



モダン 分析プラットフォームの 構成要素

Dan Kogan, Jen Underwood

「モダン分析」の特徴と必要とされている理由

私たちは技術革新が急速に発展し、国際競争が激しさを増し、今までなかった物事のあり方を変え新たに考案するチャンスに囲まれた、非常にエキサイティングな時代に生きています。デジタルテクノロジーの急速な成長、至る所で利用可能なインターネット接続、そしてスマートデバイスの激増が第4の産業革命を引き起こしており、この環境により生み出される溢れかえるほどのデータを素早く正確に理解できれば、競争的に優位な立場に立つことができます。

モダンテクノロジーの可能性には計り知れないものがあります。物理的世界と仮想世界の境界はなくなりつつあり、市場参入への壁が低くなり、全く新たなビジネスモデルが生まれています。どの業界も転換期を迎え、十年前に世の中を変えたばかりの業界ですら例外ではありません。クラウドファンディングが零細企業ローンに、オンラインリテールが小売市場全体に、そしてサブスクリプションメディアサービスが従来のケーブルテレビに与えたインパクトを考えてみてください。周りを見ればどこでも変革が急速に進んでいます。自動プロセス、メッセージャーボット、人工知能は、予想のつかないわくわくさせる未来へと導いている新たなテクノロジーのほんの一握りです。



BI で好んで使われる包括的な用語の役割の変化に関する詳細は、今日の分析を取り巻くキーとなる用語をハイライトしたホワイトペーパー、「[分析を定義する](#)」をご覧ください。

しかし、これらのトレンドのすべてに共通して絡まっているものが1つあります。それは膨大な量のデータです。

データを原料に例えるなら、分析はそれを有用な情報に精製するプロセスで、最終的にビジネスに競争利益をもたらします。今日、データは今までになく重要になっています。ビジネスが変わりゆく業界に後れをとらず進化していく上で、迅速かつ賢い意思決定を行うためには、増え続けている正確で最新のデータに頼らざるを得ません。しかし、従来の分析テクノロジーはデータの成長と形式の変化への対応が遅く、扱いにくいものです。まるで毎週のように新たなタイプのデータベースが登場し、新しいデバイスでデータが生成され、それらすべてが全く新しいテクノロジーで生まれているような現状です。分析はこの変化に全く追いついていないのです。



エンタープライズ分析へのモダンなアプローチを可能にする上で、IT とビジネスの協力体制が欠かせない理由の詳細に関しては、ホワイトペーパー「[分析の文化を築く方法](#)」をご覧ください。

今日のデジタル時代のデータドリブンな社風を作り、今日のビジネスチャレンジに正確性とスピードで立ち向かえるようにするためには、新たなテクノロジーだけでなく、このチャレンジの中をリードしていく人材を含め、情報を提供する新しい方法への投資も必要です。この並行する文化の変化は、IT とビジネスの関係を根本的に変え、これら2つの部門は協力してデータを収集し掘り出すためのパートナーであり、データを精製して適切な情報をオンデマンドで提供するためのパートナーでもあります。IT とビジネスが協力したとき、企業はセルフサービス分析を夢から現実へと変えることができるのです。



IT 組織がセルフサービス分析への変革をリードする際、規模に応じたガバナンスとセキュリティを確保することができます。そして、ビジネス部門をデータドリブンに、またアジャイルにすれば、IT 部門はビジネス部門が信頼を置くパートナーとなります。

- COLIN REES 氏、DOMINO'S PIZZA GROUP LIMITED 社 CIO

このホワイトペーパーでは、ビジネスと IT が共に企業全体にデータ、価値、そして意思決定の手段をもたらすモダン分析プラットフォームの構成要素をご紹介します。これらには今日の新しいピカピカのツールと、何十年にもわたりビジネスインテリジェンスの基礎として使用されてきた従来のツールの両方があります。また、データを収集しレポートするツールであれ、インタラクティブなインサイトを共有するための新しいテクノロジーであれ、それぞれの構成要素がデータをインサイトへと変えるより大規模なプロセスにどのように組み合わされていくのかをご紹介します。また、Tableau がどのようにモダン分析プラットフォームの基盤となると同時に、新たなデータドリブンな分析文化を展開していくきっかけになるかをご説明します。

目次

モダン分析の特徴と必要とされている理由	2
1. 今日の 3 つのデータチャレンジ	4
データはどこにでもある	4
誰もがデータを必要としている	5
データは常に変化している	5
2. モダン分析プラットフォームの構成要素	6
IT による提供	7
作成 利用	9
3. すべてを 1 つに	12
4. 付録	13
ストリームの取得	13
統合ハブオーケストレーション	13
非構造化データ、NoSQL、データレイクの詳細	13
サービスとしてのデータ	15
論理データウェアハウス	15
機械学習	17
自然言語	17
推奨データディスカバリ	17
検索	17
アラート	18
ストーリーテリング	18
Tableau について	19

1.今日の3つのデータチャレンジ

データはどこにでもあり、秒単位で大量に生成されています。たとえば、最近では歯ブラシもスマートデバイスになり、歯磨きをした時刻や磨いた時間、内部部品のステータスを記録できます。これらの情報をすべて歯科医に送ることもできます。シンプルなタスクが今では何千ものデータポイントになりました。この一例を何百万台ものデバイスの単位で考えてみてください。するとたった1つの小さな業界が数年前には存在していなかった山のようなデータを生み出しているのがわかります。それにイベントログ、API、ソーシャルメディア、Web サイトトラッキング、その他多数のインターネットテクノロジーを足せば、爆発的なデータの増加があちこちで起きています。

このモダンエコシステムは、3つのビジネスチャレンジをもたらしています：

1. データはどこにでもある
2. 誰もがデータを必要としている
3. データは常に変化している

データはどこにでもある



図1 Tableau はあらゆる場所のデータに接続可能

以前はほとんどの企業がデータをオンプレミスで保管していました。一見しっかりと定義されたデータウェアハウスで作成され保管されているすべてのデータのコントロールに、多大な労力を費やしていました。収集されていないデータがあるとしたら、それは重要なデータではなくおそらく労力を費やす価値のないものでした。



クラウドへの移行に関する最新トレンドは、[クラウドデータブリーフィング](#)をご覧ください。

Google Analytics、Splunk、ServiceNow、Salesforce など、Web サイトやモバイルデバイス、クラウドアプリケーションがすべてオフプレミスでデータを生成している今の時代、このような考え方はビジネスが終わりを迎えてしまう可能性があります。このトレンドはますます急速に進むばかりで、サードパーティプロバイダーによりさらに多くの有用な情報がクラウド上に生成されています。このクラウドへの動きに対応するため、オンプレミスのインフラストラクチャをクラウドに移行している企業もあります。

誰もがデータを必要としている

ここ 20 年間、市場のデジタルビジネスとクラウドへの移行が見られました。これはモダン分析革命の 1 つのカギです。もう 1 つのカギは、セルフサービス BI によるデータドリブンな社風への移行です。分析の文化は今日の最も革新的な企業の隅々に浸透しており、そのような企業では、ビジネスユーザーは質問があれば自分自身で答えを見つけることができます。これにより企業は急速に、デジタルゴールドとも言えるデータを情報へと精製するマスターになっています。企業が分析の文化を完全に築くには、企業の最高の資産とも言える人々とデータをまとめ、誰もが適切なデータにアクセスできるようにし、探索と協力を奨励する必要があります。

分析に対するモダンなアプローチでは、IT 部門とビジネス部門が連携して事業を進めます。IT はビジネスユーザーが信頼できるデータとコンテンツを見つけられるように一元管理した環境を提供し、誰もがそれを安全に利用し、質問したり、実験したり、考えるスピードで決断を下せるようにします。これは各分野の専門家がスムーズなアジャイル性と急速かつ継続的な改善を可能にするメタデータやビジネスルール、レポートモデルを作成する、ボトムアップ式の方法です。



Tableau を使い始めた時点で私たちが考えていたのは、ダッシュボードとレポートの作成だけでした。まさか Tableau が組織の DNA を根本から変えるだろうなどは、誰も予想していませんでした。ソリューションやテクノロジーの面だけではありません。データに対する文化がどのように変わったかということも重要です。

- ASHISH BRAGANZA 氏、グローバル BI 担当ディレクター

データは常に変化している

誰もが承知のように、常に変わらないのは「変化」し続けているということだけです。モダン分析プラットフォームは柔軟性、つまりデータをプラットフォーム間でやり取りでき、オンデマンドでインフラストラクチャを調整でき、新しいデータタイプを活用でき、新たな使い道を可能にすることを優先しています。また、機械学習や音声アシスタント、自然言語クエリなど、毎日のように世界中でデータ分析の新たなテクノロジーが生まれているようです。これらのテクノロジーの中には、少なくとも今のところは実用的というよりはフィクションのようなものもありますが、新たな方法やテクニックは後に熟成し、顧客や企業にとって価値のあるものとなります。

データが急速に進化する世の中、拡がり続けるインフラストラクチャのニーズと新たなテクノロジーの両方に対応するためには、柔軟性が最も重要となります。柔軟性は紛れもない競争利益を得て、それを維持するために欠かせません。未来に備えた分析アーキテクチャを検討する際には、将来のアジャイル性が大幅に制限されないよう、ベンダーの専用アーキテクチャに固定されてしまうことを避けるべきです。

2. モダン分析プラットフォームの構成要素

今日のビジネスが直面しているこれら3つのチャレンジは、思っているほど打ち勝ちがたいものではありません。データが今日の進化し続けるビジネス世界に共通する糸なら、モダン分析プラットフォームはその可能性を解き放つためのカギです。しかし、インフラストラクチャを構成するのはモダン分析プラットフォームだけではありません。インフラストラクチャはいくつもの独立した構成要素で成り立っています。そのうちのいくつかは従来のビジネスインテリジェンスの主要要素がモダン化したもの(データウェアハウスなど)です。他の要素は、そもそもビジネスによるデータ分析へのアプローチに革命をもたらした全く新しいコンセプト(ビジュアル分析など)です。これらの両方の要素が合わさり、企業が今日のビジネスチャレンジに立ち向かうことのできる分析プラットフォームとなるのです。

モダン分析プラットフォームは、次の2つのサイドに分けられます:

- **ITによる提供**、これにはデータの取得、キュレーション、準備が含まれます。
- **作成・利用**、これにはデータの分析と適切な関係者へのインサイトの伝達が含まれます。

従来、これら2つのサイドがITによるITだけの1つのプロセスにまとめられていました。現在、片側にあるデータソースの作成と処理はITが担当しています。もう片方の分析とデリバリーは、権限はITにより与えられるものの、ビジネスユーザー自身で実行されています。

この2つの組み合わせが、ビジネスとITの真のパートナーシップであり、データドリブンな意志決定を目指す企業の新しい運営形態です。これは時にバイモーダルBIとも呼ばれ、従来のBIと運用報告の良さをそのままに、モダン分析のセルフサービスメカニズムが採り入れられています。

このITとビジネスの関係では、ITがデータアーキテクチャをデザインし、適切なデータセキュリティとアクセスコントロールを管理します。そして、ビジネスの専門家が必要な分析資産を必要ときに作成します。その結果、ITにより誰もが重要な質問の答えを効率よく自分で見つけることができるようになり、ビジネスユーザーは質問が出た瞬間に答えを見つけることができ、モダンなビジネス環境のチャレンジに立ち向かうことのできる機敏な組織が生まれます。

このホワイトペーパーでは、モダン分析プラットフォームの2つのサイドのそれぞれを構成する様々な要素、各サイドのトピトレンド、そして念頭に置くべき重要なコンセプトを詳細にご紹介します。具体的なコンポーネントの詳細(各コンポーネントを提供する市場をリードしているベンダーなど)や、それらがお客様の環境で利用可能か確認するには、[付録](#)の各構成要素のセクションをご覧ください。



図2 基本的な構成要素

ITによる提供

従来のIT中心のビジネスインテリジェンスと異なり、今日の最も有能なIT組織は、ユーザーと専門家が作成利用できるようデータを組織化し、整理し統一することで、分析データソースを有効化することに焦点を当てています。この役割は明白ですが、どれだけ誇張してもし過ぎることはありません。データの収集、データソースの管理、他のユーザーが利用できるように処理することは、従来からビジネスインテリジェンスにとって重要でしたが、モダン分析プラットフォームでもその中心的地位は変わりません。精製する生データがなければ、インサイトを発見することはできません。

モダン分析プラットフォームの違いは、ビジネスとITのパートナーシップにあります。ビジネスユーザーに自分でデータを分析するためのツールを提供すれば、データ自体が信頼できるものと認識している上で即座に自由に質問に答えることができます。これにより、正確で機敏なレポートとダッシュボードが生まれます。そしてITはダッシュボードや変更のリクエストから解放され、やっとデータそのものを優先することができ、データガバナンスおよびセキュリティの保護、データの正確性の確保、そしてデータを収集、処理、保管するための最も効率的なパイプラインを確立することができるようになります。

データを優先するのにこれ以上ピッタリなタイミングはありません。ビジネスはその規模に関わらず、すでにデータを収集しており、分析を行っているもののそれはごく一部のデータで、残りのデータはおそらく**ダークデータ**と化しているでしょう。データの収集地点は十億とあり、できるだけ多くのデータを収集できるよう、より多くのツールが市場に登場しています。今日では、ハイボリューム、データの種類、異なるデータソースタイプなど、様々な特性のデータを扱うためのテクノロジーが数多く存在します。同じ組織は2つとなく、お客様の環境にとってどのコンポーネントが現在、そして将来的に最もふさわしいかじっくり見極める必要があります。

考慮すべき点には次のようなものがあります:

データストリーミングやDaaS (Data-as-a-Service) などの特定のテクノロジーに関する詳細 (具体的なベンダーの選択肢を含む) は、下記の付録をご覧ください。

データベースとデータウェアハウス

データベースとデータウェアハウスは、何十年にもわたりビジネスインテリジェンスの基礎となってきました。これらには、モダン分析アーキテクチャで続けて利用されているものもあれば、現在では利用されなくなってきているものもあります。

最も古いデータベースの中に、OLAP (オンライン分析処理) として知られているものがあります。OLAP はもともとデータベーステクノロジーの動作の遅さに対するソリューションとして生まれ、集計とキャッシングにより予測可能なクエリへの応答時間を短縮しました。しかし、企業の質問がより複雑になり予測が難しくなるにつれ、頻繁に全く新しい集計を必要とする OLAP は追いつくことができなくなりました。また、さらに改善されたデータベーステクノロジーを前にしてその必要性は失われつつあります。

今日のデータベースは、インメモリや MPP テクノロジーなどの最新の計算技術を活用しています。これにより、データベースの非常に高速なパフォーマンスと線形的なスケーラビリティを実現しつつ、データストレージとハードウェアメモリ使用量を最適化し、場合によってはビルトインの計算機能やデータサイエンス機能を搭載させることもできます。

また、クラウドの登場によりデータベーステクノロジーに、オンプレミスでは不可能だった新たな命が吹き込まれました。この新たな命には、ハードウェアを揃えずに利用開始できたり、ビジネスの需要の変化に合わせて自由に変更できる高いスケーラビリティ、またインフラストラクチャを管理する専門のチームが必要ないなどのメリットがあります。

モダン分析アーキテクチャからデータベースとデータウェアハウスがなくなることはなく、これらはセルフサービスレポートの作成用にガバナンスの行き届いた正確で統一されたデータを企業全体に提供する上で、継続して重要な役割を果たしていきます。他のテクノロジー (Hadoop、データレイクなど) を採り入れている企業でも、通常リレーショナルデータベースをデータソースの一部として維持しています。

NoSQL、非構造化データ、データレイク

データベースとデータウェアハウスは、予測可能なソースや形式のデータの分析をサポートする際に特に力を発揮します。もちろん、データは常に予測可能とは限りません。モダン分析アーキテクチャでは、スキーマや形式がしっかりと定義されていないデータソースを含め、どのソースからでも素早くデータを読み込めるというメリットから、NoSQL データベースがすべての企業の戦力となっています。「non-SQL」や「not-only-SQL」とも呼ばれる NoSQL データベースは、列、ドキュメント、キーバリュー、グラフストレージタイプを含む、従来のリレーショナルデータベースと異なるタイプのデータストレージを提供します。

非構造化データに関連しているのがビッグデータとデータレイクというコンセプトです。データはあらゆる場所で、時にはランダムな場所で生成されており、すべて収集して利用可能な形式にするには非常に手間がかかります。データをまず一定の形式に合うように無理やり変換するのではなく、分析ツールを生データにそのまま接続できるようにするテクノロジーが開発されています。

このようなテクノロジーの1つが、構造・非構造に関わらず大量のデータをネイティブ形式で保管できる、データレイクと呼ばれるストレージリポジトリです。ユーザーはすべてのデータを特定の形式に前処理することなく、API や SQL のような言語などの最適なプロセスメカニズムでデータをその場で変換し、分析することができます。

これらのツールはすべて、データの生成量や場所が予測不能なモノのインターネット、データサイエンス、ストリーミングデータ、その他非構造化分析の使用事例に関連したプロジェクトでよく使用されています。

NoSQL、Hadoop、データレイクに関連したデータテクノロジーのリストは、[付録](#)をご覧ください。

フラットファイル

Excel と CSV ファイルの全盛期はまだ続きます。零細企業でも大企業でも、どこからともなく現れるこれらのフラットファイルはこれからも、そしておそらく永久になくはありません。実は、これらのファイルにはこれまでよりも多くの使い道があるのです。一昔前まで、これらのファイルはユーザーのコンピューターでのみ保管されていました。今では Google Drive や Dropbox のようなクラウドストレージに保管されています。サードパーティプロバイダーは DaaS の一部としてフラットファイルを作成します。これらのファイルは素早く作成できるため、様々なデータフィールドトや追加の顧客リサーチ、または既存のデータセットを拡大するちょっとした情報の凡例として便利です。

また、フラットファイルには適切なタイミングで適切なセキュリティ対策を行うようにしましょう。必要に応じた、特に 1 度きりのシナリオでの活用を奨励しましょう。特定のファイルが多く利用されるようになったら、適切なセキュリティプロトコルを適用し、安全を確保して適切なユーザーのみがアクセスできるようにしましょう。

作成利用

モダン BI の最大の特徴は、ビジネスユーザーが BI プラットフォームを使用できることです。意思決定者が IT にレポートをリクエストして数日待った後に受け取るレポートの情報はすでに古く、問いたかったことへの答えが得られないという日々はもう終わりです。今日では、質問のある意思決定者がツールを使用して、自分自身で答えを見つけることができます。IT により企業全体で利用可能なデータを信頼できる環境が整えられているため、ビジネスユーザーはプログラミングなしでスマートかつデータドリブンな意思決定を必要に応じて行うことができます。

ここでの重要な構成要素は実際使用する分析ツールで、初日から使い始めることができるものでなければなりません。このセクションにあるどの要素が組織に最も関連性があるかを考える前に、アーキテクチャ全体を構築する準備を進めると共にデータソースを検証し素早く閲覧するためにも、あらゆる場所にあるデータに接続し始めるべきです。

データ分析には様々なコンポーネントがありますが、Tableau ではその中心にあるのがビジュアル分析、つまり組織の誰もがプログラミングの経験の有無に関わらず、データソースに直接接続してインサイトを得られる手段であると考えています。これらのツールを教師や医師、営業担当者に提供することで、組織が変更リクエストの塊からスムーズなマシンへと変わります。

ビジネスインテリジェンスの世界は、他のユーザーとのインサイトの共有方法にさらに重点的に取り組むという新たな動きを見せています。ユーザーはもうダッシュボードとレポートに縛られることなく、完全なインタラクティブアプリケーション、データやテキスト、画像を組み込んだロング形式の記事、さらにはモバイルデバイスで最適化表示されるコンテンツまで作成できます。

また、ビジネスとその所属部署が成長するにつれ、情報を素早く共有したり、データソースを探したり、最新のダッシュボードを利用したり、最も重要なメトリックスをフォローしたりするために、スマートなプロダクティビティツールを求めようになります。

このセクションでは、最高のモダン分析ツールが提供するデータデリバリー用のコンポーネントのいくつかを検討します。ストーリーテリングやアラートなどのさらなる詳細は、[付録](#)をご覧ください。

ビジュアル分析

人間の視覚系は世界で最もパワフルなツールの1つです。今日、ついにそれがデータ分析に欠くことのできない一手段となりました。ビジュアル分析は、脳が既に日常的に行っているパターン認識に基づいて、同様に上下の傾向や不規則なアクティビティの増加、外れ値となる特定の記録など、データのパターンを示すことができます。

従来のスプレッドシートの場合、行と列でデータを分析し、共有するサブセットを選択して、チャートを作成する必要がありました。作成方法が使いづらいウィザードであれ、テキストベースのコマンドであれ、これらのチャートは質問に答えることもあれば、時には全く新しい質問を生み出していましたが、どちらにせよ常に先のない最終ステップでした。それに比べビジュアル分析は、シンプルなドラッグ&ドロップ操作でビジュアルがプロセスの一部になる、エレガントでわかりやすい分析を可能にし、その結果ただ単にチャートを作成するだけでなく、インサイトを求める過程そのものが、見つけた答えと同じくらい価値のあるものになります。

ビジュアル分析は単なるきれいなビジュアライゼーションツールではありません。プログラミングなしでデータを組み合わせ、異常に気づき、計算やグループ化、what-if分析、その他多くの機能によりデータを拡大するための言語なのです。

従来のBIとレポート作成

最近では作成方法こそ異なるものの、現在でも従来のBIやダッシュボード、レポート作成は活用されています。エグゼクティブダッシュボードや財務監査など、多くの静的レポートは、分析クエリが事前に出され、頻繁に参照元データモデルの変更が必要となり、作成に技術開発のスキルが必要でした。これらすべて、開発に何日間、何週間、何か月間と要しました。

モダン分析プラットフォームでは、これらのダッシュボードとレポートの多くがアドホックの質問として始まり、答えを見つける質問の性質のためITとデータスチュワードが磨きをかけ検証し、最終的には従来の静的レポートにとって代わります。質問は進化し、変化し、全く新しい質問へと導いていくため、この新しいプロセスでは、正しい答えを求めてデータを掘り探す際にビジネスユーザーの専門知識を活用します。モダン分析はその柔軟性のため、従来のレポートの要件は未だ変わらないにも関わらず、従来のツールの代わりに使用されるようになっていきます。

パーソナルデータ準備ツール

ETLと混同しないよう、データ準備ツールとは、非ITユーザーがデータをパワフルかつ正確に操作できるようデザインされた、軽量のアプリケーションです。これらのツールは使いやすさ、スピード、ビジュアル分析ツールとしてのアジャイル性の原則に基づき開発され、一般的なビジネスユーザーでもデータセットの組み合わせ、結合の自動化、フィールド名の変更、その他の分析準備を整えるためのデータの改善が行えます。

データは使い始めるまでどのような変更が必要かわかりにくいことが多いため、これらのツールは優先順位が高いものではありません。しかし、組織の運営全期間を通して、質問の答えを求めるために費やされる時間のほとんどは、データを適切な形式に変換する作業に費やされます。パーソナルデータ準備ツールは、ITによる事前定義されたレポート作成のセマンティックレイヤーを必ずしも必要としたり無効にすることなくその時間を節約するためのパワフルな手段です。



Harvard Business Review の「**あなたのデータ戦略は?**」(英語)によると、「80%のアナリストの時間が単にデータディスカバリとデータ準備に費やされています。」

Tableau

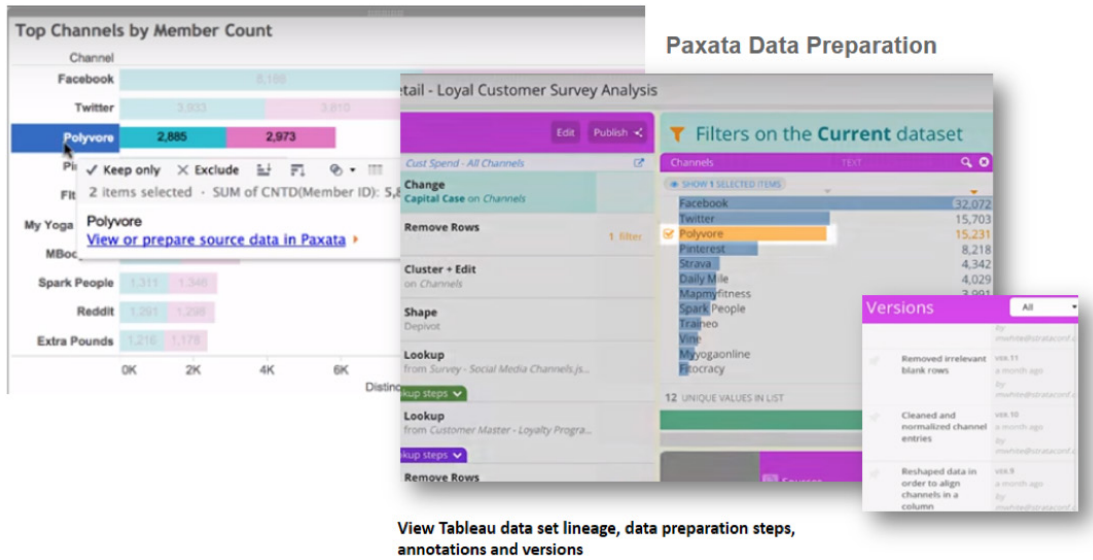


図 3 Paxata の Personal Data Preparation へのビジュアル分析の統合

高度な分析

高度な分析は、企業が統計や予測アルゴリズム、機械学習で膨大なデータセットを最大限に活用しようとする中、モダン分析アーキテクチャに欠かせないコンポーネントとして登場しました。

以前は、高度な分析は、トレーニングを受けた R、Python、SPSS、SAS などのプログラムを使うデータサイエンティストしか利用できませんでした。今日では、ビジュアル分析の発達により高度な分析がビルトイン機能として、すべてのセルフサービス分析ユーザーに利用できるようになりました。箱ヒゲ図やツリーマップ、基本的な予測モデリングテクノロジーなどの機能が一般的に普及し、クリック 1 つで利用できるようになりました。

専用の統計分析ツールは今でも使用されています。顧客に「次に注目すべきもの」などの情報を提供するために常に進化し続けるアルゴリズムが必要な場合、または顧客のクレジットカードに詐欺行為が検知されたタイミングを把握する機能を開発したい場合は、コアのビジュアル分析ツールに加えて特殊な専用ツールを検討すべきかもしれません。これらのツールを導入するにはトレーニングとツールごとのプログラミングの経験が必要で、これには何か月もかかります。

共有とコラボレーション

モダン BI プラットフォームで重要な機能は、インサイトの共有とコラボレーション、そしてソーシャル化です。コンテキストの把握から次のベストアクションの優先化まで、インサイトはコラボレーションにより最大のインパクトを發揮できます。モダン分析プラットフォームはディスカッションフォーラムや注釈、コメント、お気に入り、いいね、その他代表的なプロダクティビティアプリやポータルアプリから採り入れられたソーシャルコンセプトを提供します。分析プラットフォームから得たインサイトをツールから直接伝えられることで、分析の流れがずっとシンプルになり、より多くの探索や貴重な発見のディスカッションが行われるようになります。また、コラボレーションは埋め込みにより、外部のアプリケーションやポータルに広げることができます。

埋め込み分析

最もインパクトが大きいにもかかわらず頻繁に見過ごされているビジネスコンセプトの 1 つが、業務フローです。データで答えを探すためにビジネスユーザーが通常の業務プロセスから一旦離れるのではなく、インサイトをいつもの業務フローにシームレスに加え、既存のプロセスに組み込みましょう。

モダン分析では、データとダッシュボードは企業のポータルや他のアプリケーションに直接埋め込まれたり、プロダクティビティツールに統合されたりしているでしょう。最高の分析プラットフォームは、完成度の高いAPIやソフトウェア開発者キット、柔軟性の高いデリバリーメカニズムでこれらのシナリオをすべてサポートし、1つのツールから別のツールへと簡単に切り替えることができ、すべてのツールを1つのポータルに組み合わせることもできます。

フローは物理的な場所にも当てはまります。最近では社外に出ることも多く、ファイアウォールの内側の社内リソースにアクセスできないこともよくあります。モダン分析では、どこからでもどのようなデバイスからでもデータにアクセスできます。これにより、営業担当者はノートパソコンを取り出さなくてもモバイルデバイスから、情報に基づいたデータドリブンな意思決定を行うことができます。また、現場にいる建設管理者がVPNで社内ネットワークに接続しなくても、重要な情報にアクセスすることもできます。モバイルデバイス、そしてクラウドにより、ビジネスのやり方はすっかり変わっており、真のモダン分析プラットフォームもビジネスがこれらのメリットを活用できるものでなければなりません。

3. すべてを1つに

これらの構成要素は、ビジネスにどのようなチャレンジにでも正面から立ち向かえる力を与えるモダン分析プラットフォームの基礎となります。ビジネスとITの真のパートナーシップと組み合わせれば、このプラットフォームは従業員全員に、どのような意思決定が必要となってもそれに必要なツールが揃っていて、信頼できるデータに基づいた意思決定ができるという自信を与えてくれます。

モダン分析プラットフォーム全体を構築すること自体も、それだけでチャレンジのように思えます。しかし、導入前にエコシステム全体を一気に構築する必要はないのでご安心ください。むしろ、今日のビジネスは完全なエコシステムを一気に構築するのではなく、小さなシステムから始め、次に投資すべきものに関する情報を伝えるエコシステムを目指して少しずつ変えていくほうがうまくいっています。テクノロジーソリューションを多くの部門に展開する前に、パイロットテストから始めることができます。導入開始にあたり、計画されているすべてのコンポーネントを統合する必要はありません。たとえば、ウェアハウス全体を完成させる前に、ビジュアル分析ツールでデータパイプラインに問題がないかを確認することができます。これにより、分析から直ちに価値を引き出し、データのギャップやエラーを把握し、最終的により正確で機能性の高いデータウェアハウスを構築することができます。

重要なのはこのような少しずつの拡張を可能にするツールを使うことで、モダン分析プラットフォームとは正に、一つひとつを組み合わせながら組織のアクセス能力、アジャイル性、そして様々なデータソースからインサイトを得る能力を高めていくことのできる、構成要素の集まりです。世界クラスの分析のリーダーは正にこれを行い、ビジュアル分析を他の最高のビッグデータ分析やモノのインターネット、データサイエンスソリューションと組み合わせて戦略的に活用しています。

たとえば、**Netflix** は**総合ビッグデータプラットフォーム**とデータレイクを構築し、業務上発生する膨大な量のデータをサポートしています。TableauはS3、EMR、Sparkなどの様々なツールを組み合わせ、ビジネスをサポートする分析プラットフォームにまとめるために欠かせないコンポーネントです。

デジタルビジネス変革のどの段階にあらうとも、直ちに今あるデータを使い始めることが重要です。次の大きな市場変化を前にして、ビジネスは今以上に素早く対応しなければなりません。モダン分析のコンポーネントを導入し、ビジネスにデータドリブンな意思決定ができる力を与え、変化をもたらす側になりましょう。

4. 付録

世の中には様々な分析テクノロジーやソリューションがあり、それぞれに具体的な目的とメリットがあります。この付録では、概要で紹介されなかったオプションをご紹介します。

ストリームの取得

ストリームデータはソーシャルネットワーク、スマートメーター、ホームオートメーション、ビデオゲーム、IoT センサーなど、あちこちで接続されているデバイスやアプリから生成され続けています。多くの場合、これらのデータは半構造データのパイプラインを通じて収集されます。ストリームにはリアルタイムの分析と予測アルゴリズムが使用できますが、ストリームデータは通常ラムダアーキテクチャを使用して RAW 形式でルーティング・保管され、Hadoop のようなデータレイクで分析に使用されています。

ラムダアーキテクチャは、バッチとストリームプロセスの両方を利用して大量のデータを扱えるようデザインされたデータプロセスアーキテクチャです。このデザインによりレイテンシー、スループット、耐障害性の課題は解消されます。

現在データストリーミングには Amazon Kinesis、Storm、Flume、Kafka、Informatica Vibe Data Stream を含む様々なオプションがあります。

統合ハブオーケストレーション

ハブアンドスポークの統合パターンはわかりやすく、広く利用されているデータアーキテクチャデザインです。ハブはあらゆる場所にあるデータソースを分離し、管理するポイントツーポイントインターフェイスの数を減らすことで、より柔軟性の高い統合を可能にすることを目標としています。統合ハブのパブリッシュおよびサブスクライブ機能により、データの再利用が促進され、最適化やデータの標準化、またガバナンスのための一元管理が実現します。一元管理をすることで、あらゆる場所のデータソースをつなぐ統制の取れたデータ移動パイプライン全体の可視性が高まります。

新世代のデータ統合ハブは、従来の機能をセルフサービス分析ユーザーにも利用可能にします。誰もが IT の手をあまり借りないで、モダンな統合ハブのデータフィードにパブリッシュしたりサブスクライブすることができます。データ利用者は認証済みデータを利用し、データの系統と統合プロセスを可視化することができます。モダンデータ統合ハブには他にもシームレスなデータクオリティ機能、データソースオンボーディングの高速化、小規模または大規模なデータセットのタイムリーなデリバリーなどのメリットがあります。

Informatica と Cisco はデータ統合ハブテクノロジーの市場リーダーです。Tableau の Informatica との密接な統合により、何百種類もの異なるデータソースを Tableau データ抽出に組み合わせ、Tableau Data Server に保管し最新の状態に保ち、組織内の誰もが利用できるようにすることができます。

非構造化データ、NoSQL、データレイクの詳細

データレイクはより高速かつ柔軟性の高いデータの取得とストレージで、誰もが様々な方法で素早く生データを分析できるように、モダンビッグデータ分析の要件をサポートします。データレイクはデータウェアハウスにとって代わるものではありません。

最新のインジェストアンドロードのデザインパターンでは、あらゆるサイズや形の生データがデータレイクに送られます。データレイクは、構造化データ、半構造化データ、非構造化データに関わらず、ネイティブ形式で大量のデータを保管するストレージリポジトリです。また、データレイクは API や SQL のような言語により生データを「スキーマオンリード」機能で変換できるように最適化されたプロセスメカニズムも提供します。

初期の HDFS (Hadoop Distributed File System) のときから、データレイクにはその障害許容力の高さや低コストで Hadoop が使用されてきましたが、データレイクの導入オプションは他にもあります。スキーマの柔軟性の高い Amazon Web Services の Simple Storage Service (S3) や NoSQL データベースのようなオブジェクトストアも、データレイクに使用できます。Tableau は Amazon S3 への接続用に [Amazon の Athena](#) データサービスをサポートするようになり、NoSQL データベースに直接接続できる様々なツールが追加されました。

モダン分析アーキテクチャでは、あらゆる場所からのデータの読み込みが高速なことから、スキーマのないデータベースコンセプトから、NoSQL データベースが一般的になっています。Non-SQL や not-only-SQL とも呼ばれている NoSQL データベースは、従来に代わる形のデータストレージを提供します。一般的な NoSQL ストレージタイプには列、ドキュメント、キーバリュー、グラフなどがあります。

Tableau と組み合わせてよく使用される NoSQL データベースの例には MongoDB、Datastax、MarkLogic などがありますが、これらに限られたものではありません。

Hadoop はよくビッグデータプラットフォームに使用されますが、データベースではありません。Hadoop は、市販ハードウェアのクラスターでのデータストレージとアプリケーション実行に使用される、オープンソースのソフトウェアフレームワークです。あらゆる種類のデータを保管できる大容量のストレージと膨大なプロセッシングパワー、非常に大量のタスクやジョブを並行処理する能力を提供します。

モダン分析アーキテクチャでは、Hadoop がコストの低いストレージとデータアーカイブを提供しており、データウェアハウスから過去の古いデータをオフロードしてオンラインのコールドストレージに移すことができます。また、Hadoop は IoT やデータサイエンス、非構造分析にも利用できます。

Hadoop のフレームワーク内のデータの読み込み、整理、クエリ用の関連テクノロジーには次のものがありますが、これらに限られたものではありません:

- **Apache Spark** – 高性能のインメモリ分析が特徴の、オープンソースクラスターのコンピューティングフレームワークで、関連するプロジェクトの数は増え続けています。
- **Apache Impala** – Apache Hadoop 用のオープンソースの分析 MPP データベースです。これは、良好な結果をもたらしている Tableau の Hadoop を活用したプロジェクトで最も一般的に使用されているデータ接続です。
- **Apache Presto** – あらゆるサイズのデータセット間でインタラクティブクエリを実行するためのオープンソース分散 SQL クエリエンジンです。 [Tableau, はバージョン 10 で Presto をサポート](#)しています。
- **MapReduce** – 入力されたデータを小さな問題に区分指定し、ワーカーノードに分散する、パラレルプロセッシングソフトウェアフレームワークです。
- **Hive** – データウェアハウスで、SQL のようなクエリ言語を用います。また、Hive 2.0 には LLAP (ライブロングとプロセス) が追加され、Hive のクエリパフォーマンスが大幅に改善されました。
- **HDFS (Hadoop Distributed File System)** – 事前に整理する必要なく複数マシンにデータを保管する、スケーラブルなシステムです。
- **YARN** – (Yet Another Resource Negotiator) は、Hadoop で実行中のプロセスのリソース管理を行います。
- **Ambari** – Hadoop のサービスとコンポーネントを管理する Web インターフェイスです。
- **Cassandra** – 分散型データベースシステムです。
- **Flume** – HDFS へのデータストリーミング用のソフトウェアです。
- **Hbase** – Hadoop 上で動作する非リレーショナルの分散型データベースです。

- **HCatalog** – 表とストレージの管理レイヤーです。
- **Oozie** – Hadoop ジョブスケジューラーです。
- **Pig** – HDFS に保管されたデータを操作するためのプラットフォームです。
- **Solr** – スケーラブルな検索ツールです。
- **Sqoop** – Hadoop と リレーショナルデータベースの間でデータを移動させます。
- **Zookeeper** – 分散処理をコーディネートするアプリケーションです。

ここ 2 年間で、Apache Spark は Hadoop エコシステムのコンポーネントから、いくつもの企業に選ばれているスタンドアロンのビッグデータ分析プラットフォームになりました。Spark は Hadoop と比較してはるかに高速なデータ処理を実現します。Spark 自体にコア Apache Spark ランタイムや Spark SQL、Spark Streaming、MLlib、ML、GraphX を含む、多くの関連プロジェクトがあります。現在では 250 を超える組織の 1,000 人以上が貢献している最大のビッグデータオープンソースプロジェクトとなりました。

Tableau はビッグデータに特化した分析接続とビジュアルデータ分析の市場リーダーです。クラス最高のビッグデータ分析プログラムは、Tableau を Cloudera、Spark SQL、Amazon EMR、Hortonworks、Microsoft HDInsight/Data Lake、MapR と組み合わせ使用しています。他の多くのビッグデータテクノロジーは、これらの明らかにサポートされているテクノロジーまたはそれぞれのドライバーで Tableau と組み合わせることができます。

サービスとしてのデータ

データがゴールドのように扱われているデジタルの世界では、データは誰もが利用する商品でもあります。顧客、金融、市場、気象、地理、そして人口統計のデータは、既にデータ市場や取引プラットフォームで購入可能なサービスとして提供されています。

DaaS (Data-as-a-Service) には、クラウド経由のデータデリバリー用に柔軟性の高いサービス指向アーキテクチャ (SOA) のパターンが適用されます。SOA アーキテクチャは極度に単純化されており、このアプローチは非常に高いアジャイル性を実現します。現在、統合や外部レポート作成シナリオに、ISV、CRM、ERP が標準の DaaS REST API を提供しているケースが見られます。

Tableau の [Web データコネクタ SDK](#) は、既存のコネクタ外にあるデータへの接続を可能にします。セルフサービス分析ユーザーは、社内 Web サービス、JSON データ、REST API を含む、HTTP でアクセス可能なほぼすべてのデータに接続できます。

論理データウェアハウス

業界をリードする分析組織は、Cisco や Denodo のようなベンダーのデータ仮想化テクノロジーを用いて、あらゆる場所で柔軟性が高く論理的で、統一されたディメンションのデータのビューを提供しています。分析ユーザーにとって、論理データウェアハウスの見た目や使い勝手はリレーショナルデータウェアハウスと変わりません。Tableau ユーザーは、既製の ODBC ドライバーでこれらのツールに接続できます。

データの仮想化の主な機能の 1 つが、遠隔に分散された異種のクエリの、様々なデータソースと REST API への最適化です。また、論理データウェアハウスは、データソースの変更からレポート作成アプリケーションをバッファリングするセマンティックレイヤーにもなります。論理データウェアハウスはエンタープライズデータカタログとしてもよく利用されています。

マスターデータ管理

分析の質は使用されているデータで決まり、正しい意思決定には正確なデータが必要です。より多くの分野の専門家がボトムアップ式にデータを作成するようになり、レポートを作成するデータソースが最新ですっきり整理されていて一貫性があり正確であることを確実にするために再び極めて重要になっている、従来のデータクオリティとマスターデータ管理に新たな関心が集まっています。

よく利用されているマスターデータ管理ソリューションには Informatica、IBM、Stibo などがありますが、これらに限られたものではありません。最も頻繁に Tableau と合わせて使用されているデータクオリティソリューションには Trillium、Informatica Data Quality、Talend Data Quality 6.0、Tamr Eisenhower が挙げられます。

エンタープライズデータカタログ

最近登場しているもう1つのテクノロジーが、エンタープライズデータカタログです。エンタープライズデータカタログにより、セルフサービスでレポートを作成するユーザーが、認証済みのデータソースから意思決定に適切なデータを簡単に見つけられるようになります。エンタープライズデータカタログはビジュアル分析ソリューションに搭載されていて、Tableau とのシームレスな統合のためにデザインされたスタンドアロン型のものもあります。

エンタープライズデータカタログは取得されたデータソースをスキャンすることにより、表やビュー、ストアドプロシージャからメタデータが追加されています。新しいデータソースの自動検出やインテリジェントデータ分類、クロスデータソースエンティティマッピングにより、データカタログは基本的にデータソースと一般的なデータ定義のエンタープライズビジネスグロッサリの役割を果たします。この分野の専門家は注釈やバージョン、ドキュメントを追加して、カタログデータソースのコンテキストをより充実させることができます。

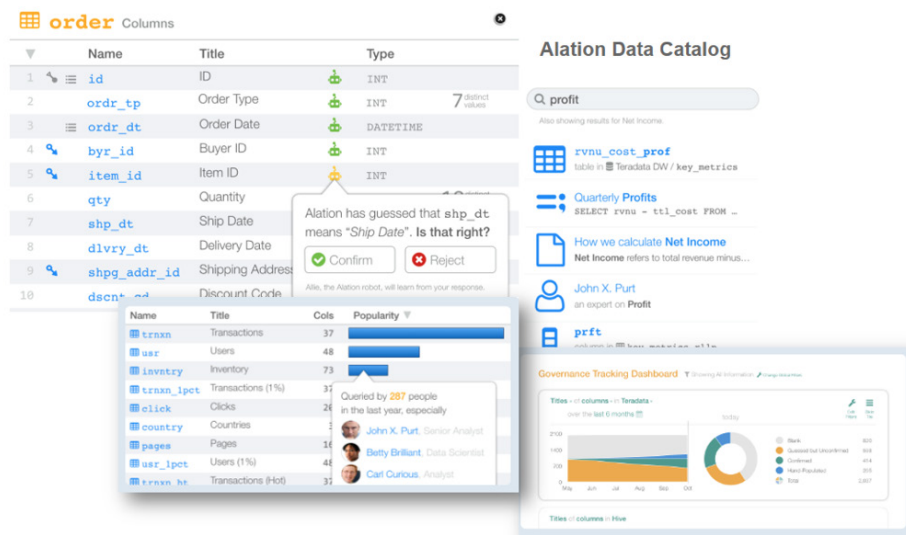


図4 Alationのエンタープライズデータカタログ

データカタログソリューションはデータキュレーションや既存のデータの効率的な再利用をサポートします。また、非常に優良なデータリネージとさらにレベルの高いデータガバナンス、セキュリティ、ログ、監査も可能にします。

