

# ビッグデータ時代を切り拓く

## —Field to Future Technology—

Accelerated Knowledge Creation through Big Data

武田 景  
Takeda Kei

加来 直子  
Kaku Naoko

あらゆるデータの収集・蓄積が可能となってデータ量が増加する一方、データのノイズの増加（S/N比の低下）が、データを扱ううえでの課題となっている。日立グループは、現場の事実から未来の業務に不可欠な情報を生成する「Field to Future Technology」を研究開発し、顧客の課題解決を図っている。データ可視化、データ仮想化、データ並列化、データ抽象化の四つのカテゴリから成る「Field to Future Technology」を通して、企業の情報利活用の進化を支えていく。

### 1. はじめに

企業の中でのIT（Information Technology）が果たすべき役割は、業務の効率化・自動化から、新たな価値創造・発見に向かって大きく変わりつつある。

ここでは、現場の事実から未来の業務に不可欠な情報を生成する技術「Field to Future Technology」について述べる。

### 2. ビッグデータがもたらす変革

#### 2.1 ITリーチャブルな世界の広がり情報S/N比の低下

本格的なインターネットの普及から15年以上が経ち、生活や企業活動の隅々までネットワークが張り巡らされてきた。モバイル機器・アプリケーションの進化、センサーデバイスの普及、画像など非構造データのデジタル化、IT機器のコスト低下、クラウドなどフィー型IT活用モデルの普及が相まって、あらゆるデータの収集・蓄積が可能となっている。しかし、人間が認知・処理できるデータ量は一定であり、かつ非常に限られている。データ量の増加は、用途が見つからないむだなデータやノイズの増加、すなわちS/N(Signal/Noise)比の低下も意味する。ビッグデータ活用をめざす企業のジレンマの一つはここにある。

#### 2.2 ビッグデータ活用は黎明(れい)明期

初期のインターネット普及を加速させたキラーコンポーネントは、Google<sup>※1)</sup>のページランク（PageRank<sup>※1)</sup>であった。Google以前の検索エンジンは、検索キーワードの出現回数などで検索結果をランキングし、欲しい結果を得るには何ページにもわたって検索結果を閲覧する必要があった。ページランクはインターネットのリンク構造を人気投票と見なすという卓越したアイデアで、ユーザーが望むページを非常に高い確率で検索結果の1ページ目に表示することができた。Googleのページランクの登場でインターネットは一気に整理され、今までノイズの山であったインターネットを宝の山に変えた。

一方、企業の現場の情報の活用は、そのデータ構造の多様さや、部門間の管理の問題などの観点で、Webよりも複雑であり、企業の現場でのデータの利活用が十分進んでいるとは言えない（図1参照）。

※1) GoogleおよびPageRankは、Google Inc.の商標または登録商標である。

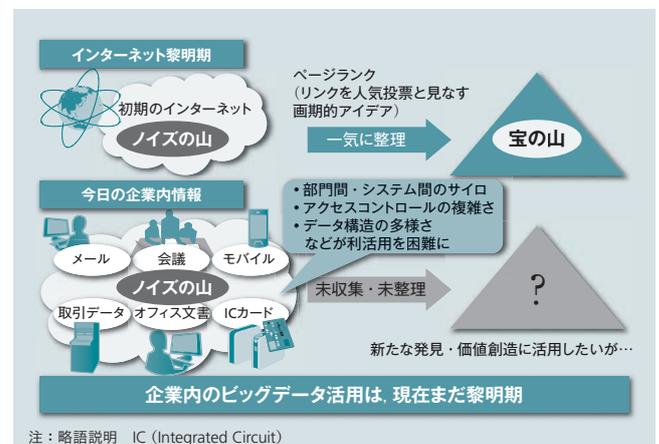


図1 | 黎明(れい)明期の企業内ビッグデータ活用

企業のビッグデータ活用には新たな技術・手法が必要となる。

Field to Future Technologyは、上記のような課題を解決する、現場の事実から未来の業務に不可欠な情報を生成するデータ利活用技術である。

### 3. Field to Future Technology

Field to Future Technologyは、(1) データ可視化技術、(2) データ仮想化技術、(3) データ並列化技術、(4) データ抽象化技術の四つの技術カテゴリーから成る(図2参照)。

