

- **Vendor: Microsoft**
- **Exam Code: DP-300**
- **Exam Name: Administering Relational Databases on Microsoft Azure**
- **New Updated Questions from [Braindump2go](https://www.braindump2go.com)**
- **(Updated in [March/2023](#))**

Visit Braindump2go and Download Full Version DP-300 Exam Dumps

QUESTION 109

You are designing a security model for an Azure Synapse Analytics dedicated SQL pool that will support multiple companies.

You need to ensure that users from each company can view only the data of their respective company. Which two objects should you include in the solution? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. a column encryption key
- B. asymmetric keys
- C. a function
- D. a custom role-based access control (RBAC) role
- E. a security policy

Answer: CE

Explanation:

Row-Level Security enables you to use group membership or execution context to control access to rows in a database table. Implement RLS by using the CREATE SECURITY POLICY Transact-SQL statement, and predicates created as inline table-valued functions.

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/security/synapse-workspace-synapse-rbac>

<https://docs.microsoft.com/en-us/security/benchmark/azure/baselines/synapse-analytics-security-baseline>

<https://docs.microsoft.com/en-us/sql/relational-databases/security/row-level-security?view=sql-server-ver15>

QUESTION 110

You have an Azure subscription that contains an Azure Data Factory version 2 (V2) data factory named df1. DF1 contains a linked service.

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

You need to encrypt df1 by using key1.

What should you do first?

- A. Disable purge protection on vault1.
- B. Remove the linked service from df1.
- C. Create a self-hosted integration runtime.
- D. Disable soft delete on vault1.

Answer: B

Explanation:

[DP-300 Exam Dumps](#) **[DP-300 Exam Questions](#)** **[DP-300 PDF Dumps](#)** **[DP-300 VCE Dumps](#)**

<https://www.braindump2go.com/dp-300.html>

A customer-managed key can only be configured on an empty data Factory. The data factory can't contain any resources such as linked services, pipelines and data flows. It is recommended to enable customer-managed key right after factory creation.

Note: Azure Data Factory encrypts data at rest, including entity definitions and any data cached while runs are in progress. By default, data is encrypted with a randomly generated Microsoft-managed key that is uniquely assigned to your data factory.

Incorrect Answers:

A, D: Should enable Soft Delete and Do Not Purge on Azure Key Vault.

Using customer-managed keys with Data Factory requires two properties to be set on the Key Vault, Soft Delete and Do Not Purge. These properties can be enabled using either PowerShell or Azure CLI on a new or existing key vault.

Reference:

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

QUESTION 111

A company plans to use Apache Spark analytics to analyze intrusion detection data.

You need to recommend a solution to analyze network and system activity data for malicious activities and policy violations. The solution must minimize administrative efforts.

What should you recommend?

- A. Azure Data Lake Storage
- B. Azure Databricks
- C. Azure HDInsight
- D. Azure Data Factory

Answer: B

Explanation:

Azure DataBricks does have integration with Azure Monitor. Application logs and metrics from Azure Databricks can be send to a Log Analytics workspace.

Reference:

<https://docs.microsoft.com/en-us/azure/architecture/databricks-monitoring/application-logs>

QUESTION 112

You have an Azure data solution that contains an enterprise data warehouse in Azure Synapse Analytics named DW1. Several users execute adhoc queries to DW1 concurrently.

You regularly perform automated data loads to DW1.

You need to ensure that the automated data loads have enough memory available to complete quickly and successfully when the adhoc queries run.

What should you do?

- A. Assign a smaller resource class to the automated data load queries.
- B. Create sampled statistics to every column in each table of DW1.
- C. Assign a larger resource class to the automated data load queries.
- D. Hash distribute the large fact tables in DW1 before performing the automated data loads.

Answer: C

Explanation:

The performance capacity of a query is determined by the user's resource class.

Smaller resource classes reduce the maximum memory per query, but increase concurrency.

Larger resource classes increase the maximum memory per query, but reduce concurrency.

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/resource-classes-for-workload-management>

QUESTION 113

You are monitoring an Azure Stream Analytics job.

You discover that the Backlogged input Events metric is increasing slowly and is consistently non-zero.

You need to ensure that the job can handle all the events.

[DP-300 Exam Dumps](#) [DP-300 Exam Questions](#) [DP-300 PDF Dumps](#) [DP-300 VCE Dumps](#)

<https://www.braindump2go.com/dp-300.html>

What should you do?

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

Answer: D

Explanation:

Backlogged Input Events: Number of input events that are backlogged. A non-zero value for this metric implies that your job isn't able to keep up with the number of incoming events. If this value is slowly increasing or consistently non-zero, you should scale out your job, by increasing the SUs.

