情報処理学会研究報告
IPSJ SIG Technical Report

Vol.2010-DPS-145 No.19
Vol.2010-GN-77 No.19
Vol.2010-EIP-50 No.19
2010/11/26

教育用PCシステムを利用した分散処理実行時におけるPC利用者への影響調査

山邉大樹[†1]　奥村勝[†2]

本論文では既存の教育用 PC システムのバックグラウンドで Hadoop を使用した分散処理を実行した際に，教育用 PC システムの利用者へ及ぼすストレス等の影響調査を行う．この時，ホスト PC 上で直接分散処理を行う手法と仮想マシン上で分散処理を行う手法の 2 パターンを想定し，PC 利用者への影響の違いを評価する．

Consideration factor affecting to PC user with distributed processing using educational PC System

Taiki YAMABE[†1] and Masaru OKUMURA[†2]

In this paper, we consider the influence of Hadoop cluster on the user, when we execute hadoop cluster operation in the background of educational PC system. To achieve the coexistence of educational use and hadoop cluster, we are evaluating two methods. One is use the virtual machine technology, and another is not to use it. We examine the difference of influence on the user with two methods.

1. はじめに

Google に代表されるように日々発生する大規模なデータを高速に処理することで提供される情報サービスが実用化し, 日常の生活に欠かせないものとなっている．これらのサービスではプログラミングモデルを制約し, 単純化することで従来の並列, 分散処理に特有の問題を除去し, 処理を行うための計算機数をスケールアウトさせることで性能の向上を図る仕組みを用いて実現されている．一方, 大学のような教育機関では, PC の導入台数が数百台に及ぶところもあり，そのような場合メンテナンス性の面から PC の構成を容易に変更できるネットブート方式を採用した教育用 PC システムを保有していることも多い．

我々は, 教育機関の保有する多数の教育用 PC を大規模分散データ処理環境に転用し, 動作させることを検討としている．これまでに行った検証の結果，オープンソースソフトウェアの分散処理フレームワーク Hadoop とネットブート型シンクライアントからなる教育用 PC システムを用いて，1000 台からなる Hadoop クラスタを構成し，実際に並列分散処が行えることを確認した．しかしながら, Hadoop クラスタ構成時は，本来の教育用 PC システムとして利用できないという制約が生じた．このため，教育用 PC システムにおいて一般の講義利用と大規模分散データ処理環境を共存，同時提供することを目指している．それらの共存，同時提供の実現方法として教育用 PC のバックグラウンドで仮想化技術を用いて Hadoop クラスタを構成する手法と，教育用 PC 上で直接 Hadoop のソフトウェアを実行させ Hadoop クラスタを構成する 2 つの実現方法を想定している．

本稿ではこの 2 つの手法を念頭において PC 利用者が PC を操作する際のストレスなどの操作感にかかる影響調査を行った．影響調査として PC に負荷を与えるベンチマークの実行時間や分散処理実行時間の変化から PC 利用者にかかる影響について考察した．また，バックグラウンドで分散処理を実行し PC に負荷をかけた状態で利用者に実習に近い PC 操作をしてもらい操作感に関する聞き取りを行った．

以下，第 2 章で教育用 PC システムを用いた Hadoop クラスタの構築について，第 3 章で教育用 PC システムと Hadoop クラスタの共存させるための方法と現状までの調査結果について述べる．第 4 章で PC 利用者へ及ぼす影響調査の概要および結果について述べる．最後に第 5 章でまとめと今後の課題について述べる．

2. 教育用 PC システムを使用した Hadoop クラスタ構築

大学のような教育機関においては, 情報処理教育等の実施のための PC 環境として教育用 PC システムが導入されているが教育機関という組織の運用上, 夜間や夏季, 冬季などの長期休暇期間は, それらの PC は遊休状態となっており, 有効活用されていないことが多い．また, PC の導入台数が数百台に及ぶ機関ではメンテナンス性の面から PC の構成を容易に変更できるネットブート方式を採用していることも多い．つまり, 通常の講義利用以外への PC システムの転用が容易に行える機能を備えていると考えることができる．そこで, 大規

