

特集

ビッグデータ： 世界を変えていく イノベーションの 原動力として

編集にあたって

石井一夫 (東京農工大学)

水田正弘 (北海道大学)

中野美由紀 (芝浦工業大学)

ビッグデータのうねりは、日本国内でも近年急速に認知されてきた。これは、政策面では2012年のObama大統領の「ビッグデータ研究開発イニシアチブ」をきっかけに、技術面では分散ファイルシステムや分散処理システムを実装するHadoopプラットフォームや、アマゾンウェブサービス(AWS)などのクラウドコンピューティングの活用促進がブレイクスルーとなっている。日々のメディアでも頻繁に取り上げられ、バズワードと語られることも多い。むしろ今となつては古くさい響きすらも感じる。

しかしながら、その利活用促進が企業の業務成績向上のキーとなっているのみならず、世界を変えていくイノベーションの原動力となるポテンシャルを秘めていることは否定できない。

ビッグデータの草分けといえ、1998年に設立されたGoogle社(以後、Google)のSergey BrinやLarry PageによるBigTableや、そこから発展してきたGoogle File SystemやMapReduceであるが、これらはまさに世界を変えたイノベーションというにふ

さわしいものである。ビッグデータ処理やビッグデータ分析が今日に至る話題を提供できるようになるまで長期間にわたりいくつもの経過を経ている。

その意味で、ビッグデータ利活用は単なる流行という次元で語られるものではなく、イノベーションを起こす一大ムーブメントと捉え、自らのパフォーマンスに取り込んで、世界を変えていく原動力として、自分のものとして身につけていくべきものであると考える。事実、Hadoopにしる、AWSにしる、その進化はとどまることなく今も続いている。新たなスキームは今も登場し続けている。その意味で、第2のSergey Brinになれるチャンスの扉はすべての研究者、エンジニアに開かれていると思いたい。まだ、冷めてしまうには早すぎる。

ビッグデータの利活用の今後の流れは、利活用がなされた各分野でどのようなイノベーションにつなげ世界を変えていくか、またビッグデータの利活用において、個々の要素技術(Hadoopやクラスタ、クラウド、ハイパフォーマンスコンピューティング(HPC)、統計学、データマイニング、機械学習など)が増え続けるデータに対してどのように進化、展開していくかということに話題がシフトしていくと思われる。

本特集は、このようなビッグデータの利活用の現状に関して、ITフォーラム「ビッグデータ活用実務フォーラム」のメンバによる執筆および編集により実現したものである。「ビッグデータ活用実務フォーラム」は、

2013年に設置され本会のなかでも最も新しいITフォーラムであるが、ソフトウェアジャパン2013などにおけるイベントなどを通じてビッグデータに関する情報交換、情報共有、情報発信を実施している。本特集もそのアクションの一環によるものである。このような背景をふまえてビッグデータの利活用に関する話題を提供するために以下のような構成で特集を組んでいる。

まず、ビッグデータの利活用事例における利用者サイドの情報提供として、マーケティング分野および医療分野におけるビッグデータの利活用に関する話題を取り上げる。

次にビッグデータプラットフォームの供給者サイドの情報提供として、AWSのクラウド環境によるビッグデータの利活用、Hadoopに関するオープンソースのディストリビューションをリリースしているCloudera社(以後、Cloudera)のソフトウェアツールの最新の話題を取り上げる。最後に、シェルスクリプトを活用したユニークなビッグデータプラットフォームに関する話題を取り上げる。

以下、本特集の各記事に関して概要を説明する。

本特集1「ビッグデータ時代の次世代ダッシュボード」(池内)は、マーケティング分野のビッグデータの活用に関する話題である。多様な大量データを用い、高い頻度かつ柔軟な軸でKPI(重要業績評価指標: Key Performance Indicator)のモニタリングを行うために、マーケティング分野におけるダッシュボードがどのように進化してきたか、ビッグデータを取り巻く環境とダッシュボードという軸で解説している。

本特集2「医療におけるビッグデータ利活用」(石井)は、医療分野のビッグデータの活用に関する話題である。次世代シーケンサーの実用化により、ゲノム配列情報の大量処理が日常的となっており、臨床情報とゲノム情報を結びつけた医療ビッグデータの利活用の促進が期待される。これら医療分野、特に、ビッグデータ利活用の現状を医療と技術の両面から概説し、特に、原因が不明でエビデンスに基づく有効な診断法の存在しない精神神経系疾患への展開について紹介している。

本特集3「クラウドサービスを用いたビッグデータ

活用事例」(吉荒)は、AWSにおけるクラウド環境を用いたビッグデータ分析の事例の紹介である。大量データを収集、保管、分析し共有するために必要な情報システム基盤には、柔軟性と拡張性にすぐれ、必要なときに必要なだけ迅速に利用できるパブリッククラウドサービスが大きな力を発揮する。AWSは、2006年のサービス開始以来、パブリッククラウドサービスを提供しているが、これらパブリッククラウドを利用したビッグデータ分析に関するさまざまな事例を紹介する。

本特集4「Hadoopによるビッグデータプラットフォームを用いた情報駆動型企業へのアプローチ」(水丸)は、ビッグデータ活用のプラットフォームとしてのHadoopに関する話題である。ビッグデータとの共生が前提となるビジネス環境において、熾烈化する企業競争を勝ち抜くためには、「情報駆動型企業」への転換が不可欠と考えている。そのため、「すべてのデータを1つの器(うつわ)に格納し、そこからさまざまな形で利用する」アプローチが有効と思われる。著者らおよび著者の勤務するClouderaは、この器となるプラットフォームとしてHadoopを中心に位置づけ、多様なデータの扱いをより統合化し、情報駆動型企業への転換を支援するための解決策として「エンタープライズデータハブ」を提唱しており、これらの取り組みを紹介している。

本特集5「シェルスクリプトを用いたビッグデータ活用の事例」(當仲)は、シェルスクリプトを用いたビッグデータ処理の話題である。ビッグデータ処理といえは、Hadoopを用いた分散ファイルシステムや分散処理が頭に浮かぶが、一方でUNIX系OSの基本機能であるシェルスクリプトを用いてテキストデータを処理し、コンピュータの性能を最大限に活かすプログラミング手法にも注目が集まっている。このような、ビッグデータ処理の事例として「ユニークな開発手法」やシェルスクリプトによる並列処理、シェルスクリプトで動作する統計コマンド群などについて紹介している。

本特集を、ビッグデータを巡る現状の理解の一助として、お役立ていただければ幸いです。

(2014年6月23日)