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.pass4lead.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

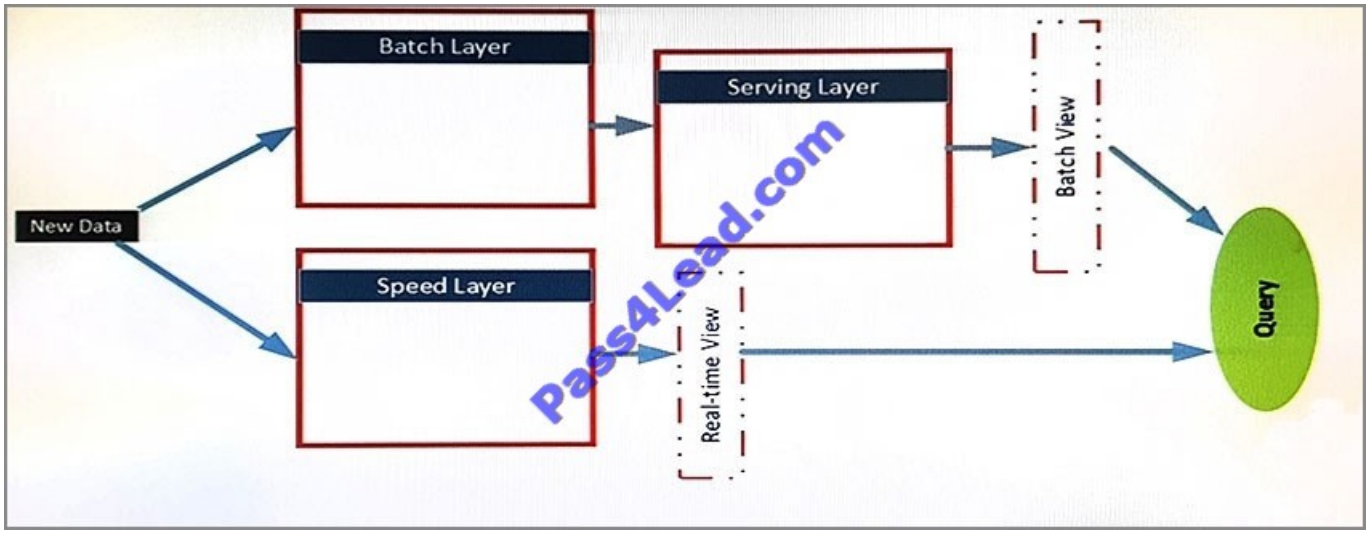
QUESTION 1

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 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 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 need to deploy an enterprise data warehouse that will support in-memory analytics. The data warehouse must support connections that use the Microsoft Hive ODBC Driver and Beeline. The data warehouse will be managed by using Apache Amrabi only.

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: F

References: https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-hadoop-use- interactive-hive

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 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 plan to consolidate all of the streams into a single timeline, even though none of the streams report events at the same interval.

You need to aggregate the data from the feeds to alight with the time interval stream. The result must be the sum of all the values for each key within a 10 second interval, with the keys being the hashtags.

Which function should you use?

A. countByWindow

B. reduceByWindow

C. reduceByKeyAndWindow

D. countByValueAndWindow

E. updateStateByKey

Correct Answer: E

QUESTION 4

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.semantikoz.com/blog/orc-intelligent-big-data-file-format-hadoop- hive/

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

Latest 70-775 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 !

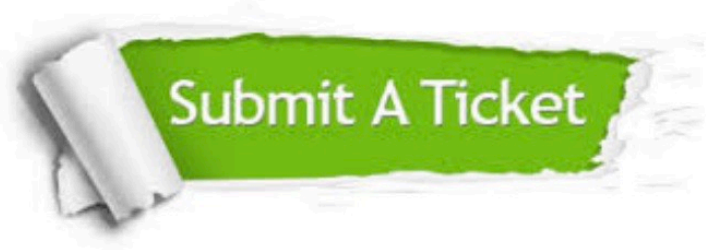
100% Guaranteed Success
100% Money Back Guarantee
365 Days Free Update
Instant Download After Purchase
24x7 Customer Support
Average 99.9% Success Rate
More than 800,000 Satisfied Customers Worldwide
Multi-Platform capabilities - Windows, Mac, Android, iPhone, iPod, iPad, Kindle

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.pass4lead.com/allproducts

Need Help

Please provide as much detail as possible so we can best assist you.
To update a previously submitted ticket:



Any charges made through this site will appear as Global Simulators Limited.
All trademarks are the property of their respective owners.
Copyright © pass4lead, All Rights Reserved.