**Braindump2go Guarantee All Exams 100% Pass
One Time!**

➢ **Vendor: Microsoft**

➢ **Exam Code: DP-201**

➢ **Exam Name: Designing an Azure Data Solution**

➢ **New Updated Questions from Braindump2go (Updated in March/2021)**

Visit Braindump2go and Download Full Version DP-201 Exam Dumps

QUESTION 1
You are designing an Azure Cosmos DB database that will contain news articles.
The articles will have the following properties: `Category, Created Datetime, Publish Datetime, Author, Headline, Body Text,` and `Publish Status`. Multiple articles will be published in each category daily, but no two stories in a category will be published simultaneously.
Headlines may be updated over time. `Publish Status` will have the following values: `draft, published, updated,` and `removed`. Most articles will remain in the `published` or `updated` status. `Publish Datetime` will be populated only when `Publish Status` is set to `published`.
You will serve the latest articles to websites for users to consume.
You need to recommend a partition key for the database container. The solution must ensure that the articles are served to the websites as quickly as possible.
Which partition key should you recommend?

A. Publish Status
B. Category + Created Datetime
C. Headline
D. Publish Date + random suffix

Answer: B
Explanation:
You can form a partition key by concatenating multiple property values into a single artificial partitionKey property.
These keys are referred to as synthetic keys.
Incorrect Answers:
D: Publish Datetime will be populated only when Publish Status is set to published.
Reference:
https://docs.microsoft.com/en-us/azure/cosmos-db/synthetic-partition-keys

QUESTION 2
You are designing a product catalog for a customer. The product data will be stored in Azure Cosmos DB. The product properties will be different for each product and additional properties will be added to products as needed.
Which Cosmos DB API should you use to provision the database?

A. Cassandra API
B. Core (SQL) API
C. Gremlin API

Answer: A
Explanation:
Cassandrsa is a type of NoSQL database.
NoSQL database (sometimes called as Not Only SQL) is a database that provides a mechanism to store and retrieve

data other than the tabular relations used in relational databases.
Incorrect Answers:
B: Core (SQL) API is a relational database which does not fit this scenario.
C: Gremlin is the graph traversal language of Apache TinkerPop. Gremlin is a functional, data-flow language that enables users to succinctly express complex traversals on (or queries of) their application's property graph.
Reference:
https://www.tutorialspoint.com/cassandra/cassandra_introduction.htm

QUESTION 3
You work for a finance company.
You need to design a business network analysis solution that meets the following requirements:
- Analyzes the flow of transactions between the Azure environments of the company's various partner organizations
- Supports Gremlin (graph) queries
What should you include in the solution?

A. Azure Cosmos DB
B. Azure Synapse
C. Azure Analysis Services
D. Azure Data Lake Storage Gen2

Answer: A
Explanation:
Gremlin is one of the most popular query languages for exploring and analyzing data modeled as property graphs. There are many graph-database vendors out there that support Gremlin as their query language, in particular Azure Cosmos DB which is one of the world's first self-managed, geo-distributed, multi-master capable graph databases.
Azure Synapse Link for Azure Cosmos DB is a cloud native hybrid transactional and analytical processing (HTAP) capability that enables you to run near real-time analytics over operational data. Synapse Link creates a tight seamless integration between Azure Cosmos DB and Azure Synapse Analytics.
Reference:
https://jayanta-mondal.medium.com/analyzing-and-improving-the-performance-azure-cosmos-db-gremlin-queries-7f68bbbac2c
https://docs.microsoft.com/en-us/azure/cosmos-db/synapse-link-use-cases

