

産業ビッグデータアプリケーションモデルを用いた RDBMS と NoSQL DBMS の電力効率の比較

西川 記史[†] 高田 実佳^{††} 茂木 和彦[†] 早水 悠登^{†††} 合田 和生^{†††}
喜連川 優^{†††,††††}

[†](株)日立製作所 〒185-8601 東京都国分寺市東恋ヶ窪

^{††}Hitachi America, Ltd. Santa Clara, CA 95054-3103

^{†††}東京大学 〒153-8505 東京都目黒区駒場

^{††††}国立情報学研究所 〒101-8430 東京都千代田区一ツ橋

E-mail: [†]{norifumi.nishikawa.mn,kazuhiko.mogi.uv}@hitachi.com, ^{††}mika.takata@hal.hitachi.com,

^{†††}{haya,kgoda,kitsure}@tkl.iis.u-tokyo.ac.jp

あらまし IoT や AI 技術の利用の拡大に伴うデータセンタのエネルギー消費の急増により、データセンタが提供する IT サービスの性能の維持と IT 機器のエネルギー消費削減の両立が求められている。しかし、IT サービスはそのシステム実装形態により提供できる性能やエネルギー消費が異なるため、それら両方を考慮した実装の選択は IT システム設計者にとって非常に困難な問題である。本研究では、データセンタの主要なアプリケーションである IoT サービスの実装の選択を容易化するために、産業ビッグデータアプリケーションのベンチマーク手法として提案されているアプリケーションモデルを異なる IT システム上に実装し、それらの性能と消費エネルギーを計測しエネルギー効率を Application Energy Effectiveness (AEE) を用いて比較する。これにより異なる実装における IoT サービスのエネルギー効率を計測・比較が可能となる。この結果、IoT サービスの性能とエネルギー効率の双方を考慮した実装の選択が容易になり、IT サービスの性能を大幅に損なわないデータセンタのエネルギー効率向上に貢献できる。

キーワード 省エネルギー、エネルギー効率、データセンタ、IoT、産業ビッグデータアプリケーション、RDBMS、NoSQL

1. はじめに

IoT や AI 技術の利用の拡大に伴うデータセンタのエネルギー消費の急増により、データセンタが提供する IT サービスの性能を維持しつつ IT 機器のエネルギー消費を削減することが求められている。LBNL のレポートによれば、米国のデータセンタでは省電力技術が採用されつつあるが、それにも関わらず米国のデータセンタの消費電力は 2014 年から 2020 年にかけて年率 4% で増加し、2020 年の消費電力は 73 billion kWh に達すると予想されている [1]。また、米国 DOE のレポートでは、IT 負荷が 100 kW 以上のデータセンタのエネルギー使用量を 5 年以内に 25% 削減するとしている [2]。

IT サービスは、実装によりその性能やエネルギー消費が変化する。例えば文献 [3] では、ストレージ内のデータ配置を変更することにより、若干の性能低下はあるもののストレージの消費電力を最大 60% 以上削減できることが、文献 [4] では RDBMS のアクセスパスにより IT 機器の消費電力が異なることがそれぞれ示されている。このように、IT サービスの性能とエネルギー消費の双方を意識した実装の選択は困難であり、その容易化が課題である。

産業アプリケーションの性能を比較するツールとして、産業向けの標準ベンチマークが多数提案されている [11, 12, 22-27]。

これらは単一あるいは少数のコンポーネントの性能を計測するものであり、IoT システム全体の end-to-end の性能を計測するものではない。

また、データセンタのエネルギー効率の計測のための指標として PUE [5] や REF [6]、ITEEsV [7] などが標準化されている。これらの指標は、データセンタの設備や IT 機器のエネルギー効率を計測するためのものであり、IT サービスの性能については考慮されていない。

本研究では、IT サービスの性能とエネルギー効率の双方を意識した実装の選択を容易化することを目的に、産業ビッグデータアプリケーションのベンチマーク手法 [9] として提案されているアプリケーションモデルの異なる実装を対象に、それらの性能と消費エネルギーを計測する。そして、それらの間のエネルギー効率を、サービス視点のエネルギー効率を計測する指標である Application Energy Efficiency (AEE) [8] を用いた比較が可能であることを示す。

これにより、IT サービスの性能と消費エネルギーが計測できること、及び AEE によりアプリケーションの実装の差異によるエネルギー効率の違いを評価できることを示す。この結果 IT サービスの性能とエネルギー効率を意識した実装の選択が容易化され、IT サービスの性能を大幅に損なわないデータセンタのエネルギー効率向上への貢献が可能となる。

2. 産業ビッグデータアプリケーションモデル

本章では、まず我々が計測に用いた産業ビッグデータアプリケーションのモデルの概要について述べる。本モデルは、鉱山の露天掘りにおける工事車両管理アプリケーションにおける問合せをモデル化したものである。詳細は文献 [9] に詳しい。本アプリケーションモデルは、企業の現場業務からバックオフィス業務まで幅広いクエリを含む。

