70-475^{Q&As}

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.pass4lead.com/70-475.html

100% Passing Guarantee
100% Money Back Assurance

Following Questions and Answers are all new published by Microsoft
Official Exam Center

QUESTION 1

Your company has 2,000 servers.

You plan to aggregate all of the log files from the servers in a central repository that uses Microsoft Azure HDInsight. Each log file contains approximately one million records. All of the files use the .log file name extension.

The following is a sample of the entries in the log files.

2017-02-03 20:26:41 SampleClass3 [ERROR] verbose detail for id

In Apache Hive, you need to create a data definition and a query capturing the number of records that have an error level of [ERROR].

What should you do? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

```
CREATE [ EXTERNAL TABLE | TABLE | VIEW ] ▼  log4jLogs (t1 string, t2 string,
               EXTERNAL TABLE
               TABLE
               VIEW

t3 string, t4 string, t5 string, t6 string, t7 string)
ROW FORMAT DELIMITED FIELDS TERMINATED BY [ .. | /T | [ERROR] ] ▼
                                           ..
                                           /T
                                           [ERROR]

STORED AS TEXTFILE LOCATION 'wasbs:///example/data/';
SELECT t4 AS sev, [ COUNT(*) | MAX(*) | SUM(*) ] ▼  AS count
                   COUNT(*)
                   MAX(*)
                   SUM(*)

FROM log4jLogs
WHERE t4 = ' [ERROR] '
AND INPUT_FILE_NAME LIKE [ '!.log' | '%.log' | '*.log' ] ▼
                          '!.log'
                          '%.log'
                          '*.log'

GROUP BY t4;
```

Correct Answer:

Answer Area

Box 1: table

Box 2: /t

Apache Hive example:

CREATE TABLE raw (line STRING)

ROW FORMAT DELIMITED FIELDS TERMINATED BY \\'\t\\' LINES TERMINATED BY \\'\n\\';

Box 3: count(*)

Box 4: \\'*.log\\'

Reference: https://cwiki.apache.org/confluence/display/Hive/CompressedStorage

QUESTION 2

You have a large datacenter.

You plan to track the hardware failure notifications that occur in the datacenter. You expect to collect approximated 2 TB of data each month.

You need to recommend a solution that meets the following requirements:

Operators must be informed by email as soon as a hardware failure occurs.

All event data associated with a hardware failure must be preserved for 24 months.

The solution must minimize costs.

What should you identify? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

Answer Area

Operators must be informed email as soon as a hardware failure occurs:

Microsoft Azure Logic App
Microsoft Azure Notification Hubs
Microsoft Azure SQL Database

All event data associated with a hardware failure must be preserved for 24 months:

Microsoft Azure Data Lake
Microsoft Azure SQL Data Warehouse
Microsoft Azure SQL Database

Correct Answer:

Answer Area

Operators must be informed email as soon as a hardware failure occurs:

Microsoft Azure Logic App
Microsoft Azure Notification Hubs
Microsoft Azure SQL Database

All event data associated with a hardware failure must be preserved for 24 months:

Microsoft Azure Data Lake
Microsoft Azure SQL Data Warehouse
Microsoft Azure SQL Database

Box 1: Microsoft Azure Logic App

With Azure Logic Apps and the Simple Mail Transfer Protocol (SMTP) connector, you can create automated tasks and workflows that send email from your SMTP account. You can also have other actions use the output from SMTP actions.

Incorrect Answers:

Notification Hubs: Notification Hubs is primarily designed to send notifications to mobile apps. It does not provide email or text message capabilities.

Box 2: Microsoft Azure Data Lake

A data lake is a storage repository that holds a vast amount of raw data in its native format until it is needed. While a hierarchical data warehouse stores data in files or folders, a data lake uses a flat architecture to store data. Each data

element in a lake is assigned a unique identifier and tagged with a set of extended metadata tags. When a business question arises, the data lake can be queried for relevant data, and that smaller set of data can then be analyzed to help

answer the question.

References:

https://searchaws.techtarget.com/definition/data-lake

QUESTION 3

A company named Fabrikam, Inc. has a Microsoft Azure web app. Billions of users visit the app daily.

