



Microsoft

Exam Questions DP-201

Designing an Azure Data Solution

NEW QUESTION 1

- (Exam Topic 1)

HOTSPOT

You need to ensure that security policies for the unauthorized detection system are met. What should you recommend? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Setting	Value
Audit log destination	<div><div></div><div>Storage queue</div><div>Event Hub</div><div>Event Grid</div><div>Blob storage</div></div>
Detection app service	<div><div></div><div>Function App</div><div>Web App</div><div>API App</div></div>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Blob storage

Configure blob storage for audit logs.

Scenario: Unauthorized usage of the Planning Assistance data must be detected as quickly as possible. Unauthorized usage is determined by looking for an unusual pattern of usage.

Data used for Planning Assistance must be stored in a sharded Azure SQL Database. Box 2: Web Apps

SQL Advanced Threat Protection (ATP) is to be used.

One of Azure's most popular service is App Service which enables customers to build and host web applications in the programming language of their choice without managing infrastructure. App Service offers auto-scaling and high availability, supports both Windows and Linux. It also supports automated deployments from GitHub, Visual Studio Team Services or any Git repository. At RSA, we announced that Azure Security Center leverages the scale of the cloud to identify attacks targeting App Service applications.

References:

<https://azure.microsoft.com/sv-se/blog/azure-security-center-can-identify-attacks-targeting-azure-app-service-ap>

NEW QUESTION 2

- (Exam Topic 1)

You need to design a sharding strategy for the Planning Assistance database. What should you recommend?

- A. a list mapping shard map on the binary representation of the License Plate column
- B. a range mapping shard map on the binary representation of the speed column
- C. a list mapping shard map on the location column
- D. a range mapping shard map on the time column

Answer: A

Explanation:

Data used for Planning Assistance must be stored in a sharded Azure SQL Database.

A shard typically contains items that fall within a specified range determined by one or more attributes of the data. These attributes form the shard key (sometimes referred to as the partition key). The shard key should be static. It shouldn't be based on data that might change.

References:

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

NEW QUESTION 3

- (Exam Topic 2)

You plan to use an Azure SQL data warehouse to store the customer data. You need to recommend a disaster recovery solution for the data warehouse. What should you include in the recommendation?

- A. AzCopy
- B. Read-only replicas
- C. AdlCopy
- D. Geo-Redundant backups

Answer: D

Explanation:

References:

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

NEW QUESTION 4

- (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 5

- (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 6

- (Exam Topic 3)

You need to optimize storage for CONT_SQL3. What should you recommend?

- A. AlwaysOn
- B. Transactional processing
- C. General
- D. Data warehousing

Answer: B

Explanation:

CONT_SQL3 with the SQL Server role, 100 GB database size, Hyper-VM to be migrated to Azure VM. The storage should be configured to optimized storage for database OLTP workloads.

Azure SQL Database provides three basic in-memory based capabilities (built into the underlying database engine) that can contribute in a meaningful way to performance improvements:

In-Memory Online Transactional Processing (OLTP)

Clustered columnstore indexes intended primarily for Online Analytical Processing (OLAP) workloads Nonclustered columnstore indexes geared towards Hybrid Transactional/Analytical Processing (HTAP) workloads

References:

<https://www.databasejournal.com/features/mssql/overview-of-in-memory-technologies-of-azure-sqldatabase.htm>

NEW QUESTION 7

- (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 8

- (Exam Topic 3)

You are designing an Azure SQL Data Warehouse for a financial services company. Azure Active Directory will be used to authenticate the users.

You need to ensure that the following security requirements are met:

- ▶ Department managers must be able to create new database.
- ▶ The IT department must assign users to databases.
- ▶ Permissions granted must be minimized.

Which role memberships should you recommend? To answer, drag the appropriate roles to the correct groups. Each role 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.

