Practice Test 1

Question 1: Skipped

You are working as a data engineer for your company and your company has an Azure Databricks workspace. This workspace will be containing 3 types of workload.

· One workload for data engineers that would make use of Python and SQL

· One workload for jobs that would run notebooks that would make use of Python, Spark, Scala and SQL

· One workload that data scientists would use to perform ad hoc analysis in Scala and R

The following standards need to be adhered for the different Databricks environments

· The data engineers need to share a cluster

· The cluster that runs jobs would be triggered via a request. The data scientists and data engineers would provide package notebooks that would need to be deployed to the cluster.

· There are three data scientists currently. Every data scientist has to be assigned their own cluster. The cluster needs to terminate automatically after 120 minutes of inactivity.

You have been given the task to create new Databrick clusters for the workloads.

Now, you decide to create a High Concurrency cluster for each data scientist, a High Concurrency cluster for the data engineers and a standard cluster for the jobs.

Do you think that this can fulfil the requirement?

- ○

  Yes

- ○

  No

  (Correct)

Explanation

Data scientists should have standard clusters.

Standard clusters

Standard clusters are recommended for a single user. Standard clusters can run workloads developed in any language: Python, R, Scala, and SQL.

High Concurrency clusters

A High Concurrency cluster is a managed cloud resource. The key benefits of High Concurrency clusters are that they provide Apache Spark-native fine-grained sharing for maximum resource utilization and minimum query latencies.

High Concurrency clusters work only for SQL, Python, and R. The performance and security of High Concurrency clusters is provided by running user code in separate processes, which is not possible in Scala.

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

Question 2: Skipped
You are working as a data engineer for your company and your company has an Azure Databricks workspace. This workspace will be containing 3 types of workload.

· One workload for data engineers that would make use of Python and SQL

· One workload for jobs that would run notebooks that would make use of Python, Spark, Scala and SQL

· One workload that data scientists would use to perform ad hoc analysis in Scala and R

The following standards need to be adhered for the different Databricks environments

· The data engineers need to share a cluster

· The cluster that runs jobs would be triggered via a request. The data scientists and data engineers would provide package notebooks that would need to be deployed to the cluster.

· There are three data scientists currently. Every data scientist has to be assigned their own cluster. The cluster needs to terminate automatically after 120 minutes of inactivity.

You have been given the task to create new Databrick clusters for the workloads.

Now, you decide to create a standard cluster for each data scientist, a standard cluster for the data engineers, and a High Concurrency cluster for the jobs.

Do you think that this can fulfil the requirement?

- ○

  Yes

- ○

Explanation

Data Engineers should have high concurrent clusters.

Standard clusters

Standard clusters are recommended for a single user. Standard clusters can run workloads developed in any language: Python, R, Scala, and SQL.

High Concurrency clusters

A High Concurrency cluster is a managed cloud resource. The key benefits of High Concurrency clusters are that they provide Apache Spark-native fine-grained sharing for maximum resource utilization and minimum query latencies.

High Concurrency clusters work only for SQL, Python, and R. The performance and security of High Concurrency clusters is provided by running user code in separate processes, which is not possible in Scala.

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

Question 3: Skipped
You are working as a data engineer for your company and your company has an Azure Databricks workspace. This workspace will be containing 3 types of workload.

· One workload for data engineers that would make use of Python and SQL

· One workload for jobs that would run notebooks that would make use of Python, Spark, Scala and SQL

· One workload that data scientists would use to perform ad hoc analysis in Scala and R

The following standards need to be adhered for the different Databricks environments

· The data engineers need to share a cluster

· The cluster that runs jobs would be triggered via a request. The data scientists and data engineers would provide package notebooks that would need to be deployed to the cluster.

· There are three data scientists currently. Every data scientist has to be assigned their own cluster. The cluster needs to terminate automatically after 120 minutes of inactivity.

You have been given the task to create new Databrick clusters for the workloads.

Now, you decide to create a Standard cluster for each data scientist, a High Concurrency cluster for the data engineers and a Standard cluster for the jobs.

Do you think that this can fulfil the requirement?

- ○

(Correct)

- ○

No

Explanation

Standard clusters

Standard clusters are recommended for a single user. Standard clusters can run workloads developed in any language: Python, R, Scala, and SQL.

High Concurrency clusters

A High Concurrency cluster is a managed cloud resource. The key benefits of High Concurrency clusters are that they provide Apache Spark-native fine-grained sharing for maximum resource utilization and minimum query latencies.

High Concurrency clusters work only for SQL, Python, and R. The performance and security of High Concurrency clusters is provided by running user code in separate processes, which is not possible in Scala.

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

Question 4: Skipped
You are working as a data engineer for your company and your company has an Azure Databricks workspace. This workspace will be containing 3 types of workload.

· One workload for data engineers that would make use of Python and SQL

· One workload for jobs that would run notebooks that would make use of Python, Spark, Scala and SQL

· One workload that data scientists would use to perform ad hoc analysis in Scala and R

The following standards need to be adhered for the different Databricks environments

· The data engineers need to share a cluster

· The cluster that runs jobs would be triggered via a request. The data scientists and data engineers would provide package notebooks that would need to be deployed to the cluster.

· There are three data scientists currently. Every data scientist has to be assigned their own cluster. The cluster needs to terminate automatically after 120 minutes of inactivity.

You have to create new Databrick clusters for the workloads

You have been given the task to create new Databrick clusters for the workloads.

Now, you decide to create a High Concurrency cluster for each data scientist, a High Concurrency cluster for the data engineers, and a High Concurrency cluster for the jobs.

Do you think that this can fulfil the requirement?

- ○

  Yes

- ○

  No

  (Correct)

Explanation

Data Scientists and Jobs should have Standard Clusters and Data Engineers should have High Concurrency clusters.

Standard clusters

Standard clusters are recommended for a single user. Standard clusters can run workloads developed in any language: Python, R, Scala, and SQL.

High Concurrency clusters

A High Concurrency cluster is a managed cloud resource. The key benefits of High Concurrency clusters are that they provide Apache Spark-native fine-grained sharing for maximum resource utilization and minimum query latencies.

High Concurrency clusters work only for SQL, Python, and R. The performance and security of High Concurrency clusters is provided by running user code in separate processes, which is not possible in Scala.

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

Question 5: Skipped
You are working as a data engineer for your company and you have been given the task to design a Hadoop Distributed File System architecture. You have to use Microsoft Azure Data Lake as the data storage repository. Also, you have to ensure that the data repository has a resilient data schema.

What should you be using to provide data access to clients?

- ○

DataNode

(Correct)

○

NameNode

○

PrimaryNode

○

SecondaryNode

Explanation

NameNode and DataNodes

HDFS has a master/slave architecture. An HDFS cluster consists of a single NameNode, a master server that manages the file system namespace and regulates access to files by clients. In addition, there are a number of DataNodes, usually one per node in the cluster, which manage storage attached to the nodes that they run on. HDFS exposes a file system namespace and allows user data to be stored in files. Internally, a file is split into one or more blocks and these blocks are stored in a set of DataNodes. The NameNode executes file system namespace operations like opening, closing, and renaming files and directories. It also determines the mapping of blocks to DataNodes. **The DataNodes are responsible for serving read and write requests from the file system's clients.** The DataNodes also perform block creation, deletion, and replication upon instruction from the NameNode.

https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html#NameNode+and+DataNodes

Question 6: Skipped

You are working as a data engineer for your company and you have been given the task to design a Hadoop Distributed File System architecture. You have to use Microsoft Azure Data Lake as the data storage repository. Also, you have to ensure that the data repository has a resilient data schema.

What should you be using to run operations on files and directories on the file system?

○

DataNode

○

NameNode

(Correct)

○

PrimaryNode

○

SecondaryNode

Explanation

NameNode and DataNodes

HDFS has a master/slave architecture. **An HDFS cluster consists of a single NameNode, a master server that manages the file system namespace and regulates access to files by clients.** In addition, there are a number of DataNodes, usually one per node in the cluster, which manage storage attached to the nodes that they run on. HDFS exposes a file system namespace and allows user data to be stored in files. Internally, a file is split into one or more blocks and these blocks are stored in a set of DataNodes. The NameNode executes file system namespace operations like opening, closing, and renaming files and directories. It also determines the mapping of blocks to DataNodes. The DataNodes are responsible for serving read and write requests from the file system's clients. The DataNodes also perform block creation, deletion, and replication upon instruction from the NameNode.

https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html#NameNode+and+DataNodes

Question 7: Skipped
You are working as a data engineer for your company and you have been given the task to design a Hadoop Distributed File System architecture. You have to use Microsoft Azure Data Lake as the data storage repository. Also, you have to ensure that the data repository has a resilient data schema.
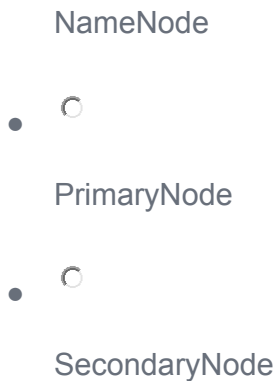
What should you be using to perform block creation, deletion, and replication?

○

DataNode

(Correct)

○

NameNode

- ○

PrimaryNode

- ○

SecondaryNode

Explanation

NameNode and DataNodes

HDFS has a master/slave architecture. An HDFS cluster consists of a single NameNode, a master server that manages the file system namespace and regulates access to files by clients. In addition, there are a number of DataNodes, usually one per node in the cluster, which manage storage attached to the nodes that they run on. HDFS exposes a file system namespace and allows user data to be stored in files. Internally, a file is split into one or more blocks and these blocks are stored in a set of DataNodes. The NameNode executes file system namespace operations like opening, closing, and renaming files and directories. It also determines the mapping of blocks to DataNodes. The DataNodes are responsible for serving read and write requests from the file system's clients. **The DataNodes also perform block creation, deletion, and replication upon instruction from the NameNode.**

https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html#NameNode+and+DataNodes

Question 8: Skipped
A company wants to use a set of services on Azure. They want to make use of Platform-as-a-service products to create a new data pipeline process. They have the following requirements

Data Ingestion

· This layer must provide access to multiple sources

· This layer must provide the ability to orchestrate a workflow

· It must also provide the capability to run SQL Server Integration Service packages

Storage

· The storage layer must be optimized for Big Data workloads

· It must provide encryption of data at rest

· There must be no size constraints

Prepare and Train

· This layer must provide a fully managed interactive workspace for exploration and visualization

· Here you should be able to program in R, SQL or Scala

· It must provide seamless user authentication with Azure Active Directory

Model and Service

· This layer must provide support for SQL language

· It must implement native columnar storage

What technology should they be using for the "Storage" layer?

- ○

  Azure Data Lake Storage

    (Correct)

- ○

  Azure Blob Storage

- ○

  Azure Files

- ○

  Azure SQL Data warehouse

Explanation

Azure Data Lake Storage Gen2 is a set of capabilities dedicated to big data analytics, built on Azure Blob storage. Data Lake Storage Gen2 is the result of converging the capabilities of our two existing storage services, Azure Blob storage and Azure Data Lake Storage Gen1. Features from Azure Data Lake Storage Gen1, such as file system semantics, directory, and file level security and scale are combined with low-cost, tiered storage, high availability/disaster recovery capabilities from Azure Blob storage. **Designed from the start to service multiple petabytes of information while sustaining hundreds of gigabits of throughput, Data Lake Storage Gen2 allows you to easily manage massive amounts of data.**

https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-introduction

Question 9: Skipped

A company wants to use a set of services on Azure. They want to make use of Platform-as-a-service products to create a new data pipeline process. They have the following requirements

Data Ingestion

· This layer must provide access to multiple sources

· This layer must provide the ability to orchestrate a workflow

· It must also provide the capability to run SQL Server Integration Service packages

Storage

· The storage layer must be optimized for Big Data workloads

· It must provide encryption of data at rest

· There must be no size constraints

Prepare and Train

· This layer must provide a fully managed interactive workspace for exploration and visualization

· Here you should be able to program in R, SQL or Scala

· It must provide seamless user authentication with Azure Active Directory

Model and Service

· This layer must provide support for SQL language

· It must implement native columnar storage

What technology should they be using for the "Data Ingestion" layer?

- ○

  Azure Logic Apps

- ○

  Azure Data Factory

  (Correct)

- ○

  Azure Automation

- ○

  Azure Functions

Explanation

It is the cloud-based ETL and data integration service that allows you to create data-driven workflows for orchestrating data movement and transforming data at scale.Using Azure Data Factory, you can create and schedule data-driven workflows (called pipelines) that can ingest data from disparate data stores. You can build complex ETL processes that transform data visually with data flows or by using compute services such as Azure HDInsight Hadoop, Azure Databricks, and Azure SQL Database.

https://docs.microsoft.com/en-us/azure/data-factory/introduction

Question 10: Skipped
Case Study

A company wants to use a set of services on Azure. They want to make use of Platform-as-a-service products to create a new data pipeline process. They have the following requirements

Data Ingestion

· This layer must provide access to multiple sources

· This layer must provide the ability to orchestrate a workflow

· It must also provide the capability to run SQL Server Integration Service packages

Storage

· The storage layer must be optimized for Big Data workloads

· It must provide encryption of data at rest

· There must be no size constraints

Prepare and Train

· This layer must provide a fully managed interactive workspace for exploration and visualization

· Here you should be able to program in R, SQL or Scala

· It must provide seamless user authentication with Azure Active Directory

Model and Service

· This layer must provide support for SQL language

· It must implement native columnar storage

What technology should they be using for the "Prepare and Train" layer?

- ○

  HDInsight Apache Spark Cluster

- ○

  Azure Databricks

  (Correct)

- ○

  HDInsight Apache Storm Cluster

- ○

  Azure SQL Data warehouse

Explanation

Azure Databricks is an Apache Spark-based analytics platform optimized for the Microsoft Azure cloud services platform. Designed with the founders of Apache Spark, Databricks is integrated with Azure to provide one-click setup, streamlined workflows, and an interactive workspace that enables collaboration between data scientists, data engineers, and business analysts.

https://docs.microsoft.com/en-us/azure/databricks/scenarios/what-is-azure-databricks

Azure Databricks now supports Azure Active Directory (Azure AD) conditional access, which allows administrators to control where and when users are permitted to sign in to Azure Databricks.

https://azure.microsoft.com/en-in/updates/azure-active-directory-conditional-access-in-azure-databricks/#:~:text=Azure%20Databricks%20now%20supports%20Azure,sign%20in%20to%20Azure%20Databricks.

Standard clusters are recommended for a single user. Standard clusters can run workloads developed in any language: Python, R, Scala, and SQL.

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

Question 11: Skipped

A company wants to use a set of services on Azure. They want to make use of Platform-as-a-service products to create a new data pipeline process. They have the following requirements

Data Ingestion

· **This layer must provide access to multiple sources**

· **This layer must provide the ability to orchestrate a workflow**

· **It must also provide the capability to run SQL Server Integration Service packages**

Storage

· **The storage layer must be optimized for Big Data workloads**

· **It must provide encryption of data at rest**

· **There must be no size constraints**

Prepare and Train

· **This layer must provide a fully managed interactive workspace for exploration and visualization**

· **Here you should be able to program in R, SQL or Scala**

· **It must provide seamless user authentication with Azure Active Directory**

Model and Service

· **This layer must provide support for SQL language**

· **It must implement native columnar storage**

What technology should they be using for the "Model and Service" layer?

- ○

  HDInsight Apache Kafta cluster

- ○

  Azure SQL Data warehouse

  (Correct)

- ○

  Azure Data Lake Storage

- ○

  Azure Blob Storage

Explanation

Dedicated SQL pool of data warehouse stores data in relational tables with columnar storage.

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

Question 12: Skipped
You are working as a data engineer for a company that make use of Azure Data Lake Gen 2 storage account. They want it to be used to store Big Data related to an application. Your company now wants to implement logging.

After discussion, they decide to create an Azure Automation runbook which would be used to copy events.

Do you think this would fulfil the requirement?

- ○

  Yes

- ○

  No

  (Correct)

Explanation

Organizations can enable diagnostic logging for their Azure Data Lake Storage Gen1 account to collect data access audit trails that provides information such as list of users accessing the data, how frequently the data is accessed, how much data is stored in the account, etc. When enabled, the diagnostics and/or requests are logged on a best-effort basis. Both Requests and Diagnostics log entries are created only if there are requests made against the service endpoint.

https://docs.microsoft.com/en-us/azure/data-lake-store/data-lake-store-diagnostic-logs

Question 13: Skipped
You are working as a data engineer for a company that make use of Azure Data Lake Gen 2 storage account. They want it to be used to store Big Data related to an application. Your company now wants to implement logging.

After the discussion, they decide to use the information that is stored in Azure Active Directory reports.

Do you think this would fulfil the requirement?

- ○

Yes

O

No

(Correct)

Explanation

Organizations can enable diagnostic logging for their Azure Data Lake Storage Gen1 account to collect data access audit trails that provides information such as list of users accessing the data, how frequently the data is accessed, how much data is stored in the account, etc. When enabled, the diagnostics and/or requests are logged on a best-effort basis. Both Requests and Diagnostics log entries are created only if there are requests made against the service endpoint.

https://docs.microsoft.com/en-us/azure/data-lake-store/data-lake-store-diagnostic-logs

Question 14: Skipped
You are working as a data engineer for a company that make use of Azure Data Lake Gen 2 storage account. They want it to be used to store Big Data related to an application. Your company now wants to implement logging.

After discussion, decide to configure Azure Data Lake Storage diagnostics to store the logs and metric data in a storage account.

Do you think this would fulfil the requirement?

O

Yes

(Correct)

O

No

Explanation

Organizations can enable diagnostic logging for their Azure Data Lake Storage Gen1 account to collect data access audit trails that provides information such as list of users accessing the data, how frequently the data is accessed, how much data is stored in the account, etc. When enabled, the diagnostics and/or requests are logged on a best-effort basis. Both Requests and Diagnostics log entries are created only if there are requests made against the service endpoint.

https://docs.microsoft.com/en-us/azure/data-lake-store/data-lake-store-diagnostic-logs

Question 15: Skipped

Contoso relies on an extensive partner network for marketing, sales, and distribution. Contoso uses external companies that manufacture everything from the actual pharmaceutical to the packaging.

The majority of the company's data reside in Microsoft SQL Server database. Application databases fall into one of the following tiers:

Larger image

Applications	Tier	Replication	Notes
Internal Contoso	1	Yes	
Internal Contoso	2	SQL Data Sync	Data Sync between databases
Internal Partner	3	Yes	Replicate to Partner
External Contoso	4,5,6	Yes	
External Partner	7,8	No	Partner managed
Internal Distribution and Sales	9	Yes, once ingested at branches	Data ingested from Contoso branches
External Distribution and Sales	10	Yes, once ingested at Contoso main office	Data is ingested from multiple sources

The company has a reporting infrastructure that ingests data from local databases and partner services. Partners services consists of distributors, wholesales, and retailers across the world. The company performs daily, weekly, and monthly reporting.

Requirements -

Tier 3 and Tier 6 through Tier 8 application must use database density on the same server and Elastic pools in a cost-effective manner.

Applications must still have access to data from both internal and external applications keeping the data encrypted and secure at rest and in transit.

A disaster recovery strategy must be implemented for Tier 3 and Tier 6 through 8 allowing for failover in the case of server going offline.

Selected internal applications must have the data hosted in single Microsoft Azure SQL Databases.

Tier 1 internal applications on the premium P2 tier

Tier 2 internal applications on the standard S4 tier

The solution must support migrating databases that support external and internal application to Azure SQL Database. The migrated databases will be supported by Azure Data Factory pipelines for the continued movement, migration and updating of data both in the cloud and from local core business systems and repositories.

Tier 7 and Tier 8 partner access must be restricted to the database only.

In addition to default Azure backup behavior, Tier 4 and 5 databases must be on a backup strategy that performs a transaction log backup eve hour, a differential backup of databases every day and a full back up every week.

Back up strategies must be put in place for all other standalone Azure SQL Databases using Azure SQL-provided backup storage and capabilities.

Databases -

Contoso requires their data estate to be designed and implemented in the Azure Cloud. Moving to the cloud must not inhibit access to or availability of data.

Databases:

Tier 1 Database must implement data masking using the following masking logic:

Larger image

Data type	Masking requirement
A	Mask 4 or less string data type characters
B	Mask first letter and domain
C	Mask everything except characters at the beginning and end

Tier 2 databases must sync between branches and cloud databases and in the event of conflicts must be set up for conflicts to be won by on-premises databases.

Tier 3 and Tier 6 through Tier 8 applications must use database density on the same server and Elastic pools in a cost-effective manner.

Applications must still have access to data from both internal and external applications keeping the data encrypted and secure at rest and in transit.

A disaster recovery strategy must be implemented for Tier 3 and Tier 6 through 8 allowing for failover in the case of a server going offline.

Selected internal applications must have the data hosted in single Microsoft Azure SQL Databases.

Tier 1 internal applications on the premium P2 tier

Tier 2 internal applications on the standard S4 tier

Reporting -

Security and monitoring -

Security -

A method of managing multiple databases in the cloud at the same time is must be implemented to streamlining data management and limiting management access to only those requiring access.

Monitoring -

Monitoring must be set up on every database. Contoso and partners must receive performance reports as part of contractual agreements.

Tiers 6 through 8 must have unexpected resource storage usage immediately reported to data engineers.

The Azure SQL Data Warehouse cache must be monitored when the database is being used. A dashboard monitoring key performance indicators (KPIs) indicated by traffic lights must be created and displayed based on the following metrics:

Larger image

Metric	Description
A	Low cache hit %, high cache usage %
B	Low cache hit %, low cache usage %
C	High cache hit %, high cache usage %

Existing Data Protection and Security compliances require that all certificates and keys are internally managed in an on-premises storage.

You identify the following reporting requirements:

Azure Data Warehouse must be used to gather and query data from multiple internal and external databases

Azure Data Warehouse must be optimized to use data from a cache

Reporting data aggregated for external partners must be stored in Azure Storage and be made available during regular business hours in the connecting regions

Reporting strategies must be improved to real time or near real time reporting cadence to improve competitiveness and the general supply chain

Tier 9 reporting must be moved to Event Hubs, queried, and persisted in the same Azure region as the company's main office

Tier 10 reporting data must be stored in Azure Blobs

Issues -

Team members identify the following issues:

Both internal and external client application run complex joins, equality searches and group-by clauses. Because some systems are managed externally, the queries will not be changed or optimized by Contoso

External partner organization data formats, types and schemas are controlled by the partner companies

Internal and external database development staff resources are primarily SQL developers familiar with the Transact-SQL language.

Size and amount of data has led to applications and reporting solutions not performing are required speeds

Tier 7 and 8 data access is constrained to single endpoints managed by partners for access

The company maintains several legacy client applications. Data for these applications remains isolated form other applications. This has led to hundreds of databases being provisioned on a per application basis

Question

What should be used to process and query the ingested data for the Tier 9 data?

- ○

  Azure Notification Hubs

- ○

  Apache Cache for Redis

- ○

  Azure Functions

- ○

  Azure Stream Analytics

    (Correct)

Explanation

The Azure Stream Analytics service makes it easy to ingest, process, and analyze streaming data from Azure Event Hubs, enabling powerful insights to drive real-time actions.

https://docs.microsoft.com/en-us/azure/event-hubs/process-data-azure-stream-analytics

Contoso relies on an extensive partner network for marketing, sales, and distribution. Contoso uses external companies that manufacture everything from the actual pharmaceutical to the packaging.

The majority of the company's data reside in Microsoft SQL Server database. Application databases fall into one of the following tiers:

Larger image

Applications	Tier	Replication	Notes
Internal Contoso	1	Yes	
Internal Contoso	2	SQL Data Sync	Data Sync between databases
Internal Partner	3	Yes	Replicate to Partner
External Contoso	4,5,6	Yes	
External Partner	7,8	No	Partner managed
Internal Distribution and Sales	9	Yes, once ingested at branches	Data ingested from Contoso branches
External Distribution and Sales	10	Yes, once ingested at Contoso main office	Data is ingested from multiple sources

The company has a reporting infrastructure that ingests data from local databases and partner services. Partners services consists of distributors, wholesales, and retailers across the world. The company performs daily, weekly, and monthly reporting.

Requirements -

Tier 3 and Tier 6 through Tier 8 application must use database density on the same server and Elastic pools in a cost-effective manner.

Applications must still have access to data from both internal and external applications keeping the data encrypted and secure at rest and in transit.

A disaster recovery strategy must be implemented for Tier 3 and Tier 6 through 8 allowing for failover in the case of server going offline.

Selected internal applications must have the data hosted in single Microsoft Azure SQL Databases.

Tier 1 internal applications on the premium P2 tier

Tier 2 internal applications on the standard S4 tier

The solution must support migrating databases that support external and internal application to Azure SQL Database. The migrated databases will be supported by Azure Data Factory pipelines for the continued movement, migration and updating of data both in the cloud and from local core business systems and repositories.

Tier 7 and Tier 8 partner access must be restricted to the database only.

In addition to default Azure backup behavior, Tier 4 and 5 databases must be on a backup strategy that performs a transaction log backup eve hour, a differential backup of databases every day and a full back up every week.

Back up strategies must be put in place for all other standalone Azure SQL Databases using Azure SQL-provided backup storage and capabilities.

Databases -

Contoso requires their data estate to be designed and implemented in the Azure Cloud. Moving to the cloud must not inhibit access to or availability of data.

Databases:

Tier 1 Database must implement data masking using the following masking logic:

Larger image

Data type	Masking requirement
A	Mask 4 or less string data type characters
B	Mask first letter and domain
C	Mask everything except characters at the beginning and end

Tier 2 databases must sync between branches and cloud databases and in the event of conflicts must be set up for conflicts to be won by on-premises databases.

Tier 3 and Tier 6 through Tier 8 applications must use database density on the same server and Elastic pools in a cost-effective manner.

Applications must still have access to data from both internal and external applications keeping the data encrypted and secure at rest and in transit.

A disaster recovery strategy must be implemented for Tier 3 and Tier 6 through 8 allowing for failover in the case of a server going offline.

Selected internal applications must have the data hosted in single Microsoft Azure SQL Databases.

Tier 1 internal applications on the premium P2 tier

Tier 2 internal applications on the standard S4 tier

Reporting -

Security and monitoring -

Security -

A method of managing multiple databases in the cloud at the same time is must be implemented to streamlining data management and limiting management access to only those requiring access.

Monitoring -

Monitoring must be set up on every database. Contoso and partners must receive performance reports as part of contractual agreements.

Tiers 6 through 8 must have unexpected resource storage usage immediately reported to data engineers.

The Azure SQL Data Warehouse cache must be monitored when the database is being used. A dashboard monitoring key performance indicators (KPIs) indicated by traffic lights must be created and displayed based on the following metrics:

Larger image

Metric	Description
A	Low cache hit %, high cache usage %
B	Low cache hit %, low cache usage %
C	High cache hit %, high cache usage %

Existing Data Protection and Security compliances require that all certificates and keys are internally managed in an on-premises storage.

You identify the following reporting requirements:

Azure Data Warehouse must be used to gather and query data from multiple internal and external databases

Azure Data Warehouse must be optimized to use data from a cache

Reporting data aggregated for external partners must be stored in Azure Storage and be made available during regular business hours in the connecting regions

Reporting strategies must be improved to real time or near real time reporting cadence to improve competitiveness and the general supply chain

Tier 9 reporting must be moved to Event Hubs, queried, and persisted in the same Azure region as the company's main office

Tier 10 reporting data must be stored in Azure Blobs

Issues -

Team members identify the following issues:

Both internal and external client application run complex joins, equality searches and group-by clauses. Because some systems are managed externally, the queries will not be changed or optimized by Contoso

External partner organization data formats, types and schemas are controlled by the partner companies

Internal and external database development staff resources are primarily SQL developers familiar with the Transact-SQL language.

Size and amount of data has led to applications and reporting solutions not performing are required speeds

Tier 7 and 8 data access is constrained to single endpoints managed by partners for access

The company maintains several legacy client applications. Data for these applications remains isolated form other applications. This has led to hundreds of databases being provisioned on a per application basis.

Question

The Azure Data Factory instance have to meet the requirements to move the data from the On-premise SQL Servers to Azure. What should you be using for as the integration runtime?

- ○

  Azure-SSIS Integration runtime

- ○

  .Net Common Language Runtime

- ○

  Azure Integration runtime

- ○

  Self-hosted integration run time

  (Correct)

Explanation

A self-hosted integration runtime can run copy activities between a cloud data store and a data store in a private network. It also can dispatch transform activities against compute

resources in an on-premises network or an Azure virtual network. The installation of a self-hosted integration runtime needs an on-premises machine or a virtual machine inside a private network.

https://docs.microsoft.com/en-us/azure/data-factory/create-self-hosted-integration-runtime

Question 17:

Contoso relies on an extensive partner network for marketing, sales, and distribution. Contoso uses external companies that manufacture everything from the actual pharmaceutical to the packaging.

The majority of the company's data reside in Microsoft SQL Server database. Application databases fall into one of the following tiers:

Larger image

Applications	Tier	Replication	Notes
Internal Contoso	1	Yes	
Internal Contoso	2	SQL Data Sync	Data Sync between databases
Internal Partner	3	Yes	Replicate to Partner
External Contoso	4,5,6	Yes	
External Partner	7,8	No	Partner managed
Internal Distribution and Sales	9	Yes, once ingested at branches	Data ingested from Contoso branches
External Distribution and Sales	10	Yes, once ingested at Contoso main office	Data is ingested from multiple sources

The company has a reporting infrastructure that ingests data from local databases and partner services. Partners services consists of distributors, wholesales, and retailers across the world. The company performs daily, weekly, and monthly reporting.

Requirements -

Tier 3 and Tier 6 through Tier 8 application must use database density on the same server and Elastic pools in a cost-effective manner.

Applications must still have access to data from both internal and external applications keeping the data encrypted and secure at rest and in transit.

A disaster recovery strategy must be implemented for Tier 3 and Tier 6 through 8 allowing for failover in the case of server going offline.

Selected internal applications must have the data hosted in single Microsoft Azure SQL Databases.

Tier 1 internal applications on the premium P2 tier

Tier 2 internal applications on the standard S4 tier

The solution must support migrating databases that support external and internal application to Azure SQL Database. The migrated databases will be supported by

Azure Data Factory pipelines for the continued movement, migration and updating of data both in the cloud and from local core business systems and repositories.

Tier 7 and Tier 8 partner access must be restricted to the database only.

In addition to default Azure backup behavior, Tier 4 and 5 databases must be on a backup strategy that performs a transaction log backup eve hour, a differential backup of databases every day and a full back up every week.

Back up strategies must be put in place for all other standalone Azure SQL Databases using Azure SQL-provided backup storage and capabilities.

Databases -

Contoso requires their data estate to be designed and implemented in the Azure Cloud. Moving to the cloud must not inhibit access to or availability of data.

Databases:

Tier 1 Database must implement data masking using the following masking logic:

Larger image

Data type	Masking requirement
A	Mask 4 or less string data type characters
B	Mask first letter and domain
C	Mask everything except characters at the beginning and end

Tier 2 databases must sync between branches and cloud databases and in the event of conflicts must be set up for conflicts to be won by on-premises databases.

Tier 3 and Tier 6 through Tier 8 applications must use database density on the same server and Elastic pools in a cost-effective manner.

Applications must still have access to data from both internal and external applications keeping the data encrypted and secure at rest and in transit.

A disaster recovery strategy must be implemented for Tier 3 and Tier 6 through 8 allowing for failover in the case of a server going offline.

Selected internal applications must have the data hosted in single Microsoft Azure SQL Databases.

Tier 1 internal applications on the premium P2 tier

Tier 2 internal applications on the standard S4 tier

Reporting -

Security and monitoring -

Security -

A method of managing multiple databases in the cloud at the same time is must be implemented to streamlining data management and limiting management access to only those requiring access.

Monitoring -

Monitoring must be set up on every database. Contoso and partners must receive performance reports as part of contractual agreements.

Tiers 6 through 8 must have unexpected resource storage usage immediately reported to data engineers.

The Azure SQL Data Warehouse cache must be monitored when the database is being used. A dashboard monitoring key performance indicators (KPIs) indicated by traffic lights must be created and displayed based on the following metrics:

Larger image

Metric	Description
A	Low cache hit %, high cache usage %
B	Low cache hit %, low cache usage %
C	High cache hit %, high cache usage %

Existing Data Protection and Security compliances require that all certificates and keys are internally managed in an on-premises storage.

You identify the following reporting requirements:

Azure Data Warehouse must be used to gather and query data from multiple internal and external databases

Azure Data Warehouse must be optimized to use data from a cache

Reporting data aggregated for external partners must be stored in Azure Storage and be made available during regular business hours in the connecting regions

Reporting strategies must be improved to real time or near real time reporting cadence to improve competitiveness and the general supply chain

Tier 9 reporting must be moved to Event Hubs, queried, and persisted in the same Azure region as the company's main office

Tier 10 reporting data must be stored in Azure Blobs

Issues -

Team members identify the following issues:

Both internal and external client application run complex joins, equality searches and group-by clauses. Because some systems are managed externally, the queries will not be changed or optimized by Contoso

External partner organization data formats, types and schemas are controlled by the partner companies

Internal and external database development staff resources are primarily SQL developers familiar with the Transact-SQL language.

Size and amount of data has led to applications and reporting solutions not performing are required speeds

Tier 7 and 8 data access is constrained to single endpoints managed by partners for access

The company maintains several legacy client applications. Data for these applications remains isolated form other applications. This has led to hundreds of databases being provisioned on a per application basis.

Question

You are implementing the following steps to encrypt the data for the external applications:

· Use the Always Encrypted Wizard in SQL Server Management Studio

· Select the column that needs to be encrypted

· Set the encryption type to Randomized

· Configure the master key to be used from the Windows Certificate Store

· Confirm the configuration and deploy the solution

Do you think that these steps will fulfil the requirement?

- ○ Yes

- ○ No

  (Correct)

Explanation

We should not use encryption type randomized as it will impact the queries. Both internal and external client application run complex joins, equality searches and group-by clauses.

Click **Next** on the **Introduction** page to open the **Column Selection** page. On this page, you will select which columns you want to encrypt, the type of encryption, and what column encryption key (CEK) to use.

Encrypt **SSN** and **BirthDate** information for each patient. The **SSN** column will use deterministic encryption, which supports equality lookups, joins, and group by.
The **BirthDate** column will use randomized encryption, which does not support operations.

Set the **Encryption Type** for the **SSN** column to **Deterministic** and the **BirthDate** column to **Randomized**. Click **Next**.

https://docs.microsoft.com/en-us/azure/azure-sql/database/always-encrypted-certificate-store-configure

Question 18: Skipped

Contoso relies on an extensive partner network for marketing, sales, and distribution. Contoso uses external companies that manufacture everything from the actual pharmaceutical to the packaging.

The majority of the company's data reside in Microsoft SQL Server database. Application databases fall into one of the following tiers:

Larger image

Applications	Tier	Replication	Notes
Internal Contoso	1	Yes	
Internal Contoso	2	SQL Data Sync	Data Sync between databases
Internal Partner	3	Yes	Replicate to Partner
External Contoso	4,5,6	Yes	
External Partner	7,8	No	Partner managed
Internal Distribution and Sales	9	Yes, once ingested at branches	Data ingested from Contoso branches
External Distribution and Sales	10	Yes, once ingested at Contoso main office	Data is ingested from multiple sources

The company has a reporting infrastructure that ingests data from local databases and partner services. Partners services consists of distributors, wholesales, and retailers across the world. The company performs daily, weekly, and monthly reporting.

Requirements -

Tier 3 and Tier 6 through Tier 8 application must use database density on the same server and Elastic pools in a cost-effective manner.

Applications must still have access to data from both internal and external applications keeping the data encrypted and secure at rest and in transit.

A disaster recovery strategy must be implemented for Tier 3 and Tier 6 through 8 allowing for failover in the case of server going offline.

Selected internal applications must have the data hosted in single Microsoft Azure SQL Databases.

Tier 1 internal applications on the premium P2 tier

Tier 2 internal applications on the standard S4 tier

The solution must support migrating databases that support external and internal application to Azure SQL Database. The migrated databases will be supported by Azure Data Factory pipelines for the continued movement, migration and updating of data both in the cloud and from local core business systems and repositories.

Tier 7 and Tier 8 partner access must be restricted to the database only.

In addition to default Azure backup behavior, Tier 4 and 5 databases must be on a backup strategy that performs a transaction log backup eve hour, a differential backup of databases every day and a full back up every week.

Back up strategies must be put in place for all other standalone Azure SQL Databases using Azure SQL-provided backup storage and capabilities.

Databases -

Contoso requires their data estate to be designed and implemented in the Azure Cloud. Moving to the cloud must not inhibit access to or availability of data.

Databases:

Tier 1 Database must implement data masking using the following masking logic:

Larger image

Data type	Masking requirement
A	Mask 4 or less string data type characters
B	Mask first letter and domain
C	Mask everything except characters at the beginning and end

Tier 2 databases must sync between branches and cloud databases and in the event of conflicts must be set up for conflicts to be won by on-premises databases.

Tier 3 and Tier 6 through Tier 8 applications must use database density on the same server and Elastic pools in a cost-effective manner.

Applications must still have access to data from both internal and external applications keeping the data encrypted and secure at rest and in transit.

A disaster recovery strategy must be implemented for Tier 3 and Tier 6 through 8 allowing for failover in the case of a server going offline.

Selected internal applications must have the data hosted in single Microsoft Azure SQL Databases.

Tier 1 internal applications on the premium P2 tier

Tier 2 internal applications on the standard S4 tier

Reporting -

Security and monitoring -

Security -

A method of managing multiple databases in the cloud at the same time is must be implemented to streamlining data management and limiting management access to only those requiring access.

Monitoring -

Monitoring must be set up on every database. Contoso and partners must receive performance reports as part of contractual agreements.

Tiers 6 through 8 must have unexpected resource storage usage immediately reported to data engineers.

The Azure SQL Data Warehouse cache must be monitored when the database is being used. A dashboard monitoring key performance indicators (KPIs) indicated by traffic lights must be created and displayed based on the following metrics:

Larger image

Metric	Description
A	Low cache hit %, high cache usage %
B	Low cache hit %, low cache usage %
C	High cache hit %, high cache usage %

Existing Data Protection and Security compliances require that all certificates and keys are internally managed in an on-premises storage.

You identify the following reporting requirements:

Azure Data Warehouse must be used to gather and query data from multiple internal and external databases

Azure Data Warehouse must be optimized to use data from a cache

Reporting data aggregated for external partners must be stored in Azure Storage and be made available during regular business hours in the connecting regions

Reporting strategies must be improved to real time or near real time reporting cadence to improve competitiveness and the general supply chain

Tier 9 reporting must be moved to Event Hubs, queried, and persisted in the same Azure region as the company's main office

Tier 10 reporting data must be stored in Azure Blobs

Issues -

Team members identify the following issues:

Both internal and external client application run complex joins, equality searches and group-by clauses. Because some systems are managed externally, the queries will not be changed or optimized by Contoso

External partner organization data formats, types and schemas are controlled by the partner companies

Internal and external database development staff resources are primarily SQL developers familiar with the Transact-SQL language.

Size and amount of data has led to applications and reporting solutions not performing are required speeds

Tier 7 and 8 data access is constrained to single endpoints managed by partners for access

The company maintains several legacy client applications. Data for these applications remains isolated form other applications. This has led to hundreds of databases being provisioned on a per application basis.

Question

You are implementing the following steps to encrypt the data for the external applications:

· Use the Always Encrypted Wizard in SQL Server Management Studio

· Select the column that needs to be encrypted

· Set the encryption type to Deterministic

· Configure the master key to be used from Azure Key vault

· Confirm the configuration and deploy the solution

Do you think that these steps will fulfil the requirement?

- ○

  Yes

- ○

No

(Correct)

Explanation

While the steps to encrypt are correct, as per the case study, all keys and certificates need to be managed in on-premise data stores

https://docs.microsoft.com/en-us/azure/sql-database/sql-database-always-encrypted

Question 19: Skipped

Contoso relies on an extensive partner network for marketing, sales, and distribution. Contoso uses external companies that manufacture everything from the actual pharmaceutical to the packaging.

