

シーク距離を考慮した VM イメージファイル配置変更による  
仮想化環境における Hadoop I/O 性能の向上  
Improving Hadoop I/O Performance  
Based on Seek Distance Aware VM Image File Location Optimization

中島 健司<sup>†</sup> 藤島 永太<sup>†</sup> 山口 実靖<sup>†</sup>  
Kenji Nakashima Eita Fujishima Saneyasu Yamaguchi

## 1. はじめに

単一物理計算機上に複数の仮想計算機を起動する様なクラウド環境上で動作する Hadoop においては、複数の仮想計算機が同時に単一の HDD にアクセスするような処理において I/O 処理がボトルネックとなってしまう問題がある。物理計算機環境や仮想計算機環境における Hadoop の I/O 性能を向上させる研究としてファイル格納位置制御手法[1][2]がある。本稿では、本格納位置制御手法の仮想化環境への適応に関する課題を述べ、その解決手法を提案する。そして、これら手法をホスト OS のみに、あるいはホスト OS とゲスト OS の両方に適用し、その性能を評価する。そして、シーク距離を考慮したストライピング手法の有効性について考察する。

## 2. ファイル格納位置の静的制御手法

定記録密度方式 HDD のシーケンシャル I/O 速度は内周側のゾーンより外周側のゾーンの方が高い。文献[1]では、この特性を考慮しファイルを外周側のゾーンに優先的に格納し、シーケンシャル I/O 性能を向上させている。この手法の実装は、オープンソースファイルシステムである ext2/ext3 を用いて行っている。これらのファイルシステムは、ディスクは 4KB のブロックを単位に管理され、複数のブロックでブロックグループを構成している。そして、ブロックグループ毎にブロックビットマップ、inode ビットマップ、inode テーブルが用意されている。ブロックが使用中か未使用であるかはビットマップで管理されている。各ブロックグループのデータブロックビットマップのうち、ディスクの外周側に相当する低アドレス部以外のブロックのビットを使用中ビットへ変更し、内周側に相当する高アドレス部にデータが格納されることを回避している。

## 3. 仮想化環境における配置位置制御手法

本章にて、仮想化環境におけるファイル格納位置制御手法[1][2]の課題を述べる。

ファイルシステム上に生成されたイメージファイルは HDD の連続領域に配置される。よって、最初に生成されたイメージファイルは HDD の外周側領域(すなわち高速領域)に生成されるが、2 個目以降のイメージファイルは既にイメージファイルが配置された領域以外(すなわち低速領域)に生成される。イメージファイルが配置された状況では、仮想 HDD およびゲスト OS ファイルシステムにおける格納位置が物理 HDD における格納位置と一致しないことになる。よって、この環境下でファイル格納位置制御

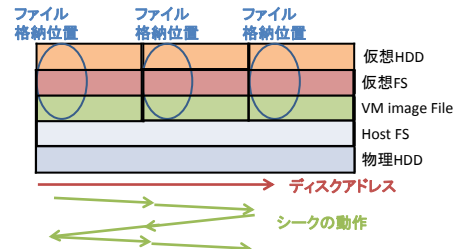
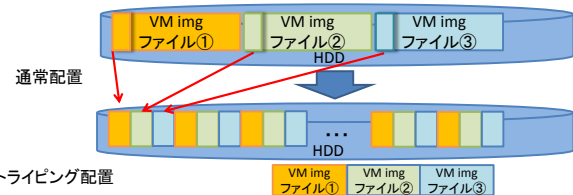


図1 仮想化環境における静的ファイル格納位置制御手法



ストライピング配置

図2 イメージファイルの通常配置とストライピング配置

手法を用いると、図1のように各仮想HDDの低アドレス部に I/O が発生することになるが、1 個目の VM 以外は物理 HDD の低速領域を使うことになる。また、ファイル格納位置が物理アドレスにて離れることになり、シーク距離が長くなってしまいう問題が生じる。

