

Pytanie 1:

Pominięto

You are designing a streaming data solution that will ingest variable volumes of data.

You need to ensure that you can change the partition count after creation.

Which service should you use to ingest the data?

- ☐ **Azure Event Hubs Dedicated**

(Poprawne)

- ☐ **Azure Stream Analytics**

- ☐ **Azure Data Factory**

- ☐ **Azure Synapse Analytics**

Wyjaśnienie

Correct Answer: A

You can't change the partition count for an event hub after its creation except for the event hub in a dedicated cluster.

Reference:

<https://docs.microsoft.com/en-us/azure/event-hubs/event-hubs-features>

Pytanie 2:

Pominięto

You are designing a date dimension table in an Azure Synapse Analytics dedicated SQL pool. The date dimension table will be used by all the fact tables.

Which distribution type should you recommend to minimize data movement?

☐

HASH

☐

REPLICATE

(Poprawne)

☐

ROUND_ROBIN

Wyjaśnienie

Correct Answer: B

A replicated table has a full copy of the table available on every Compute node. Queries run fast on replicated tables since joins on replicated tables don't require data movement. Replication requires extra storage, though, and isn't practical for large tables.

Incorrect Answers:

A: A hash distributed table is designed to achieve high performance for queries on large tables.

C: A round-robin table distributes table rows evenly across all distributions. The rows are distributed randomly. Loading data into a round-robin table is fast. Keep in mind that queries can require more data movement than the other distribution methods.

Reference:

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

Pytanie 3:

Pominięto

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.

Hot Area:

Answer Area

```
df.write
```

.bucketBy	("*")
.format	("GeographyRegionID")
.partitionBy	("GeographyRegionID", "Year", "Month", "Day")
.sortBy	("Year", "Month", "Day", "GeographyRegionID")

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


.csv("/DBTBL1")
.json("/DBTBL1")
.parquet("/DBTBL1")
.saveAsTable("/DBTBL1")

• ☐

.bucketBy


("GeographyRegionID")

.saveAsTable("/DBTBL1")

- 

```

.sortBy
("GeographyRegionID", "Year", "Month", "Day")
.json("/DBTBL1")

```
- 

```

.partitionBy
("Year", "Month", "Day", "GeographyRegionID")
.parquet("/DBTBL1")

```

(Poprawne)

Wyjaśnienie

Correct Answer:

Box 1: `.partitionBy` -

Incorrect Answers:

- `.format`:

Method: `format()`:

Arguments: "parquet", "csv", "txt", "json", "jdbc", "orc", "avro", etc.

- `.bucketBy`:

Method: `bucketBy()`

Arguments: (numBuckets, col, col..., coln)

The number of buckets and names of columns to bucket by. Uses Hive's bucketing scheme on a filesystem.

Box 2: ("Year", "Month", "Day", "GeographyRegionID")

Specify the columns on which to do the partition. Use the date columns followed by the GeographyRegionID column.

Box 3: `.parquet("/DBTBL1")`

Reference:

<https://www.oreilly.com/library/view/learning-spark-2nd/9781492050032/ch04.html>

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

Pytanie 4:

Pominięto

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 security policy**
(Poprawne)
- ☐ **a custom role-based access control (RBAC) role**
- ☐ **a function**
(Poprawne)
- ☐ **a column encryption key**
- ☐ **asymmetric keys**

Wyjaśnienie

Correct Answer: AC

A: Row-Level Security (RLS) 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.

C: a function - predicates created as inline table-valued functions

Reference:

<https://docs.microsoft.com/en-us/sql/relational-databases/security/row-level-security>

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

Pytanie 5:

Pominięto

You have a SQL pool in Azure Synapse that contains a table named `dbo.Customers`. The table contains a column name `Email`.

You need to prevent nonadministrative users from seeing the full email addresses in the `Email` column. The users must see values in a format of a

`XXX@XXXX.com` instead.

What should you do?

- ☐
From Microsoft SQL Server Management Studio, set an email mask on the `Email` column.
(Poprawne)
- ☐
From the Azure portal, set a mask on the `Email` column.
- ☐
From Microsoft SQL Server Management Studio, grant the `SELECT` permission to the users for all the columns in the `dbo.Customers` table except `Email`.
- ☐
From the Azure portal, set a sensitivity classification of `Confidential` for the `Email` column.

Wyjaśnienie

Correct Answer: A

The Email masking method, which exposes the first letter and replaces the domain with `XXX.com` using a constant string prefix in the form of an email address. `aXX@XXXX.com`

Reference:

<https://docs.microsoft.com/en-us/azure/azure-sql/database/dynamic-data-masking-overview>

Pytanie 6:

Pominięto

You are designing an Azure Synapse solution that will provide a query interface for the data stored in an Azure Storage account. The storage account is only accessible from a virtual network.

You need to recommend an authentication mechanism to ensure that the solution can access the source data.

What should you recommend?

☐

a managed identity

(Poprawne)

☐

anonymous public read access

☐

a shared key

Wyjaśnienie

Correct Answer: A

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

Reference:

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

Pytanie 7:

Pominięto

You are developing an application that uses Azure Data Lake Storage Gen2.

You need to recommend a solution to grant permissions to a specific application for a limited time period.

What should you include in the recommendation?

- ☐ **role assignments**
- ☐ **shared access signatures (SAS)**
(Poprawne)
- ☐ **Azure Active Directory (Azure AD) identities**
- ☐ **account keys**

Wyjaśnienie

Correct Answer: B

A shared access signature (SAS) provides secure delegated access to resources in your storage account. With a SAS, you have granular control over how a client can access your data. For example:

What resources the client may access.

What permissions they have to those resources.

How long the SAS is valid.

Reference:

<https://docs.microsoft.com/en-us/azure/storage/common/storage-sas-overview>

Pytanie 8:

Pominięto

You use Azure Data Lake Storage Gen2 to store data that data scientists and data engineers will query by using Azure Databricks interactive notebooks. Users will have access only to the Data Lake Storage folders that relate to the projects on which they work.

You need to recommend which authentication methods to use for Databricks and Data Lake Storage to provide the users with the appropriate access. The solution must minimize administrative effort and development effort.

Which authentication method should you recommend for each Azure service? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

Databricks:

	▼
Azure Active Directory credential passthrough	
Azure Key Vault secrets	
Personal access tokens	

Data Lake Storage:

	▼
Azure Active Directory credential passthrough	
Shared access keys	
Shared access signatures	

☐

Azure Active Directory credential passthrough

Shared access signatures

(Poprawne)

☐

Azure Key Vault secrets

Shared access keys

☐

Personal access tokens

Azure Active Directory credential passthrough

Wyjaśnienie

Correct Answer: A

Box1: Azure Active Directory credential passthrough

Access ADLS Gen2 from Databricks by running query interactively from notebooks.

Box2: 'Shared access signatures'

Users also need directly access to the Data Lake Storage for specific folders.

Accessing the ADLS via Databricks should be using Azure Active Directory with Passthrough. Accessing the files in ADLS should be SAS, based on the options provided.

Pytanie 9:

Pominięto

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?

- ☐ table partitions
- ☐ a default value
- ☐ row-level security (RLS)
- ☐ column encryption
- ☐ dynamic data masking

(Poprawne)

Wyjaśnienie

Correct Answer: E

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>

Pytanie 10:

Pominięto

A company purchases IoT devices to monitor manufacturing machinery. The company uses an Azure IoT Hub to communicate with the IoT devices.

The company must be able to monitor the devices in real-time.

You need to design the solution.

What should you recommend?

☐

Azure Data Factory instance using Azure Portal

☐

Azure Data Factory instance using Azure PowerShell

☐

Azure Stream Analytics cloud job using Azure Portal

(Poprawne)

☐

Azure Data Factory instance using Microsoft Visual Studio

Wyjaśnienie

Correct Answer: C

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. Your Stream Analytics job would ingest these events from Event Hubs and run real-time analytics queries against the streams.

Create a Stream Analytics job:

In the Azure portal, select + Create a resource from the left navigation menu. Then, select Stream Analytics job from Analytics.

Reference:

<https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-get-started-with-azure-stream-analytics-to-process-data-from-iot-devices>

Pytanie 11:

Pominięto

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.

Hot Area:

Answer Area

Number of partitions:

	▼
1	
8	
16	
32	

Partition key:

	▼
Fraud indicator	
Fraud score	
Individual line items	
Payment details	
Transaction ID	

• ☐

1

Fraud indicator

• ☐

8

Fraud score

• ☐

16

Transaction ID

(Poprawne)

• ☐

32

Payment details

Wyjaśnienie

Correct Answer:

Answer Area

Number of partitions:

	▼
1	
8	
16	
32	

Partition key:

	▼
Fraud indicator	
Fraud score	
Individual line items	
Payment details	
Transaction ID	

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>

Pytanie 12:

Pominięto

You have an on-premises data warehouse that includes the following fact tables. Both tables have the following columns: DateKey, ProductKey, RegionKey. There are 120 unique product keys and 65 unique region keys.

Table	Comments
Sales	The table is 600 GB in size. DateKey is used extensively in the WHERE clause in queries. ProductKey is used extensively in join operations. RegionKey is used for grouping. Severity-five percent of records relate to one of 40 regions.
Invoice	The table is 6 GB in size. DateKey and ProductKey are used extensively in the WHERE clause in queries. RegionKey is used for grouping.

Queries that use the data warehouse take a long time to complete.

You plan to migrate the solution to use Azure Synapse Analytics. You need to ensure that the Azure-based solution optimizes query performance and minimizes processing skew.

