DUMPSBOSS.

Designing an Azure Data Solution

Microsoft DP-201

Version Demo

Total Demo Questions: 15

Total Premium Questions: 209

Buy Premium PDF

https://dumpsboss.co

support@dumpsboss.co

support@dumpsboss.co
dumpsboss.co

Topic Break Down

Topic	No. of Questions
Topic 1, Case Study 1	5
Topic 2, Case Study 2	4
Topic 3, Case Study 3	3
Topic 4, Case Study 4	5
Topic 5, Case Study 5	3
Topic 6, Case Study 6	2
Topic 7, Case Study 7	4
Topic 8, Case Study 8	4
Topic 9, Case Study 9	4
Topic 10, Case Study 10	2
Topic 11, Case Study 11	2
Topic 12, Case Study 12	2
Topic 13, Case Study 13	2
Topic 14, Mixed Questions	167
Total	209

QUESTION NO: 1

You are designing a solution that will use Azure Databricks and Azure Data Lake Storage Gen2.

From Databricks, you need to access Data Lake Storage directly by using a service principal.

What should you include in the solution?

A. shared access signatures (SAS) in Data Lake Storage

B. access keys in Data Lake Storage

C. an organizational relationship in Azure Active Directory (Azure AD)

D. an application registration in Azure Active Directory (Azure AD)

ANSWER: D

Explanation:

Create and grant permissions to service principal

If your selected the access method requires a service principal with adequate permissions, and you do not have one, follow these steps:

1. Create an Azure AD application and service principal that can access resources. Note thefollowing properties:

- client-id: An ID that uniquely identifies the application.
- directory-id: An ID that uniquely identifies the Azure AD instance.
- service-credential: A string that the application uses to prove its identity.

2. Register the service principal, granting the correct role assignment, such as Storage Blob Data

3. Contributor, on the Azure Data Lake Storage Gen2 account.

Reference: https://docs.databricks.com/data/data-sources/azure/azure-datalake-gen2.html

QUESTION NO: 2 - (HOTSPOT)

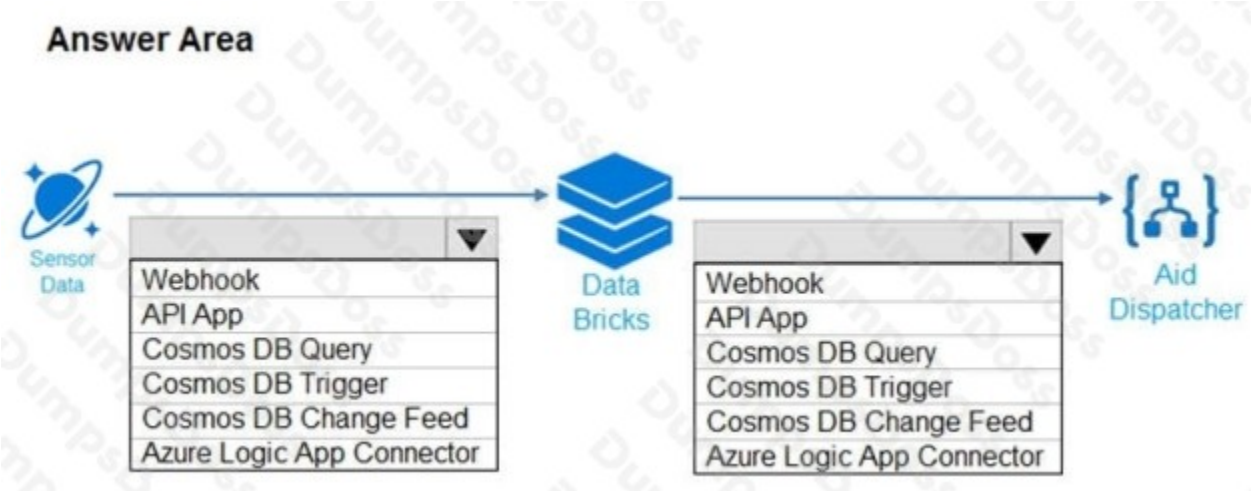
HOTSPOT

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

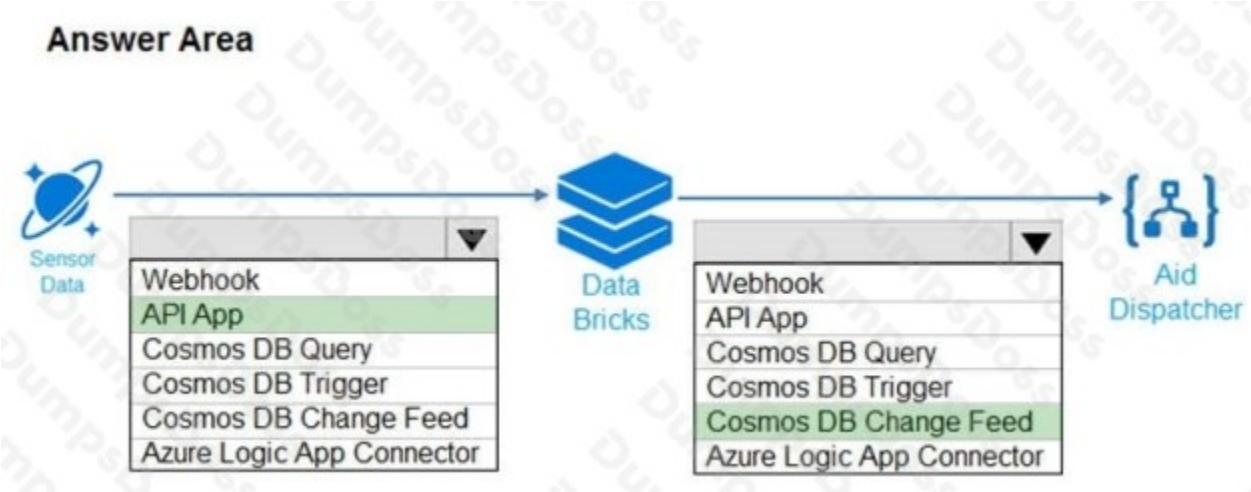
How should you design the processing system? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

