

Databricks.Databricks-Certified-Professional-Data-Engineer.v2023-11-18.q93

試験コード:	Databricks-Certified-Professional-Data-Engineer
試験名称:	Databricks Certified Professional Data Engineer Exam
認定資格:	Databricks
無料問題数:	93
バージョン:	v2023-11-18
アクセス数:	806
ページビュー数:	930
https://www.jpnpdf.com/Databricks.Databricks-Certified-Professional-Data-Engineer.v2023-11-18.q93-mondaishu.html	

最新問題: 1

データ エンジニアリング チームはデータ サイエンス チームとデータを共有する必要があり、両方のチームが同じ組織内の異なるワークスペースを使用しています。データの共有を簡素化するために使用できる手法は次のうちどれですか？

*この質問は、組織内で複数のワークスペース間でデータがどのように共有されるかを尋ねていることに注意してください。

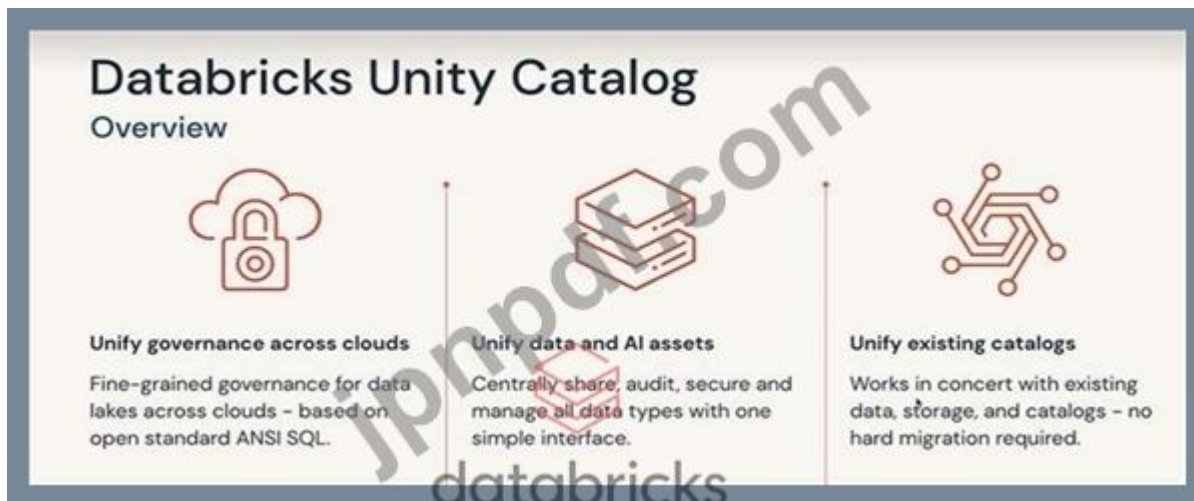
- A. データ共有
- B. Unity カタログ
- C. デルタレイク
- D. 単一の保管場所を使用します。
- E. DELTA LIVE パイプライン

Answer: ([解答を表示する](#))

説明

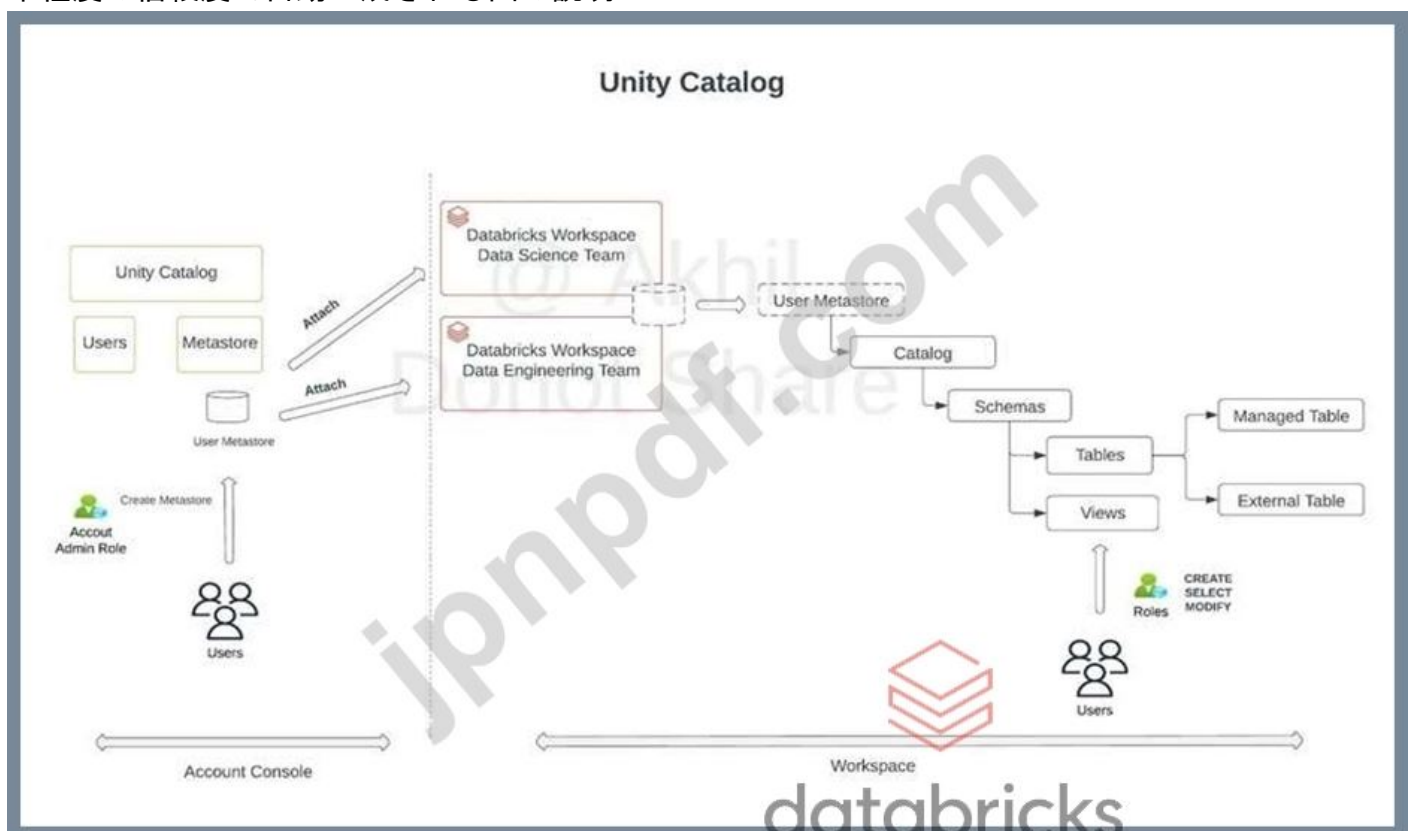
答えは Unity カタログです。

自動生成される図の説明



Unity カタログはアカウント レベルで動作し、メタ ストアを作成し、そのメタ ストアを多くのワークスペースにアタッチする機能があります。Unity カタログの仕組みを理解するには、以下の図を参照してください。メタストアは、次を使用して両方のワークスペースで共有できることがわかります。Unity Catalog では、Unity Catalog より前のオプションは、2 番目の Databricks ワークスペースに手動でマウントされる単一のクラウド オブジェクトストレージを使用することでした。ここで、Unity Catalog がそれを本当に簡素化していることがわかります。

中程度の信頼度で自動生成される図の説明



Udemyの他の人が私の質問や画像をコピーしているため、ウォーターマークが追加されてしまい、ご迷惑をおかけして申し訳ありません。

ダクトの特徴

<https://databricks.com/product/unity-catalog>

最新問題: 2

Delt テーブルはどのように保存されますか？

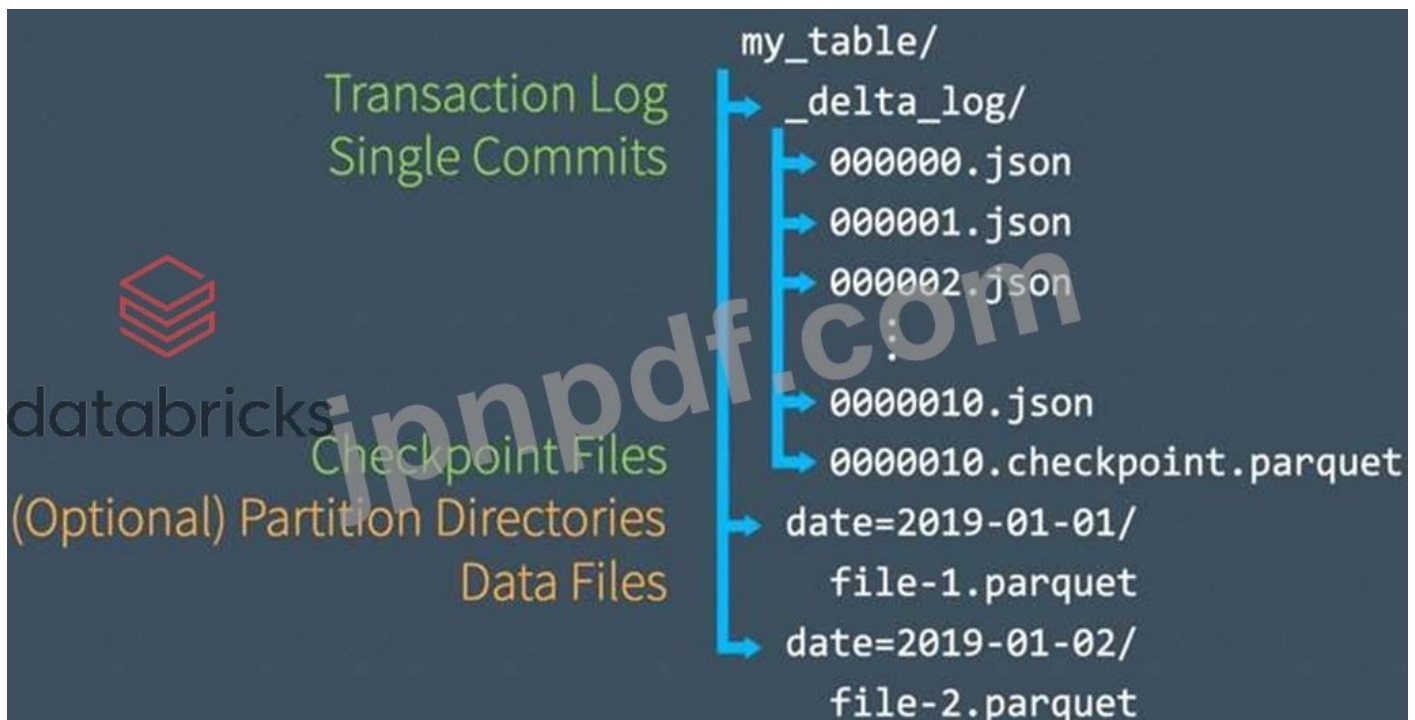
- A. パーケット データ ファイルが保存されるディレクトリ、メタ データおよびトランザクション ログが JSON ファイルとして保存されるサブ ディレクトリ `_delta_log`。
- B. 寄木細工のデータ ファイルが保存されるディレクトリ。すべてのメタ データはメモリに保存されます。
- C. 寄木細工のデータ ファイルがデータ プレーンに保存されるディレクトリ、メタ データ、履歴、ログがコントロール ペインに保存されるサブ ディレクトリ `_delta_log`。
- D. 寄木細工のデータ ファイルが保存されるディレクトリ。すべてのメタデータは寄木細工のファイルに保存されます。
- E. データはデータ プレーンに保存され、メタデータとデルタ ログはコントロール ペインに保存されます。

Answer: A ([メッセージを残す](#))

説明

答えは、寄木細工のデータ ファイルが保存されるディレクトリ、メタデータとトランザクション ログが JSON ファイルとして保存されるサブ ディレクトリ `_delta_log` です。

タイムラインの説明が自動生成される



最新問題: 3

デルタ レイク テーブルに適用できる次のテーブル制約のうち、サポートされているものはどれですか？

- A. 主キー、外部キー、Not Null、チェック制約
- B. 主キー、Null ではない、制約をチェック
- C. デフォルト、NULL ではない、制約をチェック
- D. Null ではありません。制約を確認してください
- E. 一意、NULL ではない、チェック制約

Answer: D (メッセージを残す)

説明

答えは NULL ではありません。制約を確認してください」です。

<https://docs.microsoft.com/en-us/azure/databricks/delta/delta-constraints>

* CREATE TABLE events(id LONG,

* 日付STRING、

* 場所STRING、

* 説明文字列

*) デルタを使用します。

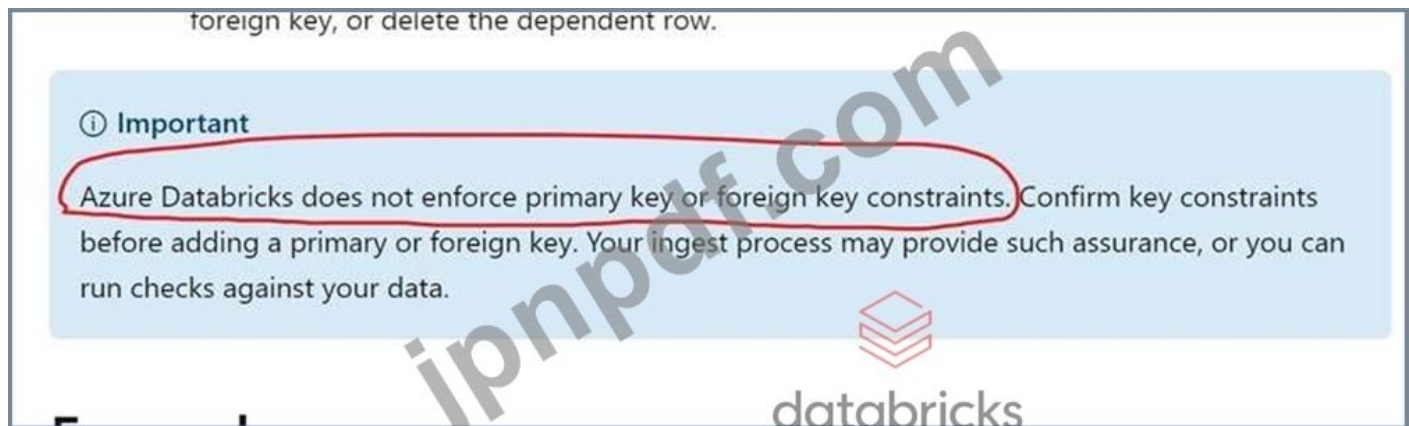
ALTER TABLE イベント CHANGE COLUMN ID SET NOT NULL;

ALTER TABLE events ADD CONSTRAINT dateWithinRange CHECK (date > '1900-01-01'); 注:

DBR 11.1 の Databricks では、Unity カタログが有効な場合の主キーと外部キーのサポートが追加されましたが、これは情報提供のみを目的としており、実際には強制されません。では、強制されていないのになぜこれらを定義するのかと疑問に思われるかもしれません。そのため、特にこれらの情報制約は、テーブル間の関係を知ることによって恩恵を受ける BI ツールを使用している場合に非常に役立ち、レポート/ダッシュボードの作成が容易になります。または、データ モデリング ツールを使用する場合のデータ モデルの理解。

主キーと外部キー

グラフィカル ユーザー インターフェイス、テキスト、アプリケーション、電子メールの説明が自動生成されます



最新問題: 4

あなたは 6 つの部門のジョブとしてセット ノートブックを設定するタスクを課されており、各部門はタスクを並行して実行できます。ノートブックは入力パラメータ部門番号を受け取り、部門ごとにデータを処理します。これをジョブで設定するにはどうすればよいですか??

A. 単一のノートブックをジョブのタスクとして使用し、dbutils.notebook.run を使用して、異なるセルのパラメーターを指定して各ノートブックを実行します。

B. ジョブ内のタスクは入力パラメーターを取得できず、ハードコードされた部門番号を持つ 6 つのノートブックを作成し、ジョブ内に線形依存関係を持つ 6 つのタスクをセットアップすることはできません。

C. タスクはキーと値のペアのパラメーターを受け入れ、すべて並列実行できるためジョブ内に依存関係がなく、各タスクのパラメーターとして部門番号を渡す6つのタスクを作成します。

(正しい)

D. パラメータはジョブレベルでのみ渡すことができます。6つのジョブを作成し、線形ジョブ依存関係を持つ各ジョブに部門番号を渡します。

E. パラメータはジョブレベルでのみ渡すことができます。6つのジョブを作成し、ジョブの依存関係を持たずに各ジョブに部門番号を渡します。

Answer: C ([メッセージを残す](#))

説明

セットアップ方法は次のとおりです

同じノートブックで1つのジョブと6つのタスクを作成し、各タスクに異なるパラメータを割り当てます。グラフィカルユーザーインターフェイス、テキスト、アプリケーション、電子メールの説明が自動的に生成されます。



databricks

Task name * ⓘ

sales

Type *

Notebook

Source * ⓘ

Workspace

Path * ⓘ

/Repos/[REDACTED]/cli-demo/notebooks/job-demo

You are using a Local repository. To automatically track the upstream repository in staging or production Jobs, select Git in the Source field.

Cluster * ⓘ

Process_all_departments_cluster (28.00 GB | 8 Cores | DBR 10.4 LTS | Spark 3.2.1 | S... | ✎ | ▼

Parameters ⓘ

UI | JSON

```
{
  "department": "accounting"
}
```

Depends on

Select task dependencies... | ▼

▼ Advanced options

Cancel

Save task

すべてのタスクは単一のジョブに追加され、単一の共有クラスターまたは個別のクラスターを使用して並列実行できます。

グラフィカル ユーザー インターフェイス、アプリケーション、Teams の説明が自動的に生成される



最新問題: 5

DLT パイプラインを構築する場合、ライブ テーブルを作成するには 2 つのオプションがあります。CREATE STREAMING LIVE TABLE と CREATE LIVE TABLE の主な違いは何ですか？

- A. CREATE STREAMING LIVE テーブルは MULTI HOP アーキテクチャで使用されます
- B. CREATE LIVE TABLE は、ストリーミング データ ソースと増分データを操作するときに使用されます。
- C. CREATE STREAMING LIVE TABLE は、ストリーミング データ ソースと増分データを操作する場合に使用されます。
- D. 違いはありません。どちらも同じです。CREATE STRAMING LIVE はまもなく廃止されます。
- E. CREATE LIVE TABLE は DELTA LIVE TABLES で使用され、CREATE STREAMING LIVE は Structured Streaming アプリケーションでのみ使用できます。

Answer: [\(解答を表示する\)](#)

説明

答えは、CREATE STREAMING LIVE TABLE は、ストリーミング データ ソースと増分データを操作するときに使用されるということです。

最新問題: 6

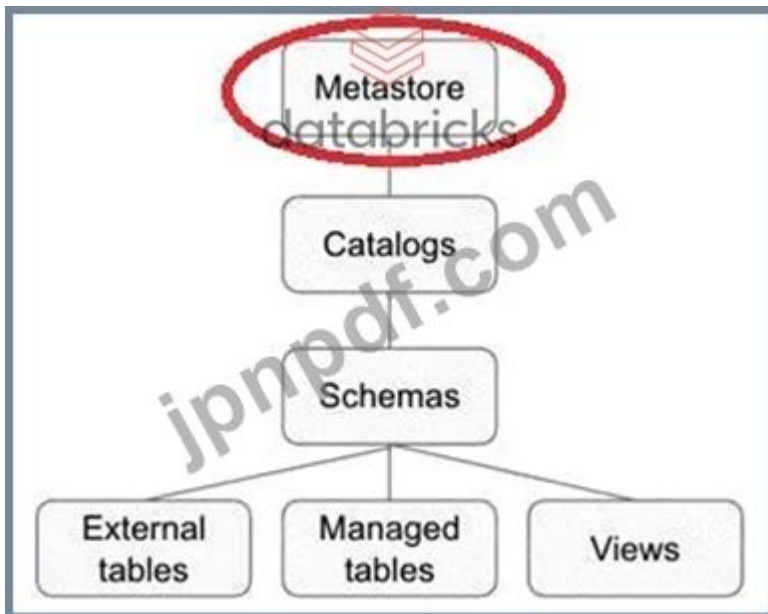
Unity カタログのトップレベルのオブジェクトは何ですか？

- A. カタログ
- B. テーブル
- C. ワークスペース
- D. データベース
- E. メタストア

Answer: [E \(メッセージを残す\)](#)

説明

主要な概念 - Azure Databricks | Microsoft ドキュメント



自動生成される図の説明

最新問題: 7

SQL 構文 `DROP TABLE table_name` を使用してマネージド テーブルを削除すると、テーブルに格納されているメタデータ、履歴、データにどのような影響がありますか？

- A. メタ ストアからテーブルを削除し、ストレージ内のメタデータ、履歴、およびデータを削除します。
- B. メタ ストアからテーブルを削除し、ストレージからデータを削除しますが、メタデータと履歴はストレージに保持します。
- C. メタ ストア、メタ データ、履歴からテーブルを削除しますが、データはストレージに保持します。
- D. テーブルを削除しますが、メタデータ、履歴、データをストレージに保持します。
- E. テーブルと履歴を削除しますが、メタデータとデータはストレージに保持します。

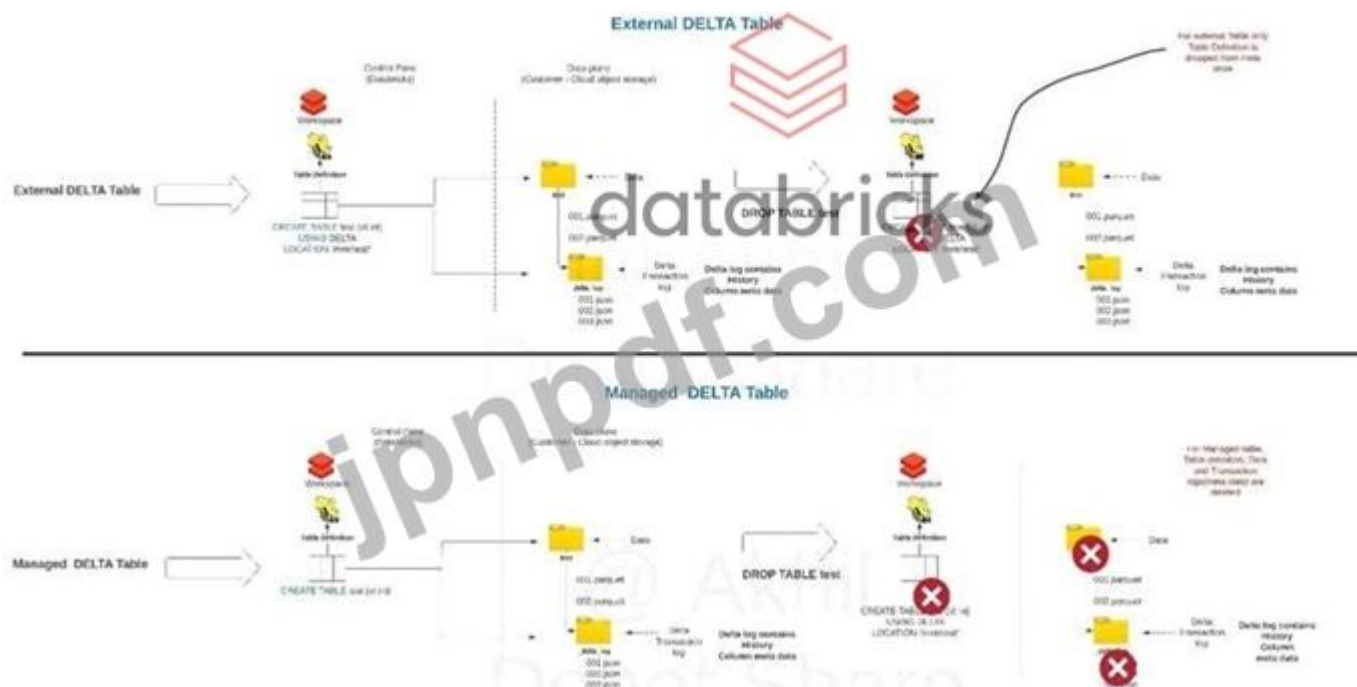
Answer: A ([メッセージを残す](#))

説明

管理対象テーブルの場合、drop コマンドはメタストアとストレージからすべてを削除します。

外部テーブルの削除の違いを理解するには、以下の図を参照してください。

自動生成される図の説明



最新問題: 8

インベントリ プロセスの最後に、ファイルがクラウド オブジェクト ストレージにアップロードされます。データを取り込むプロセスを構築するように求められます。次のいずれかの方法を使用して、データを段階的に取り込むことができます。ファイルのスキーマは変更されることが予想されます。時間外の取り込みプロセスは、これらの変更を自動的に処理できる必要があります。以下は、データをロードし、以下のコードを正常に実行するために空白を埋めるための自動ローダー コマンドです。

- 1.spark.readStream
 - 2..format("クラウドファイル")
 - 3..option("cloudfiles.format","csv")
 - 4..option("_____", 'dbfs:/location/checkpoint/')
 - 5..load(データソース)
 - 6..writeStream
 - 7..option("_____", ' dbfs:/location/checkpoint/')
 - 8..option("mergeSchema", "true")
 - 9..テーブル(テーブル名))
- A. チェックポイントの場所、スキーマの場所
 B. チェックポイントの場所、cloudfiles.schemalocation
 C. スキーマの場所、チェックポイントの場所
 D. cloudfiles.schemalocation、checkpointlocation
 E. cloudfiles.schemalocation、cloudfiles.checkpointlocation

Answer: ([解答を表示する](#))

