

White Paper

DXの支柱となるデータパイプラインをエンドツーエンドで強化する IBM ストレージソリューション

Sponsored by: IBM

鈴木 康介
April 2020

調査概要

近年の ICT 技術の進化によって、デジタルデータを利用することのハードルは大幅に下がった。その様子は「データ活用の民主化」と称されることもあり、かつてのように特権／資金を持つ限られた大企業のみしかその恩恵にあずかれないという時代ではなくなった。ただし、データ活用が電気や水道の利用並みに簡単ということではなく、その成功度は、IT リテラシー、利用アイデアだけでなく、データ収集、整理、処理、業務適用という流れを円滑化、効率化する仕組み、いわゆる「データパイプライン」をどのように整備するかにかかっている。本調査レポートでは、データ活用を促進するテクノロジーの進化が与えるビジネスへの影響と、デジタルトランスフォーメーション（DX）の継続を支える「データパイプライン」整備の必要性を考察すると共に、IBM のストレージソリューションがユーザー企業のデータパイプラインの強化にどのように寄与するかについて考察する。

デジタル化時代の企業競争力の中核をなすデータパイプライン

テクノロジーがもたらすビジネススピードの向上とそれに伴う戦略の変化

デジタル化社会となった現代では、数ある企業競争力要因の中で、特に「スピード」を重視する傾向にある。その理由はどのようなものであろうか。

ビジネススピード向上をもたらす数々のテクノロジー進化

ICT 技術の発達は、多方面からビジネススピードを向上させている。たとえば、アジャイル開発を支えるクラウド、コンテナ、マイクロサービスなどの技術は製品／サービスの市場投入スピードを、Hadoop などの分散処理技術は分析のスピードを、GPU コンピューティングやフラッシュストレージなどハードウェアテクノロジーはデータ処理スピードを、それぞれ大幅に改善したと言えるであろう。スピード向上の流れは今後も続く見込みである。データ分析の分野では、現在、分析よりも時間がかかっているデータ準備（Data Preparation）の自動化を進める開発競争が行われている。コンテナ関連技術の発達もアプリケーションの開発／展開を加速するであろうし、RPA（Robotic Process Automation）はデスクトップ業務を自動化することで業務プロセスの迅速化に貢献している。IT 運用の分野でも Infrastructure as Code の取り組みが進んでおり、プロビジョニングの時間を大幅に短縮した事例が増えている。ストレージの分野では、現在、ハードディスクからフラッシュメモリーへの交代の半ばでありながら、早くも次世代メディアとして SCM（ストレージクラスメモリー）が利用可能になっている。このようなテクノロジーの絶えざる進化によるビジネススピードの向上は、競争戦略にも影響を与えている。

ビジネススピードを重視した戦略のメリット

企業では、過去には、市場開拓やビジネスイノベーションの成功確率を高めるため、市場を精密に分析し、ニーズを把握しようと努力していた。しかし、現在、一部の企業では、開発のスピードを上げ、早期の市場投入を優先するようになってきている。その背景には、データ収集から分析ま

での時間を短縮し、ビジネスの軌道修正が速ければ、成功の確率がより上げられるという考え方がある。市場参入がたとえ1番でなくても、ビジネス状況をリアルタイム（もしくはそれに近い形）で可視化し、改善サイクルを速く回し、顧客の信頼を獲得できれば、次第に市場で優位に立つことができるという発想である。

市場や顧客など競争環境の変化が激しい時代では、このような開発とデータ分析の速さを武器としたアプローチは合理的であり、かつ、有利と考えられる。時間のかかる市場調査を行っていても、調査期間に状況が変化する可能性も排除できない。それよりも、実際に事業を展開し始めてみてからデータを収集すること、たとえばSNSからスクレイピングした製品の評判を判断したり、店舗内のカメラで来客の行動をモニタリングしてデータを取得することは、シミュレーションを繰り返したり、モニターに対するヒヤリングを行うよりも有用と考えられている。ただし、このような競争戦略を取る場合は、データ分析の速さと正確性で他社を上回る必要がある。

