Vendor: Microsoft

Exam Code:70-475

Exam Name: Designing and Implementing Big Data Analytics Solutions

Version: Demo

QUESTION 1

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

---

QUESTION 2

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

Correct Answer: A

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-

QUESTION 3

A company named Fabrikam, Inc. has a web app. Millions of users visit the app daily.

Fabrikam performs a daily analysis of the previous day\\'s logs by scheduling the following Hive query.

CREATE EXTERNAL TABLE IF NOT EXISTS UserActivity (...) Partitioned BY (LogDate string) Location MSCK REPAIR TABLE UserActivity; Select ... From UserActivity where LogDate = "{date}";

You need to recommend a solution to gather the log collections from the web app.

What should you recommend?

A. Generate a single directory that contains multiple files for each day. Name the file by using the syntax of {date}_{randomsuffix}.txt.

B. Generate a directory that is named by using the syntax of "LogDate={date}" and generate a set of files for that day.

C. Generate a directory each day that has a single file.

D. Generate a single directory that has a single file for each day.

Correct Answer: B

By default, all the data files for a table are located in a single directory. Partitioning is a technique for physically dividing the data during loading, based on values from one or more columns, to speed up queries that test those columns. References: https://www.cloudera.com/documentation/enterprise/5-8-x/topics/impala_partitioning.html#partitioning

QUESTION 4

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

Correct Answer: D

QUESTION 5

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

Correct Answer: DE

QUESTION 6

Your company has thousands of Internet-connected sensors.

You need to recommend a computing solution to perform a real-time analysis of the date 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. Microsoft Azure HDInsight HBase cluster

Correct Answer: D

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.

Reference: https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-hbase-overview https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-introduction

---

QUESTION 7

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.

Select and Place:

Actions	Answer Area
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.	

Correct Answer:

Actions	Answer Area
	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, modify the properties of the custom core-site.
	From the HDInsight Hadoop cluster, restart all of the Hadoop services.
From the HDInsight Hadoop cluster, enable maintenance mode.	
From the Azure portal, create a copy of the shared access signature (SAS) token.	

Example: scenario where a customer wishes to explore a move from an existing Hortonworks (HDP) cluster to an Azure HDInsight (HDI) cluster. With the target Storage account in place, I can now configure the HDP cluster to connect to it. The easiest way I have found to do this is using Ambari: Login to the Ambari portal and navigate to the (default) dashboard Select HDFS from the left-hand navigation On the resulting HDFS page, click on the Configs tab On the resulting page, select the Advanced option Scroll down and expand the Custom core-site node Select Add Property... from the bottom of the expanded node Enter fs.azure.account.key..blob.core.windows.net, substituting the name of the Storage account for , as the Name of the property Enter the Account key as the Value of the property Click the Add

button and verify the new property appears under the Custom core-site node Locate the notification bar at the top of the current page Click the Save button on the notification bar to push changes to the cluster Follow any remaining prompts to complete the Save process Once the Save process is completed, Ambari will indicate a restart of some services is required. Click the Restart button and Restart All Affected from the resulting drop-down. Follow any remaining prompts and monitor the process until it is successfully completed.

Reference: https://blogs.msdn.microsoft.com/data_otaku/2016/12/09/pushing-data-from-a-hortonworks-cluster-to-an-azure-hdinsight-cluster/

QUESTION 8

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

Correct Answer: B

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

---

QUESTION 9

Your company plans to deploy a web application that will display marketing data to its customers.

You create an Apache Hadoop cluster in Microsoft Azure HDInsight and an Azure data factory.

You need to implement a linked service to the cluster.

Which JSON specification should you use to create the linked service?

