

**100%** Money Back  
**Guarantee**

**Vendor:**Microsoft

**Exam Code:**DP-203

**Exam Name:**Data Engineering on Microsoft Azure

**Version:**Demo

## QUESTION 1

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

Correct Answer: B

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>

---

## QUESTION 2

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.

Correct Answer: CE

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-unitconsumption>

---

### QUESTION 3

You have an Azure Stream Analytics job that receives clickstream data from an Azure event hub.

You need to define a query in the Stream Analytics job. The query must meet the following requirements:

Count the number of clicks within each 10-second window based on the country of a visitor. Ensure that each click is NOT counted more than once.

How should you define the Query?

- A. `SELECT Country, Avg(*) AS Average FROM ClickStream TIMESTAMP BY CreatedAt GROUP BY Country, SlidingWindow(second, 10)`
- B. `SELECT Country, Count(*) AS Count FROM ClickStream TIMESTAMP BY CreatedAt GROUP BY Country, TumblingWindow(second, 10)`
- C. `SELECT Country, Avg(*) AS Average FROM ClickStream TIMESTAMP BY CreatedAt GROUP BY Country, HoppingWindow(second, 10, 2)`
- D. `SELECT Country, Count(*) AS Count FROM ClickStream TIMESTAMP BY CreatedAt GROUP BY Country, SessionWindow(second, 5, 10)`

Correct Answer: B

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.

Example:

Incorrect Answers:

A: Sliding windows, unlike Tumbling or Hopping windows, output events only for points in time when the content of the window actually changes. In other words, when an event enters or exits the window. Every window has at least one event, like in the case of Hopping windows, events can belong to more than one sliding window.

C: Hopping window functions hop forward in time by a fixed period. It may be easy to think of them as Tumbling windows that can overlap, so events can belong to more than one Hopping window result set. To make a Hopping window the same as a Tumbling window, specify the hop size to be the same as the window size.

D: Session windows group events that arrive at similar times, filtering out periods of time where there is no data.

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

---

#### QUESTION 4

##### HOTSPOT

You plan to create a real-time monitoring app that alerts users when a device travels more than 200 meters away from a designated location.

You need to design an Azure Stream Analytics job to process the data for the planned app. The solution must minimize the amount of code developed and the number of technologies used.

What should you include in the Stream Analytics job? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

## Answer Area

Input type:  ▼

Stream
Reference

Function:  ▼

Aggregate
Geospatial
Windowing

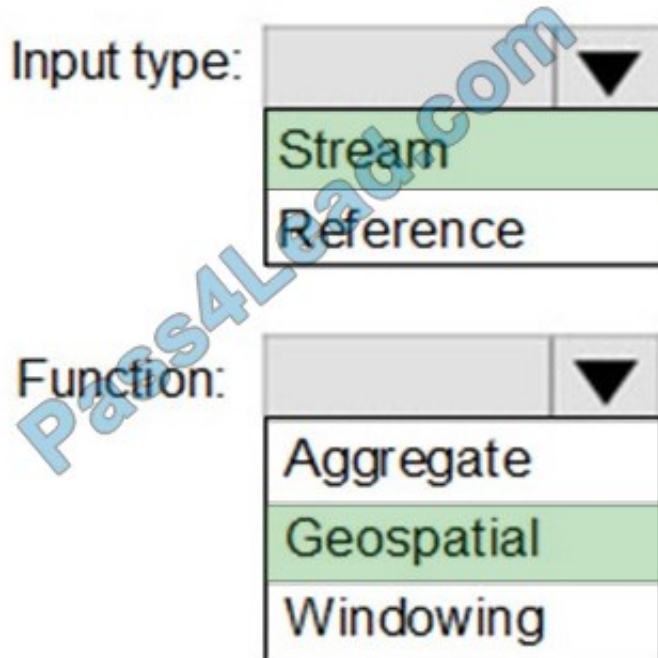
*Pass4Less.com*

Correct Answer:

# Answer Area

Input type:

Function:

The image shows a configuration interface with two dropdown menus. The first dropdown, labeled 'Input type:', has 'Stream' selected. The second dropdown, labeled 'Function:', has 'Geospatial' selected. A watermark 'Pass4it.com' is visible across the interface.

Input type: Stream

You can process real-time IoT data streams with Azure Stream Analytics.

Function: Geospatial

With built-in geospatial functions, you can use Azure Stream Analytics to build applications for scenarios such as fleet management, ride sharing, connected cars, and asset tracking.

Note: In a real-world scenario, you could have hundreds of these sensors generating events as a stream. Ideally, a gateway device would run code to push these events to Azure Event Hubs or Azure IoT Hubs.

