

# クラウドベースの データ分析プラットフォーム

## ワークロードの比較調査

John Myers  
Senior Director, Market Intelligence, Teradata



# エグゼクティブサマリ

- 3 はじめに
- 4 ワークロード比較方法
- 7 エンタープライズ価格パフォーマンスの算出方法
- 8 クエリあたりのコスト
- 9 結論

適切なクラウドベースのデータ分析プラットフォームを選択することは、企業の成功にとって極めて重要です。このホワイトペーパーでは、主要な3つのクラウドデータ分析プラットフォーム：Teradata、Snowflake、Databricksの包括的なワークロードを比較、分析しています。<sup>1</sup>

この調査では、各プラットフォームの価格パフォーマンスに焦点を当て、コスト効率と分析パフォーマンスをバランスさせる重要な指針について検討しています。これは、コスト重視のクラウド環境における高度なワークロード管理とクエリ最適化の重要性を浮き彫りとするものになっています。

## 意思決定者のための主要なインサイト

**コスト管理：**価格パフォーマンスに大きな差があることから、プラットフォームを慎重に選択することでコスト削減が可能であることが明らかになりました。これは、ビジネス関係者と情報技術 (IT) のクラウドデータ分析予算全体に大きな影響を与える可能性があります。

**拡張性：**Teradataが、処理能力に比例したコスト増を伴わずに大幅に多くのクエリを処理できることは、データ量とユーザー数の増加に対する拡張性が優れていることを示しています。これは、急成長が見込まれる企業や、分析要件が成熟するにつれて予測不可能な分析要求に直面する企業にとって極めて重要となります。

**業務の効率化：**クエリ処理能力の向上は、より迅速なインサイトの獲得と意思決定につながり、市場の変化への迅速な対応が重要なデータドリブン企業にとって、競争優位性の向上に貢献する可能性があります。

**総所有コスト (TCO) と投資収益率 (ROI)：**ワークロード比較調査の結果から、Teradataは複雑な分析ニーズを持つ企業に、より低いTCOとより高いROIを提供できる可能性があることが示唆されました。意思決定者は、プラットフォームを評価する際に、これらの長期的な影響を考慮する必要があります。

**ワークロード管理：**Teradataのパフォーマンスは、クラウド環境における高度なワークロード管理とクエリ最適化がいかに重要であるかを明らかにしています。企業は、個々のクエリ性能だけでなく、プラットフォームが多様で同時並行的なワークロードをどれだけ適切に管理できるかを評価する必要があります。

1. Teradata の競争力のあるワークロード比較の手法とプロセスの詳細については、[www.teradata.jp/how-we-help/work-load-comparisons](http://www.teradata.jp/how-we-help/work-load-comparisons) を参照するか、[John.Myers@Teradata.com](mailto:John.Myers@Teradata.com) までお問い合わせください。

# はじめに

エラスティックなクラウドベースのコンピュートリソースは、かつてない拡張性、柔軟性、コンピュー能力へのアクセスを提供することで、データ分析に革命をもたらしました。企業が分析リソースをクラウドに移行する動きが加速する中、ワークロードをオンプレミスからクラウドに移行する場合、俊敏性の向上とインフラコストの削減が期待できることが、重要な近代化の推進要因と考えられています。しかし、企業がこのような取り組みを推進するにつれ、クラウドデータ分析業務におけるコスト効率と財務ガバナンスの必要性が、さらに強く求められるようになります。

今日の競争の激しいビジネス環境では、膨大な量のデータから迅速かつ正確にインサイトを導き出す能力は、もはや単なるアドバンテージではなく、必要不可欠なものとなっています。クラウドデータ分析プラットフォームは、この需要に応えるツールとリソースを提供し、かつて想像もできなかった規模でデータを処理、分析、可視化できるようにすることができます。しかし、データ量が指数関数的に増加し、分析ワークロードが複雑化するにつれて、クラウドデータ分析に関連するコストは、効果的に管理されなければ、急速に膨張する可能性があります。

