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.pass4itsure.com/dp-203.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 an Azure subscription that contains an Azure Data Lake Storage account named myaccount1. The myaccount1 account contains two containers named container1 and contained. The subscription is linked to an Azure Active Directory (Azure AD) tenant that contains a security group named Group1.

You need to grant Group1 read access to contamer1. The solution must use the principle of least privilege. Which role should you assign to Group1?

A. Storage Blob Data Reader for container1

B. Storage Table Data Reader for container1

C. Storage Blob Data Reader for myaccount1

D. Storage Table Data Reader for myaccount1

Correct Answer: A

QUESTION 2

You have an Azure Databricks workspace named workspace! in the Standard pricing tier. Workspace contains an all-purpose cluster named cluster.

You need to reduce the time it takes for cluster 1 to start and scale up. The solution must minimize costs.

What should you do first?

A. Upgrade workspace! 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 3

You have an Azure Synapse Analytics dedicated SQL pool.

You need to create a pipeline that will execute a stored procedure in the dedicated SQL pool and use the returned result set as the input for a downstream activity. The solution must minimize development effort.

Which type of activity should you use in the pipeline?

A. U-SQL

B. Stored Procedure

C. Script

D. Notebook

Correct Answer: B

You can use the Stored Procedure Activity to invoke a stored procedure in one of the following data stores in your enterprise or on an Azure virtual machine (VM): Azure SQL Database

Azure Synapse Analytics SQL Server Database. Note: Create a Stored Procedure activity with UI

To use a Stored Procedure activity in a pipeline, complete the following steps:

Search for Stored Procedure in the pipeline Activities pane, and drag a Stored Procedure activity to the pipeline canvas.

Select the new Stored Procedure activity on the canvas if it is not already selected, and its Settings tab, to edit its details.

Select an existing or create a new linked service to an Azure SQL Database, Azure Synapse Analytics, or SQL Server.

Choose a stored procedure, and provide any parameters for its execution.

Incorrect:

* U-SQL

You can process data by running U-SQL scripts on Azure Data Lake Analytics with Azure Data Factory and Synapse Analytics.

Reference:

https://learn.microsoft.com/en-us/azure/data-factory/transform-data-using-stored-procedure

QUESTION 4

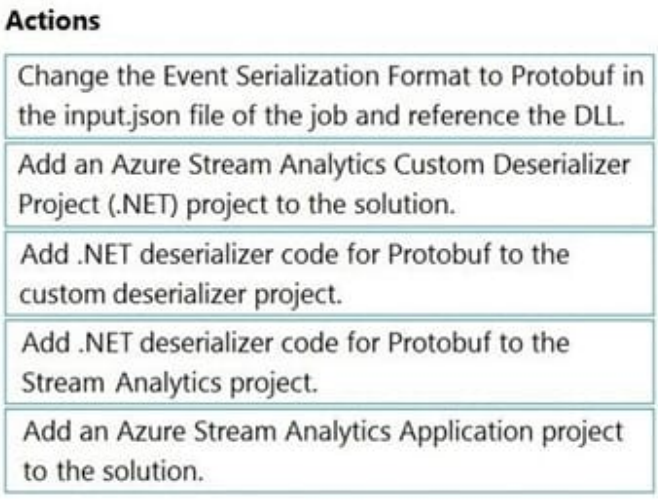
DRAG DROP

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.

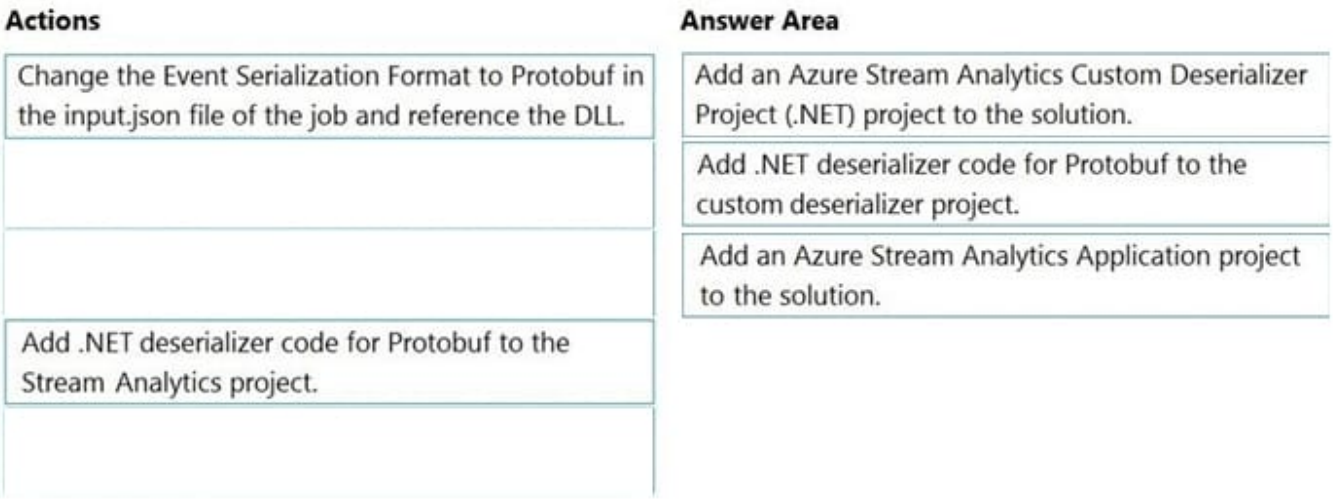
You need to modify the job to accept data generated by the IoT devices in the Protobuf format.

