Google.Associate-Data-Practitioner.v2026-01-29.q50

試験コード：	Associate-Data-Practitioner
試験名称：	Google Cloud Associate Data Practitioner
認証ベンダー：	Google
無料問題の数：	50
バージョン：	v2026-01-29
ページの閲覧量：	108
問題集の閲覧量：	736

https://www.jpnshiken.com/shiken/Google.Associate-Data-Practitioner.v2026-01-29.q50.html

質問: **1**
組織では、IoTイベントデータをPub/Subトピックに送信しています。サブスクライバーアプリケーションは、メッセージを読み取り、変換してからデータウェアハウスに保存します。トピックへのデータ書き込みが特に集中する時間帯に、サブスクライバーアプリケーションが期限内にメッセージの確認応答を行っていないことに気づきました。このようなアクティビティの急増に対応し、メッセージの処理を継続するには、パイプラインを変更する必要があります。どうすればよいでしょうか？
A. 確認応答があるまでメッセージを再試行します。
B 加入者に対してフロー制御を実装する
B. 未確認メッセージをデッドレタートピックに転送します。
C. 最後に確認されたメッセージまで戻ります。
正解: (正解を表示します)
サブスクライバーにフロー制御を実装することで、サブスクライバーアプリケーションは、アクティビティの急増時にメッセージのプルと処理の速度を制御することで、メッセージ処理を管理できます。これにより、サブスクライバーへの過負荷を防ぎ、メッセージが期限内に確実に確認応答されます。フロー制御は、トラフィック量が多い時間帯でも、メッセージを不必要にドロップしたり遅延させたりすることなく、パイプラインの安定性を維持するのに役立ちます。

質問: **2**
あなたのチームは、毎日更新される予算データを追跡するためにGoogleスプレッドシートを使用しています。チームは、予算データとBigQueryテーブルに保存されている実際の費用データを比較したいと考えています。毎日の予算と実際の費用の差を計算するソリューションを作成する必要があります。チームがGoogleスプレッドシートで毎日更新される結果にアクセスできるようにする必要があります。どうすればよいでしょうか？
A. GoogleスプレッドシートのドライブURIを使用してBigQuery外部テーブルを作成し、実際の費用テーブルを結合します。結合したテーブルをCSVファイルとして保存し、Googleスプレッドシートで開きます。

B. 予算データをCSVファイルとしてダウンロードし、そのCSVファイルをCloud Storageバケットにアップロードします。Cloud Storageから新しいBigQueryテーブルを作成し、実際の費用テーブルを結合します。結合したBigQueryテーブルをConnected Sheetsを使用して開きます。
C. 予算データをCSVファイルとしてダウンロードし、そのCSVファイルをアップロードして新しいBigQueryテーブルを作成します。実際の費用テーブルと新しいBigQueryテーブルを結合し、結果をCSVファイルとして保存します。Googleスプレッドシートで CSVファイルを開きます。
D. Google スプレッドシートのドライブ URI を使用して BigQuery 外部テーブルを作成し、実際のコストテーブルを結合します。結合したテーブルを保存し、Connected Sheets を使用して開きます。
正解**: (**正解を表示します**)**
包括的かつ詳細な説明:
D が正解である理由: Google スプレッドシートから直接 BigQuery 外部テーブルを作成すると、リアルタイムの更新が可能になります。
外部テーブルを BigQuery の実際のコスト テーブルと結合すると、計算が実行されます。
Connected Sheets を使用すると、チームは Google スプレッドシートで直接結果にアクセスして分析でき、データが更新されます。
他のオプションが間違っている理由:A: CSV ファイルとして保存すると、ライブ接続と毎日の更新が失われます。
B: CSV ファイルとしてダウンロードおよびアップロードすると、不要な手順が追加され、ライブ接続が失われます。
C: B と同じ問題で、ライブ接続が失われます。

質問**: 3**
あなたの会社には複数の小売店があり、各店舗で毎日合計売上数を追跡しています。SQLを使用して、店舗ごとの売上の週次移動平均を計算し、各店舗の傾向を把握したいと考えています。どのクエリを使用すればよいでしょうか？
A.

```
SELECT store_id, date, total_sales, AVG(total_sales) OVER (
PARTITION BY store_id
ORDER BY total_sales RANGE BETWEEN 6 PRECEDING AND CURRENT ROW ) as rolling_avg
FROM store_sales_daily
```

B.

```
SELECT store_id, date, total_sales, AVG(total_sales)
OVER (
PARTITION BY date ORDER BY store_id ROWS BETWEEN 6 PRECEDING AND CURRENT ROW ) as rolling_avg
FROM store_sales_daily
```

C.

```
SELECT store_id, date, total_sales, AVG(total_sales)
OVER (
PARTITION BY store_id
ORDER BY date ROWS BETWEEN 6 PRECEDING AND CURRENT ROW ) as rolling_avg
FROM store_sales_daily
```

D.

```
SELECT store_id, date, total_sales, AVG(total_sales)
OVER (
PARTITION BY total_sales
ORDER BY date RANGE BETWEEN 6 PRECEDING AND CURRENT ROW ) as rolling_avg
FROM store_sales_daily
```

正解: (正解を表示します)

場所別の売上の週次移動平均を計算するには:

* クエリは、store_id でグループ化する必要があります (計算を各ストアごとに分割します)。
* ORDER BY 日付により、売上が時系列で評価されます。
* ROWS BETWEEN 6 PRECEDING AND CURRENT ROW は、7 行のローリング ウィンドウを指定します (各行が日次データを表す場合は 1 週間)。
* AVG(total_sales) は、定義されたローリング ウィンドウ全体の平均売上を計算します。

選択されたクエリは次の要件を満たしています:

PARTITION BY store_id は、計算を各店舗ごとにグループ化します。

ORDER BY date は、ローリング平均に合わせて行を正しく順序付けます。

前の 6 行と現在の 6 行の間の行は、7 日間の移動平均を保証します。

Google ドキュメントからの抜粋: BigQuery の分析関数」(https://cloud.google.com)/bigquery/docs/reference/standard-sql/analytic-function-concepts):"時間列の ORDER BY を指定した ROWS BETWEEN n PRECEDING AND CURRENT ROW を使用して、store_id などのグループ化キーでパーティション化された 7 日間のウィンドウなどの固定行数の移動平均を計算します。"

質問: **4**
あなたは、ある調査会社のCloud Storageバケットの管理を担当しています。会社では、データの階層化と保持ルールが明確に定義されています。データ保持のニーズを満たしながら、ストレージコストを最適化する必要があります。どうすればよいでしょうか？
A. アーカイブ ストレージ クラスを使用するようにバケットを構成します。
B. 各バケットにライフサイクル管理ポリシーを構成して、ストレージ クラスをダウングレードし、経過時間に基づいてオブジェクトを削除します。
C. 標準ストレージ クラスを使用するようにバケットを構成し、オブジェクトのバージョン管理を有効にします。
D. Autoclass 機能を使用するようにバケットを構成します。
正解: **(**正解を表示します**)**
各Cloud Storageバケットにライフサイクル管理ポリシーを設定することで、オブジェクトの保存期間やその他の基準に基づいて、オブジェクトを低コストのストレージクラス（Nearline、Coldline、Archiveなど）に自動的に移行できます。さらに、このポリシーは不要になったオブジェクトの削除を自動化できるため、保持ルールの遵守を確保し、ストレージコストを最

適化できます。このアプローチは、明確に定義されたデータ階層化と保持のニーズに適しており、コスト効率と自動化を実現します。

質問: **5**
小売企業はさまざまなソースから顧客データを収集します。
このデータを抽出するためのデータパイプラインを設計しています
a. さらなる分析と ML モデルのトレーニングには、どの Google Cloud ストレージ システムを選択する必要がありますか?
A. 1. オンライン取引: クラウドストレージ
2. 顧客からのフィードバック: クラウドストレージ
3. ソーシャルメディアアクティビティ: クラウドストレージ
B. 1. オンライントランザクション: BigQuery
2. 顧客からのフィードバック: クラウドストレージ
3. ソーシャルメディアアクティビティ: BigQuery
C. 1. オンライントランザクション: Bigtable
2. 顧客からのフィードバック: クラウドストレージ
3. ソーシャルメディア活動: MySQL 用 CloudSQL
D. 1. オンライン トランザクション: Cloud SQL for MySQL
2. 顧客からのフィードバック: BigQuery
3. ソーシャルメディアアクティビティ: クラウドストレージ
正解: **(**正解を表示します**)**
オンライントランザクション :BigQueryは大規模な構造化データのクエリと分析に最適化されたサーバーレスデータウェアハウスであるため、トランザクションデータをBigQueryに保存するのが理想的です。SQLクエリをサポートし、構造化されたトランザクションデータに適しています。
顧客からのフィードバック：顧客からのフィードバックをCloud Storageに保存することは、非構造化テキストファイルを確実かつ低コストで保存できるため、適切です。また、Cloud Storageはデータ処理ツールやMLツールと連携し、さらなる分析を行うのにも役立ちます。
ソーシャルメディアアクティビティ :BigQueryはストリーミング挿入をサポートしており、リアルタイムのデータの取り込みと分析が可能であるため、リアルタイムのソーシャルメディアアクティビティをBigQueryに保存するのが最適です。これにより、ダッシュボードやMLパイプラインへの即時分析と統合が可能になります。

質問: **6**
組織のビジネスアナリストは、ストリーミングデータへのほぼリアルタイムのアクセスを必要としています。しかし、ダッシュボードクエリの読み込みが遅いという報告を受けています。BigQueryのクエリパフォーマンスを調査したところ、遅いダッシュボードクエリは複数の結合と集計を実行していることがわかりました。
ダッシュボードの読み込み時間を改善し、ダッシュボードのデータを可能な限り最新の状態に保つ必要があります。どうすればよいでしょうか？
A. BiqQuery クエリ結果のキャッシュを無効にします。

