

Pytanie 1:

Pominięto

Your data engineering team needs to develop a solution that will make use of Azure Stream Analytics. The stream will take in data from Azure Event Hubs. Azure Event Hubs has been configured to take in the streaming data from twitter.

You have to write the SQL statement that will be part of the Stream Analytics job. Below is the key requirement of the job

“Provide the count of tweets every 5 seconds over the last 10 seconds”

Below is the SQL statement that needs to be completed

SELECT Topic, COUNT(*) as Total

FROM Inputstream

Area 1

CreatedAt

GROUP By Topic,

Area 2

(second,10,5)

Which of the following will go into Area 1?

• ☐

ORDER BY

• ☐

COUNT

• ☐

TIMESTAMP BY

(Poprawne)

Wyjaśnienie

Here we are ensuring the events are sorted by the timestamp for each window.

Pytanie 2:

Pominięto

Your data engineering team needs to develop a solution that will make use of Azure Stream Analytics. The stream will take in data from Azure Event Hubs. Azure Event Hubs has been configured to take in the streaming data from twitter.

You have to write the SQL statement that will be part of the Stream Analytics job. Below is the key requirement of the job

“Provide the count of tweets every 5 seconds over the last 10 seconds”

Below is the SQL statement that needs to be completed

```
SELECT Topic, COUNT(*) as Total  
  
FROM Inputstream Area 1 CreatedAt  
  
GROUP By Topic, Area 2 (second,10,5)
```

Which of the following will go into Area 2?

☐

TumblingWindow

☐

HoppingWindow

(Poprawne)

☐

SlidingWindow

Wyjaśnienie

Here we need to use the Hopping window. Here each window has a size of 10 seconds. And every 5 seconds a new window is created.

This reference has been taken from the below link

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

Pytanie 3:

Pominięto

Your team has an Azure Synapse Analytics workspace in place. It also has a Dedicated SQL pool

The following table has been created in a Dedicated SQL pool

```
1 CREATE TABLE [dbo].[DimStudent]
2 (
3     studentkey int NOT NULL,
4     studentemailaddress nvarchar(500),
5     enrolmentdate date,
6     address nvarchar(100)
7 )
8 WITH
9 (
10     DISTRIBUTION = REPLICATE,
11     CLUSTERED COLUMNSTORE INDEX
12 )
13
14
```

When you issue the following statement

```
SELECT studentkey,studentemailaddress,format(enrolmentdate,'MM/dd/yyyy') AS
enrolmentdate,address FROM [dbo].[DimStudent]
```

You get the following values

ResultsMessages

View

TableChart

Export results

Search

studentkey	studentemailaddress	enrolmentdate	address
3	userC@gmail.com	05/20/2021	address3
1	userA@gmail.com	01/20/2021	address1
2	userB@gmail.com	03/20/2021	address2

Two users have been defined in the database that are given below

Username	Database role
companyusrA	Server admin
companyusrB	db_datareader

The following masking rules have been applied to the table

Masking rules

Schema	Table	Column	Mask Function
dbo	DimStudent	enrolmentdate	Default value (0, xxxx, ...
dbo	DimStudent	studentemailaddress	Email (aXXX@XXXX.co...
dbo	DimStudent	studentkey	Default value (0, xxxx, ...

What is the value that will be shown to companyusrA when they issue the following statement?

```
SELECT studentemailaddress FROM [dbo].[DimStudent]
```

```
WHERE studentkey=1
```

- ☐ **userA@gmail.com**
(Poprawne)
- ☐ **uXXX@XXXX.com**
- ☐ **userA@XXXX.com**

Wyjaśnienie

Since here the user is an admin level user for the database , the masking rules will not apply

For more information on Dynamic data masking , one can visit the below URL

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

Pytanie 4:

Pominięto

Your team has an Azure Synapse Analytics workspace in place. It also has a Dedicated SQL pool

The following table has been created in a Dedicated SQL pool

```
1 CREATE TABLE [dbo].[DimStudent]
2 (
3     studentkey int NOT NULL,
4     studentemailaddress nvarchar(500),
5     enrolmentdate date,
6     address nvarchar(100)
7 )
8 WITH
9 (
10     DISTRIBUTION = REPLICATE,
11     CLUSTERED COLUMNSTORE INDEX
12 )
13
14
```

When you issue the following statement

```
SELECT studentkey,studentemailaddress,format(enrolmentdate,'MM/dd/yyyy') AS
enrolmentdate,address FROM [dbo].[DimStudent]
```

You get the following values

ResultsMessages

View

TableChart

Export results

Search

studentkey	studentemailaddress	enrolmentdate	address
3	userC@gmail.com	05/20/2021	address3
1	userA@gmail.com	01/20/2021	address1
2	userB@gmail.com	03/20/2021	address2

Two users have been defined in the database that are given below

Username	Database role
companyusrA	Server admin
companyusrB	db_datareader

The following masking rules have been applied to the table

Masking rules

Schema	Table	Column	Mask Function
dbo	DimStudent	enrolmentdate	Default value (0, xxxx, ...
dbo	DimStudent	studentemailaddress	Email (aXXX@XXXX.co...
dbo	DimStudent	studentkey	Default value (0, xxxx, ...

What is the value that will be shown to companyusrB when they issue the following statement?

```
SELECT studentemailaddress FROM [dbo].[DimStudent]
```

```
WHERE studentkey=1
```

☐ ☐

userA@gmail.com

☐ ☐

uXXX@XXXX.com

(Poprawne)

☐ ☐

userA@XXXX.com

Wyjaśnienie

Here since the Email masking function has been used , the constant prefix of **uXXX@XXXX.com** will be returned.

