

# ハードウェアでアナリティクスは加速する

## — 成功するアナリティクス基盤の5つのハードウェア要件 —



日本アイ・ビー・エム株式会社  
システム製品事業 テクニカル・セールス  
アドバンスド・テクノロジー・センター  
システムズ&テクノロジー・エバンジェリスト

石井 旬



日本アイ・ビー・エム株式会社  
システム製品事業 テクニカル・セールス  
アドバンスド・テクノロジー・センター  
IT アーキテクト

町田 武夫

【プロフィール】

1989年日本IBM入社。2006年よりシステム製品事業で先進テクノロジー関連のアーキテクトを担当。2009年以降アナリティクス分野の担当になり、多くのお客様案件や研究活動、講演などの外部活動に従事。2012年からアナリティクス分野のシステムズ&テクノロジー・エバンジェリストも務める。

【プロフィール】

入社以来、メインフレームから分散システムといった多岐にわたるプラットフォーム上でのオートノミック・コンピューティング、SOA、クラウドといった先進ソリューションの技術支援に従事。現在はアナリティクス分野の担当として、アナリティクスのシステム基盤に関する提案・構築案件を支援している。

### 1. 分析 (アナリティクス) の動向

「ビッグデータ」というキーワードが毎日飛び交い、世界中のデータ量の急激な増加が目撃されつつある今、企業におけるデータ分析、およびデータ活用への関心も著しく高まっています。

BI (ビジネス・インテリジェンス) や DWH (データ・ウェアハウス) という言葉に代表されるようなデータ分析は、決して新しいものではありません。しかしながら、近年のデータ量の増大や、要素技術、分析ツールの進化などにより、分析 (アナリティクス) の中身も急速に変わりつつあります。

データの観点では、ビッグデータの特

性である4つのV (図1) に対応した分析能力と基盤が必要となります。2020年には全世界で35ゼタバイトになるとも言われているデータ量の急激な増加 (Volume) [1]、データ発生スピードの高速化 (Velocity) 以外にも、全体の80%以上を占めると言われる非構造データとこれまでの構造化データといった多種多様なデータの混合 (Variety)、加えて2015年までに全データでの80%を占めると言われる不確かな情報 [2] を認識した上で情報の正確さ (Veracity) といった、データ自体の特性を意識することが重要となるのです。

これらの、量 (Volume) をはじめとする

データ特性は、大量データを効率良く格納するストレージ、大量データを処理するために必要となる高速な並行処理が可能な演算装置、テクノロジーの進化や破壊的技術の登場による超高速処理機器、高速なネットワークなど、ハードウェアのより先進的な活用が不可欠になります。

これに加え、分析の手法も急速に変化してきています。よりリアルタイムに近い処理と分析、回帰分析や決定木分析といった統計手法の予測への活用、および多種多様な大量データから知見を導き出すモデルやアルゴリズムの適用などにより、レポート中心の従来のBIから、より高度な分析・最適化へと分析手法が変わ

