Microsoft

Exam Questions DP-203

Data Engineering on Microsoft Azure

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

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)
You are designing a data mart for the human resources (MR) department at your company. The data mart will contain information and employee transactions.
From a source system you have a flat extract that has the following fields:
• EmployeeID
• FirstName
• LastName
• Recipient
• GrossArnount
• TransactionID
• GovernmentID
• NetAmountPaid
• TransactionDate
You need to design a start schema data model in an Azure Synapse analytics dedicated SQL pool for the data mart.
Which two tables should you create? Each Correct answer present part of the solution.

A. a dimension table for employee
B. a fabric for Employee
C. a dimension table far EmployeeTransaction
D. a dimension table for Transaction
E. a fact table for Transaction

Answer: AE

Explanation:
Reference:
https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-overvie

NEW QUESTION 4
- (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.

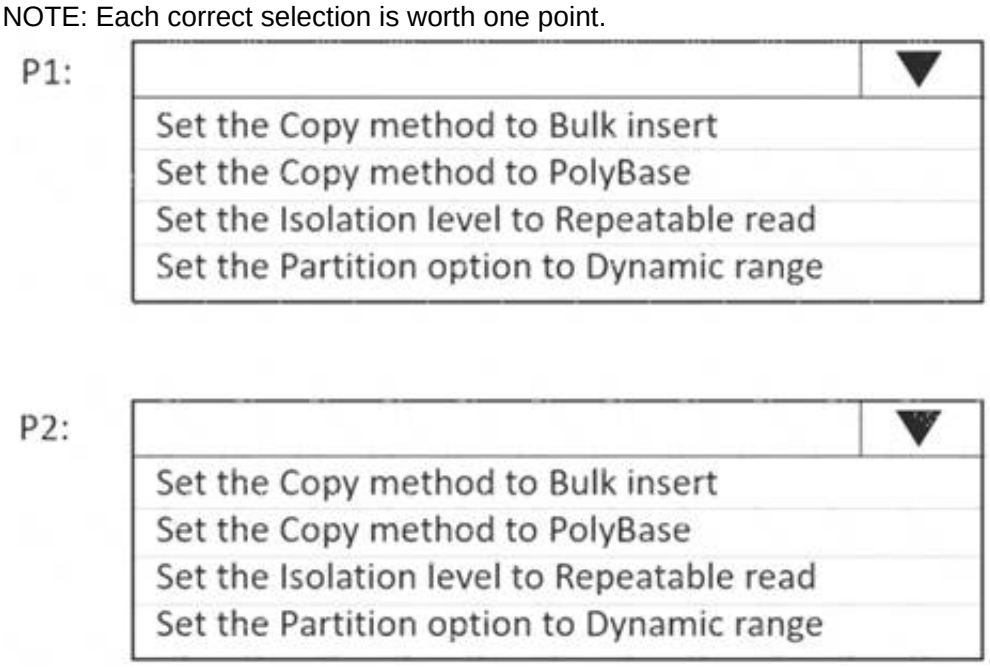
Answer: DF

Explanation:
Reference:
https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-parallelization

NEW QUESTION 5
- (Exam Topic 3)
HOTSPOT
You have an Azure Data Factory instance named ADF1 and two Azure Synapse Analytics workspaces named WS1 and WS2.
ADF1 contains the following pipelines:

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

You need to configure P1 and P2 to maximize parallelism and performance.
Which dataset settings should you configure for the copy activity if each pipeline? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.



A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: Set the Copy method to PolyBase
While SQL pool supports many loading methods including non-Polybase options such as BCP and SQL BulkCopy API, the fastest and most scalable way to load data is through PolyBase. PolyBase is a technology that accesses external data stored in Azure Blob storage or Azure Data Lake Store via the T-SQL language.
Box 2: Set the Copy method to Bulk insert
Polybase not possible for text files. Have to use Bulk insert. Reference:
https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/load-data-overview

NEW QUESTION 6
- (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 7
- (Exam Topic 3)
You are designing a dimension table for a data warehouse. The table will track the value of the dimension attributes over time and preserve the history of the data by adding new rows as the data changes.
Which type of slowly changing dimension (SCD) should use?

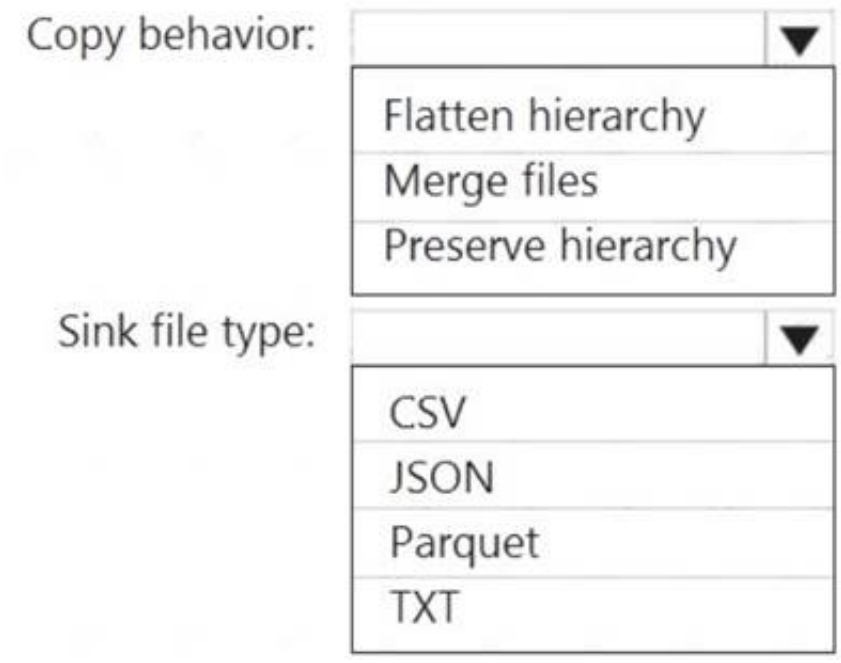
A. Type 0
B. Type 1
C. Type 2
D. Type 3

Answer: C

Explanation:

Type 2 - Creating a new additional record. In this methodology all history of dimension changes is kept in the database. You capture attribute change by adding a new row with a new surrogate key to the dimension table. Both the prior and new rows contain as attributes the natural key(or other durable identifier). Also 'effective date' and 'current indicator' columns are used in this method. There could be only one record with current indicator set to 'Y'. For 'effective date' columns, i.e. start_date and end_date, the end_date for current record usually is set to value 9999-12-31. Introducing changes to the dimensional model in type 2 could be very expensive database operation so it is not recommended to use it in dimensions where a new attribute could be added in the future.
https://www.datawarehouse4u.info/SCD-Slowly-Changing-Dimensions.html

NEW QUESTION 8
- (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.

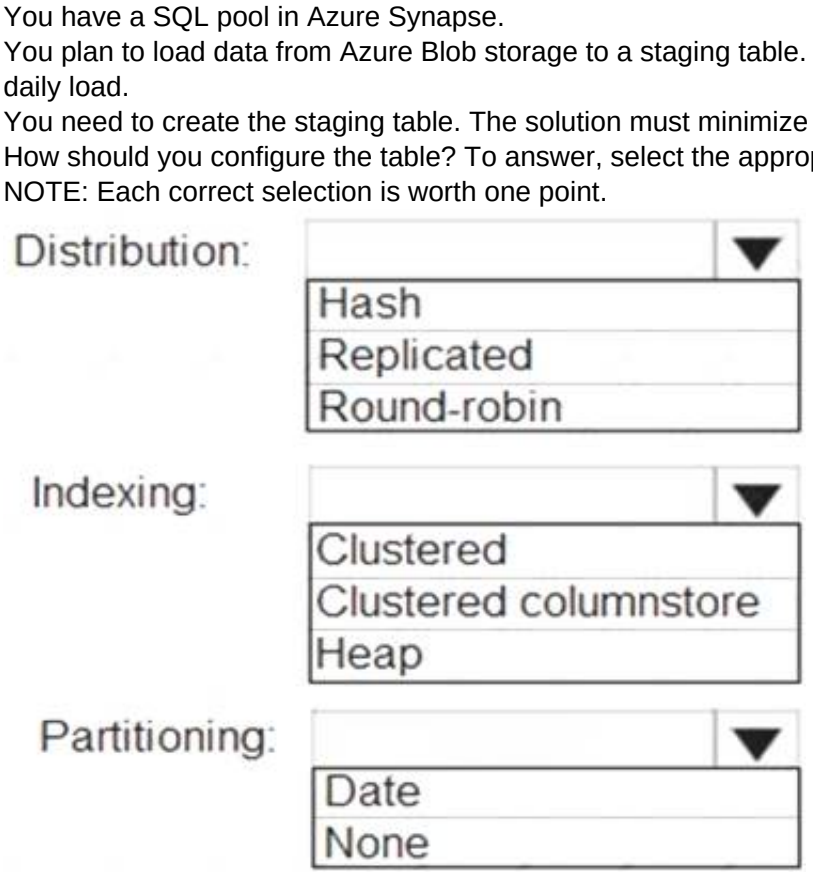


A. Mastered
B. Not Mastered

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 9
- (Exam Topic 3)
You have a SQL pool in Azure Synapse.
You plan to load data from Azure Blob storage to a staging table. Approximately 1 million rows of data will be loaded daily. The table will be truncated before each daily load.
You need to create the staging table. The solution must minimize how long it takes to load the data to the staging table.
How should you configure the table? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.



A. Mastered

B. Not Mastered

Answer: A

Explanation:
Graphical user interface, application, table Description automatically generated
Box 1: Hash
Hash-distributed tables improve query performance on large fact tables. They can have very large numbers of rows and still achieve high performance.
Box 2: Clustered columnstore
When creating partitions on clustered columnstore tables, it is important to consider how many rows belong to each partition. For optimal compression and performance of clustered columnstore tables, a minimum of 1 million rows per distribution and partition is needed.
Box 3: Date
Table partitions enable you to divide your data into smaller groups of data. In most cases, table partitions are created on a date column.
Partition switching can be used to quickly remove or replace a section of a table. Reference:
https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-partitio https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-distribu

NEW QUESTION 10
- (Exam Topic 3)
You have an Azure Synapse Analytics dedicated SQL pool that contains a table named Table1. Table1 contains the following:
- One billion rows
- A clustered columnstore index
- A hash-distributed column named Product Key
- A column named Sales Date that is of the date data type and cannot be null Thirty million rows will be added to Table1 each month.

You need to partition Table1 based on the Sales Date column. The solution must optimize query performance and data loading.
How often should you create a partition?

A. once per month
B. once per year
C. once per day
D. once per week

Answer: B

Explanation:
Need a minimum 1 million rows per distribution. Each table is 60 distributions. 30 millions rows is added each month. Need 2 months to get a minimum of 1 million rows per distribution in a new partition.
Note: When creating partitions on clustered columnstore tables, it is important to consider how many rows belong to each partition. For optimal compression and performance of clustered columnstore tables, a minimum of 1 million rows per distribution and partition is needed. Before partitions are created, dedicated SQL pool already divides each table into 60 distributions.
Any partitioning added to a table is in addition to the distributions created behind the scenes. Using this example, if the sales fact table contained 36 monthly partitions, and given that a dedicated SQL pool has 60 distributions, then the sales fact table should contain 60 million rows per month, or 2.1 billion rows when all months are populated. If a table contains fewer than the recommended minimum number of rows per partition, consider using fewer partitions in order to increase the number of rows per partition.
Reference:
https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-partitio

NEW QUESTION 10
- (Exam Topic 3)
You have an Azure Factory instance named DF1 that contains a pipeline named PL1.PL1 includes a tumbling window trigger.
You create five clones of PL1. You configure each clone pipeline to use a different data source.
You need to ensure that the execution schedules of the clone pipeline match the execution schedule of PL1. What should you do?

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

Answer: B

NEW QUESTION 14
- (Exam Topic 3)
You have an Azure Data Lake Storage account that contains a staging zone.
You need to design a daily process to ingest incremental data from the staging zone, transform the data by executing an R script, and then insert the transformed data into a data warehouse in Azure Synapse Analytics.
Solution: You use an Azure Data Factory schedule trigger to execute a pipeline that executes mapping data Flow, and then inserts the data info the data warehouse.
Does this meet the goal?

A. Yes
B. No

Answer: B

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

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

NEW QUESTION 18
- (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_YYYMMMDD.parquet
B. \Depdftment\DataSource\YYY\MM\DataFile_YYYYMMDD.parquet
C. \DD\MM\YYYY\Department\DataSource\DataFile_DDMMYY.parquet
D. \DataSource\Department\YYYYMM\DataFile_YYYYMMDD.parquet

Answer: B

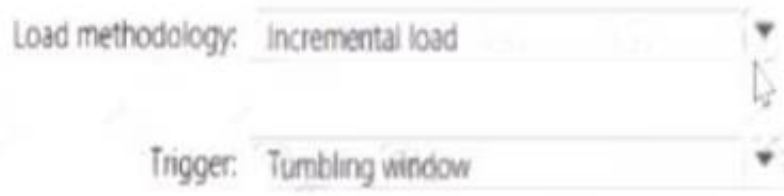
Explanation:
Department top level in the hierarchy to simplify security management.
Month (MM) at the leaf/bottom level to support fast data retrieval for data from the current month.