What should you recommend? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point

\

Hot Area:

Answer Area

Table	Distribution type	Distribution column
Sales:	<div><div></div><div>Hash-distributed</div><div>Round-robin</div></div>	<div><div></div><div>DateKey</div><div>ProductKey</div><div>RegionKey</div></div>
Invoices:	<div><div></div><div>Hash-distributed</div><div>Round-robin</div></div>	<div><div></div><div>DateKey</div><div>ProductKey</div><div>RegionKey</div></div>

• ☐

Hash-distributed

ProductKey

Hash-distributed

RegionKey

(Poprawne)

• ☐

Hash-distributed

RegionKey

Round-robin

DateKey

• ☐

Round-robin

DateKey

Round-robin

RegionKey

• ☐

Round-robin

ProductKey

Hash-distributed

RegionKey

Wyjaśnienie

Correct Answer:

Answer Area

Table	Distribution type	Distribution column
Sales:	<div><div>▼</div><div>Hash-distributed</div><div>Round-robin</div></div>	<div><div>▼</div><div>DateKey</div><div>ProductKey</div><div>RegionKey</div></div>
Invoices:	<div><div>▼</div><div>Hash-distributed</div><div>Round-robin</div></div>	<div><div>▼</div><div>DateKey</div><div>ProductKey</div><div>RegionKey</div></div>

Box 1: Hash-distributed -

Box 2: ProductKey -

ProductKey is used extensively in joins.

Hash-distributed tables improve query performance on large fact tables.

Box 3: Hash-distributed -

Box 4: RegionKey -

"The table size on disk is more than 2 GB." and you have to choose a distribution column which: "Is not used in WHERE clauses. This could narrow the query to not run on all the distributions."

Reference:

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

Pytanie 13:

Pominięto

You have a partitioned table in an Azure Synapse Analytics dedicated SQL pool.

You need to design queries to maximize the benefits of partition elimination.

What should you include in the Transact-SQL queries?

- ☐ JOIN
- ☐ WHERE
(Poprawne)
- ☐ DISTINCT
- ☐ GROUP BY

Wyjaśnienie

Correct Answer: B

Data partition elimination refers to the database server's ability to determine, based on query predicates

When you add the "WHERE" clause to your T-SQL query it allows the query optimizer accesses only the relevant partitions to satisfy the filter criteria of the query - which is what partition elimination is all about.

<https://stackoverflow.com/questions/51677471/what-is-a-difference-between-table-distribution-and-table-partition-in-sql/51677595>

Pytanie 14:**Pominięto**

You implement an enterprise data warehouse in Azure Synapse Analytics.

You have a large fact table that is 10 terabytes (TB) in size.

Incoming queries use the primary key SaleKey column to retrieve data as displayed in the following table:

SaleKey	CityKey	CustomerKey	StockItemKey	InvoiceDateKey	Quantity	UnitPrice	TotalExcludingTax
49309	90858	70	69	10/22/13	8	16	128
49313	55710	126	69	10/22/13	2	16	32
49343	44710	234	68	10/22/13	10	16	160
49352	66109	163	70	10/22/13	4	16	64
49488	65312	230	70	10/22/13	8	16	128
49646	85877	271	70	10/24/13	1	16	16
49798	41238	288	69	10/24/13	1	16	16

You need to distribute the large fact table across multiple nodes to optimize performance of the table.

Which technology should you use?

☐

hash distributed table with clustered index

☐

hash distributed table with clustered Columnstore index

(Poprawne)

☐

round robin distributed table with clustered index

☐

round robin distributed table with clustered Columnstore index

☐

heap table with distribution replicate

Wyjaśnienie

Correct Answer: B

Hash-distributed tables improve query performance on large fact tables.

Columnstore indexes can achieve up to 100x better performance on analytics and data warehousing workloads and up to 10x better data compression than traditional rowstore indexes.

Incorrect Answers:

C, D: Round-robin tables are useful for improving loading speed.

Reference:

<https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-tables-distribute> <https://docs.microsoft.com/en-us/sql/relational-databases/indexes/columnstore-indexes-query-performance>

Pytanie 15:

Pominięto

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?

- ☐ **nonclustered columnstore**
- ☐ **clustered columnstore**
(Poprawne)
- ☐ **nonclustered**
- ☐ **clustered**

Wyjaśnienie

Correct Answer: B

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>

Pytanie 16:

Pominięto

You create an Azure Databricks cluster and specify an additional library to install.

When you attempt to load the library to a notebook, the library is not found.

You need to identify the cause of the issue.

What should you review?

☐

notebook logs

☐

cluster event logs

(Poprawne)

☐

global init scripts logs

☐

workspace logs

Wyjaśnienie

Correct Answer: B

Azure Databricks provides three kinds of logging of cluster-related activity:

Cluster event logs, which capture cluster lifecycle events, like creation, termination, configuration edits, and so on.

Apache Spark driver and worker logs, which you can use for debugging.

Cluster init-script logs, valuable for debugging init scripts.

Reference:

<https://docs.microsoft.com/en-us/azure/databricks/clusters/clusters-manage#event-log>

Pytanie 17:

Pominięto

You have an Azure data factory.

You need to examine the pipeline failures from the last 60 days.

What should you use?

- ☐ the Activity log blade for the Data Factory resource
- ☐ the Monitor & Manage app in Data Factory
- ☐ the Resource health blade for the Data Factory resource
- ☐ Azure Monitor

(Poprawne)

Wyjaśnienie

Correct Answer: D

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>

Pytanie 18:

Pominięto

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?

☐

Drop late arriving events from the job.

☐

Add an Azure Storage account to the job.

☐

Increase the streaming units for the job.

(Poprawne)

☐

Stop the job.

Wyjaśnienie

Correct Answer: C

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>

Pytanie 19:

Pominięto

You are designing an Azure Databricks interactive cluster. The cluster will be used infrequently and will be configured for auto-termination.

You need to use that the cluster configuration is retained indefinitely after the cluster is terminated. The solution must minimize costs.

What should you do?

- ☐ **Pin the cluster.**
(Poprawne)
- ☐ **Create an Azure runbook that starts the cluster every 90 days.**
- ☐ **Terminate the cluster manually when processing completes.**
- ☐ **Clone the cluster after it is terminated.**

Wyjaśnienie

Correct Answer: A

30 days after a cluster is terminated, it is permanently deleted. To keep an interactive cluster configuration even after a cluster has been terminated for more than 30 days, an administrator can pin the cluster. Up to 20 clusters can be pinned.

Reference:

<https://docs.databricks.com/clusters/clusters-manage.html#pin-a-cluster>

<https://docs.azuredatabricks.net/user-guide/clusters/terminate.html>

Pytanie 20:

Pominięto

You have an Azure data solution that contains an enterprise data warehouse in Azure Synapse Analytics named DW1.

Several users execute ad hoc 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?

☐

Hash distribute the large fact tables in DW1 before performing the automated data loads.

☐

Assign a smaller resource class to the automated data load queries.

☐

Assign a larger resource class to the automated data load queries.

(Poprawne)

☐

Create sampled statistics for every column in each table of DW1.

Wyjaśnienie

Correct Answer: C

The performance capacity of a query is determined by the user's resource class. Resource classes are pre-determined resource limits in Synapse SQL pool that govern compute resources and concurrency for query execution.

Resource classes can help you configure resources for your queries by setting limits on the number of queries that run concurrently and on the compute-resources assigned to each query. There's a trade-off between memory and concurrency. 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>

Pytanie 21:

Pominięto

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?

- ☐
Connect to the built-in pool and run DBCC PDW_SHOWSPACEUSED.
- ☐
Connect to the built-in pool and run DBCC CHECKALLOC.
- ☐
Connect to Pool1 and query sys.dm_pdw_node_status.
- ☐
Connect to Pool1 and query sys.dm_pdw_nodes_db_partition_stats.

(Poprawne)

Wyjaśnienie

Correct Answer: D

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>

Pytanie 22:

Pominięto

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.

Hot Area:

Answer Area

Library:

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

Workspace:

	▼
Azure Databricks	
Azure Log Analytics	
Azure Machine Learning	

☐

Azure Databricks Monitoring Library

Azure Log Analytics

(Poprawne)

☐

PyTorch

Azure Machine Learning

☐

Microsoft Azure Management Monitoring Library

Azure Databricks



TensorFlow

Azure Log Analytics

Wyjaśnienie

Correct Answer:

Answer Area

Library:

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

Workspace:

	▼
Azure Databricks	
Azure Log Analytics	
Azure Machine Learning	

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.

Reference:

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

Pytanie 23:

Pominięto

You have a SQL pool in Azure Synapse.

You discover that some queries fail or take a long time to complete.

You need to monitor for transactions that have rolled back.

Which dynamic management view should you query?

- ☐ **sys.dm_pdw_request_steps**
- ☐ **sys.dm_pdw_nodes_tran_database_transactions**
(Poprawne)
- ☐ **sys.dm_pdw_waits**
- ☐ **sys.dm_pdw_exec_sessions**

Wyjaśnienie

Correct Answer: B

You can use Dynamic Management Views (DMVs) to monitor your workload including investigating query execution in SQL pool.

If your queries are failing or taking a long time to proceed, you can check and monitor if you have any transactions rolling back.

Example:

-- Monitor rollback

```
SELECT -  
SUM(CASE WHEN t.database_transaction_next_undo_lsn IS NOT NULL THEN 1 ELSE  
0 END), t.pdw_node_id, nod.[type]  
FROM sys.dm_pdw_nodes_tran_database_transactions t  
JOIN sys.dm_pdw_nodes nod ON t.pdw_node_id = nod.pdw_node_id  
GROUP BY t.pdw_node_id, nod.[type]
```

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-manage-monitor#monitor-transaction-log-rollback>

Pytanie 24:

Pominięto

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.

What should you do?

☐

Change the compatibility level of the Stream Analytics job.

☐

Increase the number of streaming units (SUs).

(Poprawne)

☐

Remove any named consumer groups from the connection and use \$default.

