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.pass4itsure.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

You have a data warehouse that contains the sales data of several customers.

You plan to deploy a Microsoft Azure data factory to move additional sales data to the data warehouse.

You need to develop a data factory job that reads reference data from a table in the source data.

Which type of activity should you add to the control flow of the job?

A. a ForEach activity

B. a lookup activity

C. a web activity

D. a GetMetadata activity

Correct Answer: B

References: https://docs.microsoft.com/en-us/azure/data-factory/control-flow-lookup-activity

QUESTION 2

Overview:

Litware, Inc. is a company that manufactures personal devices to track physical activity and other health-related data.

Litware has a health tracking application that sends health-related data from a user\\'s personal device to Microsoft Azure.

Litware has three development and commercial offices. The offices are located in the United States, Luxembourg, and India.

Litware products are sold worldwide. Litware has commercial representatives in more than 80 countries.

Existing Environment:

In addition to using desktop computers in all of the offices, Litware recently started using Microsoft Azure resources and services for both development and operations.

Litware has an Azure Machine Learning solution.

Litware recently extended its platform to provide third-party companies with the ability to upload data from devices to Azure. The data can be aggregated across multiple devices to provide users with a comprehensive view of their global

health activity.

While the upload from each device is small, potentially more than 100 million devices will upload data daily by using an Azure event hub. Each health activity has a small amount of data, such as activity type, start date/time, and end date/time.

Each activity is limited to a total of 3 KB and includes a customer identification key.

In addition to the Litware health tracking application, the users\\' activities can be reported to Azure by using an open API.

The developers at Litware perform Machine Learning experiments to recommend an appropriate health activity based on the past three activities of a user.

The Litware developers train a model to recommend the best activity for a user based on the hour of the day.

Requirements:

Litware plans to extend the existing dashboard features so that health activities can be compared between the users based on age, gender, and geographic region.

Minimize the costs associated with transferring data from the event hub to Azure Storage.

Litware identifies the following technical requirements:

Data from the devices must be stored for three years in a format that enables the fast processing of date fields and filtering.

The third-party companies must be able to use the Litware Machine Learning models to generate recommendations to their users by using a third-party application.

Any changes to the health tracking application must ensure that the Litware developers can run the experiments without interrupting or degrading the performance of the production environment.

Activity tracking data must be available to all of the Litware developers for experimentation. The developers must be prevented from accessing the private information of the users.

When the Litware health tracking application asks users how they feel, their responses must be reported to Azure.

You need to recommend a permanent Azure Storage solution for the activity data. The solution must meet the technical requirements.

What is the best recommendation to achieve the goal? More than one answer choice may achieve the goal. Select the BEST answer.

A. Azure SQL Database

B. Azure Queue storage

C. Azure Blob storage

D. Azure Event Hubs

Correct Answer: A

From scenario: While the upload from each device is small, potentially more than 100 million devices will upload data daily by using an Azure event hub. Each health activity has a small amount of data, such as activity type, start date/time, and end date/time. Each activity is limited to a total of 3 KB and includes a customer identification key.

QUESTION 3

Overview:

Relecloud is a social media company that processes hundreds of millions of social media posts per day and sells advertisements to several hundred companies. Relecloud has a Microsoft SQL Server database named DB1 that stores

information about the advertisers. DB1 is hosted on a Microsoft Azure virtual machine.

Relecloud has two main offices. The offices are located in San Francisco and New York City.

The offices connect to each other by using a site-to-site VPN. Each office connects directly to the Internet.

Relecloud modifies the pricing of its advertisements based on trending topics. Topics are considered to be trending if they generate many mentions in a specific country during a 15-minute time frame. The highest trending topics generate the

highest advertising revenue.

Relecloud wants to deliver reports to the advertisers by using Microsoft Power BI. The reports will provide real-time data on trending topics, current advertising rates, and advertising costs for a given month. Relecloud will analyze the trending

topics data, and then store the data in a new data warehouse for ad-hoc analysis. The data warehouse is expected to grow at a rate of 1 GB per hour or 8.7 terabytes (TB) per year. The data will be retained for five years for the purpose of

long-term trending.

Requirements:

Management at Relecloud must be able to view which topics are trending to adjust advertising rates in near real-time.

