



DP-203^{Q&As}

Data Engineering on Microsoft Azure

Pass Microsoft DP-203 Exam with 100% Guarantee

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

<https://www.lead4pass.com/dp-203.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 are designing an anomaly detection solution for streaming data from an Azure IoT hub. The solution must meet the following requirements:

1.
Send the output to Azure Synapse.
2.
Identify spikes and dips in time series data.
3.
Minimize development and configuration effort. Which should you include in the solution?
A. Azure Databricks
B. Azure Stream Analytics
C. Azure SQL Database

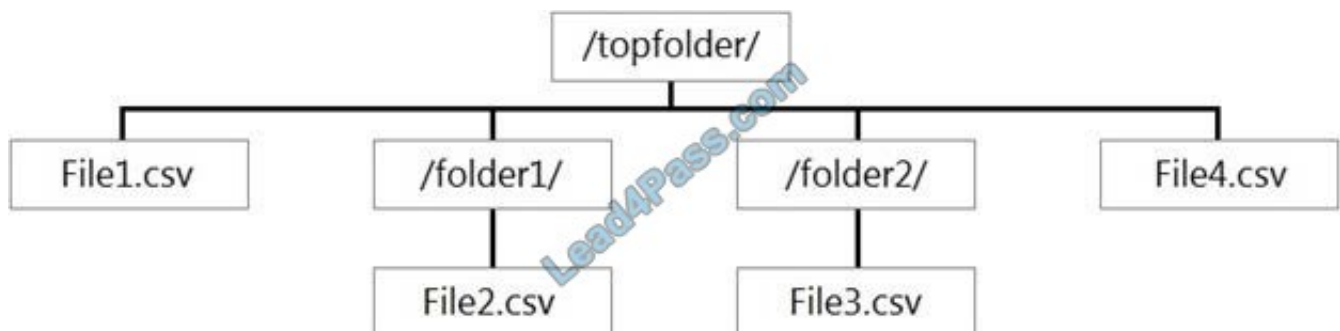
Correct Answer: B

You can identify anomalies by routing data via IoT Hub to a built-in ML model in Azure Stream Analytics.

Reference: <https://docs.microsoft.com/en-us/learn/modules/data-anomaly-detection-using-azure-iot-hub/>

QUESTION 2

You have files and folders in Azure Data Lake Storage Gen2 for an Azure Synapse workspace as shown in the following exhibit.



You create an external table named ExtTable that has LOCATION=\\'\topfolder\\'.

When you query ExtTable by using an Azure Synapse Analytics serverless SQL pool, which files are returned?

- A. File2.csv and File3.csv only
- B. File1.csv and File4.csv only



- C. File1.csv, File2.csv, File3.csv, and File4.csv
- D. File1.csv only

Correct Answer: B

To run a T-SQL query over a set of files within a folder or set of folders while treating them as a single entity or rowset, provide a path to a folder or a pattern (using wildcards) over a set of files or folders.

Reference: <https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/query-data-storage#query-multiple-files-or-folders>

QUESTION 3

HOTSPOT

You have an Azure Storage account that generates 200,000 new files daily. The file names have a format of (YYY)/(MM)/(DD)/(HH)/(CustomerID).csv.

You need to design an Azure Data Factory solution that will load new data from the storage account to an Azure Data lake once hourly. The solution must minimize load times and costs.

How should you configure the solution? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:

Load methodology
<input type="checkbox"/> incremental load
<input type="checkbox"/> placeholder

Trigger
<input type="checkbox"/> Tumbling window
<input type="checkbox"/> placeholder

Correct Answer:



Load methodology
incremental load
placeholder

Trigger
Tumbling window
placeholder

QUESTION 4

You have an Azure subscription linked to an Azure Active Directory (Azure AD) tenant that contains a service principal named ServicePrincipal1. The subscription contains an Azure Data Lake Storage account named adls1. Adls1 contains a folder named Folder2 that has a URI of <https://adls1.dfs.core.windows.net/container1/Folder1/Folder2/>.

