

# VMware ESXi 5 の I/O 性能評価

斎藤 大輔<sup>1</sup> 芳澤 伸介<sup>1</sup> 山崎 輝幸<sup>1</sup>  
高野 了成<sup>2</sup> 池上 努<sup>2</sup> 中田 秀基<sup>2</sup>  
竹房 あつ子<sup>2</sup> 田中 良夫<sup>2</sup>

## 1. 目的

近年、データセンター内のサーバからデスクトップまで、幅広い分野へ仮想化技術の活用が注目されている。仮想化には運用コストの削減、セキュリティの向上などの利点があるが、仮想化オーバーヘッドの問題から基幹業務系など性能が重視される用途への導入は進んでいない。この問題はハードウェア仮想化支援機構の進展により解決されつつあるが、その検証は十分に行われているとは言えない。特に I/O については PCI パススルーや SR-IOV など、仮想マシンモニタ (VMM) バイパス I/O 技術が有効である。我々はこれまでハイパフォーマンス計算 (HPC) 用途に対する VMM バイパス I/O 技術の効果を評価してきた[1]。本稿では、エンタープライズ分野で市場シェアの高い VMware ESXi を用いた仮想化環境の I/O 性能を評価することで、基幹業務への仮想化技術の利用拡大の実現性を検証することを目的とする。

## 2. VMware ESXi 5 の概要

VMware ESXi 5 (以下、ESXi) [2] は、ハードウェア上で直接実行される Type I (ハイパーバイザ) 型の仮想マシンモニタ (VMM) [3] である。Type I VMM のアーキテクチャは、ESXi のモノリシックカーネル型、Xen や Hyper-V のマイクロカーネル型に分類できる。なお、両者のハイブリッドである KVM のような実装も存在する。

I/O 処理に着目すると、管理 OS に処理を委譲するマイクロカーネル型と比較して、VMM 内部で処理を実行するモノリシックカーネル型は、性能面では有利だが、新規デバイスへの対応や保守、安全性などの面で不利となる。

続いて ESXi の I/O アーキテクチャについて概観する。ESXi は、完全仮想化 I/O (HVIO)、準仮想化 I/O (PVIO)、VMDirectPath I/O (DPIO) の 3 つの I/O 方式を提供する。HVIO および PVIO 方式では、VMkernel と呼ばれる VMM 内部のデバイスドライバによって I/O が処理さ

れる。DPIO 方式では、I/O 処理は VMkernel をバイパスし、ゲスト OS のデバイスドライバから直接デバイスが操作される。PVIO 方式では、ゲスト OS に ESXi 専用のデバイスドライバが必要であり、例えば PVSCSI や VMNET3 などが存在する。また、HVIO および PVIO 方式はデバイスを複数のゲスト OS から共有できるのに対して、DPIO 方式では複数のゲスト OS からの共有は不可能である。

ESXi はファイルシステムとして、VMware Virtual Machine File System (VMFS) と Raw Device Mapping (RDM) の 2 方式を提供する。VMFS では、ストレージの論理ユニット(LUN) 内に複数の VM ファイルを格納し、ゲスト OS に各 VM ファイルを割り当てる。これに対して、RDM ではゲスト OS に LUN を直接割り当てる。

## 3. 実験

### 3.1 実験環境

使用したハードウェア環境を表 1 に示す。VMM として VMware ESXi 5.0 Update1、ゲスト OS として Windows Server 2008 R2 SP1 Standard (以下、Windows)、および Red Hat Enterprise Linux Server 6.2 (以下、Linux) を用いた。仮想マシンには仮想 CPU として 1 または 8 コアを、メモリは 4GB を割り当てた。基本的な I/O 性能を把握するために、実験は次に示すように極力単純な構成で行った。ストレージ実験では、サーバと Fibre Channel (FC) ストレージを直接接続した。

表1 実験環境の諸元

Server: IBM System x3650M2	
CPU	Intel quad-core Xeon E5540/2.53GHz x2
Memory	16GB(4×4GB) DDR3-1033 LP RDIMM
Fibre Channel	QLogic ISP2532-based 8Gb FC
Ethernet	Broadcom BCM57710-based 10GbE
Storage: IBM DS3524	
Disk	15K RPM SAS 136 GB x24, RAID 5

1 株式会社アルゴグラフィックス

2 独立行政法人産業技術総合研究所情報技術研究部門

ネットワーク実験では、通信相手として、上記のサーバとほぼ同じ構成の IBM System x3550M3 に Windows を搭載し利用した。MTU は 9000 バイトに設定した。なお、実験結果の詳細な結果についてはポスタにて発表し、ここでは実験内容と結果の概略の記載に留める。

