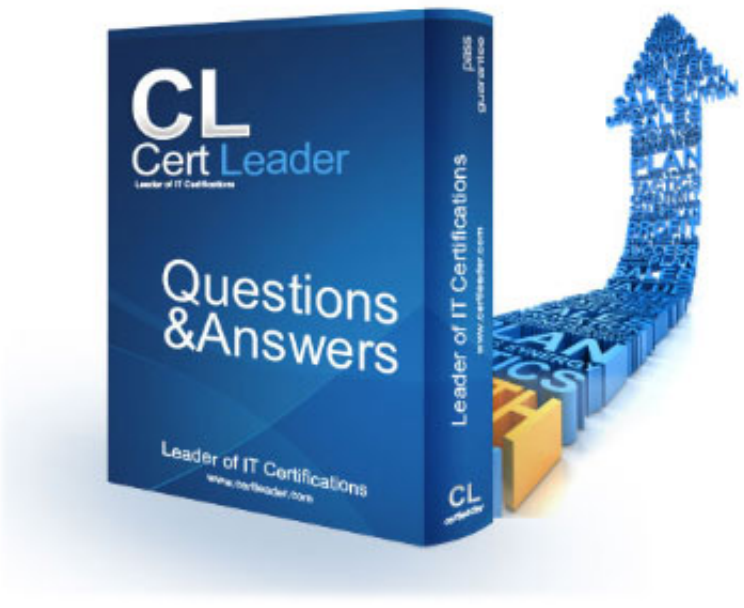
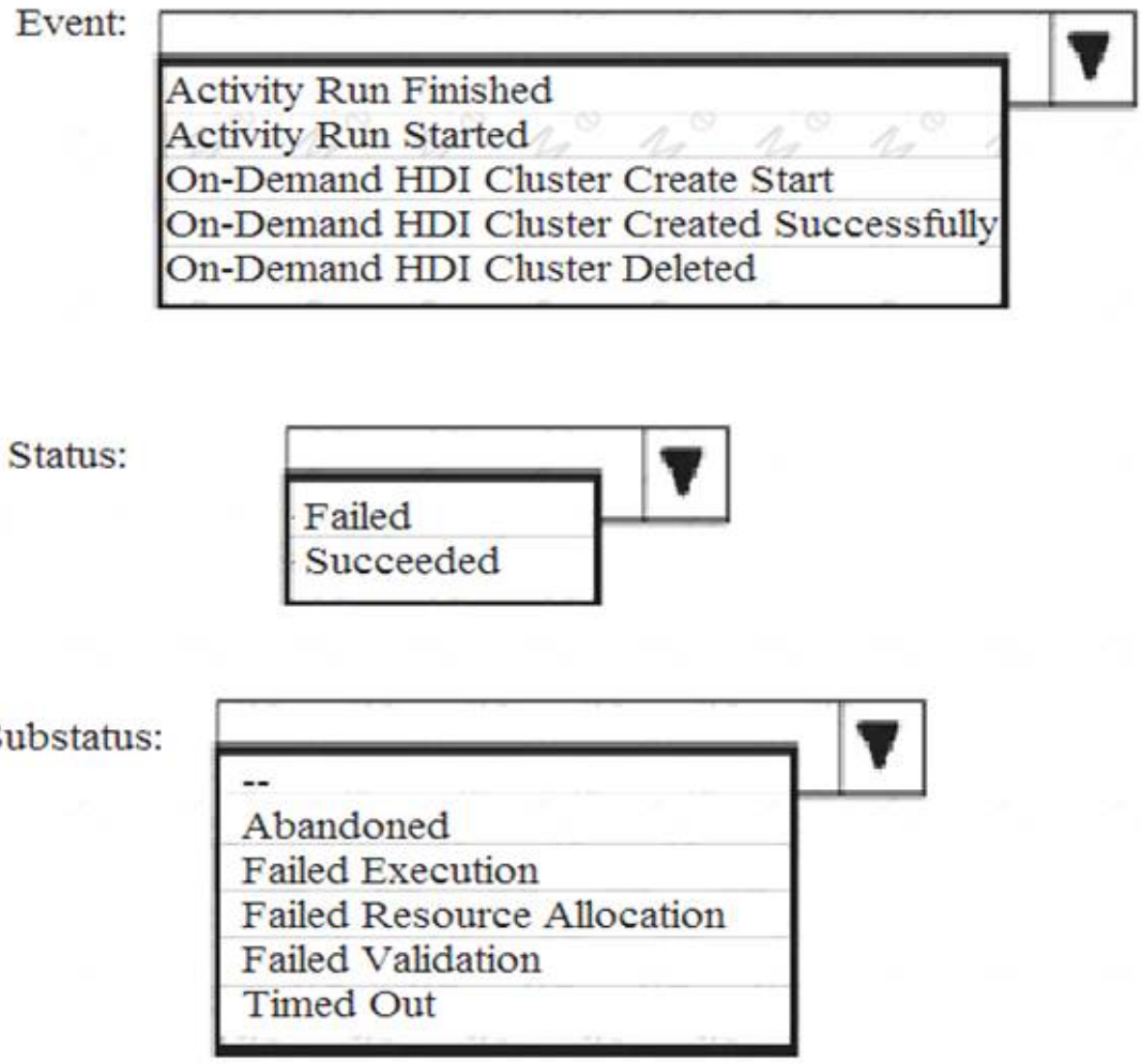
70-475 Dumps

Designing and Implementing Big Data Analytics Solutions

https://www.certleader.com/70-475-dumps.html



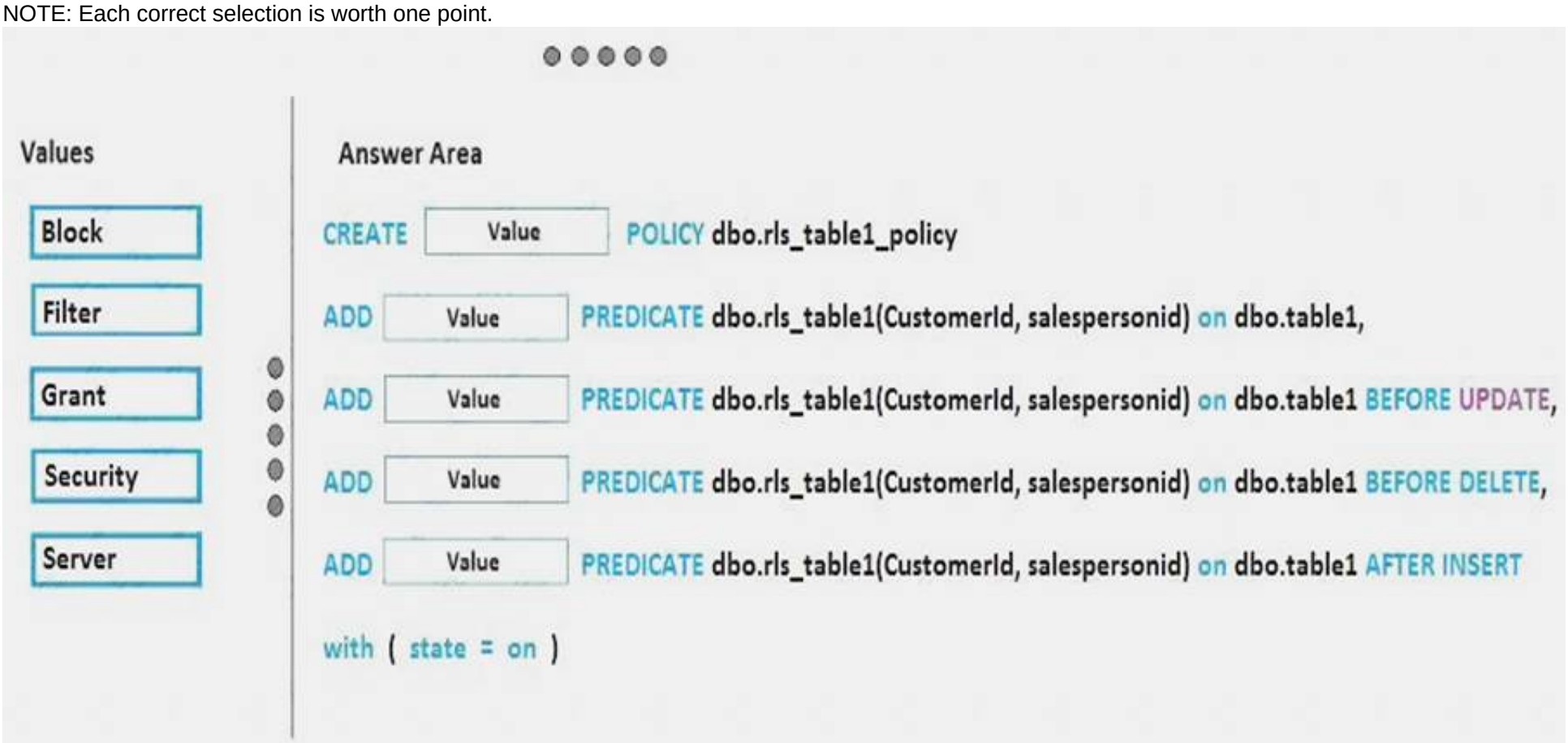
NEW QUESTION 1
You need to configure the alert to meet the requirements for ETL.
Which settings should you use for the alert? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.



Answer:

Explanation: Scenario: Relecloud identifies the following requirements for extract, transformation, and load (ETL): An email alert must be generated when a failure of any type occurs during ETL processing.

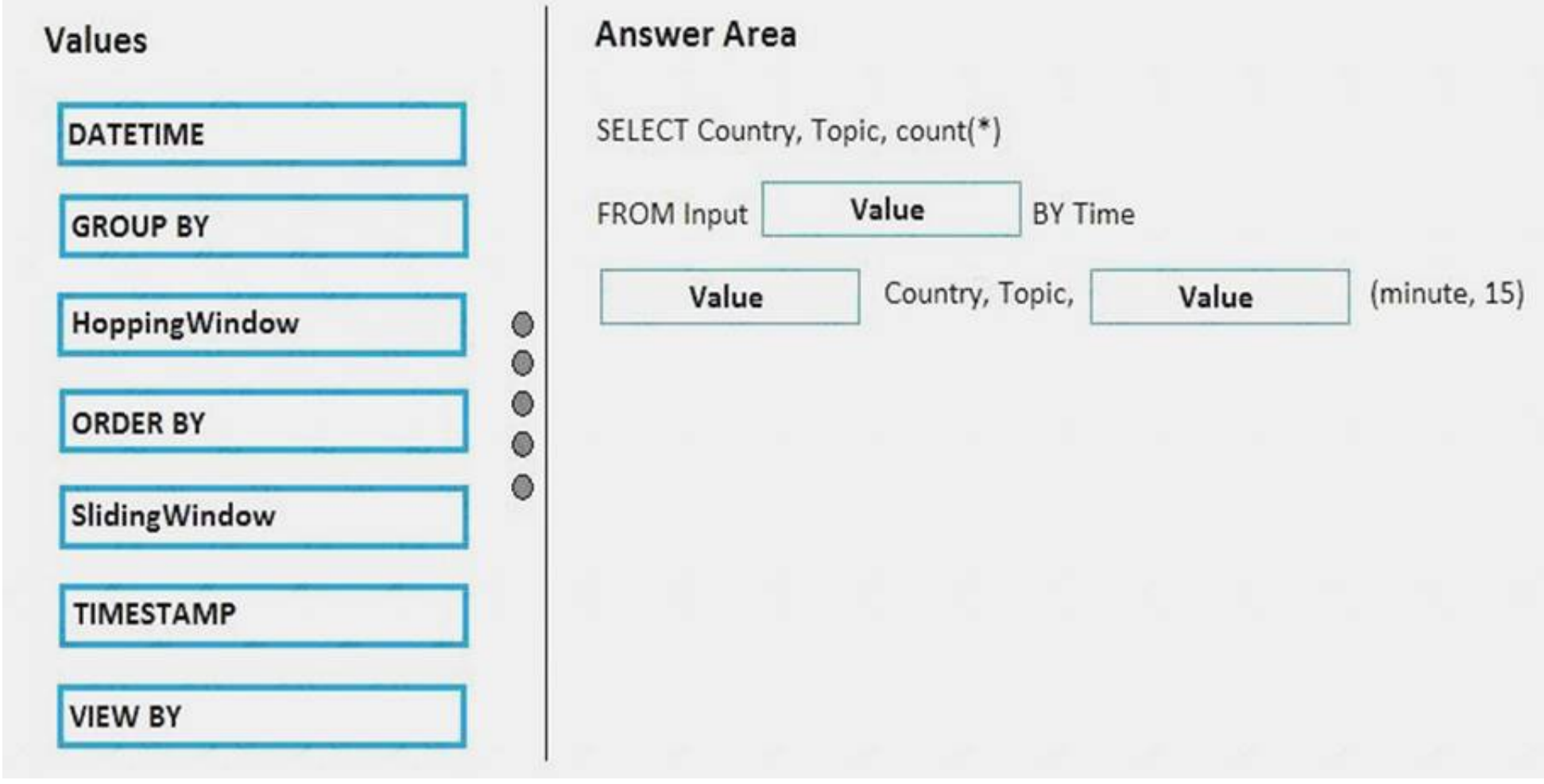
NEW QUESTION 2
You need to implement rls_table1.
Which code should you execute? To answer, drag the appropriate values to the correct targets. Each value may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view content.
NOTE: Each correct selection is worth one point.



Answer:

Explanation: Box 1: Security Security Policy
Example: After we have created Predicate function, we have to bind it to the table, using Security Policy. We will be using CREATE SECURITY POLICY command to set the security policy in place.
CREATE SECURITY POLICY DepartmentSecurityPolicy
ADD FILTER PREDICATE dbo.DepartmentPredicateFunction(UserDepartment) ON dbo.Department WITH(STATE = ON)
Box 2: Filter
[FILTER | BLOCK]
The type of security predicate for the function being bound to the target table. FILTER predicates silently filter the rows that are available to read operations.
BLOCK predicates explicitly block write operations that violate the predicate function.
Box 3: Block
Box 4: Block
Box 5: Filter

NEW QUESTION 3
You need to create a query that identifies the trending topics.
How should you complete the query? To answer, drag the appropriate values to the correct targets. Each value 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: From scenario: Topics are considered to be trending if they generate many mentions in a specific country during a 15-minute time frame.
Box 1: TimeStamp
Azure Stream Analytics (ASA) is a cloud service that enables real-time processing over streams of data flowing in from devices, sensors, websites and other live systems. The stream-processing logic in ASA is expressed in a SQL-like query language with some added extensions such as windowing for performing temporal calculations.
ASA is a temporal system, so every event that flows through it has a timestamp. A timestamp is assigned automatically based on the event's arrival time to the input source but you can also access a timestamp in your event payload explicitly using TIMESTAMP BY:
SELECT * FROM SensorReadings TIMESTAMP BY time Box 2: GROUP BY
Example: Generate an output event if the temperature is above 75 for a total of 5 seconds SELECT sensorId, MIN(temp) as temp
FROM SensorReadings TIMESTAMP BY time
GROUP BY sensorId, SlidingWindow(second, 5) HAVING MIN(temp) > 75
Box 3: SlidingWindow
Windowing is a core requirement for stream processing applications to perform set-based operations like counts or aggregations over events that arrive within a specified period of time. ASA supports three types of windows: Tumbling, Hopping, and Sliding.
With a Sliding Window, the system is asked to logically consider all possible windows of a given length and output events for cases when the content of the window actually changes – that is, when an event entered or existed the window.

NEW QUESTION 4
Which technology should you recommend to meet the technical requirement for analyzing the social media data?

