HPCアプライアンスソリューション
システム構成と仮想化テクノロジー

スケーラブルシステムズ株式会社

HPCアプライアンスソリューション

- 様々なワークロードに対応したコストパフォーマンスに優れたシステムを構築
- システムのコンセプト
  - 64ビットのインテルXeonプロセッサを搭載したx86プラットフォームのクラスタと共有メモリシステムを高速の共有ファイルシステムと共に提供
  - OSやクラスタストレージの出荷前のセットアップ
- 利点
  - インストール後、すぐにシステムの利用が可能
  - プラットフォームの各コンポーネントは、それぞれが高い拡張性を持つ
  - 複雑なワークロードに対応可能なインフラストラクチャとなります。ITインフラストラクチャに対する投資を最適化

システム導入時の課題

従来のシステム調達	対応方法	問題点
• ユーザとアプリケーション毎に異なった計算リソースの要求（プロセッサ数、メモリサイズ、IOなど）	• 要望に合わせた構成の異なった計算機システムの個別導入	• 導入コストと運用コスト
• 導入後はシステムの構成を変更するのが困難	• 導入後のシステム利用に関する検討	• 評価作業コスト
• ワークロードの変化や季節的な計算負荷の変化に対応するのが困難	• 導入時にある程度余裕のある構成でのシステム導入の検討	• コスト高
• 短期的に大きなメモリ空間が必要な場合の対応	• 大規模メモリシステムの新規導入	• コスト高

HPCアプライアンスソリューション

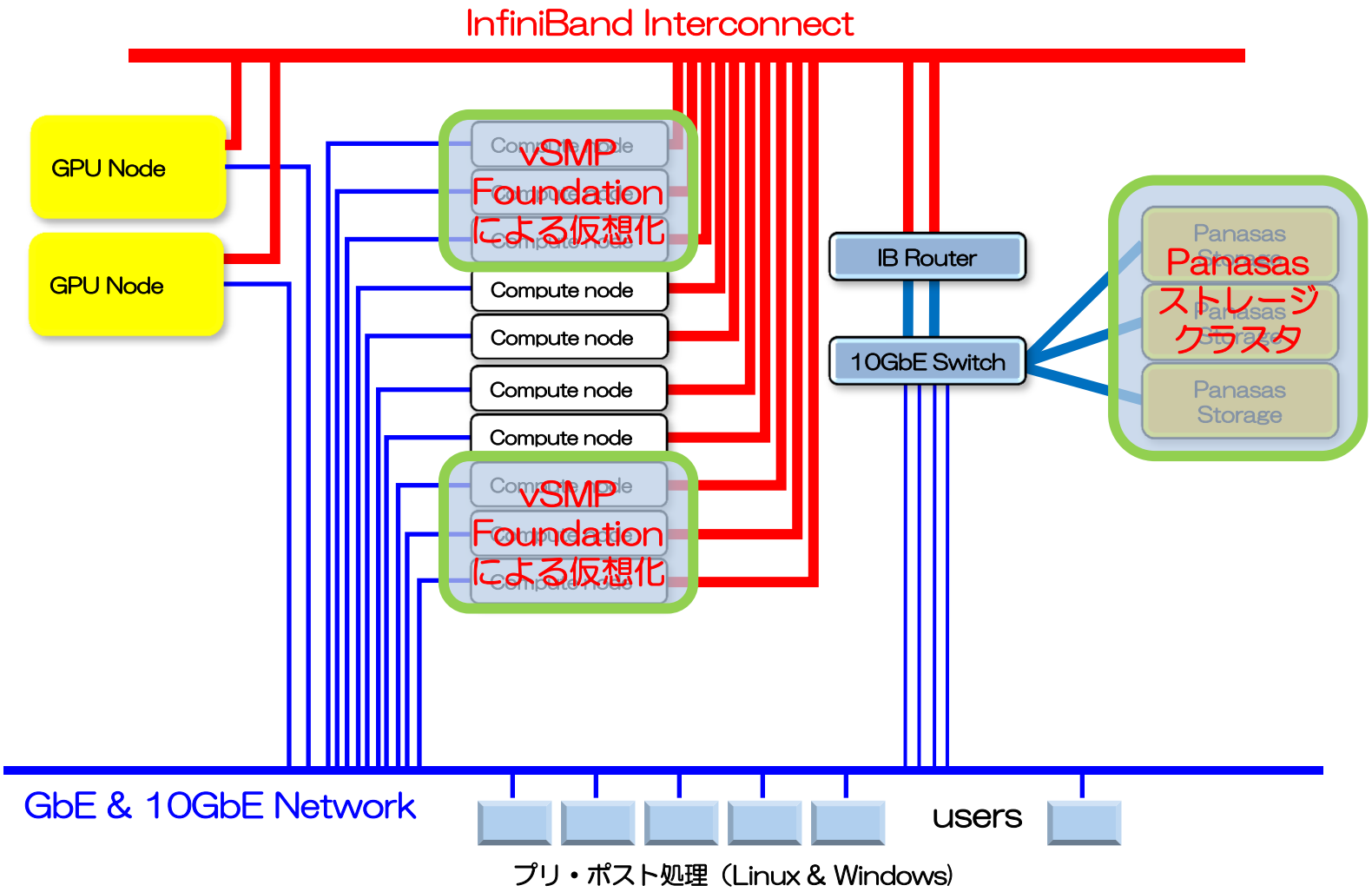
ユーザにとってのメリット

- インテルの最新プロセッサとScaleMP vSMP Foundationにより、より多くのプロセッサと大きなメモリ空間をアプリケーションが利用可能
- 高いスケーラビリティと大容量メモリまで拡張可能）を持つシステムをアプリケーションやプログラミングモデルの変更なしで利用
- パラレルNFSによるスケーラブルなIO性能とより高いIOバンド幅によるアプリケーション実行の効率化

管理者にとってのメリット

- データセンターで利用されているインテル の標準化技術を最大限に活用したスケーラブルなエンタープライズクラスのシステムとして、多くのIT部門で利用可能なシステム
- SMPシステムは、クラスタに対しては、その運用管理が容易であり、現在、最も問題となっているTCOにおけるそれらの管理費用を大幅に削減することを可能
- Panasasのストレージアプライアンスによりリソースのプロビジョニングを迅速に行うことが可能となり、システムを利用する際のインストール、設定、保守といった手間を大幅に省くことが可能となります。

HPCアプライアンスソリューション システム構成



HPCアプライアンスソリューション その利点

最新技術の利用	日々進化している最新のIT技術、そして最新のx86システムを常に利用できることが要求されます。標準化されたITインフラに採用される最新技術を即座に利用できることで、ITインフラの陳腐化を避けることが可能となります。さらに、最新技術の導入によって、より高い性能をより低コストで実現するシステムが構築できることも重要です。
容易な運用管理	導入が容易でシステムの設置後、直ぐに利用出来る計算機システムとなります。また、特別なシステムツールを必要とせず、一般的な運用管理ツールなどを利用して、スケジューリング、システムのモニター、リソース管理などを一元的に管理可能にする、高度な柔軟性を備えていることが要求されます。
拡張性と柔軟性	ITインフラでは、より低コストでの導入とシステムの容易で柔軟な拡張が重要です。また、リソースの無駄を排除すると同時に、過不足なく効率的に活用できるようにリソースがプールされていることが求められます。
オンデマンドで利用可能	ワークロードに合わせて柔軟にシステムを構成することができます。ジョブやプロジェクト、あるいは顧客毎に仮想マシンを構築、プロビジョニングすることが可能で、お客様のニーズに最適なインフラ構築が求められる場合には最適です。
シングルインフラストラクチャ	一般的なブレードサーバやラックマウントサーバで構築することができます。汎用サーバは、各ベンダーが競ってシステム開発を続けることで常に改良されており、最新システムはより優れたエネルギー効率、より高い性能、より高度なシステム実装が可能となるため、その技術革新のメリットを100%活用することが可能となります。
仮想アプライアンス	ソフトウエアによるハイエンドな仮想化技術を最新のハードウエアに適用することで、リソースのプロビジョニングを迅速に行うことが可能となり、システムを利用する際のインストール、設定、保守といった手間を大幅に省くことが可能となります。

HPCアプライアンス構成システム

HPCアプライアンスソリューション

サーバプラットフォーム

スケーラブルシステムズ株式会社

Atlasサーバライン

製品の特長

- 1U-5Uサイズ
- ラック搭載＆タワー型
- 用途に応じたシステム提案
- クラスタ対応

タワーサーバ　Atlas　T120/T220

1U ラック搭載サーバ　Atlas 120

2U ラック搭載サーバ　Atlas 530

2U 高密度ラック搭載サーバ　Atlas 3100

KRONOSS

Build by Engineers for Engineers

NEXXUS

スケーラブルシステムズ株式会社

KRONOS S810/S810R

- インテル Core i7-3960X Extreme Edition（3.30GHz）プロセッサーをオーバークロックで4.8GHzまで動作保証
- 最大64GBの高性能メモリ
- OCZ RevoDrive 3 X2 SSDを採用
- Nvidiaハイエンドグラフィックス搭載
- ビジネス用途向け高信頼性冷却システム
- 1200W電源装置（80PLUSゴールド認定）
- 静音設計

3U 19インチ標準ラック搭載

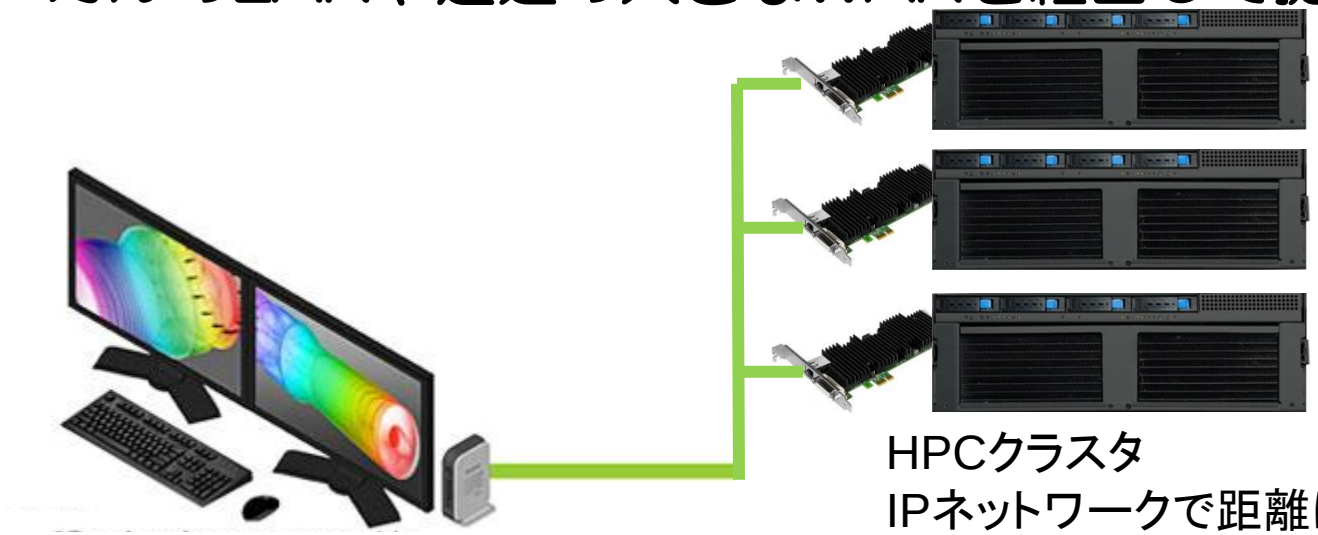
KRONOS S810R HPC

- KRONOS S810R HPC
  - KRONOS S810を計算ノードとしたクラスタシステム
  - 高いシングルノードの計算能力とInfiniBandクラスタでのスケーラビリティの提供

- 製品構成オプション
  - InfiniBand HCA (QDR/FDR)
  - InfiniBand Switch (8/18/36 ポート QDR/FDR)
  - OCZ RevoDrive 3 X2 SSD
  - OS プリインストール
  - 搭載ラック
  - PCoIPソリューション

PCoIP (PC-over-IP)ソリューション

- IT資源をデータセンターに実用的に統合
- ネットワークを経由して任意の場所で利用可能
  - データではなくピクセル(画素)を転送
  - 高解像度でフルフレームレートの3Dグラフィックス
HD メディアの転送、完全なUSBデバイスの相互運用性を、
ローカルのLANや遅延の大きなWANを経由して提供



