

ビッグデータ・アナリティクス:

CIOが今知っておくべき 将来のアーキテクチャ、 スキル、ロードマップ

2011年9月

フィリップ・カーター (Philip Carter)

Sponsored by



THE
POWER
TO KNOW.

ビッグデータ・アナリティクス：

CIOが今知っておくべき
将来のアーキテクチャ、スキル、ロードマップ

ビッグデータという「すばらしい新世界」

「ビッグデータ時代」が到来しています。ペタバイト規模のデータウェアハウス、ソーシャルメディア上の会話、各種のセンサーから発生するリアルタイムのデータフィード、地理空間情報など、さまざまな新しいデータソースの出現によって企業は幅広い課題に直面していますが、これは同時に大きなチャンスでもあります。IDCの調査によると、従来のデータベースやアーキテクチャでは対応できない膨大なデータセットを用いた処理／発見／分析に必要となる新しいテクノロジーをCIOが採用し始めています。そして、企業が生成するデータはその量、高速性、多様性のすべての面で増大し続けており、そうしたデータに対応したハイエンドの分析機能、つまり「ビッグデータ・アナリティクス」こそが真の価値を引き出すと、IDCでは確信しています。

従来型のアナリティクスと、ここで「ビッグデータ時代」という括りの中で取り上げているアナリティクスとの大きな違いは、現時点で認識している必要性の有無によらず、幅広くデータを収集するという点です。分析の観点からすると、ビッグデータでは「何がわかっていないかがわからない」ことを意味します。こうした状況では、変数やモデルは完全に新しいものとなる可能性が高いため、これまでとは異なるインフラ戦略が求められるようになり、おそらく最も重要な点として、新たなスキルセットが必要となります。

このホワイトペーパーの目的は、ビッグデータが初期段階で企業組織、とりわけIT部門に及ぼす影響を明らかにすることです。そして、この影響が及ぼす結果として避けて通れなくなるのが、アーキテクチャ、提供モデル、将来に向けたロードマップの見直しです。そこで本書では、次の領域について詳しく検討していきます。

■ ビッグデータの定義

ここでは、ビッグデータを具体的な数量やしきい値で規定するのではなく(こうした数値は常に変化しますし、業種や市場のセグメントによっても異なるためです)、その代わりに、「極めて高速なデータの取得／発見／分析を実現することを通じて、膨大かつ多様なデータからビジネス価値を引き出せるように設計された新世代のテクノロジーおよびアーキテクチャ」という観点から、ビッグデータとは何かを定義します。

■ Hadoop, MapReduce, Key Value Storeとは何か？

「ビッグデータ現象」に対応するために市場で使われ始めている新たなテクノロジーをめぐっては、多くの宣伝文句が飛び交っています。ここでは代表的なものをいくつか取り上げ、その相対的な重要性を検討します。

■ アナリティクスにおけるビッグデータの意義

最も重要な点は、増え続ける大規模データセットの処理と分析はますます複雑化しており、そのため、ビジネス・アナリティクスへの取り組みに乗り出した企業



ビッグデータ・アナリティクス：

CIOが今知っておくべき 将来のアーキテクチャ、スキル、ロードマップ

組織の多くは、本質的なレベルで、データの領域に留まらない広範な情報管理戦略の見直しを迫られるということです。

- ビッグデータ・アナリティクスが重要な(および従来と異なる)理由
「この動向の何が新しいのか？」— これは多くの人が疑問に思っている点です。このセクションでは、少し前の「ビッグデータ以前の世界」におけるビジネス・アナリティクスの活用方法と、「新世界」におけるビッグデータ・アナリティクスとの違いを明確にします。また、幅広い業種に最も一般的に普及するようになるとIDCが予測している、さまざまな用途(ユースケース)にも注目します。
- スキルの側面：データ・サイエンティストの登場
ビッグデータ現象の現実化にともない、さまざまなテクノロジーや組織構造を新たに整備する必要性が生じているため、今後は「データ・サイエンティスト」の

需要が高まっていくでしょう。これは大量のデータセットから情報を引き出し、付加価値の高いビジネス・コンテンツに加工した上でデータの専門家ではないユーザーに提示できる能力を持つと同時に、新たに整備すべき最新の分析モデルを理解できる高度なスキルも身に付けた、新世代の分析プロフェッショナルを指しています。

- ビッグデータ・アナリティクスへの取り組みの緻密な計画
ビッグデータ・アナリティクス活用の道は、反復的な取り組みになります。そのため、幅広いフレームワークの一環として緻密に計画することが重要です。このセクションの目標は、そうした計画を的確に行う方法を検討するとともに、ビッグデータ・アナリティクスという素晴らしい新世界への取り組みに乗り出すCIOに向けて推奨事項を示すことです。

現状の概要

ビジネス・アナリティクスの台頭

世界のデータ量が爆発的に増大していることは、すでに広く指摘されています。IDC Digital Universeの最近の調査によると、生成および複製される情報量は、2011年には1.9ゼットバイト(1.8兆ギガバイト)を超える見込みであり、わずか5年間で9倍というペースです。

ビッグデータは、どこからともなく現れてきた急激な動向であるかのように思われます。しかし実際には、ビッグデータは目新しい現象ではなく、主流になりつつあることで関心が高まっているだけです。昨今のビッグデータ拡大の背景には、安価なストレージ、センサーおよびデータ収集テクノロジーの普及、クラウドや仮想化ストレージ・インフラによる情報アクセスの向上、さらには革新的なソフトウェアや分析ツールの出現といった要因があります。そのため、テクノロジー領域としてのビジネス・アナリティクスが、

CIOや業務部門(LOB)幹部の注目を集めているのも不思議ではありません。これを裏付けるデータとして、米国市場における5,722名のエンドユーザーを対象とした最近の調査では、ビジネス・アナリティクスが企業におけるIT構想のトップ5にランクインしています。ビジネス・アナリティクス採用の主な推進要因が保守的な姿勢にもとづいている点は従来どおりです。コストの管理、顧客の維持、業務の最適化が重視されているのは、先の見えない経済状況が続いていることを反映していると考えられます。



ビッグデータ・アナリティクス：

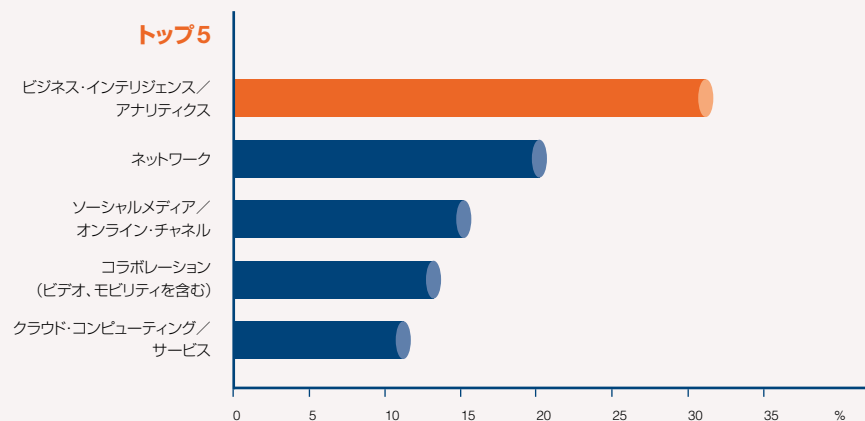
CIOが今知っておくべき 将来のアーキテクチャ、スキル、ロードマップ

とはいえ、最重要の要因に目を向けると、企業の規模と業種によって大きな違いが見られます。IDCが2011年2月に欧州の693の企業組織を対象に実施した調査でも同様に、回答者の51%が、優先順位の高いテクノロジーとしてBI（ビジネス・インテリジェンス）とアナリティクスを挙げています。アジア太平洋などの新興成長市場では、成長の次の波に乗ることが最も重視されています。

IDCが2011年2月に「Asia/Pacific C-Suite Barometer」(アジア太平洋の経営幹部に対する意識調査)の一環として行った1,000名以上のCIOと業務部門幹部へのインタビューによれば、今後1年間に競争優位性を獲得する上で重要なテクノロジー領域として、ビジネス・アナリティクスが第1位になっています。

図1：ビジネス・アナリティクスの台頭

Q：あなた(CIO/CTO)は「ICT*の活用による競争優位性の獲得」をお選びになりましたが、次のテクノロジーまたはソリューションのうち、ICTの活用度を高める上で最優先にしたいものはどれですか？



*ICT：Information and communications technology

出典：IDC、2011年

アジアでは新興成長市場における急激な成長の波に乗るうとして、ITに投資する企業が増えています。そうした企業はアナリティクス主導型のソリューションを活用することで、顧客に関する洞察を深め、リスクと財務指標の管理を効率化すると同時に、市場で自社独自の差別化を図ろうとしています。振り返ると、ほとんどの企業はビジネスプロセスの自動化とデータの収集を通じて業務効率を改善するためのアプリケーションに多大な投資を行ってきました。そうしたプロジェクトの多くは今も続いています。それは、経営幹部にとって明らかになってきた点があります。それは、経営幹部(および管理職)が適切な情報を適切なタイミングで把握し、それをもとめて組織内の適切なステークホルダーに指示を与えることができていないため(不十分なシステム統合とデータ品質の問題、およびパフォー

マンスと拡張性の問題などが原因)、経営目標を達成するために欠かすことのできない重要な意思決定プロセスが適切に機能していないということです。また、こうした状況の中、各業務部門では「影のIT」投資という新たな動向に乗る形で、ビジネス・アナリティクスを重視した部門独自のソリューションを調達・展開する動きが進んでいます。そのためCIOとしては、ITとビジネスの連携性・整合性の強化に力点を置いて、これらの問題を再検討する必要があります。この傾向は「ビッグデータ」が視野に入っていない状況でも見られますが、ビッグデータ問題が加わった場合はまさに「パーフェクト・ストーム(最悪の状況)」となり、ビッグデータ・アナリティクスの出番となります。



ビッグデータ・アナリティクス：

CIOが今知っておくべき
将来のアーキテクチャ、スキル、ロードマップ

用語について：

BIとアナリティクスの違い

ビジネス・アナリティクスに関しては、その定義や用語の使い方に注意が必要です。BI市場は成熟しているため、長い間に多くの用語が生まれ、すでに時代遅れとなったものもあれば、時間とともに定義が見直されてきたものもあります。例えば、「BI」という用語自体、クエリー／レポート／分析(QRA)テクノロジーのみを指す狭い意味で使われることもあれば、IDCがビジネス・アナリティクスと呼ぶものの全体を包括する広い意味(フロントエンド・ツールだけでなくデータウェアハウスや分析アプリケーションも含む)で使われることもあります。「アナリティクス」という用語もまた、比較的新しいこともあり、曖昧に使われがちです。最も広い意味では、予測的分析、最適化と将来予測、分析アプリケーションを含む高度な分析機能を指しているケースもあります。その一方で、一部の分野、例えばWebアナリティクスなどでは、「アナリティクス」という

用語はデータの集計値などを提示するダッシュボードを意味します。

このホワイトペーパーでは、「BI」を狭い意味ではQRAツールもしくはBI全般として、また広い意味ではIDC用語における「ビジネス・アナリティクス」として解釈します。「アナリティクス」は、高度な分析機能(データマイニング、統計、最適化、予測)、または分析アプリケーション(予実管理・経営管理[FPSM]、CRM／マーケティング分析、サプライチェーン分析など)と解釈します。そして、「ビジネス・アナリティクス」は上記(BIとアナリティクス)の組み合わせに、さらにデータウェアハウス・テクノロジーも加えたものと解釈します。この定義は、IDCの「Business Analytics Taxonomy for 2011」(2011年版ビジネス・アナリティクス分類)に明確に示されています(下の図2を参照)。

図2：Business Analytics Taxonomy for 2011 (2011年版ビジネス・アナリティクスの分類)





