DUMPSARENA

Implementing an Azure Data Solution (beta)

Microsoft DP-200

Version Demo

Total Demo Questions: 15

Total Premium Questions: 372

Buy Premium PDF

https://dumpsarena.com

sales@dumpsarena.com

Topic Break Down

Topic	No. of Questions
Topic 1, Case Study 1	2
Topic 2, Case Study 2	5
Topic 3, Case Study 3	2
Topic 4, Case Study 4	2
Topic 5, Case Study 5	3
Topic 6, Case Study 6	2
Topic 7, Case Study 7	5
Topic 8, Case Study 8	2
Topic 9, Case Study 9	3
Topic 10, Case Study 10	3
Topic 11, Mixed Questions	343
Total	372

QUESTION NO: 1

A company has an Azure SQL data warehouse. They want to use PolyBase to retrieve data from an Azure Blob storage account and ingest into the Azure SQL data warehouse. The files are stored in parquet format. The data needs to be loaded into a table called XYZ_sales.

Which of the following actions need to be performed to implement this requirement? (Choose four.)

A. Create an external file format that would map to the parquet-based files

B. Load the data into a staging table

C. Create an external table called XYZ_sales_details

D. Create an external data source for the Azure Blob storage account

E. Create a master key on the database

F. Configure Polybase to use the Azure Blob storage account

ANSWER: B C D E

Explanation:

There is an article on github as part of the Microsoft documentation that provides details on how to load data into an Azure SQL data warehouse from an Azure Blob storage account. The key steps are: Creating a master key in the database.

Creating an external data source for the Azure Blob storage account:

3. Create a master key for the MySampleDataWarehouse database. You only need to create a master key once per database.

CREATE MASTER KEY;

4. Run the following CREATE EXTERNAL DATA SOURCE statement to define the location of the Azure blob. This is the location of the external taxi cab data. To run a command that you have appended tothe query window, highlight the commands you wish to run and click Execute.

```
CREATE EXTERNAL DATA SOURCE NYTPublic
WITH
(
TYPE = Hadoop,
LOCATION = 'wasbs://2013@nytaxiblob.blob.core.windows.net/'
);
```

Next you load the data. But it is always beneficial to load the data into a staging table first:

Load the data into your data warehouse.

This section uses the external tables you just defined to load the sample data from Azure Storage Blob to SQL Data Warehouse.

[!NOTE] This tutorial loads the data directly into the final table. In a production environment, you will usually use CREATE TABLE AS SELECT to load into a staging table. While data is in the staging table you can perform any necessary transformations. To append the data in the staging table to a production table, you can use the INSERT...SELECT statement. For more information, see Inserting data into a production table.

Since this is clearly provided in the documentation, all other options are incorrect.

QUESTION NO: 2

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.

A company uses Azure Data Lake Gen 1 Storage to store big data related to consumer behavior.

You need to implement logging.

Solution: Create an Azure Automation runbook to copy events.

Does the solution meet the goal?

A. Yes

B. No

ANSWER: B

Explanation:

Instead configure Azure Data Lake Storage diagnostics to store logs and metrics in a storage account.

Note:

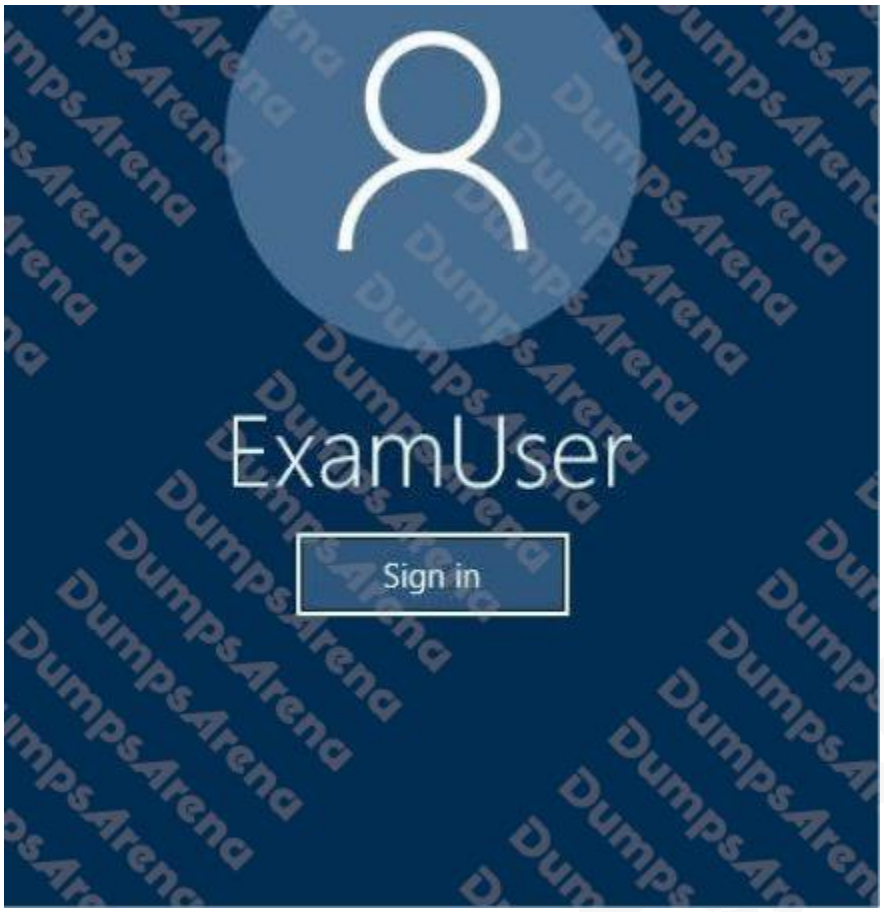
You can enable diagnostic logging for your Azure Data Lake Storage Gen1 accounts, blobs, files, queues and tables.

Diagnostic logs aren't available for Data Lake Storage Gen2 accounts [as of August 2019]. Reference:

https://docs.microsoft.com/en-us/azure/data-lake-store/data-lake-store-diagnostic-logs
https://github.com/MicrosoftDocs/azure-docs/issues/34286

QUESTION NO: 3 - (SIMULATION)

SIMULATION



Use the following login credentials as needed:

Azure Username: xxxxx Azure Password: xxxxx

The following information is for technical support purposes only: Lab Instance: 10277521

You plan to create large data sets on db2.

