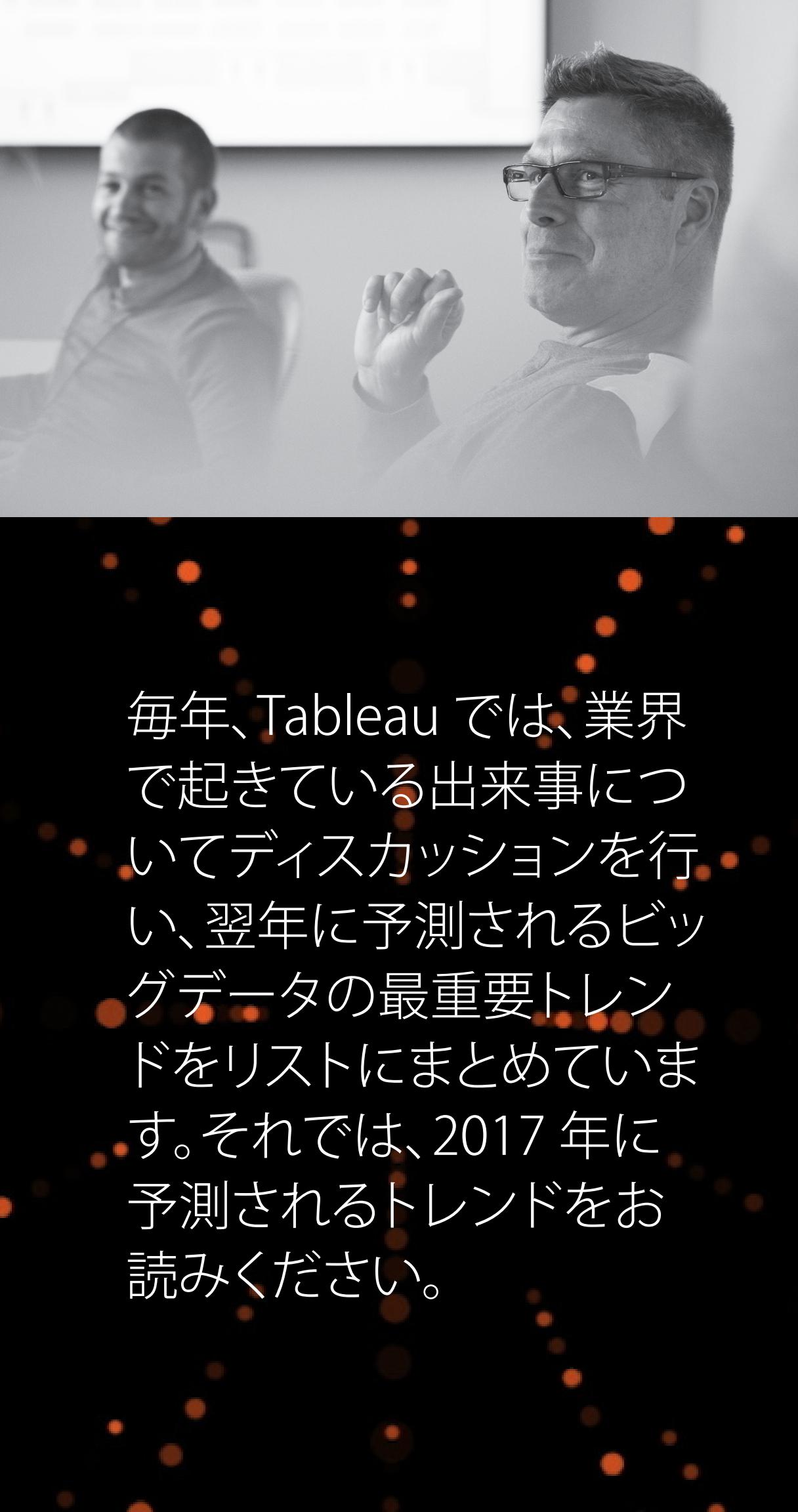


2017年
ビッグデータ
のトレンド Top 10

+ a b | e a u®



毎年、Tableau では、業界で起きている出来事についてディスカッションを行い、翌年に予測されるビッグデータの最重要トレンドをリストにまとめています。それでは、2017 年に予測されるトレンドをお読みください。

