Exam Questions 70-475

Designing and Implementing Big Data Analytics Solutions

https://www.2passeasy.com/dumps/70-475/

NEW QUESTION 1
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 2
You have a Microsoft Azure Data Factory pipeline that contains an input dataset.
You need to ensure that the data from Azure Table Storage is copied only if the table contains 1,000 records or more.
Which policy setting should you use in JSON?

```
A.    "policy":
      {
         "validation":
         {
            "minimum":1000
         }
      }

B.    "policy":
      {
         "validation":
         {
            "sqlRows":1000
         }
      }

C.    "policy":
      {
         "validation":
         {
            "minimumRows":1000
         }
      }

D.    "policy":
      {
         "validation":
         {
            "sqlMinimumRows":1000
         }
      }
```

A. Option A
B. Option B
C. Option C
D. Option D

Answer: B

Explanation: The following JSON defines a Linux-based on-demand HDInsight linked service. The Data Factory service automatically creates a Linux-based HDInsight cluster to process the required activity.

{
"name": "HDInsightOnDemandLinkedService", "properties": {
"type": "HDInsightOnDemand", "typeProperties": { "clusterType": "hadoop", "clusterSize": 1,
"timeToLive": "00:15:00", "hostSubscriptionId": "<subscription ID>", "servicePrincipalId": "<service principal ID>", "servicePrincipalKey": {
"value": "<service principal key>", "type": "SecureString"
},
"tenant": "<tenent id>",
"clusterResourceGroup": "<resource group name>", "version": "3.6",
"osType": "Linux", "linkedServiceName": {
"referenceName": "AzureStorageLinkedService", "type": "LinkedServiceReference"
}
},
"connectVia": {
"referenceName": "<name of Integration Runtime>", "type": "IntegrationRuntimeReference"
}
}
}
References: https://docs.microsoft.com/en-us/azure/data-factory/compute-linked-services

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 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 5
Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while 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 6
The settings used for slice processing are described in the following table.

Setting	Value
Retry	2
LongRetry	2
Timeout	00:10:00
Concurrency	5

If the slice processing fails, you need to identify the number of retries that will be performed before the slice execution status changes to failed.
How many retries should you identify?

A. 2

B. 3
C. 5
D. 6

Answer: C

NEW QUESTION 7
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.

On-premises data gateway
One gateway for multiple cloud services and experiences

Cloud services

| Azure Analysis Services | Azure Logic Apps | Power BI | PowerApps | Microsoft Flow |

Gateway Cloud Service
- Encrypts and stores data source credentials and on-premises data gateway details
- Routes queries and results between cloud services, on-premises data gateway, and data source

Azure Service Bus

On-premises data gateway
- Decrypts data source credentials and connects to data source
- Sends queries to data source and returns the results to gateway cloud service

On-premises data sources

| SQL Server | SQL Server Analysis Services | Other data sources | Files, SharePoint |

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

NEW QUESTION 8
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 9
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 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 10
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 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 17
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 20
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 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 AzCopy to transfer the data as text files from SQL Server to Azure Blob storage, and then you use Azure Data Factory to refresh the data warehouse database.
Does this meet the goal?