データ利活用のライフサイクルは、データ可視化に始まりデータ仮想化、並列化、抽象化と進んで可視化に戻る。以下おのおのの技術カテゴリーについて述べる。

#### 3.1 データ可視化技術

データ可視化技術は「データをどのように集めてよいか分からない」、「データと業務とがどのように関連づけるかが分からない」という課題を解決する技術である。この技術は、現場の生データをITシステムから認識できるようにするための、データ収集・既存データ連携技術と、データを人間にとって分かりやすく表現するための、データとビジネスの関連を可視化する技術から成る(図3参照)。ここでは前者の技術分野について述べる。

電子マネー、IC(Integrated Circuit)カード、RFID(Radio-frequency Identification)を活用した物流管理などから生成される現場データは、量が多く、発生スピードが速く、データの構造が不完全であるという特性がある。したがって、データを活用するためには、高速かつ連続的にデータを処理し、また周囲の情報も活用してデータに意味を与え、ITシステムからデータを扱える形に加工していく必要がある。

このような問題を解決する技術の一つとしてストリームデータ処理技術がある。ストリームデータ処理技術は、センサーデータや株取引データのように、切れ目なく時系列に流れ込んでくるデータを、逐次的に処理する技術であ

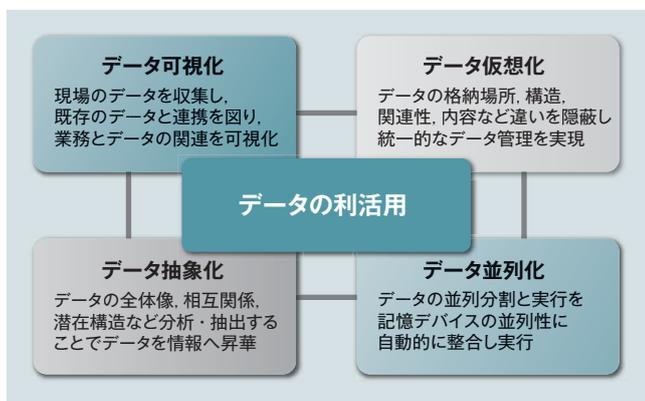


図2 | Field to Future Technology  
データ可視化、仮想化、並列化、抽象化の四つのカテゴリーから成る。

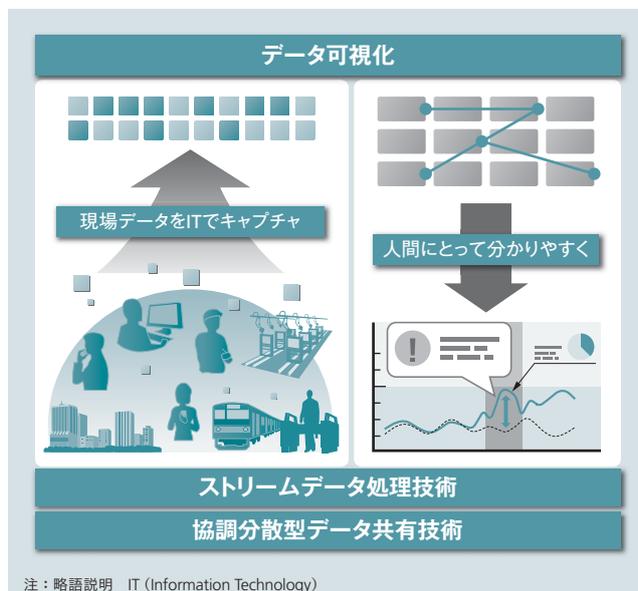


図3 | データ可視化技術  
現場データをITシステムで取り扱えるようし、またIT上のデータを人間にとって分かりやすく提示する。

り、特にインメモリで処理することで、データベースに格納する場合に比べ、より高速な処理を可能としている。この技術を活用することで、リアルタイムな状態の監視や、より鮮度の高いデータに基づいた意思決定ができるようになる。日立グループは、2002年からスタンフォード大学のSTREAMプロジェクトに参画し、データ可視化技術の開発育成に力を注いできた。また、この技術を2008年から製品として提供している。

