

次世代グリーンデータセンターの全体像

Overview of Next-Generation Green Data Center

● 久門耕一

あらまし

近年、急速に需要が拡大するデータセンターの消費エネルギーと運用コストの増大が問題となっている。著者らはこれらを低減することができるデータセンターアーキテクチャを検討し、「次世代グリーンデータセンター」と呼ぶ新しいシステムアーキテクチャの研究開発を開始した。本研究開発においては、装置レベルで実現される省エネルギー化、高機能化の限界を超えた性能を実証することを目標とし、装置役割の見直し、機能の垂直統合と再配分のほか、従来は別個に考えられていた給電や冷却といったファシリティ技術についても情報機器との統合設計を行うことで、全体を通じた重複機能の除去や、機能のミスマッチに起因する無駄の排斥により、エネルギー効率が高く、運用コストを低減できるデータセンターの実現を目標とする。具体的には、資源のプール化、ハードウェアのミドルウェア化、コモディティ活用、ファシリティを含む最適化、全体の統合運用という五つの手法により、ヒューマンセントリックでインテリジェントなICT社会の基盤を支えるデータセンターに向けた技術の開発を目指す。

Abstract

There has been a rapid growth in demand for data centers, and in recent years this has led to the problems of the increased amount of energy they consume and their higher operational costs. To solve these problems, we are developing a new data center architecture called "Next-Generation Green Data Center," which conserves energy and also reduces operational costs. In this development, we aim to realize specifications beyond those which can be achieved by optimizing equipment. Facilities are also optimized simultaneously with ICT equipment to omit duplicated functions, to amend the mismatched interfaces. To achieve this development, we concentrated on five areas: resource pooling, hardware as middleware, utilization of commodities, facility ICT co-optimization, and unified management. We hope the system will help bring about a Human-Centric Intelligent Society.

まえがき

クラウドコンピューティングの進展に伴い、インターネット上に流通する大量のデータの蓄積・加工・利用が行われ、さらに多くの情報を生み出し、流通するようになってきた。このようなインターネット上の膨大なデータに対する情報処理の中心的役割を果たすのがインターネットデータセンター（IDC）である。

IDCは大量のICT機器から構成される大規模な情報処理施設であるが、処理需要の伸びに対応するための能力増強の結果、全産業の中でIDCの消費する電力が突出した伸び率を示すという予想がある。このため、IDCの電力は将来のエネルギー消費抑制を考える上で重要な問題となっている。

そこで著者らは、今後のIDCの省電力化を実現するため、次世代グリーンデータセンターと呼ぶ新しいデータセンターアーキテクチャを提案・開発することにした。

本稿では、次世代グリーンデータセンターを方向付ける以下の五つの基本方針（図-1）について背景を含めた説明を行う。

- (1) 資源のプール化
- (2) ハードウェアのミドルウェア化
- (3) コモディティ活用

- (4) ファシリティー一体最適化
- (5) 全体統合運用

なお、次世代グリーンデータセンターとして、現在試作中のシステムの詳細については、本誌掲載の「次世代グリーンデータセンターを構成するシステム：Mangrove」で述べる。

クラウド時代のデータセンターにおける課題

地球温暖化対策として各機器の省エネルギー化が図られデータセンター全体としても省エネルギー化が行われてきた⁽¹⁾しかし、データセンターの需要は今後も伸び続けるため、ICT機器が消費する電力は2025年には国内総発電量の20%、世界では総発電量の15%に達すると予想されている（図-2）⁽²⁾

