

DP-200 Dumps

Implementing an Azure Data Solution

<https://www.certleader.com/DP-200-dumps.html>



NEW QUESTION 1

- (Exam Topic 1)

You need to provision the polling data storage account.

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

NOTE: Each correct selection is worth one point.

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

NEW QUESTION 2

- (Exam Topic 1)

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.

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

- A. Mastered
- B. Not Mastered

Answer: A

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>

NEW QUESTION 3

- (Exam Topic 1)

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.

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							

- A. Mastered
- B. Not Mastered

Answer: A

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>

NEW QUESTION 4

- (Exam Topic 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 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

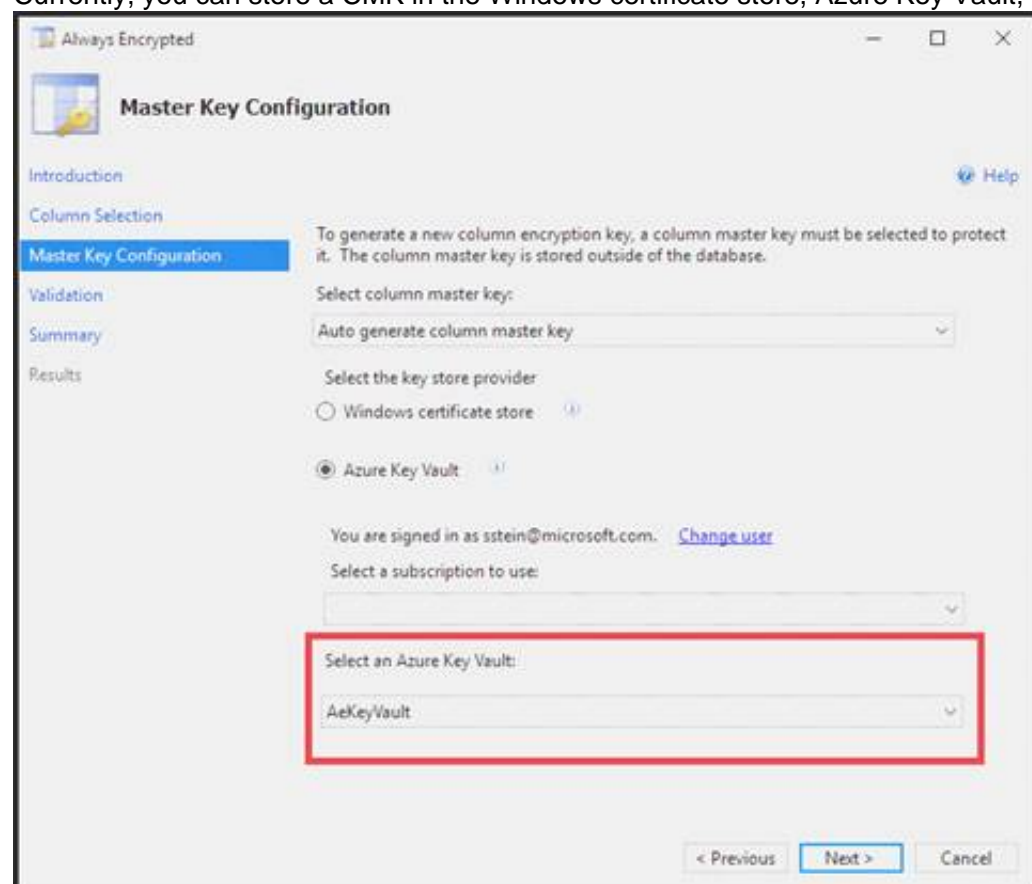
Answer: B

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>

NEW QUESTION 5

- (Exam Topic 2)

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

Answer: EF

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>

NEW QUESTION 6

- (Exam Topic 2)

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.

Data factory component	Value
Connector	<div> <div>connection string</div> <div>linked service name string</div> <div>gateway connection string</div> <div>data store name string</div> </div>
Data movement activity	<div> <div>Azure SQL Data Warehouse</div> <div>Azure Files</div> <div>Azure Blob</div> <div>Azure SQL Database</div> </div>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Connection String

To use storage account key authentication, you use the ConnectionString property, which xpecify the information needed to connect to Blobl Storage.

Mark this field as a SecureString to store it securely in Data Factory. You can also put account key in Azure Key Vault and pull the accountKey configuration out of the connection string.

Box 2: Azure Blob

Tier 10 reporting data must be stored in Azure Blobs

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

References:

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

NEW QUESTION 7

- (Exam Topic 2)

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.

Actions	Answer Area
Connect to the Database and use Azure PowerShell to create a database firewall rule	
Set the Allow Azure Services to Access Server to Disabled	
In ther Azure portal, create a database firewall rule	
In the Azure portal, create a server firewall rule	
Connect to the database and use Transact-SQL to create a database firewall rule	
Set the Allow Azure Services to Access Server setting to Enabled	

- A. Mastered
B. Not Mastered

Answer: A

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:

- In Azure portal, select SQL databases from the left-hand menu, and select your database on the SQL databases page.
- 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:

- In Object Explorer, right-click the database and select New Query.
- 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>

NEW QUESTION 8

- (Exam Topic 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 Azure Key Vault
5. Validate configuration results and deploy the solution Does the solution meet the goal?

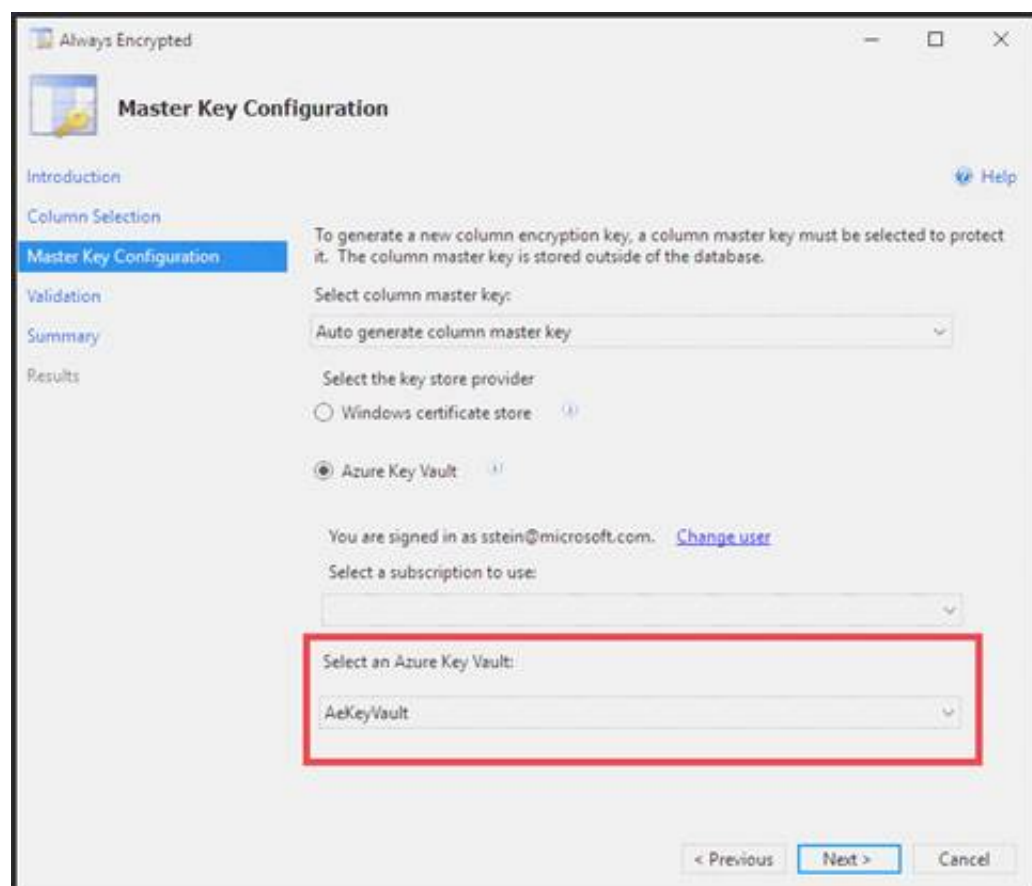
- A. Yes
B. No

Answer: A

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).



