

ITCERTMASTER

Safe, simple and fast. 100% Pass guarantee!



Most reliable providers of international IT certification exam materials, 100% pass guarantee!



<http://www.itcertmaster.com>

IT Certification Guaranteed, The Easy Way!

Exam : **DP-203J**

Title : Data Engineering on
Microsoft Azure (DP-
203日本語版)

Vendor : Microsoft

Version : DEMO

QUESTION NO: 1

統合パイプラインにバージョン管理された変更を実装する必要があります。ソリューションは、データ統合の要件を満たす必要があります。

どの順序でアクションを実行する必要がありますか？回答するには、すべてのアクションをアクションのリストから回答領域に移動し、正しい順序で配置します。

Actions	Answer Area
Publish changes.	
Create a feature branch.	
Merge changes.	
Create a repository and a main branch.	
Create a pull request.	

Answer:

Actions	Answer Area
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:

Create a repository and a main branch

Create a feature branch

Create a pull request

Merge changes

Publish changes

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

Reference:

<https://docs.microsoft.com/en-us/azure/devops/pipelines/repos/pipeline-options-for-git>

Topic 1, Contoso 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 an 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 Azure AD 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 that 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 Synapse 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.

QUESTION NO: 2

Twitterフィード用のデータ取り込みおよびストレージソリューションを設計する必要があります。ソリューションは、顧客の感情分析の要件を満たす必要があります。

ソリューションに何を含める必要がありますか回答するには、回答領域で適切なオプションを選択します。注各正しい選択は、1ポイントの価値があります。

Answer Area

To increase the throughput of ingesting the Twitter feeds:

- Configure Event Hubs partitions.
- Enable Auto-Inflate in Event Hubs.
- Use Event Hubs Dedicated.

To store the Twitter feed data, use:

- An Azure Data Lake Storage Gen2 account
- An Azure Databricks high concurrency cluster
- An Azure General-purpose v2 storage account in the Premium tier

Answer:

Answer Area

To increase the throughput of ingesting the Twitter feeds:

- Configure Event Hubs partitions.
- Enable Auto-Inflate in Event Hubs.
- Use Event Hubs Dedicated.

To store the Twitter feed data, use:

- An Azure Data Lake Storage Gen2 account
- An Azure Databricks high concurrency cluster
- An Azure General-purpose v2 storage account in the Premium tier

Explanation:

To increase the throughput of ingesting the Twitter feeds:

- Configure Event Hubs partitions.
- Enable Auto-Inflate in Event Hubs.
- Use Event Hubs Dedicated.

To store the Twitter feed data, use:

- An Azure Data Lake Storage Gen2 account
- An Azure Databricks high concurrency cluster
- An Azure General-purpose v2 storage account in the Premium tier

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).

Reference:

<https://docs.microsoft.com/en-us/azure/event-hubs/event-hubs-features>

<https://docs.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-access-control>

QUESTION NO: 3

オンプレミスのデータソースと Azure Synapse Analytics を統合する必要があります。ソリューションは、データ統合の要件を満たす必要があります。

。

どのタイプの統合ランタイムを使用する必要がありますか？

- A. Azure-SSIS 統合ランタイム
- B. セルフホステッド統合ランタイム
- C. Azure 統合ランタイム

Answer: C

QUESTION NO: 4

製品販売トランザクションのパーティションを設計する必要があります。ソリューションは、販売トランザクションデータセットの要件を満たす必要があります。

ソリューションに何を含める必要がありますか？回答するには、回答領域で適切なオプションを選択します。

注：正しい選択はそれぞれ1ポイントの価値があります。

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:

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	

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>

QUESTION NO: 5

Pool1 という名前の Azure Synapse Analytics 専用 SQL プールを含む Azure サブスクリプションがあります。

Azure Monitor を使用します。

Pool1 で実行されたクエリのパフォーマンスを監視する必要があります。

どのログをクエリする必要がありますか？

- A. SynapseSqlPoolWaits
- B. SynapseSqlPoolSqlRequests
- C. SynapseSqlPoolExecRequests
- D. SynapseSqlPoolRequestSteps

Answer: C

QUESTION NO: 6

Twitterのティードデータレコードのデータ保持ソリューションを設計する必要があります。

ソリューションは、顧客の感情分析の要件を満たす必要があります。

ソリューションにどのAzureStorage機能を含める必要がありますか？

- A. 時間ベースの保持
- B. フィードを変更する
- C. ソフト削除
- D. lifecycle管理

Answer: C

QUESTION NO: 7

小売店テーブルの代理キーを実装する必要があります。ソリューションは、販売トランザクション データセットの要件を満たす必要があります。

何を作成する必要がありますか？

- A. IDENTITY プロパティを持つテーブル
- B. システム バージョン管理されたテンポラル テーブル
- C. ユーザー定義の SEQUENCE オブジェクト
- D. FOREIGN KEY 制約のあるテーブル

Answer: A

Explanation:

Scenario: Implement a surrogate key to account for changes to the retail store addresses.

A surrogate key on a table is a column with a unique identifier for each row. The key is not generated from the table data. Data modelers like to create surrogate keys on their tables when they design data warehouse models. You can use the IDENTITY property to achieve this goal simply and effectively without affecting load performance.

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-identity>

QUESTION NO: 8

製品販売トランザクションのデータ

ストレージ構造を設計する必要があります。ソリューションは、販売トランザクションデータセットの要件を満たす必要があります。

ソリューションには何を含める必要がありますか?回答するには、回答エリアで適切なオプションを選択してください。

注: 正しい選択ごとに 1 ポイントの価値があります。

Answer Area

Table type to store the product sales transactions:

When creating the table for sales transactions:

Answer:

Answer Area

Table type to store the product sales transactions:

When creating the table for sales transactions:

Explanation:

Table type to store the product sales transactions:

When creating the table for sales transactions:

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/>

QUESTION NO: 9

販売トランザクション データを格納するための Azure Synapse Analytics データベース オブジェクトを実装する必要があります。ソリューションは、販売トランザクション データセットの要件を満たす必要があります。

販売トランザクション

データセットの要件を満たす必要があるソリューションはどれですか。

あなたは何をすべきか？回答するには、回答エリアで適切なオプションを選択してください。

注: 正しい選択ごとに 1 ポイントの価値があります。

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:

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	

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>

QUESTION NO: 10

Twitterフィードデータレコードのデータ保持ソリューションを設計する必要があります。ソリューションは、顧客の感情分析の要件を満たす必要があります。

ソリューションに含める必要があるAzureストレージ機能はどれですか？

- A. フィードを変更する
- B. ソフト削除
- C. 時間ベースの保持
- D. ライフサイクル管理

Answer: B

Explanation:

Scenario: Purge Twitter feed data records that are older than two years.

Data sets have unique lifecycles. Early in the lifecycle, people access some data often. But the need for access often drops drastically as the data ages. Some data remains idle in the cloud and is rarely accessed once stored. Some data sets expire days or months after creation, while other data sets are actively read and modified throughout their lifetimes. Azure Storage lifecycle management offers a rule-based policy that you can use to transition blob data to the appropriate access tiers or to expire data at the end of the data lifecycle.

