



Microsoft

Exam Questions DP-201

Designing an Azure Data Solution

About ExamBible

[Your Partner of IT Exam](#)

Found in 1998

ExamBible is a company specialized on providing high quality IT exam practice study materials, especially Cisco CCNA, CCDA, CCNP, CCIE, Checkpoint CCSE, CompTIA A+, Network+ certification practice exams and so on. We guarantee that the candidates will not only pass any IT exam at the first attempt but also get profound understanding about the certificates they have got. There are so many alike companies in this industry, however, ExamBible has its unique advantages that other companies could not achieve.

Our Advances

* 99.9% Uptime

All examinations will be up to date.

* 24/7 Quality Support

We will provide service round the clock.

* 100% Pass Rate

Our guarantee that you will pass the exam.

* Unique Gurantee

If you do not pass the exam at the first time, we will not only arrange FULL REFUND for you, but also provide you another exam of your claim, ABSOLUTELY FREE!

NEW QUESTION 1

- (Exam Topic 1)

You need to design the Planning Assistance database.

For each of the following statements, select Yes if the statement is true. Otherwise, select No. NOTE: Each correct selection is worth one point.

Statement	Yes	No
Including a clustered columnstore index in the design will benefit performance.	<input type="radio"/>	<input type="radio"/>
Including a nonclustered columnstore index in the design will benefit performance.	<input type="radio"/>	<input type="radio"/>
Including an index on the License Plate column will benefit performance.	<input type="radio"/>	<input type="radio"/>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: No
Data used for Planning Assistance must be stored in a sharded Azure SQL Database. Box 2: Yes
Box 3: Yes
Planning Assistance database will include reports tracking the travel of a single vehicle

NEW QUESTION 2

- (Exam Topic 1)

You need to design the vehicle images storage solution. What should you recommend?

- A. Azure Media Services
- B. Azure Premium Storage account
- C. Azure Redis Cache
- D. Azure Cosmos DB

Answer: B

Explanation:

Premium Storage stores data on the latest technology Solid State Drives (SSDs) whereas Standard Storage stores data on Hard Disk Drives (HDDs). Premium Storage is designed for Azure Virtual Machine workloads which require consistent high IO performance and low latency in order to host IO intensive workloads like OLTP, Big Data, and Data Warehousing on platforms like SQL Server, MongoDB, Cassandra, and others. With Premium Storage, more customers will be able to lift-and-shift demanding enterprise applications to the cloud.
Scenario: Traffic sensors will occasionally capture an image of a vehicle for debugging purposes. You must optimize performance of saving/storing vehicle images. The impact of vehicle images on sensor data throughout must be minimized. References:
<https://azure.microsoft.com/es-es/blog/introducing-premium-storage-high-performance-storage-for-azure-virtual>

NEW QUESTION 3

- (Exam Topic 1)

You need to ensure that performance requirements for Backtrack reports are met.

What should you recommend? To answer, drag the appropriate technologies to the correct locations. 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	Answer Area	
	Requirement	Technology
Cosmos DB TTL	Backtrack reporting	<input type="text"/>
Cosmos DB indexes	Privacy and security policy	<input type="text"/>
Cosmos DB transactions		
Cosmos DB change feed		
Cosmos DB stored procedures		

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Cosmos DB indexes

The report for Backtrack must execute as quickly as possible.

You can override the default indexing policy on an Azure Cosmos container, this could be useful if you want to tune the indexing precision to improve the query performance or to reduce the consumed storage.

Box 2: Cosmos DB TTL

This solution reports on all data related to a specific vehicle license plate. The report must use data from the SensorData collection. Users must be able to filter vehicle data in the following ways:

- vehicles on a specific road
- vehicles driving above the speed limit

Note: With Time to Live or TTL, Azure Cosmos DB provides the ability to delete items automatically from a container after a certain time period. By default, you can set time to live at the container level and override the value on a per-item basis. After you set the TTL at a container or at an item level, Azure Cosmos DB will automatically remove these items after the time period, since the time they were last modified.

NEW QUESTION 4

- (Exam Topic 1)

STION NO: 5 HOTSPOT

You need to design the authentication and authorization methods for sensors.