The majority of the company's data reside in Microsoft SQL Server database. Application databases fall into one of the following tiers:

Larger image

Applications	Tier	Replication	Notes
Internal Contoso	1	Yes	
Internal Contoso	2	SQL Data Sync	Data Sync between databases
Internal Partner	3	Yes	Replicate to Partner
External Contoso	4,5,6	Yes	
External Partner	7,8	No	Partner managed
Internal Distribution and Sales	9	Yes, once ingested at branches	Data ingested from Contoso branches
External Distribution and Sales	10	Yes, once ingested at Contoso main office	Data is ingested from multiple sources

The company has a reporting infrastructure that ingests data from local databases and partner services. Partners services consists of distributors, wholesales, and retailers across the world. The company performs daily, weekly, and monthly reporting.

Requirements -

Tier 3 and Tier 6 through Tier 8 application must use database density on the same server and Elastic pools in a cost-effective manner.

Applications must still have access to data from both internal and external applications keeping the data encrypted and secure at rest and in transit.

A disaster recovery strategy must be implemented for Tier 3 and Tier 6 through 8 allowing for failover in the case of server going offline.

Selected internal applications must have the data hosted in single Microsoft Azure SQL Databases.

Tier 1 internal applications on the premium P2 tier

Tier 2 internal applications on the standard S4 tier

The solution must support migrating databases that support external and internal application to Azure SQL Database. The migrated databases will be supported by Azure Data Factory pipelines for the continued movement, migration and updating of data both in the cloud and from local core business systems and repositories.

Tier 7 and Tier 8 partner access must be restricted to the database only.

In addition to default Azure backup behavior, Tier 4 and 5 databases must be on a backup strategy that performs a transaction log backup eve hour, a differential backup of databases every day and a full back up every week.

Back up strategies must be put in place for all other standalone Azure SQL Databases using Azure SQL-provided backup storage and capabilities.

Databases -

Contoso requires their data estate to be designed and implemented in the Azure Cloud. Moving to the cloud must not inhibit access to or availability of data.

Databases:

Tier 1 Database must implement data masking using the following masking logic:

Larger image

Data type	Masking requirement
A	Mask 4 or less string data type characters
B	Mask first letter and domain
C	Mask everything except characters at the beginning and end

Tier 2 databases must sync between branches and cloud databases and in the event of conflicts must be set up for conflicts to be won by on-premises databases.

Tier 3 and Tier 6 through Tier 8 applications must use database density on the same server and Elastic pools in a cost-effective manner.

Applications must still have access to data from both internal and external applications keeping the data encrypted and secure at rest and in transit.

A disaster recovery strategy must be implemented for Tier 3 and Tier 6 through 8 allowing for failover in the case of a server going offline.

Selected internal applications must have the data hosted in single Microsoft Azure SQL Databases.

Tier 1 internal applications on the premium P2 tier

Tier 2 internal applications on the standard S4 tier

Reporting -

Security and monitoring -

Security -

A method of managing multiple databases in the cloud at the same time is must be implemented to streamlining data management and limiting management access to only those requiring access.

Monitoring -

Monitoring must be set up on every database. Contoso and partners must receive performance reports as part of contractual agreements.

Tiers 6 through 8 must have unexpected resource storage usage immediately reported to data engineers.

The Azure SQL Data Warehouse cache must be monitored when the database is being used. A dashboard monitoring key performance indicators (KPIs) indicated by traffic lights must be created and displayed based on the following metrics:

Larger image

Metric	Description
A	Low cache hit %, high cache usage %
B	Low cache hit %, low cache usage %
C	High cache hit %, high cache usage %

Existing Data Protection and Security compliances require that all certificates and keys are internally managed in an on-premises storage.

You identify the following reporting requirements:

Azure Data Warehouse must be used to gather and query data from multiple internal and external databases

Azure Data Warehouse must be optimized to use data from a cache

Reporting data aggregated for external partners must be stored in Azure Storage and be made available during regular business hours in the connecting regions

Reporting strategies must be improved to real time or near real time reporting cadence to improve competitiveness and the general supply chain

Tier 9 reporting must be moved to Event Hubs, queried, and persisted in the same Azure region as the company's main office

Tier 10 reporting data must be stored in Azure Blobs

Issues -

Team members identify the following issues:

Both internal and external client application run complex joins, equality searches and group-by clauses. Because some systems are managed externally, the queries will not be changed or optimized by Contoso

External partner organization data formats, types and schemas are controlled by the partner companies

Internal and external database development staff resources are primarily SQL developers familiar with the Transact-SQL language.

Size and amount of data has led to applications and reporting solutions not performing are required speeds

Tier 7 and 8 data access is constrained to single endpoints managed by partners for access

The company maintains several legacy client applications. Data for these applications remains isolated form other applications. This has led to hundreds of databases being provisioned on a per application basis.

Question

What should you be using as the masking function for Data type A?

- ○

  Custom

- ○

  Default

  (Correct)

- ○

Email

- ○

Random number

Explanation

Default - Full masking according to the data types of the designated fields

• Use XXXX or fewer Xs if the size of the field is less than 4 characters for string data types (nchar, ntext, nvarchar).

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

Question 20: Skipped
Contoso relies on an extensive partner network for marketing, sales, and distribution. Contoso uses external companies that manufacture everything from the actual pharmaceutical to the packaging.

The majority of the company's data reside in Microsoft SQL Server database. Application databases fall into one of the following tiers:

Larger image

Applications	Tier	Replication	Notes
Internal Contoso	1	Yes	
Internal Contoso	2	SQL Data Sync	Data Sync between databases
Internal Partner	3	Yes	Replicate to Partner
External Contoso	4,5,6	Yes	
External Partner	7,8	No	Partner managed
Internal Distribution and Sales	9	Yes, once ingested at branches	Data ingested from Contoso branches
External Distribution and Sales	10	Yes, once ingested at Contoso main office	Data is ingested from multiple sources

The company has a reporting infrastructure that ingests data from local databases and partner services. Partners services consists of distributors, wholesales, and retailers across the world. The company performs daily, weekly, and monthly reporting.

Requirements -

Tier 3 and Tier 6 through Tier 8 application must use database density on the same server and Elastic pools in a cost-effective manner.

Applications must still have access to data from both internal and external applications keeping the data encrypted and secure at rest and in transit.

A disaster recovery strategy must be implemented for Tier 3 and Tier 6 through 8 allowing for failover in the case of server going offline.

Selected internal applications must have the data hosted in single Microsoft Azure SQL Databases.

Tier 1 internal applications on the premium P2 tier

Tier 2 internal applications on the standard S4 tier

The solution must support migrating databases that support external and internal application to Azure SQL Database. The migrated databases will be supported by Azure Data Factory pipelines for the continued movement, migration and updating of data both in the cloud and from local core business systems and repositories.

Tier 7 and Tier 8 partner access must be restricted to the database only.

In addition to default Azure backup behavior, Tier 4 and 5 databases must be on a backup strategy that performs a transaction log backup eve hour, a differential backup of databases every day and a full back up every week.

Back up strategies must be put in place for all other standalone Azure SQL Databases using Azure SQL-provided backup storage and capabilities.

Databases -

Contoso requires their data estate to be designed and implemented in the Azure Cloud. Moving to the cloud must not inhibit access to or availability of data.

Databases:

Tier 1 Database must implement data masking using the following masking logic:

Larger image

Data type	Masking requirement
A	Mask 4 or less string data type characters
B	Mask first letter and domain
C	Mask everything except characters at the beginning and end

Tier 2 databases must sync between branches and cloud databases and in the event of conflicts must be set up for conflicts to be won by on-premises databases.

Tier 3 and Tier 6 through Tier 8 applications must use database density on the same server and Elastic pools in a cost-effective manner.

Applications must still have access to data from both internal and external applications keeping the data encrypted and secure at rest and in transit.

A disaster recovery strategy must be implemented for Tier 3 and Tier 6 through 8 allowing for failover in the case of a server going offline.

Selected internal applications must have the data hosted in single Microsoft Azure SQL Databases.

Tier 1 internal applications on the premium P2 tier

Tier 2 internal applications on the standard S4 tier

Reporting -

Security and monitoring -

Security -

A method of managing multiple databases in the cloud at the same time is must be implemented to streamlining data management and limiting management access to only those requiring access.

Monitoring -

Monitoring must be set up on every database. Contoso and partners must receive performance reports as part of contractual agreements.

Tiers 6 through 8 must have unexpected resource storage usage immediately reported to data engineers.

The Azure SQL Data Warehouse cache must be monitored when the database is being used. A dashboard monitoring key performance indicators (KPIs) indicated by traffic lights must be created and displayed based on the following metrics:

Larger image

Metric	Description
A	Low cache hit %, high cache usage %
B	Low cache hit %, low cache usage %
C	High cache hit %, high cache usage %

Existing Data Protection and Security compliances require that all certificates and keys are internally managed in an on-premises storage.

You identify the following reporting requirements:

Azure Data Warehouse must be used to gather and query data from multiple internal and external databases

Azure Data Warehouse must be optimized to use data from a cache

Reporting data aggregated for external partners must be stored in Azure Storage and be made available during regular business hours in the connecting regions

Reporting strategies must be improved to real time or near real time reporting cadence to improve competitiveness and the general supply chain

Tier 9 reporting must be moved to Event Hubs, queried, and persisted in the same Azure region as the company's main office

Tier 10 reporting data must be stored in Azure Blobs

Issues -

Team members identify the following issues:

Both internal and external client application run complex joins, equality searches and group-by clauses. Because some systems are managed externally, the queries will not be changed or optimized by Contoso

External partner organization data formats, types and schemas are controlled by the partner companies

Internal and external database development staff resources are primarily SQL developers familiar with the Transact-SQL language.

Size and amount of data has led to applications and reporting solutions not performing are required speeds

Tier 7 and 8 data access is constrained to single endpoints managed by partners for access

The company maintains several legacy client applications. Data for these applications remains isolated form other applications. This has led to hundreds of databases being provisioned on a per application basis.

Question

What should you be using as the masking function for Data type B?

- ○

  Custom Text

- ○

  Default

- ○

  Email

- ○

Random Number

Explanation

Email - Masking method, which exposes the first letter and replaces the domain with XXX.com using a constant string prefix in the form of an email address.

aXX@XXXX.com

Question 21: Skipped
Contoso relies on an extensive partner network for marketing, sales, and distribution. Contoso uses external companies that manufacture everything from the actual pharmaceutical to the packaging.

The majority of the company's data reside in Microsoft SQL Server database. Application databases fall into one of the following tiers:

Larger image

Applications	Tier	Replication	Notes
Internal Contoso	1	Yes	
Internal Contoso	2	SQL Data Sync	Data Sync between databases
Internal Partner	3	Yes	Replicate to Partner
External Contoso	4,5,6	Yes	
External Partner	7,8	No	Partner managed
Internal Distribution and Sales	9	Yes, once ingested at branches	Data ingested from Contoso branches
External Distribution and Sales	10	Yes, once ingested at Contoso main office	Data is ingested from multiple sources

The company has a reporting infrastructure that ingests data from local databases and partner services. Partners services consists of distributors, wholesales, and retailers across the world. The company performs daily, weekly, and monthly reporting.

Requirements -

Tier 3 and Tier 6 through Tier 8 application must use database density on the same server and Elastic pools in a cost-effective manner.

Applications must still have access to data from both internal and external applications keeping the data encrypted and secure at rest and in transit.

A disaster recovery strategy must be implemented for Tier 3 and Tier 6 through 8 allowing for failover in the case of server going offline.

Selected internal applications must have the data hosted in single Microsoft Azure SQL Databases.

Tier 1 internal applications on the premium P2 tier

Tier 2 internal applications on the standard S4 tier

The solution must support migrating databases that support external and internal application to Azure SQL Database. The migrated databases will be supported by Azure Data Factory pipelines for the continued movement, migration and updating of data both in the cloud and from local core business systems and repositories.

Tier 7 and Tier 8 partner access must be restricted to the database only.

In addition to default Azure backup behavior, Tier 4 and 5 databases must be on a backup strategy that performs a transaction log backup eve hour, a differential backup of databases every day and a full back up every week.

Back up strategies must be put in place for all other standalone Azure SQL Databases using Azure SQL-provided backup storage and capabilities.

Databases -

Contoso requires their data estate to be designed and implemented in the Azure Cloud. Moving to the cloud must not inhibit access to or availability of data.

Databases:

Tier 1 Database must implement data masking using the following masking logic:

Larger image

Data type	Masking requirement
A	Mask 4 or less string data type characters
B	Mask first letter and domain
C	Mask everything except characters at the beginning and end

Tier 2 databases must sync between branches and cloud databases and in the event of conflicts must be set up for conflicts to be won by on-premises databases.

Tier 3 and Tier 6 through Tier 8 applications must use database density on the same server and Elastic pools in a cost-effective manner.

Applications must still have access to data from both internal and external applications keeping the data encrypted and secure at rest and in transit.

A disaster recovery strategy must be implemented for Tier 3 and Tier 6 through 8 allowing for failover in the case of a server going offline.

Selected internal applications must have the data hosted in single Microsoft Azure SQL Databases.

Tier 1 internal applications on the premium P2 tier

Tier 2 internal applications on the standard S4 tier

Reporting -

Security and monitoring -

Security -

A method of managing multiple databases in the cloud at the same time is must be implemented to streamlining data management and limiting management access to only those requiring access.

Monitoring -

Monitoring must be set up on every database. Contoso and partners must receive performance reports as part of contractual agreements.

Tiers 6 through 8 must have unexpected resource storage usage immediately reported to data engineers.

The Azure SQL Data Warehouse cache must be monitored when the database is being used. A dashboard monitoring key performance indicators (KPIs) indicated by traffic lights must be created and displayed based on the following metrics:

Larger image

Metric	Description
A	Low cache hit %, high cache usage %
B	Low cache hit %, low cache usage %
C	High cache hit %, high cache usage %

Existing Data Protection and Security compliances require that all certificates and keys are internally managed in an on-premises storage.

You identify the following reporting requirements:

Azure Data Warehouse must be used to gather and query data from multiple internal and external databases

Azure Data Warehouse must be optimized to use data from a cache

Reporting data aggregated for external partners must be stored in Azure Storage and be made available during regular business hours in the connecting regions

Reporting strategies must be improved to real time or near real time reporting cadence to improve competitiveness and the general supply chain

Tier 9 reporting must be moved to Event Hubs, queried, and persisted in the same Azure region as the company's main office

Tier 10 reporting data must be stored in Azure Blobs

Issues -

Team members identify the following issues:

Both internal and external client application run complex joins, equality searches and group-by clauses. Because some systems are managed externally, the queries will not be changed or optimized by Contoso

External partner organization data formats, types and schemas are controlled by the partner companies

Internal and external database development staff resources are primarily SQL developers familiar with the Transact-SQL language.

Size and amount of data has led to applications and reporting solutions not performing are required speeds

Tier 7 and 8 data access is constrained to single endpoints managed by partners for access

The company maintains several legacy client applications. Data for these applications remains isolated form other applications. This has led to hundreds of databases being provisioned on a per application basis.

Question

What should you be using as the masking function for Data type C?

- ○
  Custom Text
  (Correct)

- ○
  Default

- ○

Email

- ○

Random number

Explanation

Custom text - Masking method, which exposes the first and last characters and adds a custom padding string in the middle. If the original string is shorter than the exposed prefix and suffix, only the padding string is used.
prefix[padding]suffix

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

Question 22: Skipped
Contoso relies on an extensive partner network for marketing, sales, and distribution. Contoso uses external companies that manufacture everything from the actual pharmaceutical to the packaging.

The majority of the company's data reside in Microsoft SQL Server database. Application databases fall into one of the following tiers:

Larger image

Applications	Tier	Replication	Notes
Internal Contoso	1	Yes	
Internal Contoso	2	SQL Data Sync	Data Sync between databases
Internal Partner	3	Yes	Replicate to Partner
External Contoso	4,5,6	Yes	
External Partner	7,8	No	Partner managed
Internal Distribution and Sales	9	Yes, once ingested at branches	Data ingested from Contoso branches
External Distribution and Sales	10	Yes, once ingested at Contoso main office	Data is ingested from multiple sources

The company has a reporting infrastructure that ingests data from local databases and partner services. Partners services consists of distributors, wholesales, and retailers across the world. The company performs daily, weekly, and monthly reporting.

Requirements -

Tier 3 and Tier 6 through Tier 8 application must use database density on the same server and Elastic pools in a cost-effective manner.

Applications must still have access to data from both internal and external applications keeping the data encrypted and secure at rest and in transit.

A disaster recovery strategy must be implemented for Tier 3 and Tier 6 through 8 allowing for failover in the case of server going offline.

Selected internal applications must have the data hosted in single Microsoft Azure SQL Databases.

Tier 1 internal applications on the premium P2 tier

Tier 2 internal applications on the standard S4 tier

The solution must support migrating databases that support external and internal application to Azure SQL Database. The migrated databases will be supported by Azure Data Factory pipelines for the continued movement, migration and updating of data both in the cloud and from local core business systems and repositories.

Tier 7 and Tier 8 partner access must be restricted to the database only.

In addition to default Azure backup behavior, Tier 4 and 5 databases must be on a backup strategy that performs a transaction log backup eve hour, a differential backup of databases every day and a full back up every week.

Back up strategies must be put in place for all other standalone Azure SQL Databases using Azure SQL-provided backup storage and capabilities.

Databases -

Contoso requires their data estate to be designed and implemented in the Azure Cloud. Moving to the cloud must not inhibit access to or availability of data.

Databases:

Tier 1 Database must implement data masking using the following masking logic:

Larger image

Data type	Masking requirement
A	Mask 4 or less string data type characters
B	Mask first letter and domain
C	Mask everything except characters at the beginning and end

Tier 2 databases must sync between branches and cloud databases and in the event of conflicts must be set up for conflicts to be won by on-premises databases.

Tier 3 and Tier 6 through Tier 8 applications must use database density on the same server and Elastic pools in a cost-effective manner.

Applications must still have access to data from both internal and external applications keeping the data encrypted and secure at rest and in transit.

A disaster recovery strategy must be implemented for Tier 3 and Tier 6 through 8 allowing for failover in the case of a server going offline.

Selected internal applications must have the data hosted in single Microsoft Azure SQL Databases.

Tier 1 internal applications on the premium P2 tier

Tier 2 internal applications on the standard S4 tier

Reporting -

Security and monitoring -

Security -

A method of managing multiple databases in the cloud at the same time is must be implemented to streamlining data management and limiting management access to only those requiring access.

Monitoring -

Monitoring must be set up on every database. Contoso and partners must receive performance reports as part of contractual agreements.

Tiers 6 through 8 must have unexpected resource storage usage immediately reported to data engineers.

The Azure SQL Data Warehouse cache must be monitored when the database is being used. A dashboard monitoring key performance indicators (KPIs) indicated by traffic lights must be created and displayed based on the following metrics:

Larger image

Metric	Description
A	Low cache hit %, high cache usage %
B	Low cache hit %, low cache usage %
C	High cache hit %, high cache usage %

Existing Data Protection and Security compliances require that all certificates and keys are internally managed in an on-premises storage.

You identify the following reporting requirements:

Azure Data Warehouse must be used to gather and query data from multiple internal and external databases

Azure Data Warehouse must be optimized to use data from a cache

Reporting data aggregated for external partners must be stored in Azure Storage and be made available during regular business hours in the connecting regions

Reporting strategies must be improved to real time or near real time reporting cadence to improve competitiveness and the general supply chain

Tier 9 reporting must be moved to Event Hubs, queried, and persisted in the same Azure region as the company's main office

Tier 10 reporting data must be stored in Azure Blobs

Issues -

Team members identify the following issues:

Both internal and external client application run complex joins, equality searches and group-by clauses. Because some systems are managed externally, the queries will not be changed or optimized by Contoso

External partner organization data formats, types and schemas are controlled by the partner companies

Internal and external database development staff resources are primarily SQL developers familiar with the Transact-SQL language.

Size and amount of data has led to applications and reporting solutions not performing are required speeds

Tier 7 and 8 data access is constrained to single endpoints managed by partners for access

The company maintains several legacy client applications. Data for these applications remains isolated form other applications. This has led to hundreds of databases being provisioned on a per application basis.

Question

You have been given the task to implement the following:

· The Application access for Tier 7 and 8 must be restricted to the database only

What steps should you be implementing for this? Select 3 options to answer.

- ☐

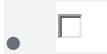 Use Azure PowerShell to create a database firewall rule

- ☐

  Configure the setting of "Allow Azure Services to Access Server" to Disabled

(Correct)

- ☐

  Configure the setting of "Allow Azure Services to Access Server" to Enabled

- ☐

  Create a database firewall rule from the Azure portal

- ☐

  Create a server firewall rule from the Azure portal

  (Correct)

- ☐

  Use Transact-SQL to create a database firewall rule

  (Correct)

Explanation

If you have the **Allow Azure Services and resources to access this server** setting enabled, this counts as a single firewall rule for the server.

You can only create and manage database-level IP firewall rules for master and user databases by using Transact-SQL statements and only after you configure the first server-level firewall.

You can configure server-level IP firewall rules by using the Azure portal, PowerShell, or Transact-SQL statements

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

Question 23: Skipped

Contoso relies on an extensive partner network for marketing, sales, and distribution. Contoso uses external companies that manufacture everything from the actual pharmaceutical to the packaging.

The majority of the company's data reside in Microsoft SQL Server database. Application databases fall into one of the following tiers:

Larger image

Applications	Tier	Replication	Notes
Internal Contoso	1	Yes	
Internal Contoso	2	SQL Data Sync	Data Sync between databases
Internal Partner	3	Yes	Replicate to Partner
External Contoso	4,5,6	Yes	
External Partner	7,8	No	Partner managed
Internal Distribution and Sales	9	Yes, once ingested at branches	Data ingested from Contoso branches
External Distribution and Sales	10	Yes, once ingested at Contoso main office	Data is ingested from multiple sources

The company has a reporting infrastructure that ingests data from local databases and partner services. Partners services consists of distributors, wholesales, and retailers across the world. The company performs daily, weekly, and monthly reporting.

Requirements -

Tier 3 and Tier 6 through Tier 8 application must use database density on the same server and Elastic pools in a cost-effective manner.

Applications must still have access to data from both internal and external applications keeping the data encrypted and secure at rest and in transit.

A disaster recovery strategy must be implemented for Tier 3 and Tier 6 through 8 allowing for failover in the case of server going offline.

Selected internal applications must have the data hosted in single Microsoft Azure SQL Databases.

Tier 1 internal applications on the premium P2 tier

Tier 2 internal applications on the standard S4 tier

The solution must support migrating databases that support external and internal application to Azure SQL Database. The migrated databases will be supported by Azure Data Factory pipelines for the continued movement, migration and updating of data both in the cloud and from local core business systems and repositories.

Tier 7 and Tier 8 partner access must be restricted to the database only.

In addition to default Azure backup behavior, Tier 4 and 5 databases must be on a backup strategy that performs a transaction log backup eve hour, a differential backup of databases every day and a full back up every week.

Back up strategies must be put in place for all other standalone Azure SQL Databases using Azure SQL-provided backup storage and capabilities.

Databases -

Contoso requires their data estate to be designed and implemented in the Azure Cloud. Moving to the cloud must not inhibit access to or availability of data.

Databases:

Tier 1 Database must implement data masking using the following masking logic:

Larger image

Data type	Masking requirement
A	Mask 4 or less string data type characters
B	Mask first letter and domain
C	Mask everything except characters at the beginning and end

Tier 2 databases must sync between branches and cloud databases and in the event of conflicts must be set up for conflicts to be won by on-premises databases.

Tier 3 and Tier 6 through Tier 8 applications must use database density on the same server and Elastic pools in a cost-effective manner.

Applications must still have access to data from both internal and external applications keeping the data encrypted and secure at rest and in transit.

A disaster recovery strategy must be implemented for Tier 3 and Tier 6 through 8 allowing for failover in the case of a server going offline.

Selected internal applications must have the data hosted in single Microsoft Azure SQL Databases.

Tier 1 internal applications on the premium P2 tier

Tier 2 internal applications on the standard S4 tier

Reporting -

Security and monitoring -

Security -

A method of managing multiple databases in the cloud at the same time is must be implemented to streamlining data management and limiting management access to only those requiring access.

Monitoring -

Monitoring must be set up on every database. Contoso and partners must receive performance reports as part of contractual agreements.

Tiers 6 through 8 must have unexpected resource storage usage immediately reported to data engineers.

The Azure SQL Data Warehouse cache must be monitored when the database is being used. A dashboard monitoring key performance indicators (KPIs) indicated by traffic lights must be created and displayed based on the following metrics:

Larger image

Metric	Description
A	Low cache hit %, high cache usage %
B	Low cache hit %, low cache usage %
C	High cache hit %, high cache usage %

Existing Data Protection and Security compliances require that all certificates and keys are internally managed in an on-premises storage.

You identify the following reporting requirements:

Azure Data Warehouse must be used to gather and query data from multiple internal and external databases

Azure Data Warehouse must be optimized to use data from a cache

Reporting data aggregated for external partners must be stored in Azure Storage and be made available during regular business hours in the connecting regions

Reporting strategies must be improved to real time or near real time reporting cadence to improve competitiveness and the general supply chain

Tier 9 reporting must be moved to Event Hubs, queried, and persisted in the same Azure region as the company's main office

Tier 10 reporting data must be stored in Azure Blobs

Issues -

Team members identify the following issues:

Both internal and external client application run complex joins, equality searches and group-by clauses. Because some systems are managed externally, the queries will not be changed or optimized by Contoso

External partner organization data formats, types and schemas are controlled by the partner companies

Internal and external database development staff resources are primarily SQL developers familiar with the Transact-SQL language.

Size and amount of data has led to applications and reporting solutions not performing are required speeds

Tier 7 and 8 data access is constrained to single endpoints managed by partners for access

The company maintains several legacy client applications. Data for these applications remains isolated form other applications. This has led to hundreds of databases being provisioned on a per application basis.

Question

You are given the task to fulfil the following requirement of this case study:

"Applications with Tiers 6 through 8 must ensure that unexpected resource storage usage is immediately reported to IT data engineers."

What should you be doing for to implement for this requirement?

- ○

    An alert rule that would be used to monitor CPU percentage for the database and then alert the IT Engineers

- ○

    An alert rule that would be used to monitor CPU percentage for the elastic pool and then alert the IT Engineers

- ○

    An alert rule that would be used to monitor storage percentage for the database and then alert the IT Engineers

- ○

    An alert rule that would be used to monitor storage percentage for the elastic pool and then alert the IT Engineers

    (Correct)

Explanation

Since the requirement asks for monitoring the storage and the databases are going to be part of an elastic pool, this option is correct

Question 24: Skipped

Contoso relies on an extensive partner network for marketing, sales, and distribution. Contoso uses external companies that manufacture everything from the actual pharmaceutical to the packaging.

The majority of the company's data reside in Microsoft SQL Server database. Application databases fall into one of the following tiers:

Larger image

Applications	Tier	Replication	Notes
Internal Contoso	1	Yes	
Internal Contoso	2	SQL Data Sync	Data Sync between databases
Internal Partner	3	Yes	Replicate to Partner
External Contoso	4,5,6	Yes	
External Partner	7,8	No	Partner managed
Internal Distribution and Sales	9	Yes, once ingested at branches	Data ingested from Contoso branches
External Distribution and Sales	10	Yes, once ingested at Contoso main office	Data is ingested from multiple sources

The company has a reporting infrastructure that ingests data from local databases and partner services. Partners services consists of distributors, wholesales, and retailers across the world. The company performs daily, weekly, and monthly reporting.

Requirements -

Tier 3 and Tier 6 through Tier 8 application must use database density on the same server and Elastic pools in a cost-effective manner.

Applications must still have access to data from both internal and external applications keeping the data encrypted and secure at rest and in transit.

A disaster recovery strategy must be implemented for Tier 3 and Tier 6 through 8 allowing for failover in the case of server going offline.

Selected internal applications must have the data hosted in single Microsoft Azure SQL Databases.

Tier 1 internal applications on the premium P2 tier

Tier 2 internal applications on the standard S4 tier

The solution must support migrating databases that support external and internal application to Azure SQL Database. The migrated databases will be supported by Azure Data Factory pipelines for the continued movement, migration and updating of data both in the cloud and from local core business systems and repositories.

Tier 7 and Tier 8 partner access must be restricted to the database only.

In addition to default Azure backup behavior, Tier 4 and 5 databases must be on a backup strategy that performs a transaction log backup eve hour, a differential backup of databases every day and a full back up every week.

Back up strategies must be put in place for all other standalone Azure SQL Databases using Azure SQL-provided backup storage and capabilities.

Databases -

Contoso requires their data estate to be designed and implemented in the Azure Cloud. Moving to the cloud must not inhibit access to or availability of data.

Databases:

Tier 1 Database must implement data masking using the following masking logic:

Larger image

Data type	Masking requirement
A	Mask 4 or less string data type characters
B	Mask first letter and domain
C	Mask everything except characters at the beginning and end

Tier 2 databases must sync between branches and cloud databases and in the event of conflicts must be set up for conflicts to be won by on-premises databases.

Tier 3 and Tier 6 through Tier 8 applications must use database density on the same server and Elastic pools in a cost-effective manner.

Applications must still have access to data from both internal and external applications keeping the data encrypted and secure at rest and in transit.

A disaster recovery strategy must be implemented for Tier 3 and Tier 6 through 8 allowing for failover in the case of a server going offline.

Selected internal applications must have the data hosted in single Microsoft Azure SQL Databases.

Tier 1 internal applications on the premium P2 tier

Tier 2 internal applications on the standard S4 tier

Reporting -

Security and monitoring -

Security -

A method of managing multiple databases in the cloud at the same time is must be implemented to streamlining data management and limiting management access to only those requiring access.

Monitoring -

Monitoring must be set up on every database. Contoso and partners must receive performance reports as part of contractual agreements.

Tiers 6 through 8 must have unexpected resource storage usage immediately reported to data engineers.

The Azure SQL Data Warehouse cache must be monitored when the database is being used. A dashboard monitoring key performance indicators (KPIs) indicated by traffic lights must be created and displayed based on the following metrics:

Larger image

Metric	Description
A	Low cache hit %, high cache usage %
B	Low cache hit %, low cache usage %
C	High cache hit %, high cache usage %

Existing Data Protection and Security compliances require that all certificates and keys are internally managed in an on-premises storage.

You identify the following reporting requirements:

Azure Data Warehouse must be used to gather and query data from multiple internal and external databases

Azure Data Warehouse must be optimized to use data from a cache

Reporting data aggregated for external partners must be stored in Azure Storage and be made available during regular business hours in the connecting regions

Reporting strategies must be improved to real time or near real time reporting cadence to improve competitiveness and the general supply chain

Tier 9 reporting must be moved to Event Hubs, queried, and persisted in the same Azure region as the company's main office

Tier 10 reporting data must be stored in Azure Blobs

Issues -

Team members identify the following issues:

Both internal and external client application run complex joins, equality searches and group-by clauses. Because some systems are managed externally, the queries will not be changed or optimized by Contoso

External partner organization data formats, types and schemas are controlled by the partner companies

Internal and external database development staff resources are primarily SQL developers familiar with the Transact-SQL language.

Size and amount of data has led to applications and reporting solutions not performing are required speeds

Tier 7 and 8 data access is constrained to single endpoints managed by partners for access

The company maintains several legacy client applications. Data for these applications remains isolated form other applications. This has led to hundreds of databases being provisioned on a per application basis.

Question

You are given the task to monitor the data warehousing solution for which you have to implement logging. What will you log from the following?

- ○

  RequestSteps

- ○

  DmsWorkers

- ○

  ExecRequests

- ○

  SQLRequests

    (Correct)

Explanation

Since the SQL requests would affect the cache, these requests need to be monitored

Question 25: Skipped

Contoso relies on an extensive partner network for marketing, sales, and distribution. Contoso uses external companies that manufacture everything from the actual pharmaceutical to the packaging.

The majority of the company's data reside in Microsoft SQL Server database. Application databases fall into one of the following tiers:

Larger image

Applications	Tier	Replication	Notes
Internal Contoso	1	Yes	
Internal Contoso	2	SQL Data Sync	Data Sync between databases
Internal Partner	3	Yes	Replicate to Partner
External Contoso	4,5,6	Yes	
External Partner	7,8	No	Partner managed
Internal Distribution and Sales	9	Yes, once ingested at branches	Data ingested from Contoso branches
External Distribution and Sales	10	Yes, once ingested at Contoso main office	Data is ingested from multiple sources

The company has a reporting infrastructure that ingests data from local databases and partner services. Partners services consists of distributors, wholesales, and retailers across the world. The company performs daily, weekly, and monthly reporting.

Requirements -

Tier 3 and Tier 6 through Tier 8 application must use database density on the same server and Elastic pools in a cost-effective manner.

Applications must still have access to data from both internal and external applications keeping the data encrypted and secure at rest and in transit.

A disaster recovery strategy must be implemented for Tier 3 and Tier 6 through 8 allowing for failover in the case of server going offline.

Selected internal applications must have the data hosted in single Microsoft Azure SQL Databases.

Tier 1 internal applications on the premium P2 tier

Tier 2 internal applications on the standard S4 tier

The solution must support migrating databases that support external and internal application to Azure SQL Database. The migrated databases will be supported by Azure Data Factory pipelines for the continued movement, migration and updating of data both in the cloud and from local core business systems and repositories.

Tier 7 and Tier 8 partner access must be restricted to the database only.

In addition to default Azure backup behavior, Tier 4 and 5 databases must be on a backup strategy that performs a transaction log backup eve hour, a differential backup of databases every day and a full back up every week.

Back up strategies must be put in place for all other standalone Azure SQL Databases using Azure SQL-provided backup storage and capabilities.

Databases -

Contoso requires their data estate to be designed and implemented in the Azure Cloud. Moving to the cloud must not inhibit access to or availability of data.

Databases:

Tier 1 Database must implement data masking using the following masking logic:

Larger image

Data type	Masking requirement
A	Mask 4 or less string data type characters
B	Mask first letter and domain
C	Mask everything except characters at the beginning and end

Tier 2 databases must sync between branches and cloud databases and in the event of conflicts must be set up for conflicts to be won by on-premises databases.

Tier 3 and Tier 6 through Tier 8 applications must use database density on the same server and Elastic pools in a cost-effective manner.

Applications must still have access to data from both internal and external applications keeping the data encrypted and secure at rest and in transit.

A disaster recovery strategy must be implemented for Tier 3 and Tier 6 through 8 allowing for failover in the case of a server going offline.

Selected internal applications must have the data hosted in single Microsoft Azure SQL Databases.

Tier 1 internal applications on the premium P2 tier

Tier 2 internal applications on the standard S4 tier

Reporting -

Security and monitoring -

Security -

A method of managing multiple databases in the cloud at the same time is must be implemented to streamlining data management and limiting management access to only those requiring access.

Monitoring -

Monitoring must be set up on every database. Contoso and partners must receive performance reports as part of contractual agreements.

Tiers 6 through 8 must have unexpected resource storage usage immediately reported to data engineers.

The Azure SQL Data Warehouse cache must be monitored when the database is being used. A dashboard monitoring key performance indicators (KPIs) indicated by traffic lights must be created and displayed based on the following metrics:

Larger image

Metric	Description
A	Low cache hit %, high cache usage %
B	Low cache hit %, low cache usage %
C	High cache hit %, high cache usage %

Existing Data Protection and Security compliances require that all certificates and keys are internally managed in an on-premises storage.

You identify the following reporting requirements:

Azure Data Warehouse must be used to gather and query data from multiple internal and external databases

Azure Data Warehouse must be optimized to use data from a cache

Reporting data aggregated for external partners must be stored in Azure Storage and be made available during regular business hours in the connecting regions

Reporting strategies must be improved to real time or near real time reporting cadence to improve competitiveness and the general supply chain

Tier 9 reporting must be moved to Event Hubs, queried, and persisted in the same Azure region as the company's main office

Tier 10 reporting data must be stored in Azure Blobs

Issues -

Team members identify the following issues:

Both internal and external client application run complex joins, equality searches and group-by clauses. Because some systems are managed externally, the queries will not be changed or optimized by Contoso

External partner organization data formats, types and schemas are controlled by the partner companies

Internal and external database development staff resources are primarily SQL developers familiar with the Transact-SQL language.

Size and amount of data has led to applications and reporting solutions not performing are required speeds

Tier 7 and 8 data access is constrained to single endpoints managed by partners for access

The company maintains several legacy client applications. Data for these applications remains isolated form other applications. This has led to hundreds of databases being provisioned on a per application basis.

Question:

You need set up the Azure Data Factory JSON definition for Tier 10 data.

What should you use?

Connector =

-connection string

-linked service name string

-gateway connection string

-data store name string

Data movement activity =

-Azure SQL data warehouse

-Azure Files

-Azure Blob

-Azure SQL database

- ○

  Connector =
  -linked service name string
  Data movement activity =
  -Azure SQL data warehouse

- ○

Connector =
-gateway connection string
Data movement activity =
-Azure SQL database

○

Connector =
-connection string
Data movement activity =
-Azure Blob

(Correct)

○

Connector =
-data store name string
Data movement activity =
-Azure Files

Explanation

Requirement in the case study :Tier 10 reporting data must be stored in Azure Blobs

The Azure Storage and Azure SQL Database linked services contain **connection strings** that Data Factory uses at runtime to connect to your Azure Storage and Azure SQL Database, respectively.

https://docs.microsoft.com/en-us/azure/data-factory/concepts-linked-services

Question 26: Skipped
Contoso relies on an extensive partner network for marketing, sales, and distribution. Contoso uses external companies that manufacture everything from the actual pharmaceutical to the packaging.

The majority of the company's data reside in Microsoft SQL Server database. Application databases fall into one of the following tiers:

Larger image

Applications	Tier	Replication	Notes
Internal Contoso	1	Yes	
Internal Contoso	2	SQL Data Sync	Data Sync between databases
Internal Partner	3	Yes	Replicate to Partner
External Contoso	4,5,6	Yes	
External Partner	7,8	No	Partner managed
Internal Distribution and Sales	9	Yes, once ingested at branches	Data ingested from Contoso branches
External Distribution and Sales	10	Yes, once ingested at Contoso main office	Data is ingested from multiple sources

The company has a reporting infrastructure that ingests data from local databases and partner services. Partners services consists of distributors, wholesales, and retailers across the world. The company performs daily, weekly, and monthly reporting.

Requirements -

Tier 3 and Tier 6 through Tier 8 application must use database density on the same server and Elastic pools in a cost-effective manner.

Applications must still have access to data from both internal and external applications keeping the data encrypted and secure at rest and in transit.

A disaster recovery strategy must be implemented for Tier 3 and Tier 6 through 8 allowing for failover in the case of server going offline.

Selected internal applications must have the data hosted in single Microsoft Azure SQL Databases.

Tier 1 internal applications on the premium P2 tier

Tier 2 internal applications on the standard S4 tier

The solution must support migrating databases that support external and internal application to Azure SQL Database. The migrated databases will be supported by Azure Data Factory pipelines for the continued movement, migration and updating of data both in the cloud and from local core business systems and repositories.

Tier 7 and Tier 8 partner access must be restricted to the database only.

In addition to default Azure backup behavior, Tier 4 and 5 databases must be on a backup strategy that performs a transaction log backup eve hour, a differential backup of databases every day and a full back up every week.

Back up strategies must be put in place for all other standalone Azure SQL Databases using Azure SQL-provided backup storage and capabilities.

Databases -

Contoso requires their data estate to be designed and implemented in the Azure Cloud. Moving to the cloud must not inhibit access to or availability of data.

Databases:

Tier 1 Database must implement data masking using the following masking logic:

Larger image

Data type	Masking requirement
A	Mask 4 or less string data type characters
B	Mask first letter and domain
C	Mask everything except characters at the beginning and end

Tier 2 databases must sync between branches and cloud databases and in the event of conflicts must be set up for conflicts to be won by on-premises databases.

Tier 3 and Tier 6 through Tier 8 applications must use database density on the same server and Elastic pools in a cost-effective manner.

Applications must still have access to data from both internal and external applications keeping the data encrypted and secure at rest and in transit.

A disaster recovery strategy must be implemented for Tier 3 and Tier 6 through 8 allowing for failover in the case of a server going offline.

Selected internal applications must have the data hosted in single Microsoft Azure SQL Databases.

Tier 1 internal applications on the premium P2 tier

Tier 2 internal applications on the standard S4 tier

Reporting -

Security and monitoring -

Security -

A method of managing multiple databases in the cloud at the same time is must be implemented to streamlining data management and limiting management access to only those requiring access.