☐

Create an additional output stream for the existing input stream.

Wyjaśnienie

Correct Answer: B

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. You should increase the Streaming Units.

Note: 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.

Reference:

<https://docs.microsoft.com/bs-cyrl-ba/azure/stream-analytics/stream-analytics-monitoring>

Pytanie 25:**Pominięto**

You are designing an inventory updates table in an Azure Synapse Analytics dedicated SQL pool. The table will have a clustered columnstore index and will include the following columns:

Table	Comment
EventDate	One million records are added to the table each day
EventTypeID	The table contains 10 million records for each event type.
WarehouseID	The table contains 100 million records for each warehouse.
ProductCategoryTypeID	The table contains 25 million records for each product category type.

You identify the following usage patterns:

- Analysts will most commonly analyze transactions for a warehouse.
- Queries will summarize by product category type, date, and/or inventory event type.

You need to recommend a partition strategy for the table to minimize query times.

On which column should you partition the table?

- ☐ **EventTypeID**
- ☐ **ProductCategoryTypeID**
- ☐ **EventDate**
- ☐ **WarehouseID**

(Poprawne)

Wyjaśnienie

Correct Answer: D

The number of records for each warehouse is big enough for a good partitioning. Note: Table partitions enable you to divide your data into smaller groups of data. In most cases, table partitions are created on a date column. 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 distributed databases.

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

Pytanie 26:

Pominięto

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.

- ☐ **Create a date dimension table that has a DateTime key.**
- ☐ **Use built-in SQL functions to extract date attributes.**
- ☒ **Create a date dimension table that has an integer key in the format of YYYYMMDD.**
(Poprawne)
- ☐ **In the fact table, use integer columns for the date fields**
(Poprawne)
- ☐ **Use DateTime columns for the date fields.**

Wyjaśnienie

Correct Answer: *CD*

You need a dimension table for the date

1) Create a date dimension table that has an integer key in the format of YYYYMMDD.

2) In the fact table, use integer columns for the date fields

Pytanie 27:

Pominięto

A company purchases IoT devices to monitor manufacturing machinery. The company uses an Azure IoT Hub to communicate with the IoT devices.

The company must be able to monitor the devices in real-time.

You need to design the solution.

What should you recommend?

- ☐ **Azure Analysis Services using Azure Portal**
- ☐ **Azure Analysis Services using Azure PowerShell**
- ☐ **Azure Stream Analytics cloud job using Azure Portal**
- ☒ **Azure Data Factory instance using Microsoft Visual Studio**

(Poprawne)

Wyjaśnienie

Correct Answer: C

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. Your Stream Analytics job would ingest these events from Event Hubs and run real-time analytics queries against the streams.

Create a Stream Analytics job:

In the Azure portal, select + Create a resource from the left navigation menu. Then, select Stream Analytics job from Analytics.

Reference:

<https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-get-started-with-azure-stream-analytics-to-process-data-from-iot-devices>

Pytanie 28:

Pominięto

You have a SQL pool in Azure Synapse.

A user reports that queries against the pool take longer than expected to complete. You determine that the issue relates to queried columnstore segments.

You need to add monitoring to the underlying storage to help diagnose the issue.

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

NOTE: Each correct selection is worth one point.

- ☐ **Snapshot Storage Size**
- ☐ **Cache used percentage**
(Poprawne)
- ☐ **DWU Limit**
- ☐ **Cache hit percentage**
(Poprawne)

Wyjaśnienie

Correct Answer: BD

D: Cache hit percentage: $(\text{cache hits} / (\text{cache hits} + \text{cache miss})) * 100$, where cache hits are the sum of all columnstore segments hits in the local SSD cache and cache miss is the columnstore segments misses in the local SSD cache summed across all nodes

B: $(\text{cache used} / \text{cache capacity}) * 100$ where cache used is the sum of all bytes in the local SSD cache across all nodes and cache capacity is the sum of the storage capacity of the local SSD cache across all nodes

Incorrect Asnwers:

C: DWU limit: Service level objective of the data warehouse.

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-concept-resource-utilization-query-activity>

Pytanie 29:

Pominięto

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?

- ☐ clusters
- (Poprawne)**
- ☐ workspace
- ☐ DBFS
- ☐ SSH
- ☐ jobs

Wyjaśnienie

Correct Answer: A

Clusters, workspace logs does not have any cluster related resource change.

<https://docs.databricks.com/administration-guide/account-settings/audit-logs.html#workspace-level-audit-log-events>

Pytanie 30:

Pimiento

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?

☐

5xx: Server Error errors

☐

Average Success E2E Latency

☐

availability

☐

Last Sync Time

(Poprawne)

Wyjaśnienie

Correct Answer: D

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>

<https://docs.microsoft.com/en-us/azure/storage/common/storage-redundancy?toc=/azure/storage/blobs/toc.json#check-the-last-sync-time-property>

Pytanie 31:

Pominięto

You configure monitoring from an Azure Synapse Analytics implementation. The implementation uses PolyBase to load data from comma-separated value (CSV) files stored in Azure Data Lake Storage Gen2 using an external table.

Files with an invalid schema cause errors to occur.

You need to monitor for an invalid schema error.

For which error should you monitor?

- ☐ **EXTERNAL TABLE access failed due to internal error: 'Java exception raised on call to HdfsBridge_Connect: Error [com.microsoft.polybase.client.KerberosSecureLogin] occurred while accessing external file.'**
- ☐ **Cannot execute the query "Remote Query" against OLE DB provider "SQLNCLI11" for linked server "(null)". Query aborted- the maximum reject threshold (0 rows) was reached while reading from an external source: 1 rows rejected out of total 1 rows processed.**
- ☒ **(Poprawne)**
- ☐ **EXTERNAL TABLE access failed due to internal error: 'Java exception raised on call to HdfsBridge_Connect: Error [Unable to instantiate LoginClass] occurred while accessing external file.'**
- ☐ **EXTERNAL TABLE access failed due to internal error: 'Java exception raised on call to HdfsBridge_Connect: Error [No FileSystem for scheme: wasbs] occurred while accessing external file.'**

Wyjaśnienie

Suggested Answer: B

Error message: Cannot execute the query "Remote Query"

Possible Reason:

The reason this error happens is because each file has different schema. The PolyBase external table DDL when pointed to a directory recursively reads all the files

in that directory. When a column or data type mismatch happens, this error could be seen in SSMS.

Reference:

<https://docs.microsoft.com/en-us/sql/relational-databases/polybase/polybase-errors-and-possible-solutions>

<https://techcommunity.microsoft.com/t5/datacat/polybase-setup-errors-and-possible-solutions/ba-p/305297>

Pytanie 32:**Pominięto**

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?

☐

All distributions data.

☐

The table contains less than 10,000 rows.

☐

The table uses round-robin distribution.

☐

The table is skewed.

(Poprawne)

Wyjaśnienie

Correct Answer: D

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-distribute>

Pytanie 33:**Pominięto**

You have two fact tables named Flight and Weather. Queries targeting the tables will be based on the join between the following columns.

Table	Column
Flight	ArrivalAirportID ArrivalDateTime
Weather	AirportID ReportDateTime

You need to recommend a solution that maximizes query performance. What should you include in the recommendation?

- ☐
In the tables use a hash distribution of ArrivalDateTime and ReportDateTime.
- ☐
In the tables use a hash distribution of ArrivalAirportID and AirportID.
(Poprawne)
- ☐
In each table, create an IDENTITY column.
- ☐
In each table, create a column as a composite of the other two columns in the table.

Wyjaśnienie

Correct Answer: B

Hash-distribution improves query performance on large fact tables.

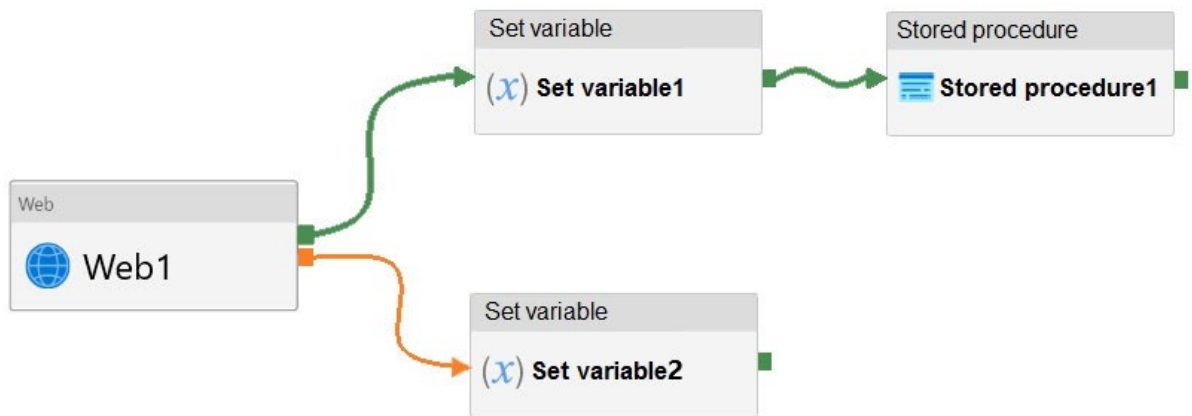
Incorrect Answers:

A: Do not use a date column for hash distribution. All data for the same date lands in the same distribution. If several users are all filtering on the same date, then only 1 of the 60 distributions do all the processing work.

Pytanie 34:

Pominięto

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.

