

Linux

Open  
Source

Kernel

Virtualization



VA LINUX  
S Y S T E M S  
J A P A N

# OSS仮想マシン環境動向 XenとKVM

VA Linux Systems Japan株式会社

**小田 逸郎**  
エンタープライズOS事業ユニット ユニット長  
VA Quest隊長

**Itsuro Oda**

- VA Linux Systems Japanとは
- 仮想マシンのアーキテクチャ
- 開発動向
- まとめ

## カーネル、ハイパーバイザの開発者

- ・開発技術をコアとしたビジネスを展開
  - VA Quest / Xen Quest  
Linuxカーネル / Xen の障害解析サービス  
正確、迅速。コードレベルでの原因解説。
  - VA Expert / Xen Expert  
Linuxカーネル / Xen 開発請負  
コミュニティとの協調ノウハウ豊富。
  - VA Zest / Xen Zest  
Linuxカーネル / Xen コンサルティング

・開発実績 (Linuxカーネル)

- NFS高速化 (zero-copy NFS)
  - メモリのホットプラグ機能
  - 資源制限機能
  - crash dump機能
  - I/O帯域制御機能 (dm-ioband)
- 仮想マシン環境での使用も想定

・開発実績 (Xen)

- IA64 対応
- メモリ管理機能を中心に、Xenのコアとなる部分を全面再設計。  
IA64アーキテクチャのメンテナを担当
- crash dump機能

クローズドソフト

フィールド部門



開発部門

オープンソフト

フィールド部門

(開発・検証部門)

VA Linux  
(開発部門の位置付け)

開発コミュニティ



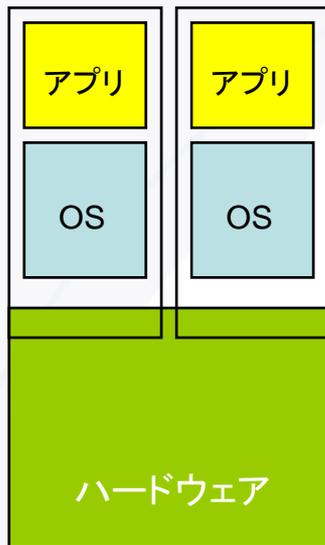
開発コミュニティとの  
橋渡しも受け持つ

# 仮想マシンのアーキテクチャ

- Java VM
- アプリケーションレベルの仮想化
  - Apache virtual hosting
- OSレベルの仮想化
- 仮想マシン環境
  - システム全体を仮想化する
  - 複数のOSを同一マシン上で動作させる技術
  - 古くは1960年代IBMメインフレームの頃に遡る
  - Virtual Machine Monitor(VMM)やHypervisorなどと呼ばれる
- ストレージの仮想化
- ネットワークの仮想化

