

DP-203 Dumps

Data Engineering on Microsoft Azure

<https://www.certleader.com/DP-203-dumps.html>



NEW QUESTION 1

- (Exam Topic 3)

A company has a real-time data analysis solution that is hosted on Microsoft Azure. The solution uses Azure Event Hub to ingest data and an Azure Stream Analytics cloud job to analyze the data. The cloud job is configured to use 120 Streaming Units (SU).

You need to optimize performance for the Azure Stream Analytics job.

Which two actions should you perform? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. Implement event ordering.
- B. Implement Azure Stream Analytics user-defined functions (UDF).
- C. Implement query parallelization by partitioning the data output.
- D. Scale the SU count for the job up.
- E. Scale the SU count for the job down.
- F. Implement query parallelization by partitioning the data input.

Answer: DF

Explanation:

Reference:

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

NEW QUESTION 2

- (Exam Topic 3)

HOTSPOT

You have an Azure Data Factory instance named ADF1 and two Azure Synapse Analytics workspaces named WS1 and WS2.

ADF1 contains the following pipelines:

- > P1: Uses a copy activity to copy data from a nonpartitioned table in a dedicated SQL pool of WS1 to an Azure Data Lake Storage Gen2 account
- > P2: Uses a copy activity to copy data from text-delimited files in an Azure Data Lake Storage Gen2 account to a nonpartitioned table in a dedicated SQL pool of WS2

You need to configure P1 and P2 to maximize parallelism and performance.

Which dataset settings should you configure for the copy activity if each pipeline? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

P1:

	▼
Set the Copy method to Bulk insert	
Set the Copy method to PolyBase	
Set the Isolation level to Repeatable read	
Set the Partition option to Dynamic range	

P2:

	▼
Set the Copy method to Bulk insert	
Set the Copy method to PolyBase	
Set the Isolation level to Repeatable read	
Set the Partition option to Dynamic range	

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Set the Copy method to PolyBase

While SQL pool supports many loading methods including non-Polybase options such as BCP and SQL BulkCopy API, the fastest and most scalable way to load data is through PolyBase. PolyBase is a technology that accesses external data stored in Azure Blob storage or Azure Data Lake Store via the T-SQL language.

Box 2: Set the Copy method to Bulk insert

Polybase not possible for text files. Have to use Bulk insert. Reference:

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

NEW QUESTION 3

- (Exam Topic 3)

You have an Azure Data Lake Storage account that contains a staging zone.

You need to design a dairy process to ingest incremental data from the staging zone, transform the data by executing an R script, and then insert the transformed data into a data warehouse in Azure Synapse Analytics.

Solution: You use an Azure Data Factory schedule trigger to execute a pipeline that copies the data to a staging table in the data warehouse, and then uses a stored procedure to execute the R script.

Does this meet the goal?

- A. Yes
- B. No

Answer: A

Explanation:

If you need to transform data in a way that is not supported by Data Factory, you can create a custom activity with your own data processing logic and use the activity in the pipeline.

Note: You can use data transformation activities in Azure Data Factory and Synapse pipelines to transform and process your raw data into predictions and insights at scale.

Reference:

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

NEW QUESTION 4

- (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 5

- (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: ▼

HubB: ▼

Database1: ▼

- 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

Reference:

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

NEW QUESTION 6

- (Exam Topic 3)

You have a SQL pool in Azure Synapse.

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

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

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

NOTE: Each correct selection is worth one point.

Distribution: ▼

Indexing: ▼

Partitioning: ▼

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Graphical user interface, application, table Description automatically generated

Box 1: Hash

Hash-distributed tables improve query performance on large fact tables. They can have very large numbers of rows and still achieve high performance.

Box 2: Clustered columnstore

When creating partitions on clustered columnstore tables, it is important to consider how many rows belong to each partition. For optimal compression and performance of clustered columnstore tables, a minimum of 1 million rows per distribution and partition is needed.

Box 3: Date

Table partitions enable you to divide your data into smaller groups of data. In most cases, table partitions are created on a date column.

Partition switching can be used to quickly remove or replace a section of a table. Reference:

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

NEW QUESTION 7

- (Exam Topic 3)

You have an Azure Data Factory pipeline that contains a data flow. The data flow contains the following expression.

```
source(output(
    License_plate as string,
    Make as string,
    Time as string
),
allowSchemaDrift: true,
```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

See below answer.

Answer Area

Number of columns: 22 ▼

Number of rows: 4 ▼

NEW QUESTION 8

- (Exam Topic 3)

The following code segment is used to create an Azure Databricks cluster.

```
{
  "num_workers": null,
  "autoscale": {
    "min_workers": 2,
    "max_workers": 8
  },
  "cluster_name": "MyCluster",
  "spark_version": "latest-stable-scala2.11",
  "spark_conf": {
    "spark.databricks.cluster.profile": "serverless",
    "spark.databricks.repl.allowedLanguages": "sql,python,r"
  },
  "node_type_id": "Standard_DS13_v2",
  "ssh_public_keys": [],
  "custom_tags": {
    "ResourceClass": "Serverless"
  },
  "spark_env_vars": {
    "PYSPARK_PYTHON": "/databricks/python3/bin/python3"
  },
  "autotermination_minutes": 90,
  "enable_elastic_disk": true,
  "init_scripts": []
}
```

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 Databricks cluster supports multiple concurrent users.	<input type="radio"/>	<input type="radio"/>
The Databricks cluster minimizes costs when running scheduled jobs that execute notebooks.	<input type="radio"/>	<input type="radio"/>
The Databricks cluster supports the creation of a Delta Lake table.	<input type="radio"/>	<input type="radio"/>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Graphical user interface, text, application Description automatically generated

Box 1: Yes

A cluster mode of 'High Concurrency' is selected, unlike all the others which are 'Standard'. This results in a worker type of Standard_DS13_v2.

Box 2: No

When you run a job on a new cluster, the job is treated as a data engineering (job) workload subject to the job workload pricing. When you run a job on an existing cluster, the job is treated as a data analytics (all-purpose) workload subject to all-purpose workload pricing.

Box 3: Yes

Delta Lake on Databricks allows you to configure Delta Lake based on your workload patterns. Reference:

<https://adatis.co.uk/databricks-cluster-sizing/> <https://docs.microsoft.com/en-us/azure/databricks/jobs>

<https://docs.databricks.com/administration-guide/capacity-planning/cmbp.html> <https://docs.databricks.com/delta/index.html>

NEW QUESTION 9

- (Exam Topic 3)

DRAG DROP

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]]))

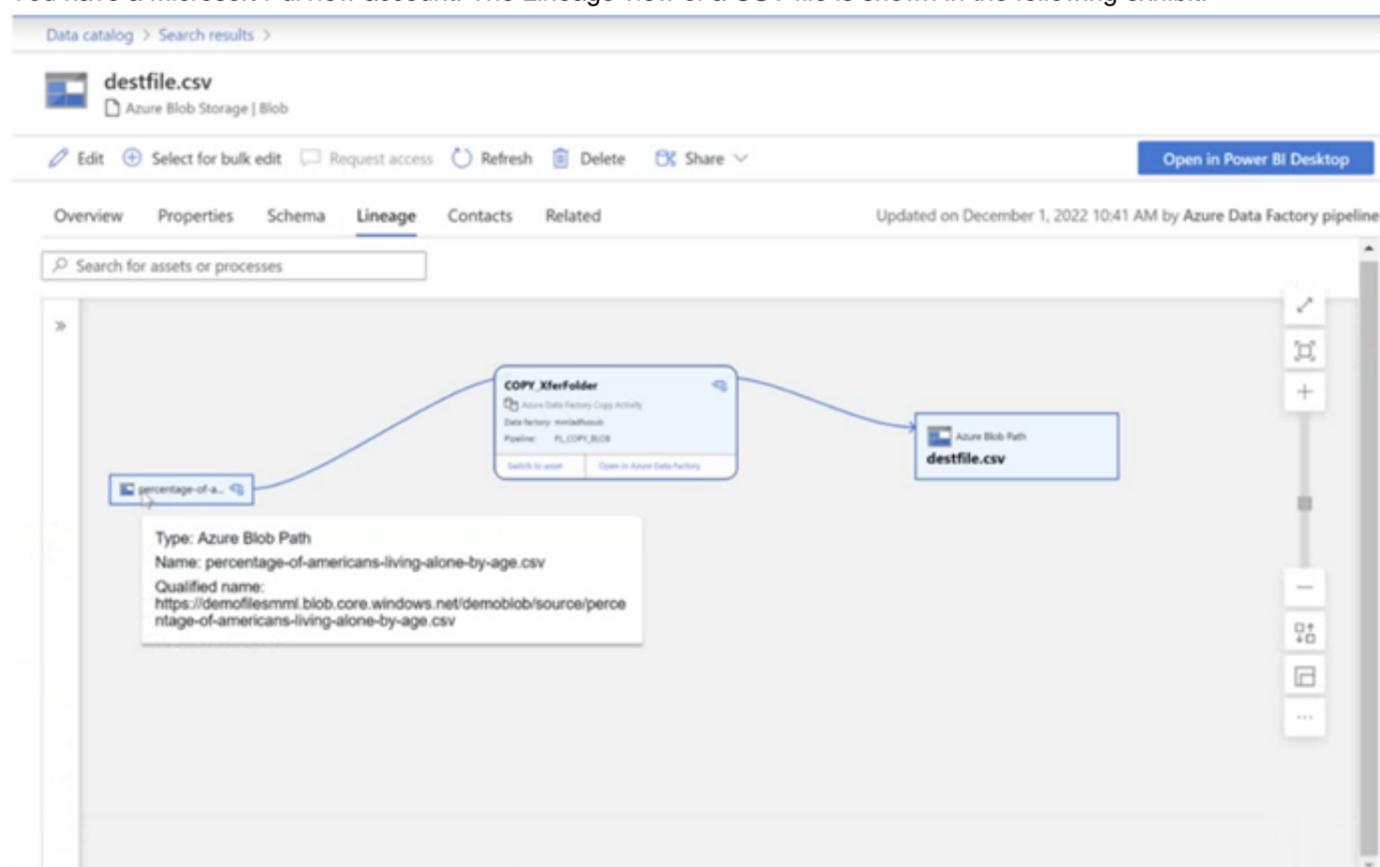
Reference:

<https://docs.microsoft.com/en-us/sql/t-sql/statements/create-table-azure-sql-data-warehouse?>

NEW QUESTION 10

- (Exam Topic 3)

You have a Microsoft Purview account. The Lineage view of a CSV file is shown in the following exhibit.



How is the data for the lineage populated?

- A. manually
- B. by scanning data stores
- C. by executing a Data Factory pipeline

Answer: B

Explanation:

According to Microsoft Purview Data Catalog lineage user guide¹, data lineage in Microsoft Purview is a core platform capability that populates the Microsoft Purview Data Map with data movement and transformations across systems². Lineage is captured as it flows in the enterprise and stitched without gaps irrespective of its source².

NEW QUESTION 10

- (Exam Topic 3)

You are performing exploratory analysis of the bus fare data in an Azure Data Lake Storage Gen2 account by using an Azure Synapse Analytics serverless SQL pool.

You execute the Transact-SQL query shown in the following exhibit.

```
SELECT
    payment_type,
    SUM(fare_amount) AS fare_total
FROM OPENROWSET (
    BULK 'csv/busfare/tripdata_2020*.csv',
    DATA_SOURCE = 'BusData',
    FORMAT = 'CSV', PARSER_VERSION = '2.0',
    FIRSTROW = 2
)
WITH (
    payment_type INT 10,
    fare_amount FLOAT 11
) AS nyc
GROUP BY payment_type
ORDER BY payment_type;
```

What do the query results include?

- A. Only CSV files in the tripdata_2020 subfolder.
- B. All files that have file names that beginning with "tripdata_2020".
- C. All CSV files that have file names that contain "tripdata_2020".
- D. Only CSV that have file names that beginning with "tripdata_2020".

Answer: D

NEW QUESTION 13

- (Exam Topic 3)

You are designing an application that will use an Azure Data Lake Storage Gen 2 account to store petabytes of license plate photos from toll booths. The account will use zone-redundant storage (ZRS).

You identify the following usage patterns:

- The data will be accessed several times a day during the first 30 days after the data is created. The data must meet an availability SU of 99.9%.
- After 90 days, the data will be accessed infrequently but must be available within 30 seconds.
- After 365 days, the data will be accessed infrequently but must be available within five minutes.

First 30 days:

Archive

Cool

Hot

After 90 days:

Archive

Cool

Hot

After 365 days:

Archive

Cool

Hot

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Hot

The data will be accessed several times a day during the first 30 days after the data is created. The data must meet an availability SLA of 99.9%.

Box 2: Cool

After 90 days, the data will be accessed infrequently but must be available within 30 seconds. Data in the Cool tier should be stored for a minimum of 30 days.

When your data is stored in an online access tier (either Hot or Cool), users can access it immediately. The Hot tier is the best choice for data that is in active use, while the Cool tier is ideal for data that is accessed less frequently, but that still must be available for reading and writing.

Box 3: Cool

After 365 days, the data will be accessed infrequently but must be available within five minutes. Reference: <https://docs.microsoft.com/en-us/azure/storage/blobs/access-tiers-overview> <https://docs.microsoft.com/en-us/azure/storage/blobs/archive-rehydrate-overview>

NEW QUESTION 16

- (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: B

Explanation:

By default, streams run in append mode, which adds new records to the table. <https://docs.databricks.com/delta/delta-streaming.html>

NEW QUESTION 21

- (Exam Topic 3)

A company plans to use Platform-as-a-Service (PaaS) to create the new data pipeline process. The process must meet the following requirements:

Ingest:

- > Access multiple data sources.
- > Provide the ability to orchestrate workflow.
- > Provide the capability to run SQL Server Integration Services packages.

Store:

Optimize storage for big data workloads. Provide encryption of data at rest. Operate with no size limits.

Prepare and Train:

- > Provide a fully-managed and interactive workspace for exploration and visualization.
 - > Provide the ability to program in R, SQL, Python, Scala, and Java.
 - > Provide seamless user authentication with Azure Active Directory.
- Model & Serve:
- > Implement native columnar storage.
 - > Support for the SQL language
 - > Provide support for structured streaming. You need to build the data integration pipeline.

Which technologies should you use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

Architecture requirement	Technology
Ingest	<div style="border: 1px solid black; padding: 5px;"> <div style="display: flex; justify-content: space-between; align-items: center;"> ▼ </div> <ul style="list-style-type: none"> Logic Apps Azure Data Factory Azure Automation </div>
Store	<div style="border: 1px solid black; padding: 5px;"> <div style="display: flex; justify-content: space-between; align-items: center;"> ▼ </div> <ul style="list-style-type: none"> Azure Data Lake Storage Azure Blob storage Azure files </div>
Prepare and Train	<div style="border: 1px solid black; padding: 5px;"> <div style="display: flex; justify-content: space-between; align-items: center;"> ▼ </div> <ul style="list-style-type: none"> HDInsight Apache Spark cluster Azure Databricks HDInsight Apache Storm cluster </div>
Model and Serve	<div style="border: 1px solid black; padding: 5px;"> <div style="display: flex; justify-content: space-between; align-items: center;"> ▼ </div> <ul style="list-style-type: none"> HDInsight Apache Kafka cluster Azure Synapse Analytics Azure Data Lake Storage </div>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Graphical user interface, application, table, email Description automatically generated

NEW QUESTION 23

- (Exam Topic 3)

You have an Azure Stream Analytics job.

You need to ensure that the job has enough streaming units provisioned. You configure monitoring of the SU % Utilization metric.

Which two additional metrics should you monitor? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. Backlogged Input Events
- B. Watermark Delay
- C. Function Events
- D. Out of order Events
- E. Late Input Events

Answer: AB

Explanation:

To react to increased workloads and increase streaming units, consider setting an alert of 80% on the SU Utilization metric. Also, you can use watermark delay and backlogged events metrics to see if there is an impact.

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, by increasing the SUs.

Reference:

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

NEW QUESTION 28

- (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

Answer Area

- CASE
- ELSE
- OVER
- PARTITION BY
- ROW_NUMBER

```

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

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Graphical user interface, text, application Description automatically generated

Box 1: CASE

CASE evaluates a list of conditions and returns one of multiple possible result expressions.

CASE can be used in any statement or clause that allows a valid expression. For example, you can use CASE in statements such as SELECT, UPDATE, DELETE and SET, and in clauses such as select_list, IN, WHERE, ORDER BY, and HAVING.

Syntax: Simple CASE expression: CASE input_expression

WHEN when_expression THEN result_expression [...n] [ELSE else_result_expression]

END

Box 2: ELSE

Reference:

<https://docs.microsoft.com/en-us/sql/t-sql/language-elements/case-transact-sql>

NEW QUESTION 32

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

Does this meet the goal?

A. Yes

B. No

Answer: B

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 34

- (Exam Topic 3)

You are building an Azure Stream Analytics job to identify how much time a user spends interacting with a feature on a webpage.

The job receives events based on user actions on the webpage. Each row of data represents an event. Each event has a type of either 'start' or 'end'.

You need to calculate the duration between start and end events.

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
[user],
feature,
DATEADD(
DATEDIFF(
DATEPART(
second,
(Time) OVER (PARTITION BY [user], feature LIMIT DURATION(hour, 1) WHEN Event = 'start'),
ISFIRST
LAST
TOPONE
Time) as duration
FROM input TIMESTAMP BY Time
WHERE
Event = 'end'
```

A. Mastered

B. Not Mastered

Answer: A

Explanation:

Box 1: DATEDIFF

DATEDIFF function returns the count (as a signed integer value) of the specified datepart boundaries crossed between the specified startdate and enddate.

