

## 仮想マシクラスタにおける リソースの負荷バランスに関する一考察

豊島 詩織†

山口 実靖‡

小口 正人†

†お茶の水女子大学

‡工学院大学

### 1 はじめに

コンピュータシステムにおいて利用可能な情報量が爆発的に増大し、それに伴う IT コストも大きな問題となっている。こうした状況から、リソース使用状況に合わせてスケーラブルなシステム構築が可能なクラウドコンピューティングへの期待が高まっている。しかしそのメリットを考えても既にクラスタシステムを所持していたり、そこに大半のデータを置いている企業はいきなりクラウドリソースに移行することは難しいと考えられる。そこで本研究では使用しているクラスタのシステム状況をモニタリングし、負荷が重い場合は外部のクラウドリソースを使用するといった、手元のクラスタからクラウドのリソースを動的に利用するシステムの実現を目指す。主にデータインテンシブアプリケーションの動作を想定しているため、ストレージアクセスに焦点を当て実験をすすめる。

### 2 研究背景

#### 2.1 仮想マシン PC クラスタ

クラスタシステムには PC クラスタのワーカノードに仮想マシンを配置した仮想マシン PC クラスタ、仮想化ソフトには Xen [1] を使用した。仮想マシン PC クラスタとすることで、動作中のアプリケーションの状態を維持したまま仮想マシンを別のノード上へマイグレーションしたり、システム負荷やサービスの需要に応じた、リソースの柔軟な調達が可能となる。

### 3 遠隔リソースの利用

#### 3.1 クラウドコンピューティング

現在注目を集めているクラウドコンピューティングは、インターネットを介してユーザが計算機やデータをサービスとして利用するという考え方である。ユーザはリソースを利用するだけでよいシステム運用コストが不要で、必要なときに必要な分だけリソースを利用することができる。そのため運用・管理コストの削減が可能となるほか、システム規模が予測しづらいつきなど、現在のシステム使用状況に合わせてキャパシティを増減できるといったメリットもある。今後は遠隔のデータセンタやクラウドコンピューティングの利用がますます増えると考えられる。

#### 3.2 IP-SAN

ストレージネットワークには SAN を使用した。SAN は、分散したストレージをネットワークで統合し、スト

レージの集中管理とディスク資源の効率的な活用を可能にする。特に IP-SAN は汎用機器を用いて構築可能であり、また専用網も含め広範囲に IP ネットワークのインフラが整備されているため長距離接続が可能となる。

IP-SAN のプロトコルとしては iSCSI (Internet Small Computer System Interface)[2] を使用した。iSCSI は SCSI コマンドを TCP/IP パケットの中にカプセル化することでブロックレベルのデータ転送を行う。Gigabit Ethernet/10Gigabit Ethernet が広く普及していくであろうことを考慮すると、IP-SAN をバックエンドに持つ PC クラスタ多くが使用されるようになって考えられる。

### 4 研究概要

#### 4.1 実験環境

表 1: Experimental setup : PCs

OS	initiator :Linux 2.6.18-53.1.14.el5(CentOS5.3)
CPU	initiator : Intel (R) Xeon(TM) 3.6GHz target : Intel (R) Xeon(TM) 3.6GHz
Main Memory	initiator(Domain0) : 2GB initiator(DomainU) : 2GB target : 4GB
iSCSI	initiator : iscsi-initiator-utils target : iSCSI-Enterprise-Target
Monitoring Tool	Ganglia

クラスタの各ノードのスペックを表 1 に示す。ストレージ通信には iSCSI を使用し、リモートアクセス環境を構築するため、ローカルサイトとリモートサイト間に人工的な遅延を挿入する Dummynet を使用し往復遅延 (RTT) を挿入した。また各計算ノードには DomainU(virtual machine) を一つずつ配置した。

#### 4.2 既存研究

既にクラスタシステムを使用している利用者がクラウドを使うことを考えた場合、まずクラウド内のストレージ機能から利用することが考えられる。文献 [3] では、構築した仮想マシン PC クラスタにおいてストレージ機能を遠隔サイトで利用するシステムを構築し、並列データマイニングの HPA[4]、データベースベンチマークの OSDL-DBT3 (Open Source Development Labs Database Test-3)[5] を動作させ評価を行なった。

HPA は大量のトランザクションデータを処理するデータマイニングではあるが、全体として CPU バウンドであるため RTT が大きくなって実行時間にあまり差は見られなかった。

一方 OSDL-DBT3 は連続的なデータベースへのアクセスが発生する I/O インテンシブなアプリケーションで

A Study of load balancing of resource in virtual machine PC cluster

† Shiori Toyoshima, Masato Oguchi

‡ Saneyasu Yamaguchi

Ochanomizu University (†)

Kogakuin University (‡)

あり, 実験からも RTT が大きくなるにつれて実行時間が增大することが確認された。

上記の既存研究より OSDL-DBT3 のような I/O バウンドなアプリケーションでは性能の劣化が著しい。このことから本実験では高遅延ネットワークを介したクラウド環境上にデータがある場合, 同一 LAN 上に仮想マシンごとアプリケーションをマイグレートし, 負荷分散とストレージアクセス最適化を行う手法を提案し, 評価を行った。