NEW QUESTION 23
- (Exam Topic 3)
You have an Azure Storage account that generates 200.000 new files daily. The file names have a format of (YYY)/(MM)/(DD)/|HH])/(CustornerID).csv.
You need to design an Azure Data Factory solution that will toad new data from the storage account to an Azure Data lake once hourly. The solution must minimize load times and costs.
How should you configure the solution? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

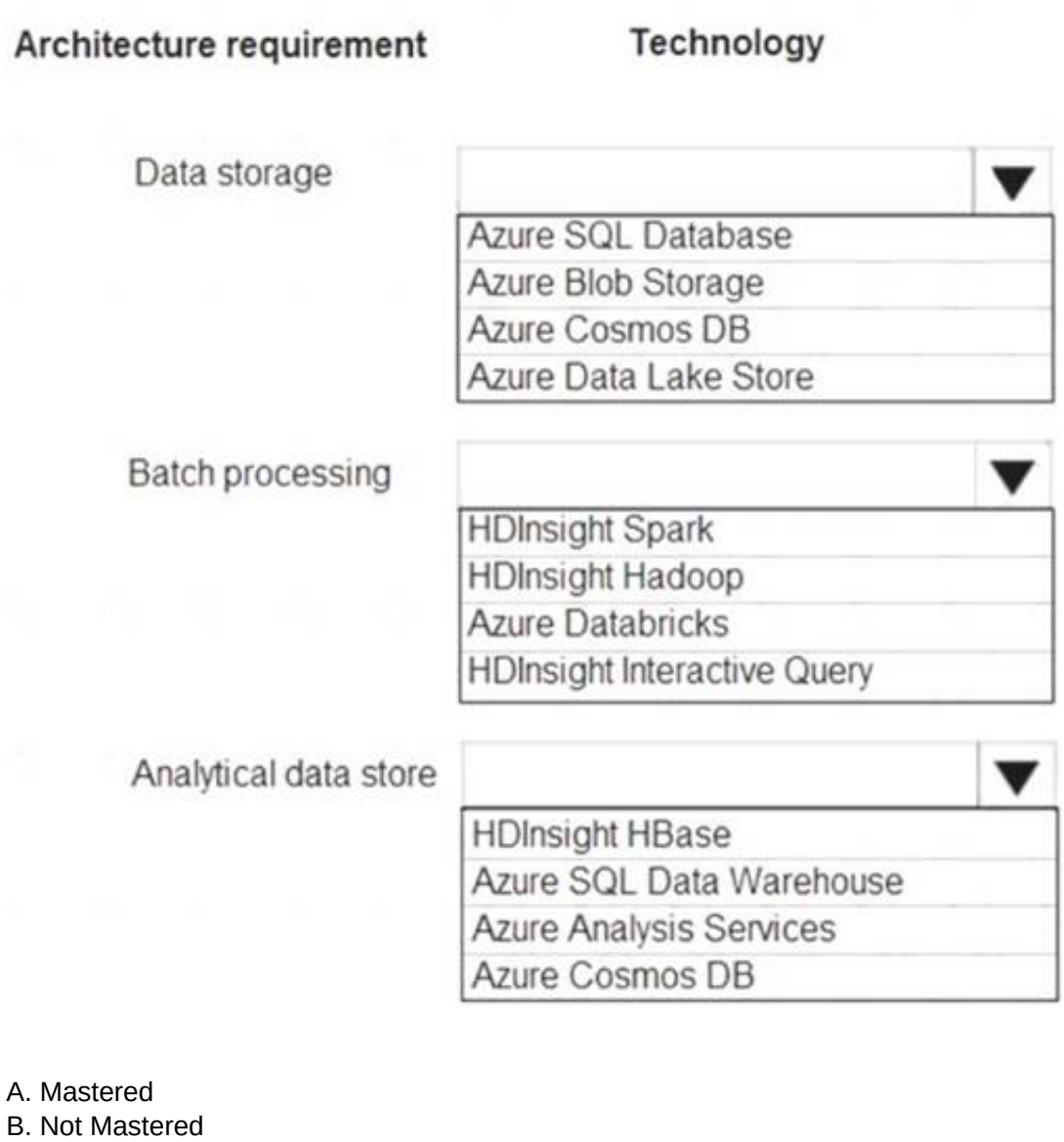
A. Mastered
B. Not Mastered

Answer: A

Explanation:
Answer Area



NEW QUESTION 27
- (Exam Topic 3)
You are developing a solution using a Lambda architecture on Microsoft Azure. The data at test layer must meet the following requirements:
Data storage:
•Serve as a repository (or high volumes of large files in various formats.
•Implement optimized storage for big data analytics workloads.
•Ensure that data can be organized using a hierarchical structure. Batch processing:
•Use a managed solution for in-memory computation processing.
•Natively support Scala, Python, and R programming languages.
•Provide the ability to resize and terminate the cluster automatically. Analytical data store:
•Support parallel processing.
•Use columnar storage.
•Support SQL-based languages.
You need to identify the correct technologies to build the Lambda architecture.
Which technologies should you use? To answer, select the appropriate options in the answer area NOTE: Each correct selection is worth one point.



A. Mastered
B. Not Mastered

Answer: A

Explanation:
Data storage: Azure Data Lake Store
A key mechanism that allows Azure Data Lake Storage Gen2 to provide file system performance at object storage scale and prices is the addition of a hierarchical namespace. This allows the collection of objects/files within an account to be organized into a hierarchy of directories and nested subdirectories in the same way that the file system on your computer is organized. With the hierarchical namespace enabled, a storage account becomes capable of providing the scalability and cost-effectiveness of object storage, with file system semantics that are familiar to analytics engines and frameworks.
Batch processing: HD Insight Spark
Aparch Spark is an open-source, parallel-processing framework that supports in-memory processing to boost the performance of big-data analysis applications. HDInsight is a managed Hadoop service. Use it deploy and manage Hadoop clusters in Azure. For batch processing, you can use Spark, Hive, Hive LLAP, MapReduce.
Languages: R, Python, Java, Scala, SQL Analytic data store: SQL Data Warehouse
SQL Data Warehouse is a cloud-based Enterprise Data Warehouse (EDW) that uses Massively Parallel Processing (MPP).
SQL Data Warehouse stores data into relational tables with columnar storage. References:
https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-namespace https://docs.microsoft.com/en-us/azure/architecture/data-guide/technology-choices/batch-processing https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-overview-what-is

NEW QUESTION 31
- (Exam Topic 3)
You have an Azure Databricks workspace named workspace! in the Standard pricing tier. Workspace1 contains an all-purpose cluster named cluster). You need to reduce the time it takes for cluster 1 to start and scale up. The solution must minimize costs. What should you do first?

A. Upgrade workspace! to the Premium pricing tier.
B. Create a cluster policy in workspace1.
C. Create a pool in workspace1.
D. Configure a global init script for workspace1.

Answer: C

Explanation:
You can use Databricks Pools to Speed up your Data Pipelines and Scale Clusters Quickly.
Databricks Pools, a managed cache of virtual machine instances that enables clusters to start and scale 4 times faster.
Reference:
https://databricks.com/blog/2019/11/11/databricks-pools-speed-up-data-pipelines.html

NEW QUESTION 32
- (Exam Topic 3)
You use Azure Stream Analytics to receive Twitter data from Azure Event Hubs and to output the data to an Azure Blob storage account. You need to output the count of tweets from the last five minutes every minute. Which windowing function should you use?

A. Sliding
B. Session
C. Tumbling
D. Hopping

Answer: D

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

NEW QUESTION 36
- (Exam Topic 3)
You develop a dataset named DBTBL1 by using Azure Databricks. DBTBL1 contains the following columns:
- SensorTypeID
- GeographyRegionID
- Year
- Month
- Day
- Hour
- Minute
- Temperature
- WindSpeed
- Other

You need to store the data to support daily incremental load pipelines that vary for each GeographyRegionID. The solution must minimize storage costs.
How should you complete the code? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

```
df.write
    [ .bucketBy | .format | .partitionBy | .sortBy ]
    [ ("*") | ("GeographyRegionID") | ("GeographyRegionID", "Year", "Month", "Day") | ("Year", "Month", "Day", "GeographyRegionID") ]
    .mode("append")
    [ .csv("/DBTBL1") | .json("/DBTBL1") | .parquet("/DBTBL1") | .saveAsTable("/DBTBL1") ]
```

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Graphical user interface, text, application Description automatically generated

NEW QUESTION 37
- (Exam Topic 3)
You have an Azure Synapse Analytics dedicated SQL pool named Pool1. Pool1 contains a fact table named Tablet. Table1 contains sales data. Sixty-five million rows of data are added to Table1 monthly.
At the end of each month, you need to remove data that is older than 36 months. The solution must minimize how long it takes to remove the data.
How should you partition Table1, and how should you remove the old data? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Answer Area

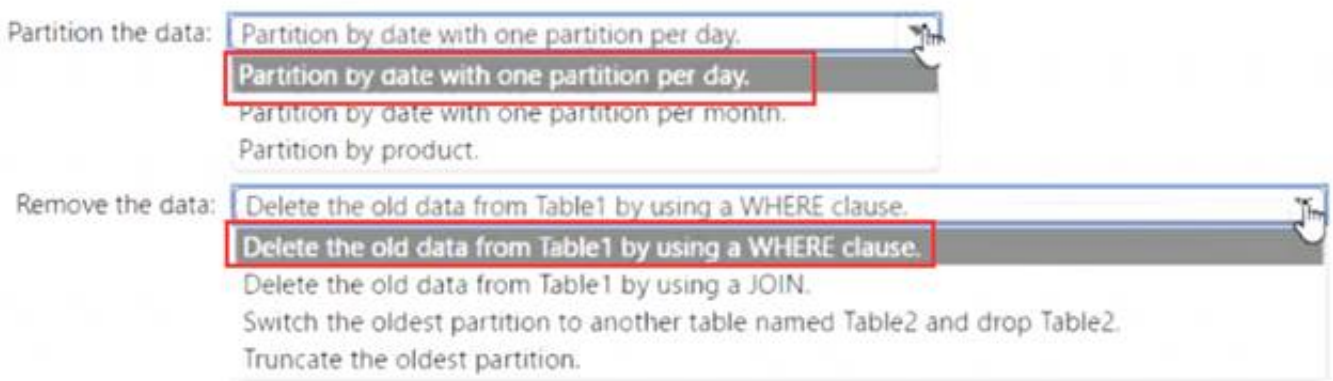
Partition the data: Partition by date with one partition per day.
- Partition by date with one partition per day.
- Partition by date with one partition per month.
- Partition by product.

Remove the data: Delete the old data from Table1 by using a WHERE clause.
- Delete the old data from Table1 by using a WHERE clause.
- Delete the old data from Table1 by using a JOIN.
- Switch the oldest partition to another table named Table2 and drop Table2.
- Truncate the oldest partition.

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Answer Area



NEW QUESTION 42
- (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 46
- (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.

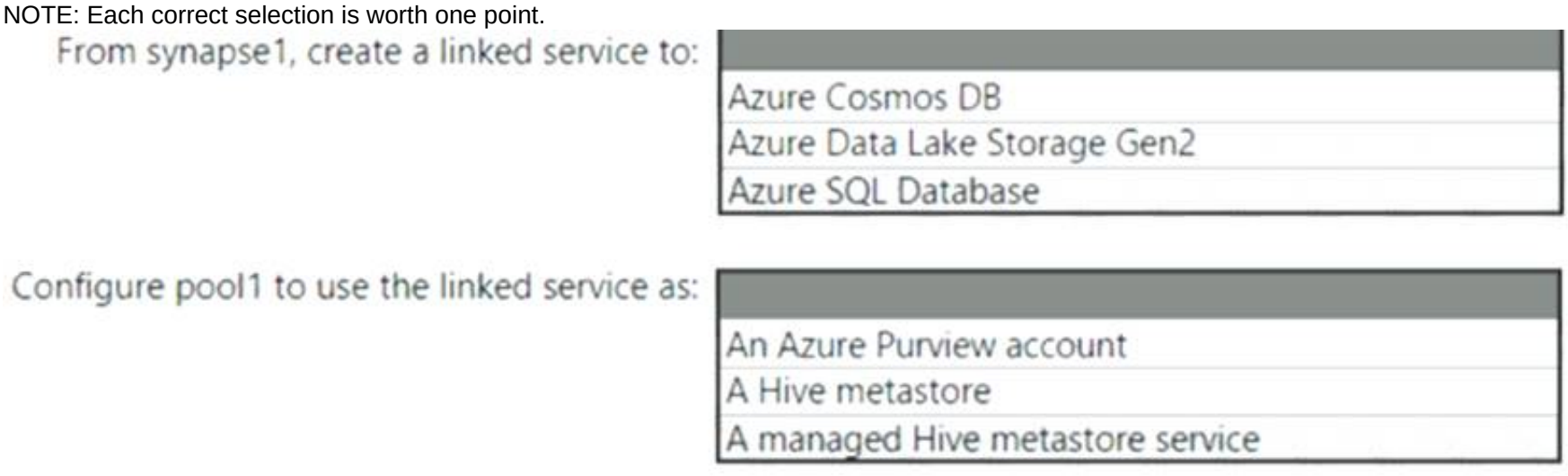
You need to create the Databricks clusters for the workloads.
Solution: You create a Standard cluster for each data scientist, a High Concurrency cluster for the data engineers, and a Standard cluster for the jobs.
Does this meet the goal?