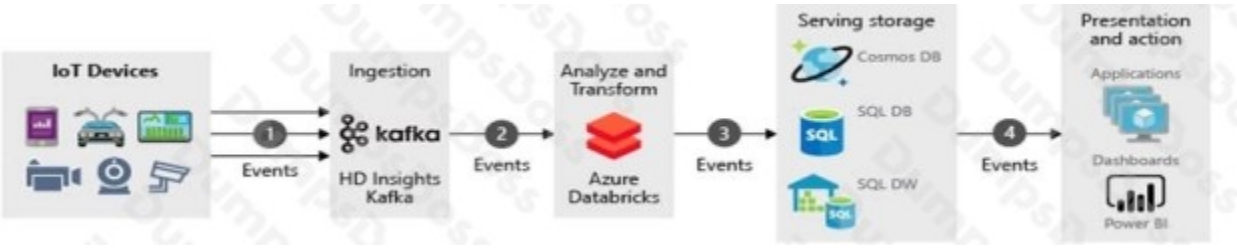


ANSWER:



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.

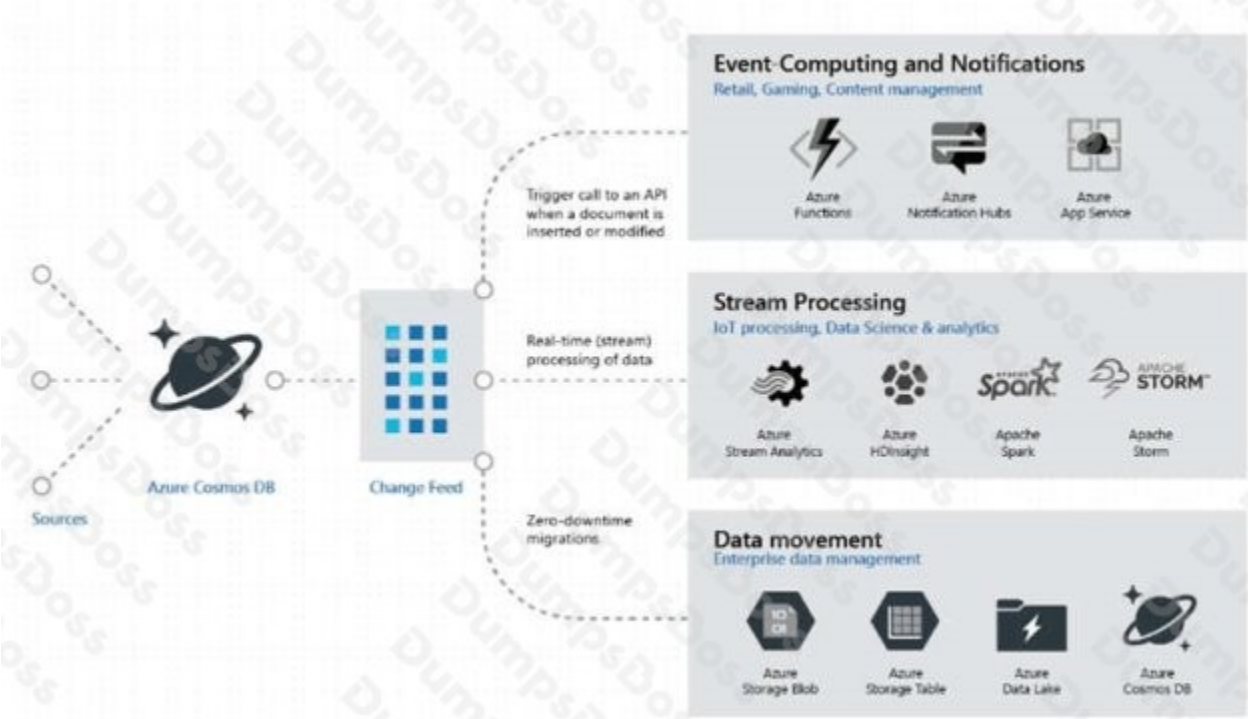
3. Downstream storage services, like Azure Cosmos DB, Azure Synapse Analytics, or Azure SQL DB, will then be a data source for presentation and action layer.

4. Business analysts can use Microsoft Power BI to analyze warehoused data. Other applications can be built upon the serving 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:



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

QUESTION NO: 3

You have a MongoDB database that you plan to migrate to an Azure Cosmos DB account that uses the MongoDB API.

During testing, you discover that the migration takes longer than expected.

You need to recommend a solution that will reduce the amount of time it takes to migrate the data.

What are two possible recommendations to achieve this goal? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

A. Increase the Request Units (RUs).

B. Turn off indexing.

C. Add a write region.

D. Create unique indexes.

E. Create compound indexes.

ANSWER: A B

Explanation:

A: Increase the throughput during the migration by increasing the Request Units (RUs).

For customers that are migrating many collections within a database, it is strongly recommend to configure database-level throughput. You must make this choice when you create the database. The minimum database-level throughput capacity is 400 RU/sec. Each collection sharing database-level throughput requires at least 100 RU/sec.

B: By default, Azure Cosmos DB indexes all your data fields upon ingestion. You can modify the indexing policy in Azure Cosmos DB at any time. In fact, it is often recommended to turn off indexing when migrating data, and then turn it back on when the data is already in Cosmos DB.

Reference: https://docs.microsoft.com/bs-latn-ba/Azure/cosmos-db/mongodb-pre-migration

QUESTION NO: 4

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

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

NOTE: Each correct selection is worth one point.

A. Provision multiple Azure Databricks workspaces in separate Azure regions.

B. Migrate users, notebooks, and cluster configurations from one workspace to another in the same region.

C. Use zone redundant storage.

D. Migrate users, notebooks, and cluster configurations from one region to another.

E. Use Geo-redundant storage.

F. Provision a second Azure Databricks workspace in the same region.

ANSWER: A D E

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.

Reference:

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

Design for high availability and disaster recovery

QUESTION NO: 5

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.

ANSWER: B C D

QUESTION NO: 6

You are designing security for administrative access to Azure Synapse Analytics.

You need to recommend a solution to ensure that administrators use two-factor authentication when accessing the data warehouse from Microsoft SQL Server Management Studio (SSMS).

What should you include in the recommendation?

A. Azure conditional access policies

B. Azure Active Directory (Azure AD) Privileged Identity Management (PIM)

C. Azure Key Vault secrets

D. Azure Active Directory (Azure AD) Identity Protection

ANSWER: A

Explanation:

Reference:

https://docs.microsoft.com/en-us/azure/sql-database/sql-database-conditional-access

QUESTION NO: 7

A company purchases IoT devices to monitor manufacturing machinery. The company uses an IoT appliance to communicate with the IoT devices.

The company must be able to monitor the devices in real-time.

You need to design the solution.

What should you recommend?

A. Azure Data Factory instance using Azure Portal

B. Azure Analysis Services using Microsoft Visual Studio

C. Azure Stream Analytics cloud job using Azure Portal

D. Azure Data Factory instance using Azure Portal

ANSWER: C

Explanation:

The Stream Analytics query language allows to perform CEP (Complex Event Processing) by offering a wide array of functions for analyzing streaming data. This query language supports simple data manipulation, aggregation and analytics functions, geospatial functions, pattern matching and anomaly detection. You can edit queries in the portal or using our development tools, and test them using sample data that is extracted from a live stream.

Note: Stream Analytics is a cost-effective event processing engine that helps uncover real-time insights from devices, sensors, infrastructure, applications and data quickly and easily.

Monitor and manage Stream Analytics resources with Azure PowerShell cmdlets and powershell scripting that execute basic Stream Analytics tasks.

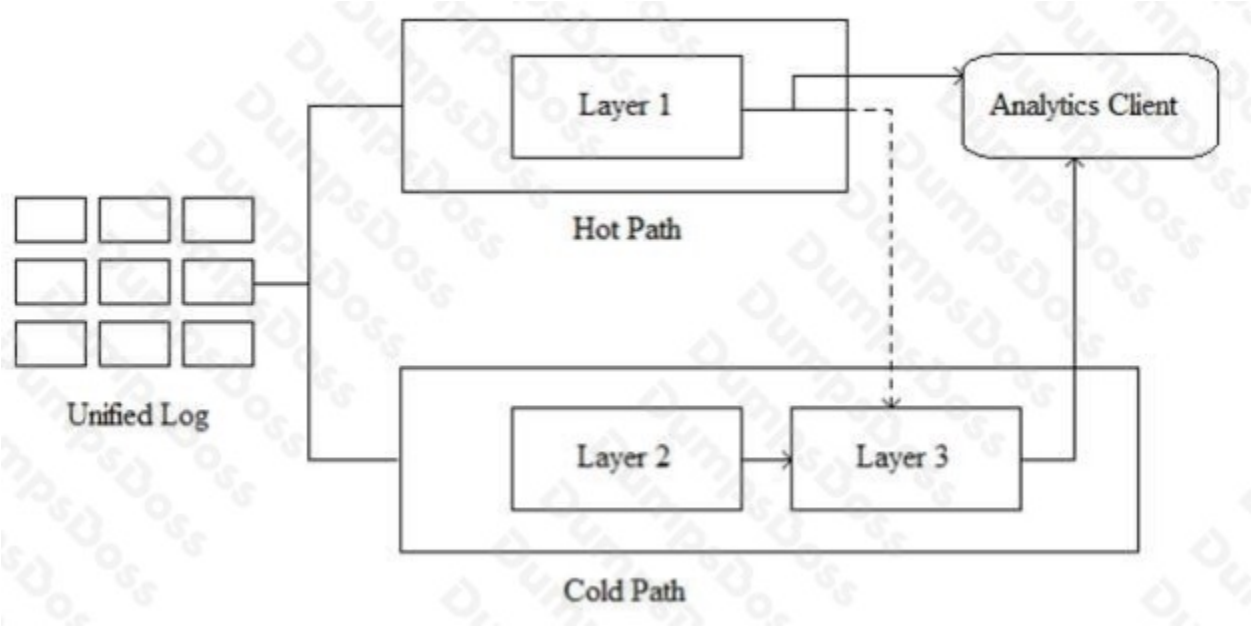
Reference:

https://cloudblogs.microsoft.com/sqlserver/2014/10/29/microsoft-adds-iot-streaming-analyticsdata-production-and-workflow-services-to-azure/ https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-introduction

