



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

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 have an initial dataset that contains the crime data from major cities.

You plan to build training models from the training data. You plan to automate the process of adding more data to the training models and to constantly tune the models by using the additional data, including data that is collected in near real-

time. The system will be used to analyze event data gathered from many different sources, such as Internet of Things (IoT) devices, live video surveillance, and traffic activities, and to generate predictions of an increased crime risk at a particular time and place.

You have an incoming data stream from Twitter and an incoming data stream from Facebook, which are event-based only, rather than time-based. You also have a time interval stream every 10 seconds.

The data is in a key/value pair format. The value field represents a number that defines how many times a hashtag occurs within a Facebook post, or how many times a Tweet that contains a specific hashtag is retweeted.

You must use the appropriate data storage, stream analytics techniques, and Azure HDInsight cluster types for the various tasks associated to the processing pipeline.

You are planning a storage strategy for a large amount of analytic data used for the crime data analytics system. The initial data load involves over 100 billion records, and more than two billion records will be added daily.

You already created an Apache Hadoop cluster in HDInsight premium.

You need to implement the storage strategy to meet the following requirements:

What should you create?

- A. a virtual machine (VM) by using the Data Science Virtual Machine template for Windows that has premium storage, a G-series size, and uses Microsoft SQL Server 2016 to store the data
- B. an Azure Data Lake Analytics service by using Azure PowerShell
- C. an Azure Data Lake Store account by using the Azure portal
- D. an Azure Blob storage account by using the Azure portal

Correct Answer: C

References: <https://docs.microsoft.com/en-us/azure/data-lake-store/data-lake-store-get-started-portal>

QUESTION 2

You have an Azure HDInsight cluster.



You need to store data in a file format that maximizes compression and increases read performance.

Which type of file format should you use?

- A. ORC
- B. Apache Parquet
- C. Apache Avro
- D. Apache Sequence

Correct Answer: A

References: <http://www.semantiko.com/blog/orc-intelligent-big-data-file-format-hadoop-hive/>

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 a NoSQL database to an HDInsight cluster. You will manage the server that host the database by using Remote Desktop. The database must use the key/value pair format in a columnar model.

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: G

References: <https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-hbase-overview>



QUESTION 4

You have an Apache Hive table that contains one billion rows.

You plan to use queries that will filter the data by using the WHERE clause. The values of the columns will be known only while the data loads into a Hive table.

You need to decrease the query runtime.

What should you configure?

- A. static partitioning
- B. bucket sampling
- C. parallel execution
- D. dynamic partitioning

Correct Answer: C

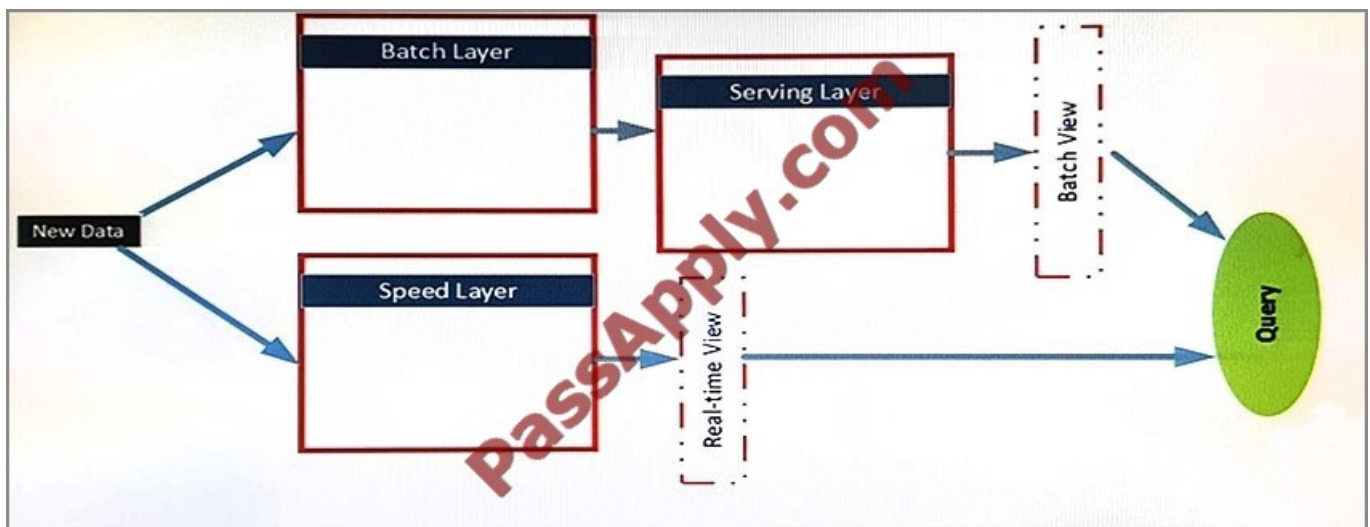
References: <https://www.qubole.com/blog/5-tips-for-efficient-hive-queries/>

QUESTION 5

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

[70-775 VCE Dumps](#)

[70-775 Practice Test](#)

[70-775 Study Guide](#)



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.