A. Yes
B. No

Answer: B

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

NEW QUESTION 48
- (Exam Topic 3)
You have an Azure subscription that contains an Azure Databricks workspace named databricks1 and an Azure Synapse Analytics workspace named synapse1.
The synapse1 workspace contains an Apache Spark pool named pool1.
You need to share an Apache Hive catalog of pool1 with databricks1.
What should you do? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.



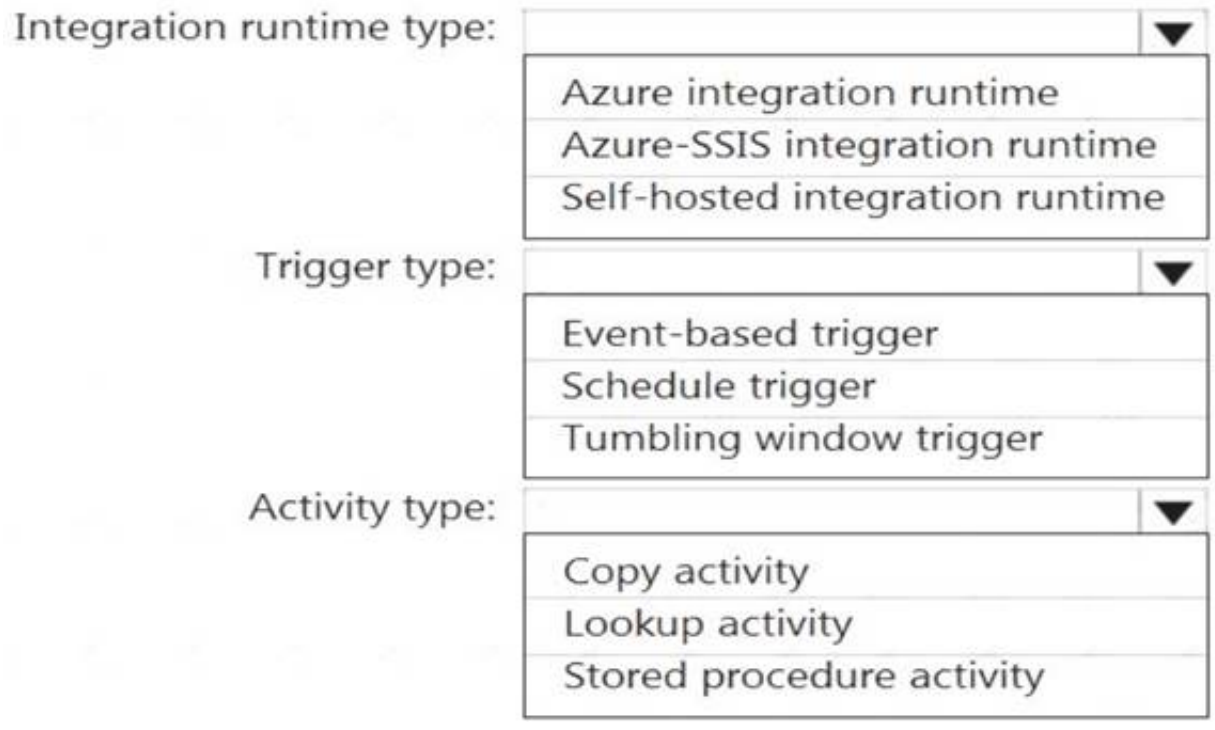
A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: Azure SQL Database
Use external Hive Metastore for Synapse Spark Pool
Azure Synapse Analytics allows Apache Spark pools in the same workspace to share a managed HMS (Hive Metastore) compatible metastore as their catalog.
Set up linked service to Hive Metastore
Follow below steps to set up a linked service to the external Hive Metastore in Synapse workspace.
- Set up Hive Metastore linked service
- Choose Azure SQL Database or Azure Database for MySQL based on your database type, click Continue.
- Provide Name of the linked service. Record the name of the linked service, this info will be used to configure Spark shortly.
- You can either select Azure SQL Database/Azure Database for MySQL for the external Hive Metastore from Azure subscription list, or enter the info manually.
- Provide User name and Password to set up the connection.
- Test connection to verify the username and password.
- Click Create to create the linked service.

Box 2: A Hive Metastore
nce: https://docs.microsoft.com/en-us/azure/synapse-analytics/spark/apache-spark-external-metastore

NEW QUESTION 51
- (Exam Topic 2)
Which Azure Data Factory components should you recommend using together to import the daily inventory data from the SQL server to Azure Data Lake Storage?
To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.



A. Mastered
B. Not Mastered

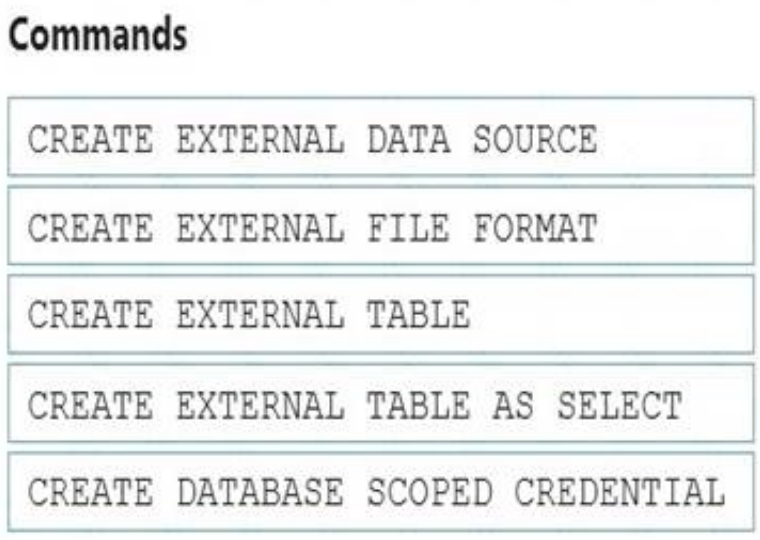
Answer: A

Explanation:
Box 1: Self-hosted integration runtime

A self-hosted IR is capable of running copy activity between a cloud data stores and a data store in private network.
Box 2: Schedule trigger Schedule every 8 hours Box 3: Copy activity Scenario:

- Customer data, including name, contact information, and loyalty number, comes from Salesforce and can be imported into Azure once every eight hours. Row modified dates are not trusted in the source table.
- Product data, including product ID, name, and category, comes from Salesforce and can be imported into Azure once every eight hours. Row modified dates are not trusted in the source table.

NEW QUESTION 54
- (Exam Topic 1)
You need to ensure that the Twitter feed data can be analyzed in the dedicated SQL pool. The solution must meet the customer sentiment analytics requirements.
Which three Transaction-SQL DDL commands should you run in sequence? To answer, move the appropriate commands from the list of commands to the answer area and arrange them in the correct order.
NOTE: More than one order of answer choices is correct. You will receive credit for any of the correct orders you select.

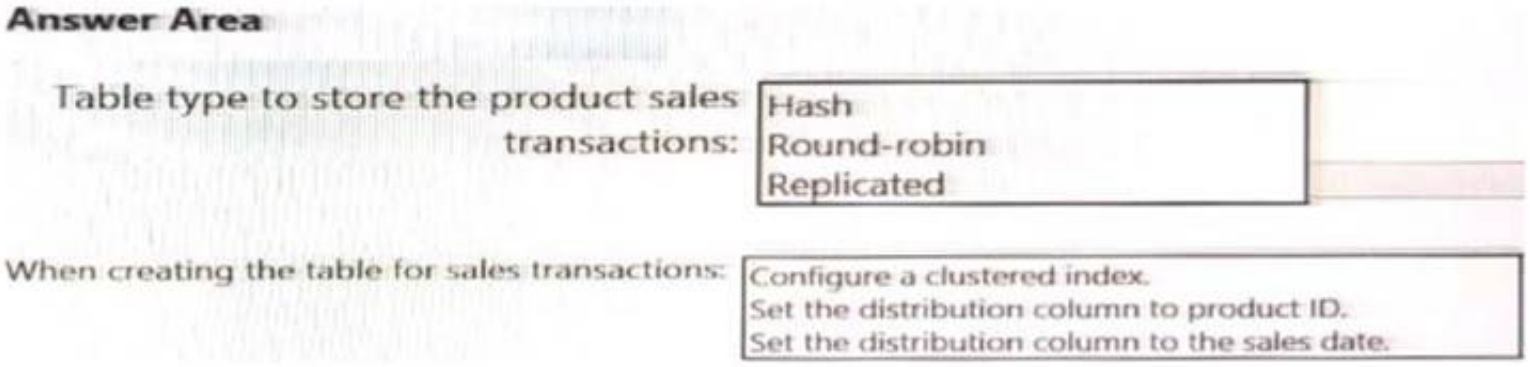


A. Mastered
B. Not Mastered

Answer: A

Explanation:
Scenario: Allow Contoso users to use PolyBase in an Azure Synapse Analytics dedicated SQL pool to query the content of the data records that host the Twitter feeds. Data must be protected by using row-level security (RLS). The users must be authenticated by using their own Azure AD credentials.
Box 1: CREATE EXTERNAL DATA SOURCE
External data sources are used to connect to storage accounts. Box 2: CREATE EXTERNAL FILE FORMAT
CREATE EXTERNAL FILE FORMAT creates an external file format object that defines external data stored in Azure Blob Storage or Azure Data Lake Storage. Creating an external file format is a prerequisite for creating an external table.
Box 3: CREATE EXTERNAL TABLE AS SELECT
When used in conjunction with the CREATE TABLE AS SELECT statement, selecting from an external table imports data into a table within the SQL pool. In addition to the COPY statement, external tables are useful for loading data.
Reference:
https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/develop-tables-external-tables

NEW QUESTION 56
- (Exam Topic 1)
You need to design a data storage structure for the product sales transactions. The solution must meet the sales transaction dataset requirements.
What should you include in the solution? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.



A. Mastered
B. Not Mastered

Answer: A

Explanation:
Graphical user interface, text, application, chat or text message Description automatically generated
Box 1: Hash Scenario:
Ensure that queries joining and filtering sales transaction records based on product ID complete as quickly as possible.
A hash distributed table can deliver the highest query performance for joins and aggregations on large tables. Box 2: Set the distribution column to the sales date.
Scenario: Partition data that contains sales transaction records. Partitions must be designed to provide efficient loads by month. Boundary values must belong to the partition on the right.
Reference:
https://rajanieshkaushikk.com/2020/09/09/how-to-choose-right-data-distribution-strategy-for-azure-synapse/

NEW QUESTION 57
- (Exam Topic 3)
You have an enterprise data warehouse in Azure Synapse Analytics.
You need to monitor the data warehouse to identify whether you must scale up to a higher service level to accommodate the current workloads
Which is the best metric to monitor?
More than one answer choice may achieve the goal. Select the BEST answer.

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

Answer: C

NEW QUESTION 58
- (Exam Topic 3)
You have an Azure Synapse Analytics dedicated SQL pool named SA1 that contains a table named Table1. You need to identify tables that have a high percentage of deleted rows. What should you run?
A)
```
sys.pdw_nodes_column_store_segments
```
B)
```
sys.dm_db_column_store_row_group_operational_stats
```
C)
```
sys.pdw_nodes_column_store_row_groups
```
D)
```
sys.dm_db_column_store_row_group_physical_stats
```

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

Answer: B

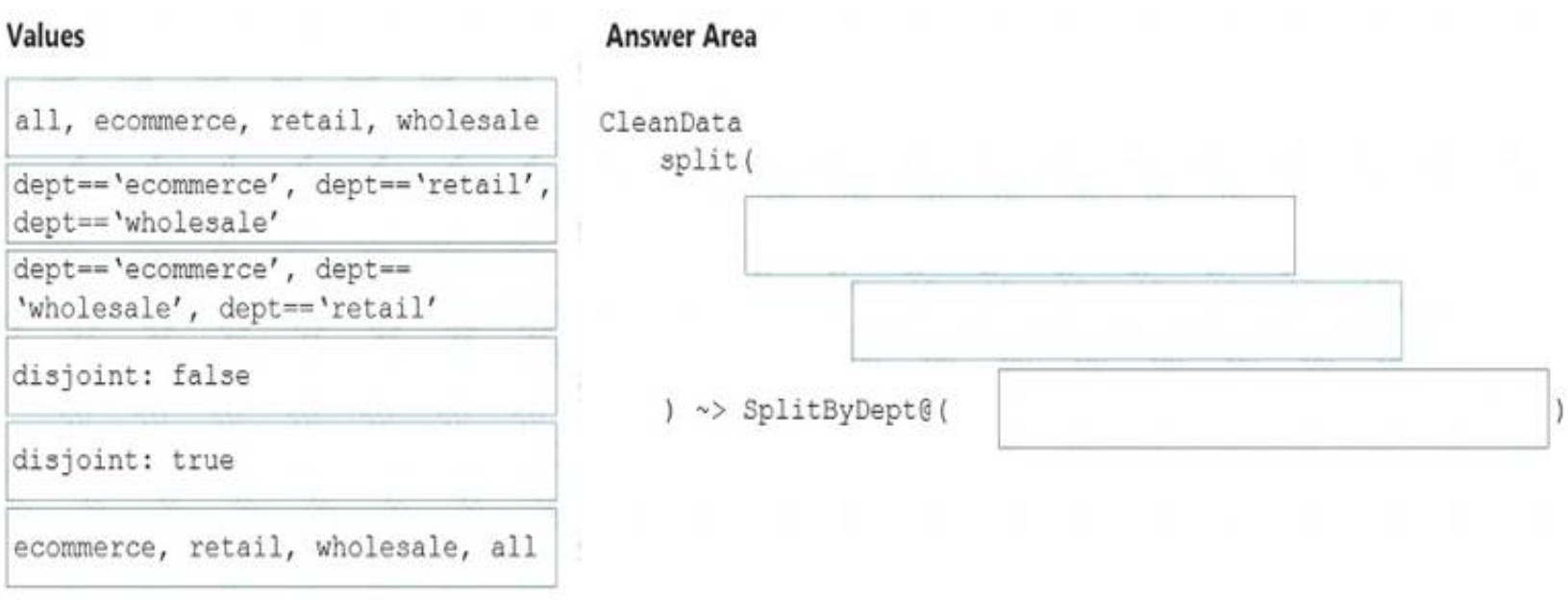
NEW QUESTION 62
- (Exam Topic 3)
You are designing a solution that will copy Parquet files stored in an Azure Blob storage account to an Azure Data Lake Storage Gen2 account.
The data will be loaded daily to the data lake and will use a folder structure of {Year}/{Month}/{Day}/. You need to design a 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 transfert 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 64
- (Exam Topic 3)
You need to create an Azure Data Factory pipeline to process data for the following three departments at your company: Ecommerce, retail, and wholesale. The solution must ensure that data can also be processed for the entire company.
How should you complete the Data Factory data flow script? To answer, drag the appropriate values to the correct targets. Each value may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.
NOTE: Each correct selection is worth one point.



A. Mastered
B. Not Mastered

Answer: A

Explanation:
The conditional split transformation routes data rows to different streams based on matching conditions. The conditional split transformation is similar to a CASE decision structure in a programming language. The transformation evaluates expressions, and based on the results, directs the data row to the specified stream.
Box 1: dept=='ecommerce', dept=='retail', dept=='wholesale'
First we put the condition. The order must match the stream labeling we define in Box 3. Syntax:
<incomingStream> split(
<conditionalExpression1>
<conditionalExpression2> disjoint: {true | false}
) ~> <splitTx>@(stream1, stream2, ..., <defaultStream>)
Box 2: discount : false
disjoint is false because the data goes to the first matching condition. All remaining rows matching the third condition go to output stream all.
Box 3: ecommerce, retail, wholesale, all Label the streams
Reference:
https://docs.microsoft.com/en-us/azure/data-factory/data-flow-conditional-split

NEW QUESTION 66
- (Exam Topic 3)
You have an Azure Synapse Analytics dedicated SQL pool that contains a table named Contacts. Contacts contains a column named Phone.
You need to ensure that users in a specific role only see the last four digits of a phone number when querying the Phone column.
What should you include in the solution?

A. a default value
B. dynamic data masking
C. row-level security (RLS)
D. column encryption
E. table partitions

Answer: B

Explanation:
Dynamic data masking helps prevent unauthorized access to sensitive data by enabling customers to designate how much of the sensitive data to reveal with minimal impact on the application layer. It's a policy-based security feature that hides the sensitive data in the result set of a query over designated database fields, while the data in the database is not changed.
Reference:
https://docs.microsoft.com/en-us/azure/azure-sql/database/dynamic-data-masking-overview

NEW QUESTION 71
- (Exam Topic 3)
You are designing a highly available Azure Data Lake Storage solution that will include geo-zone-redundant storage (GZRS).
You need to monitor for replication delays that can affect the recovery point objective (RPO). What should you include in the monitoring solution?

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

Answer: D

Explanation:
Because geo-replication is asynchronous, it is possible that data written to the primary region has not yet been written to the secondary region at the time an outage occurs. The Last Sync Time property indicates the last time that data from the primary region was written successfully to the secondary region. All writes made to the primary region before the last sync time are available to be read from the secondary location. Writes made to the primary region after the last sync time property may or may not be available for reads yet.
Reference:
https://docs.microsoft.com/en-us/azure/storage/common/last-sync-time-get

NEW QUESTION 72
- (Exam Topic 3)
You are deploying a lake database by using an Azure Synapse database template.
You need to add additional tables to the database. The solution must use the same grouping method as the template tables.
'Which grouping method should you use?

A. business area
B. size
C. facts and dimensions
D. partition style

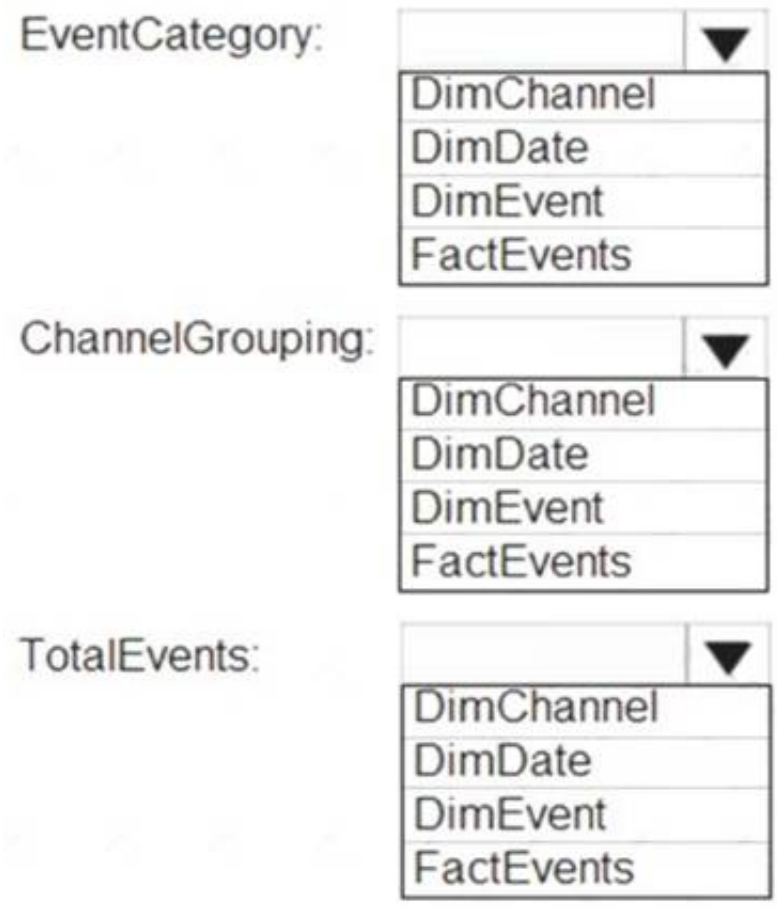
Answer: A

Explanation:
Business area: This is how the Azure Synapse database templates group tables by default. Each template consists of one or more enterprise templates that contain tables grouped by business areas. For example, the Retail template has business areas such as Customer, Product, Sales, and Store123. Using the same grouping method as the template tables can help you maintain consistency and compatibility with the industry-specific data model.
https://techcommunity.microsoft.com/t5/azure-synapse-analytics-blog/database-templates-in-azure-synapse-anal

NEW QUESTION 77
- (Exam Topic 3)
From a website analytics system, you receive data extracts about user interactions such as downloads, link clicks, form submissions, and video plays.
The data contains the following columns.

Name	Sample value
Date	15 Jan 2021
EventCategory	Videos
EventAction	Play
EventLabel	Contoso Promotional
ChannelGrouping	Social
TotalEvents	150
UniqueEvents	120
SessionWithEvents	99

You need to design a star schema to support analytical queries of the data. The star schema will contain four tables including a date dimension.
To which table should you add each column? 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:
Table Description automatically generated
Box 1: DimEvent
Box 2: DimChannel
Box 3: FactEvents
Fact tables store observations or events, and can be sales orders, stock balances, exchange rates, temperatures, etc

Reference:
https://docs.microsoft.com/en-us/power-bi/guidance/star-schema

NEW QUESTION 78
- (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 83
- (Exam Topic 3)
You plan to perform batch processing in Azure Databricks once daily. Which type of Databricks cluster should you use?

A. High Concurrency
B. automated
C. interactive

Answer: C

Explanation:
Azure Databricks has two types of clusters: interactive and automated. You use interactive clusters to analyze data collaboratively with interactive notebooks. You use automated clusters to run fast and robust automated jobs.
Example: Scheduled batch workloads (data engineers running ETL jobs)
This scenario involves running batch job JARs and notebooks on a regular cadence through the Databricks platform.
The suggested best practice is to launch a new cluster for each run of critical jobs. This helps avoid any issues (failures, missing SLA, and so on) due to an existing workload (noisy neighbor) on a shared cluster.
Reference:
https://docs.databricks.com/administration-guide/cloud-configurations/aws/cmbp.html#scenario-3-scheduled-bat

NEW QUESTION 86
- (Exam Topic 3)
You are designing a statistical analysis solution that will use custom proprietary1 Python functions on near real-time data from Azure Event Hubs.
You need to recommend which Azure service to use to perform the statistical analysis. The solution must minimize latency.
What should you recommend?

A. Azure Stream Analytics
B. Azure SQL Database
C. Azure Databricks
D. Azure Synapse Analytics

Answer: A

Explanation:
Reference:
https://docs.microsoft.com/en-us/azure/event-hubs/process-data-azure-stream-analytics

NEW QUESTION 90
- (Exam Topic 3)
You have an Azure Databricks workspace and an Azure Data Lake Storage Gen2 account named storage! New files are uploaded daily to storage1.
• Incrementally process new files as they are upkorage1 as a structured streaming source. The solution must meet the following requirements:
• Minimize implementation and maintenance effort.
• Minimize the cost of processing millions of files.
• Support schema inference and schema drift. Which should you include in the recommendation?

A. Auto Loader
B. Apache Spark FileStreamSource

C. COPY INTO
D. Azure Data Factory

Answer: D

NEW QUESTION 92
- (Exam Topic 3)
You are batch loading a table in an Azure Synapse Analytics dedicated SQL pool.
You need to load data from a staging table to the target table. The solution must ensure that if an error occurs while loading the data to the target table, all the inserts in that batch are undone.
How should you complete the Transact-SQL code? To answer, drag the appropriate values to the correct
targets. Each value may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.
NOTE Each correct selection is worth one point.

Values

- BEGIN DISTRIBUTED TRANSACTION
- BEGIN TRAN
- COMMIT TRAN
- ROLLBACK TRAN
- SET RESULT_SET_CACHING ON

Answer Area

```
[          ]
BEGIN TRY
 INSERT INTO dbo.Table1 (col1, col2, col3)
 SELECT col1, col2, col3 FROM stage.Table1;
END TRY
 BEGIN CATCH
   IF @@TRANCOUNT > 0
   BEGIN
     [          ];
   END
    END CATCH;
IF @@TRANCOUNT >0
BEGIN
  COMMIT TRAN;
  END
```

A. Mastered
B. Not Mastered

Answer: A

Explanation:

Values

- BEGIN DISTRIBUTED TRANSACTION
- BEGIN TRAN
- COMMIT TRAN
- ROLLBACK TRAN
- SET RESULT_SET_CACHING ON

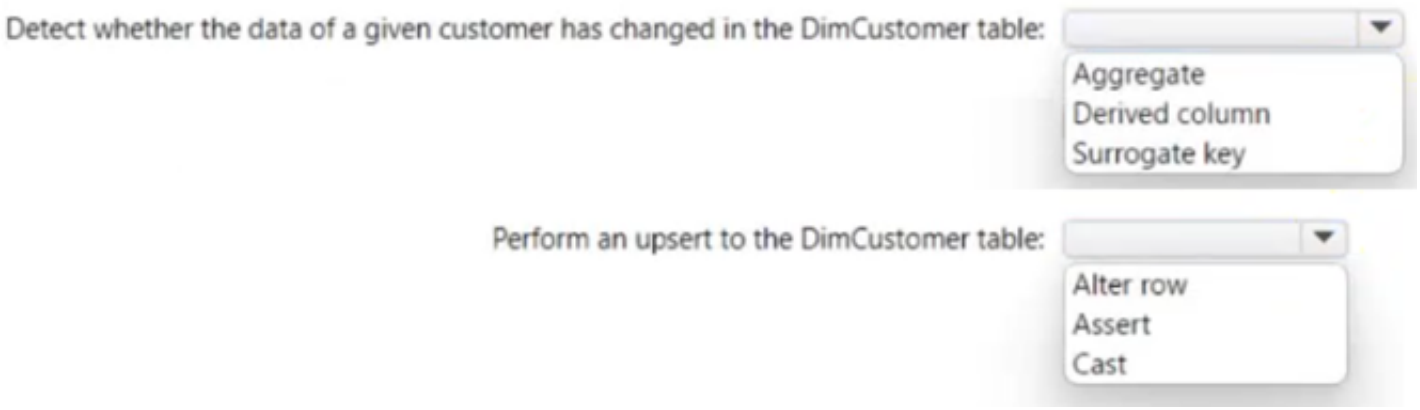
Answer Area

```
BEGIN TRAN
BEGIN TRY
 INSERT INTO dbo.Table1 (col1, col2, col3)
 SELECT col1, col2, col3 FROM stage.Table1;
END TRY
 BEGIN CATCH
   IF @@TRANCOUNT > 0
   BEGIN
     ROLLBACK TRAN;
   END
    END CATCH;
IF @@TRANCOUNT >0
BEGIN
  COMMIT TRAN;
  END
```

NEW QUESTION 97
- (Exam Topic 3)
You are developing an Azure Synapse Analytics pipeline that will include a mapping data flow named Dataflow1. Dataflow1 will read customer data from an external source and use a Type 1 slowly changing dimension (SCO) when loading the data into a table named DimCustomer1 in an Azure Synapse Analytics dedicated SQL pool.
You need to ensure that Dataflow1 can perform the following tasks:
* Detect whether the data of a given customer has changed in the DimCustomer table.
• Perform an upsert to the DimCustomer table.

