QUESTION 61
DRAG DROP
You manage a financial computation data analysis process. Microsoft Azure virtual machines (VMs) run the process in daily jobs, and store the results in virtual hard drives (VHDs.)

The VMs product results using data from the previous day and store the results in a snapshot of the VHD. When a new month begins, a process creates a new VHD.

You must implement the following data retention requirements:

- Daily results must be kept for 90 days.
- Data for the current year must be available for weekly reports.
- Data from the previous 10 years must be stored for auditing purposes.
- Data required for an audit must be produced within 10 days of a request.

You need to enforce the data retention requirements while minimizing cost.

How should you configure the lifecycle policy? To answer, drag the appropriate JSON segments to the correct locations. Each JSON segment may be used once, more than once, or not at all. You may need to drag the split bat between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

Correct Answer:

QUESTION 62
You develop data engineering solutions for a company.

You need to ingest and visualize real-time Twitter data by using Microsoft Azure.

Which three technologies should you use? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

A. Event Grid topic
B. Azure Stream Analytics Job that queries Twitter data from an Event Hub
C. Azure Stream Analytics Job that queries Twitter data from an Event Grid
D. Logic App that sends Twitter posts which have target keywords to Azure
E. Event Grid subscription
F. Event Hub instance

Correct Answer: BDF
Explanation:
You can use Azure Logic apps to send tweets to an event hub and then use a Stream Analytics job to read from event hub and send them to PowerBI.

References:
https://community.powerbi.com/t5/Integrations-with-Files-and/Twitter-streaming-analytics-step-by-

step/td-p/9594

QUESTION 63
You need to develop a pipeline for processing data. The pipeline must meet the following requirements:

- Scale up and down resources for cost reduction.
- Use an in-memory data processing engine to speed up ETL and machine learning operations.
- Use streaming capabilities.
- Provide the ability to code in SQL, Python, Scala, and R.
- Integrate workspace collaboration with Git.

What should you use?

A. HDInsight Spark Cluster
B. Azure Stream Analytics
C. HDInsight Hadoop Cluster
D. Azure SQL Data Warehouse

Correct Answer: A

QUESTION 64
An application will use Microsoft Azure Cosmos DB as its data solution. The application will use the Cassandra API to support a column-based database type that uses containers to store items.

You need to provision Azure Cosmos DB. Which container name and item name should you use? Each correct answer presents part of the solutions.

NOTE: Each correct answer selection is worth one point.

A. table
B. collection
C. graph
D. entities
E. rows

Correct Answer: AE
Explanation:
Depending on the choice of the API, an Azure Cosmos item can represent either a document in a collection, a row in a table or a node/edge in a graph. The following table shows the mapping between API-specific entities to an Azure Cosmos item:

Cosmos entity	SQL API	Cassandra API	Azure Cosmos DB's API for MongoDB	Gremlin API	Table API
Azure Cosmos item	Document	Row	Document	Node or Edge	Item

An Azure Cosmos container is specialized into API-specific entities as follows:

Azure Cosmos entity	SQL API	Cassandra API	Azure Cosmos DB's API for MongoDB	Gremlin API	Table API
Azure Cosmos container	Collection	Table	Collection	Graph	Table

References:
https://docs.microsoft.com/en-us/azure/cosmos-db/databases-containers-items

QUESTION 65
You implement 3 Azure SQL Data Warehouse instance.

You plan to migrate the largest fact table to Azure SQL Data Warehouse The table resides on Microsoft SQL Server on-premises and e 10 terabytes (TB) in size.

Incoming queues use the primary key Sale Key column to retrieve data as displayed in the following table:

SaleKey	CityKey	CustomerKey	StockItemKey	InvoiceDateKey	Quantity	UnitPrice	TotalExcludingTax
49309	90858	70	69	10/22/13	8	16	128
49313	55710	126	69	10/22/13	2	16	32
49343	44710	234	68	10/22/13	10	16	160
49352	66109	163	70	10/22/13	4	16	64
49448	65312	230	70	10/22/13	8	16	128
49646	85877	271	70	10/24/13	1	16	16
49798	41238	288	69	10/24/13	1	16	16

You need to distribute the fact table across multiple nodes to optimize performance of the table.

Which technology should you use?

