail store. The table will be approximately 2 MB. Queries for retail store sales will include the retail store addresses.
You plan to create a promotional table that will contain a promotion ID. The promotion ID will be associated to a specific product. The product will be identified by a product ID. The table will be approximately 5 GB.
Streaming Twitter Data
The ecommerce department at Contoso develops and Azure logic app that captures trending Twitter feeds referencing the company's products and pushes the products to Azure Event Hubs.
Planned Changes
Contoso plans to implement the following changes:
* Load the sales transaction dataset to Azure Synapse Analytics.
* Integrate on-premises data stores with Azure Synapse Analytics by using SSIS packages.
* Use Azure Synapse Analytics to analyze Twitter feeds to assess customer sentiments about products.
Sales Transaction Dataset Requirements
Contoso identifies the following requirements for the sales transaction dataset:
* Partition data that contains sales transaction records. Partitions must be designed to provide efficient loads by month. Boundary values must belong: to the partition on the right.
* Ensure that queries joining and filtering sales transaction records based on product ID complete as quickly as possible.
* Implement a surrogate key to account for changes to the retail store addresses.
* Ensure that data storage costs and performance are predictable.

* Minimize how long it takes to remove old records.

Customer Sentiment Analytics Requirement

Contoso identifies the following requirements for customer sentiment analytics:

* Allow Contoso users to use PolyBase in an A/ure Synapse Analytics dedicated SQL pool to query the content of the data records that host the Twitter feeds. Data must be protected by using row-level security (RLS). The users must be authenticated by using their own A/ureAD credentials.
* Maximize the throughput of ingesting Twitter feeds from Event Hubs to Azure Storage without purchasing additional throughput or capacity units.
* Store Twitter feeds in Azure Storage by using Event Hubs Capture. The feeds will be converted into Parquet files.
* Ensure that the data store supports Azure AD-based access control down to the object level.
* Minimize administrative effort to maintain the Twitter feed data records.
* Purge Twitter feed data records;itftaitJ are older than two years.

Data Integration Requirements

Contoso identifies the following requirements for data integration:

Use an Azure service that leverages the existing SSIS packages to ingest on-premises data into datasets stored in a dedicated SQL pool of Azure Synaps Analytics and transform the data.

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

NEW QUESTION: 42

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

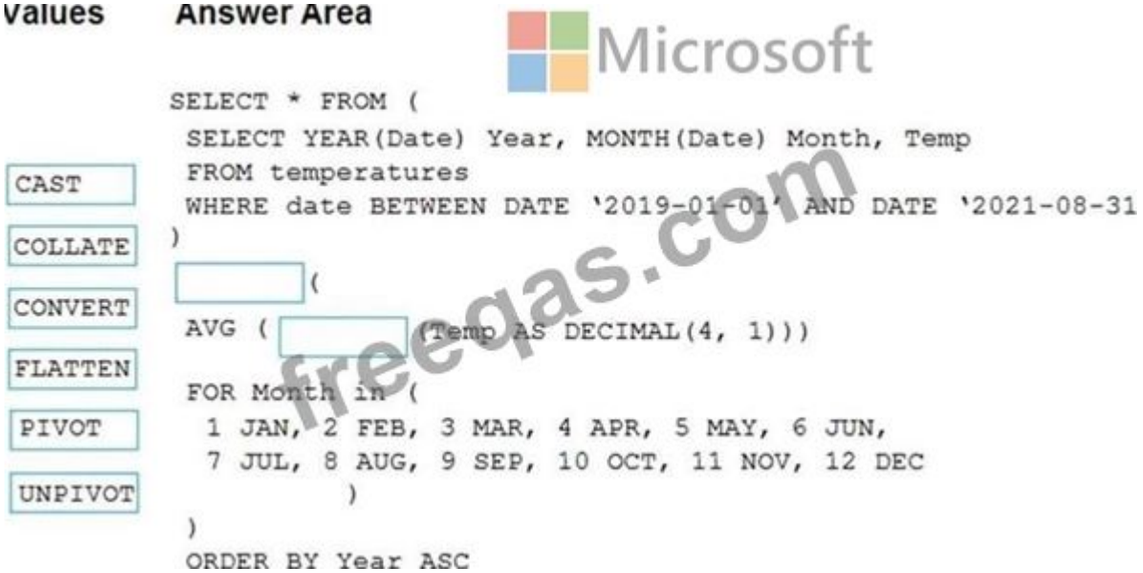
Date	Temp
...	...
18-01-2021	3
19-01-2021	4
20-01-2021	2
21-01-2021	2
...	...

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

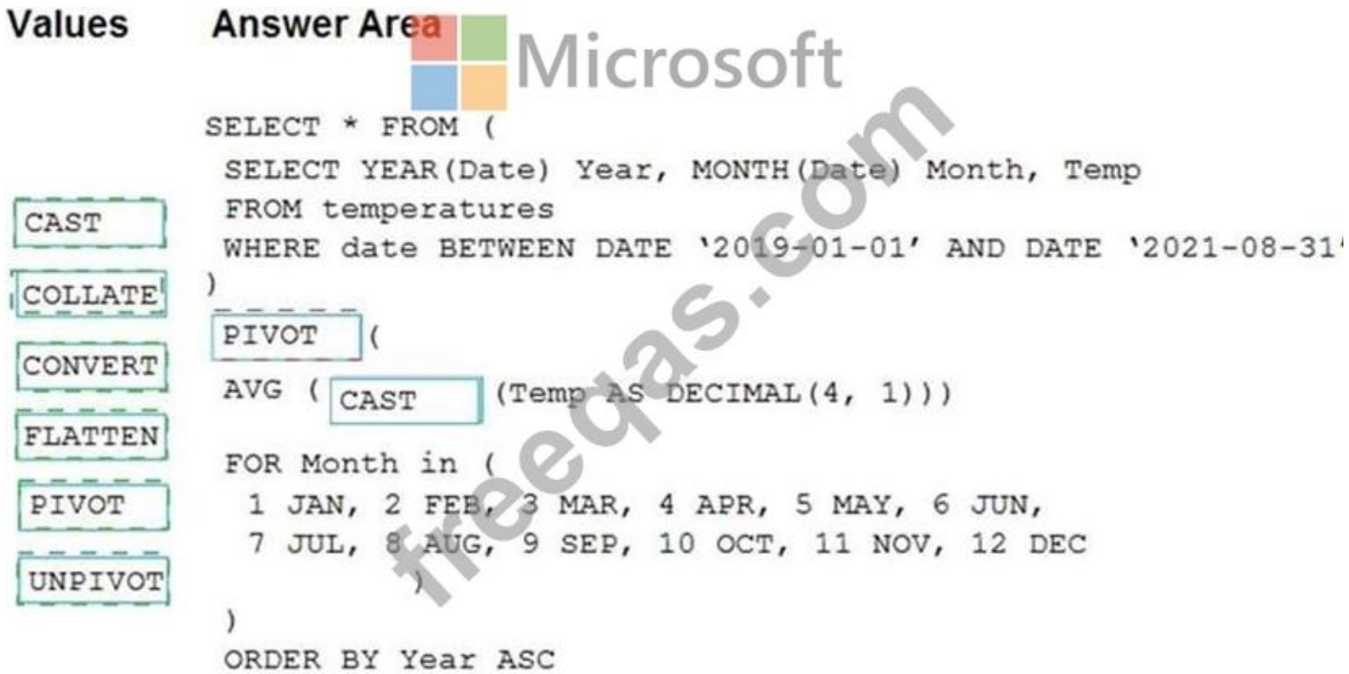
Year	JAN	FEB	MAR	APR	MAY
2019	2.3	4.1	5.2	7.6	9.2
2020	2.4	4.2	4.9	7.8	9.1
2021	2.6	5.3	3.4	7.9	9.5

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.



Answer:



Explanation

Text Description automatically generated

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: 43

You are designing a dimension table for a data warehouse. The table will track the value of the dimension attributes over time and preserve the history of the data by adding new rows as the data changes.

Which type of slowly changing dimension (SCD) should use?

A. Type 0

B. Type 1

C. Type 2

D. Type 3

Answer: C (LEAVE A REPLY)

Explanation

Type 2 - Creating a new additional record. In this methodology all history of dimension changes is kept in the database. You capture attribute change by adding a new row with a new surrogate key to the dimension table. Both the prior and new rows contain as attributes the natural key(or other durable identifier). Also

'effective date' and 'current indicator' columns are used in this method. There could be only one record with current indicator set to 'Y'. For 'effective date' columns, i.e. start_date and end_date, the end_date for current record usually is set to value 9999-12-31. Introducing changes to the dimensional model in type 2

could be very expensive database operation so it is not recommended to use it in dimensions where a new attribute could be added in the future.
https://www.datawarehouse4u.info/SCD-Slowly-Changing-Dimensions.html

NEW QUESTION: 44
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.

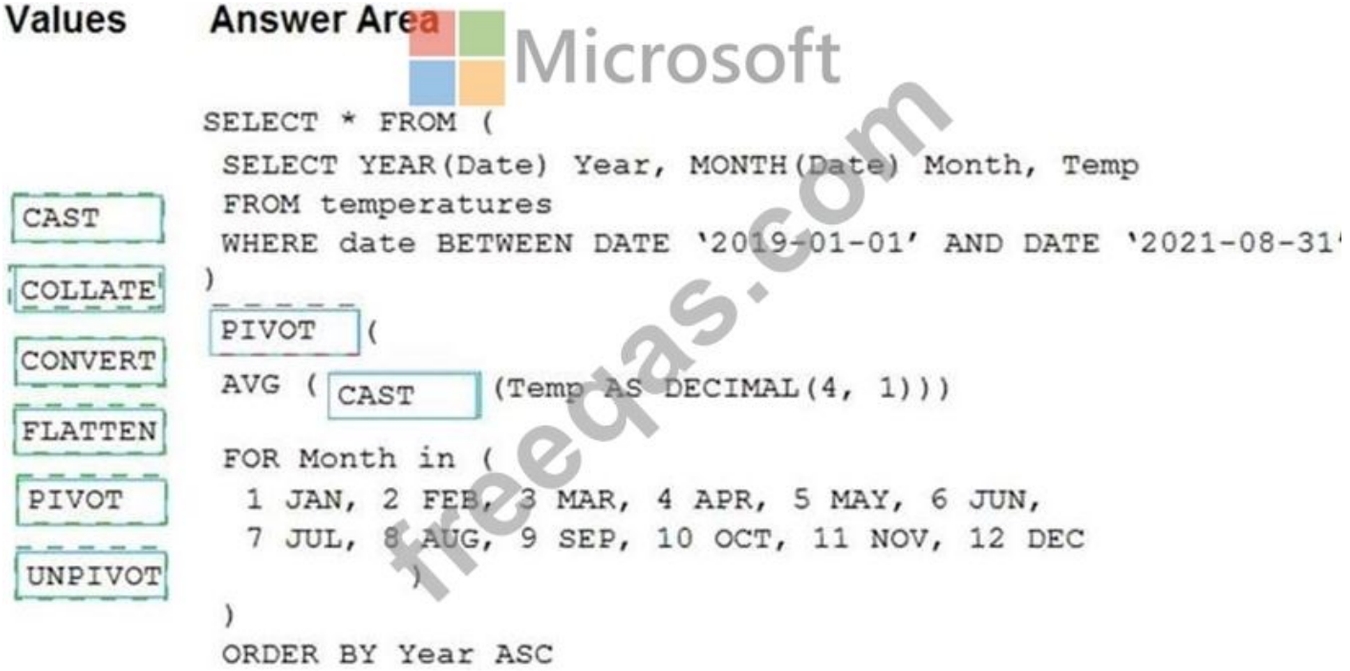
Values

CAST
COLLATE
CONVERT
FLATTEN
PIVOT
UNPIVOT

Answer Area

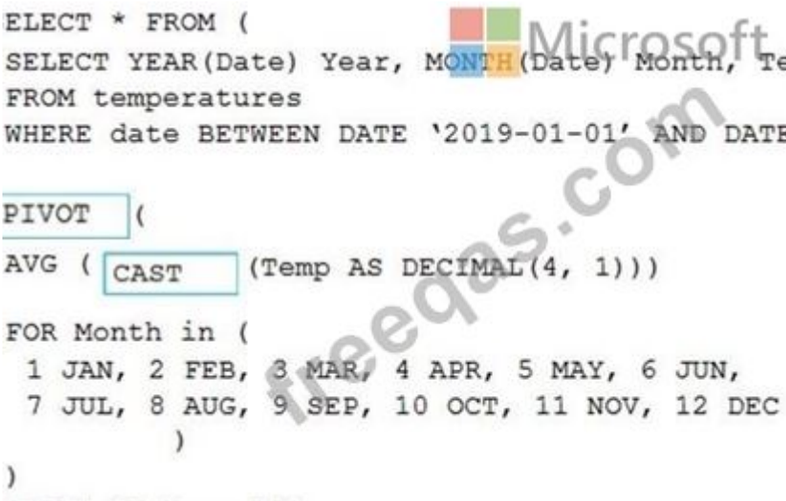
```
SELECT * FROM (
 SELECT YEAR(Date) Year, MONTH(Date) Month, Temp
 FROM temperatures
 WHERE date BETWEEN DATE '2019-01-01' AND DATE '2021-08-31
)
[          ](
 AVG ( [          ](Temp AS DECIMAL(4, 1)))
 FOR Month in (
  1 JAN, 2 FEB, 3 MAR, 4 APR, 5 MAY, 6 JUN,
  7 JUL, 8 AUG, 9 SEP, 10 OCT, 11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

Answer:



Explanation

Text Description automatically generated



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/

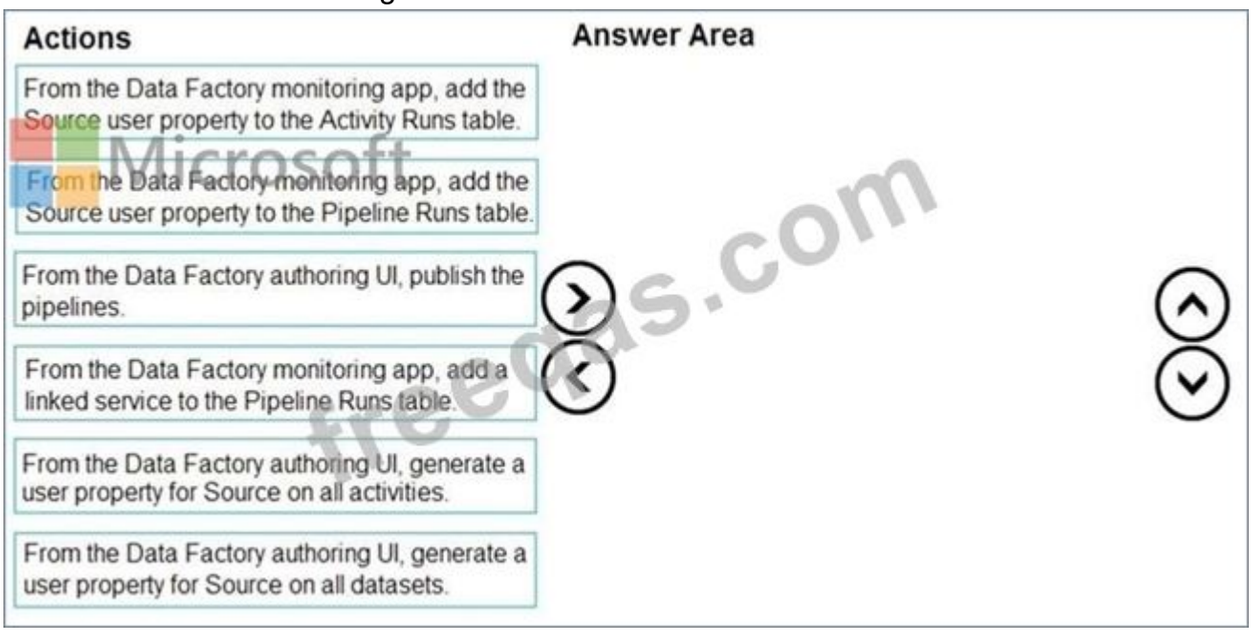
https://docs.microsoft.com/en-us/sql/t-sql/queries/from-using-pivot-and-unpivot

NEW QUESTION: 45

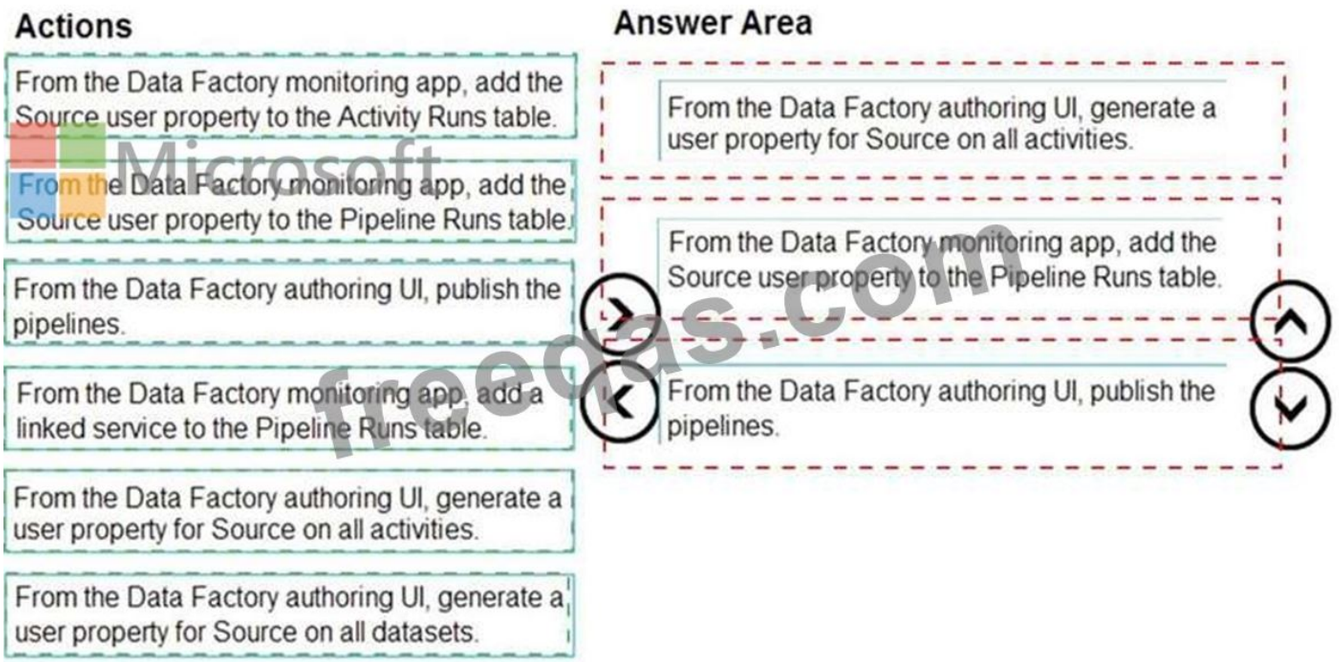
You plan to monitor an Azure data factory by using the Monitor & Manage app.

You need to identify the status and duration of activities that reference a table in a source database.

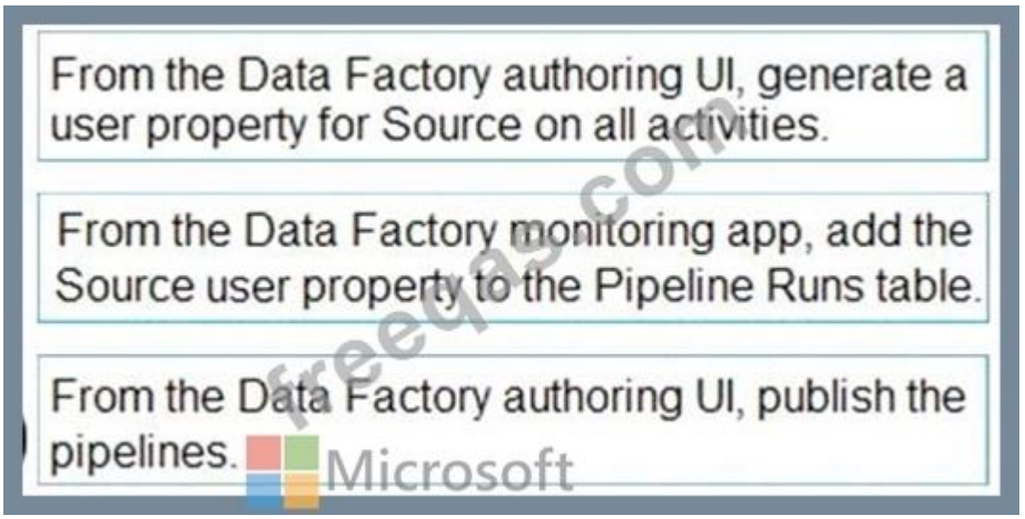
Which three actions should you perform in sequence? To answer, move the actions from the list of actions to the answer are and arrange them in the correct order.



Answer:



Explanation



Step 1: From the Data Factory authoring UI, generate a user property for Source on all activities.
Step 2: From the Data Factory monitoring app, add the Source user property to Activity Runs table.
You can promote any pipeline activity property as a user property so that it becomes an entity that you can monitor. For example, you can promote the Source and Destination properties of the copy activity in your pipeline as user properties. You can also select Auto Generate to generate the Source and Destination user properties for a copy activity.
Step 3: From the Data Factory authoring UI, publish the pipelines
Publish output data to data stores such as Azure SQL Data Warehouse for business intelligence (BI) applications to consume.
References:
https://docs.microsoft.com/en-us/azure/data-factory/monitor-visually

NEW QUESTION: 46
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)
Explanation
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: 47

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 maximizes query performance.

What should you include in the recommendation?

A. In the tables use a hash distribution of ArrivalDateTime and ReportDateTime.

B. In the tables use a hash distribution of ArrivalAirportID and AirportID.

C. In each table, create an identity column.

D. In each table, create a column as a composite of the other two columns in the table.

Answer: B (LEAVE A REPLY)

Hash-distribution improves query performance on large fact tables.

Incorrect Answers:

A: Do not use a date column for hash distribution. All data for the same date lands in the same distribution. If several users are all filtering on the same date, then only 1 of the 60 distributions do all the processing work.

NEW QUESTION: 48

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)
Explanation
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: 49
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: A (LEAVE A REPLY)
Explanation
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: 50
You configure monitoring for a Microsoft Azure SQL Data Warehouse implementation. The implementation uses PolyBase to load data from comma-separated value (CSV) files stored in Azure Data Lake Gen 2 using an external table.
Files with an invalid schema cause errors to occur.
You need to monitor for an invalid schema error.

For which error should you monitor?
A. EXTERNAL TABLE access failed due to internal error: 'Java exception raised on call to HdfsBridge_Connect: Error [com.microsoft.polybase.client.KerberosSecureLogin] occurred while accessing external files.'
B. EXTERNAL TABLE access failed due to internal error: 'Java exception raised on call to HdfsBridge_Connect: Error [No FileSystem for scheme: wasbs] occurred while accessing external file.'
C. Cannot execute the query "Remote Query" against OLE DB provider "SQLNCLI11": for linked server "(null)", Query aborted- the maximum reject threshold (o
rows) was reached while regarding from an external source: 1 rows rejected out of total 1 rows processed.
D. EXTERNAL TABLE access failed due to internal error: 'Java exception raised on call to HdfsBridge_Connect: Error [Unable to instantiate LoginClass] occurred while accessing external files.'
Answer: C (LEAVE A REPLY)
Explanation
Customer Scenario:
SQL Server 2016 or SQL DW connected to Azure blob storage. The CREATE EXTERNAL TABLE DDL points to a directory (and not a specific file) and the directory contains files with different schemas.
SSMS Error:
Select query on the external table gives the following error:
Msg 7320, Level 16, State 110, Line 14
Cannot execute the query "Remote Query" against OLE DB provider "SQLNCLI11" for linked server "(null)". Query aborted-- the maximum reject threshold (0 rows) was reached while reading from an external source: 1 rows rejected out of total 1 rows processed.
Possible Reason:
The reason this error happens is because each file has different schema. The PolyBase external table DDL when pointed to a directory recursively reads all the files in that directory. When a column or data type mismatch happens, this error could be seen in SSMS.
Possible Solution:
If the data for each table consists of one file, then use the filename in the LOCATION section prepended by the directory of the external files. If there are multiple files per table, put each set of files into different directories in Azure Blob Storage and then you can point LOCATION to the directory instead of a particular file. The latter suggestion is the best practices recommended by SQLCAT even if you have one file per table.

NEW QUESTION: 51
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.

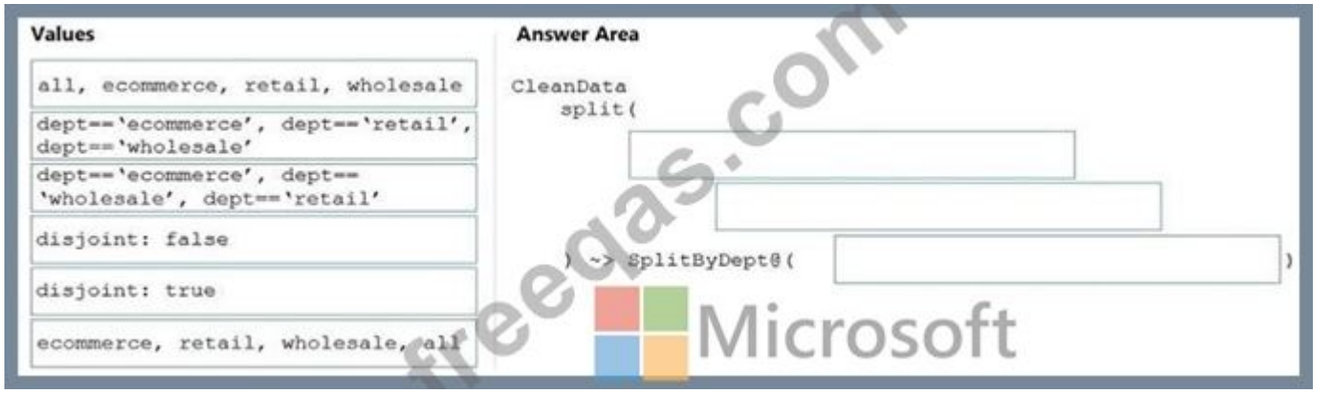
A. Azure Databricks
B. Azure Machine Learning
C. Azure Synapse Analytics
D. Azure Data Factory
E. Azure HDInsight
Answer: B,C (LEAVE A REPLY)

NEW QUESTION: 52
You need to design a data retention solution for the Twitter teed data records. The solution must meet the customer sentiment analytics requirements.
Which Azure Storage functionality should you include in the solution?
A. time-based retention
B. soft delete
C. change feed
D. lifecycle management
Answer: (SHOW ANSWER)

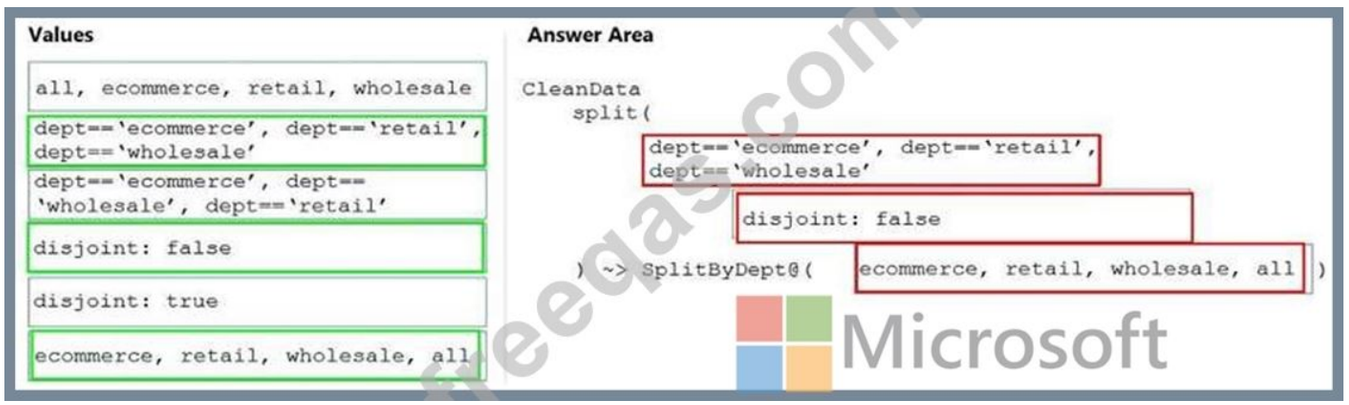
NEW QUESTION: 53
You manage an enterprise data warehouse in Azure Synapse Analytics.
Users report slow performance when they run commonly used queries. Users do not report performance changes for infrequently used queries.
You need to monitor resource utilization to determine the source of the performance issues.
Which metric should you monitor?
A. Data IO percentage
B. Local tempdb percentage
C. Cache used percentage
D. DWU percentage
Answer: C (LEAVE A REPLY)
Explanation
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-monito

NEW QUESTION: 54
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.



Answer:



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

NEW QUESTION: 55
The storage account container view is shown in the Refdata exhibit. (Click the Refdata tab.) You need to configure the Stream Analytics job to pick up the new reference data. What should you configure? To answer, select the appropriate options in the answer area NOTE: Each correct selection is worth one point.

Answer:
See the answer below in explanation.
Explanation
answer as below

NEW QUESTION: 56

You are designing an Azure Data Lake Storage Gen2 container to store data for the human resources (HR) department and the operations department at your company. You have the following data access requirements:

* After initial processing, the HR department data will be retained for seven years.
* The operations department data will be accessed frequently for the first six months, and then accessed once per month.

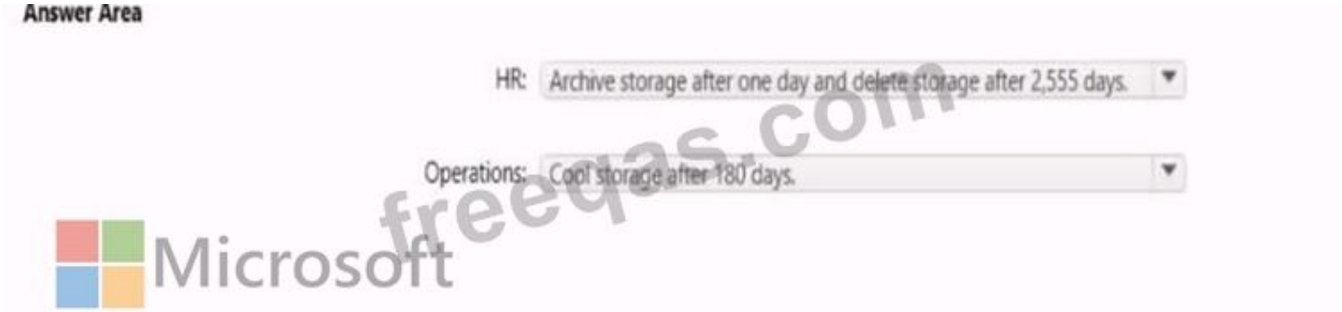
You need to design a data retention solution to meet the access requirements. The solution must minimize storage costs.

Answer:

See the answer in explanation.

Explanation

answer is below



NEW QUESTION: 57

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: A (LEAVE A REPLY)

Explanation

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: 58

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. last modified date

B. surrogate primary key

C. effective end date

D. foreign key

E. business key

F. effective start date

Answer: D,E,F (LEAVE A REPLY)

NEW QUESTION: 59

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:

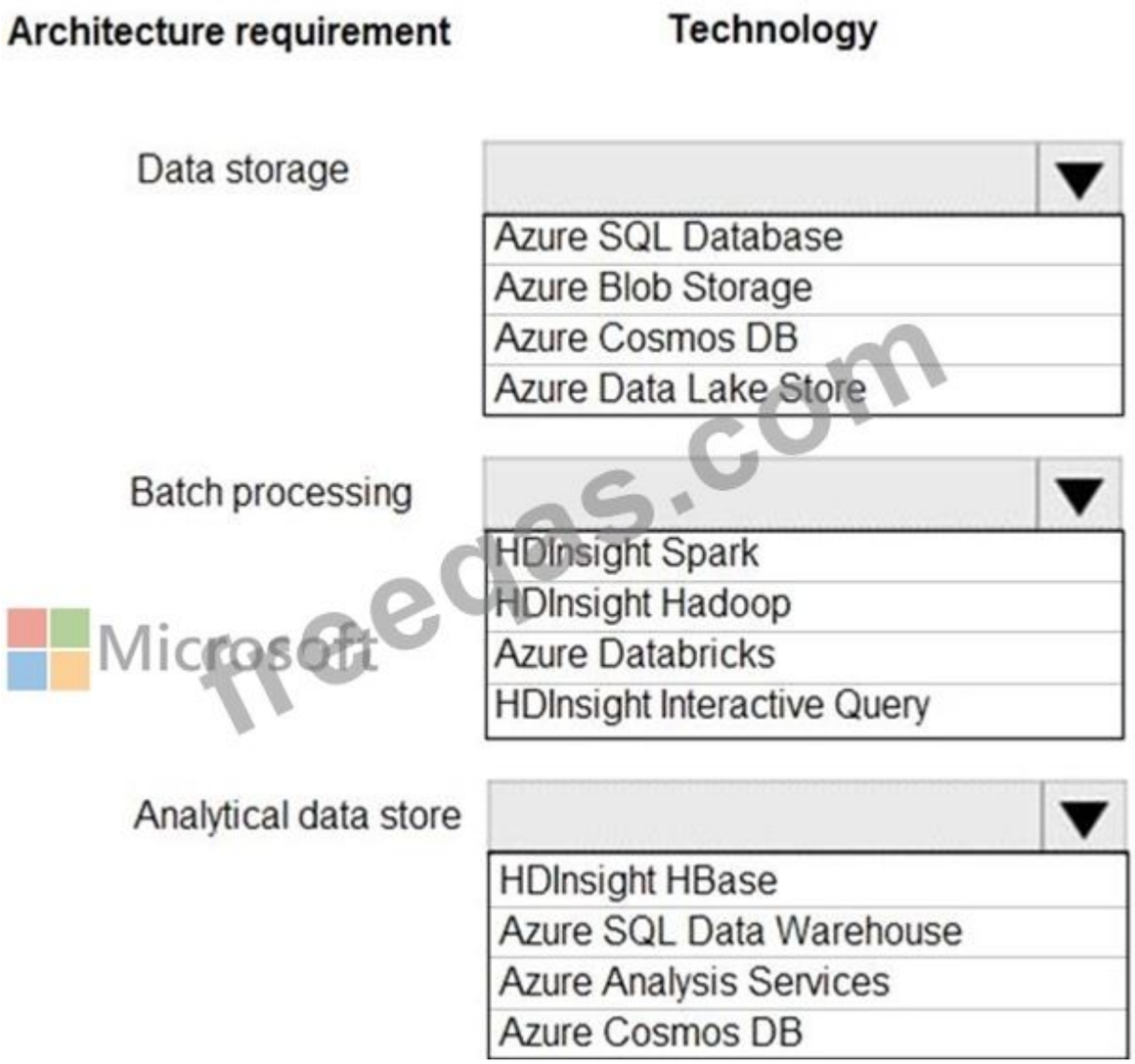
* Use a managed solution for in-memory computation processing.
* Natively support Scala, Python, and R programming languages.
* Provide the ability to resize and terminate the cluster automatically.

Analytical data store:

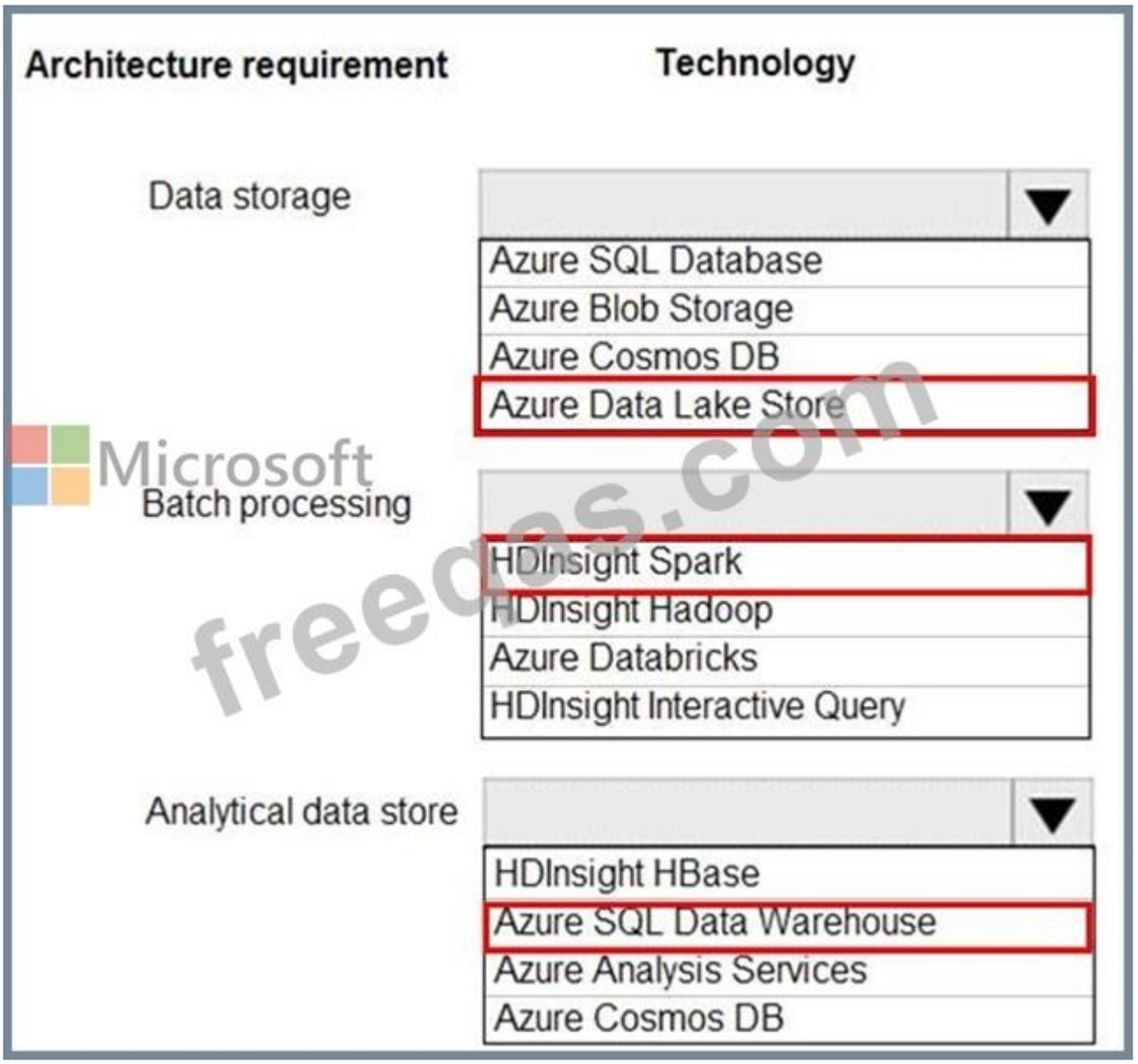
* Support parallel processing.
* Use columnar storage.
* Support SQL-based languages.

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.



Answer:



Reference:
https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-namespace
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

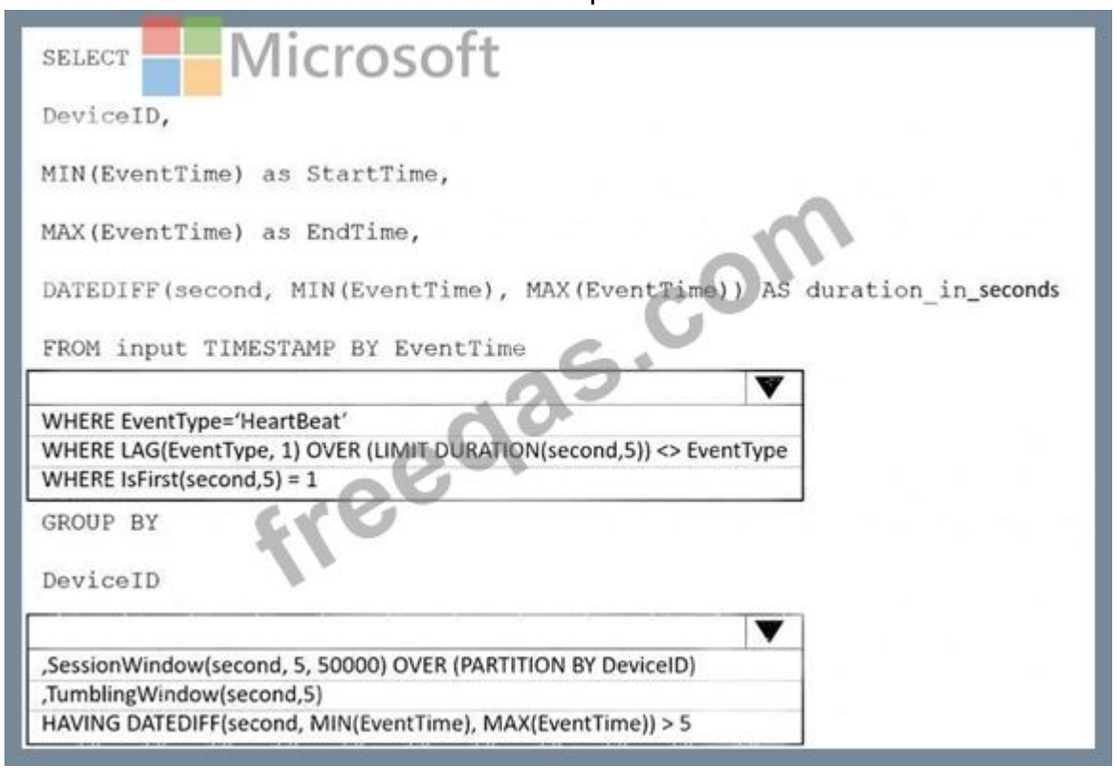
NEW QUESTION: 60
You are implementing an Azure Stream Analytics solution to process event data from devices.
The devices output events when there is a fault and emit a repeat of the event every five seconds until the fault is resolved. The devices output a heartbeat event every five seconds after a previous 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

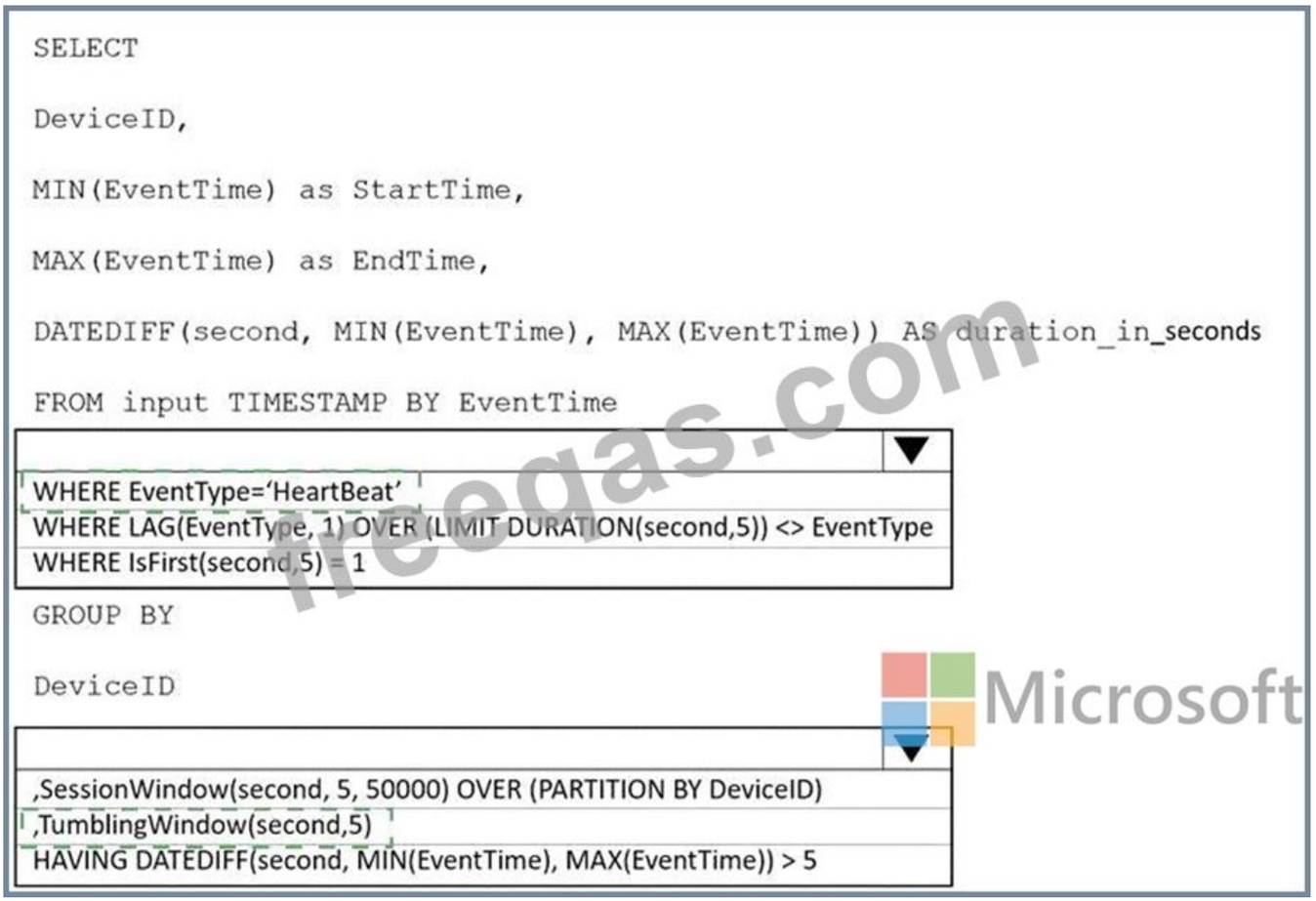
You need to calculate the uptime between the faults.

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.

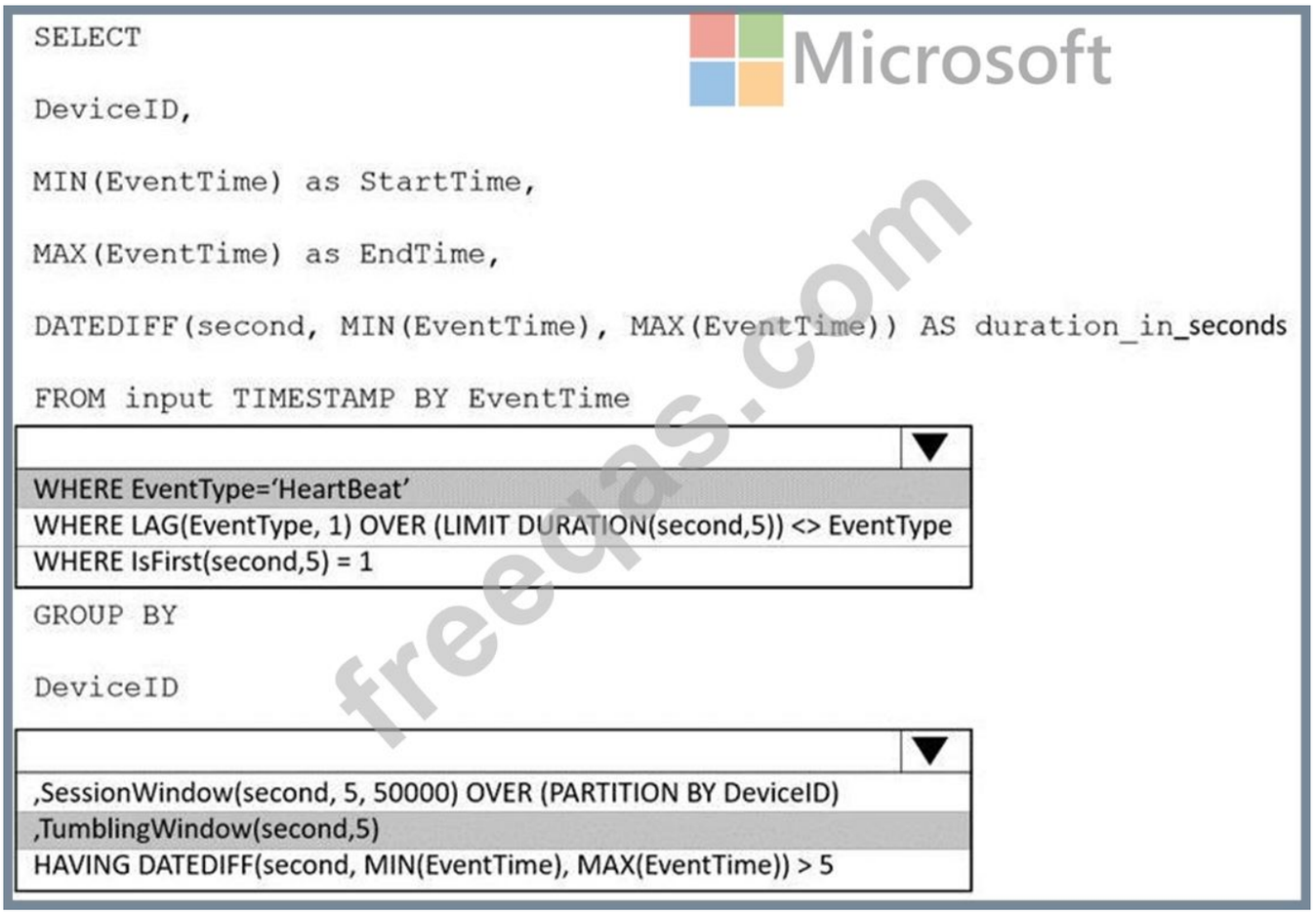


Answer:



Explanation

Graphical user interface, text, application Description automatically generated



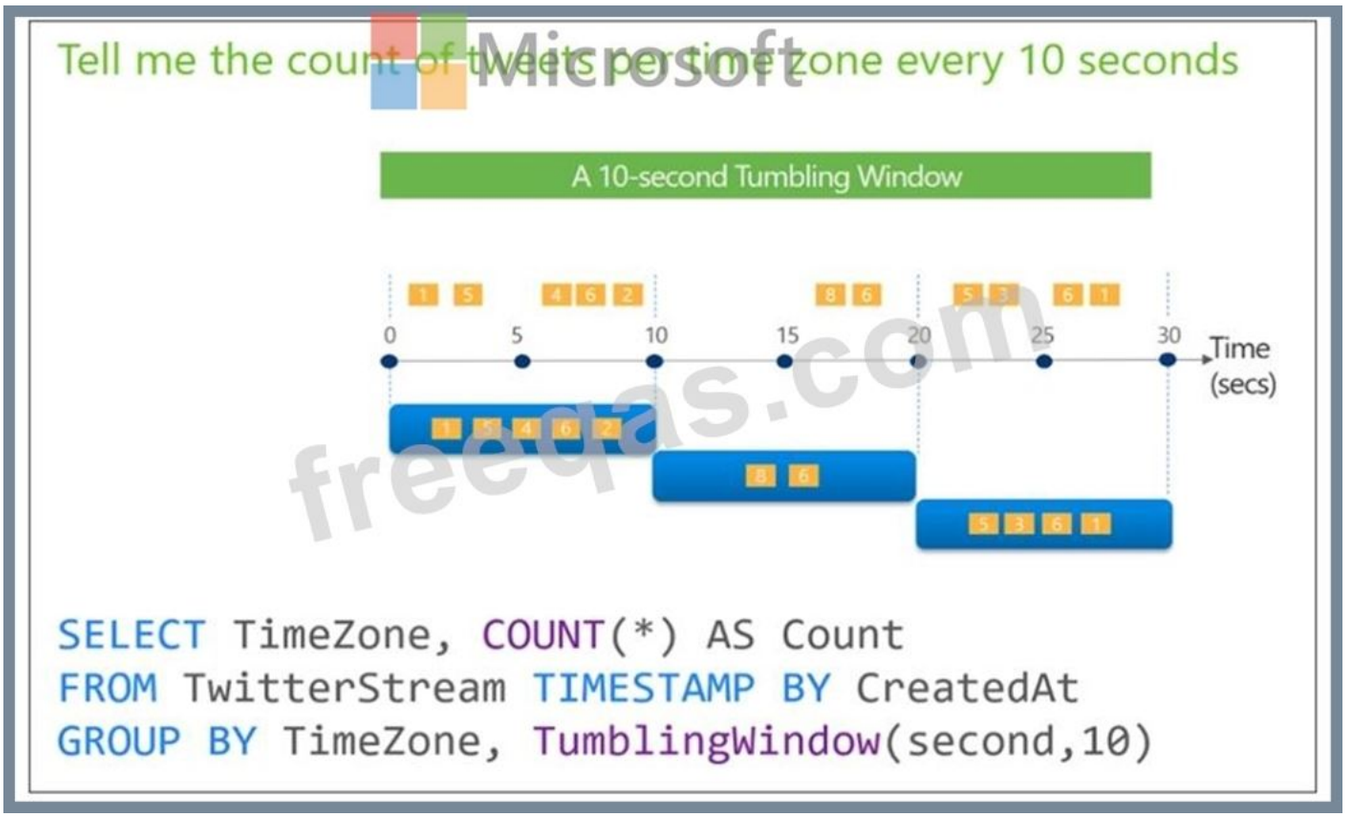
Box 1: WHERE EventType='HeartBeat'

Box 2: ,TumblingWindow(Second, 5)

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

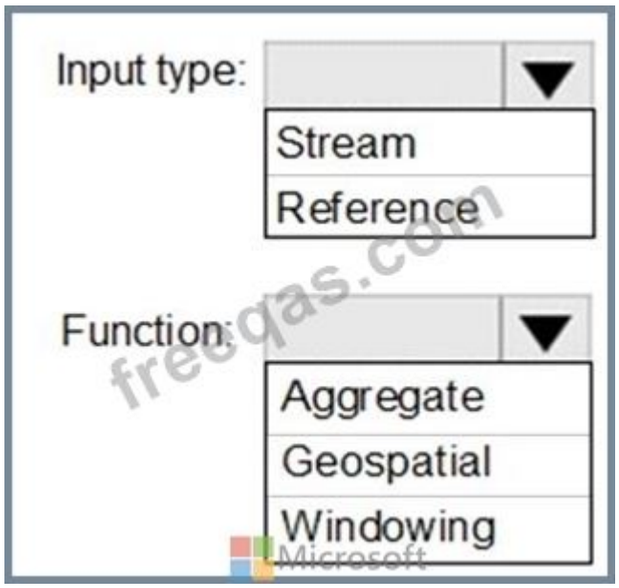
The following diagram illustrates a stream with a series of events and how they are mapped into 10-second tumbling windows.

Timeline Description automatically generated

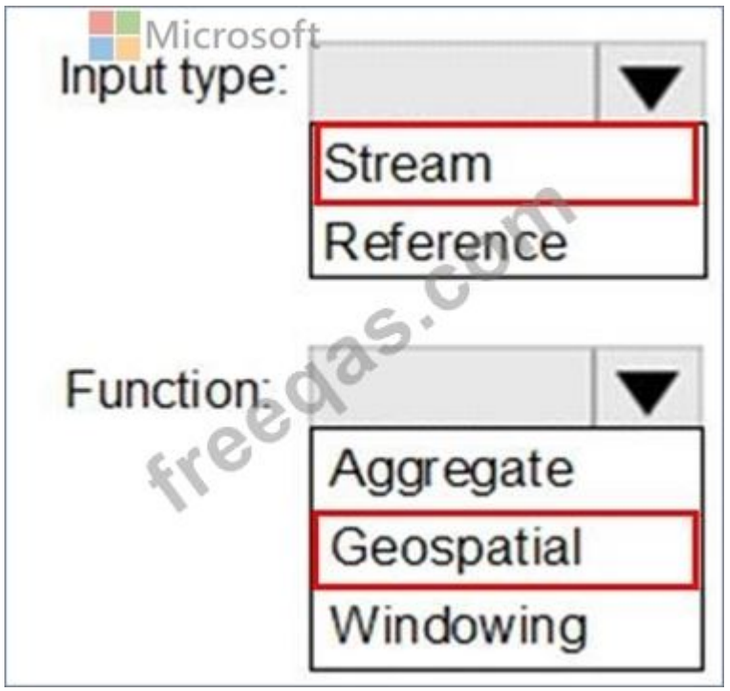


Reference:
https://docs.microsoft.com/en-us/stream-analytics-query/session-window-azure-stream-analytics
https://docs.microsoft.com/en-us/stream-analytics-query/tumbling-window-azure-stream-analytics

NEW QUESTION: 61
You plan to create a real-time monitoring app that alerts users when a device travels more than 200 meters away from a designated location.
You need to design an Azure Stream Analytics job to process the data for the planned app. The solution must minimize the amount of code developed and the number of technologies used.
What should you include in the Stream Analytics job? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.



Answer:



Reference:
https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-get-started-with-azure-stream-analytics-to-process-data-from-iot-devices
https://docs.microsoft.com/en-us/azure/stream-analytics/geospatial-scenarios

NEW QUESTION: 62
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-d

NEW QUESTION: 63
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. event
Answer: D (LEAVE A REPLY)
Explanation
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: 64
You need to design a data retention solution for the Twitter teed data records. The solution must meet the customer sentiment analytics requirements.
Which Azure Storage functionality should you include in the solution?
A. time-based retention
B. change feed
C. soft delete

D. lifecycle management

Answer: D (LEAVE A REPLY)

Topic 1, Contoso

Transactional Date

Contoso has three years of customer, transactional, operation, sourcing, and supplier data comprised of 10 billion records stored across multiple on-premises Microsoft SQL Server servers. The SQL server instances contain data from various operational systems. The data is loaded into the instances by using SQL server integration Services (SSIS) packages.

You estimate that combining all product sales transactions into a company-wide sales transactions dataset will result in a single table that contains 5 billion rows, with one row per transaction.

Most queries targeting the sales transactions data will be used to identify which products were sold in retail stores and which products were sold online during different time period. Sales transaction data that is older than three years will be removed monthly.

You plan to create a retail store table that will contain the address of each retail store. The table will be approximately 2 MB. Queries for retail store sales will include the retail store addresses.

You plan to create a promotional table that will contain a promotion ID. The promotion ID will be associated to a specific product. The product will be identified by a product ID. The table will be approximately 5 GB.

Streaming Twitter Data

The ecommerce department at Contoso develops and Azure logic app that captures trending Twitter feeds referencing the company's products and pushes the products to Azure Event Hubs.

Planned Changes

Contoso plans to implement the following changes:

* Load the sales transaction dataset to Azure Synapse Analytics.
* Integrate on-premises data stores with Azure Synapse Analytics by using SSIS packages.
* Use Azure Synapse Analytics to analyze Twitter feeds to assess customer sentiments about products.

Sales Transaction Dataset Requirements

Contoso identifies the following requirements for the sales transaction dataset:

* Partition data that contains sales transaction records. Partitions must be designed to provide efficient loads by month. Boundary values must belong: to the partition on the right.
* Ensure that queries joining and filtering sales transaction records based on product ID complete as quickly as possible.
* Implement a surrogate key to account for changes to the retail store addresses.
* Ensure that data storage costs and performance are predictable.
* Minimize how long it takes to remove old records.

Customer Sentiment Analytics Requirement

Contoso identifies the following requirements for customer sentiment analytics:

* Allow Contoso users to use PolyBase in an A/ure Synapse Analytics dedicated SQL pool to query the content of the data records that host the Twitter feeds. Data must be protected by using row-level security (RLS). The users must be authenticated by using their own A/ureAD credentials.

* Maximize the throughput of ingesting Twitter feeds from Event Hubs to Azure Storage without purchasing additional throughput or capacity units.
* Store Twitter feeds in Azure Storage by using Event Hubs Capture. The feeds will be converted into Parquet files.
* Ensure that the data store supports Azure AD-based access control down to the object level.
* Minimize administrative effort to maintain the Twitter feed data records.
* Purge Twitter feed data records;itftaitJ are older than two years.

Data Integration Requirements

Contoso identifies the following requirements for data integration:

Use an Azure service that leverages the existing SSIS packages to ingest on-premises data into datasets stored in a dedicated SQL pool of Azure Synaps Analytics and transform the data.

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

NEW QUESTION: 65

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: (SHOW ANSWER)

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: 66

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 dat

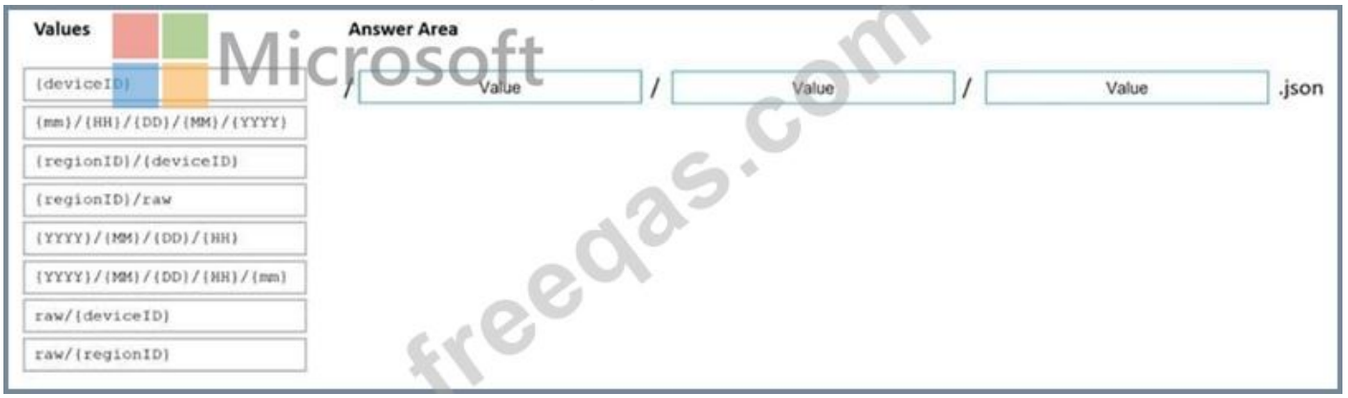
a. The solution must meet the following requirements:

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.



Answer:

Reference:

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

NEW QUESTION: 67

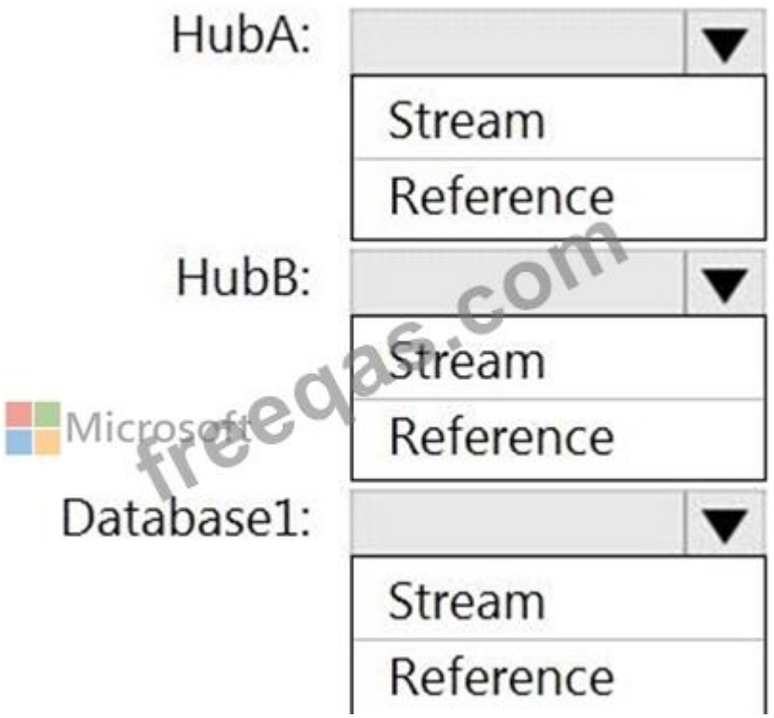
You have an Azure SQL database named Database1 and two Azure event hubs named HubA and HubB. The data consumed from each source is shown in the following table.

Source	Data
Database1	Driver's name Driver's license number
HubA	Ride route Ride distance Ride duration
HubB	Ride fare Ride payment

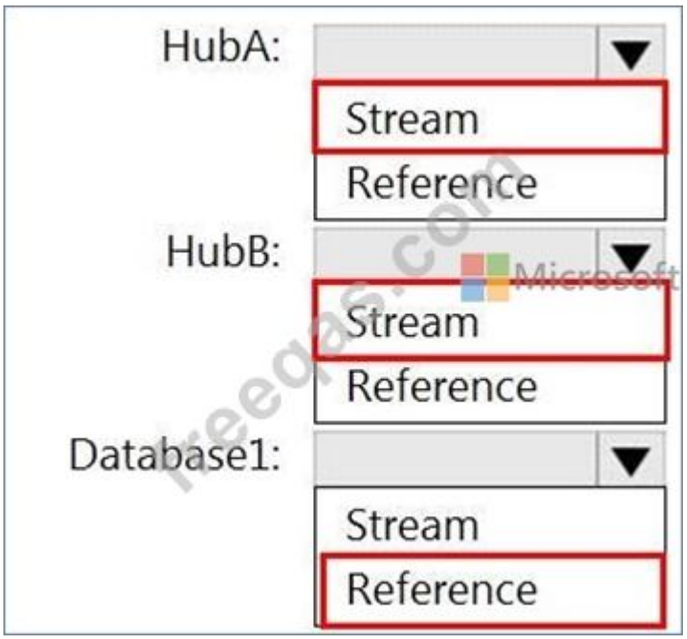
You need to implement Azure Stream Analytics to calculate the average fare per mile by driver.

How should you configure the Stream Analytics input for each source? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.



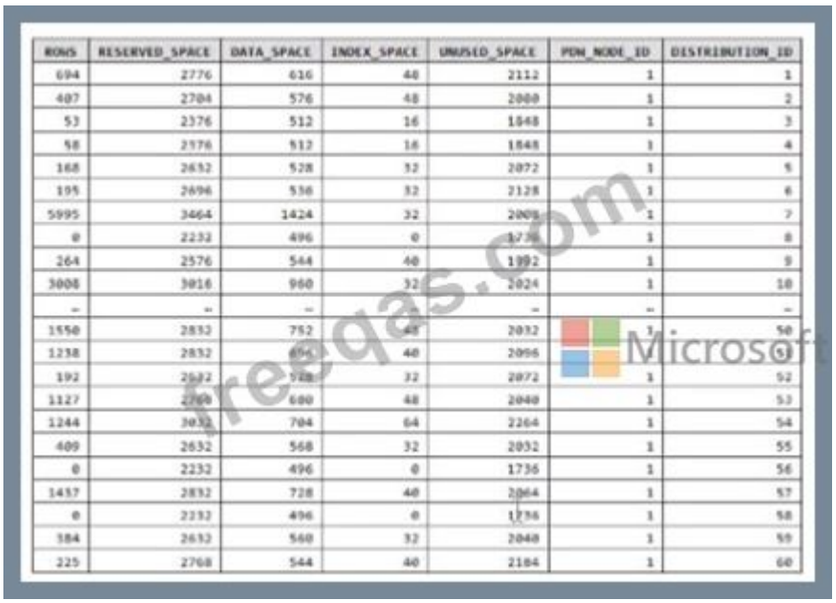
Answer:



Reference:
https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-use-reference-data

NEW QUESTION: 68
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: (SHOW ANSWER)
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: 69
You have an Azure Synapse Analytics dedicated SQL pool.
You run PDW_SHOWSPACEUSED(dbo,FactInternetSales'); and get the results shown in the following table.



ROWS	RESERVED_SPACE	DATA_SPACE	INDEX_SPACE	UNUSED_SPACE	PDW_NODE_ID	DISTRIBUTION_ID
694	2776	616	48	2112	1	1
407	2704	576	48	2080	1	2
53	2376	512	16	1848	1	3
58	2376	512	16	1848	1	4
168	2632	528	32	2072	1	5
195	2696	536	32	2128	1	6
5995	3464	1424	32	2008	1	7
0	2232	496	0	1736	1	8
264	2576	544	40	1992	1	9
3008	3016	960	32	2024	1	10
..	..	..	..	..	..	..
1550	2832	752	48	2032	1	50
1238	2832	656	40	2096	1	51
192	2632	528	32	2072	1	52
1127	2760	680	48	2040	1	53
1244	3032	704	64	2264	1	54
409	2632	568	32	2032	1	55
0	2232	496	0	1736	1	56
1457	2832	728	40	2064	1	57
0	2232	496	0	1736	1	58
384	2632	560	32	2040	1	59
225	2768	544	40	2184	1	60

Which statement accurately describes the dbo,FactInternetSales table?

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

B. All distribution contain data.

C. The table is skewed.

D. The table uses round-robin distribution.

Answer: (SHOW ANSWER)

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

Reference:

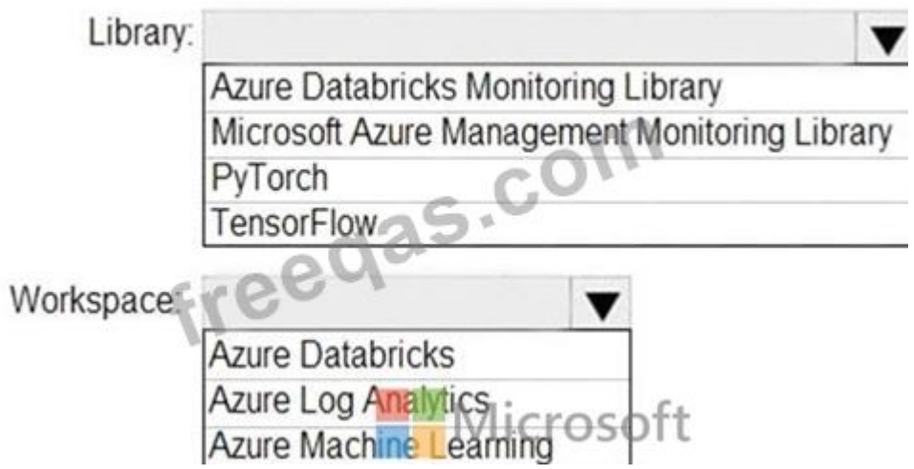
https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-distribute

NEW QUESTION: 70

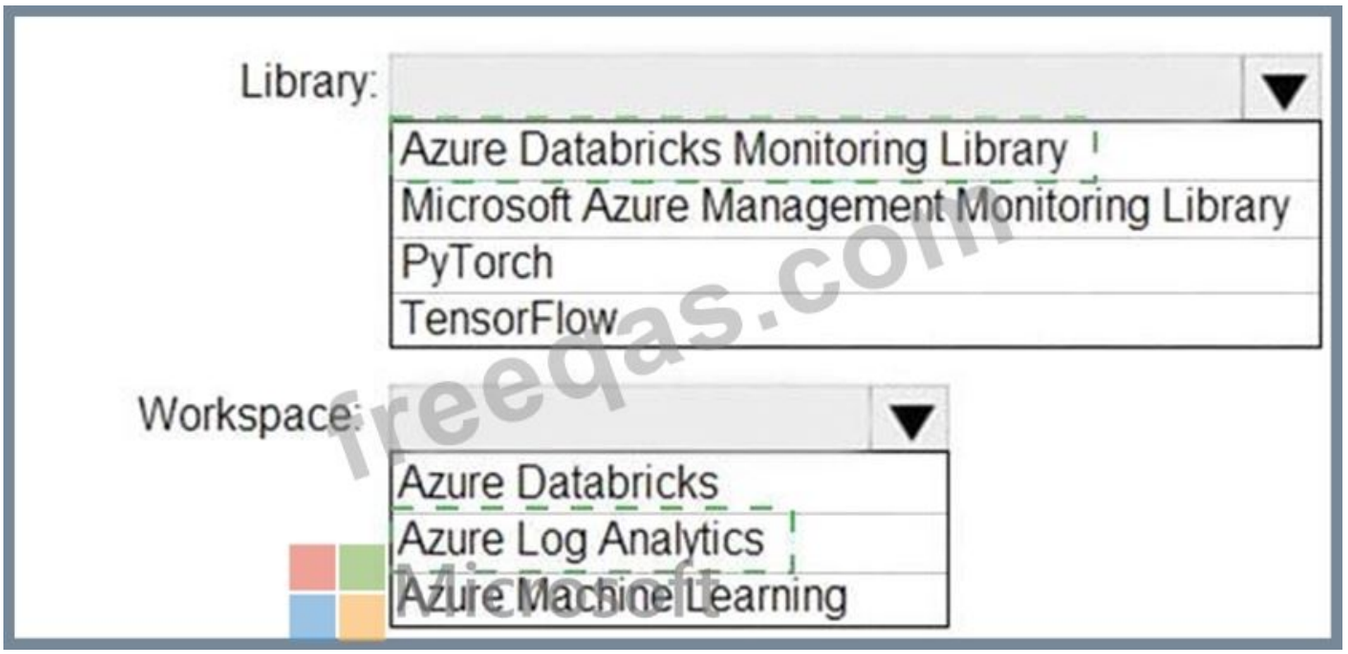
You need to collect application metrics, streaming query events, and application log messages for an Azure Databrick cluster.

Which type of library and workspace should you implement? To answer, select the appropriate options in the answer area.

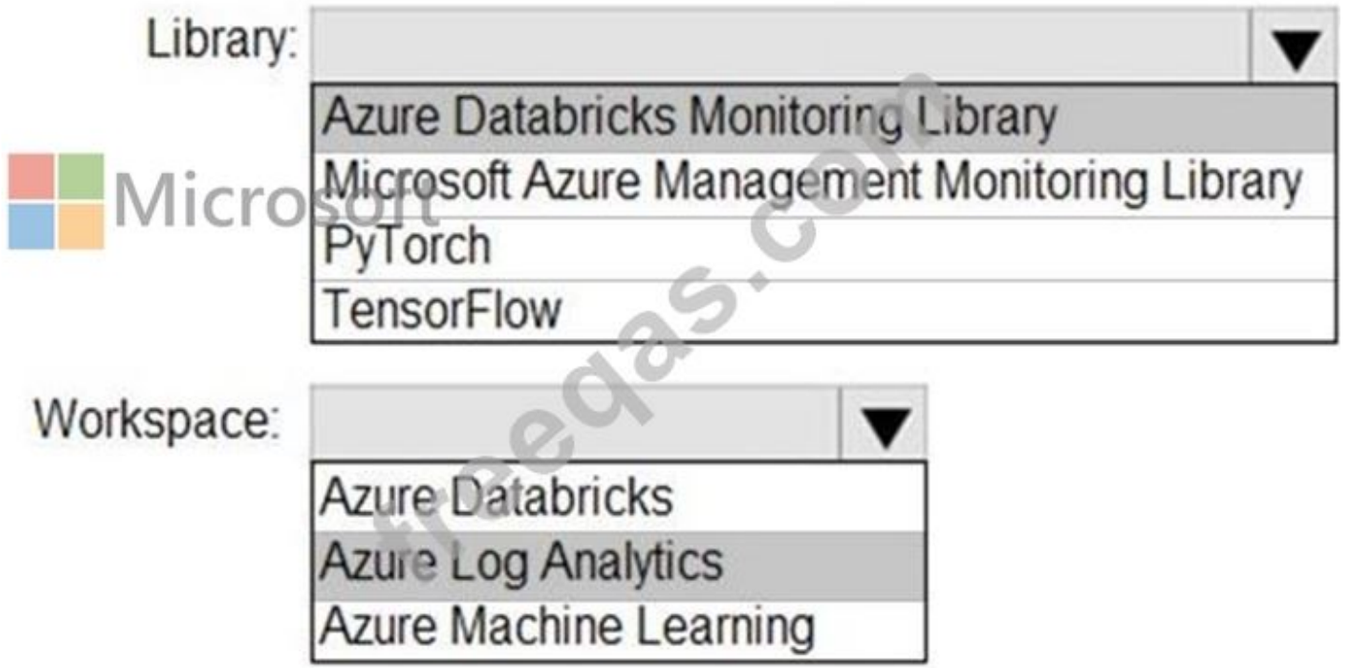
NOTE: Each correct selection is worth one point.



Answer:



Explanation



You can send application logs and metrics from Azure Databricks to a Log Analytics workspace. It uses the Azure Databricks Monitoring Library, which is available on GitHub.

References:

https://docs.microsoft.com/en-us/azure/architecture/databricks-monitoring/application-logs

NEW QUESTION: 71

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

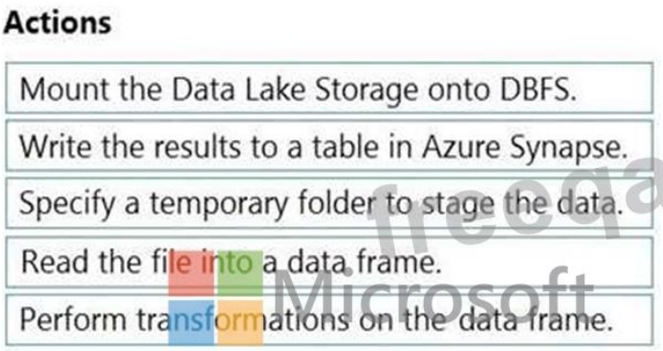
You need to configure workspace1 to support autoscaling all-purpose clusters. The solution must meet the following requirements:

Automatically scale down workers when the cluster is underutilized for three minutes.

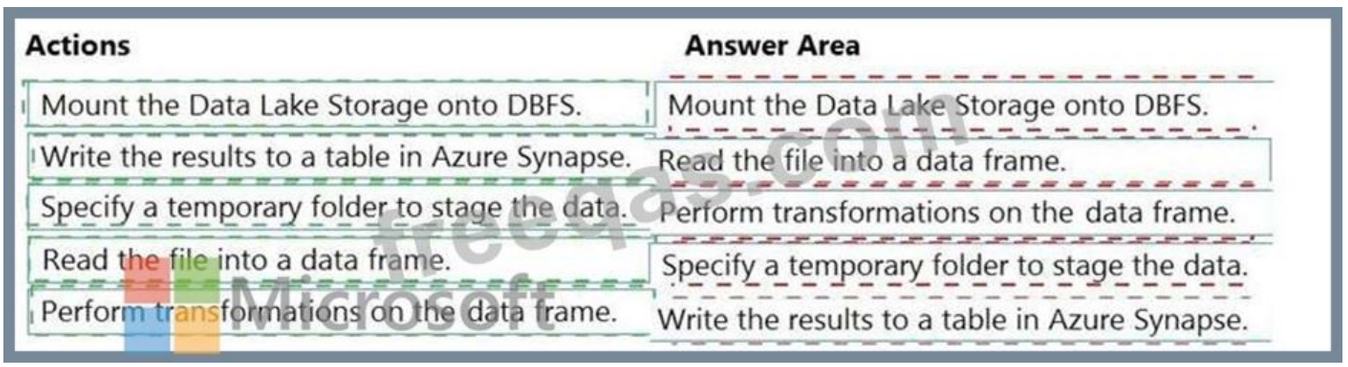
Minimize the time it takes to scale to the maximum number of workers.

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: 72
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.

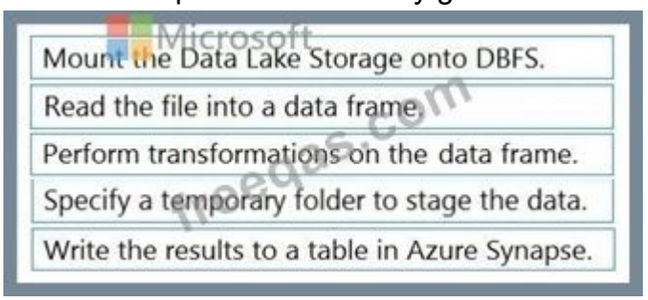


Answer:



Explanation

Table Description automatically generated



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:

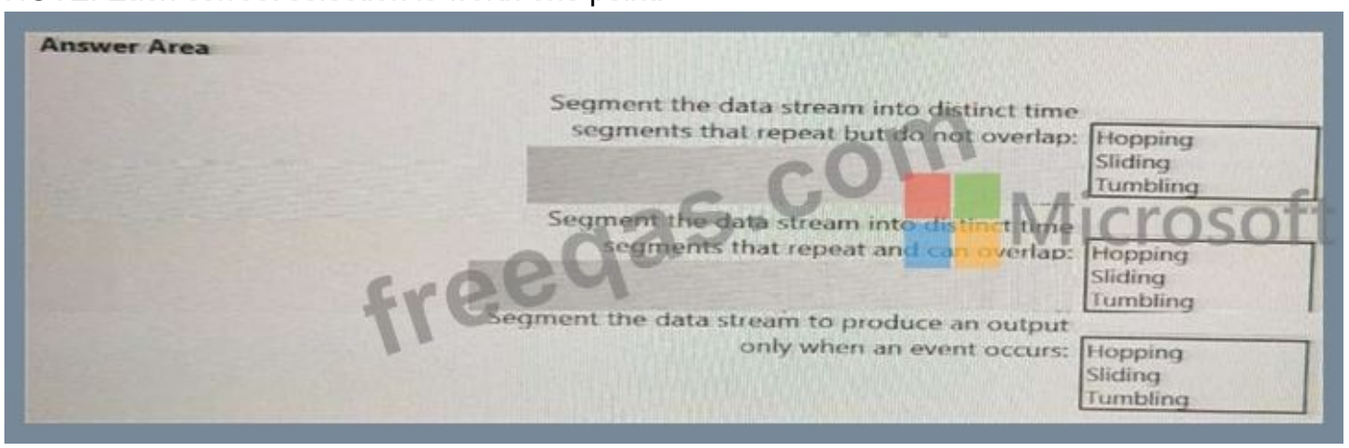
https://docs.microsoft.com/en-us/azure/azure-databricks/databricks-extract-load-sql-data-warehouse

NEW QUESTION: 73

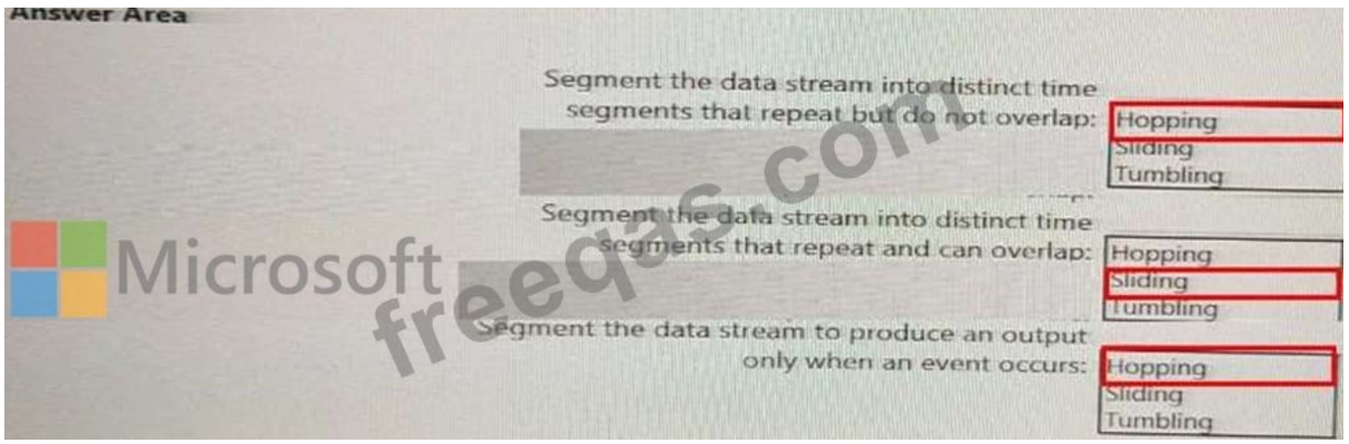
You are designing a partition strategy for a fact table in an Azure Synapse Analytics dedicated SQL pool. The table has the following specifications:

* Contain sales data for 20,000 products.
* Use hash distribution on a column named ProduclID,
* Contain 2.4 billion records for the years 20l9 and 2020.
Which number of partition ranges provides optimal compression and performance of the clustered columnstore index?
A. 2,400
B. 240
C. 40
D. 400
Answer: (SHOW ANSWER)

NEW QUESTION: 74
You are implementing Azure Stream Analytics windowing functions.
Which windowing function should you use for each requirement? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

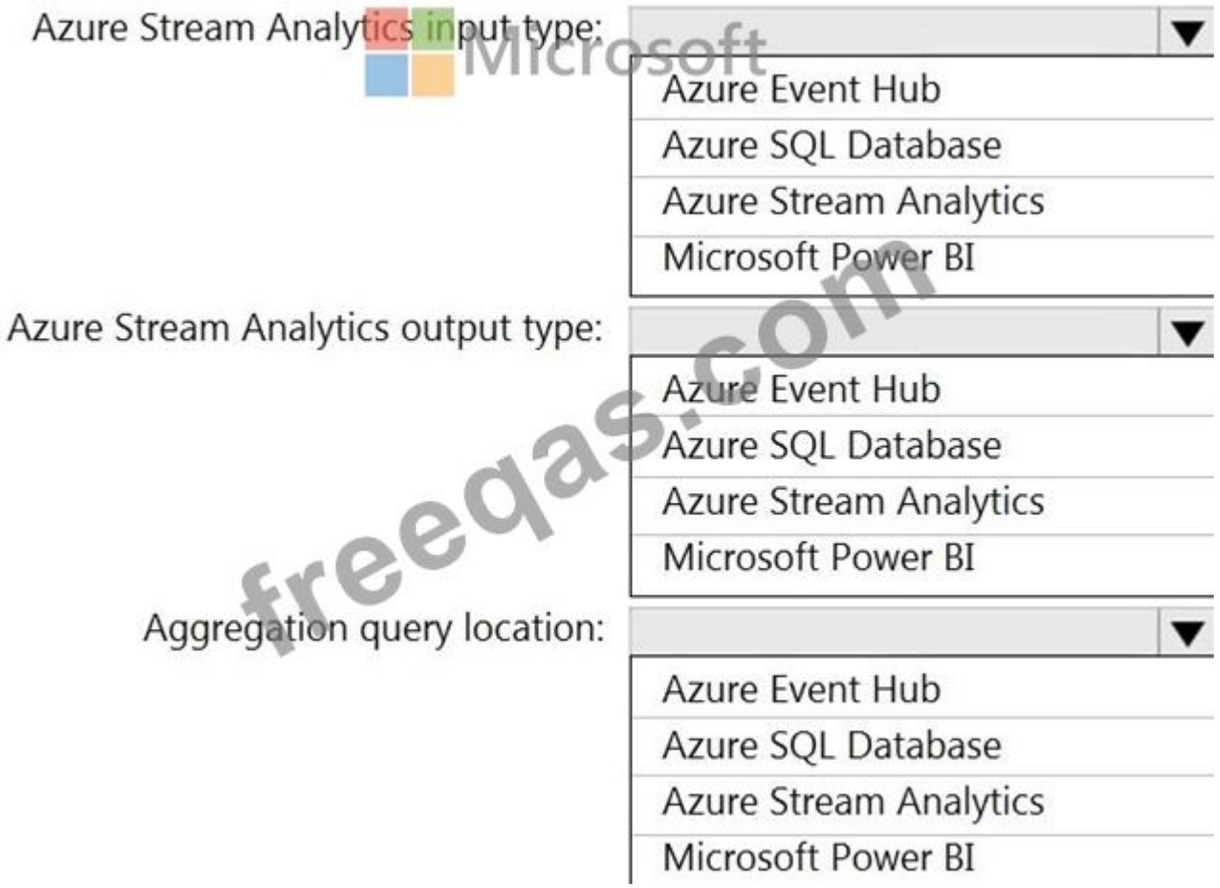


Answer:

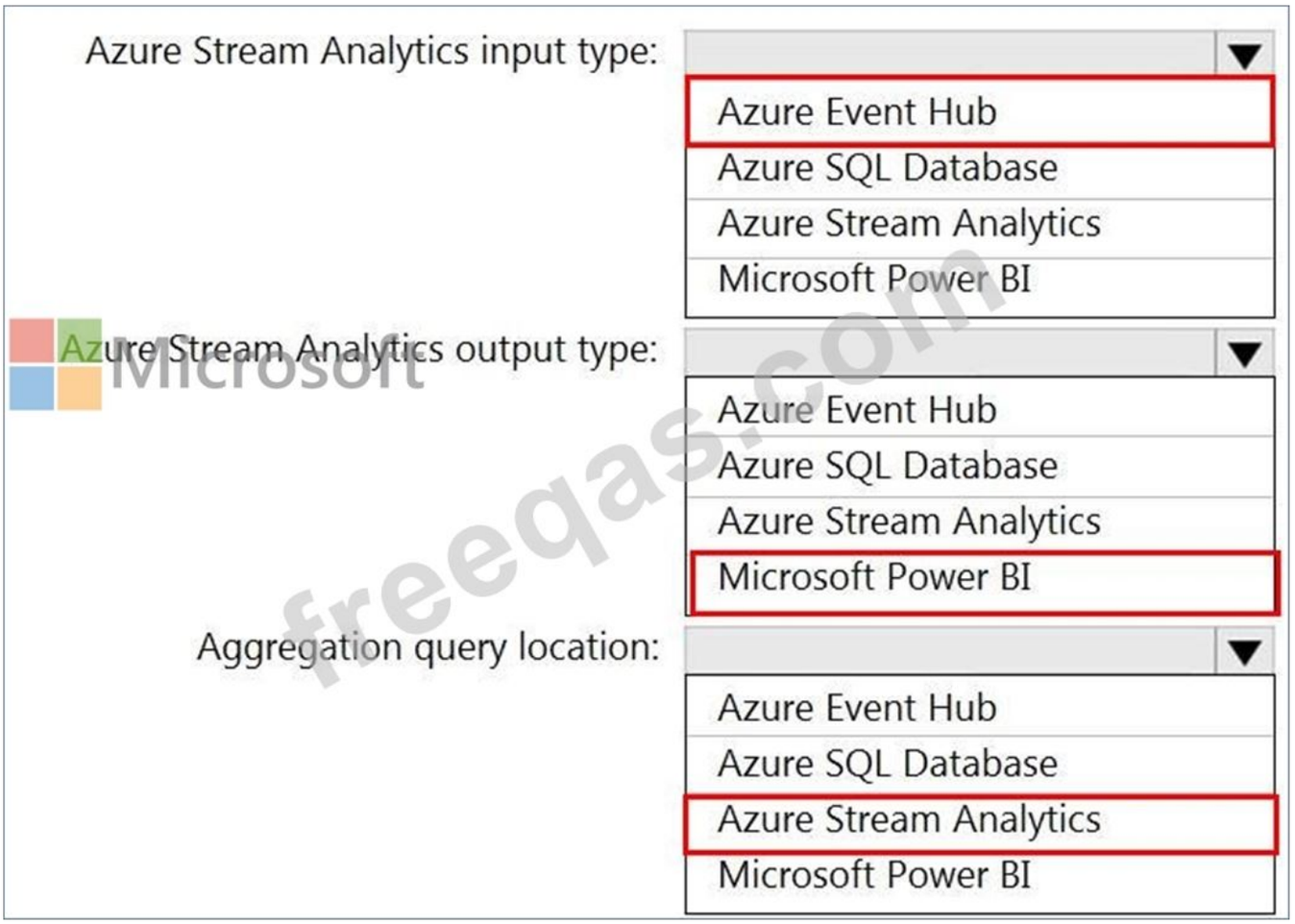


NEW QUESTION: 75
You are designing a real-time dashboard solution that will visualize streaming data from remote sensors that connect to the internet. The streaming data must be aggregated to show the average value of each 10-second interval. The data will be discarded after being displayed in the dashboard.

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



Answer:



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

NEW QUESTION: 76
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

- opendatasource
- openjson
- openquery
- openrowset

Answer Area