What should you recommend? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Requirement	Method
Authentication	<div><div>HMAC header</div><div>Resource Token</div><div>Azure Managed Identity</div><div>Storage account connection string</div></div>
Authorization	<div><div>Custom RBAC role</div><div>Cosmos DB user</div><div>Azure Active Directory user</div><div>IoT device identity</div></div>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Sensor data must be stored in a Cosmos DB named treydata in a collection named SensorData Sensors must have permission only to add items to the SensorData collection

Box 1: Resource Token

Resource tokens provide access to the application resources within a Cosmos DB database.

Enable clients to read, write, and delete resources in the Cosmos DB account according to the permissions they've been granted.

Box 2: Cosmos DB user

You can use a resource token (by creating Cosmos DB users and permissions) when you want to provide access to resources in your Cosmos DB account to a client that cannot be trusted with the master key.

References:

<https://docs.microsoft.com/en-us/azure/cosmos-db/secure-access-to-data>

NEW QUESTION 5

- (Exam Topic 2)

You need to design the image processing solution to meet the optimization requirements for image tag data. What should you configure? To answer, drag the appropriate setting to the correct drop targets.

Each source 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.

	Location	Configuration
Configurations	New York	<div></div>
	Manchester	<div></div>
	Singapore	<div></div>
	Melbourne	<div></div>

Write region

Read region

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Tagging data must be uploaded to the cloud from the New York office location.
Tagging data must be replicated to regions that are geographically close to company office locations.

NEW QUESTION 6

- (Exam Topic 2)

You need to design a backup solution for the processed customer data. What should you include in the design?

- A. AzCopy
- B. AdlCopy
- C. Geo-Redundancy
- D. Geo-Replication

Answer: C

Explanation:

Scenario: All data must be backed up in case disaster recovery is required.

Geo-redundant storage (GRS) is designed to provide at least 99.99999999999999% (16 9's) durability of objects over a given year by replicating your data to a secondary region that is hundreds of miles away from the primary region. If your storage account has GRS enabled, then your data is durable even in the case of a complete regional outage or a disaster in which the primary region isn't recoverable. References:

<https://docs.microsoft.com/en-us/azure/storage/common/storage-redundancy-grs>

NEW QUESTION 7

- (Exam Topic 3)

You need to design the disaster recovery solution for customer sales data analytics.

Which three actions should you recommend? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. Provision multiple Azure Databricks workspaces in separate Azure regions.
- B. Migrate users, notebooks, and cluster configurations from one workspace to another in the same region.
- C. Use zone redundant storage.
- D. Migrate users, notebooks, and cluster configurations from one region to another.
- E. Use Geo-redundant storage.
- F. Provision a second Azure Databricks workspace in the same region.

Answer: ADE

Explanation:

Scenario: The analytics solution for customer sales data must be available during a regional outage. To create your own regional disaster recovery topology for databricks, follow these requirements:

1. Provision multiple Azure Databricks workspaces in separate Azure regions
2. Use Geo-redundant storage.
3. Once the secondary region is created, you must migrate the users, user folders, notebooks, cluster configuration, jobs configuration, libraries, storage, init scripts, and reconfigure access control.

Note: Geo-redundant storage (GRS) is designed to provide at least 99.99999999999999% (16 9's) durability of objects over a given year by replicating your data to a secondary region that is hundreds of miles away from the primary region. If your storage account has GRS enabled, then your data is durable even in the case of a complete regional outage or a disaster in which the primary region isn't recoverable.

References:

<https://docs.microsoft.com/en-us/azure/storage/common/storage-redundancy-grs>

NEW QUESTION 8

- (Exam Topic 3)

You need to recommend the appropriate storage and processing solution? What should you recommend?

- A. Enable auto-shrink on the database.
- B. Flush the blob cache using Windows PowerShell.
- C. Enable Apache Spark RDD (RDD) caching.
- D. Enable Databricks IO (DBIO) caching.
- E. Configure the reading speed using Azure Data Studio.

Answer: C

Explanation:

Scenario: You must be able to use a file system view of data stored in a blob. You must build an architecture that will allow Contoso to use the DB FS filesystem layer over a blob store.

Databricks File System (DBFS) is a distributed file system installed on Azure Databricks clusters. Files in DBFS persist to Azure Blob storage, so you won't lose data even after you terminate a cluster.

