Microsoft

Exam Questions DP-203

Data Engineering on Microsoft Azure

About Exambible

Your Partner of IT Exam

Found in 1998

Exambible is a company specialized on providing high quality IT exam practice study materials, especially Cisco CCNA, CCDA, CCNP, CCIE, Checkpoint CCSE, CompTIA A+, Network+ certification practice exams and so on. We guarantee that the candidates will not only pass any IT exam at the first attempt but also get profound understanding about the certificates they have got. There are so many alike companies in this industry, however, Exambible has its unique advantages that other companies could not achieve.

Our Advances

* 99.9% Uptime

> All examinations will be up to date.

* 24/7 Quality Support

> We will provide service round the clock.

* 100% Pass Rate

> Our guarantee that you will pass the exam.

* Unique Gurantee

> If you do not pass the exam at the first time, we will not only arrange FULL REFUND for you, but also provide you another exam of your claim, ABSOLUTELY FREE!

NEW QUESTION 1
- (Exam Topic 3)
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.

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Answer as below

Answer Area

Path pattern: {date}/product.csv

Date format: YYYY/MM/DD

NEW QUESTION 2
- (Exam Topic 3)
You have two Azure Blob Storage accounts named account1 and account2?
You plan to create an Azure Data Factory pipeline that will use scheduled intervals to replicate newly created or modified blobs from account1 to account?
You need to recommend a solution to implement the pipeline. The solution must meet the following requirements:
• Ensure that the pipeline only copies blobs that were created of modified since the most recent replication event.
• Minimize the effort to create the pipeline. What should you recommend?

A. Create a pipeline that contains a flowlet.
B. Create a pipeline that contains a Data Flow activity.
C. Run the Copy Data tool and select Metadata-driven copy task.
D. Run the Copy Data tool and select Built-in copy task.

Answer: A

NEW QUESTION 3
- (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 4
- (Exam Topic 3)
You are implementing a batch dataset in the Parquet format.
Data tiles will be produced by using Azure Data Factory and stored in Azure Data Lake Storage Gen2. The files will be consumed by an Azure Synapse Analytics serverless SQL pool.
You need to minimize storage costs for the solution. What should you do?

A. Store all the data as strings in the Parquet tiles.
B. Use OPENROWEST to query the Parquet files.
C. Create an external table mat contains a subset of columns from the Parquet files.
D. Use Snappy compression for the files.

Answer: C

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

NEW QUESTION 5
- (Exam Topic 3)
You have an Azure Data Factory pipeline shown the following exhibit.

Execute Pipeline		Web		Set variable
Exec_COPY_BLOB	✓ →	Web_GetIP	✓ →	Set status	✓

The execution log for the first pipeline run is shown in the following exhibit.

Activity runs

Pipeline run ID 87f89922-14fa-468f-b13f-2f867606f4ff

All status

Showing 1 - 2 items

Activity name	Activity type	Run start	Duration	Status
Web_GetIP	Web	Nov 10, 2022, 11:11:36 a	00:00:02	❌ Failed
Exec_COPY_BLOB	Execute Pipeline	Nov 10, 2022, 11:11:25 a	00:00:11	✓ Succeeded

The execution log for the second pipeline run is shown in the following exhibit.

Activity runs

Pipeline run ID a7b5b522-cfaf-4c09-b3a9-f842986be984

All status

Showing 1 - 3 items

Activity name	Activity type	Run start	Duration	Status
Set status	Set variable	Nov 10, 2022, 11:13:17 a	00:00:01	✓ Succeeded
Web_GetIP	Web	Nov 10, 2022, 11:12:59 a	00:00:16	✓ Succeeded
Exec_COPY_BLOB	Execute Pipeline	Nov 10, 2022, 11:12:48 a	00:00:11	⊘ Skipped

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

Answer Area

Statements	Yes	No
The Retry property of the Web_GetIP activity is set to 1.	○	○
The waitOnCompletion property of the Exec_COPY_BLOB activity is set to true.	○	○
The Exec_COPY_BLOB activity was skipped during the second run due to pipeline dependencies.	○	○

A. Mastered
B. Not Mastered

Answer: A

Explanation:

Answer Area

Statements	Yes	No
The Retry property of the Web_GetIP activity is set to 1.	○	[○]
The waitOnCompletion property of the Exec_COPY_BLOB activity is set to true.	○	[○]
The Exec_COPY_BLOB activity was skipped during the second run due to pipeline dependencies.	○	[○]

NEW QUESTION 6
- (Exam Topic 3)
You have an Azure Synapse Analytics workspace named WS1 that contains an Apache Spark pool named
Pool1.
You plan to create a database named D61 in Pool1.
You need to ensure that when tables are created in DB1, the tables are available automatically as external tables to the built-in serverless SQL pod.
Which format should you use for the tables in DB1?

A. Parquet
B. CSV
C. ORC
D. JSON

Answer: A

Explanation:
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.
For each Spark external table based on Parquet or CSV and located in Azure Storage, an external table is created in a serverless SQL pool database.
Reference:
https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/develop-storage-files-spark-tables

NEW QUESTION 7
- (Exam Topic 3)
You have an Azure Synapse Analytics dedicated SQL pool named Pool1 that contains a table named Sales. Sales has row-level security (RLS) applied. RLS uses the following predicate filter.

```
CREATE FUNCTION Security.fn_securitypredicate(@SalesRep AS sysname)
    RETURNS TABLE
WITH SCHEMABINDING
AS
    RETURN SELECT 1 AS fn_securitypredicate_result
WHERE @SalesRep = USER_NAME() OR USER_NAME() = 'Manager';
```

A user named SalesUser1 is assigned the db_datareader role for Pool1.

A user named SalesUser1 is assigned the db_datareader role for Pool1. Which rows in the Sales table are returned when SalesUser1 queries the table?

A. only the rows for which the value in the User_Name column is SalesUser1
B. all the rows
C. only the rows for which the value in the SalesRep column is Manager
D. only the rows for which the value in the SalesRep column is SalesUser1

Answer: A

NEW QUESTION 8
- (Exam Topic 3)
You have an Azure Synapse Analytics dedicated SQL pool mat contains a table named dbo.Users.
You need to prevent a group of users from reading user email addresses from dbo.Users. What should you use?

A. row-level security
B. column-level security
C. Dynamic data masking
D. Transparent Data Encryption (TDD

Answer: B

NEW QUESTION 9
- (Exam Topic 3)
You use Azure Data Factory to prepare data to be queried by Azure Synapse Analytics serverless SQL pools. Files are initially ingested into an Azure Data Lake Storage Gen2 account as 10 small JSON files. Each file contains the same data attributes and data from a subsidiary of your company.
You need to move the files to a different folder and transform the data to meet the following requirements: ≫ Provide the fastest possible query times.

≫ Automatically infer the schema from the underlying files.
How should you configure the Data Factory copy activity? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Copy behavior:

| Flatten hierarchy |
| Merge files |
| Preserve hierarchy |

Sink file type:

| CSV |
| JSON |
| Parquet |
| TXT |

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: Preserver herarchy
Compared to the flat namespace on Blob storage, the hierarchical namespace greatly improves the performance of directory management operations, which improves overall job performance.
Box 2: Parquet
Azure Data Factory parquet format is supported for Azure Data Lake Storage Gen2. Parquet supports the schema property.
Reference:
https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-introduction https://docs.microsoft.com/en-us/azure/data-factory/format-parquet

NEW QUESTION 10
- (Exam Topic 3)
You are building an Azure Stream Analytics job to retrieve game data.
You need to ensure that the job returns the highest scoring record for each five-minute time interval of each game.
How should you complete the Stream Analytics query? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

```
SELECT  [                                          ▼ ]  as HighestScore
          Collect(Score)
          CollectTop(1) OVER(ORDER BY Score Desc)
          Game, MAX(Score)
          TopOne() OVER(PARTITION BY Game ORDER BY Score Desc)

FROM input TIMESTAMP BY CreatedAt

GROUP BY [                                          ▼ ]
          Game
          Hopping(minute,5)
          Tumbling(minute,5)
          Windows(TumblingWindow(minute,5),Hopping(minute,5))
```

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: TopOne OVER(PARTITION BY Game ORDER BY Score Desc)
TopOne returns the top-rank record, where rank defines the ranking position of the event in the window according to the specified ordering. Ordering/ranking is based on event columns and can be specified in ORDER BY clause.
Box 2: Hopping(minute,5)
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 and be emitted more often than the window size. 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.
A picture containing timeline Description automatically generated

Every 5 seconds give me the count of Tweets over the last 10 seconds

A 10-second Hopping Window with a 5-second "Hop"
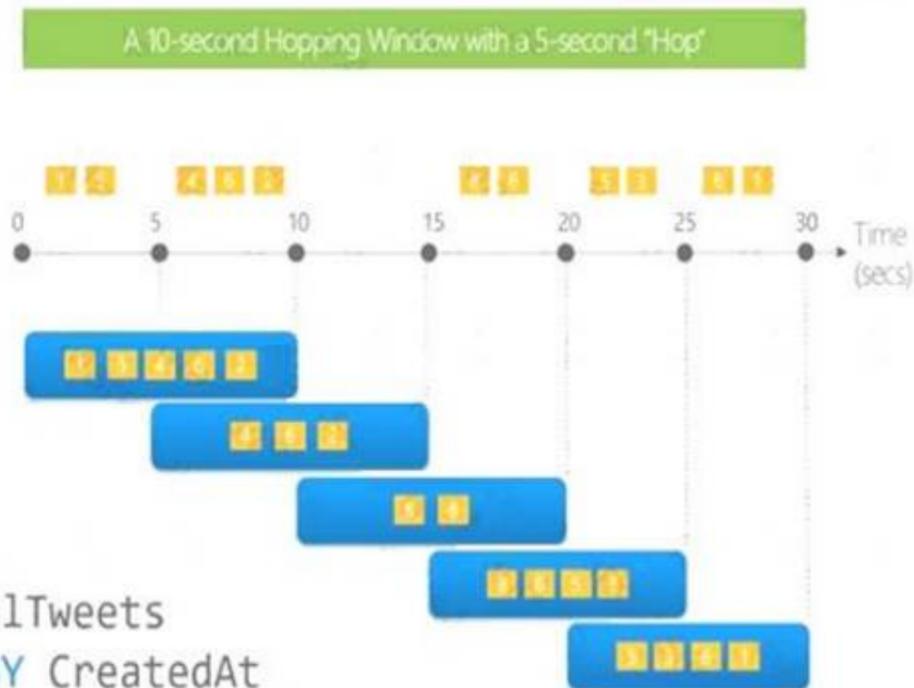
```
SELECT Topic, COUNT(*) AS TotalTweets
FROM TwitterStream TIMESTAMP BY CreatedAt
GROUP BY Topic, HoppingWindow(second, 10 , 5)
```

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

NEW QUESTION 10
- (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 12
- (Exam Topic 3)
You have an Azure Data Lake Storage Gen2 account that contains a JSON file for customers. The file contains two attributes named FirstName and LastName. You need to copy the data from the JSON file to an Azure Synapse Analytics table by using Azure Databricks. A new column must be created that concatenates the FirstName and LastName values.
You create the following components:

≫ A destination table in Azure Synapse

≫ An Azure Blob storage container

≫ A service principal

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

Actions	Answer Area
Mount the Data Lake Storage onto DBFS.	
Write the results to a table in Azure Synapse.	
Specify a temporary folder to stage the data.	
Read the file into a data frame.	
Perform transformations on the data frame.	