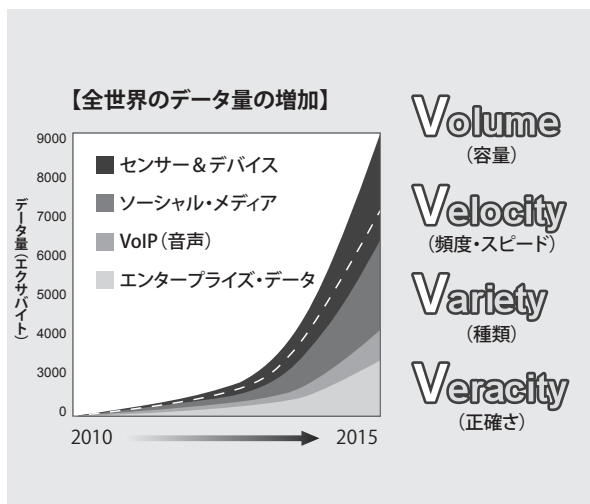


図1. ビッグデータの4つの特性

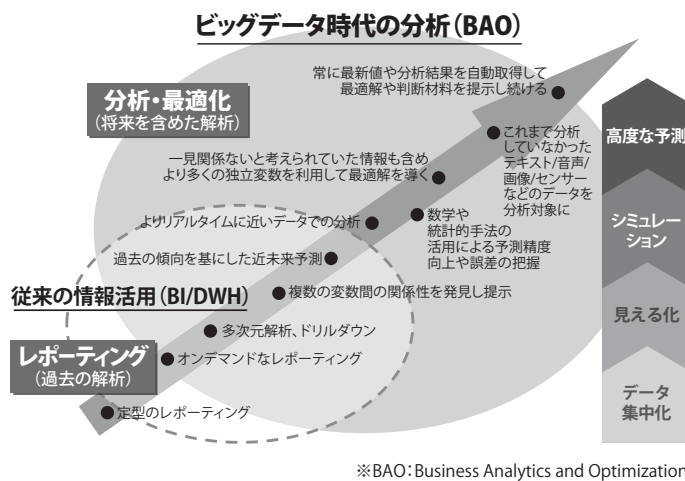


図2. 分析 (アナリティクス) の移り変わり

※BAO: Business Analytics and Optimization

りつつあります (図 2)。

この分析手法の変化とビジネス価値の増大により、アナリティクスは、ビジネス上の価値創造や企業の業績に大きく関与するようになってきています。また、これらの分析を多くの事業や業務プロセスにおいて活用することが、分析を多面的に行う企業を勝ち組にしている要因 [3] とも言われます。

## 2. アナリティクスに求められるハードウェア要件

このような分析対象のデータと分析手法の変遷、多様化に伴い、この変化を取り込んだアナリティクス・システムの最適化は、ハードウェア・テクノロジーに大きく依存します。

最近、筆者らが関わったいくつかのお客様システムの提案・構築経験から、新しい分析を支えるために重視すべきハードウェア要件は、図 3 に示す 5 つにまとめられることが分かってきました。

以降では、各要件に関し、より詳細な必要要件とそれを充足する技術・製品例を挙げて説明します。

### ①性能：圧倒的な処理スピードをもたらすハードウェア技術

データの大量化に伴い、それら大量データの格納と高速処理が必要となります。当然、システムも大規模となります。

従来、システムの大規模化は、サーバー

などの単体のシステム・コンポーネントの処理能力を上げる「スケール・アップ」と、同種のシステム・コンポーネントを並列的に複数配置することで全体の処理能力を上げる「スケール・アウト」が一般的でした。しかしながら、これら 2 つの手法によるシステム能力の向上には限界があるため、急激に処理要求が大規模化するに従い、新たにシステム密度を最大化し各処理をより低遅延 (Low latency) で実行することで、システム規模を抑えつつ大量処理を行う手法も併せて重要になってきました [4]。これらを基本とした技術の進化や、処理の発想の転換により、処理性能の側面から従来の延長とは異なる飛躍的な性能効果をもたらす技術の商用化も始まっています。

これを実現する製品の例として、IBM FlashSystem (以下、FlashSystem) が挙げられます。FlashSystem は、ストレージ技術の中で注目されつつあるフラッシュ・ストレージの 1 つです。HDD (Hard Disk Drive) に比べて高速でコンパクト、また衝撃に強いフラッシュ・メモリーを記憶装置として使ったストレージです。同じようにフラッシュ・メモリーを使っている SSD (Solid State Drive) のように HDD の代替として使うのではなく、メモリーそのものとして使うことにより、数十テラバイトのデータ記憶領域を 1U (19 インチラックの 1 ユニットのサイズ。厚さ 1.75 インチ=約 44.5mm) という小スペースで提供しつつ、マイクロ秒単位という非常に高速なデータアクセス (HDD はミリ秒単位) を実現し、

