

➤ **Vendor: Microsoft**

➤ **Exam Code: DP-200**

➤ **Exam Name: Implementing an Azure Data Solution**

➤ **New Updated Questions from [Braindump2go](#) (Updated in [January/2021](#))**

### [Visit Braindump2go and Download Full Version DP-200 Exam Dumps](#)

#### QUESTION 241

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 correct selection is worth one point.

- A. Create a managed identity.
- B. Use the shared access signature (SAS) as the credentials for the data load process.
- C. Add the managed identity to the Sales group.
- D. Add your Azure Active Directory (Azure AD) account to the Sales group.
- E. Create a shared access signature (SAS).
- F. Use the managed identity as the credentials for the data load process.

**Answer:** ACD

**Explanation:**

A: 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>

#### QUESTION 242

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 subscription that contains an Azure Storage account.

You plan to implement changes to a data storage solution to meet regulatory and compliance standards.

Every day, Azure needs to identify and delete blobs that were NOT modified during the last 100 days.

Solution: You schedule an Azure Data Factory pipeline with a delete activity.

Does this meet the goal?

- A. Yes
- B. No

**Answer:** A

[DP-200 Exam Dumps](#) [DP-200 Exam Questions](#) [DP-200 PDF Dumps](#) [DP-200 VCE Dumps](#)

<https://www.braindump2go.com/dp-200.html>

**Explanation:**

You can use the Delete Activity in Azure Data Factory to delete files or folders from on-premises storage stores or cloud storage stores.

Azure Blob storage is supported.

Note: You can also apply an Azure Blob storage lifecycle policy.

Reference:

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

<https://docs.microsoft.com/en-us/azure/storage/blobs/storage-lifecycle-management-concepts?tabs=azure-portal>

**QUESTION 243**

You plan to build a structured streaming solution in Azure Databricks. The solution will count new events in five-minute intervals and report only events that arrive during the interval. The output will be sent to a Delta Lake table.

Which output mode should you use?

- A. complete
- B. update
- C. append

**Answer: C**

**Explanation:**

Append Mode: Only new rows appended in the result table since the last trigger are written to external storage. This is applicable only for the queries where existing rows in the Result Table are not expected to change.

Incorrect Answers:

A: Complete Mode: The entire updated result table is written to external storage. It is up to the storage connector to decide how to handle the writing of the entire table.

B: Update Mode: Only the rows that were updated in the result table since the last trigger are written to external storage. This is different from Complete Mode in that Update Mode outputs only the rows that have changed since the last trigger. If the query doesn't contain aggregations, it is equivalent to Append mode.

Reference:

<https://docs.databricks.com/getting-started/spark/streaming.html>

**QUESTION 244**

You have a SQL pool in Azure Synapse.

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

You need to monitor for transactions that have rolled back.

Which dynamic management view should you query?

- A. sys.dm\_pdw\_nodes\_tran\_database\_transactions
- B. sys.dm\_pdw\_waits
- C. sys.dm\_pdw\_request\_steps
- D. sys.dm\_pdw\_exec\_sessions

**Answer: A**

**Explanation:**

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

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

Example:

--Monitor rollback

SELECT

SUM(CASE WHEN t.database\_transaction\_next\_undo\_lsn IS NOT NULL THEN 1 ELSE 0 END),

E. pdw\_node\_id,

nod.[type]

