

# 大規模VM負荷予測・配置制御技術によるシンククライアント・データセンターのグリーン化

中村 暢達・喜田 弘司  
竹村 俊徳・藤山 健一郎

## 要 旨

仮想マシン（VM）型のシンククライアントシステムでは、仮想化されたPCを任意の物理サーバ上で実行できるため、VMを適切にサーバ間で移動させて、余剰サーバ（VMを動作させなくてよくなったサーバ）の電源をオフにすることで省エネが可能となります。これに向けて、数万台のVMから負荷情報をきめ細かく収集、サーバの利用状況を把握し、VMをどの物理サーバで実行させれば良いのかを予測・制御する「大規模VM負荷予測・配置制御技術」を開発しました。本稿では、技術開発への取り組みと、6万台規模のVM運用で消費電力を約35%削減できることをシミュレーションにより検証した事例を紹介します。

## キーワード

●データセンター ●大規模データ収集・分析 ●仮想マシン ●産業 ●業務 ●ICT

## 1. はじめに

近年、企業におけるIT機器の管理あるいは業務/サービスの管理（内部統制やセキュリティへの対策を含む）を効率化する方法として、データセンターの活用が注目されています。データセンターにおいて統合的に管理することで、部門ごとに行っている情報管理の手間を削減するだけでなく、経営の見える化など企業活動の効率化が期待されています。

一方、地球温暖化への取り組みとして、IT機器の省電力化が強く求められています。データセンターの消費電力は5年間に2倍のペースで増加しており<sup>1</sup>、CO<sub>2</sub>排出量削減に向けた効率化が急務となっています。ここで、我々はデータセンターの運用管理に注目しました。データセンターにおいては、処理要求の需要は、昼と夜、あるいは平日と休日などで大きく変動するため、処理能力に余裕をもたせてサーバ運用がなされます。そのため、一般的なデータセンターにおけるサーバの稼働率は8~12%ともいわれています。このサーバの稼働率を向上させる運用管理手法を確立することで、データセンターのグリーン化を実現できると考えました。

本稿では、仮想マシン（VM）型シンククライアントシステム

のデータセンターを対象としたグリーン運用管理について紹介します。VM型シンククライアントシステムとは、データセンター側でVMを動作させ、シンククライアント側ではユーザとのインタラクションのみを行うシステムです。OSやアプリケーションの管理はデータセンター側で一括して行うことが可能となるだけでなく、VMを任意のサーバ上で実行させる機能があります。これにより、VMを適切に集約あるいは分散させることで、耐障害性や省エネルギー性に優れた運用が可能となります。

第2章より、VM型シンククライアントから収集・解析した負荷情報に基づき、データセンターで稼働しているサーバの電源を適切にオン/オフを行うなどのグリーン運用管理を実現するソフトウェア技術の詳細について説明します。

## 2. VM型シンククライアントシステム

### 2.1 VM最適配置の課題

サーバ負荷の変化に合わせて負荷の小さいサーバ上のVMを

<sup>1</sup> 米国環境保護庁（EPA）調べ（2008年）

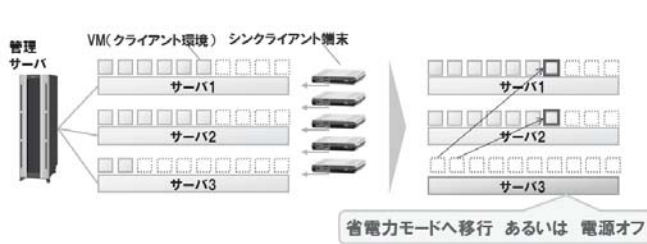


図1 VMの最適配置の例

移動させ、特定のサーバに集めることで、VMがまったく動作していない空きサーバをつくり、そのサーバの電源をオフにするといった制御を行うことで、無駄な電力の削減を実現できます。例えば、図1に示すように、サーバ3で稼働しているVMをサーバ1とサーバ2に移動させることで、サーバ3は空きサーバとなります。そして、サーバ3を省電力モードあるいは電源オフにします。

このとき、VMの負荷をできるだけ正確に把握する必要があります。もし、移動先のサーバで、VMの負荷の合計がサーバの処理能力を超えてしまうと処理速度の低下などの障害が発生します。一方、サーバの処理能力を余らせると、サーバの稼働率が低下し、無駄が生じます。また、VMを移動させるにも処理負荷及び時間がかかりますので、頻繁に移動させることはできません。移動処理後のしばらくは再度VMを移動させることのないように、VMの負荷変化をできるだけ正確に予測することが重要です。

更に、どのVMをどのサーバに移動させれば良いかの配置制御に関し、VM数、サーバ数が多くなると、配置の組合せ数が爆発的に増大し、最適な配置制御を求めることが難しくなります。サーバ3,000台、VM 60,000台の場合、1回の配置制御当たりの移動先候補数は1.8億通りとなり、この中から最適な移動先を見つける計算には多大な時間が必要です。