```
select*

FROM
[                ] (

     BULK 'https://contoso.blob.core.windows.net/contosodw',
     FORMAT= 'CSV',
     fieldterminator = '0x0b',
     fieldquote = '0x0b',
     rowterminator = '0x0b'
)
with (id varchar(50),
     contextdateventTime 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

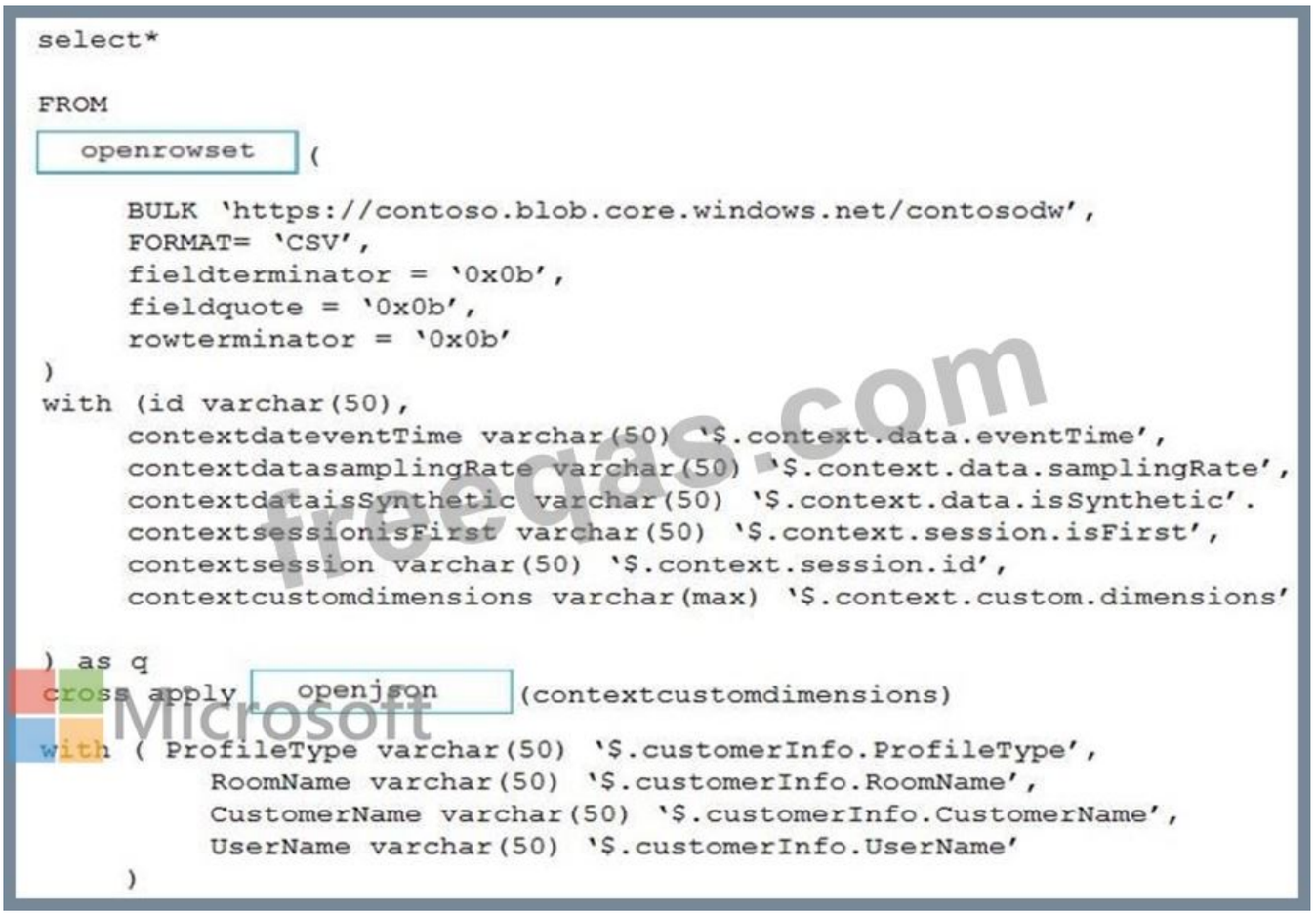
- opendatasource
- openjson
- openquery
- openrowset

Answer Area

```
select*

FROM
[ openrowset ] (

     BULK 'https://contoso.blob.core.windows.net/contosodw',
     FORMAT= 'CSV',
     fieldterminator = '0x0b',
     fieldquote = '0x0b',
     rowterminator = '0x0b'
)
with (id varchar(50),
     contextdateventTime 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'
     )
```

Explanation

Graphical user interface, text, application, email Description automatically generated



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

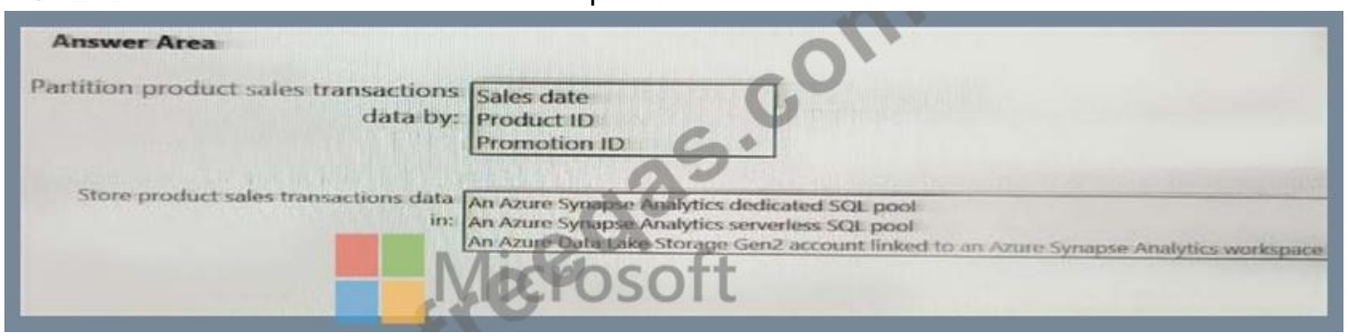
https://docs.microsoft.com/en-us/sql/relational-databases/json/import-json-documents-into-sql-server

NEW QUESTION: 77

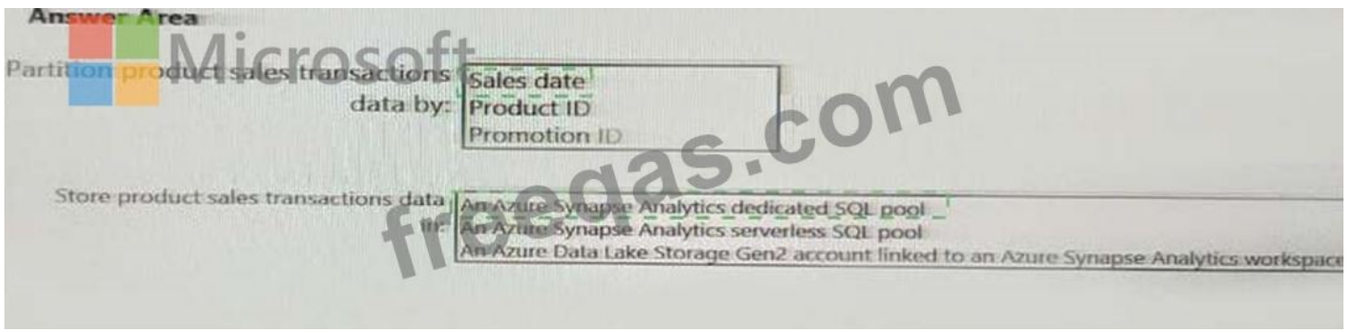
You need to design the partitions for the product sales transactions. The solution must mee the sales transaction dataset requirements.

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.



Answer:



NEW QUESTION: 78

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: You use a dedicated SQL pool to create an external table that has a additional DateTime column.
Does this meet the goal?

A. No

B. Yes

Answer: B (LEAVE A REPLY)

NEW QUESTION: 79

You have an Azure Stream Analytics job.

You need to ensure that the job has enough streaming units provisioned

You configure monitoring of the SU % Utilization metric.

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

NOTE Each correct selection is worth one point

A. Out of order Events

B. Function Events

C. Late Input Events

D. Baddogged Input Events

Answer: D (LEAVE A REPLY)

NEW QUESTION: 80

You have an on-premises data warehouse that includes the following fact tables. Both tables have the following columns: DateKey, ProductKey, RegionKey. There are 120 unique product keys and 65 unique region keys.

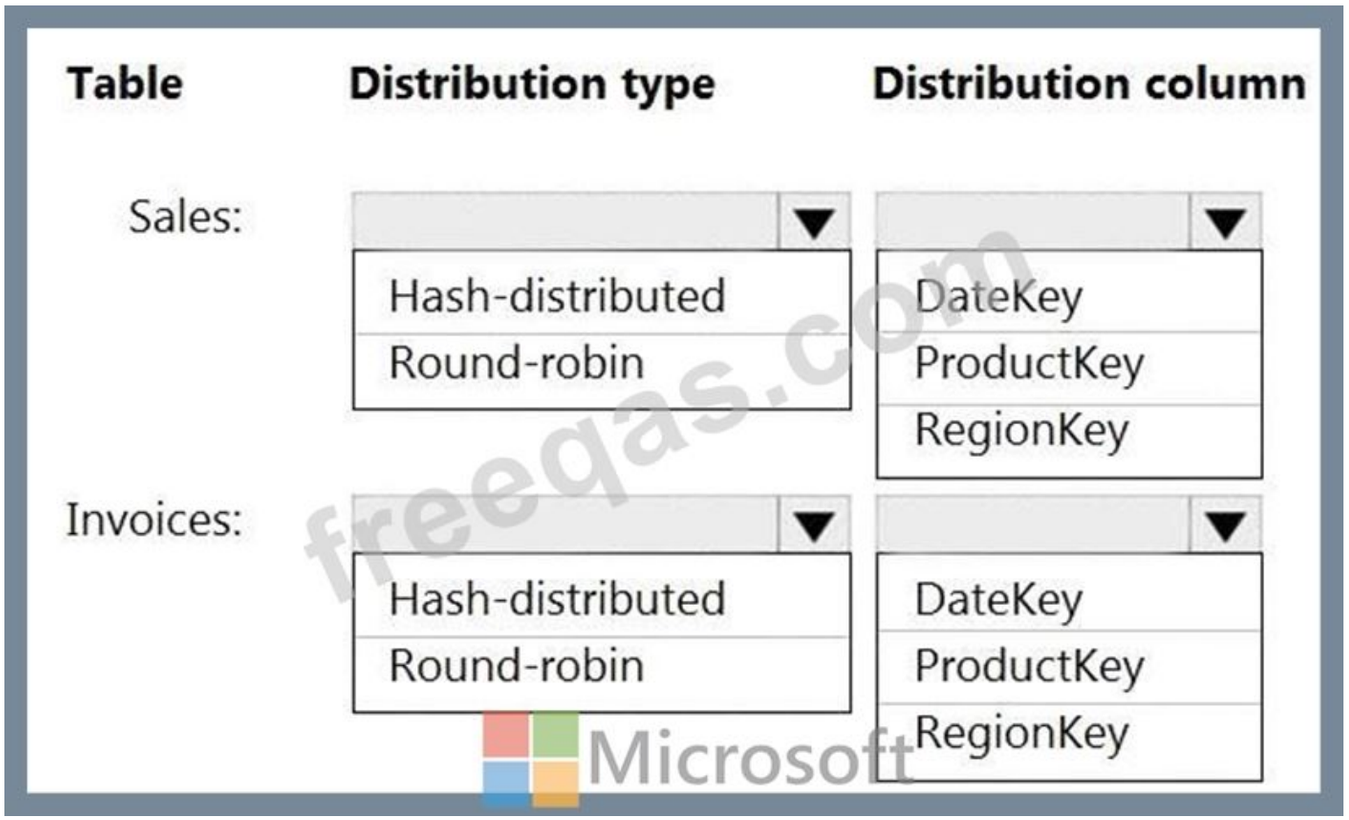
Table	Comments
Sales	The table is 600 GB in size. DateKey is used extensively in the WHERE clause in queries. ProductKey is used extensively in join operations. RegionKey is used for grouping. Severity-five percent of records relate to one of 40 regions.
Invoice	The table is 6 GB in size. DateKey and ProductKey are used extensively in the WHERE clause in queries. RegionKey is used for grouping.

Queries that use the data warehouse take a long time to complete.

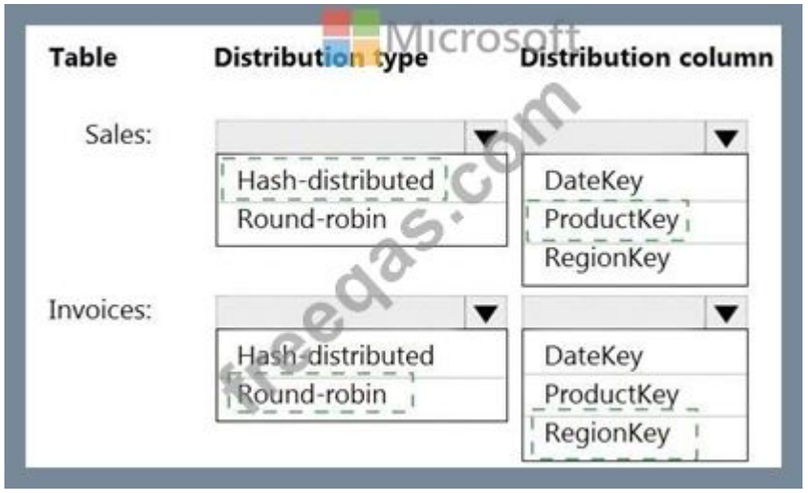
You plan to migrate the solution to use Azure Synapse Analytics. You need to ensure that the Azure-based solution optimizes query performance and minimizes processing skew.

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

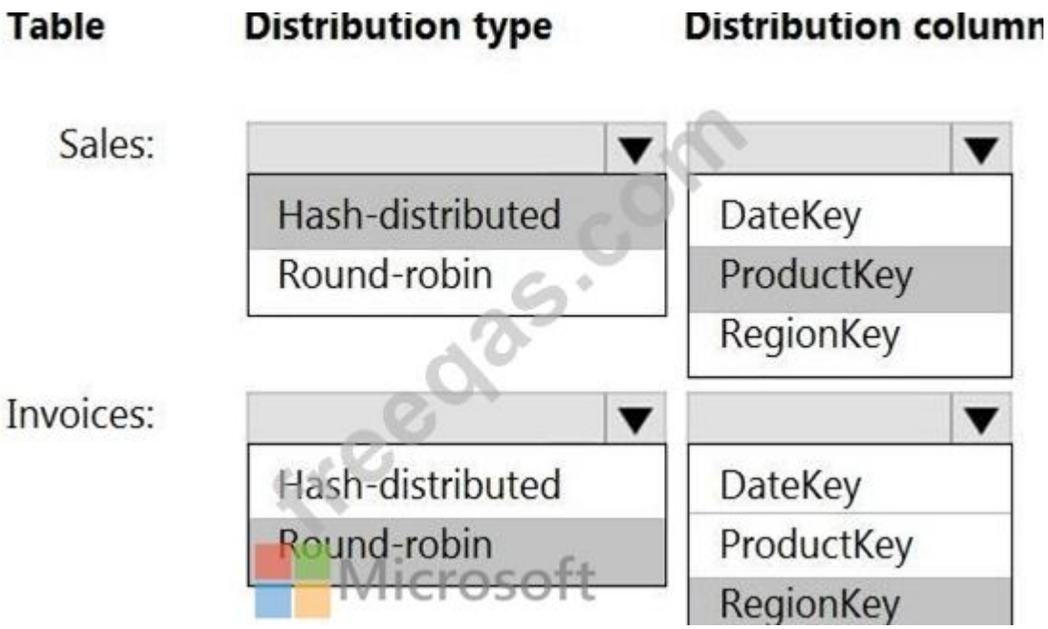
NOTE: Each correct selection is worth one point



Answer:



Explanation



Box 1: Hash-distributed

Box 2: ProductKey

ProductKey is used extensively in joins.

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

Box 3: Round-robin

Box 4: RegionKey

Round-robin tables are useful for improving loading speed.

Consider using the round-robin distribution for your table in the following scenarios:

* When getting started as a simple starting point since it is the default
* If there is no obvious joining key
* If there is not good candidate column for hash distributing the table
* If the table does not share a common join key with other tables
* If the join is less significant than other joins in the query
* When the table is a temporary staging table

Note: A distributed table appears as a single table, but the rows are actually stored across 60 distributions. The rows are distributed with a hash or round-robin algorithm.

Reference:

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

NEW QUESTION: 81

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.

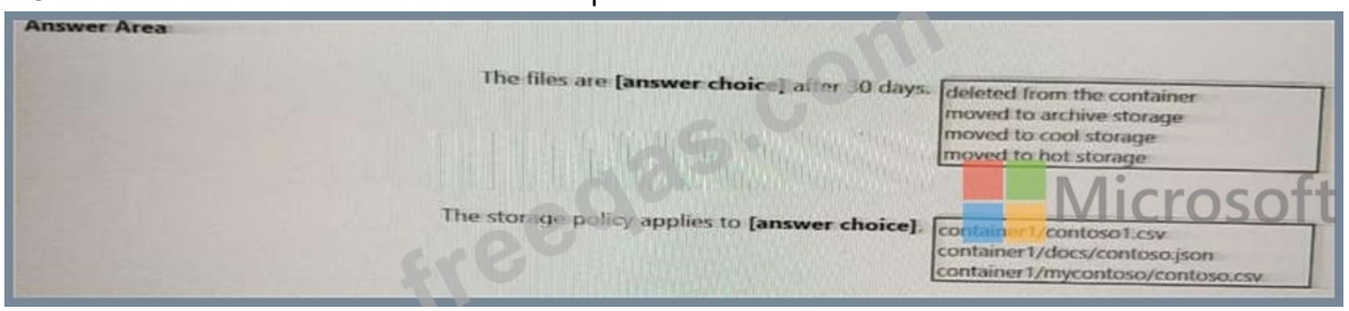
You plan to create an Azure Databricks workspace that has a tiered structure. The workspace will contain the following three workloads:

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 High Concurrency cluster for the jobs.
Does this meet the goal?
A. Yes
B. No
Answer: A (LEAVE A REPLY)
We need a High Concurrency cluster for the data engineers and 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

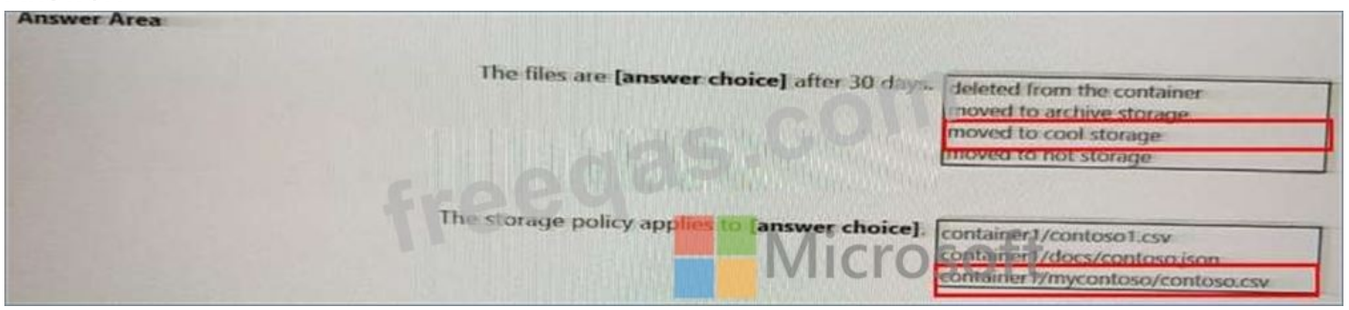
NEW QUESTION: 82
You store files in an Azure Data Lake Storage Gen2 container. The container has the storage policy shown in the following exhibit.

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.



Answer:

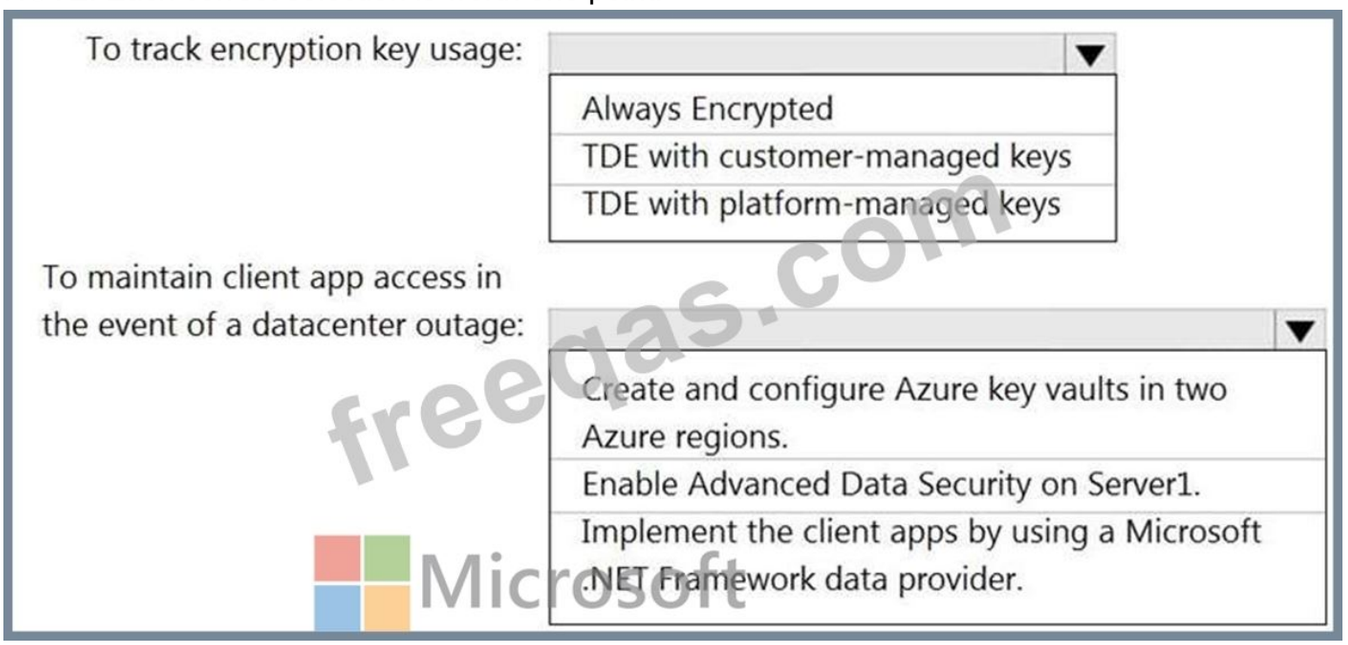


NEW QUESTION: 83

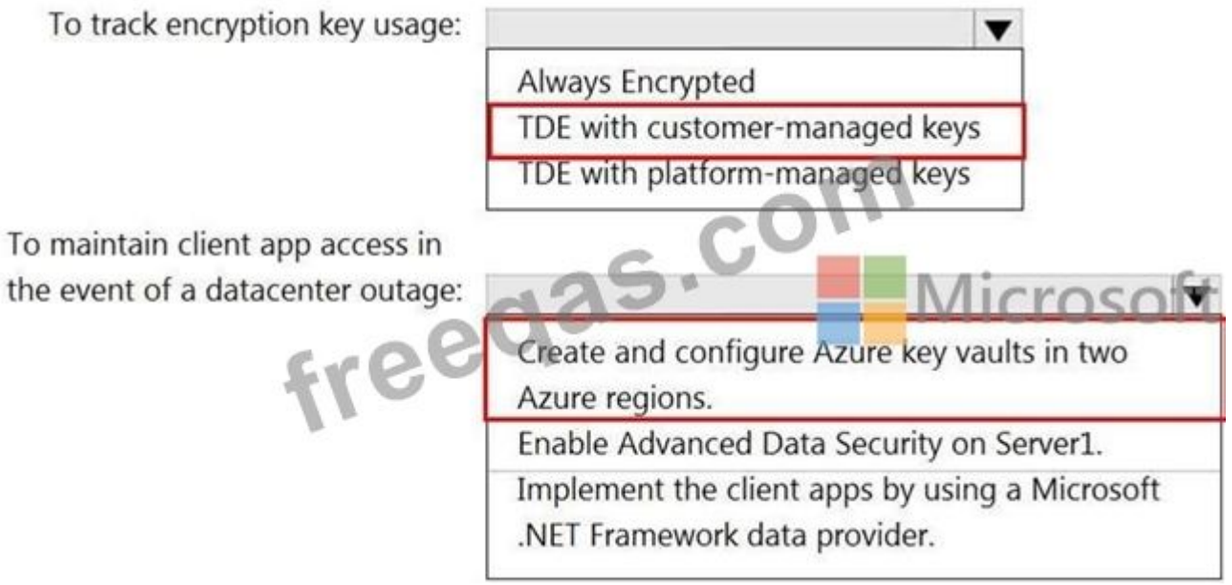
You have an Azure subscription that contains a logical Microsoft SQL server named Server1. Server1 hosts an Azure Synapse Analytics SQL dedicated pool named Pool1.

You need to recommend a Transparent Data Encryption (TDE) solution for Server1. The solution must meet the following requirements:

Track the usage of encryption keys.
Maintain the access of client apps to Pool1 in the event of an Azure datacenter outage that affects the availability of the encryption keys.
What should you include in the recommendation? To answer, select the appropriate options in the answer are a.
NOTE: Each correct selection is worth one point.