このホワイトペーパーにおけるワークロード比較調査の目的は、クラウドベースのデータ分析プラットフォームを利用する意思決定者にインサイトを提供し、強力な分析機能のニーズと費用対効果のバランスを取ることができるようにすることです。

## なぜあえて比較するのか？

クラウドデータ分析における **ワークロードの厳密な比較** は、単なる数字の評価ではありません。各企業がもつ独自のニーズや目標、特に財務ガバナンスや予算管理といった目標に照らして、あるプラットフォームを他のプラットフォームより選択することの実際的な意味を理解することです。ワークロード比較テストは次のような点で不可欠です：

- 異なるプラットフォーム間のパフォーマンス、コスト効率、拡張性を客観的に分析
- 特定の分析業務を処理する上での強みと弱みの特定
- 各ベンダーのマーケティングメッセージではなく、実用的なインサイトの提供
- 予算制約の中で、テクノロジーの選択とビジネス目標を合致

これらの評価は、特定のワークロードを処理する上での各プラットフォームの長所と短所を明らかにするのに役立ちます。個々のプラットフォームが実世界の要件にどのような影響を与えるかを示すことができ、定性的な主張を超える、プラットフォーム選択のための実用的なインサイトを提供します。パフォーマンスとコストに関して何が期待できるかを明確に把握することで、企業はテクノロジーの選択をビジネス目標に合わせるようになります。

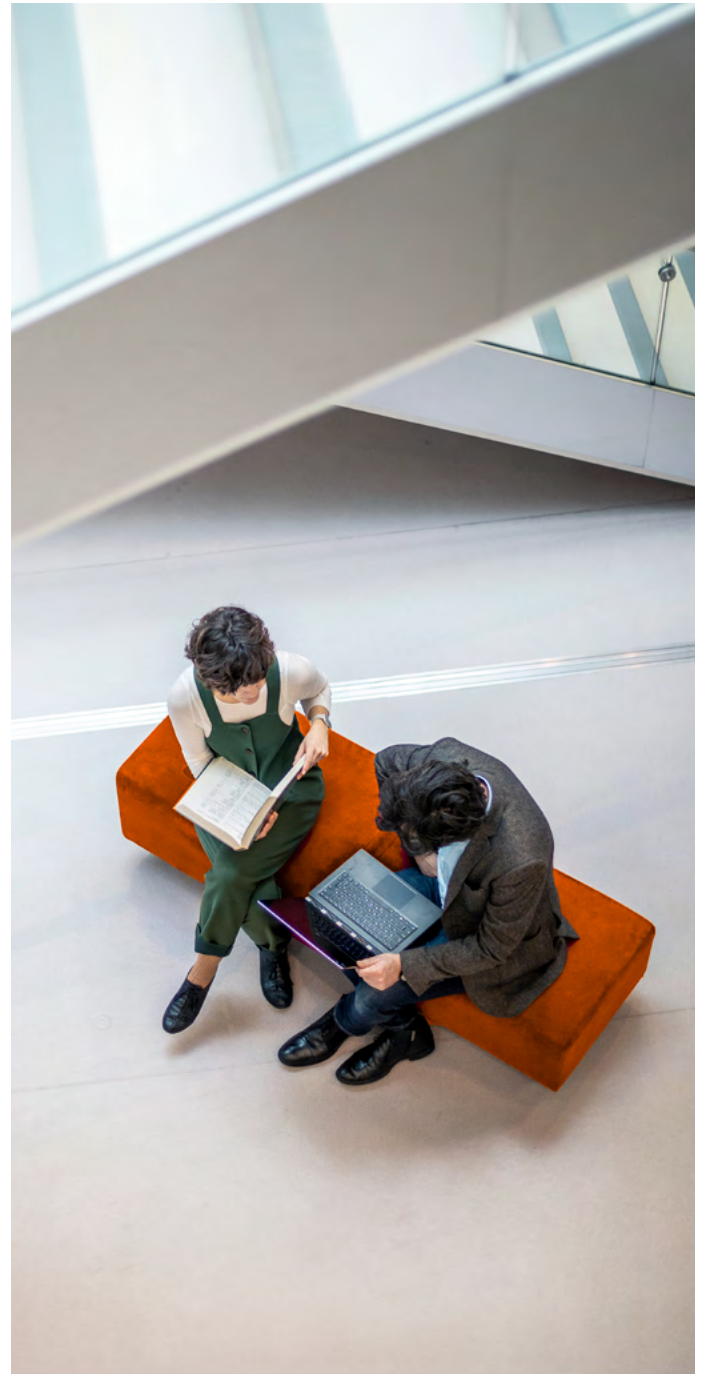


# ワークロード比較方法

今回の調査において採用した方法論は、異なるプラットフォーム間で公平かつ有意義な比較ができるように綿密に設計されたものです。これは、バランスの取れたテスト環境を作るために連動する3つの主要コンポーネントで構成されています。

1. 実際のワークロード状況を模してプラットフォームをテストするためのワークロードミックスを開発。この包括的なミックスは、業界標準の意思決定支援ベンチマークのクエリと、多様で同時並行的な業務分析クエリをシミュレートするクエリを組み合わせたものです。このクエリは、データドリブン企業がリアルタイムの意思決定システムを活用してビジネスを推進し、変化の速い業界で競争上の優位性を確立したいというニーズを反映したものです。この多様なクエリセットにより、今回の調査のワークロード分析は、歴史的にあまり洗練されていないクエリセットではなく、今日のクラウドベースのデータ分析プラットフォームに求められる複雑で多様な要求を正確に表しています。
2. 各プラットフォームを同等のコストで環境構築。このアプローチにより、企業がクラウドデータ分析ソリューションを導入する際にしばしば直面する制約を反映し、標準化された予算内で各プラットフォームの強みを評価することができます。
3. 定義されたワークロードに対して、特定のプラットフォームのチューニングは未実施。データベース管理者 (DBA) によって一般的に有効化され、プラットフォーム上で一般的に利用可能 (GA) であるパフォーマンス機能を活用しました。「マテリアライズドビュー」と呼ばれる特定の機能は、利用可能であればすべてのプラットフォームに実装しました。<sup>2</sup> この機能は、DBA のベストプラクティスに従って、一部の大きなテーブルに実装しました。

この3つのコンポーネントにより、各プラットフォーム固有のアーキテクチャの違いがあるにもかかわらず、同等に比較することができます。これらの条件を注意深くコントロールすることで、クラウドベースのデータ分析プラットフォームを評価する意思決定者に貴重で実用的なインサイトを提供するワークロード比較評価を作成しました。



2. Teradata VantageCloud Lakeは、この機能を“単一テーブル結合インデックス”と呼んでいる。Snowflakeはこれを“マテリアライズドビュー”と呼ぶ。Databricksは2024年6月まで、このコンセプトを一般に利用可能な機能 (GA) として持っておらず、構成に含まれていなかった。

## 実際のワークロード

今日の分析環境は、過去に何が起こったか、現在何が起きているか、そして将来何が起こりそうかを理解することを可能にしなければなりません。一般的に使用されているベンチマークのセットでは、データドリブン企業のすべての要件を効果的にテストすることはできません。以下に詳述する方法論は、世界で最も洗練され、しばしば最大規模のデータドリブン企業で実行されるワークロードを、より効果的に模倣したものです。

## 業界で認められた基盤の構築

Teradataの方法論は、データおよびアナリティクス業界で一般的に受け入れられている、現在利用可能なベンチマークの1つ、Transaction Processing Performance Council (TPC) のTPC-H仕様から始めています。この仕様は、ワークロード比較のための以下の基本コンポーネントを提供します：

- テーブル構造とリレーションシップを含むデータスキーマ
- 標準化されたデータ作成のためのデータ作成ツール
- ベースとなる歴史的な分析クエリ

TPC-Hは、1990年代後半から2000年代前半にかけてTPCによって開発された意思決定支援ベンチマーク仕様です。ビジネスインテリジェンスやデータウェアハウスのシナリオに典型的な、大規模データセットに対する複雑でアドホックなクエリを処理するデータベースシステムの性能を評価するために設計されました。

