



Vendor: Microsoft

Exam Code: 70-475

Exam Name: Designing and Implementing Big Data
Analytics Solutions

Version: DEMO

Case Study 1 - Relecloud (QUESTION 1 - QUESTION 5)

General Overview

Relecloud is a social media company that processes hundreds of millions of social media posts per day and sells advertisements to several hundred companies. Relecloud has a Microsoft SQL Server database named DB1 that stores information about the advertisers. DB1 is hosted on a Microsoft Azure virtual machine.

Case Study 2 - Litware, Inc (QUESTION 6 - QUESTION 11)

Overview

General Overview

Litware, Inc. is a company that manufactures personal devices to track physical activity and other health-related data.

Litware has a health tracking application that sends health-related data from a user's personal device to Microsoft Azure.

QUESTION 1

You have structured data that resides in Microsoft Azure Blob Storage.

You need to perform a rapid interactive analysis of the data and to generate visualizations of the data.

What is the best type of Azure HDInsight cluster to use to achieve the goal? More than one answer choice may achieve the goal. Select the BEST answer.

- A. Apache Storm
- B. Apache HBase
- C. Apache Hadoop
- D. Apache Spark

Answer: C

QUESTION 2

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: A

QUESTION 3

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 identify 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 Cloudera Hadoop cluster on Microsoft Azure virtual machines. Does this meet the goal?

- A. Yes
- B. No

Answer: A

QUESTION 4

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 Databricks.

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: A

QUESTION 5

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-AzureDataFactorySliceStatus

Answer: B

QUESTION 6

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: AC

QUESTION 7

You have an Apache Storm cluster.

The cluster will ingest data from a Microsoft Azure event hub.

The event hub has the characteristics described in the following table.

Setting name	Value
Message Retention	1
Namespace	storm1.servicebus.windows.net
Shared access policies	2
Partition Count	16
Region	Central US

You are designing the Storm application topology.

You need to ingest data from all of the partitions.

The solution must maximize the throughput of the data ingestion.

Which setting should you use?

- A. Partition Count
- B. Message Retention
- C. Partition Key
- D. Shared access policies

Answer: A

QUESTION 8

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>

Thank You for Trying Our Product

Lead2pass Certification Exam Features:

- ★ More than **99,900** Satisfied Customers Worldwide.
- ★ Average **99.9%** Success Rate.
- ★ **Free Update** to match latest and real exam scenarios.
- ★ **Instant Download** Access! No Setup required.
- ★ Questions & Answers are downloadable in **PDF** format and **VCE** test engine format.
- ★ Multi-Platform capabilities - **Windows, Laptop, Mac, Android, iPhone, iPod, iPad**.
- ★ **100%** Guaranteed Success or **100%** Money Back Guarantee.
- ★ **Fast**, helpful support **24x7**.



View list of all certification exams: <http://www.lead2pass.com/all-products.html>



10% Discount Coupon Code: ASTR14