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.

Solution:
Answer as below

Answer Area

Path pattern: {date}/product.csv ▼

Date format: YYYY/MM/DD ▼

Does this meet the goal?

A. Yes
B. No

Answer: A

NEW QUESTION 2
- (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.	○	○

Solution:
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.	○	[○]

Does this meet the goal?

A. Yes
B. No

Answer: A

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

A. Yes
B. No

Answer: A

Explanation:
If you need to transform data in a way that is not supported by Data Factory, you can create a custom activity with your own data processing logic and use the activity in the pipeline.
Note: You can use data transformation activities in Azure Data Factory and Synapse pipelines to transform and process your raw data into predictions and insights at scale.
Reference:
https://docs.microsoft.com/en-us/azure/data-factory/transform-data

NEW QUESTION 4
- (Exam Topic 3)
You are designing an Azure Databricks cluster that runs user-defined local processes. You need to recommend a cluster configuration that meets the following requirements:
• Minimize query latency.
• Maximize the number of users that can run queues on the cluster at the same time « Reduce overall costs without compromising other requirements
Which cluster type should you recommend?

A. Standard with Auto termination
B. Standard with Autoscaling
C. High Concurrency with Autoscaling
D. High Concurrency with Auto Termination

Answer: C

Explanation:
A High Concurrency cluster is a managed cloud resource. The key benefits of High Concurrency clusters are that they provide fine-grained sharing for maximum resource utilization and minimum query latencies.
Databricks chooses the appropriate number of workers required to run your job. This is referred to as autoscaling. Autoscaling makes it easier to achieve high cluster utilization, because you don't need to provision the cluster to match a workload.
Reference:

https://docs.microsoft.com/en-us/azure/databricks/clusters/configure

NEW QUESTION 5
- (Exam Topic 3)
You have an Azure Synapse Analytics dedicated SQL pool named pool1.
You plan to implement a star schema in pool1 and create a new table named DimCustomer by using the following code.

```
CREATE TABLE dbo.[DimCustomer](
    [CustomerKey] int NOT NULL,
    [CustomerSourceID] [int] NOT NULL,
    [Title] [nvarchar](8) NULL,
    [FirstName] [nvarchar](50) NOT NULL,
    [MiddleName] [nvarchar](50) NULL,
    [LastName] [nvarchar](50) NOT NULL,
    [Suffix] [nvarchar](10) NULL,
    [CompanyName] [nvarchar](128) NULL,
    [SalesPerson] [nvarchar](256) NULL,
    [EmailAddress] [nvarchar](50) NULL,
    [Phone] [nvarchar](25) NULL,
    [InsertedDate] [datetime] NOT NULL,
    [ModifiedDate] [datetime] NOT NULL,
    [HashKey] [varchar](100) NOT NULL,
    [IsCurrentRow] [bit] NOT NULL
)
WITH
(
    DISTRIBUTION = REPLICATE,
    CLUSTERED COLUMNSTORE INDEX
);
GO
```

You need to ensure that DimCustomer has the necessary columns to support a Type 2 slowly changing dimension (SCD). Which two columns should you add?
Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

A. [HistoricalSalesPerson] [nvarchar] (256) NOT NULL
B. [EffectiveEndDate] [datetime] NOT NULL
C. [PreviousModifiedDate] [datetime] NOT NULL
D. [RowID] [bigint] NOT NULL
E. [EffectiveStartDate] [datetime] NOT NULL

Answer: AB

NEW QUESTION 6
- (Exam Topic 3)
You have an Azure Stream Analytics query. The query returns a result set that contains 10,000 distinct values for a column named clusterID.
You monitor the Stream Analytics job and discover high latency. You need to reduce the latency.
Which two actions should you perform? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

A. Add a pass-through query.
B. Add a temporal analytic function.
C. Scale out the query by using PARTITION BY.
D. Convert the query to a reference query.
E. Increase the number of streaming units.