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:



Correct Answer:



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.

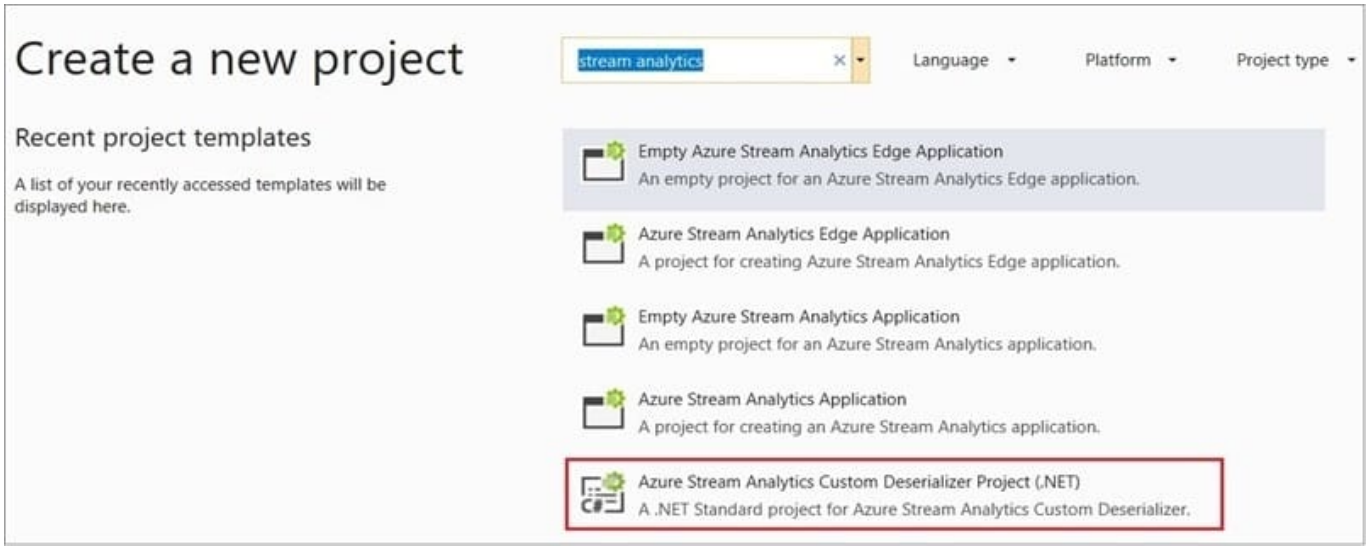
Add the MessageBodyProto class and the MessageBodyDeserializer class to your project.

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 5

DRAG DROP

You have an Azure Data Lake Storage Gen2 account that contains a JSON file for customers. The file contains two attributes named FirstName and LastName.

You need to copy the data from the JSON file to an Azure Synapse Analytics table by using Azure Databricks. A new column must be created that concatenates the FirstName and LastName values.

You create the following components:

1.

A destination table in Azure Synapse

2.

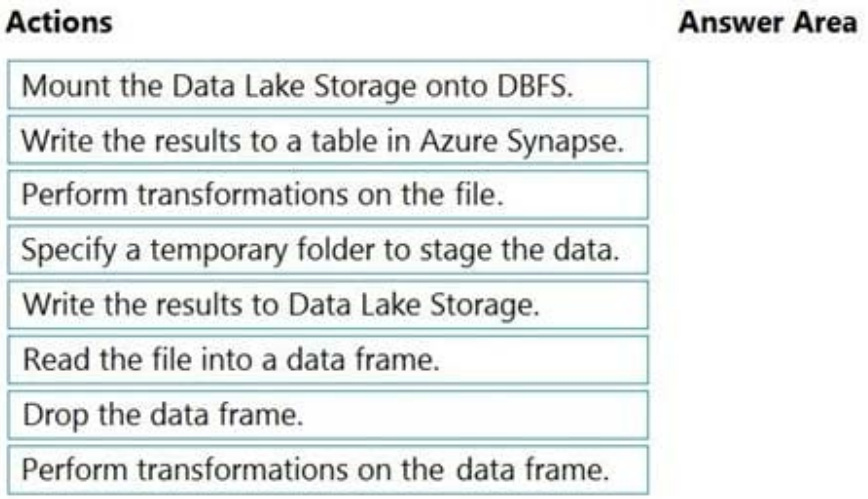
An Azure Blob storage container

3.

