
openSUSEで仮想化始めてみませんか？

moonlinux.jp

Kazuhisa Hara



moonlinux®
NEXT CREATOR & COMMUNICATION

Kazuhisa Hara

moonlinux の中の人してます

今日は横浜から参戦

moonlinuxって何だよ…



What is moonlinx ?

moonlinxとは…

デザイン × 音楽 = 新しいクリエイティション

- ・ オンラインカルチャーマガジン & webコミュニティ
 - クリエーター、ミュージシャン向け**プロモーション**サイト
- ・ ウラでは実験的なことも結構やってるよ！

Magazine



HEADLINE

Graphic Music 09/01, 2008 巨大型新人ロックバンドインタビュー

Music 07/31, 2008 HIFANA, AFRAがハイブリッドな演奏

Product 06/13, 2008 松井龍也インタビュー

Visual 05/21, 2008 写真家・森山大蓮の制作と現在を辿る展覧会

Music 04/09, 2008 井上真央インタビュー 招聘する運命を持つ人

Fashion 03/20, 2008 走馬奇なJUN(下) 山本字一インタビュー

Community

moonlinx[®] membership center

Search in the center

TOP IMAGES AUDIOS VIDEOS MEMBERS

Work:

Member:

Popular Image:

1 タンクトップ 2007-08

2 silence 2 days ago

3 eye 169 days ago

Popular Audio:

1 Pianola 59 days ago

2 Trailer 59 days ago

3 so far a way to vmi



- 仮想化について
 - 仮想化の一般的なお話
- Xenについて
 - Xen3.3 Datasheet
- openSUSE + Xen
 - Tips
 - まとめ

Xenの前に...

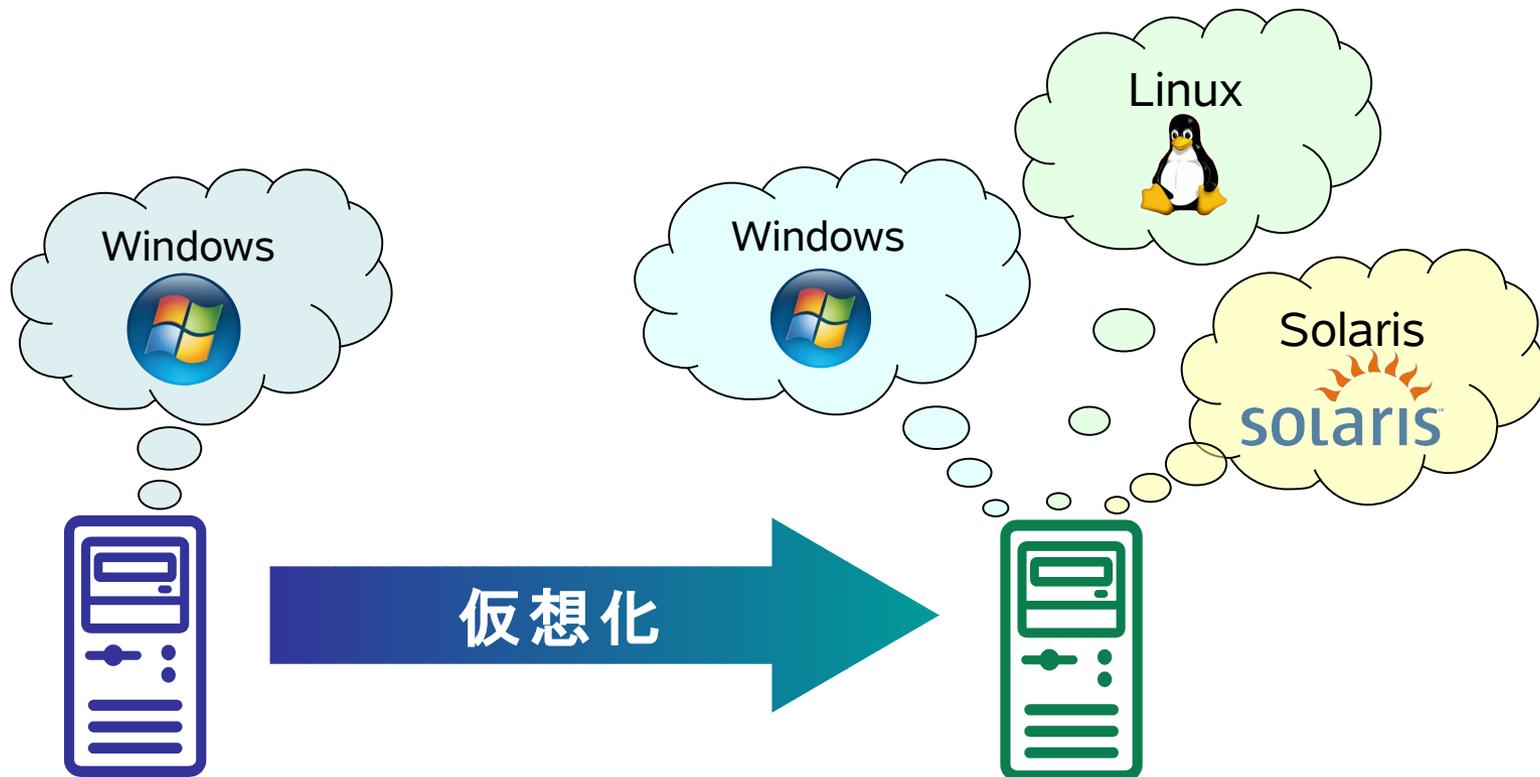
そもそも仮想
化って...



サーバ仮想化とは?

そもそもサーバ仮想化って何なのさ？

- ・ 1台のサーバコンピュータを、あたかも複数台のコンピュータであるかのように論理的に分割し、それぞれに別のOSやアプリケーションソフトを動作させる事



サーバ仮想化技術の利点

－ リソース使用率向上

- ・データセンタ設置機器の平均負荷15%以下と言われている…
 - － 物理的なリソース稼働率の向上

－ 可用性の向上

- ・複数物理サーバ上で、仮想マシンの一時停止、移行、レジュームを実現

－ 応答の迅速化

- ・リソースを動的に調整
- ・サーバの移動と統合

－ 用途別の分離

- ・1物理上で、webサーバ、メールサーバ等を個別に立て、互いに干渉しない状態を作れる
 - － アプリケーションの競合を抑える事にも繋がる

- ・ 2次的に得られる利点

- リスク軽減

- ・ 管理が容易に
 - アップグレード
 - セキュリティパッチのテスト etc...

- コスト削減

- ・ 設置面積の縮小
 - 電源
 - 冷却
 - 空間 etc...

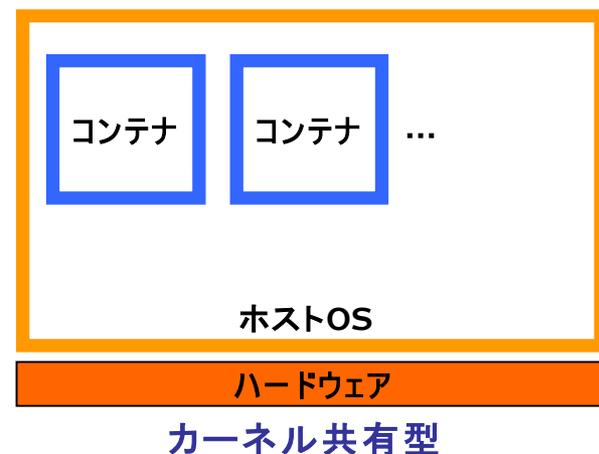
では仮想化を実現するには、どんな選択肢があるの？

商用UNIX機

- à ハイパーバイザー型
 - zVM, LPAR (IBM)
 - HP-UX Virtual Partitions (hp)
- カーネル共有型
 - Solaris Container (Sun)

x86アーキテクチャ

- ハイパーバイザー型
 - VMware ESX Server , Enterprise Server
 - Xen
- ホスト型
 - VMware Server, Virtual Server
 - User Mode Linux
- カーネル共有型
 - FreeBSD Jail



・同じハイパーバイザ型だけど…？

	Xen	VMware ESX Server
ライセンス	OSS(GPL)	商用
動作環境	x86、x64、IA-64アーキテクチャ	x86、x64アーキテクチャサーバ (ESX認定サーバ)
管理OS	Linux、NetBSD	VMware ESX Server専用OS
ゲストOS	Windows、Linux、NetBSD、solaris、Plan9	Windows、Linux、Netware、solaris
最大同時稼働VM	制限なし	仮想プロセッサ合計128個まで
1VMの最大仮想CPU数	32Way	32Way
メモリ管理	仮想マシンごとに割り当て可能(動的リサイズ可能 MAX ∞)	共有、オーバーコミット (最大200%まで MAX 64GB)
高可用性	Live Migration	VMotion

こっから本題の

Xenのお話





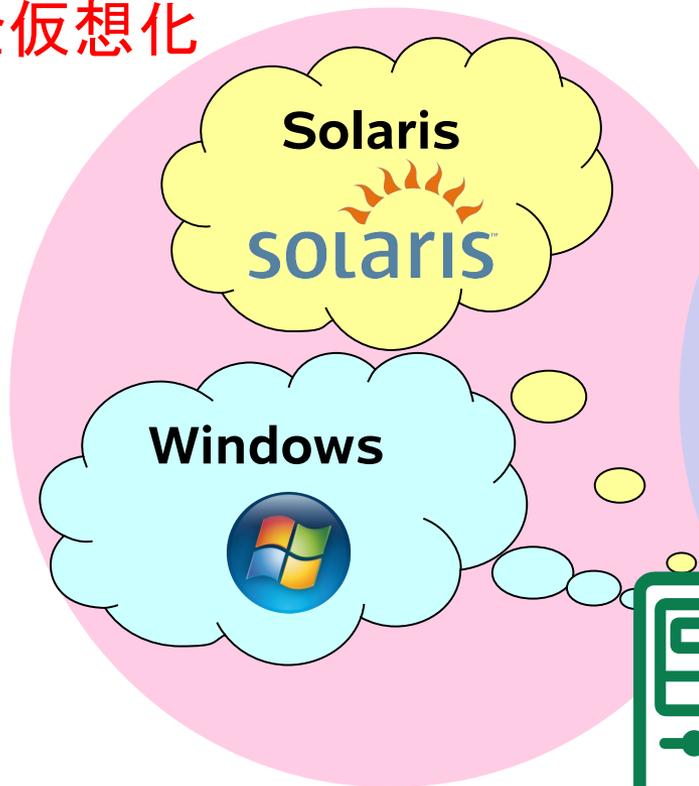
特長

- ・ オープンソース
 - GPLライセンス
 - SUSE、RedHat インストールDVDに標準同梱
- ・ 準仮想化 / 仮想マシンモニタ(ハイパーバイザ)
 - ホスト型の仮想化に比べ、オーバーヘッドが非常に少ない
- ・ 完全仮想化もサポート
 - Windowsも仮想OSとして起動できる
- ・ ライブマイグレーション
 - 高可用性(完全仮想のマシンもマイグレートできる 3.2~)

