



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.passapply.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

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/>

**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 sections, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

You are building a security tracking solution in Apache Kafka to parse security logs. The security logs record an entry each time a user attempts to access an application. Each log entry contains the IP address used to make the attempt and

the country from which the attempt originated.

You need to receive notifications when an IP address from outside of the United States is used to access the application. Solution: Create new topics. Create a file import process to send messages. Start the consumer and run the producer. Does this meet the goal?

A. Yes

B. No

Correct Answer: A

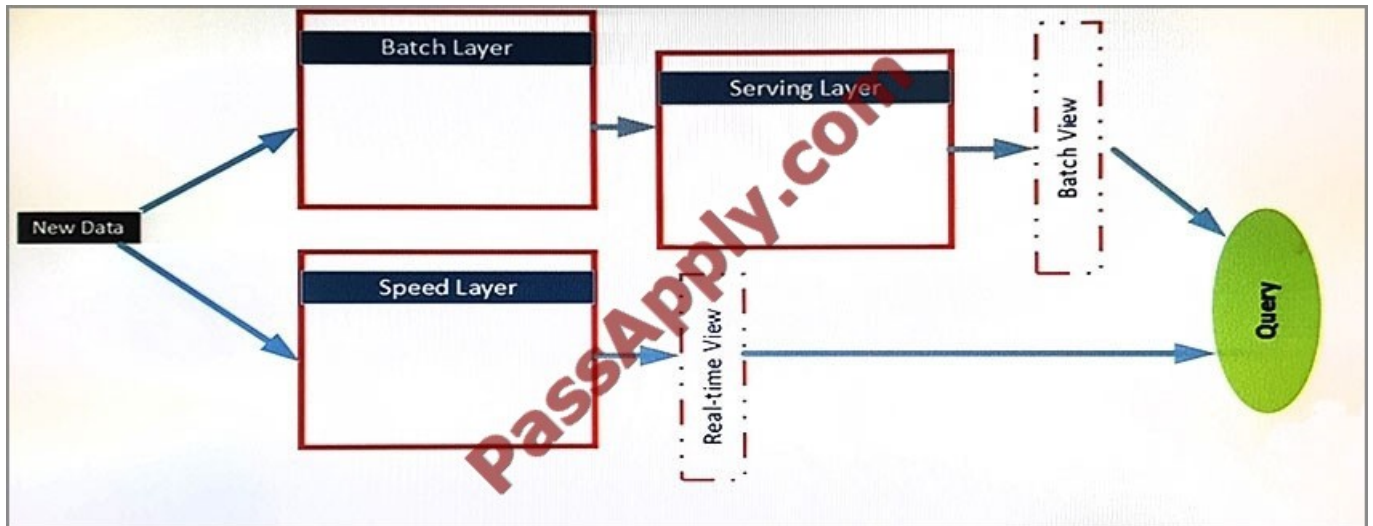
QUESTION 4

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 support analysts can develop embedded analytics applications by using the least amount of development effort.

Which technology should you implement?

- A. Zeppelin
- B. Jupyter
- C. Apache Ambari
- D. Livy

Correct Answer: D

References: <https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-apache-spark-livy-rest-interface>

QUESTION 5

You have an Apache Spark cluster in Azure HDInsight. You execute the following command.

```
%spark
import org.apache.spark.sql.hive.orc._
import org.apache.spark.sql._
```

What is the result of running the command?

- A. the Hive ORC library is imported to Spark and external tables in ORC format are created
- B. the Spark library is imported and the data is loaded to an Apache Hive table



C. the Hive ORC library is imported to Spark and the ORC-formatted data stored in Apache Hive tables becomes accessible

D. the Spark library is imported and Scala functions are executed

Correct Answer: C

[70-775 VCE 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.passapply.com/allproducts>

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.
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

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

All trademarks are the property of their respective owners.

Copyright © passapply, All Rights Reserved.