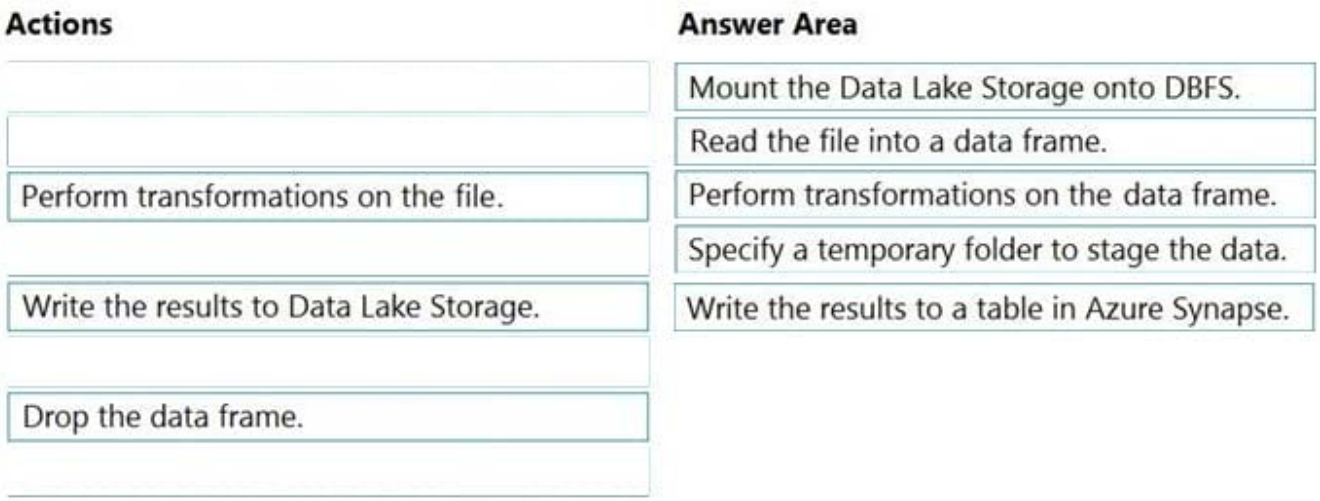
A service principal

Which five actions should you perform in sequence next in is Databricks notebook? 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:



Correct Answer:



Step 1: Mount the Data Lake Storage onto DBFS

Begin with creating a file system in the Azure Data Lake Storage Gen2 account.

Step 2: Read the file into a data frame.

You can load the json files as a data frame in Azure Databricks.

Step 3: Perform transformations on the data frame.

Step 4: Specify a temporary folder to stage the data Specify a temporary folder to use while moving data between Azure Databricks and Azure Synapse.

Step 5: Write the results to a table in Azure Synapse.

You upload the transformed data frame into Azure Synapse. You use the Azure Synapse connector for Azure

Databricks to directly upload a dataframe as a table in a Azure Synapse. https://docs.databricks.com/data/data-sources/azure/azure-datalake-gen2.html https://docs.microsoft.com/en-us/azure/databricks/scenarios/databricks-extract-load-sql-data-warehouse

QUESTION 6

You have a data warehouse in Azure Synapse Analytics.

You need to ensure that the data in the data warehouse is encrypted at rest.

What should you enable?

A. Advanced Data Security for this database

B. Transparent Data Encryption (TDE)

C. Secure transfer required

D. Dynamic Data Masking

Correct Answer: B

Azure SQL Database currently supports encryption at rest for Microsoft-managed service side and client-side encryption scenarios.

1.

Support for server encryption is currently provided through the SQL feature called Transparent Data Encryption.

2.

Client-side encryption of Azure SQL Database data is supported through the Always Encrypted feature.

Reference: https://docs.microsoft.com/en-us/azure/security/fundamentals/encryption-atrest

QUESTION 7

You have an Azure SQL database named DB1 and an Azure Data Factory data pipeline named pipeline.

From Data Factory, you configure a linked service to DB1.

In DB1, you create a stored procedure named SP1. SP1 returns a single row of data that has four columns.

You need to add an activity to pipeline to execute SP1. The solution must ensure that the values in the columns are stored as pipeline variables.

Which two types of activities can you use to execute SP1? (Refer to Data Engineering on Microsoft Azure documents or guide for Answers/available at Microsoft.com)

A. Script

B. Copy

C. Lookup

D. Stored Procedure

Correct Answer: AD

A: You use data transformation activities in a Data Factory or Synapse pipeline to transform and process raw data into predictions and insights. The Script activity is one of the transformation activities that pipelines support.

You can use the Script activity to invoke a SQL script in one of the following data stores in your enterprise or on an Azure virtual machine (VM):

Azure SQL Database Azure Synapse Analytics SQL Server Database. Oracle Snowflake

The script may contain either a single SQL statement or multiple SQL statements that run sequentially. You can use the Script task for the following purposes:

Truncate a table in preparation for inserting data.

Create, alter, and drop database objects such as tables and views.

Re-create fact and dimension tables before loading data into them.

*-> Run stored procedures. If the SQL statement invokes a stored procedure that returns results from a temporary table, use the WITH RESULT SETS option to define metadata for the result set.

Save the rowset returned from a query as activity output for downstream consumption.

D: You can transform data by using the SQL Server Stored Procedure activity in Azure Data Factory or Synapse Analytics.

You use data transformation activities in a Data Factory or Synapse pipeline to transform and process raw data into predictions and insights. The Stored Procedure Activity is one of the transformation activities that pipelines support.

You can use the Stored Procedure Activity to invoke a stored procedure in one of the following data stores in your enterprise or on an Azure virtual machine (VM):

Azure SQL Database Azure Synapse Analytics SQL Server Database.

Reference:

https://learn.microsoft.com/en-us/azure/data-factory/transform-data-using-script

https://learn.microsoft.com/en-us/azure/data-factory/transform-data-using-stored-procedure¨

QUESTION 8

HOTSPOT

You have an Azure Synapse Analytics dedicated SQL pool that contains the users shown in the following table.