B. パラメーター化されたデータ型を使用するようにスキーマを変更します。
C. 中間結果を計算して保存するためのスケジュールされたクエリを作成します。
D. マテリアライズド ビューを作成します。
正解**: D (**コメントを発表する**)**
マテリアライズド･ビューを作成することは、ダッシュボードの読み込み時間を短縮しながら、データを可能な限り最新の状態に保つための最適なソリューションです。マテリアライズド・ビューは、複雑な結合や集計の結果を事前に計算してキャッシュするため、ダッシュボードのクエリ実行時間を大幅に短縮します。また、基盤となるデータが変更されると自動的に更新されるため、ほぼリアルタイムで最新のデータにアクセスできます。このアプローチはクエリパフォーマンスを最適化し、ストリーミングデータダッシュボードに効率的でスケーラブルなソリューションを提供します。

質問**: 7**
新しいデータパイプラインを作成する必要があります。以下の要件を満たすサーバーレスソリューションが必要です。
* データは Pub/Sub からストリーミングされ、リアルタイムで処理されます。
* データは保存される前に変換されます。
* データは、Looker を使用して SQL で分析できる場所に保存されます。
パイプラインに推奨する Google Cloud サービスはどれですか?
A. 1. Dataproc サーバーレス
2. ビッグテーブル
B. 1. クラウドコンポーザー
2. MySQL 用 Cloud SQL
C. 1. ビッグクエリ
2. アナリティクスハブ
D. 1. データフロー
2. ビッグクエリ
正解**: (**正解を表示します**)**
Pub/Sub からのデータをリアルタイムで処理、変換し、Looker を使用した SQL ベースの分析用に保存するサーバーレス データ パイプラインを構築するには、Dataflow と BigQuery を使用するのが最適なソリューションです。Dataflow はリアルタイムのデータ処理と変換を可能にするフルマネージド サービスであり、BigQuery は SQL ベースのクエリをサポートし、Looker とシームレスに統合してデータ分析と可視化を実現するサーバーレス データ ウェアハウスです。この組み合わせにより、リアルタイム ストリーミング、変換、そして分析クエリのための効率的なストレージという要件を満たすことができます。

質問**: 8**

あなたは、機密性の高いデータを扱う金融サービス会社に勤めています。規制要件により、データ暗号化は完全に手動で管理する必要があります。データ保存にはどのような種類の鍵を使用することをお勧めしますか？
A. 顧客指定の暗号化キー (CSEK) を使用します。
B. 会社が選択した専用のサードパーティ キー管理システム (KMS) を使用します。
C. Google 管理の暗号鍵（GMEK）を使用します。
D. 顧客管理の暗号化キー (CMEK) を使用します。
正解: **(**正解を表示します**)**
データ暗号化の完全な手動管理を義務付ける規制要件を満たすには、顧客指定の暗号鍵（CSEK）を使用する必要があります。CSEK では、データストレージ用の暗号鍵をお客様が提供し、Google Cloud はこれらの鍵を保存または管理しません。このアプローチにより、組織は暗号化プロセスに対する完全な制御と責任を維持し、厳格な規制コンプライアンス要件を満たすことができます。

質問: **9**
組織では、リアルタイムの金融取引を処理するために Dataflow パイプラインを使用しています。Dataflow ジョブの 1 つが失敗したことがわかりました。できるだけ早く問題を解決する必要があります。どうすればよいでしょうか？
A. データ スループット、エラー率、リソース使用率などの主要な Dataflow 指標を追跡するための Cloud Monitoring ダッシュボードを設定します。
B. ジョブ ステータスの更新について Dataflow API を定期的にポーリングし、エラーが特定された場合に電子メール アラートを送信するカスタム スクリプトを作成します。
C. Google Cloud コンソールの Dataflow ジョブページに移動します。ジョブのログとワーカーログを使用してエラーを特定します。
D. gcloud CLI ツールを使用してジョブのメトリクスとログを取得し、エラーやパフォーマンスのボトルネックについて分析します。
正解: **C (**コメントを発表する**)**
失敗した Dataflow ジョブをできるだけ早くトラブルシューティングするには、Google Cloud コンソールの「Dataflow ジョブ」ページに移動してください。コンソールでは詳細なジョブログとワーカーログにアクセスでき、失敗の原因を特定するのに役立ちます。また、グラフィカル インターフェースでは、パイプラインのステージを視覚化し、パフォーマンス指標をモニタリングし、エラーが発生した場所を正確に特定できるため、問題を迅速に診断して解決するための最も効率的な方法となります。
Google ドキュメントからの抜粋: 「データフロー ジョブの監視」
(https://cloud.google.com/dataflow)
/docs/guides/monitoring-jobs):"失敗した Dataflow ジョブを迅速にトラブルシューティングするには、Google Cloud Console の Dataflow ジョブ ページに移動し、ジョブのログとワーカーのログを表示して、エラーとその根本原因を特定できます。"

質問: **10**

Cloud Storage にある小規模なデータセットを分析のために変換し、BigQuery に読み込む必要があります。変換には、単純なフィルタリングと集計操作が含まれます。最も効率的で費用対効果の高いデータ操作方法を採用したいと考えています。どうすればよいでしょうか？
A. Dataproc を使用して Apache Hadoop クラスタを作成し、Apache Spark を使用して ETL プロセスを実行し、結果を BigQuery に読み込みます。
B. BigQuery の SQL 機能を使用して、Cloud Storage からデータを読み込み、変換し、結果を新しい BigQuery テーブルに保存します。
C. Cloud Data Fusion インスタンスを作成し、Cloud Storage からデータを読み取り、組み込みの変換を使用してデータを変換し、結果を BigQuery に読み込む ETL パイプラインを視覚的に設計します。
D. Dataflow を使用して、Cloud Storage からデータを読み取り、Apache Beam を使用して変換し、結果を BigQuery に書き込む ETL プロセスを実行します。
正解: **(**正解を表示します**)**
包括的かつ詳細な説明：
単純な変換（フィルタリング、集計）を行う小規模なデータセットの場合、コストと複雑さを最小限に抑えるために、BigQuery のネイティブ SQL 機能を活用することを Google は推奨しています。
* オプション A: Spark を使用した Dataproc は小規模なデータセットには過剰であり、クラスタ管理コストとセットアップ時間が発生します。
* オプションB：BigQueryは、Cloud Storage（CSV、JSONなど）から直接データを読み込み、SQLを使用してサーバーレスで変換できるため、追加のサービスコストを回避できます。これは最も効率的で費用対効果の高いアプローチです。
* オプション C: Cloud Data Fusion は複雑な ETL に適していますが、単純なタスクには不要なオーバーヘッド (インスタンスのセットアップ、UI 設計) が追加されます。

質問: **11**
データ処理中に作成される一時ファイルを保存するCloud Storageバケットを管理しています。これらの一時ファイルは7日間のみ使用され、その後は不要になります。ストレージコストを削減し、バケットを整理された状態に保つため、これらのファイルは7日を経過すると自動的に削除したいと考えています。どうすればよいでしょうか？
A. 毎週 Cloud Run 関数を呼び出して 7 日以上経過したファイルを削除する Cloud Scheduler ジョブを設定します。
B. 7 日以上経過したオブジェクトを自動的に削除する Cloud Storage ライフサイクル ルールを構成します。
C. Dataflow を使用して、毎週実行され、ファイルの古さに基づいてファイルを削除するバッチ プロセスを開発します。
D. 毎日実行され、7 日以上経過したファイルを削除する Cloud Run 関数を作成します。
正解: **(**正解を表示します**)**

7 日以上経過したオブジェクトを自動的に削除するように Cloud Storage ライフサイクル ルールを構成することが、次の理由から最適なソリューションです。
組み込み機能: Cloud Storage ライフサイクル ルールは、経過年数に基づいてオブジェクトを自動的に削除または移行するなど、オブジェクトのライフサイクルを管理するために特別に設計されています。
追加のセットアップは不要: 外部サービスやカスタム コードは不要なので、複雑さとメンテナンスが軽減されます。
コスト効率が高い: 追加のコンピューティング コストを発生させることなく、7 日後にファイルを削除するという目標を直接達成します。

質問**: 12**
組織は既存のエンタープライズ ･データウェアハウスをBigQueryに移行することを決定しました。既存のデータパイプラインツールは既にBigQueryへのコネクタをサポートしています。移行速度を最適化するデータ移行アプローチを特定する必要があります。どうすればよいでしょうか？
A. 既存の環境から Cloud Storage へのデータ転送を容易にするために、一時ファイル システムを作成します。Storage Transfer Service を使用して、データを BigQuery に移行します。
B. Cloud Data Fusion ウェブ インターフェースを使用してデータ パイプラインを構築します。パイプラインのオーケストレーションを容易にする有向非巡回グラフ (DAG) を作成します。
C. 既存のデータ パイプライン ツールの BigQuery コネクタを使用して、データ マッピングを再構成します。
D. BigQuery Data Transfer Service を使用してデータ パイプラインを再作成し、データを BigQuery に移行します。
正解**: (**正解を表示します**)**
既存のデータパイプラインツールは既にBigQueryへのコネクタをサポートしているため、既存のデータパイプラインツールのBigQueryコネクタを使用してデータマッピングを再構成するのが最も効率的なアプローチです。これにより、既存のツールを活用し、移行の複雑さとセットアップ時間を削減しながら、移行速度を最適化できます。既存のパイプライン内でデータマッピングを再構成することで、追加のサービスや中間ステップを必要とせずに、データをシームレスにBigQueryに転送できます。

質問**: 13**
あなたの会社はLookerを使用して売上データを視覚化して分析しています
a. 地域別、商品カテゴリ別、期間別の売上など、売上指標を表示するダッシュボードを作成する必要があります。各指標は、複数のテーブルに分散された独自の属性セットに依存しています。ユーザーが特定の営業担当者でデータをフィルタリングし、個々の取引を表示できるようにする必要があります。Googleが推奨するアプローチに従う予定です。どうすればよいでしょうか？
A. 複数のExploreを作成し、それぞれが各売上指標に焦点を当てます。ドリルダウン機能を使用して、ダッシュボードでExploreをリンクします。

B. BigQuery を使用して、それぞれ特定の売上指標に焦点を当てた複数のマテリアライズドビューを作成します。これらのビューを使用してダッシュボードを構築します。
C. すべての売上指標を含む単一のExploreを作成します。このExploreを使用してダッシュボードを構築します。
D. Looker のカスタム ビジュアライゼーション機能を使用して、フィルタリング機能とドリルダウン機能を備えたすべての販売指標を表示する単一のビジュアライゼーションを作成します。
正解: **(**正解を表示します**)**
Google が推奨する方法は、すべての売上指標を網羅した単一の Explore を作成することです。この Explore は、関連するすべての属性とディメンションを含むように設計する必要があります。これにより、ユーザーは地域、商品カテゴリ、期間、営業担当者などのフィルタごとに売上データを分析できます。適切に構造化された Explore があれば、フィルタリング機能やドリルダウン機能をサポートするダッシュボードを効率的に構築できます。このアプローチにより、メンテナンスが簡素化され、一貫性のあるデータモデルが提供され、ユーザーは統一されたフレームワーク内でシームレスにデータを操作および分析できる柔軟性を確保できます。

質問: **14**
組織では、顧客の注文履歴データを保存する必要があります。このデータは分析のために月に一度だけアクセスされ、アクセス後は数秒以内にすぐに利用できる必要があります。ストレージコストを最小限に抑えながら、データの迅速な取得を保証するストレージクラスを選択する必要があります。どうすればよいでしょうか？
A. Nearline ストアを使用して Cloud Storaqe にデータを保存します。
B. Coldline ストアを使用して Cloud ストアにデータを保存します。
C. 標準ストレージを使用してデータを Cloud Storage に保存します。
D. アーカイブ ストレージを使用してデータを Cloud Storage に保存します。
正解: **A (**コメントを発表する**)**
Cloud Storage の Nearline ストレージは、アクセス頻度が低い（月に 1 回など）ものの、必要なときに数秒以内にすぐにアクセスする必要があるデータに最適です。Nearline は、低いストレージコストと迅速な取得時間のバランスを実現しており、履歴データの月次分析などのシナリオに最適です。Coldline や Archive ストレージのような高い取得コストと長いアクセス時間を回避しながら、低頻度のアクセスパターン向けに特別に設計されています。

質問: **15**
新しいアプリケーション用のデータ パイプラインを作成する必要があります。アプリケーションは、エンリッチメントとクレンジングが必要なデータをストリーミングします。最終的に、このデータは機械学習モデルのトレーニングに使用されます。このパイプラインでは、適切なデータ操作手法と、どの Google Cloud サービスを使用するかを決定する必要があります。どのようなサービスを選択すべきでしょうか？
A. ETL; データフロー -> BigQuery
B. ETL; クラウドデータフュージョン -> クラウドストレージ

C. ELT; クラウドストレージ -> Bigtable
D. ELT; Cloud SQL -> アナリティクス ハブ
正解**: A (**コメントを発表する**)**
包括的かつ詳細な説明：
ML トレーニングの前にエンリッチメントとクリーニングを必要とするストリーミング データには、リアルタイム処理と ML 用のデータ ウェアハウスに重点を置いた ETL (抽出、変換、ロード) アプローチが推奨されます。
* オプションA :Dataflow（ストリーミング変換）とBigQuery（ストレージ/MLトレーニング）を使用したETLは、Googleがストリーミング パイプラインに推奨するパターンです。Dataflowはエンリッチメント/クリーニングを処理し、BigQueryはMLモデルのトレーニング（BigQuery ML）をサポートします。
* オプション B: Cloud Data Fusion を使用した Cloud Storage への ETL はバッチ指向であり、ストリーミングに重点が置かれていません。
Cloud Storage は、ML トレーニングに直接使用するには適していません。
* オプション C: Cloud Storage を使用した ELT (ロードしてから変換) から Bigtable への変換が不適切です。Bigtable は NoSQL 用であり、ML トレーニングやロード後の変換用ではありません。

質問**: 16**
多様な商品を扱うeコマースウェブサイトを運営しています。顧客のニーズに応え、在庫切れを回避するために、将来の商品需要を正確に予測する必要があります。会社の過去の売上データはBigQueryテーブルに保存されています。季節性と過去のデータを考慮して商品需要を予測するスケーラブルなソリューションを構築する必要があります。どうすればよいでしょうか？
A. 過去の売上データを使用して、BigQuery ML 時系列モデルをトレーニングし、作成します。MLを使用します。
FORECAST 関数呼び出しにより、予測結果を新しい BigQuery テーブルに出力します。
B. Colab Enterprise を使用して Jupyter ノートブックを作成します。過去の売上データを使用して、Python でカスタム予測モデルをトレーニングします。
C. 過去の売上データを使用して、BigQuery ML線形回帰モデルをトレーニングし、作成します。MLを使用します。
PREDICT 関数呼び出しにより、予測結果を新しい BigQuery テーブルに出力します。
D. 過去の売上データを使用して、BigQuery MLロジスティック回帰モデルをトレーニングし、作成します。ML.PREDICT関数呼び出しを使用して、予測結果を新しいBigQueryテーブルに出力します。
正解**: (**正解を表示します**)**
包括的かつ詳細な説明：
季節性を考慮した製品需要の予測には時系列モデルが必要ですが、BigQuery MLはスケーラブルなサーバーレスソリューションを提供します。分析してみましょう。

* オプションA :BigQuery MLの時系列モデル（例ARIMA_PLUS）は、季節性とトレンドを考慮した予測を目的として設計されています。ML.FORECAST関数は、履歴データに基づいて予測結果を生成し、テーブルに保存します。これはスケーラブル（インフラストラクチャ不要）で、BigQueryとネイティブに統合されているため、eコマースの需要予測に最適です。
* オプション B: カスタム Python モデル (Prophet など) を使用した Colab Enterprise は柔軟性がありますが、コーディング、メンテナンス、場合によってはデータのエクスポートが必要となり、BigQuery ML のインプレース処理に比べてスケーラビリティが低下します。
* オプション C: 線形回帰は連続値を予測しますが、季節性や時系列パターンを効果的に処理しないため、需要予測には適していません。

質問: **17**
あなたはオンライン小売企業で働いています。顧客購入データをCSVファイルで収集し、10分ごとにCloud Storageにプッシュしています。分析のために、このデータを変換してBigQueryにロードする必要があります。この変換には、データのクリーニング、重複の削除、そしてBigQueryの別のテーブルにある商品情報でデータを拡充することが含まれます。ファイルがCloud Storageにロードされるとすぐにデータ処理を開始する、オーバーヘッドの少ないソリューションを実装する必要があります。どうすればよいでしょうか？
A. Cloud Composer センサーを使用して、Cloud Storage に読み込まれているファイルを検出します。Dataproc クラスタを作成し、Composer タスクを使用してクラスタ上でジョブを実行し、データを処理して BigQuery に読み込みます。
B. Cloud Composer で直接非巡回グラフ (DAG) を 1 時間ごとに実行するようにスケジュール設定し、Cloud Storage から BigQuery にデータを一括読み込みして、SQL を使用して BigQuery でデータを処理します。
C. Dataflow を使用して、Pub からの OBJECT_FINALIZE 通知を使用してストリーミング パイプラインを実装します。
/Sub を使用して Cloud Storage からデータを読み取り、変換を実行し、BigQuery にデータを書き込みます。
D. Cloud Storage から BigQuery にデータを処理して読み込む Cloud Data Fusion ジョブを作成します。Pub/Sub で OBJECT_FINALIZE 通知を作成し、新しいファイルが読み込まれるとすぐに Cloud Run 関数をトリガーして Cloud Data Fusion ジョブを開始します。
正解: **(**正解を表示します**)**

Dataflow を使用して、Pub からの OBJECT_FINALIZE 通知によってトリガーされるストリーミング パイプラインを実装します。
/Sub が最適なソリューションです。このアプローチでは、新しいファイルが Cloud Storage にアップロードされるとすぐにデータ処理が自動的に開始されるため、低レイテンシが確保されます。Dataflow は、データのクリーニング、重複排除、そして BigQuery テーブルからの商品情報によるエンリッチメントをスケーラブルかつ効率的に処理できます。Dataflow はフルマネージドサービスであるため、このソリューションはオーバーヘッドを最小限に抑え、リアルタイムまたはほぼリアルタイムのデータパイプラインに最適です。

質問: **18**
Amazon S3 から Google Cloud の Nearline Cloud Storage バケットへの週次 Storage Transfer Service 転送ジョブが既に実行されています。このジョブでは毎週、比較的小さなファイルを大量に転送しています。毎週転送するファイル数は時間の経過とともに増加しており、割り当てられた時間枠内に転送を完了できなくなるリスクがあります。プロセスを置き換えて、転送時間全体を短縮する必要があります。
ソリューションでは、可能な限りコストを最小限に抑える必要があります。どうすればよいでしょうか?
A. Google Cloud CLI を使用して転送ジョブを作成し、-custom-storage-class フラグを使用して標準ストレージ クラスを指定します。
B. include プレフィックスと exclude プレフィックスを使用して並列転送ジョブを作成します。
C. Amazon S3 から Cloud Storage にデータを移行するように毎週スケジュールされたバッチ Dataflow ジョブを作成します。
D. Compute Engine インスタンス上の複数の転送エージェントを利用するエージェントベースの転送ジョブを作成します。
正解: **(正解を表示します)**
包括的かつ詳細な説明:
B が正しい理由: include プレフィックスと exclude プレフィックスを使用して並列転送ジョブを作成すると、データを小さなチャンクに分割し、並列に転送できます。
これにより、スループットが大幅に向上し、全体的な転送時間が短縮されます。
他のオプションが間違っている理由:A: ストレージ クラスを標準に変更しても、転送速度は向上しません。
C: Dataflow は、単純なファイル転送タスクのための複雑なソリューションです。
D: エージェントベースの転送は、大きなファイルやネットワーク制限には適していますが、多数の小さなファイルには適していません。

質問: **19**
御社では、主要なビジネスインテリジェンスプラットフォームとしてLookerを使用しています。LookMLを使用して、Looker Exploreとダッシュボードで各製品の利益率を可視化したいと考えています。迅速かつ効率的にソリューションを実装する必要があります。どうすればよいでしょうか？