ServicePrincipal1 has the access control list (ACL) permissions shown in the following table.

Resource	Permission
container1	Access – Execute
Folder1	Access – Execute
Folder2	Access – Read

You need to ensure that ServicePrincipal1 can perform the following actions:

1. Traverse child items that are created in Folder2.
2. Read files that are created in Folder2.

The solution must use the principle of least privilege.



Which two permissions should you grant to ServicePrincipal1 for Folder2? Each correct answer presents part of the solution.

NOTE: Each correct selection is worth one point.

- A. Access - Read
- B. Access - Write
- C. Access - Execute
- D. Default-Read
- E. Default - Write
- F. Default - Execute

Correct Answer: DF

Execute (X) permission is required to traverse the child items of a folder. There are two kinds of access control lists (ACLs), Access ACLs and Default ACLs. Access ACLs: These control access to an object. Files and folders both have

Access ACLs.

Default ACLs: A "template" of ACLs associated with a folder that determine the Access ACLs for any child items that are created under that folder. Files do not have Default ACLs.

Reference:

<https://docs.microsoft.com/en-us/azure/data-lake-store/data-lake-store-access-control>

QUESTION 5

You have a SQL pool in Azure Synapse that contains a table named dbo.Customers. The table contains a column name Email.

You need to prevent nonadministrative users from seeing the full email addresses in the Email column. The users must see values in a format of a XXX@XXXX.com instead.

What should you do?

- A. From Microsoft SQL Server Management Studio, set an email mask on the Email column.
- B. From the Azure portal, set a mask on the Email column.
- C. From Microsoft SQL Server Management Studio, grant the SELECT permission to the users for all the columns in the dbo.Customers table except Email.
- D. From the Azure portal, set a sensitivity classification of Confidential for the Email column.

Correct Answer: A

From Microsoft SQL Server Management Studio, set an email mask on the Email column. This is because "This feature cannot be set using portal for Azure Synapse (use PowerShell or REST API) or SQL Managed Instance." So use Create table statement with Masking e.g. CREATE TABLE Membership (MemberID int IDENTITY PRIMARY KEY, FirstName



varchar(100) MASKED WITH (FUNCTION = '\partial(1,"XXXXXXX",0)\') NULL, . . upvoted 24 times

aXX@XXXX.com

Reference: <https://docs.microsoft.com/en-us/azure/azure-sql/database/dynamic-data-masking-overview>

QUESTION 6

HOTSPOT

You have two Azure Storage accounts named Storage1 and Storage2. Each account holds one container and has the hierarchical namespace enabled. The system has files that contain data stored in the Apache Parquet format.

You need to copy folders and files from Storage1 to Storage2 by using a Data Factory copy activity. The solution must meet the following requirements:

1.

No transformations must be performed.

2.

The original folder structure must be retained.

3.

Minimize time required to perform the copy activity.

How should you configure the copy activity? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:



Answer Area

Source dataset type:

	▼
Binary	
Parquet	
Delimited text	

Copy activity copy behavior:

	▼
FlattenHierarchy	
MergeFiles	
PreserveHierarchy	

Correct Answer:

Answer Area

Source dataset type:

	▼
Binary	
Parquet	
Delimited text	

Copy activity copy behavior:

	▼
FlattenHierarchy	
MergeFiles	
PreserveHierarchy	

Box 1: Parquet

For Parquet datasets, the type property of the copy activity source must be set to ParquetSource.

Box 2: PreserveHierarchy

PreserveHierarchy (default): Preserves the file hierarchy in the target folder. The relative path of the source file to the source folder is identical to the relative path of the target file to the target folder.



Incorrect Answers:

FlattenHierarchy: All files from the source folder are in the first level of the target folder. The target files have autogenerated names.

MergeFiles: Merges all files from the source folder to one file. If the file name is specified, the merged file name is the specified name. Otherwise, it's an autogenerated file name.

