

Exam Questions DP-203

Data Engineering on Microsoft Azure

<https://www.2passeasy.com/dumps/DP-203/>



NEW QUESTION 1

- (Exam Topic 1)

You need to design the partitions for the product sales transactions. The solution must mee 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:

Partition product sales transactions data by:

Sales date
 Product ID
 Promotion ID

Store product sales transactions data in:

An Azure Synapse Analytics dedicated SQL pool
 An Azure Synapse Analytics serverless SQL pool
 An Azure Data Lake Storage Gen2 account linked to an Azure Synapse Analytics workspace

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Sales date

Scenario: Contoso requirements for data integration include:

➤ Partition data that contains sales transaction records. Partitions must be designed to provide efficient loads by month. Boundary values must belong to the partition on the right.

Box 2: An Azure Synapse Analytics Dedicated SQL pool Scenario: Contoso requirements for data integration include:

➤ Ensure that data storage costs and performance are predictable.

The size of a dedicated SQL pool (formerly SQL DW) is determined by Data Warehousing Units (DWU). Dedicated SQL pool (formerly SQL DW) stores data in relational tables with columnar storage. This format significantly reduces the data storage costs, and improves query performance.

Synapse analytics dedicated sql pool Reference:

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

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 1)

You need to implement the surrogate key for the retail store table. The solution must meet the sales transaction dataset requirements.

What should you create?

- A. a table that has an IDENTITY property

- B. a system-versioned temporal table
- C. a user-defined SEQUENCE object
- D. a table that has a FOREIGN KEY constraint

Answer: A

Explanation:

Scenario: Implement a surrogate key to account for changes to the retail store addresses.

A surrogate key on a table is a column with a unique identifier for each row. The key is not generated from the table data. Data modelers like to create surrogate keys on their tables when they design data warehouse models. You can use the IDENTITY property to achieve this goal simply and effectively without affecting load performance.

Reference:

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

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 an on-premises data warehouse that includes the following fact tables. Both tables have the following columns: DateKey, ProductKey, RegionKey. There are 120 unique product keys and 65 unique region keys.

Table	Comments
Sales	The table is 600 GB in size. DateKey is used extensively in the WHERE clause in queries. ProductKey is used extensively in join operations. RegionKey is used for grouping. Severity-five percent of records relate to one of 40 regions.
Invoice	The table is 6 GB in size. DateKey and ProductKey are used extensively in the WHERE clause in queries. RegionKey is used for grouping.

Queries that use the data warehouse take a long time to complete.

You plan to migrate the solution to use Azure Synapse Analytics. You need to ensure that the Azure-based solution optimizes query performance and minimizes processing skew.

What should you recommend? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point

Table	Distribution type	Distribution column
Sales:	<div style="border: 1px solid black; padding: 5px;"> <div style="text-align: right; margin-bottom: 5px;">▼</div> <div style="padding: 5px;">Hash-distributed</div> <hr style="border: 0; border-top: 1px solid black; margin: 5px 0;"/> <div style="padding: 5px;">Round-robin</div> </div>	<div style="border: 1px solid black; padding: 5px;"> <div style="text-align: right; margin-bottom: 5px;">▼</div> <div style="padding: 5px;">DateKey</div> <hr style="border: 0; border-top: 1px solid black; margin: 5px 0;"/> <div style="padding: 5px;">ProductKey</div> <hr style="border: 0; border-top: 1px solid black; margin: 5px 0;"/> <div style="padding: 5px;">RegionKey</div> </div>
Invoices:	<div style="border: 1px solid black; padding: 5px;"> <div style="text-align: right; margin-bottom: 5px;">▼</div> <div style="padding: 5px;">Hash-distributed</div> <hr style="border: 0; border-top: 1px solid black; margin: 5px 0;"/> <div style="padding: 5px;">Round-robin</div> </div>	<div style="border: 1px solid black; padding: 5px;"> <div style="text-align: right; margin-bottom: 5px;">▼</div> <div style="padding: 5px;">DateKey</div> <hr style="border: 0; border-top: 1px solid black; margin: 5px 0;"/> <div style="padding: 5px;">ProductKey</div> <hr style="border: 0; border-top: 1px solid black; margin: 5px 0;"/> <div style="padding: 5px;">RegionKey</div> </div>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Hash-distributed
 Box 2: ProductKey
 ProductKey is used extensively in joins.
 Hash-distributed tables improve query performance on large fact tables.
 Box 3: Round-robin
 Box 4: RegionKey
 Round-robin tables are useful for improving loading speed.
 Consider using the round-robin distribution for your table in the following scenarios:

- > When getting started as a simple starting point since it is the default
- > If there is no obvious joining key
- > If there is not good candidate column for hash distributing the table
- > If the table does not share a common join key with other tables
- > If the join is less significant than other joins in the query
- > When the table is a temporary staging table

Note: A distributed table appears as a single table, but the rows are actually stored across 60 distributions. The rows are distributed with a hash or round-robin algorithm.

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

NEW QUESTION 6

- (Exam Topic 3)

You need to create an Azure Data Factory pipeline to process data for the following three departments at your company: Ecommerce, retail, and wholesale. The solution must ensure that data can also be processed for the entire company.

How should you complete the Data Factory data flow script? 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

- all, ecommerce, retail, wholesale
- dept=='ecommerce', dept=='retail', dept=='wholesale'
- dept=='ecommerce', dept=='wholesale', dept=='retail'
- disjoint: false
- disjoint: true
- ecommerce, retail, wholesale, all

Answer Area

```
CleanData
split(
    [ ]
    [ ]
    [ ]
) ~> SplitByDept@([ ])
```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

The conditional split transformation routes data rows to different streams based on matching conditions. The conditional split transformation is similar to a CASE decision structure in a programming language. The transformation evaluates expressions, and based on the results, directs the data row to the specified stream.

Box 1: dept=='ecommerce', dept=='retail', dept=='wholesale'

First we put the condition. The order must match the stream labeling we define in Box 3. Syntax:

```
<incomingStream> split(
<conditionalExpression1>
<conditionalExpression2>
disjoint: {true | false}
) ~> <splitTx>@(stream1, stream2, ..., <defaultStream>)
```

Box 2: disjoint : false

disjoint is false because the data goes to the first matching condition. All remaining rows matching the third condition go to output stream all.

Box 3: ecommerce, retail, wholesale, all Label the streams

Reference:
<https://docs.microsoft.com/en-us/azure/data-factory/data-flow-conditional-split>

NEW QUESTION 7

- (Exam Topic 3)

You have an enterprise-wide Azure Data Lake Storage Gen2 account. The data lake is accessible only through an Azure virtual network named VNET1.

You are building a SQL pool in Azure Synapse that will use data from the data lake.

Your company has a sales team. All the members of the sales team are in an Azure Active Directory group named Sales. POSIX controls are used to assign the Sales group access to the files in the data lake.

You plan to load data to the SQL pool every hour.

You need to ensure that the SQL pool can load the sales data from the data lake.

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

- A. Add the managed identity to the Sales group.
- B. Use the managed identity as the credentials for the data load process.
- C. Create a shared access signature (SAS).
- D. Add your Azure Active Directory (Azure AD) account to the Sales group.
- E. Use the snared access signature (SAS) as the credentials for the data load process.
- F. Create a managed identity.

Answer: ADF

Explanation:

The managed identity grants permissions to the dedicated SQL pools in the workspace.

Note: Managed identity for Azure resources is a feature of Azure Active Directory. The feature provides Azure services with an automatically managed identity in Azure AD Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/security/synapse-workspace-managed-identity>

NEW QUESTION 8

- (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: You use a dedicated SQL pool to create an external table that has a additional DateTime column. Does this meet the goal?

