

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 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

Answer: C

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>

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 <input type="checkbox"/>
	Domain Active Directory user <input type="checkbox"/>
	Managed Identity <input type="checkbox"/>
PolyBase	Database scoped credential <input type="checkbox"/>
	Database encryption key <input type="checkbox"/>
	Application role <input type="checkbox"/>

- 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 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 implement diagnostic logging for Data Warehouse monitoring. Which log should you use?

- A. RequestSteps
- B. DmsWorkers
- C. SqlRequests
- D. ExecRequests

Answer: C

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-r>

NEW QUESTION 4

- (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	<table border="1"> <tr><td>connection string</td><td><input checked="" type="checkbox"/></td></tr> <tr><td>linked service name string</td><td><input type="checkbox"/></td></tr> <tr><td>gateway connection string</td><td><input type="checkbox"/></td></tr> <tr><td>data store name string</td><td><input type="checkbox"/></td></tr> </table>	connection string	<input checked="" type="checkbox"/>	linked service name string	<input type="checkbox"/>	gateway connection string	<input type="checkbox"/>	data store name string	<input type="checkbox"/>
connection string	<input checked="" type="checkbox"/>								
linked service name string	<input type="checkbox"/>								
gateway connection string	<input type="checkbox"/>								
data store name string	<input type="checkbox"/>								
Data movement activity	<table border="1"> <tr><td>Azure SQL Data Warehouse</td><td><input checked="" type="checkbox"/></td></tr> <tr><td>Azure Files</td><td><input type="checkbox"/></td></tr> <tr><td>Azure Blob</td><td><input type="checkbox"/></td></tr> <tr><td>Azure SQL Database</td><td><input type="checkbox"/></td></tr> </table>	Azure SQL Data Warehouse	<input checked="" type="checkbox"/>	Azure Files	<input type="checkbox"/>	Azure Blob	<input type="checkbox"/>	Azure SQL Database	<input type="checkbox"/>
Azure SQL Data Warehouse	<input checked="" type="checkbox"/>								
Azure Files	<input type="checkbox"/>								
Azure Blob	<input type="checkbox"/>								
Azure SQL Database	<input type="checkbox"/>								

- A. Mastered
- B. Not Mastered

Answer: A

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>

NEW QUESTION 5

- (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 the 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 6

- (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?

