Custom View Settings

Question #36 *Topic 1*

HOTSPOT -

You plan to develop a dataset named Purchases by using Azure Databricks. Purchases will contain the following columns:

- ProductID
- ItemPrice
- LineTotal
- Quantity
- StoreID
- Minute
- Month
- Hour

Year -

-

- Day

You need to store the data to support hourly incremental load pipelines that will vary for each Store ID. The solution must minimize storage costs.

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

```
df.write
[ ▼ ]                                  [ ▼ ]
.bucketBy                              ("*")
.partitionBy                           ("StoreID","Hour")
.range                                 ("StoreID","Year","Month","Day","Hour")
.sortBy
.mode("append")
[ ▼ ]
.csv("/Purchases")
.json("/Purchases")
.parquet("/Purchases")
.saveAsTable("/Purchases")
```

Question #37 *Topic 1*

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

- Contain sales data for 20,000 products.

Use hash distribution on a column named ProductID.

- Contain 2.4 billion records for the years 2019 and 2020.

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

A. 40

B. 240

C. 400

D. 2,400

Question #38 *Topic 1*

HOTSPOT -

You are creating dimensions for a data warehouse in an Azure Synapse Analytics dedicated SQL pool.

You create a table by using the Transact-SQL statement shown in the following exhibit.

```
CREATE TABLE [DBO].[DimProduct](
       [ProductKey] [int] IDENTITY(1,1) NOT NULL,
       [ProductSourceID] [int] NOT NULL,
       [ProductName] [nvarchar](100) NOT NULL,
       [ProductNumber] [nvarchar](25) NOT NULL,
       [Color] [nvarchar](15) NULL,
       [Size] [nvarchar](5) NULL,
       [Weight] [decimal](8, 2) NULL,
       [ProductCategory] [nvarchar](100) NULL,
       [SellStartDate] [date] NOT NULL,
       [SellEndDate] [date] NULL,
       [RowInsertedDateTime] [datetime] NOT NULL,
       [RowUpdatedDateTime] [datetime] NOT NULL,
       [ETLAuditID] [int] NOT NULL
)
```

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:

Answer Area

DimProduct is a **[answer choice]** slowly changing dimension (SCD).

- Type 0
- Type 1
- Type 2

The ProductKey column is **[answer choice]**.

- a surrogate key
- a business key
- an audit column

Question #39 *Topic 1*

You are designing a fact table named FactPurchase in an Azure Synapse Analytics dedicated SQL pool. The table contains purchases from suppliers for a retail store. FactPurchase will contain the following columns.

Name	Data type	Nullable
PurchaseKey	Bigint	No
DateKey	Int	No
SupplierKey	Int	No
StockItemKey	Int	No
PurchaseOrderID	Int	Yes
OrderedQuantity	Int	No
OrderedOuters	Int	No
ReceivedOuters	Int	No
Package	Nvarchar(50)	No
IsOrderFinalized	Bit	No
LineageKey	Int	No

FactPurchase will have 1 million rows of data added daily and will contain three years of data.
Transact-SQL queries similar to the following query will be executed daily.

```
SELECT -
SupplierKey, StockItemKey, COUNT(*)

FROM FactPurchase -

WHERE DateKey >= 20210101 -

AND DateKey <= 20210131 -
GROUP By SupplierKey, StockItemKey
```

Which table distribution will minimize query times?

A. replicated

B. hash-distributed on PurchaseKey

C. round-robin

D. hash-distributed on DateKey

Question #40 *Topic 1*

You are implementing a batch dataset in the Parquet format.
Data files will be produced be using Azure Data Factory and stored in Azure Data Lake Storage Gen2. The files will be consumed by an Azure Synapse Analytics serverless SQL pool.
You need to minimize storage costs for the solution.
What should you do?

A. Use Snappy compression for the files.

B. Use OPENROWSET to query the Parquet files.

C. Create an external table that contains a subset of columns from the Parquet files.

D. Store all data as string in the Parquet files.

Previous Questions Next Questions