QUESTION 4
You are designing a streaming solution that must meet the following requirements:
- Accept input data from an Azure IoT hub.
- Write aggregated data to Azure Cosmos DB.
- Calculate minimum, maximum, and average sensor readings every five minutes.
- Define calculations by using a SQL query.
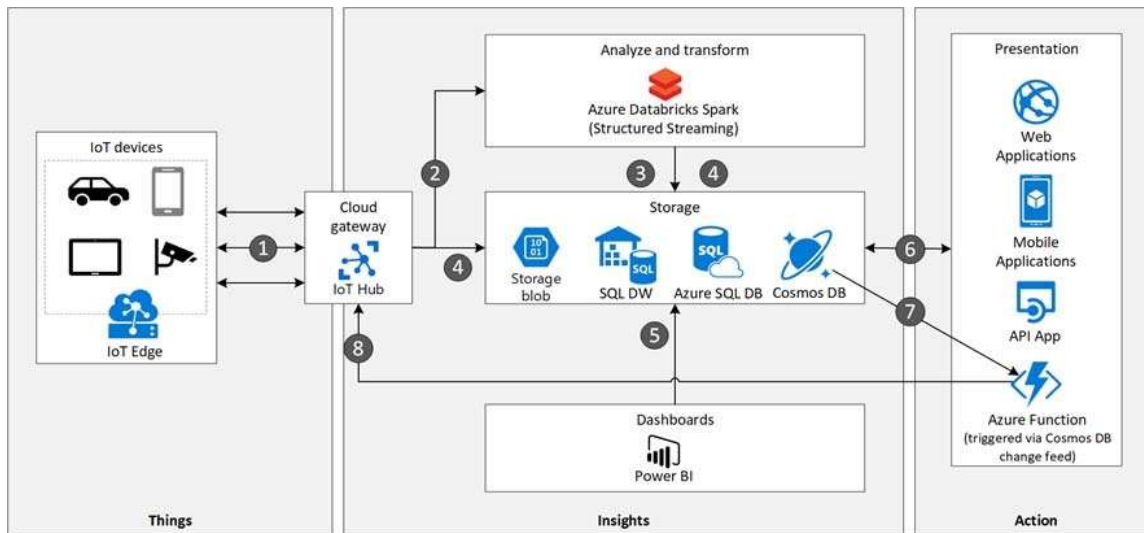- Deploy to multiple environments by using Azure Resource Manager templates.
What should you include in the solution?

A. Azure Functions
B. Azure HDInsight with Spark Streaming
C. Azure Databricks
D. Azure Stream Analytics

Answer: C
Explanation:
Cosmos DB is ideally suited for IoT solutions. Cosmos DB can ingest device telemetry data at high rates.
Architecture

Data flow
1. Events generated from IoT devices are sent to the analyze and transform layer through Azure IoT Hub as a stream of messages. Azure IoT Hub stores streams of data in partitions for a configurable amount of time.
2. Azure Databricks, running Apache Spark Streaming, picks up the messages in real time from IoT Hub, processes the data based on the business logic and sends the data to Serving layer for storage. Spark Streaming can provide real time analytics such as calculating moving averages, min and max values over time periods.
3. Device messages are stored in Cosmos DB as JSON documents.
Reference:
https://docs.microsoft.com/en-us/azure/architecture/solution-ideas/articles/iot-using-cosmos-db

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 plan to store delimited text files in an Azure Data Lake Storage account that will be organized into department folders.
You need to configure data access so that users see only the files in their respective department folder.
Solution: From the storage account, you enable a hierarchical namespace, and you use access control lists (ACLs). Does this meet the goal?

A. Yes
B. No

Answer: B
Explanation:
Azure Data Lake Storage implements an access control model that derives from HDFS, which in turn derives from the POSIX access control model.
Blob container ACLs does not support the hierarchical namespace, so it must be disabled.
Reference:
https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-known-issues

QUESTION 6
You are planning a streaming data solution that will use Azure Databricks. The solution will stream sales transaction data from an online store. The solution has the following specifications:
- The output data will contain items purchased, quantity, line total sales amount, and
line total tax amount.
- Line total sales amount and line total tax amount will be aggregated in Databricks.
- Sales transactions will never be updated. Instead, new rows will be added to adjust a

sale.
You need to recommend an output mode for the dataset that will be processed by using Structured Streaming. The solution must minimize duplicate data.
What should you recommend?

