サーバの仮想化とは？

2008年1月30日
富士通株式会社
サーバシステム事業本部
松本　一志

Copyright 2007 FUJITSU LIMITED

目次

- 仮想化とは？
- メインフレームの仮想化技術
- Xenとは？
- ストレージ仮想化
- 仮想化環境の統合管理
- まとめ
- 添付資料

Copyright 2007 FUJITSU LIMITED

仮想化とは？

Copyright 2007 FUJITSU LIMITED

仮想化とは？

- 一つの物理資源を分割して、個別のプログラムに独立した資源として見せて運用すること

または

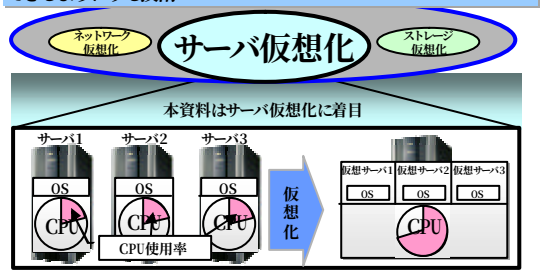
- 複数の物理資源をまとめて管理して一つの資源に見せて運用すること

この二つが基本であるが、資源の基本的な機能を抽象化した論理的な資源を利用できるようにして運用することもある。

Copyright 2007 FUJITSU LIMITED

仮想化技術とは？

物理的な性質や境界を覆い隠し、論理的な資源とし、自由に分割できるようにする技術

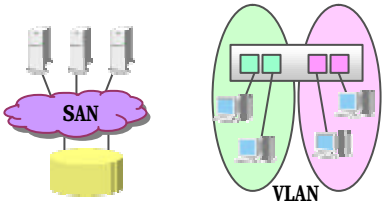


Copyright 2007 FUJITSU LIMITED

外からみた仮想化の種類

- ストレージの仮想化 ・・・・SAN
- ネットワークの仮想化 ・・・・VLAN
- サーバの仮想化

ここではサーバの仮想化に着目

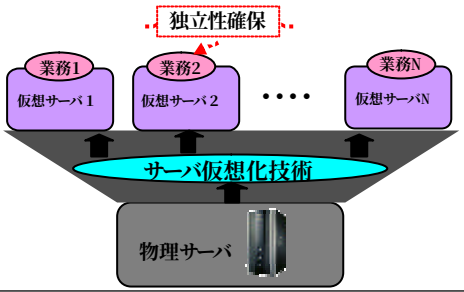


Copyright 2007 FUJITSU LIMITED

Copyright 2007 FUJITSU LIMITED

サーバの仮想化 とは？ ①

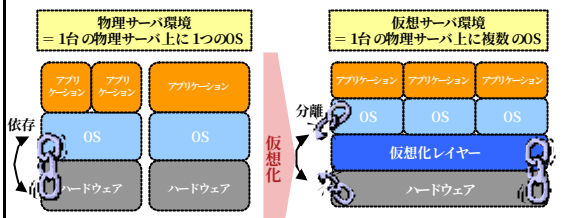
- 物理サーバに複数の仮想サーバを作ること
- 仮想サーバ上では、複数の業務システムが独立して動作



Copyright 2007 FUJITSU LIMITED

サーバの仮想化 とは？ ②

物理ハードウェア上に仮想化レイヤーを作成し、複数OSを同時に実行させる技術

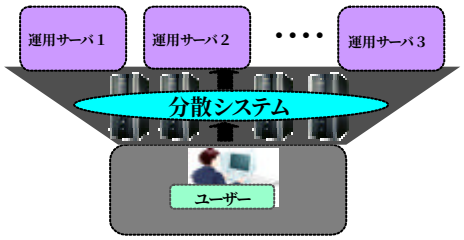


- ハードウェアとOSが密接に関係
- ハードウェア資源の配分は固定（物理環境に依存）
- ハードウェアとOSを分離
- ハードウェア資源の配分を柔軟に変更可能（仮想化レイヤーで制御可能）

Copyright 2007 FUJITSU LIMITED

サーバの仮想化 とは？ ③

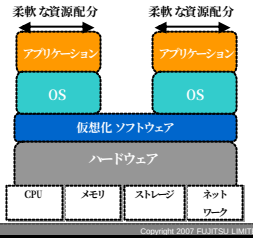
- 複数の物理資源をまとめて管理して一つの資源に見せて運用すること　今回は対象外



Copyright 2007 FUJITSU LIMITED

仮想サーバの特長

- 一つのハードウェアを複数の仮想サーバで共有
  - CPU、メモリ、ストレージ、ネットワークなどを共有
- 柔軟な資源配分
  - 仮想サーバに対するハード資源を柔軟に配分変更
- 仮想マシンの複製が容易
  - 同じ環境のサーバを簡単に作ったり、OSイメージをコピーして利用する



Copyright 2007 FUJITSU LIMITED

これからのデータセンタ運用要件

- サービスレベルの保証
  - ✓サービス継続性
  - ✓新しいサービスの迅速な立上げ
  - ✓負荷変動など環境変化への柔軟な追随
  - ✓セキュリティ
- IT投資のROI向上
  - ✓資源の有効利用
  - ✓資産保証
  - ✓最新テクノロジー導入によるコストパフォーマンス向上
  - ✓運用管理負担の軽減

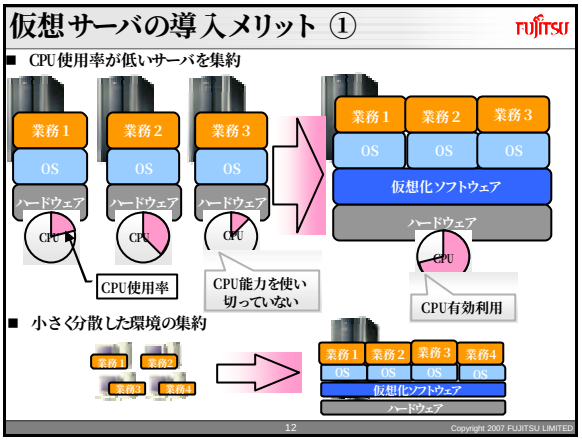
Copyright 2007 FUJITSU LIMITED

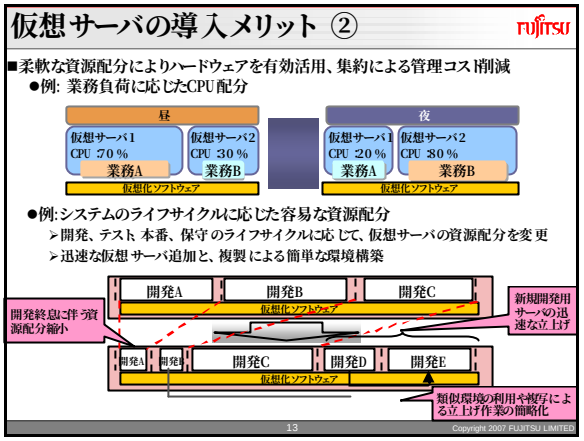
仮想化の利用目的

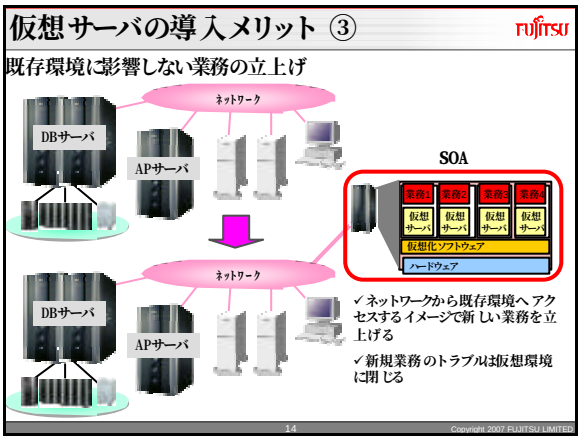
- サーバの集約
  - ✓ CPU使用率が低いサーバを集約
  - ✓ 小さく分散した環境の集約
- 負荷やセキュリティに応じた柔軟な構成で迅速に立上げ
  - ✓ 柔軟な資源配分によりハードウェアを有効活用、集約によるコスト削減
    - 業務負荷に応じたCPU配分
    - システムのライフサイクルに応じた資源配分
- 既存環境に影響しない業務立上げ
  - ✓ ネットワークから既存環境へアクセスするイメージで新しい業務を立ち上げる
  - ✓ 新しい業務トラブルは新規に立ち上げた仮想環境に閉じる

