様　式　Ｃ－１９、Ｆ－１９、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成 ２７ 年　 ９ 月 １８ 日現在

機関番号：３２６１９
研究種目：基盤研究(B)（一般）
研究期間：2012～2014
課題番号：２４３０００３４

研究課題名（和文）高伸縮性情報基盤における応用処理と連携した実行時省電力データベースの研究開発

研究課題名（英文）Application Research on Run-time Energy Management Database Systems cooperating with applications' behaviors in cloud computing

研究代表者
中野　美由紀（NAKANO, Miyuki）

芝浦工業大学・公私立大学の部局等・教授

研究者番号：３０２２７８６３

交付決定額（研究期間全体）：（直接経費）　　13,800,000円

研究成果の概要（和文）：高伸縮性情報基盤の主要な応用処理であるデータベースを対象に、アプリケーションと連携した実行時省電力技法の研究開発を行った。アプリケーションの知識（処理粒度、索引の有無など）、実行時特性（ライブラリィ挙動、入出力挙動、アクセス局所性等）を用い、アプリケーションの運用時の性能要求を満たしつつ、実行時に省電力可能なミドルウェアフレームワークの設計方式、特に複数のアプリケーション稼働時に省電力可能性の高いアクセス挙動を抽出する手法を新たに提案し、大規模エンタープライズストレージ上にて従来と比較して３０%以上の省電力が可能であること実機を用いて評価し、提案方式の有効性を明らかにした。

研究成果の概要（英文）：Numerous critical applications such as database systems or Web commerce applications are in constant operation at data centers, and the conventional approaches that only utilize storage-device-level I/O behaviors do not produce sufficient energy savings while applications are running. It may be possible to dramatically reduce storage-related energy consumption without degrading application performance levels by utilizing application-level I/O behaviors or other running behavior. We propose a universal storage energy management framework for runtime storage energy savings that can be applied to any type of application. The results of evaluations show that the use of this framework results in substantive energy savings compared with the traditional approaches that are used while applications are running.

研究分野：データベース

キーワード：実行時省電力　データベース　クラウド・コンピューティング　応用処理連携　性能評価

１．研究開始当初の背景

近年、クラウド・コンピューティングが急速に普及し、Amazon EC2,S3、Google App Engine、IBM Smart Business Cloud、Microsoft Azure 等、米国を中心に高伸縮性情報基盤(クラウド・コンピューティング基盤)が多く構築されている。様々なユーザに必要なだけの処理性能を必要なときに提供することから、高伸縮性情報基盤の特徴として「伸縮性」、「拡張性（スケールアウト）」、「マルチテナンシー」が挙げられる。多様かつ急激に変化する応用処理に対応すべく、高伸縮性情報基盤は大規模データセンタ上に構築され、その消費電力は膨大である。IDC は近い将来空調を含む電力コストがサーバ等の設備機器購入コストを上回ると報告し、低炭素社会が推進される中、高伸縮性情報基盤の省電力化が強く求められている。

大規模データセンタでは空調の効率をも考慮した集約的省電力手法が試みられているが、現状の省電力アプローチはハードウェア機器の静的な省電力化（利用率の低い夜間のノード、ストレージのコンソリデーション等）に留まっている。静的な省電力化はあらかじめ処理負荷が予想可能な場合には有効であるが、多様なサービスが稼動する高伸縮性情報基盤では必要な計算資源が短期間で変動するため適用が難しい。あるクラウドサービスでは一カ月の利用数が約 10 倍変動するとの報告がある。また、Twitter, Facebook 等のサービスでは一日数万人の単位で利用者が増える。さらに、Evernote と Google リーダー の連携等、複数のクラウドサービスが並行して利用される。そこで、「伸縮性」「拡張性」「マルチテナンシー」の３つの特徴に対応しつつ、消費電力を削減するためには、従来の省電力手法に加え、最低スループットの確保等の性能要件を満足しながら、応用処理の負荷変動に対応した新たな省電力化技法の開発が急務である。