- A. Yes
- B. No

Answer: A

NEW QUESTION 9

- (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 10

- (Exam Topic 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 Stream Analytics
- B. Azure SQL Database
- C. Azure Databricks
- D. Azure Synapse Analytics

Answer: A

NEW QUESTION 10

- (Exam Topic 3)

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

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

HubA: Stream HubB: Stream
 Database1: Reference

Reference data (also known as a lookup table) is a finite data set that is static or slowly changing in nature, used to perform a lookup or to augment your data streams. For example, in an IoT scenario, you could store metadata about sensors (which don't change often) in reference data and join it with real time IoT data streams. Azure Stream Analytics loads reference data in memory to achieve low latency stream processing

NEW QUESTION 15

- (Exam Topic 3)

You have an Azure subscription that contains a logical Microsoft SQL server named Server1. Server1 hosts an Azure Synapse Analytics SQL dedicated pool named Pool1.

You need to recommend a Transparent Data Encryption (TDE) solution for Server1. The solution must meet the following requirements:

- Track the usage of encryption keys.
- Maintain the access of client apps to Pool1 in the event of an Azure datacenter outage that affects the availability of the encryption keys.

What should you include in the recommendation? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

To track encryption key usage:

 ▼
 Always Encrypted
 TDE with customer-managed keys
 TDE with platform-managed keys

To maintain client app access in the event of a datacenter outage:

 ▼
 Create and configure Azure key vaults in two Azure regions.
 Enable Advanced Data Security on Server1.
 Implement the client apps by using a Microsoft .NET Framework data provider.

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: TDE with customer-managed keys

Customer-managed keys are stored in the Azure Key Vault. You can monitor how and when your key vaults are accessed, and by whom. You can do this by enabling logging for Azure Key Vault, which saves information in an Azure storage account that you provide.

Box 2: Create and configure Azure key vaults in two Azure regions

The contents of your key vault are replicated within the region and to a secondary region at least 150 miles away, but within the same geography to maintain high durability of your keys and secrets.

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/security/workspaces-encryption> <https://docs.microsoft.com/en-us/azure/key-vault/general/logging>

NEW QUESTION 19

- (Exam Topic 3)

You have an Azure Synapse Analytics serverless SQL pool named Pool1 and an Azure Data Lake Storage Gen2 account named storage1. The AllowedBlobpublicAccess property is disabled for storage1.

You need to create an external data source that can be used by Azure Active Directory (Azure AD) users to access storage1 from Pool1.

What should you create first?

- A. an external resource pool
- B. a remote service binding
- C. database scoped credentials
- D. an external library

Answer: C

NEW QUESTION 22

- (Exam Topic 3)

You are designing an application that will store petabytes of medical imaging data

When the data is first created, the data will be accessed frequently during the first week. After one month, the data must be accessible within 30 seconds, but files will be accessed infrequently. After one year, the data will be accessed infrequently but must be accessible within five minutes.

You need to select a storage strategy for the data. The solution must minimize costs.

Which storage tier should you use for each time frame? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

First week:

Archive
Cool
Hot

After one month:

Archive
Cool
Hot

After one year:

Archive
Cool
Hot

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

First week: Hot

Hot - Optimized for storing data that is accessed frequently. After one month: Cool

Cool - Optimized for storing data that is infrequently accessed and stored for at least 30 days.

After one year: Cool

NEW QUESTION 26

- (Exam Topic 3)

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.

Answer: B

Explanation:

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:

- > 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>

NEW QUESTION 27

- (Exam Topic 3)

You need to implement an Azure Databricks cluster that automatically connects to Azure Data Lake Storage Gen2 by using Azure Active Directory (Azure AD) integration.

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

Cluster Mode:

Advanced option to enable:

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: High Concurrency

Enable Azure Data Lake Storage credential passthrough for a high-concurrency cluster. Incorrect:

Support for Azure Data Lake Storage credential passthrough on standard clusters is in Public Preview.

Standard clusters with credential passthrough are supported on Databricks Runtime 5.5 and above and are limited to a single user.

