[NEW] Practice Tests | DP-203: Azure Data Engineer 2022

udemy.com/course/microsoft-azure-data-engineer-exam-practice-tests/learn/quiz/5461770/result/720230622

Practice Test 1: DP-203: Microsoft Azure Data Engineer Exam - Resultados

Tentativa 1

Pergunta 1: Incorreto
You plan to create an Azure Synapse Analytics dedicated SQL pool.

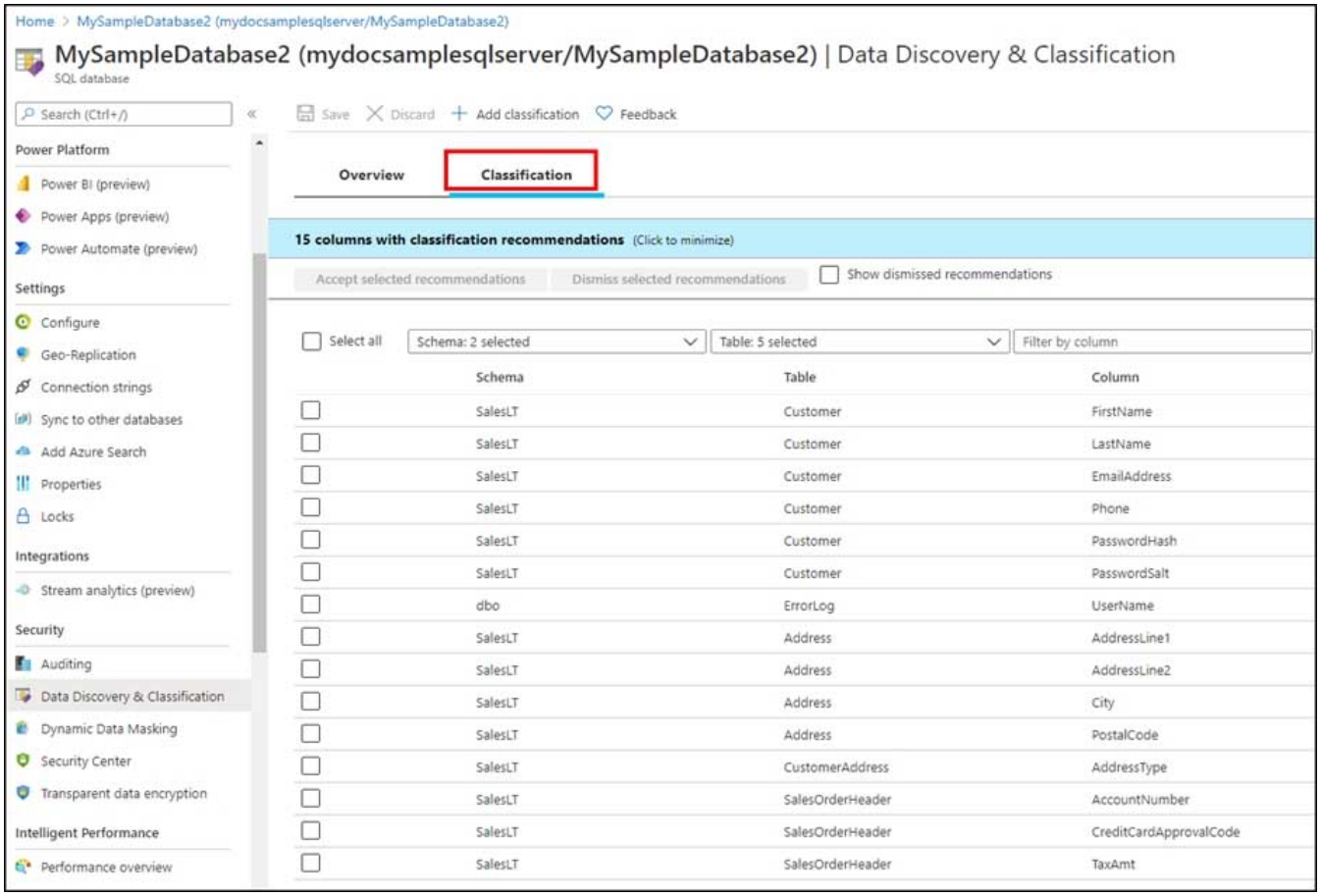
You need to minimize the time it takes to identify queries that return confidential information as defined by the company's data privacy regulations and the users who executed the queues.

Which two components should you include in the solution? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

Explicação

Correct Answer: *AC*
A: You can classify columns manually, as an alternative or in addition to the recommendation-based classification:



1. Select Add classification in the top menu of the pane.
2. In the context window that opens, select the schema, table, and column that you want to classify, and the information type and sensitivity label.
3. Select Add classification at the bottom of the context window.
C: An important aspect of the information-protection paradigm is the ability to monitor access to sensitive data. Azure SQL Auditing has been enhanced to include a new field in the audit log called data_sensitivity_information. This field logs the sensitivity classifications (labels) of the data that was returned by a query. Here's an example:

d	client_ip	application_name	duration_milliseconds	response_rows	affected_rows	connection_id	data_sensitivity_information
	[redacted]7.125	Microsoft SQL Server Management Studio - Query	1	847	847	C244A066-2271-...	Confidential - GDPR
	[redacted]7.125	Microsoft SQL Server Management Studio - Query	2	32	32	C244A066-2271-...	Confidential
	[redacted]7.125	Microsoft SQL Server Management Studio - Query	41	32	32	A7088FD4-759E-...	Confidential, Confidential - GDPR

Reference:

https://docs.microsoft.com/en-us/azure/azure-sql/database/data-discovery-and-classification-overview

Pergunta 2: Incorreto
You have an Azure Data Factory version 2 (V2) resource named Df1. Df1 contains a linked service.

You have an Azure Key vault named vault1 that contains an encryption key named key1.

You need to encrypt Df1 by using key1.

What should you do first?

Explicação

Correct Answer: *C*

Linked services are much like connection strings, which define the connection information needed for Data Factory to connect to external resources.

Incorrect Answers:

D: A self-hosted integration runtime copies data between an on-premises store and cloud storage.

Reference:

https://docs.microsoft.com/en-us/azure/data-factory/enable-customer-managed-key

https://docs.microsoft.com/en-us/azure/data-factory/concepts-linked-services

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

Pergunta 3: Incorreto
You develop data engineering solutions for a company.

A project requires the deployment of data to Azure Data Lake Storage.

You need to implement role-based access control (RBAC) so that project members can manage the Azure Data Lake Storage resources.

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

NOTE: Each correct selection is worth one point.

Explicação

Correct Answer: ACE

AC: Create security groups in Azure Active Directory. Assign users or security groups to Data Lake Storage Gen1 accounts.

E: Assign users or security groups as ACLs to the Data Lake Storage Gen1 file system

Reference:

https://docs.microsoft.com/en-us/azure/data-lake-store/data-lake-store-secure-data

Pergunta 4: Incorreto
You are designing an Azure Synapse Analytics dedicated SQL pool.

You need to ensure that you can audit access to Personally Identifiable Information (PII).

What should you include in the solution?

Explicação

Correct Answer: *D*
Data Discovery & Classification is built into Azure SQL Database, Azure SQL Managed Instance, and Azure Synapse Analytics. It provides basic capabilities for discovering, classifying, labeling, and reporting the sensitive data in your databases.
Your most sensitive data might include business, financial, healthcare, or personal information. Discovering and classifying this data can play a pivotal role in your organization's information-protection approach. It can serve as infrastructure for:
• Helping to meet standards for data privacy and requirements for regulatory compliance.
• Various security scenarios, such as monitoring (auditing) access to sensitive data.
• Controlling access to and hardening the security of databases that contain highly sensitive data.

Reference:

https://docs.microsoft.com/en-us/azure/azure-sql/database/data-discovery-and-classification-overview

Pergunta 5: Incorreto
You need to trigger an Azure Data Factory pipeline when a file arrives in an Azure Data Lake Storage Gen2 container.

Which resource provider should you enable?

Explicação

Correct Answer: C

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

Blob Storage account. Data Factory natively integrates with Azure Event Grid, which lets you trigger pipelines on such events.

