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.passapply.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

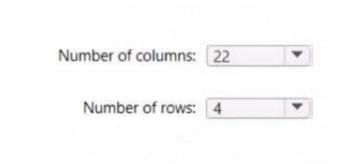⚙ **800,000+** Satisfied Customers
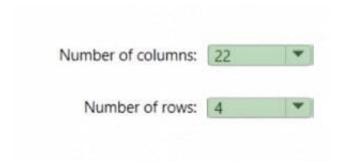
QUESTION 1

HOTSPOT

You have an Azure Data Factory pipeline that contains a data flow. The data flow contains the following expression.

```
source(output(
    License_plate as string,
    Make as string,
    Time as string
),
allowSchemaDrift: true,
```

Hot Area:

Number of columns: 22

Number of rows: 4

Correct Answer:

Number of columns: 22

Number of rows: 4

QUESTION 2

A company purchases IoT devices to monitor manufacturing machinery. The company uses an Azure IoT Hub to communicate with the IoT devices.

The company must be able to monitor the devices in real-time.

You need to design the solution.

What should you recommend?

A. Azure Analysis Services using Microsoft Visual Studio

B. Azure Data Factory instance using Azure PowerShell

C. Azure Analysis Services using Azure PowerShell

D. Azure Stream Analytics cloud job using Azure Portal

Correct Answer: D

In a real-world scenario, you could have hundreds of these sensors generating events as a stream. Ideally, a gateway device would run code to push these events to Azure Event Hubs or Azure IoT Hubs. Your Stream Analytics job would

ingest these events from Event Hubs and run real-time analytics queries against the streams.

Create a Stream Analytics job:

In the Azure portal, select + Create a resource from the left navigation menu. Then, select Stream Analytics job from Analytics.

Reference:

https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-get-started-with-azure-stream-analytics-to-process-data-from-iot-devices

QUESTION 3

DRAG DROP

You have an Apache Spark DataFrame named temperatures. A sample of the data is shown in the following table.

Date	Temp
...	...
18-01-2021	3
19-01-2021	4
20-01-2021	2
21-01-2021	2
...	...

You need to produce the following table by using a Spark SQL query.

Year	JAN	FEB	MAR	APR	MAY
2019	2.3	4.1	5.2	7.6	9.2
2020	2.4	4.2	4.9	7.8	9.1
2021	2.6	5.3	3.4	7.9	9.5

How should you complete the query? To answer, drag the appropriate values to the correct targets. Each value 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. NOTE: Each correct selection is worth one point.

Select and Place:

Values

CAST

COLLATE

CONVERT

FLATTEN

PIVOT

UNPIVOT

Answer Area

```
SELECT * FROM (
  SELECT YEAR(Date) Year, MONTH(Date) Month, Temp
  FROM temperatures
  WHERE date BETWEEN DATE '2019-01-01' AND DATE '2021-08-31'
)
  [          ] (
  AVG ( [          ] (Temp AS DECIMAL(4, 1)))

  FOR Month in (
    1 JAN, 2 FEB, 3 MAR, 4 APR, 5 MAY, 6 JUN,
    7 JUL, 8 AUG, 9 SEP, 10 OCT, 11 NOV, 12 DEC
            )
)
ORDER BY Year ASC
```

Correct Answer:

Values   Answer Area

```
                    SELECT * FROM (
                      SELECT YEAR(Date) Year, MONTH(Date) Month, Temp
    [          ]      FROM temperatures
                      WHERE date BETWEEN DATE '2019-01-01' AND DATE '2021-08-31'
    COLLATE       )
                    PIVOT  (
    CONVERT           AVG ( CAST     (Temp AS DECIMAL(4, 1)))
    FLATTEN
                      FOR Month in (
    [          ]        1 JAN, 2 FEB, 3 MAR, 4 APR, 5 MAY, 6 JUN,
                        7 JUL, 8 AUG, 9 SEP, 10 OCT, 11 NOV, 12 DEC
    UNPIVOT                   )
                    )
                    ORDER BY Year ASC
```

QUESTION 4

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: In an Azure Synapse Analytics pipeline, you use a data flow that contains a Derived Column transformation.

Does this meet the goal?

A. Yes

B. No

Correct Answer: A

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

QUESTION 5

DRAG DROP

You have an Azure subscription.

You plan to build a data warehouse in an Azure Synapse Analytics dedicated SQL pool named pool1 that will contain staging tables and a dimensional model. Pool1 will contain the following tables.

Name	Number of rows	Update frequency	Description
Common. Date	7,300	New rows inserted yearly	• Contains one row per date for the last 20 years • Contains columns named Year, Month, Quarter, and IsWeekend
Marketing.WebSessions	1,500,500,000	Hourly inserts and updates	Fact table that contains counts of and updates sessions and page views, including foreign key values for date, channel, device, and medium
Staging.WebSessions	300,000	Hourly truncation and inserts	Staging table for web session data, truncation and including descriptive fields for inserts channel, device, and medium

You need to design the table storage for pool1. The solution must meet the following requirements:

Maximize the performance of data loading operations to Staging.WebSessions. Minimize query times for reporting queries against the dimensional model.

Which type of table distribution should you use for each table? To answer, drag the appropriate table distribution types to the correct tables. Each table distribution type 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.

NOTE: Each correct selection is worth one point.

Select and Place:

Table distribution types

Hash

Replicated

Round-robin

Answer Area

Common.Data:

Marketing.Web.Sessions:

Staging. Web.Sessions:

Correct Answer:

Table distribution types

Answer Area

Common.Data: | Replicated

Marketing.Web.Sessions: | Hash

Staging. Web.Sessions: | Round-robin