クライアント
3次元画像などを円滑に処理

HPCクラスタ
IPネットワークで距離に
関係なく利用可能

デスクサイドHPCアプライアンス

NEXXUS

- シンプル
  - 完全にインテグレートされたシステムで提供されることによるユニークなシステム特性
  - ハイパフォーマンスコンピューティング向けに最適化されたシステム構成
  - オフィスでの利用を前提に開発された静音性や電源仕様（110V/15A）
- モジュール構成と柔軟性
  - 必要とするリソース（プロセッサ、GPU、メモリ、グラフィック、ストレージ）に合わせてシステムの選択が可能
  - 最新技術の利用が可能
- 経済性
  - 予算に応じたシステム構成とアップグレード
  - 容易なインストールと運用管理（管理ソフトウエア）

NEXXUS C 製品概要

- デスクサイド設置
- オフィス利用に適した静音性
- 110V/15A でも 220V/15A でも稼働可能
- モジュールデザインとスケーラビリティ
  - ワークステーションノード
  - ストレージノード
  - 計算ノード、GPUノード
  - ネットワーク
    - GbE、InfiniBand（オプション）
- Gigabit, 10 Gigabit 又は InfiniBand RAS
  - ホットスワップ可能な冗長電源
  - ホットスワップ可能なSAS/SATA ハードドライブ
  - コールドスワップ冷却ファン
  - WEBベースマネージメントソフトウエア
- 3 年間製品保証

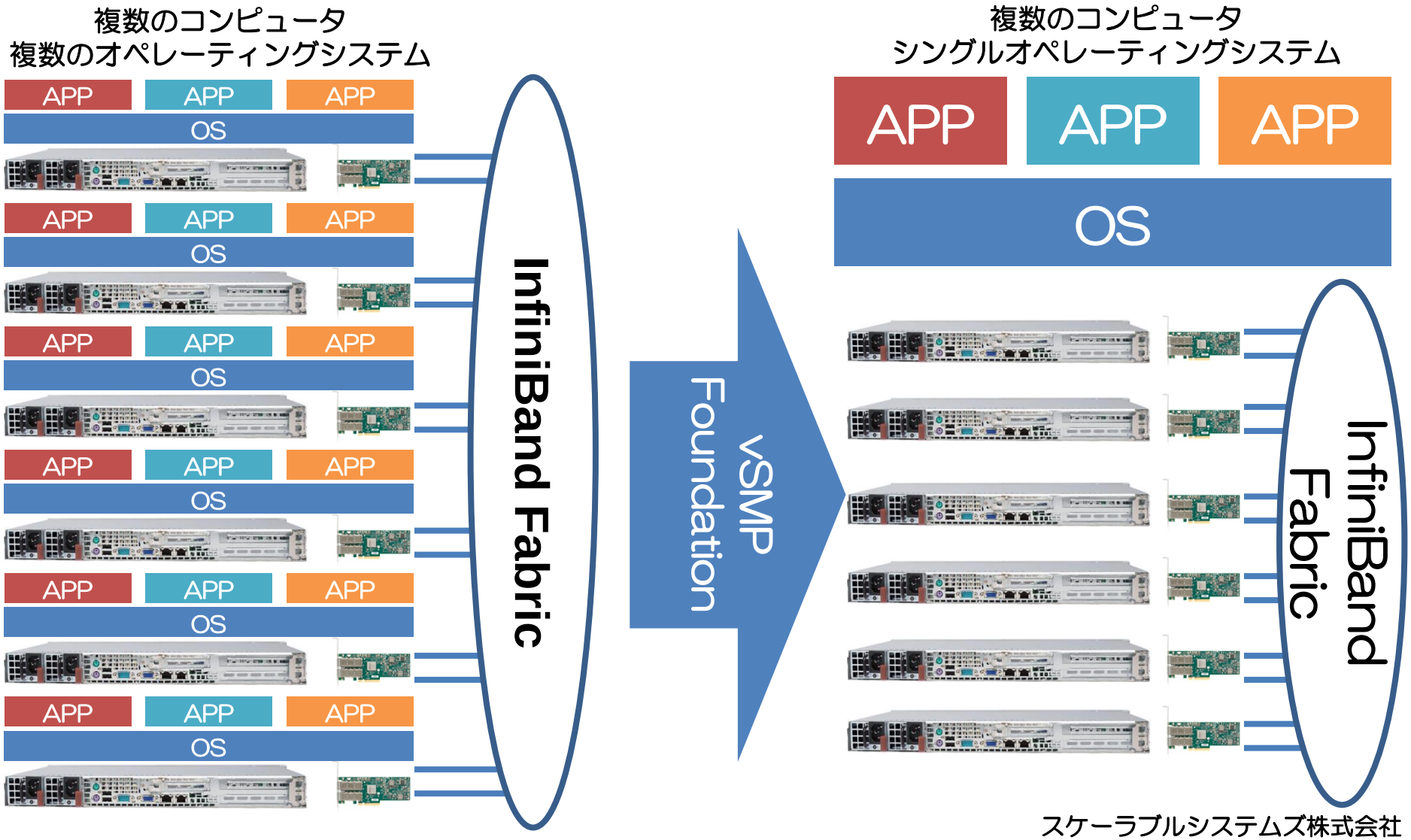
ストレージノード

HPCアプライアンスソリューション

スケーラブルSMPサーバ

スケーラブルシステムズ株式会社

vSMP Foundation 仮想化



vSMP Foundationによる ハイエンド仮想化

- Intel 64(x64/x86-64)システムによるシングルシステムを実現
- 標準Linuxをサポートし、100%のバイナリ互換

システム構築の基本

- 一般的なブレードサーバ、ラックマウントサーバで利用可能
- インターコネクトして広くInfiniBandを利用

システムコンポーネント

- 16,384プロセッサコア、最大64TBまでの拡張性

スケールアップ（最大構成）

- ユニフォームなアプリケーション実行環境を構築
- ソフトウエア階層を変更なしで提供

シングルシステム

- システム全体のメモリコヒレンシとキャッシュコヒレンシを提供
- メモリアクセスのリアルタイムでの最適化

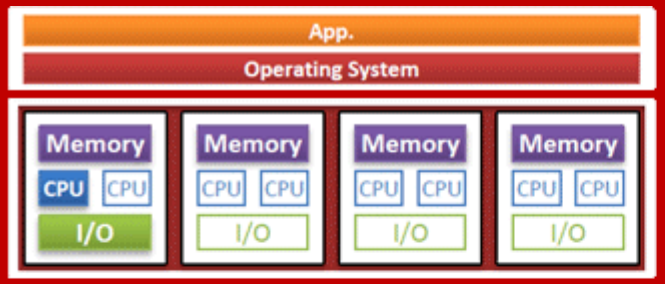
コヒレントメモリ

- すべてのIOリソースに対して、シングルOSがアクセス可能
- シングルファイルシステムを実現

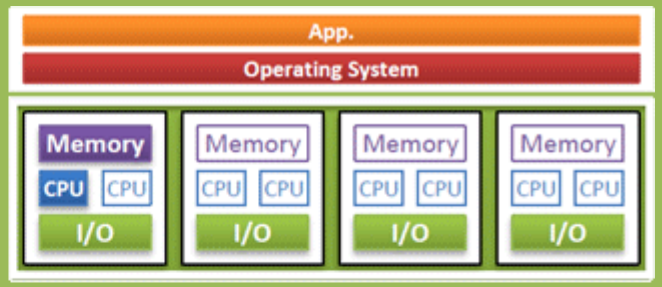
共有IO

ハイエンド仮想化の目的は?

大規模メモリ空間

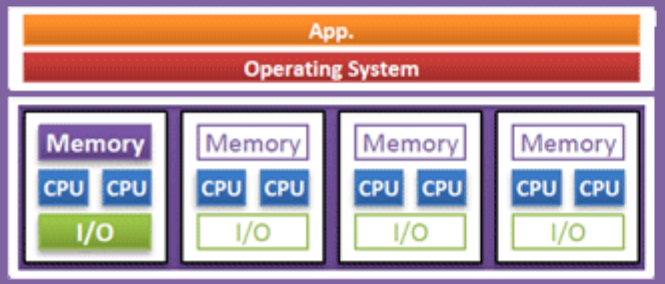


I/O処理能力

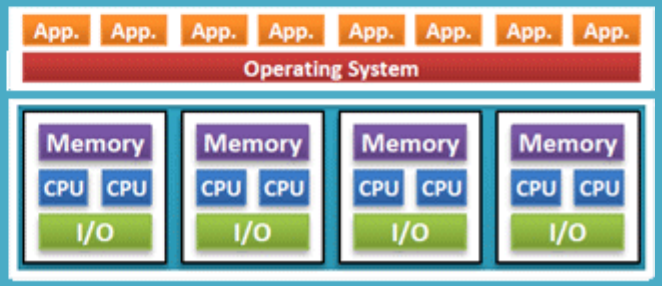


仮想化SMP

計算処理能力



システム統合



ハイエンド仮想化の目的は?

大規模メモリ空間

大規模メモリを必要とするワークロード

- 少ないコア数で、大きなメモリ空間を利用
- 複数ノードのメモリを利用

I/O処理能力

I/O負荷の大きなワークロード

- 少ないコア数で、大容量のファイル処理を行う
- ノード間でI/O処理を分散する必要がある
- メモリをI/Oのバッファに利用して速度向上を図る

仮想化SMP

CPU負荷の大きな共有メモリアプリケーション

- スレッドアプリケーション（OpenMPやPthreadsなど）

計算処理能力

アプリケーションの実行をよりシンプルに

- 複数のタスクを実行するアプリケーションの実行に際して、その実行手順をより簡便に実施

システム統合

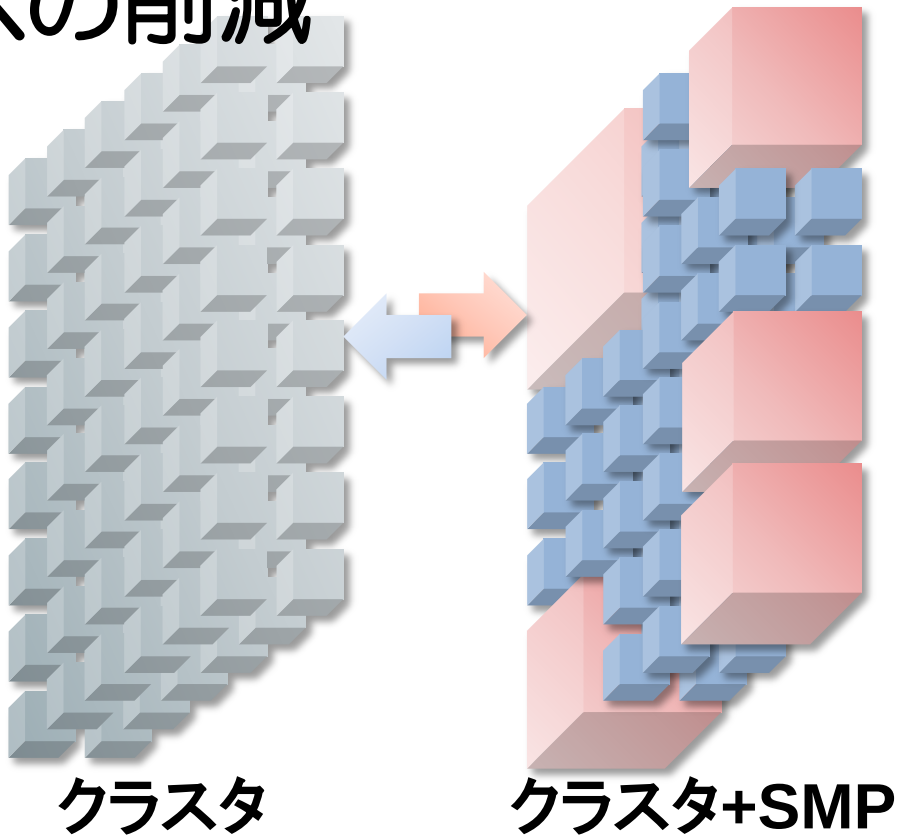
vSMP による仮想化の利点

- 物理的な計算リソースの要求に合ったシステムを構築
- OPEX(運用費用）とCAPEX(導入費用）の削減を目的

提供可能なシステム	目的	ソリューション
シングル オペレーティングシステム	OPEXの削減	サーバ・コンソリデーション
シングル （大規模）システム	CAPEXの削減	計算能力とメモリ容量の要求への柔軟な対応
シングル インフラストラクチャ	OPEXとCAPEXの削減	クラウド対応 （オンデマンドでのインフラの構築）