Reference:

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

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

Pergunta 6: Incorreto
You plan to perform batch processing in Azure Databricks once daily.

Which type of Databricks cluster should you use?

Explicação

Correct Answer: *B*
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-batch-workloads-data-engineers-running-etl-jobs

Pergunta 7: Incorreto
You are processing streaming data from vehicles that pass through a toll booth.
You need to use Azure Stream Analytics to return the license plate, vehicle make, and hour the last vehicle passed during each 10-minute window.
How should you complete the query? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.
Hot Area:

Answer Area

```
WITH LastInWindow AS
(
        SELECT
                [ COUNT | MAX | MIN | TOPONE ] (Time) AS LastEventTime
        FROM
                Input TIMESTAMP BY Time
        GROUP BY
                [ HoppingWindow | SessionWindow | SlidingWindow | TumblingWindow ] (minute, 10)
)
SELECT
        Input.License_plate,
        Input.Make,
        Input.Time
FROM
        Input TIMESTAMP BY Time
        INNER JOIN LastInWindow
        ON [ DATEADD | DATEDIFF | DATENAME | DATEPART ] (minute, Input, LastInWindow) BETWEEN 0 AND 10
AND Input.Time = LastInWindow.LastEventTime
```

Explicação

Correct Answer:

Answer Area

```
WITH LastInWindow AS
(
        SELECT
                [ COUNT | MAX | MIN | TOPONE ] (Time) AS LastEventTime
        FROM
                Input TIMESTAMP BY Time
        GROUP BY
                [ HoppingWindow | SessionWindow | SlidingWindow | TumblingWindow ] (minute, 10)
)
SELECT
        Input.License_plate,
        Input.Make,
        Input.Time
FROM
        Input TIMESTAMP BY Time
        INNER JOIN LastInWindow
        ON [ DATEADD | DATEDIFF | DATENAME | DATEPART ] (minute, Input, LastInWindow) BETWEEN 0 AND 10
AND Input.Time = LastInWindow.LastEventTime
```

Box 1: MAX -
The first step on the query finds the maximum time stamp in 10-minute windows, that is the time stamp of the last event for that window. The second step joins the results of the first query with the original stream to find the event that match the last time stamps in each window.
Query:

WITH LastInWindow AS -
(

SELECT -

MAX(Time) AS LastEventTime -

FROM -

Input TIMESTAMP BY Time -

GROUP BY -
TumblingWindow(minute, 10)
)

SELECT -
Input.License_plate,
Input.Make,

Input.Time -

FROM -

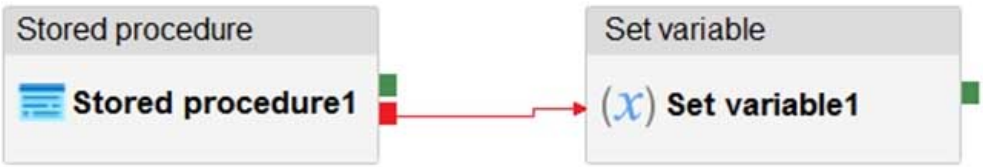
Input TIMESTAMP BY Time -

INNER JOIN LastInWindow -
ON DATEDIFF(minute, Input, LastInWindow) BETWEEN 0 AND 10
AND Input.Time = LastInWindow.LastEventTime

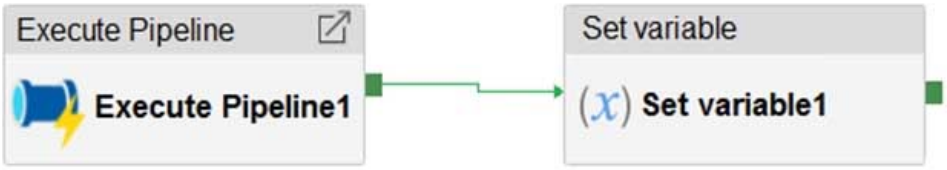
Box 2: TumblingWindow -
Tumbling windows are a series of fixed-sized, non-overlapping and contiguous time intervals.

Box 3: DATEDIFF -
DATEDIFF is a date-specific function that compares and returns the time difference between two DateTime fields, for more information, refer to date functions.
Reference:
https://docs.microsoft.com/en-us/stream-analytics-query/tumbling-window-azure-stream-analytics

Pergunta 8: Incorreto
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?

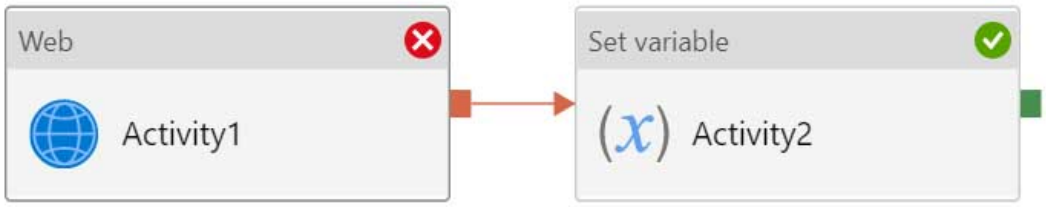
Explicação

Correct Answer: *A*

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.



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/

Pergunta 9: Correto

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.

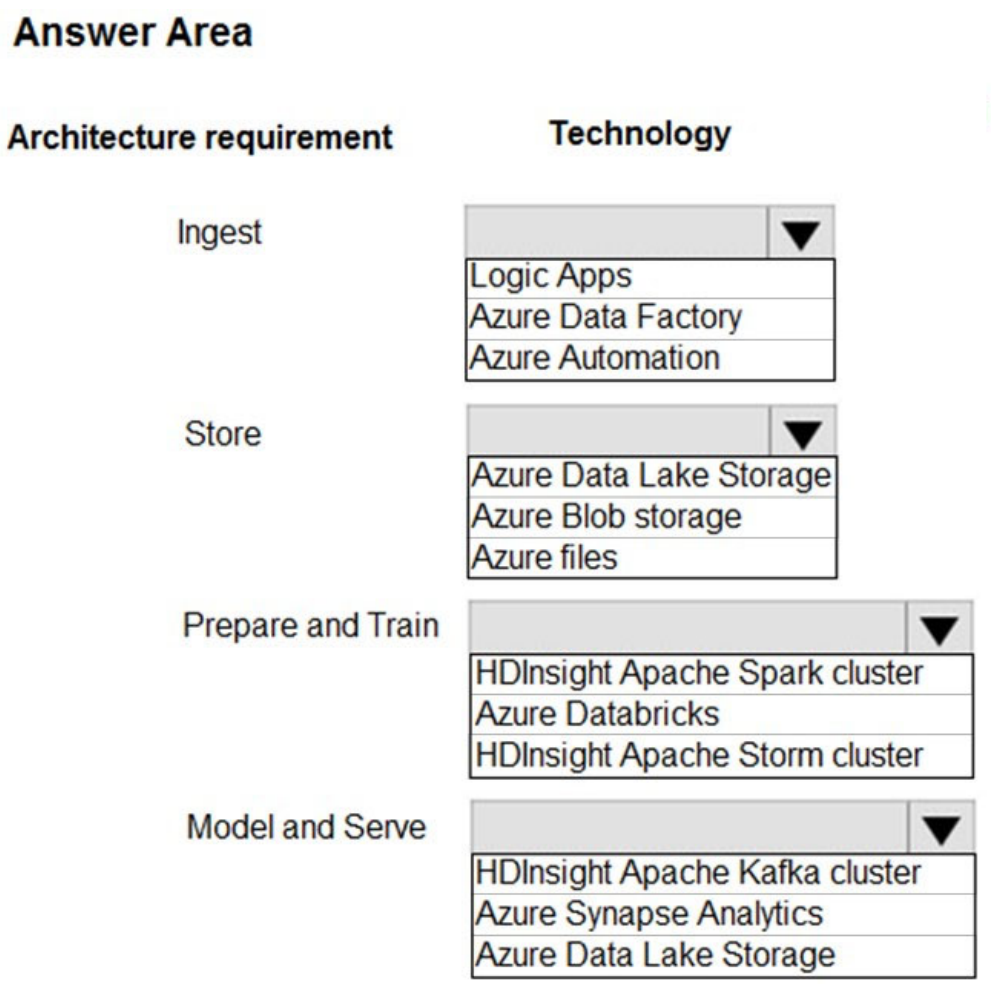
Provide seamless user authentication with Azure Active Directory.

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.

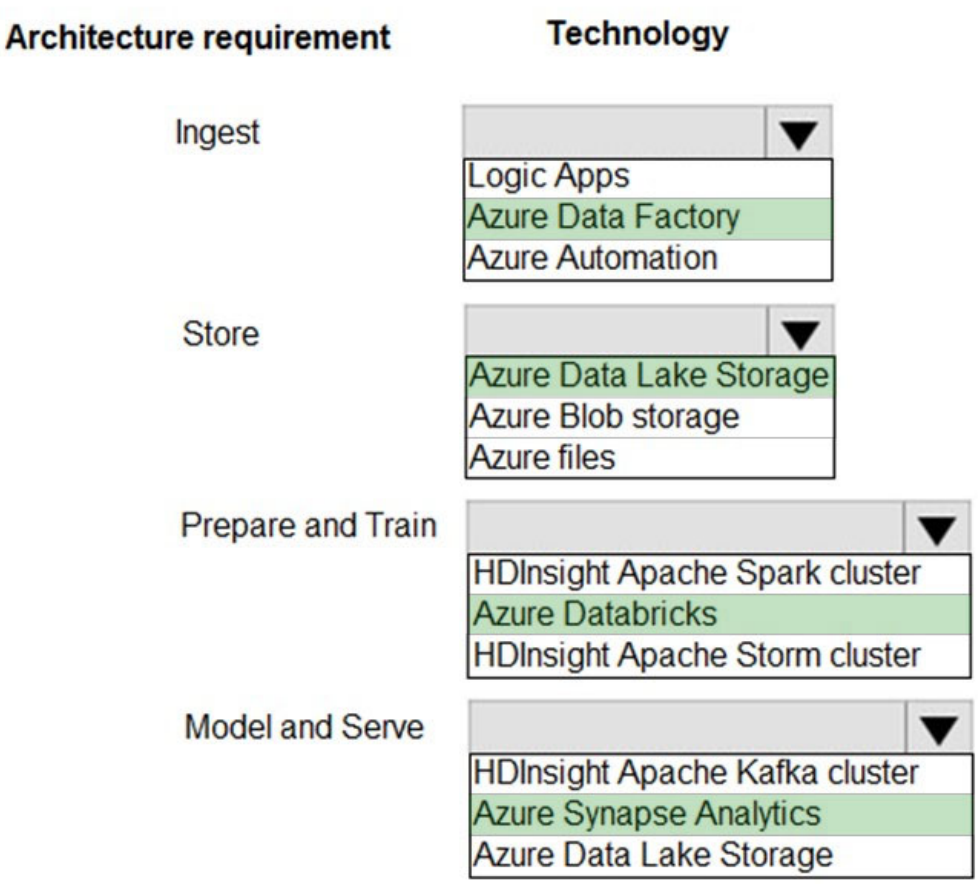
NOTE: Each correct selection is worth one point.
Hot Area:



Explicação

Correct Answer:

Answer Area



Ingest: Azure Data Factory -
Azure Data Factory pipelines can execute SSIS packages.
In Azure, the following services and tools will meet the core requirements for pipeline orchestration, control flow, and data movement: Azure Data Factory, Oozie on HDInsight, and SQL Server Integration Services (SSIS).

Store: Data Lake Storage -
Data Lake Storage Gen1 provides unlimited storage.
Note: Data at rest includes information that resides in persistent storage on physical media, in any digital format. Microsoft Azure offers a variety of data storage solutions to meet different needs, including file, disk, blob, and table storage.
Microsoft also provides encryption to protect Azure SQL Database, Azure Cosmos
DB, and Azure Data Lake.

Prepare and Train: Azure Databricks
Azure Databricks provides enterprise-grade Azure security, including Azure Active Directory integration.
With Azure Databricks, you can set up your Apache Spark environment in minutes, autoscale and collaborate on shared projects in an interactive workspace.
Azure Databricks supports Python, Scala, R, Java and SQL, as well as data science frameworks and libraries including TensorFlow, PyTorch and scikit-learn.

Model and Serve: Azure Synapse Analytics
Azure Synapse Analytics/ SQL Data Warehouse stores data into relational tables with columnar storage.
Azure SQL Data Warehouse connector now offers efficient and scalable structured streaming write support for SQL Data Warehouse. Access SQL Data
Warehouse from Azure Databricks using the SQL Data Warehouse connector.
Note: As of November 2019, Azure SQL Data Warehouse is now Azure Synapse Analytics.

Reference:
https://docs.microsoft.com/bs-latn-ba/azure/architecture/data-guide/technology-choices/pipeline-orchestration-data-movement

https://docs.microsoft.com/en-us/azure/azure-databricks/what-is-azure-databricks

Pergunta 10: Incorreto
You have an Azure Data Factory that contains 10 pipelines.

You need to label each pipeline with its main purpose of either ingest, transform, or load. The labels must be available for grouping and filtering when using the monitoring experience in Data Factory.

What should you add to each pipeline?

Explicação

Correct Answer: *D*

Annotations are additional, informative tags that you can add to specific factory resources: pipelines, datasets, linked services, and triggers. By adding annotations, you can easily filter and search for specific factory resources.

Reference:

https://www.cathrinewilhelmsen.net/annotations-user-properties-azure-data-factory/

Pergunta 11: Incorreto
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.

Explicação

Correct Answer: *BE*

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.

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/

Pergunta 12: Correto
You have an Azure Storage account and a data warehouse in Azure Synapse Analytics in the UK South region.

You need to copy blob data from the storage account to the data warehouse by using Azure Data Factory. The solution must meet the following requirements:

• Ensure that the data remains in the UK South region at all times.

• Minimize administrative effort.

Which type of integration runtime should you use?

Explicação

Correct Answer: *A*

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

Incorrect Answers:
C: Self-hosted integration runtime is to be used On-premises.
Reference:
https://docs.microsoft.com/en-us/azure/data-factory/concepts-integration-runtime

Pergunta 13: Correto
You have an Azure Stream Analytics job that receives clickstream data from an Azure event hub.

You need to define a query in the Stream Analytics job. The query must meet the following requirements:

• Count the number of clicks within each 10-second window based on the country of a visitor.

• Ensure that each click is NOT counted more than once.

How should you define the Query?

Explicação

Correct Answer: *B*

Tumbling window functions are used to segment a data stream into distinct time segments and perform a function against them, such as the example below. The key differentiators of a Tumbling window are that they repeat, do not overlap, and an event cannot belong to more than one tumbling window.

Example:

Incorrect Answers:

A: Sliding windows, unlike Tumbling or Hopping windows, output events only for points in time when the content of the window actually changes. In other words, when an event enters or exits the window. Every window has at least one event, like in the case of Hopping windows, events can belong to more than one sliding window.

C: Hopping window functions hop forward in time by a fixed period. It may be easy to think of them as Tumbling windows that can overlap, so events can belong to more than one Hopping window result set. To make a Hopping window the same as a Tumbling window, specify the hop size to be the same as the window size.

D: Session windows group events that arrive at similar times, filtering out periods of time where there is no data.

Reference:

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

Pergunta 14: Correto
You need to schedule an Azure Data Factory pipeline to execute when a new file arrives in an Azure Data Lake Storage Gen2 container.

Which type of trigger should you use?

Explicação

Correct Answer: *D*

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

Blob Storage account.

Reference:

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

Pergunta 15: Incorreto

You have two Azure Data Factory instances named ADFdev and ADFprod. ADFdev connects to an Azure DevOps Git repository.

You publish changes from the main branch of the Git repository to ADFdev.

You need to deploy the artifacts from ADFdev to ADFprod.

What should you do first?

Explicação

Correct Answer: C

In Azure Data Factory, continuous integration and delivery (CI/CD) means moving Data Factory pipelines from one environment (development, test, production) to another.

Note: The following is a guide for setting up an Azure Pipelines release that automates the deployment of a data factory to multiple environments.