ビッグデータ・アナリティクス：

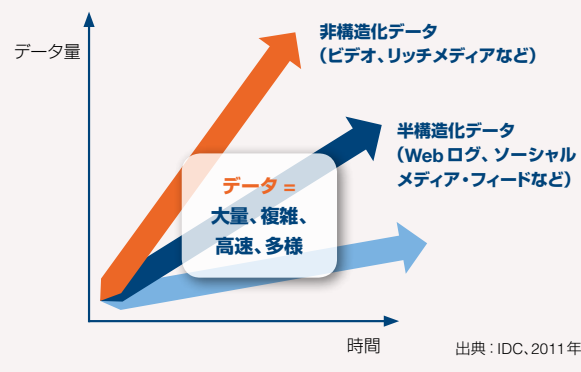
CIOが今知っておくべき
将来のアーキテクチャ、スキル、ロードマップ

「ビッグデータ」の定義

ビッグデータでは、どのようなコンテンツを作成すべきか、それをどのように利用するかは、それほど重要ではありません。より重要なのはデータの分析であり、それをどのように行うべきかという側面です。これは実体のある「モノ」というよりは、ITの中にある多くの境界を横断するための原動力／活動のようなものです。IDCではビッグデータ

を次のように定義しています。「ビッグデータ・テクノロジーとは、極めて高速なデータの取得／発見／分析を実現することを通じて、膨大かつ多様なデータから経済価値を引き出せるように設計された新世代のテクノロジーおよびアーキテクチャを意味します」

図3：「ビッグデータ」の定義



■ 量

ある種のデータは、構造化データの領域として総称されるのが一般的です。これはトランザクション用のデータストアでも見られ、急速に普及が進むオンライン上の活動で個人や企業が生成し、永続的に保持される電子的な痕跡と結びついています。マシン間(M2M)通信のセンサーデータも、この領域に分類されます。それ以外の構造化データは、既存のデータウェアハウスやデータマート内に存在し、時間の経過とともにペタバイト規模まで膨れ上がります。

■ 多様性

ビッグデータ現象のもう1つの特徴は、半構造化データと非構造化データも分析対象になることです。テキスト、ビデオ、その他の形態のメディアを対象として適切な分析を実行するためには、従来とはまったく異なるアーキテクチャとテクノロジーが必要となります。例えばソーシャルメディア現象を見ればわかるように、多くのマーケティング部門は今、Facebook、Twitter、YouTubeに投稿される内容にもとづいて評判やブランドイメージを分析する方法を模索しています。こうした動きは、中国のRenRenや韓国のNate

など、国内に大規模なソーシャルメディア・サイトがあるアジア圏では、さらに複雑な様相を呈しています。

■ 高速性

ビッグデータをより高い頻度で定期的に分析する需要も高まっています。例えば、一部を抽出したサンプルではなくトランザクション全体を分析対象とすることで、取引リスクの全体像をリアルタイムで把握したい場合などです。

以上をまとめると、ビッグデータとは、現在のデータベースやアーキテクチャでは保管や管理ができないほどの量、多様性、高速性をともなうデータセット、ということになります。IDCでは、特定のしきい値(例えばテラバイト級など)より大きいものをビッグデータと定義することを意図的に避けています。その大きな理由は、このような値は業種や規模に応じて変化する数字であることに加え、時間の経過とともに増大することが明らかであるためです。それよりも重要なのは、この現象から企業組織がどのような価値を引き出せるかという点であり、また、その結果として、価値を引き出すための情報戦略を再検討する必要が生じるという点です。



ビッグデータ・アナリティクス：

CIOが今知っておくべき
将来のアーキテクチャ、スキル、ロードマップ

その他の定義

Hadoop、MapReduce、Key Value Store

ビッグデータが重要な課題として注目されるようになるとともに、市場には多種多様な新しいテクノロジーが登場してきました。下の表は、そうしたテクノロジーの概要を開発の背景も含めてまとめたものです(ただし、すべてを網羅しているわけではないことに注意してください)。

表1：ビッグデータ関連テクノロジー(用語)

| テクノロジー | 概要と背景 |
|--|---|
| BigTable | Google File System 上に構築される独自仕様の分散データベース・システム。HBaseの着想元です。 |
| Cassandra | 分散システム上で大量のデータを処理できるように設計された、オープンソースのデータベース管理システム。このシステムは当初Facebookで開発されましたが、現在はApache Software Foundationのプロジェクトとして管理されています。 |
| データウェアハウスおよび分析プラットフォーム | サーバー、ストレージ、(複数の)オペレーティングシステム、データベース、ビジネス・インテリジェンス、データマイニング、その他のソフトウェアの統合セットで構成され、データウェアハウスに特化してプリインストールと事前の最適化を施した専用機器。 |
| 分散システム | 複数のコンピュータがネットワーク経由で相互通信しながら、ひとつの計算問題を解くシステム。問題を複数のタスクに分割し、並列して稼働する複数のコンピュータでそれぞれのタスクを解決します。価格性能比、信頼性、拡張性の向上がメリットです。 |
| Google File System | Googleが開発した独自仕様の分散ファイルシステム。Hadoopの着想元のひとつでもあります。 |
| Hadoop | 特定の種類の問題に関する大量のデータセットを分散システム上で処理するオープンソースのソフトウェア・フレームワーク。GoogleのMapReduceとGoogle File Systemが着想元になって開発されました。当初はYahoo!で開発されましたが、現在はApache Software Foundationのプロジェクトとして管理されています。 |
| HBase | GoogleのBigTableをモデルとするオープンソース(無料)の非リレーショナル分散データベース。当初はPowersetによって開発されましたが、現在はApache Software Foundationのプロジェクトとして管理され、Hadoopの一部となっています。 |
| MapReduce | 特定の種類の問題に関する大量のデータセットを分散システム上で処理するためにGoogleが導入したソフトウェア・フレームワーク。Hadoopにも実装されています。 |
| 非リレーショナル・データベース / Key Value Store | 非リレーショナル・データベースは、リレーショナル・データベースと異なり、データをテーブル(行と列)には保管しません。Key Value Storeでは、スキーマレス(非SQL)のエンティティ管理が可能です。 |