Answer:

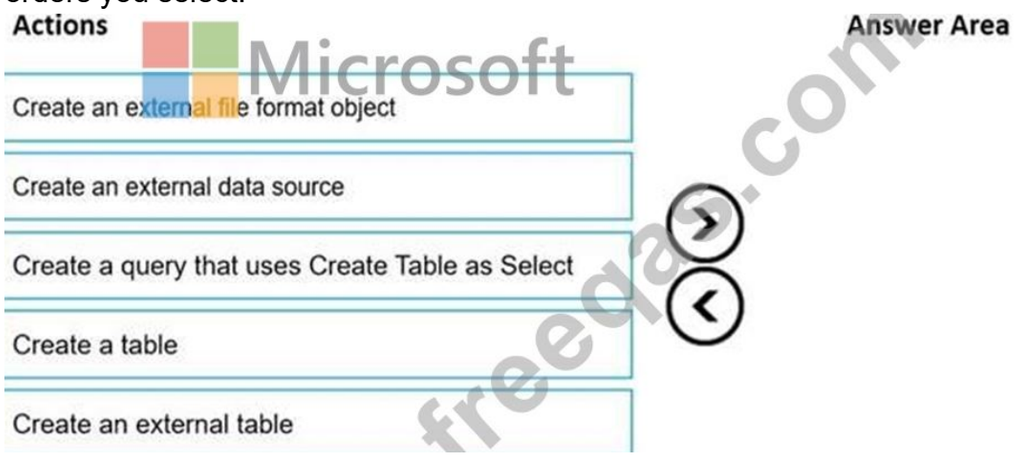


Reference:
https://docs.microsoft.com/en-us/azure/synapse-analytics/security/workspaces-encryption
https://docs.microsoft.com/en-us/azure/key-vault/general/logging

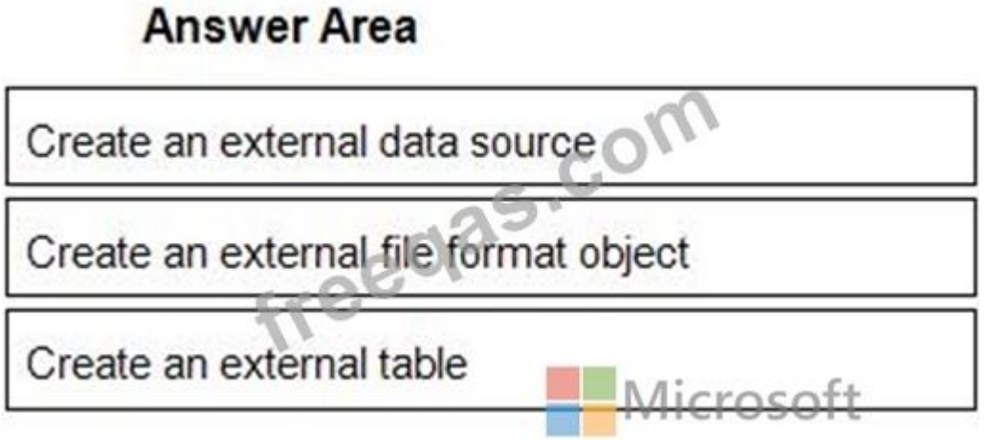
NEW QUESTION: 84
You have an Azure Synapse Analytics dedicated SQL pool that contains a table named Contacts.
Contacts contains a column named Phone.

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)
Explanation
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: 85
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.

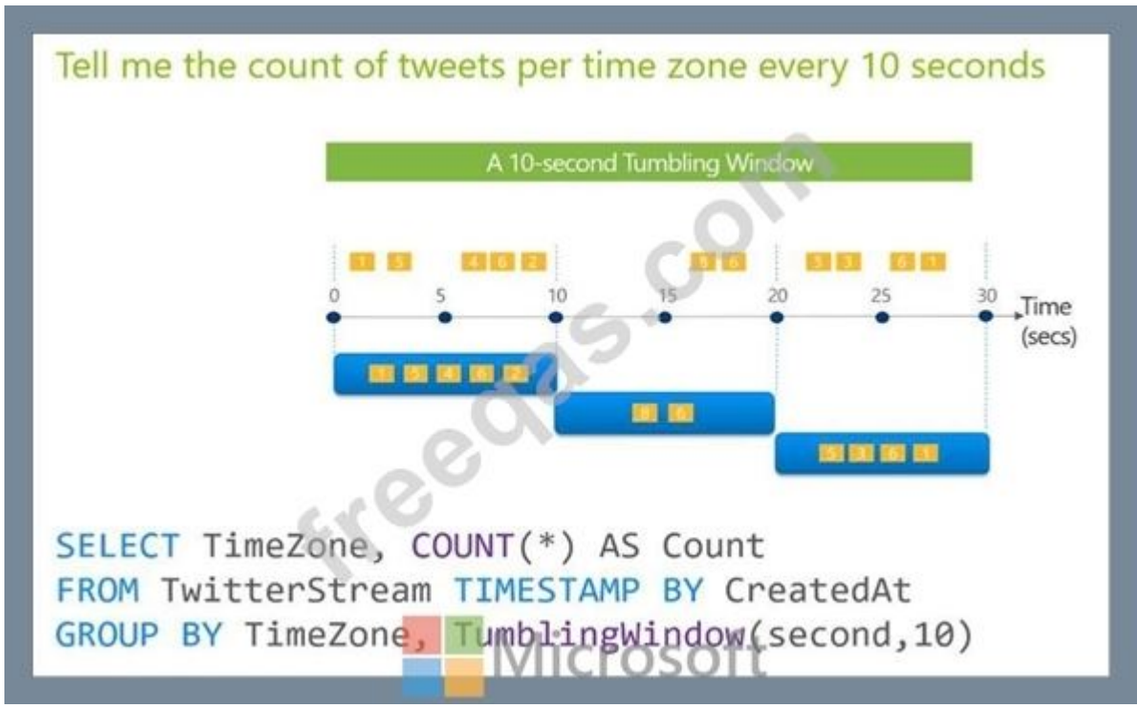


Answer:



1 - Create an external data source
2 - Create an external file format object
3 - Create an external table
Reference:
https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/develop-tables-external-tables

NEW QUESTION: 86
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.



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

NEW QUESTION: 87
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: (SHOW ANSWER)
Explanation
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: 88

You have an Azure Synapse Analystics 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. column encryption
B. dynamic data masking
C. row-level security (RLS)
D. table partitions
E. a default value
Answer: (SHOW ANSWER)

NEW QUESTION: 89
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 ProduclID,
* Contain 2.4 billion records for the years 20l9 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: A (LEAVE A REPLY)
Explanation
Each partition should have around 1 millions records. Dedication SQL pools already have 60 partitions.
We have the formula: Records/(Partitions*60)= 1 million
Partitions= Records/(1 million * 60)
Partitions= 2.4 x 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: 90
You are designing a solution that will copy Parquet files stored in an Azure Blob storage account to an Azure Data Lake Storage Gen2 account.
The data will be loaded daily to the data lake and will use a folder structure of {Year}/{Month}/{Day}/.

You need to design a daily Azure Data Factory data load to minimize the data transfer between the two accounts.
Which two configurations should you include in the design? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.
A. Delete the files in the destination before loading new data.
B. Filter by the last modified date of the source files.
C. Delete the source files after they are copied.
D. Specify a file naming pattern for the destination.
Answer: (SHOW ANSWER)
Reference:
https://docs.microsoft.com/en-us/azure/data-factory/connector-azure-data-lake-storage

NEW QUESTION: 91
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: C (LEAVE A REPLY)
Explanation
https://docs.microsoft.com/en-us/azure/architecture/data-guide/technology-choices/stream-processing

NEW QUESTION: 92
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: D (LEAVE A REPLY)
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: 93
You have an enterprise data warehouse in Azure Synapse Analytics named DW1 on a server named Server1.
You need to verify whether the size of the transaction log file for each distribution of DW1 is smaller than 160 GB.
What should you do?
A. On the master database, execute a query against the sys.dm_pdw_nodes_os_performance_counters dynamic management view.
B. From Azure Monitor in the Azure portal, execute a query against the logs of DW1.
C. On DW1, execute a query against the sys.database_files dynamic management view.
Answer: A (LEAVE A REPLY)
D. Execute a query against the logs of DW1 by using the Get-AzOperationalInsightSearchResult PowerShell cmdlet.
Explanation:
The following query returns the transaction log size on each distribution. If one of the log files is reaching 160 GB, you should consider scaling up your instance or limiting your transaction size.
-- Transaction log size
SELECT
instance_name as distribution_db,
cntr_value*1.0/1048576 as log_file_size_used_GB,
pdw_node_id
FROM sys.dm_pdw_nodes_os_performance_counters
WHERE
instance_name like 'Distribution_%'
AND counter_name = 'Log File(s) Used Size (KB)'
Reference:
https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-manage-monitor

NEW QUESTION: 94

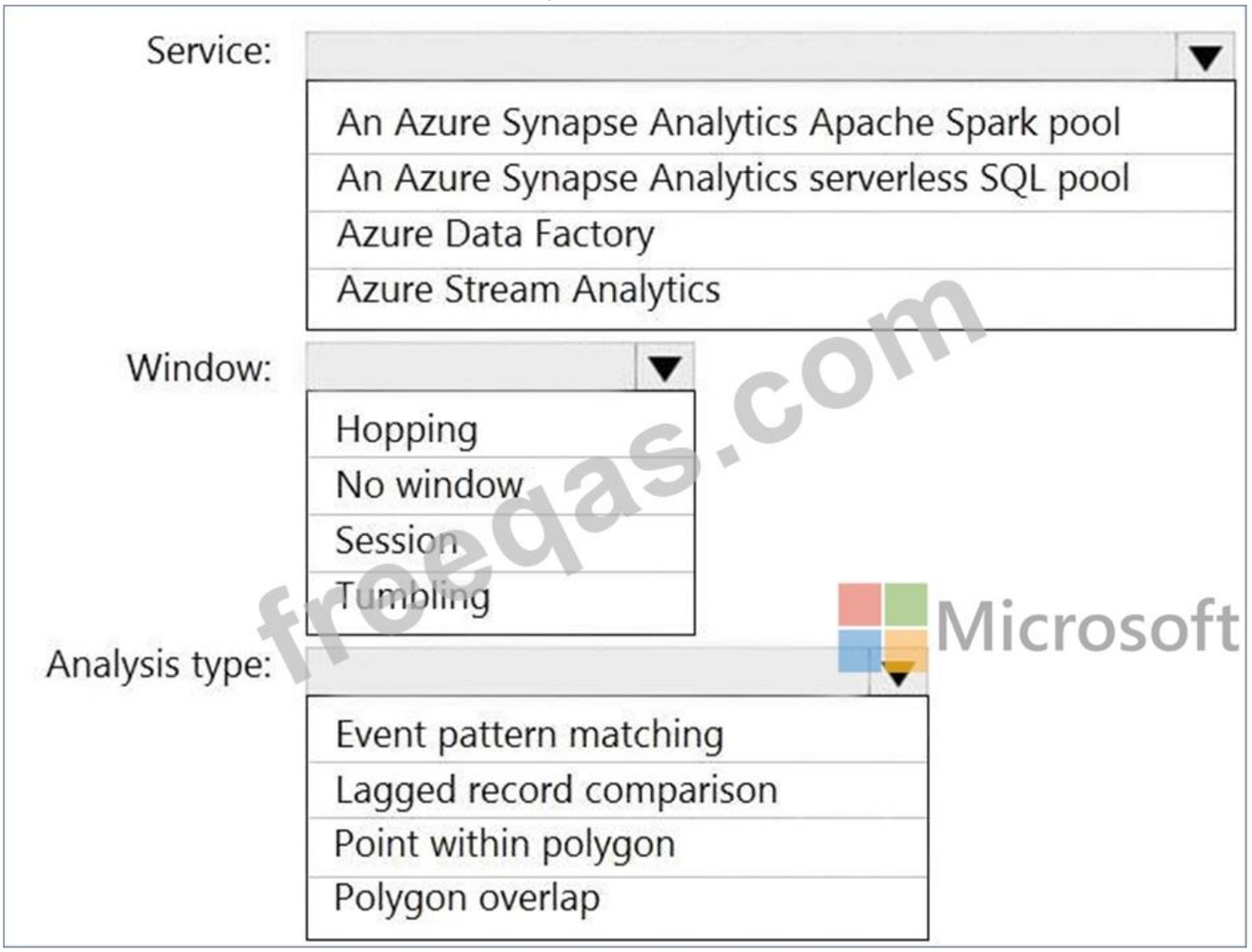
You are designing a monitoring solution for a fleet of 500 vehicles. Each vehicle has a GPS tracking device that sends data to an Azure event hub once per minute.

You have a CSV file in an Azure Data Lake Storage Gen2 container. The file maintains the expected geographical area in which each vehicle should be.

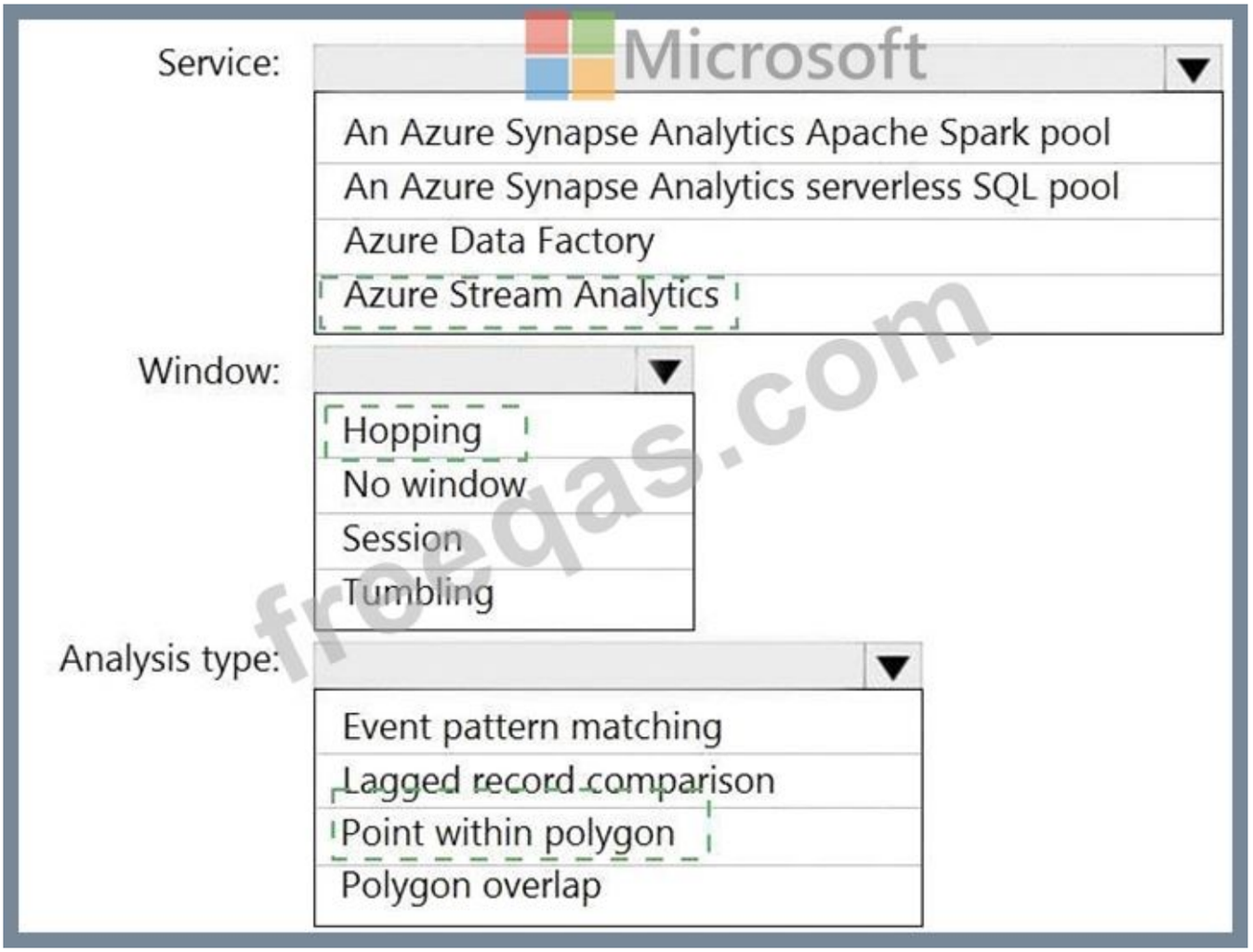
You need to ensure that when a GPS position is outside the expected area, a message is added to another event hub for processing within 30 seconds. The solution must minimize cost.

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

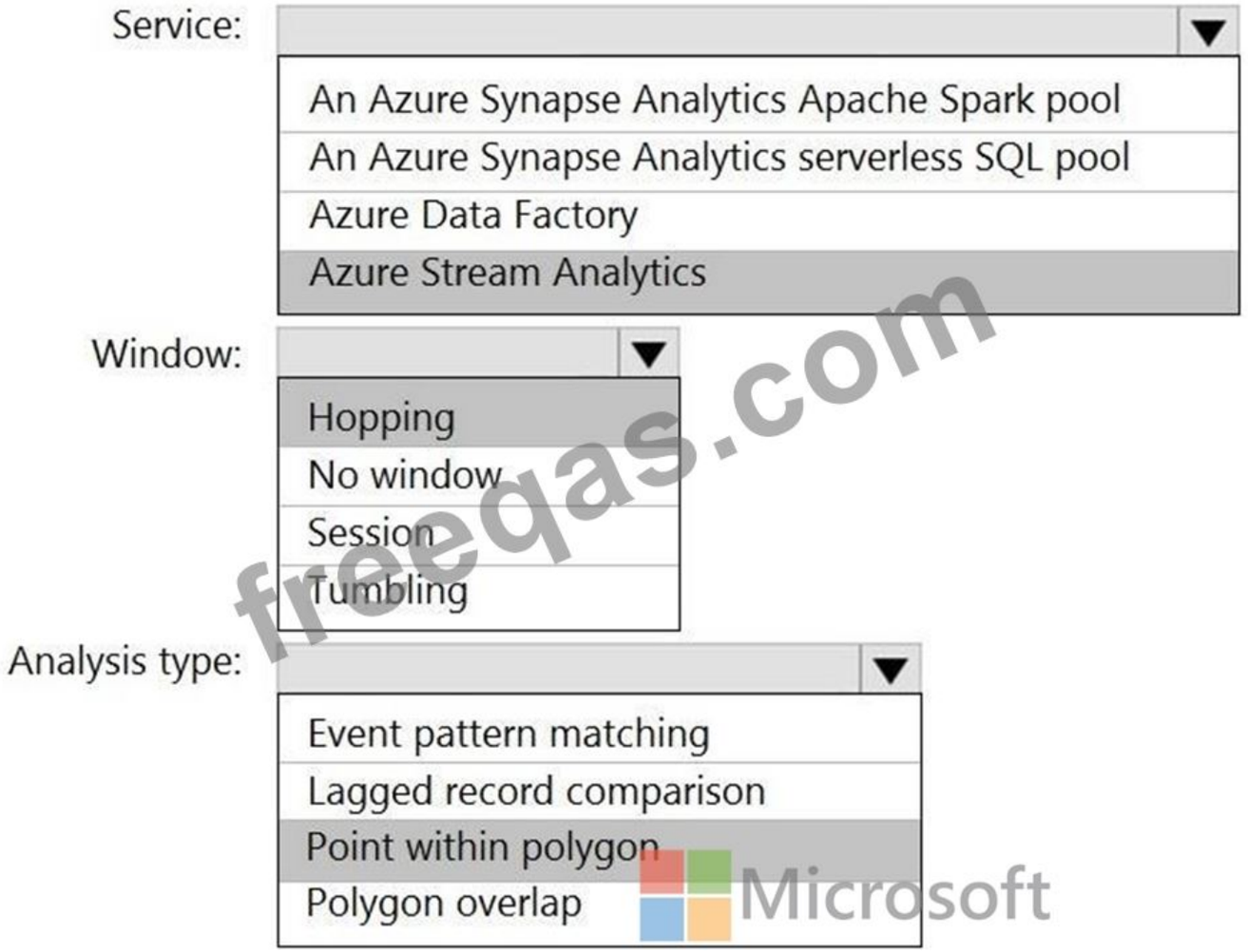
NOTE: Each correct selection is worth one point.



Answer:



Explanation



Box 1: Azure Stream Analytics

Box 2: Hopping

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.

Box 3: Point within polygon

Reference:

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

NEW QUESTION: 95

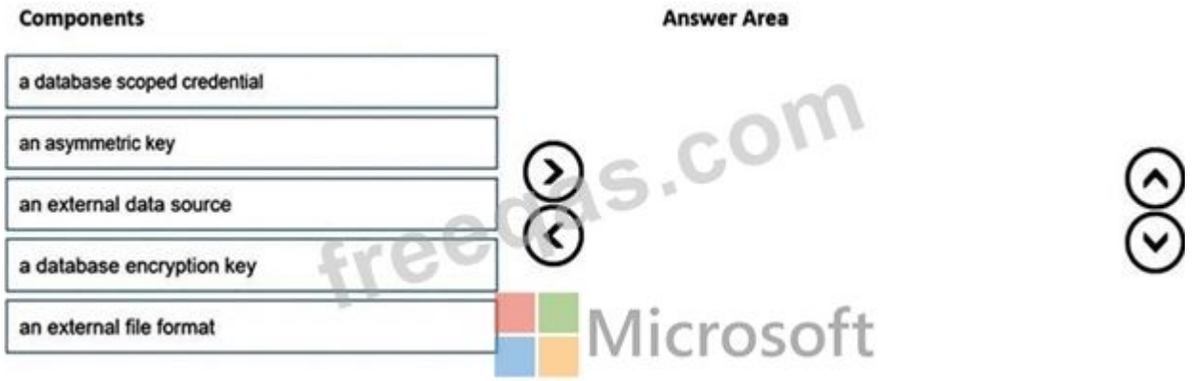
You are responsible for providing access to an Azure Data Lake Storage Gen2 account.

Your user account has contributor access to the storage account, and you have the application ID and access key.

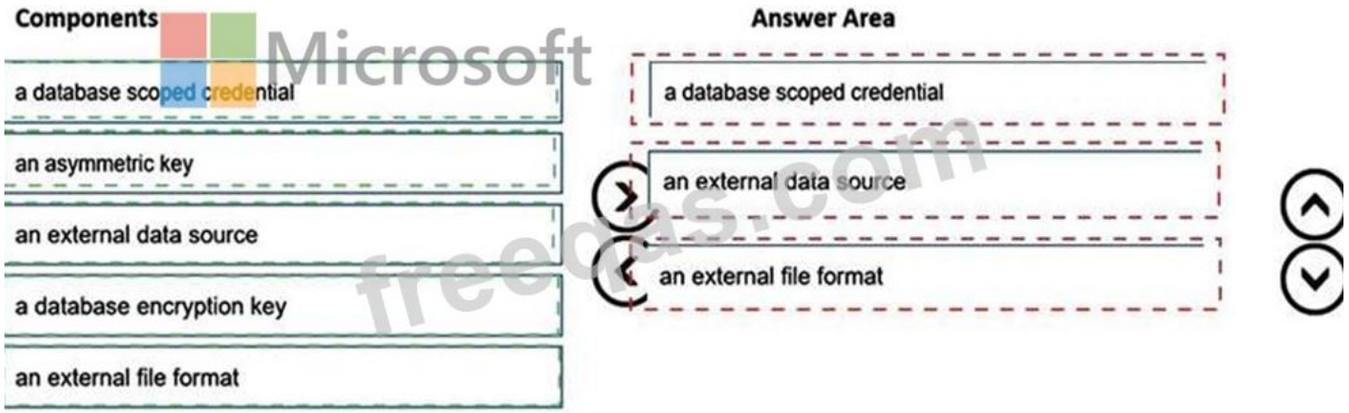
You plan to use PolyBase to load data into an enterprise data warehouse in Azure Synapse Analytics.

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.



Answer:



Explanation

NEW QUESTION: 96

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)

```
\YYYYY\WW\DataSource\SubjectArea\FileData_YYYY_MM_DD.parquet
```

B)

```
DataSource\SubjectArea\WW\YYYY\FileData_YYYY_MM_DD.parquet
```

C)

```
\DataSource\SubjectArea\YYYY\WW\FileData_YYYY_MM_DD.parquet
```

D)

```
\DataSource\SubjectArea\YYYY-WW\FileData_YYYY_MM_DD.parquet
```

E)

```
WW\YYYY\SubjectArea\DataSource\FileData_YYYY_MM_DD.parquet
```

A. Option E
B. Option A
C. Option C
D. Option D
E. Option B
Answer: D (LEAVE A REPLY)

NEW QUESTION: 97
You are designing an enterprise data warehouse in Azure Synapse Analytics that will contain a table named Customers. Customers will contain credit card information.
You need to recommend a solution to provide salespeople with the ability to view all the entries in Customers.
The solution must prevent all the salespeople from viewing or inferring the credit card information.
What should you include in the recommendation?
A. data masking
B. Always Encrypted
C. column-level security
D. row-level security
Answer: A (LEAVE A REPLY)
SQL Database dynamic data masking limits sensitive data exposure by masking it to non-privileged users.
The Credit card masking method exposes the last four digits of the designated fields and adds a constant string as a prefix in the form of a credit card.
Example: XXXX-XXXX-XXXX-1234
Reference:
https://docs.microsoft.com/en-us/azure/sql-database/sql-database-dynamic-data-masking-get-started

NEW QUESTION: 98
You have an Azure Synapse Analytics dedicated SQL pool named SA1 that contains a table named Table1. You need to identify tables that have a high percentage of deleted rows. What should you run?
A)

```
sys.pdw_nodes_column_store_segments
```

B)

```
sys.dm_db_column_store_row_group_operational_stats
```

C)

```
sys.pdw_nodes_column_store_row_groups
```

D)

```
sys.dm_db_column_store_row_group_physical_stats
```

A. Option

B. Option

C. Option

D. Option

Answer: B (LEAVE A REPLY)

NEW QUESTION: 99

You have an Azure subscription that contains the following resources:

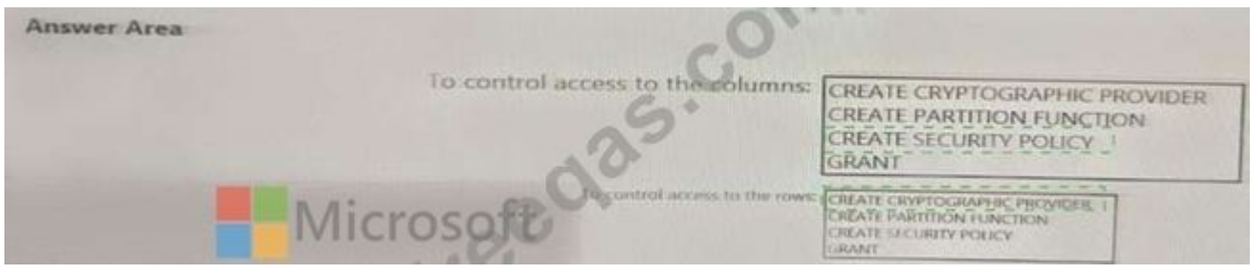
* An Azure Active Directory (Azure AD) tenant that contains a security group named Group1.

* 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.

NOTE: Each appropriate options in the answer area.

Answer:



NEW QUESTION: 100

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.

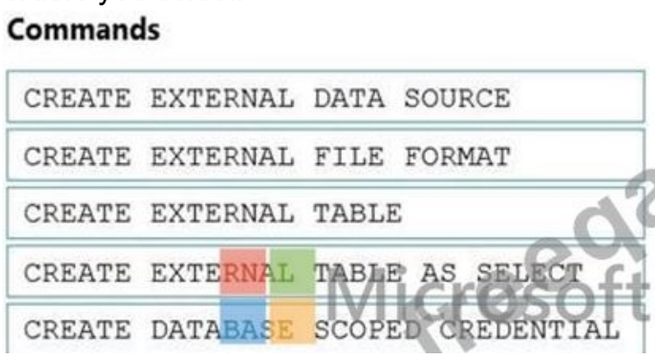
C. Deploy an Azure Stream Analytics job and use an Azure Automation runbook to check the status of the job and to start the job if it stops.

D. Set Data Lake Storage to use geo-redundant storage (GRS).

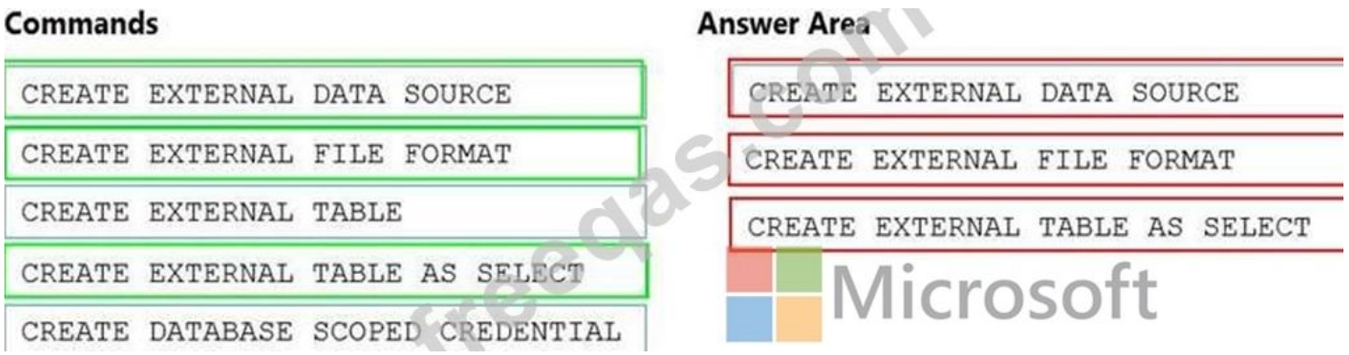
Answer: A (LEAVE A REPLY)

Explanation
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: 101
You need to ensure that the Twitter feed data can be analyzed in the dedicated SQL pool. The solution must meet the customer sentiment analytics requirements.
Which three Transaction-SQL DDL commands should you run in sequence? To answer, move the appropriate commands from the list of commands 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.



Answer:



Reference:
https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/develop-tables-external-tables
Topic 2, Litware, inc.
Overview

Litware, Inc. owns and operates 300 convenience stores across the US. The company sells a variety of packaged foods and drinks, as well as a variety of prepared foods, such as sandwiches and pizzas.
Litware has a loyalty club whereby members can get daily discounts on specific items by providing their membership number at checkout.
Litware employs business analysts who prefer to analyze data by using Microsoft Power BI, and data scientists who prefer analyzing data in Azure Databricks notebooks.

Requirements

Business Goals

Litware wants to create a new analytics environment in Azure to meet the following requirements:
See inventory levels across the stores. Data must be updated as close to real time as possible.
Execute ad hoc analytical queries on historical data to identify whether the loyalty club discounts increase sales of the discounted products.
Every four hours, notify store employees about how many prepared food items to produce based on historical demand from the sales data.

Technical Requirements

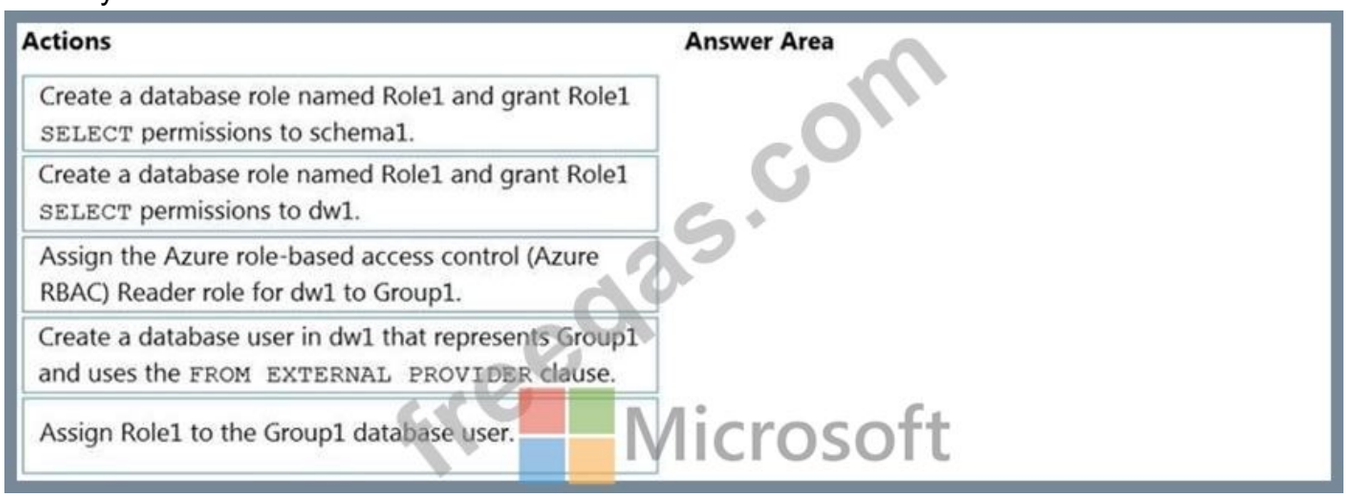
Litware identifies the following technical requirements:
Minimize the number of different Azure services needed to achieve the business goals.
Use platform as a service (PaaS) offerings whenever possible and avoid having to provision virtual machines that must be managed by Litware.
Ensure that the analytical data store is accessible only to the company's on-premises network and Azure services.
Use Azure Active Directory (Azure AD) authentication whenever possible.
Use the principle of least privilege when designing security.
Stage Inventory data in Azure Data Lake Storage Gen2 before loading the data into the analytical data store. Litware wants to remove transient data from Data Lake Storage once the data is no longer in use. Files that have a modified date that is older than 14 days must be removed.
Limit the business analysts' access to customer contact information, such as phone numbers, because this type of data is not analytically relevant.
Ensure that you can quickly restore a copy of the analytical data store within one hour in the event of corruption or accidental deletion.

Planned Environment

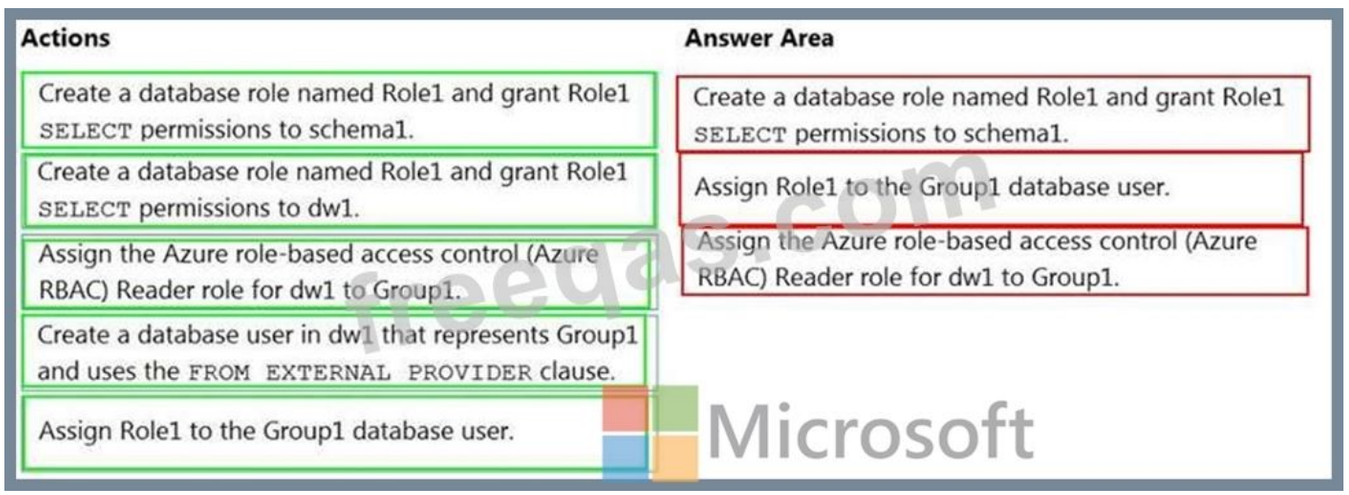
Litware plans to implement the following environment:
The application development team will create an Azure event hub to receive real-time sales data, including store number, date, time, product ID, customer loyalty number, price, and discount amount, from the point of sale (POS) system and output the data to data storage in Azure.
Customer data, including name, contact information, and loyalty number, comes from Salesforce, a SaaS application, and can be imported into Azure once every eight hours. Row modified dates are not trusted in the source table.
Product data, including product ID, name, and category, comes from Salesforce and can be imported into Azure once every eight hours. Row modified dates are not trusted in the source table.
Daily inventory data comes from a Microsoft SQL server located on a private network.

Litware currently has 5 TB of historical sales data and 100 GB of customer data. The company expects approximately 100 GB of new data per month for the next year.
Litware will build a custom application named FoodPrep to provide store employees with the calculation results of how many prepared food items to produce every four hours.
Litware does not plan to implement Azure ExpressRoute or a VPN between the on-premises network and Azure.

NEW QUESTION: 102
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.



Answer:



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

NEW QUESTION: 103
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 Storage account that contains 100 GB of files. The files contain rows of 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 copy the files to a table that has a columnstore index.
Does this meet the goal?
A. Yes
B. No
Answer: (SHOW ANSWER)
Explanation
Instead convert the files to compressed delimited text files.
Reference:
https://docs.microsoft.com/en-us/azure/sql-data-warehouse/guidance-for-loading-data

NEW QUESTION: 104
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.

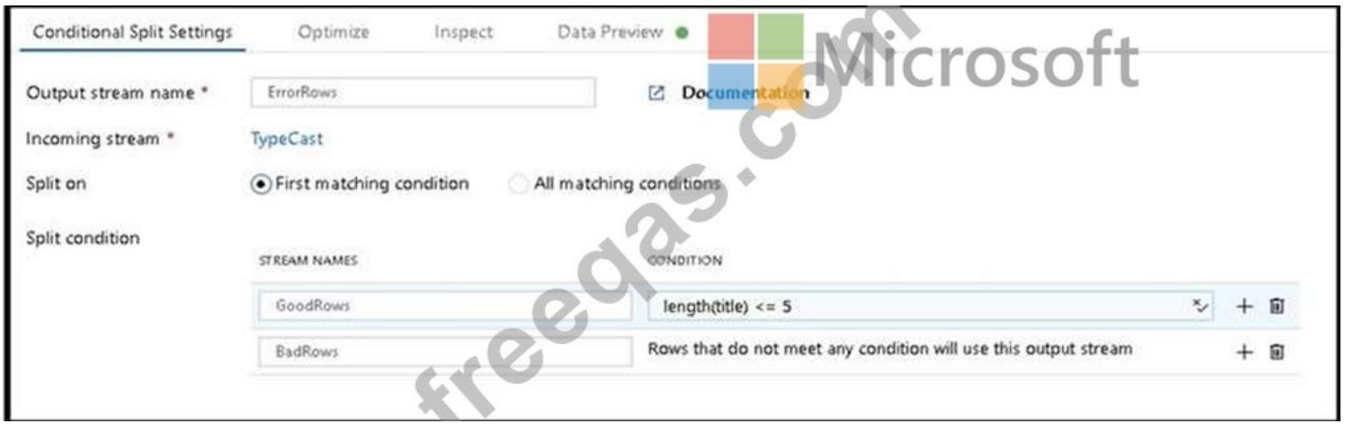
Answer:

Reference:
https://streamsets.com/documentation/datacollector/latest/help/datacollector/UserGuide/Destinations/ADLS-G2-D.html

NEW QUESTION: 105
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)

Explanation
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-a

NEW QUESTION: 106
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 Analytic dedicated SQL pool. The CSV file contains three 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. To the data flow, add a sink transformation to write the rows to a file in blob storage.
B. To the data flow, add a Conditional Split transformation to separate the rows that will cause truncation errors.
C. To the data flow, add a filter transformation to filter out rows that will cause truncation errors.
D. Add a select transformation to select only the rows that will cause truncation errors.
Answer: A,B (LEAVE A REPLY)
Explanation
B: Example:
1. This conditional split transformation defines the maximum length of "title" to be five. Any row that is less than or equal to five will go into the GoodRows stream. Any row that is larger than five will go into the BadRows stream.

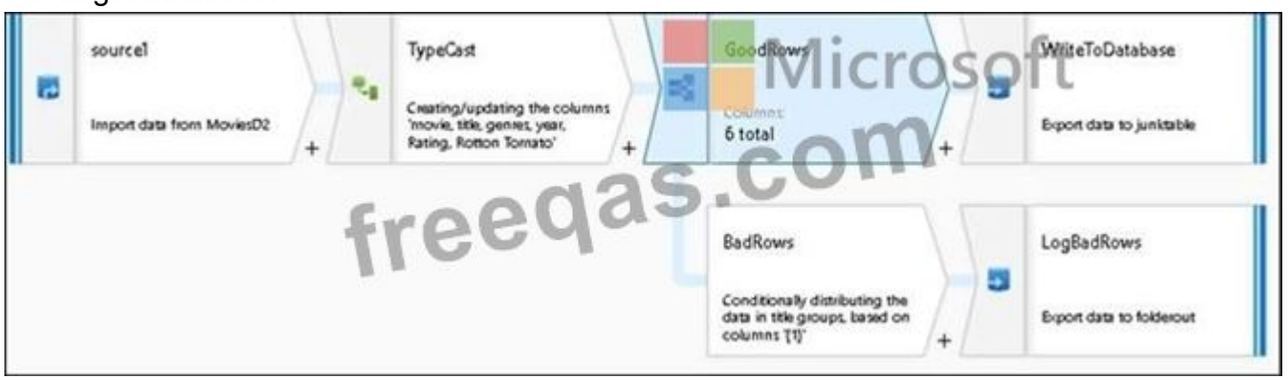


2. This conditional split transformation defines the maximum length of "title" to be five. Any row that is less than or equal to five will go into the GoodRows stream. Any row that is larger than five will go into the BadRows stream.

A:

3. Now we need to log the rows that failed. Add a sink transformation to the BadRows stream for logging. Here, we'll "auto-map" all of the fields so that we have logging of the complete transaction record. This is a text-delimited CSV file output to a single file in Blob Storage. We'll call the log file "badrows.csv".

4. The completed data flow is shown below. We are now able to split off error rows to avoid the SQL truncation errors and put those entries into a log file. Meanwhile, successful rows can continue to write to our target database.



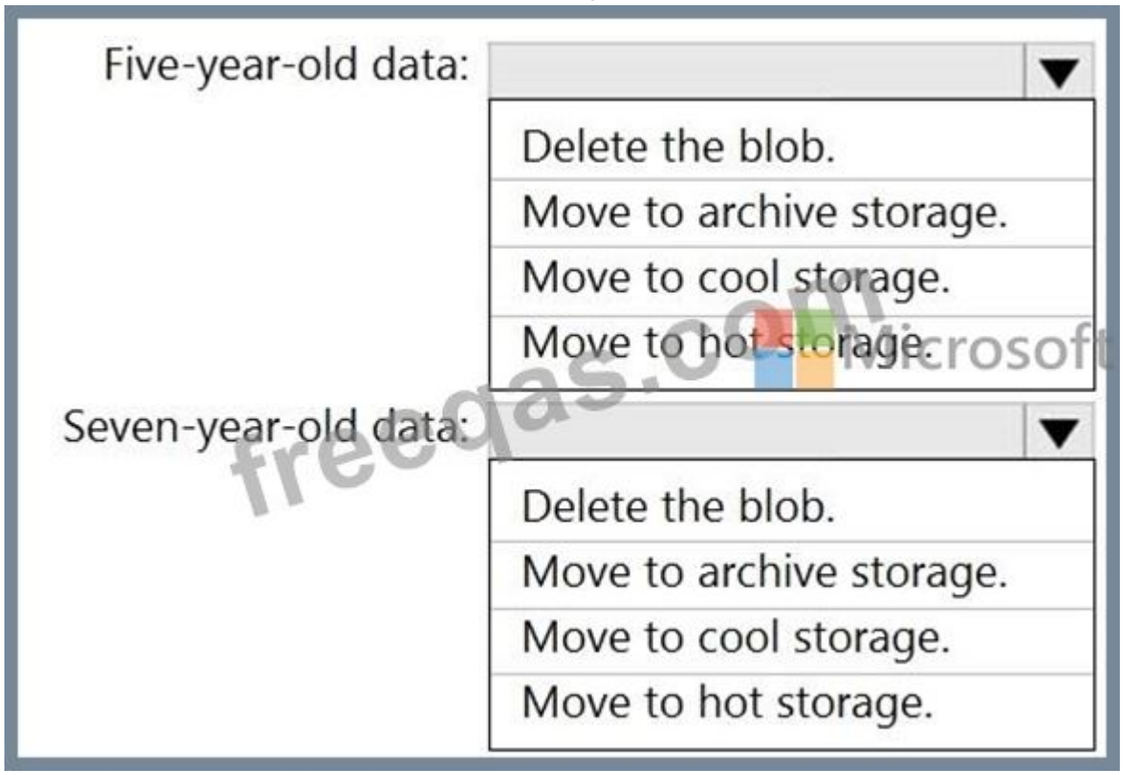
Reference:

https://docs.microsoft.com/en-us/azure/data-factory/how-to-data-flow-error-rows

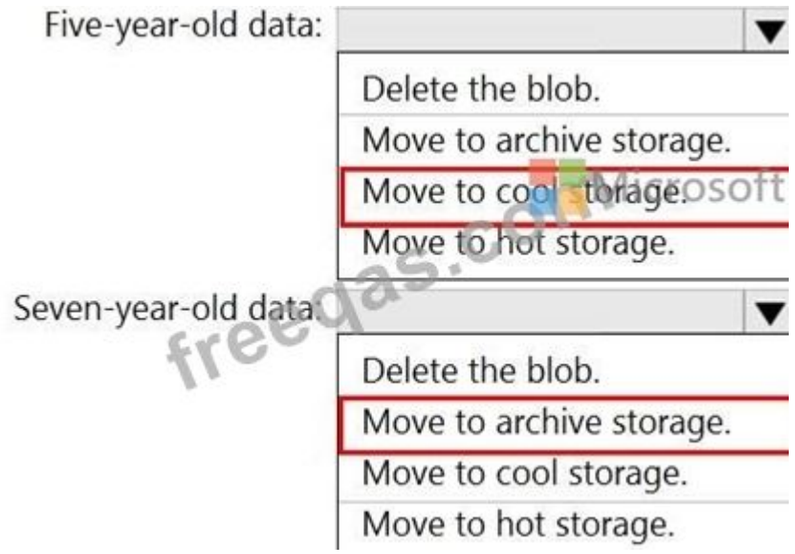
NEW QUESTION: 107
You have an Azure Data Factory pipeline that performs an incremental load of source data to an Azure Data Lake Storage Gen2 account.
Data to be loaded is identified by a column named LastUpdatedDate in the source table.
You plan to execute the pipeline every four hours.
You need to 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. event
B. on-demand
C. schedule
D. tumbling window
Answer: D (LEAVE A REPLY)
Explanation
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: 108
You have an Azure Data Lake Storage Gen2 container.
Data is ingested into the container, and then transformed by a data integration application. The data is NOT modified after that. Users can read files in the container but cannot modify the files.
You need to design a data archiving solution that meets the following requirements:
New data is accessed frequently and must be available as quickly as possible.
Data that is older than five years is accessed infrequently but must be available within one second when requested.
Data that is older than seven years is NOT accessed. After seven years, the data must be persisted at the lowest cost possible.
Costs must be minimized while maintaining the required availability.

How should you manage the data? To answer, select the appropriate options in the answer are a.
NOTE: Each correct selection is worth one point



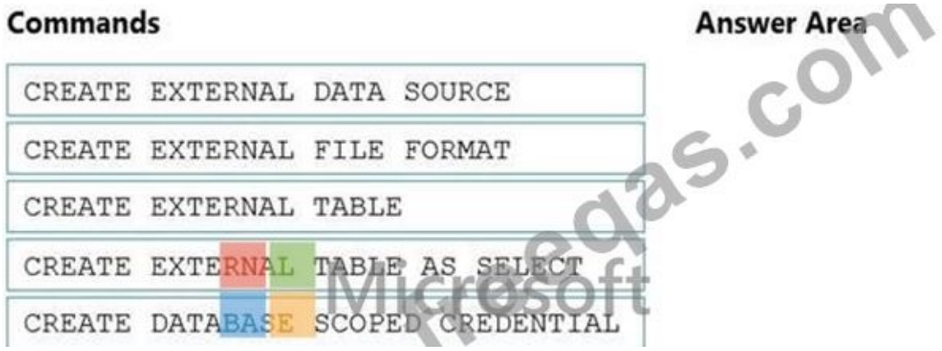
Answer:



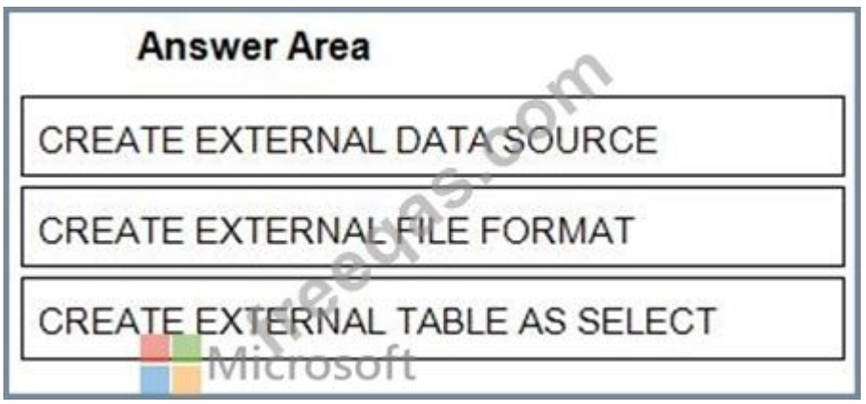
Reference:
https://docs.microsoft.com/en-us/azure/storage/blobs/storage-blob-storage-tiers
https://azure.microsoft.com/en-us/updates/reduce-data-movement-and-make-your-queries-more-efficient-with-the-general-availability-of-replicated-tables/
https://azure.microsoft.com/en-us/blog/replicated-tables-now-generally-available-in-azure-sql-data-warehouse/

NEW QUESTION: 109

You need to ensure that the Twitter feed data can be analyzed in the dedicated SQL pool. The solution must meet the customer sentiment analytics requirements.
Which three Transaction-SQL DDL commands should you run in sequence? To answer, move the appropriate commands from the list of commands 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.



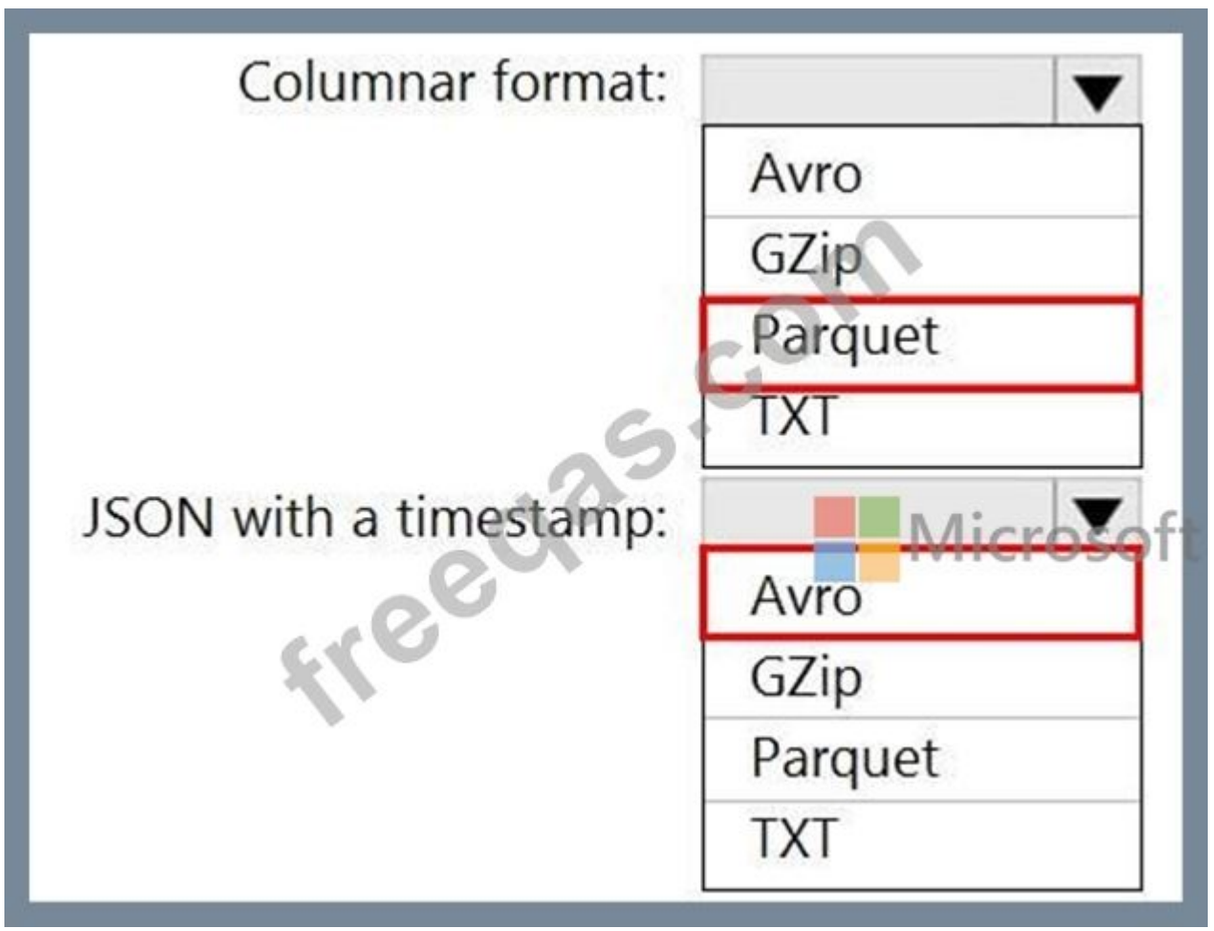
Answer:



1 - CREATE EXTERNAL DATA SOURCE
2 - CREATE EXTERNAL FILE FORMAT
3 - CREATE EXTERNAL TABLE AS SELECT
Reference:
https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/develop-tables-external-tables

NEW QUESTION: 110
You need to output files from Azure Data Factory.
Which file format should you use for each type of output? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

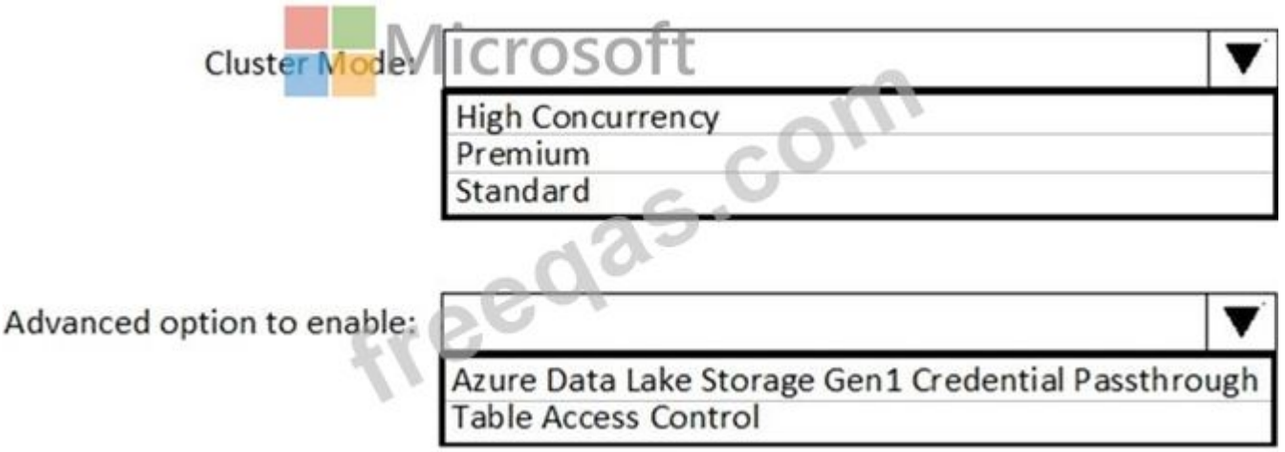
Answer:



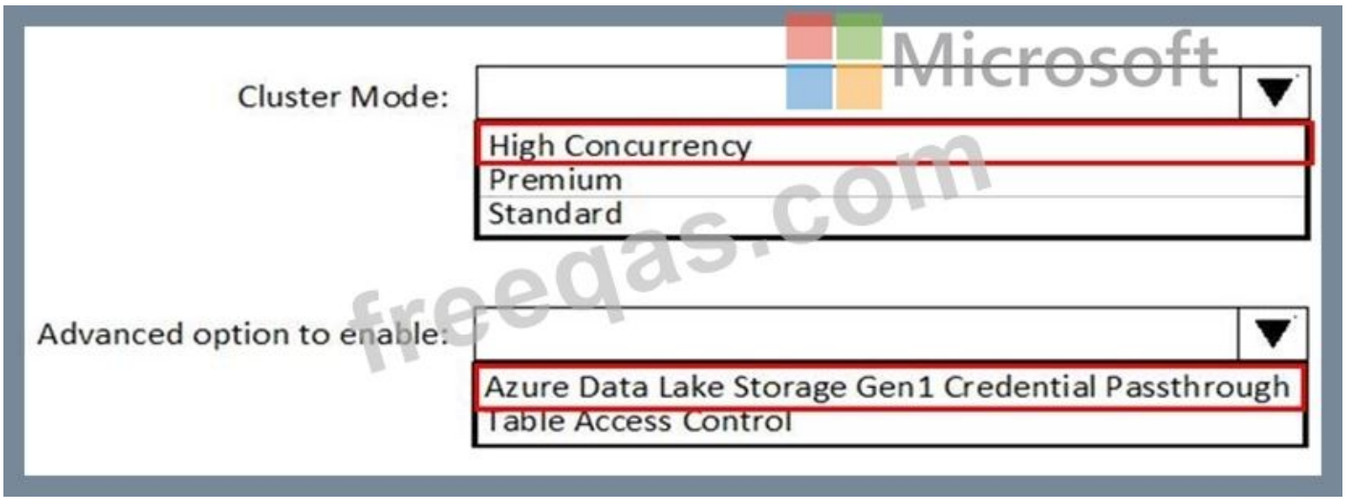
Reference:
https://www.datanami.com/2018/05/16/big-data-file-formats-demystified

NEW QUESTION: 111

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.



