

桁違いの巨大データ競争 鍵を握るのは超高速 データベース技術だ



喜連川優

国立情報学研究所所長

ネットなどで集められる莫大なデータに基づいて判断や戦略を立てる——。いわゆる「ビッグデータ時代」だが、この数年のうちにこれまで以上の巨大データ時代がやってくることは間違いない。

低コストのセンサーが家電、装身具、あるいは工場の機器など、あらゆるモノに設置され、情報を吸い上げ、インターネットでつながる時代になるからだ。このIoT (Internet of Things) が本格的にはじまるとデータ量が桁違いに増大する。本当のビッグデータ時代はこれからはじまると言っている。

私はビッグデータなる言葉が生まれる遙か以前より、巨大データからの価値創出とその高速処理に取り組んでいる地味な計算機科学者であるが、よく聞かれる

のが「ビッグデータの勝負で、日本はGoogleに勝てるのか」という問いだ。彼らは検索エンジンやSNSを通じて、テキスト、画像、閲覧情報といったビッグデータを蓄積し、マーケティングに応用するなどして莫大な利益を上げている。しかしビッグデータは他にいくらでもある。

Googleがアクセスできないビッグデータの一つが、わが国の診療報酬明細書(レセプト)の電子情報だ。われわれ東京大学のグループは、医療経済研究機構のプロジェクトに参画し、国内の全レセプトデータを解析可能とするシステムを構築した。年間約16億件、合計400億レコード、6年分で2000億を超える世界最大級のビッグデータである。

レセプト情報を解析すれば、国民が受

にブレイクが踏まれる場所があった。よく見ると、電信柱の後ろに集合住宅の通用門がある。ドライバーは電信柱付近に

来た途端、通用門を目にし、慌ててブレイクを踏むのだ。この種の情報を集めた「ヒヤリハット地図」をカーナビに搭載して、ドライバーの注意を喚起できれば、事故を予防できるはずだ。

日本には社会が抱える課題の解決に活用できるビッグデータが眠っている。我が国はビジネス展開ではGoogleやアマゾンなどに出遅れたが、ビッグデータから社会的価値を創出する新たな局面では、世界をリードできるチャンスは十分にある。社会が受容する価値が生まれればそこに必ずビジネスが生まれる。

データ基盤を制するものが勝つ

多くのベンダーがありとあらゆる分野でデータの収集に血道をあげる。解析手法の工夫よりもデータの量と質によって受ける影響の方が圧倒的に大きい。すなわちデータ基盤を制する者が勝者となる。本格的ビッグデータ時代には、データ処理の性能を桁違いに向上させることが必

要になる。

われわれはそのような観点から、データベースエンジンの開発に取り組んできた。データベースエンジンとは、データのアクセスを読み書きするためのソフトウェアのことである。コンピュータの計算を担うCPU(中央処理装置)や、データを蓄積するストレージなどのハードウェアでは、これまで凄まじい勢いで、高速化、低コスト化が進んできた。ところが、どういいうわけか、両者をつなぐ役割を果たすデータベースエンジンの基本原理はほとんど進化したままだった。

われわれは、内閣府の最先端研究開発支援プログラム(FIRST)の支援を受けて、「非順序型実行原理」という新しい方式を開発した。

従来のデータベースエンジンは、ある命令をデータベースに通知して実行し、終わったら、次の命令を実行するという順序で処理していく。それに対して、われわれの方式では、次から次へと命令を、どんどんデータベースに通知して実行する。実行結果はばらばらに返されてくるが、最終的に辻褄を合わせて一連の処理

けている医療サービスをたどることが可能だ。われわれの分析結果を医療サービスのデザインに積極的に活用しようとしているのが三重県だ。同県では、四日市市など北部の都市部の患者のほとんどが市内の病院に通院する。南部の町村に住む患者は近隣で必要な治療を提供できる病院がないため、遠方まで長時間かけて通院していた。少子高齢化により15歳から64歳の生産年齢人口が減りつつある中、地方の患者が一日の大部分を移動時間に費やして治療を受けている実態が明らかになったのだ。三重県は、こうしたデータを参考に、施策作りに取り組んでいる。

ビッグデータは交通事故の防止にも役に立つ。2016年には50万件近い交通事故が発生し、死者数は3904人である。死亡事故、重傷事故などの背景には、事故には至らないものの、ちょっとしたミスでヒヤリとしたり、何かに驚きハットしたりする「ヒヤリハット」が数多く存在すると言われる。そこでわれわれは運送事業者からドライブレコーダーのデータの提供を受け、運転操作データを解析した。たとえば、ある直線道路で頻繁

を終える。従来方式では、データの出し入れをせず、ただ待っているだけの無駄な時間が頻繁に発生していた。しかし、われわれの方式では、たくさんのリソースに命令を割り当て、一気に実行するので効率的だ。大規模環境では、それまでの1000倍の高速化を達成している。2012年にFIRSTの研究パートナーである日立製作所がわれわれのエンジンを製品化した。これは社会やビジネスの現場で広くビッグデータ処理に使われている。

ビッグデータの処理にかかる消費エネルギーもネックになる。消費エネルギーの増大で、データを処理する施設が止まるような事態も考えられるからだ。そこで、国立研究開発法人新エネルギー・産業技術総合開発機構(NEDO)のプロジェクトで、2020年度までを目途に、データ処理に掛かる消費エネルギーを従来比100分の1以下に抑える研究にも取り組んでいる。

消費エネルギーを抑えつつ高速処理できるデータ基盤は、ビッグデータ時代の競争を有利に進める戦略的武器となろう。