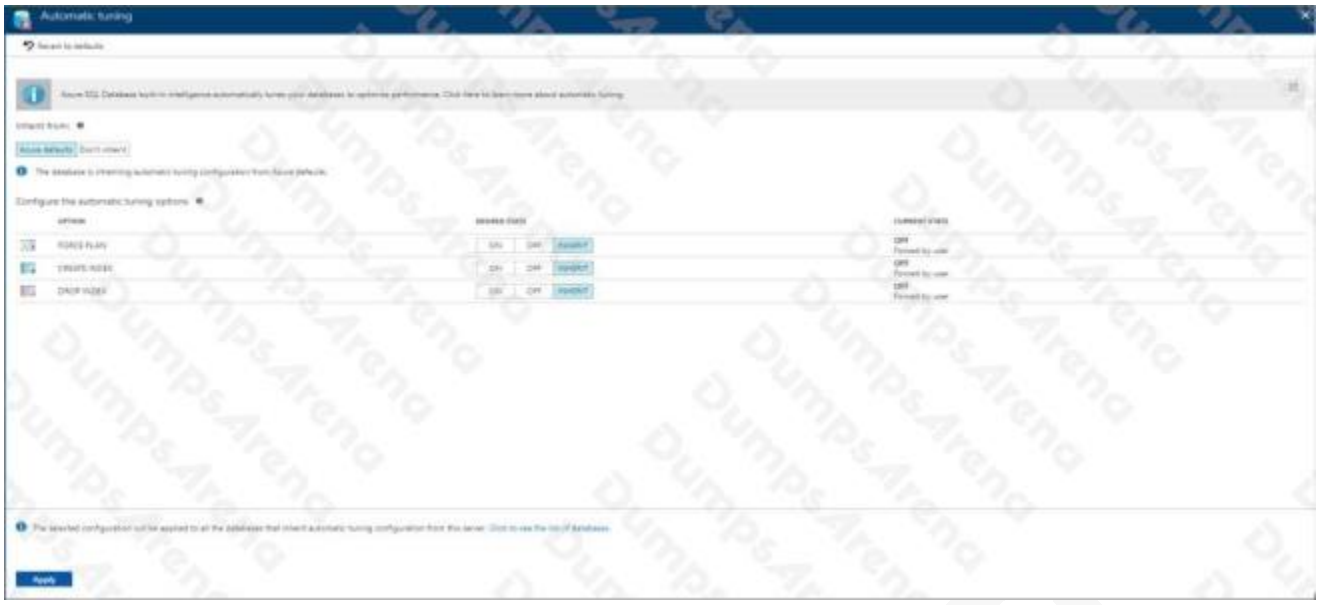
You need to ensure that missing indexes are created automatically by Azure in db2. The solution must apply ONLY to db2.

To complete this task, sign in to the Azure portal.

ANSWER: See the explanation below.

Explanation:

1. To enable automatic tuning on Azure SQL Database logical server, navigate to the server in Azure portal and then select Automatic tuning in the menu.



2. Select database db2

3. Click the Apply buttonReference:

https://docs.microsoft.com/en-us/azure/sql-database/sql-database-automatic-tuning-enable

QUESTION NO: 4

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

XYZ is an online training provider. They also provide a yearly gaming competition for their students. The competition is held every month in different locations. Current Environment

The company currently has the following environment in place:

• The racing cars for the competition send their telemetry data to a MongoDB database. The telemetry data has around 100 attributes.

• A custom application is then used to transfer the data from the MongoDB database to a SQL Server 2017 database. The attribute names are changed when they are sent to the SQL Server database.

• Another application named "XYZ workflow" is then used to perform analytics on the telemetry data to look for improvements on the racing cars.

• The SQL Server 2017 database has a table named "cardata" which has around 1 TB of data. "XYZ workflow" performs the required analytics on the data in this table. Large aggregations are performed ona column of the table.

Proposed Environment

The company now wants to move the environment to Azure. Below are the key requirements:

• The racing car data will now be moved to Azure Cosmos DB and Azure SQL database. The data must be written to the closest Azure data center and must converge in the least amount of time.

• The query performance for data in the Azure SQL database must be stable without the need of administrative overhead

• The data for analytics will be moved to an Azure SQL Data warehouse

• Transparent data encryption must be enabled for all data stores wherever possible

• An Azure Data Factory pipeline will be used to move data from the Cosmos DB database to the Azure SQL database. If there is a delay of more than 15 minutes for the data transfer, then configurationchanges need to be made to the pipeline workflow.

• The telemetry data must be monitored for any sort of performance issues.

• The Request Units for Cosmos DB must be adjusted to maintain the demand while also minimizing costs.

• The data in the Azure SQL Server database must be protected via the following requirements:

- Only the last four digits of the values in the column CarID must be shown - A zero value must be shown for all values in the column CarWeight

Which of the following would you use for the consistency level for the database?

A. Eventual

B. Session

C. Strong

D. Consistent prefix

ANSWER: A

Explanation:

Since there is a requirement for data to be written to the closest data center for Cosmos DB, we need to ensure there is a multi-master setup for Cosmos DB wherein data can be written from multiple regions. For such accounts, we can’t set the consistency level to Strong. The Microsoft documentation mentions the following:

Strong consistency and multi-master

Cosmos accounts configured for multi-master cannot be configured for strong consistency as it is not possible for a distributed system to provide an RPO of zero and an RTO of zero. Additionally, there are no write latency benefits for using strong consistency with multi-master as any write into any region must be replicated and committed to all configured regions within the account. This results in the same write latency as a single master account.

Hence if we want data to converge in the least amount of time, we need to use Eventual consistency. This offers the least latency in terms of consistency. The Microsoft documentation mentions the following on the consistency levels.

With Azure Cosmos DB, developers can choose from five well-defined consistency models on the consistency spectrum. From strongest to more relaxed, the models include strong, bounded staleness, session, consistent prefix, and eventual consistency. The models are well-defined and intuitive and can be used for specific real-world scenarios. Each model provides availability and performance tradeoffs and is backed by the SLAs. The following image shows the different consistency levels as a spectrum.

Because of the proposed logic to the consistency level, all other options are incorrect. Reference:

https://docs.microsoft.com/en-us/azure/cosmos-db/consistency-levels-tradeoffs https://docs.microsoft.com/en-us/azure/cosmos-db/consistency-levels

QUESTION NO: 5 - (SIMULATION)

SIMULATION

Use the following login credentials as needed:

Azure Username: xxxxx Azure Password: xxxxx

The following information is for technical support purposes only:

Lab Instance: 10543936

You need to ensure that users in the West US region can read data from a local copy of an Azure Cosmos DB database named cosmos10543936.