For more information on Dynamic data masking , one can visit the below URL

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

Pytanie 5:

Pominięto

Your team has an Azure Synapse Analytics workspace in place. It also has a Dedicated SQL pool

The following table has been created in a Dedicated SQL pool

```
1 CREATE TABLE [dbo].[DimStudent]
2 (
3     studentkey int NOT NULL,
4     studentemailaddress nvarchar(500),
5     enrolmentdate date,
6     address nvarchar(100)
7 )
8 WITH
9 (
10     DISTRIBUTION = REPLICATE,
11     CLUSTERED COLUMNSTORE INDEX
12 )
13
14
```

When you issue the following statement

```
SELECT studentkey,studentemailaddress,format(enrolmentdate,'MM/dd/yyyy') AS
enrolmentdate,address FROM [dbo].[DimStudent]
```

You get the following values

ResultsMessages

View

TableChart

Export results

Search

studentkey	studentemailaddress	enrolmentdate	address
3	userC@gmail.com	05/20/2021	address3
1	userA@gmail.com	01/20/2021	address1
2	userB@gmail.com	03/20/2021	address2

Two users have been defined in the database that are given below

Username	Database role
companyusrA	Server admin
companyusrB	db_datareader

The following masking rules have been applied to the table

Masking rules

Schema	Table	Column	Mask Function
dbo	DimStudent	enrolmentdate	Default value (0, xxxx, ...
dbo	DimStudent	studentemailaddress	Email (aXXX@XXXX.co...
dbo	DimStudent	studentkey	Default value (0, xxxx, ...

What is the value that will be shown to companyusrB when they issue the following statement?

```
SELECT studentkey FROM [dbo].[DimStudent]
```

```
WHERE studentkey=1
```

- ☐ 0
- ☒ (Poprawne)
- ☐ 1
- ☐ Null value

Wyjaśnienie

Here since the Default masking function has been used , and since the data type is an integer , the default value of 0 will be returned.

For more information on Dynamic data masking , one can visit the below URL

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

Pytanie 6:

Pominięto

You currently have a table named logdata in an Azure Dedicated SQL pool within an Azure Synapse Analytics workspace. Currently the table is partitioned based on the Time column in the table. You need to move a partition from the logdata onto a new table. You need to complete the below SQL statements for this requirement

Area 1

```
TABLE [logdata_new]
WITH
(
DISTRIBUTION = ROUND_ROBIN,
PARTITION ( [Time] RANGE RIGHT FOR VALUES
('2021-05-01','2021-06-01') ) )
AS
SELECT *
FROM logdata
WHERE 1=2
```

Area 2

TABLE [logdata]

Area 3

PARTITION 2 TO [logdata_new] PARTITION 1;

Which of the following would come in Area 1?

☐

CREATE

(Poprawne)

☐

UPDATE

☐

ALTER

☐

SWITCH

Wyjaśnienie

The first step is to create a new table. We then need to switch the partition from the original table to the new table

For more information on Table Partitions , one can visit the below URL

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

Pytanie 7:

Pominięto

You currently have a table named logdata in an Azure Dedicated SQL pool within an Azure Synapse Analytics workspace. Currently the table is partitioned based on the Time column in the table. You need to move a partition from the logdata onto a new table. You need to complete the below SQL statements for this requirement

Area 1

```
TABLE [logdata_new]
WITH
(
DISTRIBUTION = ROUND_ROBIN,
PARTITION ( [Time] RANGE RIGHT FOR VALUES
('2021-05-01','2021-06-01') ) )
AS
SELECT *
FROM logdata
WHERE 1=2
```

Area 2

TABLE [logdata]

Area 3

PARTITION 2 TO [logdata_new] PARTITION 1;

Which of the following would come in Area 2?

- ☐ **CREATE**
- ☐ **UPDATE**
- ☐ **ALTER**
(Poprawne)
- ☐ **SWITCH**

Wyjaśnienie

Here we need to ALTER the existing table to switch the partitions.

For more information on Table Partitions , one can visit the below URL

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

Pytanie 8:

Pominięto

You currently have a table named logdata in an Azure Dedicated SQL pool within an Azure Synapse Analytics workspace. Currently the table is partitioned based on the Time column in the table. You need to move a partition from the logdata onto a new table. You need to complete the below SQL statements for this requirement

Area 1

TABLE [logdata_new]

WITH

(

DISTRIBUTION = ROUND_ROBIN,

PARTITION ([Time] RANGE RIGHT FOR VALUES

('2021-05-01','2021-06-01')))

AS

SELECT *

FROM logdata

WHERE 1=2

Area 2

TABLE [logdata]

Area 3

PARTITION 2 TO [logdata_new] PARTITION 1;

Which of the following would come in Area 3?

☐

CREATE

☐

UPDATE

☐

ALTER

☐

SWITCH

(Poprawne)

Wyjaśnienie

Here we need to use the SWITCH command to switch the partition from the original to the new table.

For more information on Table Partitions , one can visit the below URL

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

Pytanie 9:

Pominięto

Your team currently has the following resources defined on Azure

- 1) An Azure Event Hub namespace and an Event Hub. The Event Hub is used to stream data from an external data source
- 2) An Azure Databricks cluster
- 3) An Azure Synapse Analytics workspace that contains a Dedicated SQL Pool

A Notebook is being developed in Scala in Azure Databricks. The Notebook will be used to take in the streaming data from Azure Event Hubs and add it on a table in the Dedicated SQL Pool

Below is a snippet of the code that needs to be completed

df.

Area 1

```
.format("com.databricks.spark.sqldw")  
.option("url", connection)  
.option("tempDir", tmpdir)  
.option("forwardSparkAzureStorageCredentials", "true")  
.option("dbTable", tablename)  
.option("checkpointLocation", "/tmp_location")
```

.

Area 2

Which of the following needs to go into Area 1?

- ☐ **start()**
- ☐ **update()**

- 

writeStream

(Poprawne)

- 

write

Wyjaśnienie

When it comes to streaming, we must use the `writeStream()` method.

