

Associate Data Practitioner

認定試験ガイド

Associate Data Practitioner は、Google Cloud 上のデータを保護および管理します。

データの取り込み、変換、パイプライン管理、分析、機械学習、可視化などのタスクに Google Cloud データサービスを使用した経験があります。受験者は、Infrastructure as a Service (IaaS)、Platform as a Service (PaaS)、Software as a Service (SaaS) などのクラウド コンピューティングのコンセプトに関する基本的な知識が求められます。

セクション 1: データの準備と取り込み (試験内容の 30% 以下)

1.1 データの準備と処理以下のような点を考察します。

- さまざまなデータ操作手法を区別する (例: ETL、ELT、ETLT)
- 適切なデータ転送ツールを選択する (例: Storage Transfer Service、Transfer Appliance)
- データ品質を評価する
- データ クリーニングを行う (例: Cloud Data Fusion、BigQuery、SQL、Dataflow)

1.2 データを抽出して、適切な Google Cloud ストレージ システムに読み込む。以下のような点を考察します。

- データの形式を区別する (例: CSV、JSON、Apache Parquet、Apache Avro、構造化データベース テーブル)
- 適切な抽出ツールを選択する (例: Dataflow、BigQuery Data Transfer Service、Database Migration Service、Cloud Data Fusion)
- 適切なストレージ ソリューションを選択する (例: Cloud Storage、BigQuery、Cloud SQL、Firestore、Bigtable、Spanner)
 - 適切なデータ ストレージ ロケーション タイプを選択する (例: リージョン、デュアルリージョン、マルチリージョン、ゾーン)
 - ユースケースを、構造化データ、非構造化データ、半構造化データの要件があるものに分類する
- 適切なツールを使用して、Google Cloud ストレージ システムにデータを読み込む (例: gcloud、BQ CLI、Storage Transfer Service、BigQuery Data Transfer Service、クライアント ライブラリ)

セクション 2: データ分析とプレゼンテーション(試験内容の 27% 以下)

2.1 BigQuery と Jupyter ノートブックを使用して、データの傾向、パターン、分析情報を特定する。以下のような点を考察します。

- BigQuery で SQL クエリを定義して実行し、レポートを生成して重要な分析情報を抽出する
- Jupyter ノートブックを使用してデータを分析、可視化する(例: Colab Enterprise)
- データを分析してビジネス上の疑問に答える

2.2 ビジネス要件に基づいて、Looker でデータを可視化し、ダッシュボードを作成する。以下のような点を考察します。

- ビジネス上の疑問に答えるダッシュボードを作成、変更、共有する
- さまざまな分析ユースケースにおいて Looker と Looker Studio を比較する
- 簡単な LookML パラメータを操作してデータモデルを変更する

2.3 ML モデルを定義、トレーニング、評価、使用する。以下のような点を考察します。

- BigQuery ML と AutoML を使用してモデルを開発する ML ユースケースを特定する
- BigQuery でリモート接続を使用して、Google の事前トレーニング済み大規模言語モデル(LLM)を使用する
- 標準的な ML プロジェクトを計画する(例: データ収集、モデルのトレーニング、モデルの評価、予測)
- SQL を実行して、BigQuery ML を使用してモデルを作成、トレーニング、評価する
- BigQuery ML モデルを使用して推論を実行する
- Model Registry でモデルを整理する

セクション 3: データ パイプラインのオーケストレーション(試験内容の 18% 以下)

3.1 シンプルなデータ パイプラインを設計、実装する。以下のような点を考察します。

- ビジネス要件に基づいて、データ変換ツール(例: Dataproc、Dataflow、Cloud Data Fusion、Cloud Composer、Dataform)を選択する
- ELT と ETL のユースケースを評価する
- 基本的な変換パイプラインの実装に必要なプロダクトを選択する

3.2 基本的なデータ処理タスクのスケジュール設定、自動化、モニタリングを行う。以下のような点を考察します。

- スケジュールされたクエリを作成し、管理する(例: BigQuery、Cloud Scheduler、Cloud Composer)
- Dataflow ジョブ UI を使用して Dataflow パイプラインの進行状況をモニタリングする
- Cloud Logging と Cloud Monitoring でログを確認して分析する

- ビジネス要件に基づいて、データオーケストレーションソリューション（例: Cloud Composer、スケジュールされたクエリ、Dataproc ワークフロー テンプレート、Workflows）を選択する
- Pub/Sub から BigQuery へのイベントドリブン データの取り込みのユースケースを特定する
- イベントドリブン パイプラインで Eventarc トリガーを使用する（Dataform、Dataflow、Cloud Functions、Cloud Run、Cloud Composer）

セクション 4: データ マネジメント（試験内容の 25% 以下）

4.1 アクセス制御とガバナンスを構成する。以下のような点を考察します。

- Identity and Access Management (IAM) を使用して最小権限アクセスの原則を確立する
 - 基本ロール、事前定義ロール、およびデータサービスの権限（例: BigQuery、Cloud Storage）を区別する
- Cloud Storage のアクセス制御方法（例: 公開アクセス、プライベート アクセス、均一アクセス）を比較する
- Analytics Hub を使用してデータ共有のタイミングを決定する

4.2 ライフサイクル管理を構成する。以下のような点を考察します。

- データアクセスの頻度と保持要件に基づいて適切な Cloud Storage クラスを決定する
- 指定した期間の経過後にオブジェクトを削除するルールを構成して、不要なデータを自動的に削除し、ストレージ費用を削減する（例: BigQuery、Cloud Storage）
- ビジネス要件を考慮して、データをアーカイブするための Google Cloud サービスを評価する

4.3 Cloud Storage と Cloud SQL のデータの高可用性と障害復旧の戦略を特定する。以下のような点を考察します。

- Google マネージド サービスとして提供されるバックアップと復元のソリューションを比較する
- レプリケーションを使用するタイミングを判断する
- データの冗長性を確保するために、プライマリとセカンダリのデータ ストレージ ロケーション タイプ（例: リージョン、デュアルリージョン、マルチリージョン、ゾーン）を区別する

4.4 セキュリティ対策を適用し、データ プライバシー規制を遵守する。以下のような点を考察します。

- 顧客管理の暗号鍵 (CMEK)、顧客指定の暗号鍵 (CSEK)、Google 管理の暗号鍵 (GMEK) のユースケースを特定する
- 暗号鍵を管理する Cloud Key Management Service (Cloud KMS) の役割を理解する
- 転送データの暗号化と保存データの暗号化の違いを特定する