


グレートデータ(高品質なビッグデータ)は偶然の産物ではありません。設計によって作られるものです。

意思決定を確実にサポートし、効果的なビジネスプロセスを実現するためには、高品質かつ安全で密接に連携されたデータを確保することが必要ですが、それは簡単な作業ではありません。

しかし、これに取り組まなければなりません。

A photograph of a server room with rows of server racks. The racks are filled with equipment and cables. In the background, a person is visible, possibly working or inspecting the equipment. The lighting is bright, and the overall scene is a typical data center environment.

市場のトレンドは、データプロフェッショナルにとって逆風になっています。

データを取り扱う仕事が、さらに難しくなっています。

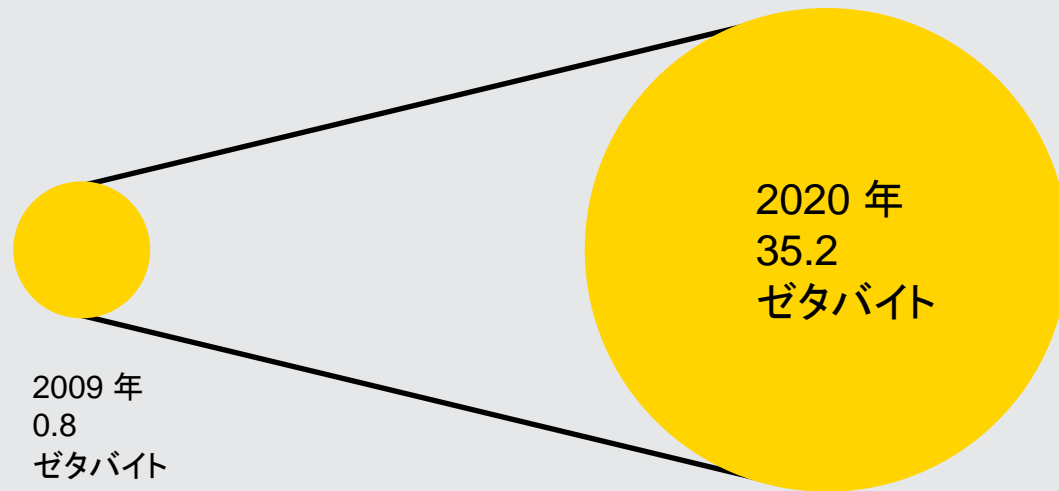
市場のトレンド

#1

より多くのデータが、より多くの
場所で、かつてないスピードで
移動しています。

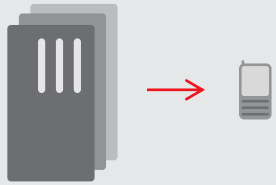
データの量やスピード、種類は、かつてないペースで増えています。今や、世界中で生み出されているデータ量は、2年毎に2倍のペースで増加しています。

新しいムーアの法則です。





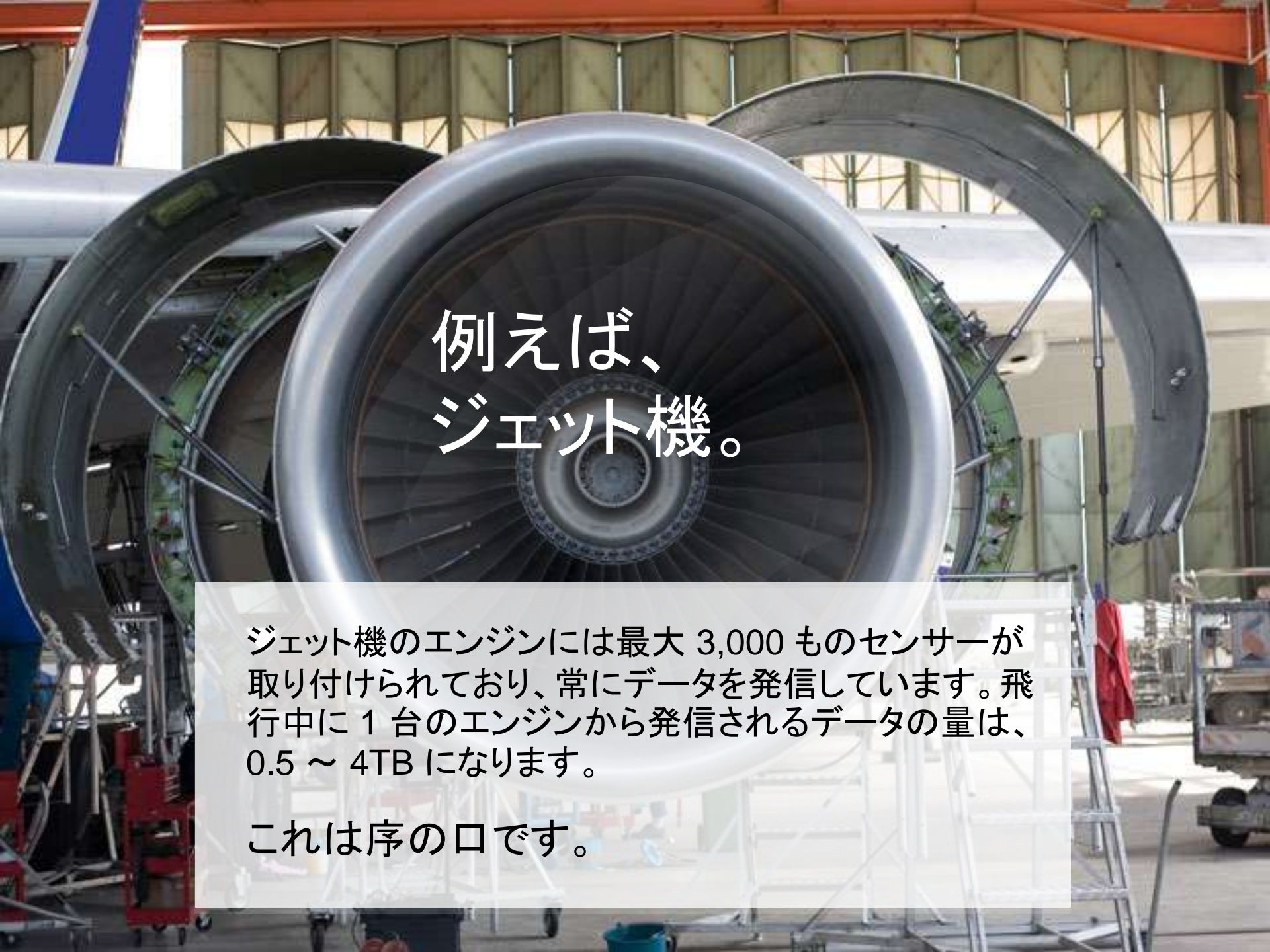
Hadoop などのテクノロジーがあれば、膨大なデータを手頃なコストで保管できます。



