

Teradata Vantage

ビジネスに必要な次世代アナリティクス・プラットフォーム

企業はアナリティクスを十分に活用できていない

「ビッグデータ」が注目され始めたのは2010年頃でした。そして今、日々蓄積されるビッグデータを分析し、新たな知見を提供できなければ、企業は厳しさを増す市場競争の中で敗者となってしまいます。企業は常にデータを活用し、ビジネスが抱える困難な課題を解決する適切な「答え」を探さなければなりません。

企業はこれまでも、答えを求めてアナリティクスに多くの投資をしてきました。それにも拘わらず、その投資に見合ったビジネス成果を得られていない現状があります。2018年にテラデータ・コーポレーションが実施^{*1}した、ビジネスおよびIT部門の幹部社員を対象にした調査において、自社のアナリティクス技術が複雑で使いにくいと答えた人が74%に達しました。これは、長年に亘る投資で複数のデータベースやBIツールなどが導入され、アナリティクス環境が複雑化しているからです。複雑さが増していることで、アナリティクスを行う組織の負担が増えていると回答した人も53%と半数を超えています。

さらに問題なのは、ユーザーが必要なデータにタイムリーにアクセスできないことです。この調査では、効率的に仕事をこなすために現状よりも多くのデータが必要だと回答が79%もあり、一方で必要なデータの約3割が失われていると考えられています。

このように企業でアナリティクスを十分に活用できていない理由の1つがデータのサイロ化です。企業では職種や役職ごとの要求に応じ、順次アナリティクス環境を用意してきました。その結果、経営層のための環境や経理部門や営業部門のための環境が別々にできあがってしまったのです。各々の環境がデータをコピーして保持しているため、データの重複や不整合も発生します。さらに、重要な情報を守るためのセキュリティやアクセスコントロールなど、データガバナンスの確保もままなりません。

あらゆるデータを蓄積するデータレイクを構築しても、この状況はあまり改善されません。そもそも増え続けるデータ全てをデータレイクに集めるには、多くの投資が必要です。仮にデータが統合できても、膨大で多様なデータに柔軟かつ迅速にアクセスできるようにするのは、非常に難しいのです。結果的に部門などの要求に応じてデータレイクから適切なデータを抽出し、アナリティクス環境を別に用意せざるを得ません。

また役職や職種ごとに分析の目的が異なるため、さまざまなアナリティクスツールを導入してきたことも大きな問題です。経営企画部門などでは、アドホックなデータ分析環境が必要ですし、データサイエンティストは、高度なデータ分析を自分好みのツールで行いたいと考えます。一方でビジネスの現場はなるべく手間をかけずにデータ分析を行いたいと考えます。結果的に複数のBIツール、さまざまな分析エンジンを運用することになるのです。IT部門はこれらの運用に多くの手間をかけています。この状況は、まさにアナリティクスのサイロ化と言わざるを得ません。

次世代アナリティクス・プラットフォームに求められる3つの要件

企業のあらゆる立場の人たちが必要なデータにタイムリーにアクセスでき、ビジネスに貢献するアナリティクスを実現するにはどうすればよいのでしょうか。そのためには、データのサイロ化、およびアナリティクスのサイロ化を解消するアナリティクス・プラットフォームが必須です。このアナリティクス・プラットフォームはデータを統合し、迅速に分析できるだけでなく、新たに満たすべき3つの要件があります。

1) オープンソース技術の柔軟性と俊敏性を取り入れること

AIや機械学習など最新のアナリティクス技術の多くは、今やオープンソース・ソフトウェアで提供されています。既存のBIツールなどはもちろん、これらオープンソースの技術を柔軟に連携できるプラットフォームが必要です。さらにオープンソース・ソフトウェアを取り込んだ環境を、企業が日常的に安心して運用できるサポート体制も必要となります。

2) すべてのデータを活用するためにデータ構造を明らかにし、データガバナンスが考慮されている

アナリティクスに利用するデータの構造は、誰にでも分かりやすい形で表現されている必要があります。データ構造が明らかであれば、ユーザーのニーズに合わせ柔軟にデータを取り込み活用できます。さらに取り込んだデータのセキュリティを保障し、適切なアクセスコントロールが行えなければなりません。データの取り込みやアクセスのした経緯などもすぐに確認できるようにして、データガバナンスも保障できなければならないのです。

*1: テラデータ・コーポレーションが依頼をしたテクノロジー専門リサーチ会社Vanson Bourneが、2018年8月～9月に実施したものです。

<https://www.teradata.jp/Press-Releases/2018/Global-Survey-Analytic-Insights-Remain-Trapp>

3) ユーザーのさまざまなニーズを満たす先進的で統合されたアーキテクチャ

シンプルで統合されたアーキテクチャによる、先進的なアナリティクス基盤が必要です。並列処理や分散処理を実現するために、仮想化技術や、コンテナ技術が必要と考えられます。統合されたアーキテクチャのアナリティクス基盤により、アナリティクスをバラバラに導入するよりも効率化がなされ、トータル・コストの管理も容易となります。