A. フィルターを適用して、利益率がプラスの製品のみを表示します。
B. 既存の収益フィールドとコスト フィールドを使用して利益率を計算する新しいメジャーを定義します。
C. 利益率の範囲 (高、中、低など) に基づいて製品を分類する新しいディメンションを作成します。
D. 各製品の利益率を事前計算する派生テーブルを作成し、それを Looker モデルに含めます。
正解**: B (**コメントを発表する**)**
包括的かつ詳細な説明:
B が正しい理由: LookML で新しいメジャーを定義することは、利益率などの集計された指標を計算して視覚化する最も効率的で直接的な方法です。
メジャーは、既存のフィールドに基づいて計算を行うために設計されています。
他のオプションが間違っている理由:A: フィルタリングでは、利益率自体は計算も視覚化もされません。
C: ディメンションはデータを分類するためのものであり、集計されたメトリックを計算するためのものではありません。
D: 派生テーブルは、メジャーを使用して実行できる利益率などの単純な計算には複雑で不要です。

質問**: 20**
機密性の高い過去の金融データを保存している Cloud SQL for PostgreSQL データベースを所有しています。プライマリリージョンが破損した場合でも、データが破損せず、復旧可能であることを保証する必要があります。データは貴重なため、目標復旧時間（RTO）よりも目標復旧ポイント（RPO）を優先する必要があります。プライマリの読み取りおよび書き込み操作のレイテンシを最小限に抑えるソリューションを推奨したいと考えています。どうすればよいでしょうか？
A. マルチリージョン バックアップ ロケーション用に Cloud SQL for PostgreSQL インスタンスを構成します。
B. Cloud SQL for PostgreSQL インスタンスをリージョン可用性（HA）に設定し、別のリージョンの Cloud Storage バケットに 1 時間ごとに Cloud SQL for PostgreSQL データベースをバックアップします。
C. 別のゾーンのセカンダリ インスタンスへの同期レプリケーションを使用して、リージョン可用性 (HA) を実現するように Cloud SQL for PostgreSQL インスタンスを構成します。
D. 別のリージョンのセカンダリ インスタンスへの非同期レプリケーションを使用して、リージョン可用性 (HA) を実現するように Cloud SQL for PostgreSQL インスタンスを構成します。
正解**: (**正解を表示します**)**
包括的かつ詳細な説明：
優先事項は、データの整合性、地域災害後の復旧可能性、低いRPO（データ損失の最小化）、そしてプライマリオペレーションの低レイテンシです。分析してみましょう。
* オプションA :マルチリージョンバックアップでは、ポイントインタイムのスナップショットを別のリージョンに保存します。自動バックアップとトランザクションログにより、RPO（目標復旧

時点）はほぼゼロ（数分程度）となり、災害発生後の復旧も可能です。プライマリオペレーションは1つのゾーンに留まるため、レイテンシは最小限に抑えられます。
* オプションB :リージョン間バックアップを1時間ごとに実行するリージョンHA（別のゾーンへのフェイルオーバー）は、ゾーン障害から保護しますが、1時間ごとのバックアップではRPO（目標復旧時点）が最大1時間となり、貴重なデータにとっては高すぎます。手動によるバックアップ管理はオーバーヘッドを増加させます。
* オプションC：別のゾーンへの同期レプリケーションは、リージョン内でのRPO（目標復旧時点）ゼロを保証しますが、リージョン全体のデータ損失は保護されません。ゾーン間の同期書き込みにより、レイテンシがわずかに増加します。

質問: **21**
組織では、BigQuery に保存されているデータの変換にスケジュールされたクエリを使用しています。スケジュールされたクエリの 1 つが失敗したことがわかりました。この問題をできるだけ早くトラブルシューティングする必要があります。どうすればよいでしょうか？
A. 管理者にBigQueryのinformation_schemaへのアクセスをリクエストします。失敗したジョブIDでジョブビューをクエリし、エラーの詳細を分析します。
B. gcloud CLI を使用してログシンクを設定し、BigQuery 監査ログを BigQuery にエクスポートします。これらのログをクエリして、失敗したジョブ ID に関連付けられたエラーを特定します。
C. Google Cloud コンソールの 「スケジュールされたクエリ」ページに移動します。失敗したジョブを選択し、エラーの詳細を分析します。
D. Cloud Logging のログ エクスプローラ ページに移動します。フィルタを使用して失敗したジョブを見つけ、エラーの詳細を分析します。
正解: **(**正解を表示します**)**

質問: **22**
組織のBigQueryデータウェアハウスには、複数のデータセットが保存されています。複数の部門にまたがるアナリストチームが、これらのデータセットを使用してクエリを実行しています。組織は、BigQueryの月間コストの変動を懸念しています。各部門が実行するクエリに関連するコストを固定予算化するソリューションを見つける必要があります。どうすればよいでしょうか？
A. BigQuery でアナリストごとにカスタム割り当てを作成します。
B. BigQuery エディションを使用して単一の予約を作成します。すべてのアナリストを予約に割り当てます。
C. 各アナリストを、担当部門に関連付けられた個別のプロジェクトに割り当てます。BigQuery エディションを使用して単一の予約を作成します。すべてのプロジェクトを予約に割り当てます。
D. 各アナリストを、担当部門に関連付けられた個別のプロジェクトに割り当てます。BigQuery エディションを使用して、部門ごとに単一の予約を作成します。適切な予約内に各プロジェクトの割り当てを作成します。
正解: **D (**コメントを発表する**)**

BigQueryエディションを使用して、各アナリストを各部門に関連付けられた個別のプロジェクトに割り当て、各部門ごとに単一の予約を作成することで、正確なコスト管理が可能になります。各プロジェクトを各部門の予約に割り当てることで、各部門に固定のコンピューティングリソースと予算を割り当てることができ、クエリコストを予測・管理できるようになります。このアプローチは、部門間の分離と責任体制を維持しながら、クエリコストに固定予算を設定するという組織の目標に合致しています。

質問: **23**
あなたのチームは、スケジュールに従って特定の順序で実行したい複雑なタスクと依存関係の集合を含む複数のデータパイプラインを構築しています。これらのタスクと依存関係は、Cloud Storage 内のファイル、Apache Spark ジョブ、BigQuery 内のデータで構成されています。これらのデータ処理タスクをフルマネージドのアプローチでスケジュールし、自動化できるシステムを設計する必要があります。どうすればよいでしょうか？
A. Cloud Scheduler を使用して、実行するジョブをスケジュールします。
B. Cloud Tasks を使用してジョブを非同期的にスケジュールおよび実行します。
C. Cloud Composer で有向非巡回グラフ（DAG）を作成します。適切な演算子を使用して、Cloud Storage、Spark、BigQuery に接続します。
D. Google Kubernetes Engine にデプロイされた Apache Airflow で有向非巡回グラフ（DAG）を作成します。適切な演算子を使用して、Cloud Storage、Spark、BigQuery に接続します。
正解: **C (**コメントを発表する**)**
Cloud Composer は Apache Airflow をベースとしたフルマネージドでスケーラブルなワークフロー オーケストレーション サービスであるため、有向非巡回グラフ（DAG）を作成するのに最適なソリューションです。Cloud Composer を使用すると、複雑なタスクの依存関係とスケジュールを定義できるだけでなく、Cloud Storage、BigQuery、Dataproc for Apache Spark ジョブなどの Google Cloud サービスとシームレスに統合できます。このアプローチは、運用オーバーヘッドを最小限に抑え、スケジュール設定と自動化をサポートし、データ パイプラインを効率的かつフルマネージドにオーケストレーションする方法を提供します。

質問: **24**
組織には5つの異なるチームに200人の従業員がいます。経営陣は、共有フォルダに保存されているすべてのLookerダッシュボードを従業員が移動または削除できることを懸念しています。組織内の5つの異なるチームが共有フォルダ内のコンテンツを閲覧できる一方で、移動または削除は各チーム固有のダッシュボードのみに制限できる、管理しやすいソリューションを構築する必要があります。どうすればよいでしょうか？
A. 1. 5つの異なるチームそれぞれを表すLookerグループを作成し、対応するグループにユーザーを追加します。2. Sharedフォルダ内に5つのサブフォルダを作成します。各グループに、対応するサブフォルダへの表示アクセスレベルを付与します。

B. 1. チーム固有のすべてのコンテンツをダッシュボード所有者の個人フォルダに移動します。2. すべてのユーザー グループの共有フォルダのアクセス レベルを [表示] に変更します。3. 各ユーザーに、ユーザーの個人フォルダにチームのコンテンツを作成するように指示します。
C. 1. 「すべてのユーザー」グループの共有フォルダのアクセス レベルを表示に変更します。2. 5 つの異なるチームそれぞれを表す Looker グループを作成し、対応するグループにユーザーを追加します。3.
共有フォルダ内に5つのサブフォルダを作成します。各グループに、対応するサブフォルダへの「アクセス管理」および「編集」アクセスレベルを付与します。
D. 1. 「すべてのユーザー」グループの共有フォルダのアクセスレベルを「表示」に変更します。2. 共有フォルダ内に5つのサブフォルダを作成します。各チームメンバーに、対応するサブフォルダへの「アクセス管理」および「編集」アクセスレベルを付与します。
正解**: C (**コメントを発表する**)**
包括的かつ詳細な説明:
C が正しい理由:共有フォルダーを「表示」に設定すると、誰でもコンテンツを見ることができます。
Looker グループを作成すると、アクセス管理が簡素化されます。
サブフォルダーを使用すると、各チームにきめ細かな権限を設定できます。
「アクセス管理、編集」を付与すると、チームは自分のコンテンツのみを変更できるようになります。
他のオプションが間違っている理由:A: 表示アクセスのみを許可するため、チームは編集できません。
B: コンテンツを個人用フォルダーに移動すると、共有の目的が達成されません。
D: チーム全体ではなく、チームのすべてのメンバーに編集アクセス権を付与しますが、これは理想的ではありません。

質問**: 25**
CSV、Avro、Parquet ファイルから Cloud Storage にデータを取り込むデータ パイプラインを設計する必要があります。データには生のユーザー入力が含まれます。BigQuery にデータを保存する前に、悪意のある SQL インジェクションをすべて削除する必要があります。どのデータ操作手法を選択すべきでしょうか。
A. ザ
B. ELT
C. ETL
D. ETLT
正解**: (**正解を表示します**)**
ETL（抽出、変換、ロード）手法は、このシナリオに最適なアプローチです。ファイルからデータを抽出し、必要なデータクレンジング（悪意のあるSQLインジェクションの除去を含む）を適用して変換し、サニタイズされたデータをBigQueryにロードできるためです。BigQueryにロードする前

にデータを変換することで、クリーンで安全なデータのみが保存されることが保証され、これはセキュリティとデータ品質にとって非常に重要です。

質問: **26**
小売企業は、BigQuery に保存されている過去の購入データを用いて顧客離脱を予測したいと考えています。データセットには、顧客の人口統計、購入履歴、そして顧客が離脱したかどうかを示すラベルが含まれています。離脱リスクのある顧客を特定するための機械学習モデルを構築したいと考えています。顧客離脱を予測するためのロジスティック回帰モデルを作成し、churned 列をターゲットラベルとして設定した customer_data テーブルを用いてトレーニングする必要があります。どの BigQuery ML クエリを使用すべきでしょうか？

A.
```
CREATE OR REPLACE MODEL churn_prediction_model
OPTIONS(model_type='logistic_reg') AS
SELECT *
FROM customer_data;
```

B.
```
CREATE OR REPLACE MODEL churn_prediction_model
OPTIONS(model_type='logistic_reg') AS
SELECT * EXCEPT(churned),
       churned AS label
FROM customer_data;
```

C.
```
CREATE OR REPLACE MODEL churn_prediction_model
OPTIONS(model_type='logistic_reg') AS
SELECT * EXCEPT(churned)
FROM customer_data;
```

D.
```
CREATE OR REPLACE MODEL churn_prediction_model
OPTIONS(model_type='logistic_reg') AS
SELECT churned as label
FROM customer_data;
```

正解: **(**正解を表示します**)**
BigQuery ML で、顧客離脱を予測するロジスティック回帰モデルを作成する場合、正しいクエリは次のようになります。

ターゲット ラベル列 (この場合は、churned) は特徴入力としてではなくトレーニングに使用されるため、特徴列から除外します。
BigQuery ML ではターゲット列に label という名前を付ける必要があるため、ターゲット ラベル列の名前を label に変更します。
選択されたクエリは次の要件を満たしています。
SELECT * EXCEPT(churned), churned AS label: 機能から churned を除外し、名前を label に変更します。
OPTIONS(model_type='logistic_reg') は、ロジスティック回帰モデルがトレーニングされていることを指定します。
この設定により、予測のために離脱列をターゲットにしながら、データセット内の機能を使用してモデルが正しくトレーニングされることが保証されます。

質問: **27**
組織ではGoogle Cloud上で新しいアプリケーションを構築しています。複数のデータファイルをCloud Storageに保存する必要があります。組織では、これらのデータファイルを保存できるクラウドリージョンを2つだけ承認しています。そのため、自動化された高可用性を含むCloud Storageバケット戦略を決定する必要があります。
何をすべきでしょうか?
A. デュアルリージョン バケットを作成し、このバケットにファイルをアップロードします。
B. 2 つのリージョンそれぞれに単一リージョン バケットを作成し、gcloud storage コマンドを使用して、両方のリージョンのバケット間でデータを複製します。
C. マルチリージョン バケットを作成し、このバケットにファイルをアップロードします。
D. 2 つのリージョンそれぞれに単一リージョン バケットを作成し、Storage Transfer Service を使用して両方のリージョンのバケット間でデータを複製します。
正解: **A (**コメントを発表する**)**
包括的かつ詳細な説明：
この戦略では、自動化された高可用性（HA）を備えた2つの特定のリージョンにストレージを配置する必要があります。Cloud Storageのロケーションオプションによってソリューションが決まります。
* オプションA :デュアルリージョンバケット（例us-west1とus-east1）は、ユーザーが指定した2つのリージョン間でデータを同期的に複製し、手動介入なしでHAを実現します。完全に自動化されており、要件を満たしています。
* オプション B: gcloud ストレージ レプリケーションを備えた 2 つの単一リージョン バケットは手動で、自動化されておらず、リアルタイム HA が欠如しています (スクリプトとモニタリングが必要)。
* オプション C: マルチリージョン バケット (例: us) は、地理内の複数のリージョンにまたがりますが、正確に 2 つのリージョンを指定することはできないため、制限に違反する可能性があります。

質問: **28**
重要な月末レポートに使用するBigQueryテーブルを管理しています。このテーブルは毎週、新しい売上データで更新されます。誤ってテーブルを削除した場合のデータ損失やレポートの問題を防ぎたいと考えています。どうすればよいでしょうか？
A. テーブルのタイムトラベル期間を正確に7日間に設定します。削除時に、タイムトラベルデータのみを使用して削除されたテーブルを再作成します。
B. テーブルの新しいスナップショットを週に1回作成するようにスケジュールします。テーブルを削除すると、スナップショットとタイムトラベルデータを使用して、削除されたテーブルが再作成されます。
C. テーブルのクローンを作成します。削除時に、クローンの内容をコピーして削除されたテーブルを再作成します。
D. テーブルのビューを作成します。削除時に、ビューとタイムトラベルデータから削除されたテーブルを再作成します。
正解: **B** (コメントを発表する)
テーブルのスナップショットを毎週作成するようにスケジュール設定することで、テーブルのポイントインタイムバックアップを確実に取得できます。誤って削除した場合でも、スナップショットからテーブルを再作成できます。さらに、BigQuery のタイムトラベル機能を使用すると、削除から最大 7 日前のデータを復元できます。スナップショットとタイムトラベルを組み合わせることで、重要なテーブルのデータ損失を防ぎ、レポートの継続性を確保する堅牢なソリューションが実現します。このアプローチは、リスクを最小限に抑えながら、柔軟なリカバリを実現します。

質問: **29**
機密性の高い過去の金融データを保存している Cloud SQL for PostgreSQL データベースを所有しています。プライマリリージョンが破損した場合でも、データが破損せず、復旧可能であることを保証する必要があります。データは貴重なため、目標復旧時間（RTO）よりも目標復旧ポイント（RPO）を優先する必要があります。プライマリの読み取りおよび書き込み操作のレイテンシを最小限に抑えるソリューションを推奨したいと考えています。どうすればよいでしょうか？
A. 別のリージョンのセカンダリ インスタンスへの非同期レプリケーションを使用して、リージョン可用性 (HA) を実現するように Cloud SQL for PostgreSQL インスタンスを構成します。
B. マルチリージョンのバックアップ場所用に Cloud SQL for PostgreSQL インスタンスを構成します。
C. Cloud SQL for PostgreSQL インスタンスをリージョン可用性（HA）用に設定します。Cloud SQL for PostgreSQL データベースを別のリージョンの Cloud Storage バケットに 1 時間ごとにバックアップします。
D. 別のゾーンのセカンダリ インスタンスへの同期レプリケーションを使用して、リージョン可用性 (HA) を実現するように Cloud SQL for PostgreSQL インスタンスを構成します。
正解: **D** (コメントを発表する)
包括的かつ詳細な説明:

D が正しい理由:同期レプリケーションにより、データがプライマリ インスタンスとセカンダリ インスタンスの両方に同時に書き込まれるため、データ損失 (RPO) が最小限に抑えられます。異なるゾーン内のリージョン可用性 (HA) により、同じリージョン内で冗長性が確保され、レイテンシが最小限に抑えられます。
他のオプションが間違っている理由:A: 非同期レプリケーションではデータが失われる可能性があります。
B: マルチリージョン バックアップは、遅延を最小限に抑えるものではなく、災害復旧を目的としています。
C: 1 時間ごとのバックアップでは、可能な限り低い RPO は提供されません。

質問: **30**
組織ではBigQueryに複数のデータセットを保有しています。これらのデータセットを外部パートナーと共有することで、外部パートナーが各自のプロジェクトにデータをコピーすることなくSQLクエリを実行できるようにしたいと考えています。各パートナーのデータは、それぞれのBigQueryデータセットに整理されています。各パートナーは、自身のデータにのみアクセスできるようにする必要があります。Googleが推奨するプラクティスに従いながら、データを共有したいと考えています。どうすればよいでしょうか？
A. Analytics Hub を使用して、各パートナーデータセットのプライベートデータエクスチェンジにリスティングを作成します。各パートナーがそれぞれのリスティングをサブスクライブできるようにします。
B. 各BigQueryデータセットからデータを読み取り、各パートナー専用のPub/Subトピックにプッシュするデータフロージョブを作成します。各パートナーにpubsub.subscriber IAMロールを付与します。
C. BigQueryデータをCloud Storageバケットにエクスポートします。パートナーにバケットに対するstorage.objectUser IAMロールを付与します。
D. パートナーに BigQuery プロジェクトに対する bigquery.user IAM ロールを付与します。
正解: **(**正解を表示します**)**
Analytics Hub を使用して各パートナー データセットのプライベート データ エクスチェンジ リストを作成することは、BigQuery データを外部パートナーと安全に共有するための Google 推奨の方法です。Analytics Hub を使用すると、大規模なデータ共有を管理できるため、パートナーはデータを独自のプロジェクトにコピーすることなく、データセットに直接クエリを実行できます。パートナー データセットごとに個別のリストを作成し、該当するパートナーのみがサブスクライブできるようにすることで、最小権限の原則に従い、パートナーが特定のデータのみにアクセスできるようにします。このアプローチは安全かつ効率的で、外部データ共有を伴うシナリオ向けに設計されています。

質問: **31**
会社ではバッチ変換パイプラインをGoogle Cloudに移行しています。SQLのみを使用したプログラムによる変換をサポートするソリューションを選択する必要があります。また、パイプライン