このため、一層の省エネルギー化を実現するための技術開発が強く求められてきている。

また、エネルギー消費の問題だけではなく、処理能力拡大のためのデータセンターの大規模化により、下記の三つのコストが問題となってくる。

- (1) 装置コスト
- (2) エネルギーコスト
- (3) 管理コスト

現在のところ、装置コストは、高コストパフォーマンスの装置の導入によって、従来のコストのま

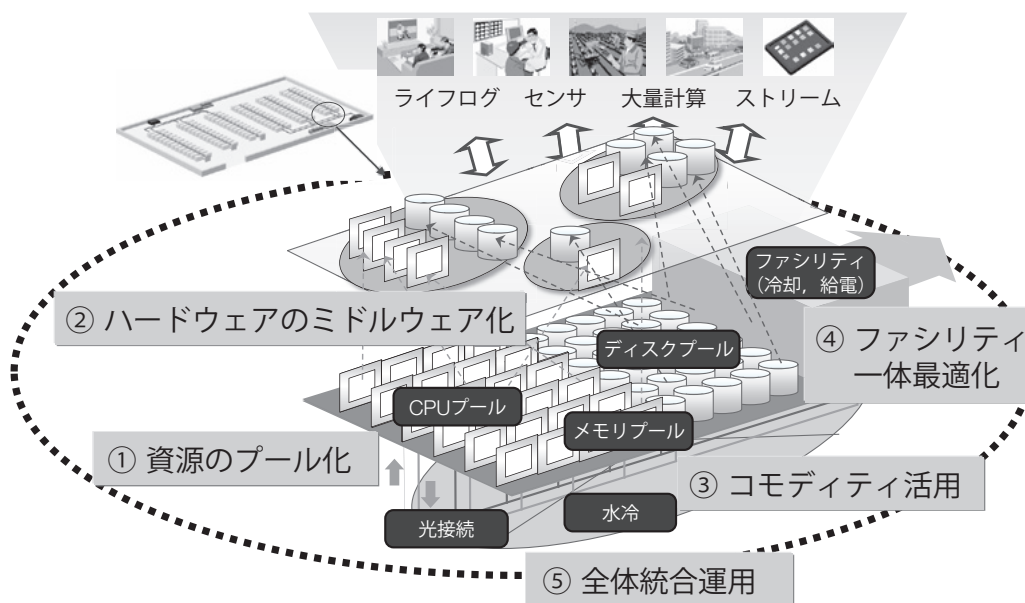


図-1 次世代グリーンデータセンターの全体像
Fig.1-Overview of Next-Generation Green Data Center.

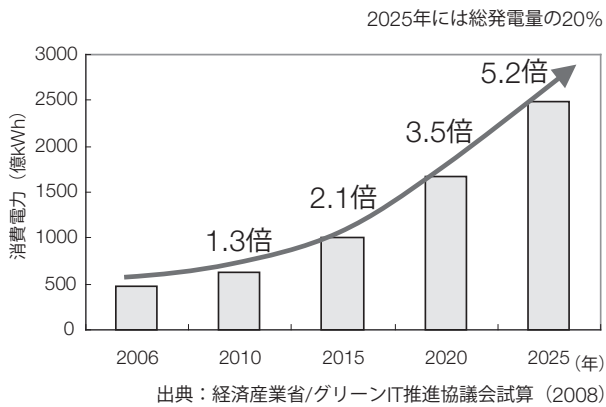


図-2 ICT機器の消費電力予想

Fig.2-Estimated power consumption of ICT equipment.

ま性能を向上させることができていた。しかし、機器コストは抑えられても、装置台数が増えることで電力消費量は増え続けている。さらに今後は、電力供給手段の世界的な見直しによってエネルギー単価が上昇する可能性が高いため、省エネルギー化に対するニーズはますます高まると予想される。

また、装置の運用管理コストについては、装置数が増えることで、システムの複雑度が増し、管理が複雑になる結果、従来の技術のままでは、管理コストが台数比例以上の割合で増えていく可能性があり、データセンターの処理性能向上の大きな障害となる。

データセンターの持つこれらの課題をより最適に解くには、ICT機器とファシリティ機器の協調など、IDC全体をシステムとして最適化することが必要となる。

次章では、ICT機器の機能構成見直しについて、具体的な例で説明する。そして、解決する手段としてプール化アーキテクチャを提案する。

資源のプール化

資源のプール化とは、次世代グリーンデータセンターを特徴付ける考え方で、単純な機能ユニットをプールと呼ぶ場所に多数用意し、必要に応じてプールから切り出して割り当て、ネットワーク経由で他ユニットと接続することによって目的を達成するシステムを構成するものである。このようなアーキテクチャをプール化アーキテクチャと呼ぶ。

呼ぶ。

従来、データセンターでは、構築するシステムの目的に合わせ、サーバのメモリ量やCPU数、ストレージ容量やスループット、ネットワーク冗長性や伝送速度といった性能や機能を設計し、最適な機器を選んで購入・構築してきた。

これに対し、クラウドサービス向けIDCは、装置の種類を少なくし、同規格の装置を複数用いることで、サービスを複数サーバやストレージに分散実行し、性能向上や信頼性向上を実現する、いわゆる、スケールアウト型ソフトウェアと組み合わせで運用される。