The Databricks Delta cache, previously named Databricks IO (DBIO) caching, accelerates data reads by creating copies of remote files in nodes' local storage using a fast intermediate data format. The data is cached automatically whenever a file has to be fetched from a remote location. Successive reads of the same data are then performed locally, which results in significantly improved reading speed.

NEW QUESTION 9

- (Exam Topic 3)

You need to recommend an Azure SQL Database service tier. What should you recommend?

- A. Business Critical
- B. General Purpose
- C. Premium
- D. Standard

E. Basic

Answer: C

Explanation:

The data engineers must set the SQL Data Warehouse compute resources to consume 300 DWUs. Note: There are three architectural models that are used in Azure SQL Database:

- ▶ General Purpose/Standard
- ▶ Business Critical/Premium
- ▶ Hyperscale

NEW QUESTION 10

- (Exam Topic 4)

A company purchases IoT devices to monitor manufacturing machinery. The company uses an IoT appliance to communicate with the IoT devices.

The company must be able to monitor the devices in real-time. You need to design the solution.

What should you recommend?

- A. Azure Stream Analytics cloud job using Azure PowerShell
- B. Azure Analysis Services using Azure Portal
- C. Azure Data Factory instance using Azure Portal
- D. Azure Analysis Services using Azure PowerShell

Answer: D

NEW QUESTION 10

- (Exam Topic 4)

You need to design the unauthorized data usage detection system. What Azure service should you include in the design?

- A. Azure Databricks
- B. Azure SQL Data Warehouse
- C. Azure Analysis Services
- D. Azure Data Factory

Answer: B

NEW QUESTION 15

- (Exam Topic 4)

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 is developing a solution to manage inventory data for a group of automotive repair shops. The solution will use Azure SQL Data Warehouse as the data store.

Shops will upload data every 10 days.

Data corruption checks must run each time data is uploaded. If corruption is detected, the corrupted data must be removed.

You need to ensure that upload processes and data corruption checks do not impact reporting and analytics processes that use the data warehouse.

Proposed solution: Configure database-level auditing in Azure SQL Data Warehouse and set retention to 10 days.

Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

Instead, create a user-defined restore point before data is uploaded. Delete the restore point after data corruption checks complete.

References:

<https://docs.microsoft.com/en-us/azure/sql-data-warehouse/backup-and-restore>

NEW QUESTION 17

- (Exam Topic 4)

You are designing a solution for a company. You plan to use Azure Databricks. You need to recommend workloads and tiers to meet the following requirements:

- ▶ Provide managed clusters for running production jobs.
- ▶ Provide persistent clusters that support auto-scaling for analytics processes.
- ▶ Provide role-based access control (RBAC) support for Notebooks.

What should you recommend? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Requirement	Workload	Tier
Provide managed clusters for running production jobs.	<div><div></div><div>Data Engineering only</div><div>Data Analytics only</div><div>Data Engineering and Data Analytics</div></div>	Standard
Provide persistent clusters that support auto-scaling for analytics processes.	<div><div></div><div>Data Engineering only</div><div>Data Analytics only</div><div>Data Engineering and Data Analytics</div></div>	<div><div></div><div>Standard</div><div>Premium</div></div>
Provide role-based access control (RBAC) support for Notebooks.	<div><div></div><div>Data Engineering only</div><div>Data Analytics only</div><div>Data Engineering and Data Analytics</div></div>	<div><div></div><div>Standard</div><div>Premium</div></div>

- A. Mastered
B. Not Mastered

Answer: A

Explanation:

Box 1: Data Engineering Only

Box 2: Data Engineering and Data Analytics Box 3: Standard

Box 4: Data Analytics only Box 5: Premium

Premium required for RBAC. Data Analytics Premium Tier provide interactive workloads to analyze data collaboratively with notebooks

References:

<https://azure.microsoft.com/en-us/pricing/details/databricks/>

NEW QUESTION 20

- (Exam Topic 4)

You are designing a data processing solution that will implement the lambda architecture pattern. The solution will use Spark running on HDInsight for data processing.

You need to recommend a data storage technology for the solution.

Which two technologies should you recommend? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

- A. Azure Cosmos DB
B. Azure Service Bus
C. Azure Storage Queue
D. Apache Cassandra
E. Kafka HDInsight

Answer: AE

Explanation:

To implement a lambda architecture on Azure, you can combine the following technologies to accelerate realtime big data analytics:

Azure Cosmos DB, the industry's first globally distributed, multi-model database service.