Roles	Group	Role
dbmanager	Department managers	
loginmanager		
dc_admin	IT	
db_securityadmin		
db_owner		
db_accessadmin		

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: dbmanager

Members of the dbmanager role can create new databases. Box 2: db_accessadmin

Members of the db_accessadmin fixed database role can add or remove access to the database for Windows logins, Windows groups, and SQL Server logins.

References:

<https://docs.microsoft.com/en-us/azure/sql-database/sql-database-manage-logins>

NEW QUESTION 9

- (Exam Topic 4)

A company manufactures automobile parts. The company installs IoT sensors on manufacturing machinery. You must design a solution that analyzes data from the sensors.

You need to recommend a solution that meets the following requirements: Data must be analyzed in real-time.

Data queries must be deployed using continuous integration. Data must be visualized by using charts and graphs.

Data must be available for ETL operations in the future. The solution must support high-volume data ingestion.

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

NOTE: Each correct selection is worth one point.

- A. Use Azure Analysis Services to query the dat
- B. Output query results to Power BI.
- C. Configure an Azure Event Hub to capture data to Azure Data Lake Storage.
- D. Develop an Azure Stream Analytics application that queries the data and outputs to Power B
- E. Use AzureData Factory to deploy the Azure Stream Analytics application.
- F. Develop an application that sends the IoT data to an Azure Event Hub.
- G. Develop an Azure Stream Analytics application that queries the data and outputs to Power B
- H. Use AzurePipelines to deploy the Azure Stream Analytics application.
- I. Develop an application that sends the IoT data to an Azure Data Lake Storage container.

Answer: BCD

NEW QUESTION 10

- (Exam Topic 4)

You design data engineering solutions for a company.

You must integrate on-premises SQL Server data into an Azure solution that performs Extract-Transform-Load (ETL) operations have the following requirements:

- ▶ Develop a pipeline that can integrate data and run notebooks.
- ▶ Develop notebooks to transform the data.
- ▶ Load the data into a massively parallel processing database for later analysis. You need to recommend a solution.

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

NOTE: Each correct selection is worth one point.

Requirement	Service
Integrate the on-premises data into the cloud.	<div>▼</div> <div> Azure Databricks Azure Data Factory Azure SQL Data Warehouse Azure Batch </div>
Develop notebooks to transform the data.	<div>▼</div> <div> Azure Databricks Azure Data Factory Azure SQL Data Warehouse Azure Batch </div>
Run notebooks.	<div>▼</div> <div> Azure Databricks Azure Data Factory Azure SQL Data Warehouse Azure Batch </div>
Load the data.	<div>▼</div> <div> Azure Databricks Azure Data Factory Azure SQL Data Warehouse Azure Batch </div>
Store the transformed data.	<div>▼</div> <div> Azure Databricks Azure Data Factory Azure SQL Data Warehouse Azure Batch </div>

- A. Mastered
 B. Not Mastered

Answer: A

Explanation:

Requirement	Service
Integrate the on-premises data into the cloud.	<div>▼</div> <div> Azure Databricks Azure Data Factory Azure SQL Data Warehouse Azure Batch </div>
Develop notebooks to transform the data.	<div>▼</div> <div> Azure Databricks Azure Data Factory Azure SQL Data Warehouse Azure Batch </div>
Run notebooks.	<div>▼</div> <div> Azure Databricks Azure Data Factory Azure SQL Data Warehouse Azure Batch </div>
Load the data.	<div>▼</div> <div> Azure Databricks Azure Data Factory Azure SQL Data Warehouse Azure Batch </div>
Store the transformed data.	<div>▼</div> <div> Azure Databricks Azure Data Factory Azure SQL Data Warehouse Azure Batch </div>

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 14

- (Exam Topic 4)

A company has many applications. Each application is supported by separate on-premises databases. You must migrate the databases to Azure SQL Database. You have the following requirements: Organize databases into groups based on database usage.

Define the maximum resource limit available for each group of databases.

You need to recommend technologies to scale the databases to support expected increases in demand. What should you recommend?

