

# Microsoft

## Exam Questions DP-203

Data Engineering on Microsoft Azure



NEW QUESTION 1

- (Exam Topic 1)

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	Answer Area
CREATE EXTERNAL DATA SOURCE	
CREATE EXTERNAL FILE FORMAT	
CREATE EXTERNAL TABLE	
CREATE EXTERNAL TABLE AS SELECT	
CREATE DATABASE SCOPED CREDENTIAL	

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Scenario: Allow Contoso users to use PolyBase in an Azure Synapse Analytics dedicated SQL pool to query the content of the data records that host the Twitter feeds. Data must be protected by using row-level security (RLS). The users must be authenticated by using their own Azure AD credentials.

Box 1: CREATE EXTERNAL DATA SOURCE

External data sources are used to connect to storage accounts. Box 2: CREATE EXTERNAL FILE FORMAT

CREATE EXTERNAL FILE FORMAT creates an external file format object that defines external data stored in Azure Blob Storage or Azure Data Lake Storage. Creating an external file format is a prerequisite for creating an external table.

Box 3: CREATE EXTERNAL TABLE AS SELECT

When used in conjunction with the CREATE TABLE AS SELECT statement, selecting from an external table imports data into a table within the SQL pool. In addition to the COPY statement, external tables are useful for loading data.

Reference:

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

NEW QUESTION 2

- (Exam Topic 1)

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.

Answer Area

Table type to store the product sales transactions:

Hash

Round-robin

Replicated

When creating the table for sales transactions:

Configure a clustered index.

Set the distribution column to product ID.

Set the distribution column to the sales date.

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area

Table type to store the product sales transactions:

Hash

Round-robin

Replicated

When creating the table for sales transactions:

Configure a clustered index.

Set the distribution column to product ID.

Set the distribution column to the sales date.

NEW QUESTION 3

- (Exam Topic 3)

You have an Azure Synapse Analytics dedicated SQL pool.

You need to ensure that data in the pool is encrypted at rest. The solution must NOT require modifying applications that query the data. What should you do?

- A. Enable encryption at rest for the Azure Data Lake Storage Gen2 account.
- B. Enable Transparent Data Encryption (TDE) for the pool.
- C. Use a customer-managed key to enable double encryption for the Azure Synapse workspace.
- D. Create an Azure key vault in the Azure subscription grant access to the pool.

**Answer:** B

**Explanation:**

Transparent Data Encryption (TDE) helps protect against the threat of malicious activity by encrypting and decrypting your data at rest. When you encrypt your database, associated backups and transaction log files are encrypted without requiring any changes to your applications. TDE encrypts the storage of an entire database by using a symmetric key called the database encryption key.

Reference:

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

**NEW QUESTION 4**

- (Exam Topic 3)

A company purchases IoT devices to monitor manufacturing machinery. The company uses an IoT appliance to communicate with the IoT devices. The company must be able to monitor the devices in real-time. You need to design the solution.

What should you recommend?

- A. Azure Stream Analytics cloud job using Azure PowerShell
- B. Azure Analysis Services using Azure Portal
- C. Azure Data Factory instance using Azure Portal
- D. Azure Analysis Services using Azure PowerShell

**Answer:** A

**Explanation:**

Stream Analytics is a cost-effective event processing engine that helps uncover real-time insights from devices, sensors, infrastructure, applications and data quickly and easily.

Monitor and manage Stream Analytics resources with Azure PowerShell cmdlets and powershell scripting that execute basic Stream Analytics tasks.

Reference:

<https://cloudblogs.microsoft.com/sqlserver/2014/10/29/microsoft-adds-iot-streaming-analytics-data-production-a>

**NEW QUESTION 5**

- (Exam Topic 3)

You have a table named SalesFact in an enterprise data warehouse in Azure Synapse Analytics. SalesFact contains sales data from the past 36 months and has the following characteristics:

- Is partitioned by month
- Contains one billion rows
- Has clustered columnstore indexes

At the beginning of each month, you need to remove data from SalesFact that is older than 36 months as quickly as possible.