Box 2: Azure Data Lake Storage Gen1 Credential Passthrough

You can authenticate automatically to Azure Data Lake Storage Gen1 and Azure Data Lake Storage Gen2 from Azure Databricks clusters using the same Azure Active Directory (Azure AD) identity that you use to log into Azure Databricks. When you enable your cluster for Azure Data Lake Storage credential passthrough, commands that you run on that cluster can read and write data in Azure Data Lake Storage without requiring you to configure service principal credentials for access to storage.

References:

<https://docs.azuredatabricks.net/spark/latest/data-sources/azure/adls-passthrough.html>

NEW QUESTION 30

- (Exam Topic 3)

You have an Azure data factory.

You need to ensure that pipeline-run data is retained for 120 days. The solution must ensure that you can query the data by using the Kusto query language.

Which four 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

Select the PipelineRuns category.

Create a Log Analytics workspace that has Data Retention set to 120 days.

Stream to an Azure event hub.

Create an Azure Storage account that has a lifecycle policy.

From the Azure portal, add a diagnostic setting.

Send the data to a Log Analytics workspace.

Select the TriggerRuns category.

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Step 1: Create an Azure Storage account that has a lifecycle policy

To automate common data management tasks, Microsoft created a solution based on Azure Data Factory. The service, Data Lifecycle Management, makes frequently accessed data available and archives or purges other data according to retention policies. Teams across the company use the service to reduce storage costs, improve app performance, and comply with data retention policies.

Step 2: Create a Log Analytics workspace that has Data Retention set to 120 days.

Data Factory stores pipeline-run data for only 45 days. Use Azure Monitor if you want to keep that data for a longer time. With Monitor, you can route diagnostic logs for analysis to multiple different targets, such as a Storage Account: Save your diagnostic logs to a storage account for auditing or manual inspection. You can use the diagnostic settings to specify the retention time in days.

Step 3: From Azure Portal, add a diagnostic setting. Step 4: Send the data to a log Analytics workspace,

Event Hub: A pipeline that transfers events from services to Azure Data Explorer. Keeping Azure Data Factory metrics and pipeline-run data.

Configure diagnostic settings and workspace.

Create or add diagnostic settings for your data factory.

- In the portal, go to Monitor. Select Settings > Diagnostic settings.
- Select the data factory for which you want to set a diagnostic setting.
- If no settings exist on the selected data factory, you're prompted to create a setting. Select Turn on diagnostics.
- Give your setting a name, select Send to Log Analytics, and then select a workspace from Log Analytics Workspace.
- Select Save. Reference:

<https://docs.microsoft.com/en-us/azure/data-factory/monitor-using-azure-monitor>

NEW QUESTION 32

- (Exam Topic 3)

You have an Azure Factory instance named DF1 that contains a pipeline named PL1. PL1 includes a tumbling window trigger.

You create five clones of PL1. You configure each clone pipeline to use a different data source.

You need to ensure that the execution schedules of the clone pipeline match the execution schedule of PL1. What should you do?

- A. Add a new trigger to each cloned pipeline
- B. Associate each cloned pipeline to an existing trigger.
- C. Create a tumbling window trigger dependency for the trigger of PL1.
- D. Modify the Concurrency setting of each pipeline.

Answer: B

NEW QUESTION 34

- (Exam Topic 3)

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?

- A. data masking
- B. Always Encrypted
- C. column-level security
- D. row-level security

Answer: A

Explanation:

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>

NEW QUESTION 36

- (Exam Topic 3)

You are planning a streaming data solution that will use Azure Databricks. The solution will stream sales transaction data from an online store. The solution has the following specifications:

* The output data will contain items purchased, quantity, line total sales amount, and line total tax amount.

* Line total sales amount and line total tax amount will be aggregated in Databricks.

* Sales transactions will never be updated. Instead, new rows will be added to adjust a sale.

You need to recommend an output mode for the dataset that will be processed by using Structured Streaming. The solution must minimize duplicate data.

What should you recommend?

- A. Append
- B. Update
- C. Complete

Answer: C

NEW QUESTION 38

- (Exam Topic 3)

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.