- A. Append
- B. Complete
- C. Update

Answer: A
Explanation:
Append Mode: Only new rows appended in the result table since the last trigger are written to external storage. This is applicable only for the queries where existing rows in the Result Table are not expected to change.
Incorrect Answers:
B: Complete Mode: The entire updated result table is written to external storage. It is up to the storage
8E6BA412E7DB6A14A62CC68E5EB6DAE3
connector to decide how to handle the writing of the entire table.
C: Update Mode: Only the rows that were updated in the result table since the last trigger are written to external storage. This is different from Complete Mode in that Update Mode outputs only the rows that have changed since the last trigger. If the query doesn't contain aggregations, it is equivalent to Append mode.
Reference:
https://docs.microsoft.com/en-us/azure/databricks/getting-started/spark/streaming

QUESTION 7
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: AC
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 8
You are designing an app that will provide a data cleaning and supplementing service for customers. The app will use Azure Data Factory to run a daily process to read and write data from Azure Storage blob containers.
You need to recommend an access mechanism for the customers to grant the app access to their data.
The solution must meet the following requirements:
- Provide access for a period of three months.

- Restrict the app's access to specific containers.
- Minimize administrative effort.
- Minimize changes to the existing access controls of the customer's Azure Storage accounts.
What should you recommend?

A. a shared key
B. anonymous public read access
C. a managed identity
D. a shared access signature (SAS)

Answer: D
Explanation:
A shared access signature (SAS) provides secure delegated access to resources in your storage account.
With a SAS, you have granular control over how a client can access your data.
For example:
- What resources the client may access.
- What permissions they have to those resources.
- How long the SAS is valid.
Reference:
https://docs.microsoft.com/en-us/azure/storage/common/storage-sas-overview

QUESTION 9
You need to implement an Azure Storage account that will use a Blob service endpoint that uses zone-redundant storage (ZRS).
The storage account must only accept connections from a virtual network over Azure Private Link.
What should you include in the implementation?

A. a private endpoint for Azure Blob storage
B. a customer-managed key
C. a shared access signature (SAS)
D. a firewall rule to allow traffic from the virtual network

Answer: A
Explanation:
You can use private endpoints for your Azure Storage accounts to allow clients on a virtual network (VNet) to securely access data over a Private Link.
When creating the private endpoint, you must specify the storage account and the storage service to which it connects.
You need a separate private endpoint for each storage service in a storage account that you need to access, namely Blobs, Data Lake Storage Gen2, Files, Queues, Tables, or Static Websites.
Note: The private endpoint uses an IP address from the VNet address space for your storage account service. Network traffic between the clients on the VNet and the storage account traverses over the VNet and a private link on the Microsoft backbone network, eliminating exposure from the public internet.
Reference:
https://docs.microsoft.com/en-us/azure/storage/common/storage-private-endpoints

QUESTION 10
You are designing a highly available Azure Data Lake Storage solution that will include geo-zone- redundant storage (GZRS).
You need to monitor for replication delays that can affect the recovery point objective (RPO).
What should you include in the monitoring solution?

A. availability
B. Average Success E2E Latency
C. 5xx: Server Error errors
D. Last Sync Time

Answer: D
Explanation:
It's important to note that account failover often results in some data loss, because geo-replication always involves latency. The secondary endpoint is typically behind the primary endpoint. So, when you initiate a failover, any data that has not yet been replicated to the secondary region will be lost.
We [Microsoft] recommend that you always check the Last Sync Time property before initiating a failover to evaluate how far the secondary is behind the primary.
Incorrect Answers:
B: Success E2E Latency: The end-to-end latency of successful requests made to a storage service or the specified API operation, in milliseconds. This value includes the required processing time within Azure Storage to read the request, send the response, and receive acknowledgment of the response.
Reference:
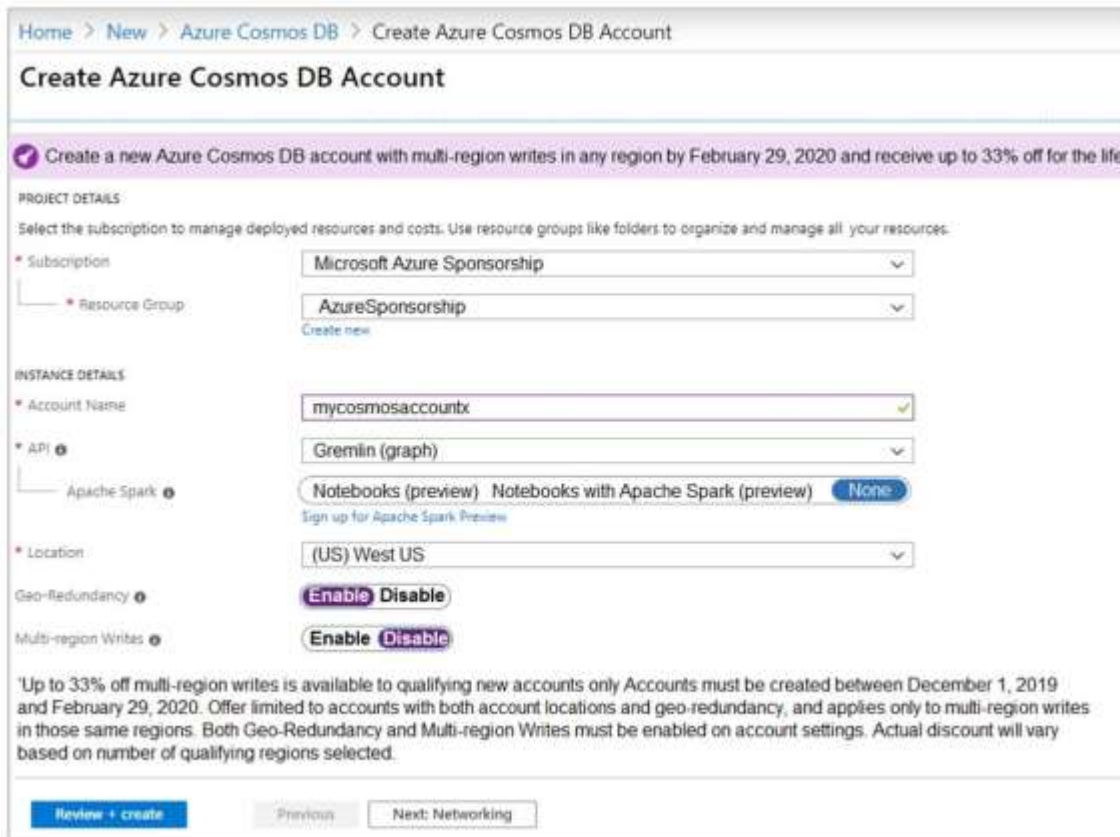https://azure.microsoft.com/en-us/blog/account-failover-now-in-public-preview-for-azure-storage/
https://docs.microsoft.com/en-us/azure/azure-monitor/essentials/metrics-supported

QUESTION 11
Hotspot Question
You are evaluating the use of an Azure Cosmos DB account for a new database.
The proposed account will be configured as shown in the following exhibit.

Use the drop-down menus to select the answer choice that completes each statement based on the information presented in the graphic.
NOTE: Each correct selection is worth one point.