#### 3.2 データ仮想化技術

データ仮想化技術は、データの格納場所、構造、関連性、内容などの違いを隠蔽し、人間やアプリケーションから見て統一的なデータ管理を実現する技術である。「いろいろな種類のデータを統一的に管理したい」、「業務ごとに異なった場所に保管されているデータを、場所を意識せずに使用したい」といった課題を解決する技術である。

企業内のデータは、放っておくと特定システムからしか利用できない、局所サイロになってしまう。業務一運用一アプリケーション一ITインフラが密に結合され、データをほかのシステムから再利用することが難しくなる。その結果、トータルでのデータ運用・管理コストが大きくなるだけでなく、別目的でのデータの再利用、すなわちビッグデータの潜在的価値を引き出すことが難しくなる。したがって、データ自体を仮想化し、格納場所やアプリケーションからの独立性を高め、自由に利用できるようにすることが重要となる(図4参照)。

この分野の技術の一つとして、日立グループはストレージの仮想化技術に取り組んできた。この技術は、実際の格

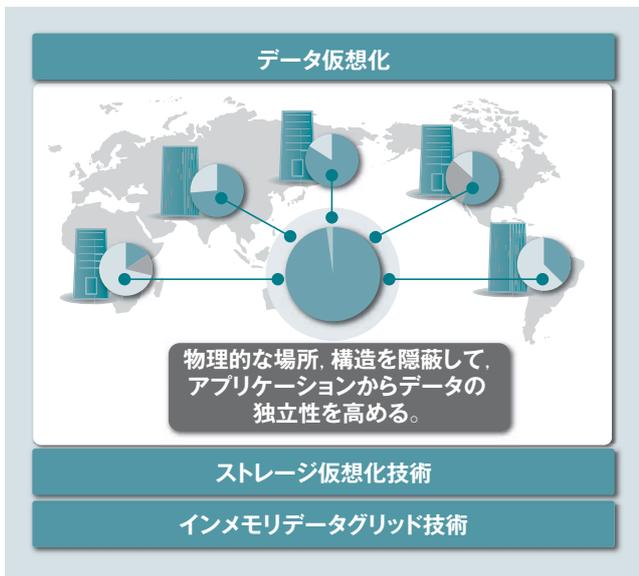


図4 | データ仮想化技術  
データとその物理的な位置やアプリケーションとの独立性を高める。

納場所や格納メディアの種別を、上位のアプリケーションから隠蔽することで、システムの柔軟性やデータの幅広い利用性を実現するものである。

例えば、ボリューム容量の仮想化技術により、データ量の拡大にあとから柔軟に対応できる格納インフラを提供してきた。

また、ストレージ階層の仮想化技術では、高速なメディア (SSD : Solid State Drive) から低速だが大容量のメディア [SATA (Serial Advanced Technology Attachment) HDD (Hard Disk Drive)] を一つのボリュームプールに混在させ、データアクセスの頻度に応じて最適な格納メディアにデータを自動的に再配置することで、さまざまなデータアクセスワークロードに応えられるストレージを実現してきた。

別の技術として、メモリ空間を仮想化するインメモリデータグリッド技術にも取り組んでいる。この技術については、次のデータ並列化技術の中で説明する。

### 3.3 データ並列化技術

データ並列化技術は、データを分散配置し、アプリケーション処理を複数の処理モジュールで並列化し、大量データを高速に処理できるようにする技術であり「業務フローで許される時間内に分析結果を得たい」、「リアルタイムに分析結果を見たい」といった課題を解決する技術である (図5参照)。

