Citrix XenServer参考資料

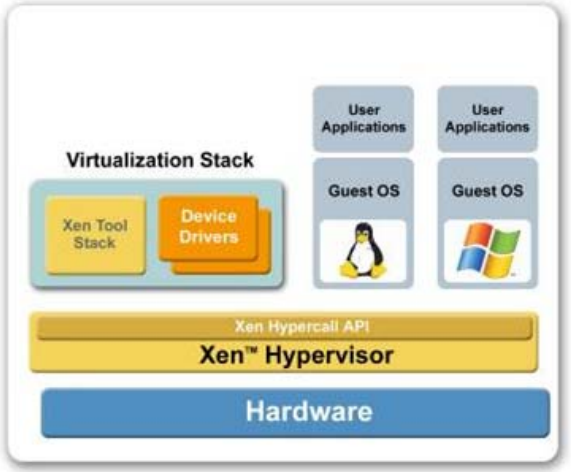
シトリックス・システムズ・ジャパン株式会社

2009年5月29日

CITRIX®

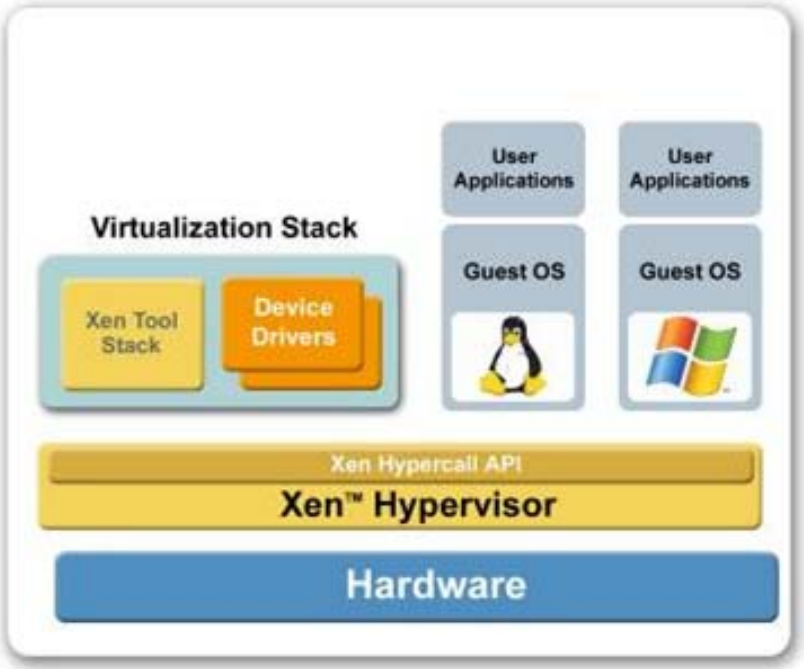
アーキテクチャと関連コンポーネント

Xen



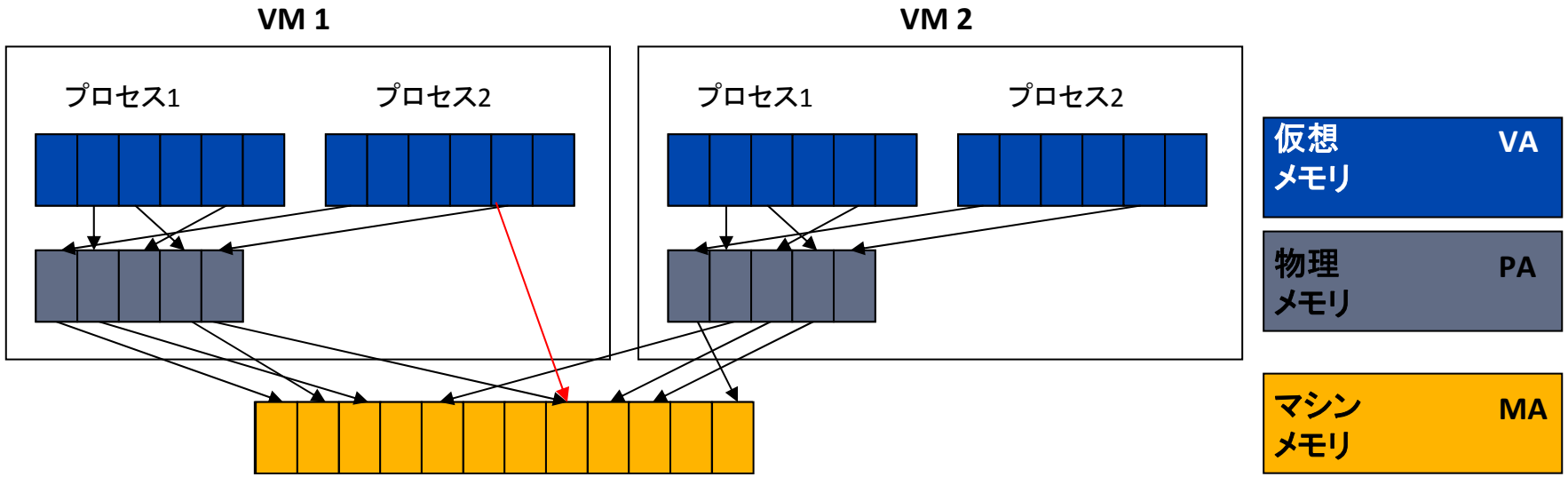
- ポリシーとメカニズムの分離
  - Xenハイパーバイザーがメカニズムを実装
  - ポリシーはDomain0に委ねる
- Xenハイパーバイザー
  - ゲストから物理デバイスへの直接アクセスが可能となるメカニズムを提供
  - 信頼性のため機能を絞り込みコードを小さくしている
- Domain0
  - デバイスの処理とユーザインターフェースの提供
  - XenServerは実績あるCent OSとLinuxデバイスドライバーを使用
  - 信頼性の向上と広範囲に及ぶハードウェアの互換性

XenServer アーキテクチャー



- 小さく効率的なハイパーバイザーはハードウェアの仮想化に適している
- ゲストはリソース管理とI/Oのため協調して実行される
- デバイスドライバーはハイパーバイザーの外に実装
- 飛躍的なパフォーマンス向上
- Xen-Hypervisor – オープンソース

仮想メモリの仮想化: シャドウページテーブル



- VMMは「シャドウページテーブル」を使い割り当てを高速化
  - シャドウはVA -> MAへ直接割り当て
  - アクセスのたびに起きるふたつのレベルの変換を回避
  - このVA -> MA割り当て用にTLBハードウェアを使用
  - ゲストOSがVA -> PAを変更した場合、VMMはシャドウページテーブルを更新

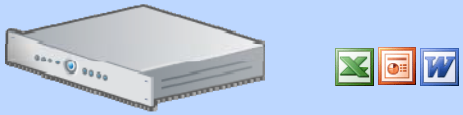
シャドウページテーブル
3つのパフォーマンストレードオフ

1. トレースのコスト
   - VMMはゲストによるプライマリページテーブルへの書き込みをインターセプトしなければならない
   - 変更をシャドウページテーブルに伝える(または無効にする)
2. ページフォルトのコスト
   - VMMはページフォルトをインターセプトしなければならない
   - シャドウページテーブルエントリ (ヒドゥンページフォルト)を有効にするか、フォルトをゲスト(実際のページフォルト)にフォワードする
3. コンテキストスイッチのコスト
   - VMMはゲストのコンテキストスイッチをインターセプトしなければならない
   - シャドウページテーブルの新たなセットをアクティベートする

パフォーマンスの実現には、良いトレードオフの発見が欠かせない

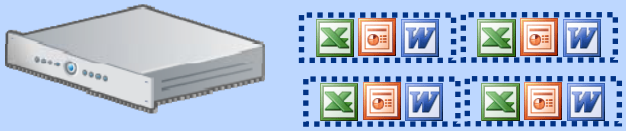
シャードー ページ テーブルの多重化

物理マシン



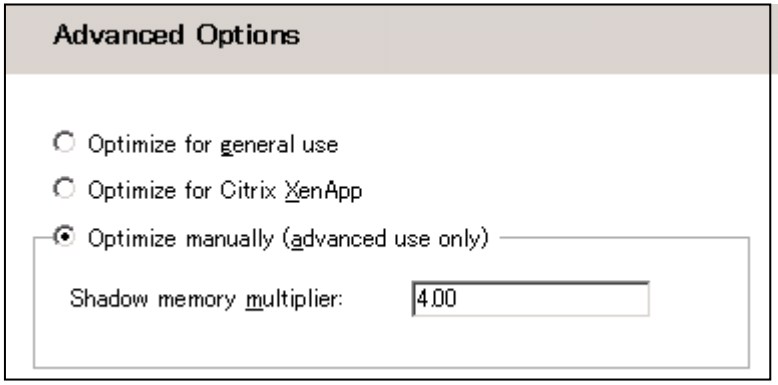
- 物理サーバ(Proliant DL360 G5)
(3～4年前の古いサーバとして設定)
- 2 CPU (2.0 Ghz)
- 4 GB メモリ
- EdgeSightのロードバランサーを使用して75ユーザをシミュレート
- ユーザスクリプト実行時間 – 4:33