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Table Description automatically generated
Step 1: Mount the Data Lake Storage onto DBFS
Begin with creating a file system in the Azure Data Lake Storage Gen2 account. Step 2: Read the file into a data frame.
You can load the json files as a data frame in Azure Databricks. Step 3: Perform transformations on the data frame.
Step 4: Specify a temporary folder to stage the data
Specify a temporary folder to use while moving data between Azure Databricks and Azure Synapse. Step 5: Write the results to a table in Azure Synapse.
You upload the transformed data frame into Azure Synapse. You use the Azure Synapse connector for Azure Databricks to directly upload a dataframe as a table in a Azure Synapse.
Reference:
https://docs.microsoft.com/en-us/azure/azure-databricks/databricks-extract-load-sql-data-warehouse

NEW QUESTION 16
- (Exam Topic 3)
You are designing a folder structure for the files m an Azure Data Lake Storage Gen2 account. The account has one container that contains three years of data.
You need to recommend a folder structure that meets the following requirements:
• Supports partition elimination for queries by Azure Synapse Analytics serverless SQL pooh
• Supports fast data retrieval for data from the current month
• Simplifies data security management by department Which folder structure should you recommend?

A. \YYY\MM\DD\Department\DataSource\DataFile_YYYMMDD.parquet
B. \Depdftment\DataSource\YYY\MM\DataFile_YYYMMDD.parquet
C. \DD\MM\YYYY\Department\DataSource\DataFile_DDMMYY.parquet
D. \DataSource\Department\YYYYMM\DataFile_YYYYMMDD.parquet

Answer: B

Explanation:
Department top level in the hierarchy to simplify security management.
Month (MM) at the leaf/bottom level to support fast data retrieval for data from the current month.

NEW QUESTION 17
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You plan to create an Azure Databricks workspace that has a tiered structure. The workspace will contain the following three workloads:
⟩ A workload for data engineers who will use Python and SQL.
⟩ A workload for jobs that will run notebooks that use Python, Scala, and SOL.
⟩ A workload that data scientists will use to perform ad hoc analysis in Scala and R.
The enterprise architecture team at your company identifies the following standards for Databricks environments:
⟩ The data engineers must share a cluster.
⟩ The job cluster will be managed by using a request process whereby data scientists and data engineers provide packaged notebooks for deployment to the cluster.
⟩ All the data scientists must be assigned their own cluster that terminates automatically after 120 minutes of inactivity. Currently, there are three data scientists.
You need to create the Databricks clusters for the workloads.
Solution: You create a Standard cluster for each data scientist, a High Concurrency cluster for the data engineers, and a High Concurrency cluster for the jobs.
Does this meet the goal?

A. Yes
B. No

Answer: A

Explanation:
We need a High Concurrency cluster for the data engineers and the jobs. Note:
Standard clusters are recommended for a single user. Standard can run workloads developed in any language: Python, R, Scala, and SQL.

A high concurrency cluster is a managed cloud resource. The key benefits of high concurrency clusters are that they provide Apache Spark-native fine-grained sharing for maximum resource utilization and minimum query latencies.
Reference: https://docs.azuredatabricks.net/clusters/configure.html

NEW QUESTION 20
- (Exam Topic 3)
You are designing the folder structure for an Azure Data Lake Storage Gen2 account. You identify the following usage patterns:
• Users will query data by using Azure Synapse Analytics serverless SQL pools and Azure Synapse Analytics serverless Apache Spark pods.
• Most queries will include a filter on the current year or week.
• Data will be secured by data source.
You need to recommend a folder structure that meets the following requirements:
• Supports the usage patterns
• Simplifies folder security
• Minimizes query times
Which folder structure should you recommend?
A)

```
\YYYYY\WW\DataSource\SubjectArea\FileData_YYYY_MM_DD.parquet
```

B)

```
DataSource\SubjectArea\WW\YYYY\FileData_YYYY_MM_DD.parquet
```

C)

```
\DataSource\SubjectArea\YYYY\WW\FileData_YYYY_MM_DD.parquet
```

D)

```
\DataSource\SubjectArea\YYYY-WW\FileData_YYYY_MM_DD.parquet
```

E)

```
WW\YYYY\SubjectArea\DataSource\FileData_YYYY_MM_DD.parquet
```

A. Option A
B. Option B
C. Option C
D. Option D
E. Option E

Answer: C

Explanation:
Data will be secured by data source. -> Use DataSource as top folder.
Most queries will include a filter on the current year or week -> Use \YYYY\WW\ as subfolders. Common Use Cases
A common use case is to filter data stored in a date (and possibly time) folder structure such as
/YYYY/MM/DD/ or /YYYY/MM/YYYY-MM-DD/. As new data is generated/sent/copied/moved to the storage account, a new folder is created for each specific time period. This strategy organises data into a maintainable folder structure.
Reference: https://www.serverlesssql.com/optimisation/azurestoragefilteringusingfilepath/

NEW QUESTION 24
- (Exam Topic 3)
You have a Microsoft Purview account. The Lineage view of a CSV file is shown in the following exhibit.
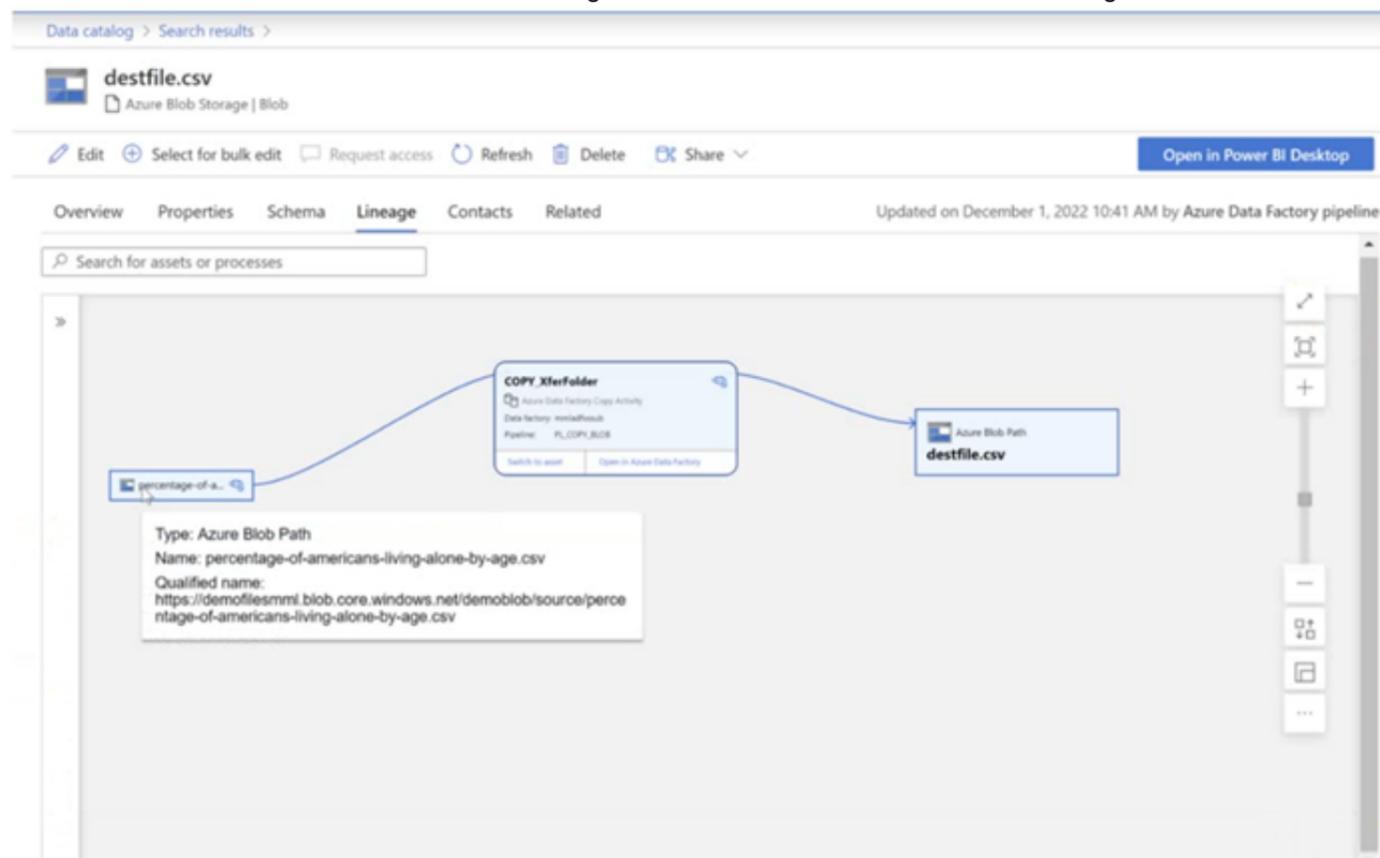
How is the data for the lineage populated?

