

ホワイトペーパー

日立サーバ論理分割機構「Virtage」と  
Intel®Virtualization Technology による  
Nested Virtualization (VMM on LPAR)の実現

2014年10月発行

株式会社 日立製作所

## 目次

1	はじめに.....	3
1.1	登録商標・商標.....	3
1.2	注意事項.....	3
1.3	免債事項.....	3
2	VMM on LPAR の狙い価値.....	4
2.1	VM と LPAR の融合.....	4
2.2	ユースケース.....	4
3	VMM on LPAR の課題と VMCS Shadowing の効果.....	6
3.1	サーバ仮想化を支援する Intel VT.....	6
3.2	VMM on LPAR の仕組みと課題.....	7
3.3	課題を解決する VMCS Shadowing.....	8
4	測定/評価結果.....	9
4.1	測定環境.....	9
4.2	評価結果.....	10
4.3	考察.....	10
5	実ユースケース.....	11
6	おわりに.....	12

1. 本書の内容は一部または全部を無断転載することは禁止されています。
2. 本書の内容に関しては将来予告なしに変更することがあります。
3. 日立製作所の許可なく複製・改変などを行うことはできません。
4. 日立製作所が製品やサービスについて行う保証は、製品添付の保証文章に記載した内容のみに限定され、本書のどの箇所であっても何ら新規の保証を行うものではありません。
5. 運用した結果の影響については、責任を負いかねますのでご了承ください。
6. 本書に技術的あるいは編集上の誤りや欠陥があったとしても、日立製作所は一切の責任を負わないものとします。

## 1 はじめに

Virtage は、日立製作所の統合サービスプラットフォーム「BladeSymphony」のブレードサーバ(BS2500、BS2000、BS500)に組み込まれたサーバ論理分割機構であり、LPAR(論理分割、Logical PARTitioning、以下、LPARとする)方式を採用している。LPAR方式は、主に基幹業務で用いられるメインフレームやUNIXサーバの仮想化方式として用いられていることでも分るように、基幹業務での利用に適した仮想化方式である。

本稿では、Virtage の LPAR 上で他社の Virtual Machine Monitor(以下、VMMとする)を稼動する Nested Virtualization (以下、VMM on LPAR とする)機能と、日立も開発に貢献した Intel Xeon Processor E5 V3 Family の Intel® Virtual Machine Control Structure(以下、VMCS とする) Shadowing による性能改善効果について示す。さらに、本技術のユーザサイトでの試行例を示す。

### 1.1 登録商標・商標

- Intel, Xeon はアメリカ合衆国および/またはその他の国における Intel Corporation の商標です。
- Microsoft, Windows Server は、米国 Microsoft Corporation の米国およびその他の国における登録商標または商標です。
- Linux は Linus Torvalds 氏の米国およびその他の国における登録商標あるいは商標です。
- Red Hat は米国およびその他の国における Red Hat, Inc の登録商標あるいは商標です。
- その他記載の会社名、製品名はそれぞれの会社の登録商標あるいは商標です。

### 1.2 注意事項

性能については、ベンチマーク値を使った机上計算のみでなく、実際の運用環境と同等構成(ハード・ソフト)において、要件を満たしていることを確認してください。ベンチマーク値の単純比較のみならず、本番相当環境(実機)を利用した事前評価等を強く推奨します。システム性能は、テスト方法やハードウェアに搭載するプロセッサの種類・動作周波数、メインメモリ・キャッシュメモリ容量、ディスク数・ディスク構成、動作ソフトウェア構成等の条件により大きく左右されます。サイジングなどにおいてはベンチマーク値を目安としてご利用いただけますが、机上計算のみでは思わぬ性能不足等に陥る可能性があります。

### 1.3 免責事項

本サイトに掲載している性能評価指標、または、その他の性能に関連する情報の使用によって、いかなる損害が生じた場合も、情報提供者である(情 IP)は責任を負いません。

2 VMM on LPAR の狙い/価値

2.1 VM と LPAR の融合

サーバ仮想化は、大別すると次の2つに分類される。

- ・VM: OSに対してハードウェア層を隠蔽し、Virtual Machine(以下、VM とする)を生成する。柔軟性、運用性に優れる。
- ・LPAR: ハードウェア資源の分割に特化し、OS にハードウェアコンポーネントの部分を見せる方式。極力直接実行させるため、堅牢性/信頼性に優れる。