2017 年ビッグデータのトレンド Top 10

2016 年は、あらゆる形態とサイズのデータを保存および処理して価値を引き出す組織が増え、ビッグデータを語るうえで重要な年となりました。2017 年も、構造化データと非構造化データのどちらでも大量に扱えるシステムが増え続けるでしょう。市場では、エンドユーザーでもビッグデータを分析できるようにしながら、データ管理者がビッグデータを管理しセキュリティを確保するためのプラットフォームが求められると予想されます。こうしたシステムは洗練され、企業の IT システムおよび標準として活用されるようになります。





ビッグデータの加速と一般化: Hadoop を高速化する選択肢の増加

1

Hadoop では機械学習とセンチメント分析を行えますが、通常、疑問点としてまず浮かぶのは、インタラクティブな SQL はどれだけ速いのか、ということでしょう。SQL は結局のところ、より高速で繰り返し利用できる KPI ダッシュボードや探索的分析で、Hadoop データを使いたいと考えているビジネスユーザーにとってのパイプ的な役割を果たしています。

速度を求めるこのニーズにより、Exasol (英語) や MemSQL (英語) といったより高速なデータベース、Kudu (英語) などの Hadoop ベースのストア、そしてより高速にクエリを実行できるテクノロジーの導入が加速しました。また、SQL on Hadoop エンジン (Apache Impala、Hive LLAP (英語)、Presto、Phoenix (英語)、Drill (英語)) や、OLAP on Hadoop テクノロジー (AtScale (英語)、Jethro Data (英語)、Kyvos Insights (英語)) といったクエリアクセラレーターを利用してことで、従来のウェアハウスとビッグデータの世界を分ける境界線はより一層その姿を消していきます。

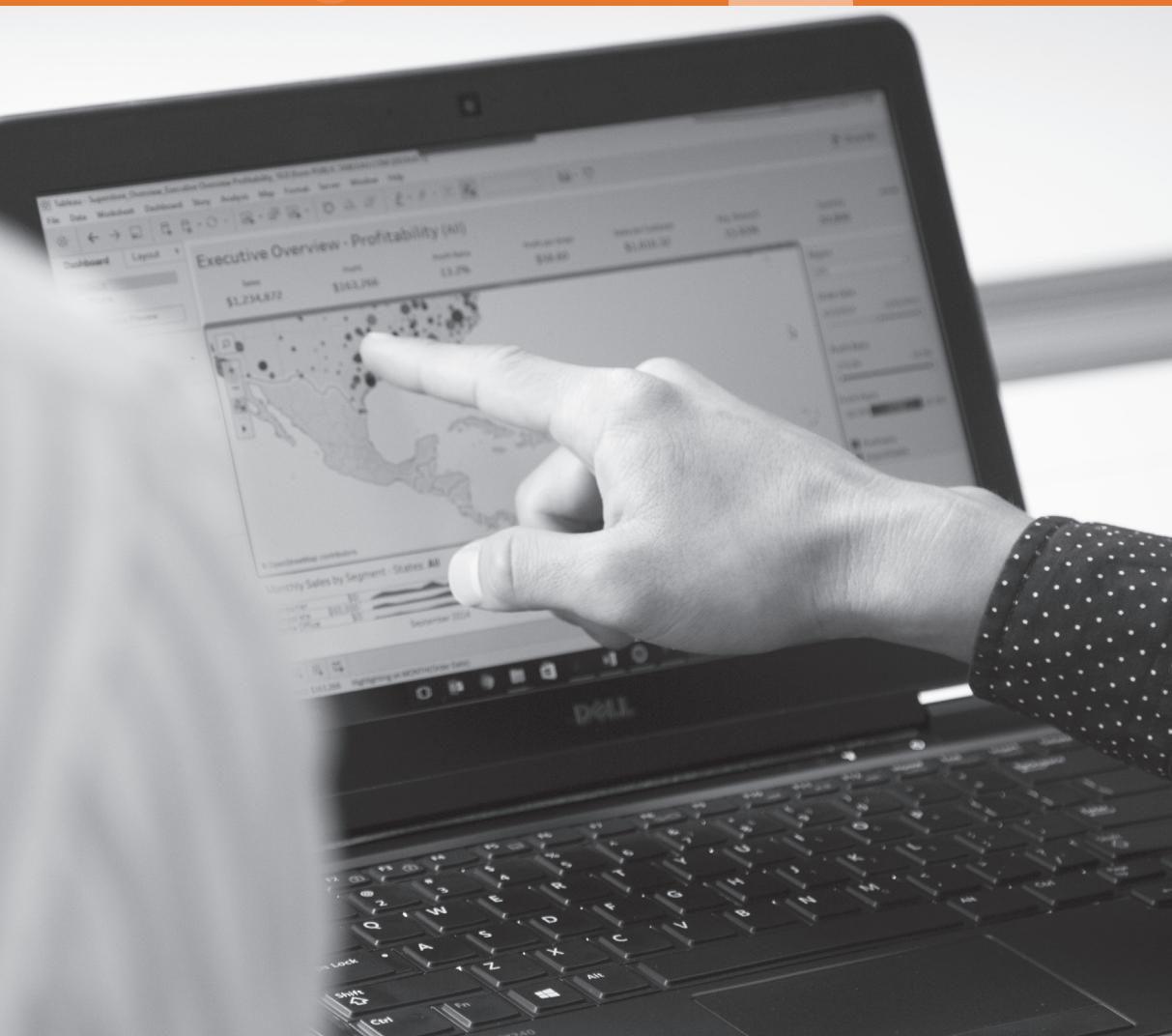
関連資料: AtScale BI on Hadoop の 2016 年第 4 四半期ベンチマーク (英語)