メインフレーム コンピューティング の処理能力を今では手のひらサイズで利用できます。



さらに、**RFID デバイス / センサー** があらゆるものに取り付けられています。

A large jet engine is shown in a hangar, with its fan blades visible. The engine is mounted on a stand, and the hangar's structure and windows are visible in the background.

例えば、 ジェット機。

ジェット機のエンジンには最大 3,000 ものセンサーが
取り付けられており、常にデータを発信しています。飛
行中に 1 台のエンジンから発信されるデータの量は、
0.5 ~ 4TB になります。

これは序の口です。

データの量や種類、
スピードは、これからもさらに
増加し続けるでしょう。

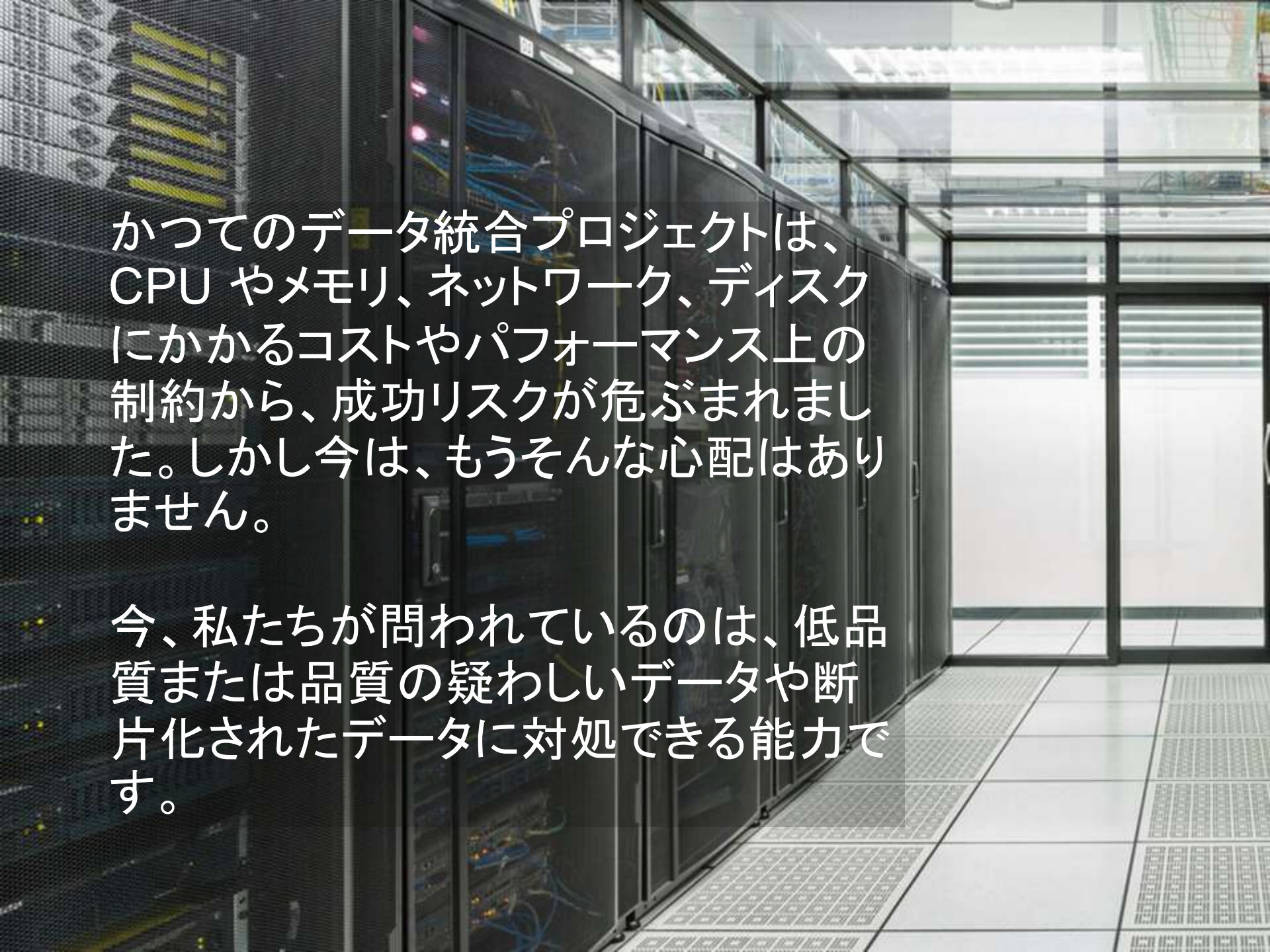
市場のトレンド

#2

データはどこにでもありますが、
その品質は疑わしいです



データは新旧あらゆるシステム内に存在します。
オンプレミスとクラウドの両方にあります。



かつてのデータ統合プロジェクトは、CPU やメモリ、ネットワーク、ディスクにかかるコストやパフォーマンス上の制約から、成功リスクが危ぶまれました。しかし今は、もうそんな心配はありません。

