



# 70-775<sup>Q&As</sup>

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 install the Microsoft Hive ODBC Driver on a computer that runs Windows 10 and has the 64-bit version of Microsoft Office 2016 installed.

You deploy a new Apache Interactive Hive cluster in Azure HDInsight. The cluster is hosted at myHDICluster.azurehdinsight.net and contains a Hive table name hivesampletable that has 200,000 rows.

You plan to use HiveQL exclusively for the queries. The queries will return from 6,000 to 10,000 rows 90 percent of the time.

You need to configure a data source to ensure that you can use Microsoft Excel to access the data. The solution must ensure that the Hive queries execute as quickly as possible.

How should you configure the Advanced Options from the Microsoft Hive ODBC Driver DSN Setup dialog box? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

## Answer Area

Rows fetched per block:

▼
3,000
6,000
10,000
20,000

Fast SQLPrepare:

▼
Disabled
Enabled

Use Native Query:

▼
Disabled
Enabled

Correct Answer:



## Answer Area

Rows fetched per block:

▼
3,000
6,000
10,000
20,000

Fast SQLPrepare:

▼
Disabled
Enabled

Use Native Query:

▼
Disabled
Enabled

---

### QUESTION 2

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

---

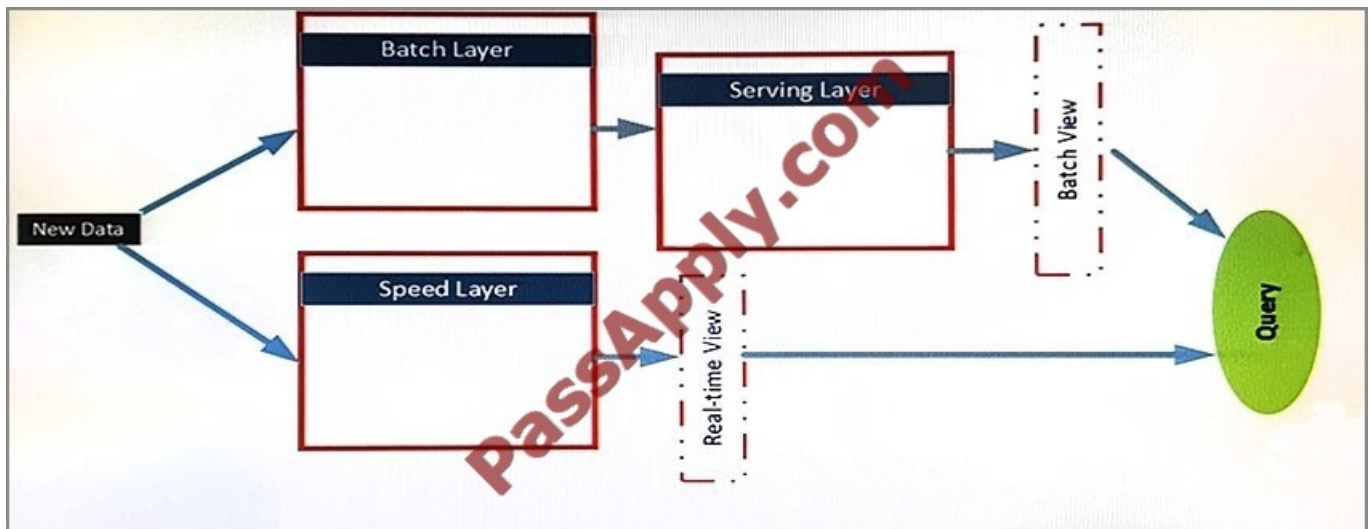
### QUESTION 3



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:

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 4

You have a domain-joined Apache Hadoop cluster in Azure HDInsight named hdcluster.

The Linux account for hdcluster is named Inxuser.



Your Active Directory account is named user1@fabrikam.com.

You need to run Hadoop commands from an SSH session.

Which credentials should you use? To answer, drag the appropriate credentials to the correct commands. Each credential may be used once, more than once, or not at all. You may need to drag the split bar between panes or scroll to view

content.

NOTE: Each correct selection is worth one point.

Select and Place:

### Credentials

lnxuser@hdicluster  
lnxuser@hdicluster-ssh.azurehdinsight.net  
user1@fabrikam.com  
user1@fabrikam.com.hdinsight.net

### Answer Area

SSH: Credential

Knit Credential

Correct Answer:

### Credentials

lnxuser@hdicluster-ssh.azurehdinsight.net  
user1@fabrikam.com.hdinsight.net

### Answer Area

SSH: lnxuser@hdicluster

Knit user1@fabrikam.com

### QUESTION 5

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 6

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 7

You use YARN to manage the resources for a Spark Thrift Server running on a Linux-based Apache Spark cluster in Azure HDInsight.

