HIGHER QUALITY
BETTER SERVICE

CERTTREE

QUESTION & ANSWER

Provide One Year
Free Update!

https://www.certtree.com

Exam : DP-203

Title : Data Engineering on Microsoft Azure

Version : DEMO

1. Topic 1, Contoso

Case Study
Transactional Date
Contoso has three years of customer, transactional, operation, sourcing, and supplier data comprised of 10 billion records stored across multiple on-premises Microsoft SQL Server servers. The SQL server instances contain data from various operational systems. The data is loaded into the instances by using SQL server integration Services (SSIS) packages.

You estimate that combining all product sales transactions into a company-wide sales transactions dataset will result in a single table that contains 5 billion rows, with one row per transaction.
Most queries targeting the sales transactions data will be used to identify which products were sold in retail stores and which products were sold online during different time period. Sales transaction data that is older than three years will be removed monthly.

You plan to create a retail store table that will contain the address of each retail store. The table will be approximately 2 MB. Queries for retail store sales will include the retail store addresses.
You plan to create a promotional table that will contain a promotion ID. The promotion ID will be associated to a specific product. The product will be identified by a product ID. The table will be approximately 5 GB.

Streaming Twitter Data
The ecommerce department at Contoso develops and Azure logic app that captures trending Twitter feeds referencing the company's products and pushes the products to Azure Event Hubs.

Planned Changes
Contoso plans to implement the following changes:
* Load the sales transaction dataset to Azure Synapse Analytics.
* Integrate on-premises data stores with Azure Synapse Analytics by using SSIS packages.
* Use Azure Synapse Analytics to analyze Twitter feeds to assess customer sentiments about products.

Sales Transaction Dataset Requirements
Contoso identifies the following requirements for the sales transaction dataset:
• Partition data that contains sales transaction records. Partitions must be designed to provide efficient loads by month. Boundary values must belong: to the partition on the right.
• Ensure that queries joining and filtering sales transaction records based on product ID complete as quickly as possible.
• Implement a surrogate key to account for changes to the retail store addresses.
• Ensure that data storage costs and performance are predictable.
• Minimize how long it takes to remove old records. Customer Sentiment Analytics Requirement

Contoso identifies the following requirements for customer sentiment analytics:
• Allow Contoso users to use PolyBase in an Aure Synapse Analytics dedicated SQL pool to query the content of the data records that host the Twitter feeds. Data must be protected by using row-level

security (RLS). The users must be authenticated by using their own AureAD credentials.
• Maximize the throughput of ingesting Twitter feeds from Event Hubs to Azure Storage without purchasing additional throughput or capacity units.
• Store Twitter feeds in Azure Storage by using Event Hubs Capture. The feeds will be converted into Parquet files.
• Ensure that the data store supports Azure AD-based access control down to the object level.
• Minimize administrative effort to maintain the Twitter feed data records.
• Purge Twitter feed data records;itftaitJ are older than two years.

Data Integration Requirements
Contoso identifies the following requirements for data integration:
Use an Azure service that leverages the existing SSIS packages to ingest on-premises data into datasets stored in a dedicated SQL pool of Azure Synaps Analytics and transform the data.
Identify a process to ensure that changes to the ingestion and transformation activities can be version controlled and developed independently by multiple data engineers.

DRAG DROP
You need to ensure that the Twitter feed data can be analyzed in the dedicated SQL pool. The solution must meet the customer sentiment analytics requirements.
Which three Transaction-SQL DDL commands should you run in sequence? To answer, move the appropriate commands from the list of commands to the answer area and arrange them in the correct order. NOTE: More than one order of answer choices is correct. You will receive credit for any of the correct orders you select.

Commands	Answer Area
CREATE EXTERNAL DATA SOURCE	
CREATE EXTERNAL FILE FORMAT	
CREATE EXTERNAL TABLE	
CREATE EXTERNAL TABLE AS SELECT	
CREATE DATABASE SCOPED CREDENTIAL	

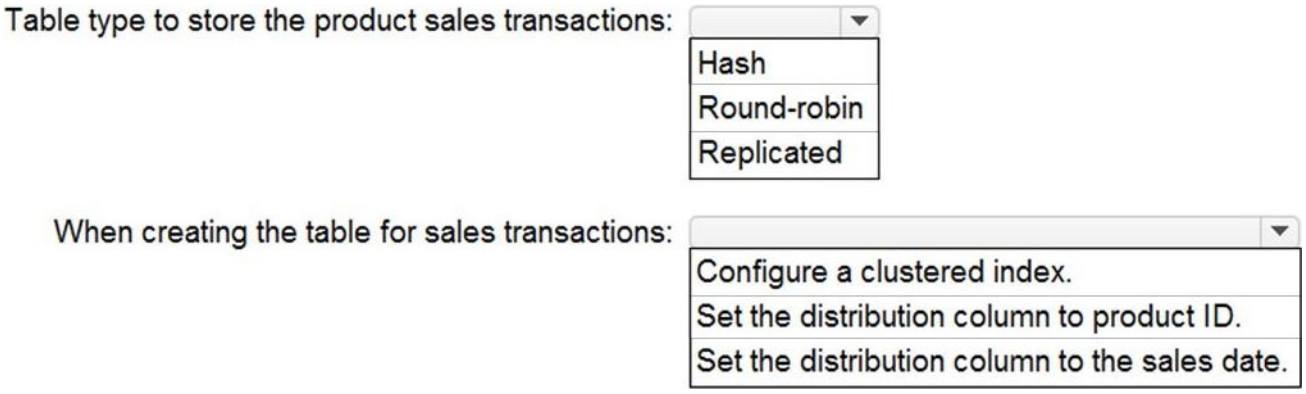
Answer:

Commands	Answer Area
CREATE EXTERNAL DATA SOURCE	CREATE EXTERNAL DATA SOURCE
CREATE EXTERNAL FILE FORMAT	CREATE EXTERNAL FILE FORMAT
CREATE EXTERNAL TABLE	CREATE EXTERNAL TABLE AS SELECT
CREATE EXTERNAL TABLE AS SELECT	
CREATE DATABASE SCOPED CREDENTIAL	

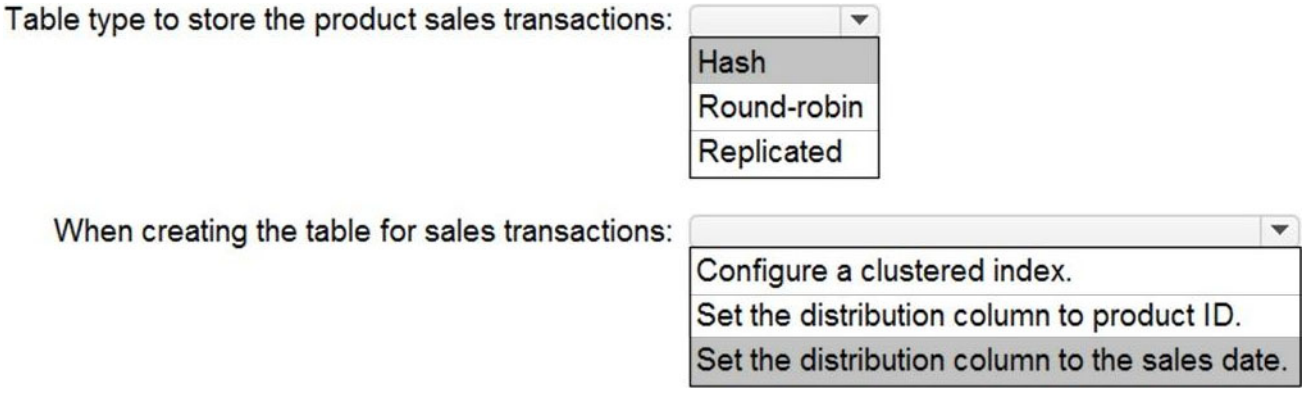
Explanation:
Scenario: Allow Contoso users to use PolyBase in an Azure Synapse Analytics dedicated SQL pool to query the content of the data records that host the Twitter feeds. Data must be protected by using row-level security (RLS). The users must be authenticated by using their own Azure AD credentials.

Box 1: CREATE EXTERNAL DATA SOURCE
External data sources are used to connect to storage accounts.
Box 2: CREATE EXTERNAL FILE FORMAT
CREATE EXTERNAL FILE FORMAT creates an external file format object that defines external data stored in Azure Blob Storage or Azure Data Lake Storage. Creating an external file format is a prerequisite for creating an external table.
Box 3: CREATE EXTERNAL TABLE AS SELECT
When used in conjunction with the CREATE TABLE AS SELECT statement, selecting from an external table imports data into a table within the SQL pool. In addition to the COPY statement, external tables are useful for loading data.
Reference: https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/develop-tables-external-tables

2.HOTSPOT
You need to design a data storage structure for the product sales transactions. The solution must meet the sales transaction dataset requirements.
What should you include in the solution? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.