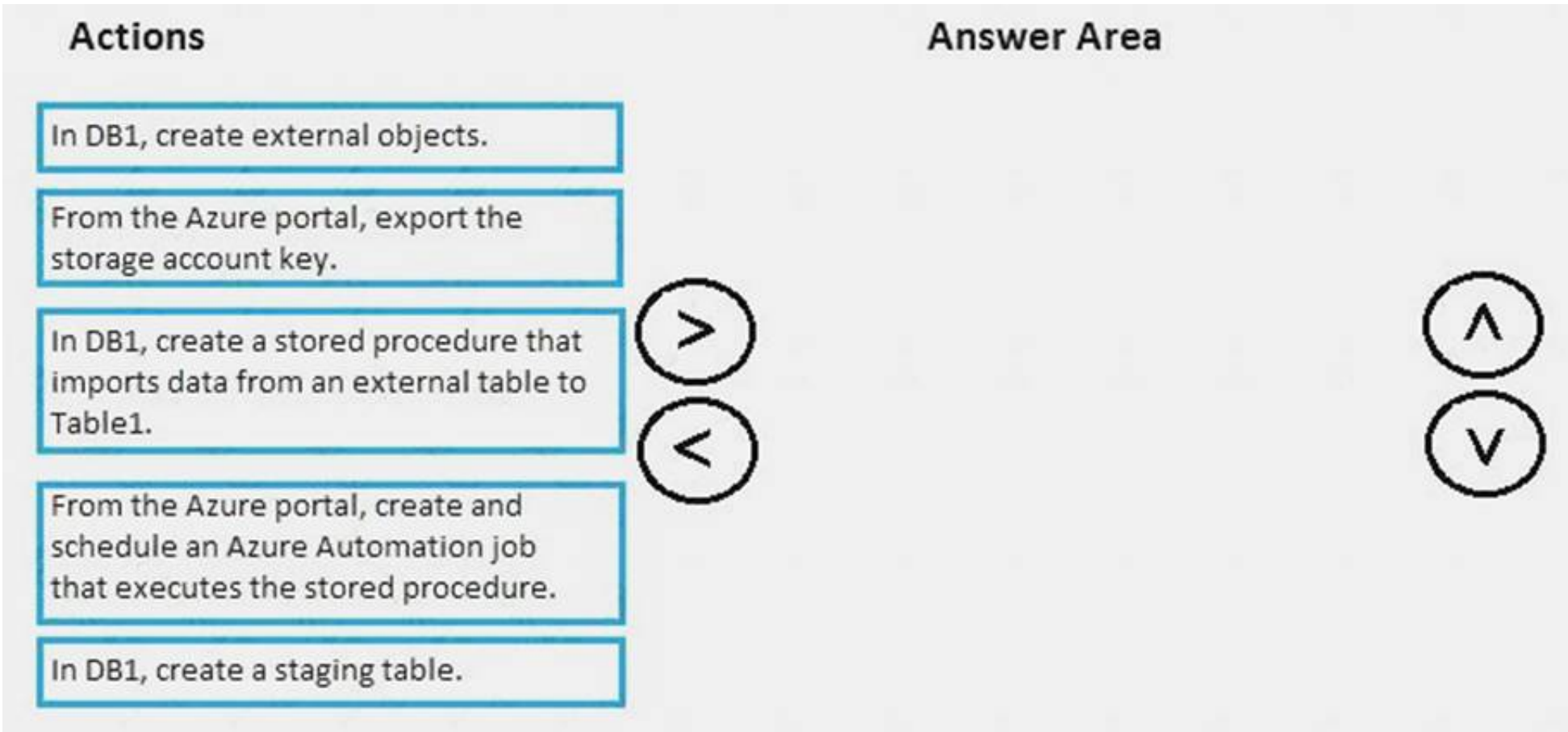
A. Azure Stream Analytics
B. Azure Data Lake Analytics
C. Azure Machine Learning

D. Azure HDInsight Storm clusters

Answer: A

Explanation: Azure Stream Analytics is a fully managed event-processing engine that lets you set up real-time analytic computations on streaming data.
Scalability
Stream Analytics can handle up to 1 GB of incoming data per second. Integration with Azure Event Hubs and Azure IoT Hub allows jobs to ingest millions of events per second coming from connected devices, clickstreams, and log files, to name a few. Using the partition feature of event hubs, you can partition computations into logical steps, each with the ability to be further partitioned to increase scalability.

NEW QUESTION 5
You need to implement a solution that meets the data refresh requirement for DB1.
Which three actions should you perform in sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.



Answer:

Explanation: Azure Data Factory can be used to orchestrate the execution of stored procedures. This allows more complex pipelines to be created and extends Azure Data Factory's ability to leverage the computational power of SQL Data Warehouse.
From scenario:
Relecloud has a Microsoft SQL Server database named DB1 that stores information about the advertisers. DB1 is hosted on a Microsoft Azure virtual machine.
Relecloud identifies the following requirements for DB1:

- Data generated by the streaming analytics platform must be stored in DB1.
- The advertisers in DB1 must be stored in a table named Table1 and must be refreshed nightly.

Topic 3, Litware, Inc
Overview
General Overview
Litware, Inc. is a company that manufactures personal devices to track physical activity and other health-related data.
Litware has a health tracking application that sends health-related data horn a user's personal device to Microsoft Azure.
Physical Locations
Litware has three development and commercial offices. The offices are located in the Untied States, Luxembourg, and India.
Litware products are sold worldwide. Litware has commercial representatives in more than 80 countries.
Existing Environment Environment
In addition to using desktop computers in all of the offices. Litware recently started using Microsoft Azure resources and services for both development and operations.
Litware has an Azure Machine Learning Solution.
Litware Health Tracking Application
Litware recently extended its platform to provide third-party companies with the ability to upload data from devices to Azure. The data can be aggregated across multiple devices to provide users with a comprehensive view of their global health activity.
While the upload from each device is small, potentially more than 100 million devices will upload data daily by using an Azure event hub.
Each health activity has a small amount of data, such as activity type, start date/time, and end date/time. Each activity is limited to a total of 3 KB and includes a customer Identification key.
In addition to the Litware health tracking application, the users' activities can be reported to Azure by using an open API.
Machine Learning Experiments
The developers at Litware perform Machine Learning experiments to recommend an appropriate health activity based on the past three activities of a user.
The Litware developers train a model to recommend the best activity for a user based on the hour of the day.
Requirements Planned Changes
Litware plans to extend the existing dashboard features so that health activities can be compared between the users based on age, gender, and geographic region.
Business Goals
Minimize the costs associated with transferring data from the event hub to Azure Storage.

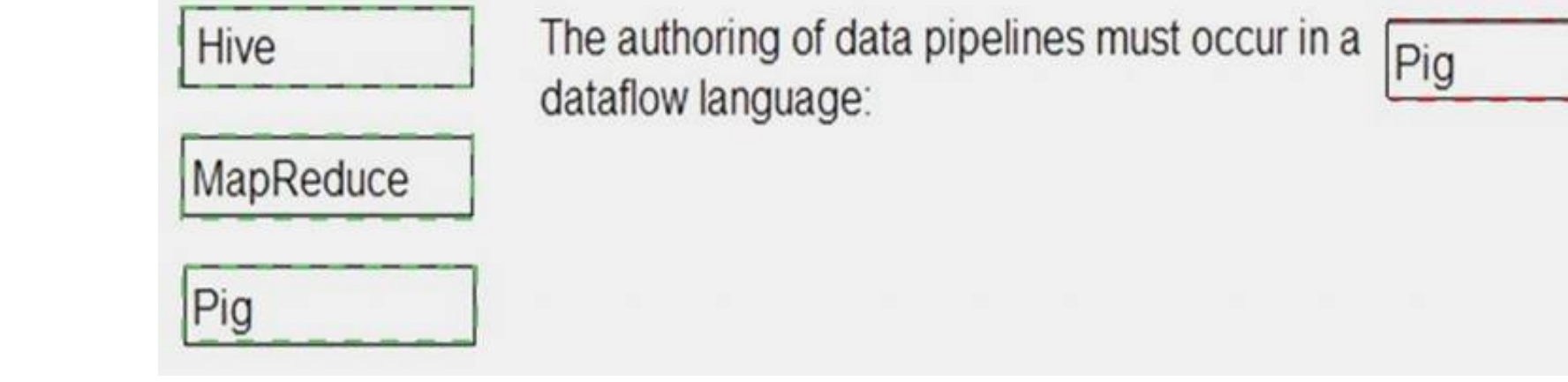
Technical Requirements
Litware identities the following technical requirements:
Data from the devices must be stored from three years in a format that enables the fast processing of data fields and Filtering.
The third-party companies must be able to use the Litware Machine learning models to generate recommendations to their users by using a third-party application.
Any changes to the health tracking application must ensure that the Litware developers can run the experiments without interrupting or degrading the performance of the production environment.
Privacy Requirements
Activity tracking data must be available to all of the Litware developers for experimentation. The developers must be prevented from accessing the private information of the users.
Other Technical Requirements
When the Litware health tracking application asks users how they feel, their responses must be reported to Azure.

Topic 2, Mix Questions

NEW QUESTION 6
You plan to deploy Microsoft Azure HDInsight clusters for business analytics and data pipelines. The clusters must meet the following requirements:

- Business users must use a language that is similar to SQL.
- The authoring of data pipelines must occur in a dataflow language. You need to identify which language must be used for each requirement.

Which languages should you identify? To answer, drag the appropriate languages to the correct requirements. Each language 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.



Answer:

Explanation:

NEW QUESTION 7
You have data pushed to Microsoft Azure Blob storage every few minutes.
You want to use an Azure Machine Learning web service to score the data hourly. You plan to deploy the data factory pipeline by using a Microsoft.NET application. You need to create an output dataset for the web service.
Which three properties should you define? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

A. Source
B. LinkedServiceName
C. TypeProperties
D. Availability
E. External

Answer: ABC

NEW QUESTION 8
You plan to implement a Microsoft Azure Data Factory pipeline. The pipeline will have custom business logic that requires a custom processing step.
You need to implement the custom processing step by using C#.
Which interface and method should you use? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Answer:

Explanation: References:
https://github.com/MicrosoftDocs/azure-docs/blob/master/articles/data-factory/v1/data-factory-use-custom-activ

NEW QUESTION 9
You have an Apache Hadoop system that contains 5 TB of data.
You need to create queries to analyze the data in the system. The solution must ensure that the queries execute as quickly as possible.
Which language should you use to create the queries?

A. Apache Pig
B. Java
C. Apache Hive
D. MapReduce

Answer: D

NEW QUESTION 10
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 deploy a Microsoft Azure SQL data warehouse and a web application.
The data warehouse will ingest 5 TB of data from an on-premises Microsoft SQL Server database daily. The web application will query the data warehouse.
You need to design a solution to ingest data into the data warehouse.
Solution: You use the bcp utility to export CSV files from SQL Server and then to import the files to Azure SQL Data Warehouse.
Does this meet the goal?

A. Yes
B. No

Answer: B

Explanation: If you need the best performance, then use PolyBase to import data into Azure SQL warehouse. References: https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-migrate-data

NEW QUESTION 10
You are designing a solution based on the lambda architecture.
You need to recommend which technology to use for the serving layer. What should you recommend?

A. Apache Storm
B. Kafka
C. Microsoft Azure DocumentDB
D. Apache Hadoop

Answer: C

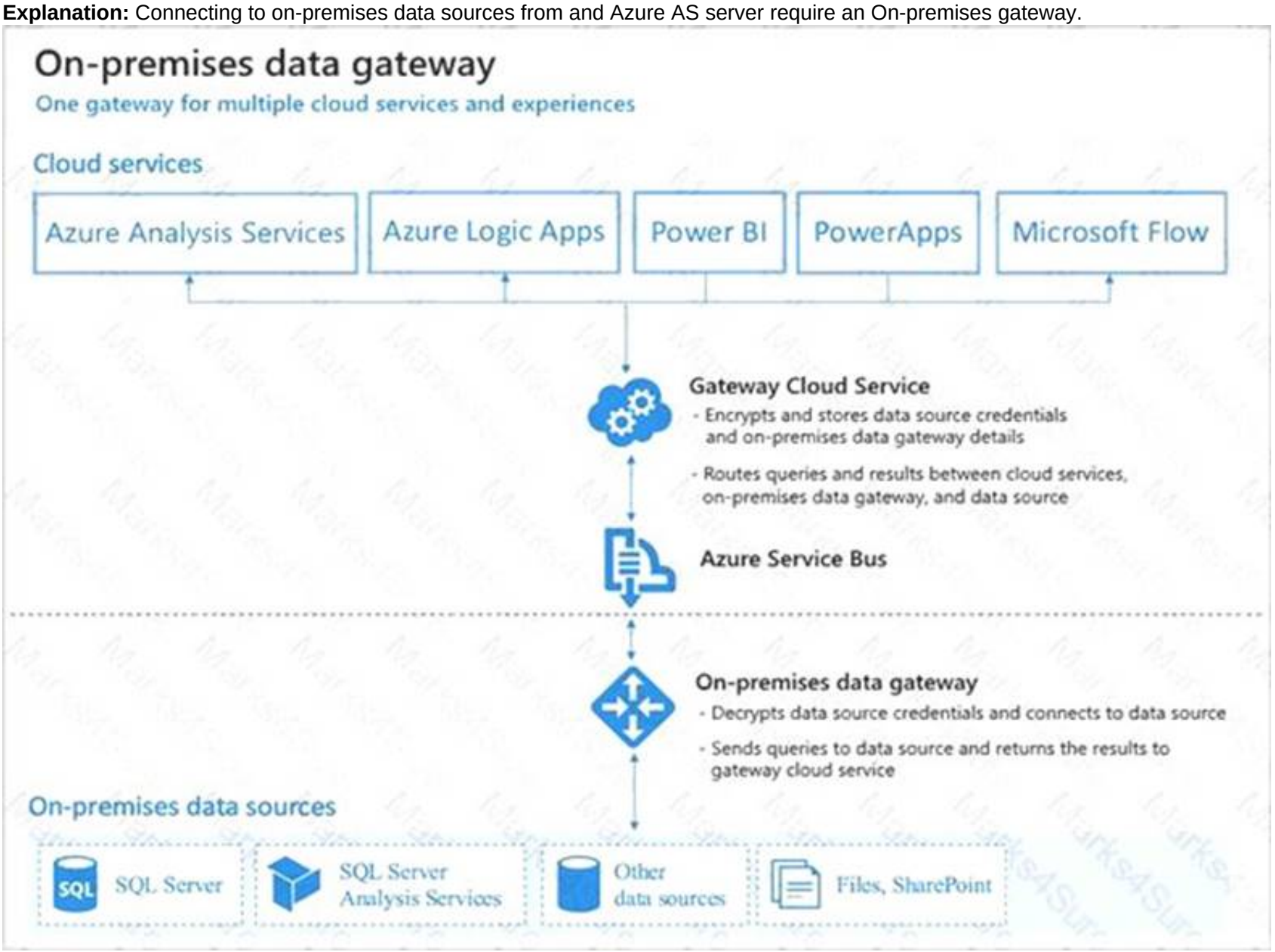
Explanation: The Serving Layer is a bit more complicated in that it needs to be able to answer a single query request against two or more databases, processing platforms, and data storage devices. Apache Druid is an example of a cluster-based tool that can marry the Batch and Speed layers into a single answerable request.

NEW QUESTION 12
You are building an Azure Analysis Services cube.
The source data for the cube is located on premises in a Microsoft SQL Server database. You need to ensure that the Azure Analysis Services service can access the source data. What should you deploy to your Azure subscription?

A. a site-to-site VPN
B. Azure Data Factory
C. a network gateway in Azure
D. a data gateway in Azure

Answer: D

Explanation: Connecting to on-premises data sources from and Azure AS server require an On-premises gateway.



References:
https://azure.microsoft.com/en-in/blog/on-premises-data-gateway-support-for-azure-analysis-services/

NEW QUESTION 16
You have a Microsoft Azure SQL database that contains Personally Identifiable Information (PII).
To mitigate the PII risk, you need to ensure that data is encrypted while the data is at rest. The solution must minimize any changes to front-end applications.
What should you use?

A. Transport Layer Security (TLS)
B. transparent data encryption (TDE)
C. a shared access signature (SAS)
D. the ENCRYPTBYPASSPHRASE T-SQL function

Answer: B

Explanation: Transparent data encryption (TDE) helps protect Azure SQL Database, Azure SQL Managed Instance, and Azure Data Warehouse against the threat of malicious activity. It performs real-time encryption and decryption of the database, associated backups, and transaction log files at rest without requiring changes to the application.
References: https://docs.microsoft.com/en-us/azure/sql-database/transparent-data-encryption-azure-sql

NEW QUESTION 17
You have an application that displays data from a Microsoft Azure SQL database. The database contains credit card numbers.
You need to ensure that the application only displays the last four digits of each credit card number when a credit card number is returned from a query. The solution must NOT require any changes to the data in the database.

What should you use?

A. Dynamic Data Masking
B. cell-level security
C. Transparent Data Encryption (TDE)
D. row-level security

Answer: A

NEW QUESTION 18
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 states goals. Some question sets might have more than one correct solution, while the 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 have an Apache Spark system that contains 5 TB of data.
You need to write queries that analyze the data in the system. The queries must meet the following requirements:

- Use static data typing.
- Execute queries as quickly as possible.
- Have access to the latest language features. Solution: You write the queries by using Scala.