2.1 データライフサイクル

文献 [9] にて示している産業ビッグデータアプリケーションのモデルは、図 1 に示すような 3 つのデータライフサイクルを有している。

最外周のサイクルは戦略サイクルであり、企業幹部やバックオフィスのスタッフが、エンタープライズデータウェアハウスに蓄積されたデータを対象に長時間(日～月)を要する意思決定を行なう。TPC-H [11] や TPC-DS [12] のような従来のアプリケーションモデル(ベンチマーク)は、この最外周サイクルのみを対象としたものである。産業ビッグデータアプリケーションモデルでは戦略サイクルに加え、中央のサイクルである日々のオペレーションサイクル、及び最内周のサイクルである制御/操作サイクルを新たに設けている。日々のオペレーションサイクルで実行されるクエリは、作業量を計測する指標であるオペレーションサイクル数やトラックの移動距離などの KPI を一定時間毎にオペレーションマネージャが確認しそれらに基づき意思決定を行なうことを想定したものである。制御サイクルで実行されるクエリは、効率的な現場リソースの管理のための現場の状態をリアルタイムに変更することが目的であり、イベントの監視や異常値の検出などの業務を行うことを想定したクエリである。

2.2 データ概要

産業ビッグデータアプリケーションで使用されるデータは、産業機器などから取得するオペレーショナルデータ(イベント・センサデータ)とビジネスで使用されるビジネスデータの 2 種類がある。オペレーショナルデータの例としては、GPS データや現時点の産出量などのようにローダーやダンプ、トラックなどの装置に取り付けられたセンサが出力するセンサデータや装置の状態を示すイベントデータがある。ビジネスデータの例としては装置の属性やビジネスプロセスなどがある。

センサデータは定期的(例えば 5 秒間隔)にデータセンタ側のサーバに送信され蓄積される。イベントデータは装置の状態が変化する毎にセンタに送信され蓄積される。装置の状態の継続時間は数十秒から数十分であるため、イベントデータは時系列データではあるがセンサデータと比べて少量である。ビジネスデータはイベントデータやセンサデータと比較すると非常に少ない。

2.3 クエリ概要

産業ビッグデータアプリケーションには、前述の 3 種類のデータライフサイクルに対応した 3 種類のクエリが定義されている。表 1 にその概要を示す。

KPI 計算クエリは、日々のオペレーションサイクルに対応す

表 1 クエリの概要

クエリ種別	クエリ概要
KPI 計算	●操作管理者向け KPI 計算(集計) ●ウィンドウ幅: Sliding, Landmark 及び Tumbling
イベント検知	●フロントオペレータ向けイベント検知
分析	●企業幹部・バックオフィスのスタッフ向けの分析・予測クエリ

るクエリであり、オペレーションマネージャが定期的に KPI を計算することを想定している。KPI 計算には sliding, landmark 及び tumbling の 3 種類のウィンドウがあり、毎分等定期的に実行される。イベント検知クエリは最内側のサイクルである制御サイクルに対応するものであり、検知されたイベントは現場リソース管理に使用される。これらのクエリは直近のデータに対するクエリであり、また短い応答時間が求められるため、一般的にストリームデータ処理基盤で処理される。

分析クエリは最外周の戦略サイクルに対応するクエリであり、状態やイベントの予測、最適化などを含む。これらのクエリは企業幹部やバックオフィススタッフが使用するものである。戦略サイクルにおける意思決定の頻度は日々～毎月一回程度である。分析クエリは過去のデータを対象とした長時間のクエリであるため、一般的に RDBMS や NoSQL DBMS などの蓄積データ処理基盤で処理される。

3. Application Energy Effectiveness (AEE)

Application Energy Effectiveness は、文献 [8] にあるように、ISO/IEC JTC 1/SC 39 にて提案中の、ある IT サービスにおける IT 機器のエネルギー効率を計測する指標であり、以下の式で表される。

$$AEE = \frac{ApplicationOutcomePerHour}{EnergyConsumptionPerHour} \quad (1)$$

ここで、ApplicationOutcomePerHour(以下アプリケーションアウトカムと呼ぶ)とは、作業計画の調整を判断する上で必要な情報を抽出するために実行する一連のクエリが単位時間当たりに完了する回数である。EnergyConsumptionPerHour(以下エネルギー消費と呼ぶ)はその一連のクエリを完了させるために IT 機器が消費した単位時間当たりのエネルギー量である。TPC-H を例とすると、TPC-H の性能指標である QphH をアプリケーションアウトカムとして用いることができる。

4. 産業ビッグデータアプリケーションモデルを用いた IT 機器のエネルギー効率の比較

4.1 計測方法

4.1.1 計測対象

データセンタにおいてデータ分析処理サービスを行う場合、その実装に用いる IT 機器としては、デスクトップ PC から数十コアの CPU を搭載する大型のサーバとエンタープライズストレージを用いる構成まで様々なものが考えられる。また、データ分析に使用されるデータ処理ソフトとしてはこれまで RDBMS が一般的に用いられてきたが、近年では Hadoop を始めとする

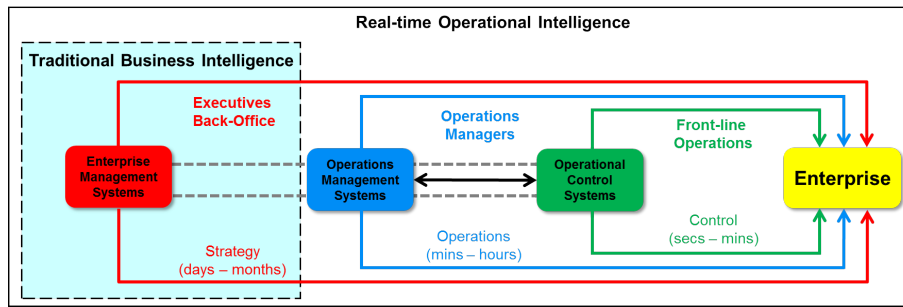


図1 Life-cycle of industrial big data applications.

