

Kernel2.6の機能にみる エンタープライズでの可能性

VA Linux Systems ジャパン 株式会社

高橋 浩和

Email:taka@valinux.co.jp

目次

1. Linux市場(2002年,2003年)
2. Linux Kernel2.6の機能
3. Linux 2.6のスケールビリティの実装例
4. VA Linux Systems Japanの取り組み

Linux市場(2001年,2002年)

2001年～2002年 Linux の動向について

ようやく利用されはじめたLinux(主な事例)

企業名	システムの概要
太平洋セメント	販売、物流管理システム
宝酒造グループ	管理会計システム
立花証券	オンライン証券システム
日本情報産業	顧客の物流システム
日本医師会	医療機関向けレセプト(診療報酬明細)システム
北陸朝日放送	食料品・工芸品のECサイト
伊勢丹	ECサイト
OBCインターネットサービス	グループウェア、会計、給与のASPサービス「落着オフィス」
麒麟シティ	POSシステム
ニュートーカー	食材発注、検品などのASPサービス
大韓航空	フライト・スケジュール照会および日収計算システムに
ローソン	マルチメディア端末「Loppi」の機能強化
株式会社シマノ	顧客情報管理システム
毎日新聞社	営業支援システム
株式会社NTT-X	リサーチデータの既存データベースシステムのクラスタ化
株式会社第一勧銀情報システム	汎用 UNIXから Linuxベースでのシステム移行
株式会社麒麟シティ	世界初の Linuxベースオーダリングシステム

2002年,大手各社のLinux対応の動き

2002年

- 1.29 zSeries(エンタープライズ) およびiSeries(統合AP)でLinux専用モデルを発表
- 2. 8サン、x86ベースのLinuxサーバ投入を含め「Linux」サポートを拡充へ
- 9.24 (プレス) IBMとRed HatがエンタープライズLinuxソリューション提供で提携を強化
- 10.23 (ZDNET) 富士通、Linuxを大規模基幹システムに適用へ
- 12.13 (ZDNET) 大手3社、64ビット版データベース出荷へ
- 12.27 (日経ItPro) オープンソースの業務利用促進で30社が仮想企業設立

2003年

- 1.23 (ZDNET) IBMとHP、Linux関連の収益増大
- 1.24 (ZDNET) 富士通とIntel、Linux対応の基幹業務システム向けIAサーバ開発で協業
- 3.12 IBM zSeries 800の出荷台数が1,000台に到達(Linuxベース)
- 3.14 (ZDNET) ベリタス、主要なストレージソフトをLinux対応に
- 3.26 (プレス) 企業基幹システム構築「NEC エンタープライズLinuxソリューション」強化

2003年現在 Linux 市場動向

Linux Kernel 2.6
リリース間近

大規模システムの基幹業務へ導入される
可能性はあるのか？

Linux Kernel 2.6の機能

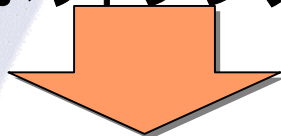
Linux Kernel 2.6 に追加される主な機能

ハイエンドシステムに向けた機能強化

- ・カーネル内部アルゴリズム変更によるスケーラビリティ向上
- ・64bit系CPU対応、大容量ハードウェア対応
- ・ファイルシステム、ネットワークの強化

商用UNIXサーバ領域へのスケールアップ

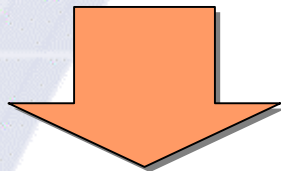
- カーネルのスケラビリティが向上
 - 多マルチプロセッサ、NUMAシステム対応強化
 - 実行並列度の向上
 - 大容量メモリ、大量のI/O機器の効率的管理
- IA64、x86対応による広いアドレス空間
- ファイルI/O性能強化
- スレッド機能強化
- ハードウェアのホットプラグ



商用ミッドレンジUNIXサーバが競合

ストレージ関連機能の強化

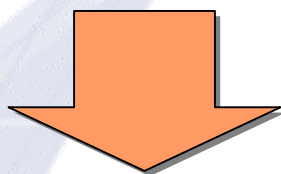
- ジャーナリングファイルシステム
- ファイルI/Oの高効率化
- 巨大ファイルシステム、巨大ファイル
- 論理ボリュームとファイルシステムの動的拡張
- スナップショット機能



バックエンド用途での適用の拡大の可能性

データベース関連機能

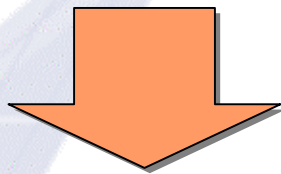
- DIRECT I/Oの性能改善
- 非同期I/O (AIO)のサポート
- readv/writevの本格的な実装
- 広大なユーザ空間 (IA64, x86-64)



商用DB、オープンソースDBとも、
かなり大きな性能向上が可能

ネットワーク

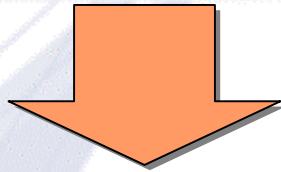
- 出来の良いIPv6
 - USAGIプロジェクトの成果
- IPSec
- SCTP



ユビキタスネットワーク社会基盤の一つ？

科学計算関連機能の強化

- HugeTLB(スーパーページ)機能
- 広大なユーザ空間 (IA64, x86-64)
- 多プロセッサ構成



クラスタ以外のソリューションもありうる？
大きなノードを結合したクラスタも可能

セキュリティ

- LSM (Linux Security Module)
 - 細かいアクセス制御が可能。ポリシーを動的に組み込むことが可能
- SELinux (実装例)

