Demo Questions

Microsoft DP-203 Exam

Azure Data Engineer Associate

Question #1 *Topic 1*

HOTSPOT -
You are creating dimensions for a data warehouse in an Azure Synapse Analyticsdedicated 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
)
```
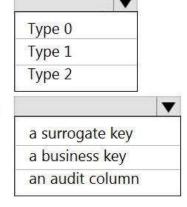
Use the drop-down menus to select the answer choice that completes each statementbased on the information presented in the graphic.
NOTE: Each correct selection is worth one point.Hot Area:

Answer Area

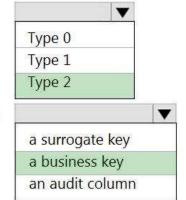DimProduct is a **[answer choice]** slowly changing dimension (SCD).

Type 0
Type 1
Type 2

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

a surrogate key
a business key
an audit column

Correct Answer:

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

Box 1: Type 2 -
A Type 2 SCD supports versioning of dimension members. Often the source system doesn't store versions, so the data warehouse load process detects and manages changes in a dimension table. In this case, the dimension table must use a surrogatekey to provide a unique reference to a version of the dimension member. It also includes columns that define the date range validity of the version (for example, StartDate and EndDate) and possibly a flag column (for example, IsCurrent) to easily filter by current dimension members.Incorrect
Answers:
A Type 1 SCD always reflects the latest values, and when changes in source data aredetected, the dimension table data is overwritten.

Box 2: a business key -
A business key or natural key is an index which identifies uniqueness of a row based oncolumns that exist naturally in a table according to business rules. For example business keys are customer code in a customer table, composite of sales order header number and sales order item line number within a sales order details table.
Reference:
https://docs.microsoft.com/en-us/learn/modules/populate-slowly-changing-dimensions- azure-synapse-analytics-pipelines/3-choose-between-dimension-types

Question #2 *Topic 1*

You are designing a fact table named FactPurchase in an Azure Synapse Analyticsdedicated 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

Correct Answer: *B*

Hash-distributed tables improve query performance on large fact tables, and are the focus of this article. Round-robin tables are useful for improving loading speed.
Incorrect:
Not D: Do not use a date column. . All data for the same date lands in the same distribution. If several users are all filtering on the same date, then only 1 of the 60distributions do all the processing work.
Reference:

Question #3 *Topic 1*

You have a table in an Azure Synapse Analytics dedicated SQL pool. The table was created by using the following Transact-SQL statement.

```
CREATE TABLE [dbo].[DimEmployee](
    [EmployeeKey] [int] IDENTITY(1,1) NOT NULL,
    [EmployeeID] [int] NOT NULL,
    [FirstName] [varchar](100) NOT NULL,
    [LastName] [varchar](100) NOT NULL,
    [JobTitle] [varchar](100) NULL,
    [LastHireDate] [date] NULL,
    [StreetAddress] [varchar](500) NOT NULL,
    [City] [varchar](200) NOT NULL,
    [StateProvince] [varchar](50) NOT NULL,
    [Portalcode] [varchar](10) NOT NULL
)
```

You need to alter the table to meet the following requirements:
✑ Ensure that users can identify the current manager of employees.
✑ Support creating an employee reporting hierarchy for your entire company.
✑ Provide fast lookup of the managersג€™ attributes such as name and job title.Which column should you add to the table?

- A. [ManagerEmployeeID] [int] NULL
- B. [ManagerEmployeeID] [smallint] NULL
- C. [ManagerEmployeeKey] [int] NULL
- D. [ManagerName] [varchar](200) NULL

Correct Answer: *A*

Use the same definition as the EmployeeID column.
Reference:
https://docs.microsoft.com/en-us/analysis-services/tabular-models/hierarchies-ssas- tabular

Question #4 *Topic 1*

You have an Azure Synapse workspace named MyWorkspace that contains an Apache Spark database named mytestdb.
You run the following command in an Azure Synapse Analytics Spark pool in MyWorkspace.
CREATE TABLE mytestdb.myParquetTable(
EmployeeID int,

EmployeeName string,
EmployeeStartDate date)

USING Parquet -
You then use Spark to insert a row into mytestdb.myParquetTable. The row containsthe following data.

EmployeeName	EmployeeID	EmployeeStartDate
Alice	24	2020-01-25

One minute later, you execute the following query from a serverless SQL pool inMyWorkspace.

SELECT EmployeeID -
FROM mytestdb.dbo.myParquetTableWHERE
name = 'Alice';
What will be returned by the query?

- A. 24
- B. an error
- C. a null value

Correct Answer: *A*

Once a database has been created by a Spark job, you can create tables in it with Spark that use Parquet as the storage format. Table names will be converted to lower case and need to be queried using the lower case name. These tables will immediatelybecome available for querying by any of the Azure Synapse workspace Spark pools.
They can also be used from any of the Spark jobs subject to permissions. Note: For external tables, since they are synchronized to serverless SQL poolasynchronously, there will be a delay until they appear.
Reference:
https://docs.microsoft.com/en-us/azure/synapse-analytics/metadata/table

Question #5 *Topic 1*

DRAG DROP -
You have a table named SalesFact in an enterprise data warehouse in Azure Synapse Analytics.
SalesFact contains sales data from the past 36 months and has the followingcharacteristics:
☞ Is partitioned by month
☞ Contains one billion rows
☞ Has clustered columnstore indexes
At the beginning of each month, you need to remove data from SalesFact that is olderthan 36 months as quickly as possible.
Which three actions should you perform in sequence in a stored procedure? To answer,move the appropriate actions from the list of actions to the answer area and arrange them in the correct order.

Select and Place:

Actions	Answer Area
Switch the partition containing the stale data from SalesFact to SalesFact_Work.	
Truncate the partition containing the stale data.	
Drop the SalesFact_Work table.	
Create an empty table named SalesFact_Work that has the same schema as SalesFact.	
Execute a DELETE statement where the value in the Date column is more than 36 months ago.	
Copy the data to a new table by using CREATE TABLE AS SELECT (CTAS).	

Correct Answer:

Actions	Answer Area
Switch the partition containing the stale data from SalesFact to SalesFact_Work.	Create an empty table named SalesFact_Work that has the same schema as SalesFact.
Truncate the partition containing the stale data.	Switch the partition containing the stale data from SalesFact to SalesFact_Work.
Drop the SalesFact_Work table.	Drop the SalesFact_Work table.
Create an empty table named SalesFact_Work that has the same schema as SalesFact.	
Execute a DELETE statement where the value in the Date column is more than 36 months ago.	
Copy the data to a new table by using CREATE TABLE AS SELECT (CTAS).	

Step 1: Create an empty table named SalesFact_work that has the same schema asSalesFact.
Step 2: Switch the partition containing the stale data from SalesFact to SalesFact_Work.
SQL Data Warehouse supports partition splitting, merging, and switching. To switchpartitions between two tables, you must ensure that the partitions align on their respective boundaries and that the table definitions match.
Loading data into partitions with partition switching is a convenient way stage new datain a table that is not visible to users the switch in the new data.

Step 3: Drop the SalesFact_Work table.
Reference:
https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-tables- partition