



70-475^{Q&As}

Designing and Implementing Big Data Analytics Solutions

Pass Microsoft 70-475 Exam with 100% Guarantee

Free Download Real Questions & Answers **PDF** and **VCE** file from:

<https://www.lead4pass.com/70-475.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

Your company has thousands of Internet-connected sensors.

You need to recommend a computing solution to perform a real-time analysis of the data generated by the sensors.

Which computing solution should you include in the recommendation?

- A. Microsoft Azure Stream Analytics
- B. Microsoft Azure Notification Hubs
- C. Microsoft Azure Cognitive Services
- D. Microsoft Azure HDInsight HBase cluster

Correct Answer: D

HDInsight HBase is offered as a managed cluster that is integrated into the Azure environment. The clusters are configured to store data directly in Azure Storage or Azure Data Lake Store, which provides low latency and increased elasticity in performance and cost choices. This enables customers to build interactive websites that work with large datasets, to build services that store sensor and telemetry data from millions of end points, and to analyze this data with Hadoop jobs. HBase and Hadoop are good starting points for big data project in Azure; in particular, they can enable real-time applications to work with large datasets.

Reference: <https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-hbase-overview> <https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-introduction>

QUESTION 2

You are designing a solution that will use Apache HBase on Microsoft Azure HDInsight.

You need to design the row keys for the database to ensure that client traffic is directed over all of the nodes in the cluster.

What are two possible techniques that you can use? Each correct answer presents a complete solution.

NOTE: Each correct selection is worth one point.

- A. padding
- B. trimming
- C. hashing
- D. salting

Correct Answer: CD

There are two strategies that you can use to avoid hotspotting:

*



Hashing keys

To spread write and insert activity across the cluster, you can randomize sequentially generated keys by hashing the keys, inverting the byte order. Note that these strategies come with trade-offs. Hashing keys, for example, makes table scans for key subranges inefficient, since the subrange is spread across the cluster.

*

Salting keys

Instead of hashing the key, you can salt the key by prepending a few bytes of the hash of the key to the actual key.

Note. Salted Apache HBase tables with pre-split is a proven effective HBase solution to provide uniform workload distribution across RegionServers and prevent hot spots during bulk writes. In this design, a row key is made with a logical key plus salt at the beginning. One way of generating salt is by calculating n (number of regions) modulo on the hash code of the logical row key (date, etc).

Reference: <https://blog.cloudera.com/blog/2015/06/how-to-scan-salted-apache-hbase-tables-with-region-specific-key-ranges-in-mapreduce/> http://maprdocs.mapr.com/51/MapR-DB/designing_row_keys_for_mapr_db_binary_tables.html

QUESTION 3

You are designing a partitioning scheme for ingesting real-time data by using Kafka. Kafka and Apache Storm will be integrated.

You plan to use four event processing servers that each run as a Kafka consumer. Each server will have a two quad-core processor.

You need to identify the minimum number of partitions required to ensure that the load is distributed evenly.

How many should you identify?

- A. 1
- B. 4
- C. 16
- D. 32

Correct Answer: C

QUESTION 4

Your company has a data visualization solution that contains a customized Microsoft Azure Stream Analytics Solution. The solution provides data to a Microsoft Power BI deployment.

Every 10 seconds, you need to query for instances that have more than three records.

How should you complete the query? To answer, drag the appropriate values to the correct targets. Each value 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:

Values

-
-
-
-
-
-



Answer Area

```

SELECT
  Value
  System.Timestamp AS Time,
  COUNT(*) AS [Count]
INTO
  AlertOutput
FROM
  Input TIMESTAMP BY Time
GROUP BY
  Make,
  
HAVING
  

```

Correct Answer:

Values

-
-
-
-
-
-



Answer Area

```

SELECT
  Value
  System.Timestamp AS Time,
  COUNT(*) AS [Count]
INTO
  AlertOutput
FROM
  Input TIMESTAMP BY Time
GROUP BY
  Make,
  
HAVING
  

```

Box 1: TumblingWindow(second, 10)

Tumbling Windows define a repeating, non-overlapping window of time.

Example: Calculate the count of sensor readings per device every 10 seconds

SELECT sensorId, COUNT(*) AS Count



FROM SensorReadings TIMESTAMP BY time

GROUP BY sensorId, TumblingWindow(second, 10)

Box 2: [Count] >= 3

Count(*) returns the number of items in a group.

Reference: <https://blogs.technet.microsoft.com/machinelearning/2015/06/01/the-azure-stream-analytics-query-language/>

QUESTION 5

You have the following script.

```
CREATE TABLE UserVisits (username string, url string, time date)
```

```
STORED AS TEXTFILE LOCATION "wasb:///Logs";
```

```
CREATE TABLE UserVisitsOrc (username string, url string, time date)
```

```
STORED AS ORC;
```

```
INSERT INTO TABLE UserVisitsOrc SELECT * FROM UserVisits
```

Use the drop-down menus to select the answer choice that completes each statement based on the information presented in the script.

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

The INSERT statement [answer choice].

moves the contents of UserVisits to the UserVisitsOrc table directory
inserts data into the UserVisitsOrc table by running a YARN application
inserts data into the UserVisitsOrc table record by record from the Hive command-line interface (CLI)

The UserVisits table type is [answer choice].

dataset
external
managed

Correct Answer:



Answer Area

The INSERT statement [answer choice].

moves the contents of UserVisits to the UserVisitsOrc table directory
inserts data into the UserVisitsOrc table by running a YARN application
inserts data into the UserVisitsOrc table record by record from the Hive command-line interface (CLI)

The UserVisits table type is [answer choice].

dataset
external
managed

A table created without the EXTERNAL clause is called a managed table because Hive manages its data. Reference: <https://cwiki.apache.org/confluence/display/Hive/LanguageManual+DDL>

[70-475 PDF Dumps](#)

[70-475 VCE Dumps](#)

[70-475 Exam Questions](#)



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](#)

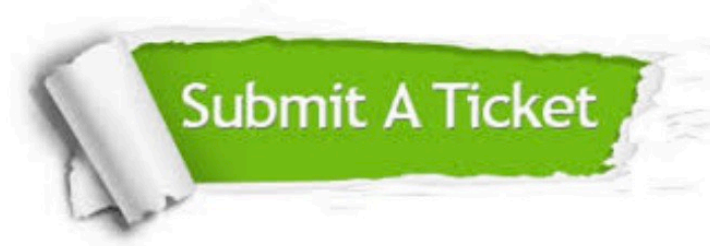
We provide exam PDF and VCE of Cisco, Microsoft, IBM, CompTIA, Oracle and other IT Certifications. You can view Vendor list of All Certification Exams offered:

<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.