データ活用を持続的競争優位の源泉とするためのデータパイプラインという概念

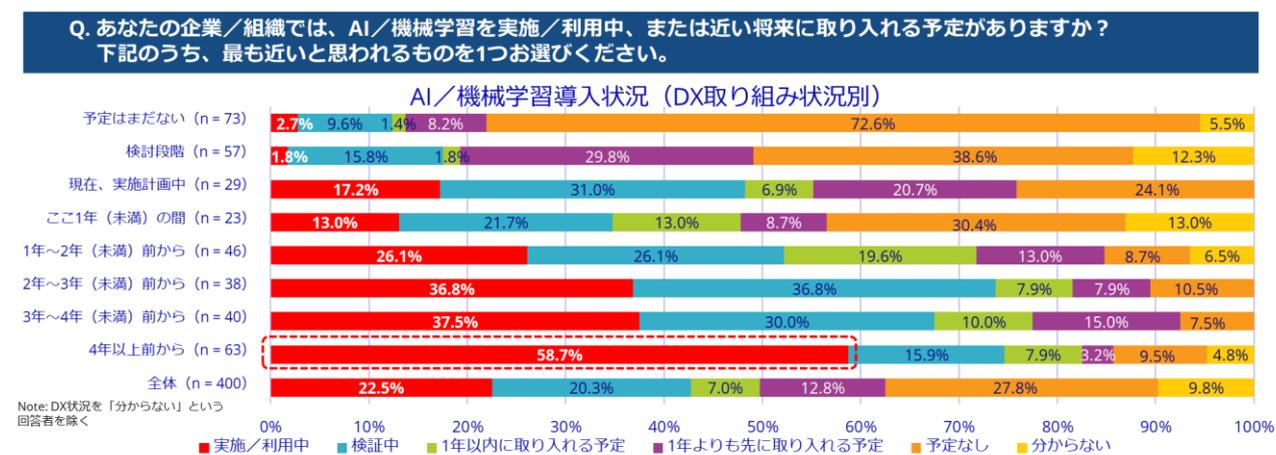
イノベーションが一過性のものであれば持続的な競争力につながりにくいと同様、データ活用についてもその分析作業が一時期のことで済むわけではない。収集から分析までの繰り返しの効率、すなわち、各段階での作業（データ整理、加工、転送など）がいかに自動化されて素早いフローになっているかが問われる。当然ながら、データ活用の仕組みには、継続的な精度向上、市場環境の変化やテクノロジーの進化に対応するための柔軟性／拡張性も必要になる。このようなデータ収集から分析までのプロセスを円滑、迅速につなぎ、有用なデータをその活用現場に供給する仕組みを、エネルギー資源の輸送インフラに例えて、データパイプラインと称されることが増えてきた。このデータパイプラインは流量が増えると運用の難易度が増す。特に現在、取り組む企業が増えている機械学習には大量の教師データが必要になり、データパイプラインをいかに構築するかが重要な鍵となる。

データ活用のステージをさらに引き上げる AI の普及

前述の分散処理技術の発展は、大量データの分析を可能にし、データ活用の有用性を格段に向上させ、ビッグデータのブームを生んだ。そして、それに続くように、AI／機械学習の進化がデータの利用価値をさらに引き上げた。データ活用がAIまで進むと、予測精度を向上させるだけでなく、ビジネスプロセスの自動化、製品／サービスの品質コントロール、顧客との信頼関係の構築、自然言語処理や画像認識を用いて人が行う業務を代替させるなど応用範囲は大きく広がる。

FIGURE 1

DXの進捗と共にAIの利用が増える相関傾向



Note: 『2020年国内マルチクラウドのデータ運用実態と課題（IDC #JPJ45137920、2020年3月発行）』からの引用

Source: IDC Japan, March 2020

DXを進めたユーザーは、AIの利用に行き着くことが多い傾向が IDC の調査でも判明している (Figure 1)。グラフの左縦軸は「DXに取り組み始めた時期」、横軸は「AI/機械学習の利用状況/導入予定」を聞いた結果の集計である。「4年以上前からDXに取り組んでいる」とした回答者の6割近くがAIを利用中と答えている。現代社会で競争を繰り広げる企業にとってDXへの取り組みが必須となっているが、その中でもAIは欠かせないツールとなってきた傾向を示している。