To complete this task, sign in to the Azure portal.

NOTE: This task might take several minutes to complete. You can perform other tasks while the task completes or end this section of the exam.

ANSWER: See the explanation below.

Explanation:

You can enable Availability Zones by using Azure portal when creating an Azure Cosmos account. You can enable Availability Zones by using Azure portal.

Step 1: enable the Geo-redundancy, Multi-region Writes

1. In Azure Portal search for and select Azure Cosmos DB.

2. Locate the Cosmos DB database named cosmos10543936

3. Access the properties for cosmos10543936

4. enable the Geo-redundancy, Multi-region Writes.

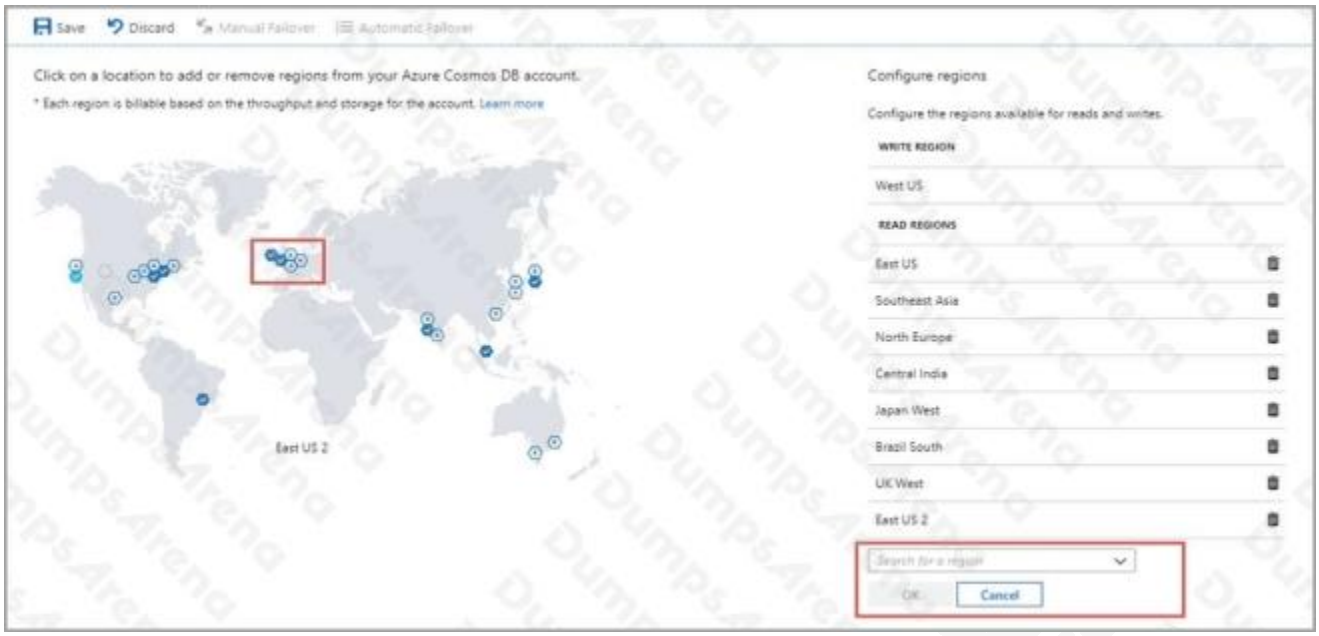
Location: West US region

Step 2: Add region from your database account

1. In to Azure portal, go to your Azure Cosmos account, and open the Replicate data globally menu.

2. To add regions, select the hexagons on the map with the + label that corresponds to your desired region(s). Alternatively, to add a region, select the + Add region option and choose a region from thedrop-down menu.

Add: West US region



3. To save your changes, select OK.

Reference: https://docs.microsoft.com/en-us/azure/cosmos-db/high-availability https://docs.microsoft.com/en-us/azure/cosmos-db/how-to-manage-database-account

QUESTION NO: 6 - (DRAG DROP)

DRAG DROP

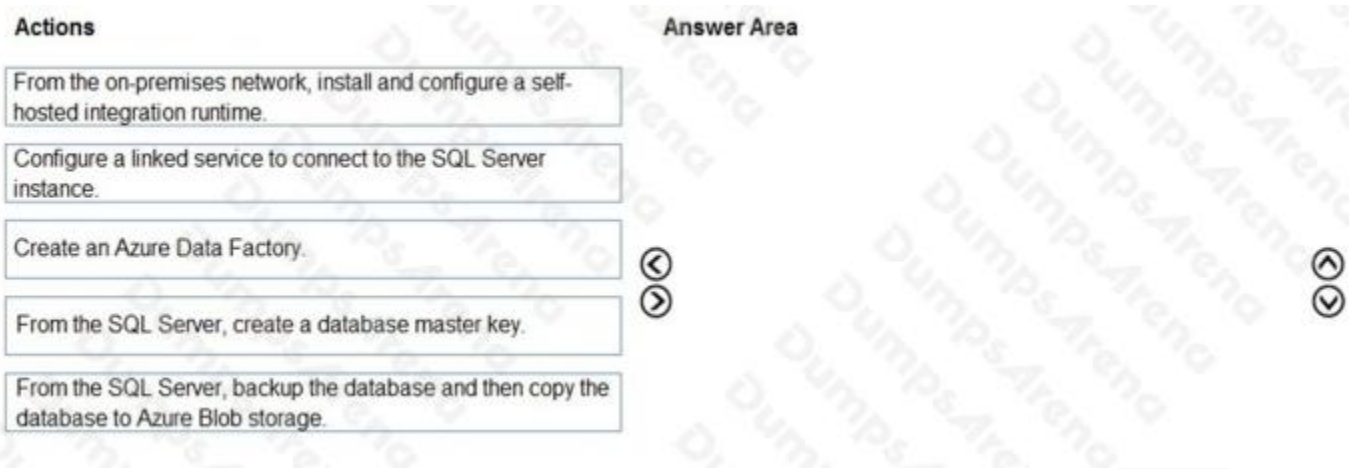
Your company has an on-premises Microsoft SQL Server instance.

The data engineering team plans to implement a process that copies data from the SQL Server instance to Azure Blob storage once a day. The process must orchestrate and manage the data lifecycle.

You need to create Azure Data Factory to connect to the SQL Server instance.

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.

