DP-200.examcollection.premium.exam.120q

Number: DP-200
Passing Score: 800
Time Limit: 120 min
File Version: 3.0

ExamCollection
free practice exam collection

DP-200

Implementing an Azure Data Solution

Version 3.0

Question Set 1

QUESTION 1
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.

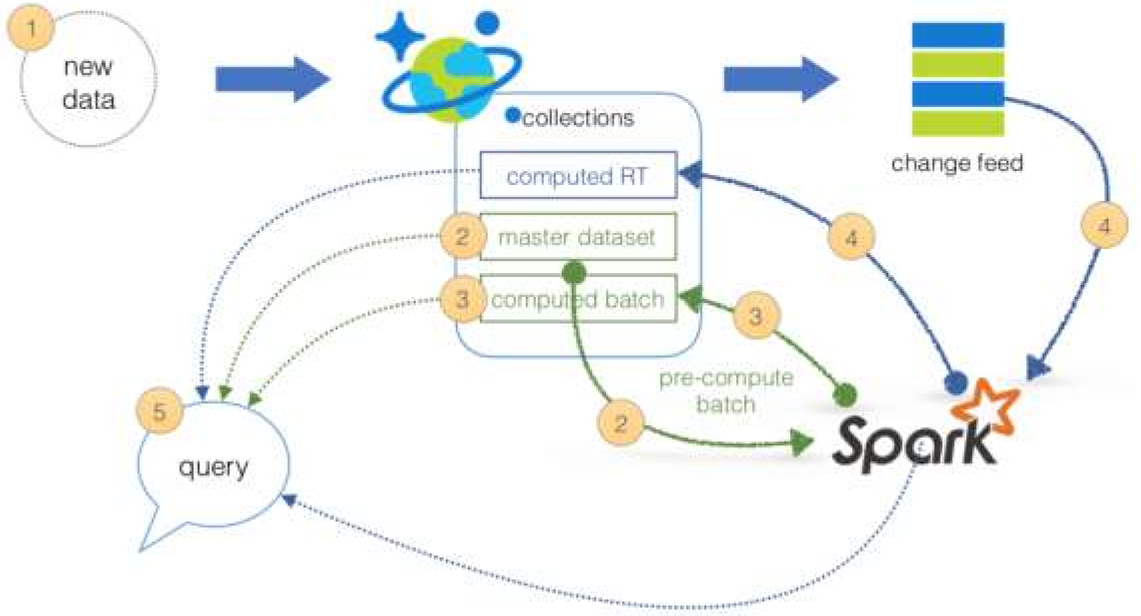
Which type of cluster should you create?

A. Interactive Query
B. Apache Hadoop
C. Apache HBase
D. Apache Spark

Correct Answer: D
Section: [none]
Explanation

Explanation/Reference:
Explanation:
Lambda Architecture with Azure:
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



Note: Lambda architecture is a data-processing architecture designed to handle massive quantities of data by taking advantage of both batch processing and stream processing methods, and minimizing the latency involved in querying big data.

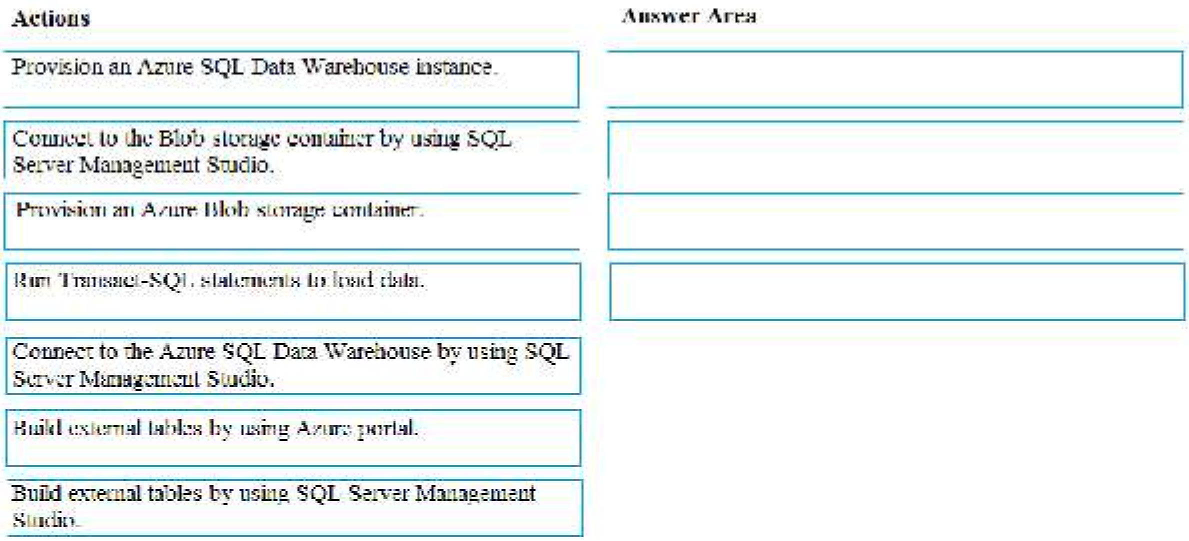
References:
https://sqlwithmanoj.com/2018/02/16/what-is-lambda-architecture-and-what-azure-offers-with-its-new-cosmos-db/

QUESTION 2
DRAG DROP

You develop data engineering solutions for a company. You must migrate data from Microsoft Azure Blob storage to an Azure SQL Data Warehouse for further transformation. You need to implement the solution.

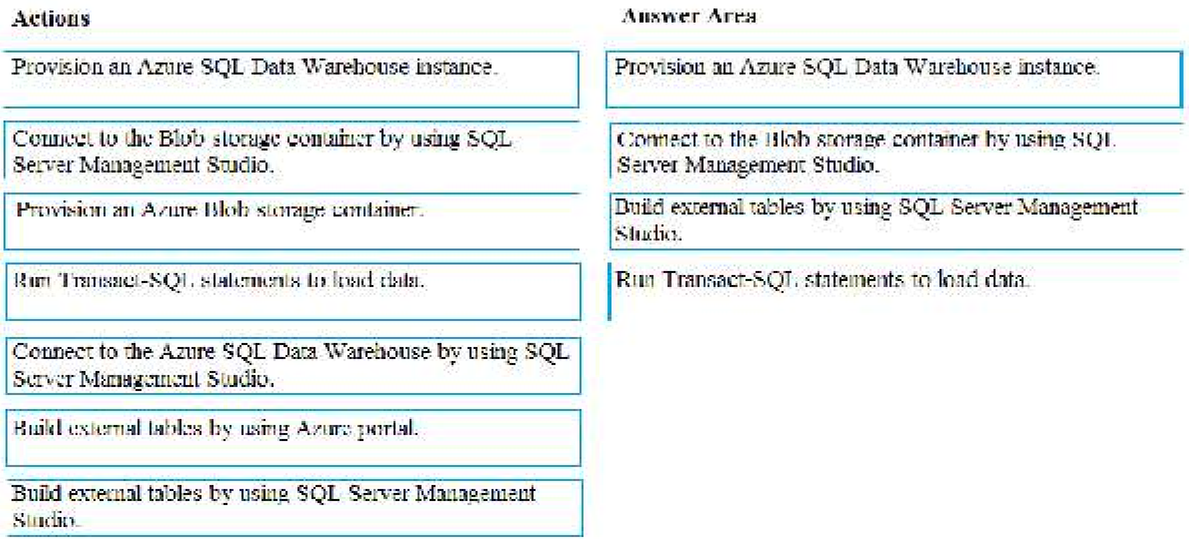
Which four actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

Select and Place:



Actions	Answer Area
Provision an Azure SQL Data Warehouse instance.	
Connect to the Blob storage container by using SQL Server Management Studio.	
Provision an Azure Blob storage container.	
Run Transact-SQL statements to load data.	
Connect to the Azure SQL Data Warehouse by using SQL Server Management Studio.	
Build external tables by using Azure portal.	
Build external tables by using SQL Server Management Studio.	

Correct Answer:



Actions	Answer Area
Provision an Azure SQL Data Warehouse instance.	Provision an Azure SQL Data Warehouse instance.
Connect to the Blob storage container by using SQL Server Management Studio.	Connect to the Blob storage container by using SQL Server Management Studio.
Provision an Azure Blob storage container.	Build external tables by using SQL Server Management Studio.
Run Transact-SQL statements to load data.	Run Transact-SQL statements to load data.
Connect to the Azure SQL Data Warehouse by using SQL Server Management Studio.	
Build external tables by using Azure portal.	
Build external tables by using SQL Server Management Studio.	

Section: [none]

Explanation

Explanation/Reference:
Explanation:

Step 1: Provision an Azure SQL Data Warehouse instance.
Create a data warehouse in the Azure portal.

Step 2: Connect to the Azure SQL Data warehouse by using SQL Server Management Studio
Connect to the data warehouse with SSMS (SQL Server Management Studio)

Step 3: Build external tables by using the SQL Server Management Studio
Create external tables for data in Azure blob storage.
You are ready to begin the process of loading data into your new data warehouse. You use external tables to load data from the Azure storage blob.

Step 4: Run Transact-SQL statements to load data.
You can use the CREATE TABLE AS SELECT (CTAS) T-SQL statement to load the data from Azure Storage Blob into new tables in your data warehouse.

References:
https://github.com/MicrosoftDocs/azure-docs/blob/master/articles/sql-data-warehouse/load-data-from-azure-blob-storage-using-polybase.md

QUESTION 3
You develop data engineering solutions for a company. The company has on-premises Microsoft SQL Server databases at multiple locations.

The company must integrate data with Microsoft Power BI and Microsoft Azure Logic Apps. The solution must avoid single points of failure during connection and transfer to the cloud. The solution must also minimize latency.

You need to secure the transfer of data between on-premises databases and Microsoft Azure.

What should you do?

A. Install a standalone on-premises Azure data gateway at each location
B. Install an on-premises data gateway in personal mode at each location
C. Install an Azure on-premises data gateway at the primary location
D. Install an Azure on-premises data gateway as a cluster at each location

Correct Answer: D
Section: [none]
Explanation

Explanation/Reference:
Explanation:
You can create high availability clusters of On-premises data gateway installations, to ensure your organization can access on-premises data resources used in Power BI reports and dashboards. Such clusters allow gateway administrators to group gateways to avoid single points of failure in accessing on-premises data resources. The Power BI service always uses the primary gateway in the cluster, unless it’s not available. In that case, the service switches to the next gateway in the cluster, and so on.

References:
https://docs.microsoft.com/en-us/power-bi/service-gateway-high-availability-clusters

QUESTION 4
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?

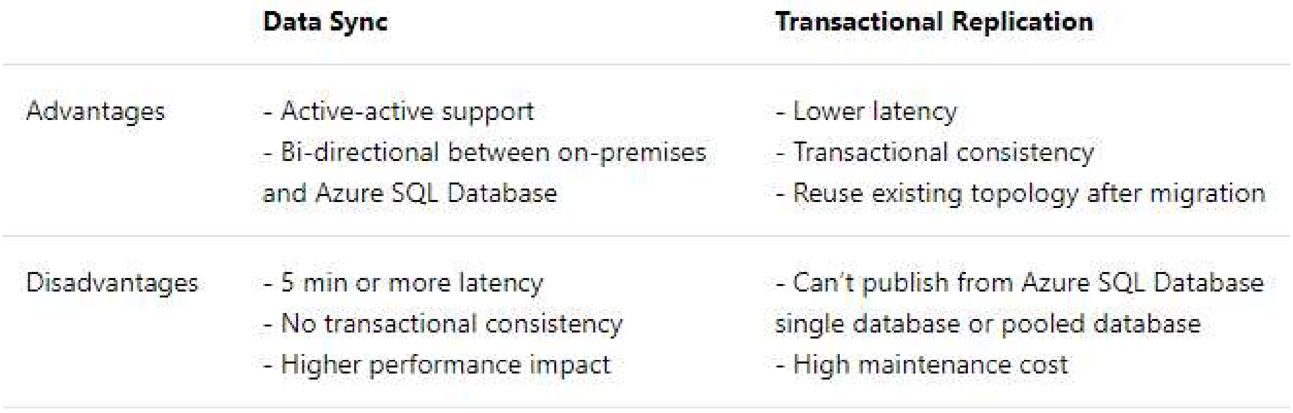
A. transactional replication
B. Data Migration Assistant (DMA)
C. backup and restore
D. SQL Server Agent job
E. Azure SQL Data Sync

Correct Answer: E
Section: [none]
Explanation

Explanation/Reference:
Explanation:
SQL Data Sync is a service built on Azure SQL Database that lets you synchronize the data you select bi-directionally across multiple SQL databases and SQL Server instances.

With Data Sync, you can keep data synchronized between your on-premises databases and Azure SQL databases to enable hybrid applications.

Compare Data Sync with Transactional Replication



	Data Sync	**Transactional Replication**
Advantages	- Active-active support - Bi-directional between on-premises and Azure SQL Database	- Lower latency - Transactional consistency - Reuse existing topology after migration
Disadvantages	- 5 min or more latency - No transactional consistency - Higher performance impact	- Can't publish from Azure SQL Database single database or pooled database - High maintenance cost

References:
https://docs.microsoft.com/en-us/azure/sql-database/sql-database-sync-data

QUESTION 5
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.

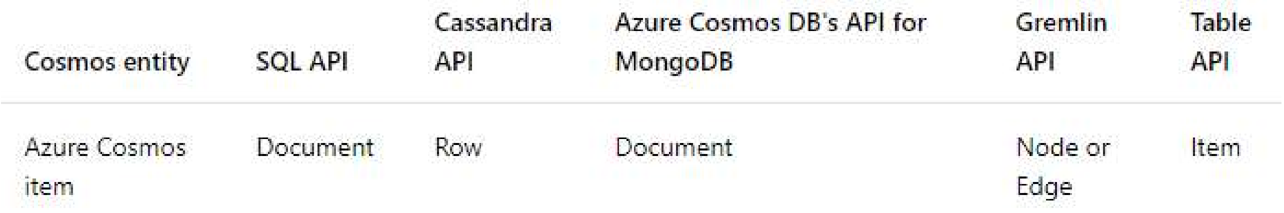
NOTE: Each correct answer selection is worth one point.

A. collection

B. rows
C. graph
D. entities
E. table

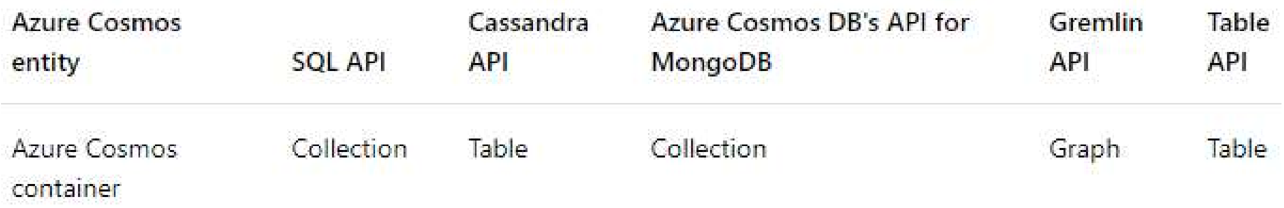
Correct Answer: BE
Section: [none]
Explanation

Explanation/Reference:
Explanation:
B: Depending on the choice of the API, an Azure Cosmos item can represent either a document in a collection, a row in a table or a node/edge in a graph. The following table shows the mapping between API-specific entities to an Azure Cosmos item:



Cosmos entity	SQL API	Cassandra API	Azure Cosmos DB's API for MongoDB	Gremlin API	Table API
Azure Cosmos item	Document	Row	Document	Node or Edge	Item

E: An Azure Cosmos container is specialized into API-specific entities as follows:



Azure Cosmos entity	SQL API	Cassandra API	Azure Cosmos DB's API for MongoDB	Gremlin API	Table API
Azure Cosmos container	Collection	Table	Collection	Graph	Table

References:
https://docs.microsoft.com/en-us/azure/cosmos-db/databases-containers-items

QUESTION 6
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?

A. Number of transactions only
B. eDTUs per database only
C. Number of databases only
D. CPU usage only
E. eDTUs and max data size

Correct Answer: E
Section: [none]
Explanation

Explanation/Reference:
Explanation:

The best size for a pool depends on the aggregate resources needed for all databases in the pool. This involves determining the following:

- Maximum resources utilized by all databases in the pool (either maximum DTUs or maximum vCores depending on your choice of resourcing model).
- Maximum storage bytes utilized by all databases in the pool.

Note: 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.

References:
https://docs.microsoft.com/en-us/azure/sql-database/sql-database-elastic-pool

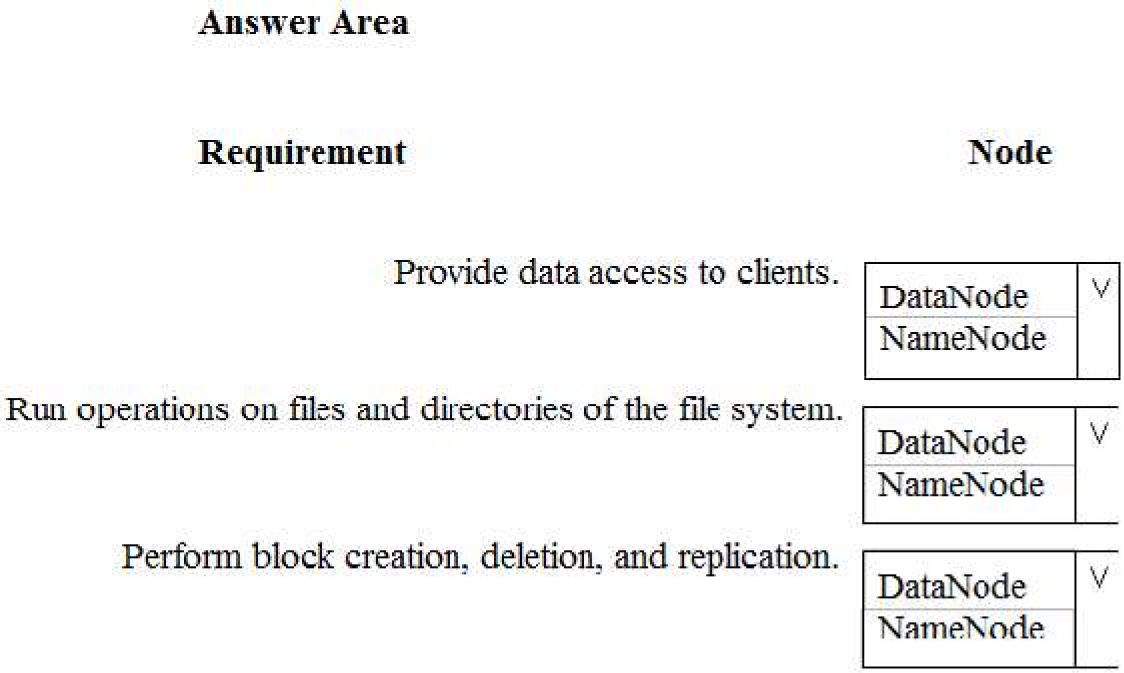
QUESTION 7
HOTSPOT

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.

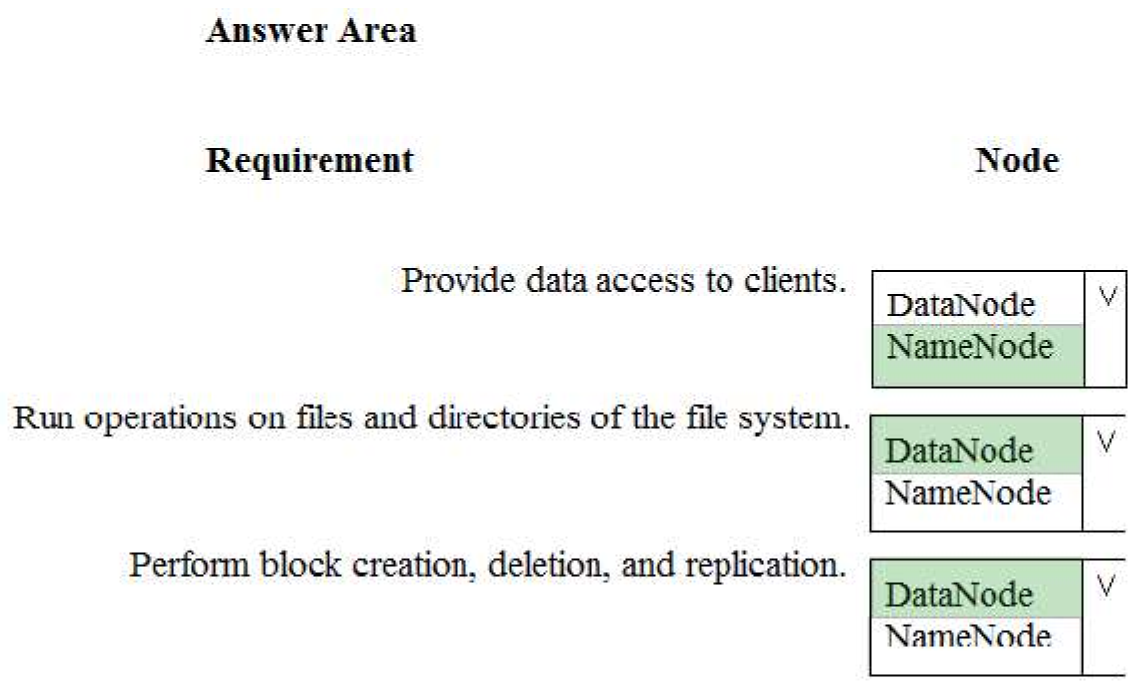
Hot Area:



Answer Area

Requirement	Node
Provide data access to clients.	DataNode NameNode
Run operations on files and directories of the file system.	DataNode NameNode
Perform block creation, deletion, and replication.	DataNode NameNode

Correct Answer:



Answer Area

Requirement	Node
Provide data access to clients.	DataNode NameNode
Run operations on files and directories of the file system.	DataNode NameNode
Perform block creation, deletion, and replication.	DataNode NameNode

Section: [none]
Explanation

Explanation/Reference:
Explanation:

Box 1: 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.

Box 2: DataNode
The DataNodes are responsible for serving read and write requests from the file system's clients.

Box 3: DataNode
The DataNodes perform block creation, deletion, and replication upon instruction from the NameNode.

Note: 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.

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

QUESTION 8
DRAG DROP

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.

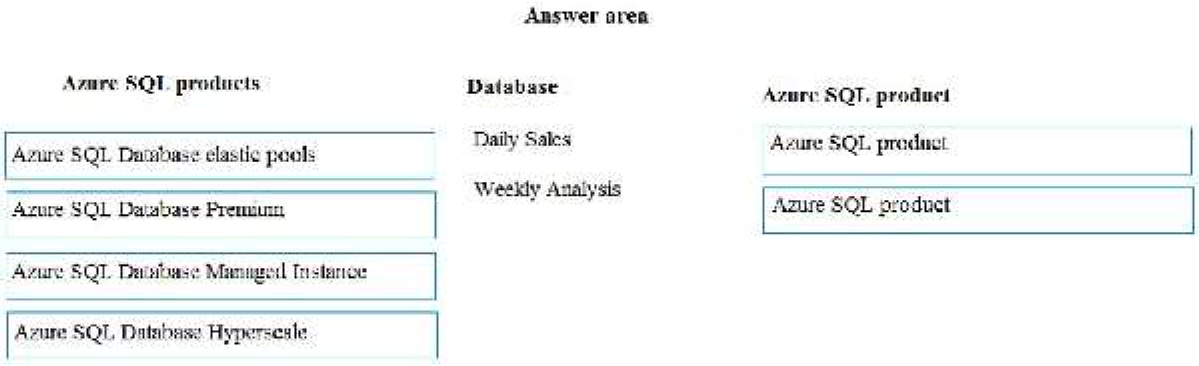
- Database provisioning must maximize performance and minimize cost
- The daily sales for each region must be stored in an Azure SQL Database instance
- Once a day, the data for all regions must be loaded in an analytical Azure SQL Database instance

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.

NOTE: Each correct selection is worth one point.

Select and Place:



Answer area

Azure SQL products

Azure SQL Database elastic pools

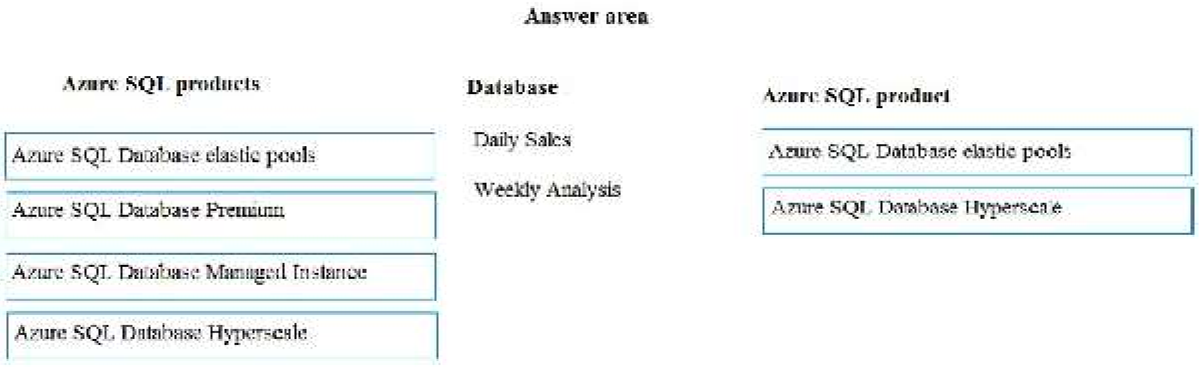
Azure SQL Database Premium

Azure SQL Database Managed Instance

Azure SQL Database Hyperscale

Database	Azure SQL product
Daily Sales	Azure SQL product
Weekly Analysis	Azure SQL product

Correct Answer:



Answer area

Azure SQL products

Azure SQL Database elastic pools

Azure SQL Database Premium

Azure SQL Database Managed Instance

Azure SQL Database Hyperscale

Database	Azure SQL product
Daily Sales	Azure SQL Database elastic pools
Weekly Analysis	Azure SQL Database Hyperscale

Section: [none]
Explanation

Explanation/Reference:
Explanation:

Box 1: 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 SQL Database enable SaaS developers to optimize the price performance for a group of databases within a prescribed budget

while delivering performance elasticity for each database.

Box 2: 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. Scaling is transparent to the application – connectivity, query processing, and so on, work like any other SQL database.

Incorrect Answers:
Azure SQL Database Managed Instance: The managed instance deployment model is designed for customers looking to migrate a large number of apps from on-premises or IaaS, self-built, or ISV provided environment to fully managed PaaS cloud environment, with as low migration effort as possible.

References:
https://docs.microsoft.com/en-us/azure/sql-database/sql-database-elastic-pool

https://docs.microsoft.com/en-us/azure/sql-database/sql-database-service-tier-hyperscale-faq

QUESTION 9
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?

A. Azure SQL Database single database
B. Azure SQL Data Warehouse
C. Azure Cosmos DB
D. Azure SQL Database Managed Instance

Correct Answer: D
Section: [none]
Explanation

Explanation/Reference:
Explanation:
Managed instance is a new deployment option of Azure SQL Database, providing near 100% compatibility with the latest SQL Server on-premises (Enterprise Edition) Database Engine, providing a native virtual network (VNet) implementation that addresses common security concerns, and a business model favorable for on-premises SQL Server customers. 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.

References:
https://docs.microsoft.com/en-us/azure/sql-database/sql-database-managed-instance

QUESTION 10
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?

A. Azure Databricks
B. Azure Traffic Manager
C. Azure Resource Manager templates
D. Ambari web user interface

Correct Answer: C
Section: [none]
Explanation

Explanation/Reference:
Explanation:
A Resource Manager template makes it easy to create the following resources for your application in a single, coordinated operation:

- HDInsight clusters and their dependent resources (such as the default storage account).
- Other resources (such as Azure SQL Database to use Apache Sqoop).

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.

References:
https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-hadoop-create-linux-clusters-arm-templates

QUESTION 11
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? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

A. Table API
B. MongoDB API
C. Gremlin API
D. SQL API
E. Cassandra API

Correct Answer: BE
Section: [none]
Explanation

Explanation/Reference:
Explanation:
B: Azure Cosmos DB is the globally distributed, multimodel database service from Microsoft for mission-critical applications. It is a multimodel database and supports document, key-value, graph, and columnar data models.

E: Wide-column stores store data together as columns instead of rows and are optimized for queries over large datasets. The most popular are Cassandra and HBase.

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

https://www.mongodb.com/scale/types-of-nosql-databases

QUESTION 12
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?

A. SQL Server Migration Assistant (SSMA)
B. Microsoft Assessment and Planning Toolkit
C. SQL Vulnerability Assessment (VA)
D. Azure SQL Data Sync
E. Data Migration Assistant (DMA)

Correct Answer: E
Section: [none]
Explanation

Explanation/Reference:
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 13
DRAG DROP

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:

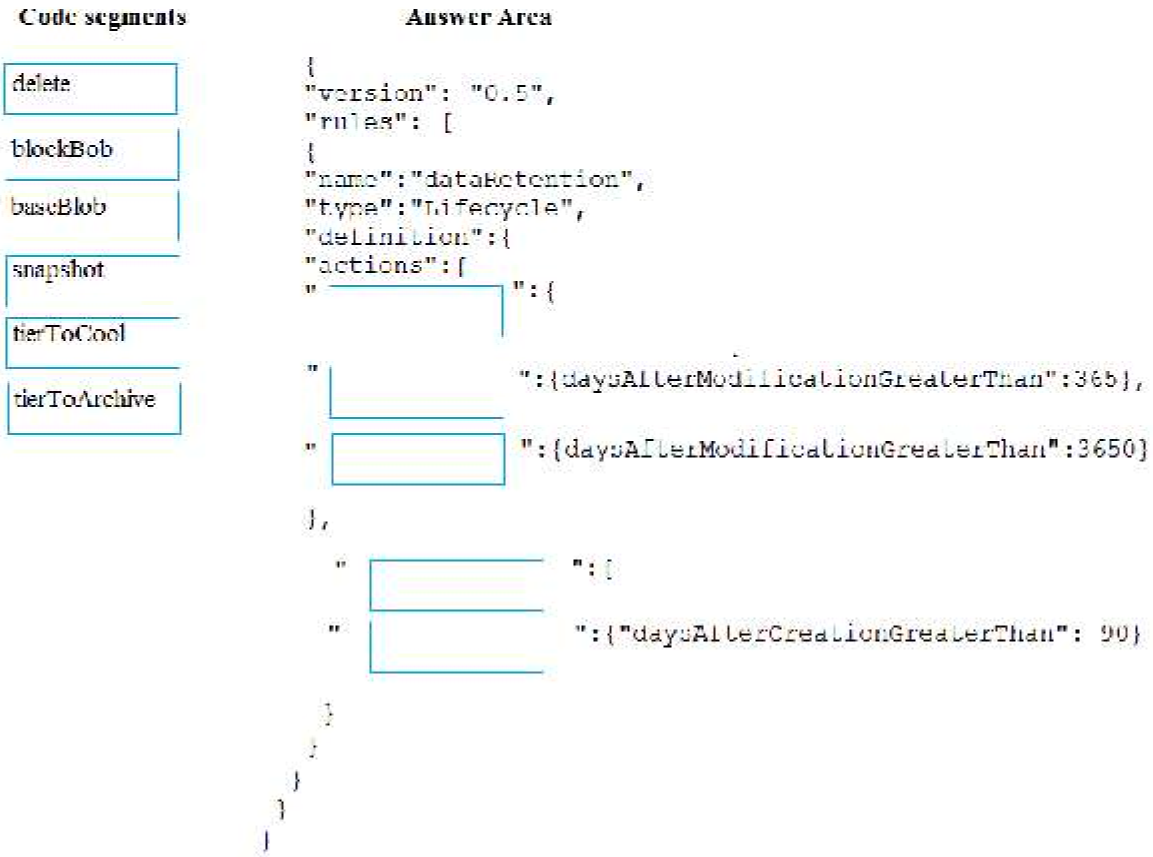
- Daily results must be kept for 90 days
- Data for the current year must be available for weekly reports
- Data from the previous 10 years must be stored for auditing purposes
- Data required for an audit must be produced within 10 days of a request.

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.

Select and Place:



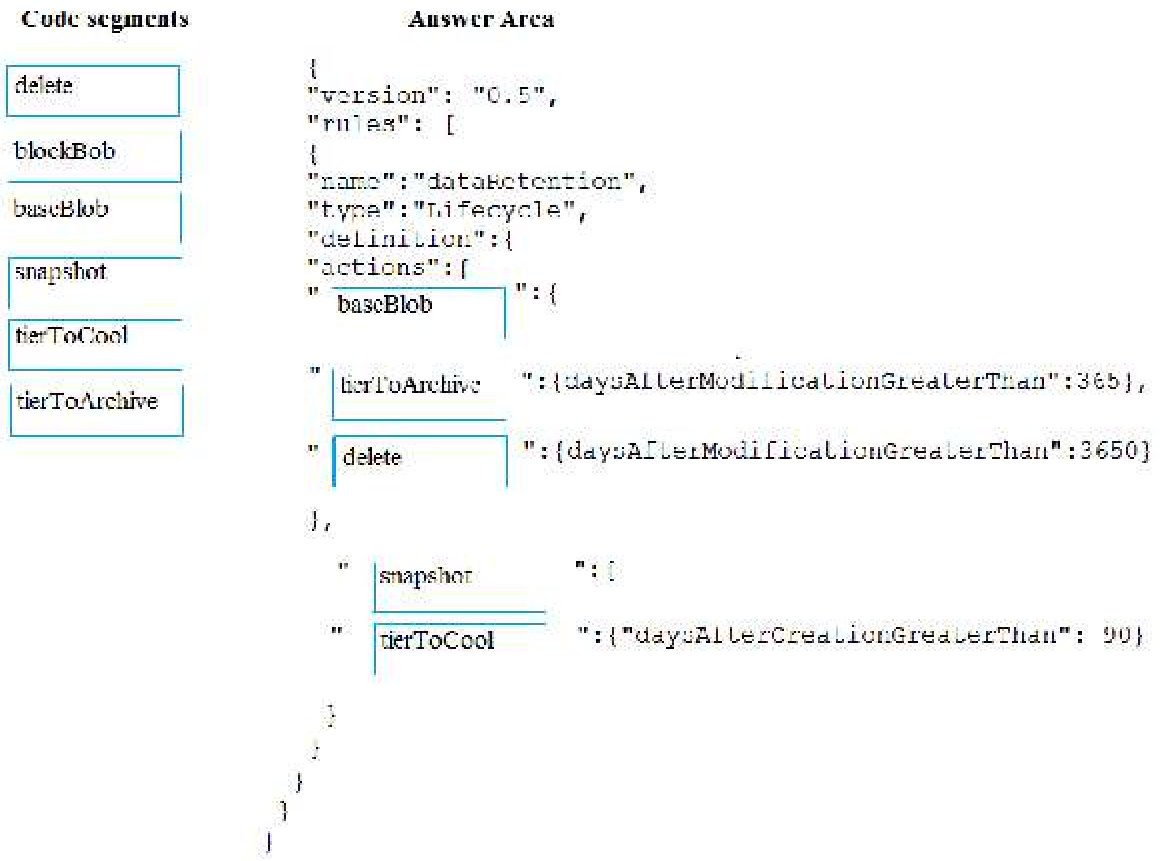
Code segments

- delete
- blockBob
- baseBlob
- snapshot
- tierToCool
- tierToArchive

Answer Area

```
{
"version": "0.5",
"rules": [
{
"name":"dataRetention",
"type":"Lifecycle",
"definition":{
"actions":[
"            ":{
"            ":{daysAfterModificationGreaterThan":365},
"            ":{daysAfterModificationGreaterThan":3650}
},
  "            ":[
  "            ":{"daysAfterCreationGreaterThan": 90}
  }
 }
}
}
}
```

Correct Answer:



Code segments

- delete
- blockBob
- baseBlob
- snapshot
- tierToCool
- tierToArchive

Answer Area

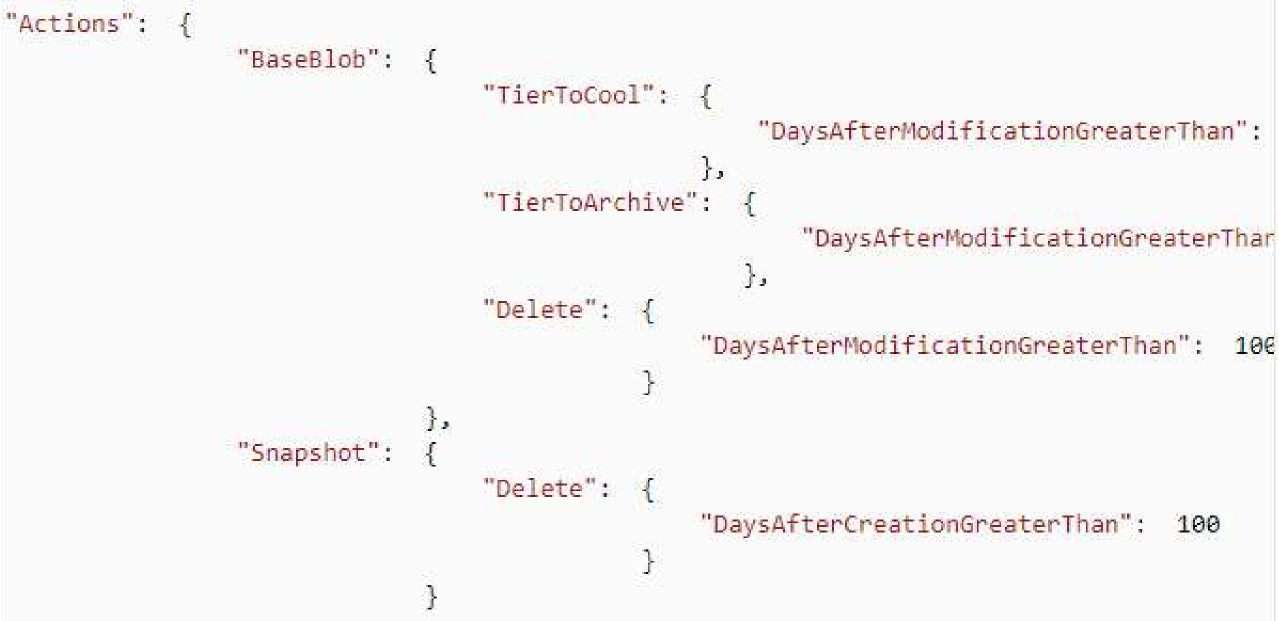
```
{
"version": "0.5",
"rules": [
{
"name":"dataRetention",
"type":"Lifecycle",
"definition":{
"actions":[
" baseBlob ":{
" tierToArchive ":{daysAfterModificationGreaterThan":365},
" delete ":{daysAfterModificationGreaterThan":3650}
},
  " snapshot ":{
  " tierToCool ":{"daysAfterCreationGreaterThan": 90}
   }
  }
 }
}
}
```

Section: [none]
Explanation

Explanation/Reference:
Explanation:

The Set-AzStorageAccountManagementPolicy cmdlet creates or modifies the management policy of an Azure Storage account.

Example: Create or update the management policy of a Storage account with ManagementPolicy rule objects.



```
"Actions":  {
                "BaseBlob":  {
                                 "TierToCool":  {
                                                    "DaysAfterModificationGreaterThan":
                                                },
                                 "TierToArchive":  {
                                                       "DaysAfterModificationGreaterThar
                                                   },
                                 "Delete":  {
                                                "DaysAfterModificationGreaterThan":  10
                                            }
                             },
                "Snapshot":  {
                                 "Delete":  {
                                                "DaysAfterCreationGreaterThan":  100
                                            }
                             }
```

Action -BaseBlobAction Delete -daysAfterModificationGreaterThan 100
PS C:\>$action1 = Add-AzStorageAccountManagementPolicyAction -InputObject $action1 -BaseBlobAction TierToArchive -daysAfterModificationGreaterThan 50
PS C:\>$action1 = Add-AzStorageAccountManagementPolicyAction -InputObject $action1 -BaseBlobAction TierToCool -daysAfterModificationGreaterThan 30
PS C:\>$action1 = Add-AzStorageAccountManagementPolicyAction -InputObject $action1 -SnapshotAction Delete -daysAfterCreationGreaterThan 100
PS C:\>$filter1 = New-AzStorageAccountManagementPolicyFilter -PrefixMatch ab,cd
PS C:\>$rule1 = New-AzStorageAccountManagementPolicyRule -Name Test -Action $action1 -Filter $filter1

PS C:\>$action2 = Add-AzStorageAccountManagementPolicyAction -BaseBlobAction Delete -daysAfterModificationGreaterThan 100
PS C:\>$filter2 = New-AzStorageAccountManagementPolicyFilter

References:
https://docs.microsoft.com/en-us/powershell/module/az.storage/set-azstorageaccountmanagementpolicy

QUESTION 14
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.

NOTE: Each correct selection is worth one point.

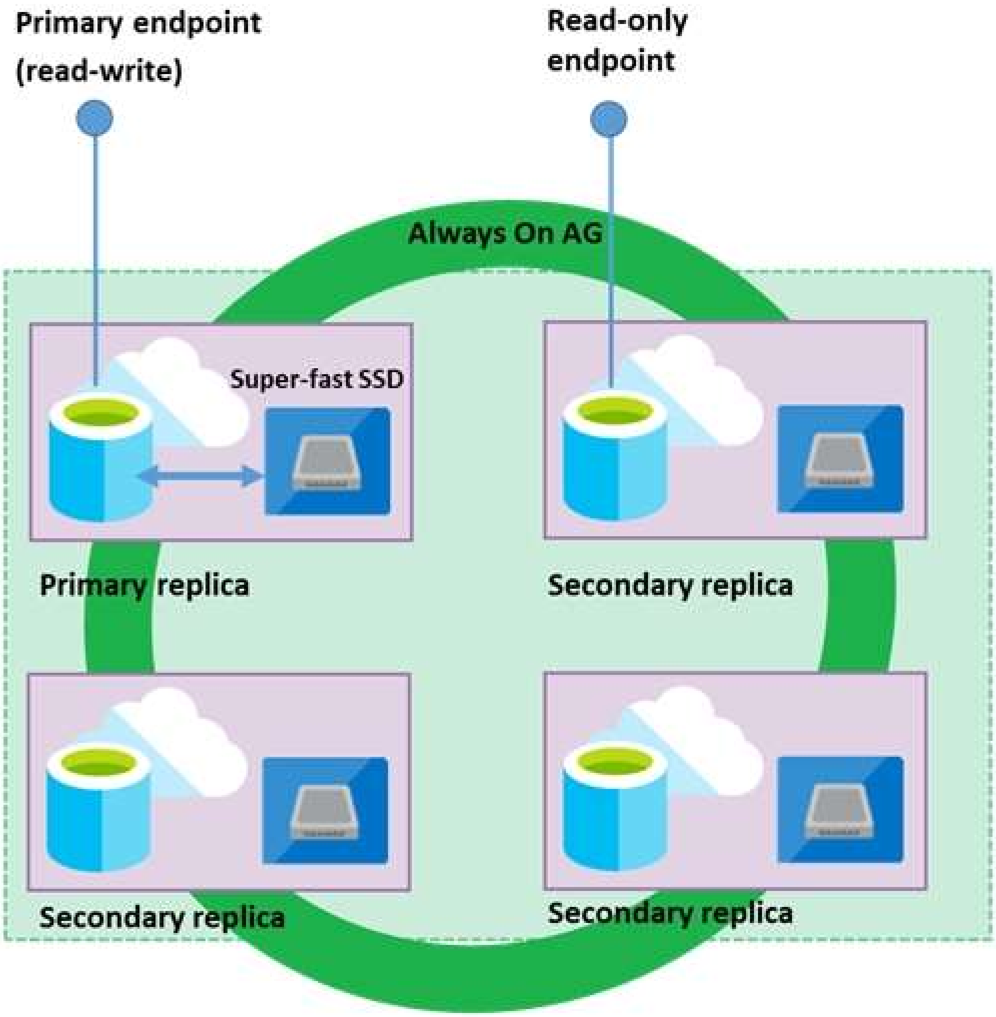
A. Premium service tier
B. Virtual machine Scale Sets
C. Basic service tier
D. SQL Data Sync
E. Always On availability groups
F. Zone-redundant configuration

Correct Answer: AEF

Section: [none]
Explanation

Explanation/Reference:
Explanation:
A: Premium/business critical service tier model that is based on a cluster of database engine processes. This architectural model relies on a fact that there is always a quorum of available database engine nodes and has minimal performance impact on your workload even during maintenance activities.

E: In the premium model, Azure SQL database integrates compute and storage on the single node. High availability in this architectural model is achieved by replication of compute (SQL Server Database Engine process) and storage (locally attached SSD) deployed in 4-node cluster, using technology similar to SQL Server Always On Availability Groups.



Business Critical service tier: collocated compute and storage

F: Zone redundant configuration
By default, the quorum-set replicas for the local storage configurations are created in the same datacenter. With the introduction of Azure Availability Zones, you have the ability to place the different replicas in the quorum-sets to different availability zones in the same region. To eliminate a single point of failure, the control ring is also duplicated across multiple zones as three gateway rings (GW).

References:
https://docs.microsoft.com/en-us/azure/sql-database/sql-database-high-availability

QUESTION 15
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.

NOTE: Each correct selection is worth one point.

A. Configure the storage account to log read, write and delete operations for service type Blob
B. Use the AzCopy tool to download log data from $logs/blob
C. Configure the storage account to log read, write and delete operations for service-type table
D. Use the storage client to download log data from $logs/table
E. Configure the storage account to log read, write and delete operations for service type queue

Correct Answer: AB
Section: [none]
Explanation

Explanation/Reference:
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 16
DRAG DROP

You are developing a solution to visualize multiple terabytes of geospatial data.

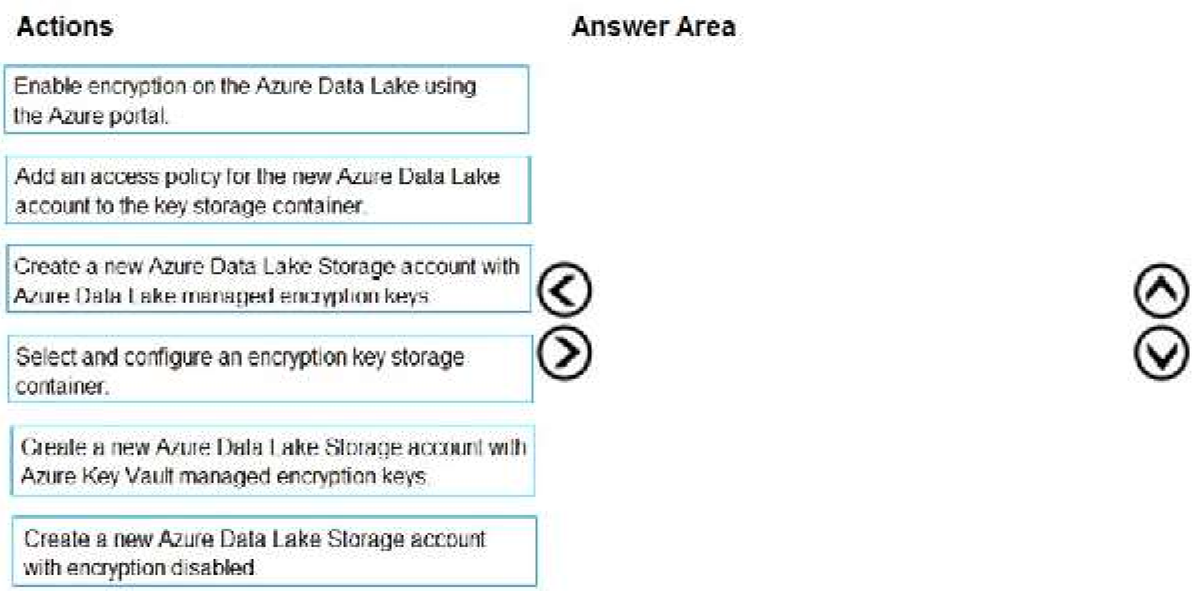
The solution has the following requirements:

- Data must be encrypted.
- Data must be accessible by multiple resources on Microsoft Azure.

You need to provision storage for the solution.

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

Select and Place:



Actions

Enable encryption on the Azure Data Lake using the Azure portal.

Add an access policy for the new Azure Data Lake account to the key storage container.

Create a new Azure Data Lake Storage account with Azure Data Lake managed encryption keys

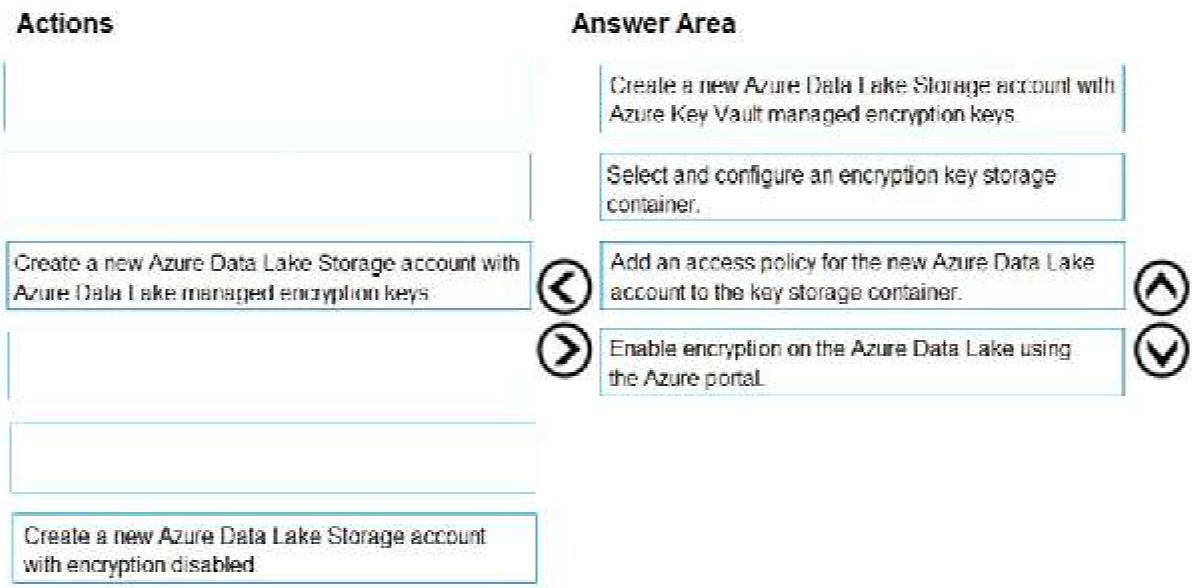
Select and configure an encryption key storage container.

Create a new Azure Data Lake Storage account with Azure Key Vault managed encryption keys

Create a new Azure Data Lake Storage account with encryption disabled

Answer Area

Correct Answer:



Actions

Create a new Azure Data Lake Storage account with Azure Data Lake managed encryption keys

Create a new Azure Data Lake Storage account with encryption disabled

Answer Area

Create a new Azure Data Lake Storage account with Azure Key Vault managed encryption keys

Select and configure an encryption key storage container.

Add an access policy for the new Azure Data Lake account to the key storage container.

Enable encryption on the Azure Data Lake using the Azure portal.

Section: [none]
Explanation

Explanation/Reference:
Explanation:

Create a new Azure Data Lake Storage account with Azure Data Lake managed encryption keys

For Azure services, Azure Key Vault is the recommended key storage solution and provides a common management experience across services. Keys are stored and managed in key vaults, and access to a key vault can be given to users or services. Azure Key Vault supports customer creation of keys or import of customer keys for use in customer-managed encryption key scenarios.

Note: Data Lake Storage Gen1 account Encryption Settings. There are three options:

- Do not enable encryption.
- Use keys managed by Data Lake Storage Gen1, if you want Data Lake Storage Gen1 to manage your

encryption keys.
- Use keys from your own Key Vault. You can select an existing Azure Key Vault or create a new Key Vault. To use the keys from a Key Vault, you must assign permissions for the Data Lake Storage Gen1 account to access the Azure Key Vault.

References:
https://docs.microsoft.com/en-us/azure/security/fundamentals/encryption-atrest

QUESTION 17
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.

A. Cosmos account-level throughput.
B. Provision an Azure Cosmos DB account with the Azure Table API. Enable geo-redundancy.
C. Configure table-level throughput.
D. Replicate the data globally by manually adding regions to the Azure Cosmos DB account.
E. Provision an Azure Cosmos DB account with the Azure Table API. Enable multi-region writes.

Correct Answer: E
Section: [none]
Explanation

Explanation/Reference:
Explanation:
Scale read and write throughput globally. You can enable every region to be writable and elastically scale reads and writes all around the world. The throughput that your application configures on an Azure Cosmos database or a container is guaranteed to be delivered across all regions associated with your Azure Cosmos account. The provisioned throughput is guaranteed up by financially backed SLAs.

References:
https://docs.microsoft.com/en-us/azure/cosmos-db/distribute-data-globally

QUESTION 18
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.

NOTE: Each correct selection is worth one point.

A. maximum data size
B. number of databases
C. eDTUs consumption
D. number of read operations
E. number of transactions

Correct Answer: AC
Section: [none]
Explanation

Explanation/Reference:
Explanation:
A: With the vCore purchase model, in the General Purpose tier, you are charged for Premium blob storage that you provision for your database or elastic pool. Storage can be configured between 5 GB and 4 TB with 1 GB increments. Storage is priced at GB/month.

C: In the DTU purchase model, elastic pools are available in basic, standard and premium service tiers. Each tier is distinguished primarily by its overall performance, which is measured in elastic Database Transaction Units (eDTUs).

