

## 企業内のビッグデータのマイニングによる ビジネス・インテリジェンスの向上

ビッグデータのマイニングと分析は、ビジネスのパターンとトレンドに関する深く豊かな知見をもたらし、製造、セキュリティ、マーケティング、IT の各分野の運用効率と競争力向上の原動力となります。

### 概要

インテル IT 部門は、インテルのビジネス・インテリジェンス (BI) の向上を目指して、ビッグデータ分析用のシステムとスキルの整備を進めています。ビッグデータの主要部分は、企業データの最大 90% を占める大規模な非構造化データセットで構成されます。多くのソースから得られるあらゆる形式のビッグデータのマイニングと分析は、ビジネスのパターンとトレンドに関する深く豊かな知見をもたらし、製造、セキュリティ、マーケティング、IT の各分野の運用効率と競争力向上の原動力となります。

ビッグデータのマイニングと処理に基づいて予測的および処方的な分析を実行するための組織的スキルの開発は、将来的なインテルの業績を促進する最大の要因であり、以下のメリットをもたらします。

- より的確な意思決定
- ビジネス速度の向上
- イノベーションの加速
- 新規市場の発見と開拓

並列コンピューティングの進化によってビッグデータの処理が可能になった結果として、情報の価値を完全に理解するよりも前に、まず情報を収集、格納することが一般的な慣行となりました。そして、従来は規模が大きすぎて扱えなかった多くのビジネス課題にも取り組めるようになりました。

2012 年、インテル IT 部門は、インテルの各ビジネスグループとの緊密な連携の下、ビッグデータ・プラットフォームを導入し、社内向けのビジネス・インテリジェンスの提供においてビッグデータ分析がいかに有効であるかを示す概念実証を実施しました。

2012 年のインテルのビッグデータ・プロジェクトは、以下の内容で構成されます。

- マルウェアの検出
- 半導体設計の検証
- マーケット・インテリジェンス
- レコメンデーション・システム

インテルはまだビッグデータによるビジネス・インテリジェンスの向上を目指す能力開発の初期段階にあります。しかし今後は、研究開発、サイバー・セキュリティ、設計、製造、運用、市場開発、人事管理の各分野で、これらの機能の急速な成長が予想されます。

### Moty Fania

インテル IT 部門  
アドバンスド・ビジネス・  
インテリジェンス・ソリューション

### John David Miller

インテル IT ラボ  
プリンシパル・エンジニア

## 目次

概要	1
ビジネス課題	2
大量のデータから 実践的な知見を得るには	2
開拓者の必要性	3
ソリューション	3
MPP データ・マネジメント・ システム・プラットフォーム	3
Hadoop*	3
ハイブリッド・プラットフォームの 利点	4
ビッグデータのスキルセットと ノウハウの開発	4
概念実証	5
マルウェアの検出	5
半導体設計の検証	6
マーケット・インテリジェンス	6
レコメンデーション・システム	6
まとめ	7
詳細情報	7
協力者	8
略語	8

## IT@Intel

IT@Intel は IT プロフェッショナル、マネージャー、エグゼクティブが、インテル IT 部門のスタッフや数多くの業界 IT リーダーを通じ、今日の困難な IT 課題に対して成果を発揮してきたツール、手法、戦略、ベスト・プラクティスについて詳しく知るための情報源です。詳細については、<http://www.intel.co.jp/itatintel/> を参照してください。あるいは御社担当のインテル社員までお問い合わせください。

## ビジネス課題

未加工データの量は世界中で急激に増大しています。その主な原因としては、ネットワーク接続機器、インターネット・サービス、ソーシャルメディア、カメラ、センサー、およびユーザーが作成したコンテンツの爆発的な増加が挙げられます。その上、文書、Web ページ、電子メールなど、企業データの約 90% は非構造化データです。非常に大量かつ複雑なデータは、一般的なデータベース・ソフトウェアの処理能力を超えています。こうした状況では、全く新しい分析手法が必要になります。

大量のデータから  
実践的な知見を得るには

McKinsey Global Institute によるレポート『Big data: The next frontier for innovation, competition, and productivity』によると、米国内の 17 業種のうち 15 業種で、1 社当たりの保存データ量は米国議会図書館の情報量を超えています。<sup>1</sup> 例えば、小売業界の最大手である Wal-Mart では、1 時間当たり 100 万件以上の顧客取引を処理し、推定で 2.5 ペタバイト以上のデータを格納しているデータベースにデータをインポートしています。この情報量は、米国議会図書館に収蔵されている全書籍の情報量の 167 倍に相当します。