vSMP による仮想化の利点

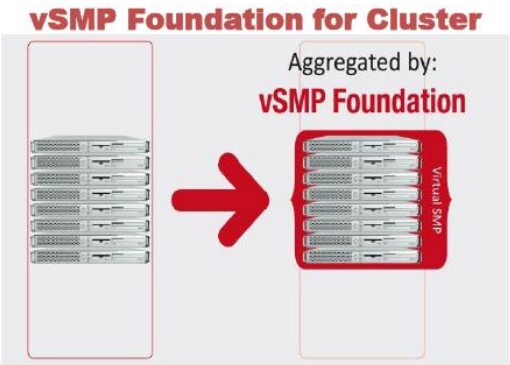
- シングルインフラストラクチャによるOPEX&CAPEXの削減
  - 同じハードウエア基盤とソフトウエアを利用して、
  - 用途とアプリケーションに応じて、システムを動的に再構成



vSMP Foundation導入のメリット

シングルオペレーティングシステム

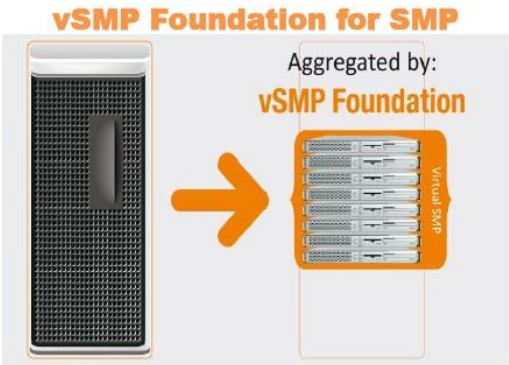
管理とサーバコンソリデーション



OPEXの削減

シングル（大規模）イメージシステム

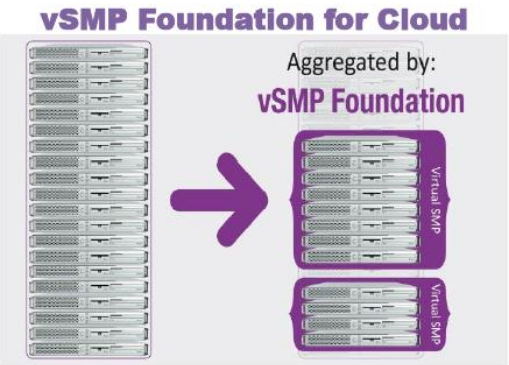
計算能力とメモリ容量の要求への柔軟な対応



CAPEXの削減

シングルインフラストラクチャ

クラウド対応（オンデマンドでのインフラの構築）

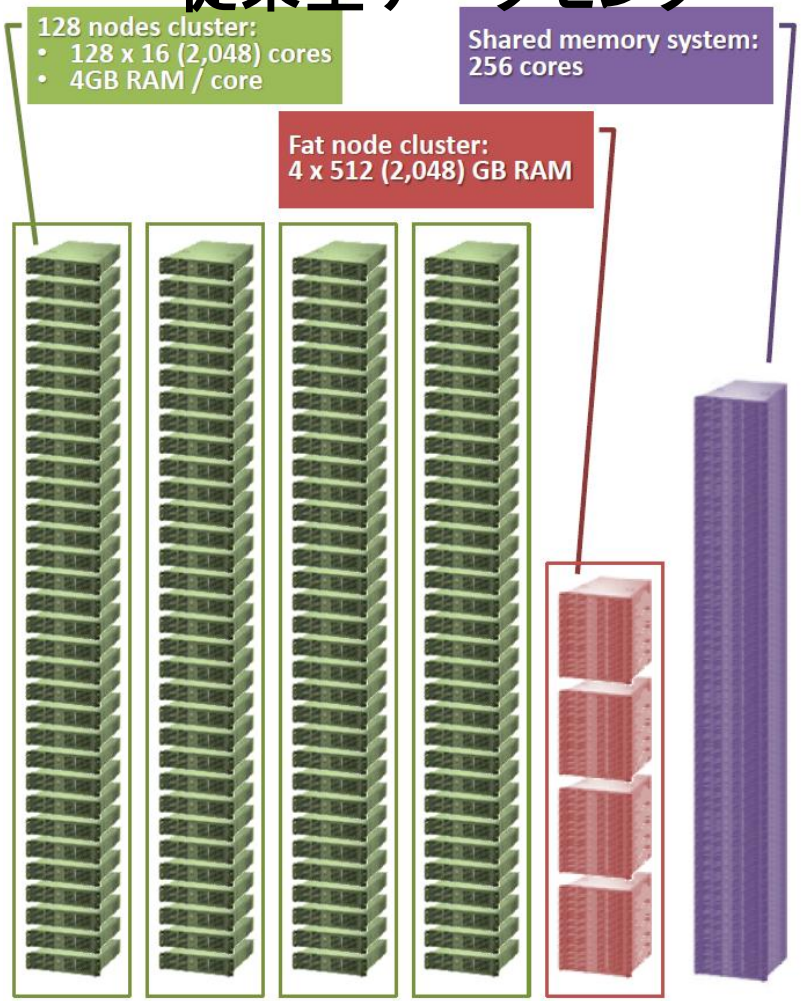


柔軟なシステム構築
OPEXとCAPEXの削減

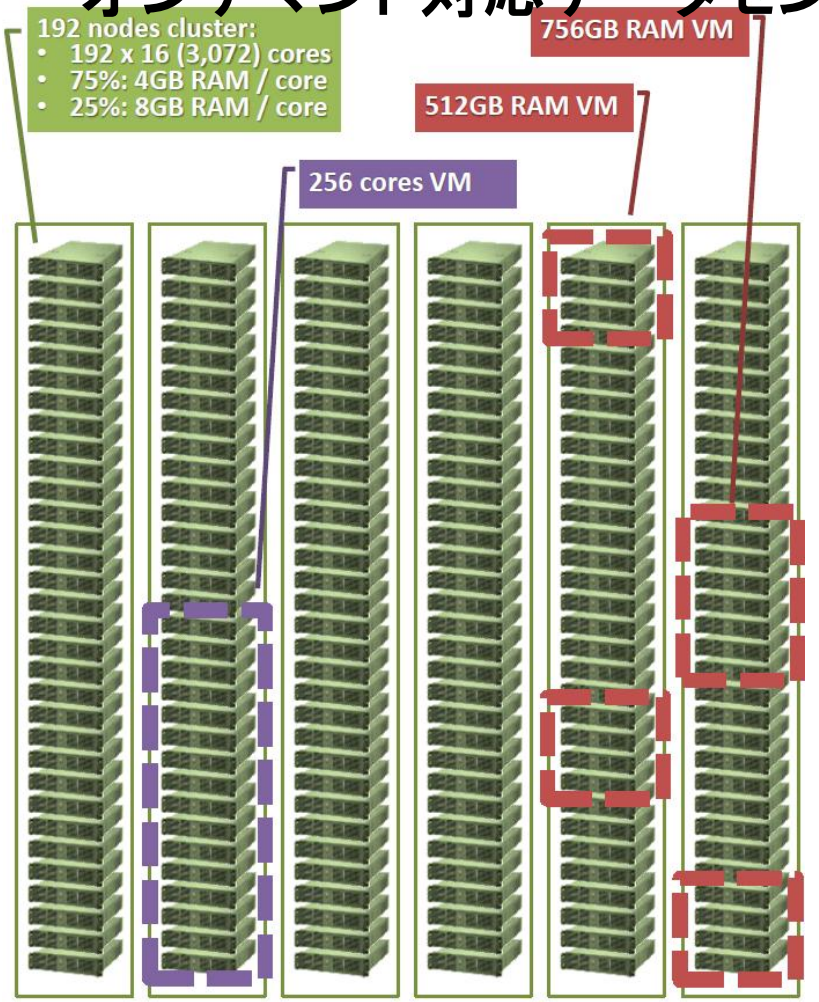
OPEX(運用費用）　CAPEX(導入費用）

従来型データセンター と オンデマンド対応データセンター

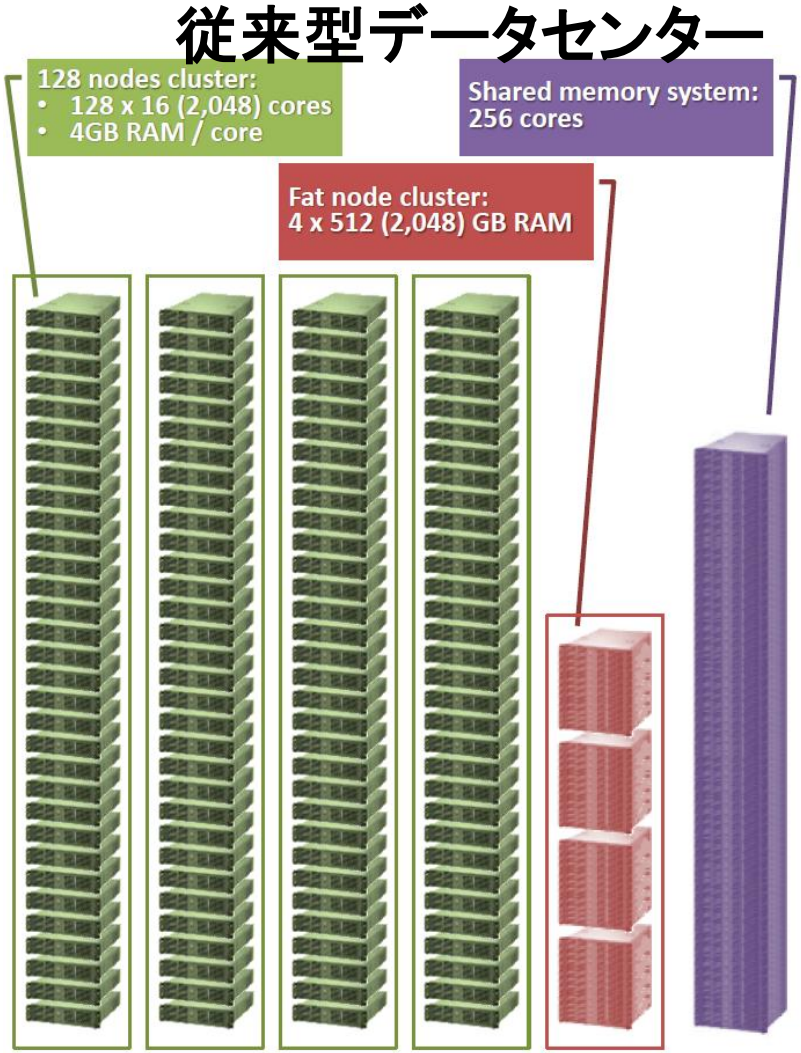
従来型データセンター



オンデマンド対応データセンター



従来型データセンター と オンデマンド対応データセンター

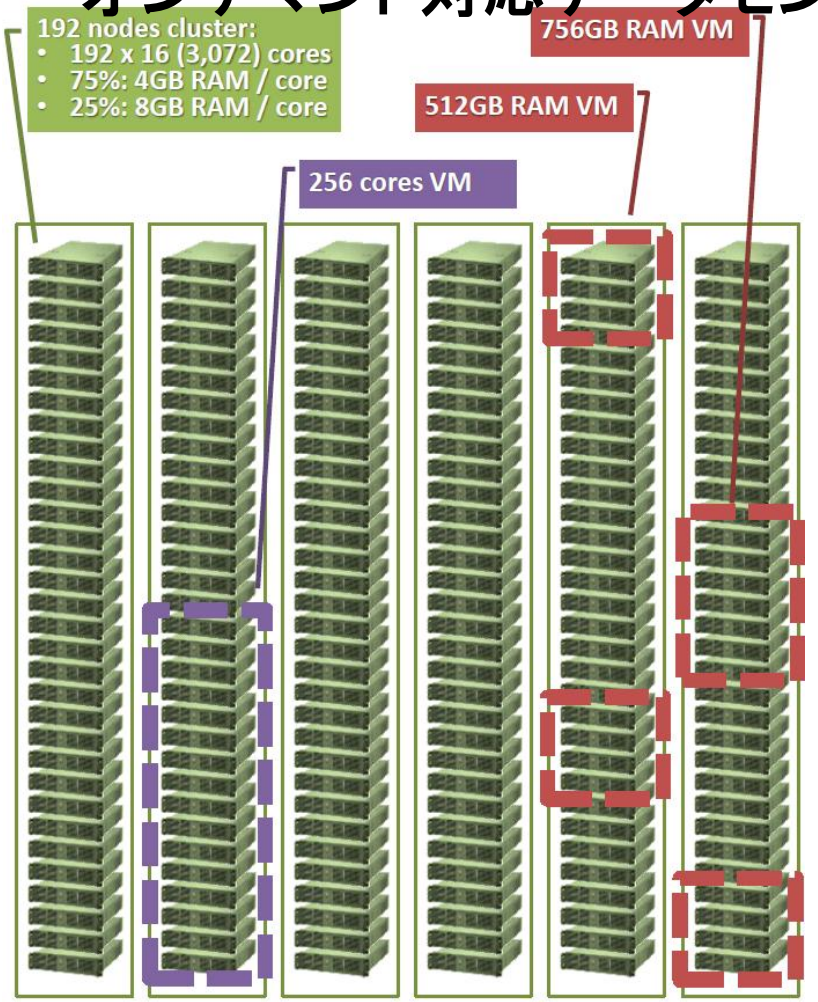


- 利用環境
  - アプリケーション毎の個別システム
- 利用効率
  - 負荷の偏在時の対応
- 運用効率
  - 利用状況の変化に柔軟に対応出来ない（必要とするリソースの偏在）
- 費用
  - クラスタシステムと比較して、SMPシステムの導入コストが高い

