

Question #61

Topic 1

HOTSPOT -

You have an Azure Synapse Analytics dedicated SQL pool.

You need to create a table named FactInternetSales that will be a large fact table in a dimensional model. FactInternetSales will contain 100 million rows and two columns named SalesAmount and OrderQuantity. Queries executed on FactInternetSales will aggregate the values in SalesAmount and OrderQuantity from the last year for a specific product. The solution must minimize the data size and query execution time. How should you complete the code? 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].[FactInternetSales]
(
    [ProductKey] int NOT NULL
,   [OrderDateKey] int NOT NULL
,   [CustomerKey] int NOT NULL
,   [PromotionKey] int NOT NULL
,   [SalesOrderNumber] nvarchar(20) NOT NULL
,   [OrderQuantity] smallint NOT NULL
,   [UnitPrice] money NOT NULL
,   [SalesAmount] money NOT NULL
)
```

WITH

(CLUSTERED COLUMNSTORE INDEX
(CLUSTERED INDEX ([OrderDateKey])
(HEAP
(INDEX on [ProductKey]

, DISTRIBUTION =

);

Hash([OrderDateKey])
Hash([ProductKey])
REPLICATE
ROUND_ROBIN

You have an Azure Synapse Analytics dedicated SQL pool that contains a table named Table1. Table1 contains the following:

- ☞ One billion rows
- ☞ A clustered columnstore index
- ☞ A hash-distributed column named Product Key
- ☞ A column named Sales Date that is of the date data type and cannot be null

Thirty million rows will be added to Table1 each month.

You need to partition Table1 based on the Sales Date column. The solution must optimize query performance and data loading.

How often should you create a partition?

- A. once per month
- B. once per year
- C. once per day
- D. once per week

You have an Azure Databricks workspace that contains a Delta Lake dimension table named Table1.

Table1 is a Type 2 slowly changing dimension (SCD) table.

You need to apply updates from a source table to Table1.

Which Apache Spark SQL operation should you use?

- A. CREATE
- B. UPDATE
- C. ALTER
- D. MERGE

You are designing an Azure Data Lake Storage solution that will transform raw JSON files for use in an analytical workload.

You need to recommend a format for the transformed files. The solution must meet the following requirements:

- ☞ Contain information about the data types of each column in the files.
- ☞ Support querying a subset of columns in the files.
- ☞ Support read-heavy analytical workloads.
- ☞ Minimize the file size.

What should you recommend?

- A. JSON
- B. CSV
- C. Apache Avro
- D. Apache Parquet

Note: This question is part of a series of questions that present the same scenario. Each question in the series contains a unique solution that might meet the stated goals. Some question sets might have more than one correct solution, while others might not have a correct solution. 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 Storage account that contains 100 GB of files. The files contain rows of text and numerical values. 75% of the rows contain description data that has an average length of 1.1 MB.

You plan to copy the data from the storage account to an enterprise data warehouse in Azure Synapse Analytics.

You need to prepare the files to ensure that the data copies quickly.

Solution: You modify the files to ensure that each row is less than 1 MB.

Does this meet the goal?

A. Yes

B. No

[← Previous Questions](#)

[Next Questions →](#)