A. Yes
B. No

Answer: B

NEW QUESTION 24
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 25
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 26
You plan to design a solution to gather data from 5,000 sensors that are deployed to multiple machines. The sensors generate events that contain data on the health status of the machines.
You need to create a new Microsoft Azure event hub to collect the event data.
Which command should you run? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.
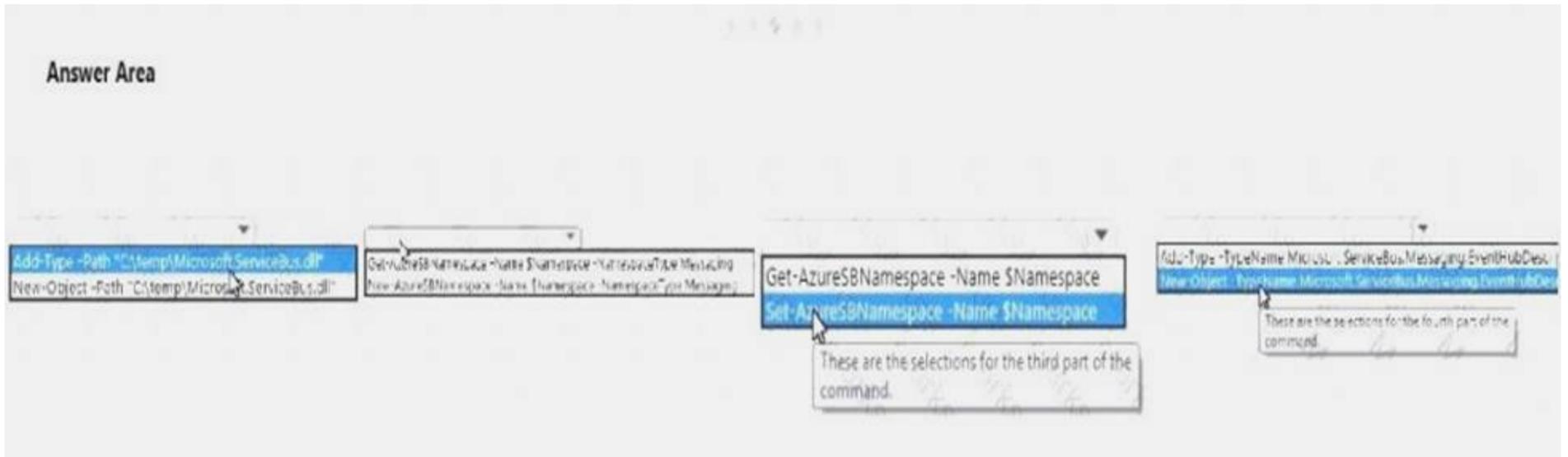
Answer:

Explanation:
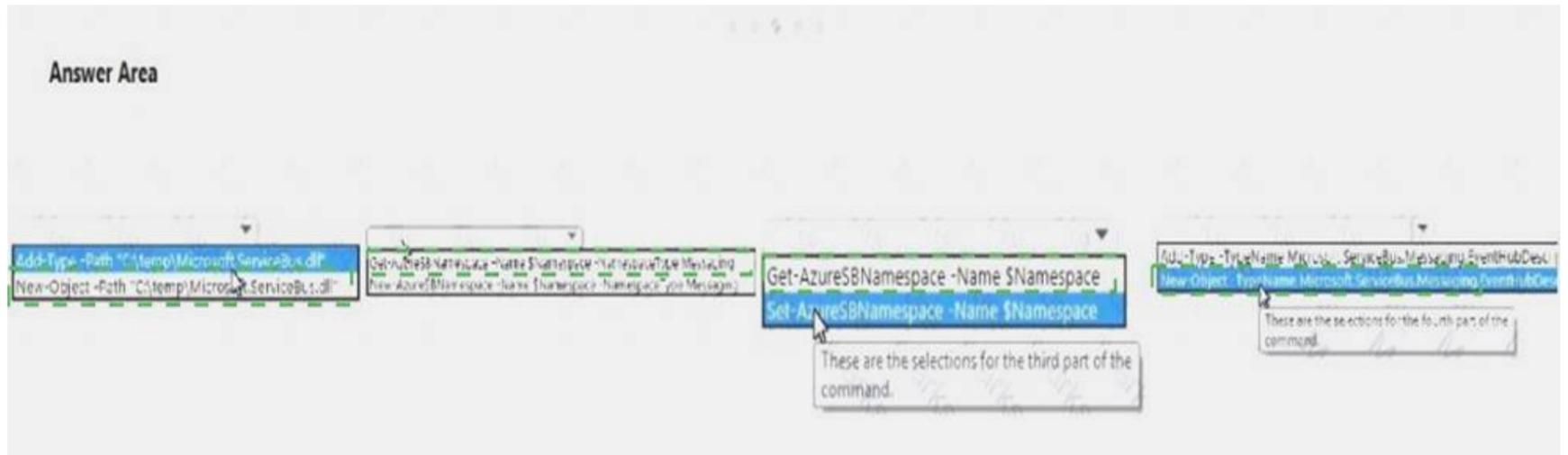
NEW QUESTION 27
A company named Fabrikam, Inc. has a Microsoft Azure web app. Billions of users visit the app daily.
The web app logs all user activity by using text files in Azure Blob storage. Each day, approximately 200 GB of text files are created.
Fabrikam uses the log files from an Apache Hadoop 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 28
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 Area:

The data ingestion process must only communicate with the batch layer:	▼ Yes / No
The query execution must communicate with both the serving layer and the speed layer:	▼ Yes / No
You can use Kafka to execute the queries:	▼ Yes / No

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.
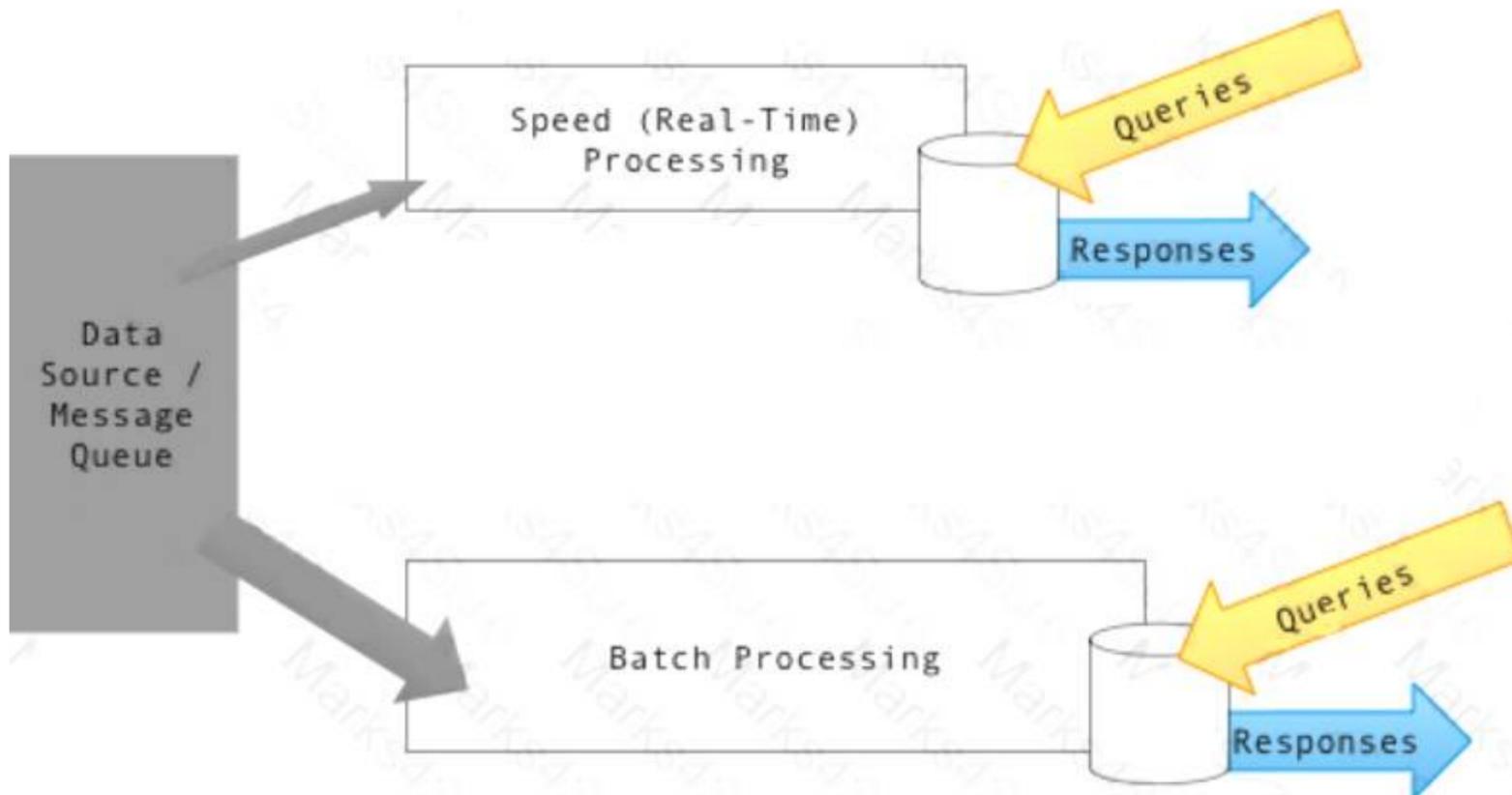
Box 3: Yes.
We are excited to announce Interactive Queries, a new feature for stream processing with Apache Kafka. Interactive Queries allows you to get more than just processing from streaming.
Note: Lambda architecture is a popular choice where you see stream data pipelines applied (speed layer). Architects 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 32
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 37
Your company supports multiple Microsoft Azure subscriptions.
You plan to deploy several virtual machines to support the services in Azure.
You need to automate the management of all the subscriptions. The solution must minimize administrative effort.
Which two cmdlets should you run? Each correct answer presents part of the solution.
NOTE: Each correct selection is worth one point.