FROM sys.dm\_pdw\_nodes\_tran\_database\_transactions t

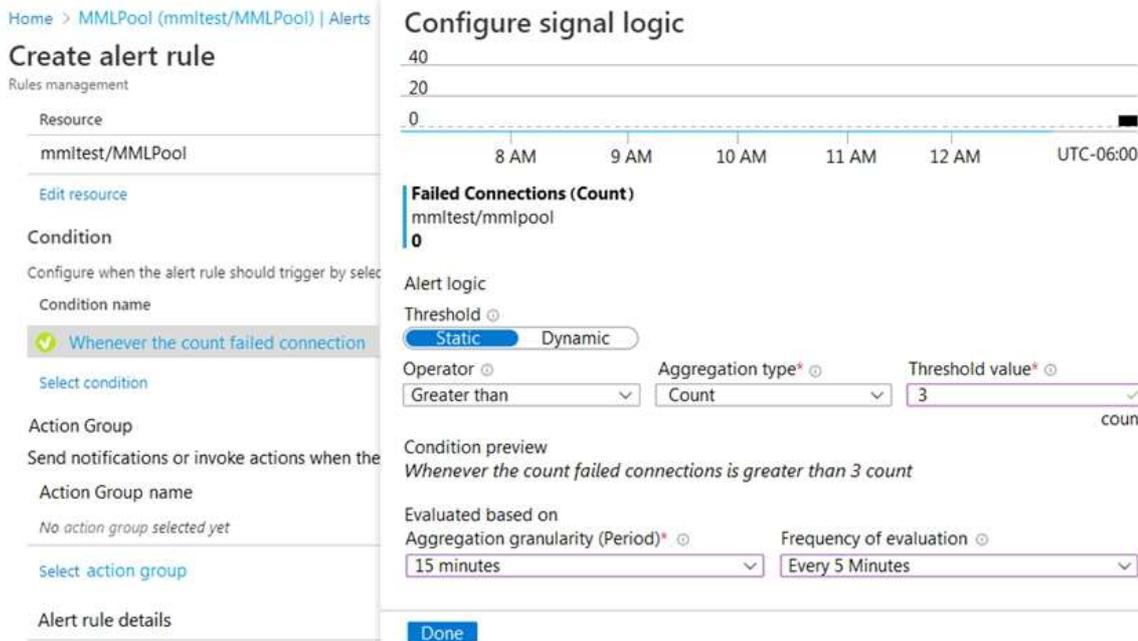
JOIN sys.dm\_pdw\_nodes nod ON t.pdw\_node\_id = nod.pdw\_node\_id GROUP BY t.pdw\_node\_id, nod.[type]

Reference:

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

**QUESTION 245**

You have an alert on a SQL pool in Azure Synapse that uses the signal logic shown in the exhibit.



The screenshot shows the 'Configure signal logic' interface for an alert rule named 'Failed Connections (Count)' on the resource 'mmltest/mmlpool'. The alert logic is configured as follows:

- Alert logic:** Static
- Threshold:** 3
- Operator:** Greater than
- Aggregation type:** Count
- Threshold value:** 3
- Condition preview:** Whenever the count failed connections is greater than 3 count
- Evaluated based on:**
  - Aggregation granularity (Period): 15 minutes
  - Frequency of evaluation: Every 5 Minutes

On the same day, failures occur at the following times:

- 08:01
- 08:03
- 08:04
- 08:06
- 08:11
- 08:16
- 08:19

The evaluation period starts on the hour.

At which times will alert notifications be sent?

- A. 08:15 only
- B. 08:10, 08:15, and 08:20
- C. 08:05 and 08:10 only
- D. 08:10 only
- E. 08:05 only

**Answer:** B

**Explanation:**

<https://docs.microsoft.com/en-us/azure/azure-sql/database/alerts-insights-configure-portal>

**QUESTION 246**

You plan to monitor the performance of Azure Blob storage by using Azure Monitor.

You need to be notified when there is a change in the average time it takes for a storage service or API operation type to process requests.

For which two metrics should you set up alerts? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. SuccessE2ELatency
- B. SuccessServerLatency

- C. UsedCapacity
- D. Egress
- E. Ingress

**Answer:** AB

**Explanation:**

Success E2E Latency: The average end-to-end latency of successful requests made to a storage service or the specified API operation. This value includes the required processing time within Azure Storage to read the request, send the response, and receive acknowledgment of the response.

Success Server Latency: The average time used to process a successful request by Azure Storage. This value does not include the network latency specified in SuccessE2ELatency.

Reference:

<https://docs.microsoft.com/en-us/azure/storage/blobs/storage-blob-scalable-app-verify-metrics>

#### **QUESTION 247**

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 is not found.

You need to identify the cause of the issue.

What should you review?

- A. workspace logs
- B. notebook logs
- C. global init scripts logs
- D. cluster event logs

**Answer:** C

**Explanation:**

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>

#### **QUESTION 248**

Hotspot Question

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 Area

Distribution: 

	▼
Hash	
Replicated	
Round-robin	

Indexing: 

	▼
Clustered	
Clustered columnstore	
Heap	

Partitioning: 

	▼
Date	
None	

Answer:

## Answer Area

Distribution:

Indexing:

Partitioning:

**Explanation:**

Box 1: Hash

Hash-distributed tables improve query performance on large fact tables. they can have very large numbers of rows and still achieve high performance.