For more information on using Azure Databricks and Azure Synapse , one can visit the below URL

<https://docs.microsoft.com/en-us/azure/databricks/data/data-sources/azure/synapse-analytics>

Pytanie 10:

Pominięto

Your team currently has the following resources defined on Azure

- 1) An Azure Event Hub namespace and an Event Hub. The Event Hub is used to stream data from an external data source
- 2) An Azure Databricks cluster
- 3) An Azure Synapse Analytics workspace that contains a Dedicated SQL Pool

A Notebook is being developed in Scala in Azure Databricks. The Notebook will be used to take in the streaming data from Azure Event Hubs and add it on a table in the Dedicated SQL Pool

Below is a snippet of the code that needs to be completed

```
df. Area 1  
  
  .format("com.databricks.spark.sqldw")  
  
  .option("url", connection)  
  
  .option("tempDir", tmpdir)  
  
  .option("forwardSparkAzureStorageCredentials", "true")  
  
  .option("dbTable", tablename)  
  
  .option("checkpointLocation", "/tmp_location")  
  
  Area 2
```

Which of the following needs to go into Area 2?

☐

`start()`

(Poprawne)

☐

update()

• 

writeStream

• 

write

Wyjaśnienie

When it comes to streaming, we must use the `start()` method to start the streaming process

For more information on using Azure Databricks and Azure Synapse , one can visit the below URL

<https://docs.microsoft.com/en-us/azure/databricks/data/data-sources/azure/synapse-analytics>

Pytanie 11:

Pominięto

Your company currently has an on-premises environment. Here they have a Microsoft SQL Server that hosts the database for an ecommerce application. They have also setup an Azure subscription and an Azure Synapse Analytics workspace.

You have to implement the following

1) Create an Orders table in the Azure Synapse Analytics pool. The Orders table will have the following schema

Column name	Column Type
OrderID	int
Quantity	int
Price	decimal
OrderDate	int
Region	varchar(100)

2) Have Azure Data Factory transfer transactional data from the on-premises SQL Server to the Orders table on a weekly basis.

3) The Orders table will contain around 2 billion rows initially

4) Queries on the Orders table will be carried out using the OrderID. It needs to be ensured that queries performing joins with the Orders table completes as quickly as possible

Which of the following would you choose as the table distribution for the Orders table?

• ☐

Hash

(Poprawne)

• ☐

Round-Robin

• ☐

Replicated

Wyjaśnienie

Since this is a large table and a Fact-based table, we should choose Hash distribution for the underlying table.

For more information on designing distributed tables , one can visit the below URL

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

Pytanie 12:

Pominięto

Your company currently has an on-premises environment. Here they have a Microsoft SQL Server that hosts the database for an ecommerce application. They have also setup an Azure subscription and an Azure Synapse Analytics workspace.

You have to implement the following

1) Create an Orders table in the Azure Synapse Analytics pool. The Orders table will have the following schema

Column name	Column Type
OrderID	int
Quantity	int
Price	decimal
OrderDate	int
Region	varchar(100)

2) Have Azure Data Factory transfer transactional data from the on-premises SQL Server to the Orders table on a weekly basis.

3) The Orders table will contain around 2 billion rows initially

4) Queries on the Orders table will be carried out using the OrderID. It needs to be ensured that queries performing joins with the Orders table completes as quickly as possible

Which of the following would you choose as the Integration runtime in Azure Data Factory for the pipeline?



Azure Integration Runtime



Self-Hosted Runtime

(Poprawne)



SSIS-Runtime

Wyjaśnienie

Here data needs to be transferred from the on-premises Microsoft SQL Server. Hence you need to install the Integration runtime on the on-premises server itself.

For more information on the self-hosted runtime , one can visit the below URL

<https://docs.microsoft.com/en-us/azure/data-factory/create-self-hosted-integration-runtime?tabs=data-factory>

Pytanie 13:

Pominięto

Your company currently has an on-premises environment. Here they have a Microsoft SQL Server that hosts the database for an ecommerce application. They have also setup an Azure subscription and an Azure Synapse Analytics workspace.

You have to implement the following

1) Create an Orders table in the Azure Synapse Analytics pool. The Orders table will have the following schema

Column name	Column Type
OrderID	int
Quantity	int
Price	decimal
OrderDate	int
Region	varchar(100)

2) Have Azure Data Factory transfer transactional data from the on-premises SQL Server to the Orders table on a weekly basis.

3) The Orders table will contain around 2 billion rows initially

4) Queries on the Orders table will be carried out using the OrderID. It needs to be ensured that queries performing joins with the Orders table completes as quickly as possible

Which of the following would you choose as the distribution column for the Orders table?

- ☒ **OrderID**
(Poprawne)
- ☐ **OrderDate**
- ☐ **Region**

Wyjaśnienie

Here since the JOINS are being performed on the OrderID, we need to ensure this is chosen as the distribution column.

For more information on designing distributed tables , one can visit the below URL

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

Pytanie 14:

Pominięto

Your company currently has an Azure Synapse Analytics workspace. They want to create a table that will hold product-related data. The data will be stored in the SQL data warehouse itself. The table needs to also have a surrogate key named ProductID in place. You have to create the table. Below is the script that you need to complete for this requirement

```
Area 1  dbo.Product  
  
(  
    ProductID INT Area 2 (1,1) NOT NULL,  
    ProductName varchar(200),  
    Price decimal  
)
```

Which of the following will come in Area 1?

- ☒ **CREATE TABLE**
(Poprawne)
- ☐ **CREATE EXTERNAL TABLE**
- ☐ **CREATE VIEW**

Wyjaśnienie

Here we need to create a table since the data needs to reside within the SQL data warehouse itself.

Pytanie 15:

Pominięto

Your company currently has an Azure Synapse Analytics workspace. They want to create a table that will hold product-related data. The data will be stored in the SQL data warehouse itself. The table needs to also have a surrogate key named

ProductID in place. You have to create the table. Below is the script that you need to complete for this requirement

```
Area 1  dbo.Product
(
    ProductID INT Area 2 (1,1) NOT NULL,
    ProductName varchar(200),
    Price decimal
)
```

Which of the following will come in Area 2?