Answer Area:

Integration runtime type:	<input type="checkbox"/> Azure integration runtime <input checked="" type="checkbox"/> Azure-SSIS integration runtime <input type="checkbox"/> Self-hosted integration runtime
Trigger type:	<input type="checkbox"/> Event-based trigger <input checked="" type="checkbox"/> Schedule trigger <input type="checkbox"/> Tumbling window trigger
Activity type:	<input checked="" type="checkbox"/> Copy activity <input type="checkbox"/> Lookup activity <input type="checkbox"/> Stored procedure activity

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area:

Integration runtime type:	<input type="checkbox"/> Azure integration runtime <input checked="" type="checkbox"/> Azure-SSIS integration runtime <input type="checkbox"/> Self-hosted integration runtime
Trigger type:	<input type="checkbox"/> Event-based trigger <input checked="" type="checkbox"/> Schedule trigger <input type="checkbox"/> Tumbling window trigger
Activity type:	<input checked="" type="checkbox"/> Copy activity <input type="checkbox"/> Lookup activity <input type="checkbox"/> Stored procedure activity

NEW QUESTION 40

- (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 42

- (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	Answer Area
CLUSTERED INDEX	CREATE TABLE table1
COLLATE	(
DISTRIBUTION	ID INTEGER,
PARTITION	col1 VARCHAR(10),
PARTITION FUNCTION	col2 VARCHAR(10)
PARTITION SCHEME) 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 44

- (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 47

- (Exam Topic 3)

You have an Azure Data Lake Storage Gen2 container that contains 100 TB of data.

You need to ensure that the data in the container is available for read workloads in a secondary region if an outage occurs in the primary region. The solution must minimize costs.

Which type of data redundancy should you use?

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

Answer: C

NEW QUESTION 49

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

▼
LAG
LAST
LEAD

▼
LIMIT DURATION
OFFSET
WHEN

- 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 50

- (Exam Topic 3)

You need to design an Azure Synapse Analytics dedicated SQL pool that meets the following requirements:

- > Can return an employee record from a given point in time.
- > Maintains the latest employee information.
- > Minimizes query complexity.

How should you model the employee data?

- A. as a temporal table
- B. as a SQL graph table
- C. as a degenerate dimension table
- D. as a Type 2 slowly changing dimension (SCD) table

Answer: D

Explanation:

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 53

- (Exam Topic 3)

You have an Azure Synapse Analytics dedicated SQL Pool1. Pool1 contains a partitioned fact table named dbo.Sales and a staging table named stg.Sales that has the matching table and partition definitions.

You need to overwrite the content of the first partition in dbo.Sales with the content of the same partition in stg.Sales. The solution must minimize load times. What should you do?

- A. Switch the first partition from dbo.Sales to stg.Sales.
- B. Switch the first partition from stg.Sales to db
- C. Sales.
- D. Update dbo.Sales from stg.Sales.
- E. Insert the data from stg.Sales into dbo.Sales.

Answer: D

NEW QUESTION 58

- (Exam Topic 3)

You are designing the folder structure for an Azure Data Lake Storage Gen2 container.

Users will query data by using a variety of services including Azure Databricks and Azure Synapse Analytics serverless SQL pools. The data will be secured by subject area. Most queries will include data from the current year or current month.

Which folder structure should you recommend to support fast queries and simplified folder security?

- A. /{SubjectArea}/{DataSource}/{DD}/{MM}/{YYYY}/{FileData}_{YYYY}_{MM}_{DD}.csv
- B. /{DD}/{MM}/{YYYY}/{SubjectArea}/{DataSource}/{FileData}_{YYYY}_{MM}_{DD}.csv

- C. /{YYYY}/{MM}/{DD}/{SubjectArea}/{DataSource}/{FileData}_{YYYY}_{MM}_{DD}.csv
- D. /{SubjectArea}/{DataSource}/{YYYY}/{MM}/{DD}/{FileData}_{YYYY}_{MM}_{DD}.csv

Answer: D

Explanation:

There's an important reason to put the date at the end of the directory structure. If you want to lock down certain regions or subject matters to users/groups, then you can easily do so with the POSIX permissions. Otherwise, if there was a need to restrict a certain security group to viewing just the UK data or certain planes, with the date structure in front a separate permission would be required for numerous directories under every hour directory. Additionally, having the date structure in front would exponentially increase the number of directories as time went on.