Hot Area:

Answer Area

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	

☐ ☐

fail

Canceled

☐ ☐

succeed

Failed

(Poprawne)

☐ ☐

fail

Failed



succeed

Succeeded

Wyjaśnienie

Correct Answer:

Answer Area

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]

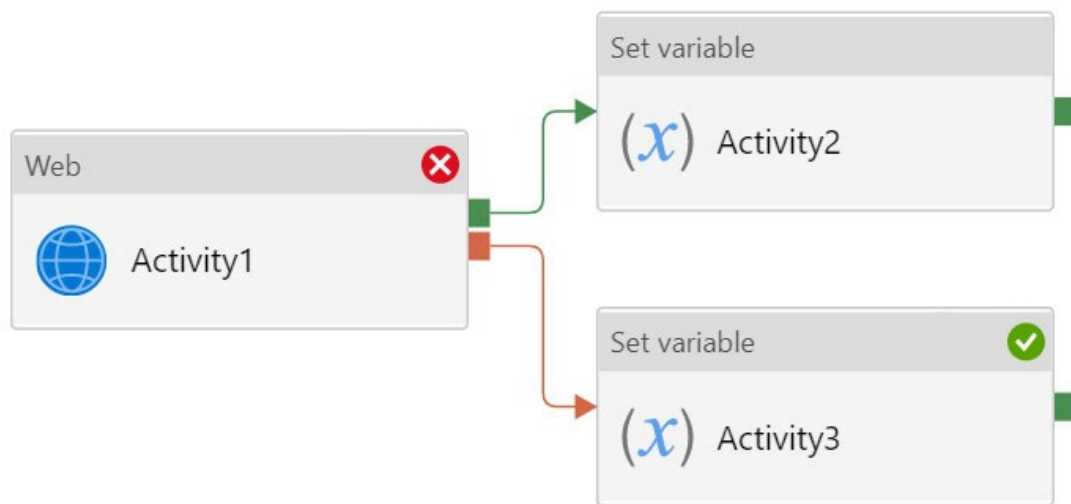
	▼
Canceled	
Failed	
Succeeded	

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/>

Pytanie 35:

Pominięto

You have several Azure Data Factory pipelines that contain a mix of the following types of activities:

- Wrangling data flow
- Notebook
- Copy
- Jar

Which two Azure services should you use to debug the activities? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point

- ☐
Azure Synapse Analytics
- ☐
Azure HDInsight
- ☐
Azure Machine Learning
- ☐
Azure Data Factory
(Poprawne)
- ☐
Azure Databricks
(Poprawne)

Wyjaśnienie

Correct Answer: DE

Notebook- azure databricks, managing activities in pipeline-datafactory

<https://docs.microsoft.com/en-us/azure/data-factory/wrangling-overview>

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

Pytanie 36:

Pominięto

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?

• ☐

Connect to the built-in pool and query sys.dm_pdw_sys_info.

• ☐

Connect to Pool1 and run DBCC CHECKALLOC.

• ☐

Connect to the built-in pool and run DBCC CHECKALLOC.

• ☐

Connect to Pool1 and query sys.dm_pdw_nodes_db_partition_stats.

(Poprawne)

Wyjaśnienie

Correct Answer: D

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>

Pytanie 37:

Pominięto

Case study -

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

To start the case study -

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an

All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

Overview -

Contoso, Ltd. is a clothing retailer based in Seattle. The company has 2,000 retail stores across the United States and an emerging online presence.

The network contains an Active Directory forest named contoso.com. The forest is integrated with an Azure Active Directory (Azure AD) tenant named contoso.com. Contoso has an Azure subscription associated to the contoso.com Azure AD tenant.

Existing Environment -

Transactional Data -

Contoso has three years of customer, transactional, operational, sourcing, and supplier data comprised of 10 billion records stored across multiple on-premises

Microsoft SQL Server servers. The SQL Server instances contain data from various operational systems. The data is loaded into the instances by using SQL

Server Integration Services (SSIS) packages.

You estimate that combining all product sales transactions into a company-wide sales transactions dataset will result in a single table that contains 5 billion rows, with one row per transaction.

Most queries targeting the sales transactions data will be used to identify which products were sold in retail stores and which products were sold online during different time periods. Sales transaction data that is older than three years will be removed monthly.

You plan to create a retail store table that will contain the address of each retail store. The table will be approximately 2 MB. Queries for retail store sales will include the retail store addresses.

You plan to create a promotional table that will contain a promotion ID. The promotion ID will be associated to a specific product. The product will be identified by a product ID. The table will be approximately 5 GB.

Streaming Twitter Data -

The ecommerce department at Contoso develops an Azure logic app that captures trending Twitter feeds referencing the company's products and pushes the products to Azure Event Hubs.

Planned Changes and Requirements

Planned Changes -

Contoso plans to implement the following changes:

Load the sales transaction dataset to Azure Synapse Analytics.

Integrate on-premises data stores with Azure Synapse Analytics by using SSIS packages.

Use Azure Synapse Analytics to analyze Twitter feeds to assess customer sentiments about products.

Sales Transaction Dataset Requirements

Contoso identifies the following requirements for the sales transaction dataset:

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.

Ensure that queries joining and filtering sales transaction records based on product ID complete as quickly as possible.

Implement a surrogate key to account for changes to the retail store addresses.

Ensure that data storage costs and performance are predictable.

Minimize how long it takes to remove old records.

Customer Sentiment Analytics Requirements

Contoso identifies the following requirements for customer sentiment analytics:

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.

Maximize the throughput of ingesting Twitter feeds from Event Hubs to Azure Storage without purchasing additional throughput or capacity units.

Store Twitter feeds in Azure Storage by using Event Hubs Capture. The feeds will be converted into Parquet files.

Ensure that the data store supports Azure AD-based access control down to the object level.

Minimize administrative effort to maintain the Twitter feed data records.

Purge Twitter feed data records that are older than two years.

Data Integration Requirements -

Contoso identifies the following requirements for data integration:

Use an Azure service that leverages the existing SSIS packages to ingest on-premises data into datasets stored in a dedicated SQL pool of Azure Synapse

Analytics and transform the data.

Identify a process to ensure that changes to the ingestion and transformation activities can be version-controlled and developed independently by multiple data engineers.

QUESTION:

You need to ensure that the Twitter feed data can be analyzed in the dedicated SQL pool. The solution must meet the customer sentiment analytic requirements. Which three Transact-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.

Select and Place:

Commands

CREATE EXTERNAL DATA SOURCE

CREATE EXTERNAL FILE FORMAT

CREATE EXTERNAL TABLE

CREATE EXTERNAL TABLE AS SELECT

CREATE DATABASE SCOPED CREDENTIAL

Answer Area

- ☐ **CREATE EXTERNAL DATA SOURCE**
- ☐ **CREATE EXTERNAL FILE FORMAT**
- ☐ **CREATE EXTERNAL TABLE AS SELECT**

(Poprawne)

- ☐ **CREATE EXTERNAL DATA SOURCE**
- ☐ **CREATE EXTERNAL FILE FORMAT**

CREATE DATABASE SCOPED CREDENTIAL



CREATE EXTERNAL TABLE

CREATE DATABASE SCOPED CREDENTIAL

CREATE EXTERNAL DATA SOURCE

Wyjaśnienie

Correct Answer: A

Commands

CREATE EXTERNAL DATA SOURCE
CREATE EXTERNAL FILE FORMAT
CREATE EXTERNAL TABLE
CREATE EXTERNAL TABLE AS SELECT
CREATE DATABASE SCOPED CREDENTIAL

Answer Area

CREATE EXTERNAL DATA SOURCE
CREATE EXTERNAL FILE FORMAT
CREATE EXTERNAL TABLE AS SELECT

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.

Incorrect Answers:

CREATE EXTERNAL TABLE -

The CREATE EXTERNAL TABLE command creates an external table for Synapse SQL to access data stored in Azure Blob Storage or Azure Data Lake Storage.

Reference:

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

<https://docs.microsoft.com/en-us/sql/t-sql/statements/create-external-table-transact-sql?view=azure-sqldw-latest&tabs=dedicated>

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/best-practices-dedicated-sql-pool#use-polybase-to-load-and-export-data-quickly>

Reference: <https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/develop-tables-external-tables?tabs=hadoop#credential>

Pytanie 38:

Pominięto

Case study -

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

To start the case study -

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an

All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

Overview -

Contoso, Ltd. is a clothing retailer based in Seattle. The company has 2,000 retail stores across the United States and an emerging online presence.

The network contains an Active Directory forest named contoso.com. The forest is integrated with an Azure Active Directory (Azure AD) tenant named contoso.com. Contoso has an Azure subscription associated to the contoso.com Azure AD tenant.

Existing Environment -

Transactional Data -

Contoso has three years of customer, transactional, operational, sourcing, and supplier data comprised of 10 billion records stored across multiple on-premises

Microsoft SQL Server servers. The SQL Server instances contain data from various operational systems. The data is loaded into the instances by using SQL

Server Integration Services (SSIS) packages.

You estimate that combining all product sales transactions into a company-wide sales transactions dataset will result in a single table that contains 5 billion rows, with one row per transaction.

Most queries targeting the sales transactions data will be used to identify which products were sold in retail stores and which products were sold online during different time periods. Sales transaction data that is older than three years will be removed monthly.

You plan to create a retail store table that will contain the address of each retail store. The table will be approximately 2 MB. Queries for retail store sales will include the retail store addresses.

You plan to create a promotional table that will contain a promotion ID. The promotion ID will be associated to a specific product. The product will be identified by a product ID. The table will be approximately 5 GB.

Streaming Twitter Data -

The ecommerce department at Contoso develops an Azure logic app that captures trending Twitter feeds referencing the company's products and pushes the products to Azure Event Hubs.

Planned Changes and Requirements

Planned Changes -

Contoso plans to implement the following changes:

Load the sales transaction dataset to Azure Synapse Analytics.

Integrate on-premises data stores with Azure Synapse Analytics by using SSIS packages.

Use Azure Synapse Analytics to analyze Twitter feeds to assess customer sentiments about products.

