

DP-200

Question1 Implement non-relational data stores – D -

Question1.1

You are a data engineer for your company. Your company wants to build an e-commerce site that allows vendors to sell their products online. Each vendor's product can have an arbitrary number of attributes. For example, Vendor A has a product named Bicycle Z that contains attributes Price, Description, Brand, PartNumber, and Size. Vendor B has the same product named Bicycle Z that contains attributes Price, Description, Weight, and Model. You want to implement a solution that does not restrict the product attributes that are used by each vendor. Your solution must allow you to use .NET to query product data.

You need to create the appropriate data store.

Solution: You create a Cosmos DB account that uses the Table API.

Does this solution meet the goal?

No

Yes

Explanation

This solution does meet the stated goal. Cosmos DB is a non-relational data store that allows you to query data by using one of five APIs. The Table API allows you to use OData and Language Integrated Query (LINQ) to query data. You can issue LINQ queries with .NET.

References

[Introduction to Azure Cosmos DB: Table API](#)

[Welcome to Azure Cosmos DB](#)

Question1.2

You are a data engineer for your company. Your company wants to build an e-commerce site that allows vendors to sell their products online. Each vendor's product can have an arbitrary number of attributes. For example, Vendor A has a product named Bicycle Z that contains attributes Price, Description, Brand, PartNumber, and Size. Vendor B has the same product named Bicycle Z that contains attributes Price, Description, Weight, and Model. You want to implement a solution that does not restrict the product attributes that are used by each vendor. Your solution must allow you to use .NET to query product data.

You need to create the appropriate data store.

Solution: You create an Azure SQL Database account that uses a managed instance.

Does this solution meet the goal?

No

Yes

Explanation

This solution does not meet the stated goal.

An Azure SQL Database account is Microsoft's Platform-as-a-Service (PaaS) offering equivalent to on-premises SQL Server. With Azure SQL Database, you do not have to manage the physical infrastructure for SQL Server. However, Azure SQL Database is a relational database. This means that every row in each table must have identical columns (or attributes). In this scenario, you need to allow varying attributes. The managed instance deployment option is useful if you have an on-premises SQL Server instance with multiple databases that must all be moved to the cloud. All databases in a managed instance deployment share the same resources.

References

[What is Azure SQL Database service](#)

Question1.3

You are a data engineer for your company. Your company wants to build an e-commerce site that allows vendors to sell their products online. Each vendor's product can have an arbitrary number of attributes. For example, Vendor A has a product named Bicycle Z that contains attributes Price, Description, Brand, PartNumber, and Size. Vendor B has the same product named Bicycle Z that contains attributes Price, Description, Weight, and Model. You want to implement a solution that does not restrict the product attributes that are used by each vendor. Your solution must allow you to use .NET to query product data.

You need to create the appropriate data store.

Solution: You create a table storage account.

Does this solution meet the goal?

No

Yes

Explanation

This solution does meet the stated goal. Azure Table storage uses NoSQL, which allows you to store keys and attributes in a schemaless data store. This is similar to Cosmos DB with the Table API. Each entity (row) can store a varying number of attributes (fields). This allows different vendors to upload products with varying attributes. You can also use .NET to query the data.

References

[Azure Table storage overview](#)

Question1.4

You are a data engineer for your company. Your company wants to build an e-commerce site that allows vendors to sell their products online. Each vendor's product can have an arbitrary number of attributes. For example, Vendor A has a product named Bicycle Z that contains attributes Price, Description, Brand, PartNumber, and Size. Vendor B has the same product named Bicycle Z that contains attributes Price, Description, Weight, and Model. You want to implement a solution that does not restrict the product attributes that are used by each vendor. Your solution must allow you to use .NET to query product data.

You need to create the appropriate data store.

Solution: You create a Cosmos DB account that uses the SQL API.

Does this solution meet the goal?

No

Yes

Explanation

This solution does meet the stated goal. Cosmos DB is a non-relational data store that allows you to query data by using one of five APIs. The SQL API allows you to use SQL to query data as JavaScript Object Notation (JSON) documents. JSON documents allow you to work with tree-shaped data instead of rows and columns, allowing you to determine the structure of results dynamically. You can use .NET to query Cosmos DB data that uses the SQL API.

References

[Welcome to Azure Cosmos DB](#)

[Getting started with SQL queries](#)

Question5 Implement non-relational data stores

Case Study

Complete the Case Study

- Overview

Company1 is a food delivery company with a global presence. Company1 data solutions use the following Azure data products:

- Azure Blob Storage to store order tracking status
- Azure Cosmos DB to store orders, restaurant (menu, address, contact info) and customer data
- Azure SQL Database elastic pool used by the company's internal systems

You need to support plans to release this software in a new Azure region.

- Requirements

- The platform must be resilient in case of an outage impacting one Azure availability zone.
- You need to provide a mobile application secure access to the orders database that is hosted in Azure Cosmos DB.
- The mobile application needs to perform read and write queries directly to the orders database.
- The internal systems use 15 separate databases configured in an elastic pool.
- The internal system databases must be readable in case of availability zone outage.
- In this new region deployment, you need to provide an external partner of Company1 direct read access to the tracking order history that is stored in blob storage.
- Access to Company1 by the external partner should be limited and should automatically expire after two months.

- Orders Database

During the first marketing campaign of the previous regional launch, the database was the cause of a bottleneck while reading the orders collection. The Request Units (RU) database throughput needed to be increased in order to eliminate the bottleneck. Collections different from orders throughput did not change during the marketing campaign.

- Internal system databases

You need to implement a daily administrative task on all databases. This administrative task is a script written in PowerShell.

Question 5.1

You need to implement replication so that the internal system databases meet the requirements.

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

Configure SQL Data Sync for each database.

Create an Azure SQL Server elastic pool in another region.

Enable geo-replication for each database.

Create a Sync Group in elastic pool.

Explanation

You should create an Azure SQL Server elastic pool in another region. This server is used as a secondary location to store the read replicas in case of an Azure availability zone outage.

You should also enable geo-replication for each database. You should create a secondary read replica for each database to start the replication process in the secondary elastic pool.

You should not create a sync group in elastic pool and configure SQL Data Sync for each database. Azure SQL Data Sync is a solution that lets you synchronize data bi-directionally across multiple SQL Databases. This is not the preferred solution for disaster recovery scenarios.

- References
- [Creating and using active geo-replication](#)
- <https://docs.microsoft.com/en-us/azure/azure-sql/database/active-geo-replication-overview>
- [Sync data across multiple cloud and on-premises databases with SQL Data Sync](#)
- <https://docs.microsoft.com/en-us/azure/azure-sql/database/sql-data-sync-data-sql-server-sql-database>

Question 5.2-

you need to implement the redundancy strategy for order tracking status storage to meet the requirements. Costs need to be kept to a minimum.

Which data redundancy strategy and storage account type should you use? To answer, select the appropriate options from the drop-down menus.

You have been given the task to implement the redundancy strategy for order tracking status storage to meet the appropriate requirements. The costs should have to be kept as minimum as possible.

What data redundancy strategy and storage account type should you be using? Select the relevant options to answer the question.

Data redundancy strategy:

Storage account type:

Data redundancy strategy:

ZRS

Storage account type:

V2

Explanation

You should use ZRS. This redundancy strategy replicates data in other availability zones in the same Azure region.

You should use a general-purpose v2 account type. This account type incorporates all general-purpose v1 features, including blob storage, and delivers the lowest per-gigabyte capacity prices for Azure storage.

You should not use GRS. This redundancy strategy will replicate data to another Azure region. This increases costs unnecessarily because replication in another availability zone meets the requirements.

You should not use LRS. This redundancy level will expose the application in case of an Azure outage in a specific availability zone.

You should not use Block blob storage accounts. This is a specialized account type used to store block blobs and appends blobs, providing low latency and higher transaction rates. Block blob storage accounts only support premium tiers, which are more expensive than general-purpose v2 account types.

References

[Azure Storage redundancy](#)

[Azure storage account overview](#)

Question 5.3-

You need to implement daily administrative tasks for the internal system databases. You need to minimize administrative efforts.

Which solution should you use?

SQL Server Agent

Azure Alerts action group

Elastic Database Jobs

Azure Data Factory

Explanation

You should use Elastic Database Jobs. This solution allows you to run jobs against all the databases in the elastic pool and supports PowerShell scripts. The job definition is stored in a job database, and the internal system databases in the elastic pool are configured in a target group.

You should not use SQL Server Agent. This solution allows you to run administrative tasks in a specific database. However, SQL Server Agent is only supported in on-premises SQL Server instances or Azure SQL managed instances.

You should not use Azure Data Factory. This solution orchestrates data workflows to move data across different data stores via data pipelines. You can create scheduled pipelines to run on a regular basis. However, Azure Data Factory is not suited for administrative tasks targeting Azure SQL elastic pool databases.

You should not use an Azure Alerts action group. An action group is a collection of notification preferences used by Azure Monitor to respond to an alert. These notification preferences could also include automated remediation actions, like executing an Azure function or running an automated runbook written in PowerShell. An action group should be attached to an alert. You need to manually create all the connection and target servers logic in a runbook script, which increases administrative efforts.

References

[Automate management tasks using database jobs](#)

[Create, configure, and manage elastic jobs](#)

[SQL Server Agent](#)

[What is Azure Data Factory?](#)

[Create and manage action groups in the Azure portal](#)

Question 5.4-

You need to create the necessary credentials for the mobile client application to connect with Azure Cosmos DB.

Which credential key and permission should you use to to meet security requirements? To answer, select the appropriate options from the drop-down menus.

Credential key:

Permission:

Resource Token

All

Explanation

You should use a resource token as the credential key. Resource tokens provide granular access to Cosmos DB limiting access to administrative tasks. This is a safe alternative for clients that cannot be trusted with a master key, like a mobile application.

You should use all as the permission. The mobile application needs permission to write and read data in Cosmos DB. A resource token can be used with all permissions in a specific database.

You should not use a master key as the credential key. Master keys provide full access to your database account, allowing access to all administrative tasks. You should not set granular permissions for a master key. This should be used only in trusted clients, which is not the case with a mobile client application.

You should not use read as the permission. The mobile application needs permission to write and read data in Cosmos DB. A read permission does not allow the mobile application to write in Cosmos DB.

You should not use an Azure AD user as the credential key. Azure AD user is used to read data in Cosmos DB from Azure Portal and Azure Storage Explorer. **You cannot use Azure AD user to authenticate from a mobile application.**

You should not use a Cosmos DB Operator role. This role is used to grant permission to manage Cosmos DB accounts. However, this role does not grant access to account keys and connection strings, preventing data access.

References

[Secure access to data in Azure Cosmos DB](#)

[Role-based access control in Azure Cosmos DB](#)

[Work with data using Azure Storage Explorer](#)

Question 5.5-

You need to grant the external partner of Company1 access to the blob containers that store the order tracking status.

Which authorization method should you use to meet security requirements?

Anonymous public read access
Shared Key authorization

Shared access signatures (SAS)

Azure Active Directory (Azure AD) role-based access control (RBAC)

Explanation

You should use SAS. SAS delegates access to blob containers in a storage account with granular control over how the client accesses your data. You can define a SAS token to allow access only to a specific blob container with a defined expiration.

You should not use Shared Key authorization. Shared Key authorization gives full administrative access to storage accounts, granting the external partner more access than necessary. Shared keys could be regenerated, but they do not expire automatically in two months.

You should not use anonymous public read access. This will give read access to everyone who knows the blob container URL, exposing the data unnecessarily.

You should not implement Azure AD RBAC. You could authorize a principal in Azure AD with an OAuth 2.0 token that allows access to data in the blob container storage and a granular RBAC role that only allows access to the blob container data. However, access will only expire if it is manually revoked.

References

[Authorizing access to Azure Storage](#)

[Grant limited access to Azure Storage resources using shared access signatures \(SAS\)](#)

[Authorize access to Azure blobs and queues using Azure Active Directory](#)

[Authorize with Shared Key](#)

[Manage anonymous read access to containers and blobs](#)

Question10 Manage data security

Case Study

Instructions

This case study contains a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some

question sets might have more than one correct solution, while others might not have a correct solution.

Note: You cannot go back or review questions of this type on the actual certification exam.

Question 10.1-

You manage an Azure SQL Database containing a column with sensitive data. The column is used for joins by an application.

You need to configure encryption for this database to protect the column.

Solution: You configure Always Encrypted with a randomized type.

Does the solution meet the goal?

No

Yes

Explanation

This solution does not meet the goal. Configuring Always Encrypted with a randomized type will encrypt the column with random generated values, preventing join operations, grouping, and indexing.

References

[Always Encrypted \(Database Engine\)](#)

Question 10.2-

You manage an Azure SQL database containing a column with sensitive data. The column is used for joins by an application.

You need to configure encryption for this database to protect the column.

Solution: You configure Always Encrypted with a deterministic type.

Does the solution meet the goal?

No

Yes

Explanation

This solution meets the goal. Configuring Always Encrypted with a deterministic type will encrypt the column and allow it to be used in join queries. Deterministic encryption generates the same value for a given input value, allowing grouping and indexing involving this column.

References

[Always Encrypted \(Database Engine\)](#)

Question 10.3-

You manage an Azure SQL database containing a column with sensitive data. The column is used for joins by an application.

You need to configure encryption for this database to protect the column.

Solution: You configure dynamic data masking (DDM) with random masking.

Does the solution meet the goal?

No

Yes

Explanation

This solution does not meet the goal. You could use DDM to limit data exposure to users by masking the field value, but it does not actually encrypt the column.

References

[Dynamic Data Masking](#)

Question 10.4-

You manage an Azure SQL database containing a column with sensitive data. The column is used for joins by an application.

You need to configure encryption for this database to protect the column.

Solution: You configure dynamic data masking (DDM) with partial masking.

Does the solution meet the goal?

No

Yes

Explanation

This solution does not meet the goal. You could use partial data masking to limit data exposure to users by partially masking the field value, but it does not actually encrypt the column.

References

[Dynamic Data Masking](#)

Question14 - Develop batch processing solutions

Case Study

Instructions

This case study contains a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

Note: You cannot go back or review questions of this type on the actual certification exam.

Question 14.1-

You keep Azure Storage access keys in Azure Key Vault.

You need to configure a reference to those keys from within Azure Databricks to enable secure access to Azure Blob Storage.

Solution: You create a secret scope using the Databricks CLI (version 0.7.1 and above).

Does this solution meet the goal?

Yes

No

Explanation

This solution does not meet the goal.

The Databricks CLI can be used to create only Databricks-backed secret scopes, for example when the keys are stored in an encrypted database owned and managed by Azure Databricks.

Currently, the setup of an Azure Key Vault-backed secret scope is supported only in Azure Databricks UI.

References

[Databricks CLI – Databricks Documentation](#)

[Tutorial: Access Azure Blob Storage from Azure Databricks using Azure Key Vault](#)

Question 14.2-

You keep Azure Storage access keys in Azure Key Vault.

You need to configure a reference to those keys from within Azure Databricks to enable secure access to Azure Blob Storage.

Solution: You create a secret scope using the Secrets API via the Azure Databricks 2.0/secrets/scopes/create endpoint.

Does this solution meet the goal?

Yes

No

Explanation

This solution does not meet the goal.

The Secrets API can be used to create only Databricks-backed secret scopes, for example when the keys are stored in an encrypted database owned and managed by Azure Databricks.

Currently, the setup of an Azure Key Vault-backed secret scope is supported only in the Azure Databricks UI.

References

[Secrets API – Databricks Documentation](#)

[Tutorial: Access Azure Blob Storage from Azure Databricks using Azure Key Vault](#)

Question 14.3-

You keep Azure Storage access keys in Azure Key Vault.

You need to configure a reference to those keys from within Azure Databricks to enable secure access to Azure Blob Storage.

Solution: You open Azure Databricks workspace from Azure portal, add #secrets/createScope to its URL, and fill in all the details to create the secret scope.

Does this solution meet the goal?

Yes

No

Explanation

The solution meets the goal.

To refer to the secret keys stored in Azure Key Vault and use them for secure access, for example to mount the Azure Blob Storage container, you need to create a secret scope in Azure Databricks. Currently, the setup of an Azure Key Vault-backed secret scope is supported only in the Azure Databricks UI.

References

[Tutorial: Access Azure Blob Storage from Azure Databricks using Azure Key Vault](#)

Question17 Develop batch processing solutions

Case Study

Instructions

This case study contains a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

Note: You cannot go back or review questions of this type on the actual certification exam.

Question 17.1-

You are a data engineer. You are developing a data ingestion solution that ingests data from large pipe-delimited text files in an Azure Data Lake Gen 1 storage account to Azure Data Warehouse.

You need to load the data.

Solution:

You do the following:

- Create an external file format and external data source.
- Create an external table that uses the external data source.
- Load the data from the external table.

Does this solution meet the goal?

Yes

No

Explanation

This solution does meet the stated goal. To import data into SQL Data Warehouse, you should first create an external file format by using the CREATE EXTERNAL FILE FORMAT statement. This defines the type of file that represents the source data. Next, you should create an external data source by using the CREATE EXTERNAL DATA SOURCE statement. This specifies the location and credentials to the Azure Data Lake Gen 1 storage account. Then you should create an external table by using the CREATE EXTERNAL TABLE statement. This defines the table fields, specifies its location in the storage account, and the file format that you created. Finally, you should load data into the table by using CREATE TABLE AS SELECT, which allows you to write a query that selects data from the source file and place it in a new table.

References

[Load data from Azure Data Lake Storage to SQL Data Warehouse](#)

Question 17.2-

You are a data engineer. You are developing a data ingestion solution that ingests data from large pipe-delimited text files in an Azure Data Lake Gen 1 storage account to Azure Data Warehouse.

You need to load the data.

Solution:

You do the following:

- Create an Azure Databricks account and a linked server.
- Create an external table that points to the Azure Databricks account.
- Load the data by running the dbutils.fs.cp command.

Does this solution meet the goal?

Yes

No

Explanation

This solution does not meet the goal. Azure Databricks uses Spark clusters to execute code in notebooks. You should not use `sp_addlinkedserver` to connect to a Databricks account. This stored procedure allows you to connect to other SQL Server instances. The `dbutils.fs.cp` command allows you to copy files in Databricks. Because you should not use Databricks, you should not run this command.

References

[Load data from Azure Data Lake Storage to SQL Data Warehouse](#)

[sp_addlinkedserver \(Transact-SQL\)](#)

[Databricks Utilities](#)

Question 17.3-

You are a data engineer. You are developing a data ingestion solution that ingests data from large pipe-delimited text files in an Azure Data Lake Gen 1 storage account to Azure Data Warehouse.

You need to load the data.

Solution:

You do the following:

- Create an Azure Cosmos DB account and a linked server.
- Create an external table that points to the Azure Cosmos DB account.
- Load the data by running the BULK IMPORT statement.

Does this solution meet the goal?

Yes

No

Explanation

This solution does not meet the goal. Azure Cosmos DB is a multi-model, non-relational database that uses one of five APIs: SQL, Table, Cassandra, MongoDB, and Gremlin. You should not use `sp_addlinkedserver` to connect to a Cosmos DB account. This stored procedure allows you to connect to other SQL Server instances. `BULK IMPORT` does allow you to bulk import data, but this command cannot import data from a Cosmos DB account.

References

[Load data from Azure Data Lake Storage to SQL Data Warehouse](#)

[Welcome to Azure Cosmos DB](#)

[sp_addlinkedserver \(Transact-SQL\)](#)

Question20 Develop streaming solutions - D

Case Study

Instructions

This case study contains a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

Note: You cannot go back or review questions of this type on the actual certification exam.

Question 20.1-

You are using Azure Stream Analytics (ASA) to process real-time defect detection events from your factory's production lines.

You need to select the right windowing function in your ASA job's `SELECT` query, so that you can:

- Group events per line if they occur within specific time intervals between each other, but not exceeding maximum duration time set for the window.
- Filter out periods of time when no defects are reported.
- Count each event only once.

Solution: In the ASA job query, you group events by using the session window.

Does this solution meet the goal?

No

Yes

Explanation

This solution meets the goal.

Session windows begin when the defect detection event occurs, and they continue to extend, including new events occurring within the set time interval (timeout). If no further events are detected, then the window will close. The window will also close if the maximum duration parameter is set for the session window, and then a new session window may begin.

The session window option will effectively filter out periods of time where no events are streamed. Each event is only counted once.

References

[Introduction to Stream Analytics windowing functions](#)

[Session window \(Azure Stream Analytics\)](#)

Question 20.2-

You are using Azure Stream Analytics (ASA) to process real-time defect detection events from your factory's production lines.

You need to select the right windowing function in your ASA job's SELECT query, so that you can:

- Group events per line if they occur within specific time intervals between each other, but not exceeding maximum duration time set for the window.
- Filter out periods of time when no defects are reported.

- Count each event only once.

Solution: In the ASA job query, you group events by using the tumbling window.

Does this solution meet the goal?

No

Yes

Explanation

This solution does not meet the goal.

Tumbling windows are a series of fixed-sized, non-overlapping and contiguous time intervals. Each event is only counted once. However, they do not check the time duration between events and do not filter out periods of time when no events are streamed.

To meet the goal, you should use Session windows, which begin when the defect detection event occurs and continue to extend, including new events occurring within the set time interval (timeout). If no further events are detected, the window will close. The window will also close if the maximum duration parameter is set for the session window, and then a new session window may begin.

The Session window option will effectively filter out periods of time where no events are streamed. Each event is only counted once.

References

[Introduction to Stream Analytics windowing functions](#)

[Tumbling Window \(Azure Stream Analytics\)](#)

[Session window \(Azure Stream Analytics\)](#)

Question 20.3-

You are using Azure Stream Analytics (ASA) to process real-time defect detection events from your factory's production lines.

You need to select the right windowing function in your ASA job's SELECT query, so that you can:

- Group events per line if they occur within specific time intervals between each other, but not exceeding maximum duration time set for the window.
- Filter out periods of time when no defects are reported.
- Count each event only once.

Solution: In the ASA job query, you group events by using the hopping window.

Does this solution meet the goal?

No

Yes

Explanation

This solution does not meet the goal.

Hopping windows are a series of fixed-sized and contiguous time intervals. They hop forward by a specified fixed time. If the hop size is less than a size of the window, hopping windows overlap, and that is why an event may be part of several windows. Hopping windows do not check the time duration between events and do not filter out periods of time when no events are streamed.

To meet the goal, you should use Session windows, which begin when the defect detection event occurs and continue to extend, including new events occurring within the set time interval (timeout). If no further events are detected, the window will close. The window will also close if the maximum duration parameter is set for the session window, and then a new session window may begin.

The Session window option will effectively filter out periods of time where no events are streamed. Each event is only counted once.

References

[Introduction to Stream Analytics windowing functions](#)

[Hopping Window \(Azure Stream Analytics\)](#)

[Session window \(Azure Stream Analytics\)](#)

Question 20.4-

You are using Azure Stream Analytics (ASA) to process real-time defect detection events from your factory's production lines.

You need to select the right windowing function in your ASA job's SELECT query, so that you can:

- Group events per line if they occur within specific time intervals between each other, but not exceeding maximum duration time set for the window.
- Filter out periods of time when no defects are reported.
- Count each event only once.

Solution: In the ASA job query, you group events by using the sliding window.

Does this solution meet the goal?

Yes

No

Explanation

This solution does not meet the goal.

Sliding windows are a series of fixed-sized and contiguous time intervals. They produce output only when an event occurs, so you can filter out periods of times where no events are streamed. However, they may overlap and that is why an event may be included in more than one window. Sliding windows also do not check the time duration between events.

To meet the goal, you should use Session windows, which begin when the defect detection event occurs and continue to extend, including new events occurring within the set time interval (timeout). If no further events are detected, the window will close. The window will also close if the maximum duration parameter is set for the session window, and then a new session window may begin.

The session window option will effectively filter out periods of time where no events are streamed. Each event is only counted once.

References

[Introduction to Stream Analytics windowing functions](#)

[Sliding Window \(Azure Stream Analytics\)](#)

Question 24 Develop streaming solutions

Case Study

Instructions

This case study contains a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

Note: You cannot go back or review questions of this type on the actual certification exam.

Question 24.1-

You are a data engineer for an autonomous vehicle manufacturer. Each vehicle contains a transmitter that submits sensor data over Advanced Message Queuing Protocol (AMQP). You want to retrieve the sensor data in real time so that you can extract relevant information, transform it, and then send it to Power BI.

You need to implement the solution.

Solution:

You do the following:

- Create an Event Hub instance.
- Create a Stream Analytics job that uses a query to extract data.

Does this solution meet the goal?

Yes

No

Explanation

This solution meets the goal. Event Hubs is an Azure resource that allows you to stream big data to the cloud. It accepts streaming data over HTTPS and AMQP. A Stream Analytics job can read data from Event Hubs and store the transformed data in a variety of output data sources, including Power BI.

References

[Analyze phone call data with Stream Analytics and visualize results in Power BI dashboard](#)

Question 24.2-

You are a data engineer for an autonomous vehicle manufacturer. Each vehicle contains a transmitter that submits sensor data over Advanced Message Queuing Protocol (AMQP). You want to retrieve the sensor data in real time so that you can extract relevant information, transform it, and then send it to Power BI.

You need to implement the solution.

Solution:

You do the following:

- Create an IoT Hub instance.
- Create a Stream Analytics job that uses a query to extract data.

Does this solution meet the goal?

Yes

No

Explanation

This solution meets the goal. IoT Hub is an Azure resource that allows you to stream big data to the cloud. It supports per-device provisioning. It accepts streaming data over HTTPS, AMQP, and Message Queue Telemetry Transport (MQTT). A Stream Analytics job can read data from Event Hubs and store the transformed data in a variety of output data sources, including Power BI.

References

[Get started with Azure Stream Analytics to process data from IoT devices](#)

[Analyze phone call data with Stream Analytics and visualize results in Power BI dashboard](#)

Question 24.3-

You are a data engineer for an autonomous vehicle manufacturer. Each vehicle contains a transmitter that submits sensor data over Advanced Message Queuing Protocol (AMQP). You want to retrieve the sensor data in real time so that you can extract relevant information, transform it, and then send it to Power BI.

You need to implement the solution.

Solution:

You do the following:

- Create an Azure Databricks instance.
- Create an Azure Automation runbook that extracts and queries data from Databricks.

Does this solution meet the goal?

No

Yes

Explanation

This solution does not meet the goal. Azure Databricks is an analytics platform that uses Spark clusters. Notebooks contain code that run on clusters of nodes. Automation runbooks execute code written in PowerShell and Python. The problem with this solution is that sensor data cannot be streamed to either of these resources.

References

[Azure Automation runbook types](#)

[What is Azure Databricks?](#)

Question 24.4-

You are a data engineer for an autonomous vehicle manufacturer. Each vehicle contains a transmitter that submits sensor data over Advanced Message Queuing Protocol (AMQP). You want to retrieve the sensor data in real time so that you can extract relevant information, transform it, and then send it to Power BI.

You need to implement the solution.

Solution:

You do the following:

- Create an Azure Relay service.
- Create an Azure Function app that extracts and queries data from Azure Relay.

Does this solution meet the goal?

Yes

No

Explanation

This solution does not meet the goal. Azure Relay allows client applications to connect to services hosted on a private network over the internet. An Azure Function app contains one or more functions that are exposed over HTTP. These functions can be invoked by triggers or on a schedule. Neither Azure Relay nor Azure Function apps accept messages over AMPQ.

References

[An introduction to Azure Functions](#)

[What is Azure Relay?](#)

Question28 Develop streaming solutions

Case Study

Complete the Case Study

• Overview

Company A delivers exams to test centers around the world. Each test center uses a web application that displays exam content. Exam content is retrieved from a web API hosted as an Azure App Service.

• Problem

It appears that some test centers are passing substantially more candidates than others. An investigation is underway to determine whether or not some test centers are allowing candidates to cheat on exams. It appears that the suspected cheating also occurs around the same time of day. You must be able to determine the number of pass results that occur at a test center within 20 minutes of each other.

- Solution

The IT Director wants you to implement a solution that sends exam data to Azure as soon as the candidate completes the exam. Data includes the candidate's name, ID, test center number, exam number, score, date, time, and length of exam. You must have this data sent to Power BI so that the business investigator can determine whether or not a physical investigation at the test centers in question should be made. Because you are proficient in SQL, you want to use a solution that allows you to take advantage of your SQL skills. You also want to provide test data for analysis before the real data is received.

- Question 1
- Question 2
- Question 3
- Question 4
- Question 5

Question 28.1-

You need to have the test center web application send exam data to Azure.

Which technology should you choose to receive the data?

Azure Relay

Azure Databricks

Event Grid

Event Hub

Explanation

You should use Event Hub. Event Hub is an Azure resource that accepts streaming telemetry data from other sources. It is basically a big data pipeline. It allows you to capture, retain, and replay telemetry data, which in this case is candidate exam data.

You should not use Event Grid. Event Grid is a publish-subscribe platform for events. Event publishers send the events to Event Grid. Subscribers subscribe to the events they want to handle.

You should not use Azure Relay. Azure Relay allows client applications to access on-premises services through Azure.

You should not use Azure Databricks. Databricks is a technology that allows you to ingest and analyze data.

References

[Choosing a real-time message ingestion technology in Azure](#)

[Choosing a stream processing technology in Azure](#)

[Choose between Azure messaging services - Event Grid, Event Hubs, and Service Bus](#)

[What is Azure Databricks?](#)

Question 28.2-

You need to choose the technology to query the data, filter it, and send it to Power BI.

Which technology should you choose?

WebJob

HDInsight

Function app

Stream Analytics

Explanation

You should choose Stream Analytics. Stream Analytics allows you to define an input data source, a query, and an output data source. The input data source can be an event hub, an IoT hub, or blob storage. The output data source in this scenario is Power BI. The query is a SQL-like query language. This allows you to take advantage of your existing skills.

You should not use HDInsight. HDInsight is a streaming technology that allows you to use C#, F#, Java, Python, and Scala. It does not allow you to use a SQL-like language.

You should not use a WebJob. A WebJob runs in the context of an Azure App Service app. It can be invoked on a schedule or by a trigger. You can use C#, Java, Node.js, PHP, Python to implement WebJobs. However, you cannot use a SQL-like language.

You should not use a function app. A function app is similar to a WebJob in that it can be invoked on a schedule or by a trigger. You can use many different languages to create a function in a function app. However, you cannot use a

SQL-like language.

References

[What is Apache Hadoop in Azure HDInsight?](#)

[Run Background tasks with WebJobs in Azure App Service](#)

[An introduction to Azure Functions](#)

[Analyze phone call data with Stream Analytics and visualize results in Power BI dashboard](#)

[What is Azure Stream Analytics?](#)

Question 28.3-

You need to choose the windowing function that is most appropriate.

Which windowing function should you use?

Sliding

Hopping

Tumbling

Session

Explanation

You should use the session windowing function. Windowing functions are native to Stream Analytics, which is what you should use to analyze the data. The session windowing function allows you to group streaming events that arrive at a similar time, and filter out time periods where no data exists. In this scenario, you want to determine the time when the suspected cheating occurs. Specifically, you want to determine the number of pass results that occur at a test center within 20 minutes of each other.

You should not use the tumbling windowing function. This function allows you to segment data into distinct time segments. This does not help in this scenario.

You should not use the hopping windowing function. A hopping window function looks backwards to determine when an event occurs.

You should not use the sliding windowing function. This function produces output only when an event occurs.

References

[Introduction to Stream Analytics windowing functions](#)

Question 28.4-

You need to add test data for analysis in Azure.

Which two data formats should you use? Each correct answer presents a complete solution.

YAML

JSON

CSV

XML

Explanation

You should use JavaScript Object Notation (JSON) or comma-separated-value (CSV). These are two of three data formats supported for test data in Stream Analytics, which is what you should use to analyze the data. The third data format is AVRO.

You should not use XML or YAML. Neither of these data formats is supported by Streaming Analytics test data.

References

[Test an Azure Stream Analytics job with sample data](#)

Question 28.5-

You are specifying the input data source for the solution that queries the data, filters it, and sends it to Power BI.

What should you do?

For each of the following statements, select Yes if the statement is true. Otherwise, select No.

•

Statement	Yes	No
Add test centers as reference input.	<input checked="" type="radio"/>	<input type="radio"/>

Explanation

You should add test centers as reference input. Reference input is data that never or rarely changes. You can do this in Stream Analytics. Stream Analytics allows you to define an input data source, a query, and an output data source. The input data source can be an event hub, an IoT hub, or blob storage. The output data source in this scenario is Power BI. The query is a SQL-like query language. This allows you to take advantage of your existing skills.

You should not add candidates as data stream input. You should add candidates as reference input, because this input represents data that never or rarely changes.

You should add scores as data stream input. Data stream input changes over time. Candidate scores are not static data.

References

[Understand inputs for Azure Stream Analytics](#)

Question33 Develop streaming solutions

Case Study

Complete the Case Study

- Solution Evaluation

- Question 1
- Question 2
- Question 3
- Question 4

Instructions

This case study contains a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

Note: You cannot go back or review questions of this type on the actual certification exam.

Question 33.1-

You use Azure Stream Analytics to stream real-time IoT sensor data for a brewery company.

You need to use a window function with a fixed-size. Events should belong to a single

window.

Solution: You analyze the stream with a session window function.

Does the solution meet the goal?

No

Yes

Explanation

This solution does not meet the goal. Session window functions group events that arrive at a similar time. However, events could belong to more than one window and session windows have a variable length.

References

[Introduction to Stream Analytics windowing functions](#)

[Session Window \(Azure Stream Analytics\)](#)

[Window Functions in Stream Analytics](#)

Question 33.2-

You use Azure Stream Analytics to stream real-time IoT sensor data for a brewery company.