Which type of transformation should you use for each task? To answer, select the appropriate options in the answer area
NOTE; Each correct selection is worth one point.

Answer Area

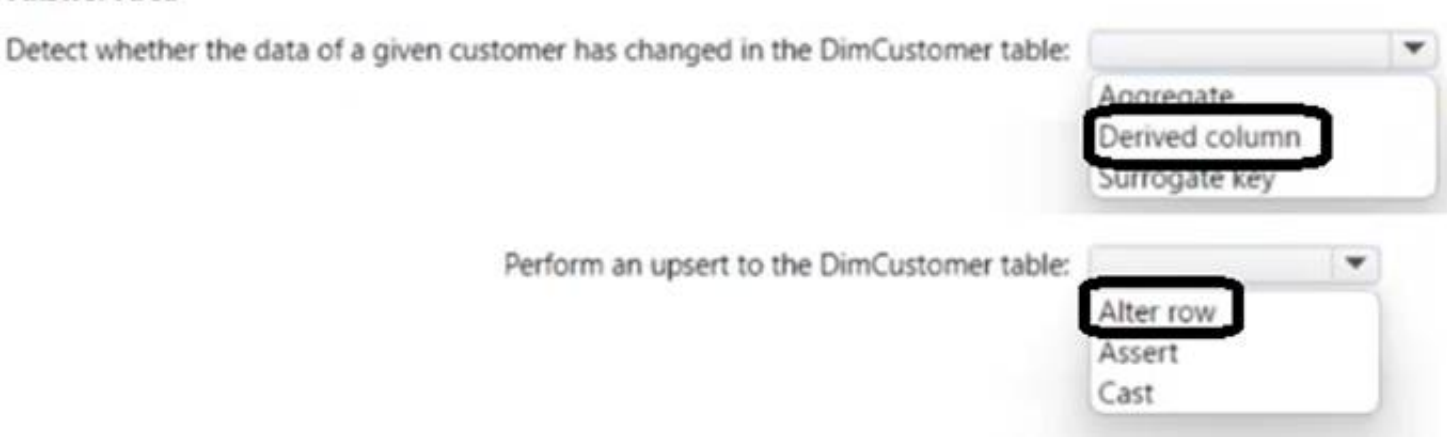


A. Mastered
B. Not Mastered

Answer: A

Explanation:

Answer Area



NEW QUESTION 102
- (Exam Topic 3)
You have an Azure Synapse Analytics dedicated SQL pool named Pool1 and a database named DB1. DB1 contains a fact table named Table1.
You need to identify the extent of the data skew in Table1. What should you do in Synapse Studio?

A. Connect to the built-in pool and query sysdm_pdw_sys_info.
B. Connect to Pool1 and run DBCC CHECKALLOC.
C. Connect to the built-in pool and run DBCC CHECKALLOC.
D. Connect to Pool! and query sys.dm_pdw_nodes_db_partition_stats.

Answer: D

Explanation:
Microsoft recommends use of sys.dm_pdw_nodes_db_partition_stats to analyze any skewness in the data. Reference:
https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/cheat-sheet

NEW QUESTION 103
- (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

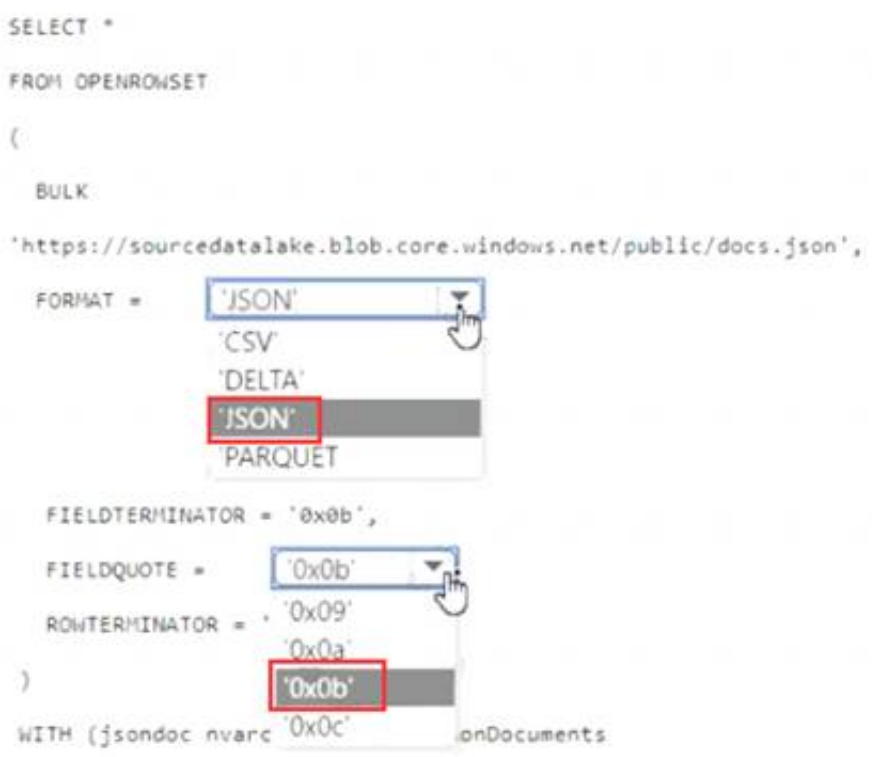
Explanation:
Parquet, an open-source file format for Hadoop, stores nested data structures in a flat columnar format. Compared to a traditional approach where data is stored in a row-oriented approach, Parquet file format is
more efficient in terms of storage and performance.
It is especially good for 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 107
- (Exam Topic 3)
You have an Azure Synapse serverless SQL pool.
You need to read JSON documents from a file by using the OPENROWSET function.
How should you complete the query? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.
Answer Area

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Answer Area



NEW QUESTION 110
- (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 114
- (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.

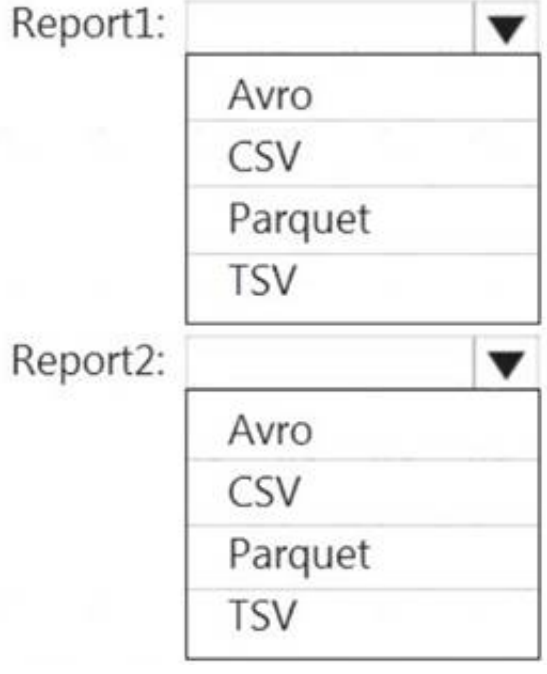
- Support for server encryption is currently provided through the SQL feature called Transparent Data Encryption.
- Client-side encryption of Azure SQL Database data is supported through the Always Encrypted feature. Reference:

https://docs.microsoft.com/en-us/azure/security/fundamentals/encryption-atrest

NEW QUESTION 117
- (Exam Topic 3)
You are planning the deployment of Azure Data Lake Storage Gen2. You have the following two reports that will access the data lake:

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

You need to recommend in which format to store the data in the data lake to support the reports. The solution must minimize read times.
What should you recommend for each report? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.



A. Mastered
B. Not Mastered

Answer: A

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

NEW QUESTION 118
- (Exam Topic 3)
You need to trigger an Azure Data Factory pipeline when a file arrives in an Azure Data Lake Storage Gen2 container.
Which resource provider should you enable?

A. Microsoft.Sql
B. Microsoft-Automation
C. Microsoft.EventGrid
D. Microsoft.EventHub

Answer: C

Explanation:
Event-driven architecture (EDA) is a common data integration pattern that involves production, detection, consumption, and reaction to events. Data integration scenarios often require Data Factory customers to trigger pipelines based on events happening in storage account, such as the arrival or deletion of a file in Azure Blob Storage account. Data Factory natively integrates with Azure Event Grid, which lets you trigger pipelines on such events.
Reference:
https://docs.microsoft.com/en-us/azure/data-factory/how-to-create-event-trigger https://docs.microsoft.com/en-us/azure/data-factory/concepts-pipeline-execution-triggers

NEW QUESTION 119
- (Exam Topic 3)
You have an Azure Databricks workspace named workspace1 in the Standard pricing tier.
You need to configure workspace1 to support autoscaling all-purpose clusters. The solution must meet the following requirements:

- Automatically scale down workers when the cluster is underutilized for three minutes.

- Minimize the time it takes to scale to the maximum number of workers.
- Minimize costs. What should you do first?

A. Enable container services for workspace1.
B. Upgrade workspace1 to the Premium pricing tier.
C. Set Cluster Mode to High Concurrency.
D. Create a cluster policy in workspace1.

Answer: B

Explanation:
For clusters running Databricks Runtime 6.4 and above, optimized autoscaling is used by all-purpose clusters in the Premium plan
Optimized autoscaling:
Scales up from min to max in 2 steps.
Can scale down even if the cluster is not idle by looking at shuffle file state. Scales down based on a percentage of current nodes.
On job clusters, scales down if the cluster is underutilized over the last 40 seconds.
On all-purpose clusters, scales down if the cluster is underutilized over the last 150 seconds.
The spark.databricks.aggressiveWindowDownS Spark configuration property specifies in seconds how often a cluster makes down-scaling decisions. Increasing the value causes a cluster to scale down more slowly. The maximum value is 600.
Note: Standard autoscaling
Starts with adding 8 nodes. Thereafter, scales up exponentially, but can take many steps to reach the max. You can customize the first step by setting the spark.databricks.autoscaling.standardFirstStepUp Spark configuration property.
Scales down only when the cluster is completely idle and it has been underutilized for the last 10 minutes. Scales down exponentially, starting with 1 node.
Reference: https://docs.databricks.com/clusters/configure.html

NEW QUESTION 124
- (Exam Topic 3)
You have an Azure Databricks workspace that contains a Delta Lake dimension table named Tablet. Table1 is a Type 2 slowly changing dimension (SCD) table. You need to apply updates from a source table to Table1. Which Apache Spark SQL operation should you use?

A. CREATE
B. UPDATE
C. MERGE
D. ALTER

Answer: C

Explanation:
The Delta provides the ability to infer the schema for data input which further reduces the effort required in managing the schema changes. The Slowly Changing Data(SCD) Type 2 records all the changes made to each key in the dimensional table. These operations require updating the existing rows to mark the previous values of the keys as old and then inserting new rows as the latest values. Also, Given a source table with the updates and the target table with dimensional data, SCD Type 2 can be expressed with the merge.
Example:
// Implementing SCD Type 2 operation using merge function customersTable
as("customers") merge(
stagedUpdates.as("staged_updates"), "customers.customerId = mergeKey")
whenMatched("customers.current = true AND customers.address <> staged_updates.address") updateExpr(Map(
"current" -> "false",
"endDate" -> "staged_updates.effectiveDate")) whenNotMatched()
insertExpr(Map(
"customerid" -> "staged_updates.customerId", "address" -> "staged_updates.address", "current" -> "true",
"effectiveDate" -> "staged_updates.effectiveDate",
"endDate" -> "null")) execute()
}
Reference:
https://www.projectpro.io/recipes/what-is-slowly-changing-data-scd-type-2-operation-delta-table-databricks

NEW QUESTION 128
- (Exam Topic 3)
You are designing an Azure Data Lake Storage Gen2 structure for telemetry data from 25 million devices distributed across seven key geographical regions. Each minute, the devices will send a JSON payload of metrics to Azure Event Hubs.
You need to recommend a folder structure for the data. The solution must meet the following requirements:

- Data engineers from each region must be able to build their own pipelines for the data of their respective region only.
- The data must be processed at least once every 15 minutes for inclusion in Azure Synapse Analytics serverless SQL pools.

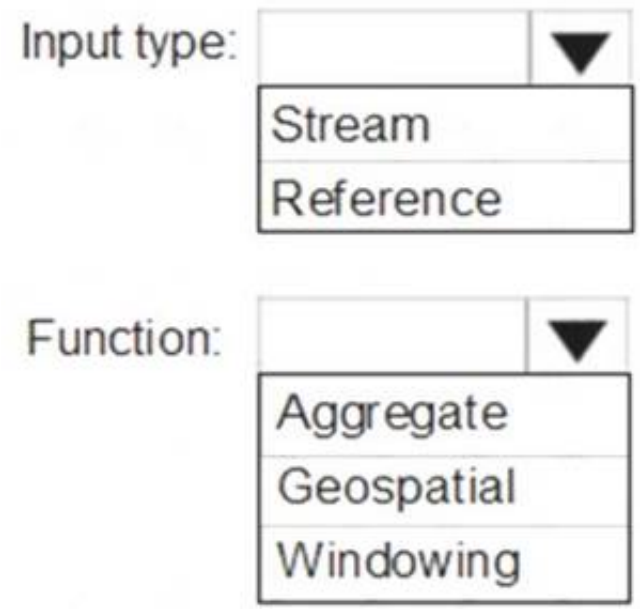
How should you recommend completing the structure? To answer, drag the appropriate values to the correct targets. Each value may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.
NOTE: Each correct selection is worth one point.

