

DP-201.VCEplus.premium.exam.63q

Number: DP-201
Passing Score: 800
Time Limit: 120 min
File Version: 1.0



Website: <https://vceplus.com>

VCE to PDF Converter: <https://vceplus.com/vce-to-pdf/>

Facebook: <https://www.facebook.com/VCE.For.All.VN/>

Twitter : https://twitter.com/VCE_Plus

DP-201

Designing an Azure Data Solution (beta)



Testlet 1

Case study

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other question on this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next sections of the exam. After you begin a new section, you cannot return to this section.

To start the case study

To display the first question on this case study, click the **Next** button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an **All Information** tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the **Question** button to return to the question.

Background

Trey Research is a technology innovator. The company partners with regional transportation department office to build solutions that improve traffic flow and safety.

The company is developing the following solutions:

Solution	Comments
Real Time Response	This solution will detect sudden changes in traffic flow including slow downs and stops that persist for more than one minute. The system will automatically dispatch emergency response vehicles to investigate issues. The solution will use a PySpark script to detect traffic flow changes. Script performance will be limited by available memory.
Backtrack	This solution will allow public safety officials to locate vehicles on roadways that implement traffic sensors. The solution must report changes in real time.
Planning Assistance	Transportation organizations will use Planning Assistance to analyze traffic data. The solution will allow users to define reports based on queries of the traffic data. The reports can be used for the following analyses: <ul style="list-style-type: none"> • current traffic load • correlation with recent local events such as sporting events • historical traffic • tracking the travel of a single vehicle

Regional transportation departments installed traffic sensor systems on major highways across North America. Sensors record the following information each time a vehicle passes in front of a sensor:

- Time
- Location in latitude and longitude
- Speed in kilometers per second (km/s)
- License plate number
- Length of vehicle in meters

Sensors provide data by using the following structure:

```
{
  "time" : "2014-09-15T23:14:25.72511732",
  "location" : {
    "type": "Point",
    "coordinates": [
      31.9,
      -4.8
    ]
  },
  "speed": 66.2,
  "license_plate": "WA-AJ0072W",
  "vehicle_length": 4.5
}
```

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

Traffic sensor data

- Sensors must have permission only to add items to the SensorData collection.
 - Traffic data insertion rate must be maximized.
 - Once every three months all traffic sensor data must be analyzed to look for data patterns that indicate sensor malfunctions.
- Sensor data must be stored in a Cosmos DB named treydata in a collection named SensorData
- The impact of vehicle images on sensor data throughout must be minimized.

Backtrack

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

Planning Assistance

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

Data from the Sensor Data collection will automatically be loaded into the Planning Assistance database once a week by using Azure Data Factory. You must be able to manually trigger the data load process. **Privacy**

and security policy

- Azure Active Directory must be used for all services where it is available.
 - For privacy reasons, license plate number information must not be accessible in Planning Assistance.
 - 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 must only be stored for seven years.

Performance and availability

- The report for Backtrack must execute as quickly as possible.
 - The SLA for Planning Assistance is 70 percent, and multiday outages are permitted.
 - All data must be replicated to multiple geographic regions to prevent data loss.
- You must maximize the performance of the Real Time Response system.

Financial requirements

Azure resource costs must be minimized where possible.

QUESTION 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

Correct Answer: B

Section: [none]

Explanation

Explanation/Reference:

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-machine-workloads/>

QUESTION 2 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



Correct Answer: A

Section: [none]

Explanation

Explanation/Reference:

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>

QUESTION 3

HOTSPOT

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.

Hot Area:

Answer Area

Setting	Value
Default consistency level	<input type="text" value=""/> strong session eventual consistent prefix bounded staleness
Partition key property	<input type="text" value=""/> Time Location Speed License plate Vehicle length



Correct Answer:

Answer Area

Setting	Value
Default consistency level	<div style="border: 1px solid gray; padding: 2px;"> <div style="background-color: #f0f0f0; padding: 2px; display: flex; justify-content: space-between; align-items: center;"> ▼ </div> <ul style="list-style-type: none"> strong session <li style="background-color: #d9ead3;">eventual consistent prefix bounded staleness </div>
Partition key property	<div style="border: 1px solid gray; padding: 2px;"> <div style="background-color: #f0f0f0; padding: 2px; display: flex; justify-content: space-between; align-items: center;"> ▼ </div> <ul style="list-style-type: none"> Time Location Speed <li style="background-color: #d9ead3;">License plate Vehicle length </div>



Section: [none]

Explanation

Explanation/Reference:

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>

QUESTION 4 You need to recommend an Azure SQL Database pricing tier for Planning Assistance.

Which pricing tier should you recommend?

- A. Business critical Azure SQL Database single database
- B. General purpose Azure SQL Database Managed Instance
- C. Business critical Azure SQL Database Managed Instance
- D. General purpose Azure SQL Database single database

Correct Answer: B

Section: [none]

Explanation

Explanation/Reference:

Explanation:

Azure resource costs must be minimized where possible.

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

The SLA for Planning Assistance is 70 percent, and multiday outages are permitted.

QUESTION 5

HOTSPOT

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.

Hot Area:

Answer Area

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"/>

Correct Answer:

Answer Area

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

Section: [none]

Explanation

Explanation/Reference:

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 **Testlet 2**



Case study

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other question on this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next sections of the exam. After you begin a new section, you cannot return to this section.

To start the case study

To display the first question on this case study, click the **Next** button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an **All Information** tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the **Question** button to return to the question.

Requirements

Business

The company identifies the following business requirements:

- You must transfer all images and customer data to cloud storage and remove on-premises servers.
- You must develop an analytical processing solution for transforming customer data.
- You must develop an image object and color tagging solution.
- Capital expenditures must be minimized.
- Cloud resource costs must be minimized.

Technical

The solution has the following technical requirements:

- 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.
 - Image data must be stored in a single data store at minimum cost.
 - Customer data must be analyzed using managed Spark clusters. ▪
- Power BI must be used to visualize transformed customer data.
- All data must be backed up in case disaster recovery is required.

Security and optimization

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

- parallel processing of customer data ▪ hyper-scale storage of images
- global region data replication of processed image data

QUESTION 1 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

Correct Answer: C

Section: [none]

Explanation

Explanation/Reference:

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>

QUESTION 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

Correct Answer: A

Section: [none]

Explanation

Explanation/Reference:

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>

Testlet 3

Case study

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other question on this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next sections of the exam. After you begin a new section, you cannot return to this section.

To start the case study

To display the first question on this case study, click the **Next** button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an **All Information** tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the **Question** button to return to the question.

Background

Current environment

The company has the following virtual machines (VMs):

VM	Roles	Database size	VM type	Destination
CONT_SQL1	Microsoft SQL Server	2 TB	Hyper-V	Azure SQL Database
CONT_SQL2	Microsoft SQL Server	2 TB	Hyper-V	Azure SQL Database
CONT_SQL3	Microsoft SQL Server	100 GB	Hyper-V	Azure VM
CONT_SAP1	SAP	1 TB	Vmware	On-premises
CONT_SAP2	SAP	1 TB	Vmware	On-premises
CPNT_SSRS	Microsoft SQL Server Reporting Services	1 TB	Hyper-V	Azure VM

Requirements

Storage and processing

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. The architecture will need to support data files, libraries, and images. Additionally, it must provide a web-based interface to documents that contain runnable command, visualizations, and narrative text such as a notebook.

CONT_SQL3 requires an initial scale of 35000 IOPS.

CONT_SQL1 and CONT_SQL2 must use the vCore model and should include replicas. The solution must support 8000 IOPS. The storage should be configured to optimized storage for database OLTP workloads.

Migration

- You must be able to independently scale compute and storage resources.
- You must migrate all SQL Server workloads to Azure. You must identify related machines in the on-premises environment, get disk size data usage information.
- Data from SQL Server must include zone redundant storage.
- You need to ensure that app components can reside on-premises while interacting with components that run in the Azure public cloud. ▪

SAP data must remain on-premises.

- The Azure Site Recovery (ASR) results should contain per-machine data.

Business requirements

- You must design a regional disaster recovery topology.
- The database backups have regulatory purposes and must be retained for seven years.
- CONT_SQL1 stores customers sales data that requires ETL operations for data analysis. A solution is required that reads data from SQL, performs ETL, and outputs to Power BI. The solution should use managed clusters to minimize costs. To optimize logistics, Contoso needs to analyze customer sales data to see if certain products are tied to specific times in the year. ▪ The analytics solution for customer sales data must be available during a regional outage.

Security and auditing

- Contoso requires all corporate computers to enable Windows Firewall.
 - Azure servers should be able to ping other Contoso Azure servers.
 - Employee PII must be encrypted in memory, in motion, and at rest. Any data encrypted by SQL Server must support equality searches, grouping, indexing, and joining on the encrypted data. ▪
- Keys must be secured by using hardware security modules (HSMs).
- CONT_SQL3 must not communicate over the default ports

Cost

- All solutions must minimize cost and resources.
- The organization does not want any unexpected charges.
- The data engineers must set the SQL Data Warehouse compute resources to consume 300 DWUs.
- CONT_SQL2 is not fully utilized during non-peak hours. You must minimize resource costs for during non-peak hours.



QUESTION 1 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

Correct Answer: C

Section: [none]

Explanation

Explanation/Reference:

Explanation:

CONT_SQL3 requires an initial scale of 35000 IOPS.

Ultra SSD Managed Disk Offerings

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>

QUESTION 2 You need to recommend an Azure SQL Database service tier.

What should you recommend?

- A. Business Critical
- B. General Purpose
- C. PremiumD. Standard
- E. Basic

Correct Answer: C

Section: [none]

Explanation

Explanation/Reference:

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

Incorrect Answers:

A: Business Critical service tier is designed for the applications that require low-latency responses from the underlying SSD storage (1-2 ms in average), fast recovery if the underlying infrastructure fails, or need to off-load reports, analytics, and read-only queries to the free of charge readable secondary replica of the primary database.

References: <https://docs.microsoft.com/en-us/azure/sql-database/sql-database-service-tier-business-critical>

QUESTION 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.

Correct Answer: C

Section: [none]

Explanation

Explanation/Reference:

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.

References: <https://docs.databricks.com/delta/delta-cache.html#delta-cache>

Question Set 4

QUESTION 1

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 larger than 250MB.

Does the solution meet the goal?