Incorrect:

Round-robin tables are useful for improving loading speed.

Box 2: Clustered columnstore

When creating partitions on clustered columnstore tables, it is important to consider how many rows belong to each partition. For optimal compression and performance of clustered columnstore tables, a minimum of 1 million rows per distribution and partition is needed.

Box 3: Date

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

Partition switching can be used to quickly remove or replace a section of a table.

Reference:

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

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

**QUESTION 249**

Hotspot Question

You have two Azure Storage accounts named Storage1 and Storage2. Each account contains an Azure Data Lake Storage file system. 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.

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

[DP-200 Exam Dumps](#) [DP-200 Exam Questions](#) [DP-200 PDF Dumps](#) [DP-200 VCE Dumps](#)

<https://www.braindump2go.com/dp-200.html>

NOTE: Each correct selection is worth one point.

### Answer Area

Source dataset type:

	▼
Binary	
Parquet	
Delimited text	

Copy activity copy behavior:

	▼
FlattenHierarchy	
MergeFiles	
PreserveHierarchy	

Answer:

### Answer Area

Source dataset type:

	▼
Binary	
Parquet	
Delimited text	

Copy activity copy behavior:

	▼
FlattenHierarchy	
MergeFiles	
PreserveHierarchy	

#### Explanation:

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.

Incorrect Answers:

FlattenHierarchy: All files from the source folder are in the first level of the target folder. The target files have autogenerated names.

MergeFiles: Merges all files from the source folder to one file. If the file name is specified, the merged file name is the specified name. Otherwise, it's an autogenerated file name.

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>

**QUESTION 250**

Hotspot Question

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.

**Answer Area**

Statements	Yes	No
The query combines two streams of partitioned data.	<input type="radio"/>	<input type="radio"/>
The stream scheme key and count must match the output scheme.	<input type="radio"/>	<input type="radio"/>
Providing 60 streaming units will optimize the performance of the query.	<input type="radio"/>	<input type="radio"/>

Answer:

## Answer Area

Statements	Yes	No
The query combines two streams of partitioned data.	<input checked="" type="radio"/>	<input type="radio"/>
The stream scheme key and count must match the output scheme.	<input checked="" type="radio"/>	<input type="radio"/>
Providing 60 streaming units will optimize the performance of the query.	<input checked="" type="radio"/>	<input type="radio"/>

### Explanation:

Box 1: Yes

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  $6 \times 10 = 60$  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>

### QUESTION 251

Hotspot Question

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 Area**

```

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:

**Answer Area**

```

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'
  
```

**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>

### QUESTION 252

Hotspot Question

You have an Azure Stream Analytics job named ASA1.

The Diagnostic settings for ASA1 are configured to write errors to Log Analytics.

ASA1 reports an error, and the following message is sent to Log Analytics.

```
{
"properties_s" : {
  "Source": "InputTelemetryData",
  "Type": "DataError",
  "DataErrorType": "InputDeserializerError.InvalidData",
  "BriefMessage": "Json input stream should either be an array of objects or
line separated objects. Found token type: Integer",
  "Message": "Error: Json input stream should either be an array of objects
or line separated objects. Found token type: Integer",
  "ExampleEvents": "",
  "FromTimestamp": "2020-07-22T22:34:18.5664937Z",
  "ToTimestamp": "2020-07-22T22:34:18.5965248Z",
  "EventCount": 2
}
}
```

You need to write a Kusto query language query to identify all instances of the error and return the message field. How should you complete the query? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

#### Answer Area

AzureDiagnostics

| where ResourceProvider == "MICROSOFT.STREAMANALYTICS" and

	- "InputDeserializerError.InvalidData"
<input type="checkbox"/> DataErrorType	
<input type="checkbox"/> Message	
<input type="checkbox"/> parse_json(properties_s).DataErrorType	
<input type="checkbox"/> parse_json(properties_s).Message	

| project TimeGenerated,

<input type="checkbox"/> DataErrorType	
<input type="checkbox"/> Message	
<input type="checkbox"/> parse_json(properties_s).DataErrorType	
<input type="checkbox"/> parse_json(properties_s).Message	

Answer:

[DP-200 Exam Dumps](#) [DP-200 Exam Questions](#) [DP-200 PDF Dumps](#) [DP-200 VCE Dumps](#)

<https://www.braindump2go.com/dp-200.html>

**Answer Area**

AzureDiagnostics

| where ResourceProvider -- "MICROSOFT.STREAMANALYTICS" and

	- "InputDeserializerError.InvalidData"
DataErrorType	
Message	
parse_json(properties_s).DataErrorType	
parse_json(properties_s).Message	

| project TimeGenerated,

	▼
DataErrorType	
Message	
parse_json(properties_s).DataErrorType	
parse_json(properties_s).Message	

**Explanation:**

Box 1: DataErrorType

The DataErrorType is InputDeserializerError.InvalidData.

Box 2: Message

Retrieve the message.

Reference:

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

**QUESTION 253**

Hotspot Question

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.

**Answer Area**

WITH LastInWindow AS