説明

答えは、cloudfiles.schemalocation、checkpointlocation です。

データを読み取るとき、cloudfiles.schemalocation は、受信データの推論されたスキーマを保存するために使用されます。

障害から回復するためにストリームを書き込む場合、checkpointlocation は、最後に処理されたバイトのオフセットを保存するために使用されます。

最新問題: 9

2021年のデータを含むorders というテーブルで作業しており、2020年のデータを含むorders_archive という2番目のテーブルがあります。2つのテーブルのデータを結合する必要があり、両方のテーブル間に同じ行が存在する可能性があります。両方のテーブルの結果を結合して重複行を削除したいと考えています。これを達成するには、次のSQLステートメントのどれが役立ちますか？

- A. SELECT * FROM 注文 UNION SELECT * FROM 注文アーカイブ
(正しい)
- B. SELECT * FROM 注文 INTERSECT SELECT * FROM 注文アーカイブ
- C. SELECT * FROM 注文 UNION ALL SELECT * FROM 注文アーカイブ
- D. SELECT * FROM 注文アーカイブ MINUS SELECT * FROM 注文
- E. SELECT unique * FROM 注文 JOIN order_archive on order.id = orders_archive.id

Answer: A ([メッセージを残す](#))

説明

答えは SELECT * FROM 注文 UNION SELECT * FROM 注文アーカイブです
UNION と UNION ALL は集合演算子です。

UNION は両方のクエリの出力を結合しますが、重複も削除します。

UNION ALL は、両方のクエリの出力を結合します。

最新問題: 10

同僚がジョブの設定方法を説明していましたが、次のような警告メッセージに気づきました。

汎用クラスターで実行されているジョブは、汎用コンピューティングとみなされます。」という同僚は、警告メッセージが表示される理由がわかりませんでした。この警告メッセージについてどう説明するのが最善ですか？

- A. パフォーマンスの問題のため、汎用クラスターはジョブクラスターには使用できません。
- B. 汎用クラスターは、ジョブクラスターと比較してクラスターの起動に時間がかかります。
- C. 汎用クラスターはジョブクラスターよりも安価です。
- D. 汎用クラスターはジョブクラスターよりも高価です
- E. 汎用クラスターは、ジョブでは表示できない対話型メッセージを提供します。

Answer: D ([メッセージを残す](#))

説明

警告メッセージ:

グラフィカル ユーザー インターフェイス、テキスト、アプリケーション、電子メールの説明が自動生成されます



Task name * ⓘ
test_job

Type *

Notebook

Source * ⓘ

Workspace

Path * ⓘ

/Users, [REDACTED] /Data Analysis

Cluster * ⓘ

SingleNode (DBR 10.4 LTS | Spark 3.2.1 | Scala 2.12)

⚠ Jobs running on all-purpose clusters are considered all-purpose compute. [Learn more](#)

Parameters ⓘ

UI | JSON

Add

Advanced options

Cancel

Create

汎用クラスターの価格はジョブ クラスターよりも高価です AWS の価格(2022 年 8 月 15 日) グラフィカル ユーザー インターフェイス 説明が自動的に生成されます

	Standard	Premium	Enterprise
	One platform for your data analytics and ML workloads	Data analytics and ML at scale across your business	Data analytics and ML for your mission critical workloads
Jobs Light Compute Run data engineering pipelines to build data lakes.	\$0.07 / DBU	\$0.10 / DBU	\$0.13 / DBU
Jobs Compute Jobs Compute Photon Run data engineering pipelines to build data lakes and manage data at scale.	\$0.10 / DBU	\$0.15 / DBU	\$0.20 / DBU
CLASSIC COMPUTE Delta Live Tables Delta Live Tables Photon Easily build high quality streaming or batch ETL pipelines using Python or SQL with the DLT Edition that is best for your workload. Learn more	\$0.20 - \$0.36 / DBU	\$0.20 - \$0.36 / DBU	\$0.20 - \$0.36 / DBU
SQL Compute Run SQL queries for BI reporting, analytics and visualization to get timely insights from data lakes.		\$0.22 / DBU	\$0.22 / DBU
All-Purpose Compute All-Purpose Compute Photon Run interactive data science and machine learning workloads. Also good for data engineering, BI and data analytics.	\$0.40 / DBU	\$0.55 / DBU	\$0.65 / DBU

US East (N. Virginia)

フォームの下部
 フォームの先頭

最新問題: 11

メダリオン建築における銀層と金層の主な違いは何ですか？

- A. シルバーには集約されたデータが含まれる可能性があります
- B. ゴールドには集計データが含まれる場合があります
- C. データ品質チェックはゴールドで適用されます
- D. シルバーはブロンズデータのコピーです。
- E. 神は銀のデータのコピー

Answer: B ([メッセージを残す](#))

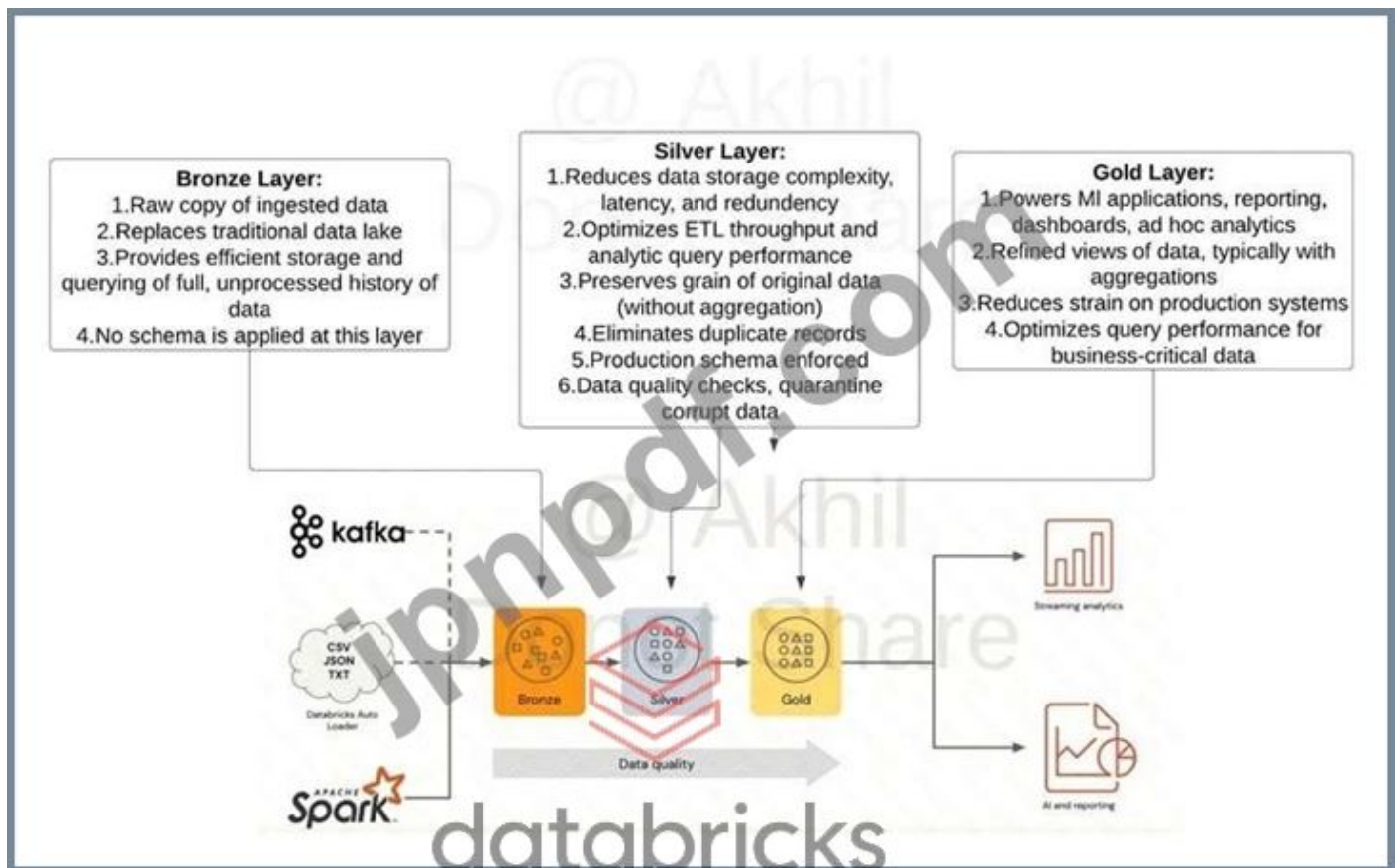
説明

メダリオン アーキテクチャ - Databricks

試験の焦点: 下の画像を確認して、メダリオン建築における各層 (ブロンズ、シルバー、ゴールド) の役割を理解してください。各層とその目的を対象としたさまざまな質問が表示されます。

Udemy 内の一部の人が私のコンテンツをコピーしているため、ウォーターマークを追加する必要がありました。

低信頼度で自動的に生成された家の説明図



最新問題: 12

2つの変数に基づいてデータを処理しようとしています。1つは部門がサプライチェーンであるかどうかを確認するため、もう1つはプロセスフラグがTrueに設定されているかどうかを確認するためです。

- A. 部門 == "サプライチェーン" または プロセス = TRUE の場合:
- B. 部門 == "サプライチェーン" の場合 | プロセス == TRUE:
- C. 部門 = "サプライチェーン" の場合 | プロセス :
- D. 部門 == "サプライチェーン" または プロセスの場合:
- E. 部門 == "サプライチェーン" の場合 | プロセス == TRUE の場合:

Answer: ([解答を表示する](#))

最新問題: 13

受信データを処理するために AUTO LOADER をセットアップするように求められます。このデータは JSON 形式で到着し、クラウドオブジェクトストレージにドロップされます。データはクラウドストレージに到着したらすぐに処理する必要があります。次のステートメントはどれですか?正しい

- A. AUTO LOADER は DELTA Lake にネイティブであり、外部のクラウドオブジェクトストレージをサポートできません
- B. ファイルがクラウドストレージに到着したときに、外部プロセスから AUTO LOADER をトリガーする必要があります
- C. AUTO LOADER を構造化ストリームプロセスに変換する必要があります

- D. AUTO LOADER は、DELTA Lake に保存されている場合のみ連続データを処理できます。
- E. AUTO LOADER はファイル通知方法をサポートできるため、データが到着したときに処理できます。

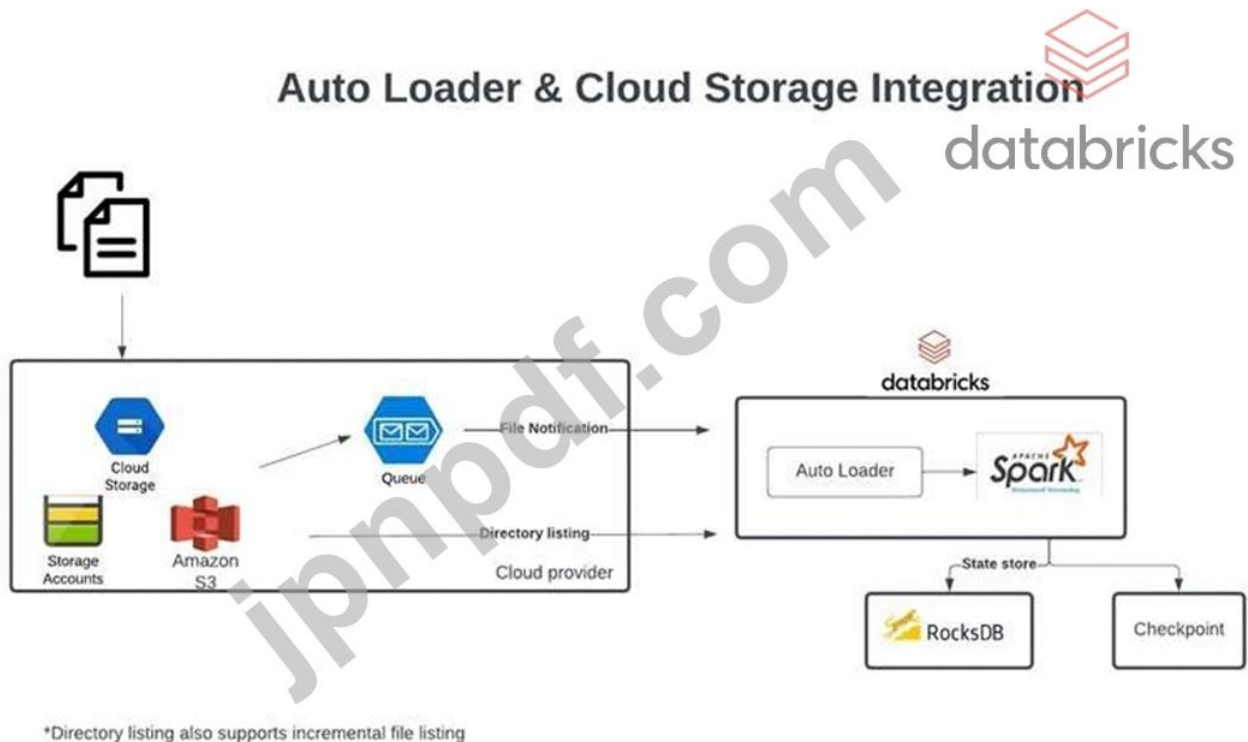
Answer: E ([メッセージを残す](#))

説明

Auto Loader は、クラウド オブジェクト ストレージから新しいファイルを取り込むときに 2 つのモードをサポートします。ディレクトリ リスト: Auto Loader は、入力ディレクトリをリストすることで新しいファイルを識別し、ディレクトリ ポーリング アプローチを使用します。

ファイル通知: Auto Loader は、入力ディレクトリからのファイル イベントをサブスクライブする通知サービスとキュー サービスを自動的にセットアップできます。

自動生成される図の説明



ファイル通知はより効率的であり、データがクラウド オブジェクト ストレージに到着するとリアルタイムでデータを処理するために使用できます。

ファイル通知モードとディレクトリ一覧モードの選択 | AWS 上のデータブリック

最新問題: 14

データサイエンス チームは、テーブルに平均価格と呼ばれる列が欠落しているとリクエストしました。これは、販売数量と売上金額を使用して計算できます。追加の列を使用してデータを再ロードできるのは、次の SQL ステートメントのどれですか。

- A. 1.INSERT 上書き売上
- 2.SELECT *, 売上高/単位販売額からの平均価格
- B. 1.テーブルの作成または置換 sales
- 2.AS SELECT *, 販売金額/単位販売額の平均価格として販売
- C. MERGE INTO sales USING (SELECT *, salesAmt/unitsSold as avgPrice FROM sales)

- D. 売上高を SELECT * として上書き、売上高/単位売上高からの平均価格として販売
- E. COPY INTO SALES AS SELECT *、売上金額/単位売上高からの平均価格として販売

Answer: ([解答を表示する](#))

説明

1. テーブル販売の作成または置換

2. AS SELECT *、販売金額/単位販売額の平均価格として販売

INSERT OVERWRITE と CREATE OR REPLACE TABLE(CRAS) の主な違いは、CRAS がテーブルのスキーマを変更できること、つまり、新しい列を追加したり、既存の列のデータ型を変更したりできることです。デフォルトでは、INSERT OVERWRITE はデータのみを上書きします。次の場合に限り、INSERT OVERWRITE を使用してスキーマを上書きすることもできます。このオプションが有効になっておらず、スキーマの不一致がある場合、コマンドは失敗する場合、spark.databricks.delta.schema.autoMerge.enabled は true に設定されます。

最新問題: 15

データ エンジニアリング チームは、ジョブの 1 つが通常は 15 分で終了しますが、リモート データベースの読み取り中にネットワーク パケット ドロップが原因でランダムに停止することに気付きました。ジョブの安定性を向上させるために使用できる手順は次のうちどれですか？

- A. Databrick REST API を使用して、長時間実行されているジョブを監視し、kill コマンドを発行します。
- B. 長時間実行されているジョブを監視および強制終了するには、ジョブの実行、アクティブな実行の UI セクションを使用します。
- C. タスクが 15 分を超えて実行される場合にジョブを強制終了するタイムアウトを含めるようにタスクを変更します。
- D. Spark UI で Spark ジョブのタイムアウト設定を使用します。
- E. ジョブ クラスター UI のクラスター タイムアウト設定を使用します。

Answer: C ([メッセージを残す](#))

説明

答えは、タスクが 15 分を超えて実行される場合にジョブを強制終了するタイムアウトを含めるようにタスクを変更することです。

<https://docs.microsoft.com/en-us/azure/databricks/data-engineering/jobs/jobs#timeout>

最新問題: 16

2 人の若手データ エンジニアが、1 つのデータ パイプライン ノートブックの別々の部分を作成しています。彼らは取り組んでいます

別々の Git ブランチを使用して、同じノートブック上のプログラムを同時にペアリングできるようにします。シニアデータエンジニア

Databricks の経験から、このタイプのコラボレーションにはより良い代替手段があることが示唆されています。

上級データ エンジニアの主張を裏付けるものは次のうちどれですか？

- A. Databricks ノートブックはコメントと通知コメントをサポートします

- B. Databricks ノートブックは、同じノートブックでの複数の言語の使用をサポートします。
- C. Databricks ノートブックは自動変更追跡とバージョン管理をサポートします。
- D. Databricks ノートブックは、単一のノートブックでのリアルタイム共同編集をサポートしません。
- E. Databricks ノートブックは、インタラクティブなデータ視覚化の作成をサポートします。

Answer: ([解答を表示する](#))

有効な **Databricks-Certified-Professional-Data-Engineer** 問題集は GoShiken.com が提供された合格しやすい Databricks-Certified-Professional-Data-Engineer 試験問題集！

GoShiken.com が最新の **Databricks-Certified-Professional-Data-Engineer** 試験問題集を提供しています。GoShiken.com Databricks-Certified-Professional-Data-Engineer 試験問題は最新で、解答が正確でございます。最新の GoShiken.com Databricks-Certified-Professional-Data-Engineer 問題集をゲットする人はこちら: <https://www.goshiken.com/Databricks/Databricks-Certified-Professional-Data-Engineer-mondaishu.html> (20430%OFF問題集溶と正解付きで 30%w 特別割引コード: **Freepdfdumps**)

最新問題: 17

CSV、JSON、TEXT、BINARY 形式を使用して外部テーブルを定義する場合、外部テーブルに対するクエリはパフォーマンス上の理由からデータと場所をキャッシュするため、特定の Spark セッション内では、到着した可能性のある新しいファイルは最初のセッション後に使用できなくなります。クエリ。この制限にどう対処すればよいでしょうか？

- A. UNCACHE TABLE table_name
- B. キャッシュ テーブル テーブル名
- C. REFRESH TABLE テーブル名
- D. ブロードキャストテーブル テーブル名
- E. CLEAR CACH テーブル名

Answer: ([解答を表示する](#))

説明

答えは REFRESH TABLE table_name です。

REFRESH TABLE table_name は、Spark に外部ファイルの可用性と変更を強制的に更新させます。

Spark が外部テーブルにクエリを実行すると、そのテーブルに関連付けられたファイルがキャッシュされるため、テーブルが再度クエリされた場合にキャッシュされたファイルを使用できるため、クラウドオブジェクトストレージから再度取得する必要がなくなります。ただし、ここでの欠点は、新しいファイルが利用可能かどうかは、Refresh コマンドが実行されるまで Spark にはわかりません。

最新問題: 18

DELTA テーブルを削除するには、次のコマンドのうちどれを使用できますか？

- A. DROP DELTA テーブル名
- B. DROP TABLE table_name FORMAT DELTA
- C. DROP TABLE テーブル名
- D. DROP テーブル名

Answer: C (メッセージを残す)

最新問題: 19

次のデータ ワークロードのうち、Bronze テーブルをソースとして利用するものはどれですか？

- A. クリーンアップされたデータを集約して標準の要約統計を作成するジョブ
- B. 集計されたデータをクエリして、重要な洞察をダッシュボードに公開するジョブ
- C. ストリーミング ソースから生データを Lakehouse に取り込むジョブ
- D. タイムスタンプを人間が判読できる形式に解析してデータを強化するジョブ
- E. 機械学習アプリケーションの機能セットを開発するジョブ

Answer: D (メッセージを残す)

最新問題: 20

営業チームはデータ エンジニアリング チームに、全店舗の販売実績を示すダッシュボードの開発を依頼しましたが、営業チームはダッシュボードを使用したいが、個々の店舗の場所を選択したいと考えています。データ エンジニアリング チームにアプローチするのは次のどれですか。を使用して、この機能をダッシュボードに組み込むことができます。

- A. ユーザーが任意の場所を選択できるようにするクエリ パラメータを使用します。
- B. 現在、ダッシュボードはパラメータをサポートしていません
- C. Databricks REST API を使用して、場所ごとにダッシュボードを作成します
- D. SQL UDF 関数を使用して、場所に基づいてデータをフィルタリングします。
- E. 動的ビューを使用して、場所に基づいてデータをフィルターします。

Answer: A (メッセージを残す)

説明

答えは

Databricks はダッシュボードで多くの種類のパラメータをサポートしており、ストアの場所の一意のリストを含むクエリに基づいてドロップダウン リストを作成できます。

以下はパラメータを受け取る単純なクエリです。

```
SELECT * FROM sales WHERE フィールド IN ( {{ 複数選択パラメータ }} )
```

また

```
SELECT * FROM sales WHERE フィールド = {{ 単一選択パラメータ }}
```

クエリパラメータのタイプ

*文章

*番号

*ドロップダウンリスト

*クエリベースのドロップダウンリスト

*日時

最新問題: 21

Lakehouse と Datawarehouse のどちらを選択する際に間違っているのはどれですか？

- A. Lakehouse は、機械学習用に最適化された特別なインデックスとキャッシュを持つことができます。
- B. Lakehouse は、BI ワークロードに対して高い信頼性を備えた低クエリ レイテンシを提供できず、バッチ ワークロードにのみ適しています。
- C. Lakehouse には、Py-thon/R/SQL を含む (ただしこれらに限定されない) さまざまな API を通じてアクセスできます。
- D. 従来のデータ ウェアハウスにはストレージとコンピューティングが結合されています。
- E. Lakehouse は Parquet などの標準データ形式を使用します。

Answer: ([解答を表示する](#))

説明

答えは、Lakehouse は BI ワークロードに対して高い信頼性を備えた低クエリ レイテンシーを提供できず、バッチ ワークロードにのみ適しているということです。

Lakehouse は、ストレージやキャッシングなどのコンピューティングの最適化を活用することで、従来のウェアハウスを置き換えて、高い信頼性を備えた低クエリ レイテンシーでサービスを提供できます。

Spark キャッシュと デルタ キャッシュの比較に焦点を当てます。

<https://docs.databricks.com/delta/optimizations/delta-cache.html>

レイクハウスとは何ですか? - Databricks ブログ

グラフィカル ユーザー インターフェイス、テキスト、アプリケーションの説明が自動生成される

A lakehouse has the following key features:

- **Transaction support:** In an enterprise lakehouse many data pipelines will often be reading and writing data concurrently. Support for ACID transactions ensures consistency as multiple parties concurrently read or write data, typically using SQL.
- **Schema enforcement and governance:** The Lakehouse should have a way to support schema enforcement and evolution, supporting DW schema architectures such as star/snowflake-schemas. The system should be able to **reason about data integrity**, and it should have robust governance and auditing mechanisms.
- **BI support:** Lakehouses enable using BI tools directly on the source data. This reduces staleness and improves recency, reduces latency, and lowers the cost of having to operationalize two copies of the data in both a data lake and a warehouse.
- **Storage is decoupled from compute:** In practice this means storage and compute use separate clusters, thus these systems are able to scale to many more concurrent users and larger data sizes. Some modern data warehouses also have this property.
- **Openness:** The storage formats they use are open and standardized, such as Parquet, and they provide an API so a variety of tools and engines, including machine learning and Python/R libraries, can efficiently access the data **directly**.
- **Support for diverse data types ranging from unstructured to structured data:** The lakehouse can be used to store, refine, analyze, and access data types needed for many new data applications, including images, video, audio, semi-structured data, and text.
- **Support for diverse workloads:** including data science, machine learning, and SQL and analytics. Multiple tools might be needed to support all these workloads but they all rely on the same data repository.
- **End-to-end streaming:** Real-time reports are the norm in many enterprises. Support for streaming eliminates the need for separate systems dedicated to serving real-time data applications.

フォームの下部

フォームの先頭

最新問題: 22

外部テーブルが作成されるのは次のどれですか？

- A. DELTA LOCATION EXTERNAL を使用した CREATE TABLE トランザクション (id int, desc string)
- B. CREATE TABLE トランザクション (id int, desc string)
- C. CREATE EXTERNAL TABLE トランザクション (id int, desc string)
- D. CREATE TABLE トランザクション (id int, desc string) TYPE EXTERNAL
- E. CREATE TABLE トランザクション (id int, desc string) LOCATION '/mnt/delta/transactions'

Answer: ([解答を表示する](#))

説明

答えは、CREATE TABLE トランザクション (id int, desc string) USING DELTA LOCATION です。

'/mnt/デルタ/トランザクション'

Location を使用してテーブルが作成されると、そのテーブルは外部テーブルとみなされます。現在の構文は次のとおりです。

構文

CREATE TABLE テーブル名 (列 列データ型...) USING 形式 LOCATION "dbfs:/"

最新問題: 23

次の SQL コマンドのうち、行が存在するかどうかを確認する条件に基づいて行を挿入、更新、または削除するために使用できるものはどれですか？

- A. テーブル名にマージします
- B. テーブル名にコピーします
- C. UPDATE テーブル名
- D. INSERT INTO OVERWRITE テーブル名
- E. EXISTS テーブル名を挿入

Answer: A ([メッセージを残す](#))

説明

ここにレビュー用の追加ドキュメントがあります。

<https://docs.databricks.com/spark/latest/spark-sql/language-manual/delta-merge-into.html>

1. MERGE INTO target_table_name [target_alias]
2. source_table_reference [source_alias] の使用
3. マージ条件をオンにする
4. [WHEN MATCHED [AND 条件] THEN matched_action] [...]
5. [WHEN NOT MATCHED [AND 条件] THEN not_matched_action] [...]
- 6.
7. 一致したアクション
8. { 削除 |
9. セットの更新 * |
10. UPDATE SET { 列 1 = 値 1 } [, ...] }
- 11.
12. 一致しないアクション
13. { 挿入 * |
14. INSERT (column1 [, ...]) VALUES (value1 [, ...])