表2 分析クエリ概要と実装

#	クエリ概要と実装
Q5	シフト終了時の生産量が目標値を満たしているかを予測 単調増加するセンサ値のシフト終了時の値を線形回帰を用いて予測する
Q6	最も高価な車両の利用度を最大化する車両配置の計算 車両のローダへの到着率とローダのサービス率を求め、 車両の平均待ち時間が一定値以下となるように車両を ローダに配置する
Q7	類似センサ値との比較による故障したセンサの検出 センサ値の統計情報をクラスタリングし、故障したセンサ を検出する
Q8	t 時間以内に保守が必要になる装置の予測 単調増加するセンサ値が、 t 時間以内に閾値に達するか否 かを線形回帰を用いて予測する
Q9	保守時に取るべきアクションの推薦 保守に入るまでのセンサ値の推移と最も似ている過去の (保守に入るまでの) センサ値の推移を求め、その時行わ れた保守のオペレータノートを求める

表3 デスクトップ PC + RDBMS

機器	概要
PC	CPU: 8 論理コア/socket × 2 socket (16 物理コア) Intel®Xeon®W5580 CPU @ 3.20GHz Memory: 16 GB HDD: 7.2 krpm 2 TB SATA × 4 台 OS: CentOS 7.2
電力計	横川電機 CW121 クランプ電力計

表4 デスクトップ PC クラスタ + Hive/Hadoop

機器	概要
クラスタ	ノード数 4
インターコネクト	Cisco SF100D-16 (1Gbps)
ノード	CPU: 4 論理コア/socket × 1 socket (4 物理コア) Intel®Core™i5-3470 CPU @ 3.20GHz Memory: 4 GB HDD: 7.2 krpm 450 GB SATA × 1 台 OS: CentOS 7.2
電力計	横川電機 CW121 クランプ電力計

NoSQL 系の DBMS が多く用いられるようになってきている。そこで、本論文では産業ビッグデータアプリケーションモデルを以下の構成で実装し、それぞれの構成のアプリケーションアウトカムと消費電力量を計測、ここから Application Energy Effectiveness (AEE) を計算し両者のエネルギー効率を求めることとした。

- デスクトップ PC + RDBMS
- デスクトップ PC クラスタ + Hadoop/Hive DBMS
- エンタープライズサーバ + ストレージ + RDBMS
- PC サーバクラスタ + Hadoop/Hive DBMS

4.1.2 評価に用いるクエリ

評価には、前章にて述べた産業ビッグデータアプリケーションモデルに定義されている分析系クエリを用いる。これは、KPI 計算及びイベント検知は主にストリームデータ処理基盤で実行される処理であり、データ蓄積基盤のエネルギー効率比較には適さないと考えたためである。文献 [9] に示されている分析系クエリは、表 2 に示す 5 種類である。なお、文献 [9] にはクエリの実装や分析に用いる手法は記載されていないため、表 2 に示す実装及び予測手法を用いる。

4.1.3 アプリケーションアウトカムの定義

AEE を用いた比較を可能とするために、産業ビッグデータベ

表5 エンタープライズサーバ+ストレージ+RDBMS

機器	概要
サーバ	CPU: 12 物理コア/socket × 2 socket (24 物理コア) Intel®Xeon®CPU E5-2690 v3 @ 2.60GHz Memory: 256 GB OS: RHEL 6.6
FC	8 Gbps Fibre channel × 4
ストレージ	15 krpm 300 GB SAS × (256 + 8) 台 (7D + 1P) × 32 RG + 8HDD spares DB を (7D + 1P) × 32 RG 上に構築 (表&索引: 31 RG, ワーク: 1RG)
電力計	横川電機 WT1800

ンチマークモデルのアプリケーションアウトカムを定義する。文献 [9] に示された分析系クエリの評価指標はクエリの応答時間である。そこで、Q5 から Q9 までの処理を直列に実行し、Q5 から Q9 までの処理の 1 時間当りの処理回数を一時間当りのアプリケーションアウトカムと定義と定義する。

4.1.4 ハードウェア及びソフトウェア構成

前述の 4 種類のハードウェア構成をそれぞれ表 3, 4, 5 及び 6 にそれぞれ示す。

ソフトウェア構成は次の通りである。RDBMS として、商用の

表6 PC サーバクラス + Hive/Hadoop

機器	概要
クラス	ノード数 128
インターコネクト	Dell Force 10 Z9000 (10Gbps)
ノード	CPU: 8 物理コア/socket × 2 socket (16 物理コア) Intel®Xeon®E5-2680 CPU @ 2.70GHz Memory: 64 GB HDD(OS): 15 krpm 300 GB SAS × 2 台 HDD(Data): 10 krpm 900 GB SAS × 24 台 OS: RHEL 5
電力計	Ralitan インテリジェントラック PDU

RDBMS である Hitachi Advanced Data Binder^(注1) version 03-06 (HADB) [13] を, Hive/Hadoop DBMS として Apache Hadoop [10] 上に Apache Hive [14] データベースを構築しをそれぞれ用いる. RDBMS 実装では JDBC 経由で SQL により DB からデータを取得し, Apache Spark [15] 上で計算を行う. Hive/Hadoop 実装では HDFS 上に Hive 形式のデータベースを作成し, Apache Spark より Spark-SQL 経由でデータを取得, Spark 上で計算する. クエリ Q5, 7, 8 には Apache Spark の MLlib を, Q9 は時系列データ間の距離を計算するために Dynamic Time Warping (DTW) を用いる.