Monitoring -

Monitoring must be set up on every database. Contoso and partners must receive performance reports as part of contractual agreements.

Tiers 6 through 8 must have unexpected resource storage usage immediately reported to data engineers.

The Azure SQL Data Warehouse cache must be monitored when the database is being used. A dashboard monitoring key performance indicators (KPIs) indicated by traffic lights must be created and displayed based on the following metrics:

Larger image

Metric	Description
A	Low cache hit %, high cache usage %
B	Low cache hit %, low cache usage %
C	High cache hit %, high cache usage %

Existing Data Protection and Security compliances require that all certificates and keys are internally managed in an on-premises storage.

You identify the following reporting requirements:

Azure Data Warehouse must be used to gather and query data from multiple internal and external databases

Azure Data Warehouse must be optimized to use data from a cache

Reporting data aggregated for external partners must be stored in Azure Storage and be made available during regular business hours in the connecting regions

Reporting strategies must be improved to real time or near real time reporting cadence to improve competitiveness and the general supply chain

Tier 9 reporting must be moved to Event Hubs, queried, and persisted in the same Azure region as the company's main office

Tier 10 reporting data must be stored in Azure Blobs

Issues -

Team members identify the following issues:

Both internal and external client application run complex joins, equality searches and group-by clauses. Because some systems are managed externally, the queries will not be changed or optimized by Contoso

External partner organization data formats, types and schemas are controlled by the partner companies

Internal and external database development staff resources are primarily SQL developers familiar with the Transact-SQL language.

Size and amount of data has led to applications and reporting solutions not performing are required speeds

Tier 7 and 8 data access is constrained to single endpoints managed by partners for access

The company maintains several legacy client applications. Data for these applications remains isolated form other applications. This has led to hundreds of databases being provisioned on a per application basis.

You need to configure data encryption for external applications.

Solution:

1. Access the Always Encrypted Wizard in SQL Server Management Studio

2. Select the column to be encrypted

3. Set the encryption type to Deterministic

4. Configure the master key to use the Windows Certificate Store

5. Validate configuration results and deploy the solution

Does the solution meet the goal?

- ○

  Yes

    (Correct)

- ○

  No

Explanation

Yes it has all the correct steps.

Click **Next** on the **Introduction** page to open the **Column Selection** page. On this page, you will select which columns you want to encrypt, the type of encryption, and what column encryption key (CEK) to use.

Encrypt **SSN** and **BirthDate** information for each patient. The **SSN** column will use deterministic encryption, which supports equality lookups, joins, and group by.
The **BirthDate** column will use randomized encryption, which does not support operations.

Set the **Encryption Type** for the **SSN** column to **Deterministic** and the **BirthDate** column to **Randomized**. Click **Next**.

https://docs.microsoft.com/en-us/azure/azure-sql/database/always-encrypted-certificate-store-configure

Question 27:

You are working as a data engineer for a company and you have been given the task to access Azure Blob Storage from Azure Databricks using secrets stored in a key vault. Your company provided you with the storage account, the blob container and Azure key vault in place.

You implement the following steps:

· Add the secret to the storage container

· Create a Databricks workspace and add the access keys

· Access the blob container from Azure Databricks

Do you think that these steps are enough to fulfil the requirement?

- ○

  Yes

    (Correct)

- ○

  No

Explanation

Create a storage account and blob container

Create an Azure Key Vault and add a secret

Create an Azure Databricks workspace and add a secret scope

Access your blob container from Azure Databricks

https://docs.microsoft.com/en-us/azure/databricks/scenarios/store-secrets-azure-key-vault

Question 28:

You are working as a data engineer for a company and you have been given the task to access Azure Blob Storage from Azure Databricks using secrets stored in a key vault. Your company provided you with the storage account, the blob container and Azure key vault in place.

You implement the following steps:

· Add the secret to the key vault

· Create a Databricks workspace and add the secret scope

· Access the blob container from Azure Databricks

Do you think that these steps are enough to fulfil the requirement?

- ○
  Yes

  (Correct)

- ○
  No

Explanation

Create a storage account and blob container

Create an Azure Key Vault and add a secret

Create an Azure Databricks workspace and add a secret scope

Access your blob container from Azure Databricks

https://docs.microsoft.com/en-us/azure/databricks/scenarios/store-secrets-azure-key-vault

Question 29: Skipped
You are working as a data engineer for a company and you have been given the task to access Azure Blob Storage from Azure Databricks using secrets stored in a key vault. Your company provided you with the storage account, the blob container and Azure key vault in place.

You implement the following steps:

· Add the secret to the key vault

· Create a Databricks workspace and add the access keys

· Access the blob container from Azure Databricks

Do you think that these steps are enough to fulfil the requirement?

- ○
  Yes

- ○
  No

  (Correct)

Explanation

Create a storage account and blob container

Create an Azure Key Vault and add a secret

Create an Azure Databricks workspace and add a secret scope

Access your blob container from Azure Databricks

https://docs.microsoft.com/en-us/azure/databricks/scenarios/store-secrets-azure-key-vault

Question 30: Skipped
Your company is creating a e-commerce website where different vendors can sell their products.

These products can have same or different attributes. For example, first vendor can have product name Bycycle which contains attributes like Price, Description, Brand, PartNumber, and Size.

And another vendor with same product as Bycycle have attributes like Price, Description, Weight, and Model.

As a database engineer, you need to design a data solution that can accommodate with product attributes for same product. You should also be able to query product data using .Net query.

Do you think, Cosmos DB with Table API will meet the goal of this solution?

- ○

  No

- ○

  Yes

    (Correct)

Explanation

Azure Cosmos DB provides the Table API for applications that are written for Azure Table storage and that need premium capabilities like:

Turnkey global distribution.

Dedicated throughput worldwide.

Single-digit millisecond latencies at the 99th percentile.

Guaranteed high availability.

Automatic secondary indexing.

Applications written for Azure Table storage can migrate to Azure Cosmos DB by using the Table API with no code changes and take advantage of premium capabilities. **The Table API has client SDKs available for .NET, Java, Python, and Node.js.**

https://docs.microsoft.com/en-us/azure/cosmos-db/table-introduction

Question 31: _{Skipped}

Your company is creating a e-commerce website where different vendors can sell their products.

These products can have same or different attributes. For example, first vendor can have product name Bycycle which contains attributes like Price, Description, Brand, PartNumber, and Size.

And another vendor with same product as Bycycle have attributes like Price, Description, Weight, and Model.

As a database engineer, you need to design a data solution that can accommodate with product attributes for same product. You should also be able to query product data using .Net query.

Do you think, Azure SQL Database account that uses a managed instance will meet the goal of this solution?

- ○

  Yes

- ○

  No

  (Correct)

Explanation

SQL Database is a relational database but here the data to be stored is Non Relational in nature.

Question 32: _{Skipped}

Your company is creating a e-commerce website where different vendors can sell their products.

These products can have same or different attributes. For example, first vendor can have product name Bycycle which contains attributes like Price, Description, Brand, PartNumber, and Size.

And another vendor with same product as Bycycle have attributes like Price, Description, Weight, and Model.

As a database engineer, you need to design a data solution that can accommodate with product attributes for same product. You should also be able to query product data using .Net query.

Do you think, if You create a table storage account, it will meet the goal of this solution?

- ○

  No

- ○

  Yes

  (Correct)

Explanation

Azure Table storage stores large amounts of structured data. The service is a NoSQL datastore which accepts authenticated calls from inside and outside the Azure cloud. Azure tables are ideal for storing structured, non-relational data. Common uses of Table storage include:

Storing TBs of structured data capable of serving web scale applications

Storing datasets that don't require complex joins, foreign keys, or stored procedures and can be denormalized for fast access

Quickly querying data using a clustered index

Accessing data using the OData protocol and LINQ queries with WCF Data Service .NET Libraries

https://docs.microsoft.com/en-us/azure/cosmos-db/table-storage-overview

Question 33: Skipped
Your company is creating a e-commerce website where different vendors can sell their products.

These products can have same or different attributes. For example, first vendor can have product name Bycycle which contains attributes like Price, Description, Brand, PartNumber, and Size.

And another vendor with same product as Bycycle have attributes like Price, Description, Weight, and Model.

As a database engineer, you need to design a data solution that can accommodate with product attributes for same product. You should also be able to query product data using .Net query.

Do you think, Cosmos DB account that uses the SQL API will meet the goal of this solution?

- ○

  Yes

  (Correct)

- ○

  No

Explanation

In Azure Cosmos DB SQL API accounts, there are two ways to read data:

Point reads - You can do a key/value lookup on a single *item ID* and partition key. The *item ID* and partition key combination is the key and the item itself is the value. For a 1 KB document, point reads typically cost 1 request unit with a latency under 10 ms. Point reads return a single item.

Here are some examples of how to do **Point reads** with each SDK:

.NET SDK

Java SDK

Node.js SDK

Python SDK

https://docs.microsoft.com/en-us/azure/cosmos-db/sql-query-getting-started

Question 34: Skipped
Overview

A food delivery organization known as FoodCorp has a globally distributed applications. The data solutions provided by FoodCorps utilize the following Azure products specialized for data.

Azure Blob Storage for storing the tracking status of each order.

Azure Cosmos DB for storing customer orders, data regarding the customer and restaurant data such as menu, address etc.

Azure SQL Database elastic pool to be utilized by internal company systems

Requirements

The platform should be adaptable and flexible enough to deal with service failure or breakdown, if it happens, that affects one of the availability zones within Azure region.

A secure mobile application access needs to be provided to the database that stores customer orders which is hosted in Cosmos DB by Azure.

The application must be able to read and write queries straight to the database that stores customer orders.

The elastic pools stores the fifteen configured databases used by the internal systems.

In case of service failure of breakdown in Azure Availability Region, the internal system databases should be in a readable format.

A new external partner of FoodCorps must be provided when deploying in a new region that has the read access to directly view the tracking order history saved in the Azure Blob Storage.

The external partner must also not have unlimited access to FoodCorps and the access should be automated to expire after two-months

Orders Database

The orders database caused a blockage in the course of reading the orders collection in the database during the marketing campaign of a preceding regional launch. The throughput of the Request Units DB had to be increased for the purpose of eliminating the blockage. No change was done to other collections having a throughput different from the throughput possessed by orders collection.

Internal system databases

An administrative task needs to be executed daily on all databases present in the system. This task is a PowerShell script.

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

Which two actions are you supposed to take? Each correct answer is part of the solution.

- ☐

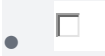   Set up SQL Data Sync for each database.

- ☐

    Create an elastic pool for Azure SQL Server in another region.

        (Correct)

- ☐

    Enable the geo-replication of each database.

(Correct)

- ☐

Create an elastic pool Sync Group.

Explanation

Active geo-replication is an Azure SQL Database feature that allows you to create readable secondary databases of individual databases on a server in the same or different data center (region).

https://docs.microsoft.com/en-us/azure/azure-sql/database/active-geo-replication-overview

Question 35: Skipped
Overview

A food delivery organization known as FoodCorp has a globally distributed applications. The data solutions provided by FoodCorps utilize the following Azure products specialized for data.

Azure Blob Storage for storing the tracking status of each order.

Azure Cosmos DB for storing customer orders, data regarding the customer and restaurant data such as menu, address etc.

Azure SQL Database elastic pool to be utilized by internal company systems

Requirements

The platform should be adaptable and flexible enough to deal with service failure or breakdown, if it happens, that affects one of the availability zones within Azure region.

A secure mobile application accesss needs to be provided to the database that stores customer orders which is hosted in Cosmos DB by Azure.

The application must be able to read and write queries straight to the database that stores customer orders.

The elastic pools stores the fifteen configured databases used by the internal systems.

In case of service failure of breakdown in Azure Availability Region, the internal system databases should be in a readable format.

A new external partner of FoodCorps must be provided when deploying in a new region that has the read access to directly view the tracking order history saved in the Azure Blob Storage.

The external partner must also not have unlimited access to FoodCorps and the access should be automated to expire after two-months

Orders Database

The orders database caused a blockage in the course of reading the orders collection in the database during the marketing campaign of a preceding regional launch. The throughput of the Request Units DB had to be increased for the purpose of eliminating the blockage. No change was done to other collections having a throughput different from the throughput possessed by orders collection.

Internal system databases

An administrative task needs to be executed daily on all databases present in the system. This task is a PowerShell script.

Question:-

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 Redendancy Strategy:

A. Geo Redundant Storage (GRS)

B. Local Redundant Storage (GRS)

C. Zone Redundant Storage (GRS)

Storage account type:

D. Block Blog Storage Account

E. General Purpose V2

- ○

  A and D

- ○

  B and E

- ○

  C and E

  (Correct)

- ○

C and D

Explanation

As per below statement in the case study we need to implement ZRS

"The platform should be adaptable and flexible enough to deal with service failure or breakdown, if it happens, that affects one of the availability zones within Azure region."

General Purpose v2 provides access to the latest Azure storage features, including Cool and Archive storage, with pricing optimised for the lowest GB storage prices.

https://azure.microsoft.com/en-in/pricing/details/storage/data-lake/

Question 36: Skipped
Overview

A food delivery organization known as FoodCorp has a globally distributed applications. The data solutions provided by FoodCorps utilize the following Azure products specialized for data.

Azure Blob Storage for storing the tracking status of each order.

Azure Cosmos DB for storing customer orders, data regarding the customer and restaurant data such as menu, address etc.

Azure SQL Database elastic pool to be utilized by internal company systems

Requirements

The platform should be adaptable and flexible enough to deal with service failure or breakdown, if it happens, that affects one of the availability zones within Azure region.

A secure mobile application access needs to be provided to the database that stores customer orders which is hosted in Cosmos DB by Azure.

The application must be able to read and write queries straight to the database that stores customer orders.

The elastic pools stores the fifteen configured databases used by the internal systems.

In case of service failure of breakdown in Azure Availability Region, the internal system databases should be in a readable format.

A new external partner of FoodCorps must be provided when deploying in a new region that has the read access to directly view the tracking order history saved in the Azure Blob Storage.

The external partner must also not have unlimited access to FoodCorps and the access should be automated to expire after two-months

Orders Database

The orders database caused a blockage in the course of reading the orders collection in the database during the marketing campaign of a preceding regional launch. The

throughput of the Request Units DB had to be increased for the purpose of eliminating the blockage. No change was done to other collections having a throughput different from the throughput possessed by orders collection.

Internal system databases

An administrative task needs to be executed daily on all databases present in the system. This task is a PowerShell script.

Question:-

You have been given the task to implement daily administrative tasks for the internal system databases. You also have to minimize administrative efforts.

What solution should you be using?

- ○

   SQL Server Agent

- ○

   Azure Alerts action group

- ○

   Elastic Database Jobs

      (Correct)

- ○

   Azure Data Factory

Explanation

You can create and schedule jobs that could be periodically executed against one or many databases to run Transact-SQL (T-SQL) queries and perform maintenance tasks.

https://docs.microsoft.com/en-us/azure/azure-sql/database/job-automation-overview

Question 37: Skipped
Overview

A food delivery organization known as FoodCorp has a globally distributed applications. The data solutions provided by FoodCorps utilize the following Azure products specialized for data.

Azure Blob Storage for storing the tracking status of each order.

Azure Cosmos DB for storing customer orders, data regarding the customer and restaurant data such as menu, address etc.

Azure SQL Database elastic pool to be utilized by internal company systems

Requirements

The platform should be adaptable and flexible enough to deal with service failure or breakdown, if it happens, that affects one of the availability zones within Azure region.

A secure mobile application access needs to be provided to the database that stores customer orders which is hosted in Cosmos DB by Azure.

The application must be able to read and write queries straight to the database that stores customer orders.

The elastic pools stores the fifteen configured databases used by the internal systems.

In case of service failure of breakdown in Azure Availability Region, the internal system databases should be in a readable format.

A new external partner of FoodCorps must be provided when deploying in a new region that has the read access to directly view the tracking order history saved in the Azure Blob Storage.

The external partner must also not have unlimited access to FoodCorps and the access should be automated to expire after two-months

Orders Database

The orders database caused a blockage in the course of reading the orders collection in the database during the marketing campaign of a preceding regional launch. The throughput of the Request Units DB had to be increased for the purpose of eliminating the blockage. No change was done to other collections having a throughput different from the throughput possessed by orders collection.

Internal system databases

An administrative task needs to be executed daily on all databases present in the system. This task is a PowerShell script.

Question:-

You have been given the task to create the required credentials for the mobile client application to connect with Azure Cosmos DB.

What credential key and permission should you be using to meet the security requirements? Select the relevant options to answer.

Credential Key:

A. Azure Active Directory (Azure AD) user

B. Maser Key

C. Resource Token

Permissions:

D. All

E. Cosmos DB Operator role

F. Read

- ○

  A and D

- ○

  B and E

- ○

  C and F

- ○

  C and D

  (Correct)

Explanation

Basis below statement,

"The application must be able to read and write queries straight to the database that stores customer orders."

the answer should be "All" to indicate Read and Write.

Resource tokens - Used for application resources: containers, documents, attachments, stored procedures, triggers, and UDFs

https://docs.microsoft.com/en-us/azure/cosmos-db/secure-access-to-data

Question 38: Skipped
Overview

A food delivery organization known as FoodCorp has a globally distributed applications. The data solutions provided by FoodCorps utilize the following Azure products specialized for data.

Azure Blob Storage for storing the tracking status of each order.

Azure Cosmos DB for storing customer orders, data regarding the customer and restaurant data such as menu, address etc.

Azure SQL Database elastic pool to be utilized by internal company systems

Requirements

The platform should be adaptable and flexible enough to deal with service failure or breakdown, if it happens, that affects one of the availability zones within Azure region.

A secure mobile application access needs to be provided to the database that stores customer orders which is hosted in Cosmos DB by Azure.

The application must be able to read and write queries straight to the database that stores customer orders.

The elastic pools stores the fifteen configured databases used by the internal systems.

In case of service failure of breakdown in Azure Availability Region, the internal system databases should be in a readable format.

A new external partner of FoodCorps must be provided when deploying in a new region that has the read access to directly view the tracking order history saved in the Azure Blob Storage.

The external partner must also not have unlimited access to FoodCorps and the access should be automated to expire after two-months

Orders Database

The orders database caused a blockage in the course of reading the orders collection in the database during the marketing campaign of a preceding regional launch. The throughput of the Request Units DB had to be increased for the purpose of eliminating the blockage. No change was done to other collections having a throughput different from the throughput possessed by orders collection.

Internal system databases

An administrative task needs to be executed daily on all databases present in the system. This task is a PowerShell script.

Question:-

You have been given the task to grant access to the blob containers that store the order tracking status for the external partner of the Company1.

What authorization method should you be using to meet the security requirements?

- ○

  Anonymous public read access

- ○

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

- ○

  Shared Key authorization

- ○

  Shared access signatures (SAS)

     (Correct)

Explanation

A shared access signature (SAS) provides secure delegated access to resources in your storage account without compromising the security of your data. With a SAS, you have granular control over how a client can access your data. You can control what resources the client may access, what permissions they have on those resources, and how long the SAS is valid, among other parameters.

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

Question 39: Skipped

You have been given the task to manage an Azure SQL Database containing a column with sensitive data and this column is used for joins by an application.

To protect the column, you have to configure encryption for this encryption.

You select the following solution: You configure Always Encrypted with a randomized type.

Do you think that this solution can meet the goal?

- ○

  Yes

- ○

  No

     (Correct)

Explanation

Deterministic encryption always generates the same encrypted value for any given plain text value. Using deterministic encryption allows point lookups, equality joins, grouping and indexing on encrypted columns.

https://docs.microsoft.com/en-us/sql/relational-databases/security/encryption/always-encrypted-database-engine?view=sql-server-ver15

Question 40: Skipped

You have been given the task to manage an Azure SQL Database containing a column with sensitive data and this column is used for joins by an application.

To protect the column, you have to configure encryption for this encryption.

You select the following solution: You configure Always Encrypted with a deterministic type.

Do you think that this solution can meet the goal?

- ○

  Yes

  (Correct)

- ○

  No

Explanation

Deterministic encryption always generates the same encrypted value for any given plain text value. Using deterministic encryption allows point lookups, equality joins, grouping and indexing on encrypted columns.

https://docs.microsoft.com/en-us/sql/relational-databases/security/encryption/always-encrypted-database-engine?view=sql-server-ver15

Question 41: Skipped

You have been given the task to manage an Azure SQL Database containing a column with sensitive data and this column is used for joins by an application.

To protect the column, you have to configure encryption for this encryption.

You select the following solution: You configure dynamic data masking (DDM) with random masking.
Do you think that this solution can meet the goal?

- ○

  Yes

- ○

No

(Correct)

Explanation

Deterministic encryption always generates the same encrypted value for any given plain text value. Using deterministic encryption allows point lookups, equality joins, grouping and indexing on encrypted columns.

https://docs.microsoft.com/en-us/sql/relational-databases/security/encryption/always-encrypted-database-engine?view=sql-server-ver15

Question 42: Skipped

You have been given the task to manage an Azure SQL Database containing a column with sensitive data and this column is used for joins by an application.

To protect the column, you have to configure encryption for this encryption.

You select the following solution: You configure dynamic data masking (DDM) with partial masking.

Do you think that this solution can meet the goal?

- ○

Yes

(Correct)

- ○

No

Explanation

Partial data masking is not randomized.

Deterministic encryption always generates the same encrypted value for any given plain text value. Using deterministic encryption allows point lookups, equality joins, grouping and indexing on encrypted columns.

https://docs.microsoft.com/en-us/sql/relational-databases/security/encryption/always-encrypted-database-engine?view=sql-server-ver15

https://docs.microsoft.com/en-us/sql/relational-databases/security/dynamic-data-masking?view=sql-server-ver15

Question 43: Skipped

You are developing a solution that will use Azure Stream Analytics. The solution will accept an Azure Blob storage file named Customers. The file will contain both in-store and online customer details. The online customers will provide a mailing address.
You have a file in Blob storage named LocationIncomes that contains based on location. The file rarely changes.
You need to use an address to look up a median income based on location. You must output the data to Azure SQL Database for immediate use and to Azure Data Lake Storage Gen2 for long-term retention.
Solution: You implement a Stream Analytics job that has one streaming input, one query, and two outputs.
Does this meet the goal?

- ○

  Yes

- ○

  No

  (Correct)

Explanation

We need one reference data input for LocationIncomes, which rarely changes.

We need two queries, on for in-store customers, and one for online customers.

For each query two outputs is needed.

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-define-outputs

Question 44: Skipped
You are developing a solution that will use Azure Stream Analytics. The solution will accept an Azure Blob storage file named Customers. The file will contain both in-store and online customer details. The online customers will provide a mailing address.

You have a file in Blob storage named LocationIncomes that contains based on location. The file rarely changes.

You need to use an address to look up a median income based on location. You must output the data to Azure SQL Database for immediate use and to Azure

Data Lake Storage Gen2 for long-term retention.

Solution: You implement a Stream Analytics job that has one streaming input, one reference input, two queries, and four outputs.

Does this meet the goal?

- ○

Yes

(Correct)

- ○

No

Explanation

We need one reference data input for LocationIncomes, which rarely changes.

We need two queries, on for in-store customers, and one for online customers.

For each query two outputs is needed.

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-define-outputs

Question 45: Skipped
You need to collect applications metrics, streaming query events, and application log messages for an Azure Databricks cluster.

What type of library should you implement?

- ○

Azure Databricks Monitoring Library

(Correct)

- ○

Microsoft Azure Management Monitoring Library

- ○

PyTorch

- ○

TensorFlow

Explanation

This article shows how to send application logs and metrics from Azure Databricks to a Log Analytics workspace. It uses the Azure Databricks Monitoring Library, which is available on GitHub.

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

Question 46: Skipped

You need to collect applications metrics, streaming query events, and application log messages for an Azure Databricks cluster.

What type of workspace should you implement?

- ○

  Azure Databricks

- ○

  Azure Log Analytics

  (Correct)

- ○

  Azure Machine Learning

Explanation

This article shows how to send application logs and metrics from Azure Databricks to a Log Analytics workspace. It uses the Azure Databricks Monitoring Library, which is available on GitHub.

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

Question 47: Skipped

You plan to build a structured streaming solution in Azure Databricks. The solution will count new events in five minutes intervals and report only events that arrive during the interval. The output will be sent to Data Lake Table.

The output mode you should you use?

- ○

  complete

- ○

  append

(Correct)

○

update

Explanation

Complete Mode: The entire updated result table is written to external storage. It is up to the storage connector to decide how to handle the writing of the entire table.

Append Mode: Only new rows appended in the result table since the last trigger are written to external storage. This is applicable only for the queries where existing rows in the Result Table are not expected to change.

Update Mode: Only the rows that were updated in the result table since the last trigger are written to external storage. This is different from Complete Mode in that Update Mode outputs only the rows that have changed since the last trigger. If the query doesn't contain aggregations, it is equivalent to Append mode.

https://docs.databricks.com/getting-started/spark/streaming.html

Question 48: Skipped
You are monitoring an Azure Steam Analytics job by using metrics in Azure Stream Analytics job by using metrics in Azure.

You discover that during the last 12 hours, the average watermark delay is consistently greater than the configured later arrival tolerance.

What should you use?

○

The late arrival policy causes events to be dropped

○

Events whose application timestamp is earlier than their arrival time by more than five minutes arrive as inputs

○

There are errors in the input data

○

The job lacks the resources to process the volume of incoming data

Explanation

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-set-up-alerts

Question 49: Skipped
You have an Azure data factory.

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

What would you use?

- ○

  Azure Monitor

  (Correct)

- ○

  the Resource health blade for the Data Factory resource

- ○

  the Activity log blade for the Data Factory resource

- ○

  the Monitor & Manage app in Data Factory

Explanation

Data Factory stores pipeline-run data for only 45 days. Use Azure Monitor if you want to keep that data for a longer time.

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

Question 50: Skipped
You are developing a data engineering solution for a company. The solution will store a large set of key-value pair data by using Microsoft Azure Cosmos DB.

The solution has following requirements:

1. Data must be partitioned into multiple containers

2. Data containers must be configured separately

3. Data must be accessible from applications hosted around the world

4. The solution must minimize latency

You need to provision cosmos DB.

Which 3 actions should you perform?

- ☐

  Provision an Azure Cosmos DB account with Azure Table API. Enable geo-redundancy

  (Correct)

- ☐

  Configure table-level throughput

  (Correct)

- ☐

  Replicate the data globally by manually adding regions to the Azure Cosmos DB account

  (Correct)

- ☐

  Provision an Azure Cosmos DB account with the Azure Table API. Enable multi-region writes

- ☐

  Configure account-level throughput

Explanation

We can answer this question by using the elimination method.

1. Throughput is set up at container level or database level

https://docs.microsoft.com/en-us/azure/cosmos-db/request-units

2. For accessing resources from around the world we need to enable geo redundancy and add the desired regions for read latency. Multi region write may or may not be enabled depending upon write latency requirements. First requirement is always multi region read followed by multi-region write

https://docs.microsoft.com/en-us/azure/cosmos-db/high-availability

Question 51: Skipped
You have an Azure Storage account.

You need to configure the storage account to send an email when an administrative action is performed on the storage account.

What should you do?

- ○

  Create an alert based on metric

- ○

  Create an alert based activity log

  (Correct)

- ○

  Create a custom role for the storage account

- ○

  Enable Advance Threat Protection

Explanation

Typically, you create activity log alerts to receive notifications when:

Specific operations occur on resources in your Azure subscription, often scoped to particular resource groups or resources. For example, you might want to be notified when any virtual machine in myProductionResourceGroup is **deleted.** Or, you might want to be notified if any **new roles are assigned** to a user in your subscription.

A service health event occurs. Service health events include notification of incidents and maintenance events that apply to resources in your subscription.

https://docs.microsoft.com/en-us/azure/azure-monitor/platform/activity-log-alerts

Question 52: Skipped
You have an enterprise wide Azure Data Lake Storage Gen2 account. The data lake is accessible only through Azure virtual network names VNET1.

You are building a SQL Pool in Azure Synapse that will use data from the data lake.

Your company has a sales team. All the members of the sales team are in an Azure Active Directory group named Sales. POSIX controls are used to assign the Sales group access to the files in the data lake.

You plan to load data in the SQL pool every hour.

You need to ensure that the SQL pool can load the sales data from the data lake.

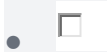**What 3 actions you need to perform.**

- ☐

  Use the managed identity as the credentials for the data load process

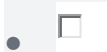 **(Correct)**

- ☐

  Add your Azure Active Directory account to the Sales Group

  (Correct)

- ☐

  Use the shared access signature as the credentials for the data load process

- ☐

  Create a shared access signature

- ☐

  Add managed identity to the Sales Group

  (Correct)

Explanation

This is a real exam question and we are slightly unsure of the answers. Do let us know if you know the right answer.

Practice Test 2

Question 1: Skipped
You are working as a data engineer for a compay and you keep Azure Storage access keys in Azure Key Vault.

To enable secure access to Azure Blob Storage, you have to configure a reference to those keys from within Azure Databricks.

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

Do you think this solution can meet the goal?

- ◯

  Yes

- ◯

  No

  (Correct)

Explanation

Process to access Azure Blob storage from Azure Databricks

Create a storage account and blob container

Create an Azure Key Vault and add a secret

Create an Azure Databricks workspace and add a secret scope

Access your blob container from Azure Databricks

https://docs.microsoft.com/en-us/azure/databricks/scenarios/store-secrets-azure-key-vault

Question 2: Skipped
You are working as a data engineer for a compay and you keep Azure Storage access keys in Azure Key Vault.

To enable secure access to Azure Blob Storage, you have to configure a reference to those keys from within Azure Databricks.

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

Do you think this solution can meet the goal?

- ◯

  Yes

- ○

No

(Correct)

Explanation

Process to access blob storage from Azure Databricks

Create a storage account and blob container

Create an Azure Key Vault and add a secret

Create an Azure Databricks workspace and add a secret scope

Once your Azure Databricks workspace is open in a separate window, append **#secrets/createScope** to the URL. The URL should have the following format:

https://<\location>.azuredatabricks.net/?o=<\orgID>#secrets/createScope.

Access your blob container from Azure Databricks

https://docs.microsoft.com/en-us/azure/databricks/scenarios/store-secrets-azure-key-vault

Question 3: Skipped
You are working as a data engineer for a compay and you keep Azure Storage access keys in Azure Key Vault.

To enable secure access to Azure Blob Storage, you have to configure a reference to those keys from within Azure Databricks.

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

Do you think this solution can meet the goal?

- ○

Yes

(Correct)

- ○

No

Explanation

Process to access Azure Storage from Databricks

Create a storage account and blob container

Create an Azure Key Vault and add a secret

Create an Azure Databricks workspace and add a secret scope

Access your blob container from Azure Databricks

https://docs.microsoft.com/en-us/azure/databricks/scenarios/store-secrets-azure-key-vault

Question 4: Skipped

You are working as a data engineer for your company. To process real-time defect detection events from your factory's production lines your company uses Azure Stream Analytics (ASA).

You now have to select the right windowing function in your ASA job's SELECT query, so that you could perform the following things:

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.

You select the following solution: You group events by using the session window in the ASA job query.

Do you think this solution will meet the goal?

- ○

  Yes

     (Correct)

- ○

  No

Explanation

Session windows group events that arrive at similar times, filtering out periods of time where there is no data. Session window function has three main parameters: timeout, maximum duration, and partitioning key (optional).

https://docs.microsoft.com/en-us/stream-analytics-query/session-window-azure-stream-analytics

Question 5: Skipped

You are working as a data engineer for your company. To process real-time defect detection events from your factory's production lines your company uses Azure Stream Analytics (ASA).

You now have to select the right windowing function in your ASA job's SELECT query, so that you could perform the following things:

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.

You select the following solution: You group events by using the tumbling window in the ASA job query.

Do you think this solution will meet the goal?

- ○

    Yes

- ○

    No

        (Correct)

Explanation

Tumbling windows are a series of fixed-sized, non-overlapping and contiguous time intervals. The following diagram illustrates a stream with a series of events and how they are mapped into 10-second tumbling windows.

https://docs.microsoft.com/en-us/stream-analytics-query/tumbling-window-azure-stream-analytics

Question 6: Skipped
You are working as a data engineer for your company. To process real-time defect detection events from your factory's production lines your company uses Azure Stream Analytics (ASA).

You now have to select the right windowing function in your ASA job's SELECT query, so that you could perform the following things:

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.

You select the following solution: You group events by using the hopping window in the ASA job query.

Do you think this solution will meet the goal?

- ○

  Yes

- ○

  No

    (Correct)

Explanation

Unlike tumbling windows, hopping windows model scheduled overlapping windows.

https://docs.microsoft.com/en-us/stream-analytics-query/hopping-window-azure-stream-analytics

Question 7: Skipped

You are working as a data engineer for your company. To process real-time defect detection events from your factory's production lines your company uses Azure Stream Analytics (ASA).

You now have to select the right windowing function in your ASA job's SELECT query, so that you could perform the following things:

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.

You select the following solution: You group events by using the sliding window in the ASA job query.

Do you think this solution will meet the goal?

- ○

  Yes

- ○

  No

    (Correct)

Explanation

When using a sliding window, the system is asked to logically consider all possible windows of a given length. As the number of such windows would be infinite, Azure Stream Analytics instead outputs events only for those points in time when the content of the window actually changes, in other words when an event entered or exits the window.

https://docs.microsoft.com/en-us/stream-analytics-query/sliding-window-azure-stream-analytics

Question 8: Skipped
You are working as a data engineer for an autonomous vehicle manufacturer company. This company fits a transmitter is each vehicle that submits sensor data over Advanced Message Queuing Protocol (AMQP). You are given the task to retrieve the data by the sensor in real time so that you can extract relevant information, transform it, and then send it to Power BI.

You have to implement this solution as soon as possible.

The solution is:

You do the following things

Create an Event Hub instance.

Create a Stream Analytics job that uses a query to extract data.

Do you think that this solution is enough to complete the task?

- ○

  Yes

    (Correct)

- ○

  No

Explanation

An Azure Stream Analytics job consists of an input, query, and an output. Stream Analytics ingests data from Azure Event Hubs (including Azure Event Hubs from Apache Kafka), Azure IoT Hub, or Azure Blob Storage.

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-introduction

Question 9: Skipped
You are working as a data engineer for an autonomous vehicle manufacturer company. This company fits a transmitter is each vehicle that submits sensor data over Advanced Message Queuing Protocol (AMQP). You are given the task to retireve

the data by the sensor in real time so that you can extract relevant information, transform it, and then send it to Power BI.

You have to implement this solution as soon as possible.

The solution is:

You do the following things

Create an IoT Hub instance.

Create a Stream Analytics job that uses a query to extract data.

Do you think that this solution is enough to complete the task?

- ○

  Yes

    (Correct)

- ○

  No

Explanation

An Azure Stream Analytics job consists of an input, query, and an output. Stream Analytics ingests data from Azure Event Hubs (including Azure Event Hubs from Apache Kafka), Azure IoT Hub, or Azure Blob Storage.

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-introduction

Question 10: Skipped
You are working as a data engineer for an autonomous vehicle manufacturer company. This company fits a transmitter is each vehicle that submits sensor data over Advanced Message Queuing Protocol (AMQP). You are given the task to retireve the data by the sensor in real time so that you can extract relevant information, transform it, and then send it to Power BI.

You have to implement this solution as soon as possible.

The solution is:

You do the following things

Create an Azure Databricks instance.

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

Do you think that this solution is enough to complete the task?

- ○

Yes

○

No

(Correct)

Explanation

We need a solution which has streaming component in it.

An Azure Stream Analytics job consists of an input, query, and an output. Stream Analytics ingests data from Azure Event Hubs (including Azure Event Hubs from Apache Kafka), Azure IoT Hub, or Azure Blob Storage.

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-introduction

Question 11: Skipped

You are working as a data engineer for an autonomous vehicle manufacturer company. This company fits a transmitter is each vehicle that submits sensor data over Advanced Message Queuing Protocol (AMQP). You are given the task to retireve the data by the sensor in real time so that you can extract relevant information, transform it, and then send it to Power BI.

You have to implement this solution as soon as possible.

The solution is:

You do the following things

Create an Azure Relay service.

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

Do you think that this solution is enough to complete the task?

○

Yes

○

No

(Correct)

Explanation

We need a solution which has streaming component in it.

An Azure Stream Analytics job consists of an input, query, and an output. Stream Analytics ingests data from Azure Event Hubs (including Azure Event Hubs from Apache Kafka), Azure IoT Hub, or Azure Blob Storage.

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-introduction

Question 12: Skipped

You are working as a data engineer for a company. To stream real-time IoT sensor data for a brewery company your company uses Azure Stream Analysis.

You are given the task to use a window function with a fixed-size. You should note that the events should belong to a single window.
You select the following solution: You analyze the stream with a session window function.

Do you think this solution will meet the goal?

- ○ Yes

- ○ No

    (Correct)

Explanation

Session window is not fixed width.

https://docs.microsoft.com/en-us/stream-analytics-query/session-window-azure-stream-analytics

Question 13: Skipped

You are working as a data engineer for a company. To stream real-time IoT sensor data for a brewery company your company uses Azure Stream Analysis.

You are given the task to use a window function with a fixed-size. You should note that the events could belong to more than one window.
You select the following solution: You analyze the stream with a tumbling window function.

Do you think this solution will meet the goal?

- ○ Yes

- ○

No

(Correct)

Explanation

Tumbling windows are a series of fixed-sized, non-overlapping and contiguous time intervals.

https://docs.microsoft.com/en-us/stream-analytics-query/tumbling-window-azure-stream-analytics

Question 14: Skipped
You are working as a data engineer for a company. To stream real-time IoT sensor data for a brewery company your company uses Azure Stream Analysis.

You are given the task to use a window function with a fixed-size. You should note that the events could belong to more than one window.
You select the following solution: You analyze the stream with a hopping window function.

Do you think this solution will meet the goal?

- ○

Yes

(Correct)

- ○

No

Explanation

Unlike tumbling windows, hopping windows model scheduled overlapping windows. A hopping window specification consist of three parameters: the *timeunit*, the *windowsize* (how long each window lasts) and the *hopsize* (by how much each window moves forward relative to the previous one).

https://docs.microsoft.com/en-us/stream-analytics-query/hopping-window-azure-stream-analytics

Question 15: Skipped
You are working as a data engineer for a company. You have been given the task to manage an Azure SQL Database service and monitor the database weekly for performance.

You just have to choose the tool or service that should monitor the database for performance opportunities through the creation or dropping of indexes.

You select the following solution: You use Query Performance Insight.

Do you think this solution will help you meet the goal?

- ○ Yes

- ○ No

  (Correct)

Explanation

Azure SQL Database learns and adapts with your application. Azure SQL Database has a number of database advisors that provide customized recommendations that enable you to maximize performance. These database advisors continuously assess and analyze the usage history and provide recommendations based on workload patterns that help improve performance.

https://docs.microsoft.com/en-us/azure/azure-sql/database/database-advisor-implement-performance-recommendations

Question 16: Skipped
You are working as a data engineer for a company. You have been given the task to manage an Azure SQL Database service and monitor the database weekly for performance.

You just have to choose the tool or service that should monitor the database for performance opportunities through the creation or dropping of indexes.

You select the following solution: You use SQL Database Advisor.
Do you think this solution will help you meet the goal?

- ○ Yes

  (Correct)

- ○ No

Explanation

Azure SQL Database learns and adapts with your application. Azure SQL Database has a number of database advisors that provide customized recommendations that enable you to maximize performance. These database advisors continuously assess and analyze the usage history and provide recommendations based on workload patterns that help improve performance.

https://docs.microsoft.com/en-us/azure/azure-sql/database/database-advisor-implement-performance-recommendations

Question 17: Skipped

You are working as a data engineer for a company. You have been given the task to manage an Azure SQL Database service and monitor the database weekly for performance.

You just have to choose the tool or service that should monitor the database for performance opportunities through the creation or dropping of indexes.

You select the following solution: You use Azure Advisor.
Do you think this solution will help you meet the goal?

- ○

  Yes

- ○

  No

      (Correct)

Explanation

Azure SQL Database learns and adapts with your application. Azure SQL Database has a number of database advisors that provide customized recommendations that enable you to maximize performance. These database advisors continuously assess and analyze the usage history and provide recommendations based on workload patterns that help improve performance.

https://docs.microsoft.com/en-us/azure/azure-sql/database/database-advisor-implement-performance-recommendations

Question 18: Skipped

You are working as a data engineer for a company that have an Azure Synapse Analytics SQL pool (data warehouse). In that, some columns store very sensitive data.