Answer:



Reference:
https://docs.azuredatabricks.net/spark/latest/data-sources/azure/adls-passthrough.html

NEW QUESTION: 112
You are designing an Azure Databricks table. The table will ingest an average of 20 million streaming events per day.
You need to persist the events in the table for use in incremental load pipeline jobs in Azure Databricks. The solution must minimize storage costs and incremental load times.
What should you include in the solution?
A. Partition by DateTime fields.
B. Sink to Azure Queue storage.
C. Include a watermark column.
D. Use a JSON format for physical data storage.
Answer: (SHOW ANSWER)

The Databricks ABS-AQS connector uses Azure Queue Storage (AQS) to provide an optimized file source that lets you find new files written to an Azure Blob storage (ABS) container without repeatedly listing all of the files.
This provides two major advantages:
Lower latency: no need to list nested directory structures on ABS, which is slow and resource intensive.
Lower costs: no more costly LIST API requests made to ABS.
Reference:
https://docs.microsoft.com/en-us/azure/databricks/spark/latest/structured-streaming/aqs

NEW QUESTION: 113
You are planning a streaming data solution that will use Azure Databricks. The solution will stream sales transaction data from an online store. The solution has the following specifications:
* The output data will contain items purchased, quantity, line total sales amount, and line total tax amount.
* Line total sales amount and line total tax amount will be aggregated in Databricks.
* Sales transactions will never be updated. Instead, new rows will be added to adjust a sale.
You need to recommend an output mode for the dataset that will be processed by using Structured Streaming.
The solution must minimize duplicate data.
What should you recommend?
A. Complete
B. Update
C. Append
Answer: (SHOW ANSWER)

NEW QUESTION: 114
You have an Azure Synapse Analytics dedicated SQL pool that contains the users shown in the following table.

Name	Role
User1	Server admin
User2	db_datereader

User1 executes a query on the database, and the query returns the results shown in the following exhibit.

```
SELECT c.name,
    tbl.name as table_name,
    typ.name as datatype,
    c.is_masked,
    c.masking_function
FROM sys.masked_columns AS c
INNER JOIN sys.tables AS tbl ON c.[object_id] = tbl.[object_id]
INNER JOIN sys.types typ ON c.user_type_id = typ.user_type_id
WHERE is_masked = 1;
```

Results Messages

	name	table_name	datatype	is_masked	masking_function
1	BirthDate	DimCustomer	date	1	default()
2	Gender	DimCustomer	nvarchar	1	default()
3	EmailAddress	DimCustomer	nvarchar	1	email()
4	YearlyIncome	DimCustomer	money	1	default()

User1 is the only user who has access to the unmasked data.

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.

When User2 queries the YearlyIncome column, the values returned will be **[answer choice]**.

- a random number
- the values stored in the database
- XXXX
- 0

When User1 queries the BirthDate column, the values returned will be **[answer choice]**.

- a random date
- the values stored in the database
- XXXX
- 1900-01-01

Answer:

When User2 queries the YearlyIncome column, the values returned will be **[answer choice]**.

- a random number
- the values stored in the database
- XXXX
- 0 ✓

When User1 queries the BirthDate column, the values returned will be **[answer choice]**.

- a random date
- the values stored in the database ✓
- XXXX
- 1900-01-01

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

NEW QUESTION: 115
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 are a.
NOTE: Each correct selection is worth one point.

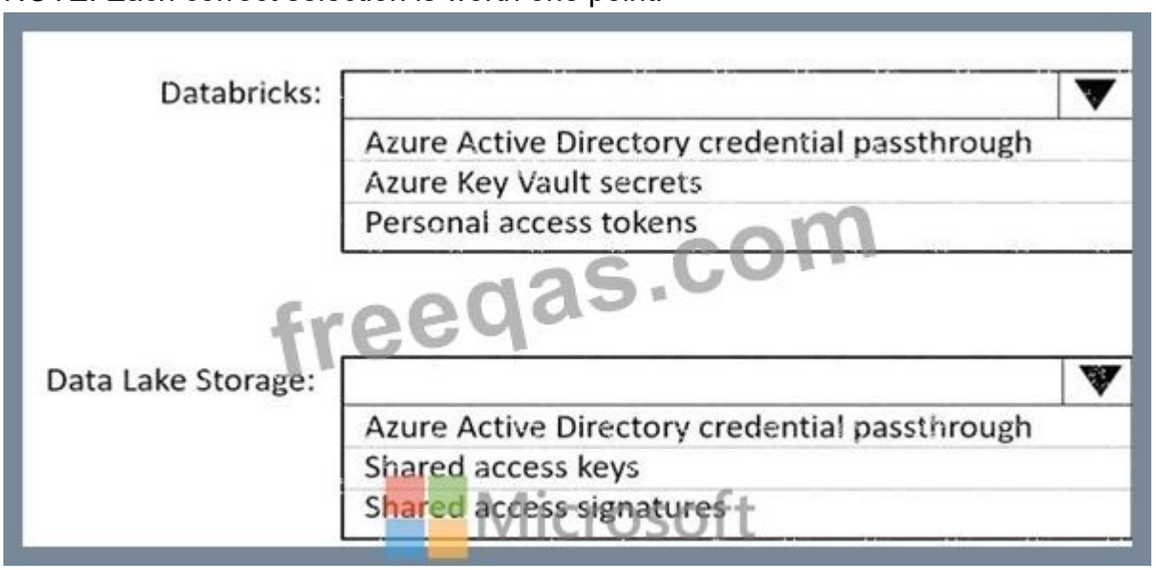
```
SELECT [ ▼ ] as HighestScore
FROM input TIMESTAMP BY CreatedAt
GROUP BY [ ▼ ]
```

- SELECT: Collect(Score) / CollectTop(1) OVER(ORDER BY Score Desc) ✓ / Game, MAX(Score) / TopOne() OVER(PARTITION BY Game ORDER BY Score Desc)
- GROUP BY: Game / Hopping(minute,5) / Tumbling(minute,5) ✓ / Windows(TumblingWindow(minute,5),Hopping(minute,5))

Answer:

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: 116
You use Azure Data Lake Storage Gen2 to store data that data scientists and data engineers will query by using Azure Databricks interactive notebooks. Users will have access only to the Data Lake Storage folders that relate to the projects on which they work.
You need to recommend which authentication methods to use for Databricks and Data Lake Storage to provide the users with the appropriate access. The solution must minimize administrative effort and development effort.
Which authentication method should you recommend for each Azure service? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.



Answer:

Explanation

Table Description automatically generated

Box 1: Personal access tokens

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

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

Box 2: Azure Active Directory credential passthrough

You can authenticate automatically to Azure Data Lake Storage Gen1 (ADLS Gen1) and Azure Data Lake Storage Gen2 (ADLS Gen2) from Azure Databricks clusters using the same Azure Active Directory (Azure AD) identity that you use to log into Azure Databricks. When you enable your cluster for Azure Data Lake Storage credential passthrough, commands that you run on that cluster can read and write data in Azure Data Lake Storage without requiring you to configure service principal credentials for access to storage.

After configuring Azure Data Lake Storage credential passthrough and creating storage containers, you can access data directly in Azure Data Lake Storage Gen1 using an adl:// path and Azure Data Lake Storage Gen2 using an abfss:// path:
Reference:
https://docs.microsoft.com/en-us/azure/databricks/data/data-sources/azure/adls-gen2/azure-datalake-gen2-sas-acc
https://docs.microsoft.com/en-us/azure/databricks/security/credential-passthrough/adls-passthrough

NEW QUESTION: 117
You have the following table named Employees.

first_name	last_name	hire_date	employee_type
Jane	Doe	2019-08-23	new
Ben	Smith	2017-12-15	Standard

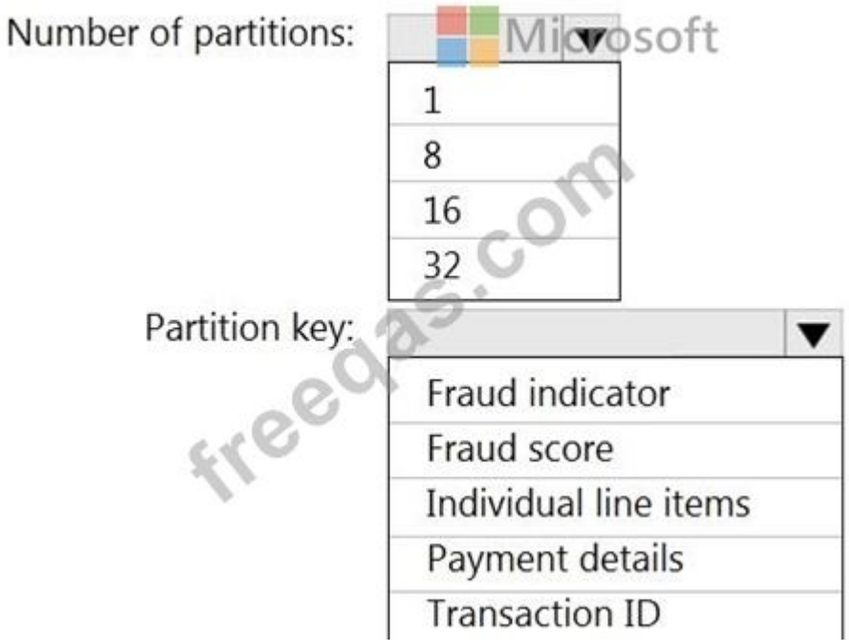
You need to calculate the employee _type value based on the hire date value.
How should you complete the Transact-SQL statement? To answer, drag the appropriate values to the correct targets. Each value may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content- NOTE: Each correct selection is worth one point.

Answer:

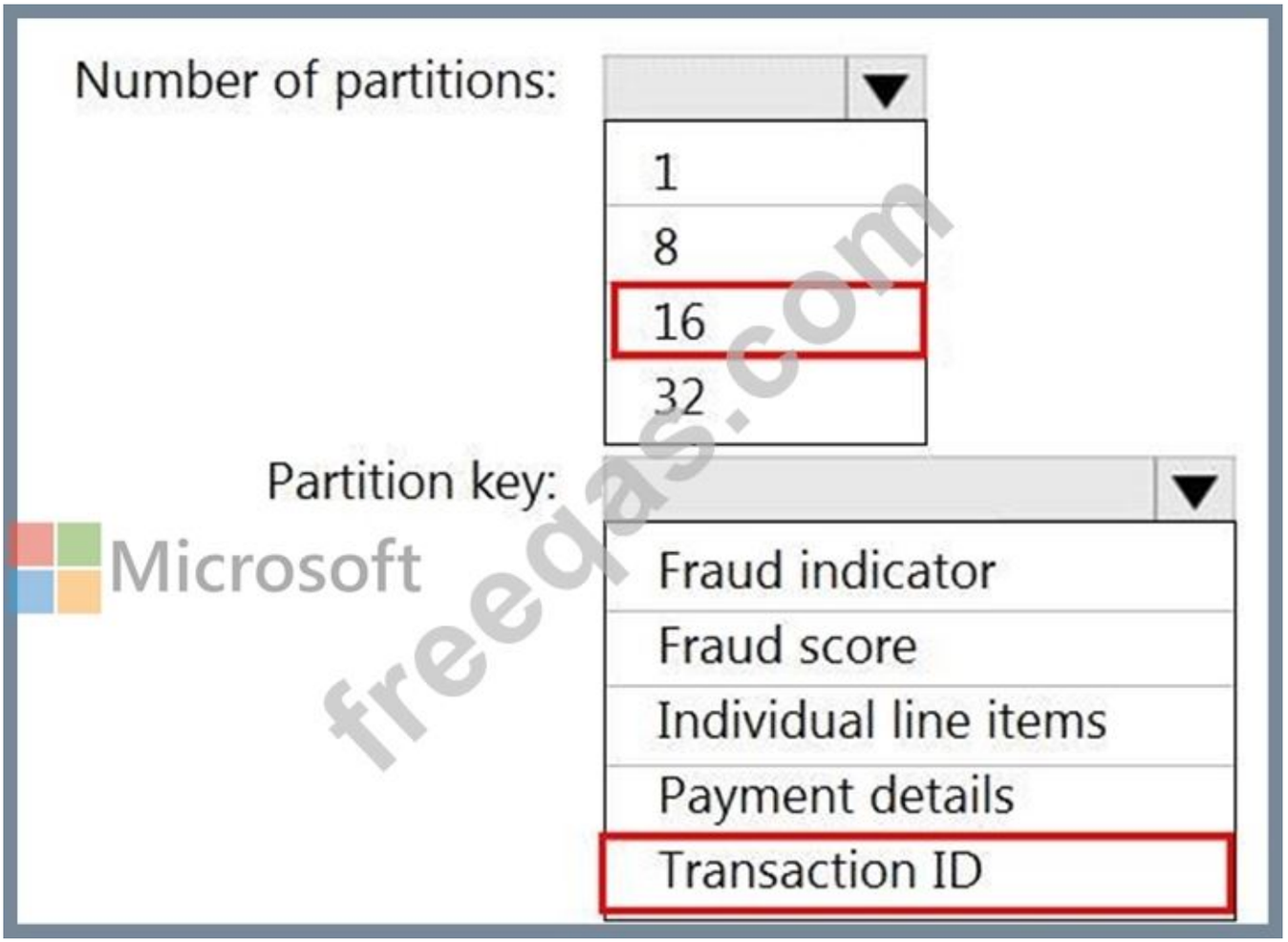
NEW QUESTION: 118
You are designing a date dimension table in an Azure Synapse Analytics dedicated SQL pool. The date dimension table will be used by all the fact tables.
Which distribution type should you recommend to minimize data movement?
A. HASH
B. REPLICATE
C. ROUND ROBIN
Answer: (SHOW ANSWER)

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: 119
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.



Answer:



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

NEW QUESTION: 120
You have a SQL pool in Azure Synapse that contains a table named dbo.Customers. The table contains a column name 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: A (LEAVE A REPLY)
Explanation
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,"XXXXXXX",0)') NULL, . .
https://docs.microsoft.com/en-us/azure/azure-sql/database/dynamic-data-masking-overview upvoted 24 times

NEW QUESTION: 121
From a website analytics system, you receive data extracts about user interactions such as downloads, link clicks, form submissions, and video plays.
The data contains the following columns.

Name	Sample value
Date	15 Jan 2021
EventCategory	Videos
EventAction	Play
EventLabel	Contoso Promotional
ChannelGrouping	Social
TotalEvents	150
UniqueEvents	120
SessionWithEvents	99

You need to design a star schema to support analytical queries of the data. The star schema will contain four tables including a date dimension.
To which table should you add each column? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Answer:

Reference:

https://docs.microsoft.com/en-us/power-bi/guidance/star-schema

NEW QUESTION: 122

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.

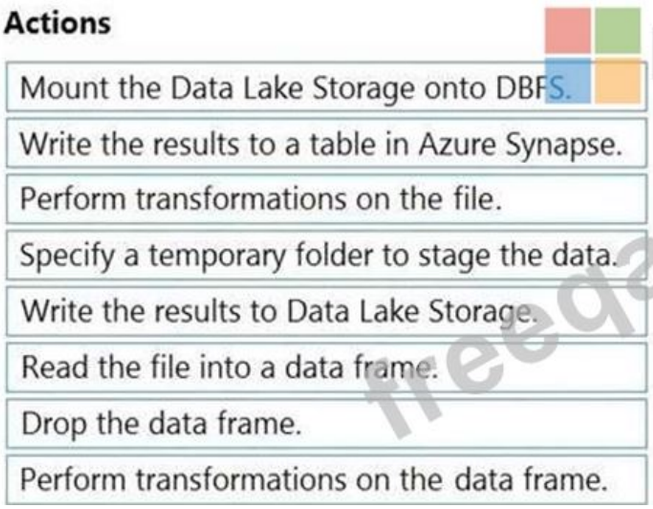
You need to copy the data from the JSON file to an Azure Synapse Analytics table by using Azure Databricks.

A new column must be created that concatenates the FirstName and LastName values.

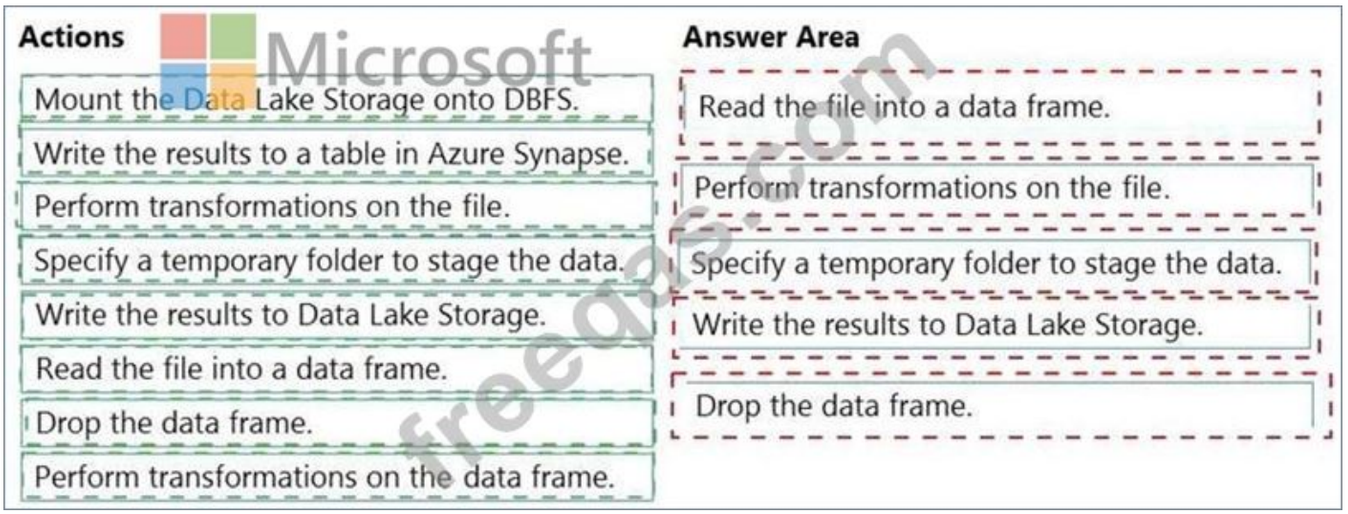
You create the following components:

* A destination table in Azure Synapse
* An Azure Blob storage container
* A service principal

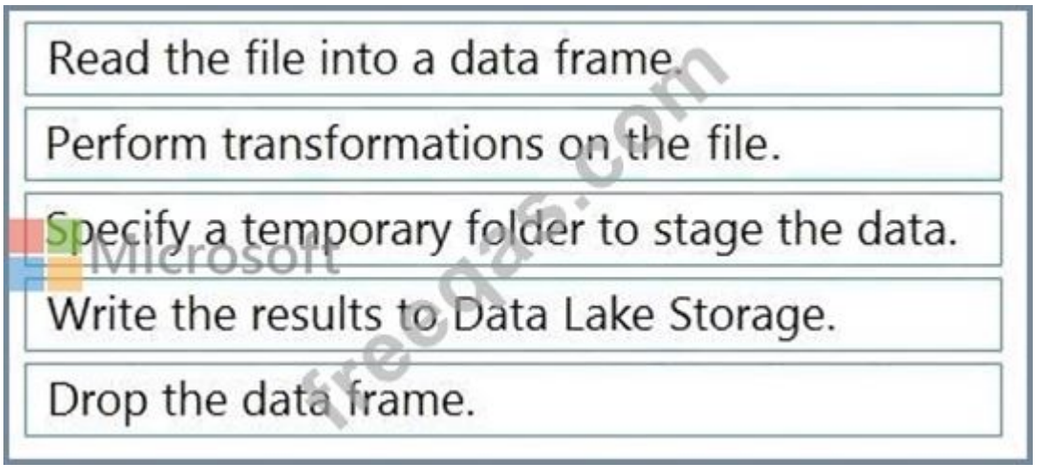
Which five actions should you perform in sequence next in is Databricks notebook? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



Answer:



Explanation



Step 1: Read the file into a data frame.

You can load the json files as a data frame in Azure Databricks.

Step 2: Perform transformations on the data frame.

Step 3:Specify a temporary folder to stage the data

Specify a temporary folder to use while moving data between Azure Databricks and Azure Synapse.

Step 4: 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.

Step 5: Drop the data frame

Clean up resources. You can terminate the cluster. From the Azure Databricks workspace, select Clusters on the left. For the cluster to terminate, under Actions, point to the ellipsis (...) and select the Terminate icon.

Reference:

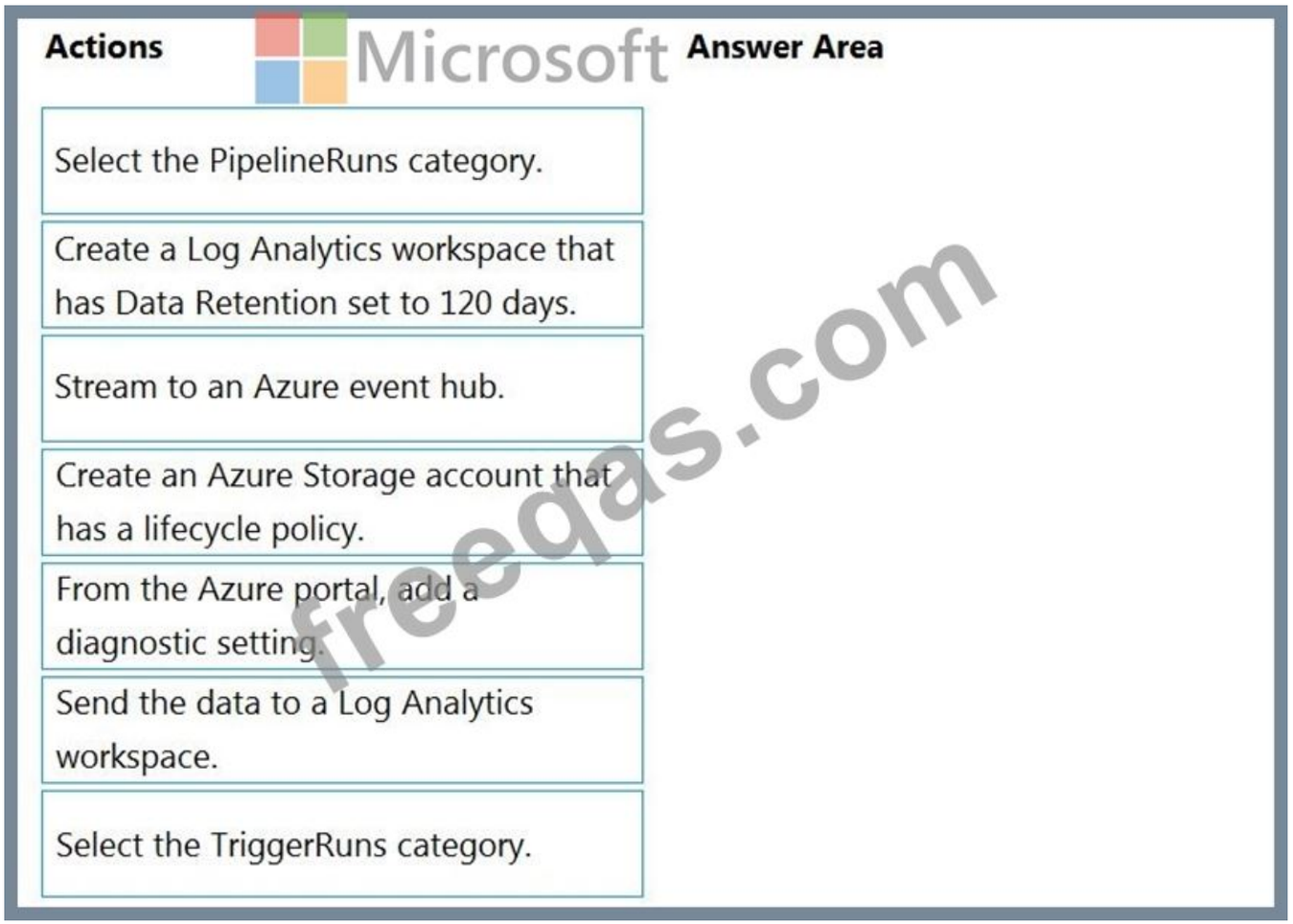
https://docs.microsoft.com/en-us/azure/azure-databricks/databricks-extract-load-sql-data-warehouse

NEW QUESTION: 123

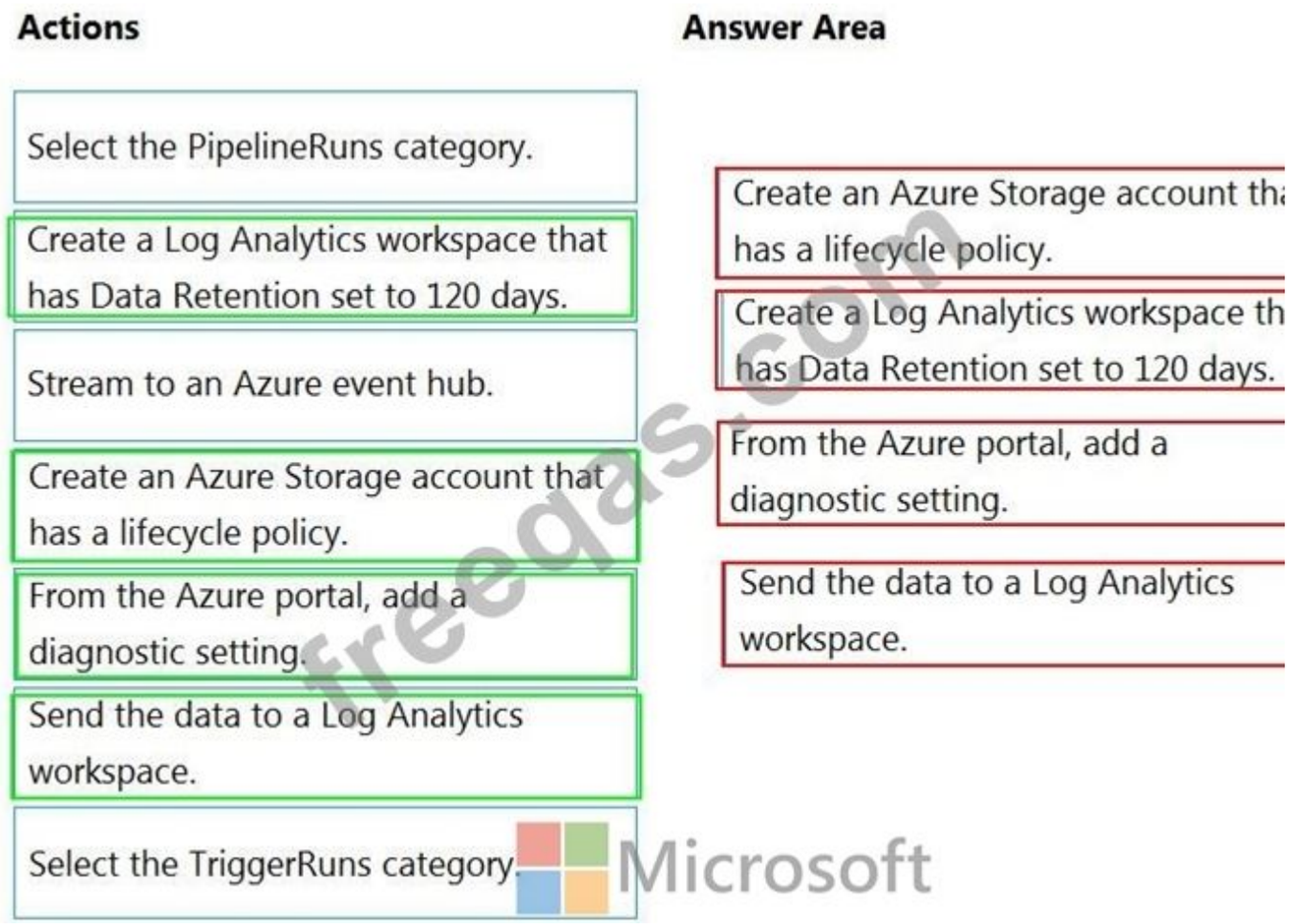
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

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 dat a.
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 session window that uses a timeout size of 10 seconds.
Does this meet the goal?
A. Yes
B. No
Answer: B (LEAVE A REPLY)
Instead use a tumbling window. Tumbling windows are a series of fixed-sized, non-overlapping and contiguous time intervals.
Reference:
https://docs.microsoft.com/en-us/stream-analytics-query/tumbling-window-azure-stream-analytics

NEW QUESTION: 124
You have an Azure data factory.
You need to ensure that pipeline-run data is retained for 120 days. The solution must ensure that you can query the data by using the Kusto query language.
Which four actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.
NOTE: More than one order of answer choices is correct. You will receive credit for any of the correct orders you select.



Answer:



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

NEW QUESTION: 125

You are designing database for an Azure Synapse Analytics dedicated SQL pool to support workloads for detecting ecommerce transaction fraud.

Data will be combined from multiple ecommerce sites and can include sensitive financial information such as credit card numbers.

You need to recommend a solution that meets the following requirements:

* Users must be able to identify potentially fraudulent transactions.
* Users must be able to use credit cards as a potential feature in models.
* Users must NOT be able to access the actual credit card numbers.

What should you include in the recommendation?

A. column-level encryption
B. row-level security (RLS)
C. Azure Active Directory (Azure AD) pass-through authentication
D. Transparent Data Encryption (TDE)

Answer: A (LEAVE A REPLY)

NEW QUESTION: 126

You have an enterprise data warehouse in Azure Synapse Analytics that contains a table named FactOnlineSales. The table contains data from the start of 2009 to the end of 2012.

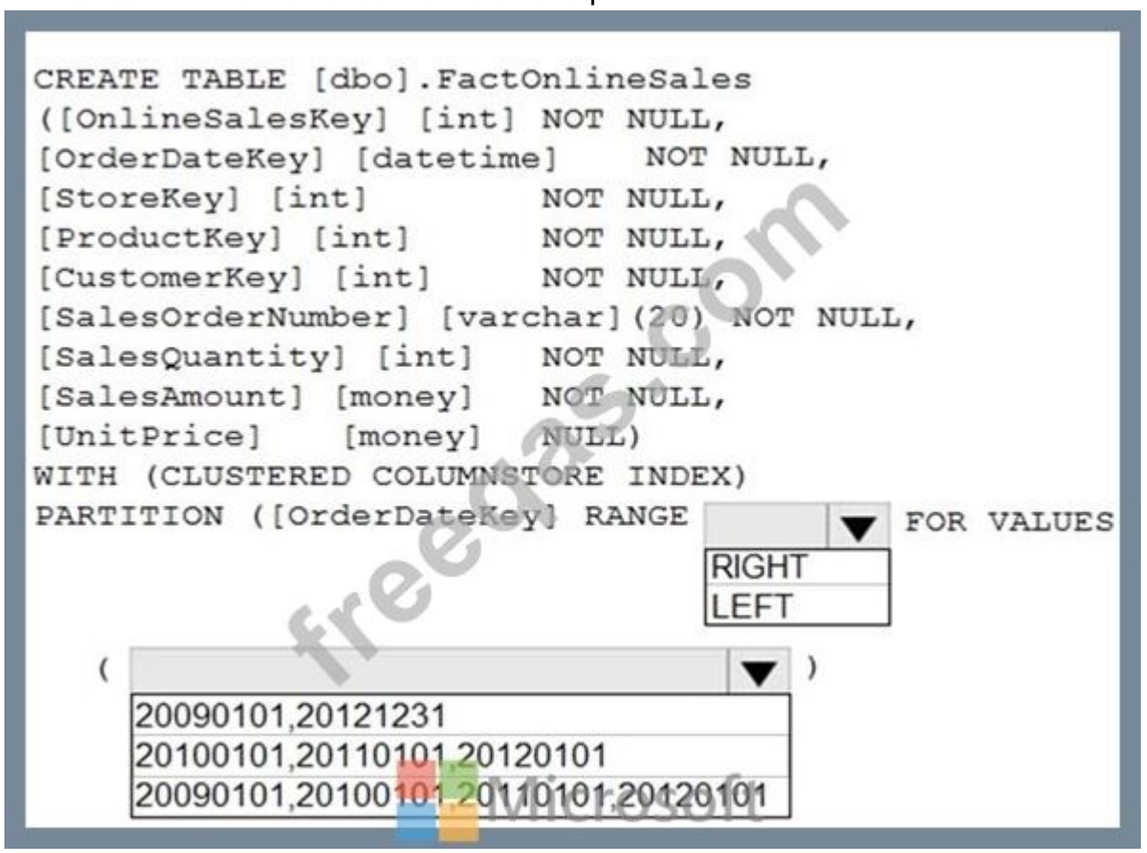
You need to improve the performance of queries against FactOnlineSales by using table partitions. The solution must meet the following requirements:

Create four partitions based on the order date.

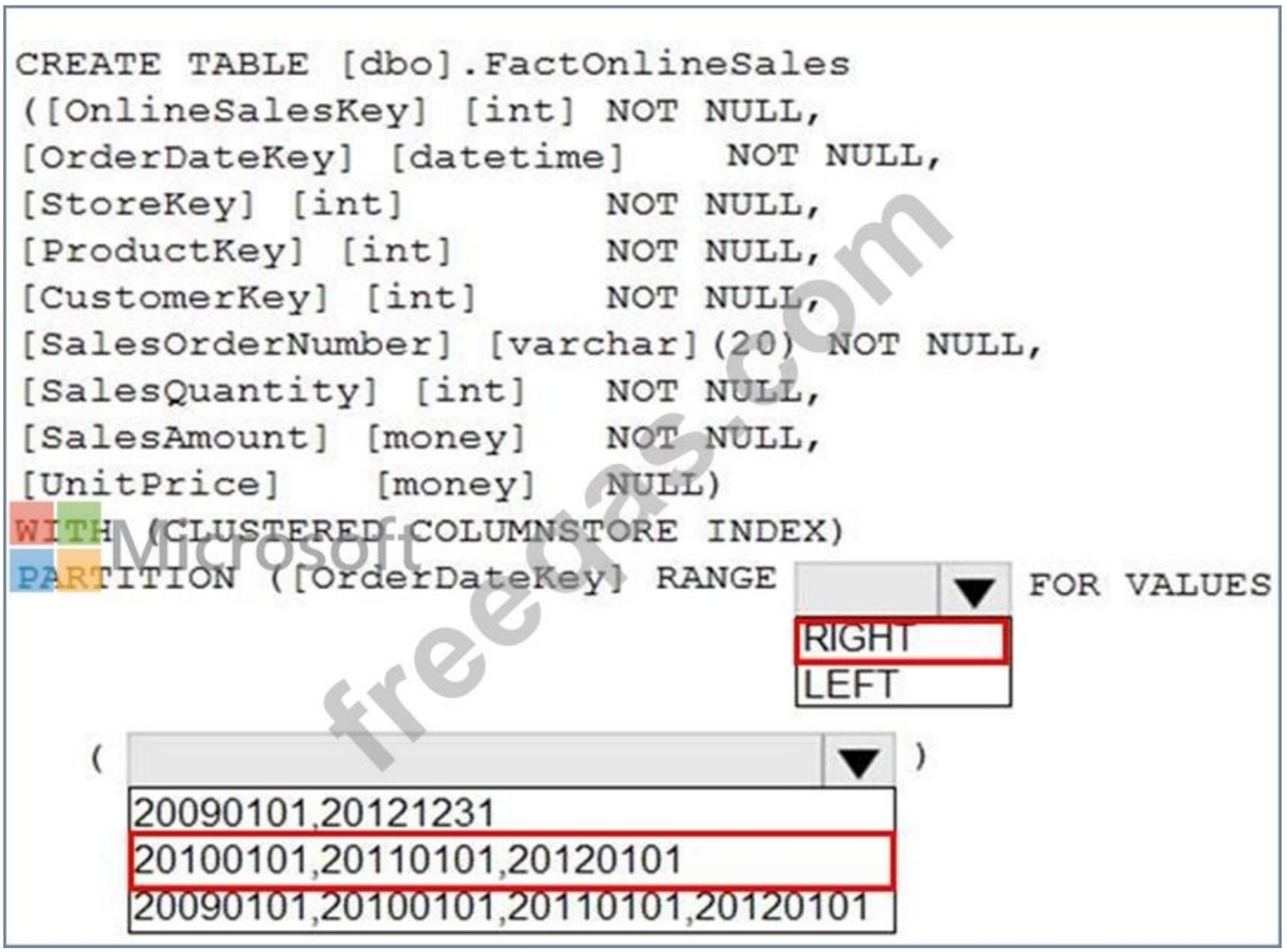
Ensure that each partition contains all the orders places during a given calendar year.

How should you complete the T-SQL command? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.



Answer:

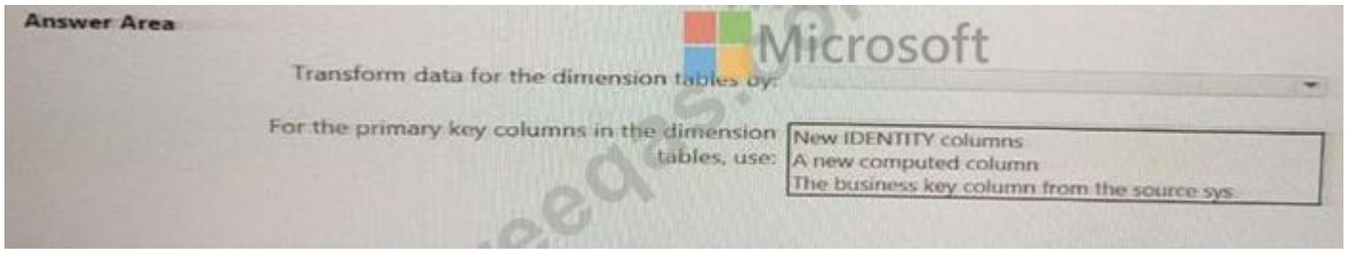


Reference:
https://docs.microsoft.com/en-us/sql/t-sql/statements/create-partition-function-transact-sql?view=sql-server-ver15

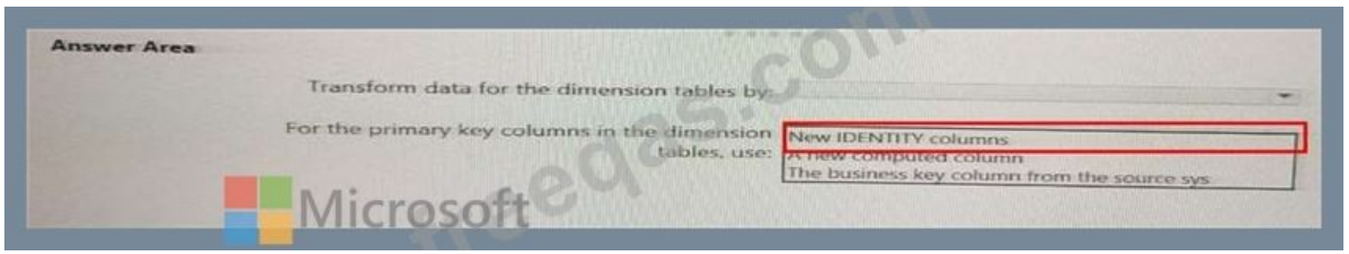
NEW QUESTION: 127
You plan to ingest streaming social media data by using Azure Stream Analytics. The data will be stored in files in Azure Data Lake Storage, and then consumed by using Azure Datiabricks and PolyBase in Azure Synapse Analytics.
You need to recommend a Stream Analytics data output format to ensure that the queries from Databricks and PolyBase against the files encounter the fewest possible errors. The solution must ensure that the tiles can be queried quickly and that the data type information is retained.
What should you recommend?
A. Parquet
B. Avro
C. CSV
D. JSON
Answer: (SHOW ANSWER)
Explanation

The Avro format is great for data and message preservation.Avro schema with its support for evolution is essential for making the data robust for streaming architectures like Kafka, and with the metadata that schema provides, you can reason on the data. Having a schema provides robustness in providing meta-data about the data stored in Avro records which are self- documenting the data.References: http://cloudurable.com/blog/avro/index.html

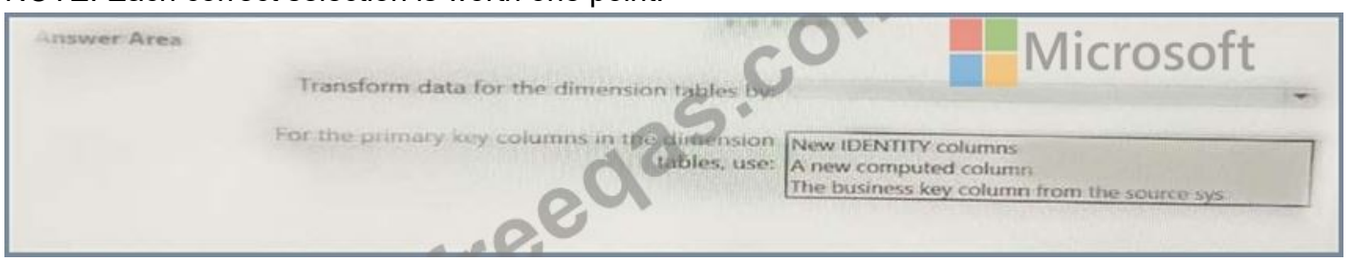
NEW QUESTION: 128
You have a Microsoft SQL Server database that uses a third normal form schema.
You plan to migrate the data in the database to a star schema in an A?\ire Synapse Analytics dedicated SQI pool.
You need to design the dimension tables. The solution must optimize read operations.
What should you include in the solution? to answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.



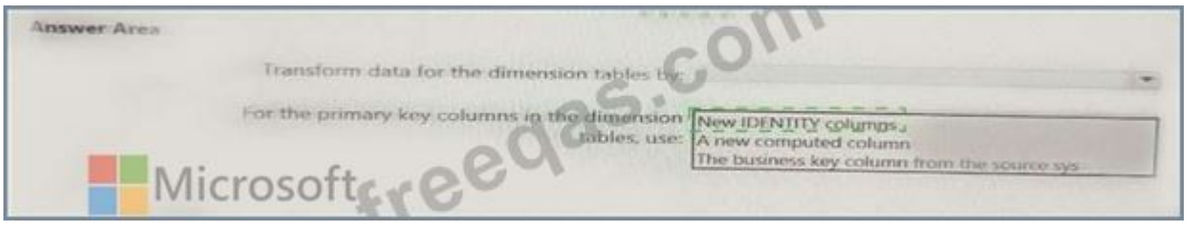
Answer:



NEW QUESTION: 129
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 A?\ire Synapse Analytics dedicated SQI 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.



Answer:



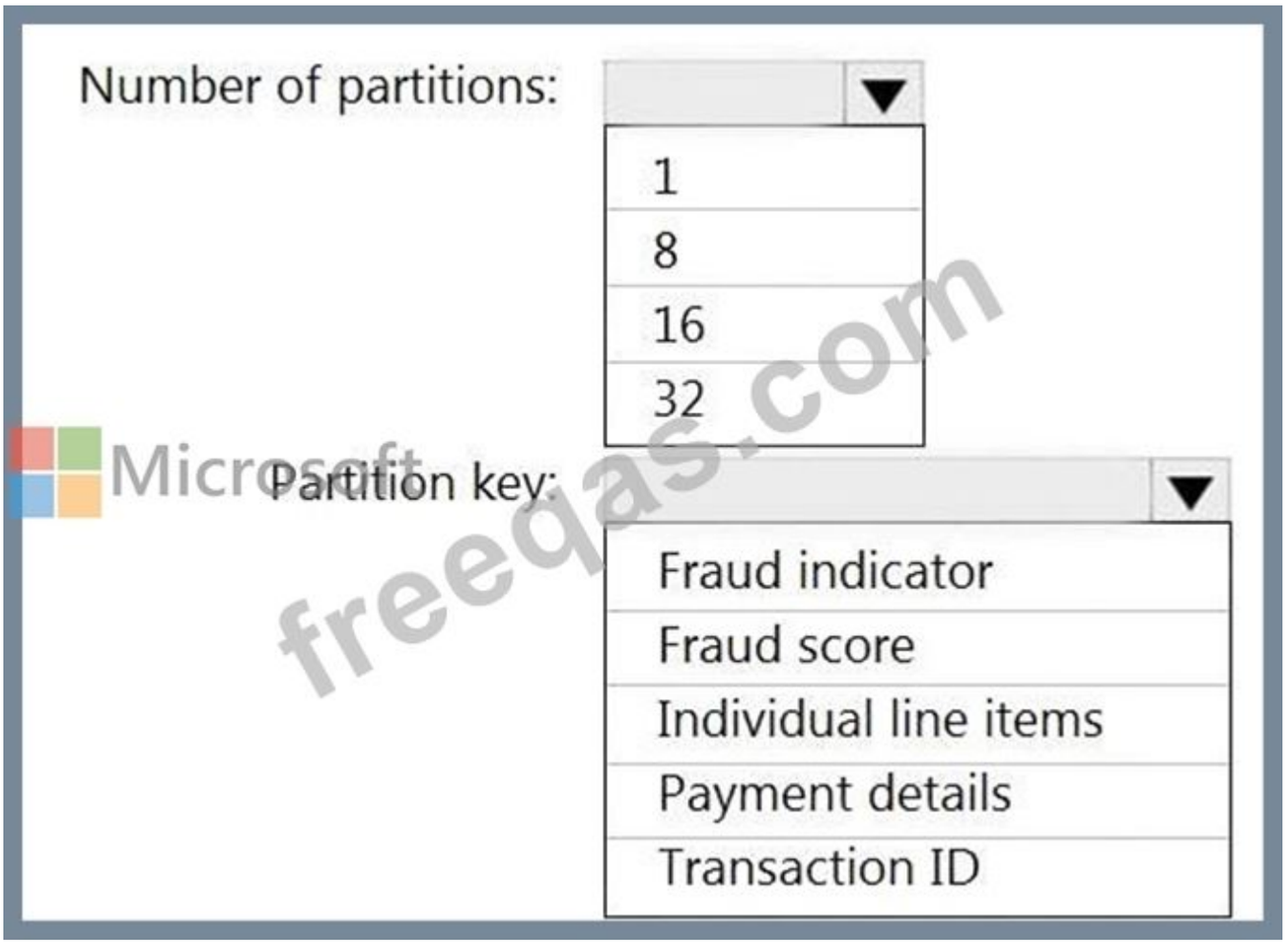
NEW QUESTION: 130
You have the following Azure Data Factory pipelines
* ingest Data from System 1
* Ingest Data from System2
* Populate Dimensions
* Populate facts
ingest Data from System1 and Ingest Data from System1 have no dependencies. Populate Dimensions must execute after Ingest Data from System1 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: 131
You plan to create an Azure Data Factory pipeline that will include a mapping data flow.
You have JSON data containing objects that have nested arrays.
You need to transform the JSON-formatted data into a tabular dataset. The dataset must have one tow for each item in the arrays.
Which transformation method should you use in the mapping data flow?
A. new branch
B. flatten
C. unpivot
D. alter row
Answer: (SHOW ANSWER)

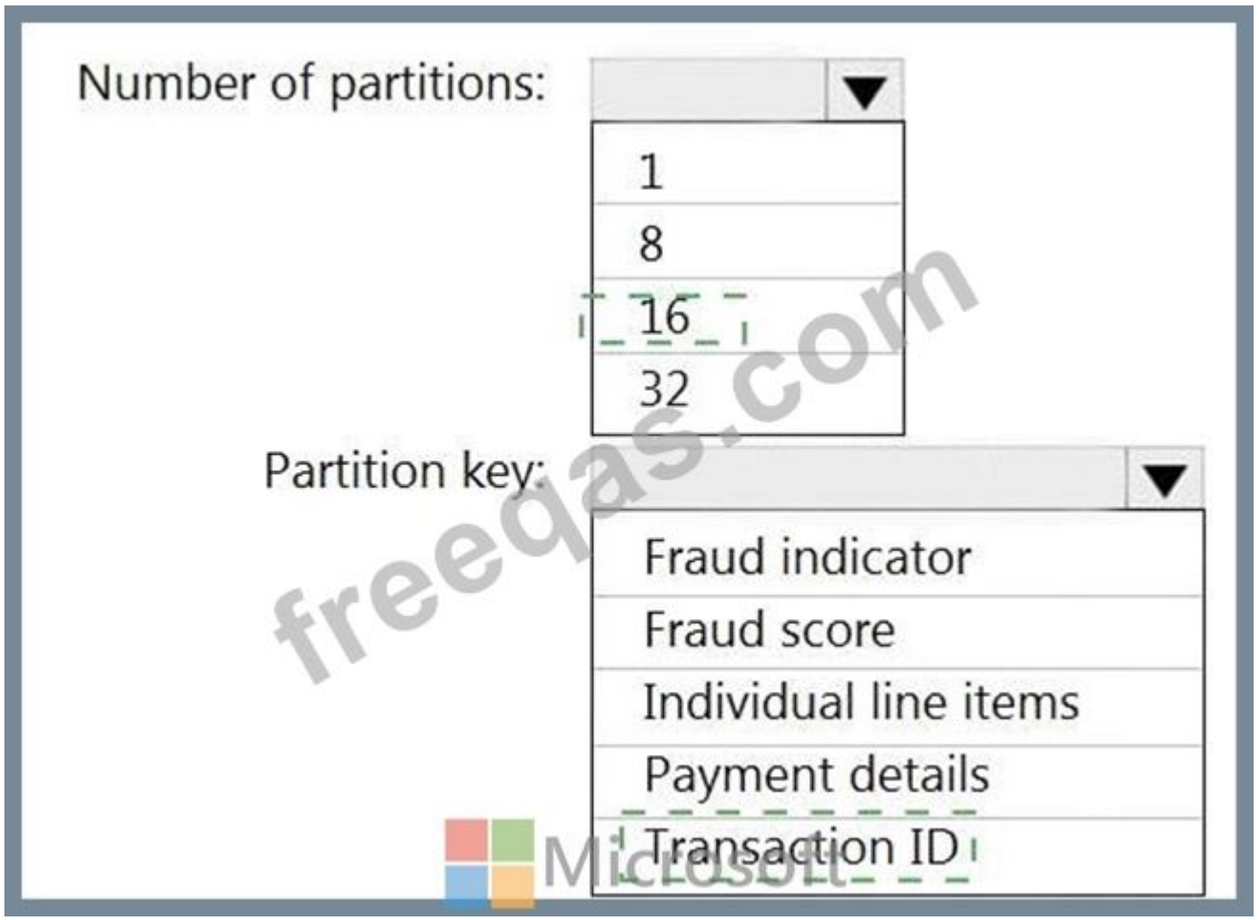
NEW QUESTION: 132
What should you recommend to prevent users outside the Litware on-premises network from accessing the analytical data store?
A. a server-level virtual network rule

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

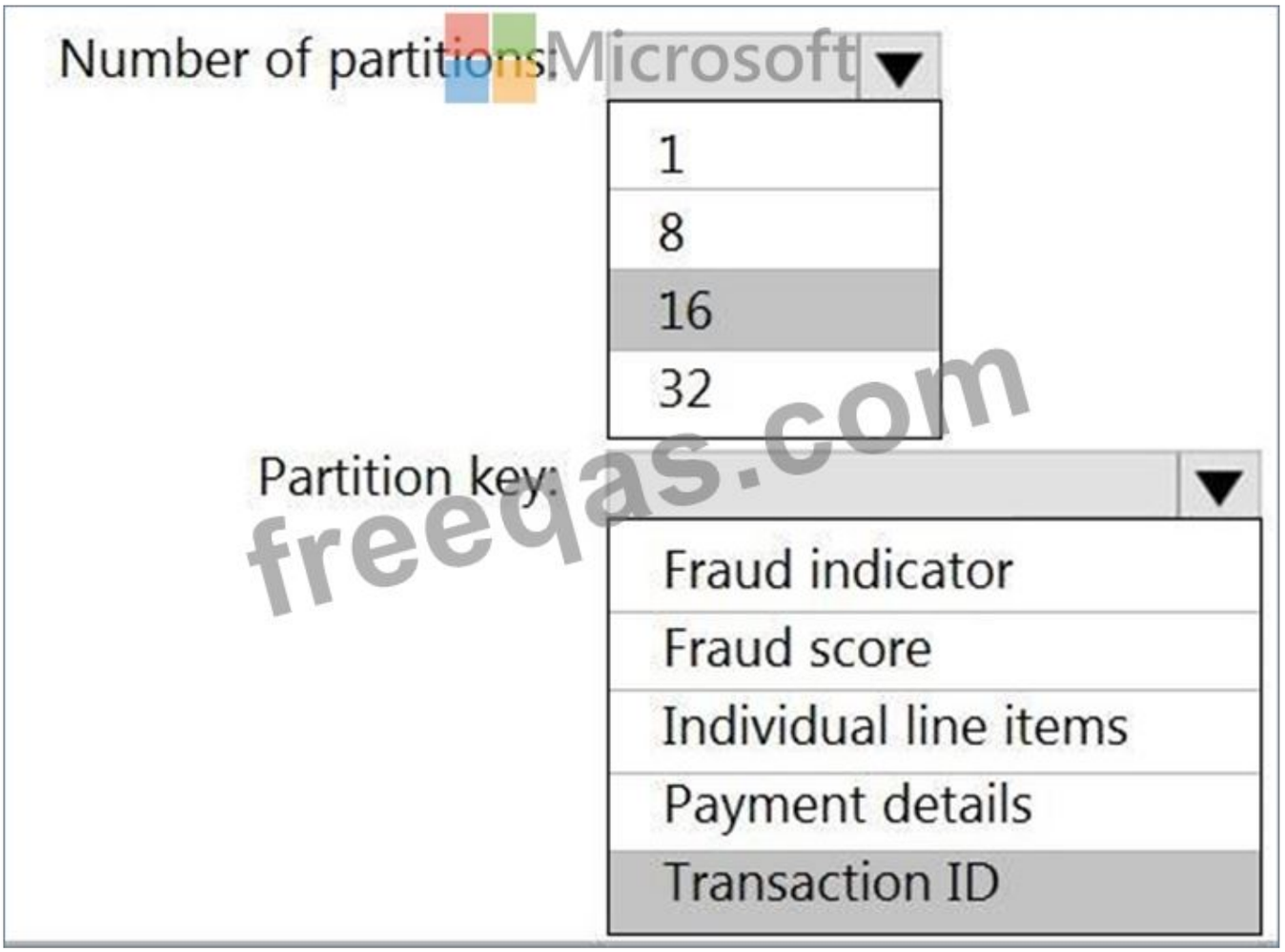
NEW QUESTION: 133
You have an Azure event hub named retailhub that has 16 partitions. Transactions are posted to retailhub. Each transaction includes the transaction ID, the individual line items, and the payment details. The transaction ID is used as the partition key.
You are designing an Azure Stream 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.



Answer:



Explanation



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: 134

You need to design an analytical storage solution for the transactional dat a. 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.

Answer:

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

NEW QUESTION: 135

You have an Azure Synapse Analytics dedicated SQL pool named Pool1 and an Azure Data Lake Storage Gen2 account named Account 1.

You plan to access the files in Accoun1l by using an external table.

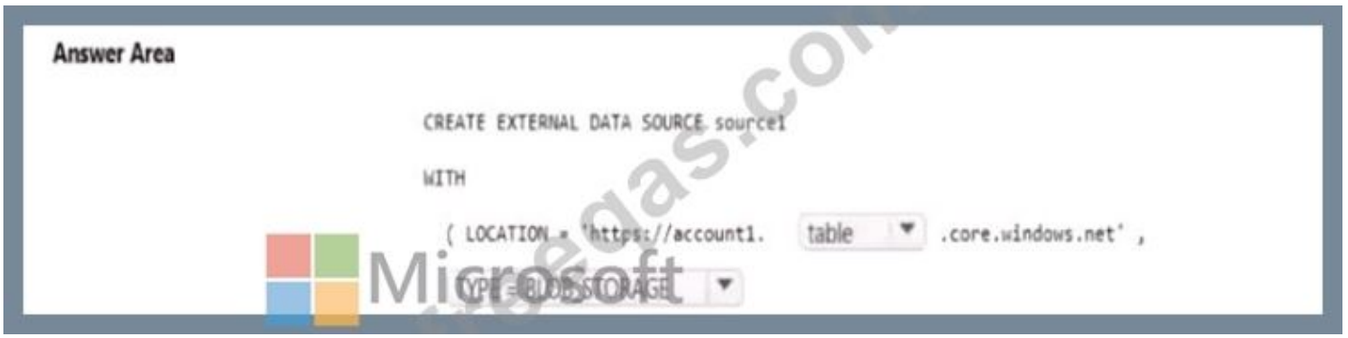
You need to create a data source in Pool1 that you can reference when you create the external table.

How should you complete the Transact-SQL statement? To answer, select the appropriate options in the answer area.

NOTE Each coned selection is worth one point.

Answer:

Answer as below



NEW QUESTION: 136
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 from the last five minutes every minute.
Which windowing function should you use?
A. Sliding
B. Session
C. Tumbling
D. Hopping
Answer: D (LEAVE A REPLY)
Explanation
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.
Reference:
https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-window-functions

NEW QUESTION: 137
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.

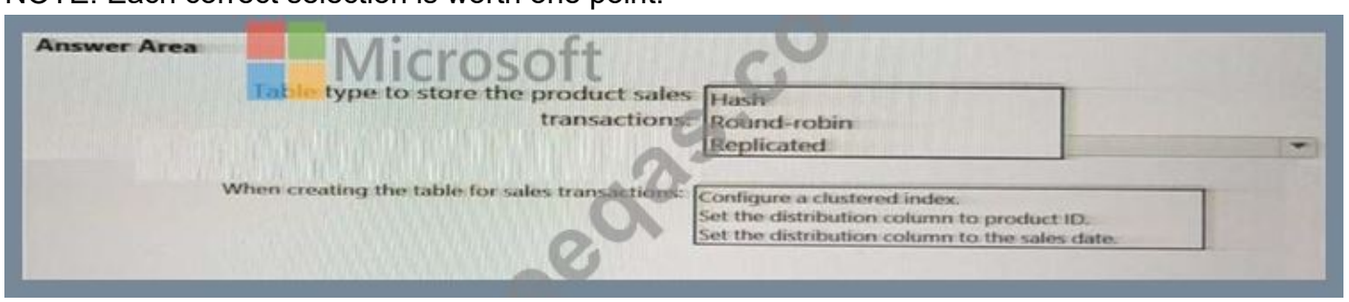
You plan to insert data from the files into Table1 and transform the data. Each row of data in the files will produce one row in the serving layer of Table1.
You need to ensure that when the source data files are loaded to container1, the DateTime is stored as an additional column in Table1.
Solution: In an Azure Synapse Analytics pipeline, you use a data flow that contains a Derived Column transformation.
A. Yes
B. No
Answer: A (LEAVE A REPLY)
Explanation
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: 138
You have two fact tables named Flight and Weather. Queries targeting the tables will be based on the join between the following columns.