従来型データセンター と オンデマンド対応データセンター

- 利用環境
  - 単一のハードウエアインフラ
  - アプリケーションに合わせてシステムを再構成可能
  - ノード毎のプロビジョニングを動的に実行
- 利用効率
  - プロジェクト毎に必要なリソースに合わせてSMPシステムを構築
  - クラスタ環境の保持
- 運用管理
  - クラスタ利用と同じように必要なリソースに合わせてSMPシステム利用のリソース管理が可能
- 費用
  - 必要なハードウエアはクラスタシステム構築に必要なもので十分

オンデマンド対応データセンター



- Major deployment at STFC - Daresbury Lab, UK
  - 512 IBM iDataPlex M4 nodes:
  - 2 x Intel E5-2670, 128GB, InfiniBand
- Maximum VM capability: 384 nodes, 48TB RAM
- On-demand SMP by Platform Computing's provisioning and scheduling tools

シングル(大規模)システム 技術革新とシステムコスト

従来型システムの問題点	vSMPでのソリューション
・独自規格でのシステム開発による最新技術への追随の遅れ	・サーバプラットフォームの技術革新を直ぐに活用可能
・システム開発期間とシステムコストの問題	・導入後もソフトウエアのアップデートによって、性能向上を恒常的に実現

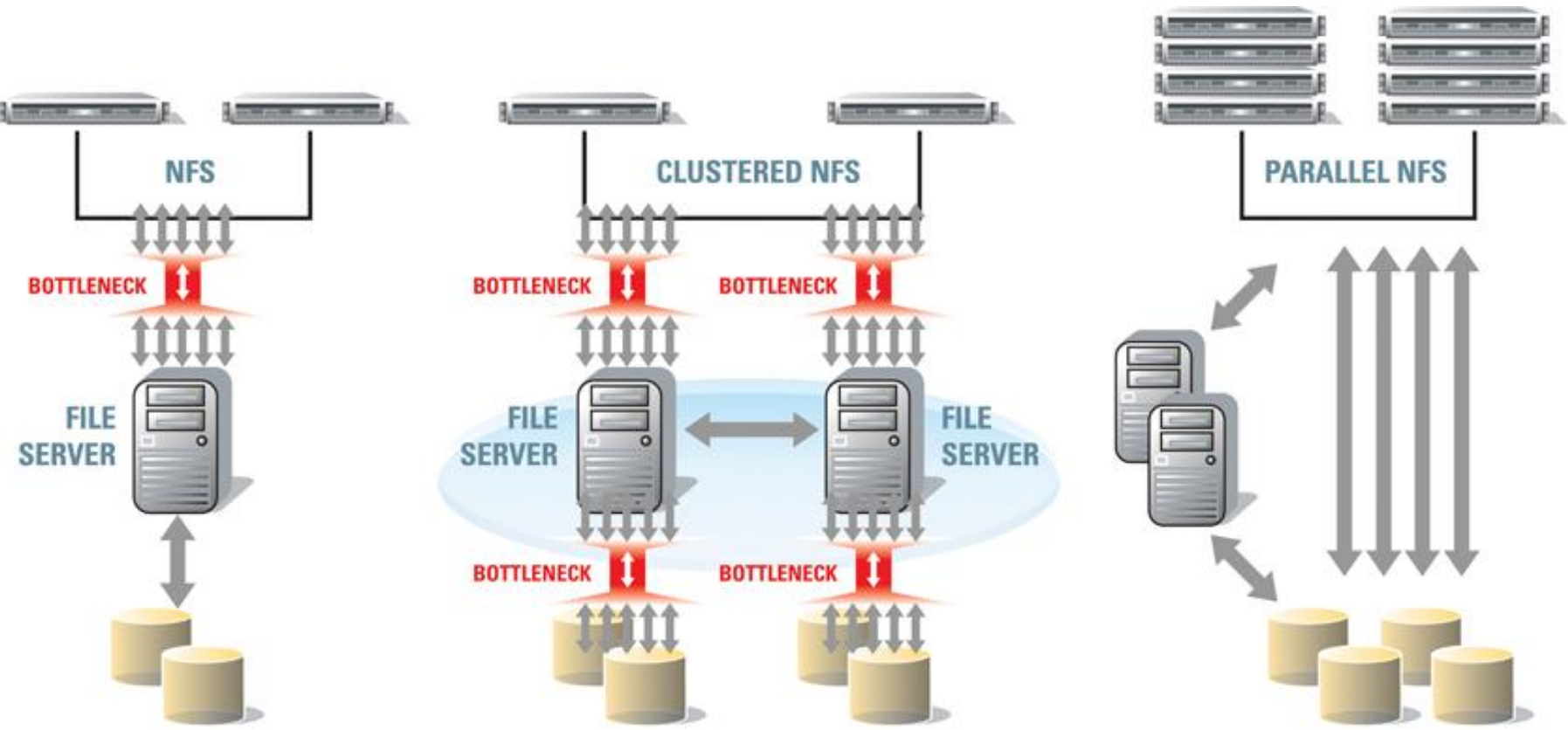
最新マイクロプロセッサ、最新インターフェイス、最新チップセットなどサーバプラットフォームの技術革新をシングルシステムとして提供可能

HPCアプライアンスソリューション

ストレージオプション

スケーラブルシステムズ株式会社

Panasasによるソリューション
パラレルNFSの実装



NAS

Network Attached Storage
シリアルI/Oがボトルネック

CLUSTERED STORAGE

複数のNASを統合的に運用管理
個々のNASサーバでのシリアルI/Oがボトルネック

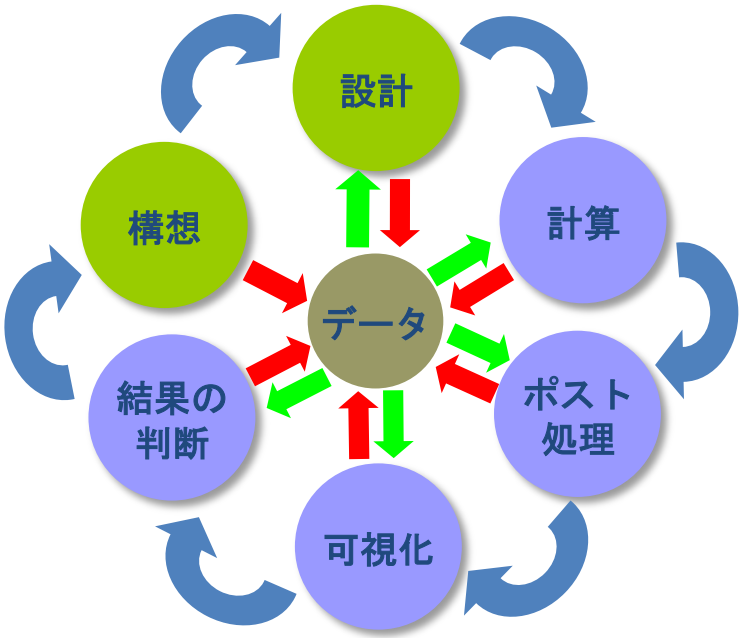
PARALLEL STORAGE

ファイルサーバを経由しないデータ転送パス
シリアルI/Oのボトルネックの解消と容易なシステム全体の運用管理

理想的なデータの流れの提案

- 一般的な業務の流れを考えてみると..
  - コンセプトの段階から、最終的な製品化の段階まで、データは、業務の流れの中心に位置する
  - 情報は、多くのグループで共有され、データは、ホストコンピュータ間で移動する
  - データセットのサイズは、各ステップ毎により大きくなる

もし、各ステップ毎で、データをコピーし、移動する必要がないのであれば、業務の効率を向上させ、本質的な問題の理解に時間を割ける



データセンターの課題

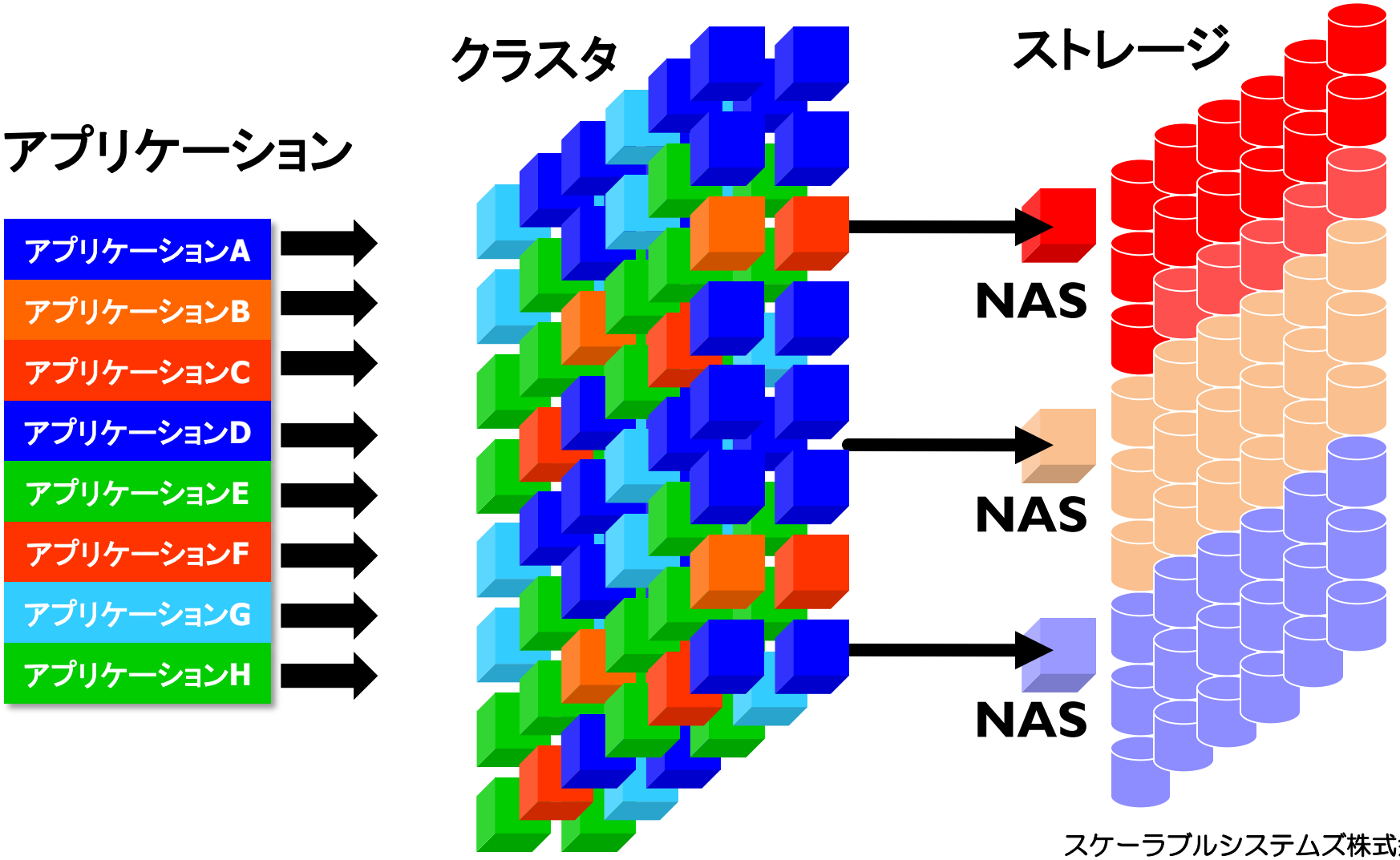
- 継続した性能向上の要求
  - クライアント数、クラスタノード数の増加
  - マルチコア化による計算リソースの強化
- 予測困難なストレージ利用量
  - 導入時に予測することは困難
- 24x7の連続稼働
- 障害からの迅速な復旧
- IT投資の最大活用
- アプリケーションの実行がシンプル

