

# Exam Questions DP-201

Designing an Azure Data Solution

<https://www.2passeasy.com/dumps/DP-201/>



### NEW QUESTION 1

- (Exam Topic 1)

You need to design the Planning Assistance database.

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

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

- A. Mastered
- B. Not Mastered

**Answer:** A

#### Explanation:

Box 1: No

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

Box 3: Yes

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

### NEW QUESTION 2

- (Exam Topic 1)

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

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

**Answer:** B

#### Explanation:

Premium Storage stores data on the latest technology Solid State Drives (SSDs) whereas Standard Storage stores data on Hard Disk Drives (HDDs). Premium Storage is designed for Azure Virtual Machine workloads which require consistent high IO performance and low latency in order to host IO intensive workloads like OLTP, Big Data, and Data Warehousing on platforms like SQL Server, MongoDB, Cassandra, and others. With Premium Storage, more customers will be able to lift-and-shift demanding enterprise applications to the cloud.

Scenario: Traffic sensors will occasionally capture an image of a vehicle for debugging purposes. You must optimize performance of saving/storing vehicle images.

The impact of vehicle images on sensor data throughout must be minimized. References:

<https://azure.microsoft.com/es-es/blog/introducing-premium-storage-high-performance-storage-for-azure-virtual>

### NEW QUESTION 3

- (Exam Topic 1)

You need to design the runtime environment for the Real Time Response system. What should you recommend?

- A. General Purpose nodes without the Enterprise Security package
- B. Memory Optimized Nodes without the Enterprise Security package
- C. Memory Optimized nodes with the Enterprise Security package
- D. General Purpose nodes with the Enterprise Security package

**Answer:** B

### NEW QUESTION 4

- (Exam Topic 2)

You need to recommend a solution for storing customer data. What should you recommend?

- A. Azure SQL Data Warehouse
- B. Azure Stream Analytics
- C. Azure Databricks
- D. Azure SQL Database

**Answer:** C

#### Explanation:

From the scenario:

Customer data must be analyzed using managed Spark clusters.

All cloud data must be encrypted at rest and in transit. The solution must support: parallel processing of customer data.

References:

<https://www.microsoft.com/developerblog/2019/01/18/running-parallel-apache-spark-notebook-workloads-on-a>

### NEW QUESTION 5

- (Exam Topic 2)

You need to recommend a solution for storing the image tagging data. What should you recommend?

- A. Azure File Storage
- B. Azure Cosmos DB
- C. Azure Blob Storage
- D. Azure SQL Database
- E. Azure SQL Data Warehouse

**Answer:** C

#### Explanation:

Image data must be stored in a single data store at minimum cost.

Note: Azure Blob storage is Microsoft's object storage solution for the cloud. Blob storage is optimized for storing massive amounts of unstructured data.

Unstructured data is data that does not adhere to a particular data model or definition, such as text or binary data.

Blob storage is designed for:

- ▶ Serving images or documents directly to a browser.
- ▶ Storing files for distributed access.
- ▶ Streaming video and audio.
- ▶ Writing to log files.
- ▶ Storing data for backup and restore, disaster recovery, and archiving.
- ▶ Storing data for analysis by an on-premises or Azure-hosted service.

References:

<https://docs.microsoft.com/en-us/azure/storage/blobs/storage-blobs-introduction>

### NEW QUESTION 6

- (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 7

- (Exam Topic 3)

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

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

NOTE: Each correct selection is worth one point.

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

**Answer:** ADE

#### Explanation:

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

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

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

References:

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

### NEW QUESTION 8

- (Exam Topic 3)

You need to design a solution to meet the SQL Server storage requirements for CONT\_SQL3. Which type of disk should you recommend?

