

Microsoft

Exam Questions DP-200

Implementing an Azure Data Solution



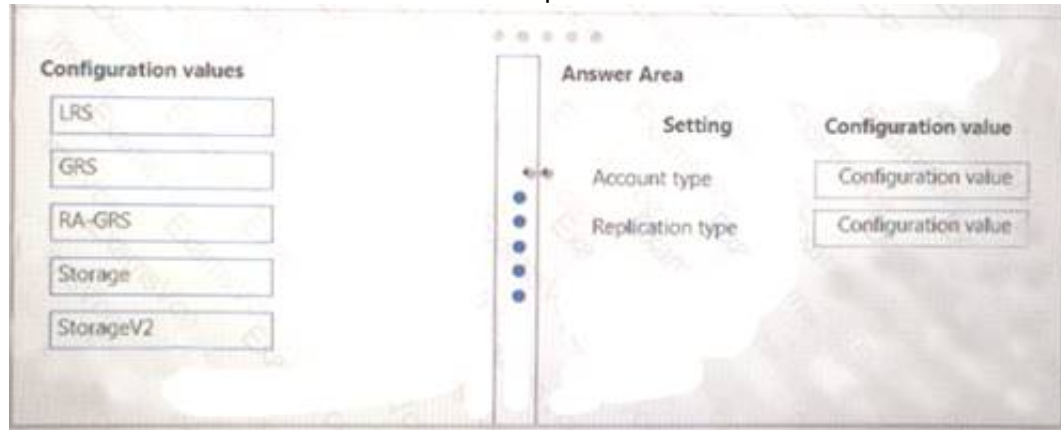
NEW QUESTION 1

- (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.

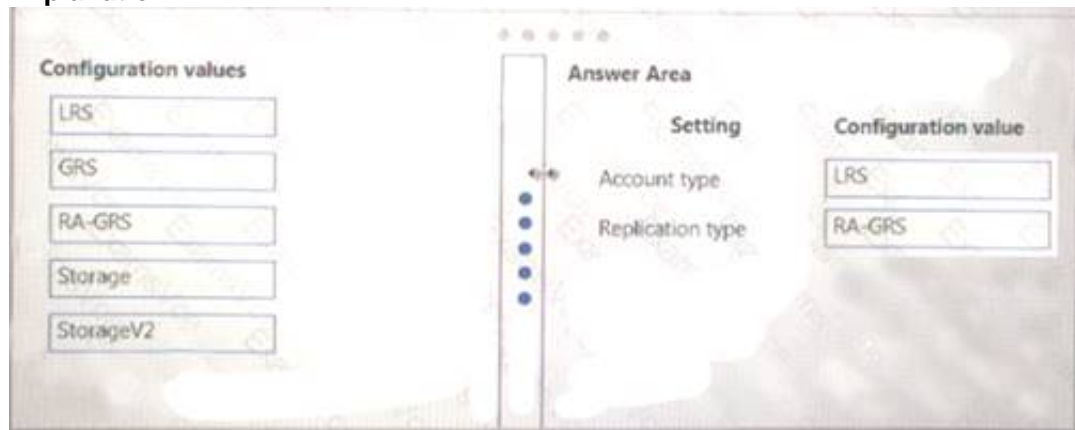
NOTE: Each correct selection is worth one point.



- A. Mastered
- B. Not Mastered

Answer: A

Explanation:



NEW QUESTION 2

- (Exam Topic 1)

You need to ensure that phone-based poling data can be analyzed in the PollingData database. How should you configure Azure Data Factory?

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

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 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 4

- (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>	<div>Strong</div> <div>Session</div> <div>Eventual</div> <div>Bounded staleness</div>
<code>--kind</code>	<div>Parse</div> <div>MongoDB</div> <div>GlobalDocumentDB</div>

