



70-775^{Q&As}

Perform Data Engineering on Microsoft Azure HDInsight

Pass Microsoft 70-775 Exam with 100% Guarantee

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

<https://www.lead4pass.com/70-775.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

Note: This question is part of a series of questions that use the same or similar answer choices. An answer choice may be correct for more than one question in the series. Each question is independent of the other questions in this series.

Information and details provided in a question apply only to that question.

You are implementing a batch processing solution by using Azure HDInsight.

You have data stored in Azure.

You need to ensure that you can access the data by using Azure Active Directory (Azure AD) identities.

What should you do?

- A. Use a shuffle join in an Apache Hive query that stores the data in a JSON format.
- B. Use a broadcast join in an Apache Hive query that stores the data in an ORC format.
- C. Increase the number of spark.executor.cores in an Apache Spark job that stores the data in a text format.
- D. Increase the number of spark.executor.instances in an Apache Spark job that stores the data in a text format.
- E. Decrease the level of parallelism in an Apache Spark job that stores the data in a text format.
- F. Use an action in an Apache Oozie workflow that stores the data in a text format.
- G. Use an Azure Data Factory linked service that stores the data in Azure Data Lake.
- H. Use an Azure Data Factory linked service that stores the data in an Azure DocumentDB database.

Correct Answer: G

References: <https://docs.microsoft.com/en-us/azure/data-factory/concepts-datasets-linked-services>

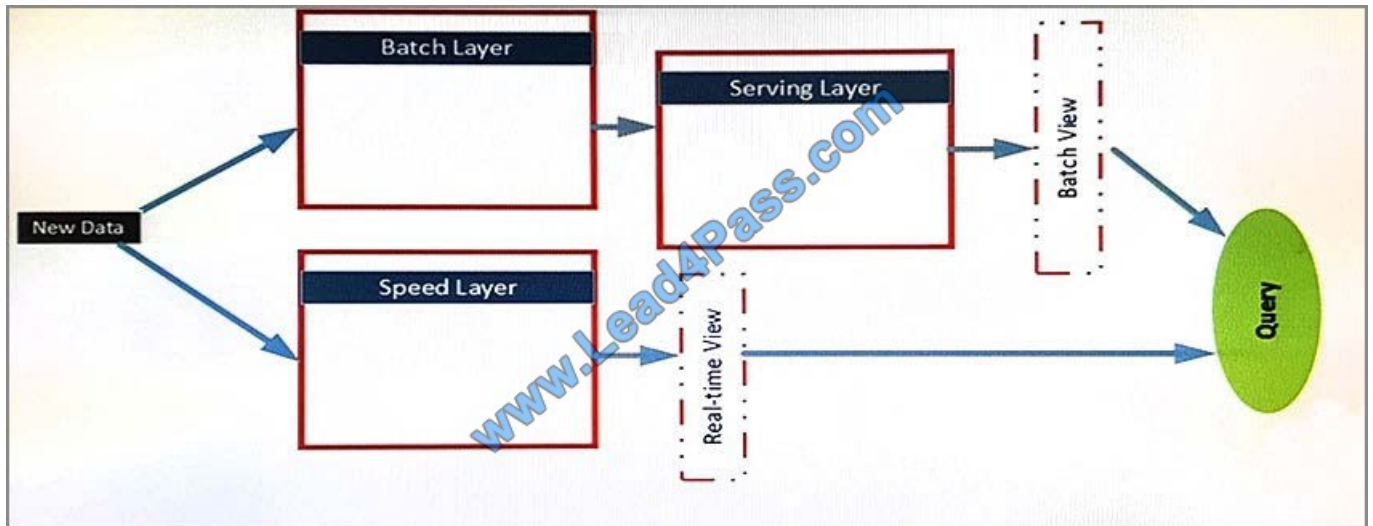
QUESTION 2

Note: This question is part of a series of questions that use the same scenario. For your convenience, the scenario is repeated in each question. Each question presents a different goal and answer choices, but the text of the scenario is

exactly the same in each question in this series.

You are planning a big data infrastructure by using an Apache Spark cluster in Azure HDInsight. The cluster has 24 processor cores and 512 GB of memory.