サーバ仮想化技術が最初に確立したメインフレームにおいては、まず VM 方式の仮想計算機が商用化されたが、ハードウェア技術の進歩とともに、業務の高い堅牢性と資源の有効活用を可能とする LPAR 方式が商用化された。さらに LPAR 方式では、VM 方式の価値を取り入れるべく、LPAR 方式で論理分割された区画上で VMM を動作させる技術が採用された。こうして生まれた技術が VMM on LPAR である。LPAR は限りなく物理のサーバに近づき、VMM も OS の一つと扱えるようになったと理解頂いて良い。

x86 においても、VM 方式が提供する柔軟性/運用性と LPAR 方式が提供する堅牢性の両立が求められている。そこで、日立ではこのような要求に応えるため、LPAR 上で VMM が動作する機能を開発している。

図1に VMM on LPAR の概念を示す。VMM on LPAR は、Virtage によって分割された LPAR 上で OS と同様に他社の VMM を稼働させる技術である。当然、複数の LPAR 上で独立して複数の VMM を稼働させることが可能である。Virtage も仮想化レイヤとみなすと、仮想化のレイヤが重なっていることから、“Nested Virtualization”とも呼ばれることもある。Virtage 直上の OS、VMM をレベル1ゲスト、レベル1の VMM 上の OS をレベル2ゲストと定義する。

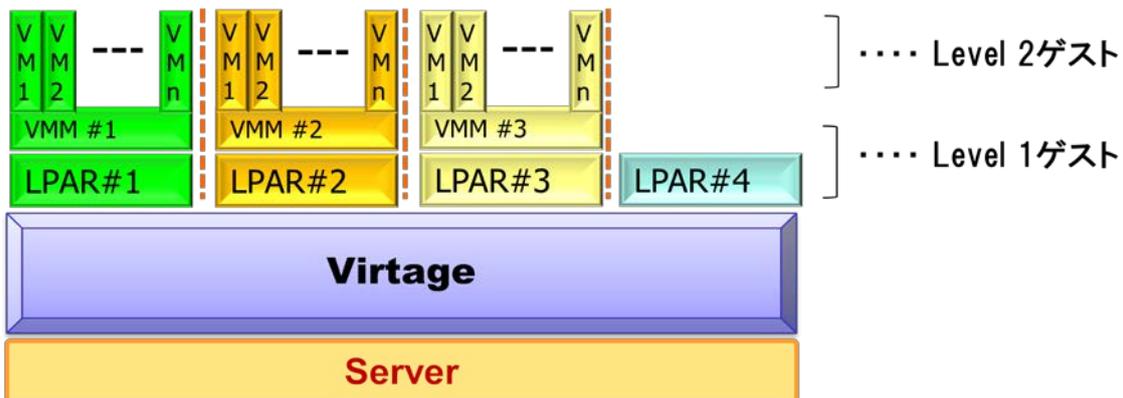


図1 VMM on LPAR の概念

2.2 ユースケース

以下に、VMM on LPAR が想定するユースケースを説明する。

(1) マルチテナント対応高信頼クラウド

クラウドシステムにおいて、複数のテナントが一つのプラットフォーム上に同居するようなケースは多い。このような他部門、他社のサーバと混在するような環境では、テナント間の隔離性(性能、運用、障害)の確保が重要な課題となる。ここで、1つのインスタンスの VMM に複数のテナントの VM を集約した場合、VMM の管理ソフトウェアで運用面は隔離することが可能となるが、例えば、他テナントの VM の負荷変動の影響の排除や、障害発生時の影響排除などについては、完全に実現はできない。そのためには、テナント毎に VMM を用意し、それぞれの別の物理サーバ上で稼働させれば問題は解決するが、コスト増につながる。これに対して、本 VMM on LPAR 技術により、テナント間の障害や負荷変動に対する隔離性を実現しつつ、複数テナント/VMM の集約によりコスト削減が可能である。(図2)