Reference:

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

<https://docs.microsoft.com/en-us/azure/stream-analytics/geospatial-scenarios>

---

## QUESTION 5

What should you do to improve high availability of the real-time data processing solution?

- A. Deploy a High Concurrency Databricks cluster.
- B. Deploy an Azure Stream Analytics job and use an Azure Automation runbook to check the status of the job and to

start the job if it stops.

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

D. Deploy identical Azure Stream Analytics jobs to paired regions in Azure.

Correct Answer: D

Guarantee Stream Analytics job reliability during service updates Part of being a fully managed service is the capability to introduce new service functionality and improvements at a rapid pace. As a result, Stream Analytics can have a service update deploy on a weekly (or more frequent) basis. No matter how much testing is done there is still a risk that an existing, running job may break due to the introduction of a bug. If you are running mission critical jobs, these risks need to be avoided. You can reduce this risk by following Azure's paired region model.

Scenario: The application development team will create an Azure event hub to receive real-time sales data, including store number, date, time, product ID, customer loyalty number, price, and discount amount, from the point of sale (POS) system and output the data to data storage in Azure

Reference: <https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-job-reliability>

---

## QUESTION 6

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

1.

Contain sales data for 20,000 products.

2.

Use hash distribution on a column named ProductID,

3.

Contain 2.4 billion records for the years 2019 and 2020.

Which number of partition ranges provides optimal compression and performance of the clustered columnstore index?

A. 40

B. 240

C. 400

D. 2400

Correct Answer: A

Each partition should have around 1 millions records. Dedicated SQL pools already have 60 partitions.

We have the formula:  $\text{Records}/(\text{Partitions} * 60) = 1 \text{ million}$   $\text{Partitions} = \text{Records}/(1 \text{ million} * 60)$

$\text{Partitions} = 2.4 \times 1,000,000,000 / (1,000,000 * 60) = 40$  Note: Having too many partitions can reduce the effectiveness of clustered columnstore indexes if each partition has fewer than 1 million rows. Dedicated SQL pools automatically

partition

your data into 60 databases. So, if you create a table with 100 partitions, the result will be 6000 partitions.

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/best-practices-dedicated-sql-pool>

---

### QUESTION 7

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

TransactionType: 40 million rows per transaction type

CustomerSegment: 4 million per customer segment

TransactionMonth: 65 million rows per month

AccountType: 500 million per account type

You have the following query requirements:

Analysts will most commonly analyze transactions for a given month.

Transactions analysis will typically summarize transactions by transaction type, customer segment, and/or account 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. CustomerSegment
- B. AccountType
- C. TransactionType
- D. TransactionMonth

Correct Answer: D

---

### QUESTION 8

You have an Azure Data Factory pipeline that is triggered hourly.

The pipeline has had 100% success for the past seven days.

The pipeline execution fails, and two retries that occur 15 minutes apart also fail. The third failure returns the following error.

```
ErrorCode=UserErrorFileNotFound,'Type=Microsoft.DataTransfer.Common.Shared.HybridDeliveryException,Message=ADLS Gen2 operation failed for:  
Operation returned an invalid status code 'NotFound'. Account: 'contosoproduksouth'. FileSystem: wwi. Path:  
'BIKES/CARBON/year=2021/month=01/day=10/hour=06'. ErrorCode: 'PathNotFound'. Message: 'The specified path does not exist.'. RequestId: '6d269b78-  
901f-001b-4924-e7a7bc000000'. TimeStamp: 'Sun, 10 Jan 2021 07:45:05
```

What is a possible cause of the error?

- A. From 06:00 to 07:00 on January 10,2021 there was no data in w1/bikes/CARBON.
- B. The parameter used to generate year.2021/month=0/day=10/hour=06 was incorrect
- C. From 06:00 to 07:00 on January 10,2021 the file format of data wi/BiKES/CARBON was incorrect
- D. The pipeline was triggered too early.

Correct Answer: C

---

## QUESTION 9

### DRAG DROP

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

Data in the table will be retained for five years. Once a year, data that is older than five years will be deleted.

You need to ensure that the data is distributed evenly across partitions. The solution must minimize the amount of time required to delete old data.

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.