# 仮想マシン環境のアーキテクチャ

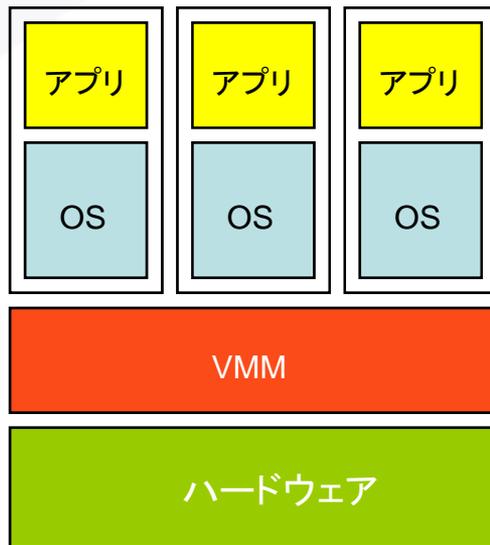
ハードウェア  
パーティション



メインフレーム  
大型UNIXサーバ

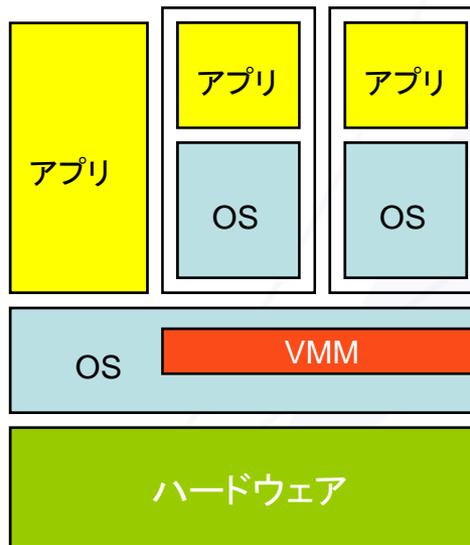
仮想マシン環境

直接VMM型



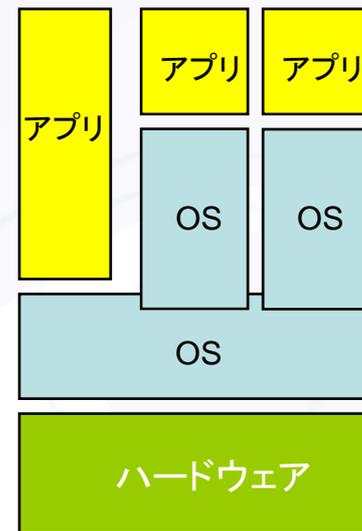
VMware ESX  
Xen  
Hyper-V

ホストOS型



VMware Server  
KVM  
QEMU  
Virtual Box

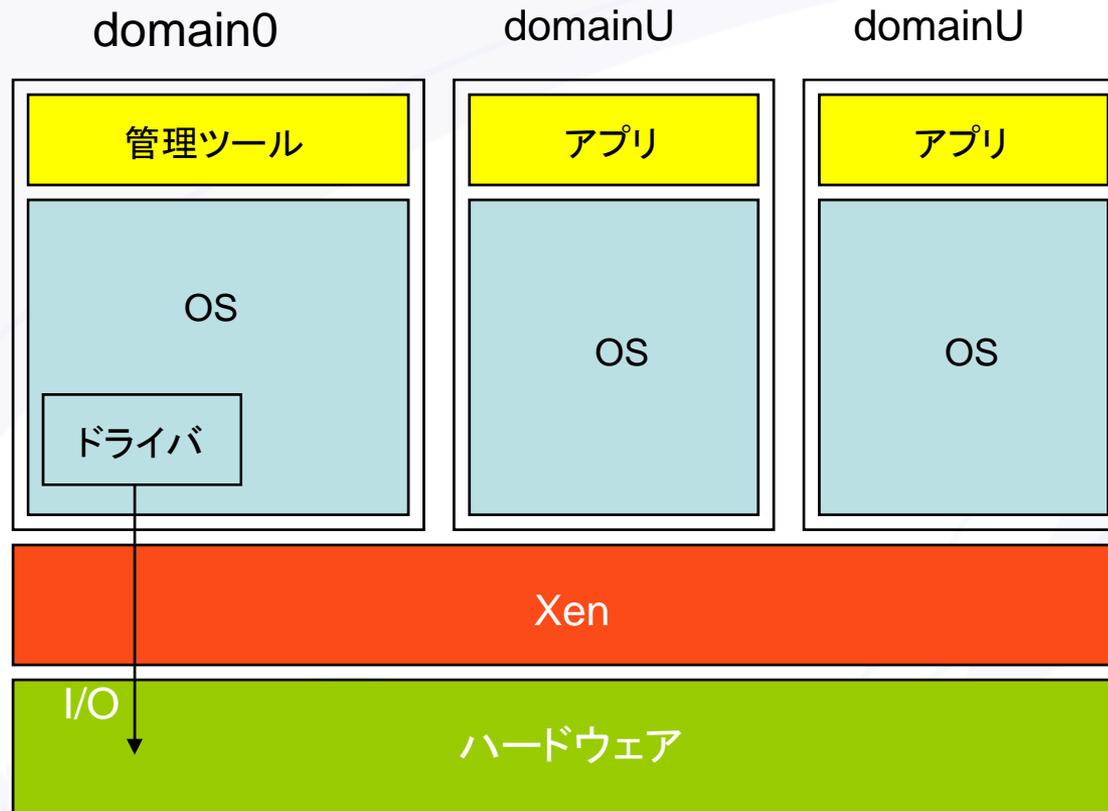
OSレベルの  
仮想化



BSD jail  
Solaris container  
Linux vserver  
OpenVZ  
Virtuozzo

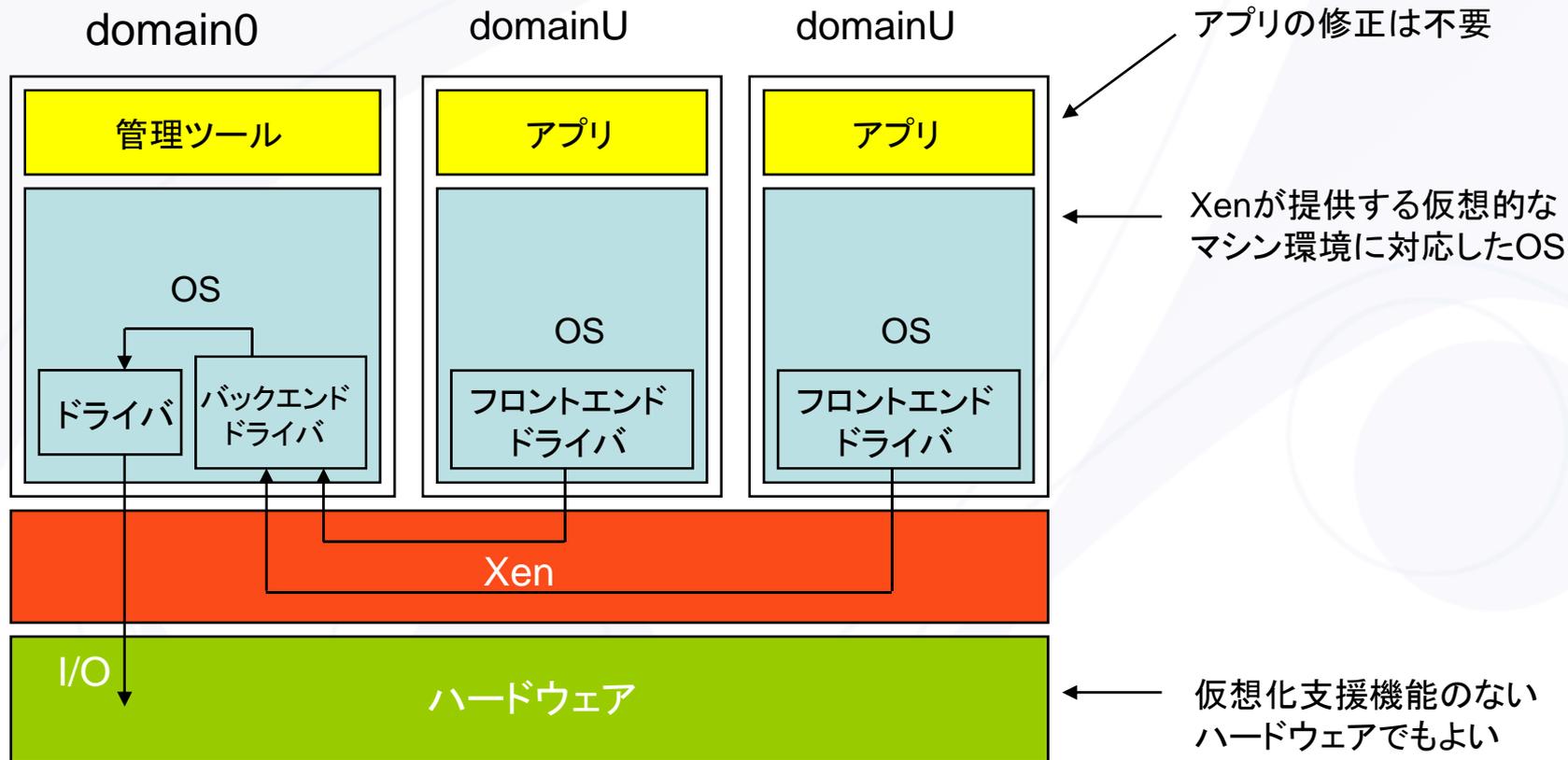
- 準仮想化 (para virtualization)
  - 新たに定義した仮想的なハードウェア環境を提供
  - OSは上記環境に対応したものが必要  
(既存OSはそのままでは動かない)
  - 性能重視
  - 例) [Xen](#)
- 完全仮想化 (full virtualization)
  - 既存ハードウェア環境を完全にエミュレートする
  - 既存OSがそのまま動作する
  - オーバヘッド大
  - 例) VMware、[KVM](#)、[QEMU](#)、[Xen](#)