Which three actions should you perform in sequence in a stored procedure? 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**

Switch the partition containing the stale data from SalesFact to SalesFact_Work.
Truncate the partition containing the stale data.
Drop the SalesFact_Work table.
Create an empty table named SalesFact_Work that has the same schema as SalesFact.
Execute a DELETE statement where the value in the Date column is more than 36 months ago.
Copy the data to a new table by using CREATE TABLE AS SELECT (CTAS).

1 2 3 4 5 6 7 8 9 10 11 12

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Step 1: Create an empty table named SalesFact\_work that has the same schema as SalesFact. Step 2: Switch the partition containing the stale data from

SalesFact to SalesFact\_Work.

SQL Data Warehouse supports partition splitting, merging, and switching. To switch partitions between two tables, you must ensure that the partitions align on their respective boundaries and that the table definitions match.

Loading data into partitions with partition switching is a convenient way stage new data in a table that is not visible to users the switch in the new data.

Step 3: Drop the SalesFact\_Work table. Reference:

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

#### NEW QUESTION 6

- (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 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 enterprise data warehouse in Azure Synapse Analytics.

You need to prepare the files to ensure that the data copies quickly. Solution: You convert the files to compressed delimited text files. Does this meet the goal?

- A. Yes
- B. No

**Answer:** A

#### Explanation:

All file formats have different performance characteristics. For the fastest load, use compressed delimited text files.

Reference:

<https://docs.microsoft.com/en-us/azure/sql-data-warehouse/guidance-for-loading-data>

#### NEW QUESTION 7

- (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 High Concurrency cluster for each data scientist, a High Concurrency cluster for the data engineers, and a Standard cluster for the jobs.

Does this meet the goal?

- A. Yes
- B. No

**Answer:** B

#### Explanation:

Need a High Concurrency cluster for the jobs.

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 8

- (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 9



- (Exam Topic 3)

You are monitoring an Azure Stream Analytics job.

The Backlogged Input Events count has been 20 for the last hour. You need to reduce the Backlogged Input Events count.

What should you do?

- A. Drop late arriving events from the job.
- B. Add an Azure Storage account to the job.
- C. Increase the streaming units for the job.
- D. Stop the job.

**Answer: C**

**Explanation:**

General symptoms of the job hitting system resource limits include:

➤ If the backlog event metric keeps increasing, it's an indicator that the system resource is constrained (either because of output sink throttling, or high CPU).

Note: Backlogged Input Events: Number of input events that are backlogged. A non-zero value for this metric implies that your job isn't able to keep up with the number of incoming events. If this value is slowly increasing or consistently non-zero, you should scale out your job: adjust Streaming Units.

Reference:

<https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-scale-jobs> <https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-monitoring>

**NEW QUESTION 10**

- (Exam Topic 3)

You use Azure Data Factory to prepare data to be queried by Azure Synapse Analytics serverless SQL pools. Files are initially ingested into an Azure Data Lake Storage Gen2 account as 10 small JSON files. Each file contains the same data attributes and data from a subsidiary of your company.

You need to move the files to a different folder and transform the data to meet the following requirements: ➤ Provide the fastest possible query times.

➤ Automatically infer the schema from the underlying files.

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

NOTE: Each correct selection is worth one point.

Copy behavior:

	▼
Flatten hierarchy	
Merge files	
Preserve hierarchy	

Sink file type:

	▼
CSV	
JSON	
Parquet	
TXT	

- A. Mastered
- B. Not Mastered

**Answer: A**

**Explanation:**

Box 1: Preserver herarchy

Compared to the flat namespace on Blob storage, the hierarchical namespace greatly improves the performance of directory management operations, which improves overall job performance.

Box 2: Parquet

Azure Data Factory parquet format is supported for Azure Data Lake Storage Gen2. Parquet supports the schema property.