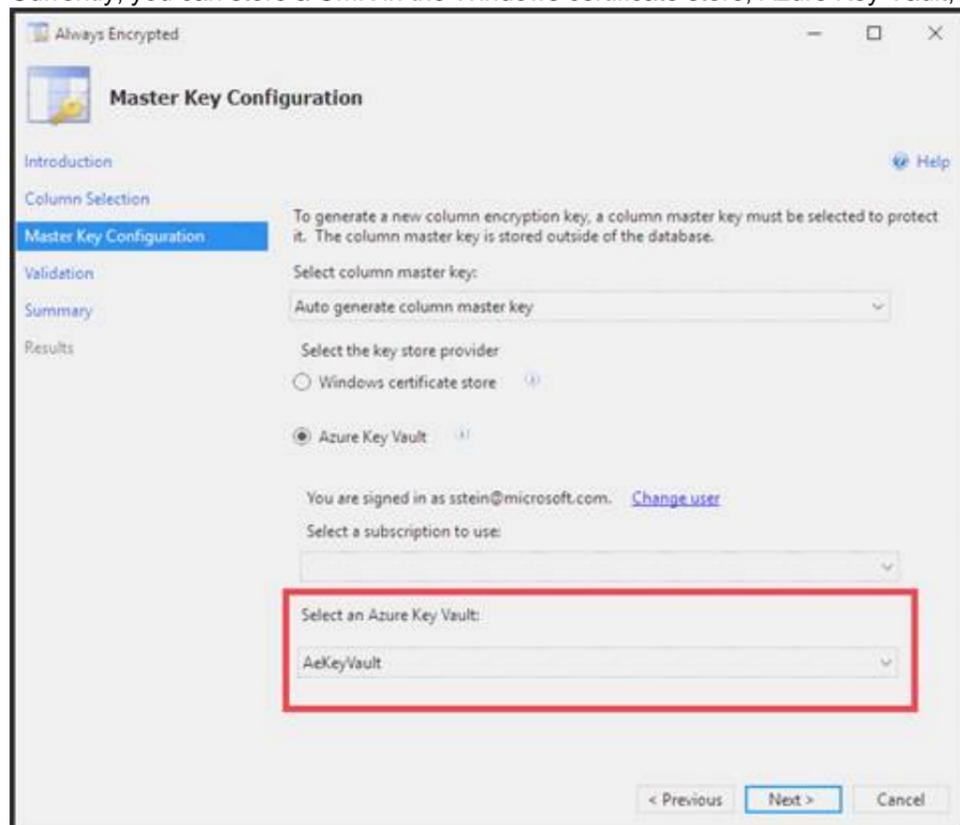
- A. Yes
- B. No

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 7

- (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><input type="checkbox"/></td></tr> <tr><td>default</td><td><input type="checkbox"/></td></tr> <tr><td>email</td><td><input type="checkbox"/></td></tr> <tr><td>random number</td><td><input type="checkbox"/></td></tr> </table>	custom text	<input type="checkbox"/>	default	<input type="checkbox"/>	email	<input type="checkbox"/>	random number	<input type="checkbox"/>
custom text	<input type="checkbox"/>								
default	<input type="checkbox"/>								
email	<input type="checkbox"/>								
random number	<input type="checkbox"/>								
B	<table border="1"> <tr><td>custom text</td><td><input type="checkbox"/></td></tr> <tr><td>default</td><td><input type="checkbox"/></td></tr> <tr><td>email</td><td><input type="checkbox"/></td></tr> <tr><td>random number</td><td><input type="checkbox"/></td></tr> </table>	custom text	<input type="checkbox"/>	default	<input type="checkbox"/>	email	<input type="checkbox"/>	random number	<input type="checkbox"/>
custom text	<input type="checkbox"/>								
default	<input type="checkbox"/>								
email	<input type="checkbox"/>								
random number	<input type="checkbox"/>								
C	<table border="1"> <tr><td>custom text</td><td><input type="checkbox"/></td></tr> <tr><td>default</td><td><input type="checkbox"/></td></tr> <tr><td>email</td><td><input type="checkbox"/></td></tr> <tr><td>random number</td><td><input type="checkbox"/></td></tr> </table>	custom text	<input type="checkbox"/>	default	<input type="checkbox"/>	email	<input type="checkbox"/>	random number	<input type="checkbox"/>
custom text	<input type="checkbox"/>								
default	<input type="checkbox"/>								
email	<input type="checkbox"/>								
random number	<input type="checkbox"/>								