結果的に秒あたり数十万 I/O 回数という驚異的なデータ処理を実現します (図 4)。

また、プロセッサの分野でも、アナリティクスやビッグデータにより良い性能をもたらすハードウェア技術があります。IBM POWER プロセッサは従来からある技術ですが、ビッグデータを処理するための並行処理を、より多くのスレッドで実行できるという観点で、ビッグデータ時代の性能を向上する演算機構の代表格と言えるでしょう。多くのアナリティクス処理のベンチマークでは、通常の x86 プロセッサより優れた結果を出しています。

### ②最適化：ワークロードに最適化された装置

ひと口にデータ処理と言っても、さまざまな種類のデータ・アクセス・パターンがあります。e- コマースの決済システムや銀行の勘定系システムに代表される OLTP (Online Transaction Processing) システムのように、一件一件の処理の取り扱うデータは小さいが、大量のユーザーからの処理要求を短い応答時間で実行が求められるシステムがあります。

一方で、OLAP (Online Analytical Processing) システムやデータ・マイニング・システムのように、ユーザーからの処理要求件数はあまり多くないが、一件一件の処理の取り扱うデータが大量なシステムもあります。ディスクから見ても、ランダムなアクセスと順次アクセスでは、最適なディスクの種類や配置が異なってきます。

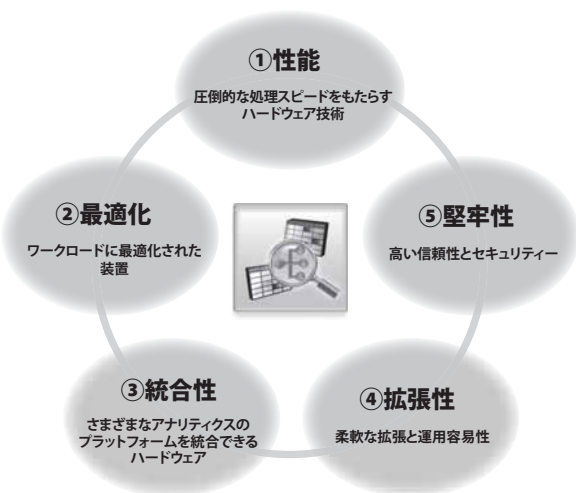


図 3. ハードウェアに求められる 5 つの要件

### 圧倒的なパフォーマンス

**2,200万 IOPS/1ラック**

ビジネスにおけるパフォーマンス、スケール、洞察力を向上させ、サービスと製品の市場への投入時間を短縮



### 高い集積率 Macro Efficiency

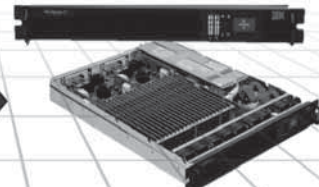
**1ペタバイト/1ラック**

ハードウェアとソフトウェアの融合による効率的運用と迅速な展開

### 超高速な応答時間 MicroLatency

**100マイクロ秒**

重要なアプリケーションの応答時間をマイクロ秒に短縮



### 高信頼性 Enterprise Reliability

**最高30倍の耐久性** ※対MLC\*比

エンタープライズクラス・フラッシュと特許取得データ保護技術による高信頼性

\*MLC (Multiple Level Cell) …… NANDフラッシュ・メモリーの記憶方式の一つ

図 4. IBM FlashSystem

特にデータベース・システムにおいては、データ・アクセスの特性に応じた、システム資源（CPU やメモリー、ディスクやネットワーク）の最適なバランスでの設計、およびサイジングやチューニングは複雑で、従来は専門家でなくては実施できないものでした。最近、製品としてあらかじめハードウェアやソフトウェアを導入・構成し、必要な専門家の知見を組み込んだ「データベース・アプライアンス」の活用が増え、従来は専門家でなくては実施できないものでした。最近、製品としてあらかじめハードウェアやソフトウェアを導入・構成し、必要な専門家の知見を組み込んだ「データベース・アプライアンス」の活用が増え、従来は専門家でなくては実施できないものでした。最近、製品としてあらかじめハードウェアやソフトウェアを導入・構成し、必要な専門家の知見を組み込んだ「データベース・アプライアンス」の活用が増え、従来は専門家でなくては実施できないものでした。