4.1.5 データ

本評価で用いたデータは, 鉱山の露天掘りにおける工事車両の動きを模したシミュレータにより生成した. シミュレーションに登場する装置は土砂あるいは鉱物資源等をトラックに積み込むローダ, 土砂あるいは鉱物資源を運搬するトラック, 土砂あるいは鉱物資源を降ろすダンプ, の3種類である. シミュレータでは, トラックがローダで資源等を積み込み, それをダンプの場所まで運び, 空になったら再度ローダに行き, 資源等を積み込むサイクルを繰り返している. ローダ及びダンプには生産量の累積値を計測するセンサが, トラックには積載量の累積値, 移動距離の累積値, 現在位置を示す GPS センサがそれぞれ取り付けられており, 一定時間間隔でデータを送信する. また, ローダとダンプの間には複数のフィールドセンサ (Beacon) が置かれており, 近くに来たトラックの ID を送信する. ローダ, ダンプは累積の生産量が一定値を超えたら, トラックは累積の積載量または累積の移動距離が閾値を超えると保守モードに入る.

本評価で生成したデータに含まれるローダ・ダンプの数はそれぞれ 100, トラック数は 2,000 台である. これは, 車両の代理店一店舗の規模に相当する [28]. CSV テキスト形式で約 550 GB であり, イベント・センサデータの小規模表は約 1,800 万行, 中規模表は約 3 億行, 大規模表約 30 億行である. その他のパラメタの詳細は表 7 に示す通りである.

次に, シミュレーションにより生成したデータの傾向を示す. 表 8 はトラック, ローダ, ダンプそれぞれのアクティビティ(イ

表7 シミュレーションパラメタ

#	Parameters	Values
1	# of activities of hauler	6
2	# of activities of loader	4 (one is idle)
3	# of activities of dump	4 (one is idle)
4	# of maintenance event for hauler (maintenance 1, 2)	2
5	# of maintenance event for loader	1
6	# of maintenance event for dump	1
7	# of field sensor per loader/dump pair	20
8	# of equipment sensor per hauler	1 (total carried object weight)
9	# of object sensor per hauler	1 (load weight per shift)
10	# of GPS sensor per hauler	1
11	# of field sensor per loader/dump	1 (includes #7)
12	Distance of two field sensors	90.5
13	Moving equipment max speed	8 (randomly selected, equal or less than 8) per 5 sec
14	Load 1 & 3 event length	Load1 = 30 sec, Load3=15 sec
15	Dump1 & 3 event length	Dump1 = 20 sec, Dump3=35 sec
16	Load weight	>0, 10 x (2 + Gauss distribution (mean 0.0, standard deviation 1.0))
17	Load2 & dump 2 length	value of load weight sec
18	Hauler skip dump frequency (for event detection 4a)	1/300
19	Hauler skip specified field sensor frequency (4b)	1/100
20	Hauler maintenance 1 occurrence condition	Total moved distance exceeds 100,000
21	Hauler maintenance 2 occurrence condition	Total weight exceeds 900
22	Length of maintenance	One hour to one day (randomly selected)

表8 Event time length (sec.).

#	Event	Max.	Min.	Avg.
1	Queuing at loader	87,027	0.0	2,548
2	Loading	119	46	65
3	Hauling Full	1,745	1,276	1,483
4	Queuing at Dump	12,688	0.0	749
5	Dumping	129	56	75
6	Hauling Empty	2,433	538	1,033
7	Loading 1	30	30	30
8	Loading 2	74	1	20
9	Loading 3	15	15	15
10	Load Idle	9,922	1	327
11	Dumping 1	20	20	20
12	Dumping 2	74	1	20
13	Dumping 3	35	35	35
14	Dump Idle	257,591	1	566
15	Hauler Maintenance 1	259,196	3,601	116,521
16	Hauler Maintenance 2	259,195	3,600	112,193
17	Loader Maintenance	259,184	3,602	119,868
18	Dump Maintenance	259,190	3,633	122,006

表9 Maintenance interval (hr.).

#	Event	Max.	Min.	Avg.
1	Hauler Maintenance 1	2,099.2	1.3	467.1
2	Hauler Maintenance 2	2,039.9	0.6	540.6
3	Loader Maintenance	510.6	3.7	100.7
4	Dump Maintenance	529.5	4.1	110.2

イベント) とその長さを示している. #1-6 はトラックの, #7-10 はローダの, #11-14 はダンプのイベントである. #15-18 はトラック, ローダ, ダンプの保守イベントである. 表 9 は保守イベントの間隔を示している. Hauler maintenance 1, hauler maintenance 2, loader maintenance 及び dump maintenance の平均間隔は, それぞれ 467.1 時間, 540.6 時間, 100.7 時間, 110.2 時間である.