Name	Role
User1	Server admin
User2	db_datereader

User1 executes a query on the database, and the query returns the results shown in the following exhibit.

```
SELECT c.name,
    tbl.name as table_name,
    typ.name as datatype,
    c.is_masked,
    c.masking_function
FROM sys.masked_columns AS c
INNER JOIN sys.tables AS tbl ON c.[object_id] = tbl.[object_id]
INNER JOIN sys.types typ ON c.user_type_id = typ.user_type_id
WHERE is_masked = 1;
```

Results Messages

	name	table_name	datatype	is_masked	masking_function
1	BirthDate	DimCustomer	date	1	default()
2	Gender	DimCustomer	nvarchar	1	default()
3	EmailAddress	DimCustomer	nvarchar	1	email()
4	YearlyIncome	DimCustomer	money	1	default()

User1 is the only user who has access to the unmasked data.

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

NOTE: Each correct selection is worth one point.

Hot Area:

When User2 queries the YearlyIncome column, the values returned will be **[answer choice]**.

- a random number
- the values stored in the database
- XXXX
- 0

When User1 queries the BirthDate column, the values returned will be **[answer choice]**.

- a random date
- the values stored in the database
- XXXX
- 1900-01-01

Correct Answer:

When User2 queries the YearlyIncome column, the values returned will be **[answer choice]**.

- a random number
- the values stored in the database
- XXXX
- 0 ✓

When User1 queries the BirthDate column, the values returned will be **[answer choice]**.

- a random date
- the values stored in the database ✓
- XXXX
- 1900-01-01

Box 1: 0

The YearlyIncome column is of the money data type.

The Default masking function: Full masking according to the data types of the designated fields

Use a zero value for numeric data types (bigint, bit, decimal, int, money, numeric, smallint, smallmoney, tinyint, float, real).

Box 2: the values stored in the database

Users with administrator privileges are always excluded from masking, and see the original data without any mask.

Reference:

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

QUESTION 9

HOTSPOT

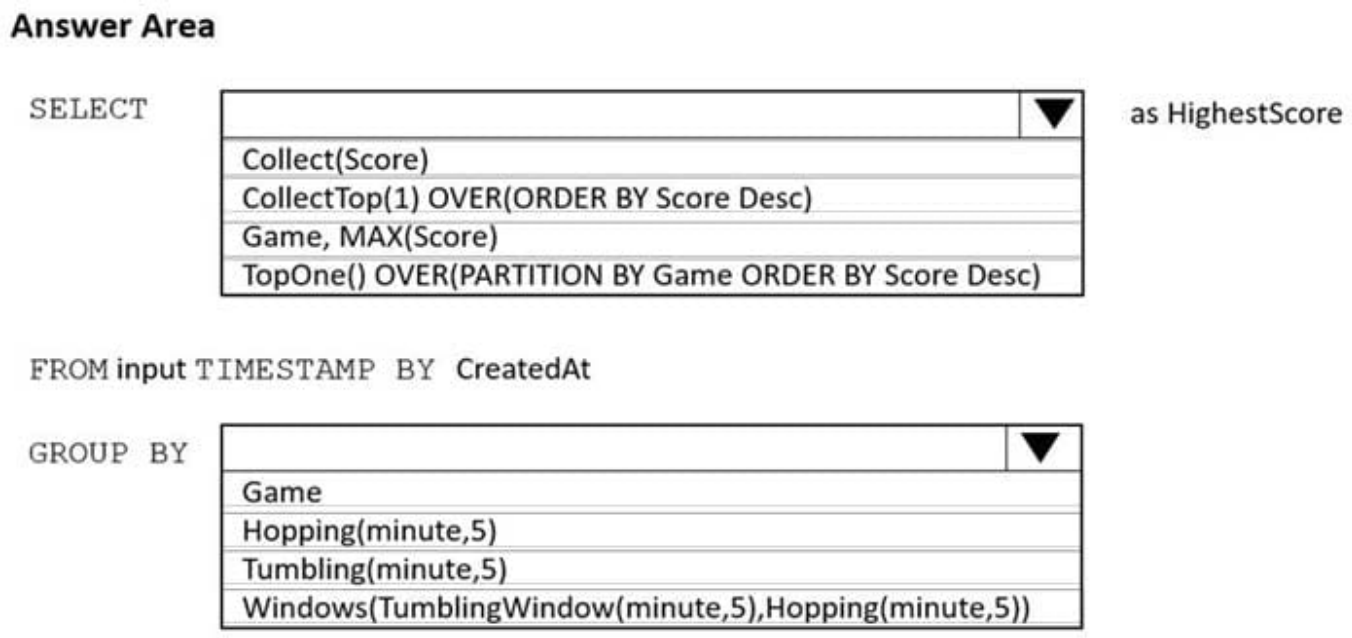
You are building an Azure Stream Analytics job to retrieve game data.

You need to ensure that the job returns the highest scoring record for each five-minute time interval of each game.

How should you complete the Stream Analytics query? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:



Correct Answer:

Answer Area

```
SELECT [ ▼ ] as HighestScore
```

- Collect(Score)
- CollectTop(1) OVER(ORDER BY Score Desc)
- Game, MAX(Score)
- TopOne() OVER(PARTITION BY Game ORDER BY Score Desc)

```
FROM input TIMESTAMP BY CreatedAt

GROUP BY [ ▼ ]
```

- Game
- Hopping(minute,5)
- Tumbling(minute,5)
- Windows(TumblingWindow(minute,5),Hopping(minute,5))