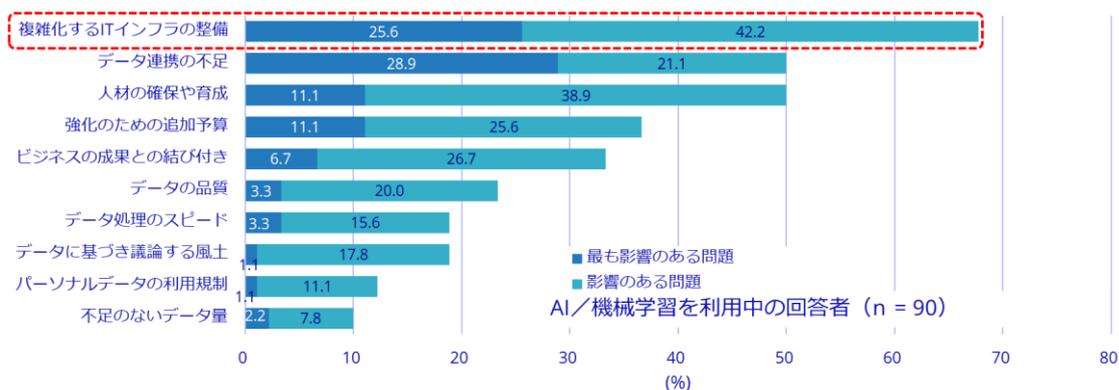
AI 利用によるデータパイプラインの要件変化

AIは有用性が高いツールであるが、扱うデータ量の大規模化には対策が必要である。市場では「AI 利用にはそれに応じた IT インフラを準備すべきである」との発想から「AI Infrastructure」という言葉が使われ始めている。前述の GPU コンピューティングもその要素技術の一つである。IDC の調査でも、AI をすでに利用しているユーザーからの回答集計として、「分析能力の強化のために何が影響を持つか」という質問に対し、「データ品質」や「人材」よりも上位の回答率で「IT インフラの整備」の影響を挙げているのは興味深い結果である (Figure 2)。

FIGURE 2

AI を利用中のユーザーが大きな影響を持つと考える課題は IT インフラ

Q. 現状よりも、企業/組織の分析能力を強化するために問題、障害となるのは何だと思われますか？
最も影響のある問題を1つと、それ以外の問題をすべてお選びください。



Notes:

- 選択肢「分からない」と「特になし」は表示していない
- 『2020年 国内マルチクラウドのデータ運用実態と課題 (IDC #JPJ45137920、2020年3月発行)』

Source: IDC Japan, March 2020

機械学習では、画像や音声などの非構造化データを扱うことも多く、その場合、データ量が桁違いに大きくなるだけでなく、ファイル名での識別が難しくなるといった課題も出てくる。また、モデル化のためのデータ学習は長期間に及ぶことが少なくない。これはストレージのランダム I/O 性能によって大きく左右される部分のため、戦略的なストレージ投資を検討する価値がある。仮に、四半期に1度のバッチジョブが5時間から1時間に短縮されたとして、ビジネス競争力にどれだけの影響を与えるかは測りにくい。機械学習の期間が6か月から半分の3か月に短縮されるのなら、その3か月はビジネスの成否を分ける可能性があるからである。また、AIを業務適用する場合に制御にリアルタイム性が要求されるならば、それに応じたレイテンシーを実現する必要があるし、実務のミッションクリティカル性に応じた可用性も確保しなければならない。

クラウドとオンプレミスをどのように組み合わせるかを検討課題となる。クラウドでAI利用をスモールスタートするのは合理的であるが、機械学習を長期的、かつ I/O インテンシブに行う場合、利用コストが肥大化する可能性がある。実際、米国市場では「Cloud