A. Yes
B. No

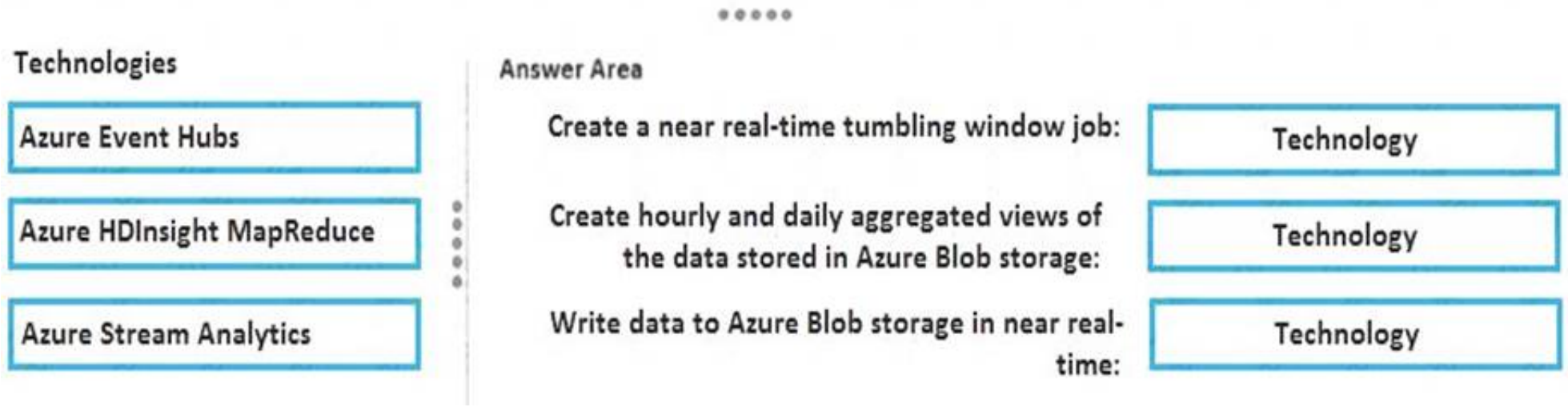
Answer: A

NEW QUESTION 19
You have a Microsoft Azure Data Factory pipeline.
You discover that the pipeline fails to execute because data is missing. You need to rerun the failure in the pipeline.
Which cmdlet should you use?

A. Set-AzureAutomationJob
B. Resume-AzureDataFactoryPipeline
C. Resume-AzureAutomationJob
D. Set-AzureDataFactotySliceStatus

Answer: B

NEW QUESTION 24
You have data generated by sensors. The data is sent to Microsoft Azure Event Hubs.
You need to have an aggregated view of the data in near real-time by using five minute tumbling windows to identity short-term trends. You must also have hourly and a daily aggregated views of the data.
Which technology should you use for each task? To answer, drag the appropriate technologies to the correct tasks. 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.



Answer:

Explanation: Box 1: Azure HDInsight MapReduce
Azure Event Hubs allows you to process massive amounts of data from websites, apps, and devices. The Event Hubs spout makes it easy to use Apache Storm on HDInsight to analyze this data in real time.
Box 2: Azure Event Hub
Box 3: Azure Stream Analytics
Stream Analytics is a new service that enables near real time complex event processing over streaming data. Combining Stream Analytics with Azure Event Hubs enables near real time processing of millions of events per second. This enables you to do things such as augment stream data with reference data and output to storage (or even output to another Azure Event Hub for additional processing).

NEW QUESTION 27
You need to ingest data from various data stores into a Microsoft Azure SQL data warehouse by using PolyBase.
You create an Azure Data Factory.
Which three components should you create next? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

A. an Azure Function
B. datasets
C. a pipeline
D. an Azure Batch account
E. linked services

Answer: AE

NEW QUESTION 31
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 states goals. Some question sets might have more than one correct solution, while the 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 implement a new data warehouse.
You have the following information regarding the data warehouse:

- The first data files for the data warehouse will be available in a few days.
- Most queries that will be executed against the data warehouse are ad-hoc.
- The schemas of data files that will be loaded to the data warehouse change often.
- One month after the planned implementation, the data warehouse will contain 15 TB of data. You need to recommend a database solution to support the planned implementation.

Solution: You recommend an Apache Hadoop system. Does this meet the goal?

A. Yes
B. No

Answer: A

NEW QUESTION 33
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.
Your company has multiple databases that contain millions of sales transactions. You plan to implement a data mining solution to identity purchasing fraud.
You need to design a solution that mines 10 terabytes (TB) of sales data. The solution must meet the following requirements:

- Run the analysis to identify fraud once per week.
- Continue to receive new sales transactions while the analysis runs.
- Be able to stop computing services when the analysis is NOT running. Solution: You create a Microsoft Azure HDlnsight cluster.

Does this meet the goal?

A. Yes
B. No

Answer: B

Explanation: HDInsight cluster billing starts once a cluster is created and stops when the cluster is deleted. Billing is pro-rated per minute, so you should always delete your cluster when it is no longer in use.

NEW QUESTION 36
A Company named Fabrikam, Inc. has a web app. Millions of users visit the app daily.
Fabrikam performs a daily analysis of the previous day's logs by scheduling the following Hive query.

```
CREATE EXTERNAL TABLE IF NOT EXISTS UserActivity (...) PARTITIONED BY (LogDate string) LOCATION 'wasb:///logs';
MSCK REPAIR TABLE UserActivity;
Select …. From UserActivity where LogDate = "{date}";
```

You need to recommend a solution to gather the log collections from the web app. What should you recommend?

A. Generate a single directory that contains multiple files for each da
B. Name the file by using the syntax of{date}_{randomsuffix}.txt.
C. Generate a directory that is named by using the syntax of "LogDate={date}" and generate a set of files for that day.
D. Generate a directory each day that has a single file.
E. Generate a single directory that has a single file for each day.

Answer: B

NEW QUESTION 40
A company named Fabrikam, Inc. has a Microsoft Azure web app. Billions of users visit the app daily.
The web app logs all user activity by using text files in Azure Blob storage. Each day, approximately 200 GB of text files are created.
Fabrikam uses the log files from an Apache Hadoop cluster on Azure DHlnsight.
You need to recommend a solution to optimize the storage of the log files for later Hive use.
What is the best property to recommend adding to the Hive table definition to achieve the goal? More than one answer choice may achieve the goal. Select the BEST answer.

A. STORED AS RCFILE
B. STORED AS GZIP
C. STORED AS ORC
D. STORED AS TEXTFILE

Answer: C

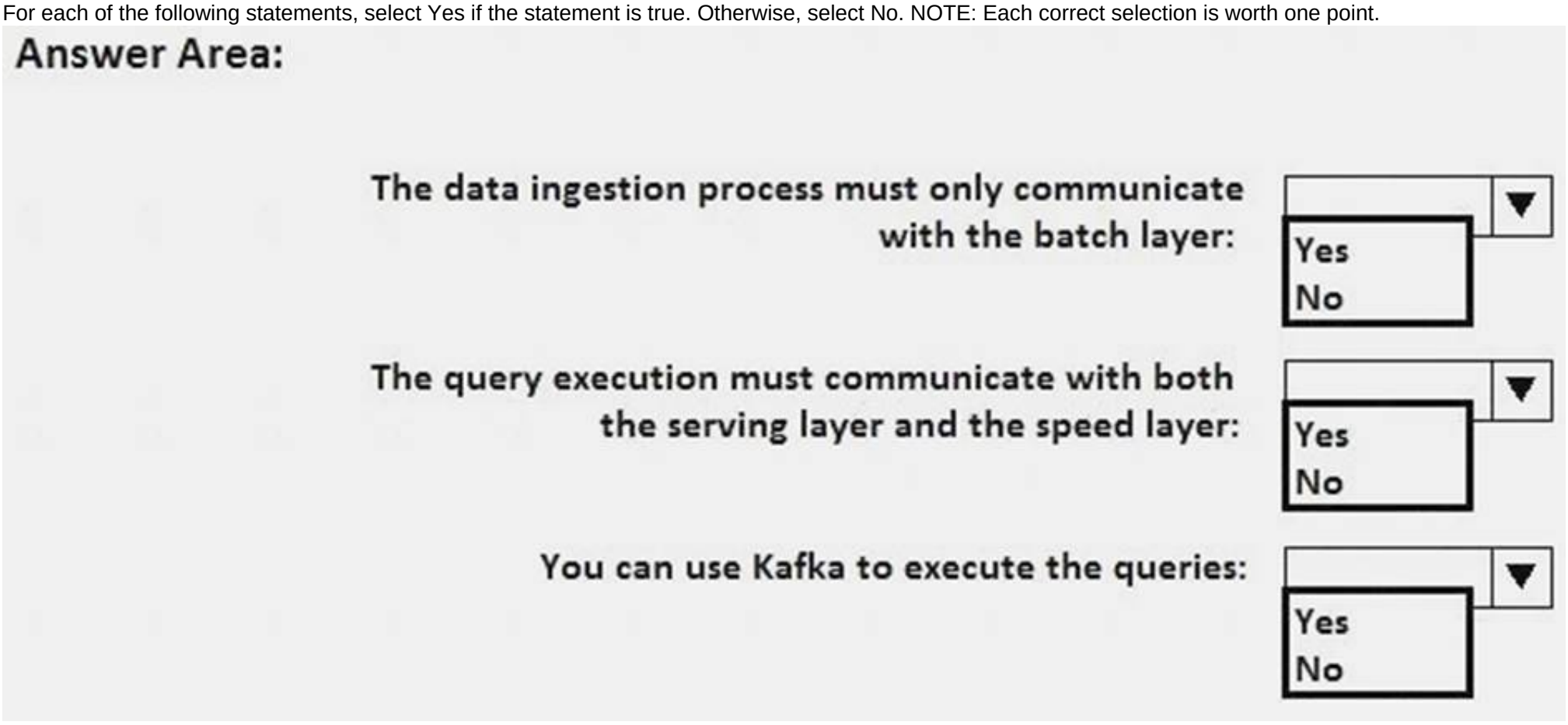
Explanation: The Optimized Row Columnar (ORC) file format provides a highly efficient way to store Hive data. It was designed to overcome limitations of the other Hive file formats. Using ORC files improves performance when Hive is reading, writing, and processing data.
Compared with RCFile format, for example, ORC file format has many advantages such as:

- a single file as the output of each task, which reduces the NameNode's load
- Hive type support including datetime, decimal, and the complex types (struct, list, map, and union)
- light-weight indexes stored within the file
- skip row groups that don't pass predicate filtering
- seek to a given row
- block-mode compression based on data type
- run-length encoding for integer columns
- dictionary encoding for string columns
- concurrent reads of the same file using separate RecordReaders
- ability to split files without scanning for markers
- bound the amount of memory needed for reading or writing
- metadata stored using Protocol Buffers, which allows addition and removal of fields

NEW QUESTION 43
You are designing a solution based on the lambda architecture. The solution has the following layers;

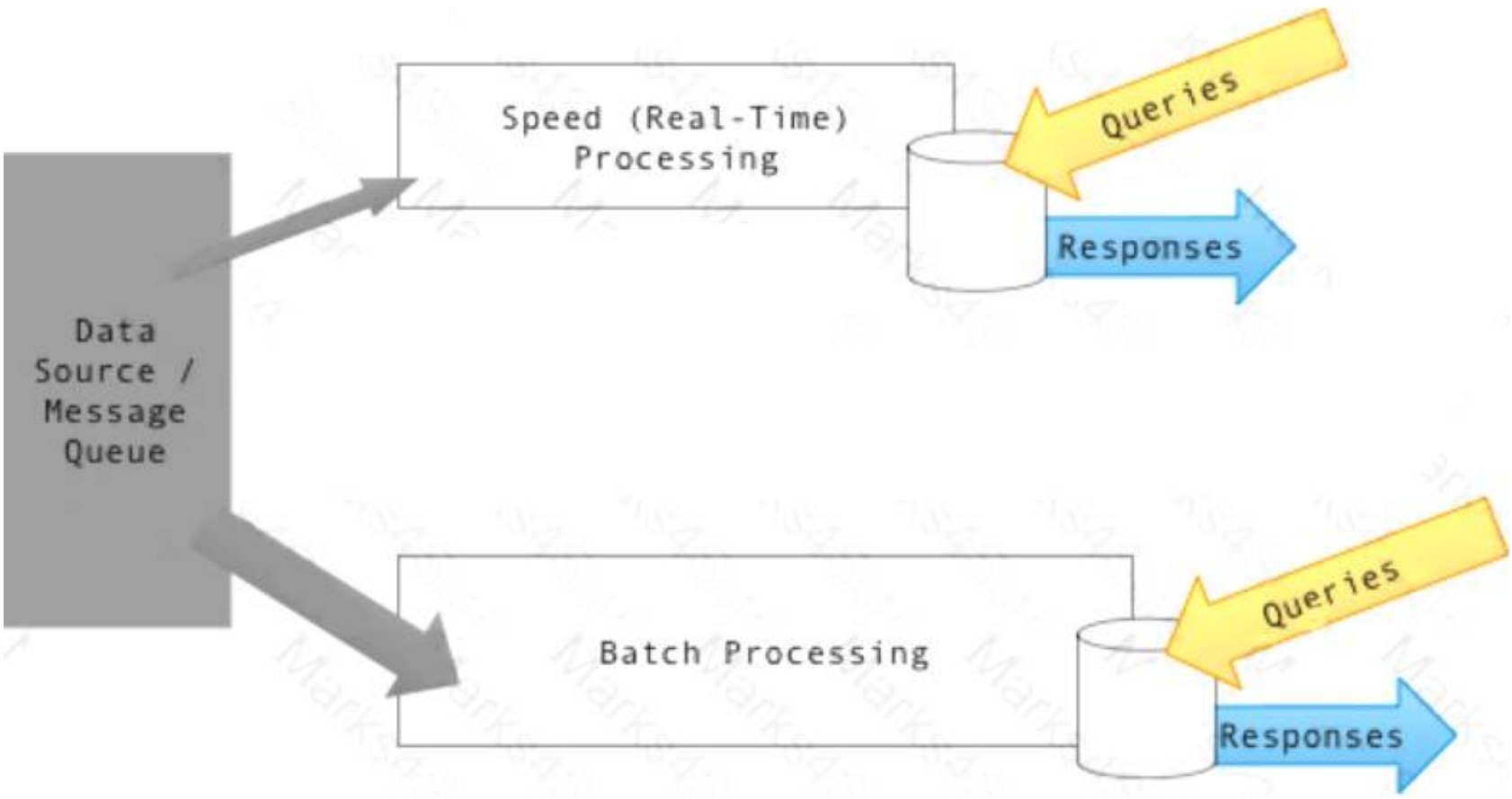
- Batch
- Speed
- Serving

You are planning the data ingestion process and the query execution.
For each of the following statements, select Yes if the statement is true. Otherwise, select No. NOTE: Each correct selection is worth one point.



Answer:

Explanation: Box 1: No
Box 2: No
Output from the batch and speed layers are stored in the serving layer, which responds to ad-hoc queries by returning precomputed views or building views from the processed data.



Box 3: Yes.
We are excited to announce Interactive Queries, a new feature for stream processing with Apache Kafka. Interactive Queries allows you to get more than just processing from streaming.
Note: Lambda architecture is a popular choice where you see stream data pipelines applied (speed layer). Architects can combine Apache Kafka or Azure Event Hubs (ingest) with Apache Storm (event processing),
Apache HBase (speed layer), Hadoop for storing the master dataset (batch layer), and, finally, Microsoft Power BI for reporting and visualization (serving layer).

NEW QUESTION 47
You are designing an Internet of Thing: (IoT) solution intended to identify trends. The solution requires the realtime analysis of data originating from sensors. The results of the analysis will be stored in a SQL database.
You need to recommend a data processing solution that uses the Transact-SQL language. Which data processing solution should you recommend?

A. Microsoft Azure Stream Analytics
B. Microsoft SQL Server Integration Services (SSIS)
C. Microsoft Azure Machine Learning
D. Microsoft Azure HDInsight Hadoop clusters

Answer: A

NEW QUESTION 52
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 states goals. Some question sets might have more than one correct solution, while the 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 implement a new data warehouse.
You have the following information regarding the data warehouse:

- The first data files for the data warehouse will be available in a few days.
- Most queries that will be executed against the data warehouse are ad-hoc.
- The schemas of data files that will be loaded to the data warehouse change often.
- One month after the planned implementation, the data warehouse will contain 15 TB of data. You need to recommend a database solution to support the planned implementation.

Solution: You recommend a Microsoft SQL server on a Microsoft Azure virtual machine. Does this meet the goal?