Reference:

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

QUESTION 114

You have an Azure Stream Analytics job.

You need to ensure that the job has enough streaming units provisioned.

You configure monitoring of the SU % Utilization metric.

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

NOTE: Each correct selection is worth one point.

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

Answer: CD

Explanation:

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

Note: Backlogged Input Events: Number of input events that are backlogged. A non-zero value for this metric implies that your job isn't able to keep up with the number of incoming events. If this value is slowly increasing or consistently non-zero, you should scale out your job, by increasing the SUs.

Reference:

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

QUESTION 115

You have an Azure Databricks resource.

You need to log actions that relate to changes in compute for the Databricks resource.

Which Databricks services should you log?

- A. clusters
- B. jobs
- C. DBFS
- D. SSH
- E. workspace

Answer: A

Explanation:

Clusters logs include information regarding changes in compute.

Incorrect:

Not E: Workspace logs do not include information related to changes in compute.

Reference:

<https://docs.microsoft.com/en-us/azure/databricks/administration-guide/account-settings/azure-diagnostic-logs#configure-diagnostic-log-delivery>

QUESTION 116

Your company uses Azure Stream Analytics to monitor devices.

The company plans to double the number of devices that are monitored.

You need to monitor a Stream Analytics job to ensure that there are enough processing resources to handle the additional load.

Which metric should you monitor?

- A. Input Deserialization Errors
- B. Late Input Events
- C. Early Input Events
- D. Watermark delay

Answer: D

Explanation:

The Watermark delay metric is computed as the wall clock time of the processing node minus the largest watermark it has seen so far.

The watermark delay metric can rise due to:

1. Not enough processing resources in Stream Analytics to handle the volume of input events.
2. Not enough throughput within the input event brokers, so they are throttled.
3. Output sinks are not provisioned with enough capacity, so they are throttled.

Reference:

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

QUESTION 117

You manage an enterprise data warehouse in Azure Synapse Analytics.

Users report slow performance when they run commonly used queries. Users do not report performance changes for infrequently used queries.

You need to monitor resource utilization to determine the source of the performance issues.

Which metric should you monitor?

- A. Local tempdb percentage
- B. DWU percentage
- C. Data Warehouse Units (DWU) used
- D. Cache hit percentage

Answer: D

Explanation:

You can use Azure Monitor to view cache metrics to troubleshoot query performance.

The key metrics for troubleshooting the cache are Cache hit percentage and Cache used percentage.

Possible scenario: Your current working data set cannot fit into the cache which causes a low cache hit percentage due to physical reads. Consider scaling up your performance level and rerun your workload to populate the cache.

Reference:

<https://docs.microsoft.com/da-dk/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-how-to-monitor-cache>

QUESTION 118

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 Pool1 and query sys.dm_pdw_nodes_db_partition_stats.
- B. Connect to the built-in pool and run DBCC CHECKALLOC.
- C. Connect to Pool1 and run DBCC CHECKALLOC.

[DP-300 Exam Dumps](#) [DP-300 Exam Questions](#) [DP-300 PDF Dumps](#) [DP-300 VCE Dumps](#)

<https://www.braindump2go.com/dp-300.html>

D. Connect to the built-in pool and query sys.dm_pdw_nodes_db_partition_stats.

Answer: A

Explanation:

First connect to Pool1, not the built-in serverless pool, then use 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>

QUESTION 119

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 | 680 | 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 |
| 384 | 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 10,000 rows.
- B. All distributions contain data.
- C. The table uses round-robin distribution
- D. The table is skewed.

Answer: D

Explanation:

The rows per distribution can vary up to 10% without a noticeable impact on performance. Here the distribution varies more than 10%. It is skewed.

Note: SHOWSPACEUSED displays the number of rows, disk space reserved, and disk space used for a specific table, or for all tables in a Azure Synapse Analytics or Parallel Data Warehouse database.

This is a very quick and simple way to see the number of table rows that are stored in each of the 60 distributions of your database. Remember that for the most balanced performance, the rows in your distributed table should be spread evenly across all the distributions.

ROUND_ROBIN distributed tables should not be skewed. Data is distributed evenly across the nodes by design.

Reference:

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

<https://github.com/rgl/azure-content/blob/master/articles/sql-data-warehouse/sql-data-warehouse-manage-distributed-data-skew.md>

QUESTION 120

[DP-300 Exam Dumps](#) [DP-300 Exam Questions](#) [DP-300 PDF Dumps](#) [DP-300 VCE Dumps](#)

<https://www.braindump2go.com/dp-300.html>

You are designing a dimension table in an Azure Synapse Analytics dedicated SQL pool. You need to create a surrogate key for the table. The solution must provide the fastest query performance. What should you use for the surrogate key?

- A. an IDENTITY column
- B. a GUID column
- C. a sequence object

Answer: A

Explanation:

Dedicated SQL pool supports many, but not all, of the table features offered by other databases. Surrogate keys are not supported. Implement it with an Identity column.

Reference:

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

QUESTION 121

You are designing a star schema for a dataset that contains records of online orders. Each record includes an order date, an order due date, and an order ship date.

You need to ensure that the design provides the fastest query times of the records when querying for arbitrary date ranges and aggregating by fiscal calendar attributes.

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

NOTE: Each correct selection is worth one point.

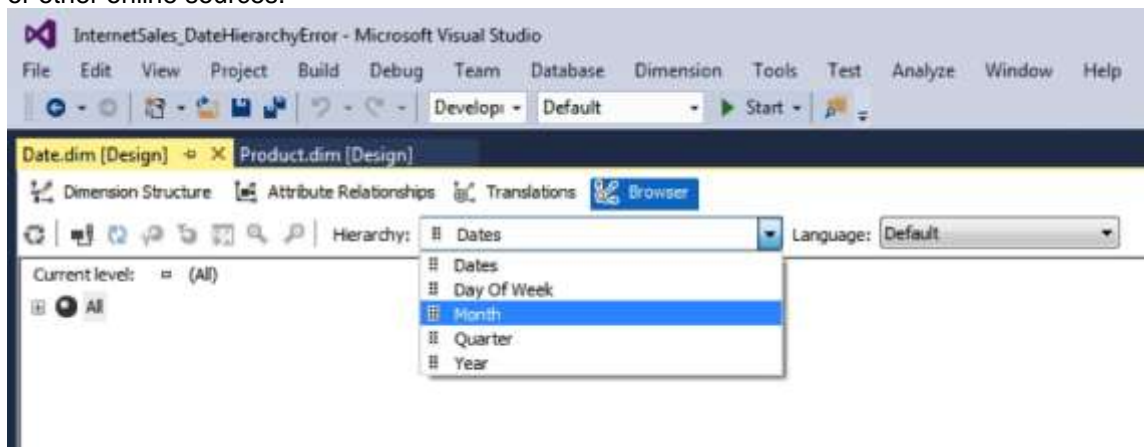
- A. Create a date dimension table that has a DateTime key.
- B. Create a date dimension table that has an integer key in the format of YYYYMMDD.
- C. Use built-in SQL functions to extract date attributes.
- D. Use integer columns for the date fields.
- E. Use DateTime columns for the date fields.

Answer: BD

Explanation:

Why use a Date Dimension Table in a Data Warehouse.

The Date dimension is one of these dimension tables related to the Fact. Here is a simple Data Diagram for a Data Mart of Internet Sales information for the Adventure Works DW database which can be obtained for free from CodePlex or other online sources.



The relationship is created by the surrogate keys columns (integer data type) rather than the date data type. The query users have to write against a Data Mart are much simpler than against a transaction database. There are less joins because of the one to many relationships between the fact dimension table(s). The dimension tables are confusing to someone who has been normalizing databases as a career. The dimension is a flattened or de-normalized table. This creates cases of duplicate data, but the simplistic query overrides the duplicate data in a dimensional model.

Reference:

<https://www.mssqltips.com/sqlservertip/3117/defining-role-playing-dimensions-for-sql-server-analysis-services/>

https://community.idera.com/database-tools/blog/b/community_blog/posts/why-use-a-date-dimension-table-in-a-data-warehouse

QUESTION 122

You have an Azure Data Factory pipeline that is triggered hourly.

The pipeline has had 100% success for the past seven days.

The pipeline execution fails, and two retries that occur 15 minutes apart also fail. The third failure returns the following error.

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

What is a possible cause of the error?

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

Answer: B

Explanation:

A file is missing.

Incorrect:

Not A, not C, not D: Time of the error is 07:45.

QUESTION 123

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

Answer: B

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>

QUESTION 124

You have the following Azure Data Factory pipelines:

- Ingest Data from System1
- Ingest Data from System2
- Populate Dimensions
- Populate Facts

Ingest Data from System1 and Ingest Data from System2 have no dependencies. Populate Dimensions must execute after Ingest Data from System1 and Ingest Data from System2. Populate Facts must execute after the Populate Dimensions pipeline. All the pipelines must execute every eight hours.

[DP-300 Exam Dumps](#) **[DP-300 Exam Questions](#) **[DP-300 PDF Dumps](#) **[DP-300 VCE Dumps](#)******

<https://www.braindump2go.com/dp-300.html>

What should you do to schedule the pipelines for execution?