今、私たちが問われているのは、低品質または品質の疑わしいデータや断片化されたデータに対処できる能力です。

データが持つ価値を最大限に引き出すためには、組織全体にわたってデータを統合できる能力が必要です。

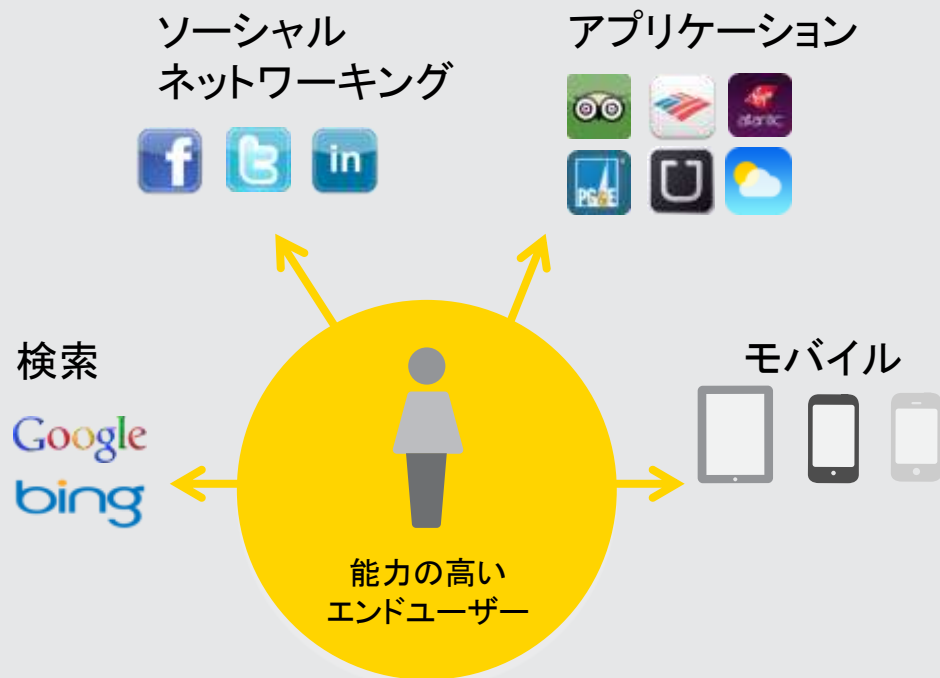
また、このプロセスに、データ品質管理を組み込む必要があります。80年代に製造業が製造プロセスに品質管理ステップを組み込んだことで転換したように、データにも品質管理を導入する必要があります。

市場のトレンド

#3

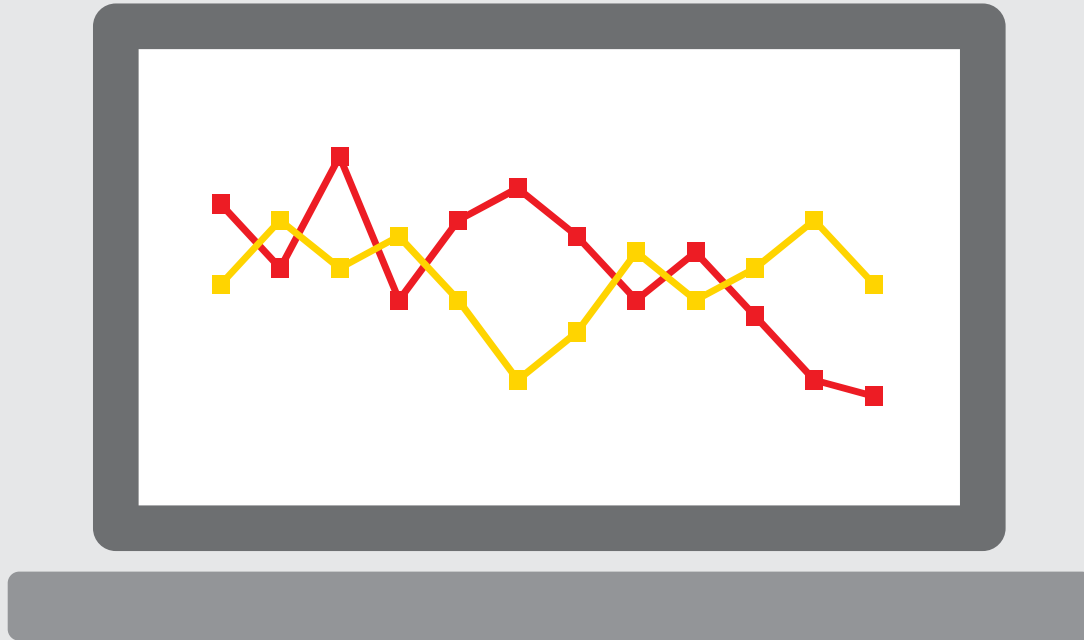
業務担当者はセルフサービスを
求めています

過去 5 年間で、業務担当者はさらに技術に詳しくなってきました。使いやすいテクノロジーが普段の生活で大きな役割を果たすようになり、何でも素早く簡単かつスマートにできるように助けてくれます。それが、個人の能力を高めたのです。そして彼らは、仕事にも同様のエクスペリエンスを期待しています。



業務担当者は、IT 部門が
高品質なビッグデータを
用意して、提供してくれる
まで待つのは、もう嫌だ
と思っています。
自分でやりたいのです。

業務担当者が、自分でデータを視覚化できる (見た目も恰好よい) セルフサービス ツールはあります。



しかし、1 回につき 1 つのデータセットしか操作できないという問題があります。

つまり、部門間でデータを利用する必要がある場合は、依然として IT 部門に頼らなければなりません。

最悪の場合、セルフサービス ツールでできることをすべてやってしまった後に、次のことに気付いて、IT 部門に助けを求める羽目になります。「自分が使用しているデータは基幹業務に関わる重要なデータで、それには基幹業務プロセスが必要...