- ☐ SURROGATE
 - ☐ UNIQUE
 - ☐ IDENTITY
- (Poprawne)

Wyjaśnienie

For implementing surrogate keys in Azure Synapse, we can make use of the IDENTITY column.

For more information on using the IDENTITY column , one can visit the below URL

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

Pytanie 16:

Pominięto

Your company has an Azure Data Lake Gen2 storage account. Currently data is loaded onto the storage account with the use of Azure Data Factory. You have to create an Azure Databricks workspace and cluster to process the data in the Azure Data Lake Gen2 storage account. You have to ensure that authentication from the Azure Databricks cluster to the storage account is done via the same Azure AD identity that is used to log into Azure Databricks. You have to also ensure that costs are minimized in the implementation.

Which of the following pricing plan would you choose for the Azure Databricks workspace?

☐

Standard

☐

Premium

(Poprawne)

Wyjaśnienie

Here you need to enable a feature known as credentials passthrough. This can only be achieved with the use of the Premium Pricing plan.

For more information on the credential passthrough feature , one can visit the below URL

<https://docs.microsoft.com/en-us/azure/databricks/security/credential-passthrough/adls-passthrough>

Pytanie 17:

Pominięto

Your company has an Azure Data Lake Gen2 storage account. Currently data is loaded onto the storage account with the use of Azure Data Factory. You have to create an Azure Databricks workspace and cluster to process the data in the Azure Data Lake Gen2 storage account. You have to ensure that authentication from the Azure Databricks cluster to the storage account is done via the same Azure AD identity that is used to log into Azure Databricks. You have to also ensure that costs are minimized in the implementation.

Which of the following must you enable when creating the cluster?

- ☐ Autoscaling
- ☐ Termination
- ☐ Credential passthrough
- ☒ (Poprawne)
- ☐ Spot Instances

Wyjaśnienie

When creating the cluster, you need to enable credential passthrough

Create Cluster

Autopilot Options

- ☒ Enable autoscaling ⓘ
- ☒ Terminate after 120 minutes of inactivity ⓘ

Worker Type ⓘ Min Workers Max Workers

Standard_DS3_v2 14 GB Memory, 4 Cores 2 8 ☐ Spot instances ⓘ

Now Configure separate pools for workers and drivers for flexibility. [Learn more](#)

Driver Type

Same as worker 14 GB Memory, 4 Cores

DBU / hour: 2.25 - 6.75 ⓘ Standard_DS3_v2

▼ Advanced Options

Azure Data Lake Storage Credential Passthrough ⓘ

☒ Enable credential passthrough for user-level data access

Single User Access ⓘ

9f4a5dde-7283-4d8a-92b1-1fdad2bb0c65 551a3618-4f31-4... ⓘ Only one user is allowed to run commands on this cluster when Credential Passthrough is enabled [Learn more](#)

For more information on the credential passthrough feature , one can visit the below URL

<https://docs.microsoft.com/en-us/azure/databricks/security/credential-passthrough/adls-passthrough>

Pytanie 18:

Pominięto

Your team has an Azure Data Lake Gen2 storage account named datalakestore. Data is streamed onto the storage account with the use of Azure Event Hubs. You have to create a lifecycle rule that will be based on the following requirements

- 1) Should be applied to base blobs in a container named data
- 2) Ensure that if objects are not modified since the last 30 days, they are moved to the Cool tier
- 3) Ensure that if objects are not modified since the last 90 days, they are moved to the Archive tier

You have to complete the below lifecycle policy for this requirement

```
{
  "rules": [
    {
      "enabled": true,
      "name": "RuleA",
      "type": "Lifecycle",
      "definition": {
        "actions": {
          "baseBlob": {
            Area 1 : {
              "daysAfterModificationGreaterThan": 30
            },
            Area 2 : {
              "daysAfterModificationGreaterThan": 90
            }
          }
        },
        "filters": {
          "blobTypes": [
            "blockBlob"
          ],
          "prefixMatch": [
            Area 3 ] } } } ]}
```

Which of the following would go into Area 1?

☐

"datalakestore/data"

☐

"data"

☐

"tierToCool"

(Poprawne)

☐

"tierToArchive"

Wyjaśnienie

Here since we need to ensure that if objects are not modified since the last 30 days , they are moved to the Cool Access tier, we need to choose the tier as tierToCool

For more information on Azure Blob storage lifecycle management , one can visit the below URL

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

Pytanie 19:

Pominięto

Your team has an Azure Data Lake Gen2 storage account named `datalakestore`. Data is streamed onto the storage account with the use of Azure Event Hubs. You have to create a lifecycle rule that will be based on the following requirements

- 1) Should be applied to base blobs in a container named `data`
- 2) Ensure that if objects are not modified since the last 30 days, they are moved to the Cool tier
- 3) Ensure that if objects are not modified since the last 90 days, they are moved to the Archive tier

You have to complete the below lifecycle policy for this requirement

```
{
  "rules": [
    {
      "enabled": true,
      "name": "RuleA",
      "type": "Lifecycle",
      "definition": {
        "actions": {
          "baseBlob": {
            Area 1 : {
              "daysAfterModificationGreaterThan": 30
            },
            Area 2 : {
              "daysAfterModificationGreaterThan": 90
            }
          }
        },
        "filters": {
          "blobTypes": [
            "blockBlob"
          ],
          "prefixMatch": [
            Area 3 ] } } ] }
```

Which of the following would go into Area 2?