のバージョン管理のためにGit統合をサポートするテクノロジーも必要です。どうすればよいでしょうか？
A. Cloud Data Fusion パイプラインを使用します。
B. Dataform ワークフローを使用します。
C. Dataflow パイプラインを使用します。
D. Cloud Composer 演算子を使用します。
正解: **B** (コメントを発表する)
Dataform ワークフローは、SQL のみを使用してプログラムによる変換を実行したい場合に、バッチ変換パイプラインを Google Cloud に移行するための理想的なソリューションです。Dataform を使用すると、データ変換用の SQL ベースのワークフローを定義でき、バージョン管理のための Git 統合をサポートしているため、パイプラインの共同作業とバージョン追跡が可能になります。このアプローチは、SQL ドリブンのデータパイプライン管理向けに特別に構築されており、お客様の要件に完全に適合します。
このソリューションでは、変換にSQLを使用し、バージョン管理にはGitと統合し、特にバッチパイプラインに重点を置く必要があります。評価してみましょう。
* オプションA: Cloud Data Fusion は、SQL のみの変換ではなく、プラグインを備えたビジュアル UI を使用します。ネイティブの Git 統合がないため（外部ツールが必要）、重要な要件を満たしていません。
* オプションB :Dataformは、BigQuery変換用のSQLベースのワークフローツールで、パイプラインをSQLXスクリプトとして定義します。バージョン管理のためにGitとネイティブに統合されており、最小限のオーバーヘッドでバッチELTプロセスをサポートします。
* オプションC: Cloud Composer は、SQL のみの変換ではなく、Python DAG と演算子を使用します。Git は可能ですが、ワークフロー設計に組み込まれているわけではありません。

質問: **32**
あなたの会社は、ユーザーが動画ファイルをアップロードして共有できるウェブサイトを開発しました。これらのファイルは、最初にアップロードされたときに最も頻繁にアクセスされ、共有されます。時間が経つにつれて、ファイルへのアクセスと共有頻度は低下しますが、古い動画ファイルの中には非常に人気のあるものもあるかもしれません。
シンプルでコスト効率の高いストレージシステムを設計する必要があります。どうすればよいでしょうか？

A. Autoclass を有効にした単一リージョン バケットを作成します。
B. 単一リージョンのバケットを作成します。24時間ごとに実行され、アップロード日に基づいてストレージクラスを変更するCloud Schedulerジョブを設定します。
C. アップロード日に基づいてカスタム オブジェクト ライフサイクル管理ポリシーを使用して単一リージョン バケットを作成します。
D. アーカイブをデフォルトのストレージ クラスとして単一リージョン バケットを作成します。
正解: **(**正解を表示します**)**
アップロード日に基づいてカスタムオブジェクトライフサイクル管理ポリシーを適用した単一リージョンバケットを作成することが最適なソリューションです。このアプローチにより、時間の経過とともにアクセス頻度が低下するオブジェクトを、より安価なストレージクラスに自動的に移行できます。例えば、頻繁にアクセスされるファイルは、最初はStandardストレージクラスに残し、人気が下がってきたらNearline、Coldline、またはArchiveストレージに移行できます。この戦略により、動画ファイルのライフサイクル管理を自動化することで、シンプルさを維持しながら、費用対効果が高く効率的なストレージシステムを実現できます。

質問: **33**
あなたはリアルタイムのプレイヤーアクティビティデータを収集するゲーム会社で働いています。このデータはPubにストリーミングされます。
/Sub で処理され、BigQuery にロードして分析を行う必要があります。この処理には、BigQuery のパーティション分割テーブルにロードする前に、データのフィルタリング、拡充、集計が含まれます。Google が推奨するアプローチに従いながら、低レイテンシと高スループットを確保するパイプラインを設計する必要があります。
何をすべきでしょうか?
A. Cloud Composer を使用して、Pub/Sub からデータを読み取り、Python スクリプトを使用してデータを処理し、BigQuery に書き込むワークフローをオーケストレートします。
B. Dataproc を使用して、Pub/Sub からデータを読み取り、データを処理して BigQuery に書き込む Apache Spark ストリーミング ジョブを作成します。
C. Dataflow を使用して、Pub/Sub からデータを読み取り、データを処理して、ストリーミング API を使用して BigQuery に書き込むストリーミング パイプラインを作成します。
D. Cloud Run 関数を使用して Pub/Sub トピックをサブスクライブし、データを処理し、ストリーミング API を使用して BigQuery に書き込みます。
正解: **(**正解を表示します**)**
包括的かつ詳細な説明:
C が正しい理由: Dataflow は、Google Cloud でのリアルタイム ストリーム処理に推奨されるサービスです。
低レイテンシと高スループットを備えたスケーラブルで信頼性の高い処理を提供します。
Dataflow のストリーミング API は、Pub/Sub 統合と BigQuery ストリーミング挿入用に最適化されています。

他のオプションが間違っている理由:A: Cloud Composer は、リアルタイム ストリーミングではなく、バッチ オーケストレーション用です。
B: Dataproc と Spark ストリーミングはより複雑であり、このタスクでは Dataflow ほど効率的ではありません。
D: Cloud Run 関数は、継続的なストリーム処理ではなく、ステートレスでイベント ドリブンなアプリケーション用です。

質問: **34**
組織には、給与や業績評価などの従業員の機密情報を含むBigQueryデータセットがあります。人事部の給与担当者は、集計された業績データへの継続的なアクセスが必要ですが、その他の機密データへの継続的なアクセスは必要ありません。給与担当者に、データセット全体へのアクセスを許可することなく、業績データへのアクセスを許可する必要があります。そのためには、最もシンプルかつ安全な方法を使用する必要があります。どうすればよいでしょうか？
A. 承認済みビューを使用して、クエリ結果を給与担当者と共有します。
B. データセット内のパフォーマンスデータを含むテーブルに対して、行レベルおよび列レベルの権限とポリシーを作成します。給与計算担当者に適切な権限セットを提供します。
C. 集計されたパフォーマンスデータを含むテーブルを作成します。テーブルレベルの権限を使用して、給与担当者にアクセス権を付与します。
D. 集計されたパフォーマンスデータを使用してSQLクエリを作成します。結果をCloud Storageバケット内のAvroファイルにエクスポートします。バケットを給与計算担当者と共有します。
正解: **A** (コメントを発表する)
承認済みビューの使用は、データセット全体を公開することなく、給与計算担当者に集計されたパフォーマンスデータへのアクセスを許可する最もシンプルかつ安全な方法です。承認済みビューを使用すると、BigQuery で集計されたパフォーマンスデータのクエリ結果のみを含むビューを作成できます。給与計算担当者は、基盤となる機密データへのアクセスを許可されることなく、ビューに対してクエリを実行できます。このアプローチはセキュリティを確保し、最小権限の原則を遵守し、複雑な行レベルまたは列レベルの権限管理の必要性を排除します。

質問: **35**
組織では、取り込み時間に基づいてパーティション分割されたBigQueryテーブルを使用しています。組織のストレージコストを削減するため、1年以上経過したデータを削除する必要があります。コストを最小限に抑えながら、最も効率的なアプローチを採用したいと考えています。どうすればよいでしょうか？
A. 1年以上前のデータに対して「deleted」列を「yes」に設定するSQL文を定期的に実行するスケジュールクエリを作成します。削除済みとしてマークされた行を除外するビューを作成します。
B. 1 年以上経過した行を除外するビューを作成します。
C. ユーザーは、SQL の alter table ステートメントを使用してパーティション フィルターを指定する必要があります。

D. SQL の ALTER TABLE ステートメントを使用して、テーブル パーティションの有効期限を 1 年に設定します。
正解: **(**正解を表示します**)**
ALTER TABLE ステートメントを使用してテーブルパーティションの有効期限を1年に設定することが、最も効率的で費用対効果の高い方法です。これにより、1年以上経過したパーティションのデータが自動的に削除され、手動による介入や追加のクエリを必要とせずにストレージコストを削減できます。管理オーバーヘッドを最小限に抑え、データ保持ポリシーの遵守を確保しながら、BigQuery のストレージ使用量を最適化できます。

質問: **36**
ソーシャルメディアの日々のデータを分析するプロジェクトに取り組んでいます。Cloud Storage には 100 GB の JSON 形式のデータが保存されており、容量は増え続けています。
このデータを変換してBigQueryにロードし、分析する必要があります。Googleが推奨するアプローチに従いたいのですが、どうすればよいでしょうか？
A. Cloud Storage からデータを手動でダウンロードします。Python スクリプトを使用してデータを変換し、BigQuery にアップロードします。
B. Cloud Run 関数を使用してデータを変換し、BigQuery に読み込みます。
C. Dataflow を使用してデータを変換し、変換されたデータを BigQuery に書き込みます。
D. Cloud Data Fusion を使用してデータを BigQuery の生のテーブルに転送し、SQL を使用して変換します。
正解: **C (**コメントを発表する**)**
包括的かつ詳細な説明:
C が正しい理由: Dataflow は、バッチ モードとストリーミング モードの両方でデータを変換および拡充するためのフルマネージド サービスです。
Dataflow は、大規模なデータセットを変換するための Google 推奨の方法です。
並列処理用に設計されているため、大規模なデータセットに適しています。
他のオプションが間違っている理由:A: 手動でのダウンロードとスクリプトはスケーラブルでも効率的でもありません。
B: Cloud Run 関数は、大規模なデータ変換ではなく、ステートレス アプリケーション用です。
D: Cloud Data fusion は機能しますが、大規模なデータ変換には Dataflow の方が最適化されています。

質問: **37**
Cloud Storage では、生データ、処理済みデータ、バックアップなど、大量のデータを管理しています。組織は、特定のデータタイプについてデータの不変性を義務付ける厳格なコンプライアンス規制の対象となっています。効率的なプロセスを用いてストレージコストを削減しつつ、ストレージ戦略が保持要件を満たしていることを確認したいと考えています。どうすればよいでしょうか？

A. アクセスパターンに基づいてオブジェクトを適切なストレージクラスに移行するためのライフサイクル管理ルールを設定します。不変性要件を満たすために、すべてのオブジェクトに対してオブジェクトのバージョン管理を設定します。
B. オブジェクトを、その経過時間とアクセスパターンに基づいて異なるストレージクラスに移動します。Cloud Key Management Service（Cloud KMS）を使用して、特定のオブジェクトを顧客管理暗号鍵（CMEK）で暗号化し、不変性の要件を満たします。
C. オブジェクトのメタデータを定期的にチェックし、経過時間とアクセスパターンに基づいてオブジェクトを適切なストレージクラスに移動する Cloud Run 関数を作成します。オブジェクトの保持を使用して、特定のオブジェクトの不変性を強化します。
D. オブジェクト保持を使用して特定のオブジェクトの不変性を強制し、ライフサイクル管理ルールを構成して、経過時間やアクセス パターンに基づいてオブジェクトを適切なストレージ クラスに移行します。
正解: **(**正解を表示します**)**
このシナリオでは、オブジェクトの保持とライフサイクル管理ルールの使用が最も効率的で準拠した戦略です。その理由は次のとおりです。
不変性: オブジェクトの保持 (一時的またはイベントベース) により、オブジェクトが削除または上書きされないことが保証され、データの不変性に関する厳格なコンプライアンス規制が満たされます。
コスト効率: ライフサイクル管理ルールは、オブジェクトの古さとアクセス パターンに基づいて、オブジェクトをよりコスト効率の高いストレージ クラスに自動的に移行します。
コンプライアンスと自動化: このアプローチでは、組み込みの Cloud Storage 機能を活用して、手作業による労力を削減しながら、保持要件へのコンプライアンスを確保します。