...つまり個人の
ラップトップでは
処理できない」

セルフサービスには限界があります。



新しい考え方が
必要です

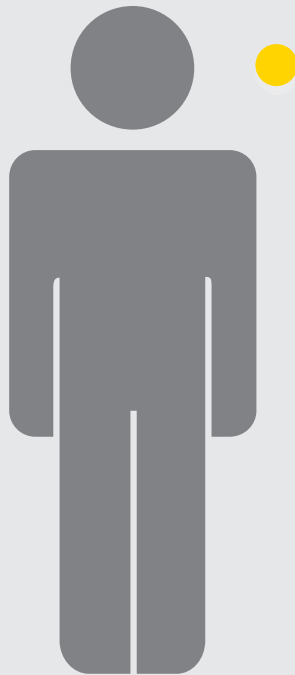
多くの企業は、コアビジネス プロセスに集中すれば、競争優位性を獲得できると考えています。



最も得意なことをやれば、競争に勝てる。

これは必ずしも間違いではありません。

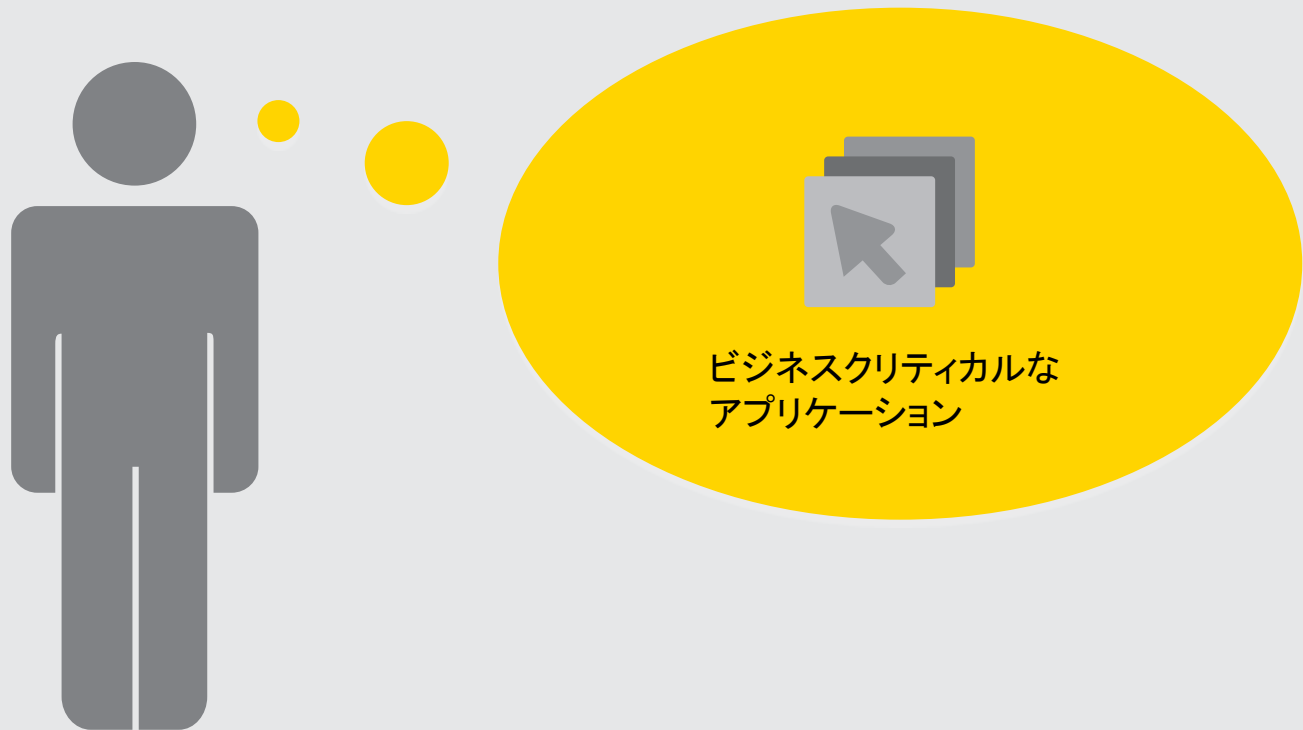
多くの企業は、彼らがコアとするビジネス プロセスを サポートするアプリケーションに投資すれば業務の 生産性が高まり、効率化によってサービスが向上す れば競争力が高まると考えています。



コアのプロセスを
自動化して改善できる
アプリケーションが必要だ。
そうすれば競争に勝てる。

これは必ずしも間違い
ではありません。

問題は、人々が依然としてビジネス アプリケーションを一枚岩だと思っていることです。

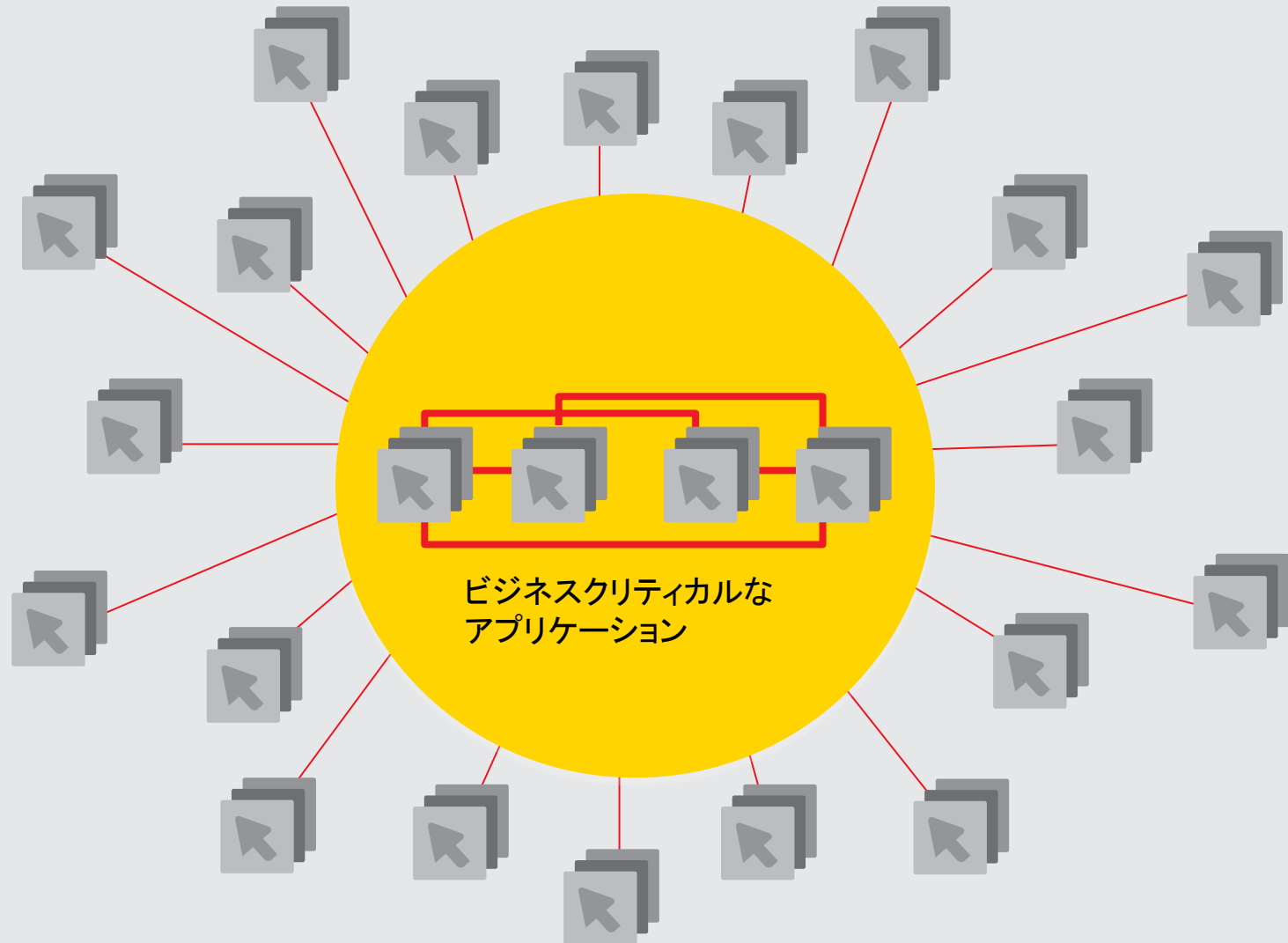


これが間違いなのです。

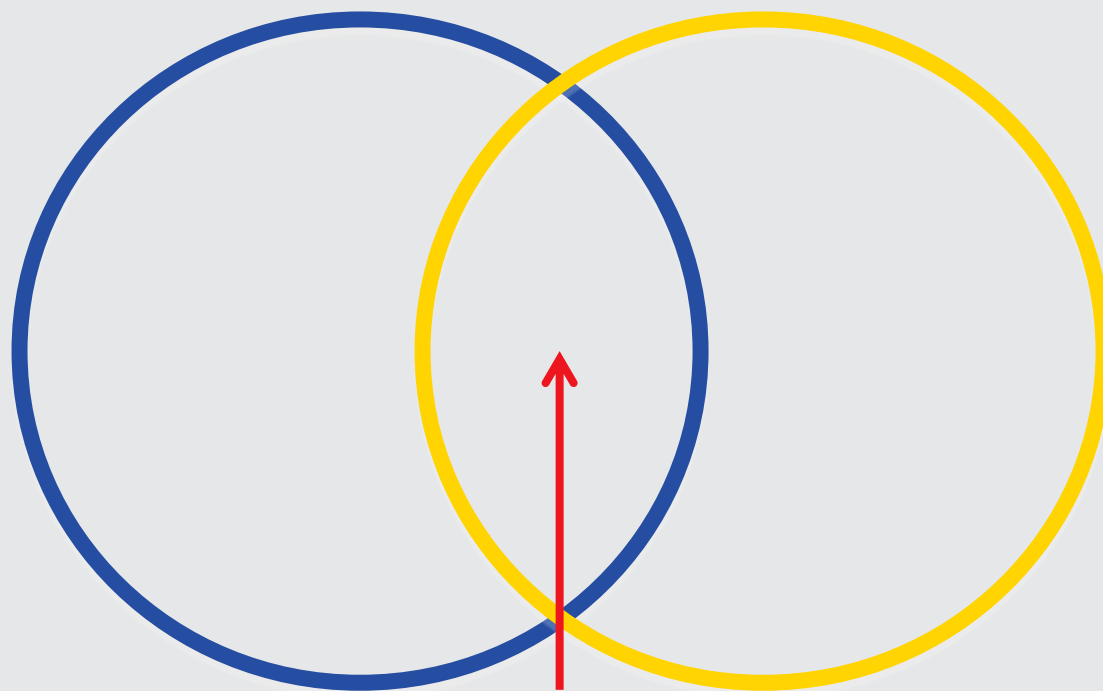
実際は、こうしたプロセスや、それを支えるコアアプリケーションは、一枚岩ではありません。今日のビジネスプロセスは、どれも複数のシステムに分散しています。



....データが出たり入ったりするシステムの数も、データポイントの数も、増加する一方です。

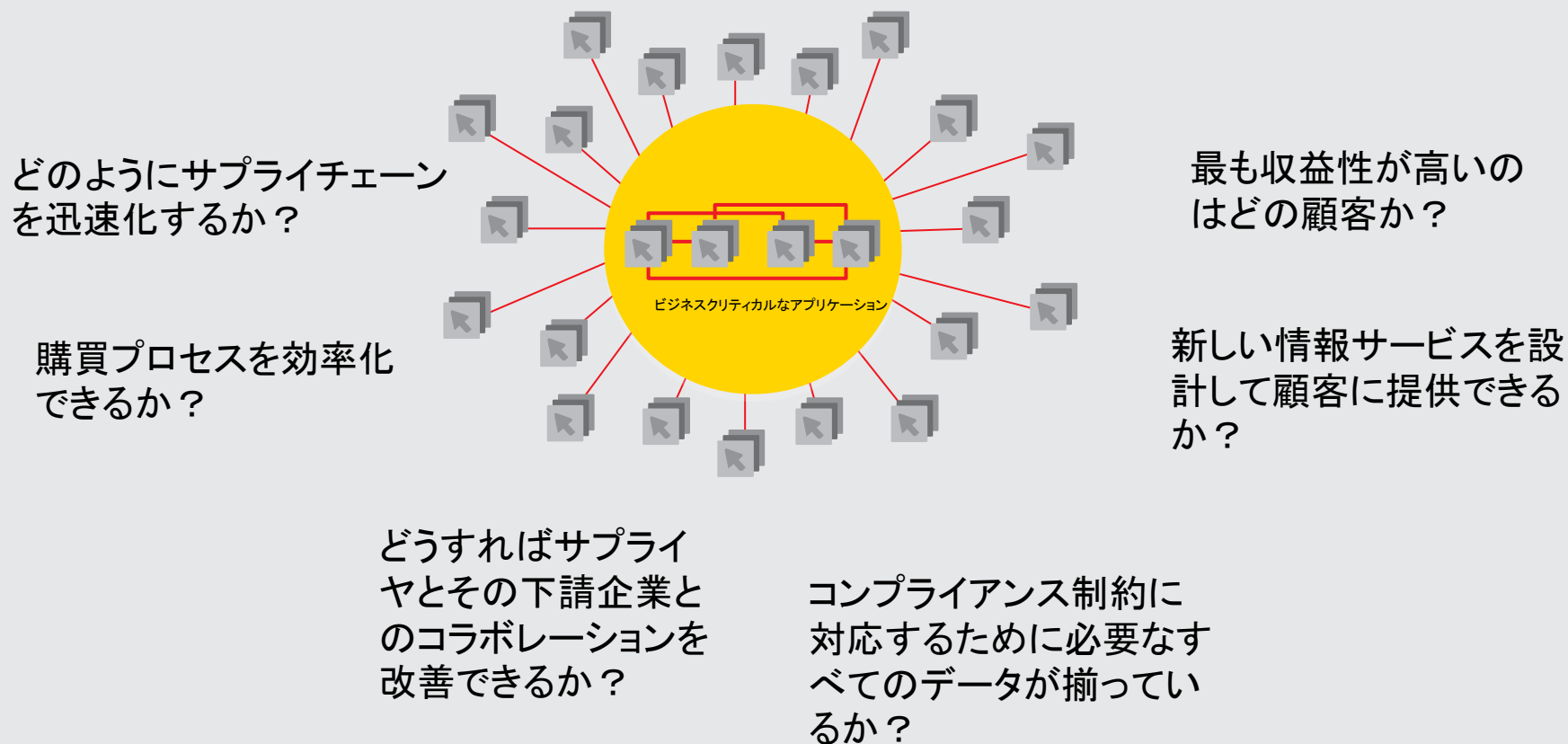


一般的にイノベーションは、異なる分野の境界線や、それらが重なる部分に存在します。




イノベーションが起きる場所

より多くのシステムの中で、より多くのデータが作成され続ければ、各分野の枠を越えてデータを統合し、交差させる能力が、次世代イノベーションの成功に不可欠となります。

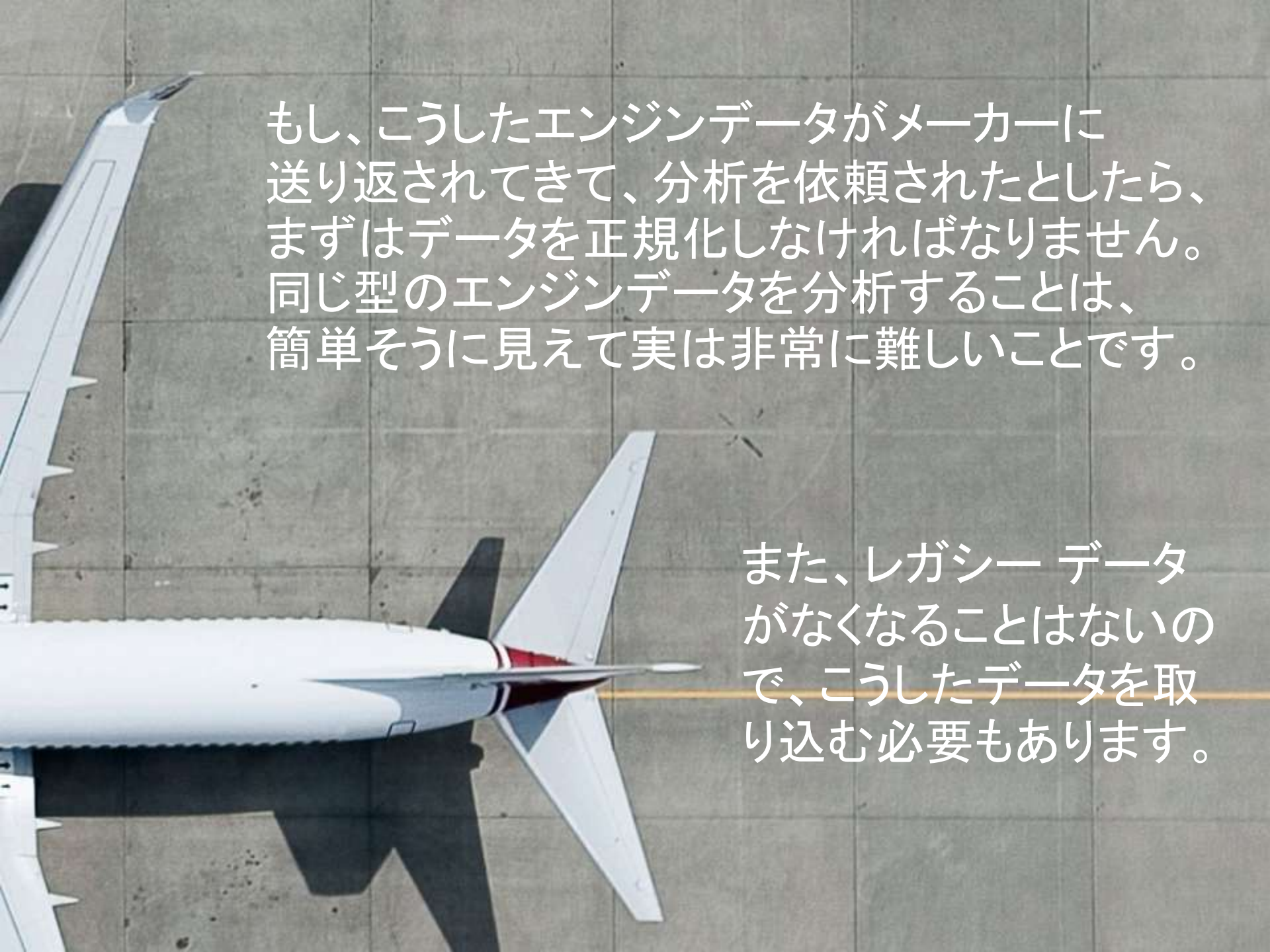


しかし、データ統合は、ほとんどの人が考えているよりも難しいものです。



ジェット機を例に考えてみましょう。どの機体もエンジンは同じかもしれませんが、エンジンに取り付けられた 3,000 ものセンサーから発信されるデータを制御するのは、エンジンを作ったメーカーではありません。航空会社です。

そして、航空会社は、こうした 3,000 もの属性を各社独自の形式で保存します。



もし、こうしたエンジンデータがメーカーに
送り返されてきて、分析を依頼されたとしたら、
まずはデータを正規化しなければなりません。
同じ型のエンジンデータを分析することは、
簡単そうに見えて実は非常に難しいことです。

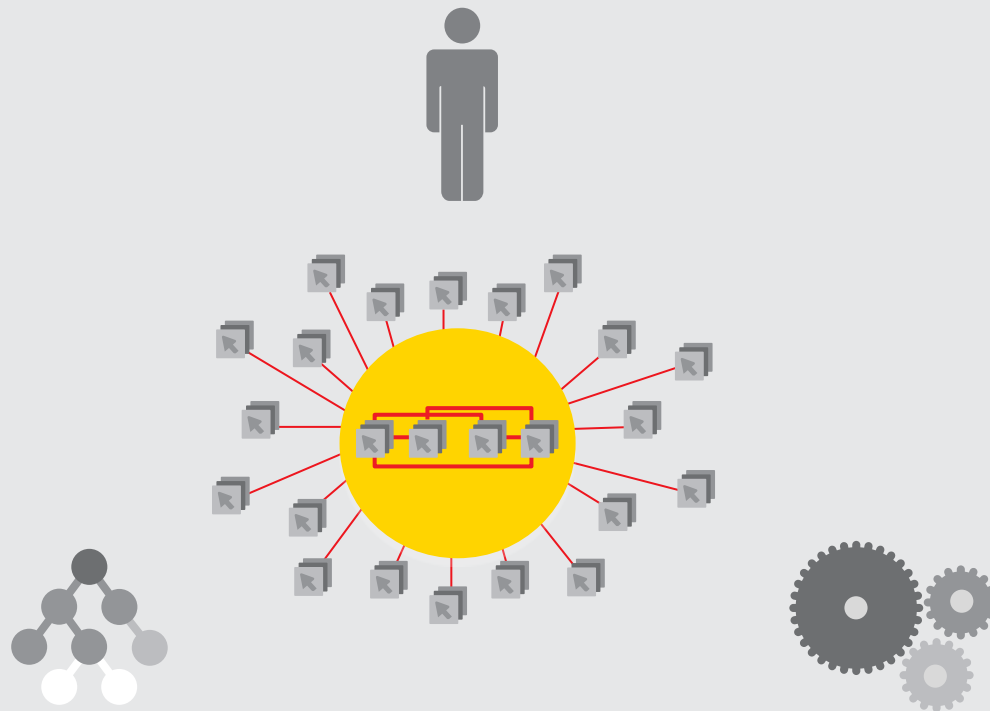
また、レガシー データ
がなくなることはないの
で、こうしたデータを取
り込む必要もあります。



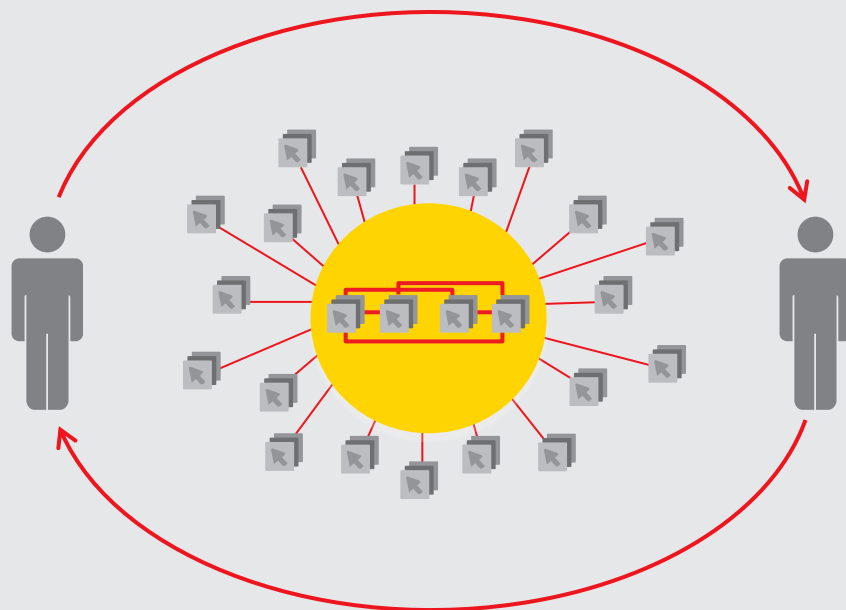
どんなデータプロジェクトにも、こうした課題が伴います。データプロジェクトは常に予想以上に難しく、その可変要素は増える一方です。