(注1) : 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サービスの実証・評価」(中心研究者: 喜連川 東大教授/国立情報学研究所所長)の成果を利用

4.2 IT 機器のエネルギー効率の初期評価

アプリケーションアウトカムの計測は、前述のクエリ Q5 から Q9 までを単一スレッドで直列に連続で一定時間実行し、その間に処理できたクエリの数を実測することにより行う。IT 機器の消費エネルギーは、前述のクエリ Q5 から Q9 を実行している間のサーバ及びストレージの消費電力量から求める。

図 2 に、産業ビッグデータアプリケーションの性能と消費エネルギーの計測結果、及び AEE の比較結果をそれぞれ示す。AEE は式 1 に示すように、単位時間当たりのアプリケーションアウトカムを単位時間当たりの消費エネルギーで除した値である。

図 2(a) から分かるように、アプリケーションアウトカムはそれぞれデスクトップ PC+RDBMS が 1.04、デスクトップ PC クラスタ+Hive/Hadoop が 0.99、エンタープライズサーバ+ストレージ+RDBMS が 15.86、PC サーバクラスタ+Hive/Hadoop が 4.79 であった。エネルギー消費量は図 2(b) に示すように、それぞれデスクトップ PC+RDBMS が 0.21、デスクトップ PC クラスタ+Hive/Hadoop が 0.23、エンタープライズサーバ+ストレージ+RDBMS が 2.7、PC サーバクラスタ+Hive/Hadoop が 30.5 であった。図 2(c) に、それぞれの実装での AEE の比較結果を示す。AEE の値は、デスクトップ PC+RDBMS が 4.96、デスクトップ PC クラスタ+Hive/Hadoop が 4.33、エンタープライズサーバ+ストレージ+RDBMS が 5.87、PC サーバクラスタ+Hive/Hadoop が 0.16 であった。

これにより、AEE を用いることで実装毎の IT サービスの視点での電力効率を比較することが可能であることが示され、電力効率を考慮した IT サービスの実装が可能となる見込みを得ることができた。

4.3 PC サーバクラスタの挙動調査

AEE により異なる実装間での IT サービス視点での電力効率の違いを比較できることを示したが、図 2(c) に示すように PC サーバクラスタ+Hive/Hadoop 構成の AEE は他の構成の AEE と比較して著しく低い。これは、図 2(a) に示すように、同構成のアプリケーションアウトカムの値が、コア数から想定される値と比べて著しく低いことが原因である。そこで、同構成の挙動について調査を行った。その結果を図 3 及び図 4 に示す。

図 3 は 128 台の PC サーバクラスタの、クエリ実行中のノード毎の CPU 使用率の推移である。ここでの値は %user と %system の合計値である。図から分かるように、ある一つのノード (midori004) のみ CPU の高騰が続いており、これに引っ張られてクエリの応答時間が劣化したことが分かる。なお、CPU 使用率が 50% となっているのは、ノード当り 16 コア中 8 コアのみを Spark のエグゼキュータに割り当てたためである。

次に、図 4 にクエリ実行時に各ノードに割り振られた Apache Spark のタスク数を示す。図から分かるように、ある一つのノード (midori004) にのみタスクが集中していることが分かる。このノードは CPU 使用率の高騰が長く続いたノードと同一である。タスク一つのノードに集中した原因は、Spark job の中に特定の分割キーを対象とするものがあり、その特定のパーティションを格納しているノードに処理が集中したためと考えられる。

表 10 RDBMS 構成のクエリの特性.

Item	Q5	Q6	Q7	Q8	Q9
# of read record (KB)	864	6,016	864	864	11,872
read size (MB)	26	367	26	26	362
response time (s)	14.6	0.9	4.1	70.3	59.7
read type	scan	scan	scan	scan	index access
read throughput per core (KB/s)	77	18,342	277	16	-
# of processed record per core	-	-	-	-	8,239
required core #	6	1	1	29	24
required HDD #					256

る。これによりアプリケーションアウトカムが想定ほど伸びず、結果 PC サーバクラスタの AEE が悪くなったと考えられる。

5. 考察

本章では、RDBMS 構成と Hadoop 構成それぞれで、ポテンシャルを最大限に引き出した場合の程度まで AEE を向上できるのかについて考察する。これらの考察結果は、今後 AEE の目標値を決定する上での参考とすることが可能である。

5.1 AEE の推定

まず我々は、前章での計測結果を用いて、AEE がどの程度まで向上できるのかを推定した。AEE の値はクエリに求められる性能により変化するので、ここではクエリの応答時間の目標値を 60 秒として見積を行った。

5.1.1 RDBMS 構成のパラメタ見積り

まず RDBMS 構成で性能要件を満たすために必要となるコア数及び HDD 数を見積る。これにはエンタープライズサーバ+ストレージ+RDBMS の計測結果を用いた。表 10 に、クエリ毎の特性を示す。表から分かるように、クエリはスキャン系のクエリ (Q5-8) と索引を使ったランダムアクセスを行うクエリ (Q9) に大別される。前者のコア当りの read 性能は 16 KB/s から 18,342 KB/s である。ばらつきが大きいのは演算処理の負荷がそれぞれ異なるためである。後者のコア当りのレコード処理件数は 8,293 件であった。この結果から、RDBMS 構成で前述の 60 秒の性能要件を満たすためには、コア数は 29 個必要であることが分かる。また、Q9 の応答時間が約 1 分であることから必要な HDD 台数はエンタープライズサーバ+ストレージ+RDBMS の構成と同じ 256 台と見積れる。

5.1.2 Hadoop 構成のパラメタ見積り

次に Hadoop 構成で前述の性能要件を満たすために必要となるコア数及び HDD 数を求める。これには、まずデスクトップ PC クラスタ+Hive/Hadoop 構成での実測値よりコア当りの read 性能と HDD 当りの read 性能を求め、表 11 に示すクエリ特性と合わせて必要となる CPU 数及び HDD 数を計算する。計算の結果、必要コア数は 352、HDD 数は 88 であった。