Relecloud plans to implement a new streaming analytics platform that will report on trending topics.

Relecloud plans to implement a data warehouse named DB2.

Relecloud identifies the following technical requirements:

Social media data must be analyzed to identify trending topics in real-time.

The use of Infrastructure as a Service (IaaS) platforms must minimized, whenever possible.

The real-time solution used to analyze the social media data must support scaling up and down without service interruption.

Relecloud identifies the following technical requirements for the advertisers:

The advertisers must be able to see only their own data in the Power BI reports.

The advertisers must authenticate to Power BI by using Azure Active Directory (Azure AD) credentials.

The advertisers must be able to leverage existing Transact-SQL language knowledge when developing the real-time streaming solution.

Members of the internal advertising sales team at Relecloud must be able to see only the sales date of the advertisers to which they are assigned.

The internal Relecloud advertising sales team must be prevented from inserting, updating, and deleting rows for the advertisers to which they are not assigned.

The internal Relecloud advertising sales team must be able to use a text file to update the list of advertisers, and then to upload the file to Azure Blob storage.

Relecloud identifies the following requirements for DB1:

Data generated by the streaming analytics platform must be stored in DB1.

The user names of the advertisers must be mapped to CustomerID in a table named Table2.

The advertisers in DB1 must be stored in a table named Table1 and must be refreshed nightly.

The user names of the employees at Relecloud must be mapped to EmployeeID in a table named Table3.

Relecloud identifies the following requirements for DB2:

DB2 must have minimal storage costs.

DB2 must run load processes in parallel.

DB2 must support massive parallel processing.

DB2 must be able to store more than 40 TB of data.

DB2 must support scaling up and down, as required.

Data from DB1 must be archived in DB2 for long-term storage.

All of the reports that are executed from DB2 must use aggregation.

Users must be able to pause DB2 when the data warehouse is not in use.

Users must be able to view previous versions of the data in DB2 by using aggregates.

Relecloud identifies the following requirements for extract, transformation, and load (ETL):

Data movement between DB1 and DB2 must occur each hour.

An email alert must be generated when a failure of any type occurs during ETL processing.

Sample code and data:

You execute the following code for a table named rls_table1.

```
create function rls_table1 (@CustomerId int, @SalesPersonId int)
       returns table
       with schemabinding
as
return
select 1 as result
from dbo.table1
join dbo.table2 on table1.customerid = Table2.CustomerId
where table2.UserName = suser_sname()
       and table1.customerid =@CustomerId
union all
select 1 as result
from dbo.table1
join dbo.table3 on table1.salespersonid = table3.EmployeeId
where table3.UserName = suser_sname()
       and table1.salespersonid = @SalesPersonId
go
```


You use the following code to create Table1.

create table table1 (customerid int, salespersonid int ...) Go

The following is a sample of the streaming data.

User	Country	Topic	Time
user1	USA	Topic1	2017-01-01T00:00:01.0000000Z
user1	USA	Topic3	2017-01-01T00:02:01.0000000Z
user2	Canada	Topic2	2017-01-01T00:01:11.0000000Z
user3	India	Topic1	2017-01-01T00:03:14.0000000Z

