



# 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**

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.

Hot Area:

**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

Correct Answer:



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

---

## QUESTION 2

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 tumbling window, and you set the window size to 10 seconds.

Does this meet the goal?

A. Yes

B. No

Correct Answer: A

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>

### QUESTION 3

You are designing a highly available Azure Data Lake Storage solution that will include geo-zone-redundant storage (GZRS).

You need to monitor for replication delays that can affect the recovery point objective (RPO).

What should you include in the monitoring solution?

- A. 5xx: Server Error errors
- B. Average Success E2E Latency
- C. availability
- D. Last Sync Time

Correct Answer: D

Because geo-replication is asynchronous, it is possible that data written to the primary region has not yet been written to the secondary region at the time an outage occurs. The Last Sync Time property indicates the last time that data from the primary region was written successfully to the secondary region. All writes made to the primary region before the last sync time are available to be read from the secondary location. Writes made to the primary region after the last sync time property may or may not be available for reads yet.



Reference: <https://docs.microsoft.com/en-us/azure/storage/common/last-sync-time-get>

---

#### QUESTION 4

You are designing an Azure Databricks table. The table will ingest an average of 20 million streaming events per day.

You need to persist the events in the table for use in incremental load pipeline jobs in Azure Databricks. The solution must minimize storage costs and incremental load times.

What should you include in the solution?

- A. Partition by DateTime fields.
- B. Sink to Azure Queue storage.
- C. Include a watermark column.
- D. Use a JSON format for physical data storage.

Correct Answer: B

- A. Partition by DateTime field

Each partition will generate a file. Loading latency may reduce, but feel storage cost will increase because generate more folders and files for different partition. Is it right???

- B. Sink to Azure Queue Storage.

Read this document. Spark table files are stored in DBFS. Mount Azure Blob storage containers to Databricks File System (DBFS). If The Databricks ABS-AQS provides these two benefits, sounds like it is a correct answer.

<https://docs.microsoft.com/en-us/azure/databricks/spark/latest/structured-streaming/aqs>

- C. Include a watermark column: For sure it is not correct

Watermarks define how long your aggregate should wait around for data delay.

- D. User a Json format for physical data storage. - ???

Don't find any documents to compare physical data storage of JSON, CSV, and Parquet.

The Databricks ABS-AQS connector uses Azure Queue Storage (AQS) to provide an optimized file source that lets you find new files written to an Azure Blob storage (ABS) container without repeatedly listing all of the files. This provides two major advantages:

1.

Lower latency: no need to list nested directory structures on ABS, which is slow and resource intensive.

2.

Lower costs: no more costly LIST API requests made to ABS.

Reference: <https://docs.microsoft.com/en-us/azure/databricks/spark/latest/structured-streaming/aqs>

---



### QUESTION 5

You are designing a streaming data solution that will ingest variable volumes of data. You need to ensure that you can change the partition count after creation.

Which service should you use to ingest the data?

- A. Azure Event Hubs Dedicated
- B. Azure Stream Analytics
- C. Azure Data Factory
- D. Azure Synapse Analytics

Correct Answer: A

You can't change the partition count for an event hub after its creation except for the event hub in a dedicated cluster.

Reference: <https://docs.microsoft.com/en-us/azure/event-hubs/event-hubs-features>

[DP-203 PDF Dumps](#)

[DP-203 VCE Dumps](#)

[DP-203 Braindumps](#)