```

(
    SELECT
        

|        |   |
|--------|---|
|        | ▼ |
| COUNT  |   |
| MAX    |   |
| MIN    |   |
| TOPONE |   |

 (Time) AS LastEventTime
    FROM
        Input TIMESTAMP BY Time
    GROUP BY
        

|                |   |
|----------------|---|
|                | ▼ |
| HoppingWindow  |   |
| SessionWindow  |   |
| SlidingWindow  |   |
| TumblingWindow |   |

 (minute, 10)
)
SELECT
    Input.License_plate,
    Input.Make,
    Input.Time
FROM
    Input TIMESTAMP BY Time
    INNER JOIN LastInWindow
ON
    

|          |   |
|----------|---|
|          | ▼ |
| DATEADD  |   |
| DATEDIFF |   |
| DATENAME |   |
| DATEPART |   |

 (minute, Input, LastInWindow) BETWEEN 0 AND 10
    AND Input.Time = LastInWindow.LastEventTime

```

**Answer:**

**Answer Area**

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

```

**Explanation:**

Box 1: MAX

The first step on the query finds the maximum time stamp in 10-minute windows, that is the time stamp of the last event for that window. The second step joins the results of the first query with the original stream to find the event that match the last time stamps in each window.

Query:

```

WITH LastInWindow AS
(
    SELECT
    MAX(Time) AS LastEventTime
    FROM
    Input TIMESTAMP BY Time
    GROUP BY
    TumblingWindow(minute, 10)
)
SELECT
Input.License_plate,
Input.Make,
Input.Time
FROM
Input TIMESTAMP BY Time

```

INNER JOIN LastInWindow  
ON DATEDIFF(minute, Input, LastInWindow) BETWEEN 0 AND 10 AND Input.Time = LastInWindow.LastEventTime

Box 2: TumblingWindow

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

Box 3: DATEDIFF

DATEDIFF is a date-specific function that compares and returns the time difference between two DateTime fields, for more information, refer to date functions.

Reference:

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

**QUESTION 254**

Drag and Drop Question

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.

**Actions**

Add .NET deserializer code for Protobuf to the Stream Analytics project

Change the Event Serialization Format to Protobuf in the Input.json file of the job and reference the DLL

Add an Azure Stream Analytics Application project to the solution

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

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

**Answer Area**

**Answer:**

**Actions**

Add .NET deserializer code for Protobuf to the Stream Analytics project

Change the Event Serialization Format to Protobuf in the Input.json file of the job and reference the DLL

**Answer Area**

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

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

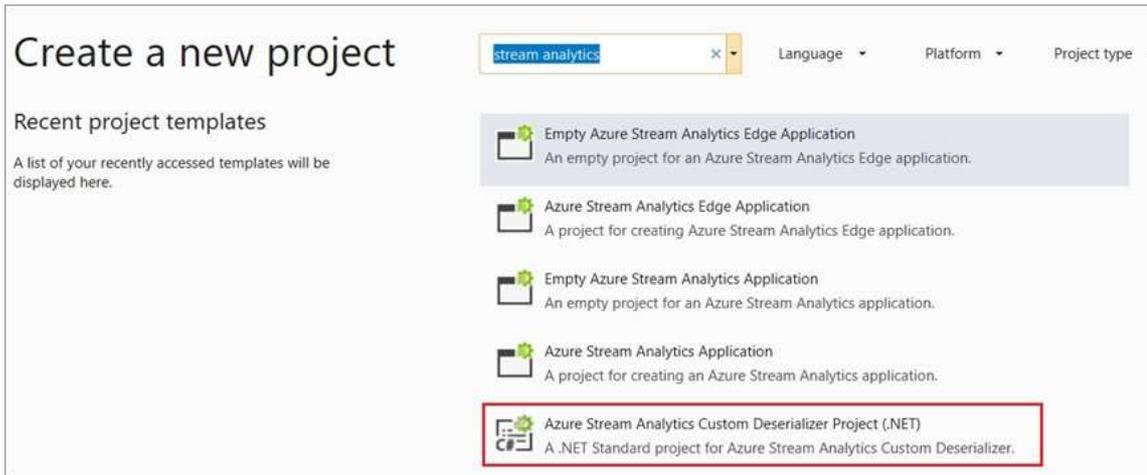
Add an Azure Stream Analytics Application project to the solution