Reference:

<https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-introduction> <https://docs.microsoft.com/en-us/azure/data-factory/format-parquet>

**NEW QUESTION 10**

- (Exam Topic 3)

You need to schedule an Azure Data Factory pipeline to execute when a new file arrives in an Azure Data Lake Storage Gen2 container.

Which type of trigger should you use?

- A. on-demand
- B. tumbling window
- C. schedule
- D. event

**Answer:** D

**Explanation:**

Event-driven architecture (EDA) is a common data integration pattern that involves production, detection, consumption, and reaction to events. Data integration scenarios often require Data Factory customers to trigger pipelines based on events happening in storage account, such as the arrival or deletion of a file in Azure Blob Storage account.

Reference:

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

**NEW QUESTION 13**

- (Exam Topic 3)

What should you recommend using to secure sensitive customer contact information?

- A. data labels
- B. column-level security
- C. row-level security
- D. Transparent Data Encryption (TDE)

**Answer:** B

**Explanation:**

Scenario: All cloud data must be encrypted at rest and in transit.

Always Encrypted is a feature designed to protect sensitive data stored in specific database columns from access (for example, credit card numbers, national identification numbers, or data on a need to know basis). This includes database administrators or other privileged users who are authorized to access the database to perform management tasks, but have no business need to access the particular data in the encrypted columns. The data is always encrypted, which means the encrypted data is decrypted only for processing by client applications with access to the encryption key.

References:

<https://docs.microsoft.com/en-us/azure/sql-database/sql-database-security-overview>

**NEW QUESTION 17**

- (Exam Topic 3)

You have a self-hosted integration runtime in Azure Data Factory.

The current status of the integration runtime has the following configurations:

- > Status: Running
- > Type: Self-Hosted
- > Version: 4.4.7292.1
- > Running / Registered Node(s): 1/1
- > High Availability Enabled: False
- > Linked Count: 0
- > Queue Length: 0
- > Average Queue Duration: 0.00s

The integration runtime has the following node details:

- > Name: X-M
- > Status: Running
- > Version: 4.4.7292.1
- > Available Memory: 7697MB
- > CPU Utilization: 6%
- > Network (In/Out): 1.21KBps/0.83KBps
- > Concurrent Jobs (Running/Limit): 2/14
- > Role: Dispatcher/Worker
- > Credential Status: In Sync

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

NOTE: Each correct selection is worth one point.

If the X-M node becomes unavailable, all  
executed pipelines will:

	▼
fail until the node comes back online	
switch to another integration runtime	
exceed the CPU limit	

The number of concurrent jobs and the  
CPU usage indicate that the Concurrent  
Jobs (Running/Limit) value should be:

	▼
raised	
lowered	
left as is	

- A. Mastered  
B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: fail until the node comes back online We see: High Availability Enabled: False

Note: Higher availability of the self-hosted integration runtime so that it's no longer the single point of failure in your big data solution or cloud data integration with Data Factory.

Box 2: lowered We see:

Concurrent Jobs (Running/Limit): 2/14 CPU Utilization: 6%

Note: When the processor and available RAM aren't well utilized, but the execution of concurrent jobs reaches a node's limits, scale up by increasing the number of concurrent jobs that a node can run

Reference:

<https://docs.microsoft.com/en-us/azure/data-factory/create-self-hosted-integration-runtime>

**NEW QUESTION 21**

- (Exam Topic 3)

You are developing a solution using a Lambda architecture on Microsoft Azure. The data at test layer must meet the following requirements:

Data storage:

- Serve as a repository (or high volumes of large files in various formats.
- Implement optimized storage for big data analytics workloads.
- Ensure that data can be organized using a hierarchical structure. Batch processing:
- Use a managed solution for in-memory computation processing.
- Natively support Scala, Python, and R programming languages.
- Provide the ability to resize and terminate the cluster automatically. Analytical data store:
- Support parallel processing.
- Use columnar storage.
- Support SQL-based languages.

