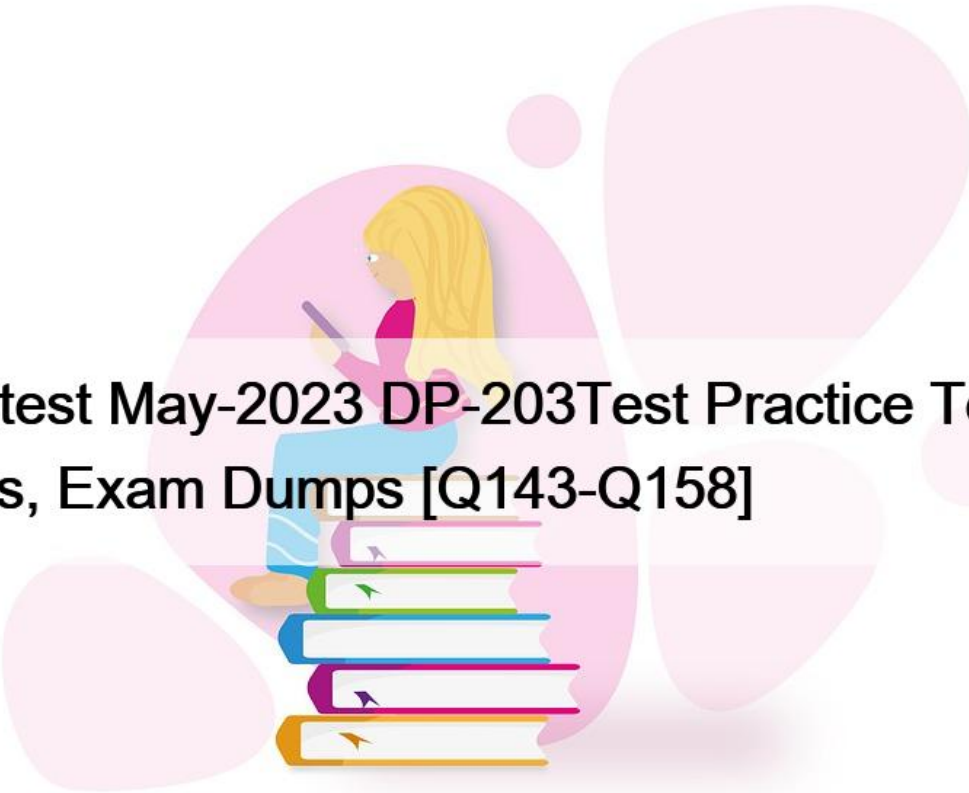


Share Latest May-2023 DP-203 Test Practice Test Questions, Exam Dumps [Q143-Q158]



Share Latest May-2023 DP-203 Test Practice Test Questions, Exam Dumps [Q143-Q158]

Share Latest May-2023 DP-203 Test Practice Test Questions, Exam Dumps Positive Aspects of Valid Dumps DP-203 Exam Dumps!

All the details of Microsoft DP-203 Exam format?

Exam Length: 40-60 question Language: English Exam Format: multiple-choice Exam Duration: 130 minutes Passing score: 75%

What is the cost of the Microsoft DP-203 Exam

The Microsoft DP-203 Exam cost is \$165 USD.

Q143. You use Azure Data Lake Storage Gen2 to store data that data scientists and data engineers will query by using Azure Databricks interactive notebooks. Users will have access only to the Data Lake Storage folders that relate to the projects on which they work.

You need to recommend which authentication methods to use for Databricks and Data Lake Storage to provide the users with the

appropriate access. The solution must minimize administrative effort and development effort.

Which authentication method should you recommend for each Azure service? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Databricks:

	▼
Azure Active Directory credential passthrough	
Azure Key Vault secrets	
Personal access tokens	

Data Lake Storage:

	▼
Azure Active Directory credential passthrough	
Shared access keys	
Shared access signatures	

Databricks:

	▼
Azure Active Directory credential passthrough	
Azure Key Vault secrets	
Personal access tokens	

Data Lake Storage:

	▼
Azure Active Directory credential passthrough	
Shared access keys	
Shared access signatures	

Reference:

<https://docs.microsoft.com/en-us/azure/databricks/data/data-sources/azure/adls-gen2/azure-datalake-gen2-sas-access>

<https://docs.microsoft.com/en-us/azure/databricks/security/credential-passthrough/adls-passthrough>

Q144. 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?

- * Yes
- * No

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>

Q145. 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

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

Explanation

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

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 start time.

ference:

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

Q146. You have a Microsoft SQL Server database that uses a third normal form schema.

You plan to migrate the data in the database to a star schema in an Azure Synapse Analytics dedicated SQI pool.

You need to design the dimension tables. The solution must optimize read operations.

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.

Transform data for the dimension tables by:

	▼
Maintaining to a third normal form	
Normalizing to a fourth normal form	
Denormalizing to a second normal form	

For the primary key columns in the dimension tables, use:

	▼
New IDENTITY columns	
A new computed column	
The business key column from the source sys	

Transform data for the dimension tables by:

	▼
Maintaining to a third normal form	
Normalizing to a fourth normal form	
Denormalizing to a second normal form	

For the primary key columns in the dimension tables, use:

	▼
New IDENTITY columns	
A new computed column	
The business key column from the source sys	

Reference:

<https://www.mssqltips.com/sqlservertip/5614/explore-the-role-of-normal-forms-in-dimensional-modeling/>

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

Q147. You have a SQL pool in Azure Synapse.

You plan to load data from Azure Blob storage to a staging table. Approximately 1 million rows of data will be loaded daily. The table will be truncated before each daily load.

You need to create the staging table. The solution must minimize how long it takes to load the data to the staging table.

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

NOTE: Each correct selection is worth one point.

Distribution: ▼
Hash
Replicated
Round-robin

Indexing: ▼
Clustered
Clustered columnstore
Heap

Partitioning: ▼
Date
None

Distribution: ▼
Hash
Replicated
Round-robin

Indexing: ▼
Clustered
Clustered columnstore
Heap

Partitioning: ▼
Date
None

Reference:

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

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

Q148. 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: ▼
Stream
Reference

HubB: ▼
Stream
Reference

Database1: ▼
Stream
Reference

HubA: ▼
 Stream
 Reference

HubB: ▼
 Stream
 Reference

Database1: ▼
 Stream
 Reference

Reference:

<https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-use-reference-data>

Q149. You are building an Azure Analytics query that will receive input data from Azure IoT Hub and write the results to Azure Blob storage.

You need to calculate the difference in readings per sensor per hour.

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 sensorId,  
       growth = reading -  
           (reading) OVER (PARTITION BY sensorId  
                           LAG (hour, 1))  
FROM input
```

```
SELECT sensorId,  
       growth = reading -  
           (reading) OVER (PARTITION BY sensorId  
                           LAST (hour, 1))  
FROM input
```

Explanation

```
SELECT sensorId,  
       growth = reading -  
           (reading) OVER (PARTITION BY sensorId  
                           LEAD (hour, 1))  
FROM input
```

Box 1: LAG

The LAG analytic operator allows one to look up a “previous” event in an event stream, within certain constraints. It is very useful for computing the rate of growth of a variable, detecting when a variable crosses a threshold, or when a condition starts or stops being true.

Box 2: LIMIT DURATION

Example: Compute the rate of growth, per sensor:

```
SELECT sensorId,  
       growth = reading -  
           LAG(reading) OVER (PARTITION BY sensorId  
                             LIMIT DURATION(hour, 1))  
FROM input
```

Reference:

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

Q150. 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 are a.

NOTE: Each correct selection is worth one point.

Select TimeZone, count (*) AS MessageCount

FROM MessageStream

	▼
LAST	
OVER	
SYSTEM.TIMESTAMP()	
TIMESTAMP BY	

CreatedAt

GROUP BY TimeZone,

	▼
HOPPINGWINDOW	
SESSIONWINDOW	
SLIDINGWINDOW	
TUMBLINGWINDOW	

(second, 15)

Select TimeZone, count (*) AS MessageCount

FROM MessageStream

	▼
LAST	
OVER	
SYSTEM.TIMESTAMP()	
TIMESTAMP BY	

CreatedAt

GROUP BY TimeZone,

	▼
HOPPINGWINDOW	
SESSIONWINDOW	
SLIDINGWINDOW	
TUMBLINGWINDOW	

(second, 15)

Reference:

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

Q151. 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 are a.

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	

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	

Reference:

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

Q152. 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?

- * data masking
- * Always Encrypted
- * column-level security
- * row-level security

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>

Q153. You need to design an analytical storage solution for the transactional data. The solution must meet the sales transaction dataset requirements.

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.

Table type to store retail store data:

	▼
Hash	
Replicated	
Round-robin	

Table type to store promotional data:

	▼
Hash	
Replicated	
Round-robin	

Table type to store retail store data:

	▼
Hash	
Replicated	
Round-robin	

Table type to store promotional data:

	▼
Hash	
Replicated	
Round-robin	

Reference:

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

Q154. 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

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

Actions

Mount the Data Lake Storage onto DBFS.

Write the results to a table in Azure Synapse.

Specify a temporary folder to stage the data.

Read the file into a data frame.

Perform transformations on the data frame.

Answer Area

Actions

Mount the Data Lake Storage onto DBFS.

Write the results to a table in Azure Synapse.

Specify a temporary folder to stage the data.

Read the file into a data frame.

Perform transformations on the data frame.

Answer Area

Mount the Data Lake Storage onto DBFS.

Read the file into a data frame.

Perform transformations on the data frame.

Specify a temporary folder to stage the data.

Reference:

<https://docs.microsoft.com/en-us/azure/azure-databricks/databricks-extract-load-sql-data-warehouse>

Q155. You need to ensure that the Twitter feed data can be analyzed in the dedicated SQL pool. The solution must meet the customer sentiment analytics requirements.

Which three Transaction-SQL DDL commands should you run in sequence? To answer, move the appropriate commands from the list of commands 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.

Commands

CREATE EXTERNAL DATA SOURCE

CREATE EXTERNAL FILE FORMAT

CREATE EXTERNAL TABLE

CREATE EXTERNAL TABLE AS SELECT

CREATE DATABASE SCOPED CREDENTIAL

Answer Area

Commands