5.1.3 消費電力量の推定

次に、前節で求めた CPU コア数及び HDD 数より、RDBMS 及び Hadoop に必要となる機器の消費電力を推定する。

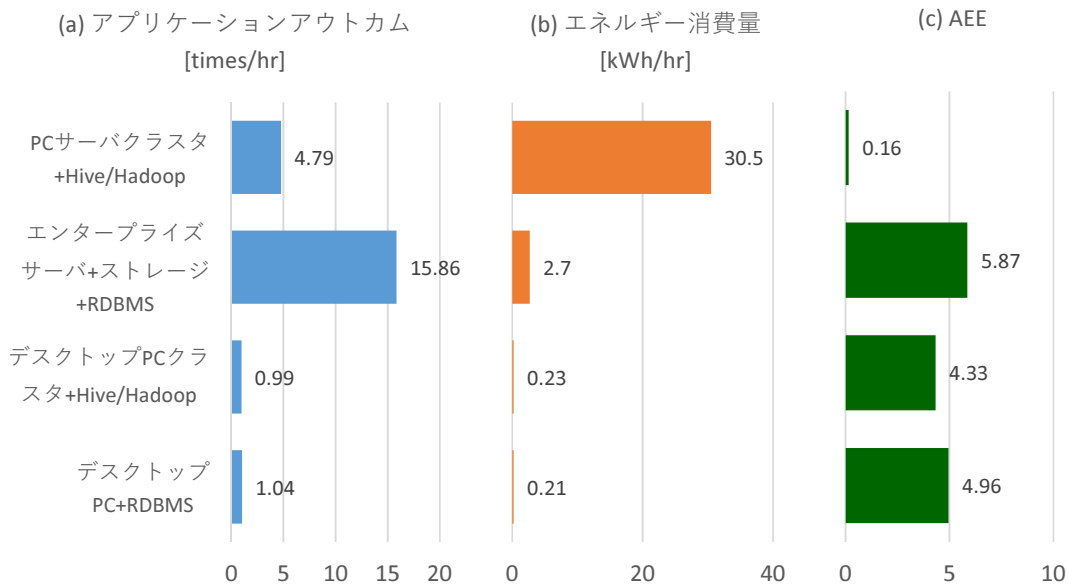


図2 エネルギー効率の初期評価結果.

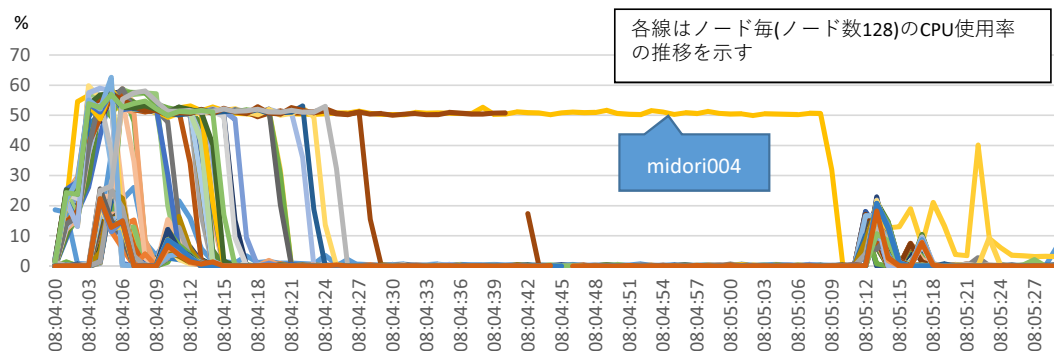


図3 ノード毎 CPU 使用率推移.

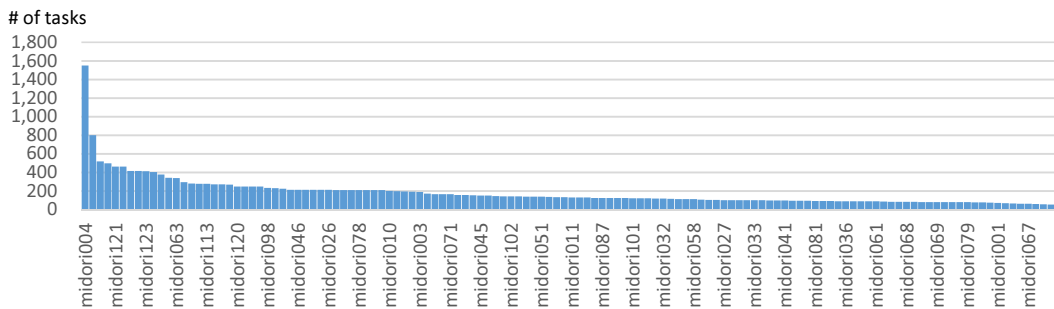


図4 ノード毎 Spark タスク数.

(1) RDBMS 構成

RDBMS 構成では、サーバとストレージの構成を、それぞれ 29 コアエンタープライズサーバと 256 台の HDD を持つストレージとした。サーバの消費電力が CPU 数に比例するとした場合、別途計測した 48 コアのエンタープライズサーバの消費電力は約 0.29kW であったことからサーバの消費電力は約 0.15 kW と見積れる。次にストレージの消費電力であるが、これは前章の計測で用いたストレージと同等であることからその消費電力は 2.44 kW と推定できる。これらの結果から、RDBMS 構成の一時時間当りの消費電力量は 2.59kWh/hr と計算できる。

(2) Hadoop 構成

次に、Hadoop 用構成の消費電力量を計算する。Hadoop 構成では 352 コア、88 HDD が必要であるが、これは 4 章の比較で用いた HADB 用のデスクトップ PC(16 コア、4HDD)16 台分に相当する。HADB 用デスクトップの消費電力は 0.21kW であったことから、Hadoop 構成での一時間当りの消費電力量は 4.6 kWh/hr と計算できる。

5.1.4 アプリケーションアウトカムの推定

表 10 の処理件数・データ処理量とコア数・HDD 数とから、RDBMS 構成の 5 クエリの合計応答時間は約 126.2 秒と計算できる。これより RDBMS 構成のアプリケーションアウトカムは 28.5 となる。また、表 11 のデータ処理量とコア数・HDD 数と

表 11 Hadoop 構成のクエリの特性.