You need to use a window function with a fixed-size. Events could belong to more than one window.

Solution: You analyze the stream with a tumbling window function.

Does the solution meet the goal?

No

Yes

Explanation

This solution meets the goal. Tumbling window functions define fixed-size, non-overlapping and contiguous time intervals. In tumbling windows, events only belong to a single window.

References

[Introduction to Stream Analytics windowing functions](#)

[Tumbling Window \(Azure Stream Analytics\)](#)

[Window Functions in Stream Analytics](#)

Question 33.3-

You use Azure Stream Analytics to stream real-time IoT sensor data for a brewery company.

You need to use a window function with a fixed-size. Events could belong to more than one window.

Solution: You analyze the stream with a hopping window function.

Does the solution meet the goal?

Yes

No

Explanation

This solution meets the goal. Hopping window functions define fixed-size, overlapping and contiguous time intervals. When defining a hopping window, you need to define the window size and hop size (how long a window will overlap with the previous one). This results in events that could belong to one or more windows.

References

[Introduction to Stream Analytics windowing functions](#)

[Hopping Window \(Azure Stream Analytics\)](#)

Question 33.4-

You use Azure Stream Analytics to stream real-time IoT sensor data for a brewery company.

You need to use a window function with a fixed-size. Events could belong to more than one window.

Solution: You analyze the stream with a sliding window function.

Does the solution meet the goal?

No

Yes

Explanation

This solution does meet the goal. Sliding window functions define fixed-size, overlapping and contiguous time intervals. When you define a window length for a sliding function, Stream Analytics will consider all possible windows for that length. This is similar to setting a hopping window function with a hop size equal to zero, with the exception that the sliding function will only produce an output when an event occurs.

References

[Introduction to Stream Analytics windowing functions](#)

[Sliding Window \(Azure Stream Analytics\)](#)

[Window Functions in Stream Analytics](#)

Question37 Monitor data storage

Case Study

Complete the Case Study

- Background

You are building a modern data warehouse solution for your company.

Consumer and sales data is currently stored in an Azure SQL Database, while the product catalog is maintained in a Cosmos DB backend. The Marketing department also has access to the market research data, published weekly by a third-party vendor on their Amazon S3 storage in CSV format.

The Marketing team wants to use PowerBI as a reporting tool and query against a single consolidated dataset in Azure SQL Data Warehouse. You have internal data scientists who can help with the data transformation and consolidation.

Your company is using Cherwell as its service desk platform. You establish a bi-directional connection between your Azure subscription and Cherwell using a connector from the Azure Marketplace, but you have not used it yet.

- Business requirements

The new data warehouse solution must meet the following business requirements:

- Unauthorized users should not be able to see the contact details of consumers in Azure SQL Data Warehouse.
- System owners want to enforce the data retention policy and every month delete consumers who were not active for more than two years.

- Technical requirements

The new data warehouse solution must meet the following technical requirements:

- Market research data must be copied over to Azure and retained in its original format, storing files in a year and month-based hierarchical structure.
- Incidents with Azure SQL Data Warehouse based on Azure Monitor alerts need to be logged automatically in Cherwell using an existing Azure connection.

- Question 1
- Question 2
- Question 3
- Question 4
- Question 5

Question 37.1-

You need to select an Azure resource to store market research raw data.

Which resource should you choose?

Azure Managed Disks

Azure Table Storage

Azure Data Lake Storage Gen2

Azure Cosmos DB

Explanation

You should use Azure Data Lake Storage Gen2. It has the capabilities of both Azure Blob Storage and Azure Data Lake Storage Gen1 and also natively supports hierarchical namespaces. The market research data can be stored in the original .CSV format and kept in a year and month-based hierarchical structure.

You should not use Azure Managed Disks. These are virtual hard disks intended to be used as a part of Virtual Machine (VM) related storage.

You should not use Azure Table Storage. This is a no-SQL data store solution, optimized for storing data in key/value pairs. It is not intended for storing .CSV files.

You should not use Azure Cosmos DB. This is a globally distributed, multi-model database service. It is not intended for storing attachments and does not meet the goal of storing them in a hierarchical structure.

References

[Introduction to Azure Data Lake Storage Gen2](#)

[Azure Data Lake Storage Gen2 hierarchical namespace](#)

[Introduction to Azure managed disks](#)

[Azure Table storage overview](#)

[Welcome to Azure Cosmos DB](#)

Question 37.2-

You need to enable the required protection of consumer contact details in Azure SQL Data Warehouse.

What should you do?

Enable Transparent Data Encryption (TDE).

Enable row level security (RLS).

Enable Dynamic Data Masking (DDM).

Create a secret in Azure Key Vault.

Explanation

You should enable DDM. This would hide consumer contact details by masking it for unauthorized users. DDM allows limiting the exposure of specific fields.

You should not enable RLS. This control operates on the row level and will hide the whole record, while you need to protect only specific fields.

You should not enable TDE. TDE is enabled in Azure SQL Data Warehouse by default, and it cannot help hide or protect specific consumer contact details.

You should not create a secret in an Azure Key Vault. Creating custom secrets does not help prevent access to consumer contact details by unauthorized people.

References

[SQL Database dynamic data masking](#)

[Row-Level Security](#)

[Transparent data encryption for SQL Database and Data Warehouse](#)

[What is Azure Key Vault?](#)

Question 37.3-

You create partitions in SQL Data Warehouse to support the monthly deletion of obsolete users.

After uploading the historical data, your team realizes that the date partition needs to be further split. You use a SQL query to perform the task but get an error message for the ALTER PARTITION statement, as shown in the exhibit.

You need to resolve the problem without emptying the target partition.

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

Possible actions

Actions in order

- Use the ALTER TABLE statement with the SWITCH clause.
- Drop the table.
- Disable the columnstore index.
- Use the ALTER TABLE statement with the SPLIT clause.
- Rebuild the columnstore index.

Explanation

You should perform the following steps in order:

1. Disable the columnstore index.
2. Use the ALTER TABLE statement with the SPLIT clause.
3. Rebuild the columnstore index.

If the table has a columnstore, as in this case, the partition can be split only when it is empty. However, the goal here is to retain existing data. That is why you should disable the columnstore index first. This allows you to use the ALTER TABLE

statement with the SPLIT clause again to split the partition. Afterwards, you can rebuild the columnstore.

You should not drop the table. That would delete all the data.

You should not use THE ALTER TABLE statement with the SWITCH clause. This is used to switch one partition to another, for example to switch to a temporary partition. This would empty the original partition.

References

[Partitioning tables in SQL Data Warehouse](#)

Question 37.4-

Your data scientists report that their queries in Azure SQL Data Warehouse often stay in Suspended mode around lunch time.

You need to monitor the execution of the queries using Dynamic Management Views (DMVs) to identify the longest running ones.

Which DMV should you use?

sys.dm_pdw_exec_requests

sys.dm_exec_sessions

sys.dm_pdw_exec_sessions

sys.dm_exec_requests

Explanation

You should use sys.dm_pdw_exec_requests. This DMV can provide information about currently or recently active queries in Azure SQL Data Warehouse and can be sorted by the elapsed execution time to identify the longest running ones.

You should not use sys.dm_pdw_exec_sessions. This DMV can provide information about current or recent logins, with details of the client app, IP address, and number of queries executed. However, it cannot help to identify the longest running queries.

You should not use sys.dm_exec_requests or sys.dm_exec_sessions. These DMVs are not applicable to Azure SQL Data Warehouse. They are available for

SQL Server and Azure SQL Database only. Azure SQL Data Warehouse DMVs include the pdw (Parallel Data Warehouse) prefix.

References

[Monitor your workload using DMVs](#)

[sys.dm_pdw_exec_requests \(Transact-SQL\)](#)

[sys.dm_pdw_exec_sessions \(Transact-SQL\)](#)

[sys.dm_exec_requests \(Transact-SQL\)](#)

[sys.dm_exec_sessions \(Transact-SQL\)](#)

Question 37.5-

You are configuring a new rule in Azure Monitor to trigger an alert if the number of the failed connections exceeds 30 within a 10-minute interval.

You need to choose the right action type to log incidents in Cherwell when an alert is fired. You need to configure it with the Portal UI and use the existing connection to the service desk platform.

Which action type should you choose?

Automation Runbook

Azure Functions

IT Service Management Connector (ITSMC)

Push Notification

Explanation

You should use ITSMC. Azure ITSMC supports various ITSM tools, including Cherwell. By using this action type, it is possible to automatically log new incidents in Cherwell.

You should not use Automation Runbook. Automation Runbook allows you to run a workflow, for example to shut down a selected service, but you cannot use it to non-programmatically log Cherwell incidents via an existing ITSM connector.

You should not use Azure Functions. Azure Functions is an event-driven serverless compute platform that allows you to code a solution that you can run on demand. However, you cannot use it to non-programmatically log Cherwell incidents via an existing ITSM connector.

You should not use a push notification. Push notifications allow sending alerts to the Azure mobile app, but you cannot use them to non-programmatically log Cherwell incidents via an existing ITSM connector.

References

[Connect Azure to ITSM tools using IT Service Management Connector](#)

[Create alerts for Azure SQL Database and Data Warehouse using Azure portal](#)

[Azure Automation runbook types](#)

[An introduction to Azure Functions](#)

[Azure mobile app](#)

Question42 Monitor data storage

Case Study

Complete the Case Study

- Solution Evaluation

- Question 1
- Question 2
- Question 3

Instructions

This case study contains a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

Note: You cannot go back or review questions of this type on the actual certification exam.

Question 42.1-

You are a data engineer for your company. You manage an Azure SQL Database service. You want to monitor the database weekly for performance.

You need to choose a tool or service that monitors the database for performance opportunities through the creation or dropping of indexes.

Solution: You use Query Performance Insight.

Does this solution meet the goal?

No

Yes

Explanation

This solution does not meet the goal. Query Performance Insight allows you to view database queries that consume the most resources and those that take the longest to run. It does not suggest when to create or drop and index.

References

[Query Performance Insight for Azure SQL Database](#)

Question 42.2-

You are a data engineer for your company. You manage an Azure SQL Database service. You want to monitor the database weekly for performance.

You need to choose a tool or service that monitors the database for performance opportunities through the creation or dropping of indexes.

Solution: You use SQL Database Advisor.

Does this solution meet the goal?

No

Yes

Explanation

This solution does meet the goal. It allows you to review recommendations for creating and dropping indexes, fixing schemas, and parameterizing queries. You can access this service by opening the Performance Recommendations blade for your SQL Database instance.

References

[Find and apply performance recommendations](#)

Question 42.3-

You are a data engineer for your company. You manage an Azure SQL Database service. You want to monitor the database weekly for performance.

You need to choose a tool or service that monitors the database for performance opportunities through the creation or dropping of indexes.

Solution: You use Azure Advisor.

Does this solution meet the goal?

No

Yes

Explanation

This solution meets the goal. Azure Advisor provides recommendations for availability, security, performance, and cost. It integrates with SQL Database Advisor to provide recommendations for creating and dropping indexes.

References

[Improve performance of Azure applications with Azure Advisor](#)

[Introduction to Azure Advisor](#)

Case Study

Complete the Case Study

- Solution Evaluation

- Question 1
- Question 2
- Question 3

Instructions

This case study contains a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

Note: You cannot go back or review questions of this type on the actual certification exam.

Question 45.1-

You have an Azure Synapse Analytics SQL pool (data warehouse). Some tables contain sensitive data.

The applicable tables/columns must be discovered and categorized. Access to the

sensitive data must be logged and reported.

You need to configure the solution to meet the requirements.

Solution: Perform the following actions:

- Create a Log Analytics workspace.
- Enable database auditing.
- Enable Advanced Data Security.
- Classify the sensitive data.
- Use the built-in Access to Sensitive Data portal dashboard.

Does this solution meet the goal?

No

Yes

Explanation

This solution meets the goal. You need to enable database auditing to log the queries executed on the database. You need to create a Log Analytics workspace to store the auditing data. You need to enable the Advanced Data Security feature. Data discovery & classification is a functionality of this feature. You need to classify the sensitive data. For every applicable column, you specify the information type and you give it a sensitivity label. Later you can use the built-in Access to Sensitive Data portal dashboard for reporting.

References

[Azure SQL Database and Azure Synapse Analytics data discovery & classification](#)

[Get started with SQL database auditing](#)

Question 45.2-

You have an Azure Synapse Analytics SQL pool (data warehouse). Some tables contain sensitive data.

The applicable tables/columns must be discovered and categorized. Access to the sensitive data must be logged and reported.

You need to configure the solution to meet the requirements.

Solution: Use the Set-AzSqlServerAuditPowerShell cmdlet.

Does this solution meet the goal?

No

Yes

Explanation

Using the Set-AzSqlServerAudit cmdlet does not meet the goal. This only activates auditing. This does not discover and classify the sensitive data. Also, using this cmdlet, only audit action groups can be created. This cmdlet can be used to activate auditing all queries using the BATCH_COMPLETED_GROUP group. Activating auditing on a selection of the tables is not possible with this cmdlet.

You need to enable database auditing to log the queries executed on the database. You need to create a Log Analytics workspace to store the auditing data. You need to enable the Advanced Data Security feature. Data discovery & classification is a functionality of this feature. You need to classify the sensitive data. For every applicable column, you specify the information type and you give it a sensitivity label. Later you can use the built-in Access to Sensitive Data portal dashboard for reporting.

References

[Get started with SQL database auditing](#)

[Set-AzSqlServerAudit](#)

[Azure SQL Database and Azure Synapse Analytics data discovery & classification](#)

Question 45.3-

You have an Azure Synapse Analytics SQL pool (data warehouse). Some tables contain sensitive data.

The applicable tables/columns must be discovered and categorized. Access to the sensitive data must be logged and reported.

You need to configure the solution to meet the requirements.

Solution: Perform the following actions:

- Enable database auditing.
- Create security policies on the tables with sensitive data.
- Write Kusto queries on the auditing data.

Does this solution meet the goal?

No

Yes

Explanation

This solution does not meet the goal. You should not create security policies on the tables with sensitive data. When a security policy is created, row-level security is activated. It can filter the rows returned to the client, acting like a hidden WHERE clause. It can be used to partially block access to sensitive data. Data access should not be blocked, but tracked and recorded.

You need to enable database auditing to log the queries executed on the database. You need to create a Log Analytics workspace to store the auditing data. You need to enable the Advanced Data Security feature. Data discovery & classification is a functionality of this feature. You need to classify the sensitive data. For every applicable column, you specify the information type and you give it a sensitivity label. Later you can use the built-in Access to Sensitive Data portal dashboard for reporting.

References

[Azure SQL Database and Azure Synapse Analytics data discovery & classification](#)

[Get started with SQL database auditing](#)

[Azure SQL Database metrics and diagnostics logging](#)

[CREATE SECURITY POLICY \(Transact-SQL\)](#)

[Row-Level Security](#)

Question48 Implement non-relational data stores – D -

You have a globally distributed application with millions of documents stored in Cosmos DB.

Your application is spread across five Azure regions and stores data as documents. An example is shown in the exhibit. The region field stores the original creation region. The sensorId is a unique field that does not repeat across the database.

The application performs thousands of write and read operations per second, resulting in high throughput. Queries originated from the application usually filter the results by region and sensorId.

You need to choose a proper partition key for Cosmos DB to meet the requirements.

What partition key should you use?

Choose the correct answer

timestamp with random suffix

region

region with pre-calculated suffix based on sensorId

sensorId

Explanation

You should use region with pre-calculated suffix based on sensorId. This partition key will distribute all the documents evenly across logical partitions. Including a pre-calculated suffix based on a known value (sensorId, for example) will greatly improve both write and read throughput across the partitions.

You should not use region. This field contains only five possible values, resulting in a small number of logical partitions and a low read and write throughput.

You should not use sensorId. This field is unique across all documents and will result in millions of logical partitions with a single document, impacting the read and write throughput.

You should not use timestamp with random suffix. This partition key will distribute all documents across logical partitions evenly, greatly increasing the write throughput. However, reading a specific item will become harder because you do not know which suffix was used, impacting the read throughput.

References

[Create a synthetic partition key](#)

[Partitioning in Azure Cosmos DB](#)

Question49 Implement non-relational data stores

You are developing an application that queries data from Azure Cosmos DB.

The application will be hosted in multiple Azure App Services that are distributed across multiple regions. Data stored in Cosmos DB is updated frequently using a replica to write

and read data in the same region.

You need to ensure that read operations will return the most recent committed version of data. The application will consume Cosmos DB with the SQL API.

The Cosmos DB account should be named GlobalDatabase in a resource group named MotorsWorldwide.

You need to complete the Azure Command Line Interface (CLI) command to provision Azure Cosmos DB according to the requirements.

How should you complete the code? To answer, select the appropriate options from the drop-down menus.

Choose the correct options

```
az cosmosdb create --name GlobalDatabase
--resource-group MotorsWorldwide
--kind " "
--default-consistency-level " "
--
```

Kind

Global Document DB

Default-consistency-level

Strong

Enable-multiple-write-location

```
az cosmosdb create --name GlobalDatabase  
--resource-group MotorsWorldwide  
--kind "GlobalDocumentDB" "  
--default-consistency-level "Strong" "  
-- enable-multiple-write-locations
```

Explanation

You should complete the command as shown below:

```
az cosmosdb create --name GlobalDatabase  
--resource-group MotorsWorldwide  
--kind "GlobalDocumentDB"  
--default-consistency-level "Strong"  
--enable-multiple-write-locations
```

You should use GlobalDocumentDB to provision a Cosmos DB with the SQL API as required by the application.

You should use the Strong consistency level. This level is guaranteed to return the most recent committed version of a records.

You should enable multiple write locations to provision Cosmos DB in multi-regions with multiple write regions to satisfy the application requirements.

You should not use MongoDB or Parse. The --kind arguments are used to provision MongoDB API and SQL API respectively.

You should not use the Eventual and Session consistency levels. These two levels do not guarantee reading the most recent version of a record across all regions.

You should not enable automatic failover. This is used to automatically failover Cosmos DB in disaster recovery scenarios.

You should not enable virtual network support. Virtual network support is used to configure access from a virtual network. Azure App Service does not require a virtual network to connect with Cosmos DB.

References

[Consistency levels in Azure Cosmos DB](#)

[az cosmosdb](#)

[Access Azure Cosmos DB from virtual networks \(VNet\)](#)

[Manage an Azure Cosmos account](#)

Question50 Implement relational data stores - D

Your company is using various SQL and no-SQL databases in the Microsoft Azure cloud and on-premises to collect and store logistics data. However, business users struggle to build their Business Intelligence (BI) reports because of inconsistency and complexity of existing data stores.

You plan to consolidate the required data and make it available for BI reporting from a centralized Azure SQL Data Warehouse. You plan to use PolyBase as a mechanism for the data load.

As a first step, you automate the data extract from the source databases to Azure Data Lake Storage Gen2. You also create a new Azure SQL Data Warehouse resource and set up a service principal in Azure Active Directory (Azure AD).

You need to define external tables in your Azure SQL Data Warehouse, so that PolyBase can understand where and how to access your extracted data files, before it can load the data.

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

Create a list in the correct order

Possible actions

Actions in order

- Create a clustered columnstore index.
- Execute DBCC OPENTRAN.
- Create a database master key.
- Create a database scoped credential.
- Create an external data source and external file format.
- Create an external table.

Explanation

You should perform the following steps in order:

1. Create a database master key.
2. Create a database scoped credential.
3. Create an external data source and external file format.
4. Create an external table.

To access Azure Data Lake Storage Gen2, SQL DW uses service principal credentials. However, first it requires the creation of the master key (an operation that is required only once per database), which is then used to encrypt those credentials and store them as a database scoped credential.

Relevant credentials will then be referenced in the newly created external data sources, which are used along with the external file formats to create external tables in the target Azure SQL Data Warehouse.

You should not create a clustered columnstore index. A columnstore index is an in-memory table used in operational analytics. It is not relevant to the setup of external tables.

You should not execute DBCC OPENTRAN. It is a database command line utility, which helps to identify active transactions that may be preventing log truncation.

References

[Designing a PolyBase data loading strategy for Azure SQL Data Warehouse](#)

[CREATE DATABASE SCOPED CREDENTIAL \(Transact-SQL\)](#)

[Indexing tables in SQL Data Warehouse](#)

[DBCC OPENTRAN \(Transact-SQL\)](#)

Question51 Implement relational data stores – D -

You are a data engineer for a utility billing solutions company. Your company has five customers that use your company for utility billing. Each customer has different peak usage periods within the year. In the legacy environment, each customer has its own database. You deploy an Azure SQL Database elastic pool.

You need to configure the number of data transaction units (DTUs) to minimize cost.

What should you do?

Choose the correct answer

Determine the number of total DTUs that are used by all five databases combined.

Determine the number of DTUs that are used by the largest database.

Determine the number of DTUs that are used by the database with the most transactions.

Determine the number of DTUs that are used by the database with the longest peak period.

Explanation

You should determine the number of total DTUs that are used by all five databases combined. This sets the DTUs for the pool. All databases in a pool share the same resources. Because some databases will be used more frequently than others during peak periods, this allows the pool to shift resources around.

You should not determine the number of DTUs that are used by the database with the most transactions. That alone does not provide enough DTUs to meet the pool. You should use enough DTUs so that resources can be shifted during peak periods.

You should not determine the number of DTUs that are used by the largest database. That alone does not provide enough DTUs to meet the pool. You should use enough DTUs so that resources can be shifted during peak periods.

You should not determine the number of DTUs that are used by the database with the longest peak period. That alone does not provide enough DTUs to meet the pool. You should use enough DTUs so that resources can be shifted during peak periods.

References

[Elastic pools help you manage and scale multiple Azure SQL databases](#)

[Service tiers in the DTU-based purchase model](#)

Question52 Implement relational data stores - D

You are a data engineer for your company. Your company is planning to upgrade a customer's web application to use a different user interface technology. You create an Azure SQL Database instance that the developers will use while developing the application. Developers must only be able to access the database while on the company's

premises.

You need to configure the database to meet these requirements.

What should you do?

Choose the correct answer

Run the New-AzSqlServerFirewallRule PowerShell cmdlet.

Add role assignments on the Access control (IAM) page of the Azure portal.

Run the az sql db audit-policy Azure CLI command.

Set the Allow access to Azure services setting to off in the Azure portal.

Explanation

You should run the New-AzSqlServerFirewallRule PowerShell cmdlet. This cmdlet creates a firewall rule that protects the logical server instance and database from unauthorized access. Only the IP addresses specified as parameters to the cmdlet can access the database from the internet. By default, internet access to Azure SQL Database is prohibited.

You should not set the Allow access to Azure services setting to off in the Azure portal. This setting allows you to execute queries and commands within the Azure portal. This does not affect external access.

You should not run the az sql db audit-policy Azure CLI command. This command allows you to manage the audit policy for an Azure SQL Database instance.

You should not add role assignments on the Access control (IAM) page of the Azure portal. Role assignments allow you to grant access to Azure resources to specific users or groups. In this scenario, you only want to ensure that users cannot access Azure SQL Database from the internet. Firewall settings provide this functionality.

References

[Azure SQL Database and Azure SQL Data Warehouse IP firewall rules](#)

[What is role-based access control \(RBAC\) for Azure resources?](#)

[az sql db audit-policy](#)

Question 53 Implement relational data stores - D

You are a data engineer for your company. Your company has an on-premises SQL Server instance that contains 16 databases. Four of the databases require Common Language Runtime (CLR) features.

You must be able to manage each database separately because each database has its own resource needs. You plan to migrate these databases to Azure. You want to migrate the databases by using a backup and restore process by using SQL commands.

You need to choose the most appropriate deployment option to migrate the databases.

What should you use?

Choose the correct answer

Azure Cosmos DB with the Table API

Azure Cosmos DB with the SQL (DocumentDB) API

Azure SQL Database managed instance

Azure SQL Database with an elastic pool

Explanation

You should use an Azure SQL Database managed instance deployment. This deployment option is almost 100% compatible with an on-premises instance, including the ability to use CLR features. When you back up the databases on-premises, you can execute a restore command to migrate the databases in Azure. This is referred to as lift and shift.

You should not use an Azure Cosmos DB with the SQL (DocumentDB) API deployment. Cosmos DB is a multimodel database that supports five APIs for storage and queries, including SQL, Table, Cassandra, Gremlin, and MongoDB. The SQL API allows you to access data by using SQL-like queries. You cannot restore SQL Server databases to Cosmos DB by using SQL commands.

You should not use an Azure Cosmos DB with the Table API deployment. The Table API is similar to Azure Tables. This deployment is useful if you are migrating an application from Azure Tables to Cosmos DB. With Azure Tables, you can access data by using Language Integrated Query (LINQ) and OData. You cannot restore SQL Server databases to Cosmos DB by using SQL commands.

You should not use an Azure SQL Database with an elastic pool deployment. An elastic pool allows you to deploy multiple databases to a single logical instance and have all databases share a pool of resources. You configure the resource usage up front by choosing a purchasing model. You cannot take advantage of CLR features with an elastic pool.

References

[Welcome to Azure Cosmos DB](#)

[Elastic pools help you manage and scale multiple Azure SQL databases](#)

[What is Azure SQL Database managed instance?](#)

Question54 Implement relational data stores – D -

You are a data engineer for your company. You deploy an Azure SQL Database instance with the single database deployment option. You want to give an existing user named Sam administrative rights to the database. Sam must be able to add and remove other users.

You need to complete the T-SQL command.

How should you complete the command? To answer, select the appropriate options from the drop-down menus.

Choose the correct options

ADD MEMBER Sam

Alter

ROLE

Db_owner

Choose the correct options

ALTER

ROLE

db_owner

ADD MEMBER Sam

Explanation

You should use the following command: ALTER ROLE db_owner ADD MEMBER Rahul

This statement adds Sam as the database owner (db_owner). The db_owner role has administrative access over the database, including the ability to add or remove other users.

The CREATE command allows you to create database objects, such as tables, logins, and users.

The GRANT command grants permissions to a user. For example, GRANT ALTER ANY USER TO Sam gives Sam the permission to create and remove other users. However, by adding Sam to the db_owner role, Sam automatically inherits that permission.

The db_datareader role allows a user to query database objects.

The db_datawriter role allows a user to write and update database objects.

References

[Controlling and granting database access to SQL Database and SQL Data Warehouse](#)

Question55 Implement relational data stores-

You are a data engineer for your company. You create an Azure SQL Database server named autoone. You want to deploy two databases to this server. You want to allow both databases to share the resources provided by the server.

How should you complete the commands? To answer, select the appropriate options from the drop-down menus.

az <<< A >>> create -n shared -s autoone

az <<< B >>> create -n database1 -s autoone -- <<< C >>> shared

az <<< D >>> create -n database2 -s autoone -- <<< E >>> shared

<<< A >>> Options

1. CosmosDB
2. SQL Elastic-Pool
3. SQL DB

<<< B >>> Options

1. CosmosDB database
2. sql db
3. sql elastic-pool

<<< C >>> Options

1. elastic-pool
2. zone-redundant

<<< D >>> Options

1. CosmosDB
2. SQL Elastic-Pool
3. SQL DB

<<< E >>> Options

1. elastic-pool
2. zone-redundant

Choose the correct options

az create -n shared -s autoone

az create -n database1 \

-s autoone -- shared

az create -n database1 \

-s autoone -- shared

Sql elastic_pool

Sql db

Elastic pool

Sql db

Elastic pool

Choose the correct options

az sql elastic-pool create -n shared -s autoone

az sql db create -n database1 \

-s autoone --elastic-pool shared

az sql db create -n database1 \

-s autoone --elastic-pool shared

Explanation

You should use the following commands:

```
az sql elastic-pool create -n shared -s autoone
```

```
az sql db create -n database1 -s autoone --elastic-pool shared
```

```
az sql db create -n database2 -s autoone --elastic-pool shared
```

The first command creates an elastic pool named shared. An elastic pool allows multiple databases to share server resources.

The second and third commands create databases and add them to the shared elastic pool, as indicated by the --elastic-pool parameter.

You should not use the az cosmosdb create command. This creates an Azure Cosmos DB account, which is a multimodel database that allows you to use one of five APIs. In this scenario, you want to add a SQL database to the server.

You should not call the az sql db create command first. This command creates a database. You should first create the elastic pool so that you can add the database to the pool when you create the database.

You should not set the zone-redundant parameter. This parameter specifies a Boolean value that indicates whether or not the database should use zone redundancy. This allows you to copy the database to multiple datacenters in the same region to provide high availability.

References

[Elastic pools help you manage and scale multiple Azure SQL databases](#)

[az sql db](#)

[az cosmosdb](#)

[az sql elastic-pool](#)

Question56 Implement relational data stores -

You are a data engineer for your company. You create an Azure SQL Data Warehouse instance to house inventory data for 30 stores. There are approximately 10,000 products among the stores, with each distinct type of product having a different stock keeping unit (SKU). The inventory data will be captured every Friday. Friday updates cause massive changes to inventory data. **You want to shard the data to improve performance.**

You need to write a T-SQL statement to generate the table.

How should you complete the statement? To answer, select the appropriate code segments from the drop-down menus.

Choose the correct options

```
CREATE TABLE [dbo].[StoreInventoryData](
```

```
    [Sku] [varchar](30) ,
```

```
    [Price] [decimal](38, 0) ,
```

```
    [StoreId] [varchar](20),
```

```
    [Quantity] [int] ,
```

```
    [Date] [datetime]
```

```
)
```

```
WITH
```

```
(
```

```
    CLUSTERED COLUMNSTORE INDEX ,
```

```
=
```

```
(
```

```
)
```

```
)
```

DISTRIBUTION

HASH

StoreID

Choose the correct options

```
CREATE TABLE [dbo].[StoreInventoryData](
    [Sku] [varchar](30) ,
    [Price] [decimal](38, 0) ,
    [StoreId] [varchar](20),
    [Quantity] [int] ,
    [Date] [datetime]
)
WITH
(
    CLUSTERED COLUMNSTORE INDEX ,
    DISTRIBUTION = HASH ( StoreId )
)
```

Explanation

You should use the following statement:

```
CREATE TABLE [dbo].[StoreInventoryData](
    [Sku] [varchar](30) ,
    [Price] [decimal](38, 0) ,
    [StoreId] [varchar](20),
    [Quantity] [int] ,
    [Date] [datetime]
)
WITH
(
    CLUSTERED COLUMNSTORE INDEX ,
    DISTRIBUTION = HASH ( StoreId )
)
```

This creates a hash-distributed table that uses StoreId as the distribution column. This allows the table to be sharded by store. Parallel queries can be run for different stores on different compute nodes.

You should not use Sku as the distribution column. This would distribute data across more than 10,000 tables.

You should not use the PARTITION table option. This creates a partitioned table.

In this scenario, you should create a distributed table.

You should not specify REPLICATE as the distribution type. This uses replicated distribution, which copies the same data across compute nodes. This is beneficial for small tables.

You should not specify ROUND_ROBIN as the distribution type. This creates a round-robin-distributed table. A round-robin distribution shards data evenly. Performance is better when using hash distribution.

References

[Guidance for designing distributed tables in Azure SQL Data Warehouse](#)

[Partitioned Tables and Indexes](#)

[Table geometries](#)

Question57 Implement relational data stores-

You are a data engineer for your company. You create an Azure SQL Data Warehouse instance to house inventory data for 30 stores. You want to shard the store data to improve performance when accessing store data. The size of the data is less than 200 megabytes (MB).

You need to write a T-SQL statement to generate the store table.

How should you complete the statement? To answer, select the appropriate code segments from the drop-down menus.

Choose the correct options

```
CREATE TABLE [dbo].[Store](
    [StoreId] [varchar](20),
    [StoreName] [varchar](200) ,
    [StoreNumber] [int] ,
    [StoreLocation] [varchar](200)
)
WITH
(
    CLUSTERED COLUMNSTORE INDEX ,
    =
)
```

Distribution

Replicate

Choose the correct options

```
CREATE TABLE [dbo].[Store](
    [StoreId] [varchar](20),
    [StoreName] [varchar](200) ,
    [StoreNumber] [int] ,
    [StoreLocation] [varchar](200)
)
WITH
(
    CLUSTERED COLUMNSTORE INDEX ,
    DISTRIBUTION = REPLICATE
)
```

Explanation

You should use the following statement:

```
CREATE TABLE [dbo].[Store](
    [StoreId] [varchar](20),
    [StoreName] [varchar](200) ,
    [StoreNumber] [int] ,
    [StoreLocation] [varchar](200)
```

```
)  
WITH  
(  
    CLUSTERED COLUMNSTORE INDEX ,  
    DISTRIBUTION = REPLICATE  
)
```

This creates a replicated table. A replicated table is copied across all the compute nodes in a data warehouse. This improves the performance of queries for data in small tables. In this scenario, the Store table is small. It is less than 200 MB.

You should not use the PARTITION table option. This creates a partitioned table. In this scenario, you should create a distributed table.

You should not specify HASH as the distribution type. This uses hash distribution. Data is sharded across compute nodes by a column that you specify.

You should not specify ROUND_ROBIN as the distribution type. This creates a round-robin-distributed table. A round-robin distribution shards data evenly. Query performance is better when using hash distribution.

References

[Guidance for designing distributed tables in Azure SQL Data Warehouse](#)

[Partitioned Tables and Indexes](#)

[Table geometries](#)

Question58 Implement relational data stores-

You are a data engineer for your company. You create an Azure SQL Data Warehouse instance to house inventory data for 1000 stores. Each store has a distinct store number. You want to shard the store data to improve performance when accessing store data. You want to ensure that the data is evenly sharded across compute nodes.

You need to write a T-SQL statement to generate the store table.

How should you complete the statement? To answer, select the appropriate code segments from the drop-down menus.

Choose the correct options

```
CREATE TABLE [dbo].[Store](
    [StoreId] [varchar](20),
    [StoreName] [varchar](200) ,
    [StoreNumber] [int] ,
    [StoreLocation] [varchar](200)
)
WITH
(
    CLUSTERED COLUMNSTORE INDEX ,
    =
)
```

Distribution

Round_Robin

Choose the correct options

```
CREATE TABLE [dbo].[Store](
    [StoreId] [varchar](20),
    [StoreName] [varchar](200) ,
    [StoreNumber] [int] ,
    [StoreLocation] [varchar](200)
)
WITH
(
    CLUSTERED COLUMNSTORE INDEX ,
    DISTRIBUTION = ROUND_ROBIN
)
```

Explanation

You should use the following statement:

```
CREATE TABLE [dbo].[Store](  
    [StoreNumber] [varchar](30) ,  
    [StoreName] [varchar](200) ,  
    [StoreId] [varchar](20)  
)  
WITH  
(  
    CLUSTERED COLUMNSTORE INDEX ,  
    DISTRIBUTION = ROUND_ROBIN  
)
```

This creates a round-robin-distributed table. A round-robin distribution shards data evenly across compute nodes.

You should not use the PARTITION table option. This creates a partitioned table. In this scenario, you should create a distributed table.

You should not specify REPLICATE as the distribution type. This uses replicated distribution, which copies the same data across compute nodes. This is beneficial for small tables.

You should not specify HASH (StoreNumber) as the distribution type. This creates a hash-distributed table that uses StoreNumber as the distribution column. This allows the table to be sharded by store. Because store number is unique per store, and because there are 1000 stores, this would create a single data row in each table across 1000 tables.

References

[Guidance for designing distributed tables in Azure SQL Data Warehouse](#)

[Partitioned Tables and Indexes](#)

[Table geometries](#)

Question59 Implement relational data stores -

You are a data engineer for your company. You create an Azure Data Warehouse database that uses round robin distribution. You write the following query to retrieve store sales data:

```
SELECT S.[Amount] AS [Sales], ST.[Name] FROM [FactSalesByStore] AS S JOIN  
[DimStore] AS ST ON S.[StoreId] = ST.[StoreId]
```

This query returns over 200,000 records, and it runs slowly. There are over 50,000 stores.

You need to improve the performance of this query.

What should you do?

For each of the following statements, select Yes if you should perform the task. Otherwise, select No.

Statement	Yes	No
Use hash distribution on StoreId for FactSalesByStore.	<input type="radio"/>	<input type="radio"/>
Use hash distribution on StoreId for DimStore.	<input type="radio"/>	<input type="radio"/>
Use replicated distribution for DimStore.	<input type="radio"/>	<input type="radio"/>
Use an outer join instead of an inner join.	<input type="radio"/>	<input type="radio"/>

Explanation

You should use hash distribution on StoreId for both FactSalesByStore and DimStore. Hash distribution shards data across compute nodes by placing all data that uses the same hash key on the same compute node. This improves the performance of the query in the scenario.

You should use replicated distribution for the DimStore table. Replicated distribution copies the same data to all the compute nodes. This is useful when you want to read data from small fact tables.

You should not use an outer join instead of an inner join. This would return more rows than necessary, and it would not improve performance.

References

[Guidance for designing distributed tables in Azure SQL Data Warehouse](#)

Table geometries

Question60 Implement relational data stores -

You are a data engineer for your company. An Azure blob storage container contains a large comma-separated-value (CSV) file. You want to load the data from that file into an Azure SQL Data Warehouse database table. You run the following SQL statements:

```
CREATE MASTER KEY;
```

```
CREATE EXTERNAL FILE FORMAT TextFile  
WITH (  
  FORMAT_TYPE = DelimitedText,  
  FORMAT_OPTIONS (FIELD_TERMINATOR = ',')  
);
```

You need to run four additional SQL statements to load the data from the blob container.

Which four SQL statements should you run? To answer, move the appropriate statements from the list of possible statements to the answer area and arrange them in the correct order.

Create a list in the correct order

Possible SQL statements

SQL statements in order

- CREATE DATABASE
- CREATE DATABASE SCOPED CREDENTIAL
- CREATE EXTERNAL DATA SOURCE
- CREATE EXTERNAL TABLE
- CREATE TABLE

Explanation

You should run the following SQL statements:

```
CREATE DATABASE SCOPED CREDENTIAL  
CREATE EXTERNAL DATA SOURCE  
CREATE EXTERNAL TABLE  
CREATE TABLE
```

The first statement creates a credential in the database. This credential should represent the access key to the blob storage container. The following is an

example:

```
CREATE DATABASE SCOPED CREDENTIAL SampleStorageCredential
WITH
IDENTITY = 'SampleStorage',
SECRET =
'aabLNS8Hj9KSkFuP6mRMhOXJIndjiExfNEkuEL9YV1dFzKesC0elpxTqxFbqRsuo
zpWiGV2NikDWvRP4EzkBaA==';
```

The second statement creates an external data source to the blob storage container, using the credential that you created. The following is an example:

```
CREATE EXTERNAL DATA SOURCE SampleStorage
WITH (
TYPE = HADOOP,
LOCATION =
'wasbs://largetextfile@samplestorageaccount.blob.core.windows.net',
CREDENTIAL = SampleStorageCredential
);
```

The third statement creates a temporary table to hold the data. The following is an example:

```
CREATE EXTERNAL TABLE dbo.Temp (
  [DataField1] varchar(100),
  [DataField2] varchar(100)
  [DataField3] varchar(100)
)
WITH (
LOCATION='/',
DATA_SOURCE=SampleStorage,
FILE_FORMAT=TextFile
);
```

The fourth statement loads the data from the temporary table into a permanent table. The following is an example:

```
CREATE TABLE [dbo].[SampleTable]
WITH (
CLUSTERED COLUMNSTORE INDEX,
DISTRIBUTION = ROUND_ROBIN
)
AS
SELECT * FROM [dbo].[Temp];
```

You should not run the CREATE DATABASE statement. This statement creates a

database on the data warehouse server. No additional databases are required to load data into a data warehouse table.

References

[Exercise - import data from blob storage to Azure SQL Data Warehouse by using PolyBase](#)

[PolyBase query scenarios](#)

Question61 Implement relational data stores -

You manage your company's Manufacturing Execution Solution (MES), which stores data in an Azure SQL Database.

Because of a recent acquisition, you are asked to move the Azure SQL Database to another Azure region, so that the data resides closer to the manufacturing sites.

You need to move the Azure SQL Database to the new region and delete the original Azure SQL Database without losing data.

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

Create a list in the correct order

Possible actions

Actions in order

- Monitor the sync process with the Get-AzSqlDatabaseFailoverGroup cmdlet and verify that its ReplicationState is equal to 0.
- Execute the Switch-AzSqlDatabaseFailoverGroup cmdlet by using the failover group's read-write listener endpoint.
- Create a failover group and add the original Azure SQL Database.
- Monitor the sync process with the Get-AzSqlDatabaseFailoverGroup cmdlet and verify that its ReplicationState is equal to 2.
- Execute the Switch-AzSqlDatabaseFailoverGroup cmdlet by using the failover group's read-only listener endpoint.
- With the NSLOOKUP command, verify the swap of IP addresses between the failover group's read-write and read-only listeners.
- Delete the failover group and the original Azure SQL Database.

Explanation

You should perform the following steps in order:

1. Create a failover group and add the original Azure SQL Database.
2. Monitor the sync process with the `Get-AzSqlDatabaseFailoverGroup` cmdlet and verify that its `ReplicationState` is equal to 2.
3. Execute the `Switch-AzSqlDatabaseFailoverGroup` cmdlet by using the failover group's read-only listener endpoint.
4. With the `NSLOOKUP` command, verify the swap of IP addresses between the failover group's read-write and read-only listeners.
5. Delete the failover group and the original Azure SQL Database.

You should start with the creation of the failover group and adding the original Azure SQL Database. This will initiate the deployment and the synchronization of the Azure SQL Database in the secondary Azure location.

You should then monitor the sync process by using the `Get-AzSqlDatabaseFailoverGroup` cmdlet. When its output parameter `ReplicationState` returns 2 ("CATCH_UP"), this indicates that the databases are in sync and can be safely failed over.

You should then execute the `Switch-AzSqlDatabaseFailoverGroup` cmdlet. It needs to be executed against a secondary database server (represented by the failover group's read-only listener endpoint) that should become primary with full sync.

Once the failover process is completed, you should use the `NSLOOKUP` command to confirm that the primary and secondary databases switched their geographical IP addresses.

Finally, you should delete the failover group and the original Azure SQL Database. Their contents have now been successfully moved to the new region.

You should not monitor the sync process with the `Get-AzSqlDatabaseFailoverGroup` cmdlet and verify that its `ReplicationState` is equal to 0. Status 0 ("SEEDING") indicates that the secondary database is not yet seeded and that is why attempts to failover will fail.

You should not execute the `Switch-AzSqlDatabaseFailoverGroup` cmdlet by using the

failover group's read-write listener endpoint. This endpoint represents the primary database server, while this cmdlet should be executed against a secondary database server (represented by the failover group's read-only listener endpoint).

References

[How to move Azure SQL resources to another region](#)

[Switch-AzSqlDatabaseFailoverGroup](#)

[Use auto-failover groups to enable transparent and coordinated failover of multiple databases](#)

Question62 Implement relational data stores -

You are deploying a new e-commerce solution in the Microsoft Azure environment that uses Azure SQL Database as its relational data store.

As a part of the business continuity plan, you enable active geo-replication of the primary database to the secondary database in another Azure region.

The Marketing team wants you to update the price list in Azure SQL Database shortly after the announcement of the new product ranges.

You need to ensure that the price changes made in the primary database are replicated synchronously to the secondary database.

What should you do?

Choose the correct answer

Run the Get-AzSqlDatabaseReplicationLink Az PowerShell cmdlet after the change, using the connection string of the primary database.

Call the sp_wait_for_database_copy_sync procedure in the primary database after the change.

Call the sp_wait_for_database_copy_sync procedure in the secondary database after the change.

Verify in the sys.geo_replication_links view of Azure SQL database that the replication_state field of the secondary database record is set to 1.

Explanation

You should call the sp_wait_for_database_copy_sync procedure in the primary database after the change. By default, active geo-replication replicates data asynchronously. Calling this procedure in the primary database will cause the application to wait until the changes are replicated and acknowledged by the active

secondary database.

You should not call the `sp_wait_for_database_copy_sync` procedure in the secondary database after the change. This procedure should be called only in the primary database.

You should not run the `Get-AzSqlDatabaseReplicationLink` Az PowerShell cmdlet after the change, using the connection string of the primary database. This cmdlet gets all geo-replication links between the specified Azure SQL Database and a resource group or Azure SQL Server, but it does not help enabling synchronous replication.

You should not verify in the `sys.geo_replication_links` view of Azure SQL Database that the `replication_state` field of the secondary database record is set to 1. This view only shows databases with the geo-replication links, and value 1 for `replication_state` indicates that the geo-replication target is being seeded, but databases are not synchronized yet.

References

[Active Geo-Replication - sp_wait_for_database_copy_sync](#)

[Get-AzSqlDatabaseReplicationLink](#)

[sys.geo_replication_links \(Azure SQL Database\)](#)

Question63 Implement relational data stores -

You manage an Azure SQL database hosted in the Central US Azure region.

You need to synchronize the database with another Azure SQL Database instance that will be created in the East US Azure region. These databases will be used by a distributed application with write and read permission.

You need to perform an initial data sync with the new database with minimal downtime and enable bi-directional sync after the initial sync.

Choose the correct answer

Azure SQL Data Sync

Azure SQL active geo-replication

Azure Database Migration Service

Data Migration Assistant (DMA)

Explanation

You should use Azure SQL Data Sync. Data Sync is a service that lets you synchronize data across multiple Azure SQL Databases and on-premises SQL Server instances bi-directionally.

You should not use Azure SQL active geo-replication. Active geo-replication is a disaster recovery solution for Azure SQL Database that allows replicating a database to another Azure region. The synchronization direction is only from the master to the replica database, and you only have read access to the replica database.

You should not use DMA. DMA is an assessment tool for migrating SQL Server instances to Azure SQL Database. It evaluates incompatibilities and recommends performance improvements for the target database.

You should not use Azure Database Migration Service. This is a fully managed service to migrate multiple database sources to Azure with minimal downtime. It does not support bi-directional synchronization or migration.

References

[Sync data across multiple cloud and on-premises databases with SQL Data Sync](#)

[Creating and using active geo-replication](#)

[Overview of Data Migration Assistant](#)

[What is Azure Database Migration Service?](#)

Question64 Implement relational data stores -

Your company manages a modern data warehouse solution. Part of this solution consists in an Azure SQL Data Warehouse and Azure Data Lake Storage Gen2.

You need to configure PolyBase to load data from Azure Data Lake Storage Gen2 in a fact table named FactCarSales. This solution should not use service principals in Azure Active Directory (Azure AD).

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

Create a list in the correct order

Possible actions

Actions in order

- Create an external data source with the BLOB_STORAGE type.
- Create a scoped credential with the Client Id and OAuth 2.0 token endpoint.
- Create a scoped credential with the Azure storage account key.
- Create an external data source with the HADOOP type.
- Create an external file format.
- Create an external table.
- Load the data into the FactCarSales table.

Explanation

You should perform the following actions in order:

1. Create a scoped credential with the Azure storage account key.
2. Create an external data source with the HADOOP type.
3. Create an external file format.
4. Create an external table.
5. Load the data into the FactCarSales table.

You should create a scoped credential with the Azure storage account key. Azure Data Lake Storage Gen2 supports the use of storage account keys for Polybase access.

You should create an external data source with the HADOOP type using the scoped credential created to connect Azure SQL Data Warehouse with the Azure Data Lake Store.

You should create an external file format to configure the file format of the data stored in Azure Data Lake Store, such as a field delimiter for the source files.

You should create an external table using the external data source and the external file format previously defined to create a table representation of the data for Azure SQL Data Warehouse.

You should then load the data into the FactCarSales table. You can use SQL queries to load the data from the external table into the FactCarSales table.

You should not create a scoped credential with the Client Id and OAuth 2.0 token endpoint. You could use this option when you are connecting with Azure Data Lake Store. However, this would require a service principal in Azure AD to generate the

Client Id and the token.

You should not create an external data source with the BLOB_STORAGE type. This type is used when executing bulk operations with an on-premises SQL server or Azure SQL Database.

References

[Load data from Azure Data Lake Storage to SQL Data Warehouse](#)

[CREATE EXTERNAL DATA SOURCE \(Transact-SQL\)](#)

Question65 Implement relational data stores -

Your company is planning to migrate an on-premises data warehouse to Azure SQL Data Warehouse.

Some SQL Data Warehouse tables need special attention before the migration:

- There is a large fact table named FactSales with 5 terabytes (TB). Queries involving this table use a primary key defined in the SaleKey column to retrieve data.
- A dimension table named DimBusinessUnits with 200 megabytes (MB) is used by almost all reports.
- A staging table named StagingFactSales is used during the extract, transform, and load (ETL) process to load new data to the FactSales table.

You need to migrate these tables and configure the appropriate distribution strategy for each table to optimize query performance.

Which distribution strategy should you use? To answer, drag the appropriate distribution to the correct table. Each distribution may be used once, more than once, or not at all.

Drag and drop the answers

Replicated

Round-robin

Hash-distributed

Replicated

Round-robin

Hash-distributed

Drag and drop the answers

Hash-distributed

FactSales

Replicated

DimBusinessUnits

Round-robin

StagingFactTable

Replicated

Round-robin

Hash-distributed

Explanation

You should use the hash-distributed distribution method for the FactSales table. Hash-distributed tables achieve the highest query performance for large tables, splitting the work across multiple compute nodes.

You should use the replicated table option for the DimBusinessUnits table. This is a small table in size, and replicating its content across all compute nodes will improve performance when joining with other tables.

You should use round-robin distribution for the StagingFactTable table. This method distributes table rows evenly across all distributions and improves loading performance during the ETL process.

References

[Guidance for designing distributed tables in Azure SQL Data Warehouse](#)

[Design guidance for using replicated tables in Azure SQL Data Warehouse](#)

[CREATE TABLE \(Azure SQL Data Warehouse\)](#)

Question66 Implement relational data stores-

Your team is implementing a new Azure SQL Data Warehouse solution in Azure.

You need to use PolyBase to load data from a parquet file stored in an Azure Blob Storage in a table named FactSaleOrders.

You need to configure the Azure SQL Data Warehouse to receive the data.

Which five actions should you perform? To answer, move the appropriate actions from the list of possible actions to the answer area and arrange them in any order.

Create a list in any order

Possible actions

Actions to perform

- Load the parquet file into a staging table.
- Enable Transparent Data Encryption (TDE).
- Enable PolyBase support with the sp_configure command.
- Create a database master key.
- Create an external data source for Azure Blob Storage.
- Create an external table.
- Create an external file format for parquet file.
- Load the data in the FactSaleOrders table using T-SQL.

Explanation

You should perform the following actions in any order:

- Create a database master key.
- Create an external data source for Azure Blob Storage.
- Create an external file format from parquet file.
- Create an external table.
- Load the data in the FactSaleOrders table using T-SQL.

You should first create a database master key. This master key stores shared keys to access Azure Blob Storage.

You should create an external data source for Azure Blob Storage to allow the connection between the Azure SQL Data Warehouse and Azure Blob Storage.

You should create an external file format for the parquet file to configure how PolyBase will import the data using the parquet files.

You should create an external table. The external table is a representation of the data stored in Azure Blob Storage as a table. You need to use the external data source and the external file format previously defined with the parquet file to create this table.

You should load the data in the FactSaleOrders table using T-SQL. The external table created previously supports common T-SQL queries. You could use an insert into select statement to load data from the external table to the FactSaleOrders table.

You should not enable PolyBase support with the sp_configure command. Enabling PolyBase support is only required for an on-premises SQL Server. Azure SQL Data Warehouse support for PolyBase is enabled by default.

You should not enable TDE. TDE is used to encrypt data in rest. TDE should not be enabled to use PolyBase.

You should not load the parquet file into a staging table. You can directly use the data in the parquet file using the T-SQL statements from the external table. You do not need to create a staging table to load the data.

References

[Designing a PolyBase data loading strategy for Azure SQL Data Warehouse](#)

[Tutorial: Load New York Taxicab data to Azure SQL Data Warehouse](#)

[Configure PolyBase to access external data in Azure Blob Storage](#)

[CREATE EXTERNAL DATA SOURCE \(Transact-SQL\)](#)

[CREATE EXTERNAL FILE FORMAT \(Transact-SQL\)](#)

Question67 Implement relational data stores-

Your company is provisioning multiple Azure SQL managed instances for an automotive industry application. The application must be resilient in case of an outage impacting an Azure region.

You create a primary instance in the West US region named AutoIndustry and a secondary instance in the East US region named AutoIndustrySecondary. The instances are provisioned in a resource group named AutoProduction.

You need to implement an auto-failover group for database disaster recovery.

How should you complete the PowerShell cmdlet? To answer, select the appropriate options from the drop-down menus.

Choose the correct options

-Name AutoFailoverGroup

- AutoProduction

- AutoIndustry

-PartnerRegion EastUS

- AutoIndustrySecondary

-FailoverPolicy Automatic

Choose the correct options

```
New-AzSqlDatabaseInstanceFailoverGroup
```

```
-Name AutoFailoverGroup
```

```
-ResourceGroupName AutoProduction
```

```
-PrimaryManagedInstanceName AutoIndustry
```

```
-PartnerRegion EastUS
```

```
-PartnerManagedInstanceName AutoIndustrySecondary
```

```
-FailoverPolicy Automatic
```

Explanation

You should complete the cmdlet as shown below:

```
New-AzSqlDatabaseInstanceFailoverGroup  
-Name AutoFailoverGroup  
-ResourceGroupName AutoProduction  
-PrimaryManagedInstanceName AutoIndustry  
-PartnerRegion EastUS  
-PartnerManagedInstanceName AutoIndustry Secondary  
-FailoverPolicy Automatic
```

You should use `New-AzSqlDatabaseInstanceFailoverGroup` to create a failover group between Azure SQL Managed Instances. This command has four required parameters, the failover group Name, `PrimaryManagedInstanceName`, `PartnerManagedInstanceName` and `PartnerRegion`. You need to properly configure these properties to create the failover group as specified.

You should configure `ResourceGroupName` for `AutoProduction`. This resource group will be used to provision your failover group.

You should configure `PrimaryManagedInstanceName` for `AutoIndustry`. This is the primary managed instance as specified in requirements.

You should configure `PartnerManagedInstanceName` for `AutoIndustrySecondary`. This is the secondary database as specified in requirements.

You should configure `PartnerRegion` as `EastUS`. This is the region where `AutoIndustrySecondary` is provisioned.

You should not use Set-AzSqlDatabaseFailoverGroup. This command is used to modify a failover group configuration, like setting a new failover group policy.

You should not use New-AzSqlDatabaseFailoverGroup. This command creates a failover group for Azure SQL Databases. You need to create a failover group for Azure SQL Managed Instances.

References

[Overview of business continuity with Azure SQL Database](#)

[Use auto-failover groups to enable transparent and coordinated failover of multiple databases](#)

[New-AzSqlDatabaseInstanceFailoverGroup](#)

[Set-AzSqlDatabaseInstanceFailoverGroup](#)

[New-AzSqlDatabaseFailoverGroup](#)

Question68 Implement relational data stores-

You are creating tables for Azure Synapse Analytics.

You need to choose the appropriate types of tables for various needs.

Which table option should you use in the following cases? To answer, drag the appropriate table option to each table category. A table option may be used once, more than once, or not at all.

Drag and drop the answers

Hash-distributed

Replicated

Round-robin

Partitioned

Hash-distributed

Replicated

Round-robin

Partitioned

Temporary

Table category	Recommended option
Large fact table	Hash-distributed
Staging table with unknown data	Round-robin
Dimension table with size of 1 GB	Replicated
Table with queries that scan a date range	Partitioned

Hash-distributed

Replicated

Round-robin

Partitioned

Temporary

Explanation

For a large fact table, you should create a hash-distributed table. Query performance improves when the table is joined with a replicated table or with a table that is distributed on the same column. This avoids data movement.

For a staging table with unknown data, you should create a round-robin distributed table. The data will be evenly distributed across the nodes, and no distribution column needs to be chosen.

For a small dimension table, you should use a replicated table. No data movement is involved when the table is joined with another table on any column.

For a table that has queries that scan a date range, you should create a partitioned table. Partition elimination can improve the performance of the scans when the scanned range is a small part of the table.

You should not create a temporary table. A temporary table is only visible to the session that creates it and will be deleted when the session ends.

References

[Designing tables in Azure SQL Data Warehouse](#)

[Guidance for designing distributed tables in SQL Analytics](#)

[Design guidance for using replicated tables in SQL Analytics](#)

[Partitioning tables in SQL Analytics](#)

[Temporary tables in SQL Data Warehouse](#)

Question69 Implement relational data stores-

You have an Azure Synapse Analytics database.

You need a list of all hash distributed tables. The list must include the table name and the column name of the distribution column.

Which three catalog views do you need to join in a query? Each correct answer presents part of the solution.

Choose the correct answers

sys.columns

sys.pdw_table_distribution_properties

sys.pdw_distributions

sys.pdw_column_distribution_properties

sys.pdw_nodes_columns

Explanation

You need the following catalog views:

- sys.tables - Gives the tables, including the name.
- sys.columns - Gives the columns, including the name.
- sys.pdw_column_distribution_properties - Gives the distribution information.

Example of a query:

```
SELECT t.name AS [tablename], c.name AS [columnname]
FROM sys.tables AS t
JOIN sys.columns AS c ON c.object_id = t.object_id
JOIN sys.pdw_column_distribution_properties AS d ON d.object_id = t.object_id AND
d.column_id = c.column_id
WHERE d.distribution_ordinal = 1
```

You do not need the following catalog views:

- sys.pdw_distributions - Gives information about the distributions on the appliance.
- sys.pdw_table_distribution_properties - Gives the distribution information for the tables.
- sys.pdw_nodes_columns - Gives columns for user-defined objects.

References

[System views supported in SQL Analytics](#)

[sys.pdw_column_distribution_properties \(Transact-SQL\)](#)

Question70 Implement relational data stores-

You have a table named Sales in Azure Synapse Analytics SQL pool with the following definition

```
CREATE TABLE Sales (  
    ID [int] identity NOT NULL,  
    CustomerKey [int] NOT NULL,  
    Amount [money] NOT NULL,  
    Date [date] NOT NULL)  
WITH  
(DISTRIBUTION = HASH (CustomerKey),  
PARTITION ( [Date] RANGE RIGHT  
FOR VALUES ('2018-01-01', '2019-01-01', '2020-01-01', '2021-01-01' )));
```

The table does not contain dates after the year 2020.

You need to archive the oldest partition (with dates before January 1, 2018). The archive table does not exist, and the name should be SalesHistory. After archiving the old data, you need to add a partition to the Sales table for the next year.

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

Create a list in the correct order

Possible actions

Actions in order

- Create the SalesHistory table with the same column definitions, data distribution and partition boundaries as the Sales table.
Create a check constraint on the Date column limiting the dates to smaller than '2018-01-01'.
- Switch partition 1 of the Sales table to partition 0 of the SalesHistory table.
- Create a new partition in the SalesHistory table with this command:
ALTER TABLE SalesHistory SPLIT RANGE ('2022-01-01');
- Create the SalesHistory table with the same column definitions and data distribution as the Sales table. Include one partition boundary as '2018-01-01'.
- Switch partition 1 of the Sales table to partition 1 of the SalesHistory table.

- Remove the boundary value '2018-01-01' in the Sales table by using:
`ALTER TABLE Sales MERGE RANGE ('2018-01-01');`
- Create a new partition in the Sales table with this command:
`ALTER TABLE Sales SPLIT RANGE ('2022-01-01');`

Explanation

You should perform the following actions in this order:

1. Create a history table with the same boundary as the partition that will be switched into it. The definition must match the Sales table:

```
CREATE TABLE SalesHistory (  
    ID [int] identity NOT NULL,  
    CustomerKey [int] NOT NULL,  
    Amount [money] NOT NULL,  
    Date [date] NOT NULL)  
WITH  
(DISTRIBUTION = HASH (CustomerKey),  
PARTITION ( [Date] RANGE RIGHT  
FOR VALUES ('2018-01-01')));
```

The physical structure of the Sales and SalesHistory tables must be exactly the same. Partition switching is a metadata-only operation, so no actual data is moved or copied. The database engine must guarantee that all dates in a partition fall within the boundaries. This is only possible if the boundaries of the partitions that are involved in the switch process are equal.

2. Switch partition 1 of the Sales table to partition 1 of the SalesHistory table:

```
ALTER TABLE Sales SWITCH PARTITION 1 TO SalesHistory PARTITION 1;
```

By doing this, you archive the data in the first partition. The data in the first partition of the Sales table is switched to the first partition of the SalesHistory table.

3. Remove the partition boundary of January 1th 2018 in the Sales table:

```
ALTER TABLE Sales MERGE RANGE ('2018-01-01');
```

The first partition in the Sales table is now empty, and it can be removed. MERGE RANGE removes a boundary value, and thus removes a partition.

4. Create a new partition in the Sales table for the next year:

```
ALTER TABLE Sales SPLIT RANGE ('2022-01-01');
```

SPLIT RANGE will create a new boundary, and thus a new partition.

You should not create a check constraint on the SalesHistory table. You cannot create a check constraint on a date range in Azure Synapse Analytics.

You should not switch partition 1 of table Sales to partition 0 of the SalesHistory table. Partition 0 does not exist.

You should not perform a partition split of the SalesHistory table. This is not a requirement.

References

[Partitioning tables in SQL Analytics](#)

[ALTER TABLE \(Transact-SQL\)](#)

[Unique Constraints and Check Constraints](#)

Question71 - Manage data security-

Your company stores your consumer contact details in an Azure SQL Database. Data is encrypted at rest with the default Microsoft-managed transparent data encryption (TDE).

Because of the company's new security policy, you are asked to re-encrypt consumer database with your company's own asymmetric key and enable auditing on its use.

To support this, you set up a new Azure Key Vault and import a custom encryption key into it.

You need to enable TDE with the new custom key from Azure Key Vault. You will use PowerShell cmdlets, not Azure Portal, for this configuration.

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

Create a list in the correct order

Possible actions

Actions in order

- Export an Azure SQL Database to a BACPAC file.
- Import a BACPAC file to an Azure SQL Database.
- Assign an Azure AD identity to the Azure SQL Database server.
- Grant Key Vault permissions to the Azure SQL Database server.
- Add the Key Vault key to the Azure SQL Database server and set it as TDE Protector.
- Turn on TDE.

Explanation

You need to perform the following steps in order:

1. Assign an Azure AD identity to the Azure SQL Database server.
2. Grant Key Vault permissions to the Azure SQL Database server.
3. Add the Key Vault key to the Azure SQL Database server and set it as TDE Protector.
4. Turn on TDE.

When you configure Azure SQL TDE in Azure portal, you get an Azure AD identity assigned and relevant permissions set to the Azure SQL Database server via the Azure Key Vault access policy automatically. However, when you use PowerShell, you need to perform these two steps manually.

Next you retrieve your custom key from the Azure Key Vault and add it to your Azure SQL Database server, and afterwards set that key as a TDE protector for all server

resources.

And finally you use the `Set-AzSqlDatabaseTransparentDataEncryption` cmdlet to turn on TDE again, which will re-encrypt your Azure SQL Database, switching from the Microsoft-managed key to your own custom-managed key.

You should not export or import the contents of your Azure SQL Database in BACPAC format. You would do that to move the contents of an Azure SQL Database between cloud and on-premises SQL instances.

References

[Azure SQL Transparent Data Encryption with customer-managed keys in Azure Key Vault: Bring Your Own Key support](#)

[PowerShell and CLI: Enable Transparent Data Encryption with customer-managed key from Azure Key Vault](#)

[Export an Azure SQL database to a BACPAC file](#)

[Quickstart: Import a BACPAC file to a database in Azure SQL Database](#)

Question72- Manage data security-

You are a data engineer for a law firm. You create the following Azure SQL Database table:

```
CREATE TABLE [Case]
(
    [CaseNumber] [int],
    [Attorney] [sysname],
    [Type] [varchar] (50),
    [Docket] [varchar] (10)
)
```

Each attorney has a database user account. The Attorney column represents a database user account. You create the following function:

```
CREATE FUNCTION attorney_cases(@Attorney as sysname)
RETURNS TABLE
```

WITH SCHEMABINDING

AS

RETURN SELECT 1 as retVal WHERE USER_NAME() = @Attorney

You need to write a query that performs the following:

- Enforce row-level security on the Case table.
- Ensure that rows are returned only for the attorney performing the query.

You need to construct the query.

How should you complete the SQL statement? To answer, select the appropriate code segments from the drop-down menus.

Choose the correct options

CREATE

CasesByAttorney

ADD

attorney_cases ()

ON

WITH (STATE = ON)

SECURITY POLICY

FilterPredicate dbo

Attorney

[dbo].[case]

How should you complete the SQL statement? To answer, select the appropriate code segments from the drop-down menus.

Choose the correct options

```
CREATE SECURITY POLICY CasesByAttorney  
  
ADD FILTER PREDICATE dbo.attorney_cases ( Attorney )  
  
ON [dbo].[Case]  
  
WITH (STATE = ON)
```

Explanation

You should use the following SQL statement:

```
CREATE SECURITY POLICY CasesByAttorney  
ADD FILTER PREDICATE dbo.attorney_cases ( Attorney )  
ON [dbo].[Case]  
WITH (STATE = ON)
```

CREATE SECURITY POLICY creates a security policy that enforces row-level security. ADD FILTER PREDICATE specifies the function that performs row-level security filtering. The function in this scenario accepts a sysname parameter. You should pass the Attorney field to this function. The ON Case clause specifies the Case table to which the security policy applies.

You should not use CREATE FUNCTION. A function that performs security policy filtering already exists. You should instead create a security policy.

You should not use CREATE FILTER PREDICATE. You must specify FILTER PREDICATE with the ADD statement.

You should not use ADD SECURITY POLICY. You must use CREATE SECURITY POLICY to create a security policy.

You should not specify CURRENT_USER or Case as the parameter to the attorney_cases function call. You should specify the name of the field that represents the attorney, which is Attorney. The function compares this value to the current user to determine if the user is allowed to access that row.

You should not use ON Attorney. The ON clause should specify the name of the

table to which the security policy applies, which is Case.

References

[Row-Level Security](#)

Question73- Manage data security-

You are a data engineer. The following query creates a table in an Azure SQL Database:

```
CREATE TABLE Employee (  
  [ID] [int],  
  [GivenName] [varchar](50)  
    COLLATE Latin1_General_BIN2  
    ENCRYPTED WITH (COLUMN_ENCRYPTION_KEY = EmployeeCEK,  
    ENCRYPTION_TYPE = RANDOMIZED ,  
    ALGORITHM = 'AEAD_AES_256_CBC_HMAC_SHA_256'),  
  [SurName] [varchar](50),  
  [SSN] [char](9)  
    COLLATE Latin1_General_BIN2  
    ENCRYPTED WITH (COLUMN_ENCRYPTION_KEY = EmployeeCEK,  
    ENCRYPTION_TYPE = DETERMINISTIC ,  
    ALGORITHM = 'AEAD_AES_256_CBC_HMAC_SHA_256'),  
)
```

You write the following query to insert data into the table:

```
DECLARE @SSN = '123456789';  
INSERT INTO Employee(ID, GivenName, SurName, SSN)  
SELECT 1, 'Sam', 'Jack', @SSN
```

You need to determine which queries will return data when you enable Always Encrypted and Parameterization for Always Encrypted.