## 4. VM イメージファイルのストライピング手法

本章にて、仮想化環境における Hadoop の I/O 性能の向上を実現する手法を提案する。本手法では、各 VM のイメージファイルを等間隔で分割し、図2のようにイメージファイルの断片を順番にストライプ状に配置する。そして、この状態にてゲスト OS ファイルシステムにファイル格納位置制御手法を適用する。各イメージファイルの断片をストライプ状に配置することにより、各 VM の仮想 HDD の低アドレス部が物理 HDD の低アドレス部に対応し、物理 HDD の高速な外周側を優先的に使用させることが可能となる。また、ファイルが格納されている物理アドレスが近づく物理シーク距離も削減することができ、I/O 性能を向上させることができる。

本手法の実装は、2 章で記述した既存手法と同様に ext3 ファイルシステムのデータブロックビットマップを書き換えることにより行うことが可能である。具体的には、各々のイメージファイルを生成する領域以外はビットを使用中に書き換えた状態でイメージファイルの生成を行う。これにより、イメージファイルの断片を図2のように順番に配置させることが可能となる。

想定環境は、仮想ストレージアドレスと物理ストレージアドレスが一致しない環境であり、ホスト OS の管理権限を有している状況である。プライベートクラウドなどで単一のストレージデバイスに複数の仮想計算機の仮想ストレージが配置される状況などがこれに該当すると考えられる。

<sup>†</sup>工学院大学大学院 工学研究科 電気・電子工学専攻  
Electrical Engineering and Electronics, Kogakuin University  
Graduate School

## 5. 性能評価

### 5.1 測定方法

Hadoop サンプルアプリケーションである TeraSort を実行し、通常手法、ファイル格納位置制御手法、ストライピング手法の性能評価を行った。

通常手法はホスト OS およびゲスト OS 上で制御を行わず、ファイル格納位置制御手法はホスト OS 上で制御を行わずゲスト OS 上でファイル格納位置制御を行い、ストライピング手法ではホスト OS 上でストライピング配置をゲスト OS 上でファイル格納位置制御を行っている。

測定環境は 1 台の物理計算機で構成され、この物理計算機上に 3 台の仮想計算機を起動した。この 3 台の仮想計算機を SlaveNode として、物理計算機を MasterNode として使用した。HDFS ブロックサイズは 64MB、ブロックの複製数は 3、各 VM のファイルシステムは ext3 とした。ストライピング手法の実装方法は前章の通りである。

TeraSort は入力データサイズ 24GB にて 5 回行った。ファイル格納位置制御手法およびストライピング手法ではファイル格納位置の静的制御手法を適用し、仮想 HDD における外周側 50 GB 以外へのファイルの配置を禁止した。

### 5.2 測定結果

TeraSort の実行時間を図 3 に示す。図より、ストライピング手法の平均実行時間は通常手法よりも約 23.5% 短く、ファイル格納位置の制御手法よりも約 20.0% 短いことが確認できる。また、3 章で述べたようにファイル格納位置制御手法を仮想化環境にナイーブに適用すると性能向上が小さいことも確認できる。

図 3 において、通常手法およびファイル格納位置制御手法の各実行時間には分散があるが、ストライピング手法では実行時間のばらつきが小さいことが分かる。

### 5.3 I/O 要求が行われたディスクアドレス

TeraSort を実行時に測定用 HDD に発行された I/O 要求を調査した。各手法において I/O 要求が発行されたディスクアドレスを図 4、図 5、図 6 に示す。

図 4 の通常手法において、1 つの VM 以外高アドレス領域(低速領域)を使用していることが分かる。また、ファイル格納位置が広域に分散してしまい、シーク距離が長くなってしまっていることが分かる。図 5 のファイル格納位置制御手法では、3 章で述べたように 1 つ目の VM 以外は物理 HDD の低速領域を使うことになっており、またファイル格納位置が離れることによりシーク距離が長くなってしまっていることが確認できる。これらに比べて図 6 のストライピング手法適用時では、高速な物理 HDD の外周側を積極的に使用できていることが分かる。また、ファイル格納位置が狭い範囲にまとまっていることによりシーク距離も短くなっていることが分かる。

