Microsoft

Exam Questions DP-201

Designing an Azure Data Solution

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.	○	○
Including a nonclustered columnstore index in the design will benefit performance.	○	○
Including an index on the License Plate column will benefit performance.	○	○

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
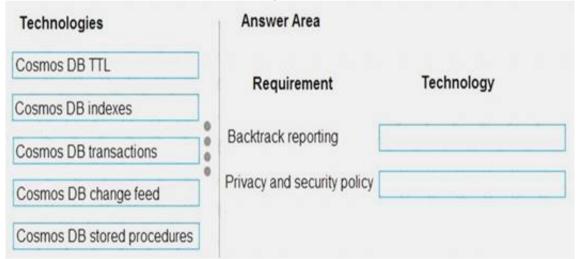Planning Assistance database will include reports tracking the travel of a single vehicle

NEW QUESTION 2
- (Exam Topic 1)
You need to 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

Cosmos DB TTL

Cosmos DB indexes

Cosmos DB transactions

Cosmos DB change feed

Cosmos DB stored procedures

Answer Area

Requirement	Technology
Backtrack reporting	
Privacy and security policy	

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 3
- (Exam Topic 1)
You need to design the SensorData collection.
What should you recommend? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Setting	Value
Default consistency level	strong
	session
	eventual
	consistent prefix
	bounded staleness
Partition key property	Time
	Location
	Speed
	License plate
	Vehicle length

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: Eventual
Traffic data insertion rate must be maximized.
Sensor data must be stored in a Cosmos DB named treydata in a collection named SensorData
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.
Box 2: License plate
This solution reports on all data related to a specific vehicle license plate. The report must use data from the SensorData collection.
References:
https://docs.microsoft.com/en-us/azure/cosmos-db/consistency-levels

NEW QUESTION 4
- (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.

Configurations	Location	Configuration
Write region	New York	
	Manchester	
Read region	Singapore	
	Melbourne	

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 5
- (Exam Topic 2)
You need to design the solution for analyzing customer data. What should you recommend?

A. Azure Databricks
B. Azure Data Lake Storage
C. Azure SQL Data Warehouse
D. Azure Cognitive Services
E. Azure Batch

Answer: A

Explanation:
Customer data must be analyzed using managed Spark clusters. You create spark clusters through Azure Databricks. References:
https://docs.microsoft.com/en-us/azure/azure-databricks/quickstart-create-databricks-workspace-portal
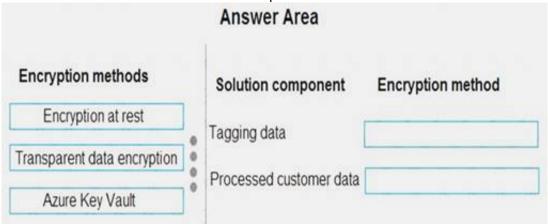
NEW QUESTION 6
- (Exam Topic 2)
You need to design the encryption strategy for the tagging data and customer data.
What should you recommend? To answer, drag the appropriate setting to the correct drop targets. Each source may be used once, more than once, or not at all.
You may need to drag the split bar between panes or scroll to view content.
NOTE: Each correct selection is worth one point.

A. Mastered
B. Not Mastered

Answer: A

Explanation:
All cloud data must be encrypted at rest and in transit. Box 1: Transparent data encryption
Encryption of the database file is performed at the page level. The pages in an encrypted database are encrypted before they are written to disk and decrypted when read into memory.
Box 2: Encryption at rest
Encryption at Rest is the encoding (encryption) of data when it is persisted. References:
https://docs.microsoft.com/en-us/sql/relational-databases/security/encryption/transparent-data-encryption?view= https://docs.microsoft.com/en-us/azure/security/azure-security-encryption-atrest

NEW QUESTION 7
- (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 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 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)
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 11
- (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.	▼ Azure Databricks / Azure Data Factory / Azure SQL Data Warehouse / Azure Batch
Develop notebooks to transform the data.	▼ Azure Databricks / Azure Data Factory / Azure SQL Data Warehouse / Azure Batch
Run notebooks.	▼ Azure Databricks / Azure Data Factory / Azure SQL Data Warehouse / Azure Batch
Load the data.	▼ Azure Databricks / Azure Data Factory / Azure SQL Data Warehouse / Azure Batch
Store the transformed data.	▼ Azure Databricks / Azure Data Factory / Azure SQL Data Warehouse / Azure Batch

A. Mastered
B. Not Mastered

Answer: A

Explanation:

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

NEW QUESTION 16
- (Exam Topic 4)
You are designing a real-time stream solution based on Azure Functions. The solution will process data uploaded to Azure Blob Storage.

The solution requirements are as follows:
New blobs must be processed with a little delay as possible. Scaling must occur automatically.
Costs must be minimized. What should you recommend?