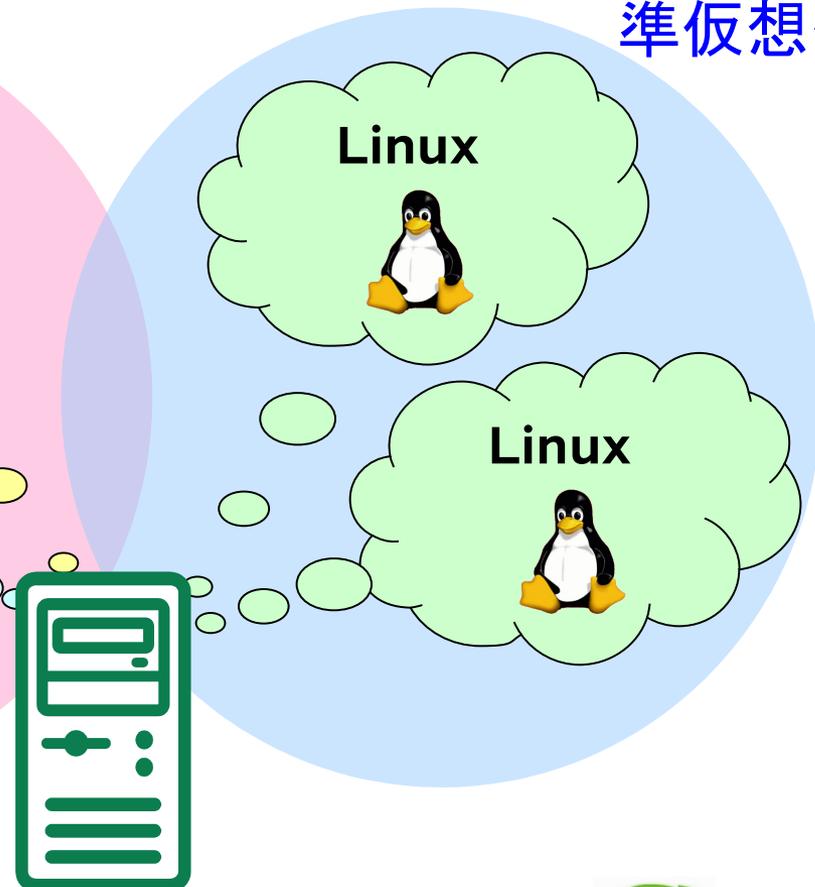
WindowsやSolaris、勿論Linuxも1つのハードで稼働できる

- 完全仮想化・準仮想化の**混合**もOK

完全仮想化

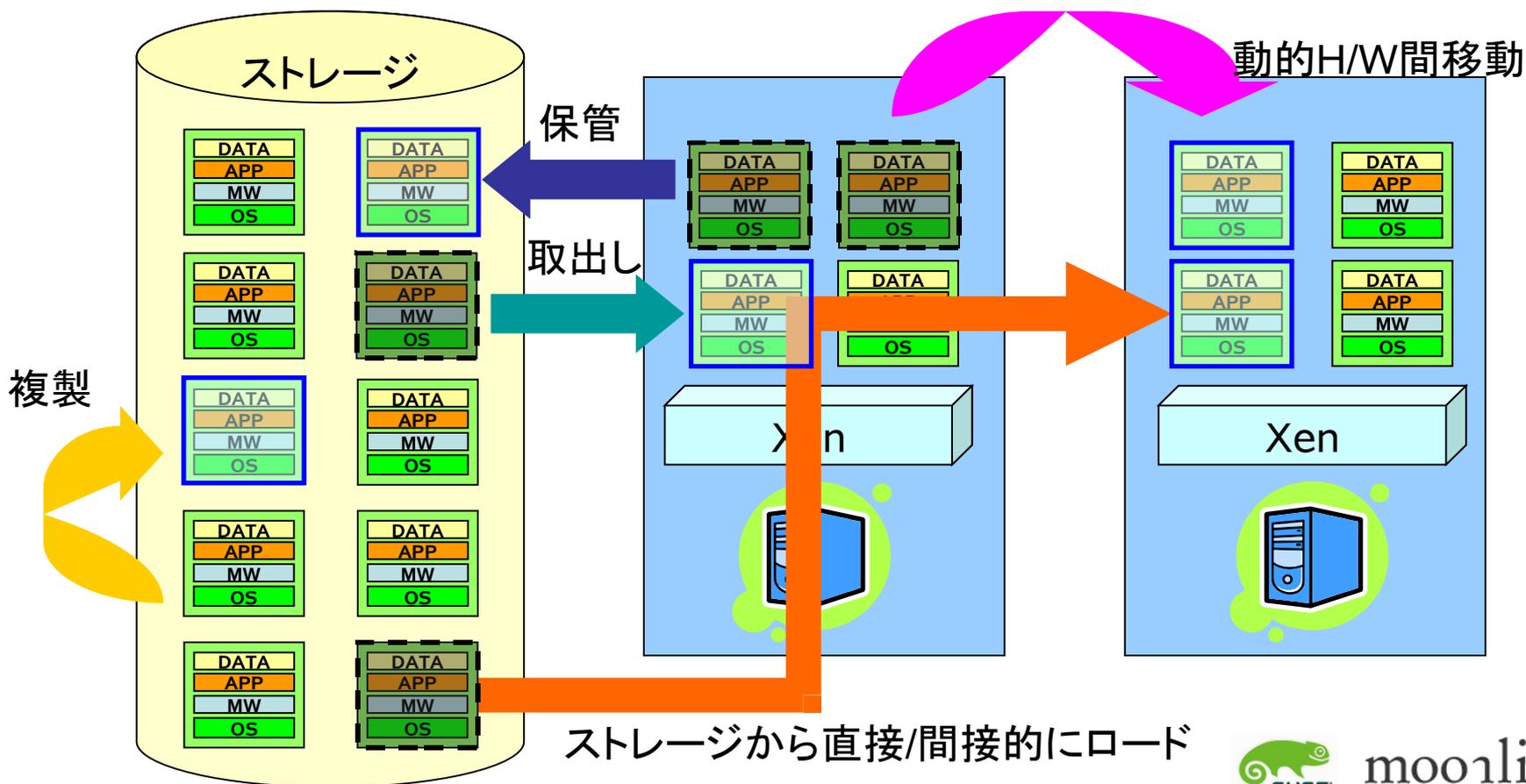


準仮想化



手軽に複製・移動可能

- 仮想マシンはカプセル化されて1ファイルとして管理
- 仮想マシンは完全にH/Wから分離して管理
- カプセル化されたイメージは保管、取り出し、複製、移動等の管理が容易



仮想OSを手軽に複製・移動可能



開発環境を複製して、本番環境にそのままコピーしたり…

逆に本番環境を複製して、カーネルアップデートやセキュリティアップデートのテストをしたりも思いのまま！

さらに…

例えば典型的なLAMPサーバを**テンプレート**として作成し、それを大量に複製して共有ストレージに**プール**しておけば…



・必要に応じて簡単にWebサーバの**デリバリー**なんて事も

- これからは試験環境を作るのに、**毎回ハードを買い足す必要はありません!**
- 基本形さえテンプレートとして持っていれば、用途に合わせてカスタマイズもすぐ行える

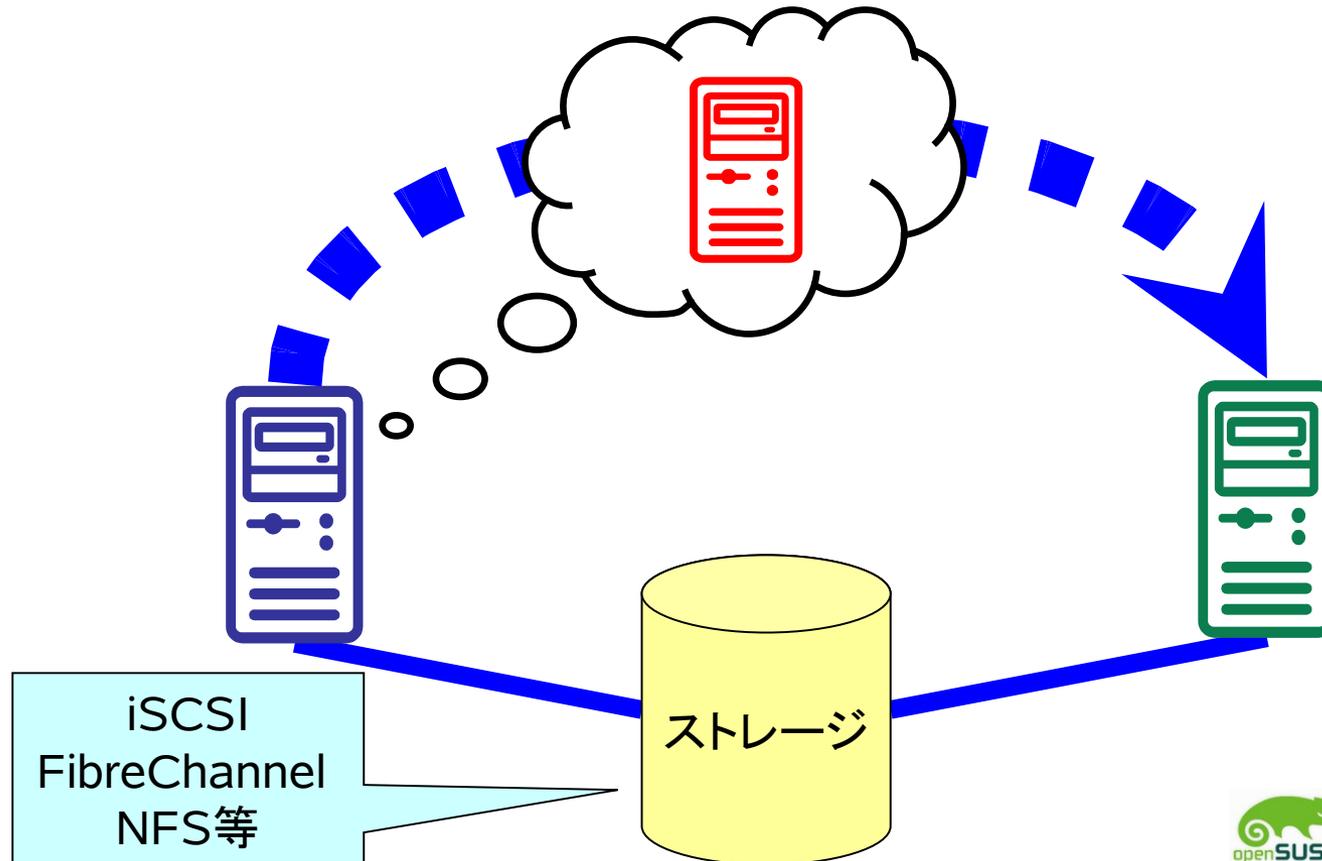
※最近ではXenを用いた仮想Webサーバのレンタルを行う事業者が出てきたようです

ライブマイグレーションとは？

・ Xen上で実行中の仮想マシンを、別のホストに**無停止**で移動させる技術

- 通常のマイグレーションには、停止時間がある

- ・ 厳密には、ライブマイグレーションにもミリ秒単位の停止時間がある
- ・ ただし十分に短いのでネットワークのセッションは維持される
- ・ しかもストレージはiSCSIやNFSでOK



ライブマイグレーションって何に使えるの？ 1

例えば…

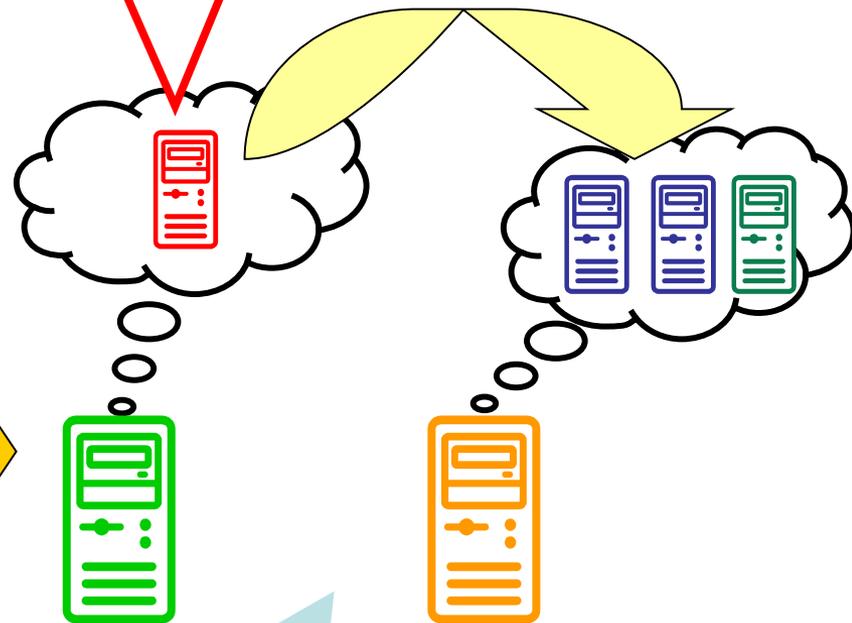
CPU使用率**100%**!
メモリ全然足りません!