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 27
- (Exam Topic 3)
You have an enterprise data warehouse in Azure Synapse Analytics named DW1 on a server named Server1. You need to determine the size of the transaction log file for each distribution of DW1.
What should you do?

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

Answer: A

Explanation:
For information about the current log file size, its maximum size, and the autogrow option for the file, you can also use the size, max_size, and growth columns for that log file in sys.database_files.
Reference:
https://docs.microsoft.com/en-us/sql/relational-databases/logs/manage-the-size-of-the-transaction-log-file

NEW QUESTION 30
- (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
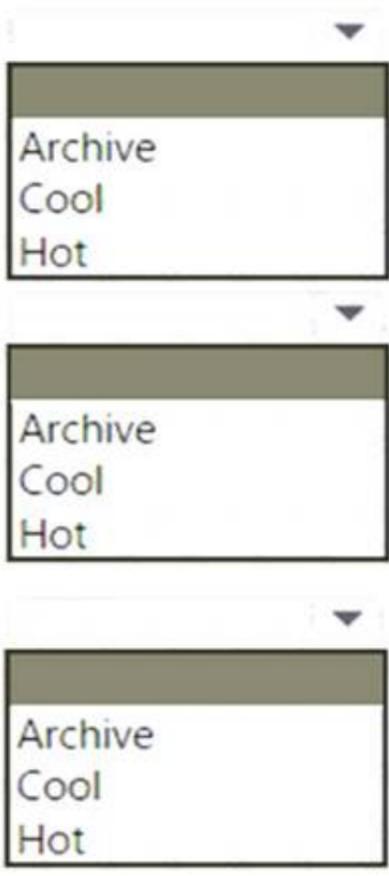
A. Mastered
B. Not Mastered

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 34
- (Exam Topic 3)
You have an Azure subscription that contains an Azure Synapse Analytics workspace named workspace1. Workspace1 contains a dedicated SQL pool named SQL Pool and an Apache Spark pool named sparkpool. Sparkpool1 contains a DataFrame named pyspark.df.
You need to write the contents of pyspark_df to a tabte in SQLPooM by using a PySpark notebook. How should you complete the code? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Answer Area

```
pyspark_df.createOrReplaceTempView("pysparkdftemptable")

%%local
%%spark        park.sqlContext.sql ("select * from pysparkdftemptable")
%%sql
                                    ("sqlpool1.dbo.PySparkTable", Constants.INTERNAL)
               jdbc
               saveAsTable
               synapsesql
```

A. Mastered
B. Not Mastered

Answer: A

Explanation:

Answer Area

```
pyspark_df.createOrReplaceTempView("pysparkdftemptable")

%%local
%%spark        park.sqlContext.sql ("select * from pysparkdftemptable")
%%sql
                                    ("sqlpool1.dbo.PySparkTable", Constants.INTERNAL)
               jdbc
               saveAsTable
               synapsesql
```

NEW QUESTION 36
- (Exam Topic 3)
You have an Azure Synapse Analytics dedicated SQL pool that contains the users shown in the following table.

Name	Role
User1	Server admin
User2	db_datereader

User1 executes a query on the database, and the query returns the results shown in the following exhibit.

```
1  SELECT c.name,
2      tbl.name as table_name,
3      typ.name as datatype,
4      c.is_masked,
5      c.masking_function
6  FROM sys.masked_columns AS c
7  INNER JOIN sys.tables AS tbl ON c.[object_id] = tbl.[object_id]
8  INNER JOIN sys.types typ ON c.user_type_id = typ.user_type_id
9  WHERE is_masked = 1;
10
```

Results Messages

	name	table_name	datatype	is_masked	masking_function
1	BirthDate	DimCustomer	date	1	default()
2	Gender	DimCustomer	nvarchar	1	default()
3	EmailAddress	DimCustomer	nvarchar	1	email()
4	YearlyIncome	DimCustomer	money	1	default()

User1 is the only user who has access to the unmasked data.
Use the drop-down menus to select the answer choice that completes each statement based on the information presented in the graphic.
NOTE: Each correct selection is worth one point.

When User2 queries the YearlyIncome column, the values returned will be **[answer choice]**.

a random number
the values stored in the database
XXXX
0

When User1 queries the BirthDate column, the values returned will be **[answer choice]**.

a random date
the values stored in the database
XXXX
1900-01-01

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Graphical user interface, text, application, email Description automatically generated
Box 1: 0
The YearlyIncome column is of the money data type.
The Default masking function: Full masking according to the data types of the designated fields

> Use a zero value for numeric data types (bigint, bit, decimal, int, money, numeric, smallint, smallmoney, tinyint, float, real).

Box 2: the values stored in the database
Users with administrator privileges are always excluded from masking, and see the original data without any mask.
Reference:
https://docs.microsoft.com/en-us/azure/azure-sql/database/dynamic-data-masking-overview

NEW QUESTION 38
- (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	▼ Logic Apps Azure Data Factory Azure Automation
Store	▼ Azure Data Lake Storage Azure Blob storage Azure files
Prepare and Train	▼ HDInsight Apache Spark cluster Azure Databricks HDInsight Apache Storm cluster
Model and Serve	▼ HDInsight Apache Kafka cluster Azure Synapse Analytics Azure Data Lake Storage

A. Mastered
B. Not Mastered

Answer: A

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

NEW QUESTION 41
- (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 45
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You plan to create an Azure Databricks workspace that has a tiered structure. The workspace will contain the following three workloads:

> A workload for data engineers who will use Python and SQL.

> A workload for jobs that will run notebooks that use Python, Scala, and SOL.

> A workload that data scientists will use to perform ad hoc analysis in Scala and R.
The enterprise architecture team at your company identifies the following standards for Databricks environments:

> The data engineers must share a cluster.

> The job cluster will be managed by using a request process whereby data scientists and data engineers provide packaged notebooks for deployment to the cluster.

>

All the data scientists must be assigned their own cluster that terminates automatically after 120 minutes of inactivity. Currently, there are three data scientists. You need to create the Databricks clusters for the workloads.
Solution: You create a Standard cluster for each data scientist, a 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 49
- (Exam Topic 3)
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.
You plan to create an Azure Databricks workspace that has a tiered structure. The workspace will contain the following three workloads:

≫ A workload for data engineers who will use Python and SQL.

≫ A workload for jobs that will run notebooks that use Python, Scala, and SOL.

≫ A workload that data scientists will use to perform ad hoc analysis in Scala and R.

The enterprise architecture team at your company identifies the following standards for Databricks environments:

≫ The data engineers must share a cluster.

≫ The job cluster will be managed by using a request process whereby data scientists and data engineers provide packaged notebooks for deployment to the cluster.

≫ All the data scientists must be assigned their own cluster that terminates automatically after 120 minutes of inactivity. Currently, there are three data scientists. You need to create the Databricks clusters for the workloads.
Solution: You create a High Concurrency cluster for each data scientist, a High Concurrency cluster for the data engineers, and a Standard cluster for the jobs.
Does this meet the goal?

A. Yes
B. No

Answer: B

Explanation:
Need a High Concurrency cluster for the jobs.
Standard clusters are recommended for a single user. Standard can run workloads developed in any language: Python, R, Scala, and SQL.
A high concurrency cluster is a managed cloud resource. The key benefits of high concurrency clusters are that they provide Apache Spark-native fine-grained sharing for maximum resource utilization and minimum query latencies.
Reference: https://docs.azuredatabricks.net/clusters/configure.html

NEW QUESTION 54
- (Exam Topic 3)
You build an Azure Data Factory pipeline to move data from an Azure Data Lake Storage Gen2 container to a database in an Azure Synapse Analytics dedicated SQL pool.
Data in the container is stored in the following folder structure.
/in/{YYYY}/{MM}/{DD}/{HH}/{mm}
The earliest folder is /in/2021/01/01/00/00. The latest folder is /in/2021/01/15/01/45. You need to configure a pipeline trigger to meet the following requirements:

≫ Existing data must be loaded.

≫ Data must be loaded every 30 minutes.

≫ Late-arriving data of up to two minutes must he included in the load for the time at which the data should have arrived.
How should you configure the pipeline trigger? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Type:
- Event
- On-demand
- Schedule
- Tumbling window

Additional properties:
- Prefix: /in/, Event: Blob created
- Recurrence: 30 minutes, Start time: 2021-01-01T00:00
- Recurrence: 30 minutes, Start time: 2021-01-01T00:00, Delay: 2 minutes
- Recurrence: 32 minutes, Start time: 2021-01-15T01:45

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: Tumbling window
To be able to use the Delay parameter we select Tumbling window. Box 2:
Recurrence: 30 minutes, not 32 minutes
Delay: 2 minutes.
The amount of time to delay the start of data processing for the window. The pipeline run is started after the expected execution time plus the amount of delay. The delay defines how long the trigger waits past the due time before triggering a new run. The delay doesn't alter the window startTime.
Reference:
https://docs.microsoft.com/en-us/azure/data-factory/how-to-create-tumbling-window-trigger

NEW QUESTION 55
- (Exam Topic 3)
You build a data warehouse in an Azure Synapse Analytics dedicated SQL pool.
Analysts write a complex SELECT query that contains multiple JOIN and CASE statements to transform data for use in inventory reports. The inventory reports will use the data and additional WHERE parameters depending on the report. The reports will be produced once daily.
You need to implement a solution to make the dataset available for the reports. The solution must minimize query times.
What should you implement?

A. a materialized view
B. a replicated table
C. in ordered clustered columnstore index
D. result set chaching

Answer: A

Explanation:
Materialized views for dedicated SQL pools in Azure Synapse provide a low maintenance method for complex analytical queries to get fast performance without any query change.
Note: When result set caching is enabled, dedicated SQL pool automatically caches query results in the user database for repetitive use. This allows subsequent query executions to get results directly from the persisted cache so recomputation is not needed. Result set caching improves query performance and reduces compute resource usage. In addition, queries using cached results set do not use any concurrency slots and thus do not count against existing concurrency limits.
Reference:
https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/performance-tuning-materialized- https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/performance-tuning-result-set-cac

NEW QUESTION 60
- (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 61
- (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 65
- (Exam Topic 1)
You need to integrate the on-premises data sources and Azure Synapse Analytics. The solution must meet the data integration requirements.
Which type of integration runtime should you use?