Answer: CE

Explanation:
C: Scaling a Stream Analytics job takes advantage of partitions in the input or output. Partitioning lets you divide data into subsets based on a partition key. A process that consumes the data (such as a Streaming Analytics job) can consume and write different partitions in parallel, which increases throughput.
E: Streaming Units (SUs) represents the computing resources that are allocated to execute a Stream Analytics job. The higher the number of SUs, the more CPU and memory resources are allocated for your job. This capacity lets you focus on the query logic and abstracts the need to manage the hardware to run your Stream Analytics job in a timely manner.
References:
https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-parallelization https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-streaming-unit-consumption

NEW QUESTION 7
- (Exam Topic 3)
You have the following Azure Stream Analytics query.

```
WITH

step1 AS (SELECT *
     FROM input1
     PARTITION BY StateID
     INTO 10),
step2 AS (SELECT *
     FROM input2
     PARTITION BY StateID
     INTO 10)

SELECT *
INTO output
FROM step1
PARTITION BY StateID
UNION
SELECT * INTO output
     FROM step2
     PARTITION BY StateID
```

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

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

Solution:
Box 1: No
Note: You can now use a new extension of Azure Stream Analytics SQL to specify the number of partitions of a stream when reshuffling the data.
The outcome is a stream that has the same partition scheme. Please see below for an example: WITH step1 AS (SELECT * FROM [input1] PARTITION BY DeviceID INTO 10),
step2 AS (SELECT * FROM [input2] PARTITION BY DeviceID INTO 10)
SELECT * INTO [output] FROM step1 PARTITION BY DeviceID UNION step2 PARTITION BY DeviceID Note: The new extension of Azure Stream Analytics SQL includes a keyword INTO that allows you to specify the number of partitions for a stream when performing reshuffling using a PARTITION BY statement.
Box 2: Yes
When joining two streams of data explicitly repartitioned, these streams must have the same partition key and partition count. Box 3: Yes
Streaming Units (SUs) represents the computing resources that are allocated to execute a Stream Analytics job. The higher the number of SUs, the more CPU and memory resources are allocated for your job.
In general, the best practice is to start with 6 SUs for queries that don't use PARTITION BY. Here there are 10 partitions, so 6x10 = 60 SUs is good.
Note: Remember, Streaming Unit (SU) count, which is the unit of scale for Azure Stream Analytics, must be adjusted so the number of physical resources available to the job can fit the partitioned flow. In general, six SUs is a good number to assign to each partition. In case there are insufficient resources assigned to the job, the system will only apply the repartition if it benefits the job.
Reference:
https://azure.microsoft.com/en-in/blog/maximize-throughput-with-repartitioning-in-azure-stream-analytics/ https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-streaming-unit-consumption

Does this meet the goal?

A. Yes
B. No

Answer: A

NEW QUESTION 8
- (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 9
- (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 10
- (Exam Topic 3)
You have the following table named Employees.

first_name	last_name	hire_date	employee_type
Jane	Doe	2019-08-23	new
Ben	Smith	2017-12-15	Standard

You need to calculate the employee_type value based on the hire_date value.
How should you complete the Transact-SQL statement? To answer, drag the appropriate values to the correct targets. Each value may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.
NOTE: Each correct selection is worth one point.

Solution:
Graphical user interface, text, application Description automatically generated
Box 1: CASE
CASE evaluates a list of conditions and returns one of multiple possible result expressions.
CASE can be used in any statement or clause that allows a valid expression. For example, you can use CASE in statements such as SELECT, UPDATE, DELETE and SET, and in clauses such as select_list, IN, WHERE, ORDER BY, and HAVING.
Syntax: Simple CASE expression: CASE input_expression
WHEN when_expression THEN result_expression [...n] [ELSE else_result_expression]
END
Box 2: ELSE
Reference:
https://docs.microsoft.com/en-us/sql/t-sql/language-elements/case-transact-sql

Does this meet the goal?

A. Yes
B. No

Answer: A

NEW QUESTION 11
- (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 are designing an Azure Stream Analytics solution that will analyze Twitter data.
You need to count the tweets in each 10-second window. The solution must ensure that each tweet is counted only once.

Solution: You use a hopping window that uses a hop size of 10 seconds and a window size of 10 seconds. Does this meet the goal?

A. Yes
B. No

Answer: B

Explanation:
Instead use a tumbling window. Tumbling windows are a series of fixed-sized, non-overlapping and contiguous time intervals.
Reference:
https://docs.microsoft.com/en-us/stream-analytics-query/tumbling-window-azure-stream-analytics

NEW QUESTION 12
- (Exam Topic 3)
You have an Azure subscription that contains an Azure Synapse Analytics dedicated SQL pool named Pool1 and an Azure Data Lake Storage account named storage1. Storage1 requires secure transfers.
You need to create an external data source in Pool1 that will be used to read .orc files in storage1. 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 EXTERNAL DATA SOURCE AzureDataLakeStore