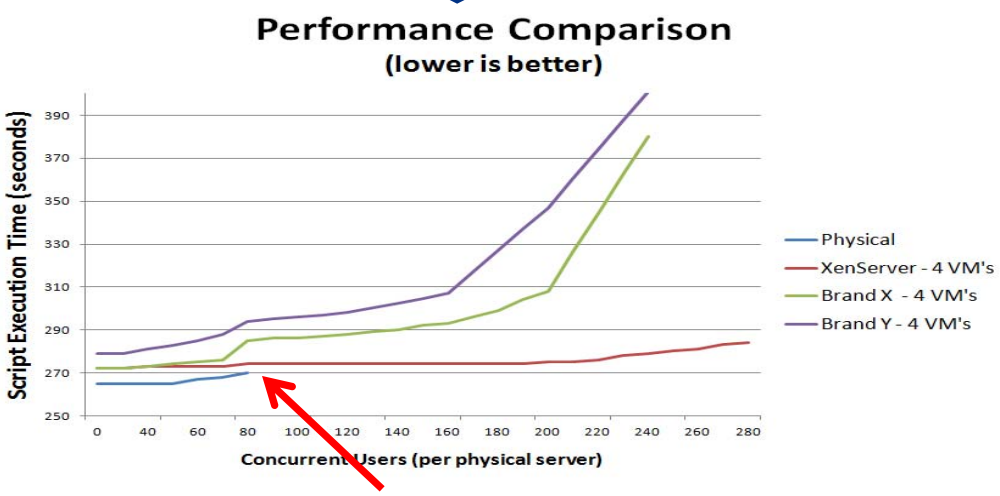
仮想マシン



- 1サーバに4仮想マシン (Proliant DL380 G5)
- 8 プロセッサ コア(VMに2 vCPU, 3.0 Ghz)
- 16 GB メモリ(VMあたり3.5 GB)
- EdgeSightのロードバランサーを使用してシミュレート（VMごとに）
- XenServer, “Brand X”と “Brand Y” でテスト

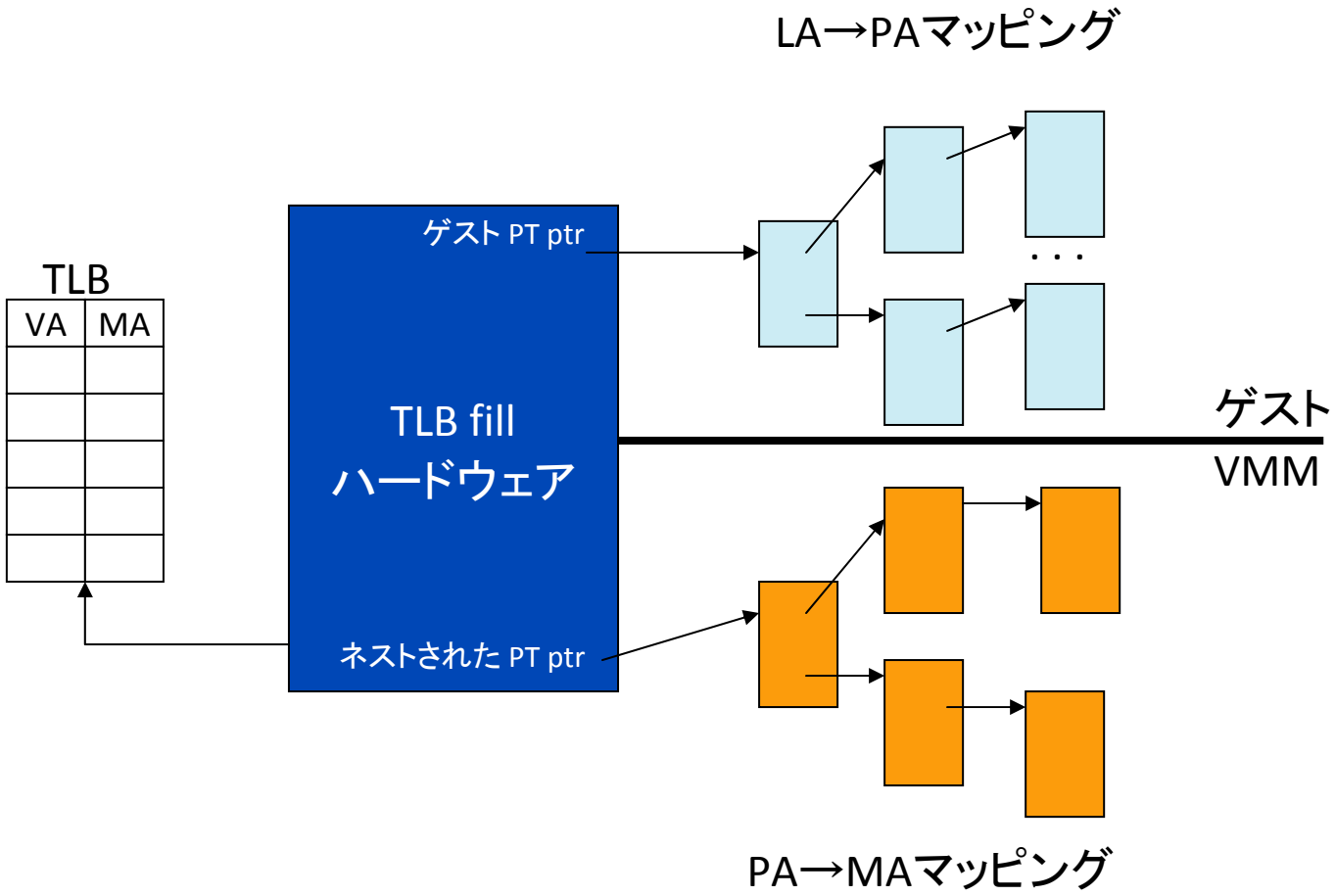
XenApp (Presentation Server)のテスト





物理サーバのメモリが90セッション以下で使いつくされた

ハードウェアサポート：EPT/RVI(NPT)



RVI (Rapid Virtualization Indexing): AMD Barcelona CPUからサポート
EPT (Extended Page Table): Intel Nehalem CPUからサポート

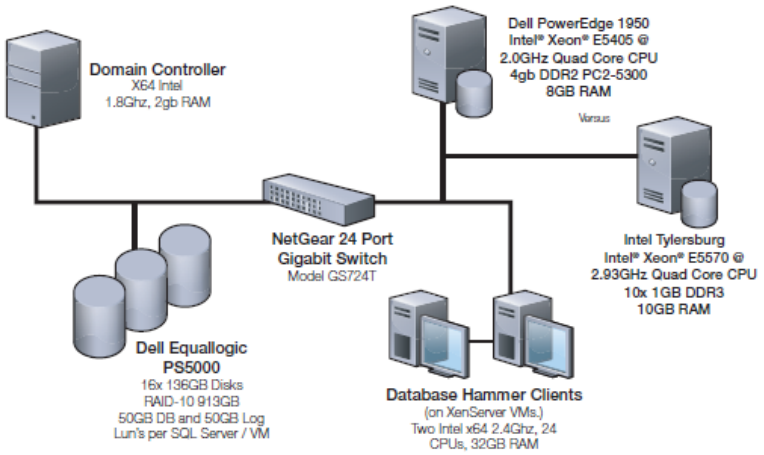
RVI/EPTの分析

- ハードウェアMMUは、オンザフライでLA->PAとPA->MAマッピングを生成
- メリット
  - 「exit頻度」の激減 – パフォーマンスへの寄与
    - トレースフォルトがない(ネイティブと同じ速度のプライマリページテーブル変更)
    - ページフォルトがexitを必要としない
    - コンテキストスイッチがexitを必要としない
  - シャドウページテーブルのメモリオーバーヘッドがない
  - パフォーマンス
    - MMUオーバーヘッドによるワークロードの劇的な改善
  - さらに多くのワークロードが仮想化の候補に

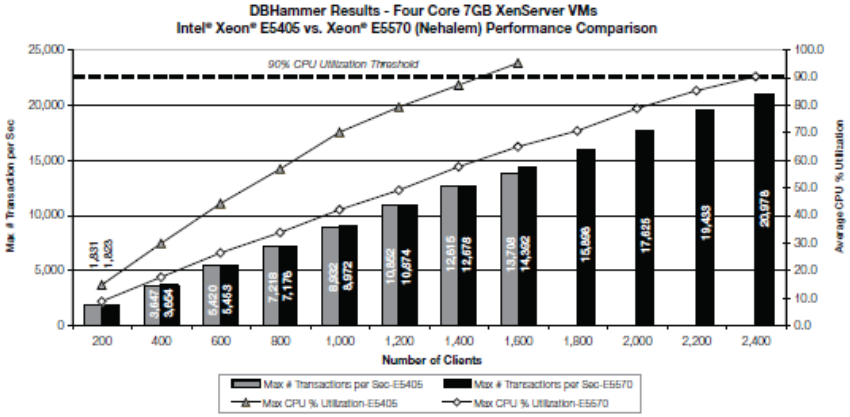
EPTパフォーマンス結果

DBHammer results for E5405 vs. E5570, Single XenServer Host VM, 4 cores, 7GB

# Client	Avg % CPU util-E5405	Avg % CPU util-E5570	Max trans per sec-E5405	Max trans per sec-E5570
200	14.7	8.7	1,831	1,832
400	29.7	17.5	3,647	3,654
600	44.2	26.3	5,420	5,453
800	56.7	33.7	7,218	7,176
1000	70.0	41.9	8,932	8,972
1200	79.1	49.0	10,852	10,874
1400	87.1	57.5	12,615	12,678
1600	95.0	64.8	13,708	14,392
1800		70.5		15,898
2000		78.6		17,625
2200		85.1		19,433
2400		90.2		20,978