最新問題: 24

以下のテーブル temp_data には、シカゴ市の 1 日の 4 時間ごとの気温を記録する JSON データが含まれる raw という列が 1 つあります。毎日の午後 12 時に記録された最高気温を計算するように求められます。JSON データを解析し、必要な配列関数を使用して最大温度を計算します。

テーブル: temp_date

列: 生

データ型: 文字列

```

{
  "chicago": [
    {"date": "01-01-2021",
     "temp": [25, 28, 45, 56, 39, 25]},
    {"date": "01-02-2021",
     "temp": [25, 28, 49, 54, 38, 25]},
    {"date": "01-03-2021",
     "temp": [25, 28, 49, 58, 38, 25]}
  ]
}

```

[12AM, 4AM, 8AM, 12PM, 4PM, 8PM]



予想される出力: 58

- A. 1.temp_data から max(raw.chicago.temp[3]) を選択します
- B. 1.temp_data から array_max(raw.chicago[*].temp[3]) を選択します
- C. 1.select array_max(from_json(raw['chicago'].temp[3],'array<int>')) temp_data
- D. 1.select array_max(from_json(raw:chicago[*].temp[3],'array<int>')) temp_data
- E. 1.select max(from_json(raw:chicago[3].temp[3],'array<int>')) from temp_data

Answer: ([解答を表示する](#))

説明

注: これは難しい質問です。これと似た簡単な質問が見つかる可能性が高くなりますが、試験の準備ができていればいるほど、試験に合格するのは容易になります。

他の例を探すには、以下のリンクを使用してください。これは間違いなく役立ちます。

<https://docs.databricks.com/optimizations/semi-structured.html>

ここでは解決策を段階的に示します

自動生成されるテキスト説明

```
select raw:chicago[*] from temp_data
```



Select Chicago and all elements inside Chicago

```
[
  {
    "date": "01-01-2021",
    "temp": [25, 28, 45, 56, 39, 25]
  },
  {
    "date": "01-02-2021",
    "temp": [25, 28, 49, 54, 38, 25]
  },
  {
    "date": "01-03-2021",
    "temp": [25, 28, 49, 58, 38, 25]
  }
]
```

```
select raw:chicago[*].temp[3] from temp_data
```



select 4th element in the temp array, index starts from 0,1,2,3

```
'[56,54,58]'
```

```
select from_json(raw:chicago[*].temp[3], 'array<int>') from temp_data
```



```
[56,54,58]
```

convert string to array using from_json

```
select array_max(from_json(raw:chicago[*].temp[3], 'array<int>')) from temp_data
```



```
58
```

use array_max to calculate the max temp

databricks

他の例を探すには、以下のリンクを使用してください。これは間違いなく役立ちます。

<https://docs.databricks.com/optimizations/semi-structured.html>

このソリューションを試したい場合は、以下の DDL を使用してください。

1. テーブル temp_data を作成または置換します

2. として選択 '{

3. シカゴ」:[

4. {"日付": "2021-01-01",

5. 温度」:[25,28,45,56,39,25]

6.},

7. {"日付": "2021年2月1日",

8. 温度」:[25,28,49,54,38,25]

9.},

10. {"日付": "2021-03-01",

11. 温度」:[25,28,49,58,38,25]

12. }]

13. }

14. 生のまま

15.

16. temp_data から array_max(from_json(raw:chicago[*].temp[3],'array<int>')) を選択します

17.

最新問題: 25

レイクハウスについて正しいのはどれですか？

A. Lakehouse は機械学習ワークロードのみをサポートし、データ ウェアハウスは BI ワークロードをサポートします。

B. Lakehouse はエンドツーエンドのストリーミング ワークロードのみをサポートし、データ ウェアハウスはバッチ ワークロードをサポートします。

C. Lakehouse は ACID をサポートしていません

D. Lakehouse ではストレージとコンピューティングが結合されています

E. Lakehouse はスキーマの適用と進化をサポートします

Answer: E (メッセージを残す)

説明

答えは、Lakehouse はスキーマの強制と進化をサポートしているということです。

Delta Lake を使用する Lakehouse は、読み取り時にのみスキーマを強制できる従来のビッグデータ システムとは対照的に、書き込み時にスキーマを強制できるだけでなく、進化を制御する機能により、時間の経過とともに進化するスキーマもサポートします。

たとえば、以下はデータフレーム ライター API で、強制と進化の 3 つのモードをサポートしています。デフォルト: 強制のみ、変更は許可されず、スキーマのドリフト/進化は失敗します。

マージ: 柔軟性があり、強制と進化をサポートします。

※新しいコラムが追加されました

* ネストされた列を進化させます

* Byte、Short、Integer、Bigint などの進化するデータ型をサポート 有効にする方法:

* DF.write.format("delta").option("mergeSchema", "true").saveAsTable("table_name")

* また

*spark.databricks.delta.schema.autoMerge = True ## Spark セッション

上書き: 強制なし

* 列の削除

* 文字列を整数に変更します

* 列の名前を変更します

有効にする方法:

* DF.write.format("delta").option("overwriteSchema", "True").saveAsTable("table_name")

レイクハウスとは何ですか? - Databricks ブログ グラフィカル ユーザー インターフェイス、テキスト、アプリケーションの説明が自動的に生成される

A lakehouse has the following key features:

- **Transaction support:** In an enterprise lakehouse many data pipelines will often be reading and writing data concurrently. Support for ACID transactions ensures consistency as multiple parties concurrently read or write data, typically using SQL.
- **Schema enforcement and governance:** The Lakehouse should have a way to support schema enforcement and evolution, supporting DW schema architectures such as star/snowflake-schemas. The system should be able to reason about data integrity, and it should have robust governance and auditing mechanisms.
- **BI support:** Lakehouses enable using BI tools directly on the source data. This reduces staleness and improves recency, reduces latency, and lowers the cost of having to operationalize two copies of the data in both a data lake and a warehouse.
- **Storage is decoupled from compute:** In practice this means storage and compute use separate clusters, thus these systems are able to scale to many more concurrent users and larger data sizes. Some modern data warehouses also have this property.
- **Openness:** The storage formats they use are open and standardized, such as Parquet, and they provide an API so a variety of tools and engines, including machine learning and Python/R libraries, can efficiently access the data directly.
- **Support for diverse data types ranging from unstructured to structured data:** The lakehouse can be used to store, refine, analyze, and access data types needed for many new data applications, including images, video, audio, semi-structured data, and text.
- **Support for diverse workloads:** including data science, machine learning, and SQL and analytics. Multiple tools might be needed to support all these workloads but they all rely on the same data repository.
- **End-to-end streaming:** Real-time reports are the norm in many enterprises. Support for streaming eliminates the need for separate systems dedicated to serving real-time data applications.

最新問題: 26

データ エンジニアは、2 つのテーブルからデータを取得してリレーショナル オブジェクトを作成したいと考えています。リレーショナル オブジェクトは、他のデータ エンジニアが他のセッションで使用することもできます。ストレージ コストを節約するために、データ エンジニアは次のことを望んでいます。

物理データのコピーや保存は避けてください。

データ エンジニアは次のリレーショナル オブジェクトのうちどれを作成する必要がありますか？

- A. ビュー
- B. デルタテーブル
- C. Spark SQL テーブル
- D. データベース
- E. 一時的なビュー

Answer: A ([メッセージを残す](#))

最新問題: 27

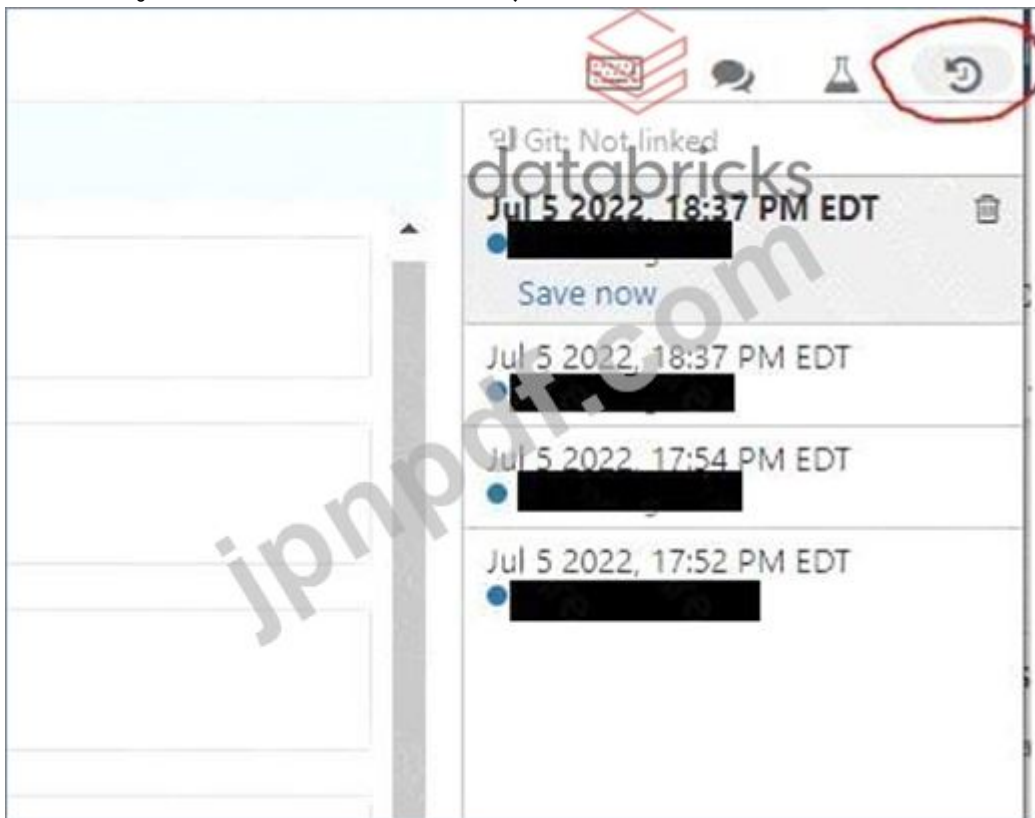
あなたは現在 2 番目のチームと協力しており、両方のチームが同じノートブックを変更しようとしています。2 番目のメンバーがノートブックを個人フォルダにコピーして、コラボレーションノートブックを編集して置き換えていることに気がきました。どのノートブック機能を使用することをお勧めしますか? プロセスの共同作業が容易になります。

- A. Databricks ノートブックをローカルマシンにコピーし、ソース管理をローカルでセットアップしてノートブックをバージョン管理する必要があります。
- B. Databricks ノートブックは、自動変更追跡とバージョン管理をサポートしています。
- C. Databricks ノートブックは、単一ノートブックでのリアルタイム共同編集をサポートします。
- D. Databricks ノートブックは dbc アーカイブファイルにエクスポートし、データレイクに保存できます。
- E. Databricks ノートブックは HTML としてエクスポートし、後でインポートできます。

Answer: C ([メッセージを残す](#))

説明

答えは、Databricks ノートブックは単一のノートブックでのリアルタイム共同編集をサポートしています。すべての変更は保存され、複数のユーザーがノートブックを変更できます。



最新問題: 28

次のステートメントのうち、ノートブック ウィジェットを正常に読み取り、Python ノートブックセル内の SQL ステートメントに Python 変数を渡すことができるのはどれですか?

A. `1.order_date = dbutils.widgets.get("widget_order_date")`

2.

3.spark.sql(f"SELECT * FROM sales WHERE orderDate = '{order_date }'")

B. 1.order_date = dbutils.widgets.get("widget_order_date")

2.

3.spark.sql(f"SELECT * FROM sales WHERE orderDate = '\${order_date }' ")

C. 1.order_date = dbutils.widgets.get("widget_order_date")

2.

3.spark.sql(f"SELECT * FROM sales WHERE orderDate = 'order_date' ")

D. 1.order_date = dbutils.widgets.get("widget_order_date")

2.

3.spark.sql(f"SELECT * FROM sales WHERE orderDate = '{order_date}' ")

(正しい)

E. 1.order_date = dbutils.widgets.get("widget_order_date")

2.

3.spark.sql("SELECT * FROM sales WHERE orderDate = order_date")

Answer: ([解答を表示する](#))

最新問題: 29

Databricks SQL クエリの実行が遅いことに気づきました。クエリの実行が遅い理由を調べて、パフォーマンスを改善する手順を特定するように求められます。問題を確認すると、すべてのクエリが並列で実行され、SQL エンドポイントを使用していることがわかりました。(SQL Warehouse) を単一クラスターで使用します。クエリのパフォーマンスや応答時間を改善するには、次のどの手順を実行できますか？

*Databricks は最近、SQL エンドポイントの名前を SQL ウェアハウスに変更したことに注意してください。

A. SQL エンドポイント (SQL ウェアハウス) のサーバーレス機能を有効にすることができます。

B. SQL エンドポイント (SQL ウェアハウス) のスケーリング範囲の上限を増やすことができます。

C. SQL エンドポイント (SQL ウェアハウス) のウェアハウス サイズを 2X-Smal から 4XLarge まで増やすことができます。

D. SQL エンドポイント (SQL ウェアハウス) の自動停止機能をオンにできます。

E. SQL エンドポイント (SQL ウェアハウス) のサーバーレス機能を有効にし、スポット インスタンス ポリシーを「信頼性の最適化」に変更できます。

Answer: ([解答を表示する](#))

説明

答えは、最大スケーリング範囲を増やすと、より多くのクラスターが追加され、クエリがキュー内で待機する代わりに、利用可能なクラスターを使用して実行を開始できるようになり、SQL エンドポイントのスケーリング範囲の最大境界を増やすことができます。詳細については、以下を参照してください。

質問は、SQL エンドポイント (SQL ウェアハウス) をスケーリングする方法を知る能力をテストすることを目的としており、手がかりとなる言葉を探るか、クエリが順次実行されているか、同時

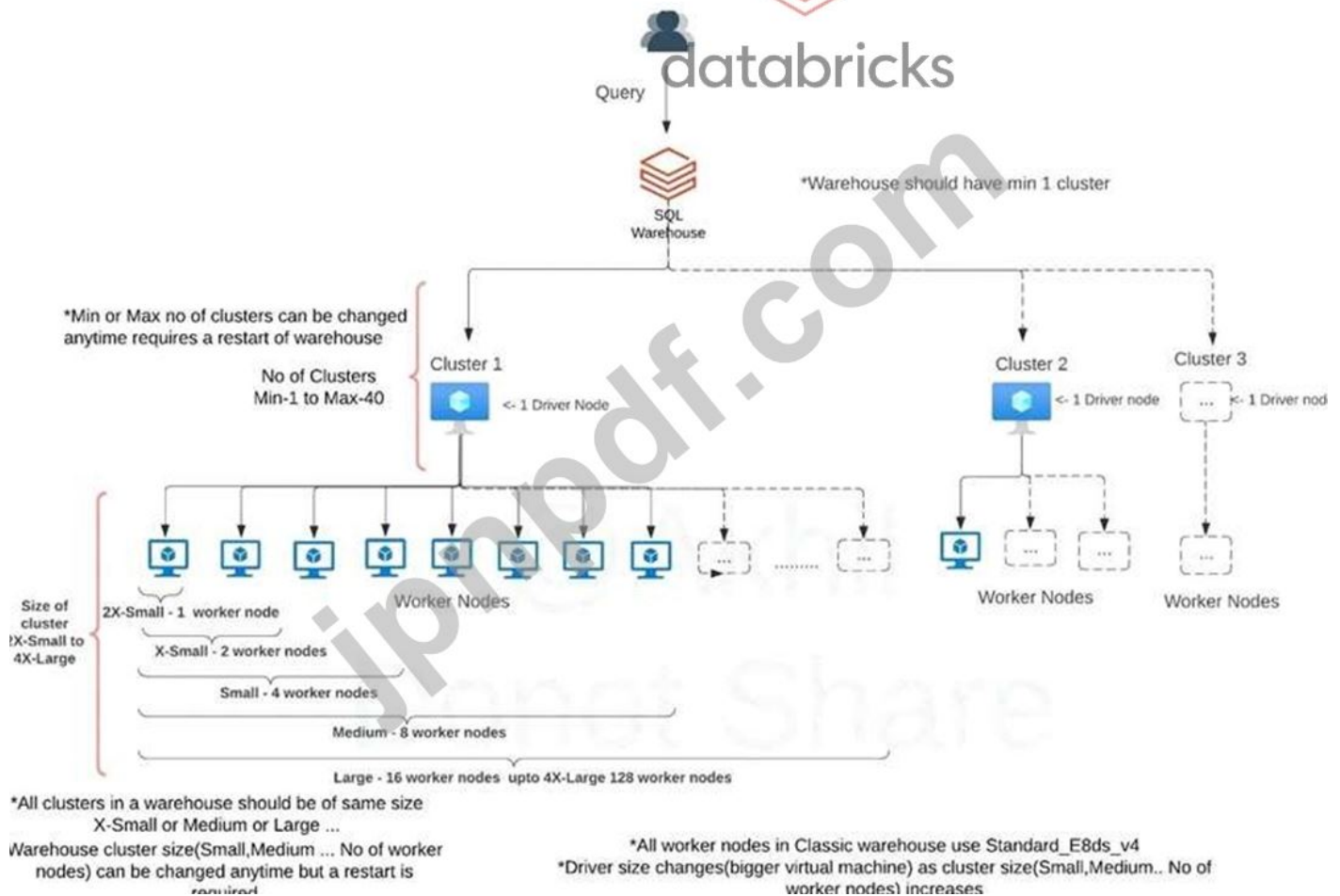
に実行されているかを理解する必要があります。クエリが順次実行されている場合はスケールアップします (クラスターのサイズが 2 倍から 4 倍まで)。クエリが同時に実行されている場合、またはより多くのユーザーが実行している場合は、スケールアウトします (クラスターを追加します)。

SQL エンドポイント (SQL ウェアハウス) の概要: (理解するには、以下のすべての点と以下の図を読んでください)

1. SQL ウェアハウスには少なくとも 1 つのクラスターが必要です
2. クラスターは 1 つのドライバー ノードと 1 つまたは複数のワーカー ノードで構成されます。
3. クラスター内のワーカー ノードの数はクラスターのサイズによって決まります (2X-Small -> 1 ワーカー、X-Small -> 2 ワーカー.... 最大 4X-Large -> 128 ワーカー)。これはクラスターのサイズと呼ばれます。拡大する
4. クラスター サイズ (2X-Smal.. ~ ...4XLarge) に関係なく、単一クラスターは、ユーザーが 3X-Large クラスター サイズとクラスターを持つウェアハウスに 20 個のクエリを一度に送信した場合、一度に 10 個のクエリのみを実行できます。スケールアップ (最小 1、max1) 10 個のクエリの実行が開始されますが、残りの 10 個のクエリはこれら 10 個のクエリが完了するまでキューで待機します。
5. ウェアハウス クラスター サイズを増やすと、クエリのパフォーマンスが向上します。たとえば、ウェアハウス サイズが 2X-Small でクエリが 1 分間実行された場合、ウェアハウス サイズを X-Small に変更すると、クエリが 30 秒で実行される可能性があります。これは、2X-Small には 1 つのワーカー ノードがあり、X-Small には 2 つのワーカー ノードがあるため、クエリがより多くのタスクを持ち、より高速に実行されるためです (注: これは理想的なケースの例であり、クエリのパフォーマンスのスケラビリティは多くの要因に依存します。必ずしも直線的であるとは限りません)
6. ウェアハウスには複数のクラスターを含めることができます。これはスケールアウトと呼ばれます。ウェアハウスがクラスター スケールアップ (Min1、Max 2) の X-Small クラスター サイズで構成されている場合、Databricks はクエリがキュー内で待機していることを検出すると追加のクラスターを起動します。ウェアハウスが 2 つのクラスター (Min1、Max 2) を実行するように構成されている場合最大 2)、ユーザーが 20 個のクエリを送信すると、10 個のクエリが実行を開始して残りをキューに保持し、Databricks が自動的に 2 番目のクラスターを起動し、キューで待機している 10 個のクエリを 2 番目のクラスターにリダイレクトし始めます。
7. 単一のクエリは複数のクラスターにまたがることはありません。クエリがクラスターに送信されると、スケールアップ可能なクラスターの数に関係なく、クエリの実行が終了するまでそのクラスターに残ります。

上記の概念を理解するには、以下の図を確認してください。

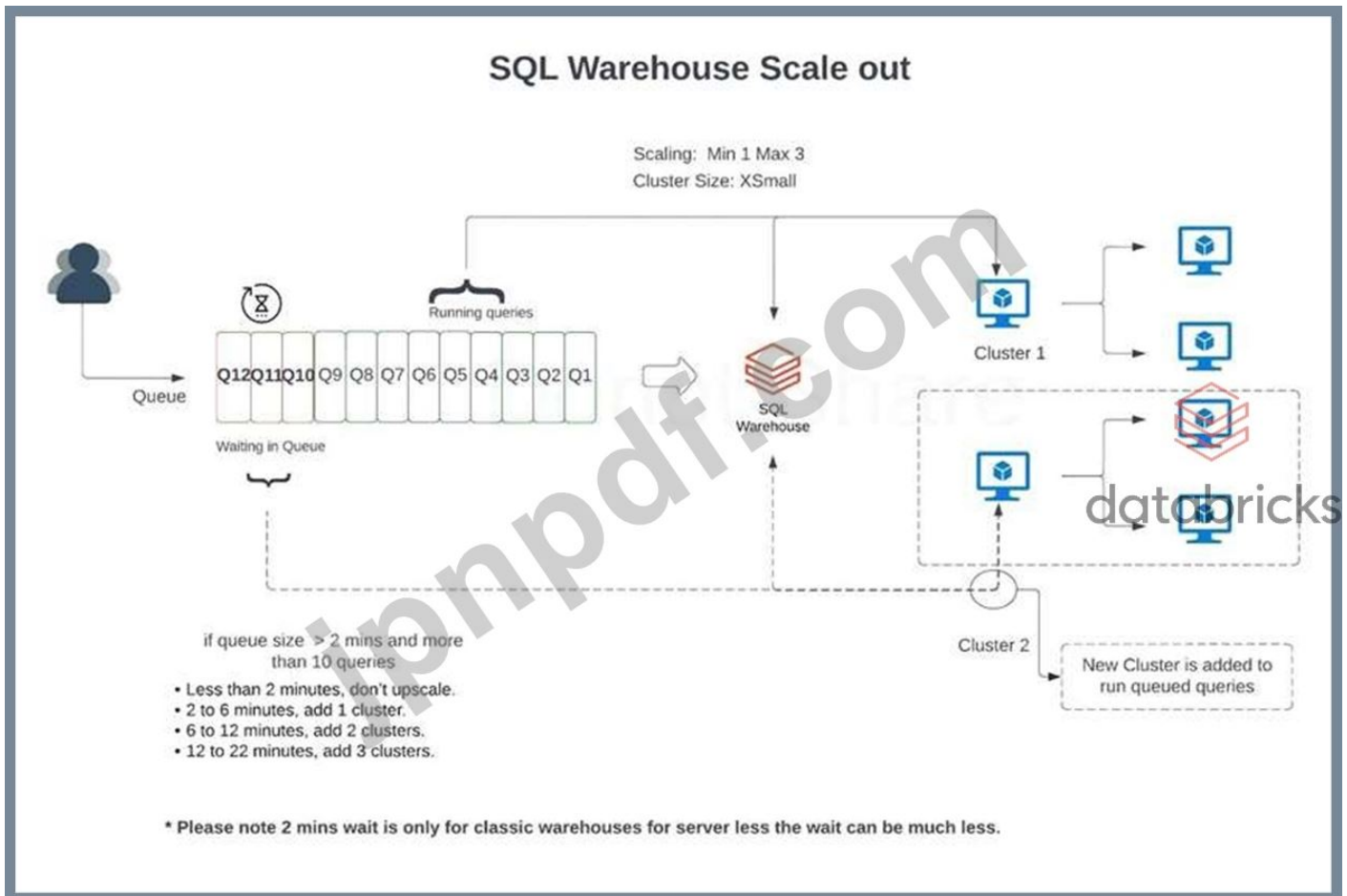
SQL Warehouse Overview



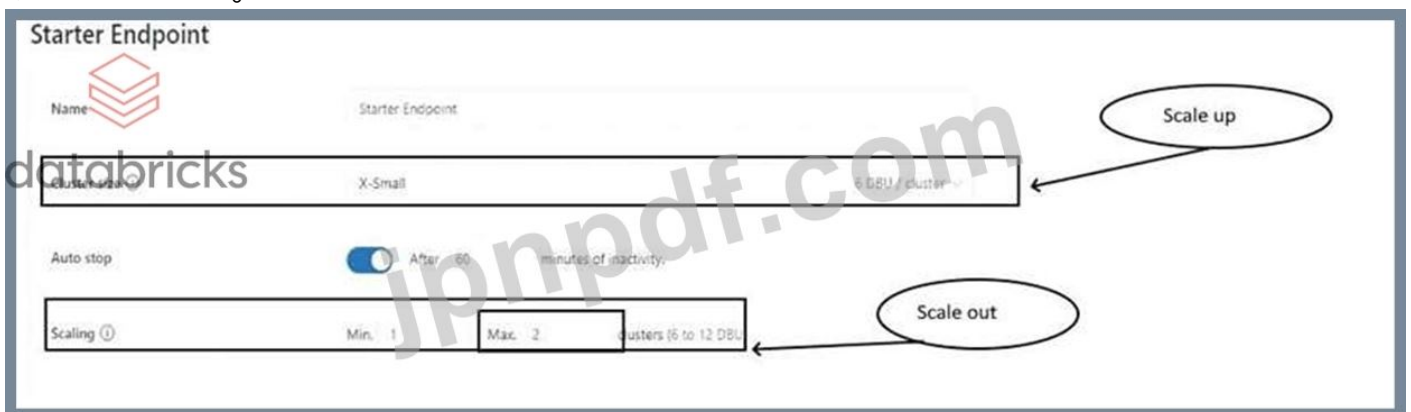
SQL エンドポイント (SQL Warehouse) は、水平方向 (スケールアウト) と垂直方向 (スケールアップ) にスケールします。いつ何を使用するかを理解する必要があります。