McKinsey は Smart City, アーバンセンシング等 IT 化された社会活動が生成・集積するデータの急増しつつあるとし、Big Data の到来を示した。これらの蓄積された大規模データの解析、管理にはデータベース化が欠かせない。ストレージ出荷量は年率 30%以上増加しており、大規模データセンタで利用されるストレージの主たる用途の 70%以上はデータベースと報告されている。Web Service の顧客管理等の例からも分かるように、高伸縮性情報基盤上の主たる応用処理はデータベース上で稼働すると考えられる。従って、応用処理の負荷変動に対応した実行時省電力化の実現には、応用処理の共通基盤（ミドルウェア）となるデータベースにおいて、応用処理の挙動と連携した実行時省電力化を導入する必要がある。オンライントランザクション処理における電力消費の解析、電力消費を抑えたデータベース・プラットフォームのハードウェア構成は検討されているが、サービス運用時におけるデータベースの実行時省電力化の研究は申請者の知るかぎり見当たらず、世界に先駆けた研究といえる。

２．研究の目的

低炭素社会が推進される今、巨大化する一方の高伸縮性情報基盤（クラウド基盤）では省電力化が強く求められている。従来から大規模データセンタのハードウェア機器の省電力は試みられているが、クリティカルな応用処理稼働時の省電力は未だ実現されていない。本研究では、高伸縮性情報基盤の主要な応用処理であるデータベースを対象に、応用処理と連携した実行時省電力技法の研究開発を行う。すなわち、応用の知識（処理粒度、索引の有無等）、実行時特性(ライブラリィ挙動、入出力挙動、アクセス局所性等)を用い、応用の性能要求は満たしながら、実行時に省電力可能なデータベースの設計、実装構築を行う。提案方式のデータベースの上で応用処理を稼働、実行時消費電力を実測し、性能要求を満たしつつ電力消費を大幅に低減できることを示す。

具体的には、高伸縮性情報基盤における多様な応用処理に動的に対応可能な実行時省電力データベースの設計、開発を行う。現在の高伸縮性情報基盤ではユーザに割り当てた計算資源の省電力化は行わない。そこで、ユーザ応用処理負荷の変動に動的に追随可能な省電力データベースを実機上で実現することを目的とする。予測の難しい負荷変動に対応するために、実行時プロファイリングで蓄積した統計情報に併せて、動的に応用処理のモニタリング（図中①）を行い、その結果から省電力の可能性（利用ノード数の縮退、データ再配置によるストレージの停止とスローダウン、スループットを維持したままの CPU クロックダウン等）を探り、動的に省電力化を実行する。この際に、応用処理をモニタリング時に SLA (Service Level Agreement)に基づいたスループット、処理性能を保持すべく、処理特性を解析し、負荷が大きく変動する応用処理の挙動と連携した実行時省電力データベースの基本設計を確立する。

高伸縮性情報基盤で利用される応用処理では大量データがストレージに保持され、ユーザは想定されるピーク値に合わせて計算資源を確保している。この予測ピーク値は変動を吸収するため大きく設定されており、応用処理実行中、タスクごとの処理負荷は比較的小さい。結果、ストレージ処理負荷がピーク値に到達しない場合、要求処理性能の維持に不要なストレージ、ノードの省電力化が期待される。しかしながら、どのデータにどのような処理がなされるのかは実行時まで定まらない。本研究では、データベース上のデータ構成(表、索引)およびアクセス履歴が応用処理挙動を反映していることを利用し、実行時モニタリングをデータベース内で行い、ミ

ドルウェアであるデータベースが応用処理実行時にストレージあるいは CPU の省電力機構（ハードウェア）を起動する「実行時省電力データベース」を開発する。TPC-C,TPC-W 等のベンチマークに加え、複数の擬似応用処理を稼働させ、提案手法で実行時に電力消費を低減可能なことを実機による電力の実測により確認する。さらに、実ビジネストレース等を入手し、スループットを維持しつつ消費電力を大幅に減らし、提案手法が実運用に適用可能なことを明らかにする。

３．研究の方法

応用処理と連携した実行時省電力データベースの設計、実装を行い、実験システム上にて応用処理（OLTP，Web Service，DSS 等）を実行、提案する手法による省電力の効果を実際の電力計測によって検証、確認する。平成 24 年度は応用処理と連携した省電力データベースの基本設計を行うと同時に、実験システムの構築を開始し、サーバ機器の基本的な消費電力等を測定する。平成 25 年度以降、実験システム上でオープンソースデータベースを基に実行時省電力データベースプロトタイプを実装する。ストレージ省電力機能とデータベース連携機構の設計、応用処理のモニタリングとデータベース連携機構の設計を行い、この二つの連携機構をプロトタイプ上に実装する。実装後、TPC-C、TPC-W 等のベンチマークや実応用処理トレースを用い、応用処理実行時の省電力効果を確認し、また既存の手法との比較解析を行い、提案手法の有効性を示す。