現場データを活用することで、今まで考慮の対象外だったデータを取り込み、新たな付加価値を創造する可能性を上げることができる。しかしながら、現場データは、発生スピードも量も非常に大きい一方で、データの単位量当た

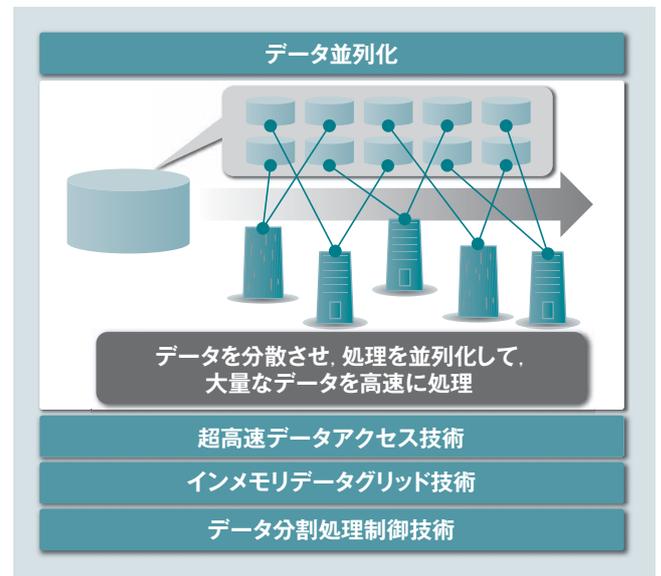


図5 | データ並列化技術  
データを分散配置し処理を並列化する。

りのバリューが小さいという特性がある。

したがって従来の処理モデルで、現場の大量データを処理することは、コストパフォーマンスの観点から得策ではない。特に、CPU (Central Processing Unit) のクロック数やHDDのI/O (Input/Output) 性能といった、処理コンポーネント単体で見た性能の伸びは長期の技術トレンドとして鈍化しており、単体で高パフォーマンスなCPUや格納装置の調達は少なくとも当面の間は高コストである。そのため、大量データ処理では、データを分散配置し、適切に並列処理していくことが重要になる。

この分野の技術の一つとして、日立グループは東京大学と共同で、超高速データアクセス技術およびそれを組み込んだデータベースエンジン<sup>※2)</sup>の開発に取り組んできた。超高速データアクセス技術は、データを多数のストレージに分散格納し、非順序でアクセス要求を発行することで、並列配置されたハードウェアリソースのそれぞれを極限まで使い切り、総体としてデータの検索を高速化する技術である。日立グループは、このデータベースエンジンと、高信頼・高性能なサーバおよびストレージ製品とを組み合わせ、事前検証済みのベストプラクティスモデルを高速データアクセス基盤として製品化し、2012年6月より提供している。

また別の技術として、インメモリデータグリッド技術にも取り組んできた。この技術は、複数の物理サーバにまたがってデータ分散格納することにより、仮想的に巨大なメモリ空間を実現する。これにより、サーバに搭載できるメ

※2) これらの技術には内閣府が創設した最先端研究開発支援プログラムで採択された「超巨大データベース時代に向けた最高速データベースエンジンの開発と当該エンジンを核とする戦略的社会サービスの実証・評価」(東京大学、日立製作所)で技術開発された成果が反映されている。

メモリ容量を大きく超えたサイズのデータを、インメモリで処理することが可能となり、高速なレスポンスが求められるリクエストに、同時に大量に応えることができるようになる。日立グループは、この技術を製品化し2012年2月より提供している。