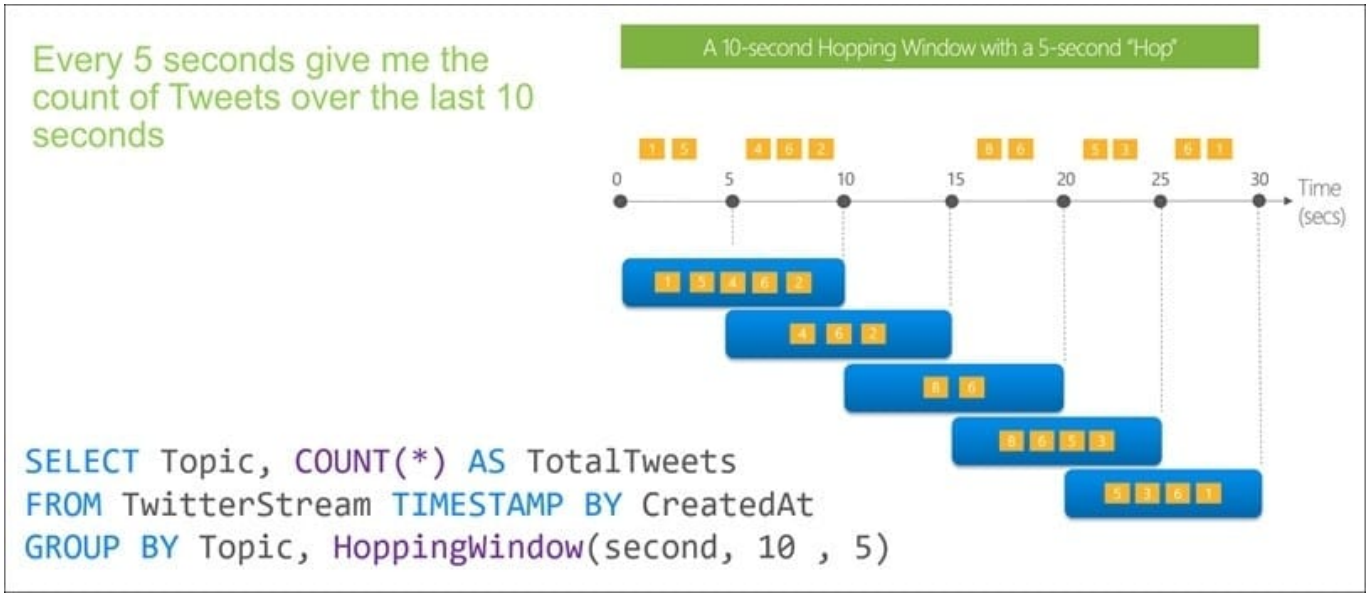
Box 1: TopOne OVER(PARTITION BY Game ORDER BY Score Desc) TopOne returns the top-rank record, where rank defines the ranking position of the event in the window according to the specified ordering. Ordering/ranking is based on

event columns and can be specified in ORDER BY clause.

Box 2: Hopping(minute,5)

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 and be emitted more often than the window size. 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.



QUESTION 10

HOTSPOT

You are developing a solution using a Lambda architecture on Microsoft Azure.

The data at test layer must meet the following requirements:

Data storage:

1.

Serve as a repository (or high volumes of large files in various formats.

2.

Implement optimized storage for big data analytics workloads.

3.

Ensure that data can be organized using a hierarchical structure.

Batch processing:

1.

Use a managed solution for in-memory computation processing.

2.

Natively support Scala, Python, and R programming languages.

3.

Provide the ability to resize and terminate the cluster automatically.

Analytical data store:

1.

Support parallel processing.

2.

Use columnar storage.

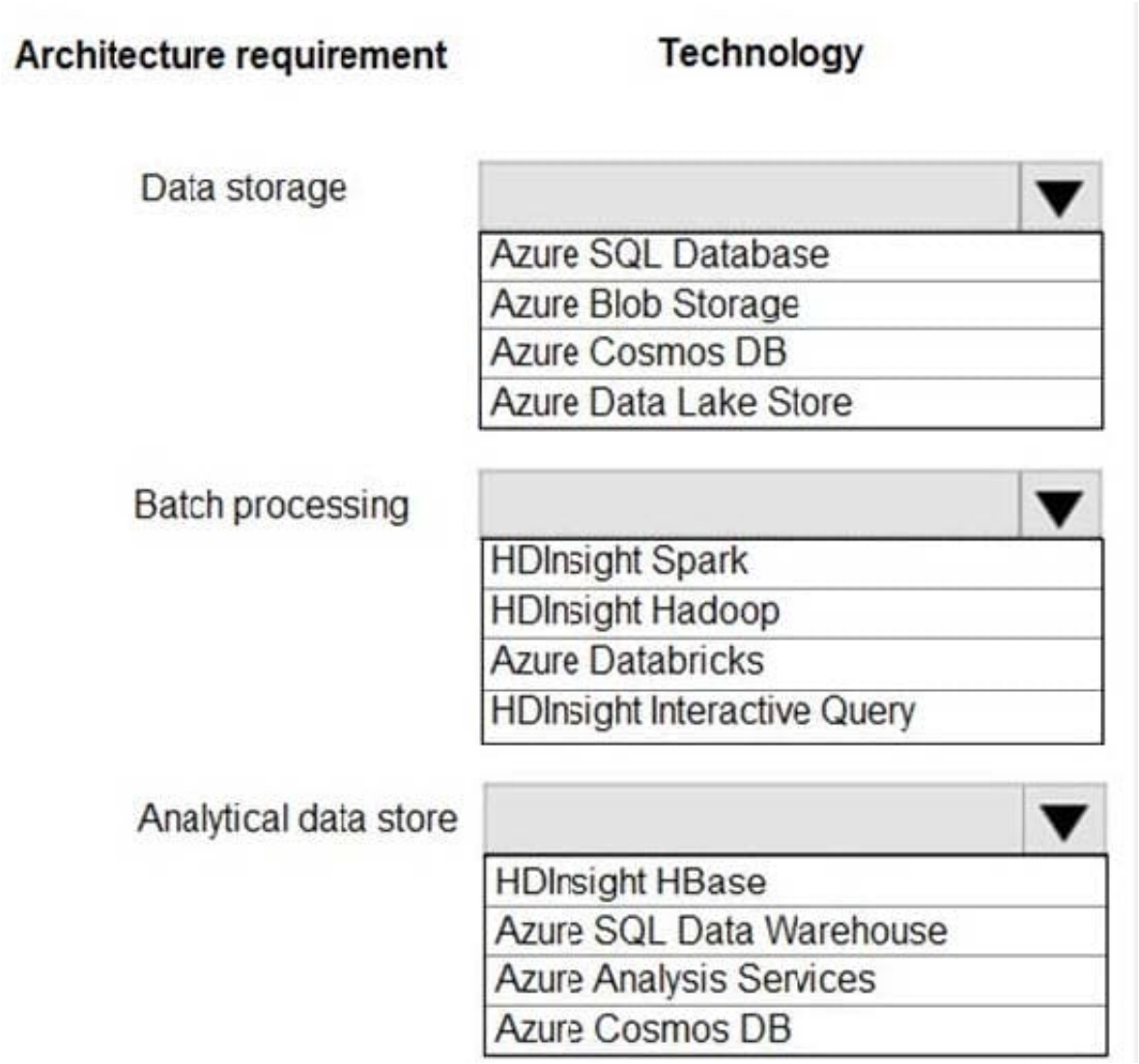
3.