QUESTION NO: 8 - (DRAG DROP)

DRAG DROP

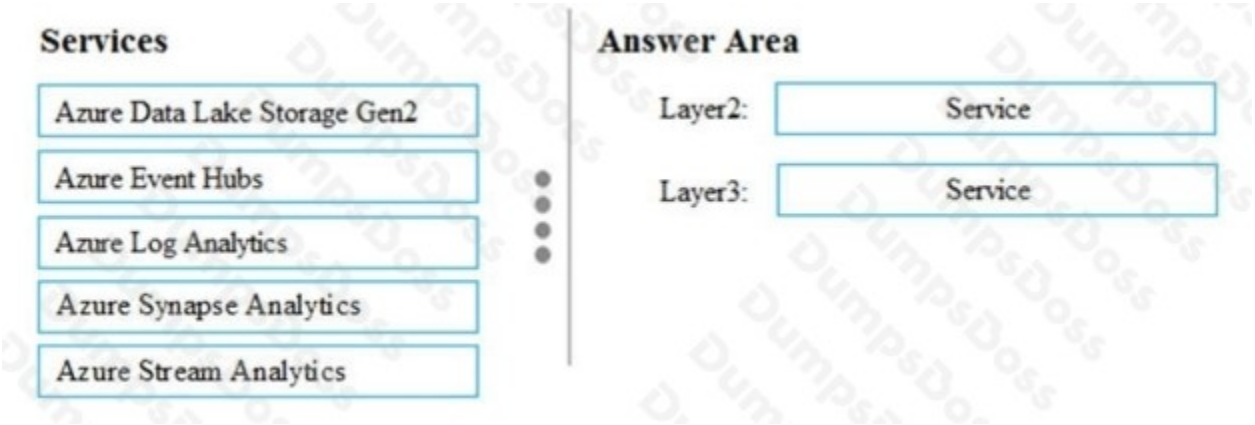
You are planning a design pattern based on the Lambda architecture as shown in the exhibit.



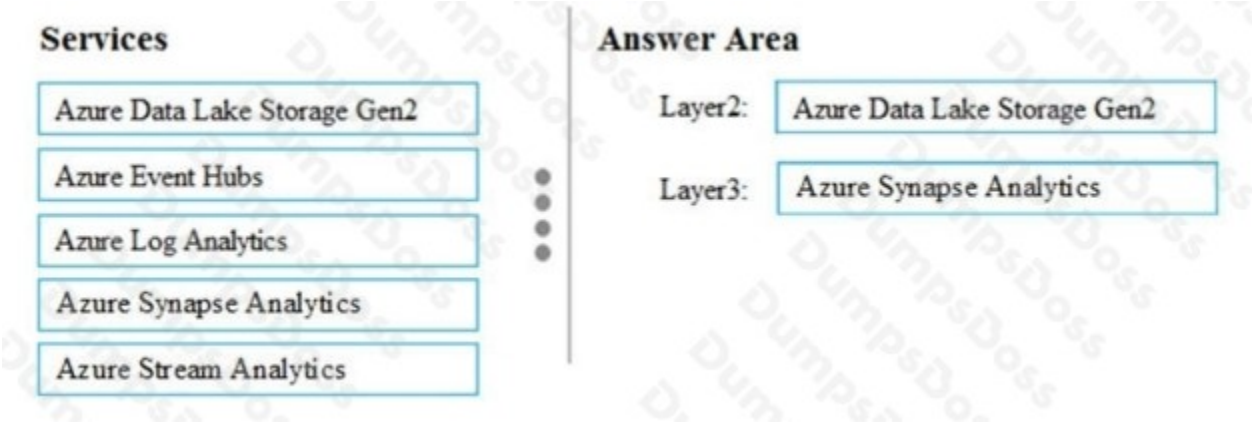
Which Azure services should you use for the cold path? To answer, drag the appropriate services to the correct layers. Each service may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

Select and Place:



ANSWER:



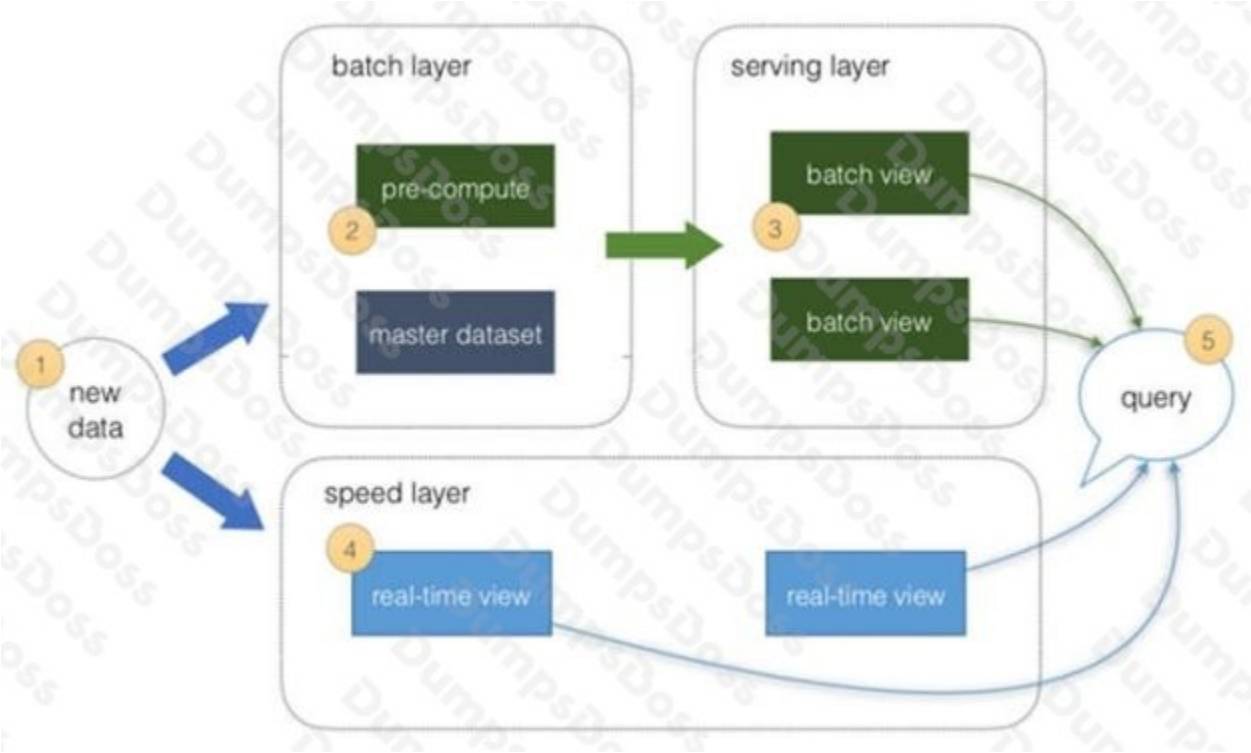
Explanation:

Layer 2: Azure Data Lake Storage Gen2

Layer 3: Azure Synapse Analytics

Azure Synapse Analytics can be used for batch processing.

Note: Lambda architectures use batch-processing, stream-processing, and a serving layer to minimize the latency involved in querying big data.



Reference: https://azure.microsoft.com/en-us/blog/lambda-architecture-using-azure-cosmosdb-faster-performance-low-tco-low-devops/ https://docs.microsoft.com/en-us/azure/architecture/data-guide/technology-choices/batch-processing

QUESTION NO: 9 - (DRAG DROP)

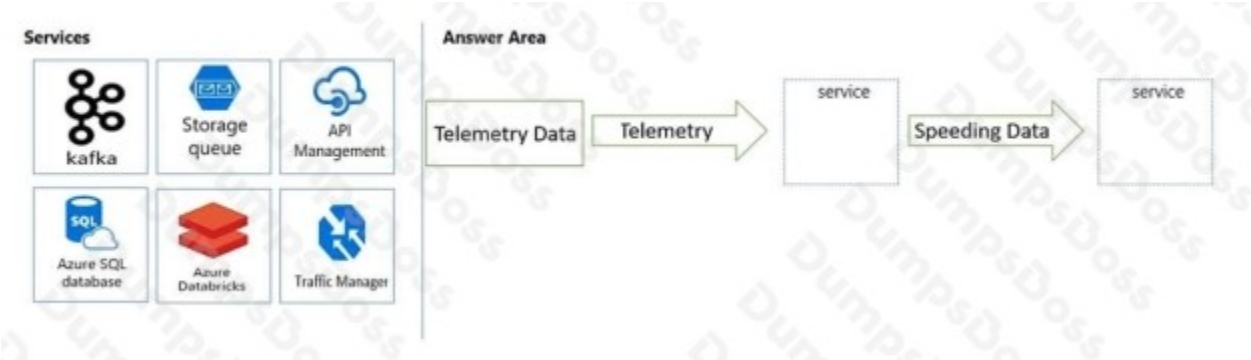
DRAG DROP

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

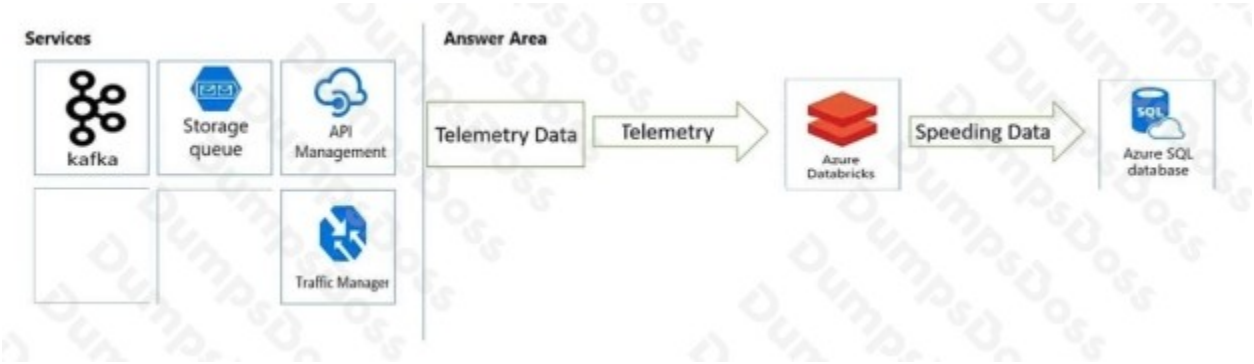
How should you design the pipeline? To answer, drag the appropriate services to the correct locations. Each service may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

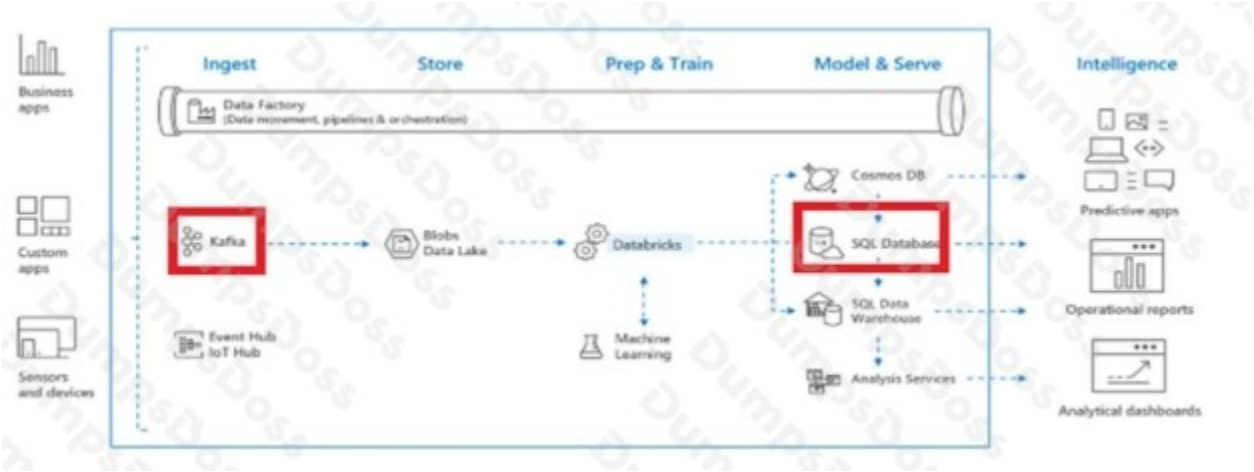
Select and Place:



ANSWER:



Explanation:



Scenario:

Information about vehicles that have been detected as going over the speed limit during the last 30 minutes must be available to law enforcement officers. Several law enforcement organizations may respond to speeding vehicles.

Telemetry Capture

The telemetry capture system records each time a vehicle passes in front of a sensor. The sensors run on a custom embedded operating system and record the following telemetry data:

- Time
- Location in latitude and longitude
- Speed in kilometers per hour (kmph) ▪ Length of vehicle in meters

Reference:

https://docs.microsoft.com/en-us/azure/azure-databricks/what-is-azure-databricks

QUESTION NO: 10

You need to design a solution that meets the business requirements of Health Insights.

What should you include in the recommendation?

A. Azure Cosmos DB that uses the Gremlin API

B. Azure Data Factory

C. Azure Cosmos DB that uses the SQL API

D. Azure Databricks

ANSWER: D

Explanation:

Azure Synapse Analytics is a cloud-based enterprise data warehouse that leverages massively parallel processing (MPP) to quickly run complex queries across petabytes of data. Use SQL Data Warehouse as a key component of a big data solution.

You can access Azure Synapse Analytics (SQL DW) from Databricks using the SQL Data

Warehouse connector (referred to as the SQL DW connector), a data source implementation for Apache Spark that uses Azure Blob Storage, and PolyBase in SQL DW to transfer large volumes of data efficiently between a Databricks cluster and a SQL DW instance.

Scenario: ADatum identifies the following requirements for the Health Insights application: ▪ The new Health Insights application must be built on a massively parallel processing (MPP) architecture that will support the high performance of joins on large fact tables

Reference:

https://docs.databricks.com/data/data-sources/azure/sql-data-warehouse.html

QUESTION NO: 11 - (HOTSPOT)

HOTSPOT

You are designing a solution that uses Azure Cosmos DB to store and serve data.