- A. Add a schedule trigger to all four pipelines.
- B. Add an event trigger to all four pipelines.
- C. Create a parent pipeline that contains the four pipelines and use an event trigger.
- D. Create a parent pipeline that contains the four pipelines and use a schedule trigger.

Answer: D

Explanation:

<https://www.mssqltips.com/sqlservertip/6137/azure-data-factory-control-flow-activities-overview/>

QUESTION 125

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

Answer: A

Explanation:

The Tumbling window trigger supports backfill scenarios. Pipeline runs can be scheduled for windows in the past.

Incorrect Answers:

D: Schedule trigger does not support backfill scenarios. Pipeline runs can be executed only on time periods from the current time and the future.

Reference:

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

QUESTION 126

You have an Azure Data Factory that contains 10 pipelines.

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

What should you add to each pipeline?

- A. an annotation
- B. a resource tag
- C. a run group ID
- D. a user property
- E. a correlation ID

Answer: A

Explanation:

Azure Data Factory annotations help you easily filter different Azure Data Factory objects based on a tag.

You can define tags so you can see their performance or find errors faster.

Reference:

<https://www.techtalkcorner.com/monitor-azure-data-factory-annotations/>

QUESTION 127

[DP-300 Exam Dumps](#) [DP-300 Exam Questions](#) [DP-300 PDF Dumps](#) [DP-300 VCE Dumps](#)

<https://www.braindump2go.com/dp-300.html>

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 use an Azure Data Factory schedule trigger to execute a pipeline that executes mapping data flow, and then inserts the data into the data warehouse.

Does this meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

Correct solution: You use an Azure Data Factory schedule trigger to execute a pipeline that executes an Azure Databricks notebook, and then inserts the data into the data warehouse.

Reference:

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

QUESTION 128

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.

Correct solution: You use an Azure Data Factory schedule trigger to execute a pipeline that executes an Azure Databricks notebook, and then inserts the data into the data warehouse.

Reference:

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

QUESTION 129

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 use an Azure Data Factory schedule trigger to execute a pipeline that executes an Azure Databricks notebook, and then inserts the data into the data warehouse.

Does this meet the goal?

[DP-300 Exam Dumps](#) [DP-300 Exam Questions](#) [DP-300 PDF Dumps](#) [DP-300 VCE Dumps](#)

<https://www.braindump2go.com/dp-300.html>

- A. Yes
- B. No

Answer: A

Explanation:

An Azure Data Factory can trigger a Databricks notebook.

Reference:

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

QUESTION 130

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 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: B

Explanation:

Azure Synapse Analytics does not support R script.

Correct solution: You use an Azure Data Factory schedule trigger to execute a pipeline that executes an Azure Databricks notebook, and then inserts the data into the data warehouse.

Reference:

<https://docs.microsoft.com/en-us/azure/architecture/data-guide/technology-choices/r-developers-guide>

QUESTION 131

You plan to perform batch processing in Azure Databricks once daily.

Which type of Databricks cluster should you use?

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

Answer: A

Explanation:

Azure Databricks makes a distinction between all-purpose clusters and job clusters. You use all-purpose clusters to analyze data collaboratively using interactive notebooks. You use job clusters to run fast and robust automated jobs. The Azure Databricks job scheduler creates a job cluster when you run a job on a new job cluster and terminates the cluster when the job is complete.

Reference:

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

QUESTION 132

Hotspot Question

You have an Azure Synapse Analytics dedicated SQL pool named Pool1 and an Azure Data Lake Storage Gen2 account named Account1.

You plan to access the files in Account1 by using an external table.

You need to create a data source in Pool1 that you can reference when you create the external table.

[DP-300 Exam Dumps](#) [DP-300 Exam Questions](#) [DP-300 PDF Dumps](#) [DP-300 VCE Dumps](#)

<https://www.braindump2go.com/dp-300.html>

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

Answer Area

```
CREATE EXTERNAL DATA SOURCE source1
```

```
WITH
```

```
( LOCATION = 'https://account1. 
```

| |
|-------|
| ▼ |
| blob |
| dfs |
| table |

| |
|---------------------|
| ▼ |
| PUSHDOWN = ON |
| TYPE = BLOB_STORAGE |
| TYPE = HADOOP |

```
)
```

Answer:

Answer Area

```
CREATE EXTERNAL DATA SOURCE source1
```

```
WITH
```

```
( LOCATION = 'https://account1. 
```

| |
|-------|
| ▼ |
| blob |
| dfs |
| table |

| |
|---------------------|
| ▼ |
| PUSHDOWN = ON |
| TYPE = BLOB_STORAGE |
| TYPE = HADOOP |

```
)
```

Explanation:

Box 1: dfs

For Azure Data Lake Store Gen 2 used the following syntax:

http[s] <storage_account>.dfs.core.windows.net/<container>/subfolders

Incorrect:

Not blob: blob is used for Azure Blob Storage. Syntax:

http[s] <storage_account>.blob.core.windows.net/<container>/subfolders

Box 2: TYPE = HADOOP

Syntax for CREATE EXTERNAL DATA SOURCE.