The applicable table/columns have to be searched and categorized and access to specific data should be logged and reported.

You perform the following solution:

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.

Do you think that this solution can meet the required goal?

- ○

  Yes

  (Correct)

- ○

  No

Explanation

Data Discovery & Classification introduces a set of advanced services and new capabilities in Azure. It forms a new information-protection paradigm for SQL Database, SQL Managed Instance, and Azure Synapse, aimed at protecting the data and not just the database. The paradigm includes:

Discovery and recommendations: The classification engine scans your database and identifies columns that contain potentially sensitive data. It then provides you with an easy way to review and apply recommended classification via the Azure portal.

Labeling: You can apply sensitivity-classification labels persistently to columns by using new metadata attributes that have been added to the SQL Server database engine. This metadata can then be used for advanced, sensitivity-based auditing and protection scenarios.

Query result-set sensitivity: The sensitivity of a query result set is calculated in real time for auditing purposes.

Visibility: You can view the database-classification state in a detailed dashboard in the Azure portal. Also, you can download a report in Excel format to use for compliance and auditing purposes and other needs.

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

Auditing for Azure SQL Database and Azure Synapse Analytics tracks database events and writes them to an audit log in your Azure storage account, Log Analytics workspace, or Event Hubs.

Question 19: Skipped

You are working as a data engineer for a company that have an Azure Synapse Analytics SQL pool (data warehouse). In that, some columns store very sensitive data.

The applicable table/columns have to be searched and categorized and access to specific data should be logged and reported.

You perform the following solution: Use the Set-AzSqlServerAuditPowerShell cmdlet. Do you think that this solution can meet the required goal?

- ○

   Yes

- ○

   No

      (Correct)

Explanation

Data Discovery & Classification introduces a set of advanced services and new capabilities in Azure. It forms a new information-protection paradigm for SQL Database, SQL Managed Instance, and Azure Synapse, aimed at protecting the data and not just the database. The paradigm includes:

Discovery and recommendations: The classification engine scans your database and identifies columns that contain potentially sensitive data. It then provides you with an easy way to review and apply recommended classification via the Azure portal.

Labeling: You can apply sensitivity-classification labels persistently to columns by using new metadata attributes that have been added to the SQL Server database engine. This metadata can then be used for advanced, sensitivity-based auditing and protection scenarios.

Query result-set sensitivity: The sensitivity of a query result set is calculated in real time for auditing purposes.

Visibility: You can view the database-classification state in a detailed dashboard in the Azure portal. Also, you can download a report in Excel format to use for compliance and auditing purposes and other needs.

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

Auditing for Azure SQL Database and Azure Synapse Analytics tracks database events and writes them to an audit log in your Azure storage account, Log Analytics workspace, or Event Hubs.

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

Question 20: Skipped

You are working as a data engineer for a company that have an Azure Synapse Analytics SQL pool (data warehouse). In that, some columns store very sensitive data.

The applicable table/columns have to be searched and categorized and access to specific data should be logged and reported.

You perform the following solution:

Enable database auditing.

Create security policies on the tables with sensitive data.

Write Kusto queries on the auditing data.

Do you think that this solution can meet the required goal?

- ○ Yes

- ○ No

  (Correct)

Explanation

Data Discovery & Classification introduces a set of advanced services and new capabilities in Azure. It forms a new information-protection paradigm for SQL Database, SQL Managed Instance, and Azure Synapse, aimed at protecting the data and not just the database. The paradigm includes:

Discovery and recommendations: The classification engine scans your database and identifies columns that contain potentially sensitive data. It then provides you with an easy way to review and apply recommended classification via the Azure portal.

Labeling: You can apply sensitivity-classification labels persistently to columns by using new metadata attributes that have been added to the SQL Server database engine. This metadata can then be used for advanced, sensitivity-based auditing and protection scenarios.

Query result-set sensitivity: The sensitivity of a query result set is calculated in real time for auditing purposes.

Question 21: Skipped

You are working as a data engineer for a company. You have been given the task to store inventory data for 30 stores. You create an instance in Azure SQL Data Warehouse. There are about 10,000 products in the stores, each of which has a different stock keeping unit (SKU) for each distinct product type. Every Friday the update takes place and it causes a massive change in the data.

You have to write T-SQL statement to create a table.

CREATE TABLE [dbo].[StoreInventoryData](

[Sku] [varchar] (30),

[Price] [decimal] (30,0),

[StoreId] [varchar] (20),

[Quantity] [int],

[Date] [datetime]

)

WITH

(CLUSTERED COLUMNSTORE INDEX

------------A---------- = ------------B------------(-------------C----------)

)

How will you fill A,B,C

- ☐

  Distribution

    (Correct)

- ☐

Partition

- ☐

Hash

(Correct)

- ☐

Replicate

- ☐

Round_Robin

- ☐

Sku

(Correct)

- ☐

StoreID

Explanation

Refer the SQL statement here

https://docs.microsoft.com/en-us/sql/t-sql/statements/create-table-azure-sql-data-warehouse?view=aps-pdw-2016-au7

A hash-distributed table distributes table rows across the Compute nodes by using a deterministic hash function to assign each row to one distribution.

https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-distribute#:~:text=A%20round%2Drobin%20distributed%20table,assigned%20to%20the%20same%20distribution.

Question 22: Skipped
You are working as a data engineer for a company. You have been given the task to store inventory data for 30 stores. You create an instance in Azure SQL Data Warehouse. To improve the performance, you have to shard the store data. Data size is under 200 Megabytes (MB).

Complete the T- SQL statements below for this

```
CREATE TABLE  [dbo].[Store](
        [StoreId] [varchar)(30),
        [StoreName][varchar](200),
        [StoreNumber][int],
        [StoreLocation][varchar](200)
)
WITH
( CLUSTERED COLUMN STORE INDEX,
-------------A-------------= ------------B-----------(StoreId)
)
```

Fill A & B

- ☐

  Distribution

    (Correct)

- ☐

  Partition

- ☐

  HASH

- ☐

  Replicate

    (Correct)

- ☐

  Round_Robin

Explanation

Refer the syntax here

Question 23: Skipped

To apply for a job in company, the company has given you a task. The task is to house inventory data for 1000 stores in an instance in Azure SQL Data Warehouse. Each store has a different store number. The task in simple, you have to shard the data to improve the performance when accessing it. You have to make sure that the data is evenly sharded across compute nodes.

Write a T-SQL statement to generate the store table.

CREATE TABLE [dbo].[Store](

[StoreId] [varchar)(30),

[StoreName][varchar](200),

[StoreNumber][int],

[StoreLocation][varchar](200)

)

WITH

(CLUSTERED COLUMN STORE INDEX,

--------------A-------------= ------------B-----------(StoreId)

)

Fill A & B

- ☐

  Distribution

    (Correct)

- ☐

  Partition

- ☐

  HASH

- ☐

  Replicate

- ☐

  Round_Robin

  (Correct)

Explanation

Refer the syntax here

https://docs.microsoft.com/en-us/sql/t-sql/statements/create-table-azure-sql-data-warehouse?view=aps-pdw-2016-au7

A round-robin distributed table distributes table rows evenly across all distributions. The assignment of rows to distributions is random. Unlike hash-distributed tables, rows with equal values are not guaranteed to be assigned to the same distribution.

https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-distribute#:~:text=A%20round%2Drobin%20distributed%20table,assigned%20to%20the%20same%20distribution.

Question 24: Skipped
You are working as a data engineer for your company that is provisioning multiple Azure SQL managed instances for an automotive industry application. This application should be resilient in case of an outage impacting an Azure region.

Your company thinks and creates a primary instance in the West US region called AutoIndustry and a second instance in the East US region called AutoIndustrySecondary. These instances are provisioned in a resource group called AutoProduction.

You have been given the task to implement an auto-failover group for database disaster recovery.

How should you be completing the PowerShell cmdlet? Select the relevant options to answer.

Larger image

Larger image

A-

1. New-AzSqlDatabaseFailoverGroup

2. New-AzSqlDatabaseInstanceFailoverGroup

3. Set-AZSqlDatabaseFailoverGroup

B-

4. PartnerManagedInstanceName

5. PrimaryManagedInstanceName

6. ResourceGroupName

C-

7. PartnerManagedInstanceName

8. PrimaryManagedInstanceName

9. ResourceGroupName

D-

10. PartnerManagedInstanceName

11. PrimaryManagedInstanceName

12. ResourceGroupName

- ○

  2,6,8,10

    (Correct)

- ○

  1, 6, 7, 11

- ○

  3, 4, 9, 12

- ○

3, 5, 8, 10

Explanation

Refer the command here:

https://docs.microsoft.com/en-us/powershell/module/az.sql/new-azsqldatabaseinstancefailovergroup?view=azps-5.0.0

Question 25:
You are implementing automatic tuning mode for Azure SQL databases.

Automatic tuning is configured as shown in the following table.

Larger image

Option	Server level	Database level
Force Plan	Inherited	Inherited
Create Index	Inherited	Inherited
Drop Index	Inherited	Inherited

Select if below statement is TRUE or FALSE?

Force Plan for database is always ON?

- ○

  Yes

    (Correct)

- ○

  No

Explanation

Automatic tuning options can be independently enabled or disabled per database, or they can be configured at the server-level and applied on every database that inherits settings from the server. **Servers can inherit Azure defaults** for automatic tuning settings. Azure defaults at this time are set to **FORCE_LAST_GOOD_PLAN is enabled**, CREATE_INDEX is enabled, and DROP_INDEX is disabled.

Question 26:
You are implementing automatic tuning mode for Azure SQL databases.

Automatic tuning is configured as shown in the following table.

Option	Server level	Database level
Force Plan	Inherited	Inherited
Create Index	Inherited	Inherited
Drop Index	Inherited	Inherited

Select if TRUE or FALSE

Create Index for database is ON?

- ○

  Yes

  (Correct)

- ○

  No

Explanation

Automatic tuning options can be independently enabled or disabled per database, or they can be configured at the server-level and applied on every database that inherits settings from the server. **Servers can inherit Azure defaults** for automatic tuning settings. Azure defaults at this time are set to FORCE_LAST_GOOD_PLAN is enabled, C**REATE_INDEX is enabled**, and DROP_INDEX is disabled.

Question 27: Skipped
You are implementing automatic tuning mode for Azure SQL databases.

Automatic tuning is configured as shown in the following table.

Option	Server level	Database level
Force Plan	Inherited	Inherited
Create Index	Inherited	Inherited
Drop Index	Inherited	Inherited

Select if TRUE or FALSE?

Drop Index for the database is ON?

- ○

  Yes

- ○

  No

  (Correct)

Explanation

Automatic tuning options can be independently enabled or disabled per database, or they can be configured at the server-level and applied on every database that inherits settings from the server. Servers can inherit Azure defaults for automatic tuning settings. Azure defaults at this time are set to FORCE_LAST_GOOD_PLAN is enabled, CREATE_INDEX is enabled, and **DROP_INDEX is disabled.**

Question 28: Skipped
Proseware, Inc, develops and manages a product named Poll Taker. The product is used for delivering public opinion polling and analysis.
Polling data comes from a variety of sources, including online surveys, house-to-house interviews, and booths at public events.

Polling data -
Polling data is stored in one of the two locations:
An on-premises Microsoft SQL Server 2019 database named PollingData
Azure Data Lake Gen 2
Data in Data Lake is queried by using PolyBase

Poll metadata -
Each poll has associated metadata with information about the poll including the date and number of respondents. The data is stored as JSON.

Phone-based polling -

Security -
Phone-based poll data must only be uploaded by authorized users from authorized devices
Contractors must not have access to any polling data other than their own
Access to polling data must set on a per-active directory user basis

Data migration and loading -
All data migration processes must use Azure Data Factory
All data migrations must run automatically during non-business hours
Data migrations must be reliable and retry when needed

Performance -
After six months, raw polling data should be moved to a storage account. The storage must be available in the event of a regional disaster. The solution must minimize costs.

Deployments -
All deployments must be performed by using Azure DevOps. Deployments must use templates used in multiple environments
No credentials or secrets should be used during deployments

Larger image
.

Reliability -
All services and processes must be resilient to a regional Azure outage.

Monitoring -
All Azure services must be monitored by using Azure Monitor. On-premises SQL Server performance must be monitored.
Question
You need to provision the polling data storage account.
How should you configure the storage account?

Larger image

Configuration values

- LRS
- GRS
- RA-GRS
- Storage
- StorageV2

Answer Area

Setting	Configuration value
Account type	
Replication type	

○

Account Type : StorageV2
Replication Type : RA-GRS

(Correct)

○

Account Type : StorageV2
Replication Type : GRS

○

Account Type : Storage
Replication Type : GRS

○

Account Type : Storage
Replication Type : LRS

Explanation

Question 29: Skipped

Proseware, Inc, develops and manages a product named Poll Taker. The product is used for delivering public opinion polling and analysis.
Polling data comes from a variety of sources, including online surveys, house-to-house interviews, and booths at public events.

Polling data -
Polling data is stored in one of the two locations:
An on-premises Microsoft SQL Server 2019 database named PollingData
Azure Data Lake Gen 2
Data in Data Lake is queried by using PolyBase

Poll metadata -
Each poll has associated metadata with information about the poll including the date and number of respondents. The data is stored as JSON.

Phone-based polling -

Security -
Phone-based poll data must only be uploaded by authorized users from authorized devices
Contractors must not have access to any polling data other than their own
Access to polling data must set on a per-active directory user basis

Data migration and loading -
All data migration processes must use Azure Data Factory
All data migrations must run automatically during non-business hours
Data migrations must be reliable and retry when needed

Performance -
After six months, raw polling data should be moved to a storage account. The storage must be available in the event of a regional disaster. The solution must minimize costs.

Deployments -
All deployments must be performed by using Azure DevOps. Deployments must use templates used in multiple environments
No credentials or secrets should be used during deployments

Larger image
.

Reliability -
All services and processes must be resilient to a regional Azure outage.

Monitoring -
All Azure services must be monitored by using Azure Monitor. On-premises SQL Server performance must be monitored.
You need to make sure polling data security requirements are met. Which security technologies would you use?

Larger image

- ○

  SQL Server : Managed Identity
  Polybase : Database encryption key

- ○

  SQL Server : Azure Active Directory user
  Polybase : Database scoped credential

  (Correct)

- ○

  SQL Server : Domain Active Directory User
  Polybase : Application role

Explanation

SQL Server uses a database scoped credential to access non-public Azure blob storage or Kerberos-secured Hadoop clusters with PolyBase.

https://docs.microsoft.com/en-us/sql/t-sql/statements/create-database-scoped-credential-transact-sql?view=sql-server-ver15

This statement implies "Access to polling data must set on a per-active directory user basis"
- Azure Active Directory

Question 30: Skipped
Proseware, Inc, develops and manages a product named Poll Taker. The product is used for delivering public opinion polling and analysis.
Polling data comes from a variety of sources, including online surveys, house-to-house interviews, and booths at public events.

Polling data -
Polling data is stored in one of the two locations:
An on-premises Microsoft SQL Server 2019 database named PollingData
Azure Data Lake Gen 2
Data in Data Lake is queried by using PolyBase

Poll metadata -
Each poll has associated metadata with information about the poll including the date and number of respondents. The data is stored as JSON.

Phone-based polling -

Security -
Phone-based poll data must only be uploaded by authorized users from authorized devices
Contractors must not have access to any polling data other than their own
Access to polling data must set on a per-active directory user basis

Data migration and loading -
All data migration processes must use Azure Data Factory
All data migrations must run automatically during non-business hours
Data migrations must be reliable and retry when needed

Performance -
After six months, raw polling data should be moved to a storage account. The storage must be available in the event of a regional disaster. The solution must minimize costs.

Deployments -
All deployments must be performed by using Azure DevOps. Deployments must use templates used in multiple environments
No credentials or secrets should be used during deployments

Larger image
▪

Reliability -

All services and processes must be resilient to a regional Azure outage.

Monitoring -
All Azure services must be monitored by using Azure Monitor. On-premises SQL Server performance must be monitored.
You need to ensure phone-based polling data upload reliability requirements are met. How should you configure monitoring?

Larger image

- ○

    Metric : BlobCapacity
    Aggregation : Avg

    (Correct)

- ○

    Metric : FileCapacity
    Aggregation : Avg

- ○

    Metric : FileCapacity
    Aggregation : Sum

Explanation

Since the data is stored in DataLake Gen which is built on top of blob we need to use BlobCapacity as the metric which support AVG as the aggregation

https://docs.microsoft.com/en-us/azure/azure-monitor/platform/metrics-supported

Question 31: Skipped
Proseware, Inc, develops and manages a product named Poll Taker. The product is used for delivering public opinion polling and analysis.

Polling data comes from a variety of sources, including online surveys, house-to-house interviews, and booths at public events.

Polling data -

Polling data is stored in one of the two locations:

An on-premises Microsoft SQL Server 2019 database named PollingData

Azure Data Lake Gen 2

Data in Data Lake is queried by using PolyBase

Poll metadata -

Each poll has associated metadata with information about the poll including the date and number of respondents. The data is stored as JSON.

Phone-based polling -

Security -

Phone-based poll data must only be uploaded by authorized users from authorized devices

Contractors must not have access to any polling data other than their own

Access to polling data must set on a per-active directory user basis

Data migration and loading -

All data migration processes must use Azure Data Factory

All data migrations must run automatically during non-business hours

Data migrations must be reliable and retry when needed

Performance -

After six months, raw polling data should be moved to a storage account. The storage must be available in the event of a regional disaster. The solution must minimize costs.

Deployments -

All deployments must be performed by using Azure DevOps. Deployments must use templates used in multiple environments

No credentials or secrets should be used during deployments

Reliability -

All services and processes must be resilient to a regional Azure outage.

Monitoring -

All Azure services must be monitored by using Azure Monitor. On-premises SQL Server performance must be monitored.

You need to ensure that phone-based polling data can be analyzed in the PollingData database.

Which three actions should you perform in sequence?

A) Parameterize deployment by using Azure Integration Runtime

B) Configure an Azure Logic App to deploy the deployment artifact

C) Configure Azure Devops to deploy the deployment artifact

D) Create a deployment artifact containing an extracted azure resource Manager template

E) Parameterize deployment by using the Azure Resource Manager template parameter file

F) Create a deployment artifact containing a SQL Server Integration Services(SSIS) package

- ○

  F-->A-->C

- ○

  D-->E-->C

  (Correct)

- ○

  D-->E-->B

Explanation

All deployments must be performed by using Azure DevOps. Deployments must use templates used in multiple environments

No credentials or secrets should be used during deployments

https://docs.microsoft.com/en-us/azure/azure-resource-manager/templates/template-tutorial-use-parameter-file?tabs=azure-powershell

Question 32: Skipped
Proseware, Inc, develops and manages a product named Poll Taker. The product is used for delivering public opinion polling and analysis.

Polling data comes from a variety of sources, including online surveys, house-to-house interviews, and booths at public events.

Polling data -

Polling data is stored in one of the two locations:

An on-premises Microsoft SQL Server 2019 database named PollingData

Azure Data Lake Gen 2

Data in Data Lake is queried by using PolyBase

Poll metadata -

Each poll has associated metadata with information about the poll including the date and number of respondents. The data is stored as JSON.

Phone-based polling -

Security -

Phone-based poll data must only be uploaded by authorized users from authorized devices

Contractors must not have access to any polling data other than their own

Access to polling data must set on a per-active directory user basis

Data migration and loading -

All data migration processes must use Azure Data Factory

All data migrations must run automatically during non-business hours

Data migrations must be reliable and retry when needed

Performance -

After six months, raw polling data should be moved to a storage account. The storage must be available in the event of a regional disaster. The solution must minimize costs.

Deployments -

All deployments must be performed by using Azure DevOps. Deployments must use templates used in multiple environments

No credentials or secrets should be used during deployments

Reliability -

All services and processes must be resilient to a regional Azure outage.

Monitoring -

All Azure services must be monitored by using Azure Monitor. On-premises SQL Server performance must be monitored.

You need to ensure that phone-based polling data can be analyzed in the PollingData database.

How should you configure Azure Data Factory?

- ○

  Use a tumbling schedule trigger

- ○

Use an event-based trigger

○

Use a schedule trigger

(Correct)

○

Use manual execution

Explanation

All data migration processes must use Azure Data Factory

All data migrations must run automatically during non-business hours

Data migrations must be reliable and retry when needed

As per above requirement, we should have a scheduled trigger.

When creating a schedule trigger, you specify a schedule (start date, recurrence, end date etc.) for the trigger, and associate with a pipeline.

https://docs.microsoft.com/en-us/azure/data-factory/how-to-create-schedule-trigger

Question 33: Skipped

Proseware, Inc, develops and manages a product named Poll Taker. The product is used for delivering public opinion polling and analysis.

Polling data comes from a variety of sources, including online surveys, house-to-house interviews, and booths at public events.

Polling data -

Polling data is stored in one of the two locations:

An on-premises Microsoft SQL Server 2019 database named PollingData

Azure Data Lake Gen 2

Data in Data Lake is queried by using PolyBase

Poll metadata -

Each poll has associated metadata with information about the poll including the date and number of respondents. The data is stored as JSON.

Phone-based polling -

Security -

Phone-based poll data must only be uploaded by authorized users from authorized devices

Contractors must not have access to any polling data other than their own

Access to polling data must set on a per-active directory user basis

Data migration and loading -

All data migration processes must use Azure Data Factory

All data migrations must run automatically during non-business hours

Data migrations must be reliable and retry when needed

Performance -

After six months, raw polling data should be moved to a storage account. The storage must be available in the event of a regional disaster. The solution must minimize costs.

Deployments -

All deployments must be performed by using Azure DevOps. Deployments must use templates used in multiple environments

No credentials or secrets should be used during deployments

Reliability -

All services and processes must be resilient to a regional Azure outage.

Monitoring -

All Azure services must be monitored by using Azure Monitor. On-premises SQL Server performance must be monitored.

You need to ensure that Azure Data Factory pipelines can be deployed. How should you configure authentication and authorization for deployments?

Larger image

Answer Area

Security requirement	Technology
Authorization	RBAC V
	DAC
	MAC
	Claims
Authentication	Service Principal ∧
	Kerberos
	Certificate-based
	Bearer Token

- ○

  Authorization : RBAC
  Authentication : Service Principle

  (Correct)

- ○

Authorization : MAC
Authentication : Certificate-based

•   ○

Authorization : DAC
Authentication : Bearer Token

•   ○

Authorization : Claims
Authentication : Kerberos

Explanation

Azure RBAC is an authorization system built on Azure Resource Manager that provides fine-grained access management of Azure resources.

A *security principal* is an object that represents a user, group, service principal, or managed identity that is requesting access to Azure resources. You can assign a role to any of these security principals.

https://docs.microsoft.com/en-us/azure/role-based-access-control/overview

Question 34: Skipped
Litware, Inc, is an international car racing and manufacturing company that has 1,000 employees. Most employees are located in Europe. The company supports racing teams that complete in a worldwide racing series.

Physical Locations -
Litware has two main locations: a main office in London, England, and a manufacturing plant in Berlin, Germany.
During each race weekend, 100 engineers set up a remote portable office by using a VPN to connect the datacentre in the London office. The portable office is set up and torn down in approximately 20 different countries each year.

Existing environment -

Race Central -
During race weekends, Litware uses a primary application named Race Central. Each car has several sensors that send real-time telemetry data to the London datacentre. The data is used for real-time tracking of the cars.
Race Central also sends batch updates to an application named Mechanical Workflow by using Microsoft SQL Server Integration Services (SSIS).
The telemetry data is sent to a MongoDB database. A custom application then moves the data to databases in SQL Server 2017. The telemetry data in MongoDB has more than 500 attributes. The application changes the attribute names when the data is moved to SQL Server 2017.
The database structure contains both OLAP and OLTP databases.

Mechanical Workflow -
Mechanical Workflow is used to track changes and improvements made to the cars during their lifetime.
Currently, Mechanical Workflow runs on SQL Server 2017 as an OLAP system.
Mechanical Workflow has a named Table1 that is 1 TB. Large aggregations are performed on a single column of Table 1.

Requirements -

Planned Changes -
Litware is the process of rearchitecting its data estate to be hosted in Azure. The company plans to decommission the London datacentre and move all its applications to an Azure datacentre.

Technical Requirements -
Litware identifies the following technical requirements:
Data collection for Race Central must be moved to Azure Cosmos DB and Azure SQL Database. The data must be written to the Azure datacentre closest to each race and must converge in the least amount of time.
The query performance of Race Central must be stable, and the administrative time it takes to perform optimizations must be minimized.
The datacentre for Mechanical Workflow must be moved to Azure SQL data Warehouse.
Transparent data encryption (IDE) must be enabled on all data stores, whenever possible.
An Azure Data Factory pipeline must be used to move data from Cosmos DB to SQL Database for Race Central. If the data load takes longer than 20 minutes, configuration changes must be made to Data Factory.
The telemetry data must migrate toward a solution that is native to Azure.
The telemetry data must be monitored for performance issues. You must adjust the Cosmos DB Request Units per second (RU/s) to maintain a performance
SLA while minimizing the cost of the Ru/s.

Data Masking Requirements -
During rare weekends, visitors will be able to enter the remote portable offices. Litware is concerned that some proprietary information might be exposed. The company identifies the following data masking requirements for the Race Central data that will be stored in SQL Database:
Only show the last four digits of the values in a column named SuspensionSprings.
Only Show a zero value for the values in a column named ShockOilWeight.

Question You are monitoring the Data Factory pipeline that runs from Cosmos DB to SQL Database for Race Central.
You discover that the job takes 45 minutes to run.
What should you do to improve the performance of the job?

- ○

Decrease parallelism for the copy activities.

○

Increase that data integration units.

(Correct)

○

Configure the copy activities to use staged copy.

○

Configure the copy activities to perform compression.

Explanation

Data Integration Units

A Data Integration Unit (DIU) is a measure that represents the power of a single unit in Azure Data Factory. Power is a combination of CPU, memory, and network resource allocation. DIU only applies to Azure integration runtime. DIU does not apply to self-hosted integration runtime. Learn more here.

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

Question 35: Skipped
Litware, Inc, is an international car racing and manufacturing company that has 1,000 employees. Most employees are located in Europe. The company supports racing teams that complete in a worldwide racing series.

Physical Locations -
Litware has two main locations: a main office in London, England, and a manufacturing plant in Berlin, Germany.
During each race weekend, 100 engineers set up a remote portable office by using a VPN to connect the datacentre in the London office. The portable office is set up and torn down in approximately 20 different countries each year.

Existing environment -

Race Central -
During race weekends, Litware uses a primary application named Race Central. Each car has several sensors that send real-time telemetry data to the London datacentre. The data is used for real-time tracking of the cars.
Race Central also sends batch updates to an application named Mechanical Workflow by using Microsoft SQL Server Integration Services (SSIS).
The telemetry data is sent to a MongoDB database. A custom application then moves

the data to databases in SQL Server 2017. The telemetry data in MongoDB has more than 500 attributes. The application changes the attribute names when the data is moved to SQL Server 2017.
The database structure contains both OLAP and OLTP databases.

Mechanical Workflow -
Mechanical Workflow is used to track changes and improvements made to the cars during their lifetime.
Currently, Mechanical Workflow runs on SQL Server 2017 as an OLAP system.
Mechanical Workflow has a named Table1 that is 1 TB. Large aggregations are performed on a single column of Table 1.

Requirements -

Planned Changes -
Litware is the process of rearchitecting its data estate to be hosted in Azure. The company plans to decommission the London datacentre and move all its applications to an Azure datacentre.

Technical Requirements -
Litware identifies the following technical requirements:
Data collection for Race Central must be moved to Azure Cosmos DB and Azure SQL Database. The data must be written to the Azure datacentre closest to each race and must converge in the least amount of time.
The query performance of Race Central must be stable, and the administrative time it takes to perform optimizations must be minimized.
The datacentre for Mechanical Workflow must be moved to Azure SQL data Warehouse.
Transparent data encryption (IDE) must be enabled on all data stores, whenever possible.
An Azure Data Factory pipeline must be used to move data from Cosmos DB to SQL Database for Race Central. If the data load takes longer than 20 minutes, configuration changes must be made to Data Factory.
The telemetry data must migrate toward a solution that is native to Azure.
The telemetry data must be monitored for performance issues. You must adjust the Cosmos DB Request Units per second (RU/s) to maintain a performance
SLA while minimizing the cost of the Ru/s.

Data Masking Requirements -
During rare weekends, visitors will be able to enter the remote portable offices. Litware is concerned that some proprietary information might be exposed. The company identifies the following data masking requirements for the Race Central data that will be stored in SQL Database:
Only show the last four digits of the values in a column named SuspensionSprings.
Only Show a zero value for the values in a column named ShockOilWeight

Question What should you implement to optimize SQL Database for Race Central to meet the technical requirements?

- ○

  the sp_updatestats stored procedure

  (Correct)

- ○

  automatic tuning

- ○

  Query Store

- ○

  the dbcc checkdb command

Explanation

Updates query optimization statistics on a table or indexed view. By default, the query optimizer already updates statistics as necessary to improve the query plan; in some cases you can improve query performance by using `UPDATE STATISTICS` or the stored procedure sp_updatestats to update statistics more frequently than the default updates.

https://docs.microsoft.com/en-us/sql/t-sql/statements/update-statistics-transact-sql?view=sql-server-ver15

Question 36: Skipped
Litware, Inc, is an international car racing and manufacturing company that has 1,000 employees. Most employees are located in Europe. The company supports racing teams that complete in a worldwide racing series.

Physical Locations -
Litware has two main locations: a main office in London, England, and a manufacturing plant in Berlin, Germany.
During each race weekend, 100 engineers set up a remote portable office by using a VPN to connect the datacentre in the London office. The portable office is set up and torn down in approximately 20 different countries each year.

Existing environment -

Race Central -
During race weekends, Litware uses a primary application named Race Central. Each car has several sensors that send real-time telemetry data to the London datacentre. The data is used for real-time tracking of the cars.
Race Central also sends batch updates to an application named Mechanical

Workflow by using Microsoft SQL Server Integration Services (SSIS).
The telemetry data is sent to a MongoDB database. A custom application then moves the data to databases in SQL Server 2017. The telemetry data in MongoDB has more than 500 attributes. The application changes the attribute names when the data is moved to SQL Server 2017.
The database structure contains both OLAP and OLTP databases.

Mechanical Workflow -
Mechanical Workflow is used to track changes and improvements made to the cars during their lifetime.
Currently, Mechanical Workflow runs on SQL Server 2017 as an OLAP system.
Mechanical Workflow has a named Table1 that is 1 TB. Large aggregations are performed on a single column of Table 1.

Requirements -

Planned Changes -
Litware is the process of rearchitecting its data estate to be hosted in Azure. The company plans to decommission the London datacentre and move all its applications to an Azure datacentre.

Technical Requirements -
Litware identifies the following technical requirements:
Data collection for Race Central must be moved to Azure Cosmos DB and Azure SQL Database. The data must be written to the Azure datacentre closest to each race and must converge in the least amount of time.
The query performance of Race Central must be stable, and the administrative time it takes to perform optimizations must be minimized.
The datacentre for Mechanical Workflow must be moved to Azure SQL data Warehouse.
Transparent data encryption (IDE) must be enabled on all data stores, whenever possible.
An Azure Data Factory pipeline must be used to move data from Cosmos DB to SQL Database for Race Central. If the data load takes longer than 20 minutes, configuration changes must be made to Data Factory.
The telemetry data must migrate toward a solution that is native to Azure.
The telemetry data must be monitored for performance issues. You must adjust the Cosmos DB Request Units per second (RU/s) to maintain a performance
SLA while minimizing the cost of the Ru/s.

Data Masking Requirements -
During rare weekends, visitors will be able to enter the remote portable offices.
Litware is concerned that some proprietary information might be exposed. The company identifies the following data masking requirements for the Race Central data that will be stored in SQL Database:
Only show the last four digits of the values in a column named SuspensionSprings.
Only Show a zero value for the values in a column named ShockOilWeight.

Question You need to build a solution to collect the telemetry data for Race Control. What should you use?

Larger image

Answer Area

API: ▼

Cassandra
Gremlin
MongoDB
SQL
Table

Consistency level: ▼

Eventual
Session
Strong

○

API : MongoDB
Consistency level : Strong

○

API : SQL
Consistency level : Eventual

○

API : Cassandra
Consistency level : Session

○

API : Gremlin
Consistency level : Session

- ○

API : MongoDB
Consistency level : Session

(Correct)

Explanation

Cosmos accounts configured with multiple write regions cannot be configured for strong consistency as it is not possible for a distributed system to provide an RPO of zero and an RTO of zero.

https://docs.microsoft.com/en-us/azure/cosmos-db/consistency-levels#strong-consistency-and-multiple-write-regions

Of the other two options, the preferable is Session then Eventual(which takes more time and is out of order). They both take same no of RU/s

Question 37: Skipped
Litware, Inc, is an international car racing and manufacturing company that has 1,000 employees. Most employees are located in Europe. The company supports racing teams that complete in a worldwide racing series.

Physical Locations -
Litware has two main locations: a main office in London, England, and a manufacturing plant in Berlin, Germany.
During each race weekend, 100 engineers set up a remote portable office by using a VPN to connect the datacentre in the London office. The portable office is set up and torn down in approximately 20 different countries each year.

Existing environment -

Race Central -
During race weekends, Litware uses a primary application named Race Central. Each car has several sensors that send real-time telemetry data to the London datacentre. The data is used for real-time tracking of the cars.
Race Central also sends batch updates to an application named Mechanical Workflow by using Microsoft SQL Server Integration Services (SSIS).
The telemetry data is sent to a MongoDB database. A custom application then moves the data to databases in SQL Server 2017. The telemetry data in MongoDB has more than 500 attributes. The application changes the attribute names when the data is moved to SQL Server 2017.
The database structure contains both OLAP and OLTP databases.

Mechanical Workflow -
Mechanical Workflow is used to track changes and improvements made to the cars during their lifetime.
Currently, Mechanical Workflow runs on SQL Server 2017 as an OLAP system. Mechanical Workflow has a named Table1 that is 1 TB. Large aggregations are performed on a single column of Table 1.

Requirements -

Planned Changes -
Litware is the process of rearchitecting its data estate to be hosted in Azure. The company plans to decommission the London datacentre and move all its applications to an Azure datacentre.

Technical Requirements -
Litware identifies the following technical requirements:
Data collection for Race Central must be moved to Azure Cosmos DB and Azure SQL Database. The data must be written to the Azure datacentre closest to each race and must converge in the least amount of time.
The query performance of Race Central must be stable, and the administrative time it takes to perform optimizations must be minimized.
The datacentre for Mechanical Workflow must be moved to Azure SQL data Warehouse.
Transparent data encryption (IDE) must be enabled on all data stores, whenever possible.
An Azure Data Factory pipeline must be used to move data from Cosmos DB to SQL Database for Race Central. If the data load takes longer than 20 minutes, configuration changes must be made to Data Factory.
The telemetry data must migrate toward a solution that is native to Azure.
The telemetry data must be monitored for performance issues. You must adjust the Cosmos DB Request Units per second (RU/s) to maintain a performance
SLA while minimizing the cost of the Ru/s.

Data Masking Requirements -
During rare weekends, visitors will be able to enter the remote portable offices. Litware is concerned that some proprietary information might be exposed. The company identifies the following data masking requirements for the Race Central data that will be stored in SQL Database:
Only show the last four digits of the values in a column named SuspensionSprings.
Only Show a zero value for the values in a column named ShockOilWeight.

Question On which data store you configure TDE to meet the technical requirements?

- ○

    Cosmos DB

- ○

- ○

SQL Database

Explanation

With the release of encryption at rest for Cosmos DB, all your databases, media attachments, and backups are encrypted. Your data is now encrypted in transit (over the network) and at rest (nonvolatile storage), giving you end-to-end encryption.

https://docs.microsoft.com/en-us/azure/cosmos-db/database-encryption-at-rest

By default, TDE is enabled for all newly deployed SQL Databases and must be manually enabled for older databases of Azure SQL Database, Azure SQL Managed Instance. TDE must be manually enabled for Azure Synapse Analytics.

https://docs.microsoft.com/en-us/azure/azure-sql/database/transparent-data-encryption-tde-overview?tabs=azure-portal

Question 38: Skipped

Litware, Inc, is an international car racing and manufacturing company that has 1,000 employees. Most employees are located in Europe. The company supports racing teams that complete in a worldwide racing series.

Physical Locations -

Litware has two main locations: a main office in London, England, and a manufacturing plant in Berlin, Germany.

During each race weekend, 100 engineers set up a remote portable office by using a VPN to connect the datacentre in the London office. The portable office is set up and torn down in approximately 20 different countries each year.

Existing environment -

Race Central -

During race weekends, Litware uses a primary application named Race Central. Each car has several sensors that send real-time telemetry data to the London datacentre. The data is used for real-time tracking of the cars.

Race Central also sends batch updates to an application named Mechanical Workflow by using Microsoft SQL Server Integration Services (SSIS).

The telemetry data is sent to a MongoDB database. A custom application then moves the data to databases in SQL Server 2017. The telemetry data in MongoDB has more than 500 attributes. The application changes the attribute names when the data is moved to SQL Server 2017.

The database structure contains both OLAP and OLTP databases.

Mechanical Workflow -

Mechanical Workflow is used to track changes and improvements made to the cars during their lifetime.

Currently, Mechanical Workflow runs on SQL Server 2017 as an OLAP system.

Mechanical Workflow has a named Table1 that is 1 TB. Large aggregations are performed on a single column of Table 1.

Requirements -

Planned Changes -

Litware is the process of rearchitecting its data estate to be hosted in Azure. The company plans to decommission the London datacentre and move all its applications to an Azure datacentre.

Technical Requirements -

Litware identifies the following technical requirements:

Data collection for Race Central must be moved to Azure Cosmos DB and Azure SQL Database. The data must be written to the Azure datacentre closest to each race and must converge in the least amount of time.

The query performance of Race Central must be stable, and the administrative time it takes to perform optimizations must be minimized.

The datacentre for Mechanical Workflow must be moved to Azure SQL data Warehouse.

Transparent data encryption (IDE) must be enabled on all data stores, whenever possible.

An Azure Data Factory pipeline must be used to move data from Cosmos DB to SQL Database for Race Central. If the data load takes longer than 20 minutes, configuration changes must be made to Data Factory.

The telemetry data must migrate toward a solution that is native to Azure.

The telemetry data must be monitored for performance issues. You must adjust the Cosmos DB Request Units per second (RU/s) to maintain a performance

SLA while minimizing the cost of the Ru/s.

Data Masking Requirements -

During rare weekends, visitors will be able to enter the remote portable offices. Litware is concerned that some proprietary information might be exposed. The

company identifies the following data masking requirements for the Race Central data that will be stored in SQL Database:

Only show the last four digits of the values in a column named SuspensionSprings.

Only Show a zero value for the values in a column named ShockOilWeight.

Question

You are building the data store solution for Mechanical Workflow.

How should you configure Table1?

Larger image

Answer Area

Table Type:

| Hash distributed |
| Replicated |
| Round-robin |

Index type:

| Clustered |
| Clustered columnstore |
| Heap |
| Nonclustered |

- ○

  Table Type: Round-robin
  Index Type: Clustered

- ○

  Table Type: Replicated
  Index Type: Heap

- ○

  Table Type: Hash Distributed
  Index Type: Clustered columnstore

  (Correct)

- ○

Explanation

Requirements - "Mechanical Workflow has a named Table1 that is 1 TB. Large aggregations are performed on a single column of Table 1"

Hash-distributed tables work well for large fact tables in a star schema. They can have very large numbers of rows and still achieve high performance.

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

Columnstore indexes are the standard for storing and querying large data warehousing fact tables. This index uses column-based data storage and query processing to achieve gains up to 10 times the query performance in your data warehouse over traditional row-oriented storage.

https://docs.microsoft.com/en-us/sql/relational-databases/indexes/columnstore-indexes-overview?view=sql-server-ver15

Question 39: Skipped
Litware, Inc, is an international car racing and manufacturing company that has 1,000 employees. Most employees are located in Europe. The company supports racing teams that complete in a worldwide racing series.

Physical Locations -
Litware has two main locations: a main office in London, England, and a manufacturing plant in Berlin, Germany.
During each race weekend, 100 engineers set up a remote portable office by using a VPN to connect the datacentre in the London office. The portable office is set up and torn down in approximately 20 different countries each year.