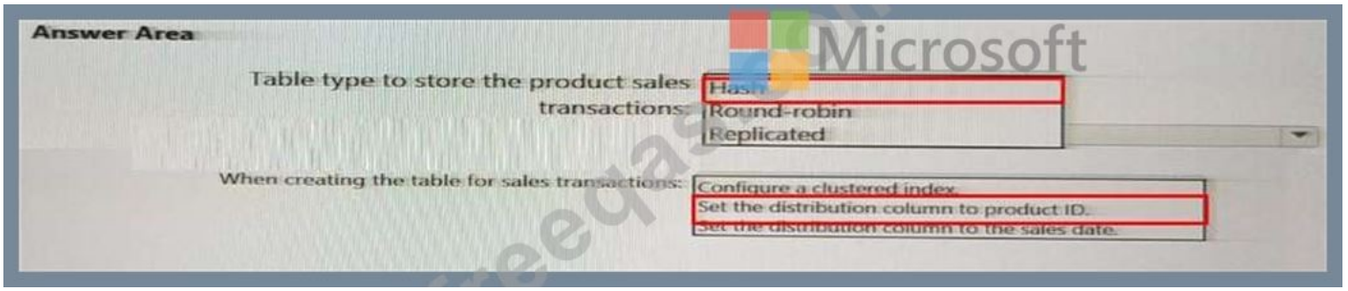
Table	Column
Flight	ArrivalAirportID ArrivalDateTime
Weather	AirportID ReportDateTime

You need to recommend a solution that maximum query performance.
What should you include in the recommendation?
A. In each table, create an IDENTITY column.
B. In each table, create a column as a composite of the other two columns in the table.
C. In the tables, use a hash distribution of ArriveDateTime and ReportDateTime.
D. In the tables, use a hash distribution of ArriveAirPortID and AirportID.
Answer: (SHOW ANSWER)

NEW QUESTION: 139
You need to design a data storage structure for the product sales transactions. 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.



Answer:



Topic 2, Litware, inc.

Requirements

Business Goals

Litware wants to create a new analytics environment in Azure to meet the following requirements:

See inventory levels across the stores. Data must be updated as close to real time as possible.

Execute ad hoc analytical queries on historical data to identify whether the loyalty club discounts increase sales of the discounted products.

Every four hours, notify store employees about how many prepared food items to produce based on historical demand from the sales data.

Technical Requirements

Litware identifies the following technical requirements:

Minimize the number of different Azure services needed to achieve the business goals.

Use platform as a service (PaaS) offerings whenever possible and avoid having to provision virtual machines that must be managed by Litware.

Ensure that the analytical data store is accessible only to the company's on-premises network and Azure services.

Use Azure Active Directory (Azure AD) authentication whenever possible.

Use the principle of least privilege when designing security.

Stage Inventory data in Azure Data Lake Storage Gen2 before loading the data into the analytical data store. Litware wants to remove transient data from Data Lake Storage once the data is no longer in use. Files that have a modified date that is older than 14 days must be removed.

Limit the business analysts' access to customer contact information, such as phone numbers, because this type of data is not analytically relevant.

Ensure that you can quickly restore a copy of the analytical data store within one hour in the event of corruption or accidental deletion.

Planned Environment

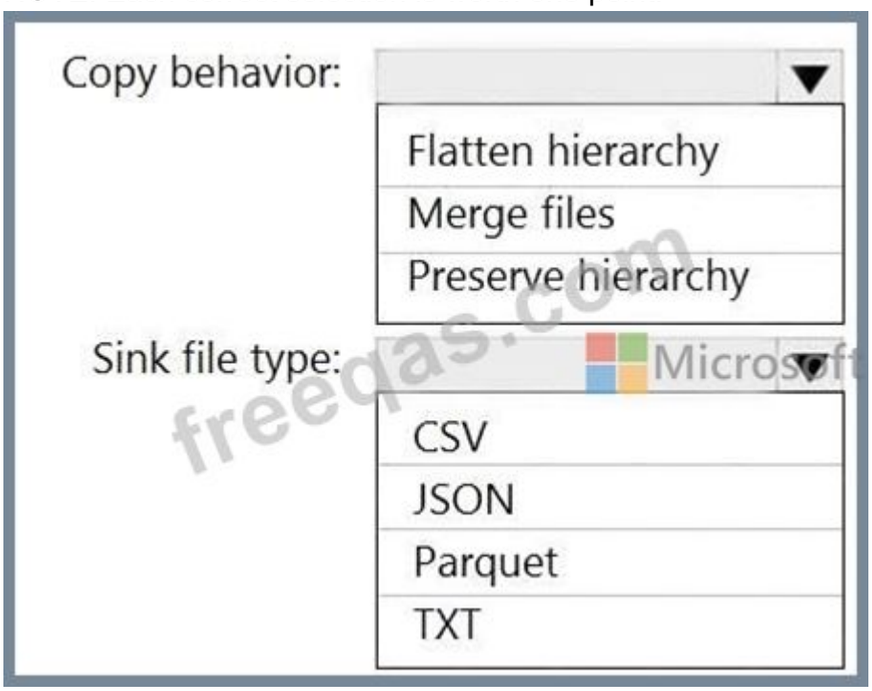
Litware plans to implement the following environment:

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.

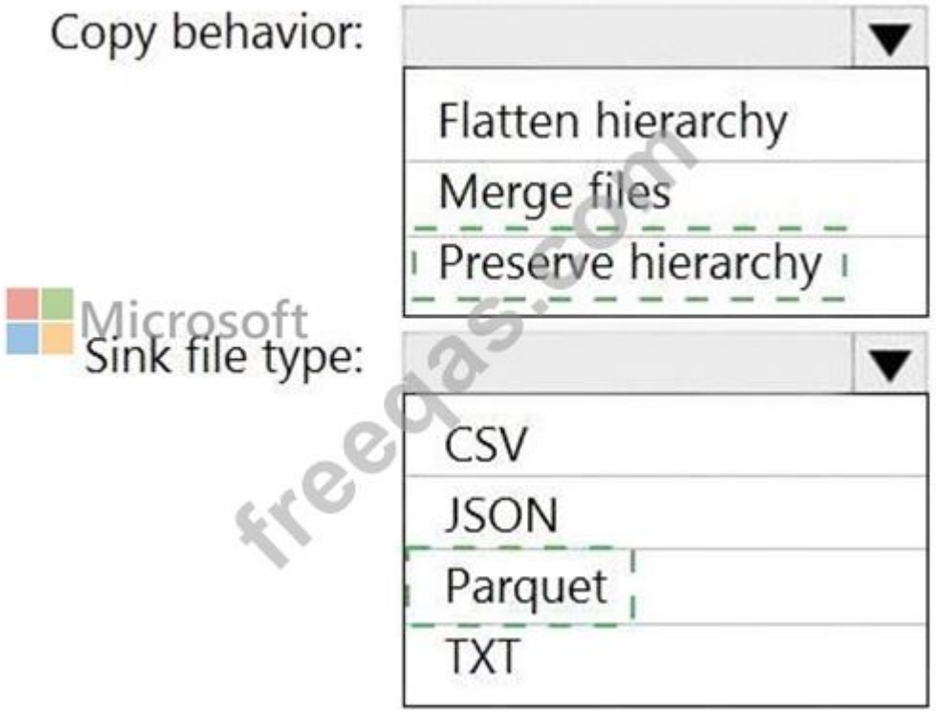
Customer data, including name, contact information, and loyalty number, comes from Salesforce, a SaaS application, and can be imported into Azure once every eight hours. Row modified dates are not trusted in the source table.

Product data, including product ID, name, and category, comes from Salesforce and can be imported into Azure once every eight hours. Row modified dates are not trusted in the source table.
Daily inventory data comes from a Microsoft SQL server located on a private network.
Litware currently has 5 TB of historical sales data and 100 GB of customer data. The company expects approximately 100 GB of new data per month for the next year.
Litware will build a custom application named FoodPrep to provide store employees with the calculation results of how many prepared food items to produce every four hours.
Litware does not plan to implement Azure ExpressRoute or a VPN between the on-premises network and Azure.

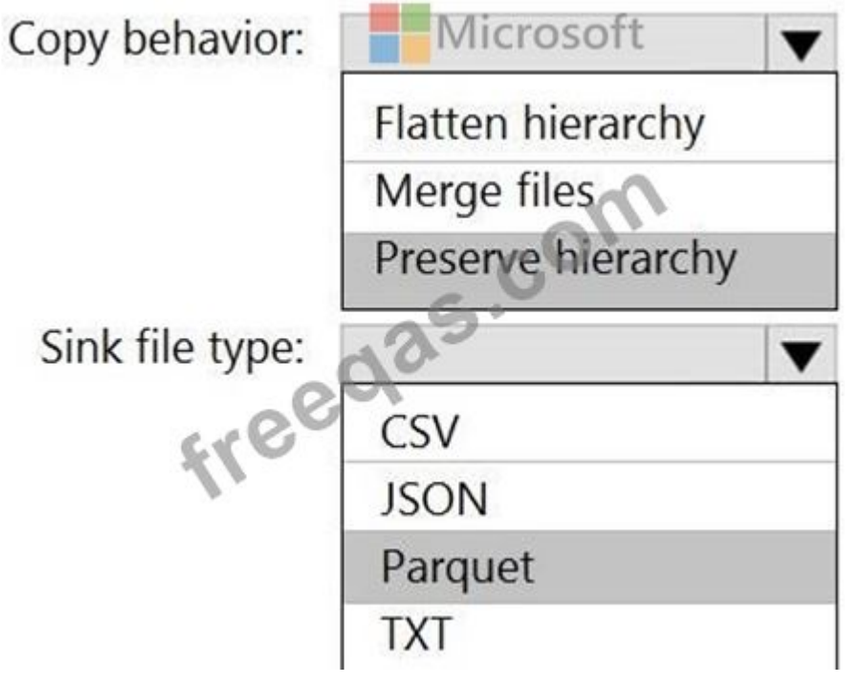
NEW QUESTION: 140
You use Azure Data Factory to prepare data to be queried by Azure Synapse Analytics serverless SQL pools.
Files are initially ingested into an Azure Data Lake Storage Gen2 account as 10 small JSON files. Each file contains the same data attributes and data from a subsidiary of your company.
You need to move the files to a different folder and transform the data to meet the following requirements:
* Provide the fastest possible query times.
* Automatically infer the schema from the underlying files.
How should you configure the Data Factory copy activity? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.



Answer:



Explanation



Box 1: Preserver herarchy

Compared to the flat namespace on Blob storage, the hierarchical namespace greatly improves the performance of directory management operations, which improves overall job performance.

Box 2: Parquet

Azure Data Factory parquet format is supported for Azure Data Lake Storage Gen2.

Parquet supports the schema property.

Reference:

https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-introduction

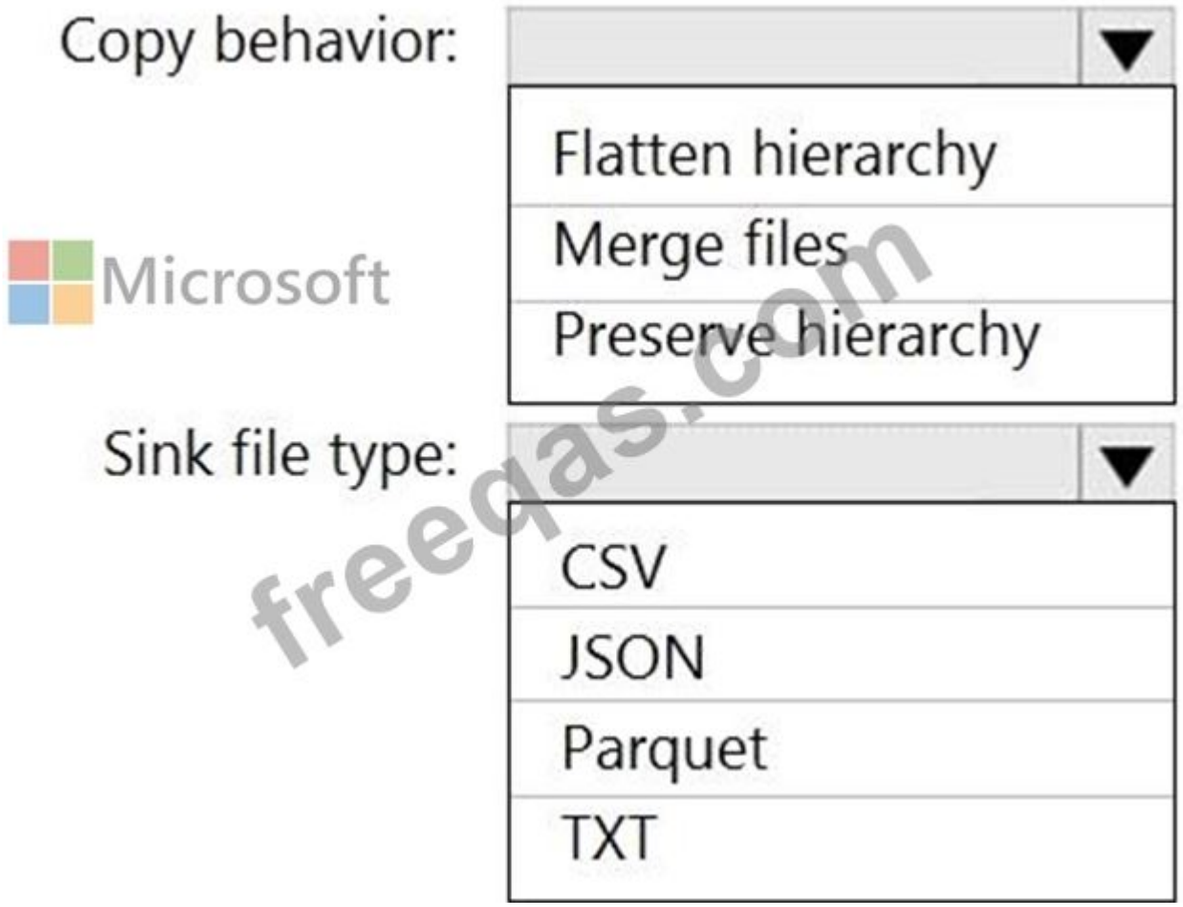
https://docs.microsoft.com/en-us/azure/data-factory/format-parquet

NEW QUESTION: 141
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 OPENROWEST to query the Parquet files.
C. Create an external table mat contains a subset of columns from the Parquet files.
D. Use Snappy compression for the files.
Answer: (SHOW ANSWER)
Explanation
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

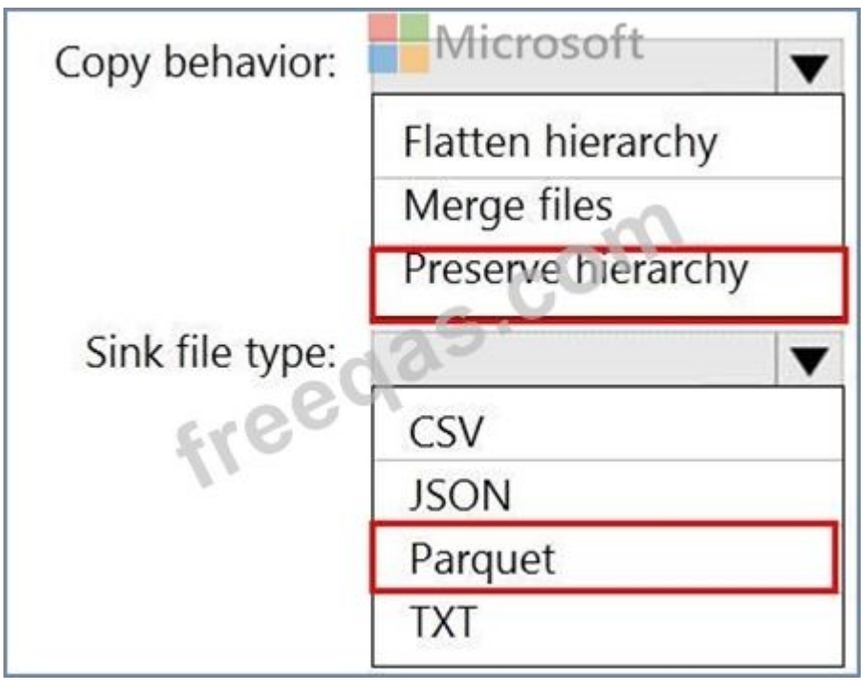
NEW QUESTION: 142
You plan to implement an Azure Data Lake Storage Gen2 container that will contain CSV files. The size of the files will vary based on the number of events that occur per hour.
File sizes range from 4.KB to 5 GB.
You need to ensure that the files stored in the container are optimized for batch processing.
What should you do?
A. Compress the files.
B. Convert the files to Avro.
C. Convert the files to JSON
D. Merge the files.
Answer: B (LEAVE A REPLY)

NEW QUESTION: 143
You use Azure Data Factory to prepare data to be queried by Azure Synapse Analytics serverless SQL pools.
Files are initially ingested into an Azure Data Lake Storage Gen2 account as 10 small JSON files. Each file contains the same data attributes and data from a subsidiary of your company.
You need to move the files to a different folder and transform the data to meet the following requirements:
Provide the fastest possible query times.
Automatically infer the schema from the underlying files.
How should you configure the Data Factory copy activity? To answer, select the appropriate options in the answer are a.

NOTE: Each correct selection is worth one point.



Answer:



Reference:

https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-introduction

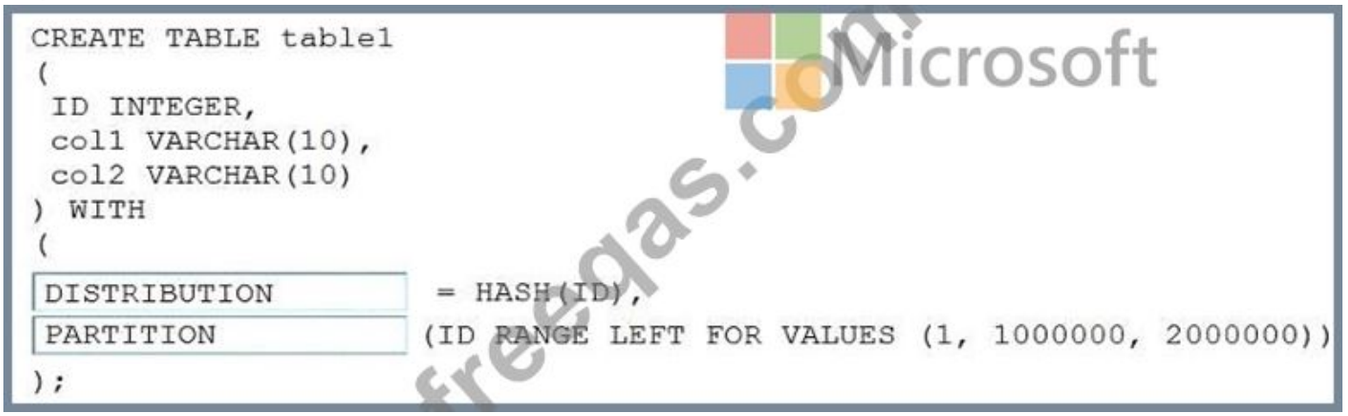
https://docs.microsoft.com/en-us/azure/data-factory/format-parquet

NEW QUESTION: 144
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: (SHOW ANSWER)
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: 145
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.

Answer:

Explanation



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: 146

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.

Answer:

See the answer below in explanation.

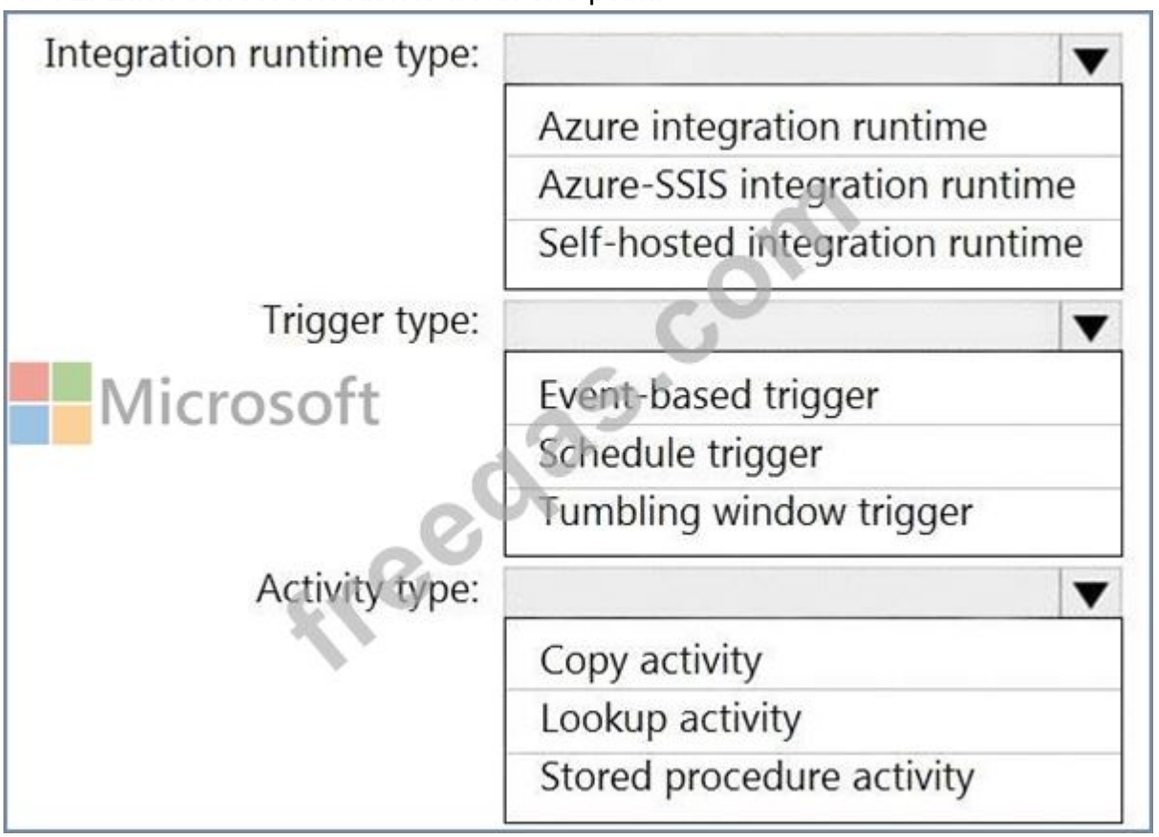
Explanation

Answer as below

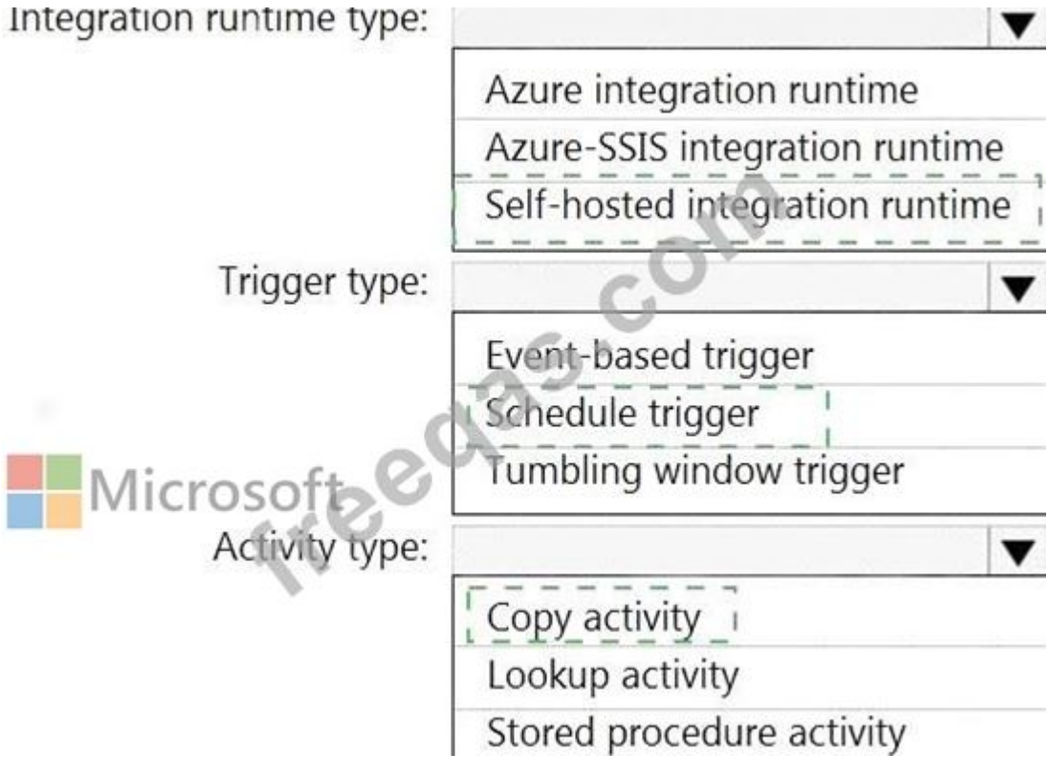
NEW QUESTION: 147

Which Azure Data Factory components should you recommend using together to import the daily inventory data from the SQL server to Azure Data Lake Storage? To answer, select the appropriate options in the answer area.

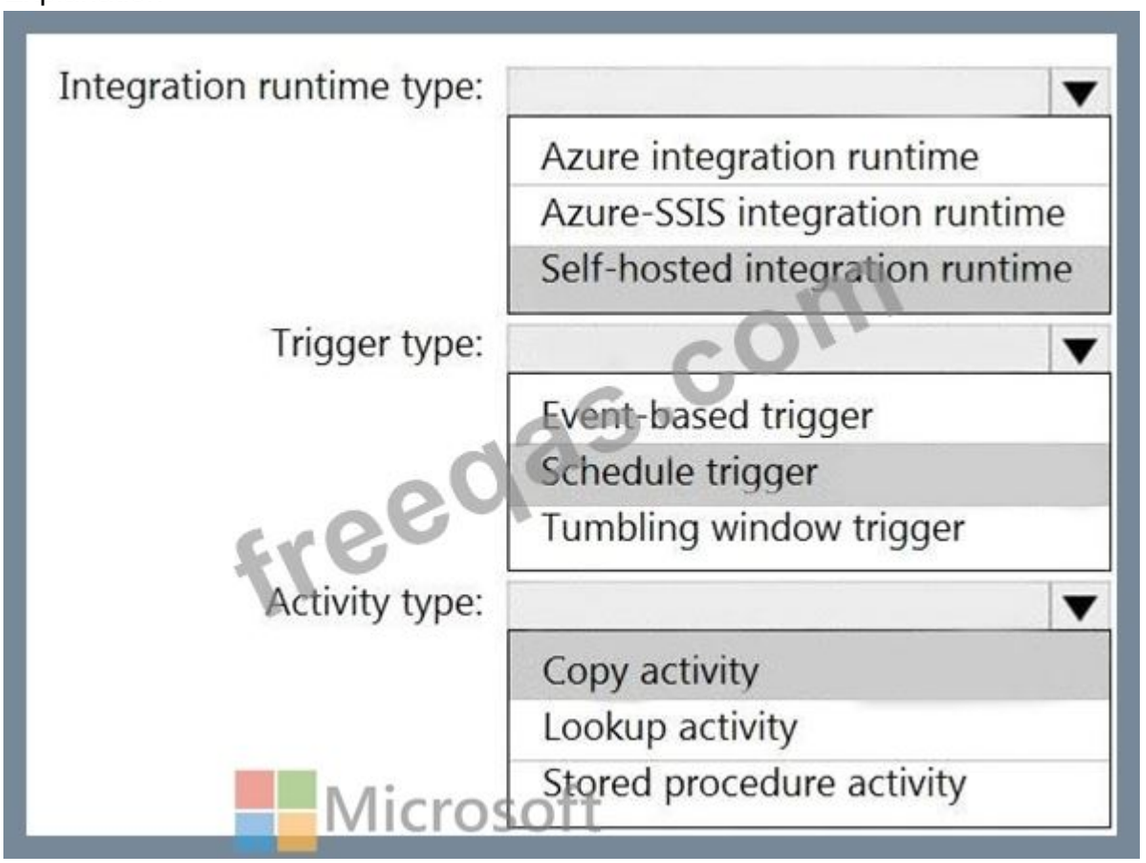
NOTE: Each correct selection is worth one point.



Answer:



Explanation



Box 1: Self-hosted integration runtime

A self-hosted IR is capable of running copy activity between a cloud data stores and a data store in private network.

Box 2: Schedule trigger

Schedule every 8 hours

Box 3: Copy activity

Scenario:

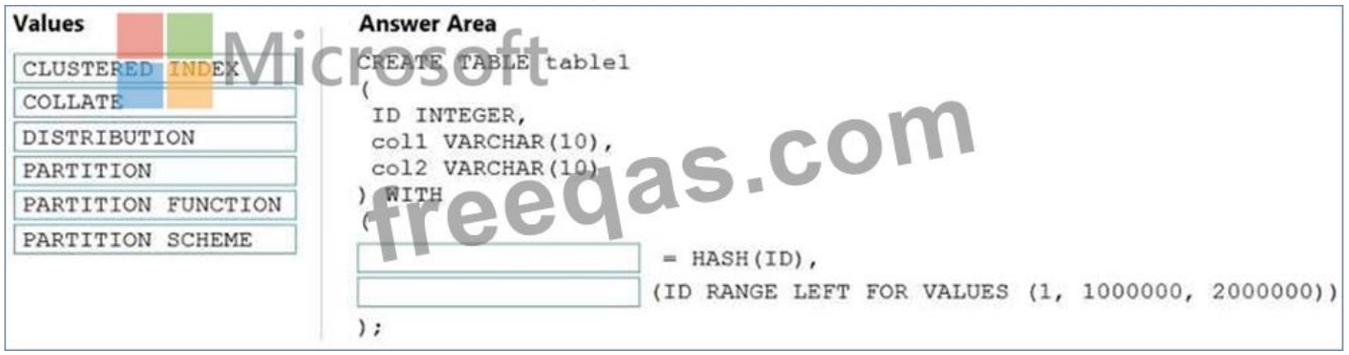
* Customer data, including name, contact information, and loyalty number, comes from Salesforce and can be imported into Azure once every eight hours. Row modified dates are not trusted in the source table.
* Product data, including product ID, name, and category, comes from Salesforce and can be imported into Azure once every eight hours. Row modified dates are not trusted in the source table.

NEW QUESTION: 148

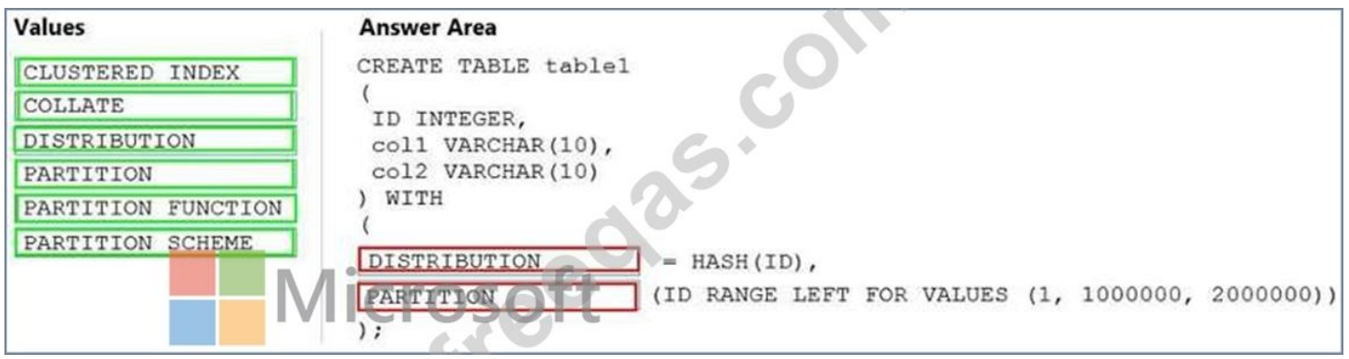
You need to create a partitioned table in an Azure Synapse Analytics dedicated SQL pool.

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.



Answer:



Reference:

https://docs.microsoft.com/en-us/sql/t-sql/statements/create-table-azure-sql-data-warehouse?

NEW QUESTION: 149

You are designing a fact table named FactPurchase in an Azure Synaps 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
```

A. round-robin
B. hash-distributed on DateKey
C. hash-distributed on PurchaseKey
D. replicated
Answer: A (LEAVE A REPLY)

NEW QUESTION: 150
You are designing an anomaly detection solution for streaming data from an Azure IoT hub. The solution must meet the following requirements:
* Send the output to Azure Synapse.
* Identify spikes and dips in time series data.
* Minimize development and configuration effort.
Which should you include in the solution?
A. Azure Databricks
B. Azure Stream Analytics
C. Azure SQL Database
Answer: (SHOW ANSWER)
Explanation
You can identify anomalies by routing data via IoT Hub to a built-in ML model in Azure Stream Analytics.
Reference:
https://docs.microsoft.com/en-us/learn/modules/data-anomaly-detection-using-azure-iot-hub/

NEW QUESTION: 151

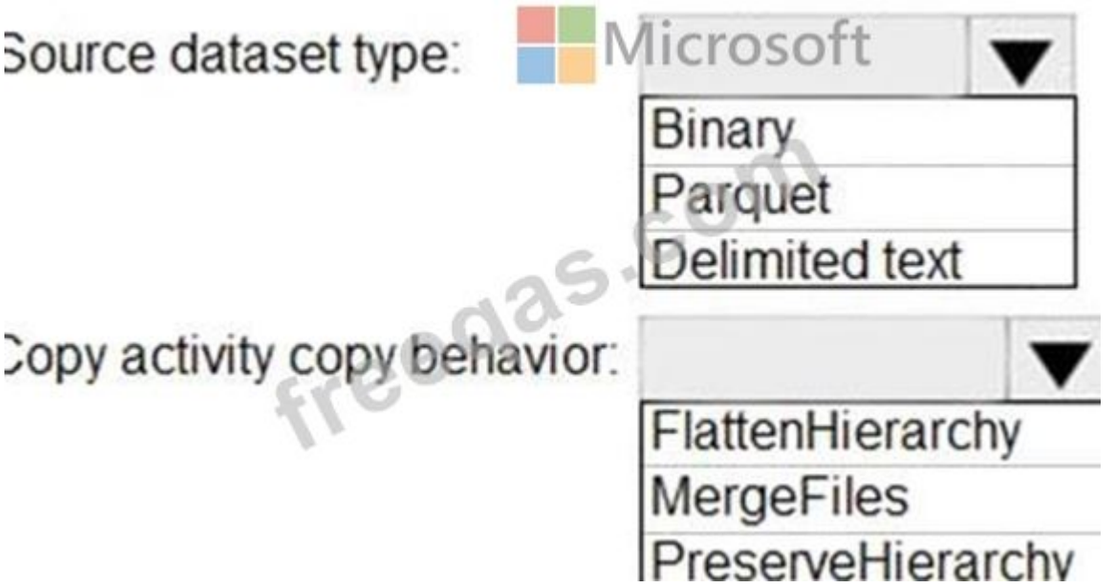
You have two Azure Storage accounts named Storage1 and Storage2. Each account holds one container and has the hierarchical namespace enabled. The system has files that contain data stored in the Apache Parquet format.

You need to copy folders and files from Storage1 to Storage2 by using a Data Factory copy activity. The solution must meet the following requirements:

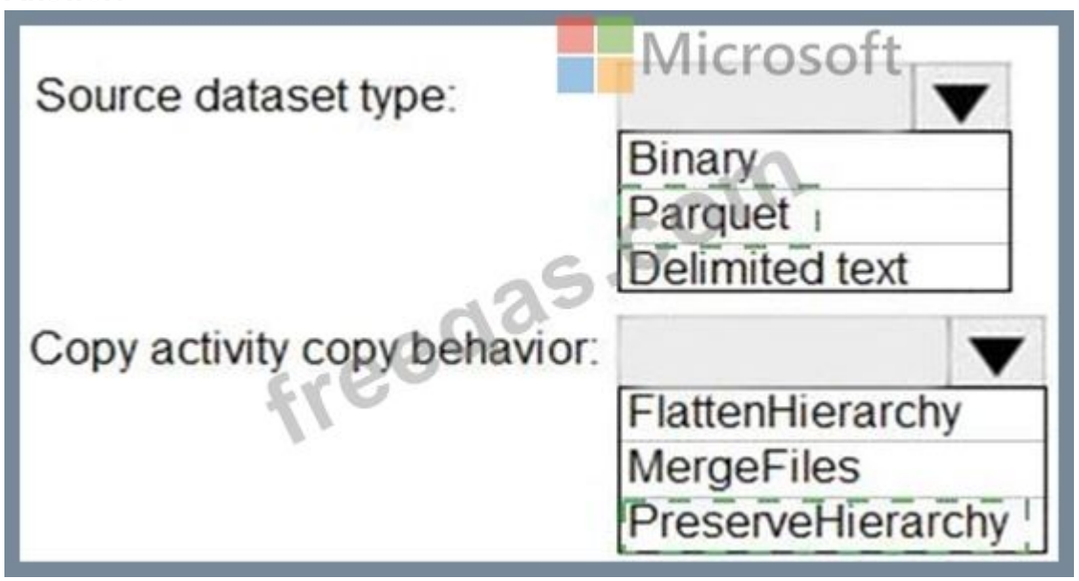
* No transformations must be performed.
* The original folder structure must be retained.
* Minimize time required to perform the copy activity.

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

NOTE: Each correct selection is worth one point.

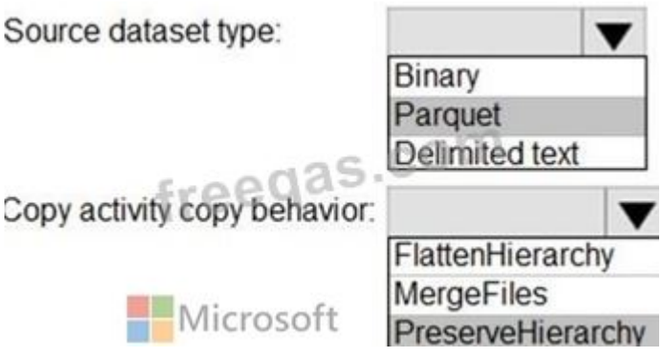


Answer:



Explanation

Graphical user interface, text, application, chat or text message Description automatically generated



Box 1: Parquet

For Parquet datasets, the type property of the copy activity source must be set to ParquetSource.

Box 2: PreserveHierarchy

PreserveHierarchy (default): Preserves the file hierarchy in the target folder. The relative path of the source file to the source folder is identical to the relative path of the target file to the target folder.

Reference:

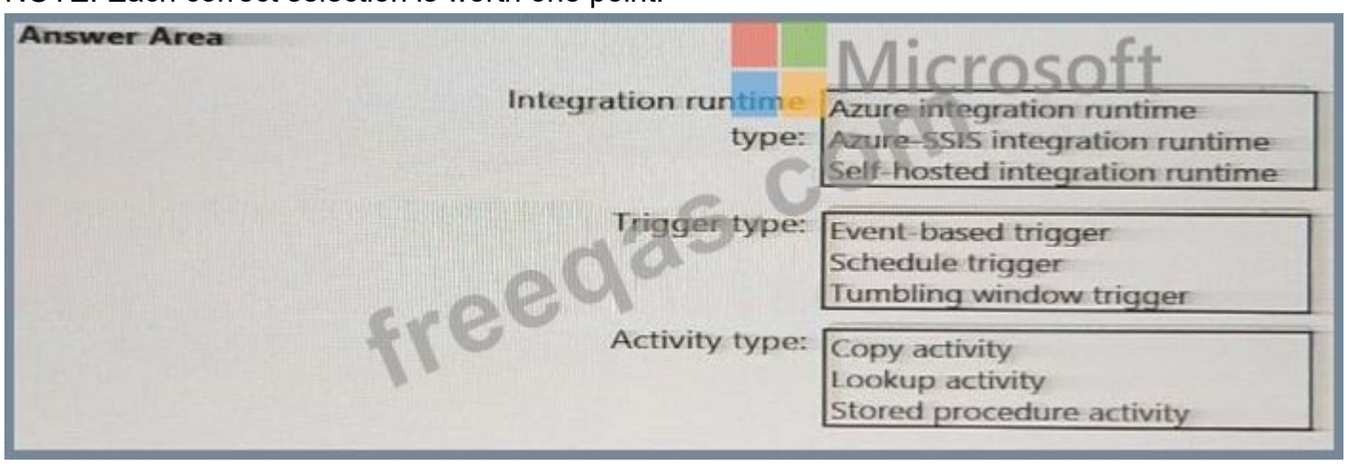
https://docs.microsoft.com/en-us/azure/data-factory/format-parquet

https://docs.microsoft.com/en-us/azure/data-factory/connector-azure-data-lake-storage

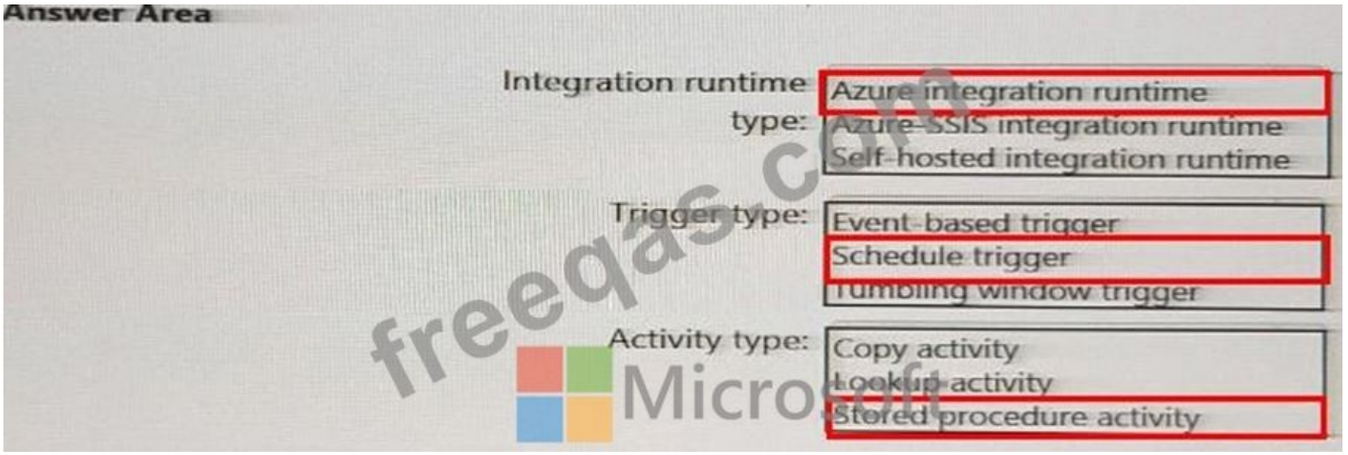
NEW QUESTION: 152

Which Azure Data Factory components should you recommend using together to import the daily inventory data from the SQL server to Azure Data Lake Storage? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.



Answer:



NEW QUESTION: 153

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.

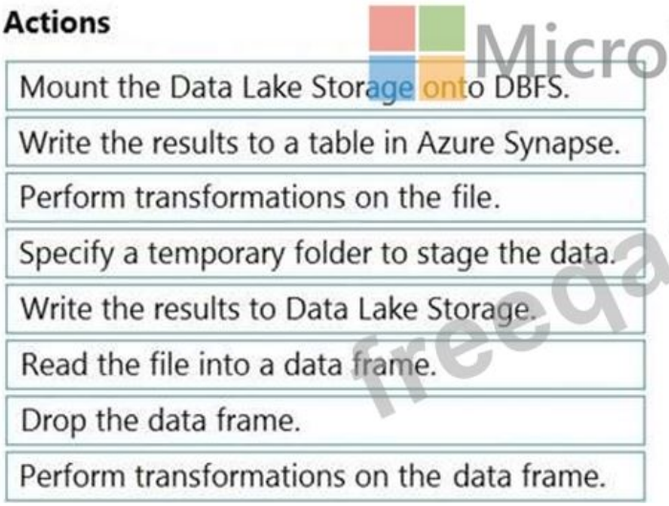
You create the following components:

A destination table in Azure Synapse

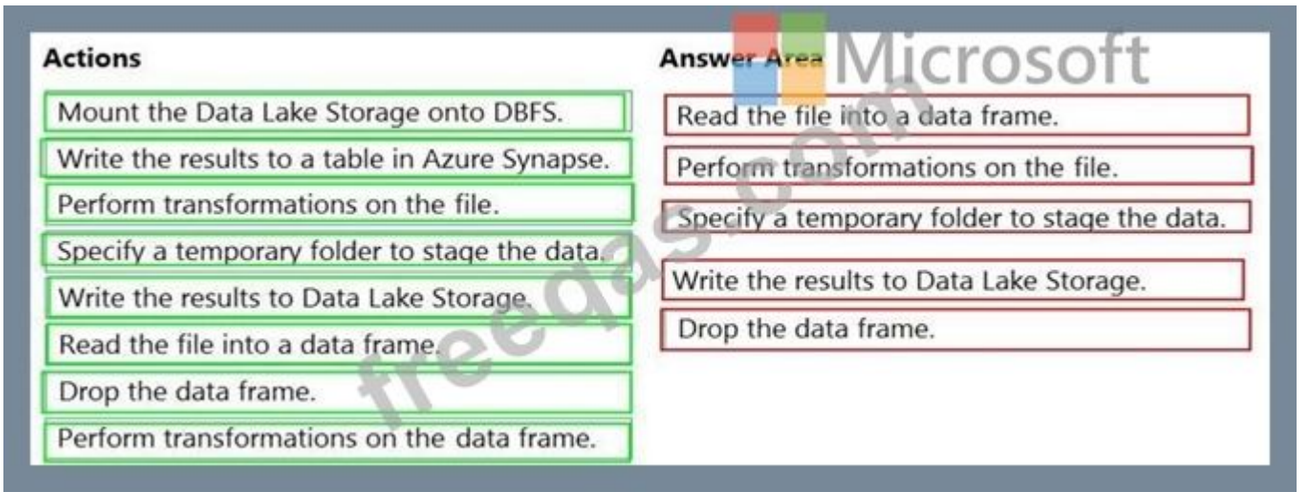
An Azure Blob storage container

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.



Answer:



Reference:
https://docs.microsoft.com/en-us/azure/azure-databricks/databricks-extract-load-sql-data-warehouse

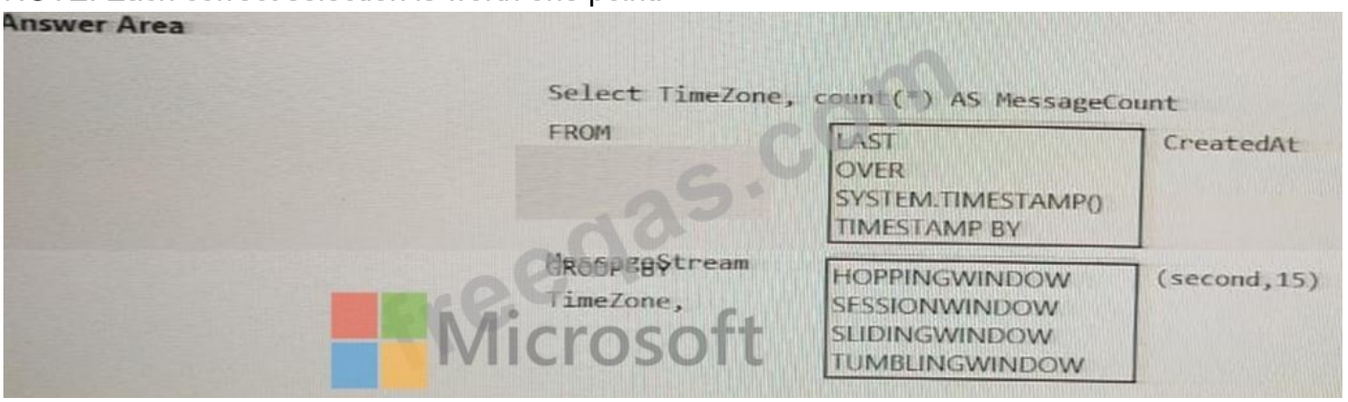
NEW QUESTION: 154

You are designing an Azure Stream Analytics solution that receives instant messaging data from an Azure event hub.

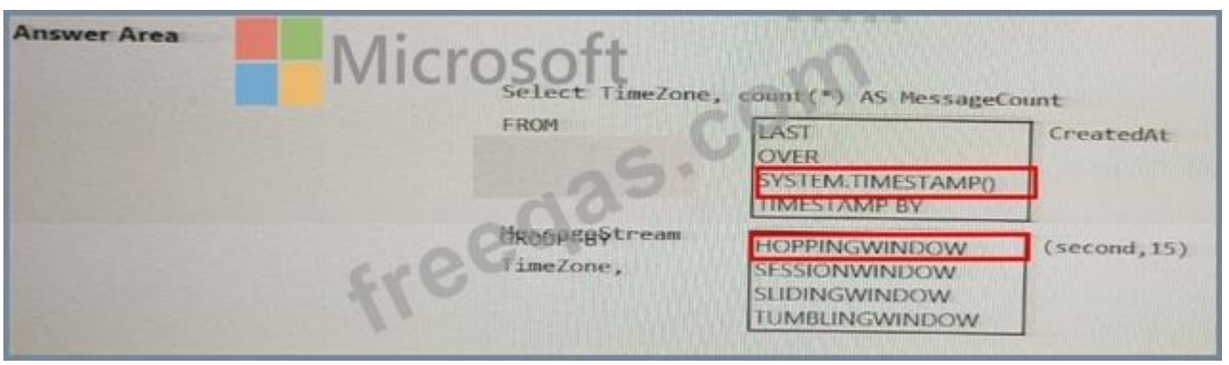
You need to ensure that the output from the Stream Analytics job counts the number of messages per time zone every 15 seconds.

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.



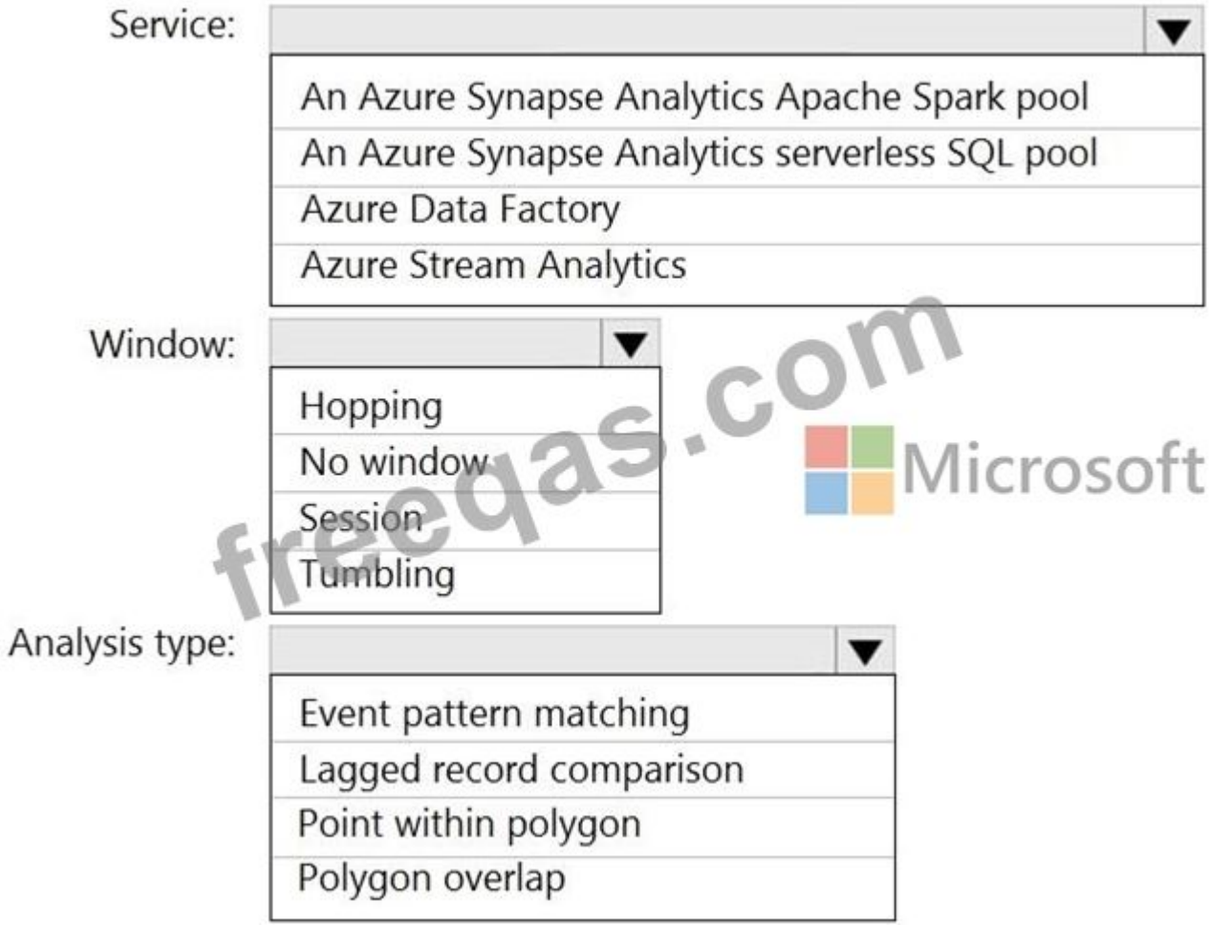
Answer:



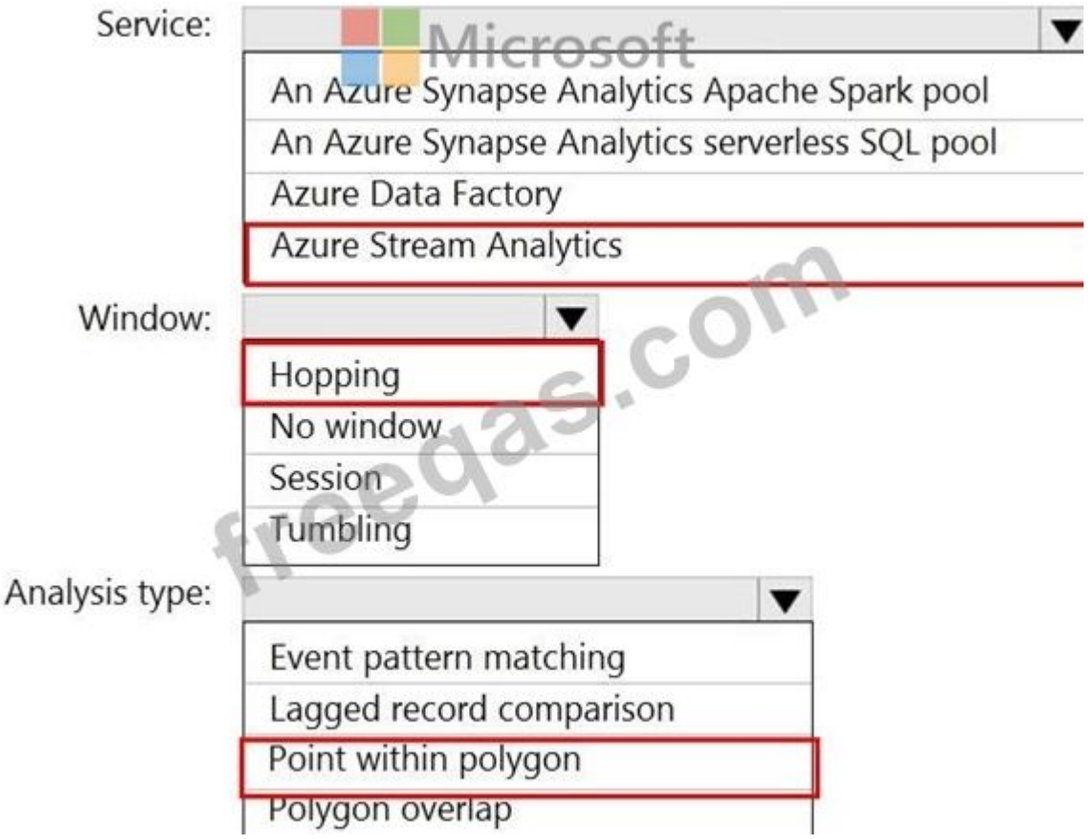
NEW QUESTION: 155

You have an Azure Data Factory version 2 (V2) resource named Df1. Df1 contains a linked service.
You have an Azure Key vault named vault1 that contains an encryption key named key1.
You need to encrypt Df1 by using key1.
What should you do first?
A. Add a private endpoint connection to vaul 1.
B. Enable Azure role-based access control on vault 1.
C. Remove the linked service from Df1.
D. Create a self-hosted integration runtime.
Answer: (SHOW ANSWER)
Linked services are much like connection strings, which define the connection information needed for Data Factory to connect to external resources.
Reference:
https://docs.microsoft.com/en-us/azure/data-factory/enable-customer-managed-key
https://docs.microsoft.com/en-us/azure/data-factory/concepts-linked-services
https://docs.microsoft.com/en-us/azure/data-factory/create-self-hosted-integration-runtime

NEW QUESTION: 156
You are designing a monitoring solution for a fleet of 500 vehicles. Each vehicle has a GPS tracking device that sends data to an Azure event hub once per minute.
You have a CSV file in an Azure Data Lake Storage Gen2 container. The file maintains the expected geographical area in which each vehicle should be.
You need to ensure that when a GPS position is outside the expected area, a message is added to another event hub for processing within 30 seconds. The solution must minimize cost.
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.



Answer:



Reference:

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

NEW QUESTION: 157
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.
Reference:
https://docs.microsoft.com/en-us/azure/sql-database/sql-database-security-overview

NEW QUESTION: 158
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: (SHOW ANSWER)
Explanation
Explanation:

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

Incorrect Answers:

C: Self-hosted integration runtime is to be used On-premises.

Reference:

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

NEW QUESTION: 159

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)

Explanation

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: 160

You have a table named SalesFact in an enterprise data warehouse in Azure Synapse Analytics.

SalesFact contains sales data from the past 36 months and has the following characteristics:

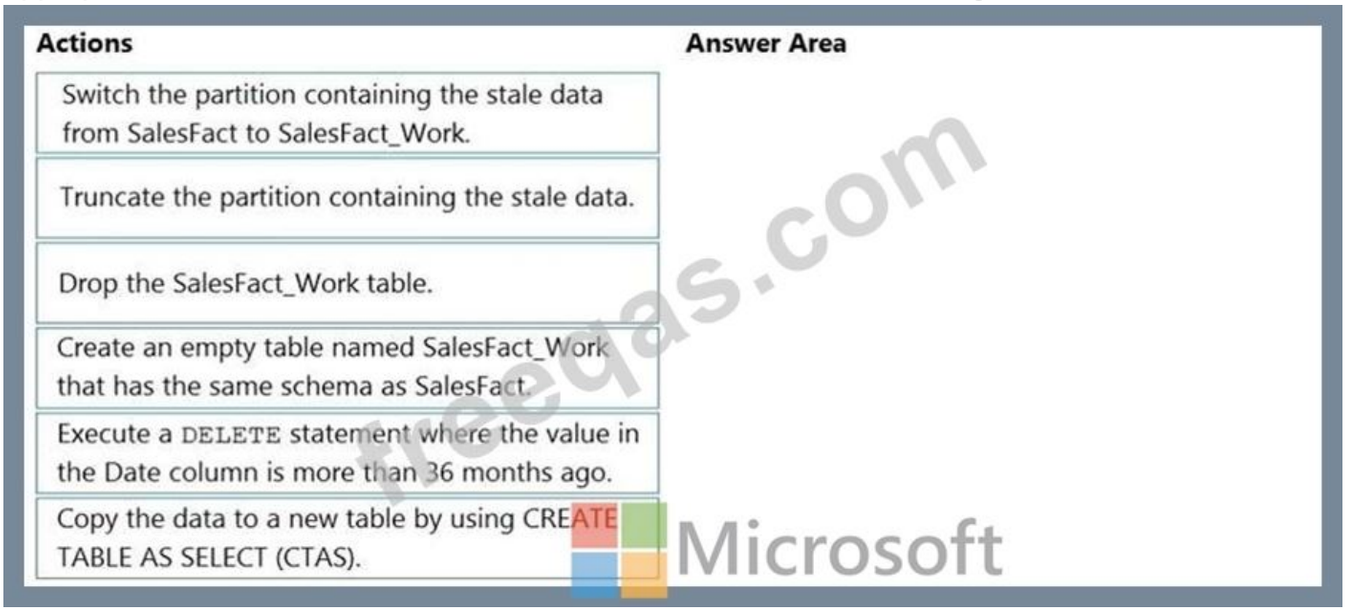
Is partitioned by month

Contains one billion rows

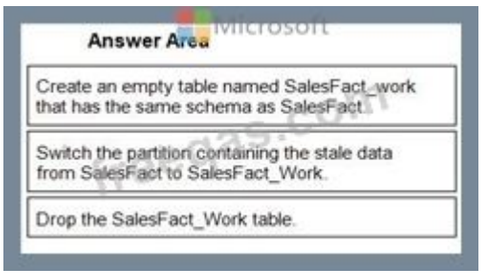
Has clustered columnstore indexes

At the beginning of each month, you need to remove data from SalesFact that is older than 36 months as quickly as possible.

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



Answer:



1 - Create an empty table named SalesFact_work that has the same schema as SalesFact.

2 - Switch the partition containing the stale data from SalesFact to SalesFact_Work.

3 - Drop the SalesFact_Work table.

Reference:

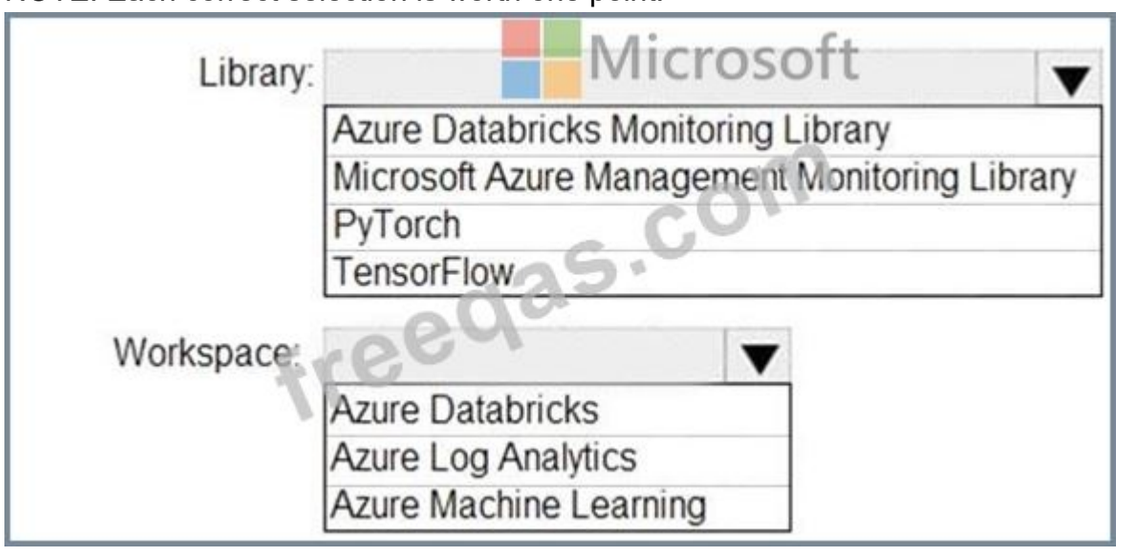
https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-tables-partition

NEW QUESTION: 161

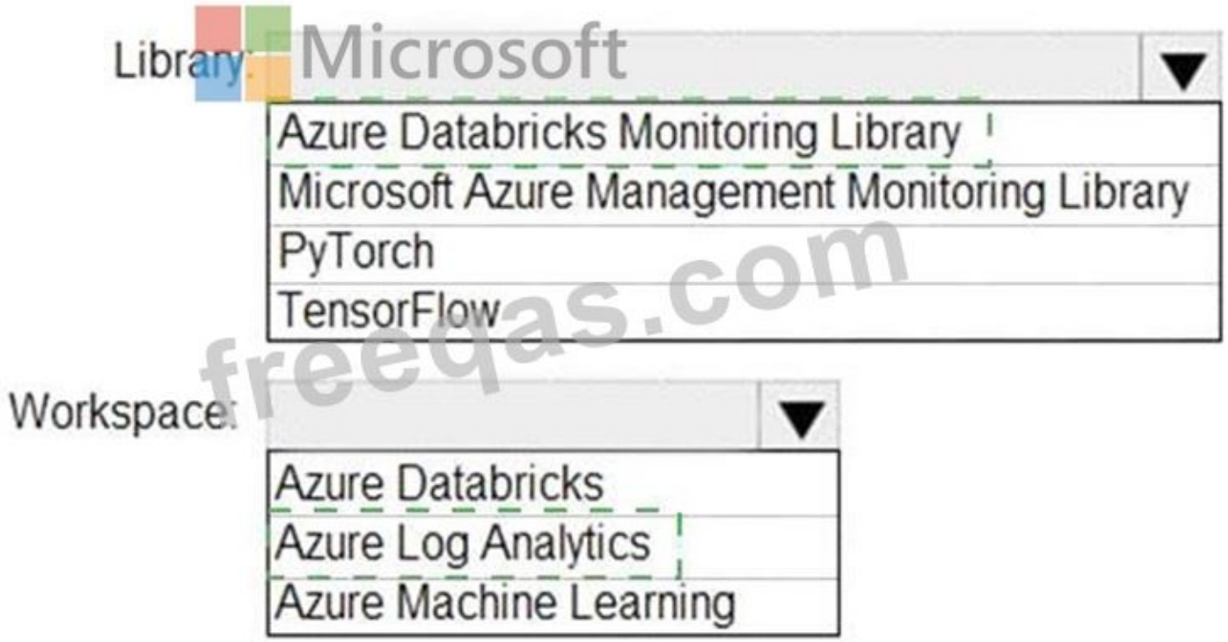
You have an Azure Synapse Analytics dedicated SQL pool that contains a table named Table1.

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 data flow that contains a Derived Column transformation.
A. No
B. Yes
Answer: (SHOW ANSWER)

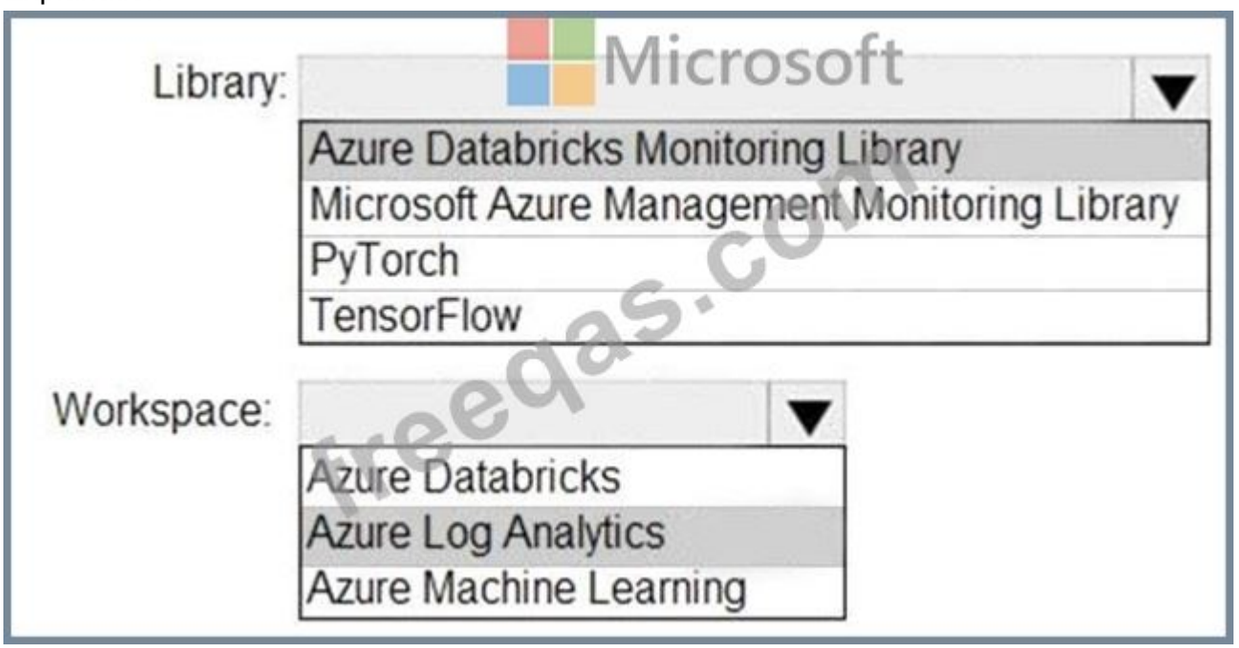
NEW QUESTION: 162
You need to collect application metrics, streaming query events, and application log messages for an Azure Databrick cluster.
Which type of library and workspace should you implement? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.



Answer:



Explanation



You can send application logs and metrics from Azure Databricks to a Log Analytics workspace. It uses the Azure Databricks Monitoring Library, which is available on GitHub.

References:

https://docs.microsoft.com/en-us/azure/architecture/databricks-monitoring/application-logs

NEW QUESTION: 163

You are designing a streaming data solution that will ingest variable volumes of data.

You need to ensure that you can change the partition count after creation.

Which service should you use to ingest the data?

A. Azure Event Hubs Dedicated

B. Azure Stream Analytics

C. Azure Data Factory

D. Azure Synapse Analytics

Answer: A (LEAVE A REPLY)

You can't change the partition count for an event hub after its creation except for the event hub in a dedicated cluster.

Reference:

https://docs.microsoft.com/en-us/azure/event-hubs/event-hubs-features

NEW QUESTION: 164

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 more than 1 MB.

Does this meet the goal?

A. Yes

B. No

Answer: (SHOW ANSWER)

Explanation

Instead modify the files to ensure that each row is less than 1 MB.

References:

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

NEW QUESTION: 165

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.

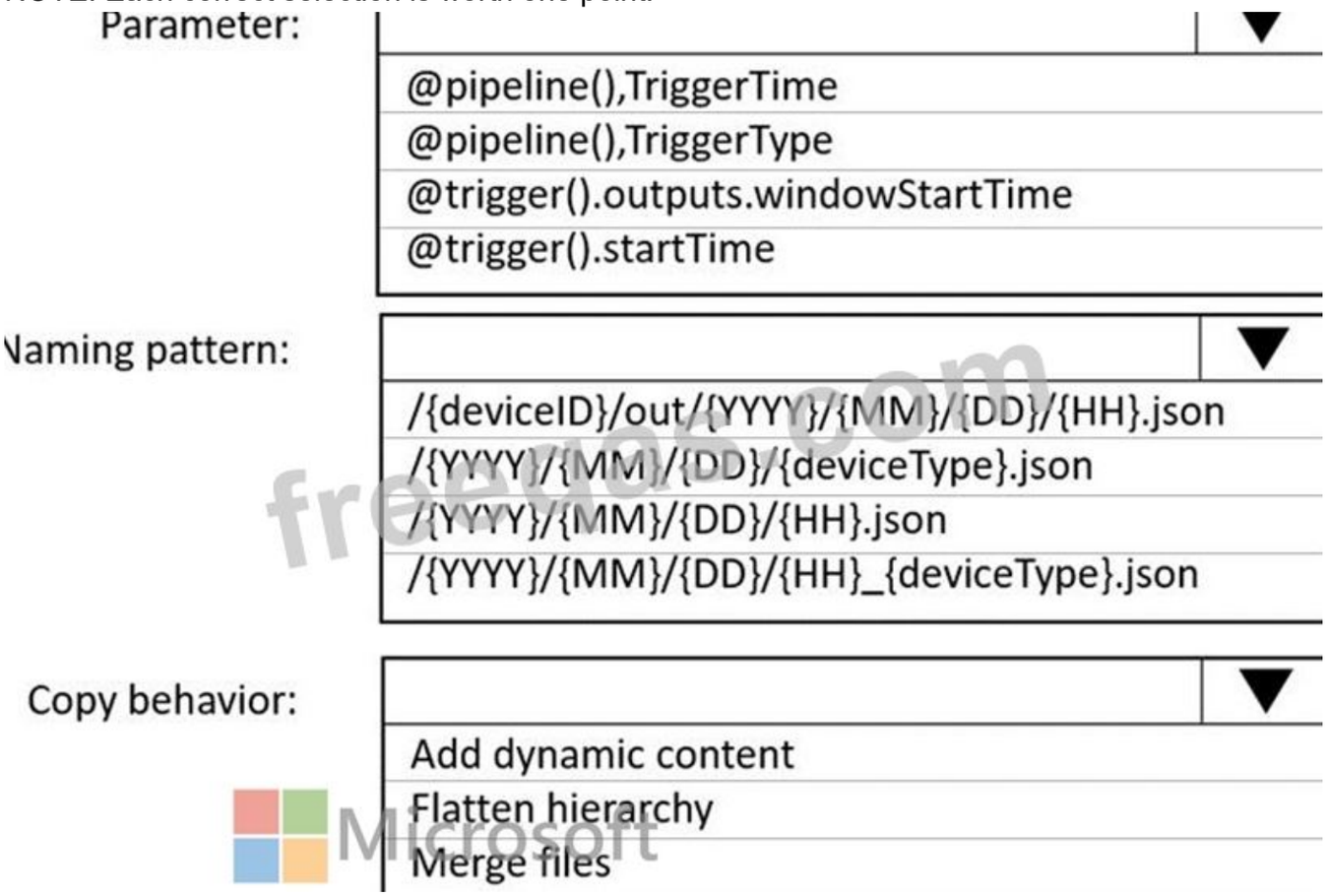
Answer: C (LEAVE A REPLY)

Explanation

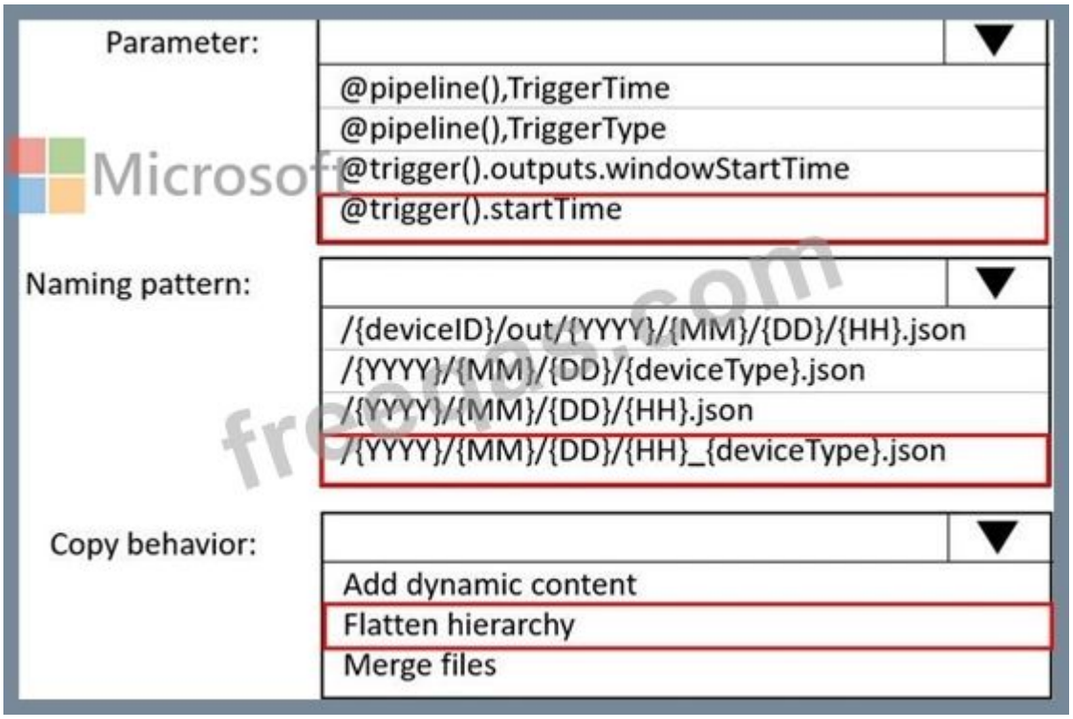
General symptoms of the job hitting system resource limits include:

* 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: 166
You are building an Azure Data Factory solution to process data received from Azure Event Hubs, and then ingested into an Azure Data Lake Storage Gen2 container.
The data will be ingested every five minutes from devices into JSON files. The files have the following naming pattern.
/{deviceType}/in/{YYYY}/{MM}/{DD}/{HH}/{deviceID}_{YYYY}{MM}{DD}HH}{mm}.json You need to prepare the data for batch data processing so that there is one dataset per hour per deviceType. The solution must minimize read times.
How should you configure the sink for the copy activity? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.



Answer:



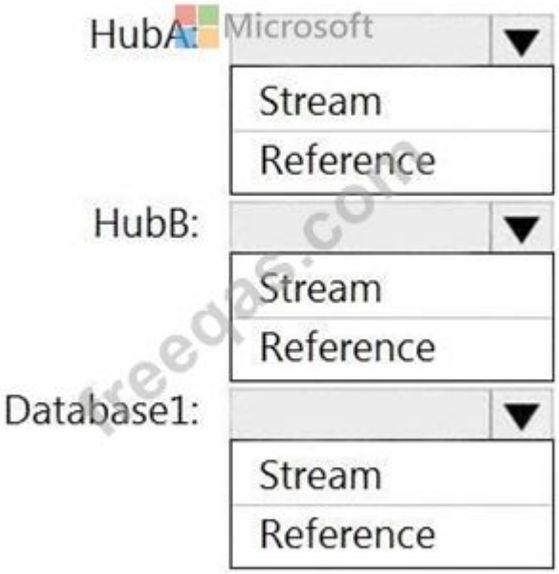
Reference:
https://docs.microsoft.com/en-us/azure/data-factory/concepts-pipeline-execution-triggers
https://docs.microsoft.com/en-us/azure/data-factory/connector-file-system

NEW QUESTION: 167
You have an Azure SQL database named Database1 and two Azure event hubs named HubA and HubB. The data consumed from each source is shown in the following table.

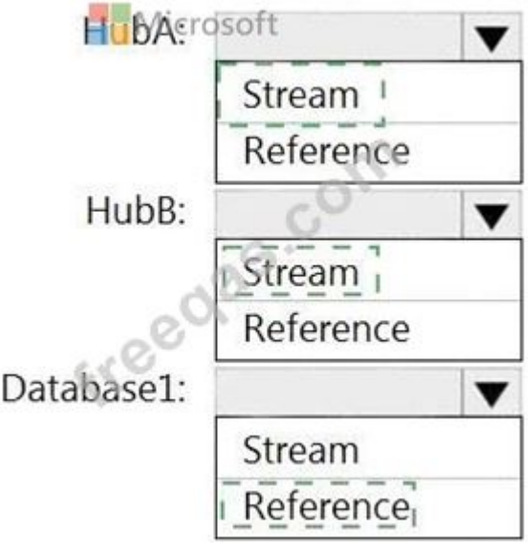
Source	Data
Database1	Driver's name Driver's license number
HubA	Ride route Ride distance Ride duration
HubB	Ride fare Ride payment

You need to implement Azure Stream Analytics to calculate the average fare per mile by driver.
How should you configure the Stream Analytics input for each source? To answer, select the appropriate options in the answer area.

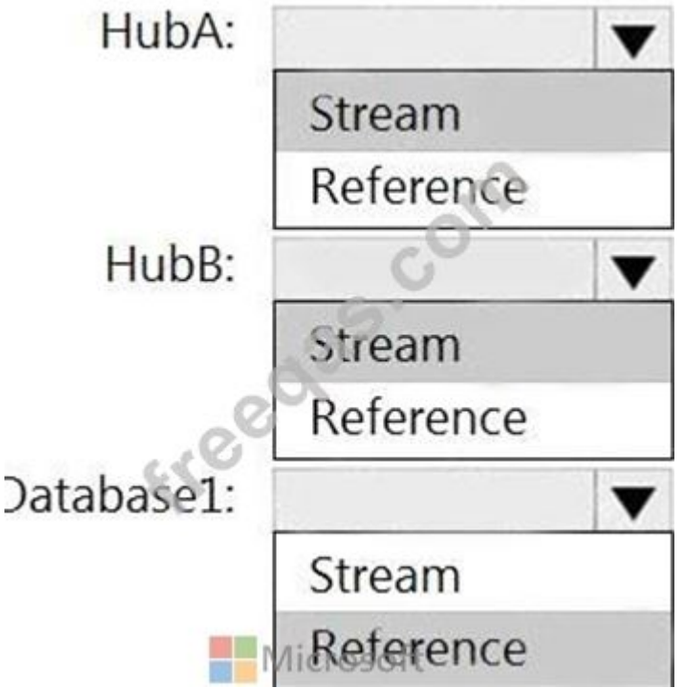
NOTE: Each correct selection is worth one point.



Answer:



Explanation

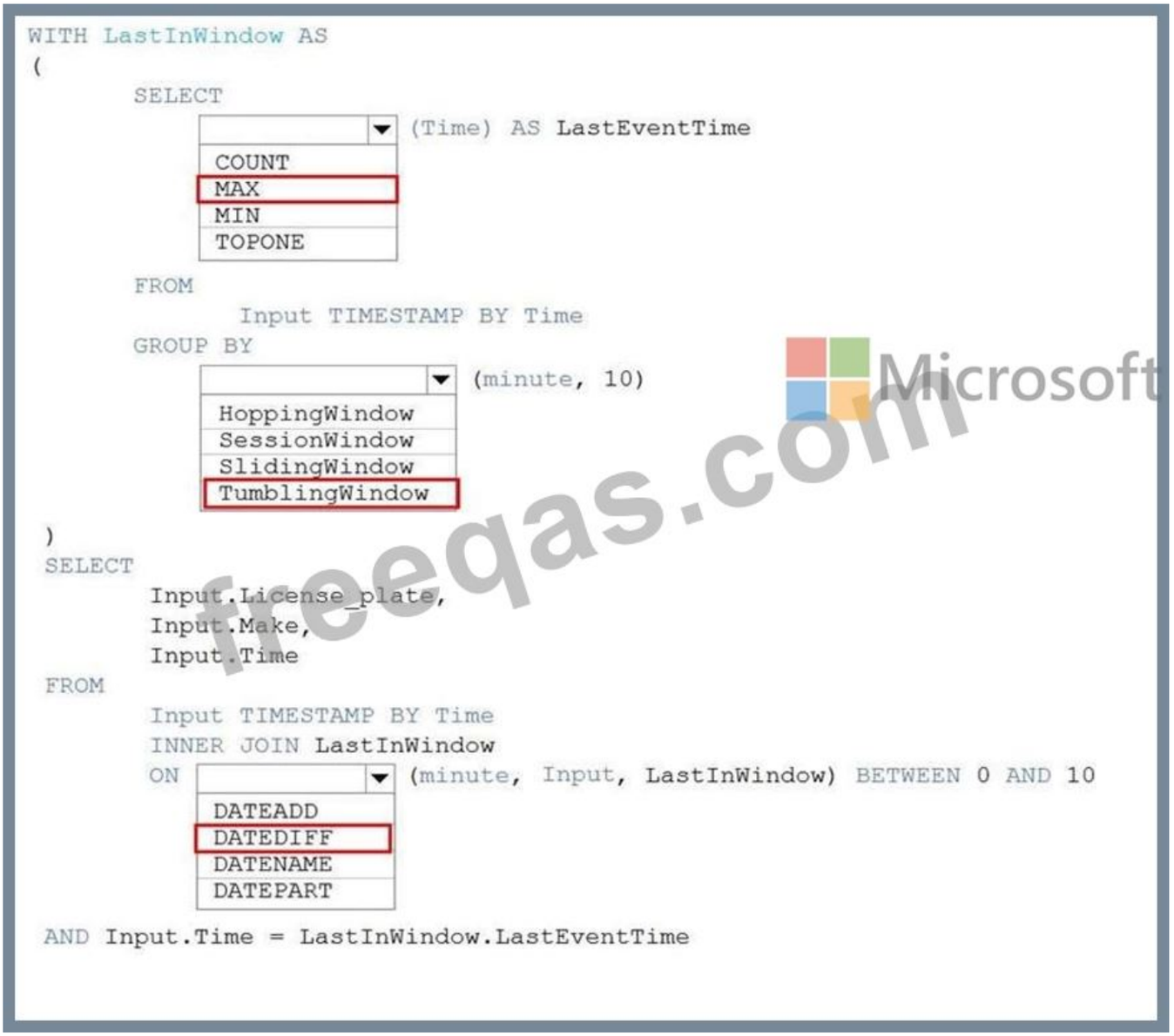


HubA: Stream
HubB: Stream
Database1: Reference
Reference data (also known as a lookup table) is a finite data set that is static or slowly changing in nature, used to perform a lookup or to augment your data streams. For example, in an IoT scenario, you could store metadata about sensors (which don't change often) in reference data and join it with real time IoT data streams. Azure Stream Analytics loads reference data in memory to achieve low latency stream processing Reference:
https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-use-reference-data

NEW QUESTION: 168
You are processing streaming data from vehicles that pass through a toll booth.
You need to use Azure Stream Analytics to return the license plate, vehicle make, and hour the last vehicle passed during each 10-minute window.
How should you complete the query? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

