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.

Event:

Activity Run Finished
Activity Run Started
On-Demand HDI Cluster Create Start
On-Demand HDI Cluster Created Successfully
On-Demand HDI Cluster Deleted

Status:

Failed
Succeeded

Substatus:

--
Abandoned
Failed Execution
Failed Resource Allocation
Failed Validation
Timed Out

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 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.

Actions

In DB1, create external objects.

From the Azure portal, export the storage account key.

In DB1, create a stored procedure that imports data from an external table to Table1.

From the Azure portal, create and schedule an Azure Automation job that executes the stored procedure.

In DB1, create a staging table.

Answer Area

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 3
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-AzureRmAutomationJob
B. Set-AzureRmDataFactorySliceStatus
C. Resume-AzureRmDataFactoryPipeline
D. Resume-AzureRmAutomationJob

Answer: B

Explanation: Use some PowerShell to inspect the ADF activity for the missing file error. Then simply set the dataset slice to either skipped or ready using the cmdlet to override the status.
For example:
Set-AzureRmDataFactorySliceStatus `
-ResourceGroupName $ResourceGroup `
-DataFactoryName $ADFName.DataFactoryName `
-DatasetName $Dataset.OutputDatasets `
-StartDateTime $Dataset.WindowStart `
-EndDateTime $Dataset.WindowEnd `
-Status "Ready" `
-UpdateType "Individual" References:
https://stackoverflow.com/questions/42723269/azure-data-factory-pipelines-are-failing-when-no-files-available-

NEW QUESTION 4
You have a Microsoft Azure Machine Learning Solution that contains several Azure Data Factory pipeline jobs.
You discover that the jobs for a dataset named CustomerSalesData fails. You resolve the issue that caused the job to fail.

You need to rerun the slices for CustomerSalesData. What should you do?

A. Run the Set-AzureRMDataFactorySliceStatus cmdlet and specify the–Status Retry parameter.
B. Run the Set-AzureRMDataFactorySliceStatus cmdlet and specify the–Status PendingExecution parameter.
C. Run the Resume-AzureRMDataFactoryPipeline cmdlet and specify the–Status Retry parameter.
D. Run the Resume-AzureRMDataFactoryPipeline cmdlet and specify the–Status PendingExecution parameter.

Answer: B

NEW QUESTION 5
You need to recommend a platform architecture for a big data solution that meets the following requirements: Supports batch processing
Provides a holding area for a 3-petabyte (PB) dataset
Minimizes the development effort to implement the solution
Provides near real time relational querying across a multi-terabyte (TB) dataset
Which two platform architectures should you include in the recommendation? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

A. a Microsoft Azure SQL data warehouse
B. a Microsoft Azure HDInsight Hadoop cluster
C. a Microsoft SQL Server database
D. a Microsoft Azure HDInsight Storm cluster
E. Microsoft Azure Table Storage

Answer: AE

Explanation: A: Azure SQL Data Warehouse is a SQL-based, fully-managed, petabyte-scale cloud data warehouse. It's highly elastic, and it enables you to set up in minutes and scale capacity in seconds. Scale compute and storage independently, which allows you to burst compute for complex analytical workloads, or scale down your warehouse for archival scenarios, and pay based on what you're using instead of being locked into predefined cluster configurations—and get more cost efficiency versus traditional data warehouse solutions.
E: Use Azure Table storage to store petabytes of semi-structured data and keep costs down. Unlike many data stores—on-premises or cloud-based—Table storage lets you scale up without having to manually shard your dataset. Perform OData-based queries.

NEW QUESTION 6
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 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 AzCopy to transfer the data as text files from SQL Server to Azure Blob storage, and then you use PolyBase to run Transact-SQL statements that refresh the data warehouse database.
Does this meet the goal?

A. Yes
B. No

Answer: A

Explanation: If you need the best performance, then use PolyBase to import data into Azure SQL warehouse.
Note: Often the speed of migration is an overriding concern compared to ease of setup and maintainability, particularly when there's a large amount of data to move. Optimizing purely for speed, a source controlled differentiated approach relying on bcp to export data to files, efficiently moving the files to Azure Blob storage, and using the Polybase engine to import from blob storage works best.
References: https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-migrate-data

NEW QUESTION 7
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 8
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.
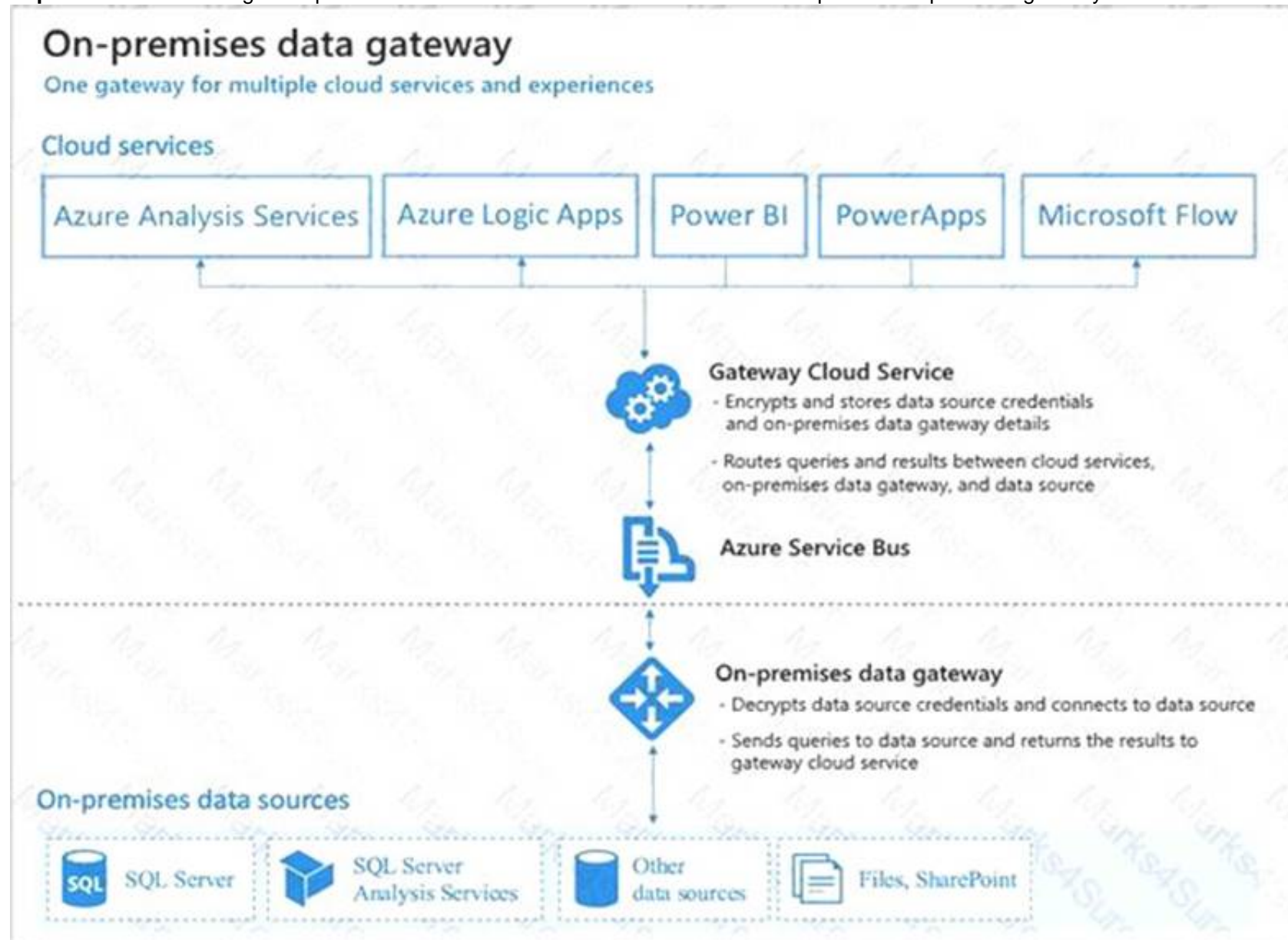
NEW QUESTION 9
You are building an Azure Analysis Services cube.
The source data for the cube is located on premises in a Microsoft SQL Server database. You need to ensure that the Azure Analysis Services service can access the source data. What should you deploy to your Azure subscription?