A. Yes
B. No

Answer: B

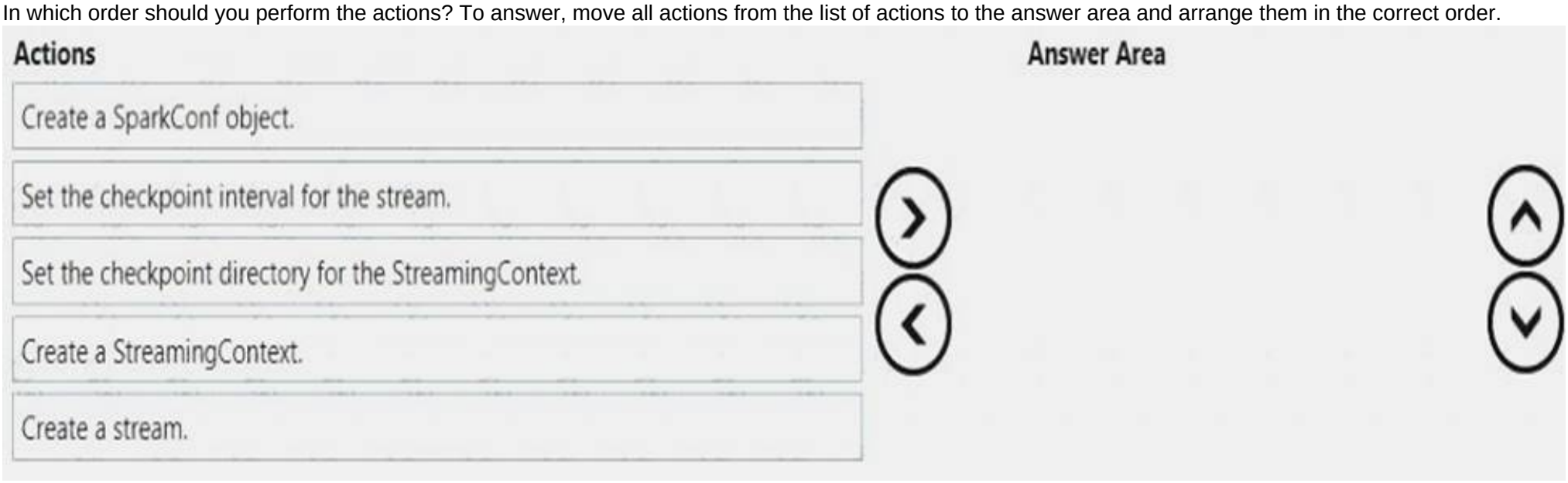
NEW QUESTION 56
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution.
After 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 deploy a Microsoft Azure SQL data warehouse and a web application.
The data warehouse will ingest 5 TB of data from an on-premises Microsoft SQL Server database daily. The web application will query the data warehouse.
You need to design a solution to ingest data into the data warehouse.
Solution: You use SQL Server Integration Services (SSIS) to transfer data from SQL Server to Azure SQL Data Warehouse.
Does this meet the goal?

A. Yes
B. No

Answer: B

Explanation: Integration Services (SSIS) is a powerful and flexible Extract Transform and Load (ETL) tool that supports complex workflows, data transformation, and several data loading options.
The main drawback is speed. We should use Polybase instead.
References: https://docs.microsoft.com/en-us/sql/integration-services/sql-server-integration-services

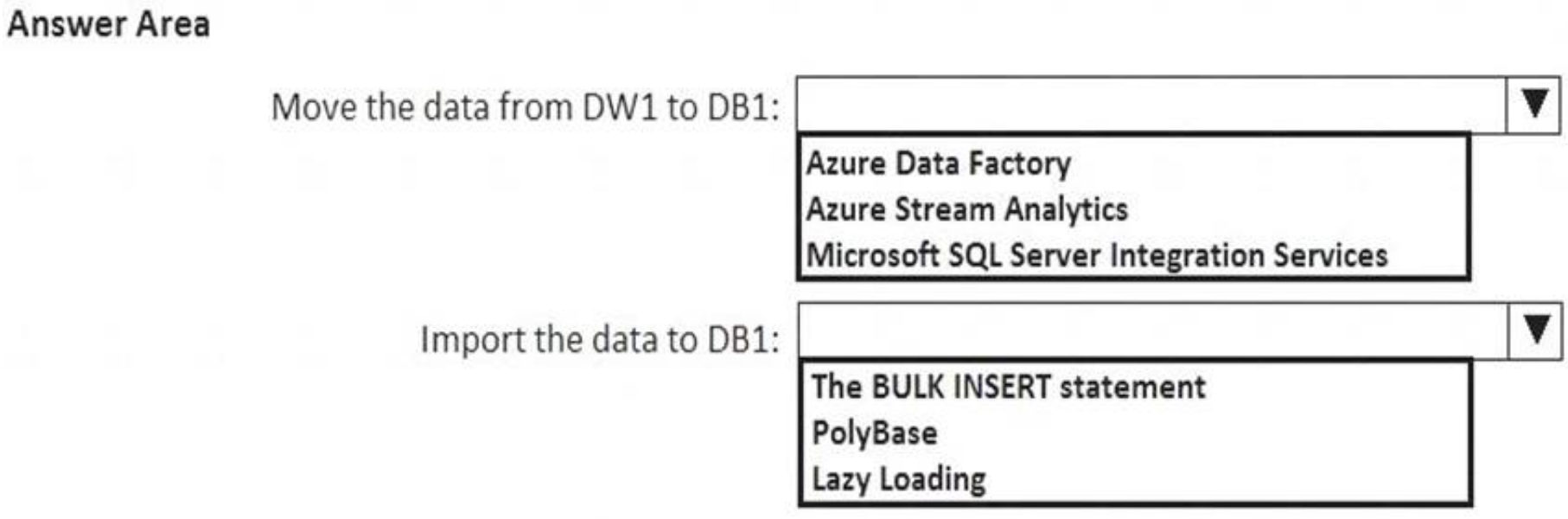
NEW QUESTION 57
You have an Apache Spark cluster on Microsoft Azure HDInsight for all analytics workloads.
You plan to build a Spark streaming application that processes events ingested by using Azure Event Hubs. You need to implement checkpointing in the Spark streaming application for high availability of the event
data.
In which order should you perform the actions? To answer, move all actions from the list of actions to the answer area and arrange them in the correct order.



Answer:

Explanation:

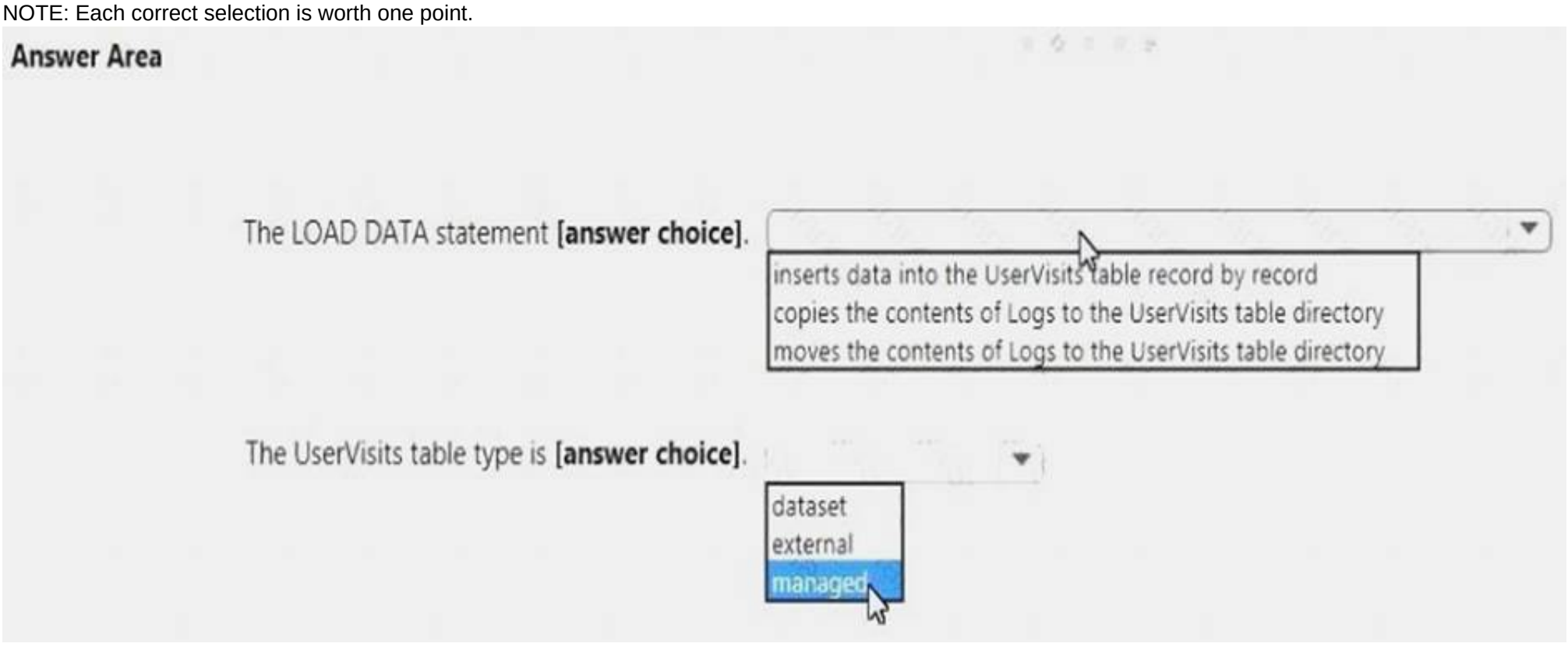
NEW QUESTION 59
You have a Microsoft Azure SQL data warehouse named DW1.
A department in your company creates an Azure SQL database named DB1. DB1 is a data mart.
Each night, you need to insert new rows Into 9.000 tables in DB1 from changed data in DW1. The solution must minimize costs.
What should you use to move the data from DW1 to DB1, and then to import the changed data to DB1? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.



Answer:

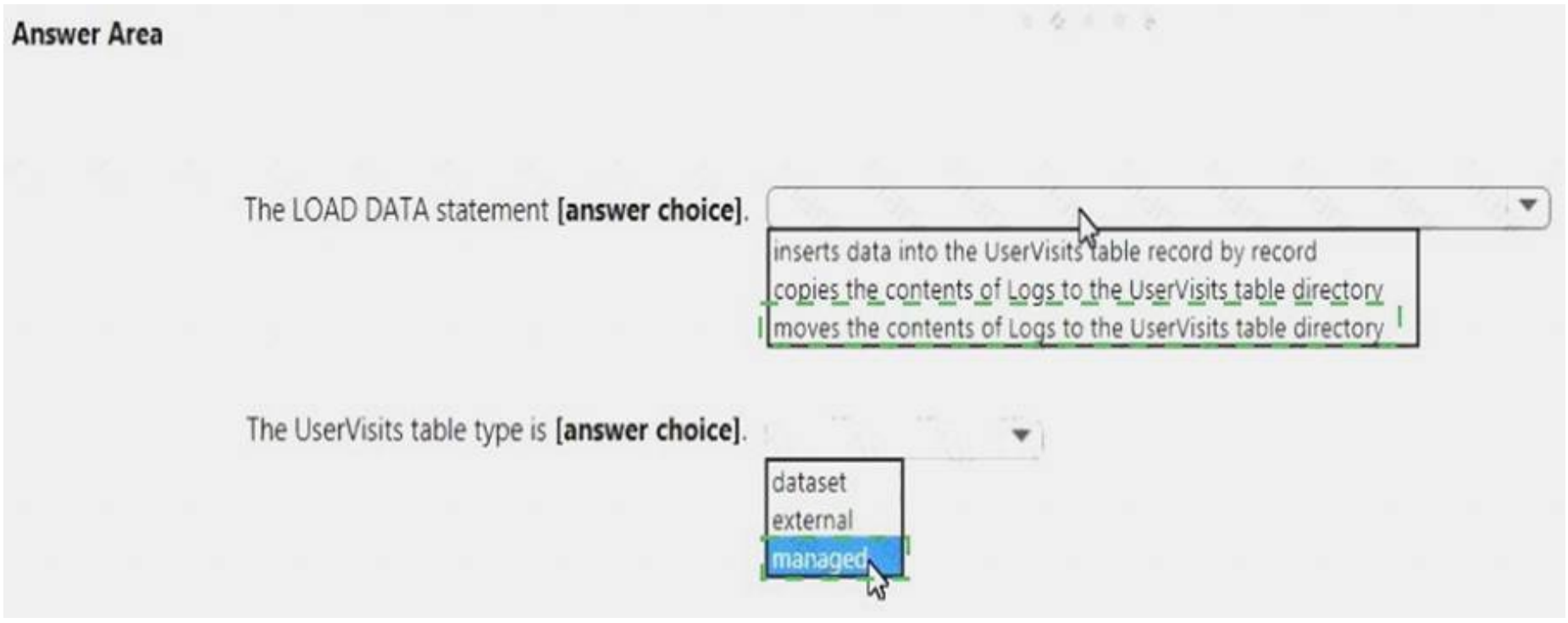
Explanation: Box 1: Azure Data Factory
Use the Copy Activity in Azure Data Factory to move data to/from Azure SQL Data Warehouse. Box 2: The BULK INSERT statement

NEW QUESTION 63
You have the following Hive query.
CREATE TABLE UserVisits (username string, urlvisited string, time date); LOAD DATA INPATH 'wasb:///Logs' OVERWRITE INTO TABLE UserVisits;
Use the drop-down menus to select the answer choice that completes each statement based on the information presented in the script.
NOTE: Each correct selection is worth one point.



Answer:

Explanation:



NEW QUESTION 67
You plan to deploy a Hadoop cluster that includes a Hive installation.
Your company identifies the following requirements for the planned deployment:

- During the creation of the cluster nodes, place JAR files in the clusters.
- Decouple the Hive metastore lifetime from the cluster lifetime.
- Provide anonymous access to the cluster nodes.

You need to identify which technology must be used for each requirement.
Which technology should you identify for each requirement? To answer, drag the appropriate technologies to the correct requirements. 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.

Technologies	Answer Area	
An Azure SQL Database External Metastore	Provide anonymous access to the cluster nodes:	
An Azure Table Storage External Metastore	During the creation of the cluster nodes, place the JAR files in the clusters:	
An Azure virtual network	Decouple the Hive metastore lifetime from the cluster lifetime:	
Script Actions		

Answer:

Explanation:

Technologies	Answer Area	
An Azure SQL Database External Metastore	Provide anonymous access to the cluster nodes:	An Azure virtual network
An Azure Table Storage External Metastore	During the creation of the cluster nodes, place the JAR files in the clusters:	Script Actions
An Azure virtual network	Decouple the Hive metastore lifetime from the cluster lifetime:	An Azure SQL Database External Metastore
Script Actions		

NEW QUESTION 71
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 the 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 have a Microsoft Azure deployment that contains the following services:

- Azure Data Lake
- Azure Cosmos DB
- Azure Data Factory
- Azure SQL Database

You load several types of data to Azure Data Lake.
You need to load data from Azure SQL Database to Azure Data Lake. Solution: You use the AzCopy utility.
Does this meet the goal?