例として、データベースのためのアプライアンス・システムである IBM PureData System が挙げられます。ワークロードに応じて最適なシステムが必要であるという設計思想から、ワークロードに最適化された 3 種類の IBM PureData System が提供されています（図 5）。

これらアプライアンス・システムは、単にハードウェアを組み合わせただけでなく、性能向上のためにフラッシュ・ストレージを適所で活用したり、データに関わる汎用的な処理を CPU で行うのではなく、FPGA（Field Programmable Gate Array）と呼ばれる書き換え可能な専用集積回路で行うといった独自の性能向上がハードウェアとして組み込まれているのです。

③統合性：さまざまなアナリティクスの

プラットフォームを統合できるハードウェア

今後、企業は顧客にとっての価値を増大し、競合他社との差別化を実現するために、さまざまなアナリティクスを投入することになるでしょう。これは一つの BI ツールやデータ処理技術では実現できず、複数で多様な分析ツールやシステムを開発・運用することになります。その結果、アナリティクス・システムは、用途の異なるさまざまなシステム・コンポーネントで構成されるでしょう。ETL（Extract/Transform/Load）サーバーや BI サーバー、ファイル転送サーバーなどのシステム連携サーバーやデータマート・サーバーがその代表例です。将来的にはデータ・マイニングやテキスト・マイニングのサーバー、メモリー内で高速処理を行うストリーミング・サーバーなどが必要となるかもしれません。

これらサーバーを個別に構築していたのではサーバーが林立し、システム運用が煩雑なものとなります。各システム資源の効率的な利用は難しいでしょう。この課題を解決するためには、ハードウェアやプラットフォームの差異を吸収し、十分な仮想化基盤上にさまざまなアナリティクス・システムを統合し、運用していく全体最適化のアプローチが推奨されます。

全体最適化のアプローチには、(A) 単一プラットフォームと仮想化機能を活用して分散サーバーを統合する方式と、(B) ハイブリッド化によって、複数プラットフォーム / OS をリソース・プールとして自

律的に管理し、分散ワークロードをプール内で適材適所に配置する方式の 2 つが考えられます。現実的には、既存システムの継承・互換性や、多様化するワークロードへの対応から、(B) のハイブリッド化が有力と考えられます。

ハイブリッド統合を実現するシステムの例として、IBM PureSystems や IBM System z が挙げられます。IBM PureSystems ファミリーの 1 製品である IBM PureFlex System は、UNIX（AIX、Linux）や Windows を仮想化基盤を使って統合的に集約・管理できるシステムです。集約性・拡張性が高く、統合管理による管理負荷の低減も実現できる次世代プライベート・クラウド環境と言えます（図 6）。

④拡張性：柔軟な拡張と運用容易性

分析の高度化とビッグデータ、特に非構造化データの活用、および利用者の増加に伴い、最初は小さく始めたアナリティクス・システムの基盤も、すぐに拡張が必要となってきます。最近のいくつかのシステムでは、その拡張のペースが非常に早くなっているにもかかわらず、初期の設計時に将来的な拡張を見据えていなかったために、拡張が困難になる、あるいは運用が困難になっているケースも多く見受けられます。このような問題に直面しないためにも、設計初期の段階から、十分に拡張可能で、かつ大規模になっても運用可能なシステムとなるような考慮が必要であると言えます。