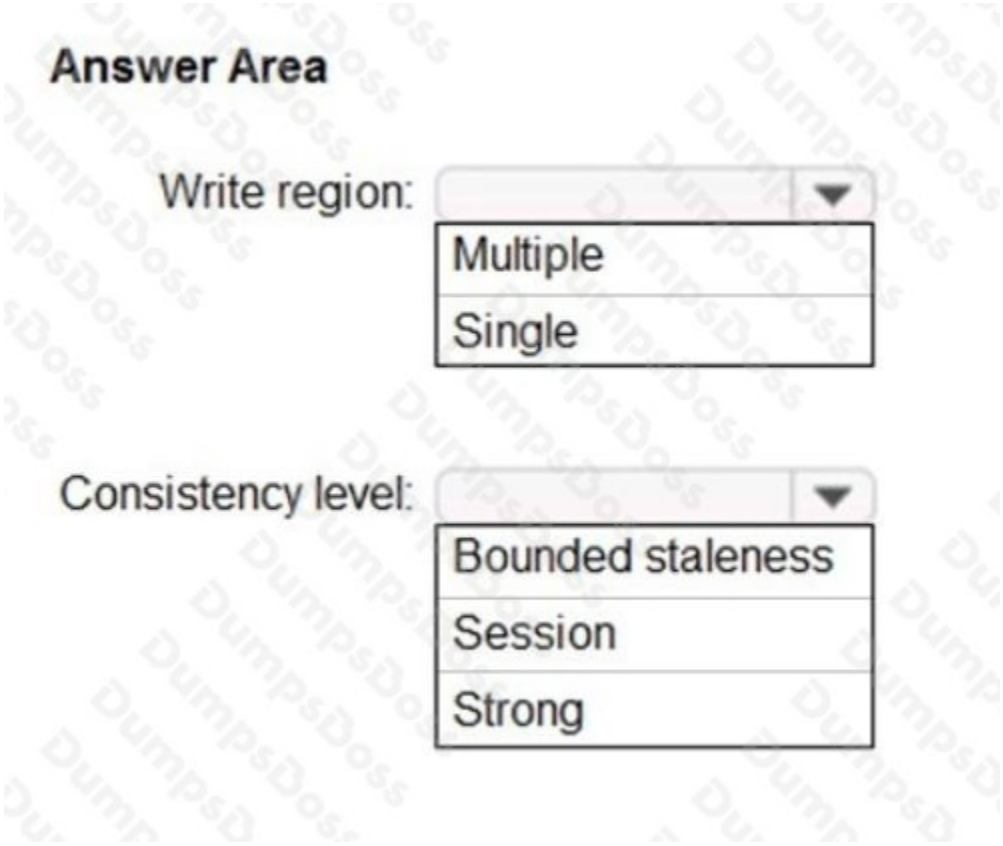
You need to design the Azure Cosmos DB storage to meet the following requirements:

- Provide high availability.
- Provide a recovery point objective (RPO) of less than 15 minutes.
- Provide a recovery time objective (RTO) of less than two minutes. ▪ Minimize data loss in the event of a disaster.

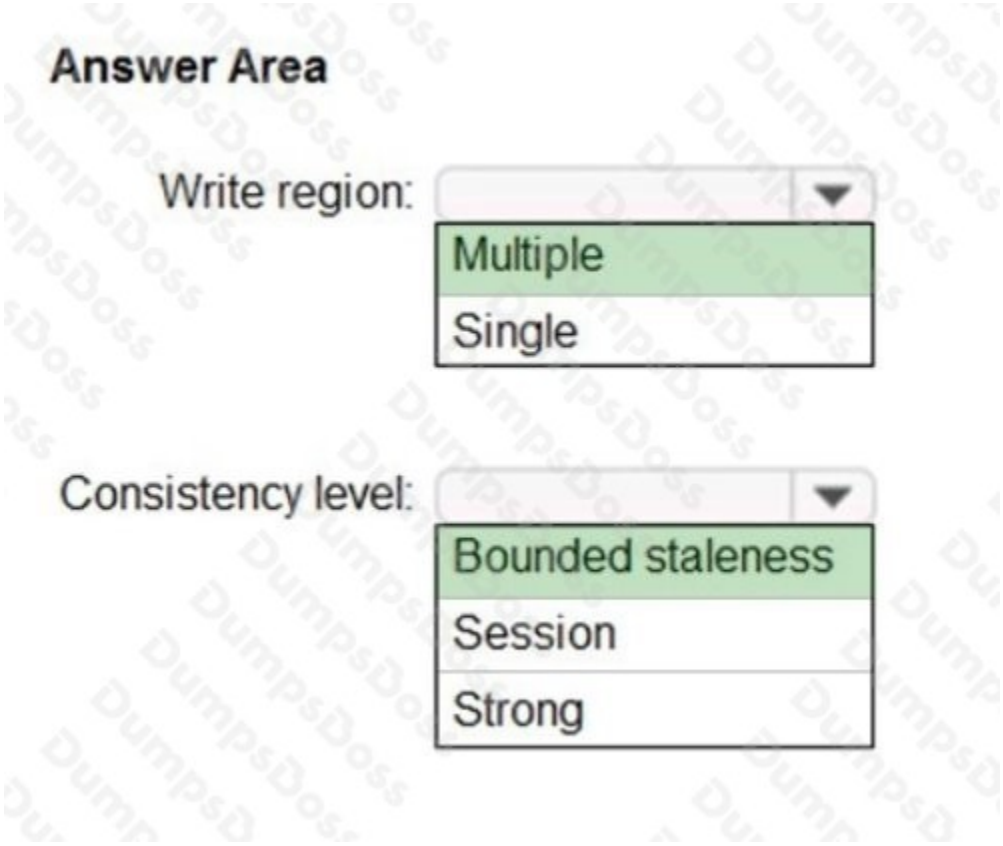
What should you include in the design? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:


Answer Area
Write region:
Multiple
Single
Consistency level:
Bounded staleness
Session
Strong

ANSWER:



Explanation:

Box 1: Multiple

For higher write availability, configure your Azure Cosmos account to have multiple write regions.