References:
https://azure.microsoft.com/en-in/pricing/details/sql-database/elastic/

QUESTION 19
HOTSPOT

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:

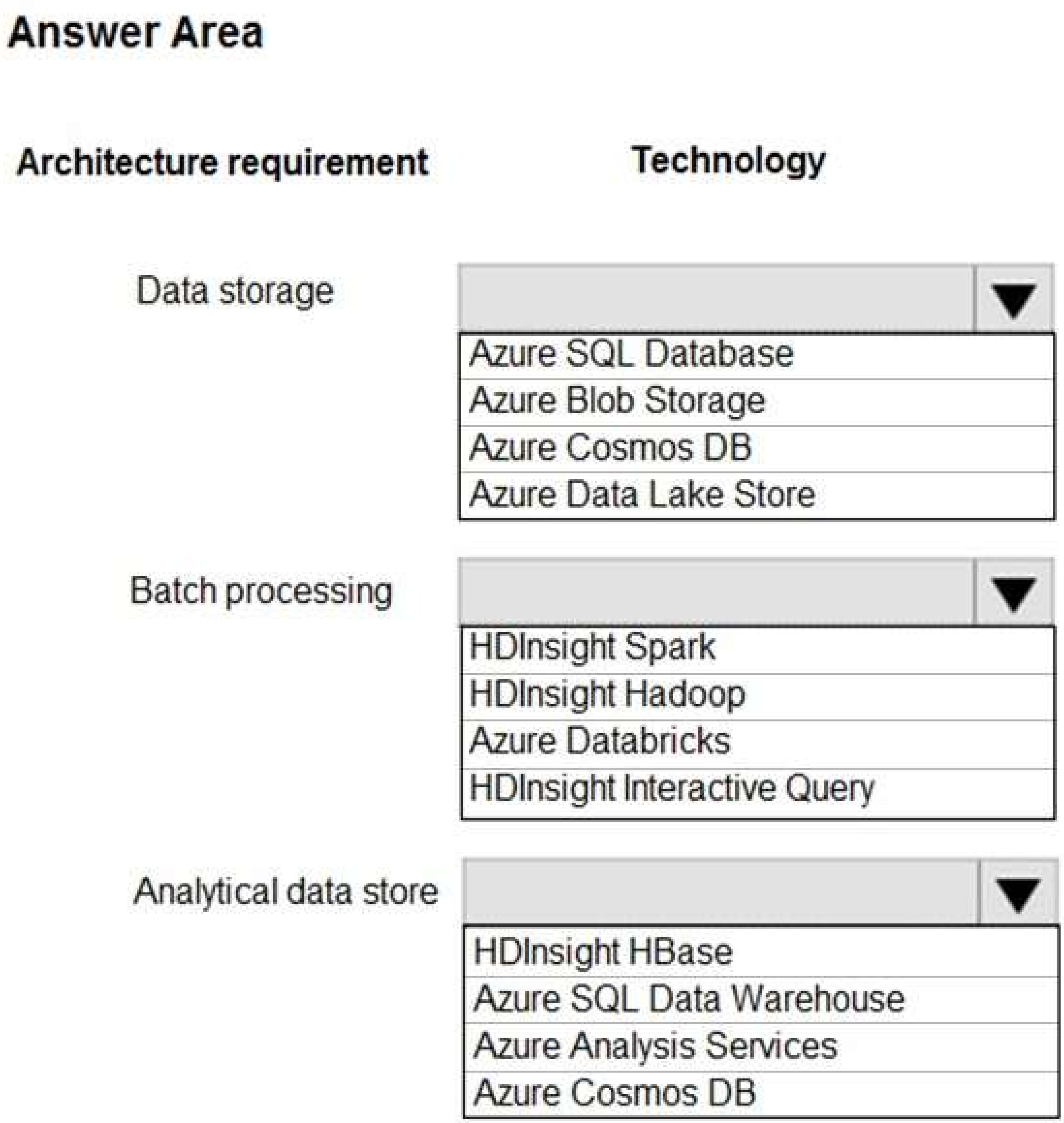
- Support parallel processing.
- Use columnar storage.
- Support SQL-based languages.

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.

NOTE: Each correct selection is worth one point.

Hot Area:



Answer Area

Architecture requirement | **Technology**

Data storage ▼

- Azure SQL Database
- Azure Blob Storage
- Azure Cosmos DB
- Azure Data Lake Store

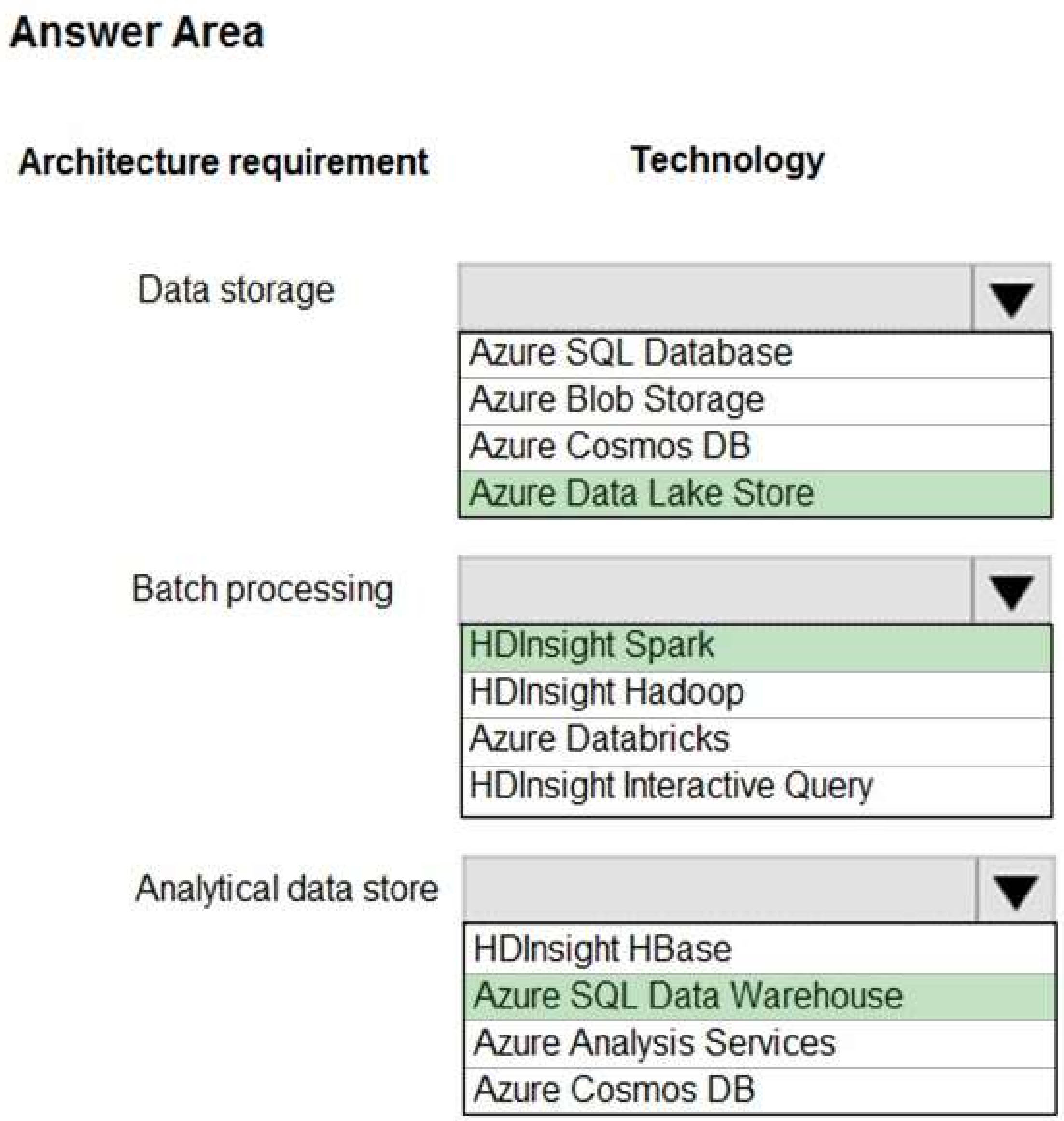
Batch processing ▼

- HDInsight Spark
- HDInsight Hadoop
- Azure Databricks
- HDInsight Interactive Query

Analytical data store ▼

- HDInsight HBase
- Azure SQL Data Warehouse
- Azure Analysis Services
- Azure Cosmos DB

Correct Answer:



Answer Area

Architecture requirement

Technology

Data storage

Azure SQL Database
Azure Blob Storage
Azure Cosmos DB
[✓ Azure Data Lake Store]

Batch processing

[✓ HDInsight Spark]
HDInsight Hadoop
Azure Databricks
HDInsight Interactive Query

Analytical data store

HDInsight HBase
[✓ Azure SQL Data Warehouse]
Azure Analysis Services
Azure Cosmos DB

Section: [none]
Explanation

Explanation/Reference:
Explanation:

Data storage: Azure Data Lake Store
A key mechanism that allows Azure Data Lake Storage Gen2 to provide file system performance at object storage scale and prices is the addition of a hierarchical namespace. This allows the collection of objects/files within an account to be organized into a hierarchy of directories and nested subdirectories in the same way that the file system on your computer is organized. With the hierarchical namespace enabled, a storage account becomes capable of providing the scalability and cost-effectiveness of object storage, with file system semantics that are familiar to analytics engines and frameworks.

Batch processing: HD Insight Spark
Aparch Spark is an open-source, parallel-processing framework that supports in-memory processing to boost the performance of big-data analysis applications.

HDInsight is a managed Hadoop service. Use it deploy and manage Hadoop clusters in Azure. For batch processing, you can use Spark, Hive, Hive LLAP, MapReduce.

Languages: R, Python, Java, Scala, SQL

Analytic data store: SQL Data Warehouse
SQL Data Warehouse is a cloud-based Enterprise Data Warehouse (EDW) that uses Massively Parallel Processing (MPP).
SQL Data Warehouse stores data into relational tables with columnar storage.

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

https://docs.microsoft.com/en-us/azure/architecture/data-guide/technology-choices/batch-processing

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

QUESTION 20
DRAG DROP

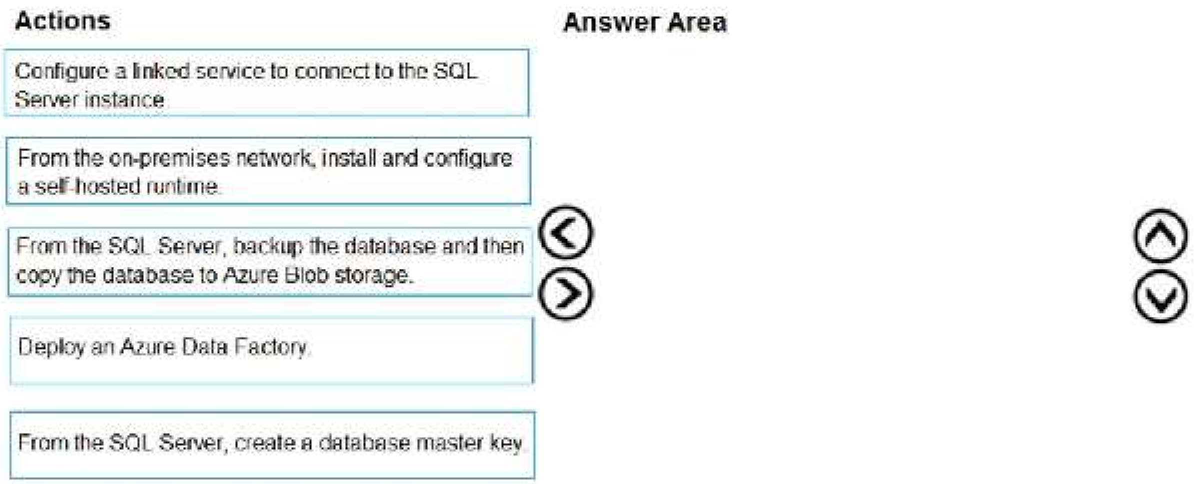
Your company has on-premises Microsoft SQL Server instance.

The data engineering team plans to implement a process that copies data from the SQL Server instance to Azure Blob storage. The process must orchestrate and manage the data lifecycle.

You need to configure Azure Data Factory to connect to the SQL Server instance.

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.

Select and Place:



Actions

Configure a linked service to connect to the SQL Server instance

From the on-premises network, install and configure a self-hosted runtime.

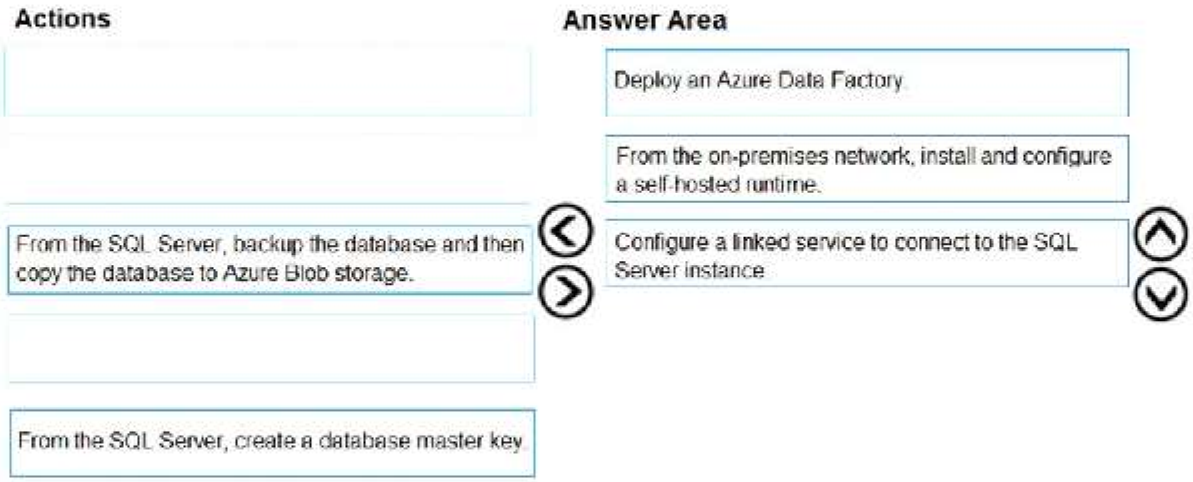
From the SQL Server, backup the database and then copy the database to Azure Blob storage.

Deploy an Azure Data Factory

From the SQL Server, create a database master key

Answer Area

Correct Answer:



Actions

From the SQL Server, backup the database and then copy the database to Azure Blob storage.

From the SQL Server, create a database master key

Answer Area

Deploy an Azure Data Factory

From the on-premises network, install and configure a self-hosted runtime.

Configure a linked service to connect to the SQL Server instance

Section: [none]
Explanation

Explanation/Reference:
Explanation:

Step 1: Deploy an Azure Data Factory
You need to create a data factory and start the Data Factory UI to create a pipeline in the data factory.

Step 2: From the on-premises network, install and configure a self-hosted runtime.
To use copy data from a SQL Server database that isn't publicly accessible, you need to set up a self-hosted integration runtime.

Step 3: Configure a linked service to connect to the SQL Server instance.

References:
https://docs.microsoft.com/en-us/azure/data-factory/connector-sql-server

QUESTION 21
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?

A. For each Azure SQL Database, set the Access Control to administrator.
B. For each Azure SQL Database server, set the Active Directory to administrator.
C. For each Azure SQL Database, set the Active Directory administrator role.
D. For each Azure SQL Database server, set the Access Control to administrator.

Correct Answer: C
Section: [none]
Explanation

Explanation/Reference:
Explanation:
There are two administrative accounts (Server admin and Active Directory admin) that act as administrators.

One Azure Active Directory account, either an individual or security group account, can also be configured as an administrator. It is optional to configure an Azure AD administrator, but an Azure AD administrator must be configured if you want to use Azure AD accounts to connect to SQL Database.

References:
https://docs.microsoft.com/en-us/azure/sql-database/sql-database-manage-logins

QUESTION 22
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this scenario, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You have an Azure 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?

A. Yes
B. No

Correct Answer: B
Section: [none]
Explanation

Explanation/Reference:
Explanation:
Must use Custom Text data masking, which exposes the first and last characters and adds a custom padding string in the middle.

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

QUESTION 23
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this scenario, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You have an Azure 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?

A. Yes
B. No

Correct Answer: B
Section: [none]
Explanation

Explanation/Reference:
Explanation:
Must use Custom Text data masking, which exposes the first and last characters and adds a custom padding string in the middle.

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

QUESTION 24
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this scenario, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You have an Azure 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?

A. Yes
B. No

Correct Answer: B
Section: [none]
Explanation

Explanation/Reference:
Explanation:
Must use Custom Text data masking, which exposes the first and last characters and adds a custom padding string in the middle.

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

QUESTION 25
DRAG DROP

You are responsible for providing access to an Azure Data Lake Storage Gen2 account.

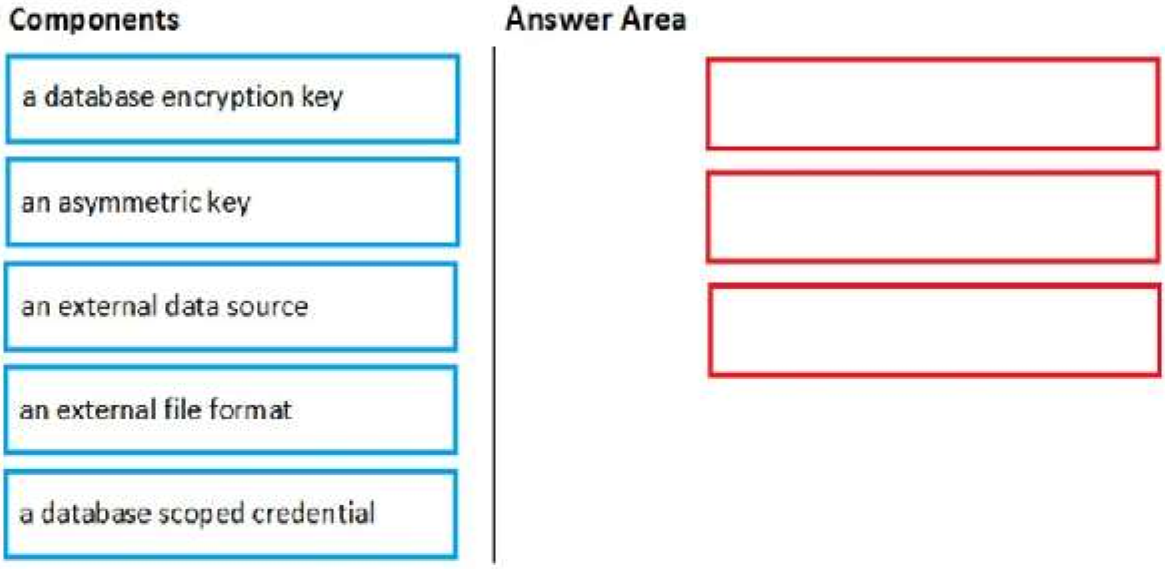
Your user account has contributor access to the storage account, and you have the application ID access key.

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.

Select and Place:

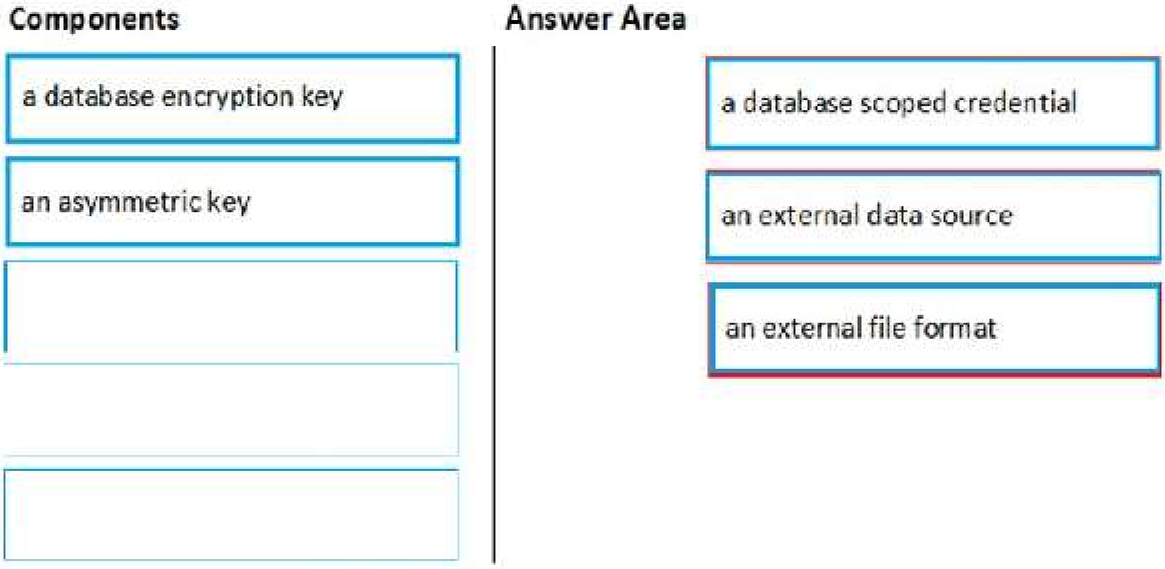


Components

- a database encryption key
- an asymmetric key
- an external data source
- an external file format
- a database scoped credential

Answer Area

Correct Answer:



Components

- a database encryption key
- an asymmetric key

Answer Area

- a database scoped credential
- an external data source
- an external file format

Section: [none]
Explanation

Explanation/Reference:
Step 1: a database scoped credential
To access your Data Lake Storage account, you will need to create a Database Master Key to encrypt your credential secret used in the next step. You then create a database scoped credential.

Step 2: an external data source
Create the external data source. Use the CREATE EXTERNAL DATA SOURCE command to store the location of the data. Provide the credential created in the previous step.

Step 3: an external file format
Configure data format: To import the data from Data Lake Storage, you need to specify the External File Format. This object defines how the files are written in Data Lake Storage.

References:
https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-load-from-azure-data-lake-store

QUESTION 26
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?

A. hash distributed
B. heap
C. replicated
D. round-robin

Correct Answer: D
Section: [none]
Explanation

Explanation/Reference:
Explanation:
Usually common dimension tables or tables that doesn't distribute evenly are good candidates for round-robin distributed table.

Note: Dimension tables or other lookup tables in a schema can usually be stored as round-robin tables. Usually these tables connect to more than one fact tables and optimizing for one join may not be the best idea. Also usually dimension tables are smaller which can leave some distributions empty when hash distributed. Round-robin by definition guarantees a uniform data distribution.

References:
https://blogs.msdn.microsoft.com/sqlcat/2015/08/11/choosing-hash-distributed-table-vs-round-robin-distributed-table-in-azure-sql-dw-service/

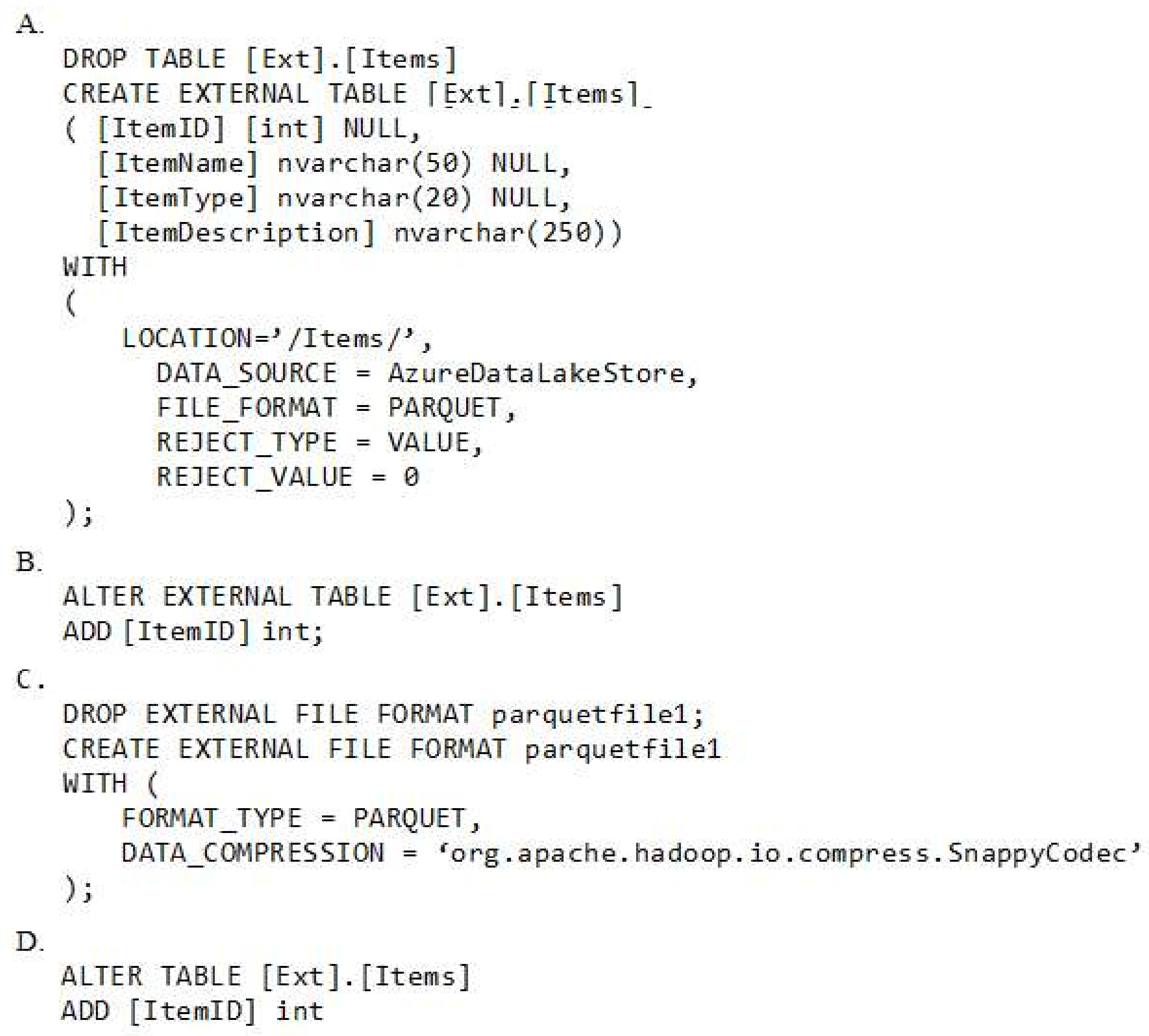
QUESTION 27
You have an Azure SQL data warehouse.

Using PolyBase, you create table named [Ext].[Items] to query Parquet files stored in Azure Data Lake Storage Gen2 without importing the data to the data warehouse.

The external table has three columns.

You discover that the Parquet files have a fourth column named ItemID.

Which command should you run to add the ItemID column to the external table?



```
A.
   DROP TABLE [Ext].[Items]
   CREATE EXTERNAL TABLE [Ext].[Items]
   ( [ItemID] [int] NULL,
     [ItemName] nvarchar(50) NULL,
     [ItemType] nvarchar(20) NULL,
     [ItemDescription] nvarchar(250))
   WITH
   (
       LOCATION=’/Items/’,
         DATA_SOURCE = AzureDataLakeStore,
         FILE_FORMAT = PARQUET,
         REJECT_TYPE = VALUE,
         REJECT_VALUE = 0
   );
B.
   ALTER EXTERNAL TABLE [Ext].[Items]
   ADD [ItemID] int;
C.
   DROP EXTERNAL FILE FORMAT parquetfile1;
   CREATE EXTERNAL FILE FORMAT parquetfile1
   WITH (
       FORMAT_TYPE = PARQUET,
       DATA_COMPRESSION = ‘org.apache.hadoop.io.compress.SnappyCodec’
   );
D.
   ALTER TABLE [Ext].[Items]
   ADD [ItemID] int
```

A. Option A
B. Option B
C. Option C
D. Option D

Correct Answer: A
Section: [none]
Explanation

Explanation/Reference:
Incorrect Answers:
B, D: Only these Data Definition Language (DDL) statements are allowed on external tables:

- CREATE TABLE and DROP TABLE
- CREATE STATISTICS and DROP STATISTICS

- CREATE VIEW and DROP VIEW

References:
https://docs.microsoft.com/en-us/sql/t-sql/statements/create-external-table-transact-sql

QUESTION 28
DRAG DROP

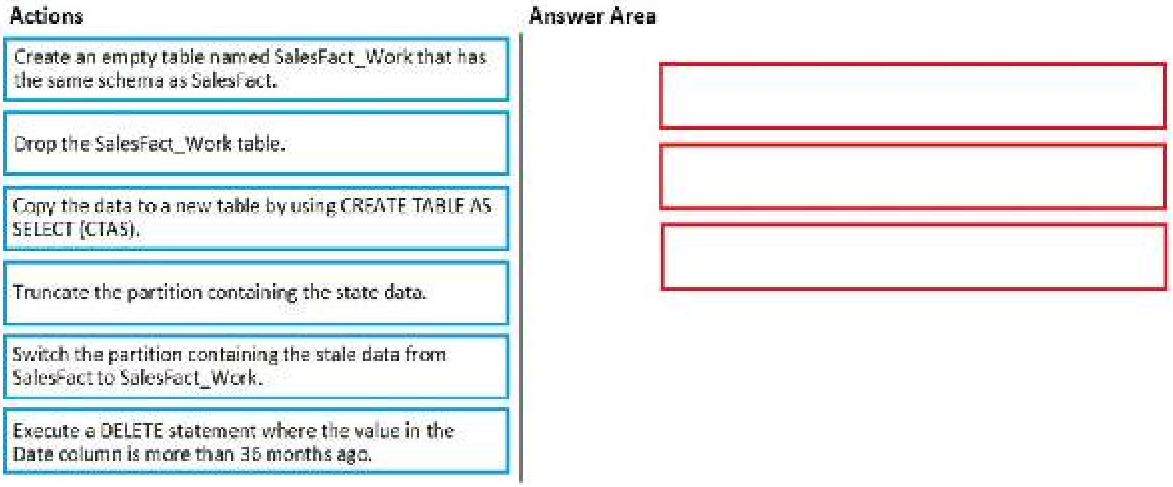
You have a table named SalesFact in an Azure SQL data warehouse. SalesFact contains sales data from the past 36 months and has the following characteristics:

- 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.

Select and Place:

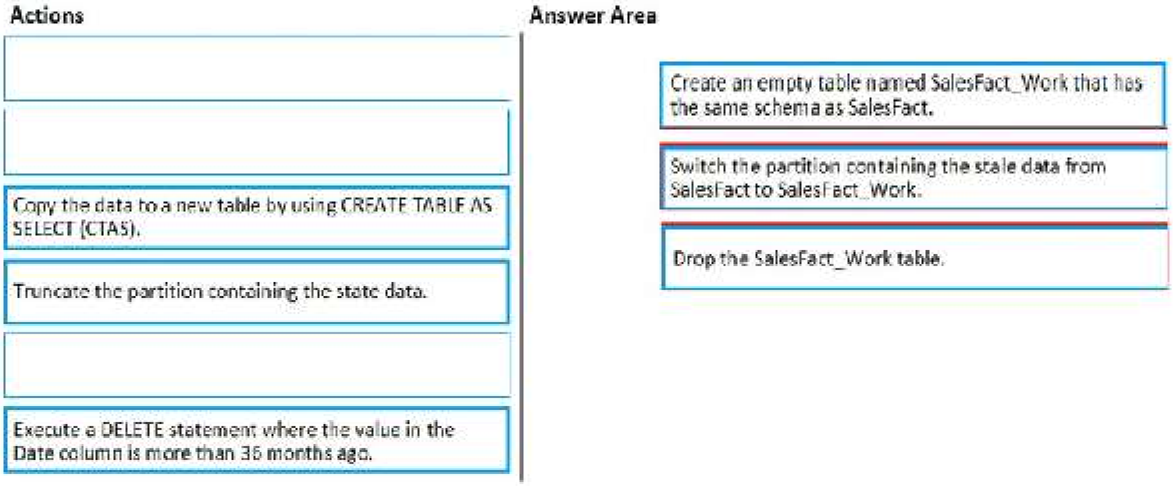


Actions

- Create an empty table named SalesFact_Work that has the same schema as SalesFact.
- Drop the SalesFact_Work table.
- Copy the data to a new table by using CREATE TABLE AS SELECT (CTAS).
- Truncate the partition containing the state data.
- Switch the partition containing the stale data from SalesFact to SalesFact_Work.
- Execute a DELETE statement where the value in the Date column is more than 36 months ago.

Answer Area

Correct Answer:



Actions

- Copy the data to a new table by using CREATE TABLE AS SELECT (CTAS).
- Truncate the partition containing the state data.
- Execute a DELETE statement where the value in the Date column is more than 36 months ago.

Answer Area

- Create an empty table named SalesFact_Work that has the same schema as SalesFact.
- Switch the partition containing the stale data from SalesFact to SalesFact_Work.
- Drop the SalesFact_Work table.

Section: [none]
Explanation

Explanation/Reference:
Step 1: Create an empty table named SalesFact_work that has the same schema as SalesFact.

Step 2: Switch the partition containing the stale data from SalesFact to SalesFact_Work.
SQL Data Warehouse supports partition splitting, merging, and switching. To switch partitions between two tables, you must ensure that the partitions align on their respective boundaries and that the table definitions match.

Loading data into partitions with partition switching is a convenient way stage new data in a table that is not visible to users the switch in the new data.

Step 3: Drop the SalesFact_Work table.

References:
https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-tables-partition

QUESTION 29
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?

A. Strong
B. Bounded Staleness
C. Eventual
D. Session
E. Consistent Prefix

Correct Answer: D
Section: [none]
Explanation

Explanation/Reference:
Explanation:
Session: Within a single client session reads are guaranteed to honor the consistent-prefix (assuming a single “writer” session), monotonic reads, monotonic writes, read-your-writes, and write-follows-reads guarantees. Clients outside of the session performing writes will see eventual consistency.

References:
https://docs.microsoft.com/en-us/azure/cosmos-db/consistency-levels

QUESTION 30
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this scenario, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You have an Azure 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?

A. Yes

B. No

Correct Answer: B
Section: [none]
Explanation

Explanation/Reference:
Explanation:
We must use Custom Text data masking, which exposes the first and last characters and adds a custom padding string in the middle.

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

Testlet 2

Background
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.

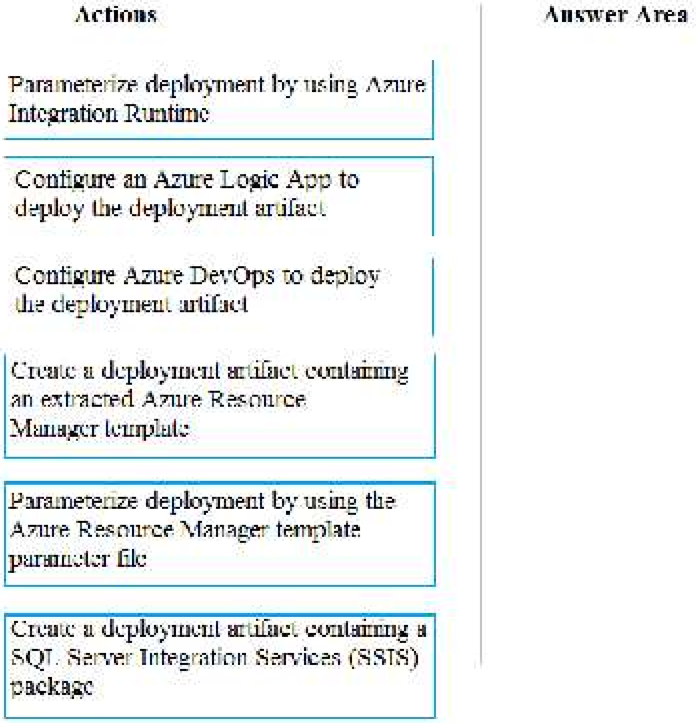
Monitoring
All Azure services must be monitored by using Azure Monitor. On-premises SQL Server performance must be monitored.

QUESTION 1
DRAG DROP

You need to ensure that phone-based polling data can be analyzed in the PollingData database.

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

Select and Place:



Actions

Parameterize deployment by using Azure Integration Runtime

Configure an Azure Logic App to deploy the deployment artifact

Configure Azure DevOps to deploy the deployment artifact

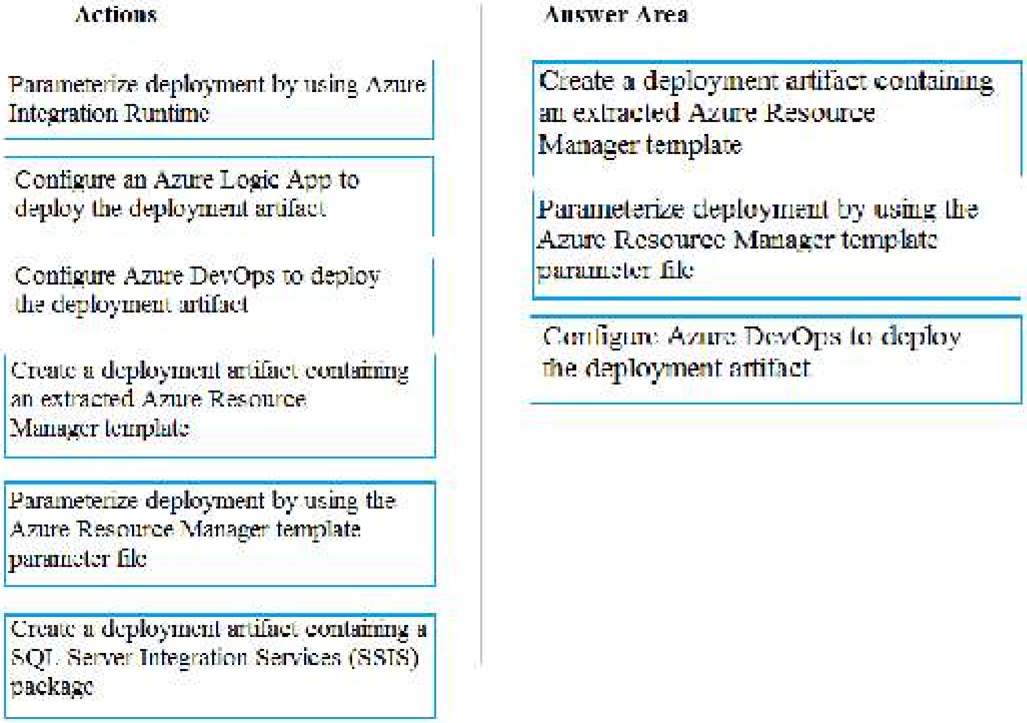
Create a deployment artifact containing an extracted Azure Resource Manager template

Parameterize deployment by using the Azure Resource Manager template parameter file

Create a deployment artifact containing a SQL Server Integration Services (SSIS) package

Answer Area

Correct Answer:



Actions

Parameterize deployment by using Azure Integration Runtime

Configure an Azure Logic App to deploy the deployment artifact

Configure Azure DevOps to deploy the deployment artifact

Create a deployment artifact containing an extracted Azure Resource Manager template

Parameterize deployment by using the Azure Resource Manager template parameter file

Create a deployment artifact containing a SQL Server Integration Services (SSIS) package

Answer Area

Create a deployment artifact containing an extracted Azure Resource Manager template

Parameterize deployment by using the Azure Resource Manager template parameter file

Configure Azure DevOps to deploy the deployment artifact

Section: [none]
Explanation

Explanation/Reference:
Explanation:
Scenario:
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

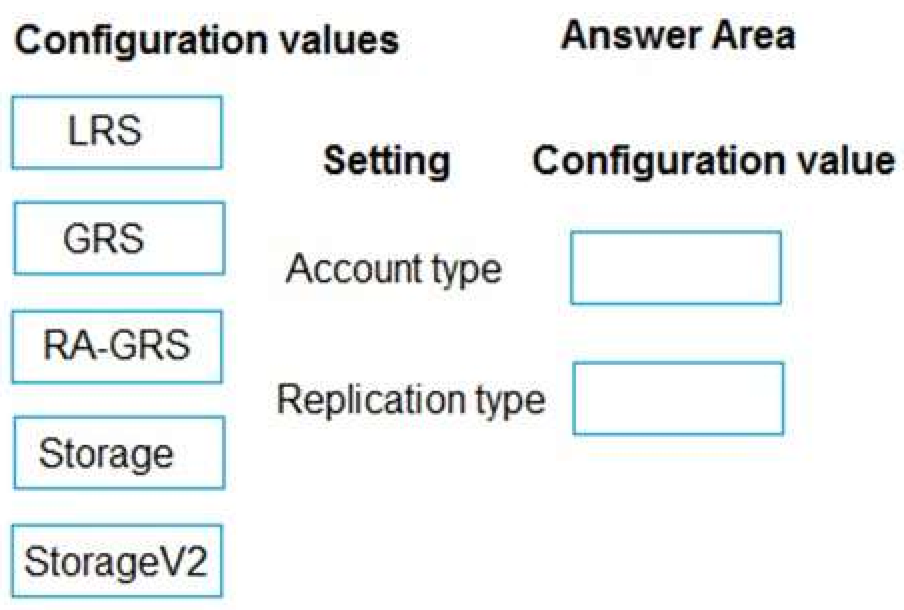
QUESTION 2
DRAG DROP

You need to provision the polling data storage account.

How should you configure the storage account? To answer, drag the appropriate Configuration Value to the correct Setting. Each Configuration Value may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

Select and Place:



Configuration values

LRS

GRS

RA-GRS

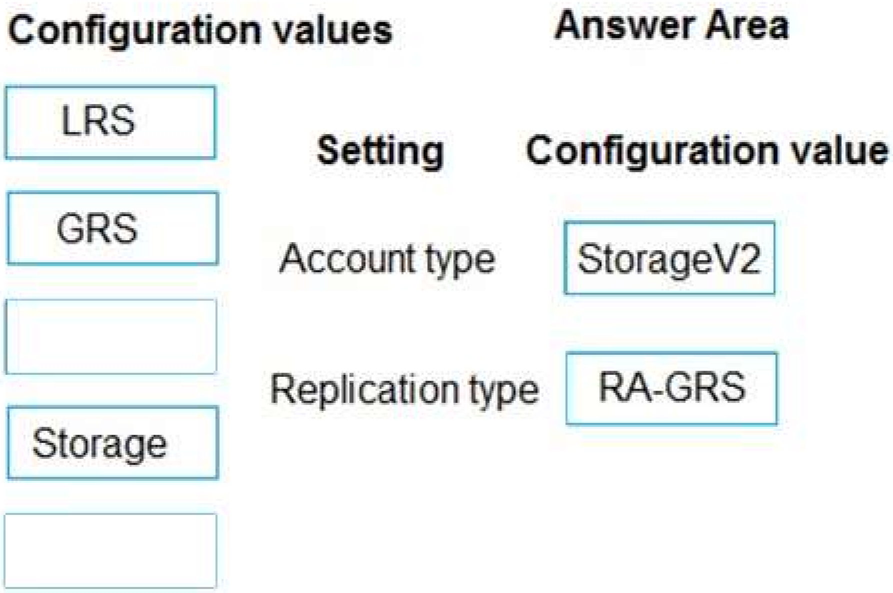
Storage

StorageV2

Answer Area

Setting	Configuration value
Account type	
Replication type	

Correct Answer:



Configuration values

LRS

GRS

Storage

Answer Area

Setting	Configuration value
Account type	StorageV2
Replication type	RA-GRS

Section: [none]
Explanation

Explanation/Reference:
Explanation:

Account type: StorageV2
You must create new storage accounts as type StorageV2 (general-purpose V2) to take advantage of Data Lake Storage Gen2 features.

Scenario: 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

Replication type: RA-GRS
Scenario: All services and processes must be resilient to a regional Azure outage.

Geo-redundant storage (GRS) is designed to provide at least 99.99999999999999% (16 9's) durability of objects over a given year by replicating your data to a secondary region that is hundreds of miles away from the primary region. If your storage account has GRS enabled, then your data is durable even in the case of a complete regional outage or a disaster in which the primary region isn't recoverable.

If you opt for GRS, you have two related options to choose from:

- GRS replicates your data to another data center in a secondary region, but that data is available to be read only if Microsoft initiates a failover from the primary to secondary region.
- Read-access geo-redundant storage (RA-GRS) is based on GRS. RA-GRS replicates your data to another data center in a secondary region, and also provides you with the option to read from the secondary region. With RA-GRS, you can read from the secondary region regardless of whether Microsoft initiates a failover from the primary to secondary region.

References:
https://docs.microsoft.com/bs-cyrl-ba/azure/storage/blobs/data-lake-storage-quickstart-create-account

https://docs.microsoft.com/en-us/azure/storage/common/storage-redundancy-grs

Testlet 3

Case Study

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

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

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

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

Overview

General Overview

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.

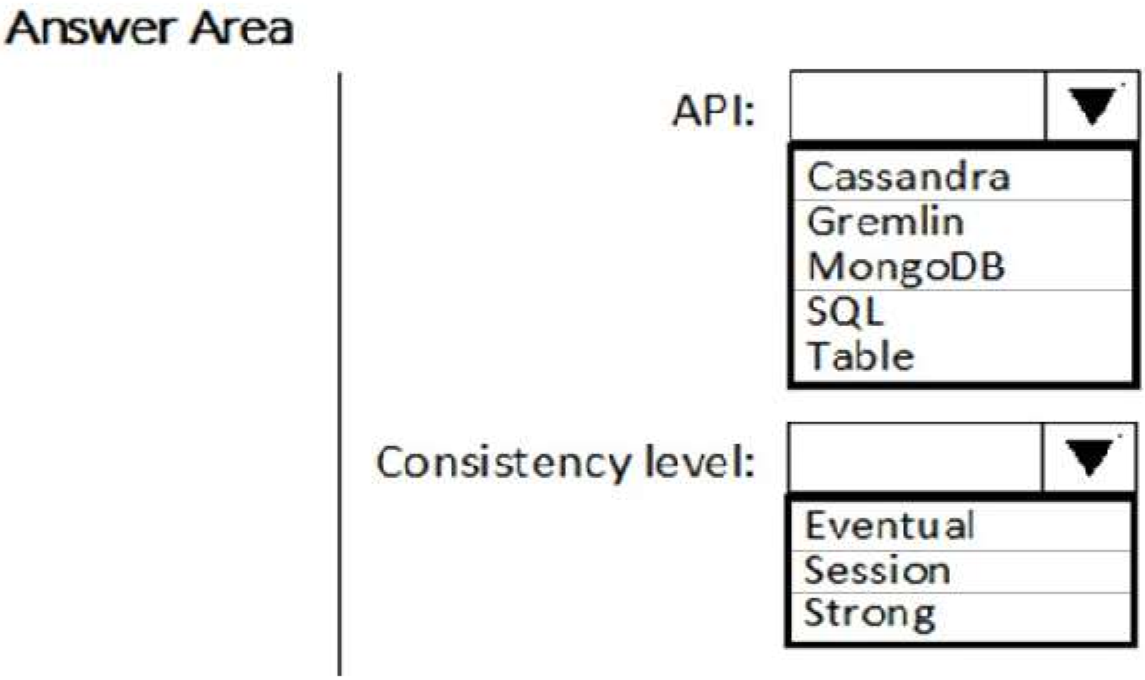
QUESTION 1

You need to build a solution to collect the telemetry data for Race Control.

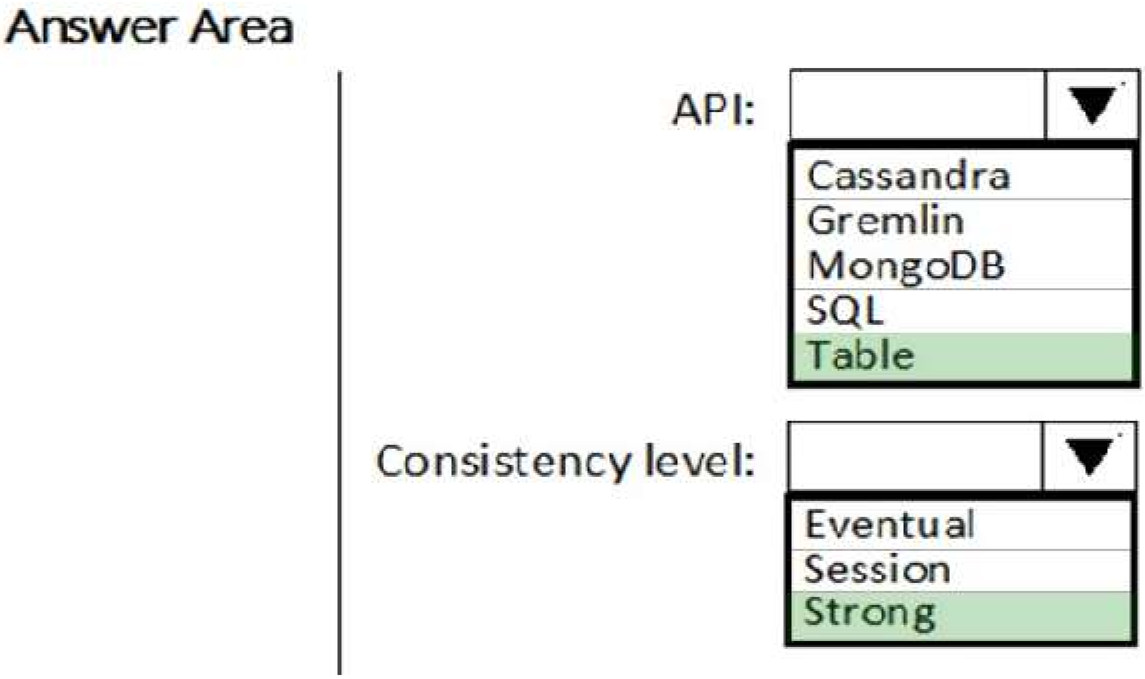
What should you use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:



Correct Answer:



Section: [none]
Explanation

Explanation/Reference:
API: Table
Azure Cosmos DB provides native support for wire protocol-compatible APIs for popular databases. These include MongoDB, Apache Cassandra, Gremlin, and Azure Table storage.

Scenario: The telemetry data must migrate toward a solution that is native to Azure.

Consistency level: Strong
Use the strongest consistency Strong to minimize convergence time.

Scenario: The data must be written to the Azure datacentre closest to each race and must converge in the least amount of time.

References:
https://docs.microsoft.com/en-us/azure/cosmos-db/consistency-levels

QUESTION 2
On which data store you configure TDE to meet the technical requirements?

A. Cosmos DB
B. SQL Data Warehouse
C. SQL Database

Correct Answer: B
Section: [none]
Explanation

Explanation/Reference:
Explanation:
Scenario: Transparent data encryption (TDE) must be enabled on all data stores, whenever possible.
The datacentre for Mechanical Workflow must be moved to Azure SQL data Warehouse.

Incorrect Answers:
A: Cosmos DB does not support TDE.

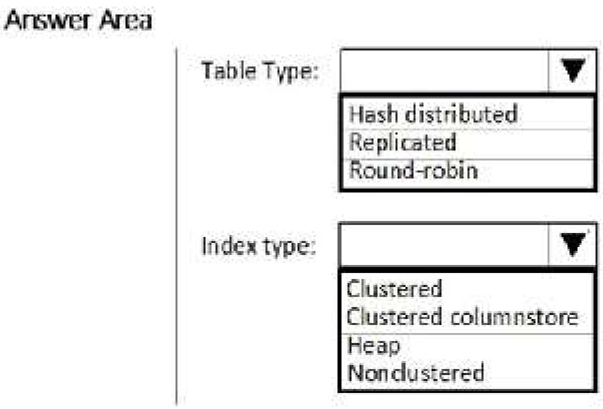
QUESTION 3
HOTSPOT

You are building the data store solution for Mechanical Workflow.

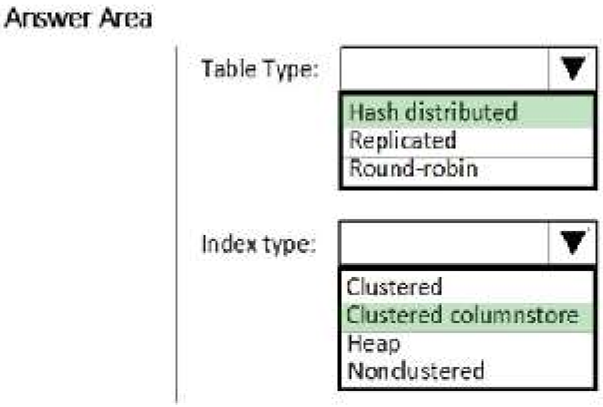
How should you configure Table1? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:



Correct Answer:



Answer Area

Table Type:

[✓ Hash distributed]
Replicated
Round-robin

Index type:

Clustered
[✓ Clustered columnstore]
Heap
Nonclustered

Section: [none]
Explanation

Explanation/Reference:
Table Type: Hash distributed.
Hash-distributed tables improve query performance on large fact tables.

Index type: Clusted columnstore

Scenario:
Mechanical Workflow has a named Table1 that is 1 TB. Large aggregations are performed on a single column of Table 1.