Item	Q5	Q6	Q7	Q8	Q9
read size (MB) (w/ partition)	32	244	309	309	221,610
read size (MB) (wo/ partition)	11,337	1,029	110,805	221,610	221,610
# of processed recode per core	1	1	1	1	352
required core # (w/ partition)	6	1	1	29	24
required HDD # (w/ partition)	1	1	1	1	88

から、Hadoop 構成のクエリ応答時間はパーティション有りの場合で約 60.3 秒と計算できる。これより Hadoop 構成のアプリケーションアウトカムは 59.7 となる。

5.1.5 AEE の推定

前節までの一時間当たり消費電力量とアプリケーションアウトカムより AEE を計算すると、RDBMS 構成、Hadoop 構成それぞれ、 $28.5/2.59 = 11.0$ 、 $59.7/4.6 = 12.3$ となる。

5.2 AEE の改善

前節での見積りはストレージに HDD を使用した場合のものであるが、近年ストレージの主流となりつつある SSD を用いることにより、消費電力量をさらに削減でき、AEE を大幅に改善できると考えられる。そこで、我々はストレージに SSD を用いた場合の AEE を見積った。見積りに使用した SSD は、RDBMS 構成がエンタープライズで用いられる SSD であるインテル®SSD DC P4600 シリーズ、4.0TB (ランダムリード 225,000 IOPS, read 時 9.6W, アイドル時 5W 未満)、Hadoop 構成が安価なデスクトップ用 SSD であるインテル®SSD Pro 7600p シリーズ (3230MB/s, アクティブ時 50mW, アイドル時 25mW) である。

RDBMS 構成、Hadoop 構成それぞれについて、ストレージをこれらの SSD に置き換えた場合、RDBMS 構成、Hadoop 構成とも SSD 数は 1 台である。これらの構成での一時間当たりの消費電力量は RDBMS 構成が 0.157 kWh, Hadoop 構成は 1.9 kWh/hr となる。

これらの結果から、RDBMS 構成、Hadoop 構成それぞれの AEE は 181.5, 31.4 と計算できる。この結果は、ストレージを SSD に置き換えることにより、IT サービスの視点でも大幅に電力効率が向上できることを示している。

6. 関連研究

RDBMS と NoSQL の比較としては、これまで主に性能比較に関する研究が報告されている。例えば文献 [16] では、MySQL と MongoDB を用いて IoT ビッグデータに対する CURD 処理の性能を比較している。また、文献 [17] では、Yahoo! Cloud Serving Benchmark (YCSB) を用いて RDBMS と MongoDB を、文献 [18] ではモバイルアプリケーションを用いて MongoDB と PostgreSQL を比較している。文献 [19] では、複数の NoSQL データベースについて、性能やスケラビリティ等複数の指標で評価を行っている。

産業向け標準ベンチマークとしては、例えば TPC-H [11,22] や TPC-DS [12,23] が、Transaction Processing Performance Council (TPC) により開発されている。近年ではビッグデータシステムをベンチマーキングするためのツール [24,25] の提案や、ハードウェア・OS 及び商用の Hadoop File System API を計測する TPC Express Benchmark for Hadoop System (TPCx-HS) [26] がリリースされている。また、NoSQL に対するベンチマークとしては Yahoo! Cloud Serving Benchmark [27] が広く知られている。我々が使用した産業ビッグデータアプリケーションモデルは、バックエンドオフィスのみならず、オペレーションマネージャや現場の制御/操作に関するクエリも含んでいる。

エネルギー効率の比較の研究では、近年では仮想化環境やクラウド環境での比較が報告されている。文献 [20] では、異なるワークロード、ハイパバイザ、サーバの組合せ毎の性能とエネルギー消費を比較し、同一サーバ・ワークロードであってもハイパバイザによって電力・エネルギー特性が異なることを示している。文献 [21] ではクラウドにおける HPC のスケジューリング方式の電力効率について比較している。

7. まとめ

本研究では、産業ビッグデータアプリケーションモデルを、種類の異なる IT 機器及びデータ処理ソフトを用いて実装し、その性能及び消費エネルギーを計測した。そして、それら異なる実装間のエネルギー効率を Application Energy Effectiveness (AEE) を用いて比較できることを示した。また、RDBMS 構成、Hadoop 構成でのポテンシャルが最大限に引き出されたことと仮定した場合の AEE の推定を行うとともに、ストレージに SSD を用いることにより、AEE を大きく改善できる見込みがあることを示した。

これにより、産業ビッグデータアプリケーションモデルが IoT システムのデータベース周辺の実装の性能・消費エネルギー比較に使用できること、及び AEE を用いることで、アプリケーションの実装の差異によるエネルギー効率の違いを評価できることを確認した。この結果、IT サービスの性能とエネルギー効率の双方を考慮した実装の選択が容易にできるようになり、IT サービスの性能を大幅に損なわないデータセンタのエネルギー効率向上への貢献が可能となる。