A. Yes
B. No

Answer: B

Explanation: Note: You can use the Copy Activity in Azure Data Factory to copy data to and from Azure Data Lake Storage Gen1 (previously known as Azure Data Lake Store). Azure SQL database is supported as source.
References: https://docs.microsoft.com/en-us/azure/data-factory/connector-azure-data-lake-store

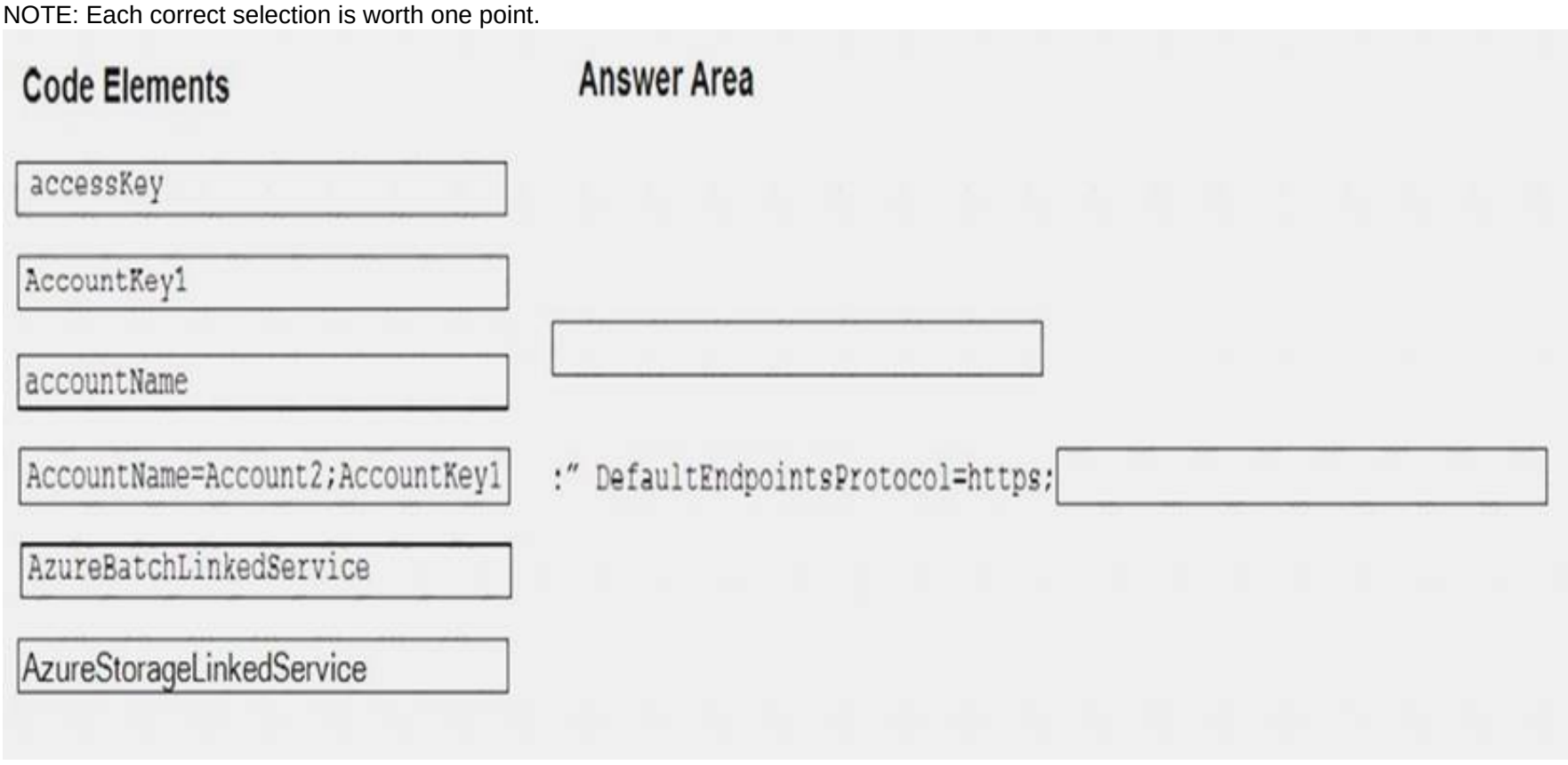
NEW QUESTION 74
You are designing an Apache HBase cluster on Microsoft Azure HDInsight. You need to identify which nodes are required for the cluster.
Which three nodes should you identify? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

A. Nimbus
B. Zookeeper
C. Region
D. Supervisor
E. Falcon
F. Head

Answer: BCF

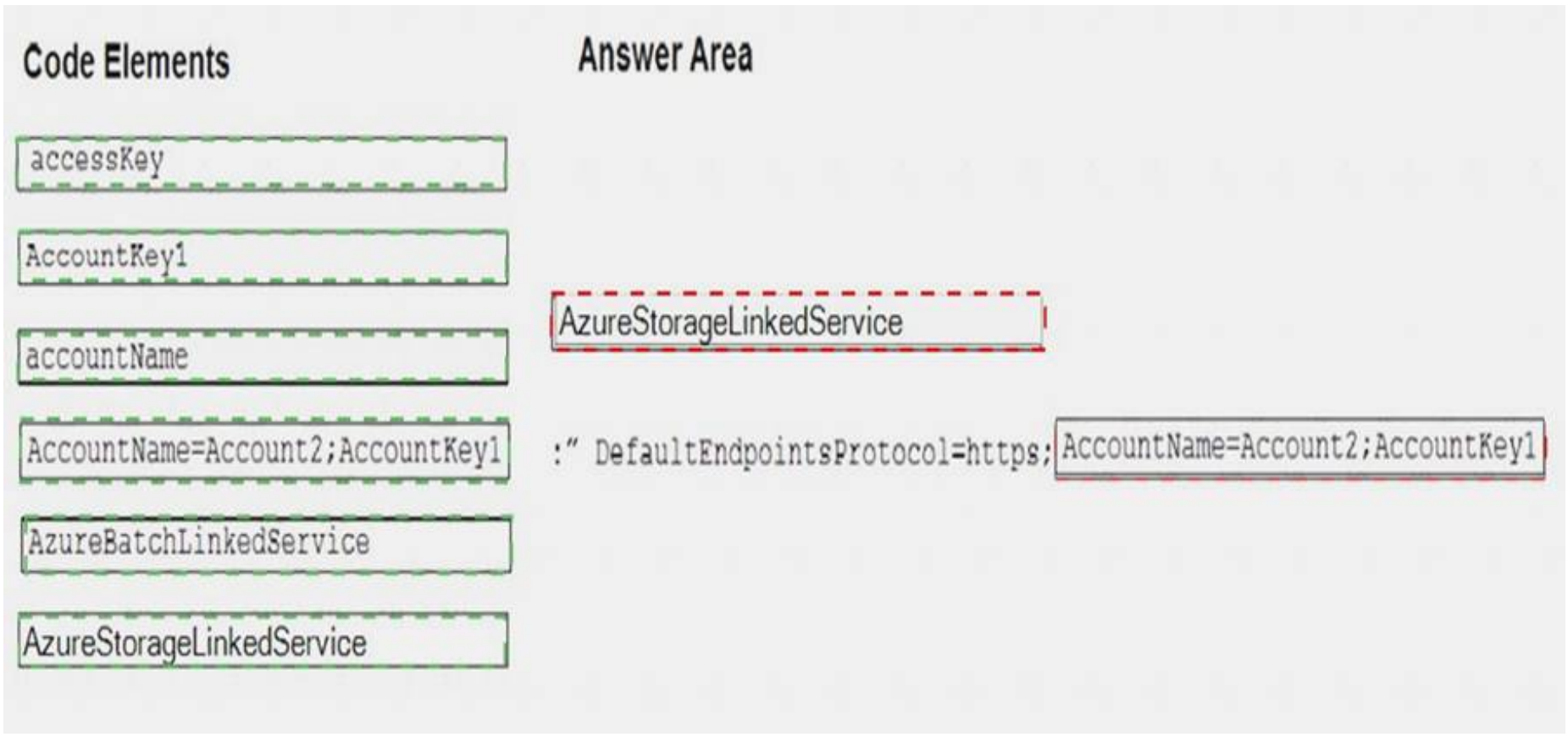
Explanation: https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-hadoop-provision-linux-clusters

NEW QUESTION 78
You are automating the deployment of a Microsoft Azure Data Factory solution. The data factory will interact with a file stored in Azure Blob storage.
You need to use the REST API to create a linked service to interact with the file.
How should you complete the request body? To answer, drag the appropriate code elements to the correct locations. Each code may be used once, more than once, or not at all. You may need to drag the slit bar between panes or scroll to view content.
NOTE: Each correct selection is worth one point.



Answer:

Explanation:



NEW QUESTION 82
Your company has two Microsoft Azure SQL databases named db1 and db2.
You need to move data from a table in db1 to a table in db2 by using a pipeline in Azure Data Factory. You create an Azure Data Factory named ADF1.
Which two types Of objects Should you create In ADF1 to complete the pipeline? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

A. a linked service
B. an Azure Service Bus
C. sources and targets
D. input and output I datasets
E. transformations

Answer: AD

Explanation: You perform the following steps to create a pipeline that moves data from a source data store to a sink data store:

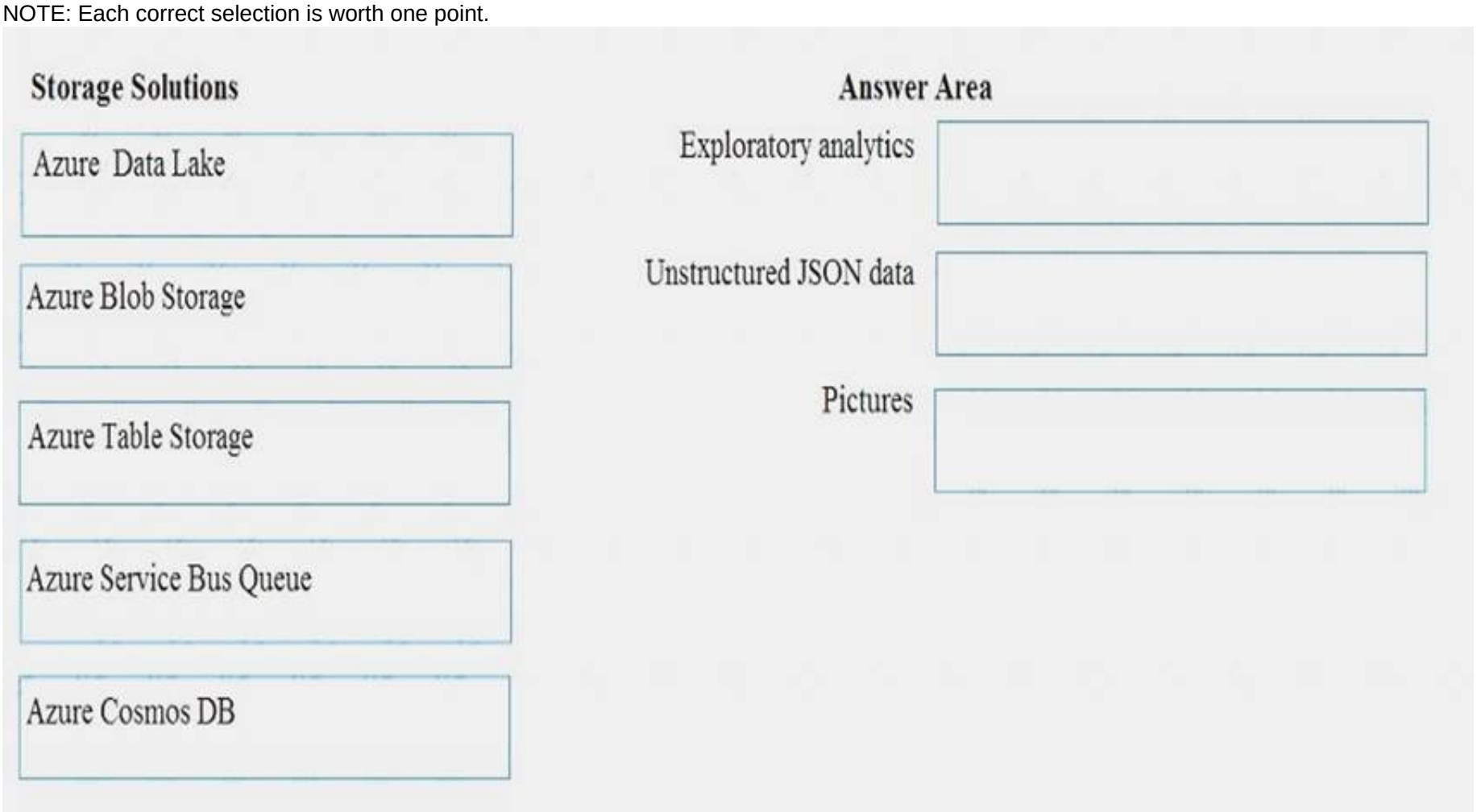
Create linked services to link input and output data stores to your data factory.

- Create datasets to represent input and output data for the copy operation.
- Create a pipeline with a copy activity that takes a dataset as an input and a dataset as an output.

NEW QUESTION 85
You plan to deploy a storage solution to store the output of stream analytics. You plan to store the data for the following three types of data streams:

- Unstructured JSON data
- Exploratory analytics
- Pictures

You need to implement a storage solution for the data stream types.
Which storage solution should you implement for each data stream type? To answer, drag the appropriate storage solutions to the correct data stream types. Each storage solution may be used once, more than once, or not at all. You may need to drag the split bar between the panes or scroll to view content.
NOTE: Each correct selection is worth one point.



Answer:

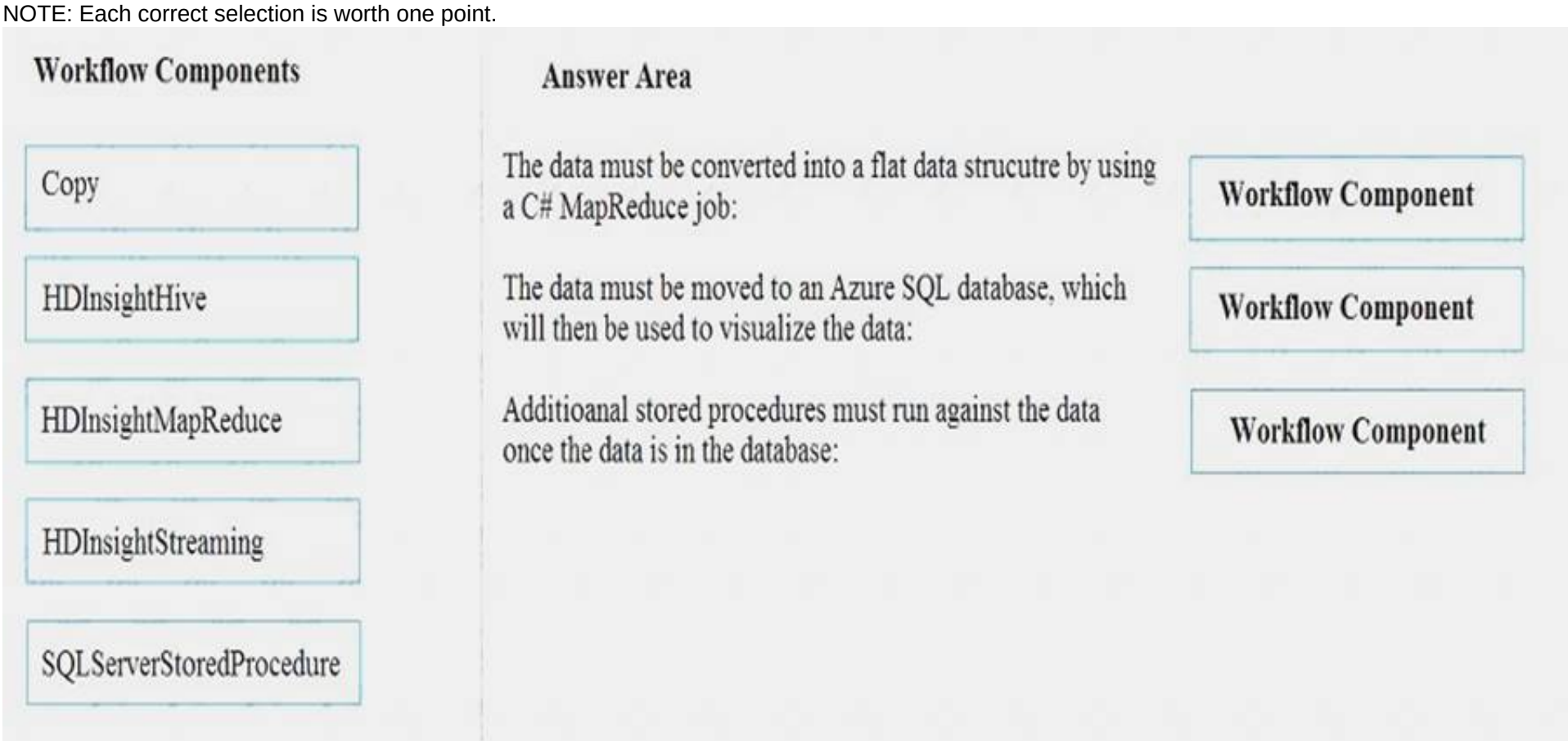
Explanation: Box 1: Azure Data Lake Store
Stream Analytics supports Azure Data Lake Store. Azure Data Lake Store is an enterprise-wide hyper-scale repository for big data analytic workloads. Data Lake Store enables you to store data of any size, type and ingestion speed for operational and exploratory analytics. Stream Analytics has to be authorized to access the Data Lake Store.
Box 2: Azure Cosmos DB
Stream Analytics can target Azure Cosmos DB for JSON output, enabling data archiving and low-latency queries on unstructured JSON data.
Box 3: Azure Blob Storage
Blob storage offers a cost-effective and scalable solution for storing large amounts of unstructured data in the cloud.
Incorrect Asnwers: Azure SQL Database:
Azure SQL Database can be used as an output for data that is relational in nature or for applications that depend on content being hosted in a relational database. Stream Analytics jobs write to an existing table in an Azure SQL Database.
Azure Service Bus Queue:
Service Bus Queues offer a First In, First Out (FIFO) message delivery to one or more competing consumers. Typically, messages are expected to be received and processed by the receivers in the temporal order in which they were added to the queue, and each message is received and processed by only one message consumer.
Azure Table Storage
Azure Table storage offers highly available, massively scalable storage, so that an application can automatically scale to meet user demand. Table storage is Microsoft's NoSQL key/attribute store, which one can leverage for structured data with fewer constraints on the schema. Azure Table storage can be used to store data for persistence and efficient retrieval.
References: https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-define-outputs

NEW QUESTION 86
You have raw data in Microsoft Azure Blob storage. Each data file is 10 KB and is the XML format. You identify the following requirements for the data:

- The data must be converted into a flat data structure by using a C# MapReduce job.
- The data must be moved to an Azure SQL database, which will then be used to visualize the data.
- Additional stored procedures must run against the data once the data is in the database.