スケールアウト -> SQL エンドポイントにクラスターを追加するには、クラスターの最大数を変更します。スループットを向上させたい場合は、できるだけ多くのクエリを実行できるようにクラスターを追加すると、パフォーマンスが向上します。

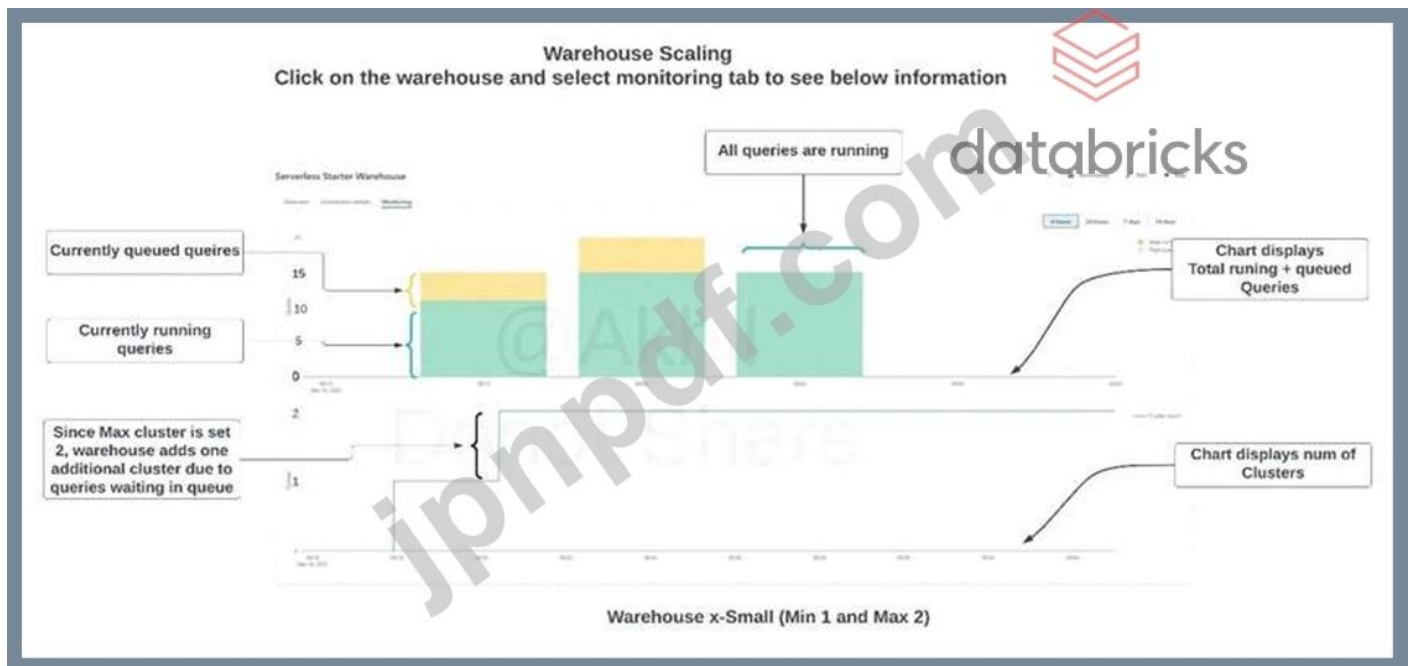
Databricks SQL は、クエリがキュー状態にあることを検出するとすぐに自動的にスケールアップします。この例では、スケールアップは最小 1 と最大 3 に設定されています。これは、クエリが待機していることを検出した場合、ウェアハウスが 3 つのクラスターを追加できることを意味します。



ウェアハウスの作成中、またはウェアハウスのサイズを変更できるようになった後 (2X-Small...から...4XLarge) を使用して、クエリのパフォーマンスを向上させ、SQL エンドポイント (SQL ウェアハウス) スケールアウトでクラスターを追加するためのスケーリング範囲を最大化します。既存のウェアハウスを変更する場合は、変更を有効にするためにウェアハウスの再起動が必要になる場合があります。



必要なクラスターの数はどのようにしてわかりますか (最大クラスターサイズを設定する方法)? 既存のウェアハウスをクリックして監視タブを選択すると、ウェアハウスの使用状況情報 (下記を参照) が表示されます。ウェアハウスの使用状況に関する重要な情報を提供する 2 つのグラフが表示されます。クエリがキューに入れられている場合は、ウェアハウスは追加のクラスターから恩恵を受けることができます。コストとパフォーマンスのバランスのとれた決定を行えるよう、クラスターの追加に関連する追加の DBU コストを確認してください。



最新問題: 30

データ エンジニアがジョブ クラスターの代わりにジョブ クラスターを使用するシナリオを説明しているものは次のどれですか。

万能クラスター？

- A. データ エンジニアは本番エラーを手動で調査する必要があります
- B. Databricks SQL クエリを上向きレポート用にスケジュールする必要があります
- C. 自動ワークフローは 30 分ごとに実行する必要があります
- D. コンピューティング コストを最小限に抑えながら、アドホック分析レポートを開発する必要があります
- E. データ チームは機械学習モデルの開発に協力する必要があります

Answer: C ([メッセージを残す](#))

最新問題: 31

以下の SQL コマンドのうち、グローバル一時ビューを作成するものはどれですか？

- A. 1.一時ビューの作成または置換 view_name
2. AS SELECT * FROM table_name
- B. 1. ローカル一時ビュー view_name を作成または置換します。
2. AS SELECT * FROM table_name
- C. 1. グローバル一時ビュー view_name を作成または置換します
2. AS SELECT * FROM table_name
(正しい)
- D. 1.ビューの作成または置換 view_name
2. AS SELECT * FROM table_name
- E. 1. ローカル ビュー view_name の作成または置換
2. AS SELECT * FROM table_name

Answer: (解答を表示する)

説明

1. グローバル一時ビュー view_name の作成または置換
2. AS SELECT * FROM table_name

作成できる一時ビューには、ローカルとグローバルの2種類があります。

*セッションスコープの一時ビューは、Spark セッションでのみ使用できるため、同じクラスター内の別のノートブックはアクセスできません。ノートブックが切り離され、再接続された場合、ローカルの一時ビューが失われます。

*グローバル一時ビューはクラスター内のすべてのノートブックで使用できますが、クラスターが再起動するとグローバル一時ビューは失われます。

有効な **Databricks-Certified-Professional-Data-Engineer** 問題集は GoShiken.com が提供された合格しやすい Databricks-Certified-Professional-Data-Engineer 試験問題集！

GoShiken.com が最新の **Databricks-Certified-Professional-Data-Engineer** 試験問題集を提供しています。GoShiken.com Databricks-Certified-Professional-Data-Engineer 試験問題は最新で、解答が正確でございます。最新の GoShiken.com Databricks-Certified-Professional-Data-Engineer 問題集をゲットする人はこちら: <https://www.goshiken.com/Databricks/Databricks-Certified-Professional-Data-Engineer-mondaishu.html> (20430%OFF問題集溶と正解付き

で 30%w 特別割引コード: **Freepdfdumps**)

最新問題: 32

AUTO LOADER と COPY INTO の主な違いは何ですか？

- A. COPY INTO はスキーマの進化をサポートします。
- B. AUTO LOADER はスキーマの進化をサポートします。
- C. COPY INTO は、増分ロードの実行時のファイル通知をサポートします。
- D. AUTO LOADER は Apache Kafka からのデータの読み取りをサポートします
- E. AUTO LOADER インクリメンタルロード実行時のファイル通知をサポートします。

Answer: E (メッセージを残す)

説明

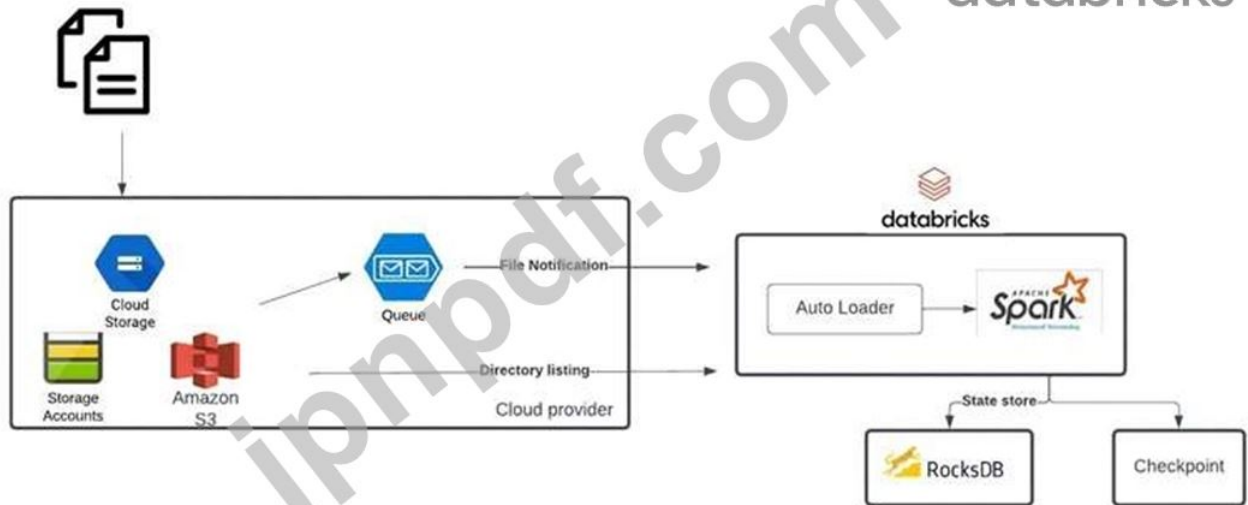
オートローダーはディレクトリリストとファイル通知の両方をサポートしますが、COPY INTO はディレクトリリストのみをサポートします。

自動ローダー ファイル通知は、Azure BLOB ストレージや S3 などのクラウド オブジェクトストレージの入力ディレクトリからファイル イベントをサブスクライブする通知サービスとキューサービスを自動的にセットアップします。ファイル通知モードは、大規模な入力ディレクトリまたは大量のファイルの場合、よりパフォーマンスが高く、スケーラブルです。

Auto Loader & Cloud Storage Integration



databricks



*Directory listing also supports incremental file listing

オートローダーとクラウドストレージの統合

Auto Loader は、データを段階的に取り込むいくつかの方法をサポートしています

1. ディレクトリリスト - ディレクトリをリストし、RocksDBで状態を維持し、増分ファイルリストをサポートします。
2. ファイル通知 - ディレクトリリストとは異なり、トリガーとキューを使用してファイル通知を保存し、後でファイルを取得するために使用できます。ファイル通知は、1日あたり数百万のファイルまでスケールアップできます。

[オプション]

オートローダーとCOPY INTO?

オートローダー

Auto Loader は、新しいデータ ファイルがクラウドストレージに到着すると、追加のセットアップを行わずに段階的かつ効率的に処理します。Auto Loader は、cloudFiles と呼ばれる新しい構造化ストリーミングソースを提供します。クラウドファイルストレージ上の入力ディレクトリパスを指定すると、cloudFiles ソースは新しいファイルが到着すると自動的に処理し、オプションでそのディレクトリ内の既存のファイルも処理します。

COPY INTO の代わりに Auto Loader を使用するのはどのような場合ですか？

*数百万以上のファイルが含まれるファイルの場所からデータをロードしたい。オートローダーは、COPY INTO SQL コマンドよりも効率的にファイルを検出でき、ファイル処理を複数のバッチに分割できます。

*以前にアップロードしたファイルのサブセットをロードする予定はありません。Auto Loader を使用すると、ファイルのサブセットを再処理することがより困難になる場合があります。ただし、COPY INTO SQL コマンドを使用すると、オートローダー ストリームの同時実行中にファイルのサブセットを再ロードできます。

自動ローダー ファイル通知は、Azure BLOB ストレージや S3 などのクラウド オブジェクト ストレージの入力ディレクトリからファイル イベントをサブスクライブする通知サービスとキューサービスを自動的にセットアップします。ファイル通知モードは、大規模な入力ディレクトリまたは大量のファイルの場合、よりパフォーマンスが高く、スケーラブルです。

COPY INTO と Auto Loader をいつ使用するかに関する追加の注意事項を次に示します。

COPY INTO を使用する場合

<https://docs.databricks.com/delta/delta-ingest.html#copy-into-sql-command> 自動ローダーを使用する場合

<https://docs.databricks.com/delta/delta-ingest.html#auto-loader>

最新問題: 33

あなたは、同僚が以前のバージョンを保存するために _bkp を使用してノートブックを手動でコピーしていることに気づきました。代わりに次の機能のうちどれをお勧めしますか。

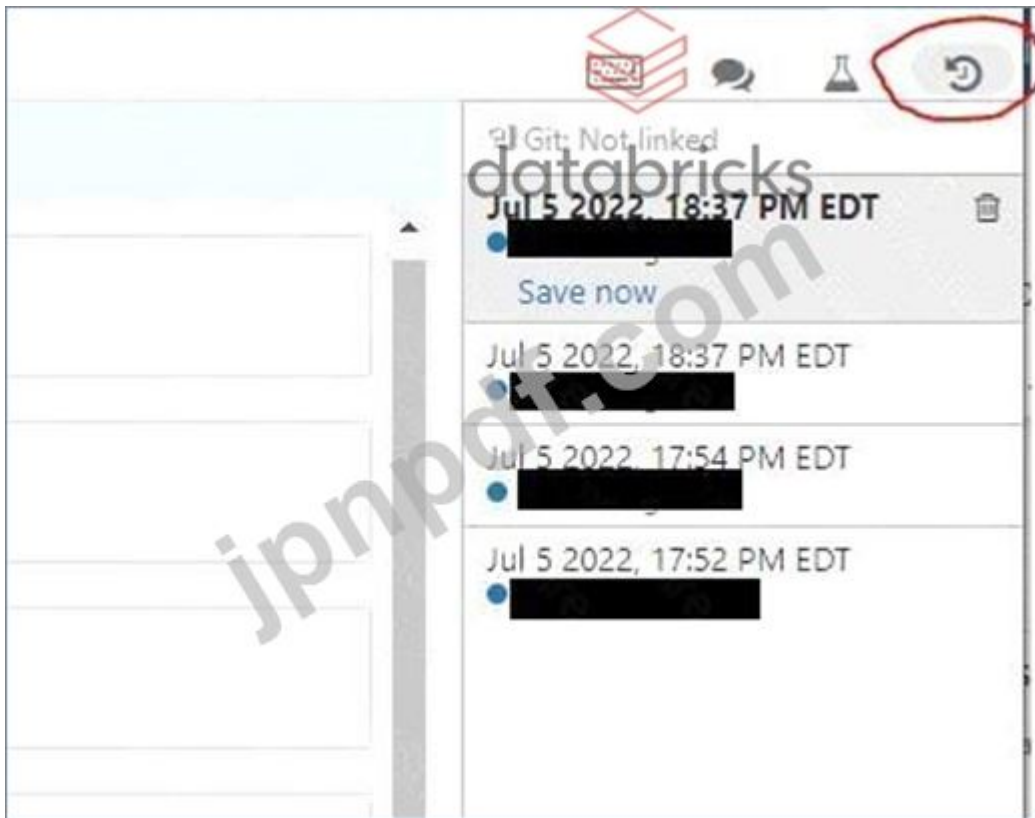
- A. Databricks ノートブックは変更の追跡とバージョン管理をサポートしています
- B. Databricks ノートブックをローカル マシンにコピーし、ソース管理をローカルでセットアップしてノートブックをバージョン管理する必要があります。
- C. Databricks ノートブックは dbc アーカイブ ファイルにエクスポートし、データ レイクに保存できます。
- D. Databricks ノートブックは HTML としてエクスポートし、後でインポートできます。

Answer: A ([メッセージを残す](#))

説明

答えは、Databricks ノートブックは自動変更追跡とバージョン管理をサポートしているということです。

右側でノートブックを編集しているときは、バージョン履歴を確認してすべての変更を表示し、加えたすべての変更がキャプチャされて保存されます。



最新問題: 34

Databricks レイクハウス オブジェクトではないものは次のうちどれですか？

- A. テーブル
- B. ビュー
- C. データベース/スキーマ
- D. カタログ
- E. 機能
- F. ストアド プロシージャ

Answer: F ([メッセージを残す](#))

説明

答えは、ストアド プロシージャです。

Databricks Lakehouse はストアド プロシージャをサポートしていません。

最新問題: 35

ユーザーのアクセスに基づいてデルタ テーブルの行と列に対するきめ細かいアクセス制御を実装するために使用できる手法は次のどれですか？

- A. Unity カタログを使用して行と列へのアクセスを許可します
- B. 行および列のアクセス制御リスト
- C. ダイナミックビュー機能を使用する
- D. データアクセス制御リスト
- E. Unity カタログを使用した動的アクセス制御リスト

Answer: C ([メッセージを残す](#))

説明

答えは、ダイナミック ビュー機能を使用することです。

これは、ユーザーがマネージャー グループに属していることに基づいて行へのアクセスを制限する例です。ユーザーがマネージャー グループのメンバーではない場合、以下のビューでは、合計金額が 1000000 以下の行のみが表示されます。動的ビュー機能により、行をフィルタリングする

1.CREATE VIEW sales_redacted AS

2.user_id、国、製品、合計を選択します

3.FROM販売_生

4.WHERE CASE WHEN is_member('managers') THEN TRUE ELSE 合計 <= 1000000 END;

ユーザーのアクセスに基づいて列データを非表示にするダイナミックビュー機能、

1.CREATE VIEW sales_redacted AS

2.ユーザーIDを選択し、

3. CASE WHEN is_member('auditors') THEN 電子メール ELSE 'REDACTED' END AS email、

4.国、

5.製品、

6.合計

7.FROM販売_生

詳細については以下をご覧ください

<https://docs.microsoft.com/en-us/azure/databricks/security/access-control/table-acls/object-privileges#dynamic-v>

最新問題: 36

あなたは、チーム メンバーが汎用クラスターを使用してノートブックを開発し始め、同じ汎用クラスターを使用して 30 分ごとに実行できるジョブを設定して、ダッシュボードで使用される基礎となるテーブルを更新できることに気づきました。。このアプローチの全体的なコストを削減するには何を推奨しますか？

A. クラスターのサイズを削減します。

B. ノード数を減らし、自動スケールを有効にします。

C. 30 分後の自動終了を有効にする

D. ジョブのスケジュール時に汎用クラスターをジョブクラスターに変更します。

E. クラスタモードを汎用モードからシングルモードに変更します。

Answer: (解答を表示する)

説明

開発中は汎用クラスターを使用しても問題ありませんが、ノートブックを操作する必要がないときは常に、特にスケジュールされたジョブの場合は、ジョブクラスターを使用した方がコストが安くなります。汎用クラスターを使用すると、ジョブクラスターの 2 倍のコストがかかります。

注: 汎用クラスターとジョブクラスター間で同じクラスター タイプとサイズに対してクラウドプロバイダーに支払うコンピューティング コストは同じであり、唯一の違いは DBU コストです。

クラスターの総コスト = VM コンピューティングの総コスト (Azure または AWS または GCP) + DBU あたりのコスト DBU あたりのコストは、汎用クラスターとジョブ クラスターの間で異なります。これは、ジョブ クラスターと汎用クラスター間の AWS からの最近のコスト見積もりです。、ジョブの場合はその値を計算します

DBU あたり 0.15 セント v 汎用の場合、DBU あたり 0.55 セント

グラフィカルユーザーインターフェース 自動生成される説明

	Standard	Premium	Enterprise
aws	One platform for your data analytics and ML workloads	Data analytics and ML at scale across your business	Data analytics and ML for your mission critical workloads
Jobs Light Compute Run data engineering pipelines to build data lakes.	\$0.07 / DBU	\$0.10 / DBU	\$0.13 / DBU
Jobs Compute Jobs Compute Photon Run data engineering pipelines to build data lakes and manage data at scale.	\$0.10 / DBU	\$0.15 / DBU	\$0.20 / DBU
Delta Live Tables Delta Live Tables Photon Easily build high quality streaming or batch ETL pipelines using Python or SQL with the DLT Edition that is best for your workload. Learn more	\$0.20 - \$0.36 / DBU	\$0.20 - \$0.36 / DBU	\$0.20 - \$0.36 / DBU
SQL Compute Run SQL queries for BI reporting, analytics and visualization to get timely insights from data lakes.	-	\$0.22 / DBU	\$0.22 / DBU
All-Purpose Compute All-Purpose Compute Photon Run interactive data science and machine learning workloads. Also good for data engineering, BI and data analytics.	\$0.40 / DBU	\$0.55 / DBU	\$0.65 / DBU

クラスターの DBU コストを確認するにはどうすればよいですか？

既存のクラスターをクリックするか、クラスターの詳細を見ると、右上隅にこれが表示されます。グラフィカル ユーザー インターフェイス、テキスト、アプリケーション、電子メール 説明が自動的に生成されます。



最新問題: 37

データサイエンスチームのメンバーは、単一クラスターを使用してデータ分析を実行していません。クラスターサイズは複数のユーザーを処理できるように選択され、自動スケーリングが有効になっていますが、チームはクエリの実行が依然として遅いことに気づきました。これに対して提案される修正は何ですか？

- A. 各チームメンバーが独自のクラスターを持つように複数のクラスターをセットアップします。
- B. オートスケール機能を無効にする
- C. 標準モードの代わりに高同時実行モードを使用します。
- D. ドライバーノードのサイズを大きくします。

Answer: C ([メッセージを残す](#))

説明

答えは、標準モードの代わりに高同時実行モードを使用することです。

<https://docs.databricks.com/clusters/cluster-config-best-practices.html#cluster-mode> 高同時実行クラスターは、リソースを共有したり、アドホックジョブを実行する必要があるユーザーのグループに最適です。

Databricks では、高同時実行クラスターの自動スケーリングを有効にすることをお勧めします。

最新問題: 38

データエンジニアは、/customer/customer360 の場所に customer360 というデータベースを作成する必要があります。の

データエンジニアは、同僚の 1 人がすでにデータベースを作成しているかどうかわかりません。データエンジニアがこのタスクを完了するには、次のコマンドのうちどれを実行する必要がありますか？

- A. データベース customer360 LOCATION '/customer/customer360' を作成します。
- B. データベース customer360 デルタロケーション '/customer/customer360' を作成します。

- C. customer360 の場所 '/customer/customer360'; が存在しない場合はデータベースを作成します。
- D. customer360 デルタ ロケーション '/customer/customer360'; が存在しない場合はデータベースを作成します。
- E. データベースが存在しない場合は作成します customer360;

Answer: C (メッセージを残す)

最新問題: 39

あなたは現在、データ パイプラインの構築に取り組むよう求められています。現在、多くのデータ品質の問題を抱えているデータ ソースを使用していることに気づきました。データ品質を監視し、データ取り込みプロセスの一環としてそれを強制する必要があります。この問題に対処するために使用できるツールは次のうちどれですか？

- A. オートローダー
- B. デルタライブテーブル
- C. ジョブとタスク
- D. UNITY カタログとデータ ガバナンス
- E. マルチホップによる構造化ストリーミング

Answer: B (メッセージを残す)

説明

答えは、DELTA LIVE TABLES です。

デルタ ライブ テーブルの期待値は、不良データの特定と隔離に使用でき、すべてのデータ品質メトリクスはイベント ログに保存され、後で分析および監視するために使用できます。

DELTA LIVE テーブルの期待

以下に 3 種類の期待値を示します。これら 3 つの違いに注意してください。

無効なレコードを保持します。

期待に反するレコードを保持したい場合は、expect 演算子を使用します。期待に違反するレコードは、有効なレコードとともにターゲット データセットに追加されます。

パイソン

```
1.@dlt.expect("有効なタイムスタンプ", "col("タイムスタンプ") > '2012-01-01')
```

SQL

```
1.CONSTRAINT valid_timestamp EXPECT (タイムスタンプ > '2012-01-01')
```

無効なレコードを削除します。

無効なレコードの処理を防ぐには、expect 演算子またはdrop 演算子を使用します。期待に違反するレコードは、ターゲット データセットから削除されます。

パイソン

```
1.@dlt.expect_or_drop("valid_current_page", "current_page_id は NULL ではなく、current_page_title も NULL ではありません") SQL
```

```
1.CONSTRAINT valid_current_page EXPECT (current_page_id IS NOT NULL および current_page_title IS NOT NULL) ON VIOLATION DROP ROW 無効なレコードで失敗します。
```

無効なレコードが受け入れられない場合は、expect または failed 演算子を使用して、レコードが検証に失敗したときに実行をただちに停止します。操作がテーブル更新の場合、システムはトランザクションをアトミックにロールバックします。

パイソン

```
1.@dlt.expect_or_fail("有効なカウント", "カウント > 0")
```

SQL

```
1.CONSTRAINT valid_count EXPECT (count > 0) ON VIOLATION FAIL UPDATE
```

最新問題: 40

Databricks リポジトリに関して正しいのは次のどれですか？

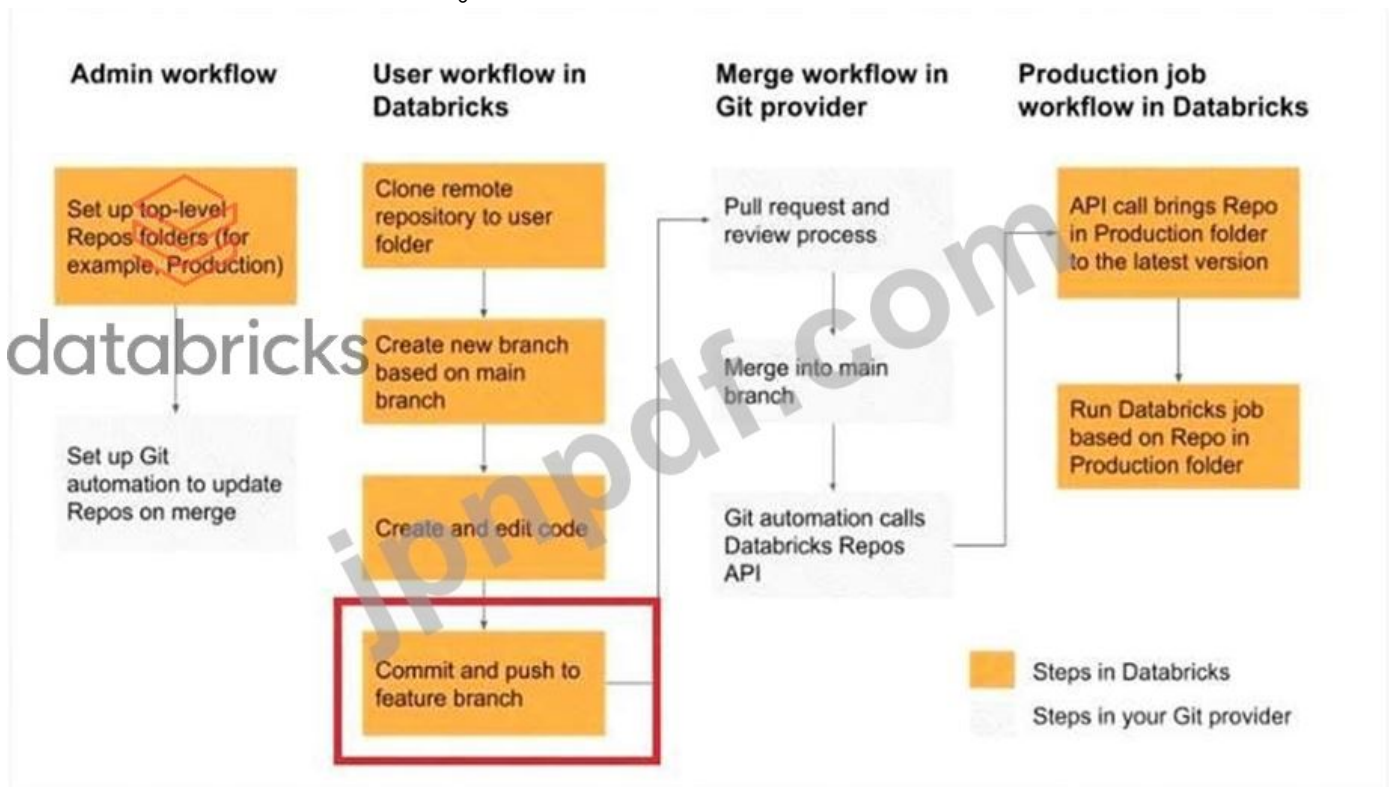
- A. Databricks リポジトリの所有者であれば、プル リクエストを承認できます。
- B. ワークスペースには git 統合のインスタンスを 1 つだけ含めることができます
- C. Databricks リポジトリとノートブックのバージョン管理は同じ機能です
- D. Databricks リポジトリに新しいブランチを作成することはできません
- E. Databricks リポジトリを使用すると、コードの変更をコメントしてコミットし、リモート ブランチにプッシュできます。

Answer: ([解答を表示する](#))

説明

CI/CD ワークフローを構築するときに Databricks Repos と Git プロバイダーが果たす役割を理解するには、以下の図を参照してください。

黄色で強調表示されているすべての手順は Databricks Repo で実行できます。灰色で強調表示されているすべての手順は、Github や Azure Devops などの Git プロバイダーで実行できます。図の説明は自動的に生成されます。



最新問題: 41

現在、ダウストリーム プロセスの消費に関するレポート テーブルを作成するノートブックを作成しています。このプロセスは 1 時間ごとのスケジュールで実行する必要があります。このジョブのセットアップにはどのような種類のクラスターを使用しますか？

- A. これは単一のジョブであり、1 時間ごとに実行する必要があるため、汎用クラスターを使用できません。
- B. ジョブ クラスターがこの目的に最適です。
- C. Azure VM を使用して Python でデルタ テーブルの読み取りと書き込みを行う
- D. デルタ ライブ テーブル パイプラインを使用して連続モードで実行します。

Answer: B (メッセージを残す)

説明

答えは、ジョブ クラスターがこの目的に最適であるということです。

特にスケジュールされたジョブの場合、実行中にノートブックを操作する必要がないため、ジョブ クラスターは合理的です。汎用クラスターを使用すると、ジョブ クラスターの 2 倍のコストがかかる可能性があります。

ご参考までに、

ジョブの完了時に新しいクラスターを作成するオプションを指定してジョブ スケジューラを実行すると、クラスターが終了します。ジョブ クラスターを再起動することはできません。

最新問題: 42

Databricks 製品アーキテクチャ内の次の場所のうち、ノートブックとジョブをホストする場所はどこですか？

- A. データプレーン
- B. コントロールプレーン
- C. Databricks ファイルシステム
- D. JDBC データソース
- E. Databricks Web アプリケーション

Answer: B (メッセージを残す)

説明

答えは「コントロール パネル」です。

Databricks は、ほとんどのサービスをコントロール プレーンとデータ プレーンから運用します。SQL エンドポイントや DLT コンピューティングなどのサーバーレス機能は、コントロール ペインで共有コンピューティングを使用することに注意してください。

コントロール プレーン: Databricks Cloud アカウントに保存

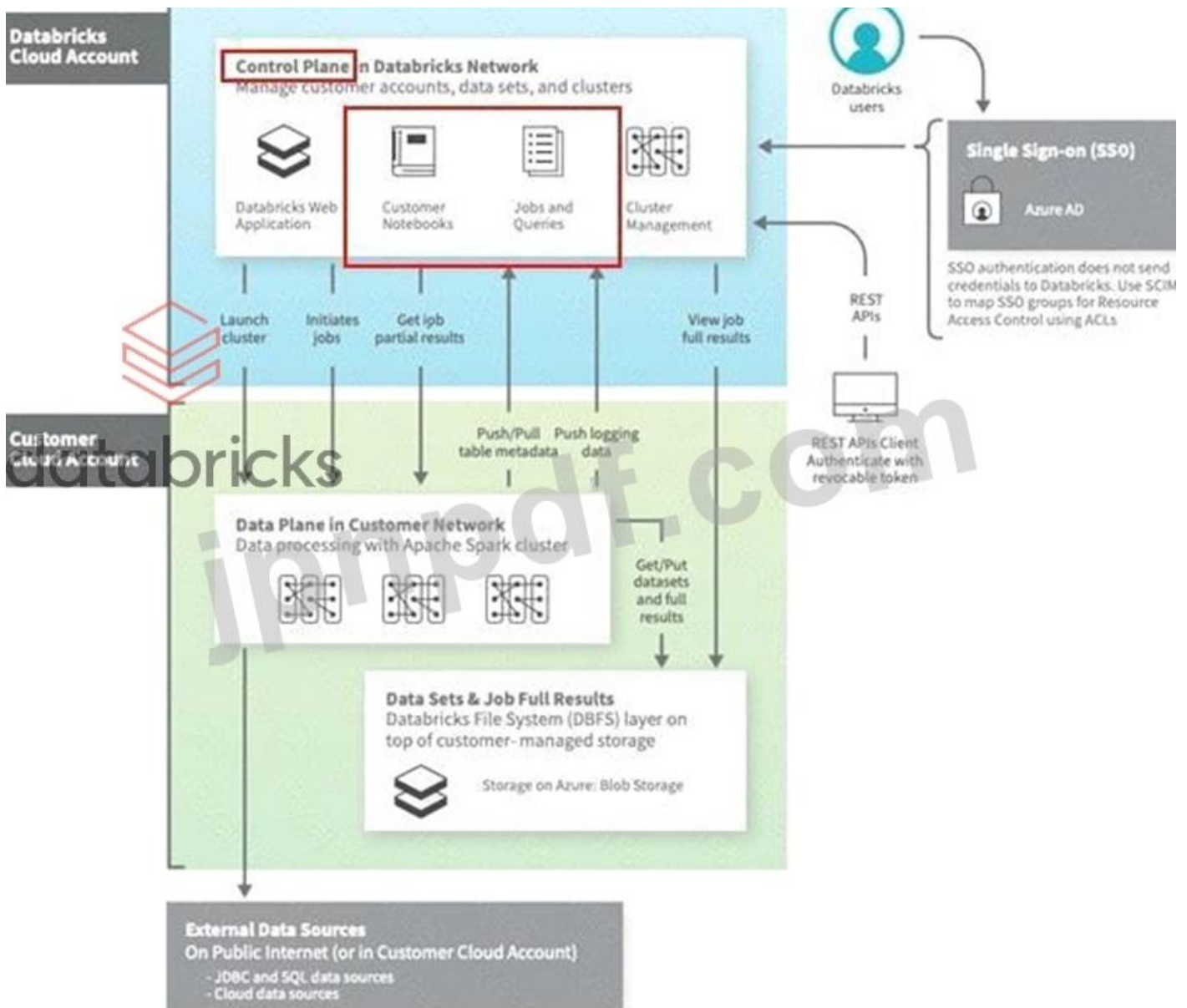
*コントロール プレーンには、Databricks が独自の Azure アカウントで管理するバックエンド サービスが含まれています。Notebook コマンドやその他の多くのワークスペース構成はコントロール プレーンに保存され、保存時に暗号化されます。

データプレーン: 顧客のクラウドアカウントに保存

*データ プレーンは Azure アカウントによって管理され、データが存在する場所です。ここでデータが処理されます。Azure Databricks コネクタを使用すると、クラスターが Azure アカウントの

外部の外部データ ソースに接続して、データを取り込んだりストレージを取得したりできるようになります。

タイムラインの説明が自動生成される



最新問題: 43

部門をパラメーターとして受け取り、それに応じてデータを処理できるノートブックを作成するように求められました。次のステートメントは、ノートブックのパラメーターを Python 変数に保存します。

- A. `SET 部門 = dbutils.widget.get("部門")`
- B. `部門を割り当て == dbutils.widget.get("部門")`
- C. `部門 = dbutils.widget.get("部門")`
- D. `部門 = notebook.widget.get("部門")`
- E. `部門 = notebook.param.get("部門")`

Answer: C ([メッセージを残す](#))

説明

答えは、部門 = `dbutils.widget.get("部門")` です。
ここで追加のドキュメントを参照してください
<https://docs.databricks.com/notebooks/widgets.html>

最新問題: 44

Databricks 製品アーキテクチャ内の次の場所のうち、ジョブ/パイプラインおよびクエリをホストする場所はどこですか？

- A. データプレーン
- B. コントロールプレーン
- C. Databricks ファイルシステム
- D. JDBC データソース
- E. Databricks Web アプリケーション

Answer: B (メッセージを残す)

説明

答えはコントロールプレーンです。

Databricks は、ほとんどのサービスをコントロール プレーンとデータ プレーンから運用します。SQL エンドポイントや DLT コンピューティングなどのサーバーレス機能は、コントロールプレーンで共有コンピューティングを使用することに注意してください。

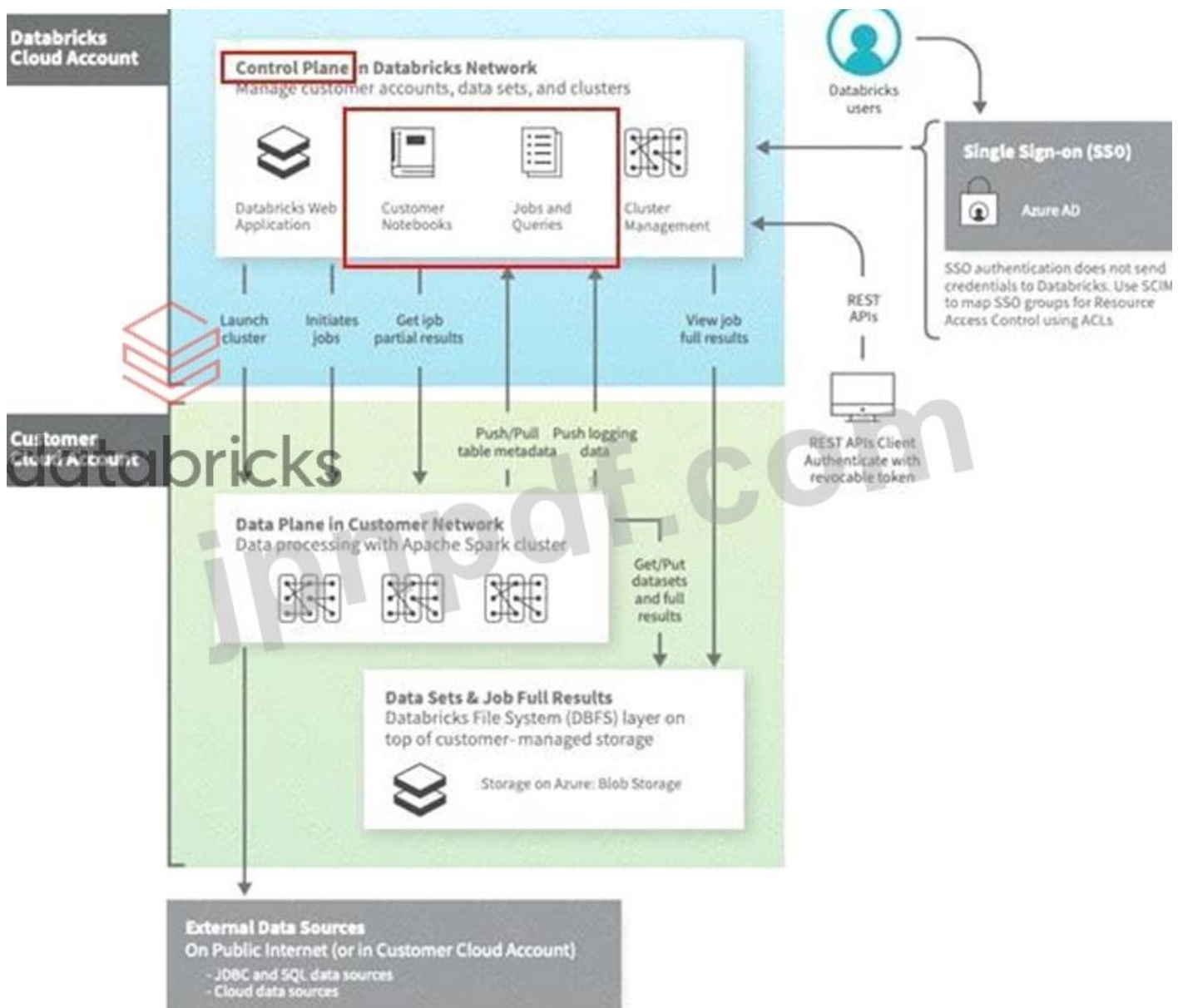
コントロール プレーン: Databricks Cloud アカウントに保存

*コントロール プレーンには、Databricks が独自の Azure アカウントで管理するバックエンドサービスが含まれています。Notebook コマンドやその他の多くのワークスペース構成はコントロール プレーンに保存され、保存時に暗号化されます。

データプレーン: 顧客のクラウドアカウントに保存

*データ プレーンは Azure アカウントによって管理され、データが存在する場所です。ここでデータが処理されます。Azure Databricks コネクタを使用すると、クラスターが Azure アカウントの外部の外部データ ソースに接続して、データを取り込んだりストレージを取得したりできるようになります。

以下は、強調表示されている製品アーキテクチャ図です。



最新問題: 45

データ エンジニアは、次のどのアプローチを使用して、ジョブのスケジュールと構成のバージョン管理可能な構成を取得できますか？

- A. ジョブ クラスター上でジョブを 1 回送信できます。
- B. ジョブのページからジョブと同等の JSON をダウンロードできます。
- C. Databricks リポジトリの一部であるノートブックにジョブをリンクできます。
- D. 汎用クラスター上でジョブを 1 回送信できます。
- E. ジョブのページからジョブの XML 記述をダウンロードできます。

Answer: D ([メッセージを残す](#))

最新問題: 46

ストリーミング データを書き込む場合、Spark の構造化ストリームは以下の書き込みモードをサポートします。

- A. 追加、デルタ、完了

- B. デルタ、完全、連続
- C. 追加、完了、更新
- D. 完全、増分、更新
- E. 追加、上書き、連続

Answer: C (メッセージを残す)

説明

答えは、追加、完了、更新です。

*追加モード (デフォルト) - これはデフォルトのモードで、最後のトリガー以降に結果テーブルに追加された新しい行のみがシンクに出力されます。これは、結果テーブルに追加される行が決して変更されないクエリでのみサポートされます。したがって、このモードでは、各行が 1 回だけ出力されることが保証されます (フォールトトレラントシンクを想定)。たとえば、select、where、map、flatMap、filter、join などのみを含むクエリは追加モードをサポートしません。

*Complete モード - すべてのトリガー後に結果テーブル全体がシンクに出力されます。これは集計クエリでサポートされています。

*更新モード - (Spark 2.1.1 以降で利用可能) 最後のトリガー以降に更新された結果テーブル内の行のみがシンクに出力されます。さらなる情報は将来のリリースで追加される予定です。

有効な **Databricks-Certified-Professional-Data-Engineer** 問題集は GoShiken.com が提供された合格しやすい Databricks-Certified-Professional-Data-Engineer 試験問題集！

GoShiken.com が最新の **Databricks-Certified-Professional-Data-Engineer** 試験問題集を提供しています。GoShiken.com Databricks-Certified-Professional-Data-Engineer 試験問題は最新で、解答が正確でございます。最新の GoShiken.com Databricks-Certified-Professional-Data-Engineer 問題集をゲットする人はこちら: <https://www.goshiken.com/Databricks/Databricks-Certified-Professional-Data-Engineer-mondaishu.html> (20430%OFF問題集溶と正解付き

で 30%w 特別割引コード: **Freepdfdumps**)

最新問題: 47

デルタ レイクは従来のデータ レイクとどう違うのでしょうか？

- A. デルタ レイクは、信頼性、セキュリティ、パフォーマンスを提供できるデータ レイク上のデータウェアハウス サービスです。
- B. デルタ レイクは、信頼性、セキュリティ、パフォーマンスを提供できるデータ レイク上のキャッシュ レイヤーです。
- C. デルタ レイクは、信頼性、セキュリティ、パフォーマンスを提供できる追加機能を備えた寄木細工のようなオープンストレージフォーマットです。
- D. デルタ レイクは、フラット ファイルを、信頼性、セキュリティ、パフォーマンスを提供できる追加機能で置き換えるように設計されたオープンストレージ形式です。

E. デルタ レイクは、信頼性、セキュリティ、パフォーマンスを提供できる Databricks によって設計された独自のソフトウェアです。

Answer: C ([メッセージを残す](#))

説明

答えは、デルタ レイクは、信頼性、セキュリティ、パフォーマンスを提供できる追加機能を備えた寄木細工のようなオープン ストレージ フォーマットです。デルタ レイクは

- * オープンソース
- * 標準データ形式に基づいて構築されます
- * クラウドオブジェクトストレージ用に最適化
- * スケーラブルなメタデータ処理のために構築

デルタ湖はそうではない

- * 独自の技術
- * 保存形式
- * 記憶媒体
- * データベース サービスまたはデータ ウェアハウス

最新問題: 48

サプライ チェーン チームが在庫と製品の注文を監視するために SQL ダッシュボードが構築されましたが、ダッシュボードに表示されるタイムスタンプはすべて UTC 形式で表示されているため、タイム ゾーンをニューヨークの場所に変更するように要求されました。この問題の解決にどのようにアプローチしますか？

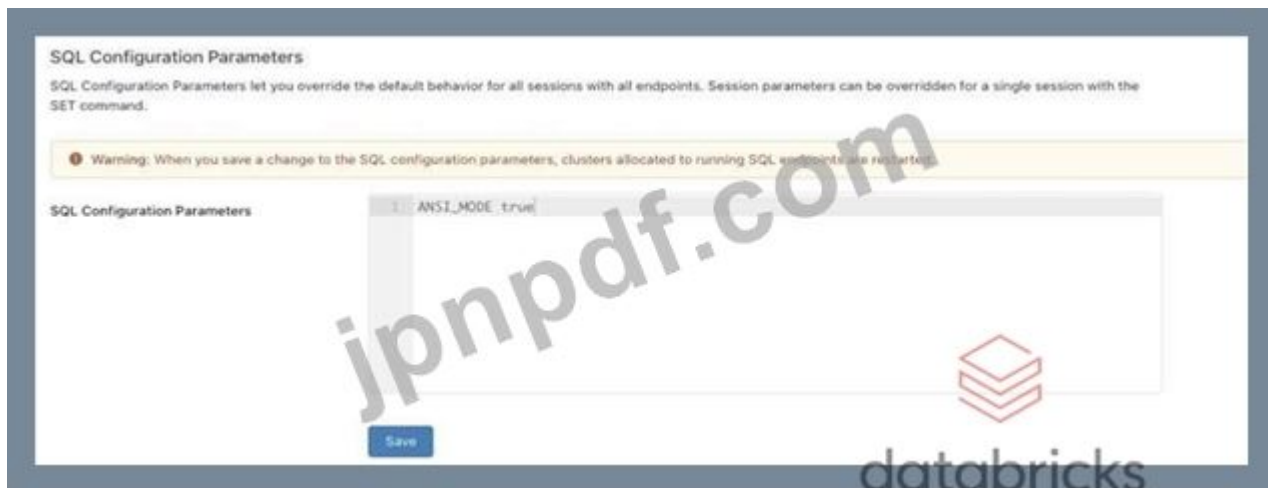
- A. ワークスペースを米国中部ゾーンから米国東部ゾーンに移動します。
- B. デルタ テーブルのタイムスタンプを America/New_York 形式に変更します。
- C. SQL エンドポイントの Spark 構成を変更して、タイムスタンプを America-ca/New_York にフォーマットします。
- D. SQL Admin Console で、SQL 構成パラメータのタイムゾーンを America/New_York に設定します。
- E. ダッシュボードのすべての SQL クエリに SET Timezone = America/New_York を追加します。

Answer: D ([メッセージを残す](#))

説明

答えは、SQL 管理コンソールで、SQL 構成パラメータのタイム ゾーンを America/New_York に設定します。これを構成する手順は次のとおりです。これにより、個々のクエリを変更せずにダッシュボード全体が変更されます。SQL パラメータの構成 SQL パラメータを使用してすべてのウェアハウスを構成するには:

1. サイドバーの下部にある「設定」をクリックし、「SQL 管理コンソール」を選択します。
2. 「SQL ウェアハウス設定」タブをクリックします。
3. 「SQL 構成パラメータ」テキストボックスで、1 行に 1 つのキーと値のペアを指定します。パラメータの名前と値をスペースで区切ります。たとえば、ANSI_MODE を有効にするには:
グラフィカル ユーザー インターフェイス、テキスト、アプリケーションの説明が自動生成される



同様に、SQL 構成パラメータに行を追加できます。
タイムゾーン アメリカ/ニューヨーク
SQL 構成パラメータ | AWS 上のデータブリック

最新問題: 49

データ エンジニアは、次のスキーマを持つテーブル raw_table に JSON ファイルを取り込みました。

- 1.transaction_id STRING、
- 2.ペイロード ARRAY<customer_id:STRING、date:TIMESTAMP、store_id:STRING>

データ エンジニアは、各トランザクションの日付を次のようなテーブルに効率的に抽出したいと考えています。

スキーマ:

- 1.transaction_id STRING、
- 2.日付タイムスタンプ

データ エンジニアがこのタスクを完了するには、次のコマンドのうちどれを実行する必要がありますか？

- A. 1.SELECT トランザクション ID、ペイロード日付
2.FROM raw_table;
- B. 1.SELECT トランザクション ID、ペイロードからの日付
2.FROM raw_table;
- C. 1.SELECT トランザクション ID、explode(ペイロード)
2.FROM raw_table;
- D. 1.SELECT トランザクション ID、日付
2.FROM raw_table;
- E. 1.SELECT トランザクション ID、ペイロード [日付]
2.FROM raw_table;

Answer: ([解答を表示する](#))

最新問題: 50

クラウドオブジェクトストレージから受信する IOT データを 15 分ごとに処理するように AUTO LOADER を構成しました。最近、処理ロジックを更新するためにノートブックコードに変更が加えられましたが、チームはその後、ノートブックが過去 24 時間にわたって障害が発生していたことに気づきました。どのような手順ですか?チームは、ノートブックが修正された後にロードされなかったデータを再処理する必要がありますか?

- A. 処理されなかったファイルを別の場所に移動し、ファイルを取り込みパスに手動でコピーして再処理します。
- B. `back_fill = TRUE` を有効にしてデータを再処理します。
- C. チェックポイント フォルダーを削除し、オートローダーを再度実行します。
- D. オートローダーはロードされなかったデータを自動的に再処理します
- E. データを手動で再読み込みします

Answer: D ([メッセージを残す](#))

説明

答えは、

オートローダーは、チェックポイントを使用してロードされなかったデータを自動的に再処理します。

最新問題: 51

ジョブ クラスターの開始に 6 ~ 8 分かかっており、これによりジョブが時間通りに終了するのが遅れています。クラスターの起動時間を短縮するために実行できる手順は何ですか?

- A. 最初のジョブの前に 2 番目のジョブをセットアップしてクラスターを開始します。これにより、ジョブの開始時にクラスターがリソースを準備できるようになります。
- B. 代わりに汎用クラスターを使用して、クラスターの起動時間を短縮します。
- C. クラスターのサイズを小さくします。クラスターのサイズが小さくなると、クラスターの起動にかかる時間が短くなります。
- D. クラスター プールを使用してジョブの起動時間を短縮します。
- E. SQL エンドポイントを使用して起動時間を短縮します。

Answer: ([解答を表示する](#)**)**

説明

答えは、クラスター プールを使用してジョブの起動時間を短縮することです。

クラスター プールを使用すると、新しいジョブ クラスターが作成されるときに VM をプールから取得するときに、事前に VM を予約できます。注: VM がクラスターによる使用を待機している場合、発生するコストは Azure のみです。Databricks の実行時間コストは、VM がクラスターに割り当てられた場合にのみ請求されます。

ここでは、セットアップ方法といくつかのベスト プラクティスに従う方法のデモを示します。

https://www.youtube.com/watch?v=FVtITxOabxg&ab_channel=DatabricksAcademy

最新問題: 52

DELTA テーブルを管理するときにストレージ内でデータがどのように編成されるかについて正しい記述は次のうちどれですか?

- A. すべてのデータは 1 つまたは複数の寄せ木ファイルに分割され、ログ ファイルは 1 つまたは複数の JSON ファイルに分割され、トランザクションごとに新しいデータ ファイルとログ ファイルが作成されます。
(正しい)
- B. すべてのデータとログは単一の寄せ木細工のファイルに保存されます
- C. すべてのデータは 1 つまたは複数の寄せ木細工のファイルに分割されますが、ログ ファイルは 1 つの json ファイルとして保存され、トランザクションごとに新しいデータ ファイルが作成され、ログ ファイルが追加されます。
- D. すべてのデータは 1 つまたは複数の寄せ木細工のファイルに分割され、トランザクションがコミットされるとログ ファイルは削除されます。
- E. すべてのデータは 1 つの寄せ木細工ファイルに保存され、ログ ファイルは 1 つまたは複数の json ファイルに分割されます。

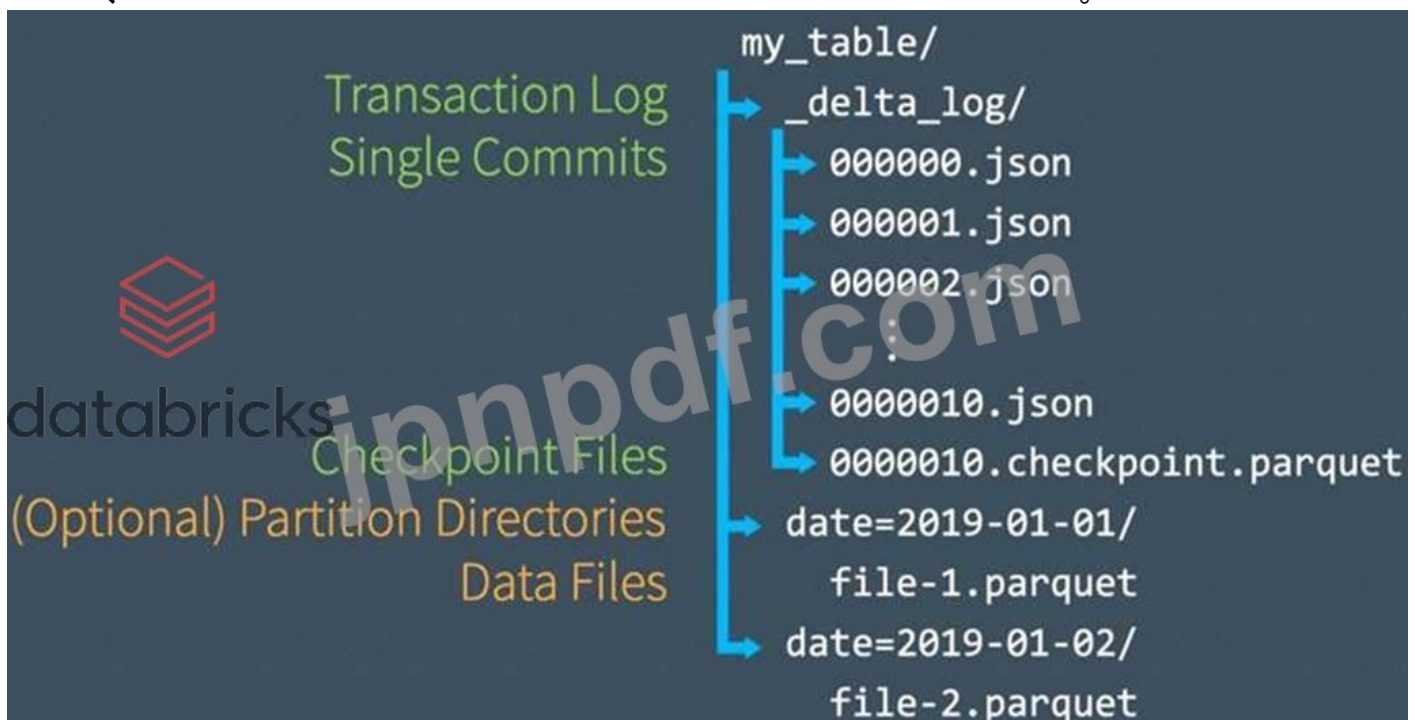
Answer: A ([メッセージを残す](#))

説明

答えは

すべてのデータは 1 つまたは複数の寄せ木細工ファイルに分割され、ログ ファイルは 1 つまたは複数の json ファイルに分割され、トランザクションごとに新しいデータ ファイルとログ ファイルが作成されます。

これは、DELTA テーブルがどのように見えるかのサンプル レイアウトです。



最新問題: 53

Databricks ジョブで 2 つのタスクを設定するように求められます。最初のタスクはノートブックを実行してリモート システムからデータをダウンロードし、2 番目のタスクはこのデータを処理できる DLT パイプラインです。ジョブでこれをどのように構成する予定ですか? UI

- A. 1つのジョブにノートブック タスクと DLT パイプライン タスクを含めることはできません。線形依存関係を持つ2つの異なるジョブを使用します。
- B. ジョブ UI は DTL パイプラインをサポートしていません。ジョブ UI を使用して最初のタスクをセットアップし、連続モードで実行するように DLT をセットアップします。
- C. ジョブ UI は DTL パイプラインをサポートしていません。ジョブ UI を使用して最初のタスクをセットアップし、トリガー モードで実行するように DLT をセットアップします。
- D. 単一のジョブを使用してノートブックと DLT パイプラインの両方をセットアップでき、線形依存関係を持つ2つの異なるタスクを使用します。
- E. DLT パイプラインに最初のステップを追加し、ジョブ UI でトリガー モードとして DLT パイプラインを実行します。

Answer: D (メッセージを残す)

説明

答えは、単一のジョブを使用してノートブックと DLT パイプラインの両方をセットアップでき、線形依存関係を持つ2つの異なるタスクを使用できます。これが JOB UI です。

1. ノートブックタスクを作成する
2. DLTタスクの作成
 - a. ノートブックタスクを依存関係として追加します

3. 完成図

ノートブックタスクを作成する

グラフィカル ユーザー インターフェイス、テキスト、アプリケーション、電子メールの説明が自動生成されます

Task name * ⓘ

Notebook_task

Type * ⓘ

Notebook | v

Source * ⓘ

Workspace

Path * ⓘ

/Users/ /Data Analysis

Cluster * ⓘ

Shared_job_cluster (126.00 GB | 36 Cores | DBR 10.4 LTS | Spark 3.2.1 | Scala 2.12) | v

Parameters ⓘ

Add

Depends on


Select task dependencies... | v

Advanced options

Cancel Save task

DLT タスク

グラフィカル ユーザー インターフェイス、テキスト、アプリケーション、電子メールの説明が自動生成されます



Task name * ⓘ databricks

DLT_task

Type * Pipeline * ⓘ

Delta Live Tables pipeline | v

Depends on

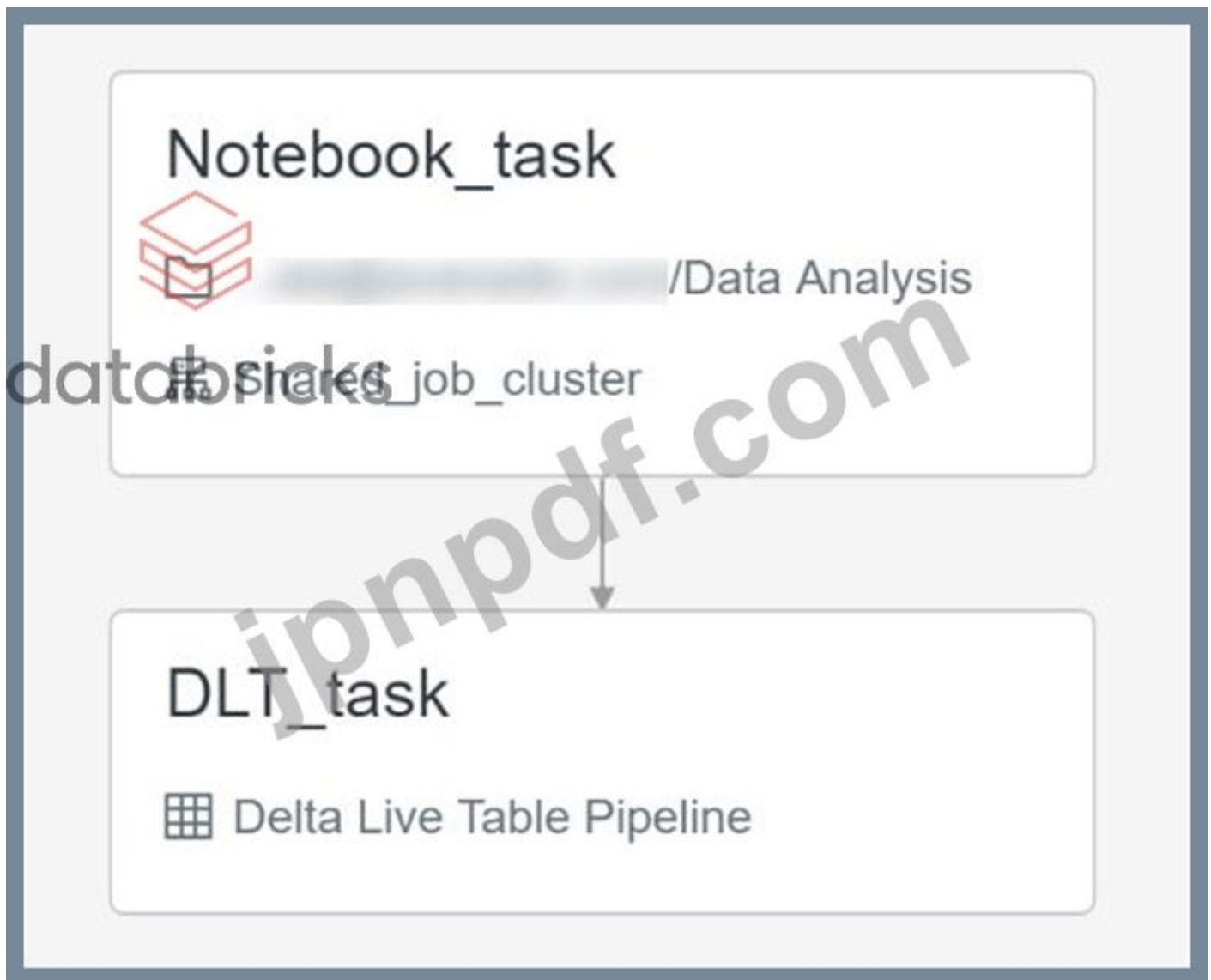
Notebook_task x | v

Advanced options

Cancel Save task

最終ビュー

グラフィカル ユーザー インターフェイス、テキスト、アプリケーション、PowerPoint の説明が自動生成されます



フォームの下部
フォームの先頭

最新問題: 54

連続確率分布は次のうちどれですか？

- A. 正規確率分布
- B. 二項確率分布
- C. ポアソン確率分布
- D. 負の二項分布

Answer: A ([メッセージを残す](#))

最新問題: 55

メダリオン建築におけるブロンズ層とシルバー層の主な違いは何ですか？

- A. ブロンズでは重複が削除され、シルバーではスキーマが適用されます。
- B. シルバーには集約されたデータが含まれる可能性があります
- C. ブロンズは取り込まれたデータの生のコピー、シルバーには実稼働スキーマを含むデータが含まれ、ELT/ETL スループット用に最適化されています。

D. 不良データはブロンズでフィルタリングされ、シルバーはブロンズ データのコピーです。

Answer: ([解答を表示する](#))

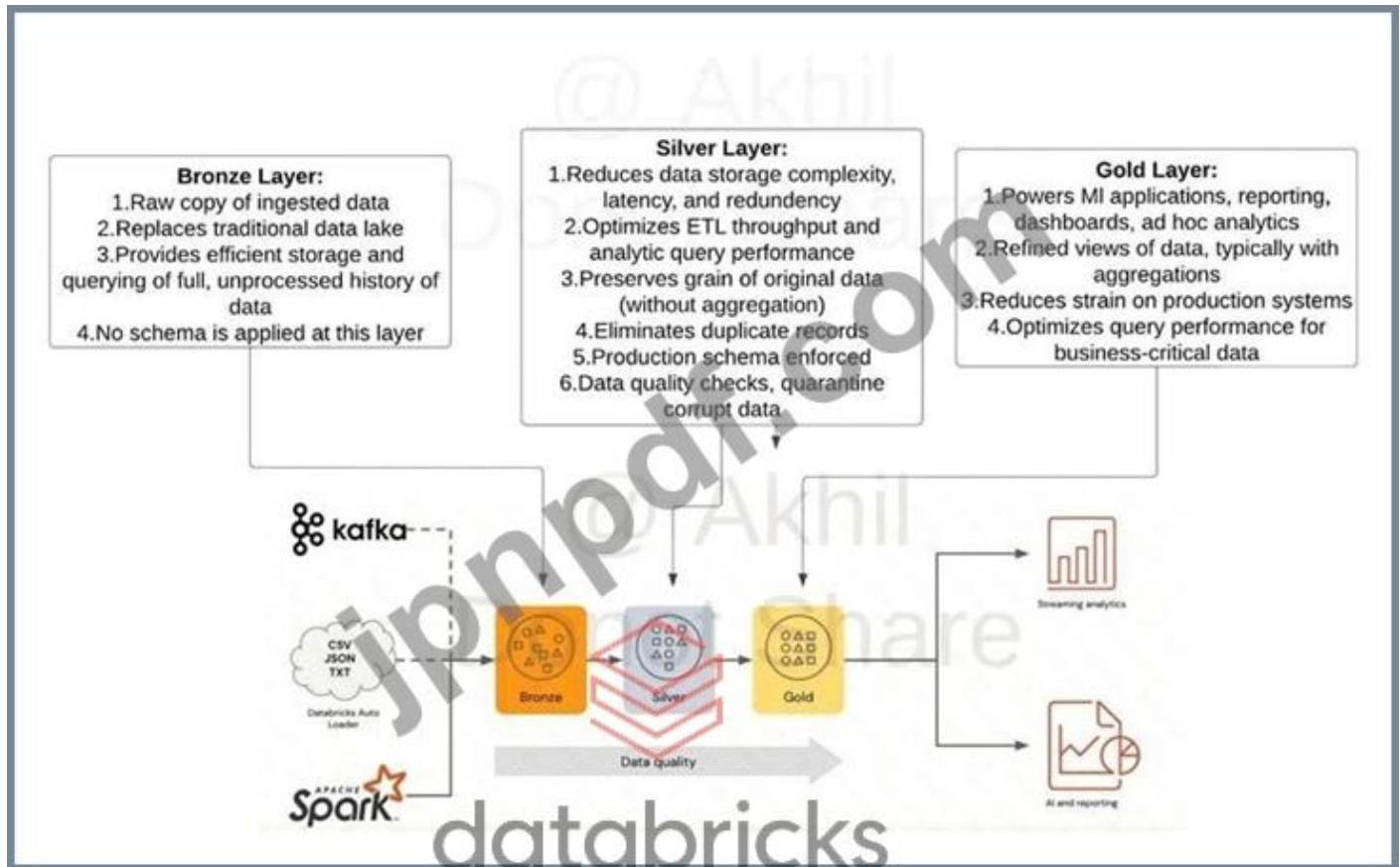
説明

メダリオン アーキテクチャ - Databricks

試験の焦点: 下の画像を確認して、メダリオン建築における各層 (ブロンズ、シルバー、ゴールド) の役割を理解してください。各層とその目的を対象としたさまざまな質問が表示されます。

Udemy 内の一部の人が私のコンテンツをコピーしているため、ウォーターマークを追加する必要がありました。

低信頼度で自動的に生成された家の説明図



最新問題: 56

データ エンジニアリング チームは、ETL ジョブを監視するために SQL クエリを使用してデータの完全性を毎日レビューしており、クエリの出力は複数のダッシュボードで使用されています。スケジュールを設定してこのプロセスを自動化するには、次のどのアプローチを使用できますか？

- A. ジョブ UI からクエリを毎日実行するようにスケジュールできます。
- B. Databricks SQL のクエリのページから毎日更新されるようにクエリをスケジュールできます。
- C. ジョブ UI から 12 時間ごとにクエリを実行するようにスケジュールできます。
- D. Databricks SQL の SQL エンドポイントのページからクエリを毎日更新するようにスケジュールできます。
- E. Databricks SQL の SQL エンドポイントのページから 12 時間ごとにクエリを更新するようにスケジュールできます。

Answer: ([解答を表示する](#))

説明

答えは、Databricks SQL の SQL エンドポイントのページから 12 時間ごとにクエリを更新するようにスケジュールできることです。Databricks SQL ワークスペースのクエリ ペイン ビューでは、実行する個々のクエリを追加または編集し、スケジュールする機能が提供されます。

スケジュールされたクエリの実行を使用して、ダッシュボードを更新したり、定期的なアラートを有効にしたりできます。デフォルトでは、クエリにはスケジュールがありません。

ノート

クエリがアラートで使用されている場合、アラートは独自の更新スケジュールで実行され、クエリ スケジュールは使用されません。

スケジュールを設定するには:

- * クエリ情報タブをクリックします。
- * グラフィカル ユーザー インターフェイス、テキスト、アプリケーション、電子メールの説明が自動生成されます
- * [更新スケジュール] の右側にあるリンクをクリックして、スケジュール間隔を指定するピッカーを開きます。
- * グラフィカル ユーザー インターフェイス、アプリケーションの説明が自動的に生成される
- * 3.スケジュールを設定します。
- * ピッカーはスクロールして次の項目を選択できます。
- * *間隔: 1 ~ 30 分、1 ~ 12 時間、1 ~ 30 日、1 ~ 2 週間
- * *時間。時間セレクターは、間隔が 1 日より大きく、曜日の選択が 1 週間より大きい場合にのみピッカーに表示されます。特定の時刻をスケジュールすると、Databricks SQL はコンピューターのタイムゾーンで入力を受け取り、それを UTC に変換します。UTC の特定の時間にクエリを実行したい場合は、ローカル オフセットによってピッカーを調整する必要があります。たとえば、クエリを毎日 00:00 UTC に実行したいが、現在のタイムゾーンが PDT (UTC-7) である場合は、ピッカーで 17:00 を選択する必要があります。
- * グラフィカル ユーザー インターフェイス 説明が自動生成されます

最新問題: 57

インベントリ プロセスの最後に、ファイルがクラウド オブジェクト ストレージにアップロードされます。データを取り込むプロセスを構築するように求められます。次のどの方法を使用してデータを増分的に取り込むことができます。ファイルのスキーマが必要です。時間外に取り込みプロセスを変更するには、これらの変更を自動的に処理する必要があります。

以下は、データをロードするコマンドを実行するオートローダーです。以下のコードを正常に実行するには、空白を埋めてください。

- 1.spark.readStream
- 2..format("クラウドファイル")
- 3..option("_____", "csv")
- 4..option("_____", 'dbfs:/location/checkpoint/')
- 5..load(データソース)
- 6..writeStream

7..option("_____", 'dbfs:/location/checkpoint/')

8..option("_____", "true")

9..テーブル(テーブル名))

A. フォーマット、チェックポイントの場所、スキーマの場所、上書き

B. cloudfiles.format、checkpointlocation、cloudfiles.schemalocation、上書き

C. cloudfiles.format、cloudfiles.schemalocation、checkpointlocation、mergeSchema

D. cloudfiles.format、cloudfiles.schemalocation、checkpointlocation、上書き

E. cloudfiles.format、cloudfiles.schemalocation、checkpointlocation、追加

Answer: C ([メッセージを残す](#))

説明

答えは、cloudfiles.format、cloudfiles.schemalocation、checkpointlocation、mergeSchema です。ストリーミング ELT のエンドツーエンド構文は次のとおりです。以下のリンクには完全なオプションが含まれています。Auto Loader オプション | AWS 上のデータブリック

1.spark.readStream

2..format("cloudfiles") # ストリーム データ ソースを返し、トリガーに基づいて到着したデータを読み取ります。

3..option("cloudfiles.format","csv") # 受信ファイルの形式

4..option("cloudfiles.schemalocation", "dbfs:/location/checkpoint/") 推論されたスキーマとその後の変更を保存する場所

5..load(データソース)

6..writeStream

7..option("checkpointlocation","dbfs:/location/checkpoint/") # ストリームのチェックポイントの場所

8..option("mergeSchema", "true") # 複数のファイルにわたるスキーマを推測し、各ファイルのスキーマをマージします。スキーマを推論するとき、Auto Loader に対してデフォルトで有効になります。

9..table(table_name)) # ターゲットテーブル

最新問題: 58

Unity カタログ機能を使用するには、Databricks ワークスペースの管理テーブル/外部テーブルで次のどの手順を実行する必要がありますか？

A. ワークスペース設定で Unity カタログ機能を有効にする

B. ワークスペース管理対象/外部テーブル/ビュー内のオブジェクトを Unity カタログに移行/アップグレードします。

C. DBR バージョン 15.0 にアップグレードします。

D. ワークスペースから Unity カタログにデータをコピーします。

E. ワークスペースを Unity カタログにアップグレードします

Answer: B ([メッセージを残す](#))

説明

テーブルとビューを Unity カタログにアップグレードする - Azure Databricks | Microsoft Docs
マネージド テーブル: マネージド テーブルを Unity カタログにアップグレードする
外部テーブル: 外部テーブルを Unity カタログにアップグレードする

最新問題: 59

次のタイプのタスクのうち、ジョブを通じてセットアップできないものはどれですか？

- A. パイソン
- B. Databricks SQL ダッシュボードの更新
- C. ノート
- D. デルタライブパイプライン
- E. スパークサブミット

Answer: ([解答を表示する](#))

最新問題: 60

デルタ湖について説明しているのは次のどれですか？

- A. Delta Lake は、完全な機械学習ライフサイクルの管理を支援するオープンソース プラットフォームです。
- B. Delta Lake は、信頼性、セキュリティ、パフォーマンスを提供するオープン フォーマットのストレージ レイヤーです
- C. Delta Lake は、分散データ用のオープンソース データ ストレージ形式です。
- D. Delta Lake は、ビッグ データ ワークロードに使用されるオープンソース分析エンジンです。
- E. デルタレイクは、データを処理するオープンフォーマットのストレージレイヤーです

Answer: ([解答を表示する](#))

説明

デルタ湖

最新問題: 61

グローバル一時ビューについて正しいのは次のうちどれですか？

- A. ノートブックをデタッチしてアタッチすると、グローバル一時ビューにアクセスできなくなります。
- B. グローバル一時ビューは多くのクラスター間でアクセス可能
- C. ノートブックが切り離されたり接続されたりしても、グローバル一時ビューには引き続きアクセスできます。
- D. クラスターが再起動されても、グローバル一時ビューには引き続きアクセスできます。
- E. グローバル一時ビューは、temp データベースと呼ばれるデータベースに作成されます。

Answer: C ([メッセージを残す](#))

説明

答えは、グローバル一時ビューには、ノートブックが切り離されてアタッチされていてもアクセスできるということです。作成できる一時ビューには、ローカルとグローバルの 2 種類があります。

* ローカル一時ビューは Spark セッションでのみ使用できるため、同じクラスター内の別のノートブックはアクセスできません。ノートブックが切り離され、再接続された場合、ローカルの一時ビューが失われます。

* グローバル一時ビューはクラスター内のすべてのノートブックで使用できます。ノートブックが切り離されて再接続された場合でも引き続きアクセスできますが、クラスターが再起動されるとグローバル一時ビューは失われます。

有効な **Databricks-Certified-Professional-Data-Engineer** 問題集は GoShiken.com が提供された合格しやすい Databricks-Certified-Professional-Data-Engineer 試験問題集！

GoShiken.com が最新の **Databricks-Certified-Professional-Data-Engineer** 試験問題集を提供しています。GoShiken.com Databricks-Certified-Professional-Data-Engineer 試験問題は最新で、解答が正確でございます。最新の GoShiken.com Databricks-Certified-Professional-Data-Engineer 問題集をゲットする人はこちら: <https://www.goshiken.com/Databricks/Databricks-Certified-Professional-Data-Engineer-mondaishu.html> (20430%OFF問題集溶と正解付きで 30%w 特別割引コード: **Freepdfdumps**)

最新問題: 62

現在カタログまたはスキーマにアクセスできない新規ユーザーが sales スキーマの customer テーブルへのアクセスを要求していますが、customer テーブルには機密情報が含まれているため、機密で許可されている列を除いたテーブルのビューを作成することにしました。

user@company.com にビュー GRANT SELECT ON view_name にアクセスしますが、ユーザーがビューをクエリしようとする、ビューが存在しないというエラーが表示されます。ユーザーがビューにアクセスできない問題は何ですか?またその修正方法は何ですか?

