



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.lead4pass.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 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 align 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 2

You have an Apache Hadoop cluster in Azure HDInsight that has a head node and three data nodes. You have a MapReduce job.

You receive a notification that a data node failed.



You need to identify which component cause the failure.

Which tool should you use?

- A. JobTracker
- B. TaskTracker
- C. ResourceManager
- D. ApplicationMaster

Correct Answer: C

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 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 4



You have an Apache Spark cluster in Azure HDInsight.

You plan to join a large table and a lookup table.

You need to minimize data transfers during the join operation.

What should you do?

- A. Use the reduceByKey function.
- B. Use a Broadcast variable.
- C. Repartition the data.
- D. Use the DISK_ONLY storage level.

Correct Answer: B

References: <https://www.dezyre.com/article/top-50-spark-interview-questions-and-answers-for-2017/208>

QUESTION 5

You have a text file named Data/examples/product.txt that contains product information.

You need to create a new Apache Hive table, import the product information to the table, and then read the top 100 rows of the table.

Which four code segments should you use in sequence? To answer, move the appropriate code segments from the list of code segments to the answer area and arrange them in the correct order.

Select and Place:



Code Segments

```
sqlContext.sql("CREATE TABLE IF NOT EXISTS product (productid INT, productname STRING)")
```

```
sqlContext.sql("SELECT productid, productname FROM product LIMIT 100").collect().foreach (println)
```

```
sqlContext.sql("LOAD DATA LOCAL INPATH 'data/examples/product.txt' INTO TABLE product")
```

```
sqlContext.sql("DROP TABLE [IF EXISTS] product")
```

```
val sqlContext = new org.apache.spark.sql.hive.HiveContext(sc)
```

```
sqlContext.sql("SELECT productid, productname FROM product WHERE productid < \"100\").collect ().foreach (println)
```

Answer Area

www.lead4pass.com



Correct Answer:



Code Segments

```
sqlContext.sql("DROP TABLE [IF EXISTS] product")
```

```
sqlContext.sql("SELECT productid, productname FROM product WHERE productid < \"100\"").collect().foreach (println)
```

Answer Area

```
val sqlContext = new org.apache.spark.sql.hive.HiveContext(sc)
```

```
sqlContext.sql("CREATE TABLE IF NOT EXISTS product (productid INT, productname STRING)")
```

```
sqlContext.sql("LOAD DATA LOCAL INPATH 'data/examples/product.txt' INTO TABLE product")
```

```
sqlContext.sql("SELECT productid, productname FROM product LIMIT 100").collect().foreach (println)
```





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](#)

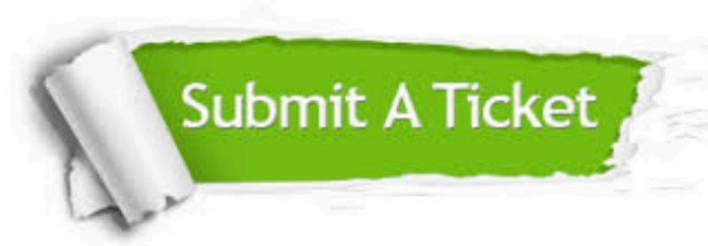
We provide exam PDF and VCE of Cisco, Microsoft, IBM, CompTIA, Oracle and other IT Certifications. You can view Vendor list of All Certification Exams offered:

<https://www.lead4pass.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 © lead4pass, All Rights Reserved.