Note: In IoT workloads, there can be a great deal of data being landed in the data store that spans across numerous products, devices, organizations, and customers. It's important to pre-plan the directory layout for organization, security, and efficient processing of the data for down-stream consumers. A general template to consider might be the following layout:

{Region}/{SubjectMatter(s)}/{yyyy}/{mm}/{dd}/{hh}/

NEW QUESTION 60

- (Exam Topic 3)

You are creating an Azure Data Factory data flow that will ingest data from a CSV file, cast columns to specified types of data, and insert the data into a table in an Azure Synapse Analytic dedicated SQL pool. The CSV file contains three columns named username, comment, and date.

The data flow already contains the following:

- > A source transformation.
- > A Derived Column transformation to set the appropriate types of data.
- > A sink transformation to land the data in the pool.

You need to ensure that the data flow meets the following requirements:

- > All valid rows must be written to the destination table.
- > Truncation errors in the comment column must be avoided proactively.
- > Any rows containing comment values that will cause truncation errors upon insert must be written to a file in blob storage.

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

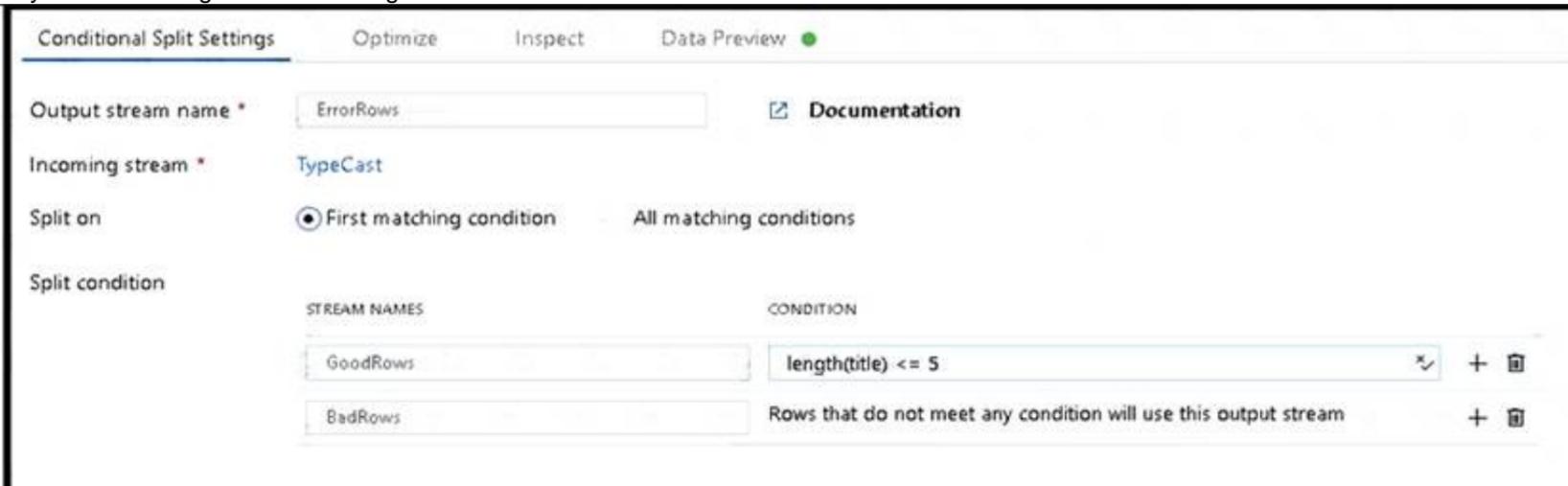
- A. To the data flow, add a sink transformation to write the rows to a file in blob storage.
- B. To the data flow, add a Conditional Split transformation to separate the rows that will cause truncation errors.
- C. To the data flow, add a filter transformation to filter out rows that will cause truncation errors.
- D. Add a select transformation to select only the rows that will cause truncation errors.