External data sources with TYPE=HADOOP are available only in dedicated SQL pools.

CREATE EXTERNAL DATA SOURCE <data_source_name>

WITH -

(LOCATION = '<prefix>://<path>'

[, CREDENTIAL = <database scoped credential>]

```
, TYPE = HADOOP
)
```

Reference:

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

QUESTION 133

Hotspot Question

You plan to develop a dataset named Purchases by using Azure Databricks. Purchases will contain the following columns:

- ProductID
- ItemPrice
- LineTotal
- Quantity
- StoreID
- Minute
- Month
- Hour
- Year
- Day

You need to store the data to support hourly incremental load pipelines that will vary for each StoreID. 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.

Answer Area

```
df.write
```

| | |
|----------------------|---|
| <input type="text"/> | <input type="text"/> |
| .bucketBy | ("*") |
| .partitionBy | ("StoreID", "Hour") |
| .range | ("StoreID", "Year", "Month", "Day", "Hour") |
| .sortBy | ("Year", "Month", "Day", "Hour", "StoreID") |

```
.mode("append")
```

| |
|---------------------------|
| <input type="text"/> |
| .csv("/Purchases") |
| .json("/Purchases") |
| .parquet("/Purchases") |
| saveAsTable("/Purchases") |

Answer:

Answer Area

```
df.write
```

| | |
|---------------------|---|
| .bucketBy | ("**") |
| .partitionBy | ("StoreID", "Hour") |
| .range | ("StoreID", "Year", "Month", "Day", "Hour") |
| .sortBy | ("Year", "Month", "Day", "Hour", "StoreID") |

```
.mode("append")
```

| |
|-------------------------------|
| .csv("/Purchases") |
| .json("/Purchases") |
| .parquet("/Purchases") |
| .saveAsTable("/Purchases") |

Explanation:

Box 1: .partitionBy

Example:

```
df.write.partitionBy("y","m","d")
```

```
.mode(SaveMode.Append)
```

```
.parquet("/data/hive/warehouse/db_name.db/" + tableName)
```

Box 2: ("Year","Month","Day","Hour","StoreID")

Box 3: .parquet("/Purchases")

Reference:

<https://intellipaat.com/community/11744/how-to-partition-and-write-dataframe-in-spark-without-deleting-partitions-with-no-new-data>

QUESTION 134

Hotspot Question

You are building a database in an Azure Synapse Analytics serverless SQL pool.

You have data stored in Parquet files in an Azure Data Lake Storage Gen2 container.

Records are structured as shown in the following sample.

```
{
  "id":123,
  "address_housenumber": "19c",
  "address_line1": "Memory Lane",
  "applicant1_name": "Jane",
  "applicant2_name": "Dev"
}
```

The records contain two applicants at most.

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

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

NOTE: Each correct selection is worth one point.

Answer Area

| | |
|-----------------------|----------------|
| | ▼ applications |
| CREATE EXTERNAL TABLE | |
| CREATE TABLE | |
| CREATE VIEW | |

```

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

| | |
|-------------|--|
| | ▼ (BULK 'https://contosol.dfs.core.windows.net/ applications/year=/*.parquet', |
| CROSS APPLY | |
| OPENJSON | |
| OPENROWSET | |

Answer:

Answer Area

| | |
|-----------------------|----------------|
| | ▼ applications |
| CREATE EXTERNAL TABLE | |
| CREATE TABLE | |
| CREATE VIEW | |

```

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

| | |
|-------------|--|
| | ▼ (BULK 'https://contosol.dfs.core.windows.net/ applications/year=/*.parquet', |
| CROSS APPLY | |
| OPENJSON | |
| OPENROWSET | |

Explanation:

Box 1: CREATE EXTERNAL TABLE

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

Syntax:

CREATE EXTERNAL TABLE { database_name.schema_name.table_name | schema_name.table_name |

[DP-300 Exam Dumps](#) **[DP-300 Exam Questions](#) **[DP-300 PDF Dumps](#) **[DP-300 VCE Dumps](#)******

<https://www.braindump2go.com/dp-300.html>


```
table_name }  
( <column_definition> [ ,...n ] )  
WITH (  
LOCATION = 'folder_or_filepath',  
DATA_SOURCE = external_data_source_name,  
FILE_FORMAT = external_file_format_name
```

Box 2. OPENROWSET

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

Example:

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

Reference:

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

QUESTION 135

Hotspot Question

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 |
| SessionsWithEvents | 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.

