

Microsoft

Exam Questions DP-203

Data Engineering on Microsoft Azure



NEW QUESTION 1

- (Exam Topic 2)

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.

Integration runtime type:

- Azure integration runtime
- Azure-SSIS integration runtime
- Self-hosted integration runtime

Trigger type:

- Event-based trigger
- Schedule trigger
- Tumbling window trigger

Activity type:

- Copy activity
- Lookup activity
- Stored procedure activity

- A. Mastered
- B. Not Mastered

Answer: A

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 2

- (Exam Topic 3)

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

Library:

- Azure Databricks Monitoring Library
- Microsoft Azure Management Monitoring Library
- PyTorch
- TensorFlow

Workspace:

- Azure Databricks
- Azure Log Analytics
- Azure Machine Learning

- A. Mastered
- B. Not Mastered

Answer: A

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 3

- (Exam Topic 3)

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.

Values

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

Answer Area

```
CleanData
split(
    [ ]
) ~> SplitByDept@[ ]
```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

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

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

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

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

Box 2: discount : false

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

Box 3: ecommerce, retail, wholesale, all Label the streams

Reference:

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

NEW QUESTION 4

- (Exam Topic 3)

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.

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

- A. Mastered
- B. Not Mastered

Answer: A

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 5

- (Exam Topic 3)

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

Explanation:

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

Always Encrypted is a feature designed to protect sensitive data stored in specific database columns from access (for example, credit card numbers, national identification numbers, or data on a need to know basis). This includes database administrators or other privileged users who are authorized to access the database to perform management tasks, but have no business need to access the particular data in the encrypted columns. The data is always encrypted, which means the encrypted data is decrypted only for processing by client applications with access to the encryption key.

References:

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

NEW QUESTION 6

- (Exam Topic 3)

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

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

What should you recommend?

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

Answer: A

NEW QUESTION 7

- (Exam Topic 3)

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)
    FROM Temperatures
    WHERE date BETWEEN DATE '2019-01-01' AND DATE
    '2021-08-31'
    Value (
        Value (Temp AS DECIMAL(4, 1)))
    AVG (
        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
    
```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Values

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

Answer Area

```

SELECT * FROM (
    SELECT YEAR(Date) Year, MONTH(Date)
    FROM Temperatures
    WHERE date BETWEEN DATE '2019-01-01' AND DATE
    '2021-08-31'
    CONVERT (
        COLLATE (Temp AS DECIMAL(4, 1)))
    AVG (
        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
    
```

NEW QUESTION 8

- (Exam Topic 3)

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

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

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

What should you include in the solution? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point

Azure Stream Analytics input type:

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

Azure Stream Analytics output type:

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

Aggregation query location:

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

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

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

NEW QUESTION 9

- (Exam Topic 3)

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

Explanation:

Azure Stream Analytics on IoT Edge empowers developers to deploy near-real-time analytical intelligence closer to IoT devices so that they can unlock the full value of device-generated data. UDF are available in C# for IoT Edge jobs
 Azure Stream Analytics on IoT Edge runs within the Azure IoT Edge framework. Once the job is created in Stream Analytics, you can deploy and manage it using IoT Hub.

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

NEW QUESTION 10

- (Exam Topic 3)

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: BC

Explanation:

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

NEW QUESTION 10

- (Exam Topic 3)

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: B

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 11

- (Exam Topic 3)

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.

```

1  SELECT c.name,
2     tbl.name as table_name,
3     typ.name as datatype,
4     c.is_masked,
5     c.masking_function
6  FROM sys.masked_columns AS c
7  INNER JOIN sys.tables AS tbl ON c.[object_id] = tbl.[object_id]
8  INNER JOIN sys.types typ ON c.user_type_id = typ.user_type_id
9  WHERE is_masked = 1;
10

```

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 Area

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

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area

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

NEW QUESTION 14

- (Exam Topic 3)

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

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 17

- (Exam Topic 3)

You have an Azure Data Lake Storage Gen2 account that contains a JSON file for customers. The file contains two attributes named FirstName and LastName. You need to copy the data from the JSON file to an Azure Synapse Analytics table by using Azure Databricks. A new column must be created that concatenates the FirstName and LastName values.

You create the following components:

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

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

Actions

Answer Area

- Mount the Data Lake Storage onto DBFS.
- Write the results to a table in Azure Synapse.
- Perform transformations on the file.
- Specify a temporary folder to stage the data.
- Write the results to Data Lake Storage.
- Read the file into a data frame.
- Drop the data frame.
- Perform transformations on the data frame.

- A. Mastered
- B. Not Mastered

Answer: A

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 22

- (Exam Topic 3)

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

