

## 特集「データエンジニアリングⅡ」の発刊に寄せて

羽 田 昭 裕

2009年に発刊した技報101号では、大規模化・多様化・複雑化するデータの集合（以下、〈データセット〉と呼ぶ）を蓄積・整理・活用する技術や製品を「データエンジニアリング」と総称して特集した。その背景には、データベースマネジメントにとっての大きな変化が念頭にあった。資材管理から始まったデータベースマネジメントは、ある品質の〈データセット〉を、必要とするときに必要な量だけ適正なコストで取り揃え、要求元へタイムリーに供給するための管理活動である。情報処理技術は、この〈データセット〉に基づいた、記号論理などの人工言語、統計的方法などの大量データの蓄積に基づく分析、巨大システム管理アプローチなどの意思決定手法、シミュレーションなど仮想空間を利用した認知などを通じて、ビジネス領域ごとに組織が統合している知と知的活動を支えてきた。しかし現在では、人間の活動はより広い範囲でデジタルデータ化され、その規模、多様性、複雑性は、この枠組みを超えるものとなりつつある。例えば、インターネットで扱われるデータは音声、静止画、動画と多様になり、CISCO VNI, 2011によれば、2015年には、全世界のIPトラフィックは年間966エクサバイト( $966 \times 10^{18}$  byte)となり、その8割以上はコンシューマインターネットトラフィックとなり、ビジネスインターネットトラフィックはIP WANを上回ると予想している。

この転回はデータベースマネジメントの新たな様式を生み出している。そして、2009年頃に出来た〈ビッグデータ〉がITキーワードの一つとなる中、101号発刊以後の進展を特集することは時宜に適うと考え、「データエンジニアリング」特集の続編を発刊した。

〈ビッグデータ〉に関連する議論の中では、既存のソフトウェアだけではデータベースマネジメントが実現できなくなっているという見方が多い。例えば、2011年5月に発表されたマツケンゼー・グローバル・インスティテュートのレポート「Big data」では“〈データセット〉であり、その全体の大きさが典型的なデータベースソフトウェアを道具として記録・格納・管理・分析できる量を超えたもの”としている。そして、この報告に限らず、〈ビッグデータ〉に関しては概念的な定義を避けるのが通例になっているようだ。それはデータに関連する領域で、従来からの連続性よりも、断続性に注目するほうが有効なさまざまな現象があり、一つの新しい言葉で表現しようという立場からの用語法なのであろう。一方で、この断続性と連続性から、さまざまな期待と懐疑が生じている。

連続性とは、電磁氣的に記録されたものという意味でのデータは、永続性、共有性、そして複製性などで特徴づけられ、結果としてネットワークを経由したデータ交換の自由度が飛躍的に高まる点にある。断続性のひとつは、情報処理の対象としての〈データセット〉の量、速度、多様性や複雑性の異質さであり、これは1980年代に生み出された分散処理技術を応用した非構造化大量データの並列処理に基づいている。もうひとつの断続性は新しい利用の仕方である。共同利用を志向する全体的な計画としてのスキーマに即して格納された、首尾一貫した

〈データセット〉を用いるデータベースと異なり、〈ビッグデータ〉では生成された文脈から切り離され、一貫性に制約されることなく、集められた断片的な〈データセット〉をそのときどきの状況的な目的に応じて用いるという探索的な利用となる。この利用方法は、Bill Inmon が1990年代に提唱したデータウェアハウスの中核的な概念であるサブジェクト指向と類似している。異なる点は、〈ビッグデータ〉の発生源が組織の外部にあるため、一元管理できないという点であろう。逆に、組織外部の大量データを利用した深い分析が可能になることから、膨大なアイデアを市場における価値や健全性の観点で評価するための解析やフィードバックに多大なコストと時間を必要としていた研究、デザイン、市場調査などの業務にとっては大きな助けとなる期待がある。

日本ユニシスは、エンタープライズ・システムに長い経験と責務のある立場から、この従来とは異質なデータの価値を活かしながらも、組織の〈データセット〉は管理可能なものになっている必要があると捉えている。それに対応するデータエンジニアリングの対象分野は広大であるが、本特集号では、次のような分野に注目して技術報告をまとめた。

まず、大規模かつ多様で複雑な〈データセット〉を統合する試みの状況を概観し、事例を示す。事例のひとつは今後の企業システムでの検索基盤を提供する技術の解説であり、もうひとつは第四期科学技術基本計画でも取り上げられているライフイノベーションに関する実証実験である。また、〈データセット〉が統合化するにつれて、生まれつつある新たなセキュリティニーズについて考察した。

次に、モデル化やルール化などデータの抽象化をめぐる問題を取り上げている。データ交換の利用形態が高度化することによって、システムが複雑性を増し運用コストを引き上げる結果を招く可能性がある。これを極力抑える方策としてのデータモデルパターンを活用したモデル化についての考察と、交換に用いられるビジネスロジックをデータ化することで安定した品質を目指すルールエンジンを実証した報告を示す。

最後に、データ統合環境整備や抽象化技術の実用化が進むにつれて発生する複合的な問題への、情報システムとしての対応に触れる。第一に、データウェアハウスを中心とした企業情報システムに関連する技術動向と活用状況、対象範囲の拡大について俯瞰する。そして、データベースマネジメントシステムの原理、歴史的背景から、データの未来環境が完備した段階で情報システムはどのように変貌し、何をもたらすかについて考察する。

日本ユニシスは、データエンジニアリングに課題解決の道を見出されたお客様とともに価値の創造が実現できるよう、技術の研鑽とソリューション開発を積極的に行っている。さらに、パートナーとともに先端分野における実証実験にも挑戦している。これらの詳細は、本特集号の各論文を読んでいただければ理解していただけるだろう。

現在の〈ビッグデータ〉に関連する技術は、関係データベースマネジメントシステムに比べて利用のしやすさという点では退行した感がある。1990年代に登場したときはダム端末に戻ったと言われた Web ブラウザが、現在では HTML5 にまで進化していることを念頭に置くと、非構造化データについても新しい技術が普及する可能性はある。しかし、実用化はまだ時間がかかりそうだ。また、事例として取り上げた医療分野で見られるように、データ中心の知識を作り上げていくには、データセットを共有することで、新しいイノベーションが生まれるよう、規制やデータの所有権などの問題を解決していく必要がある。データエンジニアリング技術や製品が、その隔たりを埋める役割を担い、データ交換の自由度を軸に顧客価値をもたらす

リビューションとなることを願っている。本特集号がデータエンジニアリングに関心のある方々の参考になれば幸いである。

(総合技術研究所長)