Copyright 2007 FUJITSU LIMITED

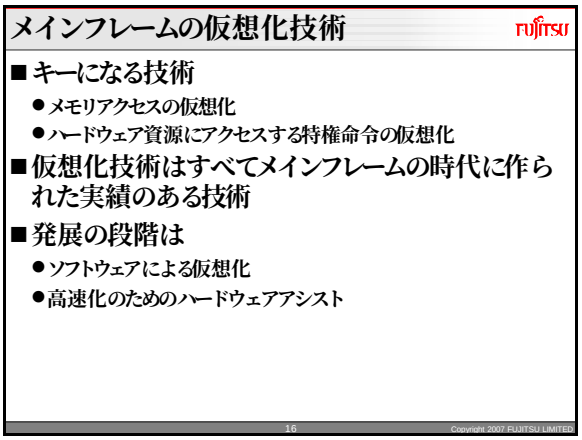
Copyright 2007 FUJITSU LIMITED

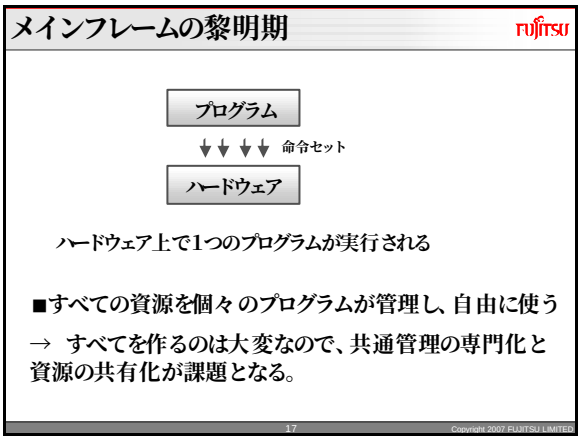

仮想サーバの導入メリット ①
■ CPU使用率が低いサーバを集約
業務1
業務2
業務3
OS
ハードウェア
CPU
CPU使用率
CPU能力を使い切っていない
仮想化ソフトウェア
CPU有効利用
■ 小さく分散した環境の集約
業務4
12
Copyright 2007 FUJITSU LIMITED


仮想サーバの導入メリット ②
■柔軟な資源配分によりハードウェアを有効活用、集約による管理コスト削減
●例: 業務負荷に応じたCPU配分
昼
夜
仮想サーバ1
CPU 70％
CPU 20％
仮想サーバ2
CPU 30％
CPU 80％
業務A
業務B
仮想化ソフトウェア
●例:システムのライフサイクルに応じた容易な資源配分
➢開発、テスト 本番、保守のライフサイクルに応じて、仮想サーバの資源配分を変更
➢迅速な仮想サーバ追加と、複製による簡単な環境構築
開発A
開発B
開発C
開発D
開発E
開発終息に伴う資源配分縮小
新規開発用サーバの迅速な立上げ
類似環境の利用や複写による立上げ作業の簡略化
13
Copyright 2007 FUJITSU LIMITED


仮想サーバの導入メリット ③
既存環境に影響しない業務の立上げ
ネットワーク
DBサーバ
APサーバ
SOA
業務1
業務2
業務3
業務4
仮想サーバ
仮想化ソフトウェア
ハードウェア
✓ネットワークから既存環境へアクセスするイメージで新しい業務を立上げる
✓新規業務のトラブルは仮想環境に閉じる
14
Copyright 2007 FUJITSU LIMITED

メインフレームの仮想化技術
15
Copyright 2007 FUJITSU LIMITED


メインフレームの仮想化技術
■ キーになる技術
●メモリアクセスの仮想化
●ハードウェア資源にアクセスする特権命令の仮想化
■仮想化技術はすべてメインフレームの時代に作られた実績のある技術
■発展の段階は
●ソフトウェアによる仮想化
●高速化のためのハードウェアアシスト
16
Copyright 2007 FUJITSU LIMITED


メインフレームの黎明期
プログラム
命令セット
ハードウェア
ハードウェア上で1つのプログラムが実行される
■すべての資源を個々のプログラムが管理し、自由に使う
→ すべてを作るのは大変なので、共通管理の専門化と資源の共有化が課題となる。
17
Copyright 2007 FUJITSU LIMITED

マルチプログラミング機能を有するOSの登場

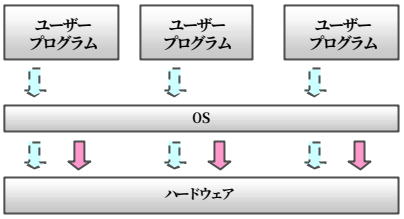
マルチプログラミング機能を有するOS(システムプログラム)の登場により、複数のプログラムが同時走行可能になる。

各プロセスが独立にメモリを要求する ⇒ 仮想記憶方式の実現

I/O処理などのシステム処理 ⇒ OSが専門に実施するための特権化

Copyright 2007 FUJITSU LIMITED

メインフレームのOS特権モード



非特権命令
四則演算、比較、分岐などを行う命令

特権命令(OS専用)
記憶装置の特定領域や入出力装置などの計算機資源の直接使用を要求する命令

Copyright 2007 FUJITSU LIMITED

仮想メモリへのアクセスに対するハードアシスト

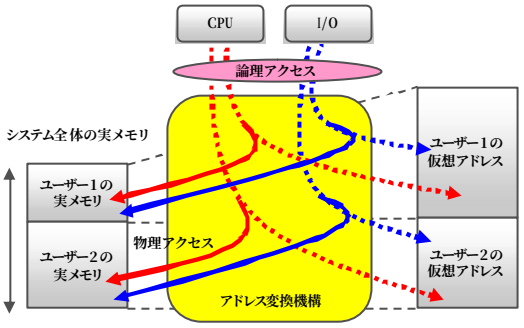
■仮想アドレスへのアクセス

- CPU :DAT*機構
- チャネル :実アドレス
  - ✓I/O発行時には、実アドレスを使う必要がある。(OSの仕事)
  - ✓VMを使うとさらにプレフィックス変換もいる!
- チャネル :仮想アドレス
  - ✓I/O発行を含めて仮想アドレスを使える。
  - ✓CPUとチャネルで同じDATテーブルを参照する方式で実現
    → 現在のDMAの基礎になる

* DAT: Dynamic Address Translation

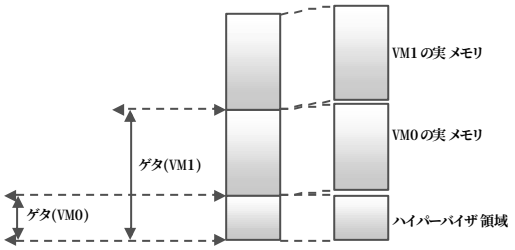
Copyright 2007 FUJITSU LIMITED

メインフレームにおける仮想化時のアドレス変換 ①



Copyright 2007 FUJITSU LIMITED

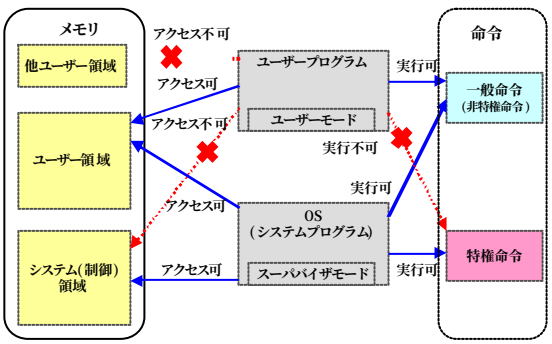
メインフレームにおける仮想化時のアドレス変換 ②



ゲストごとにゲタはかせ(アドレス加算)してメモリアクセス(プレフィックス変換)