また、VMの負荷状況は時々刻々変化しますので、その状況変化に応じて、VMを移動させることが望ましいです。そのため、負荷予測及び配置制御に要する処理はできるだけ軽量かつ高速であることが求められます。

## 2.2 VM負荷予測・配置制御方式

図2に、開発したシステムの全体構成を示します。複数のサーバがあり、各サーバには最大50台のVMが稼働します。

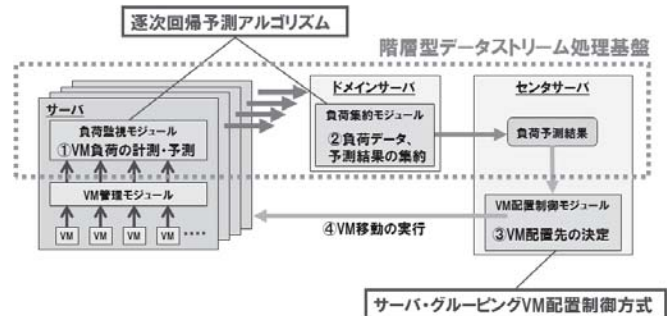


図2 VM負荷予測・配置制御の全体構成

サーバ250台ごとにドメインサーバがあり、システム全体で1台のセンタサーバがあります。サーバにおいて各VM負荷の計測・予測を行い、ドメインサーバにおいて負荷データ、予測結果の集約を行い、センタサーバで統合してVM配置先の決定を行います。サーバからドメインサーバを介してセンタサーバへ、データが処理されながら集約される機構となっており、これはNECが開発した階層型データストリーム処理基盤<sup>1)2)</sup>を用いて実現されています。また、VM負荷データの収集・予測処理及びVM配置制御の高速化・軽量化のために、それぞれ逐次回帰予測アルゴリズム、サーバ・グルーピングVM配置制御方式を開発しました。次に開発した個別技術の詳細を説明します。

データストリームとは、比較的小さいサイズの連続して発生するイベント、ログなどのデータです。センサ、端末などからのデータ量が爆発的に増大すると、サーバでこれらデータストリームを収集し、リアルタイムに解析して活用することは困難です。このような課題に対し、データストリーム処理基盤では、小サイズデータに対する連続的な処理に適したメモリ管理、複数の処理モジュール間でのスムーズなデータの受け渡しを実現するフロー管理を支援します。

VMからの負荷データもデータストリームの1つとみなすことができ、負荷データの収集・予測の処理に、データストリーム処理基盤を適用することで、大規模システムにおける処理の高速化を実現しました。

逐次回帰予測アルゴリズムでは、前に行った計算処理の計算途中の状態を保存しておき、それを使って差分の処理のみを行うことで、処理の高速化・軽量化を実現します。つまり、一般的な回帰予測手法では、データベースから必要データのすべてを読み込み、近似直線 $Y = aX + b$ に当てはめて予測値を計

算します。一般的には、近似直線の変数 $a$ は次のように計算します。

$$a_{n+1} = f(x_{m+1}, y_{m+1}) + \dots + f(x_n, y_n) + f(x_{n+1}, y_{n+1})$$

これに対し、開発した方式では、処理の途中結果として、「 $a_n, f(x_m, y_m), \dots, f(x_n, y_n)$ 」を保持し、次のデータとこの途中結果を使って次のように計算します。

$$a_{n+1} = a_n - f(x_m, y_m) + f(x_{n+1}, y_{n+1})$$

このような手法を用いることで、計算処理量を約1/3に抑え、高速かつ軽量の予測処理を実現しました。

サーバ・グルーピングVM配置制御方式では、負荷状態に近いサーバ同士をグルーピングし、移動先となるサーバの候補数を大幅に削減することで、VMの配置計算を行います。具体的には、次の手順で処理を行います。

- 1) CPU使用率でサーバをグループ分けします。図3では、10のグループに分けています。
- 2) あらかじめ設定した閾値に基づいて移動元サーバを決定します。負荷分散が必要な高負荷のサーバ及び空きサーバとするための低負荷のサーバを選択します。
- 3) 各グループの中から指定台数分（図3では3台）の代表サーバを選定します。
- 4) 代表サーバを移動先の候補とみなして、VMの移動先サーバを決定します。高負荷を回避する際、できるだけ負荷が平準化するように、また低負荷サーバを空きサーバにする際、できるだけ負荷が集中するようにします。
- 5) 残りのサーバに対して、1)～4)の処理を繰り返します。