```
{
  "rules": [
    {
      "enabled": true,
      "name": "contosorule",
      "type": "lifecycle",
      "definition": {
        "actions": {
          "version": {
            "delete": {
              "daysAfterCreationGreaterThan": 60
            }
          }
        },
        "baseBlob": {
          "tierToCool": {
            "daysAfterModificationGreaterThan": 30
          }
        }
      }
    }
  ],
  "filters": {
    "blobTypes": [
      "blockBlob"
    ],
    "prefixMatch": [
      "container1/contoso"
    ]
  }
}
```

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

NOTE: Each correct selection is worth one point.

Answer Area

The files are [answer choice] after 30 days.

The storage policy applies to [answer choice].

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area

The files are [answer choice] after 30 days.

The storage policy applies to [answer choice].

NEW QUESTION 23

- (Exam Topic 3)

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.

To track encryption key usage:

▼

Always Encrypted

TDE with customer-managed keys

TDE with platform-managed keys

To maintain client app access in the event of a datacenter outage:

▼

Create and configure Azure key vaults in two Azure regions.

Enable Advanced Data Security on Server1.

Implement the client apps by using a Microsoft .NET Framework data provider.

- A. Mastered
- B. Not Mastered

Answer: A

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 25

- (Exam Topic 3)

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

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

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

Actions

Answer Area

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

- A. Mastered

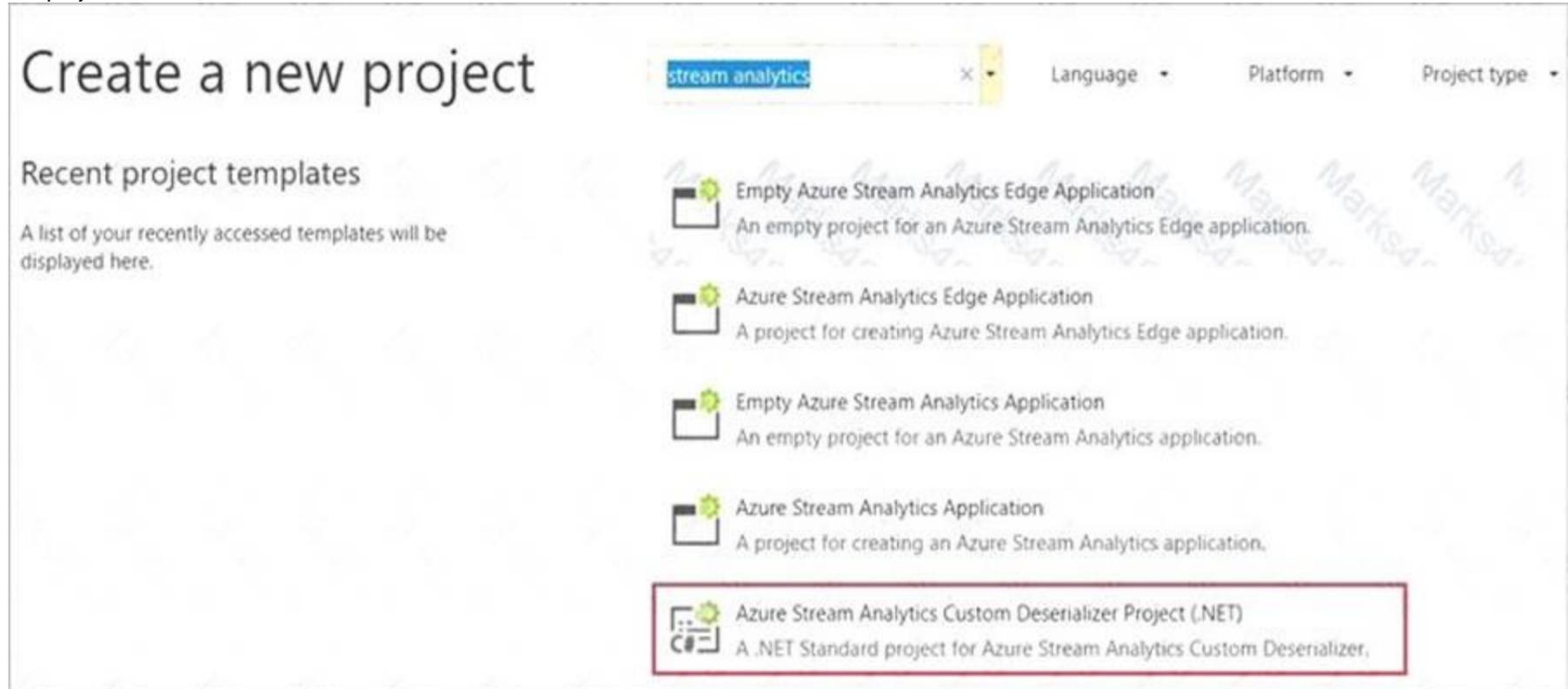
B. Not Mastered

Answer: A

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.

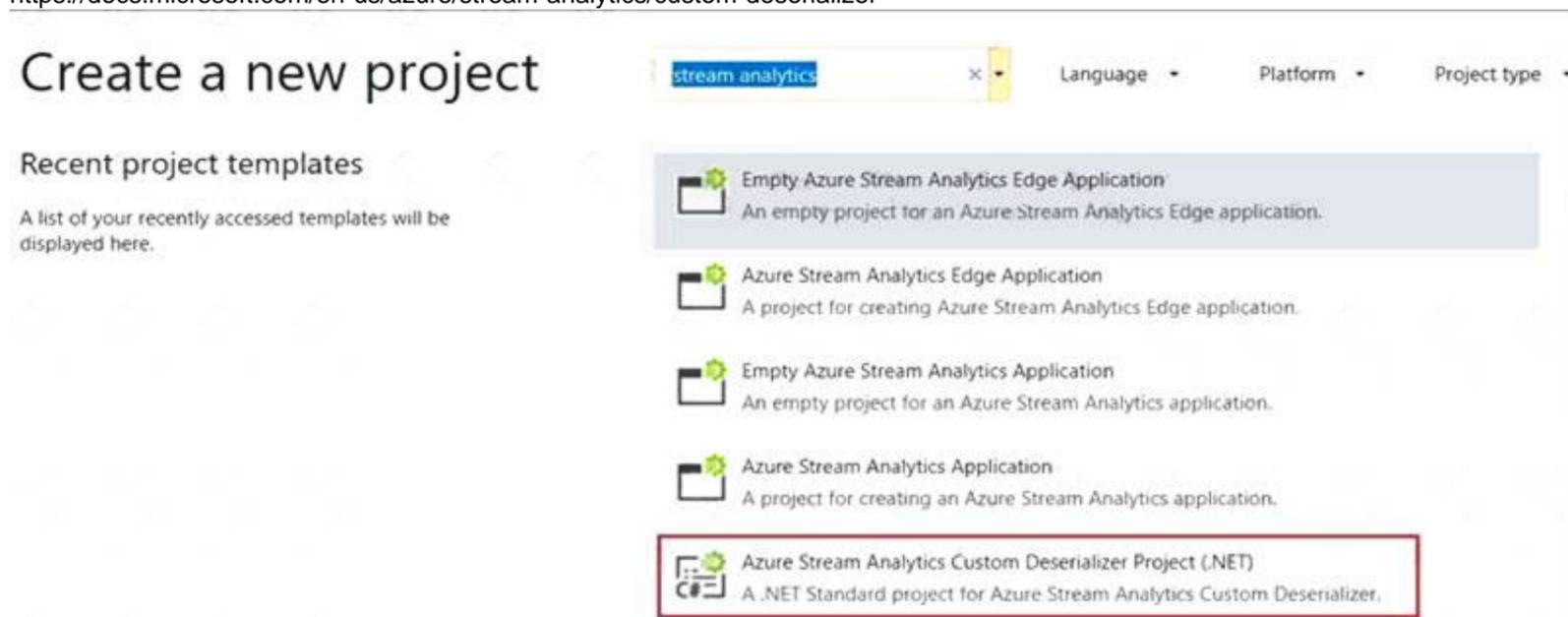
Step 3: Add an Azure Stream Analytics Application project to the solution Add an Azure Stream Analytics project

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

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

Reference:

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



NEW QUESTION 30

- (Exam Topic 3)

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.

- A. Mastered
- B. Not Mastered

Answer: A

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 34

- (Exam Topic 3)

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

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 37

- (Exam Topic 3)

You are creating dimensions for a data warehouse in an Azure Synapse Analytics dedicated SQL pool. You create a table by using the Transact-SQL statement

shown in the following exhibit.