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 10
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 10
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 13
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 a stored procedure.
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 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 have a Microsoft Azure subscription that includes Azure Data Lake and Cognitive Services. An administrator plans to deploy an Azure Data Factory.
You need to ensure that the administrator can create the data factory. Solution: You add the user to the Owner role.
Does this meet the goal?

A. Yes
B. No

Answer: B

NEW QUESTION 18
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 22
You deploy a Microsoft Azure SQL database.
You create a job to upload customer data to the database.
You discover that the job cannot connect to the database and fails. You verify that the database runs successfully in Azure.
You need to run the job successfully. What should you create?

A. a virtual network rule
B. a network security group (NSG)
C. a firewall rule
D. a virtual network

Answer: C

Explanation: If the application persistently fails to connect to Azure SQL Database, it usually indicates an issue with one of the following:
Firewall configuration. The Azure SQL database or client-side firewall is blocking connections to Azure SQL Database.
Network reconfiguration on the client side: for example, a new IP address or a proxy server.
User error: for example, mistyped connection parameters, such as the server name in the connection string. References:
https://docs.microsoft.com/en-us/azure/sql-database/sql-database-troubleshoot-common-connection-issues

NEW QUESTION 27
You plan to analyze the execution logs of a pipeline to identify failures by using Microsoft power BI. You need to automate the collection of monitoring data for the planned analysis.
What should you do from Microsoft Azure?

A. Create a Data Factory Set
B. Save a Data Factory Log
C. Add a Log Profile
D. Create an Alert Rule Email

Answer: A

Explanation: You can import the results of a Log Analytics log search into a Power BI dataset so you can take advantage of its features such as combining data from different sources and sharing reports on the web and mobile devices.
To import data from a Log Analytics workspace into Power BI, you create a dataset in Power BI based on a log search query in Log Analytics. The query is run each time the dataset is refreshed. You can then build Power BI reports that use data from the dataset.
References: https://docs.microsoft.com/en-us/azure/azure-monitor/platform/powerbi

NEW QUESTION 28
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 DHInsight.
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 32
You have a Microsoft Azure data factory.
You assign administrative roles to the users in the following table.

User name	Role
User1	Contributor
User2	Administrator
User3	Reader
User4	Automation Operator
User5	Owner

You discover that several new data factory instances were created.
You need to ensure that only User5 can create a new data factory instance.
Which two roles should you change? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

A. User2 to Reader
B. User3 to Contributor
C. User1 to Reader

D. User4 to Contributor
E. User5 to Administrator

Answer: AC

NEW QUESTION 35
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 38
You are developing a solution to ingest data in real-time from manufacturing sensors. The data will be archived. The archived data might be monitored after it is written.
You need to recommend a solution to ingest and archive the sensor data. The solution must allow alerts to be sent to specific users as the data is ingested.
What should you include in the recommendation?

A. a Microsoft Azure notification hub and an Azure function
B. a Microsoft Azure notification hub an Azure logic app
C. a Microsoft Azure Stream Analytics job that outputs data to an Apache Storm cluster in AzureHDInsight
D. a Microsoft Azure Stream Analytics job that outputs data to Azure Cosmos DB

Answer: C

NEW QUESTION 43
Your company has a Microsoft Azure environment that contains an Azure HDInsight Hadoop cluster and an Azure SQL data warehouse. The Hadoop cluster contains text files that are formatted by using UTF-8 character encoding.
You need to implement a solution to ingest the data to the SQL data warehouse from the Hadoop cluster. The solution must provide optimal read performance for the data after ingestion.
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: SQL Data Warehouse supports loading data from HDInsight via PolyBase. The process is the same as loading data from Azure Blob Storage - using PolyBase to connect to HDInsight to load data.
Use PolyBase and T-SQL Summary of loading process: Recommendations
Create statistics on newly loaded data. Azure SQL Data Warehouse does not yet support auto create or auto update statistics. In order to get the best performance from your queries, it's important to create statistics on all columns of all tables after the first load or any substantial changes occur in the data.