- 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 5

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

- 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 6

- (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 7

- (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 8

- (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 9

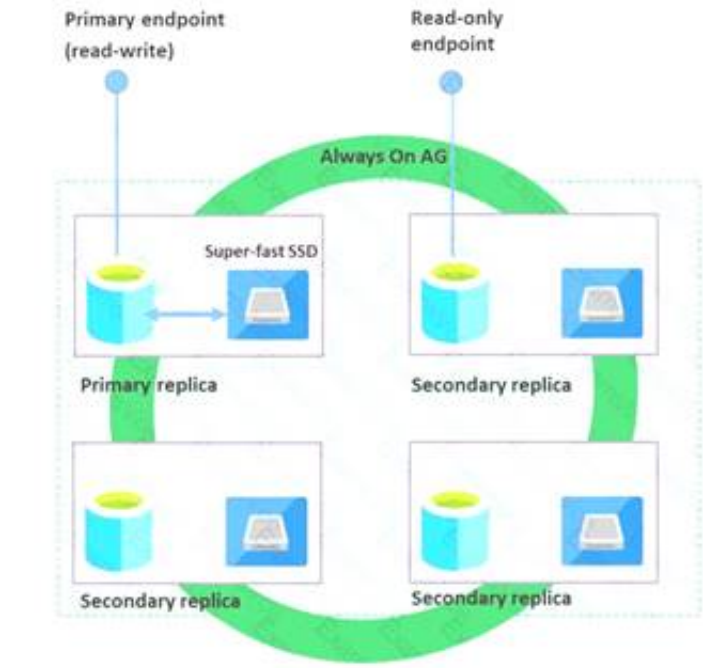
- (Exam Topic 3)
A company plans to use Azure SQL Database to support a mission-critical application. The application must be highly available without performance degradation during maintenance windows. You need to implement the solution. Which three technologies should you implement? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

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

Answer: AEF

Explanation:

Premium/business critical service tier model that is based on a cluster of database engine processes. This architectural model relies on a fact that there is always a quorum of available database engine nodes and has minimal performance impact on your workload even during maintenance activities. In the premium model, Azure SQL database integrates compute and storage on the single node. High availability in this architectural model is achieved by replication of compute (SQL Server Database Engine process) and storage (locally attached SSD) deployed in 4-node cluster, using technology similar to SQL Server Always On Availability Groups.