You need to identify the correct technologies to build the Lambda architecture.

Which technologies should you use? To answer, select the appropriate options in the answer area NOTE: Each correct selection is worth one point.

Architecture requirement	Technology
Data storage	<div><div></div><div><div>Azure SQL Database</div><div>Azure Blob Storage</div><div>Azure Cosmos DB</div><div>Azure Data Lake Store</div></div></div>
Batch processing	<div><div></div><div><div>HDInsight Spark</div><div>HDInsight Hadoop</div><div>Azure Databricks</div><div>HDInsight Interactive Query</div></div></div>
Analytical data store	<div><div></div><div><div>HDInsight HBase</div><div>Azure SQL Data Warehouse</div><div>Azure Analysis Services</div><div>Azure Cosmos DB</div></div></div>

- A. Mastered  
B. Not Mastered

Answer: A

**Explanation:**

Data storage: Azure Data Lake Store

A key mechanism that allows Azure Data Lake Storage Gen2 to provide file system performance at object storage scale and prices is the addition of a hierarchical namespace. This allows the collection of objects/files within an account to be organized into a hierarchy of directories and nested subdirectories in the same way that the file system on your computer is organized. With the hierarchical namespace enabled, a storage account becomes capable of providing the scalability and cost-effectiveness of object storage, with file system semantics that are familiar to analytics engines and frameworks.

Batch processing: HD Insight Spark

Apache Spark is an open-source, parallel-processing framework that supports in-memory processing to boost the performance of big-data analysis applications. HDInsight is a managed Hadoop service. Use it to deploy and manage Hadoop clusters in Azure. For batch processing, you can use Spark, Hive, Hive LLAP, MapReduce.

Languages: R, Python, Java, Scala, SQL Analytic data store: SQL Data Warehouse

SQL Data Warehouse is a cloud-based Enterprise Data Warehouse (EDW) that uses Massively Parallel Processing (MPP).

SQL Data Warehouse stores data into relational tables with columnar storage. References:

<https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-namespaces> <https://docs.microsoft.com/en-us/azure/architecture/data-guide/technology-choices/batch-processing> <https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-overview-what-is>

**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 Standard cluster for the jobs.

Does this meet the goal?

- A. Yes  
B. No

Answer: B

**Explanation:**

We would need a High Concurrency cluster for 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 25

- (Exam Topic 3)

You plan to implement an Azure Data Lake Gen2 storage account.

You need to ensure that the data lake will remain available if a data center fails in the primary Azure region. The solution must minimize costs.

Which type of replication should you use for the storage account?

- A. geo-redundant storage (GRS)
- B. zone-redundant storage (ZRS)
- C. locally-redundant storage (LRS)
- D. geo-zone-redundant storage (GZRS)

**Answer: A**

#### Explanation:

Geo-redundant storage (GRS) copies your data synchronously three times within a single physical location in the primary region using LRS. It then copies your data asynchronously to a single physical location in the secondary region.

Reference:

<https://docs.microsoft.com/en-us/azure/storage/common/storage-redundancy>

#### NEW QUESTION 26

- (Exam Topic 3)

You are developing a solution that will stream to Azure Stream Analytics. The solution will have both streaming data and reference data.

Which input type should you use for the reference data?

- A. Azure Cosmos DB
- B. Azure Blob storage
- C. Azure IoT Hub
- D. Azure Event Hubs

**Answer: B**

#### Explanation:

Stream Analytics supports Azure Blob storage and Azure SQL Database as the storage layer for Reference Data.

Reference:

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

#### NEW QUESTION 29

- (Exam Topic 3)

You store files in an Azure Data Lake Storage Gen2 container. The container has the storage policy shown in the following exhibit.



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

NOTE: Each correct selection is worth one point.