これら3つの要件を満たす次世代のアナリティクス・プラットフォームが、Teradata Vantage (以下、Vantage)です。Vantageはシンプルなアーキテクチャで、あらゆるデータを大規模に統合でき、迅速に価値ある答えを導き出します。Vantageはパブリッククラウド、ハイブリッドクラウド、オンプレミスなど、さまざまな環境で動かすことができます。これにより、ビジネスに直結する価値ある分析環境が迅速に実現できるのです。Vantageならビジネスにおける重要な課題を解決し、ビジネスにイノベーションをもたらすような高度なアナリティクスも容易となります。

高度で多様なニーズにも対応する Teradata Vantageの分析機能と分析エンジン

Teradata Vantageには、データ活用で豊富な実績のあるTeradata Databaseと、高度で多様な分析を得意とするTeradata Asterの技術が統合されています。Vantageの新型SQLエンジンは、従来の高速並列処理を実現する大規模並列処理(MPP)アーキテクチャを踏襲しており、膨大で多様なデータを迅速に処理できます。

Vantageは並列処理と分散処理を実現するために、コンテナ技術、Kubernetesを採用しています。これにより多様な分析ニーズにも柔軟に対応できるのです。

SQLエンジン

Vantageでは、従来データウェアハウスに蓄積してきた基幹系システムの構造化データはもちろん、それ以外のIoTのセンサーデータやソーシャルネットワークから生まれる非構造化データなども、ストレージに蓄積されます。統合的に蓄積されたデータには、VantageのSQLエンジンを經由して自由にアクセスできます。またVantageの外にあるデータについては、Teradata IntelliSphereのデータ連携機能で、Vantageのストレージにあるデータと同じように扱うことができます。つまりVantageに接続するだけでSQLエンジンが全てのリクエストの入り口となり、データサイエンティストはデータがどこにあるかを意識する必要がなくなるのです。

このVantageの新しいSQLエンジンには、以下のようなさまざまな機能が搭載されています。

4D Analytics

時系列の3次元の地理空間データ、すなわち4次元データおよび従来の時系列データを組み合わせた高度なアナリティクスを実現します。人やクルマなどの移動体の位置データ(GPSデータなど)と、それに付随する温度の変化などの時系列データを用い、移動体の傾向などの分析を容易に行うことができます。

nPath

順序を持つ列データから、指定したパターンにマッチした行シーケンスを抽出します。条件にマッチした列記号として表現され、列パターンマッチングにはその記号列を要素とする正規表現を利用できます。分析結果を工程間の流量を表現するSankeyダイアグラムで可視化でき、優良顧客や価格センシティブ顧客などを特定するパスの発見などにつなげることが可能です。

セッションイズ

Webサイトにおける訪問者のクリックを伴う行動履歴に関して、各ページの訪問者のクリック行動をユニークなセッションIDにマッピングすることによって分析します。セッションの特定には、前回のクリックから次のクリックまでの時間間隔を最大時間として設定します。この時間を過ぎたクリックは、新しいセッションとみなします。ボットによるアクセスなどを除外するために、最小クリック間隔(RAPIDFIRE)を指定することもできます。

アトリビューション

ECサイトなどの訪問者のクリックや購入といった、特定イベントに至る行動に重みを割り当てます。そしてさまざまなモデルタイプと分布モデルを仮定し、指定したコンバージョンに至るまでの各ページのアトリビューション(貢献度)を計算できます。多数のモデルタイプ、および分布モデルが用意されており、必要なパラメータを設定できます。分析では異なる2つのモデルを併せて設定できます。

時系列(Time series)

時系列データを分析する機能です。時系列モデルが実装されている関数を用い、時系列モデルのパラメータ推定およびスコアリングが行えます。

スコアリング機能

ツリー構造を用いて分類や回帰を行う、決定木モデルで推定を行うための関数です。

機械学習エンジン

Vantageにはさまざまなビジネス上の課題を解決するために、パス、パターン、統計、およびテキスト分析を行う機械学習エンジンがあります。100を超える機械学習のための関数が用意されており、それらを利用してユーザーの行動データから感情を理解したり、機器の故障予測を行ったりと高度な分析が可能です。

グラフエンジン

Vantageに搭載されているグラフエンジンを使うことで、人や製品、プロセス間の関係などを視覚化し、明らかにすることができます。事前に定義された関数群を使い、ソーシャルネットワークでの接続関係、インフルエンサーとの関連性、詐欺行為や未知の脅威の検出など、通常の分析では難しい複雑な問題も迅速に解決できます。