The architecture of the infrastructure is shown in the exhibit. (Click the Exhibit button.)



The architecture will be used by the following users:

The data sources in the batch layer share a common storage container. The following data sources are used:

The business analysts require queries to monitor the sales data. The queries must be faster and more interactive than the batch layer queries.

You need to create a new infrastructure to support the queries. The solution must ensure that you can tune the cache policies of the queries.

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.

Select and Place:



Actions

Create an Apache HBase cluster type.

Create a Windows-based HDInsight cluster.

Create a Linux-based HDInsight cluster.

Modify the hfile.block.cache.size setting.

Create an Apache Interactive Hive cluster type.

Modify the hive.llap.io.lrfu.lambda setting.

Answer Area



Correct Answer:

Actions

Create a Windows-based HDInsight cluster.

Modify the hfile.block.cache.size setting.

Create an Apache Interactive Hive cluster type.

Answer Area

Create a Linux-based HDInsight cluster.

Modify the hive.llap.io.lrfu.lambda setting.

Create an Apache HBase cluster type.





QUESTION 3

Note: This question is part of a series of questions that use the same or similar answer choices. An answer choice may be correct for more than one question in the series. Each question is independent of the other questions in this series.

Information and details provided in a question apply only to that question.

You need to deploy an HDInsight cluster that will have a custom Apache Ambari configuration. The cluster will be joined to a domain and must perform the following:

What should you do?

- A. Use an Azure PowerShell script to create and configure a premium HDInsight cluster. Specify Apache Hadoop as the cluster type and use Linux as the operating system.
- B. Use the Azure portal to create a standard HDInsight cluster. Specify Apache Spark as the cluster type and use Linux as the operating system.
- C. Use an Azure PowerShell script to create a standard HDInsight cluster. Specify Apache HBase as the cluster type and use Windows as the operating system.
- D. Use an Azure PowerShell script to create a standard HDInsight cluster. Specify Apache Storm as the cluster type and use Windows as the operating system.
- E. Use an Azure PowerShell script to create a premium HDInsight cluster. Specify Apache HBase as the cluster type and use Linux as the operating system.
- F. Use an Azure portal to create a standard HDInsight cluster. Specify Apache Interactive Hive as the cluster type and use Linux as the operating system.
- G. Use an Azure portal to create a standard HDInsight cluster. Specify Apache HBase as the cluster type and use Linux as the operating system.

Correct Answer: C

References: <https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-hadoop- introduction>