†1 福岡大学大学院
Graduate School of Engineering, Fukuoka University
†2 福岡大学総合情報処理センター
Information Technology Center, Fukuoka University

Vol.2010-DPS-145 No.19
Vol.2010-GN-77 No.19
Vol.2010-EIP-50 No.19
2010/11/26

模データ処理を実現するための仕組みである大規模分散処理フレームワークを既存の教育用 PC システムのリソースを活用して実現し, 大学が保有する PC リソースの有効化を図り, 今後ニーズが増加すると予想される大規模データ処理環境を利用者に提供する仕組みを実現することを目指す．以下，2.1 節で大規模分散処理フレームワーク Hadoop および，2.2 節でネットブート型シンクライアントの概要について述べる．2.3 節にて，ネットブート型シンクライアントを利用した Hadoop クラスタの動作検証について述べる．

2.1 Hadoop

Hadoop とは Google の基盤技術を基に作られた分散処理環境を実現するオープンソースソフトウェアである[1)]．Hadoop を使用することにより利用者は分散処理に関する難解な知識を必要とせずに分散処理を行うことができる．Hadoop は大きく分けて, 分散されたデータを管理する分散ファイルシステムである Hadoop Distributed File System(以下，HDFS) と, HDFS 上で分散処理を行う MapReduce の 2 つのコンポーネントから構成されている．

HDFS は NameNode と DataNode の 2 つのサーバから構成されており, NameNode はマスタとしてメタデータの管理を行い, DataNode はスレーブとして実際にデータを保持する役割を担う．一方, MapReduce は HDFS と同様に JobTracker と TaskTracker の 2 つのサーバから構成されている．JobTracker は Job の管理を行い, TaskTracker は実際に計算やデータ処理を行う．

2.2 ネットブート型シンクライアント

ネットブート型シンクライアントは, ネットワークを介して別のサーバに保持される仮想起動イメージからネットワークを介して起動イメージを転送し，クライアント側で OS を起動するシンクライアントの一方式である．一般にクライアント台数が多い場合，各 PC のローカル HDD の OS イメージをメンテナンスするコストが増加するが，ネットブート方式を用いるとサーバ側で保持する仮想起動イメージのみを更新することで，すべてのクライアントの起動イメージを更新することが可能となる．そのため管理コストの低減や，柔軟な PC 環境の提供を目的として大学等の教育機関で用いられている.[2)]

2.3 ネットブート型シンクライアントを利用した Hadoop クラスタの動作検証

Hadoop クラスタの構成上の特徴としてスレーブノード (DataNode および TaskTracker) が，基本的に同一ソフトウェア構成の多数のマシン群から構成されることがあげられる．一方でネットブート型シンクライアントでは一つの仮想起動イメージを準備することで多数の PC クライアントのソフトウェア構成などの動作環境を容易に統一できる．これら 2 つの特徴を組み合わせることによって, 各 PC に個別に必要なソフトウェアをインストール等することなく大規模分散処理環境を構築することができることに着目し，学内のネットブート型シンクライアントを利用した教育用 PC システムと Hadoop を組み合わせることによって, 大規模分散処理環境を構築し, 1000 台でのクラスタが動作することを確認した[3),4)].

しかし, この検証では通常の教育用 PC の仮想起動イメージを Hadoop クラスタ専用の仮想起動用に切り替えて行ったため, Hadoop クラスタを構成，利用している間, 本来の教育用 PC として使用できないという結果となった．

そのため, 仮にこの手法により Hadoop クラスタを構成したとしても，夜間や休日を利用した期間に利用が限定される，または一部の教室の PC のみを利用して構成規模を縮小した Hadoop クラスタを構成するなど，教育用 PC システムのもつ台数効果を活かせないこととなる．そこで問題解決のために，教育用 PC の利用と Hadoop を使用した分散処理を共存させる手法の検討を行うこととした．

3. 教育用 PC 環境と Hadoop クラスタ環境の共存に向けて

3.1 教育用 PC と分散処理の共存

教育用 PC の利用と Hadoop を使用した分散処理の共存のために充たすべき要件として 2 つを同時実行した際に，Hadoop を使用した分散処理によって教育用 PC の利用者に与える影響を少なくすることがあげられる．また，教育用 PC の利用者に影響を及ぼさない範囲で分散処理の性能を向上させる必要がある．

共存を実現する環境として本学のネットブート型シンクライアントである教育用 PC システムを想定している．本学の教育用 PC システムの PC に採用されている OS が Windows7 であるため，Windows 環境上に Hadoop クラスタを構築することとする．また，Hadoop の動作環境は Java が動作する Linux 環境を前提としているが，Windows 上でも動作は可能である．

上記の制約を考慮した上で共存の実現方法として次の 2 手法を検討した．まず，仮想化技術を用い Hadoop スレーブのノード機能を盛り込んだ仮想マシン（以下 VM）を教育用の Windows7 の PC（ホスト PC）のバックグラウンドで動作させる手法である．もう一つの手法は通常の講義利用のためのホスト PC 上で直接 Hadoop クラスタに必要なソフトウェアを動作させる手法である．Hadoop 関連の Java プログラムはデーモン構成を取っており，PC 利用者にそのソフトウェアの存在が直接目に触れることはない．以下，2 つの手法の得失について具体的に述べる．

ホスト PC 上で VM を使用する場合　メリットとしては Hadoop の分散処理中によりソ

Vol.2010-DPS-145 No.19
Vol.2010-GN-77 No.19
Vol.2010-EIP-50 No.19
2010/11/26

フトウェア障害が起きてもその影響は仮想マシン内にのみ限定されるため，ホスト PC を講義のために利用している利用者にはその影響が及ばず, 利用者のデータ等は保障されると考えられる．デメリットとしては，VM にホスト PC の CPU やメモリ, ディスク容量を仮想マシンに占有され, 教育用 PC が本来発揮できる性能よりも処理効率が落ちることやレスポンスが低下することが考えられる．同様に VM で構成される Hadoop クラスタも VM に割り当てられたリソースのみしか活用できず，ベースとなるホスト PC 全体の CPU やメモリ等のリソースを活用できず分散処理実行時の性能が犠牲になる．

ホスト PC 上で直接実行する場合　メリットとしては, VM の設定等を行う必要がないこと，Hadoop スレーブがホスト PC の CPU やメモリ等のリソースを十分に活用できるため分散処理時の性能が低下しにくいことが考えられる．デメリットとして Hadoop による分散処理実行時にソフトウェア障害等が起きるとホスト PC を利用している利用者が使用する領域にも影響を及ぼし, 利用者が保持するデータ等が障害と同時に一緒に失われてしまう可能性がある．

3.2 VM を利用した Hadoop クラスタ実行時のホスト PC への影響に関する評価

提案する 2 手法における Hadoop 実行時のホスト PC への影響を調査するために実機を用いた影響調査を行った[5]．その結果，Hadoop の処理性能については VM を使用して分散処理を行った場合，VM を使用しない場合よりも分散処理の性能が低下することを確認している．

ホスト PC への影響としては VM を使用しない場合，CPU の全てのコアを分散処理に使用できるため，利用者が PC を使用している間，分散処理が瞬間的に高い割合で CPU 利用率を占めることを確認した．また，使用可能な空きメモリ容量についても VM を使用した場合と異なり安定しない結果となった．

一方，VM を使用した場合，ディスク IO がボトルネックとなるため分散処理の性能が本来よりも低下するといった問題点もある．VM に割り当てたメモリは，VM で起動する OS やアプリケーション等で使用されるため，分散処理に使用できる実メモリ量はさらに少なくなる．使用できる空きメモリ量が減少するとスワップ処理等が発生し，HDD への読み書きが生じるため，分散処理性能が低下した．

4. 分散処理実行時における PC 利用者への影響調査

4.1 調 査 目 的

今回の影響調査の目的は, 3.1 節で述べた 2 つの提案手法において Hadoop を使用した分散処理によって講義利用などの教育用 PC の利用者の操作感に与えるストレスの影響を調査することである．

具体的には, Hadoop クラスタに対し分散処理とベンチマークの 2 つを同時実行させ PC に負荷を与えた状況で，実際に PC 利用者に講義等における実習に近い操作を行ってもらい操作感に対する意見を調査した．

4.2 検 証 環 境

検証環境として 5 台の PC を用いて, PC クラスタを構成した．PC クラスタ上に Hadoop クラスタをマスタ 1 台，スレーブ 4 台で構成する．VM を用いる場合は 4 台の PC 上にそれぞれ Hadoop を実装した VM を起動し Hadoop クラスタを構成する．検証環境を図 1 に, 検証に使用したマシンスペックを表 1 に, 使用したソフトウェアのバージョンを表 2 に示す．

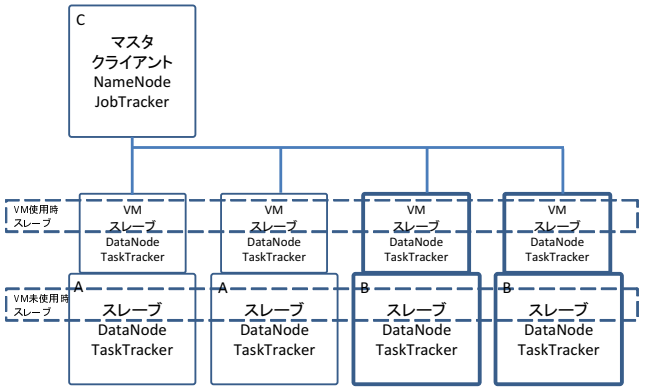

図 1　検証環境

4.3 分散処理と教育用 PC 利用の同時実行における相互への影響

Hadoop を使用した分散処理と教育用 PC 利用の同時実行における相互への影響を調査するために Hadoop を使用した分散処理と PC 利用者側のアプリケーションを同時実行した際のそれぞれの実行時間の推移を計測した．

Vol.2010-DPS-145 No.19
Vol.2010-GN-77 No.19
Vol.2010-EIP-50 No.19
2010/11/26

表 1 使用したマシンスペック

	CPU	メモリ	OS
A	Intel(R)Pentium(R)4 3.00GHz	3GB	Windows7
B	Intel(R)Core(TM)2Quad Q9550 2.83GHz	3GB	Windows7
C	Intel(R)Pentium(R)4 3.40GHz	2GB	Windows7
VM	ホスト OS 上の CPU(A 又は B)	1GB	Windows7

表 2 使用したソフトウェアのバージョン

Java(jdk)	1.6.0_20
Hadoop	0.20.2
VMware Player	3.0.1

PC 利用者側のアプリケーション (ユーザアプリケーション) としてベンチマークプログラムである Superπ を利用した[6]．Superπ は円周率を求める CPU バウンドのアプリケーションでマルチコア環境で実行すると 1 プロセス 1 コアを占有する．今回の調査で用いる 4 コアの PC では 1 プロセスが CPU 利用率 25 %を占有し，同時に 4 プロセス実行すると CPU 利用率 100%を占有する．Superπ では 3357 万桁の π の計算を実行した．一方，Hadoop での分散処理として 4 台のスレーブから成る Hadoop クラスタでランダムな数値データの並べ替え処理である Sort プログラムを 6GB のデータに対して行った．いずれの処理も単独で実行した場合の実行時間は，約 1000 秒であった．

図 2，3 は VM を使用した手法と VM を使用しない手法の 2 つの状況において，Hadoop を使用した分散処理とユーザアプリケーションを各 PC の Windows 上でそれぞれ 0〜4 本同時実行させた際の実行時間の変化を表している．Hadoop の実行時間についてはクラスタ全体での処理時間を，ユーザアプリケーションについては想定している教育用 PC の性能に近い表 1 のスペック B の PC での実行時間 (平均) である．

分散処理とユーザアプリケーションの同時実行の実行時間と，ユーザアプリケーションのみで実行させた時の実行時間を比較すると，同時実行の実行時間はベンチマークのみの実行時間の VM を使用しない場合では 1.12 倍，VM を使用した場合では 1.04 倍となっている．ユーザアプリケーションの同時実行数が増加してもユーザアプリケーションの実行時間はさほど変化していない．これは，分散処理と同時に実行されているユーザアプリケーションが CPU を占有しているからであると考えることができる．このことから，分散処理とユーザアプリケーションを同時実行した場合，分散処理の性能が低くなると考えることができ，図 2，3 からも分散処理の性能が低下していることが確認できる．

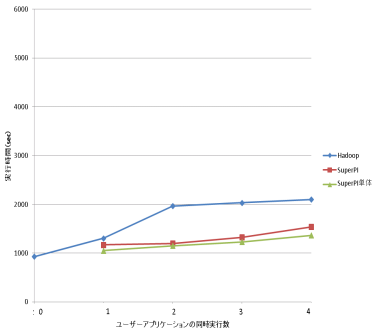

図 2 VM を使用しない場合の Hadoop とユーザアプリケーションの同時実行

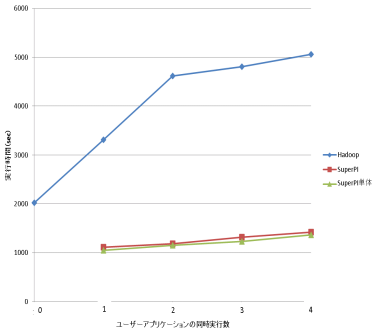

図 3 VM を使用する場合の Hadoop とユーザアプリケーションの同時実行

4.4 分散処理実行時の PC 利用者の操作感についての調査

次に Hadoop を使用した分散処理と教育用 PC 利用の同時実行時における PC 利用者の操作感に与えるストレスの影響を調査するために，図 1 の検証環境において調査を行った．調査には教育用 PC のスペックに近い表 1 の B のスペックを持つマシンを利用した．調査内容は，4.3 節と同様の VM を使用する手法と使用しない手法のそれぞれにおいてバックグラウンドで Hadoop による分散処理とユーザアプリケーションを同時実行させる．この時，

Vol.2010-DPS-145 No.19
Vol.2010-GN-77 No.19
Vol.2010-EIP-50 No.19
2010/11/26

ユーザアプリケーションの同時実行数を 0〜4 と変更させた．なお，比較用として負荷なしの条件でも行った．それぞれの条件での実行時に PC 利用者に実際に PC でアプリケーション等を操作をしてもらい操作感に対するストレスを聞き取った．PC 利用者が行った操作の内容と操作感の内容は以下の通りである．

PC 利用者が行った操作は次の 4 項目である．IE8 による Web ブラウジング，Word を使用した文書作成，動画共有サイト Youtube を利用した動画の視聴，USB メモリから PC へ 400MB のデータファイルのコピーである．また，PC 利用者が行う評価として操作感に及ぼすストレスを 4 段階評価の数値で記録してもらった．4 段階の数値の内容として分散処理もユーザアプリケーションによる負荷もかけていない状態を基準とし，通常の PC 利用と比べて問題なく使用できる場合は 4，違和感はあるがストレスを感じない場合は 3，強いストレスを感じた場合は 2，全く PC を使用できないと感じた場合は 1 である．検証に協力してもらった PC 利用者の人数は 3 人である．その時の記録してもらった数値を平均した結果を表 3 に示す．