```
WITH LastInWindow AS
(
    SELECT
        [ ▼ ] (Time) AS LastEventTime
        COUNT
        MAX
        MIN
        TOPONE
    FROM
        Input TIMESTAMP BY Time
    GROUP BY
        [ ▼ ] (minute, 10)
        HoppingWindow
        SessionWindow
        SlidingWindow
        TumblingWindow
)
SELECT
    Input.License_plate,
    Input.Make,
    Input.Time
FROM
    Input TIMESTAMP BY Time
    INNER JOIN LastInWindow
    ON [ ▼ ] (minute, Input, LastInWindow) BETWEEN 0 AND 10
        DATEADD
        DATEDIFF
        DATENAME
        DATEPART
AND Input.Time = LastInWindow.LastEventTime
```


Answer:



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

NEW QUESTION: 169

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.

Answer:

Reference:
https://streamsets.com/documentation/datacollector/latest/help/datacollector/UserGuide/Destinations/ADLS-G2-D.html

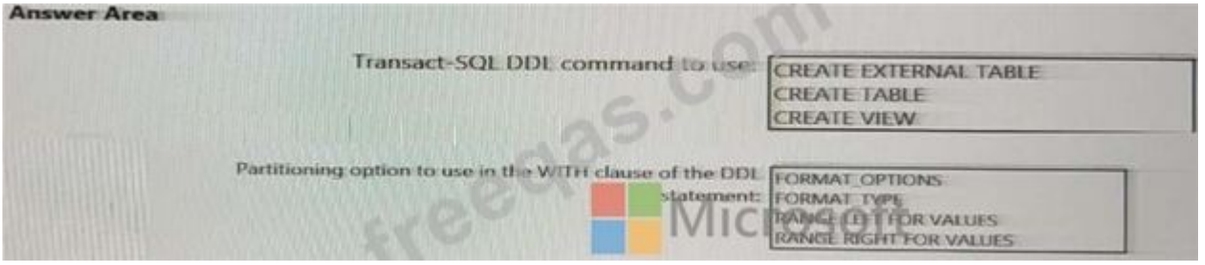
NEW QUESTION: 170

You need to implement an Azure Synapse Analytics database object for storing the sales transactions dat a. The solution must meet the sales transaction dataset requirements.

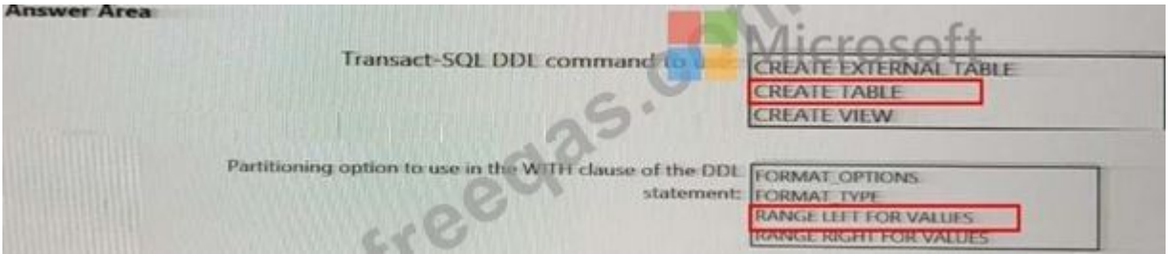
What solution must meet the sales transaction dataset requirements.

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

NOTE: Each correct selection is worth one point.



Answer:



NEW QUESTION: 171

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 Analytic dedicated SQL pool. The CSV file contains three 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. To the data flow, add a sink transformation to write the rows to a file in blob storage.

B. To the data flow, add a Conditional Split transformation to separate the rows that will cause truncation errors.

C. To the data flow, add a filter transformation to filter out rows that will cause truncation errors.

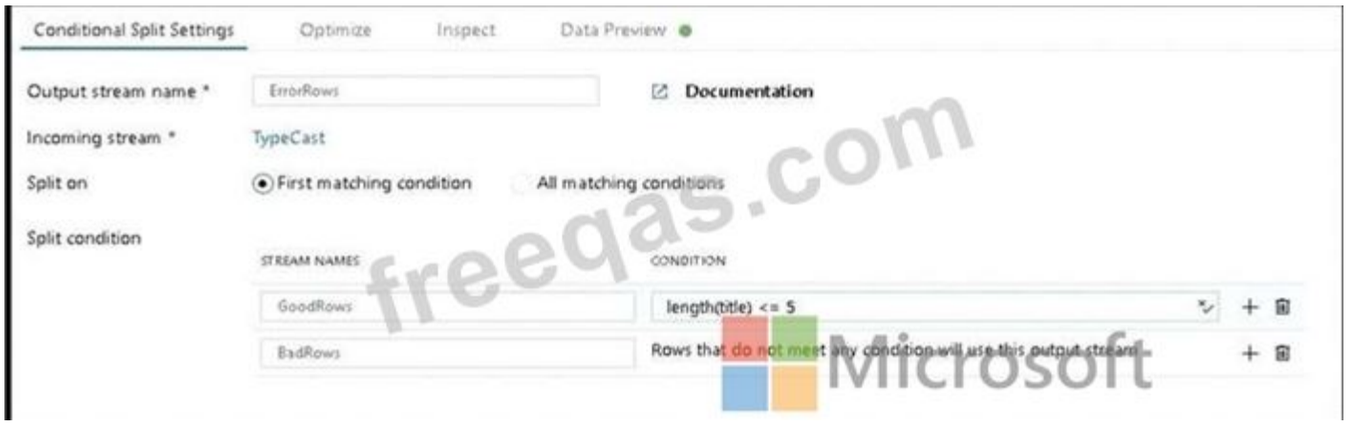
D. Add a select transformation to select only the rows that will cause truncation errors.

Answer: (SHOW ANSWER)

Explanation

B: Example:

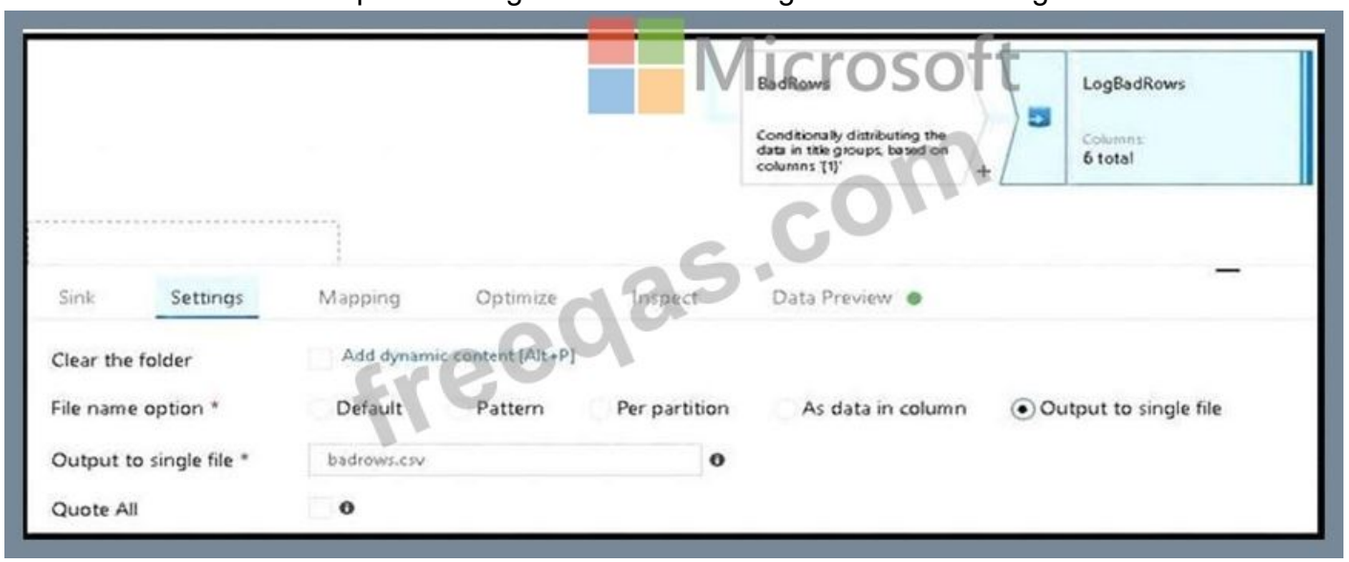
1. This conditional split transformation defines the maximum length of "title" to be five. Any row that is less than or equal to five will go into the GoodRows stream. Any row that is larger than five will go into the BadRows stream.



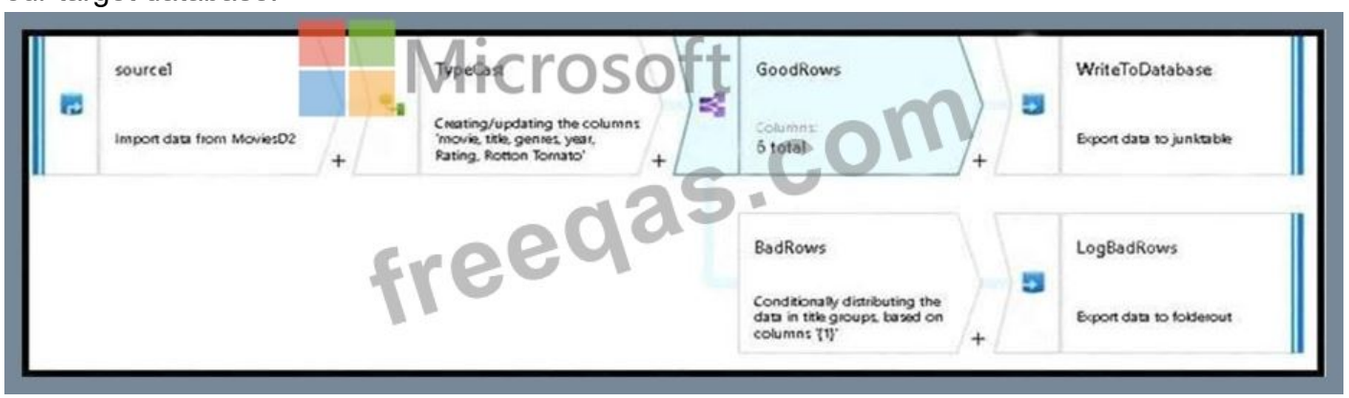
2. This conditional split transformation defines the maximum length of "title" to be five. Any row that is less than or equal to five will go into the GoodRows stream. Any row that is larger than five will go into the BadRows stream.

A:

3. Now we need to log the rows that failed. Add a sink transformation to the BadRows stream for logging. Here, we'll "auto-map" all of the fields so that we have logging of the complete transaction record. This is a text-delimited CSV file output to a single file in Blob Storage. We'll call the log file "badrows.csv".



4. The completed data flow is shown below. We are now able to split off error rows to avoid the SQL truncation errors and put those entries into a log file. Meanwhile, successful rows can continue to write to our target database.



Reference:

https://docs.microsoft.com/en-us/azure/data-factory/how-to-data-flow-error-rows

NEW QUESTION: 172

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: (SHOW ANSWER)

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.

Reference:

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

NEW QUESTION: 173

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:

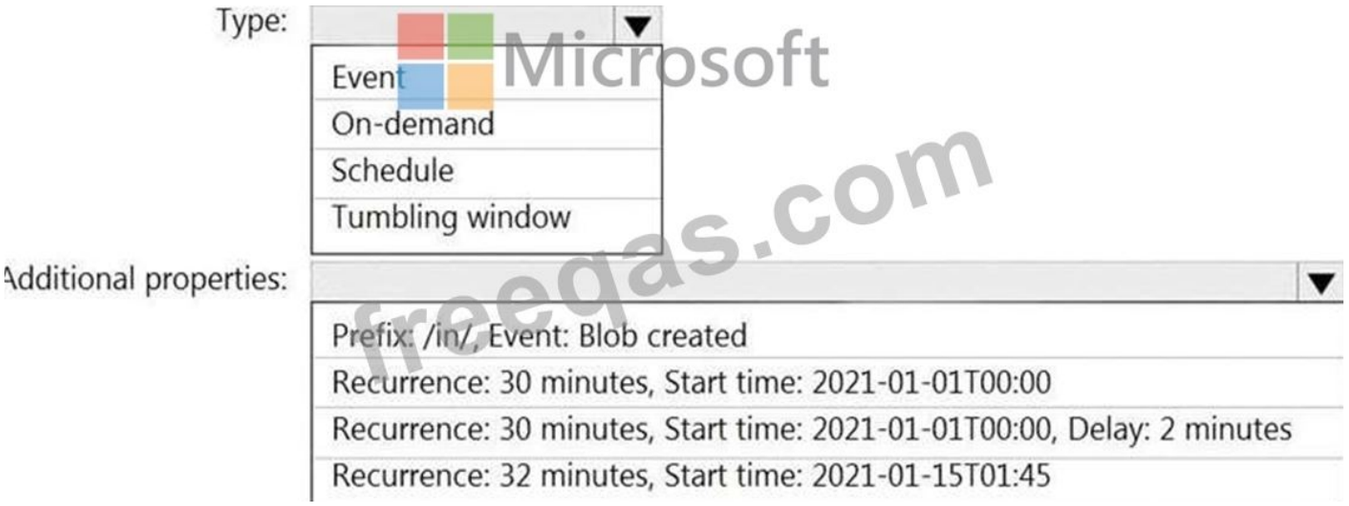
* Existing data must be loaded.

* Data must be loaded every 30 minutes.

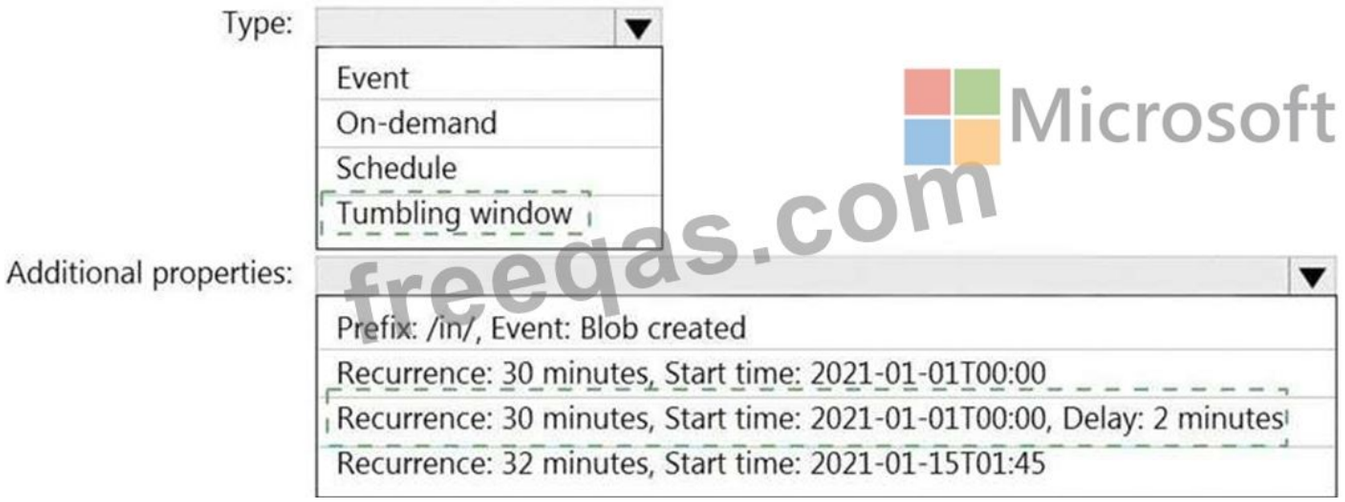
* Late-arriving data of up to two minutes must he included in the load for the time at which the data should have arrived.

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

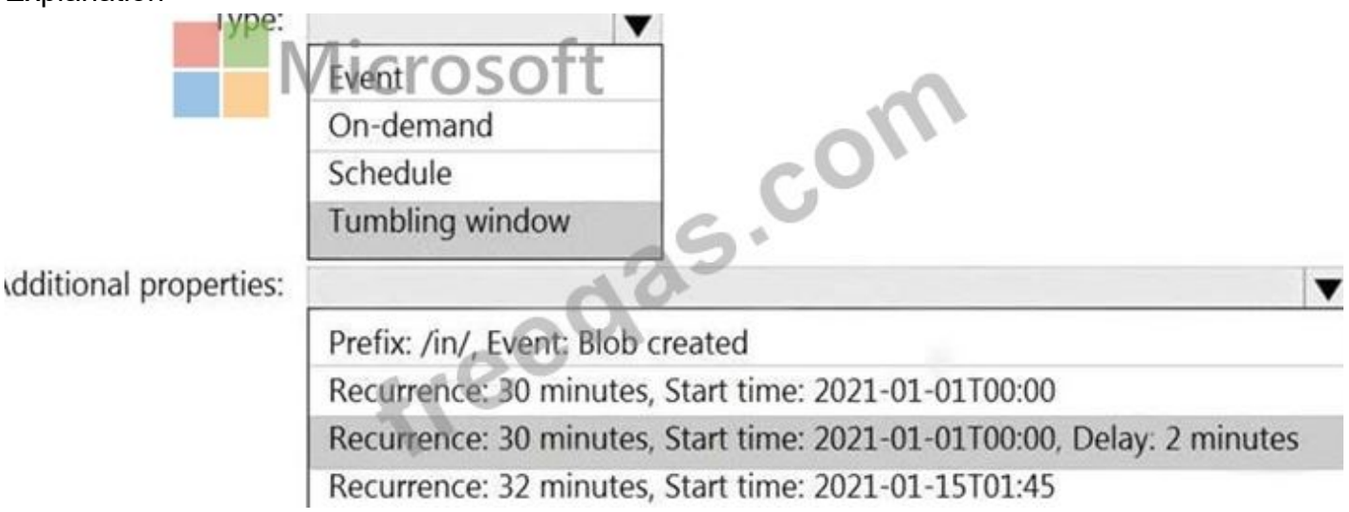
NOTE: Each correct selection is worth one point.



Answer:



Explanation



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

NEW QUESTION: 174
You are building an Azure Synapse Analytics dedicated SQL pool that will contain a fact table for transactions from the first half of the year 2020.
You need to ensure that the table meets the following requirements:
Minimizes the processing time to delete data that is older than 10 years Minimizes the I/O for queries that use year-to-date values 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.

Answer Area

```
CREATE TABLE [dbo].[FactTransaction]
(
     [TransactionTypeID]    int      NOT NULL
,    [TransactionDateID]    int      NOT NULL
,    [CustomerID]           int      NOT NULL
,    [RecipientID]          int      NOT NULL
,    [Amount]               money    NOT NU::
)
WITH
(
```

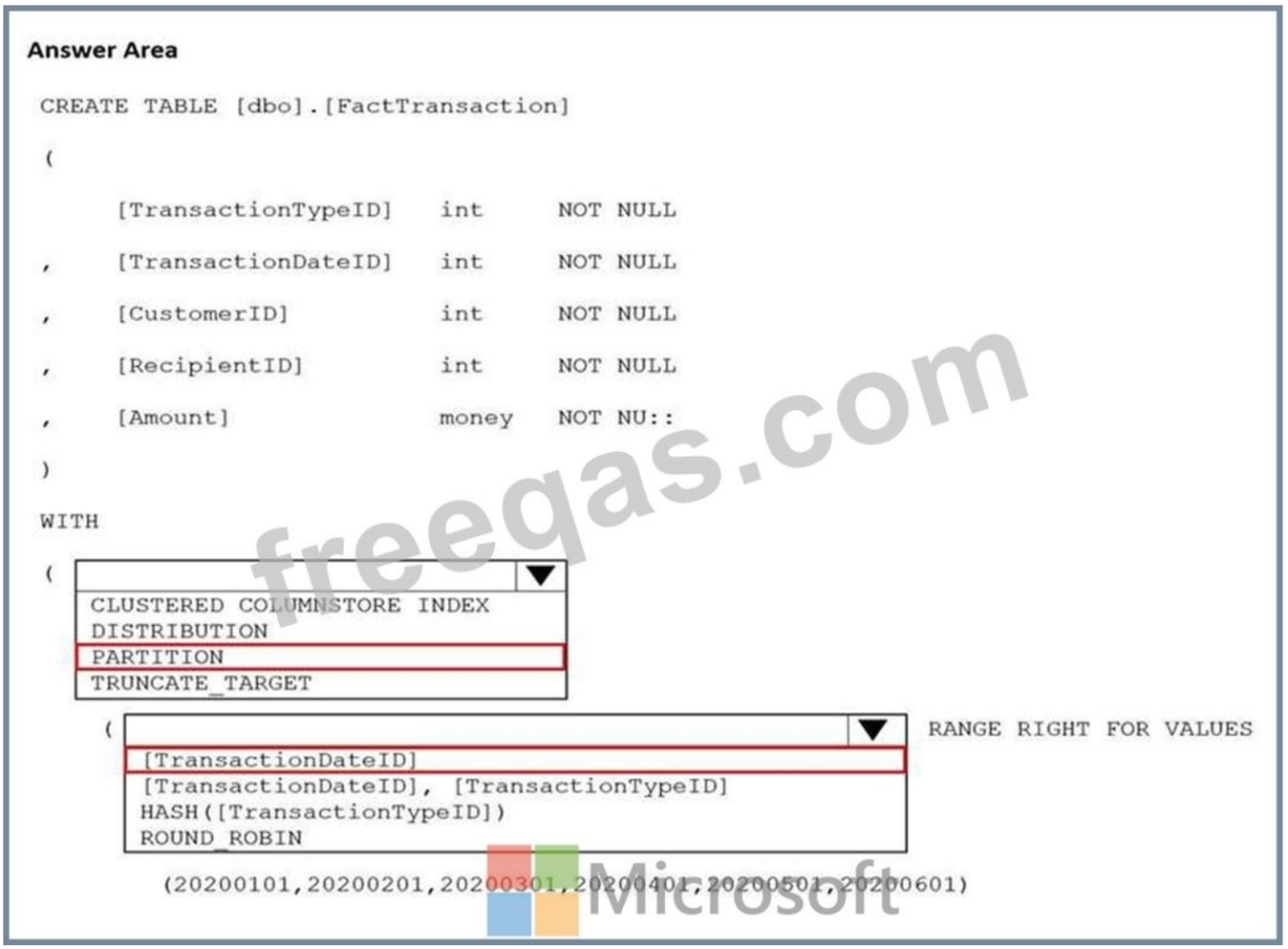
[▼]
- CLUSTERED COLUMNSTORE INDEX
- DISTRIBUTION
- PARTITION
- TRUNCATE_TARGET