WITH

( Location1 '  [          ]   ://data@newyorktaxidataset.dfs.core.windows.net' ,
                  abfs
                  abfss
                  wasb
                  wasbs

credential = ADLS_credential ,

TYPE -  [          ]
                  BLOB_STORAGE
                  HADOOP
);              RDBMS
                  SHARP MAP MANAGER
```

Solution:
Graphical user interface, text, application, email Description automatically generated
Reference:
https://docs.microsoft.com/en-us/sql/t-sql/statements/create-external-data-source-transact-sql?view=azure-sqldw

Does this meet the goal?

A. Yes
B. No

Answer: A

NEW QUESTION 13
- (Exam Topic 3)
You are designing 2 solution that will use tables in Delta Lake on Azure Databricks. You need to minimize how long it takes to perform the following:
*Queries against non-partitioned tables
* Joins on non-partitioned columns
Which two options should you include in the solution? Each correct answer presents part of the solution. (Choose Correct Answer and Give explanation and References to Support the answers based from Data
Engineering on Microsoft Azure)

A. Z-Ordering
B. Apache Spark caching
C. dynamic file pruning (DFP)
D. the clone command

Answer: AC

Explanation:
According to the information I found on the web, two options that you should include in the solution to minimize how long it takes to perform queries and joins on non-partitioned tables are:

> Z-Ordering: This is a technique to colocate related information in the same set of files. This co-locality
is automatically used by Delta Lake in data-skipping algorithms. This behavior dramatically reduces the
amount of data that Delta Lake on Azure Databricks needs to read123.

> Apache Spark caching: This is a feature that allows you to cache data in memory or on disk for faster access. Caching can improve the performance of repeated queries and joins on the same data. You can cache Delta tables using the CACHE TABLE or CACHE LAZY commands.

To minimize the time it takes to perform queries against non-partitioned tables and joins on non-partitioned columns in Delta Lake on Azure Databricks, the following options should be included in the solution:

* A. Z-Ordering: Z-Ordering improves query performance by co-locating data that share the same column values in the same physical partitions. This reduces the need for shuffling data across nodes during query execution. By using Z-Ordering, you can avoid full table scans and reduce the amount of data processed.

* B. Apache Spark caching: Caching data in memory can improve query performance by reducing the amount of data read from disk. This helps to speed up subsequent queries that need to access the same data. When you cache a table, the data is read from the data source and stored in memory. Subsequent queries can then read the data from memory, which is much faster than reading it from disk.

References:

> Delta Lake on Databricks: https://docs.databricks.com/delta/index.html

> Best Practices for Delta Lake on
Databricks: https://databricks.com/blog/2020/05/14/best-practices-for-delta-lake-on-databricks.html

NEW QUESTION 14
- (Exam Topic 3)
You plan to use an Apache Spark pool in Azure Synapse Analytics to load data to an Azure Data Lake Storage Gen2 account.
You need to recommend which file format to use to store the data in the Data Lake Storage account. The solution must meet the following requirements:
• Column names and data types must be defined within the files loaded to the Data Lake Storage account.
• Data must be accessible by using queries from an Azure Synapse Analytics serverless SQL pool.
• Partition elimination must be supported without having to specify a specific partition. What should you recommend?

A. Delta Lake
B. JSON
C. CSV
D. ORC

Answer: D

NEW QUESTION 15
- (Exam Topic 3)
You have an Azure subscription that contains an Azure Databricks workspace named databricks1 and an Azure Synapse Analytics workspace named synapse1.
The synapse1 workspace contains an Apache Spark pool named pool1.
You need to share an Apache Hive catalog of pool1 with databricks1.
What should you do? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

From synapse1, create a linked service to:

Azure Cosmos DB
Azure Data Lake Storage Gen2
Azure SQL Database

Configure pool1 to use the linked service as:

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

Solution:
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

Does this meet the goal?

A. Yes
B. No

Answer: A

NEW QUESTION 16

- (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 17
- (Exam Topic 1)
You need to design a data storage structure for the product sales transactions. The solution must meet the sales transaction dataset requirements.
What should you include in the solution? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Answer Area

Table type to store the product sales transactions:	Hash Round-robin Replicated
When creating the table for sales transactions:	Configure a clustered index. Set the distribution column to product ID. Set the distribution column to the sales date.

Solution:
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/

Does this meet the goal?

A. Yes
B. No

Answer: A

NEW QUESTION 18
- (Exam Topic 1)
You need to design an analytical storage solution for the transactional data. 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.

Table type to store retail store data:	▼ Hash Replicated Round-robin
Table type to store promotional data:	▼ Hash Replicated Round-robin

Solution:
Graphical user interface, text, application, table Description automatically generated
Box 1: Round-robin
Round-robin tables are useful for improving loading speed.
Scenario: Partition data that contains sales transaction records. Partitions must be designed to provide efficient loads by month.
Box 2: Hash
Hash-distributed tables improve query performance on large fact tables. Reference:
https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-distribu

Does this meet the goal?

A. Yes
B. No

Answer: A

NEW QUESTION 19
- (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 20
- (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 Storage account that contains 100 GB of files. The files contain rows of text and numerical values. 75% of the rows contain description data that has an average length of 1.1 MB.
You plan to copy the data from the storage account to an enterprise data warehouse in Azure Synapse Analytics.
You need to prepare the files to ensure that the data copies quickly. Solution: You copy the files to a table that has a columnstore index. Does this meet the goal?

A. Yes
B. No

Answer: B

Explanation:
Instead convert the files to compressed delimited text files. Reference:
https://docs.microsoft.com/en-us/azure/sql-data-warehouse/guidance-for-loading-data

NEW QUESTION 21
- (Exam Topic 3)
You have an Azure Data Factory pipeline named Pipeline1!. Pipelinel contains a copy activity that sends data to an Azure Data Lake Storage Gen2 account. Pipeline 1 is executed by a schedule trigger.
You change the copy activity sink to a new storage account and merge the changes into the collaboration branch.
After Pipelinel executes, you discover that data is NOT copied to the new storage account. You need to ensure that the data is copied to the new storage account. What should you do?

A. Publish from the collaboration branch.
B. Configure the change feed of the new storage account.
C. Create a pull request.
D. Modify the schedule trigger.

Answer: A

Explanation:
CI/CD lifecycle

⟩ A development data factory is created and configured with Azure Repos Git. All developers should have permission to author Data Factory resources like pipelines and datasets.

⟩ A developer creates a feature branch to make a change. They debug their pipeline runs with their most recent changes

⟩ After a developer is satisfied with their changes, they create a pull request from their feature branch to the main or collaboration branch to get their changes reviewed by peers.

⟩ After a pull request is approved and changes are merged in the main branch, the changes get published to the development factory.
Reference: https://docs.microsoft.com/en-us/azure/data-factory/continuous-integration-delivery

NEW QUESTION 22
- (Exam Topic 3)
You need to collect application metrics, streaming query events, and application log messages for an Azure Databrick cluster.
Which type of library and workspace should you implement? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Library:

```
Azure Databricks Monitoring Library
Microsoft Azure Management Monitoring Library
PyTorch
TensorFlow
```

Workspace:

```
Azure Databricks
Azure Log Analytics
Azure Machine Learning
```

Solution:
You can send application logs and metrics from Azure Databricks to a Log Analytics workspace. It uses the Azure Databricks Monitoring Library, which is available on GitHub.
References:
https://docs.microsoft.com/en-us/azure/architecture/databricks-monitoring/application-logs

Does this meet the goal?

A. Yes
B. No

Answer: A

NEW QUESTION 23
- (Exam Topic 3)
You have an Azure Synapse Analytics dedicated SQL pool named Pool1. Pool1 contains a table named table1. You load 5 TB of data intotable1.
You need to ensure that columnstore compression is maximized for table1. Which statement should you execute?

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

Answer: B

Explanation:
Columnstore and columnstore archive compression
Columnstore tables and indexes are always stored with columnstore compression. You can further reduce the size of columnstore data by configuring an additional compression called archival compression. To perform archival compression, SQL Server runs the Microsoft XPRESS compression algorithm on the data. Add or remove archival compression by using the following data compression types:
Use COLUMNSTORE_ARCHIVE data compression to compress columnstore data with archival compression.
Use COLUMNSTORE data compression to decompress archival compression. The resulting data continue to be compressed with columnstore compression.
To add archival compression, use ALTER TABLE (Transact-SQL) or ALTER INDEX (Transact-SQL) with the REBUILD option and DATA COMPRESSION = COLUMNSTORE_ARCHIVE.