- ☐ **"datalakestore/data"**
- ☐ **"data"**
- ☐ **"tierToCool"**
- ☐ **"tierToArchive"**

(Poprawne)

Wyjaśnienie

Here since we need to ensure that if objects are not modified since the last 90 days , they are moved to the Cool Access tier, we need to choose the tier as tierToArchive

For more information on Azure Blob storage lifecycle management , one can visit the below URL

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

Pytanie 20:

Pominięto

Your team has an Azure Data Lake Gen2 storage account named `datalakestore`. Data is streamed onto the storage account with the use of Azure Event Hubs. You have to create a lifecycle rule that will be based on the following requirements

- 1) Should be applied to base blobs in a container named `data`
- 2) Ensure that if objects are not modified since the last 30 days, they are moved to the Cool tier
- 3) Ensure that if objects are not modified since the last 90 days, they are moved to the Archive tier

You have to complete the below lifecycle policy for this requirement

```
{
  "rules": [
    {
      "enabled": true,
      "name": "RuleA",
      "type": "Lifecycle",
      "definition": {
        "actions": {
          "baseBlob": {
            Area 1 : {
              "daysAfterModificationGreaterThan": 30
            },
            Area 2 : {
              "daysAfterModificationGreaterThan": 90
            }
          }
        },
        "filters": {
          "blobTypes": [
            "blockBlob"
          ],
          "prefixMatch": [
            Area 3 ] } } } ]}
```

Which of the following would go into Area 3?

☐

"datalakestore/data"

☐

"data"

(Poprawne)

☐

"tierToCool"

☐

"tierToArchive"

Wyjaśnienie

Since we need to apply this to the data container, this should be the filter used in the lifecycle policy.

For more information on Azure Blob storage lifecycle management , one can visit the below URL

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

Pytanie 21:

Pominięto

Your team is designing a data engineering solution for your company. Data is going to be ingested via the use of Azure Stream Analytics. The data will then be written as files in an Azure Data Lake Gen2 storage account. Azure Databricks will be used to analyze the data in the storage account. You have to decide on the file format for the files that will be stored in the storage account. Some of the queries used within Azure Databricks will only target certain columns of data within the files. You have to ensure that the queries fired are optimized for speed and efficiency. Which of the following would you choose as the underlying file format for the files?

- ☐ CSV
 - ☐ TSV
 - ☐ JSON
 - ☐ Parquet
- (Poprawne)

Wyjaśnienie

In such a scenario , Parquet is the ideal file format. It's a columnar-based file format. So querying for just certain columns is very efficient.

For more information on Parquet file format and its benefits , one can visit the below links

<https://databricks.com/glossary/what-is-parquet>

<https://docs.microsoft.com/en-us/azure/databricks/data/data-sources/read-parquet>

Pytanie 22:**Pominięto**

You have to design a table in a Dedicated SQL pool in Azure Synapse Analytics. This will be a Fact table that will contain order information. You need to decide on the optimal number of partitions for the table. The table will contain around 2.1 billion rows. What should be the ideal number of partitions set for the table?

☐

24

☐

36

(Poprawne)

☐

360

☐

4000

Wyjaśnienie

This is based on the below Microsoft documentation link

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

Here for optimal performance, you should have a minimum of 1 million rows per distribution and partition. Since there are 60 distributions by default, you already have 60 million rows. Hence 2.1 billion/60 million will give you 36 partitions.

Pytanie 23:

Pominięto

Your team is planning on using Azure Data Factory to design several ETL workflows. You need to store the pipeline-run data for 60 days. Which of the following would you configure for this requirement?

- ☐ Integration Runtime
- ☐ Diagnostic settings
- ☒ Mapping Data Flow

(Poprawne)

Wyjaśnienie

In order to retain logs for a duration more than 45 days, you need to enable the diagnostic settings for Azure Data Factory.

For more information on monitoring aspects for Azure Data Factory , one can visit the below URL

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

Pytanie 24:

Pominięto

Your team is planning on using Azure Data Factory to design several ETL workflows. You need to store the pipeline-run data for 60 days. Which of the following could you create for this requirement?

☐

Azure SQL Database

☐

Azure Synapse Analytics

☒

Log Analytics workspace

(Poprawne)

Wyjaśnienie

With Diagnostics setting , you can stream the data onto a storage account, Azure Event Hubs or a Log Analytics workspace.

For more information on monitoring aspects for Azure Data Factory , one can visit the below URL

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

Pytanie 25:

Pominięto

Your data engineering team wants to develop a series of Notebooks for data analytics. Which of the below two services have support to develop and debug Notebooks?

- ☐
Azure Data Factory
- ☐
Azure Synapse Analytics
(Poprawne)
- ☐
Azure Databricks
(Poprawne)
- ☐
Azure Stream Analytics

Wyjaśnienie

You can develop and debug Notebooks in Azure Synapse Analytics and Azure Databricks

For more information on using Notebooks in either service , one can visit the below URL

<https://docs.microsoft.com/en-us/azure/synapse-analytics/spark/apache-spark-development-using-notebooks?tabs=classical>

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

Pytanie 26:**Pominięto**

Your company is planning on setting an Azure Synapse Analytics workspace. The workspace will contain a set of SQL data warehouses. The data is going to be ingested into these SQL data warehouse from a data source that has an external IP address of 56.10.2.10. You have to ensure that only this data source can initially connect to the data warehouses. Which of the following would you configure for this requirement?

• ☐

Server-Level IP firewall rule

(Poprawne)

• ☐

Virtual Network service endpoint

• ☐

Transparent Data Encryption

Wyjaśnienie

Here you can configure a Server-Level IP firewall rule to ensure that only a certain IP can create a connection to the data warehouses stored in the Azure Synapse Analytics workspace.

For more information on configuring firewall rules , one can visit the below URL

<https://docs.microsoft.com/en-us/azure/azure-sql/database/firewall-configure>

Pytanie 27:

Pominięto

Your team is currently developing an Azure Data Factory pipeline. You have to design a trigger for the pipeline. The pipeline will be used to ingest data from an Azure Data Lake Gen2 Storage account. Below are the key requirements

- 1) Ensure that data is loaded from the storage account every 20 minutes
- 2) Cater to late arriving events

Which of the following should you define as the trigger type?

New trigger

Name *

Newtrigger

Description

Type *

Schedule

Filter...