You need to create the workflow for the Azure Data Factory pipeline.
Which activity type should you use for each requirement? To answer, drag the appropriate workflow components to the correct requirements. Each workflow component may be used once, more than once, or not at all. You may need to drag the split bar between the panes or scroll to view content.

NOTE: Each correct selection is worth one point.



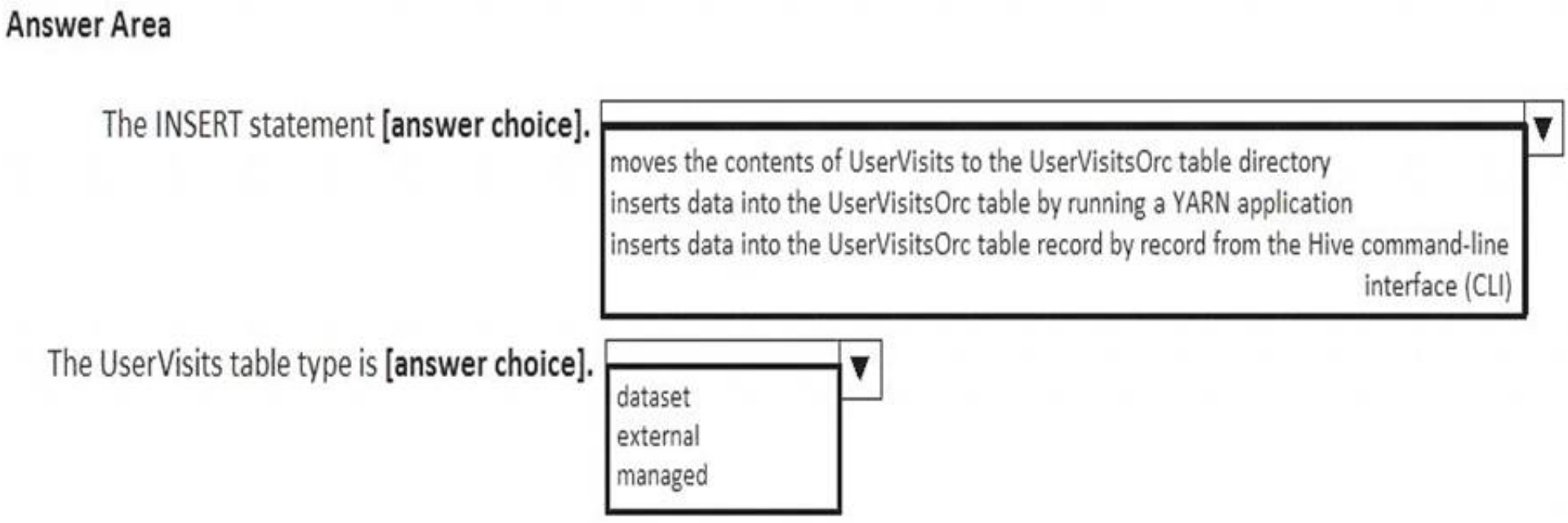
Answer:

Explanation: Box 1: HDinsightMapReduce
The HDInsight MapReduce activity in a Data Factory pipeline invokes MapReduce program on your own or on-demand HDInsight cluster.
Box 2: HDInsightStreaming
Box 3: SQLServerStoredProcedure

NEW QUESTION 91
You have the following script.

```
CREATE TABLE UserVisits (username string, url string, time date) STORED AS TEXTFILE LOCATION "wasb:///Logs";
CREATE TABLE UserVisitsOrc (username string, url string, time date) STORED AS ORC;
INSERT INTO TABLE UserVisitsOrc SELECT * FROM UserVisits
```

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



Answer:

Explanation: A table created without the EXTERNAL clause is called a managed table because Hive manages its data.

NEW QUESTION 96
A company named Fabricam, Inc, has a web app hosted in Microsoft Azure. Millions of users visit the app daily.
All of the user visits are logged in Azure Blob storage. Data analysts at Fabrikam built a dashboard that processes the user visit logs.
Fabrikam plans to use an Apache Hadoop cluster on Azure HDInsight to process queries. The queries will access the data only once.
You need to recommend a query execution strategy. What is the best to recommend using to achieve the goal?
More than one answer choice may achieve the goal. Select the BEST answer.

A. Load the text files to ORC files, and then run dashboard queries on the ORC files.
B. Load the text files to sequence files, and then run dashboard queries on the sequence files.
C. Run the queries on the text files directly.

D. Load the text files to parquet files, and then run dashboard queries on the parquet files.

Answer: B

Explanation: File format versatility and Intelligent caching: Fast analytics on Hadoop have always come with one big catch: they require up-front conversion to a columnar format like ORCFile, Parquet or Avro, which is
time-consuming, complex and limits your agility.
With Interactive Query Dynamic Text Cache, which converts CSV or JSON data into optimized in-memory format on-the-fly, caching is dynamic, so the queries determine what data is cached. After text data is cached, analytics run just as fast as if you had converted it to specific file formats.
References:
https://azure.microsoft.com/en-us/blog/azure-hdinsight-interactive-query-simplifying-big-data-analytics-architec

NEW QUESTION 100
You are designing a solution for an Internet of Things (IoT) project.
You need to recommend a data storage solution for the project. The solution must meet the following requirements:

- Allow data to be queried in real-time as it streams into the solution
- Provide the lowest amount of latency for loading data into the solution. What should you include in the recommendation?

A. a Microsoft Azure SQL database that has In-Memory OLTP enabled
B. a Microsoft Azure HDInsight Hadoop cluster
C. a Microsoft Azure HDInsight R Server cluster
D. a Microsoft Azure Table Storage solution

Answer: A

Explanation: References:
https://azure.microsoft.com/en-gb/blog/in-memory-oltp-in-azure-sql-database/

NEW QUESTION 101
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 states goals. Some question sets might have more than one correct solution, while the 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 implement a new data warehouse.
You have the following information regarding the data warehouse:

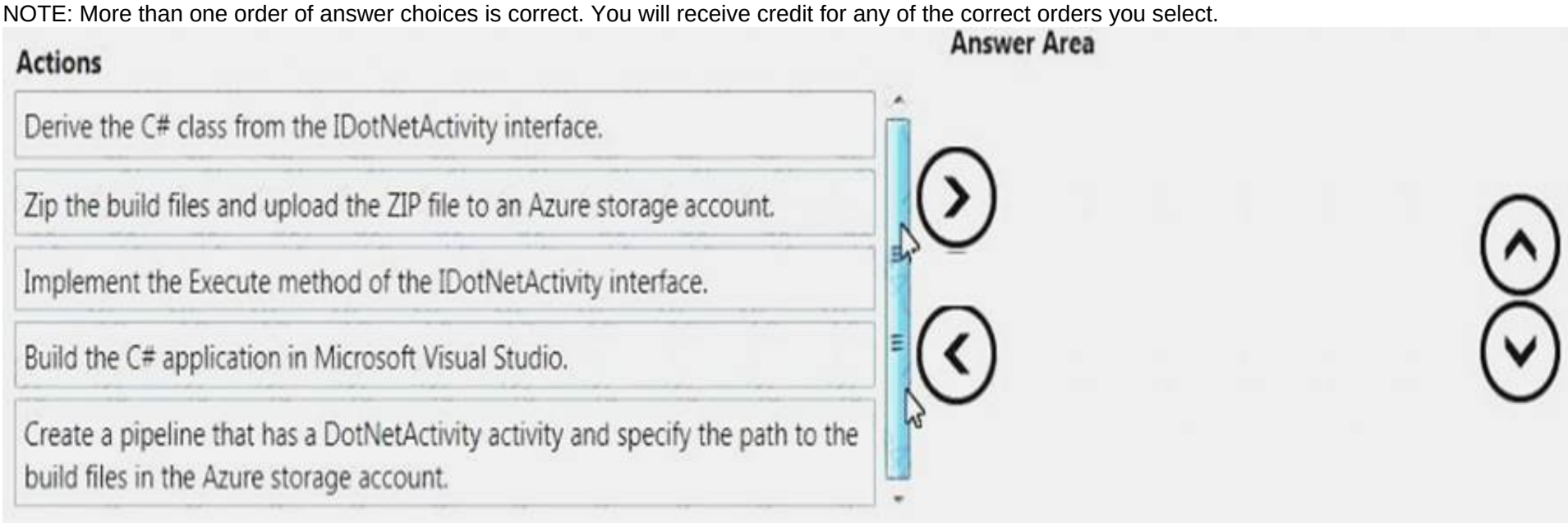
- The first data files for the data warehouse will be available in a few days.
- Most queries that will be executed against the data warehouse are ad-hoc.
- The schemas of data files that will be loaded to the data warehouse change often.
- One month after the planned implementation, the data warehouse will contain 15 TB of data. You need to recommend a database solution to support the planned implementation.

Solution: You recommend an Apache Spark system. Does this meet the goal?

A. Yes
B. No

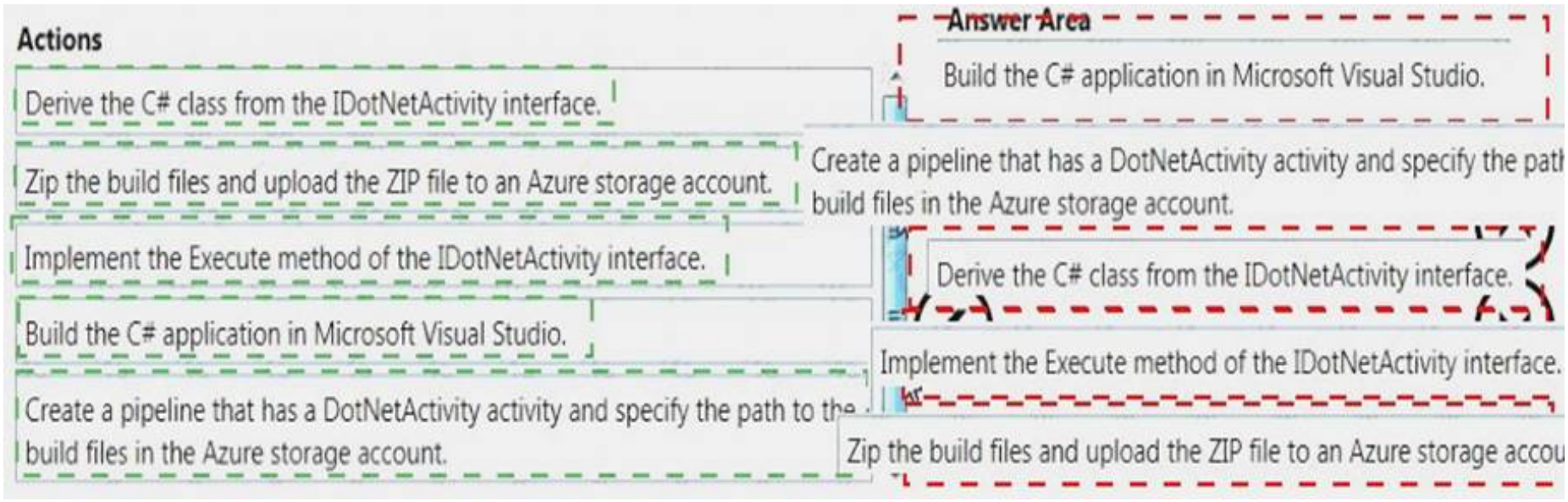
Answer: B

NEW QUESTION 106
You have a Microsoft Azure HDInsight cluster for analytics workloads. You have a C# application on a local computer.
You plan to use Azure Data Factory to run the C# application in Azure.
You need to create a data factory that runs the C# application by using HDInsight.
In which order should you perform the actions? To answer, move all actions from the list of actions to the answer area and arrange them in the correct order.
NOTE: More than one order of answer choices is correct. You will receive credit for any of the correct orders you select.



Answer:

Explanation:



NEW QUESTION 108
You have an Apache Storm cluster.
The cluster will ingest data from a Microsoft Azure event hub.
The event hub has the characteristics described in the following table.

Setting name	Value
Message Retention	1
Namespace	storm1.servicebus.windows.net
Shared access policies	2
Partition Count	16
Region	Central US

You are designing the Storm application topology.
You need to ingest data from all of the partitions. The solution must maximize the throughput of the data ingestion.
Which setting should you use?

A. Partition Count
B. Message Retention
C. Partition Key
D. Shared access policies

Answer: A

NEW QUESTION 113
You have a pipeline that contains an input dataset in Microsoft Azure Table Storage and an output dataset in Azure Blob storage. You have the following JSON data.

```
availability: {    frequency: "Day",    interval: 3,
"anchorDateTime": "2014-10-10T10:00:00Z"
   waitOnExternal:    {        retryInterval: "00:01:00",        retryTimeout: "00:10:00",        maximumRetry: 3
   } }
```

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

Answer Area

The pipeline will run **[answer choice].**
- every three days at 10:00
- every three days at 22:00
- three times a day starting at 10:00
- three times a day starting at 22:00

If the pipeline fails to run, it will retry **[answer choice].**
- every minute up to three times
- every 10 minutes up to three times
- every 10 minutes until the pipeline succeeds
- every one minute until the pipeline succeeds

Answer:

Explanation: Box 1: Every three days at 10.00
anchorDateTime defines the absolute position in time used by the scheduler to compute dataset slice boundaries.
"frequency": "<Specifies the time unit for data slice production. Supported frequency: Minute, Hour, Day, Week, Month>",
"interval": "<Specifies the interval within the defined frequency. For example, frequency set to 'Hour' and interval set to 1 indicates that new data slices should be

produced hourly>
Box 2: Every minute up to three times.
retryInterval is the wait time between a failure and the next attempt. This setting applies to present time. If the previous try failed, the next try is after the retryInterval period.
Example: 00:01:00 (1 minute)
Example: If it is 1:00 PM right now, we begin the first try. If the duration to complete the first validation check is 1 minute and the operation failed, the next retry is at 1:00 + 1min (duration) + 1min (retry interval) = 1:02 PM.
For slices in the past, there is no delay. The retry happens immediately. retryTimeout is the timeout for each retry attempt.
maximumRetry is the number of times to check for the availability of the external data.

NEW QUESTION 116
You are developing an Apache Storm application by using Microsoft Visual Studio. You need to implement a custom topology that uses a custom bolt. Which type of object should you initialize in the main class?

A. Stream
B. TopologyBuilder
C. StreamInfo
D. Logger

Answer: A

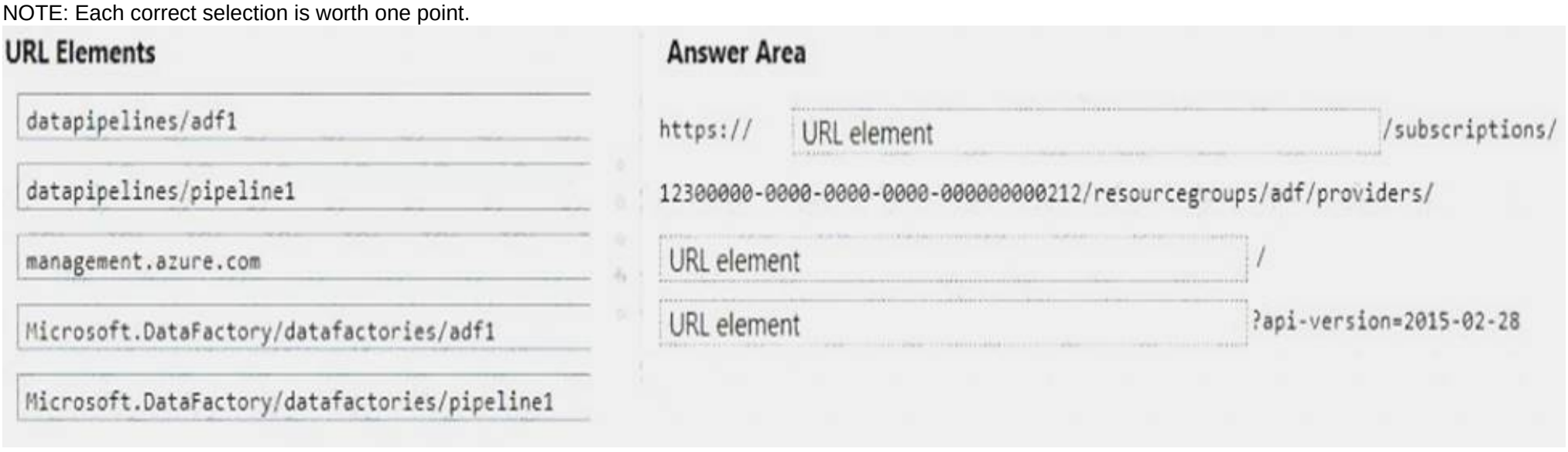
NEW QUESTION 120
You plan to use Microsoft Azure IoT Hub to capture data from medical devices that contain sensors. You need to ensure that each device has its own credentials. The solution must minimize the number of required privileges.
Which policy should you apply to the devices?

