

データはデータでしかない、 データの統合が差別化をもたらす



本書概要

- ビッグデータの取り組みの多くは、既存の分析ソリューションが放棄され、新しいものが優先されている
- 新しい分析テクニックを採用する際は、既存のインフラストラクチャを活用すべきである
- Teradata® Unified Data Architecture™ は、企業データの統合かつ360度ビューの実現を支援する

Datamation® Executive Brief

はじめに

ビッグデータ分析は、昨今多くの企業において最優先事項となっています。多くの企業は、利用可能なすべてのデータから新たな洞察を引き出すことを希望しています。その目的は、生産性の向上、コスト削減、顧客離反の減少、顧客体験の向上、そして新たなビジネスの機会をつかむことにあります。

残念なことに、ビッグデータの取り組みをゼロから始める企業は、往々にして自社の既存の分析ソリューションを放棄し、新しいものを優先する傾向があります。典型的なのが、関心のあるすべてのデータを1つの共通ファイル・システムにまとめ、そのデータに対して新たな「ビッグデータ」分析を実行するというアプローチです。このアプローチでは、多くの企業がHadoop、Pig、Python、その他のビッグデータ技術の実装を急ぎ、プロセスを統合して既存のデータ探索および分析ツールを組み込むことを考慮せずに行います。その結果、ビッグデータの取り組みから引き出される情報を必要とする人たちは、新たなツールの使用方法を習得しなければなりません。さらに、企業は特定のツールの使用にとらわれているため、しばしばデータに対して狭い見方しかできなくなり、希望していた新たな洞察を引き出すことが妨げられています。

これらの課題を回避し、ビッグデータ分析を最大限活用するために、企業は多様なデータと分析技法に適応しながらも既存のインフラストラクチャを活用するアプローチを採用する必要があります。特に、データの収集、蓄積、処理を行なう現行システムから真のビッグ・アーキテクチャへと企業が移行するためには、Hadoopだけでは目的を果たせません。企業には、柔軟でオープンなディス



カバリー・プラットフォーム、統合データウェアハウス、そして大規模なデータセットをコスト効率よく蓄積および処理するデータ・プラットフォームという、3つの主要要素が必要です。この3つの主要要素により、企業はビッグデータ分析の利点を実感できるようになります。

ビッグデータ分析の課題

ビッグデータ分析には、企業の業務改善、より迅速でよりインテリジェントな意思決定を下せるように支援する潜在能力があります。そのような理由から、ビッグデータ構想は爆発的な増加の構えをみせています。しかし、あらゆる新技術と同様に、ビッグデータがもたらす利点を企業が十分に実感するのを妨げる障害が存在します。

まず、**拡大するユーザーのニーズを満たすという課題があります。昨今、多くの企業では、ビッグデータ分析の取り組みから引き出される意思決定を支援する**

情報へのアクセスを必要とするユーザーが増加しています。データサイエンティストに加えて、企業の経営陣や各事業部門はビッグデータ分析の活用を望んでいます。

このような状況においては、適切なビッグデータ・アプローチが選択されない、いくつかの問題が発生する可能性があります。情報へのアクセスを必要とするユーザー数の増加は、レポートの作成やダッシュボードの構築をIT部門とアナリストに頼っていた従来の分析アプローチではもはや立ち行かなくなることを意味します。

データサイエンティストだけが高度な分析業務を担っていた頃には、データサイエンティスト自身が分析ワークフローのかじ取りができるようになるためのトレーニングを受け、専門知識を備えていました。彼らは、関連性の高いデータの発見、抽出、処理を行なって結果の意味を理解するための方法を習得していました。企業

の経営者やビジネス・マネージャには、必要な技法やツールすべての使用方法を習得するための時間も専門知識もありません。しかも、それらの仕事をIT部門に代行させることは実用的ではなく、どのプロセスにも数週間、数ヶ月にわたる期間が加算されることになります。