４．研究成果

本研究で提案しているアプリケーション協調型大規模ストレージ省電力手法について述べる．本手法の特長は，i）アプリケーション実行時のストレージ省電力，ii）アプリケーションレベルにおける入出力発行間隔の長さや read/write 入出力の頻度等のモニタリング結果に基づくアプリケーションレベルでの入出力挙動のパターン化，及び iii）アプリケーションレベルの入出力挙動のパターンに基づく，適切なストレージ省電力手法の選択及び適用，である．

4.1 省電力ストレージモデル

我々が提案するアプリケーション協調型大規模ストレージ省電力を実現する省電力ストレージモデルを図１に示す．

アプリケーションは，バッファに対してアプリケーションレベルの入出力を行う．そして，バッファを用いて先読み、書き込み遅延など省電力に有効な手法を各データごとに適用、その後ストレージへの入出力を行う．．本モデルでは，アプリケーションレベル、ストレージレベル入出力の監視を行うと共に、消費電力の推移を計測し、アプリケーション入出力挙動の経時変化に追随するようデータ配置、入出力パターンなどの見直しを行う。

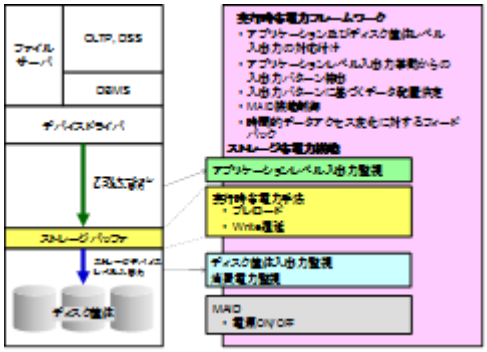

図１．省電力ストレージモデル

4.2 論理入出力パターン

ストレージの省電力にアプリケーションの入出力挙動特性を利用するために，論理入出力パターンという概念を導入する．論理入出力パターンとは，アプリケーションの入出力挙動をストレージ省電力手法を適用できるように分類・パターン化したものであり，実行時に省電力機能を適切に選択するために使用する指標である．

入出力パターンは次の 4 種類である．第一は，モニタリング期間中にアプリケーションから入出力が発行されなかったことを識別するための入出力パターン(P0)である．本パターンに該当するデータを識別することにより，容易に電源 OFF などのストレージ省電力機能を適用できる可能性が増加する．第二はストレージキャッシュを用いることで read 入出力間隔を延伸できる可能性があるデータを識別するためのパターン(P1)である．第三は同じくストレージキャッシュを用いるが，read ではなく write 入出力間隔を延伸できる可能性があるデータを識別するためのパターン(P2)である．最後は入出力の間隔が短く，ストレージ省電力機能を適用することができないデータを識別するための入出力パターン(P3)である．

論理入出力パターンに従ってストレージ省電力を行うことで、実行時に個々のアプリケーションレベルの入出力挙動ごとに適切な省電力化が図られ、アプリケーション実行時にもストレージ省電力を達成することが可能になると考えられる．

4.3 論理入出力パターンを用いたストレージ省電力

アプリケーション協調型大規模ストレージ省電力方式は，前述の論理入出力パターンを用いてデータ配置制御，及びキャッシュを利用した入出力発行制御を行う．

(1) データ配置制御

ストレージデバイスに省電力機能を適用するためには，ストレージデバイスに対する入出力間隔が，省電力機能を適用できる長さ

以上が必要となる。．このために，省電力化が期待できない P3 型のデータは同一のディスク筐体に配置し，残りのデータの入出力発行間隔を省電力機能を適用できる程度に長くすることを試みる．

(2) キャッシュを利用した入出力発行制御

提案手法はキャッシュ上にデータを保持することで、プレロード及び write 遅延による入出力間隔の延伸を行う．

プレロード： データがアプリケーションから read される前にキャッシュにロードする．これにより read 入出力間隔を伸ばす．

Write 遅延： データに対する更新を一時的にキャッシュに蓄積し，まとめてストレージデバイスに書き出すことで write 間隔を伸ばす．

4.4 ストレージ省電力管理機構の実装

アプリケーション協調型大規模ストレージ省電力管理機構とその実装を図２に示す。モニタリング機構は，論理入出力統計，及びストレージ性能・消費電力を収集・管理する．省電力管理機能は，データ要件管理，データ階層・ストレージ階層構築，ファイル配置計算，及びファイル配置変更機能を持つ．実行時省電力機能は，ユーザアプリケーション実行時にアプリケーションに動的にリンクされるストレージ省電力ライブラリ，仮想ファイルツリーファイルキャッシュ，及びストレージ電源制御から構成される．

