70-475 Q&As

Designing and Implementing Big Data Analytics Solutions

Pass Microsoft 70-475 Exam with 100% Guarantee

Free Download Real Questions & Answers **PDF** and **VCE** file from:

https://www.passapply.com/70-475.html

100% Passing Guarantee
100% Money Back Assurance

Following Questions and Answers are all new published by Microsoft Official Exam Center

- **Instant Download** After Purchase
- **100% Money Back** Guarantee
- **365 Days** Free Update
- **800,000+** Satisfied Customers

QUESTION 1

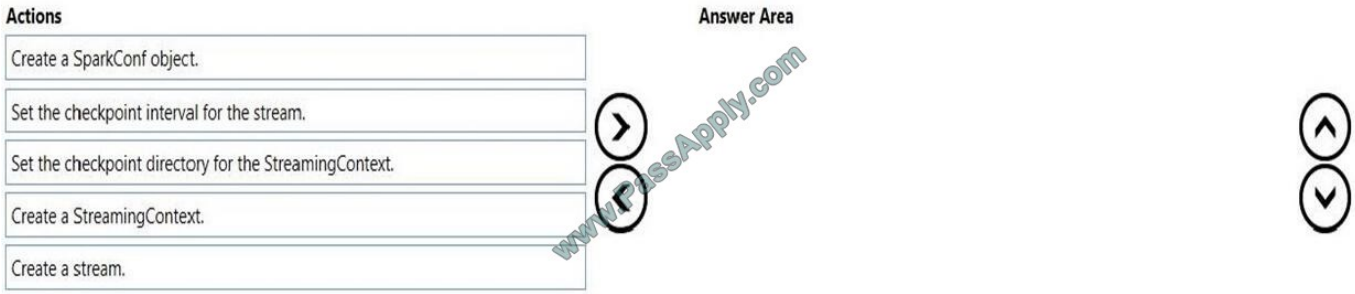
You have an Apache Spark cluster on Microsoft Azure HDInsight for all analytics workloads.

You plan to build a Spark streaming application that processes events ingested by using Azure Event Hubs.

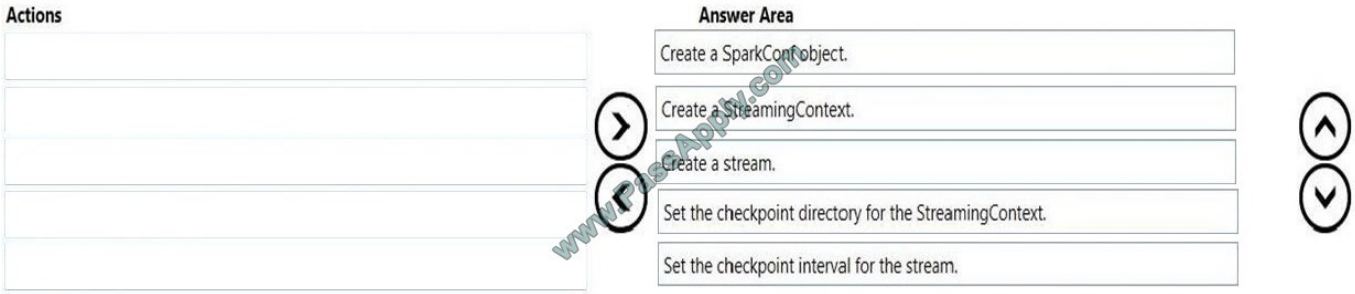
You need to implement checkpointing in the Spark streaming application for high availability of the event data.

In which order should you perform the actions? To answer, move all actions from the list of actions to the answer area and arrange them in correct order.

Select and Place:



Correct Answer:



Step 1: Create a SparkConf object

SparkConf. Spark properties control most application settings and are configured separately for each application. These properties can be set directly on a SparkConf passed to your SparkContext.

Step 2: Create a StreamingContext

Example:

val conf = new SparkConf().setMaster("local[2]").setAppName("NetworkWordCount")

val ssc = new StreamingContext(conf, Seconds(1))

Step 3: Create a stream

Step 4: Set the checkpoint directory for the StreamingContext

Set a checkpoint directory in an HDFS-compatible file system with streamingContext.checkpoint(hdfsDirectory).

Step 5: Set the checkpoint interval for the stream

tried to write in C++ but mostly fall back to the better know C

References:

https://spark.apache.org/docs/latest/streaming-programming-guide.html#checkpointing

QUESTION 2

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.