Answer Area

The data in the account will be organized and queried as: ▼

documents
rows and columns
vertices and edges

If no additional Azure regions are added by default, read access to the data in the account will be available in: ▼

US East
US West 2
US East 2

Answer:

Answer Area

The data in the account will be organized and queried as: ▼

documents
rows and columns
vertices and edges

If no additional Azure regions are added by default, read access to the data in the account will be available in: ▼

US East
US West 2
US East 2

Explanation:
Box 1: vertices and edges
Gremlin API is selected.
You can use the Gremlin language to create graph entities (vertices and edges), modify properties within those entities, perform queries and traversals, and delete entities.
Box 2: US East
The (US) West US is selected as the primary location and geo- redundancy is enabled.
The secondary location for West US is East US.
Note: When a storage account is created, the customer chooses the primary location for their storage account.
However, the secondary location for the storage account is fixed and customers do not have the ability to change this.
The following table shows the current primary and secondary location pairings:

Primary	Secondary
North Central US	South Central US
South Central US	North Central US
East US	West US
West US	East US
North Europe	West Europe
West Europe	North Europe
South East Asia	East Asia
East Asia	South East Asia
East China	North China
North China	East China

Reference:
https://docs.microsoft.com/en-us/azure/cosmos-db/gremlin-support
https://technet2.github.io/Wiki/blogs/windowsazurestorage/windows-azure-storage-redundancy-options-and-read-access-geo-redundant-storage.html

QUESTION 12
Hotspot Question
You have an Azure event hub named retailhub that has 16 partitions. Transactions are posted to retailhub. Each transaction includes the transaction ID, the individual line items, and the payment details. The transaction ID is used as the partition key.
You are designing an Azure Stream Analytics job to identify potentially fraudulent transactions at a retail store. The job will use retailhub as the input. The job will output the transaction ID, the individual line items, the payment details, a fraud score, and a fraud indicator.
You plan to send the output to an Azure event hub named fraudhub.
You need to ensure that the fraud detection solution is highly scalable and processes transactions as quickly as possible.
How should you structure the output of the Stream Analytics job? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Answer Area

Number of partitions:

1
8
16
32

Partition key:

Fraud indicator
Fraud score
Individual line items
Payment details
Transaction ID

Answer:

Answer Area

Number of partitions:

1
8
16
32

Partition key:

Fraud indicator
Fraud score
Individual line items
Payment details
Transaction ID

Explanation:
Box 1: 16
For Event Hubs you need to set the partition key explicitly.
An embarrassingly parallel job is the most scalable scenario in Azure Stream Analytics. It connects one partition of the

input to one instance of the query to one partition of the output.
Box 2: Transaction ID
Reference:
https://docs.microsoft.com/en-us/azure/event-hubs/event-hubs-features#partitions