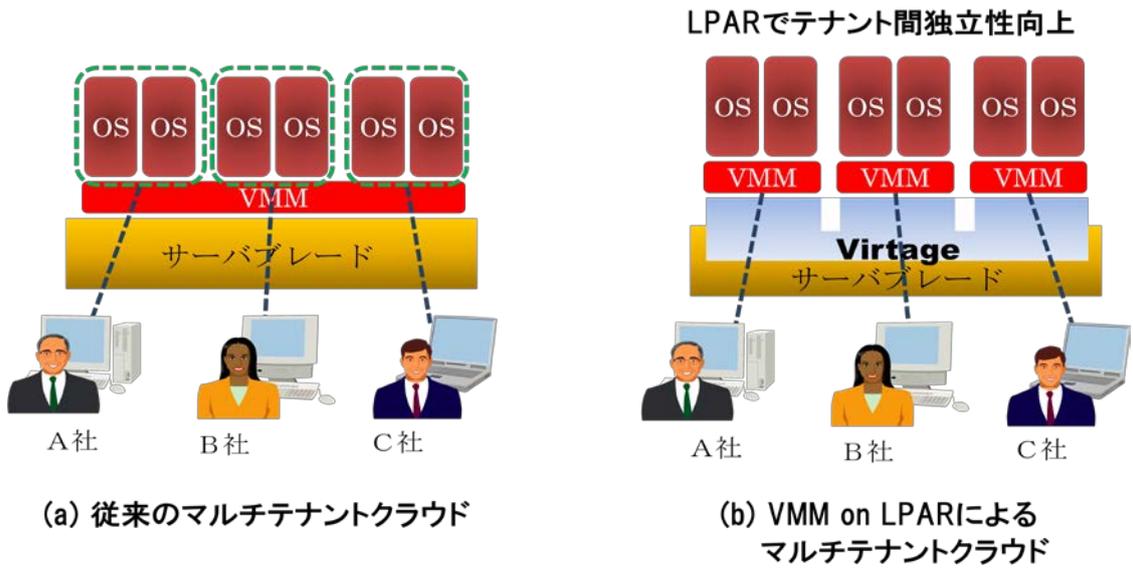


図2 VMM on LPAR によるマルチテナントクラウド基盤の提供

(2) VMM as a Service

本機能をホスティングサービスに適用すれば、従来からある VM 単位のホスティングだけでなく、VMM を丸ごとレンタルする「VMM as a Service」が可能となる(図3)。エンドユーザーはホスティング事業者から CPU〇〇個、メモリ〇〇ギガバイト、ディスク容量〇〇ギガバイトといった形のハードウェアリソースを借用し、その上で何台の仮想マシンを生成するかはユーザー自身により自由に定義可能となる利点がある。レンタルされている VM 構成に変更があった際でも毎回ホスティング事業者に変更を依頼する必要がなく、運用の柔軟性が向上する。一方、ホスティング事業者にとって従来は、それぞれの VMM 毎に物理サーバを個々に準備する必要があったが、近年のマルチコア化により物理サーバ単位にするとそのリソース粒度が大きく、リソースに無駄が生じ、コスト増を招く状況であった。これに対し、VMM on LPAR 技術により、エンドユーザーの案件の規模に合わせて細かな粒度で柔軟にリソースを割り付けることが可能となり、リソースの無駄を省くことができる。さらに、エンドユーザーの負荷変動にも柔軟に対応可能となる。

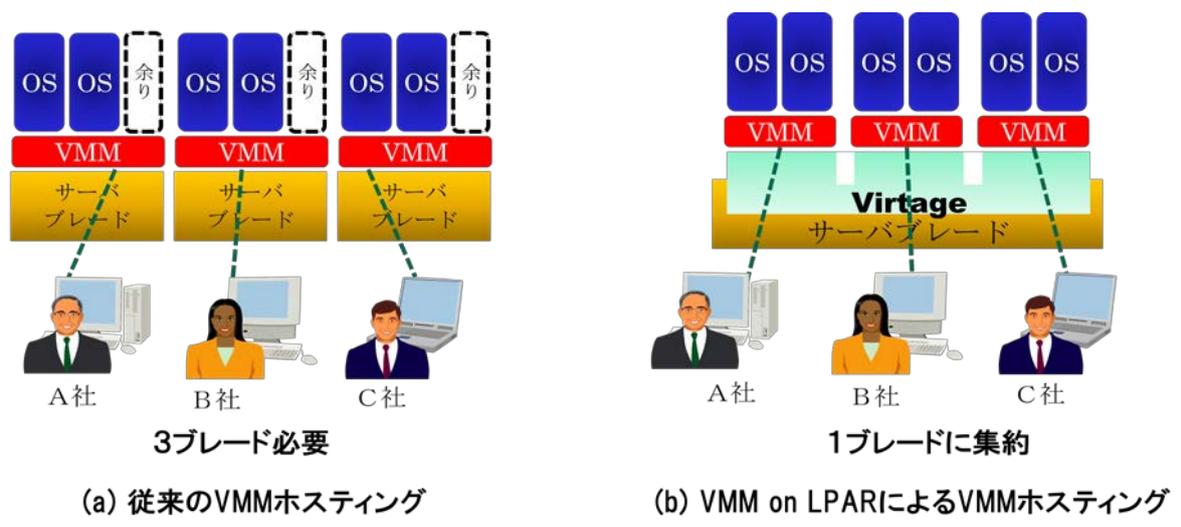


図3 VMM on LPAR による VMM ホスティング

その他、基幹/仮想混在コンソリデーションやVMMのマルチバージョンの混在やマルチVMM環境のコンソリデーションなどが想定される。

### 3 VMM on LPAR の課題と VMCS Shadowing の効果

本節では、VMM の動作に必要となる Intel VT の概要と、LPAR 上で VMM を動作させる VMM on LPAR の課題を述べ、課題を解決する VMCS Shadowing について説明する。

#### 3.1 サーバ仮想化を支援する Intel VT

一般的なサーバ仮想化では、1 台の物理サーバ上で単一の VMM と複数の OS が動作する(図4)。物理サーバは CPU や Storage などを含むが、Storage は複数の OS で共有するケースが多い。複数の OS を安全に稼働させるために Intel® Virtualization Technology(以下、Intel VT とする)は、必要に応じて VMM を呼び出す機能を提供している。また Intel VT は、OS の状態を VMCS に保持し、VMCS へのアクセスする手段を VMM に提供している。

例えば OS が Storage 操作を要求すると、Intel VT が VMM を呼び出す。呼び出された VMM は VMCS を読んで OS の状態を把握し、他 OS の存在を考慮しつつ Storage を代行操作する。VMM はその結果を VMCS に反映し、OS を再開する(図5)。

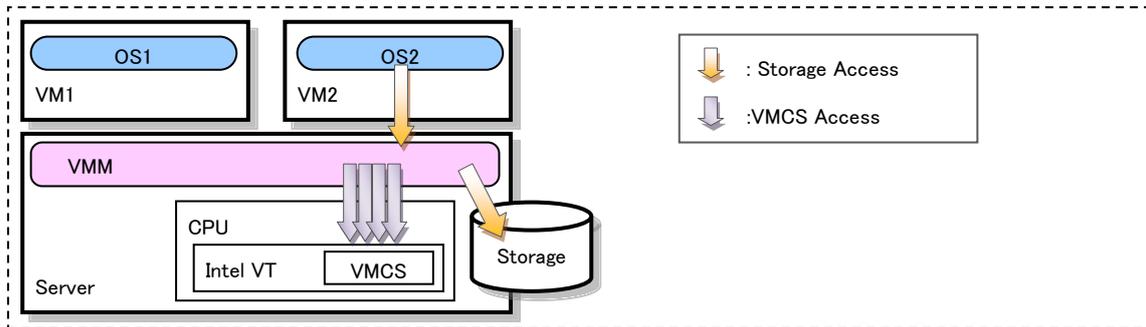


図4: 一般的なサーバ仮想化の構成図

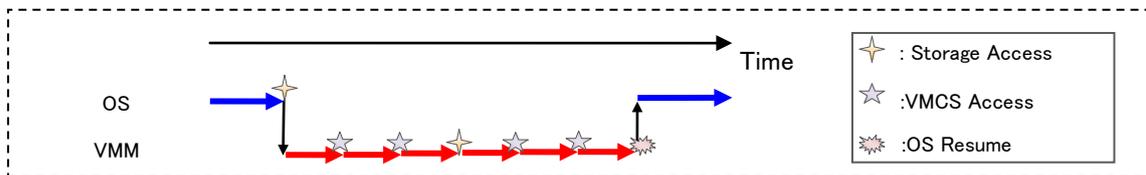


図5: 一般的なサーバ仮想化の動作タイムチャート

3.2 VMM on LPAR の仕組みと課題

VMM on LPAR では Virtage が LPAR を作成し、VMCS を使って VMM を稼働させる。更に VMM の動作にも VMCS が必要であるため、Virtage は VMM 用の VMCS (Shadow VMCS と呼ぶ) を Virtage 内部に作成する。そして VMM が VMCS にアクセスする際に Virtage が Shadow VMCS を操作する(図6)。

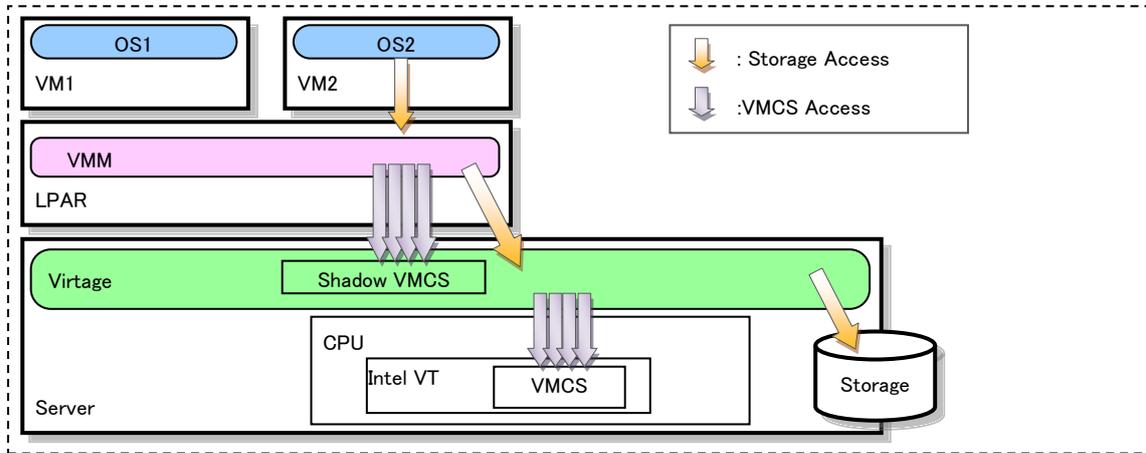


図6: 従来の VMM on LPAR 構成

例えば OS が Storage を操作する場合は、Intel VT が操作を認識して Virtage を呼ぶ。Virtage は VMCS に保存された OS の状態を Shadow VMCS にコピーしてから VMM を呼ぶ。続いて VMM が VMCS を操作する時には、Virtage が Shadow VMCS を操作する。そして VMM が OS を再開する際は、Virtage が Shadow VMCS の内容を VMCS にコピーして OS を再開させる(図7)。

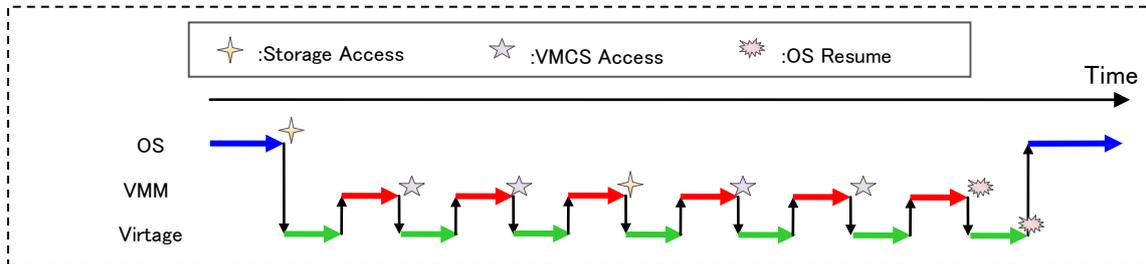


図7: 従来の VMM on LPAR の動作タイムチャート

VMM は1回のエミュレーションで VMCS を複数回参照・更新する。そのため VMM on LPAR では Storage 操作等による VMM の呼び出し頻度が高い程、Virtage の動作時間が増えて性能が下がり易いという課題があった。

### 3.3 課題を解決する VMCS Shadowing

本課題を解決するために Intel Xeon Processor E5 V3 Family は、Shadow VMCS 操作を CPU に実施させる VMCS Shadowing を実装した。本機能の定義と開発には日立のエンジニアも密接に貢献している。VMCS Shadowing に対応する CPU は VMCS だけでなく Shadow VMCS も保持する(図8)。そして LPAR 上の VMM が VMCS を操作する際に、Virtage を呼ばずに Shadow VMCS を直接読み書きする。

VMCS Shadowing を用いると、Shadow VMCS 操作に起因する Virtage 動作を省けるため、VMM on LPAR の性能が大きく改善する(図9)。