Support SQL-based languages.

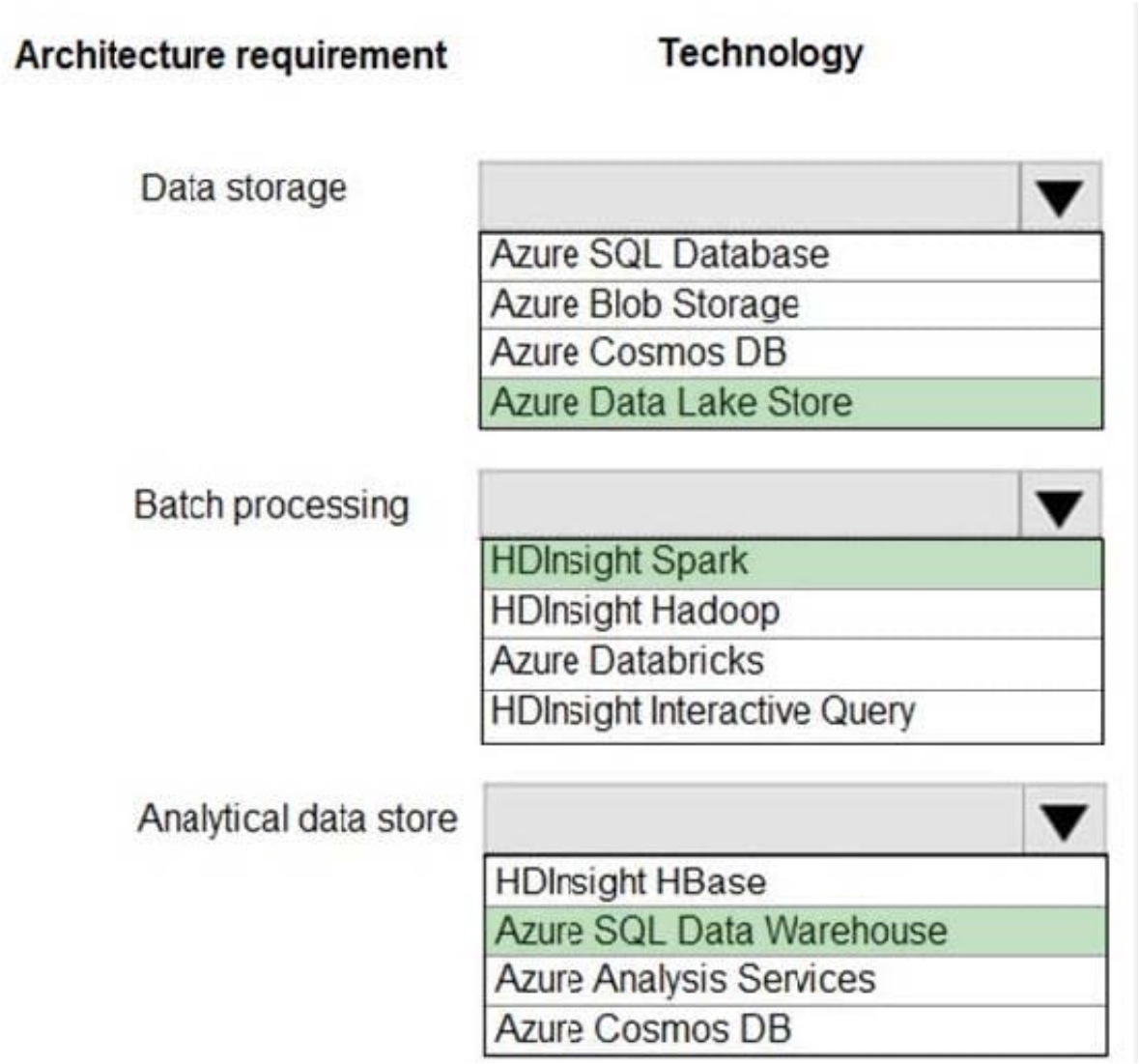
You need to identify the correct technologies to build the Lambda architecture.

