



# CompTIA Data+

## 認定資格試験

## 出題範囲

試験番号: DAO-001



# 試験について

CompTIA Data+認定資格試験は、データドリブン型の意味決定をサポートでき、ビジネス要件を変革するために必要な知識とスキルを持っていることを証明します:

- データマイニング
- データを操作する
- 基本的な統計手法の適用
- データライフサイクル全体を通じたカバランスと品質基準を守りながら、複雑なデータセットの分析を行う

CompTIA Data+ は、ビジネスインテリジェンスレポート / データアナリストとしての 18 ~ 24 月の実務経験に相当するスキルを目安に設計されています。出題範囲に掲載された項目は、認定資格試験の目的を明確にするためのものであり、試験の出題内容を完全に網羅したものではありません。

## 試験開発

CompTIAの認定資格試験は、ITプロフェッショナルに必要とされるスキルと知識に関して、専門分野のエキスパートによるワークショップ、および業界全体へのアンケートの調査結果に基づいて策定されています。

## CompTIA認定教材の使用に関するポリシー

CompTIA Certifications, LLCは、無許可の第三者トレーニングサイト(通称「ブレインダンプ」)とは提携関係がなく、これらが提供するいかなるコンテンツも公認・推薦・容認しません。CompTIAの認定資格試験の受験準備にこのような教材を使用した個人は、CompTIA受験者同意書の規定に基づいて資格認定を取り消され、その後の受験資格を停止されます。CompTIAでは、無許可教材の使用に関する試験実施ポリシーをよりよく理解していただくための取り組みを進めています。認定資格試験を受験される方は、[CompTIA認定資格試験実施ポリシー](#)をご一読ください。CompTIAの認定資格試験を受験するための学習を始める前には、必ずCompTIAが定めるすべてのポリシーをご確認ください。受験者には、[CompTIA受験者同意書](#)の規定を遵守することが求められています。個々の教材が無許可扱いになるかどうかを確認するには、CompTIA ([examsecurity@comptia.org](mailto:examsecurity@comptia.org)) までメールにてご確認ください。

## 注意事項

箇条書きで挙げられた項目は、すべての試験内容を網羅するものではありません。この出題範囲に掲載がない場合でも、各分野に関連する技術、プロセス、あるいはタスクを含む問題が出題される可能性があります。CompTIAでは、提供している認定資格試験の内容に現在必要とされているスキルを反映するため、また試験問題の信頼性維持のため、継続的な試験内容の検討と問題の改訂を行っています。必要に応じて、現在の出題範囲を基に試験を改訂する場合があります。この場合、現在の試験に関連する資料・教材等は、継続的にご利用いただくことが可能です。

## 試験情報

試験番号	DA0-001
問題数	90問
出題形式	単一/複数選択
試験時間	90分
推奨経験	<ul style="list-style-type: none"><li>・レポート/ビジネスアナリストにおける18～24ヶ月の実務経験</li><li>・データベースや分析ツールの使用経験</li><li>・統計学の基本的な知識</li><li>・データビジュアライゼーションの経験</li></ul>
合格スコア	675 (100-900のスコア形式)

## 試験の出題範囲 (試験分野)

下表は、この試験における試験分野 (ドメイン) と出題比率の一覧です：

試験分野	出題比率
1.0 データの概念と環境	15%
2.0 データマイニング	25%
3.0 データアナリシス	23%
4.0 ビジュアライゼーション	23%
5.0 データガバナンス、品質と管理	14%
<b>計</b>	<b>100%</b>



# 1.0 データの概念と環境

## 1.1 データスキーマとディメンションの基本的な概念を理解する。

- データベース
  - リレーショナル
  - 非相関的
- データマート/データウェアハウス/  
データレイク
  - オンラインランザクシオン処理 (OLTP)
  - オンライン分析処理 (OLAP)
- スキーマの概念
  - スノーフレーク
  - スター
- ゆっくりと変化するディメンション
  - 現在の情報を把握
  - 過去と現在の情報を把握

## 1.2 異なるデータタイプを比較対照することができる。

- 日付
- 数字
- 英数字
- 通貨
- テキスト
- 離散型と連続型
- カテゴリ/ディメンション
- 画像
- オーディオ
- ビデオ

## 1.3 一般的なデータ構造とファイル形式を比較対比する。

- 構成
  - 構造化
    - 定義された行/列
    - キーバリューペア
  - 非構造化
    - 未定義のフィールド
    - マシンデータ
- データファイルの形式
  - テキスト/フラットファイル
    - タブ区切り
    - カンマ区切り
  - JavaScript Object Notation (JSON)
  - Extensible Markup Language (XML)
  - Hypertext Markup Language (HTML)



## 2.0 データマイニング

### 2.1 データ収集の概念を説明することができる。

- 整合性
  - ETL: Extract (抽出)、transform (変換)、load (格納)
  - ELT: Extract (抽出)、load (格納)、transform (変換)
  - デルタロード
- アプリケーションプログラミングインターフェース (API)
- データの収集メソッド
  - Web スクレイピング
  - 公開データベース
- アプリケーションプログラミングインターフェース (API)/Web サービス
  - 調査
  - サンプリング
  - 観察

### 2.2 データベースのクレンジングとプロファイリングを行う一般的な理由を特定することができる。

- 重複データ
- 冗長データ
- 欠測値
- 無効データ
- ノンパラメトリックデータ
- データ異常値
- 仕様の不一致
- データの種類検証

### 2.3 与えられたシナリオに基づいて、データ操作技術を実行することができる。

- データの再コード化
  - 数値
  - カテゴリ
- 派生変数
- データマージ
- データブレンド
- 連結
- データ追加
- インピューテーション
- 削減/集約
- トランスポーズ
- データ正規化
- 構文解析/文字列操作

### 2.4 データ操作とクエリ最適化のための一般的な技術を説明することができる。

- データ操作
  - フィルタリング
  - ソート
  - 日付関数
  - 論理関数
  - 集約関数
  - システム機能
- クエリの最適化
  - パラメータ化
  - インデックスの作成
  - クエリセットの一時表
  - レコードのサブセット
  - 実施計画



## 3.0 データアナリシス

3.1 与えられたシナリオに基づいて、適切な記述統計手法を適用することができる。

- 中心傾向の測定
  - 平均値
  - 中央値
  - モード
- 分散の測定
  - 範囲
  - 最大
  - 最小
  - 分布
  - 分散
  - 標準偏差
- 頻度/パーセンテージ
  - 増減率
  - 変化率
  - 信頼区間

3.2 推測統計手法の目的を説明することができる。

- t検定
- Z値
- P値
- カイ二乗
- 仮説検証
  - 第一種過誤
  - 第二種過誤
- 単純線形回帰
- 相関分析

3.3 分析の種類と主要な分析技術を要約することができる。

- 分析の種類を決定するプロセス
  - ビジネスに関する質問の確認/絞り込み
  - データの必要性と情報源を決定して分析を実施する
  - スコープ分析/ギャップ分析
- 分析の種類
  - 動向分析
    - 時間の経過に伴いデータを比較
  - パフォーマンス分析
    - 定義された目標に対し測定値を追跡する
    - 目標達成のための基本的な計画
  - 探索的データ分析
    - 記述統計を利用して観察結果を判断する
  - リンク解析
    - データポイントまたはパスウェイの接続

3.4 一般的な分析ツールを活用することができる。

(\*\*この出題範囲は、特定ベンダーの機能やツールの目的をテストすることではありません。)

- Structured Query Language (SQL)
- Python
- Microsoft Excel
- R
- Rapid mining
- IBM Cognos
- IBM SPSS Modeler
- IBM SPSS
- SAS
- Tableau
- Power BI
- Qlik
- MicroStrategy
- BusinessObjects
- Apex
- Dataroma
- Domo
- AWS QuickSight
- Stata
- Minitab



## 4.0 ビジュアライゼーション

4.1 与えられたシナリオに基づいて、ビジネス要件を変換してレポートを作成することができる。

- データの内容
- フィルタリング
- ビュー
- 日付範囲
- 頻度
- レポート対象者
  - 配布リスト

4.2 与えられたシナリオに基づいて、レポートとダッシュボードに適切なデザインコンポーネントを使用することができる。

- レポートの表紙
  - 指示
  - 要約
  - 観察と洞察
- 設計要素
  - 色彩設計
  - レイアウト
  - フォントサイズとスタイル
  - 主要なグラフ要素
    - タイトル
    - ラベル
    - 凡例
- 企業レポートの規格/スタイルガイド
  - ブランディング
  - カラーコード
  - ロゴ/商標
  - 透かし
- 文書化要素
  - バージョン番号
  - 参照データソース
  - 基準日
    - レポート実行日
    - データ更新日
- よくある質問 (FAQ)
- 付録

4.3 与えられたシナリオに基づいて、ダッシュボード開発に適切な方法を使用することができる。

- ダッシュボードに関する検討事項
  - データソースと属性
    - フィールドの定義
    - デイメンション
    - Measures (基準)
  - 連続データ/ライブデータ
  - フィールド対静的データ
  - 消費者タイプ
    - 経営幹部
    - 経営者
- 外部ベンダー/利害関係者
  - 一般
  - 技術者
- 開発プロセス
  - モックアップ/ワイヤーフレーム
  - レイアウト/プレゼンテーション
  - フロー/ナビゲーション
  - データストーリー計画
- 承認
- ダッシュボード開発
- 本番環境へのデプロイ
- 配送に関する考慮事項
  - サブスクリプション
  - 配信予定
  - インタラクティブ (ドリルダウン/ロールアップ)
    - 保存された検索
    - フィルタリング
  - 静的
  - ウェブインターフェース
  - ダッシュボード最適化
  - アクセス権限