```
CREATE TABLE [DBO].[DimProduct] (
    [ProductKey] [int] IDENTITY(1,1) NOT NULL,
    [ProductSourceID] [int] NOT NULL,
    [ProductName] [nvarchar](100) NOT NULL,
    [ProductNumber] [nvarchar](25) NOT NULL,
    [Color] [nvarchar](15) NULL,
    [Size] [nvarchar](5) NULL,
    [Weight] [decimal](8, 2) NULL,
    [ProductCategory] [nvarchar](100) NULL,
    [SellStartDate] [date] NOT NULL,
    [SellEndDate] [date] NULL,
    [RowInsertedDateTime] [datetime] NOT NULL,
    [RowUpdatedDateTime] [datetime] NOT NULL,
    [ETLAuditID] [int] NOT NULL
)
```

Use the drop-down menus to select the answer choice that completes each statement based on the information presented in the graphic.
 NOTE: Each correct selection is worth one point.

DimProduct is a **[answer choice]** slowly changing dimension (SCD).

▼
Type 0
Type 1
Type 2

The ProductKey column is **[answer choice]**.

▼
a surrogate key
a business key
an audit column

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Type 2

A Type 2 SCD supports versioning of dimension members. Often the source system doesn't store versions, so the data warehouse load process detects and manages changes in a dimension table. In this case, the dimension table must use a surrogate key to provide a unique reference to a version of the dimension member. It also includes columns that define the date range validity of the version (for example, StartDate and EndDate) and possibly a flag column (for example, IsCurrent) to easily filter by current dimension members.

Reference:

<https://docs.microsoft.com/en-us/learn/modules/populate-slowly-changing-dimensions-azure-synapse-analytics>

NEW QUESTION 40

- (Exam Topic 3)

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: AE

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-resou>

NEW QUESTION 44

- (Exam Topic 3)

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. Append
- B. Update
- C. Complete

Answer: C

NEW QUESTION 45

- (Exam Topic 3)

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 Area:

Integration runtime type:	<input type="checkbox"/> Azure integration runtime <input checked="" type="checkbox"/> Azure-SSIS integration runtime <input type="checkbox"/> Self-hosted integration runtime
Trigger type:	<input type="checkbox"/> Event-based trigger <input checked="" type="checkbox"/> Schedule trigger <input type="checkbox"/> Tumbling window trigger
Activity type:	<input type="checkbox"/> Copy activity <input checked="" type="checkbox"/> Lookup activity <input type="checkbox"/> Stored procedure activity

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area:

Integration runtime type:	<input type="checkbox"/> Azure integration runtime <input checked="" type="checkbox"/> Azure-SSIS integration runtime <input type="checkbox"/> Self-hosted integration runtime
Trigger type:	<input type="checkbox"/> Event-based trigger <input checked="" type="checkbox"/> Schedule trigger <input type="checkbox"/> Tumbling window trigger
Activity type:	<input type="checkbox"/> Copy activity <input checked="" type="checkbox"/> Lookup activity <input type="checkbox"/> Stored procedure activity

NEW QUESTION 46

- (Exam Topic 3)

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. Reregister the Microsoft Data Lake Store resource provider.
- B. Reregister the Azure Storage resource provider.
- C. Create a storage policy that is scoped to a container.
- D. Register the query acceleration feature.
- E. Create a storage policy that is scoped to a container prefix filter.

Answer: BD

NEW QUESTION 47

- (Exam Topic 3)

You are designing a sales transactions table in an Azure Synapse Analytics dedicated SQL pool. The table will contain approximately 60 million rows per month and will be partitioned by month. The table will use a clustered column store index and round-robin distribution. Approximately how many rows will there be for each combination of distribution and partition?

- A. 1 million
- B. 5 million
- C. 20 million
- D. 60 million

Answer: D

Explanation:

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

NEW QUESTION 52

- (Exam Topic 3)

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

NEW QUESTION 57

- (Exam Topic 3)