A. Clear-AzureProfile
B. Add-AzureSubscription
C. Add-AzureRMAccount
D. Import-AzurePublishSettingsFile
E. Get-AzurePublishSettingsFile

Answer: DE

NEW QUESTION 42
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
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 47
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
You are designing an Internet of Things (IoT) solution intended to identify trends. The solution requires the
real-time 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 Azure HDInsight Spark clusters
C. Microsoft Azure Event Hubs
D. Microsoft Azure HDInsight Hadoop clusters

Answer: A

Explanation: For your Internet of Things (IoT) scenarios that use Event Hubs, Azure Stream Analytics can serve as a possible first step to perform near real-time analytics on telemetry data. Just like Event Hubs, Steam Analytics supports the streaming of millions of event per second. Unlike a standard database, analysis is performed on data in motion. This streaming input data can also be combined with reference data inputs to perform lookups or do correlation to assist in unlocking business insights. It uses a SQL-like language to simplify the analysis of data inputs and detect anomalies, trigger alerts or transform the data in order to create valuable outputs

NEW QUESTION 53
You have a Microsoft Azure Stream Analytics job that contains several pipelines.
The Stream Analytics job is configured to trigger an alert when the sale of products in specific categories exceeds a specified threshold.
You plan to change the product-to-category mappings next month to meet future business requirements.
You need to create the new product-to-category mappings to prepare for the planned change. The solution must ensure that the Stream Analytics job only uses the new product-to-category mappings when the
mappings are ready to be activated.
Which naming structure should you use for the file that contains the product-to-category mappings?

A. Use any date after the day the file becomes active.
B. Use any date before the day the categories become active.
C. Use the date and hour that the categories are to become active.
D. Use the current date and time.

Answer: C

NEW QUESTION 55
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 57
You manage a Microsoft Azure HDInsight Hadoop cluster. All of the data for the cluster is stored in Azure Premium Storage.
You need to prevent all users from accessing the data directly. The solution must allow only the HDInsight service to access the data.
Which five 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

| From the Azure portal, create a copy of the storage account key. |
| From the Azure portal, create a stored access policy. |
| From the HDInsight Hadoop cluster, restart all of the Hadoop services. |
| From the HDInsight Hadoop cluster, modify the properties of the custom core-site. |
| From the HDInsight Hadoop cluster, enable maintenance mode. |
| From the Azure portal, create a copy of the shared access signature (SAS) token. |

Answer Area
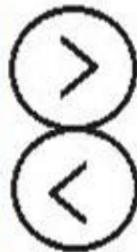
Answer:

Explanation: 1. Create Shared Access Signature policy2. Save the SAS policy token, storage account name, and container name. These values are used when associating the storage account with your HDInsight cluster.3. Update property of core-site4. Maintenance mode5. Restart all serviceshttps://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-storage-sharedaccesssignature-permissions

NEW QUESTION 59
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

| An Azure SQL Database External Metastore |
| An Azure Table Storage External Metastore |
| An Azure virtual network |
| Script Actions |

Answer Area

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

Answer:

Explanation:

Technologies

| An Azure SQL Database External Metastore |
| An Azure Table Storage External Metastore |
| An Azure virtual network |
| Script Actions |

Answer Area

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

NEW QUESTION 60
You have an Apache Storm cluster.
You need to ingest data from a Kafka queue.
Which component should you use to consume data emitted from Kaka?