References:

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

NEW QUESTION 9

- (Exam Topic 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

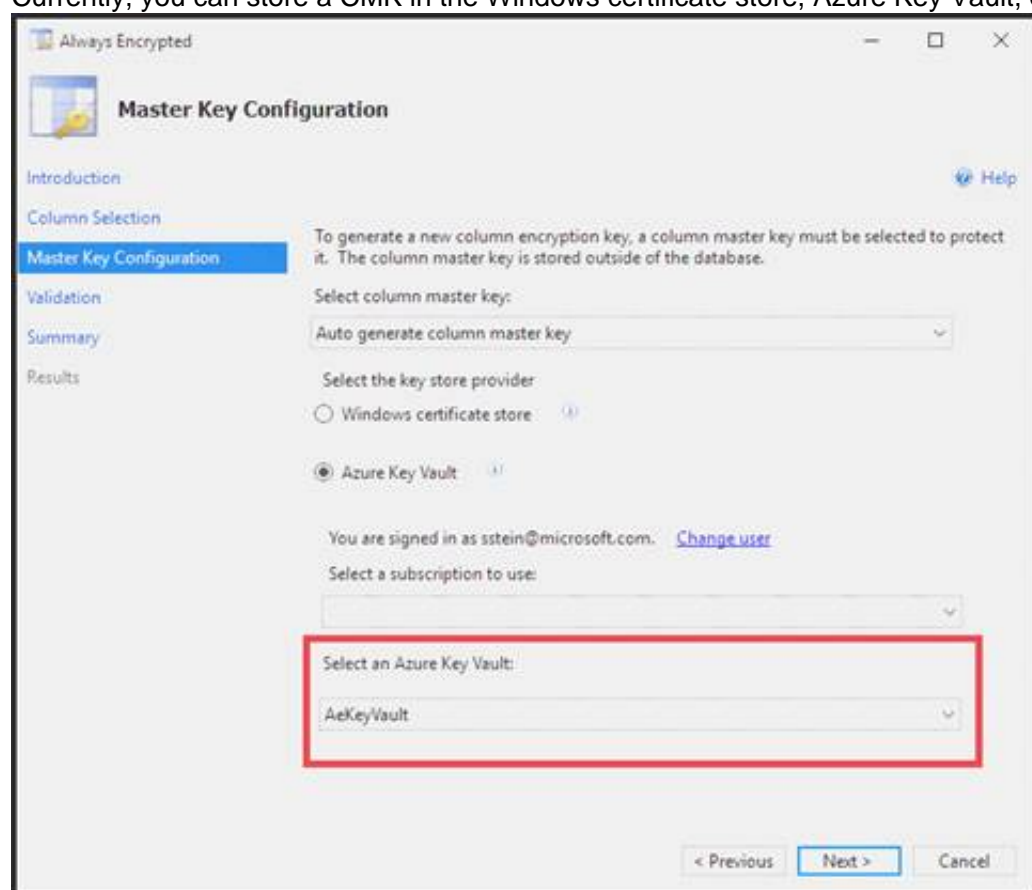
B. No

Answer: B

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>

NEW QUESTION 10

- (Exam Topic 2)

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.

Data type	Masking function					
A	<table border="1"> <tr><td>custom text</td><td rowspan="4">V</td></tr> <tr><td>default</td></tr> <tr><td>email</td></tr> <tr><td>random number</td></tr> </table>	custom text	V	default	email	random number
custom text	V					
default						
email						
random number						
B	<table border="1"> <tr><td>custom text</td><td rowspan="4">V</td></tr> <tr><td>default</td></tr> <tr><td>email</td></tr> <tr><td>random number</td></tr> </table>	custom text	V	default	email	random number
custom text	V					
default						
email						
random number						
C	<table border="1"> <tr><td>custom text</td><td rowspan="4">V</td></tr> <tr><td>default</td></tr> <tr><td>email</td></tr> <tr><td>random number</td></tr> </table>	custom text	V	default	email	random number
custom text	V					
default						
email						
random number						

- A. Mastered
B. Not Mastered

Answer: A

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>

NEW QUESTION 10

- (Exam Topic 3)

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.

Architecture requirement	Answer Area					
Ingestion	<table border="1"> <tr><td>HDInsight Kafka</td><td rowspan="4">V</td></tr> <tr><td>Azure Event Hubs</td></tr> <tr><td>HDInsight Storm</td></tr> <tr><td>HDInsight Spark</td></tr> </table>	HDInsight Kafka	V	Azure Event Hubs	HDInsight Storm	HDInsight Spark
HDInsight Kafka	V					
Azure Event Hubs						
HDInsight Storm						
HDInsight Spark						
Stream Processing	<table border="1"> <tr><td>Azure Stream Analytics</td><td rowspan="4">V</td></tr> <tr><td>HDInsight with Spark Streaming</td></tr> <tr><td>Azure Cosmos DB Change Feed</td></tr> <tr><td>Azure Analysis Services</td></tr> </table>	Azure Stream Analytics	V	HDInsight with Spark Streaming	Azure Cosmos DB Change Feed	Azure Analysis Services
Azure Stream Analytics	V					
HDInsight with Spark Streaming						
Azure Cosmos DB Change Feed						
Azure Analysis Services						
Analytical Data Store	<table border="1"> <tr><td>Hive LLAP on HDInsight</td><td rowspan="4">V</td></tr> <tr><td>Azure Analysis Services</td></tr> <tr><td>Azure Cosmos DB</td></tr> <tr><td>SQL Data Warehouse</td></tr> </table>	Hive LLAP on HDInsight	V	Azure Analysis Services	Azure Cosmos DB	SQL Data Warehouse
Hive LLAP on HDInsight	V					
Azure Analysis Services						
Azure Cosmos DB						
SQL Data Warehouse						

- A. Mastered
B. Not Mastered

Answer: A

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.

NEW QUESTION 11

- (Exam Topic 3)

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

Answer: B

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>

NEW QUESTION 12

- (Exam Topic 3)

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

Answer: E

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/>

NEW QUESTION 14

- (Exam Topic 3)

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.

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

- A. Mastered
B. Not Mastered

Answer: A

Explanation:

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

NEW QUESTION 18

- (Exam Topic 3)

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.

Azure SQL products	Database	Azure SQL product
Azure SQL Database elastic pools	Daily Sales	Azure SQL product
Azure SQL Database Premium	Weekly Analysis	Azure SQL product
Azure SQL Database Managed Instance		
Azure SQL Database Hyperscale		

- A. Mastered
B. Not Mastered

Answer: A

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.

NEW QUESTION 23

- (Exam Topic 3)

Your company has on-premises Microsoft SQL Server instance.

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

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

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

Actions

- Configure a linked service to connect to the SQL Server instance.
- From the on-premises network, install and configure a self-hosted integration runtime.
- From the SQL Server, backup the database and then copy the database to Azure Blob storage.
- Deploy an Azure Data Factory
- From the SQL Server, create a database master key

Answer Area

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Actions

- Configure a linked service to connect to the SQL Server instance.
- From the on-premises network, install and configure a self-hosted integration runtime.
- From the SQL Server, backup the database and then copy the database to Azure Blob storage.
- Deploy an Azure Data Factory
- From the SQL Server, create a database master key

Answer Area

- Deploy an Azure Data Factory
- Configure a linked service to connect to the SQL Server instance.
- From the on-premises network, install and configure a self-hosted integration runtime.

NEW QUESTION 26

- (Exam Topic 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 question sets might have more than one correct solution, while others might not have a correct solution.

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

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

Solution: Configure Azure Data Lake Storage diagnostics to store logs and metrics in a storage account. Does the solution meet the goal?

- A. Yes
- B. No

Answer: A

NEW QUESTION 28

- (Exam Topic 3)

A company has a real-time data analysis solution that is hosted on Microsoft Azure the solution uses Azure Event Hub to ingest data and an Azure Stream Analytics cloud job to analyze the data. The cloud job is configured to use 120 Streaming Units (SU).