## 6. おわりに

本稿では、ファイル格納位置制御手法の仮想化環境への適用に関する課題を示し、その解決手法を提案した。そして、これらをホスト OS のみに、あるいはホスト OS とゲスト OS の両方に適用し性能評価を行い、提案手法の有効性を

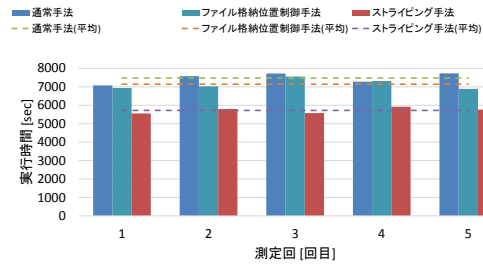


図 3 TeraSort の実行時間

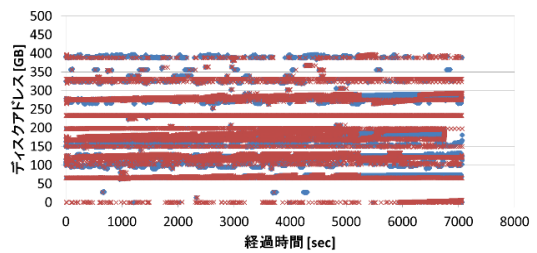


図 4 I/O 要求が行われたディスクアドレス(通常手法)

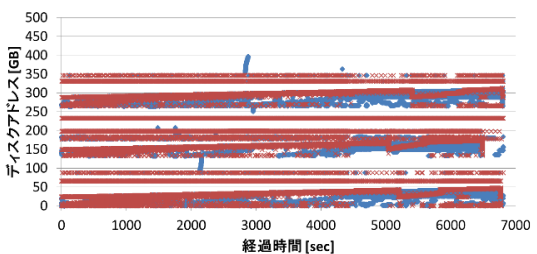


図 5 I/O 要求が行われたディスクアドレス (ファイル格納位置制御手法)

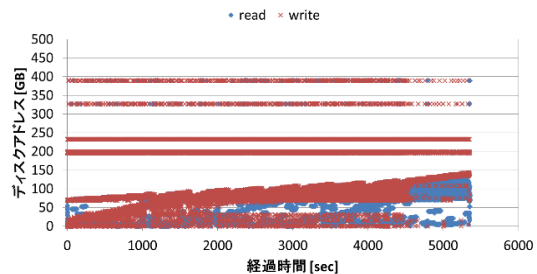


図 6 I/O 要求が行われたディスクアドレス (ストライピング手法)

を確認した。今後は、アクセス頻度を考慮した格納位置制御について考察していく予定である。

### 謝辞

本研究は JSPS 科研費 26730040, 15H02696 の助成を受けたものである。

本研究は、JST、CREST の支援を受けたものである。

### 参考文献

- [1] Eita Fujishima, Saneyasu Yamaguchi, "I/O Improving on Reduce Phase of Hadoop," International Symposium on Computing and Networking (CANDAR'15), (2015)
- [2] Eita Fujishima, Saneyasu Yamaguchi, "Dynamic File Placing Control for Improving the I/O Performance in the Reduce Phase of Hadoop", the Tenth International Conference on Ubiquitous Information Management and Communication (IMCOM2016), 8-2, (2016)
- [3] 中島 健司, 藤島 永太, 山口 実靖, "ホスト OS ファイルシステムにおける VM イメージファイルの非連続配置による仮想化環境における Hadoop I/O 性能の向上", システムソフトウェアとオペレーティング・システム研究会 第 137 回, 2016-OS-137, 4 (2016).