質問: **38**
大量のデータに基づいて、週ごとの集計売上レポートを作成する必要があります。Pythonを使用して、このレポートを効率的に生成するプロセスを設計したいと考えています。どうすればよいでしょうか？
A. NumPy を使用する Cloud Run 関数を作成します。Cloud Scheduler を使用して、関数が週に 1 回実行されるようにスケジュールを設定します。
B. Colab Enterprise ノートブックを作成し、bigframes.pandas ライブラリを使用します。ノートブックを週に 1 回実行するようにスケジュールします。
C. Cloud Data Fusion と Wrangler のフローを作成します。フローを週に 1 回実行するようにスケジュールします。
D. Python でコーディングされた Dataflow 有向非巡回グラフ（DAG）を作成します。Cloud Scheduler を使用して、コードを週に 1 回実行するようにスケジュールします。
正解: **D (**コメントを発表する**)**
Python でコーディングされた有向非巡回グラフ（DAG）と Dataflow を組み合わせることで、大量のデータに基づいて週次集計売上レポートを生成する最も効率的なソリューションを実現できます。Dataflow は大規模データ処理に最適化されており、集計を効率的に処理できます。Python で

はパイプライン ロジックをカスタマイズでき、Cloud Scheduler ではプロセスを自動化して週次で実行できます。このアプローチにより、スケーラビリティ、効率性、そして大規模なデータセットを費用対効果の高い方法で処理する能力が確保されます。

質問: **39**
Cloud SQL データベースにデータを保存するウェブアプリケーションを管理しています。プライマリデータベースインスタンスから読み取りトラフィックをオフロードすることで、アプリケーションの読み取りパフォーマンスを向上させる必要があります。労力とコストを最小限に抑えるソリューションを実装したいと考えています。どうすればよいでしょうか？
A. Cloud CDN を使用して、頻繁にアクセスされるデータをキャッシュします。
B. 頻繁にアクセスされるデータを Memorystore インスタンスに保存します。
C. データベースをより大きな Cloud SQL インスタンスに移行します。
D. 自動バックアップを有効にし、Cloud SQL インスタンスの読み取りレプリカを作成します。
正解: **D** (コメントを発表する)
自動バックアップを有効にし、Cloud SQLインスタンスのリードレプリカを作成することは、読み取りパフォーマンスを向上させるための最適なソリューションです。リードレプリカを使用すると、プライマリデータベースインスタンスから読み取りトラフィックをオフロードできるため、負荷が軽減され、全体的なパフォーマンスが向上します。このアプローチは費用対効果が高く、Cloud SQLに簡単に実装できます。プライマリインスタンスは書き込み操作に集中し、レプリカは読み取りクエリを処理するため、最小限の労力でシームレスなパフォーマンス向上が実現します。

質問: **40**
組織のeコマースウェブサイトでは、Pub/Subトピックを使用してユーザーアクティビティログを収集しています。組織の経営陣は、ユーザーエンゲージメント指標を集約したダッシュボードを希望しています。ユーザーアクティビティログを集約指標に変換し、同時に生データへのクエリを容易に実行できるソリューションを構築する必要があります。どうすればよいでしょうか？
A. Pub/SubトピックへのDataflowサブスクリプションを作成し、アクティビティログを変換します。変換されたデータをBigQueryテーブルに読み込み、レポートを作成します。
B. イベントドリブンの Cloud Run 関数を作成し、データ変換パイプラインの実行をトリガーします。変換されたアクティビティログを BigQuery テーブルに読み込み、レポートを作成します。
C. Pub/SubトピックへのCloud Storageサブスクリプションを作成します。アクティビティログをAvroファイル形式を使用してバケットに読み込みます。Dataflowを使用してデータを変換し、レポート用にBigQueryテーブルに読み込みます。
D. Pub/SubトピックへのBigQueryサブスクリプションを作成し、アクティビティログをテーブルに読み込みます。SQLを使用してBigQueryにマテリアライズドビューを作成し、レポート用にデータを変換します。
正解: **A** (コメントを発表する)

このシナリオでは、Dataflow を使用して Pub/Sub トピックをサブスクライブし、アクティビティ ログを変換するのが最適なアプローチです。Dataflow は、ストリーミング データをリアルタイムで処理および変換するために設計されたマネージド サービスです。生のアクティビティ ログから指標を効率的に集計し、変換されたデータを BigQuery テーブルにロードしてレポートを作成できます。このソリューションは、スケーラビリティを確保し、リアルタイム処理をサポートし、BigQuery で生データと集計データの両方をクエリできるため、ダッシュボードに必要な柔軟性と分析情報を提供します。

質問: **41**
あなたの会社のeコマースウェブサイトでは、顧客から製品レビューを収集しています。レビューはCSVファイルとして毎日Cloud Storageバケットに読み込まれます。レビューは複数の言語で書かれており、スペイン語に翻訳する必要があります。サーバーレスで効率的、そしてメンテナンスが最小限で済むパイプラインを構成する必要があります。どうすればよいでしょうか？
A. Dataproc を使用して BigQuery にデータを読み込みます。Apache Spark を使用して Cloud Translation API を呼び出し、レビューを翻訳します。シンクとして BigQuery を設定します。
B. Dataflow テンプレート パイプラインを使用して、Cloud Translation API でレビューを翻訳します。シンクとして BigQuery を設定します。
C. Cloud Run 関数を使用してデータを BigQuery に読み込みます。BigQuery ML の create model ステートメントを使用して翻訳モデルをトレーニングします。このモデルを使用して、BigQuery 内で製品レビューを翻訳します。
D. Cloud Run 関数を使用してデータを BigQuery に読み込みます。Cloud Translation API を呼び出す BigQuery リモート関数を作成します。スケジュールされたクエリを使用して新しいレビューを翻訳します。
正解: **D** (コメントを発表する)
Cloud Run 関数を使用して BigQuery にデータを読み込み、Cloud Translation API を呼び出す BigQuery リモート関数を作成することは、サーバーレスで効率的なアプローチです。この設定により、BigQuery でスケジュールされたクエリを使用してリモート関数を呼び出し、新製品レビューを定期的に翻訳できます。このソリューションでは、BigQuery がストレージとクエリを処理し、Cloud Translation API がカスタム ML モデルの開発を必要とせずに正確な翻訳を提供するため、メンテナンスは最小限で済みます。

質問: **42**
組織では、Pub/Sub で毎秒数千ものイベントを受信するほぼリアルタイムの分析を実装する必要があります。受信メッセージには変換が必要です。開発時間を最小限に抑えながら、データを処理、変換し、BigQuery に読み込むパイプラインを構成する必要があります。どうすればよいでしょうか？
A. Google 提供の Dataflow テンプレートを使用して Pub/Sub メッセージを処理し、変換を実行して、結果を BigQuery に書き込みます。

B. Cloud Data Fusion インスタンスを作成し、Pub/Sub をソースとして設定します。Data Fusion を使用して Pub/Sub メッセージを処理し、変換を実行して、結果を BigQuery に書き込みます。
C. Cloud Storage サブスクリプションを使用して、Pub/Sub から Cloud Storage にデータを読み込みます。Dataproc クラスタを作成し、PySpark を使用して Cloud Storage で変換を実行し、結果を BigQuery に書き込みます。
D. Cloud Run 関数を使用して Pub/Sub メッセージを処理し、変換を実行して、結果を BigQuery に書き込みます。
正解: **(**正解を表示します**)**
Google 提供の Dataflow テンプレートを使用することで、Pub/Sub メッセージの準リアルタイム分析を実装する最も効率的で開発しやすいアプローチを実現できます。Dataflow テンプレートはストリーミング データの処理用に事前に構築および最適化されているため、最小限の開発労力でパイプラインを迅速に構成およびデプロイできます。これらのテンプレートは、Pub/Sub からのメッセージの取り込み、必要な変換の実行、処理済みデータの BigQuery への読み込みに対応し、準リアルタイム分析におけるスケーラビリティと低レイテンシを実現します。

質問: **43**
Cloud Storage に保存されたデータに対してバッチ処理を行う Dataproc クラスタを運用しています。関係者にメールで送信するレポートを生成するために、Spark ジョブを毎日スケジュール設定する必要があります。導入が簡単で複雑さを最小限に抑えられるフルマネージド ソリューションが必要です。どうすればよいでしょうか？
A. Cloud Composer を使用して Spark ジョブをオーケストレートし、レポートをメールで送信します。
B. Dataproc ワークフロー テンプレートを使用して、Spark ジョブを定義およびスケジュールし、レポートをメールで送信します。
C. Cloud Run 関数を使用して Spark ジョブをトリガーし、レポートをメールで送信します。
D. Cloud Scheduler を使用して Spark ジョブをトリガーします。また、Cloud Run 関数を使用してレポートをメールで送信します。
正解: **(**正解を表示します**)**
Dataproc ワークフロー テンプレートは、Dataproc クラスタ上で Spark ジョブを定義およびスケジュールするための、フルマネージドでシンプルなソリューションです。ワークフロー テンプレートを使用すると、データ処理やレポート生成などの事前定義されたステップを使用して、Spark ジョブの実行を自動化できます。Cloud Functions や外部メールサービスなどのツールを使用してレポートを送信するステップをワークフローに追加することで、メール通知を統合できます。このアプローチにより、複雑さを最小限に抑えながら、Dataproc のマネージド機能をバッチ処理に活用できます。