今回、TPC-H仕様の22のビジネス指向クエリ群を利用しました。これらのクエリは、顧客、注文、製品を含む現実的なビジネスシナリオをモデル化した構造化データセット上で動作します。ベンチマークでは、クエリの実行時間、スループット、および価格パフォーマンスのメトリクスを測定します。

今日のデータ分析環境は、組織が過去に何が起こったか、現在何が起きているか、そして何が起こりそうかを理解することを可能にしなければなりません

## データドリブンなシナリオの考慮

この仕様に対して、実世界のデータドリブンシナリオをシミュレートするために設計された28のクエリが追加されています。TPC-H仕様の設計とは異なり、この強化されたワークロードの定義は、最新のデータドリブン環境で典型的、多様なクエリタイプとワークロードをシミュレートすることに重点を置いています。

このシナリオには、複雑な分析クエリ、レポートング、およびデータドリブン企業の要求を反映したオペレーショナル分析クエリが混在しています。この検証の目的は、同時並行的に混在するワークロードを効率的に処理するシステムの能力をテストすることにあります。このシナリオにより、現代のビジネスのダイナミックなニーズをサポートするデータベースシステムの能力を、より包括的に評価することができます。

## 質問の変更

この手法では、クエリは実行の各ラウンドで同じデータがリクエストされないようにパラメータ化されています。これは典型的なベンチマークのパフォーマンスチューニング手法で、キャッシュできるデータ量が制限されます。この手法により、データストレージ層のさまざまな部分にアクセスされるようになります。

## シリアルテストとスループットテストの比較

クラウドベースのデータ分析プラットフォームを評価する際、シリアルテストとスループットテストはそれぞれ異なる目的があります。シリアルテストは、しばしば「直線」テストと表現され、アメリカのモータースポーツのコンセプトであるドラッグレースのようなものです。ドラッグレースと同様に、シリアルテストは生のパフォーマンスメトリクスの狭い領域に焦点を当てます。このアプローチでは、個々のクエリが逐次的あるいは直列的に次々とクエリを処理するシステムの能力を評価するものです。

対照的に、スループットテストは、一定時間継続的に実行し、クエリを継続的に変化させることで、実世界のシナリオをシミュレートします。これは、ル・マン24時間レースのような耐久モータースポーツイベントのようなものです。このアプローチは、システムが並行処理を実行し、リソースを効率的に管理する能力を再現することを目的とします。本番環境レベルのワークロードをシミュレートするためのより適切な評価となります。

## 実環境ワークロードの作成

複数レベルのデータ分析コースケースに基づき、世界中のあらゆる規模のデータドリブン企業の課題に対応するため、この実環境ワークロードテストは以下の内容で構成されています：

- 複数のクエリタイプにまたがる50のクエリをシミュレートし、最新のデータドリブン型ワークロードを再現します。
- 2時間連続して実行されるスループットワークロード比較評価では、単にクエリのセットを実行するのではなく、クエリオプティマイザからワークロード管理、ストレージまで、クラウドベースのデータ分析プラットフォームのすべてのコンポーネントに負荷をかけます。

この多様なクエリの組み合わせは、複雑な分析クエリ、業務レポート、および戦術的クエリを処理する各プラットフォームの能力を、データドリブン企業の要求を反映した形でテストするために設計されています。

## コストの理解

クラウドベースのデータ分析プラットフォームでは、コスト効率が最も重要です。厳しい予算制約の下で運用するために、支出を増やすことなく分析機能を拡張するには、コスト効率の高いソリューションが不可欠です。これらのプラットフォームは、バッチ処理からリアルタイム分析まで、多様なワークロードを効率的に処理する必要があり、コストはこれらの多様な要件を管理する能力によって大きく異なります。