NEW QUESTION 44
You use Microsoft Azure Data Factory to orchestrate data movement and data transformation within Azure. You need to identify which data processing failures exceed a specific threshold. What are two possible ways to achieve the goal? Each correct answer presents a complete solution. NOTE: Each correct selection is

worth one point.

A. View the Diagram tile on the Data Factory blade of the Azure portal.
B. Set up an alert to send an email message when the number of failed validations is greater than the threshold.
C. View the data factory metrics on the Data Factory blade of the Azure portal.
D. Set up an alert to send an email message when the number of failed slices is greater than or equal to the threshold.

Answer: A

NEW QUESTION 49
A company named Fabrikam, Inc. plans to monitor financial markets and social networks, and then to correlate global stock movements to social network activity.
You need to recommend a Microsoft Azure HDInsight cluster solution that meets the following requirements: ▶ Provides continuous availability

▶ Can process asynchronous feeds
What is the best type of cluster to recommend to achieve the goal? More than one answer choice may achieve the goal. Select the BEST answer.

A. Apache Hbase
B. Apache Hadoop
C. Apache Spark
D. Apache Storm

Answer: C

NEW QUESTION 54
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 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 Area

Move the data from DW1 to DB1: ▼
| Azure Data Factory |
| Azure Stream Analytics |
| Microsoft SQL Server Integration Services |

Import the data to DB1: ▼
| The BULK INSERT statement |
| PolyBase |
| Lazy Loading |

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 64

Your company has thousands of Internet-connected sensors.
You need to recommend a computing solution to perform a real-time analysis of the data generated by the sensors.
Which computing solution should you include in the recommendation?

A. Microsoft Azure Stream Analytics
B. Microsoft Azure Notification Hubs
C. Microsoft Azure Cognitive Services
D. a Microsoft Azure HDInsight HBase cluster

Answer: D

Explanation: HDInsight HBase is offered as a managed cluster that is integrated into the Azure environment. The clusters are configured to store data directly in Azure Storage or Azure Data Lake Store, which provides low latency and increased elasticity in performance and cost choices. This enables customers to build interactive websites
that work with large datasets, to build services that store sensor and telemetry data from millions of end points, and to analyze this data with Hadoop jobs. HBase and Hadoop are good starting points for big data project in Azure; in particular, they can enable real-time applications to work with large datasets.

NEW QUESTION 65
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 68
Your company builds hardware devices that contain sensors. You need to recommend a solution to process the sensor data and. What should you include in the recommendation?

A. Microsoft Azure Event Hubs
B. API apps in Microsoft Azure App Service
C. Microsoft Azure Notification Hubs
D. Microsoft Azure IoT Hub

Answer: A

NEW QUESTION 69
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 74
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 Python.

A. Yes
B. No

Answer: B

NEW QUESTION 75
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.

Storage Solutions

| Azure Data Lake |
| Azure Blob Storage |
| Azure Table Storage |
| Azure Service Bus Queue |
| Azure Cosmos DB |

Answer Area

Exploratory analytics [ ]

Unstructured JSON data [ ]

Pictures [ ]

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 77
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.