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

Answer: C

NEW QUESTION 68
- (Exam Topic 1)
You need to design a data ingestion and storage solution for the Twitter feeds. The solution must meet the customer sentiment analytics requirements.
What should you include in the solution? To answer, select the appropriate options in the answer area NOTE: Each correct selection b worth one point.

Answer Area

To increase the throughput of ingesting the Twitter feeds:

Configure Event Hubs partitions.
Enable Auto-Inflate in Event Hubs.
Use Event Hubs Dedicated.

To store the Twitter feed data, use:

An Azure Data Lake Storage Gen2 account
An Azure Databricks high concurrency cluster
An Azure General-purpose v2 storage account in the Premium tier

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Graphical user interface, text Description automatically generated
Box 1: Configure Evegent Hubs partitions
Scenario: Maximize the throughput of ingesting Twitter feeds from Event Hubs to Azure Storage without purchasing additional throughput or capacity units.
Event Hubs is designed to help with processing of large volumes of events. Event Hubs throughput is scaled by using partitions and throughput-unit allocations.
Event Hubs traffic is controlled by TUs (standard tier). Auto-inflate enables you to start small with the minimum required TUs you choose. The feature then scales automatically to the maximum limit of TUs you need, depending on the increase in your traffic.
Box 2: An Azure Data Lake Storage Gen2 account
Scenario: Ensure that the data store supports Azure AD-based access control down to the object level. Azure Data Lake Storage Gen2 implements an access control model that supports both Azure role-based
access control (Azure RBAC) and POSIX-like access control lists (ACLs).
Reference:
https://docs.microsoft.com/en-us/azure/event-hubs/event-hubs-features https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-access-control

NEW QUESTION 73
- (Exam Topic 1)
You need to design a data retention solution for the Twitter feed data records. The solution must meet the customer sentiment analytics requirements.
Which Azure Storage functionality should you include in the solution?

A. change feed
B. soft delete
C. time-based retention
D. lifecycle management

Answer: D

Explanation:
Scenario: Purge Twitter feed data records that are older than two years.
Data sets have unique lifecycles. Early in the lifecycle, people access some data often. But the need for access often drops drastically as the data ages. Some data remains idle in the cloud and is rarely accessed once stored. Some data sets expire days or months after creation, while other data sets are actively read and modified throughout their lifetimes. Azure Storage lifecycle management offers a rule-based policy that you can use to transition blob data to the appropriate access tiers or to expire data at the end of the data lifecycle.
Reference:
https://docs.microsoft.com/en-us/azure/storage/blobs/lifecycle-management-overview

NEW QUESTION 75
- (Exam Topic 3)
You are developing a solution that will stream to Azure Stream Analytics. The solution will have both streaming data and reference data.
Which input type should you use for the reference data?

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

Answer: B

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

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

NEW QUESTION 76
- (Exam Topic 3)
You are designing a solution that will copy Parquet files stored in an Azure Blob storage account to an Azure Data Lake Storage Gen2 account.
The data will be loaded daily to the data lake and will use a folder structure of {Year}/{Month}/{Day}/. You need to design a daily Azure Data Factory data load to minimize the data transfer between the two accounts.
Which two configurations should you include in the design? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

A. Delete the files in the destination before loading new data.
B. Filter by the last modified date of the source files.
C. Delete the source files after they are copied.
D. Specify a file naming pattern for the destination.

Answer: BD

Explanation:
Copy data from one place to another. The requirements are : 1- need to minimize transfer and 2- need to adapte data to the destination folder structure. Filter on LastModifiedDate will copy everything that have changed since the latest load while minimizing the data transfert. Specifying the file naming pattern allows to copy data at the right place to the destination Data Lake.

NEW QUESTION 81
- (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 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: You use a dedicated SQL pool to create an external table that has an additional DateTime column. Does this meet the goal?

A. Yes
B. No

Answer: B

Explanation:
Instead use the derived column transformation to generate new columns in your data flow or to modify existing fields.
Reference:
https://docs.microsoft.com/en-us/azure/data-factory/data-flow-derived-column

NEW QUESTION 86
- (Exam Topic 3)
You are designing a highly available Azure Data Lake Storage solution that will induce geo-zone-redundant storage (GZRS).
You need to monitor for replication delays that can affect the recovery point objective (RPO). What should you include m the monitoring solution?

A. Last Sync Time
B. Average Success Latency
C. Error errors
D. availability

Answer: A

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 89
- (Exam Topic 3)
You are designing an Azure Databricks interactive cluster. The cluster will be used infrequently and will be configured for auto-termination.
You need to ensure that the cluster configuration is retained indefinitely after the cluster is terminated. The solution must minimize costs.
What should you do?

A. Clone the cluster after it is terminated.
B. Terminate the cluster manually when processing completes.
C. Create an Azure runbook that starts the cluster every 90 days.
D. Pin the cluster.

Answer: D

Explanation:

To keep an interactive cluster configuration even after it has been terminated for more than 30 days, an administrator can pin a cluster to the cluster list.
References:
https://docs.azuredatabricks.net/clusters/clusters-manage.html#automatic-termination

NEW QUESTION 91
- (Exam Topic 3)
You have two Azure Data Factory instances named ADFdev and ADFprod. ADFdev connects to an Azure DevOps Git repository.
You publish changes from the main branch of the Git repository to ADFdev. You need to deploy the artifacts from ADFdev to ADFprod.
What should you do first?

A. From ADFdev, modify the Git configuration.
B. From ADFdev, create a linked service.
C. From Azure DevOps, create a release pipeline.
D. From Azure DevOps, update the main branch.

Answer: C

Explanation:
In Azure Data Factory, continuous integration and delivery (CI/CD) means moving Data Factory pipelines from one environment (development, test, production) to another.
Note:
The following is a guide for setting up an Azure Pipelines release that automates the deployment of a data factory to multiple environments.

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

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

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

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

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

> Select the Empty job template. Reference:
https://docs.microsoft.com/en-us/azure/data-factory/continuous-integration-deployment

NEW QUESTION 94
- (Exam Topic 3)
You have files and folders in Azure Data Lake Storage Gen2 for an Azure Synapse workspace as shown in the following exhibit.
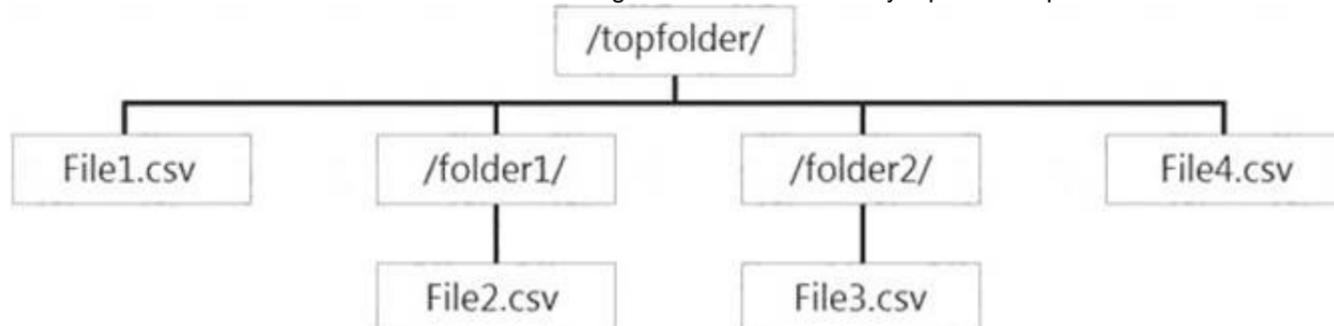
You create an external table named ExtTable that has LOCATION='/topfolder/'.
When you query ExtTable by using an Azure Synapse Analytics serverless SQL pool, which files are returned?

A. File2.csv and File3.csv only
B. File1.csv and File4.csv only
C. File1.csv, File2.csv, File3.csv, and File4.csv
D. File1.csv only

Answer: B

Explanation:
To run a T-SQL query over a set of files within a folder or set of folders while treating them as a single entity or rowset, provide a path to a folder or a pattern (using wildcards) over a set of files or folders. Reference:
https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/query-data-storage#query-multiple-files-or-folders

NEW QUESTION 95
- (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 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: You use an Azure Synapse Analytics serverless SQL pool to create an external table that has an additional DateTime column.
Does this meet the goal?

A. Yes
B. No

Answer: B

Explanation:
Instead use the derived column transformation to generate new columns in your data flow or to modify existing fields.
Reference:
https://docs.microsoft.com/en-us/azure/data-factory/data-flow-derived-column

NEW QUESTION 99
- (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

To minimize costs:
- Locally-redundant storage (LRS)
- The Archive access tier
- The Cool access tier
- Zone-redundant storage (ZRS)

To delete blobs:
- Azure Automation runbooks
- Azure Storage lifecycle management
- Soft delete

A. Mastered
B. Not Mastered

Answer: A

Explanation:
https://learn.microsoft.com/en-us/azure/storage/blobs/premium-tier-for-data-lake-storage
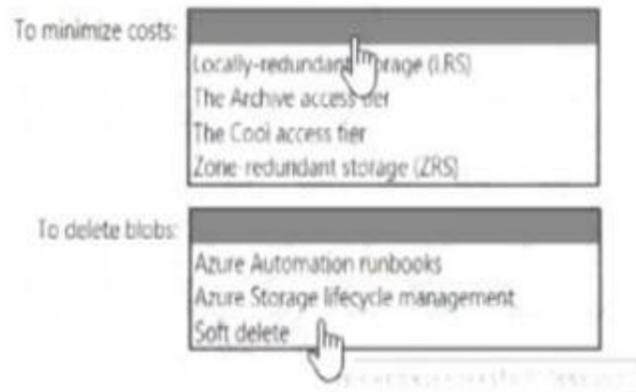
NEW QUESTION 101
- (Exam Topic 3)
You are designing an Azure Data Lake Storage solution that will transform raw JSON files for use in an analytical workload.
You need to recommend a format for the transformed files. The solution must meet the following requirements:
≫ Contain information about the data types of each column in the files.
≫ Support querying a subset of columns in the files.
≫ Support read-heavy analytical workloads.
≫ Minimize the file size.
What should you recommend?