Syntax: DATEDIFF (datepart , startdate, enddate) Box 2: LAST

The LAST function can be used to retrieve the last event within a specific condition. In this example, the condition is an event of type Start, partitioning the search by PARTITION BY user and feature. This way, every user and feature is treated independently when searching for the Start event. LIMIT DURATION limits the search back in time to 1 hour between the End and Start events.

Example: SELECT

```
[user], feature, DATEDIFF(
second,
LAST(Time) OVER (PARTITION BY [user], feature LIMIT DURATION(hour,
1) WHEN Event = 'start'), Time) as duration
FROM input TIMESTAMP BY Time
WHERE
Event = 'end'
```

Reference:

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

NEW QUESTION 39

- (Exam Topic 3)

You are monitoring an Azure Stream Analytics job.

You discover that the Backlogged Input Events metric is increasing slowly and is consistently non-zero. You need to ensure that the job can handle all the events.

What should you do?

A. Change the compatibility level of the Stream Analytics job.

- B. Increase the number of streaming units (SUs).
- C. Remove any named consumer groups from the connection and use \$default.
- D. Create an additional output stream for the existing input stream.

Answer: B

Explanation:

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. You should increase the Streaming Units.

Note: Streaming Units (SUs) represents the computing resources that are allocated to execute a Stream Analytics job. The higher the number of SUs, the more CPU and memory resources are allocated for your job.

Reference:

<https://docs.microsoft.com/bs-cyrl-ba/azure/stream-analytics/stream-analytics-monitoring>

NEW QUESTION 40

- (Exam Topic 3)

You have an Azure subscription that contains an Azure Databricks workspace named databricks1 and an Azure Synapse Analytics workspace named synapse1. The synapse1 workspace contains an Apache Spark pool named pool1.

You need to share an Apache Hive catalog of pool1 with databricks1.

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

NOTE: Each correct selection is worth one point.

From synapse1, create a linked service to:

Azure Cosmos DB
Azure Data Lake Storage Gen2
Azure SQL Database

Configure pool1 to use the linked service as:

An Azure Purview account
A Hive metastore
A managed Hive metastore service

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Azure SQL Database

Use external Hive Metastore for Synapse Spark Pool

Azure Synapse Analytics allows Apache Spark pools in the same workspace to share a managed HMS (Hive Metastore) compatible metastore as their catalog.

Set up linked service to Hive Metastore

Follow below steps to set up a linked service to the external Hive Metastore in Synapse workspace.

- > Set up Hive Metastore linked service
- > Choose Azure SQL Database or Azure Database for MySQL based on your database type, click Continue.
- > Provide Name of the linked service. Record the name of the linked service, this info will be used to configure Spark shortly.
- > You can either select Azure SQL Database/Azure Database for MySQL for the external Hive Metastore from Azure subscription list, or enter the info manually.
- > Provide User name and Password to set up the connection.
- > Test connection to verify the username and password.
- > Click Create to create the linked service.

Box 2: A Hive Metastore

nce: <https://docs.microsoft.com/en-us/azure/synapse-analytics/spark/apache-spark-external-metastore>

NEW QUESTION 45

- (Exam Topic 3)

You have an Azure Synapse Analytics dedicated SQL pool named Pool1 that contains an external table named Sales. Sales contains sales data. Each row in Sales

contains data on a single sale, including the name of the salesperson.

You need to implement row-level security (RLS). The solution must ensure that the salespeople can access only their respective sales.

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

Create:

Restrict row access by using:

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: A security policy for sale

Here are the steps to create a security policy for Sales:

- > Create a user-defined function that returns the name of the current user:
- > CREATE FUNCTION dbo.GetCurrentUser()
- > RETURNS NVARCHAR(128)
- > AS
- > BEGIN
- > RETURN SUSER_SNAME();
- > END;
- > Create a security predicate function that filters the Sales table based on the current user:
- > CREATE FUNCTION dbo.SalesPredicate(@salesperson NVARCHAR(128))
- > RETURNS TABLE
- > WITH SCHEMABINDING
- > AS
- > RETURN SELECT 1 AS access_result
- > WHERE @salesperson = SalespersonName;
- > Create a security policy on the Sales table that uses the SalesPredicate function to filter the data:
- > CREATE SECURITY POLICY SalesFilter
- > ADD FILTER PREDICATE dbo.SalesPredicate(dbo.GetCurrentUser()) ON dbo.Sales
- > WITH (STATE = ON);

By creating a security policy for the Sales table, you ensure that each salesperson can only access their own sales data. The security policy uses a user-defined function to get the name of the current user and a security predicate function to filter the Sales table based on the current user.

Box 2: table-value function

to restrict row access by using row-level security, you need to create a table-valued function that returns a table of values that represent the rows that a user can access. You then use this function in a security policy that applies a predicate on the table.

NEW QUESTION 49

- (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 50

- (Exam Topic 2)

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 51

- (Exam Topic 2)

What should you do to improve high availability of the real-time data processing solution?

- A. Deploy identical Azure Stream Analytics jobs to paired regions in Azure.
- B. Deploy a High Concurrency Databricks cluster.
- C. Deploy an Azure Stream Analytics job and use an Azure Automation runbook to check the status of the job and to start the job if it stops.
- D. Set Data Lake Storage to use geo-redundant storage (GRS).

Answer: A

Explanation:

Guarantee Stream Analytics job reliability during service updates

Part of being a fully managed service is the capability to introduce new service functionality and improvements at a rapid pace. As a result, Stream Analytics can have a service update deploy on a weekly (or more frequent) basis. No matter how much testing is done there is still a risk that an existing, running job may break due to the introduction of a bug. If you are running mission critical jobs, these risks need to be avoided. You can reduce this risk by following Azure's paired region model.

Scenario: The application development team will create an Azure event hub to receive real-time sales data, including store number, date, time, product ID, customer loyalty number, price, and discount amount, from the point of sale (POS) system and output the data to data storage in Azure

Reference:

<https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-job-reliability>

NEW QUESTION 56

- (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 57

- (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 Analytics dedicated SQL pool. The CSV file contains 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. Add a select transformation that selects only the rows which will cause truncation errors.
- B. Add a sink transformation that writes the rows to a file in blob storage.
- C. Add a filter transformation that filters out rows which will cause truncation errors.
- D. Add a Conditional Split transformation that separates the rows which will cause truncation errors.

Answer: BD

NEW QUESTION 62

- (Exam Topic 3)

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

You need to recommend which authentication methods to use for Databricks and Data Lake Storage to provide the users with the appropriate access. The solution must minimize administrative effort and development effort.

Which authentication method should you recommend for each Azure service? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Databricks:

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

Data Lake Storage:

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

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Table Description automatically generated

Box 1: Personal access tokens

You can use storage shared access signatures (SAS) to access an Azure Data Lake Storage Gen2 storage account directly. With SAS, you can restrict access to a storage account using temporary tokens with fine-grained access control.

You can add multiple storage accounts and configure respective SAS token providers in the same Spark session.

Box 2: Azure Active Directory credential passthrough

You can authenticate automatically to Azure Data Lake Storage Gen1 (ADLS Gen1) and Azure Data Lake Storage Gen2 (ADLS 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.