Which technologies should you use? To answer, select the appropriate options in the answer area NOTE: Each correct selection is worth one point.

Hot Area:



Correct Answer:



Data storage: Azure Data Lake Store

A key mechanism that allows Azure Data Lake Storage Gen2 to provide file system performance at object storage scale and prices is the addition of a hierarchical namespace.

This allows the collection of objects/files within an account to be organized into a hierarchy of directories and nested subdirectories in the same way that the file system on your computer is organized. With the hierarchical namespace enabled,

a storage account becomes capable of providing the scalability and cost-effectiveness of object storage, with

file system semantics that are familiar to analytics engines and frameworks.

Batch processing: HD Insight Spark

Aparch Spark is an open-source, parallel-processing framework that supports in-memory processing to boost the performance of big-data analysis applications.

HDInsight is a managed Hadoop service. Use it deploy and manage Hadoop clusters in Azure. For batch processing, you can use Spark, Hive, Hive LLAP, MapReduce.

Languages: R, Python, Java, Scala, SQL

Analytic data store: SQL Data Warehouse

SQL Data Warehouse is a cloud-based Enterprise Data Warehouse (EDW) that uses Massively Parallel Processing (MPP).

SQL Data Warehouse stores data into relational tables with columnar storage.

References:

https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-namespace

https://docs.microsoft.com/en-us/azure/architecture/data-guide/technology-choices/batchprocessing

https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-overviewwhat-is

QUESTION 11

You are designing a partition strategy for a fact table in an Azure Synapse Analytics dedicated SQL pool. The table has the following specifications:

1.

Contain sales data for 20,000 products.

2.

Use hash distribution on a column named ProduclID,

3.

Contain 2.4 billion records for the years 20l9 and 2020.

Which number of partition ranges provides optimal compression and performance of the clustered columnstore index?

A. 40

B. 240

C. 400

D. 2400

Correct Answer: A

Each partition should have around 1 millions records. Dedication SQL pools already have 60 partitions.

We have the formula: Records/(Partitions*60)= 1 million Partitions= Records/(1 million * 60)

Partitions= 2.4 x 1,000,000,000/(1,000,000 * 60) = 40 Note: Having too many partitions can reduce the effectiveness of clustered columnstore indexes if each partition has fewer than 1 million rows. Dedicated SQL pools automatically partition

your data into 60 databases. So, if you create a table with 100 partitions, the result will be 6000 partitions.

Reference:

https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/best-practices-dedicated-sql-pool

QUESTION 12

HOTSPOT

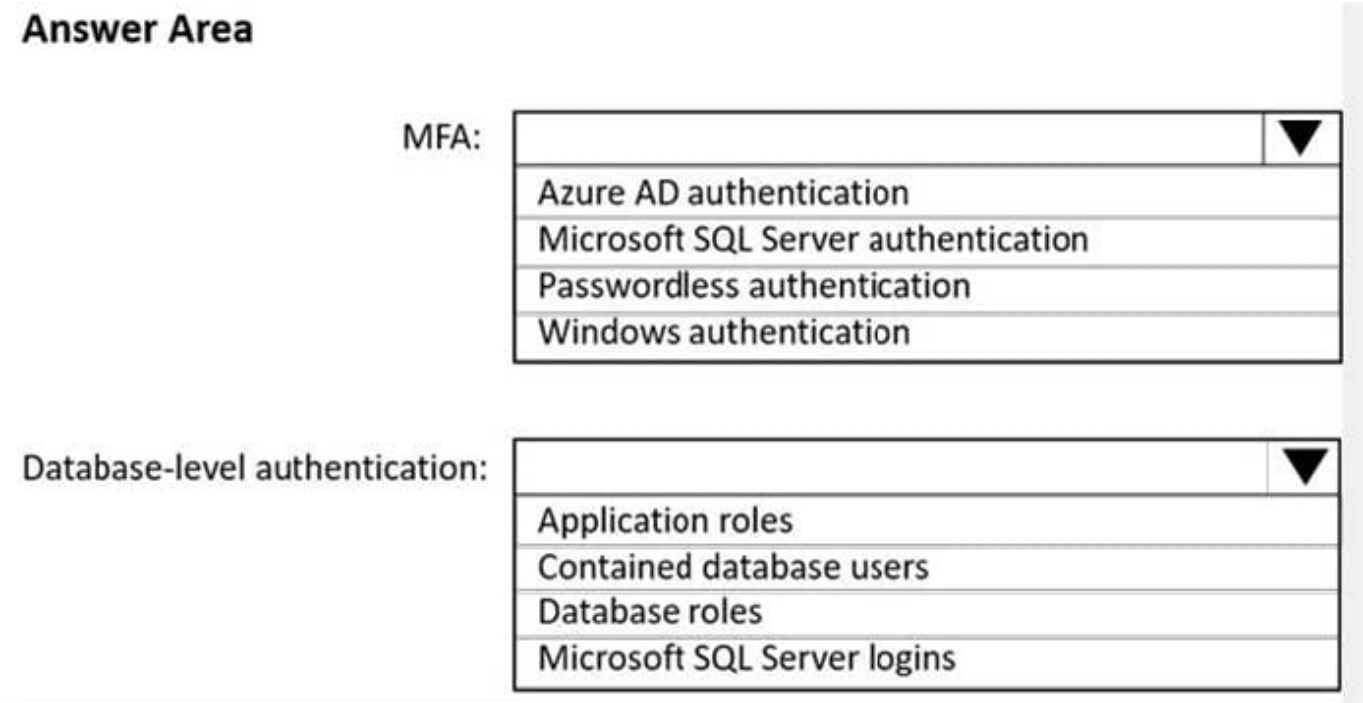
You have an Azure subscription that is linked to a hybrid Azure Active Directory (Azure AD) tenant. The subscription contains an Azure Synapse Analytics SQL pool named Pool1.

