

DP-200.VCEplus.premium.exam.58q

Number: DP-200
Passing Score: 800
Time Limit: 120 min
File Version: 1.0



Website: <https://vceplus.com>

VCE to PDF Converter: <https://vceplus.com/vce-to-pdf/>

Facebook: <https://www.facebook.com/VCE.For.All.VN/>

Twitter : https://twitter.com/VCE_Plus

DP-200

Implementing an Azure Data Solution (beta)



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

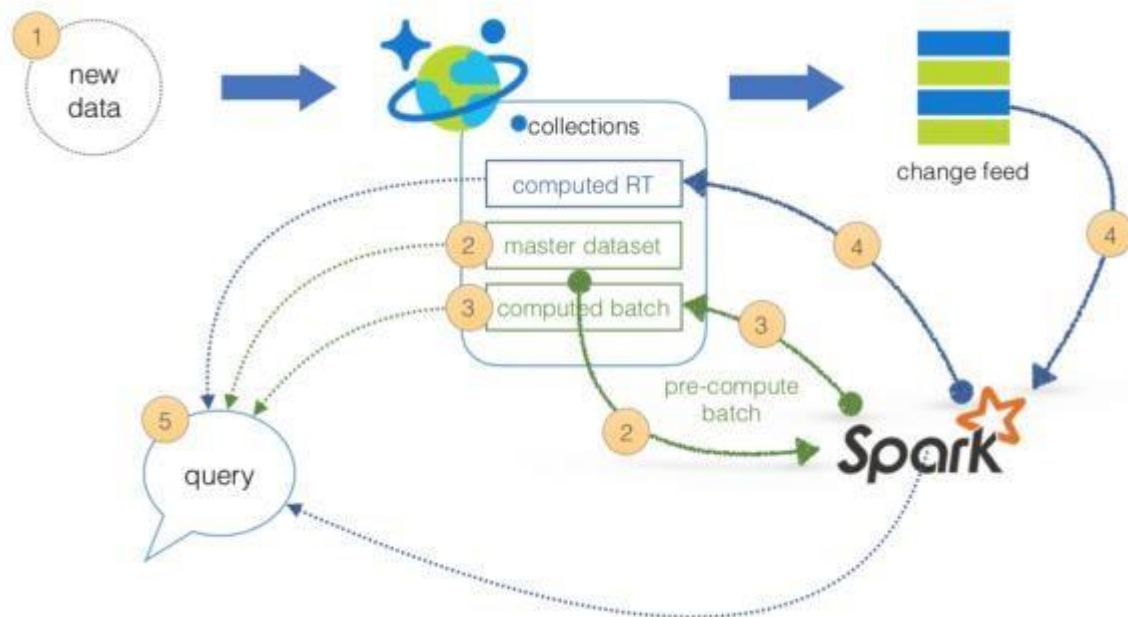
Explanation/Reference:

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:

Correct Answer:

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	<ul style="list-style-type: none"> - Active-active support - Bi-directional between on-premises and Azure SQL Database 	<ul style="list-style-type: none"> - Lower latency - Transactional consistency - Reuse existing topology after migration
Disadvantages	<ul style="list-style-type: none"> - 5 min or more latency - No transactional consistency - Higher performance impact 	<ul style="list-style-type: none"> - 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:

Correct Answer:

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 and 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

Daily Sales

Weekly Analysis

Azure SQL product

Azure SQL product

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

Daily Sales

Weekly Analysis

Azure SQL product

Azure SQL Database elastic pools

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 and restore process.

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 bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

Select and Place:

Correct Answer:

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": 30
    },
    "TierToArchive": {
      "DaysAfterModificationGreaterThan": 50
    },
    "Delete": {
      "DaysAfterModificationGreaterThan": 100
    }
  },
  "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]

