

## DP-203 Dumps

### Data Engineering on Microsoft Azure

<https://www.certleader.com/DP-203-dumps.html>



**NEW QUESTION 1**

- (Exam Topic 3)

You have several Azure Data Factory pipelines that contain a mix of the following types of activities.

- \* Wrangling data flow
- \* Notebook
- \* Copy
- \* jar

Which two Azure services should you use to debug the activities? Each correct answer presents part of the solution NOTE: Each correct selection is worth one point.

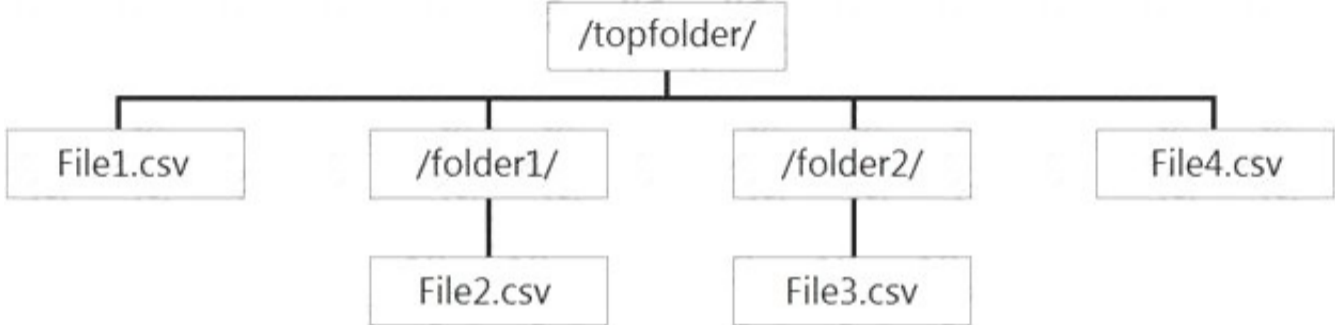
- A. Azure HDInsight
- B. Azure Databricks
- C. Azure Machine Learning
- D. Azure Data Factory
- E. Azure Synapse Analytics

**Answer:** CE

**NEW QUESTION 2**

- (Exam Topic 3)

You have files and folders in Azure Data Lake Storage Gen2 for an Azure Synapse workspace as shown in the following exhibit.



You create an external table named ExtTable that has LOCATION='/topfolder/'.

When you query ExtTable by using an Azure Synapse Analytics serverless SQL pool, which files are returned?

- A. File2.csv and File3.csv only
- B. File1.csv and File4.csv only
- C. File1.csv, File2.csv, File3.csv, and File4.csv
- D. File1.csv only

**Answer:** C

**Explanation:**

To run a T-SQL query over a set of files within a folder or set of folders while treating them as a single entity or rowset, provide a path to a folder or a pattern (using wildcards) over a set of files or folders. Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/query-data-storage#query-multiple-files-or-folders>

**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,  


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)

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 6

- (Exam Topic 3)

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

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

**Answer:** A

**Explanation:**

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

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

References:

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

**NEW QUESTION 7**

- (Exam Topic 3)

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

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

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

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

NOTE: Each correct selection is worth one point.

HubA:	<div><div></div><div>▼</div><div>Stream</div><div>Reference</div></div>
HubB:	<div><div></div><div>▼</div><div>Stream</div><div>Reference</div></div>
Database1:	<div><div></div><div>▼</div><div>Stream</div><div>Reference</div></div>