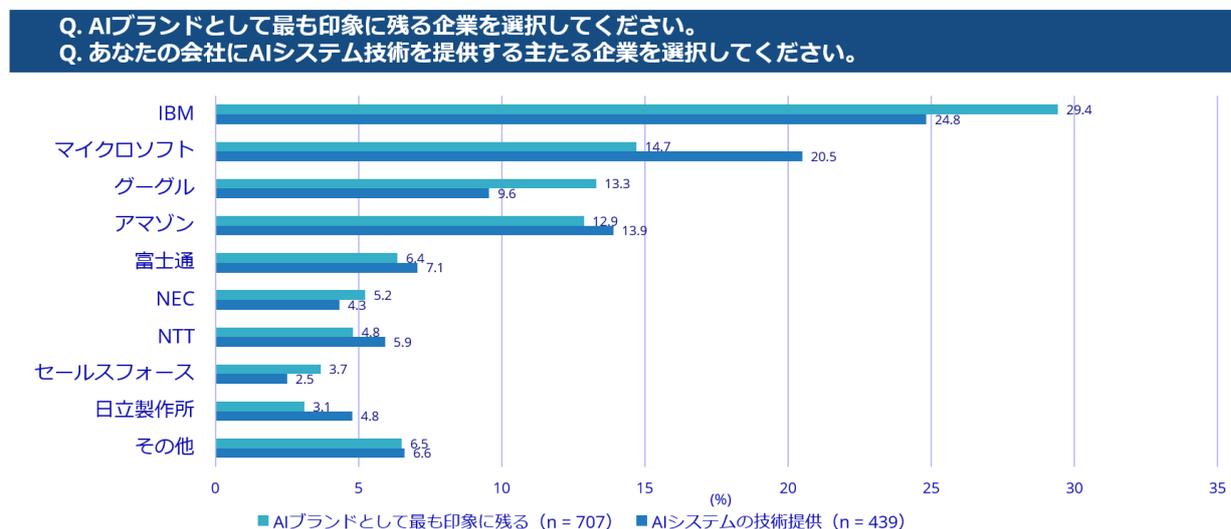
Repatriation（repatriationは帰還、本国送還などの意味）」という言葉が2019年に急速に多く聞かれるようになった。センサー／カメラ技術の発達と低価格化、ドローンの普及などを背景に、工場、店舗などデータセンター外でのデータ発生機会は増えることが見込まれるため、すべてのデータ処理がクラウドで完結することにはならず、エッジコンピューティング対応を含めたデータパイプラインの設計は複雑なものになるであろう。そのアーキテクチャ検討の際、高速性と経済性の両立は眼前の課題として明らかであるが、将来のテクノロジー進化、扱うデータ種類の増加などを考慮して柔軟性、拡張性にも十分な配慮を行うべきであろう。

AI 導入実績／認知度で国内市場をリードする IBM

国内でも導入が進む AI システムであるが、その中でも市場をリードしているのが IBM である。2019年4月に IDC が発行した調査レポートでは、技術提供の実績、AI のブランド認知で共に首位の回答率となっている（Figure 3）。IBM は AI のアーリーアダプターをサポートする中で、AI 導入のビジネス効果を長期に渡って発揮するために必要となるデータパイプラインを支えるストレージ製品のポートフォリオの充実に注力している。次セクションでは、そのビジネス価値を分析する。

FIGURE 3

AI システムのブランド認知と導入実績



Notes:

- 「AI システムを把握している」とした回答者のみを集計している。AI ブランドとして最も印象に残る企業については、回答率が3%以上の企業とし、3%未満の企業はその他に合算した
- 『2019年 国内 AI システム RPA ソフトウェア市場 企業ユーザー調査 (IDC #JPJ44005119、2019年4月発行)』からの引用