You have an Azure Stream Analytics query. The query returns a result set that contains 10,000 distinct values for a column named clusterID. You monitor the Stream Analytics job and discover high latency. You need to reduce the latency. Which two actions should you perform? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

- A. Add a pass-through query.
- B. Add a temporal analytic function.
- C. Scale out the query by using PARTITION BY.
- D. Convert the query to a reference query.
- E. Increase the number of streaming units.

Answer: CE

Explanation:

C: Scaling a Stream Analytics job takes advantage of partitions in the input or output. Partitioning lets you divide data into subsets based on a partition key. A process that consumes the data (such as a Streaming Analytics job) can consume and write different partitions in parallel, which increases throughput.

E: 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. This capacity lets you focus on the query logic and abstracts the need to manage the hardware to run your Stream Analytics job in a timely manner.

References:

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

NEW QUESTION 62

- (Exam Topic 3)

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

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

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

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

Actions

Answer Area

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

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Step 1: Create a database role named Role1 and grant Role1 SELECT permissions to schema You need to grant Group1 read-only permissions to all the tables and views in schema1.
 Place one or more database users into a database role and then assign permissions to the database role. Step 2: Assign Rol1 to the Group database user
 Step 3: Assign the Azure role-based access control (Azure RBAC) Reader role for dw1 to Group1 Reference:
<https://docs.microsoft.com/en-us/azure/data-share/how-to-share-from-sql>

NEW QUESTION 66

- (Exam Topic 3)

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.

Type:

Event
On-demand
Schedule
Tumbling window

Additional properties:

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

- A. Mastered
- B. Not Mastered

Answer: A

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.

Reference:

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

NEW QUESTION 71

- (Exam Topic 3)

You have an Azure Storage account and a data warehouse in Azure Synapse Analytics in the UK South region. You need to copy blob data from the storage account to the data warehouse by using Azure Data Factory. The solution must meet the following requirements:

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

Which type of integration runtime should you use?

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

Answer: A

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

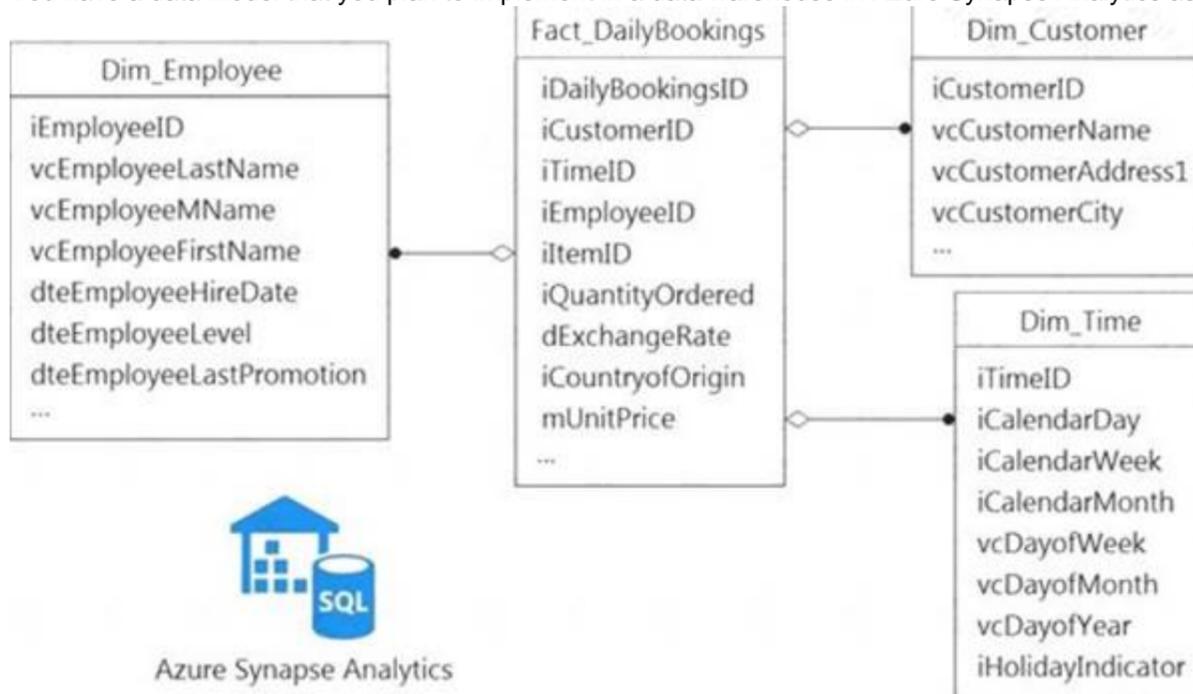
Reference:

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

NEW QUESTION 72

- (Exam Topic 3)

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

Dim_Customer: ▼
Hash distributed
Round-robin
Replicated

Dim_Employee: ▼
Hash distributed
Round-robin
Replicated

Dim_Time: ▼
Hash distributed
Round-robin
Replicated

Fact_DailyBookings: ▼
Hash distributed
Round-robin
Replicated

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:
Answer Area

Dim_Customer: ▼
Hash distributed
Round-robin
Replicated

Dim_Employee: ▼
Hash distributed
Round-robin
Replicated

Dim_Time: ▼
Hash distributed
Round-robin
Replicated

Fact_DailyBookings: ▼
Hash distributed
Round-robin
Replicated

NEW QUESTION 76
- (Exam Topic 3)

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. zone-redundant storage (ZRS)
- B. read-access geo-redundant storage (RA-GRS)
- C. locally-redundant storage (LRS)
- D. geo-redundant storage (GRS)

Answer: C

NEW QUESTION 77

- (Exam Topic 3)

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

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

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

Answer: D

NEW QUESTION 82

- (Exam Topic 3)

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 Area