Hadoop を越えて広がるビッグデータ: Hadoop 専用のツールが陳腐化

昨年、ビッグデータの波に乗って、Hadoop 上の分析ニーズを満たすテクノロジーがいくつか現れました。しかし、複雑な異種環境を抱える企業は、1つのデータソース (Hadoop) のためだけに、サイロ化された BI アクセスポイントを導入することをもはや求めてはいません。質問の答えは、Systems of Record (記録のためのシステム)、クラウドウェアハウス、そして Hadoop と非 Hadoop の両ソースにある構造化データおよび非構造化データなど、数多くのソースの中に隠れているからです。(ちなみに、リレーションナルデータベースもビッグデータに対応し始めています。たとえば SQL Server 2016 は先ごろ、JSON をサポートするようになりました。)

2017 年、あらゆるデータの分析が求められるようになるでしょう。データやソースに依存しないプラットフォームが成功する一方で、**Hadoop 専用** のプラットフォームや用途に合わせて展開できないプラットフォームは、途中で失敗に終わると予想されます。[Platfora の買収](#) (英語) は、このトレンドが早々と現れたものと言えるでしょう。

関連資料: [非常識シリーズ: ビッグデータウェアハウス \(英語\)](#)



3



最初からデータレイクを活用し、データの価値を引き出す

データレイクは人工の貯水池のようなものです。まず終端をせき止め(クラスタの構築)、次に水(データ)を貯めます。そして湖ができたら、水(データ)を発電、飲料水、レクリエーションといったさまざまな目的(予測分析、機械学習、サイバーセキュリティなど)に利用し始めます。

これまでには、湖に水を貯めること自体が目的でした。しかし2017年には、Hadoopに対しビジネス上の正当な理由を示しにくくなることから、その状況が変わるでしょう。組織は、答えをより短時間で得るために、データレイクを素早く繰り返し利用することを求めるようになると予想されます。そして、社員、データ、インフラに投資する前に、ビジネス成果を慎重に検討するようになります。その結果、**ビジネス部門とIT部門**(英語)の間で、いつそう緊密なパートナーシップが育まれます。また、セルフサービス型プラットフォームも、ビッグデータ資産を活用するためのツールとして深く認識されることになります。

関連資料: [データレイクによるデータ価値の最大化\(英語\)](#)

4

アーキテクチャが成熟し画一的なフレームワークを排斥

Hadoopは、もはやデータサイエンスの場で使う単なるバッチ処理プラットフォームではなく、アドホック分析の多目的エンジンになりました。さらに、従来はデータウェアハウスで扱われてきたような、日々の作業負荷に対する業務レポーティングでも利用されるようになっています。