好みのツールと言語の利用

日々分析を行っている経営企画部門の担当者や、高度な分析を行うデータサイエンティストなどには、使い慣れた分析ツールやプログラミング言語があります。Vantageは既存の各種BIツールなどを利用できるだけでなく、SQLでプログラミングを行うことはもちろん、さまざまなプログラミング言語を使うことが可能です。SQL、R、Pythonなどの分析で使われている主要なプログラミング言語に対応しており、GUIツールとしてはJupyter NotebookやRStudioなどオープンソースの世界で人気のあるツールや、Teradata Studio、Teradata AppCenterに対応しています。このようなVantageのアナリティクス・エコシステムにより、どのようなユーザーにとっても最良のユーザー体験を提供することができます。

Teradata Vantageで顧客が得られるメリット

1) シンプルなデータアクセス

データと分析のエンジンが1つのプラットフォームに統合されていることで、システム間のデータ移動がなくなり迅速性が得られ、さらにデータの重複や不整合の心配もなくなります。

2) 最先端の技術で高い処理能力を発揮

高速並列処理を実現する大規模並列処理 (MPP) アーキテクチャのプラットフォームとなっており、大規模なデータに対し高度で複雑な分析処理を迅速に行えます。

3) さまざまなアナリティクスツールを利用可能

オープンソースの新しいフレームワークやアナリティクスツール、よく使われている機械学習用のプログラミング言語などを自由に選択できます。コンテナ、仮想化の技術を活用し、最新の商用 (プロプライエタリ)、およびオープンソース技術を容易にVantageのプラットフォームに統合でき、最新で最良のユーザー体験が得られます。

4) アナリティクス環境の迅速な展開

信頼性、セキュリティ、可用性、データガバナンスを備えており、さらに管理性の高いプラットフォームを提供します。長年のテラデータの実績に基づいた高度な分析モデルをVantageで利用すれば、本番環境において素早くアナリティクス環境を展開できます。

多様なデータタイプに対応

Vantageは、実績あるデータストアを利用し、永続性のある堅牢なストレージを提供します。さらにAmazon S3やMicrosoft Azure Blob Storageなどのオブジェクトストレージやその他のストレージとの連携機能も用意されています。対応するデータタイプは構造化データのリレーショナルはもちろん、空間、時間、さらにはXMLやJSON、新しいデータ交換形式のAvroなどさまざまなものに対応します。さまざまなデータタイプに対応でき、好みのツール、言語を利用できることで、Vantageは多様な分析の要求に対し柔軟な選択肢を提供できるのです。

将来の変化にも柔軟に対応可能

VantageはTeradata Everywhereのコンセプトに基づいて開発された、ソフトウェア製品です。オンプレミスのTeradataのハードウェアおよび汎用ハードウェア、パブリッククラウド、Teradataクラウド*2などあらゆる環境で動かすことができます。Vantageはソフトウェアライセンスおよびサービス (as-a-Service) として提供します。また、ソフトウェアライセンスは、柔軟な環境移行が可能です。オンプレミスで始めてパブリッククラウドに移行する、あるいはパブリッククラウドで試し本番はプライ

*2 : 2019年2月現在、Teradataクラウド (IntelliCloud) はアメリカで提供しています。

ベートクラウドで運用するなど、分析環境をその用途に応じて自由に移動して利用できます。

Vantageを利用することで、ビジネスユーザーやデータサイエンティストは、アナリティクスのアルゴリズム開発や分析モデルの構築に注力できるようになります。Vantageは単に統合化されたアナリティクスプラットフォームではなく、アナリティクスをアジャイルに継続するAnalyticOpsの実現にも貢献します。これにより企業はアナリティクスのリソースやコストを最適化でき、ビジネスにインパクトを与える答えを容易に得られるようになります。

小売業でのTeradata Vantage採用事例

世界的な規模を誇るある小売企業では、Vantageを活用することでECサイトの顧客に最適化リコメンドを行えるようになりました。ECサイトは同社の売上げの25%を占めていますが、以前は不正確な検索結果を提供して顧客体験を損ない、貴重な販売機会を失っていました。そこでVantageの機械学習アルゴリズムを利用し、ユーザーの検索などの行動データを蓄積、それを解析してユーザーニーズに適合した最適な検索結果を表示するようにしたのです。新しいリコメンドエンジンにより、2ヶ月間のセール期間中に優良顧客からの収益が130万ドルも増加する結果も得られました。

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 www.teradata.jp

TeradataおよびTeradataのロゴは、米国テラデータ・コーポレーションの米国およびその他の国における登録商標です。テラデータでは、新たなテクノロジーおよびコンポーネントの導入に合わせて製品を随時改良しています。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。本書に記載された特徴、機能、および運用形態は、地域によっては販売されていない可能性があります。詳細については、テラデータ担当者にお問い合わせいただくか、Teradata.jpをご覧ください。

© 2019 Teradata Corporation All Rights Reserved. 02.19 EB9959



teradata.