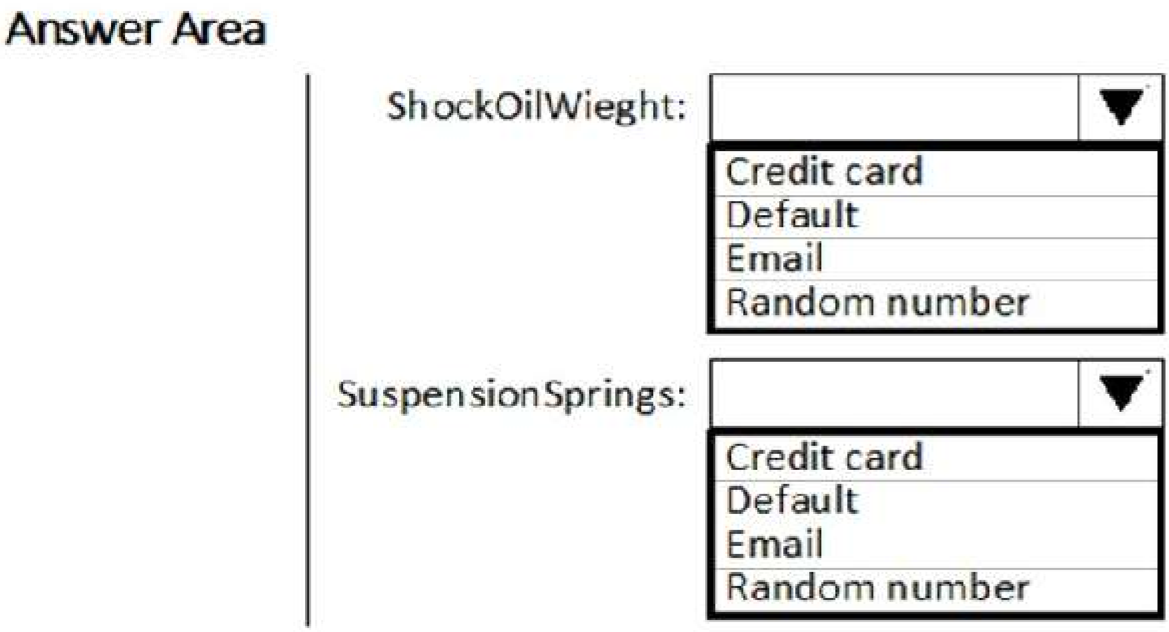
References:
https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-tables-distribute

QUESTION 4
HOTSPOT

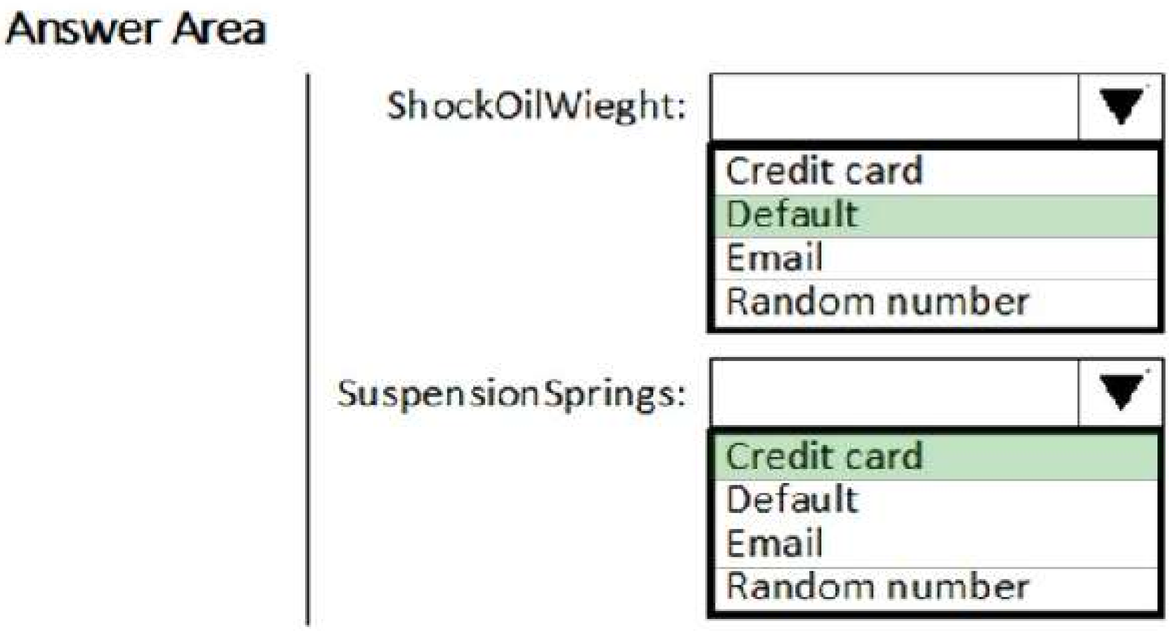
Which masking functions should you implement for each column to meet the data masking requirements? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:



Correct Answer:



Section: [none]
Explanation

Explanation/Reference:
Box 1: Default
Default uses a zero value for numeric data types (bigint, bit, decimal, int, money, numeric, smallint, smallmoney, tinyint, float, real).

- Only Show a zero value for the values in a column named ShockOilWeight.

Box 2: Credit Card

The Credit Card Masking method exposes the last four digits of the designated fields and adds a constant string as a prefix in the form of a credit card.
Example: XXXX-XXXX-XXXX-1234

- Only show the last four digits of the values in a column named SuspensionSprings.

Scenario:
The company identifies the following data masking requirements for the Race Central data that will be stored in SQL Database:

- Only Show a zero value for the values in a column named ShockOilWeight.
- Only show the last four digits of the values in a column named SuspensionSprings.

Question Set 1

QUESTION 1

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?

A. Yes

B. No

Correct Answer: B
Section: [none]
Explanation

Explanation/Reference:
Explanation:
There is no need to convert the parquet files to CSV files.
You load the data using the CREATE TABLE AS SELECT statement.

References:
https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-load-from-azure-data-lake-store

QUESTION 2

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?

A. Yes

B. No

Correct Answer: B
Section: [none]

Explanation

Explanation/Reference:
Explanation:
You load the data using the CREATE TABLE AS SELECT statement.

References:
https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-load-from-azure-data-lake-store

QUESTION 3
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?

A. Yes
B. No

Correct Answer: B
Section: [none]
Explanation

Explanation/Reference:
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.

References:
https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-load-from-azure-data-lake-store

QUESTION 4
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?

A. Use the Copy Data tool with Blob storage linked service as the source
B. Use Azure PowerShell with SQL Server linked service as a source
C. Use Azure Data Factory UI with Blob storage linked service as a source
D. Use the .NET Data Factory API with Blob storage linked service as the source

Correct Answer: C
Section: [none]
Explanation

Explanation/Reference:
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

Note: Azure Data Factory is a fully managed cloud-based data integration service that orchestrates and automates the movement and transformation of data. The key concept in the ADF model is pipeline. A pipeline is a logical grouping of Activities, each of which defines the actions to perform on the data contained in Datasets. Linked services are used to define the information needed for Data Factory to connect to the data resources.

References:
https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/move-sql-azure-adf

QUESTION 5
HOTSPOT

A company runs Microsoft Dynamics CRM with Microsoft SQL Server on-premises. SQL Server Integration Services (SSIS) packages extract data from Dynamics CRM APIs, and load the data into a SQL Server data warehouse.

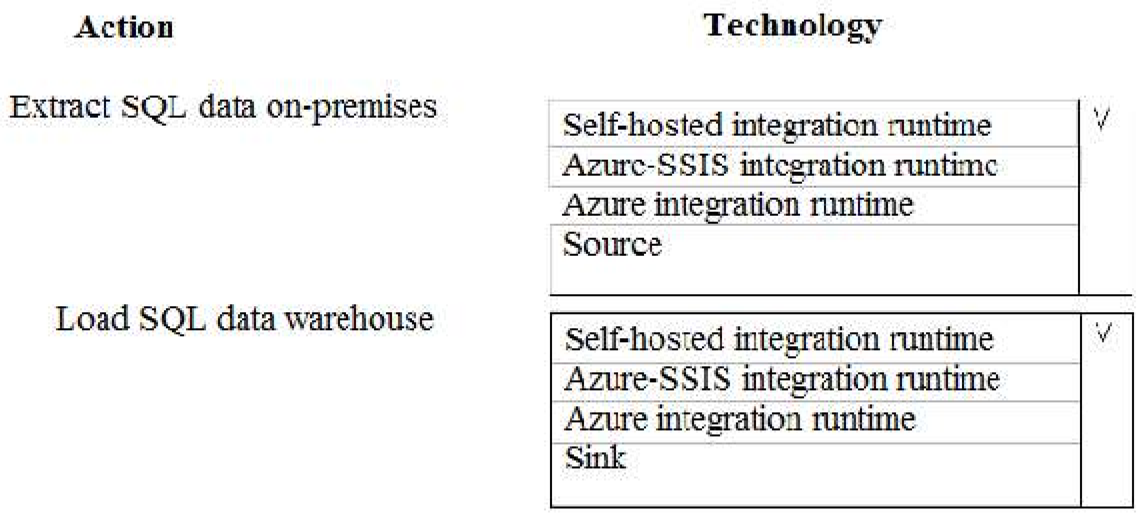
The datacenter is running out of capacity. Because of the network configuration, you must extract on premises data to the cloud over https. You cannot open any additional ports. The solution must implement the least amount of effort.

You need to create the pipeline system.

Which component should you use? To answer, select the appropriate technology in the dialog box in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:



Action

Technology

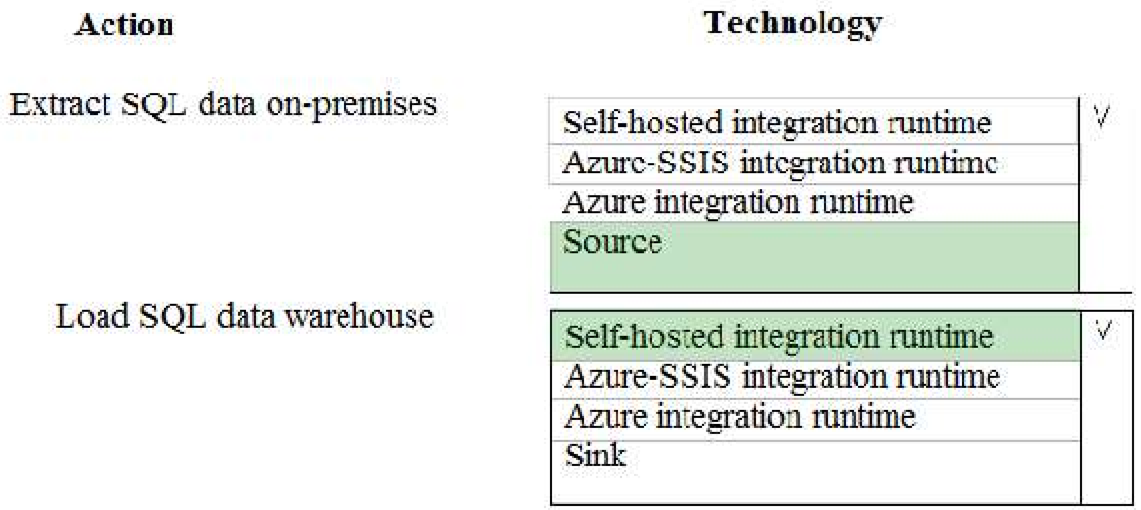
Extract SQL data on-premises

Self-hosted integration runtime
Azurc-SSIS intcgration runtimc
Azure integration runtime
Source

Load SQL data warehouse

Self-hosted integration runtime
Azure-SSIS integration runtime
Azure integration runtime
Sink

Correct Answer:



Action

Technology

Extract SQL data on-premises

Self-hosted integration runtime
Azurc-SSIS intcgration runtimc
Azure integration runtime
Source

Load SQL data warehouse

Self-hosted integration runtime
Azure-SSIS integration runtime
Azure integration runtime
Sink

Section: [none]
Explanation

Explanation/Reference:
Explanation:
Box 1: Source
For Copy activity, it requires source and sink linked services to define the direction of data flow.
Copying between a cloud data source and a data source in private network: if either source or sink linked service points to a self-hosted IR, the copy activity is executed on that self-hosted Integration Runtime.

Box 2: Self-hosted integration runtime
A self-hosted integration runtime can run copy activities between a cloud data store and a data store in a private network, and it 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 on an on-premises machine or a virtual machine (VM) inside a private network.

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

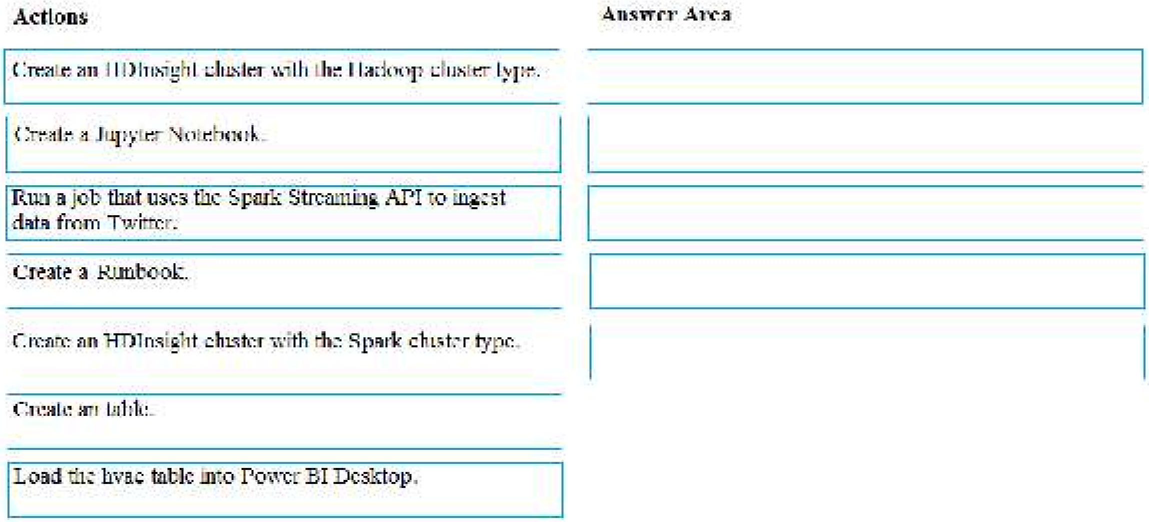
QUESTION 6
DRAG DROP

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.

Select and Place:

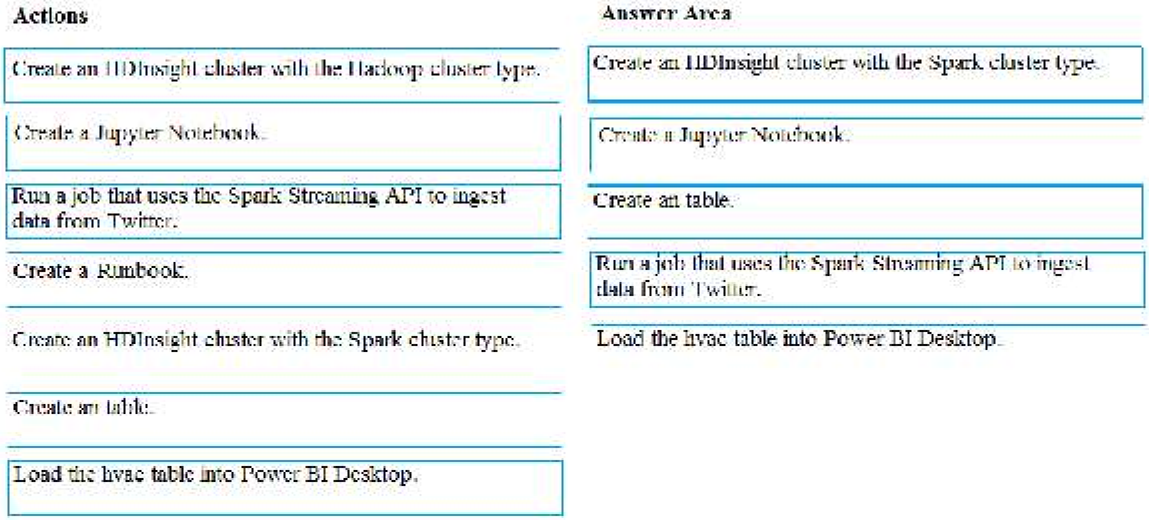


Actions

- Create an HDInsight cluster with the Hadoop cluster type.
- Create a Jupyter Notebook.
- Run a job that uses the Spark Streaming API to ingest data from Twitter.
- Create a Runbook.
- Create an HDInsight cluster with the Spark cluster type.
- Create an table.
- Load the hvac table into Power BI Desktop.

Answer Area

Correct Answer:



Actions

- Create an HDInsight cluster with the Hadoop cluster type.
- Create a Jupyter Notebook.
- Run a job that uses the Spark Streaming API to ingest data from Twitter.
- Create a Runbook.
- Create an HDInsight cluster with the Spark cluster type.
- Create an table.
- Load the hvac table into Power BI Desktop.

Answer Area

- Create an HDInsight cluster with the Spark cluster type.
- Create a Jupyter Notebook.
- Create an table.
- Run a job that uses the Spark Streaming API to ingest data from Twitter.
- Load the hvac table into Power BI Desktop.

Section: [none]
Explanation

Explanation/Reference:
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.

References:
https://acadgild.com/blog/streaming-twitter-data-using-spark
https://docs.microsoft.com/en-us/azure/hdinsight/spark/apache-spark-use-with-data-lake-store

QUESTION 7
DRAG DROP

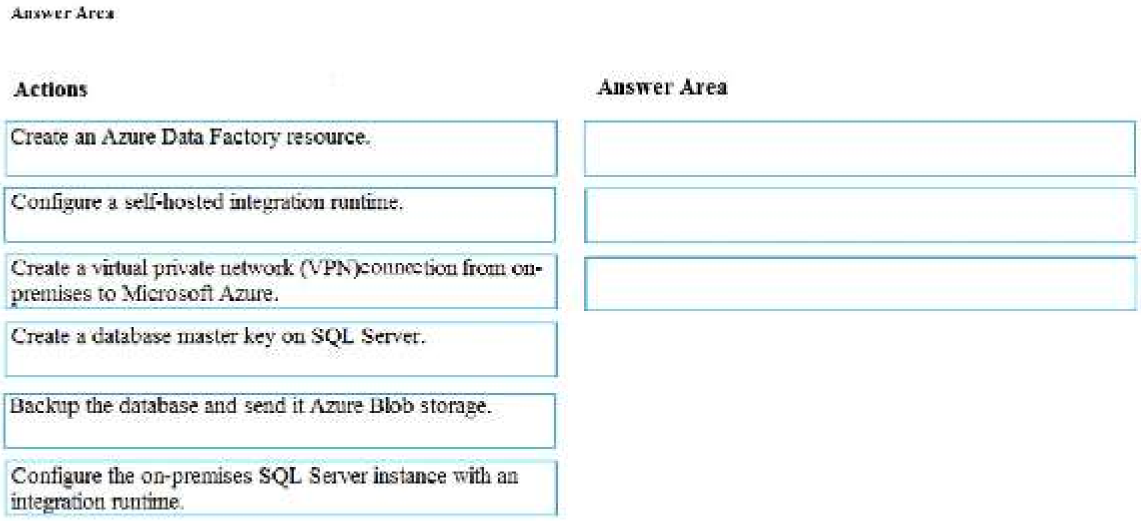
Your company manages on-premises Microsoft SQL Server pipelines by using a custom solution.

The data engineering team must implement a process to pull data from SQL Server and migrate it to Azure Blob storage. The process must orchestrate and manage the data lifecycle.

You need to configure Azure Data Factory to connect to the on-premises SQL Server database.

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.

Select and Place:



Answer Area

Actions

Create an Azure Data Factory resource.

Configure a self-hosted integration runtime.

Create a virtual private network (VPN)connection from on-premises to Microsoft Azure.

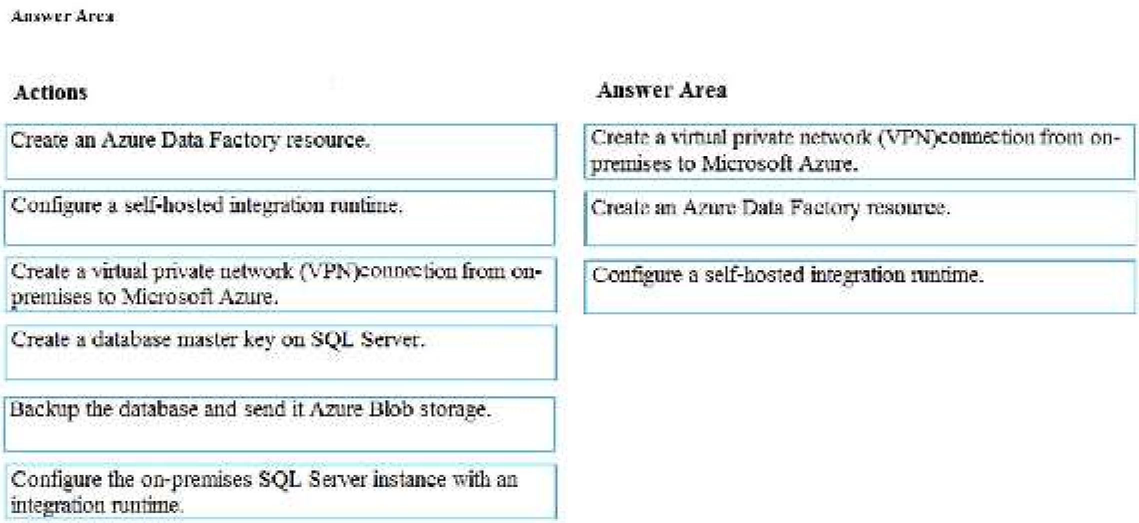
Create a database master key on SQL Server.

Backup the database and send it Azure Blob storage.

Configure the on-premises SQL Server instance with an integration runtime.

Answer Area

Correct Answer:



Answer Area

Actions	Answer Area
Create an Azure Data Factory resource.	Create a virtual private network (VPN)connection from on-premises to Microsoft Azure.
Configure a self-hosted integration runtime.	Create an Azure Data Factory resource.
Create a virtual private network (VPN)connection from on-premises to Microsoft Azure.	Configure a self-hosted integration runtime.
Create a database master key on SQL Server.	
Backup the database and send it Azure Blob storage.	
Configure the on-premises SQL Server instance with an integration runtime.	

Section: [none]
Explanation

Explanation/Reference:
Explanation:

Step 1: Create a virtual private network (VPN) connection from on-premises to Microsoft Azure.
You can also use IPSec VPN or Azure ExpressRoute to further secure the communication channel between your on-premises network and Azure.

Azure Virtual Network is a logical representation of your network in the cloud. You can connect an on-premises network to your virtual network by setting up IPSec VPN (site-to-site) or ExpressRoute (private peering).

Step 2: Create an Azure Data Factory resource.

Step 3: Configure a self-hosted integration runtime.
You create a self-hosted integration runtime and associate it with an on-premises machine with the SQL Server database. The self-hosted integration runtime is the component that copies data from the SQL Server database on your machine to Azure Blob storage.

Note: A self-hosted integration runtime can run copy activities between a cloud data store and a data store in a private network, and it 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 on an on-premises machine or a virtual machine (VM) inside a private network.

References:
https://docs.microsoft.com/en-us/azure/data-factory/tutorial-hybrid-copy-powershell

QUESTION 8
HOTSPOT

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

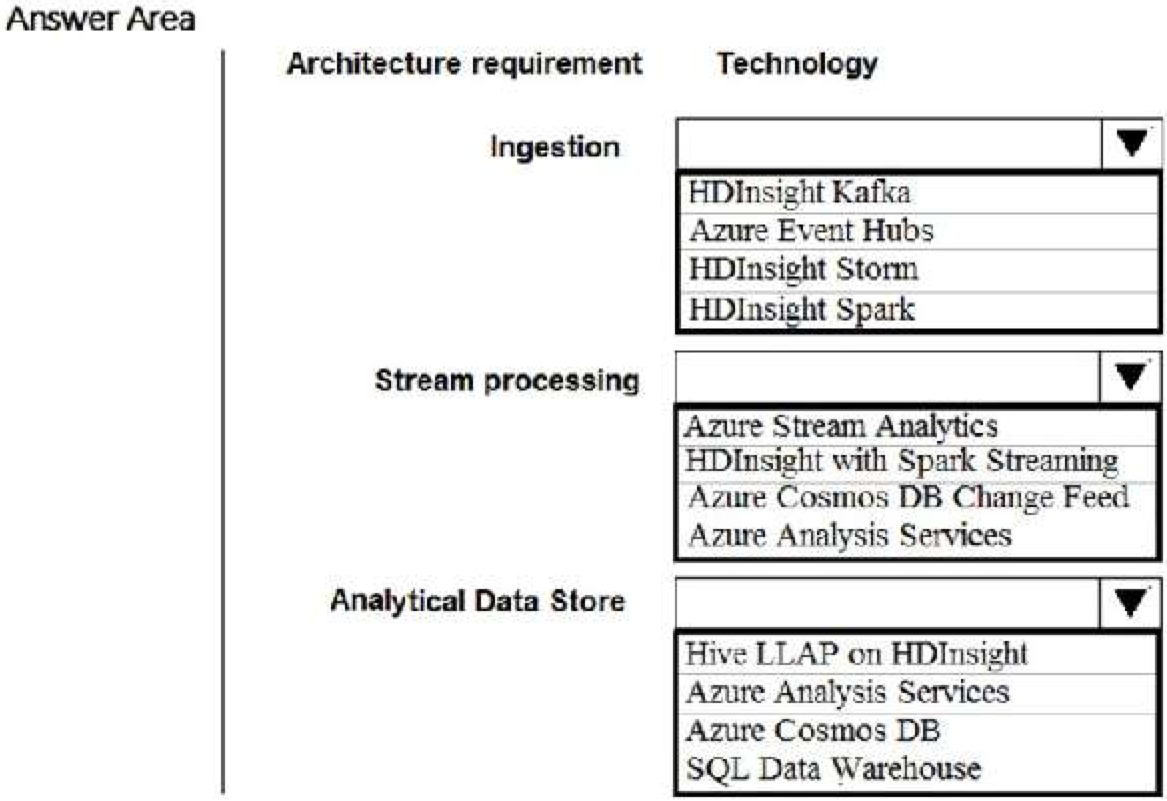
Analytical data store:

- Act as a managed service
- Use a document store
- Provide data encryption at rest

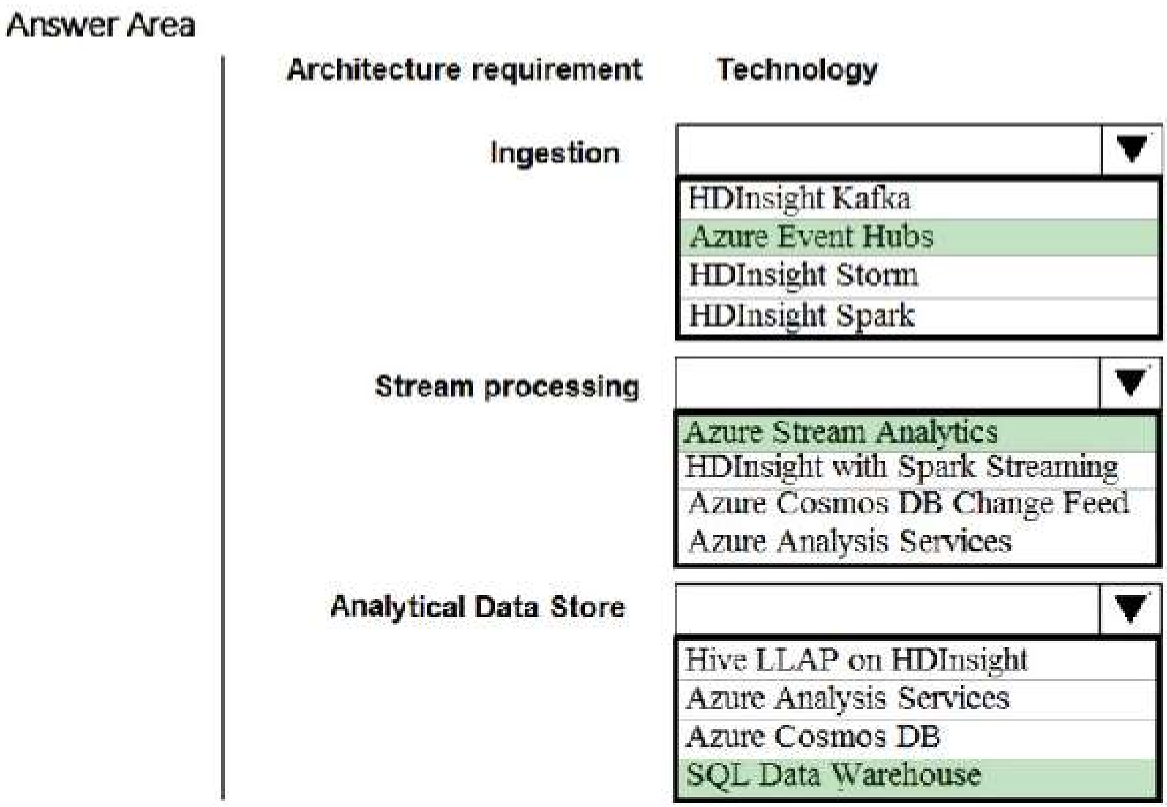
You need to identify the correct technologies to build the Lambda architecture using minimal effort. Which technologies should you use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:



Correct Answer:



Answer Area

Architecture requirement | Technology

Ingestion
- HDInsight Kafka
- [✓ Azure Event Hubs]
- HDInsight Storm
- HDInsight Spark

Stream processing
- [✓ Azure Stream Analytics]
- HDInsight with Spark Streaming
- Azure Cosmos DB Change Feed
- Azure Analysis Services

Analytical Data Store
- Hive LLAP on HDInsight
- Azure Analysis Services
- Azure Cosmos DB
- [✓ SQL Data Warehouse]

Section: [none]
Explanation

Explanation/Reference:
Explanation:

Box 1: Azure Event Hubs
This portion of a streaming architecture is often referred to as stream buffering. Options include Azure Event Hubs, Azure IoT Hub, and Kafka.

Incorrect Answers: Not HDInsight Kafka
Azure Functions need a trigger defined in order to run. There is a limited set of supported trigger types, and Kafka is not one of them.

Box 2: Azure Stream Analytics
Azure Stream Analytics provides a managed stream processing service based on perpetually running SQL queries that operate on unbounded streams.
You can also use open source Apache streaming technologies like Storm and Spark Streaming in an HDInsight cluster.

Box 3: Azure SQL Data Warehouse
Azure SQL Data Warehouse provides a managed service for large-scale, cloud-based data warehousing.
HDInsight supports Interactive Hive, HBase, and Spark SQL, which can also be used to serve data for analysis.

References:
https://docs.microsoft.com/en-us/azure/architecture/data-guide/big-data/

QUESTION 9
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.

NOTE: Each correct selection is worth one point.

A. Event Grid topic
B. Azure Stream Analytics Job that queries Twitter data from an Event Hub
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
E. Event Grid subscription
F. Event Hub instance

Correct Answer: BDF
Section: [none]
Explanation

Explanation/Reference:
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.

References:
https://community.powerbi.com/t5/Integrations-with-Files-and/Twitter-streaming-analytics-step-by-step/td-p/9594

QUESTION 10
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?

A. Azure SQL Data Warehouse
B. HDInsight Apache Storm cluster
C. Azure Stream Analytics
D. HDInsight Apache Hadoop cluster using MapReduce
E. HDInsight Spark cluster

Correct Answer: B
Section: [none]
Explanation

Explanation/Reference:
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). Data flows between the components in the graph. Each component consumes one or more data streams, and can optionally emit one or more streams.

Python can be used to develop Storm components.

References:
https://docs.microsoft.com/en-us/azure/hdinsight/storm/apache-storm-overview

QUESTION 11
HOTSPOT

A company is deploying a service-based data environment. You are developing a solution to process this data.

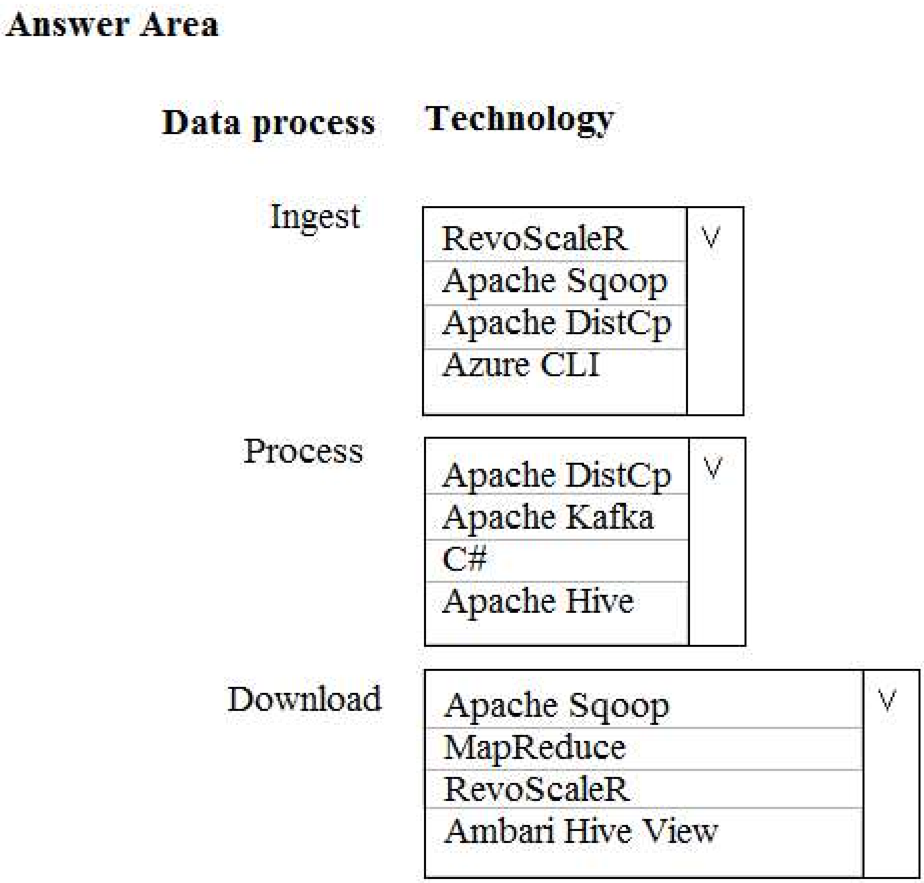
The solution must meet the following requirements:

- Use an Azure HDInsight cluster for data ingestion from a relational database in a different cloud service
- Use an Azure Data Lake Storage account to store processed data
- Allow users to download processed data

You need to recommend technologies for the solution.

Which technologies should you use? To answer, select the appropriate options in the answer area.

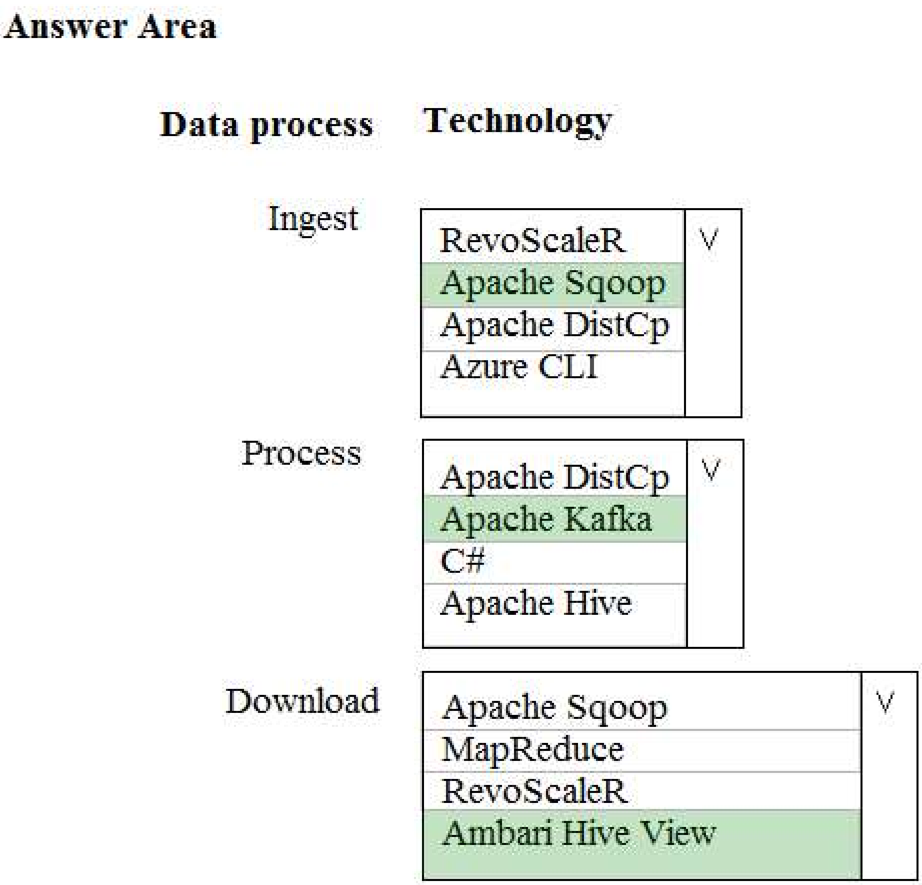
Hot Area:



Answer Area

Data process	Technology
Ingest	RevoScaleR Apache Sqoop Apache DistCp Azure CLI
Process	Apache DistCp Apache Kafka C# Apache Hive
Download	Apache Sqoop MapReduce RevoScaleR Ambari Hive View

Correct Answer:



Answer Area

Data process **Technology**

Ingest
- RevoScaleR
- [✓ Apache Sqoop]
- Apache DistCp
- Azure CLI

Process
- Apache DistCp
- [✓ Apache Kafka]
- C#
- Apache Hive

Download
- Apache Sqoop
- MapReduce
- RevoScaleR
- [✓ Ambari Hive View]

Section: [none]
Explanation

Explanation/Reference:
Explanation:

Box 1: Apache Sqoop
Apache Sqoop is a tool designed for efficiently transferring bulk data between Apache Hadoop and structured datastores such as relational databases.

Azure HDInsight is a cloud distribution of the Hadoop components from the Hortonworks Data Platform (HDP).

Incorrect Answers:
DistCp (distributed copy) is a tool used for large inter/intra-cluster copying. It uses MapReduce to effect its distribution, error handling and recovery, and reporting. It expands a list of files and directories into input to map tasks, each of which will copy a partition of the files specified in the source list. Its MapReduce pedigree has endowed it with some quirks in both its semantics and execution.

RevoScaleR is a collection of proprietary functions in Machine Learning Server used for practicing data science at scale. For data scientists, RevoScaleR gives you data-related functions for import, transformation and manipulation, summarization, visualization, and analysis.

Box 2: Apache Kafka
Apache Kafka is a distributed streaming platform.
A streaming platform has three key capabilities:
Publish and subscribe to streams of records, similar to a message queue or enterprise messaging system.
Store streams of records in a fault-tolerant durable way.
Process streams of records as they occur.

Kafka is generally used for two broad classes of applications:
Building real-time streaming data pipelines that reliably get data between systems or applications
Building real-time streaming applications that transform or react to the streams of data

Box 3: Ambari Hive View
You can 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.

References:
https://sqoop.apache.org/

https://kafka.apache.org/intro

https://docs.microsoft.com/en-us/azure/hdinsight/hadoop/apache-hadoop-use-hive-ambari-view

QUESTION 12
A company uses Azure SQL Database to store sales transaction data. Field sales employees need an offline copy of the database that includes last year's sales on their laptops when there is no internet connection available.

You need to create the offline export copy.

Which three options can you use? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

A. Export to a BACPAC file by using Azure Cloud Shell, and save the file to an Azure storage account
B. Export to a BACPAC file by using SQL Server Management Studio. Save the file to an Azure storage account
C. Export to a BACPAC file by using the Azure portal
D. Export to a BACPAC file by using Azure PowerShell and save the file locally
E. Export to a BACPAC file by using the SqlPackage utility

Correct Answer: BCE
Section: [none]
Explanation

Explanation/Reference:
Explanation:
You can export to a BACPAC file using the Azure portal.
You can export to a BACPAC file using SQL Server Management Studio (SSMS). The newest versions of SQL Server Management Studio provide a wizard to export an Azure SQL database to a BACPAC file.
You can export to a BACPAC file using the SQLPackage utility.

Incorrect Answers:
D: You can export to a BACPAC file using PowerShell. Use the New-AzSqlDatabaseExport cmdlet to submit an export database request to the Azure SQL Database service. Depending on the size of your database, the export operation may take some time to complete. However, the file is not stored locally.

References:
https://docs.microsoft.com/en-us/azure/sql-database/sql-database-export

QUESTION 13
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 Data Lake Gen 2 storage account
2. Create an external file format and external table using the external data source
3. Load the data using the `CREATE TABLE AS SELECT` statement

Does the solution meet the goal?

A. Yes
B. No

Correct Answer: A
Section: [none]
Explanation

Explanation/Reference:
Explanation:
You need to create an external file format and external table using the external data source.
You load the data using the CREATE TABLE AS SELECT statement.

References:
https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-load-from-azure-data-lake-store

QUESTION 14
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 a remote service binding pointing to the Azure Data Lake Gen 2 storage account
2. Create an external file format and external table using the external data source
3. Load the data using the `CREATE TABLE AS SELECT` statement

Does the solution meet the goal?

A. Yes
B. No

Correct Answer: B
Section: [none]
Explanation

Explanation/Reference:
Explanation:
You need to create an external file format and external table from an external data source, instead from a remote service binding pointing.

References:

https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-load-from-azure-data-lake-store

QUESTION 15
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?

A. Yes
B. No

Correct Answer: B
Section: [none]
Explanation

Explanation/Reference:
Explanation:
Use the Azure Data Lake Gen 2 storage account.

References:
https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-load-from-azure-data-lake-store

QUESTION 16
You need to develop a pipeline for processing data. The pipeline must meet the following requirements:

- Scale up and down resources for cost reduction
- Use an in-memory data processing engine to speed up ETL and machine learning operations.
- Use streaming capabilities
- Provide the ability to code in SQL, Python, Scala, and R
- Integrate workspace collaboration with Git

What should you use?

A. HDInsight Spark Cluster
B. Azure Stream Analytics
C. HDInsight Hadoop Cluster
D. Azure SQL Data Warehouse
E. HDInsight Kafka Cluster
F. HDInsight Storm Cluster

Correct Answer: A
Section: [none]
Explanation

Explanation/Reference:
Explanation:
Aparch Spark is an open-source, parallel-processing framework that supports in-memory processing to boost the performance of big-data analysis applications.

HDInsight is a managed Hadoop service. Use it deploy and manage Hadoop clusters in Azure. For batch processing, you can use Spark, Hive, Hive LLAP, MapReduce.

Languages: R, Python, Java, Scala, SQL

You can create an HDInsight Spark cluster using an Azure Resource Manager template. The template can be found in GitHub.

References:
https://docs.microsoft.com/en-us/azure/architecture/data-guide/technology-choices/batch-processing

QUESTION 17
DRAG DROP

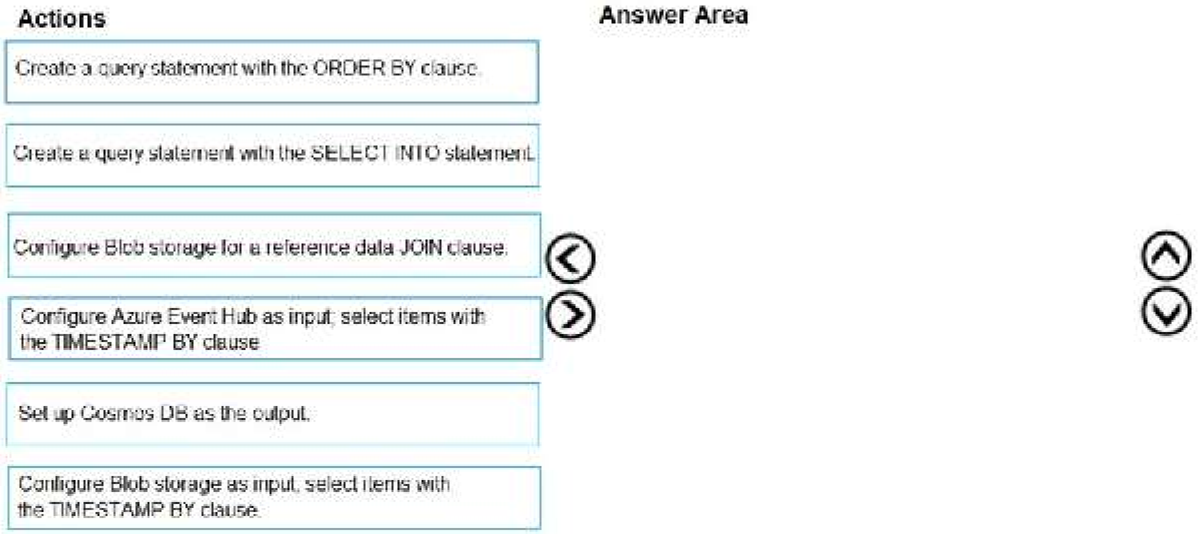
You implement an event processing solution using Microsoft Azure Stream Analytics.

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.

Select and Place:



Actions

Create a query statement with the ORDER BY clause.

Create a query statement with the SELECT INTO statement.

Configure Blob storage for a reference data JOIN clause.

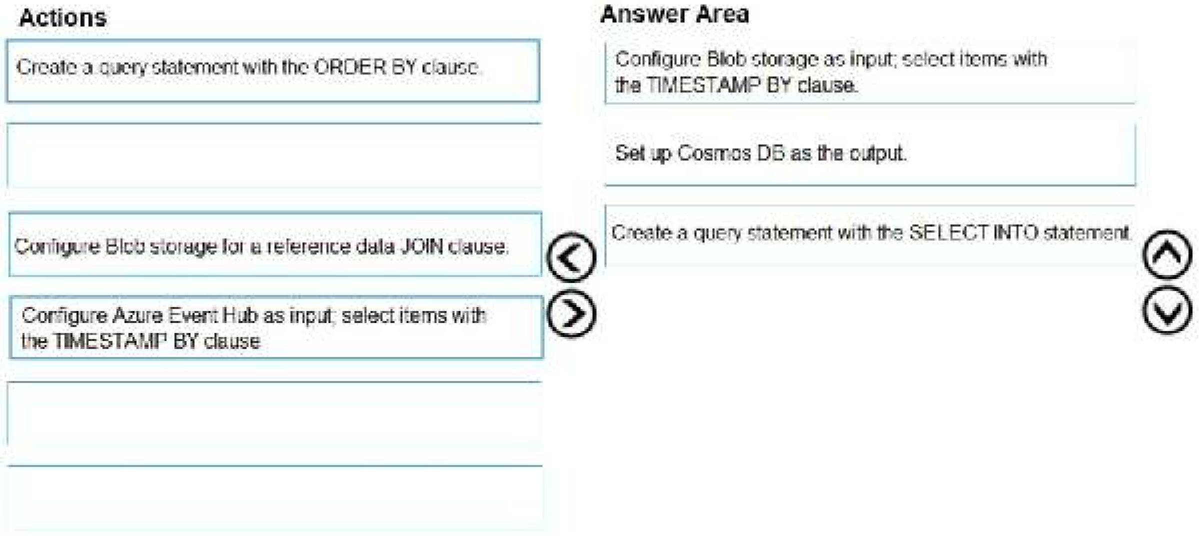
Configure Azure Event Hub as input; select items with the TIMESTAMP BY clause

Set up Cosmos DB as the output.

Configure Blob storage as input; select items with the TIMESTAMP BY clause.

Answer Area

Correct Answer:



Actions

Create a query statement with the ORDER BY clause.

Configure Blob storage for a reference data JOIN clause.

Configure Azure Event Hub as input; select items with the TIMESTAMP BY clause

Answer Area

Configure Blob storage as input; select items with the TIMESTAMP BY clause.

Set up Cosmos DB as the output.

Create a query statement with the SELECT INTO statement.

Section: [none]
Explanation

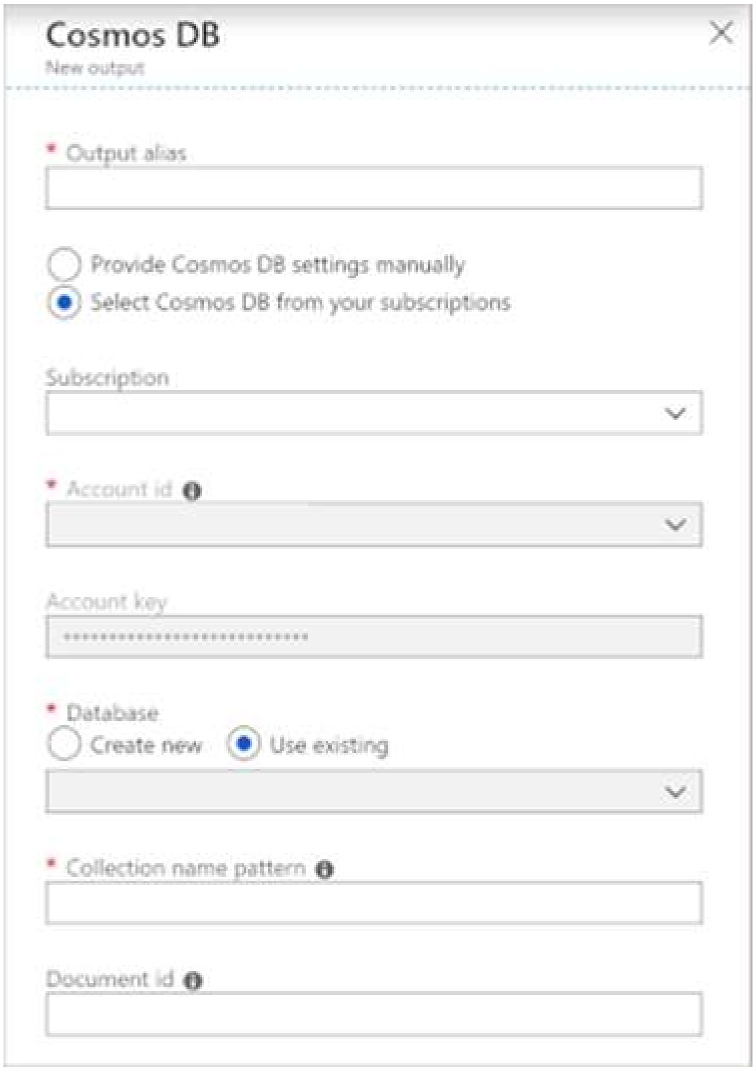
Explanation/Reference:
Explanation:

Step 1: Configure Blob storage as input; select items with the TIMESTAMP BY clause
The default timestamp of Blob storage events in Stream Analytics is the timestamp that the blob was last modified, which is BlobLastModifiedUtcTime. To process the data as a stream using a timestamp in the event payload, you must use the TIMESTAMP BY keyword.

Example:
The following is a TIMESTAMP BY example which uses the EntryTime column as the application time for events:

SELECT TollId, EntryTime AS VehicleEntryTime, LicensePlate, State, Make, Model, VehicleType, VehicleWeight, Toll, Tag
FROM TollTagEntry TIMESTAMP BY EntryTime

Step 2: Set up cosmos DB as the output
Creating Cosmos DB as an output in Stream Analytics generates a prompt for information as seen below.



Cosmos DB

New output

* Output alias

○ Provide Cosmos DB settings manually

● Select Cosmos DB from your subscriptions

Subscription

* Account id

Account key

****

* Database

○ Create new ● Use existing

* Collection name pattern

Document id

Step 3: Create a query statement with the SELECT INTO statement.

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

QUESTION 18
HOTSPOT

A company plans to use Platform-as-a-Service (PaaS) to create the new data pipeline process. The process must meet the following requirements:

Ingest:

- Access multiple data sources.
- Provide the ability to orchestrate workflow.
- Provide the capability to run SQL Server Integration Services packages.

Store:

- Optimize storage for big data workloads
- Provide encryption of data at rest.
- Operate with no size limits.

Prepare and Train:

- Provide a fully-managed and interactive workspace for exploration and visualization.
- Provide the ability to program in R, SQL, Python, Scala, and Java.
- Provide seamless user authentication with Azure Active Directory.

Model & Serve:

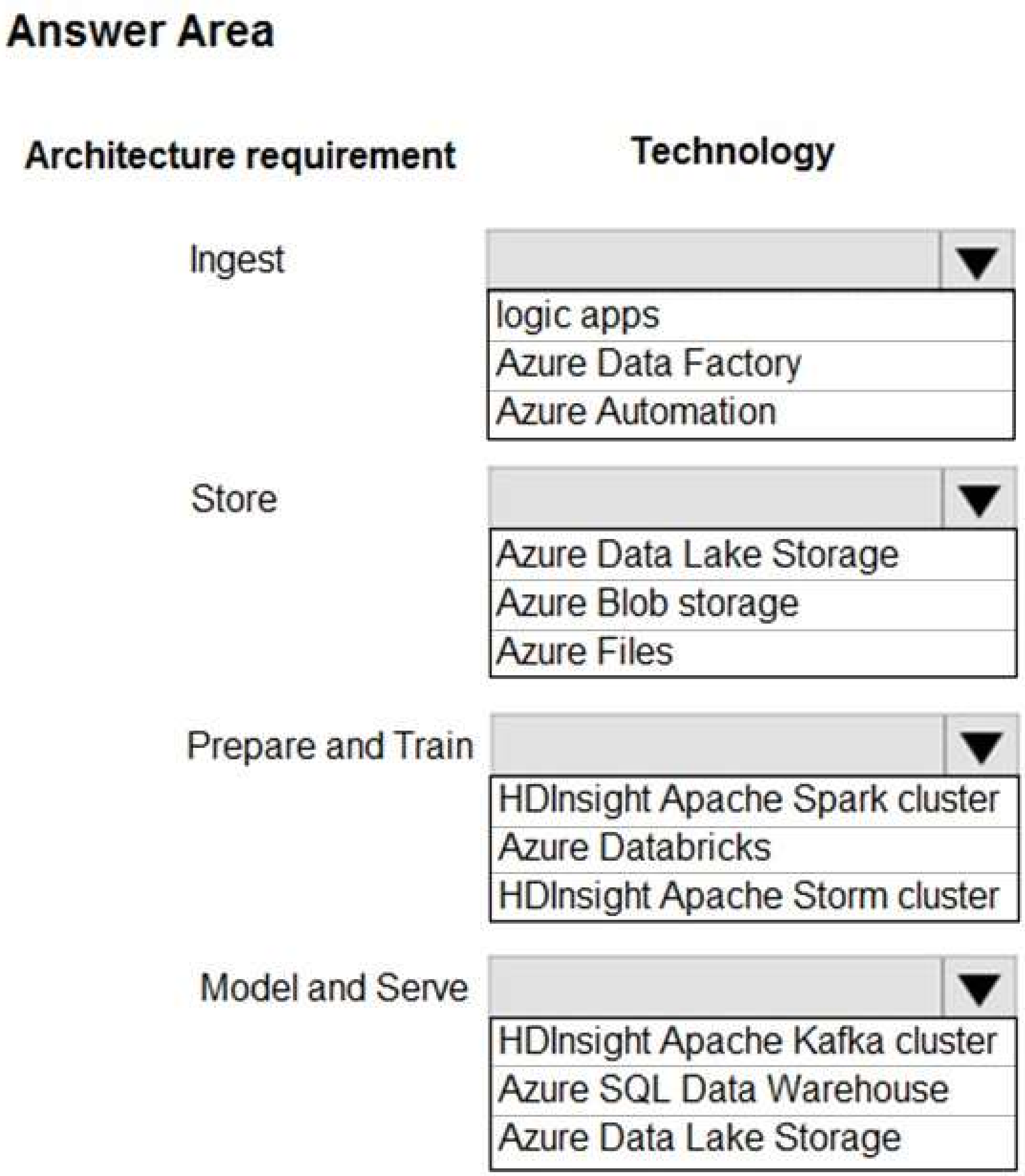
- Implement native columnar storage.
- Support for the SQL language.
- Provide support for structured streaming.

You need to build the data integration pipeline.