さらに、データの指数関数的な増大は、コストが急激に上昇する可能性があるという課題に直面します。データ量に比例してコストが上昇することなく、増大するデータを効率的に処理できるプラットフォームの価値はますます高まっています。このような状況において、コスト効率の高いデータ分析プラットフォームは、企業が少ないリソースでより多くのことを行うことを可能にし、リソースを最適化すると同時に、データドリブンで進化する分析ニーズに対応します。

クラウドベースのデータ分析プラットフォームは、アーキテクチャや設計思想が根本的に異なることが多く、直接比較することが困難です。各プラットフォームは、競合他社のものとは直接対応しない独自の機能、最適化手法、リソース配分戦略を持っている可能性があります。



## 基本単価の決定

このワークロード比較のために、Teradataは各システムの基本「処理単位」のコスト同等性を設定しました。この評価では、コンピュータアワー約 70 ドルをシステム等価性のターゲットとして使用しています：

Teradata: \$69.31	
Snowflake: \$72.00	
Databricks: \$72.80	

この目標範囲は、ベンダーの価格設定とノードの仕様の違いを反映しています。この70ドルという目標に対する正確な価格設定は、しばしば困難です。ここでいうコンピュータアワーとは、以下の条件下で各クラウドベースのプラットフォームを稼働させるための1時間あたりのコストを表しています：

- 一般的に利用可能なソフトウェアのバージョン
- 中規模プラットフォーム
- 1年間のサービス価格コミットメント

各システムは、ベストプラクティスと一般に公開されている設定ガイドンスを活用して、この予算制約の中で価値を最大化するように設定しました。

このアプローチは、各プラットフォームが与えられた予算内でその強みを活用できるようにしながら、コストベースを標準化することで公平な比較を促進します。これは、意思決定者がクラウドベースのデータ分析ソリューションを導入する際に行う実際の選択を反映したものです。

# エンタープライズ価格 パフォーマンスの算出方法

本番運用するデータ分析プラットフォームのエンタープライズ価格パフォーマンスを計算する簡単な方程式があります：

$$\text{[-x]} \text{ [+]} \text{ エンタープライズ価格パフォーマンス} = \text{運用コスト} / \text{完了したクエリ}$$

これは、クエリあたりのコストで測定され、本番ワークロードを実行したときにプラットフォームがどのようなパフォーマンスを発揮するかを表します。

VantageCloudの競合他社に対するエンタープライズ価格でのパフォーマンスを示すため、コストとクエリパフォーマンスを整理しました。ワークロード比較評価のコスト式は、上記のものとは若干異なります。ここでは、運用コストの代わりに、テスト実行にかかる1時間あたりのコストに注目します。また、テストで完了したワークロードの平均クエリ数（完了したクエリの代替）を使用します。これにより、クエリあたりのコストは以下の式となります：

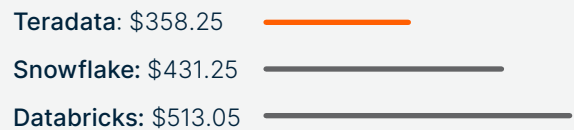
$$\text{[-x]} \text{ [+]} \text{ 価格パフォーマンス} = \text{1時間あたりのテスト費用} / \text{1時間あたりのクエリ完了数}$$

## 実際の費用

最初のベースとなる“単位”の処理については、コスト同等性を確立していますが、実際のテストコストは、プラットフォームの運用性能によって異なります。

SnowflakeやDatabricksのような、より積極的なスケーリングパラダイムを持つプラットフォームは、処理需要の増加に応じてコンピュートリソースを迅速に追加提供します。対照的に、Teradataは、より洗練されたソフトウェアにより、利用可能なコンピュートリソースを効率的かつ完全に活用した上で追加リソースを要求します。Teradata VantageCloudは、特許取得済みの機能「高度なクエリ最適化と業界をリードするワークロード管理」を活用することでこれを実現します。

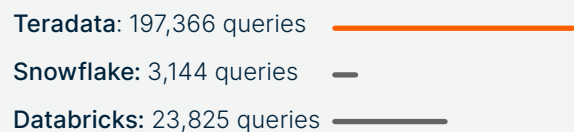
その結果、2時間の作業量を比較した場合の実際のコストは大きく異なります：



これらの数値は、各プラットフォームのスケーリング戦略とリソース利用効率が、同等のベースラインからスタートしたにもかかわらず、実際のシナリオにおいて全体のコストにどのような影響を与えるかを反映しています。

