



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

-  **Instant Download** After Purchase
-  **100% Money Back** Guarantee
-  **365 Days** Free Update
-  **800,000+** Satisfied Customers





QUESTION 1

You plan to deploy Microsoft Azure HDInsight clusters for business analytics and data pipelines. The clusters must meet the following requirements:

Business users must use a language that is similar to SQL.

The authoring of data pipelines must occur in a dataflow language.

You need to identify which language must be used for each requirement.

Which languages should you identify? To answer, drag the appropriate languages to the correct requirements. Each language 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.

Select and Place:

Languages

Apache HBase

Hive

MapReduce

Pig

Answer Area

Business users must use a language that is similar to SQL:

The authoring of data pipelines must occur in a dataflow language:

Correct Answer:

Languages

Apache HBase

MapReduce

Answer Area

Business users must use a language that is similar to SQL:

The authoring of data pipelines must occur in a dataflow language:



QUESTION 2

You have an Apache Hive cluster in Microsoft Azure HDInsight. The cluster contains 10 million data files.

You plan to archive the data.

The data will be analyzed monthly.

You need to recommend a solution to move and store the data. The solution must minimize how long it takes to move the data and must minimize costs.

Which two services should you include in the recommendation? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. Azure Queue storage
- B. Microsoft SQL Server Integration Services (SSIS)
- C. Azure Table Storage
- D. Azure Data Lake
- E. Azure Data Factory

Correct Answer: DE

D: To analyze data in HDInsight cluster, you can store the data either in Azure Storage, Azure Data Lake Storage Gen 1/Azure Data Lake Storage Gen 2, or both. Both storage options enable you to safely delete HDInsight clusters that are used for computation without losing user data.

E: The Spark activity in a Data Factory pipeline executes a Spark program on your own or on-demand HDInsight cluster. It handles data transformation and the supported transformation activities.

References: <https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-hadoop-use-data-lake-store>
<https://docs.microsoft.com/en-us/azure/data-factory/transform-data-using-spark>

QUESTION 3

You have a Microsoft Azure Data Factory pipeline that contains an input dataset.

You need to ensure that the data from Azure Table Storage is copied only if the table contains 1,000 records or more.

Which policy setting should you use in JSON?

"policy":

- A. { "validation": { "minimum":1000 } }
- B. "policy": { "validation": { "sqlRows":1000 } }
- C. "policy": { "validation": { "minimumRows":1000 } } "policy":



D. { "validation": { "sqlMinimumRows":1000 } }

Correct Answer: B

The following JSON defines a Linux-based on-demand HDInsight linked service. The Data Factory service automatically creates a Linux-based HDInsight cluster to process the required activity.

```
{
  "name": "HDInsightOnDemandLinkedService",
  "properties": {
    "type": "HDInsightOnDemand",
    "typeProperties": {
      "clusterType": "hadoop",
      "clusterSize": 1,
      "timeToLive": "00:15:00",
      "hostSubscriptionId": "",
      "servicePrincipalId": "",
      "servicePrincipalKey": {
        "value": "",
        "type": "SecureString"
      },
      "tenant": "",
      "clusterResourceGroup": "",
      "version": "3.6",
      "osType": "Linux",
      "linkedServiceName": {
        "referenceName": "AzureStorageLinkedService",
        "type": "LinkedServiceReference"
      }
    },
    "connectVia": {
      "referenceName": "",
```



```
"type": "IntegrationRuntimeReference"
```

```
}
```

```
}
```

```
}
```

References: <https://docs.microsoft.com/en-us/azure/data-factory/compute-linked-services>

QUESTION 4

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.

The health tracking application uses the features of a live dashboard to provide historical and trending.

You need to recommend which processing model must be used to process the following types of data:

The top three activities per user on rainy days

The top three activities per user during the last 24 hours

The top activities per geographic region during last 24 hours

The most common sequences of three activities in a row for all of the users

Which processing model should you recommend for each date type? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:



Answer Area

The top three activities per user on rainy days: ▼

Batch processing
Stream processing

The top three activities per user during the last 24 hours: ▼

Batch processing
Stream processing

The top activities per geographic region during last 24 hours: ▼

Batch processing
Stream processing

The most common sequences of three activities in a row for all of the users: ▼

Batch processing
Stream processing

Correct Answer:

Answer Area

The top three activities per user on rainy days: ▼

Batch processing
Stream processing

The top three activities per user during the last 24 hours: ▼

Batch processing
Stream processing

The top activities per geographic region during last 24 hours: ▼

Batch processing
Stream processing

The most common sequences of three activities in a row for all of the users: ▼

Batch processing
Stream processing

QUESTION 5

You need to automate the creation of a new Microsoft Azure data factory.

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

NOTE: Each correct selection is worth one point.