```

Select TimeZone, count(*) AS MessageCount
FROM
    [MessageStream]
    GROUP BY
    TimeZone,
    [ ]
    
```

LAST
OVER
SYSTEM.TIMESTAMP()
TIMESTAMP BY

CreatedAt

HOPPINGWINDOW
SESSIONWINDOW
SLIDINGWINDOW
TUMBLINGWINDOW

(second, 15)

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area

```

Select TimeZone, count(*) AS MessageCount
FROM
    [MessageStream]
    GROUP BY
    TimeZone,
    [ ]
    
```

LAST
OVER
SYSTEM.TIMESTAMP()
TIMESTAMP BY

CreatedAt

HOPPINGWINDOW
SESSIONWINDOW
SLIDINGWINDOW
TUMBLINGWINDOW

(second, 15)

NEW QUESTION 83

- (Exam Topic 3)

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

You run the following command in an Azure Synapse Analytics Spark pool in MyWorkspace. CREATE TABLE mytestdb.myParquetTable(EmployeeID int, EmployeeName string, EmployeeStartDate date) USING Parquet
 You then use Spark to insert a row into mytestdb.myParquetTable. The row contains the following data.

EmployeeName	EmployeeID	EmployeeStartDate
Alice	24	2020-01-25

One minute later, you execute the following query from a serverless SQL pool in MyWorkspace. SELECT EmployeeID FROM mytestdb.dbo.myParquetTable WHERE name = 'Alice';
 What will be returned by the query?

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

Answer: A

Explanation:

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

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

Reference:

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

NEW QUESTION 85

- (Exam Topic 3)

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

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

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

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

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

NOTE: Each correct selection is worth one point.

Number of partitions:

▼

1

8

16

32

Partition key:

▼

Fraud indicator

Fraud score

Individual line items

Payment details

Transaction ID

- A. Mastered
- B. Not Mastered

Answer: A

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 90

- (Exam Topic 3)

You are designing an Azure Databricks interactive cluster. The cluster will be used infrequently and will be configured for auto-termination.

You need to ensure that the cluster configuration is retained indefinitely after the cluster is terminated. The solution must minimize costs.

What should you do?

- A. Clone the cluster after it is terminated.
- B. Terminate the cluster manually when processing completes.
- C. Create an Azure runbook that starts the cluster every 90 days.
- D. Pin the cluster.

Answer: D

Explanation:

To keep an interactive cluster configuration even after it has been terminated for more than 30 days, an administrator can pin a cluster to the cluster list.

References:

<https://docs.azuredatabricks.net/clusters/clusters-manage.html#automatic-termination>

NEW QUESTION 92

- (Exam Topic 3)

You use Azure Stream Analytics to receive Twitter data from Azure Event Hubs and to output the data to an Azure Blob storage account.

You need to output the count of tweets during the last five minutes every five minutes. Each tweet must only be counted once.

Which windowing function should you use?

- A. a five-minute Session window
- B. a five-minute Sliding window
- C. a five-minute Tumbling window
- D. a five-minute Hopping window that has one-minute hop

Answer: C

Explanation:

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

References:

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

NEW QUESTION 95

.....

Thank You for Trying Our Product

We offer two products:

1st - We have Practice Tests Software with Actual Exam Questions

2nd - Questions and Answers in PDF Format

DP-203 Practice Exam Features:

- * DP-203 Questions and Answers Updated Frequently
- * DP-203 Practice Questions Verified by Expert Senior Certified Staff
- * DP-203 Most Realistic Questions that Guarantee you a Pass on Your FirstTry
- * DP-203 Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year

100% Actual & Verified — Instant Download, Please Click
[Order The DP-203 Practice Test Here](#)