同一の装置を多数用いることで、故障時に他装置に処理を肩代わりさせることが容易になる一方、少数の規格に統一してしまうため、すべての目的に対し、最適な構成を取ることができず、場合によっては提供ができないサービスや、実現できてもコストが高くなってしまう場合も起き得る。

これに対し、プール化アーキテクチャでは、プールを構成するユニットの機能は、サーバ装置やストレージ装置よりも小さく単純にすることが前提となっている。例えば、サーバの場合では、CPUとメモリから成るマザーボードや、HDD単体などがプールの構成要素として選ばれる。以下、具体的なケースを基に、小さい単位でプールを作ることによって、用途に適した構成をとることができることを述べる。

クラウド向け処理としてよく使われるHadoopを実行するシステムのアーキテクチャを例として検討する。Hadoopは、多くのサーバがストレージから並列にデータを読み出して処理することで高性能を実現する。一般に、サーバにローカル接続したHDDがストレージとして用いられ、各サーバはHDDから並列に独立にデータを読み出して処理することができる。一方、ローカルHDD接続のため、HDD本数、容量がハードウェア設置時に固定され、自由に変えることができない。

これに対し、ストレージとして、サーバに従来型の共用ストレージ装置を接続する構成の場合には、HDDの持つ物理的な容量の壁は隠蔽することができるが、ストレージ装置の処理性能限界から、複数サーバへのデータの高スループット提供が難しく、性能上のボトルネックとなる。ま

た、Hadoop処理自身が多重化による信頼性確保を行うことができるため、ストレージ装置がRAID (Redundant Array of Inexpensive Disks) 技術により提供する高信頼性はメリットとは言えず、逆に冗長化による容量低下がデメリットとして現れる。こういった問題は、Hadoop処理の場合は、サーバ装置、ストレージ装置の役割分担が適切ではないことが原因で起きる。

一方、一般の業務アプリケーションやDBMSを動作させるシステムの場合には、ストレージシステムが提供する仮想化機能や信頼性向上が重要であり、ストレージ装置をローカルHDDで代替すると可用性、信頼性ともに低下する。また、ストレージ処理にCPU能力があまり必要とされないため、一つのコントローラに接続するHDDの台数を多くすることが可能で、全体コストの低減を図ることができる。

このように、Hadoopと一般の業務アプリケーションでは、それぞれ異なった構成が最適となるが、サーバとストレージ双方の装置内部構成を比較すると、その違いはそれほど大きくない。サーバ装置、ストレージ装置とも、CPUとメモリから成るCPUボードと、データを格納する多数のHDDという2種のユニットが主要な構成要素である。ハードウェア上の違いは、ストレージ装置の接続可能HDD数が、サーバよりも多いことであり、機能面では、ストレージ装置にはストレージ機能および運用管理を実現するファームウェアが搭載され、ストレージとしての機能を提供していることが特徴である。

このように、HDD接続台数の違い、ファームウェア搭載の違いを踏まえ、現在開発中のプール化アーキテクチャでは、CPUとメモリから成るCPUボードプールと、HDDから構成されるHDDプールの2種類のプールを用意し、双方のプール間を自由に接続するため、DAN (Disk Area Network) と呼ぶシステム内インタコネクタを用意した。DANを用いることにより、サーバ機能やストレージ機能が動的に構成可能になる。

プール化されたハードウェアに対し、特定の機能を付与する役割を負うのが、装置上で実行されるミドルウェアである。このような仕組みを著者らは、「ハードウェアのミドルウェア化」と呼んで

いる。つぎにこれを説明する。

ハードウェアのミドルウェア化

現在のほぼすべてのICT機器には、操作性や機能の向上のため、ファームウェアと呼ばれる組込みソフトウェアが搭載され、機器内で動作している。ファームウェアはソフトウェアとは異なり、ユーザがインストールしたり、実行したりすることを意図していないため、開発環境や内部の仕様は公開されない。多くの場合、ファームウェア動作環境には、組込み機器用OSが使われるため、APIの違いなどから一般のオープンソースを利用したソフトウェア開発が難しく、このような閉鎖的環境のため、開発者が限定されている。

汎用のOS上でミドルウェアとして開発することが、オープンソースの活用や幅広いソフトウェア開発者の利用を可能にし、技術の開発期間の短縮と開発コストを低減させる可能性にもつながる。

ストレージの場合を例にとると、柔軟なストレージ管理を実現する、オブジェクトストレージと呼ばれる技術があり、最近注目されている。格納するデータの特性を上位のアプリケーションからストレージに伝えてもらうことにより、データ配置などの最適化を行うことができる。オブジェクトストレージを実現するオープンソースも出ており、すでに大規模でスケーラブルなファイルシステムを実現する際に用いられている。

このように、機能をソフトウェアライブラリ化し、プール化アーキテクチャとともに用いることで、装置機能を実現することをハードウェアのミドルウェア化と呼ぶ。

この考えと類似の考え方に、バーチャルプライベートクラウドがある。これは、従来プライベートクラウドとして提供されてきたものを共通ハードウェア上のソフトウェアとして実現するものである。

次世代グリーンデータセンターではこの考えを、サーバやストレージなどの基本的な装置の構成にまで広げ、同一プールから様々なタイプのサーバやストレージ機能を生成できることを目指している。

コモディティ活用

クラウド向けのデータセンターでは、サーバは特殊な装置ではなく構成部品の一つとして扱われている。これは、サーバが低価格で高性能なコモディティベースで作られるようになったため可能になった。

しかし、コモディティであるPC用に開発された技術がサーバに転用されたケースは、Intel社のCPUをはじめ数多くあるが、全く同じ部品を使うケースはあまりない。サーバに必要な信頼性要件などがコモディティであるPCとは異なるからである。

ただし、コモディティ製品の技術進展は速く、性能と機能だけで比較すると、コモディティ製品の方が上回るものが数多くある。例えば、Flash ROMによるSSD (Solid State Drive) は、耐衝撃性に優れることからノートPCで最初に使われ始めたが、ランダムアクセス性能で比較すれば、サーバ向け高性能ディスクをはるかに上回る性能を示す。このような性能をデータセンター用途で使おうとする場合、問題となるのが信頼性である。

24時間連続稼働を仮定しないコモディティ製品には十分な寿命を持たないものも多く、品質にもばらつきがある。従来、信頼性は部品、基板、装置、システムと順番に積み上げて確保するものとして扱われてきたため、信頼性に劣るコモディティ部品を何も対処せず使ったのでは、システムの信頼性を確保できない。一方、部品レベルの信頼性を高めようとするれば、部品コストが急激に高くなる可能性がある。つまり、コモディティを使いつつ、低価格で高信頼なシステムを構成するには、コモディティの使いこなし技術という独立した技術が必要となる。

使いこなし技術の典型的例が、現在ではサーバ向け技術として広く使われるRAID技術である。RAIDの本質は、冗長性を使い信頼性を向上させることと、同種のを複数並べることで性能を向上させるスケールアウトであり、ほかにも応用の利く考え方である。

使いこなし技術として重要な点に、障害モードの理解がある。RAIDの場合、HDDのデータは正常に読めるか、全く読めないかの二者択一である

ことを仮定し、障害モードではデータが読めないことを前提にしている。データが間違っ読まれる障害モードを想定するならば、RAID1やRAID5では正しい動作が保証できない。このように、障害の発生条件、発生頻度、偶発故障か摩耗故障か、一時的か永久的か、障害を完全に隠蔽するか否かなど、機能を利用する側の観点も用いる必要がある。コスト低減は、必要十分な信頼性を明確にした上で実現することが重要である。

次世代グリーンデータセンターにおいては、コモディティ化されてきた光インタコネクタ技術はDANに要求される、高バンド幅と中距離(30 m程度)を接続するインタコネクタに用いる。また、HDDには、低価格かつ大容量なSATAディスクの使用も想定し、信頼性確保の手段を検討している。

ファシリティ一体最適化

従来、ICT機器にエネルギーを供給する給電装置と、機器の排出する熱エネルギーを外部に排出する冷却装置は、建屋設備として考えられ、必要量を確保すること以上の機能は求められてこなかった。しかし、IDCの省電力化を考える上では、全体の20～50%以上のエネルギーを消費する、冷却と給電ロスを削減することは重要な課題である。