Tableau とスムーズに統合する豊富なデータカタログのあるベンダーには、Alation、Collibra、Attivio、Informatica、Waterlineなどが挙げられます。

機械学習

さらに高度な分析では、認知、ディープラーニング、そして人工知能により既存のデータやパターンから推測し、既存のナレッジベースに基づき結論を出し、その結論をナレッジベースに戻すという自己学習のサイクルが永久に続きます。

このタイプの分析結果は、通常レポート内に埋め込まれた API または統合されたアプリケーションで見ます。現在 Tableau は、CognitiveCode、Digital Reasoning、その他ベンダーからの分析結果の可視化に使用されています。

Tableau 10.3 では、データへの接続と準備の時間を節約できるよう、**表の推奨とスマート結合**が追加されました。機械学習により、データベースがより頻繁に使用されるに伴い、経時的に推奨情報が改善されていきます。

自然言語

さらに高度なデータのストーリーテリングでは、自然言語と音声クエリにより、誰もが新しくより柔軟性の高い方法でデータディスカバリを行えるようになり始めています。自然言語は、自動作成された主な結果の文脈的説明を取得したり、予測値をリクエストしたり、テキストドキュメントのボリュームを分析したりすることで、どのプラットフォームからでもより簡単に分析にアクセスできるようにします。

現在、Tableau のビジュアル分析は、Yseop、Narrative Science、Automated Insight などの主要な自然言語生成 (NLG) ソリューションと組み合わせることができます。

これらのテクノロジーは Tableau のビジュアライゼーションのコンテキストを大まかに解釈するため、最も多くの場合には自然言語ツール自体に統合されるか、JavaScript による拡張機能として統合されています。さらに、**ClearGraph との合併**により、Tableau で直接さらにスマートなデータディスカバリと分析が可能になり、自然言語によるデータの操作がより簡単になります。

推奨データディスカバリ

スマートデータディスカバリは機械学習アルゴリズムを強化して、目的に合わせてフォーマットされたデータでさらに深い分析（「何が起るか」）と処方（「どのように最適化するか」）機能を実現します。ビジュアル分析が進化するにつれ、新しい自動インサイトが生まれ、推奨データが追加されます。これらの機能は、**2016 年 Tableau Conference** のロードマップキーノートで紹介されました。R や Python、API、分析データベース機能を統合することで、高度な分析機能の結果を Tableau でビジュアル化し、探索することができます。

検索

モダン分析アーキテクチャにより、ユーザーと専門家がデータの場所に関わらず、Google のような使いやすさでデータを検索し見つけることができるようになります。データをモデリングする代わりに、分析検索エンジンインデックステクノロジーがフィールド名、データタイプ、機械学習インテリジェンスに基づき、自動的に異なるデータソースを関連付けます。そして徐々に、過去のクエリと記録された使用情報に基づき動的検索提案が生成されます。最近の Siri や Alexa のような音声認識テクノロジーの登場により、分析検索と音声クエリ機能の組み合わせが見られるようになってきました。その良い例が、Alexa の音声コントロールを Tableau に採り入れて 2016 年 Tableau Conference の**ハッカソンで優勝**した Automated Insight です。

アラート

モダン分析アーキテクチャには、常にデジタルデータの海で重要なシグナルがないか監視する、構成可能でインテリジェントなデータドリブンアラート機能が含まれます。人間が重要な値をすべて1つずつ1日24時間、年中無休で確認することはできません。そこで、モダン分析ツールには自動化とアラート機能が非常に有用になるのです。

ツールによっては定期的にスナップショットを提供するものもあれば、ログを追跡して数値が一定のしきい値を超えたかどうか確認するものもあります。これらにはそれぞれ異なる使い道があります。ダッシュボードによっては役立つ情報が提供され、ただ毎日確認したくなるものもあります。一方、重要なアクションの基になるものでも、アクションナブルなインサイトなしにダッシュボードを毎日確認することは、効率的な時間の使い方ではありません。

Tableau では、[Tableau Server のデータドリブンアラート](#)で常にビジネスの最新情報を得ることができます。しきい値を選択するだけで自分やチーム全体がアラートをメールで受信できます。す。

ストーリーテリング

インサイト、つまり「データ」だけでは十分ではないときもあります。人はデータの「理由」も知りたいものです。なぜ営業成績が上がったのか? Web トラフィックが急増したのはなぜか? 医療用品の在庫切れを防ぐのに苦心しているのはなぜか?

ビジネスは長年にわたり、分析に文字や写真、さらにはビデオなどの他のコミュニケーション方法を組み合わせて答えを見つけようとしてきました。アナリストは Powerpoint でプレゼンテーションを作ったり、長い PDF レポートを作ったり、さらに手間のかかるケースでは何ページもの文書を印刷してファイルに閉じたりします。

今日のモダン分析ツールは、これらのストーリーテリングコンセプトの良いところを取り、ファーストクラスの機能として採り入れています。インタラクティブなダッシュボードを作成したり、新しいデータが追加された際にバックグラウンドで自動的に更新されるデータの特定のスナップショットを送ったり、さらにはインタラクティブなチャートにテキストや画像を組み合わせたレポートを作成したりすることもできます。ストーリーテリングにより、ただ数字を提供するのではなくデータの分析結果を説明することができます。

Tableau は選択肢とオープンスタンダードの重要性を根本的に重視しています。分析をより素早く簡単にするための研究開発に大きく投資し、これにはパートナーエコシステムとの革新的技術の開発も含まれています。これにより、分析の世界が進化し新たなテクノロジーが登場したときに、分析のリーダーが常に Tableau を現在、そして未来のテクノロジーのどれにでも統合できるようにすることができます。

Tableau について

データの大きさ、チャンネル、保存されているデータベースに関わらず、Tableau はお客様がデータを見て理解できるように支援します。PC から iPad まで、シームレスに素早くデータの接続、ブレンド、視覚化ができます。自動データ更新機能付きのマーケティングダッシュボードの作成およびパブリッシュ、同僚、チーム、エグゼクティブリーダー、パートナー、お客様とのリアルタイムのインサイトの共有が、プログラミングのスキルなしで行えます。今すぐ[無料でお試しください!](#)

その他のリソース

[分析の文化を築く方法](#)

[分析を定義する](#)

[分析へのアプローチ \(ソリューションページ\)](#)

[クラウドデータブリーフィング](#)

[Tableau を使った高度分析 \(英語\)](#)