Schedule

1

Tumbling window

2

Storage events

3

Custom events

4

Every 15 minute(s)

☐

Schedule

☐

Tumbling

(Poprawne)

☐

Storage events



Custom events

Wyjaśnienie

Here the requirement is to continuously load events based on a schedule. But we need to additional functionality to load late-arriving events. This can be achieved with the use of the Tumbling window trigger.

For more information on the tumbling window trigger , one can visit the below URL

<https://docs.microsoft.com/en-us/azure/data-factory/how-to-create-tumbling-window-trigger?tabs=data-factory>

Pytanie 28:

Pominięto

Your team is currently developing an Azure Data Factory pipeline. You have to design a trigger for the pipeline. The pipeline will be used to ingest data from an Azure Data Lake Gen2 Storage account. Below are the key requirements

- 1) Ensure that data is loaded from the storage account every 20 minutes
- 2) Cater to late arriving events

Which of the following would you configure for the late arriving events?

☐

Delay

(Poprawne)

☐

Max concurrency

☐

Retry policy: count

Wyjaśnienie

Here you need to configure the Delay for late arriving events

For more information on the tumbling window trigger , one can visit the below URL

<https://docs.microsoft.com/en-us/azure/data-factory/how-to-create-tumbling-window-trigger?tabs=data-factory>

Pytanie 29:

Pominięto

Your team needs to develop an Azure Stream Analytics job. The job will ingest data from Azure Event Hubs. Here log data is being ingested. The log data then needs to be matched for records based on a CSV file which will be stored in an Azure Data Lake Gen2 storage account.

Which of the following would you choose as the input type for Azure Event Hubs?

- ☐

Stream

(Poprawne)

- ☐

Reference

Wyjaśnienie

Here we need to add the input type as stream type since the data is being streamed from Azure Event Hubs.

Pytanie 30:

Pominięto

Your team needs to develop an Azure Stream Analytics job. The job will ingest data from Azure Event Hubs. Here log data is being ingested. The log data then needs to be matched for records based on a CSV file which will be stored in an Azure Data Lake Gen2 storage account.

Which of the following would you choose as the input type for Azure Data Lake Gen2 storage account?

☐

Stream

☒

Reference

(Poprawne)

Wyjaśnienie

Here we need to add the input type as reference type since the data is reference data being stored in the Azure Data Lake Gen2 storage account.

For more information on reference data , one can visit the below URL

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

Pytanie 31:

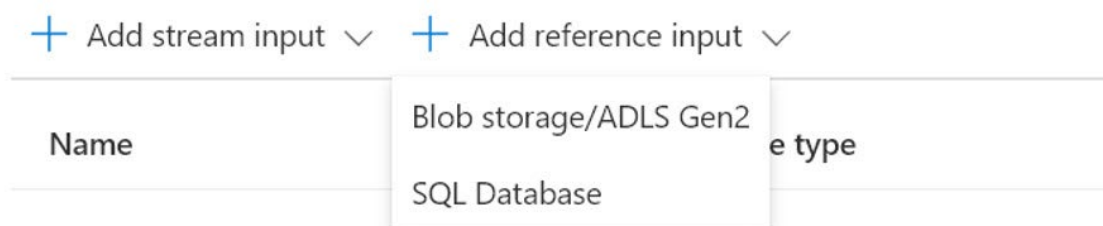
Pominięto

Your team is developing a system that will make use of an Azure Stream Analytics job. You need to add a reference stream to the Stream Analytics job. Which of the following can you add as reference input? Choose 2 answers from the options given below

- ☐
Azure SQL Database
(Poprawne)
- ☐
Azure Data Lake Gen 2 storage accounts
(Poprawne)
- ☐
Azure Synapse Analytics
- ☐
Azure Databricks

Wyjaśnienie

You can add reference input as shown below in an Azure Stream Analytics job



The screenshot shows the 'Add reference input' dropdown menu in the Azure Stream Analytics job configuration. The menu is open, displaying two options: 'Blob storage/ADLS Gen2' and 'SQL Database'. The 'Name' and 'e type' labels are visible on the left and right sides of the dropdown menu respectively.

For more information on reference data , one can visit the below URL

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

Pytanie 32:

Pominięto

You have to create a set of Azure Databricks clusters for various teams. Below are the clusters that need to be created and their requirements

Name	Requirement
Cluster 1	Here the cluster is going to be used by a set of Data Engineers. They would need to run a set of ad-hoc jobs from time to time on the cluster. The cluster should have support to create Notebooks in R and Python.
Cluster 2	Here the cluster should support autoscaling. Here the scaling down should occur is the cluster is underutilized over the last 150 seconds.

Costs should be minimized wherever possible.

Which of the following would you recommend as the cluster type for Cluster 1?

☐

Standard

☐

High Concurrency

(Poprawne)

Wyjaśnienie

When you have a number of users that need to use the cluster and run jobs, you should ideally use a High Concurrency cluster.

For more information on configuring clusters , one can visit the below URL

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

Pytanie 33:**Pominięto**

You have to create a set of Azure Databricks clusters for various teams. Below are the clusters that need to be created and their requirements

Name	Requirement
Cluster 1	Here the cluster is going to be used by a set of Data Engineers. They would need to run a set of ad-hoc jobs from time to time on the cluster. The cluster should have support to create Notebooks in R and Python.
Cluster 2	Here the cluster should support autoscaling. Here the scaling down should occur is the cluster is underutilized over the last 150 seconds.

Costs should be minimized wherever possible.

Which of the following would you choose as the Azure Databricks Plan type to host Cluster 2?

- ☐ Standard
 - ☐ Premium
- (Poprawne)

Wyjaśnienie

Here you need to use optimized autoscaling based on the requirement. For this you need to use the Premium plan when it comes to Azure Databricks.

For more information on configuring clusters , one can visit the below URL

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

Pytanie 34:

Pominięto

Your team is currently having an Azure Synapse Analytics workspace in place. Your company is also using Azure Active Directory for hosting user accounts. You have to ensure that users can use the same Azure AD user accounts to log into the SQL data warehouses defined in Azure Synapse. Which of the following step would you carry out first for this requirement?