- 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 8

- (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><input type="checkbox"/></td></tr> <tr><td>Azure Event Hubs</td><td><input checked="" type="checkbox"/></td></tr> <tr><td>HDInsight Storm</td><td><input type="checkbox"/></td></tr> <tr><td>HDInsight Spark</td><td><input type="checkbox"/></td></tr> </table>	HDInsight Kafka	<input type="checkbox"/>	Azure Event Hubs	<input checked="" type="checkbox"/>	HDInsight Storm	<input type="checkbox"/>	HDInsight Spark	<input type="checkbox"/>
HDInsight Kafka	<input type="checkbox"/>								
Azure Event Hubs	<input checked="" type="checkbox"/>								
HDInsight Storm	<input type="checkbox"/>								
HDInsight Spark	<input type="checkbox"/>								
Stream Processing	<table border="1"> <tr><td>Azure Stream Analytics</td><td><input type="checkbox"/></td></tr> <tr><td>HDInsight with Spark Streaming</td><td><input type="checkbox"/></td></tr> <tr><td>Azure Cosmos DB Change Feed</td><td><input type="checkbox"/></td></tr> <tr><td>Azure Analysis Services</td><td><input type="checkbox"/></td></tr> </table>	Azure Stream Analytics	<input type="checkbox"/>	HDInsight with Spark Streaming	<input type="checkbox"/>	Azure Cosmos DB Change Feed	<input type="checkbox"/>	Azure Analysis Services	<input type="checkbox"/>
Azure Stream Analytics	<input type="checkbox"/>								
HDInsight with Spark Streaming	<input type="checkbox"/>								
Azure Cosmos DB Change Feed	<input type="checkbox"/>								
Azure Analysis Services	<input type="checkbox"/>								
Analytical Data Store	<table border="1"> <tr><td>Hive LLAP on HDInsight</td><td><input type="checkbox"/></td></tr> <tr><td>Azure Analysis Services</td><td><input type="checkbox"/></td></tr> <tr><td>Azure Cosmos DB</td><td><input type="checkbox"/></td></tr> <tr><td>SQL Data Warehouse</td><td><input type="checkbox"/></td></tr> </table>	Hive LLAP on HDInsight	<input type="checkbox"/>	Azure Analysis Services	<input type="checkbox"/>	Azure Cosmos DB	<input type="checkbox"/>	SQL Data Warehouse	<input type="checkbox"/>
Hive LLAP on HDInsight	<input type="checkbox"/>								
Azure Analysis Services	<input type="checkbox"/>								
Azure Cosmos DB	<input type="checkbox"/>								
SQL Data Warehouse	<input type="checkbox"/>								

- 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 9

- (Exam Topic 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

Answer: D

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>

NEW QUESTION 10

- (Exam Topic 3)

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;

Answer: CD

Explanation:

C: ALTER ROLE adds or removes members to or from a database role, or changes the name of a user-defined database role.

Members of the db_owner fixed database role can perform all configuration and maintenance activities on the database, and can also drop the database in SQL Server.

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

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

References:

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

NEW QUESTION 10

- (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 14

- (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.

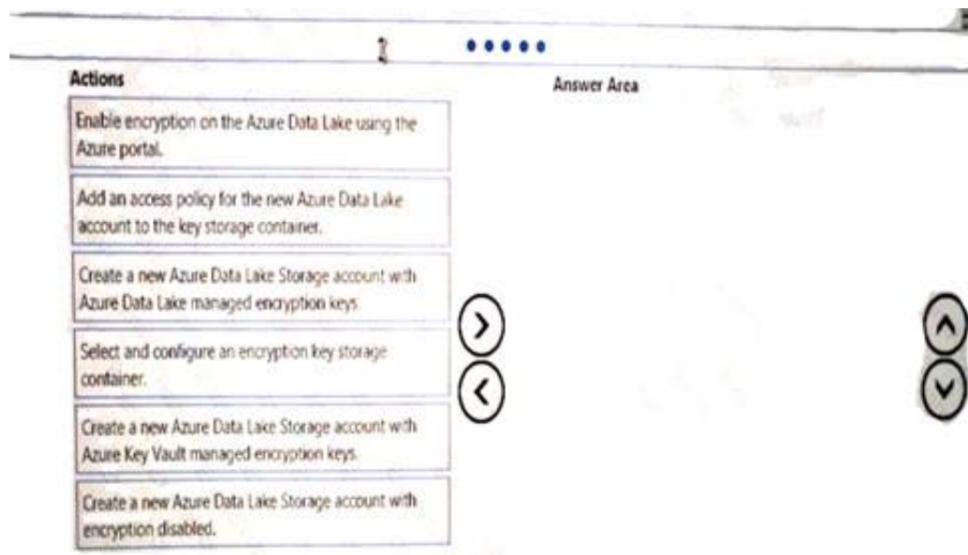
NEW QUESTION 16

- (Exam Topic 3)

You are developing a solution to visualize multiple terabytes of geospatial data. The solution has the following requirements:

- Data must be encrypted.
- Data must be accessible by multiple resources on Microsoft Azure. You need to provision storage for the solution.