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

HubA: Stream HubB: Stream

Database1: Reference

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

**NEW QUESTION 8**

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

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

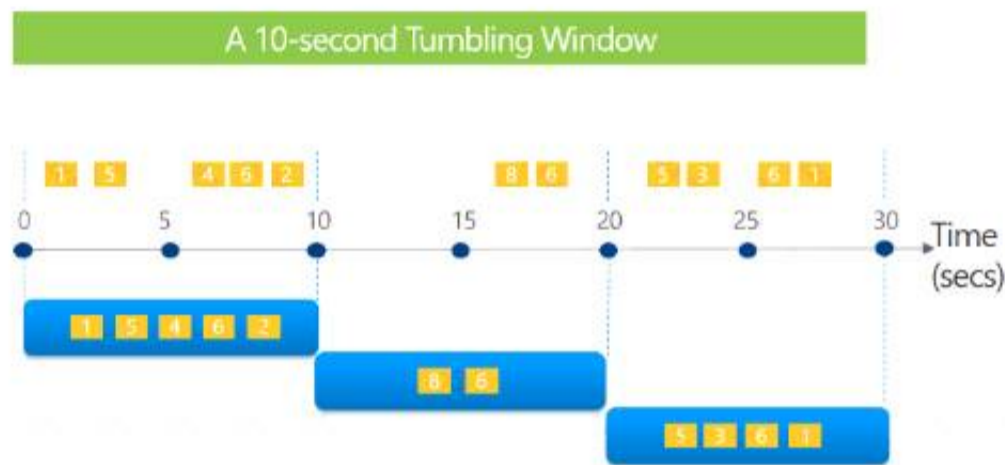
You need to count the tweets in each 10-second window. The solution must ensure that each tweet is counted only once. Does this meet the goal?

- A. Yes
- B. No

Answer: A

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

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



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

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

NEW QUESTION 9

- (Exam Topic 3)  
You are designing an Azure Synapse Analytics dedicated SQL pool.  
You need to ensure that you can audit access to Personally Identifiable information (PII). What should you include in the solution?

- A. dynamic data masking
- B. row-level security (RLS)
- C. sensitivity classifications
- D. column-level security