Business Critical service tier: collocated compute and storage

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

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

NEW QUESTION 10

- (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

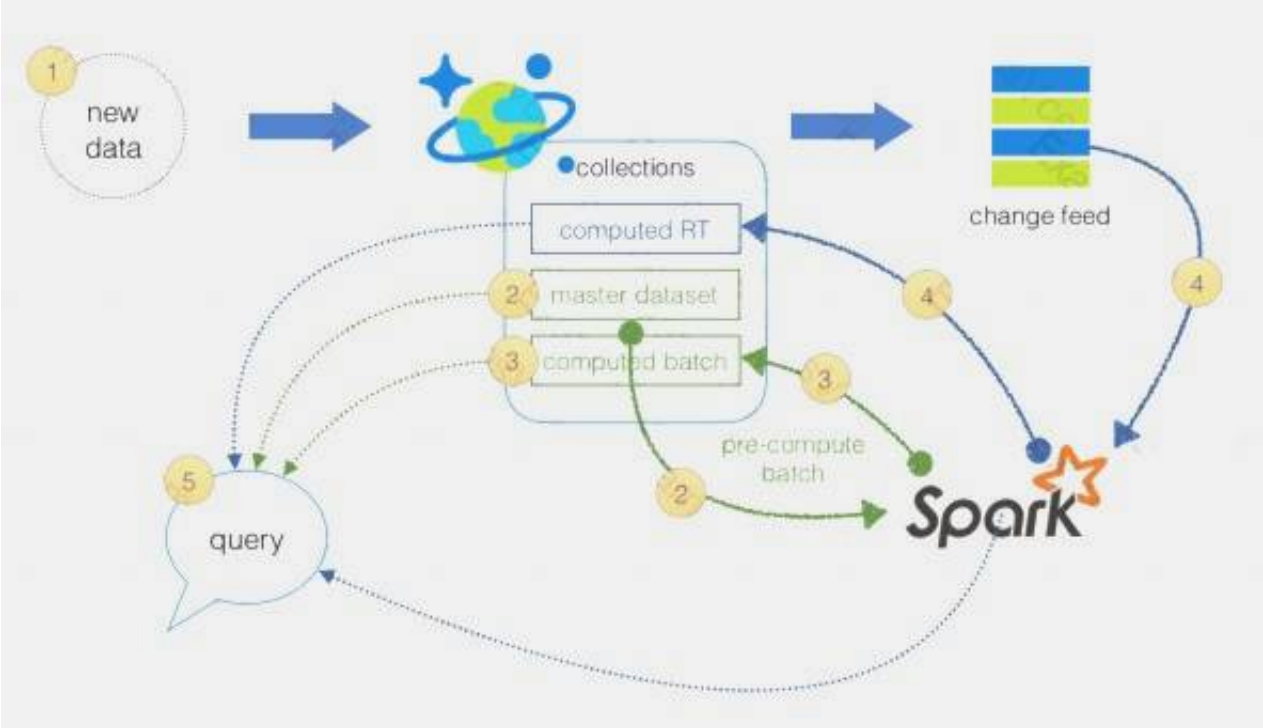
Answer: D

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 10

- (Exam Topic 3)

An application will use Microsoft Azure Cosmos DB as its data solution. The application will use the Cassandra API to support a column-based database type that uses containers to store items.

You need to provision Azure Cosmos DB. Which container name and item name should you use? Each correct answer presents part of the solutions.

NOTE: Each correct answer selection is worth one point.

- A. table
- B. collection
- C. graph
- D. entities
- E. rows

Answer: AE

Explanation:

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

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

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

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

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

NEW QUESTION 13

- (Exam Topic 3)
A company plans to analyze a continuous flow of data from a social media platform by using Microsoft Azure Stream Analytics. The incoming data is formatted as one record per row.
You need to create the input stream.
How should you complete the REST API segment? To answer, select the appropriate configuration in the answer area.
NOTE: Each correct selection is worth one point.

Answer Area

```
{
  "properties":{
    "type":"stream",
    "serialization":{
      
      "properties":{
        "fieldDelimiter":",",
        "encoding":"UTF8"
      }
    },
    "datasource":{
      
      "properties":{
        "serviceBusNamespace":"sampleServiceBus",
        "sharedAccessPolicyName":"SampleReceiver",
        "sharedAccessPolicyKey":"<PolicyKey>"
        "eventHubName":"sampleEventHub"
      }
    },
    "compression":{
      "type":"GZip"
    }
  }
}
```

Answer Area

```
{
  "properties":{
    "type":"stream",
    "serialization":{
      
      "properties":{
        "fieldDelimiter":",",
        "encoding":"UTF8"
      }
    },
    "datasource":{
      
      "properties":{
        "serviceBusNamespace":"sampleServiceBus",
        "sharedAccessPolicyName":"SampleReceiver",
        "sharedAccessPolicyKey":"<PolicyKey>"
        "eventHubName":"sampleEventHub"
      }
    }
  }
}
```

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Answer Area

```
{
  "properties":{
    "type":"stream",
    "serialization":{
      "type":"CSV",
      "type":"Avro",
      "type":"JSON",
    },
    "properties":{
      "fieldDelimiter":",",
      "encoding":"UTF8"
    }
  },
  "datasource":{
    "type":"Microsoft.Storage/Blob",
    "type":"Microsoft.ServiceBus/EventHub",
    "type":"Microsoft.Devices/IotHubs",
    "properties":{
      "serviceBusNamespace":"sampleServiceBus",
      "sharedAccessPolicyName":"SampleReceiver",
      "sharedAccessPolicyKey":"<PolicyKey>",
      "eventHubName":"sampleEventHub"
    }
  }
}
```

NEW QUESTION 17

- (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 20

- (Exam Topic 3)

Your company uses Microsoft Azure SQL Database configure with Elastic pool. You use Elastic Database jobs to run queries across all databases in the pod.

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.

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

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

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

NEW QUESTION 23

- (Exam Topic 3)

A company manages several on-premises Microsoft SQL Server databases. You need to migrate the databases to Microsoft Azure by using the backup process of Microsoft SQL Server. Which data technology should you use?

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

Answer: D

NEW QUESTION 28

- (Exam Topic 3)

You plan to create a new single database instance of Microsoft Azure SQL Database. The database must only allow communication from the data engineer's workstation. You must connect directly to the instance by using Microsoft SQL Server Management Studio. You need to create and configure the Database. Which three Azure PowerShell cmdlets should you use to develop the solution? To answer, move the appropriate cmdlets from the list of cmdlets to the answer area and arrange them in the correct order.

