ディスクストレージ省電力化型問合せ実行方式に関する一考察

合田 和生† 喜連川 優†
†東京大学 生産技術研究所

あらまし

エンタープライズシステムは、爆発的に増大する情報を確実に管理し高速に処理する必要から、多数のディスクドライブが組込まれる傾向にあり、当該システムにおいてディスクドライブによって消費される電力の削減は重要な研究課題である。本論文では、エンタープライズシステムにおけるデータベース管理システムにおいて、ディスクストレージのエネルギー消費を高効率化する問合せ実行方式を提案する。

1 はじめに

本論文では、ITシステムのうち、ストレージシステムのエネルギー消費の高効率化を議論する。今日のデータセンタには、爆発的に増大する情報を安全に管理し、また高速に処理するために、極めて多数のディスクドライブが組込まれるようになってきている。ストレージシステムの業界団体であるSNIAの統計では、平均的なデータセンタにおけるストレージシステムに関連する消費電力は、全体の28%に上るとされている[4]。殊に、エンタープライズシステムにおいてはこの傾向が顕著であり、1台のサーバに数百台以上のディスクドライブが接続されることも珍しくなく、ストレージシステムのエネルギー消費の高効率化は、重要な一課題と言えよう。

ディスクドライブの消費エネルギーは、殆どがスピンドルモータによって消費されている。ディスクドライブの消費電力を削減するためには、ディスクアクセスのある期間のみ当該モータを駆動し、ディスクアクセスのない期間に当該モータを停止させることが基本である。しかしながら、プロセッサやメモリモジュールと異なり、スピンドルモータでは、特にスピン開始に伴う時間損とエネルギー損はそれぞれ数十秒と数百ジュールにのぼる[3]。スピンドルモータの停止によって消費電力を削減するためには、多くの場合、数十秒以上に渡って全くディスクアクセスのないアイドル時間を生成を要し、しかもオンラインの制御系がこれを的確に予測してスピンドルモータの停止と再開を指令する必要がある。

図1(a)に示す通り、これまで、現行のストレージインターフェースの上でどれほどの省電力化の効果が期待できるかという観点で主にディスクストレージの制御系に関する研究が進められてきた[1, 2]。これに対して、著者らは一層のエネルギー消費の高効率化を目指し、既存のストレージインターフェースに捕われず、むしろ新しいストレージとサーバの接続関係が可能となった場合に、どれほどの省電力化の効果が期待できるかという視点に立ち、コラボラティブグリーンストレージ[6]と称する新しい消費電力削減のアプローチを提案してきている。すなわち、図1(b)に示す通り、より高いレベルのインターフェースを規定し、サーバ上のアプリケーションと高度な連携をはかることにより、サーバ上のアプリケーションまでを含んでシステム全体を見渡したグローバルな制御が可能となり、より一層の消費電力の削減効果が得られる可能性がある。

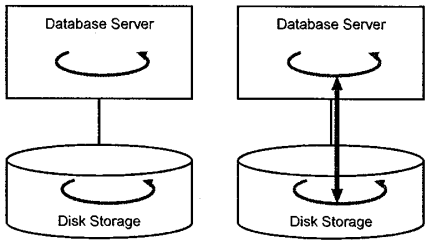

図1: ディスクストレージの省電力化手法の比較.

本論文では、当該アプローチの下で、意思決定支援システムなどに見られる比較的長時間を要する問合せ処理に関して、データベースサーバが問合せ処理に先立ち生成する問合せ実行計画なる実行情報をディスクストレージに開示するとともに、データベースサーバがディスクストレージの消費電力を意識した実行最適化を行うことにより、エネルギー消費の大幅な高効率化を目指す新しいデータベースの問合せ実行方式を提案する。

2 省電力型の問合せ実行

意思決定支援システムに代表される長時間の実行を必要とするワークロードに対する、問合せ処理に掛かるディスクストレージのエネルギー効率を向上するアイデアを示す。いずれも、データベースサーバとディスクストレージの深い相互連携が不可欠であるが、従来型のアプローチと比べて著しいエネルギー効率の改善が期待される。

A Study on Power-aware Query Execution for Improving Disk Storage Energy Efficiency
Kazuo Goda† and Masaru Kitsuregawa†
†Institute of Industrial Science, The University of Tokyo

プロアクティブな電力制御　データベースサーバに問合せが与えられると、当該サーバはまず問合せ実行計画なる実行情報を生成し、当該実行情報に従い問合せ処理を行う。当該実行情報を実行に先立ちディスクストレージに開示することにより、能動的にディスクドライブの電力制御を行うことが可能となり、よって高効率の電力制御を実現することが期待される [7]。

ハッシュ結合によってRとSの2つのテーブルを結合する場合を考えよう。RがSより小さいとすると、通常、データベースエンジンはRをディスクストレージから読み出して、データベースバッファにハッシュ表を作成し、この後、Sを読み出して、ハッシュ表を検索し、結合条件に合致したレコードを出力する。問合せ実行計画には当該手続きが記録されており、当該情報をディスクストレージに開示することにより、Rが格納されたディスクドライブとSが格納されたディスクドライブをいつどのようにアクセスするかをディスクストレージは高い確度で予測することが可能となる。RとSが異なるディスクドライブに格納されているとすると、Rを読み出している期間にはSが格納されたディスクドライブをスピンダウンし、Sを読み出し始める事前に当該ディスクドライブをスピンアップしておき、さらに、Sの読み出しが始まったあとにはRが格納されたディスクドライブをスピンダウンすることができるだろう。これは、旧来型の受動的な電力制御と比べて性能に掛かる副作用が少なく、また、高い電力削減効果があり、有益性が高い。