Enable the system-managed identity



Set the Active Directory admin

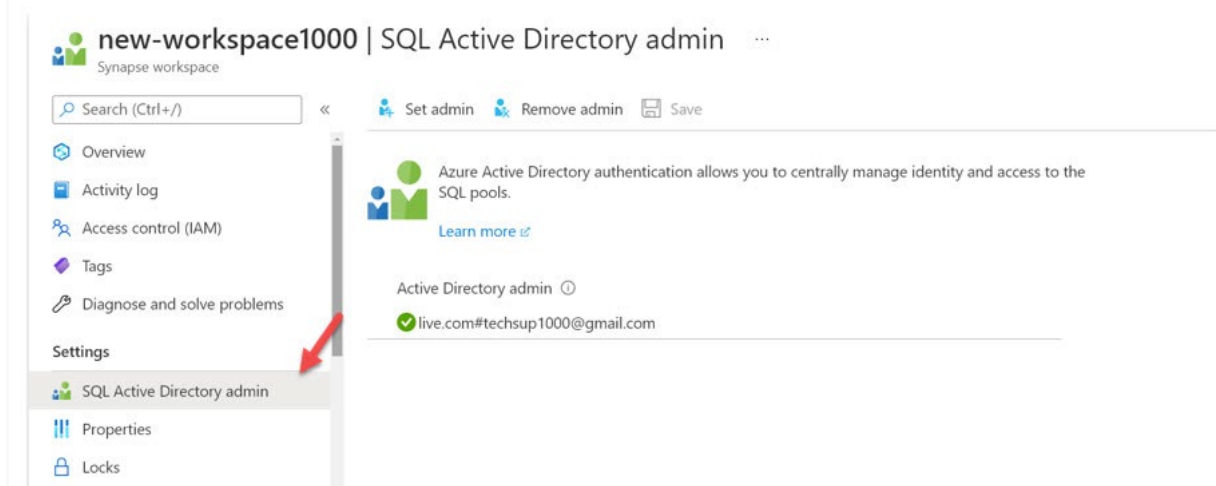
(Poprawne)



Set the Integration settings for the workspace

Wyjaśnienie

To configure Azure AD Authentication for your users in the data warehouse, you first need to set an admin if not already present for Azure Synapse.



For more information on Azure AD Authentication for Azure Synapse , one can visit the below URL

<https://docs.microsoft.com/en-us/azure/azure-sql/database/authentication-aad-configure?tabs=azure-powershell>

Pytanie 35:

Pominięto

Your team is currently having an Azure Synapse Analytics workspace in place. Your company is also using Azure Active Directory for hosting user accounts. You have to ensure that users can use the same Azure AD user accounts to log into the SQL data warehouses defined in Azure Synapse. Which of the following would you need to create in the data warehouse so that users can log in using their Azure Active Directory credentials?

- ☐ **Create a managed identity**
- ☐ **Create a contained user**
- ☒ **Create a new SQL login**

(Poprawne)

Wyjaśnienie

To ensure Azure AD users can log in, you need to create database contained users.

For more information on Azure AD Authentication for Azure Synapse , one can visit the below URL

<https://docs.microsoft.com/en-us/azure/azure-sql/database/authentication-aad-configure?tabs=azure-powershell>

Pytanie 36:

Pominięto

Your team is currently having an Azure Synapse Analytics workspace in place. A set of SQL data warehouses are going to be hosted in the Synapse workspace. The IT Security team wants to gain insights on the activities happening on the SQL data warehouses. Which of the following should be enabled on the workspace for this requirement?

☐

Private endpoint connections

☐

Encryption

☐

Azure SQL Auditing

(Poprawne)

Wyjaśnienie

For this you should enable Azure SQL Auditing. It gives you insights onto the various SQL data warehousing activities.



Save



Discard



Feedback

Azure SQL Auditing

Azure SQL Auditing tracks SQL Pool events and writes them to an audit log in your Azure storage account. [Learn more about Azure SQL Auditing](#) ↗



Azure SQL auditing settings apply only to dedicated SQL pools in this workspace.

Enable Azure SQL Auditing ⓘ



Audit log destination (choose at least one):



Storage



Log Analytics



Event Hub



Turn on Azure Defender for SQL to receive security alerts upon suspicious events.

For more information on SQL Auditing , one can visit the below URL

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

Pytanie 37:

Pominięto

Your team is currently having an Azure Synapse Analytics workspace in place. A set of SQL data warehouses are going to be hosted in the Synapse workspace. The IT Security team wants to gain insights on the activities happening on the SQL data warehouses. Which of the following needs to be additionally created for this requirement?

- ☐ **Azure Log Analytics workspace**
- ☒ **(Poprawne)**
- ☐ **Azure SQL Database**
- ☐ **Azure Databricks**

Wyjaśnienie

Here you need to make use of Azure SQL auditing. And you can stream the audit logs either onto a storage account , a Log Analytics workspace or Azure Event Hubs.



Save



Discard



Feedback

Azure SQL Auditing

Azure SQL Auditing tracks SQL Pool events and writes them to an audit log in your Azure storage account. [Learn more about Azure SQL Auditing](#)



Azure SQL auditing settings apply only to dedicated SQL pools in this workspace.

Enable Azure SQL Auditing



Audit log destination (choose at least one):

☐

Storage

☐

Log Analytics

☐

Event Hub



Turn on Azure Defender for SQL to receive security alerts upon suspicious events.

For more information on SQL Auditing , one can visit the below URL

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

Pytanie 38:

Pominięto

Your team has a Dedicated SQL Pool in Azure Synapse Analytics. You have to create a table named Orders_staging that will be used as a staging table. Around a million rows will be loaded every day into the staging table. You have to ensure that rows can be loaded quickly into the staging table.

You have to also ensure that queries against the OrderID are processed faster.