謝 辞

本研究の一部は、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) の委託業務「エネルギー・環境新技術先端プログラム／革新的な省エネルギー型データベース問合せコンパイラの研究開発」及び「IoT 推進のための横断技術開発プロジェクト／先進 IoT サービスを実現する革新的省エネルギー型ビッグデータ基盤の研究開発」の結果得られたものです。

文 献

- [1] A. Shehabi 他, United States Data Center Energy Usage Report, LBNL-1005775, Ernest Orland Lawrence Berkeley National Laboratory, <https://eta.lbl.gov/publications/united-states-data-center>

- energy, 2016.
- [2] FEMP Data Center Program Overview, U.S. Department of Energy, https://energy.gov/sites/prod/files/2016/07/f33/dc_improvement.pdf, 2016.
- [3] N. Nishikawa, et al., Application Sensitive Energy Management Framework for Storage Systems, *IEEE Transactions on Knowledge and Data Engineering*, Volume 27, Issue 9, pp.2335–2348, 2015.
- [4] 早水他, ストレージ消費電力特性に基づく関係データベース演算子の省電力指向コストモデル, 第9回データ工学と情報マネジメントに関するフォーラム (DEIM 2017), G1-3, 2017.
- [5] Information technology - Data centres - Key performance indicators - Part 2: Power usage effectiveness (PUE), ISO/IEC 30134-2:2016, 2016.
- [6] Information technology - Data centres - Key performance indicators - Part 3: Renewable energy factor (REF), ISO/IEC 31034-3:2016, 2016.
- [7] Information technology - Data centres - Key performance indicators - Part 4: IT Equipment Energy Efficiency for servers (ITEEsv), ISO/IEC 30134-4:2017, 2017.
- [8] 五十嵐他, アプリケーションを含めた IT サービスのエネルギー効率指標の提案, 第10回データ工学と情報マネジメントに関するフォーラム (DEIM 2018), C7-3, 2018.
- [9] U. Dayal, et al., An Approach to Benchmarking Industrial Big Data Applications, *Big Data Benchmarking: 5th International Workshop, WBDB 2014*, 2014.
- [10] Apache Hadoop, <http://hadoop.apache.org/>, 2017.
- [11] Transaction Processing Performance Council (TPC), TPC-H benchmark specification, <http://www.tpc.org/tpch/>, 2008.
- [12] R. O. Nambiar, et al., The making of TPC-DS, in *Proc. of the Int'l Conference on Very Large Data Bases (VLDB)*, 2006, pp. 1049–1058.
- [13] 高速データアクセス基盤 Hitachi Advanced Data Binder プラットフォーム, <http://www.hitachi.co.jp/products/it/bigdata/platform/data-binder/>
- [14] Apache Hive, <https://hive.apache.org/>, 2017.
- [15] Apache Spark, <https://spark.apache.org/>, 2017.
- [16] J. Y. Seo, et al., Performance Comparison of CRUD Operations in IoT based Big Data Computing, *International Journal on Advanced Science Engineering Information Technology*, Vol.7, No. 5, pp.1765–1770, 2017.
- [17] A. Floratou, et al., Can the Elephants Handle the NoSQL Onslaught?, *Proceedings of the VLDB Endowment*, Vol. 5, No. 12, 2012.
- [18] M. Fotache, et al., NoSQL and SQL Databases for Mobile Applications. Case Study: MongoDB versus PostgreSQL, *Informatica Economica*, 17(2), pp.41–58, 2013.
- [19] S. Venkatraman, et al., SQL Versus NoSQL Movement with Big Data Analytics, *International Journal of Information Technology and Computer Science*, 2016, 12, pp.59–66, 2016.
- [20] C. Jiang, et al., Energy Efficiency Comparison of Hypervisors, 2016 Seventh International Green and Sustainable Computing Conference (IGSC), 2016.
- [21] S. Goyal, et al., Experimental Comparison of Three Scheduling Algorithms for Energy Efficiency in Cloud Computing, 2014 IEEE International Conference on Cloud Computing in Emerging Markets, 2014.
- [22] M. Poess, et al., New TPC Benchmarks for Decision Support and Web Commerce, *SIGMOD Record*, vol. 29, no. 4, pp. 64–71, 2000.
- [23] M. Poess, et al., Why you should run TPC-DS: A workload analysis, in *Proc. of the VLDB Endowment (PVLDB)*, pp. 1138–1149, 2007.
- [24] T. Rabl, et al., Big Data Generation, in *Specifying Big Data Benchmarks*, ser. LNCS. Springer, vol. 8163, pp. 20–27, 2014.
- [25] Y. Zhu, et al., BigOP: Generating Comprehensive Big Data Workloads as a Benchmarking Framework, in *Database Systems for Advanced Applications*, ser. LNCS. Springer, vol. 8422, pp. 483–492, 2014.
- [26] Transaction Processing Performance Council, TPCx-HS, www.tpc.org/tpcxhs/, 2015.
- [27] Yahoo Cloud Serving Benchmark, <https://research.yahoo.com/news/yahoo-cloud-serving-benchmark/>, 2010.
- [28] 三輪, 世界中で稼働する建機を遠隔モニタリングする KOM-TRAX, セキュアな IT が創るつながる新産業創出セミナー, http://www.chubu.meti.go.jp/b34jyoho/shiryo/20140214yugosecurityseminar/20140214_komatsu.pdf, 2014.