

特集「データエンジニアリング」の発刊に寄せて

小川 豊

我が国で流通している情報量が年々増加を続けている。そのなかでもインターネットからの情報やセンサーからの情報などが特に増加している。総務省が2009年7月に公開した「我が国の情報流通量の指標体系と計量手法に関する報告書」では、アナログデータも含めた2007年の情報流通量は6ゼタビット（ 10^{21} ビット）、そのうち消費情報量は300ペタビット（ 10^{15} ビット）と推計している。また、2001年に比べて情報流通量は1.55倍、消費情報量は1.04倍に増加していると報告している。大きな割合を占める放送メディアを除くと情報流通量の43.4%、消費情報量の39.2%はインターネットによるもので、2001年と比べ35.7倍と放送メディアの1.6倍と比較して非常に大幅な伸びを示している。

このような急激な情報量の増加を背景として、大規模・大量データを蓄積・整理・活用する技術が新たに発展してきている。学術分野では、これら「データエンジニアリング」技術の研究が継続的に行われているが、産業界でも、「データエンジニアリング」向けの技術が開発されている。特にインターネットビジネス分野においては、Googleに代表される大規模Webシステムを運用するための高速・大容量データベース技術の開発が進んでおり、新たな形態のデータベースが出現した。また、大量データをビジネスに利用する動きも活発になってきている。

GoogleやYahoo!、Amazonは、検索連動型広告や商品レコメンデーションなどで利用するためにWeb上のユーザ行動履歴を大量に蓄積し、それを連携させて新たな情報を作り出す技術を生み出している。このような動きはWebの世界だけでなく、実世界においても見られる。スーパーにおけるPOSデータの活用、通信販売の受注センターにおける会話データの分析、IC乗車券などを用いたユーザ行動分析など、様々な企業が大量に蓄積したデータをより高度に利用し、そこから得られた知見をビジネスに活用している。

日本ユニシスでは、このような大規模化・多様化するデータを蓄積・整理・活用する「データエンジニアリング」技術と製品を提供してきた。2000年には大規模・大容量データウェアハウスを構築するエンタープライズサーバー「ES7000」を提供した。加えて2006年からはデータウェアハウスアプライアンス「Netezza」を提供することにより、更に大容量のデータウェアハウスの構築を支えてきた。また、支援ツール「MartSolution」や各種BI（Business Intelligence）ソフトウェアを使ったデータウェアハウス構築支援サービスを展開し、そこに蓄積したデータをより高度に利用するための製品やサービスも提供している。データマイニングソリューション「MiningPro21」は、小売や通販業界では優良顧客のセグメント分析・商品レコメンデーション・顧客数予測に、金融業界ではカードの解約者の事前発見・与信分析などに利用されている。

冒頭で述べたように、経済活動や社会活動などのあらゆる活動があらゆる場所で「データ」を発信するようになってきている。生産活動で発生する「データ」だけではなく、消費者の活

動から発生する「データ」も含め、増加し続ける「データ」を上手に管理する必要性が高まってきた。また、社会、ビジネスの関連性が複雑化しステークホルダーが多様化してきているが、そのような多様な関連を「データ」でつなぐ必要性が高まってきた。これまでの固定した価値の追求というニーズから、変化過程の解明と制御に多くのニーズが変化してきており、「データ」での追跡の必要性が高まってきた。

日本ユニシスでは、「情報爆発時代」とも称される現在に対応しながらビジネスの変化を進めている顧客企業に対し、今までにもまして効果的な IT サービスを提供するために「データエンジニアリング」技術分野に先進的に取り組んでいる。今回の「データエンジニアリング」特集では、日本ユニシスがこれまで技術と製品の適用を通じて蓄積してきた知見について事例を中心に述べ、更に大規模化・多様化・複雑化したデータに対して新たに取り組んでいる技術について紹介する。

本特集が、大規模・大量データの必要性・有効性を認識し、その蓄積・整理・活用を企業活動の中で積極的に行おうとしている方々への一助となれば幸いである。

(執行役員 総合技術研究所長)