これらの用語の一部はこのホワイトペーパーでも使用していますが、その詳細を深く検討することはありません。なぜなら、ある企業のIT担当幹部が最近指摘したように、「テクノロジーを知ることと、それを適切な環境に適用することは、まったく別の話」だからです。新しいテクノロジーは、純粋にテクノロジーとして検討するだけでなく、ビジネス要件との整合性を可能な限り追求する必要があります。しかしそうは言うものの、ほとんどのIT担当幹部は、この領域におけるテクノロジーと傾向の最新状況を把握していません。また、把握して

いるとしても、実際の戦略は、エンタープライズ・アーキテクチャ担当チームから数名を選び、「ビッグデータ」を扱うために活用され始めている最新テクノロジー(例：インメモリー、Hadoop、MapReduce、Key Value Storeなど)を試させる、といった程度でしかありません。



ビッグデータ・アナリティクス：

CIOが今知っておくべき
将来のアーキテクチャ、スキル、ロードマップ

ビッグデータ・アナリティクス “旧世代” 対 “新時代”

「この動向の何が新しいのか？」—これは多くの人が疑問に思っている点です。このセクションでは、少し前の「ビッグデータ以前の世代」におけるビジネス・アナリティクスの活用方法と、「すばらしい新世代」におけるビッグデータ・アナリティクスとの違いを明確にします。また、幅広い業種に最も一般的に普及するようになるとIDCが予測している、さまざまな用途（ユースケース）にも注目します。大多数のIT部門では、時間の経過とともにインフラのアーキテクチャを進化させてきました。メインフレームがすべてを支配する1980年代の環境から、1990年代にはクライアント／サーバーへと移行し、今世紀に入ってからはWebを全面的に活用し、現在では「プライベート・クラウド」として広く知られる環境へと進化しています。これがうまく機能して「ニルヴァーナ（涅槃）」と呼べるような理想的な状態を実現した場合には、インフラのリソース（サーバー、ストレージ、ネットワーク）は仮想化および統合化され、業務部門のユーザー自身が必要な環境を自動的に構

築できるようになります。しかも、SLA（サービス内容合意書）は完全に満たされ、セキュリティ、パフォーマンス、可用性、コストに関するプロファイルの透明性が、サービスカタログという形であらゆるユーザーに保証されるようになります。ただし、こうした「ニルヴァーナ」状態のインフラを実現している企業は、たとえ存在するとしても非常に少なく、そうした企業でもデータセンターでは複雑に絡み合うコンピューティング・リソースとの格闘が続いているのが実情です。また前述のとおり、昨今はビッグデータという外圧が高まっているため、CIOはインフラの再構築を迫られています。この傾向は、分析機能を全社規模で展開するにはどうすればよいかという文脈で特に顕著です。

下に示すのは、インフラの領域で今現在進行しつつあり、ビッグデータ・アナリティクスの領域にも影響を及ぼしつつある変化の概要です。

表2：“旧世代” 対 “新時代”（ビッグデータインフラ）図

| | 旧世代 | 新時代 |
|----------------|----------------|--|
| 基本形態 | 連携性を欠いたインフラ群 | リソースのプール |
| アーキテクチャ | パフォーマンスを「調整」 | 直線的な拡張性（分散並列処理や「インメモリー」ストレージの特長） |
| 提供モデル | オンプレミス（社内運用）方式 | ハイブリッド方式（クラウド・パースティング機能付き）とアプライアンスの広範な活用 |



ビッグデータ・アナリティクス：

CIOが今知っておくべき 将来のアーキテクチャ、スキル、ロードマップ

以下では、この領域におけるIDCの調査にもとづき、こうした問題に対処するCIOに向けて3つのヒントを提示します。

■ クラウド・バースティング

プライベート・クラウドは、前述した全社規模の分析環境に求められる要件とも親和性が高い取り組みですが、CIOとしてはワークロードを厳密に評価して、できるだけリスクを軽減する必要があります。この取り組みで重要となるのは、外部ベンダーが提供するクラウド・バースティング機能(すなわち、サービスとしてのインフラ：IaaS)の評価です。これはリアルタイム分析環境の活用に取りかかる段階では特に重要であり、適切な評価を通じて、インフラのリソース使用が需要と密接に対応していることや、パフォーマンスと可用性に関して問題がないことを確認する必要があります。

■ 分析アプライアンス

提供モデルに関しては、ビッグデータに対処しようとする企業にパフォーマンス面で大きなメリットをもたらすのは分析アプライアンスであることが、IDCの調査結果で明らかになっています。さらに、アプライアンスではソフトウェアが最適化され、あらかじめ統合されているため、通常は導入展開に要する期間も短くなります。CIOを対象とした最近のグローバル規模の調査によれば、回答者の10%が、2011年の提供モデルとして分析アプライアンスを検討しています。また、多くのCIOがこうしたアプライアンスを既存のデータウェアハウス環境内に統合したいと考えているため、参考にできるリファレンス・アーキテクチャの需要が高まると、IDCでは考えています。また、提供モデルとして分析アプライアンスの採用が増えるのと呼応する形で、IT部門では技術スキル(インストール、設定、管理など)に対する予算を縮小し、ビジネス上の効果を部門横断で推進するハイエンドの分析スキルに対する予算は増加するだろうというのも、IDCの見方です。

■ エンタープライズ・アーキテクチャ

エンタープライズ・アナリティクスには、企業組織の成長に対応して効果的に規模を拡張できるエンタープライズ・アーキテクチャが必要です。ビッグデータ・アナリティクスの台頭により、この問題への対応は緊急性が高まっています。企業は今、「ハイパフォーマンス分析環境」、すなわち、「In-Database分析、並列処理、インメモリ・ストレージを活用して、データの量、高速性、多様性の増大に対処するための環境」の構築を真剣に検討する必要があります。特に、非構造化データの扱いについては、Hadoopにもっと注目すべきでしょう。HadoopはApacheに含まれるオープンソースのソフトウェア・フレームワークであり、コンピュータのクラスターで大量のデータを分散処理する環境を提供します。ただし、グローバル標準とローカル要件の間には相容れない部分が常に存在することは理解しておく必要があります。Hadoopの使用もその典型例といえるでしょう。もう1つの選択肢として期待されるのは、異種混在のワークロード(例えば分析と業務)を同一のインフラ環境で処理できる能力です。前述したアプライアンスなどが、これに相当します。CIOは、特定のビジネス課題の解決に貢献する方法を検討すると同時に、グローバルなアーキテクチャ標準/仕様を把握している必要もあります。グローバル規模のガバナンス・モデルのもとでは、こうしたテクノロジーを本稼働環境で使用できない場合も想定されますが、業務部門からの期待が高まるにつれ、IT部門はエンタープライズ・アーキテクチャにおける優先課題とローカルレベルのニーズとの整合性のあり方を再検討する必要に迫られるでしょう。