Source: IDC Japan, April 2019

データ利用の高速化／大規模化の課題を解決する IBM ストレージ製品

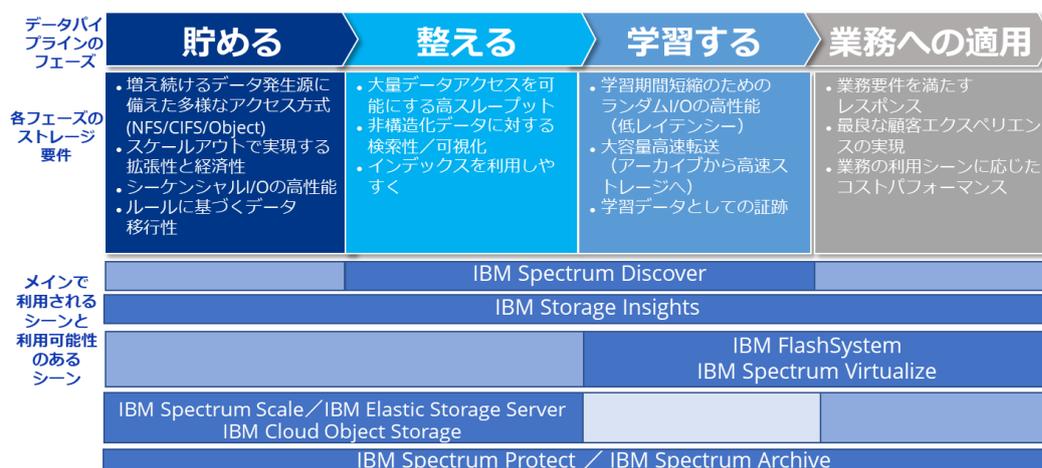
データパイプラインの各フェーズを支える IBM ソリューション

データパイプラインを、データ収集、分析に備えたデータ整理（準備）、データ学習、業務適用の4つのフェーズに分けて考えた場合、それぞれのフェーズでは、ストレージに対する要件が大きく異なるため、1種類のストレージアーキテクチャですべてを満たそうとすると無理が生じる。その点、IBM ストレージソリューションのポートフォリオは、IBM Spectrum Virtualize の先進機能を備えたブロックストレージ IBM FlashSystem の高速性、ファイルストレージ IBM Spectrum Scale / オブジェクトストレージ IBM Cloud Object Storage の拡張性、経済性といったストレージアーキテ

クチャの異なる分野でリーディングエッジに位置する製品を揃えている。その上で、それらを IBM Storage Insights で管理統合できることに特徴がある。加えて、IBM の場合は、メタデータ管理ソリューション IBM Spectrum Discover、データ保護ソリューション IBM Spectrum Protect、データアーカイビングソリューション IBM Spectrum Archive などデータパイプラインの効率性と運用安定性を高めるソフトウェアソリューションを揃えていることも強みとなっている。

FIGURE 4

IBM のソリューションポートフォリオによるデータパイプラインのサポート



Note: IBM 提供の資料に基づき作成

Source: IBM and IDC Japan, April 2020

近年はテクノロジー進化のペースが速く、競争力維持/向上には取り扱うデータ種類も増えがちで、データ管理の複雑化は、運用のリスク増加や効率低下の要因になる。頻りに要求される拡張や変更への対応スピード、データパイプラインのフェーズ間の接続性といった要求を考慮すると、ストレージインフラの標準化にはメリットが多い。Figure 4 では、IBM が持つストレージ製品を中心としたポートフォリオをデータパイプラインにマッピングしたチャートである。図の濃い色の部分は、その製品が最も活躍するフェーズ、淡い色は利用の可能性のあるフェーズを示している。

Table 1 は、Figure 4 に示した製品がどのような特徴によって、各データパイプラインのフェーズにおける課題に対応するのかを示している。IBM Spectrum Virtualize は、同社のストレージ製品だけでなく、他社を含めた約 500 機種 of ストレージ仮想化に対応している。さらに、クラウド対応版 IBM Spectrum Virtualize for Public Cloud では、AI ベースでストレージの階層化を行う Easy Tier、オンプレミスとクラウド間、および複数クラウド間のリモートコピーをサポートしており、データパイプラインの複雑性を吸収するレイヤーとして重要な役割を果たす。

IBM Spectrum Scale は、Software-Defined Storage としてビッグデータの保存用にテープストレージと併用されることもあり、国内実績も数多い。IBM Spectrum Scale 搭載のアプライアンス製品は、IBM Elastic Storage Server (ESS) として販売されている。IBM Cloud Object Storage は、HTTPS (RESTful API) によって、センサーやカメラなど多数のエッジ (データ発生源) からの同時アクセスが可能のため、IoT (Internet of Things) アプリケーションやコンテナベースの並列稼働アプリケーションとの親和性が高い。

IBM Spectrum Discover は、データパイプラインのフェーズごとの検索課題に対応するソリューションで、データ整理のフェーズではシステムメタデータの自動取り込み、データ学習のフェーズでは分析結果の証跡をカスタムメタデータで付与する機能が有用である。メタデータ管理には、分析能率の向上以外にもコンプライアンス統制へのメリットがある。IBM Storage Insights はストレ

ージ環境の可視性、可用性、利用の経済性を高める管理機能を提供するクラウドサービスである。AI ベースの予測分析によって性能や容量などの閾値を監視し、障害を未然に防ぐと共に、サポートを自動化し、迅速な問題解決を行うことでデータパイプラインの運用信頼性を高める。

