

Professional Data Engineer の更新

本更新試験ガイドは、標準の Professional Data Engineer 試験ガイドを簡略化し、職務に不可欠な主要領域に焦点を当てたものです。更新試験は、Professional Data Engineer の現行の認定資格を更新希望の資格保持者で、更新資格期間内の方にご利用いただけます。

本ガイドのセクションおよび目標番号は、標準の Professional Data Engineer 試験ガイドに準拠しています。省略された目標は、更新試験の評価対象外です。

認定試験ガイド

Professional Data Engineer は、さまざまなアプリケーションのデータを収集、変換、保存、配信することで、データドリブンな意思決定を可能にします。Professional Data Engineer は、パフォーマンスとセキュリティを最適化しながら、堅牢なデータ インフラストラクチャを設計、構築します。ビジネスニーズと規制ニーズを満たすソリューションを評価して選択し、データ プラットフォームを効果的に管理します。Professional Data Engineer は、データ処理、クリーニング、拡充、クエリの生成と変換に最新のテクノロジーを活用します。Professional Data Engineer は、データ ストレージとデータ処理の複雑さを理解し、複雑なワークロードの設計、構築、デプロイ、モニタリング、メンテナンス、最適化、保護に長けています。

セクション 1: データ処理システムの設計 (試験の 25% 程度)

1.1 セキュリティとコンプライアンスを考慮した設計。以下のような点を考察します。

- データアクセスと保存に関する地域的な考慮事項 (データ主権)
- 法令遵守、規制遵守

1.2 信頼性と確実性を考慮した設計。以下のような点を考察します。

- データの準備とクリーニング (Dataform、Dataflow、Cloud Data Fusion、クエリ生成のための LLM のプロンプト)
- データの検証

1.3 柔軟性とポータビリティを考慮した設計。以下のような点を考察します。

- データのステージング、カタログ化、プロファイリング、検出 (データ ガバナンス)

セクション 2: データの取り込みと処理 (試験の 10% 程度)

2.1 データ パイプラインの計画。以下のような点を考察します。

- データ変換とオーケストレーション ロジックの定義

2.2 パイプラインの構築。以下のような点を考察します。

- 変換
 - 処理ロジック
 - AI によるデータ拡充

2.3 パイプラインのデプロイと運用化。以下のような点を考察します。

- ジョブの自動化とオーケストレーション (例: Cloud Composer と Workflows)
- CI / CD (継続的インテグレーションおよび継続的デプロイ)

セクション 3: データの保存 (試験の 20% 程度)

3.1 ストレージ システムの選択。以下のような点を考察します。

- マネージド サービスの選択 (例: BigQuery、BigLake、AlloyDB、Bigtable、Spanner、Cloud SQL、Cloud Storage、Firestore、Memorystore)

3.3 データレイクの使用。以下のような点を考察します。

- レイクの管理 (データの検出、アクセス、費用管理の構成)

3.4 データ プラットフォームを考慮した設計。以下のような点を考察します。

- 要件に基づくデータ プラットフォームを Google Cloud のツール (例: Dataplex、Dataplex Catalog、BigQuery、Cloud Storage) で構築する
- 分散データシステム用の連携ガバナンス モデルを構築する

セクション 4: 分析用データの準備と使用(試験の 15% 程度)

4.1 可視化用データの準備。以下のような点を考察します。

- セキュリティ、データマスキング、Identity and Access Management (IAM)、Cloud Data Loss Prevention (Cloud DLP)

4.2 AI と ML のためのデータの準備。以下のような点を考察します。

- 特徴量エンジニアリング、ML モデルのトレーニングと提供のためのデータ準備(例: BigQuery ML)
- エンベディングと検索拡張生成 (RAG) のための非構造化データの準備

4.3 データの共有。以下のような点を考察します。

- BigQuery Sharing (Analytics Hub)

セクション 5: データ ワークロードの管理と自動化(試験の 18% 程度)

5.2 自動化と反復性の設計。以下のような点を考察します。

- ジョブを反復可能な方法でスケジューリングおよびオーケストレーションする

5.3 ビジネス要件に基づくワークロードの最適化。以下のような点を考察します。

- 容量管理 (BigQuery エディションと予約)

5.4 プロセスのモニタリングとトラブルシューティング。以下のような点を考察します。

- データプロセスのオブザーバビリティ(例: Cloud Monitoring、Cloud Logging、BigQuery 管理パネル)