Existing environment -

Race Central -
During race weekends, Litware uses a primary application named Race Central. Each car has several sensors that send real-time telemetry data to the London datacentre. The data is used for real-time tracking of the cars.
Race Central also sends batch updates to an application named Mechanical Workflow by using Microsoft SQL Server Integration Services (SSIS).
The telemetry data is sent to a MongoDB database. A custom application then moves the data to databases in SQL Server 2017. The telemetry data in MongoDB has more than 500 attributes. The application changes the attribute names when the data is moved to SQL Server 2017.
The database structure contains both OLAP and OLTP databases.

Mechanical Workflow -
Mechanical Workflow is used to track changes and improvements made to the cars during their lifetime.
Currently, Mechanical Workflow runs on SQL Server 2017 as an OLAP system.
Mechanical Workflow has a named Table1 that is 1 TB. Large aggregations are performed on a single column of Table 1.

Requirements -

Planned Changes -
Litware is the process of rearchitecting its data estate to be hosted in Azure. The company plans to decommission the London datacentre and move all its applications to an Azure datacentre.

Technical Requirements -
Litware identifies the following technical requirements:
Data collection for Race Central must be moved to Azure Cosmos DB and Azure SQL Database. The data must be written to the Azure datacentre closest to each race and must converge in the least amount of time.
The query performance of Race Central must be stable, and the administrative time it takes to perform optimizations must be minimized.
The datacentre for Mechanical Workflow must be moved to Azure SQL data Warehouse.
Transparent data encryption (IDE) must be enabled on all data stores, whenever possible.
An Azure Data Factory pipeline must be used to move data from Cosmos DB to SQL Database for Race Central. If the data load takes longer than 20 minutes, configuration changes must be made to Data Factory.
The telemetry data must migrate toward a solution that is native to Azure.
The telemetry data must be monitored for performance issues. You must adjust the Cosmos DB Request Units per second (RU/s) to maintain a performance
SLA while minimizing the cost of the Ru/s.

Data Masking Requirements -
During rare weekends, visitors will be able to enter the remote portable offices. Litware is concerned that some proprietary information might be exposed. The company identifies the following data masking requirements for the Race Central data that will be stored in SQL Database:
Only show the last four digits of the values in a column named SuspensionSprings.
Only Show a zero value for the values in a column named ShockOilWeight.

Question Which two metrics should you use to identify the appropriate RU/s for the telemetry data?

- ☐
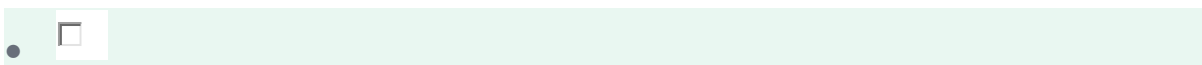
    Number of requests

(Correct)

- ☐

  Number of requests exceeded capacity

- ☐

  End to end observed read latency at the 99 th percentile

- ☐

  Session consistency

- ☐

  Data + Index storage consumed

  (Correct)

- ☐

  Avg Troughput/s

Explanation

Number of regions

Multi-region writes

Total data stored (per region)

Item size

Reads/sec per region

Writes/sec per region

https://docs.microsoft.com/en-us/azure/cosmos-db/estimate-ru-with-capacity-planner

Question 40: Skipped
Litware, Inc, is an international car racing and manufacturing company that has 1,000 employees. Most employees are located in Europe. The company supports racing teams that complete in a worldwide racing series.

Physical Locations -

Litware has two main locations: a main office in London, England, and a manufacturing plant in Berlin, Germany.

During each race weekend, 100 engineers set up a remote portable office by using a VPN to connect the datacentre in the London office. The portable office is set up and torn down in approximately 20 different countries each year.

Existing environment -

Race Central -

During race weekends, Litware uses a primary application named Race Central. Each car has several sensors that send real-time telemetry data to the London datacentre. The data is used for real-time tracking of the cars.

Race Central also sends batch updates to an application named Mechanical Workflow by using Microsoft SQL Server Integration Services (SSIS).

The telemetry data is sent to a MongoDB database. A custom application then moves the data to databases in SQL Server 2017. The telemetry data in MongoDB has more than 500 attributes. The application changes the attribute names when the data is moved to SQL Server 2017.

The database structure contains both OLAP and OLTP databases.

Mechanical Workflow -

Mechanical Workflow is used to track changes and improvements made to the cars during their lifetime.

Currently, Mechanical Workflow runs on SQL Server 2017 as an OLAP system.

Mechanical Workflow has a named Table1 that is 1 TB. Large aggregations are performed on a single column of Table 1.

Requirements -

Planned Changes -

Litware is the process of rearchitecting its data estate to be hosted in Azure. The company plans to decommission the London datacentre and move all its applications to an Azure datacentre.

Technical Requirements -

Litware identifies the following technical requirements:

Data collection for Race Central must be moved to Azure Cosmos DB and Azure SQL Database. The data must be written to the Azure datacentre closest to each race and must converge in the least amount of time.

The query performance of Race Central must be stable, and the administrative time it takes to perform optimizations must be minimized.

The datacentre for Mechanical Workflow must be moved to Azure SQL data Warehouse.

Transparent data encryption (IDE) must be enabled on all data stores, whenever possible.

An Azure Data Factory pipeline must be used to move data from Cosmos DB to SQL Database for Race Central. If the data load takes longer than 20 minutes, configuration changes must be made to Data Factory.

The telemetry data must migrate toward a solution that is native to Azure.

The telemetry data must be monitored for performance issues. You must adjust the Cosmos DB Request Units per second (RU/s) to maintain a performance

SLA while minimizing the cost of the Ru/s.

Data Masking Requirements -

During rare weekends, visitors will be able to enter the remote portable offices. Litware is concerned that some proprietary information might be exposed. The company identifies the following data masking requirements for the Race Central data that will be stored in SQL Database:

Only show the last four digits of the values in a column named SuspensionSprings.

Only Show a zero value for the values in a column named ShockOilWeight.

Question

Which masking functions should you implement for each column to meet the data masking requirements?

Larger image

Answer Area

ShockOilWieght:

| Credit card |
| Default |
| Email |
| Random number |

SuspensionSprings:

| Credit card |
| Default |
| Email |
| Random number |

ShockOilWieght: Default
SuspensionSprings: Credit Card

(Correct)

○

ShockOilWieght: Credit Card
SuspensionSprings: Default

○

ShockOilWieght: Email
SuspensionSprings: Random number

○

ShockOilWieght: Random number
SuspensionSprings: Email

Explanation

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

Question 41: Skipped
Litware, Inc, is an international car racing and manufacturing company that has 1,000 employees. Most employees are located in Europe. The company supports racing teams that complete in a worldwide racing series.

Physical Locations -
Litware has two main locations: a main office in London, England, and a manufacturing plant in Berlin, Germany.
During each race weekend, 100 engineers set up a remote portable office by using a VPN to connect the datacentre in the London office. The portable office is set up and torn down in approximately 20 different countries each year.

Existing environment -

Race Central -
During race weekends, Litware uses a primary application named Race Central. Each car has several sensors that send real-time telemetry data to the London datacentre. The data is used for real-time tracking of the cars.
Race Central also sends batch updates to an application named Mechanical Workflow by using Microsoft SQL Server Integration Services (SSIS).
The telemetry data is sent to a MongoDB database. A custom application then moves the data to databases in SQL Server 2017. The telemetry data in MongoDB has more than 500 attributes. The application changes the attribute names when the data is

moved to SQL Server 2017.
The database structure contains both OLAP and OLTP databases.

Mechanical Workflow -
Mechanical Workflow is used to track changes and improvements made to the cars during their lifetime.
Currently, Mechanical Workflow runs on SQL Server 2017 as an OLAP system.
Mechanical Workflow has a named Table1 that is 1 TB. Large aggregations are performed on a single column of Table 1.

Requirements -

Planned Changes -
Litware is the process of rearchitecting its data estate to be hosted in Azure. The company plans to decommission the London datacentre and move all its applications to an Azure datacentre.

Technical Requirements -
Litware identifies the following technical requirements:
Data collection for Race Central must be moved to Azure Cosmos DB and Azure SQL Database. The data must be written to the Azure datacentre closest to each race and must converge in the least amount of time.
The query performance of Race Central must be stable, and the administrative time it takes to perform optimizations must be minimized.
The datacentre for Mechanical Workflow must be moved to Azure SQL data Warehouse.
Transparent data encryption (IDE) must be enabled on all data stores, whenever possible.
An Azure Data Factory pipeline must be used to move data from Cosmos DB to SQL Database for Race Central. If the data load takes longer than 20 minutes, configuration changes must be made to Data Factory.
The telemetry data must migrate toward a solution that is native to Azure.
The telemetry data must be monitored for performance issues. You must adjust the Cosmos DB Request Units per second (RU/s) to maintain a performance
SLA while minimizing the cost of the Ru/s.

Data Masking Requirements -
During rare weekends, visitors will be able to enter the remote portable offices.
Litware is concerned that some proprietary information might be exposed. The company identifies the following data masking requirements for the Race Central data that will be stored in SQL Database:
Only show the last four digits of the values in a column named SuspensionSprings.
Only Show a zero value for the values in a column named ShockOilWeight.

Question What should you include in the Data Factory pipeline for Race Central?

- ○

a copy activity that uses a stored procedure as a source

○

a copy activity that contains schema mappings

(Correct)

○

a delete activity that has logging enabled

○

a filter activity that has a condition

Explanation

This article describes how the Azure Data Factory copy activity perform schema mapping and data type mapping from source data to sink data.

https://docs.microsoft.com/en-us/azure/data-factory/copy-activity-schema-and-type-mapping

Question 42: Skipped
You have an Azure Storage account that contains 100 GB of files. The files contain text and numerical values. 75% of the rows contain description data that has an average length of 1.1 MB.

You plan to copy the data from the storage account to an Azure SQL data warehouse.

You need to prepare the files to ensure that the data copies quickly.

Solution: You copy the files to a table that has a columnstore index.

Does this meet the goal?

○

Yes

○

No

(Correct)

Explanation

PolyBase loads are limited to rows smaller than 1 MB. It cannot be used to load to VARCHR(MAX), NVARCHAR(MAX), or VARBINARY(MAX). For more information, see Azure Synapse Analytics service capacity limits.

When your source data has rows greater than 1 MB, you might want to vertically split the source tables into several small ones. Make sure that the largest size of each row doesn't exceed the limit. The smaller tables can then be loaded by using PolyBase and merged together in Azure Synapse Analytics.

https://docs.microsoft.com/en-us/azure/data-factory/connector-azure-sql-data-warehouse

Question 43: Skipped

You have an Azure Storage account that contains 100 GB of files. The files contain text and numerical values. 75% of the rows contain description data that has an average length of 1.1 MB.

You plan to copy the data from the storage account to an Azure SQL data warehouse.

You need to prepare the files to ensure that the data copies quickly.

Solution: You modify the files to ensure that each row is less than 1 MB.

Does this meet the goal?

- ○

  Yes

    (Correct)

- ○

  No

Explanation

PolyBase loads are limited to rows smaller than 1 MB. It cannot be used to load to VARCHR(MAX), NVARCHAR(MAX), or VARBINARY(MAX). For more information, see Azure Synapse Analytics service capacity limits.

When your source data has rows greater than 1 MB, you might want to vertically split the source tables into several small ones. Make sure that the largest size of each row doesn't exceed the limit. The smaller tables can then be loaded by using PolyBase and merged together in Azure Synapse Analytics.

https://docs.microsoft.com/en-us/azure/data-factory/connector-azure-sql-data-warehouse

Question 44: Skipped

You have an Azure Storage account that contains 100 GB of files. The files contain text and numerical values. 75% of the rows contain description data that has an average length of 1.1 MB.

You plan to copy the data from the storage account to an Azure SQL data warehouse.

You need to prepare the files to ensure that the data copies quickly.

Solution: You modify the files to ensure that each row is more than 1 MB.

Does this meet the goal?

- ○

  Yes

- ○

  No

  (Correct)

Explanation

PolyBase loads are limited to rows smaller than 1 MB. It cannot be used to load to VARCHR(MAX), NVARCHAR(MAX), or VARBINARY(MAX). For more information, see Azure Synapse Analytics service capacity limits.

When your source data has rows greater than 1 MB, you might want to vertically split the source tables into several small ones. Make sure that the largest size of each row doesn't exceed the limit. The smaller tables can then be loaded by using PolyBase and merged together in Azure Synapse Analytics.

https://docs.microsoft.com/en-us/azure/data-factory/connector-azure-sql-data-warehouse

Question 45: Skipped
ADatum Corporation is a retailer that sells products through two sales channels: retail stores and a website.

Existing Environment -

ADatum has one database server that has Microsoft SQL Server 2016 installed. The server hosts three mission-critical databases named SALESDB, DOCDB, and REPORTINGDB.

SALESDB collects data from the store and the website.

DOCDB stored documents that connect to the sales data in SALESDB. The documents are stored in two different JSON formats based on the sales channel.

REPORTINGDB stores reporting data and contains server columnstore indexes. A daily process creates reporting data in REPORTINGDB from the data in

SALESDB. The process is implemented as a SQL Server Integration Services (SSIS) package that runs a stored procedure from SALESDB.

Requirements -

Planned Changes -

ADatum plans to move the current data infrastructure to Azure. The new infrastructure has the following requirements:

Migrate SALESDB and REPORTINGDB to an Azure SQL database.

Migrate DOCDB to Azure Cosmos DB.

The sales data including the documents in JSON format, must be gathered as it arrives and analyzed online by using Azure Stream Analytics. The analytic process will perform aggregations that must be done continuously, without gaps, and without overlapping.

As they arrive, all the sales documents in JSON format must be transformed into one consistent format.

Azure Data Factory will replace the SSIS process of copying the data from SALESDB to REPORTINGDB.

Technical Requirements -

The new Azure data infrastructure must meet the following technical requirements:

Data in SALESDB must encrypted by using Transparent Data Encryption (TDE). The encryption must use your own key.

SALESDB must be restorable to any given minute within the past three weeks.

Real-time processing must be monitored to ensure that workloads are sized properly based on actual usage patterns.

Missing indexes must be created automatically for REPORTINGDB.

Disk IO, CPU, and memory usage must be monitored for SALESDB.

Question How should you monitor SALESDB to meet the technical requirements?

- ○
  Query the sys.resource_stats dynamic management view.

  (Correct)

- ○

  Review the Query Performance Insights for SALESDB.

- ○

Query the sys.dm_os_wait_stats dynamic management view.

- ○

Review the auditing information of SALESDB.

Explanation

The **sys.resource_stats** returns historical data for CPU, IO, DTU consumption. There's one row every 5 minute for a database in an Azure logical SQL Server if there's a change in the metrics.

https://dataplatformlabs.com/monitoring-azure-sql-database-with-sys-resource_stats/

Question 46: Skipped
ADatum Corporation is a retailer that sells products through two sales channels: retail stores and a website.

Existing Environment -

ADatum has one database server that has Microsoft SQL Server 2016 installed. The server hosts three mission-critical databases named SALESDB, DOCDB, and REPORTINGDB.

SALESDB collects data from the store and the website.

DOCDB stored documents that connect to the sales data in SALESDB. The documents are stored in two different JSON formats based on the sales channel.

REPORTINGDB stores reporting data and contains server columnstore indexes. A daily process creates reporting data in REPORTINGDB from the data in SALESDB. The process is implemented as a SQL Server Integration Services (SSIS) package that runs a stored procedure from SALESDB.

Requirements -

Planned Changes -

ADatum plans to move the current data infrastructure to Azure. The new infrastructure has the following requirements:

Migrate SALESDB and REPORTINGDB to an Azure SQL database.

Migrate DOCDB to Azure Cosmos DB.

The sales data including the documents in JSON format, must be gathered as it arrives and analyzed online by using Azure Stream Analytics. The analytic process will perform aggregations that must be done continuously, without gaps, and without overlapping.

As they arrive, all the sales documents in JSON format must be transformed into one consistent format.

Azure Data Factory will replace the SSIS process of copying the data from SALESDB to REPORTINGDB.

Technical Requirements -

The new Azure data infrastructure must meet the following technical requirements:

Data in SALESDB must encrypted by using Transparent Data Encryption (TDE). The encryption must use your own key.

SALESDB must be restorable to any given minute within the past three weeks.

Real-time processing must be monitored to ensure that workloads are sized properly based on actual usage patterns.

Missing indexes must be created automatically for REPORTINGDB.

Disk IO, CPU, and memory usage must be monitored for SALESDB.

Question You need to implement event processing by using Stream Analytics to produce consistent JSON documents.

Which three actions should you perform?

- ☐

  Define an output to Cosmos DB.

- ☐

  Define a query that contains a JavaScript user-defined aggregates (UDA) function.

- ☐

  Define a reference input.

- ☐

  Define a transformation query.

    (Correct)

- ☐

  Define an output to Azure Data Lake Storage Gen2.

    (Correct)

- ☐

Question 47: Skipped

ADatum Corporation is a retailer that sells products through two sales channels: retail stores and a website.

Existing Environment -

ADatum has one database server that has Microsoft SQL Server 2016 installed. The server hosts three mission-critical databases named SALESDB, DOCDB, and REPORTINGDB.

SALESDB collects data from the store and the website.

DOCDB stored documents that connect to the sales data in SALESDB. The documents are stored in two different JSON formats based on the sales channel.

REPORTINGDB stores reporting data and contains server columnstore indexes. A daily process creates reporting data in REPORTINGDB from the data in SALESDB. The process is implemented as a SQL Server Integration Services (SSIS) package that runs a stored procedure from SALESDB.

Requirements -

Planned Changes -

ADatum plans to move the current data infrastructure to Azure. The new infrastructure has the following requirements:

Migrate SALESDB and REPORTINGDB to an Azure SQL database.

Migrate DOCDB to Azure Cosmos DB.

The sales data including the documents in JSON format, must be gathered as it arrives and analyzed online by using Azure Stream Analytics. The analytic process will perform aggregations that must be done continuously, without gaps, and without overlapping.

As they arrive, all the sales documents in JSON format must be transformed into one consistent format.

Azure Data Factory will replace the SSIS process of copying the data from SALESDB to REPORTINGDB.

Technical Requirements -

The new Azure data infrastructure must meet the following technical requirements:

Data in SALESDB must encrypted by using Transparent Data Encryption (TDE). The encryption must use your own key.

SALESDB must be restorable to any given minute within the past three weeks.

Real-time processing must be monitored to ensure that workloads are sized properly based on actual usage patterns.

Missing indexes must be created automatically for REPORTINGDB.

Disk IO, CPU, and memory usage must be monitored for SALESDB.

Question You need to ensure that the missing indexes for REPORTINGDB are added.

What should you use?

- ○

  SQL Database Advisor

- ○

  Extended events

- ○

  Query Performance Insight

- ○

  Automatic tuning

     (Correct)

Explanation

Automatic tuning for Azure SQL Database uses the **CREATE INDEX**, **DROP INDEX**, and **FORCE LAST GOOD PLAN** database advisor recommendations to optimize your database performance.

https://docs.microsoft.com/en-us/azure/azure-sql/database/automatic-tuning-overview

Question 48: Skipped
ADatum Corporation is a retailer that sells products through two sales channels: retail stores and a website.

Existing Environment -

ADatum has one database server that has Microsoft SQL Server 2016 installed. The server hosts three mission-critical databases named SALESDB, DOCDB, and REPORTINGDB.

SALESDB collects data from the store and the website.

DOCDB stored documents that connect to the sales data in SALESDB. The documents are stored in two different JSON formats based on the sales channel.

REPORTINGDB stores reporting data and contains server columnstore indexes. A daily process creates reporting data in REPORTINGDB from the data in SALESDB. The process is implemented as a SQL Server Integration Services (SSIS) package that runs a stored procedure from SALESDB.

Requirements -

Planned Changes -

ADatum plans to move the current data infrastructure to Azure. The new infrastructure has the following requirements:

Migrate SALESDB and REPORTINGDB to an Azure SQL database.

Migrate DOCDB to Azure Cosmos DB.

The sales data including the documents in JSON format, must be gathered as it arrives and analyzed online by using Azure Stream Analytics. The analytic process will perform aggregations that must be done continuously, without gaps, and without overlapping.

As they arrive, all the sales documents in JSON format must be transformed into one consistent format.

Azure Data Factory will replace the SSIS process of copying the data from SALESDB to REPORTINGDB.

Technical Requirements -

The new Azure data infrastructure must meet the following technical requirements:

Data in SALESDB must encrypted by using Transparent Data Encryption (TDE). The encryption must use your own key.

SALESDB must be restorable to any given minute within the past three weeks.

Real-time processing must be monitored to ensure that workloads are sized properly based on actual usage patterns.

Missing indexes must be created automatically for REPORTINGDB.

Disk IO, CPU, and memory usage must be monitored for SALESDB.

Question Which counter should you monitor for real-time processing to meet the technical requirements?

- ○

  Concurrent users

- ○

SU% Utilization

(Correct)

○

Data Conversion Errors

○

CPU % utilization

Explanation

The SU % utilization metric, which ranges from 0% to 100%, describes the memory consumption of your workload. For a streaming job with minimal footprint, this metric is usually between 10% to 20%. If SU% utilization is high (above 80%), or if input events get backlogged (even with a low SU% utilization since it does not show CPU usage), your workload likely requires more compute resources, which requires you to increase the number of SUs.

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-streaming-unit-consumption

Question 49: Skipped
ADatum Corporation is a retailer that sells products through two sales channels: retail stores and a website.

Existing Environment -

ADatum has one database server that has Microsoft SQL Server 2016 installed. The server hosts three mission-critical databases named SALESDB, DOCDB, and REPORTINGDB.

SALESDB collects data from the store and the website.

DOCDB stored documents that connect to the sales data in SALESDB. The documents are stored in two different JSON formats based on the sales channel.

REPORTINGDB stores reporting data and contains server columnstore indexes. A daily process creates reporting data in REPORTINGDB from the data in SALESDB. The process is implemented as a SQL Server Integration Services (SSIS) package that runs a stored procedure from SALESDB.

Requirements -

Planned Changes -

ADatum plans to move the current data infrastructure to Azure. The new infrastructure has the following requirements:

Migrate SALESDB and REPORTINGDB to an Azure SQL database.

Migrate DOCDB to Azure Cosmos DB.

The sales data including the documents in JSON format, must be gathered as it arrives and analyzed online by using Azure Stream Analytics. The analytic process will perform aggregations that must be done continuously, without gaps, and without overlapping.

As they arrive, all the sales documents in JSON format must be transformed into one consistent format.

Azure Data Factory will replace the SSIS process of copying the data from SALESDB to REPORTINGDB.

Technical Requirements -

The new Azure data infrastructure must meet the following technical requirements:

Data in SALESDB must encrypted by using Transparent Data Encryption (TDE). The encryption must use your own key.

SALESDB must be restorable to any given minute within the past three weeks.

Real-time processing must be monitored to ensure that workloads are sized properly based on actual usage patterns.

Missing indexes must be created automatically for REPORTINGDB.

Disk IO, CPU, and memory usage must be monitored for SALESDB.

Question You need to configure a disaster recovery solution for SALESDB to meet the technical requirements.

What should you configure in the backup policy?

- ○

  weekly long-term retention backups that are retained for three weeks

- ○

  failover groups

- ○

  a point-in-time restore

    (Correct)

- ○

geo-replication

Explanation

The Azure SQL Database service protects all databases with an automated backup system. These backups are retained for 7 days for Basic, 35 days for Standard and 35 days for Premium.

All Basic, Standard, and Premium databases are protected by automatic backups. Full backups are taken every week, differential backups every day, and log backups every 5 minutes.

This improvement alone may justify upgrading to one of the new service tiers. Backup and restore is lower cost (there are no charges for backups unless they are excessive, whereas you are charged for the database copy needed to ensure a transactional consistent export and for storing the BACPAC file), zero admin (backups are automatic, whereas you must manage or schedule exports yourself) and has a better RPO (**you can restore to a specific point in time with much finer (one minute) granularity** than is practical with copy/export/import) and a better recovery time (restore from backups is typically much faster than import, which involves steps to create the schema, disable indexes, load data, and then enable indexes for each table individually).

https://azure.microsoft.com/en-us/blog/azure-sql-database-point-in-time-restore/

Question 50: Skipped
You have an Azure Storage account and a data warehouse in Azure Synapse Analytics in the UK region.

You need to copy blob data from the storage account from the storage accounts to the data warehouse by using Azure Data Factory. The solution must meet the following requirements.

1.Ensure that the data remains in the UK South region at all times

2.Minimize administrative efforts

What type of Integration run time should you use?

- ○

  Azure-SSIS integration runtime

- ○

  Self-hosted integration runtime

- ○

  Azure integration runtime

Explanation

If you have strict data compliance requirements and need ensure that data do not leave a certain geography, you can explicitly create an Azure IR in a certain region and point the Linked Service to this IR using ConnectVia property. For example, if you want to copy data from Blob in UK South to Azure Synapse Analytics in UK South and want to ensure data do not leave UK, create an Azure IR in UK South and link both Linked Services to this IR.

https://docs.microsoft.com/en-us/azure/data-factory/concepts-integration-runtime

Question 51: Skipped
You have 2 Azure Data Factory Pipelines.

PipelineA has 4 activities as shown in the following exhibit.

Larger image

PipelineB has 2 activities as shown in the following exhibit.

Larger image

You create an alert for the data factory that uses Failed pieline runs metrics with following setting

Operator : Greater than

Aggregation Type : Total

Threshold value : 2

Aggregation granularity (Period) :5 minutes

Frequency of evaluation : Every 5 minutes

Data factoring monitoring records the failures shown in the following table

Larger image

Select all that is True:

- ☐

    An alert notification was sent after the failure of Activity1 on PipelineA

- ☐

  An alert notification was sent after the failure of Activity3 on PipelineA

- ☐

  An alert notification was sent after the failure of Activity1 on PipelineB

- ☐

  None of the above are TRUE

    (Correct)

Explanation

Refer how alerts are created in failed pipeline runs below:

https://azure.microsoft.com/en-in/blog/create-alerts-to-proactively-monitor-your-data-factory-pipelines/

There are many filters/criteria that needs to be defined and some are Activity Name and Pipeline Name and if they are left default then all Activity and Pipelines are used.

The answer should be very simple because at any timestamp in the last 5 minutes we have either 1 or 2 activity failure across both(aggregate) pipelines which is equal to the threshold and not greater than hence no alert will be triggered

Question 52: Skipped
You have a table dbo.Customers with column named Email containing the email ids of customer. You want to mask the column such that email appears as aXXX@XXXX.com

- ○

  From the Azure portal, set a mask on the Email

- ○

  From the Azure portal, set a sensitivity classification of Confidential for the Email column

- ○

  From Microsoft SQL Sever Management Studio, grant SELECT permission to the users for all the columns in the dbo.Customers table except Email

- ○

    From Microsoft SQL Sever Management Studio, set an email mask on the Email column

    (Correct)

Explanation

To get email masking , we have a special mask called **"Email" masking** which can be set from the portal or SQL Management Studio.

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

Question 53: Skipped

You need to implement an Azure Databricks cluster that automatically connects to Azure Data Lake Storage Gen2 by using Azure Active Directory(Azure AD) integration.

How should you configure a new cluster for Cluster Mode?

- ○

    Premium

- ○

    Data Engineering Light

- ○

    Standard

    (Correct)

Explanation

Azure Databricks supports 3 cluster modes:

1. High Concurrency

2. Standard

3. Single Node

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

Question 54: Skipped

You need to implement an Azure Databricks cluster that automatically connects to Azure Data Lake Storage Gen2 by using Azure Active Directory(Azure AD) integration.

How should you configure a new cluster for Advanced option to enable?

- ○

  Azure Data Lake Storage Credential Passthrough

  (Correct)

- ○

  Table Access Controls

Explanation

You can authenticate automatically to Azure Data Lake Storage Gen1 and Azure Data Lake Storage Gen2 from Azure Databricks clusters using the same Azure Active Directory (Azure AD) identity that you use to log into Azure Databricks. When you enable your cluster for Azure Data Lake Storage credential passthrough, commands that you run on that cluster can read and write data in Azure Data Lake Storage without requiring you to configure service principal credentials for access to storage.

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

Practice Test 3

Question 1: Skipped
While provisioning a database, one of the requirement is to have control over the timings of maintenance and patching, which database option will suit this requirement?

- ○

  SQL Server in VM

  (Correct)

- ○

  Single Database

○

Elastic Database

Explanation

SQL Server in VM is an IAAS option that gives full control over the database engine. You could control the timing of maintenance and patching of your database engine. You could also pause or stop the virtual machine whenever you don't need the SQL database to save some costs.

Question 2: Skipped
What is a collection of single databases with a shared set of resources such as CPU or memory called?

○

Single Database

○

Elastic Pool

(Correct)

○

Managed Instance

Explanation

An elastic pool is a collection of single databases with a shared set of resources, such as CPU or memory. The elastic pool has a set of resources assigned to it and we can create multiple databases within that elastic pool that can then share that bucket of resources assigned to the elastic pool.

Question 3: Skipped
Which of this is NOT a DTU Purchasing model?

○

Basic

○

Standard

- ○

Business Critical

(Correct)

- ○

Premium

Explanation

DTU purchasing model has 3 service tiers- Basic, Standard, Premium

Question 4: Skipped
One new member joined in your DB team, he is able to login in to Azure database but not able to access data from any of table, neither he is able to see any other DB objects like SP or Views.

You realized that the problem is with:

- ○

Authorization

(Correct)

- ○

Authentication

- ○

Row level security

- ○

Dynamic Data Masking

Explanation

Authorization refers to the permissions assigned to a user within an Azure SQL Database, and determines what the user is allowed to do. Permissions are controlled by adding user accounts to database roles and assigning database-level permissions to those roles or by granting the user certain object-level permissions.

Question 5: Skipped

Your SQL database is too small and needs to be scaled up. Which of the following units measures scale in a SQL database?

- ○

  DWUs

- ○

  DTUs

  (Correct)

- ○

  RUs

- ○

  SUs

Explanation

A DTU (Database Throughput Unit) is a unit of measurement used to determine scale in a SQL database

Question 6: Skipped
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?

- ○

  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.

- ○

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

  (Correct)

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.

Question 7: Skipped
You currently have multiple databases on Azure SQL Database and are looking to cut costs. Which version of SQL Database would offer the most opportunity to cut costs?

- ○

  Elastic Pool

  (Correct)

- ○

  Single Database

- ○

  Managed Instance

- ○

  General Purpose

Explanation

Elastic allows you to share resources between SQL databases so you can better utilize CPU, etc., and cut costs.

Question 8: Skipped

Which of the following terms refers to the compute scale that's used in a data warehouse in Azure Synapse Analytics?

- ○

  RU

- ○

  DTU

- ○

  DWU

  (Correct)

- ○

  SU

Explanation

DWU refers to a data warehouse unit. It is the measure of compute scale that is assigned to a data warehouse in Azure Synapse Analytics.

Question 9: Skipped

Which of the following are the appropriate steps for optimizing SQL Data Warehouse?

- ○

  Utilize PolyBase, choose the correct distribution type, and maximize transaction sizes

- ○

  Maximize partitioning, Use CREATE TABLE AS SELECT, and minimize column sizes

- ○

  Don't over-partition, utilize PolyBase, and choose the correct distribution type

○

Minimize column sizes, use round-robin distribution, and break gzip into 60+ files

Question 10: Skipped

Mike is the data engineer for Contoso and has a data warehouse created with a database named Crystal. Within the database is a table named DimSuppliers. The suppliers' information is stored in a single text file named Suppliers.txt and is 1,200 MB in size. It's currently stored in a container with an Azure blob store. Your Azure Synapse Analytics is configured as Gen 2 DW30000c. How can Mike maximize the performance of the data load?

○

Increase the Gen 2 DWU units.

○

Split the text file into 60 files of 20 MB each.

(Correct)

○

Use Gen 1 DW6000.

Explanation

That's the correct answer. Separating the single text file of Suppliers.txt into 60 files can take advantage of the fact that Gen 2 DW30000c uses 60 compute nodes and the parallelism of the data load can be evenly spread for quicker performance

Question 11: Skipped

Note: This question is part of series of questions that present the same scenario. Each question in the series contains a unique solution. Determine whether the solution meets the stated goals.

You develop a data ingestion process that will import data to a Microsoft Azure SQL Data Warehouse. The data to be ingested resides in parquet files stored in an

Azure Data Lake Gen 2 storage account.

You need to load the data from the Azure Data Lake Gen 2 storage account into the Azure SQL Data Warehouse.

Solution:

1. Create an external data source pointing to the Azure storage account

2. Create a workload group using the Azure storage account name as the pool name

3. Load the data using the INSERT"¦SELECT statement

Does the solution meet the goal?

- ○

  Yes

- ○

  No

  (Correct)

Explanation

You need to create an external file format and external table using the external data source. You then load the data using the CREATE TABLE AS SELECT statement.

Question 12: Skipped
Note: This question is part of series of questions that present the same scenario. Each question in the series contains a unique solution. Determine whether the solution meets the stated goals.

You develop a data ingestion process that will import data to a Microsoft Azure SQL Data Warehouse. The data to be ingested resides in parquet files stored in an

Azure Data Lake Gen 2 storage account.

You need to load the data from the Azure Data Lake Gen 2 storage account into the Azure SQL Data Warehouse.

Solution:

1. Create an external data source pointing to the Azure storage account

2. Create an external file format and external table using the external data source

3. Load the data using the INSERT"¦SELECT statement

Does the solution meet the goal?

- ○

  Yes

- ○

No

(Correct)

Explanation

You load the data using the CREATE TABLE AS SELECT statement.

Question 13: Skipped
Note: This question is part of series of questions that present the same scenario. Each question in the series contains a unique solution. Determine whether the solution meets the stated goals.

You develop a data ingestion process that will import data to a Microsoft Azure SQL Data Warehouse. The data to be ingested resides in parquet files stored in an

Azure Data Lake Gen 2 storage account.

You need to load the data from the Azure Data Lake Gen 2 storage account into the Azure SQL Data Warehouse.

Solution:

1. Create an external data source pointing to the Azure storage account

2. Create a workload group using the Azure storage account name as the pool name

3. Load the data using the CREATE TABLE AS SELECT statement

Does the solution meet the goal?

- ○

Yes

- ○

No

(Correct)

Explanation

You need to create external file format and external table

Question 14: Skipped
Note: This question is part of series of questions that present the same scenario. Each question in the series contains a unique solution. Determine whether the solution meets the stated goals.

You develop a data ingestion process that will import data to a Microsoft Azure SQL Data Warehouse. The data to be ingested resides in parquet files stored in an

Azure Data Lake Gen 2 storage account.

You need to load the data from the Azure Data Lake Gen 2 storage account into the Azure SQL Data Warehouse.

Solution:

1. Use Azure Data Factory to convert the parquet files to CSV files

2. Create an external data source pointing to the Azure storage account

3. Create an external file format and external table using the external data source

4. Load the data using the INSERT"¦SELECT statement

Does the solution meet the goal?

○

Yes

○

No

(Correct)

Explanation

There is no need to convert the parquet files to CSV files. You load the data using the CREATE TABLE AS SELECT statement.

Question 15: Skipped
You are as a Data Engineer wants to monitor ongoing database activities, as well as analyze and investigate historical activity to identify potential threats or suspected abuse and security violations.

Which functionality you will use?

○

Auditing

(Correct)

○

Threat Protection

○

Vulnerability Assessment

Question 16: Skipped

A company has a SaaS solution that uses Azure SQL Database with elastic pools. The solution contains a dedicated database for each customer organization.

Customer organizations have peak usage at different periods during the year.

You need to implement the Azure SQL Database elastic pool to minimize cost.

Which option or options should you configure?

○

Number of transactions only

○

eDTUs per database only

○

Number of databases only

○

CPU usage only

○

eDTUs and max data size

(Correct)

Question 17: Skipped

A company has a SaaS solution that uses Azure SQL Database with elastic pools. The solution contains a dedicated database for each customer organization. Customer organizations have peak usage at different periods during the year. You need to implement the Azure SQL Database elastic pool to minimize cost. Which option or options should you configure?

- ○

  Number of transactions only

- ○

  eDTUs per database only

- ○

  Number of databases only

- ○

  CPU usage only

- ○

  eDTUs and max data size

    (Correct)

Explanation

The best size for a pool depends on the aggregate resources needed for all databases in the pool.

Question 18: Skipped
A company plans to use Azure SQL Database to support a mission-critical application.

The application must be highly available without performance degradation during maintenance windows.

You need to implement the solution.

Which three technologies should you implement? Each correct answer presents part of the solution.

- ☐

  Premium service tier

    (Correct)

- ☐

  Virtual machine Scale Sets

- ☐

  Basic service tier

- ☐

  SQL Data Sync

- ☐

  Always On availability groups

  (Correct)

- ☐

  Zone-redundant configuration

  (Correct)

Question 19: Skipped

A company plans to use Azure Storage for file storage purposes. Compliance rules require:

☞ A single storage account to store all operations including reads, writes and deletes

☞ Retention of an on-premises copy of historical operations

You need to configure the storage account.

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

- ☐

  Configure the storage account to log read, write and delete operations for service type Blob

  (Correct)

- ☐

    Use the AzCopy tool to download log data from $logs/blob

    (Correct)

- ☐

    Configure the storage account to log read, write and delete operations for service-type table

- ☐

    Use the storage client to download log data from $logs/table

- ☐

    Configure the storage account to log read, write and delete operations for service type queue

Explanation

Storage Logging logs request data in a set of blobs in a blob container named $logs in your storage account. This container does not show up if you list all the blob containers in your account but you can see its contents if you access it directly.

To view and analyze your log data, you should download the blobs that contain the log data you are interested in to a local machine. Many storage-browsing tools enable you to download blobs from your storage account; you can also use the Azure Storage team provided command-line Azure Copy Tool (AzCopy) to download your log data.

References:

https://docs.microsoft.com/en-us/rest/api/storageservices/enabling-storage-logging-and-accessing-log-data

Question 20: Skipped
You have an Azure SQL database named DB1 that contains a table named Table1. Table1 has a field named Customer_ID that is varchar(22).

You need to implement masking for the Customer_ID field to meet the following requirements:

✎ **The first two prefix characters must be exposed.**

✎ **The last four prefix characters must be exposed.**

✎ **All other characters must be masked.**

Solution: You implement data masking and use a credit card function mask.

Does this meet the goal?

- ○

  Yes

- ○

  No

  (Correct)

Explanation

Must use Custom Text data masking, which exposes the first and last characters and adds a custom padding string in the middle.

Question 21: Skipped

You are responsible for providing access to an Azure Data Lake Storage Gen2 account.

Your user account has contributor access to the storage account, and you have the application ID access key.

You plan to use PolyBase to load data into Azure SQL data warehouse.

You need to configure PolyBase to connect the data warehouse to the storage account.

Which three components should you create in sequence? To answer, move the appropriate components from the list of components to the answer are and arrange them in the correct order.

Larger image

- ○

  a database scoped credential
  an external file format
  an external data source

- ○

  a database scoped credential
  an external data source
  an external file format

(Correct)

○

an encryption key
an external data source
an external file format

Question 22: Skipped
You develop data engineering solutions for a company.

You need to ingest and visualize real-time Twitter data by using Microsoft Azure.

Which three technologies should you use?

☐

Event Grid subscription

☐

Event Grid topic

☐

Azure Stream Analytics Job that queries Twitter data from an Event Hub

(Correct)

☐

Azure Stream Analytics Job that queries Twitter data from an Event Grid

☐

Logic App that sends Twitter posts which have target keywords to Azure

(Correct)

☐

Event Hub instance

(Correct)

Explanation

Azure Logic Apps is a cloud service that helps you schedule, automate, and orchestrate tasks, business processes, and workflows when you need to integrate apps, data, systems, and services across enterprises or organizations. With businesses moving toward digitization, logic apps help you connect legacy, modern, and cutting-edge systems more easily and quickly by providing prebuilt APIs as Microsoft-managed connectors.

https://docs.microsoft.com/en-us/azure/logic-apps/logic-apps-overview

Azure Event Hubs is a big data streaming platform and event ingestion service. It can receive and process millions of events per second.

https://docs.microsoft.com/en-us/azure/event-hubs/event-hubs-about

Azure Stream Analytics is a real-time analytics and complex event-processing engine that is designed to analyze and process high volumes of fast streaming data from multiple sources simultaneously.

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-introduction

Question 23: Skipped
You develop data engineering solutions for a company. A project requires the deployment of data to Azure Data Lake Storage.

You need to implement role-based access control (RBAC) so that project members can manage the Azure Data Lake Storage resources.

Which three actions should you perform?

- ☐

  Configure access control lists (ACL) for the Azure Data Lake Storage account.

  (Correct)

- ☐

  Configure service-to-service authentication for the Azure Data Lake Storage account.

- ☐

  Create security groups in Azure Active Directory (Azure AD) and add project members.

  (Correct)

- ☐

  Configure end-user authentication for the Azure Data Lake Storage account.

- ☐

  Assign Azure AD security groups to Azure Data Lake Storage.

    (Correct)

Explanation

You can create a basic group using the Azure Active Directory (Azure AD) portal.

https://docs.microsoft.com/en-us/azure/active-directory/fundamentals/active-directory-groups-create-azure-portal.

We recommend that you define ACLs for multiple users by using security groups. Add users to a security group, and then assign the ACLs for a file or folder to that security group.

https://docs.microsoft.com/en-us/azure/data-lake-store/data-lake-store-security-overview

Question 24: Skipped
You have a container named Sales in an Azure Cosmos DB database. Sales has 120 GB of data. Each entry in Sales has the following structure.

Larger image

The partition key is set to the OrderId attribute.

Users report that when they perform queries that retrieve data by ProductName, the queries take longer than expected to complete.

You need to reduce the amount of time it takes to execute the problematic queries.

Solution: You create a lookup collection that uses ProductName as a partition key.

Does this meet the goal?

- ○

  Yes

    (Correct)

- ○

No

Explanation

One option is to have a lookup collection "ProductName" for the mapping of "ProductName" to "OrderId" with ProductName as the partition key and row key and "OrderId" as the value

https://azure.microsoft.com/sv-se/blog/azure-cosmos-db-partitioning-design-patterns-part-1/

Question 25: Skipped
You have a container named Sales in an Azure Cosmos DB database. Sales has 120 GB of data. Each entry in Sales has the following structure

Larger image

The partition key is set to the OrderId attribute.

Users report that when they perform queries that retrieve data by ProductName, the queries take longer than expected to complete.

You need to reduce the amount of time it takes to execute the problematic queries.

Solution: You change the partition key to include ProductName.

Does this meet the goal?

- ○

  Yes

- ○

  No

  (Correct)

Explanation

One option is to have a lookup collection "ProductName" for the mapping of "ProductName" to "OrderId" with ProductName as the partition key and row key and "OrderId" as the value

https://azure.microsoft.com/sv-se/blog/azure-cosmos-db-partitioning-design-patterns-part-1/

Question 26: Skipped
You develop data engineering solutions for a company.

A project requires the deployment of resources to Microsoft Azure for batch data processing on Azure HDInsight.

Batch processing will run daily and must:

Scale to minimize costs

Be monitored for cluster performance

You need to recommend a tool that will monitor clusters and provide information to suggest how to scale.

Solution: Monitor cluster load using the Ambari Web UI. Does the solution meet the goal?

- ○

  Yes

- ○

  No

    (Correct)

Explanation

Learn how to enable Azure Monitor logs to monitor Hadoop cluster operations in HDInsight. And how to add an HDInsight monitoring solution.

HDInsight provides cluster-specific management solutions that you can add for Azure Monitor logs. Management solutions add functionality to Azure Monitor logs, providing additional data and analysis tools. These solutions collect important performance metrics from your HDInsight clusters. And provide the tools to search the metrics. These solutions also provide visualizations and dashboards for most cluster types supported in HDInsight. By using the metrics that you collect with the solution, you can create custom monitoring rules and alerts.

https://azure.microsoft.com/en-in/blog/monitoring-on-hdinsight-part-1-an-overview/

Question 27: Skipped
You develop data engineering solutions for a company.

A project requires the deployment of resources to Microsoft Azure for batch data processing on Azure HDInsight.

Batch processing will run daily and must:

Scale to minimize costs

Be monitored for cluster performance

You need to recommend a tool that will monitor clusters and provide information to suggest how to scale. Solution: Monitor clusters by using Azure Log Analytics and HDInsight cluster management solutions.

Does the solution meet the goal?

- ○

Yes

(Correct)

- ○

No

Explanation

Learn how to enable Azure Monitor logs to monitor Hadoop cluster operations in HDInsight. And how to add an HDInsight monitoring solution.

HDInsight provides cluster-specific management solutions that you can add for Azure Monitor logs. Management solutions add functionality to Azure Monitor logs, providing additional data and analysis tools. These solutions collect important performance metrics from your HDInsight clusters. And provide the tools to search the metrics. These solutions also provide visualizations and dashboards for most cluster types supported in HDInsight. By using the metrics that you collect with the solution, you can create custom monitoring rules and alerts.

https://azure.microsoft.com/en-in/blog/monitoring-on-hdinsight-part-1-an-overview/

Question 28: Skipped
You develop data engineering solutions for a company.

A project requires the deployment of resources to Microsoft Azure for batch data processing on Azure HDInsight. Batch processing will run daily and must:

Scale to minimize costs

Be monitored for cluster performance

You need to recommend a tool that will monitor clusters and provide information to suggest how to scale.

Solution: Download Azure HDInsight cluster logs by using Azure PowerShell.

Does the solution meet the goal?

- ○

Yes

- ○

No

Explanation

Learn how to enable Azure Monitor logs to monitor Hadoop cluster operations in HDInsight. And how to add an HDInsight monitoring solution.

HDInsight provides cluster-specific management solutions that you can add for Azure Monitor logs. Management solutions add functionality to Azure Monitor logs, providing additional data and analysis tools. These solutions collect important performance metrics from your HDInsight clusters. And provide the tools to search the metrics. These solutions also provide visualizations and dashboards for most cluster types supported in HDInsight. By using the metrics that you collect with the solution, you can create custom monitoring rules and alerts.

https://azure.microsoft.com/en-in/blog/monitoring-on-hdinsight-part-1-an-overview/

Question 29: Skipped
You are monitoring an Azure Stream Analytics job.

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

You need to ensure that the job can handle all the events. What should you do?

- ○

  Create an additional output stream for the existing input stream.

- ○

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

- ○

  Increase the number of streaming units (SUs).

  (Correct)

- ○

  Change the compatibility level of the Stream Analytics job.

Explanation

Streaming Units (SUs) represents the computing resources that are allocated to execute a Stream Analytics job. The higher the number of SUs, the more CPU and memory resources are allocated for your job.

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-streaming-unit-consumption

Question 30: Skipped

A company is designing a hybrid solution to synchronize data and on-premises Microsoft SQL Server database to Azure SQL Database.

You must perform an assessment of databases to determine whether data will move without compatibility issues.

You need to perform the assessment.

Which tool should you use?

- ○ SQL Vulnerability Assessment (VA)

- ○ Azure SQL Data Sync

- ○ Data Migration Assistant (DMA)

   (Correct)

- ○ SQL Server Migration Assistant (SSMA)

- ○ Microsoft Assessment and Planning Toolkit

Explanation

The Data Migration Assistant (DMA) helps you upgrade to a modern data platform by detecting compatibility issues that can impact database functionality in your new version of SQL Server or Azure SQL Database. DMA recommends performance and reliability improvements for your target environment and allows you to move your schema, data, and uncontained objects from your source server to your target server.

https://docs.microsoft.com/en-us/sql/dma/dma-overview?view=sql-server-ver15

Question 31: Skipped

A company plans to use Azure Storage for file storage purposes. Compliance rules require:

A single storage account to store all operations including reads, writes and deletes

Retention of an on-premises copy of historical operations

You need to configure the storage account. Which two actions should you perform? Each correct answer presents part of the solution.

- ☐

    Configure the storage account to log read, write and delete operations for service type queue

- ☐

    Configure the storage account to log read, write and delete operations for service-type table

- ☐

    Use the storage client to download log data from $logs/table

- ☐

    Configure the storage account to log read, write and delete operations for service type Blob

    (Correct)

- ☐

    Use the AzCopy tool to download log data from $logs/blob

    (Correct)

Explanation

All logs are stored in block blobs in a container named `$logs`, which is automatically created when Storage Analytics is enabled for a storage account. The `$logs` container is located in the blob namespace of the storage account, for example: `http://<accountname>.blob.core.windows.net/$logs`

https://docs.microsoft.com/en-us/azure/storage/common/storage-analytics-logging?tabs=dotnet

Question 32: Skipped

Your team has created a new Azure Data Factory environment. You have to analyse the pipeline executions. Trends need to be identified in execution duration over the past 30 days. You need to create a solution that would ensure that data can be queried from Azure Log Analytics.

Which of the following would you choose as the Log type when setting up the diagnostic setting for Azure Data Factory?

- ○

  ActivityRuns

- ○

  AllMetrics

- ○

  PipelineRuns

  (Correct)

- ○

  TriggerRuns

Explanation

Installing this solution creates a default set of views inside the workbooks section of the chosen Log Analytics workspace. As a result, the following metrics become enabled:

ADF Runs - 1) Pipeline Runs by Data Factory

ADF Runs - 2) Activity Runs by Data Factory

ADF Runs - 3) Trigger Runs by Data Factory

ADF Errors - 1) Top 10 Pipeline Errors by Data Factory

ADF Errors - 2) Top 10 Activity Runs by Data Factory

ADF Errors - 3) Top 10 Trigger Errors by Data Factory

ADF Statistics - 1) Activity Runs by Type

ADF Statistics - 2) Trigger Runs by Type

ADF Statistics - 3) Max Pipeline Runs Duration

Question 33: Skipped

You are working as a data engineer for a company that has an Azure SQL database. Your company wants to create an offline exported copy of the database. This is done because the company wants to ensure that the users can work with the data offline when they don't have any Internet connection on their laptops. What ways can be used to create the exported copy? Select 3 answers from the following.

- ☐

  Export to a BACPAC file by using Azure Cloud Shell and save the file to a storage account.

- ☐

  Export to a BACPAC file by using SQL Server Management Studio. Save the file to a storage account.

  (Correct)

- ☐

  Export to a BACPAC file by using the Azure portal

- ☐

  Export to a BACPAC file by using Azure PowerShell and save the file locally

  (Correct)

- ☐

  Export to a BACPAC file by using the SqlPackage utility

  (Correct)

Explanation

As per Microsoft docs, there are essentially 4 ways to achieve this , however we can select only 3 answers.

Azure portal is also correct answer.

https://docs.microsoft.com/en-us/azure/azure-sql/database/database-export

Question 34: Skipped

You are working as a data engineer for your company and you have been given the task to develop a solution which would perform the following things:

· Ingest twitter-based data into Azure

· Give the ability to visualize real-time twitter data

What will you be using to implement this solution? Select 3 options from the following.

- ☐

    Make use of an Event Grid Topic

- ☐

    Make use of Azure Stream Analytics to query twitter data from an Event Hub

    (Correct)

- ☐

    Make use of Azure Stream Analytics to query twitter data from an Event Grid

- ☐

    Have a Logic App in place that would send twitter data to Azure

    (Correct)

- ☐

    Create an Event Grid subscription

- ☐

    Create an Event Hub Instance

    (Correct)

Explanation

Find end to end twitter sentiment analysis using Azure event hub and Stream Analytics.

Question 35: Skipped

You are a data engineer and you are currently performing a task and configuring a new rule to trigger an alert in Azure Monitor when the number of the failed connections exceeds 30 within a 10-minute interval.

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

What action type should you be choosing?

- ○

  Automation Runbook

- ○

  Azure Functions

- ○

  IT Service Management Connector (ITSMC)

  (Correct)

- ○

  Push Notification

Explanation

The IT Service Management (ITSM) Connector is for connecting Azure and a supported IT Service Management (ITSM) product/service. However, Azure services like Log Analytics and Azure Monitor provide tools for detecting, analyzing and troubleshooting issues with your Azure and non-Azure resources. but, the work items related to an issue depends on an ITSM product/service.

https://www.testpreptraining.com/tutorial/connecting-azure-with-itsm-tools-using-itsm-connector/

Question 36: Skipped

In below given scenario, you as a data engineer has to choose the right partition key.

Your company has a globally distributed application which spread across five Azure regions. This application is storing millions of documents in Cosmos DB, and

performing thousands of read and write operaitons per second, that also result a very high throughput.

There are two important fields.
Region field: which stores the original creation region.
Sensor Id: Which is unique across the database.

This application usually query to filter data by region and sensor id

Which partition key would you choose?

- ○

  timestamp with random suffix

- ○

  region

- ○

  region with pre-calculated suffix based on sensorId

  (Correct)

- ○

  sensorId

Explanation

It's the best practice to have a partition key with many distinct values, such as hundreds or thousands. The goal is to distribute your data and workload evenly across the items associated with these partition key values. If such a property doesn't exist in your data, you can construct a *synthetic partition key*.

https://docs.microsoft.com/en-us/azure/cosmos-db/synthetic-partition-keys

Question 37: Skipped
You are working as a data engineer in a company and you are given to manage an Azure SQL Data Warehouse. It has 4 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 4 tables use hash distribution and queries join the DimProduct, DimCustomer, and FactSales tables. The FactOrders table contains all the data it needs.

You are given the task to optimize these tables for performance. Select all that is True.

- ☐

   You should use a replicated table for DimProduct

      (Correct)

- ☐

   You should use a replicated table for DimCustomer.

      (Correct)

- ☐

   You should use a replicated table for FactSales.

- ☐

   You should use round-robin distribution for FactOrders

      (Correct)

Explanation

Hash-distributed tables improve query performance on large fact tables, and are the focus of this article.

Round-robin tables are useful for improving loading speed. These design choices have a significant impact on improving query and loading performance.

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

A replicated table has a full copy of the table accessible on each Compute node. Replicating a table removes the need to transfer data among Compute nodes before a join or aggregation.

https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/design-guidance-for-replicated-tables

Question 38: Skipped
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?

- ○

  Run the New-AzSqlServerFirewallRule PowerShell cmdlet.

  (Correct)

- ○

  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

New-AzSqlServerFirewallRule -Creates a new server-level firewall rule

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

Question 39: Skipped
Your company is using Azure SQL single database deployment option. A new colleague named Saurabh joined in to your team. Your manager asked you to give him administrative rights so that Saurabh can also add or remove other users.

Can you complete below T-SQL command for this?

<<< A >>> <<< B >>> <<< C >>> ADD MEMBER Sam

A Options

1. Alter

2. Create

3. Grant

B Options

1. Login

2. Role

3. User

C Options

1. Db_datareader

2. Db_datawriter

3. Db_owner

- ○

  Alter, Login and Db_datareader

- ○

  Alter, Role and Db_owner

  (Correct)

- ○

  Alter, USer and Db_datawriter

- ○

  Create, Role and Db-datawriter

Explanation

-- Syntax for SQL Server

ALTER SERVER ROLE server_role_name

{

   [ADD MEMBER server_principal]

 | [DROP MEMBER server_principal]

 | [WITH NAME = new_server_role_name]

} [;]

https://docs.microsoft.com/en-us/sql/t-sql/statements/alter-server-role-transact-sql?view=sql-server-ver15

Question 40: Skipped

You 're working as a data engineer for your company. You are creating an Azure SQL Database server called autoone. You want to use two databases on this server. You want to allow both databases to share the server 's resources.

How are you supposed to complete the commands? Select the relevant options to answer.

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

- ○

CosmosDB, CosmosDB database, elastic-pool, CosmosDB and elastic-pool

○

SQL Elastic-Pool, CosmosDB database, zone-redundant, CosmosDB and zone-redundant

○

SQL DB, sql db, elastic-pool, SQL Elastic-Pool, SQL DB and zone-redundant

○

SQL Elastic-Pool, sql db, elastic-pool, SQL DB and elastic-pool

(Correct)

Question 41: Skipped
You have been given the task to show a list of all hash distributed tables. This list should also contain the table name and the column name of the distribution column.

What 3 catalog views should you be needing to join in a query?

☐

sys.columns

(Correct)

☐

sys.pdw_table_distribution_properties

☐

sys.pdw_distributions

☐

sys.pdw_column_distribution_properties

(Correct)

☐

sys.pdw_nodes_columns

- ☐

sys.tables

(Correct)

Question 42: Skipped
You are designing a new Lambda architecture on Microsoft Azure.

The real-time processing layer must meet the following requirements:

Stream processing:

✎ **Process on a per-job basis**

✎ **Provide seamless connectivity with Azure services**

✎ **Use a SQL-based query language**

You need to identify the correct technologies to build the Lambda architecture using minimal effort. Which technology should you use?

- ○

Azure Stream Analytics

(Correct)

- ○

HDInsight with Spark Streaming

- ○

Azure Cosmos DB Change Feed

- ○

Azure Analysis Services

Explanation

An Azure Stream Analytics job consists of an input, query, and an output. Stream Analytics ingests data from Azure Event Hubs (including Azure Event Hubs from Apache Kafka), Azure IoT Hub, or Azure Blob Storage. The query, which is based on SQL query language, can be used to easily filter, sort, aggregate, and join streaming data over a period of time.

Question 43: Skipped
You are designing a new Lambda architecture on Microsoft Azure.

The real-time processing layer must meet the following requirements:

Analytical data store:

⌘ Act as a managed service

⌘ Use a document store

⌘ Provide data encryption at rest

You need to identify the correct technologies to build the Lambda architecture using minimal effort. Which technology should you use?

- ○

  Hive LLAP on HDInsight

- ○

  Azure Aanalysis Services

- ○

  Azure Cosmos DB

  (Correct)

- ○

  SQL Data Warehouse

Explanation

https://docs.microsoft.com/en-us/azure/architecture/data-guide/technology-choices/analytical-data-stores#general-capabilities

Question 44: Skipped
You are designing a new Lambda architecture on Microsoft Azure.

The real-time processing layer must meet the following requirements:

Ingestion:

⌘ Receive millions of events per second

✆ Act as a fully managed Platform-as-a-Service (PaaS) solution

✆ Integrate with Azure Functions

You need to identify the correct technologies to build the Lambda architecture using minimal effort. Which technology should you use for Ingestion?

- ○

    HDInsight Kafka

- ○

    HDInsight Storm

- ○

    Azure Event Hubs

    (Correct)

- ○

    HDInsight Spark

Explanation

This article explains how to work with Azure Event Hubs trigger for Azure Functions. Azure Functions supports trigger and output bindings for Event Hubs.

https://docs.microsoft.com/en-us/azure/azure-functions/functions-bindings-event-hubs-trigger?tabs=csharp

Practice Test 4

Question 1: Skipped
You want to migrate a set of databases from on premises to cloud. You dont want to use an IAAS solution to reduce overhead.Which is the best option from the following?

- ○

    Single Database

○

Elastic Pool

○

Managed Instance

(Correct)

○

SQL database on VMs

Explanation

A managed instance is a set of databases that can be used together. This is the closest deployment option to the IaaS scenario offered by Azure SQL Database. This means easy migration from on-premises databases.

Question 2: Skipped

Your manager called you and assigned a task to configure an automatic high-availability option for recently migrated SQL Database in Azure Virtual machine.

You replied to him that SQL Server in VM can't be entirely configured automatically

is it right?

○

Yes

(Correct)

○

No

Explanation

If you require high availability, consider configuring SQL Server Availability Sets. This involves multiple SQL Server Azure VMs in a virtual network. You can configure your high availability solution manually, or you can use templates in the Azure portal for automatic configuration.

Question 3: Skipped

Which of this is not a vCore purchasing model?

- ○

  General Purpose

- ○

  Hyperscale

- ○

  Business Critical

- ○

  Premium

    (Correct)

Explanation

Vcore purchasing model has 3 service tiers - General Purpose, Hyperscale and Business Critical

Question 4: Skipped

Your company has been managing applications and databases on premises and is looking to modernize their approach and decrease their need for infrastructure management. It is also important that SQL server functionality be retained as much as possible. They would like to complete data transformations in Python and report using Power BI. You plan to recommend a managed-instance SQL warehouse. Is this an appropriate solution

- ○

  Yes

- ○

  No

    (Correct)

Explanation

A managed-instance SQL database, not warehouse, would be the best solution. Video for reference: Overview of SQL Database

Question 5: Skipped
Which encryption functionality make sure that data in motion is always encrypted?

- ○

  Transport Layer Security

  (Correct)

- ○

  Transparent Data Encryption

Explanation

The data in motion is always encrypted using transport layer security or TLS. This ensures all data is encrypted "in transit" between the client and server irrespective of the setting of Encrypt or TrustServerCertificate in the connection string.

Question 6: Skipped
Your company has a SQL Database instance loaded, and your employer is concerned about scaling. You tell your employer that SQL Database is dynamically-scaling and will auto-scale as needed. Is this an appropriate answer?

- ○

  Yes

- ○

  No

  (Correct)

Explanation

It is dynamically scaling but not auto scaling. It needs to be manually scaled

Question 7: Skipped
In Modern Database Architecture which service is most commonly used to Ingest data?

- ○

  Azure Data Factory

(Correct)

○

Azure DataLake

○

Azure Databricks

○

Azure Analysis service

Explanation

Azure Data Factory is a data ingestion service

Question 8: Skipped
You need to choose a sharding pattern for SQL Data Warehouse that offers the highest query performance for large tables. Which choice offers the best solution?

○

Round Robin

○

Hash

(Correct)

○

Replicate

Explanation

Hash is best for large tables that need high query performance.

Question 9: Skipped
You have an Azure Synapse Analytics database. Within the database, you have a dimension table named Stores that contains store information. You have a total of 263 stores nationwide. Store information is retrieved in more than half of the queries that are issued against this database. These queries include staff information per store, sales information per store, and finance information. You want to improve the

query performance of these queries by configuring the table geometry of the Stores table. Which is the appropriate table geometry to select for the Stores table?

- ○

  Round Robin

- ○

  Hash

- ○

  Replicated

  (Correct)

Explanation

A replicated table is an appropriate table geometry choice because the size of the data in the table is less than 2 GB. The table will be replicated to every distribution node of a data warehouse in Azure Synapse Analytics to improve performance.

Question 10: Skipped
You implement an Azure SQL Data Warehouse instance.

You plan to migrate the largest fact table to Azure SQL Data Warehouse. The table resides on Microsoft SQL Server on-premises and is 10 terabytes (TB) is size.

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

Larger image

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

Which technology should you use?

- ○

  Hash distributed table with clustered column Store Index

  (Correct)

- ○

Hash distributed table with clustered index

○

Replicated distribution with heap

○

Round robin distribution with clustered index

Explanation

Hash-distributed tables improve query performance on large fact tables. Columnstore indexes can achieve up to 100x better performance on analytics and data warehousing workloads and up to 10x better data compression than traditional rowstore indexes.

https://docs.microsoft.com/en-us/sql/relational-databases/indexes/columnstore-indexes-query-performance?view=sql-server-ver15

Question 11: Skipped
You are working in a Financial firm as a Data Engineer, and you are asked to label and protect sensitive data in your Data Warehouse, which functionality you will use?

○

Auditing

○

Threat Detection

○

Vulnerability Assessment

○

Data Discovery and Classification

(Correct)

Question 12: Skipped
Your manager asked your recommendation about how we can ensure safely of data in transit between the client and server. You recommend them to enable Transport layer security, is this a right recommendation?

- ○

Yes

- ○

No

    (Correct)

Explanation

The data in motion is always encrypted using transport layer security or TLS. This ensures all data is encrypted "in transit" between the client and server irrespective of the setting of Encrypt or TrustServerCertificate in the connection string.

Question 13: Skipped
You are a data engineer implementing a lambda architecture on Microsoft Azure. You use an open-source big data solution to collect, process, and maintain data. The analytical data store performs poorly.

You must implement a solution that meets the following requirements:

☞ **Provide data warehousing**

☞ **Reduce ongoing management activities**

☞ **Deliver SQL query responses in less than one second**

You need to create an HDInsight cluster to meet the requirements.

Which type of cluster should you create?

- ○

  Apache Hadoop

- ○

  Apache Spark

    (Correct)

- ○

  Apache Hbase

○

Interactive Query

Explanation

Azure offers you a combination of following technologies to accelerate real-time big data analytics:

1. Azure Cosmos DB, a globally distributed and multi-model database service.

2. Apache Spark for Azure HDInsight, a processing framework that runs large-scale data analytics applications.

3. Azure Cosmos DB change feed, which streams new data to the batch layer for HDInsight to process.

4. The Spark to Azure Cosmos DB Connector

Question 14: Skipped
You are a data engineer. You are designing a Hadoop Distributed File System (HDFS) architecture. You plan to use Microsoft Azure Data Lake as a data storage repository. You must provision the repository with a resilient data schema. You need to ensure the resiliency of the Azure Data Lake Storage.

What should you use? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Larger image

- ○

  NamedNode, NameNode, DataNode

- ○

  NameNode,DataNode,DataNode

    (Correct)

- ○

  DataNode,NameNode, DataNode

- ○

DataNode, DataNode,NameNode

Explanation

NameNode -

An HDFS cluster consists of a single NameNode, a master server that manages the file system namespace and regulates access to files by clients.

DataNode -

The DataNodes are responsible for serving read and write requests from the file system's clients.

DataNode -

The DataNodes perform block creation, deletion, and replication upon instruction from the NameNode.

Question 15: Skipped
A company manages several on-premises Microsoft SQL Server databases.

You need to migrate the databases to Microsoft Azure by using a backup process of Microsoft SQL Server.

Which data technology should you use?

- ○

  Azure SQL Database single database

- ○

  Azure SQL Data Warehouse

- ○

  Azure Cosmos DB

- ○

  Azure SQL Database Managed Instance

  (Correct)

Explanation

The managed instance deployment model allows existing SQL Server customers to lift and shift their on-premises applications to the cloud with minimal application and database changes.

Question 16: Skipped
You are developing a data engineering solution for a company. The solution will store a large set of key-value pair data by using Microsoft Azure Cosmos DB.

The solution has the following requirements:

⚭ Data must be partitioned into multiple containers.

⚭ Data containers must be configured separately.

⚭ Data must be accessible from applications hosted around the world.

⚭ The solution must minimize latency.

You need to provision Azure Cosmos DB.

- ☐

    Cosmos account-level throughput.

- ☐

    Provision an Azure Cosmos DB account with the Azure Table API. Enable geo-redundancy.

    (Correct)

- ☐

    Configure table-level throughput.

- ☐

    Replicate the data globally by manually adding regions to the Azure Cosmos DB account.

    (Correct)

- ☐

    Provision an Azure Cosmos DB account with the Azure Table API. Enable multi-region writes.

    (Correct)

Question 17: Skipped

A company has a SaaS solution that uses Azure SQL Database with elastic pools. The solution will have a dedicated database for each customer organization.

Customer organizations have peak usage at different periods during the year.

Which two factors affect your costs when sizing the Azure SQL Database elastic pools? Each correct answer presents a complete solution.

- ☐

  maximum data size

  (Correct)

- ☐

  number of databases

- ☐

  eDTUs consumption

  (Correct)

- ☐

  number of read operations

- ☐

  number of transactions

Question 18: Skipped
You are developing a solution using a Lambda architecture on Microsoft Azure.

The data at rest layer must meet the following requirements:

Data storage:

☞ **Serve as a repository for high volumes of large files in various formats.**

☞ **Implement optimized storage for big data analytics workloads.**

✎ **Ensure that data can be organized using a hierarchical structure.**

Batch processing:

✎ **Use a managed solution for in-memory computation processing.**

✎ **Natively support Scala, Python, and R programming languages.**

✎ **Provide the ability to resize and terminate the cluster automatically.**

Analytical data store:

✎ **Support parallel processing.**

✎ **Use columnar storage.**

✎ **Support SQL-based languages.**

You need to identify the correct technologies to build the Lambda architecture.

Which technologies should you use? To answer, select the appropriate options in the answer area.

Larger image

- ○

  Data storage: Azure Data Lake Store
  Batch processing: Azure Databricks
  Analytic data store: SQL Data Warehouse

  (Correct)

- ○

  Data storage: Azure Data Lake Store
  Batch processing: HDInsight Spark
  Analytic data store: SQL Data Warehouse

- ○

  Data storage: Azure Data Lake Store
  Batch processing: HDInsight Spark
  Analytic data store: Azure Cosmos DB

Question 19: Skipped
You have an Azure SQL database named DB1 that contains a table named Table1.
Table1 has a field named Customer_ID that is varchar(22).
You need to implement masking for the Customer_ID field to meet the following
requirements:
❧ The first two prefix characters must be exposed.
❧ The last four prefix characters must be exposed.
❧ All other characters must be masked.
Solution: You implement data masking and use an email function mask.
Does this meet the goal?

- ○

  Yes

- ○

  No

  (Correct)

Explanation

Must use Custom Text data masking, which exposes the first and last characters and adds
a custom padding string in the middle.

Question 20: Skipped
You have an Azure SQL database named DB1 that contains a table named Table1.
Table1 has a field named Customer_ID that is varchar(22).

You need to implement masking for the Customer_ID field to meet the following
requirements:

❧ The first two prefix characters must be exposed.

❧ The last four prefix characters must be exposed.

❧ All other characters must be masked.

Solution: You implement data masking and use a random number function mask.

Does this meet the goal?

- ○

  Yes

- ○

No

(Correct)

Explanation

Must use Custom Text data masking, which exposes the first and last characters and adds a custom padding string in the middle

Question 21: Skipped
You plan to create a dimension table in Azure Data Warehouse that will be less than 1 GB.

You need to create the table to meet the following requirements:

❧ Provide the fastest query time.

❧ Minimize data movement.

Which type of table should you use?

- ○

  hash distributed

- ○

  heal

- ○

  replicated

  (Correct)

- ○

  round robin

Question 22: Skipped
You plan to create an Azure Databricks workspace that has a tiered structure. The workspace will contain the following three workloads:

A workload for data engineers who will use Python and SQL

A workload for jobs that will run notebooks that use Python, Spark, Scala, and SQL

A workload that data scientists will use to perform ad hoc analysis in Scala and R

The enterprise architecture team at your company identifies the following standards for Databricks environments:

The data engineers must share a cluster.

The job cluster will be managed by using a request process whereby data scientists and data engineers provide packaged notebooks for deployment to the cluster.

All the data scientists must be assigned their own cluster that terminates automatically after 120 minutes of inactivity. Currently, there are three data scientists.

You need to create the Databrick clusters for the workloads. Solution: You create a Standard cluster for each data scientist, a High Concurrency cluster for the data engineers, and a High Concurrency cluster for the jobs.

Does this meet the goal?

○ Yes

   (Correct)

○ No

Explanation

Concurrent Cluster-A High Concurrency cluster is a managed cloud resource. The key benefits of High Concurrency clusters are that they provide Apache Spark-native fine-grained sharing for maximum resource utilization and minimum query latencies.High Concurrency clusters work only for SQL, Python, and R.

Standard Cluster -Standard clusters are recommended for a single user. Standard clusters can run workloads developed in any language: Python, R, Scala, and SQL.

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

Question 23: Skipped

You plan to create an Azure Databricks workspace that has a tiered structure. The workspace will contain the following three workloads:

A workload for data engineers who will use Python and SQL

A workload for jobs that will run notebooks that use Python, Spark, Scala, and SQL

A workload that data scientists will use to perform ad hoc analysis in Scala and R

The enterprise architecture team at your company identifies the following standards for Databricks environments:

The data engineers must share a cluster.

The job cluster will be managed by using a request process whereby data scientists and data engineers provide packaged notebooks for deployment to the cluster.

All the data scientists must be assigned their own cluster that terminates automatically after 120 minutes of inactivity. Currently, there are three data scientists.

You need to create the Databrick clusters for the workloads.

Solution: You create a High Concurrency cluster for each data scientist, a High Concurrency cluster for the data engineers, and a Standard cluster for the jobs.

Does this meet the goal?

- ○

  Yes

    (Correct)

- ○

  No

Explanation

There is no need for concurrent clusters for Data scientists.

Concurrent Cluster-A High Concurrency cluster is a managed cloud resource. The key benefits of High Concurrency clusters are that they provide Apache Spark-native fine-grained sharing for maximum resource utilization and minimum query latencies.High Concurrency clusters work only for SQL, Python, and R.

Standard Cluster -Standard clusters are recommended for a single user. Standard clusters can run workloads developed in any language: Python, R, Scala, and SQL.

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

Question 24: Skipped
You plan to use Microsoft Azure SQL Database instances with strict user access control. A user object must:

Move with the database if it is run elsewhere

Be able to create additional users

You need to create the user object with correct permissions. Which two Transact-SQL commands should you run?

- ☐

  CREATE USER Mary WITH PASSWORD = 'strong_password';

  (Correct)

- ☐

  GRANT ALTER ANY USER TO Mary;

- ☐

  ALTER LOGIN Mary WITH PASSWORD = 'strong_password';

- ☐

  ALTER ROLE db_owner ADD MEMBER Mary;

  (Correct)

- ☐

  CREATE LOGIN Mary WITH PASSWORD = 'strong_password';

Explanation

https://azure.microsoft.com/en-in/blog/adding-users-to-your-sql-azure-database/

https://docs.microsoft.com/en-us/sql/t-sql/statements/alter-role-transact-sql?view=sql-server-ver15

Question 25: Skipped
You develop a data ingestion process that will import data to a Microsoft Azure SQL Data Warehouse. The data to be ingested resides in parquet files stored in anAzure Data Lake Gen 2 storage account.

You need to load the data from the Azure Data Lake Gen 2 storage account into the Azure SQL Data Warehouse.

Solution:

1. Create an external data source pointing to the Azure storage account

2. Create a workload group using the Azure storage account name as the pool name

3. Load the data using the INSERT…SELECT statement

Does the solution meet the goal?

- ○

  Yes

- ○

  No

    (Correct)

Explanation

6 steps have to be performed to load data to SQL warehouse

1. Create a master key

2. Create a db scoped credential with the storage key

3. Create an external source

4. Create an external file format

5. Create an external table

6. Load from the external table

https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/load-data-overview

Question 26: Skipped
You have an Azure SQL database that has masked columns. You need to identify when a user attempts to infer data from the masked columns.

What should you use?

- ○

  Transparent Data Encryption (TDE)

- ○

  Auditing

    (Correct)

○

Azure Advanced Threat Protection (ATP)

○

Custom masking rules

Explanation

Unauthorized users can use queries to infer the masked data and hence frequent auditing of queries executed needs to be performed to understand suspicions queries

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

Question 27: Skipped

A company has a Microsoft Azure HDInsight solution that uses different cluster types to process and analyze data. Operations are continuous.

Reports indicate slowdowns during a specific time window.

You need to determine a monitoring solution to track down the issue in the least amount of time.

What should you use?

○

Ambari REST API

○

Azure Monitor Metrics

(Correct)

○

Azure Log Analytics alert rule query

○

HDInsight .NET SDK

○

Azure Log Analytics log search query

Explanation

Azure Log Analytics log search query - It's not a monitoring solution, but rather an investigation tool.

Ambari REST API - Note that it's not Ambari Web UI. It is a good portion of work to create a monitoring solution using only Ambari REST API.

Azure Monitor Metrics "- Metrics in Azure Monitor are lightweight and capable of supporting near real-time scenarios making them particularly useful for alerting and fast detection of issues."
https://docs.microsoft.com/en-us/azure/azure-monitor/platform/data-platform-metrics

HDInsight .NET SDK -This is all about managing HDInsight clusters.
https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-administer-use-dotnet-sdk

Azure Log Analytics alert rule query -Monitoring is needed, not alerts

https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-hadoop-oms-log-analytics-tutorial

Question 28: Skipped
You manage a solution that uses Azure HDInsight clusters.

You need to implement a solution to monitor cluster performance and status.

Which technology should you use?

- Ambari REST API

- Azure Log Analytics

- Ambari Web UI

    (Correct)

- Azure HDInsight .NET SDK

- ○

Azure HDInsight REST API

Explanation

Ambari is the recommended tool for monitoring utilization across the whole cluster. The Ambari dashboard shows easily glanceable widgets that display metrics such as CPU, network, YARN memory, and HDFS disk usage.

https://azure.microsoft.com/en-in/blog/monitoring-on-hdinsight-part-1-an-overview/

Question 29: Skipped

A company has a real-time data analysis solution that is hosted on Microsoft Azure. The solution uses Azure Event Hub to ingest data and an Azure Stream Analytics cloud job to analyze the data. The cloud job is configured to use 120 Streaming Units (SU).

You need to optimize performance for the Azure Stream Analytics job.

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

- ☐

  Implement Azure Stream Analytics user-defined functions (UDF)

- ☐

  Scale the SU count for the job down

- ☐

  Implement query parallelization by partitioning the data output

- ☐

  Implement query parallelization by partitioning the data input

  (Correct)

- ☐

  Implement event ordering

- ☐

  Scale the SU count for the job up

    (Correct)

Explanation

Stream Analytics supports higher performance by partitioning, allowing complex queries to be parallelized and executed on multiple streaming nodes.

As a cloud service, Stream Analytics is optimized for cost. There are no upfront costs involved - you only pay for the streaming units you consume. There is no commitment or cluster provisioning required, and you can scale the job up or down based on your business needs.

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-introduction

Question 30: Skipped
A company has a SaaS solution that uses Azure SQL Database with elastic pools. The solution contains a dedicated database for each customer organization. Customer organizations have peak usage at different periods during the year. You need to implement the Azure SQL Database elastic pool to minimize cost. Which option or options should you configure?

- ☐

  Number of transactions

- ☐

  eDTUs per database

- ☐

  Number of databases
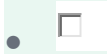
- ☐

  CPU usage

- ☐

  eDTUs

(Correct)

☐

max data size

(Correct)

Explanation

Elastic pools enable the developer to purchase resources for a pool shared by multiple databases to accommodate unpredictable periods of usage by individual databases. You can configure resources for the pool based either on the DTU-based purchasing model or the vCore-based purchasing model.

Data max size is an option that you can set up while provisioning the resource.

https://docs.microsoft.com/en-us/azure/azure-sql/database/elastic-pool-overview

Question 31: Skipped

Your team has created a new Azure Data Factory environment. You have to analyse the pipeline executions. Trends need to be identified in execution duration over the past 30 days. You need to create a solution that would ensure that data can be queried from Azure Log Analytics.

Which of the following would you use as the storage location when setting up the diagnostic setting for Azure Data Factory?

○

Azure Event Hub

○

Azure Storage Account

○

Azure Cosmos DB

○

Explanation

This might be to confuse students but this question has the answer in itself.

Data Factory stores pipeline-run data for only 45 days. Use Azure Monitor if you want to keep that data for a longer time. With Monitor, you can route diagnostic logs for analysis to multiple different targets.

Storage Account: Save your diagnostic logs to a storage account for auditing or manual inspection. You can use the diagnostic settings to specify the retention time in days.

Event Hub: Stream the logs to Azure Event Hubs. The logs become input to a partner service/custom analytics solution like Power BI.

Log Analytics: Analyze the logs with Log Analytics. The Data Factory integration with Azure Monitor is useful in the following scenarios: You want to write complex queries on a rich set of metrics that are published by Data Factory to Monitor. You can create custom alerts on these queries via Monitor. You want to monitor across data factories. You can route data from multiple data factories to a single Monitor workspace.

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

Question 32: Skipped

You are working as a data engineer for your company and you have been given the task to pull data from an on-premise SQL Server and migrate the data to Azure Blob storage. Your company is planning to use Azure Data Factory.

What steps should you be doing to implement the solution?

- ☐

  Create a new Azure Data Factory resource

  (Correct)

- ☐

  Create a Virtual Private Network Connection from the on-premise network to Azure

- ☐

  Create a self-hosted integration runtime

  (Correct)

- ☐

  Create a database master key

- ☐

  Backup the database

- ☐

  Configure the on-premise server to use an integration runtime

  (Correct)

Explanation

In this tutorial, you perform the following steps:

Create a data factory.

Create a self-hosted integration runtime. (This step also includes installation of self-hosted integration runtime on prem-server)

Create SQL Server and Azure Storage linked services.

Create SQL Server and Azure Blob datasets.

Create a pipeline with a copy activity to move the data.

Start a pipeline run.

Monitor the pipeline run.

https://docs.microsoft.com/en-us/azure/data-factory/tutorial-hybrid-copy-portal

Question 33: Skipped
You are working as a data engineer for your company and you have been given the task to migrate the data from an Azure Blob Storage Account to an Azure SQL Data warehouse. What should you be implementing for this requirement? Select 4 options from the options given below.