After configuring Azure Data Lake Storage credential passthrough and creating storage containers, you can access data directly in Azure Data Lake Storage Gen1 using an adl:// path and Azure Data Lake Storage Gen2 using an abfs:// path:

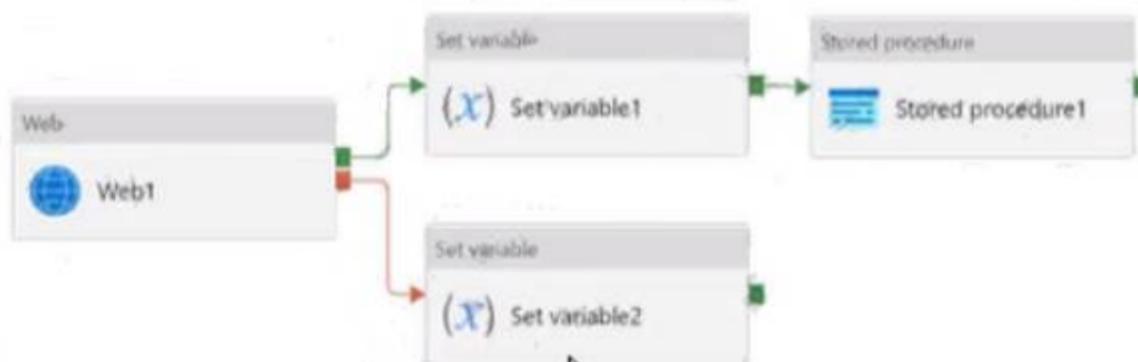
Reference:

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

NEW QUESTION 67

- (Exam Topic 3)

You have an Azure Data Factory pipeline that has the activity 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.

Answer Area

Stored procedure1 will execute if Web1 and Set variable1 [answer choice]

complete
fail
success

These are the selections for the statement Stored procedure1 will execute if Web1 and Set variable1 [answer choice]

If Web1 fails and Set variable2 succeeds, the pipeline status will be [answer choice]

Cancelled
Failed
Success

These are the selections for the statement If Web1 fails and Set variable2 succeeds, the pipeline status will be [answer choice]

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area

Stored procedure1 will execute if Web1 and Set variable1 [answer choice] succeed

If Web1 fails and Set variable2 succeeds, the pipeline status will be [answer choice] Failed

NEW QUESTION 69

- (Exam Topic 3)

You have an enterprise data warehouse in Azure Synapse Analytics.

You need to monitor the data warehouse to identify whether you must scale up to a higher service level to accommodate the current workloads

Which is the best metric to monitor?

More than one answer choice may achieve the goal. Select the BEST answer.

- A. Data 10 percentage
- B. CPU percentage
- C. DWU used
- D. DWU percentage

Answer: C

NEW QUESTION 72

- (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: discount : 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 73

- (Exam Topic 3)

You have an Azure Synapse Analytics dedicated SQL pool that contains a table named Contacts. Contacts contains a column named Phone.

You need to ensure that users in a specific role only see the last four digits of a phone number when querying the Phone column.

What should you include in the solution?

- A. a default value
- B. dynamic data masking

- C. row-level security (RLS)
- D. column encryption
- E. table partitions

Answer: B

Explanation:

Dynamic data masking helps prevent unauthorized access to sensitive data by enabling customers to designate how much of the sensitive data to reveal with minimal impact on the application layer. It's a policy-based security feature that hides the sensitive data in the result set of a query over designated database fields, while the data in the database is not changed.

Reference:

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

NEW QUESTION 76

- (Exam Topic 3)

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

You need to monitor for replication delays that can affect the recovery point objective (RPO). What should you include in the monitoring solution?

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

Answer: D

Explanation:

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

Reference:

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

NEW QUESTION 79

- (Exam Topic 3)

You have an Azure Synapse Analytics dedicated SQL pool named pool1.

You need to perform a monthly audit of SQL statements that affect sensitive data. The solution must minimize administrative effort.

What should you include in the solution?

- A. Microsoft Defender for SQL
- B. dynamic data masking
- C. sensitivity labels
- D. workload management

Answer: B

NEW QUESTION 83

- (Exam Topic 3)

You plan to perform batch processing in Azure Databricks once daily. Which type of Databricks cluster should you use?

- A. High Concurrency
- B. automated
- C. interactive

Answer: C

Explanation:

Azure Databricks has two types of clusters: interactive and automated. You use interactive clusters to analyze data collaboratively with interactive notebooks. You use automated clusters to run fast and robust automated jobs.

Example: Scheduled batch workloads (data engineers running ETL jobs)

This scenario involves running batch job JARs and notebooks on a regular cadence through the Databricks platform.

The suggested best practice is to launch a new cluster for each run of critical jobs. This helps avoid any issues (failures, missing SLA, and so on) due to an existing workload (noisy neighbor) on a shared cluster.

Reference:

<https://docs.databricks.com/administration-guide/cloud-configurations/aws/cmbp.html#scenario-3-scheduled-bat>

NEW QUESTION 86

- (Exam Topic 3)

You need to implement a Type 3 slowly changing dimension (SCD) for product category data in an Azure Synapse Analytics dedicated SQL pool.

You have a table that was created by using the following Transact-SQL statement.

```
CREATE TABLE [DBO].[DimProduct] (
[ProductKey] [int] IDENTITY(1,1) NOT NULL,
[ProductSourceID] [int] NOT NULL,
[ProductName] [nvarchar] (100) NULL,
[Color] [nvarchar] (15) NULL,
[SellStartDate] [date] NOT NULL,
[SellEndDate] [date] NULL,
[RowInsertedDateTime] [datetime] NOT NULL,
[RowUpdatedDateTime] [datetime] NOT NULL,
[ETLAuditID] [int] NOT NULL
)
```

Which two columns should you add to the table? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. [EffectiveScarcDate] [datetime] NOT NULL,
- B. [CurrentProducccAcegory] [nvarchar] (100) NOT NULL,
- C. [EffectiveEndDace] [dacecime] NULL,
- D. [ProductCategory] [nvarchar] (100) NOT NULL,
- E. [OriginalProducccAcegory] [nvarchar] (100) NOT NULL,

Answer: BE

Explanation:

A Type 3 SCD supports storing two versions of a dimension member as separate columns. The table includes a column for the current value of a member plus either the original or previous value of the member. So Type 3 uses additional columns to track one key instance of history, rather than storing additional rows to track each change like in a Type 2 SCD.

This type of tracking may be used for one or two columns in a dimension table. It is not common to use it for many members of the same table. It is often used in combination with Type 1 or Type 2 members.

Graphical user interface, application, email Description automatically generated

CustomerID	FirstName	LastName	CurrentEmail	OriginalEmail	CompanyName	InsertedDate	ModifiedDate
2	Keith	Harris	keith0@aw.com	keith0@aw.com	Progressive Sports	2021-03-20	2021-03-20
3	Donna	Carreras	donna0@aw.com	donna0@aw.com	A Bike Store	2021-03-20	2021-03-20

CustomerID	FirstName	LastName	CurrentEmail	OriginalEmail	CompanyName	InsertedDate	ModifiedDate
2	Keith	Harris	keith0@aw.com	keith0@aw.com	Progressive Sports	2021-03-20	2021-03-20
3	Donna	Carreras	dc3@aw.com	donna0@aw.com	A Bike Store	2021-03-20	2021-03-22

Reference:

<https://k21academy.com/microsoft-azure/azure-data-engineer-dp203-q-a-day-2-live-session-review/>

NEW QUESTION 89

- (Exam Topic 3)

You have an Azure Synapse Analytics dedicated SQL pool named Pool1. Pool1 contains a table named table1. You load 5 TB of data into table1. You need to ensure that columnstore compression is maximized for table1. Which statement should you execute?

- A. ALTER INDEX ALL on table1 REORGANIZE
- B. ALTER INDEX ALL on table1 REBUILD
- C. DBCC DBREINDEX (table1)
- D. DBCC INDEXDEFRAG (pool1,table1)

Answer: B

Explanation:

Columnstore and columnstore archive compression

Columnstore tables and indexes are always stored with columnstore compression. You can further reduce the size of columnstore data by configuring an additional compression called archival compression. To perform archival compression, SQL Server runs the Microsoft XPRESS compression algorithm on the data. Add or remove archival compression by using the following data compression types:

Use COLUMNSTORE_ARCHIVE data compression to compress columnstore data with archival compression.

Use COLUMNSTORE data compression to decompress archival compression. The resulting data continue to be compressed with columnstore compression.

To add archival compression, use ALTER TABLE (Transact-SQL) or ALTER INDEX (Transact-SQL) with the REBUILD option and DATA COMPRESSION = COLUMNSTORE_ARCHIVE.

Reference: <https://learn.microsoft.com/en-us/sql/relational-databases/data-compression/data-compression>

NEW QUESTION 91

- (Exam Topic 3)

You are developing an application that uses Azure Data Lake Storage Gen 2.

You need to recommend a solution to grant permissions to a specific application for a limited time period. What should you include in the recommendation?

- A. Azure Active Directory (Azure AD) identities
- B. shared access signatures (SAS)
- C. account keys
- D. role assignments

Answer: B

Explanation:

A shared access signature (SAS) provides secure delegated access to resources in your storage account. With a SAS, you have granular control over how a client can access your data. For example:

What resources the client may access.

What permissions they have to those resources. How long the SAS is valid.

Reference:

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

NEW QUESTION 95

- (Exam Topic 3)

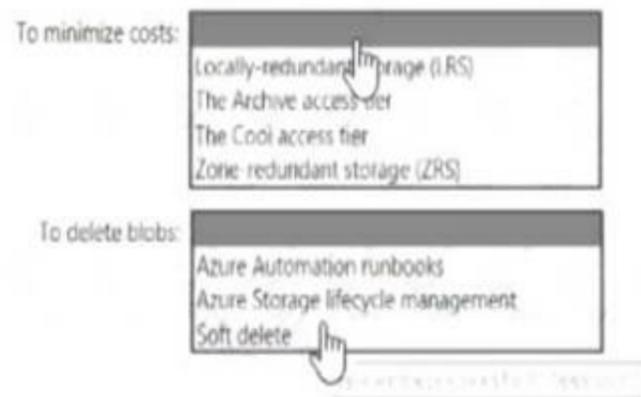
You have an Azure subscription.

You need to deploy an Azure Data Lake Storage Gen2 Premium account. The solution must meet the following requirements:

- Blobs that are older than 365 days must be deleted.
- Administrator efforts must be minimized.
- Costs must be minimized

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

Answer Area



- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

<https://learn.microsoft.com/en-us/azure/storage/blobs/premium-tier-for-data-lake-storage>

NEW QUESTION 100

- (Exam Topic 3)

You use PySpark in Azure Databricks to parse the following JSON input.

```
{
  "persons": [
    {
      "name": "Keith",
      "age": 30,
      "dogs": ["Fido", "Fluffy"]
    },
    {
      "name": "Donna",
      "age": 46,
      "dogs": ["Spot"]
    }
  ]
}
```

You need to output the data in the following tabular format.

owner	age	dog
Keith	30	Fido
Keith	30	Fluffy
Donna	46	Spot

How should you complete the PySpark code? 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

- alias
- array_union
- createDataFrame
- explode
- select
- translate

Answer Area

```

@utils.fs.put("/tmp/source.json", source_json, True)
source_df = spark.read.option("multiline", "true").json("/tmp/source.json")
persons = source_df.   ("persons").alias("persons")
persons_dogs = persons.select(col("persons.name").alias("owner"), col("persons.age").alias("age"),
explode  ("dog"))
("persons.dogs").
display(persons_dogs)
                    
```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Graphical user interface, text, application Description automatically generated

Box 1: select

Box 2: explode

Box 3: alias

pyspark.sql.Column.alias returns this column aliased with a new name or names (in the case of expressions that return more than one column, such as explode).

Reference: <https://spark.apache.org/docs/latest/api/python/reference/api/pyspark.sql.Column.alias.html> <https://docs.microsoft.com/en-us/azure/databricks/sql/language-manual/functions/explode>

NEW QUESTION 103

- (Exam Topic 3)

You have an Azure Synapse serverless SQL pool.

You need to read JSON documents from a file by using the OPENROWSET function.

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

Answer Area

```

SELECT *
FROM OPENROWSET
(
    BULK
    'https://sourcedatalake.blob.core.windows.net/public/docs.json',
    FORMAT = ,
    FIELDTERMINATOR = '0x0b',
    FIELDQUOTE = ,
    ROWTERMINATOR = '0x09',
)
WITH (jsondoc nvarchar(1000)) as jsonDocuments
                    
```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area

```

SELECT *
FROM OPENROWSET
(
    BULK
    'https://sourcedatalake.blob.core.windows.net/public/docs.json',
    FORMAT = 'JSON',
    FIELDTERMINATOR = '0x0b',
    FIELDQUOTE = '0x0b',
    ROWTERMINATOR = '0x0a',
)
WITH (jsondoc nvarchar(1000)) as jsonDocuments
    
```

NEW QUESTION 108

- (Exam Topic 3)

You have a data warehouse in Azure Synapse Analytics.

You need to ensure that the data in the data warehouse is encrypted at rest. What should you enable?

- A. Advanced Data Security for this database
- B. Transparent Data Encryption (TDE)
- C. Secure transfer required
- D. Dynamic Data Masking

Answer: B

Explanation:

Azure SQL Database currently supports encryption at rest for Microsoft-managed service side and client-side encryption scenarios.

- > Support for server encryption is currently provided through the SQL feature called Transparent Data Encryption.
- > Client-side encryption of Azure SQL Database data is supported through the Always Encrypted feature. Reference: <https://docs.microsoft.com/en-us/azure/security/fundamentals/encryption-atrest>

NEW QUESTION 109

- (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 113

- (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 117

- (Exam Topic 3)

You are designing a security model for an Azure Synapse Analytics dedicated SQL pool that will support multiple companies. You need to ensure that users from each company can view only the data of their respective company. Which two objects should you include in the solution? Each correct answer presents part of the solution

NOTE: Each correct selection it worth one point.

A. a custom role-based access control (RBAC) role.

B. asymmetric keys

C. a predicate function

D. a column encryption key

E. a security policy

Answer: AE

Explanation:

Reference:

<https://docs.microsoft.com/en-us/sql/relational-databases/security/row-level-security> <https://docs.microsoft.com/en-us/azure/synapse-analytics/security/synapse-workspace-access-control-overview>

NEW QUESTION 120

- (Exam Topic 3)

You have the following Azure Data Factory pipelines

- ingest Data from System 1
- Ingest Data from System2
- Populate Dimensions
- Populate facts

ingest Data from System1 and Ingest Data from System1 have no dependencies. Populate Dimensions must execute after Ingest Data from System1 and Ingest Data from System* Populate Facts must execute after the Populate Dimensions pipeline. All the pipelines must execute every eight hours.

What should you do to schedule the pipelines for execution?

A. Add an event trigger to all four pipelines.

B. Create a parent pipeline that contains the four pipelines and use an event trigger.

C. Create a parent pipeline that contains the four pipelines and use a schedule trigger.

D. Add a schedule trigger to all four pipelines.

Answer: C

Explanation:

Schedule trigger: A trigger that invokes a pipeline on a wall-clock schedule. Reference:

<https://docs.microsoft.com/en-us/azure/data-factory/concepts-pipeline-execution-triggers>

NEW QUESTION 125

- (Exam Topic 3)

You are designing an Azure Synapse Analytics workspace.

You need to recommend a solution to provide double encryption of all the data at rest.

Which two components should you include in the recommendation? Each coned answer presents part of the solution

NOTE: Each correct selection is worth one point.

A. an X509 certificate

B. an RSA key

C. an Azure key vault that has purge protection enabled

D. an Azure virtual network that has a network security group (NSG)

E. an Azure Policy initiative

Answer: BC

Explanation:

Synapse workspaces encryption uses existing keys or new keys generated in Azure Key Vault. A single key is used to encrypt all the data in a workspace.

Synapse workspaces support RSA 2048 and 3072 byte-sized keys, and RSA-HSM keys.

The Key Vault itself needs to have purge protection enabled. Reference:
<https://docs.microsoft.com/en-us/azure/synapse-analytics/security/workspaces-encryption>

NEW QUESTION 126

- (Exam Topic 3)
You have an Azure data factory.
You need to examine the pipeline failures from the last 180 flays. What should you use?

- A. the Activity tog blade for the Data Factory resource
- B. Azure Data Factory activity runs in Azure Monitor
- C. Pipeline runs in the Azure Data Factory user experience
- D. the Resource health blade for the Data Factory resource

Answer: B

Explanation:

Data Factory stores pipeline-run data for only 45 days. Use Azure Monitor if you want to keep that data for a longer time.
Reference:
<https://docs.microsoft.com/en-us/azure/data-factory/monitor-using-azure-monitor>

NEW QUESTION 127

- (Exam Topic 3)
You have an Apache Spark DataFrame named temperatures. A sample of the data is shown in the following table.

Date	Temp
...	...
18-01-2021	3
19-01-2021	4
20-01-2021	2
21-01-2021	2
...	...

You need to produce the following table by using a Spark SQL query.