- A. Yes
- B. No

Correct Answer: A

Section: [none]

Explanation

Explanation/Reference:

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>

QUESTION 2

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

Correct Answer: A

Section: [none]

Explanation

Explanation/Reference:

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>

QUESTION 3

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution 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



Correct Answer: B

Section: [none]

Explanation

Explanation/Reference:

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>

QUESTION 4

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

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

You need to recommend a strategy to partition data based on values in CustomerID.

Proposed Solution: Separate data into customer regions by using vertical partitioning.

Does the solution meet the goal?

- A. Yes
- B. No

Correct Answer: B

Section: [none]

Explanation

Explanation/Reference:

Explanation:

Vertical partitioning is used for cross-database queries. Instead we should use Horizontal Partitioning, which also is called charding.

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

QUESTION 5

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

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

You need to recommend a strategy to partition data based on values in CustomerID.

Proposed Solution: Separate data into customer regions by using horizontal partitioning.

Does the solution meet the goal?

- A. Yes
- B. No

Correct Answer: B

Section: [none]

Explanation

Explanation/Reference:

Explanation:

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

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

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

QUESTION 6

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

Correct Answer: A

Section: [none]

Explanation

Explanation/Reference:

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

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

QUESTION 7

HOTSPOT

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

The solution must include a page on the company intranet that displays summary information.

The summary information page must meet the following requirements:

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

You need to recommend a design for the solution.

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

Hot Area:

Answer Area

Use case	Technology
Data abstraction	<div style="border: 1px solid gray; padding: 2px;"> <div style="background-color: #f0f0f0; padding: 2px; display: flex; justify-content: space-between; align-items: center;"> ▼ </div> <div style="padding: 2px;"> Resilient Distributed Dataset (RDD) Dataset DataFrame </div> </div>
Data format	<div style="border: 1px solid gray; padding: 2px;"> <div style="background-color: #f0f0f0; padding: 2px; display: flex; justify-content: space-between; align-items: center;"> ▼ </div> <div style="padding: 2px;"> Avro parquet </div> </div>

Correct Answer:

Answer Area

Use case	Technology
Data abstraction	<div style="border: 1px solid black; padding: 2px;"> <div style="background-color: #cccccc; padding: 2px; display: flex; justify-content: space-between; align-items: center;"> ▼ </div> <div style="padding: 2px;"> <p>Resilient Distributed Dataset (RDD)</p> <p style="background-color: #d9ead3;">Dataset</p> <p>DataFrame</p> </div> </div>
Data format	<div style="border: 1px solid black; padding: 2px;"> <div style="background-color: #cccccc; padding: 2px; display: flex; justify-content: space-between; align-items: center;"> ▼ </div> <div style="padding: 2px;"> <p>Avro</p> <p style="background-color: #d9ead3;">parquet</p> </div> </div>

Section: [none]

Explanation

Explanation/Reference:

Explanation:

Box 1: DataFrame

DataFrames

Best choice in most situations.

Provides query optimization through Catalyst.

Whole-stage code generation.

Direct memory access.

Low garbage collection (GC) overhead.

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

Box 2: parquet

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

Incorrect Answers:

DataSets

Good in complex ETL pipelines where the performance impact is acceptable.

Not good in aggregations where the performance impact can be considerable.

RDDs

You do not need to use RDDs, unless you need to build a new custom RDD.

No query optimization through Catalyst.

No whole-stage code generation.

High GC overhead.

References: <https://docs.microsoft.com/en-us/azure/hdinsight/spark/apache-spark-perf>

QUESTION 8 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

Correct Answer: B

Section: [none]

Explanation

Explanation/Reference:

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>

QUESTION 9

HOTSPOT

You have an on-premises data warehouse that includes the following fact tables. Both tables have the following columns: DataKey, ProductKey, RegionKey. There are 120 unique product keys and 65 unique region keys.

Table	Comments
Sales	The table is 600 GB in size. Datekey is used extensively in the WHERE clause in queries. ProductKey is used extensively in join operations. RegionKey is used for grouping. Seventy-five percent of records relate to one of 40 regions.
Invoice	The table is 6 GB in size. DataKey and ProductKey are used extensively in the WHERE clause in queries. RegionKey is used for grouping.



Queries that use the data warehouse take a long time to complete.

You plan to migrate the solution to use Azure SQL Data Warehouse. You need to ensure that the Azure-based solution optimizes query performance and minimizes processing skew.

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

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

Table	Distribution type	Distribution column
Sales	<div style="border: 1px solid black; padding: 2px;"> <div style="background-color: #cccccc; padding: 2px;">▼</div> <div style="padding: 2px;">Hash-distributed</div> <div style="padding: 2px;">Round-robin</div> </div>	<div style="border: 1px solid black; padding: 2px;"> <div style="background-color: #cccccc; padding: 2px;">▼</div> <div style="padding: 2px;">DateKey</div> <div style="padding: 2px;">ProductKey</div> <div style="padding: 2px;">RegionKey</div> </div>
Invoices	<div style="border: 1px solid black; padding: 2px;"> <div style="background-color: #cccccc; padding: 2px;">▼</div> <div style="padding: 2px;">Hash-distributed</div> <div style="padding: 2px;">Round-robin</div> </div>	<div style="border: 1px solid black; padding: 2px;"> <div style="background-color: #cccccc; padding: 2px;">▼</div> <div style="padding: 2px;">DateKey</div> <div style="padding: 2px;">ProductKey</div> <div style="padding: 2px;">RegionKey</div> </div>

Correct Answer:

Answer Area

Table	Distribution type	Distribution column
Sales	<div style="border: 1px solid black; padding: 2px;"> <div style="background-color: #cccccc; padding: 2px;">▼</div> <div style="padding: 2px; background-color: #d9ead3;">Hash-distributed</div> <div style="padding: 2px;">Round-robin</div> </div>	<div style="border: 1px solid black; padding: 2px;"> <div style="background-color: #cccccc; padding: 2px;">▼</div> <div style="padding: 2px;">DateKey</div> <div style="padding: 2px; background-color: #d9ead3;">ProductKey</div> <div style="padding: 2px;">RegionKey</div> </div>
Invoices	<div style="border: 1px solid black; padding: 2px;"> <div style="background-color: #cccccc; padding: 2px;">▼</div> <div style="padding: 2px;">Hash-distributed</div> <div style="padding: 2px; background-color: #d9ead3;">Round-robin</div> </div>	<div style="border: 1px solid black; padding: 2px;"> <div style="background-color: #cccccc; padding: 2px;">▼</div> <div style="padding: 2px;">DateKey</div> <div style="padding: 2px;">ProductKey</div> <div style="padding: 2px; background-color: #d9ead3;">RegionKey</div> </div>

Section: [none]

Explanation

Explanation/Reference:

Explanation:

Box 1: Hash-distributed

Box 2: ProductKey

ProductKey is used extensively in joins.

Hash-distributed tables improve query performance on large fact tables.

Box 3: Round-robin

Box 4: RegionKey

Round-robin tables are useful for improving loading speed.

Consider using the round-robin distribution for your table in the following scenarios:

- When getting started as a simple starting point since it is the default
- If there is no obvious joining key
- If there is not good candidate column for hash distributing the table
- If the table does not share a common join key with other tables
- If the join is less significant than other joins in the query

When the table is a temporary staging table

Note: A distributed table appears as a single table, but the rows are actually stored across 60 distributions. The rows are distributed with a hash or round-robin algorithm.

References: <https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-tables-distribute>

QUESTION 10

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



Correct Answer: A

Section: [none]

Explanation

Explanation/Reference:

Explanation:

To implement a lambda architecture on Azure, you can combine the following technologies to accelerate real-time 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>

QUESTION 11 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 data. Output query results to Power BI.
- B. Configure an Azure Event Hub to capture data to Azure Data Lake Storage.
- C. Develop an Azure Stream Analytics application that queries the data and outputs to Power BI. Use Azure Data Factory to deploy the Azure Stream Analytics application.
- D. Develop an application that sends the IoT data to an Azure Event Hub.
- E. Develop an Azure Stream Analytics application that queries the data and outputs to Power BI. Use Azure Pipelines to deploy the Azure Stream Analytics application.
- F. Develop an application that sends the IoT data to an Azure Data Lake Storage container.

Correct Answer: BCD

Section: [none]

Explanation

Explanation/Reference:

QUESTION 12 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.

Correct Answer: B

Section: [none]

Explanation

Explanation/Reference:

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>

QUESTION 13 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.



Correct Answer: E

Section: [none]

Explanation

Explanation/Reference:

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-ipython>

Testlet 1

Case study

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other question on this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next sections of the exam. After you begin a new section, you cannot return to this section.

To start the case study

To display the first question on this case study, click the **Next** button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an **All Information** tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the **Question** button to return to the question.

Background

Trey Research is a technology innovator. The company partners with regional transportation department office to build solutions that improve traffic flow and safety.

The company is developing the following solutions:

Solution	Comments
Real Time Response	This solution will detect sudden changes in traffic flow including slow downs and stops that persist for more than one minute. The system will automatically dispatch emergency response vehicles to investigate issues. The solution will use a PySpark script to detect traffic flow changes. Script performance will be limited by available memory.
Backtrack	This solution will allow public safety officials to locate vehicles on roadways that implement traffic sensors. The solution must report changes in real time.
Planning Assistance	Transportation organizations will use Planning Assistance to analyze traffic data. The solution will allow users to define reports based on queries of the traffic data. The reports can be used for the following analyses: <ul style="list-style-type: none"> • current traffic load • correlation with recent local events such as sporting events • historical traffic • tracking the travel of a single vehicle

Regional transportation departments installed traffic sensor systems on major highways across North America. Sensors record the following information each time a vehicle passes in front of a sensor:

- Time
- Location in latitude and longitude
- Speed in kilometers per second (km/s)
- License plate number
- Length of vehicle in meters

Sensors provide data by using the following structure:

```
{
  "time" : "2014-09-15T23:14:25.72511732",
  "location" : {
    "type": "Point",
    "coordinates": [
      31.9,
      -4.8
    ]
  },
  "speed": 66.2,
  "license_plate": "WA-AJ0072W",
  "vehicle_length": 4.5
}
```

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

Traffic sensor data

- Sensors must have permission only to add items to the SensorData collection.
 - Traffic data insertion rate must be maximized.
 - Once every three months all traffic sensor data must be analyzed to look for data patterns that indicate sensor malfunctions.
- Sensor data must be stored in a Cosmos DB named treydata in a collection named SensorData ▪ The impact of vehicle images on sensor data throughout must be minimized.