Answer Area

EventCategory:

| | |
|------------|---|
| | ▼ |
| DimChannel | |
| DimDate | |
| DimEvent | |
| FactEvents | |

ChannelGrouping:

| | |
|------------|---|
| | ▼ |
| DimChannel | |
| DimDate | |
| DimEvent | |
| FactEvents | |

TotalEvents:

| | |
|------------|---|
| | ▼ |
| DimChannel | |
| DimDate | |
| DimEvent | |
| FactEvents | |

Answer:**Answer Area**

EventCategory:

| | |
|------------|---|
| | ▼ |
| DimChannel | |
| DimDate | |
| DimEvent | |
| FactEvents | |

ChannelGrouping:

| | |
|------------|---|
| | ▼ |
| DimChannel | |
| DimDate | |
| DimEvent | |
| FactEvents | |

TotalEvents:

| | |
|------------|---|
| | ▼ |
| DimChannel | |
| DimDate | |
| DimEvent | |
| FactEvents | |

Explanation:

Box 1: DimEvent

Box 2: DimChannel

Dimension tables describe business entities - the things you model. Entities can include products, people, places, and concepts including time itself. The most consistent table you'll find in a star schema is a date dimension table. A dimension table contains a key column (or columns) that acts as a unique identifier, and descriptive columns.

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>

QUESTION 136

Drag and Drop Question

You plan to create a table in an Azure Synapse Analytics dedicated SQL pool.

Data in the table will be retained for five years. Once a year, data that is older than five years will be deleted.

You need to ensure that the data is distributed evenly across partitions. The solutions must minimize the amount of time required to delete old data.

How should you complete the Transact-SQL statement? To answer, drag the appropriate values to the correct targets.

Each value may be used once, more than once, or not at all.

You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

| Values | Answer Area |
|------------------|--|
| CustomerKey | <pre>CREATE TABLE [dbo].[FactSales] ([ProductKey] int NOT NULL , [OrderDateKey] int NOT NULL , [CustomerKey] int NOT NULL , [SalesOrderNumber] nvarchar (20) NOT NULL , [OrderQuantity] smallint NOT NULL , [UnitPrice] money NOT NULL) WITH (CLUSTERED COLUMNSTORE INDEX , DISTRIBUTION = [] ([ProductKey]) , PARTITION ([]] RANGE RIGHT FOR VALUES (20170101, 20180101, 20190101, 20200101, 20210101)))</pre> |
| HASH | |
| ROUND_ROBIN | |
| REPLICATE | |
| OrderDateKey | |
| SalesOrderNumber | |

Answer:

| Values | Answer Area |
|------------------|--|
| CustomerKey | <pre>CREATE TABLE [dbo].[FactSales] ([ProductKey] int NOT NULL , [OrderDateKey] int NOT NULL , [CustomerKey] int NOT NULL , [SalesOrderNumber] nvarchar (20) NOT NULL , [OrderQuantity] smallint NOT NULL , [UnitPrice] money NOT NULL) WITH (CLUSTERED COLUMNSTORE INDEX , DISTRIBUTION = [] ([ProductKey]) , PARTITION ([]] RANGE RIGHT FOR VALUES (20170101, 20180101, 20190101, 20200101, 20210101)))</pre> |
| ROUND_ROBIN | |
| REPLICATE | |
| SalesOrderNumber | |
| | |
| | |

Explanation:

Box 1: HASH

DISTRIBUTION = HASH (distribution_column_name) Assigns each row to one distribution by hashing the value stored in distribution_column_name. The algorithm is deterministic, which means it always hashes the same value to the same distribution. The distribution column should be defined as NOT NULL because all rows that have NULL are assigned to the same distribution.

Box 2: OrderDateKey

In most cases, table partitions are created on a date column.

A way to eliminate rollbacks is to use Metadata Only operations like partition switching for data management. For example, rather than execute a DELETE statement to delete all rows in a table where the order_date was in October of 2001, you could partition your data early. Then you can switch out the partition with data for an empty partition from another table.

Reference:

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

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/best-practices-dedicated-sql-pool>

QUESTION 137

Drag and Drop Question

You are creating a managed data warehouse solution on Microsoft Azure.

You must use PolyBase to retrieve data from Azure Blob storage that resides in parquet format and load the data into a large table called FactSalesOrderDetails.

You need to configure Azure Synapse Analytics to receive the data.