You discover that the cluster does not fully utilize the resources. You want to increase resource allocation.

You need to increase the number of executors and the allocation of memory to the Spark Thrift Server driver.

Which two parameters should you modify? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. spark.dynamicAllocation.maxExecutors
- B. spark.cores.max
- C. spark.executor.memory
- D. spark\_thrift\_cmd\_opts
- E. spark.executor.instances



Correct Answer: AC

References: <https://stackoverflow.com/questions/37871194/how-to-tune-spark-executor-number-cores-and-executor-memory>

---

### QUESTION 8

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/>

---

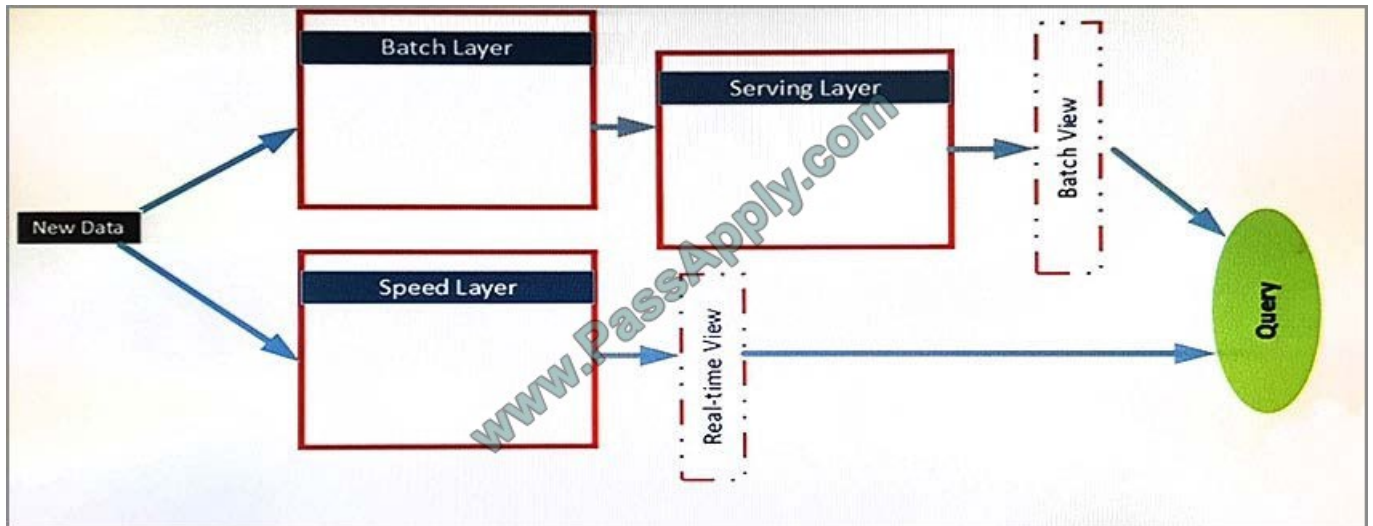
### QUESTION 9

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.lru.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.lru.lambda setting.

Create an Apache HBase cluster type.





## QUESTION 10

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 a table that contains sales data.

You plan to implement a query that will return the number of orders by zip code.

You need to minimize the execution time of the queries and to maximize the compression level of the resulting data.

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: B

---

## QUESTION 11

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 two new consumers. Create a file import process to send messages. Start the producer.

Does this meet the goal?



A. Yes

B. No

Correct Answer: B

## QUESTION 12

You are evaluating the use of Azure HDInsight clusters for various workloads.

Which type of HDInsight cluster should you create for each workload? To answer, drag the appropriate cluster types to the correct workloads. Each cluster type may be used once, more than once, or not at all. You may need to drag the split

bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

Select and Place:

### Cluster types

Apache Hadoop

Apache Spark

Apache HBase

Apache Storm

### Answer Area

ETL/ELT:

Data in motion:

Transactional processing:

Cluster type

Cluster type

Cluster type

Correct Answer:

### Cluster types

Apache Spark

### Answer Area

ETL/ELT:

Data in motion:

Transactional processing:

Apache Hadoop

Apache Storm

Apache HBase



VCE & PDF

PassApply.com

<https://www.passapply.com/70-775.html>

2021 Latest passapply 70-775 PDF and VCE dumps Download

---

[Latest 70-775 Dumps](#)

[70-775 VCE Dumps](#)

[70-775 Exam Questions](#)



VCE & PDF

PassApply.com

<https://www.passapply.com/70-775.html>

2021 Latest passapply 70-775 PDF and VCE dumps Download

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:



 <b>One Year Free Update</b> 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.	 <b>Money Back Guarantee</b> To ensure that you are spending on quality products, we provide 100% money back guarantee for 30 days from the date of purchase.	 <b>Security &amp; Privacy</b> 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.