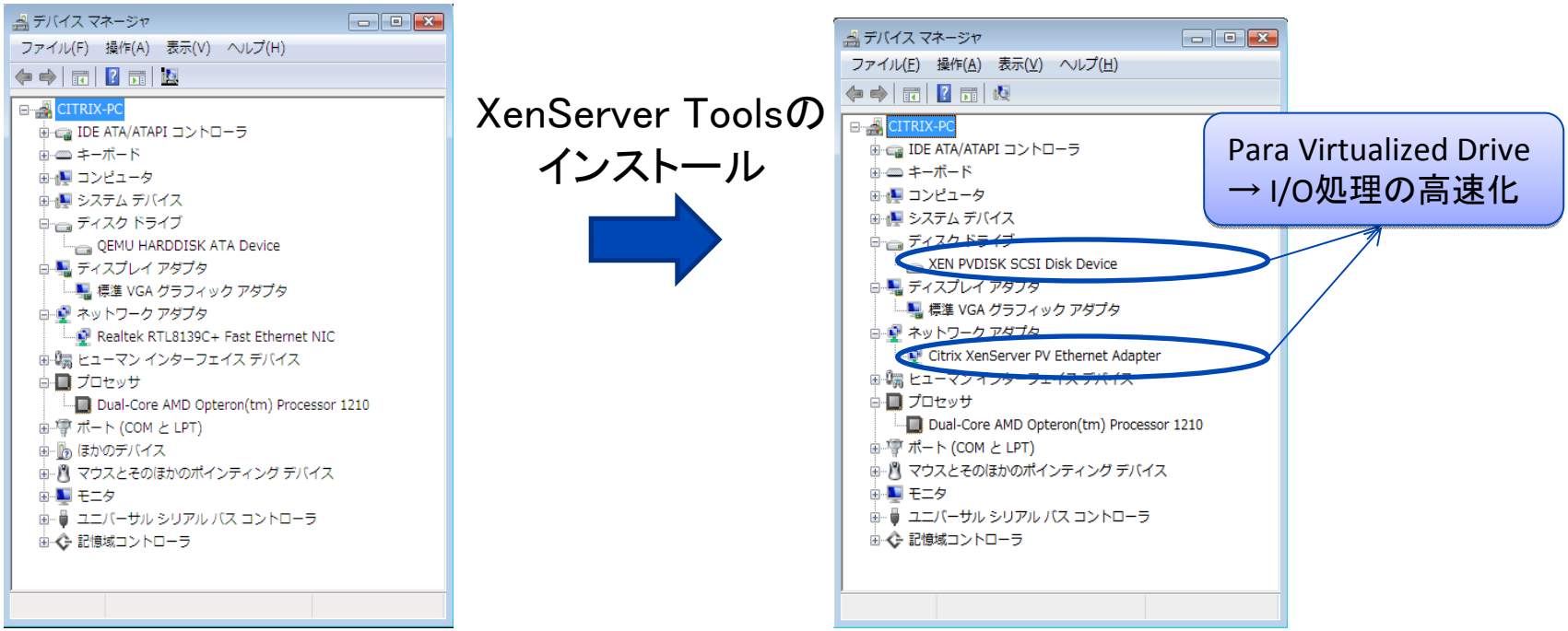


テスト環境



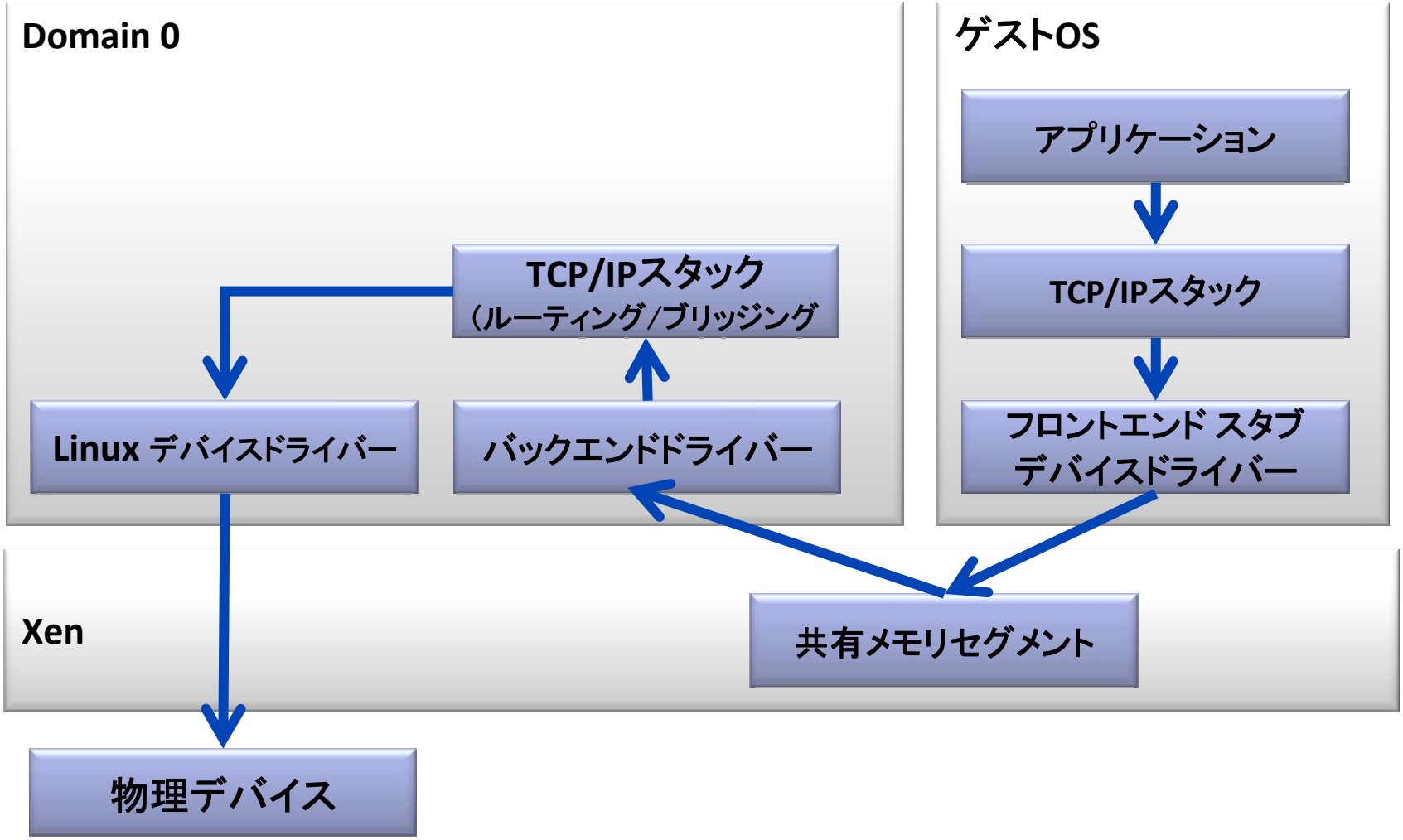
XenServer仮想マシン

仮想マシンのデバイスは仮想化されておりハードウェアの依存性がない
Para Virtualization ドライバーによる I/O の高速化



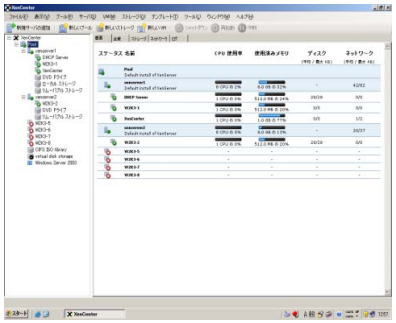
I/O処理の効率化

- メモリコピーを行う必要なく、XenのバスがDomain0にポイントする

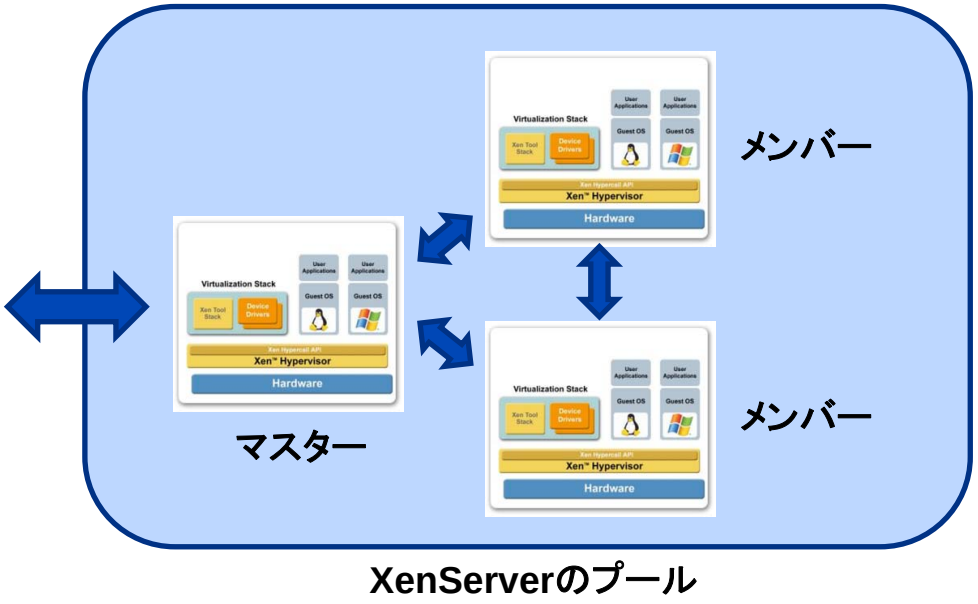


仮想化ソフトウェアの一元管理 – XenCenter

- XenServerの管理はXenCenterから実施
- XenServer プールのマスタサーバーに接続



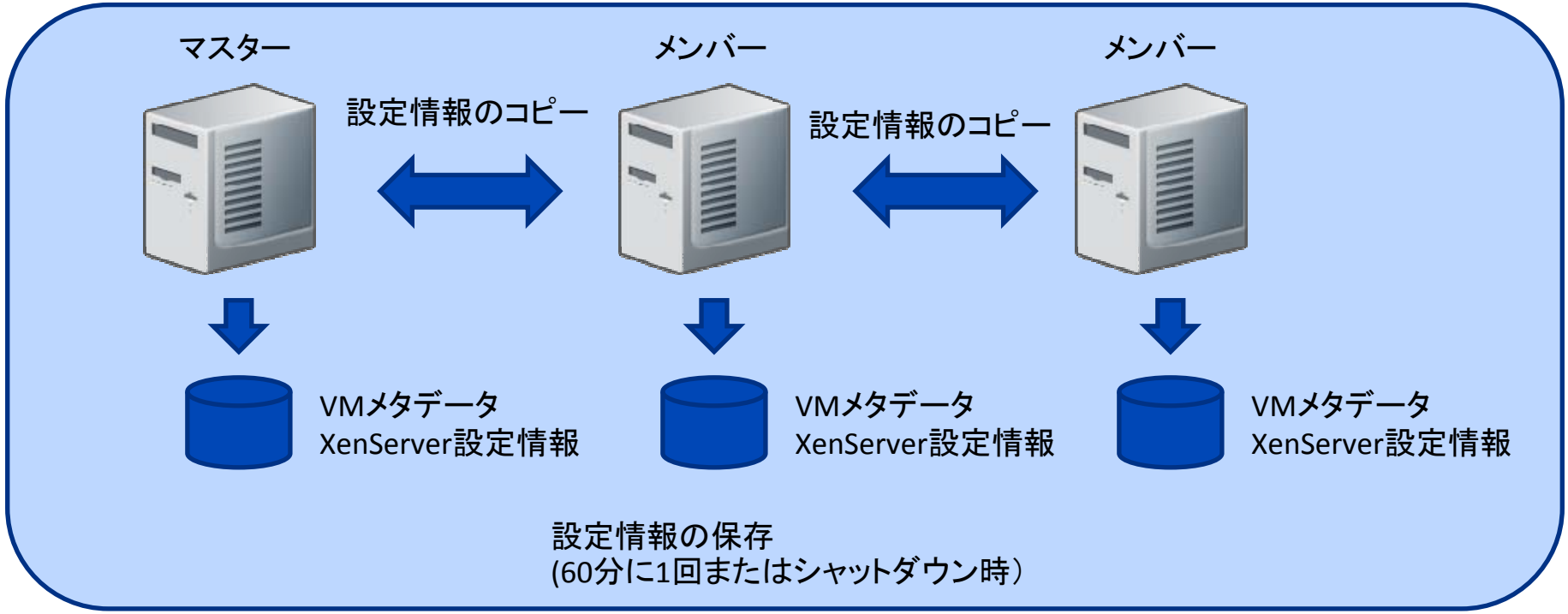
日本語版XenCenter



XenServerのプール

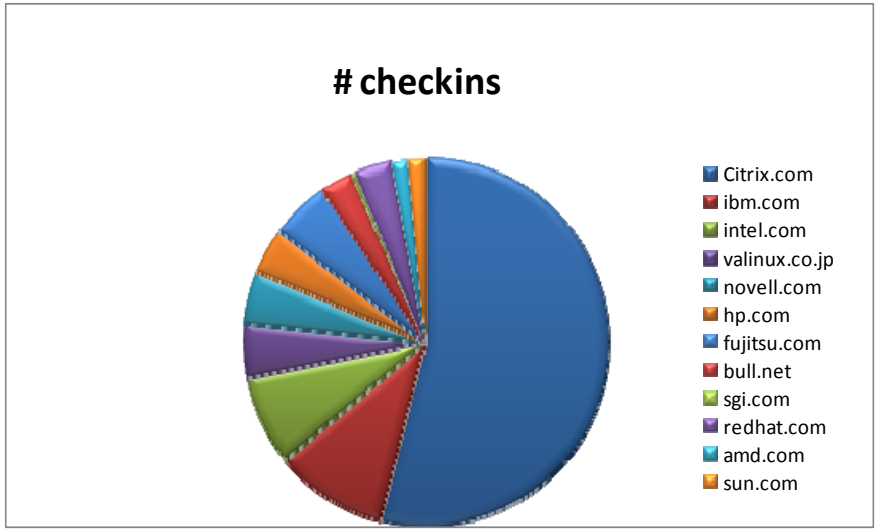
XenServer管理の仕組み

- プール内の全てのXenServerはネットワーク、共有ストレージは同じ設定となる
- 設定変更は即座に他のXenServerに通知され設定される
  - 全てのXenServerが同じ設定内容を保存する
- SPOFにならない設計を採用



XenとXenServerのリレーションシップ

- Xen コミュニティーはCitrixによってリードされている
- 新しいハードウェアと機能を敏速にマーケットに投入
- メジャーなアクティブ貢献者: IBM, Intel, valinux, HP, AMD, Red Hat, Novell



Xen

XenServer

オープンソースXenとXenServerの比較

- ✓Xen ハイパーバイザーはオープンソースとXenServer製品両方のためのコア仮想化エンジン
- ✓ハイパーバイザーはベアメタルで動く次世代の仮想化で、準仮想化とハードウェア仮想化支援機能をサポートする

オープンソースXenとXenServerとの違い

- オープンソースXenコードは不安定な、またはテストされていないコンポーネントを含んでいる
- Citrix XenServerは厳しくテストされてミッションクリティカルな本番での使用のために洗練された機能を含んでいる
- CitrixはXenハイパーバイザーにエンタープライズレベルの独自機能を追加している