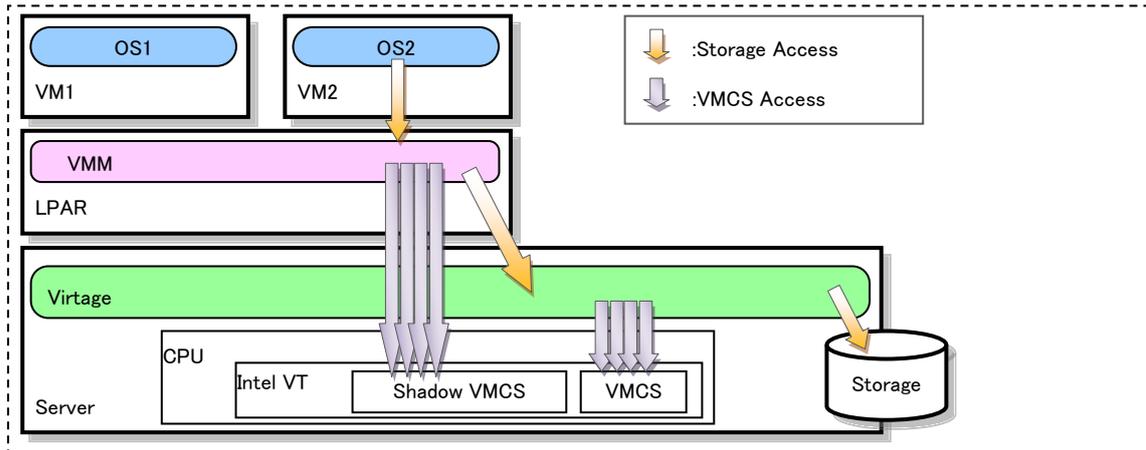


図8: VMCS Shadowing を使う VMM on LPAR 構成

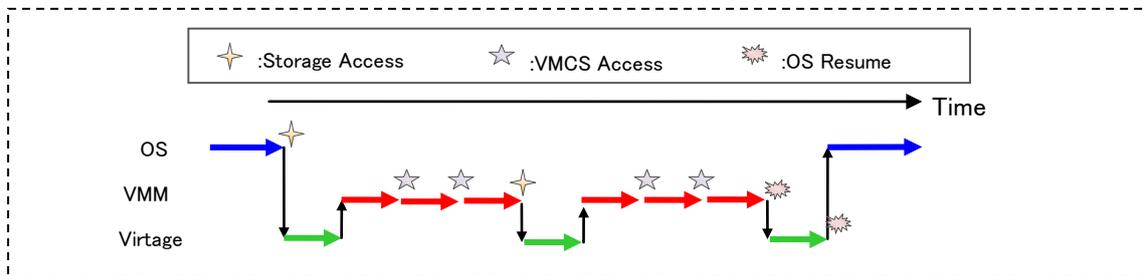


図9: VMCS Shadowing を使う VMM on LPAR の動作タイムチャート

## 4 測定/評価結果

### 4.1 測定環境

VMCS Shadowing の効果を測定するために、Intel Xeon CPU E5-2667 v3 を搭載した最新の BS500 を用意し、VMM on LPAR 及び VMCS Shadowing をサポートするプロトタイプ版 Virtage をインストールした(表1)。Virtage 上で表2に示す構成の LPAR を作成し、同 LPAR で Red Hat Enterprise Linux6.4 同梱の Kernel-based Virtual Machine(以下、KVM とする)を動作させた。また KVM 上で表2に示す構成の VM を作成し、Red Hat Enterprise Linux6.4(64bit)を稼働させた。VMM on LPAR で性能低下が生じ易い状況を代表するベンチマークとして、Storage アクセスを高頻度で繰り返す IOmeter と、Database の実運用を模擬する sysbench の性能を測定した(表3)。

表1:ハードウェア構成

サーバシャーシ	BS500
サーバブレード	BS520H B3
統合ファームウェアバージョン	08-00
Virtage	VMCS Shadowing プロトタイプ版(*1)
CPU	Intel® Xeon® CPU E5-2667 v3 @ 3.20GHz (8Core) x2 (Hyper-Threading Disabled)
メモリ	64GB
内蔵ファイバチャネルスイッチ	Brocade 8Gb ファイバチャネルスイッチモジュール
ファームウェアバージョン	072c
ストレージ	Hitachi Adaptable Modular Storage 2300
ファームウェアバージョン	08C4/F-M

\*1:先端技術の為、特別にチューンアップしており、一般提供しているバージョンと異なります。

表2:LPAR/VM 構成

OS	RHEL 6.4 x64 (Kernel 2.6.32-358.el6.x86_64)	
VM	CPU	4Core (cpu64-rhel6)
	Memory	8GB
VMM	KVM (RHEL 6.4 同梱)	
LPAR	CPU	6Core
	Memory	10GB

表3:ベンチマーク設定

benchmark	configuration
iometer-2006_07_27	1 Thread 4K; 100% Read; 0% random
sysbench 0.4.12	--test=oltp --oltp-table-size=1000000 --num-threads=4 --max-requests=0 --max-time=180 --oltp-test-mode=complex (DB Server は MySQL 5.1.66)