2番目に検討すべき課題は、データの増加に関連するものです。企業が意思決定のために利用するデータの量は爆発的に増加しています。数々の業界研究調査によると、データ量は12ヶ月から18ヶ月ごとに倍増していると報告されています。データの増加を促進している付加的要因が、非構造化データ・ソース(Eメール・メッセージ、通話記録、文書ファイル、スプレッドシート等)やソーシャル・メディア・ストリームなどの、以前は利用されていなかったデータセットから洞察を引き出すことの認知度の上昇です。結果として、日常的に収集および保存されるデータの量が桁違いに増加しています。これらのデータの多くが、長期間にわたる保持を必要とします。

さらに、ビッグデータの取り組みには非常に多種多様なデータ・タイプを利用する必要が伴うことに気付く企業が増えつつあります。あるプロジェクトでは数万のWordファイルからのテキスト分析が伴うかみならず、別のプロジェクトでは1つの巨大なデータベースからのパターン探索が伴うかみられません。各アプリケーションでは、データの収集、抽出、蓄積に関する異なる要件が設けられています。

その上、分析の取り組みの多くは、過去に注目して動向をつかむことに重点が置かれてきました。この商品はあの店舗でどれくらいよく売れたのか?このEメールを受け取った顧客のうち、メールに埋め込まれたリンクをクリックし、購入や取引を完了させた人は何人いたのか?どのサプライヤの製品が、生産ラインにおいて最も高い故障率の原因となったのか?このような種類のクエリーに対しては、構造化されたデータベースまたはデータウェアハウス内に整然と収集および格納されたデータを分析することによって応答できました。

現在では、ビッグデータ分析によりこれらのクエリーに対してさらに洗練された分析結果を生成することが可能になっています。しかし、企業からは別の種類の質問も発せられるかもしれません。ソーシャル・メディア・ストリームからの情報により、変化する顧客の感情に関して今まで以上にタイムリーな情報がもたらされるのか? 位置情報を含んで内容が豊かになった取引データが、リアルタイムのマーケティングや不正検知に役立つのか? といった例です。

3番目に、多くの企業がITの人材不足を経験しています。ビッグデータを考慮に入れなくても、IT部門は過剰な仕事でパンク寸前になっています。ここ数年、IT予算は横ばい、またはささやかな増額しかされていません。このような状態は、新たなITサービスに対する需要が急増した時に発生してきました。多くの企業において、IT部門には、スタッフの数や予算が変わらないまま、または削減

された状態で、さらに多くの仕事をすることが求められています。

ビッグデータ分析プロジェクトが構想および実装されても、IT部門は、エンドユーザーや各事業部が満足するような期間内で人材を投じることができないかもしれません。さらに、ビッグデータ構想の中には、一層のIT要員不足を引き起こし、成果が出るまでの時間を長引かせるだけの複雑性の増加をもたらす得るものがあります。

4番目に、ビッグデータの取り組みにはさまざまな新しい技術を組み込むことが必要となります。ビッグデータの取り組みの多くは、非常に広範囲にわたる技術を利用することでデータの収集とアクセスを行ない、SQL および Non-SQL 分析をベースとした新たな分析メソッドを使用し、より関連性の高い洞察をデータから引き出すための新たな技術を必要とします。単一の分析エンジンやデータ・インフラストラクチャの要素だけでは、企業全体のさまざまな業務分野を横断してビッグデータ分析を適用する柔軟性が企業にもたらされません。過去に使用したことのない特定のツールの使用を従業員に強いるためには、高コストで広範囲のトレーニングが必要になります。同時に、従業員を特定のツールの使用に縛り付けると、データの検証や確認の方法が限られてしまう可能性もあります。

「Teradata Unified Data Architectureは、企業がデータに対する統合的な360度ビューを確保するための支援をする。」

何が必要とされているか？

これらの懸念や課題には、ビッグデータ分析アーキテクチャを正しく選択することによって対処できます。

入念に計画されたアプローチは、データの抽出および分析ツールの選択肢を企業に与え、柔軟性の欠けたソリューションの利用に従業員を縛り付けるようなことはしません。さらに企業に必要なのは、ビッグデータ分析ワークフロー内において既存のソリューションに従業員に利用させるアプローチです。