質問: **44**
あなたはオンライン小売企業で働いています。顧客購入データをCSVファイルで収集し、10分ごとにCloud Storageにプッシュしています。分析のために、このデータを変換してBigQueryにロー

ドする必要があります。この変換には、データのクリーニング、重複の削除、そしてBigQueryの別のテーブルにある商品情報でデータを拡充することが含まれます。ファイルがCloud Storageにロードされるとすぐにデータ処理を開始する、オーバーヘッドの少ないソリューションを実装する必要があります。どうすればよいでしょうか？
A. Cloud Composer センサーを使用して、Cloud Storage に読み込まれているファイルを検出します。Dataproc クラスタを作成し、Composer タスクを使用してクラスタ上でジョブを実行し、データを処理して BigQuery に読み込みます。
B. Cloud Composer で直接非巡回グラフ (DAG) を 1 時間ごとに実行するようにスケジュール設定し、Cloud Storage から BigQuery にデータを一括読み込みして、SQL を使用して BigQuery でデータを処理します。
C. Dataflow を使用して、Pub/Sub からの OBJECT_FINALIZE 通知を使用してストリーミング パイプラインを実装し、Cloud Storage からデータを読み取り、変換を実行して、BigQuery にデータを書き込みます。
D. Cloud Storage から BigQuery にデータを処理して読み込む Cloud Data Fusion ジョブを作成します。Pub/Sub で OBJECT_FINALI ZE 通知を作成し、新しいファイルが読み込まれるとすぐに Cloud Run 関数をトリガーして Cloud Data Fusion ジョブを開始します。
正解: **(**正解を表示します**)**
Pub/Sub からの OBJECT_FINALIZE 通知によってトリガーされるストリーミング パイプラインを実装するには、Dataflow を使用するのが最適なソリューションです。このアプローチでは、新しいファイルが Cloud Storage にアップロードされるとすぐにデータ処理が自動的に開始されるため、低レイテンシが確保されます。Dataflow は、データのクリーニング、重複排除、そして BigQuery テーブルからの商品情報によるエンリッチメントをスケーラブルかつ効率的に処理できます。Dataflow はフルマネージド サービスであり、リアルタイムまたはほぼリアルタイムのデータ パイプラインに適しているため、このソリューションではオーバーヘッドが最小限に抑えられます。

質問: **45**
あなたはBigQueryで顧客の機密データを取り扱うデータアナリストです。最小権限の原則に従いながら、組織内の承認された担当者のみがこのデータにクエリを実行できるようにする必要があります。どうすればよいでしょうか？
A. IAM ロールを使用してアクセス制御を有効にします。
B. SQL GRANT ステートメントを使用してデータセット権限を更新します。
C. データを Cloud Storage にエクスポートし、署名付き URL を使用してアクセスを承認します。
D. 顧客管理の暗号化キー (CMEK) を使用してデータを暗号化します。
正解: **(**正解を表示します**)**
包括的かつ詳細な説明：
BigQuery はアクセス制御に IAM を使用し、必要な権限のみを付与することで最小限の権限を遵守します。

* オプション A: IAM ロール（例読み取り専用の roles/bigquery.dataViewer）は、Google のセキュリティのベスト プラクティスに沿って、承認されたユーザーへのクエリ アクセスを制限します。
* オプション B: BigQuery はデータセット権限の SQL GRANT をサポートしていません。アクセスは IAM または承認済みビューを介して管理されます。
* オプション C: 署名付き URL を使用して Cloud Storage にエクスポートすると、BigQuery のネイティブ コントロールがバイパスされ、複雑さが増します。

質問: **46**
最近、組織内の Dataflow ストリーミング パイプラインの管理タスクを引き継ぎましたが、適切なアクセス権がプロビジョニングされていないことに気付きました。パイプラインを再起動するには、Google が提供する IAM ロールをリクエストする必要があります。最小権限の原則に従う必要があります。どうすればよいでしょうか？
A. Dataflow 開発者ロールをリクエストします。
B. データフロー閲覧者ロールを要求します。
C. Dataflow Worker ロールを要求します。
D. Dataflow 管理者のロールを要求します。
正解: **(**正解を表示します**)**
Dataflow 開発者ロールは、パイプラインの再起動を含む、Dataflow ストリーミング パイプラインの管理に必要な権限を提供します。このロールは最小権限の原則に準拠しており、不要な管理アクセスを付与することなく、Dataflow ジョブの管理と操作に必要な権限のみを付与します。
Dataflow 管理者などの他のロールでは、より広範な権限が付与されますが、このシナリオでは必要ありません。

質問: **47**
会社では履歴データをCloud Storageに保存しています。すべてのデータが少なくとも3年間バケットに保存されていることを確認する必要があります。どうすればよいでしょうか？
A. オブジェクトのバージョン管理を有効にします。
B. バケットのストレージ クラスをアーカイブに変更します。
C. バケット保持ポリシーを設定します。
D. 一時的なオブジェクト保持を設定します。
正解: **(**正解を表示します**)**

包括的かつ詳細な説明:
C が正解である理由: バケット保持ポリシーは、Cloud Storage バケット内のオブジェクトに最低保持期間を適用するために特別に設計されています。これにより、指定された期間が経過する前にデータが削除または上書きされることがなくなります。
他のオプションが間違っている理由:A: オブジェクトのバージョン管理により、オブジェクトの複数のバージョンを保持できますが、最小保持期間は保証されません。
B: ストレージ クラスをアーカイブに変更するのは、データ保持の強制ではなく、コストの最適化のためです。
D: オブジェクト保留は法的保留であり、一般的な保持ではありません。

質問**: 48**
あなたの会社では、主要なビジネスインテリジェンスプラットフォームとしてLookerを使用しています。LookMLを使用して、Looker Exploreとダッシュボードで各製品の利益率を視覚化したいと考えています。
迅速かつ効率的にソリューションを実装する必要があります。どうすればよいでしょうか？
A. 各製品の利益率を事前計算する派生テーブルを作成し、それを Looker モデルに含めます。
B. 既存の収益フィールドとコスト フィールドを使用して利益率を計算する新しいメジャーを定義します。
C. 利益率の範囲 (高、中、低など) に基づいて製品を分類する新しいディメンションを作成します。
D. フィルターを適用して、利益率がプラスの製品のみを表示します。
正解**: (**正解を表示します**)**
既存の収益と費用のフィールドを用いて利益率を計算するための新しいメジャーをLookMLで定義するのが、最も効率的で簡単な解決策です。このアプローチにより、事前計算や追加テーブルの作成を必要とせず、Looker Exploreやダッシュボード内で利益率を動的に直接計算できます。メジャーは、例えば以下のようなLookML構文で定義できます。
測定基準: 利益率 {
タイプ: 数値
sql: (収益 - コスト) / 収益 ;;
値の形式: "0.0%"
}
この方法はすぐに実装でき、既存の Looker モデルにシームレスに統合されるため、製品全体の利益率を正確に視覚化できます。

質問**: 49**
あなたのチームは、BigQuery に保存されている大規模なデータセットを分析し、ユーザー行動の傾向を特定する必要があります。分析には、複雑な統計計算、Python パッケージ、可視化が含まれます。分析結果の開発と共有のために、管理された共同作業環境を推奨する必要があります。どのような環境を推奨すべきでしょうか？

A. Colab Enterprise ノートブックを作成し、BigQuery に接続します。ノートブックをチームと共有し、Colab Enterprise でデータを分析し、可視化を行います。
B. BigQuery MLを使用して統計モデルを作成します。クエリをチームと共有します。Looker Studioでデータを分析し、可視化を生成します。
C. Looker Studio ダッシュボードを作成し、BigQuery に接続します。ダッシュボードをチームと共有します。Looker Studio でデータを分析し、ビジュアライゼーションを生成します。
D. Connected Sheets を使用して Google Sheets を BigQuery に接続します。Google Sheets をチームと共有し、Google Sheets でデータを分析し、可視化を行います。
正解: **(**正解を表示します**)**
BigQueryに接続されたColab Enterpriseノートブックを使用すると、複雑な統計計算、Pythonパッケージ、可視化に最適な、管理された共同作業環境が提供されます。Colab Enterpriseは、高度な分析のためのPythonライブラリをサポートし、BigQueryとのシームレスな統合により、大規模なデータセットのクエリを可能にします。これにより、チームは共同で分析を開発・共有しながら、可視化機能を最大限に活用できます。このアプローチは、高度な計算やカスタム可視化を伴うタスクに特に適しています。

質問: **50**
オンプレミスのレガシーMySQLデータベースからGoogle Cloudにデータを移行しています。データベースには、数百万行の大規模なテーブルやトランザクションデータなど、データ型やサイズが異なる様々なテーブルが含まれています。データの整合性を維持し、ダウンタイムとコストを最小限に抑えながら、これらのデータを移行する必要があります。
何をすべきでしょうか?
A. Cloud Composer 環境を設定して、カスタムデータパイプラインをオーケストレートします。Python スクリプトを使用して MySQL データベースからデータを抽出し、Compute Engine 上の MySQL に読み込みます。
B. MySQL データベースを CSV ファイルにエクスポートし、Storage Transfer Service を使用してファイルを Cloud Storage に転送し、ファイルを Cloud SQL for MySQL インスタンスに読み込みます。
C. Database Migration Service を使用して、MySQL データベースを Cloud SQL for MySQL インスタンスに複製します。
D. Cloud Data Fusion を使用して、MySQL データベースを Compute Engine 上の MySQL に移行します。
正解: **C (**コメントを発表する**)**
MySQLデータベースをCloud SQL for MySQLインスタンスに複製するには、Database Migration Service（DMS）を使用するのが最適なアプローチです。DMSは、ダウンタイムとコストを最小限に抑えながらデータベースをGoogle Cloudに移行できるように設計されたフルマネージドサービスです。継続的なデータレプリケーションをサポートし、移行プロセス中のデータ整合性を確保し、スキーマとデータ転送を効率的に処理します。このソリューションは、ソースデータベースと

ターゲットデータベース間のリアルタイム同期を維持し、移行時のダウンタイムを最小限に抑えるため、特に大規模なテーブルやトランザクションデータに適しています。