T2R2 東京科学大学 リサーチリポジトリ
Science Tokyo Research Repository

論文 / 著書情報
Article / Book Information

論題(和文)	仮想クラスタを用いたData-Intensive Application 実行環境の 性能モデル構築と最適化に向けて
Title(English)	
著者(和文)	佐藤賢斗, 佐藤仁, 松岡聡
Authors(English)	Kento Sato, Hitoshi Sato, Satoshi Matsuoka
出典(和文)	先進的計算基盤システムシンポジウム 2008, , ,
Citation(English)	Symposium on Advanced Computing Systems and Infrastructures 2008, , ,
発行日 / Pub. date	2008, 8
権利情報 / Copyright	ここに掲載した著作物の利用に関する注意: 本著作物の著作権は（社）情報処理学会に帰属します。本著作物は著作権者である情報処理学会の許可のもとに掲載するものです。ご利用に当たっては「著作権法」ならびに「情報処理学会倫理綱領」に従うことをお願いいたします。 The copyright of this material is retained by the Information Processing Society of Japan (IPSJ). This material is published on this web site with the agreement of the author (s) and the IPSJ. Please be complied with Copyright Law of Japan and the Code of Ethics of the IPSJ if any users wish to reproduce, make derivative work, distribute or make available to the public any part or whole thereof.
Note	本原稿は著者最終版です。

Powered by T2R2 (Science Tokyo Research Repository)

仮想クラスタを用いたData-Intensive Application実行環境の性能モデル構築と最適化に向けて

佐藤賢斗[†]　佐藤　仁[†]　松岡　聡[†,††]

1. はじめに

近年、急速化する情報爆発に伴い科学技術計算で扱う実験データの大規模化の傾向があり、またこれらのデータは地理的に分散された環境で広く共有されるため、効率的なデータ管理が重要となっている。一方、仮想化技術の進展に伴い仮想マシン (VM) によるレガシーアプリケーションのサポート、リソース制御など実マシンに依存しない実行環境の構築が実用的になりつつある。

広域分散環境でのデータアクセス (Remote I/O) では、ノード間のネットワーク性能がボトルネックとなりデータアクセスのための性能低下を引き起こす (図 1 の VM1)。一方、VM ではプロセスの状態を保持もしくは動作させたまま計算基盤を移動させることができるため、VM をアクセス対象のデータの近くへ移動 (Migration-then-Local I/O) させることによりアクセスの性能低下の削減が期待される (図 1 の VM2)。しかしながら、1）Remote I/O によるコスト 2）Migration-then-Local I/O によるコストはネットワーク性能やアクセスデータサイズなど、様々な要因によってその大小が変化するため、どのような状況下でどちらのアクセス手法を選択すべきかは明らかではない。そのため、これらのコストモデルを構築し比較することによって正しい選択をする必要がある。

本研究では、コストモデル構築の予備実験として様々な環境下でこれらの手法を適用した場合の性能を比較評価することによって、Migration-then-Local I/O 手法の有効性を示すとともに、両手法にとって最適な環境を定量的に示す。

2. 広域分散環境でのデータアクセス手法

本研究で対象とするデータアクセス手法について詳しく見ていく。まず、Remote I/O では地理的に遠いデータアクセスを行う手法で、データサイズが小規模である場合やネットワーク性能が高い環境で有効であると期待される。

† 東京工業大学
†† 国立情報学研究所

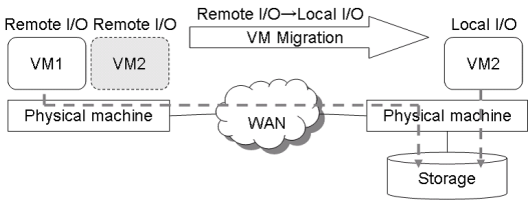

図 1　Remote I/O (VM1)　VS　Migration-then-Local I/O (VM2)

一方、Migration-then-Local I/O では Stop Migratoin と Live Migration の 2 つの手法[1] があり。前者は VM を一度サスペンドし、メモリ内容を移動先へ転送、完了後リスタートさせる方法をとり、後者は VM をサスペンドせずメモリの内容を転送し、その間に生じたダーティ・ページをある条件を満たすまで繰り返し転送した後、サスペンドさせ残りのダーティ・ページを転送し、完了後リスタートさせる方法をとる。前者の手法は、長いダウンタイムが欠点であるが VM メモリサイズの転送でマイグレーションが完了する。一方後者の手法は、ダウンタイムは短いがマイグレーション中のダーティ・ページを繰り返し転送する必要があるため、ダーティ・レートの高いアプリケーションを実行した場合、マイグレーション時間の増加を引き起こす。また、この Live Migration を本研究が対象とする問題へ適用した場合、動作中のアプリケーションがリモート・サイトのデータをメモリに読み込む一方で、そのリモート・サイトへマイグレーションを行うため、リモートアクセスとマイグレーション転送の発生によるネットワークへの負荷が問題になる。このため、アプリケーションのリモートデータ・アクセスやメモリ・アクセスの特性を考慮し、実行環境に最適なマイグレーション戦略を選択する必要がある。

3. 評　価

前章で述べた理由により本研究では、Remote I/O、Stop Migration、Live Migration の 3 つの手法の比較評価を行う。実験環境として、松岡研究室の PrestoIII クラスタの 2 ノードをネットワーク・エミュレータ GtrcNET-1[2] を介して接続 (1Gbps) し、一方のロー

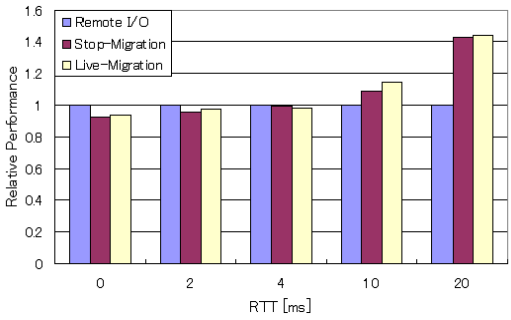

図 2　ネットワーク遅延による影響 (VM メモリサイズ:1024MB)

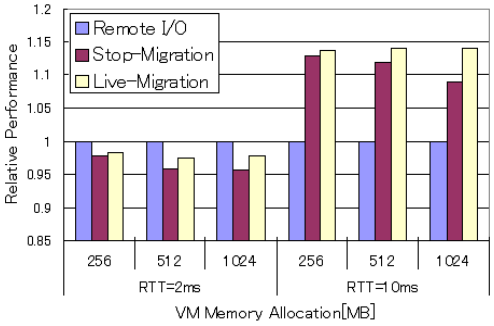

図 3　VM メモリサイズによる影響

カル・ディスクに VM イメージを、他方のノードにアクセス・データを用意し、双方のノードがそれらを NFS マウントすることにより、VM はマイグレーション後も同様の方法で目的データへアクセスできる。VM モニタは Xen3.1.0[3] を使用した。VM 上で動作させるアプリケーションとして、相同性検索バイオアプリケーションである BLAST[4] を使用し、ネットワーク遅延 (0-20ms)、データサイズ (2,7GB)、VM メモリサイズ (256,512,1024MB) を変化させ比較評価を行う。マイグレーションはアプリケーション実行後すぐに適用される場合を想定する。

Remote I/O を適応したアプリケーション性能を 1 とした Stop Migration、Live Migration の割合を図 2-4 に示す。図 2 より、RTT が増加するにつれてマイグレーションを用いる手法の有効性がより顕著に現れ、10ms 以上の環境で有効であることがわかる。また、図 3 から、VM メモリサイズにあまり影響を受けていないことがわかる。これはメモリサイズの増加による VM の性能向上と全実行時間に占めるマイグレーション時間が短いことに起因する。

しかし、今回調査した全ての環境において Stop Migration を適用したときの性能が Live Migration の性能より劣っていた。その原因は、図 5 の実行時間の内訳を見ればわかるとおり、両者のメモリ転送時間つまり SM では Downtime、Live Migration では

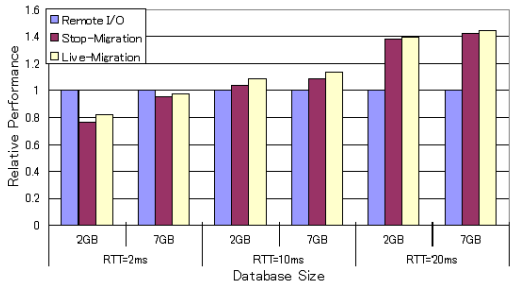

図 4　アクセスデータサイズによる影響 (VM メモリサイズ:1024MB)

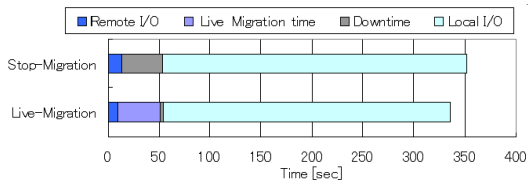

図 5　Live/Stop-Migration を適用した場合の BLAST 実行時間の内訳

Live Migration time+Downtime の時間がほぼ等しくなってしまったために、マイグレーション中もプロセスを停止させない Live Migration のほうが有利な結果となってしまった。しかし、今回使用したものよりダーティ・レートの高いアプリケーション使用した場合、Live Migration time の時間が長くなり、Local I/O 状態に遷移するまでの時間が長くなるため、Stop Migration の方が有利となることが期待される。

4.　まとめと今後の課題

本研究によって、各マイグレーション手法を比較し、Remote I/O、Live Migration が有効とされる状況を定量的に示した。今後の課題としては、マイグレーション戦略の精緻なコストモデルの構築や仮想クラスタ環境への適応などが挙げられる。

参　考　文　献

1) Christopher Clark, Keir Fraser, Steven Hand, Jacob Gorm Hanseny, Eric July, Christian Limpach, Ian Pratt, and Andrew Wareld. Live migration of virtual machines. In *NSDI*, 2005.
2) Y. Kodama, T. Kudoh, R. Takano, H. Sato, O. Tatebe, and S. Sekiguchi. Gnet-1:gigabit ethernet network testbed.
3) Xen community. http://xen.xensource.com/.
4) Ncbi blast. http://www.ncbi.nlm.nih.gov/BLAST.