- ☐

  Provision an Azure SQL Data Warehouse instance

  (Correct)

- ☐

- Connect to the Blob storage container via SQL Server Management Studio

- ☐

  Create an Azure Blob storage container

- ☐

  Connect to the Azure SQL Data warehouse via SQL Server Management Studio

  (Correct)

- ☐

  Build external tables by using Azure portal

- ☐

  Run the T-SQL statements to load the data

  (Correct)

- ☐

  Build external tables by using SQL Server Management Studio

  (Correct)

Explanation

Create a data warehouse using SQL pool in the Azure portal

Set up a server-level firewall rule in the Azure portal

Connect to the SQL pool with SSMS

Create a user designated for loading data

Create external tables that use Azure blob as the data source

Use the CTAS T-SQL statement to load data into your data warehouse

View the progress of data as it is loading

Generate a year of data in the date dimension and sales fact tables

Create statistics on the newly loaded data

https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/load-data-wi deworldimportersdw

Question 34: Skipped
You are working as a data engineer for your company that has an Azure SQL Database and an Azure Blob storage account. The company has given you the task to encrypt the data at rest on both the systems. Also, the company should be able to use their own key.

What should you be using to configure security for the Azure SQL Database?

- ○

  Always Encrypted

- ○

  Cell-level encryption

- ○

  Row-level encryption

- ○

  Transparent data encryption

  (Correct)

Explanation

Transparent Data Encryption (TDE) encrypts SQL Server, Azure SQL Database, and Azure Synapse Analytics (SQL Data Warehouse) data files. This encryption is known as encrypting data at rest.

https://docs.microsoft.com/en-us/sql/relational-databases/security/encryption/transparent-da ta-encryption?view=sql-server-ver15

Question 35: Skipped
You are working as a data engineer for your company that has an Azure SQL Database and an Azure Blob storage account. The company has given you the task to encrypt the data at rest on both the systems. Also, the company should be able to use their own key.

What should you be using to configure security for the Azure Blob storage account?

- ○

Azure Disk Encryption

- ○

Secure Transport Layer Security

- ○

Storage Account Keys

- ○

Default Storage Service Encryption

(Correct)

Explanation

Azure Storage encryption is enabled for all storage accounts, including both Resource Manager and classic storage accounts. Azure Storage encryption cannot be disabled. Because your data is secured by default, you don't need to modify your code or applications to take advantage of Azure Storage encryption.

https://docs.microsoft.com/en-us/azure/storage/common/storage-service-encryption

Question 36: Skipped
Working as a data engineer for your company, you create an Azure Data Warehouse database that uses round robin distribution. To retrieve the data on the store sales, you write the following query:

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

Now, this query returns over 200k records and it is very slow. There are over 50k stores.

What should you do to improve the performance of this query?

- ☐

Use hash distribution on StoreId for FactSalesByStore.

(Correct)

- ☐

Use hash distribution on StoreId for DimStore.

(Correct)

- ☐

  Use replicated distribution for DimStore.

- ☐

  Use an outer join instead of an inner join.

Explanation

A hash-distributed table distributes table rows across the Compute nodes by using a deterministic hash function to assign each row to one distribution.

https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-distribute#:~:text=A%20round%2Drobin%20distributed%20table,assigned%20to%20the%20same%20distribution.

Question 37: Skipped

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?

- ○

  Integrated Security = false

- ○

  Column Encryption Setting = enabled

  (Correct)

- ○

  Integrated Security = true

- ○

Column Encryption Setting = disabled

Explanation

Make sure you have enabled Always Encrypted for the database connection for the Query Editor window, from which you will run a `SELECT` query retrieving and decrypting your data. This will instruct the .NET Framework Data Provider for SQL Server (used by SSMS) to decrypt the encrypted columns in the query result set.

https://docs.microsoft.com/en-us/sql/relational-databases/security/encryption/always-encrypted-query-columns-ssms?view=sql-server-ver15

Question 38: Skipped
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?

- ○

  SaXX Jack xxxx

    (Correct)

- ○

  XXXX Jack XXX.XXX.XXXX

- ○

  SaXX Jack XXX.XXX.2222

- ○

  xx Jack XXX.XXX.2222

Explanation

Refer the default and custom text rule here.

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

Question 39: Skipped

You are working as a data engineer for a call center company that uses Azure SQL Database. The database stores customer credit card numbers in a table called PaymentInfo. To help the customer payment support telemarketing attendants will be consulting this table.

You are given the task to implement dynamic data masking (DDM) in the PaymentInfo table to mask credit card numbers for telemarketing attendants. You need to make sure that only the 1st digit and the last 4 digits are visible.

How should you be completing the T-SQL query? Select the relevant options to answer.

Larger image

Larger image

A-options

1. Default

2. Partial

3. Random

B-options

4. 0

5. 1

6. 4

C-options

7. 0

8. 2

9. 4

- ○

    A-1, B-5, C-9

- ○

  A-2, B-4, C-7

- ○

  A-2, B-5, C-9

  (Correct)

- ○

  A-3, B-6, C-9

Explanation

Refer the custom text rules here

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

Question 40: Skipped
**Your company uses Microsoft Azure SQL Database configured with Elastic pools.
You use Elastic Database jobs to run queries across all databases in the pool.
You need to analyze, troubleshoot, and report on components responsible for running Elastic Database jobs.
You need to determine the component responsible for running job service tasks.
Which components should you use for each Elastic pool job services task? To answer, drag the appropriate component to the correct task. Each component may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.**

Larger image

Answer Area

Components	Task	Component
Control Database	Execution results and diagnostics	[ ]
Azure Service Bus	Job launcher and tracker	[ ]
Azure Storage	Job metadata and state	[ ]
Job Service		

○

Execution results and diagnostics : Azure Storage
Job launcher and tracker : Azure Service Bus
Job metadata and state : Control database

○

Execution results and diagnostics : Azure Storage
Job launcher and tracker : Job Service
Job metadata and state : Control database

    (Correct)

○

Execution results and diagnostics : Azure Storage
Job launcher and tracker : Control database
Job metadata and state : Azure Service Bus

Explanation

One or more destinations to send the logs. Current destinations include Log Analytics workspace, Event Hubs, and Azure Storage.

Question 41: Skipped
You have a new Azure Data Factory environment.
You need to periodically analyze pipeline executions from the last 60 days to identify trends in execution durations. The solution must use Azure Log Analytics to query the data and create charts.
Which diagnostic settings should you configure in Data Factory?

Larger image

Answer Area

Log type:

ActivityRuns
AllMetrics
PipelineRuns
TriggerRuns

Storage location:

An Azure event hub
An Azure Storage account
Azure Cosmos DB
Azure Log Analytics

Reveal Solution

- ○

    Log type : ActivityRuns
    Storage location : An Azure Storage account

- ○

    Log type : TriggerRuns
    Storage location : Azure Cosmos DB

- ○

    Log type : PipelineRuns
    Storage location : Azure Log Analytics

    (Correct)

- ○

    Log type : AllMetrics
    Storage location : An Azure event hub

Explanation

Log Analytics workspace - Sending logs and metrics to a Log Analytics workspace allows you to analyze them with other monitoring data collected by Azure Monitor using powerful log queries and also to leverage other Azure Monitor features such as alerts and visualizations.

Retention period can be configured from 30 to 730 days (2 years)

https://docs.microsoft.com/en-us/azure/azure-monitor/platform/diagnostic-settings

https://docs.microsoft.com/en-us/azure/azure-monitor/platform/manage-cost-storage

Azure Data Factory Analytics (Preview) sends diagnostic logs to *Resource-specific* destination tables. You can write queries against the following tables: *ADFPipelineRun*, *ADFTriggerRun*, and *ADFActivityRun*.

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

Practice Test 5

Question 1: Skipped
Isolated database that is perfect for applications that need a single data source?

- ○

Single Database

(Correct)

- ○

Elastic Pool

- ○

Managed Instance

Explanation

This is a single isolated database that is perfect for applications that need a single data source. In the single database scenario, each database receives its own guaranteed compute, memory, and storage.

Question 2: Skipped
Company has a plan to migrate to Azure Cloud, they setup a meeting to figure out what is easiest and quickest way to lift and shift on-premises database to cloud, without much changes and hassle.

What option would you suggest in meeting?

- ○

SQL Server in VM

(Correct)

- ○

Azure Single Database

- ○

Azure Elastic Pool Database

Explanation

You might want to choose this option for migrating on-premises SQL Server databases and applications without any database change. This is a lift-and-shift scenario. The SQL Server installed on a VM is identical to the SQL Server you have installed on-premises.

Question 3: Skipped
As an experienced Data Engineer you wants to have a more control, flexible and transparency while choosing purchasing model. Which purchasable model would you choose?

- ○

  DTU based

- ○

  Vcore Based

  (Correct)

Explanation

vCore-based model is best for customers who need flexibility, control, and transparency. As you saw in this model, you have the flexibility to adjust individual resources, such as CPU, memory, and storage. The DTU-based model is best for customers who want simple, preconfigured resource options.

Question 4: Skipped
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?

- ○

  Azure SQL Database Managed Instance

  (Correct)

- ○

  Azure SQL Database with an Elastic Pool

- ○

  Azure Cosmos DB with Table API

- ○

  Azure Cosmos DB with SQL API

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.

Question 5: Skipped

You want to make our Azure SQL Database instance private and make sure only enable applications or resources within a specific subnet to access the database resources.

Which security setting you would configure?

- ○

  Network Security

  (Correct)

- ○

  Authorization

- ○

  Authentication

- ○

  Transparent Data Encryption

Explanation

Azure SQL Database firewall enables you to only accept requests originated from subnets inside a virtual network. This gives you the flexibility to make your Azure SQL Database instance private and only enable applications or resources within a specific subnet to access the database resources.

Question 6: Skipped
Which answer best describes a relational database?

○

Relational database systems are good for varying data types and complex queries. Typically relational databases utilize Structured Query Language.

○

Relational database systems are good for varying data types and simple, fast queries. Typically relational databases utilize Structured Query Language.

○

Relational database systems are good for non-varying data types and complex queries. Typically relational databases utilize Cassandra.

(Correct)

○

Relational database systems are good for non-varying data types and complex queries. Typically relational databases utilize Structured Query Language.

Question 7: Skipped
Even if you pause the compute, it will continue charging you for used storage?

○

Yes

(Correct)

○

No

Explanation

You will continue to be charged for Storage. The charge for compute will stop

Question 8: Skipped
Acme Warehouse is launching a new product. They need to implement an online storage solution to leverage scale-out architecture processing. Acme needs to

process highly complex queries from massive amounts of data. Which solution would you recommend?

- ○

  SQL Database

- ○

  Azure DataLake Gen2

- ○

  Azure Blob Storage

- ○

  SQL Data Warehouse

    (Correct)

Explanation

A SQL data warehouse is a massively parallel processing engine designed to handle complex queries from massive amounts of data.

Question 9: Skipped

The following query is used to retrieve the sales by business reseller, but the performance of the query is slow. The query is as follows: SELECT S.[SalesAmount] AS [Sales], R.[BusinessType],R.[ResellerName] FROM [FactResellerSales] AS S JOIN [DimReseller] AS R ON S.[ResellerKey] = R.[ResellerKey]. The tables referenced within the query are configured with a distribution of Round_Robin with a clustered columnstore index. The data engineer wants to improve the performance of the query. What operation can be used to improve the performance of the query?

- ○

  Remove the CLUSTERED COLUMNSTORE INDEX for both tables

- ○

  Change the Distribution to HASH(GeographyKey) for both tables.

- ○

Change the Distribution to HASH(ResellerKey) for both tables.

(Correct)

Explanation

Placing Hash distribution on the ResellerKey on both the FactResellerSales and DimReseller improves the performance of the query.

Question 10: Skipped
You have been asked to implement an ETL process involving Data Factory, Azure Data Lake, SQL Database, DataBricks, and PowerBI. Which order is most correct?

- ○

  Pull data from source system and move it to Data Factory with PolyBase. Then load it into Databricks with PolyBase. Next, move data back to SQL Database with PolyBase and then to PowerBI for visualization.

- ○

  Pull data from source system and move it to SQL Database with Data Factory. Then load it into Databricks with Data Factory. Next, move data back to SQL Database with Data Factory and then to PowerBI for visualization.

- ○

  Pull data from source system and move it to SQL Database with Data Factory. Then load it into PowerBI for visualization.

- ○

  Pull data from source system and move it to Databricks with Data Factory. Next, move data to SQL Database with Data Factory and then to PowerBI for visualization.

  (Correct)

Question 11: Skipped
You are working in a Health care as a Data Engineer, and you are asked to detects anomalous activities indicating unusual and potentially harmful attempts to access or exploit your database. Which functionality you will use?

- ○

  Auditing

- ○

Threat Protection

(Correct)

- ○

Vulnerability Assessment

- ○

Data Discovery and Protection

Question 12: Skipped
Which functionality you use to allow or block IP Addresses?

- ○

Firewall settings

(Correct)

- ○

Transparent Data Encryption

- ○

Transport Layer Encryption

Question 13: Skipped
Validating user credentials like User Name/ User ID and password is part of which functionality?

- ○

Authentication

(Correct)

- ○

Question 14: Skipped

You are developing the data platform for a global retail company. The company operates during normal working hours in each region. The analytical database is used once a week for building sales projections.

Each region maintains its own private virtual network.

Building the sales projections is very resource intensive are generates upwards of 20 terabytes (TB) of data.

Microsoft Azure SQL Databases must be provisioned.

☞ Database provisioning must maximize performance and minimize cost

☞ The daily sales for each region must be stored in an Azure SQL Database instance

☞ Once a day, the data for all regions must be loaded in an analytical Azure SQL Database instance

You need to provision Azure SQL database instances.

How should you provision the database instances? To answer, drag the appropriate Azure SQL products to the correct databases. Each Azure SQL product may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

Larger image

- ○

  Daily Sales-->Azure SQL Database Managed Instance
  Weekly Sales-->Azure SQL Database Premium

- ○

  Daily Sales-->Azure SQL Database Elastic pools
  Weekly Sales-->Azure SQL Database Premium

- ○

  Daily Sales-->Azure SQL Database Elastic pools
  Weekly Sales-->Azure SQL Database Hyperscale

    (Correct)

- ○

  Daily Sales-->Azure SQL Database Managed Instance
  Weekly Sales-->Azure SQL Database Hyperscale

Explanation

Azure SQL Database elastic pools

SQL Database elastic pools are a simple, cost-effective solution for managing and scaling multiple databases that have varying and unpredictable usage demands. The databases in an elastic pool are on a single Azure SQL Database server and share a set number of resources at a set price. Elastic pools in Azure

Azure SQL Database Hyperscale

A Hyperscale database is an Azure SQL database in the Hyperscale service tier that is backed by the Hyperscale scale-out storage technology. A Hyperscale database supports up to 100 TB of data and provides high throughput and performance, as well as rapid scaling to adapt to the workload requirements. S

Question 15: Skipped
The data engineering team manages Azure HDInsight clusters. The team spends a large amount of time creating and destroying clusters daily because most of the data pipeline process runs in minutes.

You need to implement a solution that deploys multiple HDInsight clusters with minimal effort.

What should you implement?

- ○

  Azure Databricks

- ○

  Azure Traffic Manager

- ○

  Azure Resource Manager templates

    (Correct)

- ○

  Ambari web user interface

Explanation

A Resource Manager template makes it easy to create the following resources for your application in a single, coordinated operation:

In the template, you define the resources that are needed for the application. You also specify deployment parameters to input values for different environments.

The template consists of JSON and expressions that you use to construct values for your deployment.

Question 16: Skipped
A company is designing a hybrid solution to synchronize data and on-premises Microsoft SQL Server database to Azure SQL Database.

You must perform an assessment of databases to determine whether data will move without compatibility issues. You need to perform the assessment.

Which tool should you use?

- ○

  SQL Server Migration Assistant (SSMA)

- ○

  Microsoft Assessment and Planning Toolkit

- ○

  SQL Vulnerability Assessment (VA)

- ○

  Azure SQL Data Sync

- ○

  Data Migration Assistant (DMA)

    (Correct)

Explanation

The Data Migration Assistant (DMA) helps you upgrade to a modern data platform by detecting compatibility issues that can impact database functionality in your new version of SQL Server or Azure SQL Database. DMA recommends performance and reliability

improvements for your target environment and allows you to move your schema, data, and uncontained objects from your source server to your target server.

References:

https://docs.microsoft.com/en-us/sql/dma/dma-overview

Question 17: Skipped
You manage a financial computation data analysis process. Microsoft Azure virtual machines (VMs) run the process in daily jobs, and store the results in virtual hard drives (VHDs.) The VMs product results using data from the previous day and store the results in a snapshot of the VHD. When a new month begins, a process creates a new VHD.

You must implement the following data retention requirements:

– Daily results must be kept for 90 days

– Data for the current year must be available for weekly reports

– Data from the previous 10 years must be stored for auditing purposes

– Data required for an audit must be produced within 10 days of a request.

You need to enforce the data retention requirements while minimizing cost.

How should you configure the lifecycle policy? To answer, drag the appropriate JSON segments to the correct locations. Each JSON segment may be used once, more than once, or not at all. You may need to drag the split bat between panes or scroll to view content. NOTE: Each correct selection is worth one point.

Larger image

- ⚪

  "actions":
  {
  "baseBlob":
  { **"tierToCool"**: { "daysAfterModificationGreaterThan": 365 },
  "delete": { "daysAfterModificationGreaterThan": 3650 }
  },
  "snapshot":
  { **"delete"**: { "daysAfterCreationGreaterThan": 90 }
  }
  }

- ⚪

  "actions":

```
{
"baseBlob":
{ "tierToArchieve": { "daysAfterModificationGreaterThan": 365 },
"delete": { "daysAfterModificationGreaterThan": 3650 }
},
"snapshot":
{ "tierToCool": { "daysAfterCreationGreaterThan": 90 }
}
}
```

○

```
"actions":
{
"baseBlob":
{ "tierToCool": { "daysAfterModificationGreaterThan": 365 },
"tierToArchieve": { "daysAfterModificationGreaterThan": 3650 }
},
"snapshot":
{ "delete": { "daysAfterCreationGreaterThan": 90 }
}
}
```

○

```
"actions":
{
"baseBlob":
{ "tierToArchieve": { "daysAfterModificationGreaterThan": 365 },
"delete": { "daysAfterModificationGreaterThan": 3650 }
},
"snapshot":
{ "delete": { "daysAfterCreationGreaterThan": 90 }
}
}
```

(Correct)

Explanation

https://docs.microsoft.com/en-us/azure/storage/blobs/storage-lifecycle-management-concepts?tabs=template#rules

"actions":

{

"baseBlob":

{ **"tierToArchieve"**: { "daysAfterModificationGreaterThan": 365 },

"delete": { "daysAfterModificationGreaterThan": 3650 }

},

"snapshot":

{ **"delete"**: { "daysAfterCreationGreaterThan": 90 }

}

}

Question 18: Skipped

A company runs Microsoft SQL Server in an on-premises virtual machine (VM).

You must migrate the database to Azure SQL Database. You synchronize users from Active Directory to Azure Active Directory (Azure AD).

You need to configure Azure SQL Database to use an Azure AD user as administrator.

What should you configure?

- ○

  For each Azure SQL Database, set the Access Control to administrator.
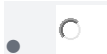
- ○

  For each Azure SQL Database server, set the Active Directory to administrator.

  (Correct)

- ○

  For each Azure SQL Database, set the Active Directory administrator role.

- ○

  For each Azure SQL Database server, set the Access Control to administrator.

Explanation

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

Question 19: Skipped

You have a table named SalesFact in an Azure SQL data warehouse. SalesFact contains sales data from the past 36 months and has the following characteristics:

☞ **Is partitioned by month**

☞ **Contains one billion rows**

☞ **Has clustered columnstore indexes**

All the beginning of each month, you need to remove data SalesFact that is older than 36 months as quickly as possible.

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

Larger image

- ○

  1. Create an empty table named SalesFact_Work that has the same schema as the SalesFact
  2. Switch the partitions containing the stale data from SalesFact to SalesFact_Work
  3. Truncate the partition containing the stale data

  (Correct)

- ○

  1. Create an empty table named SalesFact_Work that has the same schema as the SalesFact
  2. Switch the partitions containing the stale data from SalesFact to SalesFact_Work
  3. Drop the SalesFact_Work table

- ○

  1. Create an empty table named SalesFact_Work that has the same schema as the SalesFact
  2. Switch the partitions containing the stale data from SalesFact to SalesFact_Work
  3. Execute the delete statement where the value in the DATE column is more than 36 months ago

Explanation

https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-develop-best-practices-transactions

Question 20: Skipped

You plan to implement an Azure Cosmos DB database that will write 100,000 JSON every 24 hours. The database will be replicated to three regions. Only one region will be writable.

You need to select a consistency level for the database to meet the following requirements:

☞ Guarantee monotonic reads and writes within a session.

☞ Provide the fastest throughput.

☞ Provide the lowest latency.

Which consistency level should you select?

○

Strong

○

Bounded Staleness

○

Session

(Correct)

○

Consistent Prefix

Explanation

https://docs.microsoft.com/en-us/azure/cosmos-db/consistency-levels

Question 21: Skipped
You have an Azure Stream Analytics query. The query returns a result set that contains 10,000 distinct values for a column named clusterID.

You monitor the Stream Analytics job and discover high latency.

You need to reduce the latency.

Which two actions should you perform? Each correct answer presents a complete solution.

☐

Add a pass-through query.

☐

Add a temporal analytic function.

☐

Scale out the query by using PARTITION BY.

(Correct)

☐

Convert the query to a reference query.

☐

Increase the number of streaming units.

(Correct)

Explanation

Stream Analytics supports higher performance by partitioning, allowing complex queries to be parallelized and executed on multiple streaming nodes.

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-introduction

Streaming Units (SUs) represents the computing resources that are allocated to execute a Stream Analytics job. The higher the number of SUs, the more CPU and memory resources are allocated for your job.

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-streaming-unit-consumption

Question 22: Skipped
Proactive Ltd. plans to configure existing applications to use Azure SQL Database.

When security-related operations occur, the security team must be informed.

You need to configure Azure Monitor while minimizing administrative effort. Which three actions should you perform?

☐

Put all Azure SQL Database servers under one resource group

(Correct)

- ☐

Query audit log entries as a condition.

- ☐

Create a new action group to email alerts@proactive.com.

(Correct)

- ☐

Use all security operations as a condition.

(Correct)

- ☐

Use alerts@proactive.com as an alert email address.

Explanation

Although alert rules help you define the action group that triggers when the alert is generated, customers often have a common action group across their scope of operations. For example, a team responsible for the resource group **ContosoRG** will probably define the same action group for all alert rules defined within **ContosoRG**.

Action rules help you simplify this process. By defining actions at scale, an action group can be triggered for any alert that's generated on the configured scope.

https://docs.microsoft.com/en-us/azure/azure-monitor/platform/alerts-action-rules?tabs=port
al

Question 23: Skipped
You manage a process that performs analysis of daily web traffic logs on an HDInsight cluster. Each of the 250 web servers generates approximately 10 megabytes (MB) of log data each day. All log data is stored in a single folder in Microsoft Azure Data Lake Storage Gen 2.
You need to improve the performance of the process.
Which two changes should you make? Each correct answer presents a complete solution

- ☐

Combine the daily log files for all servers into one file

(Correct)

- ☐

  Increase the value of the mapreduce.map.memory parameter

- ☐

  Move the log files into folders so that each day's logs are in their own folder

  (Correct)

- ☐

  Increase the number of worker nodes

- ☐

  Increase the value of the hive.tez.container.size parameter

Explanation

Typically, analytics engines such as HDInsight and Azure Data Lake Analytics have a per-file overhead. If you store your data as many small files, this can negatively affect performance. In general, organize your data into larger sized files for better performance (256MB to 100GB in size). Some engines and applications might have trouble efficiently processing files that are greater than 100GB in size.

For Hive workloads, partition pruning of time-series data can help some queries read only a subset of the data which improves performance.

Those pipelines that ingest time-series data, often place their files with a very structured naming for files and folders. Below is a very common example we see for data that is structured by date:

\DataSet\YYYY\MM\DD\datafile_YYYY_MM_DD.tsv

Notice that the datetime information appears both as folders and in the filename.

https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-performance-tuning-guidance

Question 24: Skipped
An application will use Microsoft Azure Cosmos DB as its data solution. The application will use the Cassandra API to support a column-based database type that uses containers to store items. You need to provision Azure Cosmos DB.

Which container name and item name should you use?

Each correct answer presents part of the solutions.

- ☐

  collection

- ☐

  rows

  (Correct)

- ☐

  graph

- ☐

  entities

- ☐

  table

  (Correct)

Explanation

Read below document to understand the name of the Table and Row are each type of Cosmos DB APIs.

https://docs.microsoft.com/en-us/azure/cosmos-db/account-databases-containers-items

Question 25: Skipped

A company runs Microsoft SQL Server in an on-premises virtual machine (VM).

You must migrate the database to Azure SQL Database. You synchronize users from Active Directory to Azure Active Directory (Azure AD).

You need to configure Azure SQL Database to use an Azure AD user as administrator.

What should you configure?

- ○

For each Azure SQL Database server, set the Active Directory to administrator.

○

For each Azure SQL Database, set the Active Directory administrator role.

(Correct)

○

For each Azure SQL Database server, set the Access Control to administrator.

○

For each Azure SQL Database, set the Access Control to administrator.

Question 26: Skipped

You have to develop a solution using Azure Stream Analytics. The stream will be sued to receive Twitter data from Azure Event Hubs. The output would be sent to an Azure Blob storage account. The key requirement is to output the number of tweets during the last 3 minutes every 3 minutes. Each tweet must be counted only once. Which of the following would you use as the windowing function?

○

A three-minute Session window

○

A three-minute Sliding ion window

○

A three-minute Tumbling window

(Correct)

○

A three-minute Hopping window

Explanation

Sliding and Hopping window are ruled out because they are over lapping windows in nature.

Session window is ruled out because it ends when there is a silence of stated period.

Therefore its tumbling window which can meet all the requirements. Tumbling windows are a series of fixed-sized, non-overlapping and contiguous time intervals.

https://docs.microsoft.com/en-us/stream-analytics-query/tumbling-window-azure-stream-analytics

Question 27: Skipped

You are working as a data engineer for your company and you have been given the task to configure synchronization of data between their on-premise Microsoft SQL Server database and Azure SQL database. The synchronization process should include the following:

· Be able to perform an initial data synchronization to the Azure SQL Database with minimal downtime.

· Be able to perform bi-directional synchronization after the initial synchronization is complete

What should you be considering as the synchronization solution?

- ○

  Data Migration Assistant

- ○

  Backup and restore

- ○

  SQL Server Agent Job

- ○

  Azure SQL Data Sync

  (Correct)

Explanation

SQL Data Sync is a service built on Azure SQL Database that lets you synchronize the data you select bi-directionally across multiple databases, both on-premises and in the cloud.

https://docs.microsoft.com/en-us/azure/azure-sql/database/sql-data-sync-data-sql-server-sql-database

Question 28: Skipped

You are working as a data engineer for your company and your company has an Azure storage account named AzureCloudstore5000. The Diagnostic settings configured for the storage account are given below:

Larger image

Larger image

How long do you think that the logging data be retained for?

- ○

  7 days

  (Correct)

- ○

  365 days

- ○

  Indefinitely

- ○

  90 days

Explanation

To set the data retention policy, move the **Retention (days)** slider or enter the number of days of data to retain, from 1 to 365. The default for new storage accounts is seven days. If you do not want to set a retention policy, enter zero. **If there is no retention policy, it is up to you to delete the monitoring data.**

https://docs.microsoft.com/en-us/azure/storage/common/storage-monitor-storage-account

Question 29: Skipped
You are working as a data engineer for your company that is using an Azure SQL Data Warehouse Gen2. There have been complaints that the performance is slow when some commonly used queries are run. There have no issues like this for infrequently used queries. What should you be monitoring to find the source of performance issue?

- ○

Cache used percentage

  (Correct)

- ○

Memory percentage

- ○

CPU percentage

- ○

Failed connections

Explanation

The Gen2 storage architecture automatically tiers your most frequently queried columnstore segments in a cache residing on NVMe based SSDs designed for Gen2 data warehouses. Greater performance is realized when your queries retrieve segments that are residing in the cache.

https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-how-to-monitor-cache

Question 30: Skipped
Account that makes use of the SQL API. You are given the task to make sure that all the stale data is deleted from the database automatically.

- ○

Soft delete

- ○

Schema Read

- ○

Time to Live

(Correct)

○

CORS

Explanation

With **Time to Live** or TTL, Azure Cosmos DB provides the ability to delete items automatically from a container after a certain time period.

https://docs.microsoft.com/en-us/azure/cosmos-db/time-to-live

Question 31: Skipped
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

Select all that is True.

- ☐

  You can query data by using Gremlin API.

- ☐

  A client can see partial writes of a sales data record by default.

- ☐

  A client can set the consistency level to Eventual Consistency at connection time.

    (Correct)

- ☐

  A client can set a different consistency level during each request to sales data.

Explanation

Clients can override the default consistency level that is set by the service. Consistency level can be set on a per request, which overrides the default consistency level set at the account level.

Tip

Consistency can only be **relaxed** at the request level. To move from weaker to stronger consistency, update the default consistency for the Cosmos account.

https://docs.microsoft.com/en-us/azure/cosmos-db/how-to-manage-consistency?tabs=portal%2Cdotnetv2%2Capi-async#override-the-default-consistency-level

Question 32: Skipped

You are working as a data engineer for your company and you manage an Azure SQL Data Warehouse. You have been given the task to import data to a database form a large pipe-separated file in an Azure blob storage container.

You have to make the connection to the container.

How should you be completing the T-SQL statement? Select the relevancy code segments to answer.

Larger image

Larger image

A-options

1. Data Source

2. File Format

3. Table

B-options

4. Blob Storage

5. Hadoop

- ○

  2 and 4

- ○

  1 and 5

○

3 and 5

○

3 and 4

Explanation

-- LOCATION: Azure account storage account name and blob container name.

-- CREDENTIAL: The database scoped credential created above.

CREATE EXTERNAL DATA SOURCE AzureStorage with (

TYPE = HADOOP,

LOCATION ='wasbs://<blob_container_name>@<azure_storage_account_name>.blob.core.windows.net',

CREDENTIAL = AzureStorageCredential

);

https://docs.microsoft.com/en-us/sql/analytics-platform-system/polybase-configure-azure-blob-storage?view=aps-pdw-2016-au7

Question 33: Skipped

You're a Data Engineer for an Exam Development Company. You create an Azure Cosmos DB account that uses the level of session consistency. You create a database and a collection that allows exam developers to create and store the content of the exam.

Developer Rahul and Developer Amit are based in Los Angeles, United States. Developer Pavan is based in Bangalore, India.

At 12:00, the question entity has its difficulty attribute set to Hard. All three developers were reading the Hard value. Developer Rahul then changes the difficulty attribute to the Medium. All three developers read the entity immediately before replication occurs.

You need to answer questions about the readings.

A. Developer Rahul will read Medium.

B. Developer Amit will read Medium.

C. Developer Pavan will read Hard.

- ○

  A and B

- ○

  B and C

- ○

  A and C

  (Correct)

Explanation

Session: Within a single client session reads are guaranteed to honor the consistent-prefix, monotonic reads, monotonic writes, read-your-writes, and write-follows-reads guarantees. This assumes a single "writer" session or sharing the session token for multiple writers.

https://docs.microsoft.com/en-us/azure/cosmos-db/consistency-levels

Question 34: Skipped
You are working as a data engineer for you company. Your company uses Azure Data Factory to copy and transform data from Azure blob storage to an on-premises server.

You have been given the task to ensure that you copy the data successfully.

What 2 actions should you be performing? Each correct option provides part of the solution.

- ☐

  Create a self-hosted integration runtime in Azure Data Factory UI.

  (Correct)

- ☐

  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.

  (Correct)

- ☐

  Create an Azure-SSIS integration runtime.

Explanation

Refer the process here

https://docs.microsoft.com/en-us/azure/data-factory/create-self-hosted-integration-runtime

Question 35: Skipped

You are working as a data engineer for a company and your company is implementing an event processing to handle real-time streaming data from Twitter.

You have been given a task to configure an Azure Event hub named TweetStream to ingest the streaming data.

To analyze the streaming data from Event Hub you think of using Azure Databricks. In that you create a new workbook named TweetAnalysis.

You have to configure the TweetAnalysis workbook to connect with TweetStream.

How should you be completing the notebook code? Select the relevant options to answer.

Larger image

Larger image

A-options

1. .setEventHubName

2. .setNamesapceName

3. .setSasKeyName

B-options

4. Read

5. ReadStream

6. WriteStream

C-options

7. Eventhubs

8. Kinesis

9. memory

- ○

  A-1, B-5, C-7

  (Correct)

- ○

  A-2, B-6, C-8

- ○

  A-3, B-4, C-9

- ○

  A-1, B-4, C-8

Explanation

import org.apache.spark.eventhubs._

   import com.microsoft.azure.eventhubs._

   // Build connection string with the above information

   val namespaceName = "<EVENT HUBS NAMESPACE>"

   val eventHubName = "<EVENT HUB NAME>"

   val sasKeyName = "<POLICY NAME>"

   val sasKey = "<POLICY KEY>"

   val connStr = new com.microsoft.azure.eventhubs.ConnectionStringBuilder()

      .setNamespaceName(namespaceName)

      .setEventHubName(eventHubName)

      .setSasKeyName(sasKeyName)

      .setSasKey(sasKey)

```
val customEventhubParameters =

  EventHubsConf(connStr.toString())

  .setMaxEventsPerTrigger(5)


val incomingStream =
spark.readStream.format("eventhubs").options(customEventhubParameters.toMap).l
oad()


incomingStream.printSchema


// Sending the incoming stream into the console.

// Data comes in batches!

incomingStream.writeStream.outputMode("append").format("console").option("truncate",
false).start().awaitTermination()
```

https://docs.microsoft.com/en-us/azure/databricks/scenarios/databricks-stream-from-eventh
ubs

Question 36: Skipped
**You implement an event processing solution using Microsoft Azure Stream
Analytics.**

The solution must meet the following requirements:

⚭ Ingest data from Blob storage

⚭ Analyze data in real time

⚭ Store processed data in Azure Cosmos DB

**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 query statement with the ORDER BY clause

- ☐

  Create a query statement with the SELECT INTO clause

    (Correct)

- ☐

  Configure Blob storage for a reference data JSON clause

- ☐

  Configure Azure Event Hub as input; select items with TIMESTAMP BY clause

- ☐

  Configure Blob Storage as input; select items with TIMESTAMP BY clause

  (Correct)

- ☐

  Set up Cosmos DB as output

  (Correct)

Explanation

To process the data as a stream using a timestamp in the event payload, you must use the TIMESTAMP BY keyword.

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-define-inputs

https://docs.microsoft.com/en-us/stream-analytics-query/timestamp-by-azure-stream-analytics

Question 37: Skipped

A company plans to analyze a continuous flow of data from a social media platform by using Microsoft Azure Stream Analytics. The incoming data is formatted as one record per row.
You need to create the input stream.
How should you complete the REST API segment? To answer, select the appropriate configuration in the answer area.

Larger image

Answer Area

```
{
  "properties" : {
    "type" : "stream",
    "serialization" : {
```

▼
"type":"CSV"
"type":"Avro",
"type":"JSON",

```
      "properties" : {
        "fieldDelimiter" : ". ",
        "encoding" : "UTF8"
      }
  },
  "datasource":{
```

▼
"type":"Microsoft.Storage/Blob",
"type":"Microsoft.ServiceBus/EventHub",
"type":"Microsoft.Devices/IotHubs",

```
    "properties": {
        "serviceBusNamespace" : "sampleServiceBus",
        "sharedAccessPolicyName" : "SampleReceiver",
        "sharedAccessPolicyKey" : "<PolicyKey>"
        "eventHubName" : "sampleEventHub"
      }
    },
      "compression":{
      "type" : "GZip"
      }
    }
}
```

- ☐

CSV

- □

Avro

(Correct)

- □

Json

- □

Microsoft.Storage/Blob

- □

Microsoft.ServiceBus/EventHub

(Correct)

- □

Microsoft.Devices/IotHub

Explanation

The Azure Stream Analytics service makes it easy to ingest, process, and analyze streaming data from Azure Event Hubs, enabling powerful insights to drive real-time actions.

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-cicd-api

https://docs.microsoft.com/en-us/azure/event-hubs/process-data-azure-stream-analytics

Question 38: Skipped
Your company plans to create an event processing engine to handle streaming data from Twitter.
The data engineering team uses Azure Event Hubs to ingest the streaming data.
You need to implement a solution that uses Azure Databricks to receive the streaming data from the Azure Event Hubs.
Which three actions should you recommend be performed?

Larger image

Actions

Create and configure a Notebook that consumes the streaming data.

Import data from Blob storage.

Use Environment variables to define the Apache Spark connection.

Configure an ODBC or JDBC Connector.

Deploy the Azure Databricks service.

Deploy a Spark cluster and then attach the required libraries to the cluster.

Answer Area

(◀) (▶)   (▲) (▼)

Hide Solution

- ☐

  Create and configure a Notebook that consumes the streaming data

    (Correct)

- ☐

  Import data from Blob storage

- ☐

  Use Environments variables to define Apache spark connection

- ☐

  Configure an ODBC and JDBC connector

- ☐

  Deploy the Azure Databricks service

    (Correct)

- ☐

  Deploy a spark cluster and then attach the required libraries to the cluster

    (Correct)

Explanation

Create an Azure Databricks workspace

Create a Spark cluster in Azure Databricks

Create a Twitter app to access streaming data

Create notebooks in Azure Databricks

Attach libraries for Event Hubs and Twitter API

Send tweets to Event Hubs

Read tweets from Event Hubs

https://docs.microsoft.com/en-us/azure/databricks/scenarios/databricks-stream-from-eventhubs

Question 39: Skipped
A company plans to develop solutions to perform batch processing of multiple sets of geospatial data.
You need to implement the solutions.
Which Azure services should you use? To answer, select the appropriate configuration in the answer area.
NOTE: Each correct selection is worth one point.
Hot Area:

Larger image

Answer Area

Scenario	Tool
Use a native client application to run interactive queries and batch processes.	HDInsight Tools for Visual Studio / Hive View / HDInsight REST API / Azure Data Factory
Use a web browser to run interactive queries and batch processes.	HDInsight Tools for Visual Studio / Hive View / HDInsight REST API / Azure PowerShell
Develop batch processing applications that use Azure HDInsight	HDInsight Tools for Visual Studio / Hive View / HDInsight REST API / NoSQL database

- ○

  1. HDInsight Tools for Visual Studio 2. Hive View 3. HDInsight REST API

    (Correct)

- ○

  1. HDInsight Tools for Visual Studio 2. Hive View 3. HDInsight REST API

- ○

  1. Hive View 2. HDInsight Tools for Visual Studio 3. HDInsight REST API

- ○

1. Azure Data Factory 2. Hive View 3. NoSQL database

○

1. HDInsight Tools for Visual Studio 2. Hive View 3. HDInsight REST API

Explanation

Azure HDInsight Tools for Visual Studio Code is an extension in the Visual Studio Code Marketplace "for developing Hive Interactive Query, Hive Batch Job and PySpark Job against Microsoft HDInsight." The company describes **Azure HDInsight** as an enterprise-grade service for open source analytics

https://visualstudiomagazine.com/articles/2019/01/25/vscode-hdinsight.aspx

Learn how to run Hive queries by using Apache Ambari Hive View. The Hive View allows you to author, optimize, and run Hive queries from your web browser.

https://docs.microsoft.com/en-us/azure/hdinsight/hadoop/apache-hadoop-use-hive-ambari-view

https://docs.microsoft.com/en-us/rest/api/hdinsight/

Question 40: Skipped
You manage security for a database that supports a line of business application.

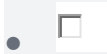**Private and personal data stored in the database must be protected and encrypted.**

You need to configure the database to use Transparent Data Encryption (TDE).

Which five actions should you perform?

- ☐

  Set the context to master database

  (Correct)

- ☐

  Set the context to company database

- ☐

  Create a master key using a password

  (Correct)

- ☐

Create or obtain a certificate protected by the master key

(Correct)

- ☐

Create a database encryption key and protect it by using the certificate

(Correct)

- ☐

Set the database to use encryption.