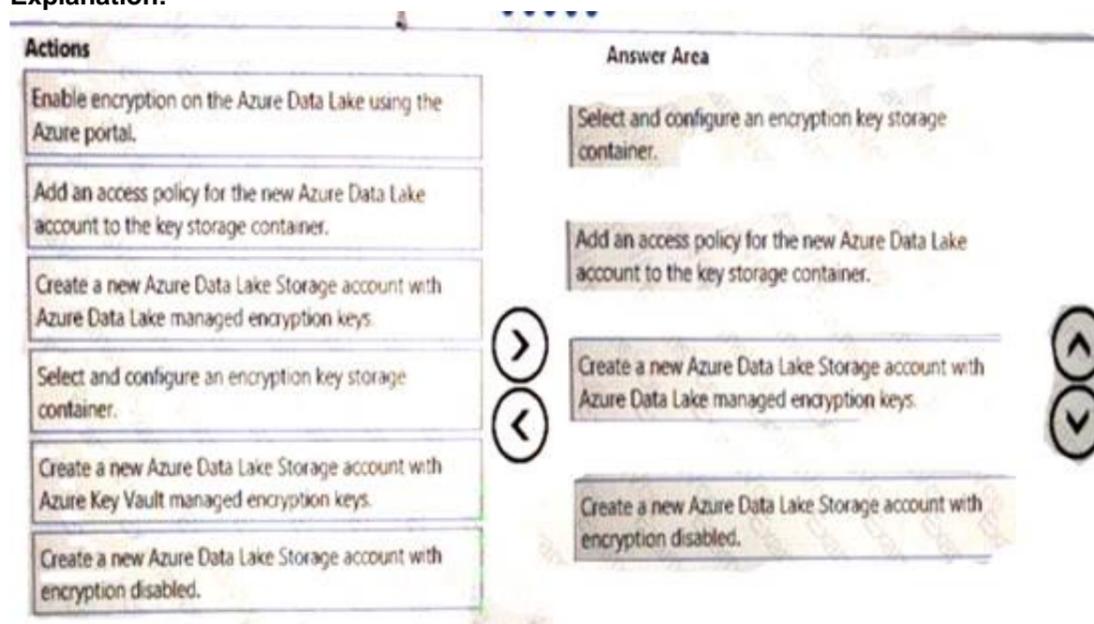
Which four actions should you perform in sequence? To answer, move the appropriate action from the list of actions to the answer area and arrange them in the correct order.



- A. Mastered
- B. Not Mastered

Answer: A

Explanation:



NEW QUESTION 19

- (Exam Topic 3)

Your company uses several Azure HDInsight clusters.

The data engineering team reports several errors with some application using these clusters. You need to recommend a solution to review the health of the clusters.

What should you include in you recommendation?

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

Answer: C

NEW QUESTION 23

- (Exam Topic 3)