ストレージシステムの課題

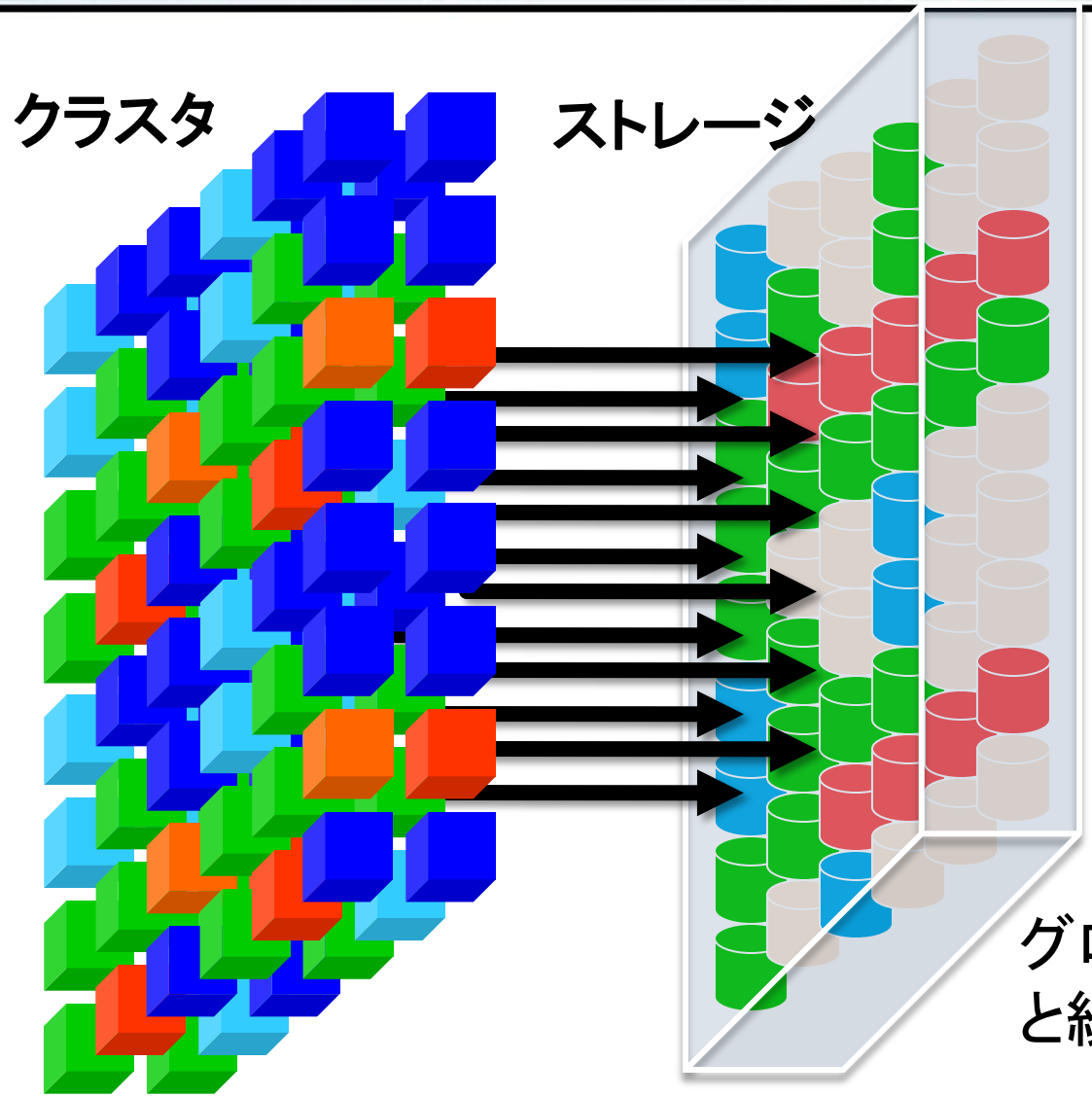
- データアクセスの問題
  - 複数のアーキテクチャ（ハードウエア、ソフトウエア）からのデータへのアクセス
  - 地理的にも分散したストレージの効率的な管理
- 管理運用の問題
  - データ移動を容易に行うことが可能であり、移動したデータに対して、透過的なアクセスが可能
  - ストレージの容量不足の場合などに容易に増設が可能
  - ユーザに負担をかけること無しで、ストレージの運用管理が可能

➡ **グローバルネームスペースによるソリューション**

NASストレージでのシステム構成



グローバルネームスペースによる　ソリューション

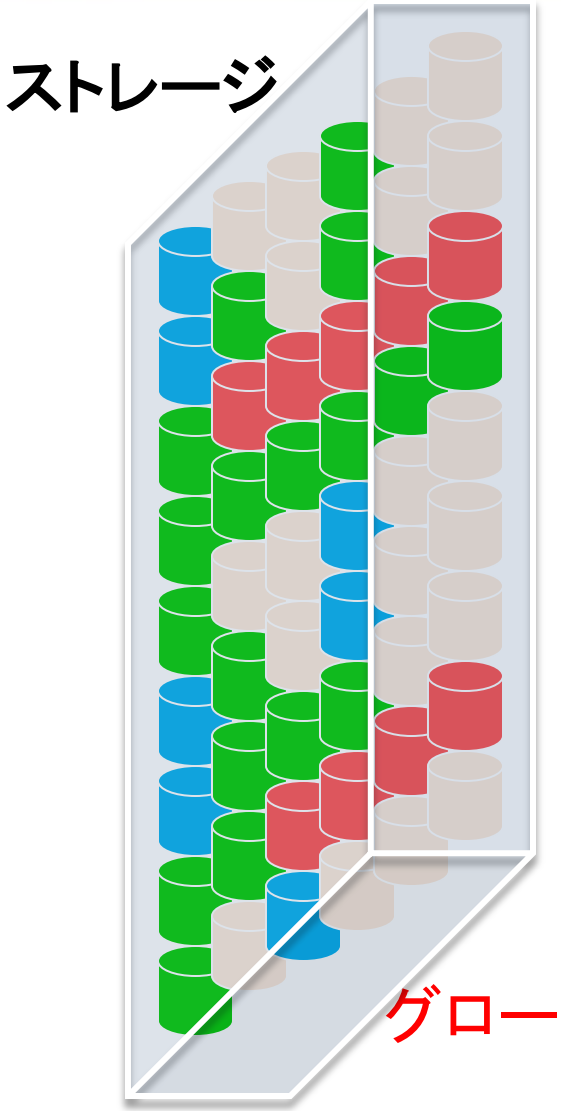


ストレージプール

- 単一の仮想リソース
- 透過的なデータアクセス
- システムの再構築が容易
- 様々なデータの格納が可能
- 可用性

グローバルネームスペースと統合された運用管理環境

グローバルネームスペース

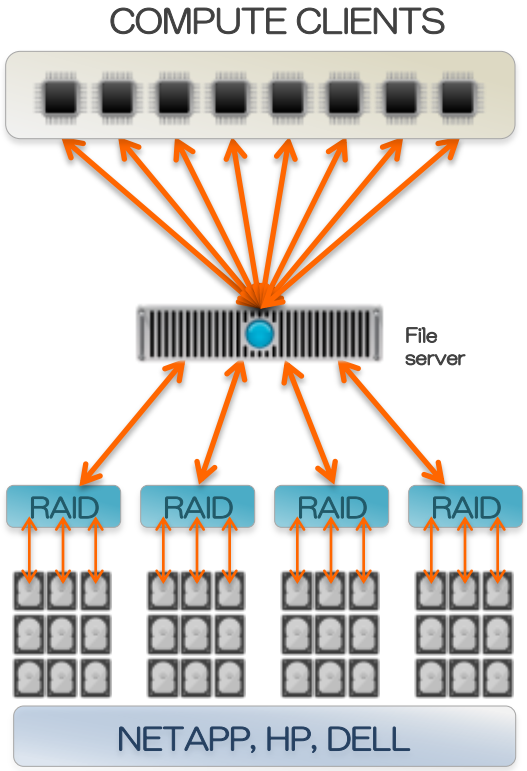


- 利用の簡便さ
  - 全クライアントが全データを見ることが可能
  - マウント・ポイント管理が不要
  - クライアント側の変更が不要
- 透過性
  - 容易な拡張
  - Failover
- スケーラビリティ
  - ネームスペースをペタバイトにまで拡張可能
  - 大規模ボリュームの容易な管理

グローバルネームスペースと統合された運用管理環境

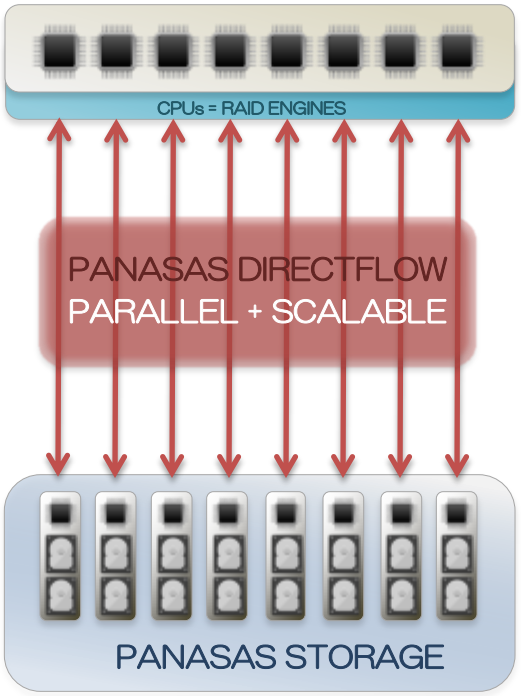
Panasas ActiveStor ストレージアーキテクチャ

Legacy Storage



- 従来型のファイルシステム
- シリアルなデータパス
- データパス上にメタデータサーバを配置
- RAID性能はファイルサーバがボトルネック
- ブロックベースのRAID構成- 全ドライブを再構成

Panasas ActiveStor



- パラレルデータパス
- パラレルデータパス
- メタデータサーバは、データパスの別（Out of Band）
- RAID性能はクライアント数に応じて増加
- オブジェクトRAID - ダメージファイルだけの再構成

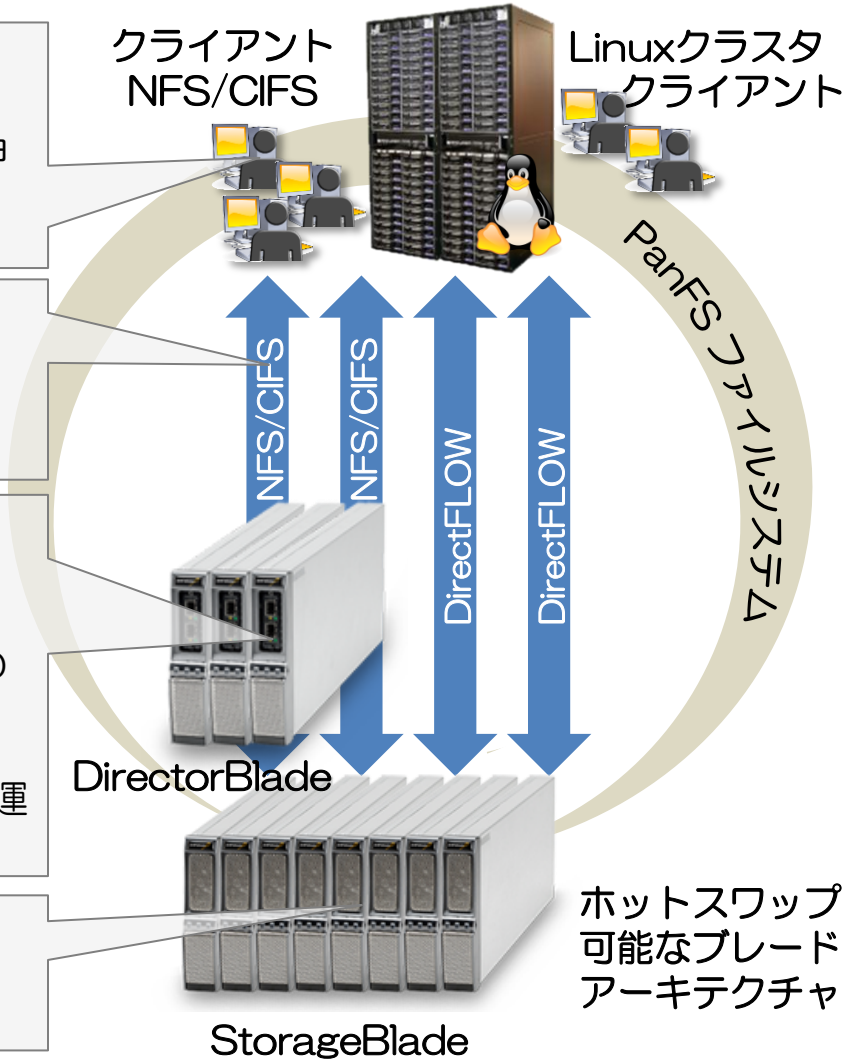
Panasasストレージクラスタ

DirectFLOW クライアントソフトウエア
- クライアントからの同時アクセスを並列に処理可能
- RedHat、SUSEなどの主要なLinuxディストリビューションで利用可能
- pNFSにも対応

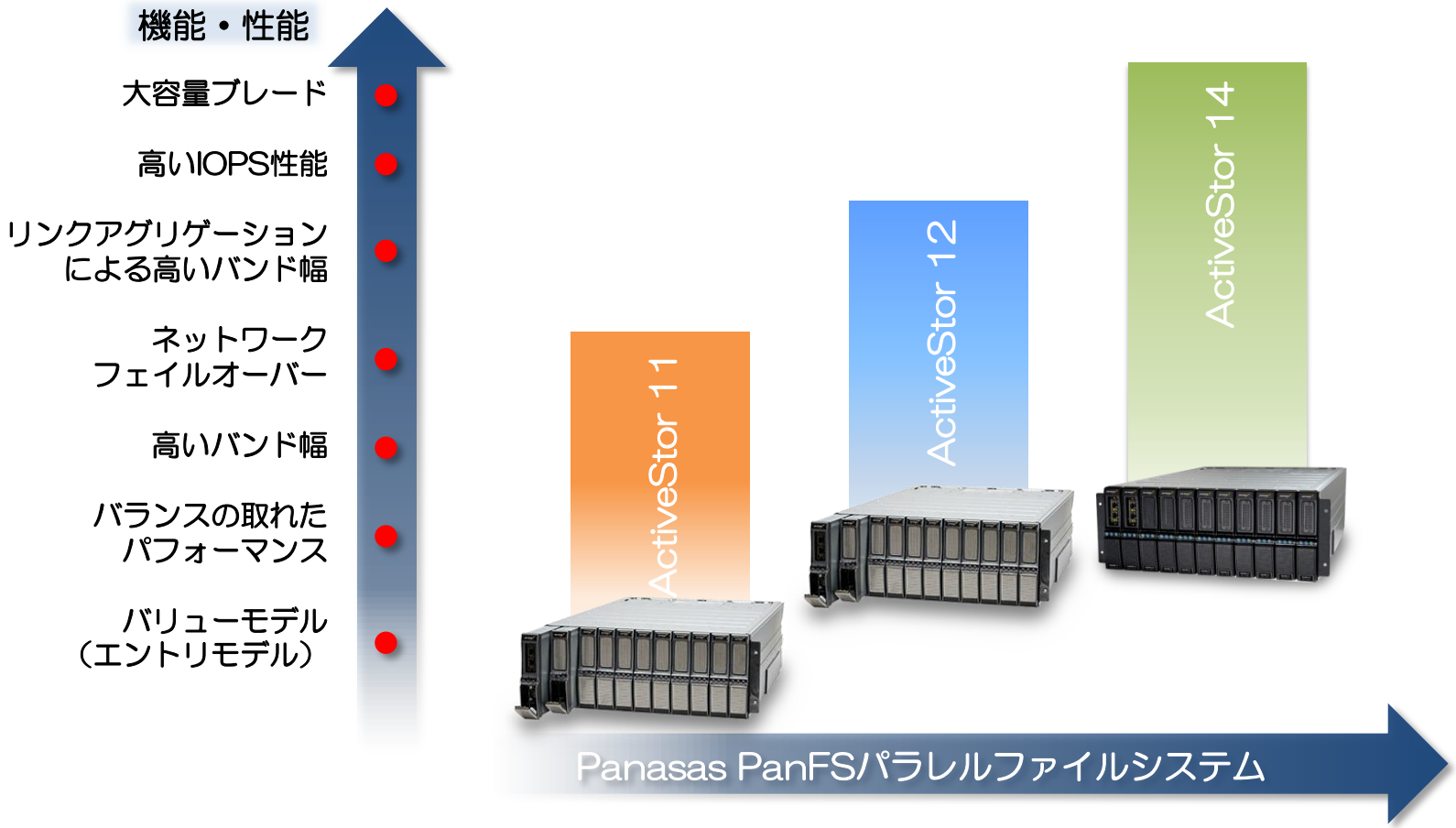
スケーラブルな NFS/CIFS/NDMPサーバ
- 負荷を自動的にストレージクラスタ全体に分散
- クライアント数の増加に合わせてスケーラブルな性能拡張
- 全てのDirectorBladeが全てのファイルにアクセス可能

シングルネームスペース
- 同一データへのいずれのプロトコルでのアクセスも可能
- シングルファイルシステム
- DirectFLOW/NFS/CIFS/NDMP間の完全なコヒレンシの実現
- 非Linuxのデバイスをシステムに統合
- グローバルネームスペースによるシステムの容易な拡張と運用の容易さ

オブジェクトベース
- 優れたスケーラビリティ、信頼性、運用管理
- Panasas Tiered Parityによるデータ保護の強化



Panasas ActiveStor製品ライン



ActiveStor 製品ライン

	ActiveStor 14	ActiveStor 12	ActiveStor 11
ActiveStor開発	第5世代	第4世代	第4世代
性能目的	スループット性能と IOP性能	スループット性能	スループット性能
容量 (TB)	83 / 81 / 45	60 / 40	60 / 40
SSD 加速オプション	Yes	No	No
スループット Write/Read (MB/sec)	1,600 / 1,500	1,600 / 1,500	950 / 1,150
4KB ファイル読み込み速度 (Read数/s)	14,150	1,350	1,300
メタデータ性能(Stat数/s)	14,150	2,750	2,650
キャッシュ (GB)	92 / 172	92	48
可用性 ネットワークフェイルオーバ	Standard	Standard	Optional
リングアグリゲーション	Yes	Yes	No

Note: A single 1+10 shelf configuration is assumed throughout except that the ActiveStor 14 4KB file reads/s and metadata performance numbers are based on a single 2+9 shelf of ActiveStor 14T (a 1+10 shelf of ActiveStor 14 would be lower than shown but still much higher than AS11 or 12). Volume fail-over was on for all configurations.

ActiveStor 14

- スループットとIOPS性能
- SATA+SSDのインテリジェンスな活用
- 複雑なワークロードに対応
- コストパフォーマンス

ActiveStor 11 & 12

- スループット
- **ActiveStor 11 はローエンドモデルとしてエントリーモデルの位置づけ**

スケーラブルシステムズ株式会社

ファイルサーバ

- 基本仕様
  - Xeonプロセッサ2台搭載
  - 大容量メモリメモリ
  - 冗長電源
  - I/O 拡張スロット
- オプション
  - 複数のSAS Raidコントローラ
  - 高速ネットワークオプション(InfiniBand/10GbE)
  - Windows/Linux双方をサポート

HPCアプライアンスソリューション

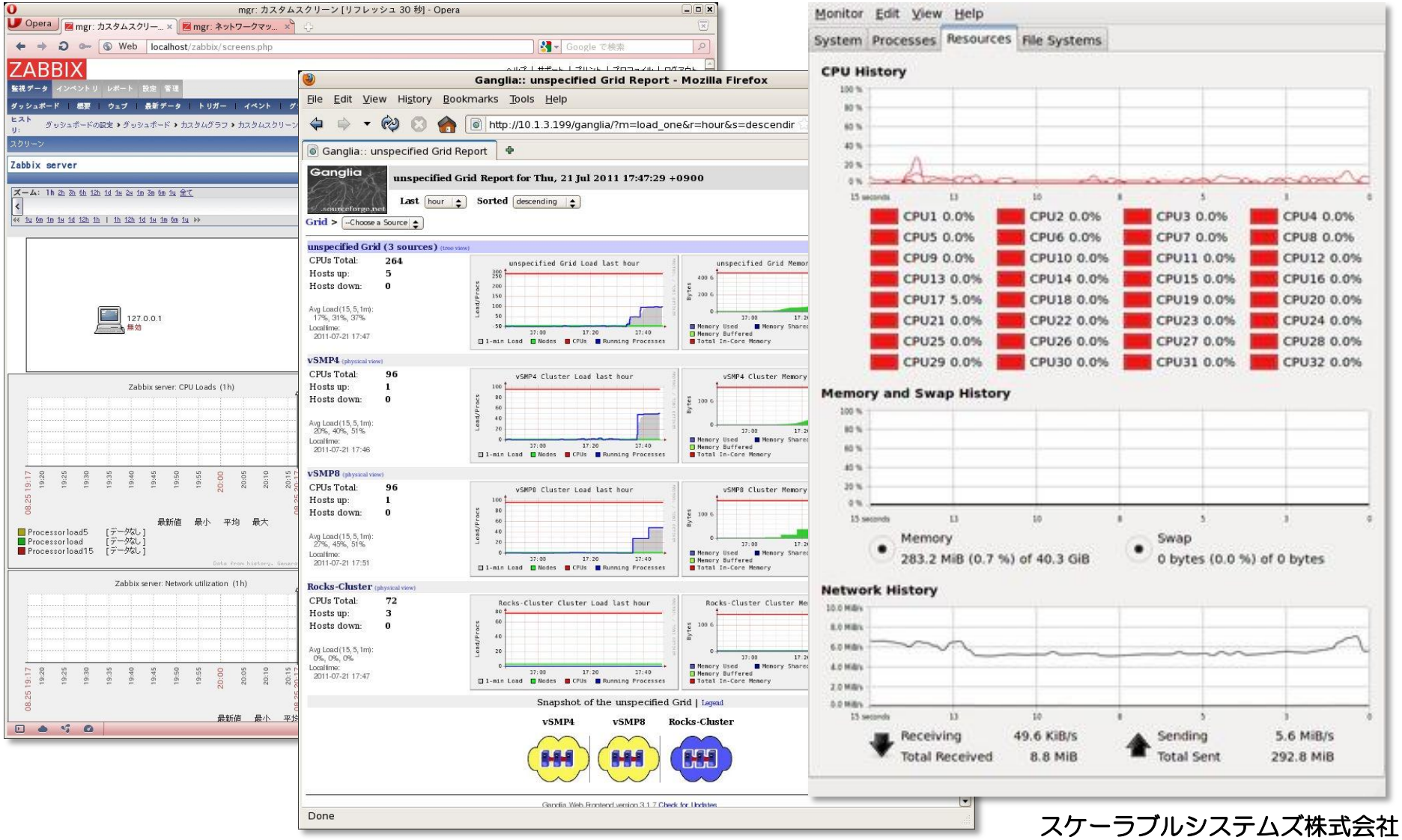
ソフトウエア

スケーラブルシステムズ株式会社

ソフトウエア環境
プリインストール

- オペレーティングシステム
  - 標準Linuxディストリビューション
- Panasasストレージクラスタ
  - ActiveScale 4.0 オペレーティングシステム
- 開発環境
  - インテルソフトウエア
- クラスタ管理
  - Rocks
  - オープンソースクラスタソフトウエアバッチシステムなども含む

