



# DP-203<sup>Q&As</sup>

Data Engineering on Microsoft Azure

**Pass Microsoft DP-203 Exam with 100% Guarantee**

Free Download Real Questions & Answers **PDF** and **VCE** file from:

<https://www.pass4itsure.com/dp-203.html>

100% Passing Guarantee  
100% Money Back Assurance

Following Questions and Answers are all new published by Microsoft  
Official Exam Center

-  **Instant Download** After Purchase
-  **100% Money Back** Guarantee
-  **365 Days** Free Update
-  **800,000+** Satisfied Customers



**QUESTION 1****HOTSPOT**

You are implementing an Azure Stream Analytics solution to process event data from devices.

The devices output events when there is a fault and emit a repeat of the event every five seconds until the fault is resolved. The devices output a heartbeat event every five seconds after a previous event if there are no faults present.

A sample of the events is shown in the following table.

DeviceID	EventType	EventTime
78cc5ht9-w357-684r-w4fr-kr16h6p9874e	HeartBeat	2020-12-01T19:00.000Z
78cc5ht9-w357-684r-w4fr-kr16h6p9874e	HeartBeat	2020-12-01T19:05.000Z
78cc5ht9-w357-684r-w4fr-kr16h6p9874e	TemperatureSensorFault	2020-12-01T19:07.000Z

You need to calculate the uptime between the faults.

How should you complete the Stream Analytics SQL query? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:



## Answer Area

```
SELECT
```

```
DeviceID,
```

```
MIN(EventTime) as StartTime,
```

```
MAX(EventTime) as EndTime,
```

```
DATEDIFF(second, MIN(EventTime), MAX(EventTime)) AS duration_in_seconds
```

```
FROM input TIMESTAMP BY EventTime
```

```
WHERE EventType='HeartBeat'
```

```
WHERE LAG(EventType, 1) OVER (LIMIT DURATION(second,5)) <> EventType
```

```
WHERE IsFirst(second,5) = 1
```

```
GROUP BY
```

```
DeviceID
```

```
,SessionWindow(second, 5, 50000) OVER (PARTITION BY DeviceID)
```

```
,TumblingWindow(second,5)
```

```
HAVING DATEDIFF(second, MIN(EventTime), MAX(EventTime)) > 5
```

Correct Answer:



## Answer Area

```
SELECT
```

```
DeviceID,
```

```
MIN(EventTime) as StartTime,
```

```
MAX(EventTime) as EndTime,
```

```
DATEDIFF(second, MIN(EventTime), MAX(EventTime)) AS duration_in_seconds
```

```
FROM input TIMESTAMP BY EventTime
```

	▼
WHERE EventType='HeartBeat'	
WHERE LAG(EventType, 1) OVER (LIMIT DURATION(second,5)) <> EventType	
WHERE IsFirst(second,5) = 1	

```
GROUP BY
```

```
DeviceID
```

	▼
,SessionWindow(second, 5, 50000) OVER (PARTITION BY DeviceID)	
,TumblingWindow(second,5)	
HAVING DATEDIFF(second, MIN(EventTime), MAX(EventTime)) > 5	

Box 1: WHERE EventType='HeartBeat'

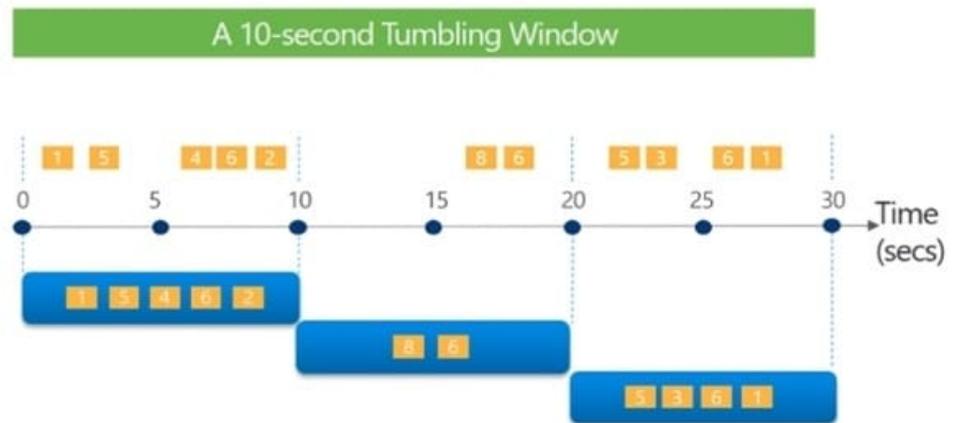
Box 2: ,TumblingWindow(Second, 5)