Reference: https://learn.microsoft.com/en-us/sql/relational-databases/data-compression/data-compression

NEW QUESTION 24
- (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 25
- (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.
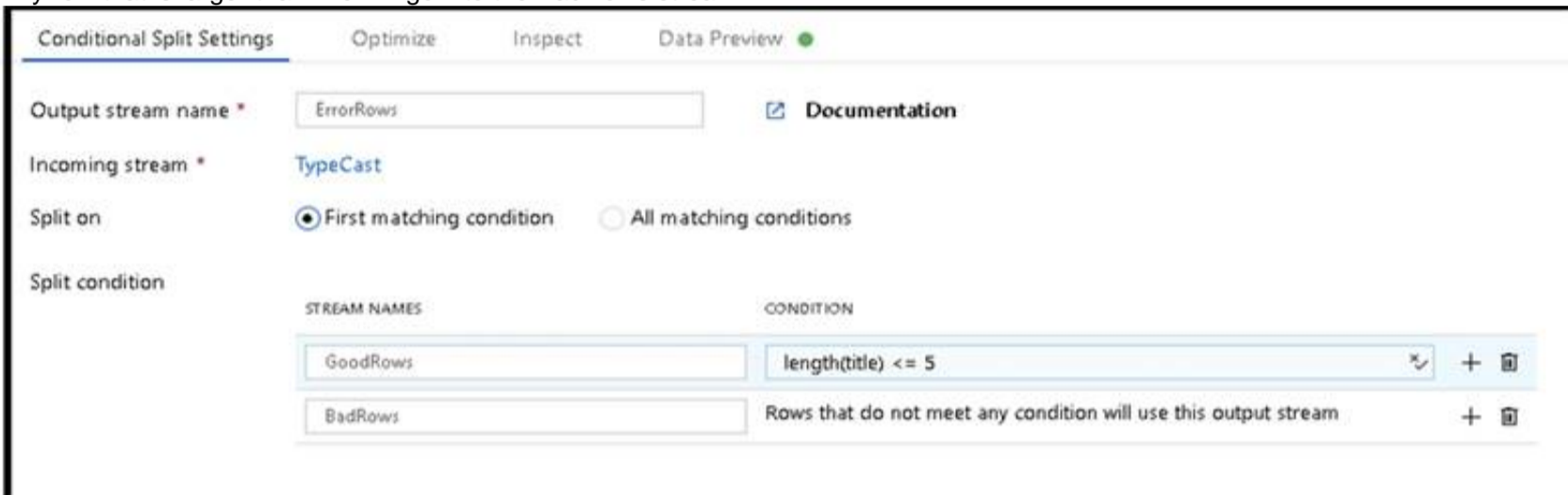
* 2. This conditional split transformation defines the maximum length of "title" to be five. Any row that is less than or equal to five will go into the GoodRows stream. Any row that is larger than five will go into the
BadRows stream. A:
* 3. Now we need to log the rows that failed. Add a sink transformation 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".
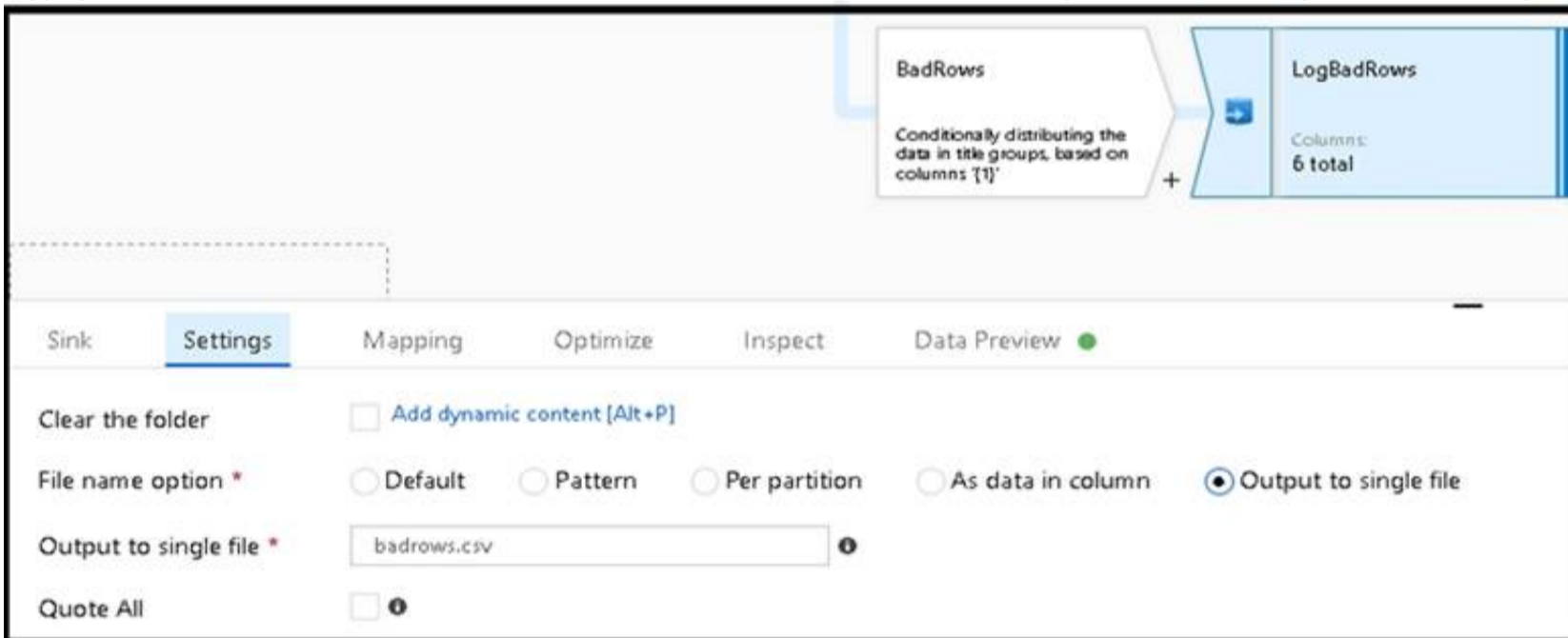
* 4. The completed data flow is shown below. We are now able to split off error rows to avoid the SQL truncation errors and put those entries into a log file. Meanwhile, successful rows can continue to write to our target database.

Reference:
https://docs.microsoft.com/en-us/azure/data-factory/how-to-data-flow-error-rows
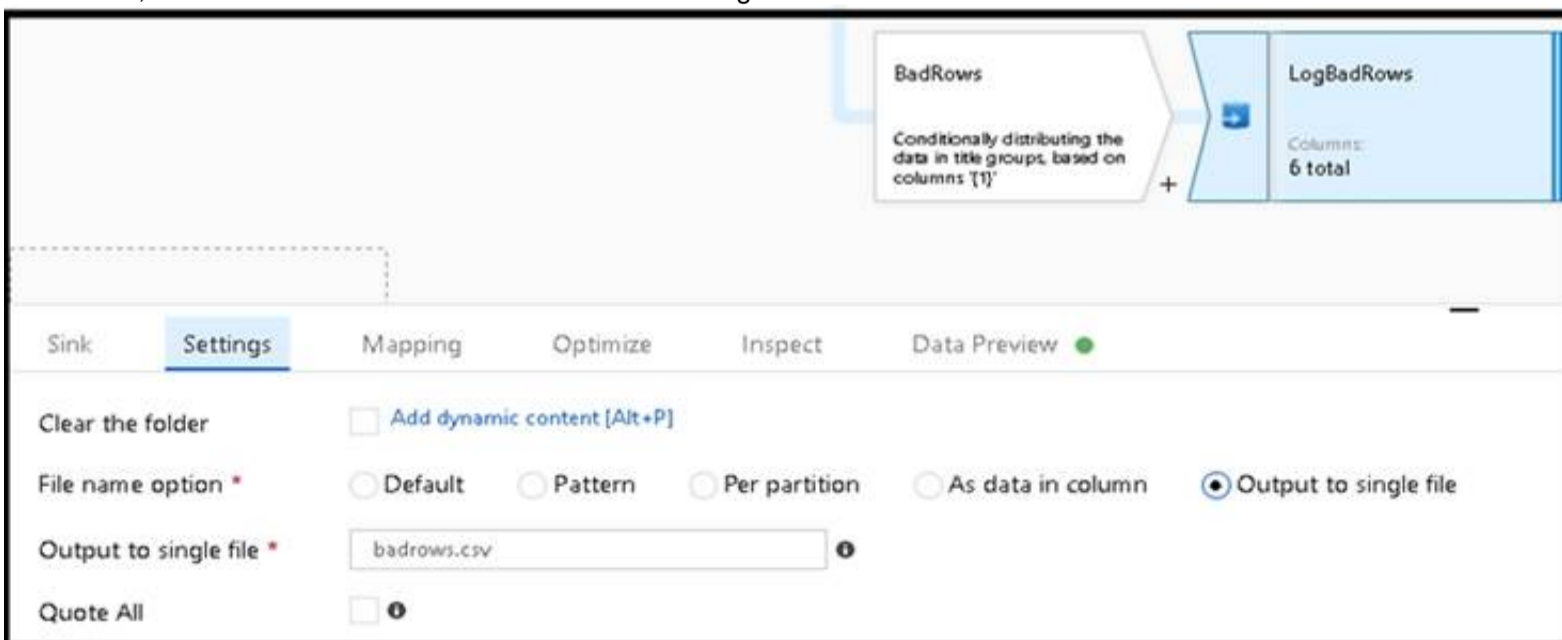
NEW QUESTION 26

- (Exam Topic 3)
You have an Azure Synapse Analytics dedicated SQL pool.
You run PDW_SHOWSPACEUSED(dbo,FactInternetSales'); and get the results shown in the following table.

ROWS	RESERVED_SPACE	DATA_SPACE	INDEX_SPACE	UNUSED_SPACE	PDW_NODE_ID	DISTRIBUTION_ID
694	2776	616	48	2112	1	1
407	2704	576	48	2080	1	2
53	2376	512	16	1848	1	3
58	2376	512	16	1848	1	4
168	2632	528	32	2072	1	5
195	2696	536	32	2128	1	6
5995	3464	1424	32	2008	1	7
0	2232	496	0	1736	1	8
264	2576	544	40	1992	1	9
3008	3016	960	32	2024	1	10
—	—	—	—	—	—	—
1550	2832	752	48	2032	1	50
1238	2832	696	40	2096	1	51
192	2632	528	32	2072	1	52
1127	2768	600	48	2040	1	53
1244	3032	704	64	2264	1	54
409	2632	568	32	2032	1	55
0	2232	496	0	1736	1	56
1437	2832	728	40	2064	1	57
0	2232	496	0	1736	1	58
584	2632	560	32	2040	1	59
225	2768	544	40	2184	1	60

Which statement accurately describes the dbo,FactInternetSales table?

A. The table contains less than 1,000 rows.
B. All distributions contain data.
C. The table is skewed.
D. The table uses round-robin distribution.

Answer: C

Explanation:
Data skew means the data is not distributed evenly across the distributions. Reference:
https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-distribu

NEW QUESTION 27
- (Exam Topic 3)
You have an Azure Data Factory pipeline that has the activities shown in the following exhibit.
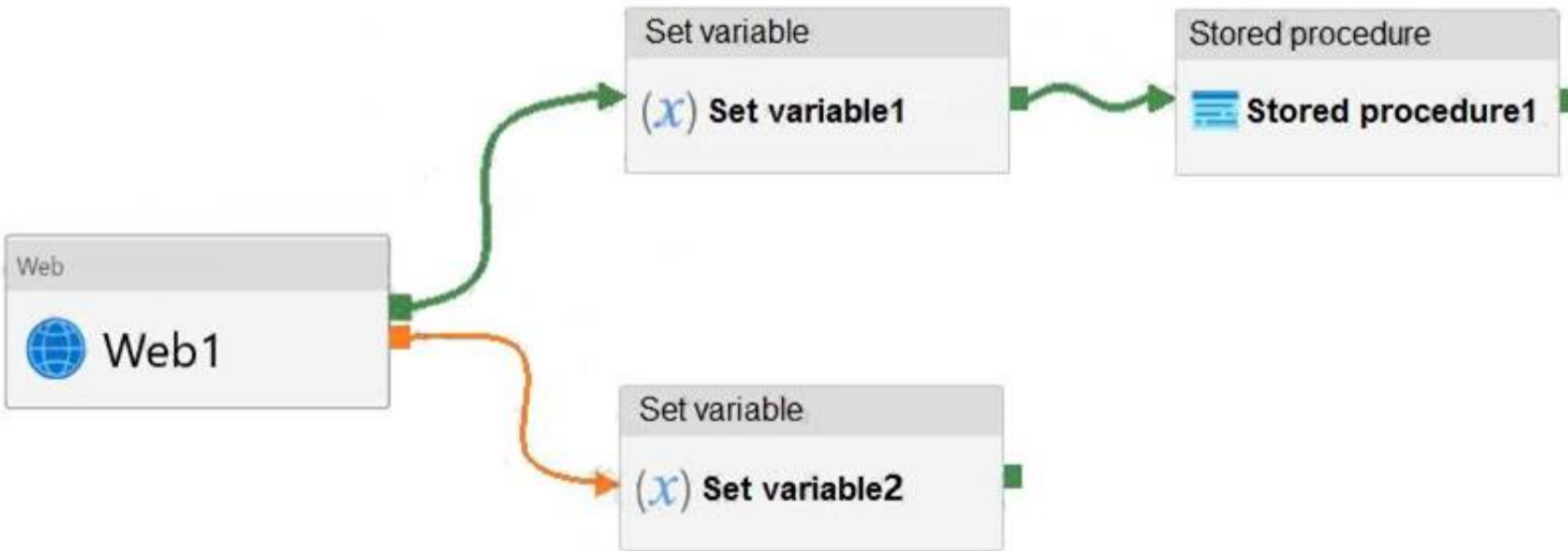
Use the drop-down menus to select the answer choice that completes each statement based on the information presented in the graphic.
NOTE: Each correct selection is worth one point.

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

complete
fail
succeed

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

Canceled
Failed
Succeeded

Solution:
Box 1: succeed
Box 2: failed Example:
Now let's say we have a pipeline with 3 activities, where Activity1 has a success path to Activity2 and a failure path to Activity3. If Activity1 fails and Activity3 succeeds, the pipeline will fail. The presence of the success path alongside the failure path changes the outcome reported by the pipeline, even though the activity executions from the pipeline are the same as the previous scenario.
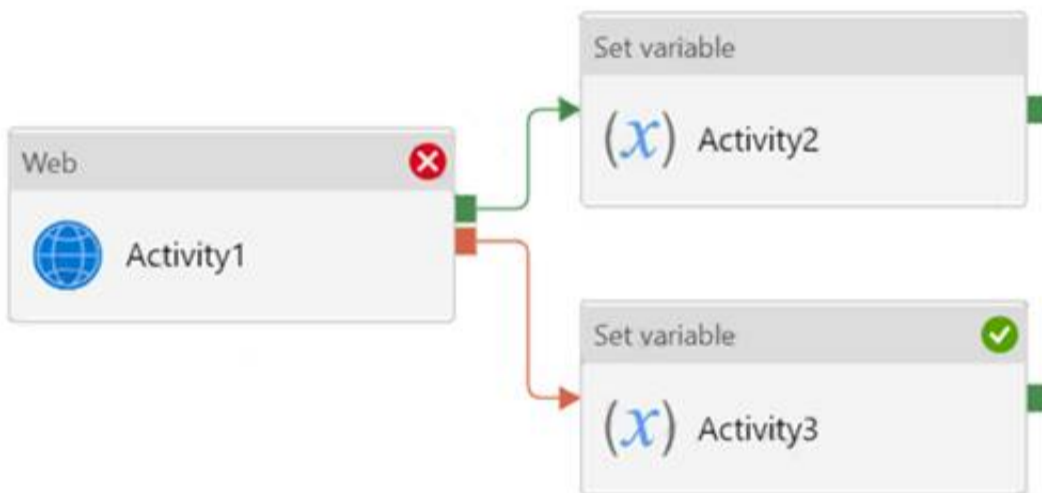
Activity1 fails, Activity2 is skipped, and Activity3 succeeds. The pipeline reports failure. Reference:
https://datasavvy.me/2021/02/18/azure-data-factory-activity-failures-and-pipeline-outcomes/

Does this meet the goal?

A. Yes
B. No

Answer: A

NEW QUESTION 28
- (Exam Topic 3)
You are monitoring an Azure Stream Analytics job.
The Backlogged Input Events count has been 20 for the last hour. You need to reduce the Backlogged Input Events count.
What should you do?

A. Drop late arriving events from the job.
B. Add an Azure Storage account to the job.
C. Increase the streaming units for the job.
D. Stop the job.

Answer: C

Explanation:
General symptoms of the job hitting system resource limits include:
> If the backlog event metric keeps increasing, it's an indicator that the system resource is constrained (either because of output sink throttling, or high CPU).
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: adjust Streaming Units.
Reference:
https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-scale-jobs https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-monitoring

NEW QUESTION 29
- (Exam Topic 3)
You have an Azure data factory.
You need to examine the pipeline failures from the last 60 days. What should you use?

A. the Activity log blade for the Data Factory resource
B. the Monitor & Manage app in Data Factory
C. the Resource health blade for the Data Factory resource

D. Azure Monitor

Answer: D

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 30
- (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 31
......

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

100% Actual & Verified — Instant Download, Please Click
[Order The DP-203 Practice Test Here](#)