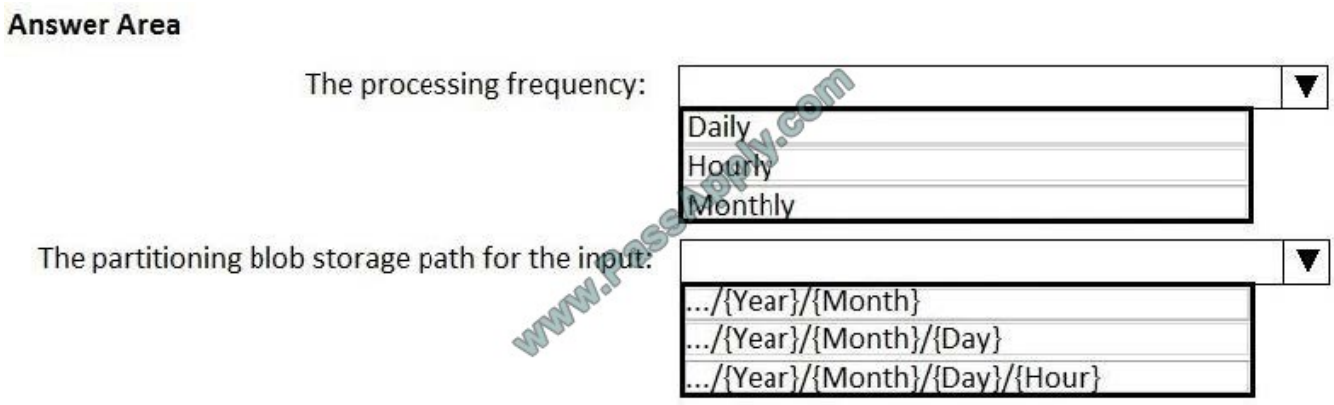
You create an Azure data factory that has blob storage as the input source and an Azure HDInsight activity that uses the input to create an output Hive table.

You need to identify a data slicing strategy for the data factory.

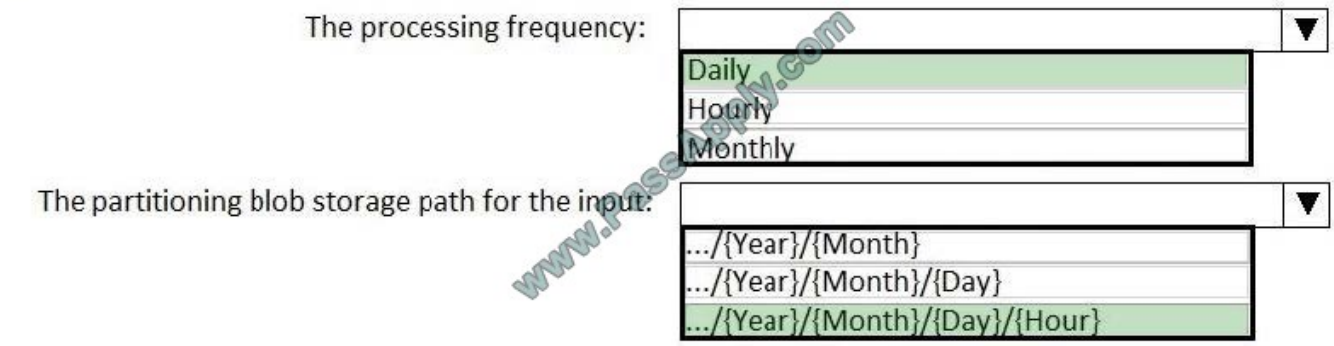
What should you identify? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:



Correct Answer:



Box 1: Daily

Data processing will occur once a day at midnight only.

Box 2: ../{Year}/{Month}/{Day}/{Hour}

You are planning a solution that will have multiple data files stored in Microsoft Azure Blob storage every hour.

QUESTION 3

You have a Microsoft Azure subscription that contains an Azure Data Factory pipeline.

You have an RSS feed that is published on a public website.

You need to configure the RSS feed as a data source for the pipeline.

Which type of linked service should you use?

A. web

B. OData

C. Azure Search

D. Azure Data Lake Store

Correct Answer: A

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

QUESTION 4

You are designing a data-driven data flow in Microsoft Azure Data Factory to copy data from Azure Blob storage to Azure SQL Database.

You need to create the copy activity.

How should you complete the JSON code? To answer, drag the appropriate code elements to the correct targets. Each 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.