(Correct)

Explanation

To use TDE, follow these steps.

Applies to: SQL Server.

Create a master key.

Create or obtain a certificate protected by the master key.

Create a database encryption key and protect it by using the certificate.

Set the database to use encryption.

https://docs.microsoft.com/en-us/sql/relational-databases/security/encryption/transparent-data-encryption?view=sql-server-ver15

Question 41: Skipped
A company is planning to use Microsoft Azure Cosmos DB as the data store for an application. You have the following Azure CLI command:

az cosmosdb create -Cname "cosmosdbdev1" C-resource-group "rgdev"

You need to minimize latency and expose the SQL API.

How should you complete the command?

Larger image

Parameter	Value
`--default-consistency-level`	Strong / Session / Eventual / Bounded staleness ▾
`--kind`	Parse / MongoDB / GlobalDocumentDB ▾

- ○

  Session
  MongoDB

- ○

  Eventual
  GlobalDocumentDB

  (Correct)

- ○

  Strong
 GlobalDocumentDB

- ○

  Bounded staleness
  Parse

- ○

  Strong
  Eventual

Explanation

Create a SQL API Cosmos DB account with session consistency and multi-region writes enabled

az cosmosdb create \

--resource-group $resourceGroupName \

--name $accountName \

**--kind GlobalDocumentDB **

--locations regionName="South Central US" failoverPriority=0

--locations regionName="North Central US" failoverPriority=1 \

--default-consistency-level "Session" \

--enable-multiple-write-locations true

https://docs.microsoft.com/en-us/azure/cosmos-db/consistency-levels
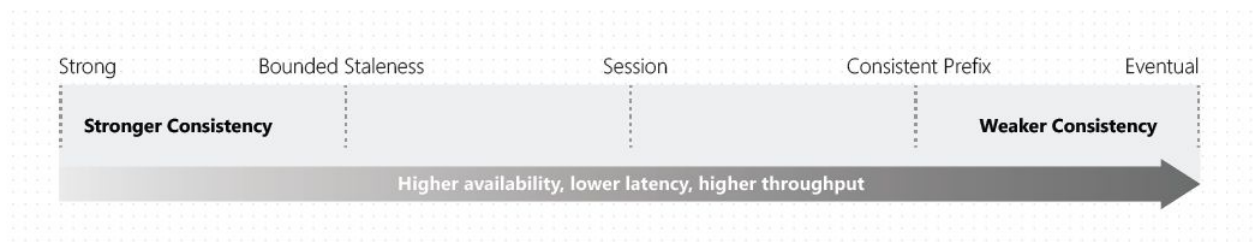
Question 42: Skipped
You need to implement complex stateful business logic within an Azure Stream Analytics service.
Which type of function should you create in the Stream Analytics topology?

○

JavaScript user-define functions (UDFs)

○

Azure Machine Learning

○

JavaScript user-defined aggregates (UDA)

(Correct)

Explanation

Azure Stream Analytics supports user-defined aggregates (UDA) written in JavaScript, it enables you to implement complex stateful business logic.

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-javascript-user-defined-aggregates

Practice Test 6

Question 1: Skipped
You develop a data ingestion process that will import data to a Microsoft Azure SQL Data Warehouse. The data to be ingested resides in parquet files stored in an

Azure Data Lake Gen 2 storage account.

You need to load the data from the Azure Data Lake Gen 2 storage account into the Azure SQL Data Warehouse.

Solution:

1. Use Azure Data Factory to convert the parquet files to CSV files

2. Create an external data source pointing to the Azure storage account

3. Create an external file format and external table using the external data source

4. Load the data using the INSERT"¦SELECT statement

Does the solution meet the goal?

- ○

  Yes

- ○

  No

  (Correct)

Explanation

There is no need to convert the parquet files to CSV files.

You load the data using the CREATE TABLE AS SELECT statement.

Question 2: Skipped
You develop a data ingestion process that will import data to a Microsoft Azure SQL Data Warehouse. The data to be ingested resides in parquet files stored in an

Azure Data Lake Gen 2 storage account.

You need to load the data from the Azure Data Lake Gen 2 storage account into the Azure SQL Data Warehouse.

Solution:

1. Create an external data source pointing to the Azure storage account

2. Create a workload group using the Azure storage account name as the pool name

3. Load the data using the INSERT"¦SELECT statement

Does the solution meet the goal?

- ◯

  Yes

- ◯

  No

  (Correct)

Explanation

You need to create an external file format and external table using the external data source.

You then load the data using the CREATE TABLE AS SELECT statement.

Question 3: Skipped
You develop data engineering solutions for a company.
You must integrate the company's on-premises Microsoft SQL Server data with Microsoft Azure SQL Database. Data must be transformed incrementally.
You need to implement the data integration solution.
Which tool should you use to configure a pipeline to copy data?

- ◯

  Use the Copy Data tool with Blob storage linked service as the source

- ◯

  Use Azure PowerShell with SQL Server linked service as a source

- ◯

○

Use the .NET Data Factory API with Blob storage linked service as the source

Explanation

The Integration Runtime is a customer managed data integration infrastructure used by Azure Data Factory to provide data integration capabilities across different network environments.

A linked service defines the information needed for Azure Data Factory to connect to a data resource. We have three resources in this scenario for which linked services are needed:

☞ On-premises SQL Server

☞ Azure Blob Storage

☞ Azure SQL database

Question 4: Skipped
You develop data engineering solutions for a company.

A project requires analysis of real-time Twitter feeds. Posts that contain specific keywords must be stored and processed on Microsoft Azure and then displayed by using Microsoft Power BI. You need to implement the solution.

Which five 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.

Larger image

○

Step 1: Create an HDInisght cluster with the Hadoop cluster type
Step 2: Create a Jyputer Notebook
Step 3: Create a table
Step 4: Run a job that uses the Spark Streaming API to ingest data from Twitter
Step 5: Load the hvac table into Power BI Desktop

○

Step 1: Create an HDInisght cluster with the Spark cluster type
Step 2: Create a Jyputer Notebook

Step 3: Create a table
Step 4: Run a job that uses the Spark Streaming API to ingest data from Twitter
Step 5: Load the hvac table into Power BI Desktop

(Correct)

- ○

Step 1: Create an HDInisght cluster with the Spark cluster type
Step 2: Create a Runbook
Step 3: Create a table
Step 4: Run a job that uses the Spark Streaming API to ingest data from Twitter
Step 5: Load the hvac table into Power BI Desktop

Explanation

Step 1: Create an HDInisght cluster with the Spark cluster type

Step 2: Create a Jyputer Notebook

Step 3: Create a table -

The Jupyter Notebook that you created in the previous step includes code to create an hvac table.

Step 4: Run a job that uses the Spark Streaming API to ingest data from Twitter

Step 5: Load the hvac table into Power BI Desktop

You use Power BI to create visualizations, reports, and dashboards from the Spark cluster data.

Question 5: Skipped
You are designing a new Lambda architecture on Microsoft Azure.

The real-time processing layer must meet the following requirements:

Ingestion:

☞ **Receive millions of events per second**

☞ **Act as a fully managed Platform-as-a-Service (PaaS) solution**

☞ **Integrate with Azure Functions**

Stream processing:

☞ **Process on a per-job basis**

☞ **Provide seamless connectivity with Azure services**

☞ **Use a SQL-based query language**

Analytical data store:

✎ **Act as a managed service**

✎ **Use a document store**

✎ **Provide data encryption at rest**

You need to identify the correct technologies to build the Lambda architecture using minimal effort. Which technologies should you use? To answer, select the appropriate options in the answer area.

Larger image

- ○

  Box 1: HDInsights Kafka
  Box 2: Azure Stream Analytics
  Box 3: Cosmos DB

- ○

  Box 1: HDInsights Kafka
  Box 2: Azure Stream Analytics
  Box 3: SQL Data Warehouse

- ○

  Box 1: Azure Event Hubs
  Box 2: Azure Stream Analytics
  Box 3: Cosmos DB

  (Correct)

Question 6: Skipped
You develop data engineering solutions for a company.

You need to ingest and visualize real-time Twitter data by using Microsoft Azure.

Which three technologies should you use? Each correct answer presents part of the solution.

- ☐

  A. Event Grid topic

- ☐

B. Azure Stream Analytics Job that queries Twitter data from an Event Hub

(Correct)

- ☐

C. Azure Stream Analytics Job that queries Twitter data from an Event Grid

- ☐

D. Logic App that sends Twitter posts which have target keywords to Azure

(Correct)

- ☐

E. Event Grid subscription

- ☐

F. Event Hub instance

(Correct)

Explanation

You can use Azure Logic apps to send tweets to an event hub and then use a Stream Analytics job to read from event hub and send them to PowerBI.

Question 7: Skipped
Each day, company plans to store hundreds of files in Azure Blob Storage and Azure Data Lake Storage. The company uses the parquet format.

You must develop a pipeline that meets the following requirements:

Process data every six hours

Offer interactive data analysis capabilities

Offer the ability to process data using solid-state drive (SSD) caching

Use Directed Acyclic Graph(DAG) processing mechanisms

Provide support for REST API calls to monitor processes

Provide native support for Python

Integrate with Microsoft Power BI

You need to select the appropriate data technology to implement the pipeline.

Which data technology should you implement?

- ○

  HDInsight Spark cluster

- ○

  HDInsight Apache Hadoop cluster using MapReduce

- ○

  Azure Stream Analytics

- ○

  HDInsight Apache Storm cluster

  (Correct)

- ○

  Azure SQL Data warehouse

Explanation

Storm runs topologies instead of the Apache Hadoop MapReduce jobs that you might be familiar with. Storm topologies are composed of multiple components that are arranged in a directed acyclic graph (DAG).

https://docs.microsoft.com/en-us/azure/hdinsight/storm/apache-storm-overview

Question 8: Skipped
You regularly perform automated data loads to DW1.

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

What should you do?

- ○

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

○

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

○

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

○

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

(Correct)

Explanation

To ensure the loading user has enough memory to achieve maximum compression rates, use loading users that are a member of a medium or large resource class.

https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/guidance-for-loading-data

Question 9: Skipped

You configure monitoring for a Microsoft Azure SQL Data Warehouse implementation. The implementation uses PolyBase to load data from comma-separated value (CSV) files stored in Azure Data Lake Gen 2 using an external table.
Files with an invalid schema cause errors to occur.
You need to monitor for an invalid schema error.
For which error should you monitor?

○

EXTERNAL TABLE access failed due to internal error: 'Java exception raised on call to HdfsBridge_Connect: Error [com.microsoft.polybase.client.KerberosSecureLogin] occurred while accessing external file.'

○

EXTERNAL TABLE access failed due to internal error: 'Java exception raised on call to HdfsBridge_Connect: Error [No FileSystem for scheme: wasbs] occurred while accessing external file.'

○

Cannot execute the query "Remote Query" against OLE DB provider "SQLNCLI11": for linked server "(null)", Query aborted- the maximum reject threshold (o rows) was reached while reading from an external source: 1 rows rejected out of total 1 rows processed.

(Correct)

○

EXTERNAL TABLE access failed due to internal error: 'Java exception raised on call to HdfsBridge_Connect: Error [Unable to instantiate LoginClass] occurred while accessing external file.'

Explanation

The reason this error happens is because each file has different schema. The PolyBase external table DDL when pointed to a directory recursively reads all the files in that directory. When a column or data type mismatch happens, this error could be seen in SSMS.

Question 10: Skipped
You have a container named Sales in an Azure Cosmos DB database. Sales has 120 GB of data. Each entry in Sales has the following structure.

Larger image

The partition key is set to the OrderId attribute.

Users report that when they perform queries that retrieve data by ProductName, the queries take longer than expected to complete.

You need to reduce the amount of time it takes to execute the problematic queries.

Solution: You increase the Request Units (RUs) for the database.

Does this meet the goal?

○

Yes

(Correct)

○

No

Explanation

Provisioned throughput mode: In this mode, you provision the number of RUs for your application on a per-second basis in increments of 100 RUs per second. To scale the provisioned throughput for your application, you can increase or decrease the number of RUs at any time in increments or decrements of 100 RUs.

https://docs.microsoft.com/en-us/azure/cosmos-db/request-units

Question 11: Skipped
You implement an Azure SQL Data Warehouse instance.

You plan to migrate the largest fact table to Azure SQL Data Warehouse. The table resides on Microsoft SQL Server on-premises and is 10 terabytes (TB) is size.

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

Larger image

Larger image

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

Which technology should you use?

- ○

  Round robin distributed table with clustered index

- ○

  heap table with distribution replicate

- ○

  hash distributed table with clustered index

- ○

  hash distributed table with clustered ColumnStore index

    (Correct)

- ○

  round robin distributed table with clustered ColumnStore index

Explanation

Hash-distributed tables work well for large fact tables in a star schema. They can have very large numbers of rows and still achieve high performance.

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

A columnstore index can provide a very high level of data compression, typically by 10 times, to significantly reduce your data warehouse storage cost. For analytics, a columnstore index offers an order of magnitude better performance than a btree index.

https://docs.microsoft.com/en-us/sql/relational-databases/indexes/columnstore-indexes-overview?view=sql-server-ver15

Question 12: Skipped
You are the data engineer for your company. An application uses a NoSQL database to store data. The database uses the key-value and wide-column NoSQL database type.
Developers need to access data in the database using an API.
You need to determine which API to use for the database model and type.
Which two APIs should you use?

- ☐

  Table API

- ☐

  MongoDB API

- ☐

  Gremlin API

- ☐

  SQL API

  (Correct)

- ☐

  Cassandra API

  (Correct)

Explanation

This question is not worded correctly and hence the answer is debatable.

For column type database we need to use Cassandra API

For key value we can use both Table API and SQL API. Now table API is more used when on prem t cloud migration is happening when the on prem database is Table NO SQL type. SQL API is preferred if we build the database from scratch on cloud. Hence can't be said with confidence what is the answer Microsoft looking for.

https://docs.microsoft.com/en-us/azure/architecture/guide/technology-choices/data-store-overview

Question 13: Skipped

A company plans to use Azure SQL Database to support a mission-critical application.

The application must be highly available without performance degradation during maintenance windows.

You need to implement the solution.

Which three technologies should you implement? Each correct answer presents part of the solution.

- ☐

  SQL Data Sync

- ☐

  Always On availability groups

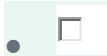 **(Correct)**

- ☐

  Zone-redundant configuration

  (Correct)

- ☐

  Basic service tier

- ☐

Virtual machine Scale Sets

☐

Premium service tier

(Correct)

Explanation

Azure SQL Database Premium tier supports multiple redundant replicas for each database that are automatically provisioned in the same datacenter within a region.

This design leverages the SQL Server AlwaysON technology and provides resilience to server failures with 99.99% availability SLA and RPO=0

https://azure.microsoft.com/en-in/blog/azure-sql-database-now-offers-zone-redundant-premium-databases-and-elastic-pools/

Question 14: Skipped
You are a data architect. The data engineering team needs to configure a synchronization of data between an on-premises Microsoft SQL Server database to Azure SQL

Database. Ad-hoc and reporting queries are being overutilized the on-premises production instance. The synchronization process must:

Perform an initial data synchronization to Azure SQL

Database with minimal downtime Perform bi-directional data synchronization after initial synchronization

You need to implement this synchronization solution. Which synchronization method should you use?

○

backup and restore

○

SQL Server Agent job

○

Azure SQL Data Sync

(Correct)

- ○

  transactional replication

- ○

  Data Migration Assistant (DMA)

Explanation

SQL Data Sync is a service built on Azure SQL Database that lets you synchronize the data you select bi-directionally across multiple databases, both on-premises and in the cloud.

https://docs.microsoft.com/en-us/azure/azure-sql/database/sql-data-sync-data-sql-server-sql-database

Question 15: Skipped

You are working as a data engineer for your company that is currently using Azure Stream Analytics to monitor devices. Your company is planning on developing more devices and all of these devices need to be monitored via the same Azure Stream Analytics instance. You also have to be ensured that there are enough processing resources to handle the load of the additional devices. What metric should you be tracking for the Stream Analytics job for this?

- ○

  Input Deserialization Errors

- ○

  Early Input Events

- ○

  Late Input Events

- ○

  Watermark delay

  (Correct)

Explanation

The watermark represents a specific timestamp in the event time timeline. This timestamp is used as a pointer or indicator of progress in the temporal computations. For example, when

Stream Analytics reports a certain watermark value at the output, it guarantees that all events prior to this timestamp were already computed. Watermark can be used as an indicator of liveliness for the data produced by the job. If the delay between the current time and watermark is small, it means the job is keeping up with the incoming data and produces results defined by the query on time.

https://azure.microsoft.com/en-in/blog/new-metric-in-azure-stream-analytics-tracks-latency-of-your-streaming-pipeline/

Question 16: Skipped

You are working as a data engineer for your company that has on-premise Microsoft SQL Server databases at several locations. The company wants to integrate the data in the databases with Microsoft Power BI and Microsoft Azure Logic Apps. You have been given the task to implement a solution that will avoid any single points of failure during the connection and transfer of data to the cloud. The latency should also be kept minimum. The transfer of data between the on-premise databases and Microsoft Azure has to be secure. What should you be implementing for this?

- ○ Install a standalone on-premise Azure data gateway at each company location

- ○ Install an on-premise data gateway in personal mode at each company location

- ○ Install an Azure on-premise data gateway at the primary company location

- ○ Install an Azure on-premise data gateway as a cluster at each location

  (Correct)

Explanation

You can create high-availability clusters of gateway installations. The clusters help ensure that your organization can access on-premises data resources from cloud services like Power BI and Power Apps. Gateway admins use such clusters to avoid single points of failure when accessing on-premises data resources.

https://docs.microsoft.com/en-us/data-integration/gateway/service-gateway-high-availability-clusters

Question 17: Skipped

You are working as a data engineer for a company that has an Azure Data Lake Storage Account. They have decided to implement role-based access control (RBAC) so that project members can manage the Azure Data Lake Storage resources. What actions should you be doing for this?

Select 3 options from options given below.

- ☐

  Ensure to assign Azure AD security groups to Azure Data Lake Storage

  (Correct)

- ☐

  Make sure to configure end-user authentication to the Azure Data Lake Storage account

- ☐

  Make sure to configure service-to-service authentication to the Azure Data Lake Storage account

- ☐

  Create security groups in Azure AD and then add the project members

  (Correct)

- ☐

  Configure Access control lists for the Azure Data Lake Storage account

  (Correct)

Explanation

Azure RBAC uses role assignments to apply sets of permissions to security principals. A security principal is an object that **represents a user, group, service principal, or managed identity that is defined in Azure Active Directory (AD).**

https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-access-control-model

You can associate a security principal with an access level for files and directories. These associations are captured in an *access control list (ACL)*. Each file and directory in your storage account has an access control list.

https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-access-control

Question 18: Skipped
You are working as a data engineer for your company that has implemented a real-time data analysis solution. The solution is using Azure Event Hub to ingest the data. Then, this data is sent to Azure Stream Analytics cloud job which has been configured to use 100 Streaming Units. What 2 actions should you be performing to optimize the performance of the Azure Stream Analytics job?

- ☐

    Scale up the Streaming Units of the job

    (Correct)

- ☐

    Make use of event ordering

- ☐

    Make use of Azure Stream Analytics user-defined functions

- ☐

    Implement query parallelization by partitioning the data input

    (Correct)

Explanation

This article shows you how to tune a Stream Analytics query to increase throughput for Streaming Analytics jobs.

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-scale-jobs

This article shows you how to take advantage of parallelization in Azure Stream Analytics. You learn how to scale Stream Analytics jobs by configuring input partitions and tuning the analytics query definition.

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-parallelization

Question 19: Skipped

You are working as a data engineer for your company and you have been given the task to create an Azure Data Lake Gen 2 storage account. Your company wants to ingest data into the storage account from various data sources.

What should you be using to ingest data from a relational data store?

- ○

  Azure Data Factory

  (Correct)

- ○

  AzCopy Tool

- ○

  Azure Event Hubs

- ○

  Azure Event Grid

Explanation

You can also source data from relational databases. Over a period of time, relational databases collect huge amounts of data which can provide key insights if processed through a big data pipeline. You can use the following tools to move such data into Data Lake Storage Gen2.

Here's a list of tools that you can use to ingest relational data.

Tool	Guidance
Azure Data Factory	Copy Activity in Azure Data Factory

https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-data-scenarios

Question 20: Skipped
You are working as a data engineer for your company and you have been given the task to create an Azure Data Lake Gen 2 storage account. Your company wants to ingest data into the storage account from various data sources.

What should you be using to ingest data from a local workstation?

- ○

Azure Data Factory

○

AzCopy Tool

(Correct)

○

Azure Event Hubs

○

Azure Event Grid

Explanation

Data Source	Ingest it using
Local computer	Azure PowerShell
	Azure CLI
	Storage Explorer
	AzCopy tool

https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-data-scenarios

Question 21: Skipped

A company has created an Azure Data Lake Gen 2 storage account. They want to ingest data into the storage account from various data sources.

Which of the following can they use to ingest data from log data stored on web servers?

○

Azure Data Factory

(Correct)

○

AzCopy Tool

- ○

  Azure Event Hubs

- ○

  Azure Event Grid

Explanation

Here's a list of tools that you can use to ingest Web server log data.

Tool	Guidance
Azure Data Factory	Copy Activity in Azure Data Factory
Azure CLI	Azure CLI
Azure PowerShell	Azure PowerShell

https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-data-scenarios

Question 22: Skipped
You have data stored in thousands of CSV files in Azure Data Lake Storage Gen2. Each file has a header row followed by a property formatted carriage return (/r) and line feed (/n).

You are implementing a pattern that batch loads the files daily into an Azure SQL data warehouse by using PolyBase.

You need to skip the header row when you import the files into the data warehouse.

Which three actions should you perform?

- ☐

  Create an external data source that uses abfs location

  (Correct)

- ☐

  Create an external data source that uses hadoop location

- ☐

  Create an external file format and set the the First_Row option

  (Correct)

- ☐

  Create a database scoped credential that uses an OAuth2 token and key

  (Correct)

- ☐

  Use Create External Table as Select (CETAS) and create a new view that removes the empty row

Explanation

-- Create a db master key if one does not already exist, using your own password.

CREATE MASTER KEY ENCRYPTION BY PASSWORD='<EnterStrongPasswordHere>';

-- Create a database scoped credential.

CREATE DATABASE SCOPED CREDENTIAL ADL_User

WITH

IDENTITY = '<client_id>@\<OAuth_2.0_Token_EndPoint>',

SECRET = '<key>'

;

https://docs.microsoft.com/en-us/sql/t-sql/statements/create-database-scoped-credential-transact-sql?view=aps-pdw-2016-au7

https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/load-data-overview

CETAS option is incorrect because its used for exporting data and not importing

https://azure.microsoft.com/en-in/blog/october-update-for-azure-sql-data-warehouse/

Question 23: Skipped
You have an Azure Data Lake Storage Gen2 account that contains JSON files for customers. The files contain two attributes named FirstName and LastName.

You need to copy the data from the JSON files to an Azure SQL data Warehouse table by using Azure Databricks. A new column must be created that concatenates the FirstName and LastName values.

You create the following components:

☞ A destination table in SQL Data Warehouse

☞ An Azure Blob storage container

☞ A service principal

Which five actions should you perform in sequence next in a Databricks notebook?

A. Write the results to Data lake storage

B. Drop the data frame

C. Perform transformations on the data frame

D. Mount the Data Lake storage onto DBFS

E. Perform transformation on the file

F. Read the file into the dataframe

G. Specify the temporary folder to stage the data

H. Write the results to a table in SQL Data Warehouse

- ○

  D,G,F,C,H

  (Correct)

- ○

  G,F,C,H,B

- ○

  D,G,F,E,H

Explanation

Create an Azure Databricks service.

Create a Spark cluster in Azure Databricks.

Create a file system in the Data Lake Storage Gen2 account.

Upload sample data to the Azure Data Lake Storage Gen2 account.

Create a service principal.-This step includes mounting the DBFS file system and creating temporary storage

Extract data from the Azure Data Lake Storage Gen2 account.

Transform data in Azure Databricks.

Load data into Azure Synapse.

Question 24: Skipped

You plan to perform batch processing in Azure Databricks once daily.
Which type of Databricks cluster should you use?

- ○

  job

  (Correct)

- ○

  interactive

- ○

  High Concurrency

Explanation

This scenario involves running batch job JARs and notebooks on a regular cadence through the Databricks platform.

The suggested best practice is to launch a new cluster for each run of critical jobs. This helps avoid any issues (failures, missing SLA, and so on) due to an existing workload (noisy neighbor) on a shared cluster. Depending on the level of criticality for the job you could go full on-demand (to meet SLAs) or even balance between spot and on-demand instances (with **Spot fall back to On-demand** option enabled for the cluster) for some cost savings.

https://docs.databricks.com/administration-guide/capacity-planning/cmbp.html#scenario-3-scheduled-batch-workloads-data-engineers-running-etl-jobs

Question 25: Skipped
You need to deploy a Microsoft Azure Stream Analytics job for an IoT solution. The solution must:

✆ **Minimize latency.**

✆ **Minimize bandwidth usage between the job and IoT device.**

Which four actions should you perform in sequence?

1. Configure routes

2. Create a data lake storage container

3. Create a blob storage container

4. Create Streaming Units

5. Create an Azure Stream Analytics Edge job and configure job definition and save location

6. Create an Azure Stream Analytics Cloud job and configure job definition and save location

7. Create an IOT Hub and add the Azure Stream Analytics module to the IOT Hub namespace

- ○

    2-->4-->7-->1

- ○

    3-->5-->7-->1

    (Correct)

- ○

    2-->6-->7-->1

- ○

    5-->3-->7-->1

Explanation

Refer the process here

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-edge

Question 26: Skipped
A company uses Microsoft Azure SQL Database to store sensitive company data. You encrypt the data and only allow access to specified users from specified locations.

You must monitor data usage, and data copied from the system to prevent data leakage.

You need to configure Azure SQL Database to email a specific user when data leakage occurs.

Which three actions should you perform

- ☐

    In Auditing, enable Auditing.

- ☐

  Configure the service to create alters for threat detections of type Data Exfiltration

  (Correct)

- ☐

  In Firewalls and virtual networks, enable Allow access to Azure services

- ☐

  Enable advanced threat protection

  (Correct)

- ☐

  Configure the service to send email alerts to security@contoso.com

  (Correct)

Explanation

Set up Advanced Threat Protection in the Azure portal

Sign into the Azure portal.

Navigate to the configuration page of the server you want to protect. In the security settings, select **Security center**.

On the **Azure Defender** configuration page:

Enable Azure Defender on the server.

In **Advanced Threat Protection Settings**, in the **Send alerts to** text box, provide the list of emails to receive security alerts upon detection of anomalous database activities.

https://docs.microsoft.com/en-us/azure/azure-sql/database/threat-detection-configure#:~:text=In%20the%20security%20settings%2C%20select,detection%20of%20anomalous%20database%20activities.

Question 27: Skipped
You develop data engineering solutions for a company. An application creates a database on Microsoft Azure. You have the following code:

Larger image

```
private static readonly string endpointUrl =
ConfigurationManager.AppSettings["EndPointUrl"];

private static readonly SecureString authorizationKey =
ToSecureString(ConfigurationManager.AppSettings
["AuthorizationKey"]);

var client = new DocumentClient(new Url(endpointUrl),
authorizationKey);

Database database = await client.CreateDatabaseAsync(

new Database

{

Id = <DatabaseName>

});
```

Which database and authorization types are used?

Azure database type:

Azure Cosmos DB

Azure SQL Database

Files Blob

Key type:

Resource token

Master key

Certificate

- ○

  Azure Cosmos DB ,Master key

    (Correct)

- ○

Azure SQL Database ,Resource token

- $\bigcirc$

Files Blob ,Certificate

Explanation

DocumentClient.CreateDatabaseAsync(Database, RequestOptions) Method

Definition

Creates a database resource as an asychronous operation in the Azure Cosmos DB service.

https://docs.microsoft.com/en-us/dotnet/api/microsoft.azure.documents.client.documentclient.createdatabaseasync?view=azure-dotnet

https://docs.microsoft.com/en-us/dotnet/api/microsoft.azure.documents.client.documentclient.createdatabaseasync?view=azure-dotnet

Azure DB key information can be found here:

https://docs.microsoft.com/en-us/azure/cosmos-db/secure-access-to-data

Question 28: Skipped
You plan to use Microsoft Azure SQL Database instances with strict user access control. A user object must:
☞ Move with the database if it is run elsewhere
☞ Be able to create additional users
You need to create the user object with correct permissions.
Which two Transact-SQL commands should you run?

- ☐

    ALTER LOGIN Mary WITH PASSWORD = 'strong_password';

- ☐

    CREATE LOGIN Mary WITH PASSWORD = 'strong_password';

- ☐

    ALTER ROLE db_owner ADD MEMBER Mary;

    (Correct)

- ☐

  CREATE USER Mary WITH PASSWORD = 'strong_password';

  (Correct)

- ☐

  GRANT ALTER ANY USER TO Mary;

Explanation

-- Syntax for SQL Server (starting with 2012) and Azure SQL Database

ALTER ROLE role_name

{

   ADD MEMBER database_principal

 | DROP MEMBER database_principal

 | WITH NAME = new_name

}

[;]

https://docs.microsoft.com/en-us/sql/t-sql/statements/alter-role-transact-sql?view=sql-server-ver15

Contained database user with password. (Not available in Azure Synapse Analytics (SQL Data Warehouse).) `CREATE USER Mary WITH PASSWORD = '********';`

https://docs.microsoft.com/en-us/sql/t-sql/statements/create-user-transact-sql?view=sql-server-ver15

Question 29: Skipped
You plan to create a new single database instance of Microsoft Azure SQL Database.

The database must only allow communication from the data engineer's workstation. You must connect directly to the instance by using Microsoft SQL Server Management Studio.

You need to create and configure the Database. Which three Azure PowerShell cmdlets should you use to develop the solution? To answer, move the appropriate cmdlets from the list of cmdlets to the answer area.

- ☐

  New-AzureSqlServer

(Correct)

- ☐

  New-AzureSqlElasticPool

- ☐

  New-AzureRmSqlServerFirewallRule

  (Correct)

- ☐

  New-AzureRmSqlServerVirtualNetworkRule

- ☐

  New-AzureRmSqlDatabase

  (Correct)

Explanation

\# Create a server with a system wide unique server name

$server = New-AzSqlServer -ResourceGroupName $resourceGroupName `

   -ServerName $serverName `

   -Location $location `

   -SqlAdministratorCredentials $(New-Object -TypeName System.Management.Automation.PSCredential -ArgumentList $adminSqlLogin, $(ConvertTo-SecureString -String $password -AsPlainText -Force))

\# Create a server firewall rule that allows access from the specified IP range

$serverFirewallRule = New-AzSqlServerFirewallRule -ResourceGroupName $resourceGroupName `

   -ServerName $serverName `

   -FirewallRuleName "AllowedIPs" -StartIpAddress $startIp -EndIpAddress $endIp

\# Create a blank database with an S0 performance level

$database = New-AzSqlDatabase -ResourceGroupName $resourceGroupName `

   -ServerName $serverName `

   -DatabaseName $databaseName `

   -RequestedServiceObjectiveName "S0" `

   -SampleName "AdventureWorksLT"

\# Clean up deployment

\# Remove-AzResourceGroup -ResourceGroupName $resourceGroupName

https://docs.microsoft.com/en-us/azure/azure-sql/database/scripts/create-and-configure-database-powershell?toc=%2Fpowershell%2Fmodule%2Ftoc.json

Question 30: Skipped
Your company uses Azure SQL Database and Azure Blob storage.

All data at rest must be encrypted by using the company's own key. The solution must minimize administrative effort and the impact to applications which use the database.

You need to configure security. What should you implement?

Larger image

- ○

  A & H

- ○

  A & G

- ○

  D & G

  (Correct)

- ○

  D & H

Explanation

TDE protector is either a service-managed certificate (service-managed transparent data encryption) or an asymmetric key stored in Azure Key Vault (customer-managed transparent data encryption).

https://docs.microsoft.com/en-us/azure/azure-sql/database/transparent-data-encryption-tde-overview?tabs=azure-portal

Data in a new storage account is encrypted with Microsoft-managed keys by default. You can continue to rely on Microsoft-managed keys for the encryption of your data, or you can manage encryption with your own keys. If you choose to manage encryption with your own keys, you have two options. You can use either type of key management, or both:

You can specify a *customer-managed key* to use for encrypting and decrypting data in Blob storage and in Azure Files.1,2 Customer-managed keys must be stored in Azure Key Vault or Azure Key Vault Managed Hardware Security Model (HSM) (preview). For more information about customer-managed keys, see Use customer-managed keys for Azure Storage encryption.

You can specify a *customer-provided key* on Blob storage operations. A client making a read or write request against Blob storage can include an encryption key on the request for granular control over how blob data is encrypted and decrypted. For more information about customer-provided keys, see Provide an encryption key on a request to Blob storage.

https://docs.microsoft.com/en-us/azure/storage/common/storage-service-encryption

Question 31: Skipped
You need to receive an alert when Azure SQL Data Warehouse consumes the maximum allotted resources.
Which resource type and signal should you use to create the alert in Azure Monitor?

Larger image

Answer Area

Resource type:	▼

Resource group
SQL server
SQL data warehouse
Subscription

Signal:	▼

CPU used
Data IO percentage
DWU limit
DWU used

Hide Solution

○

Resource Type : Resource Group
Signal : CPU used

○

Resource Type : SQL Server
Signal : DWU limit

○

Resource Type : SQL data warehouse
Signal : DWU limit

   (Correct)

○

Resource Type : SQL data warehouse
Signal : DWU used

- ○

Resource Type : Resource Group
Signal : DWU used

- ○

Resource Type : Subscription
Signal : Data IO percentage

Explanation

Resource type: SQL data warehouse

DWU limit belongs to the SQL data warehouse resource type.

Signal: DWU limit

SQL Data Warehouse capacity limits are maximum values allowed for various components of Azure SQL Data Warehouse.

Note: There is some confusion in students that this should be "DWU used" but that not the correct answer. DWU limit is a metric in azure monitor which is configured to measure if the resources are using the maximum limit.

DWU used is more preferable when we want to receive an alert at a number lower than the limit. This also means that we can use DWU used to achieve what we are trying to achieve via DWU limit by setting the value of DWU used as DWU limit.

Question 32: Skipped
A company builds an application to allow developers to share and compare code. The conversations, code snippets, and links shared by people in the application are stored in a Microsoft Azure SQL Database instance. The application allows for searches of historical conversations and code snippets.

When users share code snippets, the code snippet is compared against previously share code snippets by using a combination of Transact-SQL functions including SUBSTRING, FIRST_VALUE, and SQRT. If a match is found, a link to the match is added to the conversation.

Customers report the following issues:

✎ Delays occur during live conversations

✎ A delay occurs before matching links appear after code snippets are added to conversations

You need to resolve the performance issues.

Which technologies should you use?

Options : -columnstoreindex

-non-durable table

-materialized view

-memory-optimized table

○

Delay in conversations : columnstoreindex
Delay in matched links : non-durable table

○

Delay in conversations : memory-optimized table
Delay in matched links : materialized view

(Correct)

○

Delay in conversations : memory-optimized table
Delay in matched links : columnstoreindex

○

Delay in conversations : non-durable table
Delay in matched links : columnstoreindex

Explanation

In-Memory OLTP can significantly improve the performance of transaction processing, data ingestion and data load, and transient data scenarios.

https://docs.microsoft.com/en-us/sql/relational-databases/in-memory-oltp/in-memory-oltp-in-memory-optimization?view=sql-server-ver15

Generate prepopulated views over the data in one or more data stores when the data isn't ideally formatted for required query operations. This can help support efficient querying and data extraction, and improve application performance.

https://docs.microsoft.com/en-us/azure/architecture/patterns/materialized-view

Question 33: Skipped

You have an Azure data solution that contains an Azure SQL data warehouse named DW1.
Several users execute adhoc queries to DW1 concurrently.
You regularly perform automated data loads to DW1.
You need to ensure that the automated data loads have enough memory available to complete quickly and successfully when the adhoc queries run
What should you do?

- ○

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

- ○

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

  (Correct)

- ○

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

- ○

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

Explanation

To ensure the loading user has enough memory to achieve maximum compression rates, use loading users that are a member of a medium or large resource class.

https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/guidance-for-loading-data

Question 34: Skipped
You have an Azure virtual machine that has Microsoft SQL Server installed. The server contains a table named Table1.
You need to copy the data from Table1 to an Azure Data Lake Storage Gen2 account by using an Azure Data Factory V2 copy activity.
Which type of integration runtime should you use?

- ○

  Azure integration runtime

- ○

  self-hosted integration runtime

  (Correct)

- ○

  Azure-SSIS integration runtime

Explanation

A self-hosted integration runtime can run copy activities between a cloud data store and a data store in a private network. It also can dispatch transform activities against compute resources in an on-premises network or an Azure virtual network.

https://docs.microsoft.com/en-us/azure/data-factory/create-self-hosted-integration-runtime

Azure Virtual Network (VNet) is the fundamental building block for your private network in Azure.

https://docs.microsoft.com/en-us/azure/virtual-network/virtual-networks-overview

Question 35: Skipped
Your company manages a payroll application for its customers worldwide. The application uses an Azure SQL database named DB1. The database contains a table named Employee and an identity column named EmployeeId.
A customer requests the EmployeeId be treated as sensitive data.
Whenever a user queries EmployeeId, you need to return a random value between 1 and 10 instead of the EmployeeId value.
Which masking format should you use?

- ○

  string

- ○

  number

  (Correct)

- ○

  default

Explanation

https://docs.microsoft.com/en-us/azure/azure-sql/database/dynamic-data-masking-configure-portal

Question 36: Skipped

You have an Azure SQL server named Server1 that hosts two development databases named DB1 and DB2.
You have an administrative workstation that has an IP address of 192.168.8.8. The development team at your company has an IP addresses in the range of 192.168.8.1 to 192.168.8.5.
You need to set up firewall rules to meet the following requirements:
✆ Allows connection from your workstation to both databases.
✆ The development team must be able connect to DB1 but must be prevented from connecting to DB2.
✆ Web services running in Azure must be able to connect to DB1 but must be prevented from connecting to DB2.
Which three actions should you perform?

- ☐

  Create a firewall rule on DB1 that has a start IP address of 192.168.8.1 and an end IP address of 192.168.8.5.
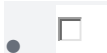
  (Correct)

- ☐

  B. Create a firewall rule on DB1 that has a start and end IP address of 0.0.0.0.

  (Correct)

- ☐

  Create a firewall rule on Server1 that has a start IP address of 192.168.8.1 and an end IP address of 192.168.8.5.

- ☐

  Create a firewall rule on DB1 that has a start and end IP address of 192.168.8.8.

- ☐

  Create a firewall rule on Server1 that has a start and end IP address of 192.168.8.8.

Explanation

When you add a database-level firewall setting where the beginning and ending IP addresses are equal to `0.0.0.0`, you enable access to your database in the SQL Database server from any Azure resource.

https://docs.microsoft.com/en-us/sql/relational-databases/system-stored-procedures/sp-set-database-firewall-rule-azure-sql-database?view=azuresqldb-current

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

Question 37: Skipped

You are designing an enterprise data warehouse in Azure Synapse Analytics. You plan to load millions of rows of data into the data warehouse each day.
You must ensure that staging tables are optimized for data loading.
You need to design the staging tables.
What type of tables should you recommend?

- ○

  Round-robin distributed table

  (Correct)

- ○

  Hash-distributed table

- ○

  Replicated table

- ○

  External table

Explanation

Use round-robin for the staging table. The load with CTAS is fast.

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

Question 38: Skipped

You have an Azure Stream Analytics job.
You need to ensure that the job has enough streaming units provisioned.
You configure monitoring of the SU% Utilization metric.
Which two additional metrics should you monitor?

- ☐

  Watermark Delay

    (Correct)

- ☐

  Late Input Events

- ☐

  Out of order Events

- ☐

  Backlogged Input Events

    (Correct)

- ☐

  Function Events

Explanation

Backlogged Input Events - Number of input events that are backlogged. A non-zero value for this metric implies that your job isn't able to keep up with the number of incoming events. If this value is slowly increasing or consistently non-zero, you should scale out your job. You can learn more by visiting Understand and adjust Streaming Units.

Watermark Delay - Indicates the delay of the streaming data processing job.

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

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