Answer: AB

Explanation:

B: Example:

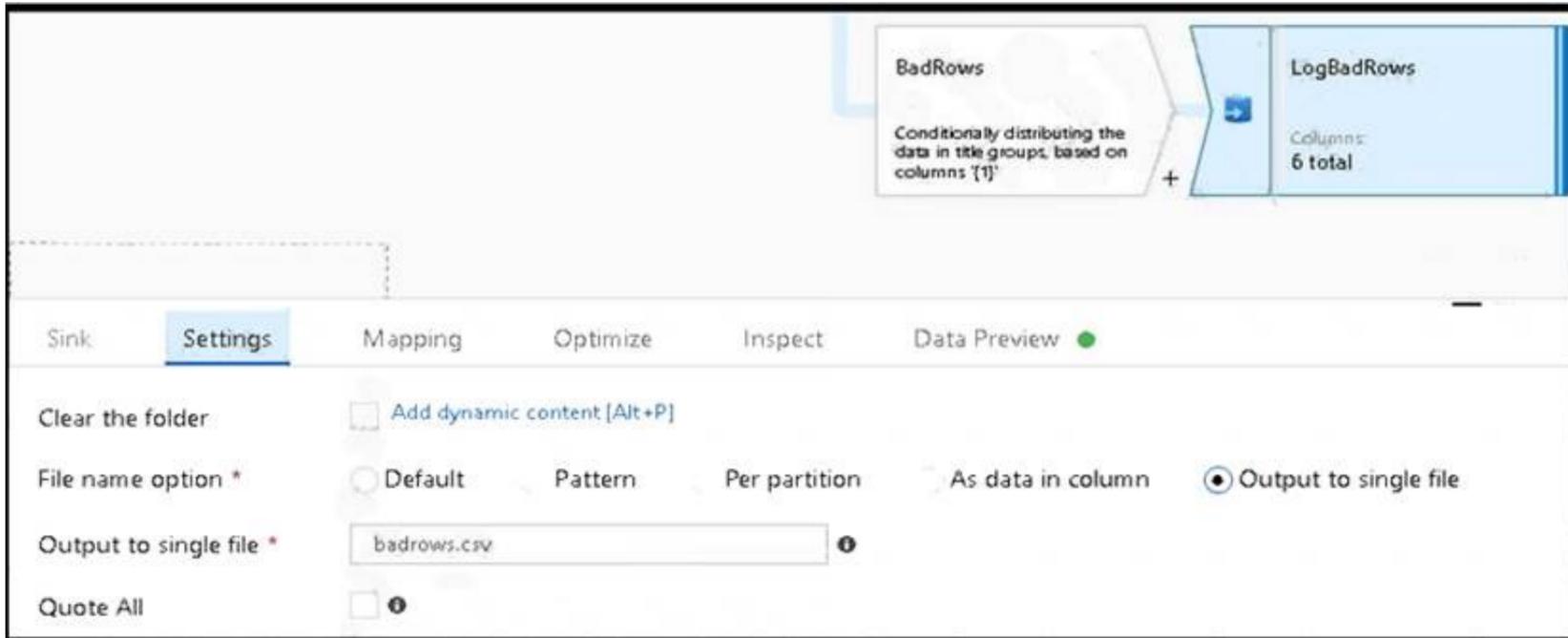
- * 1. This conditional split transformation defines the maximum length of "title" to be five. Any row that is less than or equal to five will go into the GoodRows stream. Any row that is larger than five will go into the BadRows stream.