A. Deploy the Azure Function in an App Service plan and use a Blob trigger.
B. Deploy the Azure Function in a Consumption plan and use an Event Grid trigger.
C. Deploy the Azure Function in a Consumption plan and use a Blob trigger.
D. Deploy the Azure Function in an App Service plan and use an Event Grid trigger.

Answer: C

Explanation:
Create a function, with the help of a blob trigger template, which is triggered when files are uploaded to or updated in Azure Blob storage.
You use a consumption plan, which is a hosting plan that defines how resources are allocated to your function app. In the default Consumption Plan, resources are added dynamically as required by your functions. In this serverless hosting, you only pay for the time your functions run. When you run in an App Service plan, you must manage the scaling of your function app.
References:
https://docs.microsoft.com/en-us/azure/azure-functions/functions-create-storage-blob-triggered-function

NEW QUESTION 20
- (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 24
- (Exam Topic 4)
A company has 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 25
- (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 28
- (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 33
- (Exam Topic 4)
A company installs IoT devices to monitor its fleet of delivery vehicles. Data from devices is collected from Azure Event Hub.
The data must be transmitted to Power BI for real-time data visualizations. You need to recommend a solution.
What should you recommend?

A. Azure HDInsight with Spark Streaming
B. Apache Spark in Azure Databricks
C. Azure Stream Analytics
D. Azure HDInsight with Storm

Answer: C

Explanation:
Step 1: Get your IoT hub ready for data access by adding a consumer group.
Step 2: Create, configure, and run a Stream Analytics job for data transfer from your IoT hub to your Power BI account.
Step 3: Create and publish a Power BI report to visualize the data. References:
https://docs.microsoft.com/en-us/azure/iot-hub/iot-hub-live-data-visualization-in-power-bi
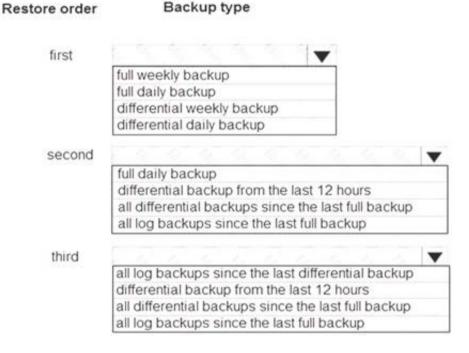
NEW QUESTION 35
- (Exam Topic 4)
You are designing a recovery strategy for your Azure SQL Databases.
The recovery strategy must use default automated backup settings. The solution must include a Point-in time restore recovery strategy.
You need to recommend which backups to use and the order in which 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.

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 39
- (Exam Topic 4)
You plan to deploy an Azure SQL Database instance to support an application. You plan to use the DTUbased purchasing model.
Backups of the database must be available for 30 days and point-in-time restoration must be possible. You need to recommend a backup and recovery policy.
What are two possible ways to achieve the goal? Each correct answer presents a complete solution.
NOTE: Each correct selection is worth one point.

A. Use the Premium tier and the default backup retention policy.

B. Use the Basic tier and the default backup retention policy.
C. Use the Standard tier and the default backup retention policy.
D. Use the Standard tier and configure a long-term backup retention policy.
E. Use the Premium tier and configure a long-term backup retention policy.

Answer: DE

Explanation:
The default retention period for a database created using the DTU-based purchasing model depends on the service tier:
- Basic service tier is 1 week.
- Standard service tier is 5 weeks.
- Premium service tier is 5 weeks.
References:
https://docs.microsoft.com/en-us/azure/sql-database/sql-database-long-term-retention

NEW QUESTION 40
- (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 45
- (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 MLib.

Answer: E

Explanation:
Spark in SQL Server big data cluster enables AI and machine learning.
You can use Apache Spark MLib to create a machine learning application to do simple predictive analysis on an open dataset.
MLib 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-ipython

NEW QUESTION 46
- (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 49
- (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 50
- (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 52
- (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: Implement compaction jobs to combine small files into larger files. Does the solution meet the goal?

A. Yes
B. No

Answer: A

Explanation:
Depending on what services and workloads are using the data, a good size to consider for files is 256 MB or greater. If the file sizes cannot be batched when landing in Data Lake Storage Gen1, you can have a separate compaction job that combines these files into larger ones.
Note: 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 56
- (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 57
- (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 60
- (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 65
- (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 69
- (Exam Topic 4)
You are designing an Azure Data Factory pipeline for processing data. The pipeline will process data that is stored in general-purpose standard Azure storage.
You need to ensure that the compute environment is created on-demand and removed when the process is completed.
Which type of activity should you recommend?