```
CREATE EXTERNAL DATA SOURCE
```

```
CREATE EXTERNAL FILE FORMAT
```

```
CREATE EXTERNAL TABLE
```

```
CREATE EXTERNAL TABLE AS SELECT
```

```
CREATE DATABASE SCOPED CREDENTIAL
```

Answer Area

```
CREATE EXTERNAL DATA SOURCE
```

```
CREATE EXTERNAL FILE FORMAT
```

```
CREATE EXTERNAL TABLE AS SELECT
```

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/develop-tables-external-tables>

Topic 2, Litware, inc.

Overview

Litware, Inc. owns and operates 300 convenience stores across the US. The company sells a variety of packaged foods and drinks, as well as a variety of prepared foods, such as sandwiches and pizzas.

Litware has a loyalty club whereby members can get daily discounts on specific items by providing their membership number at checkout.

Litware employs business analysts who prefer to analyze data by using Microsoft Power BI, and data scientists who prefer analyzing data in Azure Databricks notebooks.

Requirements

Business Goals

Litware wants to create a new analytics environment in Azure to meet the following requirements:

See inventory levels across the stores. Data must be updated as close to real time as possible.

Execute ad hoc analytical queries on historical data to identify whether the loyalty club discounts increase sales of the discounted products.

Every four hours, notify store employees about how many prepared food items to produce based on historical demand from the sales data.

Technical Requirements

Litware identifies the following technical requirements:

Minimize the number of different Azure services needed to achieve the business goals.

Use platform as a service (PaaS) offerings whenever possible and avoid having to provision virtual machines that must be managed by Litware.

Ensure that the analytical data store is accessible only to the company's on-premises network and Azure services.

Use Azure Active Directory (Azure AD) authentication whenever possible.

Use the principle of least privilege when designing security.

Stage Inventory data in Azure Data Lake Storage Gen2 before loading the data into the analytical data store. Litware wants to remove transient data from Data Lake Storage once the data is no longer in use. Files that have a modified date that is older than 14 days must be removed.

Limit the business analysts' access to customer contact information, such as phone numbers, because this type of data is not analytically relevant.

Ensure that you can quickly restore a copy of the analytical data store within one hour in the event of corruption or accidental deletion.

Planned Environment

Litware plans to implement the following environment:

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.

Customer data, including name, contact information, and loyalty number, comes from Salesforce, a SaaS application, 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.

Daily inventory data comes from a Microsoft SQL server located on a private network.

Litware currently has 5 TB of historical sales data and 100 GB of customer data. The company expects approximately 100 GB of new data per month for the next year.

Litware will build a custom application named FoodPrep to provide store employees with the calculation results of how many prepared food items to produce every four hours.

Litware does not plan to implement Azure ExpressRoute or a VPN between the on-premises network and Azure.

Q156. 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?

- * CustomerSegment
- * AccountType
- * TransactionType
- * TransactionMonth

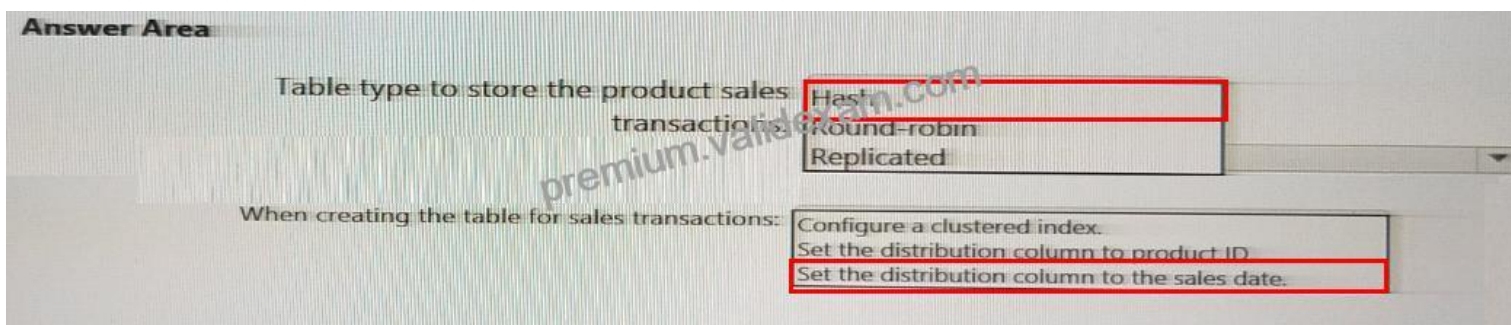
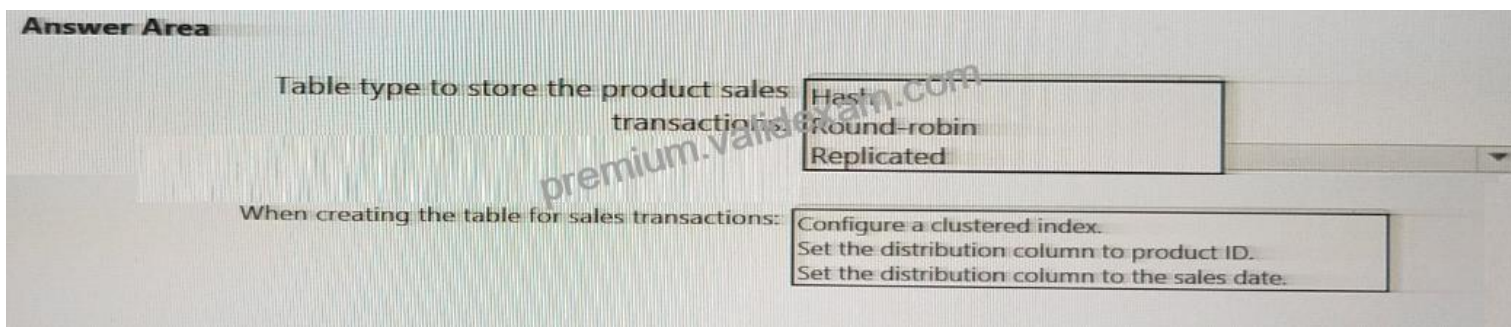
For optimal compression and performance of clustered columnstore tables, a minimum of 1 million rows per distribution and partition is needed. Before partitions are created, dedicated SQL pool already divides each table into 60 distributed databases.

Example: Any partitioning added to a table is in addition to the distributions created behind the scenes. Using this example, if the sales fact table contained 36 monthly partitions, and given that a dedicated SQL pool has 60 distributions, then the sales fact table should contain 60 million rows per month, or 2.1 billion rows when all months are populated. If a table contains fewer than the recommended minimum number of rows per partition, consider using fewer partitions in order to increase the number of rows per partition.

Q157. You need to design a data storage structure for the product sales transactions. The solution must meet the sales transaction dataset requirements.

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.



Q158. 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	

Number of partitions:

	▼
1	
8	
16	
32	

Partition key:

	▼
Fraud indicator	
Fraud score	
Individual line items	
Payment details	
Transaction ID	

Reference:

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

Practice LATEST DP-203 Exam Updated 275 Questions: <https://www.validexam.com/DP-203-latest-dumps.html>