For each of the following queries, select Yes if the query returns data. Otherwise, select No.

Query	Yes	No
-------	-----	----

DECLARE @SSN char(9) = '123456789'; SELECT GivenName FROM Employee WHERE SSN=@SSN	<input type="radio"/>	<input type="radio"/>
DECLARE @Name varchar(50) = 'Sam'; SELECT SSN FROM Employee WHERE GivenName = @Name	<input type="radio"/>	<input type="radio"/>
SELECT GivenName, SSN FROM Employee	<input type="radio"/>	<input type="radio"/>

Explanation

The following query returns data:

```
DECLARE @SSN char(9) = '123456789'; SELECT GivenName FROM Employee
WHERE SSN=@SSN
```

This query has a WHERE clause that filters data based on the SSN column. The SSN column uses deterministic encryption. With deterministic encryption, the SSN field is encrypted to the same value every time. This means that '123456789' gets encrypted to the same value each time the query is run, allowing it to match the encrypted value in the table.

The following query does not return data:

```
DECLARE @Name varchar(50) = 'Sam'; SELECT SSN FROM Employee WHERE
GivenName = @Name
```

This query has a WHERE clause that filters data based on the GivenName column. The problem is that the GivenName column uses randomized encryption. Each time the query is run, the name Sam gets encrypted to a different value. This prevents the encrypted value from matching the stored value when the query is run.

The following query returns data:

```
SELECT GivenName, SSN FROM Employee
```

This query does not use a filter in the WHERE clause that specifies encrypted columns. This causes data to be returned always.

References

[Always Encrypted \(Database Engine\)](#)

Question74- Manage data security - D

You are a data engineer. You create a table with the following query:

```
CREATE TABLE License (  
  [ID] int,  
  [Number] char(9) MASKED WITH (FUNCTION = 'partial(3, "xxxxx", 1)') NULL,  
  [GivenName] varchar(20),  
  [SurName] varchar(20)  
)
```

You insert data by using the following query:

```
INSERT INTO License (ID, Number, GivenName, SurName)  
SELECT 1, '111222333', 'Sam', 'Jack'
```

You then run the following query to return data:

```
SELECT Number FROM License where ID=1
```

You need to determine which value is returned from the query.

Which value is returned?

Choose the correct answer

22233

xxx22233x

1113

111xxxxx3

Explanation

The following value is returned:

111xxxxx3

The Number column has a mask. The statement partial (3, 'xxxxx', 1) creates a mask that causes the first three characters of the Number field to be returned, followed by five x's, and finally followed by the last character. No other value can be returned from the query.

References

[Dynamic Data Masking](#)

[SQL Database dynamic data masking](#)

Question75- Manage data security - D

You are a data engineer for your company. You manage a SQL Server 2019 database on an Azure virtual machine (VM). A developer at the company needs to connect to the database from a client application. The client application passes the credentials in the connection string.

You need to allow the developer to return decrypted values for encrypted columns.

Which parameter should the developer specify in the connection string?

Choose the correct answer

Integrated Security = false

Column Encryption Setting = enabled

Integrated Security = true

Column Encryption Setting = disabled

Explanation

The developer should specify Column Encryption Setting = enabled. This allows SQL Server to decrypt the values of encrypted columns when queries are run.

The developer should not specify Column Encryption Setting = disabled. This prevents SQL Server from decrypting values of encrypted columns when queries are run.

The developer should not specify Integrated Security. Integrated Security specifies whether or not the current Active Directory credentials are used to access a database. If it is set to false, SQL Server looks for the user id and password in the connection string.

References

[Always Encrypted \(Database Engine\)](#)

[Always Encrypted \(client development\)](#)

Question76- Manage data security-

You are a data engineer for your company. You manage a SQL Server 2019 database on an Azure virtual machine (VM). You configure the database to use Always Encrypted.

You need to grant a user named Sam permission to manage the key that is used to encrypt and decrypt column encryption keys.

How should you complete the SQL statement? To answer, select the appropriate code segments from the drop-down menus.

Choose the correct options

GRANT

ANY COLUMN

TO 'SAM'

Alter

Master

Key

Choose the correct options

GRANT

ALTER

ANY COLUMN

MASTER

KEY

TO 'SAM'

Explanation

You should use the following statement:

GRANT ALTER ANY COLUMN MASTER KEY TO 'SAM'

The key that is used to encrypt and decrypt the column encryption keys is referred to as the master key. Sam needs permission to manage the master key. This statement grants Sam that permission.

You should not use the following statement:

GRANT ALTER ANY COLUMN ENCRYPTION KEY TO 'Sam'

This grants Sam permission to manage column encryption keys. In this scenario, Sam needs permission to manage the master key.

You should not use the following statement:

GRANT VIEW ANY COLUMN MASTER KEY DEFINITION TO 'Sam'

This grants Sam permission to read metadata of column master keys and to query encrypted columns.

You should not use the following statement:

GRANT VIEW ANY COLUMN ENCRYPTION KEY DEFINITION TO 'Sam'

This grants Sam permission to read metadata of column encryption keys and to query encrypted columns.

References

[Always Encrypted \(Database Engine\)](#)

Question77- Manage data security-

You are a data engineer for your company. You manage a SQL Server 2019 database on an Azure virtual machine (VM). You configure the database to use Always Encrypted.

You need to grant a user named Sam permission to query encrypted columns.

Which minimal permissions should you grant to Sam? For each of the following statements, select Yes if the statement is true. Otherwise, select No.

Statement	Yes	No
-----------	-----	----

Grant the ALTER ANY COLUMN MASTER KEY permission.	<input type="radio"/>	<input type="radio"/>
Grant the ALTER ANY COLUMN ENCRYPTION KEY permission.	<input type="radio"/>	<input type="radio"/>
Grant the VIEW ANY COLUMN MASTER KEY DEFINITION permission.	<input type="radio"/>	<input type="radio"/>
Grant the VIEW ANY COLUMN ENCRYPTION KEY DEFINITION permission.	<input type="radio"/>	<input type="radio"/>

Explanation

You should not grant the ALTER ANY COLUMN MASTER KEY permission to Sam. This permission allows you to manage the master key, which encrypts and decrypts column encryption keys. This permission is not required to query encrypted columns.

You should not grant the ALTER ANY COLUMN ENCRYPTION KEY permission to Sam. This permission allows you to manage column encryption keys. This permission is not required to query encrypted columns.

You should grant the VIEW ANY COLUMN MASTER KEY DEFINITION permission to Sam. This permission allows Sam to view column master key metadata and to query encrypted columns.

You should grant the VIEW ANY COLUMN ENCRYPTION KEY DEFINITION permission to Sam. This permission allows Sam to view column encryption key metadata and to query encrypted columns.

References

[Always Encrypted \(Database Engine\)](#)

Question78 - Manage data security-

You are a data engineer for your company. Your company is developing a new application that uses Azure SQL Database. You are designing a security policy to help keep the company's data secure.

You need to provide instructions to developers on how to prevent SQL injection attacks.

For each of the following statements, select Yes if the statement is true. Otherwise, select No.

Statement	Yes	No
Eliminate the use of stored procedures.	<input type="radio"/>	<input type="radio"/>
Use ad-hoc queries.	<input type="radio"/>	<input checked="" type="radio"/>
Validate user input.	<input type="radio"/>	<input type="radio"/>

Explanation

You should not eliminate the use of stored procedures. Stored procedures actually enhance security by ensuring that ad-hoc queries are not sent to the database.

You should not use ad-hoc queries. Ad-hoc queries increase vulnerabilities because malicious queries can be injected.

You should validate user input before allowing the input to be used in queries. This helps prevent SQL injection.

References

[SQL Injection](#)

Question79 - Manage data security-

You manage an Azure SQL Database that has the following table definition:

```
CREATE TABLE [Order]
(
  OrderID int,
  SalesPerson sysname,
  ProductID int,
```

```
Quantity int,  
Price decimal(8, 2)  
);
```

The SalesPerson field represents the database user of a sales person responsible for the order. A sales person should only be able to query his or her own orders. You want to use a security policy and row-level security to enforce this.

You need to define the security policy predicate.

How should you complete the code? To answer, select the appropriate code segments from the drop-down menus.

Choose the correct options

```
CREATE                                predicate(@SalesPerson AS sysname      )  
  
RETURNS  
  
AS  
  
RETURN SELECT 1  
  
WHERE @SalesPerson                    USER_NAME()                ;
```

Function

Table

=

Choose the correct options

```
CREATE FUNCTION predicate(@SalesPerson AS sysname )
RETURNS TABLE
AS
RETURN SELECT 1
WHERE @SalesPerson = USER_NAME() ;
```

Explanation

You should use the following code:

```
CREATE FUNCTION predicate(@SalesPerson AS sysname )
RETURNS TABLE
AS
RETURN SELECT 1 as SalesPersonResult
WHERE @SalesPerson = USER_NAME() ;
```

This code creates a function that serves as the security policy predicate. For each row in the table, SQL Database calls this function. The function compares the value of the SalesPerson column for that row with the current user. If the two match, it returns the value 1, which indicates to SQL Database that the current row should be returned.

You should not use CREATE TABLE. The security policy predicate should be a function, not a table. It should return a table.

You should not use RETURNS FUNCTION. The security policy predicate should return a table, not a function.

References

[Row-Level Security](#)

Question80- Manage data security - D

You are a data engineer for an Azure SQL Database. You write the following SQL statements:

```
CREATE TABLE Customer (  
  CustomerID int IDENTITY PRIMARY KEY,  
  GivenName varchar(100) MASKED WITH (FUNCTION = 'partial(2,"XX",0)') NULL,  
  SurName varchar(100) NOT NULL,  
  Phone varchar(12) MASKED WITH (FUNCTION = 'default()')  
);
```

```
INSERT Customer (GivenName, SurName, Phone) VALUES ('Sammy', 'Jack',  
'555.111.2222');
```

```
SELECT * FROM Customer;
```

You need to determine what is returned by the SELECT query?

What data is returned?

Choose the correct answer

1 SaXX Jack xxxx

1 XXXX Jack XXX.XXX.XXXX

1 SaXX Jack XXX.XXX.2222

1 xx Jack XXX.XXX.2222

Explanation

The following data is returned:

1 SaXX Jack xxxx

The GivenName column is created with the mask `partial(2, "XX", 0)`. This mask indicates that the first two characters of the GivenName field should be returned, followed by two X's, followed by no more characters of the GivenName field.

The PhoneNumber column is created with the mask `default()`. This uses the default mask for string data, which uses four x's for data that is four or more characters in length, and less than four x's for data that is less than four characters in length. Because the PhoneNumber column is 12 characters, the data returned is four x's.

No other value can be returned from the query.

References

Dynamic Data Masking

Question81- Manage data security-

Your company's accounting system uses Azure SQL Database as a backend. You enable geographic redundancy between the primary and secondary Azure SQL Database instances hosted in two different Azure regions.

Your corporate IT security policy dictates that, instead of Microsoft-managed keys, you should use your own asymmetric keys for Azure SQL Database Transparent Data Encryption (TDE). You deploy Azure Key Vaults in both Azure regions, create a new encryption key (TDE protector) in the primary region's Key Vault, and clone it into the secondary region's Key Vault via backup and restore.

You need to configure TDE in your geo-redundant Azure SQL Database environment to use your TDE protector. You should perform these tasks in Azure portal.

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

Create a list in the correct order

Possible actions

Actions in order

- On the primary Azure SQL Database server, set the Advanced Data Security feature to Off.
- On the secondary Azure SQL Database server, set the Advanced Data Security feature to Off.
- On the primary Azure SQL Database server, set the Advanced Data Security feature to On.
- On the secondary Azure SQL Database server, assign a Key Vault from the same region.
- On the secondary Azure SQL Database server, assign the TDE protector.
- On the primary Azure SQL Database server, assign a Key Vault from the same region.
- On the primary Azure SQL Database server, assign the TDE protector.

Explanation

You should perform the following steps in order:

1. On the secondary Azure SQL Database server, assign a Key Vault from the same region.
2. On the secondary Azure SQL Database server, assign the TDE protector.

3. On the primary Azure SQL Database server, assign a Key Vault from the same region.
4. On the primary Azure SQL Database server, assign the TDE protector.

If you already have Azure SQL Databases with geo-redundancy enabled, you should start with the assignment of the custom TDE protector from the secondary Azure SQL Database server before updating the primary one. This ensures that the geo replication link will continue to work after the TDE protector is updated on the primary server. The TDE protector will be available to both servers.

You should not set the Advanced Data Security features to On or Off on the primary or secondary servers. Advanced Data Security is a package that helps to find and classify sensitive data. It also identifies potential database vulnerabilities and anomalous activity, which may indicate a threat to your Azure SQL Database. However, it is not relevant for the activation of a custom TDE protector in a geo-redundant Azure SQL Database environment.

References

[Azure SQL Transparent Data Encryption with customer-managed keys in Azure Key Vault: Bring Your Own Key support](#)

[Advanced data security for Azure SQL Database](#)

Question82 Implement non-relational data stores - D

You are implementing a data engineering solution for your company. You plan to use Azure table storage to store receipts. Each receipt contains a date, a category, and a unique number. Over 50 percent of the receipts have the same category. These are the 10 most popular categories. Approximately five percent of the receipts have the same date.

You need to define the row key and the partition key scheme.

For each of the following statements, select Yes if you should perform the task. Otherwise, select No.

Statement	Yes	No

Specify the date as a partition key.	<input type="radio"/>	<input type="radio"/>
Specify the category as a partition key.	<input type="radio"/>	<input type="radio"/>
Specify the unique number as the row key.	<input type="radio"/>	<input type="radio"/>

Explanation

You should not specify the date as a partition key. A partition key allows you to create logical partitions for data. Values that have the same partition key share the same logical partition. Partition keys improve performance by allowing data to be scaled out horizontally.

You should specify the category as a partition key. All receipts that have the same category share the same logical partition. When you query receipts by category, you can be sure that they all are retrieved from the same partition, which improves query performance. Because 50 percent of receipts are for the 10 most popular categories, this improves performance because most queries will be performed against those receipts.

You should specify the unique number as the row key. A row key uniquely identifies an entity.

References

[Designing a Scalable Partitioning Strategy for Azure Table Storage](#)

Question83-

You work for a call center company that uses Azure SQL Database. The database stores customer credit card numbers in a table named PaymentInfo. Telemarketing attendants will consult this table to help with customer payment support.

You need to implement dynamic data masking (DDM) in the PaymentInfo table to mask credit card numbers for telemarketing attendants. Only the two first digits and the last four digits should be visible.

How should you complete the T-SQL query? To answer, select the appropriate options from the drop-down menus.

Choose the correct options

ALTER TABLE [PaymentInfo]

ALTER COLUMN [CreditCard]

ADD MASKED WITH (FUNCTION = ' (, "XX-XXXX-XXXX-",));

Partial

2

4

Choose the correct options

ALTER TABLE [PaymentInfo]

ALTER COLUMN [CreditCard]

ADD MASKED WITH (FUNCTION = 'partial (2 , "XX-XXXX-XXXX-", 4));

Explanation

You should complete the T-SQL query as shown below:

ALTER TABLE [PaymentInfo]

ALTER COLUMN [CreditCard]

ADD MASKED WITH (FUNCTION = 'partial(2,"XX-XXXX-XXXX-",4)');

You should use the partial function to mask only parts of the credit card number.

This function receives three arguments: the first is a prefix of how many characters should be shown from the beginning of the value. The second is the pattern used to mask the value. The third is a suffix of how many characters are shown from the end of the credit card number.

You should use 2 as prefix and 4 as suffix for the function to mask the credit card number as required.

You should not use the random function mask. This is used for numeric values to mask the original value within a specified range.

You should not use the default function mask. This mask is used to fully mask the data in the configured field with XXXX for string values. You need to expose the first two digits and the last four digits from the credit card number.

References

[Dynamic Data Masking](#)

Question84-

You are a data engineer. You manage three SQL Server databases in Azure. The databases must meet the following security requirements:

Database 1 - Only specific columns in the 10 tables must be encrypted.

Database 2 - All data in the entire database must be encrypted at rest.

Database 3 - Data must be encrypted while in transit between the client application and Azure.

You need to determine which encryption technology to use for each database.

Which encryption technology should you use for each database? To answer, choose the correct encryption technology from the drop-down menus.

Choose the correct options

Database 1

Database 2

Database 3

Always Encrypted

Transparent Data Encryption

Always Encrypted

Explanation

You should use Always Encrypted for Database 1. This allows you to encrypt specific columns in a table.

You should use Transparent Data Encryption (TDE) for Database 2. TDE allows you to encrypt an entire database at rest.

You should use Always Encrypted for Database 3. Always Encrypted also allows you to encrypt data while it is in transit between the client application and the database server.

References

[Always Encrypted \(Database Engine\)](#)

[Transparent Data Encryption \(TDE\)](#)

Question85-

You manage an Azure SQL database for a financial application.

You need to configure a dynamic data mask to completely mask the data of a specific varchar field.

Which masking function should you use?

Choose the correct answer

Email

Default

Partial

Random

Explanation

You should use the default mask. This will fully mask the data in the configured field with XXXX.

You should not use the random mask. This is used for numeric values to mask the original value within a specified range.

You should not use the partial mask. These masks could be used to expose the first or last character of the value with a given mask between those values. The default mask is better suited to mask the whole field.

You should not use the email mask. This mask exposes the first letter of the address and the last part of the email domain, like aXXX@XXXX.com.

References

[Dynamic Data Masking](#)

Question86-

You manage an Azure SQL Database for a mission-critical application named ElectronicsProduction. The database stores personal information about your users.

You need to implement Transparent Data Encryption (TDE) with a customer-managed encryption key in this database. You assign an Azure Active Directory (AD) identity in Azure SQL Database.

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

Create a list in the correct order

Possible actions

Actions in order

- Create a master key in the master database.
- Create a server certificate using the master key.
- Create a database encryption key from the certificate in ElectronicsProduction.
- Create an Azure Key Vault and generate a new key.
- Grant Key Vault permissions to the Azure SQL Database server.
- Add the Key Vault key to the Azure SQL Database server.
- Set the TDE Protector to use the Key Vault key.
- Enable encryption in the ElectronicsProduction database.

Explanation

You should perform the following actions in order:

1. Create an Azure Key Vault and generate a new key.
2. Grant Key Vault permissions to the Azure SQL Database server.
3. Add the Key Vault key to the Azure SQL Database server.
4. Set the TDE Protector to use Key Vault key.
5. Enable encryption in the ElectronicsProduction database.

You should create an Azure Key Vault and generate a new key. This key will be used to encrypt your database with TDE and is managed by the customer.

You should grant Key Vault permissions to Azure SQL Database server. Azure Key Vault uses a role-based access control (RBAC) to protect keys stored in it. You need to grant get, wrapKey, and unwrapKey permissions for the Azure AD identity assigned to the Azure SQL Database server.

You should add the Key Vault key to the Azure SQL Database server and set the TDE Protector to use the Key Vault key. With the proper permissions, you can add a Key Vault key to the Azure SQL Database server and set the TDE Protector to use this key.

You should enable encryption in the ElectronicsProduction database. After setting the

TDE Protector to use your own key, you can enable TDE in the ElectronicsProduction database and protect data at rest with your encryption key.

You should not create a master key in the master database, create a server certificate using the master key, and create a database encryption key from the certificate. You could use these steps if you were configuring TDE in an on-premise SQL Server database. By default, TDE is enabled in newly provisioned Azure SQL Databases using service-managed encryption keys.

References

[Transparent Data Encryption \(TDE\)](#)

[PowerShell and CLI: Enable Transparent Data Encryption with customer-managed key from Azure Key Vault](#)

[Azure SQL Transparent Data Encryption with customer-managed keys in Azure Key Vault: Bring Your Own Key support](#)

[Azure Key Vault security](#)

Question87 Implement non-relational data stores - D

You are a data engineer for your company. You use the following Azure CLI commands to create an Azure Cosmos DB account. You plan to use this account to store sales data.

```
az cosmosdb create --resource-group 'sales-rg' --name 'sales' --kind GlobalDocumentDB \
--locations regionName="South Central US" failoverPriority=0 \
--locations regionName="North Central US" failoverPriority=1 \
--default-consistency-level "Strong" --enable-multiple-write-locations true
```

You need to answer questions regarding sales data queries and updates.

For each of the following statements, select Yes if the statement is true. Otherwise, select No.

Statement	Yes	No

You can query data by using Gremlin API.	<input type="radio"/>	<input type="radio"/>
A client can see partial writes of a sales data record by default.	<input type="radio"/>	<input type="radio"/>
A client can set the consistency level to Eventual Consistency at connection time.	<input type="radio"/>	<input type="radio"/>
A client can set a different consistency level during each request to sales data.	<input type="radio"/>	<input type="radio"/>

Explanation

You cannot query data by using Gremlin API. The `--kind` parameter of the `az cosmosdb create` command specifies the API used by the Cosmos DB account. The value `GlobalDocumentDB` means that the account uses the SQL API.

A client cannot see partial writes of a sales data record by default. This is because the account uses a strong consistency level, as indicated by the `--default-consistency-level` parameter. With a strong consistency level, reads are guaranteed to return the most recently committed version of a data record.

A client can set the consistency level to Eventual Consistency at connection time. This allows the client to specify a consistency level different from the default. Eventual consistency is the weakest consistency level. It means that data reads and writes in a distributed system will be in sync eventually.

A client can set a different consistency level during each request to sales data. This allows the client to specify a consistency level different from the default level depending on the type of request being made.

References

[Consistency levels in Azure Cosmos DB](#)

[Create an Azure Cosmos SQL \(Core\) API account, database and container using Azure CLI](#)

[Create an Azure Cosmos Gremlin API account, database and graph using Azure CLI](#)

Question88 - Develop batch processing solutions-

You are a data engineer for an Azure SQL Data Warehouse. You want to import data to a database from a large pipe-separated file in an Azure blob storage container.

You need to create the connection to the container.

How should you complete the T-SQL statement? To answer, select the appropriate code segments from the drop-down menus.

Choose the correct options

```
CREATE EXTERNAL                                BlobStorage with (  
    TYPE =                                     ,  
    LOCATION = 'wasbs://sensordata@sample.blob.core.windows.net'  
);
```

DATA SOURCE

HADOOP

Choose the correct options

```
CREATE EXTERNAL DATA SOURCE BlobStorage with (  
    TYPE = HADOOP ,  
    LOCATION = 'wasbs://sensordata@sample.blob.core.windows.net'  
);
```

Explanation

You should use the following statement:

```
CREATE EXTERNAL DATA SOURCE BlobStorage with (  
  TYPE = HADOOP ,  
  LOCATION ='wasbs://sensordata@sample.blob.core.windows.net'  
);
```

The CREATE EXTERNAL DATA SOURCE statement creates the data source from which to import data. You must set the TYPE parameter to HADOOP when accessing data from Azure blob storage. The LOCATION parameter specifies the location of the blob storage container.

You should not set the TYPE property to BLOB_STORAGE. BLOB_STORAGE does not represent Azure blob storage. It simply designates the data source as one that will be used with BULK IMPORT or OPENROWSET. It cannot be used with external data sources such as Azure blob storage. An external data source is one that is not located in Azure SQL Database.

You should not use CREATE EXTERNAL TABLE. This statement allows you to define the columns to represent the external data source. However, you must first define the external data source.

References

[Exercise - import data from blob storage to Azure SQL Data Warehouse by using PolyBase](#)

[PolyBase query scenarios](#)

[CREATE EXTERNAL DATA SOURCE \(Transact-SQL\)](#)

[Configure PolyBase to access external data in Azure Blob Storage](#)

Question89 - Develop batch processing solutions – D -

You are a data architect for your company. Your team must analyze large CSV files from Azure blob storage daily. The team must be able to generate pie charts and bar charts without writing graphics code. The data engineers on the team know Python and SQL.

You need to recommend a solution for analyzing the file.

What should you recommend?

Choose the correct answer

Azure Databricks

Azure Data Lake

Stream Analytics

Log Analytics

Explanation

You should recommend Azure Databricks. This is an Apache Spark-based technology that allows you to run code in notebooks. Code can be written in SQL, Python, Scala, and R. You can have data automatically generate pie charts and bar charts when you run a notebook.

You should not recommend Azure Data Lake. Azure Data Lake is a big data storage solution that allows you to store data of any type and size.

You should not recommend Stream Analytics. Stream Analytics is a big data analytics solution that allows you to analyze real-time events simultaneously.

You should not recommend Log Analytics. Log Analytics allows you to write queries to analyze logs in Azure.

References

[What is Azure Databricks?](#)

[What is Azure Stream Analytics?](#)

[Get started with Log Analytics in Azure Monitor](#)

[What is Azure Data Lake Storage Gen1?](#)

Question90 - Develop batch processing solutions-

You are a data engineer for your company. You create an Azure Databricks account. You add code to a notebook cell to import data from a comma-separated-value (CSV) file named sensordata.csv into a folder named /tmp.

You need to copy the data to an Azure Data Lake Storage Gen 1 account.

Which command should you run?

Choose the correct answer

```
dbutils.fs.ls("copy /tmp/sensordata.csv abfss://samplefs@sample.dfs.core.windows.net/")
```

```
spark.read.json("abfss://tmp/sensordata.csv@sample.dfs.core.windows.net")
```

```
dbutils.fs.cp("file:///tmp/sensordata.csv", "abfss://samplefs@sample.dfs.core.windows.net/")
```

```
spark.conf.set("tmp/sensordata.csv", "abfss://samplefs@sample.dfs.core.windows.net/")
```

Explanation

You should run the following command:

```
dbutils.fs.cp("file:///tmp/sensordata.csv",  
"abfss://samplefs@sample.dfs.core.windows.net/")
```

This command copies the sensordata.csv file from the tmp folder in the Spark cluster to Azure Data Lake. Azure Data Lake Storage is a repository for Big Data analytics workloads.

You should not use the following command:

```
dbutils.fs.ls("copy /tmp/sensordata.csv  
abfss://samplefs@sample.dfs.core.windows.net/")
```

The dbutils.fs.ls command lists the contents of a Databricks folder.

You should not use the following command:

```
spark.conf.set("tmp/sensordata.csv",  
"abfss://samplefs@sample.dfs.core.windows.net/")
```

The spark.conf.set command sets a configuration property.

You should not use the following command:

```
spark.read.json("abfss://tmp/sensordata.csv@sample.dfs.core.windows.net")
```

The spark.read.json file reads data from a JavaScript Object Notation (JSON) file.

References

[Tutorial: Extract, transform, and load data by using Azure Databricks](#)

[Databricks Utilities](#)

[Introduction to Azure Data Lake Storage Gen2](#)

Question91 - Develop batch processing solutions – D -

You are a data engineer for your company. You create an Azure Databricks account by using the Azure portal. You plan to ingest data from blob storage into Databricks. You import a notebook from Github.

You need to create the next resource so that you can run code to ingest the data.

What should you create next?

Choose the correct answer

Cosmos DB account

SQL Data Warehouse instance

Spark cluster

Master key

Explanation

You should create a Spark cluster. Code in notebooks run on Spark clusters, which is an execution environment that runs on a cluster of nodes. If you import a notebook, it is imported as detached. You must then create a cluster and attach the notebook to that cluster. For existing Databricks accounts that already contain clusters, you can simply attach the notebook to the existing cluster.

You should not create a SQL Data Warehouse instance. SQL Data Warehouse allows you to perform parallel queries on Big Data. In this scenario, you want to ingest data into Databricks. It is not required to use SQL Data Warehouse with Databricks.

You should not create a Cosmos DB account. Cosmos DB is a multi-model, non-relational database that supports one of five APIs: SQL, Table, Cassandra, Gremlin, and MongoDB. It is not necessary to use Cosmos DB with Databricks.

You should not create a master key. A master key is necessary when you want to use column encryption in SQL databases.

References

[Welcome to Azure Cosmos DB](#)

[Encrypt a Column of Data](#)

[What is Azure SQL Data Warehouse?](#)

[Quickstart: Run a Spark job on Azure Databricks using the Azure portal](#)

Question92 Implement non-relational data stores - D

You are a data engineer for an exam development company. You create an Azure Cosmos DB account that uses the session consistency level. You create a database and collection that allows exam developers to create and store exam content.

Developer A and Developer B reside in Virginia, United States. Developer C resides in Madrid, Spain. At 12:00, the question entity has its difficulty attribute set to Hard. All three developers read the value Hard. Developer A then changes the difficulty attribute to Medium. All three developers then immediately read the entity before replication occurs.

You need to answer questions regarding the reads.

For each of the following statements, select Yes if the statement is true. Otherwise, select No.

Statement	Yes	No
Developer A will read Medium.	<input type="radio"/>	<input type="radio"/>
Developer B will read Medium.	<input type="radio"/>	<input type="radio"/>
Developer C will read Hard.	<input type="radio"/>	<input type="radio"/>

Explanation

Developer Rahul will read Medium. With the session consistency level, the same user is guaranteed to read the same value within a single session. Even before replication occurs, the user that writes the data can read the same value.

Developer Amit and Pavan will not read the value Medium. They will read the value Hard. Only the same user within the same session is guaranteed to read the same value within a single session. This is because the data has not yet replicated.

References

Consistency levels in Azure Cosmos DB

Question93 - Develop batch processing solutions - D

You are implementing a data engineering solution for your company. Every Thursday night, a third-party company loads shipping data to an Azure blob storage container. This data is in the form of large comma-separated-value (CSV) files. You want to run analytics on that data to automatically generate bar charts that you can visualize. If the data looks good visually, you want to import that data into a SQL Data Warehouse database table.

You need to schedule a job to generate the charts.

What should you do?

Choose the correct answer

Create a scheduled WebJob.

Create a scheduled Azure Databricks job.

Create a scheduled runbook in Azure Automation.

Create a scheduled function.

Explanation

You should create a scheduled Azure Databricks job. With Databricks, you can easily visualize data from external sources in charts by using familiar languages such as Scala or SQL. To do this, you create a notebook, assign it to a cluster, and then write code in cells. With the job, you can define the notebook to run.

You should not create a scheduled runbook in Azure Automation. A runbook can run on a schedule. You can create them graphically, with PowerShell, or with Python. They are simply code components that execute within Azure. Graphical runbooks generate PowerShell code. Runbooks do not work in this scenario because they do not generate visual graphs.

You should not create a scheduled function. A function is a pay-per use service that can also run on a schedule. You can create a function by using C#, F#, JavaScript, Java, and Python. Functions do not work in this scenario because they do not generate visual graphs.

You should not create a scheduled WebJob. A WebJob can also run on a schedule. It runs in the context of an Azure App Service. You can create a WebJob by using C#, JavaScript, Java, Python, Bash, PHP, PowerShell, TypeScript, .cmd,

and .bat. WebJobs do not work in this scenario because they do not generate visual graphs.

References

[Quickstart: Run a Spark job on Azure Databricks using the Azure portal](#)

[What are Microsoft Flow, Logic Apps, Functions, and WebJobs?](#)

[Azure Automation runbook types](#)

Question94 - Develop batch processing solutions-

You are a data engineer. You have an Azure Databricks account with an imported notebook. You also create an Azure data factory.

You need to ensure that Data Factory can access Databricks.

What should you create?

Choose the correct answer

An access policy on a blob storage container

A master key

A blob storage container

An access token

Explanation

You should create an access token. An access token allows external applications to access Databricks. To generate an access token, browse to the Databricks portal, go to User Settings, and choose Generate New Token. The access token is visible only once.

You should not generate a master key. A master key is used in SQL Database to perform column encryption. A master key is not required in this scenario.

You should not create an access policy on a blob storage container. Access policies allow you to add an additional level of control to blobs. They are not used with Data Factory and Databricks.

You should not create a blob storage container. Blob storage containers allow you to store arbitrary data in Azure. Blob storage containers are not required to use Databricks and Data Factory.

References

[Transform data by using Databricks in Azure Data Factory](#)

[Encrypt a Column of Data](#)

[Introduction to Azure Blob storage](#)

[Quickstart: Create a data factory by using the Azure Data Factory UI](#)

[Define a stored access policy](#)

Question95 - Develop batch processing solutions-

You are a data engineer. You use Azure Data Factory to copy and transform data from Azure blob storage to an on-premises server.

You need to ensure that you can successfully copy data.

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

Choose the correct answers

Create a self-hosted integration runtime in Azure Data Factory UI

Create an Azure integration runtime.

Install the self-hosted integration runtime on an Azure virtual machine (VM).

Install the self-hosted integration runtime on the local network

Create an Azure-SSIS integration runtime.

Explanation

You should create a self-hosted integration runtime in Azure Data Factory UI and install the self-hosted integration runtime on the local network. The integration runtime is the execution environment that provides the compute infrastructure for Data Factory. When you use the Copy activity to copy data between Azure and a private network, you must use the self-hosted integration runtime. You do this by first creating one in Azure Data Factory UI. You must then install the runtime on your local network.

You should not create an Azure integration runtime. This is required when you need to copy data between Azure and public cloud services.

You should not create an Azure-SSIS integration runtime. This is required when

you want to run existing SSIS packages natively.

You should not install the self-hosted integration runtime on an Azure VM. Because you are copying data to an on-premises server, you must install the self-hosted integration runtime on the local network.

References

[Integration runtime in Azure Data Factory](#)

[Create and configure a self-hosted integration runtime](#)

[How to create and configure Azure Integration Runtime](#)

[Create Azure-SSIS Integration Runtime in Azure Data Factory](#)

Question96 - Develop batch processing solutions-

You are migrating a corporate research analytical solution from an internal datacenter to Azure.