Select and Place:



**Values**

CustomerKey

HASH

ROUND\_ROBIN

REPLICATE

OrderDateKey

SalesOrderNumber

**Answer Area**

```
CREATE TABLE [dbo].[FactSales]
(
    [ProductKey] int NOT NULL
, [OrderDateKey] int NOT NULL
, [CustomerKey] int NOT NULL
, [SalesOrderNumber] nvarchar ( 20 ) NOT NULL
, [OrderQuantity] smallint NOT NULL
, [UnitPrice] money NOT NULL
)
WITH
( CLUSTERED COLUMNSTORE INDEX
, DISTRIBUTION = Value ([ProductKey])
, PARTITION ( [ Value ] RANGE RIGHT FOR VALUES
(20170101,20180101,20190101,20200101,20210101)
)
)
```

Correct Answer:

**Values**

CustomerKey

ROUND\_ROBIN

REPLICATE

SalesOrderNumber

**Answer Area**

```
CREATE TABLE [dbo].[FactSales]
(
    [ProductKey] int NOT NULL
, [OrderDateKey] int NOT NULL
, [CustomerKey] int NOT NULL
, [SalesOrderNumber] nvarchar ( 20 ) NOT NULL
, [OrderQuantity] smallint NOT NULL
, [UnitPrice] money NOT NULL
)
WITH
( CLUSTERED COLUMNSTORE INDEX
, DISTRIBUTION = HASH ([ProductKey])
, PARTITION ( [ OrderDateKey ] RANGE RIGHT FOR VALUES
(20170101,20180101,20190101,20200101,20210101)
)
)
```

Box 1: HASH

Box 2: OrderDateKey

In most cases, table partitions are created on a date column.

A way to eliminate rollbacks is to use Metadata Only operations like partition switching for data management. For example, rather than execute a DELETE statement to delete all rows in a table where the order\_date was in October of 2001,

you could partition your data early. Then you can switch out the partition with data for an empty partition from another table.

---

## QUESTION 10

### HOTSPOT

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

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

1.

No transformations must be performed.

2.

The original folder structure must be retained.

3.

Minimize time required to perform the copy activity.

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

NOTE: Each correct selection is worth one point.

Hot Area:

## Answer Area

Source dataset type:

	▼
Binary	
Parquet	
Delimited text	

Copy activity copy behavior:

	▼
FlattenHierarchy	
MergeFiles	
PreserveHierarchy	

Correct Answer:

## Answer Area

Source dataset type:

	▼
Binary	
Parquet	
Delimited text	

Copy activity copy behavior:

	▼
FlattenHierarchy	
MergeFiles	
PreserveHierarchy	

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 11

### HOTSPOT

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:

1.

Track the usage of encryption keys.

2.

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.

Hot Area:

## Answer Area

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.

Correct Answer:

## Answer Area

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.

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>

---

## QUESTION 12

### HOTSPOT

You have an Azure Storage account that generates 200,000 new files daily. The file names have a format of {YYYY}/{MM}/{DD}/{HH}/{CustomerID}.csv.

You need to design an Azure Data Factory solution that will load new data from the storage account to an Azure Data Lake once hourly. The solution must minimize load times and costs.

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

NOTE: Each correct selection is worth one point.

Hot Area:

### Answer Area

Load methodology:

	▼
Full Load	
Incremental Load	
Load individual files as they arrive	

Trigger:

	▼
Fixed schedule	
New file	
Tumbling window	

Correct Answer:

## Answer Area

Load methodology:

	▼
Full Load	
Incremental Load	
Load individual files as they arrive	

Trigger:

	▼
Fixed schedule	
New file	
Tumbling window	

Box 1: Incremental load

Box 2: Tumbling window

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)
```

To Read the [Whole Q&As](#), please purchase the [Complete Version](#) from [Our website](#).

## Try our product !

**100%** Guaranteed Success

**100%** Money Back Guarantee

**365** Days Free Update

**Instant Download** After Purchase

**24x7** Customer Support

Average **99.9%** Success Rate

More than **800,000** Satisfied Customers Worldwide

Multi-Platform capabilities - **Windows, Mac, Android, iPhone, iPod, iPad, Kindle**

## Need Help

Please provide as much detail as possible so we can best assist you.

To update a previously submitted ticket:



 <p><b>One Year Free Update</b> Free update is available within One Year after your purchase. After One Year, you will get 50% discounts for updating. And we are proud to boast a 24/7 efficient Customer Support system via Email.</p>	 <p><b>Money Back Guarantee</b> To ensure that you are spending on quality products, we provide 100% money back guarantee for 30 days from the date of purchase.</p>	 <p><b>Security &amp; Privacy</b> We respect customer privacy. We use McAfee's security service to provide you with utmost security for your personal information &amp; peace of mind.</p>
---	---	--

Any charges made through this site will appear as Global Simulators Limited.

All trademarks are the property of their respective owners.