システムモニターや管理ツール
標準的なツールの利用が可能



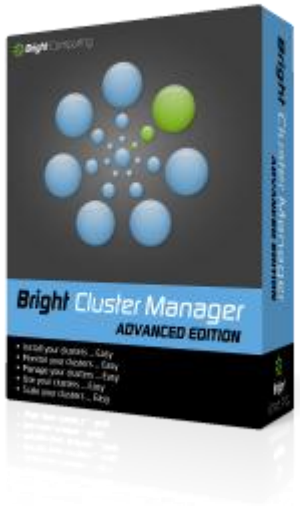
スケーラブルシステムズ株式会社

ソフトウエア環境
プリインストール

サービス	内容
os	CentOS/Red Hat Enterprise Linux など　（Linux)
Ganglia(*)	Cluster monitoring system from UCB
intel	Intel Developer Tools
Torque(*)	Torque+Maui job queueing system
PanFS(*)	Panasas PanFS ファイルシステム（DirectFlowクライアント）

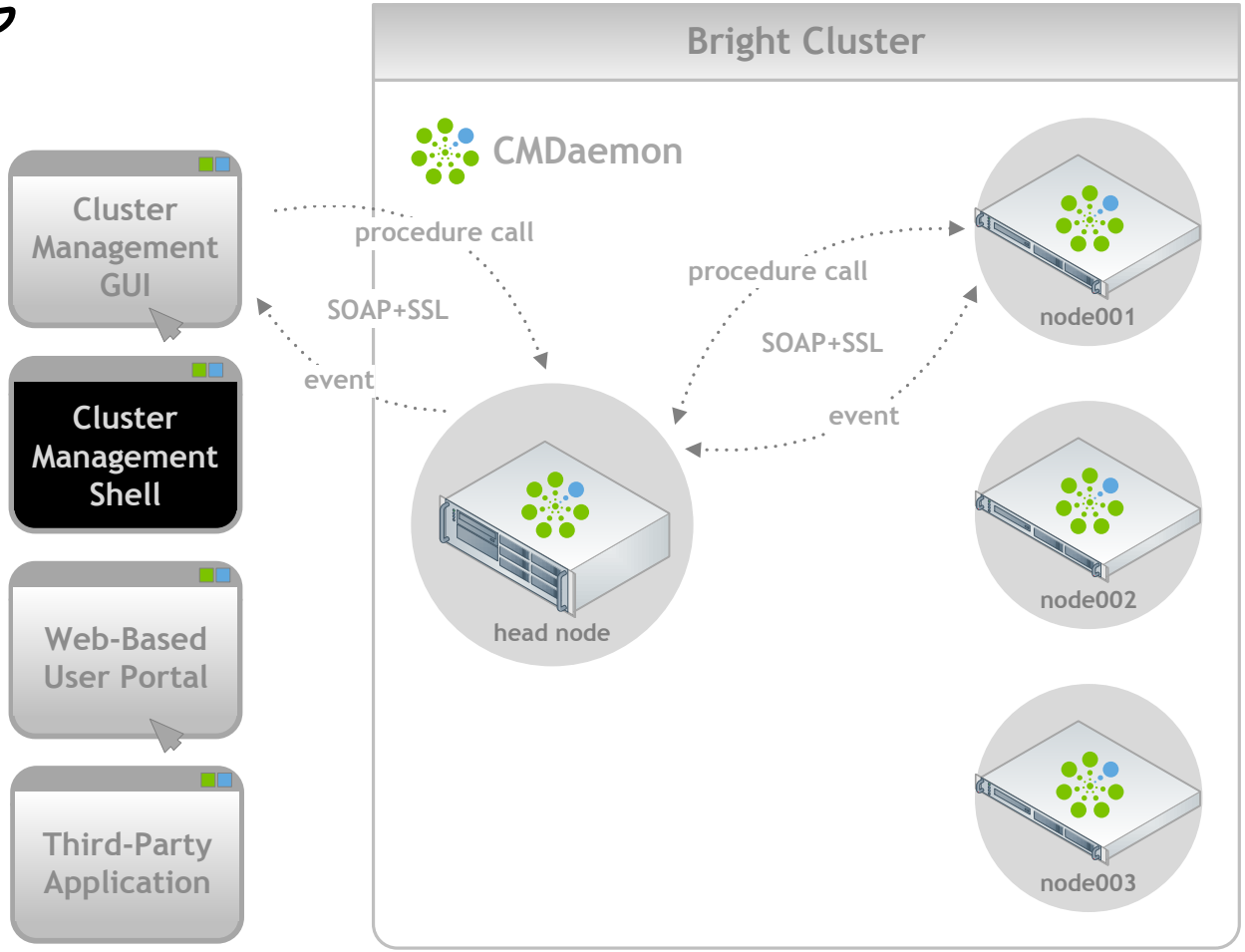
vSMPマネージメントソフトウエア

- Bright Computing – Bright Cluster Manager – Advanced Edition
  - 統合クラスタ管理ソフトウエア
    - ノードプロビジョニング
    - システムモニター
    - ワークロードマネージメント
    - ユーザポータルサービス
    - GUI & CLI サポート
  - vSMP Foundation サポート
    - プロビジョニング – GUIベース
    - システム管理・モニター

