

ハードウェア性能をフルに生かす処理構造の例 —— 未知数のメモリ型DBMS

A processing architecture for making the most of hardware resource balancing --- Feasibility of MDA (Memory based DBMS Architecture)

福市 良次[†]
Fukuichi, Ryoji[†]

[†] 有限会社ポプラ、メモリ型DBMS 推進機構
[†] Popula Ltd., Memory Base DBMS Initiative

要旨

メモリ型DBMSとは、基本的に「データベースをメモリ上に構築することを前提とした構造」を持っており、「中間処理を排除するなどによる蓄積情報の削減」の概念を持ったものである。
本論文では、実現されているメモリ型DBMSとその導入事例を参考に、メモリ型DBMSの導入効果がシステム開発者スキルに依存する範囲を超えており、明らかに技術コンセプトの転換によるものであることを述べ、さらに、その効果の見通し、必要な技術要素、今後求められる技術的検討などについて述べる。

はじめに

1970年頃、IBMのコッド博士によるRDBMSの提唱は、データへのアクセス速度は問題にならない概念モデルからスタートしている。しかし、現実的には正規化崩し、テーブルの複数媒体への配置（同一データのコピー配置）、バッチ処理などの処理分割など様々な工夫により、HDDの極めて遅いアクセスネックを乗り越える工夫が必要があった。

この努力は、HDD上のデータアクセスの高速化を導く一方、見掛けのデータ量を極端に膨張させ、通常の業務処理において、発生源データ（元々必要であるデータ、原始データ、生データ）の量の数10倍～100倍以上にも達している。

言い換えれば、テラバイト（TB）を越えるデータ量を保有している場合においても、発生源データのみを考えると、数十ギガバイト（GB）で収まっている可能性があることを示している。

1. ハードウェアの進化について

1.1. CPU 処理速度

WS、PC系のCPUクロックは数psオーダに入っており、先取り処理、並列処理などの高速化技術を駆使し、未だに「ムーアの法則」（15年間で1000倍）を維持し続けている。

1.2. 記憶装置のアクセス速度（初期アクセス）

記憶装置に関しては、15年スパンでは10倍前後に向上しているに過ぎない。しかし、メモリとHDDのアクセス速度の差は従来から10万倍程度ある（メモリ：10nsオーダ、HDD：数msオーダ）。

1.3. メモリ容量

32bitPCではここ15年で数MB搭載から数GB搭載が常識化し、1000倍になっている。さらに64bitPCの出現によりアドレス空間の拡大とメモリ価格の低下により、数10GB搭載機が出現している。

2. ソフトウェア構造の進化について

一般的な業務系の処理システムの処理構造を図1に示す。

通常は一つの処理において必要となるデータはカタマリとして媒体（HDD）上に記録したり、別媒体に分けて記録したり、並列処理のプロセスから同じ媒体へのアクセスを避けたり、様々な工夫をしながら、遅いHDDアクセスの影響を少なくする工夫がなされる。

近年、①取り扱う業務量の爆発的な増大、②事業市場の多様化に伴う分析処理の増大などによりバツ

チ処理が拡大しているため、ストレージ装置を大量に導入し、HDD 記録上の工夫はますます巧妙性が要求されている。しかし、処理構造は基本的に変わっていない。

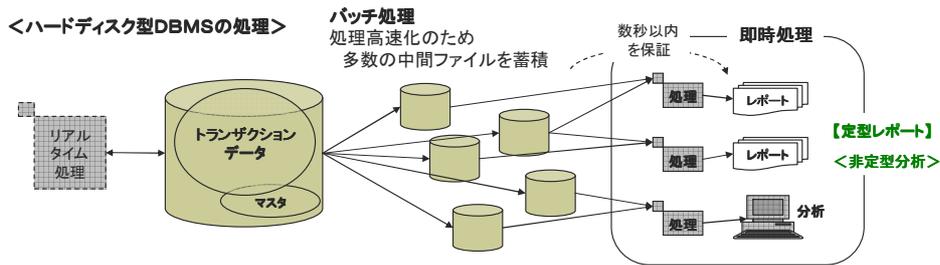


図1 一般的な業務系処理の処理構造

3. メモリ型 DBMS の発想

最近になって、メモリ容量の拡大がある限界を超えてきたことに気付いた人たちが現われた。多くの業務処理において、トランザクションデータや分析対象の発生源データのみならば、大半がメモリ上に展開可能である、という。これに従い、メモリ上でのデータ格納・アクセスを基本に設計した DBMS が製品化され始めた。