ビッグデータの大半は、数十億件ものトランザクションや、インテルのような大企業が顧客、供給メーカー、運用について毎日記録している情報に由来します。ビッグデータは、以前は主にストレージ上の問題と考えられていましたが、現在では新たな戦略的資産として認識されつつあり、企業のビジネスのあらゆる面に関する実践的な知見を秘めた金脈とさえ見なされています。

インテルでは、ビッグデータのユースケースとして、以下の 2 つの基本的なカテゴリーを認識しています。

- **ビッグ・データベース:** 従来のリレーショナル・データベース・マネジメント・システム (RDBMS) では大規模すぎて処理できない構造化データを格納します。
- **ディープ・アナリティクス:** 正解が 1 つに決められない、複雑な問題の答えを探すために使われます。通常、このような答えはソースデータ内に直接記述されていません。ビッグデータ可視化 / 分析ツールは、有益な知見が得られるように、データの継続的な調整と抽象化を行います。

従来、大半の企業は、分析のためにデータを集約したり、あるいはサンプルを収集してデータの意味を推定することしかできませんでした。こうした状況は現在も続いています。Gartner は、「2015 年にかけて、Fortune 500 企業の 85% 以上は、ビッグデータの効果的な活用による競争力の強化に失敗する可能性がある」と予測しています。<sup>2</sup> 一方で、最先端を行く一部の企業はすでにビッグデータ分析機能を導入し、大きな成果を上げています。Gartner によれば、これらの企業は、ビッグデータの処理に適した新しいツールとスキルセットを使用したビジネス・インテリジェンス、データマイニング、ビジネス分析プラクティスへの移行を急ピッチで進めています。

MIT デジタル・ビジネス・リサーチ・センターの Eric Brynjolfsson 教授は、大手株式公開企業 179 社の調査を行い、「データに基づく意思決定」を利用している企業は、競合他社に比べて生産性と収益性が約 5% 高いことを発見しました。Brynjolfsson 教授は、「ビッグデータを有効活用できる企業は、そこから多くの利益を簡単に得ることができる」と結論で述べています。<sup>3</sup>

<sup>2</sup> 『Gartner Reveals Top Predictions for IT Organizations and Users for 2012 and Beyond』Gartner のプレスリリース、2011 年 12 月 1 日。 <http://www.gartner.com/it/page.jsp?id=1862714> (英語)

<sup>3</sup> MIT Sloan Experts のブログ: 「Commentary on today's business issues」2012 年 2 月 14 日。 <http://www.mitsloanexperts.com/2012/02/15/erik-brynjolfsson-on-big-data-a-revolution-in-decision-making-improves-productivity/> (英語)

<sup>1</sup> McKinsey Global Institute, 2011 年 5 月。 [http://www.mckinsey.com/insights/mgi/research/technology\\_and\\_innovation/big\\_data\\_the\\_next\\_frontier\\_for\\_innovation/](http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation/) (英語)

## 開拓者の必要性

今日の IT マネージャーは、ビッグデータ分析用のシステムを開発し、ビッグデータ分析に基づく企業のインテリジェントな意思決定を支援するという課題に直面しています。ビッグデータは新しい分野であるため、実践している人材はまだ少なく、ベスト・プラクティスも十分に確立されていません。こうした空白部分は、ビッグデータのトレーニングやコンサルティングなどのサービスを提供する企業が補うことになりそうです。

約 2 年前、インテル IT 部門は、ビッグデータの利用方法についての検討を始めました。ビッグデータへの取り組みは、高度な分析によるインテルのビジネス改革に向けた包括的なロードマップの主要な要素になります。

## ソリューション

**インテル IT 部門は、価値の高いビジネス問題を解決する上でのビッグデータの有効性を証明するため、ビッグデータに関する複数の概念実証の開発を進めています。**

2012 年、インテル IT 部門は、インテルの調査と業界の既知の最適手法に基づいて、以下のリソースを基盤とする、コスト効率に優れたハイブリッド方式の社内ビッグデータ・プラットフォームを導入しました。