Backtrack

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

Planning Assistance

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

Data from the Sensor Data collection will automatically be loaded into the Planning Assistance database once a week by using Azure Data Factory. You must be able to manually trigger the data load process. **Privacy**

and security policy

- Azure Active Directory must be used for all services where it is available.
 - For privacy reasons, license plate number information must not be accessible in Planning Assistance.
 - 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 must only be stored for seven years.

Performance and availability

- The report for Backtrack must execute as quickly as possible.
 - The SLA for Planning Assistance is 70 percent, and multiday outages are permitted.
 - All data must be replicated to multiple geographic regions to prevent data loss.
- You must maximize the performance of the Real Time Response system.

Financial requirements

Azure resource costs must be minimized where possible.

QUESTION 1

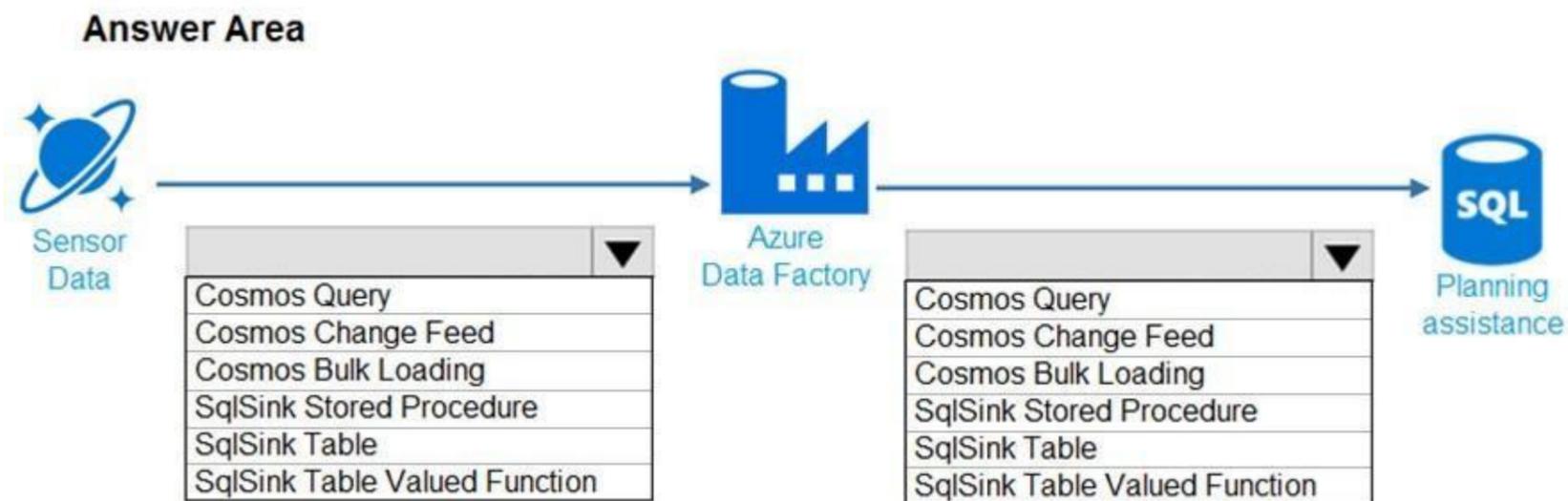
HOTSPOT

You need to design the data loading pipeline for Planning Assistance.

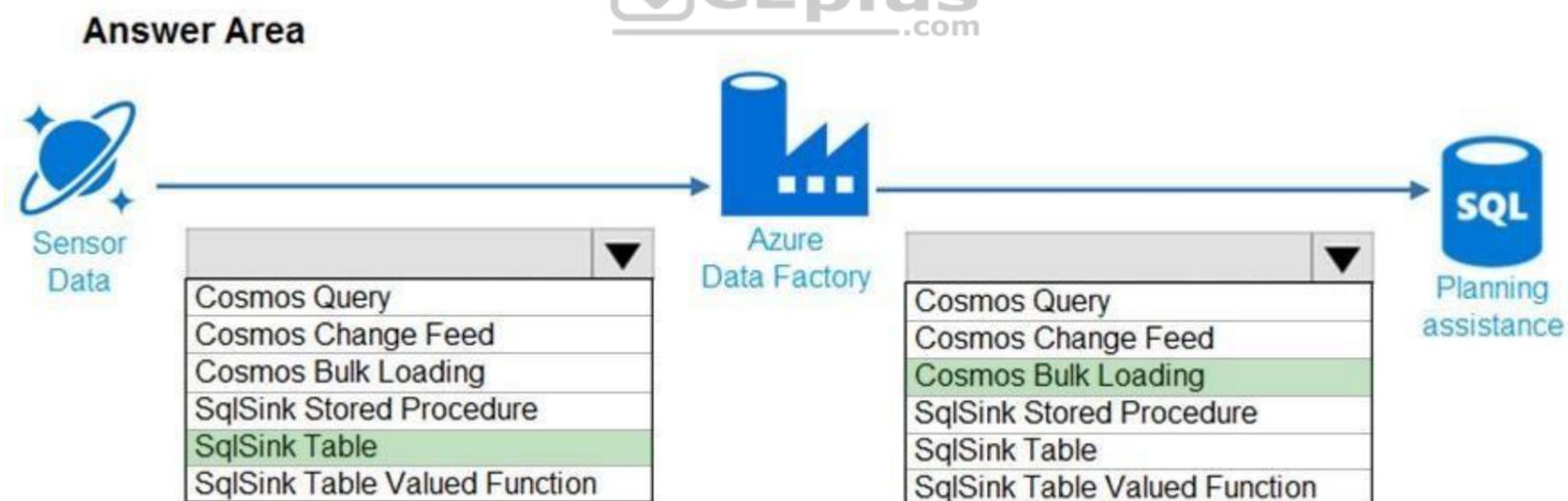
What should you recommend? To answer, drag the appropriate technologies to the correct locations. Each technology may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

Hot Area:



Correct Answer:



Section: [none]

Explanation

Explanation/Reference:

Explanation:

Box 1: SqlSink Table

Sensor data must be stored in a Cosmos DB named treydata in a collection named SensorData

Box 2: Cosmos Bulk Loading

Use Copy Activity in Azure Data Factory to copy data from and to Azure Cosmos DB (SQL API).

Scenario: Data from the Sensor Data collection will automatically be loaded into the Planning Assistance database once a week by using Azure Data Factory. You must be able to manually trigger the data load process. Data used for Planning Assistance must be stored in a sharded Azure SQL Database.

References: <https://docs.microsoft.com/en-us/azure/data-factory/connector-azure-cosmos-db>

QUESTION 2 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

Correct Answer: B

Section: [none]

Explanation

Explanation/Reference:

Explanation:

Scenario: You must maximize the performance of the Real Time Response system.

QUESTION 3

HOTSPOT

You need to ensure that emergency road response vehicles are dispatched automatically.

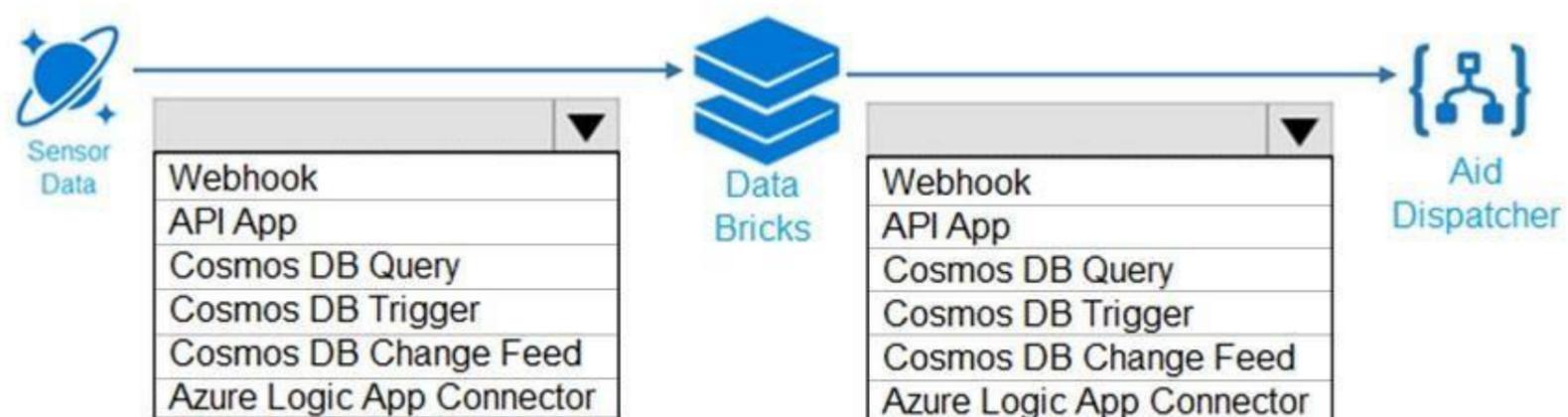
How should you design the processing system? To answer, select the appropriate options in the answer area.



NOTE: Each correct selection is worth one point.

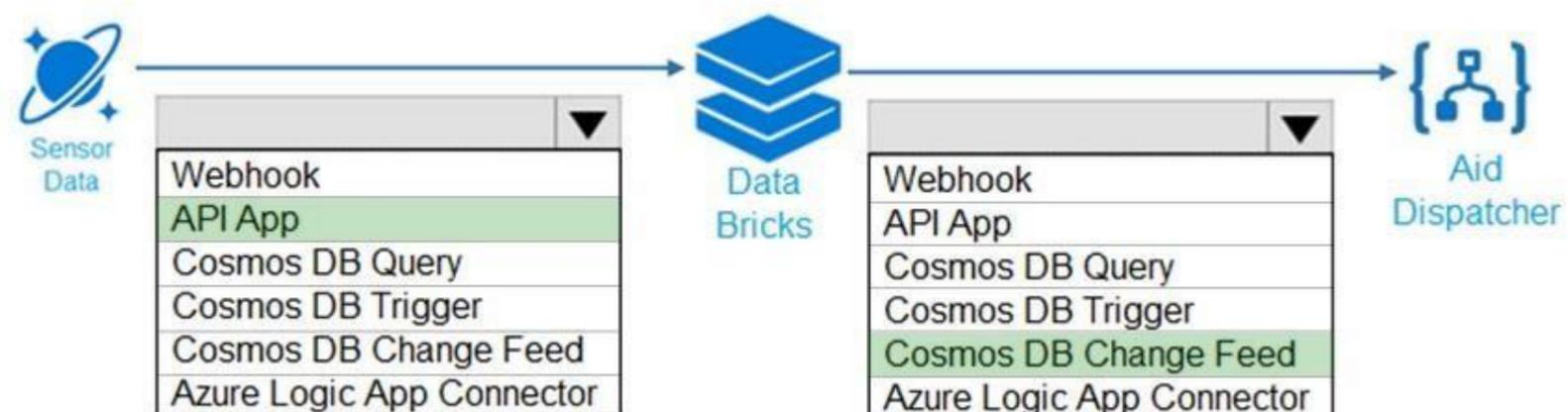
Hot Area:

Answer Area



Correct Answer:

Answer Area



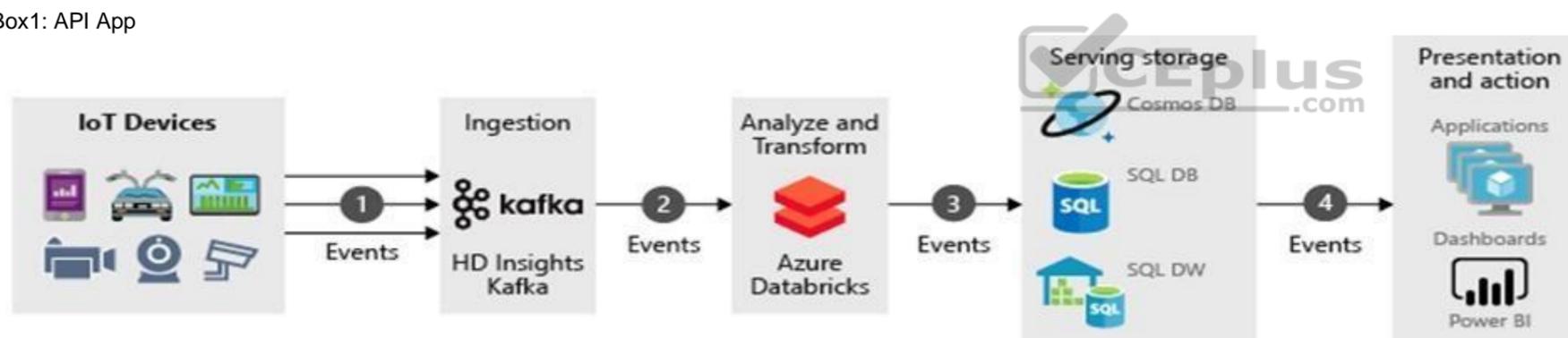
Section: [none]

Explanation

Explanation/Reference:

Explanation:

Box1: API App



1. Events generated from the IoT data sources are sent to the stream ingestion layer through Azure HDInsight Kafka as a stream of messages. HDInsight Kafka stores streams of data in topics for a configurable of time.

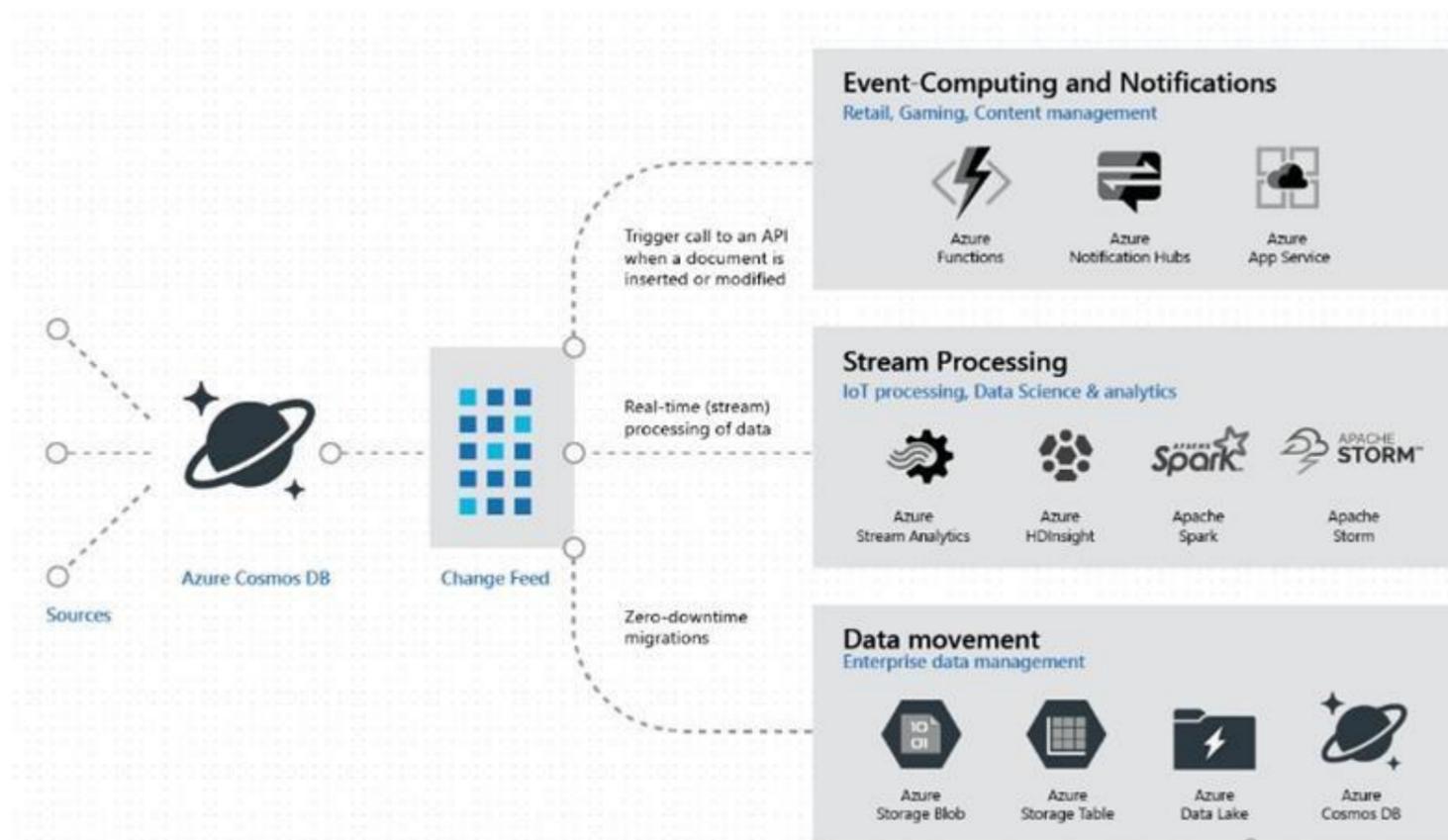
2. Kafka consumer, Azure Databricks, picks up the message in real time from the Kafka topic, to process the data based on the business logic and can then send to Serving layer for storage.

3. Downstream storage services, like Azure Cosmos DB, Azure SQL Data warehouse, or Azure SQL DB, will then be a data source for presentation and action layer.

4. Business analysts can use Microsoft Power BI to analyze warehoused data. Other applications can be built upon the service layer as well. For example, we can expose APIs based on the service layer data for third party uses.

Box 2: Cosmos DB Change Feed

Change feed support in Azure Cosmos DB works by listening to an Azure Cosmos DB container for any changes. It then outputs the sorted list of documents that were changed in the order in which they were modified. The change feed in Azure Cosmos DB enables you to build efficient and scalable solutions for each of these patterns, as shown in the following image:



References: <https://docs.microsoft.com/bs-cyrl-ba/azure/architecture/example-scenario/data/realtime-analytics-vehicle-iot?view=azuremps-4.4.1>

QUESTION 4
DRAG DROP

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.

Select and Place:

Technologies	Answer Area	
	Requirement	Technology
Cosmos DB TTL		
Cosmos DB indexes		
Cosmos DB transactions	Backtrack reporting	
Cosmos DB change feed	Privacy and security policy	
Cosmos DB stored procedures		

Correct Answer:

Technologies	Answer Area	
	Requirement	Technology
Cosmos DB TTL		
Cosmos DB indexes		
Cosmos DB transactions	Backtrack reporting	Cosmos DB indexes
Cosmos DB change feed	Privacy and security policy	Cosmos DB TTL
Cosmos DB stored procedures		

Section: [none]
Explanation

Explanation/Reference:
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.

Incorrect Answers:

Cosmos DB stored procedures: Stored procedures are best suited for operations that are write heavy. When deciding where to use stored procedures, optimize around encapsulating the maximum amount of writes possible. Generally speaking, stored procedures are not the most efficient means for doing large numbers of read operations so using stored procedures to batch large numbers of reads to return to the client will not yield the desired benefit.

References:

<https://docs.microsoft.com/en-us/azure/cosmos-db/index-policy>

<https://docs.microsoft.com/en-us/azure/cosmos-db/time-to-live>



Testlet 2

Case study

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other question on this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next sections of the exam. After you begin a new section, you cannot return to this section.

To start the case study

To display the first question on this case study, click the **Next** button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an **All Information** tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the **Question** button to return to the question.

Requirements

Business

The company identifies the following business requirements:

- You must transfer all images and customer data to cloud storage and remove on-premises servers.
- You must develop an analytical processing solution for transforming customer data.
- You must develop an image object and color tagging solution.
- Capital expenditures must be minimized.
- Cloud resource costs must be minimized.

Technical

The solution has the following technical requirements:

- 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.
- Image data must be stored in a single data store at minimum cost.
- Customer data must be analyzed using managed Spark clusters.
- Power BI must be used to visualize transformed customer data.
- All data must be backed up in case disaster recovery is required.

Security and optimization

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

- parallel processing of customer data
- hyper-scale storage of images
- global region data replication of processed image data

QUESTION 1

DRAG DROP

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.

Select and Place:



Answer Area

Configurations	Location	Configuration
<input type="checkbox"/> Write region	New York	<input type="text"/>
<input type="checkbox"/> Read region	Manchester	<input type="text"/>
	Singapore	<input type="text"/>
	Melbourne	<input type="text"/>

Correct Answer:

Answer Area

Configurations	Location	Configuration
<input type="checkbox"/> Write region	New York	<input type="text" value="Write region"/>
<input type="checkbox"/> Read region	Manchester	<input type="text" value="Read region"/>
	Singapore	<input type="text" value="Read region"/>
	Melbourne	<input type="text" value="Read region"/>

Section: [none]
Explanation

Explanation/Reference:
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.

Testlet 3

Case study

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other question on this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next sections of the exam. After you begin a new section, you cannot return to this section.

To start the case study

To display the first question on this case study, click the **Next** button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an **All Information** tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the **Question** button to return to the question.

Background

Current environment

The company has the following virtual machines (VMs):

VM	Roles	Database size	VM type	Destination
CONT_SQL1	Microsoft SQL Server	2 TB	Hyper-V	Azure SQL Database
CONT_SQL2	Microsoft SQL Server	2 TB	Hyper-V	Azure SQL Database
CONT_SQL3	Microsoft SQL Server	100 GB	Hyper-V	Azure VM
CONT_SAP1	SAP	1 TB	Vmware	On-premises
CONT_SAP2	SAP	1 TB	Vmware	On-premises
CPNT_SSRS	Microsoft SQL Server Reporting Services	1 TB	Hyper-V	Azure VM

Requirements

Storage and processing

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. The architecture will need to support data files, libraries, and images. Additionally, it must provide a web-based interface to documents that contain runnable command, visualizations, and narrative text such as a notebook.

CONT_SQL3 requires an initial scale of 35000 IOPS.

CONT_SQL1 and CONT_SQL2 must use the vCore model and should include replicas. The solution must support 8000 IOPS. The storage should be configured to optimized storage for database OLTP workloads.

Migration

- You must be able to independently scale compute and storage resources.
 - You must migrate all SQL Server workloads to Azure. You must identify related machines in the on-premises environment, get disk size data usage information.
 - Data from SQL Server must include zone redundant storage.
 - You need to ensure that app components can reside on-premises while interacting with components that run in the Azure public cloud. ▪
- SAP data must remain on-premises.
- The Azure Site Recovery (ASR) results should contain per-machine data.

Business requirements

- You must design a regional disaster recovery topology.
- The database backups have regulatory purposes and must be retained for seven years.
- CONT_SQL1 stores customers sales data that requires ETL operations for data analysis. A solution is required that reads data from SQL, performs ETL, and outputs to Power BI. The solution should use managed clusters to minimize costs. To optimize logistics, Contoso needs to analyze customer sales data to see if certain products are tied to specific times in the year. ▪ The analytics solution for customer sales data must be available during a regional outage.

Security and auditing

- Contoso requires all corporate computers to enable Windows Firewall.
- Azure servers should be able to ping other Contoso Azure servers.



- Employee PII must be encrypted in memory, in motion, and at rest. Any data encrypted by SQL Server must support equality searches, grouping, indexing, and joining on the encrypted data.
- Keys must be secured by using hardware security modules (HSMs).
- CONT_SQL3 must not communicate over the default ports

Cost

- All solutions must minimize cost and resources.
- The organization does not want any unexpected charges.
- The data engineers must set the SQL Data Warehouse compute resources to consume 300 DWUs.
- CONT_SQL2 is not fully utilized during non-peak hours. You must minimize resource costs for during non-peak hours.

QUESTION 1 You need to optimize storage for CONT_SQL3.

What should you recommend?

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

Correct Answer: B

Section: [none]

Explanation

Explanation/Reference:

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-sql-database.html>

Question Set 4

QUESTION 1

HOTSPOT

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:

Cluster	Requirement
Sales	This cluster must be optimized for ad hoc HVE queries.
Accounts	This cluster must be optimized for HVE queries that are used in batch processes.

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.

Hot Area:

Answer Area

Cluster	Cluster type
Sales	<div style="border: 1px solid black; padding: 2px;"><div style="background-color: #cccccc; padding: 2px;">▼</div><ul style="list-style-type: none">StormHadoopInteractive QueryKafka</div>
Accounts	<div style="border: 1px solid black; padding: 2px;"><div style="background-color: #cccccc; padding: 2px;">▼</div><ul style="list-style-type: none">SparkHadoopInteractive QueryKafka</div>

Correct Answer:

Answer Area

Cluster	Cluster type
Sales	<div style="border: 1px solid black; padding: 2px;"><div style="background-color: #cccccc; padding: 2px;">▼</div><ul style="list-style-type: none">StormHadoop<li style="background-color: #d9ead3;">Interactive QueryKafka</div>
Accounts	<div style="border: 1px solid black; padding: 2px;"><div style="background-color: #cccccc; padding: 2px;">▼</div><ul style="list-style-type: none">Spark<li style="background-color: #d9ead3;">HadoopInteractive QueryKafka</div>

Section: [none]

Explanation

Explanation/Reference:

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%2Fazure%2Fhdinsight%2Finteractive-query%2FTOC.json&bc=%2Fbs-latn-ba%2Fazure%2Fbread%2Ftoc.json>

QUESTION 2

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



Correct Answer: A

Section: [none]

Explanation

Explanation/Reference:

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.

Incorrect:

Not B: Consider that loading is usually a two-step process in which you first load to a staging table and then insert the data into a production data warehouse table. If the production table uses a hash distribution, the total time to load and insert might be faster if you define the staging table with the hash distribution. Loading to the staging table takes longer, but the second step of inserting the rows to the production table does not incur data movement across the distributions.

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

QUESTION 3 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.

Correct Answer: C

Section: [none]

Explanation

Explanation/Reference:

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.

Incorrect Answers:

A: Elastic pools is used if there are two or more databases.

References: https://www.skylinetechnologies.com/Blog/Skyline-Blog/August_2017/dynamically-scale-azure-sql-database

QUESTION 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

Correct Answer: C

Section: [none]

Explanation

Explanation/Reference:

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>

QUESTION 5 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

Correct Answer: C

Section: [none]

Explanation



Explanation/Reference:

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>

QUESTION 6

You have a Windows-based solution that analyzes scientific data. You are designing a cloud-based solution that performs real-time analysis of the data.

You need to design the logical flow for the solution.

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

NOTE: Each correct selection is worth one point.

- A. Send data from the application to an Azure Stream Analytics job.
- B. Use an Azure Stream Analytics job on an edge device. Ingress data from an Azure Data Factory instance and build queries that output to Power BI.
- C. Use an Azure Stream Analytics job in the cloud. Ingress data from the Azure Event Hub instance and build queries that output to Power BI.
- D. Use an Azure Stream Analytics job in the cloud. Ingress data from an Azure Event Hub instance and build queries that output to Azure Data Lake Storage.
- E. Send data from the application to Azure Data Lake Storage.
- F. Send data from the application to an Azure Event Hub instance.

Correct Answer: CF

Section: [none]

Explanation

Explanation/Reference:

Explanation:

Stream Analytics has first-class integration with Azure data streams as inputs from three kinds of resources:

- Azure Event Hubs
- Azure IoT Hub
- Azure Blob storage

References: <https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-define-inputs>

QUESTION 7

DRAG DROP

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.

Select and Place:



Answer Area

Technologies	Requirement	Technology
Livy	Triggering of jobs	
Beeline	Monitoring of jobs	
Azure Logic App		
Azure API App		

Correct Answer:

Answer Area

Technologies	Requirement	Technology
Livy	Triggering of jobs	Livy
Beeline	Monitoring of jobs	Beeline
Azure Logic App		
Azure API App		

Section: [none]
Explanation

Explanation/Reference:
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>

QUESTION 8

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.

Correct Answer: C

Section: [none]

Explanation

Explanation/Reference:

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>

QUESTION 9 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

Correct Answer: E

Section: [none]

Explanation

Explanation/Reference:

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-transactional-replication/>

QUESTION 10
HOTSPOT

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.

Hot Area:

Answer Area

Requirement	Workload	Tier
Provide managed clusters for running production jobs.	<div style="border: 1px solid gray; padding: 2px;">▼</div> <div style="border: 1px solid gray; padding: 2px;">Data Engineering only</div> <div style="border: 1px solid gray; padding: 2px;">Data Analytics only</div> <div style="border: 1px solid gray; padding: 2px;">Data Engineering and Data Analytics</div>	Standard
Provide persistent clusters that support auto-scaling for analytics processes.	<div style="border: 1px solid gray; padding: 2px;">▼</div> <div style="border: 1px solid gray; padding: 2px;">Data Engineering only</div> <div style="border: 1px solid gray; padding: 2px;">Data Analytics only</div> <div style="border: 1px solid gray; padding: 2px;">Data Engineering and Data Analytics</div>	<div style="border: 1px solid gray; padding: 2px;">▼</div> <div style="border: 1px solid gray; padding: 2px;">Standard</div> <div style="border: 1px solid gray; padding: 2px;">Premium</div>
Provide role-based access control (RBAC) support for Notebooks.	<div style="border: 1px solid gray; padding: 2px;">▼</div> <div style="border: 1px solid gray; padding: 2px;">Data Engineering only</div> <div style="border: 1px solid gray; padding: 2px;">Data Analytics only</div> <div style="border: 1px solid gray; padding: 2px;">Data Engineering and Data Analytics</div>	<div style="border: 1px solid gray; padding: 2px;">▼</div> <div style="border: 1px solid gray; padding: 2px;">Standard</div> <div style="border: 1px solid gray; padding: 2px;">Premium</div>

Correct Answer:

Answer Area

Requirement	Workload	Tier
Provide managed clusters for running production jobs.	<div style="border: 1px solid black; padding: 2px;"> <div style="background-color: #f0f0f0; padding: 2px; display: flex; justify-content: space-between; align-items: center;"> ▼ </div> <div style="padding: 2px;"> <p>Data Engineering only</p> <p>Data Analytics only</p> <p>Data Engineering and Data Analytics</p> </div> </div>	Standard
Provide persistent clusters that support auto-scaling for analytics processes.	<div style="border: 1px solid black; padding: 2px;"> <div style="background-color: #f0f0f0; padding: 2px; display: flex; justify-content: space-between; align-items: center;"> ▼ </div> <div style="padding: 2px;"> <p>Data Engineering only</p> <p>Data Analytics only</p> <p>Data Engineering and Data Analytics</p> </div> </div>	<div style="border: 1px solid black; padding: 2px;"> <div style="background-color: #f0f0f0; padding: 2px; display: flex; justify-content: space-between; align-items: center;"> ▼ </div> <div style="padding: 2px;"> <p>Standard</p> <p>Premium</p> </div> </div>
Provide role-based access control (RBAC) support for Notebooks.	<div style="border: 1px solid black; padding: 2px;"> <div style="background-color: #f0f0f0; padding: 2px; display: flex; justify-content: space-between; align-items: center;"> ▼ </div> <div style="padding: 2px;"> <p>Data Engineering only</p> <p>Data Analytics only</p> <p>Data Engineering and Data Analytics</p> </div> </div>	<div style="border: 1px solid black; padding: 2px;"> <div style="background-color: #f0f0f0; padding: 2px; display: flex; justify-content: space-between; align-items: center;"> ▼ </div> <div style="padding: 2px;"> <p>Standard</p> <p>Premium</p> </div> </div>



Section: [none]

Explanation

Explanation/Reference:

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/>

QUESTION 11 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

Correct Answer: D

Section: [none]

Explanation

Explanation/Reference:

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>

QUESTION 12

HOTSPOT

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.

Hot Area:



Answer Area

Requirement	Service				
Integrate the on-premises data into the cloud.	<div style="border: 1px solid gray; padding: 2px;">▼</div> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr><td>Azure Databricks</td></tr> <tr><td>Azure Data Factory</td></tr> <tr><td>Azure SQL Data Warehouse</td></tr> <tr><td>Azure Batch</td></tr> </table>	Azure Databricks	Azure Data Factory	Azure SQL Data Warehouse	Azure Batch
Azure Databricks					
Azure Data Factory					
Azure SQL Data Warehouse					
Azure Batch					
Develop notebooks to transform the data.	<div style="border: 1px solid gray; padding: 2px;">▼</div> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr><td>Azure Databricks</td></tr> <tr><td>Azure Data Factory</td></tr> <tr><td>Azure SQL Data Warehouse</td></tr> <tr><td>Azure Batch</td></tr> </table>	Azure Databricks	Azure Data Factory	Azure SQL Data Warehouse	Azure Batch
Azure Databricks					
Azure Data Factory					
Azure SQL Data Warehouse					
Azure Batch					
Run notebooks.	<div style="border: 1px solid gray; padding: 2px;">▼</div> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr><td>Azure Databricks</td></tr> <tr><td>Azure Data Factory</td></tr> <tr><td>Azure SQL Data Warehouse</td></tr> <tr><td>Azure Batch</td></tr> </table>	Azure Databricks	Azure Data Factory	Azure SQL Data Warehouse	Azure Batch
Azure Databricks					
Azure Data Factory					
Azure SQL Data Warehouse					
Azure Batch					
Load the data.	<div style="border: 1px solid gray; padding: 2px;">▼</div> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr><td>Azure Databricks</td></tr> <tr><td>Azure Data Factory</td></tr> <tr><td>Azure SQL Data Warehouse</td></tr> <tr><td>Azure Batch</td></tr> </table>	Azure Databricks	Azure Data Factory	Azure SQL Data Warehouse	Azure Batch
Azure Databricks					
Azure Data Factory					
Azure SQL Data Warehouse					
Azure Batch					
Store the transformed data.	<div style="border: 1px solid gray; padding: 2px;">▼</div> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr><td>Azure Databricks</td></tr> <tr><td>Azure Data Factory</td></tr> <tr><td>Azure SQL Data Warehouse</td></tr> <tr><td>Azure Batch</td></tr> </table>	Azure Databricks	Azure Data Factory	Azure SQL Data Warehouse	Azure Batch
Azure Databricks					
Azure Data Factory					
Azure SQL Data Warehouse					
Azure Batch					



Correct Answer:

Answer Area

Requirement	Service
Integrate the on-premises data into the cloud.	<div style="border: 1px solid gray; padding: 2px;"> <div style="background-color: #cccccc; padding: 2px; display: flex; justify-content: space-between; align-items: center;"> ▼ </div> <div style="padding: 2px;"> <p>Azure Databricks</p> <p style="background-color: #e0f0e0;">Azure Data Factory</p> <p>Azure SQL Data Warehouse</p> <p>Azure Batch</p> </div> </div>
Develop notebooks to transform the data.	<div style="border: 1px solid gray; padding: 2px;"> <div style="background-color: #cccccc; padding: 2px; display: flex; justify-content: space-between; align-items: center;"> ▼ </div> <div style="padding: 2px;"> <p style="background-color: #e0f0e0;">Azure Databricks</p> <p>Azure Data Factory</p> <p>Azure SQL Data Warehouse</p> <p>Azure Batch</p> </div> </div>
Run notebooks.	<div style="border: 1px solid gray; padding: 2px;"> <div style="background-color: #cccccc; padding: 2px; display: flex; justify-content: space-between; align-items: center;"> ▼ </div> <div style="padding: 2px;"> <p style="background-color: #e0f0e0;">Azure Databricks</p> <p>Azure Data Factory</p> <p>Azure SQL Data Warehouse</p> <p>Azure Batch</p> </div> </div>
Load the data.	<div style="border: 1px solid gray; padding: 2px;"> <div style="background-color: #cccccc; padding: 2px; display: flex; justify-content: space-between; align-items: center;"> ▼ </div> <div style="padding: 2px;"> <p>Azure Databricks</p> <p>Azure Data Factory</p> <p>Azure SQL Data Warehouse</p> <p style="background-color: #e0f0e0;">Azure Batch</p> </div> </div>
Store the transformed data.	<div style="border: 1px solid gray; padding: 2px;"> <div style="background-color: #cccccc; padding: 2px; display: flex; justify-content: space-between; align-items: center;"> ▼ </div> <div style="padding: 2px;"> <p>Azure Databricks</p> <p>Azure Data Factory</p> <p style="background-color: #e0f0e0;">Azure SQL Data Warehouse</p> <p>Azure Batch</p> </div> </div>



Section: [none]
Explanation

Explanation/Reference:
Testlet 1

Case study

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other question on this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next sections of the exam. After you begin a new section, you cannot return to this section.

To start the case study

To display the first question on this case study, click the **Next** button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an **All Information** tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the **Question** button to return to the question.

Background

Trey Research is a technology innovator. The company partners with regional transportation department office to build solutions that improve traffic flow and safety.

The company is developing the following solutions:

Solution	Comments
Real Time Response	This solution will detect sudden changes in traffic flow including slow downs and stops that persist for more than one minute. The system will automatically dispatch emergency response vehicles to investigate issues. The solution will use a PySpark script to detect traffic flow changes. Script performance will be limited by available memory.
Backtrack	This solution will allow public safety officials to locate vehicles on roadways that implement traffic sensors. The solution must report changes in real time.
Planning Assistance	Transportation organizations will use Planning Assistance to analyze traffic data. The solution will allow users to define reports based on queries of the traffic data. The reports can be used for the following analyses: <ul style="list-style-type: none"> • current traffic load • correlation with recent local events such as sporting events • historical traffic • tracking the travel of a single vehicle

Regional transportation departments installed traffic sensor systems on major highways across North America. Sensors record the following information each time a vehicle passes in front of a sensor:

- Time
- Location in latitude and longitude
- Speed in kilometers per second (km/s)
- License plate number
- Length of vehicle in meters

Sensors provide data by using the following structure:

```
{
  "time" : "2014-09-15T23:14:25.72511732",
  "location" : {
    "type": "Point",
    "coordinates": [
      31.9,
      -4.8
    ]
  },
  "speed": 66.2,
  "license_plate": "WA-AJ0072W",
  "vehicle_length": 4.5
}
```

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

Traffic sensor data

- Sensors must have permission only to add items to the SensorData collection.
 - Traffic data insertion rate must be maximized.
 - Once every three months all traffic sensor data must be analyzed to look for data patterns that indicate sensor malfunctions.
- Sensor data must be stored in a Cosmos DB named treydata in a collection named SensorData
- The impact of vehicle images on sensor data throughout must be minimized.

Backtrack

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

Planning Assistance

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

Data from the Sensor Data collection will automatically be loaded into the Planning Assistance database once a week by using Azure Data Factory. You must be able to manually trigger the data load process. **Privacy**

and security policy

- Azure Active Directory must be used for all services where it is available.
 - For privacy reasons, license plate number information must not be accessible in Planning Assistance.
 - 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 must only be stored for seven years.

Performance and availability

- The report for Backtrack must execute as quickly as possible.
 - The SLA for Planning Assistance is 70 percent, and multiday outages are permitted.
 - All data must be replicated to multiple geographic regions to prevent data loss.
- You must maximize the performance of the Real Time Response system.

Financial requirements

Azure resource costs must be minimized where possible.

QUESTION 1

HOTSPOT

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

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

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

Requirement	Method
Authentication	<div style="border: 1px solid black; padding: 2px;"> <div style="background-color: #f0f0f0; padding: 2px; display: flex; justify-content: space-between; align-items: center;"> ▼ </div> <div style="border-top: 1px solid black; border-bottom: 1px solid black; padding: 2px;">HMAC header</div> <div style="border-top: 1px solid black; border-bottom: 1px solid black; padding: 2px;">Resource Token</div> <div style="border-top: 1px solid black; border-bottom: 1px solid black; padding: 2px;">Azure Managed Identity</div> <div style="border-top: 1px solid black; border-bottom: 1px solid black; padding: 2px;">Storage account connection string</div> </div>
Authorization	<div style="border: 1px solid black; padding: 2px;"> <div style="background-color: #f0f0f0; padding: 2px; display: flex; justify-content: space-between; align-items: center;"> ▼ </div> <div style="border-top: 1px solid black; border-bottom: 1px solid black; padding: 2px;">Custom RBAC role</div> <div style="border-top: 1px solid black; border-bottom: 1px solid black; padding: 2px;">Cosmos DB user</div> <div style="border-top: 1px solid black; border-bottom: 1px solid black; padding: 2px;">Azure Active Directory user</div> <div style="border-top: 1px solid black; border-bottom: 1px solid black; padding: 2px;">IoT device identity</div> </div>

Correct Answer:

Answer Area

Requirement	Method
Authentication	<div style="border: 1px solid black; padding: 2px;"> <div style="background-color: #e0e0e0; padding: 2px; display: flex; justify-content: space-between; align-items: center;"> ▼ </div> <div style="padding: 2px;"> <p>HMAC header</p> <p style="background-color: #d9ead3;">Resource Token</p> <p>Azure Managed Identity</p> <p>Storage account connection string</p> </div> </div>
Authorization	<div style="border: 1px solid black; padding: 2px;"> <div style="background-color: #e0e0e0; padding: 2px; display: flex; justify-content: space-between; align-items: center;"> ▼ </div> <div style="padding: 2px;"> <p>Custom RBAC role</p> <p style="background-color: #d9ead3;">Cosmos DB user</p> <p>Azure Active Directory user</p> <p>IoT device identity</p> </div> </div>



Section: [none]
Explanation

Explanation/Reference:
Explanation:

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

Box 1: Resource Token
Resource tokens provide access to the application resources within a Cosmos DB database.
Enable clients to read, write, and delete resources in the Cosmos DB account according to the permissions they've been granted.