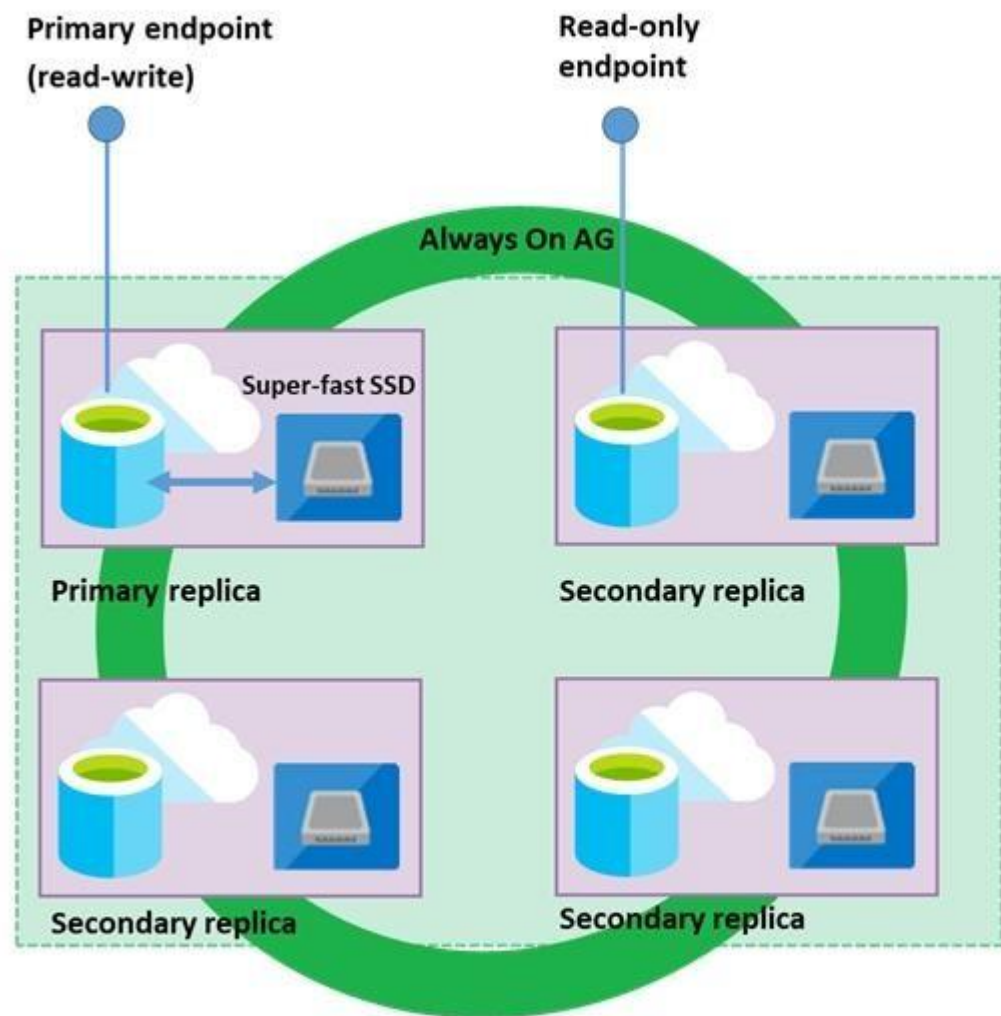
Explanation

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: colocated 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>



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 lower-cost storage solution.

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 area and arrange them in the correct order.

Select and Place:

Correct Answer:

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 Set 1

QUESTION 1

Note: This question is part of series of questions that present the same scenario. Each question in the series contain 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 contain 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 contain 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:

Correct Answer:

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:

Correct Answer:

Section: [none]
Explanation

Explanation/Reference:

Explanation:

Step 1: Create an HDInsight cluster with the Spark cluster type

Step 2: Create a Jupyter 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

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

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

Create an Azure Data Factory resource.

Configure a self-hosted 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:

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:

Correct Answer:

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>



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 lower-cost storage solution.

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 polling 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	√
DAC	
MAC	
Claims	

Authentication

Service Principal	^
Kerberos	
Certificate-based	
Bearer Token	



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 databases
Internal Partner	3	Yes	Replicate to Partner
External Contoso	4,5,6	Yes	
External Partner	7,8	No	Partner managed
Internal Distribution and Sales	9	Yes, once ingested at branches	Data ingested from Contoso branches
External Distribution and Sales	10	Yes, once ingested at Contoso main office	Data is ingested from multiple sources

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

Requirements

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

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

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

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

- Tier 1 internal applications on the premium P2 tier
- Tier 2 internal applications on the standard S4 tier

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

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

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

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

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

Databases:

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

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:

Correct Answer:

Section: [none]

Explanation

Explanation/Reference:

Explanation:

Box 1: Connection String

To use storage account key authentication, you use the ConnectionString property, which specifies the information needed to connect to Blob 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>

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:

Correct Answer:

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:

Correct Answer:

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.



ronmatwedemo - Advanced Threat Protection
SQL server

Search (Ctrl+/) « **Save** Discard Feedback

Overview
Activity log
Access control (IAM)
Tags
Diagnose and solve problems

SETTINGS

Quick start
Failover groups
Long-term backup retention
Active Directory admin
SQL databases
SQL elastic pools
Deleted databases
Import/Export history
Properties
Locks
Automation script

SECURITY

Advanced Threat Protection

Advanced Threat Protection costs 15 USD/server/month. It includes Threat Detection, Data Discovery & Classification and Vulnerability Assessment. We invite you to a trial 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 Uri(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:

Correct Answer:

Section: [none]

Explanation

Explanation/Reference:

Explanation:

Box 1: Azure Cosmos DB

The DocumentClient.CreateDatabaseAsync(Database, RequestOptions) method creates a database resource as an asynchronous 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_backupoperator).

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:

Correct Answer:

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:

Correct Answer:

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>



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 lower-cost storage solution.

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:

Correct Answer:

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 databases
Internal Partner	3	Yes	Replicate to Partner
External Contoso	4,5,6	Yes	
External Partner	7,8	No	Partner managed
Internal Distribution and Sales	9	Yes, once ingested at branches	Data ingested from Contoso branches
External Distribution and Sales	10	Yes, once ingested at Contoso main office	Data is ingested from multiple sources

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

Requirements

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

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

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

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

- Tier 1 internal applications on the premium P2 tier
- Tier 2 internal applications on the standard S4 tier

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

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

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

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

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

Databases:

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

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 at 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 from 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

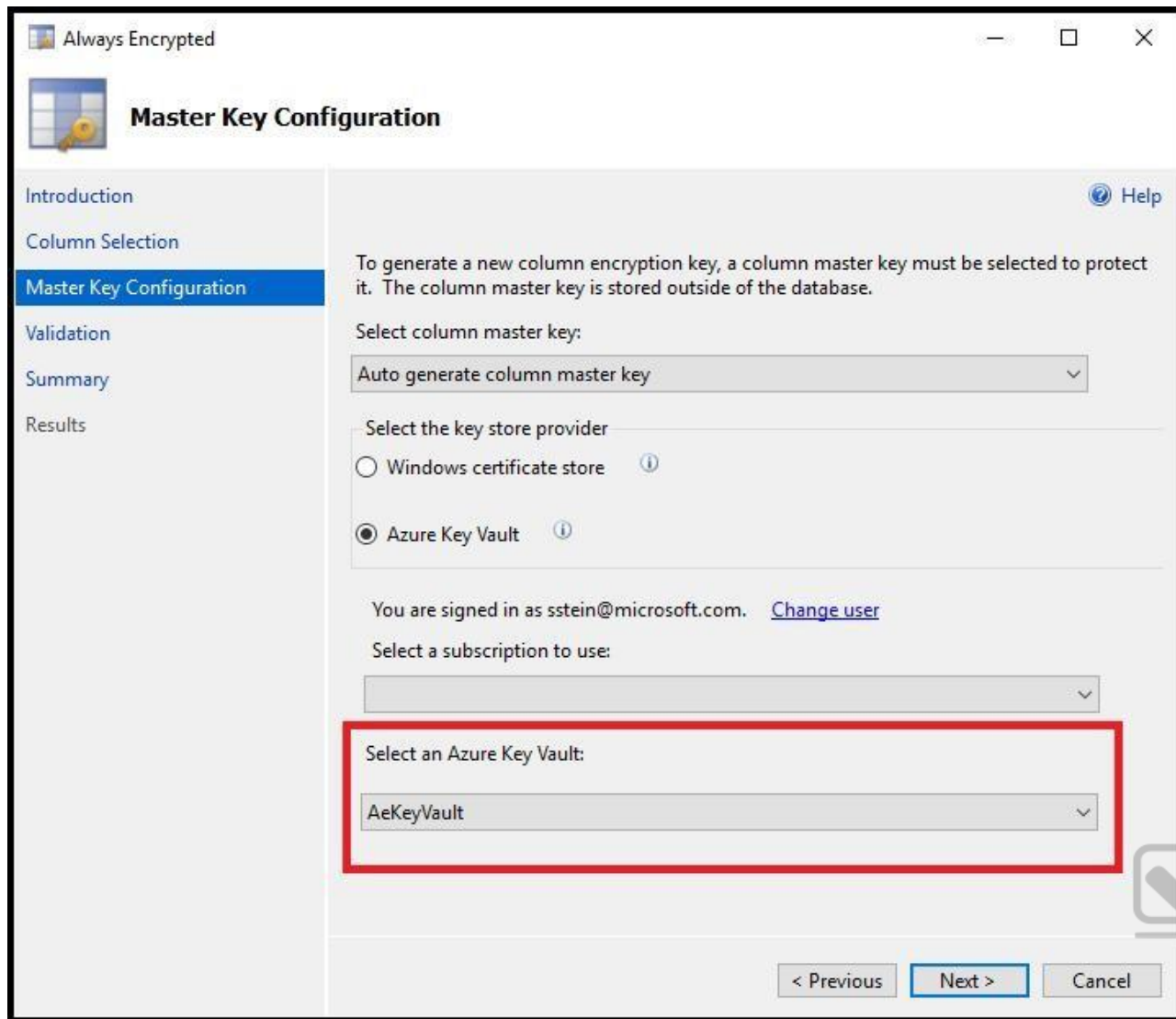
Explanation/Reference:

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).





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]

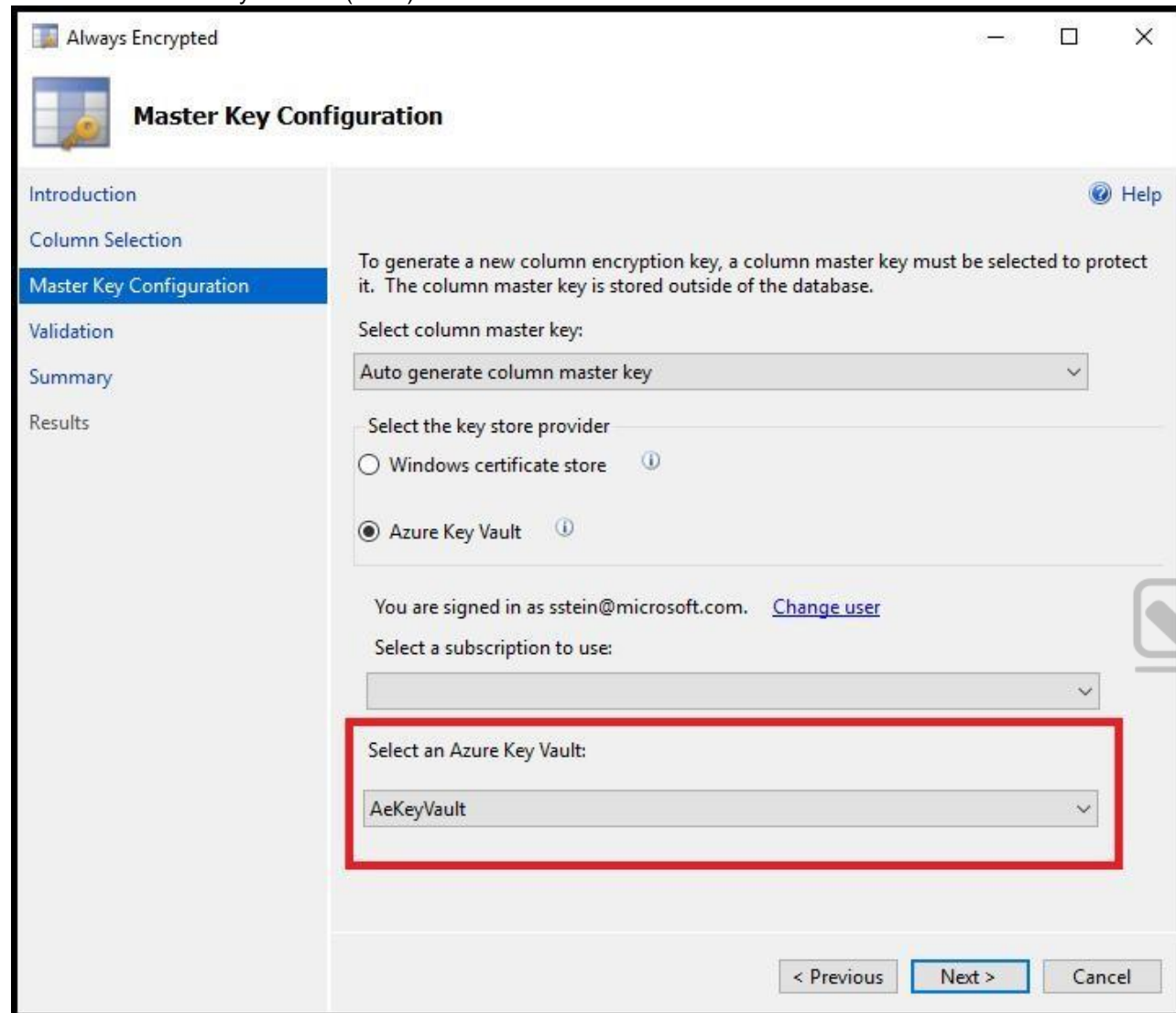
Explanation

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).



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]