You need to recommend an authentication solution for Pool1. The solution must support multi-factor authentication (MFA) and database-level authentication.

Which authentication solution or solutions should you include m the recommendation? To answer, select the appropriate options in the answer area.

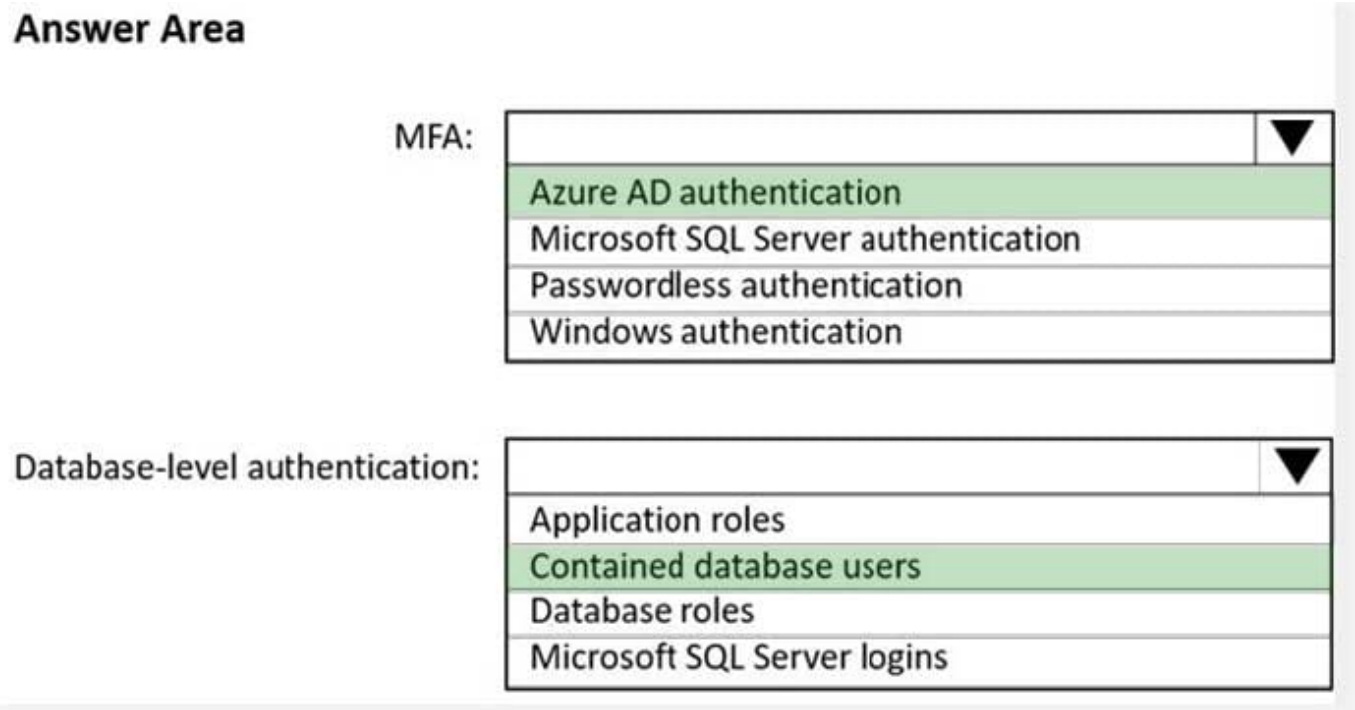
NOTE: Each correct selection is worth one point.

Hot Area:



Correct Answer:

Answer Area



Box 1: Azure AD authentication

Azure AD authentication has the option to include MFA.

Box 2: Contained database users

Azure AD authentication uses contained database users to authenticate identities at the database level.

Reference:

https://docs.microsoft.com/en-us/azure/azure-sql/database/authentication-mfa-ssms-overview

https://docs.microsoft.com/en-us/azure/azure-sql/database/authentication-aad-overview

QUESTION 13

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 an Azure Synapse Analytics dedicated SQL pool that contains a table named Table1.

You have files that are ingested and loaded into an Azure Data Lake Storage Gen2 container named container1.

You plan to insert data from the files in container1 into Table1 and transform the data. Each row of data in the files will produce one row in the serving layer of Table1.

You need to ensure that when the source data files are loaded to container1, the DateTime is stored as an additional column in Table1.

Solution: You use a dedicated SQL pool to create an external table that has an additional DateTime column.

Does this meet the goal?

A. Yes

B. No

Correct Answer: B

Instead use the derived column transformation to generate new columns in your data flow or to modify existing fields.

Reference: https://docs.microsoft.com/en-us/azure/data-factory/data-flow-derived-column