A. hash distributed table with clustered ColumnStore index
B. hash distributed table with clustered index
C. heap table with distribution replicate
D. round robin distributed table with clustered index
E. round robin distributed table with clustered ColumnStore index

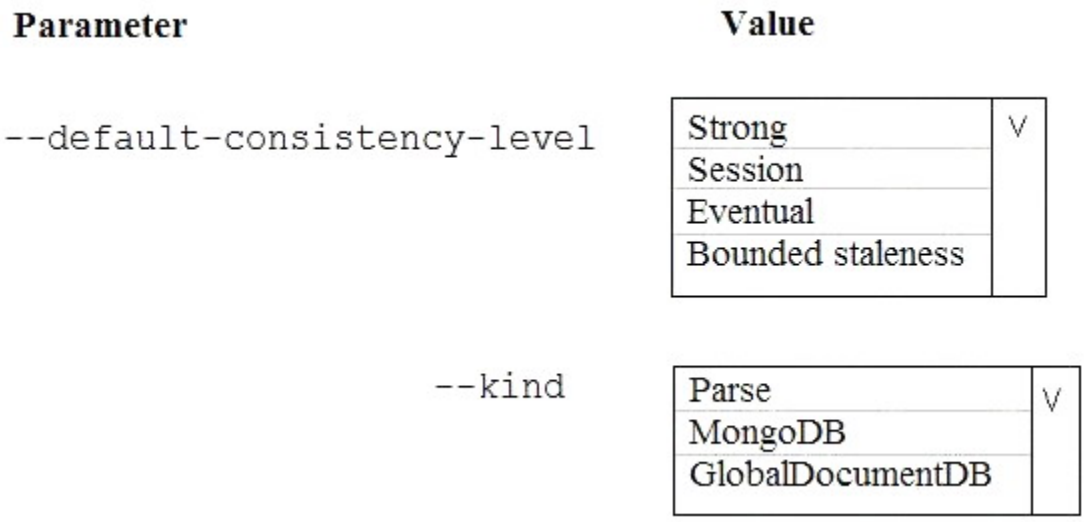
Correct Answer: A

QUESTION 66
HOTSPOT
A company is planning to use Microsoft Azure Cosmos DB as the data store for an application. You have the following Azure CLI command:

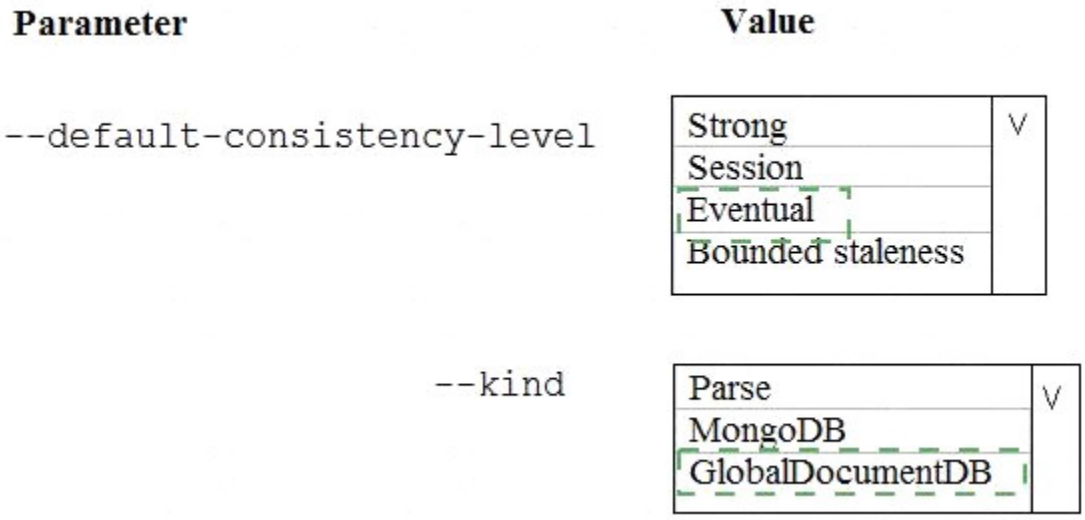
az cosmosdb create -name "cosmosdbdev1" -resource-group "rgdev"

You need to minimize latency and expose the SQL API. How should you complete the command? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.



Correct Answer:



QUESTION 67
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution. Determine whether the solution meets the stated goals.

You develop a data ingestion process that will import data to a Microsoft Azure SQL Data Warehouse.

The data to be ingested resides in parquet files stored in an Azure Data Lake Gen 2 storage account.

You need to toad the data from the Azure Data Lake Gen 2 storage account into the Azure SQL Data Warehouse.