表 3 利用者の操作感に関する評価

Hadoop		なし	あり	あり	あり	あり	あり
Superπ		なし	なし	同時 1	同時 2	同時 3	同時 4
VM なし	Web ブラウジング	4	3.3	4	3.3	3.6	3
	Word	4	3	3	4	2	2
	Youtube	4	4	4	4	2.6	2.6
	FileCopy	4	4	4	4	4	3.3
VM あり	Web ブラウジング	4	4	3	3.6	4	3
	Word	4	2.6	2	3.3	4	3
	Youtube	4	4	3.3	4	4	3.3
	FileCopy	4	4	3.6	4	3.6	3

表 3 から分散処理とユーザアプリケーションの同時実行数が 3，4 のとき PC 利用者は VM を使用しない場合にストレスを感じているが，VM を使用した場合は VM なしの時ほど PC 利用者はストレスを感じていない．また，分散処理とユーザアプリケーションの同時実行数が 1，2 のとき PC 利用者は VM を使用しない場合にストレスを感じていないが，VM を使用した場合はストレスを感じている．

これは，ユーザアプリケーションの同時実行数が少ない時，つまり，Hadoop を実行する VM が CPU を多く占有できる場合に PC 利用者がストレスを感じていると考えることができる．

5. まとめと今後の課題

調査結果から分散処理実行時に PC 利用者に与えるストレスの影響は VM を使用しないホスト PC 上で直接分散処理を行う方が少ないと考える．

検証結果からは VM を使用しない場合 PC 全体に負荷を掛けると PC 利用者の操作感に与えるストレスの影響が大きくなるが，PC 利用者が一般的な講義利用では PC 全体に過大な負荷を掛けることは少ない．そのため表 3 のユーザアプリケーションの同時実行数が 1，2 の場合を考慮すればよい．VM を使用しない場合だと，PC 利用者の操作感に及ぼすストレスは少ないが，VM を使用した場合は同様の状況ではストレスが多い．このため VM を使用しないホスト PC 上で直接 Hadoop による分散処理を行う方が PC 利用者の操作感に及ぼすストレスという観点からは望ましいと考える．また，性能面でも VM を使用しない手法は VM を使用した手法の 2 倍近く分散処理の性能が得られた．

今回教育用 PC システムを利用した分散処理実行時における PC 利用者への影響調査を行うために VM を使用する手法と VM を使用しない手法の 2 つの手法を用いて，PC に負荷のかかるユーザアプリケーションの実行時間や分散処理実行時間の変化から PC 利用者に与えるストレスの調査を行った．そして調査結果から VM を使用しない手法の方が PC 利用者に与えるストレスの影響が少ないことが分かった．

今後の課題として，VM を使用しない手法において検討すべき項目について調査を行う．最終的にはネットブート型シンクライアントである教育用 PC システムを想定した分散処理環境において分散処理性能の変化や PC 利用者への影響調査を検討している．

参 考 文 献

1) Hadoop http://hadoop.apache.org/
2) 濱田 正博: シンクライアントのすべてがわかる, 日経 BP 出版センター, 2006.
3) 山邉 大樹, 奥村 勝: 教育用 PC システムを利用した Hadoop クラスタの動作検証, 電気関係学会九州支部第 62 回連合大会, 2009.
4) 奥村 勝: 教育用 PC を用いた大規模計算フレームワークの実現に向けて, 情報処理学会研究報告, 2009-IOT-7, 2009.
5) 山邉 大樹, 奥村 勝: 仮想マシンを利用した Hadoop クラスタ実行時のホスト OS への影響に関する評価，DICOMO2010, 2010.
6) スーパー π http://www1.coralnet.or.jp/kusuto/PI/super_pi.html