Box 2: Cosmos DB user
You can use a resource token (by creating Cosmos DB users and permissions) when you want to provide access to resources in your Cosmos DB account to a client that cannot be trusted with the master key.

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

QUESTION 2
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.

Hot Area:

Answer Area

Setting	Value
Audit log destination	<div style="border: 1px solid black; padding: 2px;"><div style="background-color: #cccccc; padding: 2px; display: flex; justify-content: space-between; align-items: center;">▼</div><ul style="list-style-type: none">Storage queueEvent HubEvent GridBlob storage</div>
Detection app service	<div style="border: 1px solid black; padding: 2px;"><div style="background-color: #cccccc; padding: 2px; display: flex; justify-content: space-between; align-items: center;">▼</div><ul style="list-style-type: none">Function AppWeb AppAPI App</div>

Correct Answer:

Answer Area

Setting	Value
Audit log destination	<div style="border: 1px solid black; padding: 2px;"><div style="background-color: #cccccc; padding: 2px; display: flex; justify-content: space-between; align-items: center;">▼</div><ul style="list-style-type: none">Storage queueEvent HubEvent Grid<li style="background-color: #d9ead3;">Blob storage</div>
Detection app service	<div style="border: 1px solid black; padding: 2px;"><div style="background-color: #cccccc; padding: 2px; display: flex; justify-content: space-between; align-items: center;">▼</div><ul style="list-style-type: none">Function App<li style="background-color: #d9ead3;">Web AppAPI App</div>

Section: [none]

Explanation**Explanation/Reference:**

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-applications/>



Testlet 2

Case study

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other question on this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next sections of the exam. After you begin a new section, you cannot return to this section.

To start the case study

To display the first question on this case study, click the **Next** button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an **All Information** tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the **Question** button to return to the question.

Requirements

Business

The company identifies the following business requirements:

- You must transfer all images and customer data to cloud storage and remove on-premises servers.
- You must develop an analytical processing solution for transforming customer data.
- You must develop an image object and color tagging solution.
- Capital expenditures must be minimized.
- Cloud resource costs must be minimized.

Technical

The solution has the following technical requirements:

- 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.
- Image data must be stored in a single data store at minimum cost.
- Customer data must be analyzed using managed Spark clusters.
- Power BI must be used to visualize transformed customer data.
- All data must be backed up in case disaster recovery is required.

Security and optimization

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

- parallel processing of customer data
- hyper-scale storage of images
- global region data replication of processed image data

QUESTION 1

DRAG DROP

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.

Select and Place:



Answer Area

Encryption methods	Solution component	Encryption method
Encryption at rest	Tagging data	
Transparent data encryption	Processed customer data	
Azure Key Vault		

Correct Answer:

Answer Area

Encryption methods	Solution component	Encryption method
Encryption at rest	Tagging data	Transparent data encryption
Transparent data encryption	Processed customer data	Encryption at rest
Azure Key Vault		

Section: [none]

Explanation

Explanation/Reference:

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=sql-server-2017> <https://docs.microsoft.com/en-us/azure/security/azure-security-encryption-atrest>

Testlet 3

Case study

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other question on this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next sections of the exam. After you begin a new section, you cannot return to this section.

To start the case study

To display the first question on this case study, click the **Next** button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an **All Information** tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the **Question** button to return to the question.

Background

Current environment

The company has the following virtual machines (VMs):

VM	Roles	Database size	VM type	Destination
CONT_SQL1	Microsoft SQL Server	2 TB	Hyper-V	Azure SQL Database
CONT_SQL2	Microsoft SQL Server	2 TB	Hyper-V	Azure SQL Database
CONT_SQL3	Microsoft SQL Server	100 GB	Hyper-V	Azure VM
CONT_SAP1	SAP	1 TB	Vmware	On-premises
CONT_SAP2	SAP	1 TB	Vmware	On-premises
CPNT_SSRS	Microsoft SQL Server Reporting Services	1 TB	Hyper-V	Azure VM



Requirements

Storage and processing

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. The architecture will need to support data files, libraries, and images. Additionally, it must provide a web-based interface to documents that contain runnable command, visualizations, and narrative text such as a notebook.

CONT_SQL3 requires an initial scale of 35000 IOPS.

CONT_SQL1 and CONT_SQL2 must use the vCore model and should include replicas. The solution must support 8000 IOPS. The storage should be configured to optimized storage for database OLTP workloads.

Migration

- You must be able to independently scale compute and storage resources.
 - You must migrate all SQL Server workloads to Azure. You must identify related machines in the on-premises environment, get disk size data usage information.
 - Data from SQL Server must include zone redundant storage.
 - You need to ensure that app components can reside on-premises while interacting with components that run in the Azure public cloud. ▪
- SAP data must remain on-premises.
- The Azure Site Recovery (ASR) results should contain per-machine data.

Business requirements

- You must design a regional disaster recovery topology.
- The database backups have regulatory purposes and must be retained for seven years.
- CONT_SQL1 stores customers sales data that requires ETL operations for data analysis. A solution is required that reads data from SQL, performs ETL, and outputs to Power BI. The solution should use managed clusters to minimize costs. To optimize logistics, Contoso needs to analyze customer sales data to see if certain products are tied to specific times in the year. ▪ The analytics solution for customer sales data must be available during a regional outage.

Security and auditing

- Contoso requires all corporate computers to enable Windows Firewall.
- Azure servers should be able to ping other Contoso Azure servers.
- Employee PII must be encrypted in memory, in motion, and at rest. Any data encrypted by SQL Server must support equality searches, grouping, indexing, and joining on the encrypted data. ▪ Keys must be secured by using hardware security modules (HSMs).
- CONT_SQL3 must not communicate over the default ports

Cost

- All solutions must minimize cost and resources.
- The organization does not want any unexpected charges.
- The data engineers must set the SQL Data Warehouse compute resources to consume 300 DWUs.
- CONT_SQL2 is not fully utilized during non-peak hours. You must minimize resource costs for during non-peak hours.

QUESTION 1

HOTSPOT

You need to design network access to the SQL Server data.

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

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

Solution component	Value
Tool	<div style="border: 1px solid gray; padding: 2px;"> <div style="background-color: #f0f0f0; padding: 2px; display: flex; justify-content: space-between; align-items: center;"> ▼ </div> <div style="padding: 2px;"> 1433 8080 </div> </div>
SQL Server Network Configuration port	<div style="border: 1px solid gray; padding: 2px;"> <div style="background-color: #f0f0f0; padding: 2px; display: flex; justify-content: space-between; align-items: center;"> ▼ </div> <div style="padding: 2px;"> SQL Server Configuration Manager SQL Server Management Studio </div> </div>

Correct Answer:

Answer Area

Solution component	Value
Tool	<input type="text" value="1433"/> <ul style="list-style-type: none"> 1433 <li style="background-color: #d9ead3;">8080
SQL Server Network Configuration port	<input type="text" value="SQL Server Configuration Manager"/> <ul style="list-style-type: none"> <li style="background-color: #d9ead3;">SQL Server Configuration Manager SQL Server Management Studio

Section: [none]

Explanation

Explanation/Reference:

Explanation:

Box 1: 8080

1433 is the default port, but we must change it as CONT_SQL3 must not communicate over the default ports. Because port 1433 is the known standard for SQL Server, some organizations specify that the SQL Server port number should be changed to enhance security.

Box 2: SQL Server Configuration Manager

You can configure an instance of the SQL Server Database Engine to listen on a specific fixed port by using the SQL Server Configuration Manager.

References:

<https://docs.microsoft.com/en-us/sql/database-engine/configure-windows/configure-a-server-to-listen-on-a-specific-tcp-port?view=sql-server-2017>

Question Set 4

QUESTION 1 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

Correct Answer: B

Section: [none]

Explanation

Explanation/Reference:

Incorrect Answers:

A: Transparent Data Encryption (TDE) encrypts SQL Server, Azure SQL Database, and Azure SQL Data Warehouse data files, known as encrypting data at rest. TDE does not provide encryption across communication channels.

References: <https://cloudblogs.microsoft.com/sqlserver/2018/12/17/confidential-computing-using-always-encrypted-with-secure-enclaves-in-sql-server-2019-preview/>

QUESTION 2

DRAG DROP

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.

NOTE: Each correct selection is worth one point.

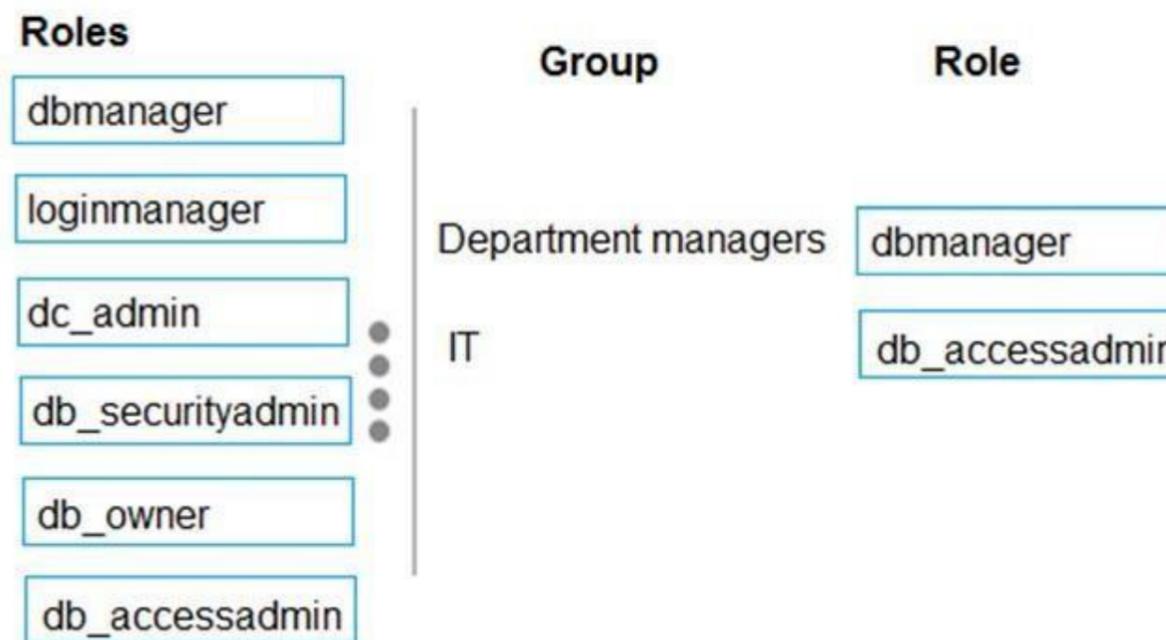
Select and Place:

Answer Area

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

Correct Answer:

Answer Area



Section: [none]
Explanation

Explanation/Reference:
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>

QUESTION 3 You plan to use Azure SQL Database to support a line of business app.