Reference:

<https://docs.microsoft.com/en-us/azure/storage/blobs/lifecycle-management-overview>

QUESTION NO: 11

トランザクションデータの分析ストレージソリューションを設計する必要があります。ソリューションは、販売トランザクションデータセットの要件を満たす必要があります。

ソリューションに何を含める必要がありますか？回答するには、回答領域で適切なオプションを選択します。

注：正しい選択はそれぞれ1ポイントの価値があります。

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:

Table type to store retail store data:

	▼
Hash	
Replicated	
Round-robin	

Table type to store promotional data:

	▼
Hash	
Replicated	
Round-robin	

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.

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql-data-warehouse/sql-data-warehouse-tables-distribute>

QUESTION NO: 12

専用の SQL プールで Twitter フィード

データを分析できることを確認する必要があります。ソリューションは、顧客感情分析の要件を満たす必要があります。

順番に実行する必要がある 3 つのトランザクション SQL DDL

コマンドはどれですか?回答するには、適切なコマンドをコマンドのリストから回答エリアに移動し、正しい順序で並べてください。

注:

複数の回答の選択肢が正しいです。選択した正しい注文のいずれかに対してクレジットを受け取ります。

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:

```
CREATE EXTERNAL DATA SOURCE
```

```
CREATE EXTERNAL FILE FORMAT
```

```
CREATE EXTERNAL TABLE AS SELECT
```

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>

QUESTION NO: 13

Pcol1 という名前の Azure Synapse Analytics 専用 SQL プールがあります。Pool1 には、tablet という名前のテーブルが含まれています。table1 に 5 TB のデータを読み込みます。

table1 の列ストア圧縮が最大化されていることを確認する必要があります。

どのステートメントを実行する必要がありますか？

- A. DBCC IIDXDEFRAG (pool1, table1)
- B. ALTER INDEX ALL on table REORGANIZE
- C. DBCC DBREINTEX (table)
- D. ALTER INDEX ALL on table REBUILD

Answer: C

Topic 2, Litware, inc. Case Study

Case study

This is a case study. Case studies are not timed separately. You can use as much exam time as you would like to complete each case. However, there may be additional case studies and sections on this exam. You must manage your time to ensure that you are able to complete all questions included on this exam in the time provided.

To answer the questions included in a case study, you will need to reference information that is provided in the case study. Case studies might contain exhibits and other resources that provide more information about the scenario that is described in the case study. Each question is independent of the other questions in this case study.

At the end of this case study, a review screen will appear. This screen allows you to review your answers and to make changes before you move to the next section of the exam. After you begin a new section, you cannot return to this section.

To start the case study

To display the first question in this case study, click the Next button. Use the buttons in the left pane to explore the content of the case study before you answer the questions. Clicking these buttons displays information such as business requirements, existing environment, and problem statements. If the case study has an All Information tab, note that the information displayed is identical to the information displayed on the subsequent tabs. When you are ready to answer a question, click the Question button to return to the question.

Overview

Litware, Inc. owns and operates 300 convenience stores across the US. The company sells a variety of packaged foods and drinks, as well as a variety of prepared foods, such as sandwiches and pizzas.

Litware has a loyalty club whereby members can get daily discounts on specific items by providing their membership number at checkout.

Litware employs business analysts who prefer to analyze data by using Microsoft Power BI, and data scientists who prefer analyzing data in Azure Databricks notebooks.

Requirements

Business Goals

Litware wants to create a new analytics environment in Azure to meet the following requirements:

See inventory levels across the stores. Data must be updated as close to real time as possible.

Execute ad hoc analytical queries on historical data to identify whether the loyalty club discounts increase sales of the discounted products.

Every four hours, notify store employees about how many prepared food items to produce based on historical demand from the sales data.

Technical Requirements

Litware identifies the following technical requirements:

Minimize the number of different Azure services needed to achieve the business goals.

Use platform as a service (PaaS) offerings whenever possible and avoid having to provision virtual machines that must be managed by Litware.

Ensure that the analytical data store is accessible only to the company's on-premises

network and Azure services.

Use Azure Active Directory (Azure AD) authentication whenever possible.

Use the principle of least privilege when designing security.

Stage Inventory data in Azure Data Lake Storage Gen2 before loading the data into the analytical data store.

Litware wants to remove transient data from Data Lake Storage once the data is no longer in use. Files that have a modified date that is older than 14 days must be removed.

Limit the business analysts' access to customer contact information, such as phone numbers, because this type of data is not analytically relevant.

Ensure that you can quickly restore a copy of the analytical data store within one hour in the event of corruption or accidental deletion.

Planned Environment

Litware plans to implement the following environment:

The application development team will create an Azure event hub to receive real-time sales data, including store number, date, time, product ID, customer loyalty number, price, and discount amount, from the point of sale (POS) system and output the data to data storage in Azure.

Customer data, including name, contact information, and loyalty number, comes from Salesforce, a SaaS application, and can be imported into Azure once every eight hours. Row modified dates are not trusted in the source table.

Product data, including product ID, name, and category, comes from Salesforce and can be imported into Azure once every eight hours. Row modified dates are not trusted in the source table.

Daily inventory data comes from a Microsoft SQL server located on a private network.

Litware currently has 5 TB of historical sales data and 100 GB of customer data. The company expects approximately 100 GB of new data per month for the next year.

Litware will build a custom application named FoodPrep to provide store employees with the calculation results of how many prepared food items to produce every four hours.

Litware does not plan to implement Azure ExpressRoute or a VPN between the on-premises network and Azure.

QUESTION NO: 14

フォルダーを含む Azure Blob ストレージ アカウントがあります。フォルダーには 120,000 個のファイルが含まれています。各ファイルには 62 列が含まれます。

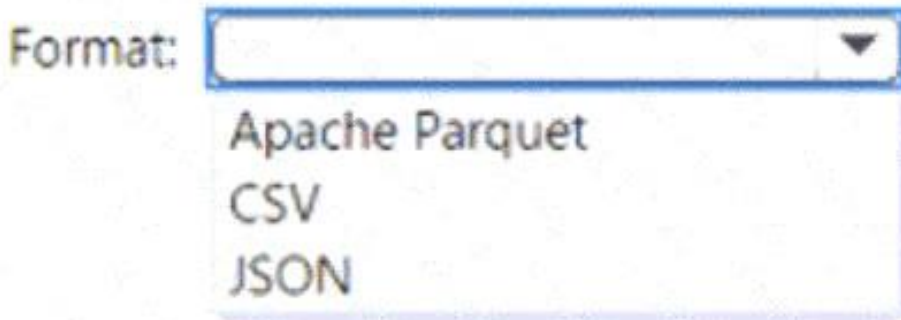
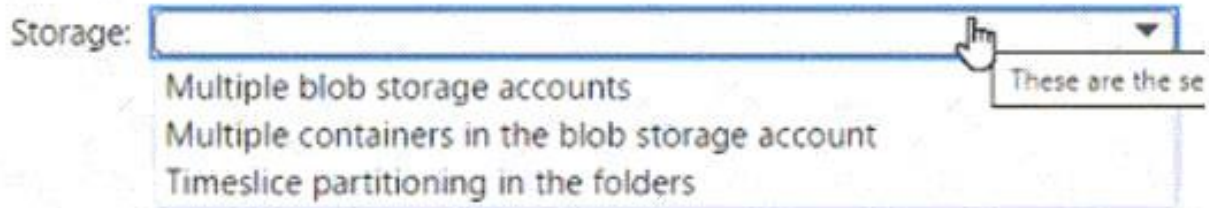
毎日、1,500 個の新しいファイルがフォルダーに追加されます。