Year	JAN	FEB	MAR	APR	MAY
2019	2.3	4.1	5.2	7.6	9.2
2020	2.4	4.2	4.9	7.8	9.1
2021	2.6	5.3	3.4	7.9	9.5

How should you complete the query? 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

- CAST
- COLLATE
- CONVERT
- FLATTEN
- PIVOT
- UNPIVOT

```

SELECT * FROM (
  SELECT YEAR(Date) Year, MONTH(Date) Month, Temp
  FROM temperatures
  WHERE date BETWEEN DATE '2019-01-01' AND DATE '2021-08-31'
)
[ ] (
  AVG ( [ ] (Temp AS DECIMAL(4, 1)))
  FOR Month in (
    1 JAN, 2 FEB, 3 MAR, 4 APR, 5 MAY, 6 JUN,
    7 JUL, 8 AUG, 9 SEP, 10 OCT, 11 NOV, 12 DEC
  )
)
ORDER BY Year ASC
    
```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Text Description automatically generated
Box 1: PIVOT
PIVOT rotates a table-valued expression by turning the unique values from one column in the expression into multiple columns in the output. And PIVOT runs aggregations where they're required on any remaining column values that are wanted in the final output.

Reference:

<https://learnsql.com/cookbook/how-to-convert-an-integer-to-a-decimal-in-sql-server/> <https://docs.microsoft.com/en-us/sql/t-sql/queries/from-using-pivot-and-unpivot>

NEW QUESTION 129

- (Exam Topic 3)

You are implementing Azure Stream Analytics windowing functions.

Which windowing function should you use for each requirement? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

Segment the data stream into distinct time segments that repeat but do not overlap:

Hopping
Sliding
Tumbling

Segment the data stream into distinct time segments that repeat and can overlap:

Hopping
Sliding
Tumbling

Segment the data stream to produce an output only when an event occurs:

Hopping
Sliding
Tumbling

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area

Segment the data stream into distinct time segments that repeat but do not overlap:

Hopping
Sliding
Tumbling

Segment the data stream into distinct time segments that repeat and can overlap:

Hopping
Sliding
Tumbling

Segment the data stream to produce an output only when an event occurs:

Hopping
Sliding
Tumbling

NEW QUESTION 131

- (Exam Topic 3)

You create an Azure Databricks cluster and specify an additional library to install. When you attempt to load the library to a notebook, the library is not found. You need to identify the cause of the issue. What should you review?

- A. notebook logs
- B. cluster event logs
- C. global init scripts logs
- D. workspace logs

Answer: C

Explanation:

Cluster-scoped Init Scripts: Init scripts are shell scripts that run during the startup of each cluster node before the Spark driver or worker JVM starts. Databricks customers use init scripts for various purposes such as installing custom libraries, launching background processes, or applying enterprise security policies. Logs for Cluster-scoped init scripts are now more consistent with Cluster Log Delivery and can be found in the same root folder as driver and executor logs for the cluster.

Reference:

<https://databricks.com/blog/2018/08/30/introducing-cluster-scoped-init-scripts.html>

NEW QUESTION 136

- (Exam Topic 3)

You have a SQL pool in Azure Synapse that contains a table named dbo.Customers. The table contains a column name Email.

You need to prevent nonadministrative users from seeing the full email addresses in the Email column. The users must see values in a format of aXXX@XXXX.com instead.

What should you do?

- A. From Microsoft SQL Server Management Studio, set an email mask on the Email column.

- B. From the Azure portal, set a mask on the Email column.
- C. From Microsoft SQL Server Management studio, grant the SELECT permission to the users for all the columns in the dbo.Customers table except Email.
- D. From the Azure portal, set a sensitivity classification of Confidential for the Email column.

Answer: D

Explanation:

From Microsoft SQL Server Management Studio, set an email mask on the Email column. This is because "This feature cannot be set using portal for Azure Synapse (use PowerShell or REST API) or SQL Managed Instance." So use Create table statement with Masking e.g. CREATE TABLE Membership (MemberID int IDENTITY PRIMARY KEY, FirstName varchar(100) MASKED WITH (FUNCTION = 'partial(1,"XXXXXXX",0)') NULL, . . .
<https://docs.microsoft.com/en-us/azure/azure-sql/database/dynamic-data-masking-overview>
upvoted 24 times

NEW QUESTION 138

- (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 139

- (Exam Topic 3)

You have an Azure Data Factory pipeline that is triggered hourly. The pipeline has had 100% success for the past seven days. The pipeline execution fails, and two retries that occur 15 minutes apart also fail. The third failure returns the following error.

```
ErrorCode=UserErrorFileNotFound,Type=Microsoft.DataTransfer.Common.Shared.HybridDeliveryException,Message=ADLS Gen2 operation failed for:
Operation returned an invalid status code 'NotFound'. Account: 'contosoproduksouth'. FileSystem: wwi. Path:
'BIKES/CARBON/year=2021/month=01/day=10/hour=06'. ErrorCode: 'PathNotFound'. Message: 'The specified path does not exist.'. RequestId: '6d269b78-
901f-001b-4924-e7a7bc000000'. TimeStamp: 'Sun, 10 Jan 2021 07:45:05'
```

What is a possible cause of the error?

- A. The parameter used to generate year=2021/month=01/day=10/hour=06 was incorrect.
- B. From 06:00 to 07:00 on January 10, 2021, there was no data in wwi/BIKES/CARBON.
- C. From 06:00 to 07:00 on January 10, 2021, the file format of data in wwi/BIKES/CARBON was incorrect.
- D. The pipeline was triggered too early.

Answer: C

NEW QUESTION 144

- (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 149

- (Exam Topic 3)

You have an Azure Stream Analytics job that is a Stream Analytics project solution in Microsoft Visual Studio. The job accepts data generated by IoT devices in the JSON format.

You need to modify the job to accept data generated by the IoT devices in the Protobuf format.

Which three actions should you perform from Visual Studio on sequence? 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

- Change the Event Serialization Format to Protobuf in the input.json file of the job and reference the DLL.
- Add an Azure Stream Analytics Custom Deserializer Project (.NET) project to the solution.
- Add .NET deserializer code for Protobuf to the custom deserializer project.
- Add .NET deserializer code for Protobuf to the Stream Analytics project.
- Add an Azure Stream Analytics Application project to the solution.

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

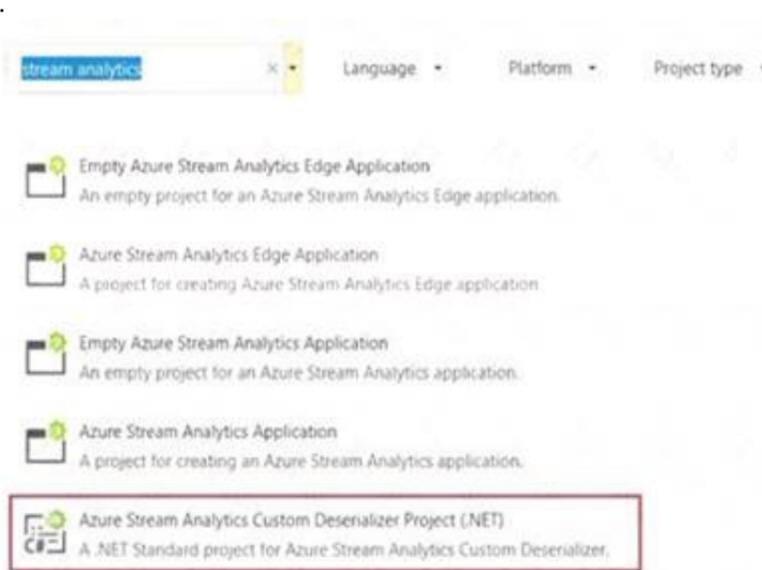
Step 1: Add an Azure Stream Analytics Custom Deserializer Project (.NET) project to the solution. Create a custom deserializer

* 1. Open Visual Studio and select File > New > Project. Search for Stream Analytics and select Azure Stream Analytics Custom Deserializer Project (.NET). Give the project a name, like Protobuf Deserializer.

Create a new project

Recent project templates

A list of your recently accessed templates will be displayed here.



* 2. In Solution Explorer, right-click your Protobuf Deserializer project and select Manage NuGet Packages from the menu. Then install the Microsoft.Azure.StreamAnalytics and Google.Protobuf NuGet packages.

* 3. Add the MessageBodyProto class and the MessageBodyDeserializer class to your project.

* 4. Build the Protobuf Deserializer project.

Step 2: Add .NET deserializer code for Protobuf to the custom deserializer project

Azure Stream Analytics has built-in support for three data formats: JSON, CSV, and Avro. With custom .NET deserializers, you can read data from other formats such as Protocol Buffer, Bond and other user defined formats for both cloud and edge jobs.