A. JSON
B. CSV
C. Apache Avro
D. Apache Parquet

Answer: D

Explanation:
Parquet, an open-source file format for Hadoop, stores nested data structures in a flat columnar format. Compared to a traditional approach where data is stored in a row-oriented approach, Parquet file format is
more efficient in terms of storage and performance.
It is especially good for queries that read particular columns from a "wide" (with many columns) table since only needed columns are read, and IO is minimized.
Reference: https://www.clairvoyant.ai/blog/big-data-file-formats

NEW QUESTION 106
- (Exam Topic 3)
You have an Azure Synapse Analytics SQL pool named Pool1 on a logical Microsoft SQL server named Server1.
You need to implement Transparent Data Encryption (TDE) on Pool1 by using a custom key named key1. Which five actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

Actions

| Answer Area |

Enable TDE on Pool1.

Assign a managed identity to Server1.

Configure key1 as the TDE protector for Server1.

Add key1 to the Azure key vault.

Create an Azure key vault and grant the managed
identity permissions to the key vault.

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Graphical user interface, text, application Description automatically generated
Step 1: Assign a managed identity to Server1
You will need an existing Managed Instance as a prerequisite.
Step 2: Create an Azure key vault and grant the managed identity permissions to the vault Create Resource and setup Azure Key Vault.
Step 3: Add key1 to the Azure key vault
The recommended way is to import an existing key from a .pfx file or get an existing key from the vault. Alternatively, generate a new key directly in Azure Key Vault.
Step 4: Configure key1 as the TDE protector for Server1 Provide TDE Protector key
Step 5: Enable TDE on Pool1 Reference:
https://docs.microsoft.com/en-us/azure/azure-sql/managed-instance/scripts/transparent-data-encryption-byok-po

NEW QUESTION 107
- (Exam Topic 3)
You have an Azure Synapse Analytics dedicated SQL pod.
You need to create a pipeline that will execute a stored procedure in the dedicated SQL pool and use the returned result set as the input (or a downstream activity. The solution must minimize development effort.
Which Type of activity should you use in the pipeline?

A. Notebook
B. U-SQL
C. Script
D. Stored Procedure

Answer: D

NEW QUESTION 111
- (Exam Topic 3)
You are creating an Azure Data Factory data flow that will ingest data from a CSV file, cast columns to specified types of data, and insert the data into a table in an Azure Synapse Analytic dedicated SQL pool. The CSV file contains three columns named username, comment, and date.
The data flow already contains the following:
➢ A source transformation.
➢ A Derived Column transformation to set the appropriate types of data.
➢ A sink transformation to land the data in the pool.
You need to ensure that the data flow meets the following requirements:
➢ All valid rows must be written to the destination table.
➢ Truncation errors in the comment column must be avoided proactively.
➢ Any rows containing comment values that will cause truncation errors upon insert must be written to a file in blob storage.
Which two actions should you perform? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

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

Answer: AB

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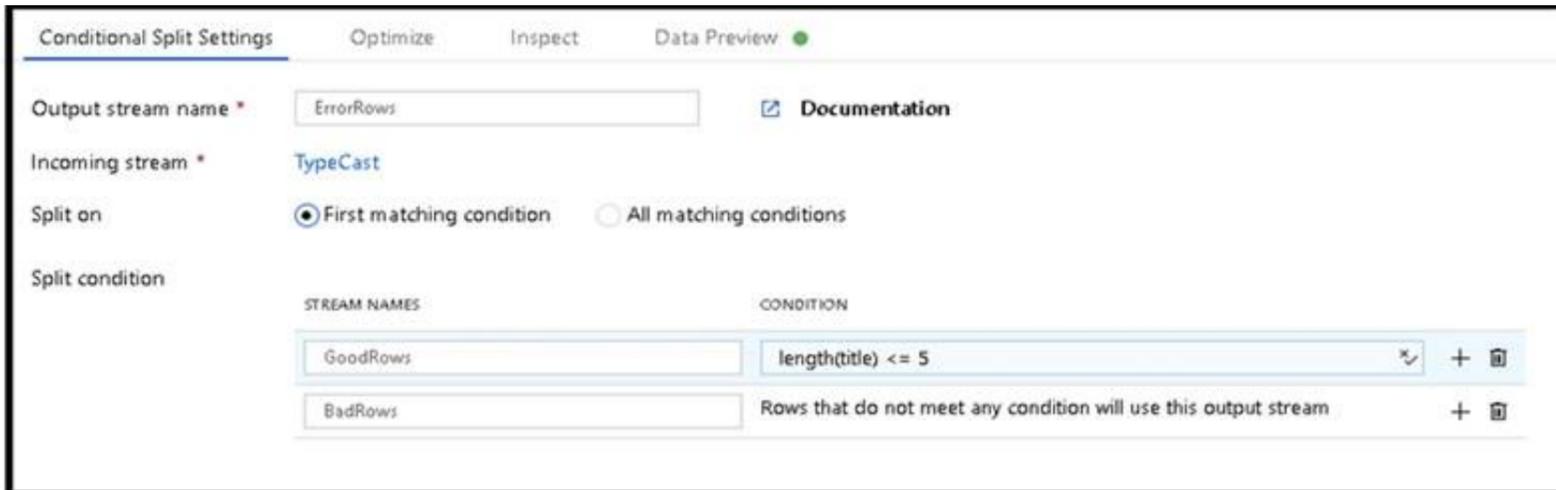
* 2. This conditional split transformation defines the maximum length of "title" to be five. Any row that is less than or equal to five will go into the GoodRows stream. Any row that is larger than five will go into the
BadRows stream. A:

* 3. Now we need to log the rows that failed. Add a sink transformation to the BadRows stream for logging. Here, we'll "auto-map" all of the fields so that we have logging of the complete transaction record. This is a text-delimited CSV file output to a single file in Blob Storage. We'll call the log file "badrows.csv".
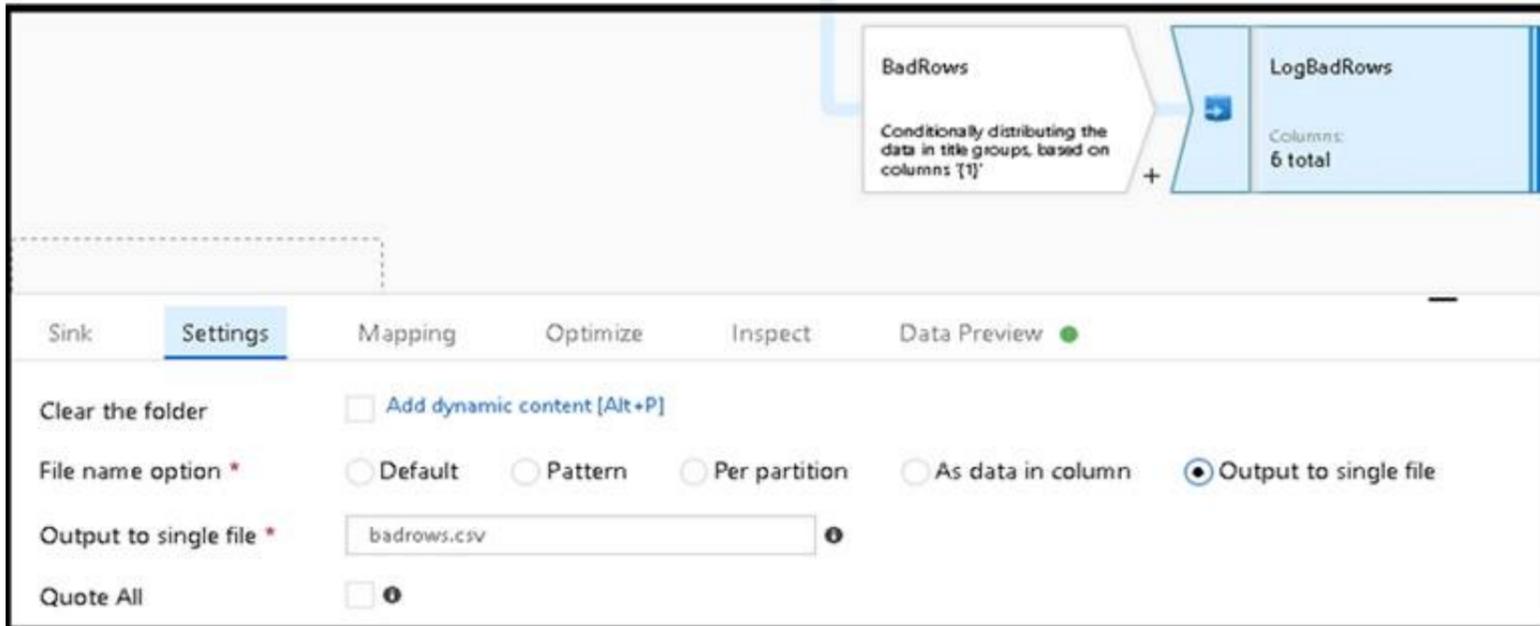
* 4. The completed data flow is shown below. We are now able to split off error rows to avoid the SQL truncation errors and put those entries into a log file. Meanwhile, successful rows can continue to write to our target database.