A. Flume
B. a bolt
C. a spout
D. a Microsoft Azure Service Bus queue

Answer: C

Explanation: To perform real-time computation on Storm, we create "topologies." A topology is a graph of a computation, containing a network of nodes called "Spouts" and "Bolts." In a Storm topology, a Spout is the source of data streams and a Bolt holds the business logic for analyzing and processing those streams. The org.apache.storm.kafka.KafkaSpout component reads data from Kafka. Example:

Storm-Kafka Pipelines

e.g. High Speed Filter Topology

References:
https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-apache-storm-with-kafka https://hortonworks.com/blog/storm-kafka-together-real-time-data-refinery/
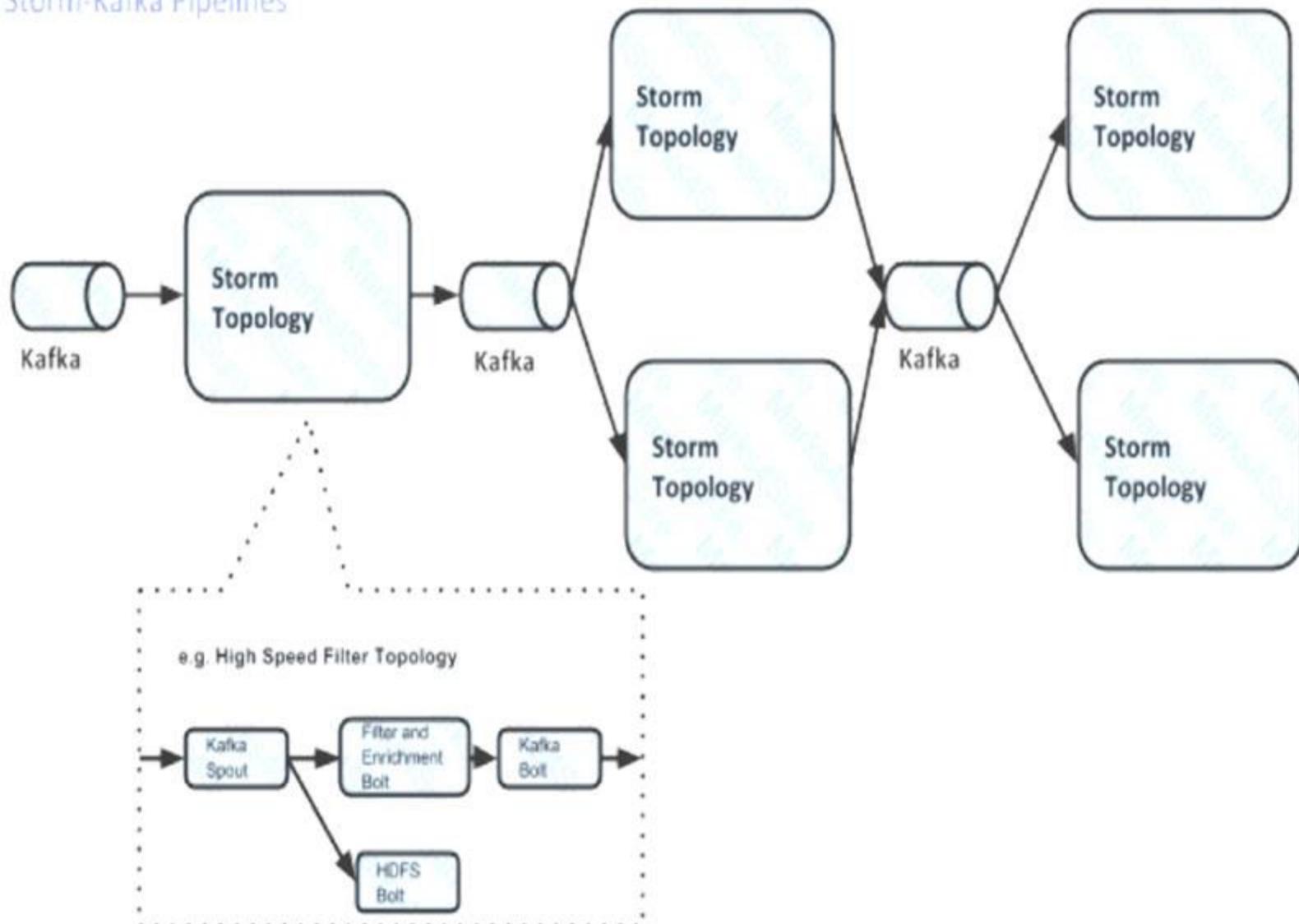
NEW QUESTION 65
Your company has thousands of Internet-connected sensors.
You need to recommend a computing solution to perform a real-time analysis of the data generated by the sensors.
Which computing solution should you include in the recommendation?

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 70
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 73
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.

