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.geekcert.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 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

Correct Answer: BCE

The dataset specifies the name of the database and the table that contains the source data.

Create a pipeline to look up the tables to be copied and another pipeline to perform the actual copy operation.

You create linked services to link your data stores and computes to a data factory. A linked service has the connection information that the Data Factory service uses to connect to the data store at runtime.

References:

https://docs.microsoft.com/en-us/azure/data-factory/tutorial-bulk-copy-portal

QUESTION 2

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.

Select and Place:

Code Elements

accessKey
AccountKey1
accountName
AccountName=Account2;AccountKey1
AzureBatchLinkedService
AzureStorageLinkedService

Answer Area

[blank box]

:" DefaultEndpointsProtocol=https; [blank box]

Correct Answer:

Code Elements

accessKey
AccountKey1
accountName
AzureBatchLinkedService

Answer Area

AzureStorageLinkedService

:" DefaultEndpointsProtocol=https; | AccountName=Account2;AccountKey1 |

QUESTION 3

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.

Correct Answer: B

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-architecture-and-operations/

QUESTION 4

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 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

Correct Answer: A

QUESTION 5

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

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

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

Hot Area:

Answer Area

The pipeline will run [answer choice].

| every three days at 10:00 |
| every three days at 22:00 |
| three times a day starting at 10:00 |
| three times a day starting at 22:00 |

If the pipeline fails to run, it will retry [answer choice].

| every minute up to three times |
| every 10 minutes up to three times |
| every 10 minutes until the pipeline succeeds |
| every one minute until the pipeline succeeds |

Correct Answer:

Answer Area

The pipeline will run [answer choice].

| every three days at 10:00 |
| every three days at 22:00 |
| three times a day starting at 10:00 |
| three times a day starting at 22:00 |

If the pipeline fails to run, it will retry [answer choice].

| every minute up to three times |
| every 10 minutes up to three times |
| every 10 minutes until the pipeline succeeds |
| every one minute until the pipeline succeeds |

Box 1: Every three days at 10.00

anchorDateTime defines the absolute position in time used by the scheduler to compute dataset slice boundaries.

"frequency": "",

"interval": "

Box 2: Every minute up to three times.

retryInterval is the wait time between a failure and the next attempt. This setting applies to present time. If the previous try failed, the next try is after the retryInterval period.

Example: 00:01:00 (1 minute)

Example: If it is 1:00 PM right now, we begin the first try. If the duration to complete the first validation check is 1 minute and the operation failed, the next retry is at 1:00 + 1min (duration) + 1min (retry interval) = 1:02 PM.

For slices in the past, there is no delay. The retry happens immediately.

retryTimeout is the timeout for each retry attempt.

maximumRetry is the number of times to check for the availability of the external data.

Reference: https://docs.microsoft.com/en-us/azure/data-factory/data-factory-create-datasets

Latest 70-475 Dumps   70-475 PDF Dumps   70-475 Practice Test

To Read the Whole Q&As, please purchase the Complete Version from Our website.
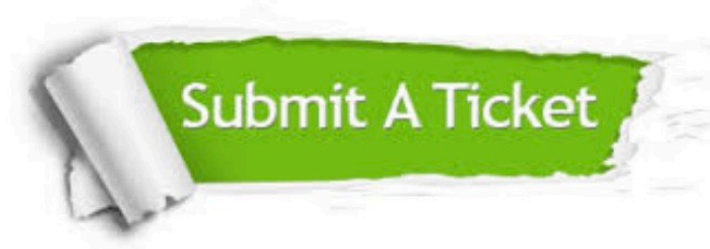
Try our product !

100% Guaranteed Success
100% Money Back Guarantee
365 Days Free Update
Instant Download After Purchase
24x7 Customer Support
Average 99.9% Success Rate
More than 800,000 Satisfied Customers Worldwide
Multi-Platform capabilities - Windows, Mac, Android, iPhone, iPod, iPad, Kindle

We provide exam PDF and VCE of Cisco, Microsoft, IBM, CompTIA, Oracle and other IT Certifications. You can view Vendor list of All Certification Exams offered:

https://www.geekcert.com/allproducts

Need Help

Please provide as much detail as possible so we can best assist you.
To update a previously submitted ticket: