

Lead2passExam

> Contact Us Login / Register Search...

Lead2passExam

HOME

ALL VENDORS

★ GUARANTEE

? FAQ

TESTIMONIALS

CART (1)

Pass Your Next Certification Exam Fast!

Everything you need to prepare, learn & pass your certification exam easily.
365 days free updates. First attempt guaranteed success.



Select a vendor...

Select an test...

Your email address

Free Download Demo

Top Certifications

- ▶ IBM Cognos ▶ Linux Essentials ▶ Magento Certified Developer Plus ▶ BCS Certification
- ▶ Citrix NetScaler ▶ Nokia Networks Certification ▶ Solutions Expert
- ▶ VCAP6-DCV Deployment ▶ Oracle Sales Cloud 2016 Certified ▶ Oracle Service Cloud
- ▶ CCP-N ▶ IBM Certified Mobile System Administrator ▶ Windows 7 ▶ APC Certification
- ▶ HPE Sales Certified

Top Vendors

- ▶ Logical Operations ▶ TIA ▶ Pegasystems ▶ IISFA ▶ Mile2 ▶ 3COM ▶ Altiris ▶ IIA
- ▶ AccessData ▶ Avaya ▶ BACB ▶ Nokia ▶ RAPS ▶ McAfee ▶ Professional Tests
- ▶ Mile2-Security ▶ CIPS ▶ Legato ▶ ASQ ▶ QlikView ▶ NSCA ▶ PSAT ▶ HRCI
- ▶ WorldatWork ▶ Guidance Software

What Client's Say

“ Passed the exam yesterday, but 10 questions new not came from this dump. every other questions are same. Totally valid. ”



Roy
★★★★★

“ This is still valid. Passed today with 80%. looked like 3-4 new questions. Many thanks! Good braindumps ”



Vic
★★★★★

<http://www.lead2passexam.com/>

Available Exam Cram and Valid Dumps - Lead2Pass Exam

Exam : **Databricks-Certified-Data-Engineer-Associate-JPN**

Title : Databricks Certified Data Engineer Associate Exam (Databricks-Certified-Data-Engineer-Associate 日本語版)

Vendor : Databricks

Version : DEMO

QUESTION NO: 1

小規模なアドホックなPythonスクリプトを高頻度で実行し、これらのクエリの実行が完了したらずに処理を終了させる必要があるシナリオでは、どのコンピューティングオプションを選択すべきでしょうか？

- A. 汎用クラスター
- B. ジョブクラスター
- C. サーバーレスコンピューティング
- D. SQLウェアハウス

Answer: C

Explanation:

Serverless Compute is ideal for small-scale, frequent, ad-hoc Python scripts because it provisions automatically, scales quickly, and terminates when queries finish, minimizing operational overhead and costs.

QUESTION NO: 2

データエンジニアが、PySparkを使用してデータセットを処理するDatabricks上のPythonノートブックをデバッグしています。ノートブックは、DataFrame変換中にエラーが発生して失敗します。エンジニアは、入力DataFrameや中間結果などの変数の状態を調べて、エラーが発生した箇所を特定したいと考えています。ノートブックをデバッグし、DataFrameなどの変数の値を検査するには、エンジニアはどのツールを使用すべきでしょうか？

- A. Databricks CLIを使用して、ドライバーログをダウンロードして詳細なエラーメッセージを分析します。
- B. Python Notebook 対話型デバッガーを使用してブレークポイントを設定し、変数の値をリアルタイムで検査します。
- C. Ganglia UI を使用してクラスターのリソース使用状況を監視し、ハードウェアの問題を特定します。
- D. Spark UI を使用して実行プランを分析し、ジョブが失敗したステージを特定します。

Answer: B

Explanation:

The Python Notebook Interactive Debugger in Databricks allows setting breakpoints and inspecting variable values, including DataFrames, in real time. This makes it the correct tool for debugging transformation errors in a PySpark notebook.

QUESTION NO: 3

データエンジニアが以下のコマンドを使用して新しいデータベースを作成しました。
customer360 データベースが存在しない場合は作成します。
customer360データベースは、以下のどの場所に配置されますか？

- A. dbfs:/user/hive/database/customer360
- B. dbfs:/user/hive/warehouse
- C. dbfs:/user/hive/customer360
- D.正しい回答を判断するには、さらに情報が必要です
- E. dbfs:/user/hive/database

Answer: B

QUESTION NO: 4

データエンジニアに新しいデータレコードが渡されました。

id STRING = 'a1'

ランク整数 = 6

評価 FLOAT = 9.4

次のSQLコマンドのうち、既存のDeltaテーブルmy_tableに新しいレコードを追加するために使用できるものはどれですか？

- A. UPDATE my_table VALUES ('a1', 6, 9.4)
- B. INSERT VALUES ('a1' , 6, 9.4) INTO my_table
- C. my_table UNION VALUES ('a1', 6, 9.4)
- D. INSERT INTO my_table VALUES ('a1', 6, 9.4)
- E. UPDATE VALUES ('a1', 6, 9.4) my_table

Answer: D

QUESTION NO: 5

データエンジニアは、各タスクが前のタスクの正常な完了に依存するワークフローで、複数のタスクをスケジュールしたいと考えています。このワークフローは、再試行と監視をサポートする必要があります。

Databricksのどの機能を使用すべきですか？

- A. デルタ湖
- B. Databricksの求人情報
- C. DBFS
- D. Spark UI

Answer: B

QUESTION NO: 6

データエンジニアがDatabricksでUnityカタログを設定しており、特定のスキーマ内のさまざまなデータオブジェクトに対する権限の付与と取り消しを行う権限を持つが、スキーマまたはそのオブジェクトに対する読み取り/書き込みアクセス権を持たないユーザーにロールを割り当てる必要があります。データエンジニアはこのユーザーにどのロールを割り当てるべきですか？

- A. テーブルオーナー
- B. カタログ所有者
- C. スキーマ所有者
- D. スキーマに対してカタログ/スキーマ権限を使用する

Answer: C

Explanation:

The Schema Owner role in Unity Catalog allows a user to grant and revoke privileges on data objects within the schema without requiring read/write access to the schema's data itself, fitting the requirement.

QUESTION NO: 7

Python の使用に慣れていないデータ エンジニアは、2 つの整数を加算してその合計を返す Python 関数を作成する必要がありますか？

データ エンジニアがこのタスクを完了するために使用できるコードブロックは次のどれですか。

A.

```
function add_integers(x, y):  
    return x + y
```

B.

```
function add_integers(x, y):  
    x + y
```

C.

```
def add_integers(x, y):  
    print(x + y)
```

D.

```
def add_integers(x, y):  
    return x + y
```

E.

```
def add_integers(x, y):  
    x + y
```

Answer: D

QUESTION NO: 8

データエンジニアが、クラウドストレージに届いたJSONファイルを読み込むストリーミングパイプラインを構築しています。このパイプラインは、新しいファイルが到着するたびに、手動操作なしで自動的に処理する必要があります。このユースケースに適したDatabricksの機能はどれですか？

A. 構造化ストリーミング自動ローダー

B. デルタタイムトラベル

C. Spark キャッシュ

D. ブロードキャスト参加

Answer: A

QUESTION NO: 9

次のSQLキーワードのうち、テーブルをロング形式からワイド形式に変換するために使用できるものはどれですか？

A. 変換

B. ピボット

C. 合計

- D. 変換
 - E. WHERE
- Answer: B**

QUESTION NO: 10

あるデータエンジニアはDatabricksにPythonノートブックを持っていますが、セル内の特定のタスクを実行するためにSQLを使用する必要があります。他のすべてのセルでは、それらのセルに変更を加えることなくPythonを使用できるようにしたいと考えています。

データ エンジニアが Python ノートブックのセル内で SQL を使用する方法を説明しているのは次のどれですか。

- A. PythonノートブックではSQLは使用できません
- B. セルをDatabricksクラスタではなくSQLエンドポイントに接続できます
- C. セル内にSQL構文を記述するだけで済みます
- D. セルの最初の行に%sqlを追加できます
- E. ノートブックのデフォルト言語をSQLに変更できます

Answer: D

QUESTION NO: 11

ある組織は、MySQL、Amazon Redshift、Google BigQueryなど、複数の外部システムにデータを保存しています。データエンジニアは、Databricksに直接取り込むことなく分析を実行し、統一されたガバナンスを確保し、データの重複を最小限に抑えたいと考えています。Databricksのどの機能を使えば、これらの外部データソースにクエリを実行しながら、一元化されたガバナンスを維持できますか？

- A. デルタ湖
- B. レイクハウス連盟
- C. MLflow
- D. Databricks Connect

Answer: B

Explanation:

Lakehouse Federation allows Databricks to query external data sources like MySQL, Redshift, and BigQuery without moving or duplicating data, while maintaining centralized governance through Unity Catalog.

QUESTION NO: 12

データエンジニアは、同じ Spark セッション内で使用する中間データを一時ビューに保存したいと考えています。セッション終了後、データは保持されないようにする必要があります。どの Spark SQL コマンドを使用すればよいでしょうか？

- A. テーブルを作成します
- B. ビューを作成
- C. 一時ビューを作成
- D. データベースを作成

Answer: C

QUESTION NO: 13

次のコマンドのうち、member_id 列の null 値の数を返すものはどれですか。

- A. my_table から count(member_id) を選択します。
- B. my_table から count(member_id) - count_null(member_id) を選択します。
- C. count_if(member_id IS NULL) を my_table から選択します。
- D. my_table から null(member_id) を選択します。
- E. count_null(member_id) を my_table から選択します。

Answer: C

QUESTION NO: 14

データエンジニアは、毎朝仕事を始める前に実行される単一タスクのジョブを持っています。上流のデータの問題を特定した後、元のタスクの前に新しいノートブックを実行するための別のタスクを設定する必要があります。

データ

エンジニアが新しいタスクを設定するために使用できるアプローチは次のどれですか。

- A. 既存のジョブ内の既存のタスクを複製し、更新して新しいノートブックを実行できます。
- B. 既存のジョブに新しいタスクを作成し、それを元のタスクの依存関係として追加できます。
- C. 既存のジョブに新しいタスクを作成し、元のタスクを新しいタスクの依存関係として追加できます。
- D. 新しいジョブを最初から作成し、両方のタスクを追加して同時に実行することができます。
- E. 既存のタスクを新しいジョブに複製し、それを編集して新しいノートブックを実行できます。

Answer: B

QUESTION NO: 15

データエンジニアは、複数のワークスペースにわたるデータ資産へのアクセスを制御し、一元化されたガバナンスポリシーを適用する必要があります。組織は、テーブル、スキーマ、カタログに対してきめ細かなアクセス制御を求めています。Databricksのどの機能がこの要件を満たしていますか？

- A. DBFS
- B. Unityカタログ
- C. デルタキャッシュ
- D. MLflow

Answer: B

QUESTION NO: 16

次のデータレイクハウス機能のうち、従来のデータレイクよりもデータ品質が向上するのはどれですか。

- A. データレイクハウスは、構造化データと非構造化データのストレージ

ソリューションを提供します。

- B. データレイクハウスは、ACID 準拠のトランザクションをサポートします。
- C. データレイクハウスでは、SQL クエリを使用してデータを検査できます。
- D. データレイクハウスはオープン形式でデータを保存します。
- E. データレイクハウスは、機械学習と人工知能のワークロードを可能にします。

Answer: B

QUESTION NO: 17

データエンジニアは、DatabricksのDelta Live

Tables (DLT) を使用して、顧客データを取り込むためのシンプルなデータパイプラインを構築しています。生の顧客データは、JSON形式でクラウドストレージに保存されています。タスクは、生のJSONデータを読み取り、Deltaテーブルに書き込んでさらに処理するDLTパイプラインを作成することです。

どのコードスニペットが生のJSONデータを正しく取り込み、DLTを使用してDeltaテーブルを作成しますか？

A.

```
import dlt
```

```
@dlt.table
```

```
def raw_customers():
```

```
    return spark.read.format("csv").load("s3://my-bucket/raw-customers/")
```

B.

```
import dlt
```

```
@dlt.view
```

```
def raw_customers():
```

```
    return spark.format.json("s3://my-bucket/raw-customers/")
```

C.

```
import dlt
```

```
@dlt.table
```

```
def raw_customers():
```

```
    return spark.read.json("s3://my-bucket/raw-customers/")
```

D.

```
import dlt

@dlt.table
def raw_customers():
    return spark.read.format("parquet").load("s3://my-bucket/raw-customers/")
```

Answer: C

Explanation:

The correct approach is to use `@dlt.table` to define a Delta Live Table and read the raw JSON data with `spark.read.json("s3://...")`. This properly ingests the JSON source data and materializes it as a managed Delta table in DLT.

QUESTION NO: 18

データエンジニアが、異なる部門のチームが共有データにアクセスする必要がある複数チームプロジェクト向けに、デルタ共有を設定しています。データエンジニアはUnityカタログメタストアの作成に成功し、現在デルタ共有の設定を行っています。目標は、内部チームが共有データに完全なアクセス権限でアクセスできるようにする一方で、外部パートナーはデータを読み取ることのみを許可することです。データエンジニアは、共有を正しく設定するためにどのような操作を行うべきでしょうか？

A. Delta

Shareを作成し、社内チームと外部パートナー向けに安全なアクセスURLを設定し、そのURLを配布して共有データへのアクセスを提供します。

B.