Answer Area

The files are [answer choice] after 30 days.

deleted from the container

moved to archive storage

moved to cool storage

moved to hot storage

The storage policy applies to [answer choice].

container1/contoso1.csv

container1/docs/contoso.json

container1/mycontoso/contoso.csv

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area

The files are [answer choice] after 30 days.

deleted from the container

moved to archive storage

moved to cool storage

moved to hot storage

The storage policy applies to [answer choice].

container1/contoso1.csv

container1/docs/contoso.json

container1/mycontoso/contoso.csv

NEW QUESTION 33

- (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

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.

Answer Area

>

<

↑

↓

- 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 34

- (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 SOL.
- 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 39

- (Exam Topic 3)

You have an Azure Synapse Analytics dedicated SQL pool that contains a large fact table. The table contains 50 columns and 5 billion rows and is a heap.

Most queries against the table aggregate values from approximately 100 million rows and return only two columns.

You discover that the queries against the fact table are very slow. Which type of index should you add to provide the fastest query times?

- A. nonclustered columnstore
- B. clustered columnstore
- C. nonclustered
- D. clustered

**Answer: B**

#### Explanation:

Clustered columnstore indexes are one of the most efficient ways you can store your data in dedicated SQL pool.

Columnstore tables won't benefit a query unless the table has more than 60 million rows. Reference:

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

#### NEW QUESTION 44

- (Exam Topic 3)

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?

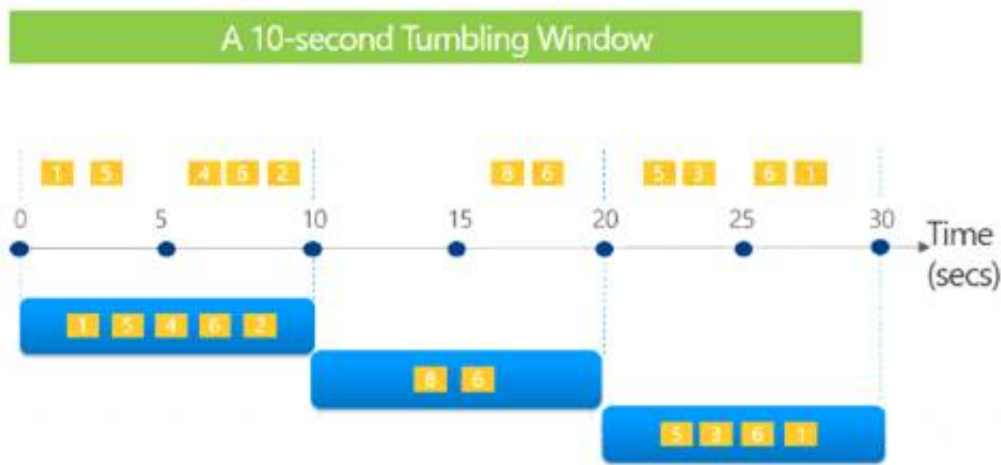
- A. snapshot
- B. tumbling
- C. hopping
- D. sliding

**Answer: B**

#### 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 48

- (Exam Topic 3)

You are planning the deployment of Azure Data Lake Storage Gen2. You have the following two reports that will access the data lake:

- > Report1: Reads three columns from a file that contains 50 columns.
- > Report2: Queries a single record based on a timestamp.

You need to recommend in which format to store the data in the data lake to support the reports. The solution must minimize read times.

What should you recommend for each report? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Report1:

▼

Avro

CSV

Parquet

TSV

Report2:

▼

Avro

CSV

Parquet

TSV

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:  
Report1: CSV  
CSV: The destination writes records as delimited data. Report2: AVRO  
AVRO supports timestamps.  
Not Parquet, TSV: Not options for Azure Data Lake Storage Gen2. Reference:  
<https://streamsets.com/documentation/datacollector/latest/help/datacollector/UserGuide/Destinations/ADLS-G2>

NEW QUESTION 51

- (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 56

- (Exam Topic 3)