Copyright 2007 FUJITSU LIMITED

メインフレームの命令の実行とメモリアクセス

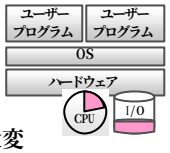


Copyright 2007 FUJITSU LIMITED

Copyright 2007 FUJITSU LIMITED

初期のメインフレームのVMの目的

■ ハードウェアの高性能化に伴い、資源が余剰するシステムが出現 ⇒ 有効活用したい



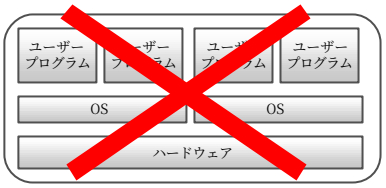
■ 導入前テストで複数の実行環境が必要

⇒ それぞれハードウェアを用意するのは大変

⇒ VMで資源を分割/共有

Copyright 2007 FUJITSU LIMITED

一つのハードウェアで複数のOSを動かすと…



実環境のOS特権モードでできることをOSにやらせると他のOSの動作と矛盾する

⇒ OSとアプリケーションの関係のように、OS間の処理を矛盾なく実行させるハイパーバイザを作る

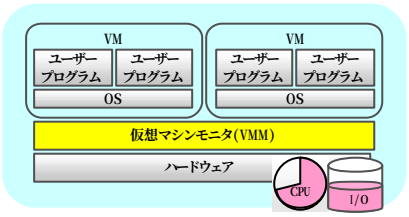
Copyright 2007 FUJITSU LIMITED

仮想マシンシステム(VMS)の出現

高い計算機使用効率を達成できるしかけの実現が望まれる

⇩

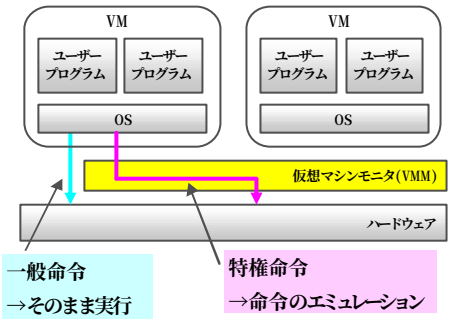
仮想マシンシステム(VMS)の出現



Copyright 2007 FUJITSU LIMITED

仮想マシンシステム(VMS)での命令の実行

システムの中の一部のユーザーとしてOSを動かす



Copyright 2007 FUJITSU LIMITED

初期のハイパーバイザ

- 特権モードの処理(命令のエミュレーション)
- I/O
- ページング
- VM管理

} OSが行うサービスの代行

✓ゲストへの仮想資源(resource)の割当管理

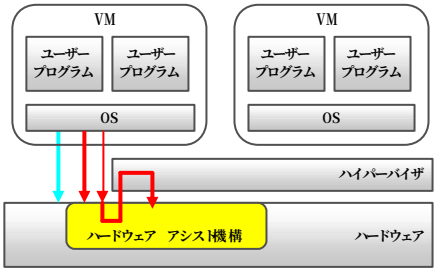
ハイパーバイザとゲスト処理の切替は、命令の例外処理(プログラムチェック割込み)で行う

⇩

モードの切替とエミュレーションのオーバヘッドが非常に大きい

Copyright 2007 FUJITSU LIMITED

ハードウェア アシスト機構による改善 ①



Copyright 2007 FUJITSU LIMITED

Copyright 2007 FUJITSU LIMITED

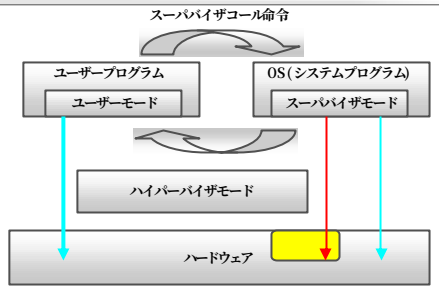
ハードウェア　アシスト機能による改善　②

■ これまでのシステムの動作
- ユーザープログラム :ユーザーモード(一般命令のみ実行可)
- OS(システムプログラム) :スーパバイザモード(特権命令／一般命令実行可)
- ユーザーモード／スーパバイザモードの切替は、スーパバイザコール命令

■ ハイパーバイザモードのサポートによるVM環境
- ユーザープログラム
  - ✓ユーザーモード(一般命令のみ実行可)
- OS(システムプログラム)
  - ✓一般命令はそのまま実行
  - ✓特権命令はそのVMに割り当てられた資源を使って実行
- ハイパーバイザ :すべての命令を実行
  - ✓そのハードウェアで動作するすべてのシステムを管理

Copyright 2007 FUJITSU LIMITED

ハードウェア　アシスト機能による改善　③

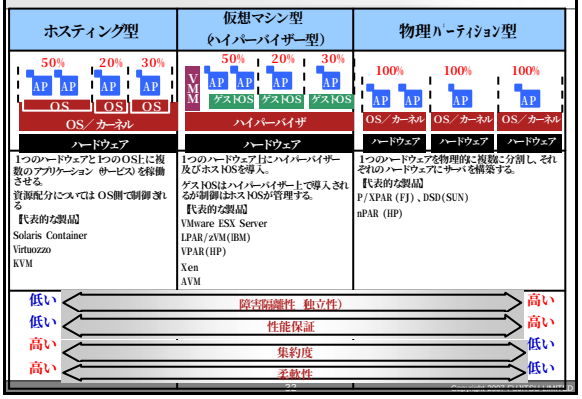


一般命令　→　そのまま実行

特権命令　→　そのVMに割り当てられた資源を使って実行

Copyright 2007 FUJITSU LIMITED

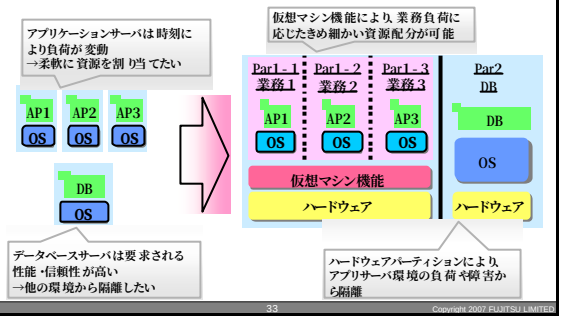
仮想マシンの種類と分類



ハードウェアパーティションとソフトウェアのVM

それぞれの特長を生かして複数の環境を1筐体に搭載

■ 構成例



Copyright 2007 FUJITSU LIMITED

Xenとは？

Copyright 2007 FUJITSU LIMITED

Xenとは

■ 英国Cambridge大学で開発され 初版2003年)，オープンソース化された仮想化ソフト

■ 現在はXenコミュニティーで開発中

■ Xen自身はGNU General Public License

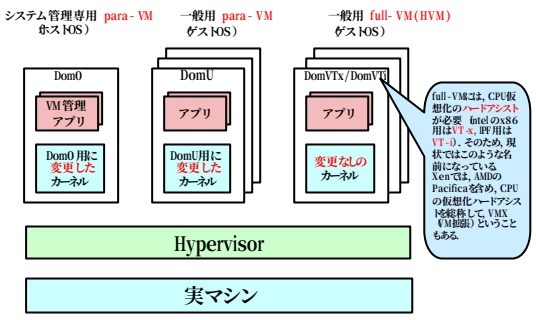
■ Xenという名前は，ギリシャ語由来
- Xen-，xeno- は，「ゲスト」「外来のもの」などの意味を持つギリシャ語由来の英語．VMでゲストOSを動作させることから，Xenという名前になった．

■ Intel，AMDの両CPUメーカー，HP，IBM，富士通などシステムメーカー，Red HatなどのOSベンダーが開発に参加．

Copyright 2007 FUJITSU LIMITED

Copyright 2007 FUJITSU LIMITED

Xenの構成



Copyright 2007 FUJITSU LIMITED

para (準)VMと, full (完全)VM

	カーネル変更	利点 欠点
Para-VM (準VM)	あり	性能がよい カーネル変更版数とHV版数を合わせる必要あり
Full-VM (完全VM) (HVM)	なし	Native OSそのまま利用可能 I/O性能が悪い 専用デバイスドライバの追加で回避可能)

- Windowsカーネルの変更は不可。 ⇔常にFull VMで動かす)
- 既存のRHEL, 既存のSLESのカーネルも事実上, 変更不可。
- カーネルを変更したものが各ディストリビューションから提供されている。

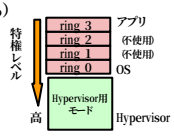
Copyright 2007 FUJITSU LIMITED

VMハードアシスト：CPU仮想化

■ Hypervisorの動作モードと, ゲストの動作モードを区別し, Hypervisorを保護するとともに, Hypervisorがゲストの動作を監視可能とする.