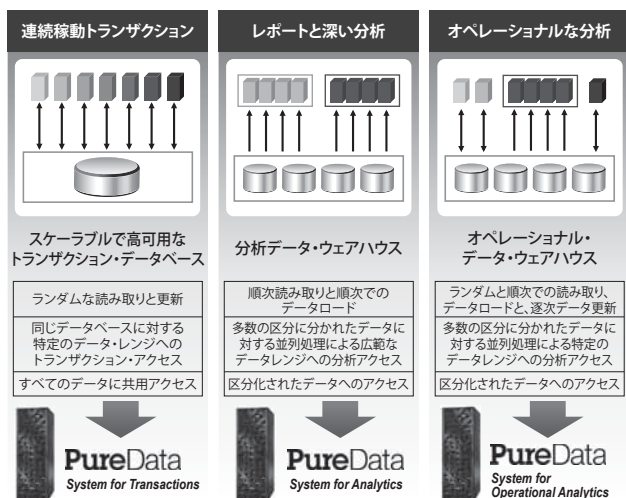


図 5. 異なるワークロードと PureData

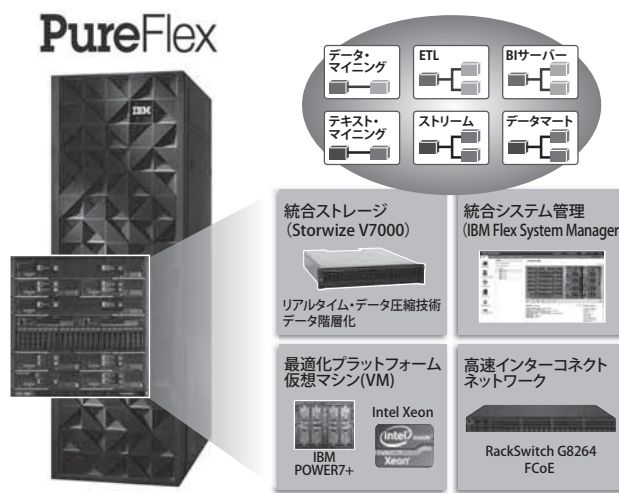


図 6. IBM PureFlex による仮想化統合

拡張性が高く、運用管理も容易なシステムの例として、IBM ストレージ製品である IBM XIV Storage System (以下、XIV) が挙げられます。

XIV は、非階層型のエンタープライズ向けストレージ装置であり、すべてのディスクを単一のプールとして保持します。ストレージ装置に書き込まれたデータは全 HDD に分散され、HDD の使用量、アクセス数が自動で平準化される設計となっています。可用性についても、従来の RAID (Redundant Arrays of Independent Disks) とは異なる IBM 独自技術の分散ミラーリングを実装しており、管理者が意識しないところで、自動冗長化と高速復旧機能を備えています。加えて、ドライブではなく「モジュール」という単位での拡張を行います。これらの技術により、高い運用管理性と拡張性を提供しています。

#### ⑤堅牢性：高い信頼性とセキュリティ

図 2 に示したとおり、これまでの分析は一部のユーザーがバッチ的に分析処理を行うケースが多かったのに対し、今後のアナリティクス・システムは、よりリアルタイム性が高く、またさまざまなユーザーからの処理要求を実行する必要があります。そのため、ひとたびアナリティクス・システムがダウンすれば、その障害による業務への影響が、それまでに比べて広範になります。つまり、信頼性面では、かつてのアナリティクス・システムに比べ、より高い可用性が求められていると言えるでしょう。

加えて、アナリティクス・システムの重要度の高まり、および個人データを含むデータ流出リスクへの対応のために、今後のアナリティクス・システムには、これまでより一層高いレベルのセキュリティが求められます。

IBM の UNIX オペレーティング・システムである AIX を搭載した IBM Power Systems は、99.997% の連続稼働性を誇り、他の UNIX システムの 2 倍以上の信頼性、x86 サーバーに対しては 10 倍以上の信頼性を提供します [5]。また、POWER7 プロセッサと AIX は、オープン・システムで最高レベルのセキュリティ環境である OSPP/EAL4+ に準拠しています。Power Systems は仮想化集約にも優れ、これと合わせて、UNIX システムで最高の信頼性とセキュリティ環境を提供していると言えるでしょう。