Answer: D

NEW QUESTION 10

- (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
```

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

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

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

#### Answer 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 10

- (Exam Topic 3)

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

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

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

Actions	Answer Area
From the Data Factory monitoring app, add the Source user property to the Activity Runs table.	
From the Data Factory monitoring app, add the Source user property to the Pipeline Runs table.	
From the Data Factory authoring UI, publish the pipelines.	⤵
From the Data Factory monitoring app, add a linked service to the Pipeline Runs table.	⤴
From the Data Factory authoring UI, generate a user property for Source on all activities.	⤶
From the Data Factory authoring UI, generate a user property for Source on all datasets.	⤵

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Step 1: From the Data Factory authoring UI, generate a user property for Source on all activities. Step 2: From the Data Factory monitoring app, add the Source user property to Activity Runs table.

You can promote any pipeline activity property as a user property so that it becomes an entity that you can monitor. For example, you can promote the Source and Destination properties of the copy activity in your pipeline as user properties. You can also select Auto Generate to generate the Source and Destination user properties for a copy activity.

Step 3: From the Data Factory authoring UI, publish the pipelines

Publish output data to data stores such as Azure SQL Data Warehouse for business intelligence (BI) applications to consume.

References:

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

NEW QUESTION 13

- (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:	<div>▼</div> <div>Always Encrypted</div> <div>TDE with customer-managed keys</div> <div>TDE with platform-managed keys</div>
To maintain client app access in the event of a datacenter outage:	<div>▼</div> <div>Create and configure Azure key vaults in two Azure regions.</div> <div>Enable Advanced Data Security on Server1.</div> <div>Implement the client apps by using a Microsoft .NET Framework data provider.</div>

- 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 16**

- (Exam Topic 3)

You have an Azure Synapse Analytics serverless SQL pool named Pool1 and an Azure Data Lake Storage Gen2 account named storage1. The AllowedBlobpublicAccess porperty is disabled for storage1.

You need to create an external data source that can be used by Azure Active Directory (Azure AD) users to access storage1 from Pool1.

What should you create first?

- A. an external resource pool
- B. a remote service binding
- C. database scoped credentials
- D. an external library

Answer: C

**NEW QUESTION 20**

- (Exam Topic 3)

You are designing an application that will store petabytes of medical imaging data

When the data is first created, the data will be accessed frequently during the first week. After one month, the data must be accessible within 30 seconds, but files will be accessed infrequently. After one year, the data will be accessed infrequently but must be accessible within five minutes.

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

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