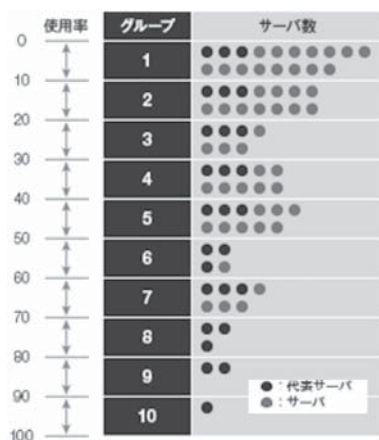


図3 サーバ・グルーピングVM配置制御方式

### 3. 評価実験と考察

上述した開発技術を評価するために、システムを試作し、稼働サーバ数をどのくらい抑制できるか、消費電力をどのくらい削減できるかの実験を行いました。

実際に多数のシンクライアント及びサーバを用意することは困難なため、シミュレーション環境を構築し、15台のPC上で、サーバ3,000台、VM60,000台相当の実験環境を用意しました。Webやデータベースなどのアプリケーションの典型的な負荷パターンを9種類用意し、各VMで負荷を模擬的に発生させました。低負荷及び高負荷とするサーバのCPU使用率の閾値は、40%及び80%とし、200秒間隔で配置制御を実施するという想定で、サーバ台数削減率を測定しました。サーバ台数削減率とは、「電源オフが可能なサーバ数÷全サーバ数」です。また、サーバ台数削減率がいくら高くとも、高負荷障害を発生させては意味がありません。そのため、高負荷発生率（CPU使用率80%以上の時間÷サーバ稼働時間）も測定しました。

図4に実験結果を示します。横軸は時間、縦軸はサーバ台数削減率です。6～8時間経過時は、システムの負荷が増大しているため、サーバ台数削減率は低下していますが、それ以外の時間帯では、約40%のサーバの電源をオフにできています。全体では、CPU使用率が29.1%に対して、サーバ台数の削減率は37.9%となりました。一方、高負荷発生率は約1.0%でした。

この実験結果から、全システムをブレードサーバ（Express5800/120Ba-4）で運用した場合の消費電力を計算します。筐体の消費電力は360W、ブレードサーバの消費電力は119W+（44W×平均CPU使用率）となりますので、全サーバ（3,000台）稼働時の消費電力は、筐体：360W×500、ブレードサーバ：（119W+44W×29.1%）×3,000から、計575,412Wで

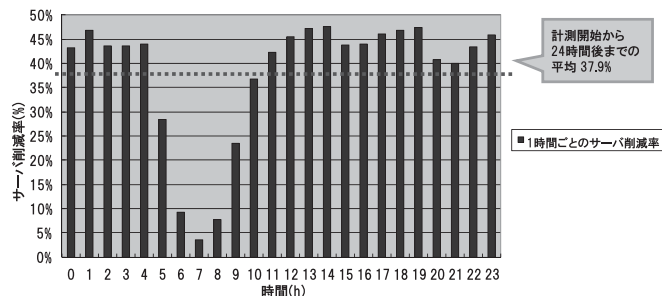


図4 実験結果

す。一方、サーバ37.9%削減時の消費電力は、筐体：360W×310、ブレードサーバ：(119W+44W×46.9%)×1,860から、計371,352Wです。削減量は約204kW、約35%の削減となります。

#### 4. おわりに

本稿で説明したVM負荷予測技術は、数万台規模のVMシンクライアントを一括管理するデータセンターにおいて、大量のVMの負荷データをリアルタイムに収集して、サーバ全体の状態を把握することを可能にします。更に、VM配置制御技術で、サーバ負荷の変化に合わせて負荷の小さいサーバ上のVMを特定のサーバへ集めることで、VMが動作していない空きサーバをつくり、そのサーバの電源をオフにするといった制御を行うことで、無駄な電力の削減を実現します。

これらの技術の有効性は、シミュレーションによる検証実験で確認しました。データセンターにおける省エネルギー運用管理を促進し、より実効性の高いグリーンITの実現に向けて研究開発を更に進めていきます。

なお、本成果の一部は、(独)新エネルギー・産業技術総合開発機構(NEDO技術開発機構)から委託を受けて実施した「エネルギー使用合理化技術戦略的開発事業」における研究開発の成果です。

#### 参考文献

- 1) 喜田弘司他、「データストリーム処理による大規模プローブカーシテムの開発と評価」、情報処理学会ITS研究会、2008年9月
- 2) K. Fujiyama etc., "High-Performance Load Forecasting on Large-scale VM-type Thin Client System using Data Stream Processing Technology", Int. Conf. on Communication Systems, Networks & Applications, 2007年10月

#### 執筆者プロフィール

中村 暢達  
サービスプラットフォーム研究所  
主任研究員

喜田 弘司  
サービスプラットフォーム研究所  
主任

竹村 俊徳  
サービスプラットフォーム研究所  
主任研究員

藤山 健一郎  
(独) 情報通信研究機構