The web app logs all user activity by using text files in Azure Blob storage. Each day, approximately 200 GB of text files are created.

Fabrikam uses the log files from an Apache Hadoop cluster on Azure HDInsight.

You need to recommend a solution to optimize the storage of the log files for later Hive use.

What is the best property to recommend adding to the Hive table definition to achieve the goal? More than one answer choice may achieve the goal. Select the BEST answer.

A. STORED AS RCFILE

B. STORED AS GZIP

C. STORED AS ORC

D. STORED AS TEXTFILE

Correct Answer: C

The Optimized Row Columnar (ORC) file format provides a highly efficient way to store Hive data. It was designed to overcome limitations of the other Hive file formats. Using ORC files improves performance when Hive is reading, writing, and processing data. Compared with RCFile format, for example, ORC file format has many advantages such as: a single file as the output of each task, which reduces the NameNode\\'s load Hive type support including datetime, decimal, and the complex types (struct, list, map, and union) light-weight indexes stored within the file skip row groups that don\\'t pass predicate filtering seek to a given row block-mode compression based on data type run-length encoding for integer columns dictionary encoding for string columns concurrent reads of the same file using separate RecordReaders ability to split files without scanning for markers bound the amount of memory needed for reading or writing metadata stored using Protocol Buffers, which allows addition and removal of fields

Reference: https://cwiki.apache.org/confluence/display/Hive/LanguageManual+ORC#LanguageManualORC-ORCFileFormat

QUESTION 4

Note: This question is part of a series htf 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 the others might not have a correct solution.

After you answer a question in this section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen. You have a Microsoft Azure deployment that contains the following services: Azure Data Lake Azure Cosmos DB Azure Data Factory Azure SQL Database

You load several types of data to Azure Data Lake.

You need to load data from Azure SQL Database to Azure Data Lake.

Solution: You use the Azure Import/Export service.

Does this meet the goal?

A. Yes

B. No

Correct Answer: B

Azure Import/Export service is used to securely import large amounts of data to Azure Blob storage and Azure Files by shipping disk drives to an Azure datacenter. This service can also be used to transfer data from Azure Blob storage to disk

drives and ship to your on-premises sites.

Note:

You can use the Copy Activity in Azure Data Factory to copy data to and from Azure Data Lake Storage Gen1 (previously known as Azure Data Lake Store). Azure SQL database is supported as source.

References: https://docs.microsoft.com/en-us/azure/data-factory/connector-azure-data-lake-store

QUESTION 5

A company named Fabrikam, Inc. has a web app. Millions of users visit the app daily.

Fabrikam performs a daily analysis of the previous day\\'s logs by scheduling the following Hive query.

```
CREATE EXTERNAL TABLE IF NOT EXISTS UserActivity (...) PARTITIONED BY (LogDate string) LOCATION 'wasb:///logs';
MSCK REPAIR TABLE UserActivity;
Select .... From UserActivity where LogDate = "{date}";
```

You need to recommend a solution to gather the log collections from the web app. What should you recommend?

A. Generate a single directory that has a single file for each day.

B. Generate a single directory that contains multiple files for each day. Name the file by using the syntax of {date}_{randomsuffix}.txt.

C. Generate a directory each day that has a single file.

D. Generate a directory that is named by using the syntax of "LogDate={date}" and generate a set of files for that day.

Correct Answer: D

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
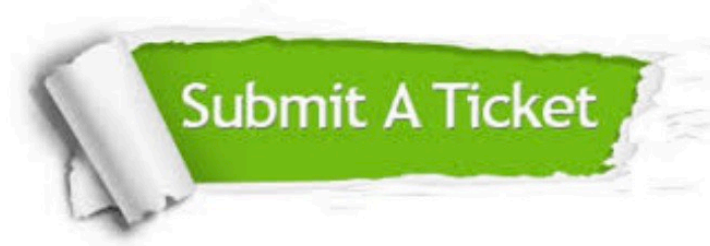Multi-Platform capabilities - Windows, Mac, Android, iPhone, iPod, iPad, Kindle

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: