



White Paper

デジタル化社会での競争に不可欠な分析力の強化を促す IBM のストレージソリューション

Sponsored by: IBM

鈴木 康介

August 2016

調査概要

モバイル機器の普及とソーシャルメディアの発達、企業のデジタルデータの利用を大きく進めた。さらに、現在急速な IoT (Internet of Things) や機械学習の進歩がデジタル化のビジネスメリットを高め、社会変化を促進しようとしている。本ホワイトペーパーでは、国内企業がデータを中心とするイノベーションをリードし、競争環境の変化に対応する必要に迫られる中で、なぜデータ分析力の強化が重要になるのか、そのために IBM が提供するストレージソリューションをどのように活用すべきかを考察する。

デジタル化によって生まれる社会変化と IT の課題

インターネット、モバイル、ビッグデータ、機械学習などの技術が相乗的に発展してきたことでデジタルデータの利用価値が大きく変わった。近年、デジタルイノベーションという言葉聞く機会が増えたのは、かつての産業革命に匹敵するほどの大規模な変化が現在起きていると人々が認識している表れであろう。ただし、多くの国内企業がその潮流に乗り、グローバルに起きている変化をリードできているのかと言えば、残念ながらそうではない。それを果たしているのは一部の先進企業にすぎず、変化に対応する難しさが浮き彫りになっている。

変化が加速する現代

デジタル化に起因する社会変化は加速している。たとえば、ビデオ録画機よりも家庭向け PC の普及速度は速かった。100年間活躍した銀塩フィルムカメラは、短期間にデジタルカメラに置き換えられたが、その後、デジタルカメラの出荷ピークはわずか10年で訪れ、スマートフォンに主流の座を譲った。最近の SNS (Social Networking Service) の新たなサービスが広まる速度も非常に速い。さまざまな分野で製品やサービスの市場優位性を発揮する時期が短くなる傾向にある。

近年、人工知能 (AI : Artificial Intelligence) の研究がついに社会に影響を及ぼすまでになってきた。ディープラーニングは現在、機械学習の代表的な手法になっているが、その基礎となるニューラルネットワークの研究は50年以上も前に一度下火になり、長い雌伏の時を経ている。その間、1997年にIBMの「Deep Blue」が現役のチェス世界チャンピオンに勝利したニュースもあったが、次に一般の話題として大きな注目を集めた「Jeopardy!」(米国の有名なクイズ番組)での「IBM Watson」の勝利は2011年のことである。その後は、AIに関連するニュースを頻繁に目にすることになる。2011年末からはスマートフォンに音声アシスタント機能が搭載された。2014年にはエンタープライズ向けのAI利用が国内でも始まり、三井住友銀行を皮切りにみずほ銀行、三菱東京UFJ銀行もコールセンター業務にIBM Watsonを導入した。

このような機械学習の発展には、アルゴリズム研究の進歩もさることながら、大量に必要な検証用データを容易に入手できるようになった社会変化も影響している。ここ数年、ソーシャルデータのマーケティング利用が先行してきたが、今後はIoT利用の拡大によって製造での品質向上や

インフラ／装置管理の現場での異常検知、予防保守などにも AI が活躍し、さらに社会を変えていくはずである。

この急速な変化に企業はどのように対応すべきか。現在、市場でのポジションを確立している企業にとって、業界の垣根を越えた競合の出現、斬新なビジネスモデルで市場浸透を図る新興企業との競合は対処しにくい課題になっている。自社のビジネス価値をどこに求めるのか、変化が緩やかであった時代とは考え方やアプローチを変える必要がある。デジタル化という大きな変化を取り入れていくには、正しい状況分析、優れた経営判断とトップダウンの改革が必要になる。

データセントリックイノベーションをいかに起こすか

企業が長期に渡って市場優位性を維持するには、Incremental Innovation（漸進的革新）と Radical Innovation（急進的、非連続な革新）が必要である。漸進的革新は通常の企業努力によってもたらされるが、大きな変化や長期的な変化に対応するには急進的革新も不可欠である。現代では2つのイノベーションの重要性は以前よりも増しているわけだが、その様相に加わった重大な変化は、データの利用がイノベーションに重要な役割を果たすようになったことである。データセントリックイノベーション（データを中心とした革新）の例を挙げると次のようになる。

- **漸進的革新**：ビジネスインテリジェンス（BI）やビッグデータ分析による業務改善、IoT 活用による製品やサービスの品質向上、パーソナルマーケティングによる収益最大化
- **急進的革新**：デジタルデータを活用した新たなビジネスの創出、API 公開（企業間のデータ共有）を通じた新機軸の協業の誕生、ビジネスに有用な新たなデータの提供（匿名化された医療データなど）

このように改めて列記すると、漸進的革新はノウハウの蓄積が可能であり、努力によって成功が見込めるが、急進的革新を生み出すことは容易ではなく、尽くした努力と成果は必ずしも比例しないことがあると推測される。実際、急進的革新を生みやすい組織風土の醸成は難しい経営課題であった。その促進のためにナレッジマネジメントに取り組む企業も少なくないが、組織の壁やキーパーソンの多忙さで情報共有が失敗しがちである。

しかし、その状況は変わるかもしれない。原因は AI の進化である。ナレッジマネジメントにおけるヒューマンリソースの問題は AI 技術の活用で解消される可能性が出てきたからである。これまで「若い社員のアイデアとベテランの知識を組み合わせてイノベーションを」と期待しても、経験豊富な第一線の社員は忙しく、時間が取れなかった。この問題は AI によるコーチングで解決できる。実は、前述の IBM Watson は専門家に知的刺激を与えるといった有用性をすでに示している。大量の料理レシピをデータとして学習させ、キーワードに応じたオリジナルレシピを IBM Watson に考案させるというデモンストレーションにおいて、IBM Watson との協働作業を経験したフランス料理のシェフが「人間の創造性への AI 支援」の可能性を肯定的に評価していた。デジタルデータを基に起こるイノベーションは、AI の進化という要因が加わることで、いっそう加速するであろう。

分析、ナレッジマネジメントを成功させるのはデータ

データセントリックイノベーションを活性化させる方法として分析／BI、ナレッジマネジメントの環境整備が必要であると分かれば、仕組みの部分は資金投入によって整えられるであろう。ただし、それがビジネスに有効な投資になるのか、競争力の強化に結び付くのかは、どのようなデータを利用するかにかかっている。何らかを契機にして方針が定まり、分析強化への投資が決まった場合、分析システムの構築は一定の時間で可能であろう。多様なクラウドサービスを利用できる現在は、準備期間も短くて済む。しかし、有効な分析結果を得るためのデータを集めるのにどれだけの期間が必要であるのか、分析内容にも依存するが、事前には知り難い。昨今、分析技術は急速に発展しており、非構造化データの分析も可能になっていることから、できるだけデータを蓄積しようとする企業が増えている。そのような企業が今、注目すべき IT インフラが、**非構造化データを含む多種多様なデータのリポジトリ「データレイク」**である。

データレイクが主流となる背景：変化への対応力

データセントリックイノベーションを支える IT インフラ構築に向けた現実的な施策として、データレイクの整備が米国ユーザーを中心に盛んである。Hadoop、Apache Spark などの分散処理技術の進歩により、大量データを低廉なコストで高速に処理できるようになったことと、ソーシャルデータやセンサーなどから大量のデータが低コストで取得可能になったことが合わさり、非構造化データからも有効な知見を導出しやすくなったためである。米国では、ソリューションベンダーもユーザーも一気にデータレイクへと向かったためにその構築が一巡し、現在はデータレイクの運用改善が提案されるフェーズに移っている。データガバナンスにより重点を置いた「Data Reservoir（データの貯水池）」「Data Refinery（データの浄水施設）」という呼び名が提唱されている理由にはそのような背景がある。