- Intel x86 (IA-32): VT-x
  - ✓VT: Virtualization Technology (旧コード名 Vander Pool)
- Intel IPF: VT-i
- AMD x86: SVM
  - ✓SVM: Secure Virtual Machine (ロード名 Pacifica)

参考 :IPMを使ったブートが定義されている)

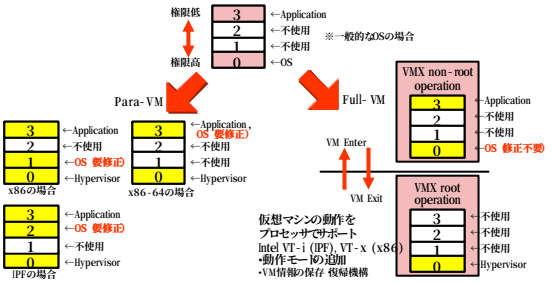


Copyright 2007 FUJITSU LIMITED

CPUの仮想化 CPUの特権レベル

■ Ringによる資源保護

- 特権命令はリング0でのみ実行可能



Copyright 2007 FUJITSU LIMITED

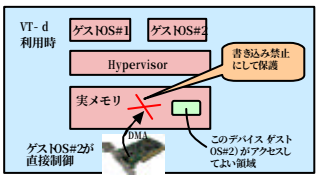
VMハードアシスト：I/O仮想化

■ ゲストOSからデバイスへの直接アクセスを安全に許し, 高性能化

- DMA (Direct Memory Access)は, CPUを介さずにデバイスとメモリの間で直接データ転送を行う機能のこと.
- つまり, デバイスは あるいはDMAコントローラーは), CPUのメモリ保護の対象外であり, 物理メモリを自由に読んだり破壊したりできてしまう問題がある 故意に, バグで, ハード故障で). これを回避.

■ 複数ゲストOSが単一ハード(HBA: Host Bus Adaptor)を安全に共有

- Intel: VT-d
- AMD: IOMMU
- 専用HBA
- PCIの拡張



Copyright 2007 FUJITSU LIMITED

I/O仮想化モデル

現状
- (1) デバイスエミュレーション方式 (full-VMの基本形)
- (2) 中継ドライバ方式 (para-VMの基本形)

必要
- (3) 中継ドライバ方式 (full-VM I/Oの高速化)

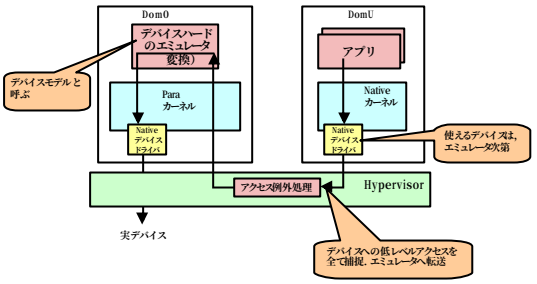
開発中
- (4) 直接I/O方式 (I/Oの最速化)
  - さらに, 複数ドメインからのデバイス直接共有
- (5) ドライバドメイン (ドライバ分離による信頼性向上)

Copyright 2007 FUJITSU LIMITED

Copyright 2007 FUJITSU LIMITED

I/O仮想化モデル(1) デバイスエミュレーション方式

Full-VMの基本形

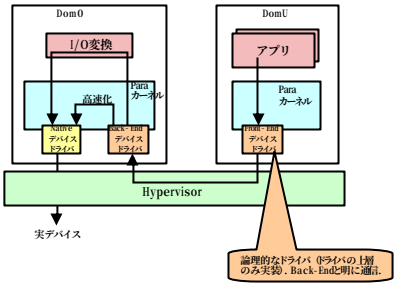


Copyright 2007 FUJITSU LIMITED

I/O仮想化モデル(2) 中継ドライバ方式 (パラドライバ)

Para-VMの基本形

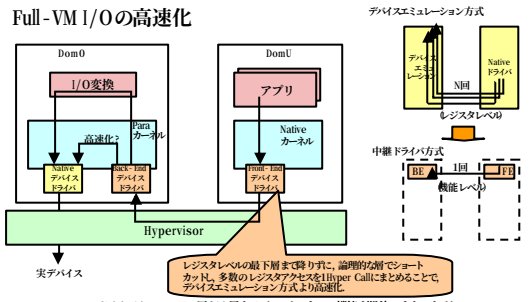
仮想NICとブロックデバイスで実現



Copyright 2007 FUJITSU LIMITED

I/O仮想化モデル(3) 中継ドライバ方式 (パラドライバ)

Full-VM I/Oの高速化



- FEドライバは、para-VM用とは異なるparaカーネルの機能を期待できないため、
- ドライバを各ゲストOSに提供する必要がある

Copyright 2007 FUJITSU LIMITED

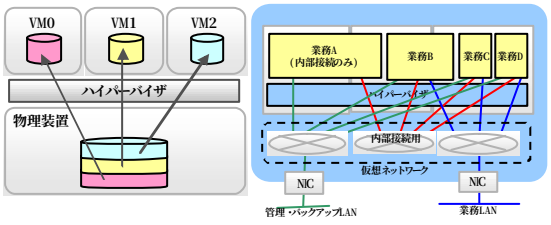
Xenの I/Oの特徴

- 抽象化されたブロックデバイスと仮想NICが中心
- 富士通の試み
  - Xenの特徴に加えて、SCSI機能をゲストから直接使用できるようにする
  - 仮想化ストレージも利用

Copyright 2007 FUJITSU LIMITED

仮想化による I/Oのサポート ①

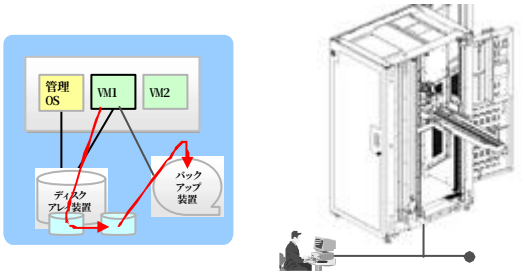
- 仮想的な装置のシミュレート(複数ゲストで物理装置を共有)
  - 仮想ブロックデバイス
  - 仮想NIC



Copyright 2007 FUJITSU LIMITED

仮想化による I/Oのサポート ②

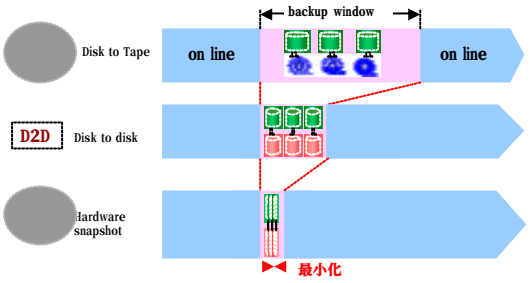
- 実装置をゲスト専用に割当
  - 専有が必要な装置(SCSIオペレーション)
  - 装置管理アプリの処理



Copyright 2007 FUJITSU LIMITED

Copyright 2007 FUJITSU LIMITED

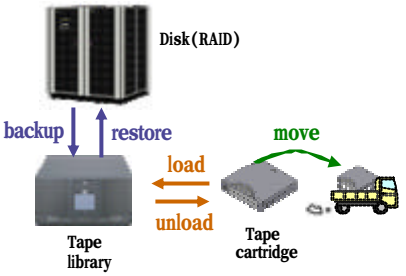
SCSIコマンドの応用例(disk)



Copyright 2007 FUJITSU LIMITED

SCSIコマンドの応用例(tape)

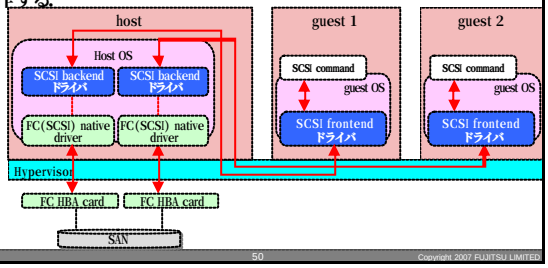
- tape robot :tape cartridgeの移動
- tape drive :load, unload, rewind



Copyright 2007 FUJITSU LIMITED

pvSCSIドライバ

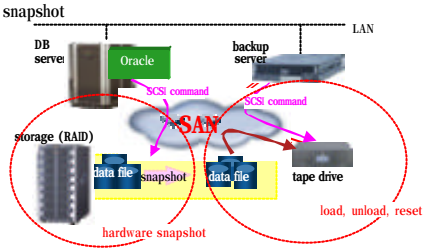
- pvSCSIドライバはSCSI frontendドライバとSCSI backendドライバから構成される.(中継ドライバ方式)
- 各ゲストはホストを介してSCSIコマンドを発行する.
- 各ゲストはそれぞれFC HBA cardを専有する.→アプリがゲスト上で動作する.