新しいファイルごとに 5 つのデータ列を Azure Synapse Analytics

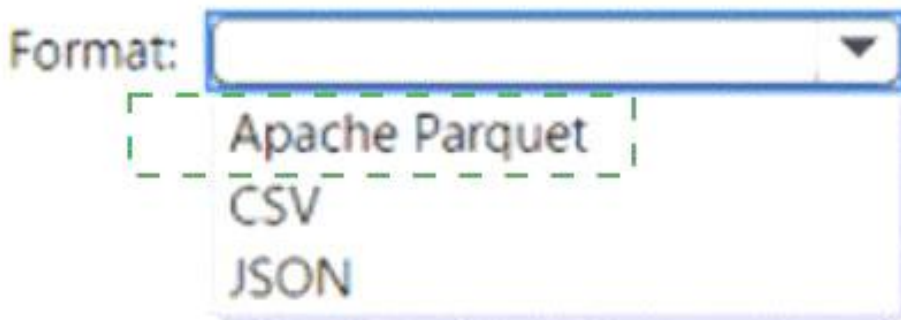
ワークスペースに段階的に読み込むことを計画しています。

増分ロードの実行にかかる時間を最小限に抑える必要があります。

ファイルとフォーマットを保存するために何を使用する必要がありますか？



Answer:



Explanation:

Box 1 = timeslice partitioning in the folders This means that you should organize your files into folders based on a time attribute, such as year, month, day, or hour. For example, you can have a folder structure like /yyyy /mm/dd/file.csv. This way, you can easily identify and load only the new files that are added each day by using a time filter in your Azure Synapse pipeline¹². Timeslice partitioning can also improve the performance of data loading and querying by reducing the number of files that need to be scanned

Box = 2 Apache Parquet This is because Parquet is a columnar file format that can efficiently store and compress data with many columns. Parquet files can also be partitioned by a time attribute, which can improve the performance of incremental loading

and querying by reducing the number of files that need to be scanned¹²³. Parquet files are supported by both dedicated SQL pool and serverless SQL pool in Azure Synapse Analytics².

QUESTION NO: 15

Apache Kafkaで発生し、Azure Data Lake

StorageGen2に出力されるストリーミングデータを集約するソリューションを計画しています。ストリーム処理ソリューションを実装する開発者はJavaを使用しますが、ストリーミングデータの処理にどのサービスを使用することをお勧めしますか？

- A. Azure Data Factory
- B. Azure Stream Analytics
- C. Azure Databricks
- D. Azure イベントハブ

Answer: C

Explanation:

<https://docs.microsoft.com/en-us/azure/architecture/data-guide/technology-choices/stream-processing>

QUESTION NO: 16

Azure データ ファクトリを含む Azure サブスクリプションがあります。

Azure Data Factory アクティビティ JSON を編集しています。

スクリプトでは、Azure Blob Storage

から複数の宛先にファイルをコピーする必要があります。ソリューションでは、ソースファイルと宛先ファイルのフォルダーパスが一貫していることを確認する必要があります

。

スクリプトをどのように完成させる必要がありますか？

回答するには、適切な値を正しいターゲットにドラッグします。各値は、1

回、複数回、またはまったく使用されない場合があります。コンテンツを表示するには、ページ間の分割バーをドラッグするか、スクロールする必要がある場合があります。

注意: 正解を選ぶごとに1ポイント獲得できます

Values

Answer Area

- FlattenHierarchy {
- ForEach
- MergeFiles
- PreserveHierarchy
- Switch
- Until

```

"name": "Pipeline1",
"properties": {
  "activities": [
    {
      "name": "Activity1",
      "type": [redacted],
      "typeProperties": {
        "isSequential": "true",
        "items": {
          "value": "@pipeline
() .parameters.mySinkDatasetFolderPath",
          "type": "Expression"},
        "activities" [
          {
            "name": "MyCopyActivity",
            "type": "Copy",
            "typeProperties": {
              "source": {
                "type": "BlobSource",
                "recursive": "false" },
              "sink": {
                "type": "BlobSink",
                "CopyBehavior": [redacted]
              }
            }
          }
        ]
      }
    }
  ]
}

```

Answer:

Values

Answer Area

- FlattenHierarchy {
- ForEach
- MergeFiles
- PreserveHierarchy
- Switch
- Until

```

"name": "Pipeline1",
"properties": {
  "activities": [
    {
      "name": "Activity1",
      "type": ForEach,
      "typeProperties": {
        "isSequential": "true",
        "items": {
          "value": "@pipeline
() .parameters.mySinkDatasetFolderPath",
          "type": "Expression"},
        "activities" [
          {
            "name": "MyCopyActivity",
            "type": "Copy",
            "typeProperties": {
              "source": {
                "type": "BlobSource",
                "recursive": "false" },
              "sink": {
                "type": "BlobSink",
                "CopyBehavior": Switch
              }
            }
          }
        ]
      }
    }
  ]
}

```

Explanation:

The screenshot shows an exam question interface. On the left, under the heading "Values", there is a list of activity types: FlattenHierarchy, ForEach, MergeFiles, PreserveHierarchy, Switch, and Until. Each item is in a box with a small icon on the left. A vertical bar with four blue dots is positioned to the right of this list. On the right, under the heading "Answer Area", there is a JSON configuration for a pipeline. The configuration is as follows:

```
{
  "name": "Pipeline1",
  "properties": {
    "activities": [
      {
        "name": "Activity1",
        "type": ForEach,
        "typeProperties": {
          "isSequential": "true",
          "items": {
            "value": "@pipeline().parameters.mySinkDatasetFolderPath",
            "type": "Expression" },
          "activities": [
            {
              "name": "MyCopyActivity",
              "type": "Copy",
              "typeProperties": {
                "source": {
                  "type": "BlobSource",
                  "recursive": "false" },
                "sink": {
                  "type": "Blobsink",
                  "CopyBehaviour": Switch
                }
              }
            }
          ]
        }
      }
    ]
  }
}
```

QUESTION NO: 17

Litware オンプレミス ネットワークの外部のユーザーが分析データストアにアクセスできないようにするには、何を推奨しますか？

- A. サーバー レベルの仮想ネットワーク ルール
- B. データベース レベルの仮想ネットワーク ルール
- C. データベース レベルのファイアウォール IP ルール
- D. サーバー レベルのファイアウォール IP ルール

Answer: A

Explanation:

Virtual network rules are one firewall security feature that controls whether the database server for your single databases and elastic pool in Azure SQL Database or for your databases in SQL Data Warehouse accepts communications that are sent from particular subnets in virtual networks.

Server-level, not database-level: Each virtual network rule applies to your whole Azure SQL Database server, not just to one particular database on the server. In other words, virtual network rule applies at the serverlevel, not at the database-level.

References:

<https://docs.microsoft.com/en-us/azure/sql-database/sql-database-vnet-service-endpoint-rule-overview>

QUESTION NO: 18

Azure Synapse Analytics にエンタープライズ データ ウェアハウスを実装します。

サイズが 10 テラバイト (TB) の大きなファクト テーブルがあります。

受信クエリは、主キーの SaleKey 列を使用して、次の表に表示されるデータを取得します

。

SaleKey	CityKey	CustomerKey	StockItemKey	InvoiceDateKey	Quantity	UnitPrice	TotalExcludingTax
49309	90858	70	69	10/22/13	8	16	128
49313	55710	126	69	10/22/13	2	16	32
49343	44710	234	68	10/22/13	10	16	160
49352	66109	163	70	10/22/13	4	16	64
49488	65312	230	70	10/22/13	8	16	128
49646	85877	271	70	10/24/13	1	16	16
49798	41238	288	69	10/24/13	1	16	16

テーブルのパフォーマンスを最適化するには、大きなファクト

テーブルを複数のノードに分散する必要があります。

どのテクノロジーを使用する必要がありますか？

- A. クラスター化インデックスを使用したハッシュ分散テーブル
- B. クラスター化された列ストア インデックスを持つハッシュ分散テーブル
- C. クラスター化インデックスを使用したラウンド ロビン分散テーブル
- D. クラスター化された列ストア インデックスを使用したラウンド ロビン分散テーブル
- E. 分散レプリケートのあるヒープ テーブル

Answer: B

Explanation:

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

Columnstore indexes can achieve up to 100x better performance on analytics and data warehousing workloads and up to 10x better data compression than traditional rowstore indexes.

Reference:

<https://docs.microsoft.com/en-us/azure/sql-data-warehouse/sql-data-warehouse-tables-distribute>

<https://docs.microsoft.com/en-us/sql/relational-databases/indexes/columnstore-indexes-query-performance>

QUESTION NO: 19

Azure Cosmos DB 分析ストアと WS 1 という名前の Azure Synapse Analytics

ワークスペースを含む Azure サブスクリプションがあります。WS1 には、サーバーレス SQL プールという名前 Pool1 があります。

Pool1 を使用して次のクエリを実行します。

```
WITH IDENTITY = 'SHARED #
SECRET = 'fed4347479872423433563653456345ddfa==' ;

SELECT clientID AS ClientID,
       client AS ClientName
FROM OPENROWSET
(
    PROVIDER = 'CosmosDB',
    CONNECTION = 'Account=account1;Database=database1',
    OBJECT = 'clients',
    SERVER_CREDENTIAL = 'AccountCred'
)
WITH
```

```
(
    clientID int,
    client varchar(50),
    streetAddress varchar(100)
) AS c;
```

次の各ステートメントについて、そのステートメントが true の場合は [はい] を選択します。それ以外の場合は、「いいえ」を選択します。
注: 正しく選択するたびに 1 ポイントの価値があります。

Answer Area

Statements	Yes	No
The query returns three columns.	<input type="radio"/>	<input type="radio"/>
The container being queried is named clients.	<input type="radio"/>	<input type="radio"/>
Authentication is performed by using an account key.	<input type="radio"/>	<input type="radio"/>

Answer:

Answer Area

Statements	Yes	No
The query returns three columns.	<input type="radio"/>	<input checked="" type="radio"/>
The container being queried is named <code>clients</code> .	<input checked="" type="radio"/>	<input type="radio"/>
Authentication is performed by using an account key.	<input type="radio"/>	<input checked="" type="radio"/>

Explanation:

Answer Area

Statements	Yes	No
The query returns three columns.	<input type="radio"/>	<input checked="" type="radio"/>
The container being queried is named <code>clients</code> .	<input checked="" type="radio"/>	<input type="radio"/>
Authentication is performed by using an account key.	<input type="radio"/>	<input checked="" type="radio"/>

QUESTION NO: 20

Group1 という名前のセキュリティ グループを含む Azure Active Directory (Azure AD) テナントがあります。 schema1 という名前のスキーマを含む dw1 という名前の Azure Synapse Analytics 専用 SQL プールがあります。

schema1 内のすべてのテーブルとビューに対する読み取り専用権限を Group1 に付与する必要があります。ソリューションでは、最小権限の原則を使用する必要があります。

順番に実行する必要がある 3

つのアクションはどれですか?回答するには、アクションのリストから適切なアクションを回答エリアに移動し、正しい順序で並べてください。

注:

複数の回答の選択肢が正しいです。選択した正しい注文のいずれかに対してクレジットを受け取ります。

Actions

Answer Area

- Create a database role named Role1 and grant Role1 SELECT permissions to schema1.
- Create a database role named Role1 and grant Role1 SELECT permissions to dw1.
- Assign the Azure role-based access control (Azure RBAC) Reader role for dw1 to Group1.
- Create a database user in dw1 that represents Group1 and uses the FROM EXTERNAL PROVIDER clause.
- Assign Role1 to the Group1 database user.

Answer:

Actions

- Create a database role named Role1 and grant Role1 SELECT permissions to schema1.
- Create a database role named Role1 and grant Role1 SELECT permissions to dw1.
- Assign the Azure role-based access control (Azure RBAC) Reader role for dw1 to Group1.
- Create a database user in dw1 that represents Group1 and uses the FROM EXTERNAL PROVIDER clause.
- Assign Role1 to the Group1 database user.

Answer Area

- Create a database role named Role1 and grant Role1 SELECT permissions to schema1.
- Assign Role1 to the Group1 database user.
- Assign the Azure role-based access control (Azure RBAC) Reader role for dw1 to Group1.

Explanation:

Create a database role named Role1 and grant Role1 SELECT permissions to schema1.

Assign Role1 to the Group1 database user.

Assign the Azure role-based access control (Azure RBAC) Reader role for dw1 to Group1.

Step 1: Create a database role named Role1 and grant Role1 SELECT permissions to schema You need to grant Group1 read-only permissions to all the tables and views in schema1.

Place one or more database users into a database role and then assign permissions to the database role.

Step 2: Assign Rol1 to the Group database user

Step 3: Assign the Azure role-based access control (Azure RBAC) Reader role for dw1 to Group1 Reference:

<https://docs.microsoft.com/en-us/azure/data-share/how-to-share-from-sql>

QUESTION NO: 21

storage1 という名前の Azure Data Lake Storage Gen 2 アカウントがあります。

storage1

のコンテンツにアクセスするためのソリューションを推奨する必要があります。ソリューションは、次の要件を満たす必要があります。

リストと読み取りのアクセス許可は、ストレージ アカウントレベルで付与する必要があります。

storage1 内の個々のオブジェクトに追加のアクセス許可を適用できます。

認証には、Microsoft Entra の一部である Microsoft Azure Active Directory (Azure AD) のセキュリティ プリンシパルを使用する必要があります。

何を使うべきですか？答えるには、適切なコンポーネントを正しい要件にドラッグします。
各コンポーネントは、1
回以上使用することも、まったく使用しないこともできます。ペイン間の分割バーをドラッグするか、コンテンツを表示するためにスクロールする必要がある場合があります。
注: それぞれの正しい選択は 1 ポイントの価値があります。

Components

- Access control lists (ACLs)
- Role-based access control (RBAC) roles
- Shared access signatures (SAS)
- Shared account keys

Answer Area

To grant permissions at the storage account level:

To grant permissions at the object level:

Answer:

Components

- Access control lists (ACLs)
- Role-based access control (RBAC) roles
- Shared access signatures (SAS)
- Shared account keys

Answer Area

To grant permissions at the storage account level:

To grant permissions at the object level:

Explanation:

Box 1: Role-based access control (RBAC) roles

List and read permissions must be granted at the storage account level.

Security principals from Microsoft Azure Active Directory (Azure AD), part of Microsoft Entra, must be used for authentication.

Role-based access control (Azure RBAC)

Azure RBAC uses role assignments to apply sets of permissions to security principals. A security principal is an object that represents a user, group, service principal, or managed identity that is defined in Azure Active Directory (AD). A permission set can give a security principal a "coarse-grain" level of access such as read or write access to all of the data in a storage account or all of the data in a container.

Box 2: Access control lists (ACLs)

Additional permissions can be applied to individual objects in storage1.

Access control lists (ACLs)

ACLs give you the ability to apply "finer grain" level of access to directories and files. An ACL is a permission construct that contains a series of ACL entries. Each ACL entry associates security principal with an access level.

Reference: <https://learn.microsoft.com/en-us/azure/storage/blobs/data-lake-storage-access-control-model>

QUESTION NO: 22

Azure Synapse サーバーレス SQL プールがあります。

OPENROWSET 関数を使用して、ファイルから JSON

ドキュメントを読み取る必要があります。

クエリをどのように完了する必要がありますか？

回答するには、回答エリアで適切なオプションを選択します。

注: それぞれの正しい選択は 1 ポイントの価値があります。

Answer Area

```
SELECT *
FROM OPENROWSET
(
    BULK
    'https://sourcedatalake.blob.core.windows.net/public/docs.json',
    FORMAT = 'JSON',
    FIELDTERMINATOR = '0x0b',
    FIELDQUOTE = '0x0b',
    ROWTERMINATOR = '0x09'
)
WITH (jsondoc nvarchar(1000) onDocuments)
```

Answer:

Answer Area

```
SELECT *
FROM OPENROWSET
(
    BULK
    'https://sourcedatalake.blob.core.windows.net/public/docs.json',
    FORMAT = 'DELTA',
    FIELDTERMINATOR = '0x0b',
    FIELDQUOTE = '0x0b',
    ROWTERMINATOR = '0x0b'
)
WITH (jsondoc nvarchar(1000) onDocuments)
```

Explanation:

Answer Area

```
SELECT *
FROM OPENROWSET
(
    BULK
    'https://sourcedatalake.blob.core.windows.net/public/docs.json',
    FORMAT = 'JSON',
    FIELDTERMINATOR = '0x0b',
    FIELDQUOTE = '0x0b',
    ROWTERMINATOR = '0x0b'
)
WITH (jsondoc nvarchar(max) AS JsonDocuments
```

QUESTION NO: 23

storage1という名前のAzureBlobStorageアカウントとPool1という名前のAzureSynapseAnalytics専用SQLプールを含むAzureサブスクリプションがあります。

storage1にデータを保存する必要があります。データはPool1によって読み取られます。ソリューションは、次の要件を満たしている必要があります。

Pool1を有効にして、クエリで不要な列と行をスキップします。

列統計を自動的に作成します。

ファイルのサイズを最小化します。

どのタイプのファイルを使用する必要がありますか？

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

Answer: B

Explanation:

Automatic creation of statistics is turned on for Parquet files. For CSV files, you need to create statistics manually until automatic creation of CSV files statistics is supported.

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/develop-tables-statistics>

QUESTION NO: 24

Azure Stream Analytics を使用して、ストリーミング ソーシャル メディア

データを取り込む予定です。データは Azure Data Lake Storage

のファイルに保存され、Azure Synapse Analytics で Azure Databricks と PolyBase を使用して使用されます。

ファイルに対する Databricks および PolyBase

からのクエリで発生するエラーが最小限になるように、Stream Analytics

データ出力形式を推奨する必要があります。ソリューションは、タイルが迅速にクエリでき、データ型情報が保持されることを保証する必要があります。

何をお勧めしますか？

- A. 寄木細工
- B. アプロ
- C.CSV
- D. JSON

Answer: A

Explanation:

<https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-define-outputs>

QUESTION NO: 25

Pool1 という名前の Azure Synapse Analytics 専用 SQL プールと storage1 という名前の Azure Data Lake Storage アカウントを含む Azure サブスクリプションがあります。Storage1 には安全な転送が必要です。storage1 の .orc ファイルを読み取るために使用する外部データ ソースを Pool1 に作成する必要があります。

コードをどのように完成させる必要がありますか？

回答するには、回答エリアで適切なオプションを選択します。

注: それぞれの正しい選択は 1 ポイントの価値があります。

Answer Area


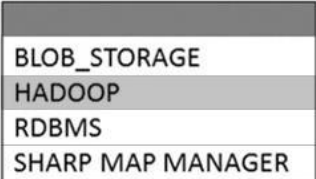
```
CREATE EXTERNAL DATA SOURCE AzureDataLakeStore
WITH
( Location1 = 'abfs://data@newyorktaxidataset.dfs.core.windows.net' ,
credential = ADLS_credential ,
TYPE = BLOB_STORAGE
);
```

Answer:

Answer Area

```
CREATE EXTERNAL DATA SOURCE AzureDataLakeStore
WITH
( Location1 `  ://data@newyorktaxidataset.dfs.core.windows.net' ,
credential = ADLS_credential ,
TYPE - 
);
```

Explanation:

```
CREATE EXTERNAL DATA SOURCE AzureDataLakeStore
WITH
( Location1 `  ://data@newyorktaxidataset.dfs.core.windows.net' ,
credential = ADLS_credential ,
TYPE - 
);
```

Reference:

<https://docs.microsoft.com/en-us/sql/t-sql/statements/create-external-data-source-transact-sql?view=azure-sqldw-latest&preserve-view=true&tabs=dedicated>

QUESTION NO: 26

Azure Databricks 上の Delta Lake のテーブルを使用する 2 つのソリューションを設計しています。

以下の作業にかかる時間を最小限に抑える必要があります。

*パーティション化されていないテーブルに対するクエリ

*パーティション化されていない列の結合

解決策にはどの 2 つのオプションを含める必要がありますか?

それぞれの正解は解決策の一部を示しています。

(正しい回答を選択し、Microsoft Azure のデータエンジニアリングに基づいて回答を裏付ける説明と参考資料を提供してください)

- A. Zオーダー
- B. Apache Spark キャッシュ
- C. 動的ファイルプルーニング (DFP)
- D. クローンコマンド

Answer: A C

Explanation:

According to the information I found on the web, two options that you should include in the solution to minimize how long it takes to perform queries and joins on non-partitioned tables are:

Z-Ordering: This is a technique to colocate related information in the same set of files. This co-locality is automatically used by Delta Lake in data-skipping algorithms. This behavior dramatically reduces the amount of data that Delta Lake on Azure Databricks needs to read.

Apache Spark caching: This is a feature that allows you to cache data in memory or on disk for faster access.

Caching can improve the performance of repeated queries and joins on the same data. You can cache Delta tables using the `CACHE TABLE` or `CACHE LAZY` commands.

To minimize the time it takes to perform queries against non-partitioned tables and joins on non-partitioned columns in Delta Lake on Azure Databricks, the following options should be included in the solution:

A:Z-Ordering: Z-Ordering improves query performance by co-locating data that share the same column values in the same physical partitions. This reduces the need for shuffling data across nodes during query execution. By using Z-Ordering, you can avoid full table scans and reduce the amount of data processed.

B:Apache Spark caching: Caching data in memory can improve query performance by reducing the amount of data read from disk. This helps to speed up subsequent queries that need to access the same data. When you cache a table, the data is read from the data source and stored in memory. Subsequent queries can then read the data from memory, which is much faster than reading it from disk.

References:

Delta Lake on Databricks: <https://docs.databricks.com/delta/index.html>

Best Practices for Delta Lake on Databricks: <https://databricks.com/blog/2020/05/14/best-practices-for-delta-lake-on-databricks.html>

QUESTION NO: 27

storage1 という名前の Azure Data Lake Storage Gen2 アカウントと Workspace1 という名前の Azure Synapse Analytics ワークスペースを含む Azure サブスクリプションがあります。Workspace1 にはサーバーレス SQL プールがあります。サーバーレス SQL プールを使用して、ストレージ 1 内のファイルから顧客の注文を照会します。次のクエリを実行します。

選択*

```
FROM OPENROWSET(BULK 'https://storage1.blob.core.windows.net/data/orders/year ='
```

/month =* / *.*', FORMAT = 'parquet') AS customerorders WHERE customerorders.
filepath(1) = '2024' AND customerorders.filepath(2) IN ('3','4');
次の各ステートメントについて、ステートメントが true の場合は [はい]
を選択します。それ以外の場合は [いいえ] を選択します。
注意: 正しい選択ごとに 1 ポイントが付与されます。

Answer Area

Statements	Yes	No
storage1 provides a hierarchical namespace.	<input type="radio"/>	<input type="radio"/>
Files from March 2025 will be included.	<input type="radio"/>	<input type="radio"/>
Only files that have a Parquet file extension will be included.	<input type="radio"/>	<input type="radio"/>

Answer:

Answer Area

Statements	Yes	No
storage1 provides a hierarchical namespace.	<input checked="" type="radio"/>	<input type="radio"/>
Files from March 2025 will be included.	<input type="radio"/>	<input checked="" type="radio"/>
Only files that have a Parquet file extension will be included.	<input checked="" type="radio"/>	<input type="radio"/>

Explanation:

Storage1 provides a hierarchical namespace: Yes

Files from March 2025 will be included: No

Only files that have a Parquet file extension will be included: Yes

Query Breakdown

* Data Source:

* The OPENROWSET function queries data stored in Azure Data Lake Storage Gen2 (storage1) using the serverless SQL pool in Synapse Analytics.

* The data is stored in Parquet files in the folder structure data/orders/year=YYYY/month=MM/.

* Query Filter:

* The filter conditions in the query are:

* customerorders.filepath(1) = '2024': Limits the query to files in the folder year=2024.

* customerorders.filepath(2) IN ('3', '4'): Limits the query to files in the subfolders month=3 or month=4.

* File Format:

* The FORMAT = 'parquet' clause specifies that only Parquet files will be queried.

Statements Analysis

* Storage1 provides a hierarchical namespace.answer: Yes

* Azure Data Lake Storage Gen2 supports a hierarchical namespace, which enables folder-based organization.

* The folder structure (e.g., data/orders/year=2024/month=3/) demonstrates the use of a hierarchical namespace.

* Files from March 2025 will be included.answer: No

- * The query explicitly filters for year=2024, so files from 2025 will not be included in the results.
- * Only files that have a Parquet file extension will be included. answer: Yes
- * The FORMAT = 'parquet' clause in the query ensures that only Parquet files are queried. Files with other extensions (e.g., .csv or .json) will not be included.

QUESTION NO: 28

あなたは、500 台の車両の監視ソリューションを設計しています。各車両には、1 分に 1 回 Azure イベント ハブにデータを送信する GPS 追跡デバイスがあります。

Azure Data Lake Storage Gen2 コンテナに CSV

ファイルがあります。このファイルは、各車両が存在するはずの予想される地理的領域を維持します。

GPS 位置が予想範囲外にある場合、メッセージが別のイベント ハブに追加され、30 秒以内に処理されるようにする必要があります。ソリューションは、コストを最小限に抑える必要があります。

ソリューションには何を含める必要がありますか? 回答するには、回答エリアで適切なオプションを選択してください。

注: 正しい選択ごとに 1 ポイントの価値があります。

Service:

	▼
An Azure Synapse Analytics Apache Spark pool	
An Azure Synapse Analytics serverless SQL pool	
Azure Data Factory	
Azure Stream Analytics	

Window:

	▼
Hopping	
No window	
Session	
Tumbling	

Analysis type:

	▼
Event pattern matching	
Lagged record comparison	
Point within polygon	
Polygon overlap	

Answer:

Service: ▼

An Azure Synapse Analytics Apache Spark pool
An Azure Synapse Analytics serverless SQL pool
Azure Data Factory
Azure Stream Analytics

