



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.geekcert.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

You plan to copy data from Azure Blob storage to an Azure SQL database by using Azure Data Factory. Which file formats can you use?

- A. binary, JSON, Apache Parquet, and ORC
- B. OXPS, binary, text and JSON
- C. XML, Apache Avro, text, and ORC
- D. text, JSON, Apache Avro, and Apache Parquet

Correct Answer: D

References: <https://docs.microsoft.com/en-us/azure/data-factory/supported-file-formats- and-compression-codecs>

QUESTION 2

You are configuring the Hive views on an Azure HDInsight cluster that is configured to use Kerberos.

You plan to use the YARN logs to troubleshoot a query that runs against Apache Hadoop.

You need to view the method, the service, and the authenticated account used to run the query.

Which method call should you view in the YARN logs?

- A. HQL
- B. WebHDFS
- C. HDFS C* API
- D. Ambari RESR API

Correct Answer: D

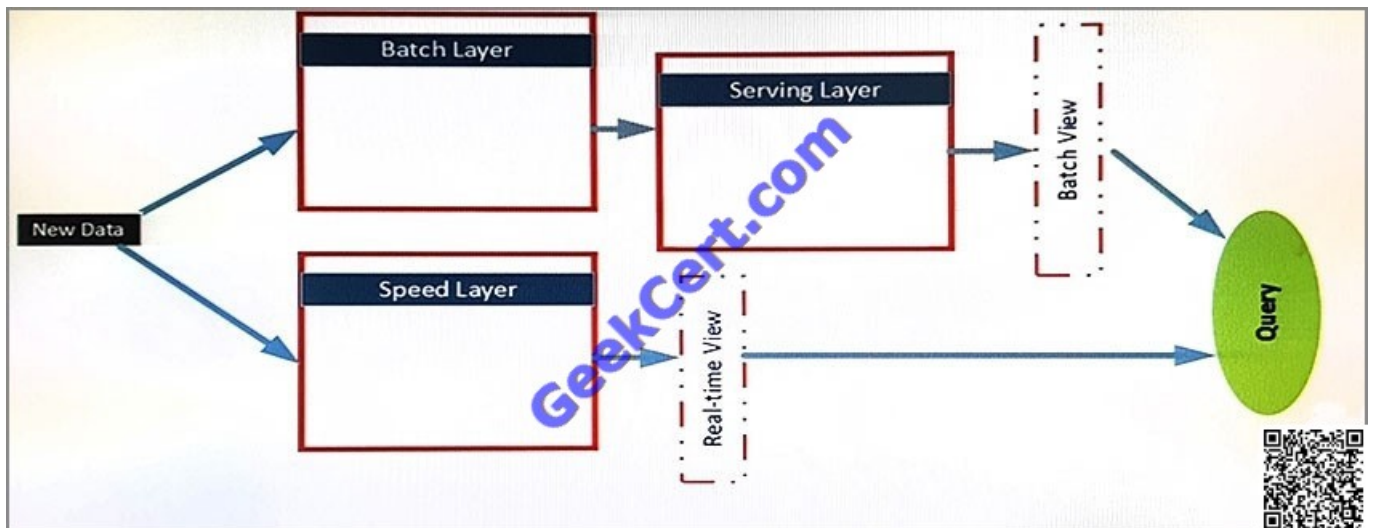
QUESTION 3

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 report that they experience performance issues when they run the monitoring queries.

You troubleshoot the performance issues and discover that the intermediate tables generated when the analysts run the queries cause pressure for the Java Virtual Machine (JVM) garbage collection per job.

Which configuration settings should you modify to alleviate the performance issues?

- A. `spark.sql.inMemoryColumnarStorage.batchSize`
- B. `spark.sql.broadcastTimeout`
- C. `spark.sql.files.openCostInBytes`
- D. `spark.sql.shuffle.partitions`

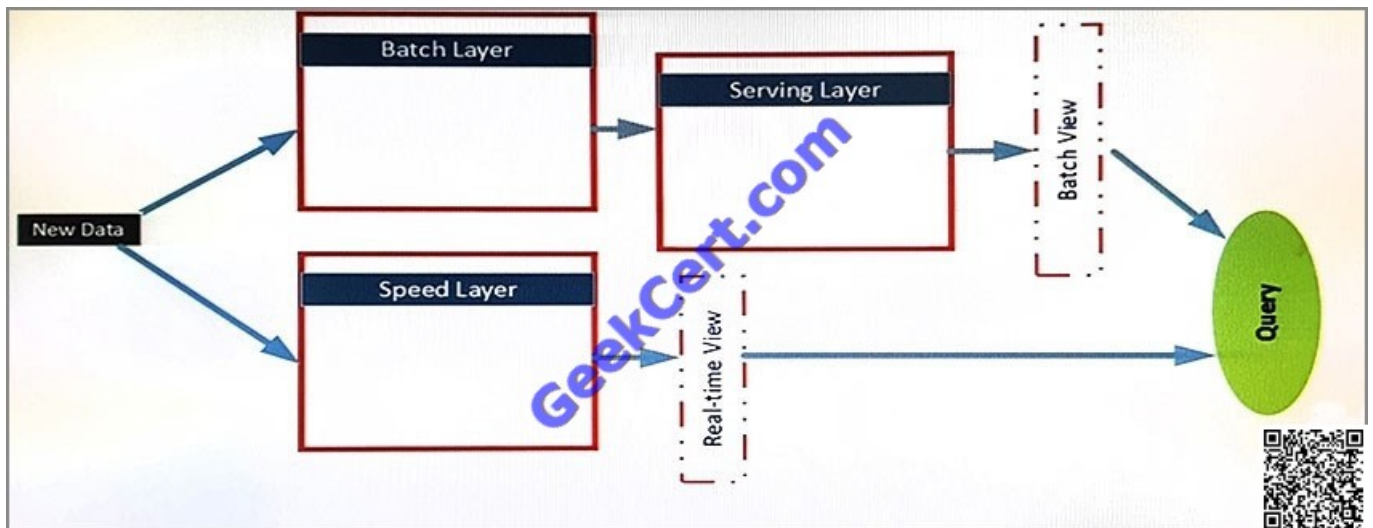
Correct Answer: D

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 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>

[70-775 PDF Dumps](#)

[70-775 Exam Questions](#)

[70-775 Braindumps](#)



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:



 <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 © geekcert, All Rights Reserved.