NOTE: Each correct selection is worth one point.

First week:

Archive

Cool

Hot

After one month:

Archive

Cool

Hot

After one year:

Archive

Cool

Hot

- A. Mastered
- B. Not Mastered

Answer: A

**Explanation:**

First week: Hot

Hot - Optimized for storing data that is accessed frequently. After one month: Cool

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

After one year: Cool

**NEW QUESTION 22**

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

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

- A workload for data engineers who will use Python and SQL.
- A workload for jobs that will run notebooks that use Python, Scala, and SQL.
- A workload that data scientists will use to perform ad hoc analysis in Scala and R.

The enterprise architecture team at your company identifies the following standards for Databricks environments:

- The data engineers must share a cluster.
  - The job cluster will be managed by using a request process whereby data scientists and data engineers provide packaged notebooks for deployment to the cluster.
  - All the data scientists must be assigned their own cluster that terminates automatically after 120 minutes of inactivity. Currently, there are three data scientists.
- You need to create the Databricks clusters for the workloads.

Solution: You create a Standard cluster for each data scientist, a High Concurrency cluster for the data engineers, and a High Concurrency cluster for the jobs. Does this meet the goal?

- A. Yes
- B. No

**Answer:** A

**Explanation:**

We need a High Concurrency cluster for the data engineers and the jobs. Note:

Standard clusters are recommended for a single user. Standard can run workloads developed in any language: Python, R, Scala, and SQL.

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

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

**NEW QUESTION 27**

- (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.

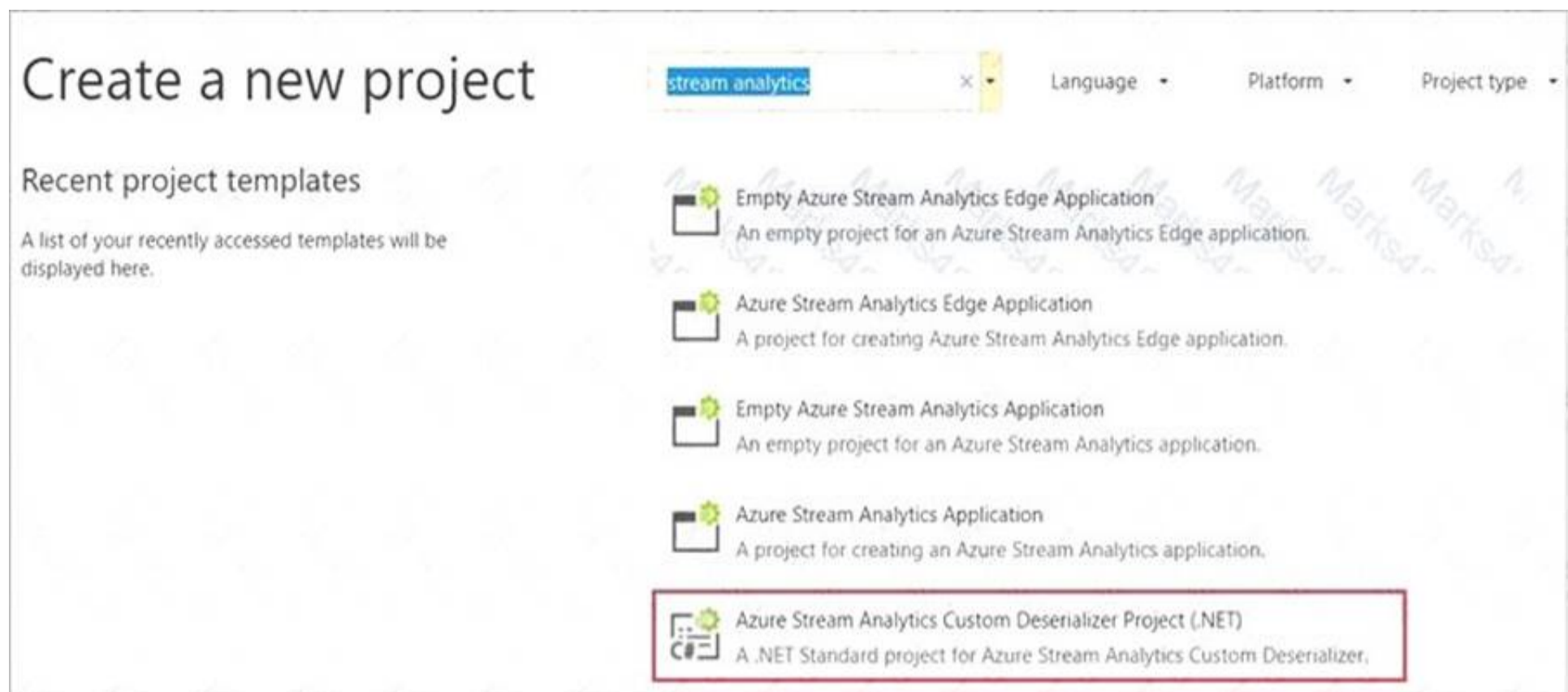
- A. Mastered
- B. Not Mastered

**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.

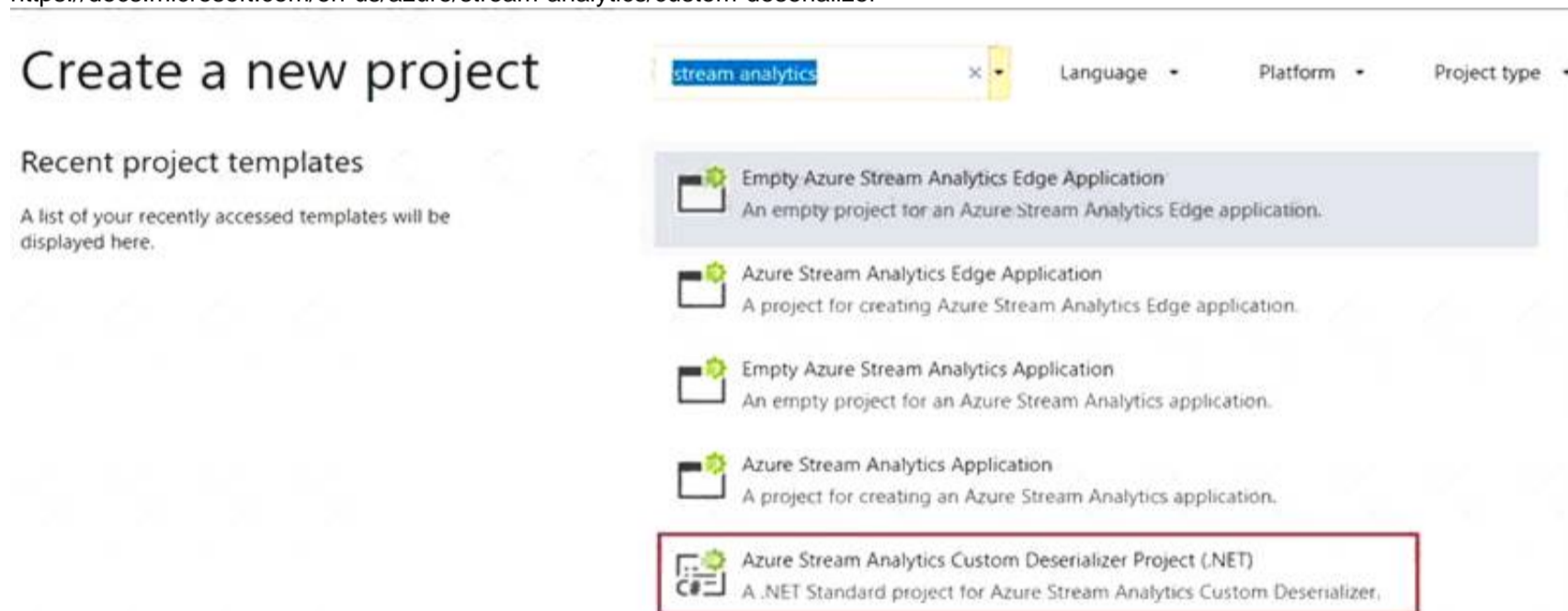
Step 3: Add an Azure Stream Analytics Application project to the solution Add an Azure Stream Analytics project

> 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 28

- (Exam Topic 3)

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

- EventDate: 1 million per day
  - EventTypeID: 10 million per event type
  - WarehouseID: 100 million per warehouse
  - ProductCategoryTypeID: 25 million per product category type
- You identify the following usage patterns:

Analyst will most commonly analyze transactions for a warehouse.

Queries will summarize by product category type, date, and/or inventory event type. You need to recommend a partition strategy for the table to minimize query times. On which column should you recommend partitioning the table?

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

**Answer: D**

#### NEW QUESTION 31

- (Exam Topic 3)

You have the following table named Employees.

first name	last name	hire date	employee type
Jane	Doe	2019-08-23	new
Ben	Smith	2017-12-15	Standard

You need to calculate the employee \_type value based on the hire date value.

How should you complete the Transact-SQL statement? To answer, drag the appropriate values to the correct targets. Each value may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content

NOTE: Each correct selection is worth one point.

**Values**

CASE
ELSE
OVER
PARTITION BY
ROW\_NUMBER

**Answer Area**

```

SELECT
    *,
    Value
    WHEN hire_date >= '2019-01-01' THEN
    'New' Value 'Standard'
    END AS employee_type
FROM
    employees;

```

- A. Mastered
- B. Not Mastered

**Answer: A**

**Explanation:**

**Values**

CASE
ELSE
OVER
PARTITION BY
ROW\_NUMBER

**Answer Area**

```

SELECT
    *,
    CASE
    WHEN hire_date >= '2019-01-01' THEN
    'New' PARTITION 'Standard'
    END AS employee_type
FROM
    employees;

```

### NEW QUESTION 33

- (Exam Topic 3)

You have two Azure Data Factory instances named ADFdev and ADFprod. ADFdev connects to an Azure DevOps Git repository.

You publish changes from the main branch of the Git repository to ADFdev. You need to deploy the artifacts from ADFdev to ADFprod.

What should you do first?

- A. From ADFdev, modify the Git configuration.
- B. From ADFdev, create a linked service.
- C. From Azure DevOps, create a release pipeline.
- D. From Azure DevOps, update the main branch.

**Answer: C**

**Explanation:**

In Azure Data Factory, continuous integration and delivery (CI/CD) means moving Data Factory pipelines from one environment (development, test, production) to another.

Note:

The following is a guide for setting up an Azure Pipelines release that automates the deployment of a data factory to multiple environments.

- > In Azure DevOps, open the project that's configured with your data factory.
  - > On the left side of the page, select Pipelines, and then select Releases.
  - > Select New pipeline, or, if you have existing pipelines, select New and then New release pipeline.
  - > In the Stage name box, enter the name of your environment.
  - > Select Add artifact, and then select the git repository configured with your development data factory.
- Select the publish branch of the repository for the Default branch. By default, this publish branch is adf\_publish.
- > Select the Empty job template. Reference:

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

### NEW QUESTION 38

- (Exam Topic 3)

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

How should you complete the Transact-SQL statement? To answer, drag the appropriate values to the correct targets. Each value may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

Values	Answer Area
CLUSTERED INDEX	CREATE TABLE table1
COLLATE	(
DISTRIBUTION	ID INTEGER,
PARTITION	col1 VARCHAR(10),
PARTITION FUNCTION	col2 VARCHAR(10)
PARTITION SCHEME	) WITH
	(
	= HASH(ID),
	(ID RANGE LEFT FOR VALUES (1, 1000000, 2000000))
	);

- A. Mastered
- B. Not Mastered

**Answer:** A

#### Explanation:

Box 1: DISTRIBUTION

Table distribution options include DISTRIBUTION = HASH ( distribution\_column\_name ), assigns each row to one distribution by hashing the value stored in distribution\_column\_name. Box 2: PARTITION

Table partition options. Syntax:

PARTITION ( partition\_column\_name RANGE [ LEFT | RIGHT ] FOR VALUES ( [ boundary\_value [,...n] ] ))

Reference:

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

### NEW QUESTION 41

- (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 be included in the load for the time at which the data should have arrived.

How should you configure the pipeline trigger? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Type: ▼

Event

On-demand

Schedule

Tumbling window

Additional properties: ▼

Prefix: /in/, Event: Blob created

Recurrence: 30 minutes, Start time: 2021-01-01T00:00

Recurrence: 30 minutes, Start time: 2021-01-01T00:00, Delay: 2 minutes

Recurrence: 32 minutes, Start time: 2021-01-15T01:45

- 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 42

- (Exam Topic 3)

You develop data engineering solutions for a company.

A project requires the deployment of data to Azure Data Lake Storage.

You need to implement role-based access control (RBAC) so that project members can manage the Azure Data Lake Storage resources.

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

- A. Assign Azure AD security groups to Azure Data Lake Storage.
- B. Configure end-user authentication for the Azure Data Lake Storage account.
- C. Configure service-to-service authentication for the Azure Data Lake Storage account.
- D. Create security groups in Azure Active Directory (Azure AD) and add project members.
- E. Configure access control lists (ACL) for the Azure Data Lake Storage account.

**Answer:** ADE

#### Explanation:

References:

<https://docs.microsoft.com/en-us/azure/data-lake-store/data-lake-store-secure-data>

#### NEW QUESTION 44

- (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 49

- (Exam Topic 3)

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 step2
BY StateID
```