- SQL ベースの超並列処理 (MPP) データ・ウェアハウス・アプライアンス
- 複数のコンピューター・クラスターによる大規模データセットの分散処理用の Hadoop\*

### MPP データ・マネジメント・システム・プラットフォーム

従来のビジネス分析ソリューションは、分析用ではなくオンライン・トランザクション処理用に設計されたデータ・ウェアハウス・システムを利用します。これらのシステムは、汎用データベース、サーバー、ストレージを使用して構築されますが、そのためのプラットフォームは、絶え間なく成長し変化する数テラバイト (TB) のデータの処理に特化した機能を備えていません。

一方、現在の MPP プラットフォームは、SQL プログラミングのみに限定されることなく、Java\*、C/C++、R などのプログラミング言語のサポートも珍しいことではありません。これらのプラットフォームに、充実した高度分析機能およびインデックスのデータマイニング機能を搭載すれば、既存の BI や SQL のスキルを活用する一方で、特定の目的に最適なプログラミング言語を使用して新しいスキルを習得できる柔軟性も得られます。

インテルが使用しているソリューションは、従来のシステムに比べて最大 100 倍高速に大量のデータを分析できる、非対称型超並列アーキテクチャーを備えたサードパーティー製データ・ウェアハウス・ソリューションです。このようなシステムは、現在、数多くのベンダーから提供されています。

インテルは、ソリューションの選択にあたって、次のような要因を検討しました。

- 価格当たりのパフォーマンスとキャパシティから見たコスト効率の高さ
- 簡素さと導入の迅速さ
- 数テラバイトから数ペタバイトの大容量データに対応する、ストレージとパフォーマンスの拡張性
- 内蔵された高度な分析機能と、オープンソースの R 統計コンピューティング言語の完全なサポート
- インテルの既存の BI エコシステムとの親和性または全体的な適合性
- エンタープライズ・エコシステムとの相互運用性

データ分析に特化して開発されたこのシステムは、独自規格に基づくデータ・フィルタリング技術と、インテル® Xeon® プロセッサー E7 ファミリー搭載ブレードとコモディティー・ディスク・ドライブを組み合わせたもので、低コストでメンテナンスもほとんど必要なく、高いビッグデータ処理性能を発揮します。このシステムは、ブレードの増設によって、そのパフォーマンスとキャパシティを拡張できるように設計されています。各ブレードは、データを並列でストリーミングする、近接した複数のディスクドライブに接続されます。これにより、別個のディスク・ストレージ・システムを

使用する従来のソリューションに比べて、アクセス・レイテンシーが大幅に削減されます。

ベンダーの選定プロセスでは、データ・ウェアハウス・アプライアンス・ベンダー 11 社の書類審査を行った後、ベンダー 5 社に対して提案依頼書 (RFP) を提示しました。そして、提案書の分析と技術評価に基づいて、最終的にベンダー 1 社を選定しました。

## Hadoop\*

Hadoop\* は、大量データ処理用のオープンソース・フレームワークです。Hadoop\* は、1 台の大型スーパーコンピューターの代わりに、クラスターとして動作する複数のサーバーのローカルストレージと演算処理を調整し、各サーバーにデータの一部を処理させます。

Hadoop\* は、Apache Software Foundation のトップレベルのオープンソース・プロジェクトです。市販ディストリビューションもさまざまな種類が流通しています。

Hadoop\* は、単独では、以下の 2 つの基本サービスを提供する分散コンピューティング OS です。

- **Hadoop\* 分散ファイルシステム:** この分散ファイルシステムは、Hadoop\* クラスター内のすべてのノードに分散される、UNIX\* に似たファイル・システム・ストレージを提供します。Hadoop\* は他のファイルシステムも使用できます。
- **MapReduce:** この分散型演算処理機能は、Hadoop\* の基盤となるものです。MapReduce は、クラスター内の各サーバーを調整し、全体的な処理タスクの一部を並列処理させます。

このコアの上に、次のような多数の市販アプリケーションとオープンソース・アプリケーション、ツールキット、データレイヤーが実装されます。

- **Hive:** Hadoop\* データのクエリー用の SQL 言語
- **HBase:** 高速、読み出し / 書き込み、列指向のデータベースで、数十億行 × 数百万列を処理できます。

- **Pig** : データ処理用のインタラクティブ・スクリプト環境
- **Mahout** : クラスタリング、協調フィルタリング、類似性認識用のアルゴリズムを提供する機械学習ライブラリー
- **Sqoop** : RDBMS データベースとのインポート / エクスポートのやり取り
- **Oozie** : 複雑なデータ処理操作を調整するためのワークフロー環境
- **Cassandra** : ドキュメント指向データベース

Hadoop\* にはリニアな拡張性があります。例えば、クラスター内のマシン数を 2 倍にすると、処理時間が約 2 分の 1 に短縮され、同じ時間で 2 倍のデータ量を処理できるようになります。

Hadoop\* は Java\* で作成され、Linux\* 上で動作します。Hadoop\* アプリケーションも通常は Java\* で作成されますが、他の言語で作成することも可能です。Hive や Pig などの一部の Hadoop\* ツールは、クライアント・コンピューター上で実行され、直ちに MapReduce プログラムを生成します。

Hadoop\* は Hadoop\* クラスター内のすべてのサーバーのストレージを集約し、これらのサーバーはコモディティー・ハードディスク・ドライブを使用できます。したがって、スト

レージ容量 1 テラバイト当たりのコストは非常に低く、クラスター内のストレージの増設によってペタバイト単位のデータに対応できます。このように、Hadoop\* は、従来は廃棄されていたデータを、コスト効率良く収集し、保存します。また、Hadoop\* により、まだ十分に理解されていないものの、価値がありそうなデータの収集と保管も可能になります。テキスト分析などの分野では、使用するアルゴリズムが単純なものであっても、(データ量が少ないよりも) データ量が多いほど優れた結果が得られることが実証されています。サイバー・セキュリティなどの分野では、Hadoop\* の大量データ処理能力により、長期間にわたる分析が可能になります。

一般的に、Hadoop\* とそれに関連するテクノロジーは、オンライン・トランザクション処理システムや従来の RDBMS に代わるものではありません。Hadoop\* の一番の強みは、テラバイト / ペタバイト単位のデータのバッチ処理にあります。

### ハイブリッド・プラットフォームの利点

インテル IT 部門では、非対称超並列アーキテクチャー・コンポーネントを搭載したサードパーティー製データ・ウェアハウス・アプリケーションと Hadoop\* を組み合わせることで(図 1 を参照)、コスト効率に優れ、拡張性が高く、各コンポーネントの強みを活用できるビッグデータ・プラットフォームを構築しました。高速ネットワーク接続と高速データ

ローダーを使用してコンポーネントを配置、接続したことで、ビッグデータ・プラットフォームは、必要に応じてデータの各部分をプラットフォーム間で効率的に移動できます。

### ビッグデータのスキルセットとノウハウの開発

ビッグデータ分析の大きな問題の 1 つは、熟練した専門家が不足していることです。先に引用した McKinsey Global Institute のレポートによると、2018 年までに米国内だけでディープ・アナリティクスのスキルを持つ技術者が 14 万人 ~ 19 万人不足すると予想されています。また、ビッグデータ分析を使用した効果的な意思決定の方法を理解しているマネージャーとアナリストは、150 万人も不足する見通しです。

統計学、数学、機械学習、ビジュアル・アナリティクスなどのビッグデータ科学のスキルはもちろん不可欠ですが、データとビジネスとを関連付けて、分析結果をビジネスの成果に結び付けるスキルも重要です。つまり、インテルの各ビジネスグループのような IT 部門の顧客がビッグデータから利益を得るには、社内におけるビッグデータ利用スキルを開発する必要があります。

Hadoop\* などのビッグデータ技術の多くはオープンソースであり、大量の構造化データと非構造化データをコスト効率よく処理する目的で、インターネット企業によって開発さ

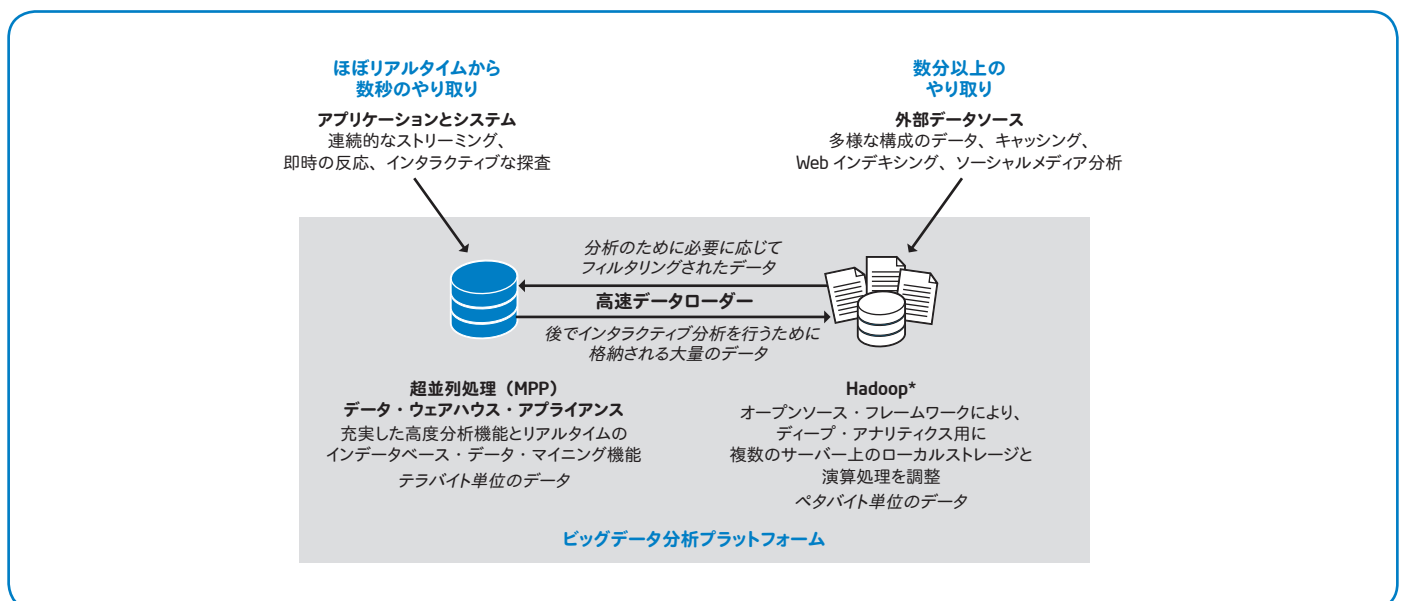


図 1. インテル IT 部門のビジネス・インテリジェンス・ビッグデータ・プラットフォームは、超並列処理 (MPP) データ・ウェアハウス・アプリケーションと、Apache Hadoop\* を実行する業界標準規格に基づくサーバークラスターの組み合わせを基盤としています。

れたものです。これらのテクノロジーは急速に成熟していますが、現時点では、Linux\*、Java\* 開発環境、分散コンピューティングなど複数の分野における高度な技術スキルが必要です。ビッグデータ・テクノロジーを導入する企業は、こうしたさまざまなスキルの開発にも取り組む必要があります。

### 知識とスキルの不足を補う方策

新しいスキルを獲得することは、テクノロジーを導入することより難しい場合もあります。インテル IT 部門とインテルの各ビジネスグループは、調査とトレーニング、ビッグデータを扱うのに不可欠なスキルセットを持つ人材の獲得、この資料で報告しているようなテストや概念実証 (PoC) の実施、特定のユースケースの導入を通して、ビッグデータに関する専門知識と経験の不足を補おうと努めています。

直接の経験が不足している状況では、実際にデータを処理し、ビッグデータ分析を実行することを通じて、経験を積んでいくことになります。インテルでは、予測的および処方的な BI の実現にビッグデータを利用できるようになれば、最終的には、ハードウェア、ソフトウェア、トレーニング、時間への投資に十分に合う利益が得られると考えています。

## 概念実証

**インテルは現在、ビッグデータに関する多くの概念実証を進めています。ここでは代表的な 4 つの事例を紹介します。**

### マルウェアの検出

サイバー犯罪の脅威は、攻撃者とツールの高度化とともに増大し続けています。セキュリティ監視機能の目標は、ユーザーやセキュリティ対応スタッフが対策をとれる時間内に脅威を発見することです。

従来のマルウェア対策の最も一般的な手法であるウイルス定義に基づくファイルスキャンは、作成されるマルウェアの数の急増によって有効性が低下しつつあります。より良い手法は、常にマルウェアの一步先を行き、マルウェアの動作とマルウェアの発信元を深いレベルで分析し、将来マルウェアが発生する場所まで予測することです。このような高度で詳細な監視と予測を行うには、サーバーの動作を常に観察し、システム、ネットワーク、アプリケーションの各レベルに異常がないかをチェックする必要があります。

このような脅威を示すパターンは、多くの場合、プロキシ、DNS、DHCP、VPN などの各種のネットワーク・ログやサーバーログの

中に隠れていますが、これらのログには大量のデータが含まれています。異常の現れ方は、一般的なマルウェア・シグネチャーから、悪質であることが判明している URL との通信、異常な検索など、疑わしい活動と動作の幅広いグローバルなパターンにいたるまで、あらゆる形態にわたります。そして、この分析には一連の複雑な手順が必要です。異常なアクティビティを検出するためには、多くのソースから得られるデータを相互に関連付けたり、正常なネットワーク・アクティビティとパターンのベースラインを定義するなどの手順を実行する必要があります。

インテルは、決められたタイムスケールに従ってこれらの異常を特定するため、ビッグデータ・テクノロジーを使用して、未加工の非構造化データを収集してそれを構造化した後、予測分析などの統計モデルを使用して異常な活動パターンを検出します。

この PoC では、マルウェアを迅速に特定し、隔離できるように、異常な動作をリアルタイムで特定することを目指しています。数カ月あるいは数年にわたって蓄積されたデータを収集して分析できれば、セキュリティ侵害の発生源と性質の予測精度を向上し、より有効な予防手段と予防システムの導入が可能になります。

## ビッグデータの課題とメリット

「ビッグデータ」とは、量が膨大であったりソースや種類が多様化したために、従来のリレーショナル・データベース手法では処理することが難しい情報に関する問題全般を指す、非常に広範な意味を持った用語です。テキスト、音声、動画、クリックストリーム、ログファイルなどのデータの多様性は、ビッグデータの重要な特徴であり、このために数テラバイト程度の非構造化データをビッグデータと呼ぶこともあります。

ビッグデータ・ツールを初めて開発したのは、Google や Yahoo などのインターネット検索企業であり、彼らは Web インデキシングのためにこれらのツールを必要としていました。まもなく他のインターネット企業が後に続き、注文やリコメンデーション (推奨)、Facebook タイプのメッセージなどのインターネット・スケールの問題を処理するために、その他のコンポーネントが開発されていきました。現在では、大企業の IT 部門がこれらのツールを使用して、これまで分析や解決が難しかった、価値の高いビジネス問題に取り組んでいます。

ただし、これらの新しい技法の導入には、さまざまな困難を伴います。これらの新しいツールは、統合、導入、保守に多くの課題を抱えています。大半のツールは依然として成熟途上であり、Linux\* や Java\* など、最新の IT スキルを必要とします。ビッグデータ・ソリューションを開発し、最適化するには、MapReduce などの並列コンピューティング構造の観点から問題を再検討する必要があります。問題によっては、Web インデキシングのような高度な並列処理ができないものもあります。さらに、インターネット・スケールの問題を解決するには、データベース・プログラマーが利用してきた ACID セマンティクス (A : 原子性、C : 一貫性、I : 独立性、D : 永続性) を緩和し、低レイテンシーの代償として高スループットを得る必要があります。

現行のデータシステムでも十分に解決可能な問題については、あえてシステムを切り替える必要はありません。しかし、既存のシステムの能力を超えている問題については、ビッグデータ・ソリューションが解決策になるでしょう。明確なユースケースがまだ定義されていない領域でも、企業は低コストのビッグデータ・ストレージを使用して、収集可能なほぼすべてのデータを収集、格納し、後から潜在的な価値を抽出することができます。

## 半導体設計の検証

半導体設計では、設計をシリコンに実装する前に、広範囲にわたるテストを行う必要があります。このテストはシリコン実装の各ステージに入っても続けられ、数百個のセンサーが1秒当たり数千回のサンプリング・レートでデータを収集します。こうした広範囲にわたるテストの結果として、大量のデータが生成されます。

このPoCでは、Intel IT部門は、ビッグデータ・プラットフォームを使用して検証プロセスを最適化する方法を調査します。数十億行の構造化データと非構造化データを分析して、設計プロセスの迅速化と量産までの期間短縮を支援し、迅速な市場投入を可能にします。

この利用モデルの一例として、カバレッジと呼ばれる概念が挙げられます。ポストシリコン検証の世界では、チップをリリースするタイミングに関する明確なルールはありません。バグのあるチップをリリースすると、企業の評価に大きなダメージを与えかねない一方で、必要以上にテストを繰り返していると、チップの発売が遅れ、数百万ドルもの売上を失うことにもなりかねません。カバレッジの概念は、こうした極端な状況の回避を目指すものです。プロセッサがテストされた（カバーされた）論理的状態と物理的状态に関するデータを収集することで、テストとテストツールがどのように機能しているかを理解し、チップを市場に投入できるかを判断できます。

また、ビッグデータ分析により、特定された欠陥を自動的にクラスター化して並べ替え、大量のヒストリカル・テストに基づく根本原因分析を実行することで、デバッグプロセスを支援できます。収集された大量のデータ（サンプルではなく）の広範囲にわたる分析を通して、各ステージの進捗状況を包括的に把握することができ、設計プロセスの改善と合理化の方法を見つけて、最終的に製品そのものの改善につなげることができます。

## マーケット・インテリジェンス

Intelのように世界的な販売網とグローバルなサプライチェーンを持つ企業にとって、刻々と変化する市場環境を予見し、翌月、6カ月後、5～10年後に何が起きそうかを正しく予測することは非常に重要です。グローバル企業は、気象トレンド、世界経済データ、ディスカッション・フォーラム、ニュースサイト、ソーシャル・ネットワーク、Wiki、ツイート、ブログなどの大量のデータの中から、必要なデータを選別しなければなりません。こうしたデータに基づいて、正確な予測の作成、販売戦略の策定、競合他社の脅威の評価、消費者行動の変化の予想、サプライチェーンの強化、ビジネス継続性計画の改善が可能になります。

このPoCでは、Intel IT部門は、Intelの各ビジネスグループと協力して、さまざまなソースから得られるデータを分析し、以下の成果を達成する見通しです。

- 世界のさまざまな地域の売上予測を改善し、生産レベルを微調整し、より正確な予測を株主に提供する。
- 起こり得る世界規模のイベントに基づくシナリオの作成とテストにより、Intelの市場、サプライチェーン、市場の要求や競合他社の挑戦への対応力に与える影響を評価する。
- Intel製品の新しいユーザーと新しい使い方を発見する。

## レコメンデーション・システム

コンテンツの量が急激に増大した場合、ユーザーが調査対象や自分の関心に合った情報を効率良く見つけるためには、何らかの手助けが必要となります。こうした理由から、社内アプリケーションと外部のアプリケーションのいずれについても、Intel全体でレコメンデーション・ベース・サービスへの需要が増大しています。AmazonやNetflixが

顧客に提供しているシステムと同様のレコメンデーション・システムが、検索やナビゲーションの時間を短縮し、パーソナライズされた、ターゲットを絞った結果を提供することで、ユーザーを支援します。これにより、生産性、信頼性、全体的なユーザー体験の質が向上します。

拡張性の高いレコメンデーション・システムを実現するには、予測分析とビッグデータ分析に関する専門知識が必要です。これは、大量の履歴データに対して、大量のリソースを使う複雑なアルゴリズムを実行する必要があります。

このPoCの中核は、ビッグデータ・プラットフォームの上に2レイヤー（オフラインおよびオンライン）アーキテクチャーを実装した、再利用可能な汎用レコメンデーション・エンジンの構築にあります。オフライン・コンポーネントは、レコメンデーション・アルゴリズムのコアを実行するバッチ指向のプロセスです。このコンポーネントは、拡張性の高い環境でビッグデータ処理を実行することで、モデルをスケールアップできることを保証します。オンライン・コンポーネントは、任意のサービス要求に対するサービスレイヤーとして機能します。このコンポーネントは、オフラインフェーズで計算された関連する中間計算をロードし、アルゴリズムの最後のステップを実行してレコメンデーションそのものを生成します。また、このコンポーネントは、コンテキスト構成済みのロジックを適用して、要求のコンテキストに従って最終的なレコメンデーションをフィルタリングし、調整します。

このソリューションの拡張性は、Mahoutを使用してアルゴリズムのコアを実装することによって得られます。Mahoutは、Hadoop\*上で動作する、Java\*で作成されたオープンソースのデータ・マイニング・ライブラリです。Mahoutは、シェアードナッシング環境内のコモディティ・ハードウェアのクラスターの内側で並列ジョブを実行することにより、Hadoop\*アーキテクチャーを利用します。すべての中間結果は、オンライン・コン

ポーネントが高速で検索できるように、MPP RDBMS に書き込まれます。

このレコメンデーション・サービスの導入は、パーソナライズされたコンテンツをジャストインタイムで提供できるようにする上で重要な役割を果たします。これにより、社員がインテルの社内アプリケーションを利用する際の生産性は向上します。また、競争力が向上し、社外の顧客がインテル製品を選びやすくなり、インテルの収益に貢献します。インテル IT 部門は、大量のデータに基づく複雑な予測分析の提供から得られる経験と知識を応用して、将来も同様のソリューションを提供していきます。

---

## まとめ

インテル IT 部門は、2012 年、体系的な手法に基づいてビッグデータ分析を全社的な BI の取り組みに統合するため、複数の概念実証を開始しました。ビッグデータのマイニングと分析の機能を追加することで、インテルの BI 機能は記述的分析から予測的および処方的な分析へと進化し、ビジネスのパターンとトレンドに関する深く豊かな知見をもたらすことが期待されます。

このプロジェクトの最初のステップはすでに完了しています。このステップでは、サードパーティー製データ・ウェアハウス・アプライアンスと Hadoop\*（複数のサーバー上での大量のデータ処理を可能にするオープンソース・フレームワーク）を組み合わせたビッグデータ・プラットフォームの設計と構築を行いました。このソリューションは、構造化データの超並列処理と、業界標準規格に基づくサーバー上での大規模データセットの分散処理を実行できます。さらにインテルでは、IT 部門の BI スタッフと各ビジネスグループとの間で、ビッグデータ分析に必要なスキルセット、専門知識、熟練度を高める取り組みも継続して行っています。

これらの概念実証が成功した段階で、ビッグデータ・プラットフォームを実稼動環境に移行する予定です。そして、実際に価値の高いビジネス問題の解決に利用して、これまで以上の運用効率を達成するとともに、収益の増大と新たな収益源の獲得に結び付けていきます。数年後には、インテルのビッグデータ分析プログラムはさらに成長し、高度な BI が実現されて、製造、セキュリティ、マーケティング、市場開発、IT の各分野に新たな競争力をもたらすことが期待されます。

---

## 詳細情報

関連トピックのホワイトペーパーについては、<http://www.intel.co.jp/itatintel/> を参照してください。

- 『Roadmap for Transforming Intel's Business with Advanced Analytics』

## 協力者

**Jessica Brindle**

インテル IT 部門  
ビジネス・インテリジェンス  
戦略プランナー

## 略語

<b>BI</b>	ビジネス・インテリジェンス
<b>MPP</b>	超並列処理
<b>PB</b>	ペタバイト
<b>PoC</b>	概念実証
<b>RDBMS</b>	リレーショナル・ データベース・マネジメント・ システム
<b>RFP</b>	提案依頼書
<b>TB</b>	テラバイト

インテル IT 部門のベスト・プラクティスの詳細については、  
<http://www.intel.co.jp/itatintel/> を参照してください。

この文書は情報提供のみを目的としています。この文書は現状のまま提供され、いかなる保証もいたしません。ここにいう保証には、商品適格性、他者の権利の非侵害性、特定目的への適合性、また、あらゆる提案書、仕様書、見本から生じる保証を含みますが、これらに限定されるものではありません。インテルはこの仕様の情報の使用に関する特許、著作権、知的財産権の侵害を含む、いかなる責任も負いません。また、明示されているか否かにかかわらず、また禁反言によるとよらずにかかわらず、いかなる知的財産権のライセンスも許諾するものではありません。

Intel、インテル、Intel ロゴ、Xeon は、アメリカ合衆国および/またはその他の国における Intel Corporation の商標です。

\* その他の社名、製品名などは、一般に各社の表示、商標または登録商標です。

インテル株式会社

〒100-0005 東京都千代田区丸の内 3-1-1  
<http://www.intel.co.jp/>

©2012 Intel Corporation. 無断での引用、転載を禁じます。  
2012年11月

327474-001JA  
JPN/1211/PDF/SE/IT/TC