TABLE 1

製品特徴とデータパイプラインの要件へのマッピング

| 製品 | IBM FlashSystem | IBM Spectrum Virtualize | IBM Spectrum Scale/ESS |
|------|--|---|--|
| 適用 | 業務適用/学習 | 業務適用/学習 | 収集/整理 (/学習/業務適用) |
| 製品特徴 | 高速性: NVMe、IBM FCM/SCM ドライブ対応、ハイブリッド、マルチクラウド対応 | 同期/非同期コピーサービス、高可用性構成、ストレージ階層化と自動データ移動、重複排除 | 高速性と経済性、多様なメディアの自動階層管理 (テープを含む)、NVMe 対応、SDS とアプライアンスの 2 つの提供形態 |
| 解決課題 | リアルタイム制御、学習期間の短縮 | 運用標準化、経済性向上、ハイブリッド/マルチクラウド対応 | 経済性向上、分析後のアーカイブ |
| 製品 | IBM Cloud Object Storage | IBM Spectrum Discover | IBM Storage Insights |
| 適用 | 収集/整理 (/学習/業務適用) | 整理/学習/業務適用 (/収集) | 収集/整理/学習/業務適用 |
| 製品特徴 | 大規模拡張性、Notification 機能、Range GET 機能 (部分抜き出し)、Aspera の高速データ転送機能、IBM Cloud SQL Query による即時照会 | メタデータを統合的に管理 Action Agent API によるアプリケーション連携 他社のファイル/オブジェクトストレージ対応 | AI ベースのストレージ管理クラウドサービス、IBM ストレージ以外の機種にも対応する一元管理、グローバルで収集したデータを活用し洞察を提供 |
| 解決課題 | 大容量保存の経済性、Edge/クラウドとの親和性 | 検索課題への対応、分析結果の証跡、コンプライアンス統制、データガバナンス強化 | 運用可視化、利用効率向上、予防保守/保守対応自動化、ベストプラクティスの推奨 |

Note: IBM 提供の資料に基づき作成

Source: IBM and IDC Japan, April 2020

ビジネス高速化の要件とデータ運用の安定を両立する IBM FlashSystem

ストレージの高速性は、データパイプラインにおける学習フェーズの期間の長短を左右する重要な要素であり、IBM のポートフォリオでは IBM FlashSystem (以下、FlashSystem) がその役割を担っている。IBM は、2020 年 2 月に FlashSystem の新たなラインナップとして、IBM FlashSystem 9200R、IBM FlashSystem 9200 (FlashSystem 9100 の後継)、IBM FlashSystem 7200 (FlashSystem 900、Storwise V7000 の後継)、IBM FlashSystem 5000 と 5100 (Storwise V5000E、V5100 からの名称変更、FlashSystem ブランドへの統一) の 5 機種を発表した (Figure 5)。新製品の開発思想としては、高速性だけでなく、ハイブリッド、マルチクラウド環境での利用を前提に、複雑性の排除とコスト削減が目指されている。

高速性を実現するテクノロジー

FlashSystem で採用されている同社の独自設計である IBM FlashCore モジュール (FCM) は高速性を実現すると共に、2U のスペースで最大 4PB の容量を提供する優れた容量密度を持っている。また、I/O パフォーマンスに影響を与えないハードウェアによるデータ圧縮と FIPS 140-2 認証の暗号化機能が組み込まれている。また、FlashSystem ではインテルおよびサムソンの SCM SSD (Solid State Drive) に対応しており、キャッシュが有効に使いにくいワークロードに対しても低いレイテンシーを実現する。また、サーバーと外付型ストレージシステムの間をエンドツーエンドで接続する NVMe over Fabrics は、現時点ではハイエンドの接続インターフェースと市場で位置付けられるが、FlashSystem では一部のエントリー機種から対応しており、IBM がストレージシステム間のデータ転送パフォーマンスを重視する姿勢がうかがわれる。

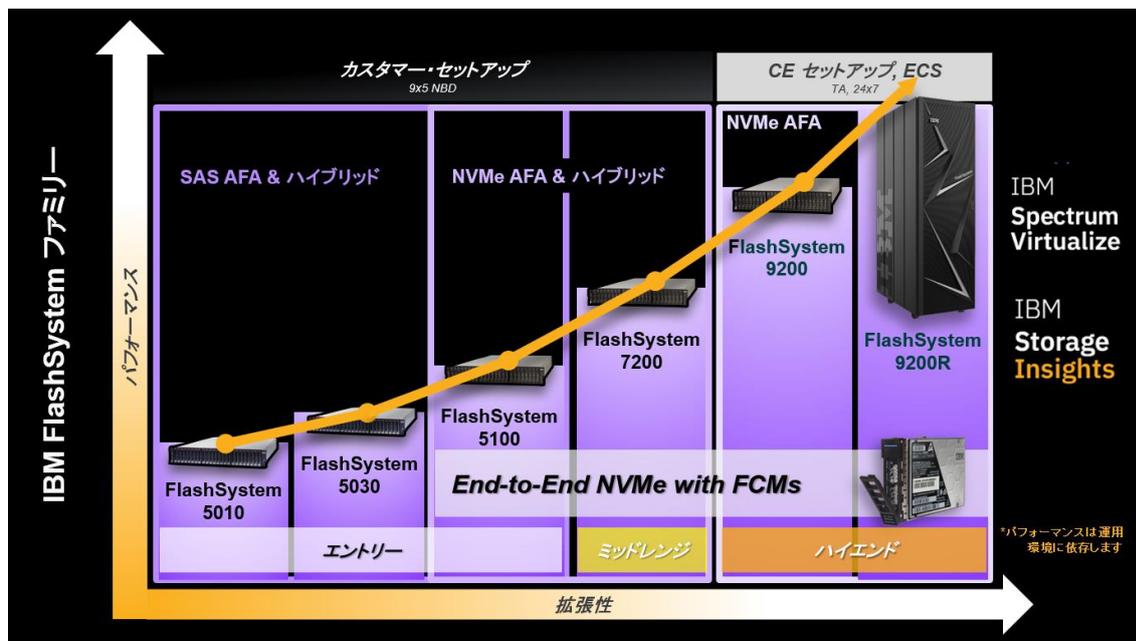
運用性を高める多種のソフトウェア機能

IBM FlashSystem は IBM Spectrum Virtualize と統合された形で利用できるため、多種のデータサービス機能（AI による自動階層化、データ圧縮、重複データ排除、暗号化、レプリケーションなど）によって、高い可用性と運用の容易性を享受できる。たとえば、可用性を担う HyperSwap 機能では、ストレージ側で自動的に障害を検知して切り替えることで、サーバーからは 1 台が無停止で稼働しているように見せられる。ハイパーバイザー側でのフェイルオーバー制御は必要ないために、サーバーの CPU 負荷が軽くなる。この設定はボリューム単位に行えるため、業務のサービスレベルに応じた運用が可能になる。

また、AI を活用したクラウドベースの管理ツール IBM Storage Insights を利用することで、世界中のベストプラクティスを機械学習し、保守に役立つ洞察を得ることができる。その管理対象は IBM のストレージ製品だけでなく、一部の他社ストレージ製品、そして IBM Spectrum Virtualize/IBM Spectrum Virtualize for Public Cloud を通じて他社ストレージ製品に加えてクラウドストレージにまで及んでおり、過去のストレージ資産を有効活用するだけでなく、データパイプラインの構築の要素としてクラウドを用いた場合にも一元的な運用が維持でき、複雑性を排除できる。

FIGURE 5

IBM FlashSystem ファミリー



Source: IBM, April 2020

IBM のビジネス機会と課題

前述のユーザー調査からも分かる通り、IBM は国内でも AI 導入実績を多く持つことから、その有効活用のノウハウを豊富に蓄積していると考えられる。ストレージ製品を始めとするポートフォリオも、データパイプラインをシームレスにつなぐためのソリューションが網羅的に揃っているため、個々の顧客の課題が各フェーズで異なるとしても、そのいずれのケースからも包括的なソリューションの提供につなげる提案、サポートが可能であると考えられ、今後のビジネス機会は広がると見込める。IBM の課題は、AI 利用をクラウドでスモールスタートしたものの、規模拡大時のコスト負担が問題となるユーザーをいかに支援するかであろう。国内市場における AI/機械学習の幅広い実績に基づき、事例内容の積極的な公開やベストプラクティスの提示などを通じ、早期段階から大規模化した場合の経済的な効率性を検討することをユーザーに対して啓発すべきであると IDC では考えている。

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、アドバイザリーサービス、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 2020 IDC Japan 無断複製を禁じます。