1. In Azure DevOps, open the project that's configured with your data factory.

2. On the left side of the page, select Pipelines, and then select Releases.

3. Select New pipeline, or, if you have existing pipelines, select New and then New release pipeline.

4. In the Stage name box, enter the name of your environment.

5. Select Add artifact, and then select the git repository configured with your development data factory. Select the publish branch of the repository for the Default branch. By default, this publish branch is adf_publish.

6. Select the Empty job template.

Reference:

https://docs.microsoft.com/en-us/azure/data-factory/continuous-integration-deployment

Pergunta 16: Incorreto
You are developing a solution that will stream to Azure Stream Analytics. The solution will have both streaming data and reference data.

Which input type should you use for the reference data?

Explicação

Correct Answer: B

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

Reference:

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

Pergunta 17: Incorreto
You are designing an Azure Stream Analytics job to process incoming events from sensors in retail environments.

You need to process the events to produce a running average of shopper counts during the previous 15 minutes, calculated at five-minute intervals.

Which type of window should you use?

Explicação

Correct Answer: *C*

Hopping, as we need to calculate running average, which means it will have overlapping.

Reference:

https://docs.microsoft.com/en-us/stream-analytics-query/hopping-window-azure-stream-analytics

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

Pergunta 18: Incorreto
You are monitoring an Azure Stream Analytics job by using metrics in Azure.

You discover that during the last 12 hours, the average watermark delay is consistently greater than the configured late arrival tolerance.

What is a possible cause of this behavior?

Explicação

Correct Answer: D

Watermark Delay indicates the delay of the streaming data processing job.

There are a number of resource constraints that can cause the streaming pipeline to slow down. The watermark delay metric can rise due to:

1. Not enough processing resources in Stream Analytics to handle the volume of input events. To scale up resources, see Understand and adjust Streaming

Units.

2. Not enough throughput within the input event brokers, so they are throttled. For possible solutions, see Automatically scale up Azure Event Hubs throughput units.

3. Output sinks are not provisioned with enough capacity, so they are throttled. The possible solutions vary widely based on the flavor of output service being used.

Reference:

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-time-handling

Pergunta 19: Correto
You plan to build a structured streaming solution in Azure Databricks. The solution will count new events in five-minute intervals and report only events that arrive during the interval. The output will be sent to a Delta Lake table.

Which output mode should you use?

Explicação

Correct Answer: *C*
Append Mode: Only new rows appended in the result table since the last trigger are written to external storage. This is applicable only for the queries where existing rows in the Result Table are not expected to change.
Incorrect Answers:
B: Complete Mode: The entire updated result table is written to external storage. It is up to the storage connector to decide how to handle the writing of the entire table.
A: Update Mode: Only the rows that were updated in the result table since the last trigger are written to external storage. This is different from Complete Mode in that Update Mode outputs only the rows that have changed since the last trigger. If the query doesn't contain aggregations, it is equivalent to Append mode.

Reference:

https://docs.databricks.com/getting-started/spark/streaming.html

Pergunta 20: Correto
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?

Explicação

Correct Answer: *B*
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

Pergunta 21: Incorreto
You have an Azure Synapse Analytics Apache Spark pool named Pool1.

You plan to load JSON files from an Azure Data Lake Storage Gen2 container into the tables in Pool1. The structure and data types vary by file.

You need to load the files into the tables. The solution must maintain the source data types.

What should you do?

Explicação

Correct Answer: C

Serverless SQL pool can automatically synchronize metadata from Apache Spark. A serverless SQL pool database will be created for each database existing in serverless Apache Spark pools.

Serverless SQL pool enables you to query data in your data lake. It offers a T-SQL query surface area that accommodates semi-structured and unstructured data queries.

To support a smooth experience for in place querying of data that's located in Azure Storage files, serverless SQL pool uses the OPENROWSET function with additional capabilities.

The easiest way to see to the content of your JSON file is to provide the file URL to the OPENROWSET function, specify csv FORMAT.

Reference:

https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/query-json-files https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/query-data-storage

Pergunta 22: Incorreto
You have an Azure Databricks workspace named workspace1 in the Standard pricing tier. Workspace1 contains an all-purpose cluster named cluster1.

You need to reduce the time it takes for cluster1 to start and scale up. The solution must minimize costs.

What should you do first?

Explicação

Correct Answer: D

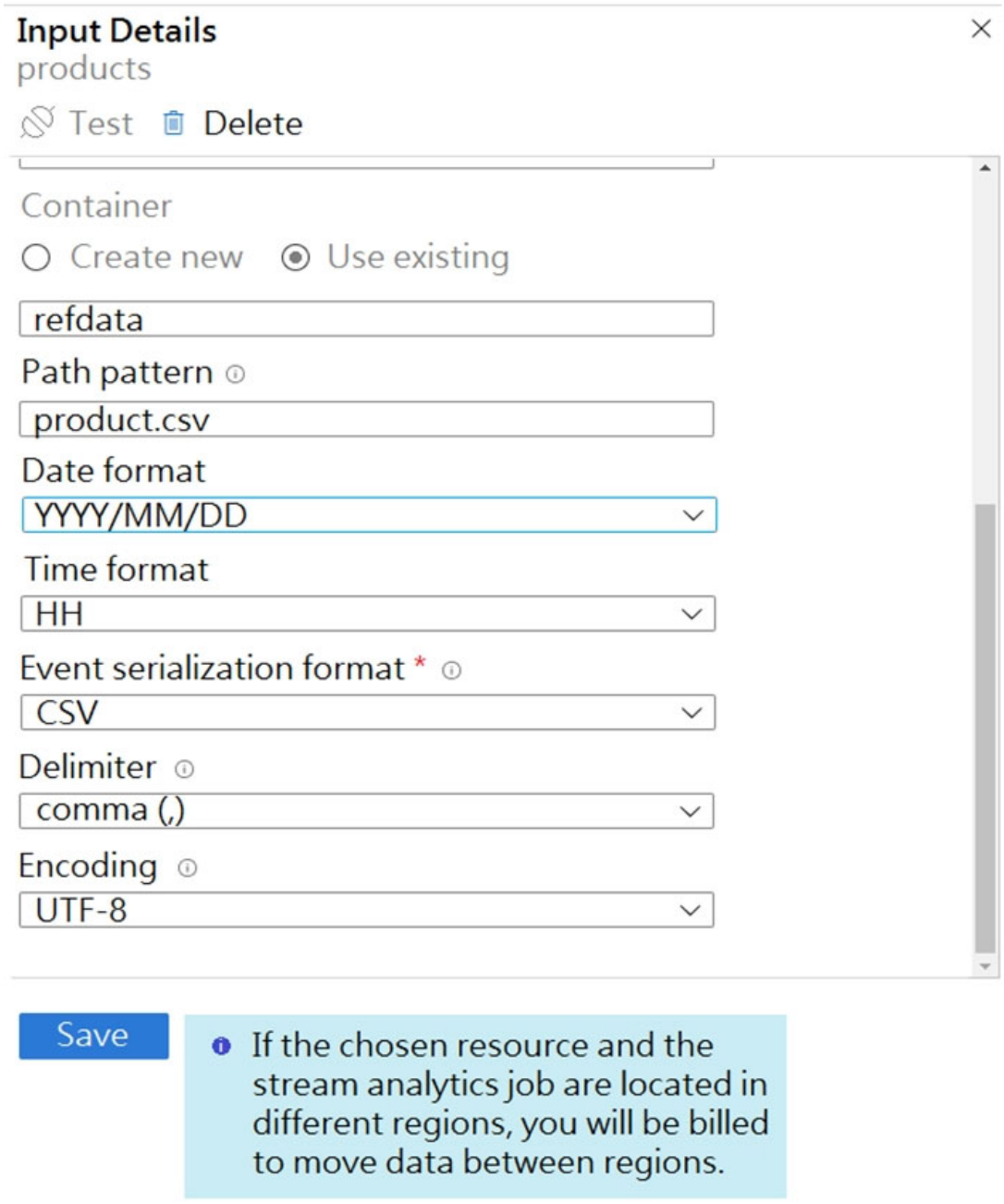
You can use Databricks Pools to Speed up your Data Pipelines and Scale Clusters Quickly.

Databricks Pools, a managed cache of virtual machine instances that enables clusters to start and scale 4 times faster.