Workflow Components	Answer Area	
Copy	The data must be converted into a flat data strucutre by using a C# MapReduce job:	Workflow Component
HDInsightHive	The data must be moved to an Azure SQL database, which will then be used to visualize the data:	Workflow Component
HDInsightMapReduce	Additioanal stored procedures must run against the data once the data is in the database:	Workflow Component
HDInsightStreaming		
SQLServerStoredProcedure		

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 79
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 Area

The INSERT statement [answer choice].
- moves the contents of UserVisits to the UserVisitsOrc table directory
- inserts data into the UserVisitsOrc table by running a YARN application
- inserts data into the UserVisitsOrc table record by record from the Hive command-line interface (CLI)

The UserVisits table type is [answer choice].
- dataset
- external
- managed

Answer:

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

NEW QUESTION 82
You have structured data that resides in Microsoft Azure Blob Storage.
You need to perform a rapid interactive analysis of the data and to generate visualizations of the data.
What is the best type of Azure HDInsight cluster to use to achieve the goal? More than one answer choice may achieve the goal. Select the BEST answer.

A. Apache Storm

B. Apache HBase
C. Apache Hadoop
D. Apache Spark

Answer: D

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

NEW QUESTION 86
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 90
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 91
Your company has several thousand sensors deployed.

You have a Microsoft Azure Stream Analytics job that receives two data streams Input1 and Input2 from an Azure event hub. The data streams are portioned by using a column named SensorName. Each sensor is identified by a field named SensorID.
You discover that Input2 is empty occasionally and the data from Input1 is ignored during the processing of the Stream Analytics job.
You need to ensure that the Stream Analytics job always processes the data from Input1.
How should you modify the query? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

```
SELECT I1.SensorID, I1,EntryTime, I2.ExitTime, [DATEDIFF
(minute, I1.EntryTIme, I2.ExitTime)] as Duration
FROM Input I1 TIMESTAMP BY EntryTime PARTITION BY SensorName
```

```
                                    ▼    Input I2 TIMESTAMP BY ExitTime PARTITION BY SensorName

    JOIN
    LEFT OUTER JOIN
    RIGHT OUTER JOIN


    JOIN table1 R
    ON R.SensorID = I1.SensorID
    AND [DATEDIFF(minute, I1, I2) BETWEEN 0 AND 10]
    WHERE R.IsValid = 1
```

```
                                                                ▼
    ON I1.SensorId=I2.SensorID
    ON I1.SensorName=I2.SensorName
    ON I1.SensorId=I2.SensorId OR I1.SensorName=I2.SensorName
```

Answer:

Explanation: Box 1: LEFT OUTER JOIN
LEFT OUTER JOIN specifies that all rows from the left table not meeting the join condition are included in the result set, and output columns from the other table are set to NULL in addition to all rows returned by the inner join.
Box 2: ON I1.SensorID= I2.SensorID
References: https://docs.microsoft.com/en-us/stream-analytics-query/join-azure-stream-analytics

NEW QUESTION 95
You are using a Microsoft Azure Data Factory pipeline to copy data to an Azure SQL database. You need to prevent the insertion of duplicate data for a given dataset slice.
Which two actions should you perform? Each correct answer presents part of the solution. NOTE: Each correct selection is worth one point.

A. Set the External property to true.
B. Add a column named SliceIdentifierColumnName to the output dataset.
C. Set the SqlWriterCleanupScript property to true.
D. Remove the duplicates in post-processing.
E. Manually delete the duplicate data before running the pipeline activity.

Answer: BC

NEW QUESTION 97
You are building a streaming data analysis solution that will process approximately 1 TB of data weekly. You plan to use Microsoft Azure Stream Analytics to create alerts on real-time data. The data must be preserved for deeper analysis at a later date.
You need to recommend a storage solution for the alert data. The solution must meet the following requirements:

▶ Support scaling up without any downtime

▶ Minimize data storage costs.
What should you recommend using to store the data?

A. Azure Data Lake
B. Azure SQL Database
C. Azure SQL Data Warehouse
D. Apache Kafka

Answer: A

NEW QUESTION 99
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 101
You have an analytics solution in Microsoft Azure that must be operationalized.
You have the relevant data in Azure Blob storage. You use an Azure HDInsight Cluster to process the data. You plan to process the raw data files by using Azure HDInsight. Azure Data Factory will operationalize the solution.
You need to create a data factory to orchestrate the data movement. Output data must be written back to Azure Blob storage.
Which four 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.