You are creating dimensions for a data warehouse in an Azure Synapse Analytics dedicated SQL pool. You create a table by using the Transact-SQL statement shown in the following exhibit.

```
CREATE TABLE [DBO].[DimProduct] (  
    [ProductKey] [int] IDENTITY(1,1) NOT NULL,  
    [ProductSourceID] [int] NOT NULL,  
    [ProductName] [nvarchar](100) NOT NULL,  
    [ProductNumber] [nvarchar](25) NOT NULL,  
    [Color] [nvarchar](15) NULL,  
    [Size] [nvarchar](5) NULL,  
    [Weight] [decimal](8, 2) NULL,  
    [ProductCategory] [nvarchar](100) NULL,  
    [SellStartDate] [date] NOT NULL,  
    [SellEndDate] [date] NULL,  
    [RowInsertedDateTime] [datetime] NOT NULL,  
    [RowUpdatedDateTime] [datetime] NOT NULL,  
    [ETLAuditID] [int] NOT NULL  
)
```

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

NOTE: Each correct selection is worth one point.

DimProduct is a **[answer choice]** slowly changing dimension (SCD).

	▼
Type 0	
Type 1	
Type 2	

The ProductKey column is **[answer choice]**.

	▼
a surrogate key	
a business key	
an audit column	

- A. Mastered
- B. Not Mastered

**Answer: A**

#### Explanation:

Box 1: Type 2

A Type 2 SCD supports versioning of dimension members. Often the source system doesn't store versions, so the data warehouse load process detects and manages changes in a dimension table. In this case, the dimension table must use a surrogate key to provide a unique reference to a version of the dimension member. It also includes columns that define the date range validity of the version (for example, StartDate and EndDate) and possibly a flag column (for example, IsCurrent) to easily filter by current dimension members.

Reference:

<https://docs.microsoft.com/en-us/learn/modules/populate-slowly-changing-dimensions-azure-synapse-analytics>

#### NEW QUESTION 61

- (Exam Topic 3)