```
(
```

[▼] RANGE RIGHT FOR VALUES
- [TransactionDateID]
- [TransactionDateID], [TransactionTypeID]
- HASH([TransactionTypeID])
- ROUND_ROBIN

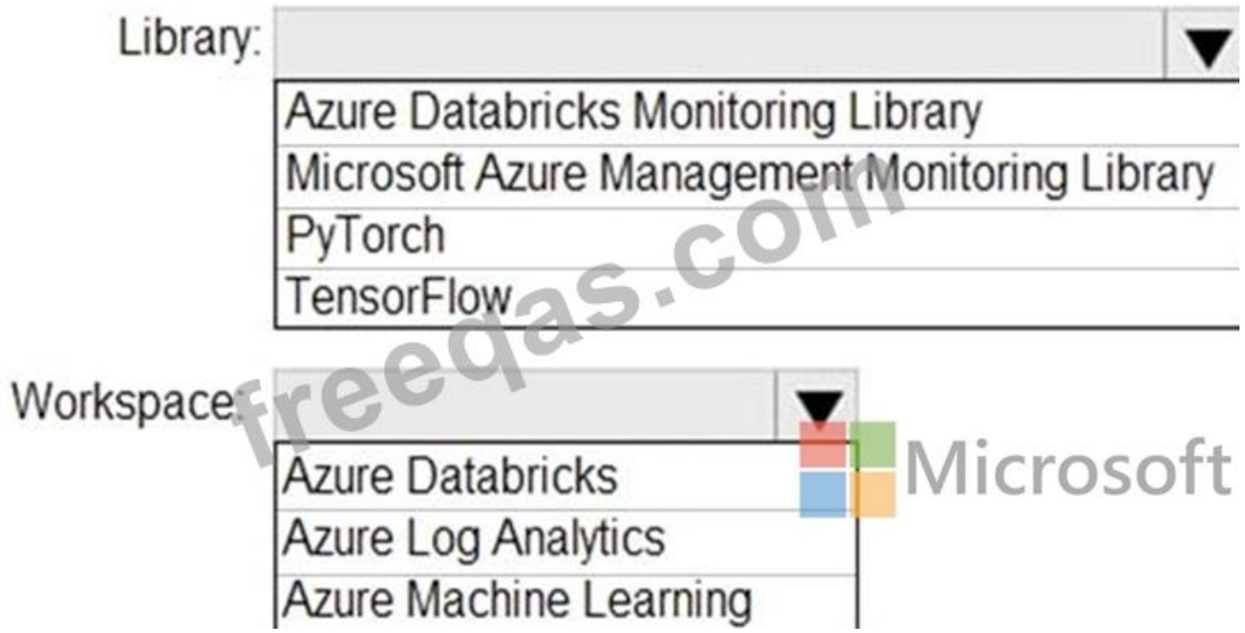
```
(20200101,20200201,20200301,20200401,20200501,20200601)
```


Answer:

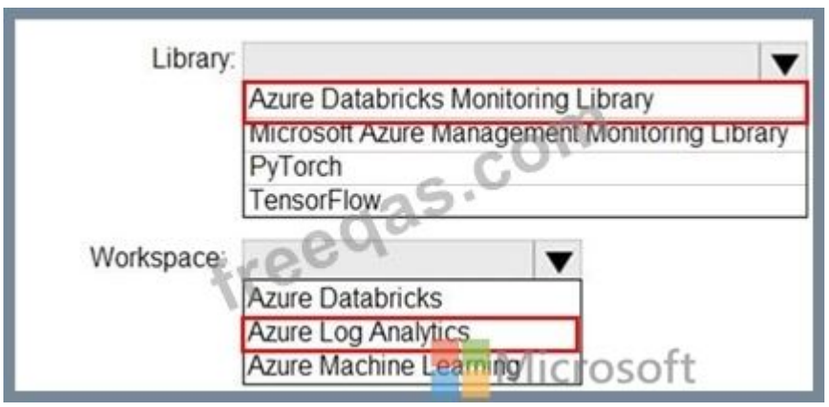


Reference:
https://docs.microsoft.com/en-us/sql/t-sql/statements/create-partition-function-transact-sql

NEW QUESTION: 175
You need to collect application metrics, streaming query events, and application log messages for an Azure Databrick cluster.
Which type of library and workspace should you implement? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.



Answer:



References:
https://docs.microsoft.com/en-us/azure/architecture/databricks-monitoring/application-logs

NEW QUESTION: 176

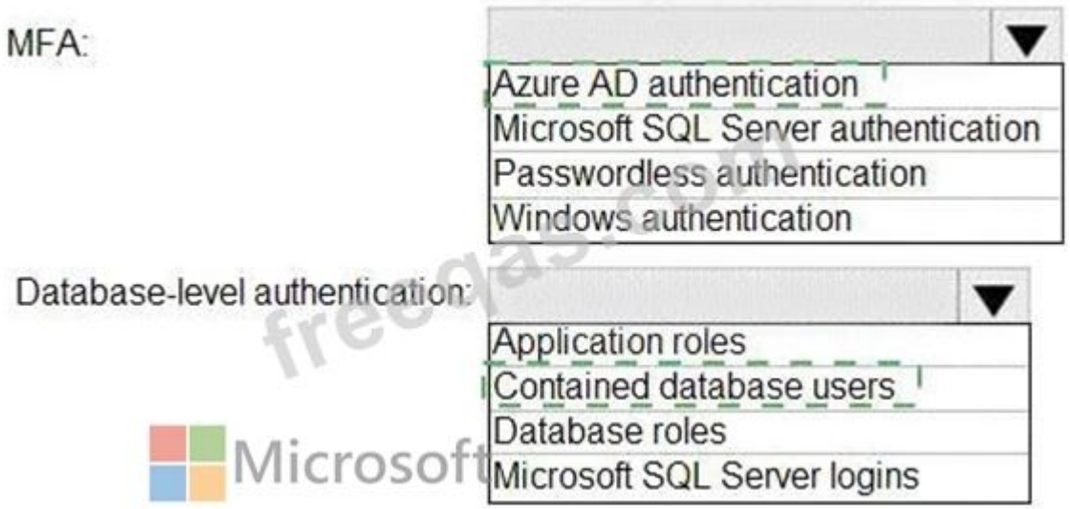
You have an Azure subscription that is linked to a hybrid Azure Active Directory (Azure AD) tenant. The subscription contains an Azure Synapse Analytics SQL pool named Pool1.

You need to recommend an authentication solution for Pool1. The solution must support multi-factor authentication (MFA) and database-level authentication.

Which authentication solution or solutions should you include in the recommendation? To answer, select the appropriate options in the answer area.

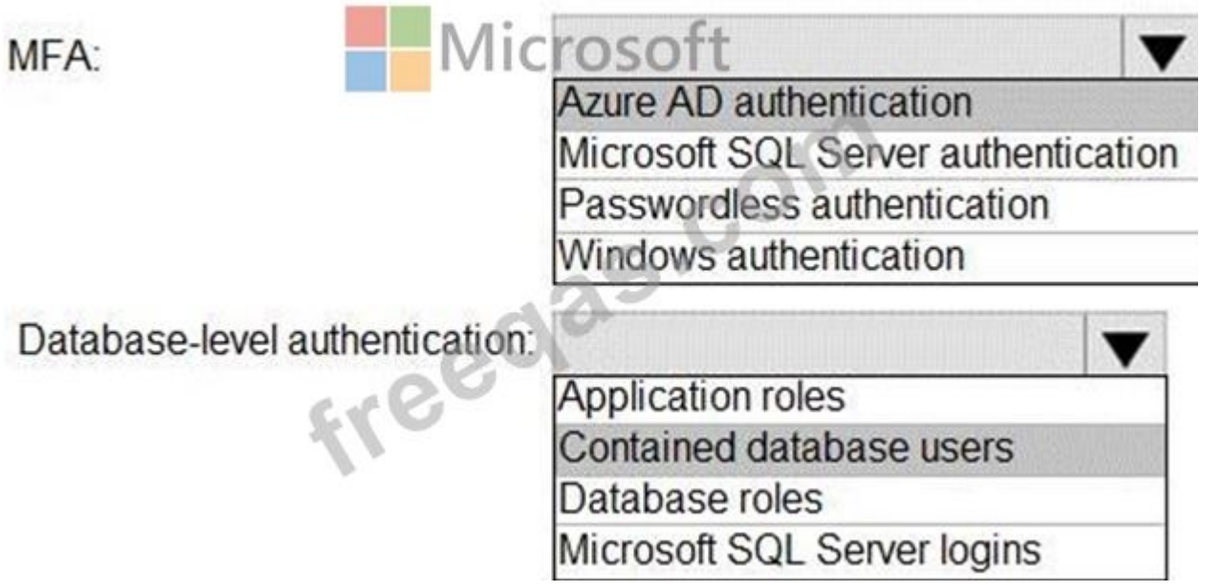
NOTE: Each correct selection is worth one point.

Answer:



Explanation

Graphical user interface, text, application, chat or text message Description automatically generated

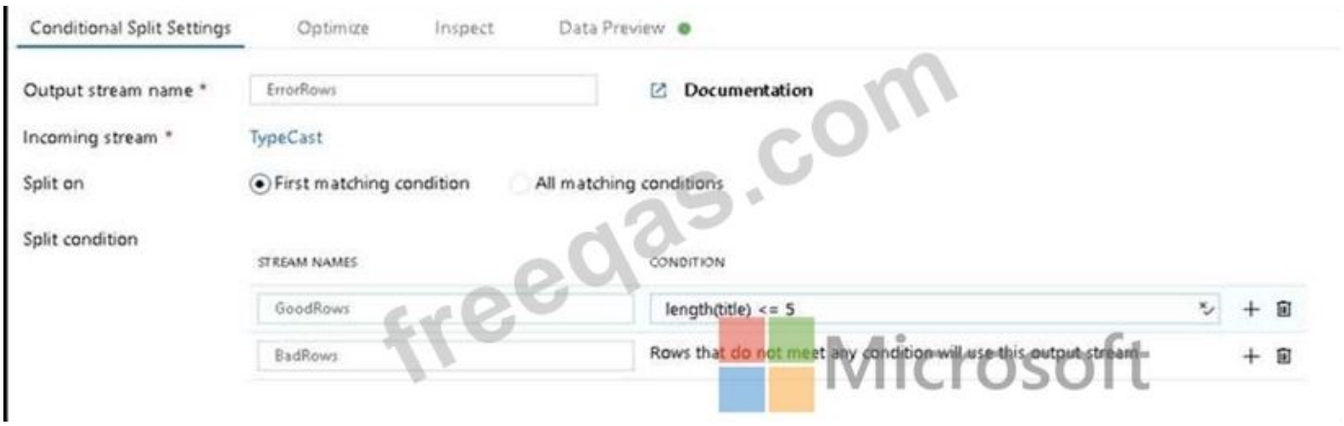


Box 1: Azure AD authentication
Azure Active Directory authentication supports Multi-Factor authentication through Active Directory Universal Authentication.
Box 2: Contained database users
Azure Active Directory Uses contained database users to authenticate identities at the database level.
Reference:
https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-authentication

NEW QUESTION: 177
You create an Azure Databricks cluster and specify an additional library to install.
When you attempt to load the library to a notebook, the library in not found.
You need to identify the cause of the issue.
What should you review?
A. notebook logs
B. cluster event logs
C. global init scripts logs
D. workspace logs
Answer: (SHOW ANSWER)
Cluster-scoped Init Scripts: Init scripts are shell scripts that run during the startup of each cluster node before the Spark driver or worker JVM starts. Databricks customers use init scripts for various purposes such as installing custom libraries, launching background processes, or applying enterprise security policies.
Logs for Cluster-scoped init scripts are now more consistent with Cluster Log Delivery and can be found in the same root folder as driver and executor logs for the cluster.
Reference:
https://databricks.com/blog/2018/08/30/introducing-cluster-scoped-init-scripts.html

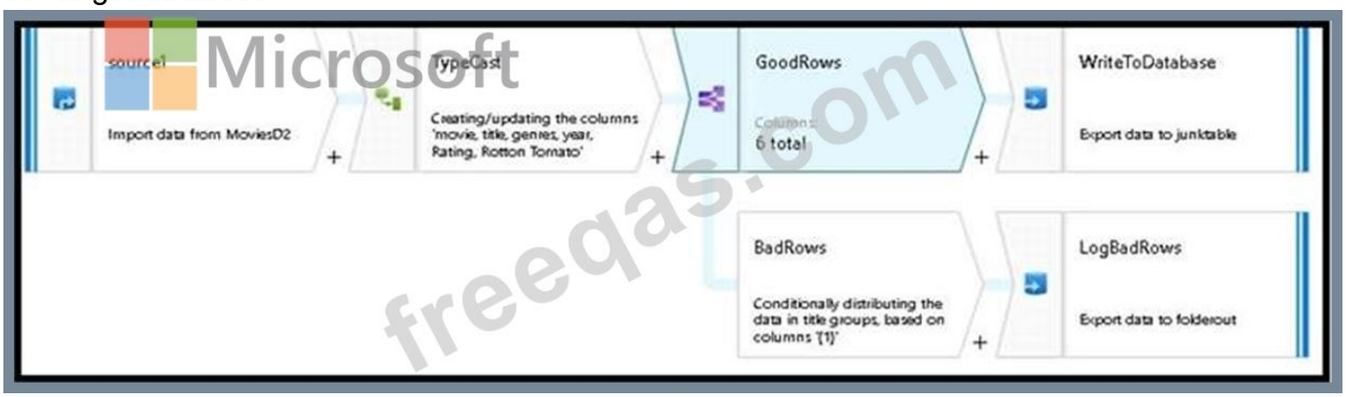
NEW QUESTION: 178
You are creating an Azure Data Factory data flow that will ingest data from a CSV file, cast columns to specified types of data, and insert the data into a table in an Azure Synapse Analytic dedicated SQL pool.
The CSV file contains three columns named username, comment, and date.
The data flow already contains the following:
A source transformation.
A Derived Column transformation to set the appropriate types of dat
a.
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. To the data flow, add a sink transformation to write the rows to a file in blob storage.
B. To the data flow, add a Conditional Split transformation to separate the rows that will cause truncation errors.
C. To the data flow, add a filter transformation to filter out rows that will cause truncation errors.
D. Add a select transformation to select only the rows that will cause truncation errors.
Answer: A,B (LEAVE A REPLY)
B: Example:
1. This conditional split transformation defines the maximum length of "title" to be five. Any row that is less than or equal to five will go into the GoodRows stream. Any row that is larger than five will go into the BadRows stream.



2. This conditional split transformation defines the maximum length of "title" to be five. Any row that is less than or equal to five will go into the GoodRows stream. Any row that is larger than five will go into the BadRows stream.
A:
3. Now we need to log the rows that failed. Add a sink transformation to the BadRows stream for logging. Here, we'll "auto-map" all of the fields so that we have logging of the complete transaction record. This is a text-delimited CSV file output to a single file in Blob Storage. We'll call the log file "badrows.csv".

4. The completed data flow is shown below. We are now able to split off error rows to avoid the SQL truncation errors and put those entries into a log file. Meanwhile, successful rows can continue to write to our target database.



Reference:
https://docs.microsoft.com/en-us/azure/data-factory/how-to-data-flow-error-rows

NEW QUESTION: 179
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. foreign key

B. effective start date

C. surrogate primary key

D. business key

E. last modified date

F. effective end date

Answer: A,B,D (LEAVE A REPLY)

NEW QUESTION: 180

You have two Azure Storage accounts named Storage1 and Storage2. Each account holds one container and has the hierarchical namespace enabled. The system has files that contain data stored in the Apache Parquet format.

You need to copy folders and files from Storage1 to Storage2 by using a Data Factory copy activity. The solution must meet the following requirements:

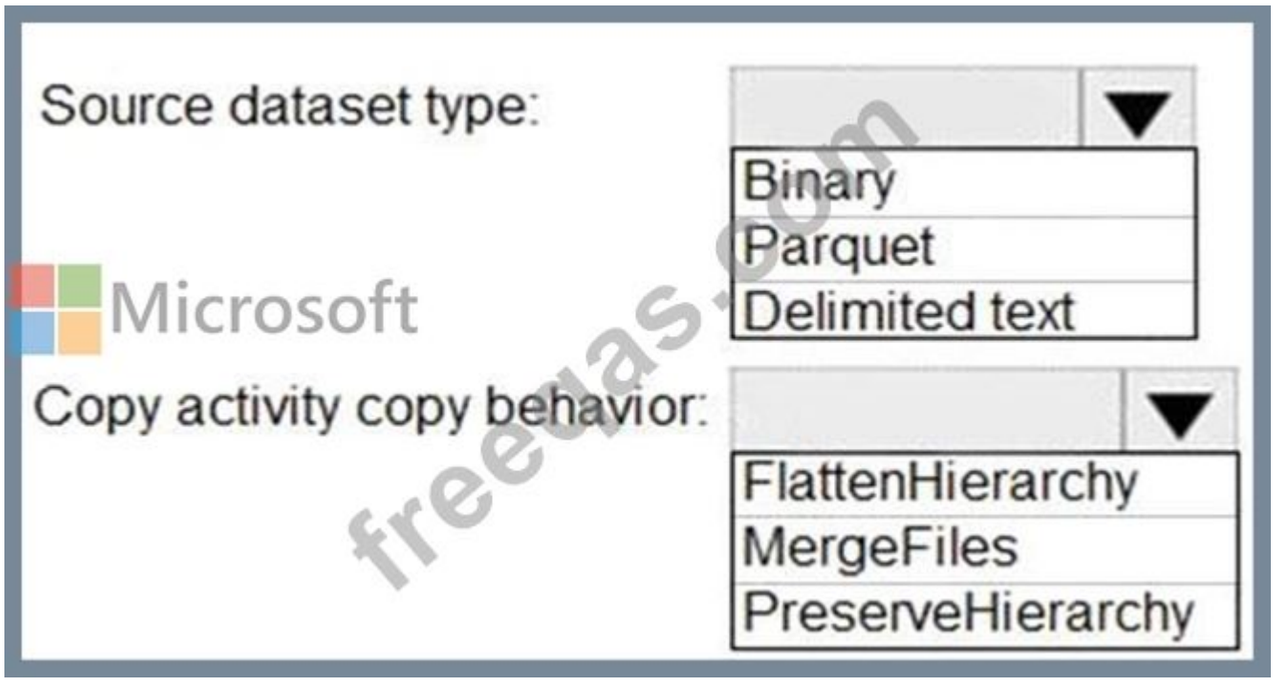
* No transformations must be performed.

* The original folder structure must be retained.

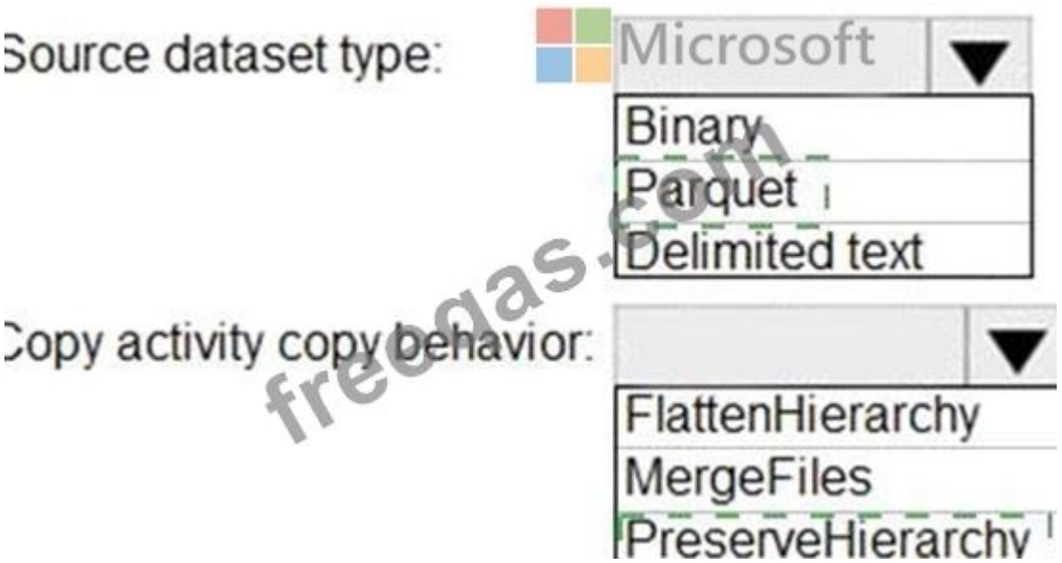
* Minimize time required to perform the copy activity.

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

NOTE: Each correct selection is worth one point.



Answer:

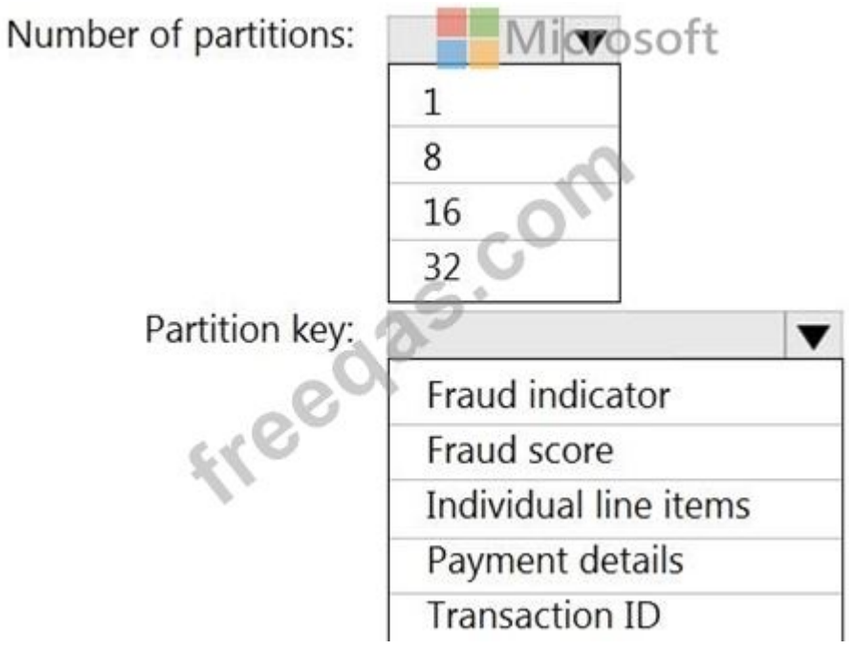


Explanation

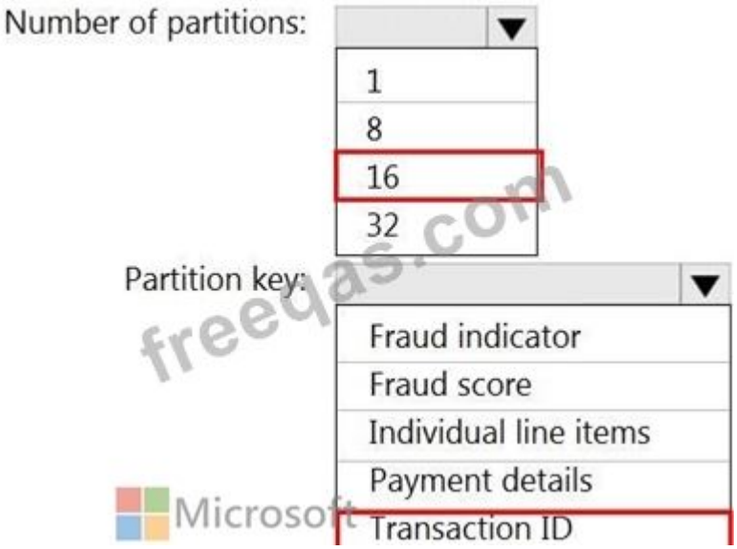
Graphical user interface, text, application, chat or text message Description automatically generated

Box 1: Parquet
For Parquet datasets, the type property of the copy activity source must be set to ParquetSource.
Box 2: PreserveHierarchy
PreserveHierarchy (default): Preserves the file hierarchy in the target folder. The relative path of the source file to the source folder is identical to the relative path of the target file to the target folder.
Reference:
https://docs.microsoft.com/en-us/azure/data-factory/format-parquet
https://docs.microsoft.com/en-us/azure/data-factory/connector-azure-data-lake-storage

NEW QUESTION: 181
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.



Answer:



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

NEW QUESTION: 182
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 dat a.
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

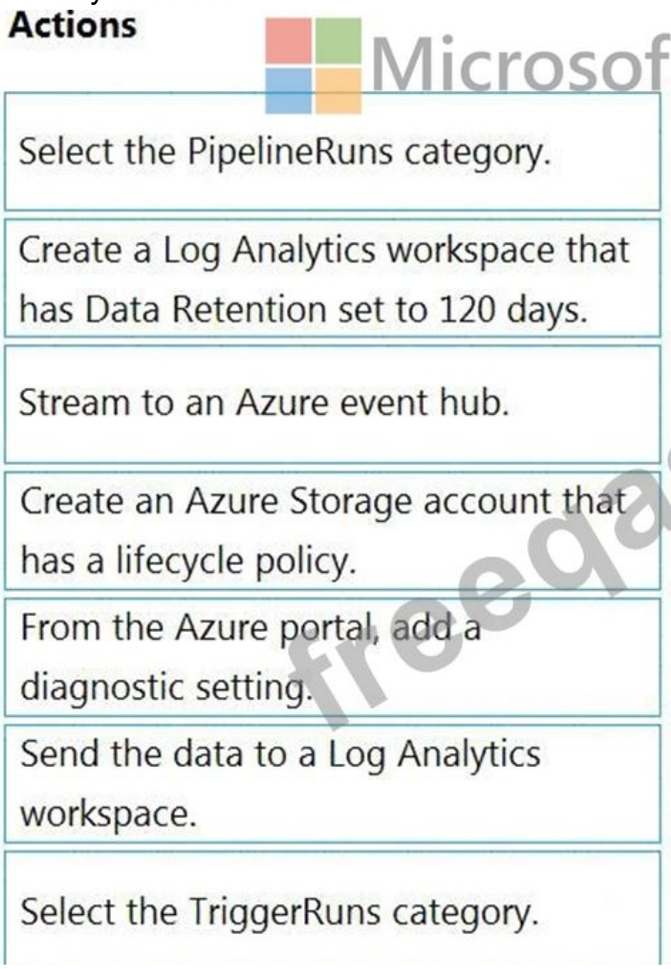
NEW QUESTION: 183

You have an Azure data factory.

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

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

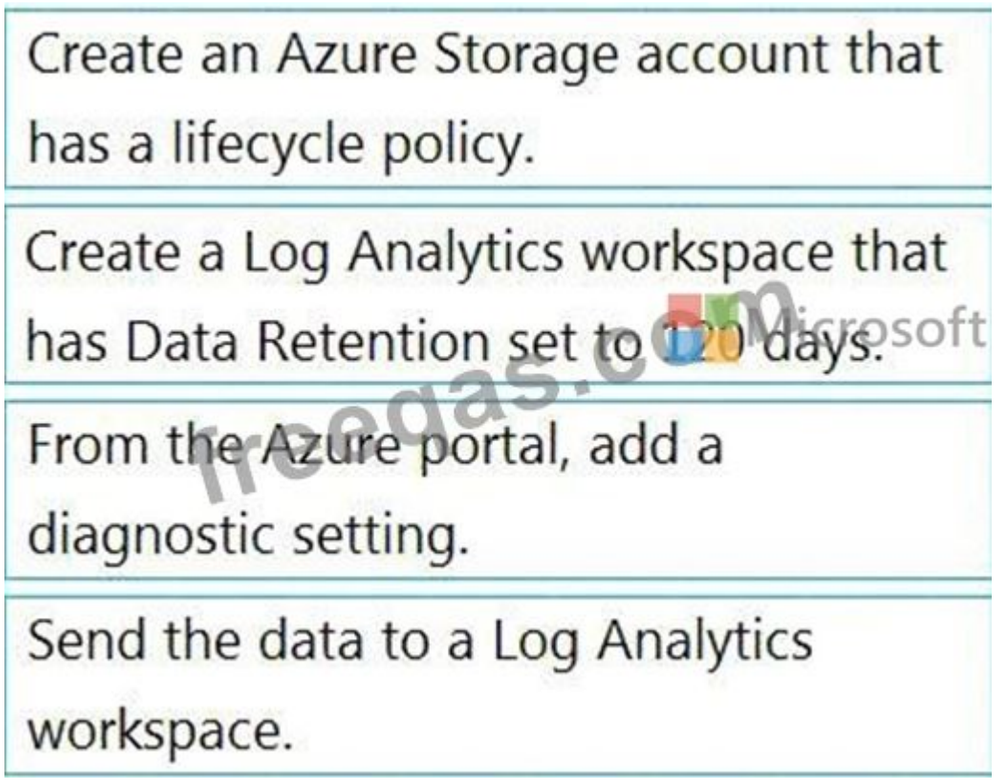
NOTE: More than one order of answer choices is correct. You will receive credit for any of the correct orders you select.



Answer:

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

Explanation



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

To automate common data management tasks, Microsoft created a solution based on Azure Data Factory. The service, Data Lifecycle Management, makes frequently accessed data available and archives or purges other data according to retention policies. Teams across the company use the service to reduce storage costs, improve app performance, and comply with data retention policies.
Step 2: Create a Log Analytics workspace that has Data Retention set to 120 days.
Data Factory stores pipeline-run data for only 45 days. Use Azure Monitor if you want to keep that data for a longer time. With Monitor, you can route diagnostic logs for analysis to multiple different targets, such as a Storage Account: Save your diagnostic logs to a storage account for auditing or manual inspection. You can use the diagnostic settings to specify the retention time in days.
Step 3: From Azure Portal, add a diagnostic setting.
Step 4: Send the data to a log Analytics workspace,
Event Hub: A pipeline that transfers events from services to Azure Data Explorer.
Keeping Azure Data Factory metrics and pipeline-run data.
Configure diagnostic settings and workspace.
Create or add diagnostic settings for your data factory.
* In the portal, go to Monitor. Select Settings > Diagnostic settings.
* Select the data factory for which you want to set a diagnostic setting.
* If no settings exist on the selected data factory, you're prompted to create a setting. Select Turn on diagnostics.
* Give your setting a name, select Send to Log Analytics, and then select a workspace from Log Analytics Workspace.
* Select Save.
Reference:
https://docs.microsoft.com/en-us/azure/data-factory/monitor-using-azure-monitor

NEW QUESTION: 184
You are designing an Azure Data Lake Storage Gen2 container to store data for the human resources (HR) department and the operations department at your company. You have the following data access requirements:
* After initial processing, the HR department data will be retained for seven years.
* The operations department data will be accessed frequently for the first six months, and then accessed once per month.
You need to design a data retention solution to meet the access requirements. The solution must minimize storage costs.
Answer:
See the answer in explanation.
Explanation
Answer is below

NEW QUESTION: 185

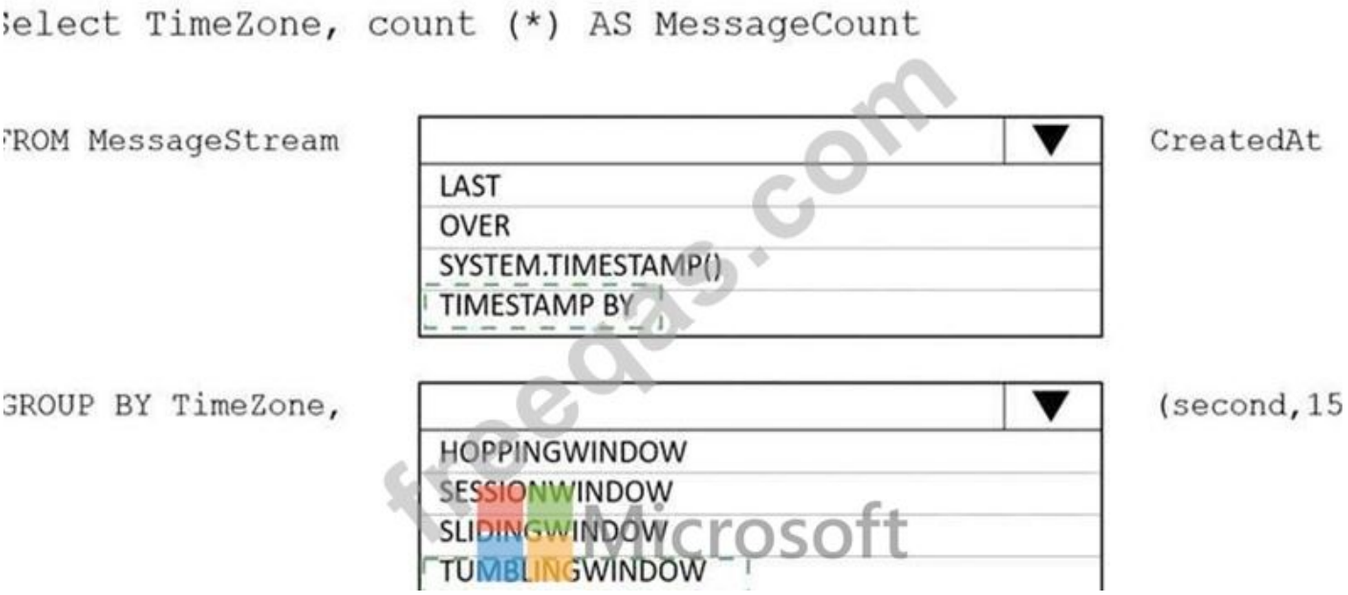
You are designing an Azure Stream Analytics solution that receives instant messaging data from an Azure Event Hub.

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.

Answer:



Explanation

Table Description automatically generated

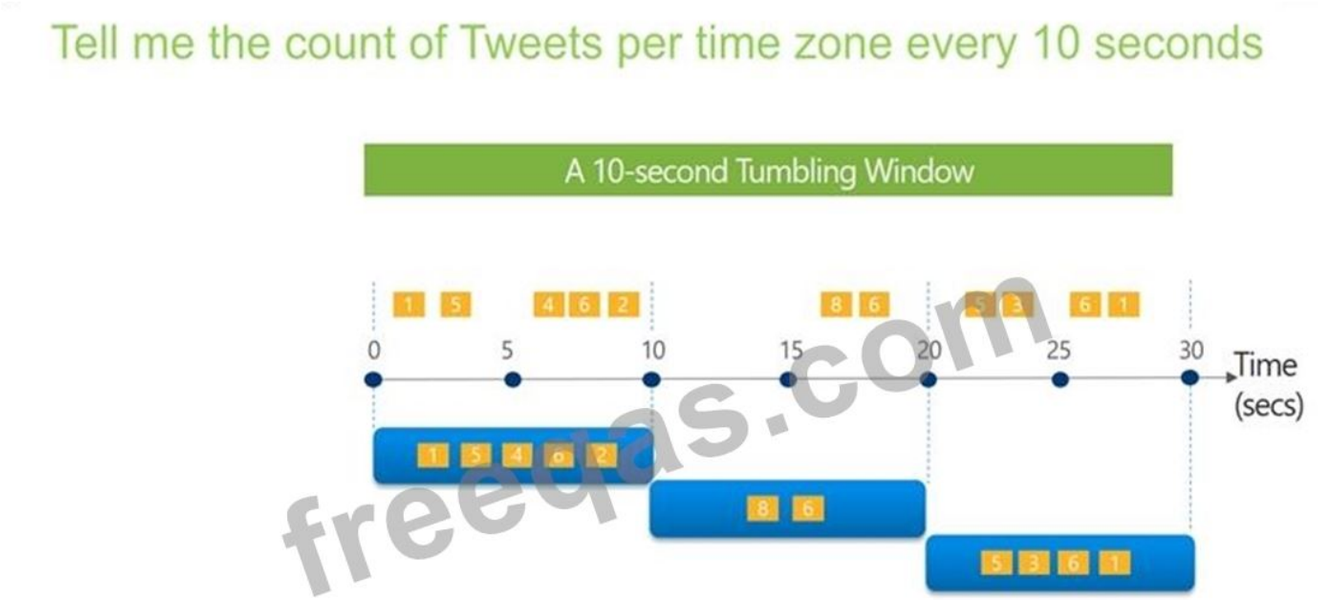
```
Select TimeZone, count (*) AS MessageCount

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.

Timeline Description automatically generated



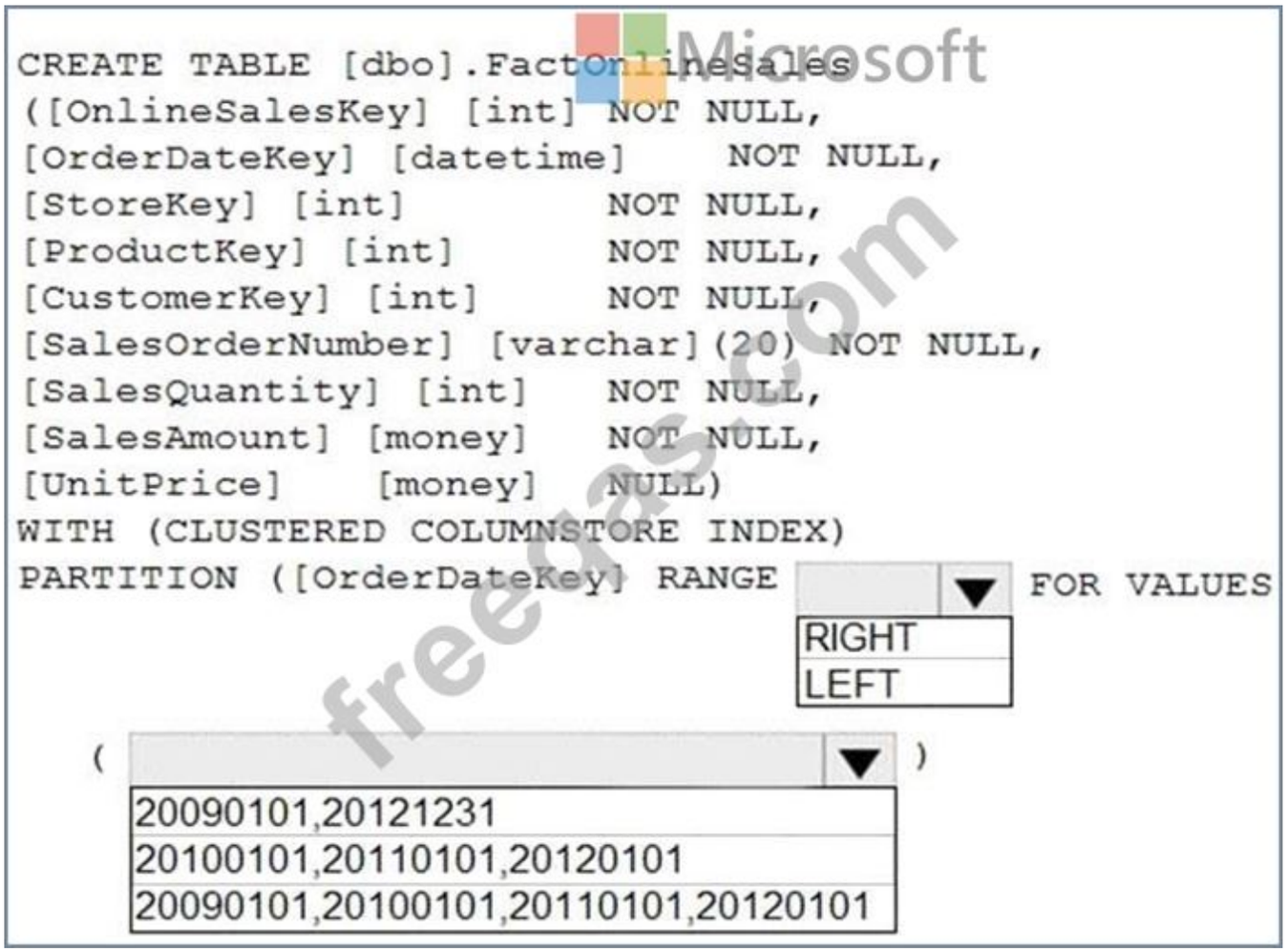

```
SELECT TimeZone, COUNT(*) AS Count
FROM TwitterStream TIMESTAMP BY CreatedAt
GROUP BY TimeZone, TumblingWindow(second,10)
```

Reference:

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

NEW QUESTION: 186
You have a data warehouse in Azure Synapse Analytics.
You need to ensure that the data in the data warehouse is encrypted at rest.
What should you enable?
A. Advanced Data Security for this database
B. Transparent Data Encryption (TDE)
C. Secure transfer required
D. Dynamic Data Masking
Answer: B (LEAVE A REPLY)
Azure SQL Database currently supports encryption at rest for Microsoft-managed service side and client-side encryption scenarios.
Support for server encryption is currently provided through the SQL feature called Transparent Data Encryption.
Client-side encryption of Azure SQL Database data is supported through the Always Encrypted feature.
Reference:
https://docs.microsoft.com/en-us/azure/security/fundamentals/encryption-atrest

NEW QUESTION: 187
You have an enterprise data warehouse in Azure Synapse Analytics that contains a table named FactOnlineSales. The table contains data from the start of 2009 to the end of 2012.
You need to improve the performance of queries against FactOnlineSales by using table partitions. The solution must meet the following requirements:
* Create four partitions based on the order date.
* Ensure that each partition contains all the orders places during a given calendar year.
How should you complete the T-SQL command? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.


CREATE TABLE [dbo].FactOnlineSales
([OnlineSalesKey] [int] NOT NULL,
[OrderDateKey] [datetime] NOT NULL,
[StoreKey] [int] NOT NULL,
[ProductKey] [int] NOT NULL,
[CustomerKey] [int] NOT NULL,
[SalesOrderNumber] [varchar](20) NOT NULL,
[SalesQuantity] [int] NOT NULL,
[SalesAmount] [money] NOT NULL,
[UnitPrice] [money] NULL)
WITH (CLUSTERED COLUMNSTORE INDEX)
PARTITION ([OrderDateKey] RANGE
RIGHT
LEFT
FOR VALUES
(
20090101,20121231
20100101,20110101,20120101
20090101,20100101,20110101,20120101
)

Answer:

```
CREATE TABLE [dbo].FactOnlineSales
([OnlineSalesKey] [int] NOT NULL,
[OrderDateKey] [datetime]    NOT NULL,
[StoreKey] [int]         NOT NULL,
[ProductKey] [int]       NOT NULL,
[CustomerKey] [int]      NOT NULL,
[SalesOrderNumber] [varchar](20) NOT NULL,
[SalesQuantity] [int]    NOT NULL,
[SalesAmount] [money]    NOT NULL,
[UnitPrice]    [money]   NULL)
WITH (CLUSTERED COLUMNSTORE INDEX)
PARTITION ([OrderDateKey] RANGE [ ▼ ] FOR VALUES
```

[▼]
- RIGHT
- LEFT

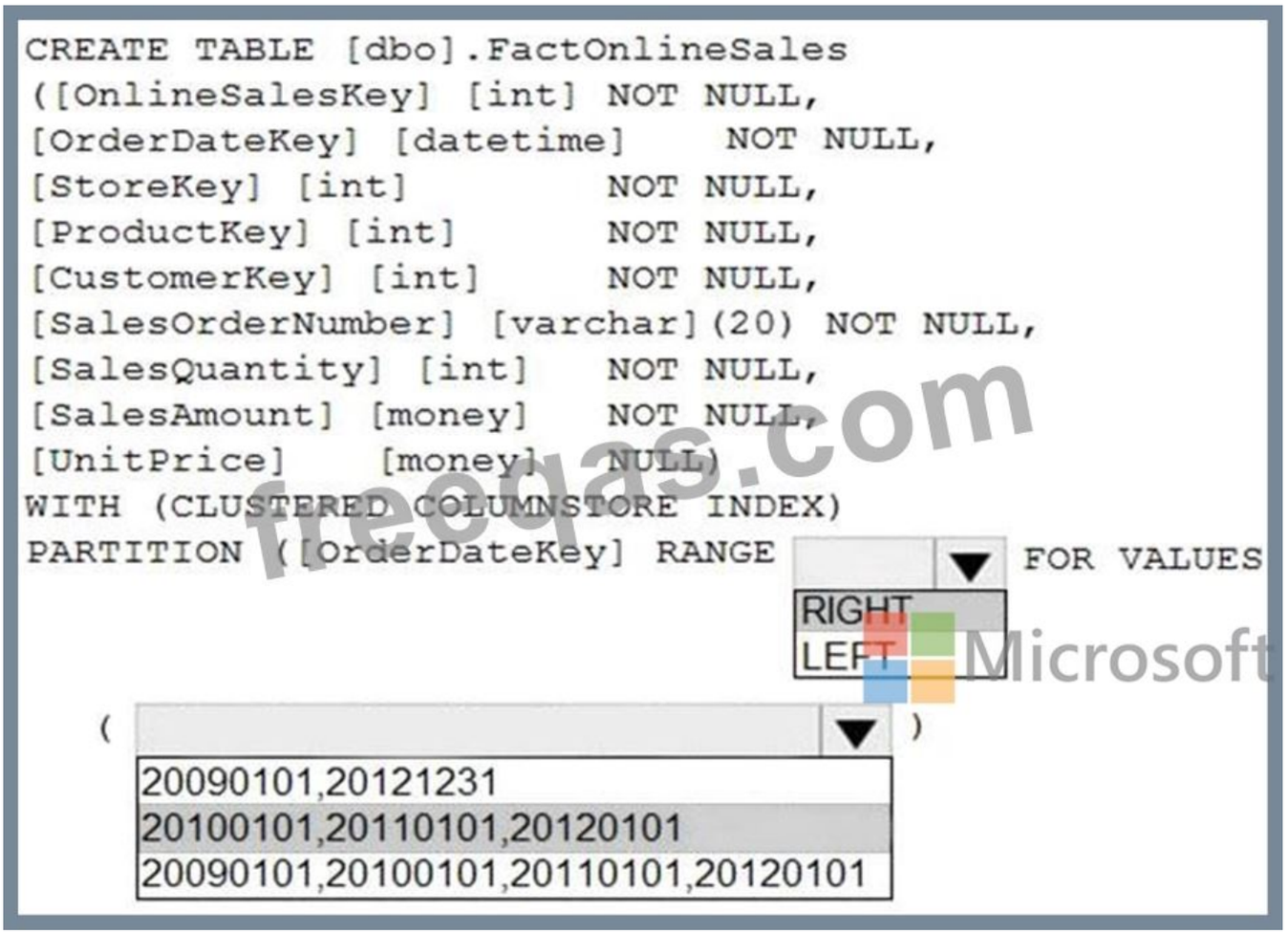
(

[▼]
- 20090101,20121231
- 20100101,20110101,20120101
- 20090101,20100101,20110101,20120101

)

Explanation

Text Description automatically generated



Range Left or Right, both are creating similar partition but there is difference in comparison For example: in this scenario, when you use LEFT and 20100101,20110101,20120101 Partition will be, datecol<=20100101, datecol>20100101 and datecol<=20110101, datecol>20110101 and datecol<=20120101, datecol>20120101 But if you use range RIGHT and 20100101,20110101,20120101 Partition will be, datecol<20100101, datecol>=20100101 and datecol<20110101, datecol>=20110101 and datecol<20120101, datecol>=20120101 In this example, Range RIGHT will be suitable for calendar comparison Jan 1st to Dec 31st Reference: https://docs.microsoft.com/en-us/sql/t-sql/statements/create-partition-function-transact-sql?view=sql-server-ver1

NEW QUESTION: 188

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 serwrless SQL pod.

Which format should you use for the tables in DB1?

A. JSON

B. CSV

C. ORC

D. Parquet

Answer: (SHOW ANSWER)

NEW QUESTION: 189

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: (SHOW ANSWER)

Explanation

Use the same definition as the EmployeeID column.

Reference:

https://docs.microsoft.com/en-us/analysis-services/tabular-models/hierarchies-ssas-tabular

NEW QUESTION: 190

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: 191
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: 192
You develop data engineering solutions for a company.
A project requires the deployment of data to Azure Data Lake Storage.
You need to implement role-based access control (RBAC) so that project members can manage the Azure Data Lake Storage resources.
Which three actions should you perform? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.
A. Assign Azure AD security groups to Azure Data Lake Storage.
B. Configure end-user authentication for the Azure Data Lake Storage account.

C. Configure service-to-service authentication for the Azure Data Lake Storage account.
D. Create security groups in Azure Active Directory (Azure AD) and add project members.
E. Configure access control lists (ACL) for the Azure Data Lake Storage account.
Answer: A,D,E (LEAVE A REPLY)
Explanation
References:
https://docs.microsoft.com/en-us/azure/data-lake-store/data-lake-store-secure-data

NEW QUESTION: 193
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. replicated
B. hash-distributed on PurchaseKey
C. round-robin

D. hash-distributed on DateKey

Answer: B (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.

Incorrect:

Not D: Do not use a date column. . All data for the same date lands in the same distribution. If several users are all filtering on the same date, then only 1 of the 60 distributions do all the processing work.

Reference:

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

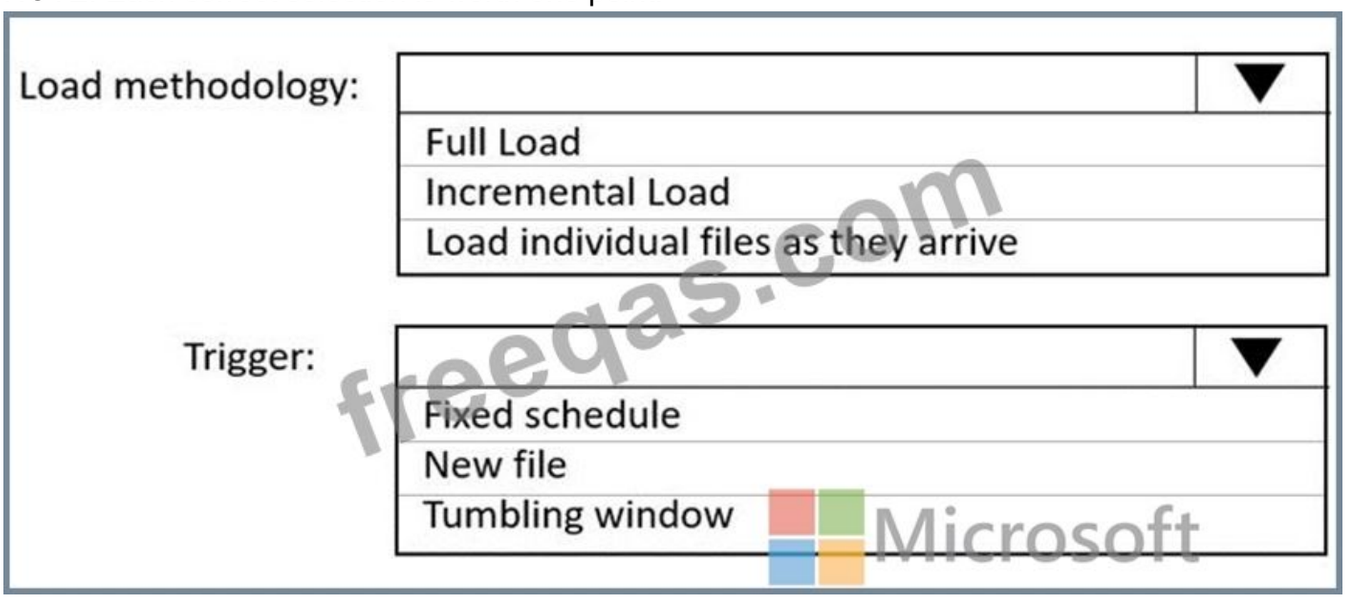
NEW QUESTION: 194

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 are a.

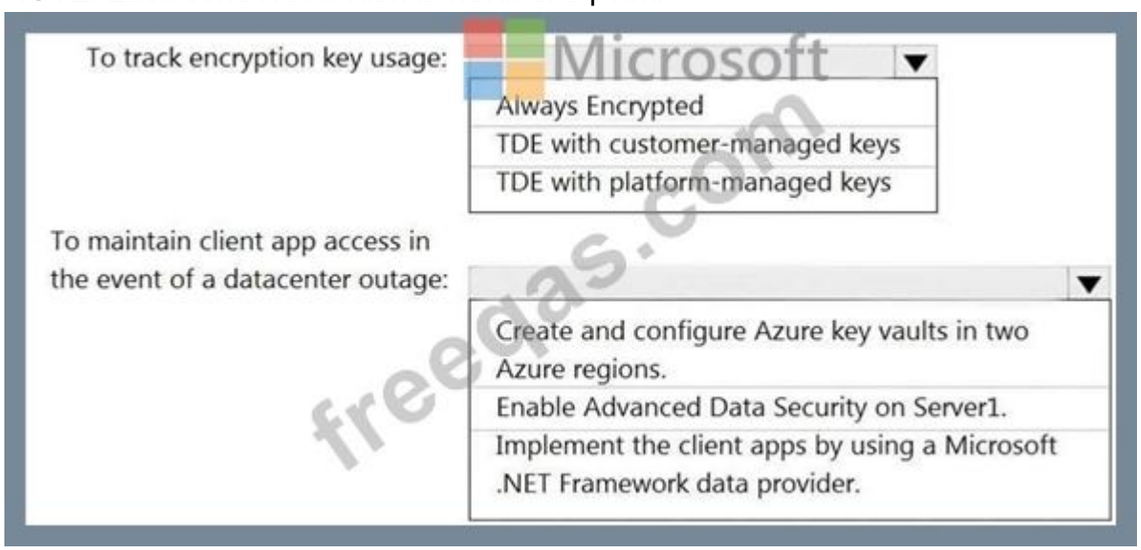
NOTE: Each correct selection is worth one point.



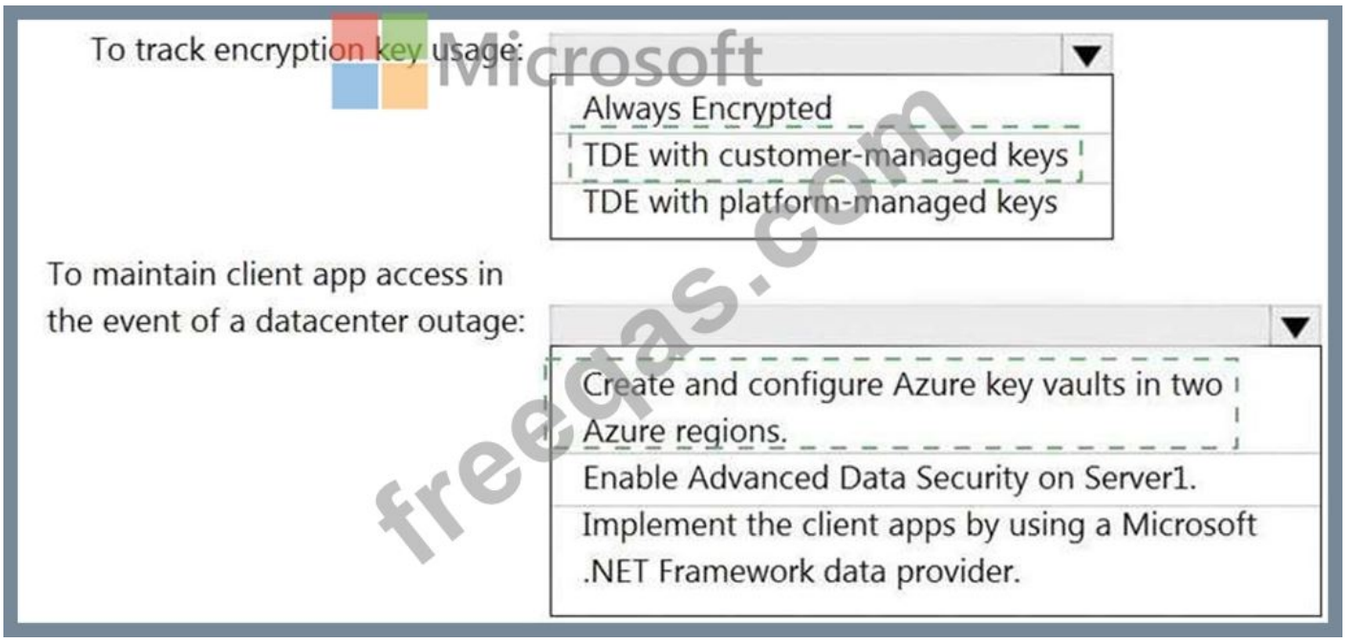
Answer:

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

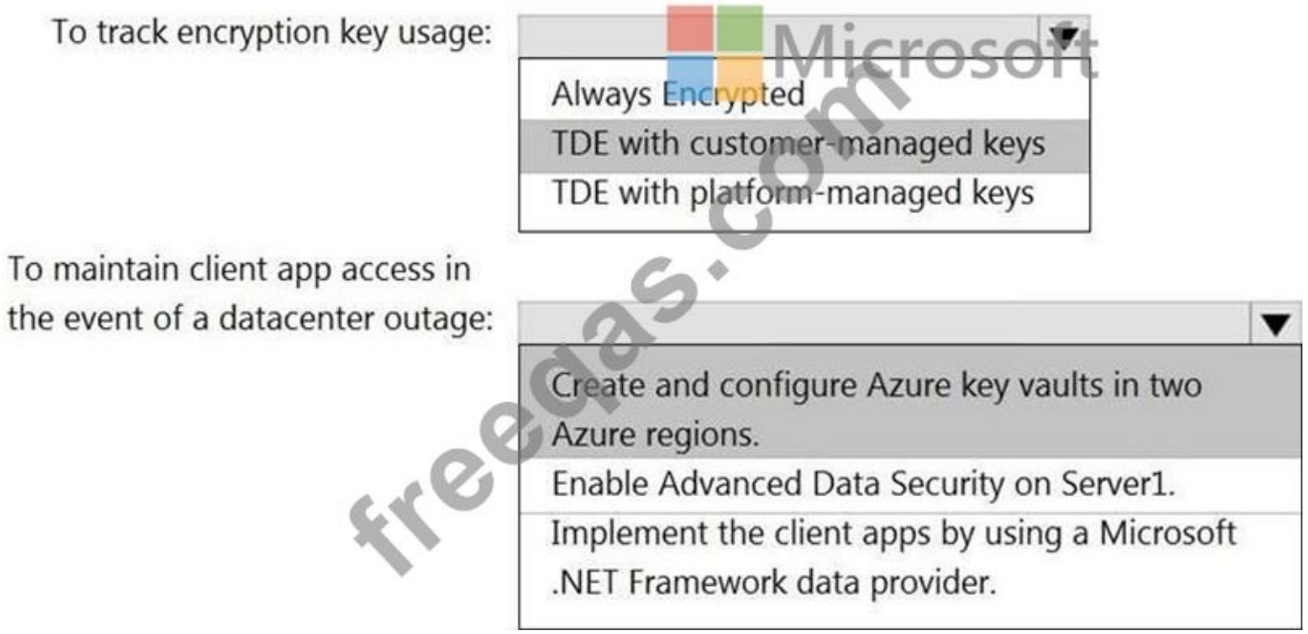
NEW QUESTION: 195
You have an Azure subscription that contains a logical Microsoft SQL server named Server1. Server1 hosts an Azure Synapse Analytics SQL dedicated pool named Pool1.
You need to recommend a Transparent Data Encryption (TDE) solution for Server1. The solution must meet the following requirements:
* Track the usage of encryption keys.
* Maintain the access of client apps to Pool1 in the event of an Azure datacenter outage that affects the availability of the encryption keys.
What should you include in the recommendation? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.



Answer:



Explanation



Box 1: TDE with customer-managed keys

Customer-managed keys are stored in the Azure Key Vault. You can monitor how and when your key vaults are accessed, and by whom. You can do this by enabling logging for Azure Key Vault, which saves information in an Azure storage account that you provide.

Box 2: Create and configure Azure key vaults in two Azure regions

The contents of your key vault are replicated within the region and to a secondary region at least 150 miles away, but within the same geography to maintain high durability of your keys and secrets.

Reference:

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

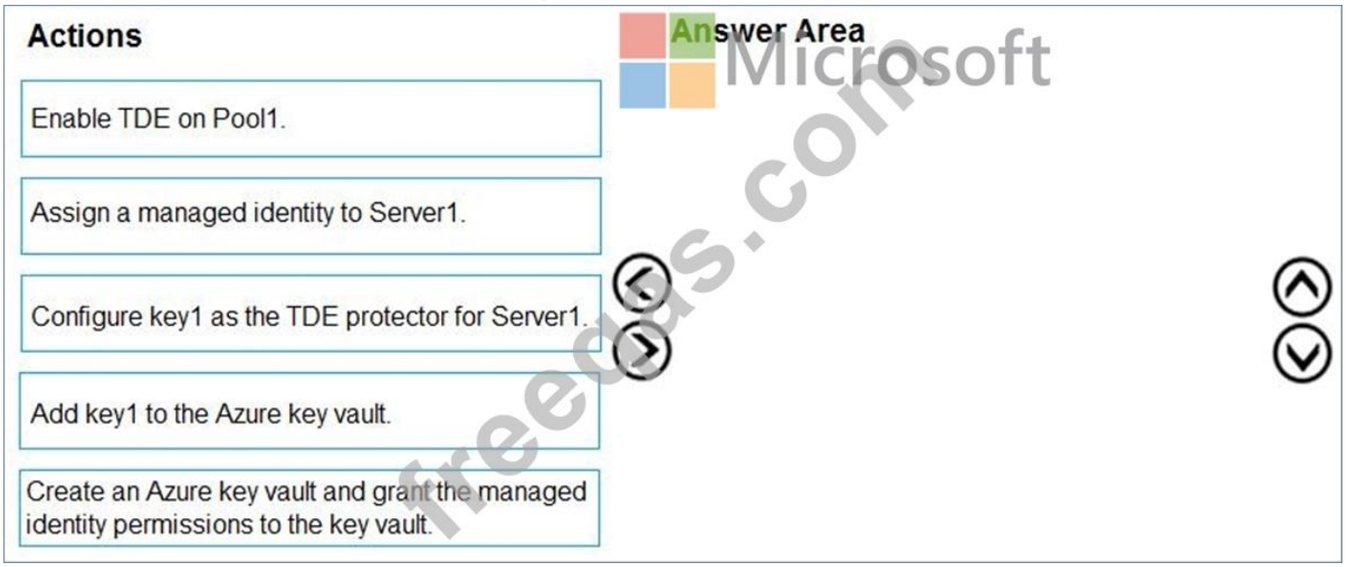
https://docs.microsoft.com/en-us/azure/key-vault/general/logging

NEW QUESTION: 196

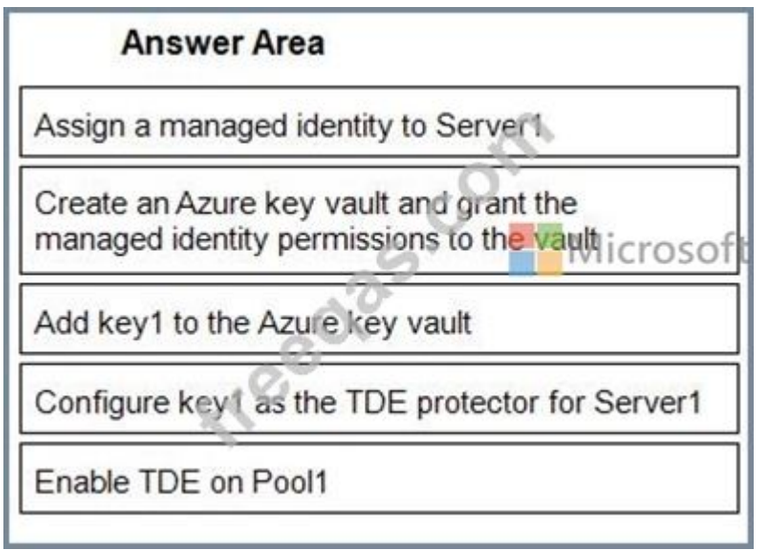
You have an Azure Synapse Analytics SQL pool named Pool1 on a logical Microsoft SQL server named Server1.

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.



Answer:



1 - Assign a managed identity to Server1

2 - Create an Azure key vault and grant the managed identity permissions to the vault

3 - Add key1 to the Azure key vault

4 - Configure key1 as the TDE protector for Server1

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: 197
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: (SHOW ANSWER)
Explanation
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

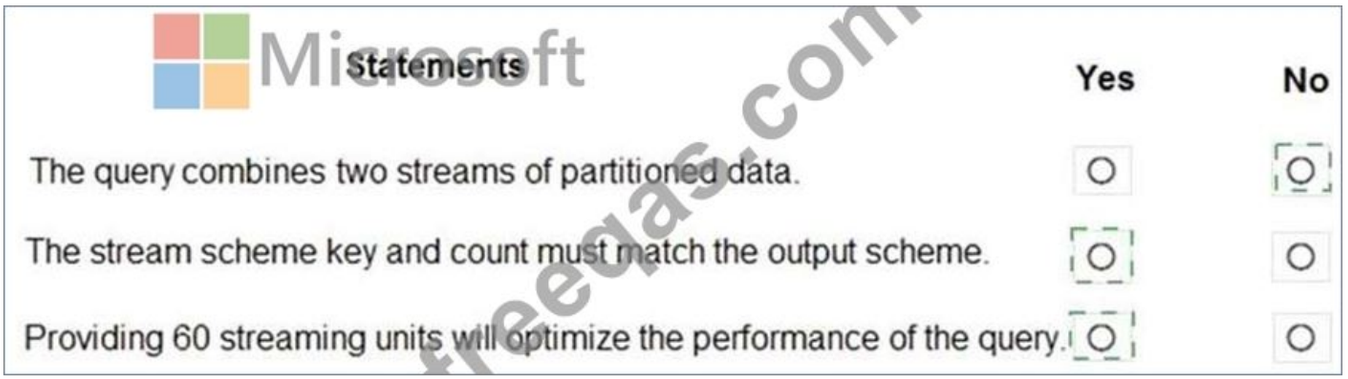
NEW QUESTION: 198
You have the following Azure Stream Analytics query.

```
WITH

step1 AS (SELECT *
      FROM input1
      PARTITION BY StateID
      INTO 10),
step2 AS (SELECT *
      FROM input2
      PARTITION BY StateID
      INTO 10)

SELECT *
INTO output
FROM step1
PARTITION BY StateID
UNION
SELECT * INTO output
       FROM step2
       PARTITION BY StateID
```

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
The query combines two streams of partitioned data.	○	○
The stream scheme key and count must match the output scheme.	○	○
Providing 60 streaming units will optimize the performance of the query.	○	○

Answer:



Explanation

Box 1: No

Note: You can now use a new extension of Azure Stream Analytics SQL to specify the number of partitions of a stream when reshuffling the data.

The outcome is a stream that has the same partition scheme. Please see below for an example:

WITH step1 AS (SELECT * FROM [input1] PARTITION BY DeviceID INTO 10),

step2 AS (SELECT * FROM [input2] PARTITION BY DeviceID INTO 10)

SELECT * INTO [output] FROM step1 PARTITION BY DeviceID UNION step2 PARTITION BY DeviceID

Note: The new extension of Azure Stream Analytics SQL includes a keyword INTO that allows you to specify the number of partitions for a stream when performing reshuffling using a PARTITION BY statement.

Box 2: Yes

When joining two streams of data explicitly repartitioned, these streams must have the same partition key and partition count.

Box 3: Yes

Streaming Units (SUs) represents the computing resources that are allocated to execute a Stream Analytics job. The higher the number of SUs, the more CPU and memory resources are allocated for your job.

In general, the best practice is to start with 6 SUs for queries that don't use PARTITION BY.

Here there are 10 partitions, so 6x10 = 60 SUs is good.

Note: Remember, Streaming Unit (SU) count, which is the unit of scale for Azure Stream Analytics, must be adjusted so the number of physical resources available to the job can fit the partitioned flow. In general, six SUs is a good number to assign to each partition. In case there are insufficient resources assigned to the job, the system will only apply the repartition if it benefits the job.

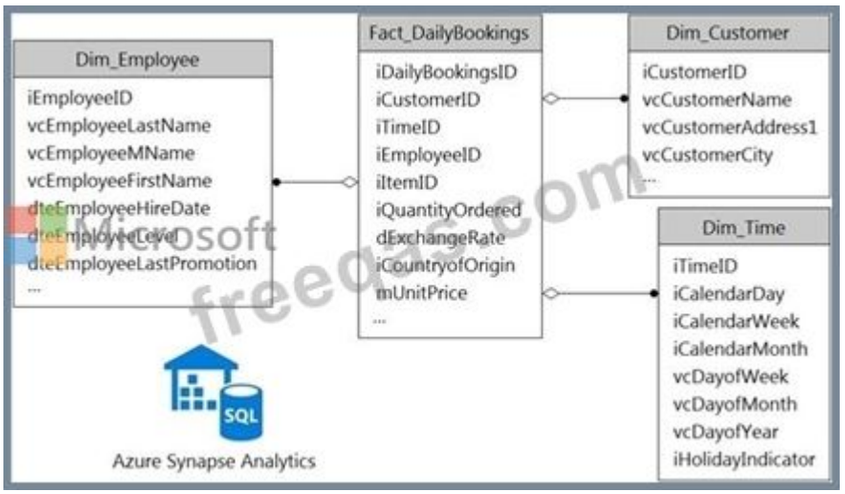
Reference:

https://azure.microsoft.com/en-in/blog/maximize-throughput-with-repartitioning-in-azure-stream-analytics/

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-streaming-unit-consumption

NEW QUESTION: 199

You have a data model that you plan to implement in a data warehouse in Azure Synapse Analytics as shown in the following exhibit.