Select and Place:



ANSWER:

Explanation:

Step 1: Create an Azure Data Factory

You need to create a data factory and start the Data Factory UI to create a pipeline in the data factory.

Step 2: From the on-premises network, install and configure a self-hosted runtime.

To use copy data from a SQL Server database that isn't publicly accessible, you need to set up a self-hosted integration runtime.

Step 3: Configure a linked service to connect to the SQL Server instance.

Reference: https://docs.microsoft.com/en-us/azure/data-factory/connector-sql-server
https://www.mssqltips.com/sqlservertip/5812/connect-to-onpremises-data-in-azure-data-factory-with-the-selfhosted-integration-runtime--part-1/

QUESTION NO: 7

A company has a set of Azure SQL Databases. They want to ensure that their IT Security team is informed when any security related operation occurs on the database. You need to configure Azure Monitor while ensuring administrative efforts are reduced.

Which of the following actions would you perform for this requirement? (Choose three.)

A. Create a new action group which send email alerts to the IT Security team

B. Make sure to use all security operations as the condition

C. Ensure to query audit log entries as the condition

D. Use all the Azure SQL Database servers as the resource

ANSWER: A B D

Explanation:

You can setup alerts based on all the security conditions in Azure Monitor. When any security operation is performed, an alert can be sent to the IT Security team. Option “Ensure to query audit log entries as the condition” is incorrect since you need to monitor all security related events. Reference:

https://docs.microsoft.com/en-us/azure/sql-databa se/sql-database-insights-alerts-portal

QUESTION NO: 8

You have an Azure subscription that contains an Azure Data Factory version 2 (V2) data factory 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. Disable purge protection on vault1.

B. Create a self-hosted integration runtime.

C. Disable soft delete on vault1.

D. Remove the linked service from df1.

ANSWER: D

Explanation:

Linked services are much like connection strings, which define the connection information needed for Data Factory to connect to external resources.

Incorrect Answers:

A, C: Data Factory requires two properties to be set on the Key Vault, Soft Delete and Do Not Purge B: A self-hosted integration runtime copies data between an on-premises store and cloud storage. 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

QUESTION NO: 9 - (DRAG DROP)

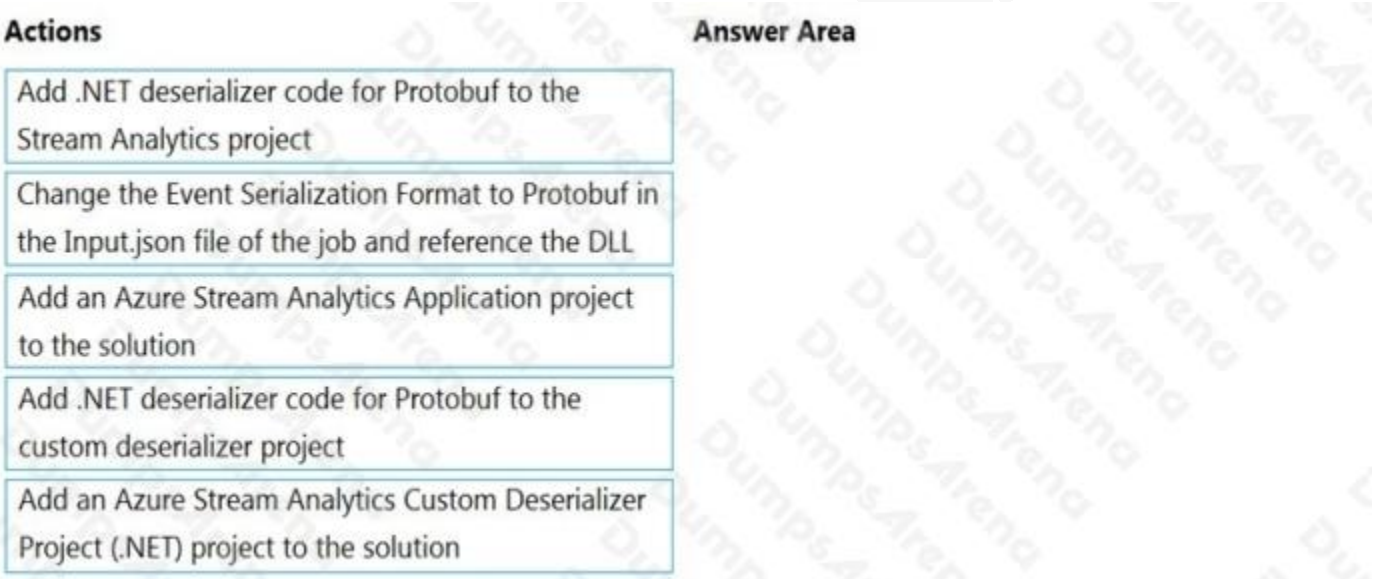
DRAG DROP

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

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

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

Select and Place:



ANSWER:

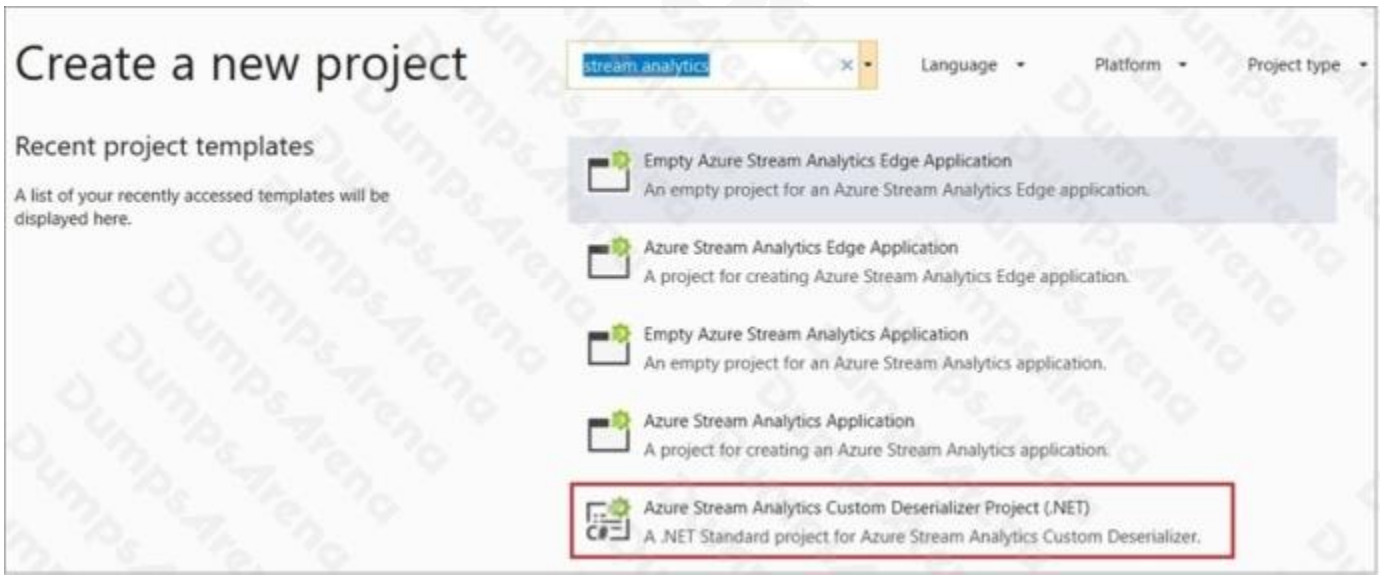
Actions	Answer Area
Add .NET deserializer code for Protobuf to the Stream Analytics project	Add an Azure Stream Analytics Custom Deserializer Project (.NET) project to the solution
Change the Event Serialization Format to Protobuf in the Input.json file of the job and reference the DLL	Add .NET deserializer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