- A. ユーザーは基になるテーブルに対して SELECT を必要とします。
- B. ユーザーは、PII データにアクセスできる特別なグループに所属する必要があります
- C. ユーザーはビューの所有者である必要があります
- D. ユーザーには Sales スキーマに対する USAGE 権限が必要です
- E. ユーザーにはビューに対する ADMIN 権限が必要です

Answer: D (メッセージを残す)

説明

答えは、ユーザーには Sales スキーマに対する USAGE 権限が必要です。

データ オブジェクトの権限 - Azure Databricks | Microsoft ドキュメント

スキーマ販売での使用を user@company.com に許可します。

*USAGE: いかなる能力も与えませんが、スキーマ オブジェクトに対して何らかのアクションを実行するための追加要件です。

最新問題: 63

クラスターが削除されたら、管理者は以下の追加のアクションを実行する必要があります。

- A. 仮想マシンを削除しますが、ストレージとネットワークは自動的に削除されません
- B. ストレージ ディスクは削除されますが、仮想マシンとネットワークは自動的に削除されません
- C. ネットワークを削除しますが、仮想マシンとストレージ ディスクは自動的に削除されません
- D. ログを削除します
- E. 何もする必要はありません。すべてのリソースは自動的に削除されます。

Answer: E (メッセージを残す)

説明

デルタとは何ですか？

デルタ湖は

- * オープンソース
- * 標準データ形式に基づいて構築されます
- * クラウドオブジェクトストレージ用に最適化
- * スケーラブルなメタデータ処理のために構築

デルタ湖はそうではない

- * 独自の技術
- * 保存形式
- * 記憶媒体
- * データベース サービスまたはデータ ウェアハウス

最新問題: 64

2つのテーブルを結合してオブジェクトを作成しようとしています。そのテーブルはデータサイエンティストのチームがアクセスできるため、クラスターが再起動したり、ノートブックが切り離されてもオブジェクトは削除されません。どのような種類のオブジェクトを作成しようとしていますか？

- A. 一時的なビュー
- B. グローバル一時ビュー
- C. キャッシュ オプションを使用したグローバル一時ビュー
- D. 外観図
- E. ビュー

Answer: E (メッセージを残す)

説明

答えはビューです。ビューは複数のテーブルを結合するために使用できますが、他のユーザーがアクセスできるようにメタストアに永続化することもできます。

最新問題: 65

データの問題を調査しているときに、プロセスが誤ってテーブルを更新したことに気づきました。前のバージョンがどのようになっているかを確認できるように、昨日のバージョンのデータで同じテーブルをクエリしたいと考えています。履歴データをクエリする最適な方法は何ですか。あなたの分析はできますか？

- A. `SELECT * FROM TIME_TRAVEL(テーブル名) WHERE time_stamp = 'タイムスタンプ'`

- B. TIME_TRAVEL FROM table_name WHERE time_stamp = date_sub(current_date(), 1)
- C. SELECT * FROM table_name TIMESTAMP AS OF date_sub(current_date(), 1)
- D. date_sub(current_date(), 1) 時点の履歴テーブル名をディスクリプト
- E. show HISTORY table_name AS OF date_sub(current_date(), 1)

Answer: C ([メッセージを残す](#))

説明

答えは、SELECT * FROM table_name TIMESTAMP as of date_sub(current_date(), 1) 参考までに、タイム トラベルでは、タイムスタンプを使用する方法と、バージョン番号のタイムスタンプを使用する方法の2つがサポートされています。

- 1.SELECT count(*) FROM my_table TIMESTAMP AS OF "2019-01-01"
- 2.SELECT count(*) FROM my_table TIMESTAMP AS OF date_sub(current_date(), 1)
- 3.SELECT count(*) FROM my_table TIMESTAMP AS OF "2019-01-01 01:30:00.000" バージョン番号:

- 1.SELECT count(*) FROM my_table VERSION AS OF 5238
- 2.my_table@v5238 から count(*) を選択します
- 3.SELECT count(*) FROM delta.`/path/to/my/table@v5238`

<https://databricks.com/blog/2019/02/04/introducing-delta-time-travel-for-large-scale-data-lakes.html>

最新問題: 66

湖畔の家について正しいのは次のうちどれですか？

- A. Lakehouse は機械学習ワークロードのみをサポートし、データ ウェアハウスはBI ワークロードをサポートします。
- B. Lakehouse はエンドツーエンドのストリーミング ワークロードのみをサポートし、データ ウェアハウスはバッチ ワークロードをサポートします。
- C. Lakehouse は ACID をサポートしていません
- D. Lakehouse は SQL をサポートしていません
- E. Lakehouse はトランザクションをサポートします

Answer: ([解答を表示する](#)**)**

説明

レイクハウスとは何ですか? - Databricks ブログ
自動生成されるテキスト説明

A lakehouse has the following key features:

- **Transaction support:** In an enterprise lakehouse many data pipelines will often be reading and writing data concurrently. Support for ACID transactions ensures consistency as multiple parties concurrently read or write data, typically using SQL.
- **Schema enforcement and governance:** The Lakehouse should have a way to support schema enforcement and evolution, supporting DW schema architectures such as star/snowflake-schemas. The system should be able to **reason about data integrity**, and it should have robust governance and auditing mechanisms.
- **BI support:** Lakehouses enable using BI tools directly on the source data. This reduces staleness and improves recency, reduces latency, and lowers the cost of having to operationalize two copies of the data in both a data lake and a warehouse.
- **Storage is decoupled from compute:** In practice this means storage and compute use separate clusters, thus these systems are able to scale to many more concurrent users and larger data sizes. Some modern data warehouses also have this property.
- **Openness:** The storage formats they use are open and standardized, such as Parquet, and they provide an API so a variety of tools and engines, including machine learning and Python/R libraries, can efficiently access the data **directly**.
- **Support for diverse data types ranging from unstructured to structured data:** The lakehouse can be used to store, refine, analyze, and access data types needed for many new data applications, including images, video, audio, semi-structured data, and text.
- **Support for diverse workloads:** including data science, machine learning, and SQL and analytics. Multiple tools might be needed to support all these workloads but they all rely on the same data repository.
- **End-to-end streaming:** Real-time reports are the norm in many enterprises. Support for streaming eliminates the need for separate systems dedicated to serving real-time data applications.

最新問題: 67

データ エンジニアは、クエリの一部として 2 つのテーブルを水平方向に結合したいと考えています。彼らは共有を使用したいと考えています

列をキー列として使用し、クエリ結果にはキー列の値が両方のテーブルに存在します。

このタスクを達成するために使用できる SQL コマンドは次のうちどれですか？

- A. マージ
- B. 内部結合
- C. 左結合
- D. 外部結合
- E. ユニオン

Answer: ([解答を表示する](#))

最新問題: 68

インタラクティブ ノートブックの結果は Databricks 製品アーキテクチャのどこに保存されますか？

- A. データプレーン
- B. コントロールプレーン
- C. データおよびコントロール プレーン
- D. JDBC データソース
- E. Databricks Web アプリケーション

Answer: ([解答を表示する](#))

説明

答えはデータ プレーンとコントロール プレーンです。

ジョブの結果のみがデータ プレーン (お客様のストレージ) に保存され、インタラクティブ ノートブックの結果は、コントロール プレーン (UI で表示される部分的な結果) と顧客のストレージの組み合わせに保存されます。

<https://docs.microsoft.com/en-us/azure/databricks/getting-started/overview#--high-level-architecture> 上記ドキュメントの抜粋、グラフィカル ユーザー インターフェイス、アプリケーションの説明が自動的に生成

Job results reside in storage in your account.

Interactive notebook results are stored in a combination of the control plane (partial results for presentation in the UI) and your Azure storage. If you want interactive notebook results stored only in your cloud account storage, you can ask your Databricks representative to enable *interactive notebook results in the customer account* for your workspace. Note that some metadata about results, such as chart column names, continues to be stored in the control plane. This feature is in Public Preview.

この行動を変えるにはどうすればよいでしょうか？

この動作は、そのワークスペースのワークスペース/管理コンソール設定を使用して変更できます。有効にすると、Databricks が最近導入した新しいノートブック視覚化機能を除き、すべてのインタラクティブな結果が顧客アカウント (データ プレーン) に保存されます。これは依然として一部のメタデータを以下の設定に関係なく、コントロール ペインに表示されます。詳細については、ドキュメントを参照してください。

グラフィカル ユーザー インターフェイス、テキスト、アプリケーション、電子メールの説明が自動生成されます

Admin Console



> DBFS File Browser: Enabled	<input checked="" type="checkbox"/>
> Databricks Autologging: Disabled	<input type="checkbox"/>
> MLflow Run Artifact Download: Enabled	<input checked="" type="checkbox"/>
> MLflow Classic Model Serving Endpoint Creation: Enabled	<input checked="" type="checkbox"/>
> MLflow Model Registry Email Notifications: Enabled	<input checked="" type="checkbox"/>
> RStudio Home Directory: /home	<input type="text" value="Enter valid home directory for RStudio, eg: /home"/> Save
▼ Store Interactive Notebook Results in Customer Account: Disabled	<input type="checkbox"/>
When enabled, all interactive notebook results are stored in the customer account.	
> Increased number of jobs: Disabled	<input type="checkbox"/>
> Verbose Audit Logs (Temporarily disabled in Databricks SQL): Disabled	<input type="checkbox"/>

これを知ることがなぜ重要なのでしょうか？

私は最近、顧客の機密情報を扱う必要があるプロジェクトに取り組んでおり、ノートブックの結果を含むすべてのデータをデータ プレーンに保存する必要があるというセキュリティ要件がありました。

最新問題: 69

次のデータ ワークロードのうち、シルバー テーブルをソースとして利用するものはどれですか？

- A. タイムスタンプを人間が判読できる形式に解析してデータを強化するジョブ
- B. すでにダッシュボードにフィードされている集計データをクエリするジョブ
- C. ストリーミング ソースから生データを Lakehouse に取り込むジョブ
- D. クリーンアップされたデータを集約して標準の要約統計を作成するジョブ
- E. 不正なフォーマットのレコードを削除してデータをクリーンアップするジョブ

Answer: D (メッセージを残す)

説明

答えは、クリーンなデータを集約して標準の概要統計を作成するジョブです。シルバー ゾーンは元のデータの粒度を維持します。このシナリオでは、ジョブはシルバー ゾーンからデータをソースとして取得し、それらを集約してゴールド ゾーンに保存します。

メダリオン アーキテクチャ - Databricks

シルバー層:

1. データストレージの複雑さ、遅延、冗長性を軽減します。
2. ETL スループットと分析クエリのパフォーマンスを最適化します。
3. 元のデータの粒度を維持します (集計なし)
4. 重複レコードの削除

5. 運用スキーマの適用

6. データ品質チェック、破損したデータの隔離

試験の焦点: 下の画像を確認して、メダリオン建築における各層 (ブロンズ、シルバー、ゴールド) の役割を理解してください。各層とその目的を対象としたさまざまな質問が表示されます。

Udemy 内の一部の人が私のコンテンツをコピーしているため、ウォーターマークを追加する必要がありました。

メダリオンアーキテクチャの各層の目的

最新問題: 70

データ エンジニアリング チームには現在、毎日午前 8 時にレポート テーブルにデータをロードするタスクを実行するジョブが設定されており、約 20 分かかります。運用チームは、そのデータを使用して 2 番目のジョブを実行することを計画しており、最新の完全なセットにアクセスします。データの。このジョブ設定を調整する最善の方法は何ですか？

- A. 同じジョブにオペレーション レポート タスクを追加し、データ エンジニアリング タスクがオペレーション レポート タスクに依存するように設定します。
- B. 同じワークスペースで午前 8 時 20 分に実行される 2 番目のジョブをセットアップします。
- C. 同じジョブにオペレーション レポート タスクを追加し、データ エンジニアリング タスクに依存するようにオペレーション レポート タスクを設定します。
- D. 自動ローダーを使用して 20 分ごとに実行し、最初のテーブルを読み取り、トリガーを 1 回に設定し、2 番目のジョブを作成します。
- E. 最初のテーブルに基づいてデルタ ライブをテーブルに設定し、ジョブを連続モードで実行するように設定します。

Answer: C ([メッセージを残す](#))

説明

答えは、同じジョブに操作レポート タスクを追加し、データ エンジニアリング タスクに依存するように操作レポート タスクを設定することです。

Task name * ⓘ

OperationsReporting

Type * | **Source *** ⓘ

Notebook | Workspace

Path * ⓘ

/Users/ [REDACTED]

Cluster * ⓘ

cluster (126.00 GB | 36 Cores | DBR 10.4 LTS | Spark 3.2.1 | Scala 2.12)

Parameters ⓘ UI | JSON

Add

Depends on

Dataengineering x

Advanced options

Cancel Create task

中程度の信頼度で自動生成される図の説明



最新問題: 71

データセットはデルタ ライブ テーブルを使用して定義されており、期待句が含まれています。

1. CONSTRAINT valid_timestamp EXPECT (タイムスタンプ > '2020-01-01')

これらの制約に違反するデータを含むデータのバッチが次のような場合に予想される動作は何ですか？

加工された？

- A. 期待に反するレコードによりジョブが失敗します。
- B. 期待に違反するレコードはターゲット データセットに追加され、ターゲット データセットに追加されたフィールドで無効としてフラグが立てられます。
- C. 期待に違反するレコードはターゲット データセットから削除され、隔離テーブルにロードされます。
- D. 期待に反するレコードはターゲット データセットから削除され、イベント ログに無効として記録されます。
- E. 期待に反するレコードがターゲット データセットに追加され、イベント ログに無効として記録されます。

Answer: E ([メッセージを残す](#))

最新問題: 72

現在の ELT パイプラインは運用チームから 1 日に 1 回データを受信しているため、トリガー (Once = True) を使用して 1 日に 1 回実行するように AUTO LOADER プロセスを設定し、1 日に 1 回実行するジョブをスケジュールしました。運用チームは最近、1 分ごとにデータを送信できる新機能ですが、1 分ごとにデータを処理するには AUTO LOADER にどのような変更を加える必要がありますか。

- A. AUTO LOADER トリガーを .trigger(ProcessingTime = "1 minutes") に変更します。
- B. オートローダートリガーを(『分』)に変更します。
- C. AUTO LOADER を構造化ストリーミングに変換する
- D. ストリーム処理を有効にする
- E. ジョブ クラスターをセットアップし、ノートブックを 1 分に 1 回実行します。

Answer: A ([メッセージを残す](#))

最新問題: 73

Databricks で管理されるクラスターのドライバー ノードとワーカー ノードをホストする場所は次のどれですか？

- A. データプレーン
- B. コントロールプレーン
- C. Databricks ファイルシステム
- D. Databricks Web アプリケーション
- E. JDBC データソース

Answer: A ([メッセージを残す](#))

説明

Databricks の高レベルのアーキテクチャを参照してください。

最新問題: 74

CREATE DATABASE sample_db ステートメントを使用してデータベース sample_db を作成する場合、DBFS 内のデータベースのデフォルトの場所はどこになりますか？

- A. デフォルトの場所、DBFS:/user/
- B. デフォルトの場所、/user/db/
- C. デフォルトのストレージ アカウント
- D. ステートメントは失敗します。場所がないとデータベースを作成できません」
- E. デフォルトの場所、dbfs:/user/hive/warehouse

Answer: E ([メッセージを残す](#))

説明

答えは dbfs:/user/hive/warehouse です。これは、Spark がユーザー データベースを保存するデフォルトの場所です。デフォルトは、spark.sql.warehouse.dir パラメータを使用して変更できません。LOCATION キーワードを使用してカスタムの場所を指定することもできます。

これがどのように機能するかは次のとおりです。

グラフィカル ユーザー インターフェイス、テキスト、アプリケーション、電子メールの説明が自動生成されます

Cmd 1



databricks

```
1 spark.conf.get("spark.sql.warehouse.dir")
```

```
Out[1]: 'dbfs:/user/hive/warehouse'
```

デフォルトの場所

```
1 %sql
2 create database sample_db
```

OK

Command took 0.32 seconds -- by akhil.vangala@ at 9/1/2022, 1:07:29 AM on SingleNode

md 4

```
1 %sql describe database sample_db
```

Table Data Profile

	database_description_item	database_description_value
1	Namespace Name	sample_db
2	Comment	
3	Location	dbfs:/user/hive/warehouse/sample_db.db
4	Owner	root

参考までに、これはクラスターのスパーク構成またはセッション構成を使用して変更できます。デフォルトの場所を変更するには、spark.sql.warehouse.dir の場所を変更します。グラフィカル ユーザー インターフェイス、テキスト、アプリケーションの説明が自動生成される

▼ Advanced options databricks

Azure Data Lake Storage credential passthrough 

Enable credential passthrough for user-level data access

Spark config 

```
spark.sql.warehouse.dir dbfs:/tmp/customDBlocation
```

```
1 %sql
2 create database sample_db
```

OK

Command took 1.32 seconds -- by akhil.vangala at 9/1/2022 12:23:45 AM on SingleNode

```
Cmd 3
1 %sql describe database sample_db
```

Table	Data Profile
database_description_item	database_description_value
1 Namespace Name	sample_db
2 Comment	
3 Location	dbfs:/tmp/customDBlocation/sample_db.db
4 Owner	root

最新問題: 75

以下のsparkコマンドは、customerIdとcustomerIdがevent_logデルタテーブルに存在する回数に基づいてサマリーテーブルを作成し、サマリーテーブルに1回限りのマイクロバッチを書き込み、空白を埋めてクエリを完了します。

1. スパーク _____
2. .format("デルタ")
3. .table("イベントログ")

4. .groupBy("顧客ID")
 5. .count()
 6. _____
 7. .format("デルタ")
 8. .outputMode("完全")
 9. .option("checkpointLocation", "/tmp/delta/eventsByCustomer/_checkpoints/")
 10. .トリガー(_____)
 11. .table("ターゲットテーブル")
- A. writeStream、readStream、1 回
B. readStream、writeStream、1 回
C. writeStream、処理時間 = 1 回
D. writeStream、readStream、1 回 = True
E. readStream、writeStream、1 回 = True

Answer: E (メッセージを残す)

説明

答えは、readStream、writeStream、once = True です。

スパーク.readStream

フォーマット("デルタ")

テーブル("イベントログ")

groupBy("顧客ID")

カウント ()

ライトストリーム

フォーマット("デルタ")

出力モード("完全")

option("checkpointLocation", "/tmp/delta/eventsByCustomer/_checkpoints/") トリガー(1 回 = True) table("target_table")

最新問題: 76

既存の差分テーブルに行を追加するために使用される SQL コマンドは次のうちどれですか？

- A. デルタテーブル名に追加
- B. テーブル名に追加
- C. デルタをテーブル名にコピーします
- D. INSERT INTO テーブル名
- E. UPDATE テーブル名

Answer: (解答を表示する)

説明

答えは INSERT INTO table_name です。

Insert は既存のテーブルに行を追加します。これは、従来のデータベースまたはデータウェアハウスに行を追加するのと非常に似ています。

有効な **Databricks-Certified-Professional-Data-Engineer** 問題集は GoShiken.com が提供された合格しやすい Databricks-Certified-Professional-Data-Engineer 試験問題集！

GoShiken.com が最新の **Databricks-Certified-Professional-Data-Engineer** 試験問題集を提供しています。GoShiken.com Databricks-Certified-Professional-Data-Engineer 試験問題は最新で、解答が正確でございます。最新の GoShiken.com Databricks-Certified-Professional-Data-Engineer 問題集をゲットする人はこちら: <https://www.goshiken.com/Databricks/Databricks-Certified-Professional-Data-Engineer-mondaishu.html> (20430%OFF問題集溶と正解付きで 30%w 特別割引コード: **Freepdfdumps**)

最新問題: 77

次の SQL ステートメントのどれを使用してトランザクション テーブルを更新し、テーブルのフラグを Y から N に設定できますか？

- A. トランザクションの変更 SET active_flag = 'N' WHERE active_flag = 'Y'
- B. MERGE トランザクション SET active_flag = 'N' WHERE active_flag = 'Y'
- C. UPDATE トランザクション SET active_flag = 'N' WHERE active_flag = 'Y'
- D. REPLACE トランザクション SET active_flag = 'N' WHERE active_flag = 'Y'

Answer: ([解答を表示する](#))

説明

答えは

UPDATE トランザクション SET active_flag = 'N' WHERE active_flag = 'Y'

Delta Lake はデルタ テーブルの UPDATE ステートメントをサポートしており、更新の一部としてのすべての変更は ACID に準拠しています。

最新問題: 78

Databricks SQL ダッシュボードを構築する場合、次のうち正しいのはどれですか？

- A. ダッシュボードは 1 つのクエリの結果のみを使用できます。
- B. 1 つのクエリ結果で開発できるビジュアライゼーションは 1 つだけです
- C. ダッシュボードは 1 つのスキーマ/データベースにのみ接続できます
- D. 単一のクエリ結果を使用して複数のビジュアライゼーションを開発できます。
- E. ダッシュボードに設定できる更新スケジュールは 1 つだけです

Answer: ([解答を表示する](#))

説明

答えは、単一のクエリ結果を使用して複数のビジュアライゼーションを開発できるということです。

クエリ エディター ペインの [視覚化の追加] タブを使用して、単一のクエリ結果に対する多数の視覚化を行うことができます。

グラフィカル ユーザー インターフェイス、テキスト、アプリケーションの説明が自動生成される



最新問題: 79

データ エンジニアリング チームは、大量の SQL クエリを使用してデータ品質を確認し、ETL ジョブを毎日監視しています。スケジュールを設定し、このプロセスを自動化するには、次のアプローチのうちどれを使用できますか？

- A. ジョブ UI からクエリを 1 日ごとに実行するようにスケジュールできます。
- B. Databricks SQL のクエリのページから 1 日ごとにクエリを更新するようにスケジュールできます。
- C. ジョブ UI から 12 時間ごとにクエリを実行するようにスケジュールできます。
- D. Databricks SQL の SQL エンドポイントのページからクエリを 1 日ごとに更新するようにスケジュールできます。
- E. Databricks SQL の SQL エンドポイントのページから 12 時間ごとにクエリを更新するようにスケジュールできます。

Answer: B ([メッセージを残す](#))

説明

説明

個々のクエリはスケジュールに基づいて更新できます。

スケジュールを設定するには:

1. 「クエリ情報」タブをクリックします。

グラフィカル ユーザー インターフェイス、テキスト、アプリケーション、電子メールの説明が自動生成されます

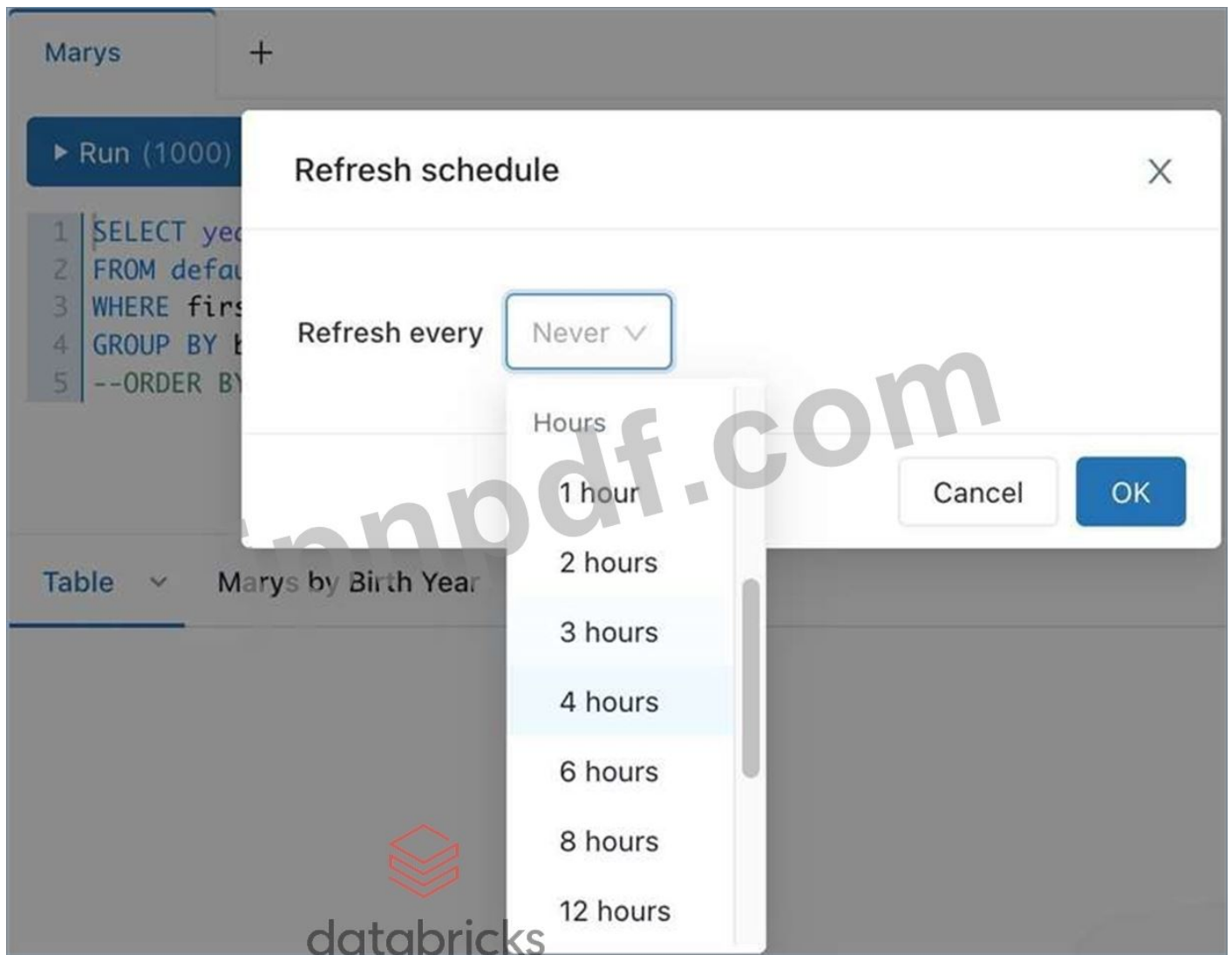
Query info  X

databricks

created an hour ago
updated an hour ago

* [更新スケジュール] の右側にあるリンクをクリックして、スケジュール間隔を指定するピッカーを開きます。

グラフィカル ユーザー インターフェイス、アプリケーションの説明が自動的に生成される



* スケジュールを設定します。

ピッカーがスクロールして次の項目を選択できます。

※ 間隔 :1~30分、1~12時間、1~30日、1~2週間

* 時間。時間セレクターは、間隔が1日より大きく、曜日の選択が1週間より大きい場合にのみピッカーに表示されます。特定の時刻をスケジュールすると、Databricks SQLはコンピューターのタイムゾーンで入力を受け取り、それをUTCに変換します。UTCの特定の時間にクエリを実行したい場合は、ローカルオフセットによってピッカーを調整する必要があります。たとえば、クエリを毎日00:00 UTCに実行したいが、現在のタイムゾーンがPDT (UTC-7)である場合は、ピッカーで17:00を選択する必要があります。

グラフィカルユーザーインターフェース 自動生成される説明

Refresh schedule

Refresh every 1 week

At time 15:35 (21:35 UTC)

On day S M T W T F S

Cancel OK

databricks

* 「OK」をクリックします。

クエリは自動的に実行されます。

スケジュールされたクエリがスケジュールに従って実行されない場合は、クエリが失敗しないように手動でクエリをトリガーする必要があります。ただし、次の点に注意する必要があります。

* 「5分ごと」などの間隔をスケジュールした場合、その間隔は最後に成功した実行から計算されます。クエリを手動で実行する場合、スケジュールされたクエリは、間隔が経過するまで実行されません。

* 時間をスケジュールした場合、Databricks SQL は結果が「古くなる」まで待機します。たとえば、毎週木曜日に更新されるようにクエリを設定し、それを水曜日に手動で実行した場合、木曜日までに結果はまだ「有効」とみなされ、クエリは新たに実行されるようにスケジュールされません。

したがって、たとえば、週単位のスケジュールを設定する場合は、最後のクエリ実行時刻を確認し、スケジュールされたクエリがその実行から1週間経過した後の選択した日に実行されることを期待します。この間はクエリを手動で実行しないようにしてください。

クエリの実行が失敗した場合、Databricks SQL はバックオフ アルゴリズムを使用して再試行します。失敗が多いほど、次の再試行は遠くなります (更新間隔を超える可能性があります)。

追加情報についてはドキュメントを参照してください。

<https://docs.microsoft.com/en-us/azure/databricks/sql/user/queries/schedule-query>

最新問題: 80

デルタ テーブルのクエリに使用できるコマンドは次のうちどれですか?

A. 1.%Python

2.spark.sql("select * from table_name")

B. 1.%sql

2.テーブル名から * を選択します

C. A と B の両方

(正しい)

D. 1.%Python

2.execute.sql("テーブルから * を選択")

E. 1.%Python

2.delta.sql("テーブルから * を選択")

Answer: ([解答を表示する](#))

説明

答えは選択肢AとBの両方です

Spark にはexecute.sql または delta.sql というコマンドがないため、オプション C と D は正しくありません。

最新問題: 81

多次元データセットを投影すると、どのベクトルに最大の分散が生じますか？

A. 第一主成分

B. 第一固有ベクトル

C. 回答するには情報が不十分です

D. 第 2 固有ベクトル

E. 第 2 主成分

Answer: A ([メッセージを残す](#))

説明

主成分分析 (PCA) に基づく方法では、次の投影に従って特徴を評価します。

初期次元の相関行列の最大固有ベクトル、フィッシャーの線形に基づく方法

判別分析で評価します。判別式の成分の大きさに応じて

ベクター。

最初の主成分は、定義上、データ内の最大の分散に対応します。投影すると

データを最初の主成分線上に投影すると、データは上に投影した場合よりも広がります (分散が高くなります)。

他の主成分を含む他の行。

最新問題: 82

あなたは現在、特定のノートブックで SQL と Python を使用する必要があるプロジェクトに取り組んでいます。どのようなアプローチをとりますか？

A. 2 つの別々のノートブックを作成します。1 つは SQL 用、もう 1 つは Python 用です。

B. 1 つのノートブックで複数の言語をサポートできます。マジック コマンドを使用して 2 つの言語を切り替えることができます。

C. Python には汎用クラスター、SQL には SQL エンドポイントを使用する

D. ジョブ クラスターを使用して Python と SQL エンドポイントを実行します。

Answer: ([解答を表示する](#))

説明

答えは、1つのノートブックで複数の言語をサポートできるため、マジック コマンドを使用して2つの言語を切り替えることができます。

同じノートブック内で %sql および %python マジック コマンドを使用する

最新問題: 83

SQL DDL コマンド CREATE TABLE sales (id int, Units int) を発行したときに作成されるテーブルのタイプは何ですか?

A. 場所が見つからないため、クエリは失敗します。

B. 形式が欠落しているため、クエリは失敗します。

C. マネージドデルタテーブル

D. 外部テーブル

E. マネージド Parquet テーブル

Answer: C ([メッセージを残す](#))

説明

答えはマネージドデルタテーブルです

Location キーワードを指定せずにテーブルが作成されると、そのテーブルは管理テーブルとみなされます。デフォルトでは、すべての管理テーブル DELTA テーブル 構文 CREATE TABLE table_name (column column_data_type...)

最新問題: 84

DBFSストレージにある外部CSVファイルにクエリを実行し、SQLを使用してデータを検査する最良の方法は何ですか?

A. SELECT * FROM 'dbfs:/location/csv_files/' FORMAT = 'CSV'

B. CSVを選択します。* dbfs:/location/csv_files/」から

C. CSV から * を選択します。dbfs:/場所/csv_files/」

D. 外部ファイルを直接クエリすることはできません。最初に COPY INTO を使用してデータをテーブルにロードします。

E. SELECT * FROM 'dbfs:/location/csv_files/' USING CSV

Answer: C ([メッセージを残す](#))

説明

答えは、SELECT * FROM CSV です。dbfs:/場所/csv_files/」

以下の構文を使用して、ストレージに保存されている外部ファイルをクエリできます

SELECT * FROM 形式。`/Location`

形式 - CSV、JSON、PARQUET、TEXT

最新問題: 85

レポート用の概要データ セットを生成するノートブックを作成しました。ノートブックはジョブ クラスターを使用してスケジュールされましたが、クラスターの起動に 8 分かかることに気付きました。クラスターをタイムリーに起動するにはどのような機能を使用すれば、ジョブを効率化できるでしょうか。すぐに実行できますか？

- A. 実際のジョブに先立って実行する追加のジョブをセットアップして、クラスターが 2 番目のジョブの開始を実行するようにします。
- B. Databricks クラスター プール機能を使用して起動時間を短縮します。
- C. Databricks Standard Edition の代わりに Databricks Premium Edition を使用します。
- D. クラスター UI ページにクラスターを固定して、ジョブで常にクラスターを使用できるようにします。
- E. 自動終了を無効にして、クラスターが常に実行されるようにします。

Answer: B ([メッセージを残す](#))

説明

クラスター プールを使用すると、新しいジョブ クラスターが作成されるたびに VM をプールから取得するときに、事前に VM を予約できます。注: VM がクラスターによる使用を待機している場合、発生するコストは Azure のみです。Databricks の実行時間コストは、VM がクラスターに割り当てられた場合にのみ請求されます。

ここでは、プールを設定し、いくつかのベスト プラクティスに従う方法のデモを示します。

グラフィカル ユーザー インターフェイス、自動生成されるテキスト説明



最新問題: 86

ジュニア データ エンジニアは、Spark がデータとデータの両方を管理する Spark SQL テーブル `my_table` を作成する必要があります。

メタデータ。メタデータとデータも Databricks ファイルシステム (DBFS) に保存する必要があります。

上級データ エンジニアがジュニア データ エンジニアと共有すべきコマンドは次のどれですか。このタスクを完了しますか？

- A. 1. 管理テーブル `my_table` (ID STRING、値 STRING) USING を作成します。

2. org.apache.spark.sql.parquet オプション (PATH "storage-path");
- B. 1. DBFS を使用してテーブル my_table (ID STRING、値 STRING) を作成します。
- C. 1. 管理テーブル my_table を作成します (ID STRING、値 STRING);
- D. 1. CREATE TABLE my_table (ID STRING、値 STRING);
- E. 1. CREATE TABLE my_table (id STRING, value STRING) USING
2. org.apache.spark.sql.parquet オプション (PATH "storage-path")

Answer: D ([メッセージを残す](#))

最新問題: 87

入力パラメータ 1、3 を使用して以下の関数を実行した場合の出力は次のとおりです。

1. def check_input(x,y):

2. x < y の場合:

3. x = x+1

4. x < y の場合:

5. x = x+1

6. x < y の場合:

7. x = x+1

8. x を返す

check_input(1,3)

A. 2

B. 3

(正しい)

C. 5

D. 4

E. 1

Answer: B ([メッセージを残す](#))

最新問題: 88

あなたは、e コマース プラットフォームからの構造化データと、Web サイトのトラフィックやアプリストアからの非構造化データを保存する分析を設計しています。このデータを保存する場所にはどのようにアプローチしますか？

A. 構造化データには従来のデータ ウェアハウスを使用し、非構造化データにはデータ レイクハウスを使用します。

B. データ レイクハウスは非構造化データのみを保存できますが、スキーマを強制することはできません

C. データ レイクハウスは構造化データと非構造化データを保存でき、スキーマを適用できます。

D. 従来のデータ ウェアハウスは、構造化データの保存とスキーマの適用に適しています。

Answer: (解答を表示する)

説明

答えは、データ レイクハウスは構造化データと非構造化データを保存でき、スキーマを強制できるということです。- Databricks ブログ グラフィカル ユーザー インターフェイス、テキスト、アプリケーションの説明が自動的に生成される

A lakehouse has the following key features:

- **Transaction support:** In an enterprise lakehouse many data pipelines will often be reading and writing data concurrently. Support for ACID transactions ensures consistency as multiple parties concurrently read or write data, typically using SQL.
- **Schema enforcement and governance:** The Lakehouse should have a way to support schema enforcement and evolution, supporting DW schema architectures such as star/snowflake-schemas. The system should be able to reason about data integrity, and it should have robust governance and auditing mechanisms.
- **BI support:** Lakehouses enable using BI tools directly on the source data. This reduces staleness and improves recency, reduces latency, and lowers the cost of having to operationalize two copies of the data in both a data lake and a warehouse.
- **Storage is decoupled from compute:** In practice this means storage and compute use separate clusters, thus these systems are able to scale to many more concurrent users and larger data sizes. Some modern data warehouses also have this property.
- **Openness:** The storage formats they use are open and standardized, such as Parquet, and they provide an API so a variety of tools and engines, including machine learning and Python/R libraries, can efficiently access the data **directly**.
- **Support for diverse data types ranging from unstructured to structured data:** The lakehouse can be used to store, refine, analyze, and access data types needed for many new data applications, including images, video, audio, semi-structured data, and text.
- **Support for diverse workloads:** including data science, machine learning, and SQL and analytics. Multiple tools might be needed to support all these workloads but they all rely on the same data repository.
- **End-to-end streaming:** Real-time reports are the norm in many enterprises. Support for streaming eliminates the need for separate systems dedicated to serving real-time data applications.

最新問題: 89

「mnt/delta/bronze」の場所を使用して、bronze という名前のスキーマを作成し、作成する前にスキーマが存在するかどうかを確認します。

- A. ブロンズの場所 「mnt/delta/bronze」が存在しない場合はスキーマを作成します
- B. LOCATION 'mnt/delta/bronze' が存在しない場合は、スキーマ ブロンズを作成します。
- C. IS_SCHEMA('bronze') の場合: スキーマ ブロンズ LOCATION 'mnt/delta/bronze' を作成します。
- D. スキーマの作成はメタストアでは使用できません。Unity カタログ UI でのみ実行できます。
- E. データベースがないとスキーマを作成できません

Answer: A (メッセージを残す)

説明

<https://docs.databricks.com/sql/language-manual/sql-ref-syntax-ddl-create-schema.html>

1.スキーマの作成 [存在しない場合] スキーマ名 [LOCATION スキーマディレクトリ]

最新問題: 90

KPI インジケーターがしきい値を超えて増加するたびに電子メールで通知するアラートを設定するように求められます。また、チームは、アラート電子メール通知に実際の値を含めるように求められます。

- A. ノートブックと Python コードを使用して毎分実行し、Python 変数を使用して情報をキャプチャし、電子メールで送信します。
- B. アラートを設定しますが、デフォルトのテンプレートを使用して電子メールの件名のメッセージを通知します。
- C. アラートを設定しますが、カスタム テンプレートを使用して電子メールの件名のメッセージを通知します。
- D. 代わりに Webhook 宛先を使用して、アラート メッセージをカスタマイズできるようにします。
- E. カスタム電子メールフックを使用してメッセージをカスタマイズします

Answer: C (メッセージを残す)

説明

アラートのサポート カスタム テンプレートでは、変数を使用してデフォルト メッセージをカスタマイズし、KPI の現在値をしきい値と比較するクエリを設定し、変数 QUERY_RESULT_VALUE を使用して電子メール通知に値を表示することがサポートされています。これは単純なアラートです。カスタム テンプレートの変数を使用して、これらの値を電子メール通知メッセージに表示します。アラートが起動されると、これらの変数は実際の値に置き換えられます。

カスタムテンプレートを使用したアラート

グラフィカル ユーザー インターフェイス、アプリケーションの説明が自動的に生成される

Query: sample query

Value column	Condition	Threshold
utilization_kpi	>	75

Top row value is 78. Only the first row of results is evaluated, so consider adding an aggregation or sorting your query.

When triggered, send notification: Just once

Template: Use custom template

Subject / Body: Preview

Current Utilization : {{QUERY_RESULT_VALUE}}

Utilization is {{QUERY_RESULT_VALUE}} more than the {{ALERT_THRESHOLD}}

Formatting guide

プレビューを有効にすると、変数を置換したときにアラートがどのように表示されるかを確認できます。

グラフィカル ユーザー インターフェイス、テキスト、アプリケーション、電子メールの説明が自動生成されます

Subject / Body: Preview

Current Utilization : 78

Utilization is 78 more than the 75

Formatting guide

以下は、カスタム テンプレートで使用できる追加のテンプレート変数です。

アラート | AWS 上のデータブリック

グラフィカル ユーザー インターフェイス、テキスト、アプリケーション、電子メールの説明が自動生成されます

5. In the **Template** drop-down, choose a template:

- **Use default template:** Alert notification is a message with links to the Alert configuration screen and the Query screen.
- **Use custom template:** Alert notification includes more specific information about the alert.

a. A box displays, consisting of input fields for subject and body. Any static content is valid, and you can incorporate built-in template variables:

- **ALERT_STATUS:** The evaluated alert status (string).
- **ALERT_CONDITION:** The alert condition operator (string).
- **ALERT_THRESHOLD:** The alert threshold (string or number).
- **ALERT_NAME:** The alert name (string).
- **ALERT_URL:** The alert page URL (string).
- **QUERY_NAME:** The associated query name (string).
- **QUERY_URL:** The associated query page URL (string).
- **QUERY_RESULT_VALUE:** The query result value (string or number).
- **QUERY_RESULT_ROWS:** The query result rows (value array).
- **QUERY_RESULT_COLS:** The query result columns (string array).

An example subject, for instance, could be: Alert, "{{ALERT_NAME}}" changed status to {{ALERT_STATUS}}.

b. Click the **Preview** toggle button to preview the rendered result.

最新問題: 91

メダリオン構造における銀層と金層の主な違いは何ですか？

- A. シルバーは ETL を実行するように最適化され、ゴールドはクエリ パフォーマンスが最適化されます。
- B. Gold は ETL の実行に最適化され、Silver はクエリのパフォーマンスに最適化されます。
- C. シルバーはブロンズのコピー、ゴールドはシルバーのコピー
- D. シルバーはデルタレイクに保存され、ゴールドはメモリに保存されます。
- E. シルバーには集約されたデータが含まれる場合があります、ゴールドには元のデータの粒度が保存される場合があります。

Answer: A (メッセージを残す)

説明

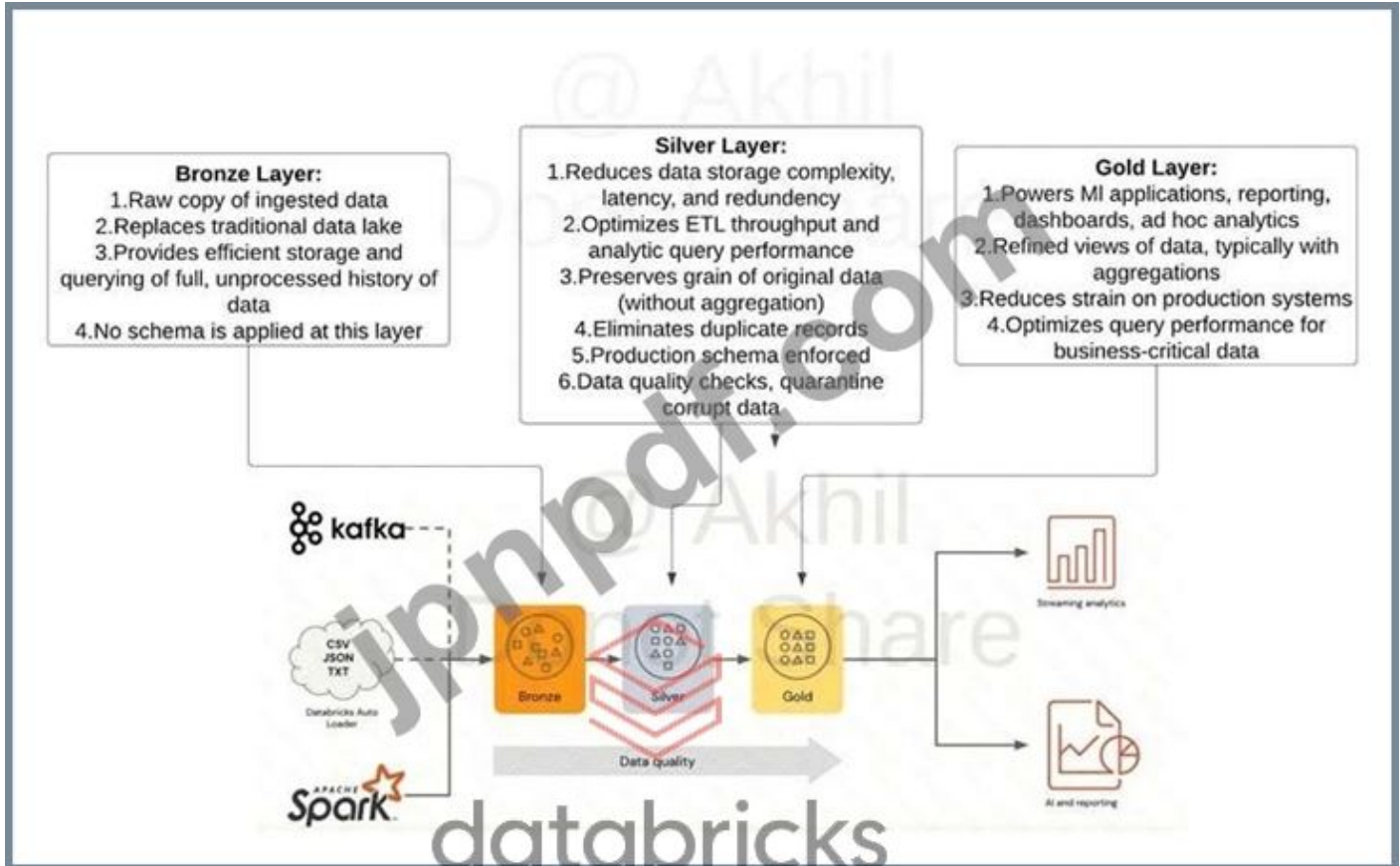
メダリオン アーキテクチャ - Databricks

ゴールドレイヤー:

1. ML アプリケーション、レポート、ダッシュボード、アドホック分析を強化します
2. データの洗練されたビュー (通常は集計を使用)
3. 実稼働システムの負担を軽減します
4. ビジネスクリティカルなデータのクエリパフォーマンスを最適化します。

試験の焦点: 下の画像を確認して、メダリオン建築における各層 (ブロンズ、シルバー、ゴールド) の役割を理解してください。各層とその目的を対象としたさまざまな質問が表示されます。Udemy 内の一部の人が私のコンテンツをコピーしているため、ウォーターマークを追加する必要がありました。

低信頼度で自動的に生成された家の説明図



有効な **Databricks-Certified-Professional-Data-Engineer** 問題集は GoShiken.com が提供された合格しやすい Databricks-Certified-Professional-Data-Engineer 試験問題集！

GoShiken.com が最新の **Databricks-Certified-Professional-Data-Engineer** 試験問題集を提供しています。GoShiken.com Databricks-Certified-Professional-Data-Engineer 試験問題は最新で、解答が正確でございます。最新の GoShiken.com Databricks-Certified-Professional-Data-Engineer 問題集をゲットする人はこちら: <https://www.goshiken.com/Databricks/Databricks-Certified-Professional-Data-Engineer-mondaishu.html> (20430%OFF問題集溶と正解付きで 30%w 特別割引コード: **Freepdfdumps**)

最新問題: 92

このテーブルに対するクエリ `SELECT COUNT (DISTINCT *) FROM user` の予想される出力は何ですか？

userId	username	email
1	john.smith	john.smith@com
2	NULL	david.clear@com
3	kevin.smith	kevin.smith@com

- A. 3
- B. 2
- (正しい)
- C. 1
- D. 0
- E. NULL

Answer: B (メッセージを残す)

説明

答えは2です、

Count(DISTINCT *) は、NULL 値を持つ列を含む行を削除します。

最新問題: 93

コマンド VACUUM を実行すると、トランザクションは 0 時間保持されますか? このコマンドの結果はどうなるでしょうか?

- A. コマンドは成功しますが、データは削除されません。
- B. アクティブなトランザクションが実行されている場合、コマンドは失敗します。
- C. コマンドは失敗します。retentionDurationcheck が有効になっているとコマンドを実行できません。
- D. コマンドは成功しますが、履歴データは削除されます。
- E. コマンドは正常に実行され、テーブル内のすべてのデータが圧縮されます。

Answer: C (メッセージを残す)

説明

答えは、

コマンドは失敗します。retentionDurationcheck を有効にしてコマンドを実行することはできません。

1.VACUUM [[データベース名.]テーブル名 | パス] [保持時間数] [ドライラン]

*デルタ テーブルに関連付けられたディレクトリを再帰的にバキュームし、テーブルのトランザクション ログの最新の状態ではなくなり、保持しきい値より古いデータ ファイルを削除します。デフォルトは 7 日です。

*このチェックが有効になっている理由は、DELTA が履歴の意図しない削除を防止しようとしているためです。また、重要な点の 1 つは、保持期間が 0 時間の場合、データ損失の可能性があることです (以下の KB を参照) VACUUM のドキュメント <https://docs.delta.io/latest/delta-utility.html>
<https://kb.databricks.com/delta/data-missing-vacuum-Parallel-write.html>

Valid Databricks-Certified-Professional-Data-Engineer Dumps shared by GoShiken.com for Helping Passing Databricks-Certified-Professional-Data-Engineer Exam! GoShiken.com now offer the **newest Databricks-Certified-Professional-Data-Engineer exam dumps**, the GoShiken.com Databricks-Certified-Professional-Data-Engineer exam **questions have been updated** and **answers have been corrected** get the **newest** GoShiken.com Databricks-Certified-Professional-Data-Engineer dumps with Test Engine here:
<https://www.goshiken.com/Databricks/Databricks-Certified-Professional-Data-Engineer-mondaishu.html> (**204** Q&As Dumps, **30%OFF** Special Discount: **Freepdfdumps**)