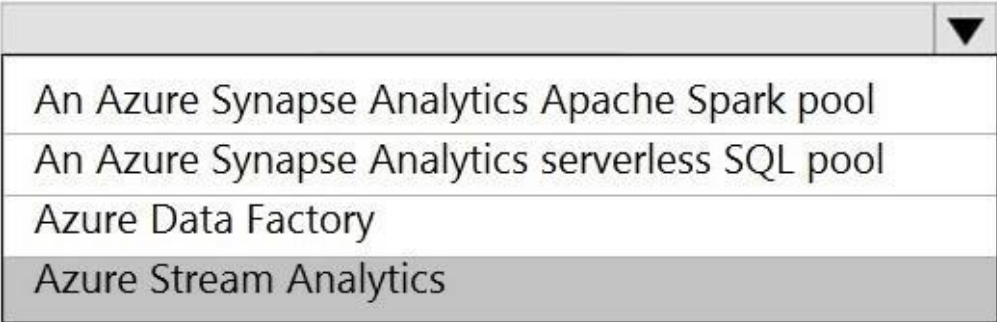
Window: ▼

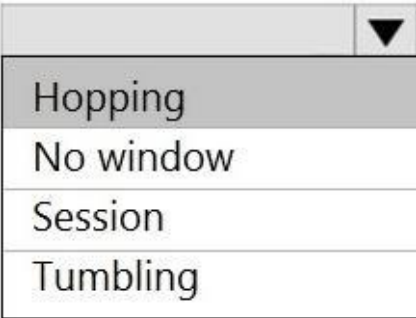
Hopping
No window
Session
Tumbling

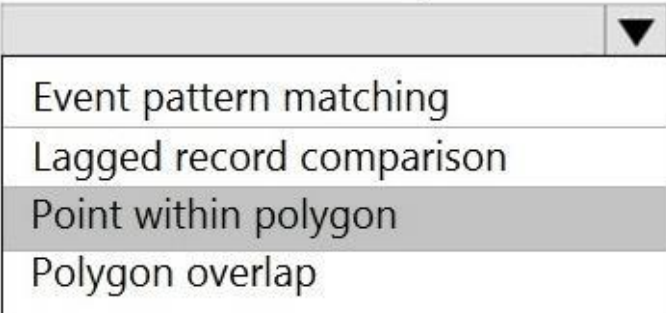
Analysis type: ▼

Event pattern matching
Lagged record comparison
Point within polygon
Polygon overlap

Explanation:

Service: 
An Azure Synapse Analytics Apache Spark pool
An Azure Synapse Analytics serverless SQL pool
Azure Data Factory
Azure Stream Analytics

Window: 
Hopping
No window
Session
Tumbling

Analysis type: 
Event pattern matching
Lagged record comparison
Point within polygon
Polygon overlap

Box 1: Azure Stream Analytics

Box 2: Hopping

Hopping window functions hop forward in time by a fixed period. It may be easy to think of them as Tumbling windows that can overlap and be emitted more often than the window size. Events can belong to more than one Hopping window result set. To make a Hopping window the same as a Tumbling window, specify the hop size to be the same as the window size.

Box 3: Point within polygon

Reference:

<https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-window-functions>

QUESTION NO: 29

顧客用のJSONファイルを含むAzureData Lake

StorageGen2アカウントがあります。このファイルには、FirstNameとLastNameという名前の2つの属性が含まれています。

Azure Databricksを使用して、JSONファイルからAzure SynapseAnalyticsテーブルにデータをコピーする必要があります。

FirstNameとLastNameの値を連結する新しい列を作成する必要があります。

次のコンポーネントを作成します。

- * AzureSynapseの宛先テーブル
- * AzureBlobストレージコンテナー
- * サービスプリンシパル

アクションを実行する順序はどれですか。回答するには、適切なアクションをアクションのリストから回答領域に移動し、正しい順序に並べます。

Actions

Answer Area

- Mount the Data Lake Storage onto DBFS.
- Write the results to a table in Azure Synapse.
- Specify a temporary folder to stage the data.
- Read the file into a data frame.
- Perform transformations on the data frame.

Answer:

Actions

Answer Area

Mount the Data Lake Storage onto DBFS.	Mount the Data Lake Storage onto DBFS.
Write the results to a table in Azure Synapse.	Read the file into a data frame.
Specify a temporary folder to stage the data.	Perform transformations on the data frame.
Read the file into a data frame.	Specify a temporary folder to stage the data.
Perform transformations on the data frame.	Write the results to a table in Azure Synapse.

Explanation:

Mount the Data Lake Storage onto DBFS.

Read the file into a data frame.

Perform transformations on the data frame.

Specify a temporary folder to stage the data.

Write the results to a table in Azure Synapse.

Step 1: Mount the Data Lake Storage onto DBFS

Begin with creating a file system in the Azure Data Lake Storage Gen2 account.

Step 2: Read the file into a data frame.

You can load the json files as a data frame in Azure Databricks.

Step 3: Perform transformations on the data frame.

Step 4: Specify a temporary folder to stage the data

Specify a temporary folder to use while moving data between Azure Databricks and Azure Synapse.

Step 5: Write the results to a table in Azure Synapse.

You upload the transformed data frame into Azure Synapse. You use the Azure Synapse connector for Azure Databricks to directly upload a dataframe as a table in a Azure Synapse.

Reference:

<https://docs.microsoft.com/en-us/azure/azure-databricks/databricks-extract-load-sql-data-warehouse>

QUESTION NO: 30

Pool1という名前のAzureSynapse Analytics ApacheSparkプールがあります。

JSONファイルをAzureData Lake

StorageGen2コンテナからPool1のテーブルにロードすることを計画しています。構造とデータ型はファイルによって異なります。

ファイルをテーブルにロードする必要があります。ソリューションは、ソースデータ型を維持する必要があります。

あなたは何をするべきか？

A.Azure DataFactoryでメタデータの取得アクティビティを使用します。

B.AzureSynapseデータフローで条件付き分割変換を使用します。

C.Azure SynapseAnalyticsサーバーレスSQLプールでOPENROWSETTransact-SQLコマンドを使用してデータを読み込みます。

D.PySparkを使用してデータをロードします。

Answer: A

Explanation:

Serverless SQL pool can automatically synchronize metadata from Apache Spark. A serverless SQL pool database will be created for each database existing in serverless Apache Spark pools.

Serverless SQL pool enables you to query data in your data lake. It offers a T-SQL query surface area that accommodates semi-structured and unstructured data queries.

To support a smooth experience for in place querying of data that's located in Azure Storage files, serverless SQL pool uses the OPENROWSET function with additional capabilities.

The easiest way to see to the content of your JSON file is to provide the file URL to the OPENROWSET function, specify csv FORMAT.

Reference:

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/query-json-files>

<https://docs.microsoft.com/en-us/azure/synapse-analytics/sql/query-data-storage>

QUESTION NO: 31

ステージングゾーンを含むAzureDataLakeストレージアカウントがあります。

ステージングゾーンから増分データを取り込み、Rスクリプトを実行してデータを変換し、変換されたデータをAzure

SynapseAnalyticsのデータウェアハウスに挿入する毎日のプロセスを設計する必要があります。

解決策AzureData Factoryのスケジュールトリガーを使用して、Azure

Databricksノートブックを実行するパイプラインを実行し、データウェアハウスにデータを挿入します。これで目標を達成できますか？

A.はい

B.いいえ

Answer: B

Explanation:

If you need to transform data in a way that is not supported by Data Factory, you can create a custom activity, not an Azure Databricks notebook, with your own data processing logic and use the activity in the pipeline.

You can create a custom activity to run R scripts on your HDInsight cluster with R installed.