You need to optimize performance for the Azure Stream Analytics job.

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

Answer: BF

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>

NEW QUESTION 33

- (Exam Topic 3)

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.

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

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

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).

NEW QUESTION 37

- (Exam Topic 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: Monitor clusters by using Azure Log Analytics and HDInsight cluster management solutions. Does the solution meet the goal?

- A. Yes
- B. No

Answer: A

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.

NEW QUESTION 38

- (Exam Topic 3)

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.

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

- A. Mastered
- B. Not Mastered

Answer: A

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>

NEW QUESTION 43

- (Exam Topic 3)

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

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

- A. HDInsight Spark Cluster
- B. Azure Stream Analytics
- C. HDInsight Hadoop Cluster
- D. Azure SQL Data Warehouse

Answer: B

NEW QUESTION 45

- (Exam Topic 3)

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

Answer: C

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>

NEW QUESTION 50

- (Exam Topic 3)

You develop data engineering solutions for a company.

A project requires the deployment of data to Azure Data Lake Storage.

You need to implement role-based access control (RBAC) so that project members can manage the Azure Data Lake Storage resources. Which three actions should you perform? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

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

Answer: ADE

NEW QUESTION 54

- (Exam Topic 3)

You manage a process that performs analysis of daily web traffic logs on an HDInsight cluster. Each of 250 web servers generates approximately gigabytes (GB) 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

Answer: AC

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>

NEW QUESTION 56

- (Exam Topic 3)

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

Answer: C

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>

NEW QUESTION 58

- (Exam Topic 3)

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

Answer: E

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>

NEW QUESTION 61

- (Exam Topic 3)

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

Answer: AB

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>

NEW QUESTION 62

- (Exam Topic 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. 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

Answer: B

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>

NEW QUESTION 66

- (Exam Topic 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

Answer: B

Explanation:

Reference:

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>

NEW QUESTION 70

- (Exam Topic 3)

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

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

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

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

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

NEW QUESTION 74

- (Exam Topic 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 question sets might have more than one correct solution, while others might not have a correct solution.

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

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

Solution: Use information stored in Azure Active Directory reports.

Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

NEW QUESTION 77

- (Exam Topic 3)

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

Answer: BE

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>

NEW QUESTION 81

- (Exam Topic 3)

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

Answer: E

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>

NEW QUESTION 85

- (Exam Topic 3)

You implement 3 Azure SQL Data Warehouse instance.

You plan to migrate the largest fact table to Azure SQL Data Warehouse The table resides on Microsoft SQL Server on-premises and is 10 terabytes (TB) in size.

Incoming queries use the primary key SaleKey column to retrieve data as displayed in the following table:

SaleKey	CityKey	CustomerKey	StockItemKey	InvoiceDateKey	Quantity	UnitPrice	TotalExcludingTax
49309	90858	70	89	10/22/13	8	16	128
49313	55710	126	69	10/22/13	2	16	32
49343	44710	234	88	10/22/13	10	16	160
49352	66109	163	70	10/22/13	4	16	64
49448	65312	230	70	10/22/13	8	16	128
49546	85877	271	70	10/24/13	1	16	16
49798	41238	288	89	10/24/13	1	16	16

You need to distribute the fact table across multiple nodes to optimize performance of the table. Which technology should you use?

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

Answer: A

NEW QUESTION 87

- (Exam Topic 3)

A company has a SaaS solutions that will use Azure SQL Database with elastic pools. The solution will have a dedicated database for each customer organization. Customer organizations have peak usage at different periods during the year.

Which two factors affect your costs when sizing the Azure SQL Database elastic pools? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

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

Answer: AC

NEW QUESTION 91

- (Exam Topic 3)

You manage a Microsoft Azure SQL Data Warehouse Gen 2.

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

You need to monitor resource utilization to determine the source of the performance issues. Which metric should you monitor?

- A. Cache used percentage
- B. Local tempdb percentage
- C. WU percentage
- D. CPU percentage

Answer: B

NEW QUESTION 92

- (Exam Topic 3)

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.