Reference:
https://docs.microsoft.com/en-us/azure/data-factory/how-to-data-flow-error-rows
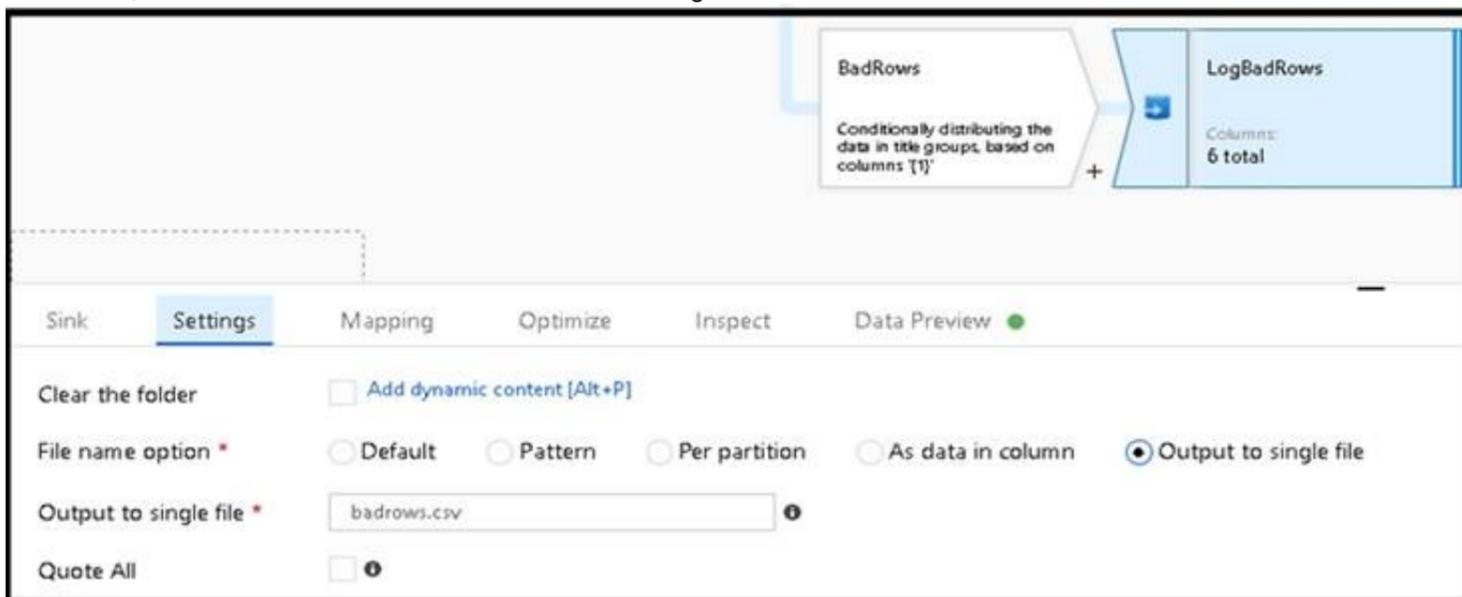
NEW QUESTION 112
- (Exam Topic 3)
You are planning a solution to aggregate streaming data that originates in Apache Kafka and is output to Azure Data Lake Storage Gen2. The developers who will implement the stream processing solution use Java, Which service should you recommend using to process the streaming data?

A. Azure Data Factory
B. Azure Stream Analytics
C. Azure Databricks
D. Azure Event Hubs

Answer: C

Explanation:
https://docs.microsoft.com/en-us/azure/architecture/data-guide/technology-choices/stream-processing

NEW QUESTION 116
- (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 121
- (Exam Topic 3)
You have an Azure subscription linked to an Azure Active Directory (Azure AD) tenant that contains a service principal named ServicePrincipal1. The subscription contains an Azure Data Lake Storage account named adls1. Adls1 contains a folder named Folder2 that has a URI of
https://adls1.dfs.core.windows.net/container1/Folder1/Folder2/.
ServicePrincipal1 has the access control list (ACL) permissions shown in the following table.

Resource	Permission
container1	Access – Execute
Folder1	Access – Execute
Folder2	Access – Read

You need to ensure that ServicePrincipal1 can perform the following actions:
➤ Traverse child items that are created in Folder2.
➤ Read files that are created in Folder2.
The solution must use the principle of least privilege.
Which two permissions should you grant to ServicePrincipal1 for Folder2? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

A. Access - Read
B. Access - Write
C. Access - Execute
D. Default-Read
E. Default - Write
F. Default - Execute

Answer: DF

Explanation:
Execute (X) permission is required to traverse the child items of a folder.
There are two kinds of access control lists (ACLs), Access ACLs and Default ACLs. Access ACLs: These control access to an object. Files and folders both have Access ACLs.
Default ACLs: A "template" of ACLs associated with a folder that determine the Access ACLs for any child items that are created under that folder. Files do not have Default ACLs.
Reference:
https://docs.microsoft.com/en-us/azure/data-lake-store/data-lake-store-access-control

NEW QUESTION 123
- (Exam Topic 3)
You are responsible for providing access to an Azure Data Lake Storage Gen2 account.
Your user account has contributor access to the storage account, and you have the application ID and access key.
You plan to use PolyBase to load data into an enterprise data warehouse in Azure Synapse Analytics. You need to configure PolyBase to connect the data warehouse to storage account.
Which three components should you create in sequence? To answer, move the appropriate components from the list of components to the answer area and arrange them in the correct order.

Components

a database scoped credential

an asymmetric key

an external data source

a database encryption key

an external file format

Answer Area

A. Mastered
B. Not Mastered

Answer: A

Explanation:

Components

a database scoped credential

an asymmetric key

an external data source

a database encryption key

an external file format

Answer Area

a database scoped credential

an external data source

an external file format

NEW QUESTION 126

- (Exam Topic 3)

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.

Values

CustomerKey

HASH

ROUND_ROBIN

REPLICATE

OrderDateKey

SalesOrderNumber

Answer Area

```
CREATE TABLE [dbo].[FactSales]
(
    [ProductKey]        int       NOT NULL
,   [OrderDateKey]      int       NOT NULL
,   [CustomerKey]       int       NOT NULL
,   [SalesOrderNumber]  nvarchar ( 20 )   NOT NULL
,   [OrderQuantity]     smallint    NOT NULL
,   [UnitPrice]         money       NOT NULL
)
WITH
(   CLUSTERED    COLUMNSTORE    INDEX
,   DISTRIBUTION =  [ Value ]   ([ProductKey])
,   PARTITION  ( [   Value   ] RANGE RIGHT FOR VALUES
            (20170101,20180101,20190101,20200101,20210101)
            )
)
```

A. Mastered
B. Not Mastered

Answer: A

Explanation:
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

NEW QUESTION 127
- (Exam Topic 3)
You are designing an enterprise data warehouse in Azure Synapse Analytics that will contain a table named Customers. Customers will contain credit card information.
You need to recommend a solution to provide salespeople with the ability to view all the entries in Customers. The solution must prevent all the salespeople from viewing or inferring the credit card information.
What should you include in the recommendation?

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

Answer: A

Explanation:
SQL Database dynamic data masking limits sensitive data exposure by masking it to non-privileged users. The Credit card masking method exposes the last four digits of the designated fields and adds a constant string as a prefix in the form of a credit card.
Example: XXXX-XXXX-XXXX-1234
Reference:
https://docs.microsoft.com/en-us/azure/sql-database/sql-database-dynamic-data-masking-get-started

NEW QUESTION 128
- (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

```
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
```

CAST

COLLATE

CONVERT

FLATTEN

PIVOT

UNPIVOT

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 133
- (Exam Topic 3)
You are designing a monitoring solution for a fleet of 500 vehicles. Each vehicle has a GPS tracking device that sends data to an Azure event hub once per minute.
You have a CSV file in an Azure Data Lake Storage Gen2 container. The file maintains the expected geographical area in which each vehicle should be.
You need to ensure that when a GPS position is outside the expected area, a message is added to another event hub for processing within 30 seconds. The solution must minimize cost.
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.

Service:
- An Azure Synapse Analytics Apache Spark pool
- An Azure Synapse Analytics serverless SQL pool
- Azure Data Factory
- Azure Stream Analytics

Window:
- Hopping
- No window
- Session
- Tumbling

Analysis type:
- Event pattern matching
- Lagged record comparison
- Point within polygon
- Polygon overlap

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: Azure Stream Analytics Box 2: Hopping
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 and be emitted more often than the window size. 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.
Box 3: Point within polygon Reference:

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

NEW QUESTION 137
- (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 in 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 139
- (Exam Topic 3)
You have an Azure subscription that contains an Azure Synapse Analytics dedicated SQL pool. You plan to deploy a solution that will analyze sales data and include the following:
• A table named Country that will contain 195 rows
• A table named Sales that will contain 100 million rows
• A query to identify total sales by country and customer from the past 30 days
You need to create the tables. The solution must maximize query performance.
How should you complete the script? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Answer Area

```
CREATE TABLE [dbo].[Sales]
(
        [OrderDate]        date         NOT NULL
,       [CustomerId] int NOT NULL
,       [CountryId] int NOT NULL
,       [Total] money NOT NULL
)

WITH
(
        DISTRIBUTION =    HASH([CustomerId])        ▼
                          HASH([CustomerId])
        CLUSTERED COLUMN  HASH([OrderDate])
)                         REPLICATE
CREATE TABLE [dbo].[Country]  ROUND_ROBIN
,
```

A. Mastered
B. Not Mastered

Answer: A

Explanation:

Answer Area

```
CREATE TABLE [dbo].[Sales]
(
        [OrderDate]         date          NOT NULL
,       [CustomerId] int NOT NULL
,       [CountryId] int NOT NULL
,       [Total] money NOT NULL
)

WITH
(
        DISTRIBUTION =    HASH([CustomerId])          ▼
                          HASH([CustomerId])
        CLUSTERED COLUMN  HASH([OrderDate])
)                         REPLICATE
                          ROUND_ROBIN
CREATE TABLE [dbo].[Country]
,
```

NEW QUESTION 142
- (Exam Topic 3)
You have a SQL pool in Azure Synapse.
A user reports that queries against the pool take longer than expected to complete. You need to add monitoring to the underlying storage to help diagnose the issue.
Which two metrics should you monitor? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

A. Cache used percentage
B. DWU Limit
C. Snapshot Storage Size
D. Active queries
E. Cache hit percentage

Answer: AE

Explanation:
A: Cache used is the sum of all bytes in the local SSD cache across all nodes and cache capacity is the sum of the storage capacity of the local SSD cache across all nodes.
E: Cache hits is the sum of all columnstore segments hits in the local SSD cache and cache miss is the columnstore segments misses in the local SSD cache summed across all nodes
Reference:
https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-concept-resou

NEW QUESTION 144
- (Exam Topic 3)
You have a Microsoft SQL Server database that uses a third normal form schema.
You plan to migrate the data in the database to a star schema in an Azure Synapse Analytics dedicated SQI pool.
You need to design the dimension tables. The solution must optimize read operations.
What should you include in the solution? to answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Transform data for the dimension tables by: ▼
Maintaining to a third normal form
Normalizing to a fourth normal form
Denormalizing to a second normal form

For the primary key columns in the dimension tables, use: ▼
New IDENTITY columns
A new computed column
The business key column from the source sys

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Text, table Description automatically generated
Box 1: Denormalize to a second normal form
Denormalization is the process of transforming higher normal forms to lower normal forms via storing the join of higher normal form relations as a base relation.