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

Answer: D

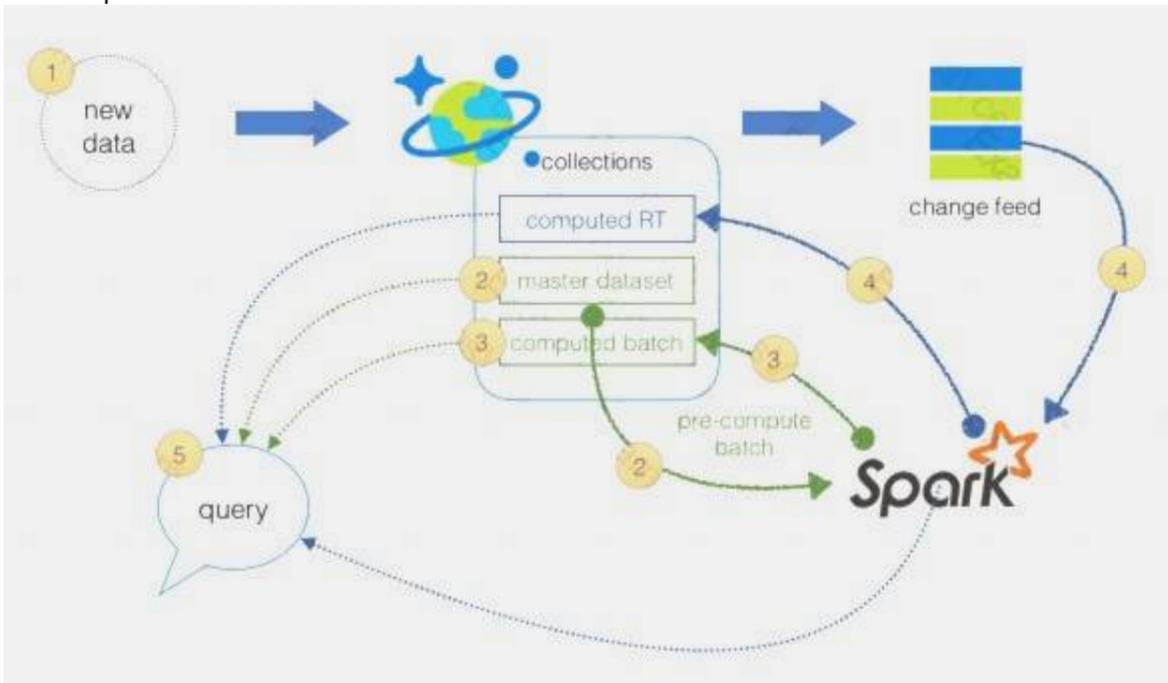
Explanation:

Lambda Architecture with Azure:

Azure offers you a combination of following technologies to accelerate real-time big data analytics:

- Azure Cosmos DB, a globally distributed and multi-model database service.

- ▶ Apache Spark for Azure HDInsight, a processing framework that runs large-scale data analytics applications.
- ▶ 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->

NEW QUESTION 24

- (Exam Topic 3)

You develop data engineering solutions for a company.

You need to deploy a Microsoft Azure Stream Analytics job for an IoT solution. The solution must:

- Minimize latency.
- Minimize bandwidth usage between the job and IoT device.

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

The screenshot shows an exam interface with two main sections: 'Actions' and 'Answer Area'. The 'Actions' list includes:

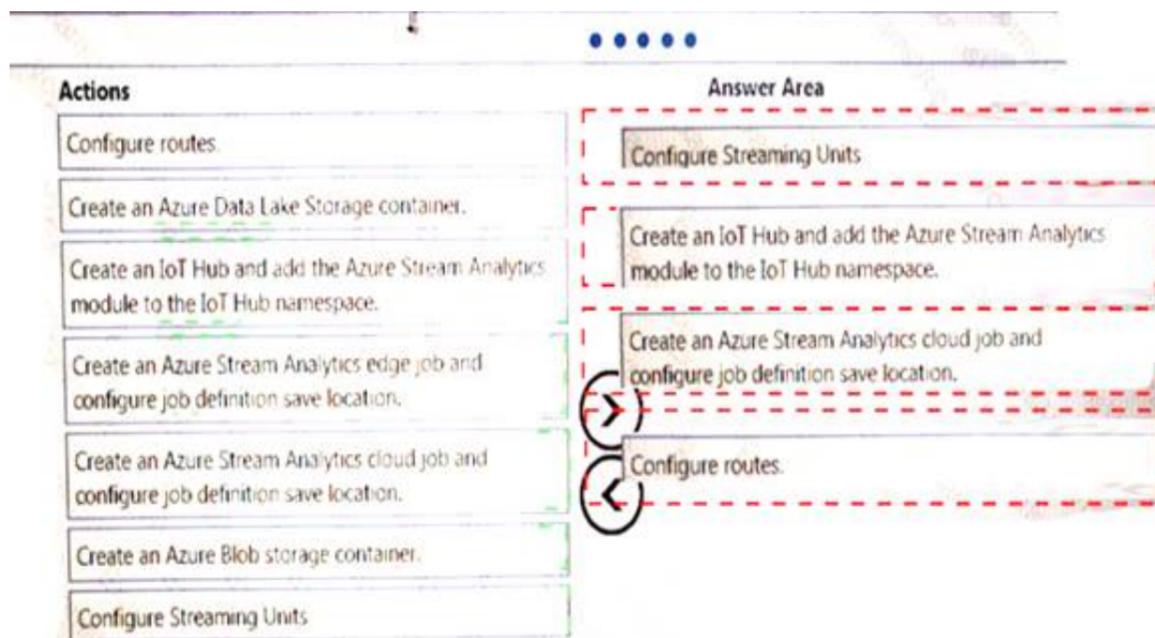
- Configure routes.
- Create an Azure Data Lake Storage container.
- Create an IoT Hub and add the Azure Stream Analytics module to the IoT Hub namespace.
- Create an Azure Stream Analytics edge job and configure job definition save location.
- Create an Azure Stream Analytics cloud job and configure job definition save location.
- Create an Azure Blob storage container.
- Configure Streaming Units.

 The 'Answer Area' is currently empty. Navigation arrows are visible between the two sections.

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:



NEW QUESTION 28

- (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	<input type="checkbox"/>
	Apache Sqoop	<input checked="" type="checkbox"/>
	Apache DistCp	<input type="checkbox"/>
	Azure CLI	<input type="checkbox"/>
Process	Apache DistCp	<input checked="" type="checkbox"/>
	Apache Kafka	<input type="checkbox"/>
	C#	<input type="checkbox"/>
	Apache Hive	<input type="checkbox"/>
Download	Apache Sqoop	<input checked="" type="checkbox"/>
	MapReduce	<input type="checkbox"/>
	RevoScaleR	<input type="checkbox"/>
	Ambari Hive View	<input type="checkbox"/>

- 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 30

- (Exam Topic 3)

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.

Action	Technology					
Extract SQL data on-premises	<table border="1"> <tr><td>Self-hosted integration runtime</td><td rowspan="4">v</td></tr> <tr><td>Azure-SSIS integration runtime</td></tr> <tr><td>Azure integration runtime</td></tr> <tr><td>Source</td></tr> </table>	Self-hosted integration runtime	v	Azure-SSIS integration runtime	Azure integration runtime	Source
Self-hosted integration runtime	v					
Azure-SSIS integration runtime						
Azure integration runtime						
Source						
Load SQL data warehouse	<table border="1"> <tr><td>Self-hosted integration runtime</td><td rowspan="4">v</td></tr> <tr><td>Azure-SSIS integration runtime</td></tr> <tr><td>Azure integration runtime</td></tr> <tr><td>Sink</td></tr> </table>	Self-hosted integration runtime	v	Azure-SSIS integration runtime	Azure integration runtime	Sink
Self-hosted integration runtime	v					
Azure-SSIS integration runtime						
Azure integration runtime						
Sink						

- A. Mastered
- B. Not Mastered

Answer: A

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>

NEW QUESTION 32

- (Exam Topic 3)

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 (orows) was reached while regarding 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] occurredwhile accessing external files.'

Answer: C

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.

NEW QUESTION 33

- (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 38

- (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 39

- (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 44

- (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 rowspan="2">✓</td> </tr> <tr> <td>NameNode</td> </tr> </table>	DataNode	✓	NameNode
DataNode	✓			
NameNode				
Run operations on files and directories of the file system.	<table border="1"> <tr> <td>DataNode</td> <td rowspan="2">✓</td> </tr> <tr> <td>NameNode</td> </tr> </table>	DataNode	✓	NameNode
DataNode	✓			
NameNode				
Perform block creation, deletion, and replication.	<table border="1"> <tr> <td>DataNode</td> <td rowspan="2">✓</td> </tr> <tr> <td>NameNode</td> </tr> </table>	DataNode	✓	NameNode
DataNode	✓			
NameNode				

- 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 48

.....

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>