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

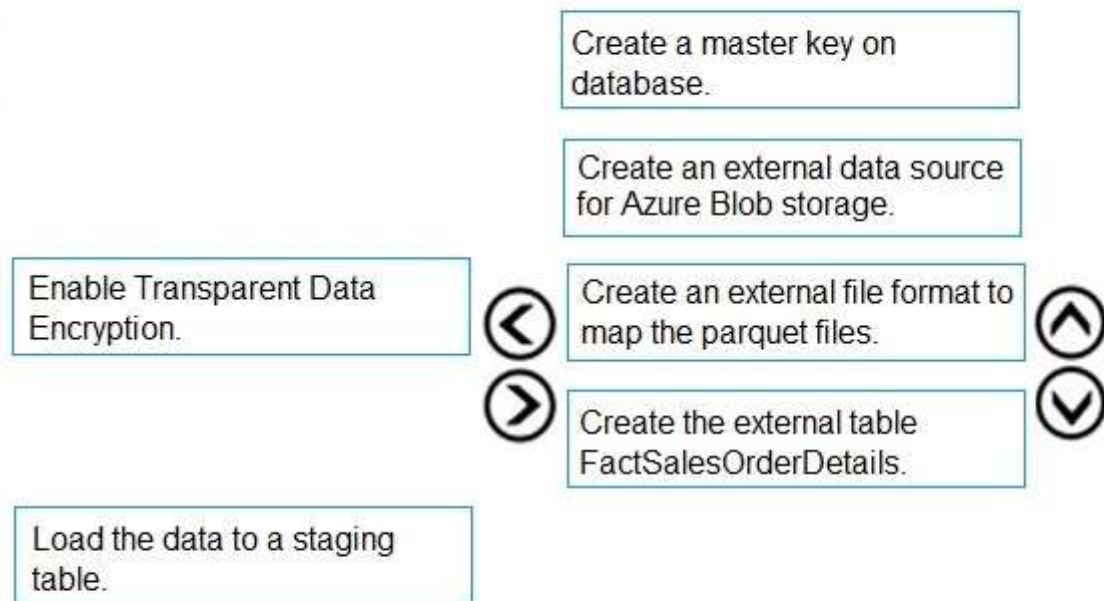
Actions

Answer Area

- Create an external data source for Azure Blob storage.
- Create a master key on database.
- Enable Transparent Data Encryption.
- Create the external table FactSalesOrderDetails.
- Load the data to a staging table.
- Create an external file format to map the parquet files.



Answer:

Actions**Answer Area****Explanation:**

To query the data in your Hadoop data source, you must define an external table to use in Transact-SQL queries. The following steps describe how to configure the external table.

Step 1: Create a master key on database.

1. Create a master key on the database. The master key is required to encrypt the credential secret.
(Create a database scoped credential for Azure blob storage.)

Step 2: Create an external data source for Azure Blob storage.

2. Create an external data source with CREATE EXTERNAL DATA SOURCE..

Step 3: Create an external file format to map the parquet files.

3. Create an external file format with CREATE EXTERNAL FILE FORMAT.

Step 4: Create an external table FactSalesOrderDetails

4. Create an external table pointing to data stored in Azure storage with CREATE EXTERNAL TABLE.

Reference:

<https://docs.microsoft.com/en-us/sql/relational-databases/polybase/polybase-configure-azure-blob-storage>

QUESTION 138

Hotspot Question

You configure version control for an Azure Data Factory instance as shown in the following exhibit.

- Home
- Linked services
- Integration runtimes
- Azure Purview (Preview)
- Source control
 - Git configuration**
 - ARM template
 - Parameterization template
- Author
 - Triggers
 - Global parameters
- Security
 - Customer managed key
 - Managed private endpoints

Git repository

Git repository information associated with your data factory. [CI/CD best practices](#)

[Setting](#) [Disconnect](#)

| | |
|----------------------|------------------|
| Repository type | Azure DevOps Git |
| Azure DevOps Account | CONTOSO |
| Project name | Data |
| Repository name | dwh_batchetl |
| Collaboration branch | main |
| Publish branch | adf_publish |
| Root folder | / |

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.