CPU使用率**20%**



実メモリ: 2GB
各仮想OSメモリ: 512MBずつ
実CPU: 4コア
各仮想OS配分: 1コアずつ

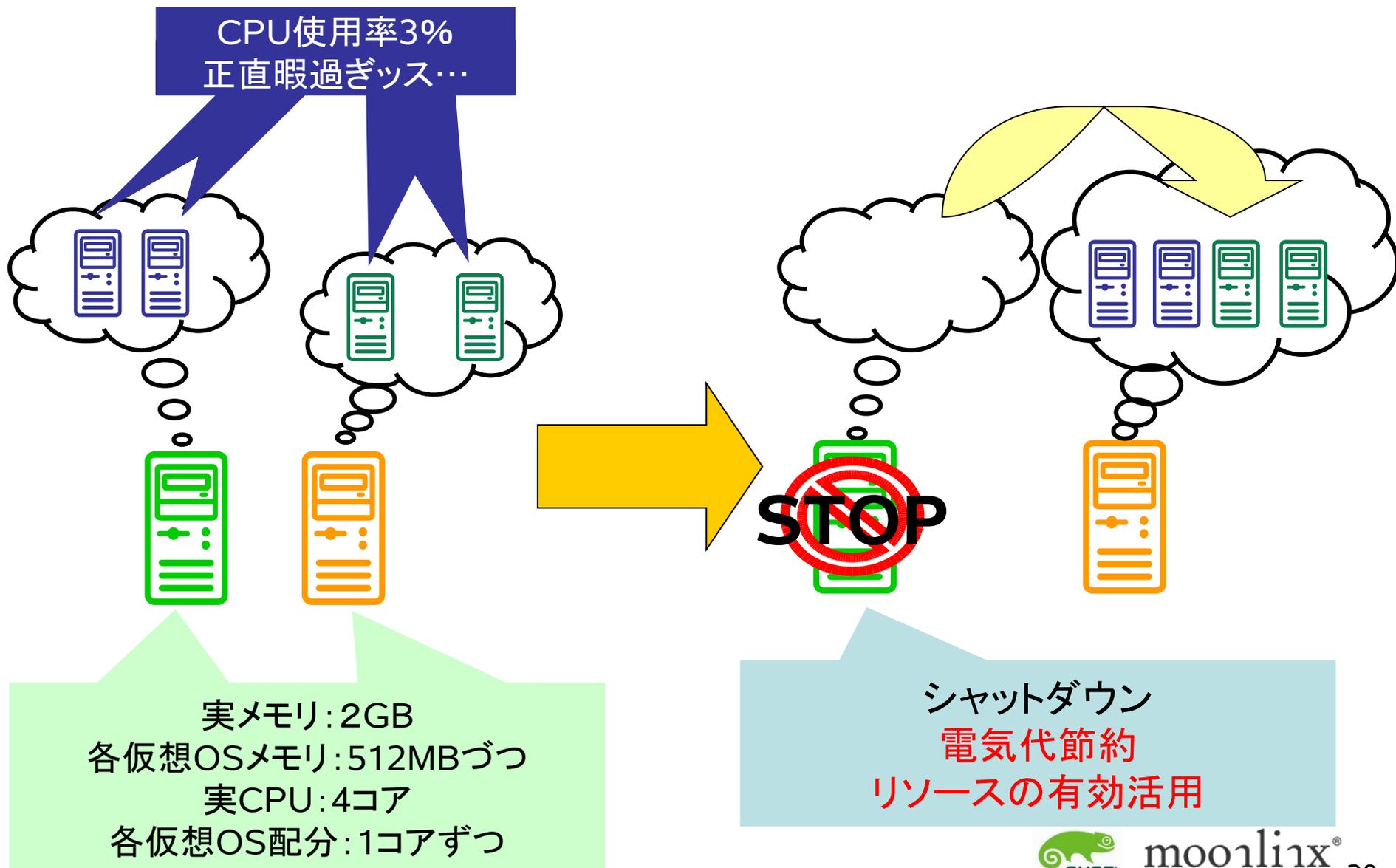
メモリ: **1.5GB**
CPU配分: **2コア**
CPU使用率**ダウン!**



危ない時も、**サービスを止めずに**
柔軟に対応できる

ライブマイグレーションって何に使えるの？ 2

- 例えば…



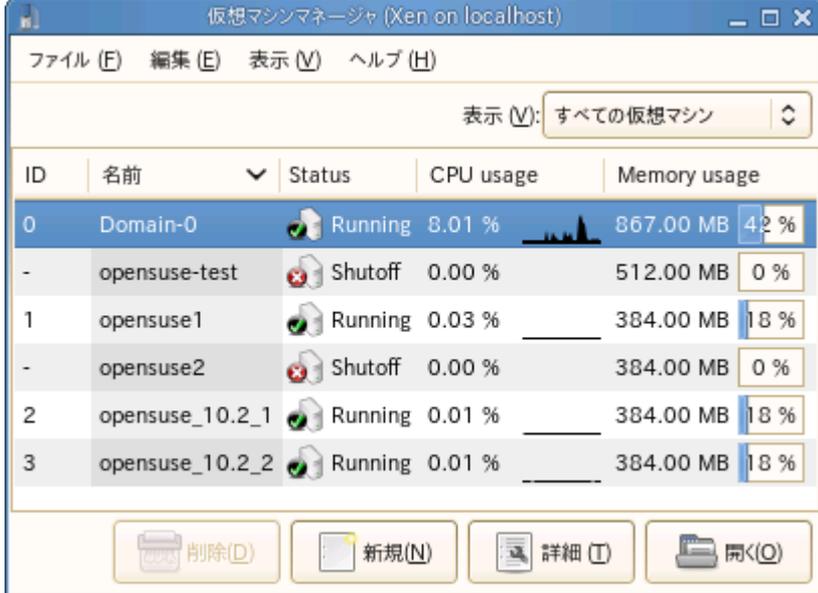
Xenの

ちょっと細かい話

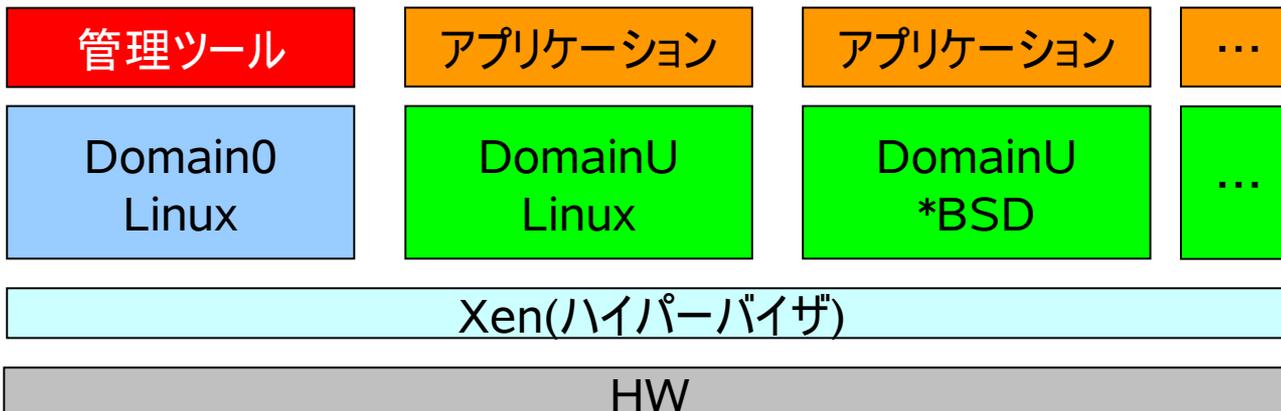


ドメインとは

- ・ XenではVMをドメインと呼ぶ(※windowsネットワーク等のドメインとは無関係)
- ・ Domain0
 - 特権あり
 - 起動時に必ず起動
 - 管理ツールを動作させる
 - 実デバイスの制御
- ・ DomainU
 - ゲストOS、ユーザーに応じて作成/削除
 - Xenデバイスを利用



ID	名前	Status	CPU usage	Memory usage
0	Domain-0	Running	8.01 %	867.00 MB 42 %
-	opensuse-test	Shutoff	0.00 %	512.00 MB 0 %
1	opensuse1	Running	0.03 %	384.00 MB 18 %
-	opensuse2	Shutoff	0.00 %	384.00 MB 0 %
2	opensuse_10.2_1	Running	0.01 %	384.00 MB 18 %
3	opensuse_10.2_2	Running	0.01 %	384.00 MB 18 %



アプリケーションは修正不要

Xenで行える仮想化手法の分類

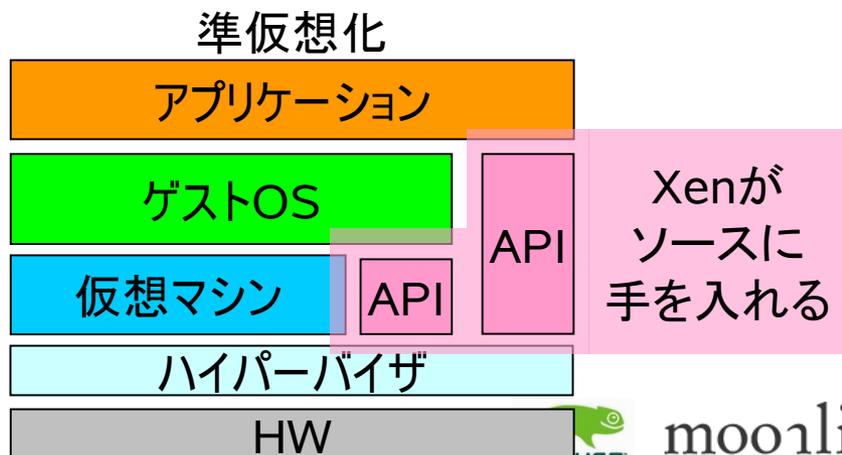
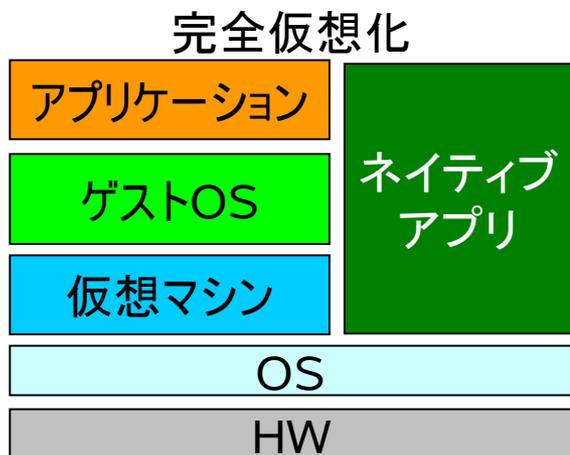
完全仮想化

- マシンを完全にエミュレートする
 - ・ オーバヘッドが大きい
- OSがそのまま動く(windowsマシンをLinux上で動かせる)
- ハードウェアのサポートが必須(VT-x/AMD -V)

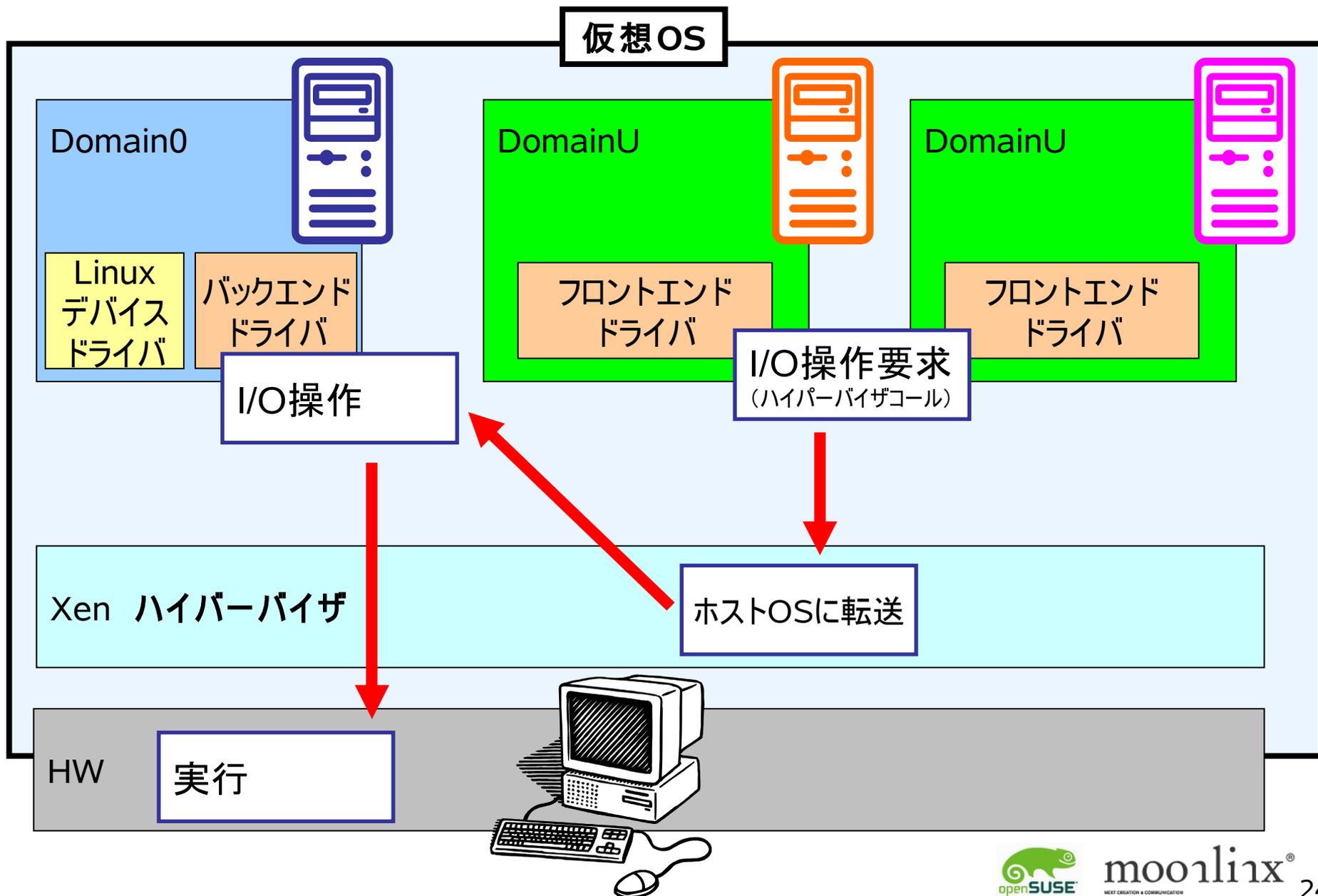
準仮想化

- OSに手を加え、VMマシンに都合の良い仮想的なハードウェア情報を再定義
- オーバヘッドが完全仮想化に比べ小さい
- ゲストOSに修正が必要
 - ・ ゲストOSのソースコードが必要
 - ・ アプリケーションの修正は不要

ちょっと極端な言い方だけど。



準仮想化 デバイスドライバモデル

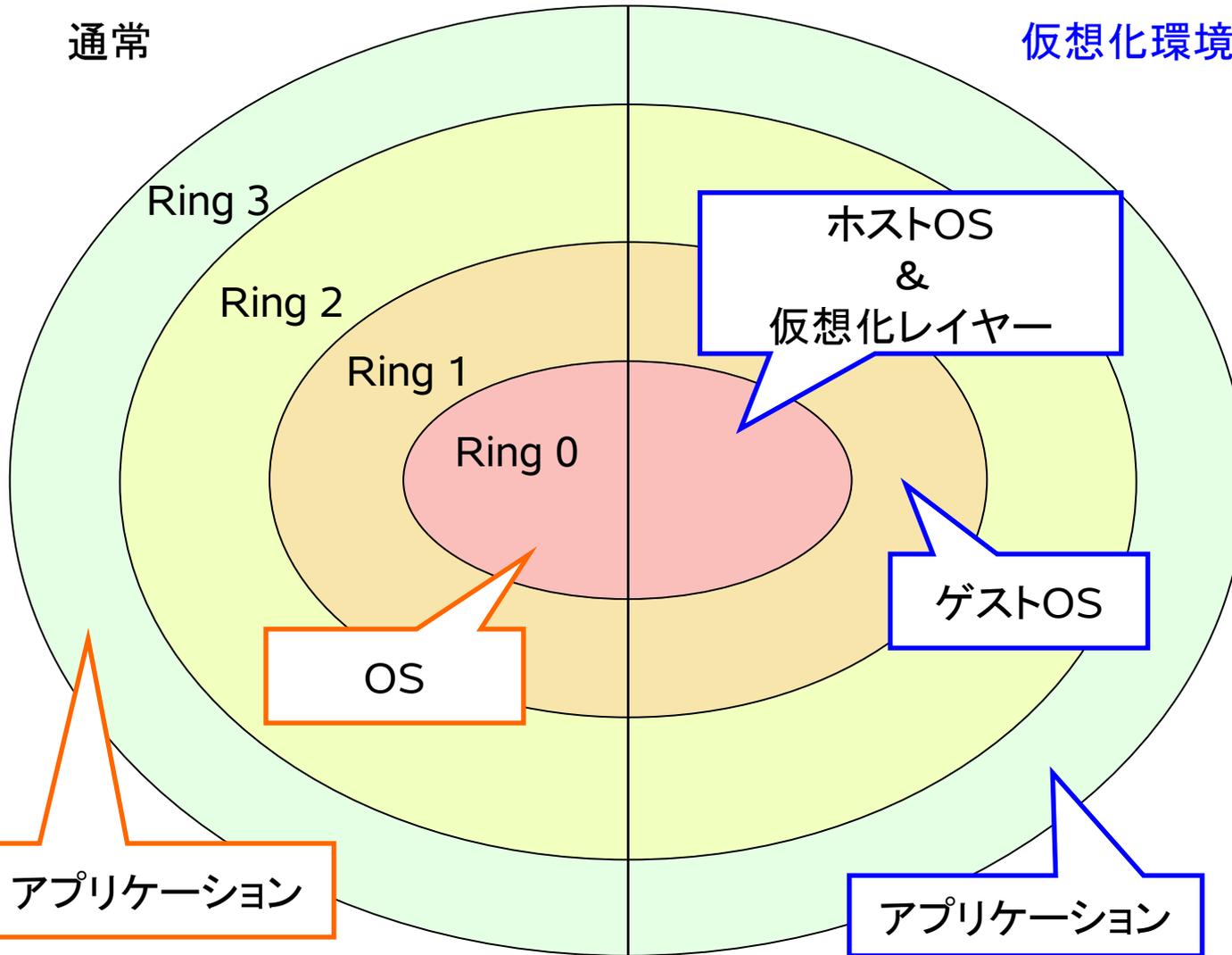


参考: VMware Server 仮想レイヤー処理(バイナリトランスレーション)

プロセッサの動作モード

通常

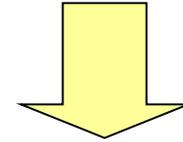
仮想化環境



通常
Ring-0 (特権) ~ Ring-3 の4段階

これまでのRingsの使い分け
Ring-0ではOSカーネルが動作
Ring-1, Ring-2 は未使用
Ring-3ではユーザプログラムが動作

問題点
Ring-0で複数のOSを動かすことが原則できない
※H/Wの設計というよりはOSの設計上の問題



VMware環境でのRingsの使い分け
Ring-0: ホストOSとVMM(仮想マシンモニタ)
Ring-1: ゲストOSカーネルで動かす

Ring Aliasing
本来Ring-0で動くプログラムをRing-1で動かすゲストOSはまるで自分がRing-0に居るように錯覚する
Ring-1での特権操作は、例外としてRing-0のVMMIにトラップされる

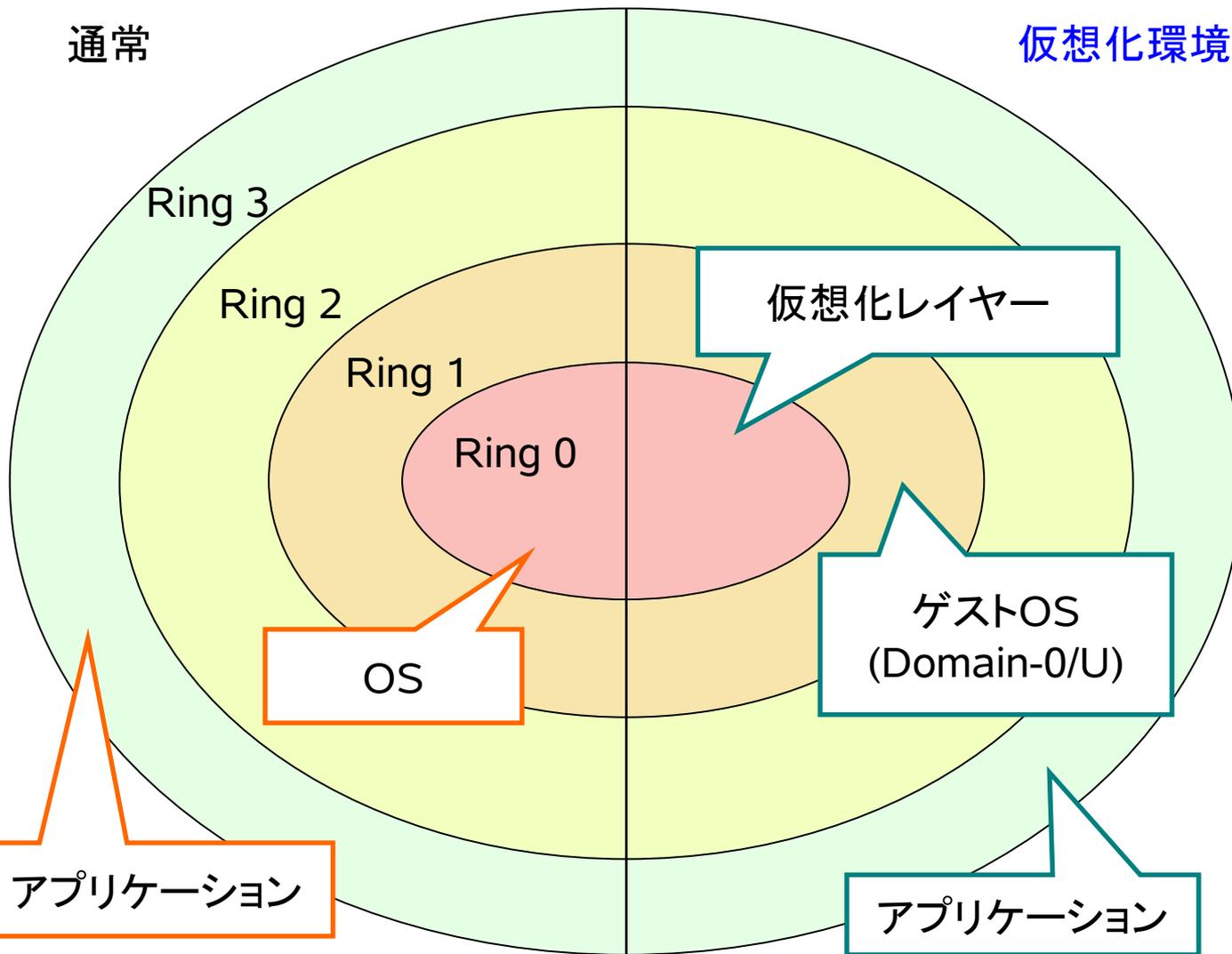
Binary Translation
Ring-1での不都合なコードは動的に書き換え

参考: Xen 仮想レイヤー処理(パラバーチャリゼーション)