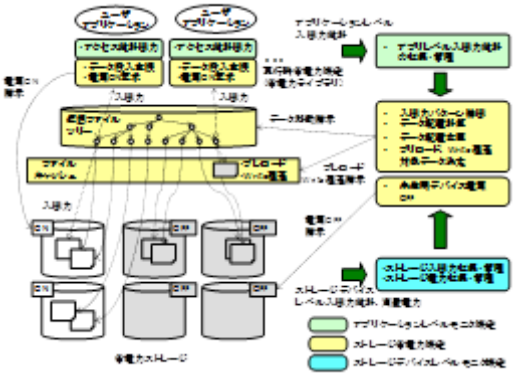

図 2．ストレージ省電力管理機構の実装

4.4 ストレージ省電力管理機構の評価

ストレージ省電力管理機構を実装したストレージ上で商用 DBMS を用いて TPC-H を実行し，ストレージの消費電力とクエリの応答時間を計測した．

4.4.1 評価環境

評価に用いたサーバのプロセッサは Intel Xeon X5670 2.93GHz (合計 24 コア)，主記憶は 48GB である．サーバの OS は Red Hat Enterprise Linux 5.4 (64 ビット版)，ファイルシステムは EXT2 である．ストレージは (株) 日立製作所製の Hitachi Adaptive Modular Storage 2500(AMS2500)を用いた．AMS2500 は 15 台 7200 回転の SATA HDD を 15 台搭載したディスク筐体(13D+2P RAID6 構成)を 11 台，及び RAID コントローラ 1 台を搭載している．RAID コントローラのキャッシュ容量は 2GB，RAID 構成前のディスク筐体の容量は 11.25TB である．サーバとストレージは 4 本の 2Gbit Fibre Channel 1 本で接続されている．

DSS プログラムとして， DSS の代表的ベンチマークである TPC-H ベンチマーク[3]を用いて計測を行った．DB サイズは約 1.2TB(Scale Factor 300)，DBMS バッファサイズは 40GB とした．ログ及び作業表をディスク筐体 1 台に，表と索引を残りのディスク筐体 10 台にキーレンジ分割機能を用いて分散配置した．上記環境において，単一スレッドにて TPC-H のクエリ 1 から 22 までを順次実行し，ストレージの消費電力とクエリの応答時間を計測した．

4.4.2 評価結果

省電力ストレージ管理機構を利用しない場合(省電力制御なし) と省電力ストレージ管理機構を利用した場合(提案手法) のストレージ消費電力の平均値及びクエリ Q2，Q7 の計測結果をそれぞれ図 3 及び 4 にそれぞれ示す．

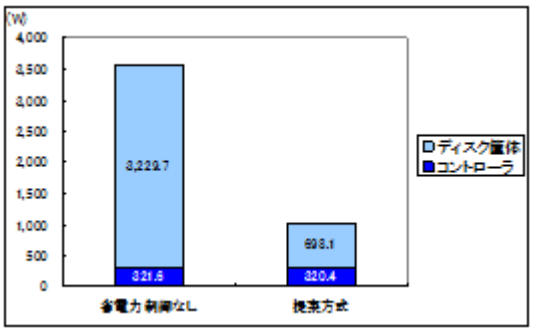

図 3．ストレージの平均消費電力

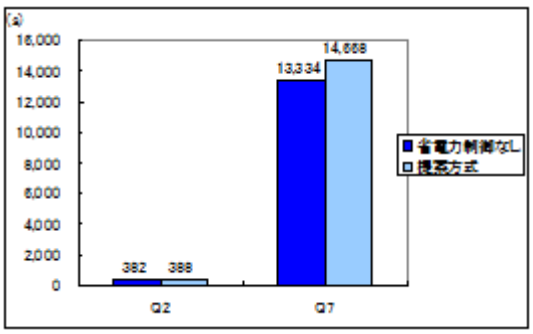

図 4． クエリ応答時間(Q2，Q7)