Sales Transaction Dataset Requirements

Contoso identifies the following requirements for the sales transaction dataset:

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.

Ensure that queries joining and filtering sales transaction records based on product ID complete as quickly as possible.

Implement a surrogate key to account for changes to the retail store addresses.

Ensure that data storage costs and performance are predictable.

Minimize how long it takes to remove old records.

Customer Sentiment Analytics Requirements

Contoso identifies the following requirements for customer sentiment analytics:

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.

Maximize the throughput of ingesting Twitter feeds from Event Hubs to Azure Storage without purchasing additional throughput or capacity units.

Store Twitter feeds in Azure Storage by using Event Hubs Capture. The feeds will be converted into Parquet files.

Ensure that the data store supports Azure AD-based access control down to the object level.

Minimize administrative effort to maintain the Twitter feed data records.

Purge Twitter feed data records that are older than two years.

Data Integration Requirements -

Contoso identifies the following requirements for data integration:

Use an Azure service that leverages the existing SSIS packages to ingest on-premises data into datasets stored in a dedicated SQL pool of Azure Synapse

Analytics and transform the data.

Identify a process to ensure that changes to the ingestion and transformation activities can be version-controlled and developed independently by multiple data engineers.

QUESTION:

You need to design the partitions 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.

Hot Area:

Answer Area

Partition product sales transactions data by:

	▼
Sales date	
Product ID	
Promotion ID	

Store product sales transactions data in:

	▼
An Azure Synapse Analytics dedicated SQL pool	
An Azure Synapse Analytics serverless SQL pool	
An Azure Data Lake Storage Gen2 account linked to an Azure Synapse Analytics workspace	



Sales date

An Azure Synapse Analytics dedicated SQL pool

(Poprawne)



Product ID

to an Azure Synapse Analytics workspace

• ☐

Promotion ID

An Azure Data Lake Storage Gen2 account linked

• ☐

Sales date

An Azure Synapse Analytics serverless SQL pool

Wyjaśnienie

Correct Answer:

Answer Area

Partition product sales
transactions data by:

	▼
Sales date	
Product ID	
Promotion ID	

Store product sales
transactions data in:

	▼
An Azure Synapse Analytics dedicated SQL pool	
An Azure Synapse Analytics serverless SQL pool	
An Azure Data Lake Storage Gen2 account linked to an Azure Synapse Analytics workspace	

Box 1: Sales date -

Scenario: Contoso requirements for data integration include:

- 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.

Box 2: An Azure Synapse Analytics Dedicated SQL pool

Scenario: Contoso requirements for data integration include:

- Ensure that data storage costs and performance are predictable.

The size of a dedicated SQL pool (formerly SQL DW) is determined by Data Warehousing Units (DWU).

Dedicated SQL pool (formerly SQL DW) stores data in relational tables with columnar storage. This format significantly reduces the data storage costs, and improves

query performance.

Synapse analytics dedicated sql pool

Reference:

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

Pytanie 39:

Pominięto

Case study -

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

To start the case study -

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an

All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

Overview -

Contoso, Ltd. is a clothing retailer based in Seattle. The company has 2,000 retail stores across the United States and an emerging online presence.

The network contains an Active Directory forest named contoso.com. The forest is integrated with an Azure Active Directory (Azure AD) tenant named contoso.com. Contoso has an Azure subscription associated to the contoso.com Azure AD tenant.

Existing Environment -

Transactional Data -

Contoso has three years of customer, transactional, operational, sourcing, and supplier data comprised of 10 billion records stored across multiple on-premises

Microsoft SQL Server servers. The SQL Server instances contain data from various operational systems. The data is loaded into the instances by using SQL

Server Integration Services (SSIS) packages.

You estimate that combining all product sales transactions into a company-wide sales transactions dataset will result in a single table that contains 5 billion rows, with one row per transaction.

Most queries targeting the sales transactions data will be used to identify which products were sold in retail stores and which products were sold online during different time periods. Sales transaction data that is older than three years will be removed monthly.

You plan to create a retail store table that will contain the address of each retail store. The table will be approximately 2 MB. Queries for retail store sales will include the retail store addresses.

You plan to create a promotional table that will contain a promotion ID. The promotion ID will be associated to a specific product. The product will be identified by a product ID. The table will be approximately 5 GB.

Streaming Twitter Data -

The ecommerce department at Contoso develops an Azure logic app that captures trending Twitter feeds referencing the company's products and pushes the products to Azure Event Hubs.

Planned Changes and Requirements

Planned Changes -

Contoso plans to implement the following changes:

Load the sales transaction dataset to Azure Synapse Analytics.

Integrate on-premises data stores with Azure Synapse Analytics by using SSIS packages.

Use Azure Synapse Analytics to analyze Twitter feeds to assess customer sentiments about products.

Sales Transaction Dataset Requirements

Contoso identifies the following requirements for the sales transaction dataset:

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.

Ensure that queries joining and filtering sales transaction records based on product ID complete as quickly as possible.

Implement a surrogate key to account for changes to the retail store addresses.

Ensure that data storage costs and performance are predictable.

Minimize how long it takes to remove old records.

Customer Sentiment Analytics Requirements

Contoso identifies the following requirements for customer sentiment analytics:

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.

Maximize the throughput of ingesting Twitter feeds from Event Hubs to Azure Storage without purchasing additional throughput or capacity units.

Store Twitter feeds in Azure Storage by using Event Hubs Capture. The feeds will be converted into Parquet files.

Ensure that the data store supports Azure AD-based access control down to the object level.

Minimize administrative effort to maintain the Twitter feed data records.

Purge Twitter feed data records that are older than two years.

Data Integration Requirements -

Contoso identifies the following requirements for data integration:

Use an Azure service that leverages the existing SSIS packages to ingest on-premises data into datasets stored in a dedicated SQL pool of Azure Synapse

Analytics and transform the data.

Identify a process to ensure that changes to the ingestion and transformation activities can be version-controlled and developed independently by multiple data engineers.

QUESTION:

You need to implement the surrogate key for the retail store table. The solution must meet the sales transaction dataset requirements.

What should you create?

- ☒ **a table that has an IDENTITY property**
- ☐ **a system-versioned temporal tab**
- ☐ **a user-defined SEQUENCE object**
- ☐ **a table that has a FOREIGN KEY constraint**

(Poprawne)

Wyjaśnienie

Correct Answer: A

Scenario: Implement a surrogate key to account for changes to the retail store addresses.

A surrogate key on a table is a column with a unique identifier for each row. The key is not generated from the table data. Data modelers like to create surrogate keys on their tables when they design data warehouse models. You can use the IDENTITY property to achieve this goal simply and effectively without affecting load performance.

Reference:

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

Design and implement data storage

Pytanie 40:

Pominięto

Case study -

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

To start the case study -

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an

All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

Overview -

Contoso, Ltd. is a clothing retailer based in Seattle. The company has 2,000 retail stores across the United States and an emerging online presence.

The network contains an Active Directory forest named contoso.com. The forest is integrated with an Azure Active Directory (Azure AD) tenant named contoso.com. Contoso has an Azure subscription associated to the contoso.com Azure AD tenant.

Existing Environment -

Transactional Data -

Contoso has three years of customer, transactional, operational, sourcing, and supplier data comprised of 10 billion records stored across multiple on-premises

Microsoft SQL Server servers. The SQL Server instances contain data from various operational systems. The data is loaded into the instances by using SQL

Server Integration Services (SSIS) packages.

You estimate that combining all product sales transactions into a company-wide sales transactions dataset will result in a single table that contains 5 billion rows, with one row per transaction.

Most queries targeting the sales transactions data will be used to identify which products were sold in retail stores and which products were sold online during different time periods. Sales transaction data that is older than three years will be removed monthly.

You plan to create a retail store table that will contain the address of each retail store. The table will be approximately 2 MB. Queries for retail store sales will include the retail store addresses.

You plan to create a promotional table that will contain a promotion ID. The promotion ID will be associated to a specific product. The product will be identified by a product ID. The table will be approximately 5 GB.

Streaming Twitter Data -

The ecommerce department at Contoso develops an Azure logic app that captures trending Twitter feeds referencing the company's products and pushes the products to Azure Event Hubs.

Planned Changes and Requirements

Planned Changes -

Contoso plans to implement the following changes:

Load the sales transaction dataset to Azure Synapse Analytics.

Integrate on-premises data stores with Azure Synapse Analytics by using SSIS packages.

Use Azure Synapse Analytics to analyze Twitter feeds to assess customer sentiments about products.

Sales Transaction Dataset Requirements

Contoso identifies the following requirements for the sales transaction dataset:

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.

Ensure that queries joining and filtering sales transaction records based on product ID complete as quickly as possible.

Implement a surrogate key to account for changes to the retail store addresses.

Ensure that data storage costs and performance are predictable.

Minimize how long it takes to remove old records.

Customer Sentiment Analytics Requirements

Contoso identifies the following requirements for customer sentiment analytics:

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.

Maximize the throughput of ingesting Twitter feeds from Event Hubs to Azure Storage without purchasing additional throughput or capacity units.

Store Twitter feeds in Azure Storage by using Event Hubs Capture. The feeds will be converted into Parquet files.

Ensure that the data store supports Azure AD-based access control down to the object level.

Minimize administrative effort to maintain the Twitter feed data records.

Purge Twitter feed data records that are older than two years.

Data Integration Requirements -

Contoso identifies the following requirements for data integration:

Use an Azure service that leverages the existing SSIS packages to ingest on-premises data into datasets stored in a dedicated SQL pool of Azure Synapse

Analytics and transform the data.

Identify a process to ensure that changes to the ingestion and transformation activities can be version-controlled and developed independently by multiple data engineers.

QUESTION:

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.