Xenプラットフォーム比較

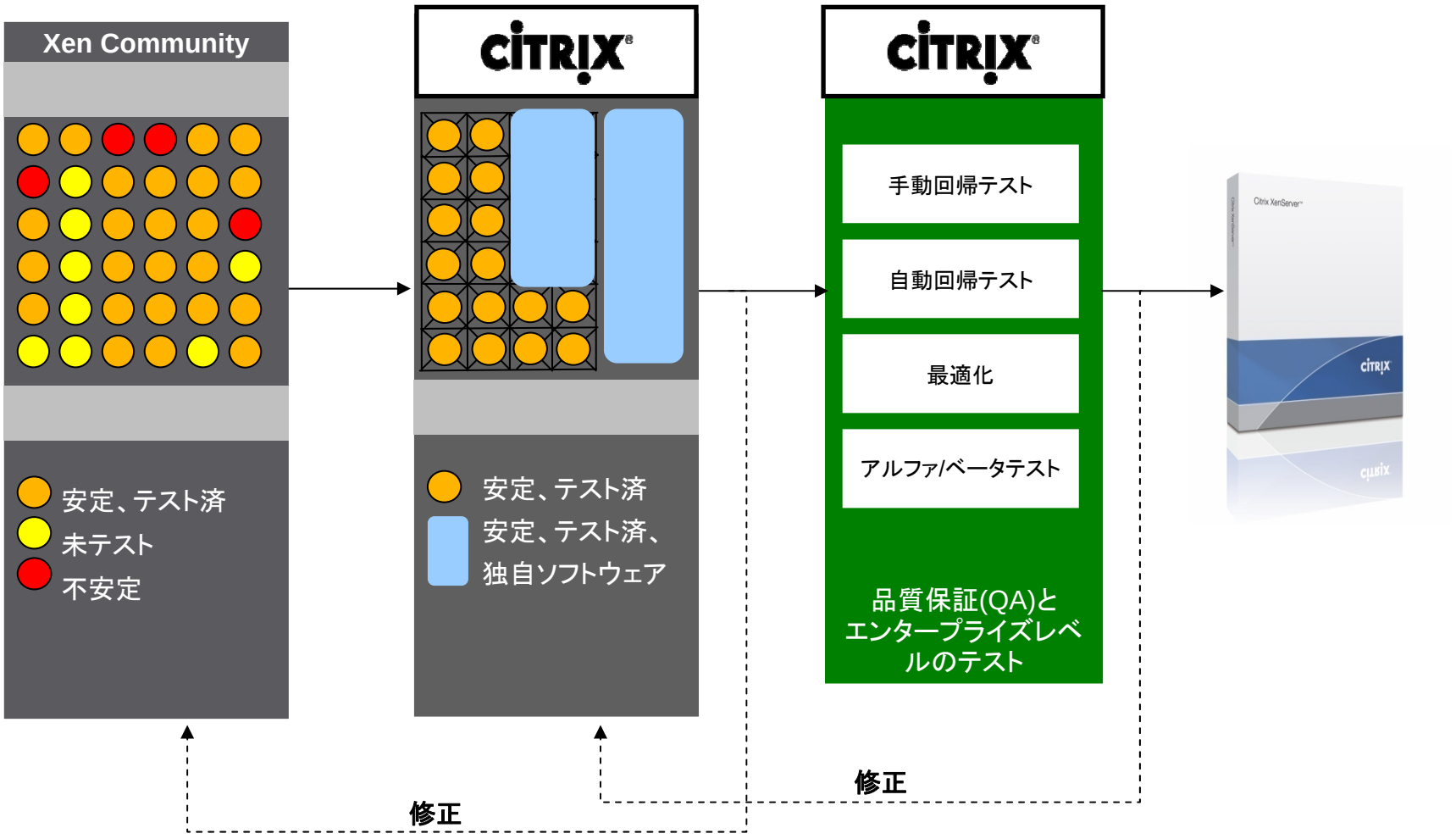
XenServer

- Citrix XenServer 5.0 は64-bitハイパーバイザーを使用する
- Citrix XenServerは以下のCPUとメモリの制限がある:
  - 32 CPU コア
  - 128 GB メモリ
- 75%のコードはXenServerを容易に利用するために独自に開発されている。インストールは通常10分以内で終了する

Xen Open Source

- Xenは以下のハイパーバイザーバージョンをサポート:
  - 32-bit
  - 32-bit-PAE
  - 64-bit
- Xenは理論的に以下のCPUとメモリの制限を持つ:
  - 128 CPU コア
  - テラバイトのメモリ
- XenはインストールにLinuxの知識を必要とする。インストールは通常1日かかる

Xen™ vs. XenServer

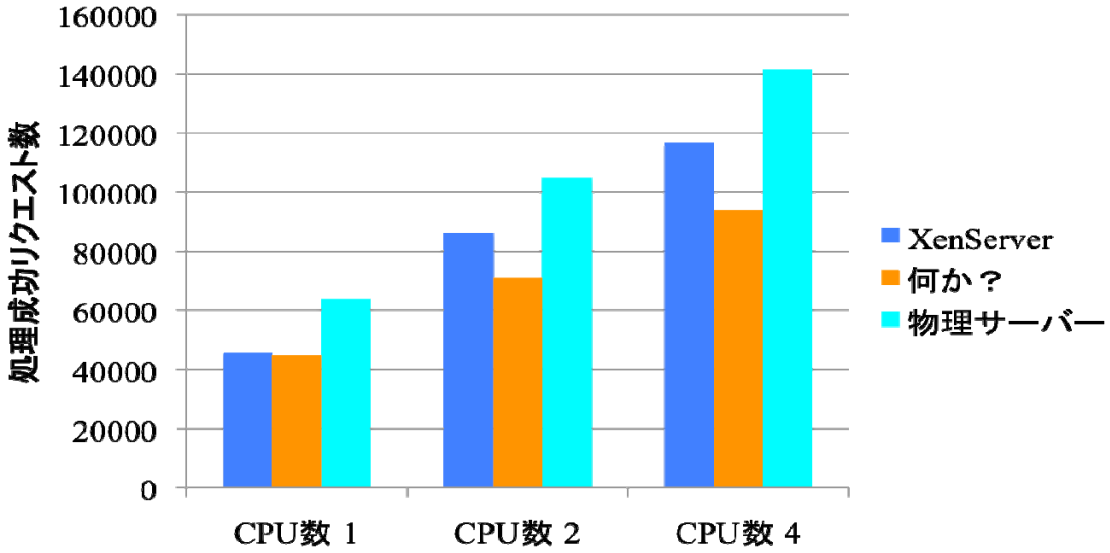


CITRIX®

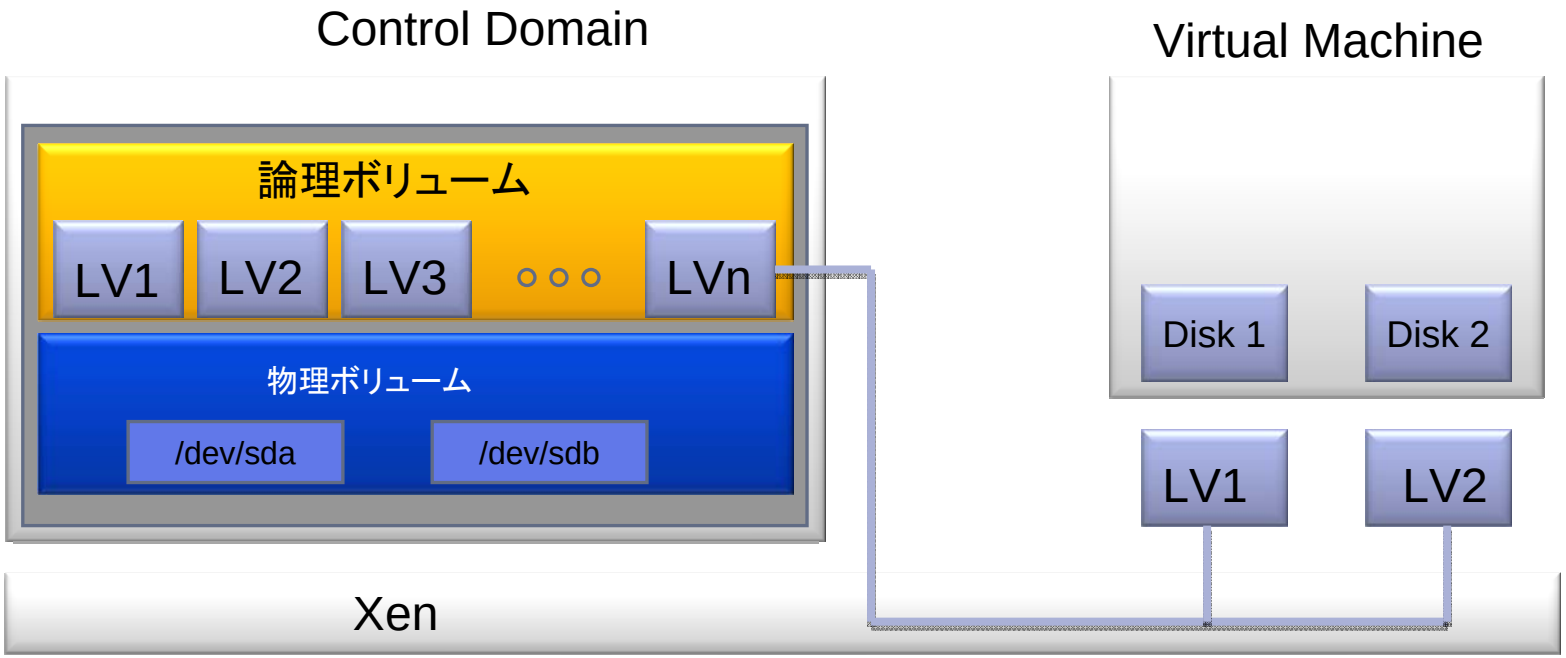
物理リソースの有効活用

マルチプロセッサー オーバーヘッド

- SPECwebによるベンチマークを実施
- 仮想マシンの割当仮想CPU数を変更
  - それぞれ1仮想CPU、2仮想CPU、4仮想CPU
- 物理マシン、XenServer、何か？を比較
- ピーク値性能で比較
  - ピーク値での同時接続数はそれぞれ異なるが、処理成功リクエスト数と相関関係にあるので考慮しない