Which technologies should you use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

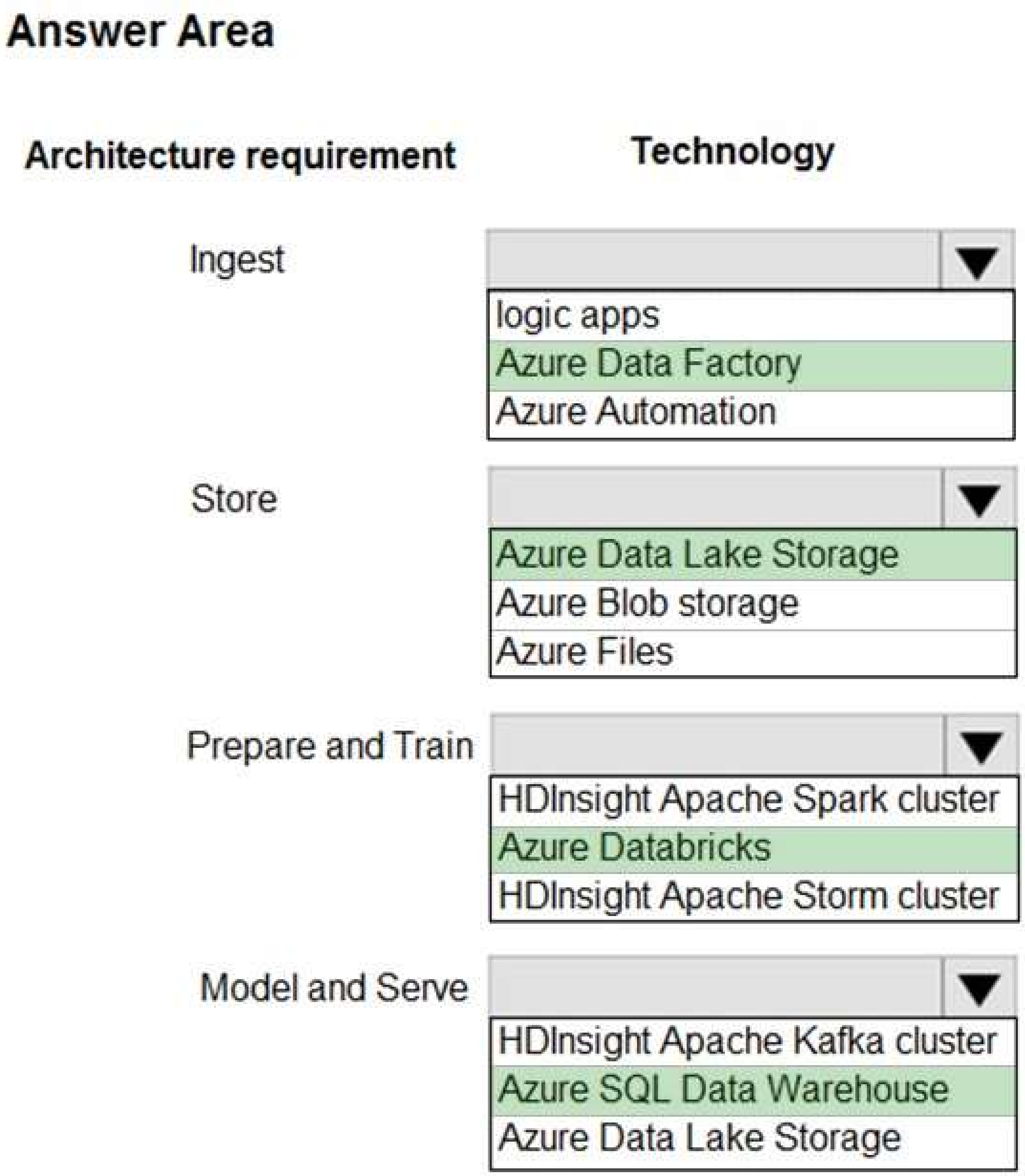
Hot Area:



Answer Area

Architecture requirement	Technology
Ingest	▼ logic apps Azure Data Factory Azure Automation
Store	▼ Azure Data Lake Storage Azure Blob storage Azure Files
Prepare and Train	▼ HDInsight Apache Spark cluster Azure Databricks HDInsight Apache Storm cluster
Model and Serve	▼ HDInsight Apache Kafka cluster Azure SQL Data Warehouse Azure Data Lake Storage

Correct Answer:



Answer Area

Architecture requirement | Technology

Ingest
- logic apps
- [✓ Azure Data Factory]
- Azure Automation

Store
- [✓ Azure Data Lake Storage]
- Azure Blob storage
- Azure Files

Prepare and Train
- HDInsight Apache Spark cluster
- [✓ Azure Databricks]
- HDInsight Apache Storm cluster

Model and Serve
- HDInsight Apache Kafka cluster
- [✓ Azure SQL Data Warehouse]
- Azure Data Lake Storage

Section: [none]
Explanation

Explanation/Reference:
Explanation:

Ingest: Azure Data Factory
Azure Data Factory pipelines can execute SSIS packages.
In Azure, the following services and tools will meet the core requirements for pipeline orchestration, control flow, and data movement: Azure Data Factory, Oozie on HDInsight, and SQL Server Integration Services (SSIS).

Store: Data Lake Storage
Data Lake Storage Gen1 provides unlimited storage.

Note: Data at rest includes information that resides in persistent storage on physical media, in any digital format. Microsoft Azure offers a variety of data storage solutions to meet different needs, including file, disk, blob, and table storage. Microsoft also provides encryption to protect Azure SQL Database, Azure Cosmos DB, and Azure Data Lake.

Prepare and Train: Azure Databricks
Azure Databricks provides enterprise-grade Azure security, including Azure Active Directory integration.
With Azure Databricks, you can set up your Apache Spark environment in minutes, autoscale and collaborate on shared projects in an interactive workspace. Azure Databricks supports Python, Scala, R, Java and SQL, as well as data science frameworks and libraries including TensorFlow, PyTorch and scikit-learn.

Model and Serve: SQL Data Warehouse
SQL Data Warehouse stores data into relational tables with columnar storage.
Azure SQL Data Warehouse connector now offers efficient and scalable structured streaming write support for SQL Data Warehouse. Access SQL Data Warehouse from Azure Databricks using the SQL Data Warehouse connector.

References:
https://docs.microsoft.com/bs-latn-ba/azure/architecture/data-guide/technology-choices/pipeline-orchestration-data-movement

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

QUESTION 19
HOTSPOT

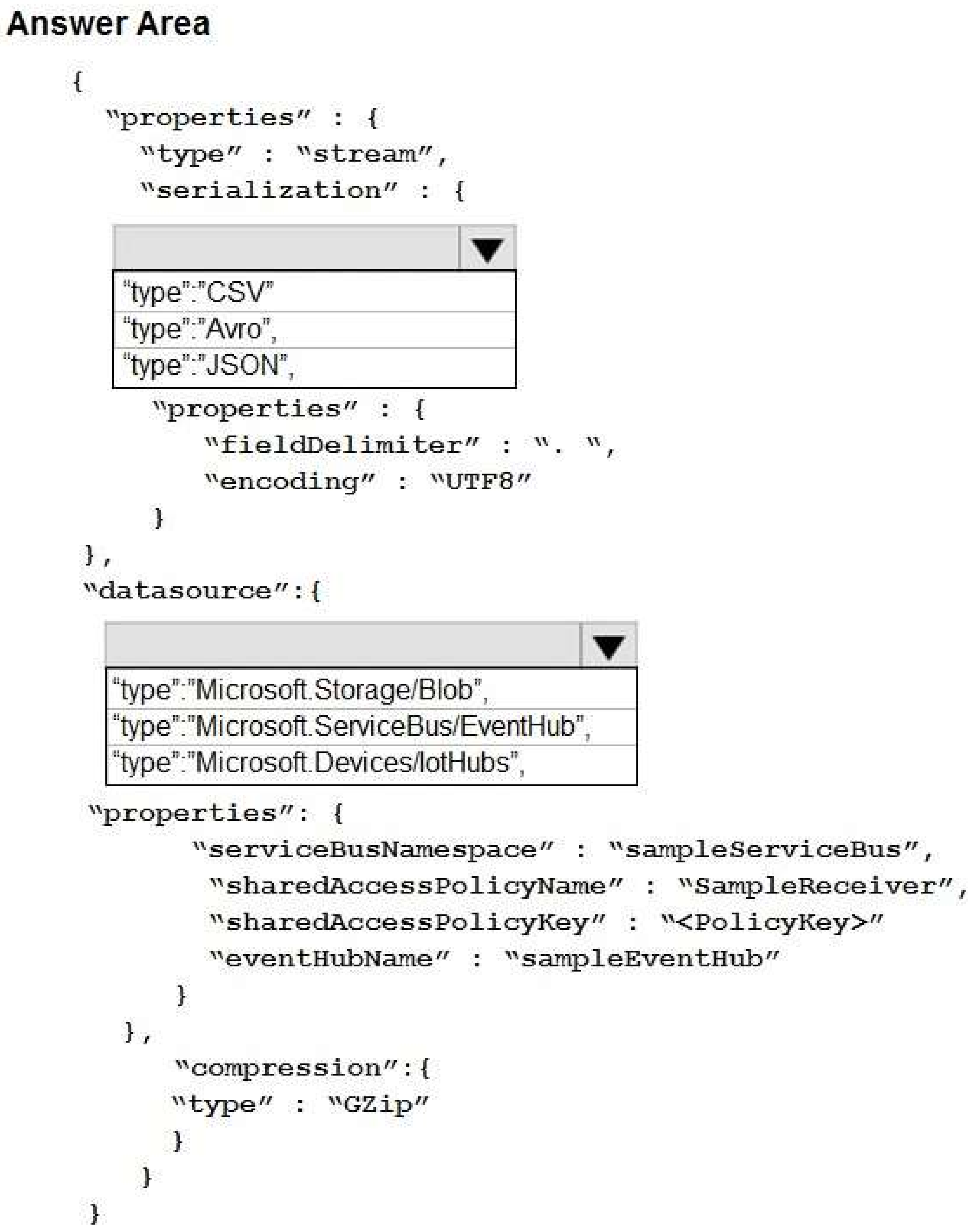
A company plans to analyze a continuous flow of data from a social media platform by using Microsoft Azure Stream Analytics. The incoming data is formatted as one record per row.

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.

NOTE: Each correct selection is worth one point.

Hot Area:



Answer Area

```
{
  “properties” : {
    “type” : “stream”,
    “serialization” : {
```

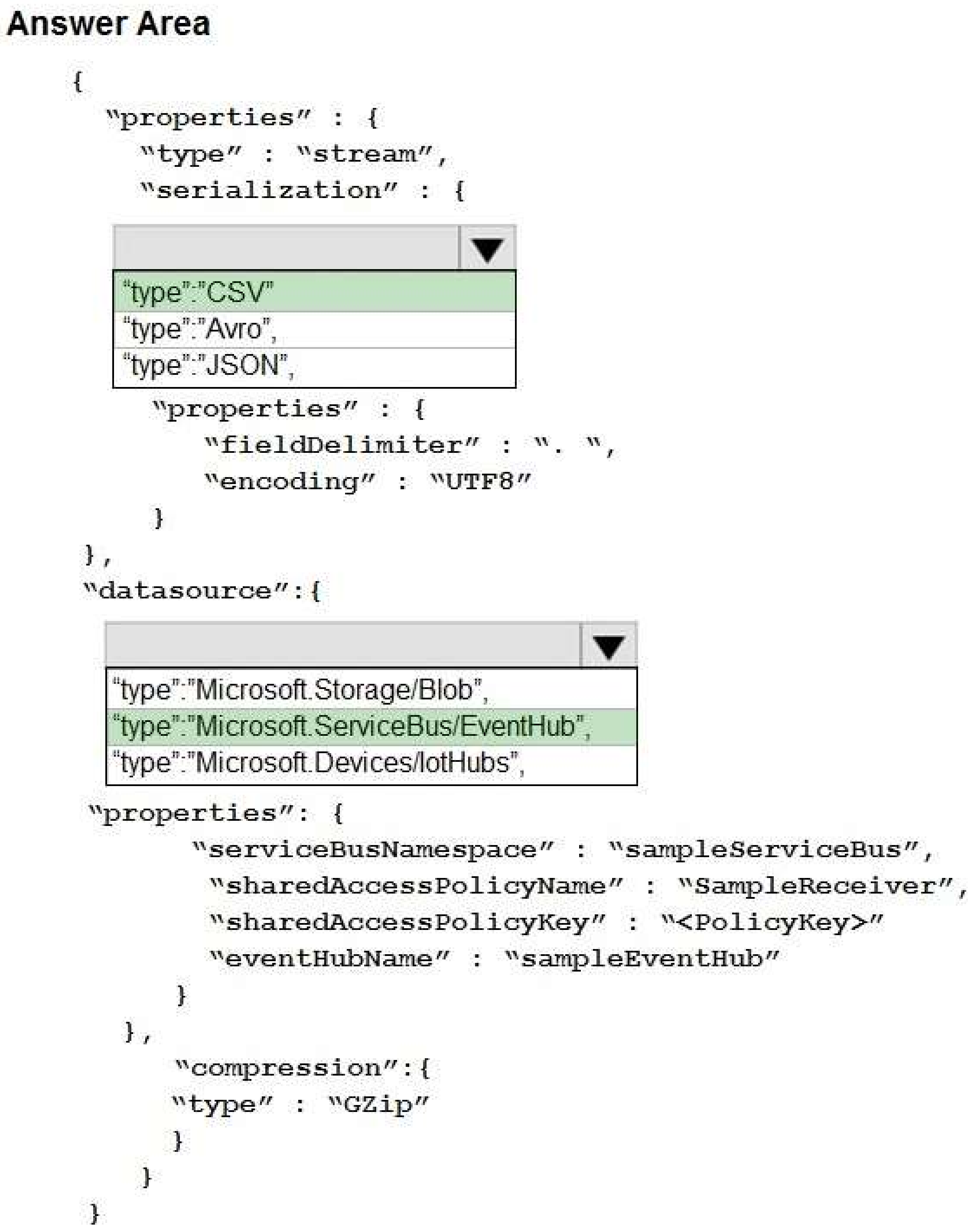
“type”:”CSV”
“type”:”Avro”,
“type”:”JSON”,

```
     “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”
      }
    }
}
```

Correct Answer:



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"
      }
    }
}
```

Section: [none]
Explanation

Explanation/Reference:
Explanation:

Box 1: CSV
A comma-separated values (CSV) file is a delimited text file that uses a comma to separate values. A CSV file stores tabular data (numbers and text) in plain text. Each line of the file is a data record.

JSON and AVRO are not formatted as one record per row.

Box 2: "type":"Microsoft.ServiceBus/EventHub",
Properties include "EventHubName"

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

https://en.wikipedia.org/wiki/Comma-separated_values

QUESTION 20
DRAG DROP

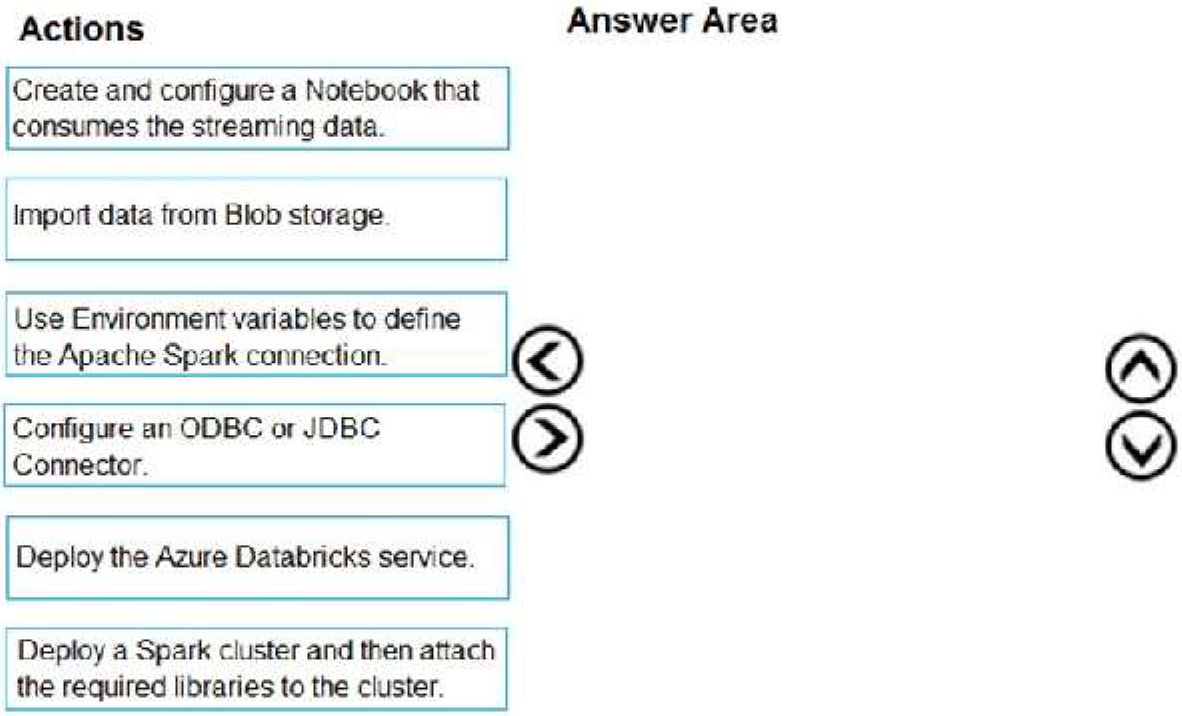
Your company plans to create an event processing engine to handle streaming data from Twitter.

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

Select and Place:



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.

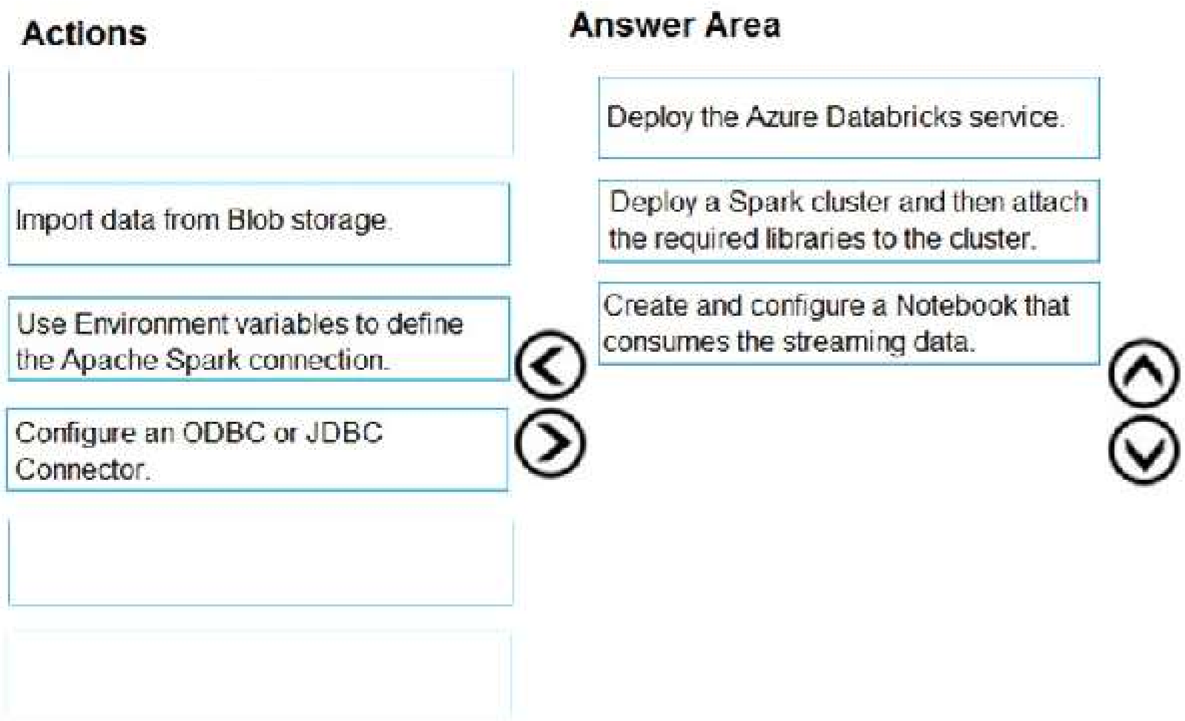
Configure an ODBC or JDBC Connector.

Deploy the Azure Databricks service.

Deploy a Spark cluster and then attach the required libraries to the cluster.

Answer Area

Correct Answer:



Actions

Import data from Blob storage.

Use Environment variables to define the Apache Spark connection.

Configure an ODBC or JDBC Connector.

Answer Area

Deploy the Azure Databricks service.

Deploy a Spark cluster and then attach the required libraries to the cluster.

Create and configure a Notebook that consumes the streaming data.

Section: [none]
Explanation

Explanation/Reference:
Explanation:

Step 1: Deploy the Azure Databricks service
Create an Azure Databricks workspace by setting up an Azure Databricks Service.

Step 2: Deploy a Spark cluster and then attach the required libraries to the cluster.
To create a Spark cluster in Databricks, in the Azure portal, go to the Databricks workspace that you created, and then select Launch Workspace.

Attach libraries to Spark cluster: you use the Twitter APIs to send tweets to Event Hubs. You also use the Apache Spark Event Hubs connector to read and write data into Azure Event Hubs. To use these APIs as part of your cluster, add them as libraries to Azure Databricks and associate them with your Spark cluster.

Step 3: Create and configure a Notebook that consumes the streaming data.
You create a notebook named ReadTweetsFromEventhub in Databricks workspace.
ReadTweetsFromEventHub is a consumer notebook you use to read the tweets from Event Hubs.

References:
https://docs.microsoft.com/en-us/azure/azure-databricks/databricks-stream-from-eventhubs

QUESTION 21
HOTSPOT

You develop data engineering solutions for a company.

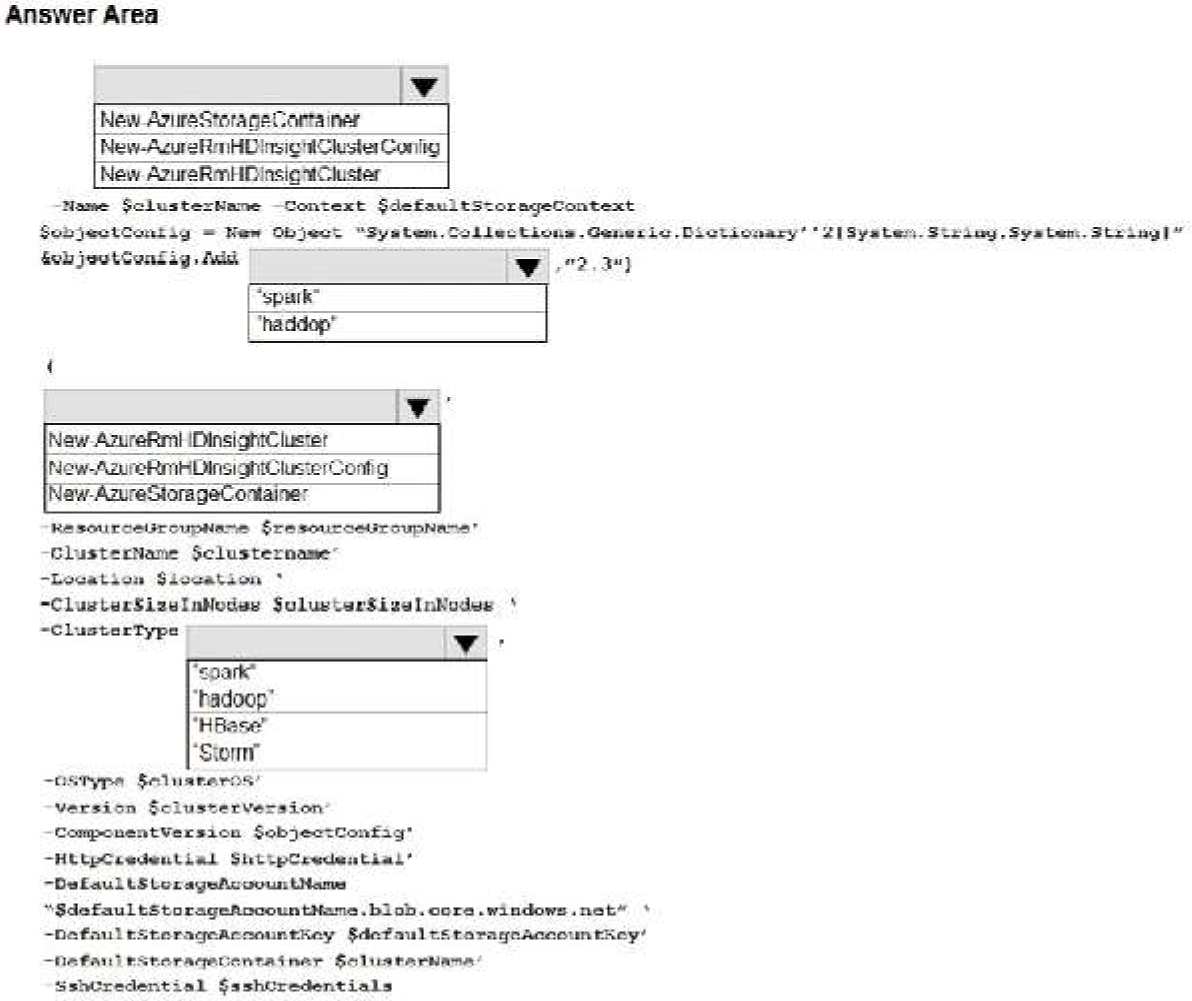
A project requires an in-memory batch data processing solution.

You need to provision an HDInsight cluster for batch processing of data on Microsoft Azure.

How should you complete the PowerShell segment? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

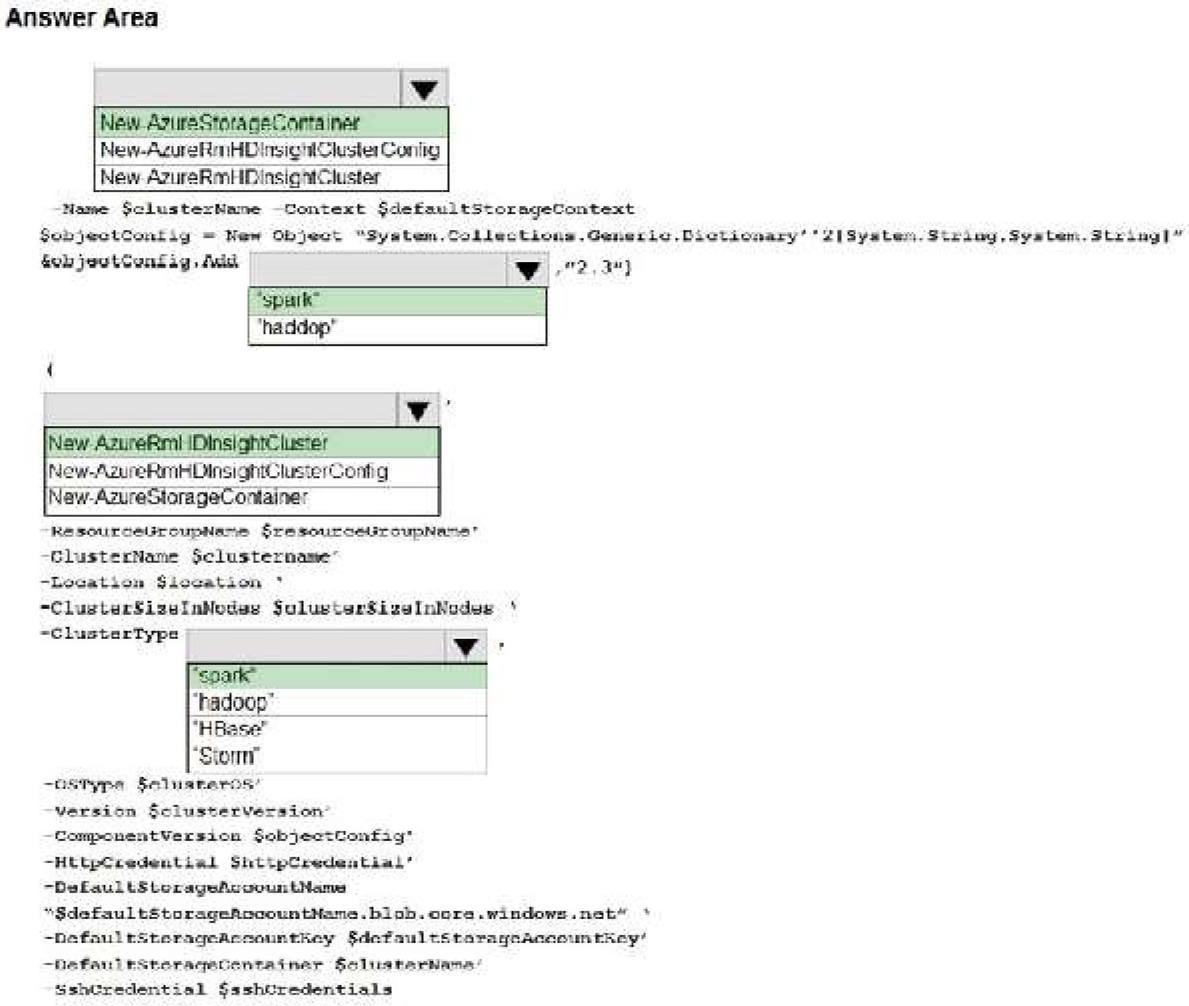
Hot Area:



Answer Area

```
New-AzureStorageContainer
New-AzureRmHDInsightClusterConfig
New-AzureRmHDInsightCluster
 -Name $clusterName -Context $defaultStorageContext
$objectConfig = New Object "System.Collections.Generic.Dictionary''2[System.String.System.String]"
&objectConfig.Add                   ,"2.3")
                  "spark"
                  'haddop'
(
New-AzureRmHDInsightCluster
New-AzureRmHDInsightClusterConfig
New-AzureStorageContainer
-ResourceGroupName $resourceGroupName'
-ClusterName $clustername'
-Location $location '
-ClusterSizeInNodes $clusterSizeInNodes '
-ClusterType                     '
              "spark"
              "hadoop"
              "HBase"
              "Storm"
-OSType $clusterOS'
-Version $clusterVersion'
-ComponentVersion $objectConfig'
-HttpCredential $httpCredential'
-DefaultStorageAccountName
"$defaultStorageAccountName.blob.core.windows.net" '
-DefaultStorageAccountKey $defaultStorageAccountKey'
-DefaultStorageContainer $clusterName'
-SshCredential $sshCredentials
```

Correct Answer:



Answer Area

```
New-AzureStorageContainer
New-AzureRmHDInsightClusterConig
New-AzureRmHDInsightCluster
 -Name $clusterName -Context $defaultStorageContext
$objectConfig = New Object "System.Collections.Generic.Dictionary''2[System.String,System.String]"
&objectConfig.Add                                              ,"2.3")
                   "spark"
                   'haddop'
(
New-AzureRmHDInsightCluster                     '
New-AzureRmHDInsightClusterConfig
New-AzureStorageContainer
-ResourceGroupName $resourceGroupName'
-ClusterName $clustername'
-Location $location '
-ClusterSizeInNodes $clusterSizeInNodes '
-ClusterType                                    '
             "spark"
             'hadoop'
             'HBase'
             'Storm'
-OSType $clusterOS'
-Version $clusterVersion'
-ComponentVersion $objectConfig'
-HttpCredential $httpCredential'
-DefaultStorageAccountName
"$defaultStorageAccountName.blob.core.windows.net" '
-DefaultStorageAccountKey $defaultStorageAccountKey'
-DefaultStorageContainer $clusterName'
-SshCredential $sshCredentials
```

Section: [none]
Explanation

Explanation/Reference:
Explanation:

Box 1: New-AzStorageContainer
# Example: Create a blob container. This holds the default data store for the cluster.
New-AzStorageContainer `
    -Name $clusterName `
    -Context $defaultStorageContext

$sparkConfig = New-Object "System.Collections.Generic.Dictionary``2[System.String,System.String]"
$sparkConfig.Add("spark", "2.3")

Box 2: Spark
Spark provides primitives for in-memory cluster computing. A Spark job can load and cache data into memory and query it repeatedly. In-memory computing is much faster than disk-based applications than disk-based applications, such as Hadoop, which shares data through Hadoop distributed file system (HDFS).

Box 3: New-AzureRMHDInsightCluster
# Create the HDInsight cluster. Example:
New-AzHDInsightCluster `
    -ResourceGroupName $resourceGroupName `

```
    -ClusterName $clusterName `
    -Location $location `
    -ClusterSizeInNodes $clusterSizeInNodes `
    -ClusterType $"Spark" `
    -OSType "Linux" `
```

Box 4: Spark
HDInsight is a managed Hadoop service. Use it deploy and manage Hadoop clusters in Azure. For batch processing, you can use Spark, Hive, Hive LLAP, MapReduce.

References:
https://docs.microsoft.com/bs-latn-ba/azure/hdinsight/spark/apache-spark-jupyter-spark-sql-use-powershell

https://docs.microsoft.com/bs-latn-ba/azure/hdinsight/spark/apache-spark-overview

QUESTION 22
HOTSPOT

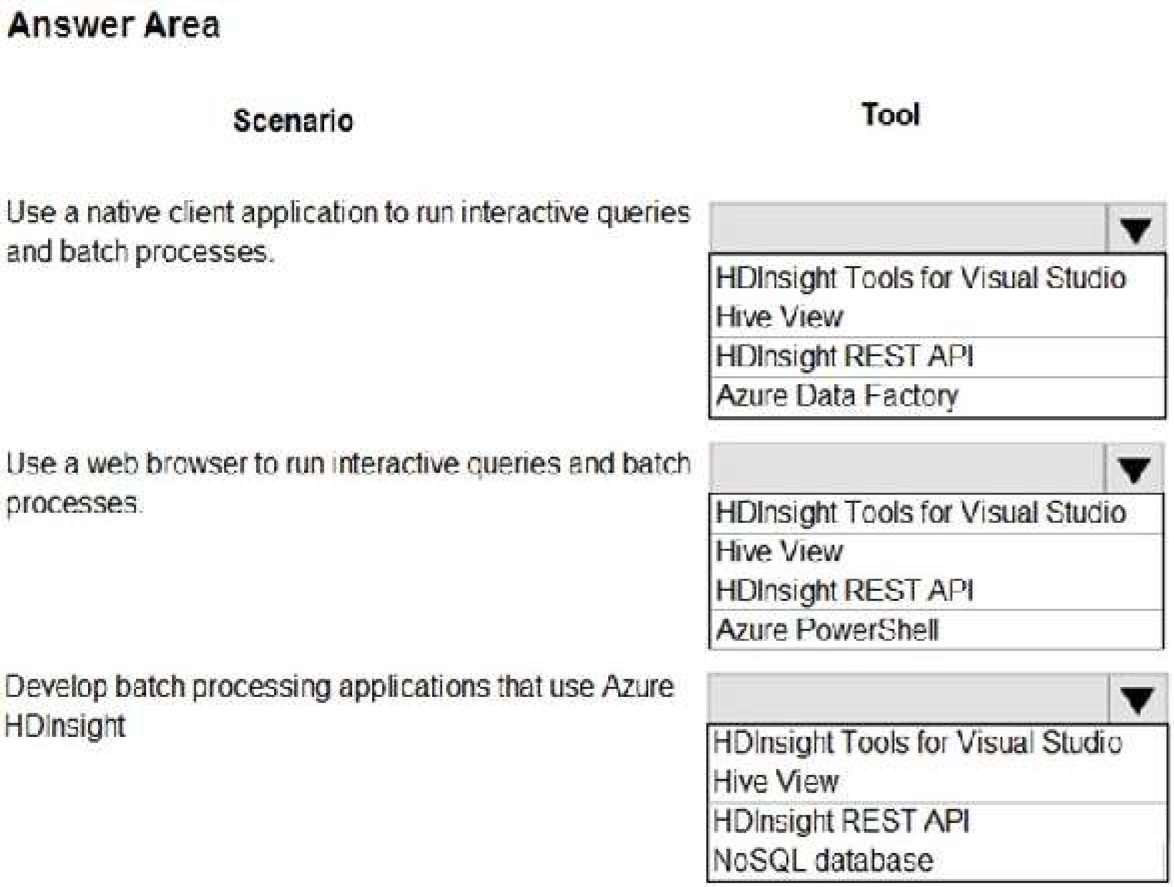
A company plans to develop solutions to perform batch processing of multiple sets of geospatial data.

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:



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

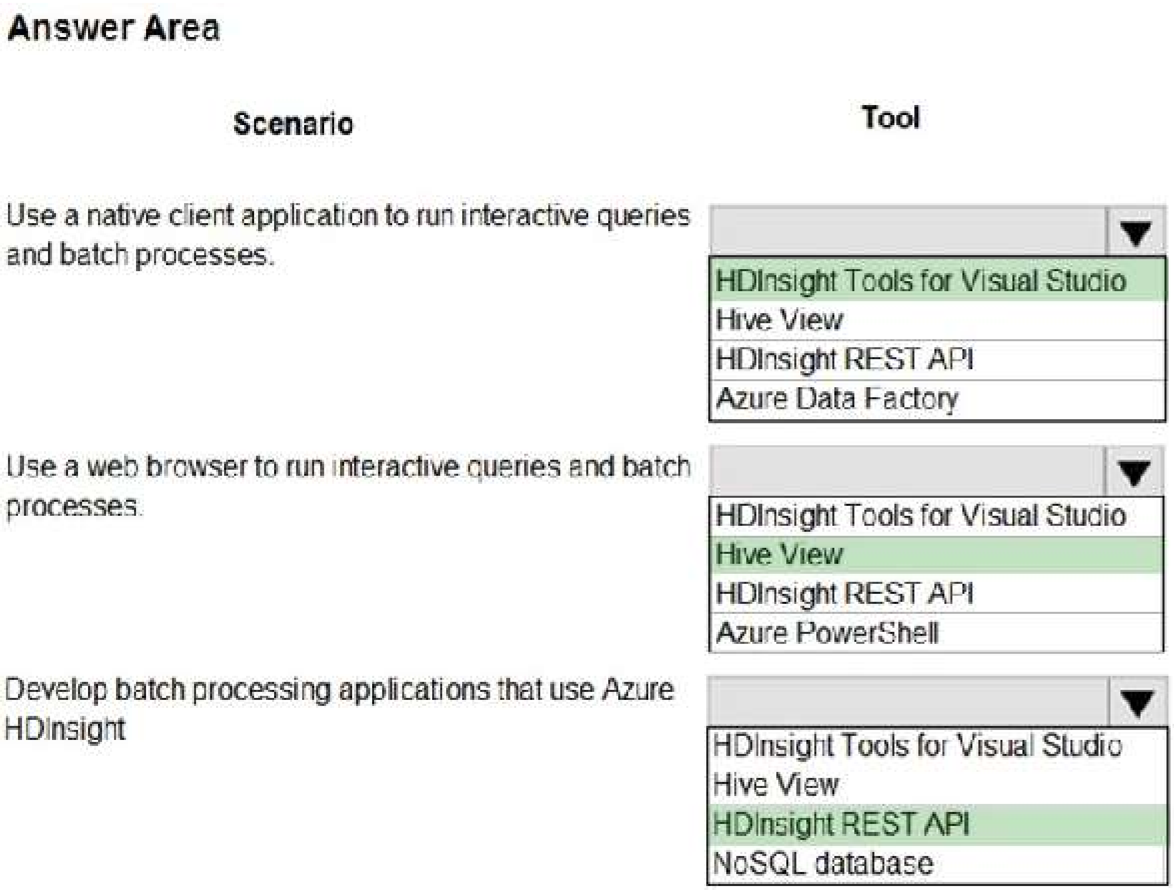
Use a web browser to run interactive queries and batch processes.

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

Correct Answer:



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

Section: [none]
Explanation

Explanation/Reference:
Explanation:

Box 1: HDInsight Tools for Visual Studio
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.

Box 2: Hive View
You can use Apache Ambari Hive View with Apache Hadoop in HDInsight. The Hive View allows you to author, optimize, and run Hive queries from your web browser.

Box 3: HDInsight REST API
Azure HDInsight REST APIs are used to create and manage HDInsight resources through Azure Resource Manager.

References:
https://visualstudiomagazine.com/articles/2019/01/25/vscode-hdinsight.aspx

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 23
DRAG DROP

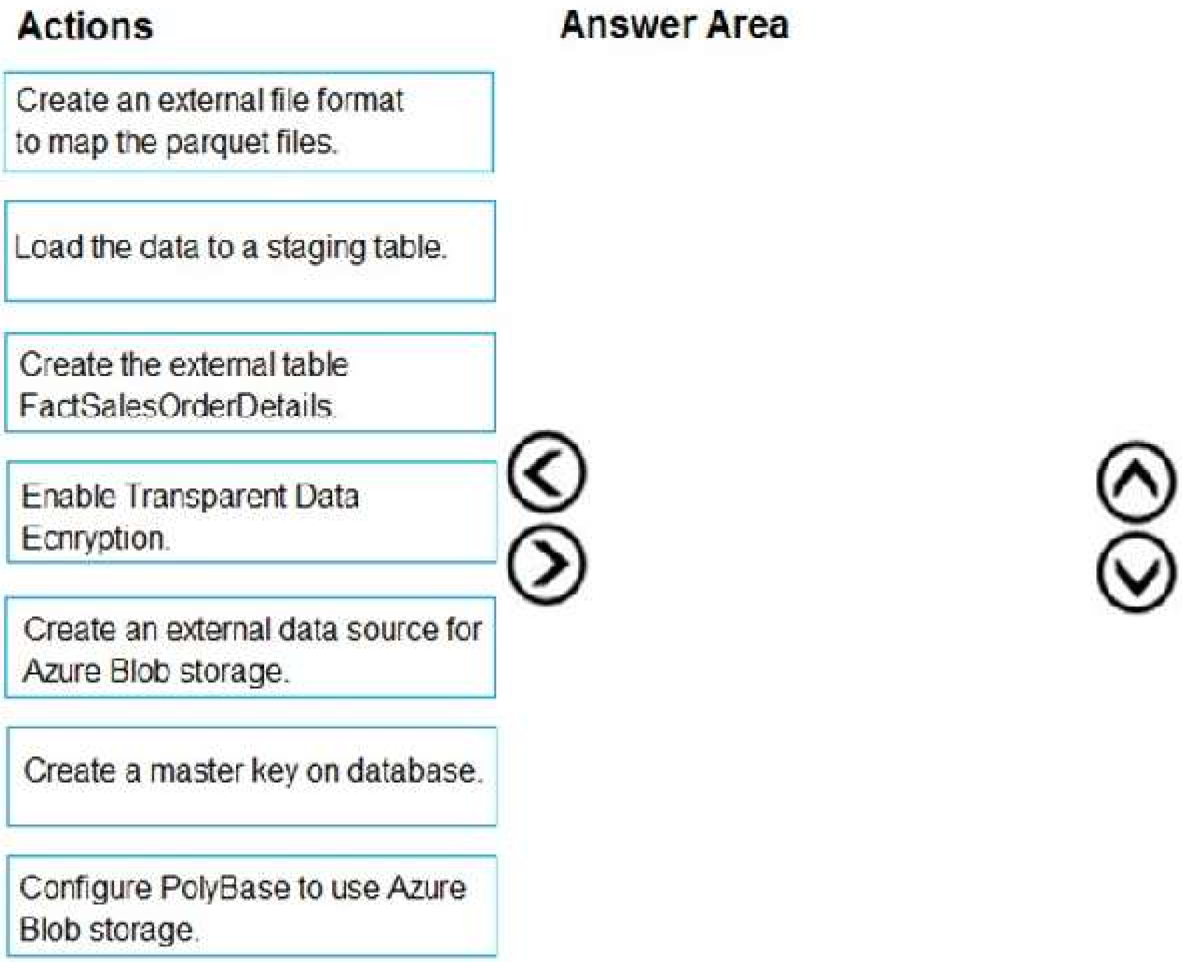
You are creating a managed data warehouse solution on Microsoft Azure.

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

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

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

Select and Place:



Actions

Create an external file format to map the parquet files.

Load the data to a staging table.

Create the external table FactSalesOrderDetails.

Enable Transparent Data Ecnryption.

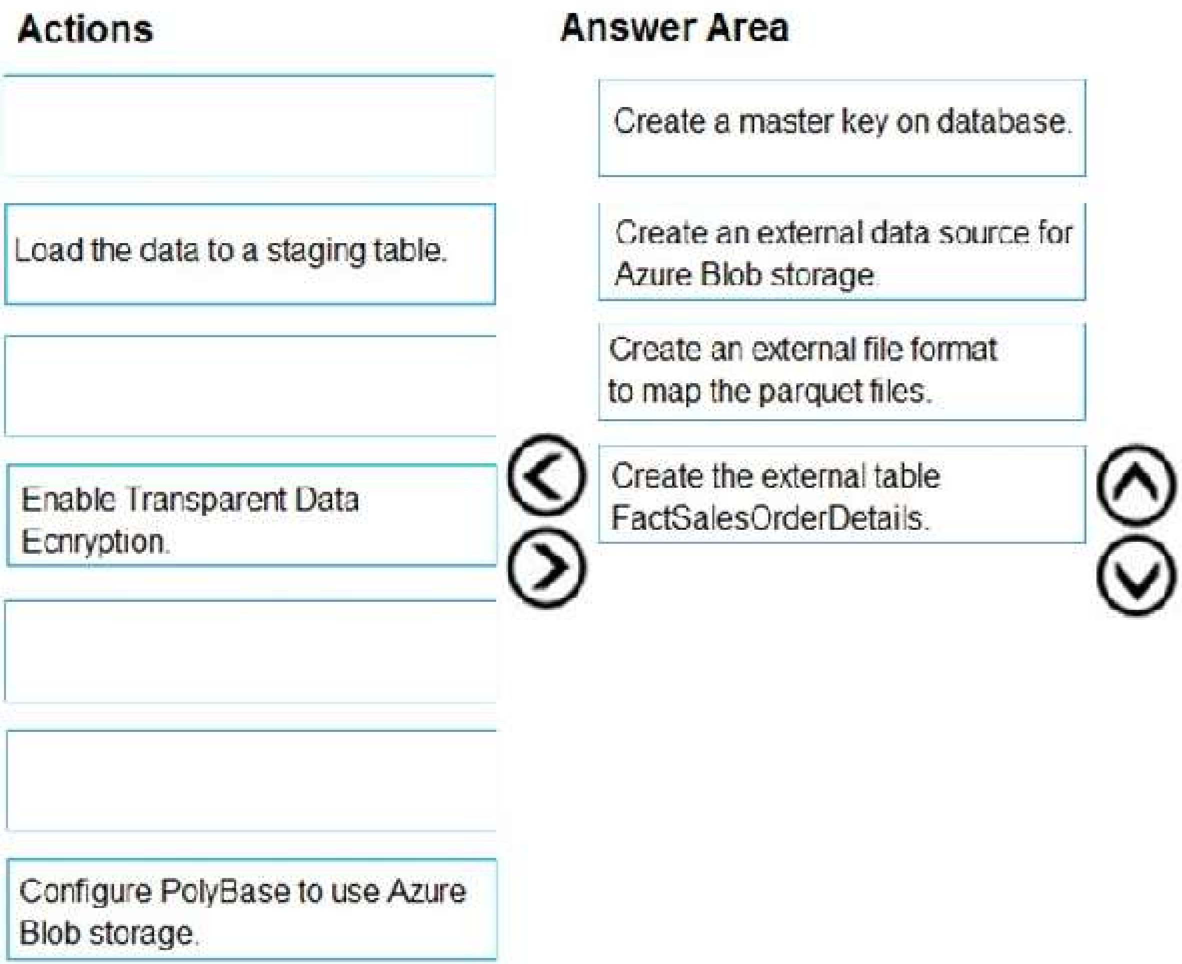
Create an external data source for Azure Blob storage.

Create a master key on database.

Configure PolyBase to use Azure Blob storage.

Answer Area

Correct Answer:



Actions

Load the data to a staging table.

Enable Transparent Data Ecnryption.

Configure PolyBase to use Azure Blob storage.

Answer Area

Create a master key on database.

Create an external data source for Azure Blob storage.

Create an external file format to map the parquet files.

Create the external table FactSalesOrderDetails.

Section: [none]
Explanation

Explanation/Reference:
Explanation:

Step 1: Create a master key on the database
Create a master key on the database. This is required to encrypt the credential secret.

Step 2: Create an external data source for Azure Blob storage
Create an external data source with CREATE EXTERNAL DATA SOURCE..

Step 3: Create an external file format to map parquet files.
Create an external file format with CREATE EXTERNAL FILE FORMAT.
FORMAT TYPE: Type of format in Hadoop (DELIMITEDTEXT, RCFILE, ORC, PARQUET).

Step 4: Create the external table FactSalesOrderDetails
To query the data in your Hadoop data source, you must define an external table to use in Transact-SQL queries.
Create an external table pointing to data stored in Azure storage with CREATE EXTERNAL TABLE.

Note: PolyBase is a technology that accesses and combines both non-relational and relational data, all from within SQL Server. It allows you to run queries on external data in Hadoop or Azure blob storage.

References:

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

QUESTION 24
DRAG DROP

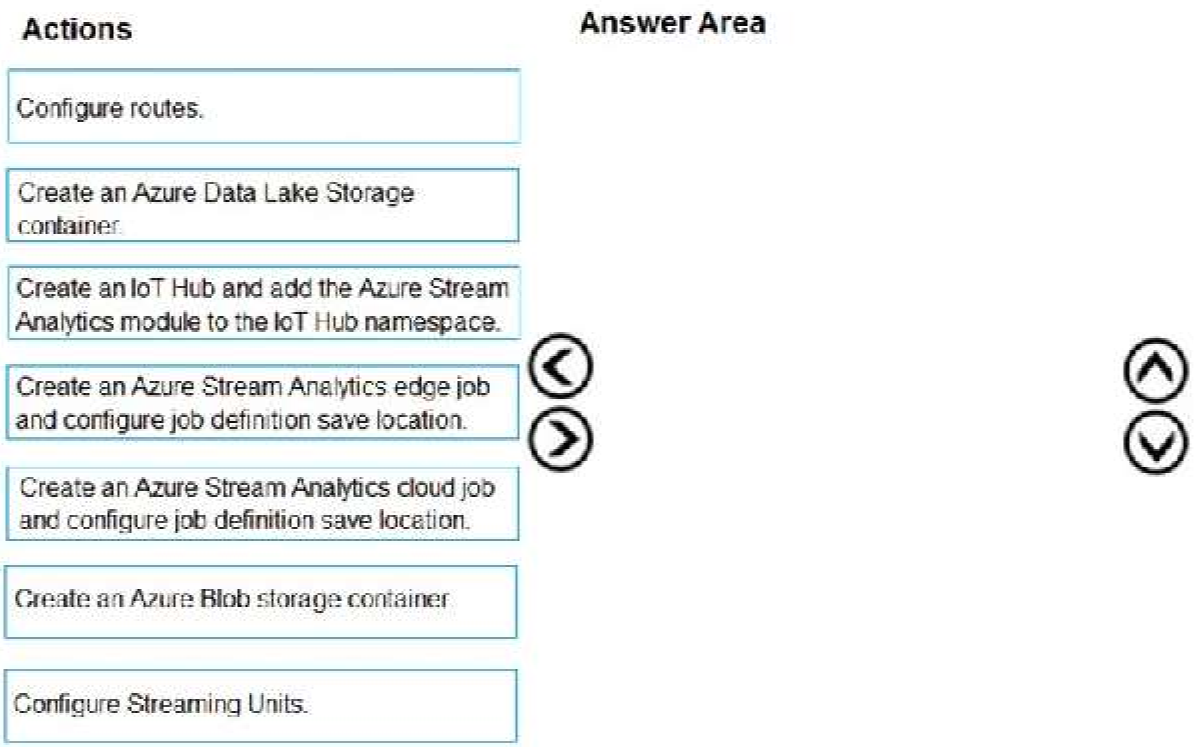
You develop data engineering solutions for a company.

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? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

Select and Place:



Actions

Configure routes.

Create an Azure Data Lake Storage container.

Create an IoT Hub and add the Azure Stream Analytics module to the IoT Hub namespace.

Create an Azure Stream Analytics edge job and configure job definition save location.

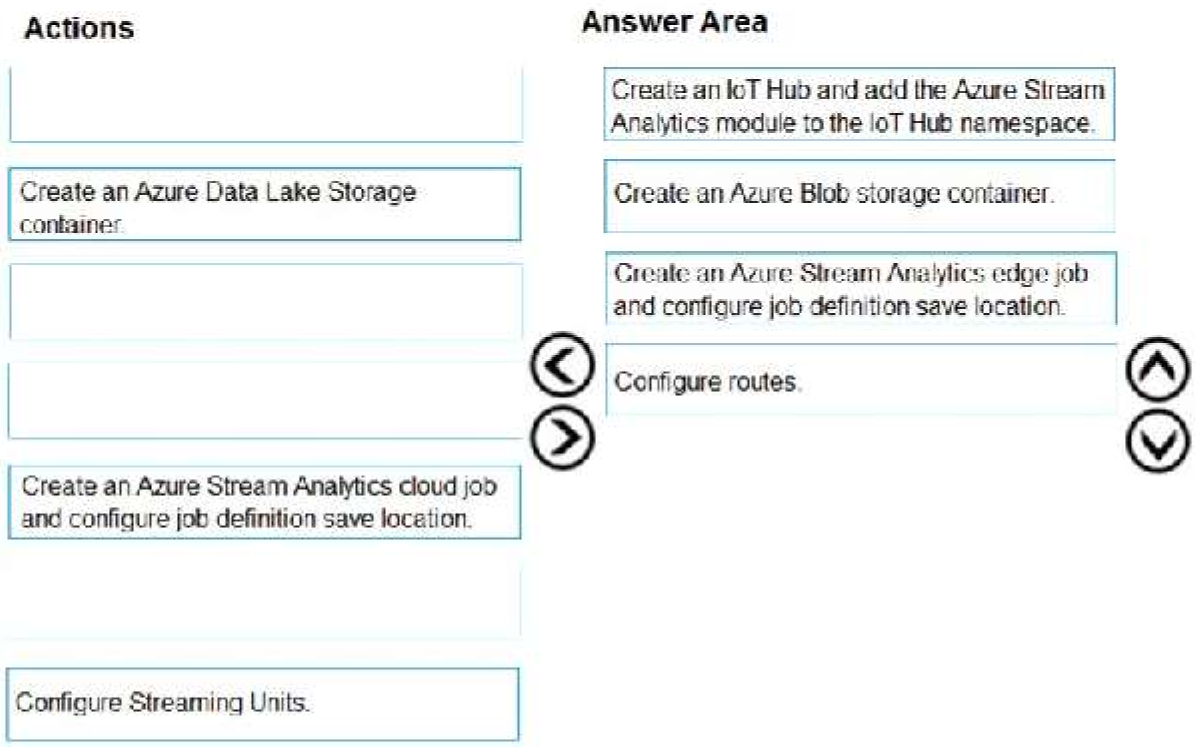
Create an Azure Stream Analytics cloud job and configure job definition save location.

Create an Azure Blob storage container

Configure Streaming Units.

Answer Area

Correct Answer:



Actions

Create an Azure Data Lake Storage container.

Create an Azure Stream Analytics cloud job and configure job definition save location.

Configure Streaming Units.

Answer Area

Create an IoT Hub and add the Azure Stream Analytics module to the IoT Hub namespace.

Create an Azure Blob storage container.

Create an Azure Stream Analytics edge job and configure job definition save location.

Configure routes.

Section: [none]
Explanation

Explanation/Reference:
Explanation:

Step 1: Create and IoT hub and add the Azure Stream Analytics module to the IoT Hub namespace
An IoT Hub in Azure is required.

Step 2: Create an Azure Blob Storage container
To prepare your Stream Analytics job to be deployed on an IoT Edge device, you need to associate the job with a container in a storage account. When you go to deploy your job, the job definition is exported to the storage container.

Stream Analytics accepts data incoming from several kinds of event sources including Event Hubs, IoT Hub, and Blob storage.

Step 3: Create an Azure Stream Analytics edge job and configure job definition save location
When you create an Azure Stream Analytics job to run on an IoT Edge device, it needs to be stored in a way that can be called from the device.

Step 4: Configure routes
You are now ready to deploy the Azure Stream Analytics job on your IoT Edge device.
The routes that you declare define the flow of data through the IoT Edge device.
Example:

References:
https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-add-inputs

https://docs.microsoft.com/en-us/azure/iot-edge/tutorial-deploy-stream-analytics

QUESTION 25
DRAG DROP

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).

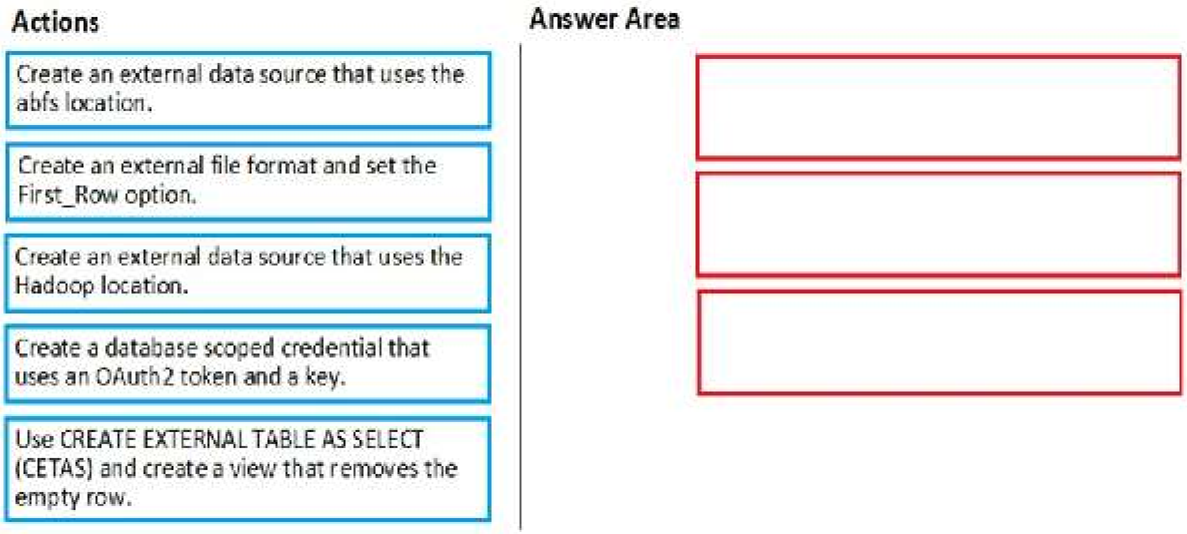
You are implementing a pattern that batch loads the files daily into an Azure SQL data warehouse by using PolyBase.

You need to skip the header row when you import the files into the data warehouse.

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

Which three actions 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.

Select and Place:

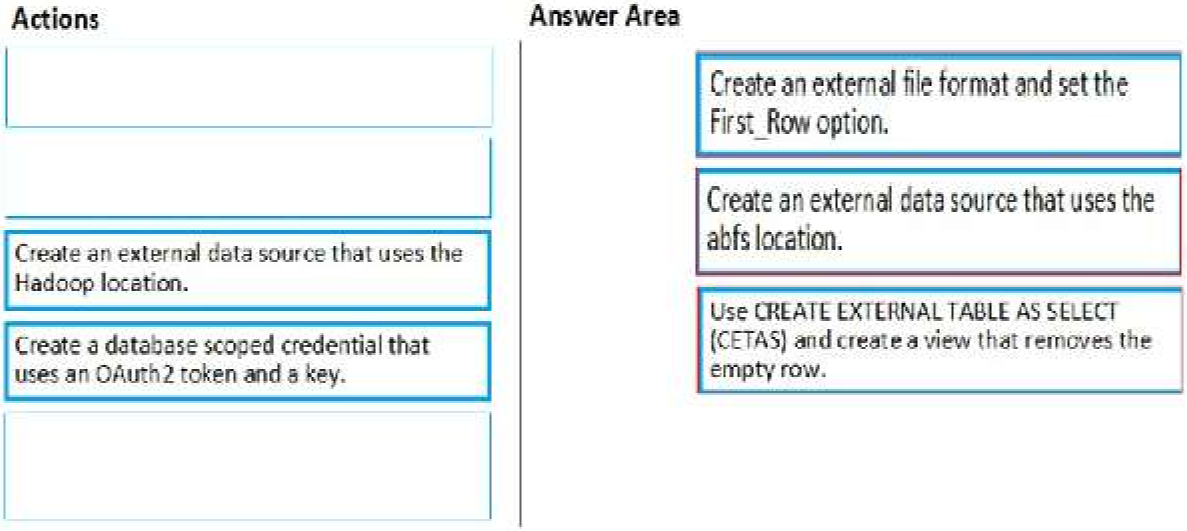


Actions

- Create an external data source that uses the abfs location.
- Create an external file format and set the First_Row option.
- Create an external data source that uses the Hadoop location.
- Create a database scoped credential that uses an OAuth2 token and a key.
- Use CREATE EXTERNAL TABLE AS SELECT (CETAS) and create a view that removes the empty row.

Answer Area

Correct Answer:



Actions

- Create an external data source that uses the Hadoop location.
- Create a database scoped credential that uses an OAuth2 token and a key.

Answer Area

1. Create an external file format and set the First_Row option.
2. Create an external data source that uses the abfs location.
3. Use CREATE EXTERNAL TABLE AS SELECT (CETAS) and create a view that removes the empty row.

Section: [none]
Explanation

Explanation/Reference:
Step 1: Create an external data source and set the First_Row option.
Creates an External File Format object defining external data stored in Hadoop, Azure Blob Storage, or Azure Data Lake Store. Creating an external file format is a prerequisite for creating an External Table.

FIRST_ROW = First_row_int
Specifies the row number that is read first in all files during a PolyBase load. This parameter can take values 1-15. If the value is set to two, the first row in every file (header row) is skipped when the data is loaded. Rows are skipped based on the existence of row terminators (/r/n, /r, /n).

Step 2: Create an external data source that uses the abfs location
The hadoop-azure module provides support for the Azure Data Lake Storage Gen2 storage layer through the "abfs" connector

Step 3: Use CREATE EXTERNAL TABLE AS SELECT (CETAS) and create a view that removes the empty row.

References:
https://docs.microsoft.com/en-us/sql/t-sql/statements/create-external-file-format-transact-sql

https://hadoop.apache.org/docs/r3.2.0/hadoop-azure/abfs.html

QUESTION 26
You are creating a new notebook in Azure Databricks that will support R as the primary language but will also support Scola and SQL.

Which switch should you use to switch between languages?

A. `%<language>`
B. `\\[<language>]`
C. `\\(<language>)`
D. `@<Language>`

Correct Answer: A
Section: [none]
Explanation

Explanation/Reference:
Explanation:
You can override the primary language by specifying the language magic command %<language> at the beginning of a cell. The supported magic commands are: %python, %r, %scala, and %sql.

References:
https://docs.databricks.com/user-guide/notebooks/notebook-use.html#mix-languages

QUESTION 27
You use Azure Stream Analytics to receive Twitter data from Azure Event Hubs and to output the data to an Azure Blob storage account.

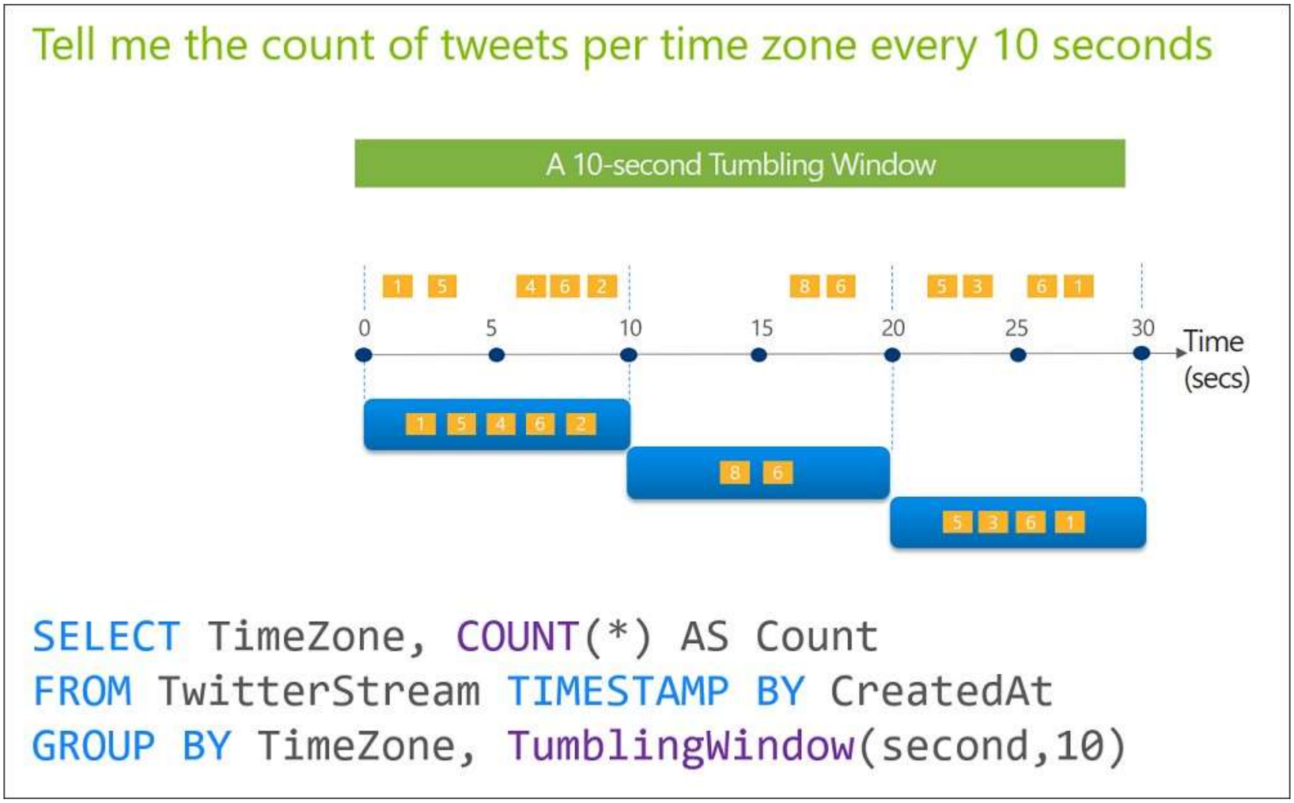
You need to output the count of tweets during the last five minutes every five minutes. Each tweet must only be counted once.

Which windowing function should you use?

A. a five-minute Session window
B. a five-minute Sliding window
C. a five-minute Tumbling window
D. a five-minute Hopping window that has one-minute hop

Correct Answer: C
Section: [none]
Explanation

Explanation/Reference:
Explanation:
Tumbling window functions are used to segment a data stream into distinct time segments and perform a function against them, such as the example below. The key differentiators of a Tumbling window are that they repeat, do not overlap, and an event cannot belong to more than one tumbling window.



Tell me the count of tweets per time zone every 10 seconds