Complete the following script to designing the best way to achieve this

```
CREATE TABLE [dbo].[Orders_staging]
(
    OrderID int NOT NULL,
    OrderName varchar(1000),
    price decimal,
    quantity int
)
WITH
```

Area 1

```
CREATE Area 2 OrderIDIndex ON [dbo].[Staging] (OrderID);
```

Which of the following would go into Area 1?

☐

HASH DISTRIBUTION

☐

CLUSTERED INDEX

☐

HEAP

(Poprawne)

Wyjaśnienie

Since this is a staging table, the ideal recommendation would be to create the table as a Heap table.

Pytanie 39:

Pominięto

Your team has a Dedicated SQL Pool in Azure Synapse Analytics. You have to create a table named Orders_staging that will be used as a staging table. Around a million rows will be loaded every day into the staging table. You have to ensure that rows can be loaded quickly into the staging table.

You have to also ensure that queries against the OrderID are processed faster.

Complete the following script to designing the best way to achieve this

```
CREATE TABLE [dbo].[Orders_staging]
(
    OrderID int NOT NULL,
    OrderName varchar(1000),
    price decimal,
    quantity int
)
WITH
```

Area 1

```
CREATE [Area 2] OrderIDIndex ON [dbo].[Staging] (OrderID);
```

Which of the following would go into Area 2?

- ☐ CLUSTERED INDEX
- ☐ INDEX
- ☒ (Poprawne)
- ☐ PRIMARY INDEX

Wyjaśnienie

For the heap table , we can use the INDEX command to create a non-clustered index on the OrderID column of the table. This can help improve the query performance.

For more information on Table Indexes , one can visit the below URL

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

Pytanie 40:

Pominięto

You have an Azure Synapse workspace in place. You have to use the Serverless SQL pool to query for data from an Azure Data Lake Gen2 storage account. The file is a JSON-based file. You need to complete the below script for this requirement

```
SELECT
    TOP 20 *
FROM
    Area 1 (
        bulk 'https://storage10000.blob.core.windows.net/data/order.json',
        format = 'csv',
        fieldterminator = '0x0b',
        fieldquote = '0x0b'
    ) with (doc nvarchar(max)) as rows
```

Which of the following would come in Area 1?

☐

WHERE

☐

JOIN

☐

OPENROWSET

(Poprawne)

Wyjaśnienie

Here we can use the OPENROWSET command to query data in JSON-based files.

For more information on querying JSON-based files , one can visit the below URL

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/query-json-files>

Pytanie 41:

Pominięto

You have an Azure Data Lake Gen2 storage account and an Azure Data Factory resource. You need to create a pipeline that will pick up data from the Azure Data Lake Gen2 storage account. You want to ensure that authorization from Azure Data Factory to the storage account is done via Azure Active Directory. Which of the following would you use as the authentication type when setting up the connector?

- ☐ **Access Keys**
- ☐ **Shared Access Signature**
- ☐ **Managed Identity**

(Poprawne)

Wyjaśnienie

You can use Managed Identities to authenticate onto the storage accounts. Here the identities are managed by Azure Active Directory.

For more information on connecting Azure Data Lake Gen2 in Azure Data Factory , one can visit the below URL

<https://docs.microsoft.com/en-us/azure/data-factory/connector-azure-data-lake-storage>

Pytanie 42:

Pominięto

Your team needs to work with an Azure Stream Analytics job. Here the job would ingest data from Azure Event Hubs. The data being ingested is log data from an application. You need to write the query for the job that looks at the growth of CPU percentage log data that is being streamed in a one hour duration. You need to complete the below query for this requirement

```
SELECT metricID,  
       change = CPUpercent -  
       [Area 1] (CPUpercent) OVER (PARTITION BY metricID [Area 2] (hour,1))  
FROM input
```

Which of the following would go into Area 1?

☐

WHERE

☐

LAG

(Poprawne)

☐

LAST

☐

LIMIT DURATION

Wyjaśnienie

Here we need to use the LAG operator to look at the previous CPUpercent reading.

This question is based on the example in the below Microsoft documentation link

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

Pytanie 43:**Pominięto**

Your team needs to work with an Azure Stream Analytics job. Here the job would ingest data from Azure Event Hubs. The data being ingested is log data from an application. You need to write the query for the job that looks at the growth of CPU percentage log data that is being streamed in a one hour duration. You need to complete the below query for this requirement

```
SELECT metricID,  
       change = CPUpercent -  
       Area 1 (CPUpercent) OVER (PARTITION BY metricID Area 2 (hour,1))  
FROM input
```

Which of the following would go into Area 2?

- ☐ **WHERE**
- ☐ **LAG**
- ☐ **LAST**
- ☐ **LIMIT DURATION**

(Poprawne)

Wyjaśnienie

Since we need to see the data over an hour, we can use the LIMIT DURATION within the SELECT statement

This question is based on the example in the below Microsoft documentation link

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

Pytanie 44:

Pominięto

Your team has workloads running in a Dedicated SQL Pool in Azure Synapse Analytics. You need to find the top 5 longest running queries. Which of the following view can help to get this information?

☐

`sys.dm_pdw_exec_requests`

(Poprawne)

☐

`sys.dm_pdw_request_steps`

☐

`sys.dm_pdw_dms_workers`

Wyjaśnienie

You can get this information from the `sys.dm_pdw_exec_requests` view

For more information on monitoring your SQL data warehouse , one can visit the below URL

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

Pytanie 45:

Pominięto

Your team has an Azure Databricks workspace. You need to create a local table from a DataFrame named df. Which of the following statement can be used to create the local table?

- ☐ `df.write.saveAsTable("Orders")`
- ☐ `df.createOrReplaceTempView.saveAsTable("Orders")`
- ☒ **(Poprawne)**
- ☐ `df.createLocalTable("Orders")`

Wyjaśnienie

You can create a local table with the method of `createOrReplaceTempView.saveAsTable`

For more information on creating tables in Azure Databricks , one can visit the below URL

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