Copyright 2007 FUJITSU LIMITED

データセンタ

- データセンタ管理
  - エンタープライズシステムにおいて,データはFC/SCSIデバイスに格納される.
  - 信頼性が要求されている.
- SCSIを使った操作.
  - hardware snapshot
  - tape操作



Copyright 2007 FUJITSU LIMITED

データセンタでのVMに対する要件

- 業務アプリの動作保証
- サーバ集約
  - **全ての資源をVMに集約.** → **pvSCSI driver + NPIV**
- 対ハード故障耐性
- 性能

Copyright 2007 FUJITSU LIMITED

想定されるサーバ運用 - VM(仮想マシン)運用 -

- CPU資源の有効利用(能力向上、コア数増加への対応)
  - ✓1つの物理CPU上に複数のVMを立ち上げ複数業務を実行
- 最新のサーバハードウェア上で旧OS・既存業務を実行
  - ✓業務資産を保証しつつ最新ハードウェアのコストパフォーマンスを享受
- 業務無停止のプロビジョニング
  - ✓VMのマイグレーション機能の活用

等々

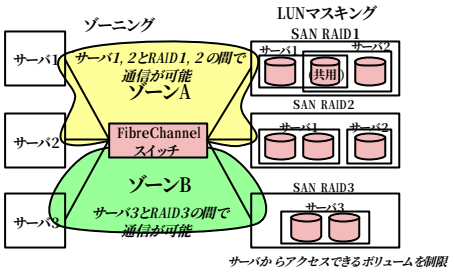
⇒VMを想定した環境でも、サーバ資源とストレージ資源の柔軟な関連付けを実現することがポイント

Copyright 2007 FUJITSU LIMITED

Copyright 2007 FUJITSU LIMITED

SAN上のサーバとストレージの関連付け

- ゾーニング、LUNマスキングなどの方法で、サーバ資源とストレージ資源の関係を規定
  - セキュリティ、サービスレベル保証などを実現

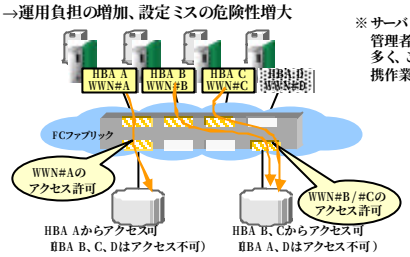


Copyright 2007 FUJITSU LIMITED

サーバとストレージの関連付けの課題

- ゾーンやホストアフィニティは、物理的な識別子（サーバに搭載したHBAのWWN）をキーとして設定
  - サーバハードウェアを交換するとWWNが変わるため再設定が必要
  - 新サーバの追加時もゾーン設定の変更が必要
  - →運用負担の増加、設定ミスの危険性増大

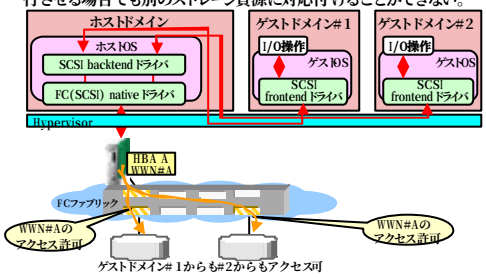
※サーバとストレージの運用管理者が別であることも多く、この場合は両者の連携作業が必要



Copyright 2007 FUJITSU LIMITED

SAN上のVMの見え方

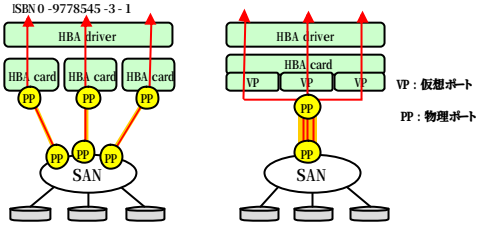
- 複数のゲストが1枚のHBAを共有
  - 現状では、SAN側からは1枚のHBA、すなわち1個のWWNしか見えない。
  - →すべてのゲストが同一ゾーンに属することになり、ゲストごとに別業務を実行させる場合でも別のストレージ資源に対応付けることができない。



Copyright 2007 FUJITSU LIMITED

NPIV（N-Port Identifier Virtualization）

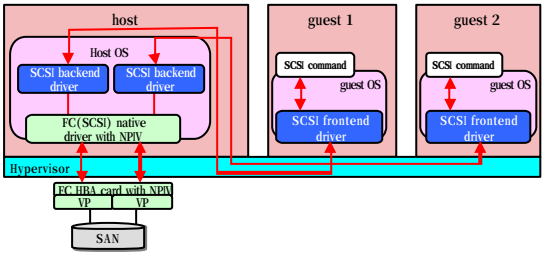
- 仮想ポートは物理ポートと同様に独立にSANに接続できる。
- NPIVはSNIAにより標準化されている。
- 「単一の物理的なファイバチャネルのエンドポイント（N_Port）が、複数の一意に識別できる論理的なエンドポイントをサポートすることができる機能。NPIVをサポートするプロトコルは、ANSI INCITS T11ファイバチャネル標準で規定されている。」
  （出典 SNIA用語集 http://www.snia.org/education/dictionary/n）
  参考資料：Fibre Chanel Association, "Fibre Channel Advances", ISBN 0-9778545-3-1



Copyright 2007 FUJITSU LIMITED

NPIVサポート

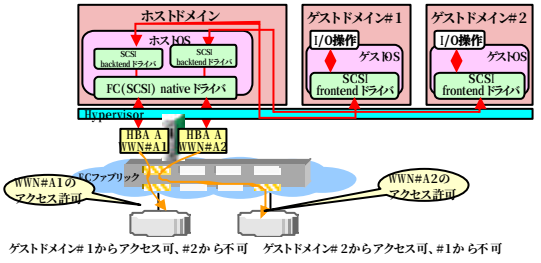
- NPIV： 1つのHBAカードを複数の仮想HBAカードに見せる.
  - ●各ゲストは自分自身の仮想HBAを持つことができる.
  - → 物理HBAカードの数を削減.
  - メインフレームの共用チャネルと同等の効果



Copyright 2007 FUJITSU LIMITED

VMとNPIVの組合せ

- 複数のゲストがそれぞれのWWNを持って1枚のHBAを共有
  - SAN側からはそれぞれのゲストに対応した複数のWWNが見える。
  - → ゲストごとに別業務を実行する場合、それぞれ別のゾーンに設定することが可能となり、業務とストレージ資源の関係をきめ細かく制御できる。



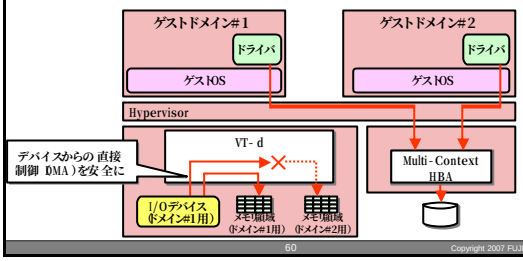
Copyright 2007 FUJITSU LIMITED

Copyright 2007 FUJITSU LIMITED

直接I/O方式（将来構想）

■ ドメインからのI/O制御をハードで管理

- VT-d
  - ✓各ドメイン管理のメモリ領域を分離，保護する機能を持つチップセット
- Multi-Context HBA（Host Bus Adapter）
  - ✓複数ドメインでI/Oデバイスを安全に共有する機構