メモリ型 DBMS は「超高速」だけではない。図2に示すとおり、保管要データは、トランザクションデータ（発生源データ）とマスタファイルのみとなってしまう。

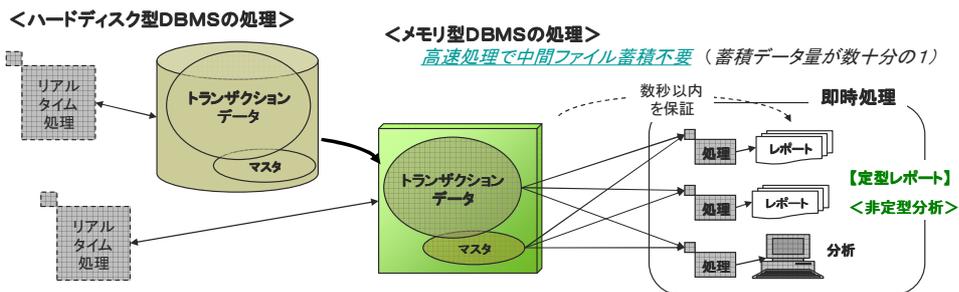


図2 メモリ型 DBMS の狙い

<補足>

メモリ型 DBMS を活用した場合、プロセス処理速度に差はない。データベースアクセス頻度や設計適正によって差が生じるが、経験的に通常業務処理においては 100 程度のコスト・パフォーマンスの改善が期待される。

4. 導入事例

メモリ型 DBMS の活用事例として最も典型的なものはCGCジャパン社のシステムである。本事例に用いた事例の概要、数値等は、日経コンピュータ 2006.3.6 より引用したものである[1][2]。

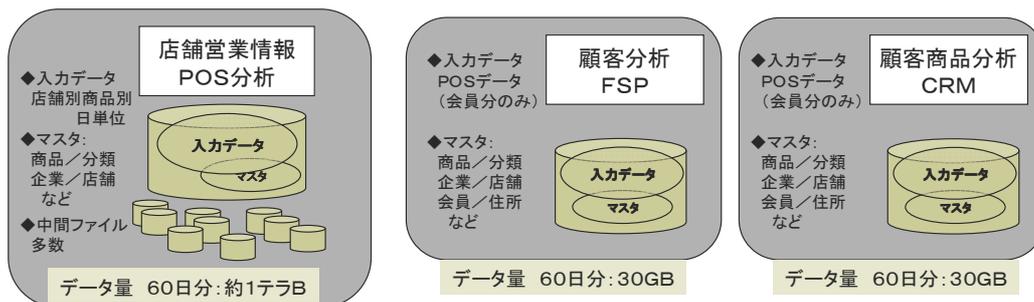


図3 CGCジャパン社の従来システム

従来 DBMS の性能限界により 3 つに分割していたシステムを、64bitLinux マシン 1 台（予備系は装備）でメモリ型 DBMS を活用して一つにと統合した。

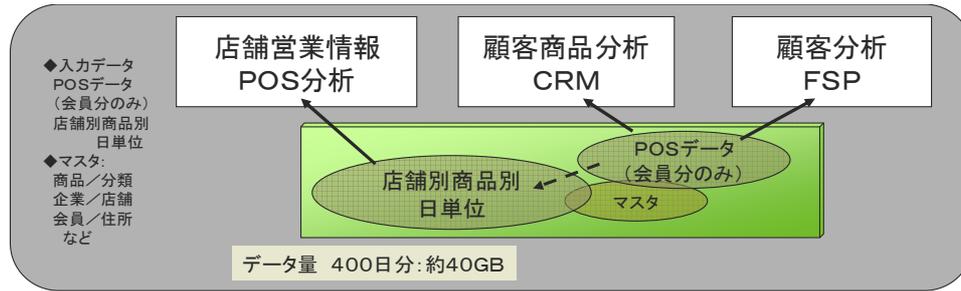


図4 CGCジャパン社の統合後のシステム

改善効果は、

- ① 3 つのシステムが一つに統合された。
- ② 夜間のバッチ処理が不要となった。
- ③ 蓄積データ量が 1 / 100 以下に低減した。
- ④ 分オーダー（Web のタイムアウトになるほど待たされた）処理が、2 ~ 3 秒に短縮した。

経験的ではあるが、開発者スキルの優位性で改善される効果は 10 倍程度といわれており、これを遥かに越えた効果を上げている。これは単にスキル差の問題を越えており、メモリ型 DBMS という方式（技術）概念がコスト・パフォーマンス上の---極めて大きい---一定の効果を引き出すした、と考えられる。

