EXAMTOPICS

- Expert Verified, Online, **Free**.

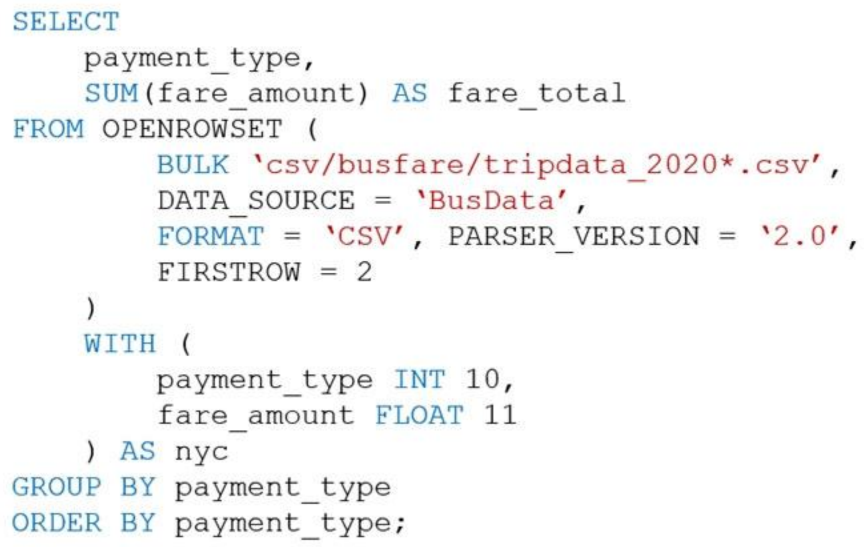
Custom View Settings

Question #46

Topic 1

You are performing exploratory analysis of the bus fare data in an Azure Data Lake Storage Gen2 account by using an Azure Synapse Analytics serverless SQL pool.

You execute the Transact-SQL query shown in the following exhibit.



```
SELECT
    payment_type,
    SUM(fare_amount) AS fare_total
FROM OPENROWSET (
        BULK 'csv/busfare/tripdata_2020*.csv',
        DATA_SOURCE = 'BusData',
        FORMAT = 'CSV', PARSER_VERSION = '2.0',
        FIRSTROW = 2
    )
    WITH (
        payment_type INT 10,
        fare_amount FLOAT 11
    ) AS nyc
GROUP BY payment_type
ORDER BY payment_type;
```

What do the query results include?

A. Only CSV files in the tripdata_2020 subfolder.

B. All files that have file names that beginning with "tripdata_2020".

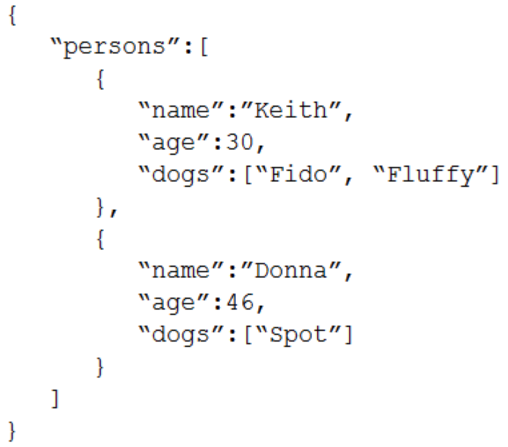
C. All CSV files that have file names that contain "tripdata_2020".

D. Only CSV that have file names that beginning with "tripdata_2020".

Question #47 *Topic 1*

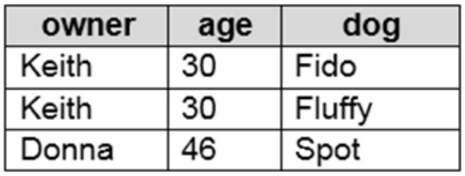
DRAG DROP -

You use PySpark in Azure Databricks to parse the following JSON input.



```
{
   "persons":[
      {
         "name":"Keith",
         "age":30,
         "dogs":["Fido", "Fluffy"]
      },
      {
         "name":"Donna",
         "age":46,
         "dogs":["Spot"]
      }
   ]
}
```

You need to output the data in the following tabular format.