Select and Place:

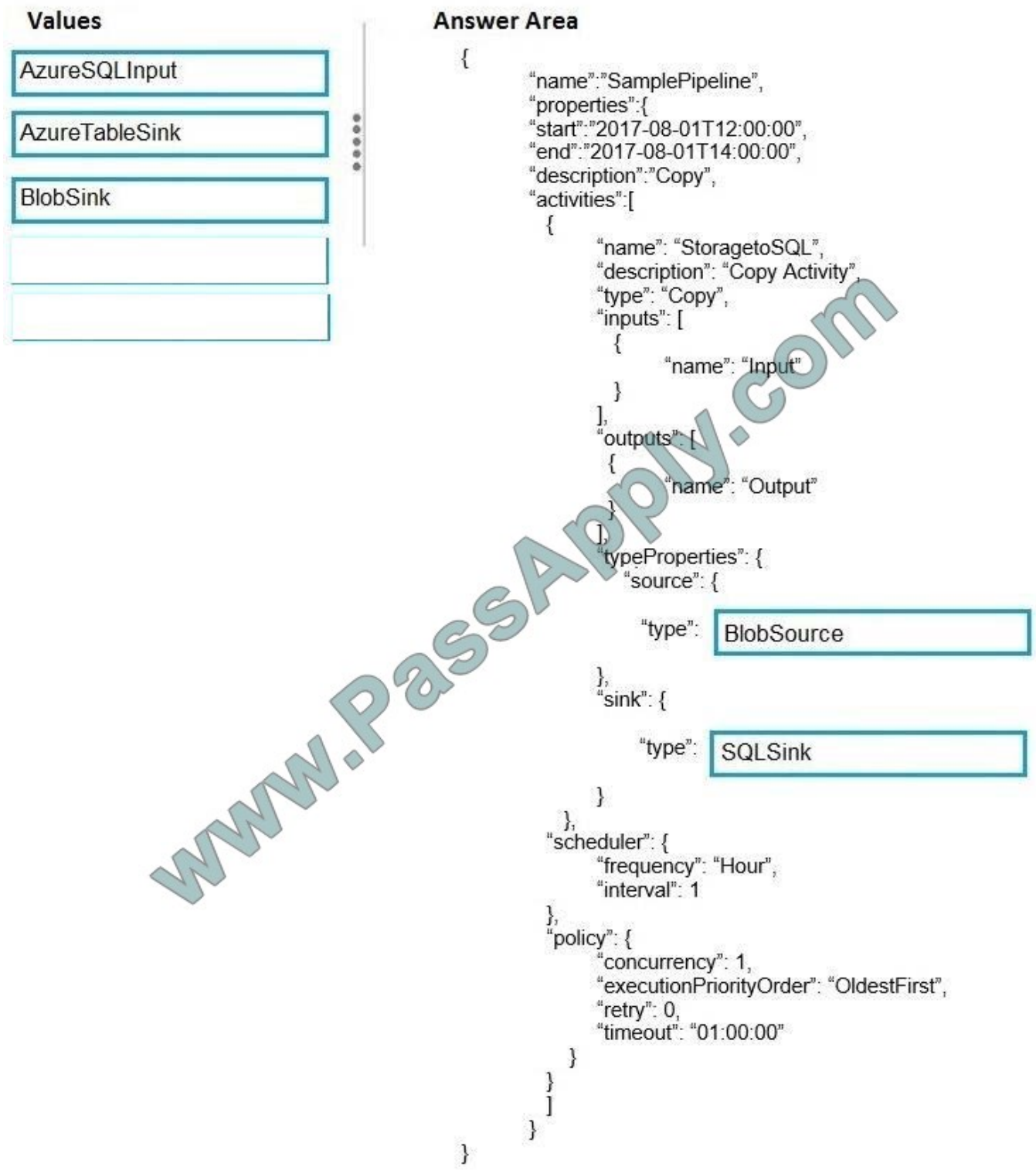
Values

- AzureSQLInput
- AzureTableSink
- BlobSink
- BlobSource
- SQLSink

Answer Area

```
{
        "name":"SamplePipeline",
        "properties":{
        "start":"2017-08-01T12:00:00",
        "end":"2017-08-01T14:00:00",
        "description":"Copy",
        "activities":[
          {
                "name": "StoragetoSQL",
                "description": "Copy Activity",
                "type": "Copy",
                "inputs": [
                  {
                        "name": "Input"
                  }
                ],
                "outputs": [
                  {
                        "name": "Output"
                  }
                ],
                "typeProperties": {
                  "source": {
                        "type":  Value
                  },
                  "sink": {
                        "type":  Value
                  }
            },
          "scheduler": {
                "frequency": "Hour",
                "interval": 1
          },
          "policy": {
                "concurrency": 1,
                "executionPriorityOrder": "OldestFirst",
                "retry": 0,
                "timeout": "01:00:00"
            }
          }
          ]
        }
}
```


Correct Answer:



In the typeProperties section, BlobSource is specified as the source type and SqlSink is specified as the sink type. Box 1: BlobSource

Box 2: SqlSink

References: https://docs.microsoft.com/en-us/azure/data-factory/concepts-pipelines-activities

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 the bcp utility to export CSV files from SQL Server and then to import the files to Azure SQL Data Warehouse.

Does this meet the goal?

A. Yes

B. No

Correct Answer: B

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