Code Elements

accessKey

AccountKey1

accountName

AccountName=Account2;AccountKey1

AzureBatchLinkedService

AzureStorageLinkedService

Answer Area

:" DefaultEndpointsProtocol=https;

Answer:

Explanation:

Code Elements

accessKey

AccountKey1

accountName

AccountName=Account2;AccountKey1

AzureBatchLinkedService

AzureStorageLinkedService

Answer Area

AzureStorageLinkedService

:" DefaultEndpointsProtocol=https; AccountName=Account2;AccountKey1

NEW QUESTION 77
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 79
Your Microsoft Azure subscription contains several data sources that use the same XML schema. You plan to process the data sources in parallel.
You need to recommend a compute strategy to minimize the cost of processing the data sources. What should you recommend including in the compute strategy?

A. Microsoft SQL Server Integration Services (SSIS) on an Azure virtual machine
B. Azure Batch
C. a Linux HPC cluster in Azure
D. a Windows HPC cluster in Azure

Answer: A

NEW QUESTION 84
Your company has a data visualization solution that contains a customized Microsoft Azure Stream Analytics solution. The solution provides data to a Microsoft Power BI deployment.
Every 10 seconds, you need to query for instances that have more than three records.
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: Box 1: TumblingWindow(second, 10)
Tumbling Windows define a repeating, non-overlapping window of time. Example: Calculate the count of sensor readings per device every 10 seconds SELECT sensorId, COUNT(*) AS Count
FROM SensorReadings TIMESTAMP BY time GROUP BY sensorId, TumblingWindow(second, 10) Box 2: [Count] >= 3
Count(*) returns the number of items in a group.

NEW QUESTION 89
Your company has 2000 servers.
You plan to aggregate all of the log files from the servers in a central repository that uses Microsoft Azure HDInsight. Each log file contains approximately one million records. All of the files use the .log file name extension.
The following is a sample of the entries in the log files.
20:26:41 SampleClass3 (ERROR) verbose detail for id 1527353937
In Apache Hive, you need to create a data definition and a query capturing tire number of records that have an error level of [ERROR].
What should you do? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Answer Area

CREATE [▼] log4jLogs (t1 string, t2 string,
EXTERNAL TABLE
TABLE
VIEW

t3 string, t4 string, t5 string, t6 string, t7 string)
ROW FORMAT DELIMITED FIELDS TERMINATED BY [▼]
..
/T
[ERROR]

STORED AS TEXTFILE LOCATION 'wasbs:///example/data/';
SELECT t4 AS sev, [▼] AS count
COUNT(*)
MAX(*)
SUM(*)

FROM log4jLogs
WHERE t4 = ' [ERROR] '
AND INPUT_FILE_NAME LIKE [▼]
'!.log'
'%.log'
'*.log'

GROUP BY

Answer:

Explanation: Box 1: table
Box 2: /t
Apache Hive example:
CREATE TABLE raw (line STRING)
ROW FORMAT DELIMITED FIELDS TERMINATED BY '\t' LINES TERMINATED BY '\n';
Box 3: count(*)
Box 4: '*.log'

NEW QUESTION 91
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 95
You are planning a solution that will have multiple data files stored in Microsoft Azure Blob storage every hour. Data processing will occur once a day at midnight only.
You create an Azure data factory that has blob storage as the input source and an Azure HD Insight activity that uses the input to create an output Hive table.
You need to identify a data slicing strategy for the data factory.
What should you identify? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Answer:

Explanation:

Answer Area

The processing frequency:

Daily
Hourly
Monthly

The partitioning blob storage path for the input:

.../{Year}/{Month}
.../{Year}/{Month}/{Day}
.../{Year}/{Month}/{Day}/{Hour}
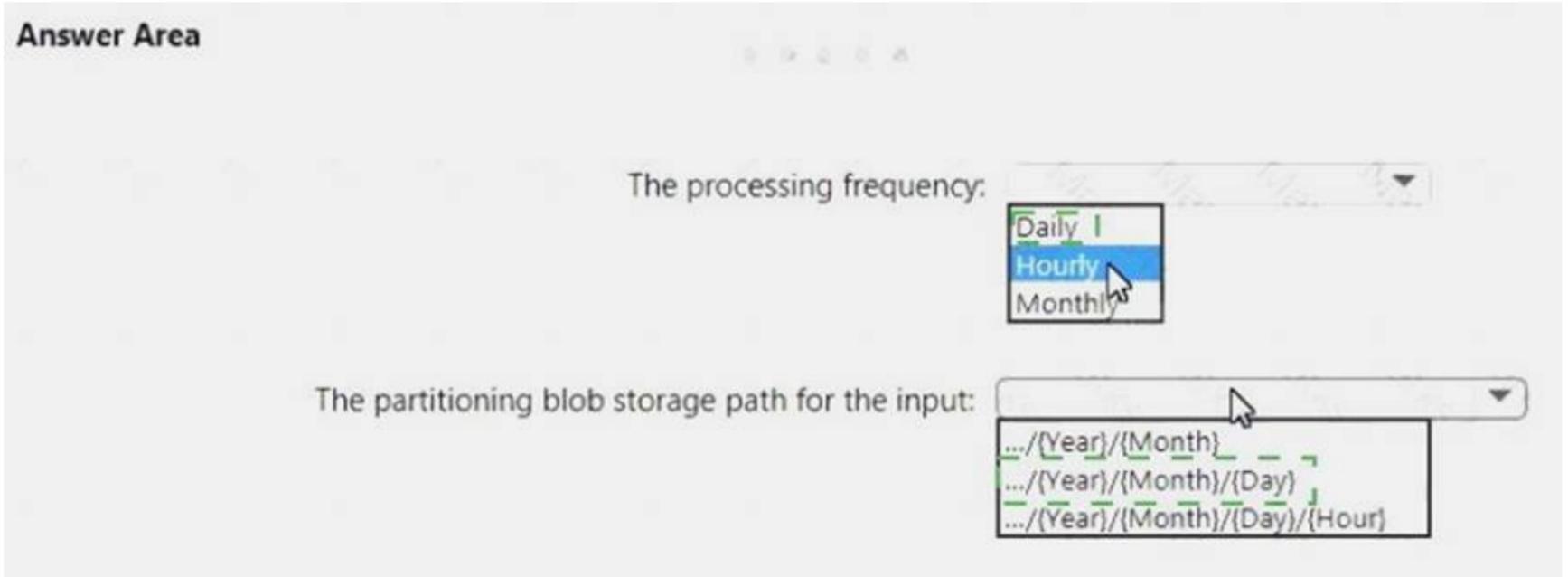
NEW QUESTION 99
You have raw data in Microsoft Azure Blob storage. Each data file is 10 KB and is the XML format. You identify the following requirements for the data:

▶ 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

Copy

HDInsightHive

HDInsightMapReduce

HDInsightStreaming

SQLServerStoredProcedure

Answer Area

The data must be converted into a flat data strucutre by using a C# MapReduce job:

Workflow Component

The data must be moved to an Azure SQL database, which will then be used to visualize the data:

Workflow Component

Additioanal stored procedures must run against the data once the data is in the database:

Workflow Component

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 102
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 Data Factory Contributor role. Does this meet the goal?

A. Yes
B. No

Answer: A

NEW QUESTION 107
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 108
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 110
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 Azure Import/Export service.
Does this meet the goal?

A. Yes
B. No

Answer: A

NEW QUESTION 114
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 Data Lake job.
Does this meet the goal?

A. Yes

B. No

Answer: B

NEW QUESTION 118
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 120
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 123
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 128
You have a financial model deployed to an application named finance1. The data from the financial model is stored in several data files.
You need to implement a batch processing architecture for the financial model. You upload the data files and finance1 to a Microsoft Azure Storage account.
Which three components should you create in sequence next? To answer, move the appropriate components from the list of components to the answer area and arrange them in the correct order.

Answer:

Explanation:

NEW QUESTION 131
You use Microsoft Azure Data Factory to orchestrate data movements and data transformations within Azure. You plan to monitor the data factory to ensure that all of the activity slices run successfully. You need to identify a solution to rerun failed slices. What should you do?

A. From the Diagram tile on the Data Factory blade of the Azure portal, double-click the pipeline that has a failed slice.
B. Move the data factory to a different resource group.
C. From the Azure portal, select the Data slice blade, and then click Run.
D. Delete and recreate the data factory.

Answer: B

NEW QUESTION 136
You need to create a new Microsoft Azure data factory by using Azure PowerShell. The data factory will have a pipeline that copies data to and from Azure Storage.
Which four cmdlets should you use in sequence? To answer, move the appropriate cmdlets from the list of cmdlets to the answer area and arrange them in the correct order.