## 完了したクエリ

次に、各プラットフォームを、先に説明した実環境のワークロードを、固定された2時間にわたって実行します。現実のクラウドの弾力性をシミュレートするために、システムはコンピュートリソースの需要を管理するために、最大4つの「ユニット」まで処理をスケールアップすることが許可されています。結果は、1時間あたりの平均クエリ完了数で集計しています：



クエリスループットの違いは、リソース利用、クエリ処理、拡張性の管理と最適化におけるプラットフォームの能力の違いを浮き彫りにしています。Teradataの完了したクエリ数が多いことは、このワークロードの処理効率が優れていることを示唆しています。

# クエリあたりのコスト

次に、上記のコストとクエリスループットのデータを組み合わせ、価格パフォーマンス方程式を作成します：

**[-x]** **価格パフォーマンス=**  
**[+三]** **1時間あたりのテスト費用 / 1時間あたりのクエリ完了数**

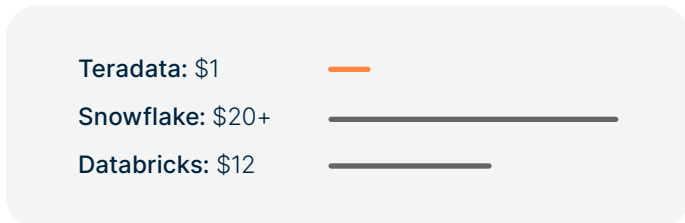
これで、各クラウドベースのデータ分析プラットフォームの価格パフォーマンスを判断できるようになります。この指標は、分析機能を提供し、大規模ワークロードを処理する際の各プラットフォームのコスト効率を定量化するものです。最初の結果は、クエリあたりのコスト（ワークロード処理のコスト効率）で評価されます：



この結果をより理解しやすくし、価格パフォーマンスをよりよく示すために、Teradataのクエリあたりのコストを基準として、コストを1.00ドルに正規化します。この正規化により、相対的なコスト効率をより明確に把握することができます。

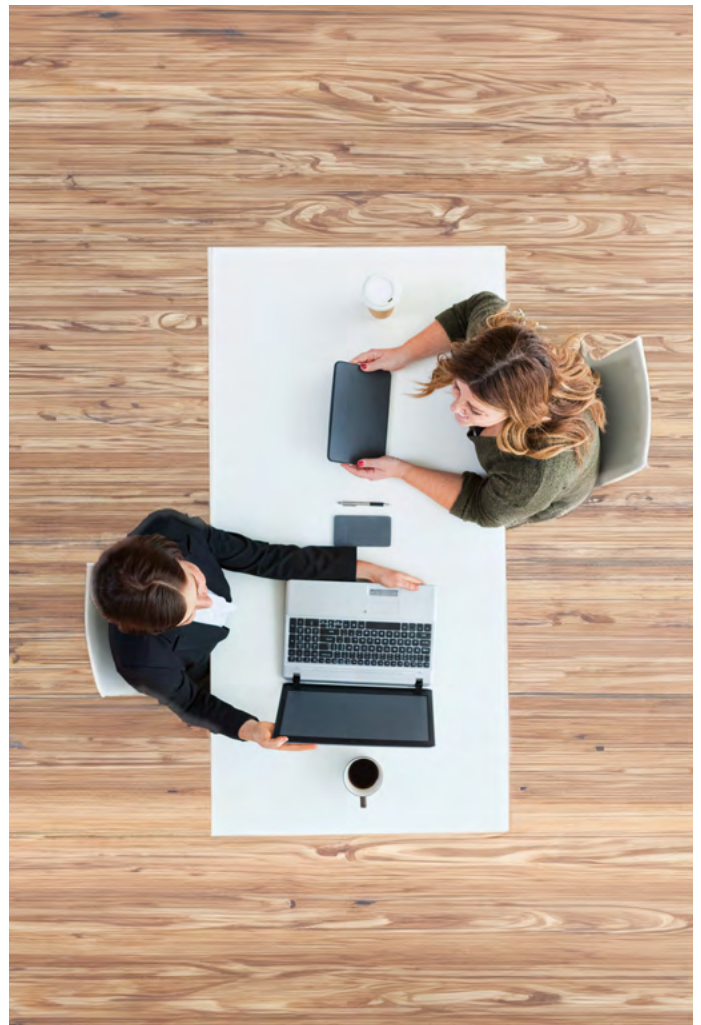
各プラットフォームは、ビジネス関係者からIT部門、財務部門に至るまで、さまざまな意思決定者の複雑で多様なクエリを処理することができます。