You need to identify sensitive data that is stored in the database and monitor access to the data.

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

NOTE: Each correct selection is worth one point.

- A. Enable Data Discovery and Classification.
- B. Implement Transparent Data Encryption (TDE).
- C. Enable Auditing.
- D. Run Vulnerability Assessment.
- E. Use Advanced Threat Protection.

Correct Answer: CDE

Section: [none]

Explanation

Explanation/Reference:

Testlet 1

Case study

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other question on this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next sections of the exam. After you begin a new section, you cannot return to this section.

To start the case study

To display the first question on this case study, click the **Next** button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an **All Information** tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the **Question** button to return to the question.

Requirements

Business

The company identifies the following business requirements:

- You must transfer all images and customer data to cloud storage and remove on-premises servers.
- You must develop an analytical processing solution for transforming customer data.
- You must develop an image object and color tagging solution.
- Capital expenditures must be minimized.
- Cloud resource costs must be minimized.



Technical

The solution has the following technical requirements:

- 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.
- Image data must be stored in a single data store at minimum cost.
- Customer data must be analyzed using managed Spark clusters.
- Power BI must be used to visualize transformed customer data.
- All data must be backed up in case disaster recovery is required.

Security and optimization

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

- parallel processing of customer data
- hyper-scale storage of images
- global region data replication of processed image data

QUESTION 1 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

Correct Answer: C

Section: [none]

Explanation

Explanation/Reference:

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>



Testlet 2

Case study

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other question on this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next sections of the exam. After you begin a new section, you cannot return to this section.

To start the case study

To display the first question on this case study, click the **Next** button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an **All Information** tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the **Question** button to return to the question.

Background

Current environment

The company has the following virtual machines (VMs):

VM	Roles	Database size	VM type	Destination
CONT_SQL1	Microsoft SQL Server	2 TB	Hyper-V	Azure SQL Database
CONT_SQL2	Microsoft SQL Server	2 TB	Hyper-V	Azure SQL Database
CONT_SQL3	Microsoft SQL Server	100 GB	Hyper-V	Azure VM
CONT_SAP1	SAP	1 TB	Vmware	On-premises
CONT_SAP2	SAP	1 TB	Vmware	On-premises
CPNT_SSRS	Microsoft SQL Server Reporting Services	1 TB	Hyper-V	Azure VM

Requirements

Storage and processing

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. The architecture will need to support data files, libraries, and images. Additionally, it must provide a web-based interface to documents that contain runnable command, visualizations, and narrative text such as a notebook.

CONT_SQL3 requires an initial scale of 35000 IOPS.

CONT_SQL1 and CONT_SQL2 must use the vCore model and should include replicas. The solution must support 8000 IOPS. The storage should be configured to optimized storage for database OLTP workloads.

Migration

- You must be able to independently scale compute and storage resources.
- You must migrate all SQL Server workloads to Azure. You must identify related machines in the on-premises environment, get disk size data usage information.
- Data from SQL Server must include zone redundant storage.
- You need to ensure that app components can reside on-premises while interacting with components that run in the Azure public cloud. ▪
- SAP data must remain on-premises.
- The Azure Site Recovery (ASR) results should contain per-machine data.

Business requirements

- You must design a regional disaster recovery topology.
- The database backups have regulatory purposes and must be retained for seven years.

- CONT_SQL1 stores customers sales data that requires ETL operations for data analysis. A solution is required that reads data from SQL, performs ETL, and outputs to Power BI. The solution should use managed clusters to minimize costs. To optimize logistics, Contoso needs to analyze customer sales data to see if certain products are tied to specific times in the year.
- The analytics solution for customer sales data must be available during a regional outage.

Security and auditing

- Contoso requires all corporate computers to enable Windows Firewall.
- Azure servers should be able to ping other Contoso Azure servers.
- Employee PII must be encrypted in memory, in motion, and at rest. Any data encrypted by SQL Server must support equality searches, grouping, indexing, and joining on the encrypted data.
- Keys must be secured by using hardware security modules (HSMs).
- CONT_SQL3 must not communicate over the default ports

Cost

- All solutions must minimize cost and resources.
- The organization does not want any unexpected charges.
- The data engineers must set the SQL Data Warehouse compute resources to consume 300 DWUs.
- CONT_SQL2 is not fully utilized during non-peak hours. You must minimize resource costs for during non-peak hours.

QUESTION 1

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.

Correct Answer: B

Section: [none]

Explanation

Explanation/Reference:

Explanation:

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

QUESTION 2 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.

Correct Answer: ADE

Section: [none]

Explanation

Explanation/Reference:

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>



Question Set 3

QUESTION 1

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

Correct Answer: B

Section: [none]

Explanation

Explanation/Reference:

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>



QUESTION 2

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

Correct Answer: A

Section: [none]

Explanation

Explanation/Reference:

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>

QUESTION 3

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

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

Shops will upload data every 10 days.

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

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

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

Does the solution meet the goal?

- A. Yes
- B. No

Correct Answer: B
Section: [none]
Explanation



Explanation/Reference:

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>

QUESTION 4

HOTSPOT

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.

Hot Area:

Answer Area

Restore order	Backup type
first	<div style="border: 1px solid black; padding: 2px;"> <div style="background-color: #f0f0f0; padding: 2px; display: flex; justify-content: space-between; align-items: center;"> ▼ </div> <div style="padding: 2px;"> <ul style="list-style-type: none"> full weekly backup full daily backup differential weekly backup differential daily backup </div> </div>
second	<div style="border: 1px solid black; padding: 2px;"> <div style="background-color: #f0f0f0; padding: 2px; display: flex; justify-content: space-between; align-items: center;"> ▼ </div> <div style="padding: 2px;"> <ul style="list-style-type: none"> full daily backup differential backup from the last 12 hours all differential backups since the last full backup all log backups since the last full backup </div> </div>
third	<div style="border: 1px solid black; padding: 2px;"> <div style="background-color: #f0f0f0; padding: 2px; display: flex; justify-content: space-between; align-items: center;"> ▼ </div> <div style="padding: 2px;"> <ul style="list-style-type: none"> all log backups since the last differential backup differential backup from the last 12 hours all differential backups since the last full backup all log backups since the last full backup </div> </div>

Correct Answer:

Answer Area

Restore order	Backup type
first	<div style="border: 1px solid black; padding: 2px;"> <div style="background-color: #e0e0e0; padding: 2px;">▼</div> <div style="background-color: #d9ead3; padding: 2px;">full weekly backup</div> <div style="padding: 2px;">full daily backup</div> <div style="padding: 2px;">differential weekly backup</div> <div style="padding: 2px;">differential daily backup</div> </div>
second	<div style="border: 1px solid black; padding: 2px;"> <div style="background-color: #e0e0e0; padding: 2px;">▼</div> <div style="padding: 2px;">full daily backup</div> <div style="background-color: #d9ead3; padding: 2px;">differential backup from the last 12 hours</div> <div style="padding: 2px;">all differential backups since the last full backup</div> <div style="padding: 2px;">all log backups since the last full backup</div> </div>
third	<div style="border: 1px solid black; padding: 2px;"> <div style="background-color: #e0e0e0; padding: 2px;">▼</div> <div style="background-color: #d9ead3; padding: 2px;">all log backups since the last differential backup</div> <div style="padding: 2px;">differential backup from the last 12 hours</div> <div style="padding: 2px;">all differential backups since the last full backup</div> <div style="padding: 2px;">all log backups since the last full backup</div> </div>

Section: [none]

Explanation

Explanation/Reference:

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/>

QUESTION 5 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.

Correct Answer: BC

Section: [none]

Explanation

Explanation/Reference:

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>

QUESTION 6

A company is developing a mission-critical line of business app that uses Azure SQL Database Managed Instance.

You must design a disaster recovery strategy for the solution/

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

What should you recommend?

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



Correct Answer: A

Section: [none]

Explanation

Explanation/Reference:

Explanation:

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

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

QUESTION 7

HOTSPOT

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.

Hot Area:

Answer Area

Option	Value
Service tier	<input type="text" value=""/> Basic Standard General Premium
Redundancy type	<input type="text" value=""/> SQL Sync Zone-redundancy Geo-redundant storage



Correct Answer:

Answer Area

Option	Value
Service tier	<div style="border: 1px solid gray; padding: 2px;"> <div style="background-color: #f0f0f0; padding: 2px; display: flex; justify-content: space-between; align-items: center;"> ▼ </div> <ul style="list-style-type: none"> Basic <li style="background-color: #d9ead3;">Standard General Premium </div>
Redundancy type	<div style="border: 1px solid gray; padding: 2px;"> <div style="background-color: #f0f0f0; padding: 2px; display: flex; justify-content: space-between; align-items: center;"> ▼ </div> <ul style="list-style-type: none"> SQL Sync Zone-redundancy <li style="background-color: #d9ead3;">Geo-redundant storage </div>



Section: [none]
Explanation

Explanation/Reference:
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>

QUESTION 8 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

Correct Answer: C

Section: [none]

Explanation

Explanation/Reference:

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>

QUESTION 9 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

Correct Answer: B

Section: [none]

Explanation

Explanation/Reference:

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>



QUESTION 10 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

Correct Answer: C

Section: [none]

Explanation

Explanation/Reference:

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 vCore-based purchasing model.

Incorrect Answers:

D: Database sharding is a type of horizontal partitioning that splits large databases into smaller components, which are faster and easier to manage.

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

QUESTION 11 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

Correct Answer: A

Section: [none]

Explanation

Explanation/Reference:

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>

QUESTION 12

You plan to deploy an Azure SQL Database instance to support an application. You plan to use the DTU-based 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.

Correct Answer: DE

Section: [none]

Explanation

Explanation/Reference:

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.

Incorrect Answers:

B: Basic tier only allows restore points within 7 days.

References: <https://docs.microsoft.com/en-us/azure/sql-database/sql-database-long-term-retention>