A. Databricks Python activity
B. Data Lake Analytics U-SQL activity
C. HDInsight Pig activity
D. Databricks Jar activity

Answer: C

Explanation:
The HDInsight Pig activity in a Data Factory pipeline executes Pig queries on your own or on-demand HDInsight cluster.
References:
https://docs.microsoft.com/en-us/azure/data-factory/transform-data-using-hadoop-pig
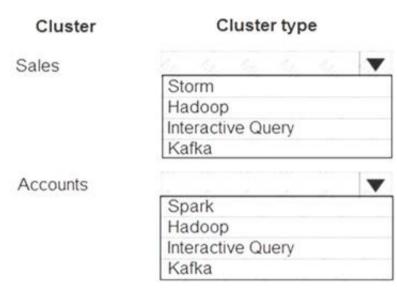
NEW QUESTION 74
- (Exam Topic 4)
A company stores large datasets in Azure, including sales transactions and customer account information. You must design a solution to analyze the data. You plan to create the following HDInsight clusters:
You need to ensure that the clusters support the query requirements.
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	▼
	Storm
	Hadoop
	Interactive Query
	Kafka
Accounts	▼
	Spark
	Hadoop
	Interactive Query
	Kafka

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 79
- (Exam Topic 4)
A company has locations in North America and Europe. The company uses Azure SQL Database to support business apps.
Employees must be able to access the app data in case of a region-wide outage. A multi-region availability solution is needed with the following requirements:
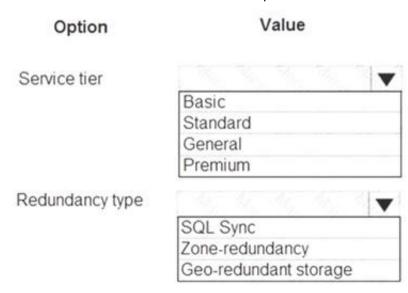🔵 Read-access to data in a secondary region must be available only in case of an outage of the primary region.
🔵 The Azure SQL Database compute and storage layers must be integrated and replicated together.
You need to design the multi-region high availability solution.
What should you recommend? To answer, select the appropriate values in the answer area.
NOTE: Each correct selection is worth one point.

Option	Value
Service tier	▼
	Basic
	Standard
	General
	Premium
Redundancy type	▼
	SQL Sync
	Zone-redundancy
	Geo-redundant storage

A. Mastered
B. Not Mastered

Answer: A

Explanation:
Box 1: Standard
The following table describes the types of storage accounts and their capabilities:

Storage account type	Supported services	Supported performance tiers	Supported access tiers	Replication options	Deployment model[1]
General-purpose V2	Blob, File, Queue, Table, and Disk	Standard, Premium[5]	Hot, Cool, Archive[3]	LRS, ZRS[4], GRS, RA-GRS	Resource Manager
General-purpose V1	Blob, File, Queue, Table, and Disk	Standard, Premium[5]	N/A	LRS, GRS, RA-GRS	Resource Manager, Classic

Box 2: Geo-redundant storage
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.
Note: If you opt for GRS, you have two related options to choose from:
GRS replicates your data to another data center in a secondary region, but that data is available to be read only if Microsoft initiates a failover from the primary to secondary region.
Read-access geo-redundant storage (RA-GRS) is based on GRS. RA-GRS replicates your data to another data center in a secondary region, and also provides you with the option to read from the secondary region. With RA-GRS, you can read from the secondary region regardless of whether Microsoft initiates a failover from the primary to secondary region.

Scenario	LRS	ZRS	GRS	RA-GRS
Node unavailability within a data center	Yes	Yes	Yes	Yes
An entire data center (zonal or non-zonal) becomes unavailable	No	Yes	Yes	Yes
A region-wide outage	No	No	Yes	Yes

References:
https://docs.microsoft.com/en-us/azure/storage/common/storage-introduction https://docs.microsoft.com/en-us/azure/storage/common/storage-redundancy-grs

NEW QUESTION 84
- (Exam Topic 4)
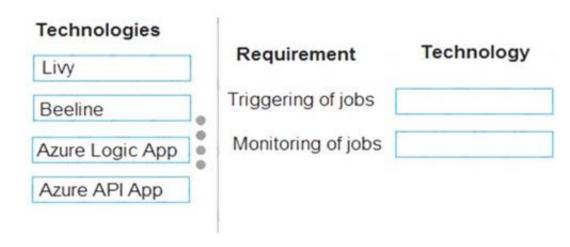You are designing a Spark job that performs batch processing of daily web log traffic.
When you deploy the job in the production environment, it must meet the following requirements:
○ Run once a day.
○ Display status information on the company intranet as the job runs. You need to recommend technologies 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

Livy

Beeline

Azure Logic App

Azure API App

Requirement	Technology
Triggering of jobs	
Monitoring of jobs	

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 88
- (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 93
......

Thank You for Trying Our Product

We offer two products:

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

2nd - Questons and Answers in PDF Format

DP-201 Practice Exam Features:

* DP-201 Questions and Answers Updated Frequently

* DP-201 Practice Questions Verified by Expert Senior Certified Staff

* DP-201 Most Realistic Questions that Guarantee you a Pass on Your FirstTry

* DP-201 Practice Test Questions in Multiple Choice Formats and Updatesfor 1 Year

100% Actual & Verified — Instant Download, Please Click
Order The DP-201 Practice Test Here