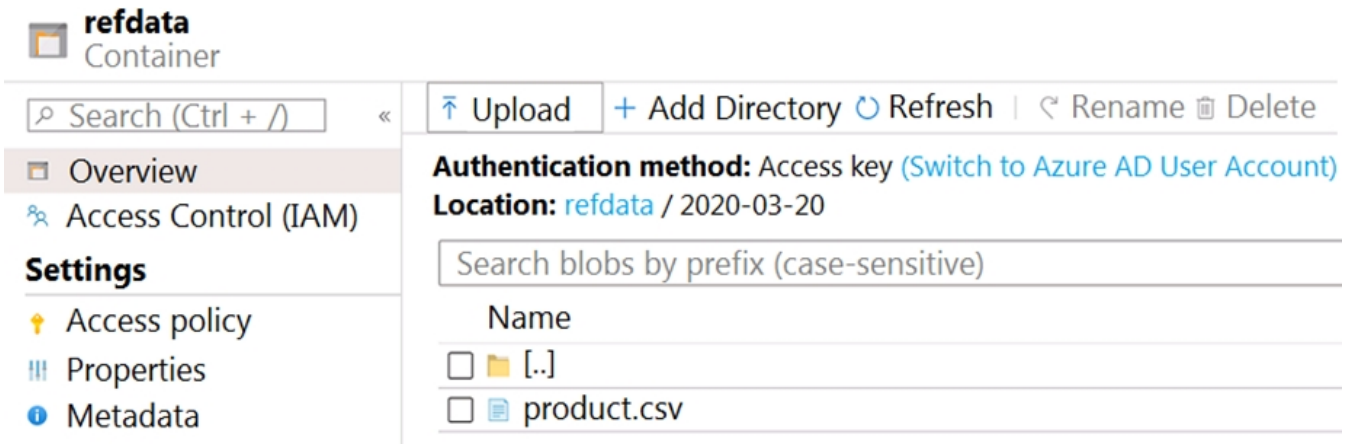
Reference:

https://databricks.com/blog/2019/11/11/databricks-pools-speed-up-data-pipelines.html

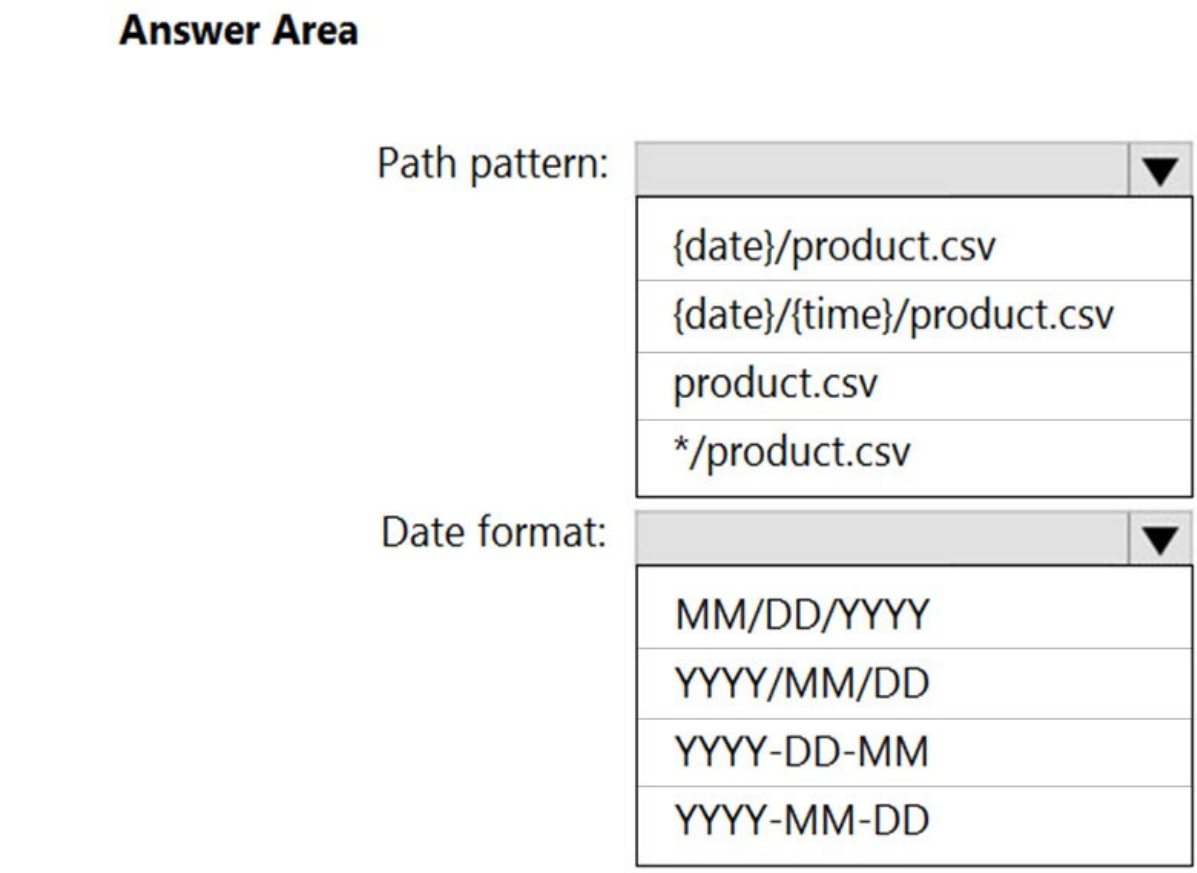
Pergunta 23: Correto
You are building an Azure Stream Analytics job that queries reference data from a product catalog file. The file is updated daily.
The reference data input details for the file are shown in the Input exhibit. (Click the Input tab.)



The storage account container view is shown in the Refdata exhibit. (Click the Refdata tab.)



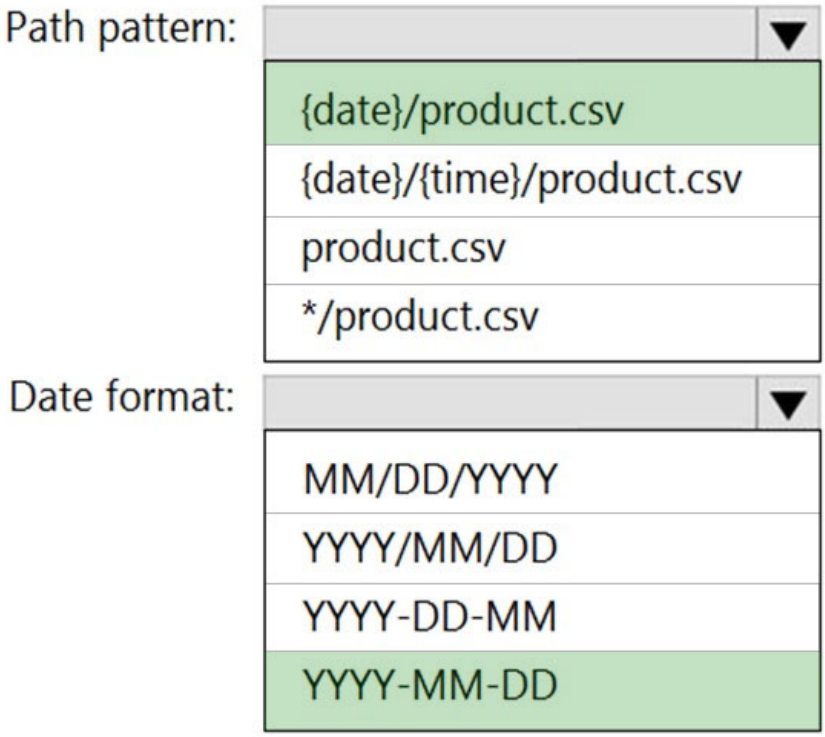
You need to configure the Stream Analytics job to pick up the new reference data.
What should you configure? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.
Hot Area:



Explicação

Correct Answer:

Answer Area



Box 1: {date}/product.csv -
In the 2nd exhibit we see: Location: refdata / 2020-03-20
Note: Path Pattern: This is a required property that is used to locate your blobs within the specified container. Within the path, you may choose to specify one or more instances of the following 2 variables:
{date}, {time}
Example 1: products/{date}/{time}/product-list.csv
Example 2: products/{date}/product-list.csv

Example 3: product-list.csv -

Box 2: YYYY-MM-DD -
Note: Date Format [optional]: If you have used {date} within the Path Pattern that you specified, then you can select the date format in which your blobs are organized from the drop-down of supported formats.
Example: YYYY/MM/DD, MM/DD/YYYY, etc.
Reference:
https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-use-reference-data

Pergunta 24: Incorreto
You have the following Azure Stream Analytics query.