デルタ共有を作成し、内部チームのテーブルとビューを追加し、外部パートナーと内部チームの両方に読み取り/書き込み権限を割り当てます。

C. Delta Share を通じて外部パートナーに READ 権限を、内部チームに READ/WRITE 権限を割り当て、正しいテーブルとビューが共有されていることを確認します。

D. Delta Share

を介して、外部パートナーには読み取り権限を、内部チームには読み取り/書き込み権限を付与します。

Answer: C

Explanation:

In Delta Sharing, permissions must be managed separately for each recipient. To achieve the goal, the data engineer should create the Delta Share, add the appropriate tables/views, assign READ permissions to external partners, and assign READ/WRITE permissions to internal teams.

This ensures external users only consume data, while internal teams can both read and modify it.

QUESTION NO: 19

あるプロジェクトに新しいデータエンジニアリングチームが配属されました。チームは、既存のテーブルを確認するためにデータベースの顧客にアクセスする必要があります。チームには独自のグループチームがあります。

新しいチームにデータベース全体に対する必要な権限を付与するには、次のコマンドのどれ

を使用できますか？

- A. カタログ顧客のビューをチームに付与します。
- B. データベース 顧客に対する CREATE 権限を team に付与します。
- C. カタログ チームの使用権限を顧客に付与します。
- D. データベース team に対する CREATE 権限を customer に付与します。
- E. データベース customers の使用権限を team に付与します。

Answer: E

QUESTION NO: 20

データエンジニアは、ETLプロセスの統合テストを開発し、バージョン管理されたパッケージ化されたワークフローを外部ジョブスケジューラを使用して本番環境にデプロイする必要があります。この作業には、データエンジニアはどのツールを使用すべきでしょうか？

- A. Databricks Connect
- B. Databricksアセットバンドル
- C. Databricks コマンドラインインターフェース
- D. Databricksソフトウェア開発キット

Answer: B

Explanation:

Databricks Asset Bundles enable packaging of workflows with version control, allowing integration testing and deployment into production while supporting orchestration by external job schedulers.

QUESTION NO: 21

データエンジニアには、パイプラインの迅速な反復開発に加え、データ取り込みの失敗後には確実にロールバックを行い、規制遵守のための監査証跡を確保し、AIとBIワークロードの両方に対して単一の信頼できる情報源への一貫したアクセスを提供する能力が求められます。

データエンジニアは、ニーズを満たすためにどのような戦略を採用すべきでしょうか？

- A. Delta Lake ACID トランザクションとタイムトラベル。一貫したアクセスと履歴のために Unity Catalog によって管理されます。
- B. 手動ファイルバージョン管理とロールバックのための毎晩のコピーを備えたDBFS CSVストレージ。
- C. 監査証跡および BI 配布用の一時的なインメモリ DataFrame。
- D. クラウドオブジェクトストレージのみで、復旧とガバナンスのためにアドホックSQLクエリを使用します。

Answer: A

Explanation:

Delta Lake provides ACID transactions and time travel for safe pipeline iteration and reliable rollbacks, while Unity Catalog ensures governed, auditable access with lineage and a single source of truth shared consistently across AI and BI workloads.

QUESTION NO: 22

データエンジニアは、数十億件のレコードを格納するDeltaテーブルのクエリパフォーマンスを最適化したいと考えている。

クエリでは、transaction_date列でフィルタリングを行うことがよくあります。クエリ実行中にスキャンされるデータ量を削減し、データを物理的に整理するには、Delta Lakeのどの機能を使用すればよいでしょうか？

- A. Zオーダー
- B. 真空
- C. キャッシュ
- D. ブロードキャスト参加

Answer: A

QUESTION NO: 23

データエンジニアがUnityカタログ内のサンドボックススキーマを廃止しようとしています。一部のテーブルは一時的なステージング出力であり、完全に削除しても安全ですが、いくつかのテーブルはDatabricks外部のダウンストリームジョブで使用される共有クラウドストレージを参照しています。エンジニアはカタログオブジェクトをクリーンアップする際に、共有ファイルを削除しないように注意する必要があります。

Unityカタログは、管理対象テーブルと外部テーブルを削除した場合、どのように動作しますか？

- A. すべてのテーブルを削除します。Databricksは、管理対象テーブルと外部テーブルの両方のメタデータのみを削除します。
- B. 一時的な管理対象テーブルと外部テーブルを削除します。両方のファイルは7日間保持されます。
- C. 管理対象ステージング テーブルを DROP してデータとメタデータを削除し、外部テーブルを DROP してメタデータのみを削除します
- D. 外部テーブルを最初に削除してファイルを削除し、次に管理対象テーブルを削除してリカバリのためにファイルを保持します。

Answer: C

Explanation:

In Unity Catalog, dropping a managed table deletes both the table metadata and the underlying data files, while dropping an external table removes only the catalog metadata and leaves the referenced cloud storage data intact.