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, COUNT(*)
FROM FactPurchase
WHERE DateKey >= 20210101
AND DateKey <= 20210131
GROUP BY SupplierKey, StockItemKey
```

Which table distribution will minimize query times?

- A. round-robin
- B. replicated
- C. hash-distributed on DateKey
- D. hash-distributed on PurchaseKey

**Answer:** D

**Explanation:**

Hash-distributed tables improve query performance on large fact tables, and are the focus of this article. 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-distribu>

**NEW QUESTION 64**

- (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 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

**Answer:** A

**Explanation:**

When exporting data into an ORC File Format, you might get Java out-of-memory errors when there are large text columns. To work around this limitation, export only a subset of the columns.

References:

<https://docs.microsoft.com/en-us/azure/sql-data-warehouse/guidance-for-loading-data>

**NEW QUESTION 65**

- (Exam Topic 3)

You have a SQL pool in Azure Synapse.

A user reports that queries against the pool take longer than expected to complete. You need to add monitoring to the underlying storage to help diagnose the issue.

Which two metrics should you monitor? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

- A. Cache used percentage
- B. DWU Limit

- C. Snapshot Storage Size
- D. Active queries
- E. Cache hit percentage

Answer: AE

Explanation:

A: Cache used is the sum of all bytes in the local SSD cache across all nodes and cache capacity is the sum of the storage capacity of the local SSD cache across all nodes.

E: Cache hits is the sum of all columnstore segments hits in the local SSD cache and cache miss is the columnstore segments misses in the local SSD cache summed across all nodes

Reference:

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

NEW QUESTION 69

- (Exam Topic 3)

You need to output files from Azure Data Factory.

Which file format should you use for each type of output? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Columnar format:

▼

Avro

GZip

Parquet

TXT

JSON with a timestamp:

▼

Avro

GZip

Parquet

TXT

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Parquet

Parquet stores data in columns, while Avro stores data in a row-based format. By their very nature, column-oriented data stores are optimized for read-heavy analytical workloads, while row-based databases are best for write-heavy transactional workloads.

Box 2: Avro

An Avro schema is created using JSON format.

AVRO supports timestamps.

Note: Azure Data Factory supports the following file formats (not GZip or TXT).

- > Avro format
- > Binary format
- > Delimited text format
- > Excel format
- > JSON format
- > ORC format
- > Parquet format
- > XML format

Reference:

<https://www.datanami.com/2018/05/16/big-data-file-formats-demystified>

NEW QUESTION 71

- (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

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

ROW\_NUMBER

Answer Area

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

NEW QUESTION 76

- (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 81**

- (Exam Topic 3)

You use Azure Data Lake Storage Gen2.

You need to ensure that workloads can use filter predicates and column projections to filter data at the time the data is read from disk.

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

- A. Reregister the Microsoft Data Lake Store resource provider.
- B. Reregister the Azure Storage resource provider.
- C. Create a storage policy that is scoped to a container.
- D. Register the query acceleration feature.
- E. Create a storage policy that is scoped to a container prefix filter.

**Answer:** BD

**NEW QUESTION 84**

- (Exam Topic 3)

You are designing a sales transactions table in an Azure Synapse Analytics dedicated SQL pool. The table will contains approximately 60 million rows per month and will be partitioned by month. The table will use a clustered column store index and round-robin distribution.

Approximately how many rows will there be for each combination of distribution and partition?

- A. 1 million
- B. 5 million
- C. 20 million
- D. 60 million

**Answer:** D

**Explanation:**

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

**NEW QUESTION 86**

- (Exam Topic 3)

You have an Azure Active Directory (Azure AD) tenant that contains a security group named Group1. You have an Azure Synapse Analytics dedicated SQL pool named dw1 that contains a schema named schema1.

You need to grant Group1 read-only permissions to all the tables and views in schema1. The solution must use the principle of least privilege.

Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions 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.

**Actions**

**Answer Area**

- Create a database role named Role1 and grant Role1 SELECT permissions to schema1.
- Create a database role named Role1 and grant Role1 SELECT permissions to dw1.
- Assign the Azure role-based access control (Azure RBAC) Reader role for dw1 to Group1.
- Create a database user in dw1 that represents Group1 and uses the FROM EXTERNAL PROVIDER clause.
- Assign Role1 to the Group1 database user.

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Step 1: Create a database role named Role1 and grant Role1 SELECT permissions to schema You need to grant Group1 read-only permissions to all the tables and views in schema1.

Place one or more database users into a database role and then assign permissions to the database role. Step 2: Assign Rol1 to the Group database user

Step 3: Assign the Azure role-based access control (Azure RBAC) Reader role for dw1 to Group1 Reference:

<https://docs.microsoft.com/en-us/azure/data-share/how-to-share-from-sql>

**NEW QUESTION 88**

- (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

CLUSTERED INDEX
COLLATE
DISTRIBUTION
PARTITION
PARTITION FUNCTION
PARTITION SCHEME

### Answer Area

```
CREATE TABLE table1
(
    ID INTEGER,
    col1 VARCHAR(10),
    col2 VARCHAR(10)
) 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 89

- (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 93

- (Exam Topic 3)

You have an Azure Storage account and a data warehouse in Azure Synapse Analytics in the UK South region. You need to copy blob data from the storage

account to the data warehouse by using Azure Data Factory. The solution must meet the following requirements:

- Ensure that the data remains in the UK South region at all times.
- Minimize administrative effort.

Which type of integration runtime should you use?