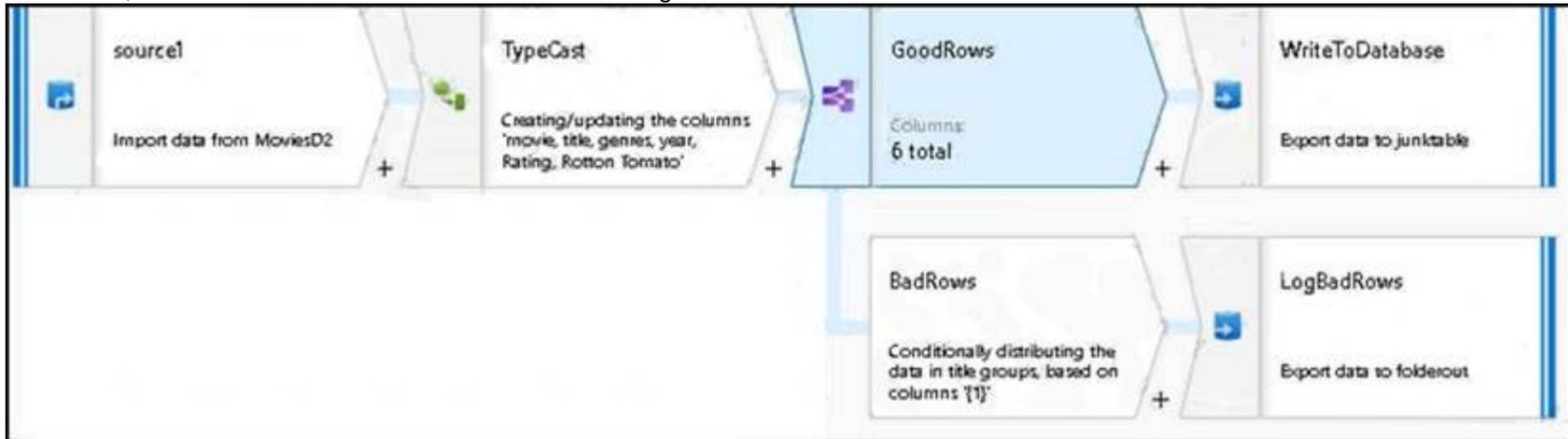
- * 2. This conditional split transformation defines the maximum length of "title" to be five. Any row that is less than or equal to five will go into the GoodRows stream. Any row that is larger than five will go into the BadRows stream.

A:

- * 3. Now we need to log the rows that failed. Add a sink transformation to the BadRows stream for logging. Here, we'll "auto-map" all of the fields so that we have logging of the complete transaction record. This is a text-delimited CSV file output to a single file in Blob Storage. We'll call the log file "badrows.csv".



* 4. The completed data flow is shown below. We are now able to split off error rows to avoid the SQL truncation errors and put those entries into a log file. Meanwhile, successful rows can continue to write to our target database.



Reference:
<https://docs.microsoft.com/en-us/azure/data-factory/how-to-data-flow-error-rows>

NEW QUESTION 62

- (Exam Topic 3)

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 area.

NOTE: Each correct selection is worth one point.

Answer Area

```

Select TimeZone, count(*) AS MessageCount
FROM
    MessageStream
GROUP BY
    TimeZone,
    [ ] CreatedAt
    [ ] (second, 15)
    
```

Options for the first []:

- LAST OVER SYSTEM.TIMESTAMP() TIMESTAMP BY

Options for the second []:

- HOPPINGWINDOW
- SESSIONWINDOW
- SLIDINGWINDOW
- TUMBLINGWINDOW

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area

```
Select TimeZone, count(*) AS MessageCount
FROM
MessageStream
GROUP BY
TimeZone,
```

LAST
OVER
SYSTEM.TIMESTAMP() |
TIMESTAMP BY

CreatedAt

HOPPINGWINDOW
SESSIONWINDOW
SLIDINGWINDOW
TUMBLINGWINDOW

(second, 15)

NEW QUESTION 66

- (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 67

.....

THANKS FOR TRYING THE DEMO OF OUR PRODUCT

Visit Our Site to Purchase the Full Set of Actual DP-203 Exam Questions With Answers.

We Also Provide Practice Exam Software That Simulates Real Exam Environment And Has Many Self-Assessment Features. Order the DP-203 Product From:

<https://www.2passeasy.com/dumps/DP-203/>

Money Back Guarantee

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