2017年、組織は利用場面に合わせたアーキテクチャデザインを追求し、多様なニーズに対応するようになるでしょう。また、データ戦略に取り組む前に、ユーザーのペルソナ、質問、数、アクセス頻度、データ速度、集約度といったさまざまな要素を調査するでしょう。このような最新のリファレンスアーキテクチャはニーズドリブンです。そして、ニーズの変化に合わせて再構成できるような形で、最も優れたセルフサービス型データ準備ツール、Hadoop Core、およびエンドユーザー用の分析プラットフォームを組み合わせます。最終的には、そのアーキテクチャの柔軟性でテクノロジーの選択が促進されるようになります。

関連資料: [コールド/ウォーム/ホットなフレームワークとそれが Hadoop 戦略に与える影響](#)

5



量や速度ではなく種類がビッグデータへの投資を促進

ガートナー社 (英語) は、ビッグデータを 3 つの V、つまり大量 (high Volume)、高速 (high Velocity)、多様 (high Variety) で表わされる情報資産として定義しています。3 つの V はすべて大きくなっていますが、多様性がビッグデータへの投資を促進する最大の要因となりつつあり、NewVantage Partners 社による最近の調査結果 (英語) にそれが現れています。このトレンドは、企業がより多くのソースを統合し、ビッグデータの「ロングテール」(英語) を重視しようと努めるにつれて、ますます高まっていくでしょう。スキーマのない JSON、他のデータベース (リレーションナルと NoSQL) のネストされたタイプ、フラットではないデータ (Avro、Parquet、XML) などデータ形式は増加の一途をたどり、コネクタの重要性が高まりつつあります。2017 年に分析プラットフォームは、そうした多種多様なソースにライブで直接接続できる能力で評価されるようになると予想されます。

関連資料: [量ではなく種類がビッグデータニシアチブの推進力に \(英語\)](#)

Spark と機械学習がビッグデータを活性化

Apache Spark (英語) はかつて Hadoop エコシステムのコンポーネントでしたが、今では企業に選ばれるビッグデータプラットフォームとなりつつあります。データ設計者、IT マネージャー、BI アナリストを対象に行われたアンケート調査 (英語) では、回答者のほぼ 70% が、バッチ指向でありインタラクティブなアプリケーションやリアルタイムのストリーム処理に適さない現在の MapReduce より、Spark を好んでいるという結果が出ました。

ビッグデータのビッグコンピュート機能は、計算集約型の機械学習、AI、グラフの各アルゴリズムを内蔵したプラットフォームの向上につながりました。とりわけ Microsoft Azure ML は、未経験者が使いやすく、既存の Microsoft プラットフォームとの統合も容易なことから、売上が伸びています。機械学習が一般に広まると、ペタバイト規模でデータを生成するモデルやアプリケーションがさらに生み出されていくでしょう。機械が学びシステムがスマートになるに従い、セルフサービスソフトウェアのプロバイダーが、このデータをエンドユーザーにとって利用しやすくなる方法に注目が集まる予想されます。

関連資料: [機械学習に Spark を利用すべき理由 \(英語\)](#)

合により、セルフサービス分析で新たなビジネスチャンスが誕生

2017年、すべての物にセンサーがつけられ、情報を送信するようになると予想されます。IoTは膨大な量の構造化データと非構造化データを生み出しており、クラウドサービスに置かれ共有されるIoTデータ(英語)の割合も増加の一途をたどっています。IoTデータは不均質であることが多く、HadoopクラスタやNoSQLデータベースなど、複数のリレーショナルシステムと非リレーショナルシステムに分散して保存されています。ストレージサービスやマネージドサービスのイノベーションにより収集プロセスは高速化されましたが、データ自体に対するアクセスと理解の面ではいまだに大きなラストマイル問題が残っています。結果として、クラウドでホスティングされている幅広いデータソースにシームレスに接続し組み合わせられる分析ツールは、需要が高まりつつあります。そのようなツールにより、企業はどこにあるどのような種類のデータでも分析して視覚化し、IoTへの投資における隠れたチャンスを見出すことができるようになります。