Cmdlets	Answer Area
New-AzureRMDataFactoryLinkedService	
New-AzureDataFactoryRMDataFactoryHub	
New-AzureRMDataFactory	
New-AzureRMDataFactoryDataset	
New-AzureDataFactoryRMDataFactoryGateway	
New-AzureRMDataFactoryPipeline	

Answer:

Explanation: Perform these operations in the following order:

▶ Create a data factory.

▶ Create linked services.

▶ Create datasets.

▶ Create a pipeline.

Step 1: New-AzureRMDataFactory Create a data factory
The New-AzureRmDataFactory cmdlet creates a data factory with the specified resource group name and location.
Step 2: New-AzureRMDataFactoryLinkedService
Create linked services in a data factory to link your data stores and compute services to the data factory. The New-AzureRmDataFactoryLinkedService cmdlet links a data store or a cloud service to Azure Data
Factory.
Step 3: New-AzureRMDataFactoryDataset
You define a dataset that represents the data to copy from a source to a sink. It refers to the Azure Storage linked service you created in the previous step.
The New-AzureRmDataFactoryDataset cmdlet creates a dataset in Azure Data Factory.
Step 4: New-AzureRMDataFactoryPipeline You create a pipeline.
The New-AzureRmDataFactoryPipeline cmdlet creates a pipeline in Azure Data Factory. References:
https://docs.microsoft.com/en-us/azure/data-factory/quickstart-create-data-factory-powershell https://docs.microsoft.com/en-us/powershell/module/azurerm.datafactories/new-azurermdatafactory

NEW QUESTION 140
You work for a telecommunications company that uses Microsoft Azure Stream Analytics. You have data related to incoming calls.
You need to group the data in the following ways:

▶ Group A: Every five minutes for a duration of five minutes

▶ Group B: Every five minutes for a duration of 10 minutes

Which type of window should you use for each group? To answer, drag the appropriate window types to the correct groups. Each window type 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.

Window Types	Answer Area
Hopping	Group A: Window Type
Sliding	Group B: Window Type
Tumbling	

Answer:

Explanation: Group A: Tumbling
Tumbling Windows define a repeating, non-overlapping window of time. Group B: Hopping
Like Tumbling Windows, Hopping Windows move forward in time by a fixed period but they can overlap with one another.

NEW QUESTION 143
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 147
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 152
Your company deploys thousands of sensors.
You plan to join the data from the sensors by using Azure Data Factory. The reference data file refreshes every 30 minutes.
You need to include the path to the reference data in Data Factory. Which path should you include?

A. products/{date}/{time}/product_listjson
B. products/{sensor_name}/product_list.json
C. products/{batch}/product_listjson
D. products/{time}/product_list.json

Answer: A

NEW QUESTION 154
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 157
You are designing a solution that will use Apache HBase on Microsoft Azure HDInsight.
You need to design the row keys for the database to ensure that client traffic is directed over all of the nodes in the cluster.
What are two possible techniques that you can use? Each correct answer presents a complete solution. NOTE: Each correct selection is worth one point.

A. padding
B. trimming
C. hashing
D. salting

Answer: CD

Explanation: There are two strategies that you can use to avoid hotspotting:
* Hashing keys
To spread write and insert activity across the cluster, you can randomize sequentially generated keys by hashing the keys, inverting the byte order. Note that these strategies come with trade-offs. Hashing keys, for example, makes table scans for key subranges inefficient, since the subrange is spread across the cluster.
* Salting keys
Instead of hashing the key, you can salt the key by prepending a few bytes of the hash of the key to the actual key.
Note. Salted Apache HBase tables with pre-split is a proven effective HBase solution to provide uniform workload distribution across RegionServers and prevent hot spots during bulk writes. In this design, a row key is made with a logical key plus salt at the beginning. One way of generating salt is by calculating n (number of regions) modulo on the hash code of the logical row key (date, etc).
Reference:
https://blog.cloudera.com/blog/2015/06/how-to-scan-salted-apache-hbase-tables-with-region-specific-key-range
http://maprdocs.mapr.com/51/MapR-DB/designing_row_keys_for_mapr_db_binary_tables.html

NEW QUESTION 160
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 163
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 164
......