5. 応用面で何が起こるか

処理の高速化、保管データ量の縮小

- ✓ 処理待からの解放
 - ・ 長時間、夜間バッチ処理運用なし
 - ・ 何時でも処理可能 / 処理指定ミスの再処理
 - ・ 過去データで遡り処理 / 組織変更、管理方式変更での仮想処理
 - ・ 臨時の処理、特別の処理 / 試行錯誤環境充実 / 非定型処理の定着
- ✓ データ管理の複雑さからの解放
 - ・ 量の削減（捨てられないものしか保管していない）
 - ・ 種類の削減（発生源データとマスタファイルデータ）

6. 技術的側面で何が起こるか

6.1. 現在開発されているメモリ型 DBMS の共通的技术事項

現在開発されているメモリ型 DBMS に、比較的共通して使われていると思われる技術は、

- ✓ 「列」管理 …… メモリアクセスの特性から
- ✓ 固定長項目 …… メモリアドレス管理の特性から

などである[3][4]。

6.2. 期待される技術的事項の検討項目

今後、メモリ上に展開されることを前提とすれば、検討が必要な技術テーマの範囲は、

- ✓ アルゴリズムの検討
 - $O(n \log n)$ 、 $O(n)$ の世界
 - HDD アクセス阻害のない世界（確率計算、 $O(n^2)$ 、 $O(n^3)$ の世界からの解放）
- ✓ データベース構造の研究開発
 - メモリ上のデータの圧縮化、処理目的に適合した DB 構造の開発
- ✓ DBMS 構造の研究開発

故障時のデータ保証、マルチ構造
不揮発性メモリ、組込み型など、メモリハードウェア特性との関係
セキュリティ方式との親和性
処理に適した API、汎用 API の開発と標準化

- ✓ アプリケーション処理構造の研究開発
メモリ型 DBMS を活用した場合のアプリケーション構造論
 - ✓ メモリ型 DBMS のオープンソース化
- など、検討されなければならないことは多い。

6.3. 将来技術への繋ぎと誘導

また、将来的にはハードウェアとソフトウェアの構造が一体として研究される可能性を示唆する。

- ✓ ハードウェア構造とソフトウェア構造の融合
反復論理マシン、連想メモリなど

7. 課題

7.1. 活用上の課題

処理構造概念の根本的な変更を伴うため、利用側に躊躇を呼ぶ。このために、ツールとして使える検索エンジン型の利用法が期待される。

7.2. IT 市場の課題

急激な普及はハードウェア市場、技術者の保守志向、バッチ処理運用市場への外乱を与える。

7.3. 研究開発上の課題

三層モデル、DOA、SOA、データウェアハウス、近年においてはグリッド・コンピューティングなど、切り口は異なるが技術概念、方式概念の普及がその分野の研究や製品開発を推進している。メモリ型DBMSの場合においても、社会的知名度の向上が今後の発展につながるものと考えられる。

8. 既開発のメモリ型 DBMS

現在開発されているメモリ型 DBMS は、リアルタイム用超高速型(非 SQL)、リアルタイム用 SQL 型、検索エンジン型などがある。詳細は、参考文献[5]に記載。

まとめ

メモリ型DBMSは単に「メモリ上のデータ格納」というだけでなく、ソフトウェア構造に大きな変革をもたらす。一方、この大変革の可能性自体が、普及の抵抗力にもなっている。

今後、各分野で同様の発想で「データのメモリ格納型の処理」について、様々の方が試行され、ある一定の効果を上げられると思う。このときに、「高速処理」というだけでなく、様々な効果が付随し、また、そのためのデータ構造に一定の工夫が行われる。これらは、個人個人のノウハウに収めてしまうことなく --- 学会などにおいて --- 「技術体系化」し、社会技術資産として共有されるべきである。

参考文献

- [1] 岡本藍, “特集2 データベース性能の限界を超える「作れなかったシステム」を実現”, 日経コンピュータ 2006.3.6, P.62-69
- [2] 大和田尚孝, 渡辺享靖, 小原忍, “CGC ジャパン情報系DBの鮮度を高める”, 日経コンピュータ 2006.8.21, P43
- [3] “高速屋テクノロジー” ホームページ <http://www.kousokuya.co.jp/technology.html#1>
- [4] 古庄晋二, “汎用超高速データベース処理技術”, 東大総研, 2005.8.
- [5] 福市良次, “メモリ型DBMS” はコロンブスの卵!, イデア出版, 2005.7.