## 4.2 評価結果

VMCS Shadowing を無効化した LPAR(従来 LPAR)と、VMCS Shadowing を有効化した LPAR でそれぞれ測定した結果を図10に掲載する。縦軸は従来 LPAR を 1.0 とする相対性能である。VMCS Shadowing によって、Storage アクセスを高頻度で繰り返す IOmeter では 1.8 倍もの性能向上が確認された。Database の実運用を模擬する sysbench でも 1.4 倍の性能向上が確認された。

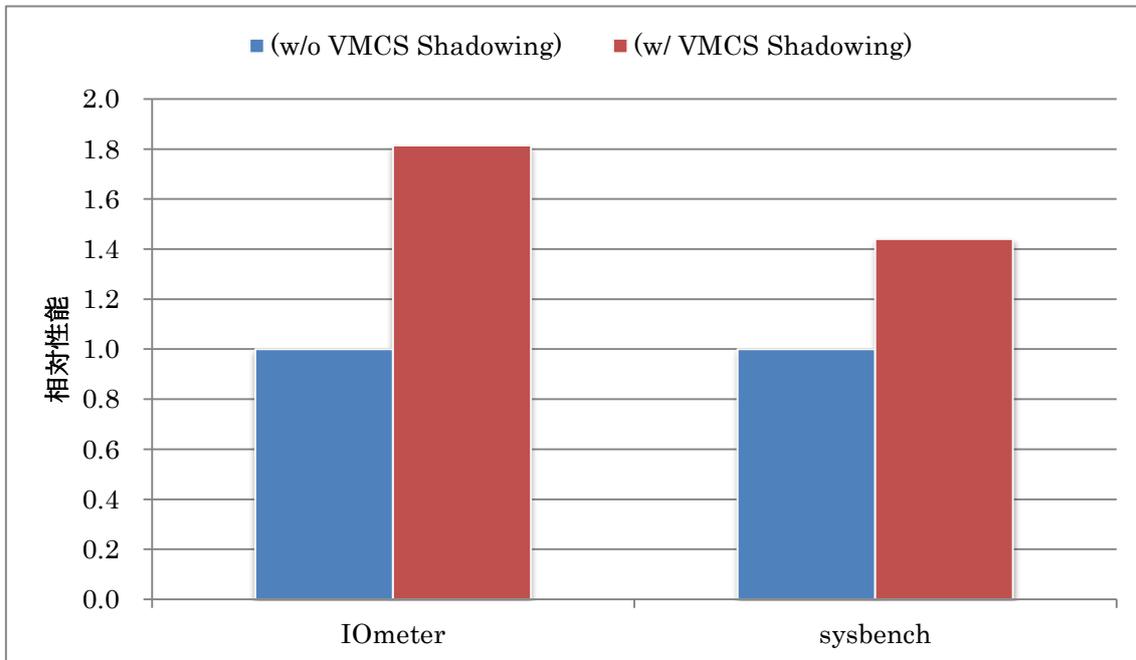


図10: VMCS Shadowing の有無による性能差

## 4.3 考察

今回の測定では、Storage 操作頻度が高い IOmeter の方が、VMCS Shadowing の効果がより大きく表れている。これは VMM の VMCS 操作頻度が高い程、VMCS Shadowing の恩恵を受け易いためである。この様に VMCS Shadowing を用いると、負荷の大小による性能差を縮める効果が得られる。従来の VMM on LPAR では、Storage 操作の頻度によって性能に大きな影響が出るため、ピーク負荷に耐える VM・LPAR 設計が難しかった。しかし、VMCS Shadowing によって負荷の大小に起因する性能差が縮まるため、VMM on LPAR の適用対象が拡大する見込みである。

## 5 実ユースケース

本章では、本機能を活用し、投資・運用コストを削減された株式会社東光高岳殿の事例を紹介する。株式会社東光高岳殿は、電力会社向け受変電設備や公共産業向け電力設備を中心に事業を展開する重電設備のリーディングカンパニーである。

従来より、経理や生産管理などの基幹業務システムを、先進技術を導入しながら、日立ブレードサーバ BS500 や日立ストレージソリューション HUS 110 を活用し構築している。高い信頼性や堅牢性が必要な業務は LPAR、レガシーな環境で稼働している業務集約には VM を活用するなど、業務に応じ適切なサーバ仮想化技術を活用し、システム構築を行ってきた。しかし、M&A などによる業務拡大に伴うシステム拡張などにより、その投資・運用コストが大きな課題となってきた。