Reference:

<https://docs.microsoft.com/en-us/azure/data-factory/format-parquet>

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

QUESTION 7

You have an Azure Databricks workspace named workspace1 in the Standard pricing tier. Workspace contains an all-purpose cluster named cluster1. You need to reduce the time it takes for cluster1 to start and scale up. The solution must minimize costs. What should you do first?

- A. Upgrade workspace1 to the Premium pricing tier.
- B. Create a cluster policy in workspace1.
- C. Create a pool in workspace1.
- D. Configure a global init script for workspace1.

Correct Answer: C

You can use Databricks Pools to Speed up your Data Pipelines and Scale Clusters Quickly.

Databricks Pools, a managed cache of virtual machine instances that enables clusters to start and scale 4 times faster.

Reference:

<https://databricks.com/blog/2019/11/11/databricks-pools-speed-up-data-pipelines.html>

QUESTION 8

You have an Azure Stream Analytics job that receives clickstream data from an Azure event hub.

You need to define a query in the Stream Analytics job. The query must meet the following requirements:

Count the number of clicks within each 10-second window based on the country of a visitor. Ensure that each click is NOT counted more than once.

How should you define the Query?

- A. `SELECT Country, Avg(*) AS Average FROM ClickStream TIMESTAMP BY CreatedAt GROUP BY Country, SlidingWindow(second, 10)`
- B. `SELECT Country, Count(*) AS Count FROM ClickStream TIMESTAMP BY CreatedAt GROUP BY Country,`



TumblingWindow(second, 10)

C. SELECT Country, Avg(*) AS Average FROM ClickStream TIMESTAMP BY CreatedAt GROUP BY Country, HoppingWindow(second, 10, 2)

D. SELECT Country, Count(*) AS Count FROM ClickStream TIMESTAMP BY CreatedAt GROUP BY Country, SessionWindow(second, 5, 10)

Correct Answer: B

Tumbling window functions are used to segment a data stream into distinct time segments and perform a function against them, such as the example below. The key differentiators of a Tumbling window are that they repeat, do not overlap,

and an event cannot belong to more than one tumbling window.

Example:

Incorrect Answers:

A: Sliding windows, unlike Tumbling or Hopping windows, output events only for points in time when the content of the window actually changes. In other words, when an event enters or exits the window. Every window has at least one event, like in the case of Hopping windows, events can belong to more than one sliding window.

C: Hopping window functions hop forward in time by a fixed period. It may be easy to think of them as Tumbling windows that can overlap, so events can belong to more than one Hopping window result set. To make a Hopping window the same as a Tumbling window, specify the hop size to be the same as the window size.

D: Session windows group events that arrive at similar times, filtering out periods of time where there is no data.

Reference: <https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-window-functions>

QUESTION 9

You have an Azure Data Lake Storage Gen2 account named adls2 that is protected by a virtual network.

You are designing a SQL pool in Azure Synapse that will use adls2 as a source.

What should you use to authenticate to adls2?

A. an Azure Active Directory (Azure AD) user

B. a shared key

C. a shared access signature (SAS)

D. a managed identity

Correct Answer: D

Managed identity for Azure resources is a feature of Azure Active Directory. The feature provides Azure services with an automatically managed identity in Azure AD. You can use the Managed Identity capability to authenticate to any service that support Azure AD authentication.

Managed Identity authentication is required when your storage account is attached to a VNet.



Reference: <https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/quickstart-bulk-load-copy-tsql-examples>

QUESTION 10

DRAG DROP

You have an Azure Stream Analytics job that is a Stream Analytics project solution in Microsoft Visual Studio. The job accepts data generated by IoT devices in the JSON format.

You need to modify the job to accept data generated by the IoT devices in the Protobuf format.