- A. Read scale-out
- B. Managed instances
- C. Elastic pools
- D. Database sharding

Answer: C

Explanation:

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.

You can configure resources for the pool based either on the DTU-based purchasing model or the vCorebased purchasing model.

NEW QUESTION 15

- (Exam Topic 4)

You plan to migrate data to Azure SQL Database.

The database must remain synchronized with updates to Microsoft Azure and SQL Server. You need to set up the database as a subscriber.

What should you recommend?

- A. Azure Data Factory
- B. SQL Server Data Tools
- C. Data Migration Assistant
- D. SQL Server Agent for SQL Server 2017 or later
- E. SQL Server Management Studio 17.9.1 or later

Answer: E

Explanation:

To set up the database as a subscriber we need to configure database replication. You can use SQL Server Management Studio to configure replication. Use the latest versions of SQL Server Management Studio in order to be able to use all the features of Azure SQL Database.

References:

<https://www.sqlshack.com/sql-server-database-migration-to-azure-sql-database-using-sql-server-transactionalrep>

NEW QUESTION 19

- (Exam Topic 4)

A company is evaluating data storage solutions.

You need to recommend a data storage solution that meets the following requirements: Minimize costs for storing blob objects.

Optimize access for data that is infrequently accessed. Data must be stored for at least 30 days.

Data availability must be at least 99 percent. What should you recommend?

- A. Premium
- B. Cold
- C. Hot
- D. Archive

Answer: B

Explanation:

Azure's cool storage tier, also known as Azure cool Blob storage, is for infrequently-accessed data that needs to be stored for a minimum of 30 days. Typical use cases include backing up data before tiering to archival systems, legal data, media files, system audit information, datasets used for big data analysis and more.

The storage cost for this Azure cold storage tier is lower than that of hot storage tier. Since it is expected that the data stored in this tier will be accessed less frequently, the data access charges are high when compared to hot tier. There are no additional changes required in your applications as these tiers can be accessed using

APIs in the same manner that you access Azure storage. References:

<https://cloud.netapp.com/blog/low-cost-storage-options-on-azure>

NEW QUESTION 21

- (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: Create a user-defined restore point before data is uploaded. Delete the restore point after data corruption checks complete.
Does the solution meet the goal?

- A. Yes
- B. No

Answer: A

Explanation:

User-Defined Restore Points

This feature enables you to manually trigger snapshots to create restore points of your data warehouse before and after large modifications. This capability ensures that restore points are logically consistent, which provides additional data protection in case of any workload interruptions or user errors for quick recovery time.

Note: A data warehouse restore is a new data warehouse that is created from a restore point of an existing or deleted data warehouse. Restoring your data warehouse is an essential part of any business continuity and disaster recovery strategy because it re-creates your data after accidental corruption or deletion.

References:

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

NEW QUESTION 22

- (Exam Topic 4)

You are designing an Azure SQL Data Warehouse. You plan to load millions of rows of data into the data warehouse each day.

You must ensure that staging tables are optimized for data loading. You need to design the staging tables.

What type of tables should you recommend?

- A. Round-robin distributed table
- B. Hash-distributed table
- C. Replicated table
- D. External table

Answer: A

Explanation:

To achieve the fastest loading speed for moving data into a data warehouse table, load data into a staging table. Define the staging table as a heap and use round-robin for the distribution option.

References:

<https://docs.microsoft.com/en-us/azure/sql-data-warehouse/guidance-for-loading-data>

NEW QUESTION 25

- (Exam Topic 4)

You are designing a solution for a company. The solution will use model training for objective classification. You need to design the solution.

What should you recommend?

- A. an Azure Cognitive Services application
- B. a Spark Streaming job
- C. interactive Spark queries
- D. Power BI models
- E. a Spark application that uses Spark MLlib.

Answer: E

Explanation:

Spark in SQL Server big data cluster enables AI and machine learning.