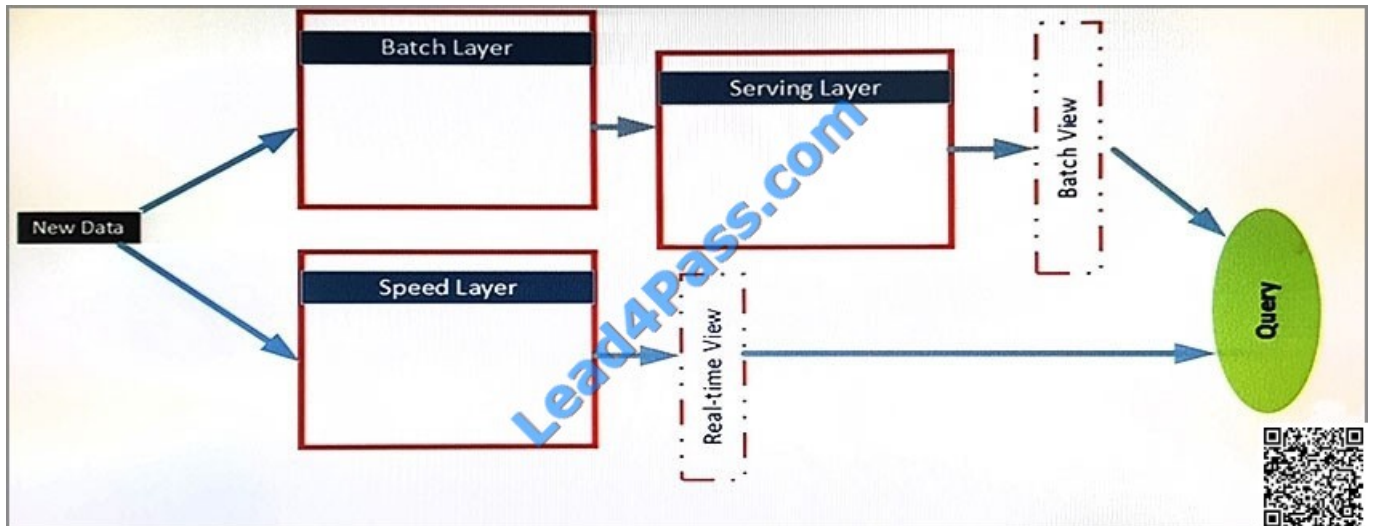
QUESTION 4

Note: This question is part of a series of questions that use the same scenario. For your convenience, the scenario is repeated in each question. Each question presents a different goal and answer choices, but the text of the scenario is

exactly the same in each question in this series.

You are planning a big data infrastructure by using an Apache Spark cluster in Azure HDInsight. The cluster has 24 processor cores and 512 GB of memory.

The architecture of the infrastructure is shown in the exhibit. (Click the Exhibit button.)



The architecture will be used by the following users:

The data sources in the batch layer share a common storage container. The following data sources are used:

You need to ensure that the analysts can query the logistics data by using JDBC APIs and SQL APIs.

Which technology should you implement?

- A. Apache Phoenix
- B. Apache Spark
- C. Apache Storm
- D. Apache Hive

Correct Answer: A

References: <http://phoenix.apache.org/>

QUESTION 5

Note: This question is part of a series of questions that use the same or similar answer choices. An answer choice may be correct for more than one question in the series. Each question is independent of the other questions in this series.

Information and details provided in a question apply only to that question.

You are implementing a batch processing solution by using Azure HDInsight.

You plan to import 300 TB of data.

You plan to use one job that has many concurrent tasks to import the data in memory.

You need to maximize the amount of concurrent tasks for the job.

What should you do?



- A. Use a shuffle join in an Apache Hive query that stores the data in a JSON format.
- B. Use a broadcast join in an Apache Hive query that stores the data in an ORC format.
- C. Increase the number of spark.executor.cores in an Apache Spark job that stores the data in a text format.
- D. Increase the number of spark.executor.instances in an Apache Spark job that stores the data in a text format.
- E. Decrease the level of parallelism in an Apache Spark job that stores the data in a text format.
- F. Use an action in an Apache Oozie workflow that stores the data in a text format.
- G. Use an Azure Data Factory linked service that stores the data in Azure Data Lake.
- H. Use an Azure Data Factory linked service that stores the data in an Azure DocumentDB database.

Correct Answer: C

References: <https://blog.cloudera.com/blog/2015/03/how-to-tune-your-apache-spark-jobs- part-2/>

[70-775 PDF Dumps](#)

[70-775 VCE Dumps](#)

[70-775 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](#)

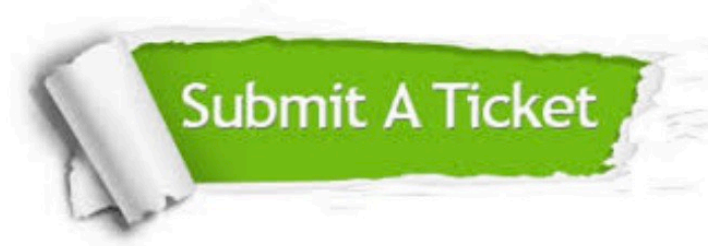
We provide exam PDF and VCE of Cisco, Microsoft, IBM, CompTIA, Oracle and other IT Certifications. You can view Vendor list of All Certification Exams offered:

<https://www.lead4pass.com/allproducts>

Need Help

Please provide as much detail as possible so we can best assist you.

To update a previously submitted ticket:



 <p>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.</p>	 <p>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.</p>	 <p>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.</p>
---	---	--

Any charges made through this site will appear as Global Simulators Limited.

All trademarks are the property of their respective owners.

Copyright © lead4pass, All Rights Reserved.