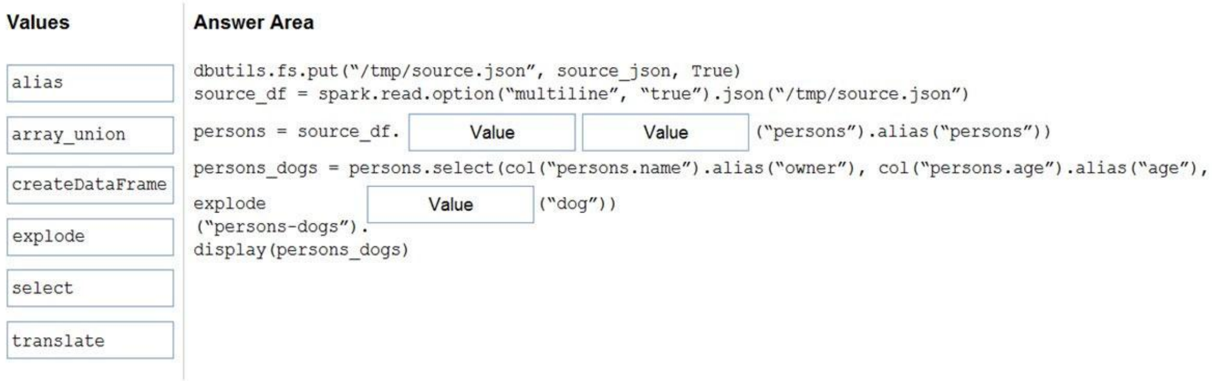


owner	age	dog
Keith	30	Fido
Keith	30	Fluffy
Donna	46	Spot

How should you complete the PySpark code? 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 spit bar between panes or scroll to view content.

NOTE: Each correct selection is worth one point.

Select and Place:



Values

- alias
- array_union
- createDataFrame
- explode
- select
- translate

Answer Area

```
dbutils.fs.put("/tmp/source.json", source_json, True)
source_df = spark.read.option("multiline", "true").json("/tmp/source.json")
persons = source_df. [Value] [Value] ("persons").alias("persons"))
persons_dogs = persons.select(col("persons.name").alias("owner"), col("persons.age").alias("age"),
explode [Value] ("dog"))
("persons-dogs").
display(persons_dogs)
```

Question #48

Topic 1

HOTSPOT -

You are designing an application that will store petabytes of medical imaging data.

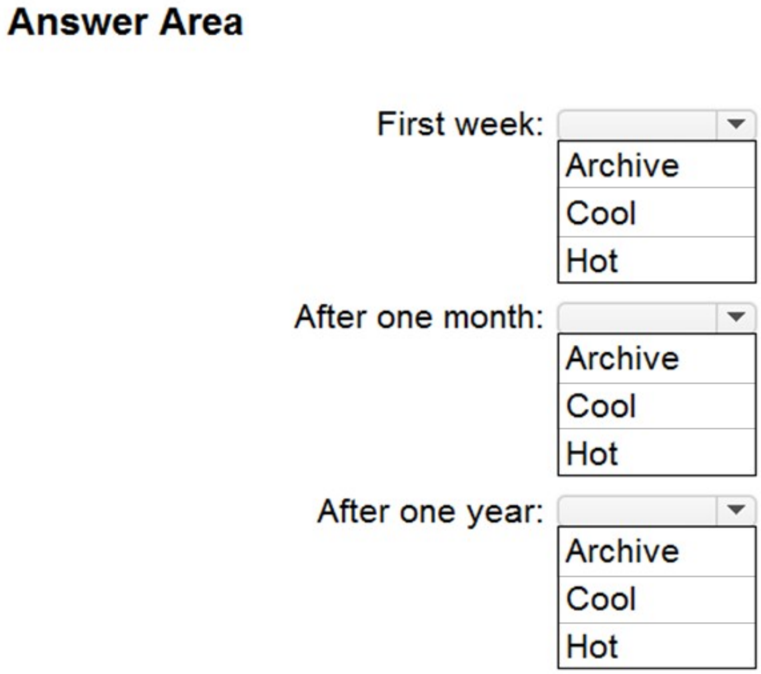
When the data is first created, the data will be accessed frequently during the first week. After one month, the data must be accessible within 30 seconds, but files will be accessed infrequently. After one year, the data will be accessed infrequently but must be accessible within five minutes.

You need to select a storage strategy for the data. The solution must minimize costs.

Which storage tier should you use for each time frame? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection is worth one point.

Hot Area:



Question #49

Topic 1

You have an Azure Synapse Analytics Apache Spark pool named Pool1.

You plan to load JSON files from an Azure Data Lake Storage Gen2 container into the tables in Pool1. The structure and data types vary by file.

You need to load the files into the tables. The solution must maintain the source data types.

What should you do?

A. Use a Conditional Split transformation in an Azure Synapse data flow.

B. Use a Get Metadata activity in Azure Data Factory.

C. Load the data by using the OPENROWSET Transact-SQL command in an Azure Synapse Analytics serverless SQL pool.

D. Load the data by using PySpark.

Question #50 *Topic 1*

You have an Azure Databricks workspace named workspace1 in the Standard pricing tier. Workspace1 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. Configure a global init script for workspace1.

B. Create a cluster policy in workspace1.

C. Upgrade workspace1 to the Premium pricing tier.

D. Create a pool in workspace1.

← Previous Questions

Next Questions →