You can use Apache Spark MLlib to create a machine learning application to do simple predictive analysis on an open dataset.

MLlib is a core Spark library that provides many utilities useful for machine learning tasks, including utilities that are suitable for:

- ▶ Classification
- ▶ Regression
- ▶ Clustering
- ▶ Topic modeling
- ▶ Singular value decomposition (SVD) and principal component analysis (PCA)
- ▶ Hypothesis testing and calculating sample statistics

References:

<https://docs.microsoft.com/en-us/azure/hdinsight/spark/apache-spark-machine-learning-mllib-ipynb>

NEW QUESTION 28

- (Exam Topic 4)

You have an on-premises MySQL database that is 800 GB in size.

You need to migrate a MySQL database to Azure Database for MySQL. You must minimize service interruption to live sites or applications that use the database.

What should you recommend?

- A. Azure Database Migration Service
- B. Dump and restore
- C. Import and export
- D. MySQL Workbench

Answer: A

Explanation:

You can perform MySQL migrations to Azure Database for MySQL with minimal downtime by using the newly introduced continuous sync capability for the Azure Database Migration Service (DMS). This functionality limits the amount of downtime that is incurred by the application. References:

<https://docs.microsoft.com/en-us/azure/mysql/howto-migrate-online>

NEW QUESTION 32

- (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 35

- (Exam Topic 4)

A company has an application that uses Azure SQL Database as the data store.

The application experiences a large increase in activity during the last month of each year.

You need to manually scale the Azure SQL Database instance to account for the increase in data write operations.

Which scaling method should you recommend?

- A. Scale up by using elastic pools to distribute resources.
- B. Scale out by sharding the data across databases.
- C. Scale up by increasing the database throughput units.

Answer: C

Explanation:

As of now, the cost of running an Azure SQL database instance is based on the number of Database Throughput Units (DTUs) allocated for the database. When determining the number of units to allocate for the

solution, a major contributing factor is to identify what processing power is needed to handle the volume of expected requests.

Running the statement to upgrade/downgrade your database takes a matter of seconds.

NEW QUESTION 40

- (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: Insert data from shops and perform the data corruption check in a transaction. Rollback transfer if corruption is detected.

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 45

- (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 47

- (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 51

- (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 52

- (Exam Topic 4)

A company stores large datasets in Azure, including sales transactions and customer account information. You must design a solution to analyze the data. You plan to create the following HDInsight clusters:

You need to ensure that the clusters support the query requirements.

Which cluster types should you recommend? To answer, select the appropriate configuration in the answer area.

NOTE: Each correct selection is worth one point.

Cluster	Cluster type
Sales	<div><div>Storm</div><div>Hadoop</div><div>Interactive Query</div><div>Kafka</div></div>
Accounts	<div><div>Spark</div><div>Hadoop</div><div>Interactive Query</div><div>Kafka</div></div>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Interactive Query

Choose Interactive Query cluster type to optimize for ad hoc, interactive queries. Box 2: Hadoop

Choose Apache Hadoop cluster type to optimize for Hive queries used as a batch process.

Note: In Azure HDInsight, there are several cluster types and technologies that can run Apache Hive queries. When you create your HDInsight cluster, choose the appropriate cluster type to help optimize performance for your workload needs.

For example, choose Interactive Query cluster type to optimize for ad hoc, interactive queries. Choose Apache Hadoop cluster type to optimize for Hive queries used as a batch process. Spark and HBase cluster types can also run Hive queries.

References:

<https://docs.microsoft.com/bs-latn-ba/azure/hdinsight/hdinsight-hadoop-optimize-hive-query?toc=%2Fko-kr%2>

NEW QUESTION 56

- (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
<div>Livy</div>	Triggering of jobs	<div></div>
<div>Beeline</div>	Monitoring of jobs	<div></div>
<div>Azure Logic App</div>		
<div>Azure API App</div>		

- 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 59

- (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 62

.....

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)

HOTSPOT

You need to ensure that security policies for the unauthorized detection system are met. What should you recommend? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Setting	Value
Audit log destination	<div><div></div><div>Storage queue</div><div>Event Hub</div><div>Event Grid</div><div>Blob storage</div></div>
Detection app service	<div><div></div><div>Function App</div><div>Web App</div><div>API App</div></div>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Blob storage

Configure blob storage for audit logs.

Scenario: Unauthorized usage of the Planning Assistance data must be detected as quickly as possible. Unauthorized usage is determined by looking for an unusual pattern of usage.

Data used for Planning Assistance must be stored in a sharded Azure SQL Database. Box 2: Web Apps

SQL Advanced Threat Protection (ATP) is to be used.

One of Azure's most popular service is App Service which enables customers to build and host web applications in the programming language of their choice without managing infrastructure. App Service offers auto-scaling and high availability, supports both Windows and Linux. It also supports automated deployments from GitHub, Visual Studio Team Services or any Git repository. At RSA, we announced that Azure Security Center leverages the scale of the cloud to identify attacks targeting App Service applications.

References:

<https://azure.microsoft.com/sv-se/blog/azure-security-center-can-identify-attacks-targeting-azure-app-service-ap>

NEW QUESTION 2

- (Exam Topic 1)

You need to design a sharding strategy for the Planning Assistance database. What should you recommend?

- A. a list mapping shard map on the binary representation of the License Plate column
- B. a range mapping shard map on the binary representation of the speed column
- C. a list mapping shard map on the location column
- D. a range mapping shard map on the time column

Answer: A

Explanation:

Data used for Planning Assistance must be stored in a sharded Azure SQL Database.

A shard typically contains items that fall within a specified range determined by one or more attributes of the data. These attributes form the shard key (sometimes referred to as the partition key). The shard key should be static. It shouldn't be based on data that might change.

References:

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

NEW QUESTION 3

- (Exam Topic 2)

You plan to use an Azure SQL data warehouse to store the customer data. You need to recommend a disaster recovery solution for the data warehouse. What should you include in the recommendation?

- A. AzCopy
- B. Read-only replicas
- C. AdlCopy
- D. Geo-Redundant backups

Answer: D

Explanation:

References:

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

NEW QUESTION 4

- (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 5

- (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 6

- (Exam Topic 3)

You need to optimize storage for CONT_SQL3. What should you recommend?

- A. AlwaysOn
- B. Transactional processing
- C. General
- D. Data warehousing

Answer: B

Explanation:

CONT_SQL3 with the SQL Server role, 100 GB database size, Hyper-VM to be migrated to Azure VM. The storage should be configured to optimized storage for database OLTP workloads.

Azure SQL Database provides three basic in-memory based capabilities (built into the underlying database engine) that can contribute in a meaningful way to performance improvements:

In-Memory Online Transactional Processing (OLTP)

Clustered columnstore indexes intended primarily for Online Analytical Processing (OLAP) workloads Nonclustered columnstore indexes geared towards Hybrid Transactional/Analytical Processing (HTAP) workloads

References:

<https://www.databasejournal.com/features/mssql/overview-of-in-memory-technologies-of-azure-sqldatabase.htm>

NEW QUESTION 7

- (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 8

- (Exam Topic 3)

You are designing an Azure SQL Data Warehouse for a financial services company. Azure Active Directory will be used to authenticate the users.

You need to ensure that the following security requirements are met:

- ▶ Department managers must be able to create new database.
- ▶ The IT department must assign users to databases.
- ▶ Permissions granted must be minimized.

Which role memberships should you recommend? To answer, drag the appropriate roles to the correct groups. Each role 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.

Roles	Group	Role
dbmanager	Department managers	
loginmanager		
dc_admin	IT	
db_securityadmin		
db_owner		
db_accessadmin		

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: dbmanager

Members of the dbmanager role can create new databases. Box 2: db_accessadmin

Members of the db_accessadmin fixed database role can add or remove access to the database for Windows logins, Windows groups, and SQL Server logins.