### 5 実験結果と考察

図 1 にマイグレーション前の実験環境を示す。iSCSI 遠隔アクセス環境を想定し, RTT1msec, 2msec, 4msec, 8msec, 16msec を挿入した環境で実験を行なった。まずそれぞれの RTT において仮想マシンをローカルサイトからリモートサイトにマイグレートする時間を測定した。結果を図 2 に示す。RTT0msec から 4msec までは 21 秒となり, 最も大きい RTT16msec の環境では 52 秒という結果になった。図 3 にマイグレートにかかる時間とその後 OSDL-DBT3 を実行した合計時間を示す。比較対象として既存研究において DomainU にジョブを与えた際の実行時間を示す。この結果から RTT が長くなるにつれて既存研究との差が大きくなる事が分かる。

アプリケーション実行はサーバをリモートサイトにマイグレートすることで遅延無しの iSCSI 環境で行なうことができる。そのため RTT が長くなるにつれてサーバをデータが配置されているリモートサイトにマイグレートした後計算処理を行なう効果が大きい。

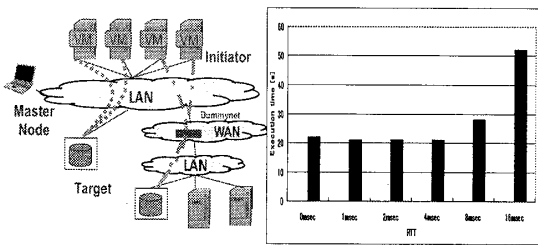


図 1: 実験環境 (マイグレーション前)

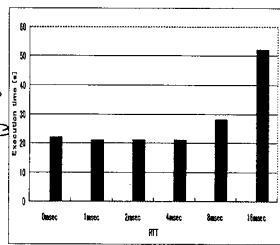


図 2: マイグレーション時間

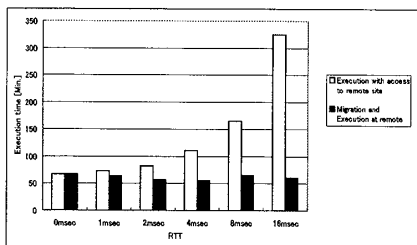


図 3: 実行時間

### 6 Amazon EC2

本研究ではこれまでクラウドコンピューティングを模擬するため Dummynet によりリモートアクセス環境を構築していた。今後は実際のクラウドサービスである Amazon

EC2(Elastic Compute Cloud)[6] を使用し, EC2 とクラスタでリソースを協調する実験を行なうため, EC2 とクラスタを接続しネットワークスループットを測定した。スループットは上りと下りが非対称な結果となった。

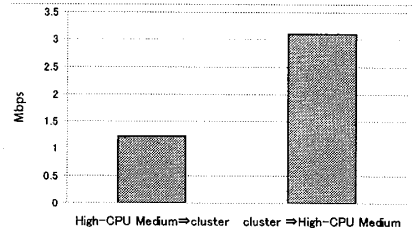


図 4: EC2 とクラスタ間の Network Throughput

### 7 まとめと今後の課題

I/O インテンシブなアプリケーションにおいて高遅延ネットワーク越しにデータに iSCSI 遠隔ストレージアクセスを行ったところ, iSCSI 遠隔アクセスにおいて RTT が大きくなると実行時間が長くなる事が確認された。このためローカルサイトの仮想マシンを遠隔サイトへマイグレートしストレージアクセスと計算処理を実行したところ, 本研究で使用した I/O インテンシブなアプリケーションでは, リモートストレージにアクセスする負荷が大きい場合はサーバをリモートサイトに再配置し, そこで計算処理を行なう本手法は有効であることを確認した。

今後はクラスタの負荷に応じて自動的に仮想マシンを遠隔サイトにマイグレートし, 動的に負荷分散を行うシステムを構築する。また仮想的に作り出したクラウド環境ではなく, Amazon EC2 といった実際のクラウド環境で実験を行なっていく。

本研究は一部, 文部科学省科学研究費特定領域研究課題番号 18049013 によるものである。

#### 参考文献

- [1] Xen : <http://www.xen.org/>
- [2] iSCSI RFC: <http://www.ietf.org/rfc/rfc3722.txt>
- [3] 豊島詩織, 山口実晴, 小口正人: "仮想マシンマイグレーションによるストレージアクセス最適化に関する性能評価", CPSY, 信学技報, Vol.109, No.296, CPSY2009-37, pp.13-18, 京都, 2009 年 11 月。
- [4] 小口正人, 喜連川優: "ATM 結合 PC クラスタにおける動的リモートメモリ利用方式を用いた並列データマイニングの実行", 電子情報通信学会論文誌, Vol.J84-D-I, No.9, pp.1336-1349, 2001 年 9 月
- [5] OSDL-DBT3: <http://ldn.linuxfoundation.org/>
- [6] Amazon Elastic Compute Cloud: <http://aws.amazon.com/ec2/>
- [7] 谷村勇輔, 小川宏高, 中田秀基, 田中良夫, 関口智嗣: "仮想クラスタに対する IP ストレージの提供方法の比較", 「ハイパフォーマンスコンピューティングとアーキテクチャの評価」に関する北海道ワークショップ (HOKKE), 2007 年