プロセッサの動作モード

通常

仮想化環境



Xen環境でのRingsの使い分け

Ring-0: 仮想マシンモニタ
Ring-1: ゲストOSカーネルで動かす

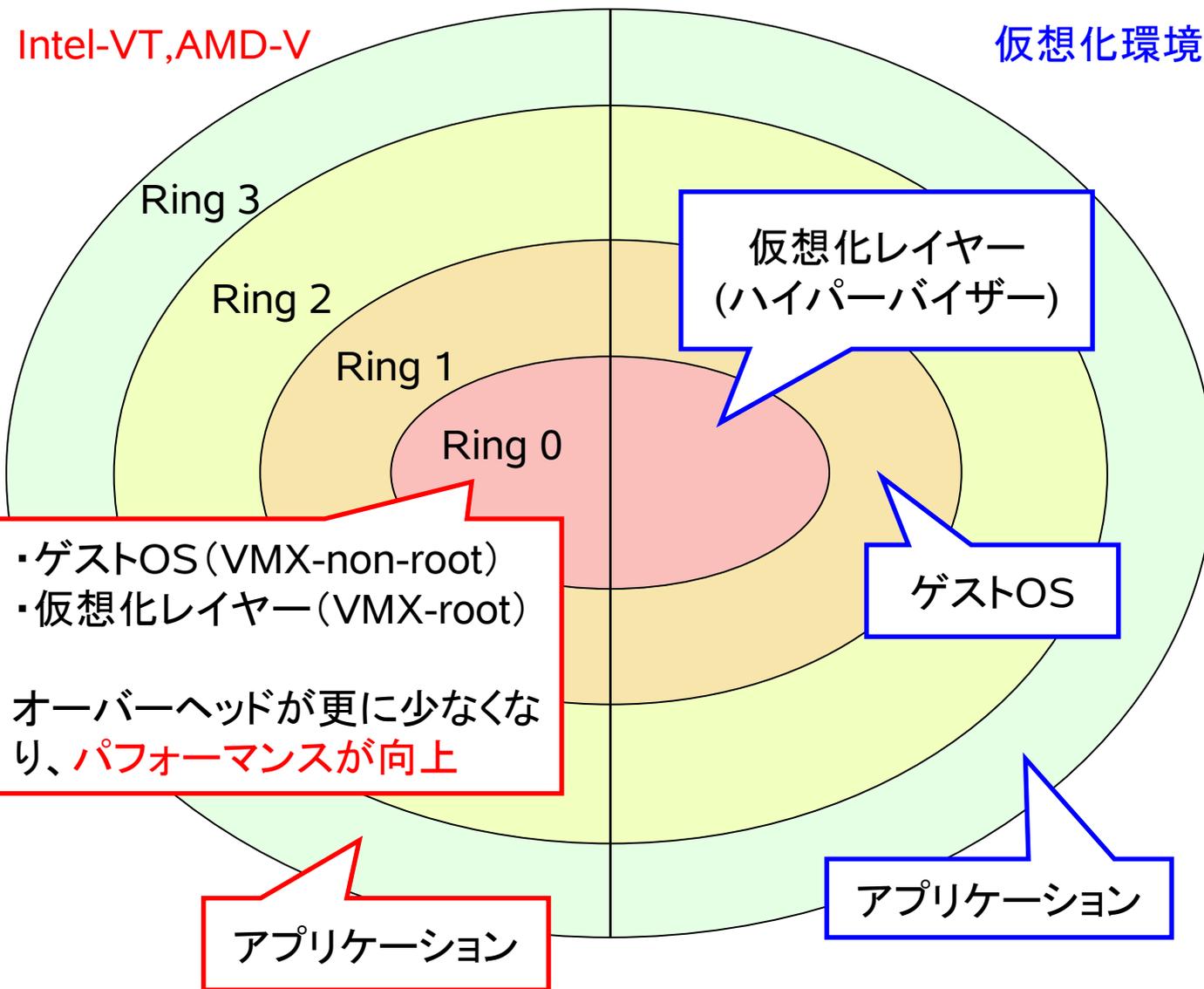
Ring Aliasing

本来Ring-0で動くプログラムをRing-1で動かすゲストOSはまるで自分がRing-0に居るように錯覚する

ParaVirtualization

あらかじめカーネルを修正したゲストOSのみ動かすことになるので、命令のトランスレート処理を行う必要が無く、結果オーバーヘッドを最小限に抑えることが出来る

プロセッサの動作モード



CPUの対応

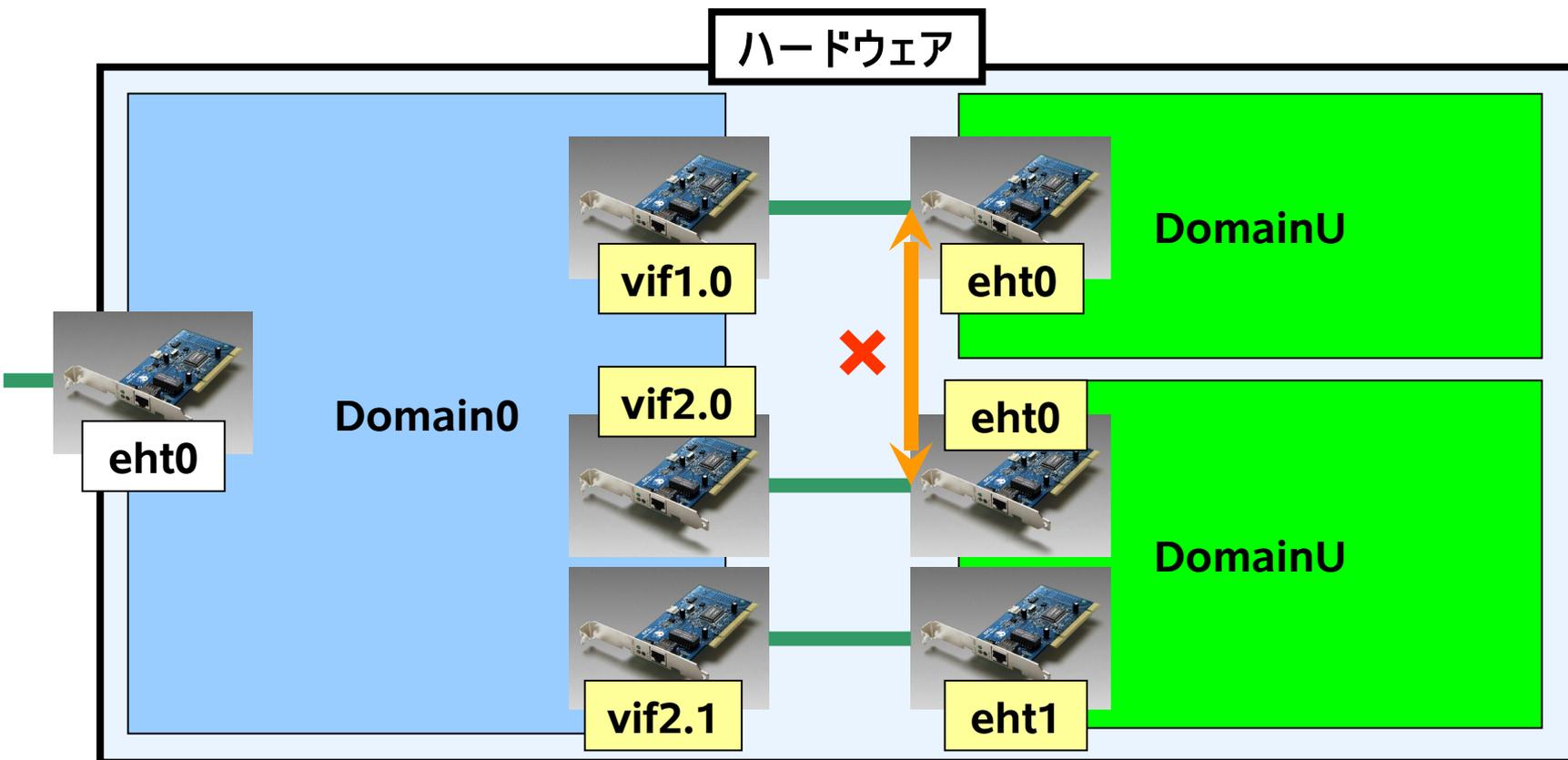
ソフトウェアの力のみで、半ば強引に仮想環境を作成していたものに対し、CPU自体の特権命令に対する柔軟性を持たせた形になる
結果オーバーヘッドを極力減らせることに成功、スムーズに仮想環境を作成出来る

Xenで完全仮想化を実現するのに不可欠

Xenでは基本的に準仮想化(ゲストOSカーネルに手を加える形で仮想化を実現する)のため、ソースコードを改変できないタイプのOSの場合、動作させることが出来なかったが、CPUの対応により完全仮想化を実現することが出来る

ネットワーク

Xenには「ブリッジ」「NAT」「ルーティング」の3つのネットワークモードが存在する
Domain0とDomainUは仮想的なネットワークインターフェースで接続されるが、
DomainUはDomain0に接続されている物理的なネットワークに直接接続できない
また、DomainU同士を繋ぐネットワークも無いため、通常DomainU同士の通信も直接は行えない



- **Performance and Scalability**

- CPUID Levelling
- Shadow 3 Page Table Optimizations
- EPT/NPT 2MB Page Support
- Hardware-accelerated HVM video memory updates tracking
- PVSCSI -- SCSI Support for PV Guests
- Full 16-bit Emulation on Intel VT
- Support for memory overcommit allowing more VMs per physical machine for some workloads

Xen.org /
xen3.3datasheetより

- **Security**

- PVGRUB Secure Replacement for PYGRUB
- IO Emulation “stub domains” for HVM IO

- **Green Computing**

- Enhanced C & P State Power Management

- **Graphics Support**

- VT-d Device Pass-Through Support
- Direct OpenGL-accelerated video output scaling

- **Miscellaneous**

- Upgrade QEMU Version
- Removal of Domain Lock for PV Guests
- Message Signaled Interrupts
- Greatly improved precision for time-sensitive SMP VMs