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)
```

## QUESTION 2

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while

others might not have a correct solution.

After you answer a question in this 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 less than 1 MB.

Does this meet the goal?

A. Yes

B. No

Correct Answer: A



Polybase loads rows that are smaller than 1 MB.

Note on Polybase Load: PolyBase is a technology that accesses external data stored in Azure Blob storage or Azure Data Lake Store via the T-SQL language.

Extract, Load, and Transform (ELT)

Extract, Load, and Transform (ELT) is a process by which data is extracted from a source system, loaded into a data warehouse, and then transformed.

The basic steps for implementing a PolyBase ELT for dedicated SQL pool are:

Extract the source data into text files.

Land the data into Azure Blob storage or Azure Data Lake Store.

Prepare the data for loading.

Load the data into dedicated SQL pool staging tables using PolyBase.

Transform the data.

Insert the data into production tables.

Reference: <https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-service-capacity-limits>

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/load-data-overview>

---

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

Correct Answer: C

Reference: <https://docs.microsoft.com/en-us/azure/event-hubs/process-data-azure-stream-analytics>

---

### QUESTION 4



You are designing an Azure Stream Analytics job to process incoming events from sensors in retail environments.

You need to process the events to produce a running average of shopper counts during the previous 15 minutes, calculated at five-minute intervals.

Which type of window should you use?

- A. snapshot
- B. tumbling
- C. hopping
- D. sliding

Correct Answer: C

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

A 10-second Tumbling Window

The diagram illustrates a 10-second tumbling window. A horizontal timeline from 0 to 30 seconds shows data points (1, 5, 4, 6, 2, 8, 6, 5, 3, 6, 1) and blue bars representing 10-second windows starting at 0, 10, and 20 seconds.

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

Unlike tumbling windows, hopping windows model scheduled overlapping windows. A hopping window specification consist of three parameters: the timeunit, the windowsize (how long each window lasts) and the hopsize (by how much each window moves forward relative to the previous one).

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

#### QUESTION 5



You are designing a fact table named FactPurchase in an Azure Synapse Analytics dedicated SQL pool. The table contains purchases from suppliers for a retail store. FactPurchase will contain the following columns.

Name	Data type	Nullable
PurchaseKey	Bigint	No
DateKey	Int	No
SupplierKey	Int	No
StockItemKey	Int	No
PurchaseOrderID	Int	Yes
OrderedQuantity	Int	No
OrderedOuters	Int	No
ReceivedOuters	Int	No
Package	Nvarchar(50)	No
IsOrderFinalized	Bit	No
LineageKey	Int	No

FactPurchase will have 1 million rows of data added daily and will contain three years of data. Transact-SQL queries similar to the following query will be executed daily.

```
SELECT
```

```
SupplierKey, StockItemKey, IsOrderFinalized, COUNT(*)FROM FactPurchase WHERE DateKey >= 20210101
```

AND DateKey 1 unique values while others may end with zero values.

2.

Does not have NULLs, or has only a few NULLs.

3.

Is not a date column. Incorrect Answers:

C: Round-robin tables are useful for improving loading speed.

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

[Latest DP-203 Dumps](#)

[DP-203 Practice Test](#)

[DP-203 Study Guide](#)