- A. Standard SSD Managed Disk
- B. Premium SSD Managed Disk
- C. Ultra SSD Managed Disk

**Answer:** C

**Explanation:**

CONT\_SQL3 requires an initial scale of 35000 IOPS.

| Disk size (GiB)       | 4         | 8         | 16        | 32        | 64         | 128        | 256        | 512         | 1,024-65,536 (in increments of 1 TiB) |
|-----------------------|-----------|-----------|-----------|-----------|------------|------------|------------|-------------|---------------------------------------|
| IOPS range            | 100-1,200 | 100-2,400 | 100-4,800 | 100-9,600 | 100-19,200 | 100-38,400 | 100-76,800 | 100-153,600 | 100-160,000                           |
| Throughput Cap (MBps) | 300       | 600       | 1,200     | 2,000     | 2,000      | 2,000      | 2,000      | 2,000       | 2,000                                 |

The following table provides a comparison of ultra solid-state-drives (SSD) (preview), premium SSD, standard SSD, and standard hard disk drives (HDD) for managed disks to help you decide what to use.

|                | Ultra SSD (preview)   | Premium SSD                                    | Standard SSD   | Standard HDD                            |
|----------------|---|--|--|---|
| Disk type      | SSD   | SSD  | SSD  | HDD                                     |
| Scenario       | IO-intensive workloads such as SAP HANA, top tier databases (for example, SQL Oracle), and other transaction-heavy workloads. | Production and performance sensitive workloads | Web servers, lightly used enterprise applications and dev/test | Backup, non-critical, infrequent access |
| Disk size      | 65,536 gibibyte (GiB) (Preview)   | 32,767 GiB                                     | 32,767 GiB   | 32,767 GiB                              |
| Max throughput | 2,000 MiB/s (Preview)   | 900 MiB/s                                      | 750 MiB/s  | 500 MiB/s                               |
| Max IOPS       | 160,000 (Preview)   | 20,000   | 6,000  | 2,000                                   |

References:

<https://docs.microsoft.com/en-us/azure/virtual-machines/windows/disks-types>

**NEW QUESTION 9**

- (Exam Topic 3)

A company stores sensitive information about customers and employees in Azure SQL Database. You need to ensure that the sensitive data remains encrypted in transit and at rest.

What should you recommend?

- A. Transparent Data Encryption
- B. Always Encrypted with secure enclaves
- C. Azure Disk Encryption
- D. SQL Server AlwaysOn

**Answer:** B

**Explanation:**

References:

<https://cloudblogs.microsoft.com/sqlserver/2018/12/17/confidential-computing-using-always-encrypted-withsec>

**NEW QUESTION 10**

- (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 10**

- (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 15**

- (Exam Topic 3)

You need to recommend a backup strategy for CONT\_SQL1 and CONT\_SQL2. What should you recommend?

- A. Use AzCopy and store the data in Azure.
- B. Configure Azure SQL Database long-term retention for all databases.
- C. Configure Accelerated Database Recovery.
- D. Use DWLoader.

**Answer:** B

**Explanation:**

Scenario: The database backups have regulatory purposes and must be retained for seven years.

**NEW QUESTION 20**

- (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 21**

- (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 22**

- (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<br/>                     Azure Data Factory<br/>                     Azure SQL Data Warehouse<br/>                     Azure Batch                 </div> |
| Develop notebooks to transform the data.       | <div>▼</div> <div>                     Azure Databricks<br/>                     Azure Data Factory<br/>                     Azure SQL Data Warehouse<br/>                     Azure Batch                 </div> |
| Run notebooks.                                 | <div>▼</div> <div>                     Azure Databricks<br/>                     Azure Data Factory<br/>                     Azure SQL Data Warehouse<br/>                     Azure Batch                 </div> |
| Load the data.                                 | <div>▼</div> <div>                     Azure Databricks<br/>                     Azure Data Factory<br/>                     Azure SQL Data Warehouse<br/>                     Azure Batch                 </div> |
| Store the transformed data.                    | <div>▼</div> <div>                     Azure Databricks<br/>                     Azure Data Factory<br/>                     Azure SQL Data Warehouse<br/>                     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<br/>                     Azure Data Factory<br/>                     Azure SQL Data Warehouse<br/>                     Azure Batch                 </div> |
| Develop notebooks to transform the data.       | <div>▼</div> <div>                     Azure Databricks<br/>                     Azure Data Factory<br/>                     Azure SQL Data Warehouse<br/>                     Azure Batch                 </div> |
| Run notebooks.                                 | <div>▼</div> <div>                     Azure Databricks<br/>                     Azure Data Factory<br/>                     Azure SQL Data Warehouse<br/>                     Azure Batch                 </div> |
| Load the data.                                 | <div>▼</div> <div>                     Azure Databricks<br/>                     Azure Data Factory<br/>                     Azure SQL Data Warehouse<br/>                     Azure Batch                 </div> |
| Store the transformed data.                    | <div>▼</div> <div>                     Azure Databricks<br/>                     Azure Data Factory<br/>                     Azure SQL Data Warehouse<br/>                     Azure Batch                 </div> |