ビッグデータ・アナリティクス：

CIOが今知っておくべき
将来のアーキテクチャ、スキル、ロードマップ

ここで最も重要な点は、こうした増え続ける大規模かつ複雑なデータセットの処理と分析はますます複雑化しており、そのため、ビジネス・アナリティクスへの取り組みに乗り出した企業の多くは、本質的なレベルで、データの領域に留まらない広範な情報管理戦略の見直しを迫られるということです。しかし、その結果として期待できる効果

は計り知れません。グローバル規模の小売チェーン企業において全品目の価格最適化や不正行為の検知をリアルタイムで行える状況を想像してみれば、ビッグデータ・アナリティクスを活用して解決できる問題がどのようなものか理解できるでしょう。

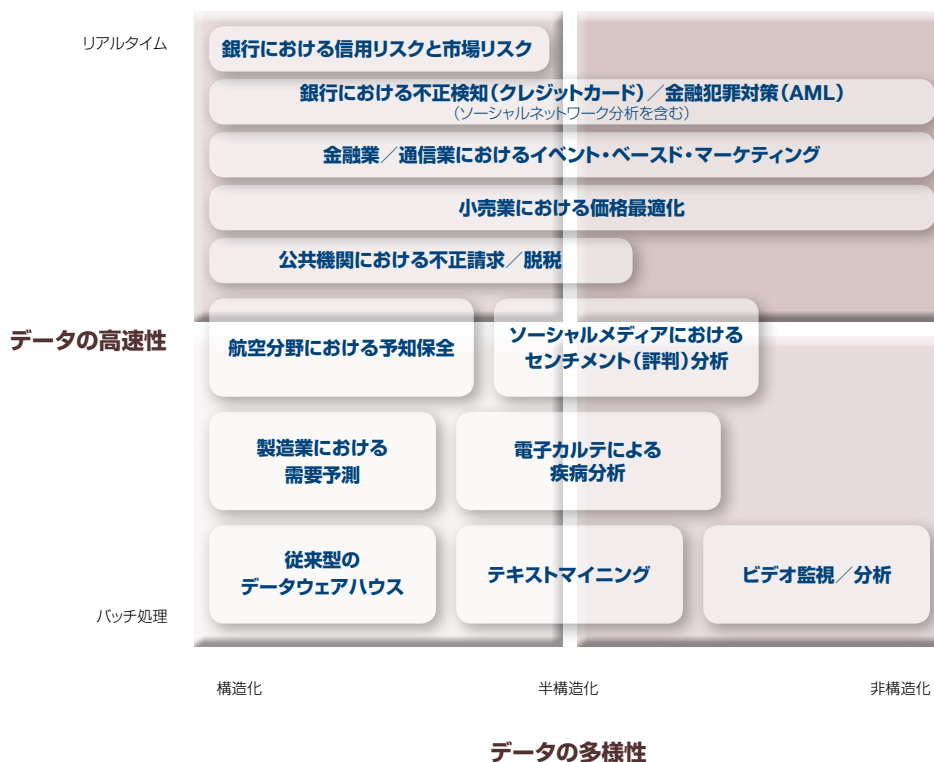
表3：“旧世代”対“新時代”（ビッグデータ・アナリティクス）

| | 旧世代 | 新時代 |
|----------|--------------|--------------------------------|
| データセット | 事前定義 | 包括的かつ反復的 |
| データの高速度性 | バッチ処理 | プロアクティブかつ動的 (適切な場合にはリアルタイム) |
| データ分析 | 基本的に履歴データが対象 | 予測と最適化 |

ただし、こうしたアナリティクスの可能性の高さは明らかであるものの、あらゆる用途に適切または適用可能とは限らない点を理解しておくことが重要です。下の図に示

すように、どのような用途に適しているかは、ビッグデータの2つの次元、つまり高速度性と多様性を軸に考えることができます。IDCでは考えています。

図4：ビッグデータ・アナリティクスの潜在的な用途





ビッグデータ・アナリティクス：

CIOが今知っておくべき 将来のアーキテクチャ、スキル、ロードマップ

ビッグデータ・アナリティクスの導入がビジネス価値の飛躍的な向上にどのように貢献するかについて理解を深めるためには、これらの用途の詳細を知るのが一番でしょう。

■ 銀行におけるリアルタイム不正検知

これは銀行業において、複数の金融商品、業務部門、販売チャネルにまたがって不正行為を検知、防止、管理するための機能を指します。これを実現するには、トランザクションに関与する各種の関連情報(カード、口座、顧客、端末ID、IPアドレスなど)について履歴データを収集する仕組みを整え、POS(販売時点管理)トランザクションの処理中に顧客の異常な行動をとらえ検知の精度を高める必要があります。収集した情報は複数の予測モデルで使用し、不正検知と信用リスク管理に活用できます。

■ 小売業における価格最適化

小売業では、プロモーション、季節変動、重要なイベン

トの影響を考慮した需要予測シナリオにもとづいて幅広い商品の価格をリアルタイムで最適化できれば、利益を大きく伸ばすことができます。ソーシャルメディアのセンチメント(評判)分析を併用すれば、この機能をさらに強化し、特定の商品に関する顧客の需要をリアルタイムに近いタイミングで把握できるようになります。

■ 電子カルテによる疾病分析

現在では、医療サービスの進歩により、患者の病歴をすべて電子形式で把握できるようになっています。これはビッグデータ・アナリティクスにとって大きなチャンスです。例えば、糖尿病のような病気の場合、患者の病歴を食事のデータと関連付けることができ(おそらく小売業のマーケット・バスケット分析との連携も可能)、運動のスケジュールも最適化できるとしたら、医者は従来なら夢物語でしかなかった斬新な知見・洞察を得ることができるでしょう。

スキルの側面

前述したように、IDCでは、企業組織が生成するデータはその量、高速性、多様性のすべての面で増大し続けており、そうしたデータに対応したハイエンドの分析機能、つまり「ビッグデータ・アナリティクス」こそが真の価値を引き出すことになると確信しています。アジアでは(米国や欧州に本拠を置く一部の多国籍企業を除けば)、どのようなタイプとレベルのスキルが必要になるかを認識している企業はほとんどありません。この状況は、(ビッグデータ現象に関係なく)一般的な傾向として、ハイエンドの分析機能という領域に関する問題意識とスキルが歴史的に欠如していたことに関係していると、IDCでは考えています。

ハイエンドの分析機能を活用するには、主に2つの
カテゴリーで新たなスキルセットが必要となります。

□ 技術スキル

従来のデータベースやアーキテクチャでは対応できない膨大なデータセットの処理、発見、分析に必要な新しいクラスのテクノロジー(インメモリー、Hadoop、MapReduce、Key Value Storeなど)を扱うためのスキル。こうしたテクノロジーの一部はアプライアンスとして提供されるため、データを活用するためには、ソフトウェアがハードウェアとどのように連携するかを正しく理解できるスキルが欠かせません。

□ 新しいタイプのアナリスト/データサイエンティスト
“旧世代”のアナリティクスと、ここで「ビッグデータ時代」という括りの中で取り上げているアナリティクスとの大きな違いのひとつは、現時点で認識している必要性の有無によらず、幅広くデータを収集するという点です。分析の観点からすると、これは「何がわかっていないかがわからない」ことを意味します。こうした状況では、大量の非構造化データが分析対象になるため、変数やモデルは完全に新しいものとなる可能性が高くなります。これが何を意味するかといえば、常に発見が出发点となる「サンドボックス・メンタリティ」(砂場遊びのような自由で実験的な知性)を推奨する



ビッグデータ・アナリティクス：

CIOが今知っておくべき 将来のアーキテクチャ、スキル、ロードマップ

方向性で、パワーユーザーが分析機能を業務に活かす方法を再考する必要があるということです。一般論としては、データマイニングや統計分析の経験があれば、このタイプの分析の優れた出発点となるでしょう。そして今後は、「データ・サイエンティスト」の需要が高まっていくでしょう。これは強力な統計分析スキルを持つ新世代の分析プロフェッショナルです。大量のデータセットから情報を引き出してデータの専門家ではないユーザーに価値を提示できる能力と、短期間でビジネスに多大な効果をもたらす最新のアルゴリズムと分析モデルを理解できる高度なスキルを兼ね備えていることが大前提となります。IDCの調査でも、こうした分析重視のスキルセットに対する関心の高まりが世界中で見られます。その役割と責任はまだ明確に定義されてはいませんが、基本的には、前述した「何がわかっていないかがわからない」（つまり、大量の非構造化データが分析対象となるため、変数やモデルは完全に新しいものとなる可能性が高くなる）状況に最適な人材であるのは確かです。こうした新しいデータタイプやデータ構造を相手に的確な分析を行うためには、まさに「即戦力」となるスキルと創造性が求められます。

例えば、いわゆる「ソーシャルメディア現象」（ビッグデータのうち半構造化データと非構造化データの部分に大きく関わっています）を見ればわかるように、多くのマーケティング部門は今、Facebook、Twitter、YouTubeに投稿される内容（ご想像どおりの膨大な量です）にもとづいて評判やブランドイメージを分析する方法を模索しています。こうした動きは、中国のRenRenや韓国のNateなど、国内に大規模なソーシャルメディア・サイトがあるアジア圏では、さらに複雑な様相を呈しています。これに関しても、現状のIT部門は、最高マーケティング責任者（CMO）がまず頼りにしたいと思う存在ではありません。ほとんどのIT部門には、ビジネスニーズを理解するスキルが欠けているためです（また、多くのIT部門はまだ、ソーシャルメディアの利用に関するポリシーやガバナンスの面でのような役割を果たすべきかを模索している段階にあります）。そのため、技術、ビジネス、人間関係に関するスキルの観点から、IT部門の人員構成を見直すことから始める必要があります。

次に示す成熟度モデルは、企業におけるビジネス・アナリティクスの導入過程に応じて、こうしたスキル（技術とビジネスの両方）の獲得をどのように計画し、ビッグデータ・アナリティクスの時代にふさわしく進化させていけばよいかに関するIDCの考え方を示しています。



ビッグデータ・アナリティクス：

CIOが今知っておくべき
将来のアーキテクチャ、スキル、ロードマップ

図5：ビッグデータ・アナリティクスの成熟度モデル図

| 段階 効果 | 旧世代 | | | 新時代 |
|------------------------|-----------------------------------|---|--|--|
| | パイロット | 部門レベルのアナリティクス | 全社規模のアナリティクス | ビッグデータ・アナリティクス |
| スタッフのスキル (IT部門) | 分析の専門知識はほとんどないか皆無で、BIツールの基本的な知識のみ | パフォーマンス、可用性、セキュリティに焦点を絞ったデータウェアハウス・チーム | 高度なデータモデルの作成者とデータ管理者をIT部門の主力として位置付け | 「データ・サイエンティスト」を配属したビジネス・アナリティクス・コンピテンシー・センター(BACC) |
| スタッフのスキル (ビジネス部門/IT部門) | BIツールの機能に関する知識 | ビジネス・アナリストが少なく、高度な分析の使用は限定的 | 豊富な知識を持つ分析モデル作成者と統計専門家が能力を発揮 | 複雑な問題の解決をビジネス・アナリティクス・コンピテンシー・センター(BACC)に統合 |
| テクノロジーとツール | 単純な履歴データにもとづくBIレポートおよびダッシュボード | データウェアハウスを導入し、BIツールを幅広く活用。分析用データマートは限定的 | In-Database方式のデータマイニング。並列処理と分析アプライアンスの利用は限定的 | 複数のワークロードに対応したアプライアンスを広範に導入。最新テクノロジーにも対応できるアーキテクチャとガバナンス |
| 収支面の影響 | 収益への目立った影響はなし。ROIモデルは未導入 | 収益創出に関するKPIをいくつか導入し、ROIを明確に理解 | 収益に大きな効果が及ぶ(定期的な計測と監視も実現) | 分析にもとづいてビジネス戦略を立て、競合他社との差別化を強化 |
| データガバナンス | ほとんどないか皆無(特命の調査研究チームのレベル) | データウェアハウス・モデル/アーキテクチャは初期段階 | データ定義とデータモデルを標準化 | 明確なマスターデータ管理戦略 |
| 業務部門(LOB) | 不満がつのる | 可視化が実現 | 連携が実現(業務部門担当役員を含む) | 部門横断が実現(CEOが状況を把握) |
| CIOの関与 | 裏方のみ | 限定的 | 積極的 | 変革を推進する柔軟性が実現 |
| 企業の割合 (IDCの推定) | 20% | 65% | 10% | 5% |

ビッグデータ・アナリティクス時代にふさわしいスキルの獲得と開発に関しては、ビジネス部門とIT部門を横断してこの技術の活用を推進するための中心組織として、ビジネス・アナリティクス・コンピテンシー・センター(BACC)を創設することが重要です。こうした組織構造を整備することで、この変革に関与する主なステークホルダーの役割と責任が明確になるだけでなく、社内の可視化が進み、教育の仕組みが整い、IT部門とビジネス部門の溝が埋まることになると、IDCは考えています(ビジネス部門側ではマーケティングと販売が特に重要であり、これらの部門の重要人物がBACCに関与する必要があります)。なぜなら、このようなプロジェクトでは、最前線のスタッフの意思決定の改善こそが第一目標となるからです。

スキルの次元に関連して、こうした中心組織は以下の領域に関与する必要があるとIDCは考えています。

- テクノロジーの特定と開発
- ビジネスケース(投資対効果検討書)の作成とROIの裏付け

- マスターデータ管理、データ品質、データモデルに関する明確なポリシーとガイドラインを備えたデータガバナンス・フレームワークの整備
- 重要なステークホルダーが適切なタイミングで関与できる仕組みを整え、IT部門とビジネス部門の連携を確立
- 必要な変革の実現と、その結果としてのビジネス目標の達成を、ITの観点からサポートする役員としてCIOに関与させること

ビッグデータ・アナリティクスが秘める可能性を真に活用できる成熟度レベルに達している企業はごく限られており、実際問題、すべての関連領域で一気に改革を進めるのは簡単ではありません。しかし、現在のような経済環境で企業が本当の意味で差別化を図ろうとするなら、この変革を避けて通ることはできません。そして、この変革で重大な役割を果たす必要があるのは、CIO(およびIT部門)です。次のセクションでは、IDCがこの取り組みで留意すべきと考えている点について、いくつかヒントを示します。



ビッグデータ・アナリティクス：

CIOが今知っておくべき 将来のアーキテクチャ、スキル、ロードマップ

CIOのための

「ビッグデータ・アナリティクス」チェックリスト

■ 将来を見据えたアーキテクチャの構築

分析業務では長年にわたり、基盤となるハードウェアの拡張性が限られていたため、さまざまな「回避策」の特定と活用に注力してきました。その結果、多くのIT部門では、マテリアライズド・ビュー、つまり事前に抽出と計算を済ませたデータ構造を作成するようになりました。ビジネスユーザーがこれらを使用しても、元データを扱う業務システムのパフォーマンスに悪影響が及ばないようにするためです。他方、クラスタリング、並列処理、インメモリーなどのテクノロジーが登場してきた結果、現在では、元になる業務データのすべてを分析環境で活用できるようになっています。ただし、可能だからといってデータ容量をやみくもに増やすのが適切でないように、分析でもそれと同じ罫にはまらないことが重要です。特定のインフラ規模がすべての状況に最適であるとは限らないため、状況に応じて複数の提供モデルを評価する必要があります。こうしたモデルとしては、クラウド(特にそのパースティング機能)、アプライアンス、従来のクライアント/サーバーや3層Webアーキテクチャなどが考えられます。

■ 「サンドボックス・メンタリティ」の醸成

旧式のバッチモードによる従来のアナリティクスと、ここで「ビッグデータ時代」という括りの中で取り上げているアナリティクスとの大きな違いのひとつは、現時点で認識している必要性の有無によらず幅広くデータを収集するという点です。分析の観点からすると、これは「何がわかっていないかがわからない」ことを意味します。こうした状況では、大量の非構造化データが分析対象になるため、変数やモデルは完全に新しいものとなる可能性が高くなります。これが何を意味するかといえば、常に発見のプロセスが出発点となる「サンドボックス・メンタリティ」(砂場遊びのような自由で実験的な知性)を推奨する方向性で、パワーユーザーが分析モデルを構築する方法を再考する必要があります。これは特に、非構造化データ、半構造化データ、構造化データを結びつけて活用するという観点から重要です。その一環として、ソーシャルメディアにおけるニュアンス(その多くはジェネ

レーションY、ジェネレーションZ、さらには2000年世代の感覚)を理解するために、新しい種類のスキルを採り入れていく必要もあります。

■ 「つぎはぎ」導入に走らない自制心

新しい優れたテクノロジーが市場に登場すると、IT部門はビジネスへの即効性を求めて、「つぎはぎ」式に導入しようとする傾向があります。ある程度の実験は良いことです(これは前項で説明した「サンドボックス・メンタリティ」の考え方にも通じるもので、HadoopとMapReduceはまさにこれに該当します)が、CIOとしては目先のビジネス価値だけを意識した実験に時間を取られすぎないように注意する必要があります。