Step 3: Add an Azure Stream Analytics Application project to the solution Add an Azure Stream Analytics project

> In Solution Explorer, right-click the Protobuf Deserializer solution and select Add > New Project. Under Azure Stream Analytics > Stream Analytics, choose Azure Stream Analytics Application. Name it ProtobufCloudDeserializer and select OK.

> Right-click References under the ProtobufCloudDeserializer Azure Stream Analytics project. Under Projects, add Protobuf Deserializer. It should be automatically populated for you.

Reference:

<https://docs.microsoft.com/en-us/azure/stream-analytics/custom-deserializer>

NEW QUESTION 151

- (Exam Topic 3)

You are designing an enterprise data warehouse in Azure Synapse Analytics that will store website traffic analytics in a star schema.

You plan to have a fact table for website visits. The table will be approximately 5 GB.

You need to recommend which distribution type and index type to use for the table. The solution must provide the fastest query performance.

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

NOTE: Each correct selection is worth one point.

Distribution:

Hash
Round robin
Replicated

Index:

Clustered columnstore
Clustered
Nonclustered

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Hash

Consider using a hash-distributed table when: The table size on disk is more than 2 GB.

The table has frequent insert, update, and delete operations. Box 2: Clustered columnstore

Clustered columnstore tables offer both the highest level of data compression and the best overall query performance.

Reference:

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

NEW QUESTION 152

- (Exam Topic 3)

You are building a database in an Azure Synapse Analytics serverless SQL pool. You have data stored in Parquet files in an Azure Data Lake Storage Gen2 container. Records are structured as shown in the following sample.

```
{
  "id": 123,
  "address_housenumber": "19c", "address_line": "Memory Lane", "applicant1_name": "Jane", "applicant2_name": "Dev"
}
```

The records contain two applicants at most.

You need to build a table that includes only the address fields.

How should you complete the Transact-SQL statement? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

CREATE EXTERNAL TABLE
CREATE TABLE
CREATE VIEW

```
WITH (
  LOCATION = 'applications/',
  DATA_SOURCE = applications_ds,
  FILE_FORMAT = applications_file_format
)
AS
SELECT id, [address_housenumber] as addresshousenumber, [address_line1] as addressline1
FROM
  (BULK 'https://contosol.dfs.core.windows.net/applications/year=*/*.parquet',
  CROSS APPLY
  OPENJSON
  OPENROWSET
  FORMAT='PARQUET') AS [r]
GO
```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: CREATE EXTERNAL TABLE

An external table points to data located in Hadoop, Azure Storage blob, or Azure Data Lake Storage. External tables are used to read data from files or write data to files in Azure Storage. With Synapse SQL, you can use external tables to read external data using dedicated SQL pool or serverless SQL pool.

Syntax:
 CREATE EXTERNAL TABLE { database_name.schema_name.table_name | schema_name.table_name | table_name }
 (<column_definition> [,...n]) WITH (
 LOCATION = 'folder_or_filepath', DATA_SOURCE = external_data_source_name, FILE_FORMAT = external_file_format_name
 Box 2. OPENROWSET
 When using serverless SQL pool, CETAS is used to create an external table and export query results to Azure Storage Blob or Azure Data Lake Storage Gen2.
 Example: AS
 SELECT decennialTime, stateName, SUM(population) AS population FROM
 OPENROWSET(BULK
 'https://azureopendatastorage.blob.core.windows.net/censusdatacontainer/release/us_population_county/year=*/
 FORMAT='PARQUET') AS [r]
 GROUP BY decennialTime, stateName GO
 Reference:
<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/develop-tables-external-tables>

NEW QUESTION 154

- (Exam Topic 3)
 You are designing database for an Azure Synapse Analytics dedicated SQL pool to support workloads for detecting ecommerce transaction fraud. Data will be combined from multiple ecommerce sites and can include sensitive financial information such as credit card numbers. You need to recommend a solution that meets the following requirements:

- > Users must be able to identify potentially fraudulent transactions.
- > Users must be able to use credit cards as a potential feature in models.
- > Users must NOT be able to access the actual credit card numbers.

What should you include in the recommendation?

- A. Transparent Data Encryption (TDE)
- B. row-level security (RLS)
- C. column-level encryption
- D. Azure Active Directory (Azure AD) pass-through authentication

Answer: C

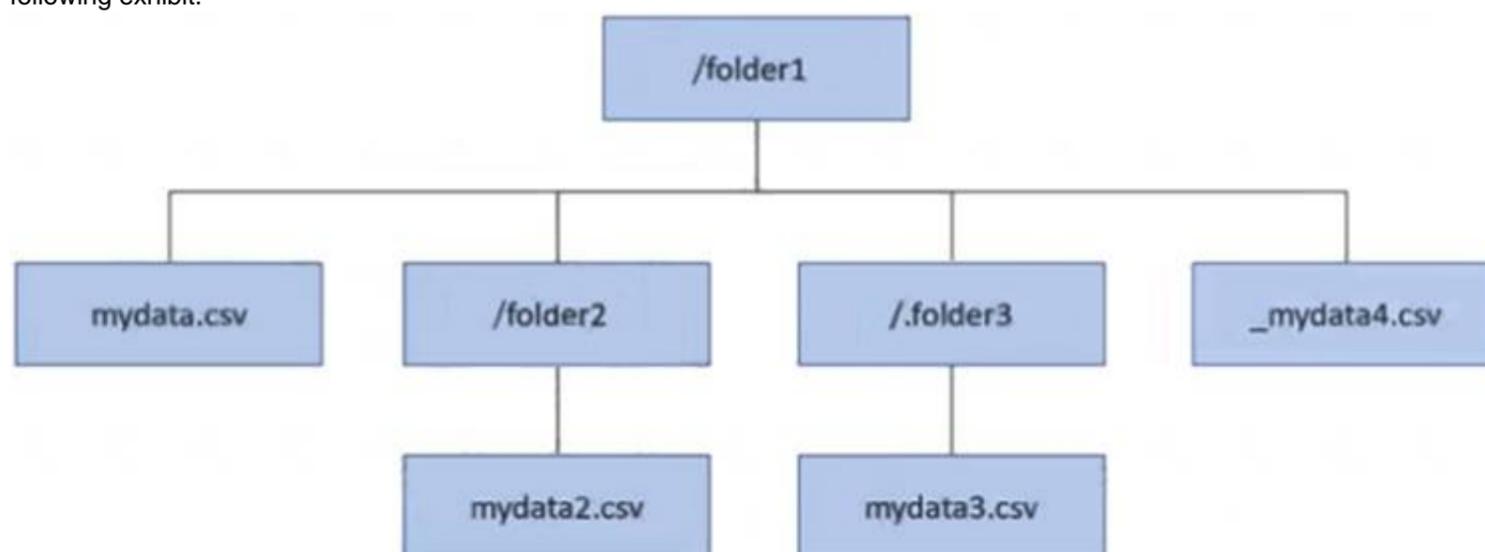
Explanation:

Use Always Encrypted to secure the required columns. You can configure Always Encrypted for individual database columns containing your sensitive data. Always Encrypted is a feature designed to protect sensitive data, such as credit card numbers or national identification numbers (for example, U.S. social security numbers), stored in Azure SQL Database or SQL Server databases.

Reference:
<https://docs.microsoft.com/en-us/sql/relational-databases/security/encryption/always-encrypted-database-engine>

NEW QUESTION 156

- (Exam Topic 3)
 You have an Azure Data Lake Storage Gen2 account that contains a container named container1. You have an Azure Synapse Analytics serverless SQL pool that contains a native external table named dbo.Table1. The source data for dbo.Table1 is stored in container1. The folder structure of container1 is shown in the following exhibit.



The external data source is defined by using the following statement.

```

CREATE EXTERNAL DATA SOURCE DataLake
WITH
(
    LOCATION = 'https://mydatalake.dfs.core.windows.net/container1/folder1/**'
    , CREDENTIAL = DataLakeCred
);
    
```

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
When selecting all the rows in dbo.Table1, data from the mydata2.csv file will be returned.	<input type="radio"/>	<input type="radio"/>
When selecting all the rows in dbo.Table1, data from the mydata3.csv file will be returned.	<input type="radio"/>	<input type="radio"/>
When selecting all the rows in dbo.Table1, data from the _mydata4.csv file will be returned.	<input type="radio"/>	<input type="radio"/>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Yes

In the serverless SQL pool you can also use recursive wildcards /logs/** to reference Parquet or CSV files in any sub-folder beneath the referenced folder.