References:

<https://docs.microsoft.com/en-us/azure/sql-database/sql-database-manage-logins>

NEW QUESTION 9

- (Exam Topic 4)

A company manufactures automobile parts. The company installs IoT sensors on manufacturing machinery. You must design a solution that analyzes data from the sensors.

You need to recommend a solution that meets the following requirements: Data must be analyzed in real-time.

Data queries must be deployed using continuous integration. Data must be visualized by using charts and graphs.

Data must be available for ETL operations in the future. The solution must support high-volume data ingestion.

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

NOTE: Each correct selection is worth one point.

- A. Use Azure Analysis Services to query the dat
- B. Output query results to Power BI.
- C. Configure an Azure Event Hub to capture data to Azure Data Lake Storage.
- D. Develop an Azure Stream Analytics application that queries the data and outputs to Power B
- E. Use AzureData Factory to deploy the Azure Stream Analytics application.
- F. Develop an application that sends the IoT data to an Azure Event Hub.
- G. Develop an Azure Stream Analytics application that queries the data and outputs to Power B
- H. Use AzurePipelines to deploy the Azure Stream Analytics application.
- I. Develop an application that sends the IoT data to an Azure Data Lake Storage container.

Answer: BCD

NEW QUESTION 10

- (Exam Topic 4)

You design data engineering solutions for a company.

You must integrate on-premises SQL Server data into an Azure solution that performs Extract-Transform-Load (ETL) operations have the following requirements:

- ▶ Develop a pipeline that can integrate data and run notebooks.
- ▶ Develop notebooks to transform the data.
- ▶ Load the data into a massively parallel processing database for later analysis. You need to recommend a solution.

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

NOTE: Each correct selection is worth one point.

Requirement	Service
Integrate the on-premises data into the cloud.	<div>▼</div> <div> Azure Databricks Azure Data Factory Azure SQL Data Warehouse Azure Batch </div>
Develop notebooks to transform the data.	<div>▼</div> <div> Azure Databricks Azure Data Factory Azure SQL Data Warehouse Azure Batch </div>
Run notebooks.	<div>▼</div> <div> Azure Databricks Azure Data Factory Azure SQL Data Warehouse Azure Batch </div>
Load the data.	<div>▼</div> <div> Azure Databricks Azure Data Factory Azure SQL Data Warehouse Azure Batch </div>
Store the transformed data.	<div>▼</div> <div> Azure Databricks Azure Data Factory Azure SQL Data Warehouse Azure Batch </div>

- A. Mastered
 B. Not Mastered

Answer: A

Explanation:

Requirement	Service
Integrate the on-premises data into the cloud.	<div>▼</div> <div> Azure Databricks Azure Data Factory Azure SQL Data Warehouse Azure Batch </div>
Develop notebooks to transform the data.	<div>▼</div> <div> Azure Databricks Azure Data Factory Azure SQL Data Warehouse Azure Batch </div>
Run notebooks.	<div>▼</div> <div> Azure Databricks Azure Data Factory Azure SQL Data Warehouse Azure Batch </div>
Load the data.	<div>▼</div> <div> Azure Databricks Azure Data Factory Azure SQL Data Warehouse Azure Batch </div>
Store the transformed data.	<div>▼</div> <div> Azure Databricks Azure Data Factory Azure SQL Data Warehouse Azure Batch </div>

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 14

- (Exam Topic 4)

A company has many applications. Each application is supported by separate on-premises databases. You must migrate the databases to Azure SQL Database. You have the following requirements: Organize databases into groups based on database usage.

Define the maximum resource limit available for each group of databases.

You need to recommend technologies to scale the databases to support expected increases in demand. What should you recommend?

- A. Read scale-out
- B. Managed instances
- C. Elastic pools
- D. Database sharding

Answer: C

Explanation:

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.