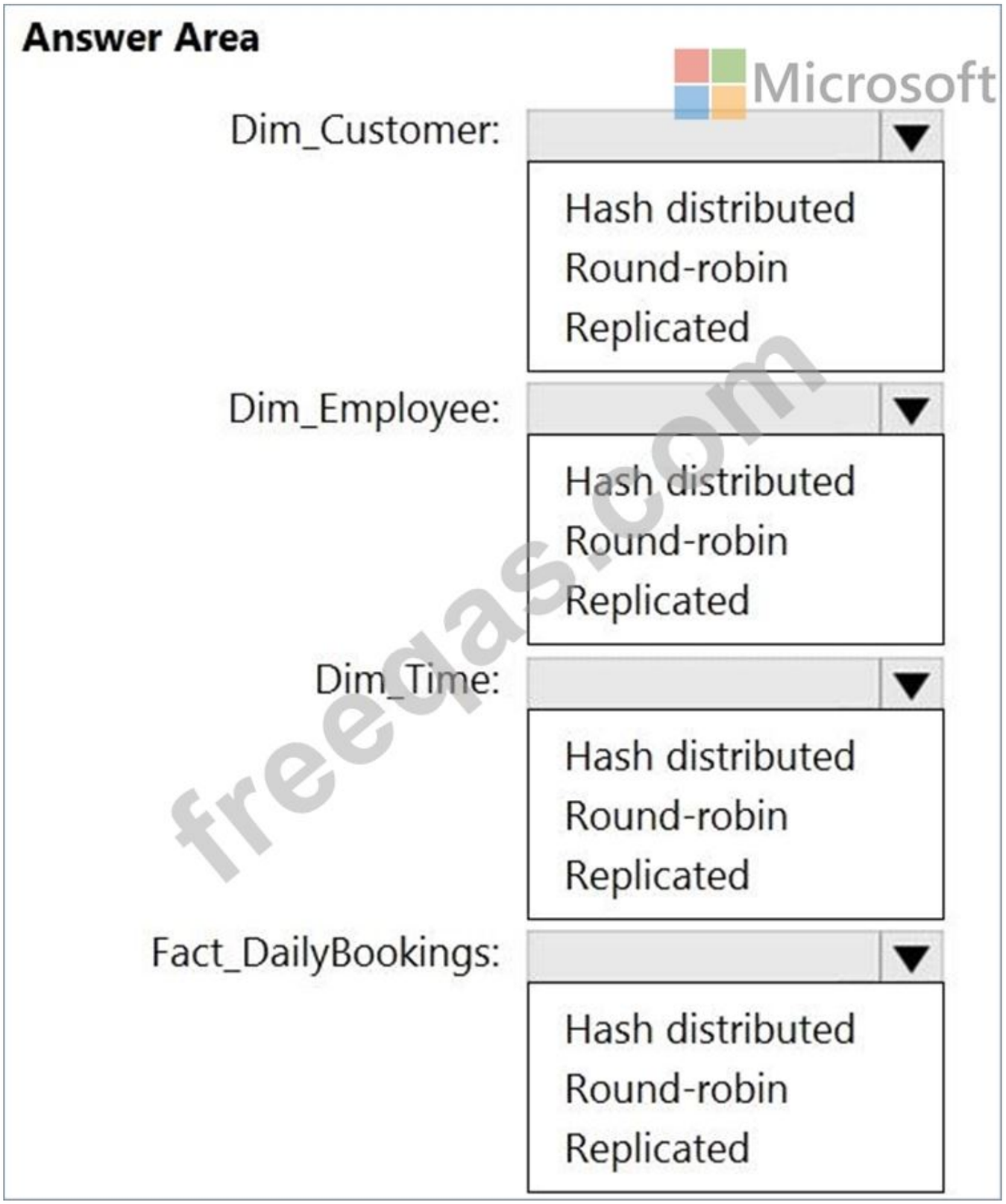


All the dimension tables will be less than 2 GB after compression, and the fact table will be approximately 6 TB.

Which type of table should you use for each table? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area



Answer:

NEW QUESTION: 200

You have an Azure Data Lake Storage Gen2 container that contains 100 TB of data.

You need to ensure that the data in the container is available for read workloads in a secondary region if an outage occurs in the primary region. The solution must minimize costs.

Which type of data redundancy should you use?

A. locally-redundant storage (LRS)

B. geo-redundant storage (GRS)

C. zone-redundant storage (ZRS)

D. read-access geo-redundant storage (RA-GRS)

Answer: A (LEAVE A REPLY)

NEW QUESTION: 201

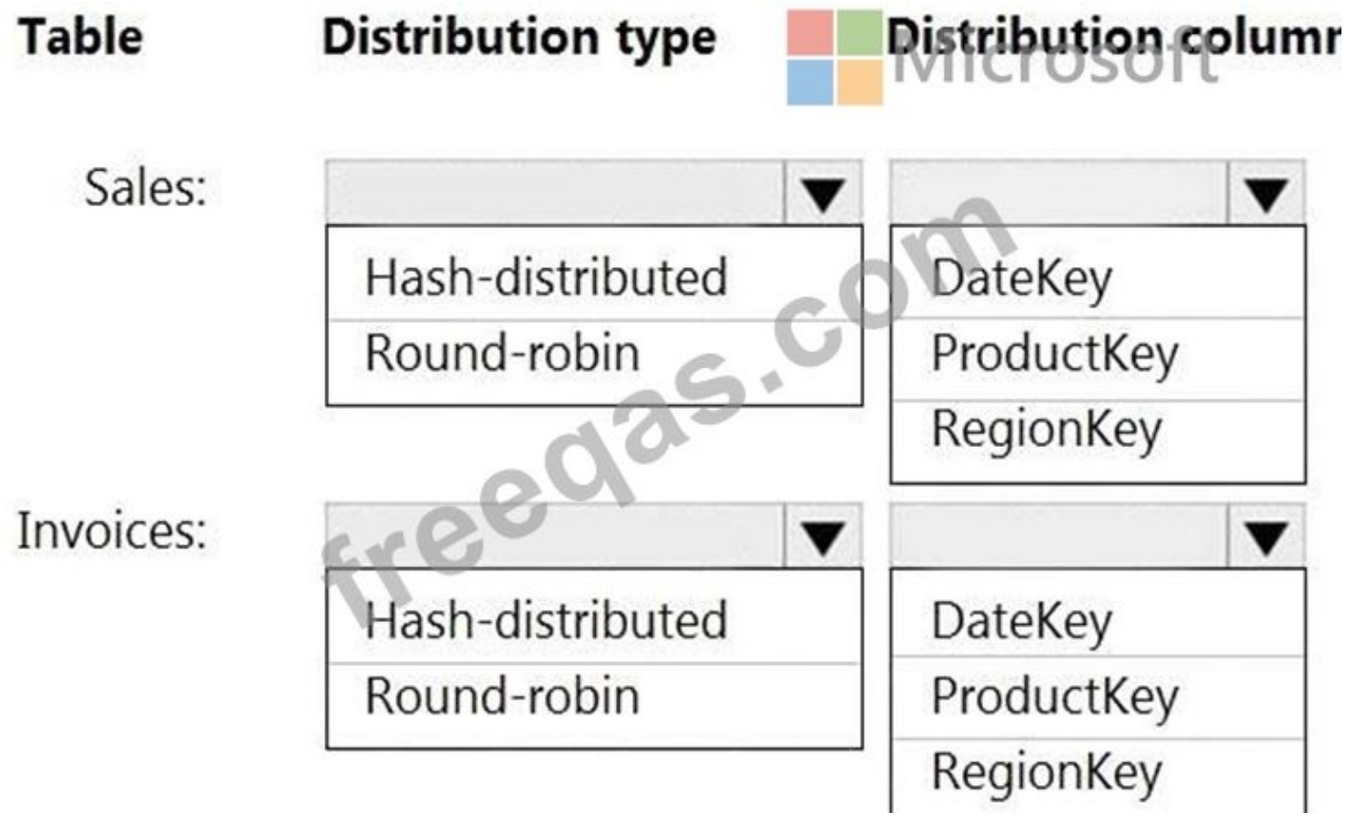
You have an Azure Data Lake Storage account that contains a staging zone.

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 info the data warehouse.
Does this meet the goal?
A. Yes
B. No
Answer: B (LEAVE A REPLY)

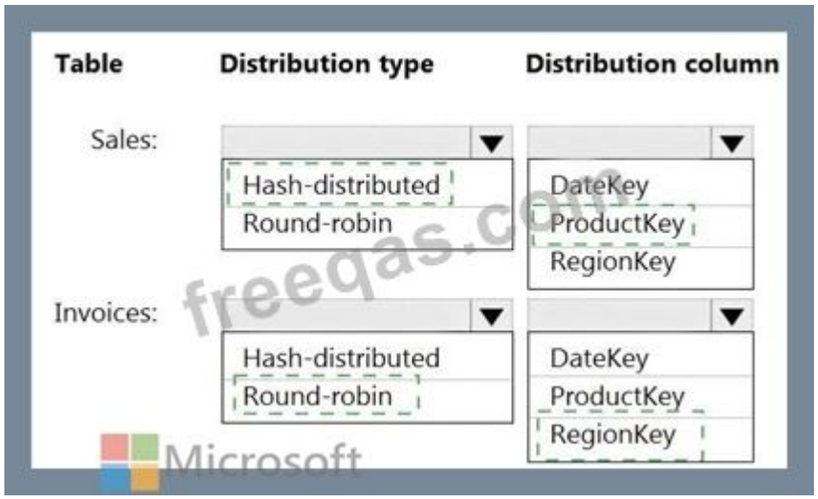
NEW QUESTION: 202
You have an on-premises data warehouse that includes the following fact tables. Both tables have the following columns: DateKey, ProductKey, RegionKey. There are 120 unique product keys and 65 unique region keys.

Table	Comments
Sales	The table is 600 GB in size. DateKey is used extensively in the WHERE clause in queries. ProductKey is used extensively in join operations. RegionKey is used for grouping. Severity-five percent of records relate to one of 40 regions.
Invoice	The table is 6 GB in size. DateKey and ProductKey are used extensively in the WHERE clause in queries. RegionKey is used for grouping.

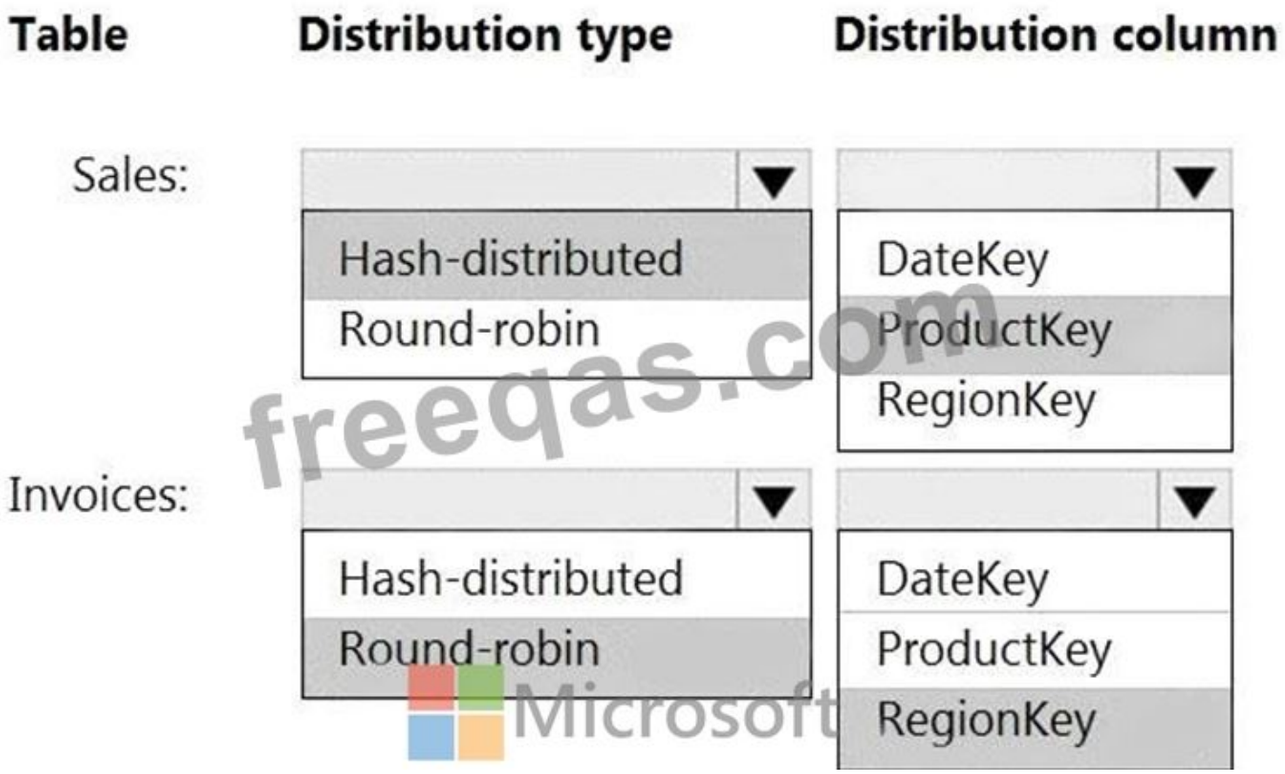
Queries that use the data warehouse take a long time to complete.
You plan to migrate the solution to use Azure Synapse Analytics. You need to ensure that the Azure-based solution optimizes query performance and minimizes processing skew.
What should you recommend? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point



Answer:



Explanation



Box 1: Hash-distributed

Box 2: ProductKey

ProductKey is used extensively in joins.

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

Box 3: Round-robin

Box 4: RegionKey

Round-robin tables are useful for improving loading speed.

Consider using the round-robin distribution for your table in the following scenarios:

* When getting started as a simple starting point since it is the default
* If there is no obvious joining key
* If there is not good candidate column for hash distributing the table
* If the table does not share a common join key with other tables
* If the join is less significant than other joins in the query
* When the table is a temporary staging table

Note: A distributed table appears as a single table, but the rows are actually stored across 60 distributions. The rows are distributed with a hash or round-robin algorithm.

Reference:

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

NEW QUESTION: 203

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 Dow this meet the goal?

A. Yes

B. No

Answer: (SHOW ANSWER)

Explanation

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: 204

You have a table named SalesFact in an enterprise data warehouse in Azure Synapse Analytics.

SalesFact contains sales data from the past 36 months and has the following characteristics:

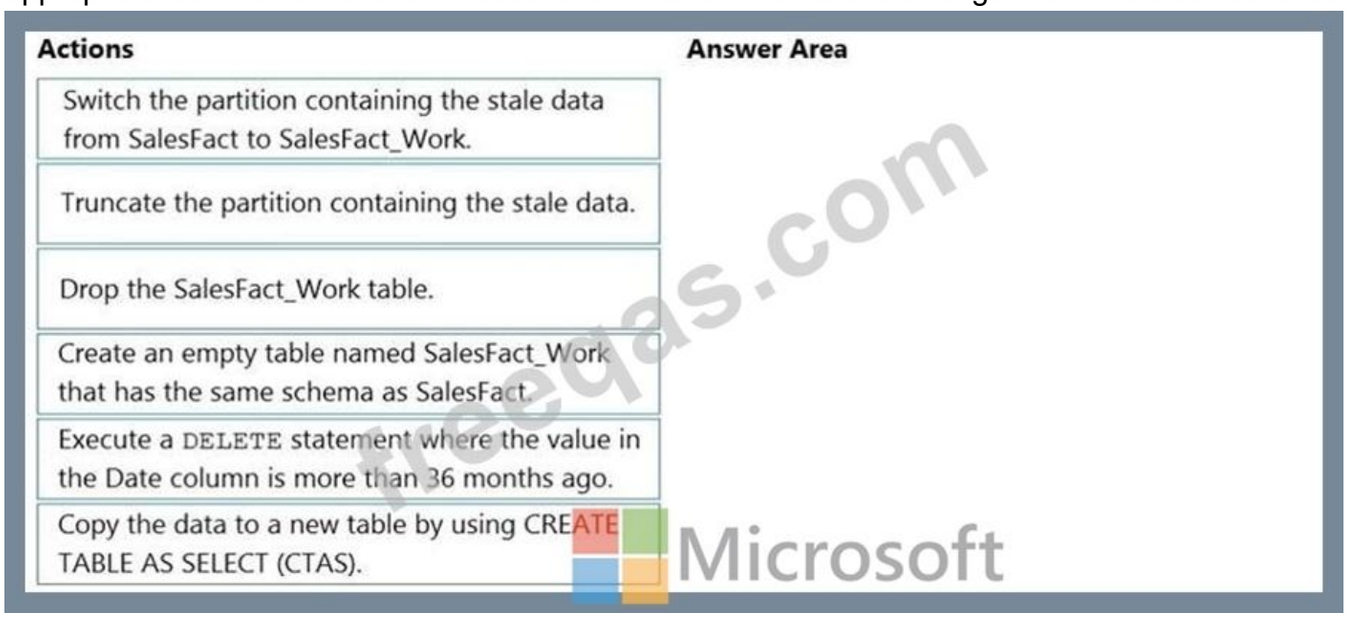
Is partitioned by month

Contains one billion rows

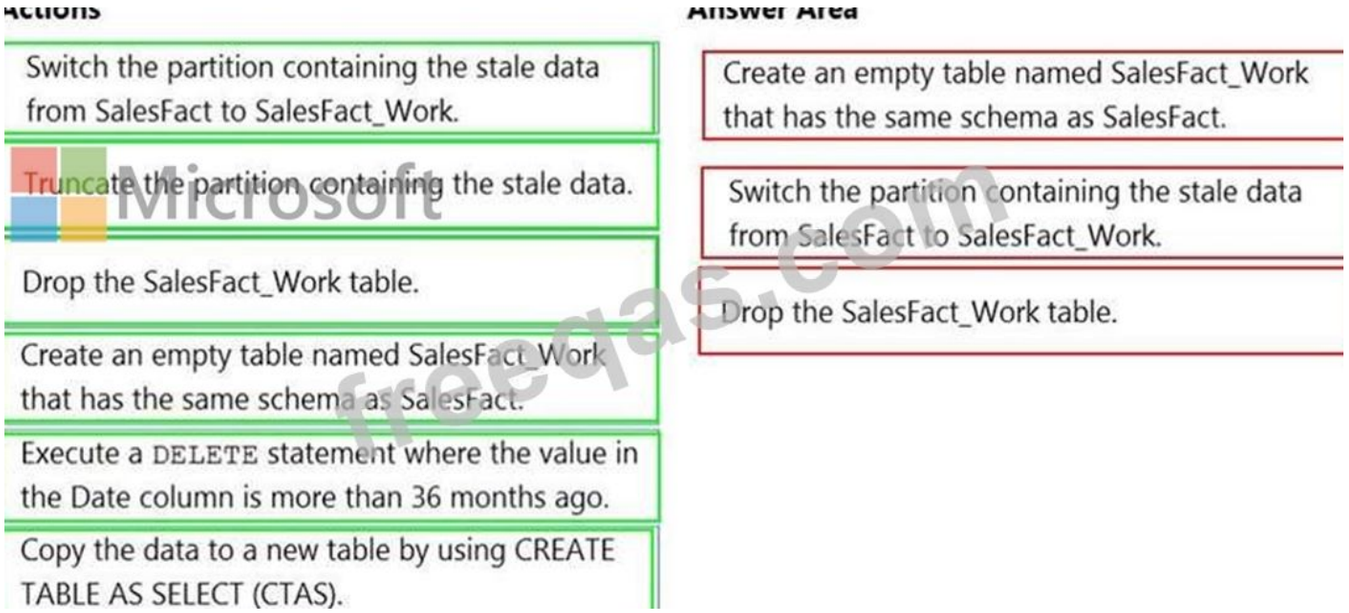
Has clustered columnstore indexes

At the beginning of each month, you need to remove data from SalesFact that is older than 36 months as quickly as possible.

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



Answer:



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

NEW QUESTION: 205

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 dat

a. 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.

Answer:

References:
https://docs.microsoft.com/en-us/azure/storage/blobs/storage-blob-storage-tiers

NEW QUESTION: 206
You are building an Azure Stream Analytics job to identify how much time a user spends interacting with a feature on a webpage.
The job receives events based on user actions on the webpage. Each row of data represents an event. Each event has a type of either 'start' or 'end'.
You need to calculate the duration between start and end events.
How should you complete the query? To answer, select the appropriate options in the answer are a.
NOTE: Each correct selection is worth one point.

```
SELECT
    [user],
    feature,
    [DATEADD( | DATEDIFF( | DATEPART(]
        second,
        [ISFIRST | LAST | TOPONE] (Time) OVER (PARTITION BY [user], feature LIMIT DURATION(hour, 1) WHEN Event = 'start'),
        Time) as duration
FROM input TIMESTAMP BY Time
WHERE
    Event = 'end'
```

Answer:

```
SELECT
    [user],
    feature,
    [DATEADD( | DATEDIFF( | DATEPART(]
        second,
        [ISFIRST | LAST | TOPONE] (Time) OVER (PARTITION BY [user], feature LIMIT DURATION(hour, 1) WHEN Event = 'start'),
        Time) as duration
FROM input TIMESTAMP BY Time
WHERE
    Event = 'end'
```

Reference:
https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-stream-analytics-query-patterns

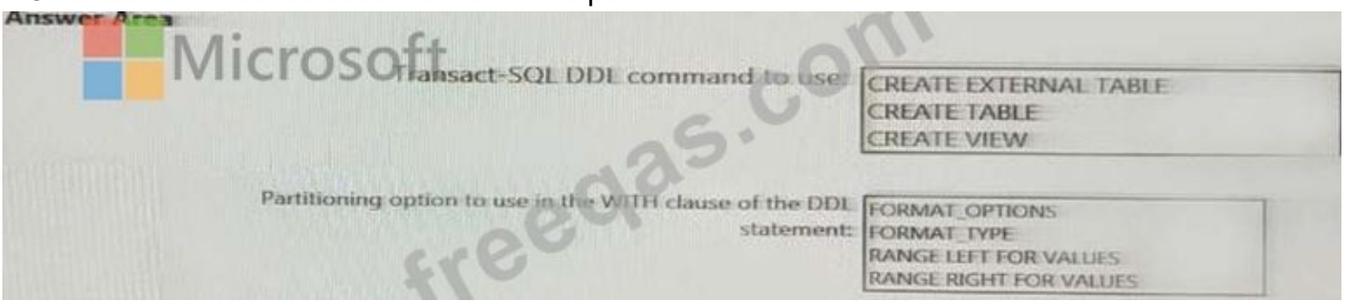
NEW QUESTION: 207
You use Azure Data Lake Storage Gen2.
You need to ensure that workloads can use filter predicates and column projections to filter data at the time the data is read from disk.
Which two actions should you perform? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.
A. Register the query acceleration feature.
B. Reregister the Microsoft Data Lake Store resource provider.
C. Create a storage policy that is scoped to a container prefix filter.
D. Create a storage policy that is scoped to a container.
E. Reregister the Azure Storage resource provider.
Answer: A,E (LEAVE A REPLY)

NEW QUESTION: 208
You are designing an enterprise data warehouse in Azure Synapse Analytics that will contain a table named Customers. Customers will contain credit card information.
You need to recommend a solution to provide salespeople with the ability to view all the entries in Customers.
The solution must prevent all the salespeople from viewing or inferring the credit card information.
What should you include in the recommendation?
A. data masking
B. Always Encrypted
C. column-level security
D. row-level security
Answer: A (LEAVE A REPLY)
Explanation
SQL Database dynamic data masking limits sensitive data exposure by masking it to non-privileged users.
The Credit card masking method exposes the last four digits of the designated fields and adds a constant string as a prefix in the form of a credit card.
Example: XXXX-XXXX-XXXX-1234
Reference:
https://docs.microsoft.com/en-us/azure/sql-database/sql-database-dynamic-data-masking-get-started

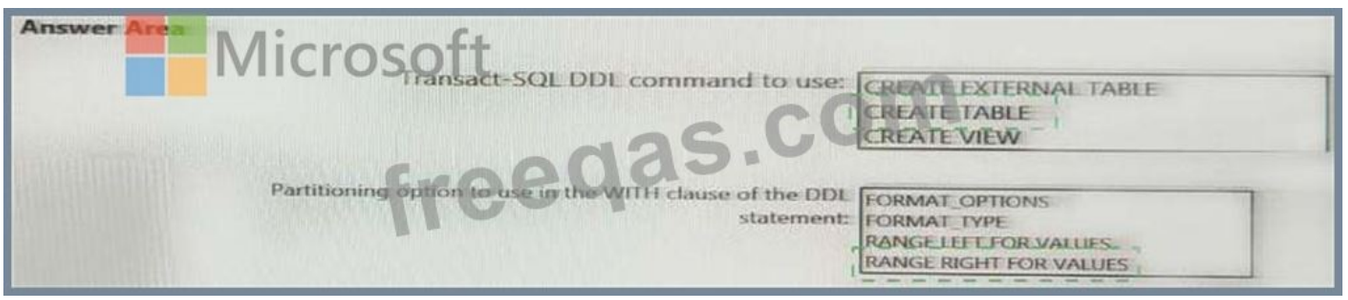
NEW QUESTION: 209
You need to schedule an Azure Data Factory pipeline to execute when a new file arrives in an Azure Data Lake Storage Gen2 container.
Which type of trigger should you use?
A. on-demand

B. tumbling window
C. schedule
D. 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: 210
You need to implement an Azure Synapse Analytics database object for storing the sales transactions data.
The solution must meet the sales transaction dataset requirements.
What solution must meet the sales transaction dataset requirements.
What should you do? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.



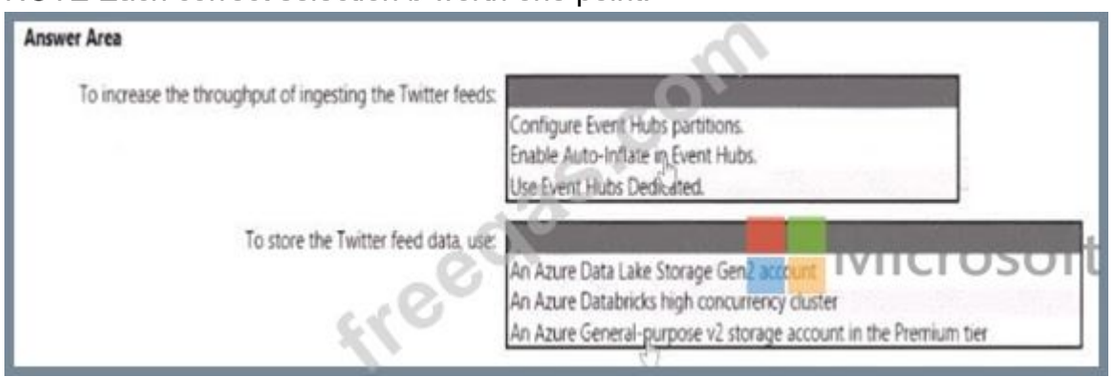
Answer:



NEW QUESTION: 211
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.

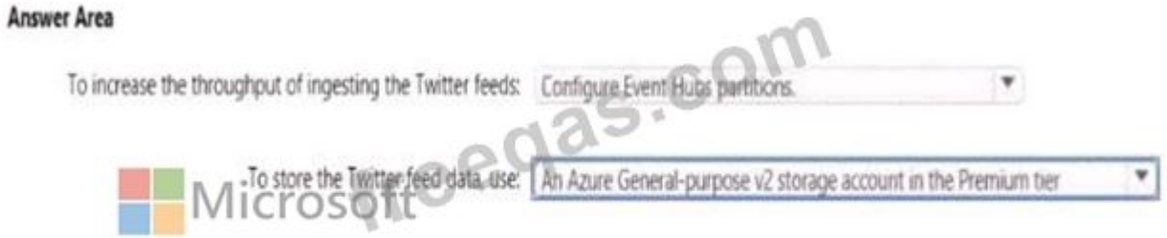
You plan to copy the data from the storage account to an Azure SQL data warehouse.
You need to prepare the files to ensure that the data copies quickly.
Solution: You modify the files to ensure that each row is more than 1 MB.
Does this meet the goal?
A. Yes
B. No
Answer: (SHOW ANSWER)
Instead modify the files to ensure that each row is less than 1 MB.
References:
https://docs.microsoft.com/en-us/azure/sql-data-warehouse/guidance-for-loading-data

NEW QUESTION: 212
You need to design a data ingestion and storage solution for the Twitter feeds. The solution must meet the customer sentiment analytics requirements.
What should you include in the solution To answer, select the appropriate options in the answer area
NOTE Each correct selection b worth one point.



Answer:

Explanation



NEW QUESTION: 213

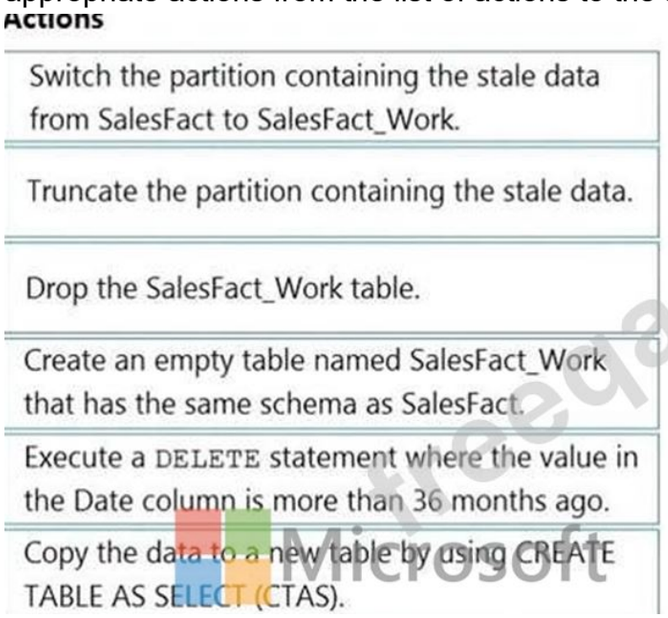
You have a table named SalesFact in an enterprise data warehouse in Azure Synapse Analytics.

SalesFact contains sales data from the past 36 months and has the following characteristics:

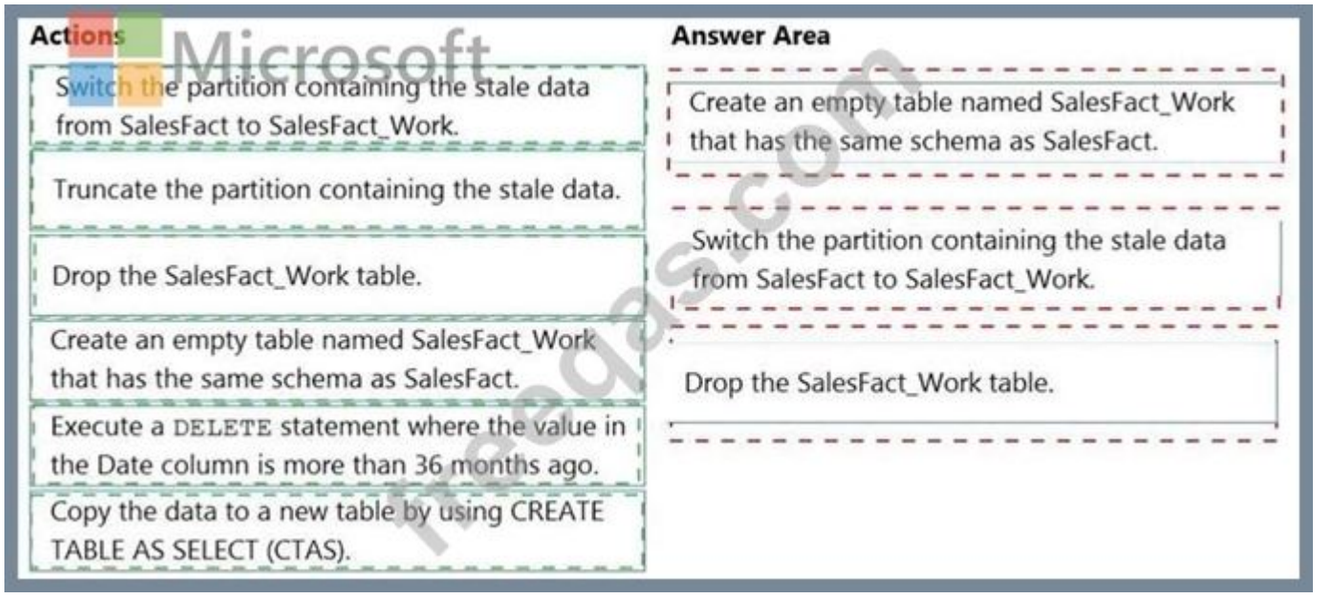
* Is partitioned by month
* Contains one billion rows
* Has clustered columnstore indexes

At the beginning of each month, you need to remove data from SalesFact that is older than 36 months as quickly as possible.

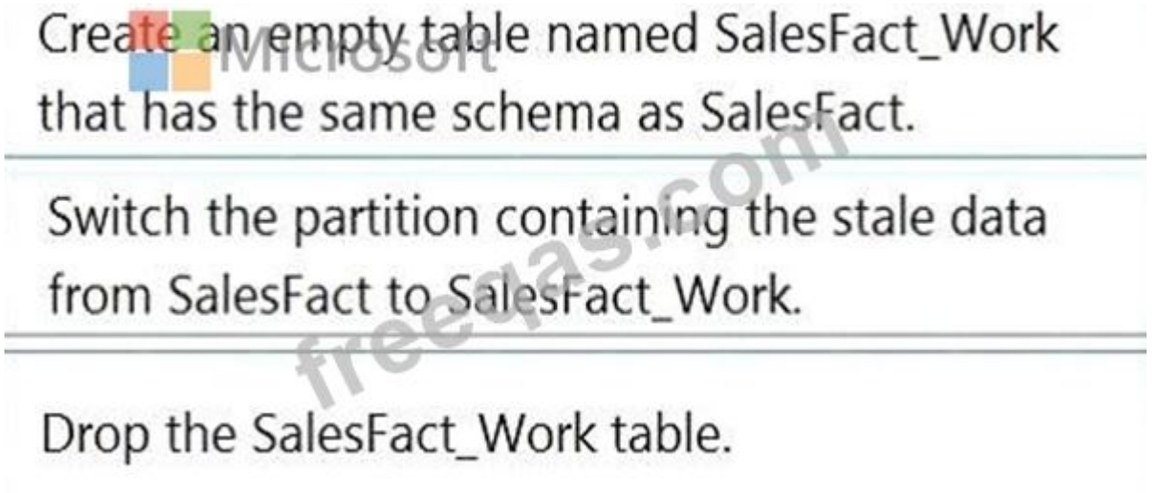
Which three actions should you perform in sequence in a stored procedure? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



Answer:



Explanation



Step 1: Create an empty table named SalesFact_work that has the same schema as SalesFact.
Step 2: Switch the partition containing the stale data from SalesFact to SalesFact_Work.
SQL Data Warehouse supports partition splitting, merging, and switching. To switch partitions between two tables, you must ensure that the partitions align on their respective boundaries and that the table definitions match.
Loading data into partitions with partition switching is a convenient way stage new data in a table that is not visible to users the switch in the new data.
Step 3: Drop the SalesFact_Work table.
Reference:
https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-tables-partition

NEW QUESTION: 214
You are performing exploratory analysis of the bus fare data in an Azure Data Lake Storage Gen2 account by using an Azure Synapse Analytics serverless SQL pool.
You execute the Transact-SQL query shown in the following exhibit.

```
SELECT
   payment_type,
   SUM(fare_amount) AS fare_total
FROM OPENROWSET(
         BULK 'csv/busfare/tripdata_2020*.csv',
         DATA_SOURCE = 'BusData',
         FORMAT = 'CSV', PARSER_VERSION = '2.0',
         FIRSTROW = 2
   )
   WITH (
         payment_type INT 10,
         fare_amount FLOAT 11
   ) AS nyc
GROUP BY payment_type
ORDER BY payment_type;
```

What do the query results include?

A. All CSV files that have file names that contain "tripdata_2020".

B. Only CSV that have file names that beginning with "tripdata_2020".

C. Only CSV files in the tripdata_2020 subfolder.

D. All files that have file names that beginning with "tripdata_2020".

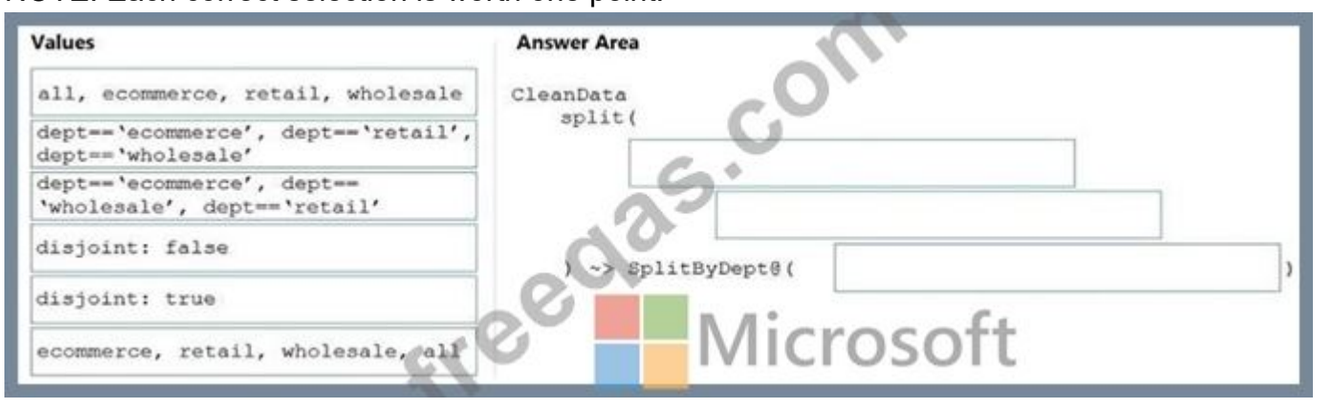
Answer: B (LEAVE A REPLY)

NEW QUESTION: 215

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.



Answer:

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

NEW QUESTION: 216
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 hopping window that uses a hop size of 10 seconds and a window size of 10 seconds.
Does this meet the goal?
A. Yes
B. No
Answer: B (LEAVE A REPLY)
Explanation
Instead use a tumbling window. Tumbling windows are a series of fixed-sized, non-overlapping and contiguous time intervals.
Reference:
https://docs.microsoft.com/en-us/stream-analytics-query/tumbling-window-azure-stream-analytics

NEW QUESTION: 217
You are building an Azure Analytics query that will receive input data from Azure IoT Hub and write the results to Azure Blob storage.
You need to calculate the difference in readings per sensor per hour.
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:

Explanation

Box 1: LAG

The LAG analytic operator allows one to look up a "previous" event in an event stream, within certain constraints. It is very useful for computing the rate of growth of a variable, detecting when a variable crosses a threshold, or when a condition starts or stops being true.

Box 2: LIMIT DURATION

Example: Compute the rate of growth, per sensor:

SELECT sensorId,

growth = reading -

LAG(reading) OVER (PARTITION BY sensorId LIMIT DURATION(hour, 1))

FROM input

Reference:

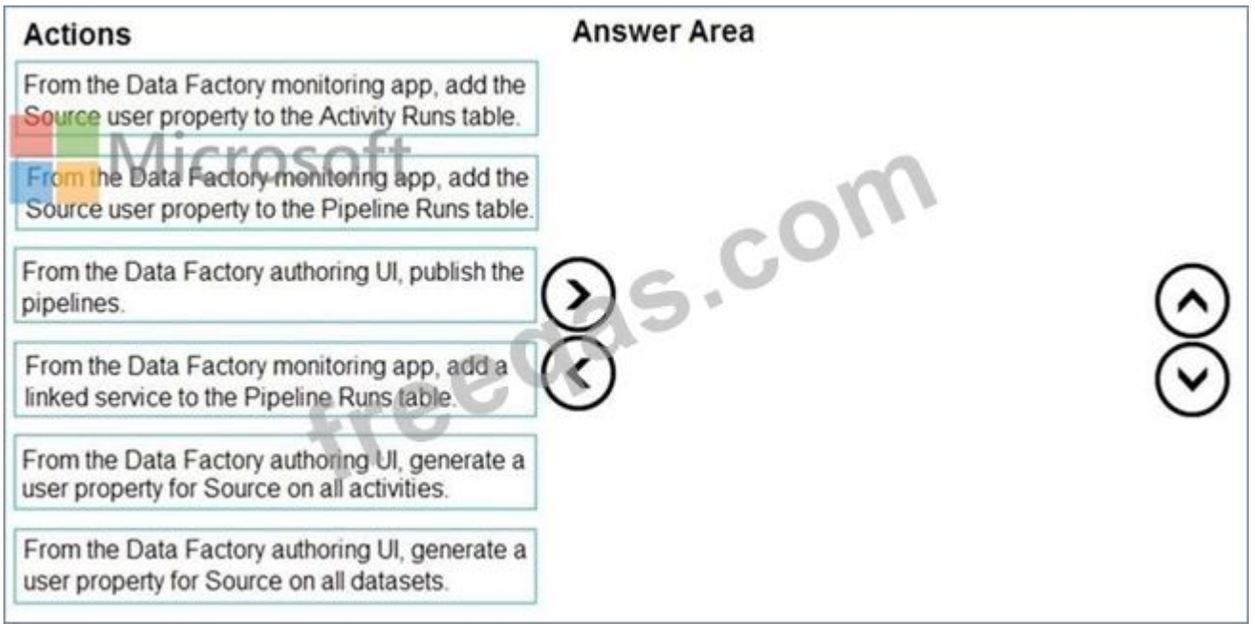
https://docs.microsoft.com/en-us/stream-analytics-query/lag-azure-stream-analytics

NEW QUESTION: 218

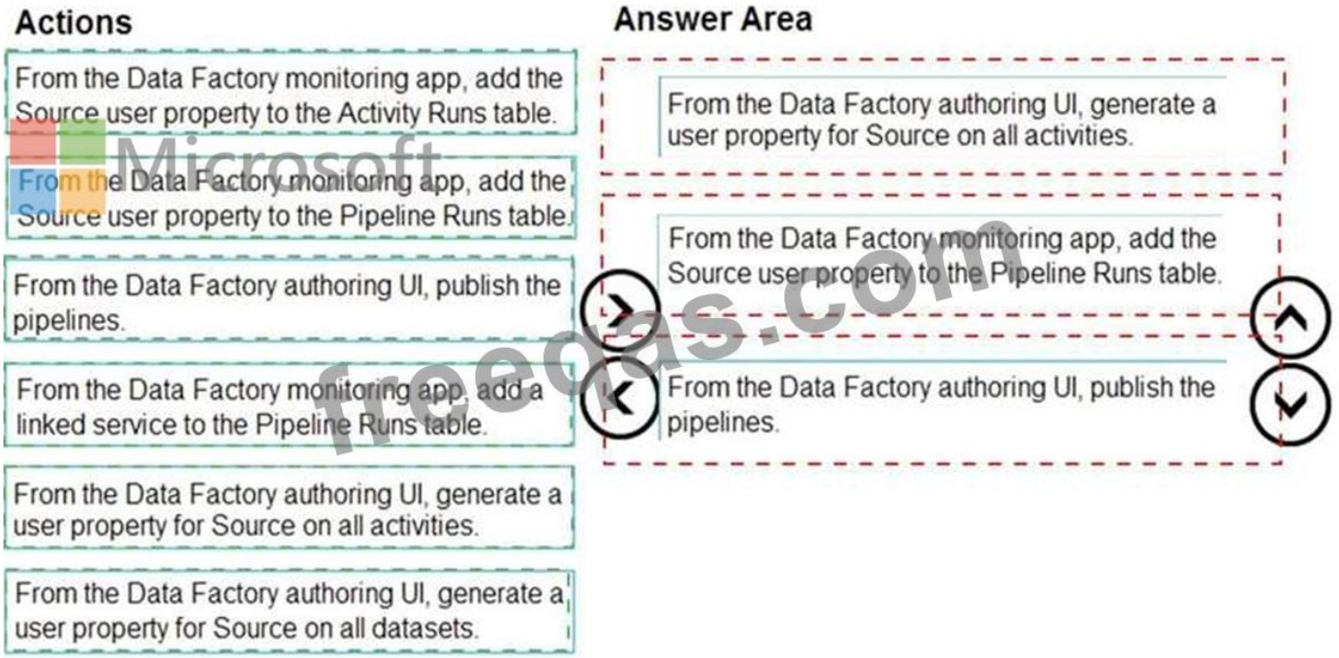
You plan to monitor an Azure data factory by using the Monitor & Manage app.

You need to identify the status and duration of activities that reference a table in a source database.

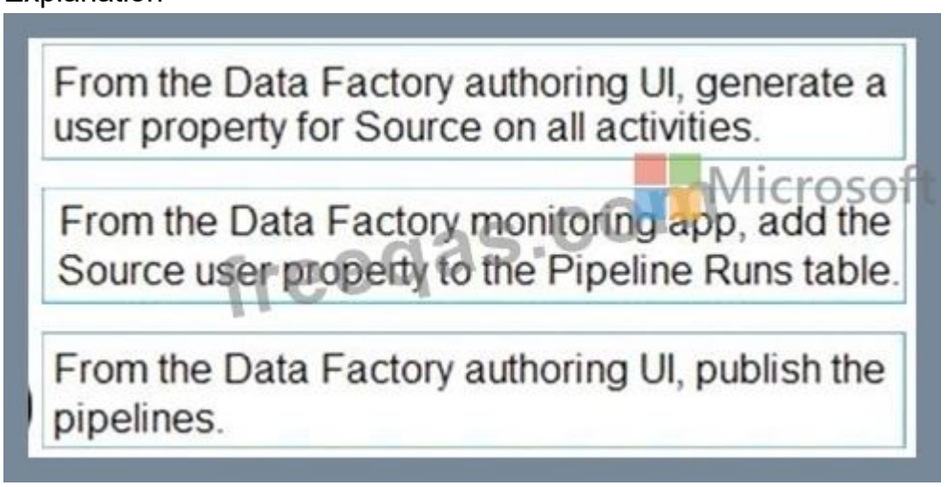
Which three actions should you perform in sequence? To answer, move the actions from the list of actions to the answer are and arrange them in the correct order.



Answer:



Explanation



Step 1: From the Data Factory authoring UI, generate a user property for Source on all activities.

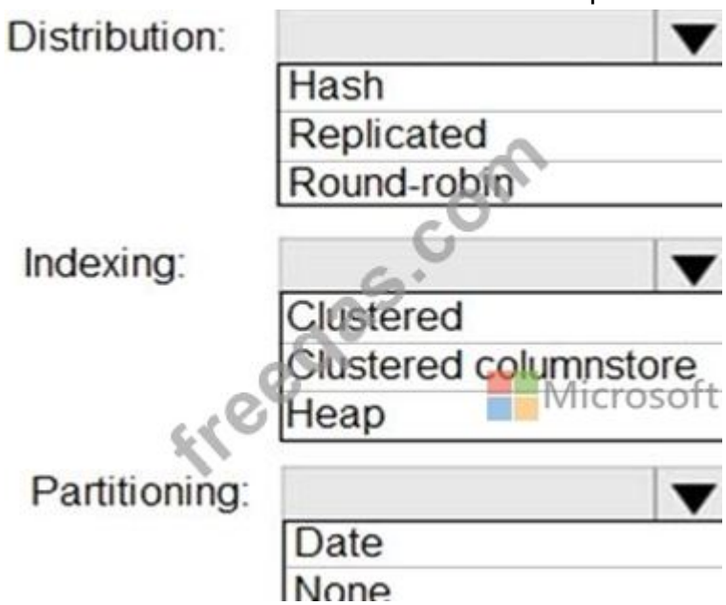
Step 2: From the Data Factory monitoring app, add the Source user property to Activity Runs table.
You can promote any pipeline activity property as a user property so that it becomes an entity that you can monitor. For example, you can promote the Source and Destination properties of the copy activity in your pipeline as user properties. You can also select Auto Generate to generate the Source and Destination user properties for a copy activity.
Step 3: From the Data Factory authoring UI, publish the pipelines
Publish output data to data stores such as Azure SQL Data Warehouse for business intelligence (BI) applications to consume.
References:
https://docs.microsoft.com/en-us/azure/data-factory/monitor-visually

NEW QUESTION: 219
You are building an Azure Stream Analytics job to identify how much time a user spends interacting with a feature on a webpage.
The job receives events based on user actions on the webpage. Each row of data represents an event. Each event has a type of either 'start' or 'end'.
You need to calculate the duration between start and end events.
How should you complete the query? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

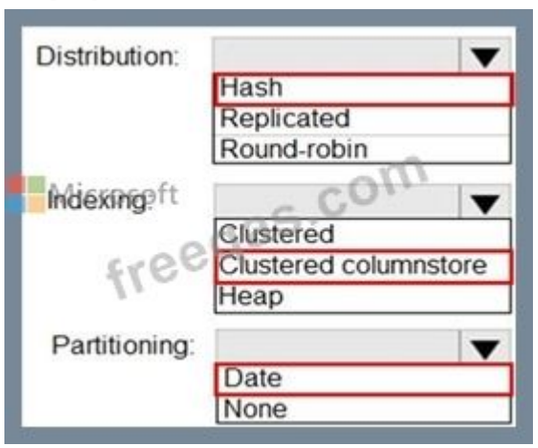
Answer:

Reference:
https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-stream-analytics-query-patterns

NEW QUESTION: 220
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.



Answer:



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: 221
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.

Reference:

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-edge

NEW QUESTION: 222

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

NEW QUESTION: 223
You have an Azure Stream Analytics job.
You need to ensure that the job has enough streaming units provisioned.
You configure monitoring of the SU % Utilization metric.
Which two additional metrics should you monitor? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.
A. Backlogged Input Events
B. Watermark Delay
C. Function Events
D. Out of order Events
E. Late Input Events
Answer: A,B (LEAVE A REPLY)
Explanation
To react to increased workloads and increase streaming units, consider setting an alert of 80% on the SU Utilization metric. Also, you can use watermark delay and backlogged events metrics to see if there is an impact.
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, by increasing the SUs.
Reference:
https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-monitoring

NEW QUESTION: 224
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)
Explanation
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: 225

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

opendatasource

openjson

openquery

openrowset

Answer Area

```
select*

FROM

[                ] (

     BULK 'https://contoso.blob.core.windows.net/contosodw',
     FORMAT= 'CSV',
     fieldterminator = '0x0b',
     fieldquote = '0x0b',
     rowterminator = '0x0b'
)
with (id varchar(50),
     contextdateventTime 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

opendatasource

openjson

openquery

openrowset

Answer Area

```
select*

FROM
  [openrowset] (

     BULK 'https://contoso.blob.core.windows.net/contosodw',
     FORMAT= 'CSV',
     fieldterminator = '0x0b',
     fieldquote = '0x0b',
     rowterminator = '0x0b'
)
with (id varchar(50),
     contextdateventTime 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'
     )
```

Reference:
https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/query-single-csv-file
https://docs.microsoft.com/en-us/sql/relational-databases/json/import-json-documents-into-sql-server

NEW QUESTION: 226
You need to integrate the on-premises data sources and Azure Synapse Analytics. The solution must meet the data integration requirements.
Which type of integration runtime should you use?
A. Azure-SSIS integration runtime
B. self-hosted integration runtime
C. Azure integration runtime
Answer: (SHOW ANSWER)
Topic 1, Contoso
Transactional Date
Contoso has three years of customer, transactional, operation, sourcing, and supplier data comprised of 10 billion records stored across multiple on-premises Microsoft SQL Server servers. The SQL server instances contain data from various operational systems. The data is loaded into the instances by using SQL server integration Services (SSIS) packages.
You estimate that combining all product sales transactions into a company-wide sales transactions dataset will result in a single table that contains 5 billion rows, with one row per transaction.
Most queries targeting the sales transactions data will be used to identify which products were sold in retail stores and which products were sold online during different time period. Sales transaction data that is older than three years will be removed monthly.
You plan to create a retail store table that will contain the address of each retail store. The table will be approximately 2 MB. Queries for retail store sales will include the retail store addresses.
You plan to create a promotional table that will contain a promotion ID. The promotion ID will be associated to a specific product. The product will be identified by a product ID. The table will be approximately 5 GB.
Streaming Twitter Data
The ecommerce department at Contoso develops and Azure logic app that captures trending Twitter feeds referencing the company's products and pushes the products to Azure Event Hubs.
Planned Changes
Contoso plans to implement the following changes:
* Load the sales transaction dataset to Azure Synapse Analytics.
* Integrate on-premises data stores with Azure Synapse Analytics by using SSIS packages.
* Use Azure Synapse Analytics to analyze Twitter feeds to assess customer sentiments about products.
Sales Transaction Dataset Requirements
Contoso identifies the following requirements for the sales transaction dataset:
* Partition data that contains sales transaction records. Partitions must be designed to provide efficient loads by month. Boundary values must belong: to the partition on the right.

* Ensure that queries joining and filtering sales transaction records based on product ID complete as quickly as possible.
* Implement a surrogate key to account for changes to the retail store addresses.
* Ensure that data storage costs and performance are predictable.
* Minimize how long it takes to remove old records.

Customer Sentiment Analytics Requirement

Contoso identifies the following requirements for customer sentiment analytics:
* Allow Contoso users to use PolyBase in an A/ure Synapse Analytics dedicated SQL pool to query the content of the data records that host the Twitter feeds. Data must be protected by using row-level security (RLS). The users must be authenticated by using their own A/ureAD credentials.
* Maximize the throughput of ingesting Twitter feeds from Event Hubs to Azure Storage without purchasing additional throughput or capacity units.
* Store Twitter feeds in Azure Storage by using Event Hubs Capture. The feeds will be converted into Parquet files.
* Ensure that the data store supports Azure AD-based access control down to the object level.
* Minimize administrative effort to maintain the Twitter feed data records.
* Purge Twitter feed data records;itftaitJ are older than two years.

Data Integration Requirements

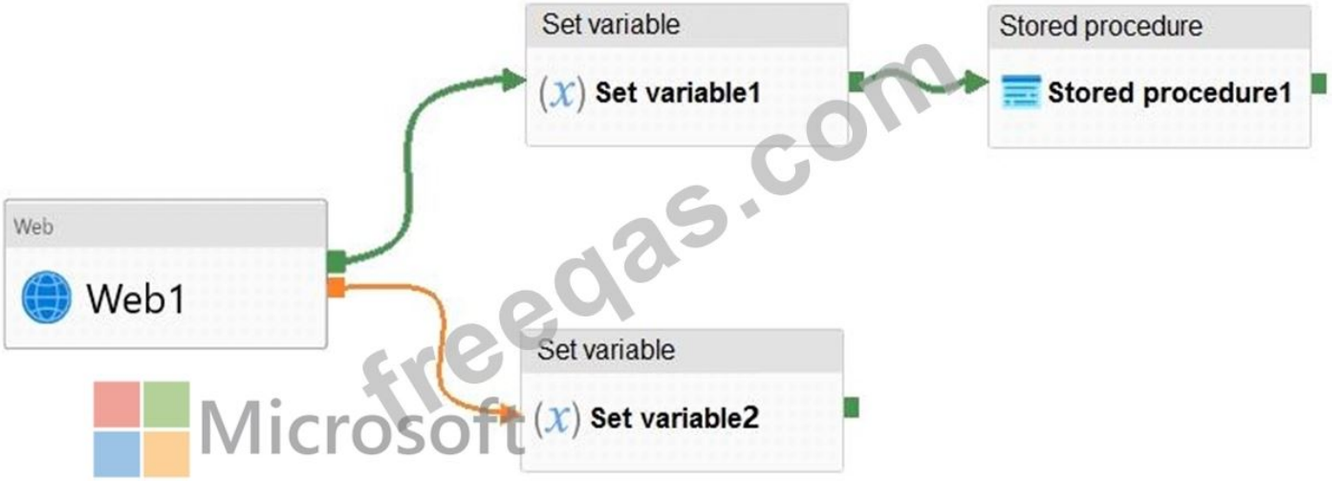
Contoso identifies the following requirements for data integration:

Use an Azure service that leverages the existing SSIS packages to ingest on-premises data into datasets stored in a dedicated SQL pool of Azure Synaps Analytics and transform the data.

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

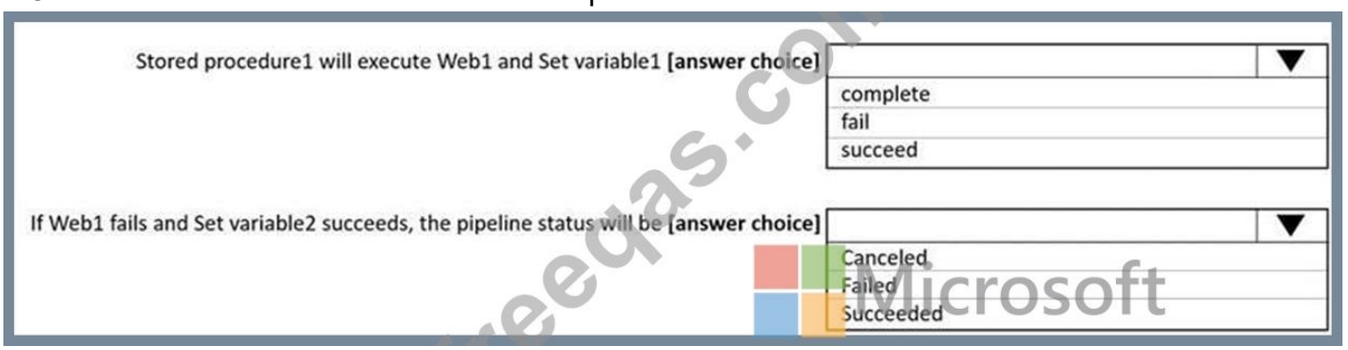
NEW QUESTION: 227

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.



Answer:

Reference:
https://datasavvy.me/2021/02/18/azure-data-factory-activity-failures-and-pipeline-outcomes/