

## 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.

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
Parameterize deployment by using Azure Integration Runtime	
Configure an Azure Logic App to deploy the deployment artifact	
Configure Azure DevOps to deploy the deployment artifact	
Create a deployment artifact containing an extracted Azure Resource Manager template	
Parameterize deployment by using the Azure Resource Manager template parameter file	
Create a deployment artifact containing a SQL Server Integration Services (SSIS) package	

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

**Answer Area**

- Create a deployment artifact containing an extracted Azure Resource Manager template
- Parameterize deployment by using the Azure Resource Manager template parameter file
- Configure Azure DevOps to deploy the deployment artifact

Scenario:

All deployments must be performed by using Azure DevOps. Deployments must use templates used in multiple environments  
No credentials or secrets should be used during deployments

**NEW QUESTION 2**

- (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 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

FileCount	
BlobCapacity	
FileCapacity	

Aggregation

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

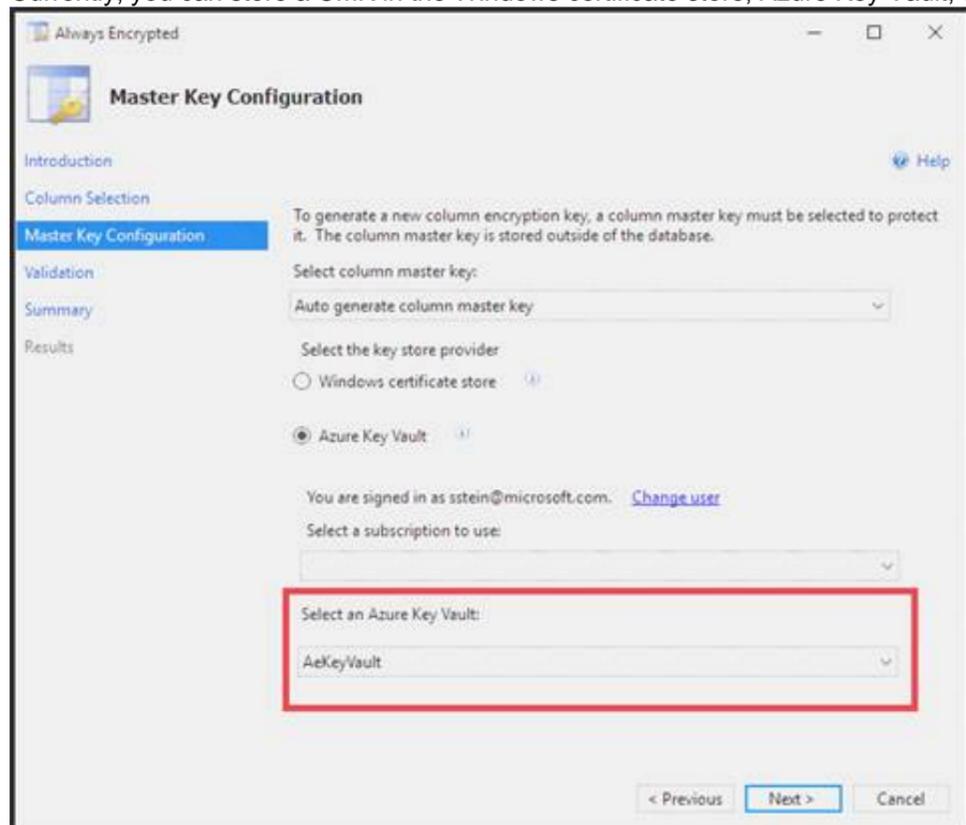
**Answer:** B

**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 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 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 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 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 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 type="checkbox"/>	Azure Files	<input type="checkbox"/>	Azure Blob	<input type="checkbox"/>	Azure SQL Database	<input type="checkbox"/>
Azure SQL Data Warehouse	<input 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 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 6**

- (Exam Topic 2)

You need to set up Azure Data Factory pipelines to meet data movement requirements. Which integration runtime should you use?