200 TB of research data is currently stored in an on-premises Hadoop cluster. You plan to copy it to Azure Storage. Your internal datacenter is connected to your Azure Virtual Network (VNet) with Express Route private peering. The Azure Storage service endpoint is accessible from the same VNet.

Corporate policy dictates that the research data cannot be transferred over public internet.

You need to securely migrate the research data online.

What should you do?

Choose the correct answer

Transfer the data using Azure Data Box Heavy devices.

Transfer the data using Azure Data Box Disk devices.

Transfer the data using Azure Data Factory in distributed copy (DistCopy) mode, with an Azure Data Factory self-hosted Integration Runtime (IR) machine installed in the on-premises datacenter.

Transfer the data using Azure Data Factory in native Integration Runtime (IR) mode, with an Azure Data Factory self-hosted IR machine installed on the Azure VNet.

Explanation

You should transfer the data using Azure Data Factory in native IR mode, with an Azure Data Factory self-hosted IR machine installed on the Azure VNet. This approach supports data transfer via Express Route and uses Azure Data Factory IR as an engine to copy the data.

You should not transfer the data using Azure Data Factory in DistCopy mode, with an Azure Data Factory self-hosted IR machine installed in the on-premises datacenter. The DistCp tool does not support Express Route private peering with an Azure Storage VNet endpoint.

You should not transfer the data using Azure Data Box Disk devices. This approach is for offline transfer scenarios. Also, an Azure Data Box Disk device has a total capacity of only 40 TB.

You should not transfer the data using Azure Data Box Heavy devices. While these devices have a capacity of 1 PB, they are intended for offline transfer scenarios only.

References

[Integration runtime in Azure Data Factory](#)

[Use Azure Data Factory to migrate data from on-premises Hadoop cluster to Azure Storage](#)

[What is Azure Data Box Disk?](#)

[What is Azure Data Box Heavy?](#)

Question97 - Develop batch processing solutions-

Your company is implementing a new Azure Databricks workspace.

This environment should be able to access private data stored in Azure Data Lake Store Gen2. Credentials for Azure Databricks should be available in a secure way.

You create a secret scope in the workspace and add the storage account key and a service principal secret.

You need to configure a notebook in this workspace to read from Azure Data Lake Store Gen2.

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

Choose the correct answers

Generate a Shared Access Signature (SAS).

Read from the storage account using the RDD API.

Mount a filesystem using a service principal.

Configure the storage account key with `spark.conf.set()`.

Configure the storage account key with `spark.sparkContext.hadoopConfiguration()`.

Explanation

You should configure the storage account key with `spark.conf.set()` to ensure that the storage account reads the storage account key from the secret scope with the `dbutils.secrets.get()` method.

You should also mount a filesystem using a service principal to securely mount the filesystem using the service principal secret stored in the secret scope. After that, you can read the mounted filesystem with the `spark.read.text()` method and other Spark APIs.

You should not configure the storage account key with `spark.sparkContext.hadoopConfiguration()`. You could use this to directly read from Azure Data Lake Store Gen2. However, your storage account key will be exposed to all users who access the cluster, resulting in a security breach.

You should not read from the storage account using RDD API. RDD API requires you to use `spark.sparkContext.hadoopConfiguration()`. This results in unnecessary exposure of the storage account key to all users who access the cluster.

You should not generate a SAS. You could use a SAS to authenticate with Azure Blob Storage. Azure Data Lake Store Gen2 requires using storage account keys.

References

[Azure Data Lake Storage Gen2](#)

[Secret Scopes](#)

[Secrets](#)

Question98 - Develop batch processing solutions-

Your team is implementing event processing to handle real-time streaming data from Twitter.

You configure an Azure Event hub named TweetStream to ingest the streaming data. You use Azure Databricks to analyze the streaming data from Event Hub. You create a new workbook named TweetAnalysis.

You need to configure the TweetAnalysis workbook to connect with TweetStream.

How should you complete the notebook code? To answer, select the appropriate options from the drop-down menus.

Choose the correct options

```
import org.apache.spark.eventhubs.{ConnectionStringBuilder, EventHubsConf, EventPosition }
```

```
val connectionString = ConnectionStringBuilder("Endpoint=<endpoint>;SharedAccessKeyName=<key-  
name>;SharedAccessKey=<key>")  
    ("TweetStream")
```

```
.build
```

```
val eventHubsConf = EventHubsConf(connectionString)
```

```
.setStartingPosition(EventPosition.fromEndOfStream)
```

```
val eventhubs = spark.
```

```
.format("      ")
```

```
.options(eventHubsConf.toMap)
```

```
.load()
```

setEventHubName

readStream

eventHubs

Choose the correct options

```
import org.apache.spark.eventhubs.{ConnectionStringBuilder, EventHubsConf, EventPosition }

val connectionString = ConnectionStringBuilder("Endpoint=<endpoint>;SharedAccessKeyName=<key-
name>;SharedAccessKey=<key>")
.setEventHubName ("TweetStream")
.build

val eventHubsConf = EventHubsConf(connectionString)
.setStartingPosition(EventPosition.fromEndOfStream)

val eventhubs = spark.readStream
.format("eventhubs ")
.options(eventHubsConf.toMap)
.load()
```

Explanation

You should complete the command as shown below:

```
import org.apache.spark.eventhubs.{ConnectionStringBuilder, EventHubsConf,
EventPosition }

val connectionString =
ConnectionStringBuilder("Endpoint=<endpoint>;SharedAccessKeyName=<key-na
me>;SharedAccessKey=<key>")
.setEventHubName("TweetStream")
.build
val eventHubsConf = EventHubsConf(connectionString)
.setStartingPosition(EventPosition.fromEndOfStream)
val eventhubs = spark.readStream
.format("eventhubs")
.options(eventHubsConf.toMap)
.load()
```

You should use the `.setEventHubName()` method first. You can omit this method if you define an entity path in the connection string directly.

You should use the `spark.readStream()` method to read the data. This method is

used by Spark instance in Databricks to read data from a streaming source, which is the source type to read from Event Hub.

You should use the eventhubs format. This defines the format that Databricks will use to read the stream data. Combining readStream, format and options methods, with the later receiving the eventHubsConf variable with the connection string, you are ready to load and read streaming data from Event Hub.

You should not use the .setNamespaceName or .setSasKeyName method in ConnectionStringBuilder. These methods are used to manually configure your connection string with the builder.

You should not use the writeStream method. This method writes data in a stream and does not read data.

You should not use the read method. This method reads batches of data, not data from a continuous stream.

You should not use the memory format. This defines an output for a stream as an in-memory table.

You should not use the Kinesis format. This loads streams from the Amazon Kinesis streaming service.

References

[Azure Event Hubs](#)

[Structured Streaming](#)

[Continuous Applications: Evolving Streaming in Apache Spark 2.0](#)

Question99 - Develop batch processing solutions-

You work as a data engineer in a company that uses Azure Data Factory for data pipelines.

The company needs to connect with an on-premises database and move data periodically to an Azure SQL Database. The data pipeline is allowed to run at specific, fixed-size time intervals.

You need to implement the Azure Data Factory component to connect with the on-premises database and use the appropriate pipeline execution.

Which component and execution should you use? To answer, select the appropriate options from the drop-down menus.

Choose the correct options

Component:

Execution:

linked service

tumbling window

Explanation

You should configure a linked service as the component. A linked service stores the connection information from the source dataset, like user credentials, server address and database name. You need to configure a linked service for the on-premises database and this linked service will be used by the dataset.

You should configure a tumbling window as the pipeline execution trigger. A tumbling window can define the starting time in the WindowStart setting and the ending time in the WindowEnd setting, defining a time frame to run the data pipeline.

You should not configure a pipeline or an activity. An activity is the task that is executed, like copying data or performing a lookup. A pipeline is a group of activities linked together to form a data pipeline. Activities use datasets to read or write data as the result of a pipeline. Neither is used to connect with an on-premises database.

You should not configure a manual execution or schedule an execution trigger. These executions allow you to manually start pipelines or schedule their execution. However, you cannot define a specific, fixed-size time interval to run the pipelines.

References

[Linked services in Azure Data Factory](#)

[Pipelines and activities in Azure Data Factory](#)

[Datasets in Azure Data Factory](#)

[Pipeline execution and triggers in Azure Data Factory](#)

[Create a trigger that runs a pipeline in response to an event](#)

[Create a trigger that runs a pipeline on a schedule](#)

[Create a trigger that runs a pipeline on a tumbling window](#)

Question100 - Develop batch processing solutions-

Your company hosts an enterprise resource planning (ERP) system with an on-premises SQL Server configured with SQL Server Integration Services (SSIS) packages to extract data from the ERP to an on-premises SQL Server Data Warehouse.

You need to integrate the SSIS packages with Azure Data Factory by configuring the self-hosted integration runtime (IR) as a proxy for Azure-SSIS IR. You already created an Azure Blob storage for the integration.

Which five actions should you perform? To answer, move the appropriate actions from the list of possible actions to the answer area and arrange them in any order.

Create a list in any order

Possible actions

Actions to perform

- Install the self-hosted IR in the on-premises data warehouse.
- Create a linked service in Azure Data Factory with an on-premises data warehouse.
- Create a linked service in Azure Data Factory with Azure Blob Storage.
- Create an Azure-SSIS IR in Azure Data Factory.
- Install the self-hosted IR in the on-premises SSIS.
- Set up the self-hosted IR as a proxy for your Azure-SSIS IR.
- Register the self-hosted IR with the authentication key.

Explanation

You should perform the following actions:

- Create an Azure-SSIS IR in Azure Data Factory.

- Install the self-hosted IR in the on-premises SSIS.
- Register the self-hosted IR with the authentication key.
- Create a linked service in Azure Data Factory with Azure Blob Storage.
- Set up the self-hosted IR as a proxy for your Azure-SSIS IR.

You should create an Azure-SSIS IR in Azure Data Factory to start the configuration for the self-hosted IR as a proxy for Azure-SSI IR.

You should then install the self-hosted IR in the on-premises SSIS. This will prepare the on-premises SSIS to be registered as a self-hosted IR in Azure Data Factory.

Next, you should create and register the self-hosted IR with the authentication key. This will create the self-hosted IR and configure it with the on-premises SSIS instance.

You should then create a linked service in Azure Data Factory with Azure Blob Storage. A self-hosted IR as a proxy for Azure-SSI IR will split the SSIS package data flow task in two stages. In the first stage, the on-premises SSIS will move the on-premises data into a staging area in Azure Blob Storage. In the next stage, the Azure-SSIS IR will move the data from the staging Blob Storage to the destination.

Finally, you should set up the self-hosted IR as a proxy for your Azure-SSIS IR. After your Azure-SSIS and self-hosted IR are created and configured, you can configure the Azure-SSIS IR to use the self-hosted IR as a proxy. In this step, you also need to select the Azure Blob Storage that will be used as staging area to move the data from the self-hosted IR to the Azure-SSIS IR in cloud.

You should not install the self-hosted IR or create a linked service in Azure Data Factory with the on-premises data warehouse. The data warehouse will not be used to process Azure Data Factory pipelines as a self-hosted IR or as a linked service. You only need to configure a self-hosted IR as a proxy for the Azure-SSIS IR.

References

[Configure Self-Hosted IR as a proxy for Azure-SSIS IR in ADF](#)

[Create and configure a self-hosted integration runtime](#)

[Provision the Azure-SSIS Integration Runtime in Azure Data Factory](#)

Question101 - Develop batch processing solutions-

Your team manages a data pipeline in Azure Data Factory that is configured with an on-premises SQL server and an Azure SQL database as linked services.

The data pipeline will be used to incrementally copy data from an on-premises SQL server table named Customers to an Azure SQL database. The Customers table was created with the T-SQL statement shown in the exhibit.

You implement a watermark approach to load delta data from the Customers table to Azure SQL Database. You create the Watermark table and a stored procedure to update the watermark table in the on-premises table.

You need to implement the activities in the data pipeline to incrementally copy data.

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

Create a list in any order

Possible actions

Actions to perform

- Look up the new watermark value in the Customers table from the PersonID field.
- Look up the old watermark value from the Customers table.
- Copy the delta data using the watermark values.
- Look up the new watermark value in the Customers table from the LastModifyTime field.
- Execute a stored procedure to update the watermarks in the Watermark table.
- Look up the old watermark value from the Watermark table.

Explanation

You should perform the following actions:

- Look up the old watermark value from the Watermark table.
- Look up the new watermark value in the Customers table from the LastModifyTime field.
- Copy the delta data using the watermark values.
- Execute a stored procedure to the update watermarks in the Watermark table.

You should look up the old watermark from the Watermark table to read the last inserted data moment and also look up the new watermark value in the Customers table's LastModifyTime field. This will create a time range that retrieves the delta data that needs to be copied.

You should then copy the delta data using the watermark values. You can select the data from the previous watermark lookup values and copy them to the Azure SQL Database destination.

Finally, you should execute a stored procedure to update the watermarks in the Watermark table. This will update the Watermark table with the last timestamp where the data pipeline was executed. When the pipeline runs again, this new timestamp will be used in lookup, loading only new data into Azure SQL Database.

You should not look up the old watermark value from the Customers table. You cannot determine which reference time to use for the old watermark value when reading from the Customers table.

You should not look up the new watermark value in the Customers table from the PersonID field. A watermark value is better determined using the maximum value of a column that controls where the latest modification occurred, like the LastModifyTime column in the Customers table. In other scenarios, you could use a numeric id if the latest modification time is not available, using the latest id as a reference.

References

[Incrementally load data from an Azure SQL database to Azure Blob storage](#)

[Incrementally load data from a source data store to a destination data store](#)

Question102 - Develop batch processing solutions-

You use a data pipeline in Azure Data Factory to move data from Azure Blob Storage to an Azure SQL Data Warehouse using PolyBase.

The data is stored in a CSV file named zipCode.csv and has a structure as shown in the exhibit. The CSV file is stored in Azure Blob storage in a path named ZipCode inside a blob storage container.

You need to complete the T-SQL query to ingest the data from Azure Blob Storage with PolyBase.

How should you complete the code? To answer, select the appropriate options from the drop-down menus.

Choose the correct options

```
CREATE EXTERNAL FILE FORMAT TextFileFormat
```

```
WITH (
```

```
    FORMAT_TYPE = DELIMITEDTEXT,
```

```
    FORMAT_OPTIONS (
```

```
        FIELD_TERMINATOR = '          ',
```

```
        STRING_DELIMITER = '"',
```

```
        FIRST_ROW =
```

```
    ));
```

```
CREATE EXTERNAL TABLE [dbo].[ZipCodeExternal] (
```

```
    [          ] [char](5) NOT NULL,
```

```
    [          ] [nvarchar](255) NOT NULL,
```

```
    [          ] [nvarchar](255) NOT NULL
```

```
) WITH (LOCATION='/ZipCode/', DATA_SOURCE = AzureStorage, FILE_FORMAT = TextFileFormat);
```

```
CREATE TABLE [dbo].[ZipCode] AS SELECT * FROM [dbo].[ZipCodeExternal];
```

1

2

Zip code

city

country

```
WITH (  
    FORMAT_TYPE = DELIMITEDTEXT,  
    FORMAT_OPTIONS (  
        FIELD_TERMINATOR = ' ' ,  
        STRING_DELIMITER = ",  
        FIRST_ROW = 2  
    );  
CREATE EXTERNAL TABLE [dbo].[ZipCodeExternal] (  
    [ Zip Code ] [char](5) NOT NULL,  
    [ City ] [nvarchar](255) NOT NULL,  
    [ County ] [nvarchar](255) NOT NULL  
    ) WITH (LOCATION='/ZipCode/', DATA_SOURCE = AzureStorage, FILE_FORMAT = TextFileFormat);  
CREATE TABLE [dbo].[ZipCode] AS SELECT * FROM [dbo].[ZipCodeExternal];
```

Explanation

You should complete the command as shown below:

```
CREATE EXTERNAL FILE FORMAT TextFileFormat  
WITH (  
    FORMAT_TYPE = DELIMITEDTEXT,  
    FORMAT_OPTIONS (  
        FIELD_TERMINATOR = ',',  
        STRING_DELIMITER = "  
        FIRST_ROW = 2  
    );  
CREATE EXTERNAL TABLE [dbo].[ZipCodeExternal] (  
    [Zip Code] [char](5) NOT NULL,  
    [City] [nvarchar](255) NOT NULL,
```

```
[County] [nvarchar](255) NOT NULL
) WITH (LOCATION='/ZipCode/', DATA_SOURCE = AzureStorage,
FILE_FORMAT = TextFileFormat);
CREATE TABLE [dbo].[ZipCode] AS SELECT * FROM [dbo].[ZipCodeExternal];
```

You should use a comma (,) as the field terminator to separate the values of each field as shown as exhibit.

You should set 2 as the first row for the external file format to skip the first row with the data header.

Finally, you should create the external table with the fields Zip Code, City, and County in sequence to map with the fields as they are listed in the CSV file.

References

[Load Contoso Retail data to Azure SQL Data Warehouse](#)

[CREATE EXTERNAL FILE FORMAT \(Transact-SQL\)](#)

[CREATE EXTERNAL TABLE \(Transact-SQL\)](#)

Question103 - Develop batch processing solutions-

You have an Azure Synapse Analytics SQL pool. You need to create two Azure Data Factory (ADF) pipelines to load data into the SQL pool.

- A pipeline to migrate data from SQL Server Analysis Services (SSAS)
- A pipeline for a daily incremental load from an Azure SQL Database

For each of the following statements, select Yes if the statement is true. Otherwise, select No.

Statement	Yes	No
You can use the SSAS data source in an ADF Copy Activity.	<input type="radio"/>	<input type="radio"/>

You can implement the incremental load from the Azure SQL Database by using the change tracking feature combined with an ADF Copy Activity.	<input type="radio"/>	<input type="radio"/>
An ADF Copy Activity can invoke the Polybase feature to load the Azure Synapse Analytics SQL pool.	<input type="radio"/>	<input type="radio"/>

Explanation

The SSAS data source cannot be used in an ADF Copy Activity. SSAS is not a supported data store for the Copy Activity. A possible approach would be to export the data to CSV files first and then use the Copy activity on the CSV files.

The incremental load from the Azure SQL Database can be implemented by using the change tracking feature combined with an ADF Copy Activity. By enabling change tracking, the changed rows in the database are recorded. This can be used in a source query in ADF to periodically transfer the changes to a data warehouse.

The Polybase feature to load the Azure Synapse Analytics SQL pool can be invoked by an ADF Copy Activity. The ADF Copy Activity has an option to enable Polybase.

References

[Copy activity in Azure Data Factory](#)

[Incrementally load data from Azure SQL Database to Azure Blob Storage using change tracking information](#)

[Copy and transform data in Azure Synapse Analytics \(formerly Azure SQL Data Warehouse\) by using Azure Data Factory](#)

Question104 - Develop batch processing solutions - D

You are working with Azure Data Factory.

Sales data from two regions must be imported into an Azure Synapse Analytics SQL pool. The data is stored in two CSV files.

You have the following requirements:

- All data from the CSV files must be stored in a single destination table.
- Duplicate records must be inserted into the destination table.

You need to implement a mapping data flow to import the data.

Which data flow transformation should you use?

Choose the correct answer

Join

Union

Aggregate

Lookup

Explanation

You should use a union transformation. This transformation collects data from multiple streams. Input columns are mapped to output columns based on a name or a position.

You should not use a join transformation. This transformation combines multiple input rows into a single output row based on a join type and condition.

You should not use a lookup transformation. This transformation is used for adding data to rows from another stream.

You should not use an aggregate transformation. This transformation is used for aggregate calculations. Aggregation involves two steps: First multiple rows are grouped into one row. Then an aggregate calculation such as SUM or COUNT is applied to the group of rows.

References

[Azure Data Factory mapping data flow union transformation](#)

[Join transformation in mapping data flow](#)

[Azure Data Factory mapping data flow Lookup Transformation](#)

[Aggregate transformation in mapping data flow](#)

Question105 Implement non-relational data stores -D

Users named Alan and Kerry are members of the Marketing Azure Active Directory (Azure AD) security group, set as a primary one for their Azure AD accounts. Kerry and another user named David are members of the Finance security group, set as a primary for David's account.

You set up a new directory in Azure Data Lake Storage and set the owning group to Finance. Kerry creates a new text file in that directory as an extract from the Sales database. Your audit report indicates that the access control list (ACL) for that file is set to 640 in the POSIX format.

You need to determine what access permissions Alan, Kerry, and David all have to the newly uploaded text file.

What permissions do they have? To answer, drag the appropriate permission option to each user's ACL column. A permission option may be used once, more than once, or not at all.

Drag and drop the answers

Read + Write

Read Only

No Access

Read + Write

Read Only

No Access

Read + Write + Execute

Read + Execute

Write Only

Employee	Access Control List
Alan	No Access
Kerry	Read + Write
David	Read Only

Read + Write

Read Only

No Access

Read + Write + Execute

Explanation

Alan has No Access, Kerry has Read + Write access, and David has Read Only.

File permissions consist of three digits: Owner (The user who created the item automatically becomes an Owner), Owner Group, and Everyone Else. In Azure Data Lake Storage Gen2, the Owner Group is inherited from the parent folder.

Azure Data Lake Storage Gen2 supports settings ACLs in the POSIX-compatible format, which assigns the following numbers to the different permission combinations:

- Read Only: 4
- Write Only: 2
- Execute Only: 1
- No Access: 0
- Read + Write: $4 + 2 = 6$
- Read + Execute: $4 + 1 = 5$
- Etc.

The 640 permission set to the text file created by Kerry can be translated as:

- Owner (Kerry): Read + Write (6)
- Owner Group (Finance, which includes David): Read Only (4)
- Everyone Else (Alan): No Access (0)

References

[Access control in Azure Data Lake Storage Gen2](#)

Question106 Implement non-relational data stores - D

You are a data engineer for an insurance company. You create an Azure Cosmos DB account that uses the strong consistency level. You create a database and a collection that allows personnel to manage insurance claims. Employee A and Employee B reside in Virginia, United States. Employee C resides in Madrid, Spain. At 12:00, a claim entity has its status attribute set to Pending. Employee A then changes the status attribute to Closed. All three employees then immediately read the entity before it is committed and before synchronization occurs.

You need to answer questions regarding the reads.

To answer, select Yes if the statement is true. Otherwise, select No.

Statement	Yes	No
Employee A reads Closed.	<input type="radio"/>	<input type="radio"/>
Employee B reads Pending.	<input type="radio"/>	<input type="radio"/>
Employee C reads Pending.	<input type="radio"/>	<input type="radio"/>

Explanation

Employee A does not read Closed. With the strong consistency level, reads are guaranteed to return the most committed version of an item. Even if Employee A is the user that changed the attribute, that employee still reads the original value until the item is committed and synchronization takes place. Therefore, Employee A,

Employee B, and Employee C all read Pending until the item is committed and synchronization takes place.

References

[Choose the right consistency level](#)

Question107 Implement non-relational data stores - D

You are a data engineer for your company. You are creating an Azure Cosmos DB account to store an existing product catalog. The existing product catalog currently exists as a single Oracle database table. Approximately 20 percent of the columns in the table are empty.

Each product type can have different attribute names and a different attribute count. You must be able to search the catalog by product id and category. You must be able to search for products in the Clothing category by size. You must be able to search for products in the Laptop category by CPU speed. You also must be able to query data by using the following syntax from a web application:

```
SELECT p.productName FROM Products p
```

You need to choose the most appropriate API.

Which API should you choose?

```
SELECT p.productName FROM Products p
```

You need to choose the most appropriate API.

Which API should you choose?

Choose the correct answer

MongoDB API

Table API

Core SQL API

Gremlin API

Explanation

You should choose the Core SQL API. This API is recommended because it supports a schema-less data store. In this scenario, different types of products can have different attributes. This is indicated by the fact that not all columns in the existing database are used. Some products can have different columns populated. This is referred to as semi-structured data. This API allows you to use SQL-like

queries to access and filter data. The data is returned as JavaScript Object Notation (JSON) documents.

You should not use the Gremlin API. This API does not support SQL-like queries. This API uses the Gremlin Graph Traversal Language to query data from a graph database.

You should not use the Table API. This API allows you to query data by using OData and Language Integrated Query (LINQ). It does not support SQL-like queries from web applications.

You should not use the MongoDB API. This API does not allow you to use SQL-like queries to access and filter data.

References

[Identify the technology options](#)

[Use Core \(SQL\) to store a product catalog](#)

[Non-relational data and NoSQL](#)

Question108 Develop streaming solutions - D

You are a data engineer for your company. You are creating a Stream Analytics query. You want to use a windowing function that allows you to capture events that repeat and that do not overlap. You also want to capture time periods when there are no events.

You need to choose the appropriate windowing function.

Which windowing function should you choose?

Choose the correct answer

Session

Tumbling

Sliding

Hopping

Explanation

You should use the tumbling windowing function. This function segments event data into distinct time segments and then performs a function against them. It allows events to repeat but not overlap.

You should not use the hopping windowing function. This function is similar to a

tumbling window function, but events can overlap.

You should not use the sliding windowing function. This function produces output only when events occur.

You should not use the session windowing function. This function groups together streaming events that arrive at similar times. It filters out time periods when there is no data.

References

[Introduction to Stream Analytics windowing functions](#)

Question109 Develop streaming solutions-

A car manufacturer implements an IoT solution in its production line. The solution uses Azure IoT Hub to connect and manage IoT devices. The IoT devices are capable of running Docker images.

You need to deploy an Azure Stream Analytics job to provide real-time analytics.

The solution must minimize latency and bandwidth usage between the job and IoT devices. The Stream Analytics job needs to stream events to the IoT Hub. In the future, an Azure function will be implemented to process data from the IoT Hub.

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

Create a list in the correct order

Possible actions

Actions in order

- Configure Streaming Units (SUs).
- Create a Stream Analytics job with cloud hosting.
- Create an Azure Blob Storage container.
- Create a Stream Analytics job with edge hosting.
- Configure the Azure Blob Storage container as save location for the job definition.
- Set up an IoT Edge environment on the IoT devices and add a Stream Analytics module.
- Configure routes in IoT Edge.

Explanation

You should perform the following actions in order:

1. Create an Azure Blob Storage container.
2. Create a Stream Analytics job with edge hosting.
3. Configure the Azure Blob Storage container as save location for the job definition.
4. Set up an IoT Edge environment on the IoT devices and add a Stream Analytics module.
5. Configure routes in IoT Edge.

To minimize latency and bandwidth usage between the Stream Analytics job and the IoT devices, you should configure IoT Edge. IoT Edge is a solution that analyzes data on devices instead of in the cloud, allowing a quicker reaction to events.

You should create an Azure Blob Storage container and a Stream Analytics job with edge hosting and configure the Azure Blob Storage container as save location for the job definition. This is required by IoT Edge to create your jobs and synchronize the job definition to IoT devices via an Azure Blob Storage container.

You should set up an IoT Edge environment on the IoT devices and add a Stream Analytics module. After installing IoT Edge on the IoT devices and adding a Stream Analytics module, you should select the Stream Analytics job to run directly on the IoT devices.

Finally, you should configure routes in IoT Edge. This will upstream events from a Stream Analytics job to an IoT Hub in the cloud, allowing Azure Functions to process the events.

You should not create a Stream Analytics job with cloud hosting. Cloud hosting is used to run a Stream Analytics job in the cloud. You should run the Stream Analytics job directly in the IoT device to minimize latency and bandwidth usage.

You should not configure Streaming Units (SUs). A Stream Analytics job with edge hosting does not consume SUs. You only need to configure SUs with cloud hosted Stream Analytics jobs.

References

[Azure Stream Analytics on IoT Edge](#)

[Tutorial: Deploy Azure Stream Analytics as an IoT Edge module](#)

[What is Azure IoT Edge](#)

Question 110 Implement non-relational data stores - D

You are building a new CRM solution with Azure Cosmos DB as a backend. You plan to use an Apache TinkerPop compatible framework.

Your consumers, promotions, and products will become vertices in your CRM's graph structure, while the references to the events your consumers attended and the specific products they bought will form the graph's edges.

You need to choose the right API in Azure Cosmos DB to build the graph database algorithms.

What should you do?

Choose the correct answer

Use Table API.

Use MongoDB API.

Use SQL API.

Use Cassandra API.

Use Gremlin API.

Explanation

You should use Gremlin API because it is based on the Apache TinkerPop graph database standard and meets the solution's requirements.

You should not use Cassandra, MongoDB, SQL or Table APIs, because they are not compatible with the required TinkerPop standard and are not intended for graph operations.

References

[Introduction to Azure Cosmos DB: Gremlin API](#)

[Introduction to the Azure Cosmos DB Cassandra API](#)

[Azure Cosmos DB's API for MongoDB](#)

[Introduction to Azure Cosmos DB: Table API](#)

[Getting started with SQL queries](#)

Question 111 Develop streaming solutions-

You need to implement an event processing solution using Azure Stream Analytics to analyze phone calls and identify fraudulent calls.

This solution must ingest phone calls from an Azure Event hub, analyze the data in real time, and visualize the fraudulent calls with Power BI.

You need to implement the Stream Analytics job.

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

Create a list in any order

Possible actions

Actions to perform

- Configure Azure Blob Storage as the reference input.
- Set up Azure Blob Storage as the stream output.
- **Configure Azure Event Hubs as the stream input.**
- **Define a query to identify fraudulent calls.**
- **Set up Power BI as the stream output.**
- **Start the job.**

Explanation

You should perform the following actions in order:

- Configure Azure Event Hubs as the stream input.
- Set up Power BI as the stream output.
- Define a query to identify fraudulent calls.
- Start the job.

You should configure Azure Event Hubs as the stream input. The Stream Analytics job requires a stream input to receive data. The phone call data to be analyzed is generated by Azure Event Hubs.

You should set up Power BI as the stream output. Power BI can directly access Stream Analytics if an output is defined and it receives real-time information about fraudulent calls.

You should define a query to identify fraudulent calls. You need to use Stream Analytics Query Language, a T-SQL subset syntax, to analyze the phone call data to determine if it is a legitimate or fraudulent call.

You should start the job. To start ingesting data from Azure Event Hubs, analyze it, and view the real-time data in Power BI, you need to start the Stream Analytics job. After that, you will see the real-time data in the Power BI dashboard.

You should not configure Azure Blob Storage as the reference input. Reference data is complimentary data that could be used to look up extra data in the Stream Analytics job. You do not need to import reference data to identify fraudulent phone calls.

You should not set up Azure Blob Storage as the stream output. You could use this to log the fraudulent call analysis in a Blob Storage. However, you need to analyze the data in Power BI.

References

[Analyze phone call data with Stream Analytics and visualize results in Power BI dashboard](#)

[Stream Analytics Query Language Reference](#)

[Using reference data for lookups in Stream Analytics](#)

[Understand outputs from Azure Stream Analytics](#)

Question112 Develop streaming solutions-

Your company is implementing Azure Stream Analytics to analyze a continuous flow of data from a social media platform.

This incoming social media data stream uses an Azure service that is compatible with

Apache Kafka protocols and streams events in binary format. The social media platform provides a Schema Registry to define the data schema. The data stream platform is managed by the social media company.

You need to create a new input stream in Azure Stream Analytics to directly consume the social media data stream.

Which input stream type and event serialization format should you implement to meet the requirements? To answer, select the appropriate options from the drop-down menus.

Choose the correct options

Input stream type:

Event Hub

Avro

Explanation

You should use Event Hub as the input stream type. Event Hub is a managed streaming platform and event ingestion service with the ability to process millions of events per second. Event Hub provides compatibility with the Kafka protocol in the standard and dedicated pricing tiers. Event Hub could be used directly by Kafka applications and Azure Stream Analytics as well without configuring custom connectors.

You should use Avro as the serialization format. Avro is a binary serialization format that relies on schemas to define what fields are present and their type. Event Hub focuses on the streaming itself, but you can use the Schema Registry in the Kafka applications to validate the schema.

You should not use IOT Hub or Blob Storage. Both input streams are supported by Azure Stream Analytics. However, they are not compatible with Apache Kafka protocols. You can configure a custom connector in Kafka Connect to integrate a producer with IoT Hub from an Apache Kafka cluster. However, you need to directly consume from the social media data stream.

You should not use CSV or JSON as the serialization format. These formats are plain-text and the social media platform sends data in binary format.

References

[Understand inputs for Azure Stream Analytics](#)

[Data streaming with Event Hubs using the Kafka protocol](#)

[Process Apache Kafka for Event Hubs events using Stream analytics](#)

[Use Azure Event Hubs from Apache Kafka applications](#)

[Event Hubs pricing](#)

[Kafka Connect](#)

[Azure/toketi-kafka-connect-iothub](#)

Question 113 Implement non-relational data stores - D

Your team is planning to use Azure Cosmos DB as the data store for a multi-region application.

You need to choose a default consistency model with the lowest latency between the application and Cosmos DB.

Which consistency model should you use?

Choose the correct answer

Eventual

Consistent Prefix

Session

Strong

Bounded Staleness

Explanation

You should use the eventual consistency level. This is the lowest consistency level available and will result in as little latency as possible, without the guarantee of reading operations using the latest committed write.

Azure Cosmos DB has five levels of data consistency options:

- Strong
- Bounded staleness
- Session
- Consistent prefix
- Eventual

Higher consistency levels guarantee that the latest committed write was read by the application, but you have a performance tradeoff of a few milliseconds using a higher consistency level.

References

[Consistency levels in Azure Cosmos DB](#)

[Consistency, availability, and performance tradeoffs](#)

Question114 Monitor data storage -

You are a data engineer. You manage an Azure SQL Database named Sample.