スケーラビリティ向上のための実装例

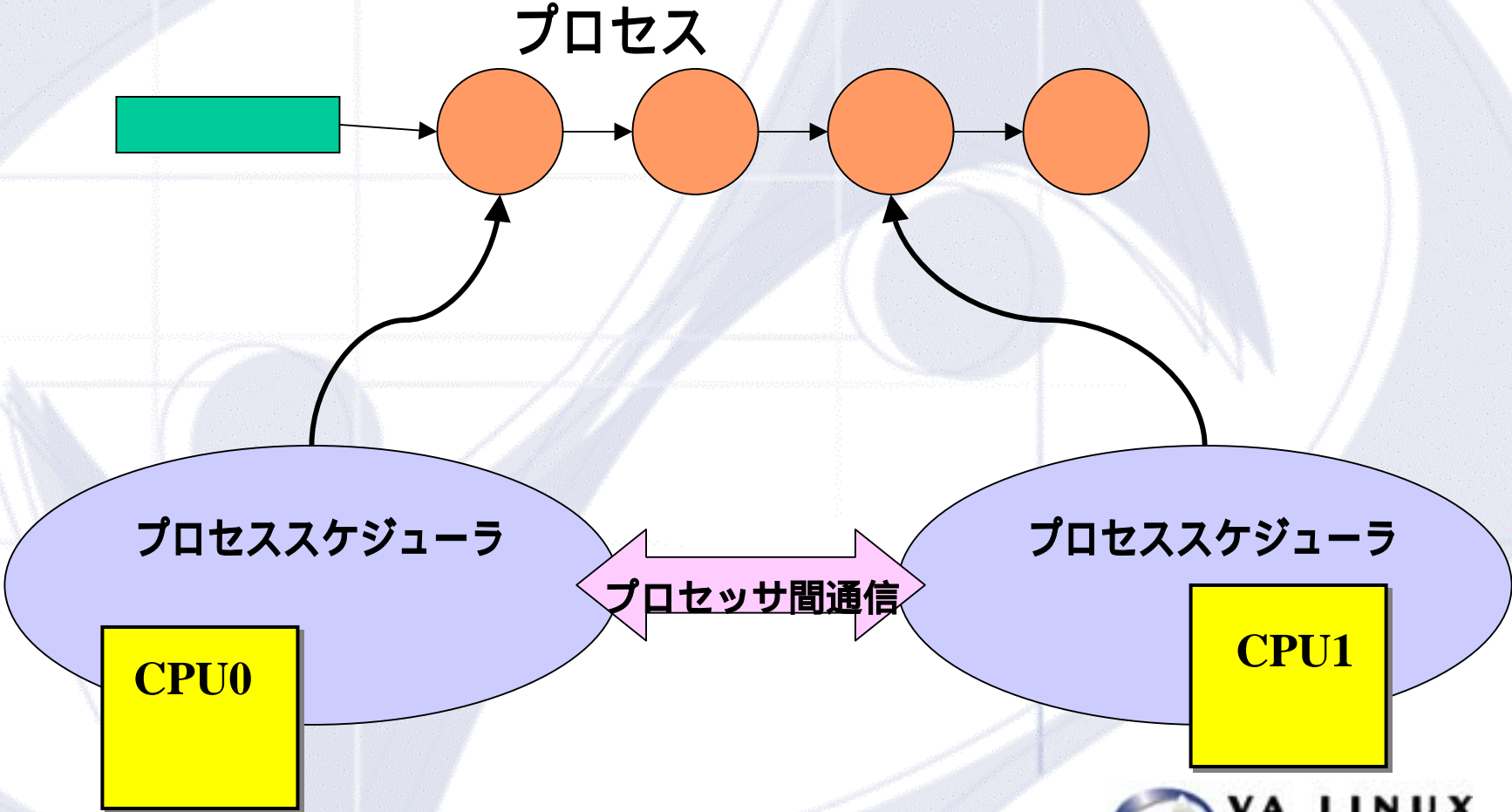
スケーラビリティ向上の実装例

- ▶ 多プロセッサ向けスケジューリング
- ▶ 高効率のファイルI/O

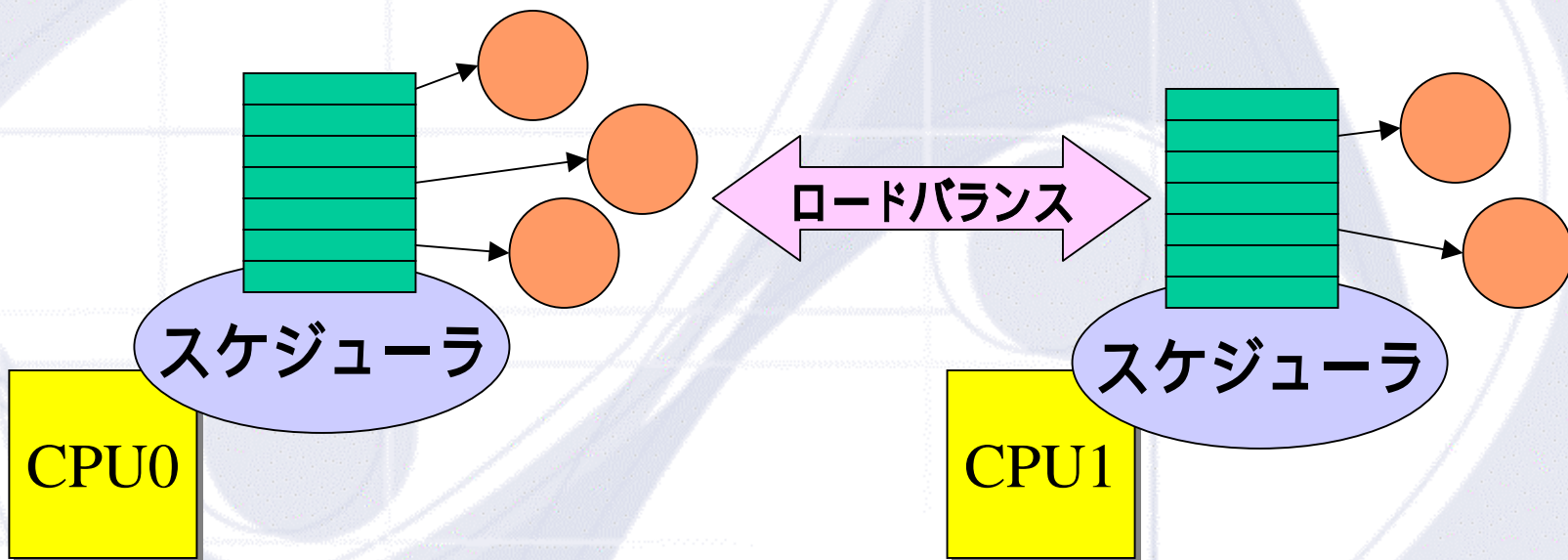
多プロセッサ向けスケジューリング

- CPU毎に独立したプロセススケジューラ
- CPU毎に並列動作するソフト割り込み処理
- CPU毎に並列動作するタイマー処理
- NUMAを意識したスケジューリング
- スケジューリングへの影響が小さな資源排他機構
 - 細かな資源排他、軽い資源排他機構、排他の発生しない資源分配

