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.

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 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 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. Premium
D. 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 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.

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 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 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 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 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 4
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.

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

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

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

Proposed Solution: Separate data into shards by using horizontal partitioning.

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) selfsharding. An elastic query is used to query or compile reports across many shards.

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

QUESTION 5
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 6
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 7
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:
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

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 2**

QUESTION 1
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 tableD. 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 roundrobin 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 2
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 3
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 4
A company installs IoT devices to monitor its fleet of delivery vehicles. Data from devices is collected from Azure Event Hub.

The data must be transmitted to Power BI for real-time data visualizations.

You need to recommend a solution.

What should you recommend?

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

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 5
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 6
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:

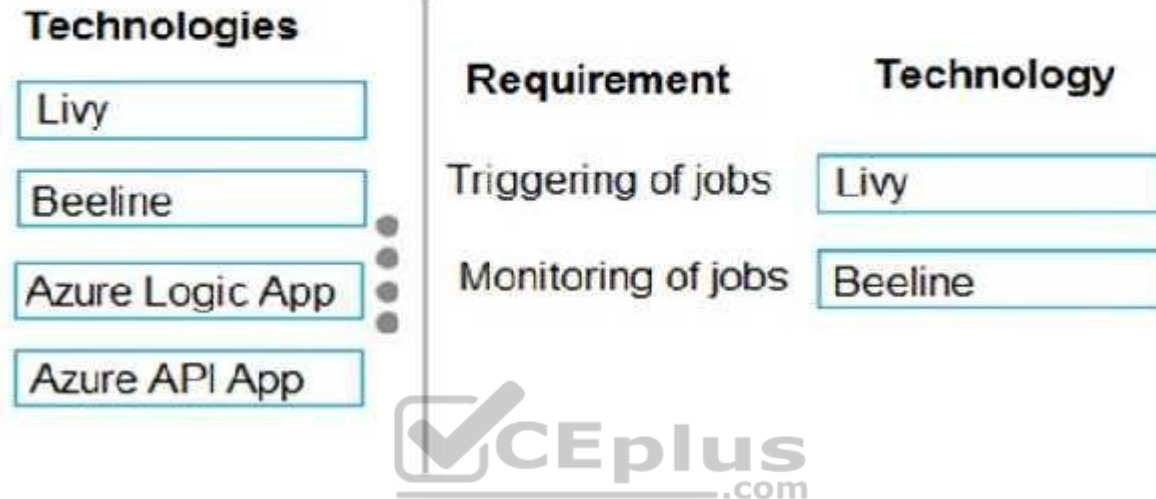## Answer Area

Technologies

| Livy |

| Beeline |

| Azure Logic App |

| Azure API App |

Requirement   **Technology**

Triggering of jobs   [   ]

Monitoring of jobs   [   ]

Correct Answer:

Answer Area

Technologies

Livy

Beeline

Azure Logic App

Azure API App

Requirement	**Technology**
Triggering of jobs	Livy
Monitoring of jobs	Beeline

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 7**

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 8
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 9
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

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: 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 2
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	▼
	full weekly backup
	full daily backup
	differential weekly backup
	differential daily backup
second	▼
	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
third	▼
	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

Correct Answer:

Answer Area

Restore order	Backup type
first	**full weekly backup** ▼
	full weekly backup
	full daily backup
	differential weekly backup
	differential daily backup
second	**differential backup from the last 12 hours** ▼
	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
third	**all log backups since the last differential backup** ▼
	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

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 3
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 4
A company is designing a solution that uses Azure Databricks.

The solution must be resilient to regional Azure datacenter outages.

You need to recommend the redundancy type for the solution.

What should you recommend?

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

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 5
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 6
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