ストレージ アーキテクチャ

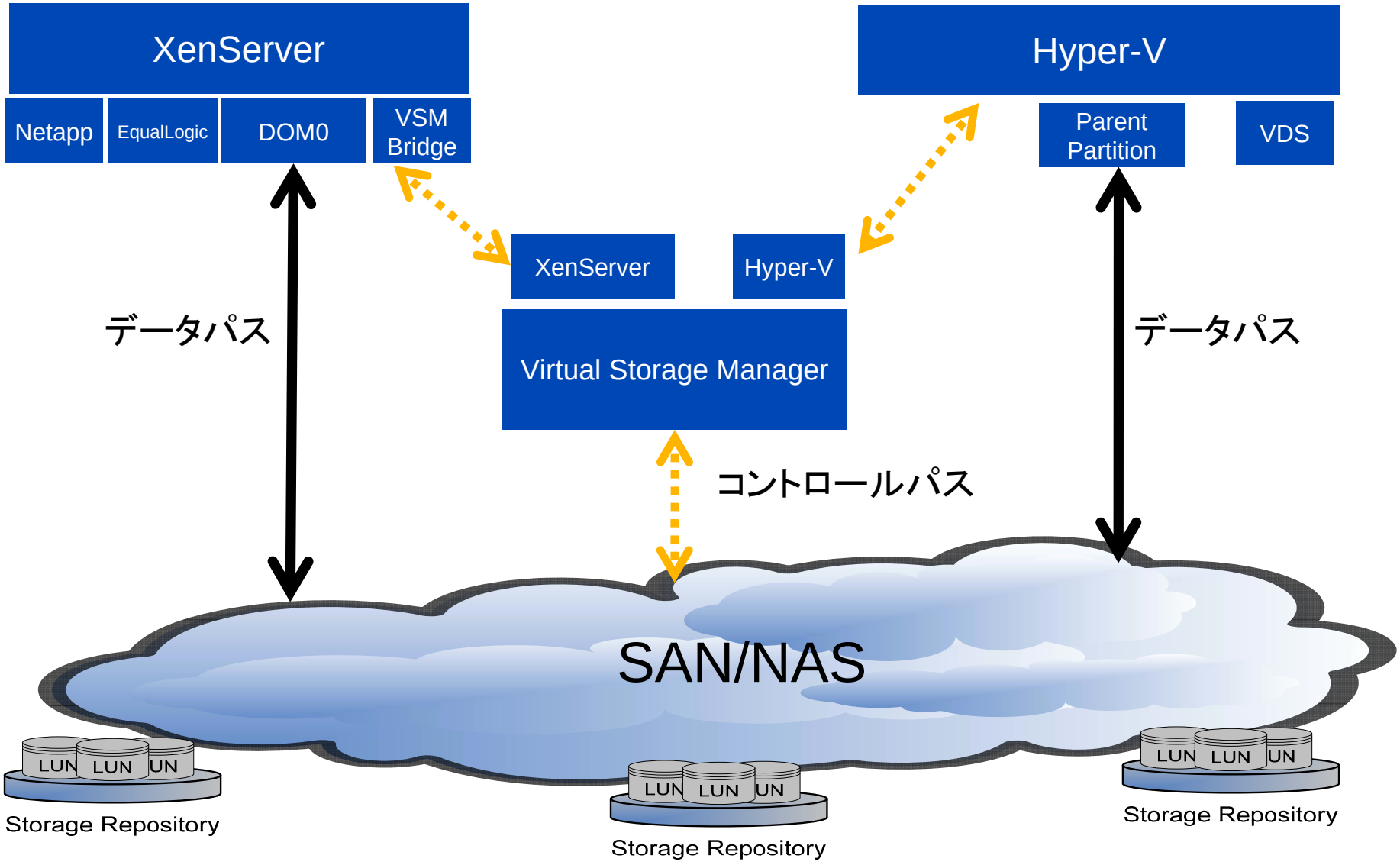


- XenServerはストレージ管理にLinuxの論理ボリュームマネージャ(LVM)を使用
  - 論理ボリューム、仮想ディスクの拡張が容易
- 論理ボリューム（パーティション）はプールから作成され仮想マシンに個別の仮想ディスクドライブとしてエクスポートされる
- 論理ボリュームは仮想マシンのOSによってフォーマットすることにより仮想マシンからのI/Oを高速する
- 仮想マシンのフロントエンド スタブ デバイスドライバーはOSのためのデバイスドライバーに似ている
- フロントエンドドライバーはコントロールドメインのバックエンドドライバーのディスクリソースにXenを経由して接続される

StorageLink

StorageLink Gateway：
仮想コンソールからストレージの機能を直接制御
StorageLink Converter：
ファイバーチャネルとiSCSI間で自動的にコンバージョン
StorageLink Resource Manager：
既存のアレイを利用した仮想化管理
StorageLink Image Manager：
集中化されたイメージライブラリから仮想マシンを敏速にデリバリ
StorageLink HyperPass：
XenServerとHyper-V間で仮想マシンをシームレスに移動
StorageLink Connect 容易にあらゆるサードパーティのバックアップまたは管理フレームワークにリンク

Storage Link/ オーバービュー



単純化されたストレージ管理

既存のストレージシステムとネイティブのストレージサービスを利用することによって、コストと複雑さを削減

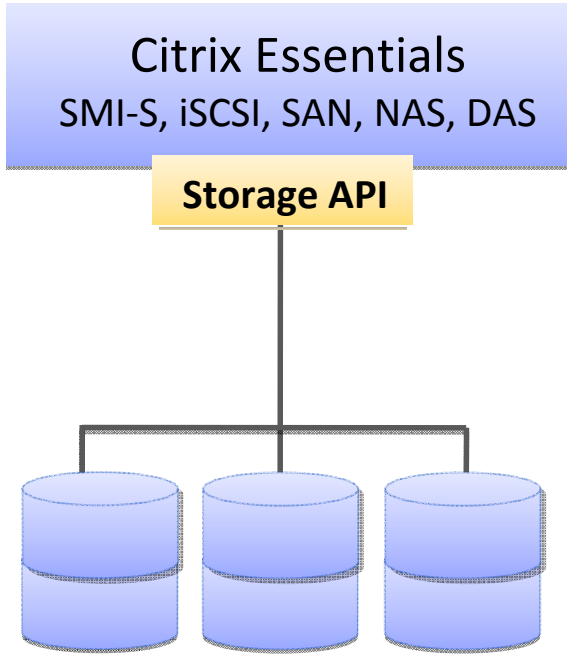
どんなストレージとも機能する(シンプル)

様々なストレージハードウェアプラットフォームでシームレスな互換性を提供すために全てのストレージアーキテクチャーと共に動作する

ネイティブのストレージサービスにワンクリックでアクセス

管理を単純化し既存のストレージアレイベースのサービスとテクノロジーを利用

既存のWindows ストレージ管理のプロダクトと動作する



エンタープライズストレージシステムの機能を利用

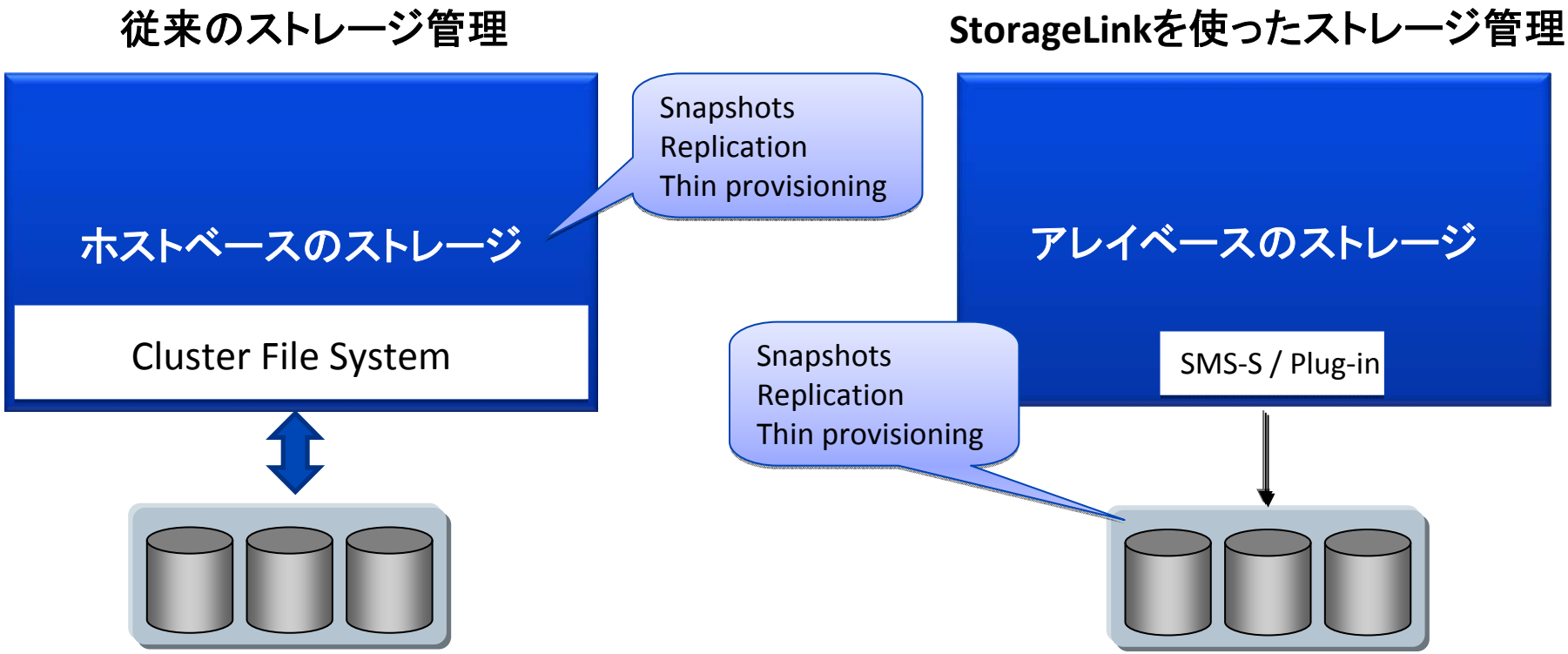
時間の節約 – 強力なストレージ機能へのアクセスに対し複雑さの排除を可能にする

SMI-S storage: EMC Clariion, EMC Symmetrix, HP EVA, IBM ESS DS6000/DS8000, LSI / IBM DS4000, Fujitsu