```
SELECT TimeZone, COUNT(*) AS Count
FROM TwitterStream TIMESTAMP BY CreatedAt
GROUP BY TimeZone, TumblingWindow(second,10)
```

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

QUESTION 28
You are developing a solution that will stream to Azure Stream Analytics. The solution will have both streaming data and reference data.

Which input type should you use for the reference data?

A. Azure Cosmos DB
B. Azure Event Hubs
C. Azure Blob storage
D. Azure IoT Hub

Correct Answer: C
Section: [none]
Explanation

Explanation/Reference:
Explanation:

Stream Analytics supports Azure Blob storage and Azure SQL Database as the storage layer for Reference Data.

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

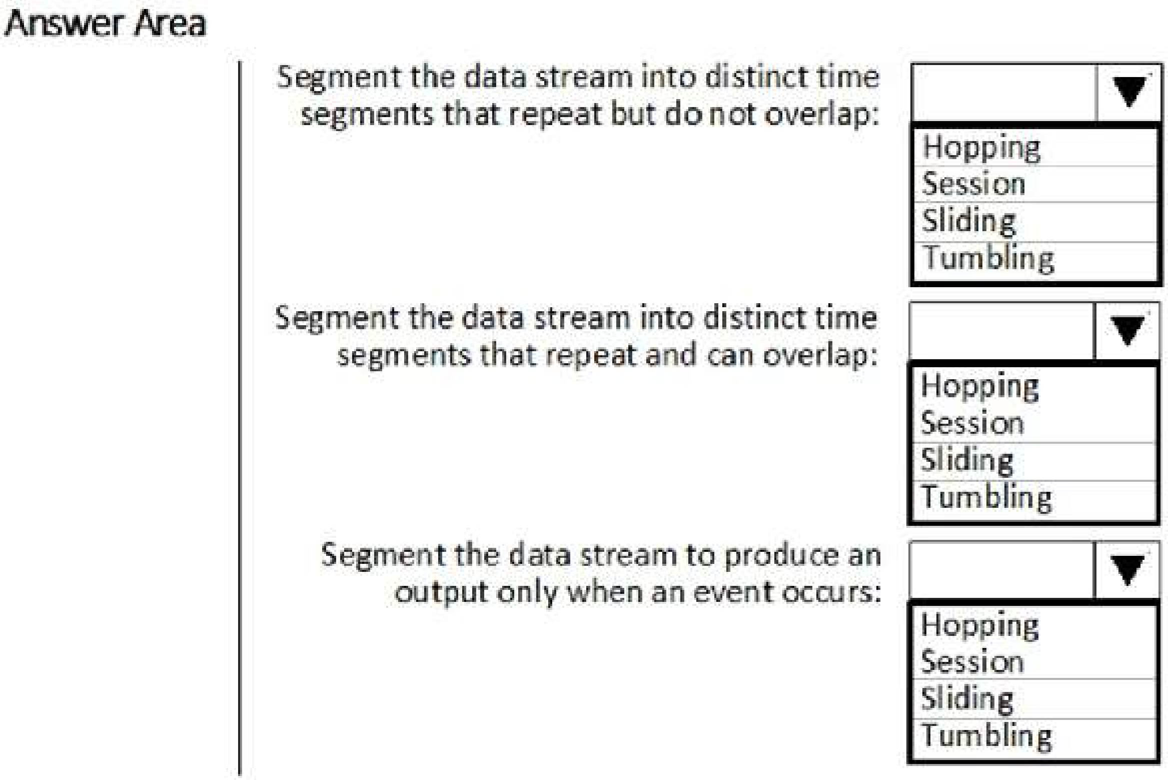
QUESTION 29
HOTSPOT

You are implementing Azure Stream Analytics functions.

Which windowing function should you use for each requirement? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:



Answer Area

Segment the data stream into distinct time segments that repeat but do not overlap:

- Hopping
- Session
- Sliding
- Tumbling

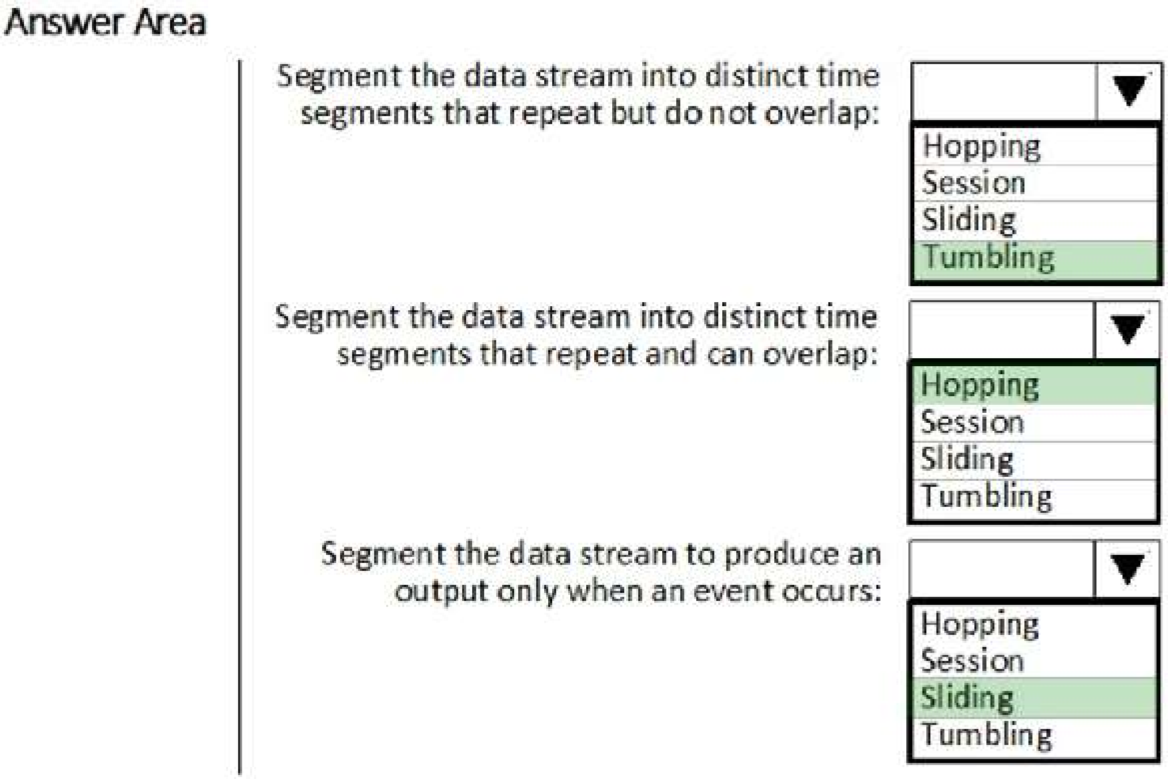
Segment the data stream into distinct time segments that repeat and can overlap:

- Hopping
- Session
- Sliding
- Tumbling

Segment the data stream to produce an output only when an event occurs:

- Hopping
- Session
- Sliding
- Tumbling

Correct Answer:



Answer Area

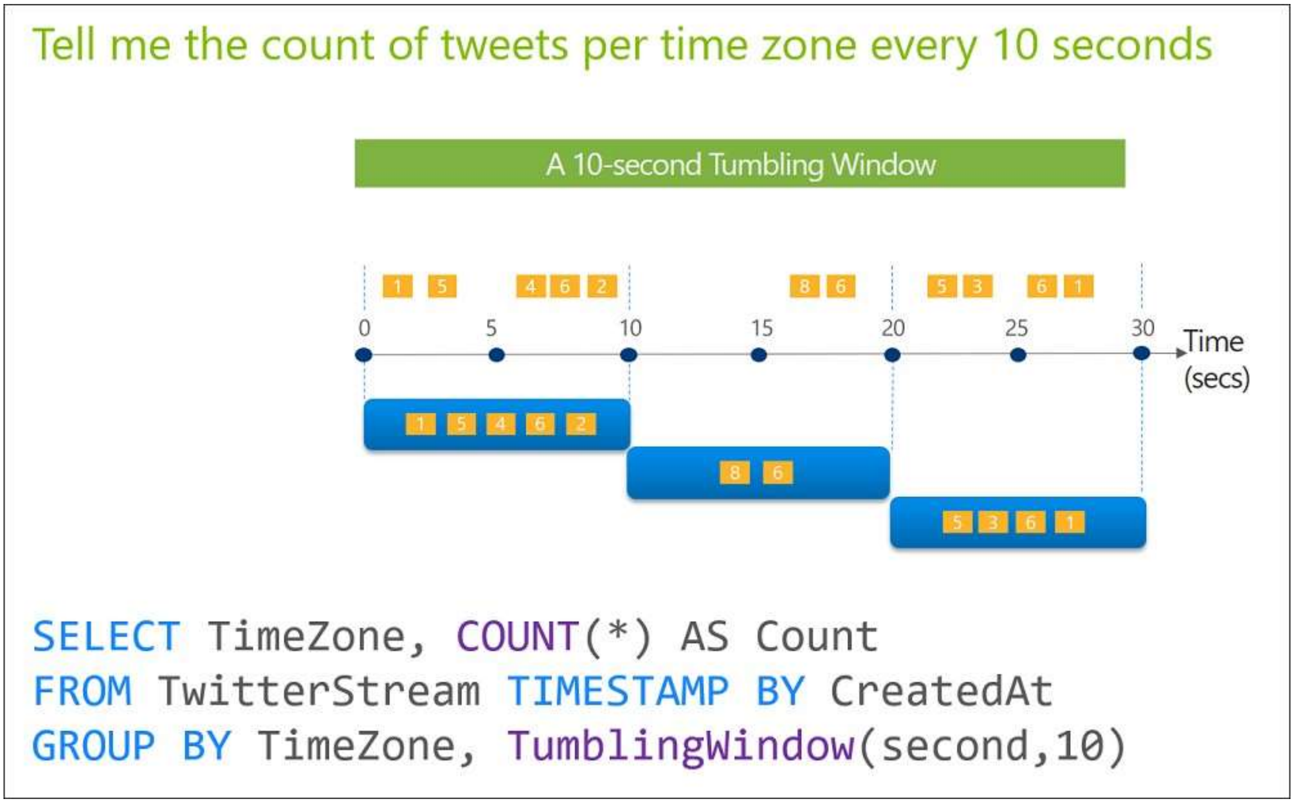
Segment the data stream into distinct time segments that repeat but do not overlap:

Segment the data stream into distinct time segments that repeat and can overlap:

Segment the data stream to produce an output only when an event occurs:

Section: [none]
Explanation

Explanation/Reference:
Box 1: Tumbling
Tumbling window functions are used to segment a data stream into distinct time segments and perform a function against them, such as the example below. The key differentiators of a Tumbling window are that they repeat, do not overlap, and an event cannot belong to more than one tumbling window.

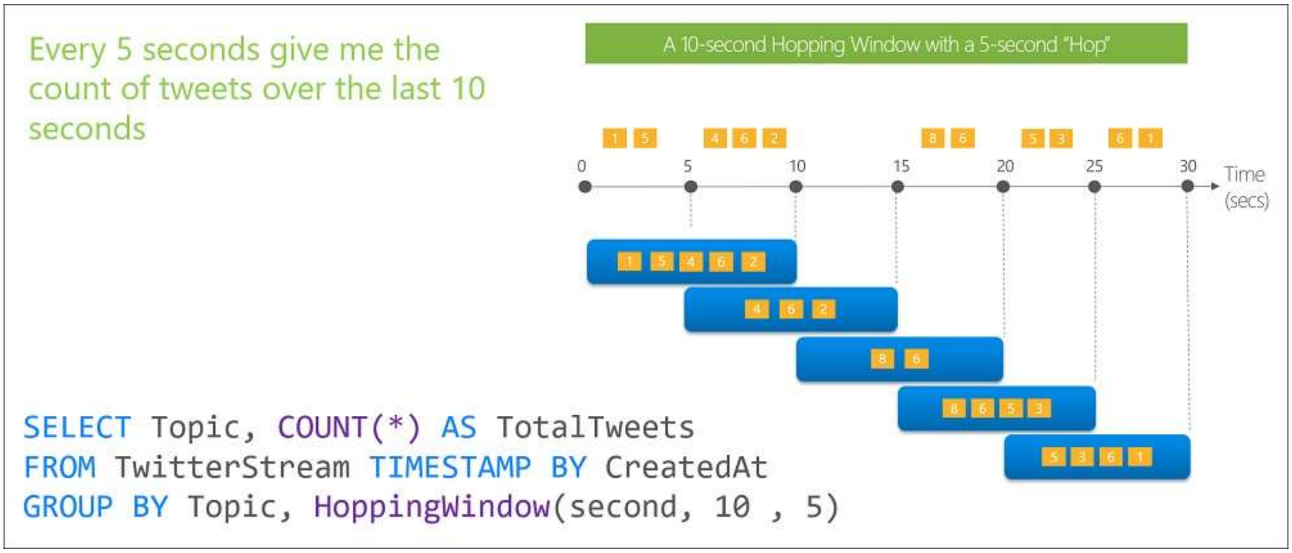


Tell me the count of tweets per time zone every 10 seconds

A 10-second Tumbling Window

```
SELECT TimeZone, COUNT(*) AS Count
FROM TwitterStream TIMESTAMP BY CreatedAt
GROUP BY TimeZone, TumblingWindow(second,10)
```

Box 2: Hoppping
Hopping window functions hop forward in time by a fixed period. It may be easy to think of them as Tumbling windows that can overlap, so events can belong to more than one Hopping window result set. To make a Hopping window the same as a Tumbling window, specify the hop size to be the same as the window size.

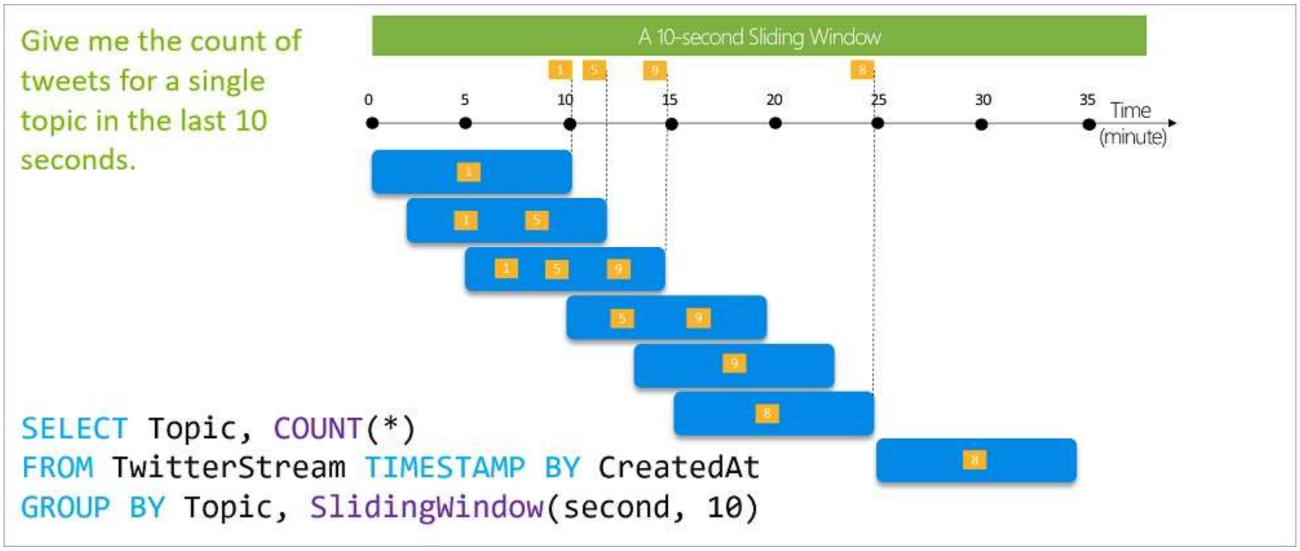


Every 5 seconds give me the count of tweets over the last 10 seconds

A 10-second Hopping Window with a 5-second "Hop"

```
SELECT Topic, COUNT(*) AS TotalTweets
FROM TwitterStream TIMESTAMP BY CreatedAt
GROUP BY Topic, HoppingWindow(second, 10 , 5)
```

Box 3: Sliding
Sliding window functions, unlike Tumbling or Hopping windows, produce an output only when an event occurs. Every window will have at least one event and the window continuously moves forward by an € (epsilon). Like hopping windows, events can belong to more than one sliding window.



Give me the count of tweets for a single topic in the last 10 seconds.

```
SELECT Topic, COUNT(*)
FROM TwitterStream TIMESTAMP BY CreatedAt
GROUP BY Topic, SlidingWindow(second, 10)
```

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

QUESTION 30
DRAG DROP

You have an Azure Data Lake Storage Gen2 account that contains JSON files for customers. The files contain two attributes named `FirstName` and `LastName`.

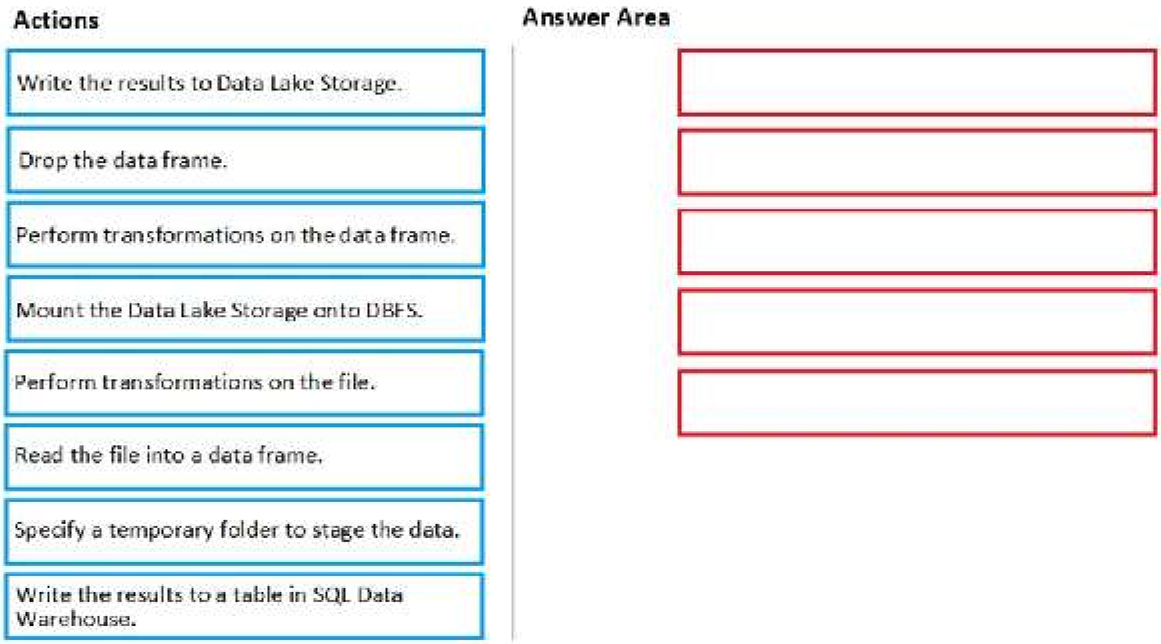
You need to copy the data from the JSON files to an Azure SQL data Warehouse table by using Azure Databricks. A new column must be created that concatenates the `FirstName` and `LastName` values.

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? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

Select and Place:

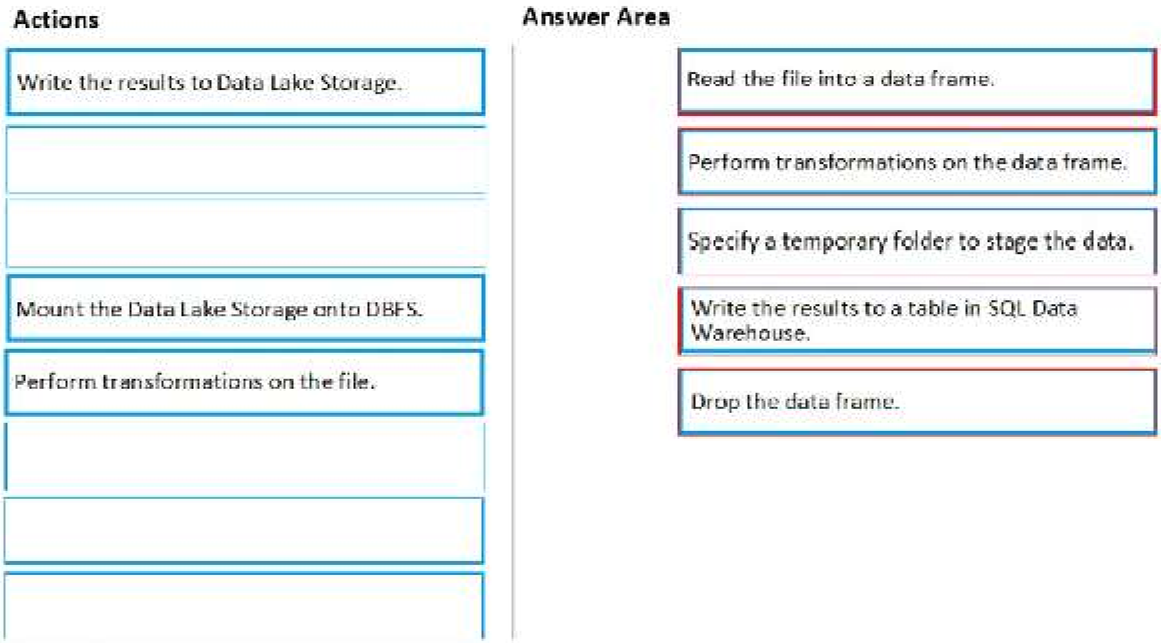


Actions

- Write the results to Data Lake Storage.
- Drop the data frame.
- Perform transformations on the data frame.
- Mount the Data Lake Storage onto DBFS.
- Perform transformations on the file.
- Read the file into a data frame.
- Specify a temporary folder to stage the data.
- Write the results to a table in SQL Data Warehouse.

Answer Area

Correct Answer:



Actions

- Write the results to Data Lake Storage.
- Mount the Data Lake Storage onto DBFS.
- Perform transformations on the file.

Answer Area

- Read the file into a data frame.
- Perform transformations on the data frame.
- Specify a temporary folder to stage the data.
- Write the results to a table in SQL Data Warehouse.
- Drop the data frame.

Section: [none]
Explanation

Explanation/Reference:
Step 1: Read the file into a data frame.
You can load the json files as a data frame in Azure Databricks.

Step 2: Perform transformations on the data frame.

Step 3:Specify a temporary folder to stage the data
Specify a temporary folder to use while moving data between Azure Databricks and Azure SQL Data

Warehouse.

Step 4: Write the results to a table in SQL Data Warehouse
You upload the transformed data frame into Azure SQL Data Warehouse. You use the Azure SQL Data Warehouse connector for Azure Databricks to directly upload a dataframe as a table in a SQL data warehouse.

Step 5: Drop the data frame
Clean up resources. You can terminate the cluster. From the Azure Databricks workspace, select Clusters on the left. For the cluster to terminate, under Actions, point to the ellipsis (...) and select the Terminate icon.

References:
https://docs.microsoft.com/en-us/azure/azure-databricks/databricks-extract-load-sql-data-warehouse

QUESTION 31
You have an Azure Storage account and an Azure SQL data warehouse by using Azure Data Factory. The solution must meet the following requirements:

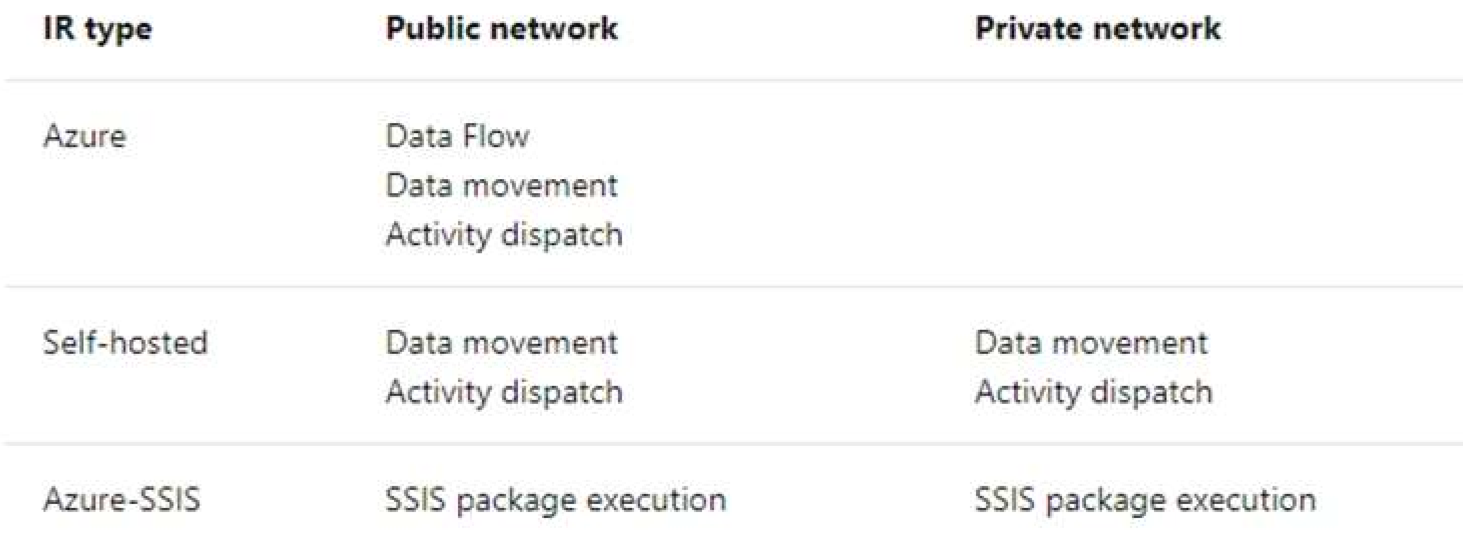
- Ensure that the data remains in the UK South region at all times.
- Minimize administrative effort.

Which type of integration runtime should you use?

A. Azure integration runtime
B. Self-hosted integration runtime
C. Azure-SSIS integration runtime

Correct Answer: A
Section: [none]
Explanation

Explanation/Reference:
Explanation:



IR type	Public network	Private network
Azure	Data Flow Data movement Activity dispatch	
Self-hosted	Data movement Activity dispatch	Data movement Activity dispatch
Azure-SSIS	SSIS package execution	SSIS package execution

Incorrect Answers:
B: Self-hosted integration runtime is to be used On-premises.

References:
https://docs.microsoft.com/en-us/azure/data-factory/concepts-integration-runtime

QUESTION 32
You plan to perform batch processing in Azure Databricks once daily.

Which type of Databricks cluster should you use?

A. job
B. interactive
C. High Concurrency

Correct Answer: A
Section: [none]
Explanation

Explanation/Reference:
Explanation:
Example: Scheduled batch workloads (data engineers running ETL jobs)
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.

Note: Azure Databricks has two types of clusters: interactive and automated. You use interactive clusters to analyze data collaboratively with interactive notebooks. You use automated clusters to run fast and robust automated jobs.

References:
https://docs.databricks.com/administration-guide/cloud-configurations/aws/cmbp.html#scenario-3-scheduled-batch-workloads-data-engineers-running-etl-jobs

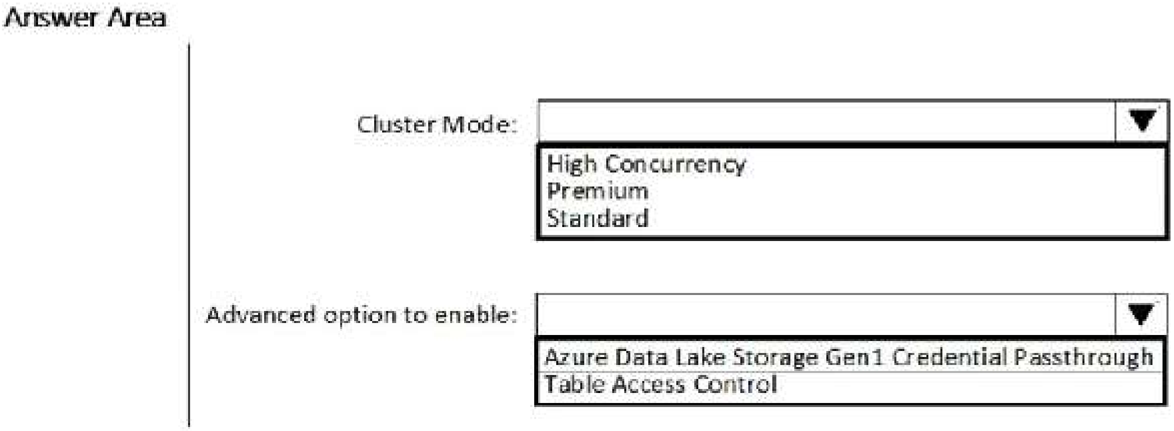
QUESTION 33
HOTSPOT

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.

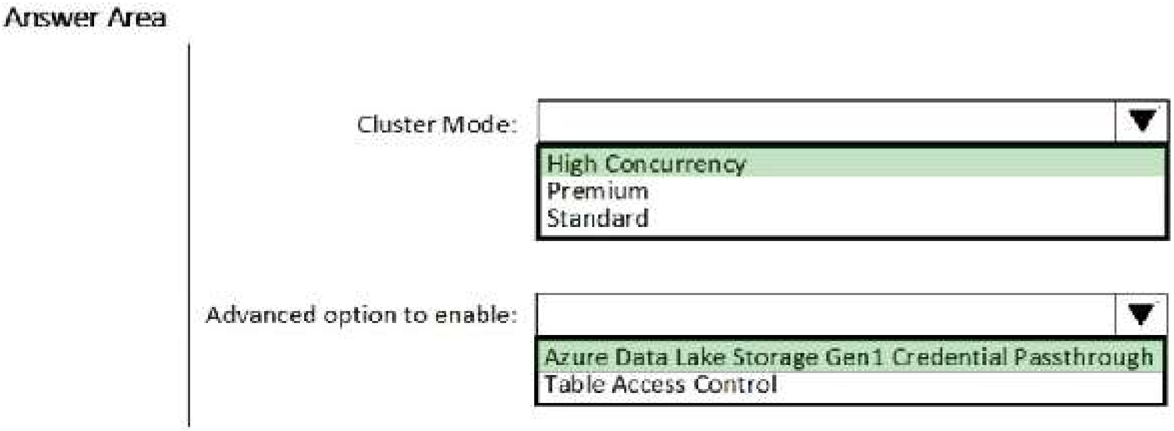
How should you configure the new cluster? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:


Answer Area
Cluster Mode:
High Concurrency
Premium
Standard
Advanced option to enable:
Azure Data Lake Storage Gen1 Credential Passthrough
Table Access Control

Correct Answer:



Answer Area

Cluster Mode:

- High Concurrency
- Premium
- Standard

Advanced option to enable:

- Azure Data Lake Storage Gen1 Credential Passthrough
- Table Access Control

Section: [none]
Explanation

Explanation/Reference:
Box 1: High Concurrency
Enable Azure Data Lake Storage credential passthrough for a high-concurrency cluster.

Incorrect:
Support for Azure Data Lake Storage credential passthrough on standard clusters is in Public Preview.

Standard clusters with credential passthrough are supported on Databricks Runtime 5.5 and above and are limited to a single user.

Box 2: Azure Data Lake Storage Gen1 Credential Passthrough
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.

References:
https://docs.azuredatabricks.net/spark/latest/data-sources/azure/adls-passthrough.html

Testlet 2

Background
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.

QUESTION 1
You need to ensure that phone-based poling data can be analyzed in the PollingData database.

How should you configure Azure Data Factory?

A. Use a tumbling schedule trigger
B. Use an event-based trigger
C. Use a schedule trigger
D. Use manual execution

Correct Answer: C
Section: [none]
Explanation

Explanation/Reference:
Explanation:
When creating a schedule trigger, you specify a schedule (start date, recurrence, end date etc.) for the trigger, and associate with a Data Factory pipeline.

Scenario:
All data migration processes must use Azure Data Factory
All data migrations must run automatically during non-business hours

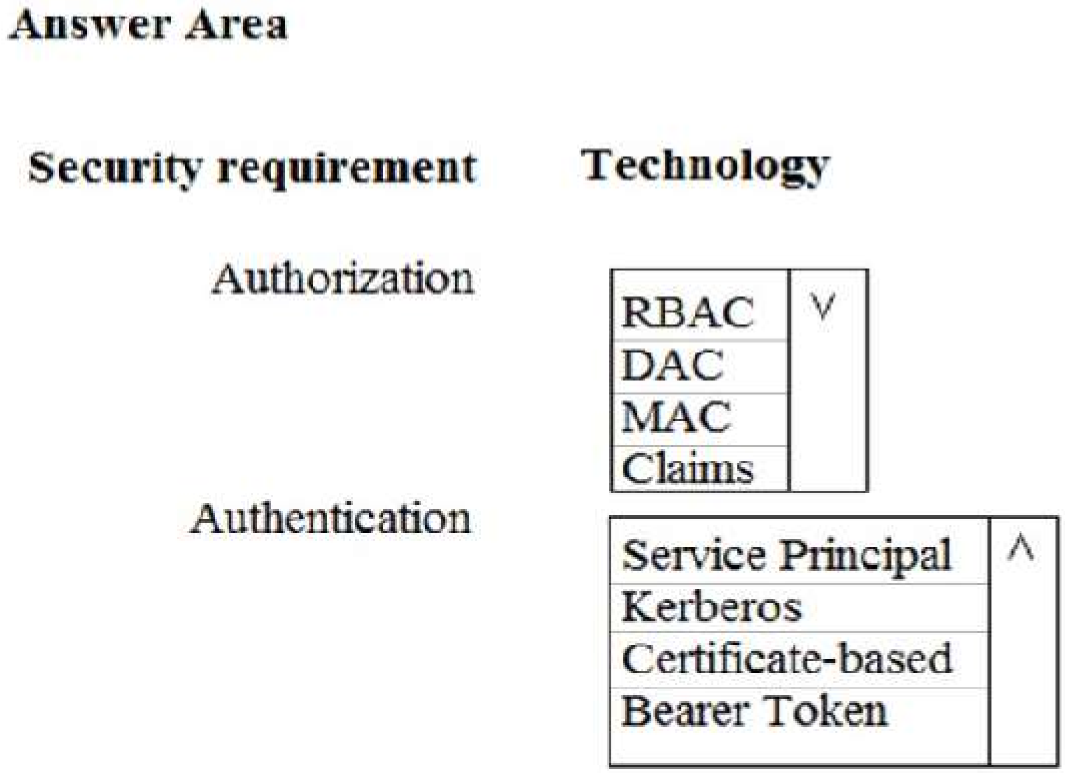
References:
https://docs.microsoft.com/en-us/azure/data-factory/how-to-create-schedule-trigger

QUESTION 2
HOTSPOT

You need to ensure that Azure Data Factory pipelines can be deployed. How should you configure authentication and authorization for deployments? To answer, select the appropriate options in the answer choices.

NOTE: Each correct selection is worth one point.

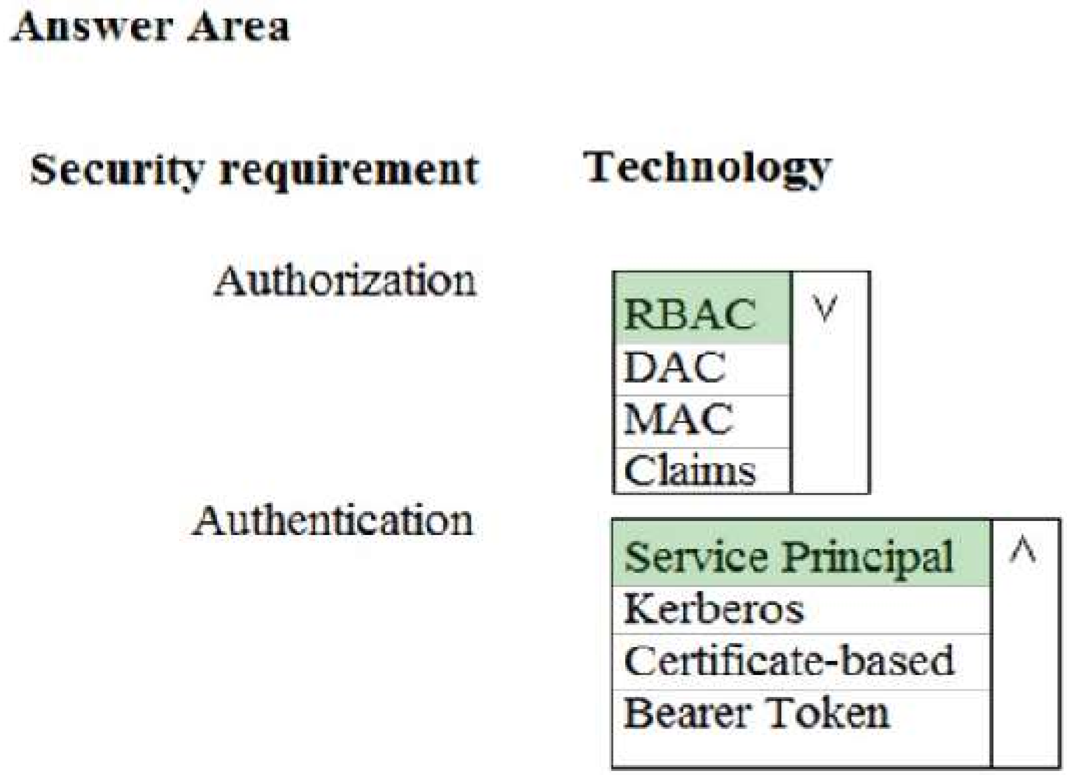
Hot Area:



Answer Area

Security requirement	Technology
Authorization	RBAC DAC MAC Claims
Authentication	Service Principal Kerberos Certificate-based Bearer Token

Correct Answer:



Answer Area

Security requirement	Technology
Authorization	RBAC
Authentication	Service Principal

Section: [none]
Explanation

Explanation/Reference:
Explanation:

The way you control access to resources using RBAC is to create role assignments. This is a key concept to understand – it’s how permissions are enforced. A role assignment consists of three elements: security principal, role definition, and scope.

Scenario:
No credentials or secrets should be used during deployments

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

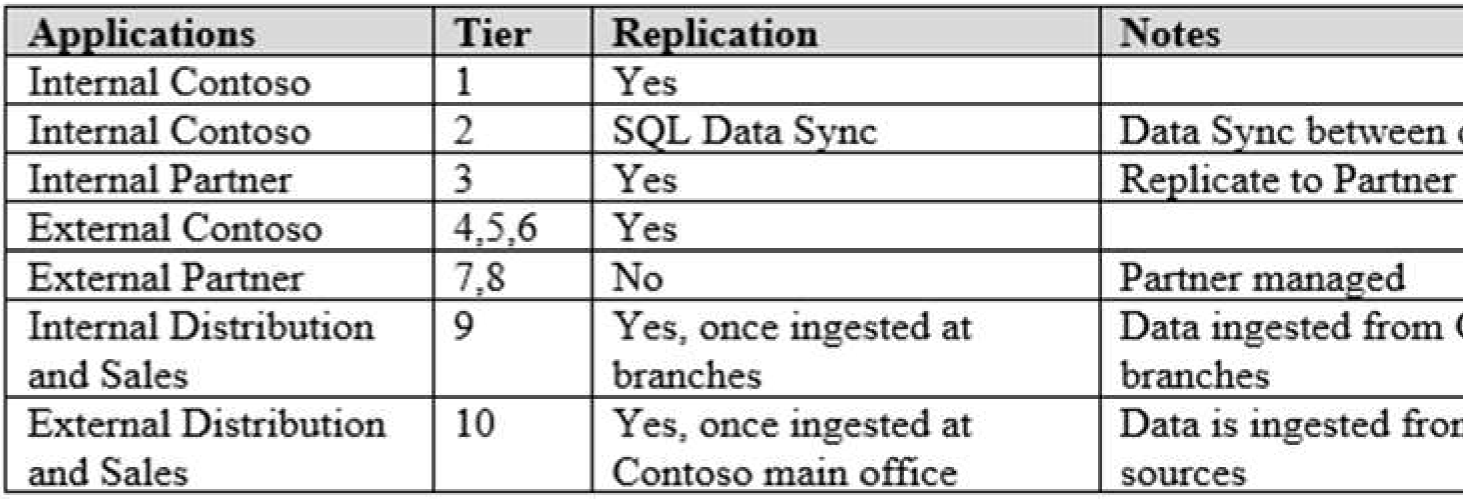
References:
https://docs.microsoft.com/en-us/azure/role-based-access-control/overview

Testlet 3

Overview

Current environment
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:



Applications	Tier	Replication	Notes
Internal Contoso	1	Yes	
Internal Contoso	2	SQL Data Sync	Data Sync between
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 branches
External Distribution and Sales	10	Yes, once ingested at Contoso main office	Data is ingested from 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.

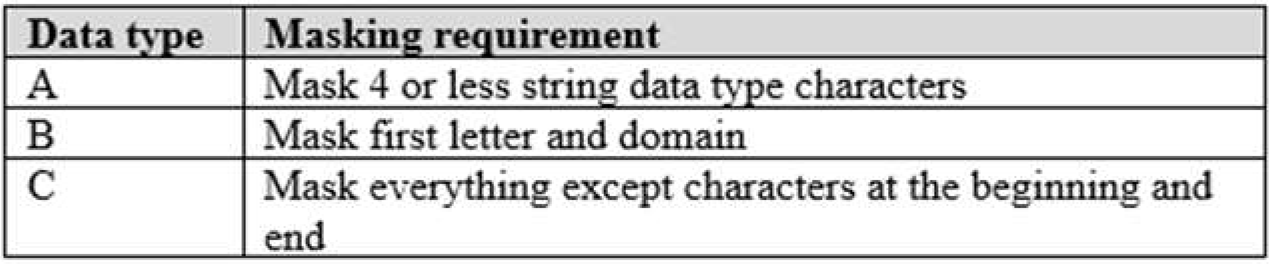
Back up strategies must be put in place for all other standalone Azure SQL Databases using Azure SQL-

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:



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

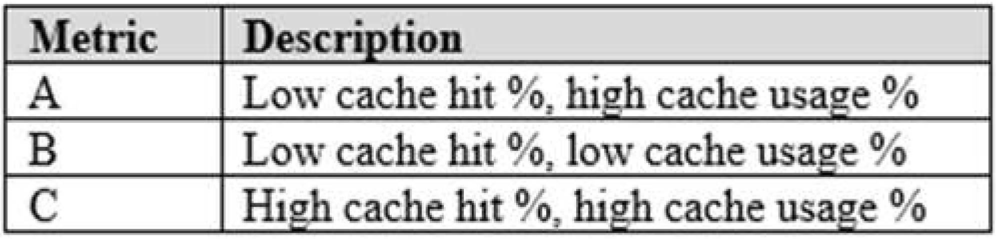
Security and monitoring

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:



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 1

You need to process and query ingested Tier 9 data.

Which two options should you use? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

A. Azure Notification Hub
B. Transact-SQL statements
C. Azure Cache for Redis
D. Apache Kafka statements
E. Azure Event Grid
F. Azure Stream Analytics

Correct Answer: EF
Section: [none]
Explanation

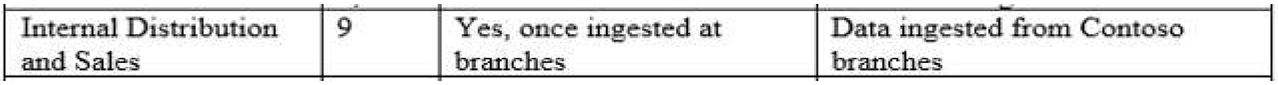
Explanation/Reference:
Explanation:

Event Hubs provides a Kafka endpoint that can be used by your existing Kafka based applications as an alternative to running your own Kafka cluster.

You can stream data into Kafka-enabled Event Hubs and process it with Azure Stream Analytics, in the following steps:

- Create a Kafka enabled Event Hubs namespace.
- Create a Kafka client that sends messages to the event hub.
- Create a Stream Analytics job that copies data from the event hub into an Azure blob storage.

Scenario:



Internal Distribution and Sales	9	Yes, once ingested at branches	Data ingested from Contoso branches

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

References:
https://docs.microsoft.com/en-us/azure/event-hubs/event-hubs-kafka-stream-analytics

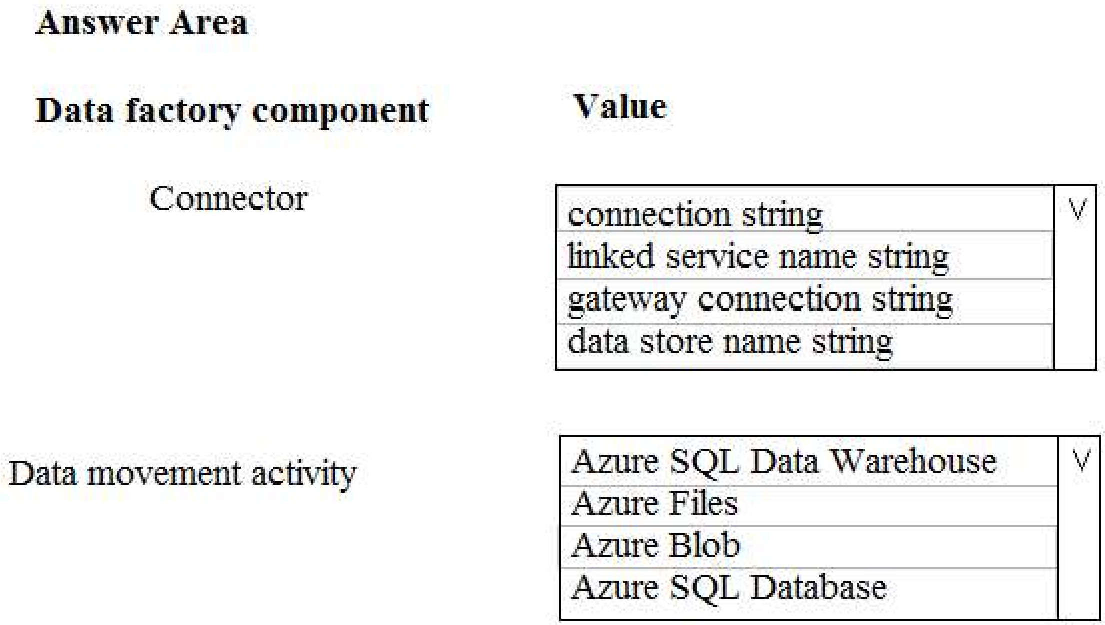
QUESTION 2
HOTSPOT

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

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

NOTE: Each correct selection is worth one point.

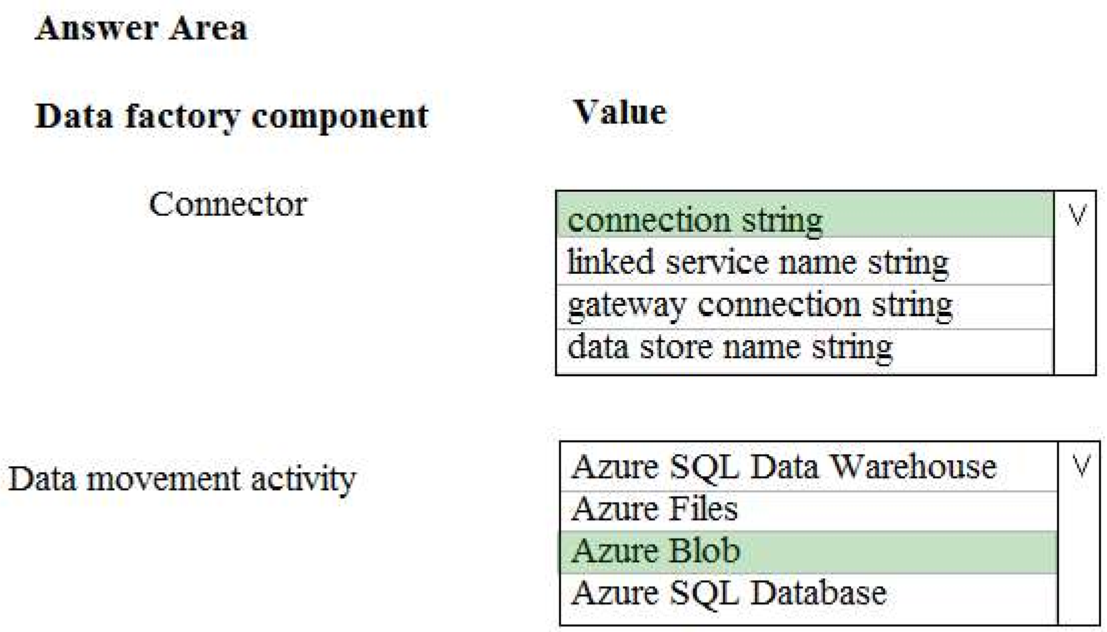
Hot Area:



Answer Area

Data factory component	Value
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

Correct Answer:



Answer Area

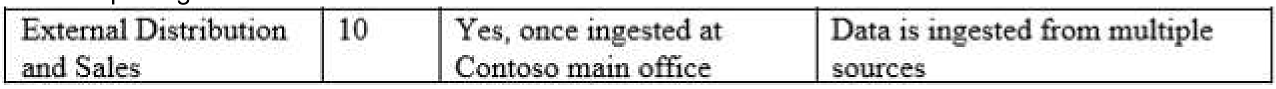
Data factory component	Value
Connector	[✓ connection string] linked service name string gateway connection string data store name string
Data movement activity	Azure SQL Data Warehouse Azure Files [✓ Azure Blob] Azure SQL Database

Section: [none]
Explanation

Explanation/Reference:
Explanation:

Box 1: Connection String
To use storage account key authentication, you use the ConnectionString property, which xpecify the information needed to connect to Blobl Storage.
Mark this field as a SecureString to store it securely in Data Factory. You can also put account key in Azure Key Vault and pull the accountKey configuration out of the connection string.

Box 2: Azure Blob
Tier 10 reporting data must be stored in Azure Blobs



External Distribution and Sales	10	Yes, once ingested at Contoso main office	Data is ingested from multiple sources

References:
https://docs.microsoft.com/en-us/azure/data-factory/connector-azure-blob-storage

QUESTION 3
You need to set up Azure Data Factory pipelines to meet data movement requirements.

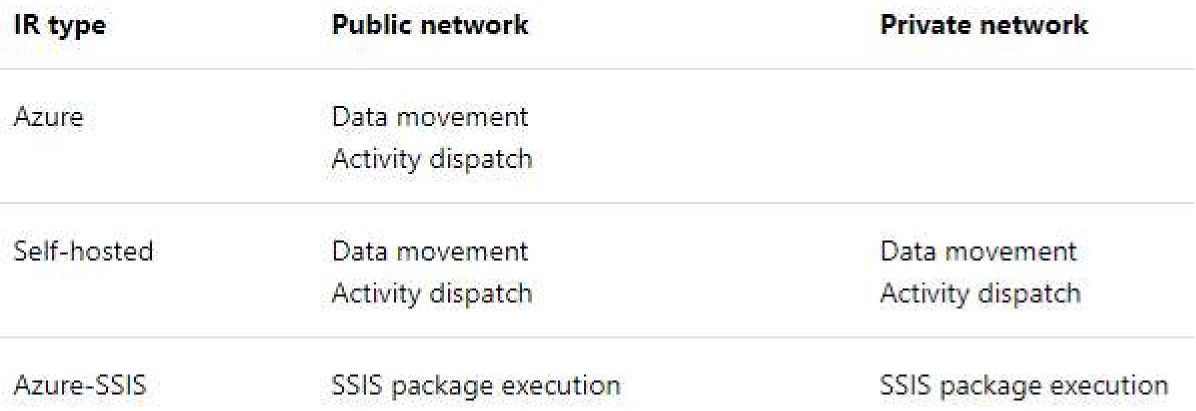
Which integration runtime should you use?

A. self-hosted integration runtime
B. Azure-SSIS Integration Runtime
C. .NET Common Language Runtime (CLR)
D. Azure integration runtime

Correct Answer: A
Section: [none]
Explanation

Explanation/Reference:

Explanation:
The following table describes the capabilities and network support for each of the integration runtime types:



IR type	Public network	Private network
Azure	Data movement Activity dispatch	
Self-hosted	Data movement Activity dispatch	Data movement Activity dispatch
Azure-SSIS	SSIS package execution	SSIS package execution

Scenario: 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.

References:
https://docs.microsoft.com/en-us/azure/data-factory/concepts-integration-runtime

Testlet 4

Case Study

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

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

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

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

Overview

General Overview

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 1
What should you include in the Data Factory pipeline for Race Central?

A. a copy activity that uses a stored procedure as a source

B. a copy activity that contains schema mappings

C. a delete activity that has logging enabled

D. a filter activity that has a condition

Correct Answer: B
Section: [none]
Explanation

Explanation/Reference:
Explanation:
Scenario:
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 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.

You can copy data to or from Azure Cosmos DB (SQL API) by using Azure Data Factory pipeline.
Column mapping applies when copying data from source to sink. By default, copy activity map source data to sink by column names. You can specify explicit mapping to customize the column mapping based on your need. More specifically, copy activity:

Read the data from source and determine the source schema

1. Use default column mapping to map columns by name, or apply explicit column mapping if specified.
2. Write the data to sink
3. Write the data to sink

References:
https://docs.microsoft.com/en-us/azure/data-factory/copy-activity-schema-and-type-mapping

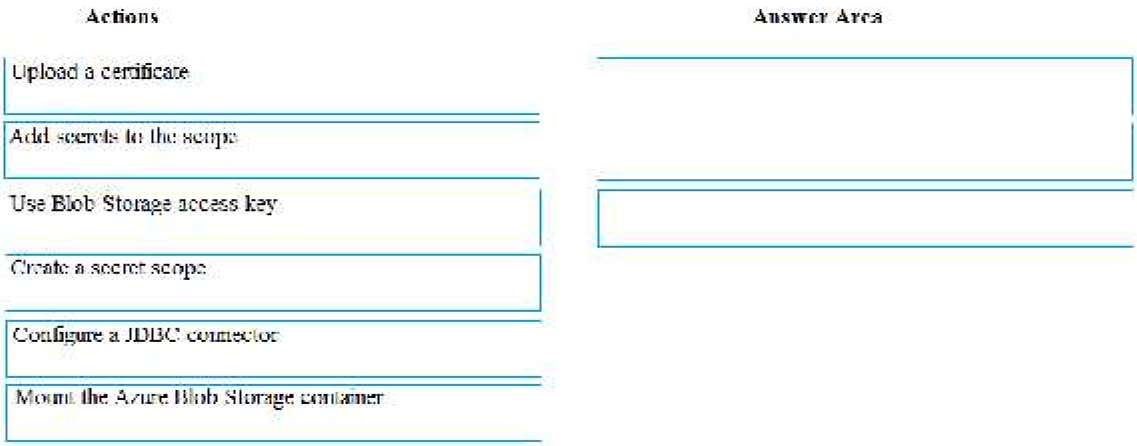
Question Set 1

QUESTION 1
DRAG DROP

You manage the Microsoft Azure Databricks environment for a company. You must be able to access a private Azure Blob Storage account. Data must be available to all Azure Databricks workspaces. You need to provide the data access.

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.

Select and Place:



Actions

Upload a certificate

Add secrets to the scope

Use Blob Storage access key

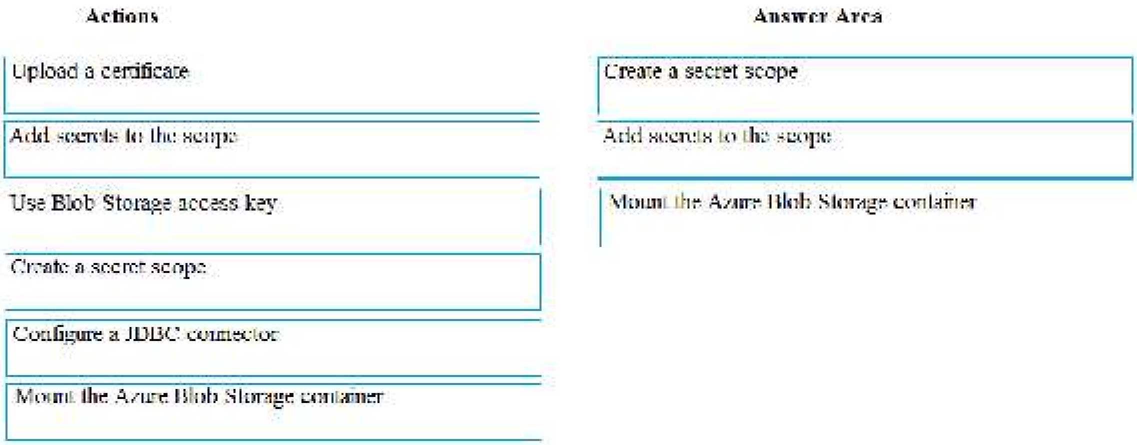
Create a secret scope

Configure a JDBC connector

Mount the Azure Blob Storage container

Answer Area

Correct Answer:



Actions

Upload a certificate

Add secrets to the scope

Use Blob Storage access key

Create a secret scope

Configure a JDBC connector

Mount the Azure Blob Storage container

Answer Area

Create a secret scope

Add secrets to the scope

Mount the Azure Blob Storage container

Section: [none]
Explanation

Explanation/Reference:
Explanation:

Step 1: Create a secret scope

Step 2: Add secrets to the scope
Note: dbutils.secrets.get(scope = "<scope-name>", key = "<key-name>") gets the key that has been stored as a secret in a secret scope.

Step 3: Mount the Azure Blob Storage container
You can mount a Blob Storage container or a folder inside a container through Databricks File System - DBFS. The mount is a pointer to a Blob Storage container, so the data is never synced locally.

Note: To mount a Blob Storage container or a folder inside a container, use the following command:

```
Python
dbutils.fs.mount(
 source = "wasbs://<your-container-name>@<your-storage-account-name>.blob.core.windows.net",
 mount_point = "/mnt/<mount-name>",
 extra_configs = {"<conf-key>":dbutils.secrets.get(scope = "<scope-name>", key = "<key-name>")})
```

where:
dbutils.secrets.get(scope = "<scope-name>", key = "<key-name>") gets the key that has been stored as a secret in a secret scope.

References:
https://docs.databricks.com/spark/latest/data-sources/azure/azure-storage.html

QUESTION 2
DRAG DROP

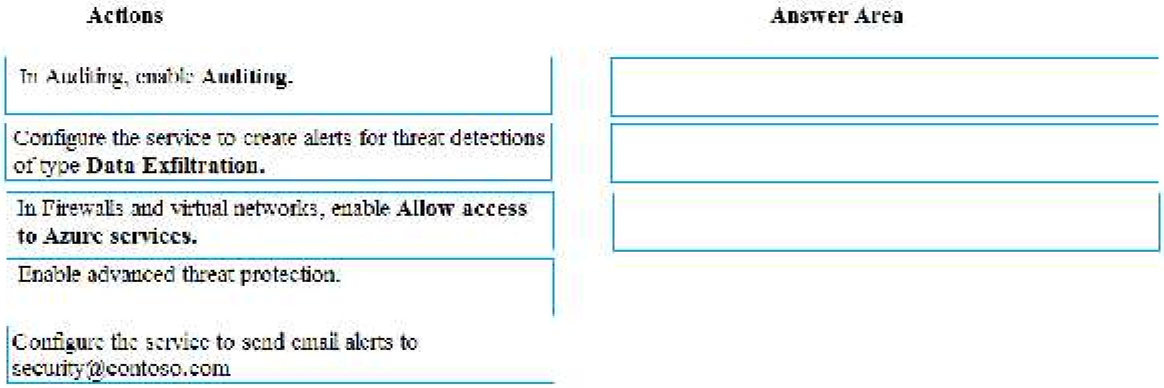
A company uses Microsoft Azure SQL Database to store sensitive company data. You encrypt the data and only allow access to specified users from specified locations.

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

Select and Place:

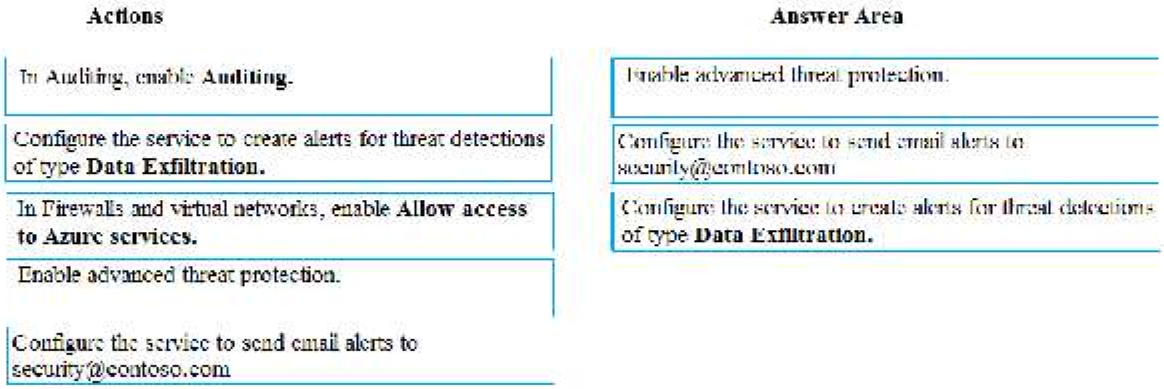


Actions

- In Auditing, enable **Auditing.**
- Configure the service to create alerts for threat detections of type **Data Exfiltration.**
- In Firewalls and virtual networks, enable **Allow access to Azure services.**
- Enable advanced threat protection.
- Configure the service to send email alerts to security@contoso.com

Answer Area

Correct Answer:



Actions

In Auditing, enable **Auditing.**

Configure the service to create alerts for threat detections of type **Data Exfiltration.**

In Firewalls and virtual networks, enable **Allow access to Azure services.**

Enable advanced threat protection.

Configure the service to send email alerts to security@contoso.com

Answer Area

Enable advanced threat protection.

Configure the service to send email alerts to security@contoso.com

Configure the service to create alerts for threat detections of type **Data Exfiltration.**

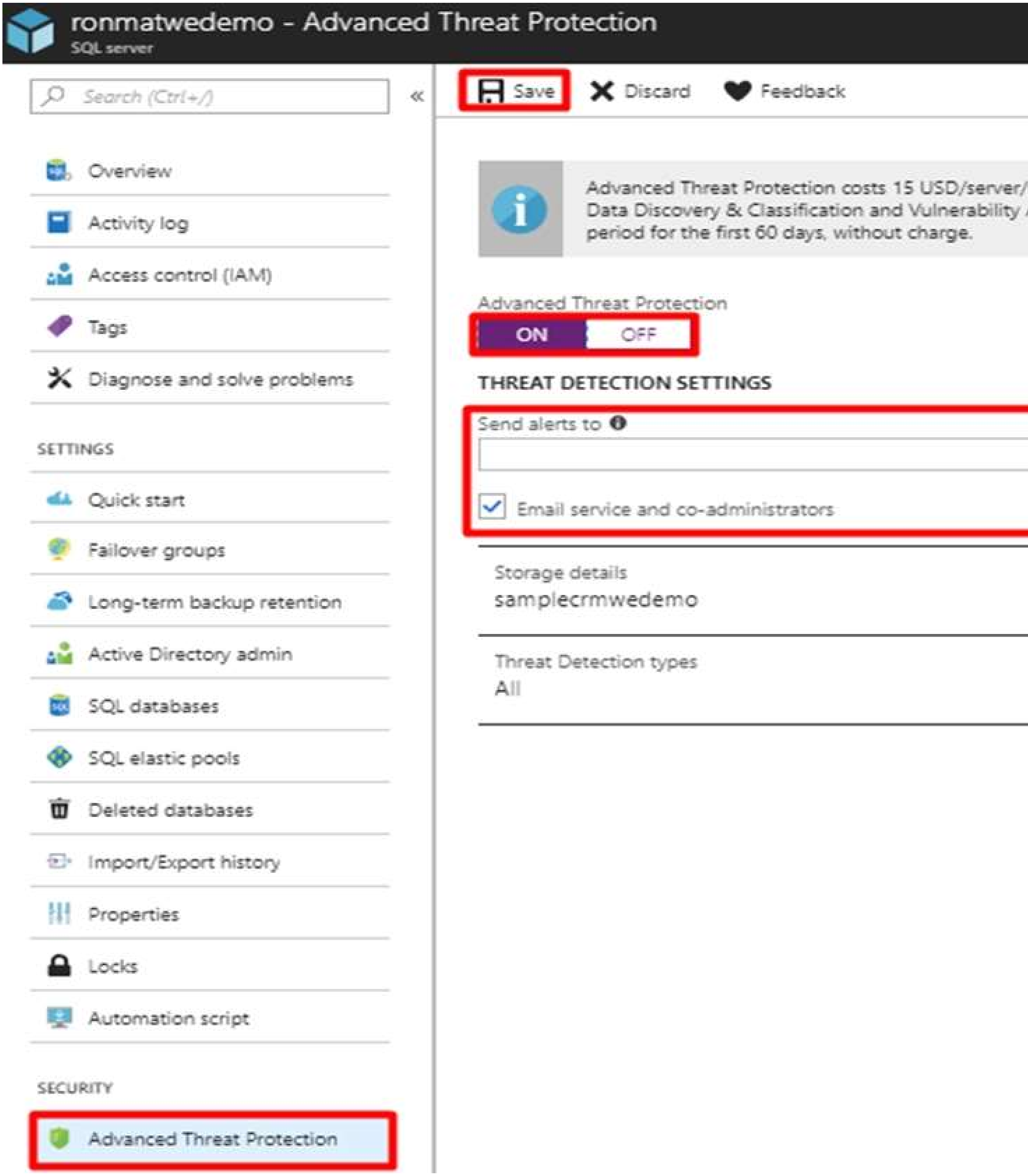
Section: [none]
Explanation

Explanation/Reference:
Explanation:

Step 1: Enable advanced threat protection
Set up threat detection for your database in the Azure portal
1. Launch the Azure portal at https://portal.azure.com.

2. Navigate to the configuration page of the Azure SQL Database server you want to protect. In the security settings, select Advanced Data Security.

3. On the Advanced Data Security configuration page:
Enable advanced data security on the server.
In Threat Detection Settings, in the Send alerts to text box, provide the list of emails to receive security alerts upon detection of anomalous database activities.



Advanced Threat Protection costs 15 USD/server/
Data Discovery & Classification and Vulnerability
period for the first 60 days, without charge.

Advanced Threat Protection: ON | OFF

THREAT DETECTION SETTINGS

Send alerts to

☑ Email service and co-administrators

Storage details
samplecrmwedemo

Threat Detection types
All

Step 2: Configure the service to send email alerts to security@contoso.team

Step 3:..of type data exfiltration
The benefits of Advanced Threat Protection for Azure Storage include:
Detection of anomalous access and data exfiltration activities.

Security alerts are triggered when anomalies in activity occur: access from an unusual location, anonymous access, access by an unusual application, data exfiltration, unexpected delete operations, access permission change, and so on.

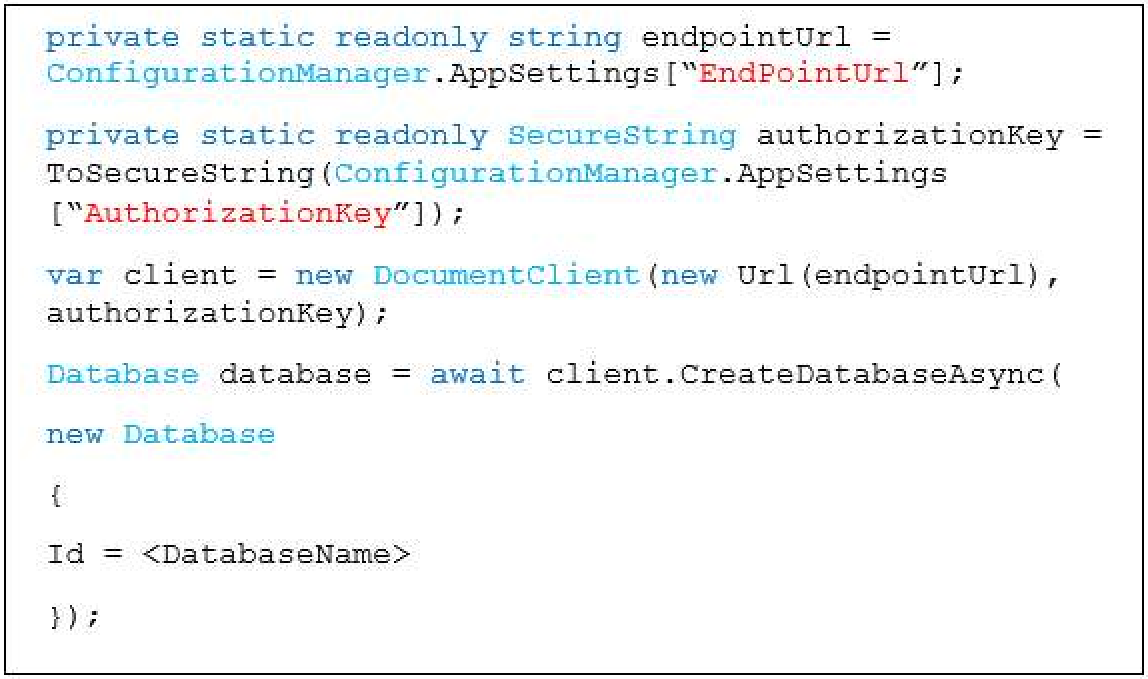
Admins can view these alerts via Azure Security Center and can also choose to be notified of each of them via email.

References:
https://docs.microsoft.com/en-us/azure/sql-database/sql-database-threat-detection

https://www.helpnetsecurity.com/2019/04/04/microsoft-azure-security/

QUESTION 3
HOTSPOT

You develop data engineering solutions for a company. An application creates a database on Microsoft Azure. You have the following code:

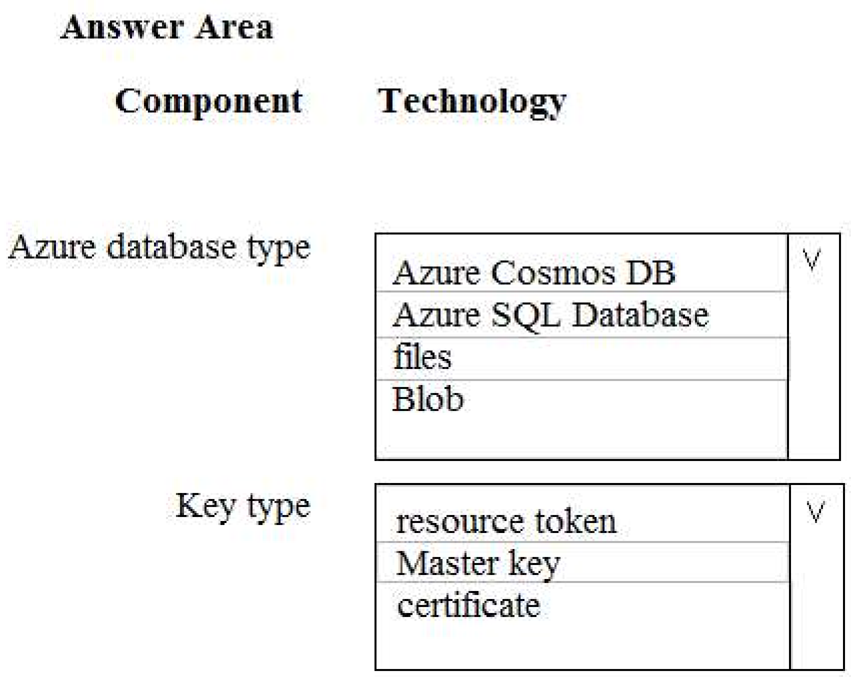


```
private static readonly string endpointUrl =
ConfigurationManager.AppSettings[“EndPointUrl”];

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? To answer, select the appropriate option in the answer area.

NOTE: Each correct selection is worth one point.

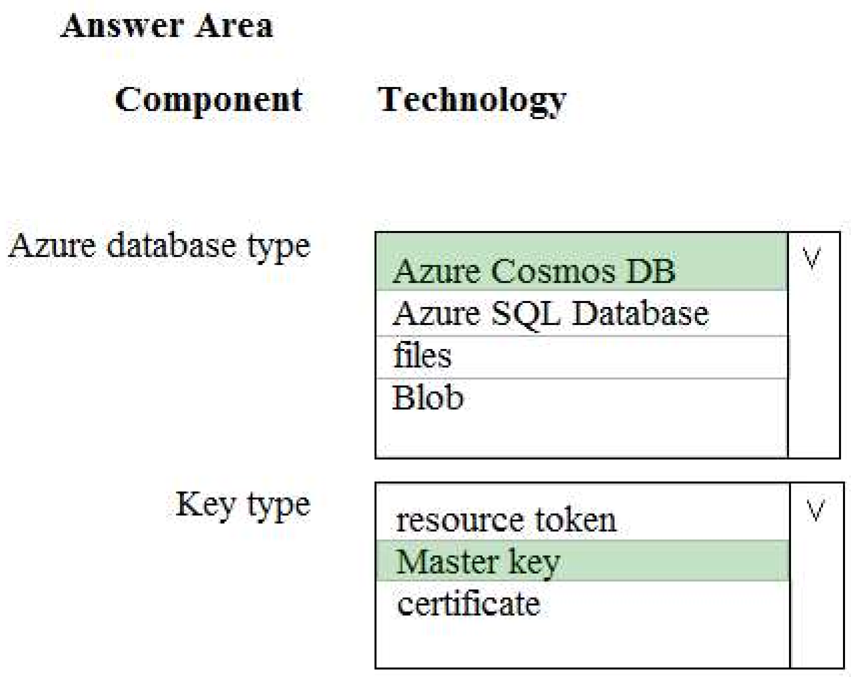
Hot Area:



Answer Area

Component	Technology
Azure database type	Azure Cosmos DB Azure SQL Database files Blob
Key type	resource token Master key certificate

Correct Answer:



Answer Area

Component	Technology
Azure database type	[✓ Azure Cosmos DB] Azure SQL Database files Blob
Key type	resource token [✓ Master key] certificate

Section: [none]
Explanation

Explanation/Reference:

Explanation:

Box 1: Azure Cosmos DB
The DocumentClient.CreateDatabaseAsync(Database, RequestOptions) method creates a database resource as an asychronous operation in the Azure Cosmos DB service.

Box 2: Master Key
Azure Cosmos DB uses two types of keys to authenticate users and provide access to its data and resources: Master Key, Resource Tokens

Master keys provide access to the all the administrative resources for the database account. Master keys:

- Provide access to accounts, databases, users, and permissions.
- Cannot be used to provide granular access to containers and documents.
- Are created during the creation of an account.
- Can be regenerated at any time.

Incorrect Answers:
Resource Token: Resource tokens provide access to the application resources within a database.

References:
https://docs.microsoft.com/en-us/dotnet/api/microsoft.azure.documents.client.documentclient.createdatabaseasync

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

QUESTION 4
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? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

A. `ALTER LOGIN Mary WITH PASSWORD = 'strong_password';`

B. `CREATE LOGIN Mary WITH PASSWORD = 'strong_password';`

C. `ALTER ROLE db_owner ADD MEMBER Mary;`

D. `CREATE USER Mary WITH PASSWORD = 'strong_password';`

E. `GRANT ALTER ANY USER TO Mary;`

Correct Answer: CD
Section: [none]
Explanation

Explanation/Reference:
Explanation:
C: ALTER ROLE adds or removes members to or from a database role, or changes the name of a user-defined database role.
Members of the db_owner fixed database role can perform all configuration and maintenance activities on the database, and can also drop the database in SQL Server.

D: CREATE USER adds a user to the current database.

Note: Logins are created at the server level, while users are created at the database level. In other words, a login allows you to connect to the SQL Server service (also called an instance), and permissions inside the database are granted to the database users, not the logins. The logins will be assigned to server roles (for example, serveradmin) and the database users will be assigned to roles within that database (eg. db_datareader, db_bckupoperator).

References:
https://docs.microsoft.com/en-us/sql/t-sql/statements/alter-role-transact-sql

https://docs.microsoft.com/en-us/sql/t-sql/statements/create-user-transact-sql

QUESTION 5
DRAG DROP

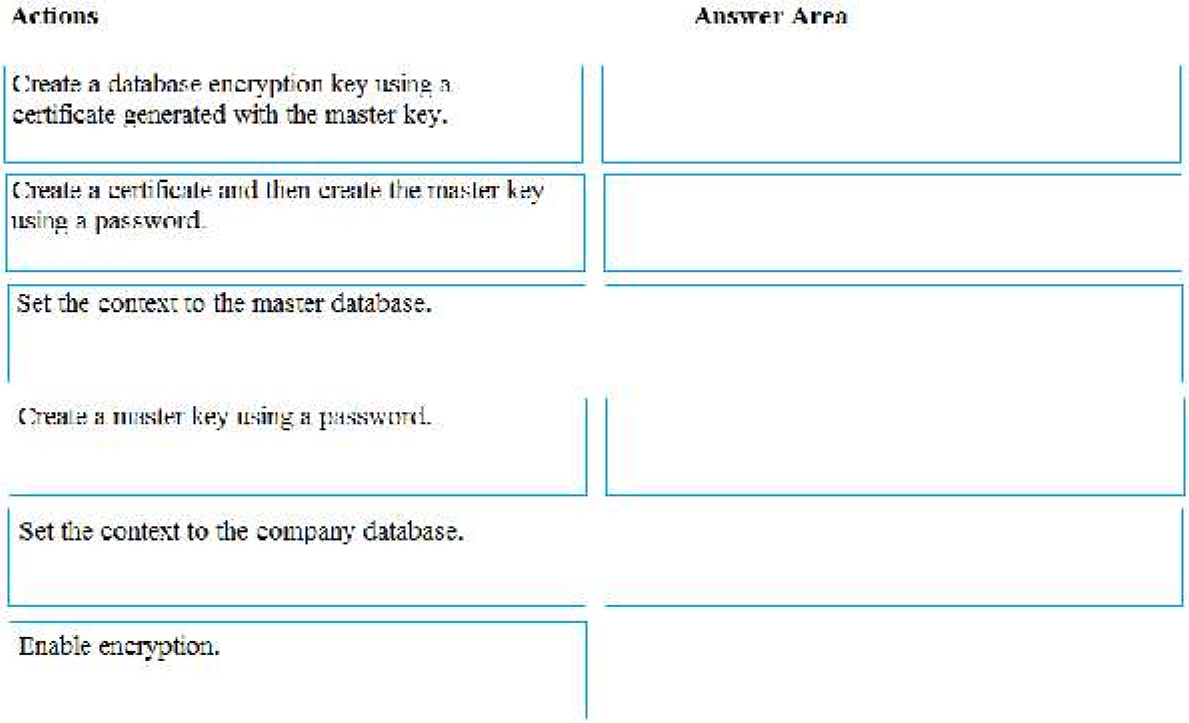
You manage security for a database that supports a line of business application.

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 in sequence? To answer, select the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

Select and Place:



Create a database encryption key using a certificate generated with the master key.

Create a certificate and then create the master key using a password.

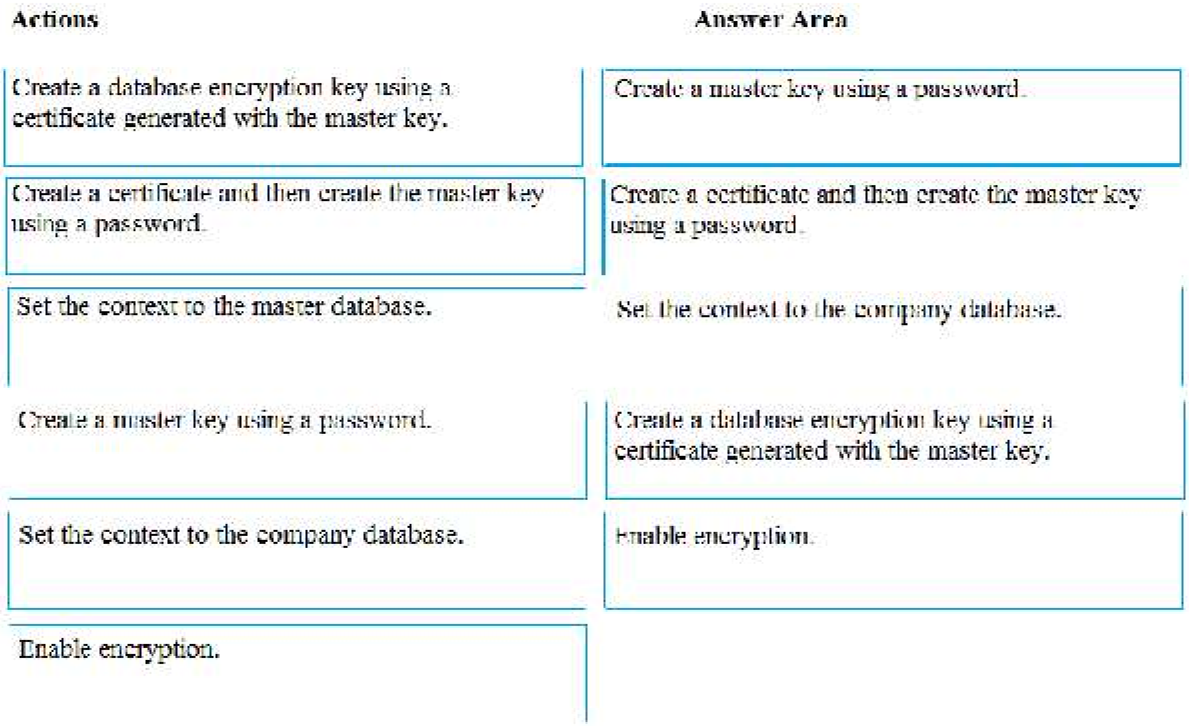
Set the context to the master database.

Create a master key using a password.

Set the context to the company database.

Enable encryption.

Correct Answer:



Actions	Answer Area
Create a database encryption key using a certificate generated with the master key.	Create a master key using a password.
Create a certificate and then create the master key using a password.	Create a certificate and then create the master key using a password.
Set the context to the master database.	Set the context to the company database.
Create a master key using a password.	Create a database encryption key using a certificate generated with the master key.
Set the context to the company database.	Enable encryption.
Enable encryption.	

Section: [none]
Explanation

Explanation/Reference:
Explanation:

Step 1: Create a master key

Step 2: Create or obtain a certificate protected by the master key

Step 3: Set the context to the company database

Step 4: Create a database encryption key and protect it by the certificate

Step 5: Set the database to use encryption

Example code:

```
USE master;
GO
CREATE MASTER KEY ENCRYPTION BY PASSWORD = '<UseStrongPasswordHere>';
go
CREATE CERTIFICATE MyServerCert WITH SUBJECT = 'My DEK Certificate';
go
USE AdventureWorks2012;
GO
CREATE DATABASE ENCRYPTION KEY
WITH ALGORITHM = AES_128
ENCRYPTION BY SERVER CERTIFICATE MyServerCert;
GO
ALTER DATABASE AdventureWorks2012
SET ENCRYPTION ON;
GO
```

References:

https://docs.microsoft.com/en-us/sql/relational-databases/security/encryption/transparent-data-encryption

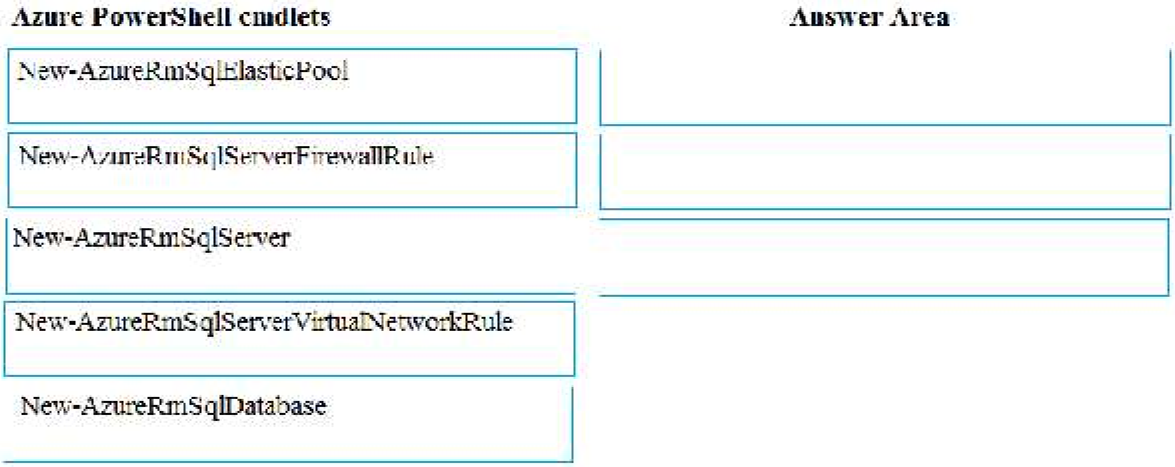
QUESTION 6
DRAG DROP

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 and arrange them in the correct order.

Select and Place:



Azure PowerShell cmdlets

New-AzureRmSqlElasticPool

New-AzureRmSqlServerFirewallRule

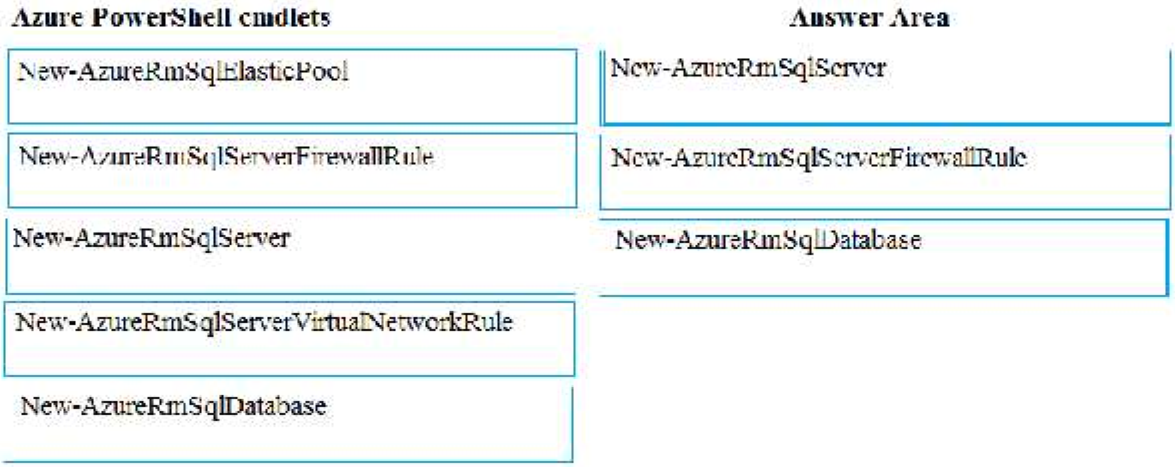
New-AzureRmSqlServer

New-AzureRmSqlServerVirtualNetworkRule

New-AzureRmSqlDatabase

Answer Area

Correct Answer:



Azure PowerShell cmdlets

New-AzureRmSqlElasticPool

New-AzureRmSqlServerFirewallRule

New-AzureRmSqlServer

New-AzureRmSqlServerVirtualNetworkRule

New-AzureRmSqlDatabase

Answer Area

New-AzureRmSqlServer

New-AzureRmSqlServerFirewallRule

New-AzureRmSqlDatabase

Section: [none]
Explanation

Explanation/Reference:
Explanation:

Step 1: New-AzureSqlServer
Create a server.

Step 2: New-AzureRmSqlServerFirewallRule
New-AzureRmSqlServerFirewallRule creates a firewall rule for a SQL Database server.
Can be used to create a server firewall rule that allows access from the specified IP range.

Step 3: New-AzureRmSqlDatabase
Example: Create a database on a specified server

PS C:\>New-AzureRmSqlDatabase -ResourceGroupName "ResourceGroup01" -ServerName "Server01" -DatabaseName "Database01

References:
https://docs.microsoft.com/en-us/azure/sql-database/scripts/sql-database-create-and-configure-database-powershell?toc=%2fpowershell%2fmodule%2ftoc.json

QUESTION 7
HOTSPOT

Your company uses Azure SQL Database and Azure Blob storage.

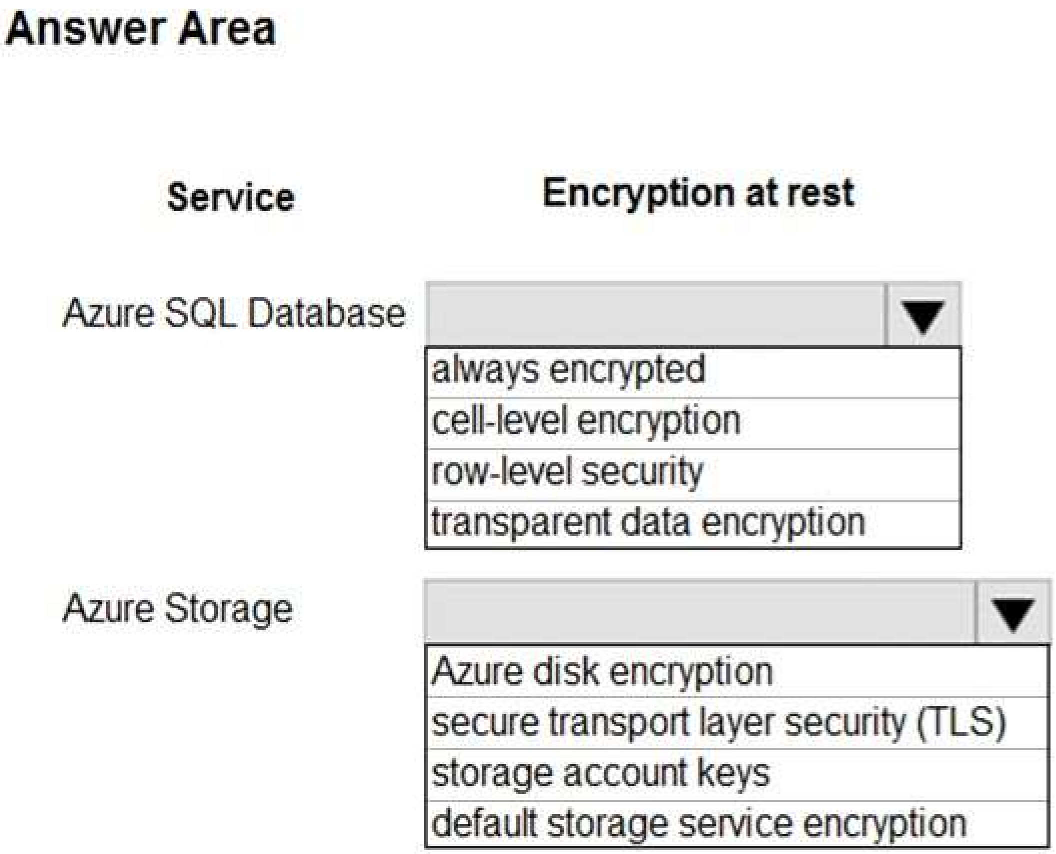
All data at rest must be encrypted by using the company's own key. The solution must minimize administrative effort and the impact to applications which use the database.

You need to configure security.

What should you implement? To answer, select the appropriate option in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:



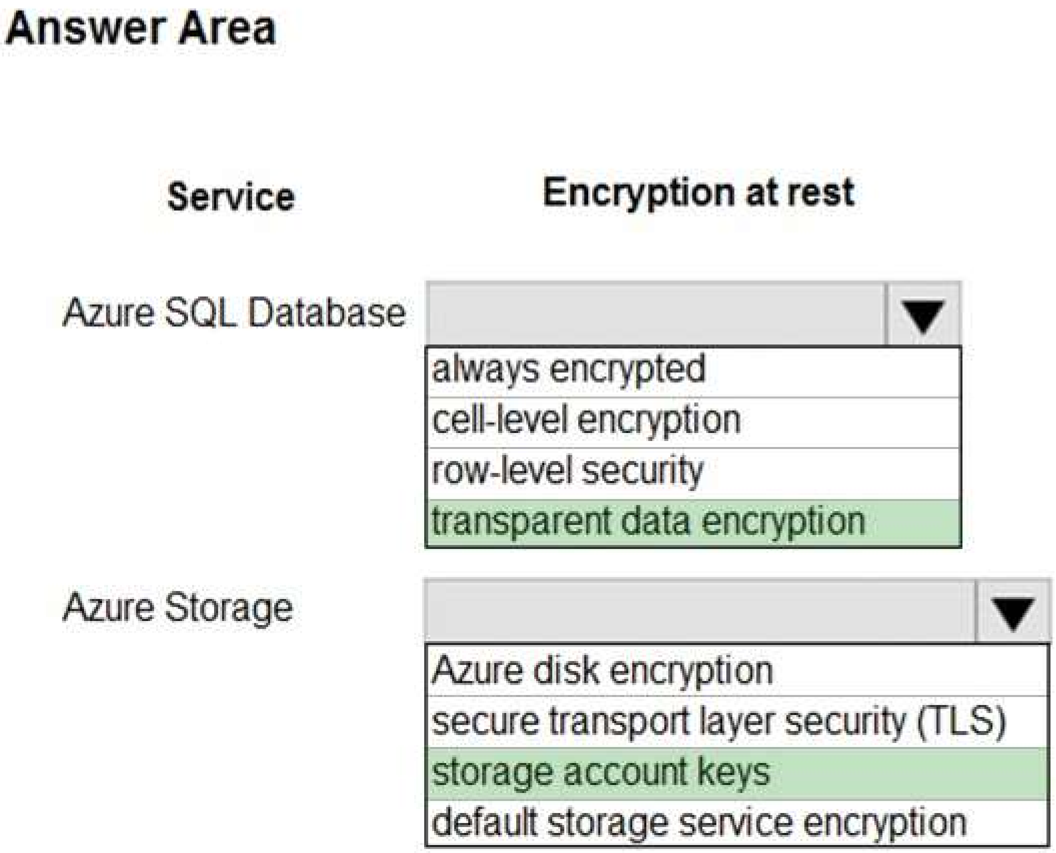
Answer Area

Service | Encryption at rest

Azure SQL Database
- always encrypted
- cell-level encryption
- row-level security
- transparent data encryption

Azure Storage
- Azure disk encryption
- secure transport layer security (TLS)
- storage account keys
- default storage service encryption

Correct Answer:



Answer Area

Service	Encryption at rest
Azure SQL Database	always encrypted cell-level encryption row-level security [✓ transparent data encryption]
Azure Storage	Azure disk encryption secure transport layer security (TLS) [✓ storage account keys] default storage service encryption

Section: [none]
Explanation

Explanation/Reference:
Explanation:

Box 1: transparent data encryption
TDE with customer-managed keys in Azure Key Vault allows to encrypt the Database Encryption Key (DEK) with a customer-managed asymmetric key called TDE Protector. This is also generally referred to as Bring Your Own Key (BYOK) support for Transparent Data Encryption.

Note: Transparent data encryption encrypts the storage of an entire database by using a symmetric key called the database encryption key. This database encryption key is protected by the transparent data encryption protector.

Transparent data encryption (TDE) helps protect Azure SQL Database, Azure SQL Managed Instance, and Azure Data Warehouse against the threat of malicious offline activity by encrypting data at rest. It performs real-time encryption and decryption of the database, associated backups, and transaction log files at rest without requiring changes to the application.

Box 2: Storage account keys
You can rely on Microsoft-managed keys for the encryption of your storage account, or you can manage encryption with your own keys, together with Azure Key Vault.

References:
https://docs.microsoft.com/en-us/azure/sql-database/transparent-data-encryption-azure-sql

https://docs.microsoft.com/en-us/azure/storage/common/storage-service-encryption

QUESTION 8
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? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

A. Assign Azure AD security groups to Azure Data Lake Storage.
B. Configure end-user authentication for the Azure Data Lake Storage account.
C. Configure service-to-service authentication for the Azure Data Lake Storage account.
D. Create security groups in Azure Active Directory (Azure AD) and add project members.
E. Configure access control lists (ACL) for the Azure Data Lake Storage account.

Correct Answer: ADE
Section: [none]
Explanation

Explanation/Reference:
Explanation:
AD: Create security groups in Azure Active Directory. Assign users or security groups to Data Lake Storage Gen1 accounts.

E: Assign users or security groups as ACLs to the Data Lake Storage Gen1 file system

References:
https://docs.microsoft.com/en-us/azure/data-lake-store/data-lake-store-secure-data

Testlet 2

Background
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.