# Xenのアーキテクチャ(1)

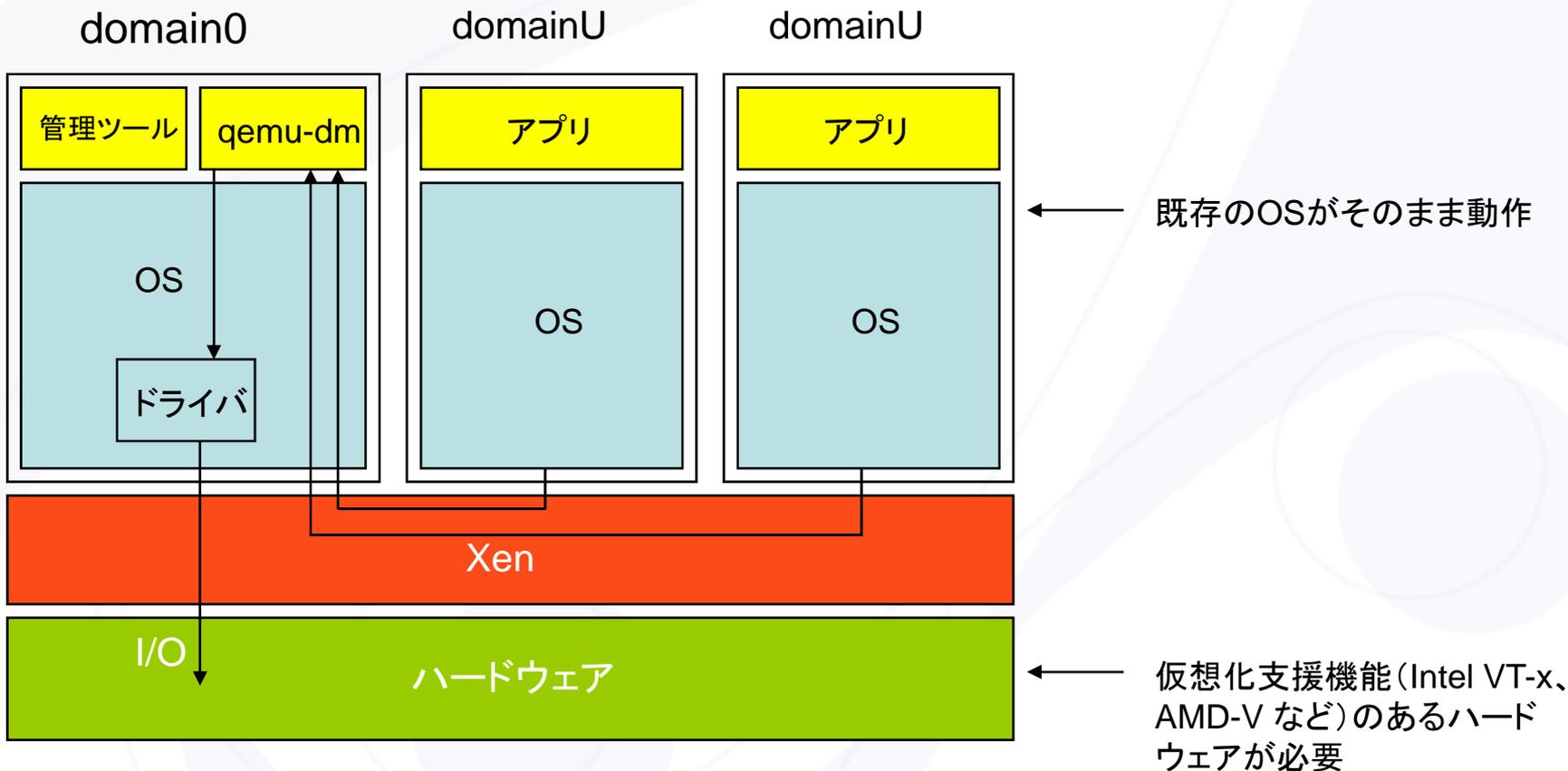


- ・管理用の仮想マシン (domain0) を設け、仮想マシンの管理とデバイスの制御を任せる。
- ・domain0用OSは通常、Linuxを使用。NetBSDやOpenSolarisなども使用可。
- ・準仮想化と完全仮想化の両方をサポート(混載可)