A. iothubowner
B. service
C. registryReadWrite
D. device

Answer: D

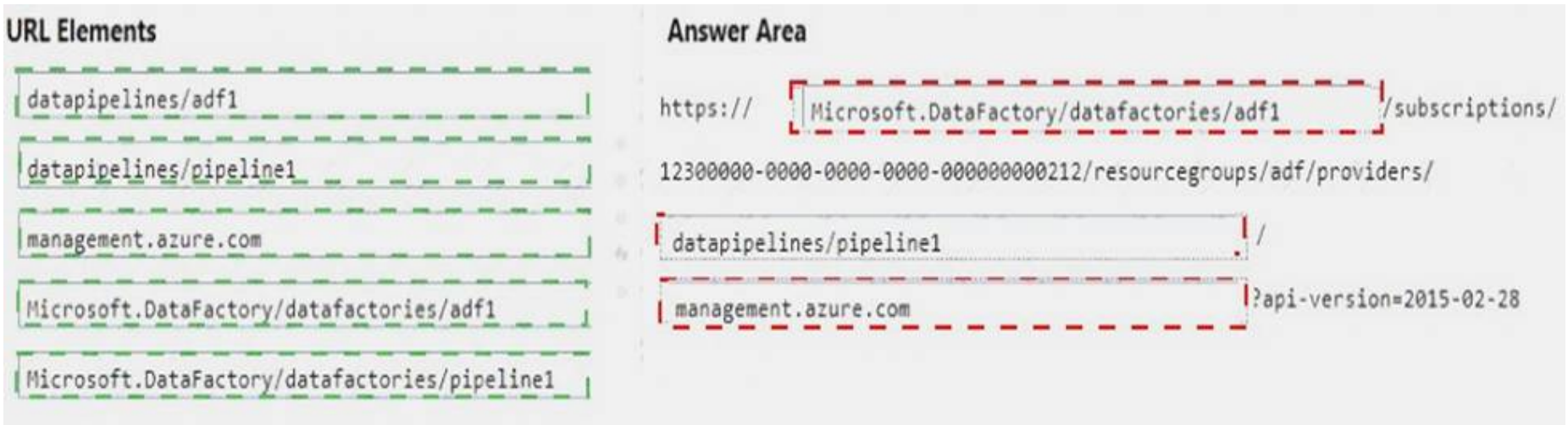
Explanation: Per-Device Security Credentials. Each IoT Hub contains an identity registry For each device in this identity registry, you can configure security credentials that grant DeviceConnect permissions scoped to the corresponding device endpoints.

NEW QUESTION 121
You have a Microsoft Azure data factory named ADF1 that contains a pipeline named Pipeline1. You plan to automate updates to Pipeline1.
You need to build the URL that must be called to update the pipeline from the REST API.
How should you complete the URL? To answer, drag the appropriate URL elements to the correct locations. Each URL element 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:

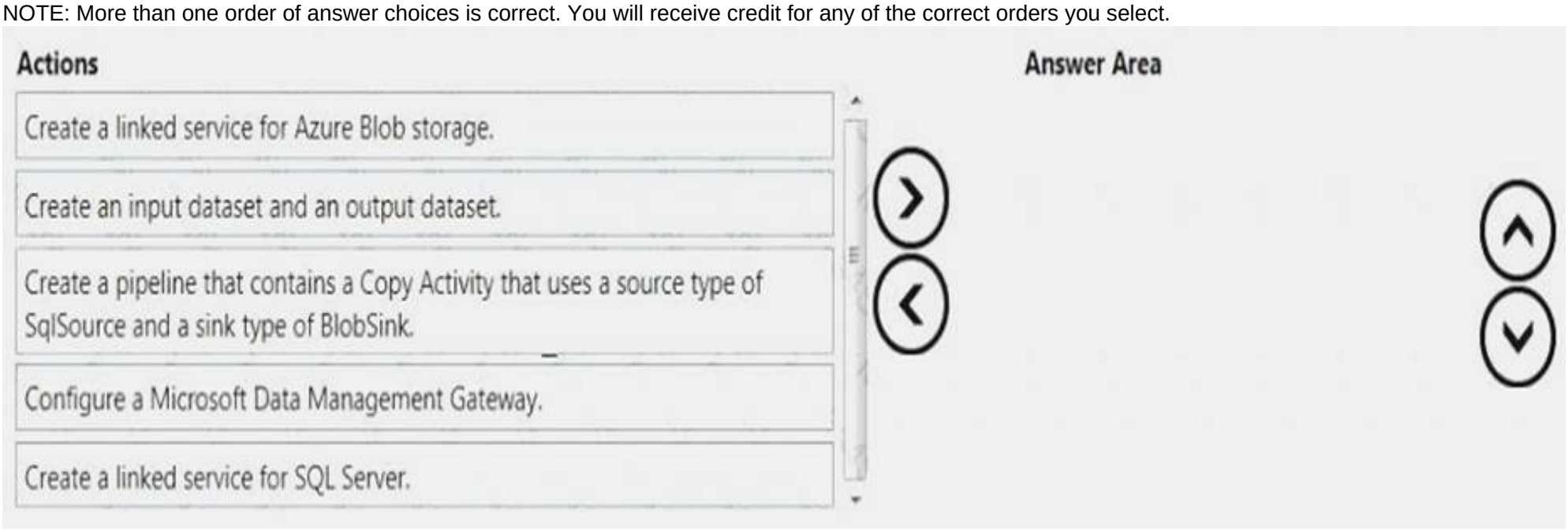
Explanation:



NEW QUESTION 124
You have a Microsoft Azure Machine Learning application named App1 that is used by several departments in your organization.
App 1 connects to an Azure database named DB1. DB1 contains several tables that store sensitive information. You plan to implement a security solution for the tables.
You need to prevent the users of App1 from viewing the data of users in other departments in the tables. The solution must ensure that the users can see only data of the users in their respective department.
Which feature should you implement?
A. Cell-level encryption
B. Row-Level Security (RLS)
C. Transparent Data Encryption (TDE)
D. Dynamic Data Masking

Answer: D

NEW QUESTION 129
You have data in an on-premises Microsoft SQL Server database.
You must ingest the data in Microsoft Azure Blob storage from the on-premises SQL Server database by using Azure Data Factory.
You need to identify which tasks must be performed from Azure.
In which sequence should you perform the actions? To answer, move all of the actions from the list of actions to the answer area and arrange them in the correct order.
NOTE: More than one order of answer choices is correct. You will receive credit for any of the correct orders you select.



Answer:

Explanation: Step 1: Configure a Microsoft Data Management Gateway Install and configure Azure Data Factory Integration Runtime.
The Integration Runtime is a customer managed data integration infrastructure used by Azure Data Factory to provide data integration capabilities across different network environments. This runtime was formerly called "Data Management Gateway".
Step 2: Create a linked service for Azure Blob storage
Create an Azure Storage linked service (destination/sink). You link your Azure storage account to the data factory.
Step 3: Create a linked service for SQL Server
Create and encrypt a SQL Server linked service (source)
In this step, you link your on-premises SQL Server instance to the data factory. Step 4: Create an input dataset and an output dataset.
Create a dataset for the source SQL Server database. In this step, you create input and output datasets. They represent input and output data for the copy operation, which copies data from the on-premises SQL Server database to Azure Blob storage.
Step 5: Create a pipeline..
You create a pipeline with a copy activity. The copy activity uses SqlServerDataset as the input dataset and AzureBlobDataset as the output dataset. The source type is set to SqlSource and the sink type is set to BlobSink.
References: https://docs.microsoft.com/en-us/azure/data-factory/tutorial-hybrid-copy-powershell

NEW QUESTION 132
You have a Microsoft Azure subscription that contains an Azure Data Factory pipeline. You have an RSS feed that is published on a public website.

You need to configure the RSS feed as a data source for the pipeline. Which type of linked service should you use?

A. web
B. OData
C. Azure Search
D. Azure Data Lake Store

Answer: A

Explanation: Reference: https://docs.microsoft.com/en-us/azure/data-factory/data-factory-web-table-connector

NEW QUESTION 134
You need to automate the creation of a new Microsoft Azure data factory.
What are three possible technologies that you can use? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point

A. Azure PowerShell cmdlets
B. the SOAP service
C. T-SQL statements
D. the REST API
E. the Microsoft .NET framework class library

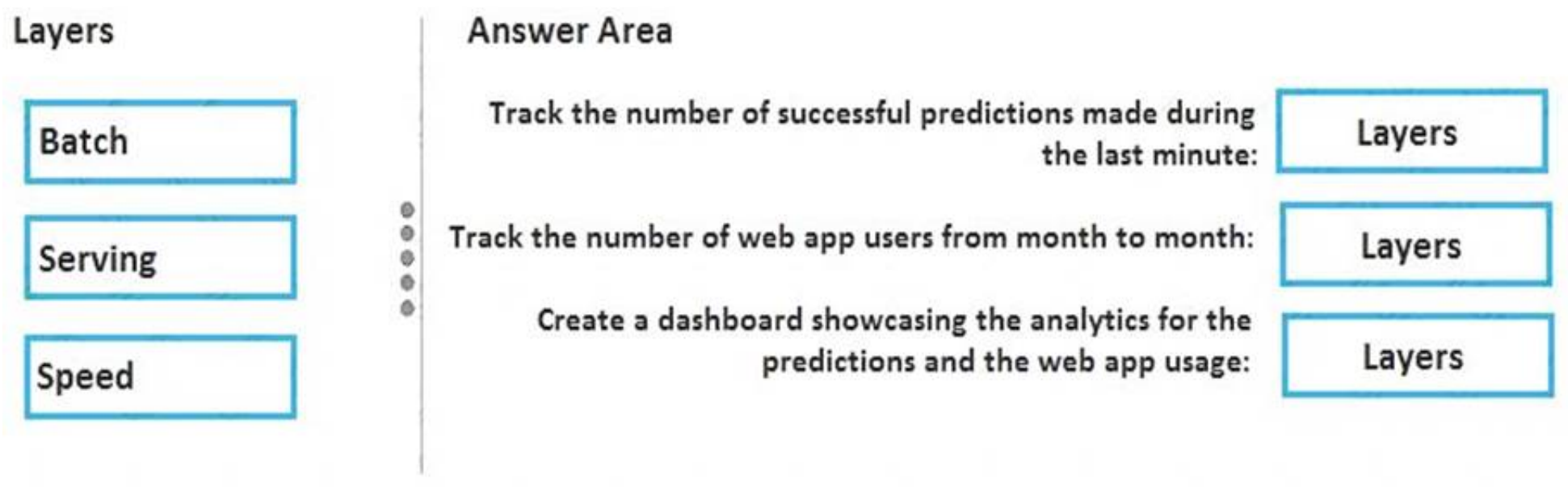
Answer: ADE

Explanation: https://docs.microsoft.com/en-us/azure/data-factory/data-factory-introduction

NEW QUESTION 139
You have a web app that accepts user input, and then uses a Microsoft Azure Machine Learning model to predict a characteristic of the user.
You need to perform the following operations:

- Track the number of web app users from month to month.
- Track the number of successful predictions made during the last minute.
- Create a dashboard showcasing the analytics tor the predictions and the web app usage.

Which lambda layer should you query for each operation? To answer, drag the appropriate layers to the correct operations. Each layer 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: Lambda architecture is a data-processing architecture designed to handle massive quantities of data by taking advantage of both batch- and stream-processing methods. This approach to architecture attempts to balance latency, throughput, and fault-tolerance by using batch processing to provide comprehensive and accurate views of batch data, while simultaneously using real-time stream processing to provide views of online data. The two view outputs may be joined before presentation
Box 1: Speed
The speed layer processes data streams in real time and without the requirements of fix-ups or completeness. This layer sacrifices throughput as it aims to minimize latency by providing real-time views into the most recent data.
Box 2: Batch
The batch layer precomputes results using a distributed processing system that can handle very large quantities of data. The batch layer aims at perfect accuracy by being able to process all available data when generating views.
Box 3: Serving
Output from the batch and speed layers are stored in the serving layer, which responds to ad-hoc queries by returning precomputed views or building views from the processed data.

NEW QUESTION 140
You have an Apache Hive cluster in Microsoft Azure HDInsight. The cluster contains 10 million data files. You plan to archive the data.

The data will be analyzed monthly.
You need to recommend a solution to move and store the data. The solution must minimize how long it takes to move the data and must minimize costs.
Which two services should you include in the recommendation? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

A. Azure Queue storage
B. Microsoft SQL Server Integration Services (SSIS)
C. Azure Table Storage
D. Azure Data Lake
E. Azure Data Factory

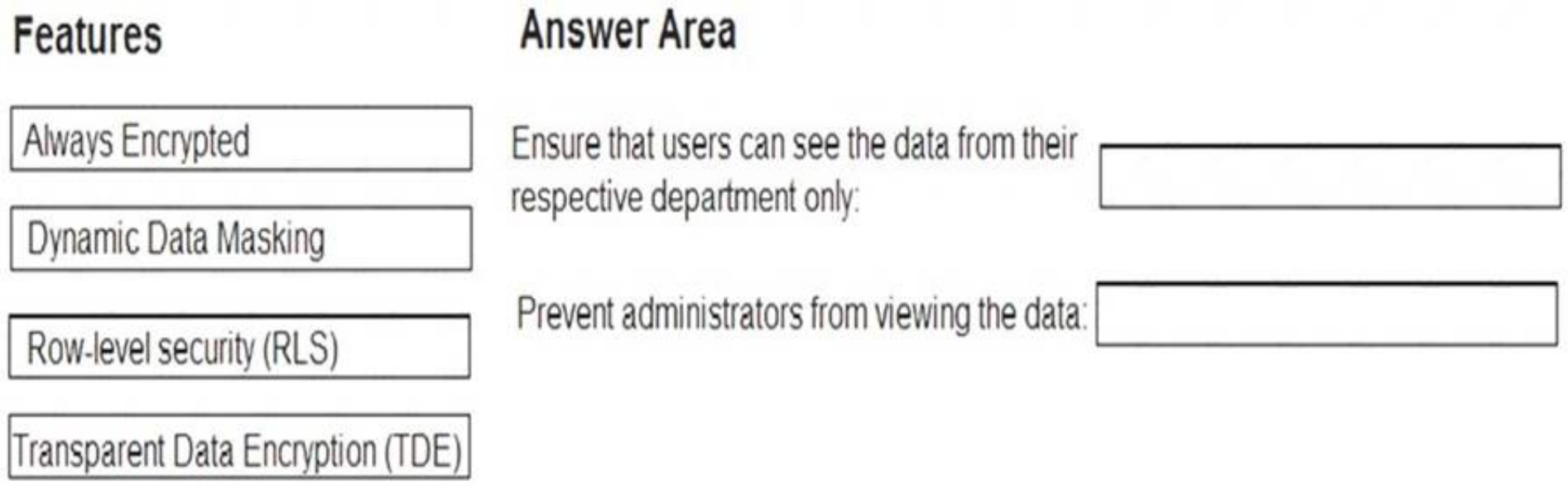
Answer: DE

Explanation: D: To analyze data in HDInsight cluster, you can store the data either in Azure Storage, Azure Data Lake Storage Gen 1/Azure Data Lake Storage Gen 2, or both. Both storage options enable you to safely delete HDInsight clusters that are used for computation without losing user data.
E: The Spark activity in a Data Factory pipeline executes a Spark program on your own or on-demand HDInsight cluster. It handles data transformation and the supported transformation activities.
References:
https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-hadoop-use-data-lake-store https://docs.microsoft.com/en-us/azure/data-factory/transform-data-using-spark

NEW QUESTION 141
You need to implement a security solution for Microsoft Azure SQL database. The solution must meet the following requirements:

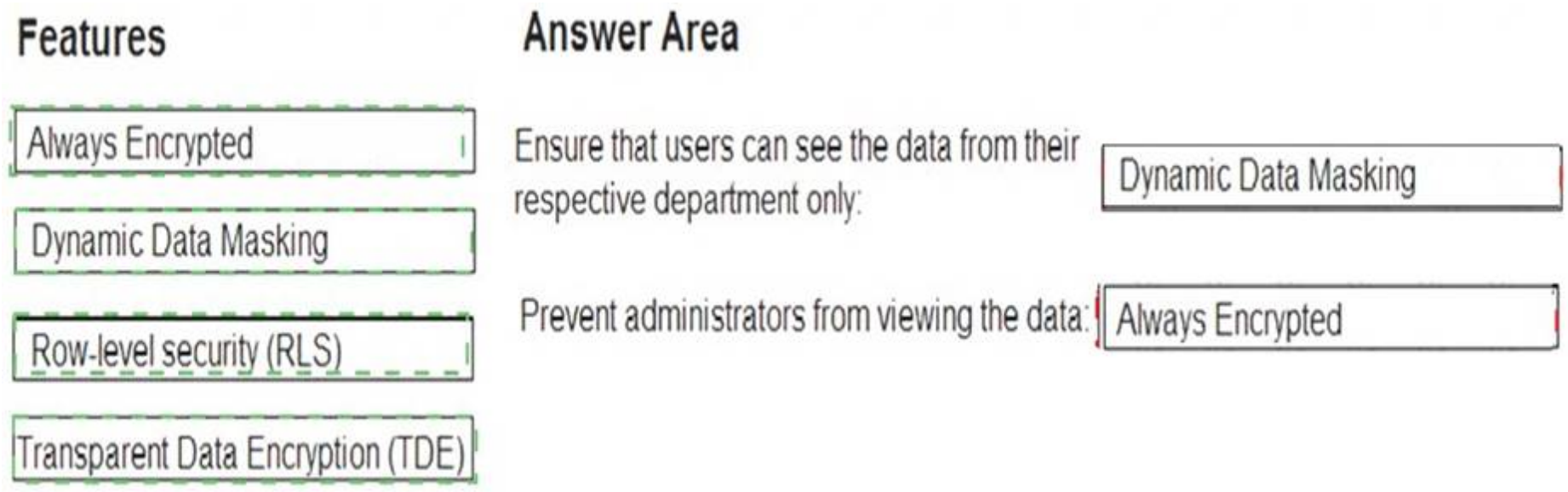
- Ensure that users can see the data from their respective department only.
- Prevent administrators from viewing the data.