Explanation:

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

Create a custom deserializer

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



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

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

4. Build the Protobuf Deserializer project.

Step 2: Add .NET deserializer code for Protobuf to the custom deserializer project

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

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

Add an Azure Stream Analytics project

1. In Solution Explorer, right-click the Protobuf Deserializer solution and select Add > New Project. Under Azure Stream Analytics > Stream Analytics, choose Azure Stream AnalyticsApplication. Name it ProtobufCloudDeserializer and select OK.

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

Reference:

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

Manage and develop data processing

QUESTION NO: 10

A company has a SaaS solution that uses Azure SQL Database with elastic pools. The solution will have a dedicated database for each customer organization. Customer organizations have peak usage at different periods during the year.

Which two factors affect your costs when sizing the Azure SQL Database elastic pools? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

A. maximum data size

B. number of databases

C. eDTUs consumption

D. number of read operations

E. number of transactions

ANSWER: A C

Explanation:

A: With the vCore purchase model, in the General Purpose tier, you are charged for Premium blob storage that you provision for your database or elastic pool. Storage can be configured between 5 GB and 4 TB with 1 GB increments. Storage is priced at GB/month.

C: In the DTU purchase model, elastic pools are available in basic, standard and premium service tiers. Each tier is distinguished primarily by its overall performance, which is measured in elastic Database Transaction Units (eDTUs).

References:

https://azure.microsoft.com/en-in/pricing/details/sql-database/elastic/

QUESTION NO: 11

A company wants to set up a NoSQL database in Azure to store data. They want to have a database that can be used to store key-value pairs. They also want to have a database that can store widecolumn data values.

Which of the following API types would you choose for Cosmos DB for these requirements? (Choose two.)

A. Table API

B. MongoDB API

C. Gremlin API

D. SQL API

E. Cassandra API

ANSWER: A E

Explanation:

The Table API can be used to store key-value pairs.

The Microsoft documentation mentions the following:

Tables, Entities, and Properties

Tables store data as collections of entities. Entities are similar to rows. An entity has a primary key and a set of properties. A property is a name, typed-value pair, similar to a column.

The Cassandra database type can be used to store wide-column data values.

Option “MongoDB API” is incorrect since this is a document-based API.

Option “Gremlin API” is incorrect since this is a graph-based database.

Option “SQL API” is incorrect since this is used to store JSON based data. Reference:

https://docs.microsoft.com/en -us/azure/cosmos-db/cassandra-introduction https://docs.microsoft.com/en- us/azure/cosmos-db/table-introduction

QUESTION NO: 12

A company has an Azure SQL Data Warehouse defined as part of their Azure subscription. The company wants to ensure their support department gets an alert when the Data Warehouse consumes the maximum allotted resources to it.

Which of the following would they use as the resource type when configuring the alert in Azure Monitor?

A. Resource Group

B. SQL server

C. SQL Data Warehouse

D. Subscription

ANSWER: C

Explanation:

Here since we need to create an alert based on the consumption of the Data Warehouse itself, we should create an alert on the Warehouse itself. The Microsoft documentation mentions the following:

Monitoring resource utilization and query activity in Azure SQL Data Warehouse

Azure SQL Data Warehouse provides a rich monitoring experience within the Azure portal to surface insights to your data warehouse workload. The Azure portal is the recommended tool when monitoring your data warehouse as it provides configurable retention periods, alerts, recommendations, and customizable charts and dashboards for metrics and logs. The portal also enables you to integrate with other Azure monitoring services such as Operations Management Suite (OMS) and Azure Monitor (logs) to provide a holistic monitoring experience for not only your data warehouse but also your entire Azure analytics platform for an integrated monitoring experience. This documentation describes what monitoring capabilities are available to optimize and manage your analytics platform with SQL Data Warehouse.

The other options are incorrect since we need to ensure the monitoring is enabled on the Data Warehouse itself. Reference:

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

QUESTION NO: 13

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 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

Explanation:

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

Reference:

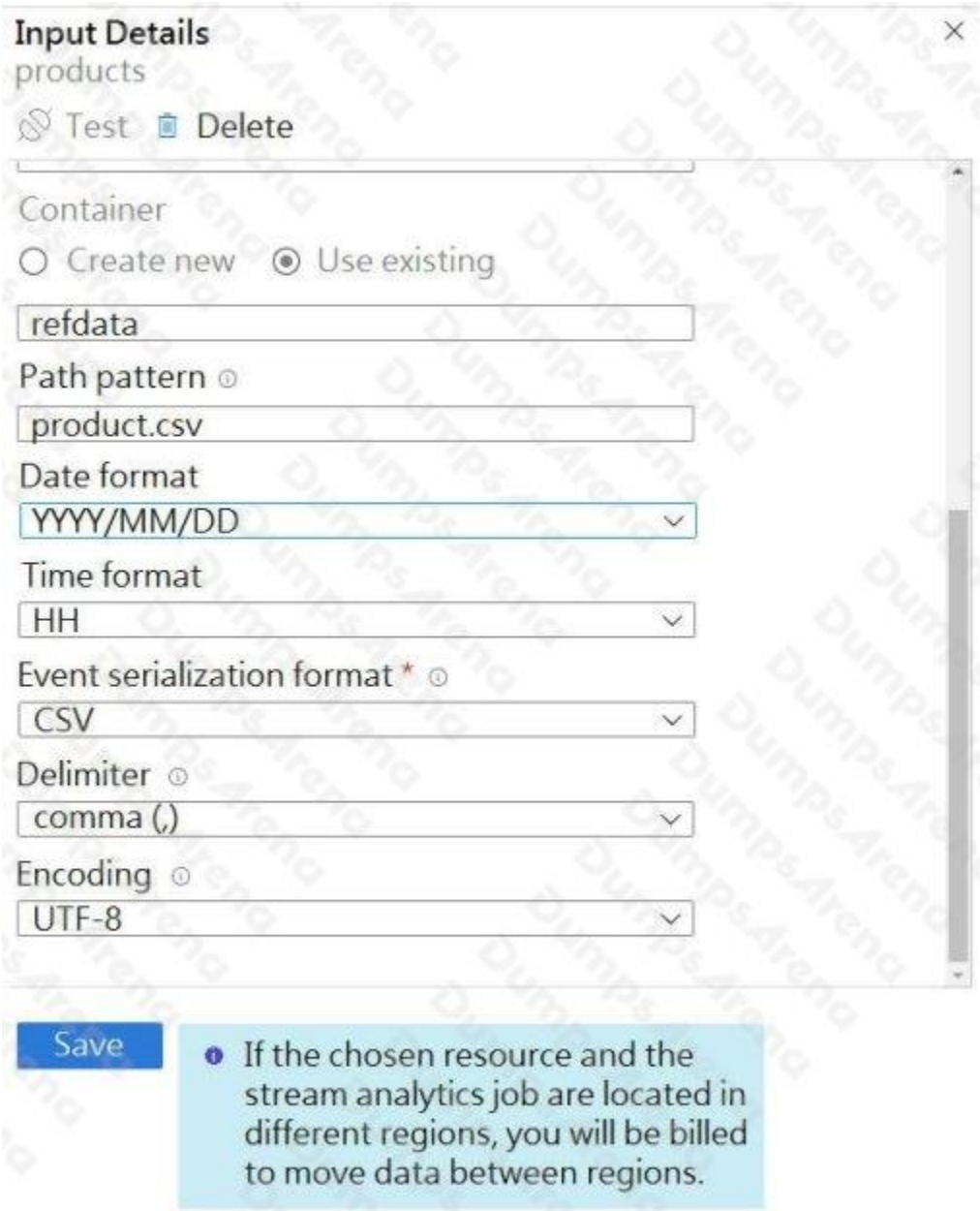
https://docs.microsoft.com/en-us/azure/sql-data-warehouse/guidance-for-loading-data

QUESTION NO: 14 - (HOTSPOT)

HOTSPOT

You are building an Azure Stream Analytics job that queries reference data from a product catalog file. The file is updated daily.

The reference data input details for the file are shown in the Input exhibit.



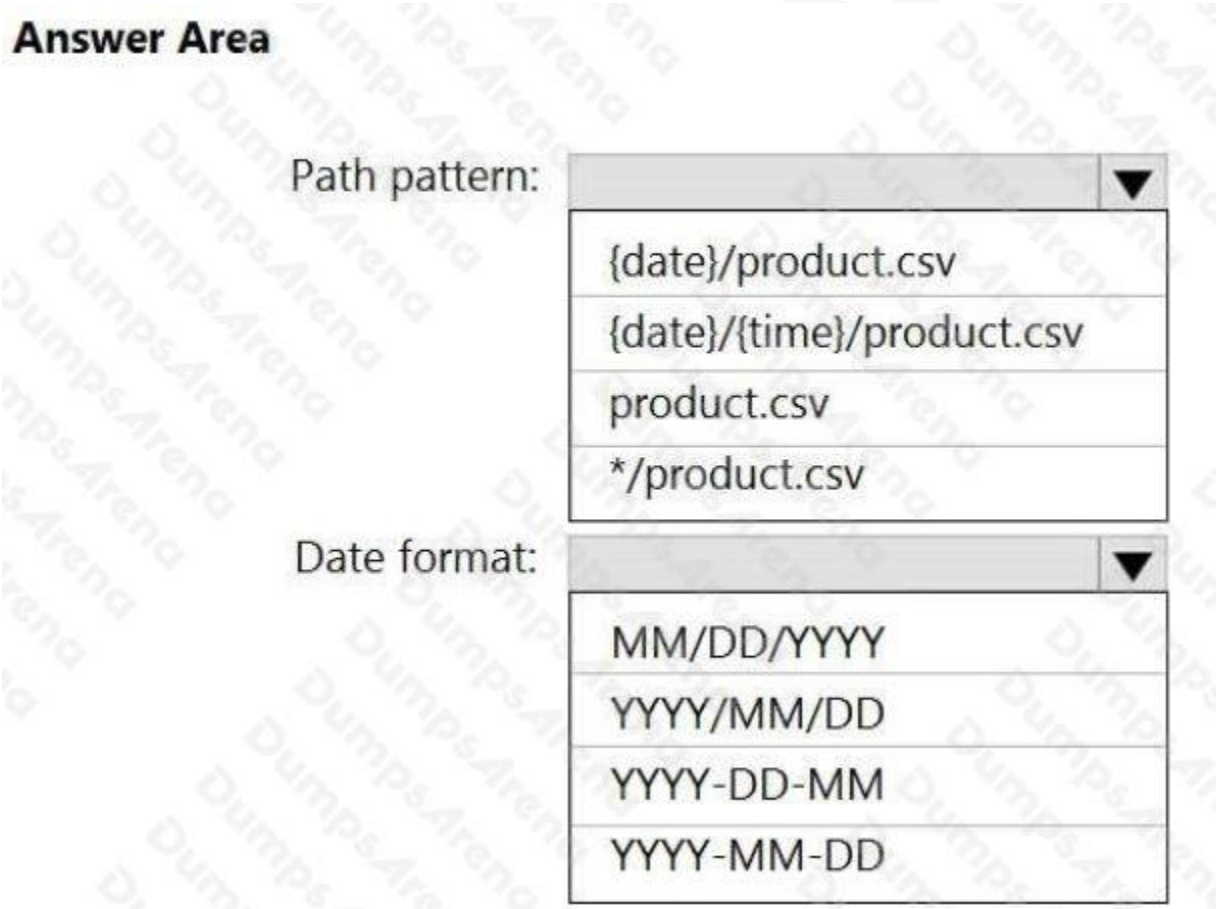
The storage account container view is shown in the Refdata exhibit.