#### NEW QUESTION 24

- (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 26**

- (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 27**

- (Exam Topic 4)

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

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

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

NOTE: Each correct selection is worth one point.

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

- A. Mastered
- B. Not Mastered

**Answer: A**

**Explanation:**

Box 1: Data Engineering Only

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

Box 4: Data Analytics only Box 5: Premium

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

References:

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

**NEW QUESTION 28**

- (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 29**

- (Exam Topic 4)

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

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

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

NOTE: Each correct selection is worth one point.

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

**Answer:** AE

**Explanation:**

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

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

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

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

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

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

**NEW QUESTION 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 38**

- (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 43

- (Exam Topic 4)

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

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

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

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

NOTE: Each correct selection is worth one point.

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

- A. Mastered
- B. Not Mastered

**Answer:** A

#### Explanation:

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

References:

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

### NEW QUESTION 45

- (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 48

- (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)

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 53**

- (Exam Topic 4)

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

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

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

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

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

**Answer: B**

**Explanation:**

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

References:

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

**NEW QUESTION 58**

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

You need to recommend a strategy to partition data based on values in CustomerID. Proposed Solution: Separate data into shards by using horizontal partitioning.

Does the solution meet the goal?

- A. Yes
- B. No

**Answer: A**

**Explanation:**

Horizontal Partitioning - Sharding: Data is partitioned horizontally to distribute rows across a scaled out data

tier. With this approach, the schema is identical on all participating databases. This approach is also called “sharding”. Sharding can be performed and managed using (1) the elastic database tools libraries or (2) selfsharding.

An elastic query is used to query or compile reports across many shards. References:

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

**NEW QUESTION 62**

- (Exam Topic 4)

You are developing a solution that performs real-time analysis of IoT data in the cloud. The solution must remain available during Azure service updates. You need to recommend a solution.

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

NOTE: Each correct selection is worth one point.

- A. Deploy an Azure Stream Analytics job to two separate regions that are not in a pair.
- B. Deploy an Azure Stream Analytics job to each region in a paired region.
- C. Monitor jobs in both regions for failure.
- D. Monitor jobs in the primary region for failure.
- E. Deploy an Azure Stream Analytics job to one region in a paired region.

**Answer: BC**

**Explanation:**

Stream Analytics guarantees jobs in paired regions are updated in separate batches. As a result there is a sufficient time gap between the updates to identify potential breaking bugs and remediate them.

Customers are advised to deploy identical jobs to both paired regions.

In addition to Stream Analytics internal monitoring capabilities, customers are also advised to monitor the jobs as if both are production jobs. If a break is identified to be a result of the Stream Analytics service update, escalate appropriately and fail over any downstream consumers to the healthy job output. Escalation to support will prevent the paired region from being affected by the new deployment and maintain the integrity of the paired jobs.

References:

<https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-job-reliability>

**NEW QUESTION 64**

- (Exam Topic 4)

You need to design the system for notifying law enforcement officers about speeding vehicles.

How should you design the pipeline? To answer, drag the appropriate services to the correct locations. Each service may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.



- A. Mastered
- B. Not Mastered

**Answer: A**

**Explanation:**



**NEW QUESTION 68**

- (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 69

.....



## THANKS FOR TRYING THE DEMO OF OUR PRODUCT

Visit Our Site to Purchase the Full Set of Actual DP-201 Exam Questions With Answers.

We Also Provide Practice Exam Software That Simulates Real Exam Environment And Has Many Self-Assessment Features. Order the DP-201 Product From:

<https://www.2passeasy.com/dumps/DP-201/>

## Money Back Guarantee

### 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