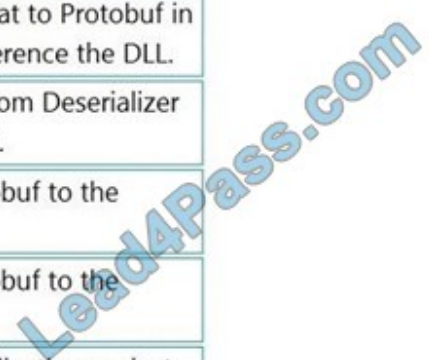
Which three actions should you perform from Visual Studio on sequence? To answer, move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

Select and Place:

Actions

- Change the Event Serialization Format to Protobuf in the input.json file of the job and reference the DLL.
- Add an Azure Stream Analytics Custom Deserializer Project (.NET) project to the solution.
- Add .NET deserializer code for Protobuf to the custom deserializer project.
- Add .NET deserializer code for Protobuf to the Stream Analytics project.
- Add an Azure Stream Analytics Application project to the solution.

Answer Area



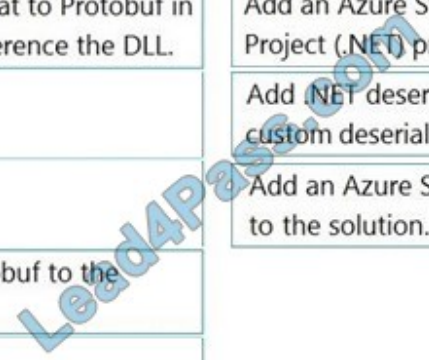
Correct Answer:

Actions

- Change the Event Serialization Format to Protobuf in the input.json file of the job and reference the DLL.
-
-
- Add .NET deserializer code for Protobuf to the Stream Analytics project.
-

Answer Area

- Add an Azure Stream Analytics Custom Deserializer Project (.NET) project to the solution.
- Add .NET deserializer code for Protobuf to the custom deserializer project.
- Add an Azure Stream Analytics Application project to the solution.



Step 1: Add an Azure Stream Analytics Custom Deserializer Project (.NET) project to the solution. Create a custom deserializer



1.

Open Visual Studio and select File > New > Project. Search for Stream Analytics and select Azure Stream Analytics Custom Deserializer Project (.NET). Give the project a name, like Protobuf Deserializer.

2.

In Solution Explorer, right-click your Protobuf Deserializer project and select Manage NuGet Packages from the menu. Then install the Microsoft.Azure.StreamAnalytics and Google.Protobuf NuGet packages.

3.

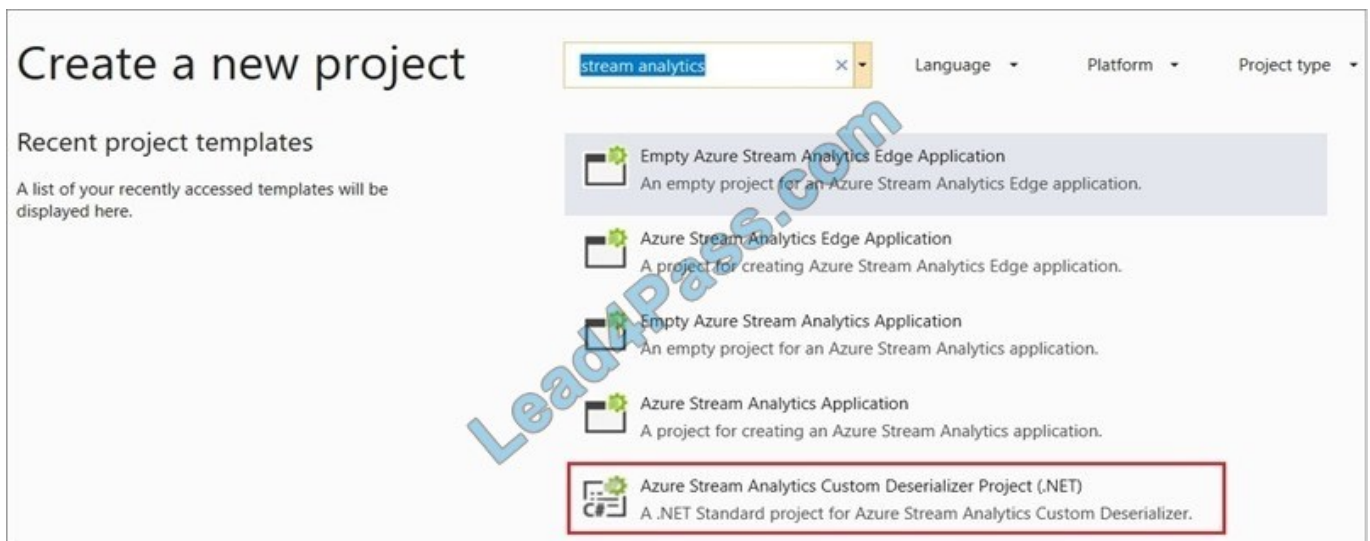
Add the MessageBodyProto class and the MessageBodyDeserializer class to your project.