Answer Area

Azure Resource Manager (ARM) templates for the pipeline assets as stored in

▼

/

adf_publish

main

Parameterization template

A Data Factory Azure Resource Manager (ARM) template named contososales can be found in

▼

/contososales

/dwh_batchetl/adf_publish/contososales

/main

Answer:

Answer Area

Azure Resource Manager (ARM) templates for the pipeline assets as stored in

| | |
|---------------------------|---|
| | ▼ |
| / | |
| adf_publish | |
| main | |
| Parameterization template | |

A Data Factory Azure Resource Manager (ARM) template named contososales can be found in

| | |
|--|---|
| | ▼ |
| /contososales | |
| /dwh_batchetl/adf_publish/contososales | |
| /main | |

Explanation:

Box 1: adf_publish

By default, data factory generates the Resource Manager templates of the published factory and saves them into a branch called adf_publish. To configure a custom publish branch, add a publish_config.json file to the root folder in the collaboration branch. When publishing, ADF reads this file, looks for the field publishBranch, and saves all Resource Manager templates to the specified location. If the branch doesn't exist, data factory will automatically create it. And example of what this file looks like is below:

```
{  
  "publishBranch": "factory/adf_publish"  
}
```

Box 2: /dwh_barchlet/ adf_publish/contososales

RepositoryName: Your Azure Repos code repository name. Azure Repos projects contain Git repositories to manage your source code as your project grows. You can create a new repository or use an existing repository that's already in your project.

Reference:

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

QUESTION 139

Hotspot Question

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

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

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

Use the drop-down menus to select the answer choice that completes each statement based on the information presented in the graphic.

Answer Area

The query results include only [answer choice] in the csv/busfare folder.

| | |
|---|---|
| | ▼ |
| CSV files in the tripdata_2020 subfolder | |
| files that have files names beginning with "tripdata_2020" | |
| CSV files that have file names containing "tripdata_202" | |
| CSV files that have file named beginning with "tripdata_2020" | |

The query assumes that the first row in a CSV file is [answer choice] row.

| | |
|----------|---|
| | ▼ |
| a header | |
| a data | |
| an empty | |

Answer:

Answer Area

The query results include only [answer choice] in the csv/busfare folder.

| | |
|---|---|
| | ▼ |
| CSV files in the tripdata_2020 subfolder | |
| files that have files names beginning with "tripdata_2020" | |
| CSV files that have file names containing "tripdata_202" | |
| CSV files that have file named beginning with "tripdata_2020" | |

The query assumes that the first row in a CSV file is [answer choice] row.

| | |
|----------|---|
| | ▼ |
| a header | |
| a data | |
| an empty | |

Explanation:

Box 1: CSV files that have file named beginning with "tripdata_2020"

Box 2: a header

FIRSTROW = 'first_row'

Specifies the number of the first row to load. The default is 1 and indicates the first row in the specified data file. The row numbers are determined by counting the row terminators. FIRSTROW is 1-based.

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/develop-openrowset>

QUESTION 140

Hotspot Question

You have an Azure subscription that is linked to a hybrid Azure Active Directory (Azure AD) tenant. The subscription contains an Azure Synapse Analytics SQL pool named Pool1.

You need to recommend an authentication solution for Pool1. The solution must support multi-factor authentication (MFA) and database-level authentication.

Which authentication solution or solutions should you include in the recommendation? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Answer Area

MFA:

| | |
|-------------------------------------|---|
| | ▼ |
| Azure AD authentication | |
| Microsoft SQL Server authentication | |
| Passwordless authentication | |
| Windows authentication | |

Database-level authentication:

| | |
|-----------------------------|---|
| | ▼ |
| Application roles | |
| Contained database users | |
| Database roles | |
| Microsoft SQL Server logins | |

Answer:

Answer Area

MFA:

| | |
|-------------------------------------|---|
| | ▼ |
| Azure AD authentication | |
| Microsoft SQL Server authentication | |
| Passwordless authentication | |
| Windows authentication | |

Database-level authentication:

| | |
|-----------------------------|---|
| | ▼ |
| Application roles | |
| Contained database users | |
| Database roles | |
| Microsoft SQL Server logins | |

Explanation:

Box 1: Azure AD authentication

Azure Active Directory authentication supports Multi-Factor authentication through Active Directory Universal Authentication.

Box 2: Contained database users

Azure Active Directory Uses contained database users to authenticate identities at the database level.

Incorrect:

SQL authentication: To connect to dedicated SQL pool (formerly SQL DW), you must provide the following information:

[DP-300 Exam Dumps](#) [DP-300 Exam Questions](#) [DP-300 PDF Dumps](#) [DP-300 VCE Dumps](#)

<https://www.braindump2go.com/dp-300.html>



Braindump2go Guarantee All Exams 100% Pass
One Time!

- Fully qualified servername
- Specify SQL authentication
- Username
- Password
- Default database (optional)

Reference:

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

[DP-300 Exam Dumps](#) **[DP-300 Exam Questions](#)** **[DP-300 PDF Dumps](#)** **[DP-300 VCE Dumps](#)**

<https://www.braindump2go.com/dp-300.html>