Requirement	Node				
Provide data access to clients.	<table border="1"> <tr> <td>DataNode</td> <td><input type="checkbox"/></td> </tr> <tr> <td>NameNode</td> <td><input type="checkbox"/></td> </tr> </table>	DataNode	<input type="checkbox"/>	NameNode	<input type="checkbox"/>
DataNode	<input type="checkbox"/>				
NameNode	<input type="checkbox"/>				
Run operations on files and directories of the file system.	<table border="1"> <tr> <td>DataNode</td> <td><input type="checkbox"/></td> </tr> <tr> <td>NameNode</td> <td><input type="checkbox"/></td> </tr> </table>	DataNode	<input type="checkbox"/>	NameNode	<input type="checkbox"/>
DataNode	<input type="checkbox"/>				
NameNode	<input type="checkbox"/>				
Perform block creation, deletion, and replication.	<table border="1"> <tr> <td>DataNode</td> <td><input type="checkbox"/></td> </tr> <tr> <td>NameNode</td> <td><input type="checkbox"/></td> </tr> </table>	DataNode	<input type="checkbox"/>	NameNode	<input type="checkbox"/>
DataNode	<input type="checkbox"/>				
NameNode	<input type="checkbox"/>				

- A. Mastered
- B. Not Mastered

Answer: A

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

NEW QUESTION 97

- (Exam Topic 3)

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 lime 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

Answer: B

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/>

NEW QUESTION 102

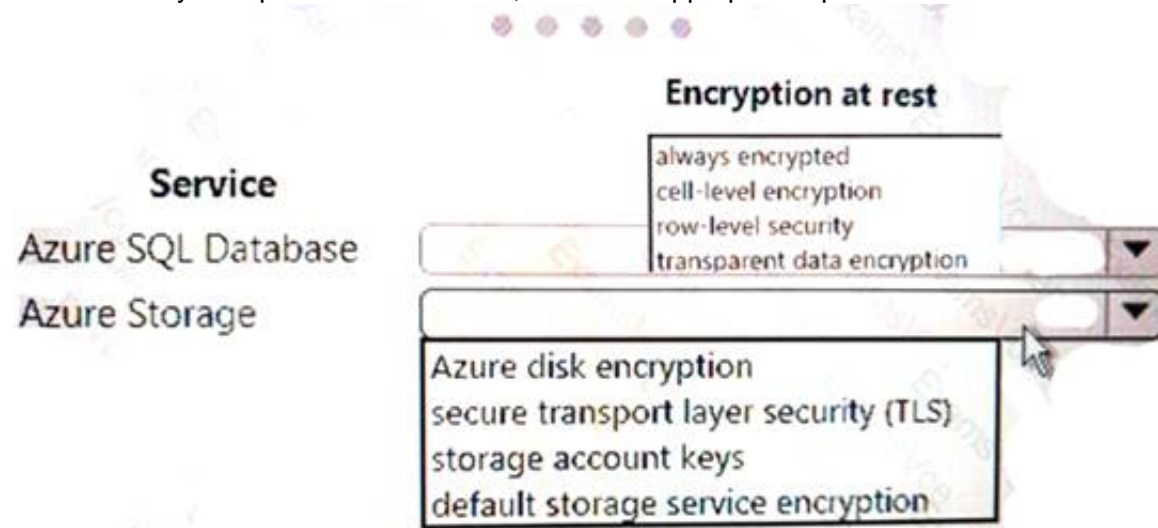
- (Exam Topic 3)

Your company uses Azure SQL Database and Azure Blob storage.

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

You need to configure security.

What should you implement? To answer, select the appropriate option in the answer area. NOTE: Each correct selection is worth one point.



- A. Mastered
- B. Not Mastered

Answer: A

Explanation:



NEW QUESTION 105

.....

Thank You for Trying Our Product

* 100% Pass or Money Back

All our products come with a 90-day Money Back Guarantee.

* One year free update

You can enjoy free update one year. 24x7 online support.

* Trusted by Millions

We currently serve more than 30,000,000 customers.

* Shop Securely

All transactions are protected by VeriSign!

100% Pass Your DP-200 Exam with Our Prep Materials Via below:

<https://www.certleader.com/DP-200-dumps.html>