You can configure resources for the pool based either on the DTU-based purchasing model or the vCorebased purchasing model.

NEW QUESTION 15

- (Exam Topic 4)

You plan to migrate data to Azure SQL Database.

The database must remain synchronized with updates to Microsoft Azure and SQL Server. You need to set up the database as a subscriber.

What should you recommend?

- A. Azure Data Factory
- B. SQL Server Data Tools
- C. Data Migration Assistant
- D. SQL Server Agent for SQL Server 2017 or later
- E. SQL Server Management Studio 17.9.1 or later

Answer: E

Explanation:

To set up the database as a subscriber we need to configure database replication. You can use SQL Server Management Studio to configure replication. Use the latest versions of SQL Server Management Studio in order to be able to use all the features of Azure SQL Database.

References:

<https://www.sqlshack.com/sql-server-database-migration-to-azure-sql-database-using-sql-server-transactionalrep>

NEW QUESTION 19

- (Exam Topic 4)

A company is evaluating data storage solutions.

You need to recommend a data storage solution that meets the following requirements: Minimize costs for storing blob objects.

Optimize access for data that is infrequently accessed. Data must be stored for at least 30 days.

Data availability must be at least 99 percent. What should you recommend?

- A. Premium
- B. Cold
- C. Hot
- D. Archive

Answer: B

Explanation:

Azure's cool storage tier, also known as Azure cool Blob storage, is for infrequently-accessed data that needs to be stored for a minimum of 30 days. Typical use cases include backing up data before tiering to archival systems, legal data, media files, system audit information, datasets used for big data analysis and more.

The storage cost for this Azure cold storage tier is lower than that of hot storage tier. Since it is expected that the data stored in this tier will be accessed less frequently, the data access charges are high when compared to hot tier. There are no additional changes required in your applications as these tiers can be accessed using

APIs in the same manner that you access Azure storage. References:

<https://cloud.netapp.com/blog/low-cost-storage-options-on-azure>

NEW QUESTION 21

- (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: Create a user-defined restore point before data is uploaded. Delete the restore point after data corruption checks complete.
Does the solution meet the goal?

- A. Yes
- B. No

Answer: A

Explanation:

User-Defined Restore Points

This feature enables you to manually trigger snapshots to create restore points of your data warehouse before and after large modifications. This capability ensures that restore points are logically consistent, which provides additional data protection in case of any workload interruptions or user errors for quick recovery time.

Note: A data warehouse restore is a new data warehouse that is created from a restore point of an existing or deleted data warehouse. Restoring your data warehouse is an essential part of any business continuity and disaster recovery strategy because it re-creates your data after accidental corruption or deletion.

References:

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

NEW QUESTION 22

- (Exam Topic 4)

You are designing an Azure SQL Data Warehouse. You plan to load millions of rows of data into the data warehouse each day.

You must ensure that staging tables are optimized for data loading. You need to design the staging tables.

What type of tables should you recommend?

- A. Round-robin distributed table
- B. Hash-distributed table
- C. Replicated table
- D. External table

Answer: A

Explanation:

To achieve the fastest loading speed for moving data into a data warehouse table, load data into a staging table. Define the staging table as a heap and use round-robin for the distribution option.

References:

<https://docs.microsoft.com/en-us/azure/sql-data-warehouse/guidance-for-loading-data>

NEW QUESTION 25

- (Exam Topic 4)

You are designing a solution for a company. The solution will use model training for objective classification. You need to design the solution.

What should you recommend?

- A. an Azure Cognitive Services application
- B. a Spark Streaming job
- C. interactive Spark queries
- D. Power BI models
- E. a Spark application that uses Spark MLlib.

Answer: E

Explanation:

Spark in SQL Server big data cluster enables AI and machine learning.

You can use Apache Spark MLlib to create a machine learning application to do simple predictive analysis on an open dataset.