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

Azure Cosmos DB change feed, which streams new data to the batch layer for HDInsight to process The Spark to Azure Cosmos DB Connector

E: You can use Apache Spark to stream data into or out of Apache Kafka on HDInsight using DStreams. References:

<https://docs.microsoft.com/en-us/azure/cosmos-db/lambda-architecture>

NEW QUESTION 25

- (Exam Topic 4)

You are designing a recovery strategy for your Azure SQL Databases.

The recovery strategy must use default automated backup settings. The solution must include a Point-in time restore recovery strategy.

You need to recommend which backups to use and the order in which to restore backups.

What should you recommend? To answer, select the appropriate configuration in the answer area.

NOTE: Each correct selection is worth one point.

Restore order	Backup type
first	<div><div></div><div></div><div></div><div></div><div></div></div> <div><div>full weekly backup</div><div>full daily backup</div><div>differential weekly backup</div><div>differential daily backup</div></div>
second	<div><div></div><div></div><div></div><div></div><div></div></div> <div><div>full daily backup</div><div>differential backup from the last 12 hours</div><div>all differential backups since the last full backup</div><div>all log backups since the last full backup</div></div>
third	<div><div></div><div></div><div></div><div></div><div></div></div> <div><div>all log backups since the last differential backup</div><div>differential backup from the last 12 hours</div><div>all differential backups since the last full backup</div><div>all log backups since the last full backup</div></div>

- A. Mastered
B. Not Mastered

Answer: A

Explanation:

All Basic, Standard, and Premium databases are protected by automatic backups. Full backups are taken every week, differential backups every day, and log backups every 5 minutes.

References:

<https://azure.microsoft.com/sv-se/blog/azure-sql-database-point-in-time-restore/>

NEW QUESTION 28

- (Exam Topic 4)

You are designing an Azure Databricks interactive cluster.

You need to ensure that the cluster meets the following requirements: Enable auto-termination

Retain cluster configuration indefinitely after cluster termination. What should you recommend?

- A. Start the cluster after it is terminated.
B. Pin the cluster
C. Clone the cluster after it is terminated.
D. Terminate the cluster manually at process completion.

Answer: B

Explanation:

To keep an interactive cluster configuration even after it has been terminated for more than 30 days, an administrator can pin a cluster to the cluster list.

References:

<https://docs.azuredatabricks.net/user-guide/clusters/terminate.html>

NEW QUESTION 33

- (Exam Topic 4)

You are designing a data processing solution that will run as a Spark job on an HDInsight cluster. The solution will be used to provide near real-time information about online ordering for a retailer.

The solution must include a page on the company intranet that displays summary information. The summary information page must meet the following requirements:

- Display a summary of sales to date grouped by product categories, price range, and review scope.
- Display sales summary information including total sales, sales as compared to one day ago and sales as compared to one year ago.
- Reflect information for new orders as quickly as possible. You need to recommend a design for the solution.

What should you recommend? To answer, select the appropriate configuration in the answer area.

Use case	Technology
Data abstraction	<div><div></div><div></div><div></div><div></div><div></div></div> <div><div>Resilient Distributed Dataset (RDD)</div><div>Dataset</div><div>DataFrame</div></div>
Data format	<div><div></div><div></div><div></div><div></div><div></div></div> <div><div>Avro</div><div>parquet</div></div>

- A. Mastered
B. Not Mastered

Answer: A

Explanation:

Box 1: DataFrame

DataFrames

Best choice in most situations.

Provides query optimization through Catalyst. Whole-stage code generation.

Direct memory access.

Low garbage collection (GC) overhead.

Not as developer-friendly as DataSets, as there are no compile-time checks or domain object programming. Box 2: parquet

The best format for performance is parquet with snappy compression, which is the default in Spark 2.x. Parquet stores data in columnar format, and is highly optimized in Spark.

NEW QUESTION 38

- (Exam Topic 4)

You are designing an Azure Databricks cluster that runs user-defined local processes. You need to recommend a cluster configuration that meets the following requirements:

- Minimize query latency.
- Reduce overall costs.
- Maximize the number of users that can run queries on the cluster at the same time. Which cluster type should you recommend?