For each of the following statements, select Yes if the statement is true. Otherwise, select No.
NOTE: Each correct selection is worth one point.
Hot Area:

```
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
SELECT * INTO output
       FROM step2
       PARTITION BY StateID
```

Answer Area

Statements	Yes	No
The query combines two streams of partitioned data.	○	○
The stream scheme key and count must match the output scheme.	○	○
Providing 60 streaming units will optimize the performance of the query.	○	○

Explicação

Correct Answer:

Answer Area

Statements	Yes	No
The query combines two streams of partitioned data.	○	○
The stream scheme key and count must match the output scheme.	○	○
Providing 60 streaming units will optimize the performance of the query.	○	○

Box 1: No -
Note: 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 -
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.
In general, the best practice is to start with 6 SUs for queries that don't use PARTITION BY.
Here there are 10 partitions, so 6x10 = 60 SUs is good.
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/
https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-streaming-unit-consumption

Pergunta 25: Correto
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 Storege 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.
Hot Area:

Answer Area

Explicação

Correct Answer:

Answer Area

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

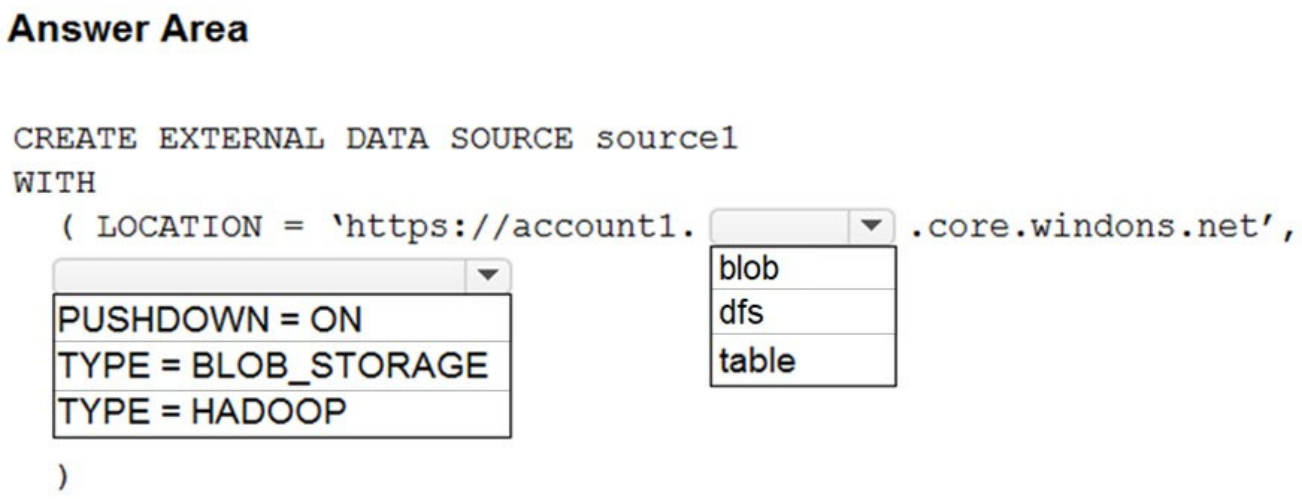
Box 2. OPENROWSET -
When using serverless SQL pool, CETAS is used to create an external table and export query results to Azure Storage Blob or Azure Data Lake Storage Gen2.
Example:

AS -
SELECT decennialTime, stateName, SUM(population) AS population

FROM -
OPENROWSET(BULK
'https://azureopendatastorage.blob.core.windows.net/censusdatacontainer/release/us_population_county/year=*/*.parquet',
FORMAT='PARQUET') AS [r]
GROUP BY decennialTime, stateName

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

Pergunta 26: Correto
You have an Azure Synapse Analytics dedicated SQL pool named Pool1 and an Azure Data Lake Storage Gen2 account named Account1.
You plan to access the files in Account1 by using an external table.
You need to create a data source in Pool1 that you can reference when you create the external table.
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.
Hot Area:



Explicação

Correct Answer:

Box 1: dfs-

Box 2: HADOOP
-- Creates a Hadoop external data source in dedicated SQL pool

CREATE EXTERNAL DATA SOURCE AzureDataLakeStore

WITH

(LOCATION = 'abfss://data@newyorktaxidataset.dfs.core.windows.net' ,

CREDENTIAL = ADLS_credential ,

TYPE = HADOOP

)

-- Creates an external data source for Azure Data Lake Gen2

CREATE EXTERNAL DATA SOURCE YellowTaxi

WITH

(LOCATION = 'https://azureopendatastorage.blob.core.windows.net/nyctlc/yellow/',

TYPE = HADOOP

)

The question asks to create a data source in Pool1. So the answer is dfs & HADOOP.

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

Pergunta 27: Correto
You have an Azure subscription that contains an Azure Blob Storage account named storage1 and an Azure Synapse Analytics dedicated SQL pool named

Pool1.

You need to store data in storage1. The data will be read by Pool1. The solution must meet the following requirements:

Enable Pool1 to skip columns and rows that are unnecessary in a query.

• Automatically create column statistics.

• Minimize the size of files.

Which type of file should you use?

Explicação

Correct Answer: B

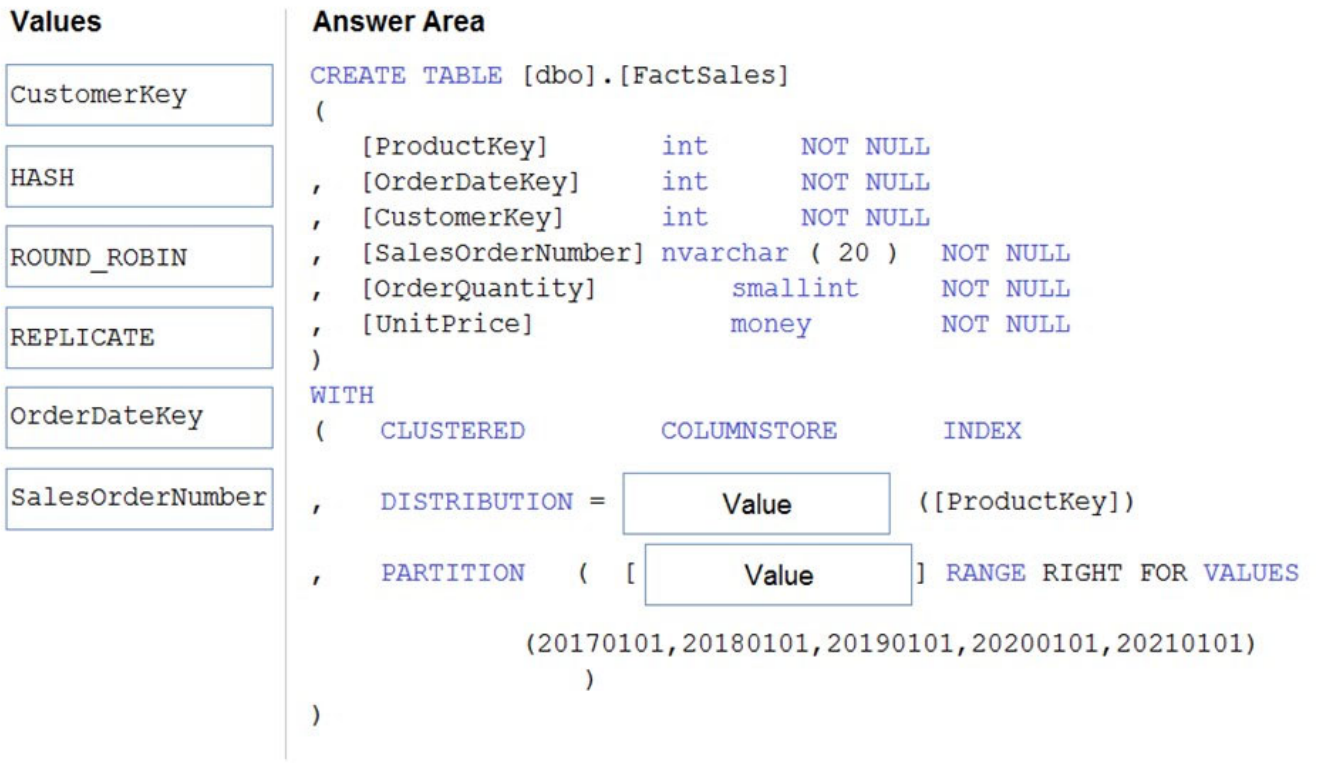
Automatic creation of statistics is turned on for Parquet files. For CSV files, you need to create statistics manually until automatic creation of CSV files statistics is supported.

