Custom View Settings

Question #10 *Topic 2*

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
```

Question #11 *Topic 2*

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

Question #12 *Topic 2*

HOTSPOT -
The following code segment is used to create an Azure Databricks cluster.

```
{
     "num_workers": null,
     "autoscale": {
          "min_workers": 2,
          "max_workers": 8
     },
     "cluster_name": "MyCluster",
     "spark_version": "latest-stable-scala2.11",
     "spark_conf": {
          "spark.databricks.cluster.profile": "serverless",
          "spark.databricks.repl.allowedLanguages": "sql,python,r"
     },
     "node_type_id": "Standard_DS13_v2",
     "ssh_public_keys": [],
     "custom_tags": {
          "ResourceClass": "Serverless"
     },
     "spark_env_vars": {
          "PYSPARK_PYTHON": "/databricks/python3/bin/python3"
     },
     "autotermination_minutes": 90,
     "enable_elastic_disk": true,
     "init_scripts": []
}
```

For each of the following statements, select Yes if the statement is true. Otherwise, select No.
NOTE: Each correct selection is worth one point.
Hot Area:

Answer Area

Statements	Yes	No
The Databricks cluster supports multiple concurrent users.	○	○
The Databricks cluster minimizes costs when running scheduled jobs that execute notebooks.	○	○
The Databricks cluster supports the creation of a Delta Lake table.	○	○

Question #13 *Topic 2*

You are designing a statistical analysis solution that will use custom proprietary Python functions on near real-time data from Azure Event Hubs.
You need to recommend which Azure service to use to perform the statistical analysis. The solution must minimize latency.
What should you recommend?

A. Azure Synapse Analytics

B. Azure Databricks

C. Azure Stream Analytics

D. Azure SQL Database

Question #14 *Topic 2*

HOTSPOT -
You have an enterprise data warehouse in Azure Synapse Analytics that contains a table named FactOnlineSales. The table contains data from the start of 2009 to the end of 2012.
You need to improve the performance of queries against FactOnlineSales by using table partitions. The solution must meet the following requirements:
- Create four partitions based on the order date.
- Ensure that each partition contains all the orders placed during a given calendar year.

How should you complete the T-SQL command? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.
Hot Area:

Answer Area

```
CREATE TABLE [dbo].FactOnlineSales
([OnlineSalesKey] [int] NOT NULL,
[OrderDateKey] [datetime] NOT NULL,
[StoreKey] [int] NOT NULL,
[ProductKey] [int] NOT NULL,
[CustomerKey] [int] NOT NULL,
[SalesOrderNumber] [nvarchar](20) NOT NULL,
[SalesQuantity] [int] NOT NULL,
[SalesAmount] [money] NOT NULL,
[UnitPrice] [money] NULL)
WITH (CLUSTERED COLUMNSTORE INDEX)
PARTITION ([OrderDateKey] RANGE [ ▼ ] FOR VALUES
```

[▼] RIGHT / LEFT

```
(
```

[▼] 20090101,20121231 / 20100101,20110101,20120101 / 20090101,20100101,20110101,20120101

```
)
```

Previous Questions

Next Questions