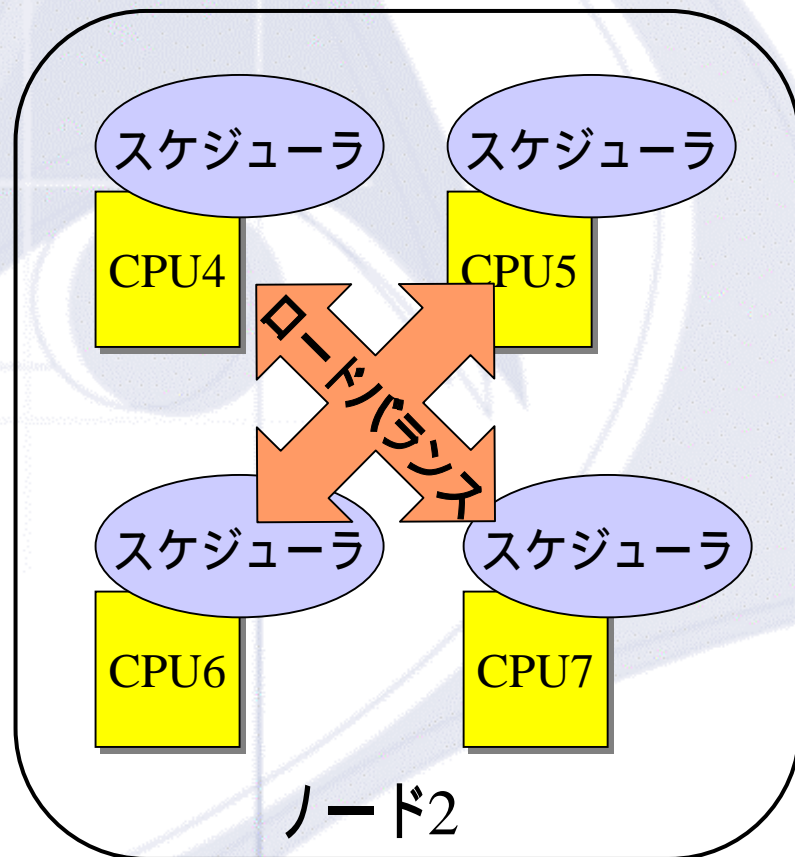
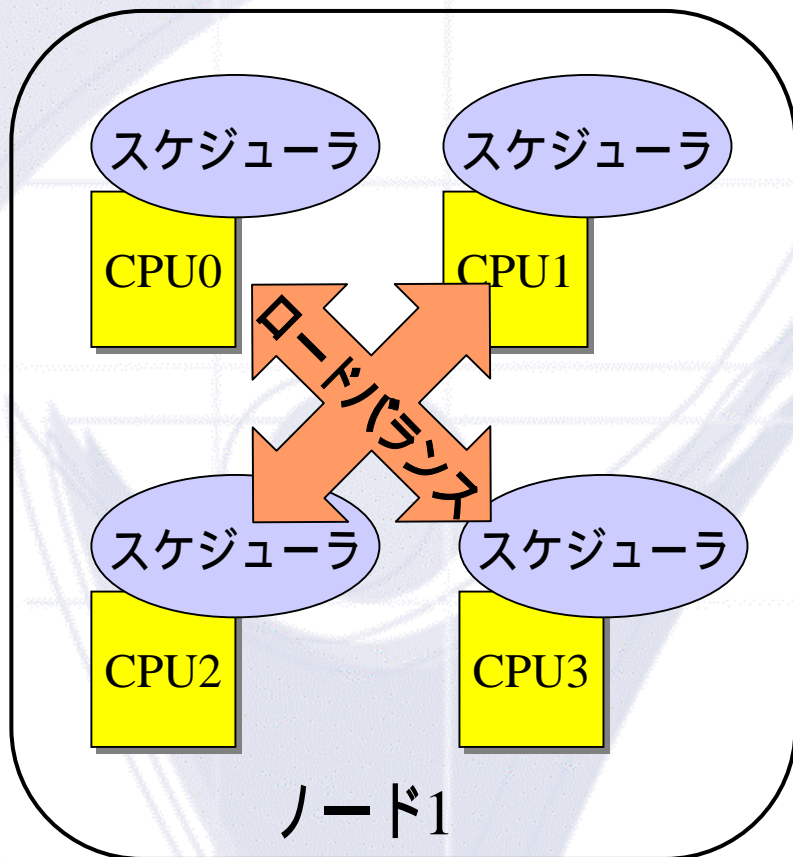
Linux2.4のスケジューラ



O(1)スケジューラ (1/2)



O(1)スケジューラ (2/2)

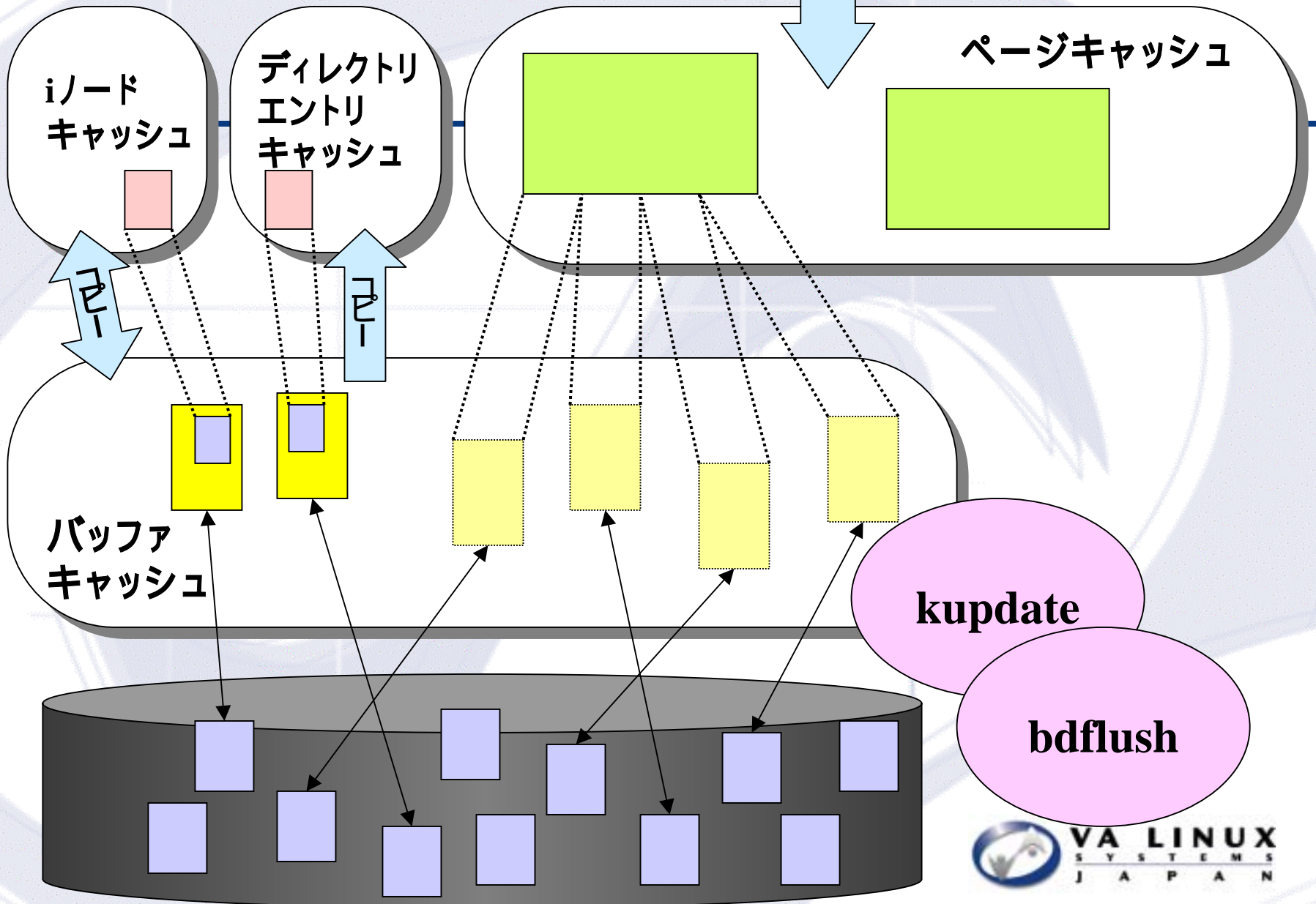


高効率のファイルI/O

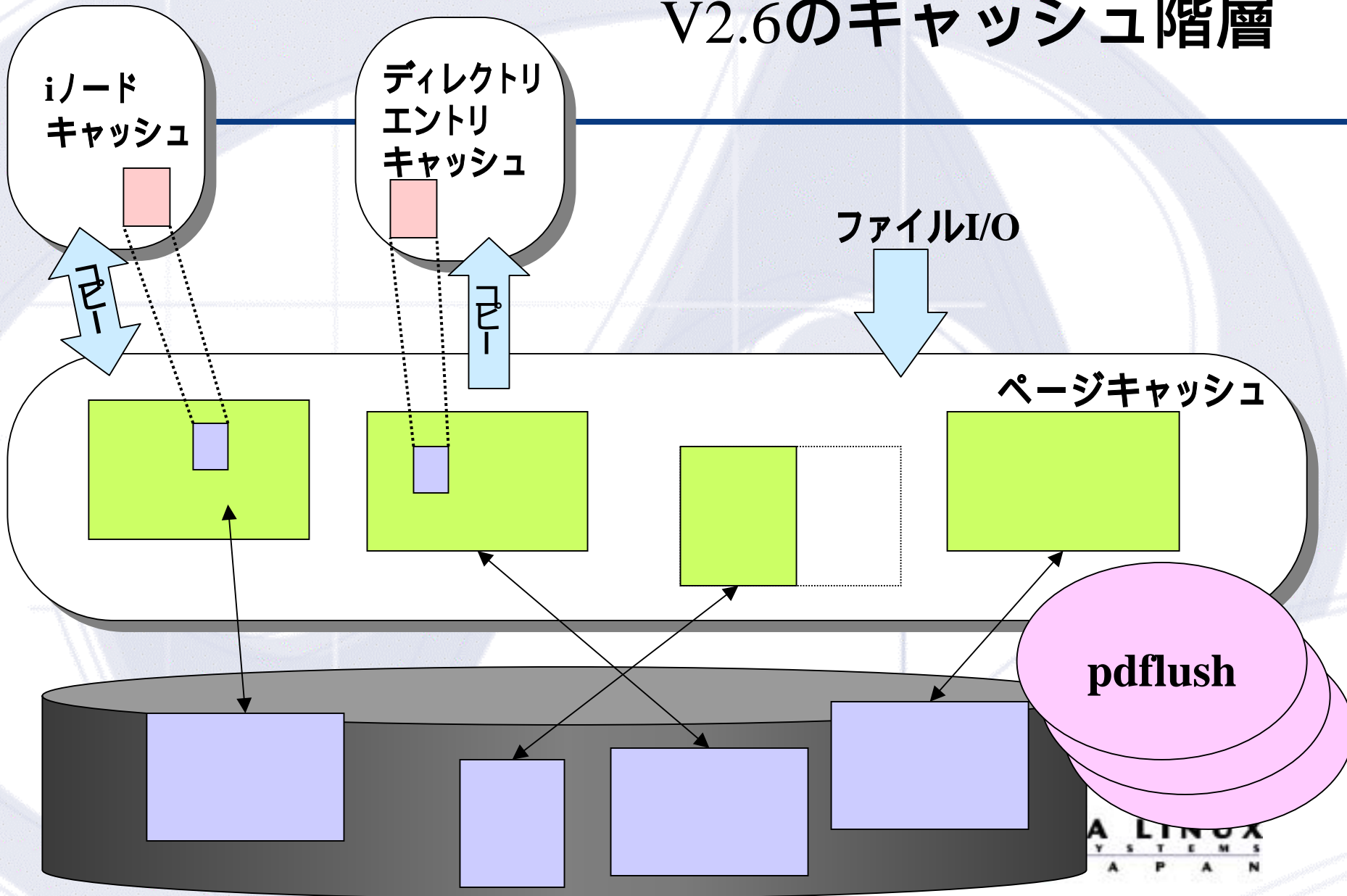
- ▶ ファイルI/Oデータのクラスタ化によるI/O効率化
- ▶ CPUが高並列実行可能なファイルシステムとブロックI/Oサブシステム
- ▶ VMとの統合

V2.4のキャッシュ階層

ファイルI/O

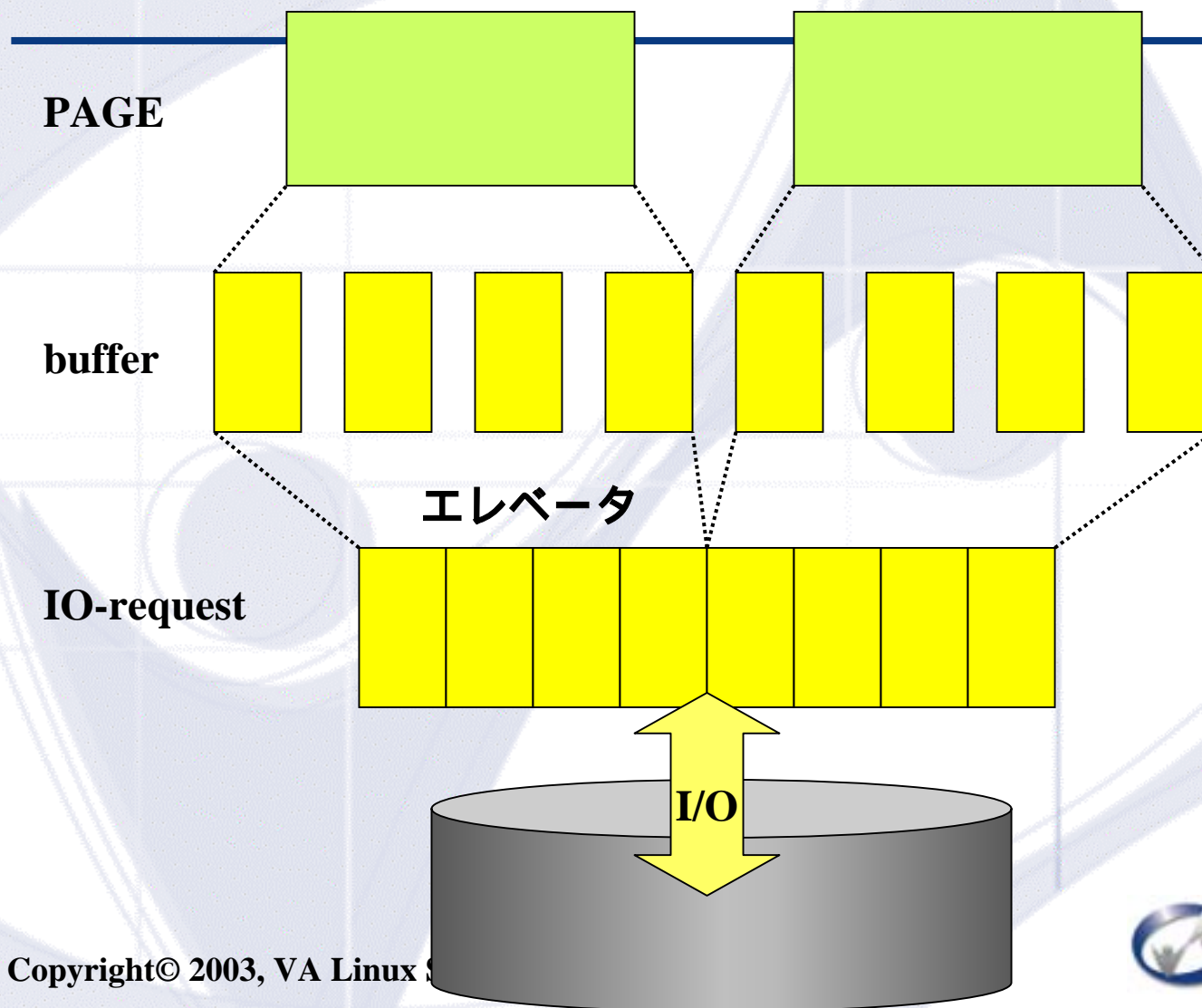


V2.6のキャッシュ階層



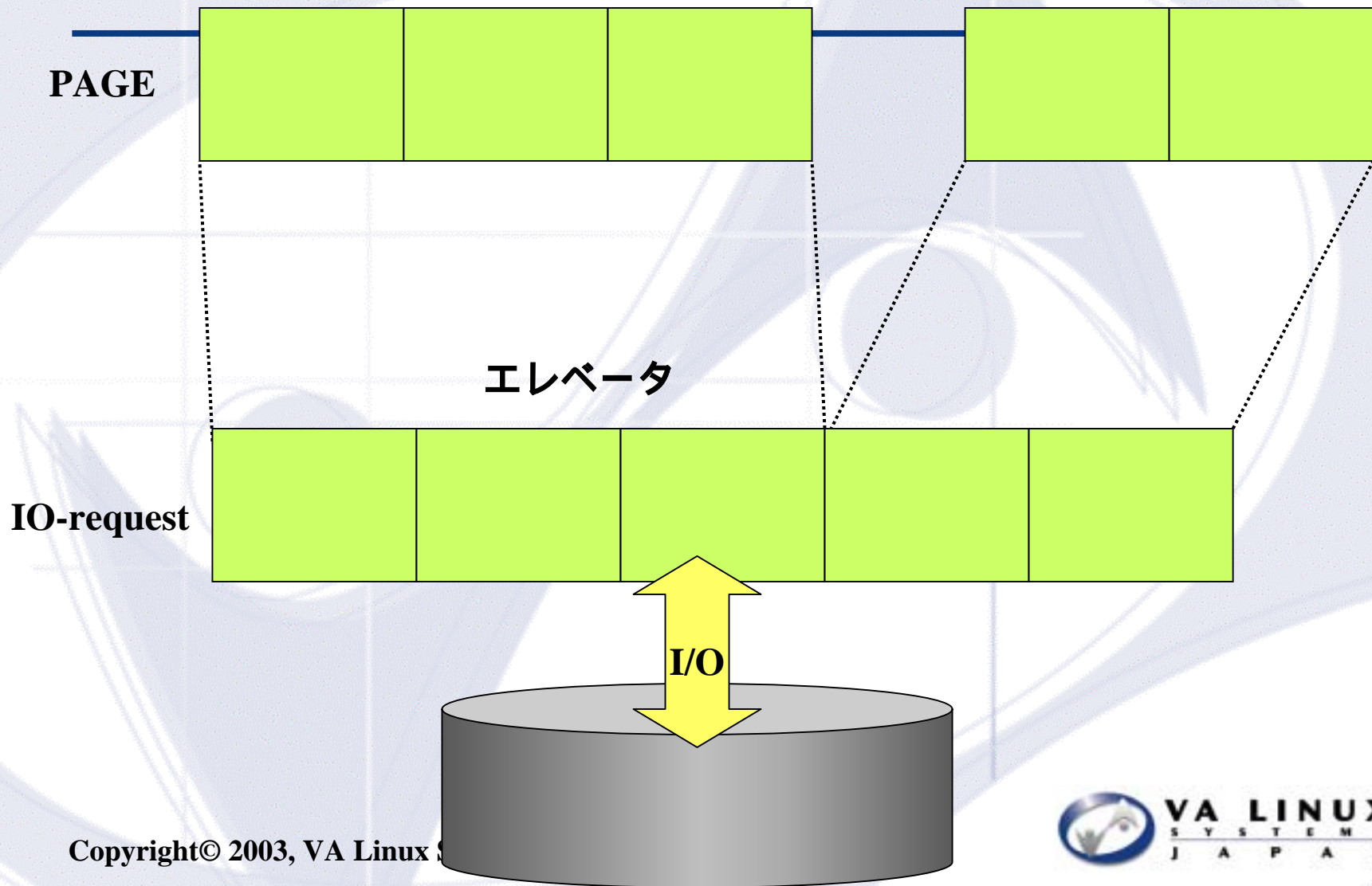
V2.4

キャッシュに対するI/O



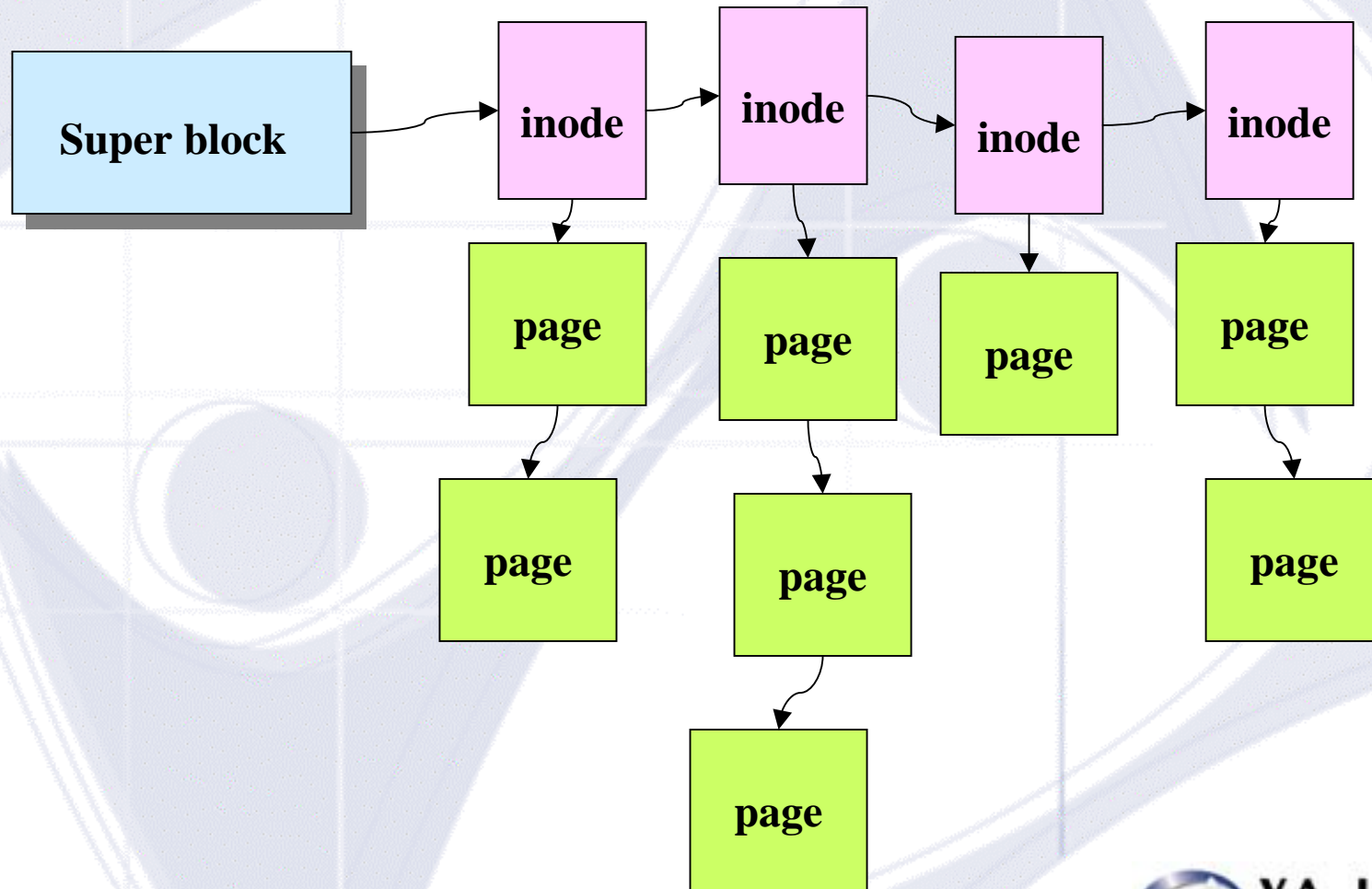
V2.6

キャッシュに対するI/O



V2.6

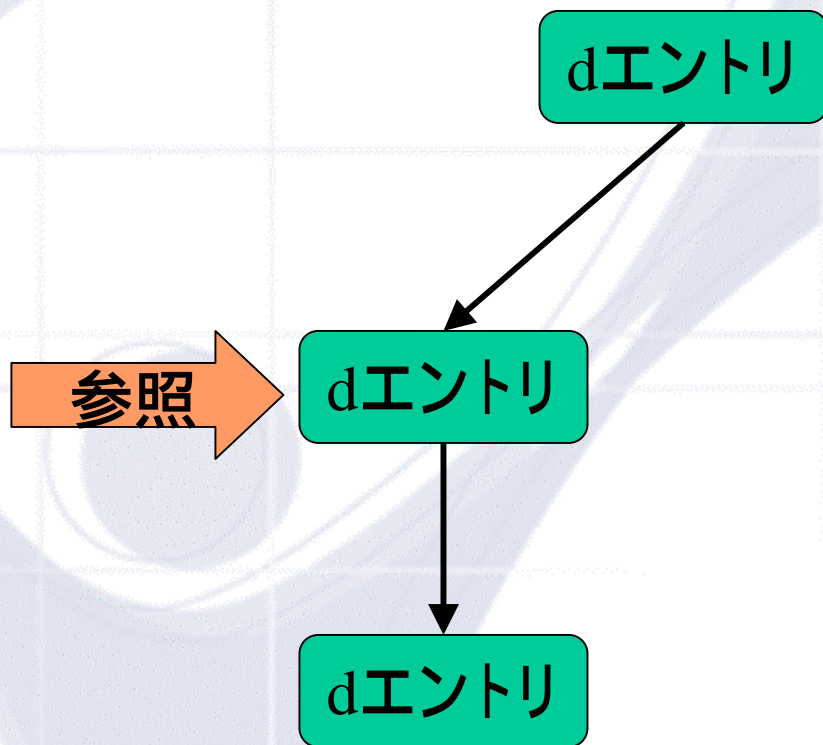
汚れたキャッシュの管理



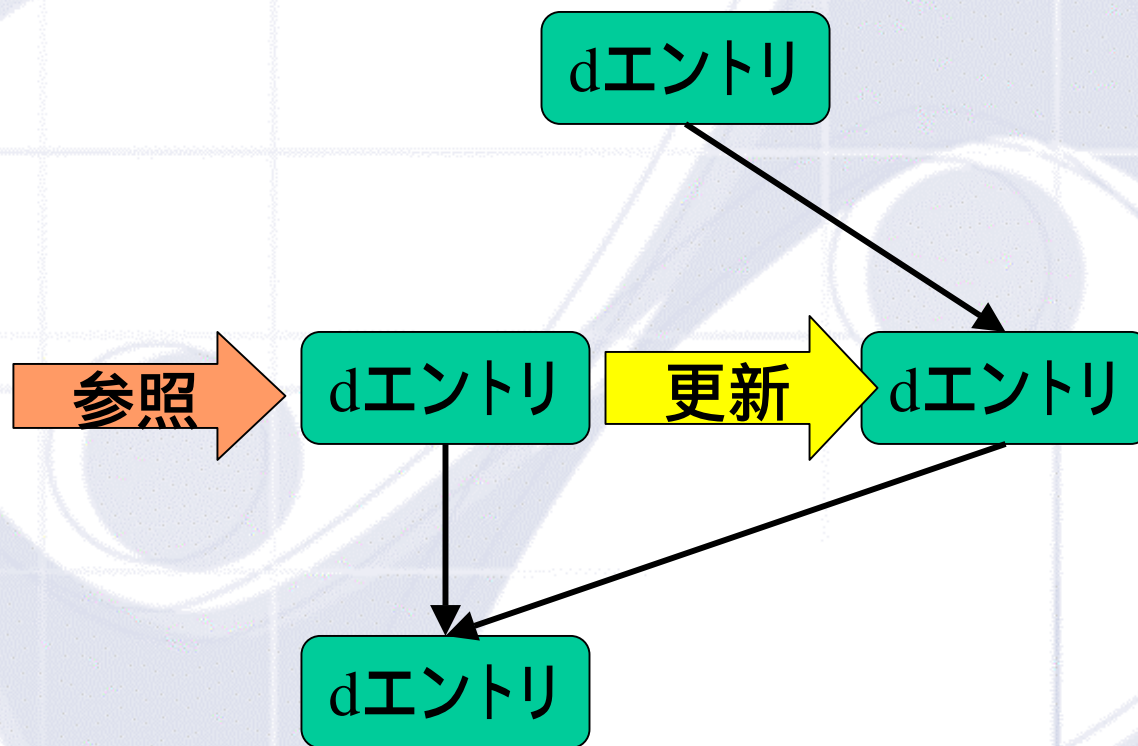
RCU(Read-Copy-Update)高速排他

- ◆ スピンロックに代わる新しい排他方式
- ◆ 最近のCPU実装に依存した排他機構
- ◆ 単純な読み書きスピンロックでは、CPU数が増えてくると
キャッシュライン共有により性能劣化
- ◆ 特に、**dエントリの排他処理**への適用で、ファイルパスの
解析処理に大きな効果 (Linux/UNIX OSにおいて最も頻
繁に発生する最も重い処理の一つ)

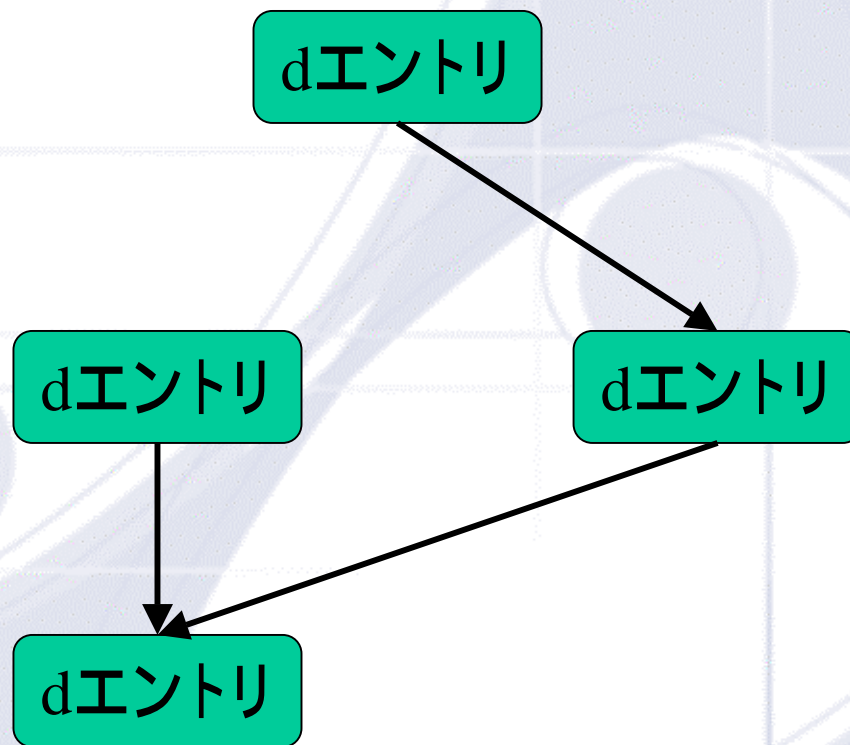
RCUの動作(1/4)



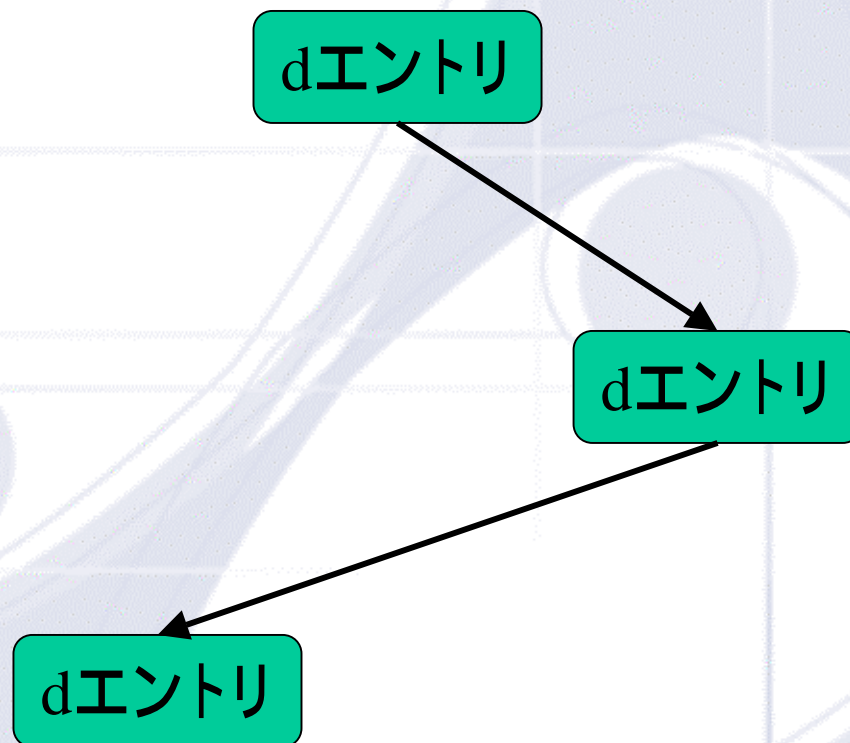
RCUの動作(2/4)



RCUの動作(3/4)



RCUの動作(4/4)



Linuxの課題

- 正確な情報
- 保証
- 長期サポート

開発状況

力強い開発が継続

- Linux 2.6 リリース直前
 - Linus氏としての機能取り込み作業は完了。他のツリーのマージと安定化 (半年くらいかかるかも)を残すのみ。
- Linux 2.7は？
 - 更にハイエンドサーバ向けの機能強化に向けて進むものと思われる
- Linus氏が、transmeta社からOSDLへ移籍
 - 初めて本業としてLinuxに携わる

VA Linux systems ジャパン
Linux カーネルへの取り組み

大規模システム向けLinuxカーネルのサポート

システム環境にあわせたLinuxカーネル

Linuxカーネルに特化したグローバルサポート

- カーネル情報、開発情報の提供
- カーネルの調査、分析
- 基幹業務システム向け独自チューニング
 - カーネル本体、AP動作解析支援
- カーネル障害解析、カーネル視点からのAP障害分析

おまけ

- ▶ **詳解Linuxカーネル第二版出版
(Oreillyジャパン社)**
 - Linux 2.4の解析文献
 - Linux 2.6の実装にも軽く触れる
- ▶ **会場の入り口にて先行発売**
 - 10%引き、消費税サービス

ご静聴ありがとうございました。