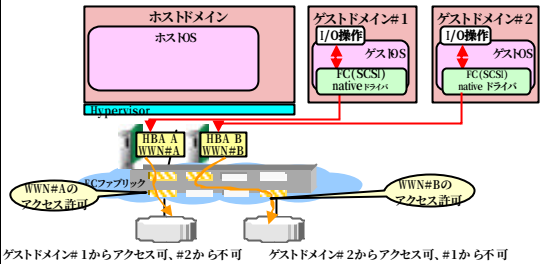
Copyright 2007 FUJITSU LIMITED

直接I/O

■ VMのゲストOSがHBAなどI/Oデバイスを直接制御できる機能

- IntelチップセットのVT-dなどを利用
- 複数のHBAをサーバに搭載し，ゲストに占有させれば，NPIVがなくとも，ゲストに対応した複数のWWNを使用することができる。

→HBA搭載数の制約，HBAの有効利用の点で，本質的な解にはならない。

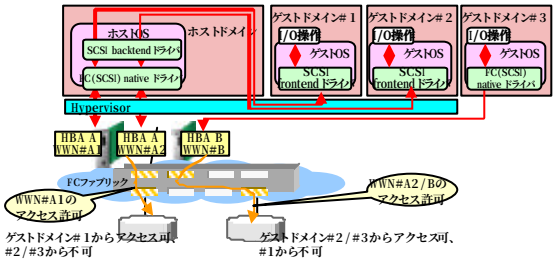


ゲストドメイン#1からアクセス可，#2から不可　ゲストドメイン#2からアクセス可，#1から不可

Copyright 2007 FUJITSU LIMITED

直接I/OとNPIVの組合せ

■ I/O性能，HBAの有効利用，業務とストレージ資源の関係のきめ細かい制御が可能となる。



Copyright 2007 FUJITSU LIMITED

ストレージ仮想化

Copyright 2007 FUJITSU LIMITED

ストレージの仮想化とは

■SNIA（Storage Networking Industry Association）の定義するストレージの仮想化とは？

(http://www.snia-j.org/resource/ronbun/snw041221_8.html)

- ストレージ資源の抽象化によってサーバからの要求に対して，よりオン・デマンドなサービスを提供する。
- ストレージは全て論理ビューとなる(管理対象を単純化)。
  - ✓ ホストからみえなくなるもの
    - 物理パス
    - デバイスの特性
    - データの物理的位置
- 動的な処理
  - ✓ 実行中の再構成を透過的に行うことが可能
  - ✓ データの格納場所の変更をホスト環境に対して透過的に行うことが可能
- ストレージの仮想化は多様なアプローチで実現可能

Copyright 2007 FUJITSU LIMITED

ストレージ仮想化のメリット

■ ストレージ装置の物理的な境界／属性を意識させない。

■ ストレージを一元化して管理（ストレージプール運用）できる。

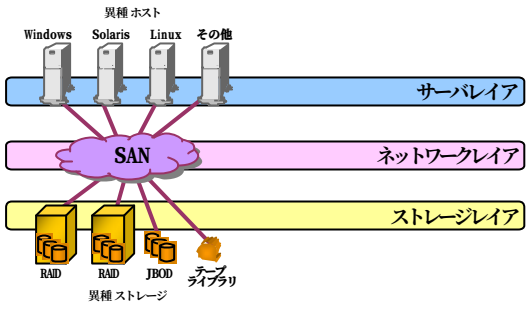
■ 簡単に何時でも追加スペースを提供できる。

■ ビットコスト，運用コストを削減できる。

Copyright 2007 FUJITSU LIMITED

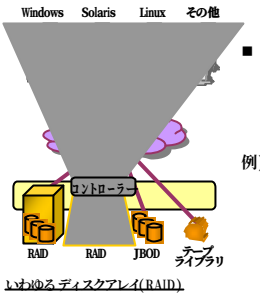
Copyright 2007 FUJITSU LIMITED

ストレージ仮想化の実装位置



Copyright 2007 FUJITSU LIMITED

ストレージレイア



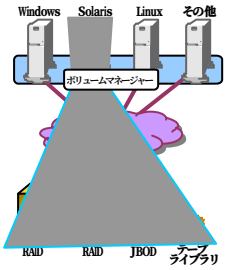
- 利点
  - サーバ種に制限されずに論理ユニットを提供できる。
- 欠点
  - 仮想化の範囲はストレージ装置内部に制限される。
  - 仮想化機能を持つストレージ装置は一般的には高価である。

例)

- Fujitsu ETERNUS
- Hitachi SANRISE
- EMC Symmetrix　他多数

Copyright 2007 FUJITSU LIMITED

サーバレイア



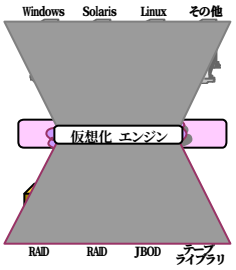
- 利点
  - 仮想化の範囲がストレージ装置に制限されない。
  - ストレージに比べて安価である。
- 欠点
  - 仮想化の範囲がサーバ種に制限される。
  - サーバのリソースに負荷がかかる場合がある。

例)

- Fujitsu SafeDISK、GDS
- VERITAS VxVM

Copyright 2007 FUJITSU LIMITED

ネットワークレイア



- 利点
  - 仮想化の範囲が、サーバやストレージ装置に制限されない。
  - SANのストレージ共有環境に適する。
- 欠点
  - ストレージ資源の管理が複雑になる（資源管理ソフトウェアによって解決可能）。

Copyright 2007 FUJITSU LIMITED

仮想化環境の統合管理

Copyright 2007 FUJITSU LIMITED

仮想化環境における統合管理の重要性

- 仮想化による管理対象数の増加
  - 例. VMゲスト数、ゾーン数
- 仮想化による仮想資源と物理資源の関係管理の必要性
- 仮想化技術による資源間のきめ細かい関係付け
  - 例. 業務対応のストレージ資源、業務対応のゾーン
- 仮想化環境は本質的にマルチベンダー環境
  - 例. サーバ、VM、HBA、スイッチ、ストレージ

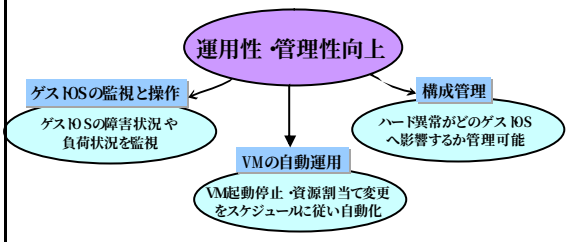
→仮想環境を的確に管理し、その利点を享受するには、ソフトウェアによる統合管理が本質的に重要

Copyright 2007 FUJITSU LIMITED

Copyright 2007 FUJITSU LIMITED

運用管理ソフトとの連携

物理環境と仮想環境の統合管理



Copyright 2007 FUJITSU LIMITED

プロビジョニング

■サーバの負荷増大への対処
- ✓ より能力の高いサーバ上に業務を移動
- ✓ 新規追加サーバ、プールされているサーバの割当て、開発機など別用途のサーバの流用などにより負荷分散

■サーバのハードウェア障害への対処
- ✓ 代替サーバ上で業務を再立上げして継続

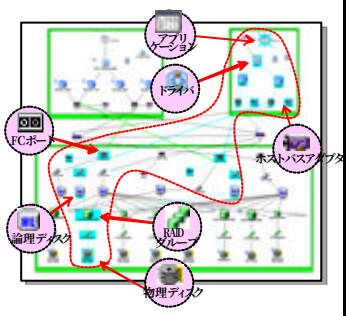
■新規業務の迅速な立上げ
- ✓ サーバ資源やストレージ資源を少ない運用負担で確実に割当て等々