A. { "name":"AzureBlobOutput", "properties": { "type":"AzureBlob", "linkedServiceName":"StorageLinkedService", "typeProperties": { "folderPath":"adfgetstarted/partitioneddata", "format": { "type":"TextFormat", "columnDeliminater":"," } } "availability":" { "frequency":"Month", "interval": 1 } } }

B. { "name":"HDInsightOnDemandLinkedService", "properties": { "type":"HDInsightOnDemand", "typeProperties": { "version":"3.2", "clusterSize": 1, "timeToLive":"00:30:00", "linkedServiceName":"StorageLinkedService" } } } {

C. "name":"StorageLinkedService", "properties": { "type":"AzureStorage", "description": "", "typeProperties": { "connectionString": "DefaultEndpointProtocol=https;AccountName=; AccountKey=" } } } {

D. "name":"AzureBlobInput", "properties": { "type":"AzureBlob", "linkedServiceName":"StorageLinkedService", "typeProperties": { "folderPath":"adfgetstarted/inputdata", "format": { "type":"TextFormat", "columnDeliminater":"," } } "availability":" { "frequency":"Month", "interval": 1 }, "external":true, "policy":{} } }

Correct Answer: B

QUESTION 10

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 from a user\\'s personal device to Microsoft Azure.

Litware has three development and commercial offices. The offices are located in the United States, Luxembourg, and India.

Litware products are sold worldwide. Litware has commercial representatives in more than 80 countries.

Existing 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 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.

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:

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.

Minimize the costs associated with transferring data from the event hub to Azure Storage.

Litware identifies the following technical requirements:

Data from the devices must be stored for three years in a format that enables the fast processing of date 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.

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.

When the Litware health tracking application asks users how they feel, their responses must be reported to Azure.

You need to recommend a data handling solution to support the planned changes to the dashboard.

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

Correct Answer: C

From scenario: 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. The developers must be prevented from accessing the private information of the users.

Dynamic Data Masking can be used to hide or obfuscate sensitive data, by controlling how the data appears in the output of database queries. Dynamic Data Masking rules can be defined on particular columns, indicating how the data in those columns will appear when queried. There are no physical changes to the data in the database itself; the data remains intact and is fully available to authorized users or applications. Database operations remain unaffected, and the masked data has the same data type as the original data, so DDM can often be applied without making any changes to database procedures or application code.

Reference: https://blogs.technet.microsoft.com/dataplatforminsider/2016/01/25/use-dynamic-data-masking-to-obfuscate-your-sensitive-data/

QUESTION 11

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 datasets

E. transformations

Correct Answer: AD

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.

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

---

QUESTION 12

Note: The 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 identify purchasing fraud.

You need to design a solution that mines 10 terabytes (TB) of sales date. 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 HDInsight cluster.

Does this meet the goal?

A. Yes

B. No

Correct Answer: B

HDInsight cluster billing starts once a cluster is created and stops when the cluster is deleted. Billing is pro-rated per minute, so you should always delete your cluster when it is no longer in use. Reference: https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-hadoop-provision-linux-clusters

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
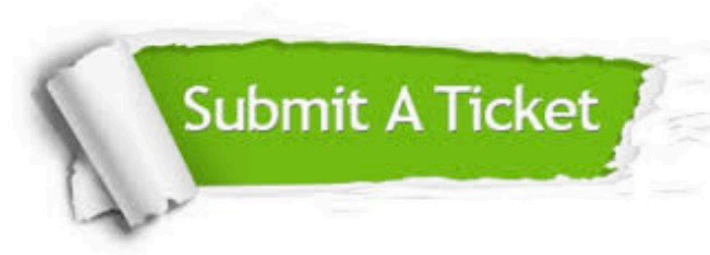
Average 99.9% Success Rate

More than 800,000 Satisfied Customers Worldwide

Multi-Platform capabilities - Windows, Mac, Android, iPhone, iPod, iPad, Kindle

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

One Year Free Update
Free update is available within One Year after your purchase. After One Year, you will get 50% discounts for updating. And we are proud to boast a 24/7 efficient Customer Support system via Email.

Money Back Guarantee
To ensure that you are spending on quality products, we provide 100% money back guarantee for 30 days from the date of purchase.

Security & Privacy
We respect customer privacy. We use McAfee's security service to provide you with utmost security for your personal information & peace of mind.