Denormalization increases the performance in data retrieval at cost of bringing update anomalies to a database.
Box 2: New identity columns
The collapsing relations strategy can be used in this step to collapse classification entities into component entities to obtain at dimension tables with single-part keys that connect directly to the fact table. The single-part key is a surrogate key generated to ensure it remains unique over time.
Example:
Diagram Description automatically generated

DimDate
- ? DateKey
- SqlDate
- DayNumberOfWeek
- DayNumberOfMonth
- DayNumberOfYear
- WeekNumberOfYear
- MonthNumberOfYear
- CalendarQuarter
- CalendarYear
- CalendarSemester

FactSales

DimCustomer
- ? CustomerKey
- CustomerNumber
- CustomerCity
- CustomerState

DimProduct
- ? ProductKey
- ProductNumber
- ProductDescription
- UnitPrice
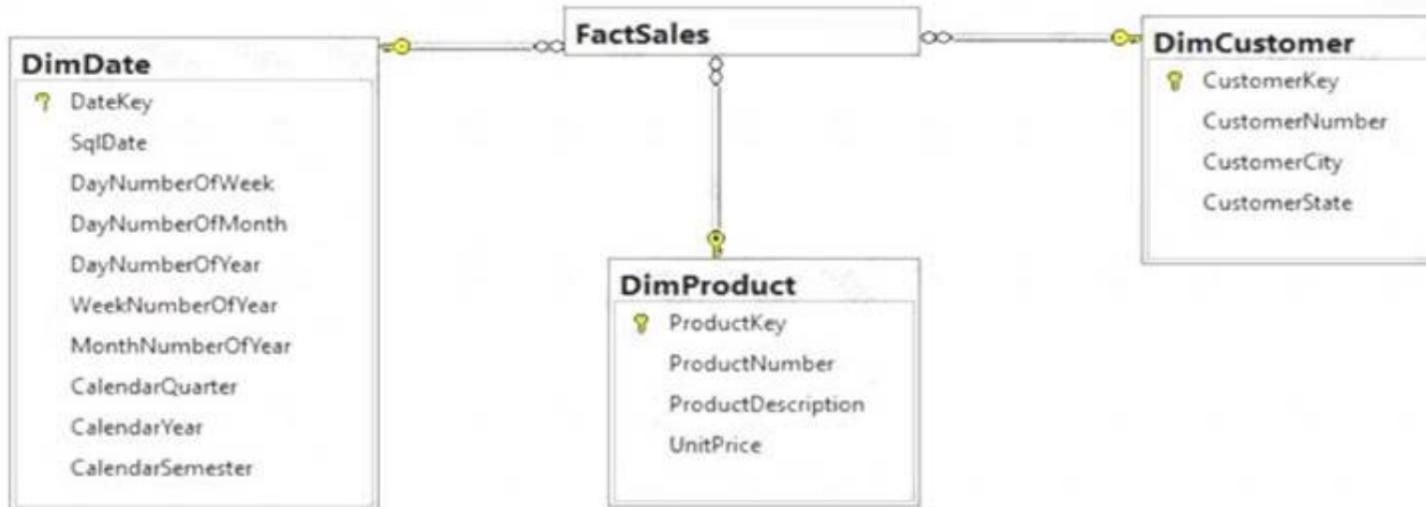
Note: A surrogate key on a table is a column with a unique identifier for each row. The key is not generated from the table data. Data modelers like to create surrogate keys on their tables when they design data warehouse models. You can use the IDENTITY property to achieve this goal simply and effectively without affecting load performance.
Reference:
https://www.mssqltips.com/sqlservertip/5614/explore-the-role-of-normal-forms-in-dimensional-modeling/ https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-identity

NEW QUESTION 149
- (Exam Topic 3)
You have an Azure subscription that contains the following resources:

≫ An Azure Active Directory (Azure AD) tenant that contains a security group named Group1

≫ An Azure Synapse Analytics SQL pool named Pool1
You need to control the access of Group1 to specific columns and rows in a table in Pool1.
Which Transact-SQL commands should you use? To answer, select the appropriate options in the answer area.

To control access to the columns:
- CREATE CRYPTOGRAPHIC PROVIDER
- CREATE PARTITION FUNCTION
- CREATE SECURITY POLICY
- GRANT

To control access to the rows:
- CREATE CRYPTOGRAPHIC PROVIDER
- CREATE PARTITION FUNCTION
- CREATE SECURITY POLICY
- GRANT

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Text Description automatically generated
Box 1: GRANT
You can implement column-level security with the GRANT T-SQL statement. Box 2: CREATE SECURITY POLICY
Implement Row Level Security by using the CREATE SECURITY POLICY Transact-SQL statement Reference:
https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/column-level-security

NEW QUESTION 151
- (Exam Topic 3)
You are designing an Azure Databricks table. The table will ingest an average of 20 million streaming events per day.
You need to persist the events in the table for use in incremental load pipeline jobs in Azure Databricks. The solution must minimize storage costs and incremental load times.
What should you include in the solution?

A. Partition by DateTime fields.
B. Sink to Azure Queue storage.
C. Include a watermark column.
D. Use a JSON format for physical data storage.

Answer: A

Explanation:
The Databricks ABS-AQS connector uses Azure Queue Storage (AQS) to provide an optimized file source that lets you find new files written to an Azure Blob storage (ABS) container without repeatedly listing all of the files.
This provides two major advantages:

≫ Lower costs: no more costly LIST API requests made to ABS.
Reference:
https://docs.microsoft.com/en-us/azure/databricks/spark/latest/structured-streaming/aqs

NEW QUESTION 156
- (Exam Topic 3)
You have an Azure Blob storage account that contains a folder. The folder contains 120,000 files. Each file contains 62 columns.
Each day, 1,500 new files are added to the folder.
You plan to incrementally load five data columns from each new file into an Azure Synapse Analytics workspace.
You need to minimize how long it takes to perform the incremental loads.
What should you use to store the files and format?

Storage: [▼]
Multiple blob storage accounts
Multiple containers in the blob storage account
Timeslice partitioning in the folders

These are the se

Format: [▼]
Apache Parquet
CSV
JSON
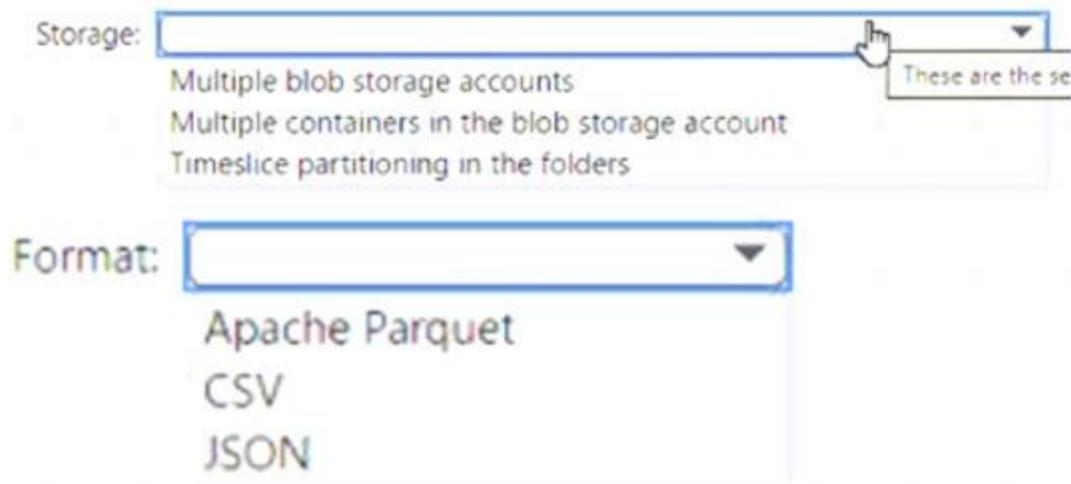
A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1 = timeslice partitioning in the foldersThis means that you should organize your files into folders based on a time attribute, such as year, month, day, or hour. For example, you can have a folder structure like
/yyyy/mm/dd/file.csv. This way, you can easily identify and load only the new files that are added each day by using a time filter in your Azure Synapse pipeline12.
Timeslice partitioning can also improve the performance of data loading and querying by reducing the number of files that need to be scanned
Box = 2 Apache Parquet This is because Parquet is a columnar file format that can efficiently store and compress data with many columns. Parquet files can also be partitioned by a time attribute, which can improve the performance of incremental loading and querying by reducing the number of files that need to be scanned1 23. Parquet files are supported by both dedicated SQL pool and serverless SQL pool in Azure Synapse Analytics2.

NEW QUESTION 159
- (Exam Topic 3)
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?

A. JSON
B. Parquet
C. Avro
D. CSV

Answer: B

Explanation:
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

NEW QUESTION 164
- (Exam Topic 3)
You have an Azure Synapse Analytics dedicated SQL pool.
You need to create a table named FactInternetSales that will be a large fact table in a dimensional model. FactInternetSales will contain 100 million rows and two columns named SalesAmount and OrderQuantity. Queries executed on FactInternetSales will aggregate the values in SalesAmount and OrderQuantity from the last year for a specific product. The solution must minimize the data size and query execution time.
How should you complete the code? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Answer Area

```
CREATE TABLE [dbo].[FactInternetSales]
( [ProductKey] int NOT NULL
, [OrderDateKey] int NOT NULL
, [CustomerKey] int NOT NULL
, [PromotionKey] int NOT NULL
, [SalesOrderNumber] nvarchar(20) NOT NULL
, [OrderQuantity] smallint NOT NULL
, [UnitPrice] money NOT NULL
, [SalesAmount] money NOT NULL
)
WITH
```

| (CLUSTERED COLUMNSTORE INDEX |
| (CLUSTERED INDEX ([OrderDateKey]) |
| (HEAP |
| (INDEX on [ProductKey] |

```
, DISTRIBUTION =
);
```

| Hash([OrderDateKey]) |
| Hash([ProductKey]) |
| REPLICATE |
| ROUND_ROBIN |

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: (CLUSTERED COLUMNSTORE INDEX CLUSTERED COLUMNSTORE INDEX
Columnstore indexes are the standard for storing and querying large data warehousing fact tables. This index uses column-based data storage and query processing to achieve gains up to 10 times the query performance in your data warehouse over traditional row-oriented storage. You can also achieve gains up to 10 times the data compression over the uncompressed data size. Beginning with SQL Server 2016 (13.x) SP1, columnstore indexes enable operational analytics: the ability to run performant real-time analytics on a transactional workload.
Note: Clustered columnstore index
A clustered columnstore index is the physical storage for the entire table. Diagram Description automatically generated
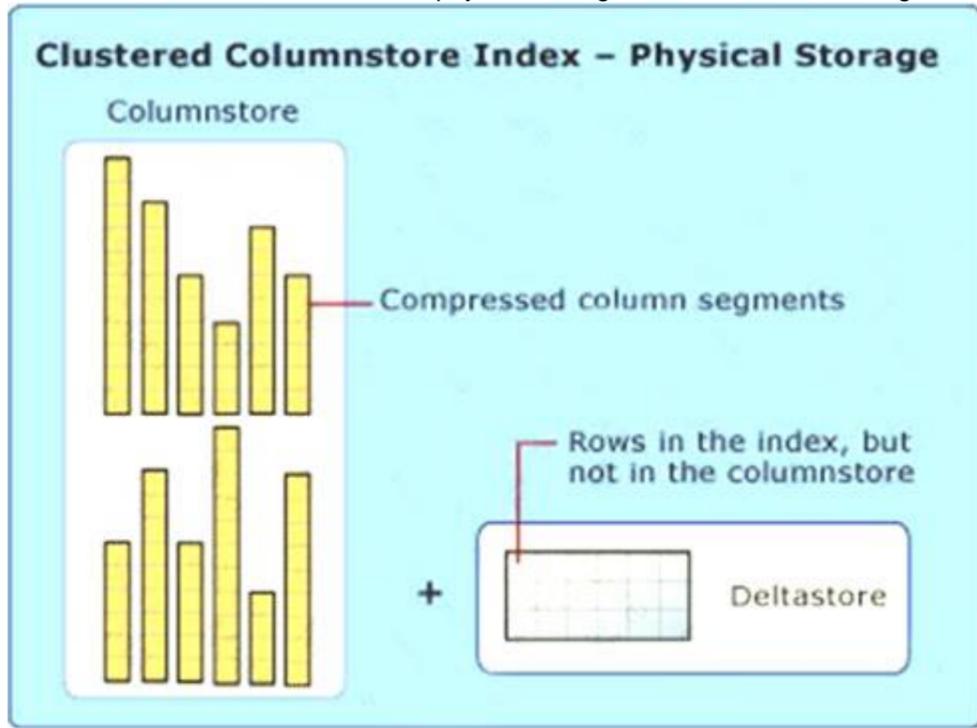
To reduce fragmentation of the column segments and improve performance, the columnstore index might store some data temporarily into a clustered index called a deltastore and a B-tree list of IDs for deleted rows. The deltastore operations are handled behind the scenes. To return the correct query results, the clustered columnstore index combines query results from both the columnstore and the deltastore.
Box 2: HASH([ProductKey])
A hash distributed table distributes rows based on the value in the distribution column. A hash distributed table is designed to achieve high performance for queries on large tables.
Choose a distribution column with data that distributes evenly
Reference: https://docs.microsoft.com/en-us/sql/relational-databases/indexes/columnstore-indexes-overview https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-overvie https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-distribu

NEW QUESTION 166
- (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

```
                              ▼  applications
┌─────────────────────────┐
│ 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://contoso1.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 171
- (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 176
- (Exam Topic 3)

You need to implement an Azure Databricks cluster that automatically connects to Azure Data lake Storage Gen2 by using Azure Active Directory (Azure AD) integration. How should you configure the new clutter? To answer, select the appropriate options in the answers area. NOTE: Each correct selection is worth one point.

Answer Area

Tier:	Premium / Standard
Advanced option to enable:	Azure Data Lake Storage Credential Passthrough / Table Access Control

A. Mastered
B. Not Mastered

Answer: A

Explanation:
https://docs.azuredatabricks.net/spark/latest/data-sources/azure/adls-passthrough.html
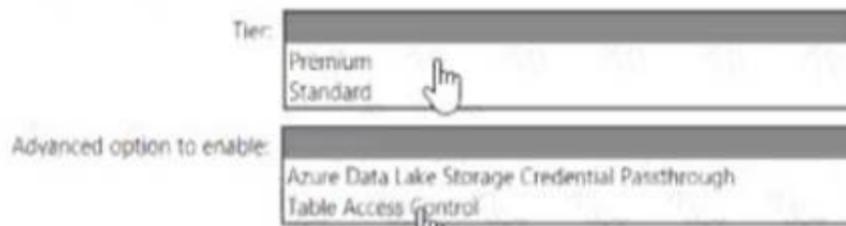
NEW QUESTION 178
- (Exam Topic 3)
You are implementing a star schema in an Azure Synapse Analytics dedicated SQL pool. You plan to create a table named DimProduct.
DimProduct must be a Type 3 slowly changing dimension (SCO) table that meets the following requirements:
• The values in two columns named ProductKey and ProductSourceID will remain the same.
• The values in three columns named ProductName, ProductDescription, and Color can change. You need to add additional columns to complete the following table definition.

```
CREATE TABLE [dbo].[dimproduct]
  (
     [ProductKey]          INT NOT NULL,
     [ProductSourceID]     INT NOT NULL,
     [ProductName]         NVARCHAR(100) NOT NULL,
     [ProductDescription]  NVARCHAR(2000) NOT NULL,
     [Color]               NVARCHAR(50) NOT NULL
  )
WITH
(
     DISTRIBUTION = REPLICATE,
     CLUSTERED COLUMNSTORE INDEX
);
```

A)
```
[OriginalProductDescription] NVARCHAR(2000) NOT NULL
```

B)
```
[IsCurrentRow] [bit] NOT NULL
```

C)
```
[EffectiveStartDate] [datetime] NOT NULL
```

D)
```
[EffectiveEndDate] [datetime] NOT NULL
```

E)
```
[OriginalProductName] NVARCHAR(100) NULL
```

F)
```
[OriginalColor] NVARCHAR(50) NOT NULL
```

A. Option A
B. Option B
C. Option C
D. Option D
E. Option E
F. Option F

Answer: ABC

NEW QUESTION 180
- (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 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?

A. Yes
B. No

Answer: B

Explanation:
Must use an Azure Data Factory, not an Azure Databricks job. Reference:
https://docs.microsoft.com/en-US/azure/data-factory/transform-data

NEW QUESTION 183
- (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 184
- (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 186
- (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.
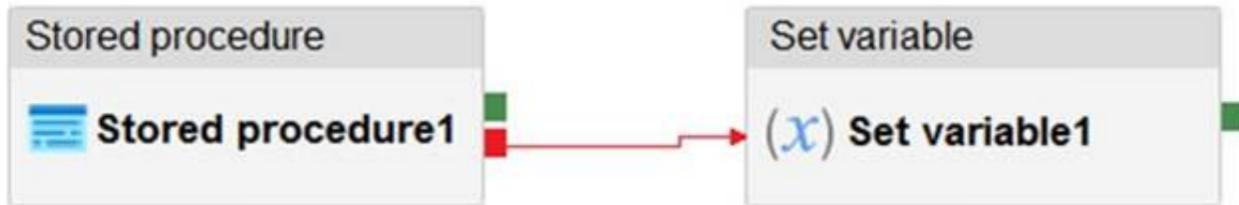
Pipeline2 has the activities shown in the following exhibit.

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
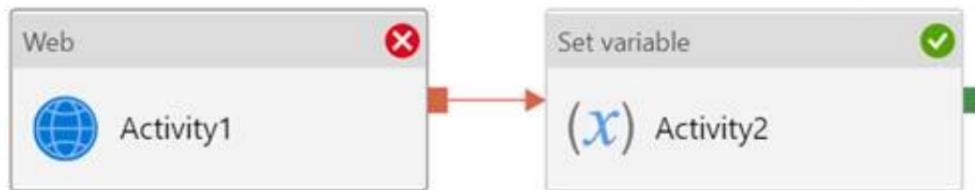
The failure dependency means this pipeline reports success. Note:
If we have a pipeline containing Activity1 and Activity2, and Activity2 has a success dependency on Activity1, it will only execute if Activity1 is successful. In this scenario, if Activity1 fails, the pipeline will fail.
Reference:
https://datasavvy.me/category/azure-data-factory/

NEW QUESTION 191
- (Exam Topic 3)
You have a C# application that process data from an Azure IoT hub and performs complex transformations. You need to replace the application with a real-time solution. The solution must reuse as much code as possible from the existing application.

A. Azure Databricks
B. Azure Event Grid
C. Azure Stream Analytics
D. Azure Data Factory

Answer: C

Explanation:
Azure Stream Analytics on IoT Edge empowers developers to deploy near-real-time analytical intelligence closer to IoT devices so that they can unlock the full value of device-generated data. UDF are available in C# for IoT Edge jobs
Azure Stream Analytics on IoT Edge runs within the Azure IoT Edge framework. Once the job is created in Stream Analytics, you can deploy and manage it using IoT Hub.
References:
https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-edge

NEW QUESTION 192
......

Relate Links

100% Pass Your DP-203 Exam with Exambible Prep Materials

https://www.exambible.com/DP-203-exam/

Contact us

We are proud of our high-quality customer service, which serves you around the clock 24/7.

Viste - https://www.exambible.com/