Azure PowerShell cmdlets

New-AzureRmSqlElasticPool

New-AzureRmSqlServerFirewallRule

New-AzureRmSqlServer

New-AzureRmSqlServerVirtualNetworkRule

New-AzureRmSqlDatabase

Answer Area

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Step 1: New-AzureSqlServer Create a server.
Step 2: New-AzureRmSqlServerFirewallRule
New-AzureRmSqlServerFirewallRule creates a firewall rule for a SQL Database server. Can be used to create a server firewall rule that allows access from the specified IP range. Step 3: New-AzureRmSqlDatabase
Example: Create a database on a specified server
PS C:\>New-AzureRmSqlDatabase -ResourceGroupName "ResourceGroup01" -ServerName "Server01" -DatabaseName "Database01
References:
https://docs.microsoft.com/en-us/azure/sql-database/scripts/sql-database-create-and-configure-database-powersh

NEW QUESTION 31

- (Exam Topic 3)

A company builds an application to allow developers to share and compare code. The conversations, code snippets, and links shared by people in the application are stored in a Microsoft Azure SQL Database instance. The application allows for searches of historical conversations and code snippets. When users share code snippets, the code snippet is compared against previously share code snippets by using a combination of Transact-SQL functions including SUBSTRING, FIRST_VALUE, and SQRT. If a match is found, a link to the match is added to the conversation. Customers report the following issues:

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

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

Technologies

columnstore index

non-durable table

meterialized view

memory-optimized table

Answer Area

IssueTechnology

Delays in conversations

Delays in match links

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: memory-optimized table

In-Memory OLTP can provide great performance benefits for transaction processing, data ingestion, and transient data scenarios.

Box 2: materialized view

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

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

References:

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

NEW QUESTION 33

- (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 (0 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] 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 35

- (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 36

- (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 39

- (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 40

- (Exam Topic 3)

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

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

Actions	Answer Area
Upload a certificate	
Add secrets to the scope	
Use Blob Storage access key	
Create a secret scope	
Configure a JDBC connector	
Mount the Azure Blob Storage container	

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Step 1: Create a secret scope Step 2: Add secrets to the scope

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

Step 3: Mount the Azure Blob Storage container

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

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

Python `dbutils.fs.mount(`

`source = "wasbs://<your-container-name>@<your-storage-account-name>.blob.core.windows.net", mount_point = "/mnt/<mount-name>",`

`extra_configs = {"<conf-key>":dbutils.secrets.get(scope = "<scope-name>", key = "<key-name>")})` where:

dbutils.secrets.get(scope = "<scope-name>", key = "<key-name>") gets the key that has been stored as a secret in a secret scope.
References:
<https://docs.databricks.com/spark/latest/data-sources/azure/azure-storage.html>

NEW QUESTION 44

- (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 49

- (Exam Topic 3)

Your company plans to create an event processing engine to handle streaming data from Twitter. The data engineering team uses Azure Event Hubs to ingest the streaming data.
You need to implement a solution that uses Azure Databricks to receive the streaming data from the Azure Event Hubs.
Which three actions should you recommend be performed in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

12345

Actions

Answer Area

Create and configure a Notebook that consumes the streaming data.

Import data from Blob storage.

Use Environment variables to define the Apache Spark connection.

Configure an ODBC or JDBC Connector.

Deploy the Azure Databricks service.

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

>

<

- A. Mastered
B. Not Mastered

Answer: A

Explanation:

12345

Actions

Answer Area

Create and configure a Notebook that consumes the streaming data.

Import data from Blob storage.

Use Environment variables to define the Apache Spark connection.

Configure an ODBC or JDBC Connector.

Deploy the Azure Databricks service.

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

Import data from Blob storage.

Configure an ODBC or JDBC Connector.

Use Environment variables to define the Apache S connection.

NEW QUESTION 52

- (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.

...

Actions

Create an external file format to map the parquet files.

Load the data to a staging table

Create the external table FactSalesOrderDetails.

Enable Transparent Data Encryption.

Create an external data source for Azure Blob storage.

Create a master key on database

Configure PolyBase to use Azure Blob storage.

Answer Area

>

<

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

...

Actions

Create an external file format to map the parquet files.

Load the data to a staging table.

Create the external table FactSalesOrderDetails.

Enable Transparent Data Encryption.

Create an external data source for Azure Blob storage.

Create a master key on database.

Configure PolyBase to use Azure Blob storage.

Answer Area

Enable Transparent Data Encryption.

Configure PolyBase to use Azure Blob storage.

Load the data to a staging table.

Create an external file format to map the parquet files.

NEW QUESTION 55

- (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 a workload group using the Azure storage account name as the pool name
3. Load the data using the INSERT...SELECT statement

Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

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

References:

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

NEW QUESTION 58

- (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 63

- (Exam Topic 3)

A company has a SaaS solution 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 68

- (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 69

- (Exam Topic 3)

You manage a financial computation data analysis process. Microsoft Azure virtual machines (VMs) run the process in daily jobs, and store the results in virtual hard drives (VHDs.)

The VMs produce results using data from the previous day and store the results in a snapshot of the VHD. When a new month begins, a process creates a new VHD.

You must implement the following data retention requirements:

- ▶ Daily results must be kept for 90 days
- ▶ Data for the current year must be available for weekly reports
- ▶ Data from the previous 10 years must be stored for auditing purposes
- ▶ Data required for an audit must be produced within 10 days of a request. You need to enforce the data retention requirements while minimizing cost.

How should you configure the lifecycle policy? To answer, drag the appropriate JSON segments to the correct locations. Each JSON segment may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

Code segments	Answer Area
<div>delete</div>	<pre>{ "version": "0.5", "rules": [{ "name": "dataRetention", "type": "Lifecycle", "definition": { "actions": { "": { <div></div> }, "": { "daysAfterModificationGreaterThan": 365 }, "": { "daysAfterModificationGreaterThan": 3650 } }, "": { <div></div> }, "": { "daysAfterCreationGreaterThan": 90 } } }] }</pre>
<div>blockBob</div>	
<div>baseBlob</div>	
<div>snapshot</div>	
<div>tierToCool</div>	
<div>tierToArchive</div>	

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

The Set-AzStorageAccountManagementPolicy cmdlet creates or modifies the management policy of an Azure Storage account.
Example: Create or update the management policy of a Storage account with ManagementPolicy rule objects.
Action -BaseBlobAction Delete -daysAfterModificationGreaterThan 100
PS C:\>\$action1 = Add-AzStorageAccountManagementPolicyAction -InputObject \$action1 -BaseBlobAction TierToArchive -daysAfterModificationGreaterThan 50
PS C:\>\$action1 = Add-AzStorageAccountManagementPolicyAction -InputObject \$action1 -BaseBlobAction TierToCool -daysAfterModificationGreaterThan 30
PS C:\>\$action1 = Add-AzStorageAccountManagementPolicyAction -InputObject \$action1 -SnapshotAction Delete -daysAfterCreationGreaterThan 100
PS C:\>\$filter1 = New-AzStorageAccountManagementPolicyFilter -PrefixMatch ab,cd
PS C:\>\$rule1 = New-AzStorageAccountManagementPolicyRule -Name Test -Action \$action1 -Filter \$filter1
PS C:\>\$action2 = Add-AzStorageAccountManagementPolicyAction -BaseBlobAction Delete -daysAfterModificationGreaterThan 100
PS C:\>\$filter2 = New-AzStorageAccountManagementPolicyFilter References:
<https://docs.microsoft.com/en-us/powershell/module/az.storage/set-azstorageaccountmanagementpolicy>

NEW QUESTION 70

.....

Thank You for Trying Our Product

We offer two products:

1st - We have Practice Tests Software with Actual Exam Questions

2nd - Questions and Answers in PDF Format

DP-200 Practice Exam Features:

- * DP-200 Questions and Answers Updated Frequently
- * DP-200 Practice Questions Verified by Expert Senior Certified Staff
- * DP-200 Most Realistic Questions that Guarantee you a Pass on Your FirstTry
- * DP-200 Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year

100% Actual & Verified — Instant Download, Please Click
[Order The DP-200 Practice Test Here](#)