

特集「データエンジニアリングV」の発行に寄せて

羽田 昭裕

当社のルーツであり、世界で最初に稼働した汎用コンピュータである ENIAC は数理的なデータ活用が主な用途であった。一九四六年二月、該機のお披露目の日に、最初の国連総会が開かれた。それ以来、国境を越える社会の動きと、コンピュータの歴史は連動しているように見える。二十一世紀に入り、再資源化やリアルタイムマーケティングなど消費から生産への経路が注目され、循環的な経済を指向するようになり、最終消費者という概念が薄れ、社会も変わってきた。それと同時期に、数理的なデータ活用が再び計算機利用の主役となった。

そこで、本技報ではデータエンジニアリングという課題を設定して、二〇〇九年から特集を組んできた。一連の特集は、ソフトウェアエンジニアリングが、需要の増加に対して、ソフトウェアの供給が質的に、あるいは量的に追いつかなくなる“ソフトウェア危機”という認識から生まれてきたのと同様に、データサイエンスなどのデータ活用においても、需要と供給のバランスが取れなくなるだろうという見通しに基づいている。従来のソフトウェアエンジニアリングと似たように、データエンジニアリングは実行環境や開発環境を提供して、データを活用する組織や人のケイパビリティを拡張する。

さて、循環的な経済に対応して、DevOps など xOps と呼ばれる、システムやサービスのユーザと提供運用者を結びつけるソフトウェアエンジニアリング手法が生まれてきた。この xOps の観点で見た場合、データエンジニアリングに求められるのは、データ管理の自動化と分権化であろう。

そこを目がけて、さまざまなツール、フレームワーク、サービスなどが提供され、最近ではデータメッシュ、リバーズ ETL、メトリック・ストア、フィーチャー・ストア、ModelOps などが注目されている。しかし、国家のデータ主権、データの個人主権などを含めて、何が許容されるかを定めるポリシーについては、非ソフトウェア技術的な側面が強く、自動的なデータ管理はまだ途上であり、そのことがデータに関わる分権化やエコシステムの形成を妨げている。

今回のデータエンジニアリング特集号は、このような状況の中で社会と企業、そして産業を変えていくデータ流通と高度なデータ利活用に関する論文を採りあげた。

各論文を通じて、IoT と無線通信の発達によりさまざまなデータが取得され、AI による認識技術によって多くの事象がデジタルデータとして記録共有されるようになった結果として、データを根拠としたオペレーションや意思決定が行われるようになったこと、センサーデータやビジネスデータの分析がますます多くの企業で行われ、さらに企業を超えてデータを共有し活用されていること、そして、一連の活動が広がり、繋がって、将来、企業や街を変え、産業や社会を変えていく可能性のあることなどを、データ蓄積、共有、活用の技術やフレームワークとともに事例で紹介している。

本特集が、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。

(フェロー, 最高技術責任者)