### 3.4 データ抽象化技術

データ抽象化技術は、「何をすれば業務に有益な分析結果が得られるのかが分からない」、「数理統計知識がなくても分析をしたい」という課題を解決する技術である。

データ抽象化技術は、大量、複雑なデータを分析し、ビジネスに活用できる情報を取り出す技術である。従来、情報の抽象化は人間の得意分野であり、コンピュータは不得意な分野であったが、一方で大量の現場データを対象に、人間がすべて目を通すことは事実上不可能となっており、やはりITによる支援が重要となる(図6参照)。

この分野の技術の一つとして、日立グループは、障害予兆検知技術に取り組んできている。この技術は、単なる閾(しきい)値に基づく障害検知ではなく、過去のシステム稼働状況トレンドに基づいて、実測値が外れ値となっていないかをダイナミックに監視する技術である。

この技術は、統合システム運用管理製品「JP1」で蓄積してきたシステム稼働監視のノウハウと、前述のストリームデータ処理技術との組み合わせにより実現している。また、この技術により、システムの利用者がサービス低下に気がつく前に、その予兆を検知することができるようになる。

### 4. おわりに

ここでは、現場の事実から未来の業務に不可欠な情報を生成する技術「Field to Future Technology」について述べた。

これらの技術は、現場のデータをITで適切に処理することによって、人間の創造性を拡大し、未来の業務に不可欠な情報を生成することを意図したものである。日立グループは、ビジネスの現場・作業の現場・生活の現場、さまざまな現場の情報を、正しく処理・抽出し、必要とされる人や場面に適切に届けて業務やプロセスに反映していくことで、未来の大きな価値を生み出すことができると考えている。「Field to Future Technology」という名前には、その思いを込めている。この技術を通して、今後もさらに情報利活用の進化を支えていく。

#### 参考文献など

- 1) ビッグデータ時代を切り拓く -Field to Future Technology-  
[http://www.hitachi.co.jp/Prod/comp/soft1/data\\_tech/](http://www.hitachi.co.jp/Prod/comp/soft1/data_tech/)

#### 執筆者紹介



##### 武田 景

1996年日立製作所入社、情報・通信システム社 ITプラットフォーム事業本部 開発統括本部 ソフトウェア本部 先端情報システム研究開発本部 所属  
 現在、主にビッグデータのソフトウェア研究開発に従事



##### 加来 直子

1983年日立製作所入社、情報・通信システム社 ITプラットフォーム事業本部 事業統括本部 企画本部 ソフト企画部 所属  
 現在、主にブランドの企画、管理に従事

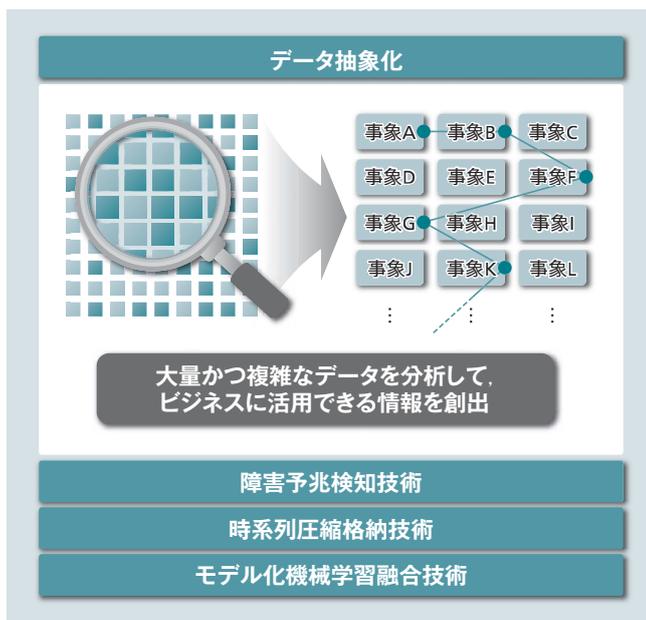


図6 | データ抽象化技術

大量なデータからビジネスに活用できるデータを創出する。