ハードウェアは、ブレードサーバなど柔軟な構成技術が進展
⇒サーバ資源とストレージ資源の柔軟な関連付けがポイント

Copyright 2007 FUJITSU LIMITED

統合管理ソフトウェアの例

- サーバ資源の管理
- ストレージ資源の管理
- ネットワーク資源の管理
- 資源の相互関係付け
- 可視化
- プロビジョニング



Copyright 2007 FUJITSU LIMITED

まとめ

Copyright 2007 FUJITSU LIMITED

まとめ

- キーとなる技術はメインフレーム時代に登場
  - メモリとI/Oの仮想化
  - ハードウェア資源をアクセスする特権命令の仮想化
- 実現方法
  - まずはソフトエミュレーション
  - ハードウェアによるアシスト
- 管理ソフト、高速処理
  - 実ハードを直接操作
- 仮想化環境の統合管理が重要
  - 現在は、ハード、ソフト、サーバ、ストレージをばらばらに管理しているが、連携して統合管理することが課題

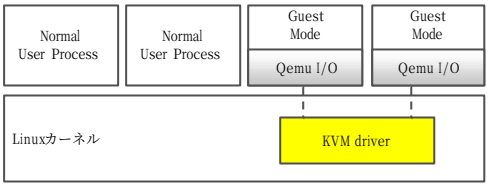
Copyright 2007 FUJITSU LIMITED

添付資料

Copyright 2007 FUJITSU LIMITED

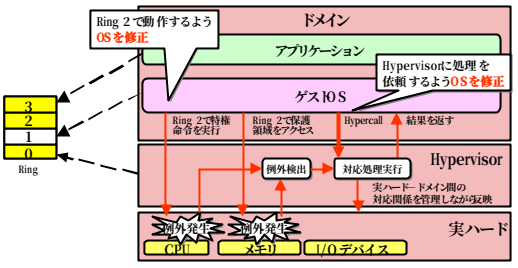
Copyright 2007 FUJITSU LIMITED

Linuxサーバの仮想化 KVM



78 Copyright 2007 FUJITSU LIMITED

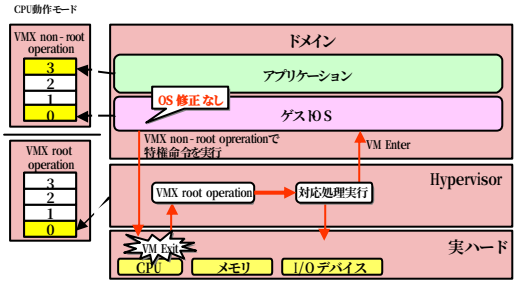
Para-Virtualization (PF)



※Ring 1はHPのvBladesが使用していたため、
Xenではring 2を使用

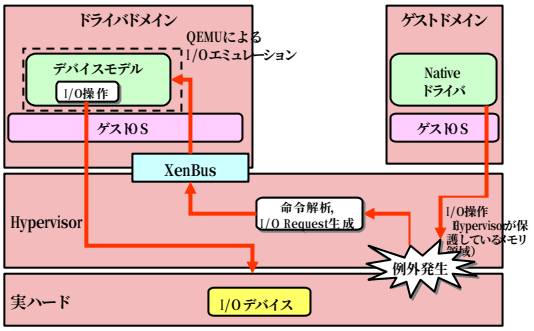
79 Copyright 2007 FUJITSU LIMITED

Full-Virtualization



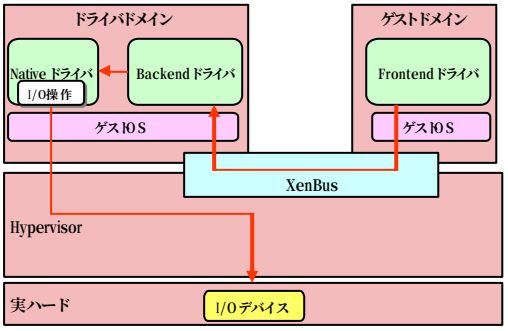
80 Copyright 2007 FUJITSU LIMITED

デバイスエミュレーション方式の動作



81 Copyright 2007 FUJITSU LIMITED

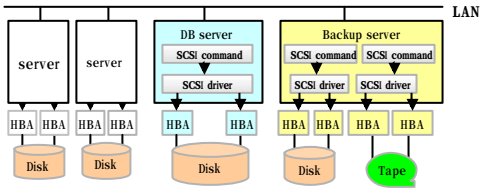
中継ドライバ方式の動作



82 Copyright 2007 FUJITSU LIMITED

データセンター(Enterprise システム)

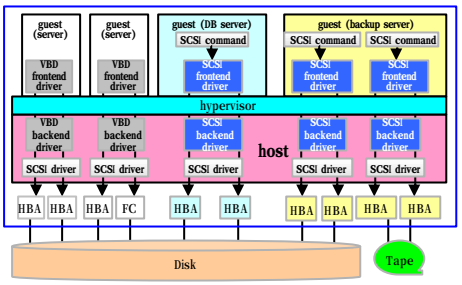
■データセンターには多くのサーバがあり, 各サーバに多数のストレージが接続されている.



83 Copyright 2007 FUJITSU LIMITED

VMによるデータセンター(Enterprise システム)

■ VMによるサーバ集約
→ 多くのHBAカードが必要.



Copyright 2007 FUJITSU LIMITED

スケールアウト型サーバの仮想化 VMware

X86系サーバのサーバ仮想化のメジャー製品。
・VMotionやHAなどの技術により運用性／信頼性の向上を実現する

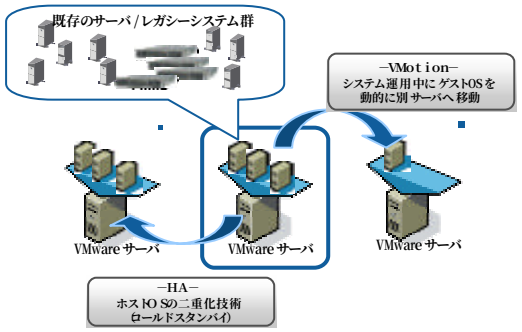
富士通のビジネス実績
- 2006年1月からOEM販売開始、250社を超えるお客様に提供

富士通の取り組み
- 設計・構築・運用の各フェーズで支援体制を構築
- 検証センター(*Platform Solution Center*)を無償で提供
- ハードウェア／ソフトウェアのサポートをワンストップで利用可能
- VMware社認定の高性能、高信頼なPCサーバ「PRIMERGY」とストレージ「ETERNUS」でお客様が安心してVMwareをご利用頂けるシステムを提供

Copyright 2007 FUJITSU LIMITED

VMwareの適用 / 機能

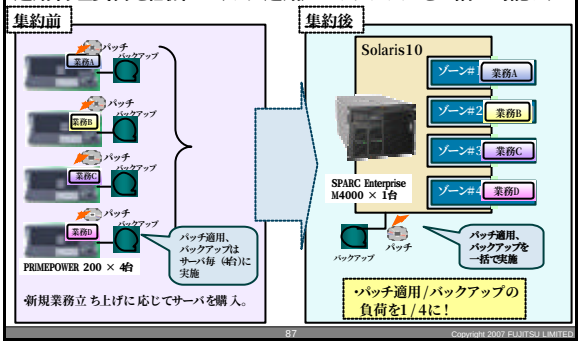
既存サーバを集約し、VMware の各種機能により「運用性の向上」「信頼性の向上」を実現



Copyright 2007 FUJITSU LIMITED

小規模な業務サーバの集約例 (Solarisコンテナ)

Solarisコンテナで小規模な業務サーバを集約し、
運用管理負荷を低減！ (パッチ適用／バックアップを一括で可能に)

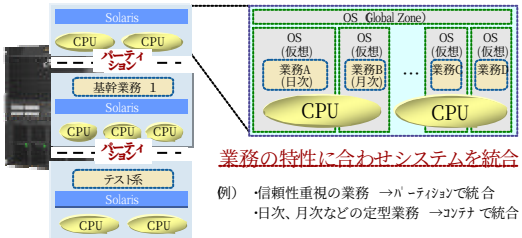


Copyright 2007 FUJITSU LIMITED

業務の特性に応じたサーバ集約

ハードウェアパーティショニング (ミッドレンジ以上)
高信頼(ハードウェア隔離性)