図から分かるように，提案手法はディスク筐体の平均消費電力を 3229.7 W から 698.1W に約 79%削減できた．これは，入出力が常時行われるデータを 2 台のディスク筐体に集めたことによる効果である．またストレージ応答時間は Q2 はほぼ同等，Q7 が約 9.1%倍増加した．Q7 の応答時間が増加したのは，クエリ実行中のディスク筐体の起動待ち(約 120 秒)，及び入出力が 2 台のディスク筐体

の集約されたことによる入出力応答時間の増加のためである．これらの結果は，提案方式が大規模なストレージ上でも動作することを示している．

５．主な発表論文等

〔雑誌論文〕（計 6 件）

1) N.Nishikawa, M.Nakano, M.Kitsuregawa, “Application Sensitive Energy Management Framework for Storage Systems”, Knowledge and Data Engineering, IEEE Transactions on (Volume:PP , Issue: 99), ISSN :1041-4347
DOI:10.1109/TKDE.2015.2416737, 27 March 2015, Publisher:IEEE

2) 鈴木恵介，早水悠登，横山大作，中野美由紀，喜連川優、SSD を用いた大規模データベースにおける複数問い合わせ処理高速化手法とその評価、日本データベース学会和文論文誌 Vol.13, No.2, pp.19-25, 2015 年 2 月

3) Naho Iimura, Norifumi Nishikawa, Miyuki Nakano, and Masato Oguchi, “A Proposal of Storage Power Control Method with Data Placement Control”, DBSJ Journal, Vol.13, No.1, pp.50-56, March 2015.

4) 石田渉，横山大作，中野美由紀，喜連川優、大規模データベースにおけるアクセス局所性を利用した VM ライブマイグレーションスケジューリング手法の提案と評価，日本データベース学会論文誌，Vol.12, No.1, pp.73-78, 2013

5) 後藤厚宏，クラウド事業者の相互連携でサービス継続性を高めるインタークラウド技術の動向，電子情報通信学会会誌、Vol.197-2, pp.127-132, 2013

〔学会発表〕（計 79 件）

1) Naho Iimura, Norifumi Nishikawa, Miyuki Nakano, and Masato Oguchi:"Evaluation of Data Placement Method in Database Run-Time Processing Considering Energy Saving and Application Performance", GPCDP2014 in conjunction with IGCC2014, Dallas, Texas, USA, 2014/11/3-5

2) Naho Iimura, Norifumi Nishikawa, Miyuki Nakano, and Masato Oguchi:"A Proposal of Storage Power Control Method with Data Placement in an Environment using Many HDDs", IMCOM2015, 9-4, Bali, Indonesia, 2015/1/8-10.

3) Naoko Hishinuma, Atsuko Takefusa, Hidemoto Nakada, Masato Oguchi, “Implementation of Data Affinity-based Distributed Parallel Processing on a Distributed Key Value Store”, IMCOM2014, Siem Rcap, Cambodia, 2014

4) Yuta Nakamura, Shun Nomura, Kyosuke Ngata, Saneyasu Yamaguchi, “I/O Scheduling in Android Devices with Flash Sotrage”, IMCOM2014, Siem Rcap, Cambodia, 2014

5) N.Nishikawa, M.Nakano, M.Kitsuregawa, Energy Efficient Storage management Cooperated with Large Data Intensive Applications, IEEE International Conference on Data Engineering 2012, April 2012, Washington D.C., USA

〔その他〕

ホームページ等

６．研究組織

(1)研究代表者

中野 美由紀(Miyuki Nakano) 芝浦工業大学・教育イノベーション推進センター・教授

研究の統括、省電力データベースのシステム設計

研究者番号：30227863

(2)研究分担者

・小口 正人(Masato Oguchi) お茶の水女子大学・基幹研究院・教授

応用処理・データベース連携省電力技法の設計

研究者番号： 60328036

・山口 実靖(Saneyasu Yamaguchi) 工学院大学・工学部・准教授

データベース・ストレージ連携省電力技法の設計

研究者番号： 50439262

・後藤厚宏(Atsuhiro Goto) 情報セキュリティ大学院大学・情報セキュリティ研究科・教授

応用処理を用いた省電力データベースの評価方式の検討と実施

研究者番号：90558868

・菅谷みどり(Midori Sugaya) 芝浦工業大学・工学部・准教授

応用処理実行時の応用処理性能および省電力結果の蓄積・解析システムの実装

研究者番号：50434288