Hot Area:

Answer Area

Table type to store retail store data:

	▼
Hash	
Replicated	
Round-robin	

Table type to store promotional data:

	▼
Hash	
Replicated	
Round-robin	

☐

Hash

Round-robin

☐

Replicated

Hash

(Poprawne)

• ☐

Round-robin

Replicated

• ☐

Replicated

Round-robin

Wyjaśnienie

Correct Answer: B

Box 1: Replicated

Box 2: Hash -

Hash-distributed tables improve query performance on large fact tables.

Scenario:

- You plan to create a promotional table that will contain a promotion ID. The promotion ID will be associated to a specific product. The product will be identified by a product ID. The table will be approximately 5 GB.
- Ensure that queries joining and filtering sales transaction records based on product ID complete as quickly as possible.

Reference:

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

by

Pytanie 41:

Pominięto

Case study -

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

To start the case study -

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an

All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

Overview -

Contoso, Ltd. is a clothing retailer based in Seattle. The company has 2,000 retail stores across the United States and an emerging online presence.

The network contains an Active Directory forest named contoso.com. The forest is integrated with an Azure Active Directory (Azure AD) tenant named contoso.com. Contoso has an Azure subscription associated to the contoso.com Azure AD tenant.

Existing Environment -

Transactional Data -

Contoso has three years of customer, transactional, operational, sourcing, and supplier data comprised of 10 billion records stored across multiple on-premises

Microsoft SQL Server servers. The SQL Server instances contain data from various operational systems. The data is loaded into the instances by using SQL

Server Integration Services (SSIS) packages.

You estimate that combining all product sales transactions into a company-wide sales transactions dataset will result in a single table that contains 5 billion rows, with one row per transaction.

Most queries targeting the sales transactions data will be used to identify which products were sold in retail stores and which products were sold online during different time periods. Sales transaction data that is older than three years will be removed monthly.

You plan to create a retail store table that will contain the address of each retail store. The table will be approximately 2 MB. Queries for retail store sales will include the retail store addresses.

You plan to create a promotional table that will contain a promotion ID. The promotion ID will be associated to a specific product. The product will be identified by a product ID. The table will be approximately 5 GB.

Streaming Twitter Data -

The ecommerce department at Contoso develops an Azure logic app that captures trending Twitter feeds referencing the company's products and pushes the products to Azure Event Hubs.

Planned Changes and Requirements

Planned Changes -

Contoso plans to implement the following changes:

Load the sales transaction dataset to Azure Synapse Analytics.

Integrate on-premises data stores with Azure Synapse Analytics by using SSIS packages.

Use Azure Synapse Analytics to analyze Twitter feeds to assess customer sentiments about products.

Sales Transaction Dataset Requirements

Contoso identifies the following requirements for the sales transaction dataset:

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.

Ensure that queries joining and filtering sales transaction records based on product ID complete as quickly as possible.

Implement a surrogate key to account for changes to the retail store addresses.

Ensure that data storage costs and performance are predictable.

Minimize how long it takes to remove old records.

Customer Sentiment Analytics Requirements

Contoso identifies the following requirements for customer sentiment analytics:

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.

Maximize the throughput of ingesting Twitter feeds from Event Hubs to Azure Storage without purchasing additional throughput or capacity units.

Store Twitter feeds in Azure Storage by using Event Hubs Capture. The feeds will be converted into Parquet files.

Ensure that the data store supports Azure AD-based access control down to the object level.

Minimize administrative effort to maintain the Twitter feed data records.

Purge Twitter feed data records that are older than two years.

Data Integration Requirements -

Contoso identifies the following requirements for data integration:

Use an Azure service that leverages the existing SSIS packages to ingest on-premises data into datasets stored in a dedicated SQL pool of Azure Synapse

Analytics and transform the data.

Identify a process to ensure that changes to the ingestion and transformation activities can be version-controlled and developed independently by multiple data engineers.

QUESTION:

You need to implement an Azure Synapse Analytics database object for storing the sales transactions data. The solution must meet the sales transaction dataset requirements.

What should you do? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

Transact-SQL DDL command to use:

	▼
CREATE EXTERNAL TABLE	
CREATE TABLE	
CREATE VIEW	

Partitioning option to use in the WITH clause of the DDL statement:

	▼
FORMAT_OPTIONS	
FORMAT_TYPE	
RANGE LEFT FOR VALUES	
RANGE RIGHT FOR VALUES	

- ☐ **CREATE EXTERNAL TABLE**
- ☐ **FORMAT_OPTIONS**
- ☐ **CREATE TABLE**
- ☐ **FORMAT_TYPE**
- ☐ **CREATE VIEW**

RANGE LEFT FOR VALUES



CREATE TABLE

RANGE RIGHT FOR VALUES

(Poprawne)

Wyjaśnienie

Correct Answer:

Answer Area

Transact-SQL DDL command to use:

	▼
CREATE EXTERNAL TABLE	
CREATE TABLE	
CREATE VIEW	

Partitioning option to use in the WITH clause of the DDL statement:

	▼
FORMAT_OPTIONS	
FORMAT_TYPE	
RANGE LEFT FOR VALUES	
RANGE RIGHT FOR VALUES	

Box 1: Create table -

Scenario: Load the sales transaction dataset to Azure Synapse Analytics

Box 2: RANGE RIGHT FOR VALUES -

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.

RANGE RIGHT: Specifies the boundary value belongs to the partition on the right (higher values).

FOR VALUES (boundary_value [...n]): Specifies the boundary values for the partition.

Scenario: Load the sales transaction dataset to Azure Synapse Analytics.

Contoso identifies the following requirements for the sales transaction dataset:

- 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.
- Ensure that queries joining and filtering sales transaction records based on product ID complete as quickly as possible.
- Implement a surrogate key to account for changes to the retail store addresses.
- Ensure that data storage costs and performance are predictable.
- Minimize how long it takes to remove old records.

Reference:

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

Pytanie 42:

Pominięto

Case study -

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

To start the case study -

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an

All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

Overview -

Contoso, Ltd. is a clothing retailer based in Seattle. The company has 2,000 retail stores across the United States and an emerging online presence.

The network contains an Active Directory forest named contoso.com. The forest is integrated with an Azure Active Directory (Azure AD) tenant named contoso.com. Contoso has an Azure subscription associated to the contoso.com Azure AD tenant.

Existing Environment -

Transactional Data -

Contoso has three years of customer, transactional, operational, sourcing, and supplier data comprised of 10 billion records stored across multiple on-premises

Microsoft SQL Server servers. The SQL Server instances contain data from various operational systems. The data is loaded into the instances by using SQL

Server Integration Services (SSIS) packages.

You estimate that combining all product sales transactions into a company-wide sales transactions dataset will result in a single table that contains 5 billion rows, with one row per transaction.

Most queries targeting the sales transactions data will be used to identify which products were sold in retail stores and which products were sold online during different time periods. Sales transaction data that is older than three years will be removed monthly.

You plan to create a retail store table that will contain the address of each retail store. The table will be approximately 2 MB. Queries for retail store sales will include the retail store addresses.

You plan to create a promotional table that will contain a promotion ID. The promotion ID will be associated to a specific product. The product will be identified by a product ID. The table will be approximately 5 GB.

Streaming Twitter Data -

The ecommerce department at Contoso develops an Azure logic app that captures trending Twitter feeds referencing the company's products and pushes the products to Azure Event Hubs.

Planned Changes and Requirements

Planned Changes -

Contoso plans to implement the following changes:

Load the sales transaction dataset to Azure Synapse Analytics.

Integrate on-premises data stores with Azure Synapse Analytics by using SSIS packages.

Use Azure Synapse Analytics to analyze Twitter feeds to assess customer sentiments about products.

Sales Transaction Dataset Requirements

Contoso identifies the following requirements for the sales transaction dataset:

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.

Ensure that queries joining and filtering sales transaction records based on product ID complete as quickly as possible.

Implement a surrogate key to account for changes to the retail store addresses.

Ensure that data storage costs and performance are predictable.

Minimize how long it takes to remove old records.

Customer Sentiment Analytics Requirements

Contoso identifies the following requirements for customer sentiment analytics:

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.

Maximize the throughput of ingesting Twitter feeds from Event Hubs to Azure Storage without purchasing additional throughput or capacity units.

Store Twitter feeds in Azure Storage by using Event Hubs Capture. The feeds will be converted into Parquet files.

Ensure that the data store supports Azure AD-based access control down to the object level.

Minimize administrative effort to maintain the Twitter feed data records.

Purge Twitter feed data records that are older than two years.

Data Integration Requirements -

Contoso identifies the following requirements for data integration:

Use an Azure service that leverages the existing SSIS packages to ingest on-premises data into datasets stored in a dedicated SQL pool of Azure Synapse

Analytics and transform the data.

Identify a process to ensure that changes to the ingestion and transformation activities can be version-controlled and developed independently by multiple data engineers.

QUESTION:

You need to design a data retention solution for the Twitter feed data records. The solution must meet the customer sentiment analytics requirements.

Which Azure Storage functionality should you include in the solution?

☐

change feed

☐

soft delete

☐

time-based retention

☒

lifecycle management

(Poprawne)

Wyjaśnienie

Correct Answer: D

Scenario: Purge Twitter feed data records that are older than two years.

Data sets have unique lifecycles. Early in the lifecycle, people access some data often. But the need for access often drops drastically as the data ages. Some data remains idle in the cloud and is rarely accessed once stored. Some data sets expire days or months after creation, while other data sets are actively read and modified throughout their lifetimes. Azure Storage lifecycle management offers a rule-based policy that you can use to transition blob data to the appropriate access tiers or to expire data at the end of the data lifecycle.

Reference:

<https://docs.microsoft.com/en-us/azure/storage/blobs/lifecycle-management-overview>

Pytanie 43:

Pominięto

Case study

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

To start the case study

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

Overview

Litware, Inc. owns and operates 300 convenience stores across the US. The company sells a variety of packaged foods and drinks, as well as a variety of prepared foods, such as sandwiches and pizzas.