4.

Build the Protobuf Deserializer project.

Step 2: Add .NET deserializer code for Protobuf to the custom deserializer project Azure Stream Analytics has built-in support for three data formats: JSON, CSV, and Avro. With custom .NET deserializers, you can read data from other

formats such as Protocol Buffer, Bond and other user defined formats for both cloud and edge jobs.



Step 3: Add an Azure Stream Analytics Application project to the solution Add an Azure Stream Analytics project In Solution Explorer, right-click the Protobuf Deserializer solution and select Add > New Project. Under Azure Stream Analytics > Stream Analytics, choose Azure Stream Analytics Application. Name it ProtobufCloudDeserializer and select OK. Right-click References under the ProtobufCloudDeserializer Azure Stream Analytics project. Under Projects, add Protobuf Deserializer. It should be automatically populated for you.

QUESTION 11

You use Azure Stream Analytics to receive Twitter data from Azure Event Hubs and to output the data to an Azure Blob storage account.

You need to output the count of tweets during the last five minutes every five minutes.

Each tweet must only be counted once.



Which windowing function should you use?

- A. a five-minute Session window
- B. B. a five-minute Sliding window
- C. a five-minute Tumbling window
- D. a five-minute Hopping window that has one-minute hop

Correct Answer: C

Tumbling window functions are used to segment a data stream into distinct time segments and perform a function against them, such as the example below. The key differentiators of a Tumbling window are that they repeat, do not overlap, and an event cannot belong to more than one tumbling window.

References: <https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-window-functions>

QUESTION 12

You have an Azure Data Factory that contains 10 pipelines.

You need to label each pipeline with its main purpose of either ingest, transform, or load. The labels must be available for grouping and filtering when using the monitoring experience in Data Factory.

What should you add to each pipeline?

- A. a resource tag
- B. a correlation ID
- C. a run group ID
- D. an annotation

Correct Answer: D

Annotations are additional, informative tags that you can add to specific factory resources: pipelines, datasets, linked services, and triggers. By adding annotations, you can easily filter and search for specific factory resources.

Reference: <https://www.cathrinewilhelmsen.net/annotations-user-properties-azure-data-factory/>

QUESTION 13

HOTSPOT You plan to create an Azure Data Lake Storage Gen2 account. You need to recommend a storage solution that meets the following requirements:

1.
Provides the highest degree of data resiliency
- 2.



Ensures that content remains available for writes if a primary data center fails

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

Hot Area:

Replication mechanism:
<input type="checkbox"/> Zone-redundant storage (ZRS)
<input type="checkbox"/> Placeholder

Failover process:
<input type="checkbox"/> Failover manually initiated by the customer
<input type="checkbox"/> Placeholder

Correct Answer:

Replication mechanism:
<input checked="" type="checkbox"/> Zone-redundant storage (ZRS)
<input type="checkbox"/> Placeholder

Failover process:
<input checked="" type="checkbox"/> Failover manually initiated by the customer
<input type="checkbox"/> Placeholder



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.