Answer:



Explanation:
Box 1: Hash
Scenario:
Ensure that queries joining and filtering sales transaction records based on product ID complete as quickly as possible.
A hash distributed table can deliver the highest query performance for joins and aggregations on large

tables.
Box 2: Set the distribution column to the sales date.
Scenario: Partition data that contains sales transaction records. Partitions must be designed to provide efficient loads by month. Boundary values must belong to the partition on the right.
Reference: https://rajanieshkaushikk.com/2020/09/09/how-to-choose-right-data-distribution-strategy-for-azure-synapse/

3.HOTSPOT
You need to design the partitions for the product sales transactions. The solution must meet the sales transaction dataset requirements.
What should you include in the solution? To answer, select the appropriate options in the answer area.
NOTE: Each correct selection is worth one point.

Partition product sales transactions data by:
- Sales date
- Product ID
- Promotion ID

Store product sales transactions data in:
- An Azure Synapse Analytics dedicated SQL pool
- An Azure Synapse Analytics serverless SQL pool
- An Azure Data Lake Storage Gen2 account linked to an Azure Synapse Analytics workspace

Answer:

Partition product sales transactions data by:
- Sales date
- Product ID
- Promotion ID

Store product sales transactions data in:
- An Azure Synapse Analytics dedicated SQL pool
- An Azure Synapse Analytics serverless SQL pool
- An Azure Data Lake Storage Gen2 account linked to an Azure Synapse Analytics workspace

Explanation:
Box 1: Sales date
Scenario: Contoso requirements for data integration include:

➪ Partition data that contains sales transaction records. Partitions must be designed to provide efficient loads by month. Boundary values must belong to the partition on the right.
Box 2: An Azure Synapse Analytics Dedicated SQL pool
Scenario: Contoso requirements for data integration include:
➪ Ensure that data storage costs and performance are predictable.
The size of a dedicated SQL pool (formerly SQL DW) is determined by Data Warehousing Units (DWU).
Dedicated SQL pool (formerly SQL DW) stores data in relational tables with columnar storage. This format significantly reduces the data storage costs, and improves query performance.
Synapse analytics dedicated sql pool
Reference: https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-overview-what-is

4.HOTSPOT
You need to implement an Azure Synapse Analytics database object for storing the sales transactions data. The solution must meet the sales transaction dataset requirements.
What solution must meet the sales transaction dataset requirements.
What should you do? To answer, select the appropriate options in the answer area. NOTE: Each correct selection is worth one point.

Transact-SQL DDL command to use:

- CREATE EXTERNAL TABLE
- CREATE TABLE
- CREATE VIEW

Partitioning option to use in the WITH clause of the DDL statement:

- FORMAT_OPTIONS
- FORMAT_TYPE
- RANGE LEFT FOR VALUES
- RANGE RIGHT FOR VALUES

Answer:

Transact-SQL DDL command to use:

- CREATE EXTERNAL TABLE
- CREATE TABLE
- CREATE VIEW

Partitioning option to use in the WITH clause of the DDL statement:

- FORMAT_OPTIONS
- FORMAT_TYPE
- RANGE LEFT FOR VALUES
- RANGE RIGHT FOR VALUES

Explanation:
Box 1: Create table
Scenario: Load the sales transaction dataset to Azure Synapse Analytics
Box 2: RANGE RIGHT FOR VALUES
Scenario: Partition data that contains sales transaction records. Partitions must be designed to provide efficient loads by month. Boundary values must belong to the partition on the right.
RANGE RIGHT: Specifies the boundary value belongs to the partition on the right (higher values).
FOR VALUES (boundary_value [,...n]): Specifies the boundary values for the partition.
Scenario: Load the sales transaction dataset to Azure Synapse Analytics.
Contoso identifies the following requirements for the sales transaction dataset:
➪ Partition data that contains sales transaction records. Partitions must be designed to provide efficient loads by month. Boundary values must belong to the partition on the right.
➪ Ensure that queries joining and filtering sales transaction records based on product ID complete as quickly as possible.
➪ Implement a surrogate key to account for changes to the retail store addresses.
➪ Ensure that data storage costs and performance are predictable.
➪ Minimize how long it takes to remove old records.
Reference: https://docs.microsoft.com/en-us/sql/t-sql/statements/create-table-azure-sql-data-warehouse

5.You need to integrate the on-premises data sources and Azure Synapse Analytics. The solution must meet the data integration requirements.
Which type of integration runtime should you use?
A. Azure-SSIS integration runtime
B. self-hosted integration runtime
C. Azure integration runtime
Answer: C