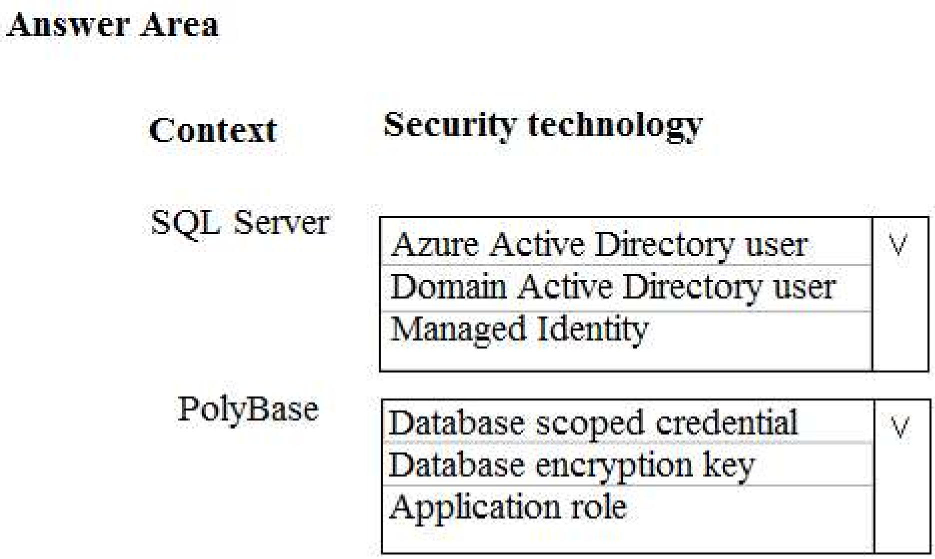
QUESTION 1
HOTSPOT

You need to ensure polling data security requirements are met.

Which security technologies should you use? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

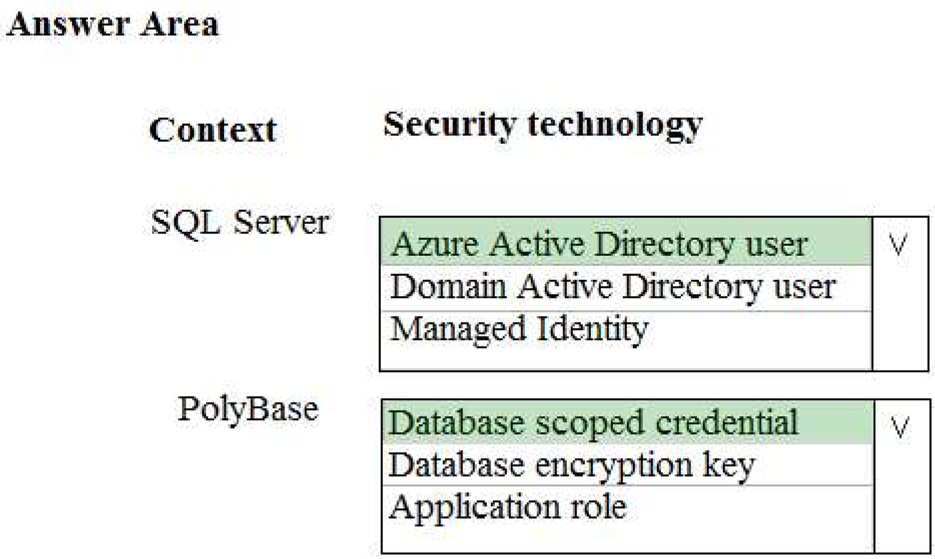
Hot Area:



Answer Area

Context	Security technology
SQL Server	Azure Active Directory user Domain Active Directory user Managed Identity ∨
PolyBase	Database scoped credential Database encryption key Application role ∨

Correct Answer:



Answer Area

Context	Security technology
SQL Server	[✓ Azure Active Directory user] Domain Active Directory user Managed Identity ∨
PolyBase	[✓ Database scoped credential] Database encryption key Application role ∨

Section: [none]
Explanation

Explanation/Reference:
Explanation:

Box 1: Azure Active Directory user
Scenario:
Access to polling data must set on a per-active directory user basis

Box 2: DataBase Scoped Credential
SQL Server uses a database scoped credential to access non-public Azure blob storage or Kerberos-secured Hadoop clusters with PolyBase.

PolyBase cannot authenticate by using Azure AD authentication.

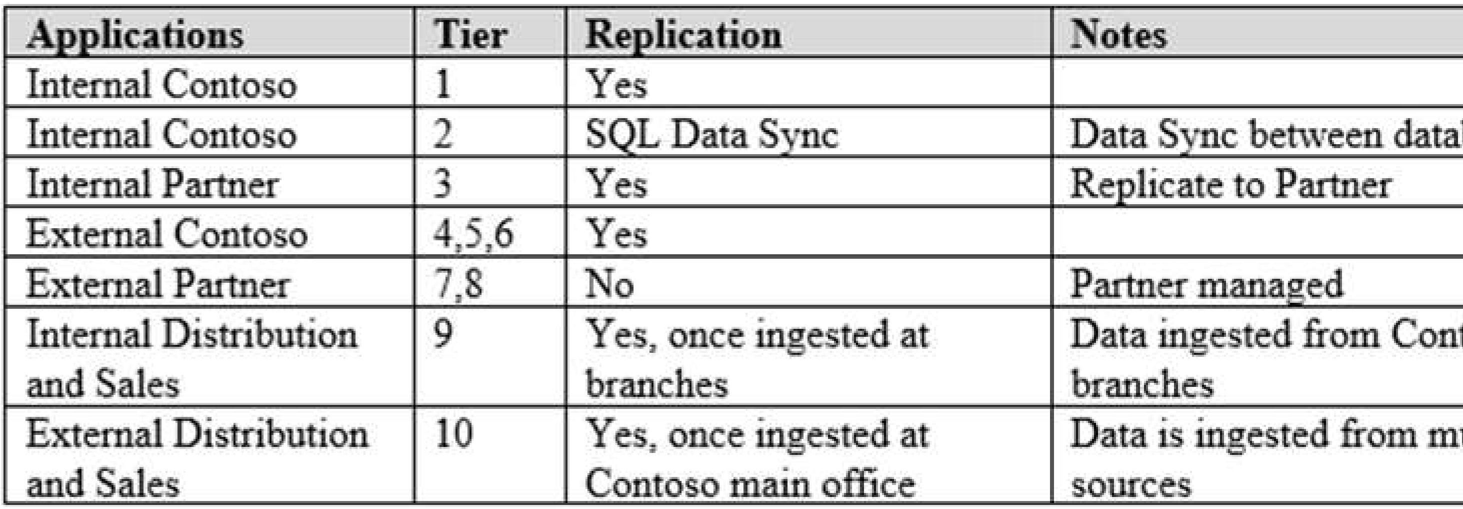
References:
https://docs.microsoft.com/en-us/sql/t-sql/statements/create-database-scoped-credential-transact-sql

Testlet 3

Overview

Current environment
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:



Applications	Tier	Replication	Notes
Internal Contoso	1	Yes	
Internal Contoso	2	SQL Data Sync	Data Sync between data
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 Con branches
External Distribution and Sales	10	Yes, once ingested at Contoso main office	Data is ingested from m 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.

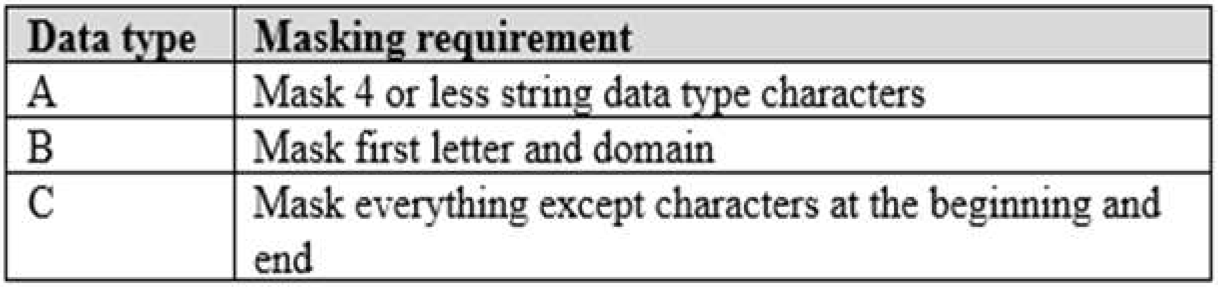
Back up strategies must be put in place for all other standalone Azure SQL Databases using Azure SQL-

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:



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

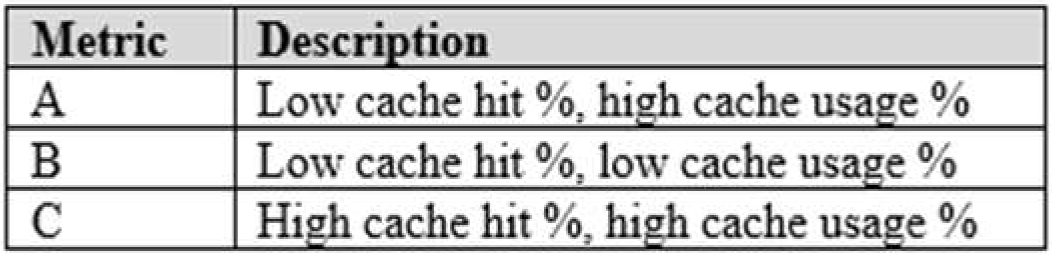
Security and monitoring

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:



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 1
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some questions sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You 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 Randomized
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?

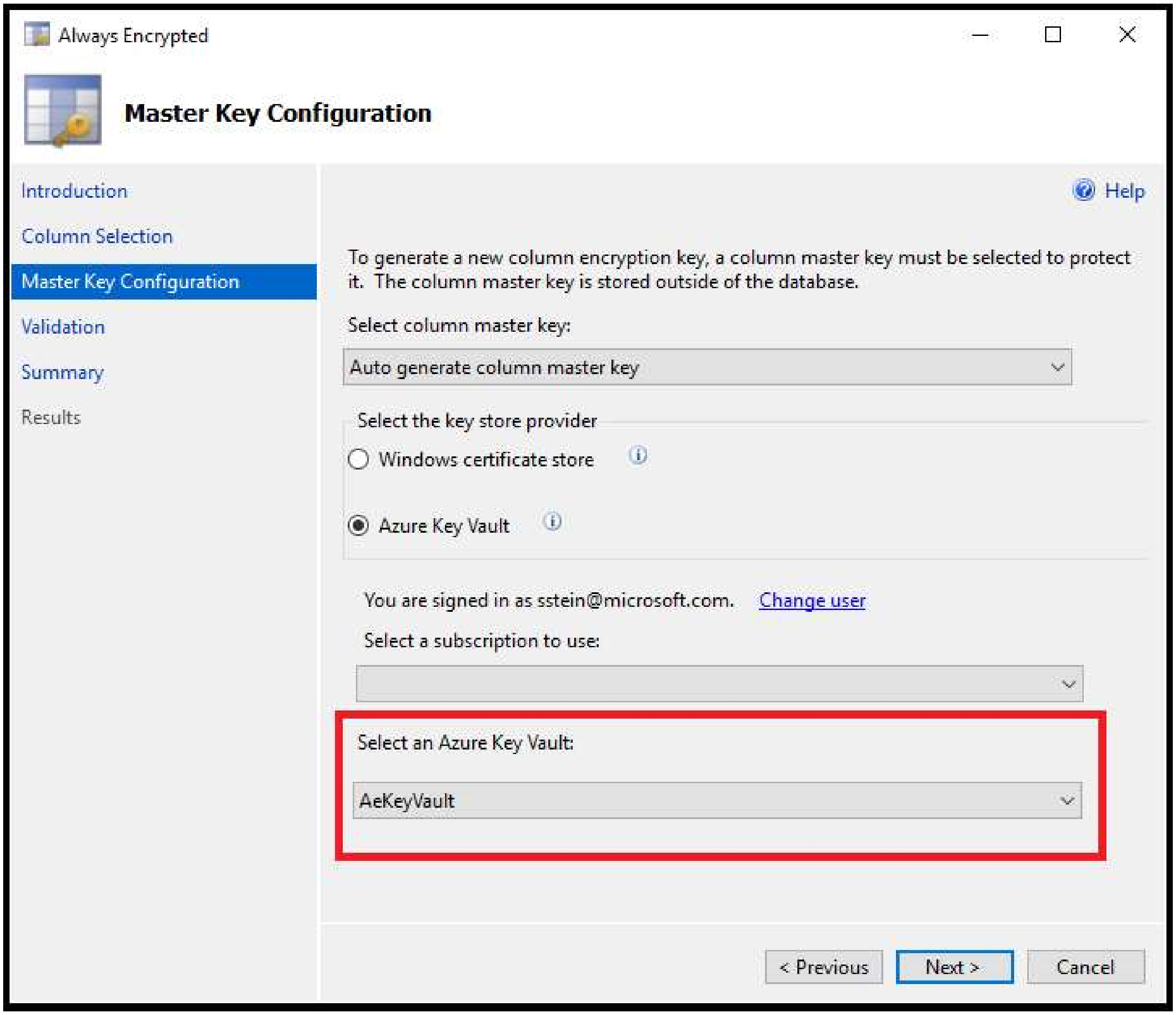
A. Yes

B. No

Correct Answer: B
Section: [none]
Explanation

Explanation/Reference:
Explanation:
Use the Azure Key Vault, not the Windows Certificate Store, to store the master key.

Note: The Master Key Configuration page is where you set up your CMK (Column Master Key) and select the key store provider where the CMK will be stored. Currently, you can store a CMK in the Windows certificate store, Azure Key Vault, or a hardware security module (HSM).



To generate a new column encryption key, a column master key must be selected to protect it. The column master key is stored outside of the database.

Select column master key: Auto generate column master key

Select the key store provider

○ Windows certificate store

● Azure Key Vault

You are signed in as sstein@microsoft.com. Change user

Select a subscription to use:

Select an Azure Key Vault: AeKeyVault

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

QUESTION 2
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some questions sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You 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?

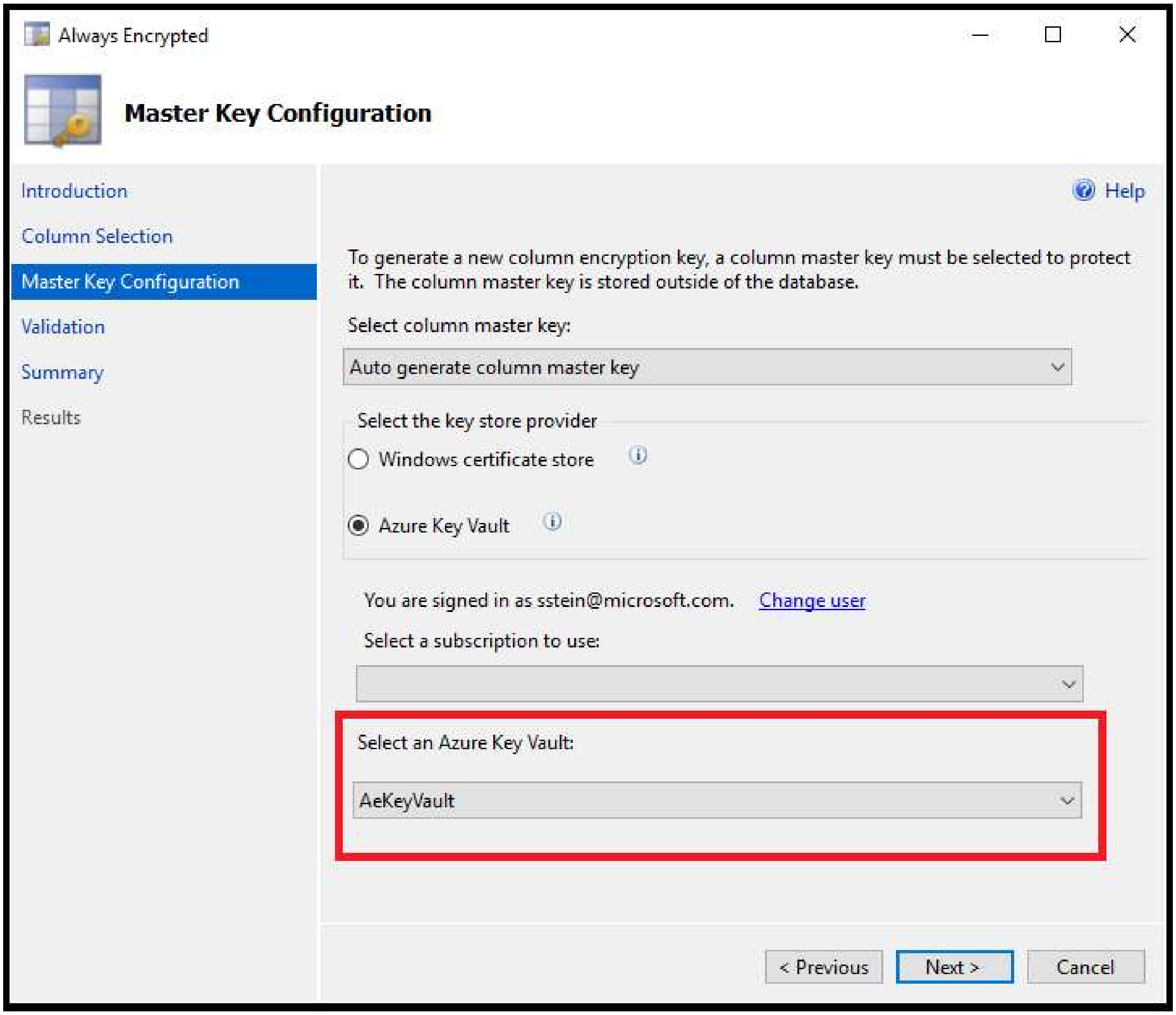
A. Yes

B. No

Correct Answer: B
Section: [none]
Explanation

Explanation/Reference:
Explanation:
Use the Azure Key Vault, not the Windows Certificate Store, to store the master key.

Note: The Master Key Configuration page is where you set up your CMK (Column Master Key) and select the key store provider where the CMK will be stored. Currently, you can store a CMK in the Windows certificate store, Azure Key Vault, or a hardware security module (HSM).



Master Key Configuration

To generate a new column encryption key, a column master key must be selected to protect it. The column master key is stored outside of the database.

Select column master key: Auto generate column master key

Select the key store provider

○ Windows certificate store

● Azure Key Vault

You are signed in as sstein@microsoft.com. Change user

Select a subscription to use:

Select an Azure Key Vault: AeKeyVault

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

QUESTION 3
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some questions sets might have more than one correct solution, while others might not have a correct solution.
After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You 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 Azure Key Vault
5. Validate configuration results and deploy the solution

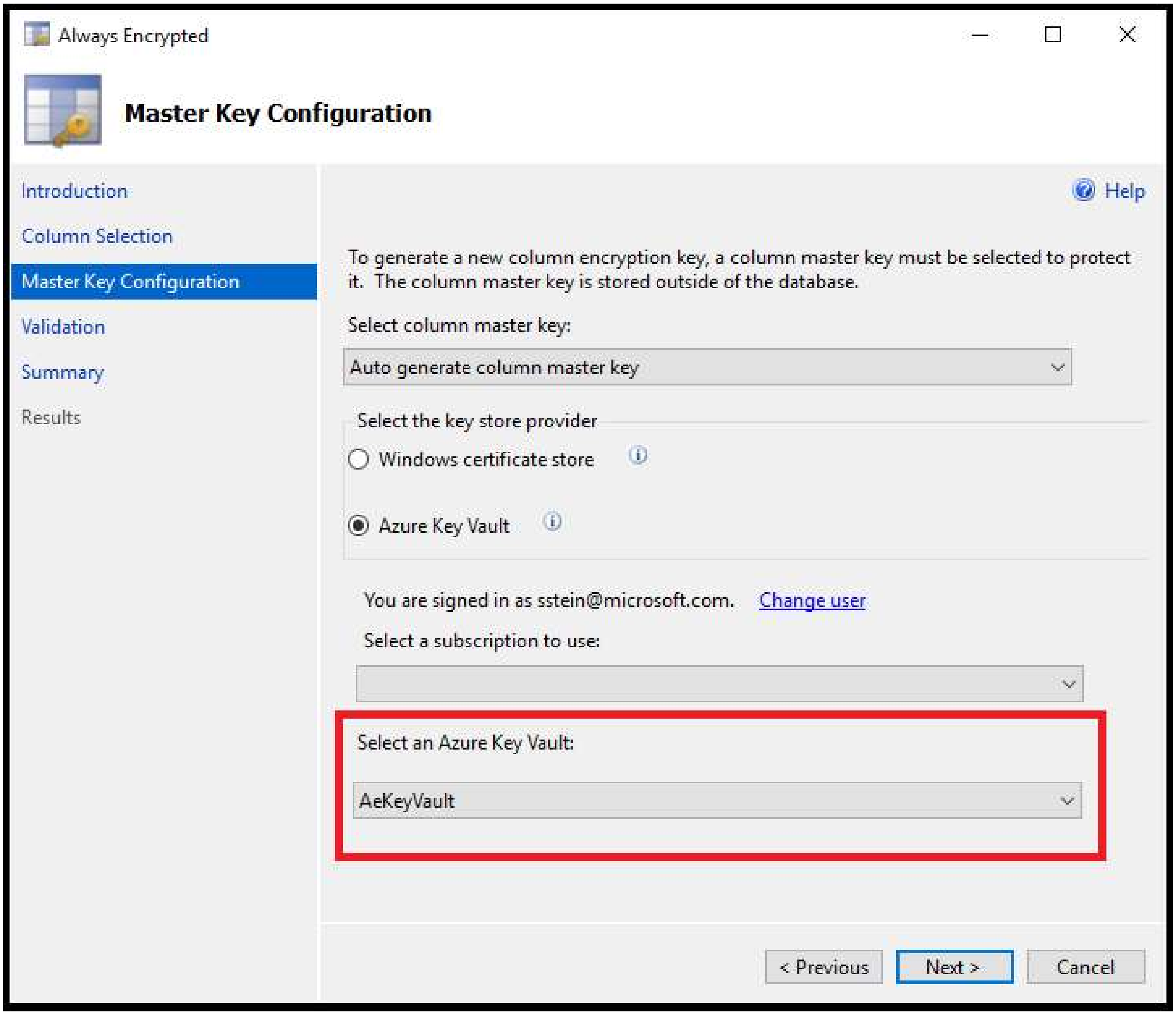
Does the solution meet the goal?

A. Yes
B. No

Correct Answer: A
Section: [none]
Explanation

Explanation/Reference:
Explanation:
We use the Azure Key Vault, not the Windows Certificate Store, to store the master key.

Note: The Master Key Configuration page is where you set up your CMK (Column Master Key) and select the key store provider where the CMK will be stored. Currently, you can store a CMK in the Windows certificate store, Azure Key Vault, or a hardware security module (HSM).



To generate a new column encryption key, a column master key must be selected to protect it. The column master key is stored outside of the database.

Select column master key: Auto generate column master key

Select the key store provider

○ Windows certificate store

● Azure Key Vault

You are signed in as sstein@microsoft.com. Change user

Select a subscription to use:

Select an Azure Key Vault: AeKeyVault

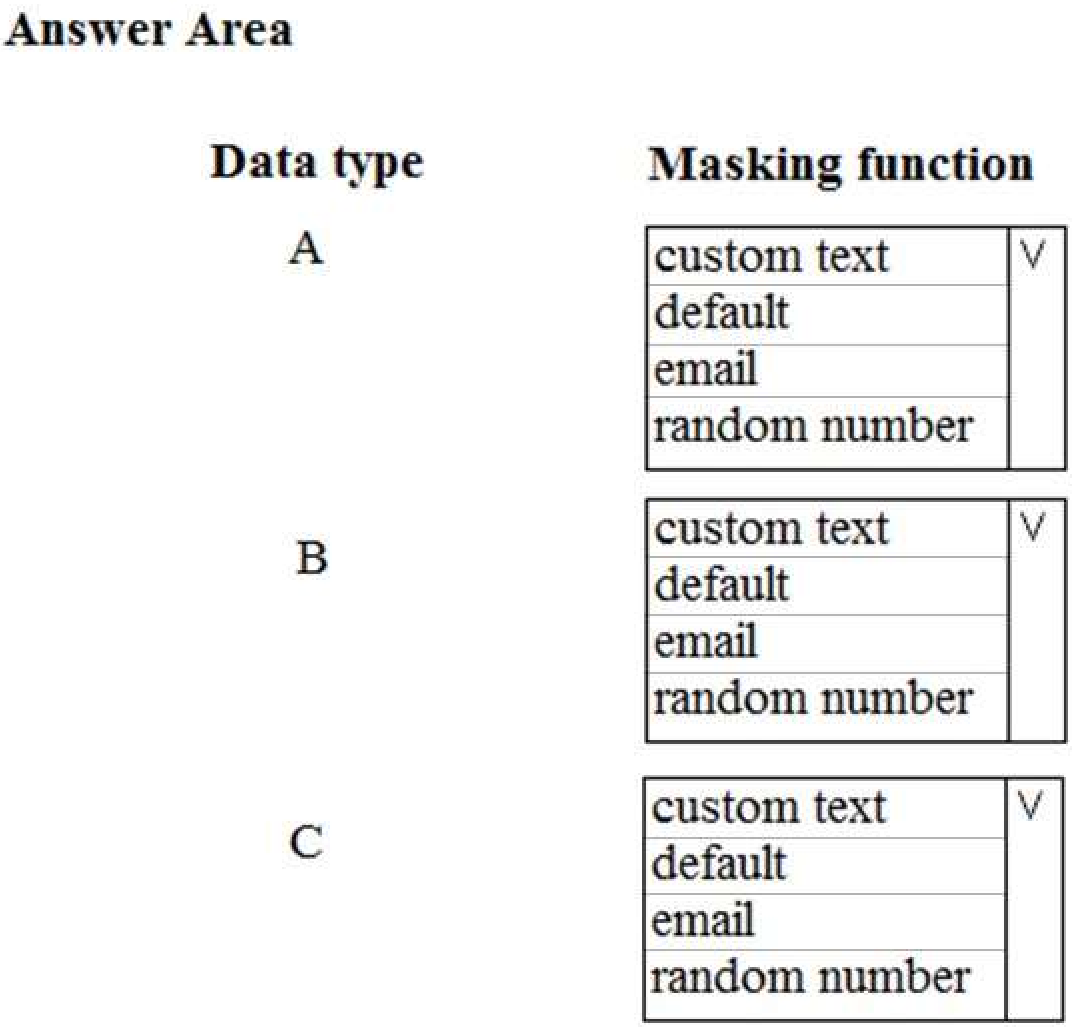
References:
https://docs.microsoft.com/en-us/azure/sql-database/sql-database-always-encrypted-azure-key-vault

QUESTION 4
HOTSPOT

You need to mask tier 1 data. Which functions should you use? To answer, select the appropriate option in the answer area.

NOTE: Each correct selection is worth one point.

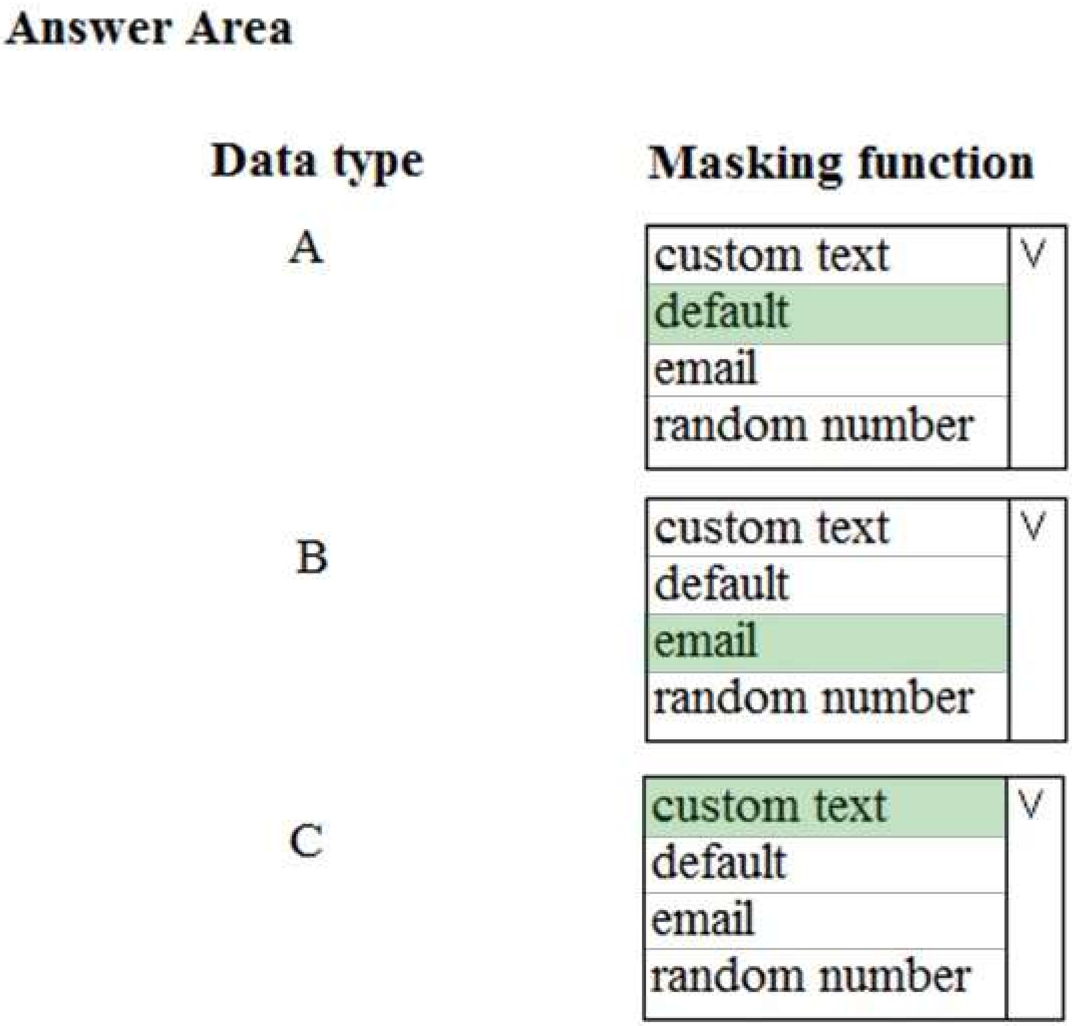
Hot Area:



Answer Area

Data type	Masking function
A	custom text default email random number
B	custom text default email random number
C	custom text default email random number

Correct Answer:



Answer Area

Data type	Masking function
A	custom text [✓ default] email random number
B	custom text default [✓ email] random number
C	[✓ custom text] default email random number

Section: [none]
Explanation

Explanation/Reference:
Explanation:

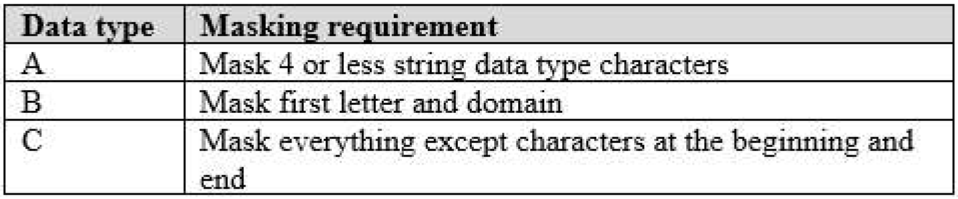
A: Default
Full masking according to the data types of the designated fields.

For string data types, use XXXX or fewer Xs if the size of the field is less than 4 characters (char, nchar, varchar, nvarchar, text, ntext).

B: email

C: Custom text
Custom StringMasking method which exposes the first and last letters and adds a custom padding string in the middle. prefix,[padding],suffix

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



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

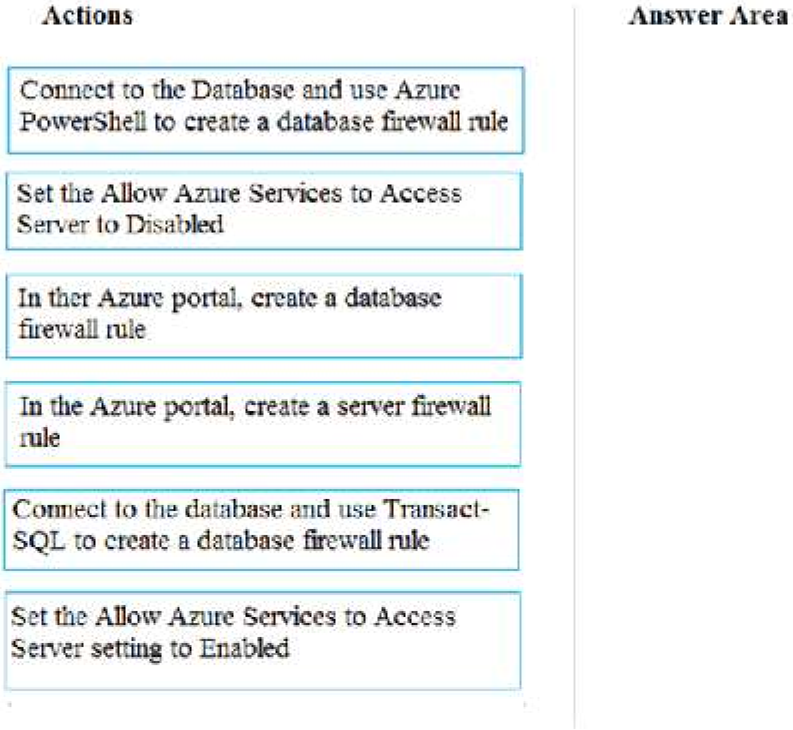
References:
https://docs.microsoft.com/en-us/sql/relational-databases/security/dynamic-data-masking

QUESTION 5
DRAG DROP

You need to set up access to Azure SQL Database for Tier 7 and Tier 8 partners.

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

Select and Place:



Actions

Connect to the Database and use Azure PowerShell to create a database firewall rule

Set the Allow Azure Services to Access Server to Disabled

In ther Azure portal, create a database firewall rule

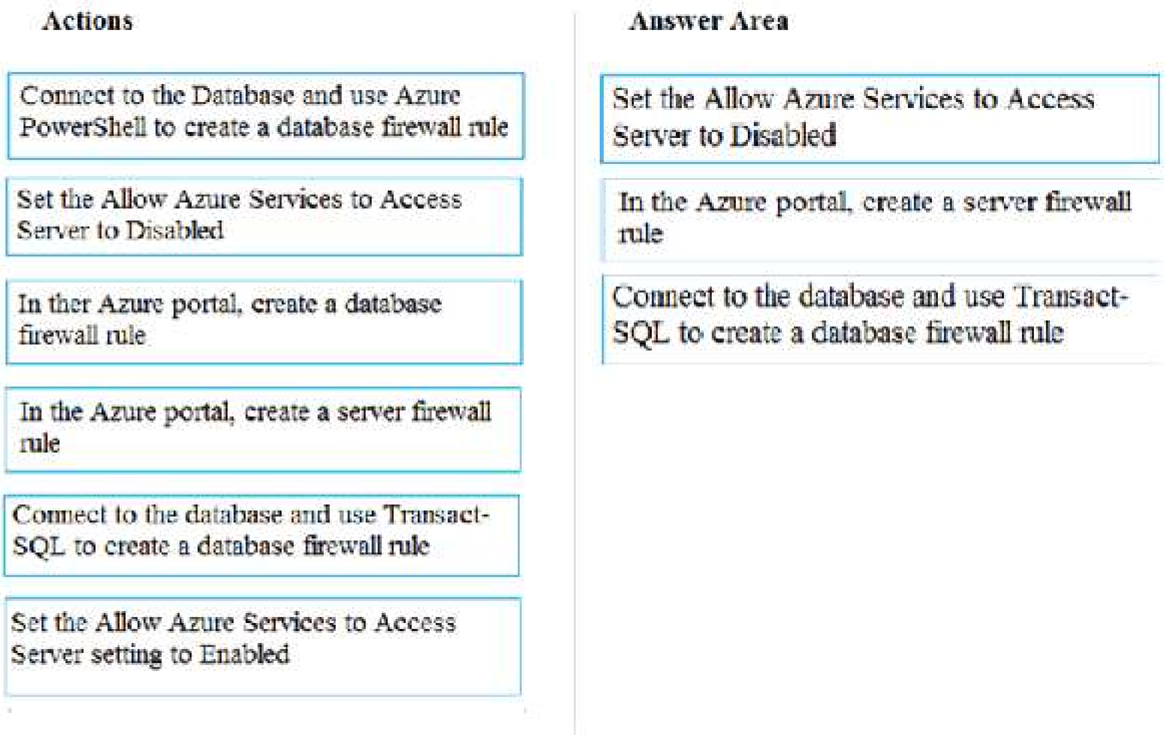
In the Azure portal, create a server firewall rule

Connect to the database and use Transact-SQL to create a database firewall rule

Set the Allow Azure Services to Access Server setting to Enabled

Answer Area

Correct Answer:



Actions

Connect to the Database and use Azure PowerShell to create a database firewall rule

Set the Allow Azure Services to Access Server to Disabled

In ther Azure portal, create a database firewall rule

In the Azure portal, create a server firewall rule

Connect to the database and use Transact-SQL to create a database firewall rule

Set the Allow Azure Services to Access Server setting to Enabled

Answer Area

Set the Allow Azure Services to Access Server to Disabled

In the Azure portal, create a server firewall rule

Connect to the database and use Transact-SQL to create a database firewall rule

Section: [none]
Explanation

Explanation/Reference:
Explanation:

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

Step 1: Set the Allow Azure Services to Access Server setting to Disabled
Set Allow access to Azure services to OFF for the most secure configuration.
By default, access through the SQL Database firewall is enabled for all Azure services, under Allow access to Azure services. Choose OFF to disable access for all Azure services.

Note: The firewall pane has an ON/OFF button that is labeled Allow access to Azure services. The ON setting allows communications from all Azure IP addresses and all Azure subnets. These Azure IPs or subnets might not be owned by you. This ON setting is probably more open than you want your SQL Database to be. The virtual network rule feature offers much finer granular control.

Step 2: In the Azure portal, create a server firewall rule
Set up SQL Database server firewall rules
Server-level IP firewall rules apply to all databases within the same SQL Database server.

To set up a server-level firewall rule:

1. In Azure portal, select SQL databases from the left-hand menu, and select your database on the SQL databases page.
2. On the Overview page, select Set server firewall. The Firewall settings page for the database server opens.

Step 3: Connect to the database and use Transact-SQL to create a database firewall rule
Database-level firewall rules can only be configured using Transact-SQL (T-SQL) statements, and only after you've configured a server-level firewall rule.

To setup a database-level firewall rule:

1. Connect to the database, for example using SQL Server Management Studio.

2. In Object Explorer, right-click the database and select New Query.
3. In the query window, add this statement and modify the IP address to your public IP address:

- EXECUTE sp_set_database_firewall_rule N'Example DB Rule','0.0.0.4','0.0.0.4';

4. On the toolbar, select Execute to create the firewall rule.

References:
https://docs.microsoft.com/en-us/azure/sql-database/sql-database-security-tutorial

Question Set 1

QUESTION 1

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 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?

A. Yes
B. No

Correct Answer: B
Section: [none]
Explanation

Explanation/Reference:
Explanation:
Ambari Web UI does not provide information to suggest how to scale.

Instead monitor clusters by using Azure Log Analytics and HDInsight cluster management solutions.

References:
https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-hadoop-oms-log-analytics-tutorial

https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-hadoop-manage-ambari

QUESTION 2

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 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?

A. Yes
B. No

Correct Answer: A
Section: [none]
Explanation

Explanation/Reference:
Explanation:
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.

References:
https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-hadoop-oms-log-analytics-tutorial

QUESTION 3
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 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?

A. Yes
B. No

Correct Answer: B
Section: [none]
Explanation

Explanation/Reference:
Explanation:
Instead monitor clusters by using Azure Log Analytics and HDInsight cluster management solutions.

References:
https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-hadoop-oms-log-analytics-tutorial

QUESTION 4
HOTSPOT

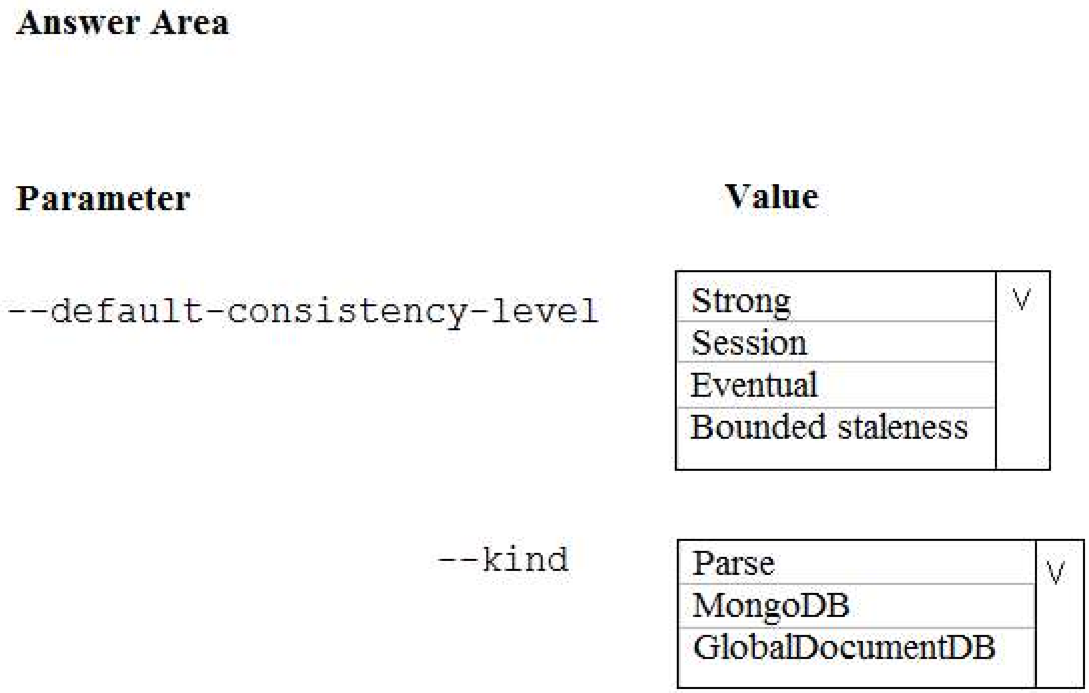
A company is planning to use Microsoft Azure Cosmos DB as the data store for an application. You have the following Azure CLI command:

```
az cosmosdb create --name "cosmosdbdev1" --resource-group "rgdev"
```

You need to minimize latency and expose the SQL API. How should you complete the command? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

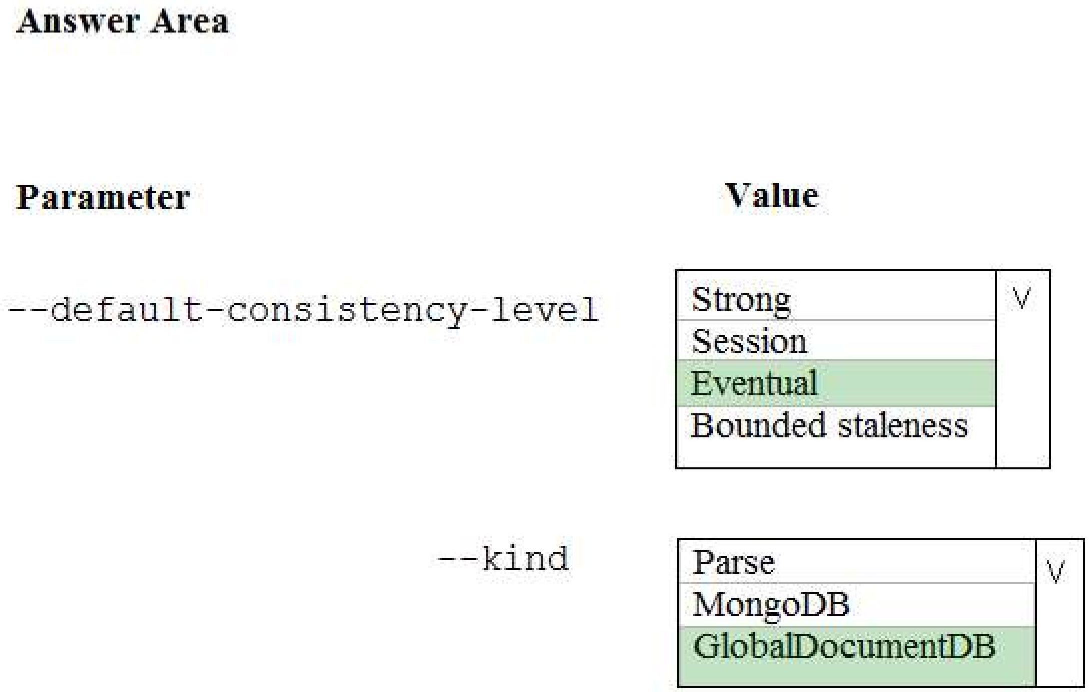
Hot Area:



Answer Area

Parameter	Value
`--default-consistency-level`	Strong Session Eventual Bounded staleness
`--kind`	Parse MongoDB GlobalDocumentDB

Correct Answer:



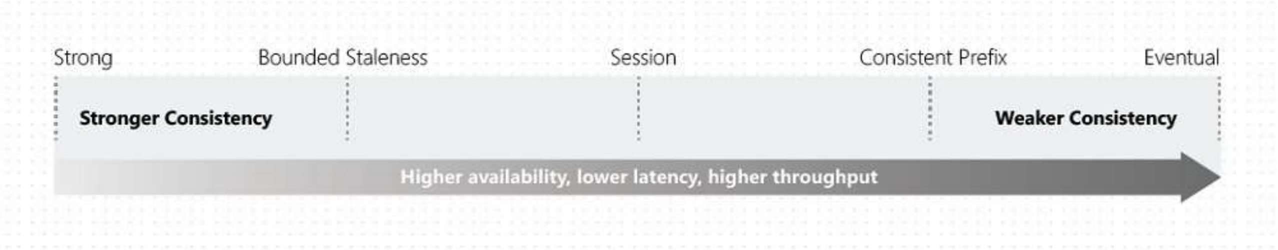
Answer Area

Parameter	Value
`--default-consistency-level`	Strong Session [✓ Eventual] Bounded staleness
`--kind`	Parse MongoDB [✓ GlobalDocumentDB]

Section: [none]
Explanation

Explanation/Reference:
Explanation:

Box 1: Eventual
With Azure Cosmos DB, developers can choose from five well-defined consistency models on the consistency spectrum. From strongest to more relaxed, the models include strong, bounded staleness, session, consistent prefix, and eventual consistency.
The following image shows the different consistency levels as a spectrum.



Box 2: GlobalDocumentDB
Select Core(SQL) to create a document database and query by using SQL syntax.

Note: The API determines the type of account to create. Azure Cosmos DB provides five APIs: Core(SQL) and MongoDB for document databases, Gremlin for graph databases, Azure Table, and Cassandra.

References:
https://docs.microsoft.com/en-us/azure/cosmos-db/consistency-levels

https://docs.microsoft.com/en-us/azure/cosmos-db/create-sql-api-dotnet

QUESTION 5
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?

A. Azure Log Analytics log search query
B. Ambari REST API
C. Azure Monitor Metrics
D. HDInsight .NET SDK
E. Azure Log Analytics alert rule query

Correct Answer: B
Section: [none]
Explanation

Explanation/Reference:
Explanation:
Ambari is the recommended tool for monitoring the health for any given HDInsight cluster.

Note: Azure HDInsight is a high-availability service that has redundant gateway nodes, head nodes, and ZooKeeper nodes to keep your HDInsight clusters running smoothly. While this ensures that a single failure will not affect the functionality of a cluster, you may still want to monitor cluster health so you are alerted when an issue does arise. Monitoring cluster health refers to monitoring whether all nodes in your cluster and the components that run on them are available and functioning correctly.
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. The specific metrics shown depend on cluster type. The “Hosts” tab shows metrics for individual nodes so you can ensure the load on your cluster is evenly distributed.

References:
https://azure.microsoft.com/en-us/blog/monitoring-on-hdinsight-part-1-an-overview/

QUESTION 6
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?

A. Azure HDInsight .NET SDK
B. Azure HDInsight REST API
C. Ambari REST API
D. Azure Log Analytics
E. Ambari Web UI

Correct Answer: E
Section: [none]
Explanation

Explanation/Reference:
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. The specific metrics shown depend on cluster type. The “Hosts” tab shows metrics for individual nodes so you can ensure the load on your cluster is evenly distributed.

The Apache Ambari project is aimed at making Hadoop management simpler by developing software for provisioning, managing, and monitoring Apache Hadoop clusters. Ambari provides an intuitive, easy-to-use Hadoop management web UI backed by its RESTful APIs.

References:
https://azure.microsoft.com/en-us/blog/monitoring-on-hdinsight-part-1-an-overview/

https://ambari.apache.org/

QUESTION 7
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?