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

**Answer: A**

**Explanation:**

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

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

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

References:

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

**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 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 8**

- (Exam Topic 3)

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

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

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

NOTE: Each correct selection is worth one point.

Parameter	Value					
<code>--default-consistency-level</code>	<table border="1"> <tr><td>Strong</td><td rowspan="4">V</td></tr> <tr><td>Session</td></tr> <tr><td>Eventual</td></tr> <tr><td>Bounded staleness</td></tr> </table>	Strong	V	Session	Eventual	Bounded staleness
Strong	V					
Session						
Eventual						
Bounded staleness						
<code>--kind</code>	<table border="1"> <tr><td>Parse</td><td rowspan="3">V</td></tr> <tr><td>MongoDB</td></tr> <tr><td>GlobalDocumentDB</td></tr> </table>	Parse	V	MongoDB	GlobalDocumentDB	
Parse	V					
MongoDB						
GlobalDocumentDB						

- A. Mastered
- B. Not Mastered

**Answer:** A

**Explanation:**

Box 1: Eventual

With Azure Cosmos DB, developers can choose from five well-defined consistency models on the consistency spectrum. From strongest to more relaxed, the models include strong, bounded staleness, session, consistent prefix, and eventual consistency.

The following image shows the different consistency levels as a spectrum.



Box 2: GlobalDocumentDB

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

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

References:

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

**NEW QUESTION 9**

- (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								
<b>Ingestion</b>	<table border="1"> <tr><td>HDInsight Kafka</td><td><input checked="" type="checkbox"/></td></tr> <tr><td>Azure Event Hubs</td><td><input 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 checked="" type="checkbox"/>	Azure Event Hubs	<input type="checkbox"/>	HDInsight Storm	<input type="checkbox"/>	HDInsight Spark	<input type="checkbox"/>
HDInsight Kafka	<input checked="" type="checkbox"/>								
Azure Event Hubs	<input type="checkbox"/>								
HDInsight Storm	<input type="checkbox"/>								
HDInsight Spark	<input type="checkbox"/>								
<b>Stream Processing</b>	<table border="1"> <tr><td>Azure Stream Analytics</td><td><input checked="" 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 checked="" 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 checked="" type="checkbox"/>								
HDInsight with Spark Streaming	<input type="checkbox"/>								
Azure Cosmos DB Change Feed	<input type="checkbox"/>								
Azure Analysis Services	<input type="checkbox"/>								
<b>Analytical Data Store</b>	<table border="1"> <tr><td>Hive LLAP on HDInsight</td><td><input checked="" 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 checked="" 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 checked="" 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 10**

- (Exam Topic 3)

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

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

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

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

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

- A. Mastered

B. Not Mastered

**Answer:** A

**Explanation:**

Step 1: Create a virtual private network (VPN) connection from on-premises to Microsoft Azure.

You can also use IPsec VPN or Azure ExpressRoute to further secure the communication channel between your on-premises network and Azure.

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

Step 2: Create an Azure Data Factory resource. Step 3: Configure a self-hosted integration runtime.

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

Note: A self-hosted integration runtime can run copy activities between a cloud data store and a data store in a private network, and it can dispatch transform activities against compute resources in an on-premises network or an Azure virtual network. The installation of a self-hosted integration runtime needs on an on-premises machine or a virtual machine (VM) inside a private network.

References:

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

**NEW QUESTION 10**

- (Exam Topic 3)

You develop data engineering solutions for a company.

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

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

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

NOTE: Each correct selection is worth one point.

**Answer Area**

New-AzureStorageContainer  
New-AzureRmHDInsightClusterConfig  
New-AzureRmHDInsightCluster

```

-Name $clusterName -Context $defaultStorageContext
$objectConfig = New-Object "System.Collections.Generic.D
$objectConfig.Add (
(
"spark"
"hadoop"
-ResourceGroupName $resourceGroupName `
-ClusterName $clusterName `
-Location $location `
-ClusterSizeInNodes $clusterSizeInNodes `
-ClusterType "spark"
".hadoop"
"HBase"
"Storm"
-OSType $clusterOS `
-Version $clusterVersion `
-ComponentVersion $objectConfig

```

A. Mastered

B. Not Mastered

**Answer:** A

**Explanation:**

**Answer Area**

New-AzureStorageContainer  
New-AzureRmHDInsightClusterConfig  
New-AzureRmHDInsightCluster

```

-Name $clusterName -Context $defaultStorageContext
$objectConfig = New-Object "System.Collections.Generic.D
$objectConfig.Add (
(
"spark"
"hadoop"
-ResourceGroupName $resourceGroupName `
-ClusterName $clusterName `
-Location $location `
-ClusterSizeInNodes $clusterSizeInNodes `
-ClusterType "spark"
".hadoop"
"HBase"
"Storm"
-OSType $clusterOS `
-Version $clusterVersion `
-ComponentVersion $objectConfig

```

**NEW QUESTION 12**

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

**NEW QUESTION 19**

- (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 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 your recommendation?

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

**Answer:** C

**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 30**

- (Exam Topic 3)

A company runs Microsoft SQL Server in an on-premises virtual machine (VM).

You must migrate the database to Azure SQL Database. You synchronize users from Active Directory to Azure Active Directory (Azure AD).

You need to configure Azure SQL Database to use an Azure AD user as administrator. What should you configure?

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

**Answer:** A

**NEW QUESTION 31**

- (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

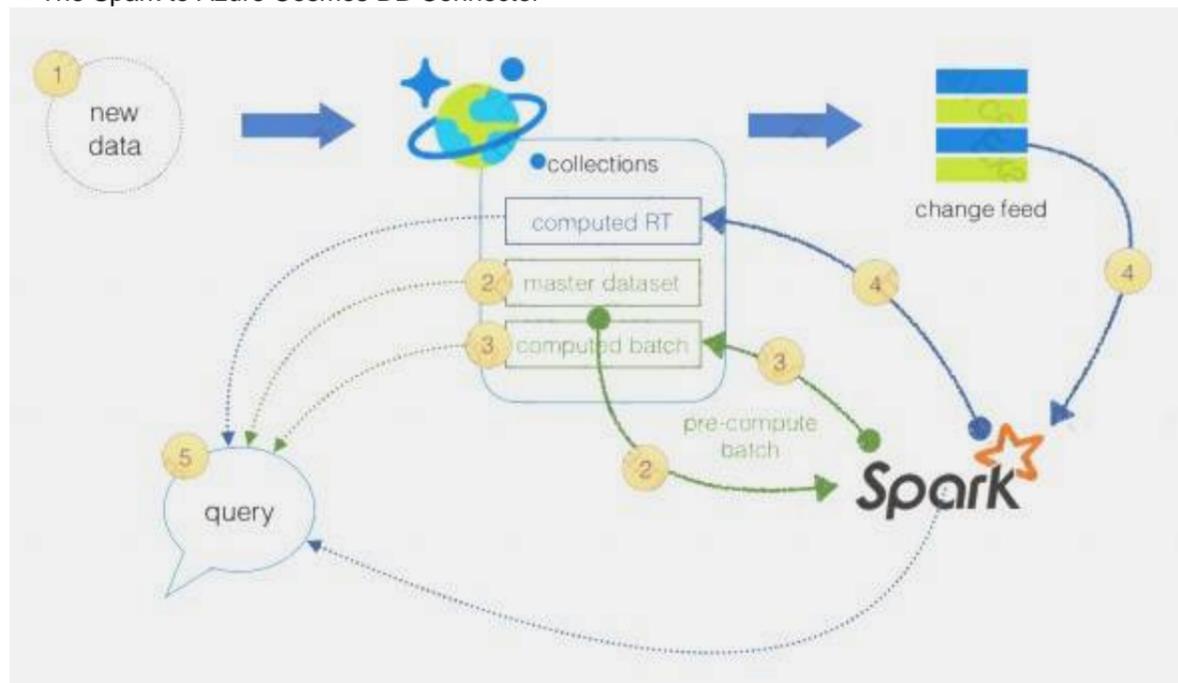
**Answer: D**

**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 32**

- (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 port.

