

新原理「非順序型実行」で、データ処理速度1000倍も可能に

ビッグデータの利活用を進める上で欠かせないのが、巨大なデータを高速に処理できるデータベースエンジンだ。一般的なデータベースエンジンの技術は近年、成熟していると考えられてきたが、東京大学 生産技術研究所の喜連川優教授が中心となって開発を進めているのが新原理「非順序型実行」に基づく超高速データベースエンジンである。「非決定的な実行」により、従来の1000倍高速な検索を目指すという、最先端研究開発支援プログラムの現状について、喜連川教授に訊く。

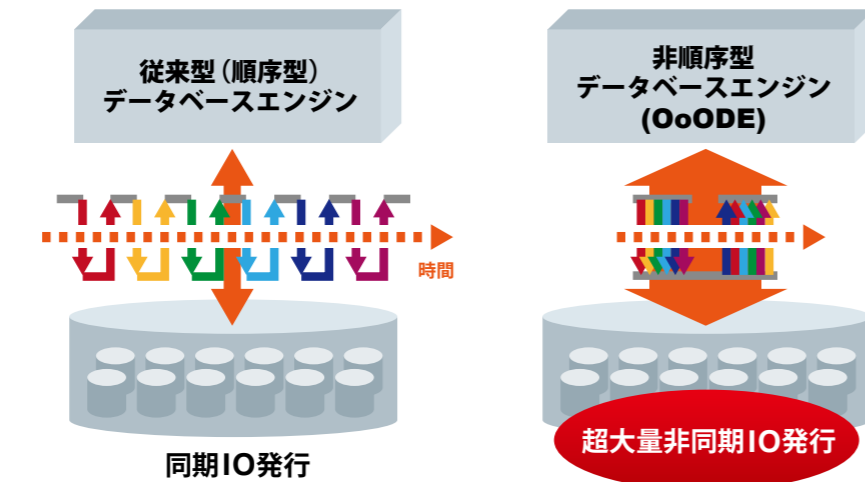
「非順序型実行」はなぜ速い？

喜連川優教授が中心研究者として推進している、内閣府の最先端研究開発支援プログラム「超巨大データベース時代に向けた最高速データベースエンジンの開発と当該エンジンを核とする戦略的社会サービスの実証・評価」において、超高速データベースエンジンの核となっている実行原理が「非順序型実行」である。データ処理を従来よりも1000倍速くすることも可能な実行原理とは、どのようなものなのか。そして、なぜ速いのか――。

喜連川教授は、次のように語る。

「現在の一般的なデータベースというのは、データの処理を『順序型』の実行原理に基づいて行っています。ある命令がくると、それを実行し、終わったら次の命令を実行するという処理を順番に繰り返していくというものです。したがって、命令も順序だててしなければいけない。それに対して、必要な命令を最初に全部言ってしまうというのが『非順序型』の実行原理です。これとあれのデータがほしいと返事が返ってくる前に、必要になったらどんどんと命令する。すると、データはばらばらな順序で返ってくるけれども、それで構わない。最後につじつまが合えばいいだろうという発想なんです。だから速くなります」

従来の実行原理では、通常は事前に決められた順序に従ってデータ処理を行うため、一つ処理しては戻し、とい



従来型データベースエンジン(左)の問題と非順序型データベースエンジン(右)

うやりとり時間に時間を要する。また、あるパートが一つの処理を行っている間、別のパートは待機していることになる。データの処理にかかる時間のほとんどは、じつはこうした待ち時間であり、ハードウェアも使われずに休んでいる部分が多い。しかし、たくさんのパートが一気に働いて処理を行えば、時間は劇的に短縮できる。

ソースコードを一から書き直す

複数のパートが同時に処理を行うと聞くと「分散処理」や「並列処理」と同じようなものではないかと思えてくる。しかし、喜連川教授によると、その両者とはまったく異なるのだという。

「分散も並列も順序実行を分割して処理するもので、あらかじめ決められた順

番に従って処理するという点では従来方式と変わりません。ところが、非順序型実行は、命令を実行するたびに処理のパターンが変わるのです。でも最後にはつじつまが合うのです」