A. 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.'
B. 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.'
C. 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.
```

D.
```
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.'
```

Correct Answer: C
Section: [none]
Explanation

Explanation/Reference:
Explanation:
Customer Scenario:
SQL Server 2016 or SQL DW connected to Azure blob storage. The CREATE EXTERNAL TABLE DDL points to a directory (and not a specific file) and the directory contains files with different schemas.

SSMS Error:
Select query on the external table gives the following error:
Msg 7320, Level 16, State 110, Line 14
Cannot execute the query "Remote Query" against OLE DB provider "SQLNCLI11" for linked server "(null)". Query aborted-- the maximum reject threshold (0 rows) was reached while reading from an external source: 1 rows rejected out of total 1 rows processed.

Possible Reason:
The reason this error happens is because each file has different schema. The PolyBase external table DDL when pointed to a directory recursively reads all the files in that directory. When a column or data type mismatch happens, this error could be seen in SSMS.
Possible Solution:
If the data for each table consists of one file, then use the filename in the LOCATION section prepended by the directory of the external files. If there are multiple files per table, put each set of files into different directories in Azure Blob Storage and then you can point LOCATION to the directory instead of a particular file. The latter suggestion is the best practices recommended by SQLCAT even if you have one file per table.

Incorrect Answers:
A: Possible Reason: Kerberos is not enabled in Hadoop Cluster.

References:
https://techcommunity.microsoft.com/t5/DataCAT/PolyBase-Setup-Errors-and-Possible-Solutions/ba-p/305297

QUESTION 8
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some questions sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

A company uses Azure Data Lake Gen 1 Storage to store big data related to consumer behavior.

You need to implement logging.

Solution: Use information stored in Azure Active Directory reports.

Does the solution meet the goal?

A. Yes
B. No

Correct Answer: B
Section: [none]

Explanation

Explanation/Reference:
Explanation:
Instead configure Azure Data Lake Storage diagnostics to store logs and metrics in a storage account.

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

QUESTION 9
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some questions sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

A company uses Azure Data Lake Gen 1 Storage to store big data related to consumer behavior.

You need to implement logging.

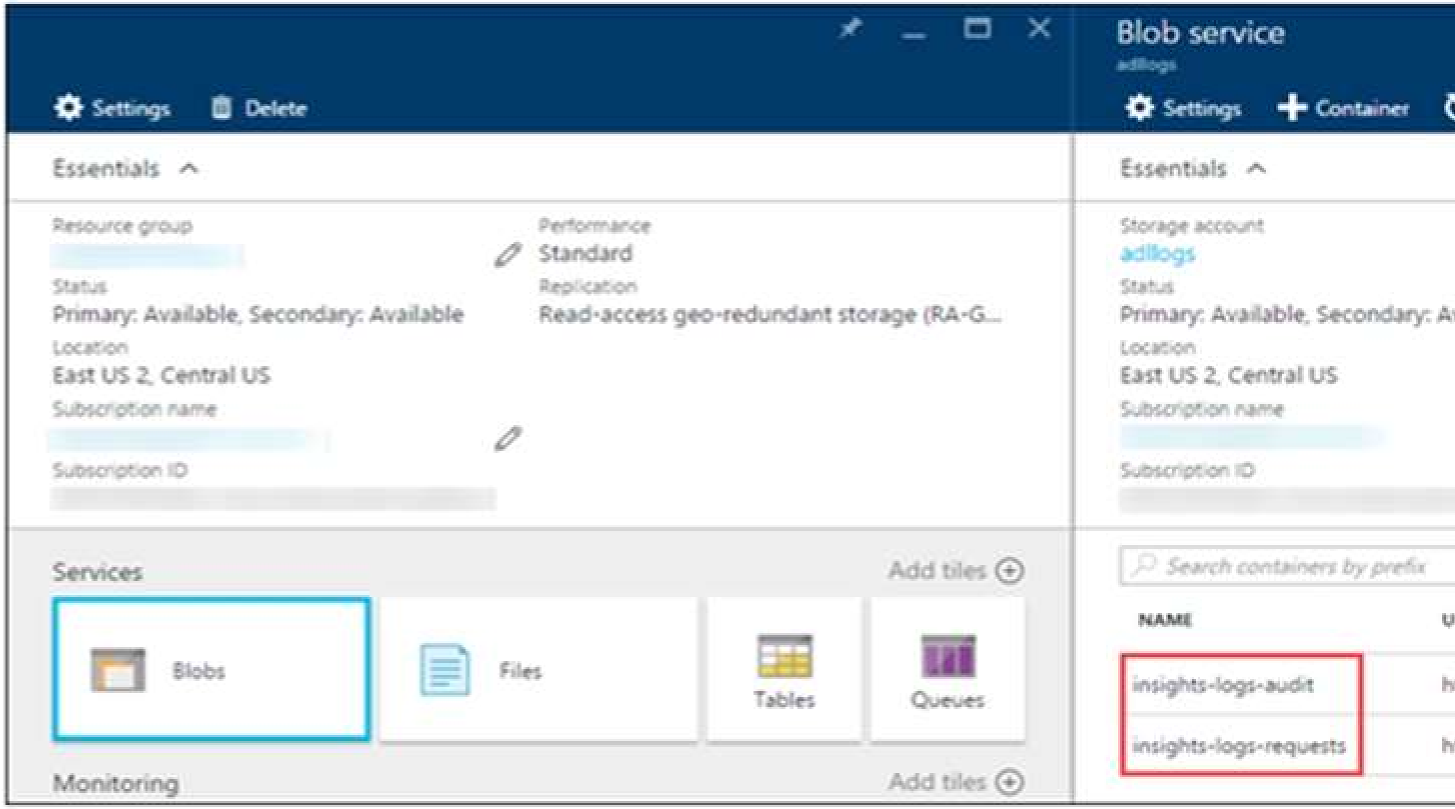
Solution: Configure Azure Data Lake Storage diagnostics to store logs and metrics in a storage account.

Does the solution meet the goal?

A. Yes
B. No

Correct Answer: A
Section: [none]
Explanation

Explanation/Reference:
Explanation:
From the Azure Storage account that contains log data, open the Azure Storage account blade associated with Data Lake Storage Gen1 for logging, and then click Blobs. The Blob service blade lists two containers.



Resource group
[redacted]
Status
Primary: Available, Secondary: Available
Location
East US 2, Central US
Subscription name
[redacted]
Subscription ID
[redacted]

Performance
Standard
Replication
Read-access geo-redundant storage (RA-G...

Storage account
adllogs
Status
Primary: Available, Secondary: A
Location
East US 2, Central US
Subscription name
[redacted]
Subscription ID
[redacted]

insights-logs-audit
insights-logs-requests

References:

QUESTION 10
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some questions sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

A company uses Azure Data Lake Gen 1 Storage to store big data related to consumer behavior.

You need to implement logging.

Solution: Configure an Azure Automation runbook to copy events.

Does the solution meet the goal?

A. Yes
B. No

Correct Answer: B
Section: [none]
Explanation

Explanation/Reference:
Explanation:
Instead configure Azure Data Lake Storage diagnostics to store logs and metrics in a storage account.

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

QUESTION 11
Your company uses several Azure HDInsight clusters.

The data engineering team reports several errors with some applications using these clusters.

You need to recommend a solution to review the health of the clusters.

What should you include in your recommendation?

A. Azure Automation
B. Log Analytics
C. Application Insights

Correct Answer: B
Section: [none]
Explanation

Explanation/Reference:
Explanation:
Azure Monitor logs integration. Azure Monitor logs enables data generated by multiple resources such as HDInsight clusters, to be collected and aggregated in one place to achieve a unified monitoring experience.

As a prerequisite, you will need a Log Analytics Workspace to store the collected data. If you have not already created one, you can follow the instructions for creating a Log Analytics Workspace.

You can then easily configure an HDInsight cluster to send many workload-specific metrics to Log Analytics.

References:
https://azure.microsoft.com/sv-se/blog/monitoring-on-azure-hdinsight-part-2-cluster-health-and-availability/

QUESTION 12
DRAG DROP

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.

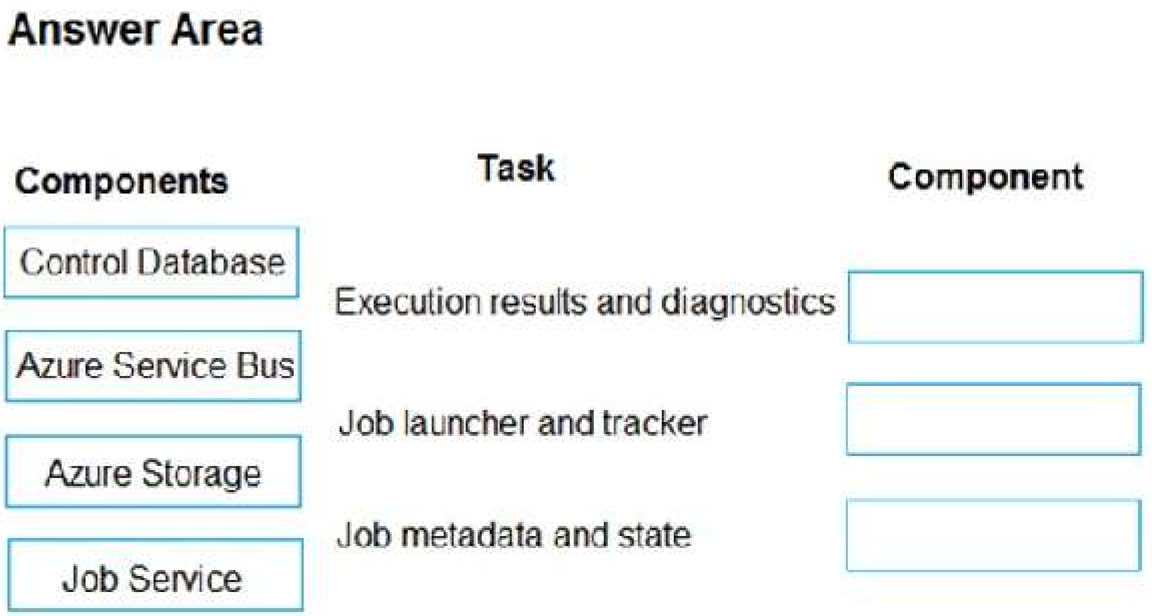
You need to analyze, troubleshoot, and report on components responsible for running Elastic Database jobs.

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.

NOTE: Each correct selection is worth one point.

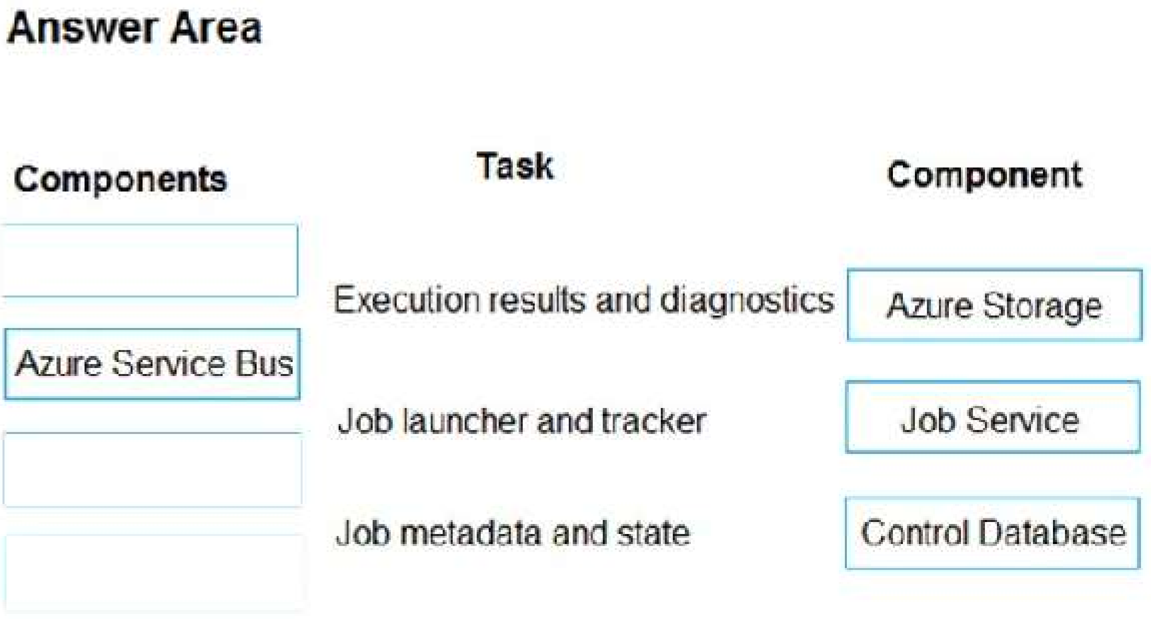
Select and Place:



Answer Area

Components	Task	Component
Control Database		
Azure Service Bus	Execution results and diagnostics	
Azure Storage	Job launcher and tracker	
Job Service	Job metadata and state	

Correct Answer:



Answer Area

Components	Task	Component
Azure Service Bus	Execution results and diagnostics	Azure Storage
	Job launcher and tracker	Job Service
	Job metadata and state	Control Database

Section: [none]
Explanation

Explanation/Reference:
Explanation:

Execution results and diagnostics: Azure Storage

Job launcher and tracker: Job Service

Job metadata and state: Control database

The Job database is used for defining jobs and tracking the status and history of job executions. The Job database is also used to store agent metadata, logs, results, job definitions, and also contains many useful stored procedures, and other database objects, for creating, running, and managing jobs using T-SQL.

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

QUESTION 13
Contoso, 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? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

A. Create a new action group to email alerts@contoso.com.
B. Use alerts@contoso.com as an alert email address.
C. Use all security operations as a condition.
D. Use all Azure SQL Database servers as a resource.
E. Query audit log entries as a condition.

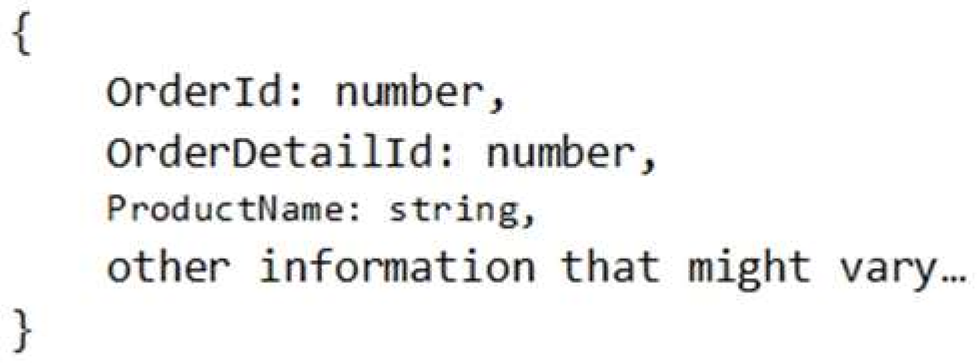
Correct Answer: ACD
Section: [none]
Explanation

Explanation/Reference:
References:
https://docs.microsoft.com/en-us/azure/azure-monitor/platform/alerts-action-rules

QUESTION 14
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this scenario, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

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.



```
{
    OrderId: number,
    OrderDetailId: number,
    ProductName: string,
    other information that might vary…
}
```

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?

A. Yes
B. No

Correct Answer: B
Section: [none]
Explanation

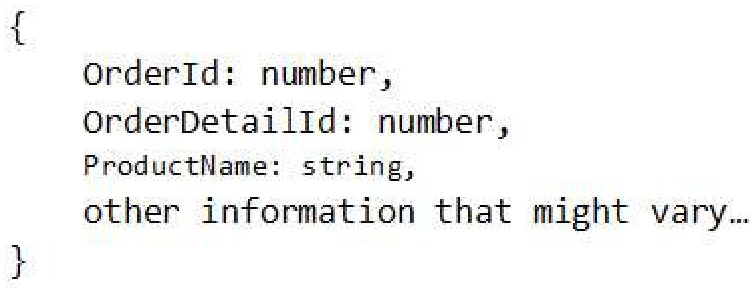
Explanation/Reference:
Explanation:
One option is to have a lookup collection "`ProductName`" for the mapping of "`ProductName`" to "OrderId".

References:
https://azure.microsoft.com/sv-se/blog/azure-cosmos-db-partitioning-design-patterns-part-1/

QUESTION 15
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this scenario, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

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.



```
{
    OrderId: number,
    OrderDetailId: number,
    ProductName: string,
    other information that might vary…
}
```

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 and `OrderId` as a value.

Does this meet the goal?

A. Yes

B. No

Correct Answer: A
Section: [none]
Explanation

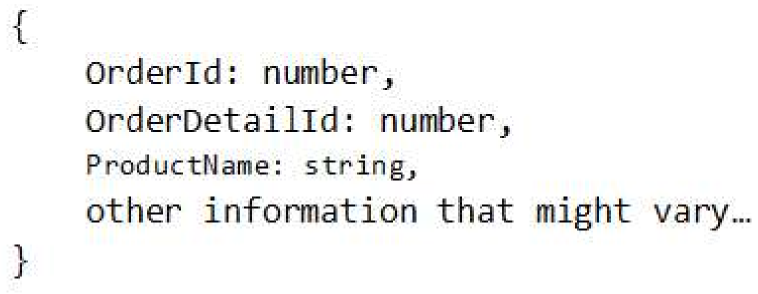
Explanation/Reference:
Explanation:
One option is to have a lookup collection “`ProductName`” for the mapping of “`ProductName`” to “OrderId”.

References:
https://azure.microsoft.com/sv-se/blog/azure-cosmos-db-partitioning-design-patterns-part-1/

QUESTION 16
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this scenario, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

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.



```
{
    OrderId: number,
    OrderDetailId: number,
    ProductName: string,
    other information that might vary…
}
```

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?

A. Yes
B. No

Correct Answer: B
Section: [none]
Explanation

Explanation/Reference:
Explanation:
One option is to have a lookup collection “`ProductName`” for the mapping of “`ProductName`” to “OrderId”.

References:

https://azure.microsoft.com/sv-se/blog/azure-cosmos-db-partitioning-design-patterns-part-1/

QUESTION 17
HOTSPOT

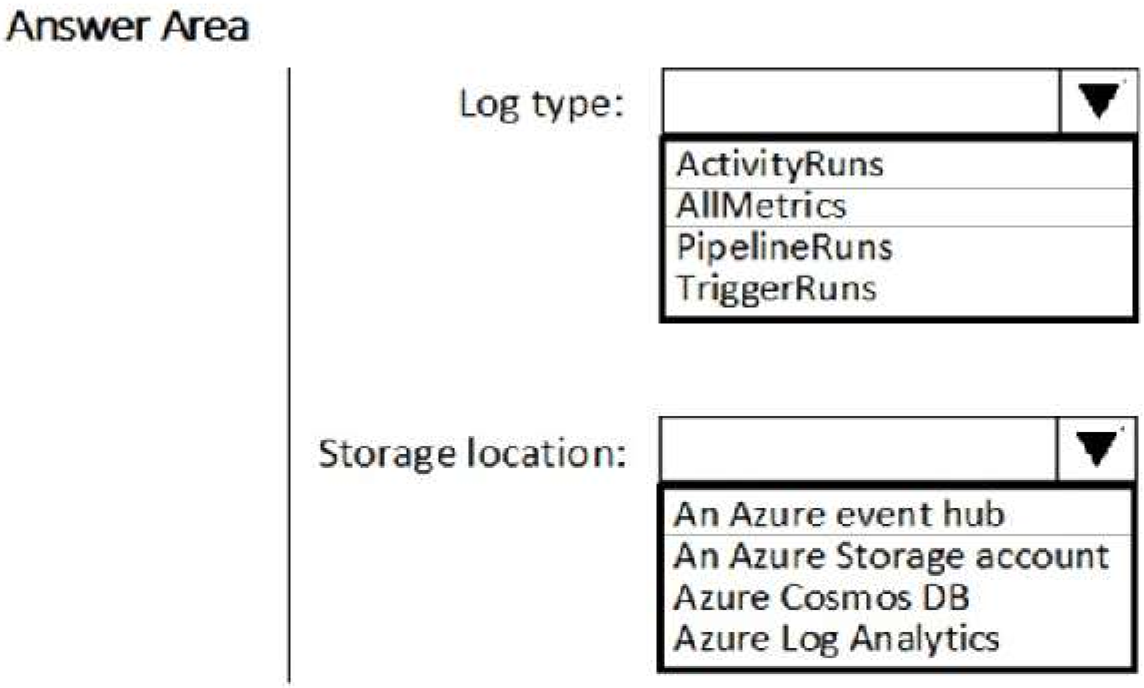
You have a new Azure Data Factory environment.

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? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

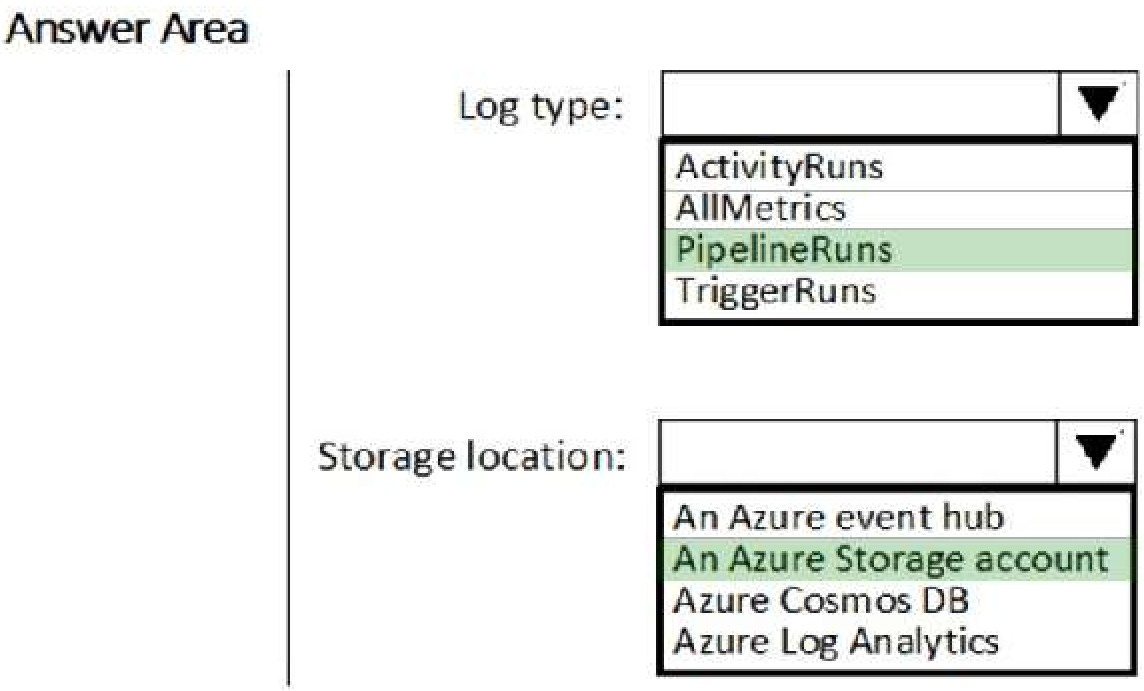


Answer Area

Log type: ▼
ActivityRuns
AllMetrics
PipelineRuns
TriggerRuns

Storage location: ▼
An Azure event hub
An Azure Storage account
Azure Cosmos DB
Azure Log Analytics

Correct Answer:



Answer Area

Log type:

- ActivityRuns
- AllMetrics
- [✓ PipelineRuns]
- TriggerRuns

Storage location:

- An Azure event hub
- [✓ An Azure Storage account]
- Azure Cosmos DB
- Azure Log Analytics

Section: [none]
Explanation

Explanation/Reference:
Log type: PipelineRuns
A pipeline run in Azure Data Factory defines an instance of a pipeline execution.

Storage location: An Azure Storage account
Data Factory stores pipeline-run data for only 45 days. Use Monitor if you want to keep that data for a longer time. With Monitor, you can route diagnostic logs for analysis. You can also keep them in a storage account so that you have factory information for your chosen duration.

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.

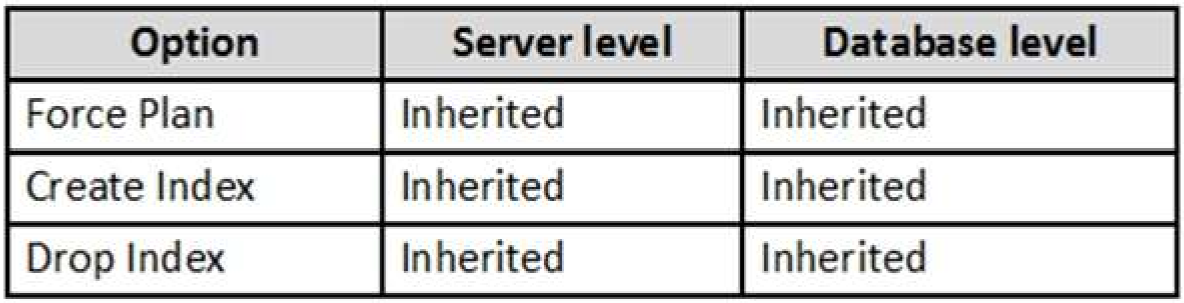
References:
https://docs.microsoft.com/en-us/azure/data-factory/concepts-pipeline-execution-triggers

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

QUESTION 18
HOTSPOT

You are implementing automatic tuning mode for Azure SQL databases.

Automatic tuning is configured as shown in the following table.

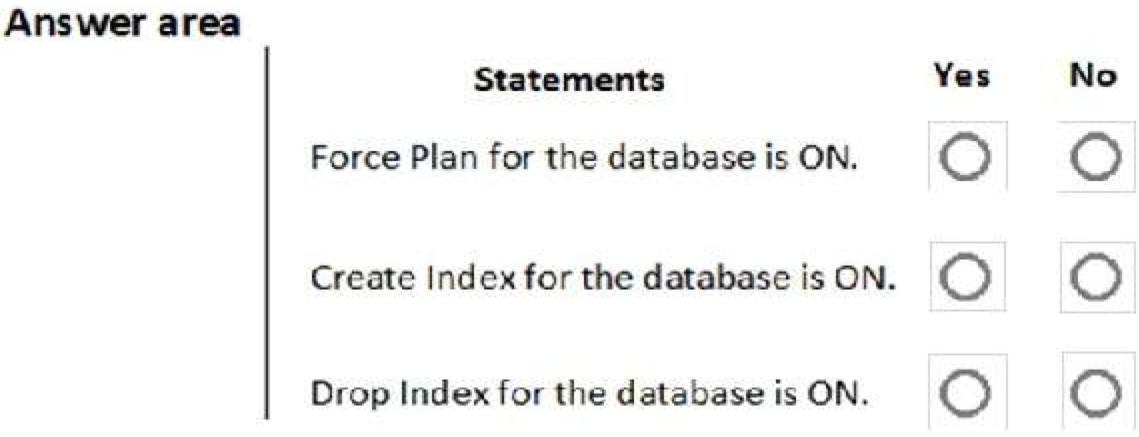


Option	Server level	Database level
Force Plan	Inherited	Inherited
Create Index	Inherited	Inherited
Drop Index	Inherited	Inherited

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

NOTE: Each correct selection is worth one point.

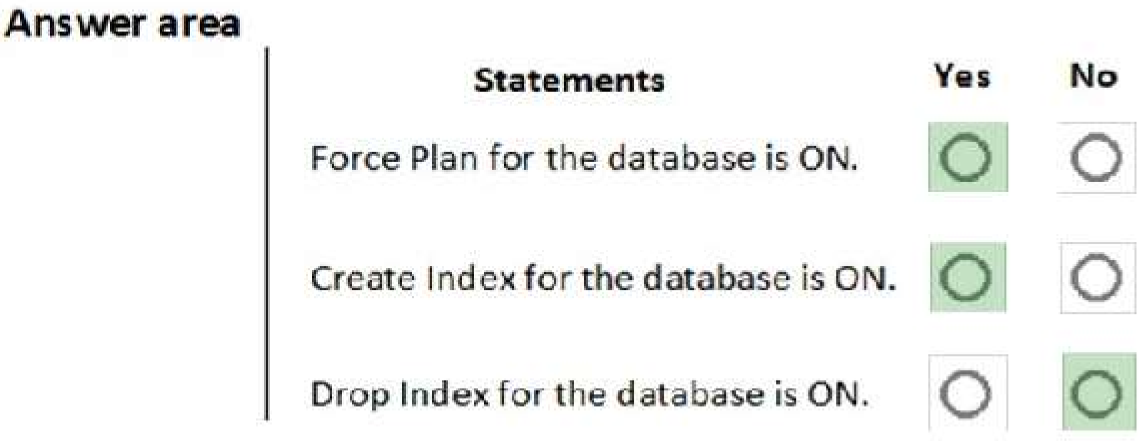
Hot Area:



Answer area

Statements	Yes	No
Force Plan for the database is ON.	○	○
Create Index for the database is ON.	○	○
Drop Index for the database is ON.	○	○

Correct Answer:



Answer area

Statements	Yes	No
Force Plan for the database is ON.	[✓ ○]	○
Create Index for the database is ON.	[✓ ○]	○
Drop Index for the database is ON.	○	[✓ ○]

Section: [none]
Explanation

Explanation/Reference:
Automatic tuning options can be independently enabled or disabled per database, or they can be configured on SQL Database servers and applied on every database that inherits settings from the server. SQL Database

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.

References:
https://docs.microsoft.com/en-us/azure/sql-database/sql-database-automatic-tuning

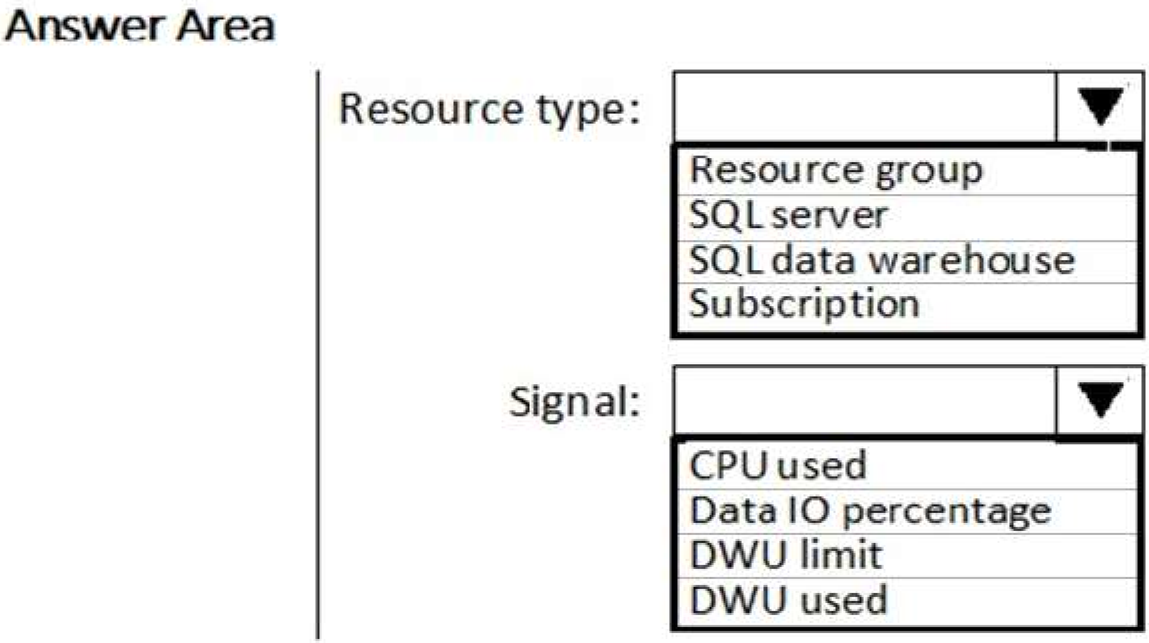
QUESTION 19
HOTSPOT

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? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

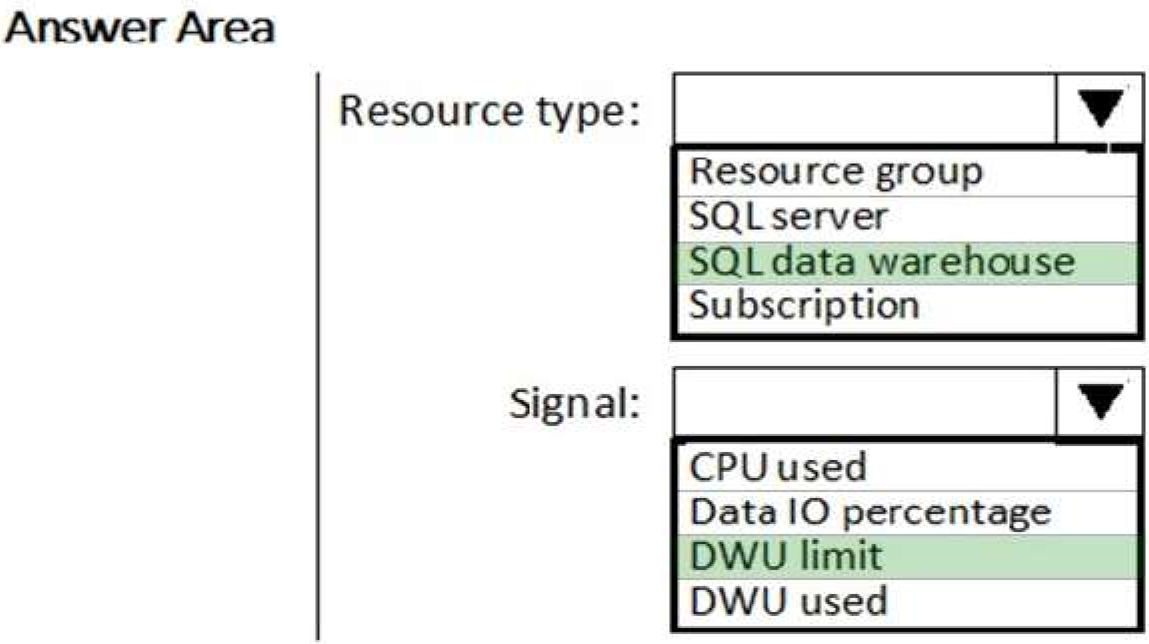


Answer Area

Resource type:
- Resource group
- SQL server
- SQL data warehouse
- Subscription

Signal:
- CPU used
- Data IO percentage
- DWU limit
- DWU used

Correct Answer:



Answer Area

Resource type:

- Resource group
- SQL server
- SQL data warehouse
- Subscription

Signal:

- CPU used
- Data IO percentage
- DWU limit
- DWU used

Section: [none]
Explanation

Explanation/Reference:
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.

References:
https://docs.microsoft.com/en-us/azure/sql-database/sql-database-insights-alerts-portal

QUESTION 20
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?

A. Azure Advanced Threat Protection (ATP)
B. custom masking rules
C. Transparent Data Encryption (TDE)
D. auditing

Correct Answer: D
Section: [none]
Explanation

Explanation/Reference:
Explanation:
Dynamic Data Masking is designed to simplify application development by limiting data exposure in a set of pre-defined queries used by the application. While Dynamic Data Masking can also be useful to prevent accidental

exposure of sensitive data when accessing a production database directly, it is important to note that unprivileged users with ad-hoc query permissions can apply techniques to gain access to the actual data. If there is a need to grant such ad-hoc access, Auditing should be used to monitor all database activity and mitigate this scenario.

References:
https://docs.microsoft.com/en-us/sql/relational-databases/security/dynamic-data-masking

Testlet 2

Background
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.

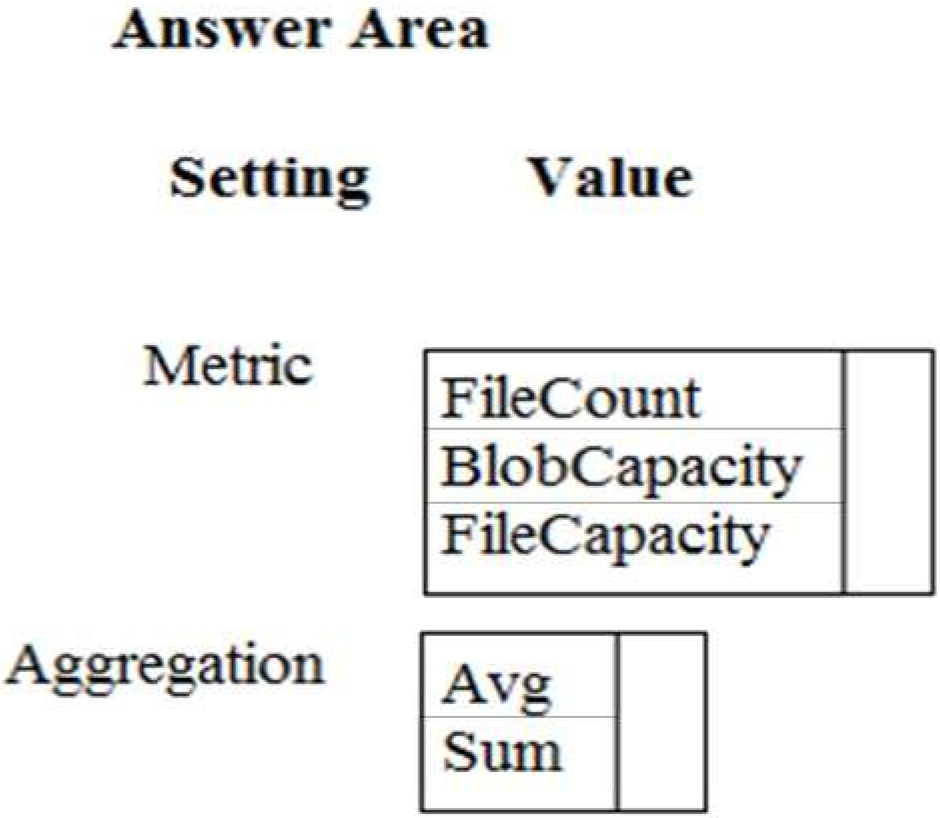
Monitoring
All Azure services must be monitored by using Azure Monitor. On-premises SQL Server performance must be monitored.

QUESTION 1
HOTSPOT

You need to ensure phone-based polling data upload reliability requirements are met. How should you configure monitoring? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

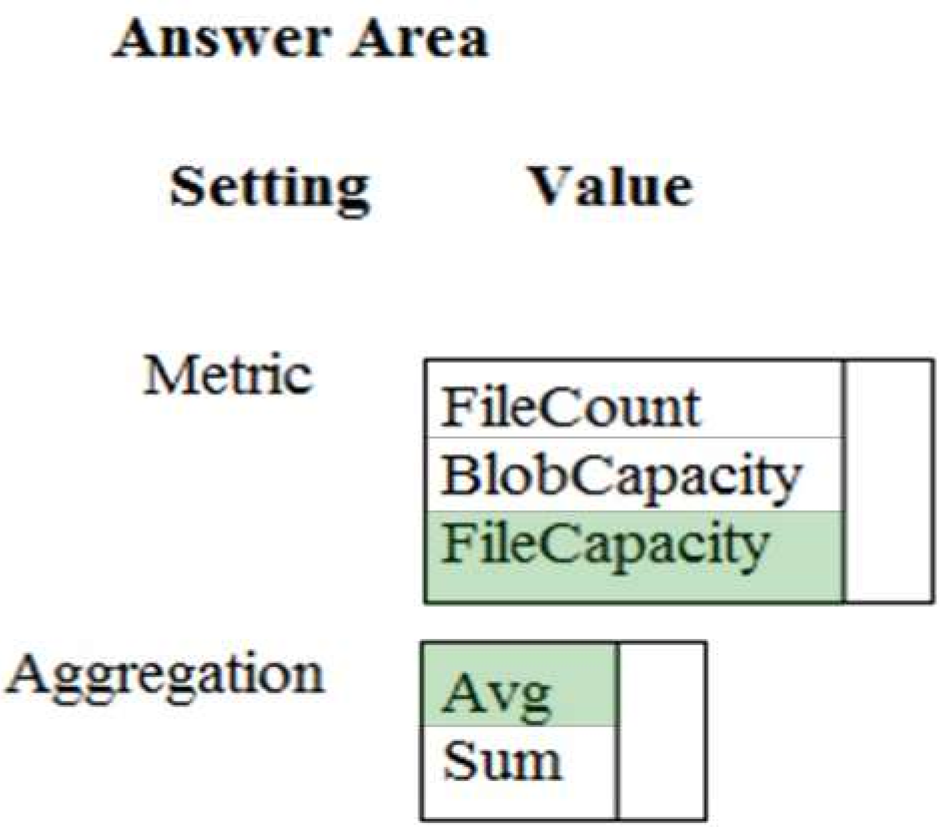
Hot Area:



Answer Area

Setting	Value
Metric	FileCount BlobCapacity FileCapacity
Aggregation	Avg Sum

Correct Answer:



Answer Area

Setting	Value
Metric	FileCount BlobCapacity 🟩 FileCapacity
Aggregation	🟩 Avg Sum

Section: [none]
Explanation

Explanation/Reference:
Explanation:

Box 1: FileCapacity
FileCapacity is the amount of storage used by the storage account’s File service in bytes.

Box 2: Avg
The aggregation type of the FileCapacity metric is Avg.

Scenario:
All services and processes must be resilient to a regional Azure outage.

All Azure services must be monitored by using Azure Monitor. On-premises SQL Server performance must be monitored.

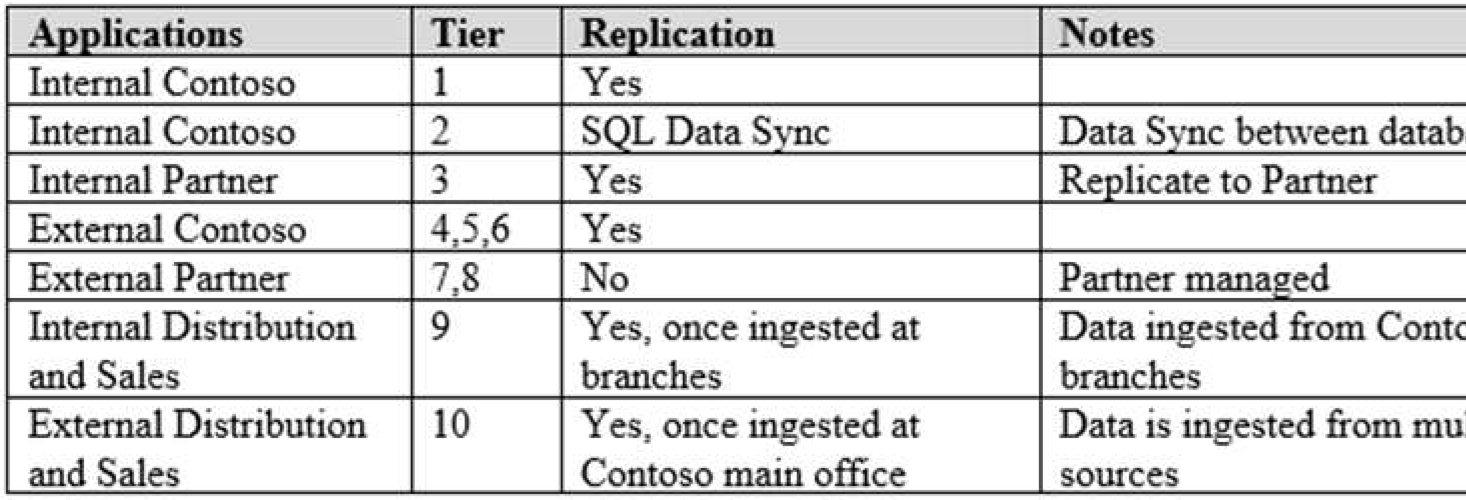
References:
https://docs.microsoft.com/en-us/azure/azure-monitor/platform/metrics-supported

Testlet 3

Overview

Current environment
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:



Applications	Tier	Replication	Notes
Internal Contoso	1	Yes	
Internal Contoso	2	SQL Data Sync	Data Sync between datab
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 Conto branches
External Distribution and Sales	10	Yes, once ingested at Contoso main office	Data is ingested from mu 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.

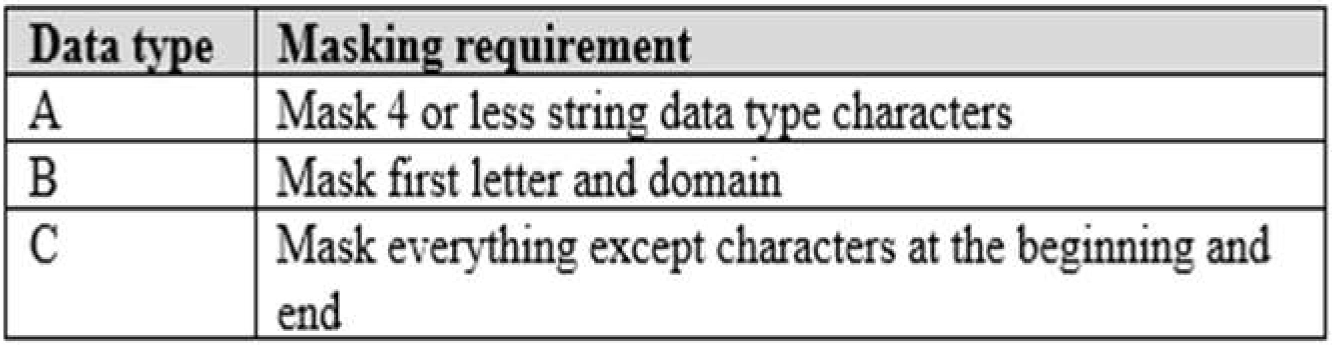
Back up strategies must be put in place for all other standalone Azure SQL Databases using Azure SQL-

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:



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

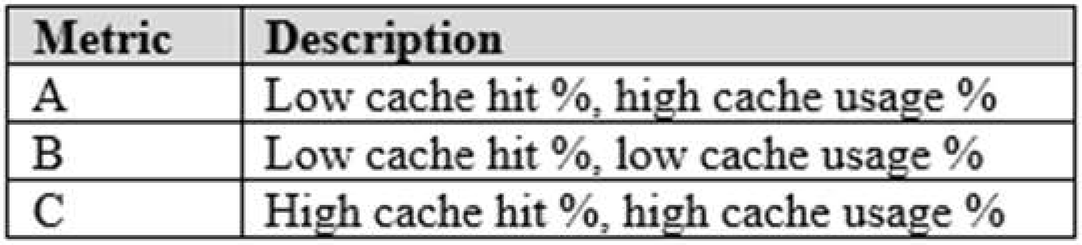
Security and monitoring

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:



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 1

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

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You need to implement diagnostic logging for Data Warehouse monitoring.

Which log should you use?

A. RequestSteps

B. DmsWorkers

C. SqlRequests

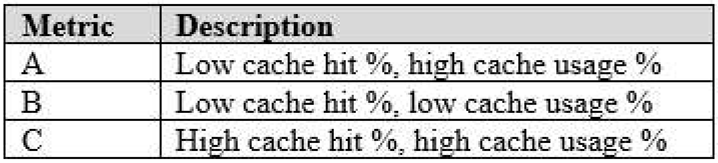
D. ExecRequests

Correct Answer: C

Section: [none]

Explanation

Explanation/Reference:
Explanation:
Scenario:
The Azure SQL Data Warehouse cache must be monitored when the database is being used.



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

References:
https://docs.microsoft.com/en-us/sql/relational-databases/system-dynamic-management-views/sys-dm-pdw-sql-requests-transact-sq

QUESTION 2
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some questions sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You need setup monitoring for tiers 6 through 8.

What should you configure?

A. extended events for average storage percentage that emails data engineers
B. an alert rule to monitor CPU percentage in databases that emails data engineers
C. an alert rule to monitor CPU percentage in elastic pools that emails data engineers
D. an alert rule to monitor storage percentage in databases that emails data engineers
E. an alert rule to monitor storage percentage in elastic pools that emails data engineers

Correct Answer: E
Section: [none]
Explanation

Explanation/Reference:
Explanation:
Scenario:
Tiers 6 through 8 must have unexpected resource storage usage immediately reported to data engineers.

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.

Testlet 4

Case Study

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

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

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

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

Overview

General Overview

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 1
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?

A. Decrease parallelism for the copy activities.
B. Increase that data integration units.
C. Configure the copy activities to use staged copy.
D. Configure the copy activities to perform compression.

Correct Answer: B
Section: [none]
Explanation

Explanation/Reference:
Explanation:

Performance tuning tips and optimization features. In some cases, when you run a copy activity in Azure Data Factory, you see a "Performance tuning tips" message on top of the copy activity monitoring, as shown in the following example. The message tells you the bottleneck that was identified for the given copy run. It also guides you on what to change to boost copy throughput. The performance tuning tips currently provide suggestions like:

- Use PolyBase when you copy data into Azure SQL Data Warehouse.
- Increase Azure Cosmos DB Request Units or Azure SQL Database DTUs (Database Throughput Units) when the resource on the data store side is the bottleneck.
- Remove the unnecessary staged copy.

References:
https://docs.microsoft.com/en-us/azure/data-factory/copy-activity-performance

QUESTION 2
What should you implement to optimize SQL Database for Race Central to meet the technical requirements?

A. the `sp_update` stored procedure
B. automatic tuning
C. Query Store
D. the `dbcc checkdb` command

Correct Answer: A
Section: [none]
Explanation

Explanation/Reference:
Explanation:
Scenario: The query performance of Race Central must be stable, and the administrative time it takes to perform optimizations must be minimized.

`sp_update` 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.

Incorrect Answers:
D: `dbcc checkdc`hecks the logical and physical integrity of all the objects in the specified database

QUESTION 3
Which two metrics should you use to identify the appropriate RU/s for the telemetry data? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

A. Number of requests
B. Number of requests exceeded capacity
C. End to end observed read latency at the 99th percentile
D. Session consistency
E. Data + Index storage consumed
F. Avg Troughput/s

Correct Answer: AE
Section: [none]
Explanation

Explanation/Reference:
Explanation:

Scenario: 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.

With Azure Cosmos DB, you pay for the throughput you provision and the storage you consume on an hourly basis.

While you estimate the number of RUs per second to provision, consider the following factors:

Item size: As the size of an item increases, the number of RUs consumed to read or write the item also increases.

Question Set 1

QUESTION 1
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 present part of the solution.

NOTE: Each correct selection is worth one point.

A. Implement event ordering
B. Scale the SU count for the job up
C. Implement Azure Stream Analytics user-defined functions (UDF)
D. Scale the SU count for the job down
E. Implement query parallelization by partitioning the data output
F. Implement query parallelization by partitioning the data input

Correct Answer: BF
Section: [none]
Explanation

Explanation/Reference:
Explanation:
Scale out the query by allowing the system to process each input partition separately.

F: A Stream Analytics job definition includes inputs, a query, and output. Inputs are where the job reads the data stream from.

References:
https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-parallelization

QUESTION 2
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.

NOTE: Each correct selection is worth one point.

A. Combine the daily log files for all servers into one file
B. Increase the value of the `mapreduce.map.memory` parameter
C. Move the log files into folders so that each day's logs are in their own folder
D. Increase the number of worker nodes
E. Increase the value of the `hive.tez.container.size` parameter

Correct Answer: AC
Section: [none]
Explanation

Explanation/Reference:
Explanation:

A: 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.

C: 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.

References:
https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-performance-tuning-guidance

QUESTION 3
DRAG DROP

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:

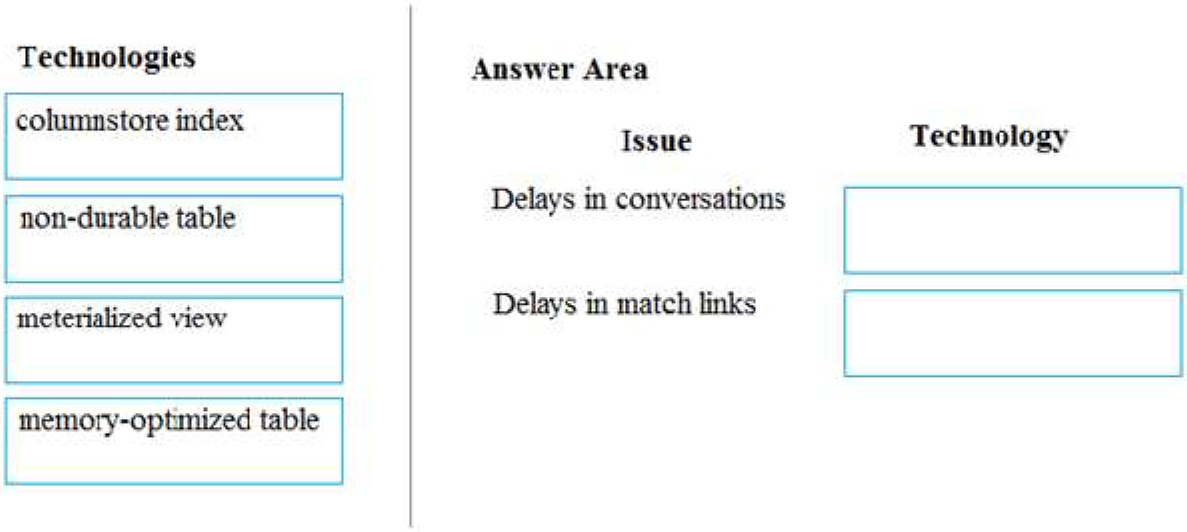
- Delays occur during live conversations
- A delay occurs before matching links appear after code snippets are added to conversations

You need to resolve the performance issues.

Which technologies should you use? To answer, drag the appropriate technologies to the correct issues. Each technology may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

Select and Place:



Technologies

columnstore index

non-durable table

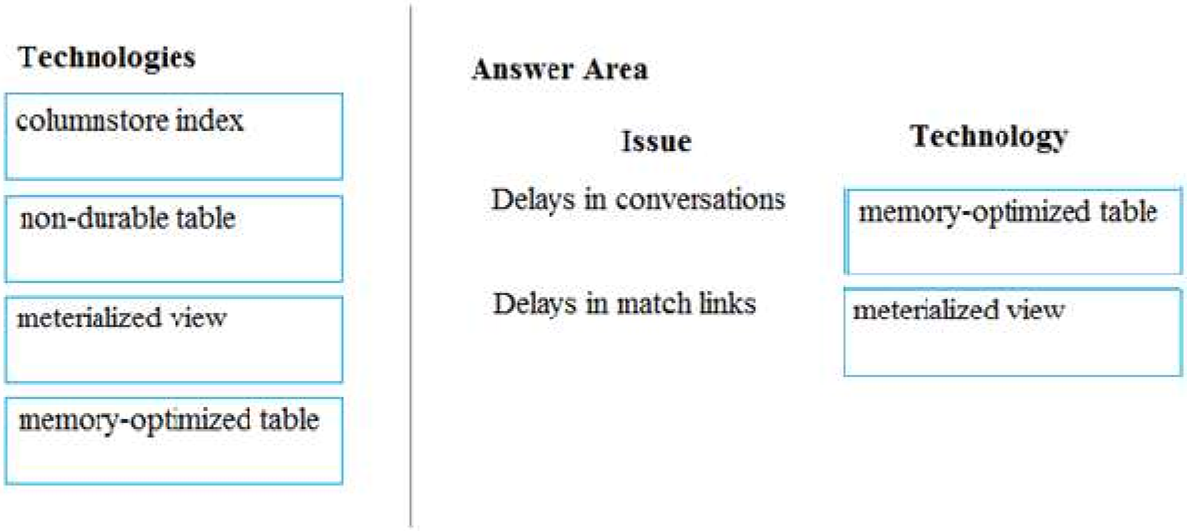
meterialized view

memory-optimized table

Answer Area

Issue	Technology
Delays in conversations	
Delays in match links	

Correct Answer:



Technologies

columnstore index

non-durable table

meterialized view

memory-optimized table

Answer Area

Issue	Technology
Delays in conversations	memory-optimized table
Delays in match links	meterialized view

Section: [none]
Explanation

Explanation/Reference:
Explanation:

Box 1: memory-optimized table
In-Memory OLTP can provide great performance benefits for transaction processing, data ingestion, and transient data scenarios.

Box 2: materialized view
To support efficient querying, a common solution is to generate, in advance, a view that materializes the data in a format suited to the required results set. The Materialized View pattern describes generating prepopulated views of data in environments where the source data isn't in a suitable format for querying, where generating a suitable query is difficult, or where query performance is poor due to the nature of the data or the data store.

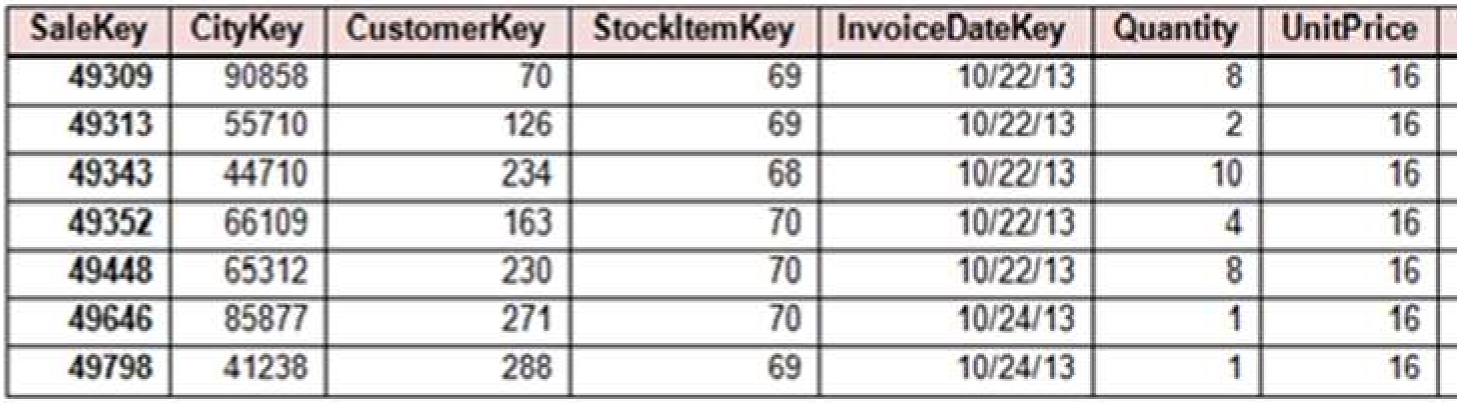
These materialized views, which only contain data required by a query, allow applications to quickly obtain the information they need. In addition to joining tables or combining data entities, materialized views can include the current values of calculated columns or data items, the results of combining values or executing transformations on the data items, and values specified as part of the query. A materialized view can even be optimized for just a single query.

References:
https://docs.microsoft.com/en-us/azure/architecture/patterns/materialized-view

QUESTION 4
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:



SaleKey	CityKey	CustomerKey	StockItemKey	InvoiceDateKey	Quantity	UnitPrice
49309	90858	70	69	10/22/13	8	16
49313	55710	126	69	10/22/13	2	16
49343	44710	234	68	10/22/13	10	16
49352	66109	163	70	10/22/13	4	16
49448	65312	230	70	10/22/13	8	16
49646	85877	271	70	10/24/13	1	16
49798	41238	288	69	10/24/13	1	16

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

Which technology should you use?

A. hash distributed table with clustered ColumnStore index
B. hash distributed table with clustered index
C. heap table with distribution replicate
D. round robin distributed table with clustered index
E. round robin distributed table with clustered ColumnStore index

Correct Answer: A
Section: [none]
Explanation

Explanation/Reference:
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.

Incorrect Answers:
D, E: Round-robin tables are useful for improving loading speed.

References:
https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-tables-distribute

https://docs.microsoft.com/en-us/sql/relational-databases/indexes/columnstore-indexes-query-performance

QUESTION 5
You manage a Microsoft Azure SQL Data Warehouse Gen 2.

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

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

Which metric should you monitor?

A. Cache used percentage
B. Local tempdb percentage
C. DWU percentage
D. CPU percentage
E. Data IO percentage

Correct Answer: A
Section: [none]
Explanation

Explanation/Reference:
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. You can monitor and troubleshoot slow query performance by determining whether your workload is optimally leveraging the Gen2 cache.

Reference:
https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-how-to-monitor-cache