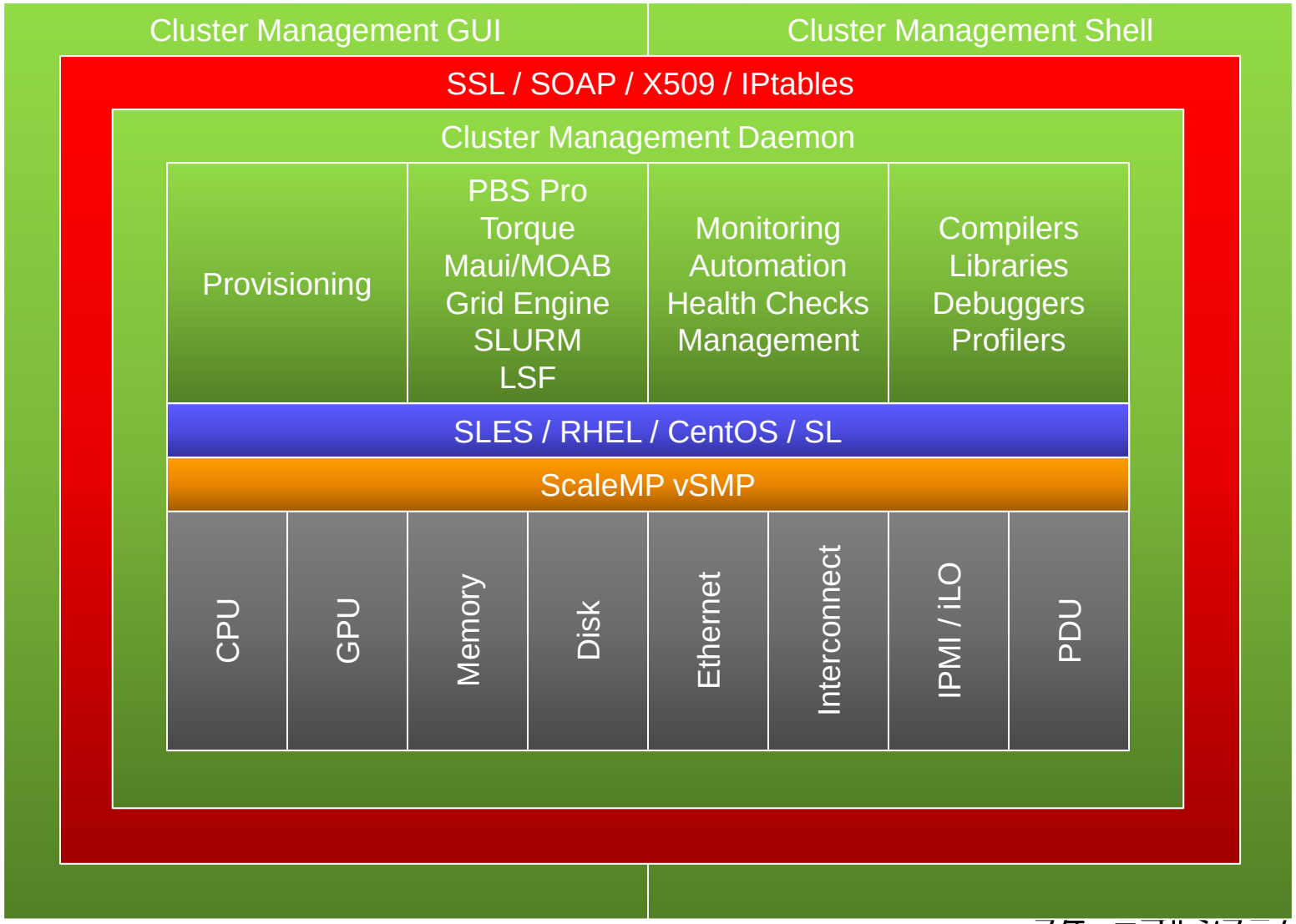


Bright Cluster Manager

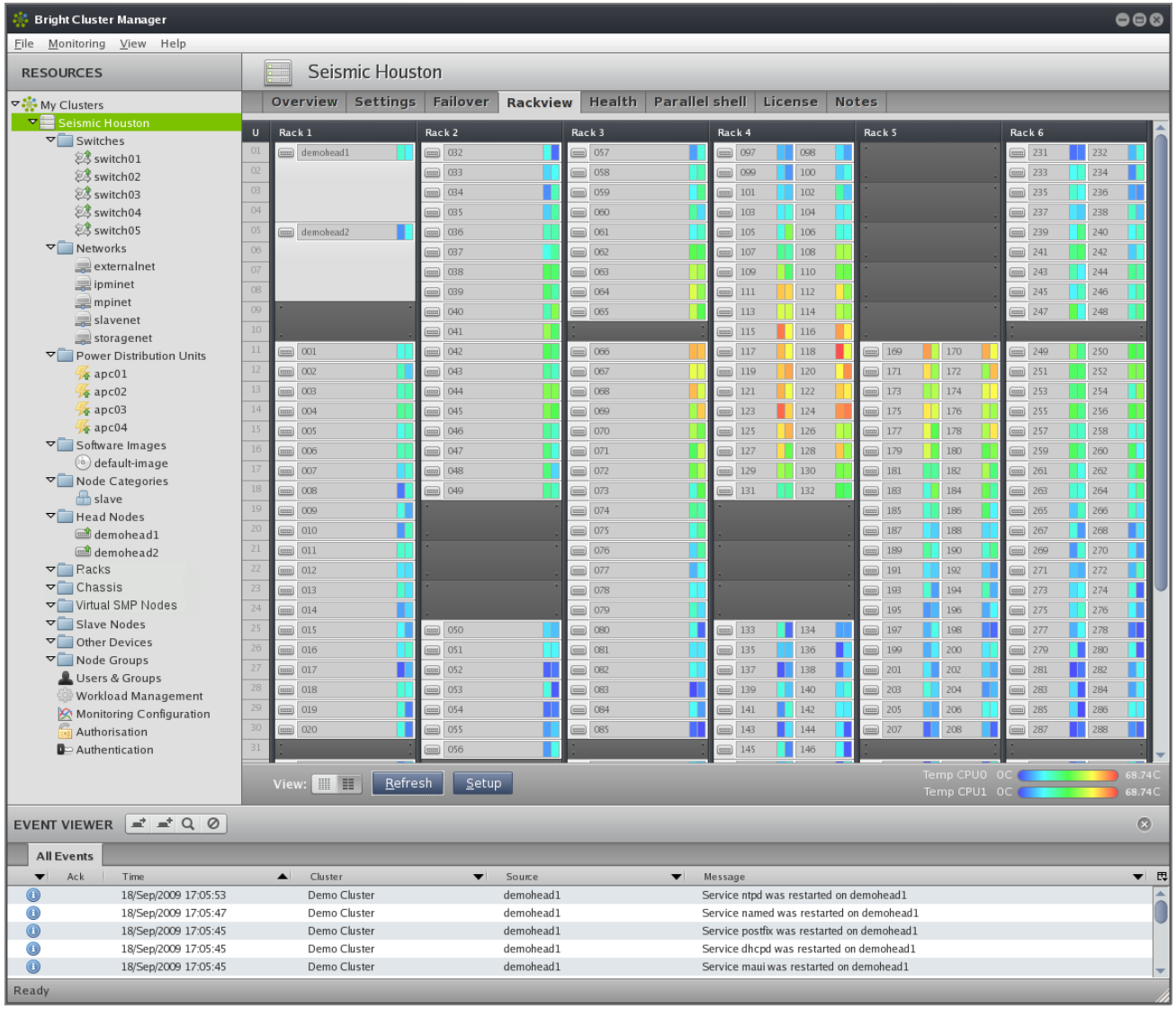
- アーキテクチャ



Bright Cluster Manager

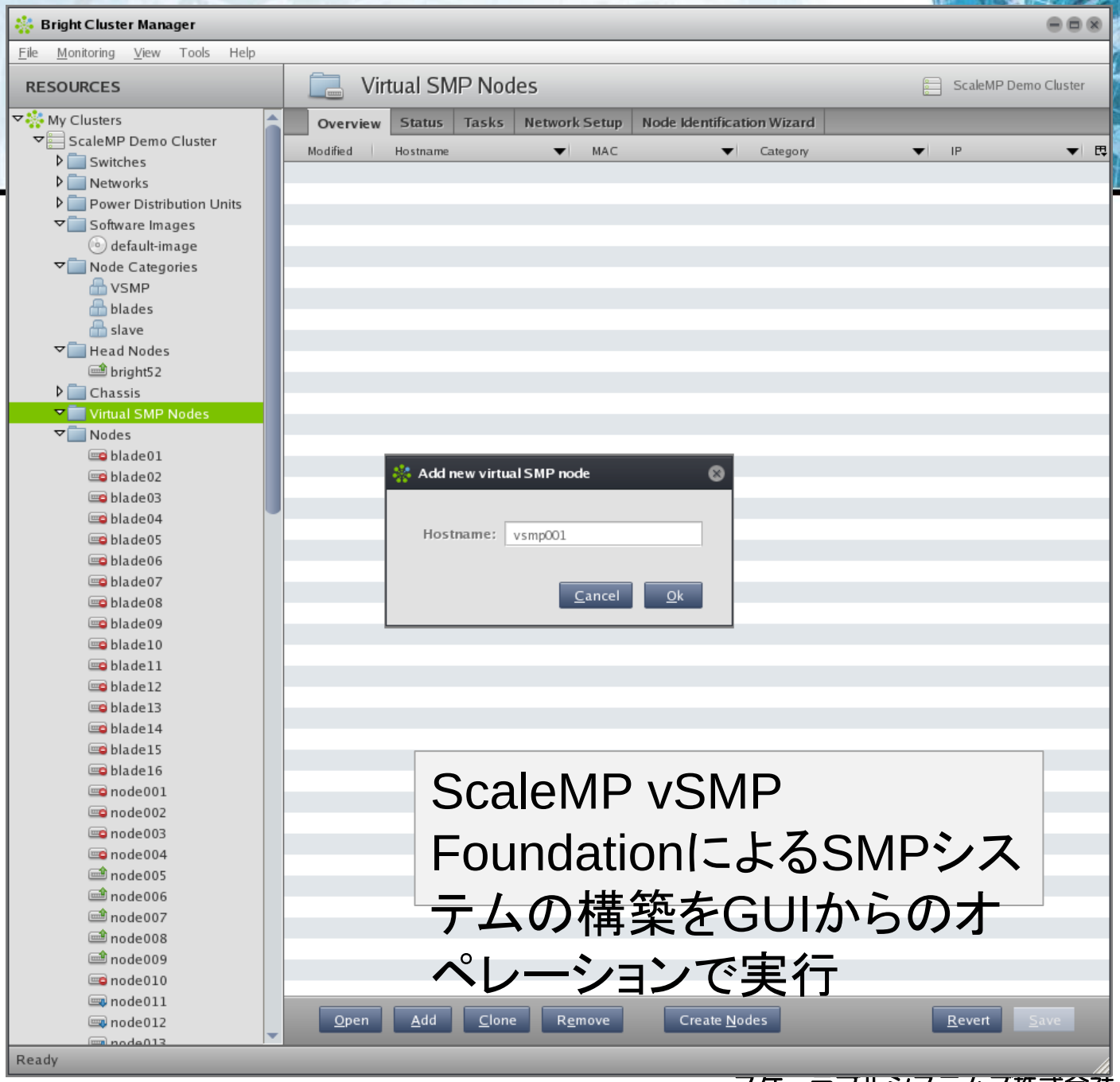


Bright Cluster Manager



クラスタモニター
ラック毎、機材毎
のモニター画面

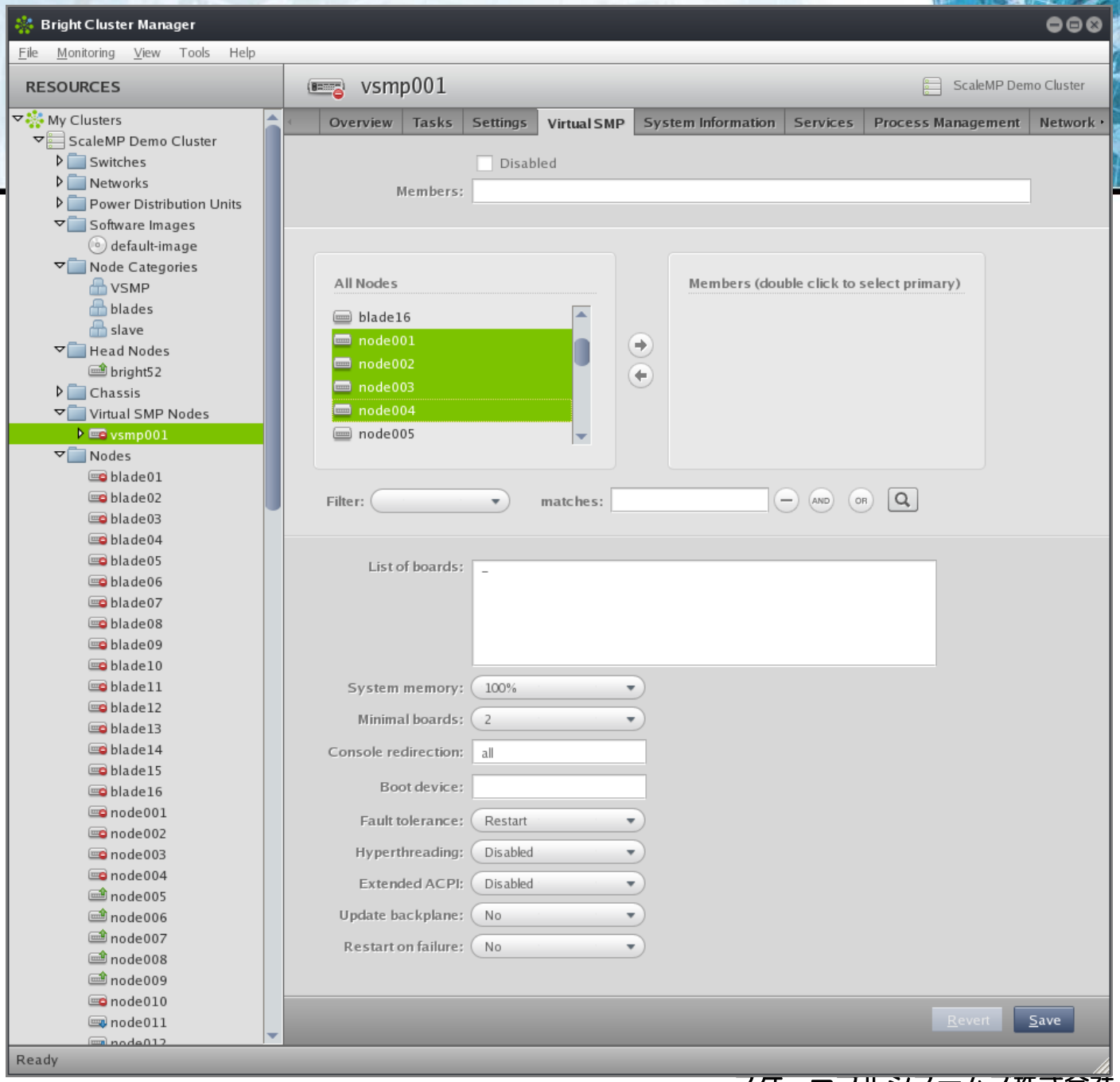
様々なメトリック
スでのシステム
管理とモニターが
可能



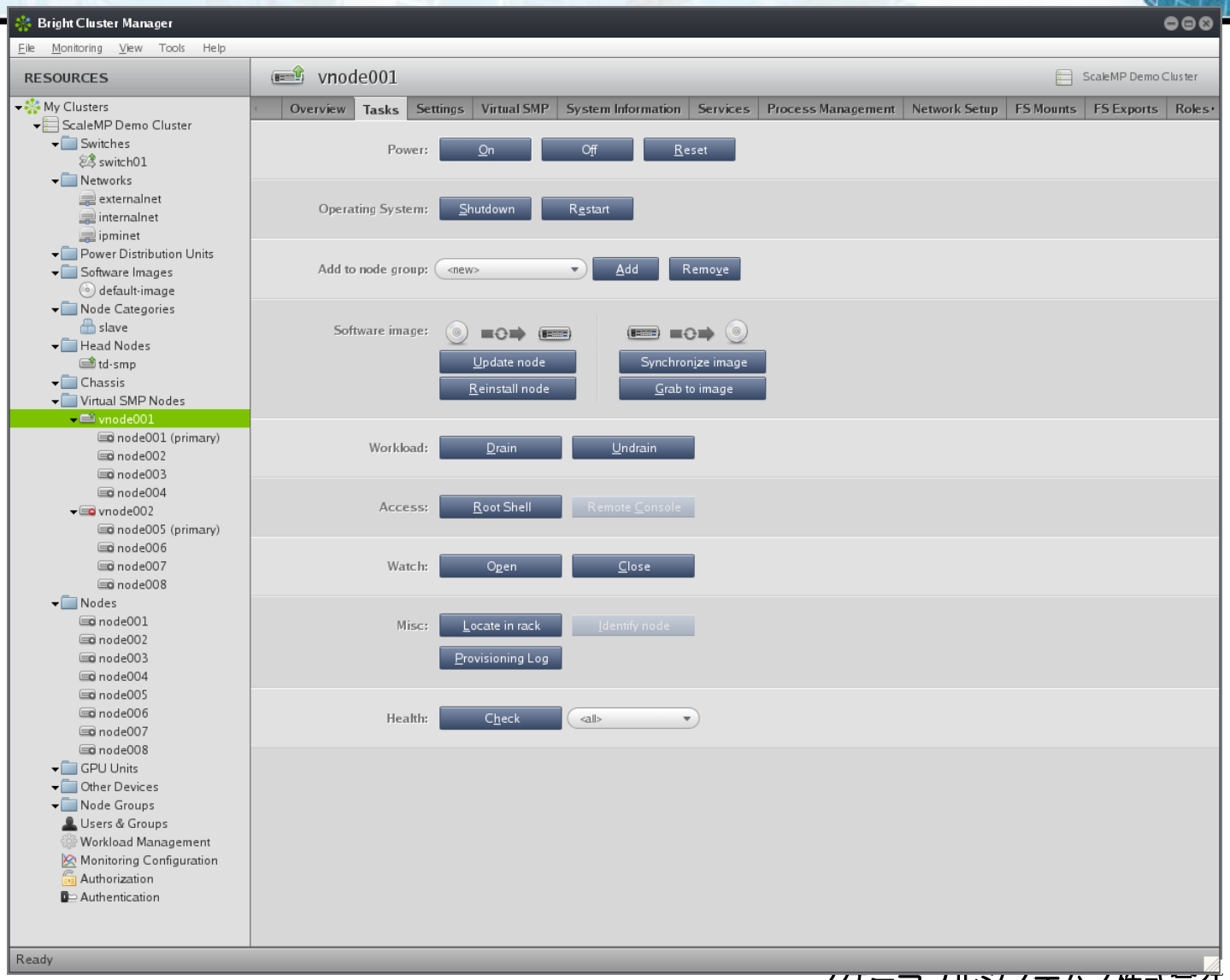
ScaleMP vSMP FoundationによるSMPシステムの構築をGUIからのオペレーションで実行

vSMP Foundationでの起動時のオプションを選択可能

vSMP Foundationでの起動時のオプションを選択可能



vSMP Foundationでの構築したSMPシステムの管理機能



お問い合わせ

0120-090715

携帯電話・PHSからは（有料）

03-5875-4718

9:00-18:00 (土日・祝日を除く)

WEBでのお問い合わせ

www.sstc.co.jp/contact

この資料の無断での引用、転載を禁じます。
社名、製品名などは、一般に各社の商標または登録商標です。なお、本文中では、特に®、TMマークは明記しておりません。

In general, the name of the company and the product name, etc. are the trademarks or, registered trademarks of each company.
Copyright Scalable Systems Co., Ltd. , 2011. Unauthorized use is strictly forbidden.