■ 適切なチームの編成

このプロセスの最初のステップは、CIOが自分の率いるIT部門を評価し、関連するスキルレベルと組織構造の妥当性を精査することです。場合によっては、IT部門が変わったことをビジネス部門に認識してもらうために、何らかの内部変革も必要になるでしょう。次に、適切な人材に必要な権限を与え、IT部門としてアナリティクス戦略を実行できるようにする必要があります。ここでは、関連するプロセスとガバナンス構造を整備して、ビジネス部門の期待に効果的に応えていくための土台を築きます。その一環として、CIO自身には、基盤となる分析テクノロジーの機能を深く理解することが求められます。その上で、各業務部門の担当役員と連携しながら、基盤となる分析テクノロジーの機能を最適なレベルで活用できる高度なスキルとマインドを備えた適切なマネージャーとナレッジワーカーを雇用する必要があります。

■ アナリティクスの全社的な展開

これまで、この領域におけるITプロジェクトの多くは、データウェアハウスを構築してさまざまなBIツールと組み合わせることにより、データに潜む情報を可視化してエンドユーザーに提供することを重視してきました。しかし、高度な分析機能の観点からすると、ITスキルが不足している場合、こうしたプロジェクトの大半は部門別の戦術的なレベルにことどまり、いわゆる「サイロ化」(個別化、非連携化)の状態が満足し、それ以上を求めない雰囲気になりがちです。その結果、リスク調整後資本利益率(財務、信用リスクスコア、顧客属性データを組み合わせた指標)のような総合的な



ビッグデータ・アナリティクス：

CIOが今知っておくべき 将来のアーキテクチャ、スキル、ロードマップ

評価は不可能になってしまいます。この状況を変えるためには、IT部門とビジネス部門のコラボレーションを新たなレベルに引き上げることが必要です。ここでのCIO自身の役割は、アナリティクスを全社的に導入するための手法に集中し、関連のプロジェクトを確実に成功に導くことです。

■ ガバナンスと活用支援

これについては、データウェアハウス・テクノロジーへの既存の投資が適切なものであれば、期待どおりのメリットを実現できます。IT部門が整備するデータモデルとリファレンス・アーキテクチャにより、さまざまなビジネス部門を横断してデータ定義とデータ標準の一貫性が確保されます。データガバナンスをめぐる業務と分析のギャップを解消するためには、マスターデータ管理(MDM)の領域で高度な取り組みが必要になりますが、基本的には、この(IT部門が整備した)プラットフォームによって、IT部門が必要とする管

理と制御も実現するはずで、ビジネスにおける活用支援の面では、ビジネス・アナリティクスをビジネスプロセス管理機能と組み合わせた新たなクラスのプロジェクトが登場し始めていると、IDCでは見ています。具体的には、ルール管理、データマイニング、クエリーとレポート、複合イベント処理(CEP)、コラボレーション、BPMスイート、検索、コンテンツ分析などのツールを含む、意思決定管理のソフトウェア・コンポーネント群に関するプロジェクトです。IT部門は、社内の各組織の意思決定プロセスとその基盤をなす意思決定管理ソフトウェアについて理解を深めることで、データウェアハウスやBIに対する既存の投資を補完することができます。そして、それを実際に行ったIT部門こそが、ITガバナンスとビジネスにおける活用支援とのジレンマを最も効果的に管理できる位置に立つことになるでしょう。IDCではそう確信しています。

結論

ビジネス・アナリティクスの成熟度と導入状況はさまざまですが、多くの企業が今、この領域でより高度なソリューションや製品を活用しようと動き出しているのは間違いありません。この方向に進むためには、ビジネス・アナリティクスの導入に先立って、まずは戦略的な計画を立て、手堅いロードマップを作成する必要があります。業務部門の新世代のマネージャーたちは、ビジネス・アナリティクスの活用がもたらす競争上のメリットを十分に理解しており、この領域のテクノロジーを積極的に導入したいと考えています。そしてその方向性に進むにあたっては、目標とする変革の達成を能動的に推進していくために、特に以下の領域に重点を置いた新たな取り組みが求められていると、IDCは見ています。

- CIOの役割の強化。全社的なアナリティクス戦略の策定において重要な役割を果たし、分析テクノロジーがビジネスに対して期待どおりの効果を実証にもたらすように努めることで、組織変革への影響力を高める必要があります。
- 現状に代わる提供モデルの評価。ビッグデータに対応したアプライアンス、インメモリー、Hadoopなどを幅広く検討します。
- 業務部門による関心度合いの向上と可視化の強化。ビジネス・アナリティクス・プロジェクトにおける次の達成目標にCEP(複合イベント処理)とBAM(ビジネス・アクティビティ・モニタリング)のテクノロジーを統合し、IDCが「意思決定管理」と呼ぶ領域について新たなクラスのプロジェクトを推進します。

役員会におけるCIOの役割の重要性は高まりつつあり、ビジネス・アナリティクスなどの高度なアプリケーションの購買判断の鍵を握る存在となりつつあります。さらに、CIOとIT部門には今、幅広いビジネス・アナリティクス機能の活用を通じて、新たなビッグデータ時代に対応した新しい情報管理戦略を確立するとともに、より高度な意思決定支援機能を全社規模でビジネス関係者に提供することが求められているのです。

#AP14962U

この刊行物について

この刊行物は IDC の Go-to-Market Services 部門が作成しました。IDC の Go-to-Market Services 部門は、さまざまな企業が配布できるように IDC のコンテンツを幅広い形式で提供しています。IDC のコンテンツを配布するライセンスは、ライセンサーによる支持やライセンサーに対する意見を示唆するものではありません。

著作権と制限事項

IDC に関する情報や IDC への言及を広告、プレスリリース、宣伝資料で使用する場合には、事前に IDC から書面による承認を受けていただく必要があります。承認申請については、GMS 部門の情報担当者までお問い合わせください(+65-6829-7757 または gmsap@idc.com)。この文書の翻訳および/またはローカライズには IDC からの追加ライセンスが必要です。

IDC の詳細については、www.idc.com をご覧ください。IDC GMS の詳細については、www.idc.com/gms をご覧ください。

IDC Asia/Pacific, 80 Anson Road, #38-00 Fuji Xerox Towers, Singapore 079970。電話：+65-6226-0330 FAX：+65-6220-6116 www.idc.com。

Copyright 2011 IDC. Reproduction is forbidden unless authorized. All rights reserved.