Box 2: Yes

Box 3: No

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

NEW QUESTION 159

- (Exam Topic 3)

You have an Azure Data Lake Storage account that has a virtual network service endpoint configured.

You plan to use Azure Data Factory to extract data from the Data Lake Storage account. The data will then be loaded to a data warehouse in Azure Synapse Analytics by using PolyBase.

Which authentication method should you use to access Data Lake Storage?

- A. shared access key authentication
- B. managed identity authentication
- C. account key authentication
- D. service principal authentication

Answer: B

Explanation:

Reference:

<https://docs.microsoft.com/en-us/azure/data-factory/connector-azure-sql-data-warehouse#use-polybase-to-load-d>

NEW QUESTION 161

- (Exam Topic 3)

You are building a data flow in Azure Data Factory that upserts data into a table in an Azure Synapse Analytics dedicated SQL pool.

You need to add a transformation to the data flow. The transformation must specify logic indicating when a row from the input data must be upserted into the sink.

Which type of transformation should you add to the data flow?

- A. join
- B. select
- C. surrogate key
- D. alter row

Answer: D

Explanation:

The alter row transformation allows you to specify insert, update, delete, and upsert policies on rows based on expressions. You can use the alter row transformation to perform upserts on a sink table by matching on a key column and setting the appropriate row policy

NEW QUESTION 164

- (Exam Topic 3)

You have an enterprise data warehouse in Azure Synapse Analytics named DW1 on a server named Server1. You need to verify whether the size of the transaction log file for each distribution of DW1 is smaller than 160 GB.

What should you do?

- A. On the master database, execute a query against the sys.dm_pdw_nodes_os_performance_counters dynamic management view.
- B. From Azure Monitor in the Azure portal, execute a query against the logs of DW1.
- C. On DW1, execute a query against the sys.database_files dynamic management view.
- D. Execute a query against the logs of DW1 by using the Get-AzOperationalInsightSearchResult PowerShell cmdlet.

Answer: A

Explanation:

The following query returns the transaction log size on each distribution. If one of the log files is reaching 160 GB, you should consider scaling up your instance or limiting your transaction size.

```
-- Transaction log size SELECT
```

```
instance_name as distribution_db, cntr_value*1.0/1048576 as log_file_size_used_GB, pdw_node_id
```

FROM sys.dm_pdw_nodes_os_performance_counters WHERE
instance_name like 'Distribution_%'

AND counter_name = 'Log File(s) Used Size (KB)'

References:

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

NEW QUESTION 165

- (Exam Topic 3)

You have an Azure Synapse Analytics dedicated SQL pool named Pool1 and a database named DB1. DB1 contains a fact table named Table1. You need to identify the extent of the data skew in Table1. What should you do in Synapse Studio?

- A. Connect to the built-in pool and run dbcc pdw_showspaceused.
- B. Connect to the built-in pool and run dbcc checkalloc.
- C. Connect to Pool1 and query sys.dm_pdw_node_scacus.
- D. Connect to Pool1 and query sys.dm_pdw_nodes_db_partition_scacs.

Answer: A

Explanation:

A quick way to check for data skew is to use DBCC PDW_SHOWSPACEUSED. The following SQL code returns the number of table rows that are stored in each of the 60 distributions. For balanced performance, the rows in your distributed table should be spread evenly across all the distributions.

DBCC PDW_SHOWSPACEUSED('dbo.FactInternetSales'); Reference:

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

NEW QUESTION 170

- (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 171

- (Exam Topic 3)

You are designing a real-time dashboard solution that will visualize streaming data from remote sensors that connect to the internet. The streaming data must be aggregated to show the average value of each 10-second interval. The data will be discarded after being displayed in the dashboard.

The solution will use Azure Stream Analytics and must meet the following requirements:

- > Minimize latency from an Azure Event hub to the dashboard.
- > Minimize the required storage.
- > Minimize development effort.

What should you include in the solution? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point

Azure Stream Analytics input type:

	▼
Azure Event Hub	
Azure SQL Database	
Azure Stream Analytics	
Microsoft Power BI	

Azure Stream Analytics output type:

	▼
Azure Event Hub	
Azure SQL Database	
Azure Stream Analytics	
Microsoft Power BI	

Aggregation query location:

	▼
Azure Event Hub	
Azure SQL Database	
Azure Stream Analytics	
Microsoft Power BI	

- A. Mastered
- B. Not Mastered

Answer: A

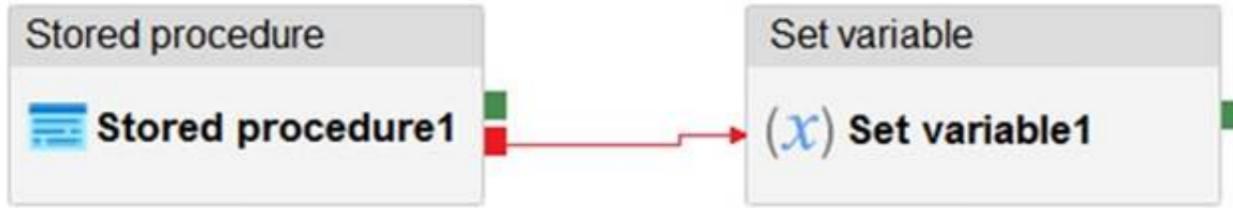
Explanation:

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

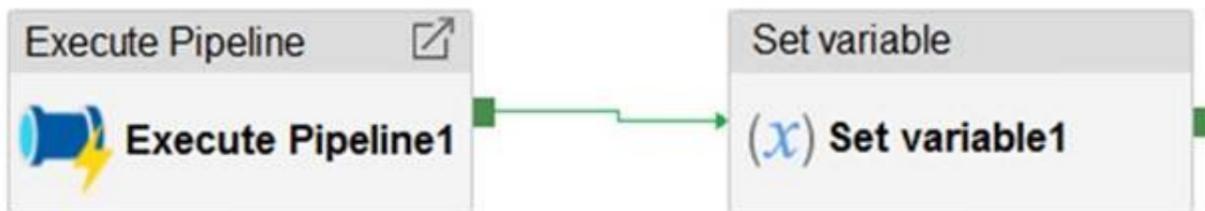
NEW QUESTION 174

- (Exam Topic 3)

You have an Azure Data Factory instance that contains two pipelines named Pipeline1 and Pipeline2. Pipeline1 has the activities shown in the following exhibit.



Pipeline2 has the activities shown in the following exhibit.



You execute Pipeline2, and Stored procedure1 in Pipeline1 fails. What is the status of the pipeline runs?

- A. Pipeline1 and Pipeline2 succeeded.
- B. Pipeline1 and Pipeline2 failed.
- C. Pipeline1 succeeded and Pipeline2 failed.
- D. Pipeline1 failed and Pipeline2 succeeded.

Answer: A

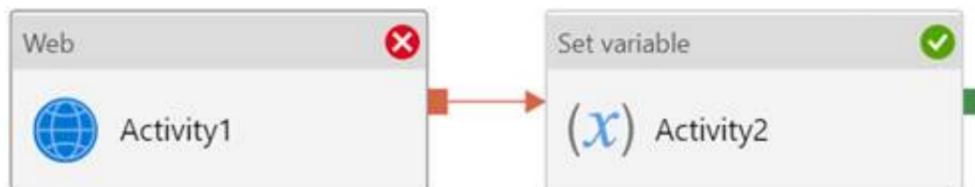
Explanation:

Activities are linked together via dependencies. A dependency has a condition of one of the following: Succeeded, Failed, Skipped, or Completed.

Consider Pipeline1:

If we have a pipeline with two activities where Activity2 has a failure dependency on Activity1, the pipeline will not fail just because Activity1 failed. If Activity1 fails and Activity2 succeeds, the pipeline will succeed. This scenario is treated as a try-catch block by Data Factory.

Waterfall chart Description automatically generated with medium confidence



The failure dependency means this pipeline reports success. Note:

If we have a pipeline containing Activity1 and Activity2, and Activity2 has a success dependency on Activity1, it will only execute if Activity1 is successful. In this scenario, if Activity1 fails, the pipeline will fail.

Reference:

<https://datasavvy.me/category/azure-data-factory/>

NEW QUESTION 177

.....

Thank You for Trying Our Product

* 100% Pass or Money Back

All our products come with a 90-day Money Back Guarantee.

* One year free update

You can enjoy free update one year. 24x7 online support.

* Trusted by Millions

We currently serve more than 30,000,000 customers.

* Shop Securely

All transactions are protected by VeriSign!

100% Pass Your DP-203 Exam with Our Prep Materials Via below:

<https://www.certleader.com/DP-203-dumps.html>