Box 2: Bounded staleness

Region(s)	Replication mode	Consistency level	RPO	RTO
1	Single or Multiple write regions	Any Consistency Level	< 240 Minutes	<1 Week
>1	Single write region	Session, Consistent Prefix, Eventual	< 15 minutes	< 15 minutes
>1	Single write region	Bounded Staleness	*K* & *T*	< 15 minutes
>1	Single write region	Strong	0	< 15 minutes
>1	Multiple write regions	Session, Consistent Prefix, Eventual	< 15 minutes	0
>1	Multiple write regions	Bounded Staleness	*K* & *T*	0

K = The number of "*K*" versions (i.e., updates) of an item.

T = The time interval "*T*" since the last update.

Reference: https://docs.microsoft.com/en-us/azure/cosmos-db/high-availability https://docs.microsoft.com/en-us/azure/cosmos-db/consistency-levels#consistency-levels-and-throughput

QUESTION NO: 12

You are designing a storage solution to store CSV files.

You need to grant a data scientist access to read all the files in a single container of an Azure Storage account. The solution must use the principle of least privilege and provide the highest level of security.

What are two possible ways to achieve the goal? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

A. Provide an access key.

B. Assign the Storage Blob Data Reader role at the container level.

C. Assign the Reader role to the storage account.

D. Provide an account shared access signature (SAS).

E. Provide a user delegation shared access signature (SAS).

ANSWER: B E

Explanation:

B: When an Azure role is assigned to an Azure AD security principal, Azure grants access to those resources for that security principal. Access can be scoped to the level of the subscription, the resource group, the storage account, or an individual container or queue.

The built-in Data Reader roles provide read permissions for the data in a container or queue.

Note: Permissions are scoped to the specified resource.

For example, if you assign the Storage Blob Data Reader role to user Mary at the level of a container named sample-container, then Mary is granted read access to all of the blobs in that container.

E: A user delegation SAS is secured with Azure Active Directory (Azure AD) credentials and also by the permissions specified for the SAS. A user delegation SAS applies to Blob storage only.

Reference: https://docs.microsoft.com/en-us/azure/storage/common/storage-auth-aad-rbac-portal
https://docs.microsoft.com/en-us/azure/storage/common/storage-sas-overview

QUESTION NO: 13

You have an Azure subscription that contains an Azure virtual machine and an Azure Storage account. The virtual machine will access the storage account.

You are planning the security design for the storage account.

You need to ensure that only the virtual machine can access the storage account.

Which two actions should you include in the design? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

A. Select Allow trusted Microsoft services to access this storage account.

B. Select Allow read access to storage logging from any network.

C. Enable a virtual network service endpoint.

D. Set the Allow access from setting to Selected networks.

ANSWER: A C

Explanation:

C: Virtual Network (VNet) service endpoint provides secure and direct connectivity to Azure services over an optimized route over the Azure backbone network. Endpoints allow you to secure your critical Azure service resources to only your virtual networks. Service Endpoints enables private IP addresses in the VNet to reach the endpoint of an Azure service without needing a public IP address on the VNet.

A: You must have Allow trusted Microsoft services to access this storage account turned on under the Azure Storage account Firewalls and Virtual networks settings menu.

Incorrect Answers:

D: Virtual Network (VNet) service endpoint policies allow you to filter egress virtual network traffic to Azure Storage accounts over service endpoint, and allow data exfiltration to only specific Azure Storage accounts. Endpoint policies provide granular access control for virtual network traffic to Azure Storage when connecting over service endpoint.

Reference:

https://docs.microsoft.com/en-us/azure/virtual-network/virtual-network-service-endpoints-overview

QUESTION NO: 14

You are designing a big data storage solution. The solution must meet the following requirements:

- Provide unlimited account sizes.
- Support a hierarchical file system.
- Be optimized for parallel analytics workloads.

Which storage solution should you use?

A. Azure Data Lake Storage Gen2

B. Azure Blob storage

C. Apache HBase in Azure HDInsight

D. Azure Cosmos DB

ANSWER: A

Explanation:

Azure Data Lake Storage is optimized performance for parallel analytics workloads

A key mechanism that allows Azure Data Lake Storage Gen2 to provide file system performance at object storage scale and prices is the addition of a hierarchical namespace. This allows the collection of objects/files within an account to be organized into a hierarchy of directories and nested subdirectories in the same way that the file system on your computer is organized.

Reference: https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-namespace

QUESTION NO: 15

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

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

You are designing an HDInsight/Hadoop cluster solution that uses Azure Data Lake Gen1 Storage.

The solution requires POSIX permissions and enables diagnostics logging for auditing.

You need to recommend solutions that optimize storage.

Proposed Solution: Ensure that files stored are smaller than 250MB.

Does the solution meet the goal?

A. Yes

B. No

ANSWER: B

Explanation:

Ensure that files stored are larger, not smaller than 250MB.

You can have a separate compaction job that combines these files into larger ones.

Note: The file POSIX permissions and auditing in Data Lake Storage Gen1 comes with an overhead that becomes apparent when working with numerous small files. As a best practice, you must batch your data into larger files versus writing thousands or millions of small files to Data Lake Storage Gen1. Avoiding small file sizes can have multiple benefits, such as:

- Lowering the authentication checks across multiple files
- Reduced open file connections
- Faster copying/replication
- Fewer files to process when updating Data Lake Storage Gen1 POSIX permissions

Reference: https://docs.microsoft.com/en-us/azure/data-lake-store/data-lake-store-best-practices