Citrix StorageLink

- StorageLink, XenServer, Hyper-Vからストレージを管理
  - 仮想ディスクのアレイベースのスナップショット
    - 管理者が異なる時間の仮想ディスクのスナップショットを取得できる
    - アレイベースのためスナップショットによるI/Oパフォーマンスの影響はない
  - 仮想ディスクのプロビジョニングと高速クローンが分から秒に
    - 仮想ディスクのクローンを作成はアレイによって高速におこなわれる
    - アレイはクローンを作成するための機能を持っており、StorageLinkはその機能を使用するため、サーバのCPUに負荷をかけない
  - シンプロビジョニングによるディスクスペースの有効利用
    - 仮想ディスク容量のオーバーアロケート防ぎ、必要な分だけの容量を確保
  - データの重複排除
    - ONTAPによって仮想ディスク間の同一データ排除機能(A-SIS 重複排除機能)によりデータ容量を70%以上削減

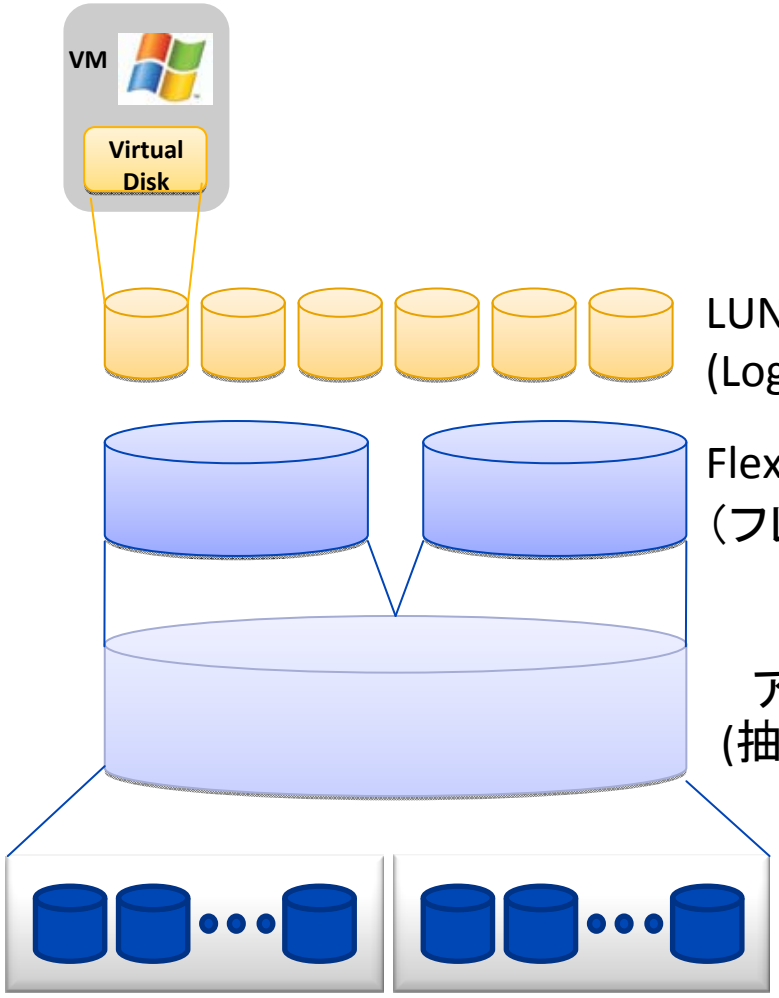
StorageLinkの考え方



- ホストCPUとストレージバスを使用
- インテリジェントストレージは単なるJBOD
- 仮想化ソフトウェアに最適化

- ストレージCPUを使用
- ストレージが持っている本来の機能を使用
- ストレージシステムに最適化

StorageLink機能(NetAppの場合)



仮想マシン作成時に自動的にLUNが作成される

ストレージ接続後自動的に作成される

XenServerからiSCSI/SANストレージとして接続

RAID-DPとアグリゲートはストレージ側で設定

XenServer ストレージ リポジトリー

SR タイプ	説明	共有	スパース (Thin provisioning)	仮想ディスク リサイズ	高速 クローン	スナップ ショット
LVM	ローカルディスク上の LVM			○		
EXT3	ローカルディスク上の VHD		○		○	○
LVHD (5.5)	LVM上のVHD		○	○	○	○
NetApp	NetApp上のLUN	○	○	○	○	○
Dell EqualLogic	Dell EqualLogic上のLUN	○	○	○	○	○
SMI-S Storage (5.5)	StorageLinkによるLUN	○	○	○	○	○
LVMoFC	FC LUN上のLVHD	○	○	○	○	○
LVMoiSCSI	iSCSI上のLVHD	○	○	○	○	○
NFS	Network ファイルシステム上のVHD	○	○		○	○