当然ながら、ビッグデータ分析ソリューションはデータの収集、蓄積、加工するために Hadoop のような新しい技術を利用できるようにするものでなければなりません。ビジネス・マネージャが自らの分析作業において使用することが見込まれる広範囲のデータの管理に役立てるためには、新たなディスカバリー・プラットフォームを利用する必要があります。そして、いかなるソリューションであっても、データがアクセス可能で管理しやすいものにするためには、統合データウェアハウスを利用する必要があります。

何よりも、エンドユーザーの課題に対処する必要があります。ソリューションは、使いやすいものでなければならず、従来のデータサイエンティスト以外にも、ビジネス・アナリスト、マーケティング・エグゼクティブ、事業部門のマネージャ、顧客対応担当の従業員にまで、ビッグデータ分析が拡大していくことを可能にする必要があります。

そのような特徴を示すソリューションは、以下のような数々の利点をもたらします。

- 俊敏性 — 従来型のアプローチではよく見られていた、IT部門にかかる負担の大部分を取り除く

- 単純性 — 未加工形式のままデータを利用できるようにすることで、時間を短縮し、エラーを削減する
- 柔軟性 — データに対してさまざまな分析技法(分析エンジン)を使用できるようにし、企業がデータを異なる角度から見ることを可能にし、新たな質問をして新たな洞察を引き出せるように支援する
- アクセス性 — ユーザーが既存のツールやスキルを使ってあらゆるデータにセルフサービスでアクセスできるようにする

企業の技術パートナーとしての テラデータ

多くの企業では、ビッグデータ分析の実装に困難を感じています。分析ワークフローのほぼ全段階に課題が存在します。

特に、多くの企業が困難に感じているのは以下になります。

- 各事業部が使用を望んでいる膨大な量の多構造化データを扱い、処理すること
- 既存インフラストラクチャを通じて重要なデータにアクセスし、つなぎ合わせる
- 将来のビジネス構想を推進するために、利用可能なデータから洞察を引き出すこと
- 競合優位性を生み出す目的で、すべての新たな技術を効果的に利用、管理、サポートすること

必要なのは、企業が利用可能なデータを活用し、データを理解した上で業務を遂行できるようにする統合的なソリューシ

ョンです。既存の分析とビッグデータの取り組みに特有の新しいソリューションが付随するインフラストラクチャのソリューションとを相乗的に組み合わせるものでなければなりません。

これこそ、テラデータが役に立てる分野です。

テラデータは、ビッグデータという名称が使われる前からビッグデータ市場に参入していました。テラデータは、大規模なデータセットを使用して有意義なビジネス情報を引き出せるように企業を支援するソリューションの提供を何十年間も経験してきました。

ビッグデータの取り組みを開始または準備するにあたり企業が現在直面している課題に対処するために、テラデータでは Teradata Unified Data Architecture を提供しています。これは、3種類のエンジンを統合し、いかなる種類のデータに対しても最も目的に合った分析能力を提供するアーキテクチャです。

Teradata Unified Data Architecture は、企業が自社のビッグデータの取り組みにおいて最高クラスのソリューションを活用できるようにする統合化されたソリューションです。これを実現する段階に至るまでにテラデータは、Teradata QueryGrid™、BYNET®、Teradata Viewpoint、Teradata Unity、SQL Assistant などの主要な実現技術と、透過的なアクセスやシームレスなデータ移動、Unified Data Architecture の一元管理を可能にするグローバル・サポートモデルを統合しています。

Teradata Unified Data Architecture により企業は、広範かつ大規模なデータセットに対する反復的な分析を迅速に実行できるようになります。分析には、

SQL、SQL以外の言語、そしてビジネス・ユーザーによる標準SQLでのMapReduce関数の実行を可能にするAster SQL-MapReduce®の使用が可能になります。

大容量の多構造化データに対してSQL分析やNon-SQL分析を実行するためには、特化した技術が必要となります。いくつかの他のアプローチにおいては1つのソリューションにまとめて閉じ込めることが必要とされますが、テラデータではビジネス・ニーズに対処する最高クラスのソリューションを3種類提供しています。