QUESTION 13
Drag and Drop Question
You have a CSV file in Azure Blob storage. The file does NOT have a header row.
You need to use Azure Data Factory to copy the file to an Azure SQL database. The solution must minimize how long it takes to copy the file.
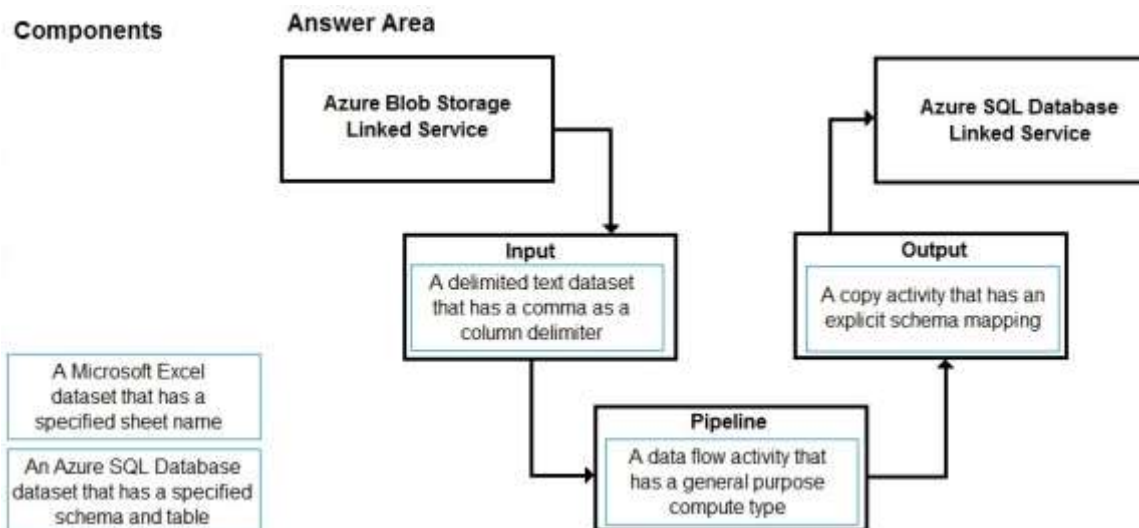How should you configure the copy process? To answer, drag the appropriate components to the correct locations.
Each component may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.
NOTE: Each correct selection is worth one point.
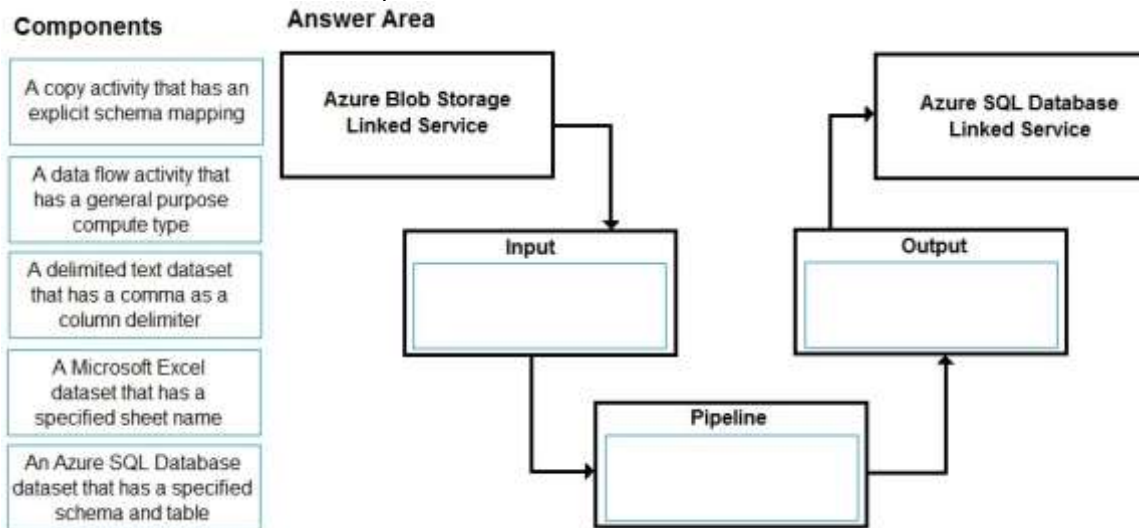
Answer:

Explanation:
Input: A delimited text dataset that has a comma a column delimiter columnDelimiter: The character(s) used to separate columns in a file. The default value is comma ,. When the column delimiter is defined as empty string, which means no delimiter, the whole line is taken as a single column.
Pipeline: A data flow activity that has a general purpose compute type When you're transforming data in mapping data flows, you can read and write files from Azure Blob storage.
Output: A copy activity that has an explicit schema mapping Use Copy Activity in Azure Data Factory to copy data from and to Azure SQL Database, and use Data Flow to transform data in Azure SQL Database.

DP-201 Exam Dumps DP-201 Exam Questions DP-201 PDF Dumps DP-201 VCE Dumps

https://www.braindump2go.com/dp-201.html

Reference:
https://docs.microsoft.com/en-us/azure/data-factory/format-delimited-text
https://docs.microsoft.com/en-us/azure/data-factory/connector-azure-sql-database