

EXAMGOOD

QUESTION & ANSWER

Exam Good provides update free of charge in one year!

Accurate study guides
High passing rate!

<http://www.examgood.com>

Exam : **DP-203**

Title : Data Engineering on
Microsoft Azure

Version : DEMO

1. Topic 1, Contoso Case

Case Study

Transactional Data

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 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 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.

HOTSPOT

You need to design an analytical storage solution for the transactional data. 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.

Table type to store retail store data:

	▼
Hash	
Replicated	
Round-robin	

Table type to store promotional data:

	▼
Hash	
Replicated	
Round-robin	

Answer:

Table type to store retail store data:

	▼
Hash	
Replicated	
Round-robin	

Table type to store promotional data:

	▼
Hash	
Replicated	
Round-robin	

Explanation:

Graphical user interface, text, application, table

Description automatically generated

Box 1: Round-robin

Round-robin tables are useful for improving loading speed.

Scenario: Partition data that contains sales transaction records. Partitions must be designed to provide efficient loads by month.

Box 2: Hash

Hash-distributed tables improve query performance on large fact tables.

2.DRAG DROP

You need to implement versioned changes to the integration pipelines. The solution must meet the data integration requirements.

In which order should you perform the actions? To answer, move all actions from the list of actions to the answer area and arrange them in the correct order.

- Publish changes
- Create a feature branch
- Merge changes
- Create a repository and a main branch
- Create a pull request

Answer:

- | | |
|---------------------------------------|---------------------------------------|
| Publish changes | Create a repository and a main branch |
| Create a feature branch | Create a feature branch |
| Merge changes | Create a pull request |
| Create a repository and a main branch | Merge changes |
| Create a pull request | Publish changes |

Explanation:

Graphical user interface, application

Description automatically generated

Scenario: Identify a process to ensure that changes to the ingestion and transformation activities can be version-controlled and developed independently by multiple data engineers.

Step 1: Create a repository and a main branch

You need a Git repository in Azure Pipelines, TFS, or GitHub with your app.

Step 2: Create a feature branch

Step 3: Create a pull request

Step 4: Merge changes

Merge feature branches into the main branch using pull requests.

Step 5: Publish changes

3.HOTSPOT

You need to design a data ingestion and storage solution for the Twitter feeds. The solution must meet the customer sentiment analytics requirements.

What should you include in the solution? To answer, select the appropriate options in the answer area.

NOTE: Each correct selection b worth one point.

To increase the throughput of ingesting the Twitter feeds:

<input type="checkbox"/> Configure Event Hubs partitions.
<input type="checkbox"/> Enable Auto-Inflate in Event Hubs.
<input type="checkbox"/> Use Event Hubs Dedicated.

To store the Twitter feed data, use:

<input type="checkbox"/> An Azure Data Lake Storage Gen2 account
<input type="checkbox"/> An Azure Databricks high concurrency cluster
<input type="checkbox"/> An Azure General-purpose v2 storage account in the Premium tier

Answer:

To increase the throughput of ingesting the Twitter feeds:

<input checked="" type="checkbox"/> Configure Event Hubs partitions.
<input type="checkbox"/> Enable Auto-Inflate in Event Hubs.
<input type="checkbox"/> Use Event Hubs Dedicated.

To store the Twitter feed data, use:

<input checked="" type="checkbox"/> An Azure Data Lake Storage Gen2 account
<input type="checkbox"/> An Azure Databricks high concurrency cluster
<input type="checkbox"/> An Azure General-purpose v2 storage account in the Premium tier

Explanation:

Graphical user interface, text

Description automatically generated

Box 1: Configure Event Hubs partitions

Scenario: Maximize the throughput of ingesting Twitter feeds from Event Hubs to Azure Storage without purchasing additional throughput or capacity units.

Event Hubs is designed to help with processing of large volumes of events. Event Hubs throughput is scaled by using partitions and throughput-unit allocations.

Event Hubs traffic is controlled by TUs (standard tier). Auto-inflate enables you to start small with the

minimum required TUs you choose. The feature then scales automatically to the maximum limit of TUs you need, depending on the increase in your traffic.

Box 2: An Azure Data Lake Storage Gen2 account

Scenario: Ensure that the data store supports Azure AD-based access control down to the object level. Azure Data Lake Storage Gen2 implements an access control model that supports both Azure role-based access control (Azure RBAC) and POSIX-like access control lists (ACLs).

4.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

- CREATE EXTERNAL DATA SOURCE
- CREATE EXTERNAL FILE FORMAT
- CREATE EXTERNAL TABLE
- CREATE EXTERNAL TABLE AS SELECT
- CREATE DATABASE SCOPED CREDENTIAL

Answer Area

Answer:

Commands

- CREATE EXTERNAL DATA SOURCE
- CREATE EXTERNAL FILE FORMAT
- CREATE EXTERNAL TABLE
- CREATE EXTERNAL TABLE AS SELECT
- CREATE DATABASE SCOPED CREDENTIAL

Answer Area

- CREATE EXTERNAL DATA SOURCE
- CREATE EXTERNAL FILE FORMAT
- CREATE EXTERNAL TABLE AS SELECT

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.

5.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.

Table type to store the product sales transactions:

▼
Hash
Round-robin
Replicated

When creating the table for sales transactions:

▼
Configure a clustered index.
Set the distribution column to product ID.
Set the distribution column to the sales date.

Answer:

Table type to store the product sales transactions:

▼
Hash
Round-robin
Replicated

When creating the table for sales transactions:

▼
Configure a clustered index.
Set the distribution column to product ID.
Set the distribution column to the sales date.

Explanation:

Graphical user interface, text, application, chat or text message
Description automatically generated

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.