openSUSEの話は

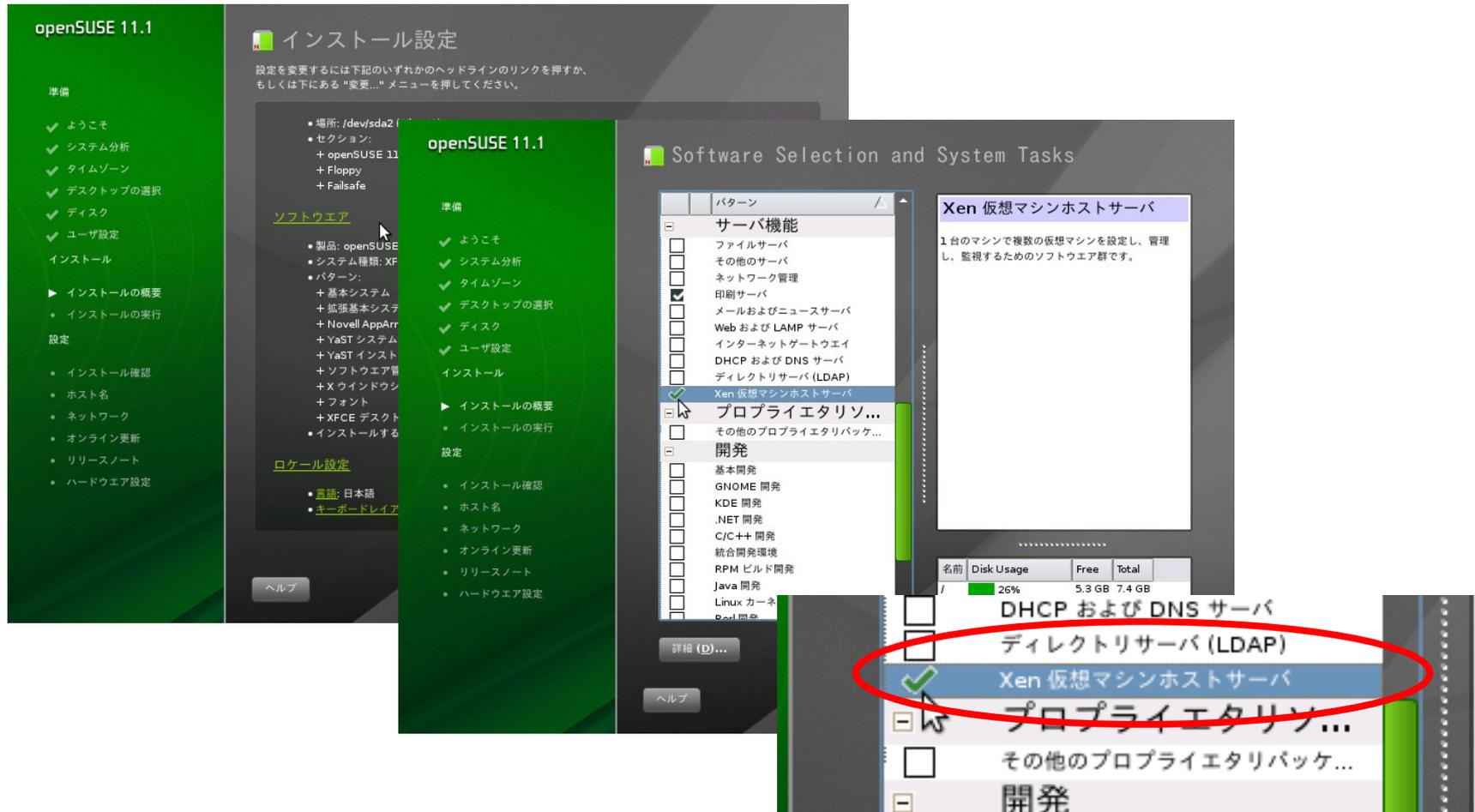
どこいったんだ



インストールは
超絶カンタン！

Xenの導入

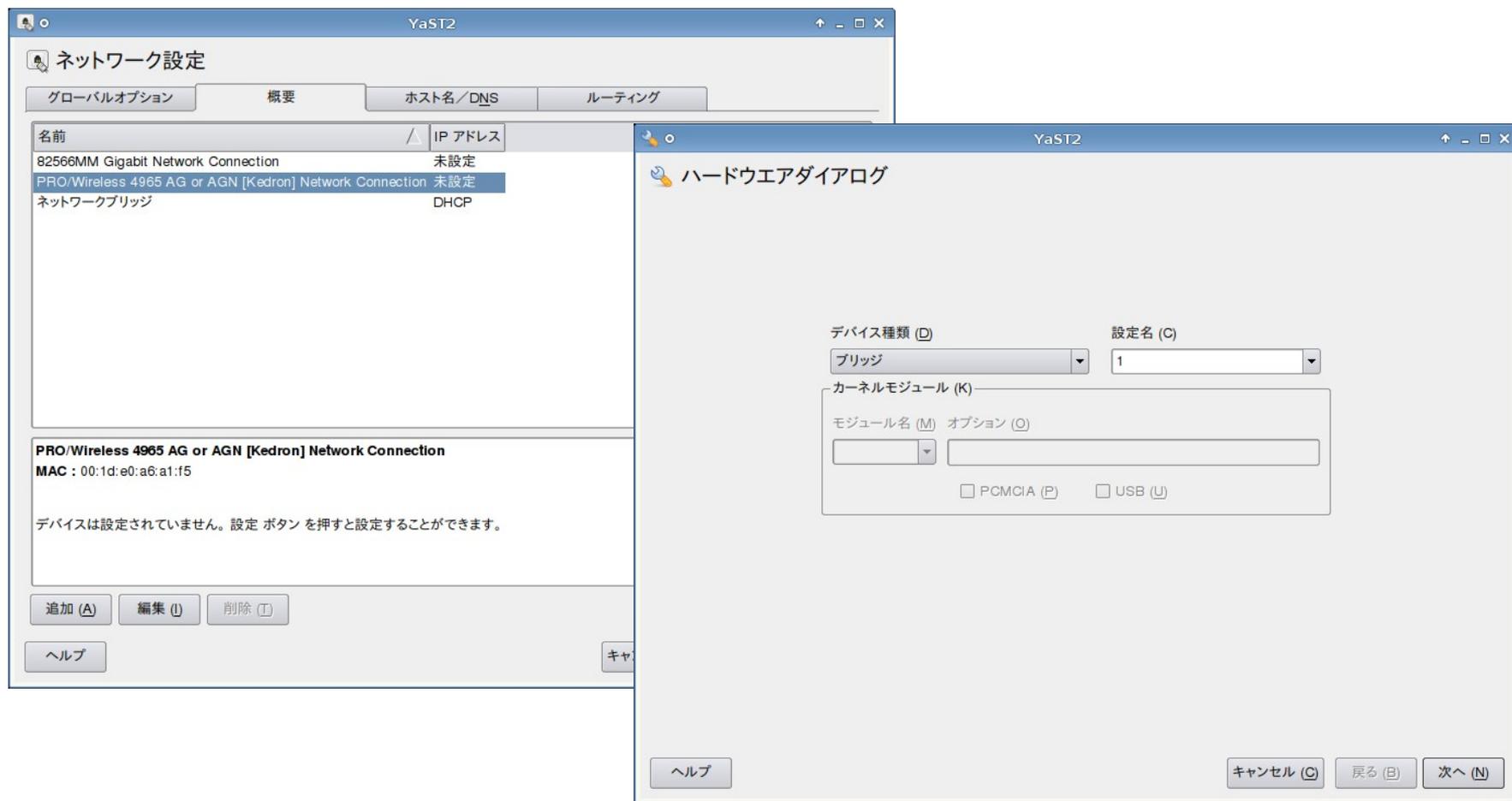
OSインストール時にインストール



OSインストール後でも導入できる。その場合…

- zypperを使う(yum,apt-getみたいなもの)
- YaST(いわゆるコントロールパネル)の『ソフトウェア管理』よりインストール
- ソースからコンパイルしても良いけど、今回は省きます