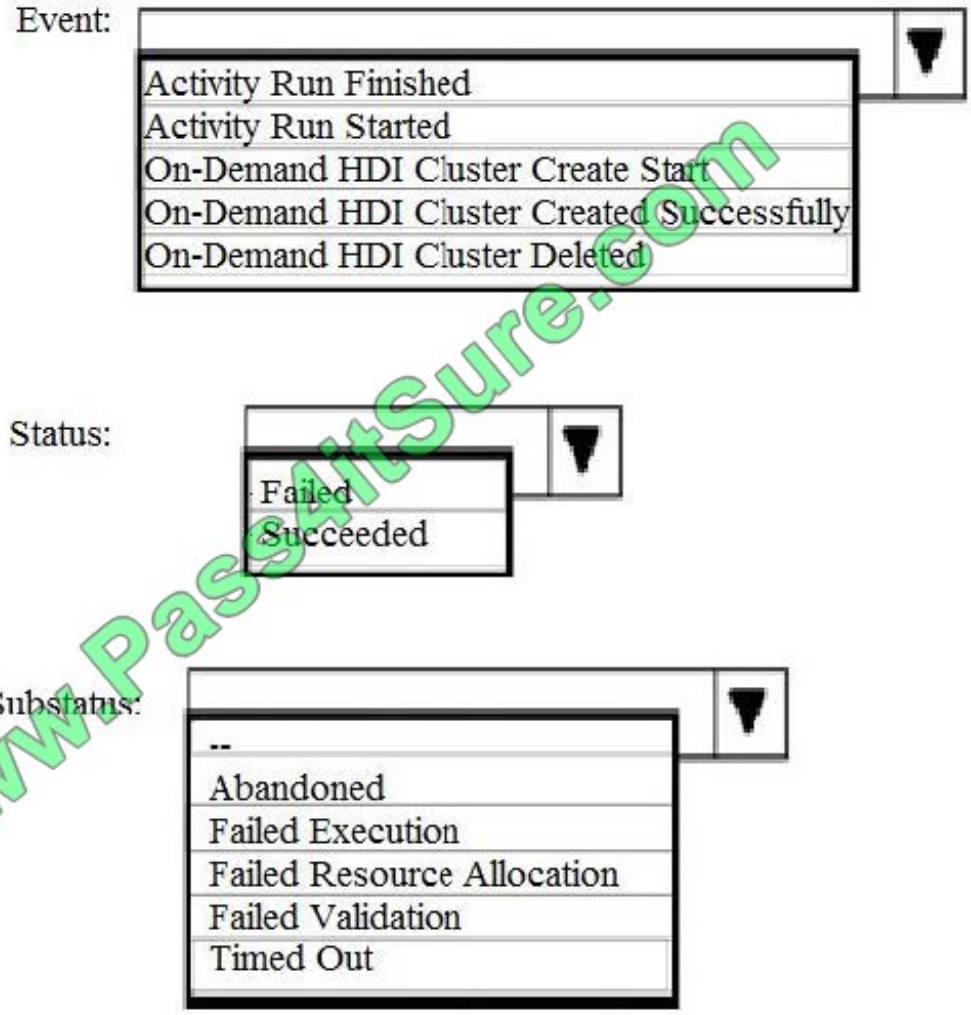
You need to configure the alert to meet the requirements for ETL.

Which settings should you use for the alert? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

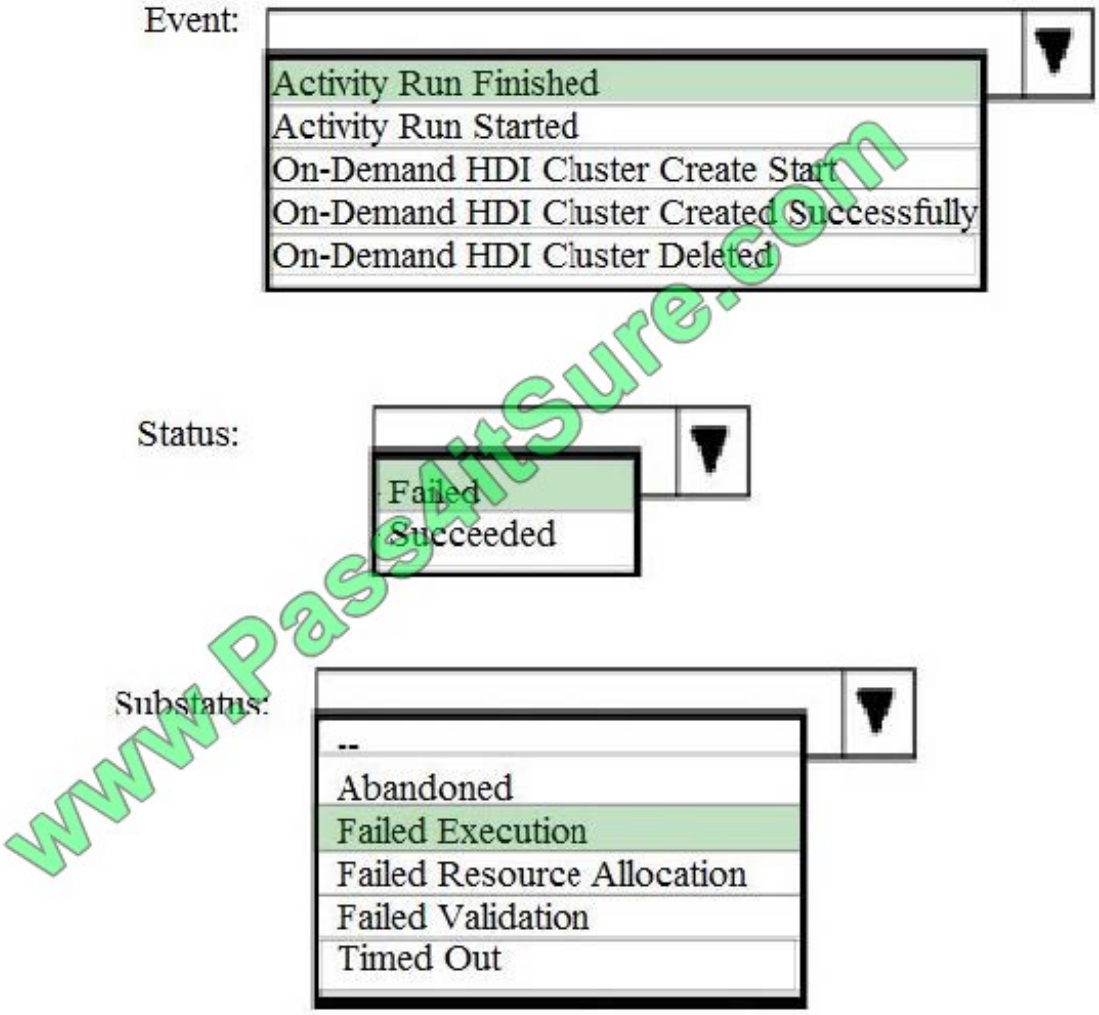
Hot Area:

Answer Area



Correct Answer:

Answer Area



QUESTION 4

Note: The 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 section, you will NOT be able to return to it. As a result, these questions will not appear in the review screen.

Your company has multiple databases that contain millions of sales transactions.

You plan to implement a data mining solution to identify purchasing fraud.

You need to design a solution that mines 10 terabytes (TB) of sales date. The solution must meet the following requirements:

Run the analysis to identify fraud once per week.

Continue to receive new sales transactions while the analysis runs.

Be able to stop computing services when the analysis is NOT running.

Solution: You create a Cloudera Hadoop cluster on Microsoft Azure virtual machines.

Does this meet the goal?

A. Yes

B. No

Correct Answer: A

Processing large amounts of unstructured data requires serious computing power and also maintenance effort. As load on computing power typically fluctuates due to time and seasonal influences and/or processes running on certain times, a cloud solution like Microsoft Azure is a good option to be able to scale up easily and pay only for what is actually used.

Reference: http://blog.cloudera.com/blog/2016/02/how-to-install-cloudera-enterprise-on-microsoft-azure-part-1/

QUESTION 5

You have data generated by sensors. The data is sent to Microsoft Azure Event Hubs.

You need to have an aggregated view of the data in near real time by using five-minute tumbling windows to identify short-term trends. You must also have hourly and a daily aggregated views of the data.

Which technology should you use for each task? To answer, drag the appropriate technologies to the correct tasks. Each technology may be used once, more than once, or not at all. You may need to drag the spilt bar between panes or scroll

to view content.

NOTE: Each correct selection is worth one point.

Select and Place:

Correct Answer:

Box 1: Azure HDInsight MapReduce

Azure Event Hubs allows you to process massive amounts of data from websites, apps, and devices. The Event Hubs spout makes it easy to use Apache Storm on HDInsight to analyze this data in real time.

Box 2: Azure Event Hub

Box 3: Azure Stream Analytics

Stream Analytics is a new service that enables near real time complex event processing over streaming data. Combining Stream Analytics with Azure Event Hubs enables near real time processing of millions of events per second. This

enables you to do things such as augment stream data with reference data and output to storage (or even output to another Azure Event Hub for additional processing).

Reference:

https://blogs.msdn.microsoft.com/kaevans/2015/02/26/using-stream-analytics-with-event-hubs/

https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-storm-develop-csharp-event-hub-topology