問合せの実行最適化　現行の問合せ最適化機構は、与えられた問合せに対して、その処理性能を最大化するような実行計画を生成する。これに対して、新たに、問合せに掛かるディスクストレージの消費電力量や、ピーク電力を考慮した新たな最適化方式が必要となる。

前者については、従来型の問合せ最適化においては、例えば、表の選択率によって、二次索引を用いた表のアクセスを行うか、表の単純走査を行うかを決定していた。選択率が十分に低い場合には、ディスクドライブのランダムアクセス性能は、シーケンシャルアクセス性能と比べて2桁以上低いものの、対象となるレコード数も少ないことから、二次索引を用いる表のアクセスが優位であり、一方、選択率が一定以上の場合には、むしろ、単純に表を走査することが、性能上優位である。ディスクストレージの消費電力を新たに考慮する場合、スピンドルモータの回転エネルギーの活用効率を最大化する必要があり、ディスクアクセスのバースト化が基本となるため、従来型の問合せ最適化と比べて、いくらか、シーケンシャル側に最適ポイントがシフトする可能性があるだろう。

一方、ピーク電力の考慮については、問合せを部分問合せに分割して、部分問合せの実行をスケジュールすることができるだろう。例えば、小さい表Rと大きい表Sを索引結合する場合を考えたい。Rは全てボリュームVRに格納されているが、Sについてはその部分表S1がボリュームVS1に、S2がVS2に、S3がVS3にそれぞれ格納されているものとする。この際、結合は、$R \bowtie S = (R \bowtie S1) \cup (R \bowtie S2) \cup (R \bowtie S3)$ と展開することができる。左辺の通りに索引結合する場合に比べて、右辺に従って3つの部分問合せを逐次的に実行する場合、Rへの冗長なアクセスがある分、性能はやや低下するものの、$R \times S1$ を実行している間は、VS2ならびにVS3をスピンダウンすることができるため、ピーク電力を抑制することができるだろう。

複数問合せの実行最適化　複数の問合せが同時に処理される場合、個別にプロアクティブ電力制御を行うことが可能であるものの、各々の問合せが独立の処理されていたのでは、ある問合せにおいては特定のディスクドライブがアクセスされないのでこれをスピンダウンして消費電力を削減しようとするのと同時に、別の問合せが同じディスクドライブをアクセスしてしまい、総じての消費電力の削減効果が低くなる可能性がある。同時に処理される問合せ間でディスクドライブの電力状態を意識した相互スケジューリングを行うことにより、総じて消費電力の削減が期待されるだろう。例えば、著者らは、[5] において、ディスクドライブをスピンアップすることのできる権限を一部の問合せ系列に制限するスケジューリング方式を提案した。

3　まとめ

本論文では、コラボラティブアティブグリーンストレージなる新しいディスクストレージのエネルギー消費を高効率化するアプローチとして、意思決定支援システムに代表される比較的長時間を要する問合せ処理に関して、プロアクティブ電力制御、単一問合せ／複数問合せの実行最適化なるアイデアを示した。今後は、当該アイデアに基づき、シミュレーション実験を行い、アイデアの潜在的な利得を明らかにすると共に、実際のシステムに実装を行い、実システムへの適用に際する諸問題の検討を進めたい。

参考文献

[1] D. Colarelli, D. Grunwald, and M. Neufeld. The Case for Massic Arrays of Idle Disks (MAID). In *Proc. USENIX Conf. on File and Storage Tech.*, 2002.

[2] R. A. Golding, P. Bosch, C. Staelin, T. Sullivan, and J.Wilkes. Idleness is not sloth. In *Proc. USENIX Tech. Conf.*, pp. 201–212, 1995.

[3] S. Gurumurthi, A. Sivasubramaniam, M. Kandemir, and H. Franke. Reducing Disk Power Consumption in Servers with DRPM. *IEEE Computer*, 36(12):59–66, 2003.

[4] S. Worth. Green Storage. SNIA Education, 2006.

[5] 合田和生, W. Qu, 喜連川優. 複数問合せ処理を意識したディスクストレージ省電力化に関する一考察. 電子情報通信学会 データ工学ワークショップ, pp. D5–2, 2008.

[6] 合田和生, 喜連川優. コラボラティブグリーンストレージ：データベースシステムとの連携によるディスクストレージ省電力化の構想. 電子情報通信学会インターネットアーキテクチャ研究会, 電子情報通信学会技術報告, 109(351 IA2009-67):13–16, 2009.

[7] 上野裕也, 合田和生, 喜連川優. データベースシステムの問い合わせ実行計画を利用したディスクアレイ省電力化に関する一考察. 日本データベース学会 Letters, 6(1):85–88, 2007.