Solarisコンテナ (全モデル)
柔軟性(最大8,191分割)
※マルチコア化にも対応



業務の特性に合わせシステムを統合

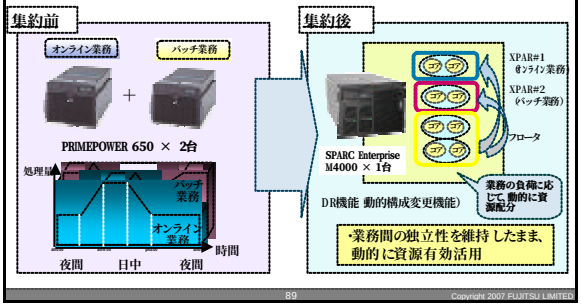
例) ・信頼性重視の業務 →パーティションで統合
・日次、月次などの定型業務 →コンテナで統合

Copyright 2007 FUJITSU LIMITED

中規模な業務サーバの集約例 (ハードウェアパーティション)

・業務ピークの異なるサーバを集約。業務の負荷に応じた資源配分で、サーバ資源の利用率を向上。
・ミッドレンジ価格で、ハイエンド並の高信頼なサーバ統合を実現

ミッドレンジ～

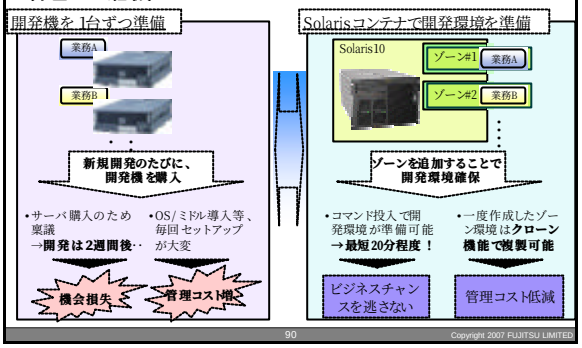


Copyright 2007 FUJITSU LIMITED

Copyright 2007 FUJITSU LIMITED

開発環境の迅速な立ち上げ（Solarisコンテナ）

・ビジネスのスピードアップ
・管理コスト低減



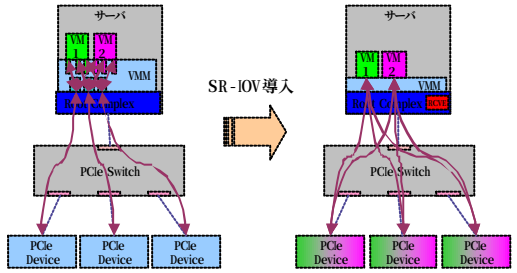
Copyright 2007 FUJITSU LIMITED

PCI-SIG IOV spec

- PCI-SIG IOVの仕様
  - Single Root I/O Virtualization（SR-IOV）→V 1.0（策定済）
  - Multi-Root I/O Virtualization（MR-IOV）→V 0.9（策定中）
  - Address Translation Services（ATS）
- 各機能の概要
  - SR-IOV： 単体サーバ上の各Guest VMからの，①直接I/O，②PCIeデバイス共有を実現
  - MR-IOV： SR-IOVを複数サーバ（ブレードサーバなど）に拡張
  - ATS： SR-IOVの実現に必要な，デバイスからのメモリアクセス時のアドレス変換機能

Copyright 2007 FUJITSU LIMITED

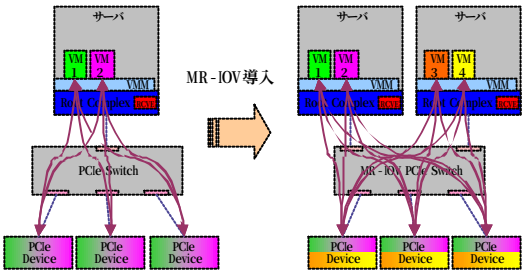
SR-IOV（Single Root I/O Virtualization）



- VMはVMM経由でPCIe Deviceに間接アクセス
- VMはPCIe Deviceに直接アクセス
- 複数VMからPCIe Deviceを共有

Copyright 2007 FUJITSU LIMITED

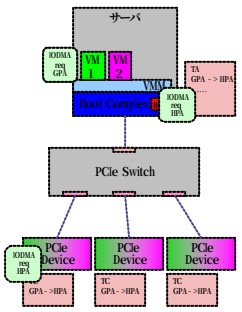
MR-IOV（Multi-Root I/O Virtualization）



- 単体サーバからのみ，①直接I/O ②PCIeデバイス共有が可能
- 複数サーバからも，①直接I/O ②PCIeデバイス共有が可能

Copyright 2007 FUJITSU LIMITED

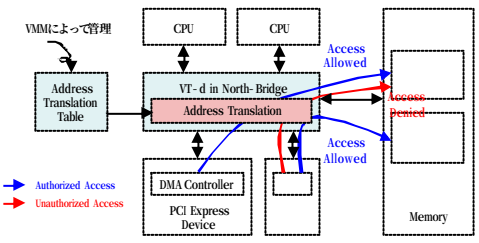
ATS（Address Translation Services）



- DMA-remappingとは？
  - Guest OSはGPA（Guest Physical Address）で動作
  - Guest OSのIODMA制御はGPA
  - DMAがGPAで実施されると物理アドレスに矛盾
  - そこでGPA->HPA（Host Physical Address）変換の仕組みをRoot Complexに入れる（TA）
    - ✓これがVT-dに相当する
- ATSとは？
  - GPA->HPA アドレス変換をEndpoint側でも実施できるようにする
  - Endpoint側にTranslation Cache（TC）を設ける
  - TCの内容によってEndpointがGPA->HPA アドレス変換をする
  - TCの内容は専用プロトコルでRC内のTAに問い合わせる

Copyright 2007 FUJITSU LIMITED

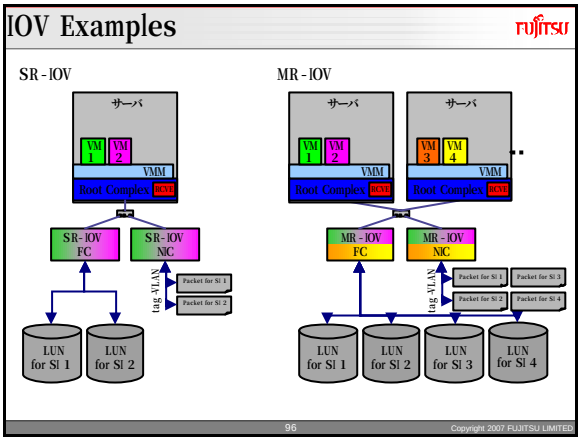
VT-d

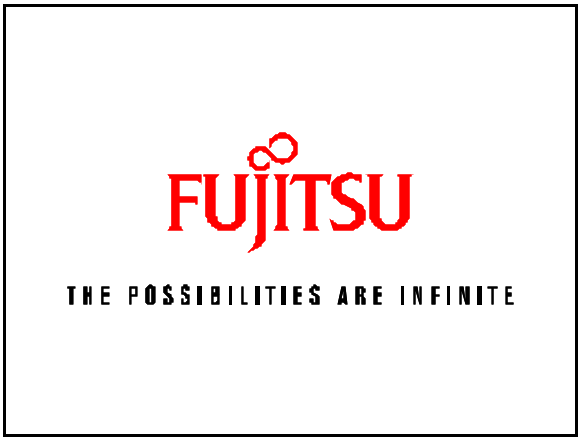


Root Complexに相当するNorth Bridgeにおいて，アドレス変換を実施

Copyright 2007 FUJITSU LIMITED

Copyright 2007 FUJITSU LIMITED


IOV Examples
SR - IOV
MR - IOV
サーバ
VM 1
VM 2
VM 3
VM 4
VMM
Root Complex
RCVT
SR-IOV FC
SR-IOV NIC
MR-IOV FC
MR-IOV NIC
VLAN tag
Packet for SI 1
Packet for SI 2
Packet for SI 3
Packet for SI 4
LUN for SI 1
LUN for SI 2
LUN for SI 3
LUN for SI 4
96
Copyright 2007 FUJITSU LIMITED


FUJITSU
THE POSSIBILITIES ARE INFINITE