### 3.2 ストレージ性能

ベンチマークプログラムとして IOMeter 1.10-rc1 [4] を用いて、非仮想化環境 (BMM: Bare Metal Machine)、VMFS、RDM、DPIO の性能評価を比較した。比較対象のシステム構成を図 1 に示す。なお、図は Windows 環境の構成を示しているが、Linux 環境でも基本的に同様である。転送要求サイズは 512B、4KB、16KB、32KB、アクセスパターンはシーケンシャルとランダム、Read:Write の割合は 100:0 から 0:100 の範囲を 25% 刻み、Outstanding I/O 数は 1、4、16、64、256 の組合せで実験を行った。なお、VMkernel 内部でのキャッシュの影響を極力避けるために、総 I/O 量は 40GB と物理メモリサイズよりも十分大きく設定した。さらに 1 台の物理マシンに 3VM を起動したときの性能を評価した。この際、FC ストレージ上に 3 つの LUN を作成し、各 VM に 1 つずつ割り当てた。なお、DPIO の場合は、1 つの VM が DPIO 経由で LUN を割り当て、残りの 2 つの VM は VMFS 経由で LUN を割り当てた。

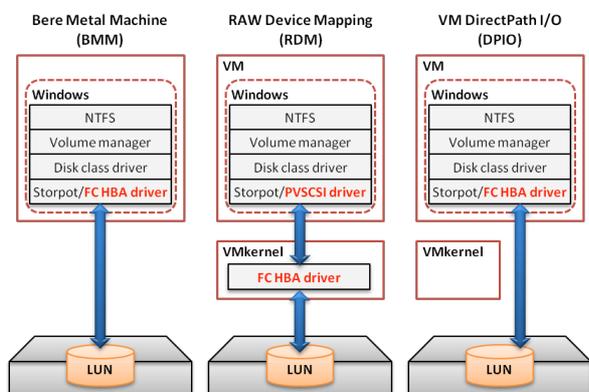


図 1 ストレージ性能比較

### 3.3 ネットワーク性能ストレージ性能

ベンチマークプログラムとして JPerf 2.0.2 [5] を用いて、BMM、VMNET3、DPIO の TCP/IO 性能を比較した。比較対象のシステム構成を図 2 に示す。メッセージサイズは 32KB、送信ウィンドウサイズは 2MB、受信ウィンドウサイズは自動チューニングに設定した。さらに 1 台の物理マシンに 3VM を起動したときの性能を評価した。ストレージ性能評価と同様に、DPIO の場合は、1VM が DPIO 経由で通信を行い、残りの 2VM は VMNET3 経由で行った。

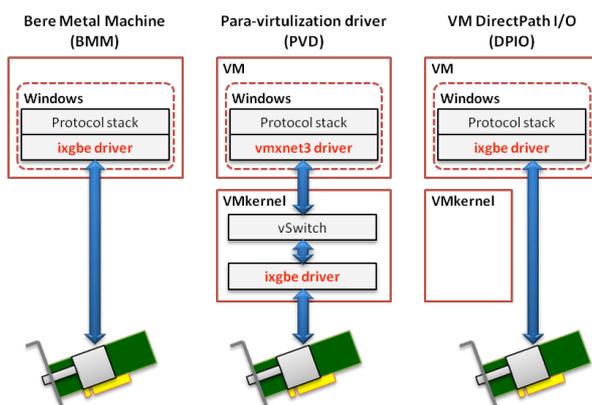


図 2 ネットワーク性能比較

## 4. 議論

ストレージ性能に関しては、Windows および Linux 共に、概ね BMM よりも ESXi の方が性能が高かった。これは我々の予想と反した意外な結果であった。原因として準仮想化デバイスの最適化による効果が考えられるが、詳細な解析は今後の課題である。また、Linux のランダム I/O 性能が低かった。ファイルシステムの影響もあるので、さらに raw ディスク性能を比較する必要があると考える。ネットワーク性能に関しては、VMNET3 使用時でも BMM に匹敵する性能が得られることが分かった。

論文[6]によると、ノード間通信 I/O が発生しない SMP 環境における HPC ベンチマーク評価では、ESXi は BMM に匹敵する性能を示している。本稿ではマイクロベンチマークによる基本 I/O 性能を評価したが、仮想化による性能低下はみられなかった。これらの結果より、仮想化環境は性能が要求される分野においても十分に実用的であるとの見込みを得られたが、実アプリケーションに近い負荷を用いた評価が今後の課題である。

## 参考文献

- [1] 高野了成, 池上努, 広瀬崇宏, 田中良夫: HPC クラウドの実現に向けた仮想化クラスターの性能評価, 情報処理学会論文誌コンピューティングシステム (ACS), Vol.5, No.2, pp. 111-120 (2012)..
- [2] VMware ESXi: <http://www.vmware.com/products/vsphere/esxi-and-esx/overview.html>.
- [3] R. P. Goldberg: Architectural Principles for Virtual Computer Systems, Ph.D. thesis, Harvard University, (1972)..
- [4] IOMeter: <http://www.iometer.org>.
- [5] JPerf: <http://code.google.com/p/xjperf/>.
- [6] Q. Ali, V. Kiriansky, J. Simons, and P. Zaro: Performance Evaluation of HPC Benchmarks on VMware's ESXiServer, 5th Workshop on System-Level Virtualization for High-Performance Computing (HPCVirt), pp. 213-222 (2011).