MLlib is a core Spark library that provides many utilities useful for machine learning tasks, including utilities that are suitable for:

- ▶ Classification
- ▶ Regression
- ▶ Clustering
- ▶ Topic modeling
- ▶ Singular value decomposition (SVD) and principal component analysis (PCA)
- ▶ Hypothesis testing and calculating sample statistics

References:

<https://docs.microsoft.com/en-us/azure/hdinsight/spark/apache-spark-machine-learning-mllib-ipynb>

NEW QUESTION 28

- (Exam Topic 4)

You have an on-premises MySQL database that is 800 GB in size.

You need to migrate a MySQL database to Azure Database for MySQL. You must minimize service interruption to live sites or applications that use the database.

What should you recommend?

- A. Azure Database Migration Service
- B. Dump and restore
- C. Import and export
- D. MySQL Workbench

Answer: A

Explanation:

You can perform MySQL migrations to Azure Database for MySQL with minimal downtime by using the newly introduced continuous sync capability for the Azure Database Migration Service (DMS). This functionality limits the amount of downtime that is incurred by the application. References:

<https://docs.microsoft.com/en-us/azure/mysql/howto-migrate-online>

NEW QUESTION 32

- (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 35

- (Exam Topic 4)

A company has an application that uses Azure SQL Database as the data store.

The application experiences a large increase in activity during the last month of each year.

You need to manually scale the Azure SQL Database instance to account for the increase in data write operations.

Which scaling method should you recommend?

- A. Scale up by using elastic pools to distribute resources.
- B. Scale out by sharding the data across databases.
- C. Scale up by increasing the database throughput units.

Answer: C

Explanation:

As of now, the cost of running an Azure SQL database instance is based on the number of Database Throughput Units (DTUs) allocated for the database. When determining the number of units to allocate for the

solution, a major contributing factor is to identify what processing power is needed to handle the volume of expected requests.

Running the statement to upgrade/downgrade your database takes a matter of seconds.

NEW QUESTION 40

- (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: Insert data from shops and perform the data corruption check in a transaction. Rollback transfer if corruption is detected.

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 45

- (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 47

- (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 51

- (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 52

- (Exam Topic 4)

A company stores large datasets in Azure, including sales transactions and customer account information. You must design a solution to analyze the data. You plan to create the following HDInsight clusters:

You need to ensure that the clusters support the query requirements.

Which cluster types should you recommend? To answer, select the appropriate configuration in the answer area.

NOTE: Each correct selection is worth one point.

Cluster	Cluster type
Sales	<div><div>Storm</div><div>Hadoop</div><div>Interactive Query</div><div>Kafka</div></div>
Accounts	<div><div>Spark</div><div>Hadoop</div><div>Interactive Query</div><div>Kafka</div></div>

- A. Mastered
- B. Not Mastered

Answer: A

Explanation:

Box 1: Interactive Query

Choose Interactive Query cluster type to optimize for ad hoc, interactive queries. Box 2: Hadoop

Choose Apache Hadoop cluster type to optimize for Hive queries used as a batch process.

Note: In Azure HDInsight, there are several cluster types and technologies that can run Apache Hive queries. When you create your HDInsight cluster, choose the appropriate cluster type to help optimize performance for your workload needs.

For example, choose Interactive Query cluster type to optimize for ad hoc, interactive queries. Choose Apache Hadoop cluster type to optimize for Hive queries used as a batch process. Spark and HBase cluster types can also run Hive queries.

References:

<https://docs.microsoft.com/bs-latn-ba/azure/hdinsight/hdinsight-hadoop-optimize-hive-query?toc=%2Fko-kr%2>

NEW QUESTION 56

- (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
<div>Livy</div>	Triggering of jobs	<div></div>
<div>Beeline</div>	Monitoring of jobs	<div></div>
<div>Azure Logic App</div>		
<div>Azure API App</div>		

- 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 59

- (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 62

.....

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/>