では改めて、従来、業務データ分析に使われてきたデータウェアハウス（DWH）と新しい概念であるデータレイクとの違いを確認しよう（Table 1 参照）。本来、分析の生産性を優先するのであれば、どのようなデータでも対象にしたいし、データ量にも制限を設けたくない。しかし、データ処理能力が低かった過去には、そのような理想をシステムに具現化する方法はなかった。現実的な処理時間で分析しようとすれば、その演算を短縮するための準備（チューニング）を DWH 側に施すことで、分析時の負荷を軽減するよりほかなかったのである。しかし、処理時間の短縮については多くの技術が登場した。Hadoop を始めとする分散処理や NoSQL データベースの技術、CPU 性能の向上、フラッシュストレージの普及などは処理の高速化に大きく貢献した。その結果、分析用のインフラは対象データの制限を取り払い、理想に近付けたのである。

データレイクの構造的なコンセプトを最もよく表す言葉は「schema on read」である。DWH はデータを格納する際に形を整え、取り出しやすい構造を維持する必要があった（schema on write）。それに対しデータレイクは、データを加工せずに保存する。そのまま自然の形で貯めておくという意味合いでレイク（湖）という言葉が使われたのである。そのため、データレイクのテクノロジーの焦点は、多様なデータ形式の分析を可能にするための方法をいかに幅広くするのかに向けられている。

TABLE 1

データウェアハウスとデータレイクの比較

	データウェアハウス	データレイク
データ	組織内の構造化データが中心	非構造化データ／社外のデータも扱う
運用の焦点	データ増加に対応するチューニング	データ多様性（新たなデータ）への対応
分析の柔軟性	低い	高い
構築コンセプト	schema on write	schema on read
用途	現状分析	現状分析（リアルタイム分析を含む）と将来予測
イノベーションタイプ	Incremental（漸進的）	Incremental & Radical（漸進的と急進的）
データ増や高速化要求に対するストレージソリューション	ハイエンド製品によるスケールアップで対応	<ul style="list-style-type: none"> ・スケールアウト型による拡張性を利用 ・性能と経済性を柔軟に選択できる Software-Defined Storage による構築

Source: IDC Japan, August 2016

あらかじめ分析内容を想定してデータを保存する必要がなく、分析ニーズが生まれてから方法を工夫しようというのがデータレイクの考え方であるため、利用者にとって自由度が高く、**市場に変化が生じたリアクションとしての分析**にも活用しやすい。アーキテクチャとしてはスケールアウトしやすいため、データが増えた場合の拡張性にも優れている。現在は分析のために構築されているが、AIが発達すればナレッジマネジメントのための活用も見えてくるであろう。AIを相手にデータレイクを使ったブレインストーミングを行い、イノベーションを生むというのは遠い未来のことではないと思われる。

分析の強化に向けたストレージ選択の重要性

分析対象データとして、どこまでの範囲にすべきかは、データの保存にかかるコスト、つまりストレージインフラにかかる予算にも依存する。分析速度の向上についても選択肢はいく通りかあるため、データ処理のタイプに応じて最適なストレージを選ぶことが投資効率の向上につながる。かつては、速度や容量の要求に対してハイエンド製品にスケールアップするより手段がなかったが、今日では変化に対応しやすいアーキテクチャ、Software-Defined Storage を利用することであらかじめ準備ができる。その意味でIBMのストレージソリューションは、Software-Defined Storage を始めとする分析プラットフォームやデータレイクに利用価値の高いストレージ製品、クラウドサービスを包括的に取り揃えており、検討する価値は高い。

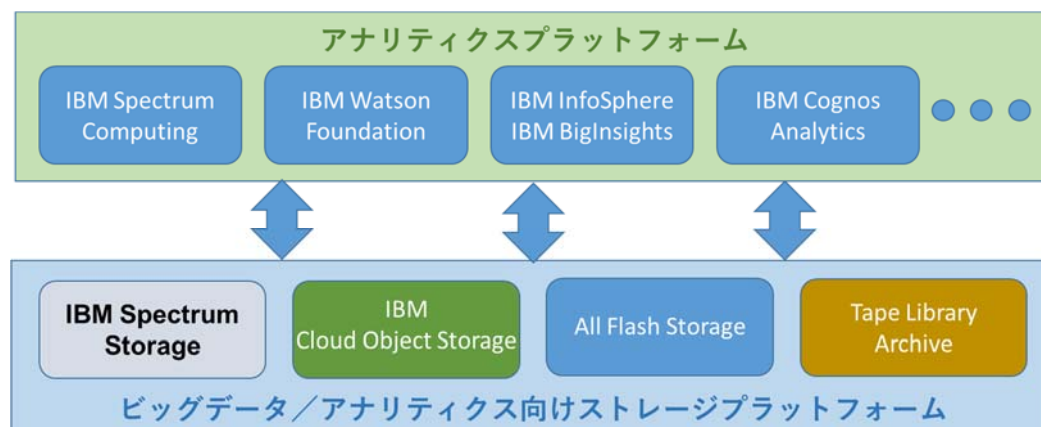
IBMのストレージソリューションによる変化への対応

社会変化に直面する顧客の課題に着目し、そのソリューションに積極的な先行投資を行っているのがIBMである。近年、「Smarter Planet」ではよりスマートな社会を実現させるテクノロジーを結集させていたが、さらに直近では「コグニティブ・コンピューティング」をキーワードに、人工知能／機械学習の商用化を牽引している。そのIBMのストレージ戦略（開発／M&A）も、やはり顧客課題の変化を見通したものになっている。

IBMのストレージ製品ポートフォリオの第一の特徴は幅広さである。そのためワークロードに応じて、数種類のストレージを組み合わせる柔軟性の高い統合ストレージインフラを構成しやすい。IBMが提供するストレージ管理ソフトウェアを使えば、適材適所なストレージを組み合わせることと、多様化による複雑性を回避することが両立できるからである。また、スケールアウト型ストレージやSoftware-Defined Storageの製品を揃えているため、スモールスタートが可能であり、要件の変更にも応じやすい。以下では、IBMのストレージポートフォリオの中でも分析ニーズに対応するストレージ製品に絞ってその利用価値を検証する。

FIGURE 1

IBMのアナリティクス向けストレージ製品群と分析プラットフォーム



Source: IBM 資料を基に IDC 作成, August 2016

多様なデータ分析ニーズをカバーするストレージラインナップ

企業がデータ分析を行うのは、状況の変化を把握したり、施策の効果を検証したり、新たなニーズを探索したり、いずれもスピードが要求される場面である。また、先にデータレイクのセクションでも考察した通り、分析プラットフォームにはデータの多様性、大容量化への対応も必要となる。こうした分析ニーズに応える IBM のストレージ製品として、IBM Spectrum Storage、IBM Cloud Object Storage (Cleversafe)、オールフラッシュストレージ、テープライブラリーを挙げることができる (Figure 1 参照)。これらのストレージは必ずしもすべてを階層的に組み合わせる必要はなく、ユースケースに応じて取捨選択しながら柔軟にインフラ構築ができる。

そのストレージを利用するアナリティクスプラットフォームは、分散処理、コグニティブからセルフサービス BI まで豊富に用意されていることも IBM を使うことのメリットである。同社のアナリティクスプラットフォームには 2016 年 6 月発表の「IBM Spectrum Computing」という最近ラインナップに加わったソフトウェア群もある。同製品群は、性能が求められるアナリティクスや機械学習を高速化するため、コンピューティング資源を最大限に使用するためのワークロード管理を行うと共に Hadoop や Apache Spark などのオープンソースフレームワークの導入を簡素化する。

IBM Spectrum Storage : 変化対応のコアとなる Software-Defined Storage

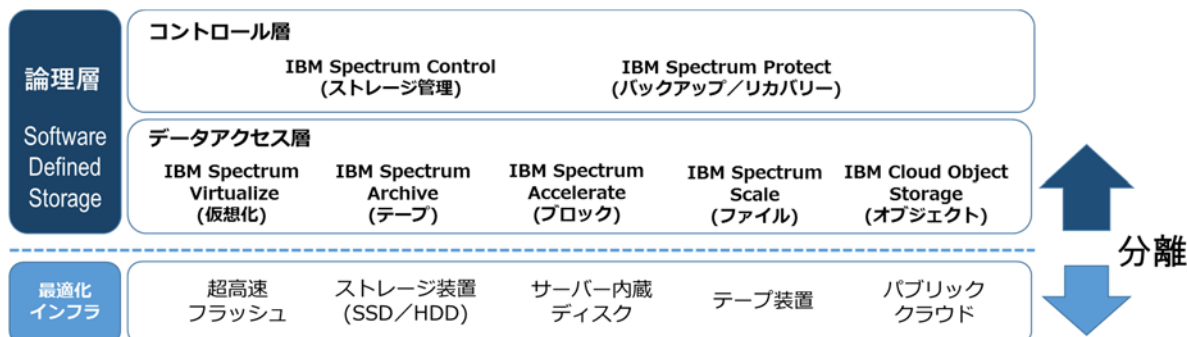
Software-Defined Storage を中心に多様なストレージ管理製品を加えた「IBM Spectrum Storage ファミリー」は、ソフトウェア製品であることを生かし、データ種類、データ処理ニーズの多様化とデータの大容量化への対応を両立する役目を担う。ゆえに IBM Spectrum Storage は、分析プラットフォームの柔軟性を保つ必要があるデータレイクを構築する上で核となる存在になる。

その第一要件、柔軟性は、SAN ストレージ (ブロック)、スケールアウトストレージ、分散ファイルシステム、オブジェクトストレージ、テープアーカイブという多様なストレージ製品の組み合わせをコントロール層で管理統合することで実現される (Figure 2 参照)。経済的にも Software-Defined Storage としてコモディティサーバーを利用でき、さらには大量のデータを低い運用コストで長期保存できるテープアーカイブを組み込めることは有利である。

ユーザーが将来の変化に備えたストレージプラットフォームを構築したいと考える場合にも、「IBM Spectrum Storage」は多くの選択肢を提供することでそれに 대응することができる。当初、対象としていなかったデータの分析ニーズにも追加投資を抑えながらストレージインフラ改革を可能にするプラットフォームであり、データレイク的设计コンセプトを具現化しやすい。

FIGURE 2

IBM Software Defined Storage と IT インフラソリューション



Source: IBM, June 2016

拡張性と柔軟性でデータレイクにも適するオブジェクトストレージ

「IBM Cloud Object Storage (Cleversafe)」は高い拡張性を備えたオブジェクトストレージである。2015年11月、買収によりIBMのポートフォリオに加わった。同製品の分散保存技術は管理情報も同時に分散させることで拡張性の制限を回避しており、構造的な容量制限を持たない。また、データ冗長性を確保するためのイレージャーコーディング技術は効率が高く、3重の冗長性を実容量の1.7倍のディスク領域で実現している。しかもそれは地域に分散したデータセンター間で実装できる冗長構成のため、従来方式でサービスプロバイダーが行っている拠点間／拠点内のミラーリングを組み合わせた合計4倍のディスク領域から考えると大幅な削減と言える。

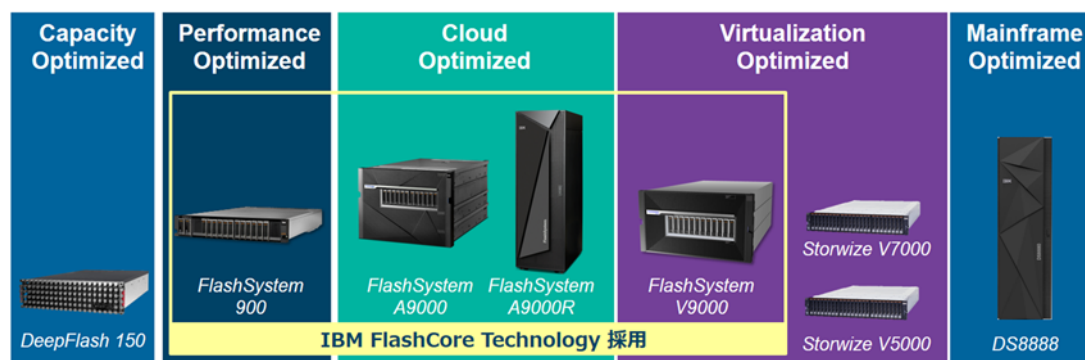
データ分析用のストレージとしてIBM Cloud Object Storageが利用しやすいもう一つの利点は、対応するアクセスプロトコルが豊富なことである。REST形式のAPI (Application Programming Interface)、Amazon S3 互換 API、OpenStack のオブジェクトストレージ「Swift」互換 APIに加え、ビッグデータ分析に適する「HDFS」の互換 API を備えている。データレイクの要件であるアクセスの柔軟性はIBM Cloud Object Storageの機能として実現されているため、分析の際に他のストレージにデータを移す必要がない。これは大容量のデータ分析を行う際には大きな利点となる。

分析の高速化を担うIBMのオールフラッシュアレイ製品

今からちょうど60年前、市場に初めてハードディスクドライブ (HDD) を投入したIBMはディスクストレージのパイオニアとして業界を牽引してきた。媒体がフラッシュメモリーに切り替わろうとしている現在も、IBMは国内オールフラッシュアレイ市場において3年連続で売上額シェア首位 (『IDC Japan Quarterly Enterprise Storage Systems Tracker 2016Q1』を参照) となり、ポートフォリオ全体のフラッシュ化を図るなどパラダイムシフトを先導している。IBMのオールフラッシュアレイ製品は用途／目的別に、5つに分けられている (Figure 3 参照)。容量志向、パフォーマンス志向から、クラウド向け、仮想化環境向け、さらにはメインフレーム用まで幅広いニーズに対応している。その中心的な製品ファミリー「IBM FlashSystem」は、市場にあるさまざまなオールフラッシュアレイ製品の中でも、高速性と可用性に競争力の核を置く製品である。「IBM FlashCore テクノロジー」はコスト優位性のあるMLCフラッシュを使いながらも独自のハードウェア設計によって高速性を実現しているほか、暗号化、データ冗長性、無停止のファームウェア更新など歴史の長いエンタープライズストレージのテクノロジーが埋め込まれている。

FIGURE 3

IBM オールフラッシュストレージソリューション



Source: IBM, June 2016

「IBM FlashSystem V9000」は、そのベースに「IBM SAN ボリューム・コントローラー (SVC)」として長い市場実績を持つ仮想化ソフトウェア「IBM Spectrum Virtualize」を組み合わせた柔軟性と高度なデータサービス機能を両立させた製品である。SVCが採用されているシステム全体を高速化する手段として最適とも言えるが、Spectrum Virtualizeの機能を生かしてデータレイクの中心的な役割

を担うこともできる。「IBM FlashSystem A9000/A9000R」は2016年5月に提供開始された製品で、スケールアウト型ストレージとして独自の分散 GRID アーキテクチャを持つ「XIV」をソフトウェア化した「IBM Spectrum Accelerate」が組み合わされているため、運用管理が容易である。

このように IBM はフラッシュストレージのラインナップの中に IBM Spectrum Storage の Software-Defined Storage ならではの柔軟性を取り入れ、多様なニーズに応えている。データレイクにフラッシュストレージを取り込む場合でも、他のストレージとの連携を損なうことなく、分析システムの高速化にフラッシュ技術のアドバンテージを適用できる仕組みを用意している。

ビッグデータの保存コスト負荷を軽減するテープストレージ

かつてテープストレージはバックアップ用途のストレージとして主に使われていたが、その後、バックアップ用 2 次ストレージとしてのディスクストレージ利用が普及した。一般の認識としてはテープストレージが過去の技術として扱われる風潮もあるが、実はそれは大きな誤解である。IBM の最新テープドライブのデータ転送は、エンタープライズ向け HDD を上回る速度、300Mbps に達している。また、高密度化技術に関しても、IBM は富士フイルムとの共同研究でカートリッジ当たり 220TB の容量を実証済みである（現在の最大容量は 1 巻で 10TB）。さらに利便性も大きく向上している。「IBM Spectrum Archive」はテープ内のデータへのダイレクトアクセスを可能にすると共に、分散ストレージ「IBM Spectrum Scale」との組み合わせではネームスペースを統合し、ポリシーベースで運用可能な階層ストレージを形成できる。

これにより、システムレベルでもテープストレージを巡る新たな動きが見られる。フラッシュメモリーの大容量化、価格低下の流れを受けて、1 次ストレージをすべてフラッシュストレージに置き換える「オールフラッシュデータセンター」が現実検討されており、その際に組み合わせられて使用される 2 次ストレージには HDD ではなく、テープが適しているとの考え方がされ始めた。実際、米国の大手サービスプロバイダー企業ではテープストレージの利用が増えた。一般企業の分析データのライフサイクルを考えた場合も、分析終了後、データをテープに格納するのは、以降のアクセス頻度と保存コスト、特に電力コストを考慮した場合、合理的な選択と言える。分析後のデータ保存ニーズはユーザーごとに異なるが、ディスクによるアクティブアーカイブ、テープによる経済的なアーカイブの柔軟な組み合わせはペタバイトを超えるデータを持つユーザーには価値あるソリューションであろう。

IBM の市場機会と課題

IBM はクラウド、コグニティブなどエンタープライズソリューションの成長分野でも市場に先駆けて製品、サービスの充実を図っており、顧客課題の解決策を数多く持っている。ユーザーにリーチする接点となる商材、特にデータマネジメント、データ分析の関連ソリューションが豊富なため、ストレージソリューションを提案する機会も多い。国内ではまだデータレイクの構築が本格化していないが、いずれ確実にその需要が高まるであろう。個々のユーザーが多様なデータレイク要件を持っているとしても、IBM のストレージのポートフォリオは幅広く、柔軟な提案が可能のため、その都度、顧客ニーズに合ったデータレイクを構築する機会があるとみられる。社内にコンサルタントを持つ IBM は、ストレージの要件を単に速度と容量として把握するのではなく、顧客やストレージインフラの改革によって達成したい目標やデータの性質を理解した上でのソリューションを提案できることが強みになるであろう。

IBM は先進ソリューションを他社に先駆けて取り揃えることで、社会変化の先端に行くユーザー層に対し強みを発揮する。言い換えると、価格競争にあるコモディティ段階の市場を得意とするのではなく、市場開拓者として先行者利益を追求するビジネスモデルと言える。国内では一部の先進ユーザーがデータ活用でグローバル企業に負けないスピードで変革を行う一方、多数のユーザーはその動きに遅れ、格差が開いている。データ活用についても例外ではなく、活用度が低いゆえにデータ保有量が少ないユーザーが多い傾向にある。このような国内ストレージビジネスで IBM が成功するには、積極的な啓発活動やプロモーションによってユーザー全体のデータ活用度を向上させることが必須であろう。特に、多数派ユーザーにリーチするには、販売パートナーに対してもデータ活用の提案を促し、ソリューション提供力を高めることが重要である。

IDC 社 概要

International Data Corporation (IDC) は、IT および通信分野に関する調査・分析、アドバイザーサービス、イベントを提供するグローバル企業です。50年にわたり、IDCは、世界中の企業経営者、IT 専門家、機関投資家に、テクノロジー導入や経営戦略策定などの意思決定を行う上で不可欠な、客観的な情報やコンサルティングを提供してきました。

現在、110 か国以上を対象として、1,100 人を超えるアナリストが、世界規模、地域別、国別での市場動向の調査・分析および市場予測を行っています。

IDC は世界をリードするテクノロジーメディア（出版）、調査会社、イベントを擁する IDG（インターナショナル・データ・グループ）の系列会社です。

IDC Japan

IDC Japan (株) 〒 102-0073 東京都千代田区九段北 1-13-5
81.3.3556.4760
Twitter: @IDC
idc-community.com
www.idc.com

Copyright Notice

本レポートは、IDC の製品として提供されています。本レポートおよびサービスの詳細は、IDC Japan 株式会社セールス（Tel : 03-3556-4761、jp-sales@idcjapan.co.jp）までお問い合わせ下さい。また、本書に掲載される「Source: IDC Japan」および「Source: IDC」と出典の明示された Figure や Table の著作権は IDC が留保します。

Copyright 2016 IDC Japan 無断複製を禁じます。