Reference:

<https://docs.microsoft.com/en-US/azure/data-factory/transform-data>

QUESTION NO: 32

分析ワークロードで使用するために raw JSON ファイルを変換する Azure Data Lake Storage ソリューションを設計しています。

変換されたファイルの形式を推奨する必要があります。ソリューションは、次の要件を満たす必要があります。

ファイル内の各列のデータ型に関する情報が含まれています。

ファイル内の列のサブセットのクエリをサポートします。

読み取り負荷の高い分析ワークロードをサポートします。

ファイル サイズを最小化します。

何をお勧めしますか？

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

Answer: D

Explanation:

Parquet, an open-source file format for Hadoop, stores nested data structures in a flat columnar format.

Compared to a traditional approach where data is stored in a row-oriented approach, Parquet file format is more efficient in terms of storage and performance.

It is especially good for queries that read particular columns from a "wide" (with many columns) table since only needed columns are read, and IO is minimized.

Reference: <https://www.clairvoyant.ai/blog/big-data-file-formats>

QUESTION NO: 33

注:

この質問は、同じシナリオを提示する一連の質問の一部です。シリーズの各質問には、記載された目標を達成する可能性のある独自のソリューションが含まれています。一部の質問セットには複数の正解がある場合もあれば、正解がない場合もあります。

このシナリオで質問に答えた後は、その質問に戻ることはできません。その結果、これらの質問はレビュー画面に表示されません。

100 GB のファイルを含む Azure ストレージ

アカウントがあります。ファイルにはテキストと数値が含まれています。行の 75% には、平均長が 1.1 MB の説明データが含まれています。

ストレージ アカウントから Azure Synapse Analytics のエンタープライズ データウェアハウスにデータをコピーする予定です。

データが迅速にコピーされるように、ファイルを準備する必要があります。

解決策: ファイルを圧縮された区切りテキスト ファイルに変換します。

これは目標を達成していますか？

A. はい

B. いいえ

Answer: A

Explanation:

All file formats have different performance characteristics. For the fastest load, use compressed delimited text files.

Reference:

<https://docs.microsoft.com/en-us/azure/sql-data-warehouse/guidance-for-loading-data>

QUESTION NO: 34

温度という名前の Apache Spark DataFrame があります。データのサンプルを次の表に示します。

Date	Temp
...	...
18-01-2021	3
19-01-2021	4
20-01-2021	2
21-01-2021	2
...	...

Spark SQL クエリを使用して、次のテーブルを生成する必要があります。

Year	JAN	FEB	MAR	APR	MAY
2019	2.3	4.1	5.2	7.6	9.2
2020	2.4	4.2	4.9	7.8	9.1
2021	2.6	5.3	3.4	7.9	9.5

どのようにクエリを完成させますか?答えるには、適切な値を正しいターゲットにドラッグします。各値は、複数回使用することも、まったく使用しないこともできます。コンテンツを表示するには、分割バーをペイン間でドラッグするか、スクロールする必要がある場合があります。

注: 正しい選択ごとに 1 ポイントの価値があります。

Values **Answer Area**

```
SELECT * FROM (  
  SELECT YEAR(Date) Year, MONTH(Date) Month, Temp  
  FROM temperatures  
  WHERE date BETWEEN DATE '2019-01-01' AND DATE '2021-08-31'  
)  
  (   
  AVG ( ( ) (Temp AS DECIMAL(4, 1)))  
  FOR Month in (  
    1 JAN, 2 FEB, 3 MAR, 4 APR, 5 MAY, 6 JUN,  
    7 JUL, 8 AUG, 9 SEP, 10 OCT, 11 NOV, 12 DEC  
  )  
)  
ORDER BY Year ASC
```

- CAST
- COLLATE
- CONVERT
- FLATTEN
- PIVOT
- UNPIVOT

Answer:

Values **Answer Area**

```
SELECT * FROM (  
  SELECT YEAR(Date) Year, MONTH(Date) Month, Temp  
  FROM temperatures  
  WHERE date BETWEEN DATE '2019-01-01' AND DATE '2021-08-31'  
)  
  PIVOT (  
  AVG ( CAST (Temp AS DECIMAL(4, 1)))  
  FOR Month in (  
    1 JAN, 2 FEB, 3 MAR, 4 APR, 5 MAY, 6 JUN,  
    7 JUL, 8 AUG, 9 SEP, 10 OCT, 11 NOV, 12 DEC  
  )  
)  
ORDER BY Year ASC
```

- CAST
- COLLATE
- CONVERT
- FLATTEN
- PIVOT
- UNPIVOT

Explanation:

```
SELECT * FROM (  
  SELECT YEAR(Date) Year, MONTH(Date) Month, Temp  
  FROM temperatures  
  WHERE date BETWEEN DATE '2019-01-01' AND DATE '2021-08-31'  
)  
PIVOT (  
  AVG ( CAST (Temp AS DECIMAL(4, 1)))  
  
  FOR Month in (  
    1 JAN, 2 FEB, 3 MAR, 4 APR, 5 MAY, 6 JUN,  
    7 JUL, 8 AUG, 9 SEP, 10 OCT, 11 NOV, 12 DEC  
  )  
)  
ORDER BY Year ASC
```

Box 1: PIVOT

PIVOT rotates a table-valued expression by turning the unique values from one column in the expression into multiple columns in the output. And PIVOT runs aggregations where they're required on any remaining column values that are wanted in the final output.

Reference:

<https://learnsql.com/cookbook/how-to-convert-an-integer-to-a-decimal-in-sql-server/>

<https://docs.microsoft.com/en-us/sql/t-sql/queries/from-using-pivot-and-unpivot>

QUESTION NO: 35

ある企業は Azure Data Lake Storage Gen2 サービスを使用しています。

次の要件を満たすデータ アーカイブ ソリューションを設計する必要があります。

5 年より古いデータにはアクセスされる頻度は低くなりますが、要求された場合には 1 秒以内に利用できるようにする必要があります。

7 年以上古いデータにはアクセスできません。

必要な可用性を維持しながら、コストを最小限に抑える必要があります。

データはどのように管理すればよいでしょうか? 回答するには、回答エリアで適切なオプションを選択してください。

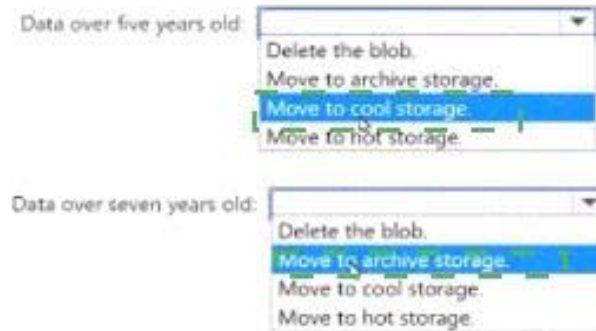
注: 正しく選択するたびに 1 ポイントの価値があります。

Answer Area



Answer:

Answer Area



Explanation:

Answer Area



Microsoft DP-203J Exam questions are updated recently, and 100% guarantee that you pass the exam successfully!