Reference:

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

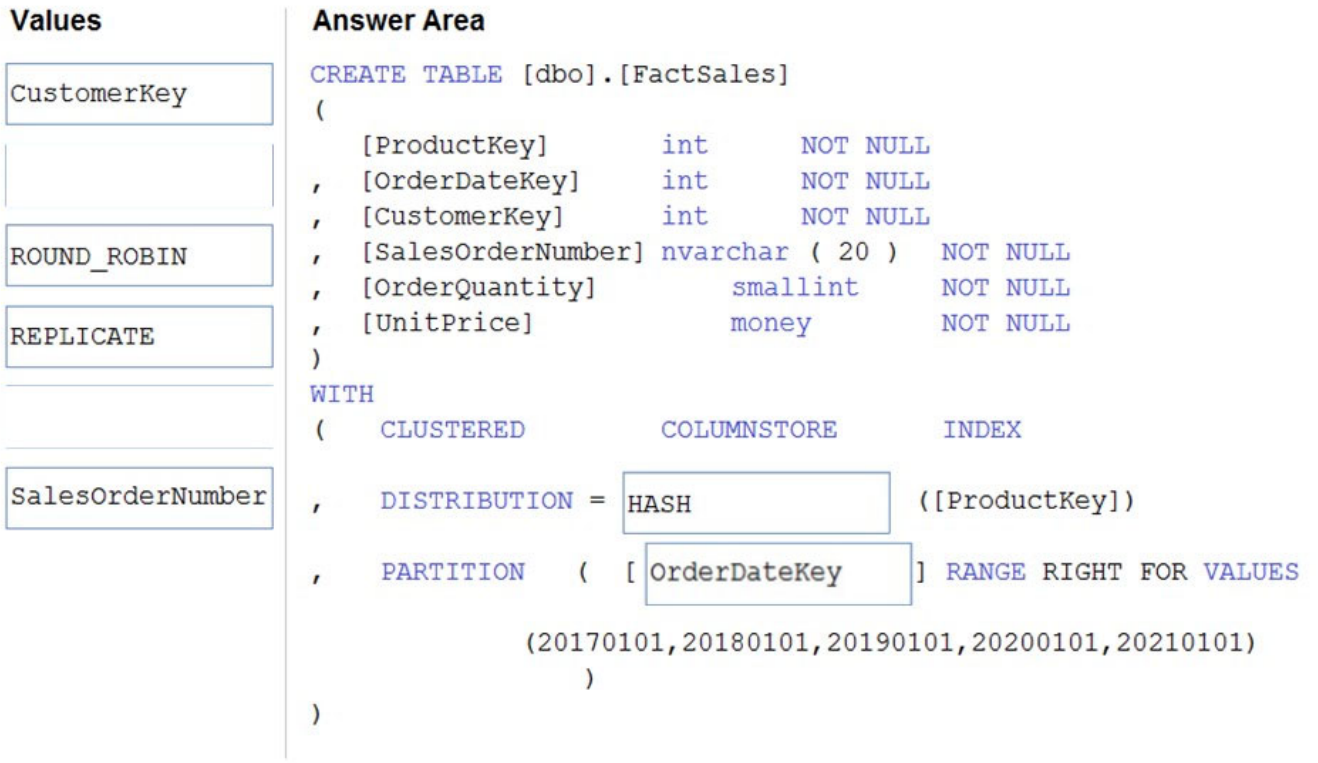
Pergunta 28: Incorreto
You plan to create a table in an Azure Synapse Analytics dedicated SQL pool.
Data in the table will be retained for five years. Once a year, data that is older than five years will be deleted.
You need to ensure that the data is distributed evenly across partitions. The solution must minimize the amount of time required to delete old data.
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.
Select and Place:



Explicação

Correct Answer:



Box 1: HASH -

Box 2: OrderDateKey -
In most cases, table partitions are created on a date column.
A way to eliminate rollbacks is to use Metadata Only operations like partition switching for data management. For example, rather than execute a DELETE statement to delete all rows in a table where the order_date was in October of 2001, you could partition your data early. Then you can switch out the partition with data for an empty partition from another table.

Reference:
https://docs.microsoft.com/en-us/sql/t-sql/statements/create-table-azure-sql-data-warehouse
https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/best-practices-dedicated-sql-pool

Pergunta 29: Incorreto
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 in container1 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 Get Metadata activity that retrieves the DateTime of the files.

Does this meet the goal?

Explicação

Correct Answer: B

Instead use a serverless SQL pool to create an external table with the extra column.

Reference:

https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/create-use-external-tables

Pergunta 30: Correto
You have an Azure Data Lake Storage account that contains a staging zone.

You need to design a daily 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 executes an Azure Databricks notebook, and then inserts the data into the data warehouse.

Does this meet the goal?

Explicação

Correct Answer: *A*

You can execute R code in a notebook, and then call it from Data Factory.

You can check it at "Databricks Notebook activity" header:

https://docs.microsoft.com/en-US/azure/data-factory/transform-data

https://docs.microsoft.com/en-us/azure/databricks/spark/latest/sparkr/overview

Pergunta 31: Incorreto
You have an Azure Data Lake Storage account that contains a staging zone.

You need to design a daily 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 executes mapping data flow, and then inserts the data into the data warehouse.

Does this meet the goal?

Explicação

Correct Answer: B

If you need to transform data in a way that is not supported by Data Factory, you can create a custom activity, not a mapping flow,5 with your own data processing logic and use the activity in the pipeline. You can create a custom activity to run R scripts on your HDInsight cluster with R installed.

Reference:

https://docs.microsoft.com/en-US/azure/data-factory/transform-data

Pergunta 32: Incorreto
You have an Azure Data Lake Storage account that contains a staging zone.

You need to design a daily 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 schedule an Azure Databricks job that executes an R notebook, and then inserts the data into the data warehouse.

Does this meet the goal?

Explicação

Correct Answer: B

Must use an Azure Data Factory, not an Azure Databricks job.

Reference:

https://docs.microsoft.com/en-US/azure/data-factory/transform-data

Pergunta 33: Incorreto
You plan to create an Azure Data Factory pipeline that will include a mapping data flow.

You have JSON data containing objects that have nested arrays.

You need to transform the JSON-formatted data into a tabular dataset. The dataset must have one row for each item in the arrays.

Which transformation method should you use in the mapping data flow?

Explicação

Correct Answer: D

Use the flatten transformation to take array values inside hierarchical structures such as JSON and unroll them into individual rows. This process is known as denormalization.