また、IBM System z は、99.999% という最高レベルの連続稼働環境を提供します。セキュリティ面でも、OSPP/EAL5 に準拠しており、優れた仮想化機能と合わせて、最高レベルの信頼性とセキュリティを提供します。加えて、IBM System z 上のデータベース・サーバーである DB2 for z/OS 環境では、IBM DB2 Analytics Accelerator という分析ワークロード高速化のためのソリューションが提供されています。これは IBM PureData System for Analytics を DB2 のアクセラレーターとして実装したもので、複雑で重い検索処理を動的に IBM

PureData System for Analytics に割り振ることにより、システム資源の最適活用と高速化を実現しているものです (図 7)。IBM System z も引き続き先進的な分析環境を提供すべく機能拡張を続けており、有力なアナリティクス・プラットフォームの一つと言えるでしょう。

### 3. おわりに ～5つの要件を満たす視野が成功への鍵

分析に関する技術や製品は、日々、急速に進歩しています。技術の革新により、これまで実現できていなかった分析が可能になった、あるいは長時間かかっていた処理が短時間で完了したなどの事例が多く見られます。

しかしながら、例えば超高速といった、ある一つの側面だけに偏った技術や製品の選定、あるいは個別最適のアプローチにより、結果的に拡張性や信頼性が確保できなくなった、運用が回らなくなったといった例も少なからず見受けられます。これからの情報活用の時代には、これらの問題を事前に回避し、企業が価値を生む多様なアナリティクスを効率良く活用していく必要があります。

そのためには技術や製品の正しい理解と鳥瞰的な視野での選択が必要であり、ハードウェアについては、ここで紹介した 5 つの視点を持って企業における情報活用インフラを考えていくことが成功の条件となるでしょう。

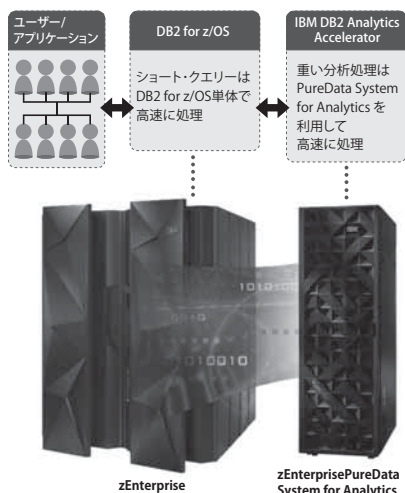


図 7. IBM DB2 Analytics Accelerator

#### [参考文献]

- [1] IDC :The Digital Universe Decade ? Are You Ready? (May2010)  
<http://www.emc.com/collateral/analyst-reports/idc-digital-universe-are-you-ready.pdf>
- [2] IBM : Global Technology Outlook 2012  
[http://www.zurich.ibm.com/pdf/isl/infoportal/GTO\\_2012\\_Booklet.pdf](http://www.zurich.ibm.com/pdf/isl/infoportal/GTO_2012_Booklet.pdf)
- [3] C.H. ダベンポート、J.G. ハリス：分析力を武器とする企業、日経 BP 社 (2008)
- [4] IBM Global Technology Outlook 2011  
<https://www.ibm.com/developerworks/community/wikis/form/anonymous/api/wiki/75d170a0-863b-4618-b557-7981a3228c5a/page/b3a03e38-8edf-44d9-b94f-e6b86bb7d5ab/attachment/0b631d82-b2a1-4568-998a-40a18248897f/media/gto2011booklet%20-%20EB%B0%B0ED%8F%AC%EC%9A%A9.pdf>
- [5] ITIC.:ITIC 2009 Global Server Hardware & Server OS Reliability Survey Results (July, 2009)  
<http://itic-corp.com/blog/2009/07/itic-2009-global-server-hardware-server-os-reliability-survey-results/>