Litware has a loyalty club whereby members can get daily discounts on specific items by providing their membership number at checkout.

Litware employs business analysts who prefer to analyze data by using Microsoft Power BI, and data scientists who prefer analyzing data in Azure Databricks notebooks.

Requirements

Business Goals

Litware wants to create a new analytics environment in Azure to meet the following requirements:

- See inventory levels across the stores. Data must be updated as close to real time as possible.
- Execute ad hoc analytical queries on historical data to identify whether the loyalty club discounts increase sales of the discounted products.
- Every four hours, notify store employees about how many prepared food items to produce based on historical demand from the sales data.

Technical Requirements

Litware identifies the following technical requirements:

- Minimize the number of different Azure services needed to achieve the business goals.
- Use platform as a service (PaaS) offerings whenever possible and avoid having to provision virtual machines that must be managed by Litware.
- Ensure that the analytical data store is accessible only to the company's on-premises network and Azure services.
- Use Azure Active Directory (Azure AD) authentication whenever possible.
- Use the principle of least privilege when designing security.
- Stage Inventory data in Azure Data Lake Storage Gen2 before loading the data into the analytical data store. Litware wants to remove transient data from Data Lake Storage once the data is no longer in use. Files that have a modified date that is older than 14 days must be removed.
- Limit the business analysts' access to customer contact information, such as phone numbers, because this type of data is not analytically relevant.
- Ensure that you can quickly restore a copy of the analytical data store within one hour in the event of corruption or accidental deletion.

Planned Environment

Litware plans to implement the following environment:

- The application development team will create an Azure event hub to receive real-time sales data, including store number, date, time, product ID, customer loyalty number, price, and discount amount, from the point of sale (POS) system and output the data to data storage in Azure.
- Customer data, including name, contact information, and loyalty number, comes from Salesforce, a SaaS application, 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.
- Daily inventory data comes from a Microsoft SQL server located on a private network.
- Litware currently has 5 TB of historical sales data and 100 GB of customer data. The company expects approximately 100 GB of new data per month for the next year.
- Litware will build a custom application named FoodPrep to provide store employees with the calculation results of how many prepared food items to produce every four hours.
- Litware does not plan to implement Azure ExpressRoute or a VPN between the on-premises network and Azure.

QUESTION:

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.

Hot Area:

Answer Area

Integration runtime type:	<div><div></div><div>Azure integration runtime</div><div>Azure-SSIS integration runtime</div><div>Self-hosted integration runtime</div></div>
Trigger type:	<div><div></div><div>Event-based trigger</div><div>Schedule trigger</div><div>Tumbling window trigger</div></div>
Activity type:	<div><div></div><div>Copy activity</div><div>Lookup activity</div><div>Stored procedure activity</div></div>

- ☐ Azure integration runtime
Event-based trigger
Lookup activity
- ☐ Azure-SSIS integration runtime
Tumbling window trigger
Stored procedure activity
- ☐ Self-hosted integration
Schedule trigger
Copy activity

(Poprawne)

Wyjaśnienie

Correct Answer:

Answer Area

Integration runtime type:	<div><div></div><div>Azure integration runtime</div><div>Azure-SSIS integration runtime</div><div>Self-hosted integration runtime</div></div>
Trigger type:	<div><div></div><div>Event-based trigger</div><div>Schedule trigger</div><div>Tumbling window trigger</div></div>
Activity type:	<div><div></div><div>Copy activity</div><div>Lookup activity</div><div>Stored procedure activity</div></div>

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.

Design and develop data processing

Pytanie 44:

Pominięto

Case study

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

To start the case study

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

Overview

Litware, Inc. owns and operates 300 convenience stores across the US. The company sells a variety of packaged foods and drinks, as well as a variety of prepared foods, such as sandwiches and pizzas.

Litware has a loyalty club whereby members can get daily discounts on specific items by providing their membership number at checkout.

Litware employs business analysts who prefer to analyze data by using Microsoft Power BI, and data scientists who prefer analyzing data in Azure Databricks notebooks.

Requirements

Business Goals

Litware wants to create a new analytics environment in Azure to meet the following requirements:

- See inventory levels across the stores. Data must be updated as close to real time as possible.
- Execute ad hoc analytical queries on historical data to identify whether the loyalty club discounts increase sales of the discounted products.
- Every four hours, notify store employees about how many prepared food items to produce based on historical demand from the sales data.

Technical Requirements

Litware identifies the following technical requirements:

- Minimize the number of different Azure services needed to achieve the business goals.
- Use platform as a service (PaaS) offerings whenever possible and avoid having to provision virtual machines that must be managed by Litware.
- Ensure that the analytical data store is accessible only to the company's on-premises network and Azure services.
- Use Azure Active Directory (Azure AD) authentication whenever possible.
- Use the principle of least privilege when designing security.
- Stage Inventory data in Azure Data Lake Storage Gen2 before loading the data into the analytical data store. Litware wants to remove transient data from Data Lake Storage once the data is no longer in use. Files that have a modified date that is older than 14 days must be removed.
- Limit the business analysts' access to customer contact information, such as phone numbers, because this type of data is not analytically relevant.
- Ensure that you can quickly restore a copy of the analytical data store within one hour in the event of corruption or accidental deletion.

Planned Environment

Litware plans to implement the following environment:

- The application development team will create an Azure event hub to receive real-time sales data, including store number, date, time, product ID, customer loyalty number, price, and discount amount, from the point of sale (POS) system and output the data to data storage in Azure.
- Customer data, including name, contact information, and loyalty number, comes from Salesforce, a SaaS application, 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.
- Daily inventory data comes from a Microsoft SQL server located on a private network.
- Litware currently has 5 TB of historical sales data and 100 GB of customer data. The company expects approximately 100 GB of new data per month for the next year.
- Litware will build a custom application named FoodPrep to provide store employees with the calculation results of how many prepared food items to produce every four hours.
- Litware does not plan to implement Azure ExpressRoute or a VPN between the on-premises network and Azure.

QUESTION:

What should you recommend to prevent users outside the Litware on-premises network from accessing the analytical data store?

- ☐ a server-level virtual network rule
- ☐ a database-level virtual network rule
- ☐ a server-level firewall IP rule
- ☒ (Poprawne)

a database-level firewall IP rule

Wyjaśnienie

Correct Answer: C

As mentioned Litware does not plan to implement VPN.

they will have to create firewall ip rules to allow connection from the IP ranges of the on-premise network. They can also use the firewall rule 0.0.0.0 to allow access from Azure services.

Pytanie 45:

Pominięto

Case study

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

To start the case study

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

Overview

Litware, Inc. owns and operates 300 convenience stores across the US. The company sells a variety of packaged foods and drinks, as well as a variety of prepared foods, such as sandwiches and pizzas.

Litware has a loyalty club whereby members can get daily discounts on specific items by providing their membership number at checkout.

Litware employs business analysts who prefer to analyze data by using Microsoft Power BI, and data scientists who prefer analyzing data in Azure Databricks notebooks.

Requirements

Business Goals

Litware wants to create a new analytics environment in Azure to meet the following requirements:

- See inventory levels across the stores. Data must be updated as close to real time as possible.
- Execute ad hoc analytical queries on historical data to identify whether the loyalty club discounts increase sales of the discounted products.
- Every four hours, notify store employees about how many prepared food items to produce based on historical demand from the sales data.

Technical Requirements

Litware identifies the following technical requirements:

- Minimize the number of different Azure services needed to achieve the business goals.
- Use platform as a service (PaaS) offerings whenever possible and avoid having to provision virtual machines that must be managed by Litware.
- Ensure that the analytical data store is accessible only to the company's on-premises network and Azure services.
- Use Azure Active Directory (Azure AD) authentication whenever possible.
- Use the principle of least privilege when designing security.
- Stage Inventory data in Azure Data Lake Storage Gen2 before loading the data into the analytical data store. Litware wants to remove transient data from Data Lake Storage once the data is no longer in use. Files that have a modified date that is older than 14 days must be removed.
- Limit the business analysts' access to customer contact information, such as phone numbers, because this type of data is not analytically relevant.
- Ensure that you can quickly restore a copy of the analytical data store within one hour in the event of corruption or accidental deletion.

Planned Environment

Litware plans to implement the following environment:

- The application development team will create an Azure event hub to receive real-time sales data, including store number, date, time, product ID, customer loyalty number, price, and discount amount, from the point of sale (POS) system and output the data to data storage in Azure.
- Customer data, including name, contact information, and loyalty number, comes from Salesforce, a SaaS application, 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.
- Daily inventory data comes from a Microsoft SQL server located on a private network.
- Litware currently has 5 TB of historical sales data and 100 GB of customer data. The company expects approximately 100 GB of new data per month for the next year.
- Litware will build a custom application named FoodPrep to provide store employees with the calculation results of how many prepared food items to produce every four hours.
- Litware does not plan to implement Azure ExpressRoute or a VPN between the on-premises network and Azure.

QUESTION:

What should you recommend using to secure sensitive customer contact information?

- ☐ **Transparent Data Encryption (TDE)**
- ☐ **row-level security**
- ☐ **column-level security**
- ☒ **(Poprawne)**
- ☐

data sensitivity labels

Wyjaśnienie

Correct Answer: C

You can use CLS(column-level security) to manage user access to specific columns in your tables in a simpler manner, without having to redesign your data warehouse. CLS eliminates the need to maintain access restriction logic away from the data in another application or introduce views for filtering out sensitive columns for a subset of users.