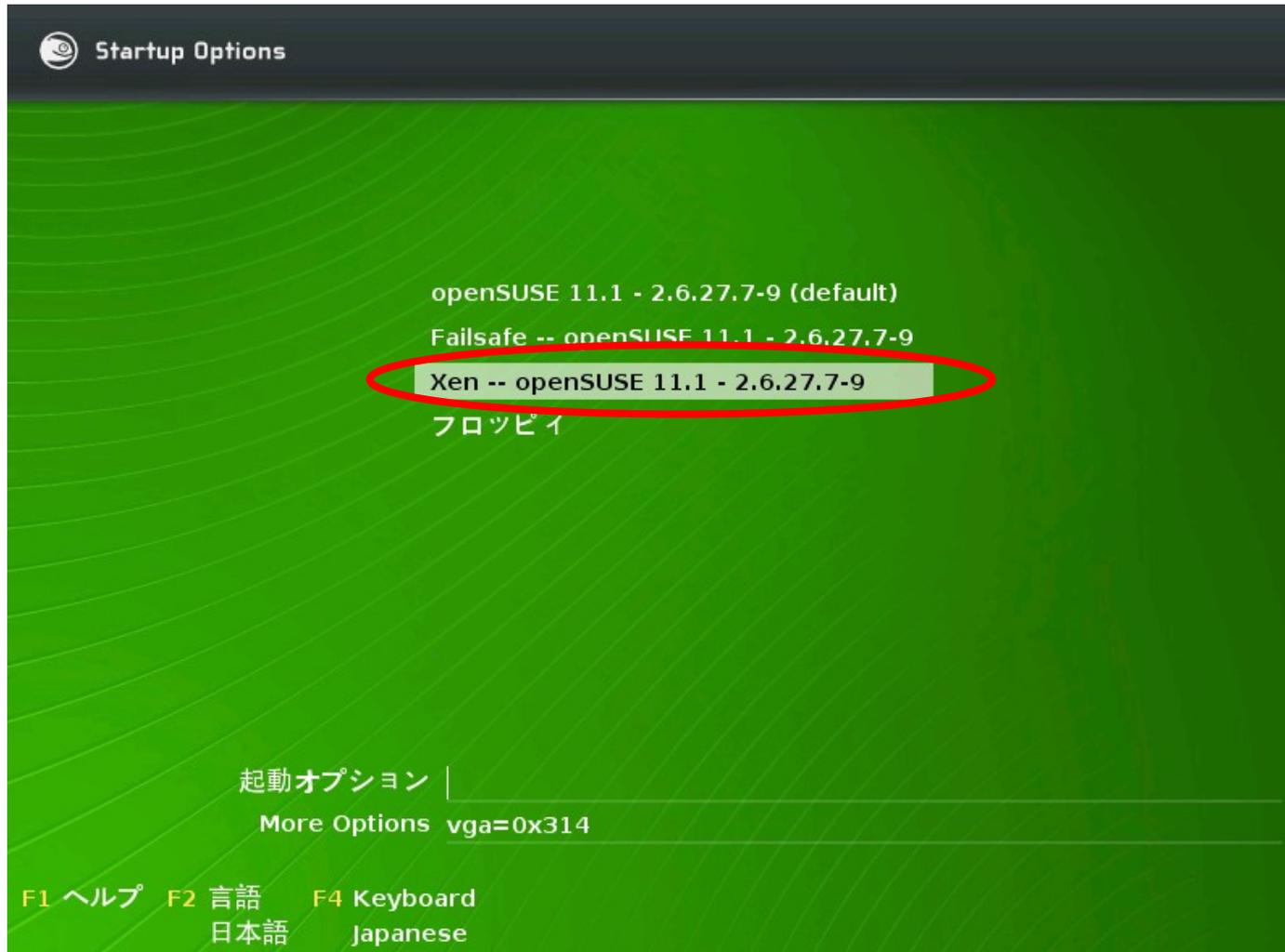
ネットワークの設定



openSUSE11.1より、仮想ネットワーク・ブリッジやNATの設定もインストール時に出来るように

- ・ 後で変更したい場合はYaST -> ネットワークデバイス -> ネットワークの設定

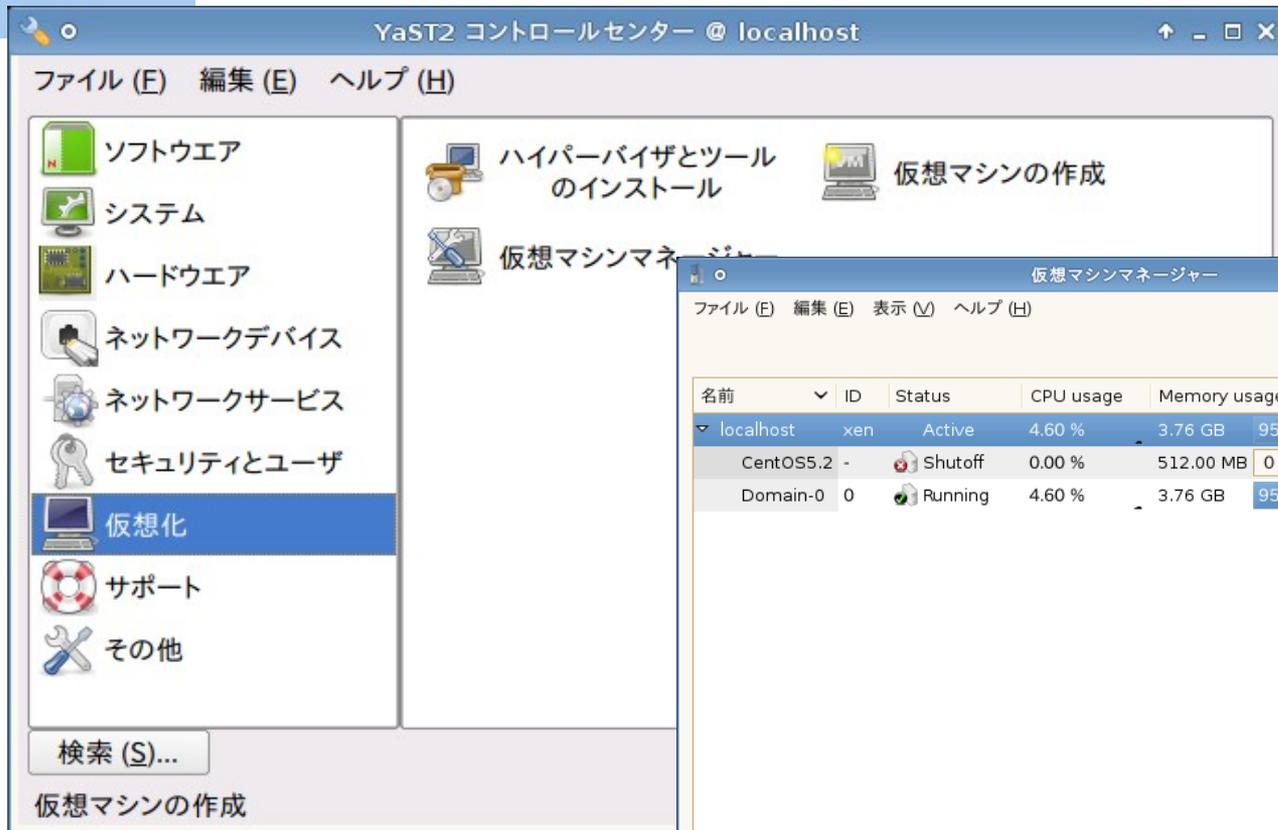
再起動



OSインストールが完了した後、再起動しXenカーネルで立ち上げなおす

管理も
親切設計！

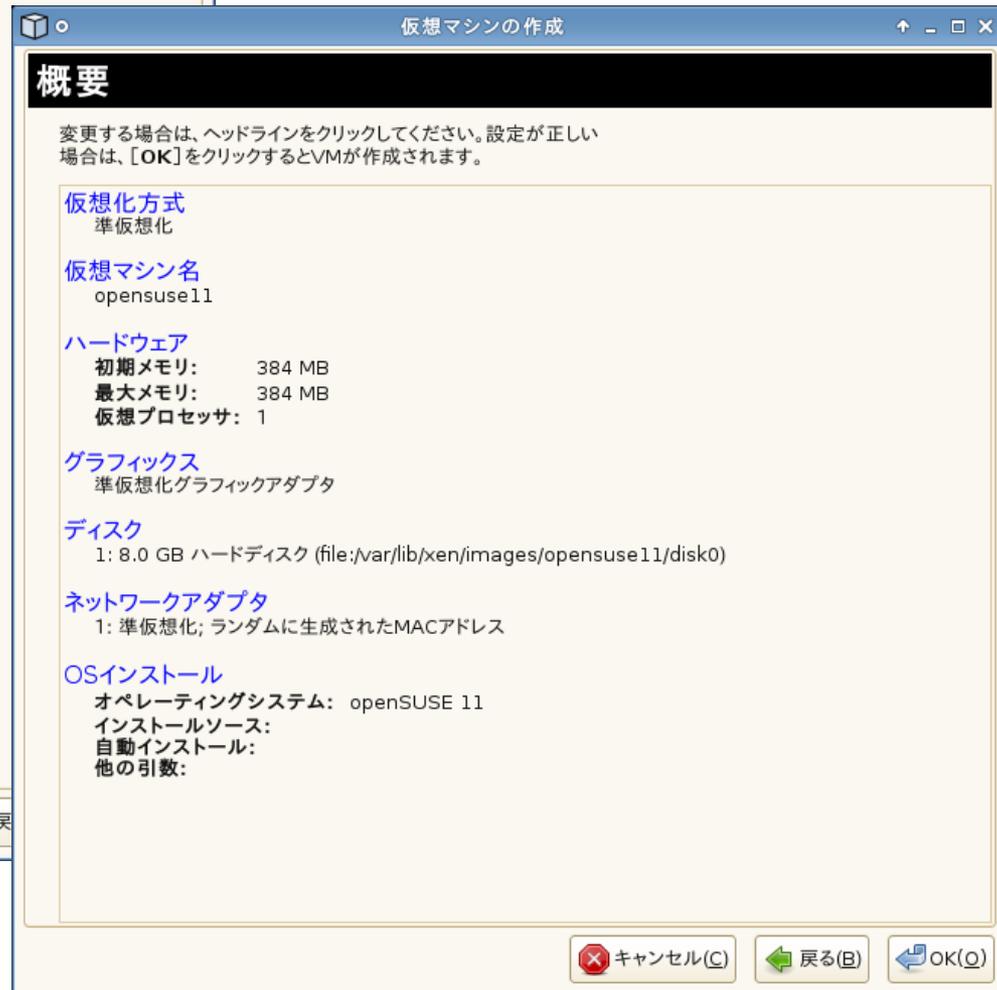
仮想マシンの管理



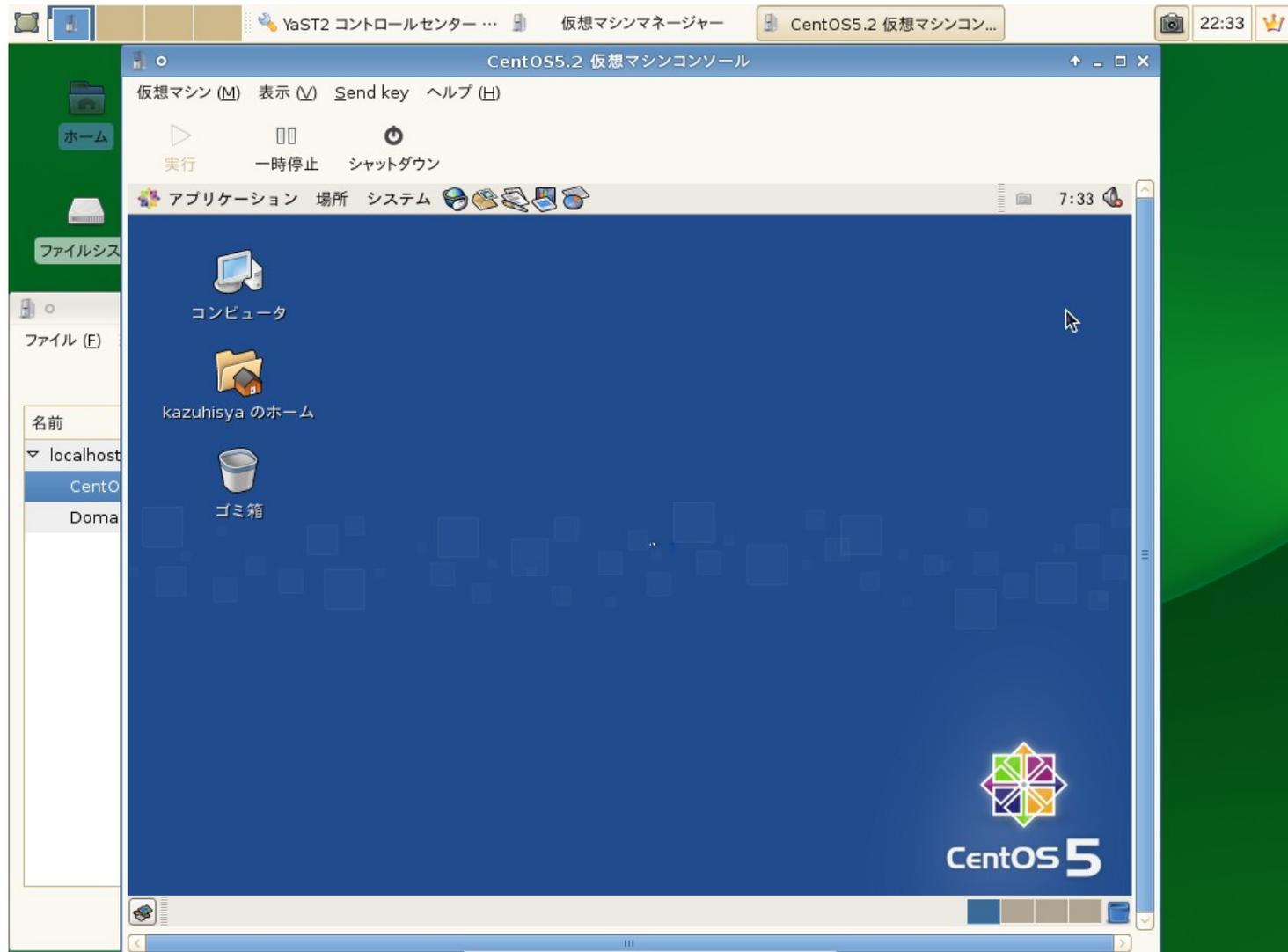
YaST → 仮想化 → 仮想マシンマネージャー or 仮想マシンの作成

作成・削除・メモリ割り当ての変更・DISK交換等々…GUIでほぼ管理できる。

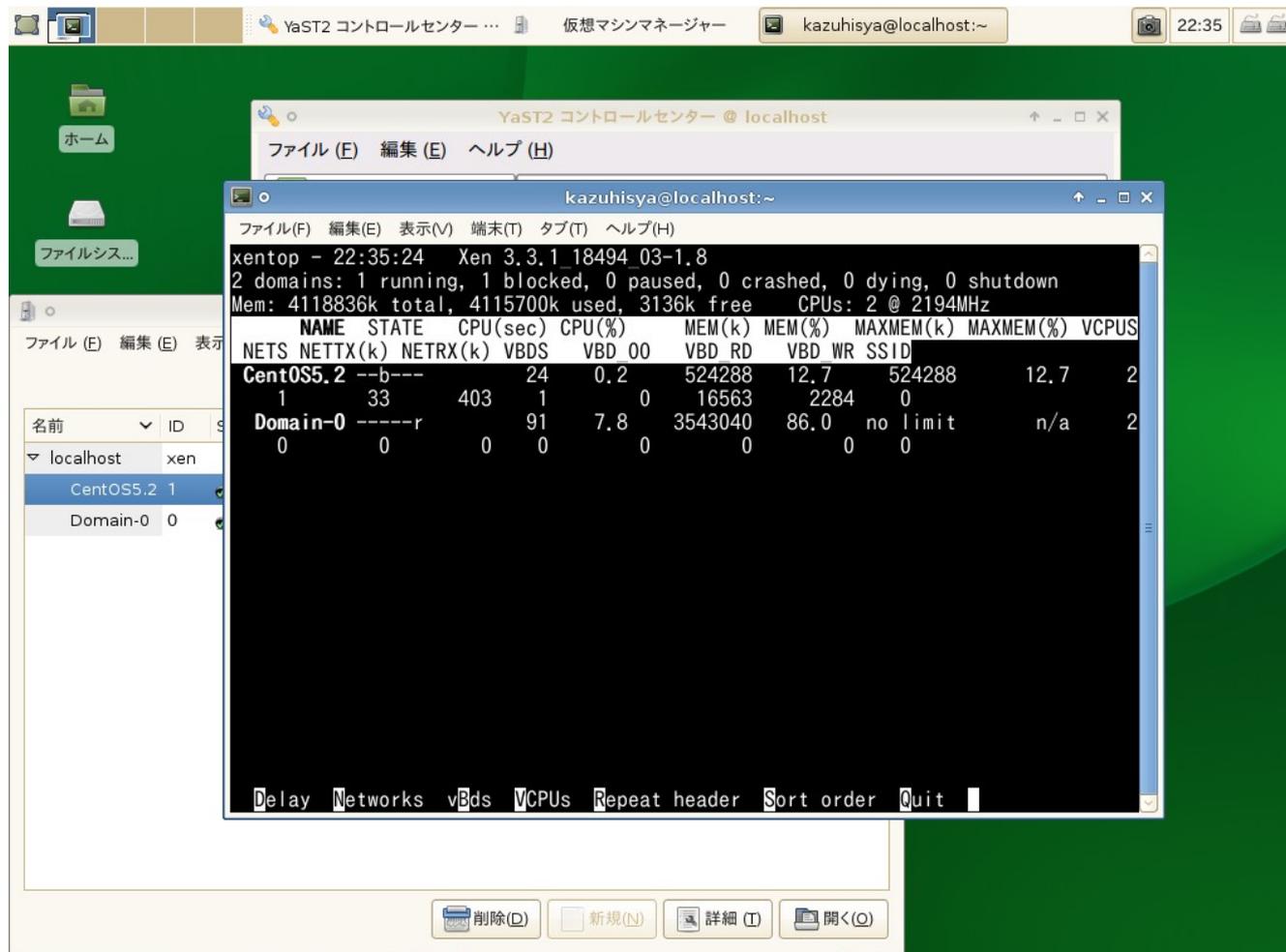
仮想マシンの作成



起動



ゲストOSのX(GUI)も普通に使えます



勿論CUIからでも、仮想マシンの管理が出来る。ゲストマシンのコンソールに直に入ったりも。また、マイグレーションなど特定の作業はコマンドラインでしか実行できない。

moonlinuxでも

やってるお勧め



運用監視

- ・ Xenの状況は**Nagios**を使って常に監視
- ・ Dom0もDomUも監視しちゃえば良いよ

Nagios

一般

- ホーム
- ドキュメント

モニタリング

- 総合監視状態
- サービス稼働状態
- ホスト稼働状態
- ホストグループオーバービュー
- ホストグループサマリ
- ホストグループグリッド
- サービスグループオーバービュー
- サービスグループサマリ
- サービスグループグリッド
- ステータスマップ
- 3Dステータスマップ
- サービス障害
- ホスト障害
- ネットワーク障害