- A. Standard with Autoscaling
- B. High Concurrency with Auto Termination
- C. High Concurrency with Autoscaling
- D. Standard with Auto Termination

Answer: C

Explanation:

High Concurrency clusters allow multiple users to run queries on the cluster at the same time, while minimizing query latency. Autoscaling clusters can reduce overall costs compared to a statically-sized cluster.

References:

<https://docs.azuredatabricks.net/user-guide/clusters/create.html> <https://docs.azuredatabricks.net/user-guide/clusters/high-concurrency.html#high-concurrency>

<https://docs.azuredatabricks.net/user-guide/clusters/terminate.html> <https://docs.azuredatabricks.net/user-guide/clusters/sizing.html#enable-and-configure-autoscaling>

NEW QUESTION 39

- (Exam Topic 4)

A company is designing a solution that uses Azure Databricks.

The solution must be resilient to regional Azure datacenter outages. You need to recommend the redundancy type for the solution. What should you recommend?

- A. Read-access geo-redundant storage
- B. Locally-redundant storage
- C. Geo-redundant storage
- D. Zone-redundant storage

Answer: C

Explanation:

If your storage account has GRS enabled, then your data is durable even in the case of a complete regional outage or a disaster in which the primary region isn't recoverable.

References:

<https://medium.com/microsoftazure/data-durability-fault-tolerance-resilience-in-azure-databricks-95392982bac7>

NEW QUESTION 41

- (Exam Topic 4)

You design data engineering solutions for a company.

A project requires analytics and visualization of large set of data. The project has the following requirements:

- ▶ Notebook scheduling
- ▶ Cluster automation
- ▶ Power BI Visualization

You need to recommend the appropriate Azure service. Which Azure service should you recommend?

- A. Azure Batch
- B. Azure Stream Analytics
- C. Azure ML Studio
- D. Azure Databricks
- E. Azure HDInsight

Answer: D

Explanation:

A databrick job is a way of running a notebook or JAR either immediately or on a scheduled basis.

Azure Databricks has two types of clusters: interactive and job. Interactive clusters are used to analyze data collaboratively with interactive notebooks. Job clusters are used to run fast and robust automated workloads using the UI or API.

You can visualize Data with Azure Databricks and Power BI Desktop.

References:

<https://docs.azuredatabricks.net/user-guide/clusters/index.html> <https://docs.azuredatabricks.net/user-guide/jobs.html>

NEW QUESTION 43

- (Exam Topic 4)

You need to design the storage for the telemetry capture system. What storage solution should you use in the design?

- A. Azure Databricks
- B. Azure SQL Data Warehouse
- C. Azure Cosmos DB

Answer: C

NEW QUESTION 47

- (Exam Topic 4)

You are evaluating data storage solutions to support a new application.

You need to recommend a data storage solution that represents data by using nodes and relationships in graph structures.

Which data storage solution should you recommend?

- A. Blob Storage
- B. Cosmos DB
- C. Data Lake Store
- D. HDInsight

Answer: B

Explanation:

For large graphs with lots of entities and relationships, you can perform very complex analyses very quickly. Many graph databases provide a query language that you can use to traverse a network of relationships efficiently.

Relevant Azure service: Cosmos DB

References:

<https://docs.microsoft.com/en-us/azure/architecture/guide/technology-choices/data-store-overview>

NEW QUESTION 48

- (Exam Topic 4)

You are designing an application. You plan to use Azure SQL Database to support the application.

The application will extract data from the Azure SQL Database and create text documents. The text documents will be placed into a cloud-based storage solution.

The text storage solution must be accessible from an SMB network share.

You need to recommend a data storage solution for the text documents. Which Azure data storage type should you recommend?

- A. Queue
- B. Files
- C. Blob
- D. Table

Answer: B

Explanation:

Azure Files enables you to set up highly available network file shares that can be accessed by using the standard Server Message Block (SMB) protocol.

References:

<https://docs.microsoft.com/en-us/azure/storage/common/storage-introduction> <https://docs.microsoft.com/en-us/azure/storage/tables/table-storage-overview>

NEW QUESTION 50

- (Exam Topic 4)

A company is developing a mission-critical line of business app that uses Azure SQL Database Managed Instance. You must design a disaster recovery strategy for the solution.