この正規化によって、Teradata VantageCloud Lake が現実のワークロードを処理するために費やした1ドルにつき、他の比較プラットフォームでは、以下のコストが必要であることが明らかになりました：



これらの数字は、プラットフォーム間のコスト効率の大きな違いを明確に示しています。Teradataはこの現実のワークロードシナリオにおいて、価格と性能の面で大きな優位性を示しています。

これは、分析機能が大規模に提供される際のコスト効率が優れていることを示しています。この分析は、企業がクラウドベースのデータ分析プラットフォームを選択する際に、TCOとROIに与える潜在的な影響を理解するのに役立ちます。





# 結論

この調査は、主要なクラウドベースのデータ分析プラットフォームのエンタープライズ価格パフォーマンスに関する貴重なインサイトを提供します。実際のワークロードミックスとコストについて同等環境での厳密なテストを通じて、これらのプラットフォーム間で効率性、拡張性、費用対効果に大きな違いがあることを実証しました。

この比較において、Teradata VantageCloud Lakeは、優れたコスト効率とクエリスループットを示し、明確なリーダーとして浮上しました。競合他社の数分の一のコストで非常に多くのクエリを処理するその能力は、クラウド環境における高度なワークロード管理とクエリ最適化機能の重要性を強調しています。

これらの調査結果は、マーケティングの広告メッセージを鵜呑みにせず、コスト効率、拡張性、多様なワークロードへの適応性など、総合的な価値提案を検討することの重要性を浮き彫りにしています。価格パフォーマンスの劇的な違いは、慎重なプラットフォームの選択によってコストを大幅に削減し、パフォーマンスを改善できる可能性を示しています。

より多くのクエリを実行できる企業は、より多くのデータを分析することができ、より正確な予測とより良い顧客体験を実現することができます。また、顧客離反や詐欺行為などのリスクを未然に回避することも可能になります。

データドリブンな意思決定を追求する上で、インサイト力を導き出すコスト効率は、インサイト力そのものと同じくらい重要なのです。

クラウドベースのデータアナリティクスに対する要求が進化するにつれて、このような評価は、企業のデジタルトランスフォーメーションやクラウドモダナイゼーションの旅の指針となります。

Teradataは、優れたソフトウェア、効率的なリソース管理、特許取得済みのワークロード管理によって、パフォーマンスと費用対効果の両面で大きな利点を提供できるのです。



## Teradataについて

Teradataは、信頼できる情報によって企業、人々が成長すると信じています。そのために、AI向けの最も包括的な機能を搭載するクラウドデータ分析プラットフォームを提供しています。

統合された信頼できるデータ、信頼できるAI、より迅速なイノベーションを提供することで、Teradataはお客様、そしてお客様のお客様が、より良く、より確信に満ちた意思決定を行えるよう、支援しています。Teradataは、あらゆる主要業種において、世界のトップ企業から信頼を獲得しています。

Teradataは、より迅速で柔軟なデータ統合と、信頼性が高くコスト効率の高いAIを活用したイノベーションによって、世界中で毎日何億人もの人々にポジティブなインパクトを与えています。

Teradataがお客様のデータ分析をどのように改善できるか、また[ワークロード比較の詳細について](#)、[エキスパートにご相談](#)ください。