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

NOTE: Each correct selection is worth one point.

Statements	Yes	No
The query joins 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"/>

- A. Mastered
- B. Not Mastered

Answer: A

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

10 partitions x six SUs = 60 SUs is fine.

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/>

NEW QUESTION 52

- (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
    [Stream]
    HOPPINGWINDOW
    SESSIONWINDOW
    SLIDINGWINDOW
    TUMBLINGWINDOW
    CreatedAt
    (second, 15)
    TimeZone,

```

### NEW QUESTION 56

- (Exam Topic 3)

You plan to ingest streaming social media data by using Azure Stream Analytics. The data will be stored in files in Azure Data Lake Storage, and then consumed by using Azure Databricks and PolyBase in Azure Synapse Analytics.

You need to recommend a Stream Analytics data output format to ensure that the queries from Databricks and PolyBase against the files encounter the fewest possible errors. The solution must ensure that the files can be queried quickly and that the data type information is retained.

What should you recommend?

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

**Answer: B**

#### Explanation:

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

### NEW QUESTION 58

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

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

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

Solution: You use a session window that uses a timeout size of 10 seconds. Does this meet the goal?

- A. Yes
- B. No

**Answer: B**

#### Explanation:

Instead use a tumbling window. Tumbling windows are a series of fixed-sized, non-overlapping and contiguous time intervals. Reference:

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

### NEW QUESTION 63

- (Exam Topic 3)

You have an Azure Synapse Analytics job that uses Scala. You need to view the status of the job.

What should you do?

- A. From Azure Monitor, run a Kusto query against the AzureDiagnostics table.
- B. From Azure Monitor, run a Kusto query against the SparkLogging1 Event.CL table.
- C. From Synapse Studio, select the workspace
- D. From Monitor, select Apache Sparks applications.
- E. From Synapse Studio, select the workspace
- F. From Monitor, select SQL requests.

**Answer: C**

### NEW QUESTION 68

- (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 73**

.....

## Thank You for Trying Our Product

\* 100% Pass or Money Back

All our products come with a 90-day Money Back Guarantee.

\* One year free update

You can enjoy free update one year. 24x7 online support.

\* Trusted by Millions

We currently serve more than 30,000,000 customers.

\* Shop Securely

All transactions are protected by VeriSign!

**100% Pass Your DP-203 Exam with Our Prep Materials Via below:**

<https://www.certleader.com/DP-203-dumps.html>