Reference:

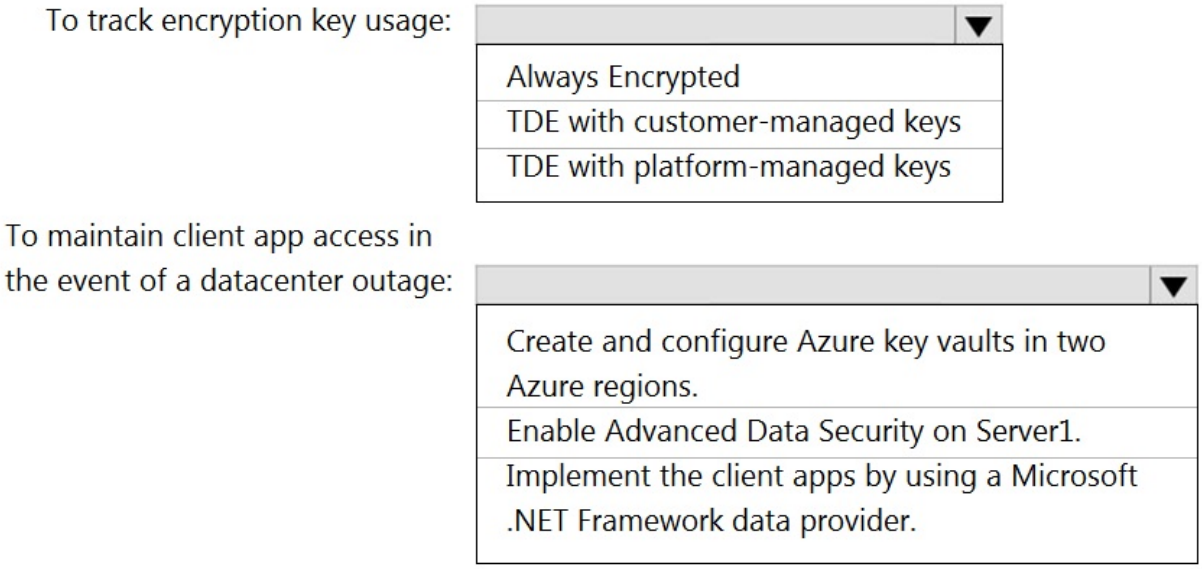
https://docs.microsoft.com/en-us/azure/data-factory/data-flow-flatten

Pergunta 34: Incorreto

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.
Hot Area:

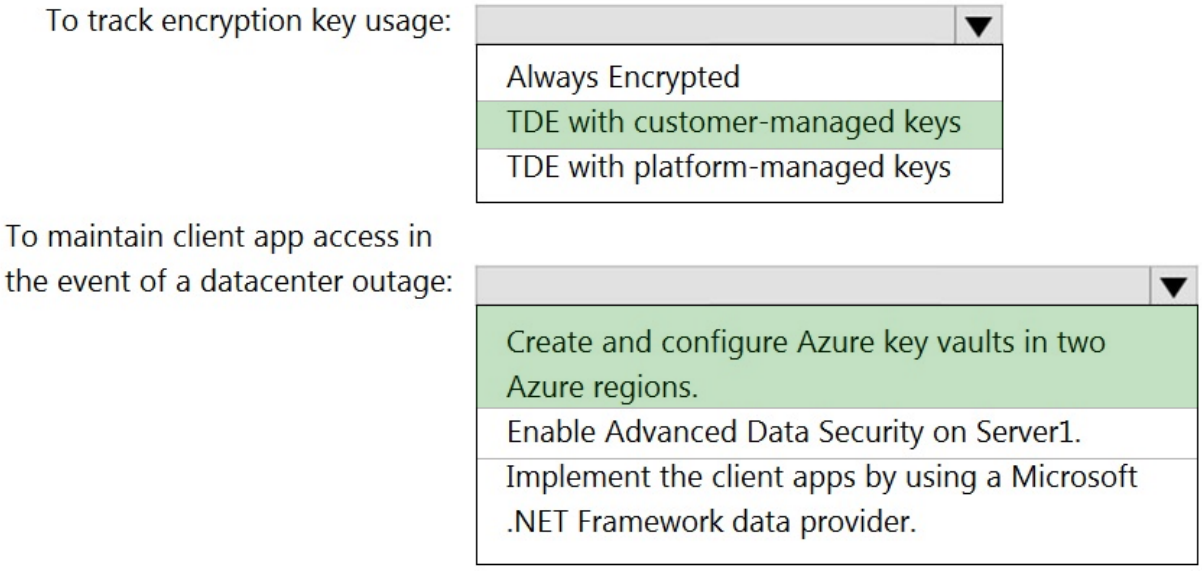
Answer Area



Explicação

Correct Answer:

Answer Area



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

Pergunta 35: Incorreto
You have an Azure Data Lake Storage Gen2 account named adls2 that is protected by a virtual network.

You are designing a SQL pool in Azure Synapse that will use adls2 as a source.

What should you use to authenticate to adls2?

Explicação

Suggested Answer: D

Managed Identity authentication is required when your storage account is attached to a VNet.

Reference:

https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/quickstart-bulk-load-copy-tsql-examples

Conteúdo do curso

- Iniciar
  Simulado 1: Practice Test 1: DP-203: Microsoft Azure Data Engineer Exam

- Iniciar
  Simulado 2: `Practice Test 2: DP-203: Microsoft Azure Data Engineer Exam
- Iniciar
  Simulado 3: Practice Test 3: DP-203: Microsoft Azure Data Engineer Exam
- Iniciar
  Simulado 4: Practice Test 4: DP-203: Microsoft Azure Data Engineer Exam

---

Sobre este curso

Professional Practice Exam | 185 Questions | May 2022 Updated Version | Maestro DP-203 in First Attempt

Pelos números

Nível de experiência: Todos os níveis

Alunos: 615

Idiomas: Inglês

Legendas: Não

Descrição

DP-203: Microsoft Azure Data Engineer Associate exam is one of the recent additions in the Azure role-based certification model. These full-length mock tests with 185 different questions.

We offer the following resources to supplement your experience and help you prepare for DP-203: Microsoft Azure Data Engineer Associate Certification, after passing DP-203 exam, you will earn the Microsoft Azure Data Engineer Associate certification.

These practice sets measure your ability to accomplish the following technical tasks: design and implement data storage, design and develop data processing, design and implement data security, monitor and optimize data storage, data processing and governance–this role should manage how decisions in each area affect an overall solution.

These practice exams supplement each exam blueprint guide and help an individual test their knowledge before taking the final exam

This exam should have subject matter expertise in integrating, transforming, and consolidating data from various structured and unstructured data systems into a structure that is suitable for building analytics solutions.

This practice test course contains four practice tests so that you can prove your skills in Azure Architecture. A perfect tool to assess your readiness, and find those one or two spots that you can study in the days before taking the test.

Azure DP-203 Exam Objectives

* Design and Implement Data Storage (40-45%)

* Design and Develop Data Processing (25-30%)

* Design and Implement Data Security (10-15%)

* Monitor and Optimize Data Storage and Data Processing (10-15%)

You should immediately catch hold of DP-203 practice tests once you are done with your preparation. Do not forget to follow the official study guide for DP-203 that you can find easily on Microsoft's official website.

Happy Learning and All the Best!

O que você aprenderá

- It contains 2022 exam version questions which are likely to be asked in the Real Exam.
- These practice tests help you with Self-Study and Self-Assessment in Exam.
- These practice tests helps you check your knowledge and upgrade your skills.
- It is compatible with iPhone and Android mobiles.
- Lifetime access practice tests with all the updates for DP-203 Certification Exam.

Há algum requisito ou pré-requisito para o curso?

- General knowledge of IT architecture
- Intermediate to strong knowledge of most Azure offerings

Para quem é este curso:

- People who want to become Azure Data Engineer Associate
- People preparing for Microsoft's DP-203 exam
- Good technical exposure with Azure Data Engineer
- Those people who interested in passing the Azure DP-203 exam
- IT teams who want to learn more about Azure Data Engineering

Instrutor

Shivam Gupta is Microsoft Certified Trainer and a Cloud Solutions Architect/DevOps Expert at an MNC company in India with multiple information technology certifications including

Microsoft Certified Azure Solutions Architect Expert,

Microsoft Certified Azure DevOps Expert,

Microsoft Certified Azure Security Engineer,

Microsoft Certified Azure Data Engineer,

Microsoft Certified Azure Network Engineer,

Microsoft Certified Azure Administrator,

Microsoft Certified Azure Fundamentals,

Microsoft Certified Azure Data Fundamentals,

Certified Kubernetes Administrator &

AWS Certified Solutions Architect.

Shivam has been privileged enough to have several roles for more than 9 years as DevOps engineer, senior infrastructure/Cloud engineer, solutions architect as well as cloud security expert.

Shivam has hands-on experience in architecting/automating and optimizing mission-critical deployment over small & large infrastructure. Proficient with Configuration Management tools and in developing CI/CD pipelines.