You need to configure the Stream Analytics job to pick up the new reference data.

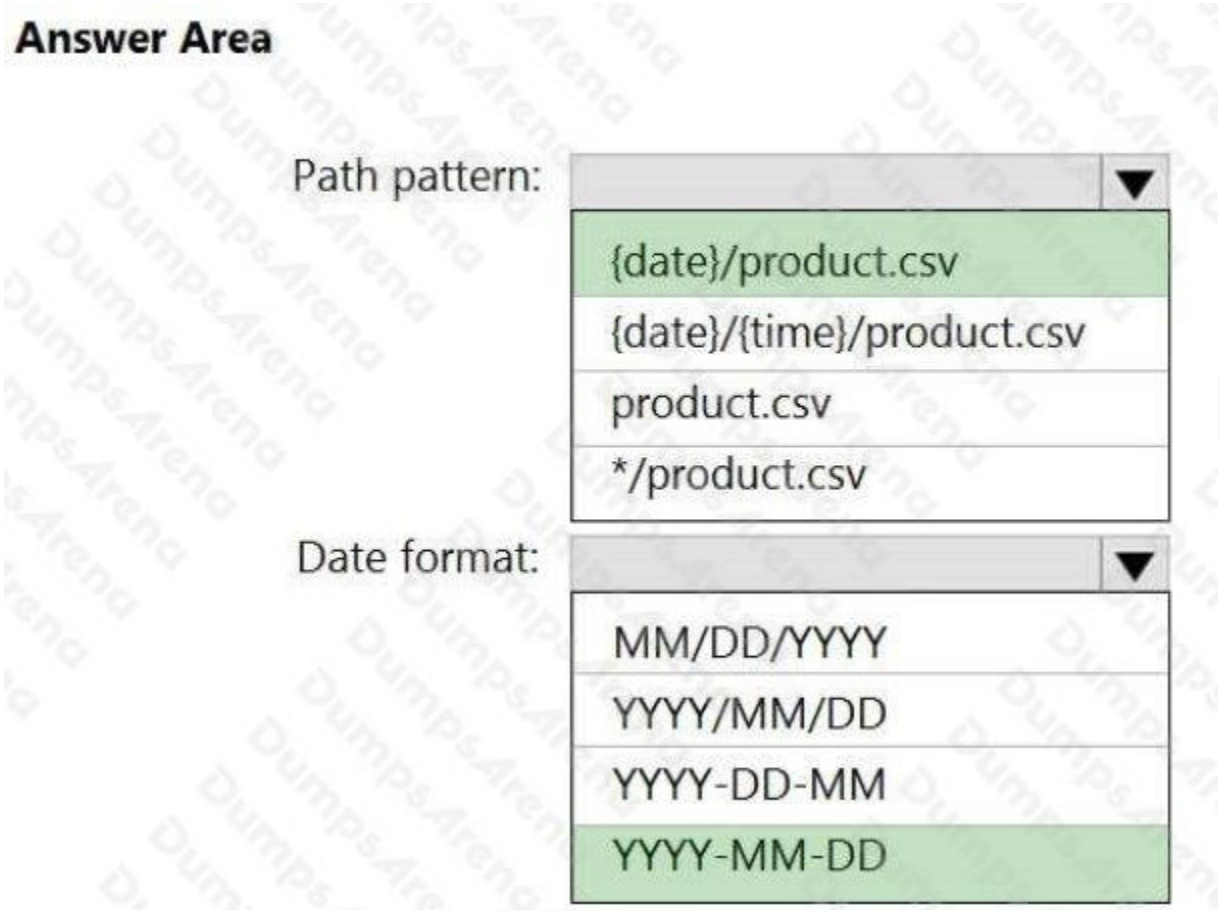
What should you configure? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:



ANSWER:



Explanation:

Box 1: {date}/product.csv

In the 2nd exhibit we see: Location: refdata / 2020-03-20

Note: Path Pattern: This is a required property that is used to locate your blobs within the specified container. Within the path, you may choose to specify one or more instances of the following 2 variables:

{date}, {time}

Example 1: products/{date}/{time}/product-list.csv

Example 2: products/{date}/product-list.csv

Example 3: product-list.csv

Box 2: YYYY-MM-DD

Note: Date Format [optional]: If you have used {date} within the Path Pattern that you specified, then you can select the date format in which your blobs are organized from the drop-down of supported formats.

Example: YYYY/MM/DD, MM/DD/YYYY, etc.

Reference:

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

Manage and develop data processing

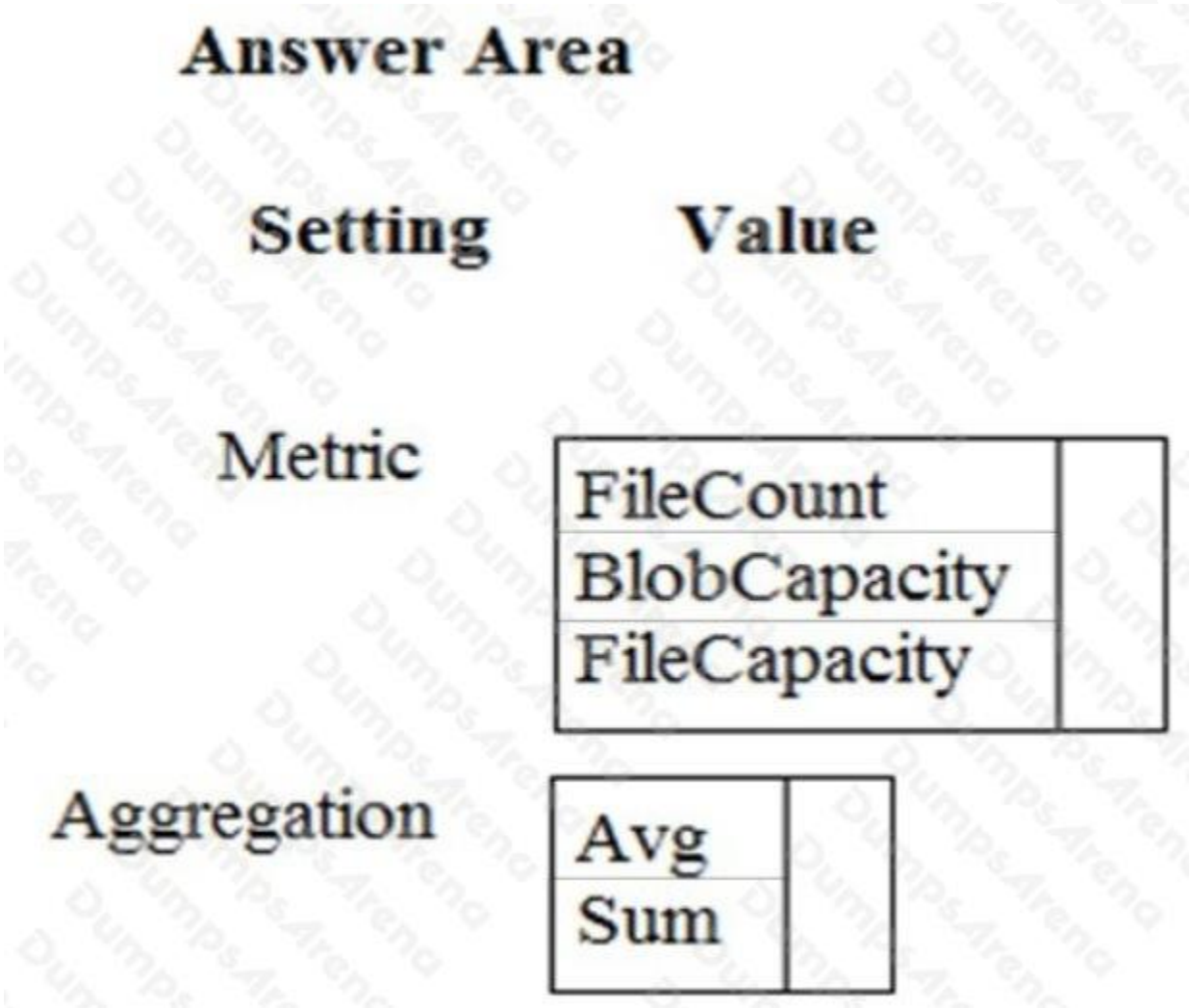
QUESTION NO: 15 - (HOTSPOT)

HOTSPOT

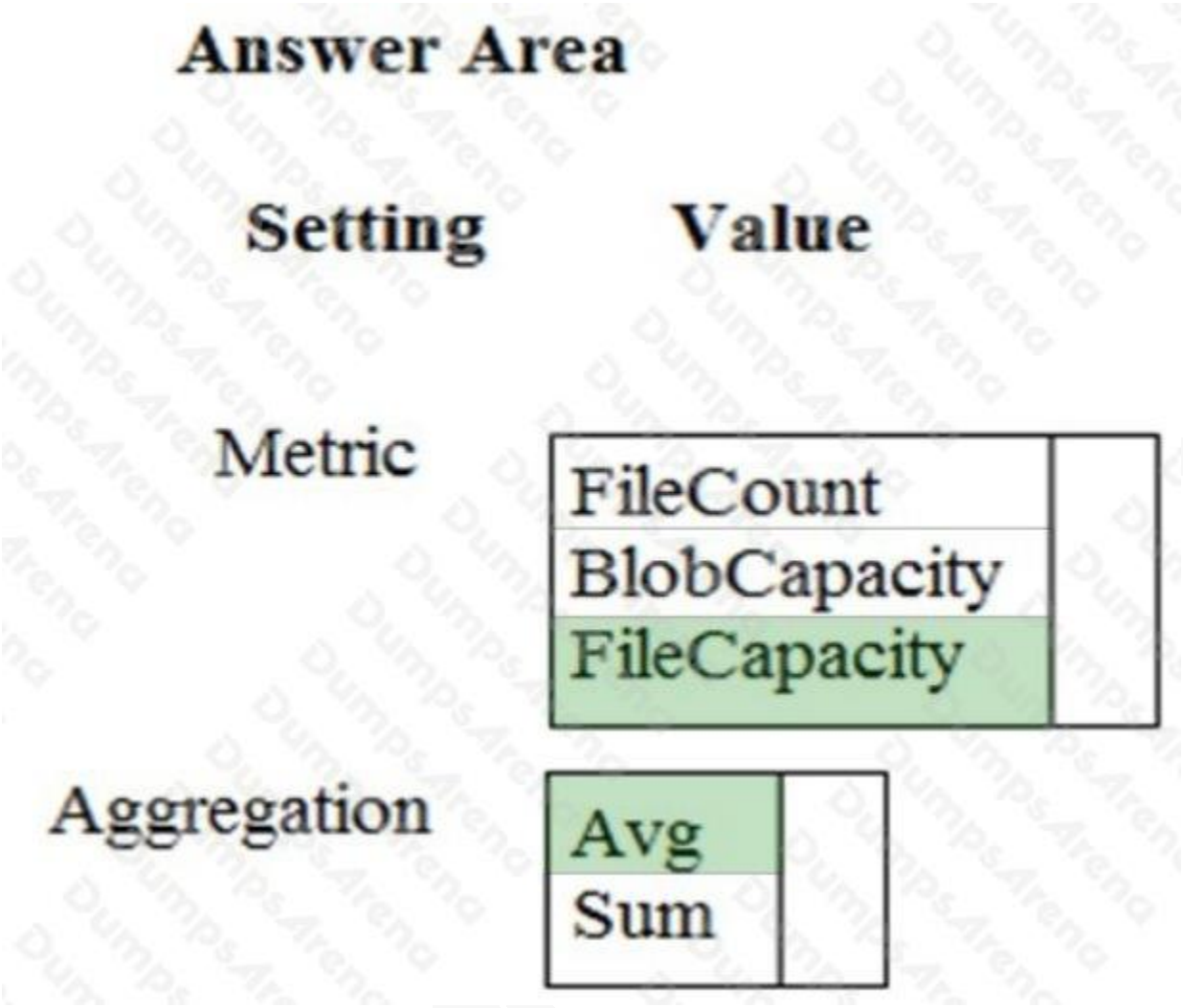
You need to ensure phone-based polling data upload reliability requirements are met. How should you configure monitoring? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:



ANSWER:



Explanation:

Box 1: FileCapacity

FileCapacity is the amount of storage used by the storage account's File service in bytes.

Box 2: Avg

The aggregation type of the FileCapacity metric is Avg.

Scenario:

All services and processes must be resilient to a regional Azure outage.

All Azure services must be monitored by using Azure Monitor. On-premises SQL Server performance must be monitored.

References: https://docs.microsoft.com/en-us/azure/azure-monitor/platform/metrics-supported Monitor and optimize data solutions