- A. Azure integration runtime
- B. Azure-SSIS integration runtime
- C. Self-hosted integration runtime

Answer: A

Explanation:

IR type	Public network	Private network
Azure	Data Flow Data movement Activity dispatch	
Self-hosted	Data movement Activity dispatch	Data movement Activity dispatch
Azure-SSIS	SSIS package execution	SSIS package execution

Reference:  
https://docs.microsoft.com/en-us/azure/data-factory/concepts-integration-runtime

NEW QUESTION 97

- (Exam Topic 3)  
You have an Azure Synapse Analytics dedicated SQL pool that contains a table named Table1. You have files that are ingested and loaded into an Azure Data Lake Storage Gen2 container named container1.  
You plan to insert data from the files into Table1 and azure Data Lake Storage Gen2 container named container1.  
You plan to insert data from the files into Table1 and transform the data. Each row of data in the files will produce one row in the serving layer of Table1.  
You need to ensure that when the source data files are loaded to container1, the DateTime is stored as an additional column in Table1.  
Solution: In an Azure Synapse Analytics pipeline, you use a data flow that contains a Derived Column transformation.

- A. Yes
- B. No

Answer: B

NEW QUESTION 101

- (Exam Topic 3)  
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 -

LAG

LAST

LEAD

(reading) OVER (PARTITION BY sensorId

LIMIT DURATION

OFFSET

WHEN

(hour,1))

FROM input

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

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

### NEW QUESTION 105

- (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 107

- (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 112

- (Exam Topic 3)

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:

Partition key:

- A. Mastered
- B. Not Mastered

Answer: A

#### Explanation:

Box 1: 16

For Event Hubs you need to set the partition key explicitly.

An embarrassingly parallel job is the most scalable scenario in Azure Stream Analytics. It connects one partition of the input to one instance of the query to one partition of the output.

Box 2: Transaction ID Reference:

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

#### NEW QUESTION 116

- (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 118

- (Exam Topic 3)

You have a partitioned table in an Azure Synapse Analytics dedicated SQL pool.

You need to design queries to maximize the benefits of partition elimination. What should you include in the Transact-SQL queries?

- A. JOIN
- B. WHERE
- C. DISTINCT
- D. GROUP BY

Answer: B

NEW QUESTION 119

- (Exam Topic 3)

You have an Azure Data Lake Storage Gen2 container.

Data is ingested into the container, and then transformed by a data integration application. The data is NOT modified after that. Users can read files in the container but cannot modify the files.

You need to design a data archiving solution that meets the following requirements: ➤ New data is accessed frequently and must be available as quickly as possible.

➤ Data that is older than five years is accessed infrequently but must be available within one second when requested.

➤ Data that is older than seven years is NOT accessed. After seven years, the data must be persisted at the lowest cost possible.

➤ Costs must be minimized while maintaining the required availability.

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

Five-year-old data:

▼

Delete the blob.

Move to archive storage.

Move to cool storage.

Move to hot storage.

Seven-year-old data:

▼

Delete the blob.

Move to archive storage.

Move to cool storage.

Move to hot storage.

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

:

Box 1: Replicated

Replicated tables are ideal for small star-schema dimension tables, because the fact table is often distributed on a column that is not compatible with the connected dimension tables. If this case applies to your schema, consider changing small dimension tables currently implemented as round-robin to replicated.

Box 2: Replicated

Box 3: Replicated

Box 4: Hash-distributed

For Fact tables use hash-distribution with clustered columnstore index. Performance improves when two hash tables are joined on the same distribution column.

Reference:

<https://azure.microsoft.com/en-us/updates/reduce-data-movement-and-make-your-queries-more-efficient-with-th> <https://azure.microsoft.com/en-us/blog/replicated-tables-now-generally-available-in-azure-sql-data-warehouse/>

NEW QUESTION 124

.....

## Thank You for Trying Our Product

### We offer two products:

1st - We have Practice Tests Software with Actual Exam Questions

2nd - Questions and Answers in PDF Format

### DP-203 Practice Exam Features:

- \* DP-203 Questions and Answers Updated Frequently
- \* DP-203 Practice Questions Verified by Expert Senior Certified Staff
- \* DP-203 Most Realistic Questions that Guarantee you a Pass on Your FirstTry
- \* DP-203 Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year

**100% Actual & Verified — Instant Download, Please Click**  
**[Order The DP-203 Practice Test Here](#)**