ソリューションは、以下のとおりです。

データウェアハウス: 30年以上にわたる開発努力が背景にあるTeradataデータベースは、全社規模での戦略的分析と実務的分析を実現し、ユーザーが単一ソースの一貫した集中型統合データにアクセスができる業界をリードするプラットフォームです。

データ・ディスカバリー: Aster SQL-MapReduceは、多構造化データと構造化データに対する反復分析を通じたデータ探索機能を、広範かつ大多数のビジネス・ユーザーに提供します。事前にパッケージされた分析関数により、ビジネスにおけるデータ主導型の探索を即時に開始することが可能になります。企業は標準のSQLを使用して、再利用可能なカスタム分析コードを作成し、幅広いユーザーに対して展開することができます。

Teradata QueryGridは、複数システムの分析に対してユニークなアプローチを採用しており、システム間でのSQLを通じてすべてのデータおよび分析にアクセスする際のシームレスなセルフサービスの使用感をユーザーに提供します。ユーザーは、既存のSQLスキルやツールの利用を続けることができます。そのため、改めてトレーニングを受けなくても、まったく新しいタイプの分析を実行できるようになります。

データ・プラットフォーム: Teradata統合ビッグデータ・プラットフォームは、データ単位当たりのコストを抑えた上でビジネス分析を大量の詳細データに適用するための、完全な統合ソリューションを提供します。オープンソースのHadoopソリューションが必要とされる場合、テラデータでは、標準のHadoopアプライアンスの枠を越えて設計およびサポートされているHadoopアプライアンスを提供しています。テラデータでは、全Teradata製品に対するシステム管理、データ・アクセス、ワンストップ・サポートといった強力なツールを、Hortonworks Hadoopにも適用し統合しています。

Teradata Unified Data Architectureは、企業がデータに対する統合的な360度ビューを確保するための支援をします。これにより企業は、以前は見えていなかった機会を、実用的な洞察へと変えることができるようになります。

このソリューションは、より賢明なビジネス上の決定を下すために使用されるデータを、企業がより効果的に編成、蓄積、活用する上で役立ちます。データセット

に対して多様な分析エンジンを使用することで、企業は、以前は獲得が不可能であったより深い新たな洞察を引き出せるようになります。

基本的には、Teradata Unified Data Architectureは、未加工のデータ・ソースと企業内でそのデータを利用する人たちの間に位置します。このソリューションは、企業がデータを移動、管理、アクセスする支援をします。これにより、あらゆる種類のユーザーが自分の分析ツールおよびアプリケーションをさらに容易に実行できるようになり、ワークフローが一層最適化され、洞察獲得までの時間が短縮されます。

Teradata Unified Data Architectureの主要な価値は、データ(大規模なものと同規模なもの、およびあらゆる組み合わせを含む)を有用で実用的な洞察へと転換することにあります。これには、以前は未知であったパターンを発見することを目的とした分析アプローチや、商品購入または解約の決定のような顧客の行動の引き金となる主なイベントの識別などが含まれます。



企業は、自社のビッグデータ分析の取り組みをサポートする複数の中核技術をまとめ上げることもできますが、テラデータでは、ビジネス価値を実現するように設計された統合ソリューションを提供しています。端的に言えば、テラデータは、

強力な製品およびサービスから成る幅広いポートフォリオを提供して、企業が分析によって最高の価値を獲得できるように支援します。ユーザーに対してはデータへのアクセスを透過的にし、管理者に対してはデータの移動とシステムの管

理を簡略化します。

Teradata Unified Data Architectureの詳細については、<http://jpn.teradata.jp/product/unified-data-architecture/>にアクセスしてください。

「Teradata Unified Data Architectureによって企業は、以前は見えていなかった機会を、実用的な洞察へと変えることができるようになる。」

お問い合わせ

日本テラデータ株式会社

〒107-0052 東京都港区赤坂2-23-1 アークヒルズ フロントタワー

<http://www.teradata.jp/>

- Unified Data Architecture、Aster SQL-MapReduce、およびTeradataはTeradata Corporationの登録商標です。本文中の社名、製品名は各社の商標または登録商標です。
- 本資料に掲載されている情報は予告なしに変更されることがあります。
- 最新情報については Teradata 担当者にお尋ねください。