You need to monitor performance by capturing a history of query plan changes over time.

What should you do?

Choose the correct answer

Run the following SQL statement:

```
ALTER DATABASE Sample SET QUERY_STORE = ON (OPERATION_MODE = READ_WRITE);
```

Open SQL Server Profiler from SQL Server Management Studio and choose the Performance statistics event.

Run the following SQL statement:

```
CREATE STATISTICS Sample WITH FULLSCAN
```

Open SQL Server Profiler from SQL Server Management Studio and choose the Showplan All event.

Explanation

You should run the following SQL statement:

```
ALTER DATABASE Sample SET QUERY_STORE = ON (OPERATION_MODE = READ_WRITE);
```

This statement enables Query Store, which provides statistics on query performance. It helps you identify performance differences that are caused by query changes. It is disabled by default.

You should not open SQL Server Profiler from SQL Server Management Studio (SSMS) and choose the Showplan All event or the Performance statistics event. SQL Server Profiler provides a graphical interface that allows you to trace SQL-related events, such as query executions and logins.

You should not run the following SQL statement:

```
CREATE STATISTICS Sample WITH FULLSCAN
```

The CREATE STATISTICS statement allows you to create optimization statistics on one or more columns of a table. You should specify a table name with columns, not a database name.

References

[Monitoring performance by using the Query Store](#)

[Operating the Query Store in Azure SQL Database](#)

[SQL Server Profiler](#)

[CREATE STATISTICS \(Transact-SQL\)](#)

Question115 Monitor data storage-

You are a data engineer for your company. You manage a Microsoft SQL Server 2019 database that is hosted on an Azure virtual machine (VM) and used by an on-premises web application. The application is undergoing a major change. You consider using Query Store to examine performance before and after the change.

You need to determine the scenarios where Query Store can help.

For each of the following scenarios, select Yes if Query Store can help. Otherwise, select No.

Scenario	Yes	No

Creating more indexes on 10 tables	<input type="radio"/>	<input type="radio"/>
Adding additional databases for the web application	<input type="radio"/>	<input type="radio"/>
Increasing the size of the VM	<input type="radio"/>	<input type="radio"/>

Explanation

Query Store can help you in the following scenarios:

- Creating more indexes on 10 tables
- Increasing the size of the VM

Query Store allows you to compare performance before and after an anticipated change. The size of the VM determines the amount of memory available.

Query Store cannot help you when you want to add additional databases for the web application. It can monitor only one database. To monitor a database, you would run the following statement:

```
ALTER DATABASE <database_name> SET QUERY_STORE = ON
```

References

[Memory optimized virtual machine sizes](#)

[Query Store Usage Scenarios](#)

Question116 Monitor data storage-

You manage an Azure SQL Data Warehouse. You want to use Dynamic Management Views (DMVs) to monitor your workloads.

You need to find the top 10 longest running queries.

Which view should you use?

Choose the correct answer

sys.dm_pdw_request_steps

sys.dm_pdw_exec_sessions

sys.dm_pdw_sql_requests

sys.dm_pdw_exec_requests

Explanation

You should use sys.dm_pdw_exec_requests. This view returns all queries that are currently running or that were recently running. You can use the following SQL statement to return the top 10 longest running queries:

```
SELECT TOP 10 * FROM sys.dm_pdw_exec_requests ORDER BY  
total_elapsed_time DESC;
```

You should not use sys.dm_pdw_exec_sessions. This view returns the last 10,000 connection logins.

You should not use sys.dm_pdw_request_steps. This view returns the query plan steps for a specific query.

You should not use sys.dm_pdw_sql_requests. This view returns the distribution run times for a specific query step.

References

[Monitor your workload using DMVs](#)

Question117 Monitor data storage-

You manage an Azure SQL Database. You want to use Dynamic Management Views (DMVs) to monitor your workloads.

You need to be able to easily identify queries within the views.

What should you do?

Choose the correct answer

Run SET SHOWPLAN_ALL OFF.

Enable Query Store.

Use the LABEL option to assign a comment to the query.

Run SET SHOWPLAN_ALL ON.

Explanation

You should use the LABEL option to assign a comment to the query. This adds a label to the query. For example, you can add the label to a query as follows:

```
SELECT * FROM FactStoreSales OPTION ( LABEL = 'Q4' );
```

You can then easily locate the query's execution steps with the following statement:

```
SELECT * FROM sys.dm_pdw_exec_requests WHERE [label] = 'Q4';
```

You should not enable Query Store. Query Store allows you to compare database performance against an anticipated change.

You should not run SET SHOWPLAN_TEXT ON. This statement allows you to display query execution information without actually executing the query. This statement is intended for applications that display text.

You should not run SET SHOWPLAN_ALL ON. This statement allows you to display query execution information without actually executing the query. This statement is intended for applications that can display text. It provides additional columns of information for each row that is output.

References

[OPTION Clause \(Transact-SQL\)](#)

[SET SHOWPLAN_TEXT \(Transact-SQL\)](#)

[SET SHOWPLAN_ALL \(Transact-SQL\)](#)

[Monitoring performance by using the Query Store](#)

[Monitor your workload using DMVs](#)

Question118 Monitor data storage-

You are a data engineering manager for your company. You manage an Azure SQL Data Warehouse. One of your employees wants to retrieve a list of the last 100 user logins to SQL Data Warehouse.

You need to ensure that the employee has the correct permission.

Which permission should you grant?

Choose the correct answer

VIEW DATABASE STATE

VIEW DEFINITION

ALTER ANY CONNECTION

ALTER ANY USER

Explanation

You should grant the VIEW DATABASE STATE permission. This permission is required to access Dynamic Management Views (DMVs), which allow you to investigate query execution in Azure SQL Data Warehouse. The view that contains logins is sys.dm_pdw_exec_sessions. It actually contains the last 10,000 logins.

You should not grant the VIEW DEFINITION permission. This allows the employee to view metadata of an object. For example, the employee can view table metadata in the sys.objects catalog.

You should not grant the ALTER ANY CONNECTION permission. This allows the employee to manage the database server.

You should not grant the ALTER ANY USER permission. This allows the employee to manage database users.

References

[Monitor your workload using DMVs](#)

[Grant T-SQL permissions for Parallel Data Warehouse](#)

[GRANT Database Permissions \(Transact-SQL\)](#)

Question119 Monitor data storage-

You are upgrading your company's online e-commerce solution. You plan to use the In-Memory features of Azure SQL Database to improve the solution's backend performance.

You convert some disk-based tables into the memory-optimized ones and select the relevant service tier for your Azure SQL Database.

You need to monitor in-memory storage use so that you can verify that Azure SQL Database does not exceed the In-Memory Online Transactional Processing (OLTP) storage cap set for the selected service tier.

Which two actions should you perform to achieve this goal? Each correct answer presents a complete solution.

Choose the correct answers

From the Database -> Monitoring -> Metrics blade, select the In-Memory OLTP Storage percentage metric.

Use the `SELECT xtp_storage_percent FROM sys.dm_db_resource_stats` query.

From the Database -> Monitoring -> Metrics blade, select the DTU Limit metric.

Use the `SELECT max_worker_percent FROM sys.dm_db_resource_stats` query.

Use the `SELECT max_session_percent FROM sys.dm_db_resource_stats` query.

Explanation

You should use the `SELECT xtp_storage_percent FROM sys.dm_db_resource_stats` query to show in-memory storage utilization. It shows storage utilization for In-Memory OLTP as a percentage of the limit of the service tier.

You can also use the In-Memory OLTP Storage percentage metric on the Database -> Monitoring -> Metrics blade. It will create a chart, showing in-memory storage use as a percentage of the storage cap of Azure SQL Database's selected service tier.

You should not use the `SELECT max_worker_percent FROM sys.dm_db_resource_stats` query. It shows maximum concurrent workers (requests) as a percentage of the limit of the database's service tier, which does not meet the goal.

You should not use the `SELECT max_session_percent FROM sys.dm_db_resource_stats` query. It shows the maximum concurrent sessions as a percentage of the limit of the database's service tier, which does not meet the goal.

You should not use the DTU Limit metric. It will create a graph showing the average DTU limit, but it cannot provide information about In-Memory OLTP storage use.

References

[sys.dm_db_resource_stats \(Azure SQL Database\)](#)

[Monitor In-Memory OLTP storage](#)

Question 120 Monitor data storage-

You have a Microsoft SQL Server 2019 database hosted on an Azure virtual machine (VM). The database is the data store for a web application. When customers visit the shopping cart page of the application, the page loads slowly.

You need to determine the stored procedure that is being called when this page is accessed.

What should you do?

Choose the correct answer

Choose Include Actual Execution Plan from the Query menu.

Choose Display Estimated Execution Plan from the Query menu.

Create a SQL Server Profiler trace.

Call the SET SHOWPLAN_TEXT statement in Query Analyzer.

Explanation

You should create a SQL Server Profiler trace. You should view the TextData output of the SQL:BatchCompleted event. This output contains the stored procedure that is executed at the time the event occurred.

You should not choose Display Estimated Execution Plan from the Query menu. This displays a graphical view of the estimated query plan before you run a query. A query plan shows a series of execution steps that are involved in executing queries. This does not allow you to determine the actual stored procedure that is causing the problem.

You should not choose Include Actual Execution Plan from the Query menu. This displays a graphical view of the query plan when you run queries. A query plan shows a series of execution steps that are involved in executing queries. This does not allow you to determine the actual stored procedure that is causing the problem.

You should not call the SET SHOWPLAN_TEXT statement in Query Analyzer. This displays a text view of the query plan when you run queries.

References

[SQL Server Profiler](#)

[SET SHOWPLAN_TEXT \(Transact-SQL\)](#)

[Display the Estimated Execution Plan](#)

[Display an Actual Execution Plan](#)

Question121 Monitor data storage - D

You are a data engineer. You manage an Azure blob storage account for your company.

You need to monitor the availability of the account for the past four hours.

What should you do?

Choose the correct answer

Open the blob storage account and create a new alert rule.

Open Azure Monitor and select Storage Accounts from the Insights section.

Open Azure Advisor and create a new advisor alert.

Open the blob storage account and select the Usage tab.

Explanation

You should open Azure Monitor and select Storage Accounts from the Insights section. This allows you to monitor storage availability and other metrics over a period of time.

You should not open the blob storage account and create a new alert rule. Alert rules allow you to be notified when a particular event occurs.

You should not open Azure Advisor and create a new advisor alert. Azure Advisor provides you with best practices recommendations with regards to security, cost and availability. You cannot use it to monitor historical availability for an individual storage account.

You should not open the blob storage account and select the Usage tab. This displays the amount of storage your account is using as a percentage of the total capacity. It does not display availability.

References

[Introduction to Azure Advisor](#)

[Monitor a storage account in the Azure portal](#)

[Monitoring your storage service with Azure Monitor for Storage \(preview\)](#)

Question122 Monitor data storage-

You manage an Azure Data Lake Gen1 storage account. You want to log API calls to the account. These logs should be stored in a blob storage account. They should only include detailed operations about the API calls.

You need to configure the Diagnostic settings blade.

To answer, select the appropriate Log and Metric options in the answer area.

Choose the correct options

Diagnostics settings

Save

X

Discard

Delete

*

Name

firstSetting

Archive to a storage account

Stream to an event hub

Send to Log Analytics

LOG

Audit

Requests

METRIC

AllMetrics

Diagnostics settings

 Save  Discard  Delete

* Name

firstSetting 

☐ Archive to a storage account

☐ Stream to an event hub

☐ Send to Log Analytics

LOG

☒ Audit

☒ Requests

METRIC

☐ AllMetrics

 Save  Discard  Delete

* Name

firstSetting 

☐ Archive to a storage account

☐ Stream to an event hub

☐ Send to Log Analytics

LOG

☒ Audit

☒ Requests

METRIC

☐ AllMetrics

You should check the Requests checkbox. This allows you to capture every API request made to the account.

You should also check the Audit checkbox to enable audit logging. This allows you to capture a breakdown of the operations made to the APIs.

You should not check the AllMetrics checkbox. This includes metrics other than API calls and operations. In this scenario, you only want to include operation details for the APIs.

References

[Accessing diagnostic logs for Azure Data Lake Storage Gen1](#)

Question123 Implement non-relational data stores - D

Your team is developing an application with Azure Cosmos DB as the data solution. The application will use the MongoDB API for a document-based database to store items.

You need to provision Cosmos DB with the correct container and item types.

Which container and item types should you use? Each correct answer presents part of the solution.

Drag and drop the answers

document

collection

document

collection

edge

graph

table

Drag and drop the answers

collection

Container type

document

Item type

document

collection

edge

graph

table

Explanation

You should use collection for the container type and document for the item type. Cosmos DB refers to container and item types according to which API is used in your account. The MongoDB API refers to the container as a collection and document for each item inside a collection.

You should not use table for the container type. Table is the container type if you are using the Cassandra API or the Table API in Cosmos DB.

You should not use graph for the container type or edge for the item type. Graph is used as a container type and edge is used as an item type if you are using the Gremlin API in Cosmos DB.

References

[Work with databases, containers, and items in Azure Cosmos DB](#)

Question124 Monitor data storage-

You manage an Azure SQL Data Warehouse Gen2 with caching enabled.

Business users report slow performance while running reports. You try to troubleshoot the issue and discover that your working data set does not fit completely into the cache. You solve the issue by scaling up your data warehouse.

You need to implement a monitoring alert that anticipates this situation using Cache hit percentage and Cache used percentage metrics.

How should you configure these metrics? To answer, select the appropriate options from the drop-down menus.

Choose the correct options

Cache hit percentage:

Cache used percentage:

low

high

Explanation

You should configure the monitoring alert with a low cache hit percentage and a high cache used percentage. In this scenario, a large data set that is rarely used could be filling your cache space, leaving higher frequently accessed data out of the cache and resulting in a low cache hit.

You should not configure a high cache hit percentage. Caching with a high hit percentage denotes a good use of caching, with an optimal use when the cache used percentage is too high. You should troubleshoot other areas because the cache is probably running ok.

You should not configure a low cache used percentage. If you have space available to store cache and your queries still run slowly, the issue is unlikely to be cache related, particularly with a high cache hit percentage. You should troubleshoot other areas.

References

[Monitoring resource utilization and query activity in Azure SQL Data Warehouse](#)

[How to monitor the Gen2 cache](#)

Question125 Monitor data storage - D

Your company is responsible for a globally-distributed application to support foreign trade operators. This application uses Cosmos DB as the main database, with throughput provisioned on a container level.

Customers are grouped in different tables based on which trade rules they work with. Customers that use a particular trade rule report errors using the application.

You discover that this issue was caused by under-provisioned Requests Units (RU) in this specific trade rule table.

You need to implement an alert using a metric that best anticipates these issues.

Which metric should you use to implement this alert?

Choose the correct answer

Provisioned throughput by container

Total Request Units by container

Request Count by HTTP status code

Max consumed RU per second by partition

Explanation

You should use the Total Request Units by container metric. This metric indicates how many Request Units were consumed. You could configure an alert for this particular trade rules table, and if RU consumption is above 80 percent, you would run a function to increase the RU for this table, avoiding errors for your customers.

You should not use the Request Count by HTTP status code metric. Configuring an alert based on the 429 HTTP Status code indicates that the container exceeded the configured throughput limit. However, you cannot anticipate when a container is close to reaching the RU limit, resulting in errors for customers.

You should not use the Max consumed RU per second by partition metric. You could use this metric to identify which partition is consuming the most provisioned RU for a given container or database.

You should not use the Provisioned throughput by container metric. This metric is used to monitor how much throughput is provisioned by container. However, you cannot determine how much throughput is consumed.

References

[Monitor and debug with metrics in Azure Cosmos DB](#)

[HTTP Status Codes for Azure Cosmos DB](#)

[Monitoring Azure Cosmos DB](#)

Question 126 Monitor data storage-

You implement Azure SQL Database to store sensitive data for your company. Only users from a specified region can access this database.

Because of a compliance requirement, you need to detect unusual access by monitoring access patterns from the database.

You need to configure the Azure SQL Database server to email the compliance department when a potential risk access occurs.

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

Create a list in the correct order

Possible actions

Actions in order

- Add a server-level firewall rule.
- Set the Advanced Threat Protection type only for anomalous client login.
- Enable auditing.
- Enable Advanced Data Security.
- Configure the feature to send email alerts to compliance.
- Set the Advanced Threat Protection type only for unsafe action.

Explanation

You need to perform the following steps in order:

1. Enable Advanced Data Security.
2. Configure the feature to send email alerts to compliance.
3. Set the Advanced Threat Protection type only for anomalous client login.

You should enable Advanced Data Security. This feature allows you to detect and respond to potential threats when they occur based on anomalous activities.

You should configure the feature to send email alerts to compliance. When you enable Advanced Data Security in Advanced Threat Protection settings, you have the option to send alerts by email. You can configure the email alert here so the compliance department can receive alerts when a threat occurs.

You should set the Advanced Threat Protection type only for anomalous client login. This type of detection raises an alert when a login pattern anomaly is detected,

including when the database is accessed from an unusual location, unusual Azure datacenter, or unfamiliar principal, meeting the compliance team notification requirement.

You should not enable auditing. Auditing stores administrative actions in Azure SQL Database in Azure Blob Storage or Log Analytics for further analysis. You could not determine if an unallowed access to database occurred by checking auditing logs.

You should not set the Advanced Threat Protection type only for unsafe action. This is used to detect the execution, or execution attempt, of SQL commands that could be used in malicious way, like xp_cmdshell.

You should not add a server-level firewall rule. This could limit which IP addresses or ranges have access to the database. However, this could prevent legitimate access from allowed users if they are outside this IP address range.

References

[Azure SQL Database Advanced Threat Protection for single or pooled databases](#)

[Azure SQL Database and Azure SQL Data Warehouse IP firewall rules](#)

[Get started with SQL database auditing](#)

[xp_cmdshell \(Transact-SQL\)](#)

[Alerts listed on this page don't match the protection types in the portal](#)

Question127 Monitor data storage-

You work as a Data Engineer for an image processing company.

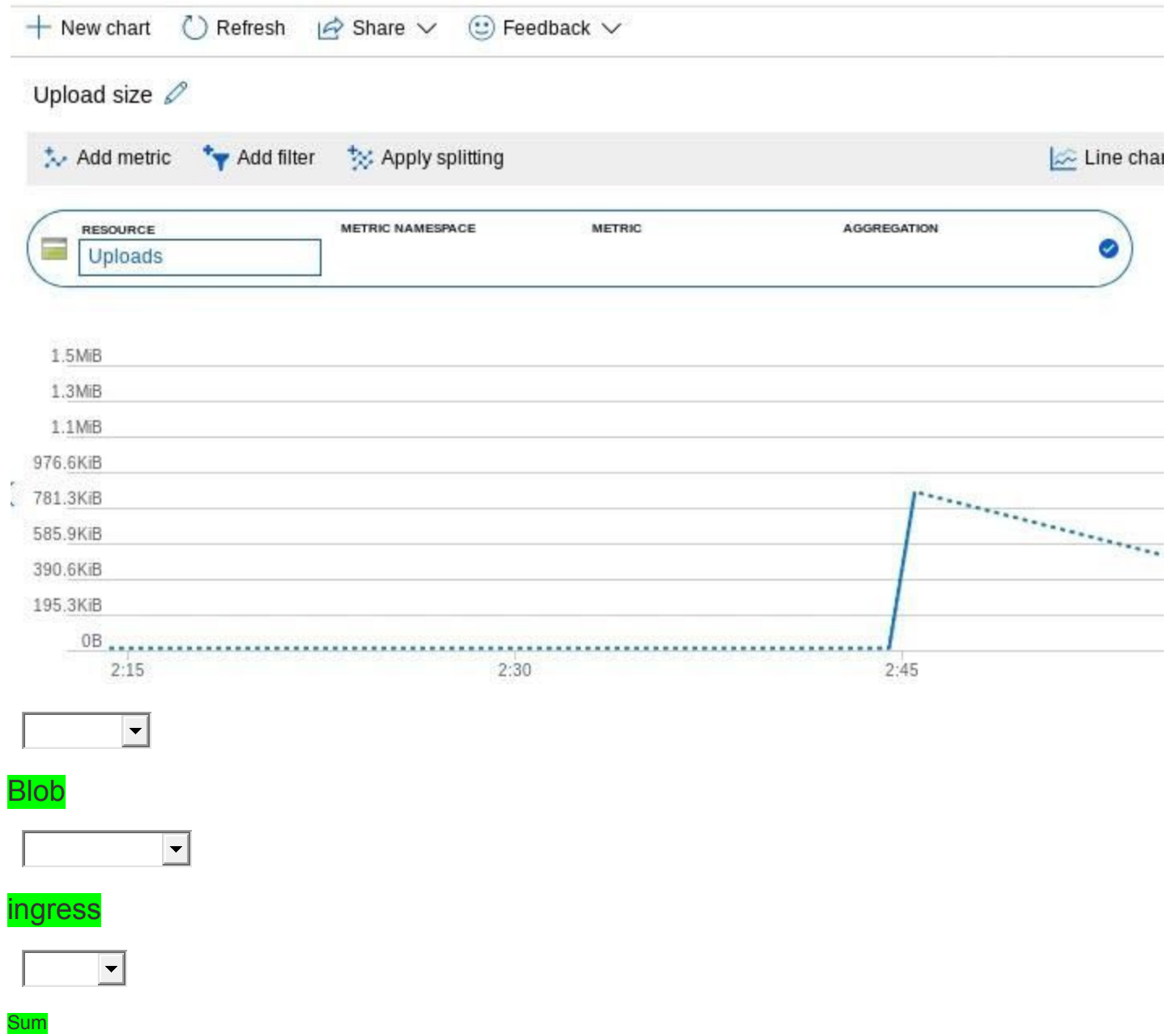
Users upload files to be analyzed by Azure HDInsight in an Azure Storage Account named Uploads. Files are stored in a blob storage container.

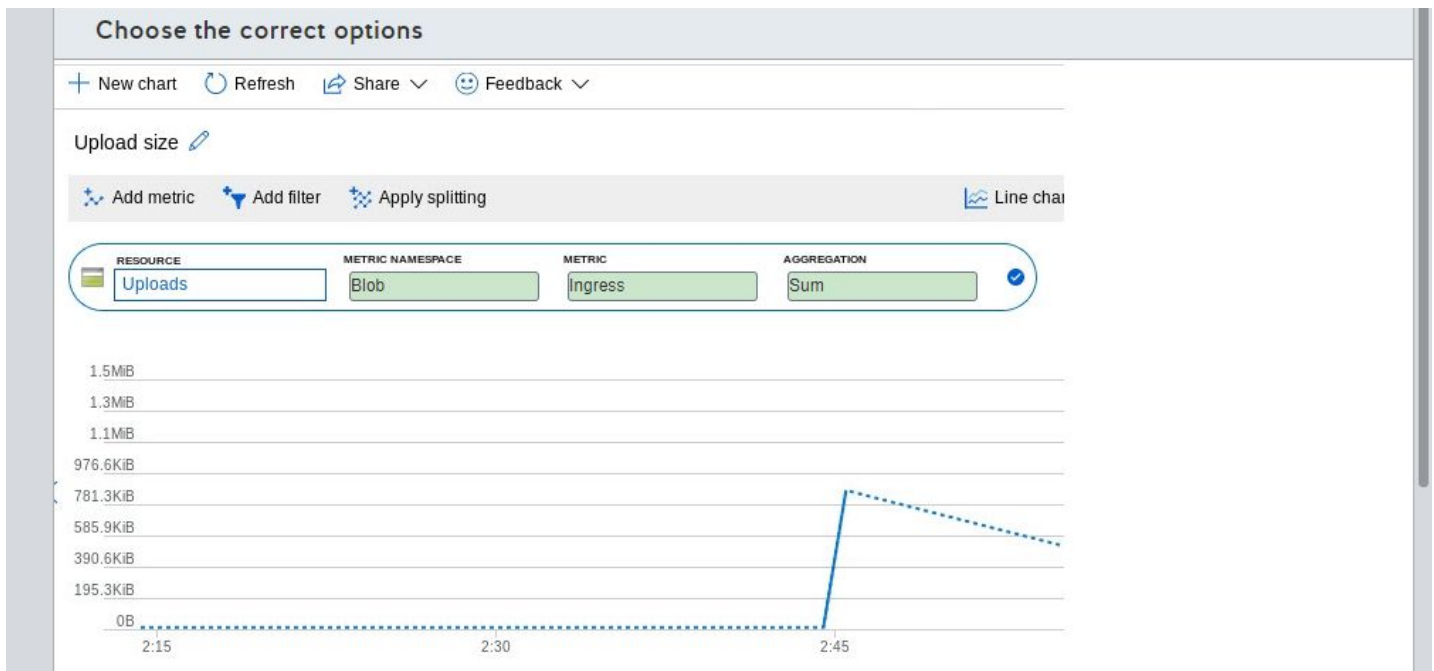
You need to create a report that shows how much data was uploaded in each time frame.

You create a new chart in Azure monitor for this report.

How should you configure this chart? To answer, select the appropriate options from the drop-down menus.

Choose the correct options





Explanation

You should use Blob as the metric namespace. Files uploaded by users are stored in blob storage, and your reports are based in these files.

You should use ingress as the metric. Ingress corresponds to the amount of ingress data in bytes. With this metric, you can collect the necessary data.

You should use sum as the aggregation. This will sum ingress data for a given time frame, resulting in the necessary chart for your report.

You should not use account as the metric namespace. The account metric namespace combines blob, table, file and queue metrics. You only need blob data to create the report.

You should not use file as the metric namespace. The file metric namespace is used for file sharing in an Azure Storage Account. User data is uploaded to a blob container.

You should not use transactions as the metric. This metric is used to determine the number of transactions that occurred, not the type and size of the transaction.

You should not use capacity as the metric. This metric reports the total of storage in the storage account. You are interested in how much data was uploaded in a given time frame.

You should not use max and min as the aggregation. This aggregation shows the maximum and minimum data ingresses in Azure Storage.

References

[Monitor, diagnose, and troubleshoot Microsoft Azure Storage](#)

[Monitor a storage account in the Azure portal](#)

[Azure Storage metrics in Azure Monitor](#)

Question128 Monitor data storage - D

Your company offers a Software-as-a-Service (SaaS) application for a support ticket system.

Each customer of the application runs in a multi-tenant, isolated database hosted on Azure SQL Elastic Pool Database and uses a Database Transaction Unit (DTU) based provisioning model.

The Sales team starts a campaign to attract more customers, so increasing DTU capacity could be necessary.

You need to monitor this database based on a metric that best anticipates out-of-capacity and performance issues. You also need to minimize the administrative effort to manage this alert.

Which metric should you implement in the monitoring alert?

Choose the correct answer

CPU percentage

DTU percentage

Data IO percentage

eDTU used

Explanation

You should use DTU percentage. This metric represents the percentage of DTU that is consumed by database, independent of the number of provisioned DTUs. If you increase the DTU capacity of the database, you do not need to edit the alert for the new capacity.

You should not use eDTU used. This metric could be used to anticipate performance issues when DTU consumption is close to the limit. However, you would need to edit this alert if you increase the DTU.

You should not use CPU or Data IO percentage. For DTU based models, CPU and Data IO and storage are bundled together in a simplified performance metric.

Monitoring only CPU or Data IO would not completely anticipate out-of-capacity issues.

References

[Create alerts for Azure SQL Database and Data Warehouse using Azure portal](#)

[Resources limits for elastic pools using the DTU-based purchasing model](#)

Question 129 Monitor data storage-

You manage multiple Azure SQL Databases for an insurance company. These databases are provisioned in a tier of a vCore-based purchasing model.

The Security team needs to audit all databases and correlate audit logs with logs generated by other Azure services. This analysis should be used to send alerts when suspicious operations occur.

You need to implement auditing in these databases to meet the security requirements while minimizing implementation efforts.

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

Choose the correct answers

Configure Azure Event Hub as the audit log destination.

Configure Azure Blob Storage as the audit log destination.

Enable auditing at the database level.

Configure Azure Log Analytics as the audit log destination.

Enable auditing at the server level.

Explanation

You should enable auditing at the server level. Server-level auditing enables auditing in all databases hosted on the server, meeting the security requirement to audit all databases.

You should also configure Azure Log Analytics as the audit log destination. Azure Log Analytics can centralize auditing logs and logs from other Azure services. You can create a log query to correlate logs stored in Log Analytics and create alerts based on these queries.

You should not enable auditing at the database level. Database-level auditing is enabled per database and is recommended only if you need to log only specific

events for a given database. Enabling server-level auditing is recommended for multiple databases on the same server.

You should not configure Azure Event Hub as the audit log destination. Audit logs can be streamed in real time by Event Hub and consumed by other services like Azure Stream Analytics or Azure Functions. However, you would need to manually correlate them with other Azure services logs and create the necessary alerts, increasing the implementation effort.

You should not configure Azure Blob Storage as the audit log destination. Storing audit logs in Blob Storage is useful for long-term retention. However, you need to consume these logs with other services like Databricks or HDInsight to correlate and create alerts.

References

[Get started with SQL database auditing](#)

Question 130 Monitor data storage-

You manage a big data solution for a pharmaceutical group. Stores scan customer prescriptions into image files. Each prescription scan is usually 1 MB. The files are stored in Azure Data Lake Storage Gen 2.

The solution uses an HDInsight cluster to analyze the prescriptions in Data Lake Storage. After data is processed, files are moved to the cool tier in Data Lake Storage.

On some occasions, stores upload prescription scans with a higher resolution, resulting in a file size greater than 20 MB. This causes performance degradation in the HDInsight analysis process.

You need to monitor when files uploaded to Data Lake Storage are larger than 20 MB to anticipate HDInsight performance degradation.

Which metric and dimension should you use? To answer, select the appropriate options from the drop-down menus.

Choose the correct options

Metric:

Dimension:

Blob Capacity Average

BlobTier

Explanation

You should use the Blob Capacity Average metric. This metric shows the average blob size stored in Azure Data Lake Storage. This metric could be configured in an alert to detect when the average file size in Data Lake Store is greater than 20 MB, anticipating performance degradation with HDInsight. However, if you only use this metric, the already processed files will be considered in the average file size, making it harder to anticipate performance degradation caused by bigger files. You need to filter already processed files by data tier using a metric dimension.

You should use the BlobTier dimension. You need to separate recent data from the already processed data to get a more accurate metric. Filtering by blob tier separates new files from already processed files because files are moved to the cool tier after HDInsights process them. Considering only files stored in the hot tier can anticipate performance degradation in HDInsight because you are monitoring for files to be analyzed.

You should not use the File Capacity Average metric. This metric shows the average file size stored in a file share. Consolidated files are not saved in a file share.

You should not use the Used Capacity Average metric. This metric shows the average capacity for the Azure Storage Account. You need to monitor data stored only in Azure Data Lake Storage.

You should not use the BlobType dimension. This dimension shows the type of the stored blob, like block blob, page blob, and Azure Data Lake Storage. You could use this metric to separate Blob Storage blobs from Azure Data Lake Storage blobs. However, Azure Data Lake Storage is provisioned in an exclusive storage account, and all blobs already have the correct type.

You should not use the ResponseType dimension. When a request is made to

Cosmos DB, a response is sent back to the client. This response contains an HTTP status code and a more detailed description, like Success, AuthorizationError, ServerTimeoutError, and others. You can filter for specific response type errors with this dimension. However, you cannot determine if the file consolidation process failed by the response type.

References

[Best practices for using Azure Data Lake Storage Gen2](#)

[Azure Storage metrics in Azure Monitor](#)

Question131 Monitor data storage-

You are developing a monitoring solution for an Azure Synapse Analytics SQL pool.

The monitoring solution has the following requirements:

- Query executions must be logged.
- Waits (including queues and blocking waits) must be logged.
- Kusto queries must be used to analyze the logged data.

You create a Log Analytics workspace.

You need to implement the monitoring solution.

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

Create a list in the correct order

Actions

Actions in order

- Write Kusto queries using the dm_pdw_sql_requests and dm_pdw_waits tables.
- Enable capturing metrics.
- Write Kusto queries using the AzureMetrics table.
- Enable auditing on the Synapse SQL pool (data warehouse) pane.
- Add a diagnostic setting in the Synapse SQL pool (data warehouse) pane.
- Select Send to Log Analytics.

- Select the ExecRequests and Waits log options.
- Write Kusto queries using the AzureDiagnostics table.

Explanation

You should perform these actions in this sequence:

1. Add a diagnostic setting in the Synapse SQL pool (data warehouse) pane.
Azure diagnostic logs are not collected by default. Capturing and storing of the logs needs to be activated by creating a diagnostic setting.
2. Select Send to Log Analytics.
Diagnostic logs can be sent to different destinations, such as Log Analytics, a storage account, or an event hub. According to the requirements, the data must be queried with Kusto query language. To make this possible, Log Analytics has to be the destination.
3. Select the ExecRequests and Waits log options.
ExecRequest needs to be selected to capture query executions. Waits needs to be selected to capture wait states of query executions.
4. Write Kusto queries using the AzureDiagnostics table.
The diagnostic logs of the Synapse SQL pool (data warehouse) are stored in the AzureDiagnostics table.

You should not enable auditing on the Synapse SQL pool (data warehouse) pane. By enabling auditing, query executions will be logged. However, waits will not be logged. This will not meet the requirements.

You should not write Kusto queries using the AzureMetrics table. You should not query the AzureMetrics table. Metrics (for example CPU percentage and active connections) will be sent to this table, not logs. Metrics will only be available if this is selected in a diagnostic setting.

You should not write Kusto queries using the dm_pdw_sql_requests and dm_pdw_waits tables. The objects sys.dm_pdw_sql_requests and sys.dm_pdw_waits are Dynamic Management Views and can be queried using T-SQL while connected to the SQL pool. They are not available in Log Analytics. The sys.dm_pdw_sql_requests view holds information about executed and currently executing queries. sys.dm_pdw_waits gives the query waits.

You should not enable capturing metrics. It is not a requirement to capture metrics

(for example CPU percentage).

References

[Monitor workload - Azure portal](#)

[sys.dm_pdw_exec_requests \(Transact-SQL\)](#)

[sys.dm_pdw_waits \(Transact-SQL\)](#)

[Overview of Azure platform logs](#)

Question132 Monitor data storage-

You have an Azure Synapse Analytics SQL pool.

You are implementing a data loading process using Polybase. The data must be loaded from Data Lake Storage Gen2. The data is stored in a directory as text files.

One of the external tables has this definition:

```
CREATE EXTERNAL TABLE [ext].[fact_Order] (  
  [Order Key] [bigint] NOT NULL,  
  [City Key] [int] NOT NULL,  
  [Customer Key] [int] NOT NULL,  
  [Stock Item Key] [int] NOT NULL,  
  [Order Date Key] [date] NOT NULL,  
  [Picked Date Key] [date] NULL,  
  [Salesperson Key] [int] NOT NULL,  
  [Picker Key] [int] NULL,  
  [WWI Order ID] [int] NOT NULL,  
  [WWI Backorder ID] [int] NULL,  
  [Description] [nvarchar](100) NOT NULL,  
  [Package] [nvarchar](50) NOT NULL,
```

```
[Quantity] [int] NOT NULL,  
[Unit Price] [decimal](18, 2) NOT NULL,  
[Tax Rate] [decimal](18, 3) NOT NULL,  
[Total Excluding Tax] [decimal](18, 2) NOT NULL,  
[Tax Amount] [decimal](18, 2) NOT NULL,  
[Total Including Tax] [decimal](18, 2) NOT NULL,  
[Lineage Key] [int] NOT NULL
```

```
)
```

```
WITH ( LOCATION = '/fact_Order/',
```

```
DATA_SOURCE = WWIStorage,
```

```
FILE_FORMAT = TextFileFormat,
```

```
REJECT_TYPE = VALUE,
```

```
REJECT_VALUE = 10
```

```
);
```

You receive this error when querying the external table in SSMS:

Rows were rejected while reading from external source(s).

11 rows rejected from external table [fact_Order] in plan step 2 of query execution:

Location: '/fact_Order/fact_Order.csv' Column ordinal: 15, Expected data type:
DECIMAL(18,2).

Location: '/fact_Order/fact_Order.csv' Column ordinal: 15, Expected data type:
DECIMAL(18,2).

Location: '/fact_Order/fact_Order.csv' Column ordinal: 15, Expected data type:
DECIMAL(18,2).

Location: '/fact_Order/fact_Order.csv' Column ordinal: 15, Expected data type:
DECIMAL(18,2).

Location: '/fact_Order/fact_Order.csv' Column ordinal: 15, Expected data type:
DECIMAL(18,2).

Location: '/fact_Order/fact_Order.csv' Column ordinal: 15, Expected data type: DECIMAL(18,2).

Location: '/fact_Order/fact_Order.csv' Column ordinal: 15, Expected data type: DECIMAL(18,2).

Location: '/fact_Order/fact_Order.csv' Column ordinal: 15, Expected data type: DECIMAL(18,2).

Location: '/fact_Order/fact_Order.csv' Column ordinal: 15, Expected data type: DECIMAL(18,2).

Location: '/fact_Order/fact_Order.csv' Column ordinal: 15, Expected data type: DECIMAL(18,2).

and 1 more...

Msg 107090, Level 16, State 1, Line 1

107090;Query aborted-- the maximum reject threshold (10 rows) was reached while reading from an external source: 11 rows rejected out of total 11 rows processed.

For each of the following statements, select Yes if the statement is true. Otherwise, select No.

Statement	Yes	No
The data to import contains more than 10 rows that do not conform to the specified external file format (dirty records).	<input type="radio"/>	<input type="radio"/>
You can use the sys.dm_pdw_errors Dynamic Management View to monitor for this error.	<input type="radio"/>	<input type="radio"/>
Failed rows will be written to a subdirectory named '_rejectedrows' in the Azure storage account.	<input type="radio"/>	<input type="radio"/>

Explanation

The data to import contains more than 10 rows that do not conform to the specified external file format (dirty records). REJECT_VALUE is a literal value because REJECT_TYPE is set to VALUE. Queries will fail with an error if the number of failed rows is greater than 10 (REJECT_VALUE).

You can use the sys.dm_pdw_errors Dynamic Management View to monitor for this error. All errors of failed queries can be read from this Dynamic Management View.

Failed rows will not be written to a subdirectory named '_rejectedrows' in the Azure storage account. Rejected rows will only be written to the storage container if REJECTED_ROW_LOCATION is included in the WITH phrase in the CREATE EXTERNAL TABLE statement. This is not the case. Rejected rows would only be written to '_rejectedrows' if the maximum reject threshold (10 rows) is not reached.

References

[Best practices for loading data for data warehousing](#)

[PolyBase Setup Errors and Possible Solutions](#)

[CREATE EXTERNAL TABLE \(Transact-SQL\)](#)

[sys.dm_pdw_errors \(Transact-SQL\)](#)

[Tutorial: Load data to Azure Synapse Analytics SQL pool](#)

Question 133 Monitor data processing-

You deploy Azure Databricks for your organization in the Microsoft Azure cloud to support development of a new big data analytics solution.

You already use Azure Monitor to monitor the performance of other corporate solutions.

You need to enable visualization of Azure Databricks metrics in Azure Monitor as well.

Which options describe the way you can enable your strategy? To answer, select the appropriate options from the drop-down menus.

Choose the correct options

By default, Azure Databricks metrics are available in the format for:

Azure Databricks supports sending log data to Azure Monitor

Metrics transferred to Azure will be stored in

Ganglia

use a third-party library

Azure Log Analytics workspace

Choose the correct options

By default, Azure Databricks metrics are available in the format for:

Azure Databricks supports sending log data to Azure Monitor

Metrics transferred to Azure will be stored in

Explanation

You should choose Ganglia as a default format for metrics data used by Azure Databricks. Integration with Azure Monitor and Datadog requires additional manual work with deployment of relevant drivers or libraries.

You should use a third-party library to transfer Azure Databricks metrics to Azure Monitor because it is not supported natively at the time of writing.

You should use Azure Log Analytics workspace as the target destination for uploaded Azure Databricks metrics in Azure Monitor. Each workspace has its own data repository, and data sources like Azure Databricks can be configured to store their metrics in a particular Azure Log Analytics workspace.

You should not use a custom Cosmos DB instance. Azure Monitor relies on a dedicated Azure Log Analytics workspace.

You should not use the ONNX runtime. It is used to exchange Machine Learning models and is not relevant to transferring performance-related metrics data.

You should not use Azure Network Watcher. It is a centralized tool for monitoring Azure networking.

References

[Monitoring Azure Databricks](#)

[Cluster Configurations](#)

[Get started with Log Analytics in Azure Monitor](#)

[ONNX Runtime is now open source](#)

[What is Azure Network Watcher?](#)

Question134 Monitor data processing-

You are managing several HDInsight clusters in your company's Azure subscription. Your line manager asks you to enable specific monitoring tasks for HDInsight clusters.

You need to choose a relevant tool for each task.

Which tool should you use? To answer, select the appropriate options from the drop-down menus.

Choose the correct options

Monitoring Task	Tool
Access YARN Queue Manager to monitor resource constraints and distribution.	
Get a consolidated view of CPU, RAM, and storage utilization across all HDInsight clusters.	
Receive an SMS message or voice call when specific alerts are fired for the monitored condition in HDInsight.	

Apache Ambari

Azure Monitor logs

Azure Monitor

Monitoring Task	Tool
Access YARN Queue Manager to monitor resource constraints and distribution.	Apache Ambari
Get a consolidated view of CPU, RAM, and storage utilization across all HDInsight clusters.	Azure Monitor logs
Receive an SMS message or voice call when specific alerts are fired for the monitored condition in HDInsight.	Azure Monitor

Explanation

You should use Apache Ambari to access YARN Queue Manager and to monitor job distribution across the queues. Ambari comes pre-installed with HDInsight and has built-in support to access YARN Queue Manager.

You should use Azure Monitor logs to get a consolidated view across all HDInsight clusters in the scope. The metrics and logs from the multiple HDInsight clusters can be sent to the Log Analytics workspace set up in Azure Monitor. Ambari is enabled separately for each HDInsight cluster and is accessible via each cluster URL. It cannot provide a consolidated view across several HDInsight clusters.

You should use Azure Monitor to receive an SMS message or voice call for the fired alert. In the Alert rule configuration, you should select the Email/SMS/Push/Voice action type and then specify the mobile number that is to receive the SMS message or automatic voice call. Ambari only supports e-mail and SNMP as a part of the alert notification configuration.

You should not use Windows Event Viewer or Linux dmesg. These tools are used for these specific operating systems and allow you to see system/application events or kernel messages. They cannot be used for HDInsight monitoring tasks.

References

[How to monitor cluster availability with Ambari and Azure Monitor logs](#)

[Manage HDInsight clusters by using the Apache Ambari Web UI](#)

Question135 Monitor data processing-

You manage an Azure SQL Database instance that uses the Database Throughput Unit (DTU) pricing model. It has a maximum of 5 DTUs and is two gigabytes (GB) in size. You must upgrade the instance's tier if its CPU reaches 50 percent, its memory usage reaches one GB, or if its disk reads or writes per second reach 50 percent of its capacity.

You need to create an alert that monitors the instance over a 12-hour period.

How should you configure the alert? To answer, select the appropriate configuration values from the drop-down menus.

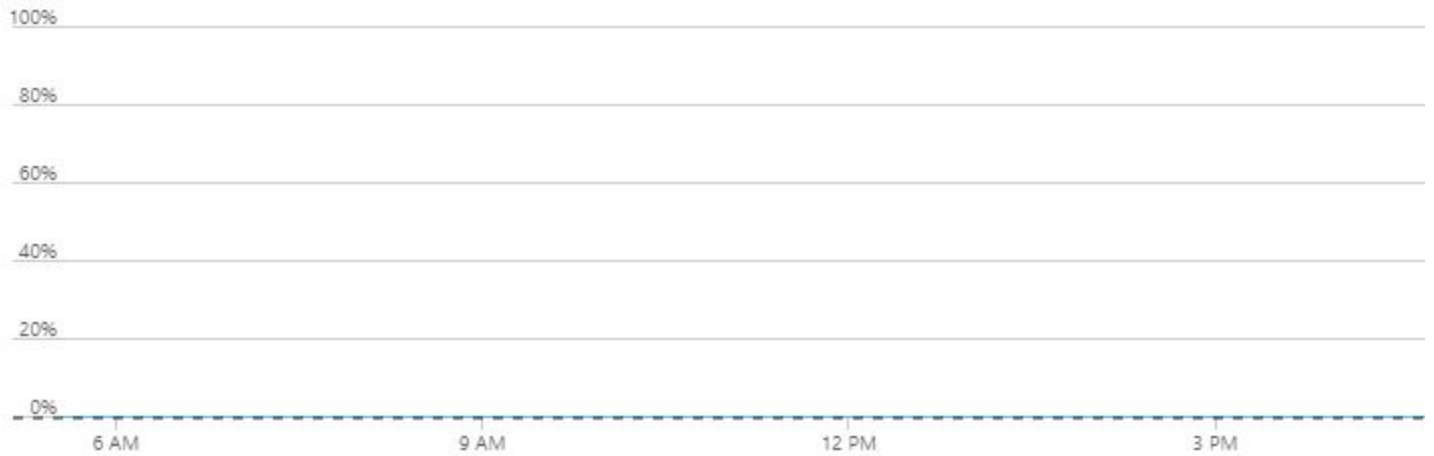
Choose the correct options

Configure signal logic

Metric:

Chart period ⓘ

Over the last 12 hours ▾



Alert logic

Threshold ⓘ

Static

Dynamic

Operator ⓘ

* Aggregation type ⓘ

* Threshold value ⓘ

DTU Percentage

Greater than or equal to

Maximum

50



Explanation

You should monitor the DTU used metric, which represents the number of DTUs that are used by the SQL Database instance. This number represents the amount of resources used by SQL Database, including compute power, storage, and I/O.

You should set the Operator to Greater than or equal to, the Aggregation type to Maximum, and the Threshold to 50. This causes Azure to monitor the instance to determine if it reaches 50 percent of its maximum DTU with regards to database size.

You should not set the Operator to Less than. This causes Azure to trigger the alert if the value is less than the specified threshold.

You should not set the Data space used. This allows you to monitor only the disk space, not the CPU percentage.

You should not set the Threshold to 5. The threshold must represent a percentage. Although 2.5 is half of the DTU value of 5, you must set this value to 50 percent.

References

[Monitoring and performance tuning](#)

Question136 Monitor data processing-

You have an Azure SQL Database instance with a single database. You are enabling diagnostics logging. You want to gather information about performance recommendations, queries that are taking the longest to run, and CPU and Database Throughput Unit (DTU) limits and percentages.

You need to determine which three logs to enable.

Which three logs should you enable? To answer, move the appropriate logs from the list of possible logs to the answer area and arrange them in any order.

Create a list in any order

Possible logs

Logs to monitor

- DatabaseWaitStatistics
- QueryStoreWaitStatistics
- Basic metrics
- QueryStoreRuntimeStatistics
- SQLInsights

Explanation

You should enable the following logs:

- SQLInsights
- QueryStoreRuntimeStatistics
- Basic metrics

SQLInsights gathers performance information and provides recommendations. QueryStoreRuntimeStatistics provides information about CPU usage and query duration. Basic metrics provides CPU and DTU usage and limits.

You should not enable QueryStoreWaitStatistics. This provides information about the resources that caused queries to wait, such as the CPU, logs, or locks.

You should not enable DatabaseWaitStatistics. This provides information about the time a database spent on waiting.

References

[Azure SQL Database metrics and diagnostics logging](#)

Question137 Monitor data processing-

You are a data engineer for your company. You have an HDInsight cluster in Azure.

You need to monitor the cluster for performance.

What should you do?

Choose the correct answer

Create a Log Analytics workspace and use Azure Advisor.

Create a Data Lake Analytics account and use Azure Advisor.

Create a Data Lake Analytics account and use Azure Monitor.

Create a Log Analytics workspace and use Azure Monitor.

Explanation

You should create a Log Analytics workspace and use Azure Monitor. You enable monitoring in the HDInsight cluster. From there, you can choose the Log Analytics workspace that should receive the logs. You can then view these logs in Azure Monitor.

You should not create a Data Lake Analytics account and use Azure Monitor. Data Lake Analytics automatically scales big data analytics jobs in Data lake.

You should not create a Log Analytics workspace and use Azure Advisor. Azure Advisor provides recommendations on cost, availability, performance, and security.

You should not create a Data Lake Analytics account and use Azure Advisor. Azure Advisor provides recommendations on cost, availability, performance, and security.

References

[What is Azure Data Lake Analytics?](#)

[Introduction to Azure Advisor](#)

[Use Azure Monitor logs to monitor HDInsight clusters](#)

Question138 Monitor data processing-

You are a data engineer for your company. You have an Azure Databricks account. You want to use Azure Monitor to monitor the Databricks account.

You need to choose a solution for sending application metrics to Azure Monitor.

Which library or tool should you use?

Choose the correct answer

Log4j

Dropwizard

Azure CLI

Powershell

Explanation

You should use Dropwizard. Dropwizard is a Java library. Spark, which is the cluster engine that is used to run Databricks, uses a configurable metrics system that is based on the Dropwizard Metrics Library.

You should not use Log4j. This Java library is used to send application logs to Azure Monitor.

You should not use Powershell or Azure CLI. Both of these are command-line utilities that allow you to administer Azure.

References

[Get started with Azure CLI](#)

[Getting Started with Windows PowerShell](#)

[Send Azure Databricks application logs to Azure Monitor](#)

Question139 Monitor data processing-

You are a data engineer. You want to set up an alert to monitor a Stream Analytics job. You want to receive an alert when the job stops unexpectedly.

You need to configure the alert.

How should you configure the alert? To answer, select the appropriate configurations from

the drop-down menus.

Choose the correct options

5 PM	6 PM	7 PM	8 PM
Alert logic			
Event Level ⓘ		Status ⓘ	
<input type="text"/>		<input type="text"/>	
All			
<input type="text"/>		<input type="text"/>	
Failed			

Choose the correct options			
5 PM	6 PM	7 PM	8 PM
Alert logic			
Event Level ⓘ		Status ⓘ	
All		Failed	

Explanation

You should set the Event Level to All and the Status to Failed. This allows you to receive an alert when the job stops.

You should not use any other combination. The Warning Event Level specifies events that are considered a warning but not catastrophic. The Succeeded event level specifies successful events.

References

[Set up alerts for Azure Stream Analytics jobs](#)

Question140 Monitor data processing-

You are a data engineer. You manage an HDInsight cluster. You want to monitor the cluster's performance by using a web user interface.

You need to choose the appropriate tool for monitoring.

Which tool should you use?

Choose the correct answer

Azure Stream Analytics

Azure Log Analytics

Apache Ambari

Apache Spark

Explanation

You should use Apache Ambari. Apache Ambari is a web user interface that allows you to monitor HDInsight Hadoop clusters. Apache also offers Ambari REST SDK, which allows developers to integrate monitoring into their applications.

You should not use Apache Spark. Spark is a parallel processing framework for big data analytics applications.

You should not use Azure Log Analytics. Log Analytics allows you to write queries to retrieve event data from event logs.

You should not use Azure Stream Analytics. Stream Analytics allows you to process telemetry data in real time.

References

[What is Azure Stream Analytics?](#)

[Get started with Log Analytics in Azure Monitor](#)

[What is Apache Spark in Azure HDInsight](#)

[Manage HDInsight clusters by using the Apache Ambari Web UI](#)

Question 141 Monitor data processing - D

You are a data engineer. You have an Azure Data Factory pipeline in the US East region. You realize that pipeline run-data is deleted after 45 days.

You need to keep the run-data from the pipeline for longer than 45 days.

What should you do?

Choose the correct answer

Configure diagnostics logs to send data to a blob storage account.

Re-create the Data Factory pipeline in the US East 2 region.

Add a lock with the CanNotDelete lock type.

Re-create the Data Factory pipeline and enable Git.

Explanation

You should configure diagnostics logs to send data to a blob storage account. By default, Azure Data Pipeline stores run-data for only 45 days. To store the data longer than that, you must configure diagnostics logs. With diagnostics logs, you can choose to store the run-data in a blob storage account, an event hub, or a Log Analytics workspace.

You should not re-create the pipeline in the US East 2 region. The region does not affect the number of days run-data is stored.

You should not re-create the pipeline and enable Git. Git is an online version control system. Git integration does not affect the number of days run-data is stored.

You should not add a lock with the CanNotDelete lock type. Azure locks allow you to control resources to prevent unexpected changes. The CanNotDelete lock type indicates that the Data Factory resource cannot be deleted. It has no effect on run-data.

References

[Introduction to Azure Data Factory](#)

[Lock resources to prevent unexpected changes](#)

[Continuous integration and delivery \(CI/CD\) in Azure Data Factory](#)

[Alert and monitor data factories by using Azure Monitor](#)

Question142 Monitor data processing-

You manage an Azure SQL Data Warehouse. You have two tables named DimEmployee and FactWorkResults.

You need to determine the top 10 longest running queries across both tables.

How should you write the SQL statement? To answer, select the appropriate code segments from the drop-down menus.

Choose the correct options

SELECT TOP 10 *

FROM

ORDER BY

sys.dm_pdw_exec_requests

total_elapsed_time

DESC

Choose the correct options

SELECT TOP 10 *

FROM sys.dm_pdw_exec_requests

ORDER BY total_elapsed_time

DESC

Explanation

You should use the following statement:

```
SELECT TOP 10 *  
FROM sys.dm_pdw_exec_requests  
ORDER BY total_elapsed_time DESC
```

This statement returns the top 10 longest-running queries. The sys.dm_pdw_exec_requests DMV shows all queries to the data warehouse. The total_elapsed column represents how long it took the query to run. By using DESC, you return the queries in decreasing order from slowest to fastest.

You should not return data from the DimEmployee and FactWorkResults tables. These tables contain data, but not the actual queries for that data.

You should not use ASC. This returns the queries in ascending order, from slowest to fastest.

You should not use the submit_time column. This represents the time a query was submitted, not how long it was run.

References

[Monitor your workload using DMVs](#)

Question143 Monitor data processing-

Your company has an Azure HDInsight solution that uses an Apache Hadoop cluster to process and analyze data.

Your management reports slowdowns during a specific time window.

You need to provide a monitoring solution that tracks cluster performance issues and alerts your team as soon as possible.

Which two monitoring solutions should you use? Each correct answer presents a complete solution.

Choose the correct answers

Azure Monitor metrics

Azure Monitor Log Analytics

Apache Ambari

Explanation

You should use Apache Ambari. Ambari is an open-source project that provisions, manages and monitors Hadoop clusters. Ambari is provisioned by default with HDInsight and is the primary source to manage and monitor your clusters. It supports alerting functionalities with Ambari Alert Framework and Ambari Metrics System.

You should also use Azure Monitor Log Analytics. You can centralize your metrics and alerts in Azure Monitor, enable your cluster to send metrics to a Log Analytics workspace, and create an Azure Monitor alert based on a log search query.

You should not use Azure Monitor metrics. The metrics available in Azure Monitor metrics are limited to the gateway request, the number of active workers, and scaling requests. You should use Ambari to manage HDInsight clusters.

You should not use the HDInsight Cluster REST API. The REST API manages your cluster resources in Azure, like creating a new cluster, deleting, and changing configurations. You cannot use the HDInsight Cluster REST API for monitoring purposes. However, you can use the Ambari REST API to manage alerts and monitor your cluster.

References

[Monitor cluster performance in Azure HDInsight](#)

[Manage HDInsight clusters by using the Apache Ambari Web UI](#)

[Use Azure Monitor logs to monitor HDInsight clusters](#)

[Cluster REST API in Azure HDInsight](#)

[Apache Ambari](#)

Question144 Monitor data processing - D

You manage an Azure Stream Analytics job for a real-time log streaming solution.

The logs are consumed by PowerBI to generate a dashboard. Users of this dashboard report that new data is not being streamed to PowerBI.

You troubleshoot this issue and discover that the Stream Analytics job stopped unexpectedly.

You need to configure an alert to monitor this issue.

Which metric or activity log should you implement for this alert?

Choose the correct answer

Streaming Units (SU) percentage utilization

Runtime Errors

All Administrative operations

Output Events

Explanation

You should use the All Administrative operations signal logic log. This activity log, when configured in signal logic to monitor any level with failed status, reports when a job enters any failed state. This would issue an alert if the Stream Analytics job stopped unexpectedly.

You should not use the Output Events or SU percentage utilization metrics. A sum of zero SU percentage utilization or output events is possible if the job is not receiving any input and the job is running properly. This could result in false positive alerts.

You should not use the Runtime Errors metric. An alert monitoring runtime errors is used when the job can receive the data but is generating errors while processing the query. If a job stops running unexpectedly, runtime errors will not occur, and the alert would not be effective.

References

[Set up alerts for Azure Stream Analytics jobs](#)

Question145 Monitor data processing-

Your company uses an Azure Stream Analytics job for a streaming solution. The job resource in Azure Portal is named FraudAnalysis.

Other services in this solution send diagnostic logs to an Azure Monitor Log Analytics workspace. This Log Analytics workspace resource is named LogWorkspace.

Both the Log Analytics and the Stream Analytics job are provisioned in a resource group named Production.

You need to complete the PowerShell command to integrate the job diagnostic logs with Azure Monitor to correlate log data with other services.

How should you complete the PowerShell command? To answer, select the appropriate options from the drop-down menus.

Choose the correct options

```
$logAnalytics = Get-AzResource -ResourceGroupName "Production"
```

```
-ResourceName "LogWorkspace" -ResourceType "
```

```
$analyticsJob = Get-AzResource -ResourceGroupName "Production"
```

```
-ResourceName "FraudAnalysis" -ResourceType "
```

```
- $analyticsJob.ResourceId
```

```
- $logAnalytics.ResourceId
```

```
-Enabled $true
```

Microsoft.OperationalInsights/workspaces

Microsoft.StreamAnalytics/streamingjobs

Set-AzDiagnosticSetting

ResourceId

WorkspaceId

Choose the correct options

```
$logAnalytics = Get-AzResource -ResourceGroupName "Production"  
-ResourceName "LogWorkspace" -ResourceType "Microsoft.OperationalInsights/workspaces"
```

```
$analyticsJob = Get-AzResource -ResourceGroupName "Production"  
-ResourceName "FraudAnalysis" -ResourceType "Microsoft.StreamAnalytics/streamingjobs"
```

```
Set-AzDiagnosticSetting -ResourceId $analyticsJob.ResourceId  
-WorkspaceId $logAnalytics.ResourceId  
-Enabled $true
```

Explanation

You should complete the command as shown below:

```
$logAnalytics = Get-AzResource -ResourceGroupName "Production"  
-ResourceName "LogWorkspace" -ResourceType  
"Microsoft.OperationalInsights/workspaces"
```

```
$analyticsJob = Get-AzResource -ResourceGroupName "Production"  
-ResourceName "FraudAnalysis" -ResourceType  
"Microsoft.StreamAnalytics/streamingjobs"
```

```
Set-AzDiagnosticSetting -ResourceId $analyticsJob.ResourceId  
-WorkspaceId $logAnalytics.ResourceId  
-Enabled $true
```

To send diagnostic logs from Stream Analytics job to Log Analytics, you need to use the resource id from both resources in the Set-AzDiagnosticSetting command. You should use the Microsoft.OperationalInsights/workspaces and Microsoft.StreamAnalytics/streamingjobs resource types, respectively, with Get-AzResource to reference the correct resources in the PowerShell object variable.

You should use the Set-AzDiagnosticSetting command to configure the diagnostic logs setting from a resource. You should reference the Stream Analytics job

resource id as -ResourceId in this command to configure the location of the diagnostic log.

You should configure the Log Analytics resource id as -WorkspaceId in the Set-AzDiagnosticSetting command. This will configure Log Analytics as the destination for diagnostic logs, integrating Databricks with Azure Monitor Log Analytics.

You should not use the Microsoft.Insights/diagnosticSettings resource type. This resource type is used to reference diagnostics settings. A diagnostics setting is created after running the Set-AzDiagnosticSetting command. You do not need to reference other diagnostics settings to integrate Databricks with Log Analytics.

You should not use the Add-AzLogProfile command. This command is used to create a new activity log profile and archive the activity log to Azure Blob storage or stream it to an Azure Event Hub. You need to collect Databricks diagnostics logs.

You should not use the Enable-AzActivityLogAlert command. This enables a previously created activity log alert. Activity log alerts are fired when a configured condition action occurs. You need to collect Databricks diagnostics logs.

You should not use the -StorageAccountId parameter in the Set-AzDiagnosticSetting command. This parameter is used to send diagnostics logs to Azure Blob Storage. You need to configure the diagnostics logs destination to Log Analytics.

References

[Create diagnostic setting to collect platform logs and metrics in Azure](#)

[Set-AzDiagnosticSetting](#)

[Add-AzLogProfile](#)

[Enable-AzActivityLogAlert](#)

[Alerts on activity log](#)

Question146 Monitor data processing - D

Your team manages a data pipeline in Azure Data Factory.

This pipeline is executed every week on Sunday and moves data to an Azure SQL Data Warehouse that is used by the whole company.

You have the following monitoring requirements for this data pipeline:

- If a pipeline run fails, your team needs to be notified so it can remediate the problem.
- Pipeline run metrics from the last 90 days must be stored for analytics and historical data.

You need to implement a monitoring solution to satisfy these requirements with minimal implementation efforts.

Which solution should you use?

Choose the correct answer

Azure Data Factory Monitor

Azure Data Factory REST API

Azure Monitor

.NET Software Development Kit (SDK)

Explanation

You should use Azure Monitor. Azure Data Factory can send diagnostics logs containing metrics about pipeline runs to Azure Monitor. You can store the metrics in Azure Monitor to analyze past pipeline runs and create an alert for your team if a pipeline run fails.

You should not use Azure Data Factory Monitor. Azure Data Factory includes monitoring capabilities for your pipeline runs with execution metrics and pipeline status. You can define alerts directly in Azure Data Factory Monitor. However, Azure Data Factory data retention is limited to 45 days. You need to use Azure Monitor for longer retention.

You should not use the Azure Data Factory REST API or .NET SDK. Both solutions could be used to create a pipeline run and monitor the status. However, you would need to manually store pipeline run data and create alerts, increasing the implementation effort.

References

[Visually monitor Azure data factories](#)

[Alert and monitor data factories by using Azure Monitor](#)

[Programmatically monitor an Azure data factory](#)

Question147 Monitor data processing-

You manage an Azure Stream Analytics job to analyze banking system transactions for fraud detection.

You recently optimized the Streaming Unit (SU) consumption for the job by implementing query parallelization. Banking system transactions are expected to increase radically during an upcoming big sales event.

You need to create an alert if SU consumption is higher than 80 percent.

How should you configure the alert? To answer, select the appropriate options from the drop-down menus.

Choose the correct options

Metric:

Operator:

Aggregation:

SU % utilization

greater than

Maximum

Choose the correct options

Metric: SU % utilization

Operator: Greater than

Aggregation: Maximum

Explanation

You should use the SU % utilization metric. This metric indicates how much of the provisioned SU % is in use. If this indicator reaches 80%, there is a high probability that the job will stop.

You should use the greater than operator for this alert. You need to be alerted if SU consumption is higher than expected.

You should use the maximum aggregation for this alert. You need to be notified as soon as possible if SU consumption reaches a higher utilization. With maximum aggregation, peak consumption events will be detected.

You should not use the Input events metric. This metric counts the number of records deserialized from the input events. This count is expected to increase during the big sales event. However, this is not the best metric to monitor SU consumption.

You should not use the Runtime errors metric. This metric is the total number of errors during query processing. A Stream Analytics job operating close to the SU limit could present query processing errors. However, SU % utilization is a more precise metric.

You should not use the less than or less than or equal operator. The alert needs to be triggered when SU consumption is above the threshold, not below.

You should not use Average or Minimum aggregation. Depending on the aggregation time granularity, these aggregations could detect a high SU % utilization later than expected, reducing the reaction time to act on the alert.

References

[Understand Stream Analytics job monitoring and how to monitor queries](#)

[Set up alerts for Azure Stream Analytics jobs](#)

Question148 Monitor data processing - D

You create an Azure Data Factory pipeline. You also create a Log Analytics workspace. In Azure Monitor you set the diagnostic settings as shown in the exhibit.

Which tables can you query using KQL in the Log Analytics workspace?

Choose the correct answer

PipelineRuns and ActivityRuns

ADFPipelineRun and AzureMetrics

ADFPipelineRun and PipelineSucceededRuns

AzureDiagnostics and ActivityRuns

ADFPipelineRun, ADFTriggerRun, and ADFActivityRun

Explanation

You can query the ADFPipelineRun table because Destination table is set to Resource specific and log PipelineRuns is selected. ADFPipelineRun contains rows for status changes like InProgress and Succeeded.

You can query the AzureMetrics table because the AllMetrics option is selected. AzureMetrics contains metrics like PipelineSucceededRuns.

You cannot query the ADFTriggerRun table. The TriggerRuns option is not selected.

You cannot query the ADFActivityRun table. The ActivityRuns option is not selected.

You cannot query the PipelineRuns table. PipelineRuns is a category in the AzureDiagnostics table that contains the pipeline runs log.

You cannot query the ActivityRuns table. ActivityRuns is a category in the AzureDiagnostics table that contains logs of activities in pipeline runs.

You cannot query the AzureDiagnostics table. This table can be used if the Destination table option is set to Azure diagnostics.

You cannot query the PipelineSucceededRuns table. PipelineSucceededRuns is a metric. It counts the number of succeeded pipeline runs. You can use the AzureMetrics table to query this data.

References

[Alert and monitor data factories by using Azure Monitor](#)

[Supported services, schemas, and categories for Azure Resource Logs](#)

[Azure Data Factory logs now available as dedicated tables in Azure Monitor Logs](#)

Question149 Monitor data processing-

You create an Azure Data Factory pipeline. You have a Log Analytics workspace.

In Azure Monitor you set the diagnostic settings for the Data Factory as shown in the exhibit.

You need to create a log alert in the Azure portal using Azure Monitor with a KQL query. The query should be executed every 20 minutes and have a time window of 20 minutes.

How should you configure the settings for the alert in Azure Monitor? To answer, drag the appropriate value to each setting. Each value may be used once, more than once, or not at all.

Drag and drop the answers

```
ADFPipelineRun  
| where Status == "Failed"
```

20

20

```
ADFPipelineRun  
| where Status == "Failed"
```

20

```
ADFPipelineRun  
| where Status == "Failed"  
| where Start > ago(20m)
```

0

40

```
search in (ADFPipelineRun) "failed"
```

Setting	Value
KQL query	<pre>ADFPipelineRun where Status == "Failed"</pre>
Period (in minutes)	20
Frequency (in minutes)	20

```
ADFPipelineRun  
| where Status == "Failed"
```

20

Explanation

You should use the following KQL query:

```
ADFPipelineRun
```

```
| where Status == "Failed"
```

When a pipeline fails, the ADFPipelineRun table will receive a row with the failed status because in the diagnostics settings for the Data Factory, the Destination table option is set to Resource specific and PipelineRuns is selected.