次世代グリーンデータセンターでは、ファシリティ部で消費するエネルギー削減には、給電、冷却ともICT機器との接続方法の見直しによりエネルギーの変換回数を削減することが必要と考えている。例えば、給電インタフェースでの見直し対象は、交流/直流給電、給電電圧の選択、停電時の信頼性確保方法などで、これらの見直しによりエネルギー消費の無駄を低減する。冷却インタフェースの見直しにおいては、チップで発生した熱を輸送する熱媒体、高温動作による冷却電力の低減を目指したICT機器の部品レベルでの動作温度見直しに相当する。また、熱輸送効率が高い水を熱媒体として利用するため、機器に従来とは異なる排熱インタフェースの利用を検討したり、生成された熱を、廃熱として外部に排出するのではなく、熱エネルギーとして利活用する技術の検討も行っている。

ICT機器とファシリティ機器を統合して最適化することは、従来のファシリティと装置との関係を大きく変える。ファシリティは、物理現象に基づ

く動作であるため実現できることに強い制約がつくが、仮想化技術を用いればICT上の処理を実行する装置の物理位置を集中や分散させることが自由にできる。ファシリティの持つ情報を使いこのような指示を与えることにより、消費するエネルギーを最小にすることが可能になる。つぎの章では、このような制御を含む全体管理を行う統合運用管理について述べる。

統合運用管理

本プロジェクトでは、機能を要素に分解し再統合することが無駄の削除と柔軟性の獲得につながるということを基本的考えに据えている。しかし、期待される効果を得るためには、大量の資源の最適配置、ファシリティを含む装置の動作監視と制御を行う運用管理システムが必要となる。

一方、システム運用者の視点からは、従来のシステムからの連続性があることが望ましい。そこで、第一段階としては、プール化アーキテクチャを隠蔽し、あたかも通常の装置が無数に並んでいるというビューを与えることからスタートする。この場合でも、動作監視、障害調査、保守対応の機能は必要である。

ファシリティ制御については、リアルタイム多点温度センサ技術、IDC全体高速熱シミュレーション技術、IDC消費電力シミュレーション技術を組み合わせることで、最適な処理性能を実現する構成を簡単に管理者が検討できるツールを用意していく。

最適運用については、研究が始まったばかりであるが、すでに従来型ハードウェアに対するVM (Virtual Machine) 配置で、コストに基づく配置最適化を行えば、単純にVMを配置するのに比べてコスト低減が可能であるという評価結果が得られている。さらに、プール化アーキテクチャが提供するハードウェア的な柔軟性を加えれば、従来よりも多様なサービスを低コスト、省エネルギーで、柔軟に提供できるようになると著者らは考える。

今後、与えた制約内でポリシーに基づき最適な

配置と制御により処理を実行するために、富士通の持つICT技術を最大限に使い、グリーン化を図るための技術開発を行っていく予定である。

む す び

次世代グリーンデータセンターは、従来の装置の役割分担を見直すことで、無駄の排斥による省エネルギー、低コスト化と柔軟性の確保を行おうというプロジェクトである。今回紹介したシステムは、機器役割の見直しとして、サーバとストレージ、およびICT機器とファシリティを対象にした。その結果、CPUとメモリから成るCPUプール、ディスク群から成るディスクプールの二つのプールを持つシステムを実現することにした。しかし、どの機能をどのレベルまで分解し再構成することが適切であるかは、分解により得られるメリットと、その時点で利用可能な技術、とくにインタコネクタ技術の利用可能性に依存する。例えば、複数のサーバでPCI-eを共通化するI/Oプールは、必要性が高ければ、現時点の技術で作ることができる。一方、CPUとメモリを分離し、プール化することは、数10 ns以下の遅延と、数100 Gバイト/秒のスループットを実現するCPUメモリ間インタコネクタが必要となり、現時点ではコストとサイズの両面で実現が困難である。

プール化実現に必要な要素技術開発は今後も進展するため、プールする対象を随時見直すことにより、さらに効率良いアーキテクチャを実現し、将来のICTを支える基盤作りに役立てたいと考える。

参考文献

- (1) 小田切 充ほか：データセンターの省エネ動向と富士通グループの取組み. *FUJITSU*, Vol.61, No.3, p.235-240 (2010).
- (2) 経済産業省商務情報政策局：グリーンITイニシアティブ (第2回), 2008年5月.
<http://www.meti.go.jp/committee/materials/downloadfiles/g80520c03j.pdf>

著者紹介



久門耕一（くもん こういち）

ITシステム研究所 所属

現在、次世代グリーンデータセンターシステム、および高性能計算機システムの研究開発に従事。