**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, like Protobuf Deserializer.



2. In Solution Explorer, right-click your Protobuf Deserializer project and select Manage NuGet Packages from the menu. Then install the Microsoft.Azure.StreamAnalytics and Google.Protobuf NuGet packages.
3. Add the MessageBodyProto class and the MessageBodyDeserializer class to your project.
4. Build the Protobuf Deserializer project.

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

Step 3: Add an Azure Stream Analytics Application project to the solution Add an Azure Stream Analytics project  
1. In Solution Explorer, right-click the Protobuf Deserializer solution and select Add > New Project. Under Azure Stream Analytics > Stream Analytics, choose Azure Stream Analytics Application. Name it ProtobufCloudDeserializer and select OK.

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>

### QUESTION 255

Drag and Drop Question

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.

**Actions**

**Answer Area**

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

**Answer:**

**Actions**

**Answer Area**

- Select the PipelineRuns category.
- Stream to an Azure event hub.
- Select the TriggerRuns category.

- Create an Azure Storage account that has a lifecycle policy.
- Create a Log Analytics workspace that has Data Retention set to 120 days.
- From the Azure portal, add a diagnostic setting.
- Send the data to a Log Analytics workspace.

**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.

1. In the portal, go to Monitor. Select Settings > Diagnostic settings.

2. Select the data factory for which you want to set a diagnostic setting.

3. If no settings exist on the selected data factory, you're prompted to create a setting. Select Turn on diagnostics.

4. Give your setting a name, select Send to Log Analytics, and then select a workspace from Log Analytics Workspace.

5. Select Save.

Reference:

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

**QUESTION 256**

Hotspot Question

You have an Azure data factory that has two pipelines named PipelineA and PipelineB.

PipelineA has four activities as shown in the following exhibit.



PipelineB has two activities as shown in the following exhibit.



You create an alert for the data factory that uses Failed pipeline runs metrics for both pipelines and all failure types.

The metric has the following settings:

- Operator: Greater than
- Aggregation type: Total
- Threshold value: 2
- Aggregation granularity (Period): 5 minutes
- Frequency of evaluation: Every 5 minutes

Data Factory monitoring records the failures shown in the following table.

Pipeline	Activity	Time
PipelineA	Activity1	31-Jan-2020 10:44:00
PipelineA	Activity3	31-Jan-2020 10:47:00
PipelineB	Activity1	31-Jan-2020 10:50:00

For each of the following statements, select yes if the statement is true. Otherwise, select no.

**NOTE:** Each correct answer selection is worth one point.

Statements	Yes	No
An alert notification was sent after the failure of Activity1 in PipelineA.	<input type="radio"/>	<input type="radio"/>
An alert notification was sent after the failure of Activity3 in PipelineA.	<input type="radio"/>	<input type="radio"/>
An alert notification was sent after the failure of Activity1 in PipelineB.	<input type="radio"/>	<input type="radio"/>

**Answer:**

Statements	Yes	No
An alert notification was sent after the failure of Activity1 in PipelineA.	<input type="radio"/>	<input checked="" type="radio"/>
An alert notification was sent after the failure of Activity3 in PipelineA.	<input type="radio"/>	<input checked="" type="radio"/>
An alert notification was sent after the failure of Activity1 in PipelineB.	<input checked="" type="radio"/>	<input type="radio"/>

**Explanation:**

Box 1: No

Only one failure at this point.

Box 2: No

Only two failures within 5 minutes.

Box 3: Yes

More than two (three) failures in 5 minutes

Reference:

<https://docs.microsoft.com/en-us/azure/azure-sql/database/alerts-insights-configure-portal>