関連資料: IoTのラストマイル問題を解決する

エンドユーザーがビッグデータを 形作り始め、セルフサービスによる データ準備が主流に

現在、最大の課題の中に数えられるのが、ビジネスユーザーが Hadoop データを利用できるようにすることです。セルフサービス分析プラットフォームの広まりによって、その状況は改善を見せました。ただしビジネスユーザーは、分析前のデータ準備に伴う時間と複雑さをさらに減らしたいと考えており、多様なデータ型やデータ形式を扱う際はとりわけその点が重要になります。

セルフサービスによるアジャイル（俊敏）なデータ準備ツールでは、ソースレベルで Hadoop データが準備できるようになるだけでなく、スナップショットとしてデータが利用できるようになり、より短時間で簡単な探索が実現します。この分野では、エンドユーザーによるビッグデータのデータ準備に注力する [Alteryx](#)（英語）、[Trifacta](#)（英語）、[Paxata](#)（英語）といった企業が数々のイノベーションを起こしました。こうしたツールは Hadoop のレイトアダプターやラガード（英語）に対して導入のハードルを下げつつあり、2017 年も勢いは止まらないでしょう。

関連資料: [セルフサービスによるデータ準備がビッグデータのキラーアプリケーションである理由（英語）](#)



ビッグデータの成長: Hadoop が企業のスタンダードに

企業の IT 環境において、Hadoop が中心的な存在になるというトレンドはますます広がりつつあります。そして 2017 年、エンタープライズシステムではセキュリティとガバナンスのコンポーネントへの投資が増大するでしょう。Apache Sentry は、Hadoop クラスタに保存されたデータとメタデータに対して、ロールベースのきめ細かな権限設定を行うことができるシステムを提供しています。[Apache Atlas](#) (英語) は、データガバナンスイニシアチブの一環として誕生したものであり、組織がデータエコシステムの全体で一貫したデータ分類を行えるようにします。[Apache Ranger](#) (英語) は、Hadoop に対し一元化されたセキュリティ管理を行うためのツールです。

エンタープライズグレードの RDBMS プラットフォームでは、このようなタイプの機能が期待されるようになっています。これらの機能が新たなビッグデータテクノロジーの最先端となり、企業における導入のハードルがまた 1 つ取り除かれることになるでしょう。

関連資料: [Hadoop の成熟度: その将来](#) (英語)

メタデータカタログの登場で見つけやすくなる、分析する価値のあるビッグデータ

企業は長年にわたって、データ量が処理しきれないほど多かつたためにデータを破棄していました。Hadoop を利用すると大量のデータを処理できるようになりますが、データは通常、見つけやすい形に整理されることはありません。

メタデータカタログは、セルフサービスツールを使って、分析する価値がある適切なデータを見出し、理解するのに役立ちます。ユーザーのニーズのこのギャップは、[Alation 社](#) (英語) や [Waterline 社](#) (英語) などの、Hadoop 内のデータを見つける作業を機械学習機能で自動化する企業により埋まりつつあります。こうした企業は、タグを使ったファイルのカタログ化、データ資産の間にある関係の発見、さらには検索に対応した UI によるクエリの提案を実現しています。これにより、データ利用者とデータスチュワードの両者が、データで信頼性を確認し、データを見出し、正確なクエリを行うのにかかる時間を短縮できるようになります。そして 2017 年、セルフサービス分析の自然な延長線上で成長していくセルフサービス型のディスカバリーに対して、認知度と需要が高まるでしょう。

関連資料: [データレイクの戦略要件としてのデータカタログ](#) (英語)



Tableauについて

データビジュアライゼーションを小売業のプログラムやプロセスに取り入れるのは、意外と簡単です。

Tableau Software のミッションは、どれほどデータ量が大きくても、データの保存先が分散されても、お客様がデータを見て理解できるように支援することです。PC からでも iPad からでもシームレスにデータダッシュボードに接続し、ブレンド、視覚化、そして共有を短時間で行えるようになります。ダッシュボードを作成してパブリッシュすれば、データは自動更新され、同僚、パートナー、お客様と共有できます。プログラミング技術は必要ありません。無料トライアル版をぜひ今すぐお試しください。

TABLEAU.COM/TRIAL