CITRIX®

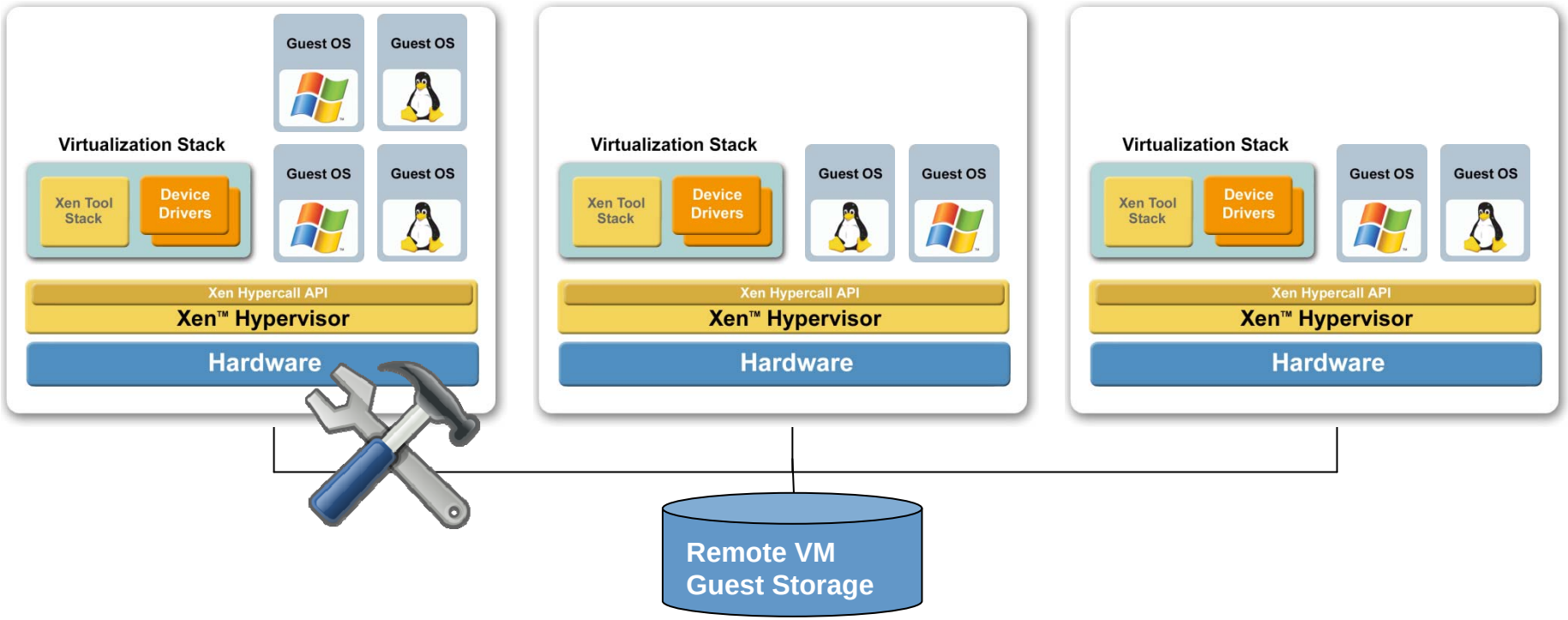
仮想環境の高度な機能

ライブマイグレーション

XenMotionはサービスのダウンタイムなしで仮想マシンを移動させることができる

計画停止をゼロダウンタイムで

異なるサーバ間でのロードバランシング

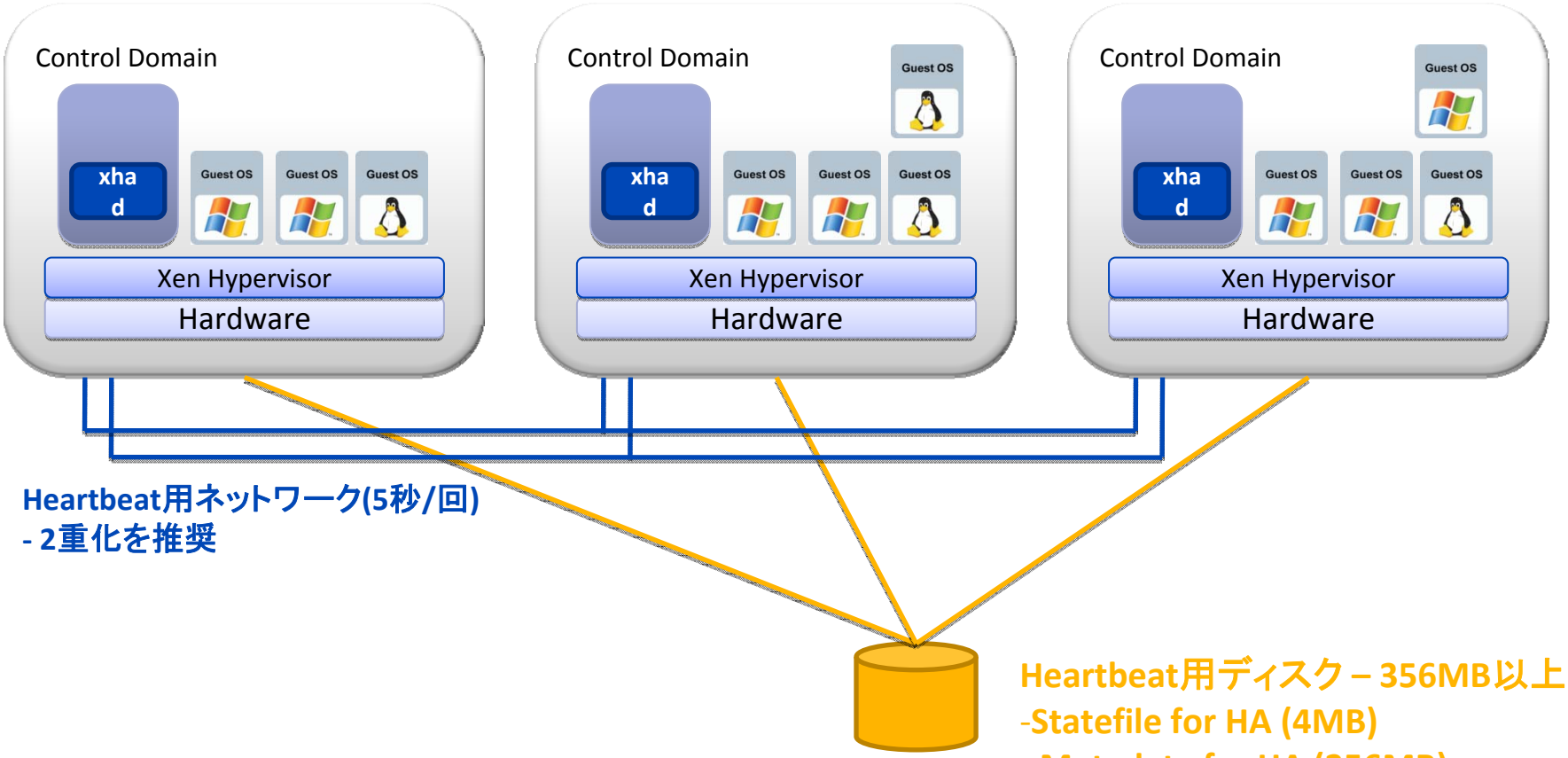


XenMotionのロジック

1. VM稼働中のXenServerが移動先XenServerに対してXenMotionのリクエストを送信
2. 移動先XenServerはそのリクエストに対して十分なリソースがあるか確認を行う(リザベーションステージ)
3. 移動先XenServerに十分なリソースがあれば、すべてのメモリページをTCPソケットを使用してコピーを行う(反復事前コピー)
4. コピー中にいくつかのメモリページは更新されてダーティーページとなったページは再度コピーを行う
5. ダーティページが数ページになったところで、稼働中VMを停止させ、残りのページのコピーを行う(停止とコピー)
6. このステージが終了した時点で移動先XenServerのVMを稼働する

- 仮想マシン停止時間は、わずか150ms ～ 2000msのためアプリケーションへの影響はない
- クライアントとのネットワーク遮断はスイッチのMACアドレステーブル再作成による遅延のみ

XenServer HA

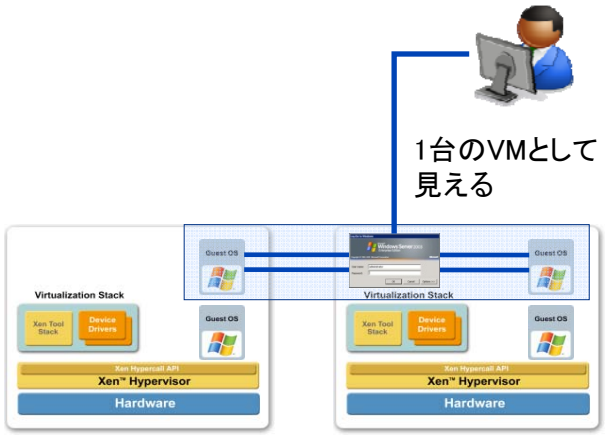


リスタートの保護レベル

- XenCenter
  - Protect (保護する)
    - 選択した仮想マシンの高可用性による保護を保証する
  - Restart if Possible
    - 仮想マシンの自動的な再起動が不可欠ではない場合に設定
  - Do Not Restart
    - 仮想マシンを自動的に再起動しない
- コマンド (xe vm-param-set uuid=<*vm_uuid*> *ha-restart-priority=<1> ha-always-run=true)*
  - 1 | 2 | 3
    プールがオーバーコミット状態になると、高可用性機能により再起動優先度の高いVMから再起動が試行される(1が最高優先度)
  - best-effort
    この優先度を設定したVMは、システムがVMの再起動を試行した場合のみ再起動する
  - ha-always-run=false
    このパラメータを設定したVMは再起動しない
- アジャイル
  - アジャイルでない仮想マシンに設定できる保護レベルは[可能なら再起動またはbest-effort]のみである。アジャイルでない仮想マシンとは、ローカルストレージを持つ、ローカルDVDドライブへの接続が設定されている、またはローカル仮想ネットワークインターフェイスを持つ仮想マシンのこと。

フォルトトレラント (FT) システム

- 複数台の物理サーバで動く2台の仮想マシンが同期を取って実行
  - ロックステップ方式
    - 仕組みが複雑: everRun VM (Marathon Technology)
  - チェックポイント方式
    - デバイスへの出力が遅延: kemari (Xen)
- 完全な冗長化による連続可用性
  - ゼロダウンタイム
- 特殊なハードウェアが不要
  - 顧客ニーズに合ったハードウエアを選択可能



CITRIX®

運用

容易な利用

- XenServer/XenCenterは誰でも容易に使用可能
  - 仮想化運用に必要な機能だけを厳選してXenCenter GUIに追加
  - より高度な機能はCUIで提供
    - 容易なスクリプトの作成
    - Linuxエキスパートはよりコマンドを好む
- XenServerトレーニングコースはわずか2日間
  - 通常共有ディスクを使用しない構成であればトレーニングコースは不要
- 日本語管理ツールXenCenter
  - XenCenterから直接XenServerを管理（管理サーバは不要）

Essentials for XenServer Edition

機能	XenServer	Essentials for XenServer, Enterprise Edition	Essentials for XenServer, Platinum Edition
ネイティブ 64-bit Xen ハイパーバイザー	✔	✔	✔
Windows とLinux ゲスト	✔	✔	✔
無制限の仮想マシン、メモリ、CPU	✔	✔	✔
Active Directory との統合(5.5)	✔	✔	✔
Consolidated backup (5.5)	✔	✔	✔
XenMotion ライブマイグレーション	✔	✔	✔
XenCenter 管理コンソール	✔	✔	✔
複数サーバーの管理	✔	✔	✔
XenCenter サーチ, パフォーマンスの保存とモニタリング	✔	✔	✔
簡素化されたストーレジ管理 (Storage Link)		✔	✔
ハイ アベイラビリティ (HA)		✔	✔
ワークロード バランシング (5.5)		✔	✔
Workflow Studio オーケストレーション *1		✔	✔
Provisioning services (仮想マシンのみ)		✔	✔
Provisioning services (物理マシンと仮想マシン)			✔
ライフサイクル管理(ラボとステージング) *1			✔
	無償	¥467,500	

*1 日本語環境未サポート