Which feature should you use for each requirement? To answer, drag the appropriate features to the correct requirements. Each feature 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.

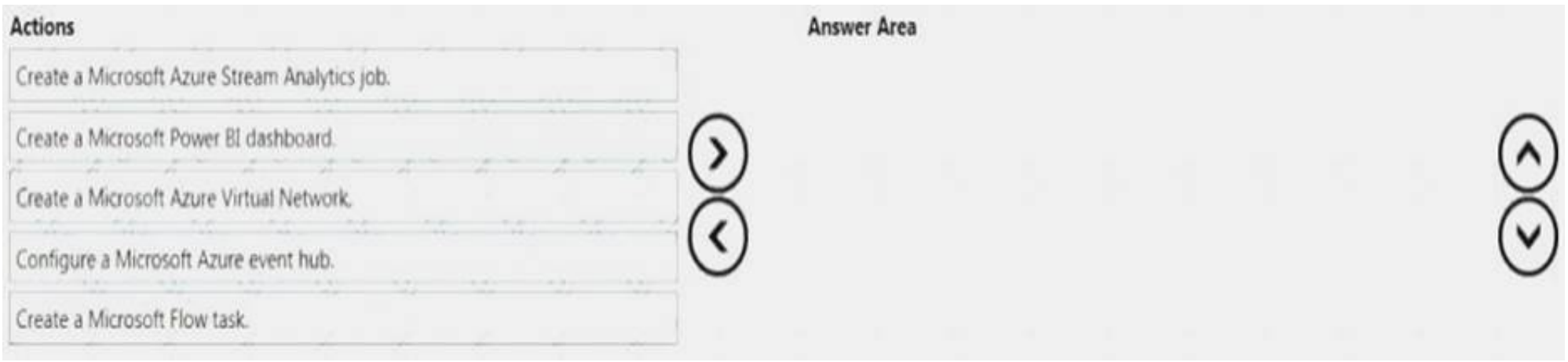


Answer:

Explanation:



NEW QUESTION 145
You are creating a retail analytics system for a company that manufactures equipment.
The company manufactures thousands of loT devices that report their status over the Internet
You need to recommend a solution to visualize notifications from the devices on a mobile-ready dashboard. Which three actions should you recommend be performed in sequence? To answer, move the appropriate
actions from the list of actions to the answer area and arrange them in the correct order.



Answer:

Explanation: References: https://docs.microsoft.com/en-us/azure/iot-hub/iot-hub-live-data-visualization-in-power-bi

NEW QUESTION 146
You have four on-premises Microsoft SQL Server data sources as described in the following table.

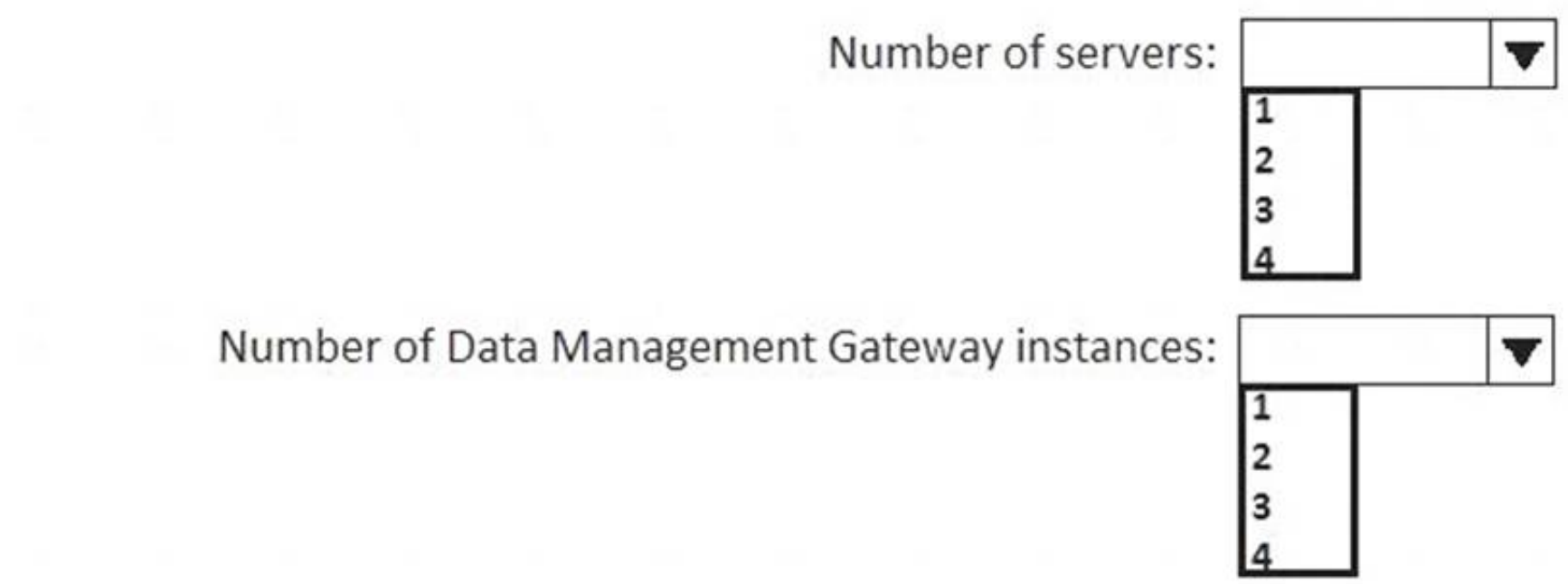
Data source name	Server name
DS1	SQL1
DS2	SQL2
DS3	SQL3
DS4	SQL4

You plan to create three Azure data factories that will interact with the data sources as described in the following table.

Data factory name	Data source used
ADF1	DS1 and DS2
ADF2	DS3
ADF3	DS4

You need to deploy Microsoft Data Management Gateway to support the Azure Data Factory deployment. The solution must use new servers to host the instances of Data Management Gateway.
What is the minimum number of new servers and data management gateways you should you deploy? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Answer Area



Answer:

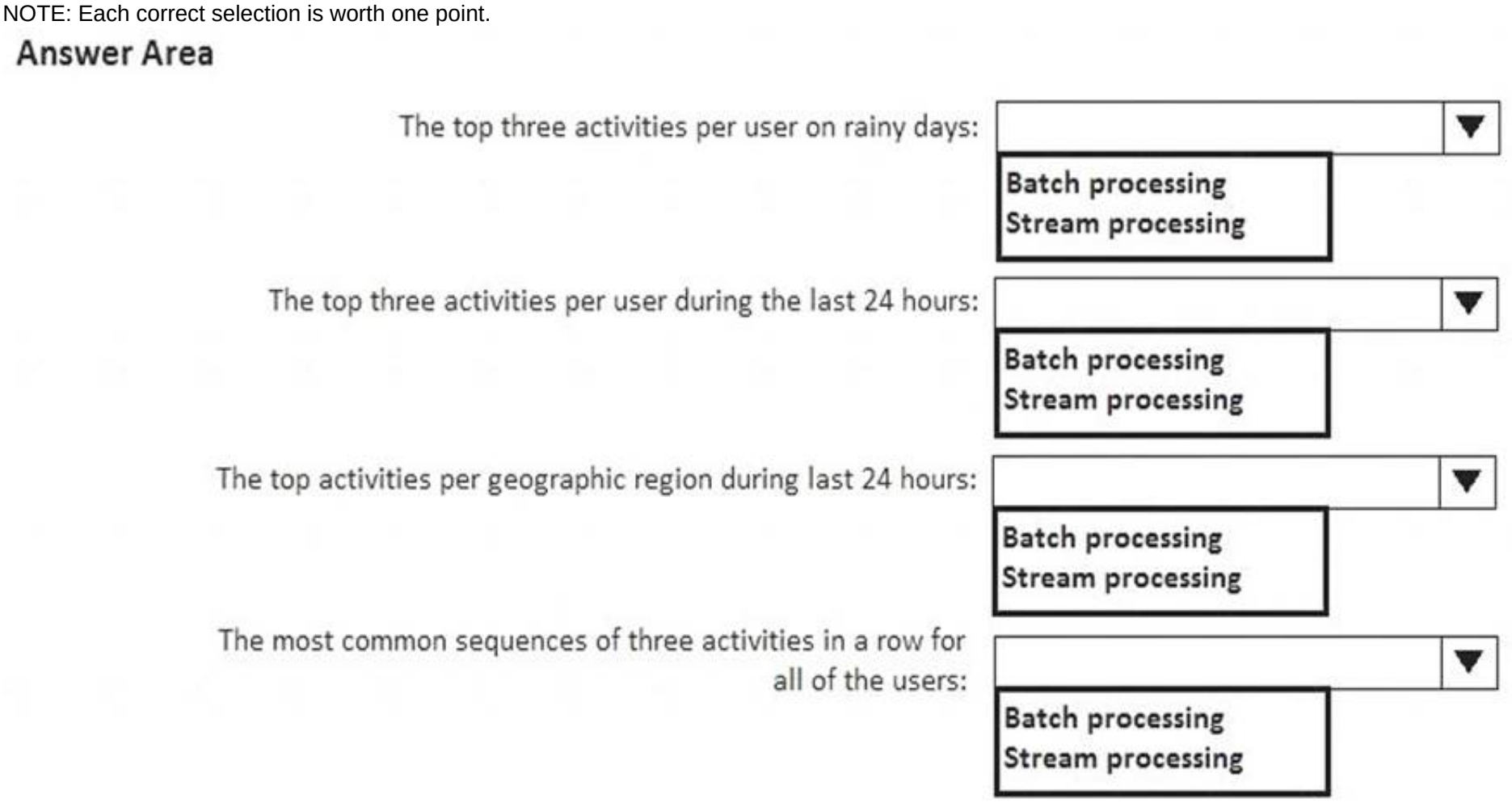
Explanation: Box 1: 3
Box 2: 3
Considerations for using gateway

NEW QUESTION 147
You extend the dashboard of the health tracking application to summarize fields across several users. You need to recommend a file format for the activity data in Azure that meets the technical requirements.
What is the best recommendation to achieve the goal? More than one answer choice may achieve the goal. Select the BEST answer.

A. ORC
B. TSV
C. CSV
D. JSON
E. XML

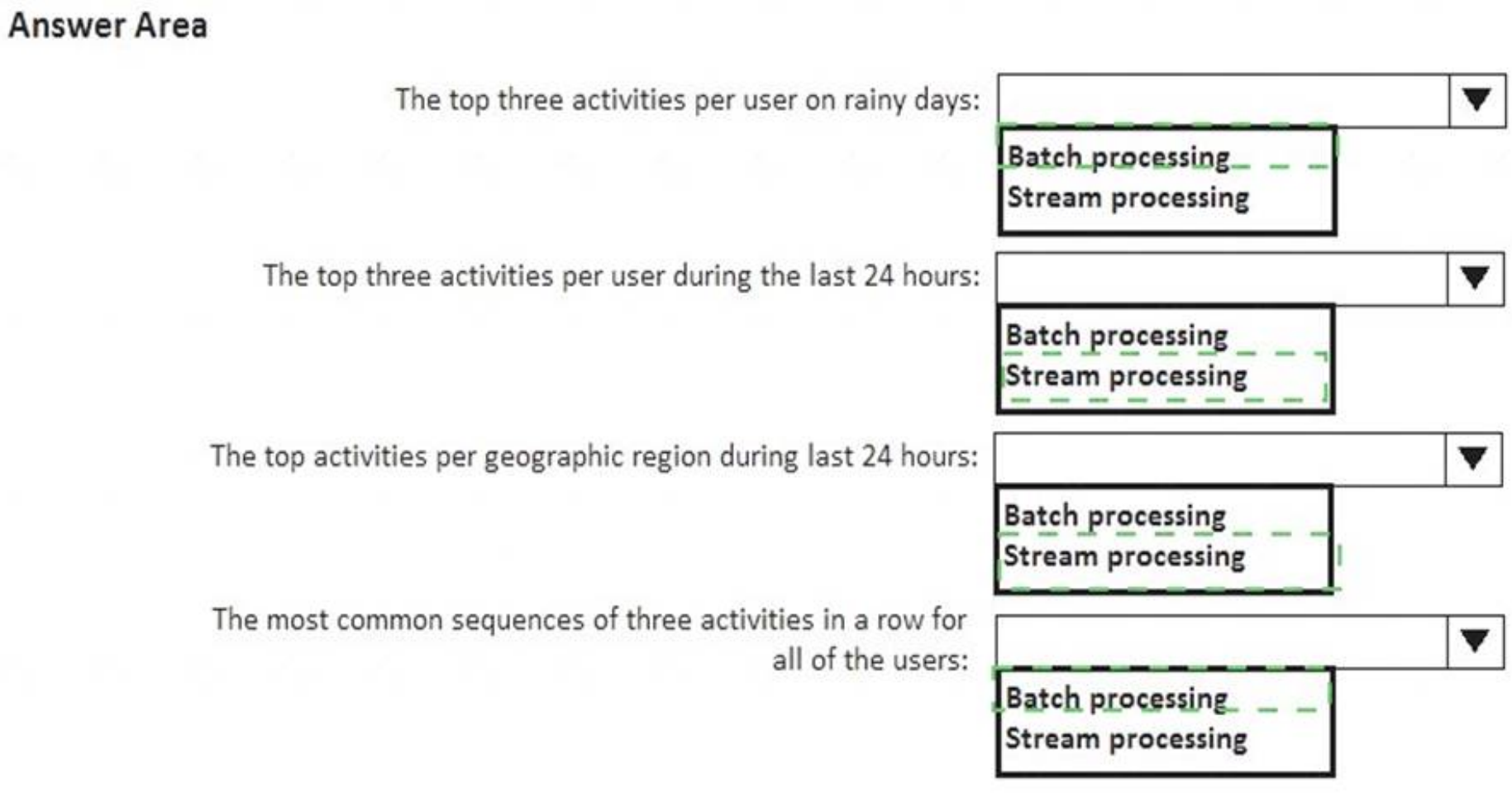
Answer: E

NEW QUESTION 151
The health tracking application uses the features of a live dashboard to provide historical and trending data based on the users activities.
You need to recommend which processing model must be used to process the following types of data: The top three activities per user on rainy days
The top three activities per user during the last 24 hours
The top activities per geographic region during last 24 hours
The most common sequences of three activities in a row for all of the users
Which processing model should you recommend for each date type? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.



Answer:

Explanation:



NEW QUESTION 155
.....

Thank You for Trying Our Product

* 100% Pass or Money Back

All our products come with a 90-day Money Back Guarantee.

* One year free update

You can enjoy free update one year. 24x7 online support.

* Trusted by Millions

We currently serve more than 30,000,000 customers.

* Shop Securely

All transactions are protected by VeriSign!

100% Pass Your 70-475 Exam with Our Prep Materials Via below:

https://www.certleader.com/70-475-dumps.html