そこで同社は、VMM on LPAR を活用し、図 11 に示すシステムを構築した。本システムでは 2 つの LPAR が動作しており、LPAR1 では、基幹業務である社内経理システムが稼働している。一方、LPAR2 では、他社の VMM が動作しており、その Level 2 ゲストである Microsoft Windows Server 2003 上で前世代の経理業務が稼働している。本業務は、移行コストの理由により、新 OS への移行が困難であるが、一方で、負荷は高くないものの、継続した稼働が求められている。また、同様にレガシーシステムである社内情報システムの問い合わせサービスシステムも Level 2 ゲストとして稼働しており、他社 VMM 環境での運用と同様に VM の追加などの運用も可能である。

同社より、以下のコメントを頂いている。

『日立が提供する VMM on LPAR 技術にはとても満足している。基幹業務の統合に適した LPAR とレガシー環境の集約に適した VM との融合により、サーバリソースのさらなる有効活用により物理サーバ数を削減できるため、投資・運用コストの双方の削減が可能となった。特にスペース・消費電力の削減に効果を発揮している。まさに、我々が望んでいたことが実現できたと考えている。』

『Intel 社と日立の密接な技術的協業により、プロセッサの進化と合わせて、VMM on LPAR 技術の適用範囲がますます拡大すると聞いており、今後の技術革新にますます期待している。』

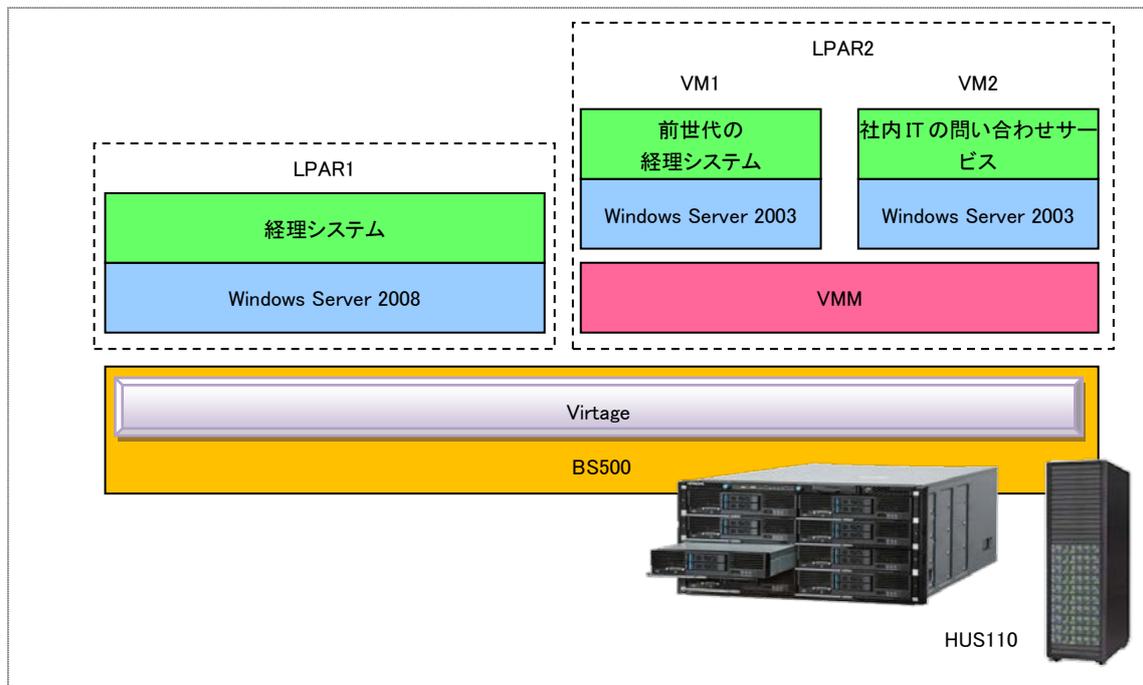


図 11: 株式会社東光高岳殿のシステム構成

## 6 おわりに

本稿では、Virtage の LPAR 上で他社の VMM を稼動する Nested Virtualization (VMM on LPAR)機能と、日立も開発に貢献した Intel Xeon Processor E5 V3 Family の VMCS Shadowing による性能改善効果について示した。さらに、本技術のユーザサイトでの試行例を紹介した。日立は、引き続き本技術の発展を Intel とともに図っていく所存である。