こうした課題を克服するためには、グレートデータを設計し、それをビジネスプロセスに組み込む必要があります。

コアのビジネス プロセスをサポートする人材やプロセス、テクノロジーに投資するのと同様に、こうしたプロセスをサポートするための分散データの統合に従事する人材、プロセス、テクノロジーにも投資する必要があります。



なぜなら、今やビジネスの俊敏性は、データ統合の俊敏性で決まるからです。そしてデータ統合の俊敏性は、いかに関係者全員の協力を得られるか、つまり業務部門とIT部門が適切なツールを使ってうまくコラボレーションできるかにかかっています。実際に、業務部門とIT部門のコラボレーションが実現している企業では、そうでない企業の5倍のスピードでデータ統合プロジェクトを実行しています。



5

グレートデータの
設計で考慮すべき 5
つの検討事項

#1

すべてのデータに接続するか

RDBMS やフラットファイル、XML、Hadoop、NOSQL、ソーシャルメディア、メインフレーム、マシンデータなどを含みます。

データ統合を実現すれば、数多くの豊富なデータソースからデータを組み合わせることで、1つのソースでは得られないような新しいビジネス情報を生み出すことができます。そのためには、RDBMS や NOSQL、メインフレーム、テキスト、アプリケーションなど、現在使用しているデータソースだけでなく、(新旧) すべてのデータソースに接続できるようなデータ統合ツールを選択しましょう。すべてのデータをひとつに集められるのは、こうしたユニバーサルな接続なのです。

#2

適切な形式とレイテンシに対応

バッチ、リアルタイム、準リアルタイム、構造化、非構造化、半構造化を含む。

数多くの異なるデータソースからデータを収集できるのと同じように、構造化、半構造化、非構造化などの多様かつ複数のデータタイプを、バッチおよびリアルタイムモードで使用できるようにデータ統合を設計しなければなりません。どんなデータにでも対応できる柔軟性のあるツールが必要です。

#3

データの構造と内容を理解する

データ プロファイリングをメソドロジーに組み込む

数多くの異なるデータソースに対応するには、データが想定通りの内容であることを保証する手段が必要です。一定レベルのデータ プロファイリングが実行できるツールを使って、システムに入出力されるデータをチェックすることで、期待通りの結果を確実に得られるように設計することが重要です。

#4

業務部門と IT 部門の効果的な コラボレーションを実現する リーンかつアジャイルであること

従来のように孤立した開発手法でプロジェクトを計画し、実行するような余裕はもうありません。リーンかつアジャイルな統合プロセスに対応し、業務部門と IT 部門のコラボレーションを実現することで、迅速かつインタラクティブに開発を進められるようなデータ統合ツールでなければなりません。

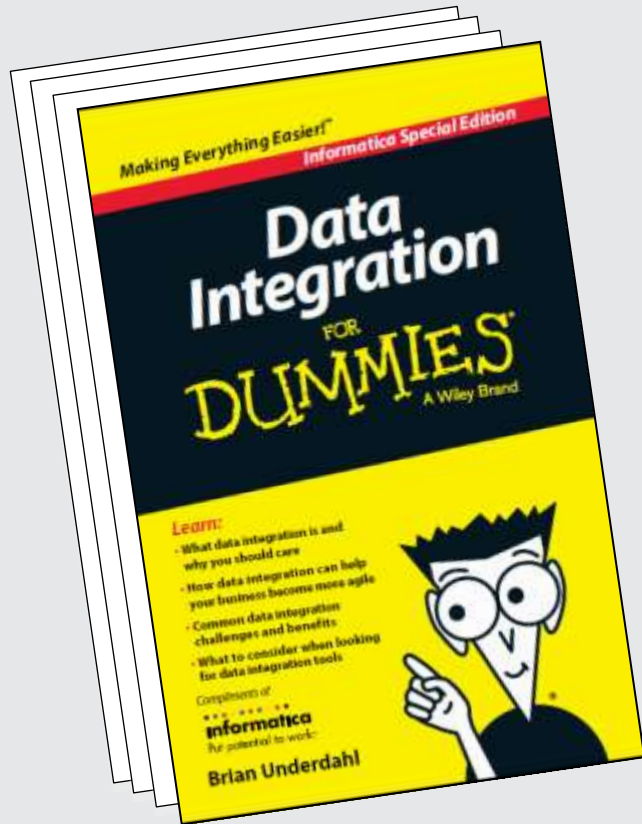
#5

企業の成長と拡大を サポート

スケールアップとダウンが可能であること

企業が成長すれば、プロジェクトの規模も大きくなります。今あるプロジェクトだけに適したツールに縛られたくはないでしょう。むしろ、部門レベルの小規模で革新的なプロジェクトから、全社レベルの大規模でミッションクリティカルな環境に至るまで、プロジェクトの規模や成長、移動に対応できるようなツールが欲しいはずです。

グレートデータを設計する方法
について、詳しくは以下の資料
をご一読ください。



[ダウンロードはこちら](#)