A. Azure PowerShell cmdlets



- B. the SOAP service
- C. T-SQL statements
- D. the REST API
- E. the Microsoft .NET Framework class library

Correct Answer: ADE

You can create a pipeline with a copy activity that moves data to/from an Azure Table Storage by using different tools/APIs.

The easiest way to create a pipeline is to use the Copy Wizard. See Tutorial: Create a pipeline using Copy Wizard for a quick walkthrough on creating a pipeline using the Copy data wizard. You can also use the following tools to create a pipeline: Azure portal, Visual Studio, Azure PowerShell, Azure Resource Manager template, .NET API, and REST API.

Reference: <https://docs.microsoft.com/en-us/azure/data-factory/data-factory-azure-table-connector>

QUESTION 6

You have a Microsoft Azure SQL database that contains Personally Identifiable Information (PII).

To mitigate the PII risk, you need to ensure that data is encrypted while the data is at rest. The solution must minimize any changes to front-end applications.

What should you use?

- A. Transport Layer Security (TLS)
- B. transparent data encryption (TDE)
- C. a shared access signature (SAS)
- D. the ENCRYPTBYPASSPHRASE T-SQL function

Correct Answer: B

Transparent data encryption (TDE) helps protect Azure SQL Database, Azure SQL Managed Instance, and Azure Data Warehouse against the threat of malicious activity. It performs real-time encryption and decryption of the database, associated backups, and transaction log files at rest without requiring changes to the application.

References: <https://docs.microsoft.com/en-us/azure/sql-database/transparent-data-encryption-azure-sql>

QUESTION 7

You are using a Microsoft Azure Data Factory pipeline to copy data to an Azure SQL database.

You need to prevent the insertion of duplicate data for a given dataset slice.

Which two actions should you perform? Each correct answer presents part of the solution.



NOTE: Each correct selection is worth one point.

- A. Set the External property to true.
- B. Add a column named SliceIdentifierColumnName to the output dataset.
- C. Set the SqlWriterCleanupScript property to true.
- D. Remove the duplicates in post-processing.
- E. Manually delete the duplicate data before running the pipeline activity.

Correct Answer: BC

QUESTION 8

You have structured data that resides in Microsoft Azure Blob storage.

You need to perform a rapid interactive analysis of the data and to generate visualizations of the data.

What is the best type of Azure HDInsight cluster to use to achieve the goal? More than one answer choice may achieve the goal. Choose the BEST answer.

- A. Apache Storm
- B. Apache HBase
- C. Apache Hadoop
- D. Apache Spark

Correct Answer: D

A Spark cluster provides in-memory processing, interactive queries, micro-batch stream processing-

Reference: <https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-hadoop-provision-linux-clusters>

QUESTION 9

You have a Microsoft Azure Machine Learning Solution that contains several Azure Data Factory pipeline jobs.

You discover that the jobs for a dataset named CustomerSalesData fails.

You resolve the issue that caused the job to fail.

You need to rerun the slices for CustomerSalesData.

What should you do?

- A. Run the Set-AzureRMDataFactorySliceStatus cmdlet and specify the--Status Retry parameter.
- B. Run the Set-AzureRMDataFactorySliceStatus cmdlet and specify the--Status PendingExecution parameter.



C. Run the Resume-AzureRMDDataFactoryPipeline cmdlet and specify the--Status Retry parameter.

D. Run the Resume-AzureRMDDataFactoryPipeline cmdlet and specify the--Status PendingExecution parameter.

Correct Answer: B

QUESTION 10

Note: This 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

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 plan to implement a new data warehouse.

You have the following information regarding the data warehouse:

The first data files for the data warehouse will be available in a few days.

Most queries that will be executed against the data warehouse are ad-hoc.

The schemas of data files that will be loaded to the data warehouse change often.

One month after the planned implementation, the data warehouse will contain 15 TB of data.

You need to recommend a database solution to support the planned implementation.

Solution: You recommend an Apache Hadoop system.

Does this meet the goal?

A. Yes

B. No

Correct Answer: A

QUESTION 11

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 12

You deploy a Microsoft Azure SQL database.

You create a job to upload customer data to the database.

You discover that the job cannot connect to the database and fails.

You verify that the database runs successfully in Azure.

You need to run the job successfully.

What should you create?

- A. a virtual network rule
- B. a network security group (NSG)
- C. a firewall rule
- D. a virtual network

Correct Answer: C

If the application persistently fails to connect to Azure SQL Database, it usually indicates an issue with one of the following:

Firewall configuration. The Azure SQL database or client-side firewall is blocking connections to Azure SQL Database.

Network reconfiguration on the client side: for example, a new IP address or a proxy server.

User error: for example, mistyped connection parameters, such as the server name in the connection string.

References: <https://docs.microsoft.com/en-us/azure/sql-database/sql-database-troubleshoot-common-connection-issues>



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.pass4itsure.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 © pass4itsure, All Rights Reserved.