You need to ensure that the database automatically recovers when full or partial loss of the Azure SQL Database service occurs in the primary region.

What should you recommend?

- A. Failover-group
- B. Azure SQL Data Sync
- C. SQL Replication
- D. Active geo-replication

Answer: A

Explanation:

Auto-failover groups is a SQL Database feature that allows you to manage replication and failover of a group of databases on a SQL Database server or all databases in a Managed Instance to another region (currently in public preview for Managed Instance). It uses the same underlying technology as active geo-replication. You can initiate failover manually or you can delegate it to the SQL Database service based on a user-defined policy.

References:

<https://docs.microsoft.com/en-us/azure/sql-database/sql-database-auto-failover-group>

NEW QUESTION 55

- (Exam Topic 4)

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.

You are designing an Azure SQL Database that will use elastic pools. You plan to store data about customers in a table. Each record uses a value for CustomerID.

You need to recommend a strategy to partition data based on values in CustomerID. Proposed Solution: Separate data into customer regions by using horizontal partitioning. Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

We should use Horizontal Partitioning through Sharding, not divide through regions.

Note: Horizontal Partitioning - Sharding: Data is partitioned horizontally to distribute rows across a scaled out data tier. With this approach, the schema is identical on all participating databases. This approach is also called “sharding”. Sharding can be performed and managed using (1) the elastic database tools libraries or (2)

self-sharding. An elastic query is used to query or compile reports across many shards.

References:

<https://docs.microsoft.com/en-us/azure/sql-database/sql-database-elastic-query-overview>

NEW QUESTION 59

- (Exam Topic 4)

You are designing a Spark job that performs batch processing of daily web log traffic.

When you deploy the job in the production environment, it must meet the following requirements:

- Run once a day.
- Display status information on the company intranet as the job runs. You need to recommend technologies for triggering and monitoring jobs.

Which technologies should you recommend? To answer, drag the appropriate technologies to the correct locations. 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	Requirement	Technology
Livy	Triggering of jobs	
Beeline	Monitoring of jobs	
Azure Logic App		
Azure API App		

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Livy

You can use Livy to run interactive Spark shells or submit batch jobs to be run on Spark. Box 2: Beeline

Apache Beeline can be used to run Apache Hive queries on HDInsight. You can use Beeline with Apache Spark.

Note: Beeline is a Hive client that is included on the head nodes of your HDInsight cluster. Beeline uses JDBC to connect to HiveServer2, a service hosted on your HDInsight cluster. You can also use Beeline to access Hive on HDInsight remotely over the internet.

References:

<https://docs.microsoft.com/en-us/azure/hdinsight/spark/apache-spark-livy-rest-interface> <https://docs.microsoft.com/en-us/azure/hdinsight/hadoop/apache-hadoop-use-hive-beeline>

NEW QUESTION 61

- (Exam Topic 4)

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.

You are designing an HDInsight/Hadoop cluster solution that uses Azure Data Lake Gen1 Storage. The solution requires POSIX permissions and enables diagnostics logging for auditing.

You need to recommend solutions that optimize storage.

Proposed Solution: Ensure that files stored are smaller than 250MB. Does the solution meet the goal?

- A. Yes
- B. No

Answer: B

Explanation:

Ensure that files stored are larger, not smaller than 250MB.

You can have a separate compaction job that combines these files into larger ones.

Note: The file POSIX permissions and auditing in Data Lake Storage Gen1 comes with an overhead that becomes apparent when working with numerous small files. As a best practice, you must batch your data into larger files versus writing thousands or millions of small files to Data Lake Storage Gen1. Avoiding small file sizes can have multiple benefits, such as:

Lowering the authentication checks across multiple files Reduced open file connections

Faster copying/replication

Fewer files to process when updating Data Lake Storage Gen1 POSIX permissions References:

<https://docs.microsoft.com/en-us/azure/data-lake-store/data-lake-store-best-practices>

NEW QUESTION 65

.....

Relate Links

100% Pass Your DP-201 Exam with Examible Prep Materials

<https://www.exambible.com/DP-201-exam/>

Contact us

We are proud of our high-quality customer service, which serves you around the clock 24/7.

Viste - <https://www.exambible.com/>