素人考えだと、そんな曖昧なことでもうまくいくのだろうかと思ってしまう。「この原理を最初に思いついたのは2005年頃のことですが、当初はわれわれ自身もうまくいくのか半信半疑でした。けれども、実際にこの原理をデータベースエンジンに適用してみると、なかなかいい手応えがある。そこで日立製作所と一緒に、この原理でデータベースエンジンを開発しようということになった。ただ、それは今あるものをちょっと改良するというレベルのことではありません。ソースコードを一から書き直すという、大きな決断と労力を要するチャレンジでした」と喜連川教授は振り返る。

3桁速くなれば、世界が変わる

そのチャレンジが実を結び、日立製作所が2012年5月末に製品化したデータベースでは、従来の100倍という驚異的な処理速度の向上を達成した。ハードウェアはそのまま、データベースエンジンのソフトウェアを変えただけで、である。

「ただ、目標はあくまでも1000倍です。3桁速くなれば、世界が変わる。今より1000倍大きいデータも、今と同じ感覚で活用できるようになるぐらいでなければ、ビッグデータの時代をリードすることはできないでしょう。そのためにはハードウェアにも能力が必要になります

から、現在はマルチコアサーバを利用して、目標達成に向けて開発の最終段階に入っています」

データベースで最も一般的なのが、リレーショナルデータベースと呼ばれる管理方法に基づいたものである。その基本的原理は30年以上変わらず、ある程度完成されたものと考えられていたが、非順序型実行原理でその根っこ部分にメスを入れようとしている。ギガからテラ、ペタ、エクサ、ゼタと、爆発的に扱うデータ量が増えていくビッグデータ時代。1000倍高速なデータベースエンジンが、その時代の先端を切り拓いていくことが期待されよう。

(取材・文＝関亜希子)



喜連川 優
東京大学 生産技術研究所 教授

University Industry Collaboration

パラダイム転換でビッグデータ利活用を促進する日本発の原理を、高速データアクセス基盤に活用

「東京大学との共同研究開発成果を製品化」

2012年5月末に株式会社日立製作所(以下、日立)が発表した高速データアクセス基盤が、IT業界の大きな注目を集めている。従来自社比約100倍^{※1}という高速検索能力を持つ「Hitachi Advanced Data Binder プラットフォーム」。それは、最先端研究開発支援プログラムにおいて東京大学と日立が共同で推進している超高速データベースエンジンの研究開発^{※2}の成果を製品化したものだ。

ビジネスや社会活動におけるビッグデータの利活用が注目を集めているなか、Hitachi Advanced Data Binder プラットフォームは、データの検索と分析を強力に支援するツールとして製品化された。超高速データベースエンジンの鍵を握っているのは、もちろん喜連川教授が中心となって考案した非順序型実行原理^{※3}である。開発陣は、日本発のすば

らしい原理をいち早く製品として世に送り出していくことで、ビッグデータの利活用に貢献したいという思いから開発に臨んだという。ただ、それは決して簡単なことではなかった。

開発は、非順序型実行原理を、日立が培ってきたデータベースエンジンHiRDBに投入するという形で行われた。しかし、そのためには、日立がこれまで営々と築いてきたコードボディを捨て、新原理に基づいて書き直す必要があった。ITベンダーにとって、それは非常に勇気の要ることである。しかし、「それぐらいのことをしなければイノベティブなもの作れない」との覚悟で挑戦を決めたという。成功すれば、データベースのパラダイムを変えるかもしれない挑戦。若手研究者・技術者にとっても、やりがいのあるものとなったことは間違いのないだろう。

今回は、これまでの研究開発成果を反映し、従来自社比で約100倍のデータ検索性能を発揮するデータベースエンジンを製品化した。さらに東京大学と日立は、研究開発における最終的な目標として、2013年度中に従来型データベースエンジン比で1000倍のデータ検索性能達成をめざし研究を続けている。

超高速データベースエンジンが拓く新たな地平は、ビジネスや社会にどのような変革をもたらすのだろうか。期待を込めて注視したい。

※1 解析系データベースに関する標準的なベンチマークを元に作成した、各種のデータ解析要求の実行性能を計測。データ解析要求の種類によって高速化率に差は見られるが、データベースにおいて特定の条件を満たす一定量のデータを絞り込んで解析を行うデータ解析要求を対象とした結果。

※2 喜連川FIRSTプロジェクトにて実施。

※3 喜連川教授・合田特任准教授(東京大学)が考案した新原理。