4.4 与えられたシナリオに基づいて、適切な種類のビジュアライゼーションを適用することができる。

- ・ラインチャート
- ・円グラフ
- ・バブルチャート
- ・散布図
- ・棒グラフ
- ・ヒストグラム
- ・ウォーターフォール
- ・ヒートマップ
- ・地形図
- ・ツリーマップ
- ・積み上げグラフ
- ・インフォグラフィック
- ・ワードクラウド

4.5 さまざまなレポートのタイプを比較対照することができる。

- ・静的レポートと動的レポート
  - 基準時点
  - リアルタイム
- ・アドホック/ワンタイムレポート
- ・セルフサービス/オンデマンド
- ・定期的なレポート
  - コンプライアンスレポート  
(例財務、健康、安全など)
  - リスクおよび規制に関するレポート
  - 運用レポート [例パフォーマンス、  
主要業績評価指標 (KPI)]
- ・戦術/調査レポート



## 5.0 データガバナンス、品質と管理

### 5.1 データガバナンスの重要な概念のまとめ。

- アクセス要件
  - ロールベース
  - ユーザーグループベース
  - データ利用契約
  - リリース承認
- セキュリティ要件
  - データの暗号化
  - データ送信
  - データの非識別化/データマスキング
- ストレージ環境の要件
  - 共有ドライブ、クラウドベース、ローカルストレージの比較
- 使用条件
  - 利用規約
  - データ処理
  - データ削除
  - データ保持
- エンティティ関係の要件
  - レコードリンク制限
  - データ制約
  - カーディナリティ
- データ分類
  - Personally identifiable information (PII: 個人を特定可能な情報)
- Personal health information (PHI: 個人健康情報)
- Payment card industry (PCI: クレジットカード業界)
- 管轄区域の要件
  - 業界および政府の規制の影響
- 情報漏洩の報告
  - 適切な機関へエスケーション

### 5.2 与えられたシナリオに基づいて、データ品質管理の概念を適用することができる。

- 品質を確認する状況
  - データ取得/データソース
  - データ変換/イントラホップ (intra hops)
    - パススルー
    - 変換
  - データ操作
  - 最終製品 (レポート/ダッシュボードなど)
- 自動検証
  - データフィールドからデータ型の検証
  - データポイント数
- データ品質の次元
  - データの一貫性
  - データの精度
  - データ完全性
  - データの整合性
  - データ属性の制限
- データ品質ルールと指標
  - 適合性
  - 不適合性
  - 合格の行
  - 不合格の行
- 品質検証の方法
  - 相互検証
  - サンプル/スポットチェック
  - 合理的な予測
  - データプロファイリング
  - データ監査

### 5.3 マスターデータ管理 (MDM) の概念を説明することができる。

- プロセス
  - 複数のデータフィールドの統合
  - データフィールド名の標準化
  - データディクショナリ
- MDMの状況
  - 合併と買収
- ポリシーおよび規制の遵守
- データアクセスの合理化

## CompTIA Data+ (DA0-001) 略語リスト

下記はCompTIA Data+認定資格試験で使用される略語の一覧です。受験者には、試験準備の一環として、これらの用語を復習し、理解することをお勧めします。

略語	定義
API	Application Programming Interface
AWS	Amazon Web Services
BI	Business Intelligence
ELT	Extract, Load, Transform
ETL	Extract, Transform, Load
FAQ	Frequently Asked Questions
GDPR	General Data Protection Regulation
HTML	Hypertext Markup Language
JSON	JavaScript Object Notation
KPI	Key Performance Indicator
MDM	Master Data Management
OLAP	Online Analytical Processing
OLTP	Online Transaction Processing
P&L	Profit and Loss
PCI	Payment Card Industry
PHI	Personal Health Information
PII	Personally Identifiable Information
RDBMS	Relational Database Management System
SDLC	Software Development Life Cycle
SQL	Structured Query Language
XML	Extensible Markup Language

# CompTIA Data+の推奨ハードウェアとソフトウェアのリスト

本リストは、CompTIA Data+の受験準備として役立てていただくためのハードウェアとソフトウェアのリストです。トレーニングを実施している企業でも、トレーニングの提供に必要な実習室コンポーネントを作成したい場合に役立ちます。各トピックの下の箇条書きリストは例であり、すべてを網羅するものではありません。

## ハードウェア

- デスクトップ/ノートパソコン
  - 大容量分析のための高い処理能力
  - 少量分析のための低い処理能力
- インターネットアクセス
- クラウド環境

## ソフトウェア

- スクリプト実行のためのSQL環境 (SQL Lite、Management Studioなど)
- Eclipse
- Anaconda
- R Studio
- Databaseモデリングツール
- Microsoft Office Suite
- ビジューライゼーションツール (Tableau、Power BIなど)
- 報告ツール
- サンプルデータセット (Kaggle)