Actions	Answer Area
Create input and output datasets for the files in Azure.	1
Rename the data factory hub.	2
Create a data factory.	3
Create a client to connect to Azure Blob storage.	4
Create linked services for Azure Blob storage and the Azure HDInsight cluster.	
Create an Azure HDInsight activity in a pipeline to process the data.	

Answer:

Explanation:

NEW QUESTION 102
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 106
You have a web application that generates several terabytes (TB) of financial documents each day. The application processes the documents in batches.
You need to store the documents in Microsoft Azure. The solution must ensure that a user can restore the previous version of a document.
Which type of storage should you use for the documents?

A. Azure Cosmos DB
B. Azure File Storage
C. Azure Data Lake
D. Azure Blob storage

Answer: A

NEW QUESTION 111
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 115
You need to recommend a data analysis solution for 20,000 Internet of Things (IoT) devices. The solution must meet the following requirements:
• Each device must be identified by using its own credentials.
• Each device must be able to route data to multiple endpoints.
• The solution must require the minimum amount of customized code. What should you recommend?

A. Microsoft Azure Notification Hubs
B. Microsoft Azure IoT Hub
C. Microsoft Azure Service Bus
D. Microsoft Azure Event Hubs

Answer: D

NEW QUESTION 118
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 120
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 124
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

Number of servers:

1
2
3
4

Number of Data Management Gateway instances:

1
2
3
4

Answer:

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

NEW QUESTION 126
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 130
Users report that when they access data that is more than one year old from a dashboard, the response time is slow.
You need to resolve the issue that causes the slow response when visualizing older data. What should you do?

A. Process the event hub data first, and then process the older data on demand.
B. Process the older data on demand first, and then process the event hub data.
C. Aggregate the older data by time, and then save the aggregated data to reference data streams.
D. Store all of the data from the event hub in a single partition.

Answer: C

NEW QUESTION 131
You need to recommend a permanent Azure Storage solution for the activity data. The solution must meet 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. Azure SQL Database
B. Azure Queue storage
C. Azure Blob storage
D. Azure Event Hubs

Answer: A

NEW QUESTION 134
You need to recommend a data transfer solution to support the business goals.
What should you recommend?

A. Configure the health tracking application to cache data locally for 24 hours.
B. Configure the health tracking application to Aggregate activities in blocks of 128 KB.
C. Configure the health tracking application to cache data locally tor 12 hours.
D. Configure the health tracking application to aggregate activities in blocks of 64 KB.

Answer: D

NEW QUESTION 136
You need to recommend a data handling solution to support the planned changes to the dashboard. The solution must meet the privacy requirements.
What is the best recommendation to achieve the goal? More than one answer choice may achieve the goal. Select the BEST answer.

A. anonymization
B. encryption
C. obfuscation
D. compression

Answer: C

NEW QUESTION 137
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 Area

The top three activities per user on rainy days:

| ▼ |
| Batch processing |
| Stream processing |

The top three activities per user during the last 24 hours:

| ▼ |
| Batch processing |
| Stream processing |

The top activities per geographic region during last 24 hours:

| ▼ |
| Batch processing |
| Stream processing |

The most common sequences of three activities in a row for all of the users:

| ▼ |
| Batch processing |
| Stream processing |

Answer:

Explanation: Answer Area

The top three activities per user on rainy days:

| ▼ |
| Batch processing |
| Stream processing |

The top three activities per user during the last 24 hours:

| ▼ |
| Batch processing |
| Stream processing |

The top activities per geographic region during last 24 hours:

| ▼ |
| Batch processing |
| Stream processing |

The most common sequences of three activities in a row for all of the users:

| ▼ |
| Batch processing |
| Stream processing |

NEW QUESTION 139
......

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