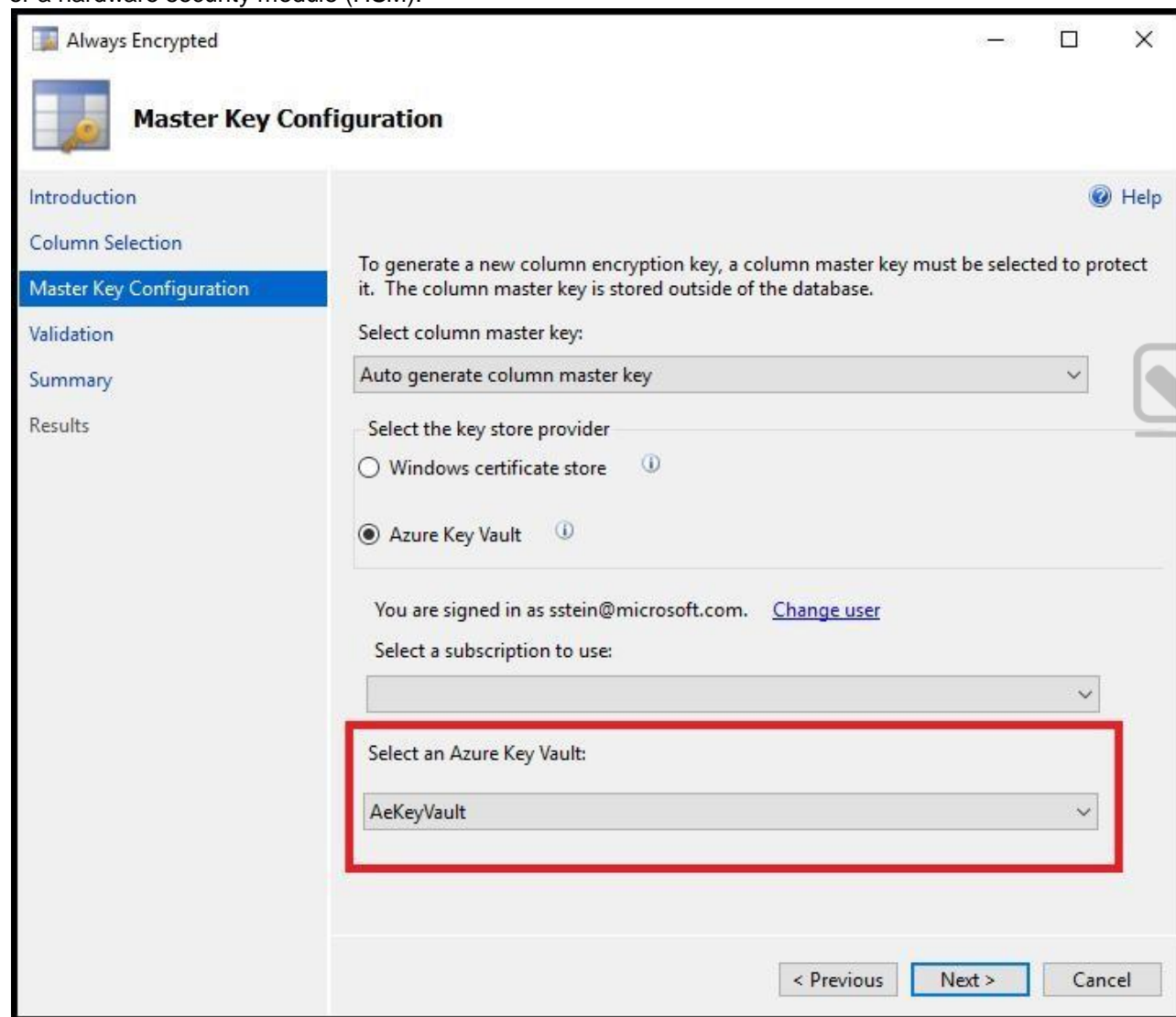
Explanation

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).



Master Key Configuration

Introduction
Column Selection
Master Key Configuration
Validation
Summary
Results

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:

< Previous Next > Cancel

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	V
default	
email	
random number	

B

custom text	V
default	
email	
random number	

C

custom text	V
default	
email	
random number	

Correct Answer:

Answer Area

Data type

Masking function

A

custom text	V
default	
email	
random number	

B

custom text	V
default	
email	
random number	

C

custom text	V
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:

Correct Answer:

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'; ▪ 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 contain 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 contain 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 contain 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:

Correct Answer:

Section: [none]

Explanation

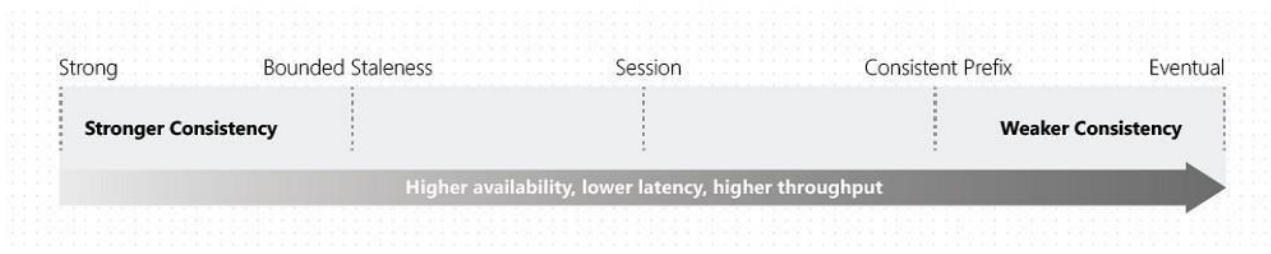
Explanation/Reference:

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 files.'
- 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 (0 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 files.'

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>



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 lower-cost storage solution.

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	<table> <tr><td>FileCount</td><td></td></tr> <tr><td>BlobCapacity</td><td></td></tr> <tr><td>FileCapacity</td><td></td></tr> </table>	FileCount		BlobCapacity		FileCapacity	
FileCount							
BlobCapacity							
FileCapacity							
Aggregation	<table> <tr><td>Avg</td><td></td></tr> <tr><td>Sum</td><td></td></tr> </table>	Avg		Sum			
Avg							
Sum							

Correct Answer:

Answer Area

Setting	Value						
Metric	<table> <tr><td>FileCount</td><td></td></tr> <tr><td>BlobCapacity</td><td></td></tr> <tr><td>FileCapacity</td><td></td></tr> </table>	FileCount		BlobCapacity		FileCapacity	
FileCount							
BlobCapacity							
FileCapacity							
Aggregation	<table> <tr><td>Avg</td><td></td></tr> <tr><td>Sum</td><td></td></tr> </table>	Avg		Sum			
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 databases
Internal Partner	3	Yes	Replicate to Partner
External Contoso	4,5,6	Yes	
External Partner	7,8	No	Partner managed
Internal Distribution and Sales	9	Yes, once ingested at branches	Data ingested from Contoso branches
External Distribution and Sales	10	Yes, once ingested at Contoso main office	Data is ingested from multiple sources

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

Requirements

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

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

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

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

- Tier 1 internal applications on the premium P2 tier
- Tier 2 internal applications on the standard S4 tier

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

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

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

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

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

Databases:

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

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 at 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 from 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.



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:

Correct Answer:

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>