- 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 35**

- (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 38**

- (Exam Topic 3)

A company is designing a hybrid solution to synchronize data and on-premises Microsoft SQL Server database to Azure SQL Database. You must perform an assessment of databases to determine whether data will move without compatibility issues. You need to perform the assessment. Which tool should you use?

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

**Answer: E**

**Explanation:**

The Data Migration Assistant (DMA) helps you upgrade to a modern data platform by detecting compatibility issues that can impact database functionality in your new version of SQL Server or Azure SQL Database. DMA recommends performance and reliability improvements for your target environment and allows you to move your schema, data, and uncontained objects from your source server to your target server.

References:

<https://docs.microsoft.com/en-us/sql/dma/dma-overview>

**NEW QUESTION 39**

- (Exam Topic 3)

Your company uses Microsoft Azure SQL Database configured with Elastic pool. You use Elastic Database jobs to run queries across all databases in the pool. You need to analyze, troubleshoot, and report on components responsible for running Elastic Database jobs. You need to determine the component responsible for running job service tasks.

Which components should you use for each Elastic pool job services task? To answer, drag the appropriate component to the correct task. Each component may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content. NOTE: Each correct selection is worth one point.

- A. Mastered
- B. Not Mastered

**Answer: A**

**Explanation:**

**NEW QUESTION 42**

- (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 (rows) 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] occurred while 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 46

- (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 48

- (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. Determine whether the solution meets the stated goals.

You develop a data ingestion process that will import data to a Microsoft Azure SQL Data Warehouse. The data to be ingested resides in parquet files stored in an Azure Data Lake Gen 2 storage account. You need to load the data from the Azure Data Lake Gen 2 storage account into the Azure SQL Data Warehouse.

Solution:

1. Create an external data source pointing to the Azure storage account
2. Create an external file format and external table using the external data source
3. Load the data using the INSERT...SELECT statement

Does the solution meet the goal?

- A. Yes
- B. No

**Answer:** B

#### Explanation:

You load the data using the CREATE TABLE AS SELECT statement. References:

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

#### NEW QUESTION 52

- (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">V</td> </tr> <tr> <td>NameNode</td> </tr> </table>	DataNode	V	NameNode
DataNode	V			
NameNode				
Run operations on files and directories of the file system.	<table border="1"> <tr> <td>DataNode</td> <td rowspan="2">V</td> </tr> <tr> <td>NameNode</td> </tr> </table>	DataNode	V	NameNode
DataNode	V			
NameNode				
Perform block creation, deletion, and replication.	<table border="1"> <tr> <td>DataNode</td> <td rowspan="2">V</td> </tr> <tr> <td>NameNode</td> </tr> </table>	DataNode	V	NameNode
DataNode	V			
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](https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html#NameNode+and+DataNodes)

**NEW QUESTION 57**

- (Exam Topic 3)

You develop data engineering solutions for a company.

You need to ingest and visualize real-time Twitter data by using Microsoft Azure.

Which three technologies should you use? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. Event Grid topic
- B. Azure Stream Analytics Job that queries Twitter data from an Event Hub
- C. Azure Stream Analytics Job that queries Twitter data from an Event Grid
- D. Logic App that sends Twitter posts which have target keywords to Azure
- E. Event Grid subscription
- F. Event Hub instance

**Answer:** BDF

**Explanation:**

You can use Azure Logic apps to send tweets to an event hub and then use a Stream Analytics job to read from event hub and send them to PowerBI.

References:

<https://community.powerbi.com/t5/Integrations-with-Files-and/Twitter-streaming-analytics-step-by-step/td-p/95>

**NEW QUESTION 59**

.....

## 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>