<https://azure.microsoft.com/en-ca/updates/column-level-security-is-now-supported-in-azure-sql-data-warehouse/>

Pytanie 46:

Pominięto

Case study

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

To start the case study

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

Overview

Litware, Inc. owns and operates 300 convenience stores across the US. The company sells a variety of packaged foods and drinks, as well as a variety of prepared foods, such as sandwiches and pizzas.

Litware has a loyalty club whereby members can get daily discounts on specific items by providing their membership number at checkout.

Litware employs business analysts who prefer to analyze data by using Microsoft Power BI, and data scientists who prefer analyzing data in Azure Databricks notebooks.

Requirements

Business Goals

Litware wants to create a new analytics environment in Azure to meet the following requirements:

- See inventory levels across the stores. Data must be updated as close to real time as possible.
- Execute ad hoc analytical queries on historical data to identify whether the loyalty club discounts increase sales of the discounted products.
- Every four hours, notify store employees about how many prepared food items to produce based on historical demand from the sales data.

Technical Requirements

Litware identifies the following technical requirements:

- Minimize the number of different Azure services needed to achieve the business goals.
- Use platform as a service (PaaS) offerings whenever possible and avoid having to provision virtual machines that must be managed by Litware.
- Ensure that the analytical data store is accessible only to the company's on-premises network and Azure services.
- Use Azure Active Directory (Azure AD) authentication whenever possible.
- Use the principle of least privilege when designing security.
- Stage Inventory data in Azure Data Lake Storage Gen2 before loading the data into the analytical data store. Litware wants to remove transient data from Data Lake Storage once the data is no longer in use. Files that have a modified date that is older than 14 days must be removed.
- Limit the business analysts' access to customer contact information, such as phone numbers, because this type of data is not analytically relevant.
- Ensure that you can quickly restore a copy of the analytical data store within one hour in the event of corruption or accidental deletion.

Planned Environment

Litware plans to implement the following environment:

- The application development team will create an Azure event hub to receive real-time sales data, including store number, date, time, product ID, customer loyalty number, price, and discount amount, from the point of sale (POS) system and output the data to data storage in Azure.
- Customer data, including name, contact information, and loyalty number, comes from Salesforce, a SaaS application, 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.
- Daily inventory data comes from a Microsoft SQL server located on a private network.
- Litware currently has 5 TB of historical sales data and 100 GB of customer data. The company expects approximately 100 GB of new data per month for the next year.
- Litware will build a custom application named FoodPrep to provide store employees with the calculation results of how many prepared food items to produce every four hours.
- Litware does not plan to implement Azure ExpressRoute or a VPN between the on-premises network and Azure.

QUESTION:

What should you do to improve high availability of the real-time data processing solution?

☐

Deploy a High Concurrency Databricks cluster.

☐

Deploy an Azure Stream Analytics job and use an Azure Automation runbook to check the status of the job and to start the job if it stops.

☐

Set Data Lake Storage to use geo-redundant storage (GRS).

☐

Deploy identical Azure Stream Analytics jobs to paired regions in Azure.

(Poprawne)

Wyjaśnienie

Correct Answer: D

Guarantee Stream Analytics job reliability during service updates

Part of being a fully managed service is the capability to introduce new service functionality and improvements at a rapid pace. As a result, Stream Analytics can have a service update deploy on a weekly (or more frequent) basis. No matter how much testing is done there is still a risk that an existing, running job may break due to the introduction of a bug. If you are running mission critical jobs, these risks need to be avoided. You can reduce this risk by following Azure's paired region model.

Scenario: The application development team will create an Azure event hub to receive real-time sales data, including store number, date, time, product ID, customer loyalty number, price, and discount amount, from the point of sale (POS) system and output the data to data storage in Azure

Reference:

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

Pytanie 47:

Pominięto

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?

- ☐ as a temporal table
- ☐ as a SQL graph table
- ☐ as a degenerate dimension table
- ☐ as a Type 2 slowly changing dimension (SCD) table

(Poprawne)

Wyjaśnienie

Correct Answer: D

Temporal table is better than SCD2, but it is not supported in Synapse yet

Temporal tables mentioned in the link below are supported in Azure SQL Database(PaaS) and Azure Managed Instance, where as in this question Dedicated SQL Pools are mentioned so no temporal tables can be used

Reference:

<https://docs.microsoft.com/en-us/azure/azure-sql/temporal-tables>

Pytanie 48:

Pominięto

You are designing an Azure Synapse Analytics dedicated SQL pool.

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

What should you include in the solution?

☐

column-level security

☐

dynamic data masking

☐

row-level security (RLS)

☐

sensitivity classifications

(Poprawne)

Wyjaśnienie

Correct Answer: D

Data Discovery & Classification is built into Azure SQL Database, Azure SQL Managed Instance, and Azure Synapse Analytics. It provides basic capabilities for discovering, classifying, labeling, and reporting the sensitive data in your databases. Your most sensitive data might include business, financial, healthcare, or personal information. Discovering and classifying this data can play a pivotal role in your organization's information-protection approach. It can serve as infrastructure for:

- Helping to meet standards for data privacy and requirements for regulatory compliance.
- Various security scenarios, such as monitoring (auditing) access to sensitive data.
- Controlling access to and hardening the security of databases that contain highly sensitive data.

Reference:

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

Pytanie 49:**Pominięto**

You use PySpark in Azure Databricks to parse the following JSON input.

```
{
  "persons": [
    {
      "name": "Keith",
      "age": 30,
      "dogs": ["Fido", "Fluffy"]
    },
    {
      "name": "Donna",
      "age": 46,
      "dogs": ["Spot"]
    }
  ]
}
```

You need to output the data in the following tabular format.

owner	age	dog
Keith	30	Fido
Keith	30	Fluffy
Donna	46	Spot

How should you complete the PySpark 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.

Select and Place:

Values	Answer Area
alias	dbutils.fs.put("/tmp/source.json", source_json, True) source_df = spark.read.option("multiline", "true").json("/tmp/source.json")
array_union	persons = source_df. <input type="text" value="Value"/> <input type="text" value="Value"/> ("persons").alias("persons")
createDataFrame	persons_dogs = persons.select(col("persons.name").alias("owner"), col("persons.age").alias("age"),
explode	explode <input type="text" value="Value"/> ("dog"))
explode	("persons-dogs").
select	display(persons_dogs)
translate	

- ☐
 - alias
 - translate
 - explode
- ☐
 - select
 - explode
 - alias
 - (Poprawne)

- ☐
 - array_union
 - createDataFrame
 - explode
- ☐
 - select
 - createDataFrame
 - alias

Wyjaśnienie
Correct Answer:

Values	Answer Area
<input type="text"/>	dbutils.fs.put("/tmp/source.json", source_json, True) source_df = spark.read.option("multiline", "true").json("/tmp/source.json")
<input type="text" value="array_union"/>	persons = source_df. <input type="text" value="select"/> <input type="text" value="explode"/> ("persons").alias("persons"))
<input type="text" value="createDataFrame"/>	persons_dogs = persons.select(col("persons.name").alias("owner"), col("persons.age").alias("age"), explode <input type="text" value="alias"/> ("dog")) ("persons-dogs"). display(persons_dogs)
<input type="text"/>	
<input type="text" value="translate"/>	

Box 1: select -

Box 2: explode -

Box 3: alias -

`pyspark.sql.Column.alias` returns this column aliased with a new name or names (in the case of expressions that return more than one column, such as `explode`).

Reference:

<https://spark.apache.org/docs/latest/api/python/reference/api/pyspark.sql.Column.alias.html> <https://docs.microsoft.com/en-us/azure/databricks/sql/language-manual/functions/explode>

Pytanie 50:

Pominięto

You are designing an application that will store petabytes of medical imaging data. When the data is first created, the data will be accessed frequently during the first week. After one month, the data must be accessible within 30 seconds, but files will be accessed infrequently. After one year, the data will be accessed infrequently but must be accessible within five minutes.

You need to select a storage strategy for the data. The solution must minimize costs. Which storage tier should you use for each time frame? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

First week:

	▼
Archive	
Cool	
Hot	

After one month:

	▼
Archive	
Cool	
Hot	

After one year:

	▼
Archive	
Cool	
Hot	



Hot

Cool

Cool

(Poprawne)

•



Hot

Cool

Archive

•



Cool

Hot

Cool

•



Hot

Archive

Archive

Wyjaśnienie

Correct Answer:

Answer Area

First week:

Archive
Cool
Hot

After one month:

Archive
Cool
Hot

After one year:

Archive
Cool
Hot

Box 1: Hot -

Hot tier - An online tier optimized for storing data that is accessed or modified frequently. The Hot tier has the highest storage costs, but the lowest access costs.

Box 2: Cool -

Cool tier - An online tier optimized for storing data that is infrequently accessed or modified. Data in the Cool tier should be stored for a minimum of 30 days. The Cool tier has lower storage costs and higher access costs compared to the Hot tier.

Box 3: Cool -

Not Archive tier - An offline tier optimized for storing data that is rarely accessed, and that has flexible latency requirements, on the order of hours. Data in the Archive tier should be stored for a minimum of 180 days.

	Premium performance	Hot tier	Cool tier	Archive tier
Availability	99.9%	99.9%	99%	Offline
Availability (RA-GRS reads)	N/A	99.99%	99.9%	Offline
Usage charges	Higher storage costs, lower access, and transaction cost	Higher storage costs, lower access, and transaction costs	Lower storage costs, higher access, and transaction costs	Lowest storage costs, highest access, and transaction costs
Minimum object size	N/A	N/A	N/A	N/A
Minimum storage duration	N/A	N/A	30 days ¹	180 days
Latency (Time to first byte)	Single-digit milliseconds	milliseconds	milliseconds	hours ²

Reference:

<https://docs.microsoft.com/en-us/azure/storage/blobs/access-tiers-overview>

<https://www.altaro.com/hyper-v/azure-archive-storage/>