You should set the period to 20 minutes. This is the time window that will be used for the KQL query.

You should set the frequency to 20 minutes. This is the time between two executions of the query.

You should not use the following KQL query:

```
ADFPipelineRun
```

```
| where Status == "Failed"
```

```
| where Start > ago(20m)
```

KQL queries for alerts should not use a time filter. A time filter is set by the period. Also Start is the start of the pipeline run, which is not the same as the failure time.

You should not use the following KQL query:

```
search in (ADFPipelineRun) "failed"
```

You cannot use search in a KQL query for alerts, because it is not supported by the platform. This query returns the rows in the ADFPipelineRun table where any of the columns has the value failed. The result of the query suits the purpose. However, using search on all the columns has more processing overhead compared to using a filter on the Status column.

References

[Log alerts in Azure Monitor](#)

[Log alert queries in Azure Monitor](#)

Question 150 Implement non-relational data stores - D

You are a data engineer for a practice test provider. You plan to use an Azure Cosmos DB to store exam content. You must be able to retrieve exam content by using a query similar to the following:

```
SELECT * FROM Exams e WHERE e.number = "DP-200"
```

You need to generate the Cosmos DB account and database.

How should you complete the Azure CLI commands? To answer, select the appropriate options from the drop-down menus.

Choose the correct options

```
az cosmosdb create \
--resource-group 'mup-rg' --name 'mup' \
--locations 'North Central US' \
--default-consistency-level "Session"
```

```
az cosmosdb create \
--resource-group 'mup-rg' \
--name 'mup' \
--db-name 'mup-database'
```

cosmosdb

kind GlobalDocumentDB

cosmosdb database

Choose the correct options

```
az cosmosdb create \
--resource-group 'mup-rg' --name 'mup' \
--kind GlobalDocumentDB \
--locations regionName='North Central US' \
--default-consistency-level "Session"
```

```
az cosmosdb database create \
--resource-group 'mup-rg' \
--name 'mup' \
--db-name 'mup-database'
```

Explanation

You should use the following commands:

```
az cosmosdb create \
--resource-group 'mup-rg' --name 'mup' \
--kind GlobalDocumentDB \
--locations regionName='North Central US' \
--default-consistency-level "Session"
```

```
az cosmosdb database create \
--resource-group 'mup-rg' \
--name 'mup' \
--db-name 'mup-database'
```

Before you can create a database and collection, you must first create an account. The `az cosmosdb create` command creates the Cosmos DB account. Each

CosmosDB account supports one of five APIs for querying data: SQL, Gremlin, MongoDB, Table, and Cassandra. In this scenario, the query that is presented uses SQL. To use the SQL API, you must specify GlobalDocumentDB as the value of the --kind parameter.

The az cosmosdb database create command creates a database in the Cosmos DB account. You specify the name of the database as the --name parameter.

You should not call az cosmosdb database create before you call az cosmosdb create. This attempts to create the database before the account is created.

You should not call az cosmosdb collection create before you call az cosmosdb database create. A collection in Cosmos DB is analogous to a table in a SQL database.

You should not set the --kind parameter to MongoDB. MongoDB uses JavaScript to query data. In this scenario, you want to use SQL.

You should not set the --capabilities parameter to EnableCassandra. This configures the Cosmos DB account to use the Cassandra API, which allows you to use Cassandra Query Language (CQL) to query data.

You should not set the --capabilities parameter to EnableGremlin. This configures the Cosmos DB account to use the Gremlin API, which allows you to use Gremlin to query data.

You should not set the --capabilities parameter to EnableTable. This configures the Cosmos DB account to use the Table API, which allows you to use OData and Language Integrated Query (LINQ) queries.

References

[Azure CLI samples for Azure Cosmos DB SQL \(Core\) API](#)

[Tutorial: Query data from a Cassandra API account in Azure Cosmos DB](#)

[Tutorial: Query Azure Cosmos DB Gremlin API by using Gremlin](#)

[Tutorial: Query data from a Cassandra API account in Azure Cosmos DB](#)

Question151 Monitor data processing-

You have an Azure resource group named resourcegroup1. This resource group contains the following resources:

- sqlserver1 of type (logical Azure) SQL server
- synapsesqlpool1 of type Synapse SQL pool (data warehouse)

You need to use the Azure portal to set up an alert rule when 80 percent of the maximum resource capacity is used.

How should you configure the alert rule in Azure Monitor? To answer, drag the appropriate value to each alert rule setting. A value may be used once, more than once, or not at all.

Drag and drop the answers

sqlserver1/synapsesqlpool1

Metrics

DWU used

sqlserver1/synapsesqlpool1

Metrics

DWU used

sqlserver1

Log

CPU used

DWU limit

resourcegroup1

Drag and drop the answers

Alert rule setting	Value
Resource	sqlserver1/synapsesqlpool1
Signal type	Metrics
Signal name	DWU used

sqlserver1/synapsesqlpool1

Metrics

DWU used

sqlserver1

Log

CPU used

DWU limit

resourcegroup1

Explanation

You should set the resource to sqlserver1/synapsesqlpool1. The measurement for resource usage and capacity of a Synapse Analytics SQL pool is Data Warehouse Unit (DWU). It is a unit of performance (a combination of CPU, memory, and IO). The measurement is applicable to the SQL pool synapsesqlpool1.

You should set the signal type to Metrics. A metric has a numeric value and activates the alert when the resource usage crosses the 80 percent boundary.

You should set the signal name to DWU used. DWU used is a measurement of the resources consumed by the workload.

You should not set the resource to sqlserver1. This is a logical SQL Server, and

the metric that needs to be used is not available on this resource.

You should not set the resource to resourcegroup1. This is a resource group, and the resource needs to be the pool.

You should not set the signal type to Log. This is used for custom log search queries. According to the requirements, a log search query is not applicable here.

You should not use DWU limit as a signal name in the condition of the alert. DWU limit is a measurement of the currently allocated resources to the SQL Data Warehouse.

You should not use CPU used in the condition of the alert. This is only a measurement of CPU, and you need to use a measurement of all resources.

References

[Create alerts for Azure SQL Database and Data Warehouse using Azure portal](#)

[Data Warehouse Units \(DWUs\)](#)

[Azure SQL Data Warehouse: February 2016 update](#)

[Create, view, and manage metric alerts using Azure Monitor](#)

[Create, view, and manage log alerts using Azure Monitor](#)

Question152 Monitor data processing-

You have an Azure SQL Database. Auditing is enabled, using a Log Analytics workspace as a destination.

You need to write a Kusto query giving an overview of the failed login attempts of the last week. The result of the query must show one row per IP address (of the client) and the number of failed logins attempts per IP address.

How should you configure the query? To answer, select the appropriate options from the drop-down menus.

Choose the correct options

```
| where TimeGenerated > ago(7d)
| where      == "SQLSecurityAuditEvents"
| where action_name_s == "DATABASE AUTHENTICATION FAILED"
|           failed_login_attempts=count() by client_ip_s
```

AzureDiagnostics

Category

Summarize

Choose the correct options

```
AzureDiagnostics
| where TimeGenerated > ago(7d)
| where Category == "SQLSecurityAuditEvents"
| where action_name_s == "DATABASE AUTHENTICATION FAILED"
| summarize failed_login_attempts=count() by client_ip_s
```

Explanation

The first option should be AzureDiagnostics. This defines the table to read from. Auditing data of an Azure SQL Database is written to table AzureDiagnostics.

The second option should be Category. Many different types of data are written to the AzureDiagnostics table. The Category field can be used to filter the auditing events of the databases.

The third option should be summarize. Summarize can be used to aggregate data, in this case a count per client IP address.

You should not use SecurityEvent. The SecurityEvent table contains logging of failed logins, but this data applies to virtual machines.

You should not use Event. This table holds data for Windows event logs.

You should not use \$table. The result of the search includes a column \$table with the name of the table.

You should not use select. Kusto does not have a select keyword. Select is a SQL keyword.

You should not use project. Project defines the columns to return. In this case, aggregation is required.

References

[Azure Monitor log query examples](#)

[Summarize operator](#)

[AZURE SQL DB AND LOG ANALYTICS BETTER TOGETHER – PART #1](#)

[Windows event log data sources in Azure Monitor](#)

[Project operator](#)

Question153 Optimize Azure data solutions-

Your company's quality control system uses a single container in Azure Cosmos DB to store information about quality checks (successful checks and defect detection) of finished products.

Review of the data shows little value in storing successful check records long term. Management asks you to enable a data retention policy in Azure Cosmos DB, so that successful check records are automatically deleted after one hour, while defect detection records are retained indefinitely.

You need to update Time-to-Live (TTL) settings in Azure Cosmos DB to enable the requested data retention policy.

How should you configure TTL settings? To answer, select the appropriate options from the drop-down menus.

Choose the correct options

1. Set TTL of Azure Cosmos DB's container to

2. Programmatically set TTL of

to

3600

defect detection records

-1

1. Set TTL of Azure Cosmos DB's container to
2. Programmatically set TTL of
to

Explanation

You should set the TTL settings of the Azure Cosmos DB container to 3600, and programmatically set the TTL of defect detection records to -1. This sets the TTL for all items in the container to 1 hour, except for the defect detection items, which are kept indefinitely.

You should not set the TTL of the container to -1 (infinity) because the items would never expire by default.

You should not set the TTL of the records to null (default) because it would inherit TTL settings from the upper container level.

References

[Time to Live \(TTL\) in Azure Cosmos DB](#)

[Configure time to live in Azure Cosmos DB](#)

Question154 Optimize Azure data solutions-

You are configuring a star-schema model in Azure SQL Data Warehouse for a new Operations Analytics solution.

Telemetry data from factories will be loaded into the fact table, and grow by about 5 GB every month. Product details, which change infrequently, will be loaded into dimension tables, with the total size of about 50 MB.

Daily shipment updates from an external website are to be fast loaded into the temporary staging table first, before being processed and then moved into the fact table.

You need to choose the right data distribution method for the fact, dimension, and staging tables to optimize future data query operations.

Which methods should you use? To answer, drag the appropriate data distribution method to each table type. A distribution method may be used once, more than once, or not at all.

Drag and drop the answers

Hash-distributed

Replicated

Round-robin

Hash-distributed

Replicated

Round-robin

Tables	Data Distribution Method
Fact	Hash-distributed
Dimension	Replicated
Staging	Round-robin

Hash-distributed

Replicated

Round-robin

Explanation

You should set the hash-distributed method for the fact table. Because its data changes frequently, it will grow fast, and this distribution method will help to ensure optimal query performance.

You should set the replicated method for the dimension table. Data changes are relatively small, with content changed infrequently. This method replicates the data across compute nodes, simplifying maintenance and optimizing join and aggregation activities with locally-stored copies.

You should set the round-robin distribution method for the staging table. This will optimize load operations with the data pulled from the external source and processed in Azure SQL Data Warehouse before the update of the fact table records.

References

[Designing tables in Azure SQL Data Warehouse](#)

[Guidance for designing distributed tables in Azure SQL Data Warehouse](#)

[Design guidance for using replicated tables in Azure SQL Data Warehouse](#)

Question 155 Optimize Azure data solutions - D

Your company's quality control system uses Azure Cosmos DB to store information about quality checks (successful checks and defect detection) of finished products.

Management asks you to enable a data retention policy in Azure Cosmos DB, so that successful check entries are automatically deleted after one hour, while defect detection records are retained indefinitely.

You need to update the Time-To-Live (TTL) settings in Azure Cosmos DB to enable this data retention policy.

What should you do?

Choose the correct answer

Set the TTL settings of the Azure Cosmos DB container to -1, and programmatically set the TTL of defect detection items to null.

Set the TTL settings of the Azure Cosmos DB container to -1, and programmatically set the TTL of successful check items to null.

Set the TTL settings of the Azure Cosmos DB container to 3600, and programmatically set the TTL of defect detection items to -1.

Set the TTL settings of the Azure Cosmos DB container to 3600, and programmatically set the TTL of successful check items to -1.

Explanation

You should set the TTL settings of the Azure Cosmos DB container to 3600, and programmatically set the TTL of defect detection items to -1. This sets the TTL for all items in the container to 1 hour, except for the defect detection ones, which are kept indefinitely.

You should not set the TTL settings of the Azure Cosmos DB container to 3600, and programmatically set the TTL of successful check items to -1. This sets the TTL for all items in the container to 1 hour, except for the successful check ones, which are kept indefinitely.

You should not set the TTL settings of the Azure Cosmos DB container to -1, and programmatically set the TTL of defect detection items to null. With container TTL set to -1 (infinity) and item TTL set to null (default), items will never expire.

You should not set the TTL settings of the Azure Cosmos DB container to -1, and programmatically set the TTL of successful check items to null. With container TTL set to -1 (infinity) and item TTL set to null (default), items will never expire.

References

[Time to Live \(TTL\) in Azure Cosmos DB](#)

[Configure time to live in Azure Cosmos DB](#)

Question156 Optimize Azure data solutions - D

You manage an Azure SQL Database service that uses a single database. A particular query is taking a long time to run.

You need to determine what about the query is causing it to take so long.

What should you do?

Choose the correct answer

Run SQL Profiler.

Display the actual execution plan.

Drop all clustered indexes.

Display the estimated execution plan.

Explanation

You should display the actual execution plan. Azure SQL Database analyzes the most efficient way to access data. An execution plan describes how Azure SQL Database should access data, such as which tables to query first. By displaying the actual execution plan, you can graphically determine what is causing the query to take so long.

You should not display the estimated execution plan. The estimated execution plan shows how it expects Azure SQL Database to access tables. But it does not include the actual statistics.

You should not run SQL Profiler. SQL Profiler allows you to view the queries that are being run. However, it cannot identify the bottlenecks within a query.

You should not drop all clustered indexes. A clustered index helps Azure SQL Database to organize rows in a table sequentially, which can sometimes improve performance.

References

[Display the Estimated Execution Plan](#)

[Manual tune query performance in Azure SQL Database](#)

[Execution Plans](#)

Question 157 Optimize Azure data solutions – Not complete question

You manage an Azure SQL Database. You run the query and display the execution plan shown in the exhibit. There are 10 rows in the Department table.

You need to determine whether you can improve the performance of the query.

What should you conclude?

Choose the correct answer

You need to remove the clustered index from the Department table.

Performance is optimal

You need to add a non-clustered index on the Name column of the Department table.

You need to remove the clustered index from the Employee table.

Explanation

You should conclude that performance is optimal. The seek performed on the Department table has a cost (53%) that is almost equal to the scan on the Employee table (47%). There is nothing more than can be done to improve this query.

You should not remove the clustered index from the Employee or Department table. A clustered index organizes rows physically based on the value of the index. However, in this scenario, the cost of both index scans is approximately the same, so there is nothing more that can be done to improve this query.

You should not add a non-clustered index on the Name column of the Department table. On a large table, this would improve performance. A non-clustered index creates a separate index table that references the main table. This speeds up searches when you filter by specific values, such as the Name column of the Department table. However, on small tables this can actually decrease performance, because both the index table and the main table need to be searched.

References

[Manual tune query performance in Azure SQL Database](#)

Question158 Optimize Azure data solutions-

You manage an Azure SQL Data Warehouse. You have three tables:

DimProduct: Small table that changes infrequently

DimCustomer: Small table that changes infrequently

FactSales: Large table that changes frequently

All three tables use round-robin distribution. Queries join all three tables.

You need to optimize these tables for performance.

For each of the following statements, select Yes if the statement is true. Otherwise, select No.

Statement	Yes	No
You should use a replicated table for DimProduct.	<input type="radio"/>	<input type="radio"/>
You should use a replicated table for DimCustomer.	<input type="radio"/>	<input type="radio"/>
You should use a replicated table for FactSales.	<input type="radio"/>	<input type="radio"/>

Explanation

You should use a replicated table for DimProduct and DimCustomer. Both of these tables are small dimension tables that change infrequently. By replicating the tables across compute nodes, you eliminate the need to transfer data during joins with the FactSales table. This improves performance.

You should not use a replicated table for FactSales. You should instead use hash distribution for the FactSales table. Hash distribution copies rows across compute nodes by using a hashing function against a column. Rows with columns that have the same hash value are copied to the same compute node. This improves performance when querying large fact tables.

References

[Guidance for designing distributed tables in Azure SQL Data Warehouse](#)

[Design guidance for using replicated tables in Azure SQL Data Warehouse](#)

Question159 Optimize Azure data solutions-

You manage an Azure SQL Data Warehouse. You have three tables:

DimExam: 2000 rows, 200 megabytes (MB), changes infrequently

DimCandidate: 400 rows, 40 MB, changes infrequently

FactScores: 100,000,000 rows, 500 gigabytes (GB), changes frequently

All three tables use round-robin distribution. Queries join all three tables.

You need to optimize these tables for performance.

For each of the following statements, select Yes if the statement is true. Otherwise, select No.

Statement	Yes	No
You should use hash distribution for DimExam.	<input type="radio"/>	<input type="radio"/>
You should use hash distribution for DimCandidate.	<input type="radio"/>	<input type="radio"/>
You should use hash distribution for FactScores.	<input type="radio"/>	<input type="radio"/>

Explanation

You should use hash distribution for the FactScores table. Hash distribution copies rows across compute nodes by using a hashing function against a column. Rows with columns that have the same hash value are copied to the same compute node. This improves performance when querying large fact tables that change frequently.

You should not use hash distribution for DimExam and DimCandidate. You should use replicated tables. Both of these tables are small dimension tables that change infrequently. By replicating the tables across compute nodes, you eliminate the need to transfer data during joins with the FactScores table. This improves performance.

References

[Design guidance for using replicated tables in Azure SQL Data Warehouse](#)

[Guidance for designing distributed tables in Azure SQL Data Warehouse](#)

Question160 Optimize Azure data solutions-

You manage an Azure SQL Data Warehouse. You have four tables:

DimProduct: 2000 rows, 200 megabytes (MB), changes infrequently

DimCustomer: 400 rows, 40 megabytes (MB), changes infrequently

FactSales: 100,000,000 rows, 500 gigabytes (GB), changes frequently

FactOrders: 100,000,000 rows, 500 gigabytes (GB), changes frequently

All three tables use hash distribution. Queries join the DimProduct, DimCustomer, and FactSales tables. The FactOrders table contains all the data it needs.

You need to optimize these tables for performance.

For each of the following statements, select Yes if the statement is true. Otherwise, select No.

Statement	Yes	No
You should use a replicated table for DimProduct.	<input type="radio"/>	<input type="radio"/>
You should use a replicated table for DimCustomer.	<input type="radio"/>	<input type="radio"/>

You should use a replicated table for FactSales.	<input type="radio"/>	<input type="radio"/>
You should use round-robin distribution for FactOrders.	<input type="radio"/>	<input type="radio"/>

Explanation

You should use a replicated table for DimProduct and DimCustomer. Both of these tables are small dimension tables that change infrequently. By replicating the tables across compute nodes, you eliminate the need to transfer data during joins with the FactSales table. This improves performance.

You should not use a replicated table for FactSales. This table should use hash distribution because it is a large fact table with a column that has many distinct values, and the table changes frequently. Hash distribution copies rows across compute nodes by using a hashing function against a column. Rows with columns that have the same hash value are copied to the same compute node. This improves performance when querying large fact tables that change frequently.

You should use round-robin distribution for the FactOrders table. Round-robin distribution copies rows equally across all compute nodes. This is beneficial when the table does not need to join other tables. It also improves performance during data loads.

References

[Guidance for designing distributed tables in Azure SQL Data Warehouse](#)

[Design guidance for using replicated tables in Azure SQL Data Warehouse](#)

Question161 Optimize Azure data solutions-

You manage an Azure Streaming Analytics job.

You need to administer the job so that it uses optimal performance.

For each of the following statements, select Yes if the statement is true. Otherwise, select No.

Statement	Yes	No
You should start with two Streaming Units (SUs) for queries that do not use PARTITION BY.	<input type="radio"/>	<input type="radio"/>
You should keep the Streaming Unit (SU) metric below 80 percent.	<input type="radio"/>	<input type="radio"/>
You should allocate more SUs than you need.	<input type="radio"/>	<input type="radio"/>

Explanation

You should not start with two SUs for queries that do not use PARTITION BY. You should start with six SUs for queries that do not use PARTITION BY. This is considered a best practice.

You should allocate more SUs than you need. This is another best practice.

You should keep the SU metric below 80 percent. This allows Streaming Analytics to account for usage spikes.

References

[Understand and adjust Streaming Units](#)

Question162 Optimize Azure data solutions - D

You manage an Azure Data Lake Storage Gen2 account. Your source data is stored on file servers and SQL Server on-premises. You anticipate that it will take a long time to copy the data from your company to Data Lake over the public internet.

You need to ensure optimal performance when copying the source data to Azure.

What should you do?

Choose the correct answer

Create an Azure Data Lake Analytics account.

Create an Azure Data Factory account.

Install Active Directory (AD) Connect on-premises.

Use ExpressRoute.

Explanation

You should use ExpressRoute. This creates a dedicated link between your on-premises datacenter and Azure. This improves performance when copying data to Azure.

You should not install AD Connect on-premises. AD Connect allows you to synchronize user accounts between on-premises AD and Azure AD.

You should not create an Azure Data Lake Analytics account. This is an analytics service that you can use to write queries that improve big data performance.

You should not create an Azure Data Factory account. Data Factory allows you to design data pipelines to copy and transform data.

References

[Tuning Azure Data Lake Storage Gen2 for performance](#)

[What is Azure Data Factory?](#)

[What is Azure Data Lake Analytics?](#)

[Azure AD Connect sync: Understand and customize synchronization](#)

Question163 Optimize Azure data solutions-

You are a data engineer for your company. You have several blob storage accounts that you need to manage.

You need to match the access tier with the blob usage.

Which access tier should you use for each scenario? To answer, drag the appropriate access tier to each blob usage. An access tier may be used once, more than once, or not at all.

Drag and drop the answers

Archive

Cool

Hot

Archive

Cool

Hot

Drag and drop the answers

You need to access employee records once annually for auditing purposes.

Archive

You need to access employee payroll hours every month.

Cool

You need to access employee login and logout times daily.

Hot

Archive

Cool

Hot

Explanation

You should use the Archive tier when you need to access employee records once annually for auditing purposes. The Archive tier is optimized for storing data that is rarely accessed and that is kept for at least 180 days.

You should use the Cool tier when you need to access employee payroll hours every month. The Cool tier is optimized for storing data that is accessed infrequently and that is kept for at least 30 days.

You should use the Hot tier when you need to access employee login and logout times daily. The Hot tier is optimized for storing data that is accessed frequently.

References

[Azure Blob storage: hot, cool, and archive access tiers](#)

Question164 Optimize Azure data solutions-

Your company has an application for technical writers to share, review and publish exam questions. The application allows live reviews for questions and provides reports about

question performance after the questions are published. All questions and reviews are stored in Azure SQL Database.

Writers are able to report question metrics in an exam results table with millions of rows. Reports include the percentage of correct answers, average question rating and answer count per question.

After a question is published, live reviews are disabled and previous reviews are archived in a separate database.

Writers report the following performance issues with the application:

- Some delays are occurring during the live reviews.
- Reports are taking too long to be generated.

You need to resolve the performance issues.

Which two features should you use? Each correct answer presents a complete solution.

Choose the correct answers

Heap table

Memory-optimized table

Nonclustered index

Columnstore index

Partitioned view

Explanation

You should use a columnstore index in the exam results table. A columnstore index stores column values and increases the aggregate queries that use these indexes. Question reports consist in aggregate queries.

You could also use a memory-optimized table for the reviews table. Memory-optimized tables store all data and schema in memory, increasing the performance for queries involving this table.

You should not use a partitioned view. You could use this feature to split large tables across multiple smaller tables. However, this does not solve the reported performance issues.

You should not use a nonclustered index. This index is generally used to increase

filter performance and to lookup rows with specific values. Using a clustered index does not solve the performance issues.

You should not use a heap table. A heap is a table without a clustered index with table data stored without any specific order. Every query in a heap should perform a table scan. A full table scan has better performance than a table with a clustered index for small tables, like a list of nine question types. However, this scenario does not apply to live reviews and results tables. Using a heap table does not solve the performance issues.

References

[Monitoring and performance tuning](#)

[Columnstore indexes: Overview](#)

[Heaps \(Tables without Clustered Indexes\)](#)

[Clustered and Nonclustered Indexes Described](#)

[Introduction to Memory-Optimized Tables](#)

[Views](#)

Question165 Optimize Azure data solutions-

Your company manages an ecommerce platform for a large retailer. The platform is composed of thousands of web servers running on Azure virtual machines (VMs). The platform generates approximately three gigabytes (GB) of log data in a day.

The log files for each server are stored in Azure Data Lake Storage Gen2 inside a single folder. This data is processed and analyzed on an HDInsight cluster.

You need to provide some performance improvements while minimizing cost.

Which two changes should you implement? Each correct answer presents a complete solution.

Choose the correct answers

Combine the daily log files into one file.

Use a cool tier for Azure Data Lake Storage Gen2.

Increase the number of worker nodes.

Separate the log files into a daily generated folder.

Explanation

You should combine the daily log files into one file. Analytics engines such as HDInsight have a per-file overhead during processing. Using small files will quickly utilize the available throughput for Azure Data Lake Storage Gen2 to read data. Organizing your data in larger files with sizes from 256MB to 100GB results in better performance.

You should separate the log files into a daily generated folder. Partitioning your files in time series folders helps HDInsight to load only a subset of the data, thus improving performance. You should separate your data in a hierarchic folder structure like \DataSet\YYYY\MM\DD\datafile_YYYY_MM_DD.tsv.

You should not increase the number of worker nodes. This increases the throughput for HDInsights, resulting in better performance. However, this solution also increases costs. You could optimize the file size combining them, reducing high throughput need.

You should not use a cool tier for Azure Data Lake Storage Gen2. This tier is optimized to store data with infrequent access and could be used to reduce storage costs. However, if you frequently access data in this tier, data access and transaction costs are higher than if you would use a hot tier.

References

[Tuning Azure Data Lake Storage Gen2 for performance](#)

[Capacity planning for HDInsight clusters](#)

[Azure Storage scalability and performance targets for storage accounts](#)

Question166 Optimize Azure data solutions-

Your team manages an image recognition process for a security firm. The solution runs on an Azure virtual machine (VM) at a daily frequency.

Images are uploaded to an Azure Blob Storage container. The VM processes the results using the data uploaded from the previous day and stores the results in the same blob storage.

You must meet the following data retention requirements:

- Daily results must be accessible for one week.

- Data for the current month must be available but is rarely used.
- Current year data must be stored for auditing purposes.
- Audit data must be requested at least one day before being accessed.

You need to use a lifecycle policy that minimizes cost.

How should you configure the lifecycle policy? To answer, drag a JSON segment to the most appropriate location. A JSON segment may be used once, more than once, or not at all.

Drag and drop the answers

delete

baseBlob

tierToArchive

7

30

delete

baseBlob

tierToArchive

7

30

snapshot

1

Drag and drop the answers

```
{
  "rules": [{
    "name": "lifecycleRule", "enabled": true, "type": "Lifecycle",
    "definition": {
      "actions": {
        "baseBlob": {
          "tierToArchive": { "daysAfterModificationGreaterThan": 30 },
          "delete": { "daysAfterModificationGreaterThan": 365 },
          "tierToCool": { "daysAfterModificationGreaterThan": 7 }
        }
      }
    }
  }]
}
```

delete

baseBlob

tierToArchive

7

30

snapshot

1

Explanation

You should configure baseBlob to manage the blob storage life cycle. This JSON object contains the rules that you need to enforce in your data storage.

You should configure tierToCool and set it to seven days. This rule moves data to a cool tier after seven days and satisfies the infrequent data access to the monthly data requirement, minimizing storage costs.

You should configure tierToArchive and set it to 30 days. This moves data older than a month to the archive tier. In an archive tier, you cannot access data directly. This requires a rehydration process to a cool or hot tier. This process takes a few hours to complete. Because audit data should be requested at least one day before, you have enough time to wait for the rehydration process.

You should configure delete for the 365 days policy. Only the current year data needs to be stored for auditing. Data older than a year should be deleted to minimize storage costs.

You should not configure a snapshot segment. If versioning is enabled in a blob container, each time a blob is modified, a new snapshot version is created. You can only delete an older snapshot of your files. Snapshot data retention for this solution is not applicable.

References

[Manage the Azure Blob storage lifecycle](#)

[Rehydrate blob data from the archive tier](#)

Question167 Optimize Azure data solutions-

You manage a real-time data analysis solution hosted in Azure.

The solution uses Azure Event Hubs to ingest data and analyze the data in real-time with an Azure Stream Analytics job.

The Stream Analytics job uses 18 Streaming Units (SU). Stream Analytics job metrics and SU percentage utilization metrics have been 90% in average over the last month.

You need to optimize the Azure Stream Analytics job performance.

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

Choose the correct answers

Configure event ordering.

Decrease the SU count for the job.

Partition data for query parallelization.

Increase the SU count for the job.

Use Stream Analytics JavaScript user-defined functions.

Explanation

You should increase the SU count for the job. SU represents the compute resources allocated to a Stream Analytics job. Too high SU % utilization indicates a high memory consumption and higher latency to process events. It is recommended to keep SU utilization below 80% for occasional spikes.

You should partition data for query parallelization. Partitioning the data increases job event throughput and allows an optimized use of the available SU. Combining query parallelization with a higher SU count results in optimized job performance.

You should not configure event ordering. Event ordering delays the job output of Streaming Analytics jobs, increasing job latency and performance.

You should not use Stream Analytics JavaScript user-defined functions. This allows you to use JavaScript functions for query analysis. Implementing JavaScript

user-defined functions does not increase job performance.

You should not decrease the SU count for the job. You could optimize Stream Analytics job performance if the allocated resources are over-provisioned. Current SU % utilization is above the recommended 80% and reducing SU could cause performance issues.

References

[Leverage query parallelization in Azure Stream Analytics](#)

[Understand and adjust Streaming Units](#)

[Tutorial: Azure Stream Analytics JavaScript user-defined functions](#)

[Configuring event ordering policies for Azure Stream Analytics](#)

Question168 Optimize Azure data solutions-

You need to load data into an Azure Synapse Analytics SQL pool. The data is stored in a general purpose v2 Azure storage account as text files.

You need to use the fastest data loading option.

Which two options can you use to meet your goal? Each correct answer presents a complete solution.

Choose the correct answers

Use SQL Server Integration Services (SSIS).

Use a Copy Activity in Azure Data Factory.

Use the Bulk Copy Program utility (BCP).

Write and run PolyBase T-SQL commands.

Use a Copy Activity in Azure Data Factory with the bulk insert option.

Use the SQL BulkCopy API.

Explanation

You can write and run PolyBase T-SQL commands. This is a fully parallel operation and is the fastest option.

You can also use a Copy Activity in Azure Data Factory with the copy method set

to PolyBase. This option creates and executes the Polybase commands automatically. This also offers the fastest performance.

The following options are not parallel operations and are thus slower:

- BCP
- SQL BulkCopy API
- SSIS
- Azure Data Factory using a Copy Activity and the bulk insert option

References

[Data loading strategies for data warehousing](#)

[Tutorial: Load the New York Taxicab dataset](#)

[CREATE TABLE AS SELECT \(Azure SQL Data Warehouse\)](#)

[Copy and transform data in Azure Synapse Analytics \(formerly Azure SQL Data Warehouse\) by using Azure Data Factory](#)

[bcp Utility](#)

[SQL Server Integration Services](#)

Question169 Optimize Azure data solutions – Not complete question

You have an Azure Synapse Analytics SQL pool.

The output of the EXPLAIN command is shown in the exhibit.

You need to optimize the performance of a query while minimizing cost.

What should you do?

Choose the correct answer

Update the statistics on the FactInternetSales table.

Upgrade the performance level.

Re-create the FactInternetSales table using hash-distribution on the ProductKey column

Re-create the FactInternetSales table using round-robin distribution.

Re-create the FactInternetSales table using page compression.

Explanation

You should re-create the FactInternetSales table using hash-distribution on the ProductKey column. The EXPLAIN command shows the execution plan of the control node. The query has a SHUFFLE_MOVE operation with a high relative cost (0.007584336). It has the same value as the total cost of the query. This indicates that data movement between the 60 distributions of the SQL pool is consuming most of the resources.

The query to optimize is:

```
SELECT [ProductKey], COUNT(*)  
FROM [dbo].[FactInternetSales]  
GROUP BY [ProductKey]
```

Data movement for this query can be avoided by re-creating the table as a hash-distributed table using ProductKey as the hash key. By doing this, all data with the same ProductKey will be stored on the same distribution node. Each node can perform the grouping on a part of the table, and this only needs to be combined on the control node to get the final aggregated resultset.

You cannot change the distribution of an existing table. The result can be achieved by re-creating the table.

You should not re-create the FactInternetSales table using round-robin distribution. The query plan indicates that data movement between the compute nodes is using the resources. In this case data movement is caused by using round-robin distribution, because rows with the same ProductKey will not be grouped together on the same compute node. This can cause a longer duration of the query.

You should not re-create the FactInternetSales table using page compression. By default the Azure Synapse Analytics SQL pool uses a clustered columnstore index. This will achieve the best performance for aggregate queries like this. Using page compression will slow down performance.

You should not upgrade the performance level. This will increase cost and is not the best option to meet the requirements.

You should not update the statistics on the FactInternetSales table. Updating statistics on the table will not prevent data movement between the compute nodes from happening. This is not the best option to meet the requirements.

References

[Guidance for designing distributed tables in SQL Analytics](#)

[CREATE TABLE AS SELECT \(Azure SQL Data Warehouse\)](#)

[EXPLAIN \(Transact-SQL\)](#)