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: {YYYY}/{MM}/{DD}/{HH}
Date Format [optional]: if the date token is used in the prefix path, you can select the date format in which your files are organized. Example: YYYY/MM/DD
Time Format [optional]: if the time token is used in the prefix path, specify the time format in which your files are organized. Currently the only supported value is HH.
Box 2: {regionID}/raw
Data engineers from each region must be able to build their own pipelines for the data of their respective region only.
Box 3: {deviceID} Reference:
https://github.com/paolosalvatori/StreamAnalyticsAzureDataLakeStore/blob/master/README.md

NEW QUESTION 130
- (Exam Topic 3)
You plan to create a real-time monitoring app that alerts users when a device travels more than 200 meters away from a designated location.
You need to design an Azure Stream Analytics job to process the data for the planned app. The solution must minimize the amount of code developed and the number of technologies used.
What should you include in the Stream Analytics job? 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:
Diagram, table Description automatically generated
Input type: Stream
You can process real-time IoT data streams with Azure Stream Analytics. Function: Geospatial
With built-in geospatial functions, you can use Azure Stream Analytics to build applications for scenarios such as fleet management, ride sharing, connected cars, and asset tracking.
Note: In a real-world scenario, you could have hundreds of these sensors generating events as a stream. Ideally, a gateway device would run code to push these events to Azure Event Hubs or Azure IoT Hubs.
Reference:
https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-get-started-with-azure-stream-analytic https://docs.microsoft.com/en-us/azure/stream-analytics/geospatial-scenarios

NEW QUESTION 132
- (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 137
- (Exam Topic 3)
You plan to create a dimension table in Azure Synapse Analytics that will be less than 1 GB. You need to create the table to meet the following requirements:
• Provide the fastest Query time.
• Minimize data movement during queries. Which type of table should you use?

A. hash distributed
B. heap
C. replicated
D. round-robin

Answer: C

Explanation:
A replicated table has a full copy of the table accessible on each Compute node. Replicating a table removes the need to transfer data among Compute nodes before a join or aggregation. Since the table has multiple copies, replicated tables work best when the table size is less than 2 GB compressed. 2 GB is not a hard limit.
Reference:
https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/design-guidance-for-replicated-tab

NEW QUESTION 138
- (Exam Topic 3)
You have an Azure data factory.
You need to examine the pipeline failures from the last 180 flays. What should you use?

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

Answer: B

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

NEW QUESTION 143
- (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

CAST
COLLATE
CONVERT
FLATTEN
PIVOT
UNPIVOT

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
```

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 145
- (Exam Topic 3)
You are implementing Azure Stream Analytics windowing functions.
Which windowing function should you use for each requirement? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Answer Area

Segment the data stream into distinct time segments that repeat but do not overlap: Hopping / Sliding / Tumbling

Segment the data stream into distinct time segments that repeat and can overlap: Hopping / Sliding / Tumbling

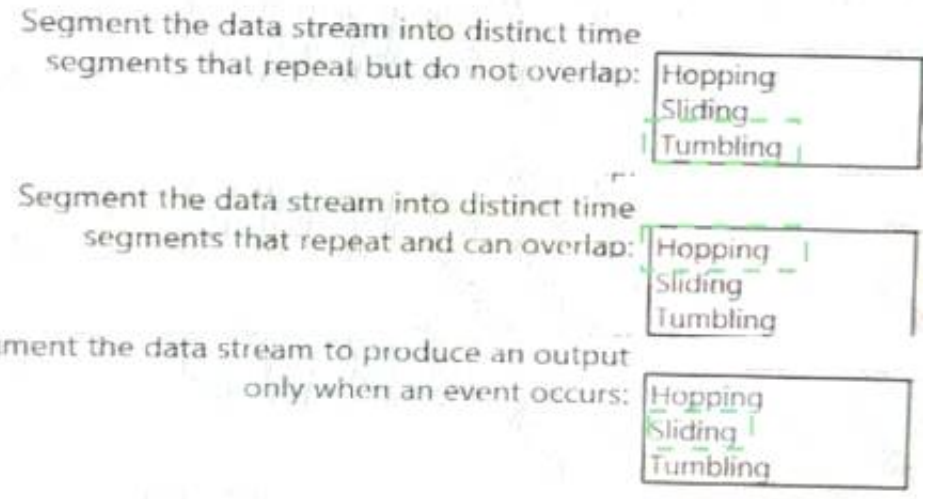
Segment the data stream to produce an output only when an event occurs: Hopping / Sliding / Tumbling

A. Mastered
B. Not Mastered

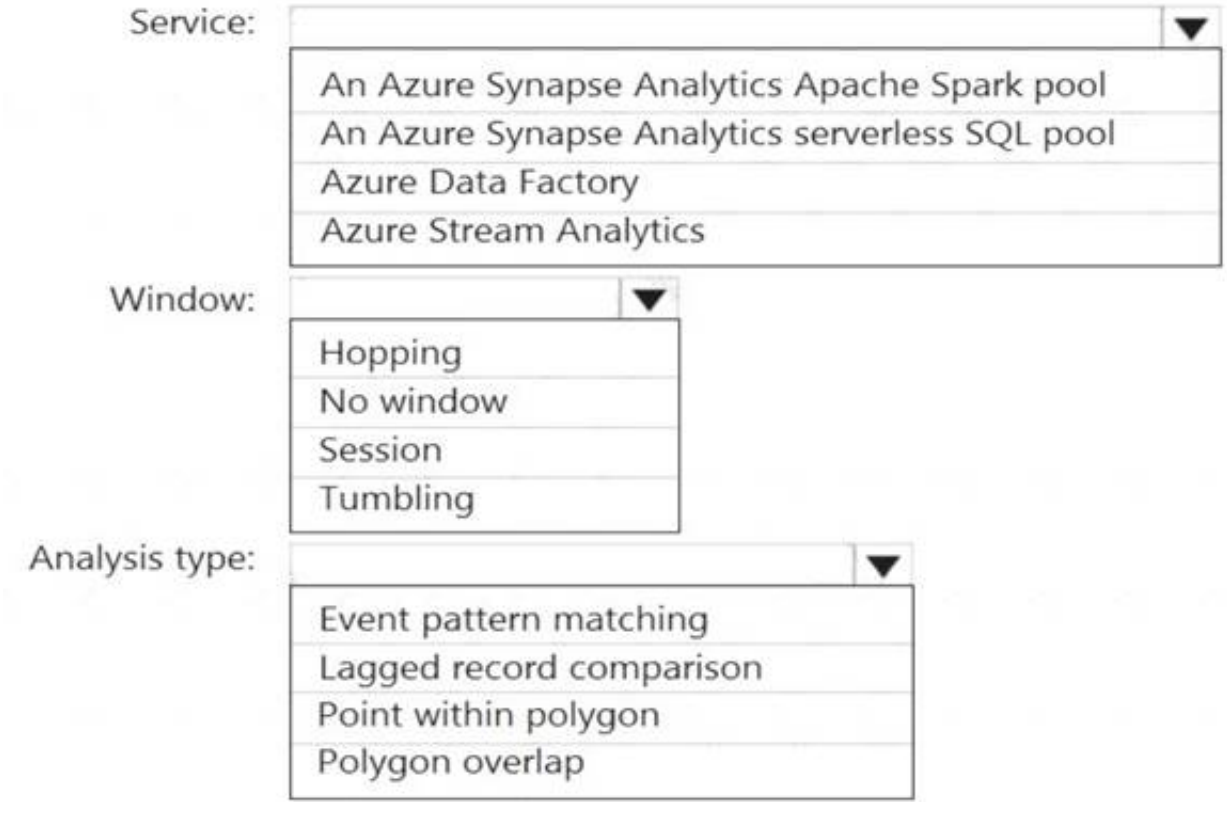
Answer: A

Explanation:

Answer Area



NEW QUESTION 150
- (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.



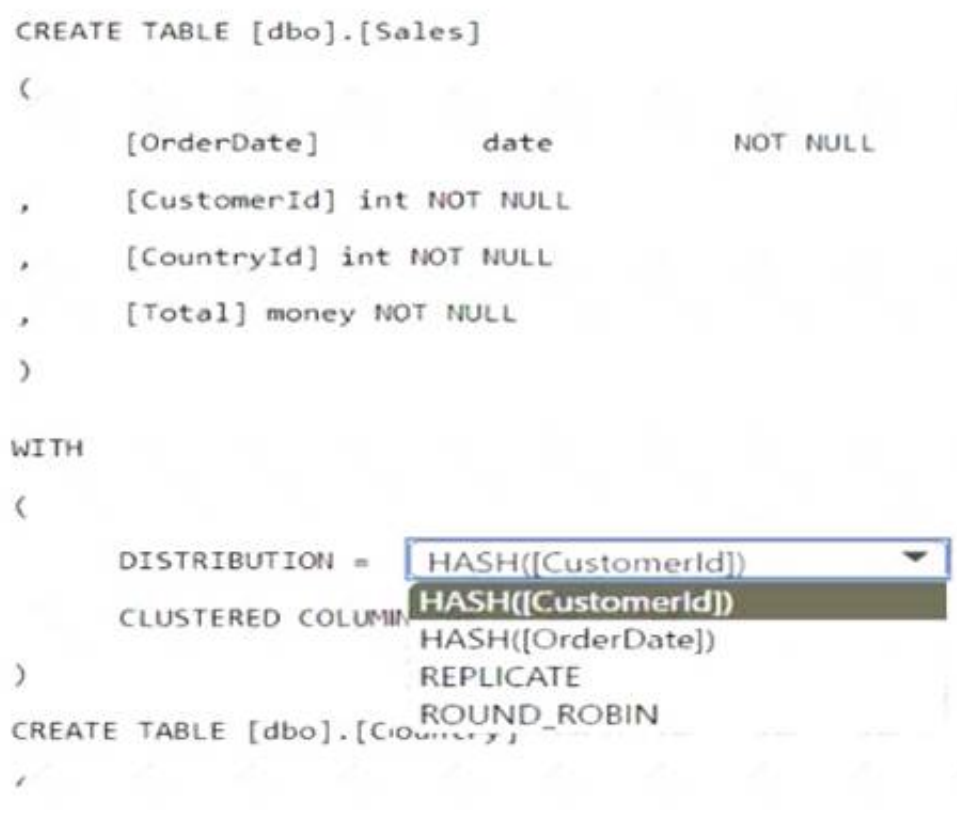
A. Mastered
B. Not Mastered

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 155
- (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



A. Mastered
B. Not Mastered

Answer: A

Explanation:
Answer Area

NEW QUESTION 157
- (Exam Topic 3)
You need to design an Azure Synapse Analytics dedicated SQL pool that meets the following requirements: Can return an employee record from a given point in time.
Maintains the latest employee information.
Minimizes query complexity.
How should you model the employee data?

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

Answer: D

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

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

NEW QUESTION 161
- (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:

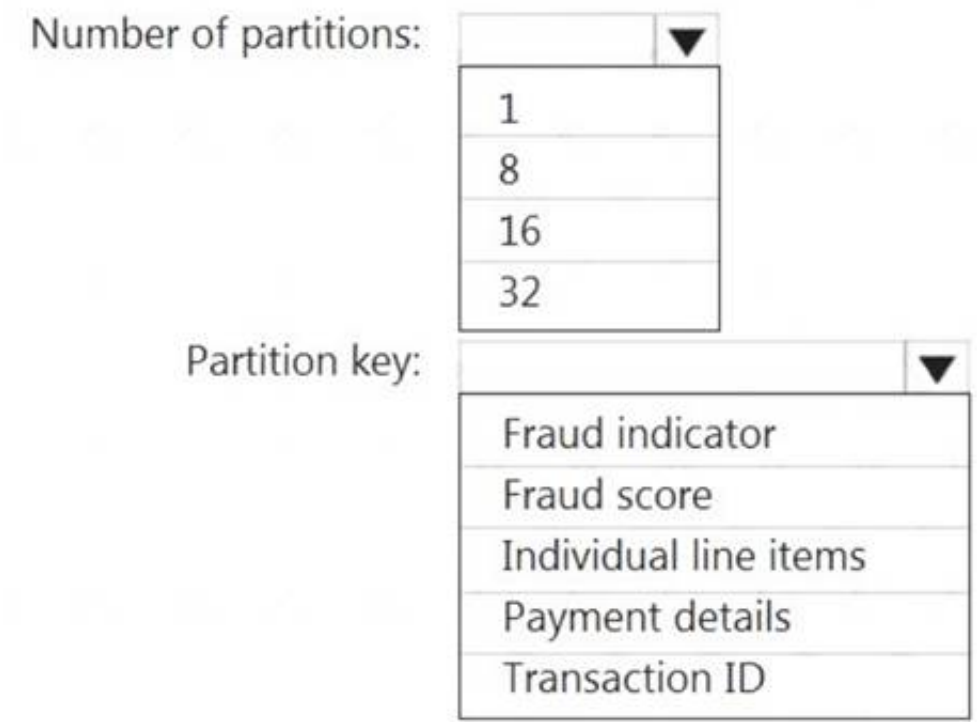
- Enable Pool1 to skip columns and rows that are unnecessary in a query.
- Automatically create column statistics.
- Minimize the size of files. Which type of file should you use?

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 165
- (Exam Topic 3)
You have an Azure event hub named retailhub that has 16 partitions. Transactions are posted to retailhub. Each transaction includes the transaction ID, the individual line items, and the payment details. The transaction ID is used as the partition key.
You are designing an Azure Stream Analytics job to identify potentially fraudulent transactions at a retail store. The job will use retailhub as the input. The job will output the transaction ID, the individual line items, the payment details, a fraud score, and a fraud indicator.
You plan to send the output to an Azure event hub named fraudhub.
You need to ensure that the fraud detection solution is highly scalable and processes transactions as quickly as possible.
How should you structure the output of the Stream Analytics job? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

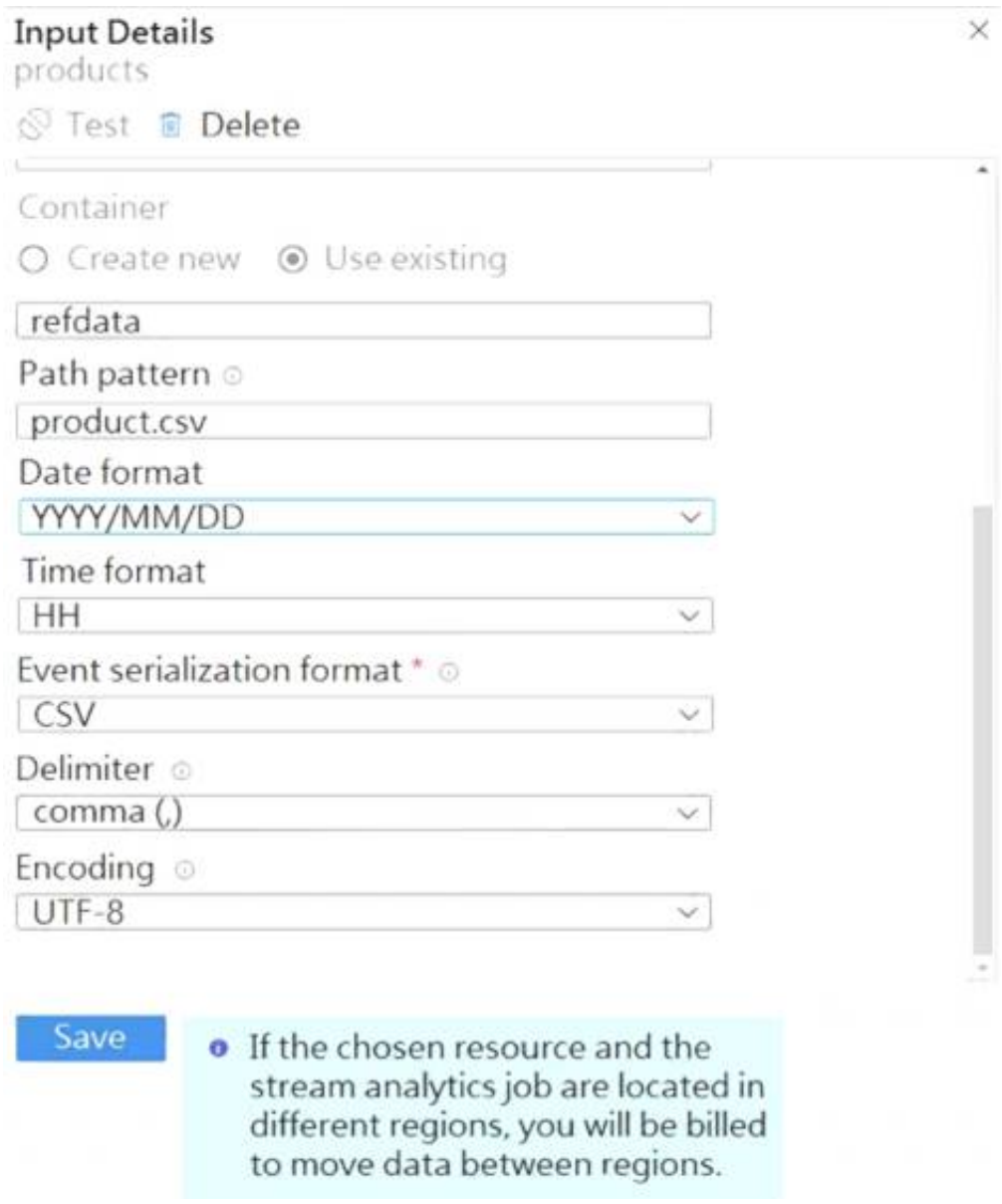


A. Mastered
B. Not Mastered

Answer: A

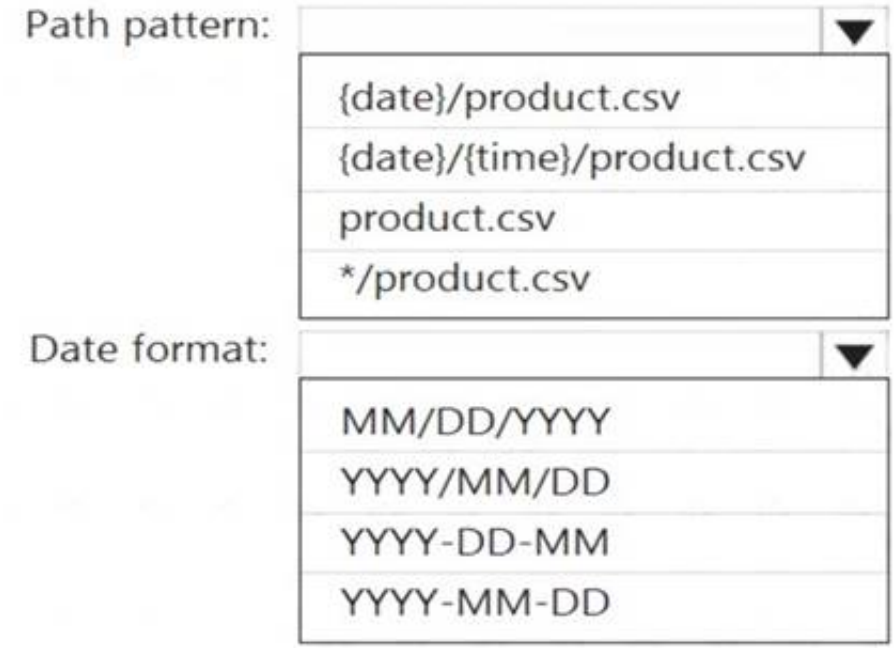
Explanation:
Box 1: 16
For Event Hubs you need to set the partition key explicitly.
An embarrassingly parallel job is the most scalable scenario in Azure Stream Analytics. It connects one partition of the input to one instance of the query to one partition of the output. Box 2: Transaction ID
Reference:
https://docs.microsoft.com/en-us/azure/event-hubs/event-hubs-features#partitions

NEW QUESTION 166
- (Exam Topic 3)
You are building an Azure Stream Analytics job that queries reference data from a product catalog file. The file is updated daily.
The reference data input details for the file are shown in the Input exhibit. (Click the Input tab.)



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

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



A. Mastered
B. Not Mastered

Answer: A

Explanation:
Graphical user interface, application, table Description automatically generated
Box 1: {date}/product.csv
In the 2nd exhibit we see: Location: refdata / 2020-03-20

Note: Path Pattern: This is a required property that is used to locate your blobs within the specified container. Within the path, you may choose to specify one or more instances of the following 2 variables:
{date}, {time}
Example 1: products/{date}/{time}/product-list.csv
Example 2: products/{date}/product-list.csv
Example 3: product-list.csv
Box 2: YYYY-MM-DD
Note: Date Format [optional]: If you have used {date} within the Path Pattern that you specified, then you can select the date format in which your blobs are organized from the drop-down of supported formats.
Example: YYYY/MM/DD, MM/DD/YYYY, etc. Reference:
https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-use-reference-data

NEW QUESTION 168
- (Exam Topic 3)
You have an Azure subscription that contains the resources shown in the following table.

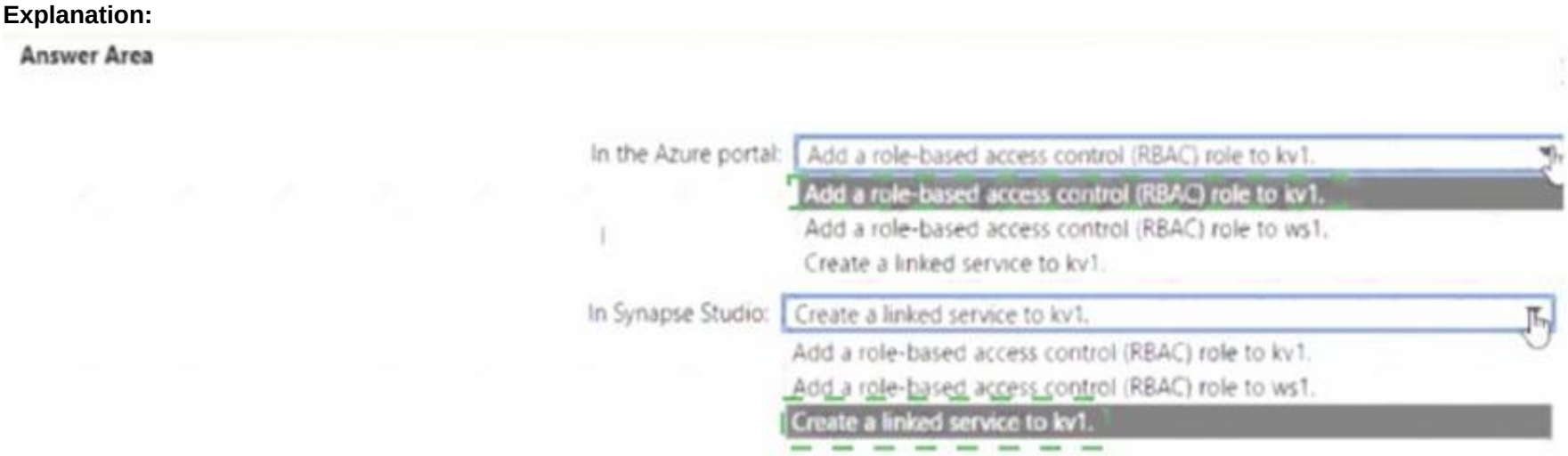
Name	Type	Description
ws1	Azure Synapse Analytics workspace	None
kv1	Azure Key Vault	None
UAMI1	User-assigned managed identity	Associated with ws1
sp1	Apache Spark pool in Azure Synapse Analytics	Associated with ws1

You need to ensure that you can Spark notebooks in ws1. The solution must ensure secrets from kv1 by using UAMI1. What should you do? 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:



NEW QUESTION 169
- (Exam Topic 3)
You have an Azure Stream Analytics job that receives clickstream data from an Azure event hub.
You need to define a query in the Stream Analytics job. The query must meet the following requirements: Count the number of clicks within each 10-second window based on the country of a visitor.
Ensure that each click is NOT counted more than once. How should you define the Query?

A. SELECT Country, Avg(*) AS AverageFROM ClickStream TIMESTAMP BY CreatedAt GROUP BY Country, SlidingWindow(second, 10)
B. SELECT Country, Count(*) AS CountFROM ClickStream TIMESTAMP BY CreatedAt GROUP BY Country, TumblingWindow(second, 10)
C. SELECT Country, Avg(*) AS AverageFROM ClickStream TIMESTAMP BY CreatedAt GROUP BY Country, HoppingWindow(second, 10, 2)
D. SELECT Country, Count(*) AS CountFROM ClickStream TIMESTAMP BY CreatedAt GROUP BY Country, SessionWindow(second, 5, 10)

Answer: B

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

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

NEW QUESTION 172
- (Exam Topic 3)
You are creating dimensions for a data warehouse in an Azure Synapse Analytics dedicated SQL pool. You create a table by using the Transact-SQL statement shown in the following exhibit.

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

Use the drop-down menus to select the answer choice that completes each statement based on the information presented in the graphic.
NOTE: Each correct selection is worth one point.

DimProduct is a **[answer choice]** slowly changing dimension (SCD).
- Type 0
- Type 1
- Type 2

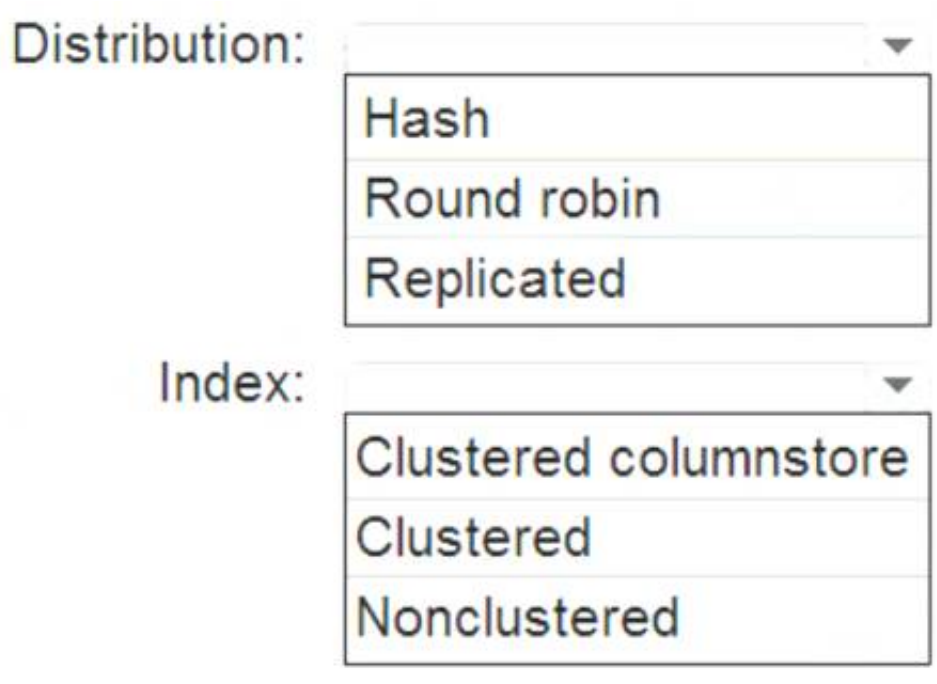
The ProductKey column is **[answer choice]**.
- a surrogate key
- a business key
- an audit column

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: Type 2
A Type 2 SCD supports versioning of dimension members. Often the source system doesn't store versions, so the data warehouse load process detects and manages changes in a dimension table. In this case, the dimension table must use a surrogate key to provide a unique reference to a version of the dimension member. It also includes columns that define the date range validity of the version (for example, StartDate and EndDate) and possibly a flag column (for example, IsCurrent) to easily filter by current dimension members.
Reference:
https://docs.microsoft.com/en-us/learn/modules/populate-slowly-changing-dimensions-azure-synapse-analytics

NEW QUESTION 173
- (Exam Topic 3)
You are designing an enterprise data warehouse in Azure Synapse Analytics that will store website traffic analytics in a star schema.
You plan to have a fact table for website visits. The table will be approximately 5 GB.
You need to recommend which distribution type and index type to use for the table. The solution must provide the fastest query performance.
What should you recommend? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.



A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: Hash
Consider using a hash-distributed table when: The table size on disk is more than 2 GB.
The table has frequent insert, update, and delete operations. Box 2: Clustered columnstore
Clustered columnstore tables offer both the highest level of data compression and the best overall query performance.
Reference:
https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-distribu https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-index

NEW QUESTION 177
- (Exam Topic 3)
You plan to implement an Azure Data Lake Storage Gen2 container that will contain CSV files. The size of the files will vary based on the number of events that occur per hour.
File sizes range from 4.KB to 5 GB.
You need to ensure that the files stored in the container are optimized for batch processing. What should you do?

A. Compress the files.
B. Merge the files.
C. Convert the files to JSON
D. Convert the files to Avro.

Answer: D

Explanation:
Avro supports batch and is very relevant for streaming.
Note: Avro is framework developed within Apache's Hadoop project. It is a row-based storage format which is widely used as a serialization process. AVRO stores its schema in JSON format making it easy to read and interpret by any program. The data itself is stored in binary format by doing it compact and efficient.
Reference:
https://www.adaltas.com/en/2020/07/23/benchmark-study-of-different-file-format/

NEW QUESTION 180
- (Exam Topic 3)
You have an Azure Synapse Analytics job that uses Scala. You need to view the status of the job.
What should you do?

A. From Azure Monitor, run a Kusto query against the AzureDiagnostics table.
B. From Azure Monitor, run a Kusto query against the SparkLogying1 Event.CL table.
C. From Synapse Studio, select the workspac
D. From Monitor, select Apache Sparks applications.
E. From Synapse Studio, select the workspac
F. From Monitor, select SQL requests.

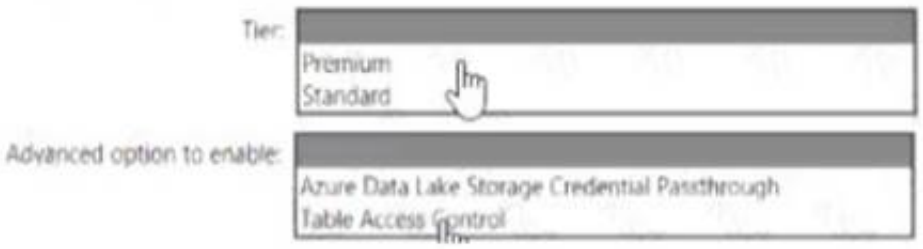
Answer: C

Explanation:
Use Synapse Studio to monitor your Apache Spark applications. To monitor running Apache Spark application Open Monitor, then select Apache Spark applications. To view the details about the Apache Spark applications that are running, select the submitting Apache Spark application and view the details. If the Apache Spark application is still running, you can monitor the progress.
Reference:
https://docs.microsoft.com/en-us/azure/synapse-analytics/monitoring/apache-spark-applications

NEW QUESTION 183
- (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



A. Mastered
B. Not Mastered

Answer: A

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

NEW QUESTION 184
- (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 186
- (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.

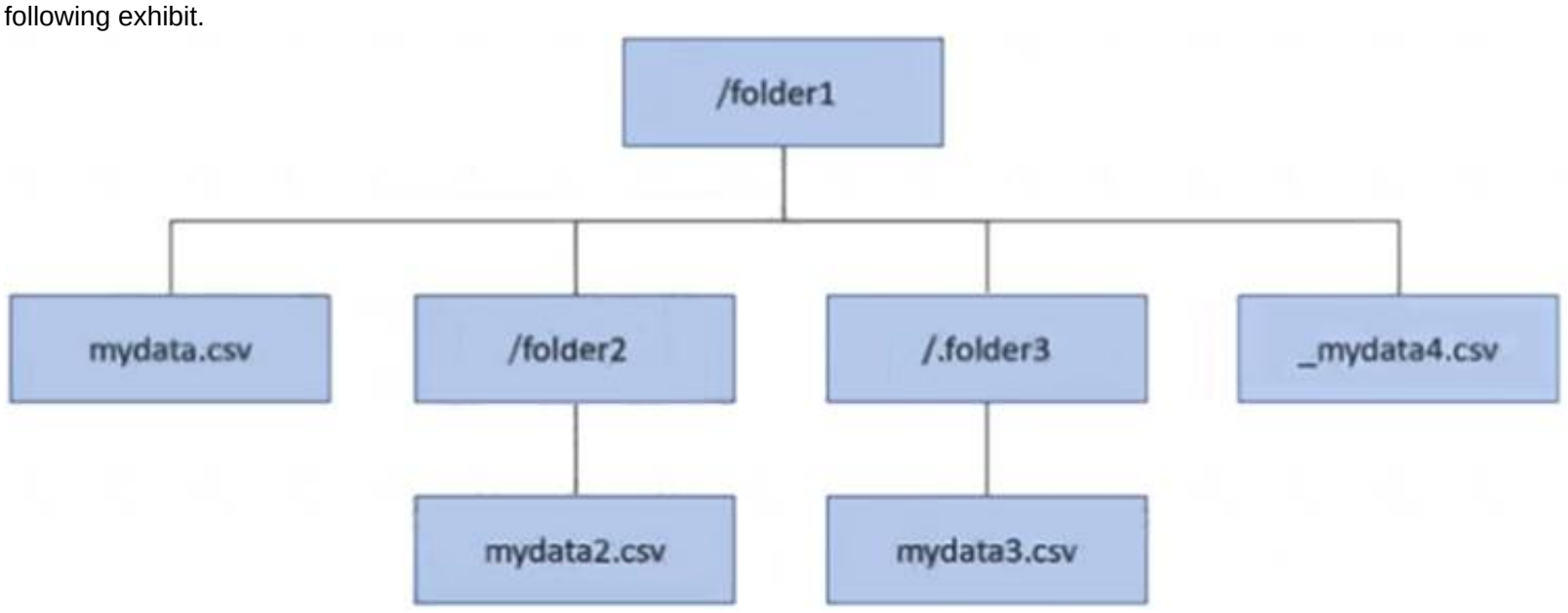
Solution: You schedule an Azure Databricks job that executes an R notebook, and then inserts the data into the data warehouse.
Does this meet the goal?

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 187
- (Exam Topic 3)
You have an Azure Data Lake Storage Gen2 account that contains a container named container1. You have an Azure Synapse Analytics serverless SQL pool that contains a native external table named dbo.Table1. The source data for dbo.Table1 is stored in container1. The folder structure of container1 is shown in the following exhibit.



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

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

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

Statements	Yes	No
When selecting all the rows in dbo.Table1, data from the mydata2.csv file will be returned.	O	O
When selecting all the rows in dbo.Table1, data from the mydata3.csv file will be returned.	O	O
When selecting all the rows in dbo.Table1, data from the _mydata4.csv file will be returned.	O	O

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: Yes
In the serverless SQL pool you can also use recursive wildcards /logs/** to reference Parquet or CSV files in any sub-folder beneath the referenced folder.
Box 2: Yes
Box 3: No
Reference: https://learn.microsoft.com/en-us/azure/synapse-analytics/sql/develop-tables-external-tables

NEW QUESTION 188
- (Exam Topic 3)
You have an Azure Data Lake Storage Gen2 account named adls2 that is protected by a virtual network. You are designing a SQL pool in Azure Synapse that will use adls2 as a source.
What should you use to authenticate to adls2?

A. a shared access signature (SAS)
B. a managed identity
C. a shared key
D. an Azure Active Directory (Azure AD) user

Answer: B

Explanation:
Managed identity for Azure resources is a feature of Azure Active Directory. The feature provides Azure services with an automatically managed identity in Azure AD. You can use the Managed Identity capability to authenticate to any service that support Azure AD authentication.
Managed Identity authentication is required when your storage account is attached to a VNet. Reference:
https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/quickstart-bulk-load-copy-tsql-exa

NEW QUESTION 191
- (Exam Topic 3)
You have an Azure Data Factory pipeline that performs an incremental load of source data to an Azure Data Lake Storage Gen2 account.
Data to be loaded is identified by a column named LastUpdatedDate in the source table. You plan to execute the pipeline every four hours.
You need to ensure that the pipeline execution meets the following requirements:

- Automatically retries the execution when the pipeline run fails due to concurrency or throttling limits.
- Supports backfilling existing data in the table.

Which type of trigger should you use?

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

Answer: D

Explanation:
In case of pipeline failures, tumbling window trigger can retry the execution of the referenced pipeline automatically, using the same input parameters, without the user intervention. This can be specified using the property "retryPolicy" in the trigger definition.
Reference:
https://docs.microsoft.com/en-us/azure/data-factory/how-to-create-tumbling-window-trigger

NEW QUESTION 195
- (Exam Topic 3)
You have an enterprise data warehouse in Azure Synapse Analytics named DW1 on a server named Server1. You need to verify whether the size of the transaction log file for each distribution of DW1 is smaller than 160 GB.
What should you do?

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

Answer: A

Explanation:
The following query returns the transaction log size on each distribution. If one of the log files is reaching 160 GB, you should consider scaling up your instance or limiting your transaction size.
-- Transaction log size SELECT
instance_name as distribution_db, cntr_value*1.0/1048576 as log_file_size_used_GB, pdw_node_id
FROM sys.dm_pdw_nodes_os_performance_counters WHERE
instance_name like 'Distribution_%'
AND counter_name = 'Log File(s) Used Size (KB)'
References:
https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-manage-monitor

NEW QUESTION 196
- (Exam Topic 3)
You have an Azure Synapse Analytics dedicated SQL pool that contains a large fact table. The table contains 50 columns and 5 billion rows and is a heap.
Most queries against the table aggregate values from approximately 100 million rows and return only two columns.
You discover that the queries against the fact table are very slow. Which type of index should you add to provide the fastest query times?

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

Answer: B

Explanation:
Clustered columnstore indexes are one of the most efficient ways you can store your data in dedicated SQL pool.
Columnstore tables won't benefit a query unless the table has more than 60 million rows. Reference:
https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/best-practices-dedicated-sql-pool

NEW QUESTION 201
- (Exam Topic 3)
You have an Azure Synapse Analytics dedicated SQL pool.
You need to Create a fact table named Table1 that will store sales data from the last three years. The solution must be optimized for the following query operations:
Show order counts by week.

• Calculate sales totals by region.
• Calculate sales totals by product.
• Find all the orders from a given month. Which data should you use to partition Table1?

A. region
B. product
C. week
D. month

Answer: D

Explanation:
Table partitions enable you to divide your data into smaller groups of data. In most cases, table partitions are created on a date column.
Benefits to queries
Partitioning can also be used to improve query performance. A query that applies a filter to partitioned data can limit the scan to only the qualifying partitions. This method of filtering can avoid a full table scan and only scan a smaller subset of data. With the introduction of clustered columnstore indexes, the predicate elimination performance benefits are less beneficial, but in some cases there can be a benefit to queries.
For example, if the sales fact table is partitioned into 36 months using the sales date field, then queries that filter on the sale date can skip searching in partitions that don't match the filter.
Note: Benefits to loads
The primary benefit of partitioning in dedicated SQL pool is to improve the efficiency and performance of loading data by use of partition deletion, switching and merging. In most cases data is partitioned on a date column that is closely tied to the order in which the data is loaded into the SQL pool. One of the greatest benefits of using partitions to maintain data is the avoidance of transaction logging. While simply inserting, updating, or deleting data can be the most straightforward approach, with a little thought and effort, using partitioning during your load process can substantially improve performance.
Reference:
https://learn.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-partitio

NEW QUESTION 202
.....

Thank You for Trying Our Product

We offer two products:

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

2nd - Questons and Answers in PDF Format

DP-203 Practice Exam Features:

* DP-203 Questions and Answers Updated Frequently

* DP-203 Practice Questions Verified by Expert Senior Certified Staff

* DP-203 Most Realistic Questions that Guarantee you a Pass on Your FirstTry

* DP-203 Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year