## 準仮想化

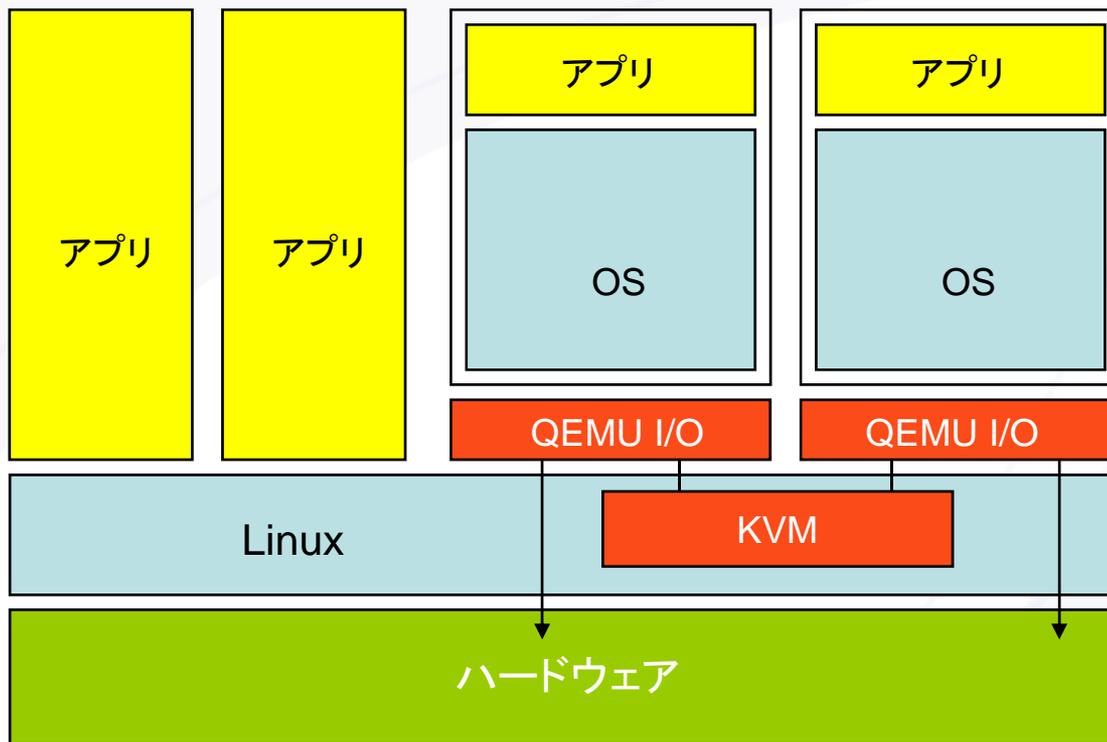


## 完全仮想化



- ・QEMUのデバイスエミュレーション部分を利用。domain0上のQEMUがデバイスをエミュレート。domainUからは、QEMUがエミュレートしたデバイスが見える。

# KVMのアーキテクチャ



- ・Linuxカーネルのモジュールとして実現  
Kernel based Virtualization Machine
- ・完全仮想化  
ハードウェアの仮想化支援機構 (Intel VT-x、AMD-Vなど) が前提
- ・デバイスのエミュレーションは、QEMUを利用

# 開発動向

- ・ケンブリッジ大学のIan Prattが中心となり開発。  
2003年～
- ・オープンソース。開発はインターネット上で行われており、誰でも参加可能。  
現在、Xen.org、Xen Advisory Board が取りまとめ。  
<http://www.xen.org>
- ・Linuxカーネル、QEMUなど他のオープンソースの成果を利用。
- ・CPUベンダ各社、米国有力ベンダ各社の技術者が参加、共同でオープンに開発。
- ・開発者メーリングリストは活発。30～100通/日
- ・Xen 3.2 2008年1月リリース。Xen 3.3 2008年8月リリース。

- Xen開発者向けの技術的な集まり。  
弊社技術者も参加
- 過去6回開催  
2005年4月、2006年1月、2006年9月、2007年4月、2007年11月、  
2008年6月と開催されている。  
2008年6月の資料は  
[http://www.xen.org/xensummit/xensummit\\_summer\\_2008.html](http://www.xen.org/xensummit/xensummit_summer_2008.html)  
から取得可能。
- 第7回、2008年11月20/21  
日本で開催。(幕張)  
[http://www.xen.org/xensummit/xensummit\\_fall\\_2008.html](http://www.xen.org/xensummit/xensummit_fall_2008.html)
- 第8回、2009年2月 (Redwood, CA)

基本機能の開発は完了し、品質も安定。  
スケーラビリティ向上や様々なデバイスのサポートに開発の中心が  
移っている。

ラップトップやモバイル向けにもターゲットを拡大

- ・完全仮想化のサポート範囲拡大
  - x86 リアルモードサポート
  - EFIサポート
  - I/Oエミュレーション性能向上、資源の隔離
- ・大規模システムでの性能向上
  - NUMAサポート
  - I/Oのスケーラビリティ向上
- ・デスクトップシステム向け機能  
Xen Client Initiative
  - グラフィックカード、3D、USBなど
  - セキュリティ
- ・仮想化支援機能を持ったハードのサポート

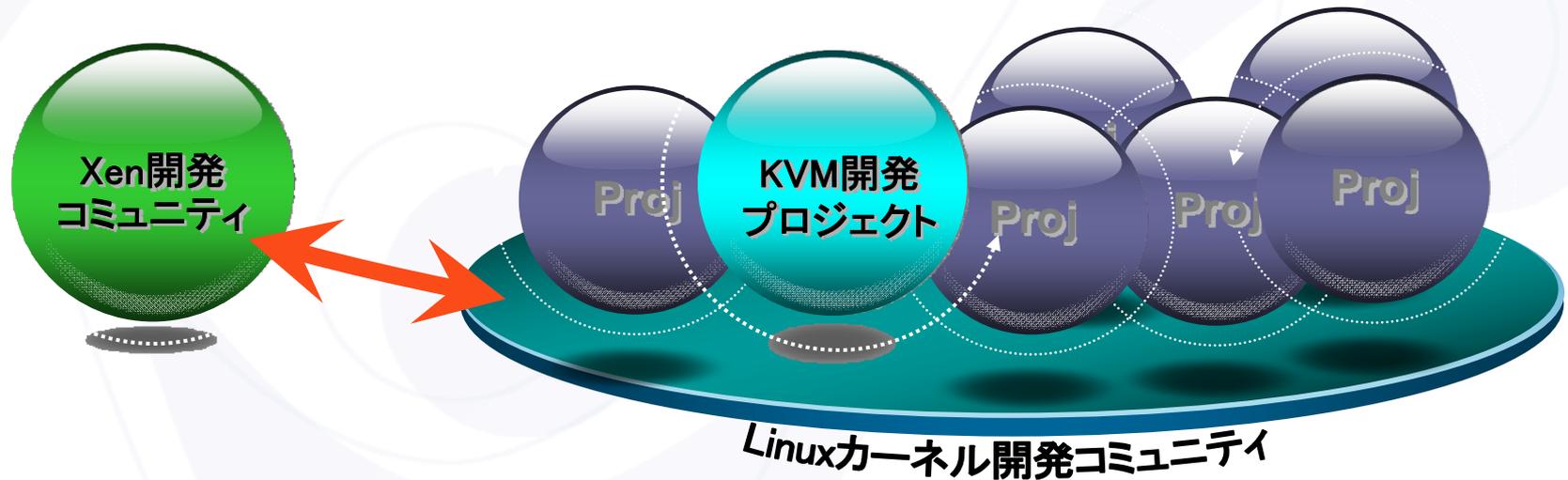
# プロダクションシステムでも十分に使用できる安定性

- ・無償Linuxディストリビューション
  - CentOS
  - Debian
- ・商用Linuxディストリビューション
  - SUSE
  - Windowsとの相互運用性
  - RedHat
  - Ubuntu
- ・商用
  - Citrix Xen Server
  - Windows、管理ツール
  - Virtual Iron
  - Oracle VM
- ・その他(Linux以外)
  - SUN xVM

- QUMRANET社が開発。GPLで公開。2006年～  
<http://kvm.qumranet.com/kvmwiki>  
(QUMRANETは、2008年9月にRed Hat に買収された)
- 標準Linuxカーネルに取り込まれたこともあり、大きなコミュニティを持つことになった。開発速度が非常に速い。  
(kernel 2.6.20 (2006年12月))
- KVM Forum  
開発者会議。過去2回開催(2007年8月、2008年6月)
- 安定化途上の段階  
コードの更新量が多い。  
I/Oの性能改善が課題。  
仮想化支援機能を持ったハードのサポートもやはり大きなテーマ

- Linuxディストリビューションで使用できるようになってきた
  - Fedora、Ubuntu
  - まだ安定性に欠ける。
- Red Hat は重点をXenからKVMに大きくシフトしている
  - 2008年9月、QUMRANETを買収
  - 仮想化システム管理「oVirt」を発表。(KVMの技術を採用)
- Xen vs KVM
  - 今、使用するならXen
  - 機能比較はあまり意味がない
  - オープンソースでエンハンスが速い。

- Xenは、Linuxカーネル開発コミュニティと連携する形で、独自の開発コミュニティを形成。
- KVMはLinuxカーネルの1機能として実装することにより、Linuxカーネル開発コミュニティの1プロジェクトとして活動。



- ・カーネル2.6.23でXenの準仮想化に対応 (domainU)  
(domain0用は、2.6.28 or 29 の予定)
- ・仮想マシン用フレームワーク paravirt\_ops を導入 (2.6.23)
  - 様々なハイパーバイザの準仮想化インタフェースに対応するためのフレームワーク  
vmwareの対応もあり
  - LinuxがNativeハードウェア上でもハイパーバイザ上でも  
同一バイナリで動作
- ・仮想I/Oインタフェース virtio (2.6.25)  
各種ハイパーバイザは共通のドライバを用いてI/Oが行える

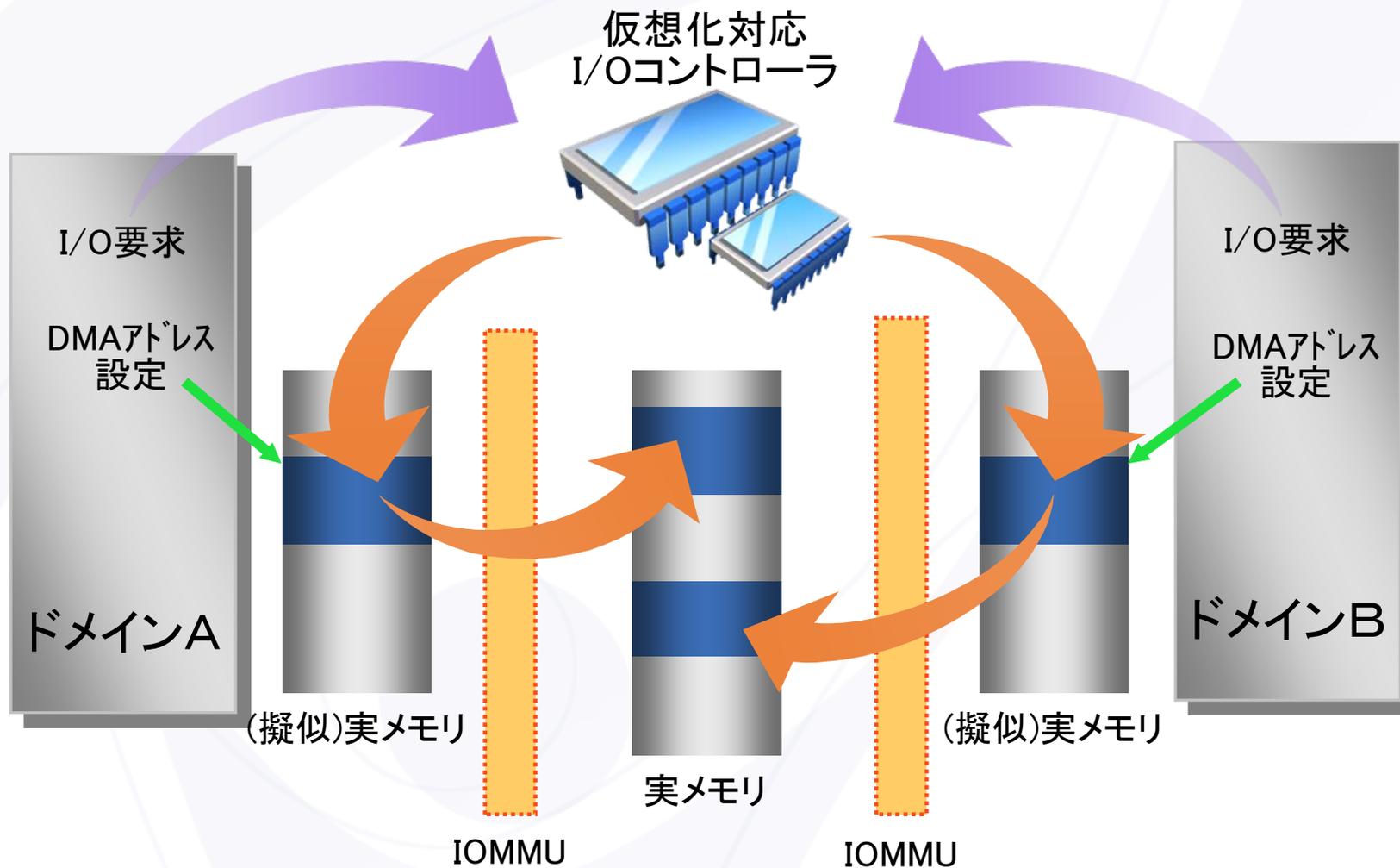
仮想マシン環境を支援する機能がどんどん追加される  
トレンドにある。

### ・CPUの仮想化支援

- Intel VT-x、AMD-V (2006年～)
- VMM用のモード遷移をサポート  
完全仮想化が可能に。
- ゲスト物理アドレスと実物理アドレスの自動変換  
EPT (Intel VT-x)、NPT (AMD-V)

- ・チップセットの仮想化支援
  - Intel VT-d、AMD IOMMU
  - 仮想マシンから安全にDMA転送が可能  
ゲストマシンからPCIデバイスに対する直接I/Oが可能  
(PCI passthrough)
  - 割り込みリマッピング (Intel VT-d)  
ゲストマシンの割り込み処理の効率化

# ハードウェアの動向(3)



- PCI Expressの仮想化支援
  - SR-IOV: PCI-SIGにて仕様策定中
  - デバイスを仮想的に複数のデバイスに見せることが可能
  - Intel VT-c、マルチキューNIC  
対応機器が出だしてきている
- Xen、KVMともハードウェアの仮想化支援機能のサポートは、Intel、AMDが大きな貢献を行っている。
- オープンソース故にまず、新しいハードウェアのリファレンス実装に使用される。  
オープンソースにより、ハードウェアの開発も加速している。

- 最近のOSは準仮想化対応済み
  - 「OSの修正が必要」はもはや問題ではない
  - ハードウェアの仮想化支援機能を利用して、より効率化
- PVドライバ
  - 完全仮想化ゲストでドライバのみを準仮想化モデルに置き換え。I/Oの高速化

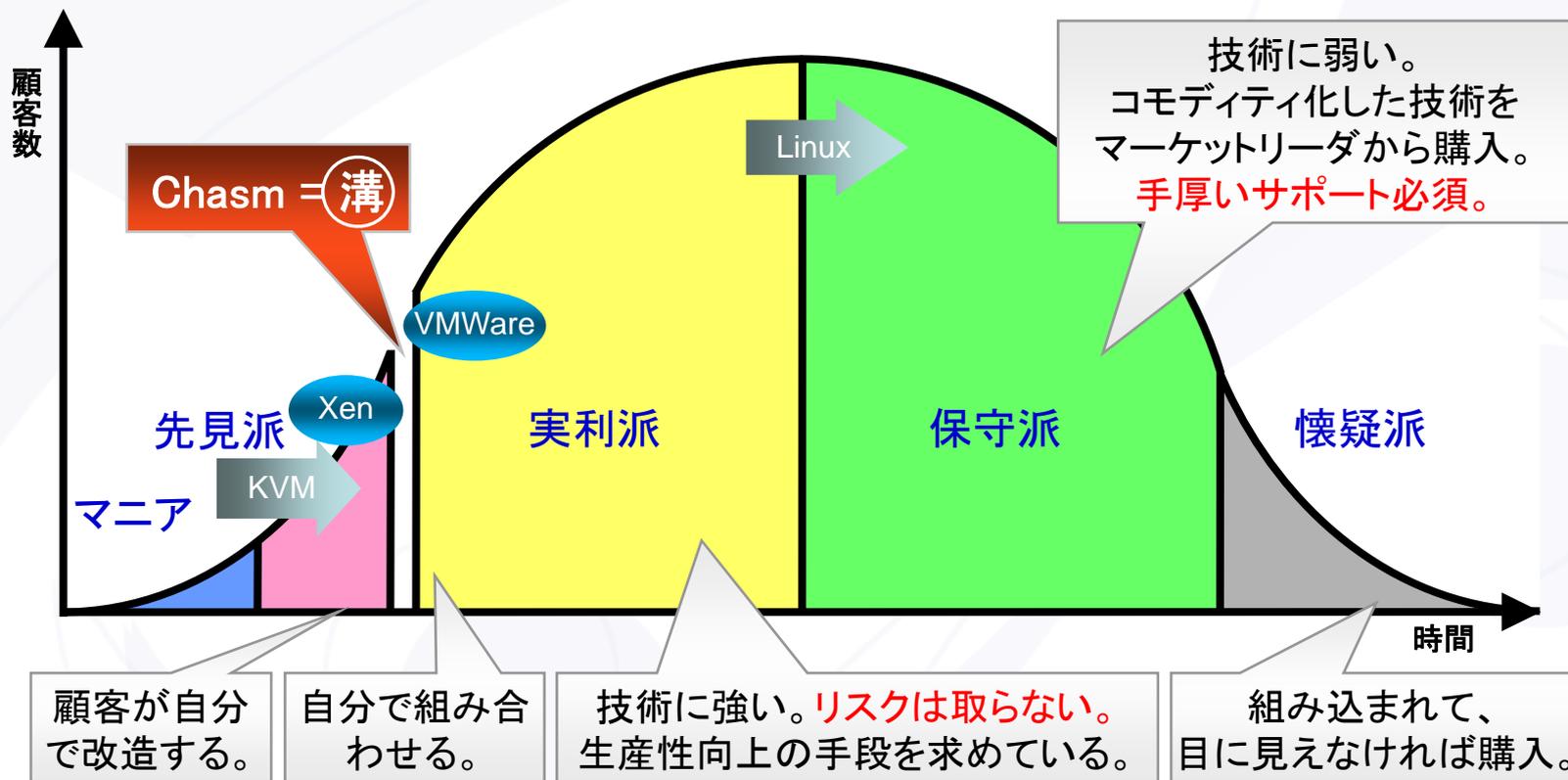
### ➡ ハイブリッド化のトレンド

- 純粋な完全仮想化ゲストは、主にレガシーなOSの動作用

# まとめ

# 仮想マシン環境とライフサイクル

- Linuxは、既にある当たり前のコモディティ化した技術になりつつある。
- PCサーバの世界では、多くの仮想化技術そのものがChasmを渡る直前に集中している。



VMwareはChasmを渡り、急激に普及が始まった。XenはChasmを渡る直前にいる。KVMは開発初期段階だが、Linuxの大きなコミュニティの力により推進力が大きい。KVMもXenに迫ってきている。

※本書掲載内容の複写・無断転載を禁じます。

Copyright © VA Linux Systems Japan. All rights reserved.

- 本書は2008年10月7日現在の情報に基づいて作成しております。
- VALINUXはVA Linux Systems Japan株式会社の日本国内における商標です。
- LinuxはLinus Torvalds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。
- その他記載されている会社名、製品名は一般に各社の商標または登録商標です。

# END



VA Linux Systems Japan株式会社

〒104-0053 東京都中央区晴海1-8-8

晴海アイランドトリトンスクエア オフィスタワー W棟16F

TEL:03-6204-4071 FAX:03-6204-4072

sales@valinux.co.jp

<http://www.valinux.co.jp/>