ホスト検索:

- コメント
- ダウンタイム

現在のネットワーク状態

最終更新: 2008年9月3日(水)
14:07:39
更新間隔: 90秒毎

ログイン名: moonlinux

Nagios@ - www.nagios.org

[全ホストの履歴を見る](#)
[全ホストの通知履歴を見る](#)
[全ホストのホスト稼働状態を見る](#)

ホスト稼働状態の概況				サービス稼働状態の概況				
稼働 (UP)	停止 (DOWN)	未到達 (UNREACHABLE)	保留 (PENDING)	正常 (OK)	警告 (WARNING)	不明 (UNKNOWN)	異常 (CRITICAL)	保留 (PENDING)
0	1	0	0	3	0	0	2	0

全稼働	全タイプ
1	1

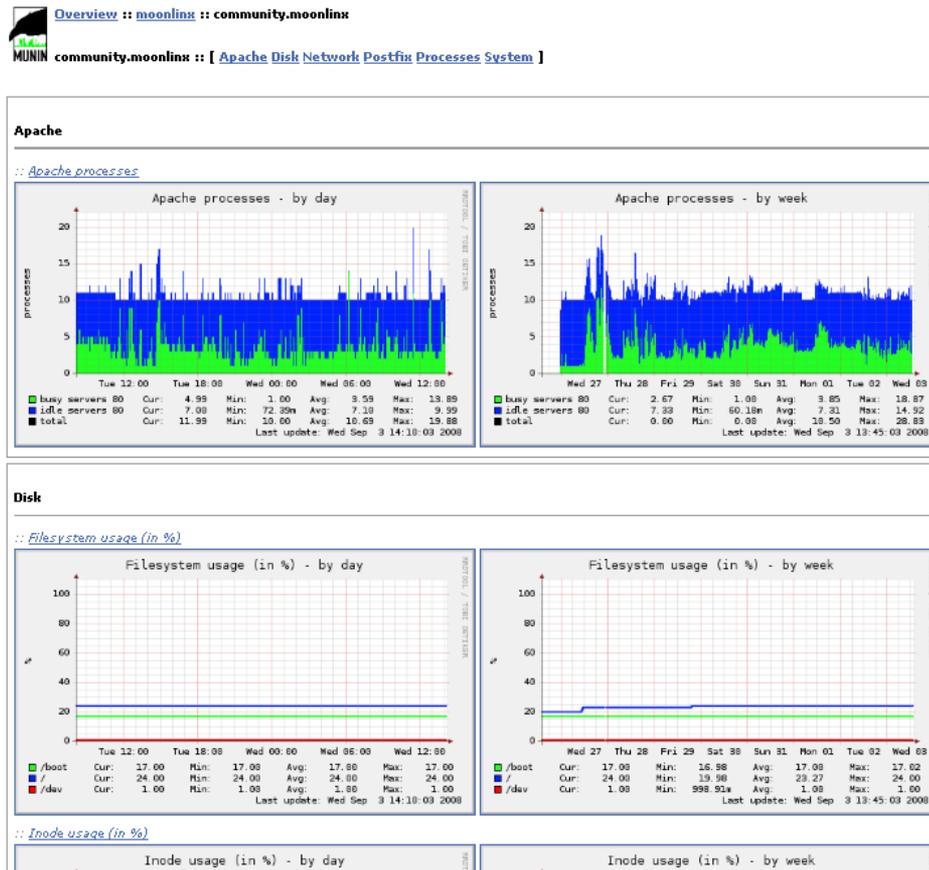
全稼働	全タイプ
2	5

全ホストの稼働状態

ホスト名 ↑↓	エイリアス	サービス名 ↑↓	状態 ↑↓	最終チェック時刻 ↑↓	経過時間 ↑↓	試行回数 ↑↓	ステータス情報
server219	Nagios server219	219_DISK	正常 (OK)	07-24-2008 10:55:46	99日と20時間 32分 46秒	1/3	Disk space OK - 28419876 kB free (59 % used)
	Nagios server219	219_HTTP	正常 (OK)	07-24-2008 10:56:49	69日と19時間 1分 27秒	1/3	HTTP OK HTTP/1.1 200 OK - 6917 bytes in 0.005 seconds
	Nagios server219	219_HTTP_v2	正常 (OK)	07-24-2008 10:57:46	99日と20時間 57分 16秒	1/3	HTTP OK HTTP/1.1 200 OK - 6917 bytes in 0.009 seconds
	Nagios server219	219_HTTP_v3	異常 (CRITICAL)	07-24-2008 10:58:53	48日と18時間 18分 53秒	1/3	Connection refused
	Nagios server219	219_PING	異常 (CRITICAL)	07-24-2008 10:59:46	48日と18時間 25分 13秒	1/3	PING CRITICAL - Packet loss = 0%, RTT = 0.02 ms

モニタリング

- サーバ / サービスのモニタリングは**Munin**を利用
- DomU自体にエージェント仕込むもよし、Dom0用のプラグイン使うもよし



- ・ **Xenカーネルで起動したら、Xが立ち上がりません** > <
 - グラフィック・ドライバの可能性大、3Dアクセレータは切った方が吉かも
 - 3Dアクセレータの設定は、YaST -> ハードウェア -> グラフィックカードとモニタ
- ・ **VTやAMD-V対応CPUなのに完全仮想化できません** > <
 - BIOSの設定でOFFになっていませんか？デフォルト=OFFの場合も結構あるようです
- ・ **ゲストマシンのローカル時間が激しくずれるわけだが**
 - ホストマシン+9時間、更にソコからゲストマシンは+9時間してるっぽい
 - ・ 差異をある程度吸収した設定にしたなら、NTPでこまめにとるのがベター
- ・ **ゲストマシンのマウスが激しくずれるわけだが**
 - VNCやリモートデスクトップで接続すると、結構快適に使えます

おさらいしましょー



- ・ **サーバ統合**
 - サーバの運用コストの削減
- ・ **サービス独立性の維持**
 - ネットワーク、アプリケーション、ドライバ…各仮想マシンが独立したOS環境
- ・ **セキュリティの確保**
 - 各仮想マシンが独立しているため、他の仮想マシンに影響を与えることは無い
- ・ **システムの拡張性と柔軟性**
 - 仮想的にルータを構築できるなど、ハードウェアの物理的制約を緩和
- ・ **仮想マシンの移動が出来る**
 - ホストマシンさえ動いてしまえば、他のマシンに簡単に移行できる

まとめ2 - Xen導入のメリット

- ・ **サービスへの資源量の確保**
 - 各仮想マシンに**独立した資源を割り振る**ため、CPU時間、メモリ量を保障できる
- ・ **仮想マシンの能力変更**
 - 仮想マシンを停止することなく、**動的に**資源を割り振ることが出来る
- ・ **仮想マシンの移動が出来る**
 - ライブマイグレーションを用いれば、ほぼ**無停止で**仮想マシンを移動できる
- ・ **ハイパーバイザー型である**
 - ホスト型に比べ**パフォーマンスに優れる**
- ・ **完全・準仮想に対応**
 - Linux、Solarisに加え**Windowsも動かせる**
- ・ **OSS**
 - 商用ソフトに比べ、**コスト**が掛からない、最悪ソースコードからバグを追求できる

・ 簡単、ラクチン

- OSインストールの時点で、Xenの設定も完結できる
- 仮想ネットワークの設定、GUIでの仮想マシン管理…YaSTが優秀で楽が出来る

・ 親和性

- Novell(&SUSE)はXen全体での拡張・修正コードの正式採用数のシェア、第3位
- その他ディストリビューションはDebian9位、Red Hat10位とのこと

・ 実務応用可

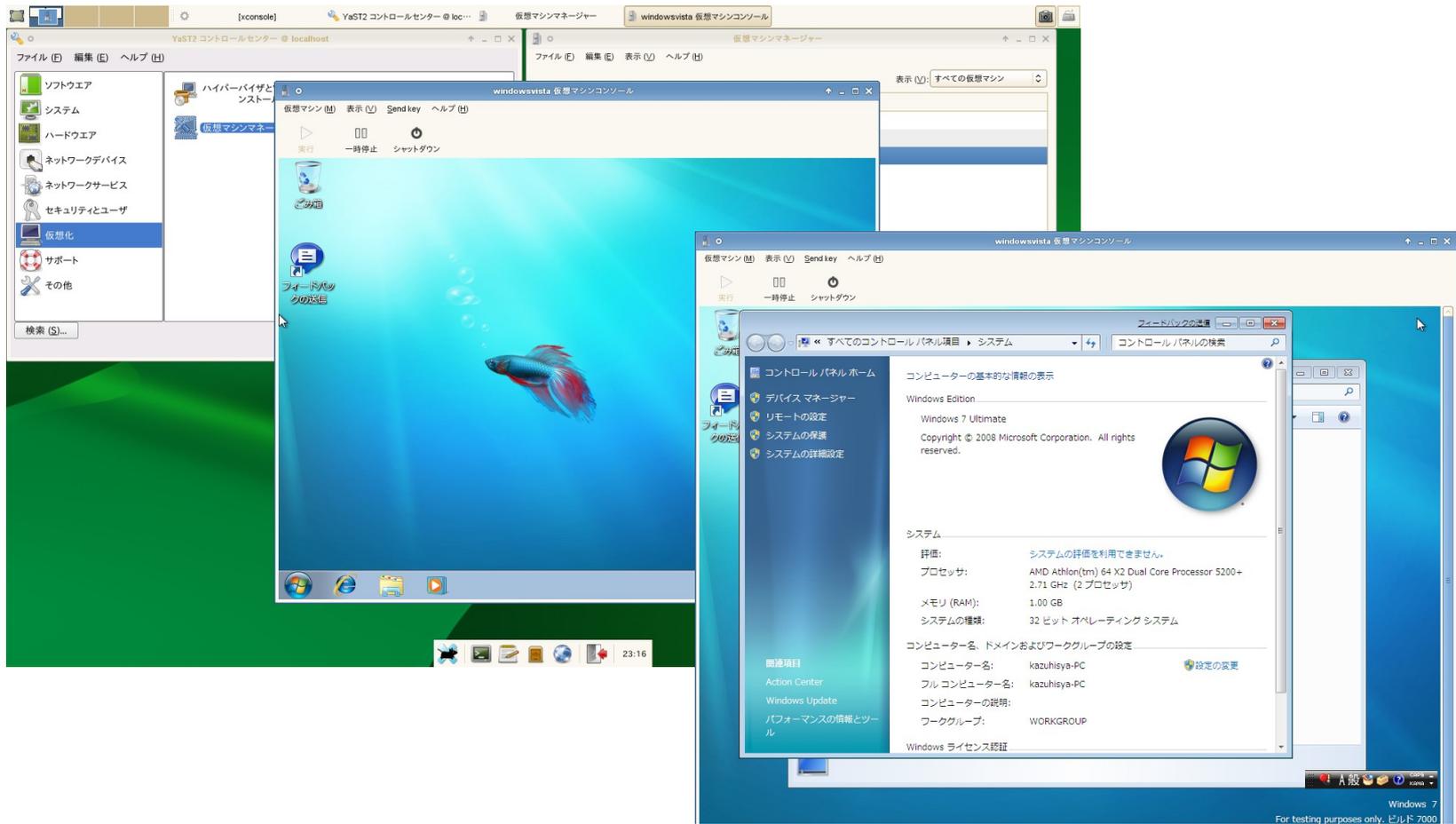
- YaST管理ツールや、仮想マシンの管理等は、ほぼSLES※でも活用できる
 - ・ SLES + Xen の組み合わせは結構実績豊富
 - ・ また、MS製品以外で仮想マシン上のWindowsが公式にサポートされている
 - Windows用の準仮想化ドライバもある

※ SUSE Linux Enterprise Server

・ Xenと組み合わせて使えるものが揃っている

- Xen + DRBD + Heartbeat2 なんかの構成もYaSTからのインストールで割りとしん
なり実現

おまけ



完全仮想化、OSタイプ Vista で Windows 7 Beta(Build 7000) も動きますよー

Thank you for listening !

moonlinux[®]

ここでQ&A Time !

<http://moonlinux.jp>