

- ▼ ミッションクリティカルなシステムに求められるもの
- ▼ インテル Itanium 2 プロセッサ
- ▼ インテルバーチャライゼーションテクノロジー
- ▼ レッドハットEL5での新たな機能
 - Virtualization, カーネル強化など
- ▼ XEN, KDUMP, KEXECのご紹介
- ▼ デモ

ミッションクリティカルなシステム



▼ 24時間365日、止まらないことを要求される基幹業務、あるいは、そのような業務遂行のために使用されるコンピュータシステムのこと。企業の経理などの金銭に関わる業務や、電子商取引などを支える基幹システムに誤りや中断、セキュリティ上の問題が発生すると、業務の中断だけでなく巨額の損失の発生や信用の失墜を招く危険性がある。このため、このような業務に使用されるシステムには、極めて高い信頼性や耐障害性、障害発生時に被害を最小に食い止める様々な機能、万全のサポート体制などが必要となる。このような性質をミッションクリティカルと呼ぶ。 IT用語辞典 e-wordsより

▼ その存在が、任務や業務の遂行にとって必要不可欠であることを言う。ミッション・クリティカルであることは、24時間365日、正常に機能し続けなければならないことを意味し、社会の基盤システムや企業の基幹システムが備えるべきものとして求められる。障害の発生による中断や停止が社会に多大な影響を及ぼすシステムとしては、金融機関や交通機関のオンライン・システムをあげることができるが、極めて高い信頼性が必要とされるこれらはミッション・クリティカル・システムと呼ばれる。

Wisdom より

ミッションクリティカルなシステム



❖ 高い可用性のためにはダウンタイムを極端に少なくしなければいけない

- ダウンしないシステム
- ダウンしても素早く復旧(立ち上がる)する

可用性

1年間のダウンタイム

90%	876 時間 (36.5 日)
95%	438 時間 (18.25 日)
99%	87.6 時間 (3.65 日)
99.90%	8.76 時間
99.99%	52.56 分
99.999% (ファイブ ナイン)	5.256 分
99.9999% (シックス ナイン)	31.536 秒

Linux OSのミッションクリティカルシステムをサポートするための機能拡張

- 高負荷な処理に対応できるカーネル
 - カーネルの改良(プリエンプティブ、スケジューラ)
- システムがダウンしたときに素早く立ち上がるための手段
 - KEXEC
- ダンプ機能、問題発生時のデバッグ手段
 - KDUMP、ダンプ解析ツール
- 性能向上のためのツール
 - SystemTAP, Frysk

カーネルコミュニティとの協調



▼ リナックスの歴史

▼ リナックス開発のしくみ

▼ エンタープライズ向けカーネルの強化

- OSDLを中心にコミュニティと企業との連携

カーネルコミュニティとの協調



📖 始まり Linusのメール (1991)

```
From: torvalds@klaava.Helsinki.FI (Linus Benedict Torvalds)
Newsgroups: comp.os.minix
Subject: Gcc-1.40 and a posix-question
Message-ID: <1991Jul3.100050.9886@klaava.Helsinki.FI>
Date: 3 Jul 91 10:00:50 GMT
Hello netlanders,
Due to a project I'm working on (in minix), I'm interested in the posi
standard definition. Could somebody please point me to a (preferably)
machine-readable format of the latest posix rules? Ftp-sites would be
nice.
```

📖 2年後12000人のユーザー

📖 現在Linux, Andrewのメインメンテナー及び数十人のサブシステムメンテナー、世界中の数万の開発者によって日々開発が進められている。

カーネルコミュニティとの協調



MINIXを使っている皆さんこんにちは

いま、386(486)AT互換機用の(フリーな)オペレーティングシステムをやっています。(ただの趣味で、GNUのような大きくて本格的なものにはならないと思いますが。)4月からやっていて、もうすぐ出来ます。みなさんがMINIXのどこが好きか、どこが嫌いかを知りたいのです、私のオペレーティングシステムはMINIXにいくらか似ていますので。現実的な理由からファイル・システムの物理的な配置は同じです。

bash(1.08)とgcc(1.40)が動いています。数ヶ月したら、いくらか実用的なものが出来ると思うので、どんな機能が欲しいかを知りたいのです。どんな意見も歓迎ですが、それを実現するかどうかは約束できません。

Linus

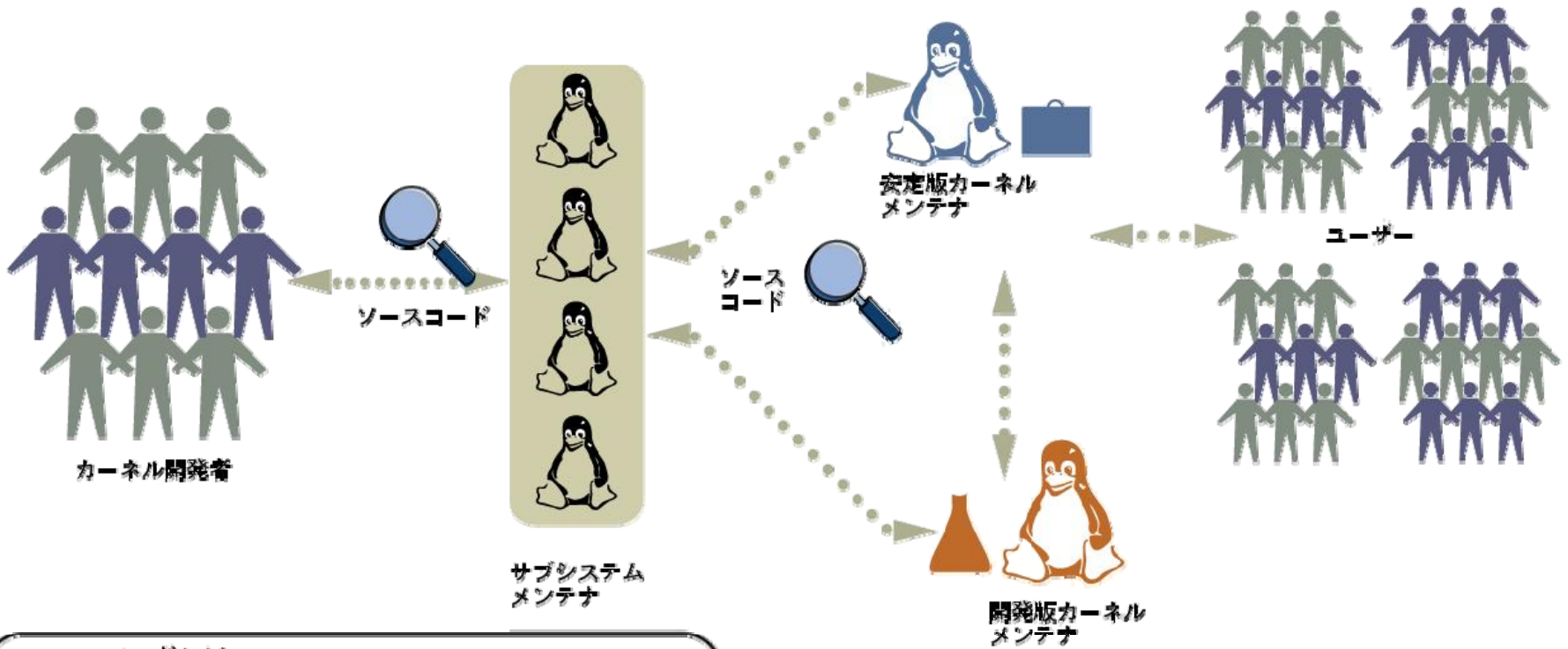
PS: そう、MINIXのどんなコードも入っていません。そして、マルチスレッドのファイル・システムがあります。386のタスクスイッチなどを使っているので移植は出来ません。AT互換のハードディスクしかサポートしないと思います。私はそれしか持っていないので。

カーネルコミュニティとの協調



開発コミュニティ

Linuxカーネル 開発コミュニティ



コードレビュー
各開発者によって作成されたコードは、他の開発者によって評価され、Linuxカーネルに採用される。この作業は、主にネットワーク上の公開された場所（メーリングリスト）で行われている。

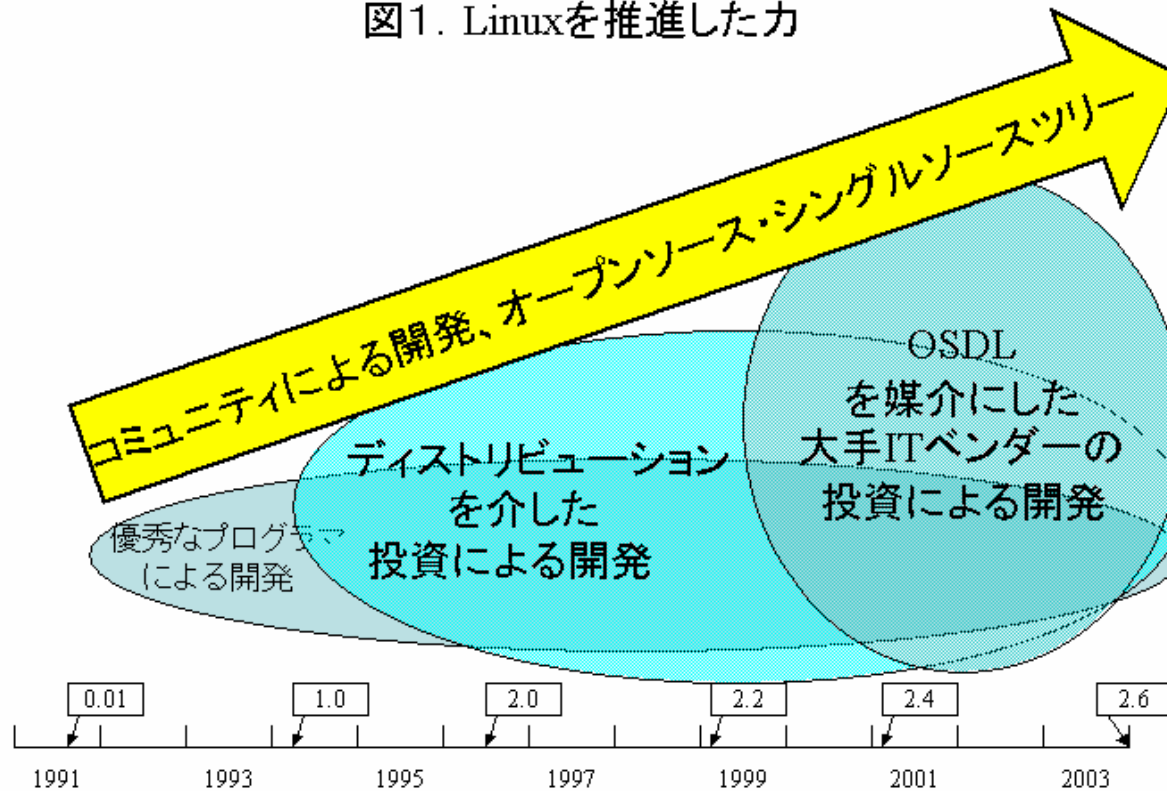
OSDLの資料より

カーネルコミュニティとの協調



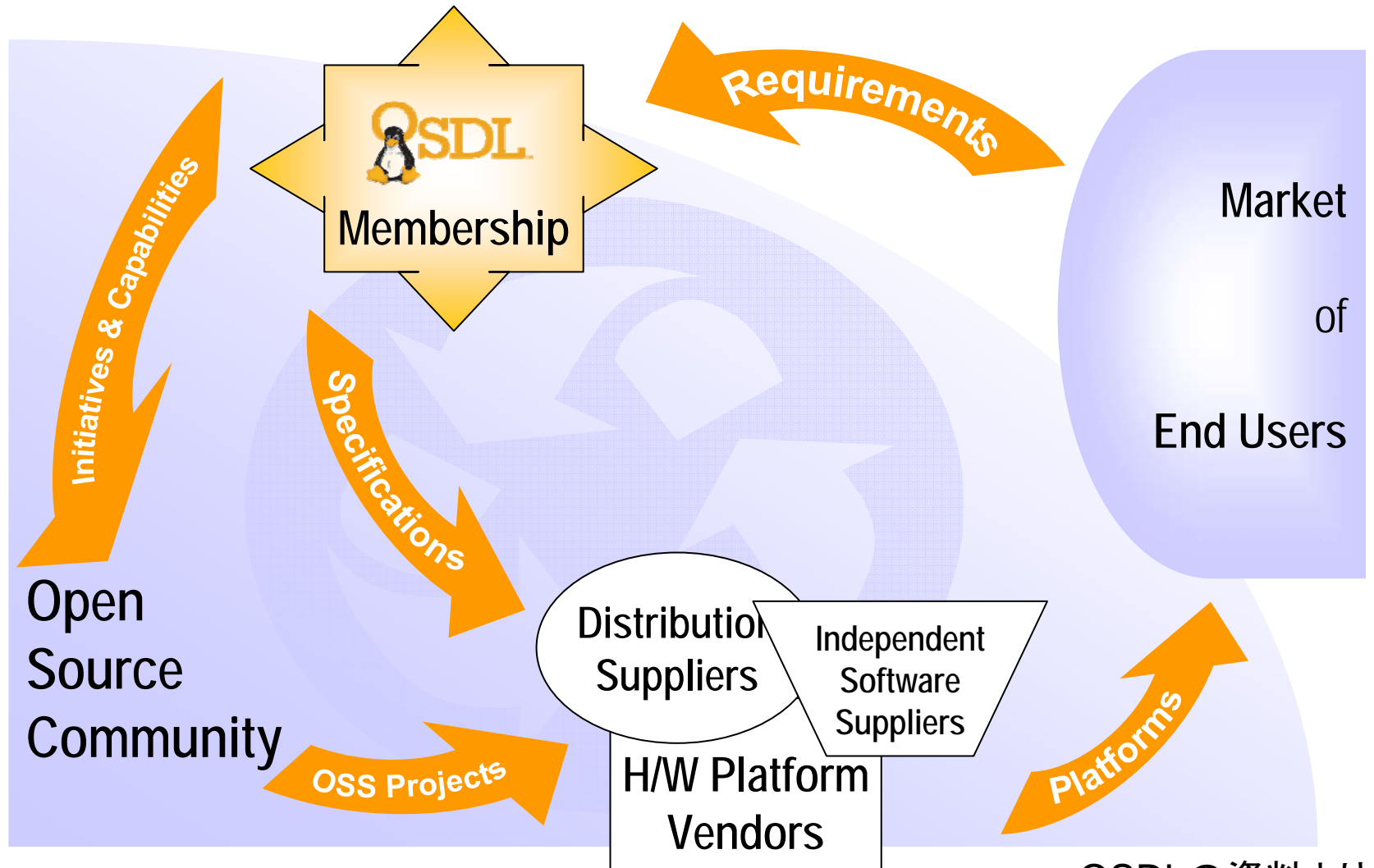
OSDL(現The Linux foundation)の効果

図1. Linuxを推進した力



(財)武田計測先端知財団 Linux開発経緯に関する調査より

カーネルコミュニティとの強調



OSDLの資料より

大規模システム向けカーネルパッチ

- 高速ブート: *kexec ...*
 - ロック機構の改善: *robust mutex ...*
 - I/Oサブシステム: *async I/O, multipath I/O ...*
 - ネットワークスタック: *netem & bridging ...*
 - セキュリティ: *LSM (Linux Security Module) ...*
- 大規模システム向けプロジェクト
 - クラスタリング: *TIPC (Transparent InterProcess Communication) ...*
 - 標準化: *OpenAIS (Application Interface Specification) ...*
- 信頼性強化のテスト環境/プロジェクト
 - *PLM/STP*
 - *2.6安定化プロジェクト*
 - *Bugzillaの運営*

現在のOSDL (The Linux Foundation)



OSDL + FSG (Free Standard Group)

Linuxの普及促進、標準化の活動

The screenshot shows the homepage of The Linux Foundation. At the top left is the logo "THE LINUX FOUNDATION". To the right is a search bar and a "Log in / create a" link. Below the logo is a navigation menu with items: "About", "Join", "Linux Standard Base", "Workgroups", "Linux Protection", "Collaboration Forum", and "Developer Services". The main content area is divided into three columns. The first column, "The Linux Foundation", describes the organization's mission and includes a link to "More on the Linux Foundation". The second column, "Standardization & Collaboration", describes programs for promoting standardization and technical collaboration, with links to "More on Standardization" and "More on Collaboration". The third column, "Protection & Promotion", describes services for protecting the future of Linux, with a link to "More on Linux Protection". Below these columns are three sections: "Announcements" with a list of recent news items and a link to "More Announcements"; "News" with a list of news items and a link to "More News"; and "Blogs" with a list of blog posts and a link to "More Blogs". At the bottom of the page, there are links for "[Article]", "[Discussion]", "[View source]", and "[History]".

Copyright © 2007 Linux Foundation. All rights reserved.
LSB is a trademark of the Linux Foundation. Linux is a registered trademark of Linus Torvalds.
Please see our privacy policy.

レッドハットEL5での新たな機能例



 Product structure

 yum

 Virtualization

 Kernel 2.6.18

 KEXEC/KDUMP

 SystemTAP/Frysk

 Security

 JBOSS



❖ ミッションクリティカルなシステムには

- 高い可用性を実現するための堅牢なシステムが要求される
- 仮に問題があった場合でも速やかにリカバリーし、原因を究明できるしかけが必要である。

❖ これらを目指すためにLinuxコミュニティおよびインテル、レッドハットを含めた企業コミュニティの数年にわたる改良を加えてきた

- これらの1つの成果がRed Hat Enterprise Linux 5である

Itaniumアーキテクチャー概要

- EPIC

- 明示的並列計算

- レジスタセット

- 128個の64ビットの汎用レジスタ

- 128個の82ビットの浮動小数点レジスタ

- レジスタースタック

- 命令グループ

- IA32のサポート

インテル Itanium 2 プロセッサ



Mckinley (2002年)

- 第一世代のItanium2
- 900MHz-1GHz

Madison (2003年)

- 第2世代のItanium2
- 130nmプロセス
- L3キャッシュ 最大9MB

Montecito (2006年)

- デュアルコア
- バーチャライゼーションテクノロジー
- ハイパースレッディングテクノロジー

インテル Itanium 2 プロセッサ



Itanium 2 プロセッサ

- EPIC
- デュアルプロセッシング
- ハイパースレッディングテクノロジー
- 24MBのL3キャッシュ(デュアルコア)
- バーチャライゼーションテクノロジー
- キャッシュセーフテクノロジー
- セキュリティー機能
- IA32アプリケーションのサポート

信頼性

- キャッシュセーフテクノロジー
 - ➔ キャッシュエラーからの回復機能
 - 改良型MCA(マシーンチェックアーキテクチャー)
 - ➔ バスデータエラーの自動検出、記録、訂正
 - ホットプラグプラットフォームコンポーネント
 - ➔ 信頼性、管理性、保守性の向上
- システム稼働時間の向上

インテル Itanium 2 プロセッサ



RAS機能の比較

RAS 機能 ^a	インテル® Itanium® 2 プロセッサ搭載システム	メインフレーム	RISCシステム
キャッシュの信頼性	✓(インテル® キャッシュ・セーフ・テクノロジー—新機能 ^b)	✓	✓
プロセッサ・ロックステップのサポート	✓新機能 ^b	✓	
データ・バス・エラーの回復	✓	✓	✓
キャッシュ ECC カバレッジ	✓	✓	✓
改良型マシン・チェック・アーキテクチャ (MCA)	✓		
不良データの隔離	✓	✓	✓
メモリー・シングル・デバイス・エラー・コレクション	✓	✓	✓
ダブルビット・エラー検出時のメモリーリトライ	✓	✓	✓
メモリー・スペアリング	✓	✓	✓
ハードウェア・パーティショニング	✓ノード	✓コア	✓ノード
電氣的に隔離されたパーティション	✓ノード	✓	✓ノード

a 記載されている機能は、1社以上のシステムベンダーによってサポートされています。

b 最新のデュアルコア インテル® Itanium® 2 プロセッサをベースにした一部のシステムで利用できます。

 Red Hat Enterprise Linux 5

 +

 インテルItanium2 プロセッサ

 ミッションクリティカルなシステム構築への基盤

ガートナーが見る、2006年の戦略的技術トップ10



1. バーチャリゼーション(仮想化)
2. グリッドコンピューティング
3. サービスとしてのソフトウェア(SAS)
4. パーベイシブコンピューティング
5. 有機発光ダイオード(OLED) & 発光ポリマー(LEP)ディスプレイ
6. 位置認識サービス
7. ミッションクリティカル向けLinux
8. インスタントメッセージ(IM)
9. 情報アクセス
10. 少額電子商取引

出典:

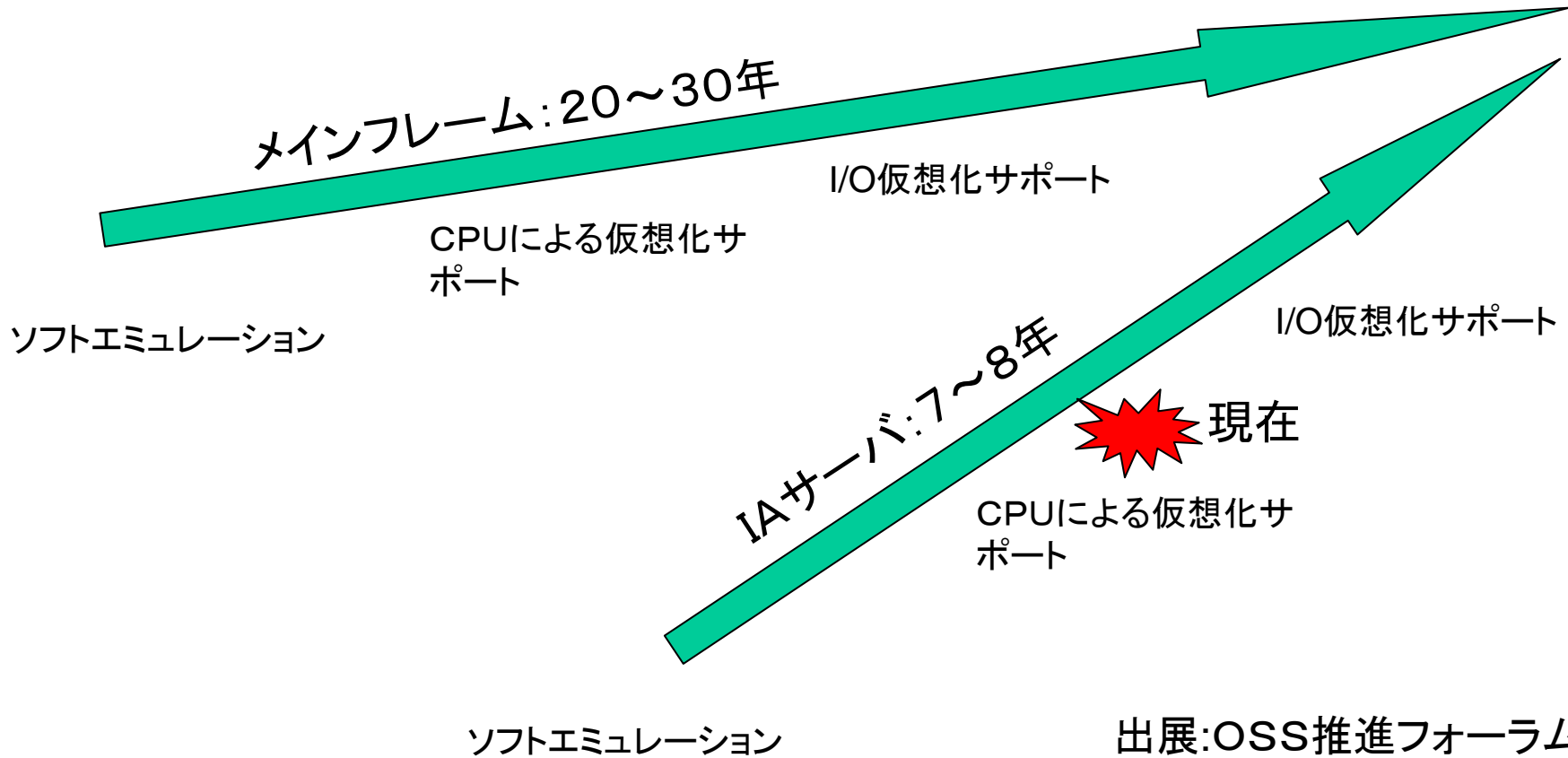
米Gartnerバイスプレジデント兼ガートナーフェローのBob Hayward氏

2005年12月2日、同社主催のGartner Symposium/ITxpo2005


<http://japan.zdnet.com/news/itm/story/0,2000052525,20092095,00.htm>

- ▼ 従業員500人以上の企業では4分の3以上がバーチャルサーバを導入している
- ▼ 調査でサーバ仮想化技術を使っていると回答者は、来年購入する新サーバのうち45%を仮想化する見通し
- ▼ 全仮想サーバのうち50%以上で、基幹業務を含むプロダクションレベルのアプリケーションを実行している
- ▼ サーバの作業負荷の種類に応じた仮想サーバのパフォーマンス最適化には大きなチャンスがある
- ▼ 現在の仮想サーバへの出費はS390、OS400、UNIXシステムが主流だが、WindowsとLinuxサーバも急増している

メインフレームのVM機能を、IAサーバーが急速にキャッチアップ



出展:OSS推進フォーラム サーバ部会 仮想化技術の最新動向
向:実運用でのメリットと課題より

 上位層のリソースが下位の仮想化されたリソースを利用することで実現

- ITサービス ITサービスレベルの仮想化
 - ➡ SOA, Webサービス
- アプリケーション アプリケーションレベルの仮想化
 - ➡ グリッド技術、分散処理
- システムシステムレベルの仮想化
 - ➡ ワークロード管理、クラスタ技術
- サーバー サーバーレベルの仮想化
 - ➡ 物理パーティション、論理パーティション、リソース・パーティション
- ストレージ・リソース ストレージの仮想化
 - ➡ 仮想ストレージ技術

IT管理者のための仮想化技術
入門 アットマークIT

物理パーティション

- pPAR, nPAR

論理パーティション

- vPAR, LPAR

仮想マシン

- VMware, XEN VMM(仮想マシンモニタ)

仮想OS

- UML

ホスティング

- Solaris Container

仮想化による利点



◆ 柔軟なサーバ・コンソリデーション:

- サーバ仮想化によって、さまざまなオペレーティング・システムやアプリケーションを短時間で簡単に 2-way ~ 16-way 以上のプラットフォームに統合できる

◆ 可用性とセキュリティの向上:

- ソフトウェア障害やデジタル攻撃を仮想パーティションに隔離したり、フェイルオーバー・パーティションを設置して簡単かつ経済的にニーズに合った可用性を実現できる

◆ テストおよび開発環境の合理化:

- 単一のプラットフォームでソフトウェア・スタックごとに複数のテスト環境をホスティングし、繰り返し利用できる

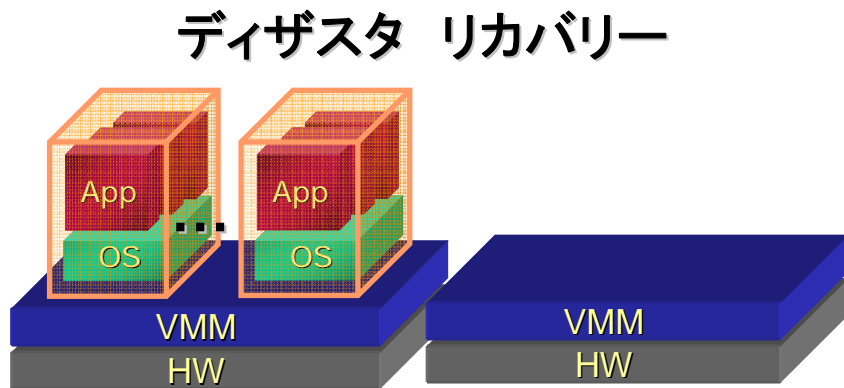
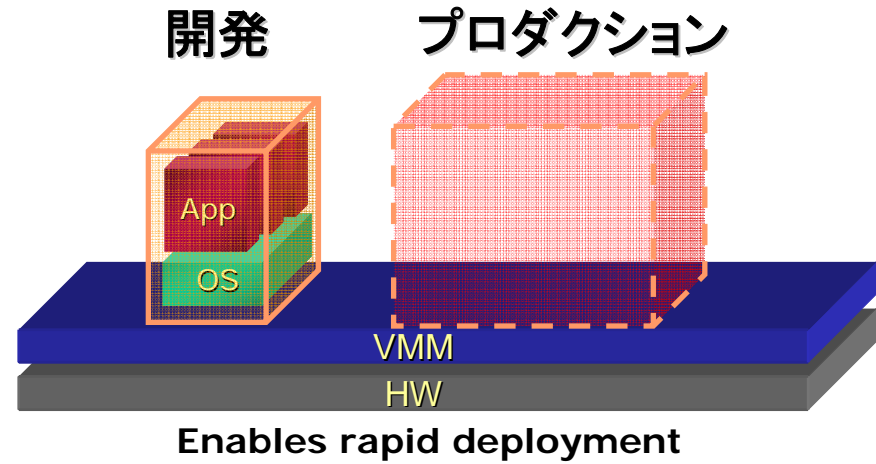
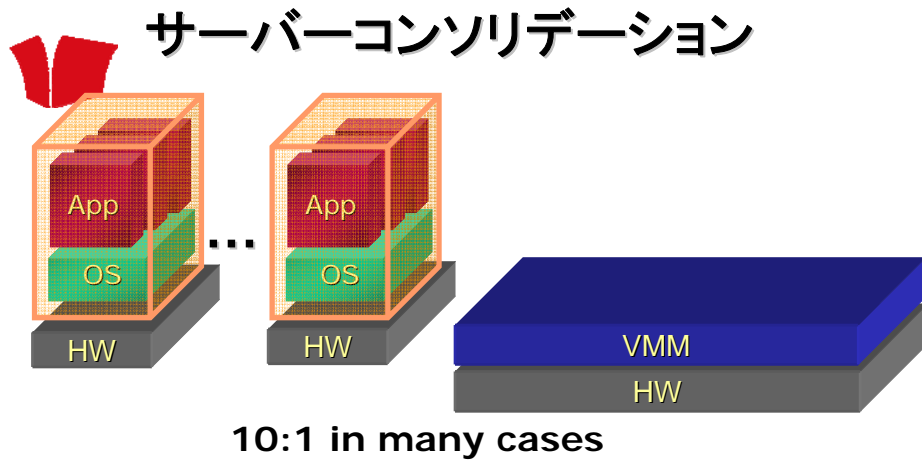
◆ OS およびハードウェア移行の簡略化:

- サーバ仮想化によって、レガシ・アプリケーションや既存 OS のバージョンを変更せずに仮想パーティションに移行できる

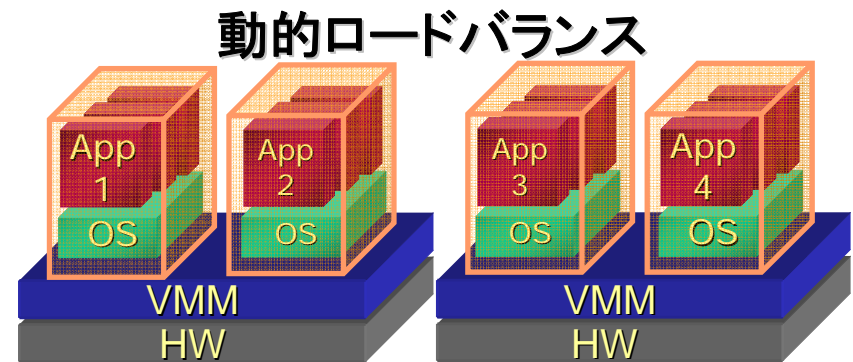
➤ ビジネスの機敏性の向上:


- 仮想パーティションのプロビジョニングやサイズ変更を簡単に行って、新しいアプリケーションやワークロードの増大、システム保守に対応できる

バーチャライゼーションの使用例



 Upholding high-levels of business continuity

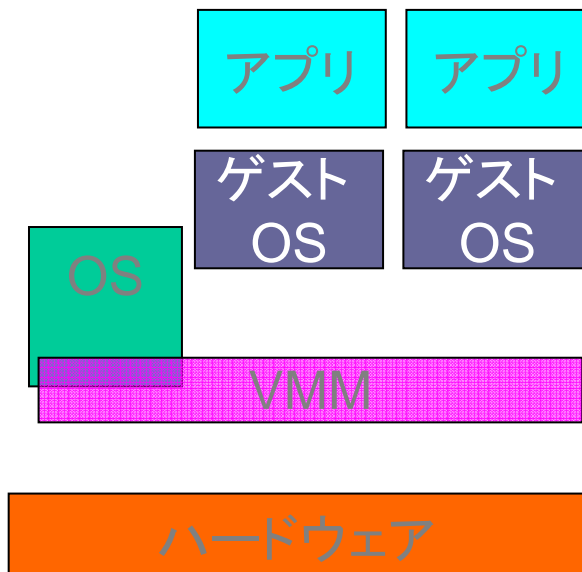


 Balancing utilization with head room

VMM(仮想マシンモニタ) ソフトウェア・アーキテクチャの種類



VMware Workstation
VMWare GSX Server




VMware ESX Server

Microsoft Virtual PC
Microsoft Virtual Server



Xen

 英国ケンブリッジ大学で研究・開発されたオープンソース

 2種類の仮想化

- パラバーチャライゼーション

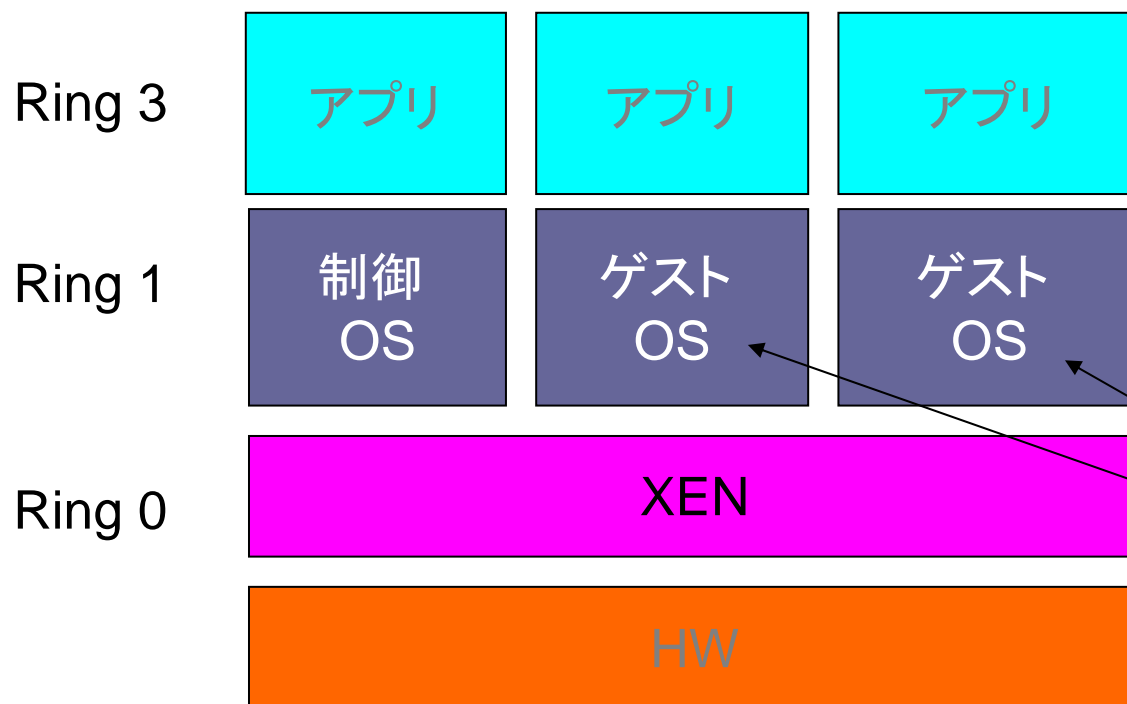
 - ➡ ゲストOSをXen用にモディファイしなければならない

- フルバーチャライゼーション

 - ➡ ゲストOSのモディファイは必要なし

 - ➡ ただしインテルバーチャライゼーションテクノロジーなどのCPUでの仮想化の機能が必要

パラバーチャライゼーション

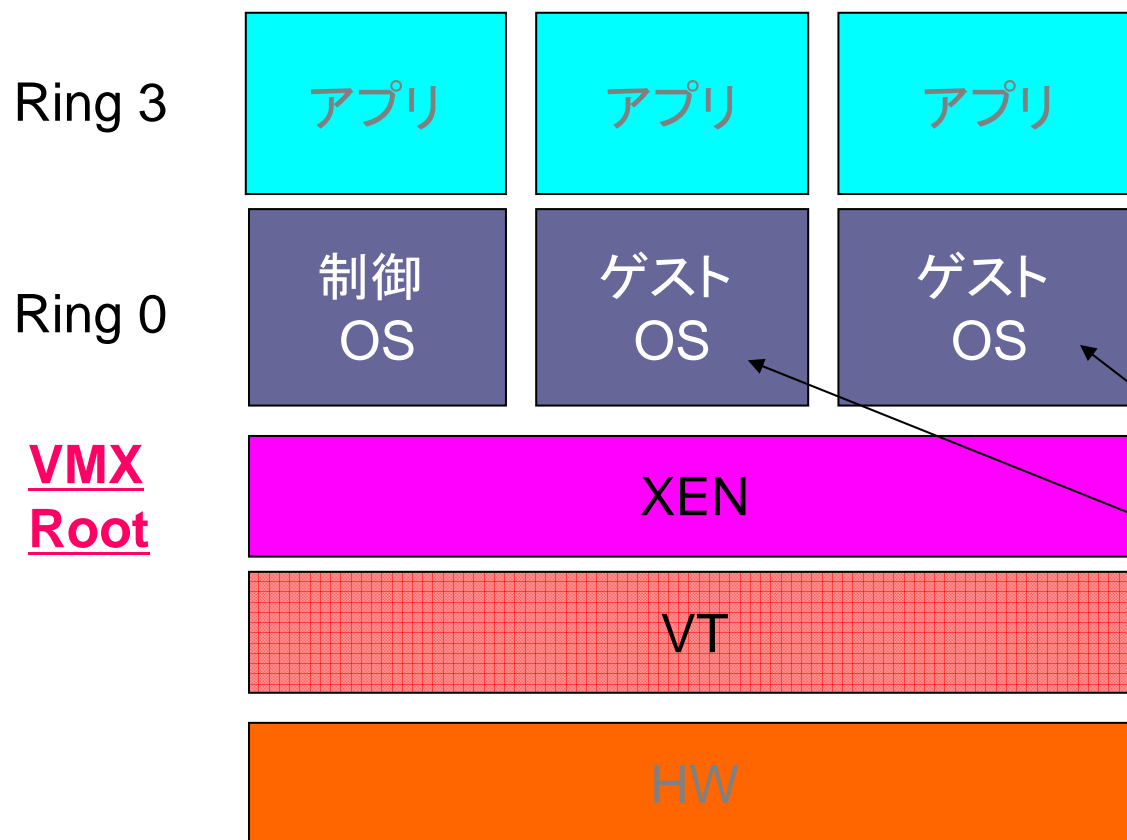


Xen対応OSが必要
になってくる。

例: Red Hat
Enterprise Linux 5

FedoreCore 6など

フルバーチャライゼーション



完全な仮想化HWを実現しているので
Xenに対応していないOSもインストール可能

例: WindowsXP,
RedHat9

- ❖ Xen+ インテルバーチャライゼーションテクノロジーでLinuxOS上でのゲストOSの選択幅がひろくなる
 - たとえば
 - Linux上でWindows OSを稼動
 - 古いOSを最新のシステムで稼動させる

実機での操作

- CPU Core2Duo

- ➡ インテルバーチャライゼーションテクノロジー

- フルバーチャライゼーション

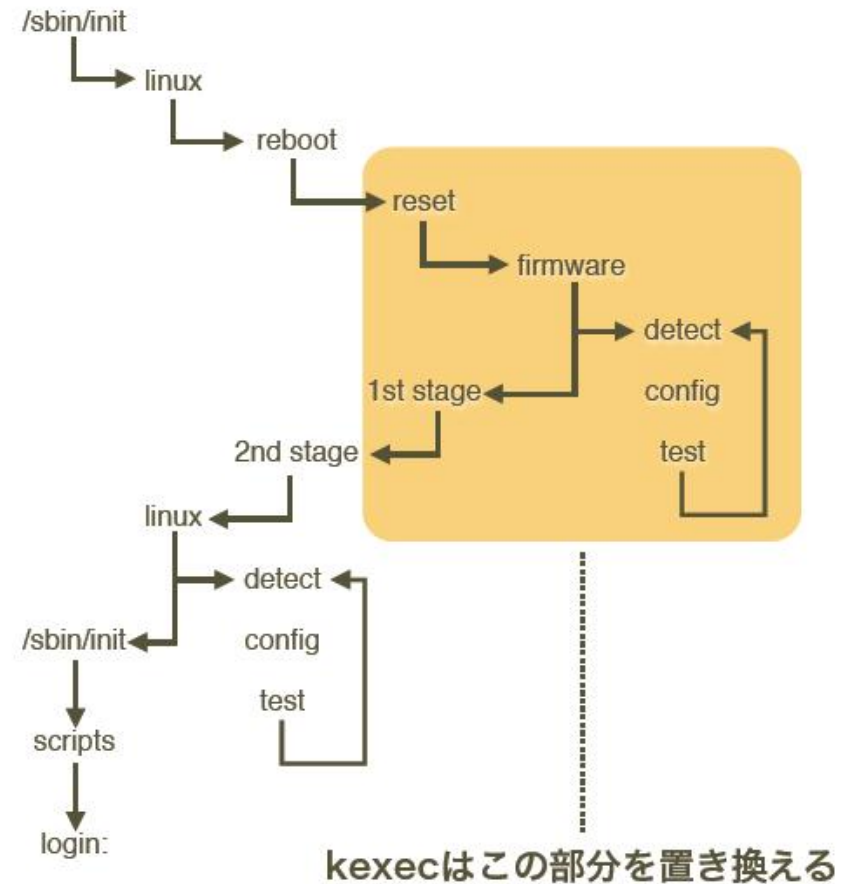
- ➡ Xen用に修正していないOSを起動

KEXECの紹介



KEXEC

- OSの起動を早める仕組み
- 起動時にBISOをバイパスする



参考資料: "Reducing System Reboot Time With kexec"

Linuxでのブート

- Shutdown
- Reset
- Pre-OS Platform Configuration (Firmware)
- Firmware to OS handoff
- 2nd stage Boot strap
- OS device Detection
- OS specific Initialization
-

ブート時間の計測

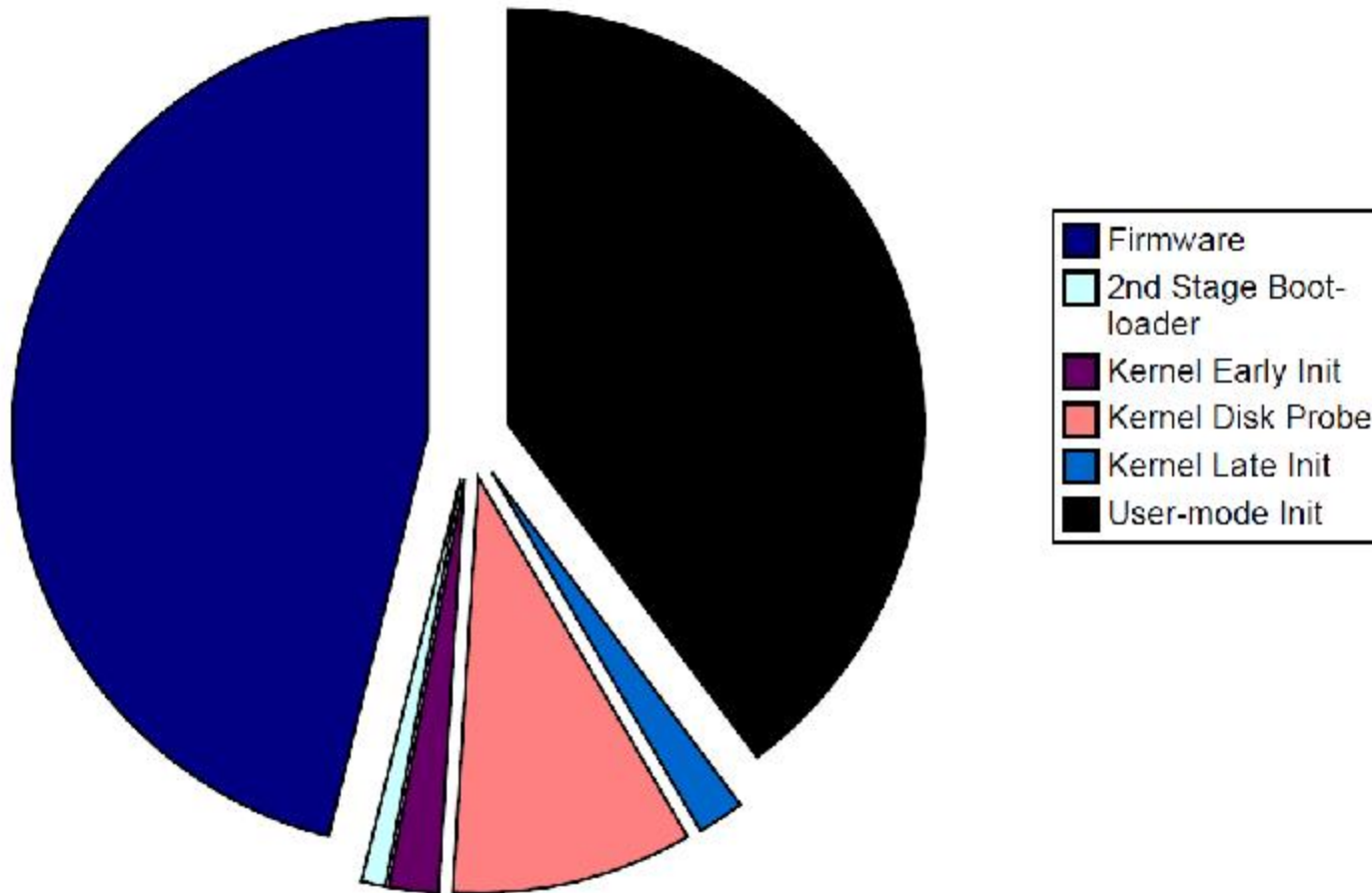


Illustration 1. Proportion of time spent during a typical reboot.

ブート時間の計測(KEXEC使用時)

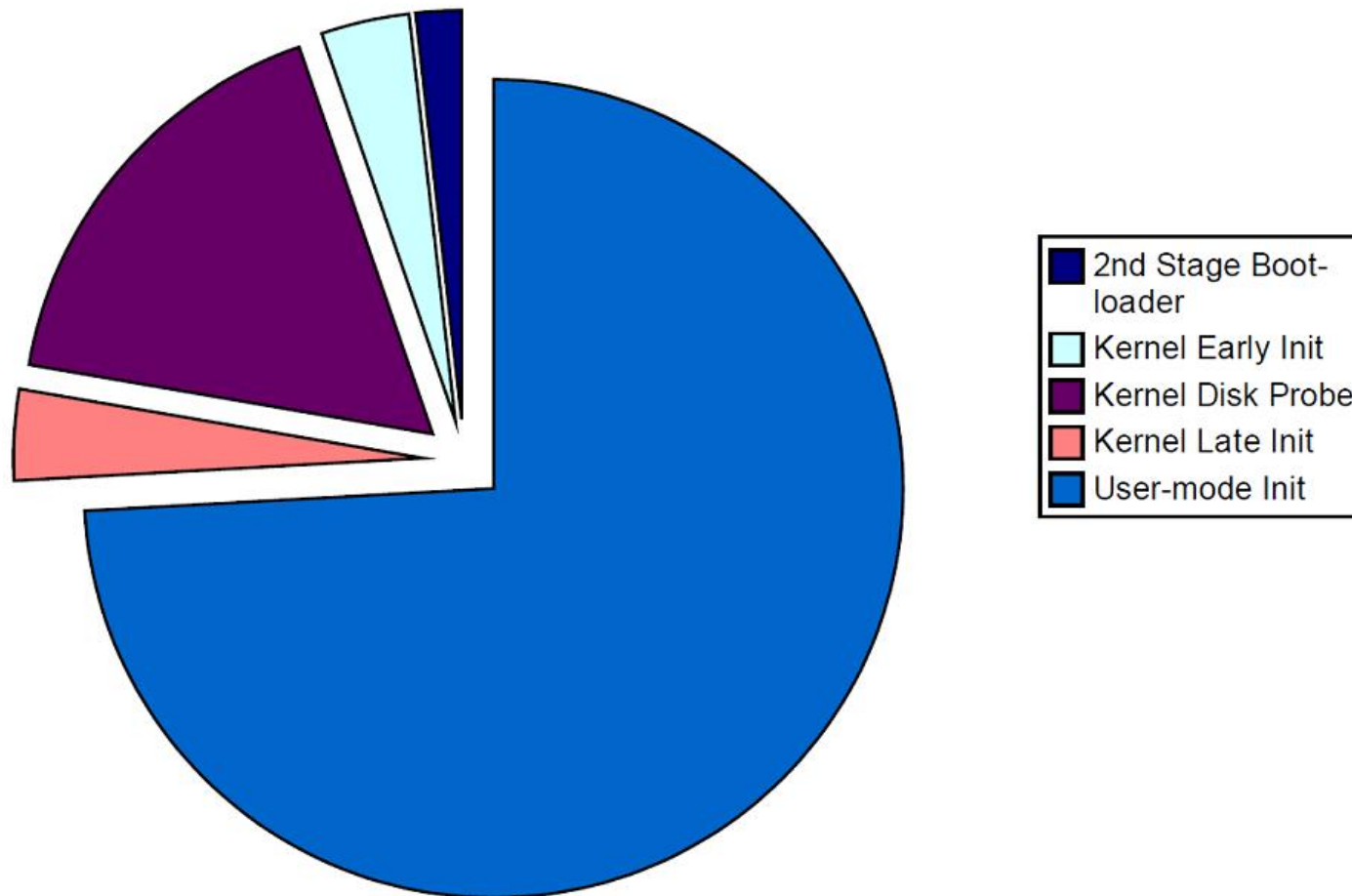


Illustration 4. Proportion of time spent during a kexec reboot.

KEXECの効果

- System 1 PentiumIII
- System 2 Xeon
- System 3 PentiumIII

<i>System</i>	<i>Reboot Time (seconds)</i>	<i>Kexec Time (seconds)</i>	<i>Time Saved (seconds)</i>	<i>Relative Percentage</i>	<i>Kernel Boot Time (seconds)</i>
System 1	112	63	49	56.25%	23
System 2	88	60	28	68.18%	16
System 3	226	56	170	24.78%	15
System 4	NEED DATA	N/A	N/A	N/A	NEED DATA

Table 1. Timing results for the entire reboot operation.

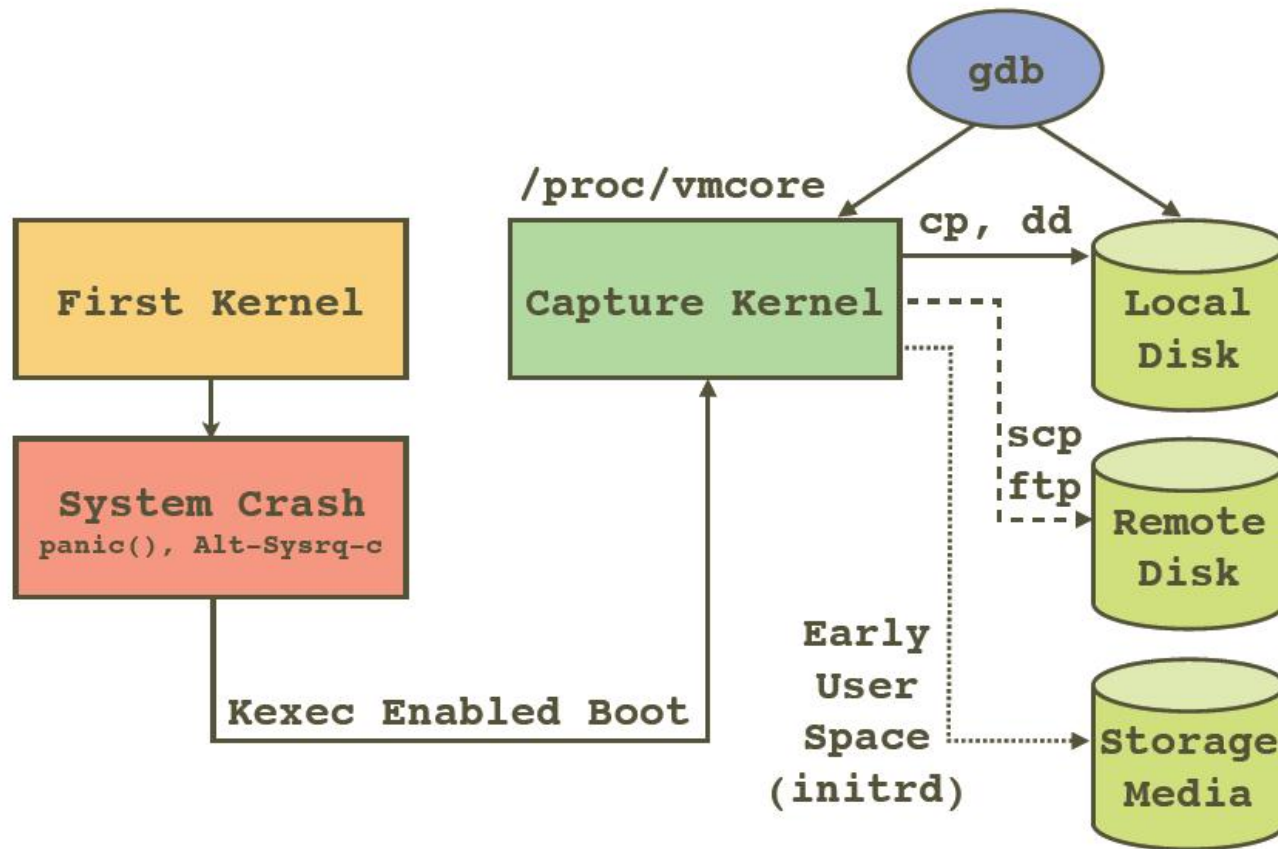
KEXECの紹介

- OS稼動時にカーネルイメージ、initrdをメモリー上にロード
- 何かのトリガーをきっかけにそのカーネルに制御を移行



Illustration 3. Simplified boot time memory layout.

KDUMPの紹介



参考資料: "Kdump - A Kexec Based Kernel Crash Dumping Mechanism"

- ▼ システムがパニックになったときにKEXECによりリブート
- ▼ あらかじめ読み込んでおいたカーネル(Capture kernel)が起動、ダンプデータを取り再起動
- ▼ /var/vmcoreにダンプデータが自動的に保存される
- ▼ Crashなどで解析を行う。

KDUMPの紹介



 OSのインストール時に
KDUMPの設定ができる



The screenshot shows a window titled "カーネルダンプの設定" (Kernel Dump Settings). It contains the following elements:

- A checked checkbox labeled "kdump を有効にする" (Enable kdump).
- System memory information: "システムメモリの合計 (MB): 519", "kdump メモリ (MB): 128" (with a spinner control), and "使用可能なメモリ (MB): 391".
- A text field for "場所:" (Location) containing "file:///var/crash", with a "場所の編集" (Edit Location) button to its right.
- A dropdown menu for "デフォルトの動作:" (Default Action) set to "rootfs をマウントして /sbin/init を実行" (Mount rootfs and execute /sbin/init).
- A text field for "コアコレクタ:" (Core Collector) containing "makedumpfile -c".
- An empty text field for "パス:" (Path).
- Buttons for "キャンセル(C)" (Cancel) and "OK(O)" (OK).

Red Hat Enterprise Linux 5

- 改良されたカーネル
- Xen, KEXEC, KDUMPなどの新しいツール類

インテル Itanium2 プロセッサー

- RAS機能
- バーチャライゼーションテクノロジー

ミッションクリティカルシステムに最適な組み合わせ

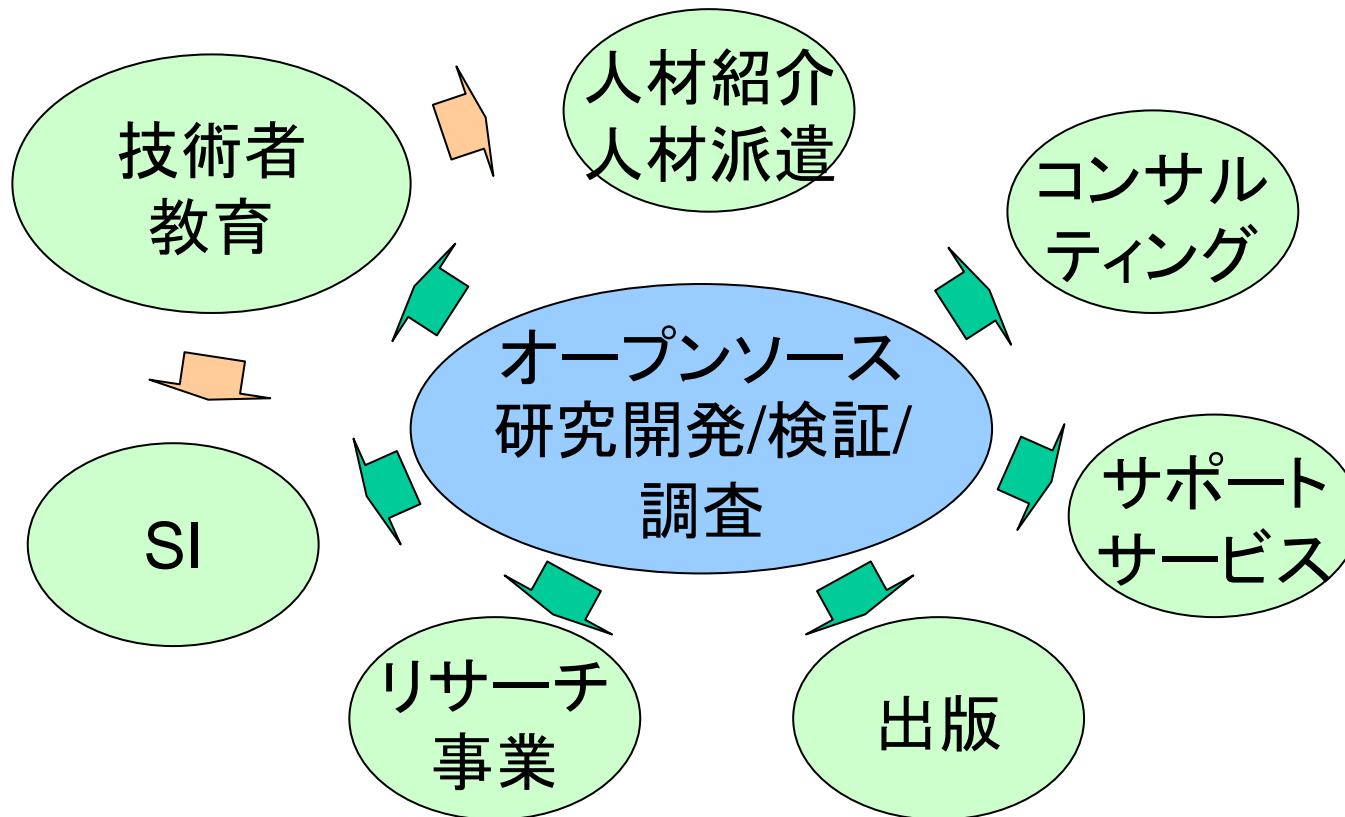
(株)オーブンドリームの紹介



📅 設立 平成19年1月15日

👤 創業者 荒谷 浩二(IPA研究員)、小菌井 康志

📌 オープンソース研究開発および検証などを通し、そこから発生するノウハウを教育/コンサルティング/SI/サポートサービス/出版事業などを通じ供給していく。



(株)オープンドリームの紹介



▼たとえば下記の教育コースを準備中

- Linux上での障害解析、仮想化
- 組み込みLinux関連
- Ruby on Rails, Strutsなどのフレームワーク
- プロジェクトマネジメント

▼ Yasushi.osonoi@opendream.co.jp

▼ありがとうございました