



プライベートクラウドのための 仮想化ソリューション

スケーラブルシステムズ株式会社

クラウド環境

“ハードウェア（物理マシン）とアプリケーションの対応関係を切り離し、アプリケーションを実行するハードウェアを必要に応じて変更できるのが、プライベート・クラウドを初めとするクラウド環境の大きな特徴である。…プライベート・クラウドの場合、どの物理サーバがどのアプリケーションを実行するのか、あらかじめ決められているわけではない。アプリケーションを実行する時点で、リソースに余裕のある物理サーバでアプリケーションを実行するのである。

プライベート・クラウドには、ITオペレーションを担当する人々だけでなく、アプリケーションを利用したいと希望するビジネス・ユーザーも関与する。ビジネス・ユーザーが希望するタイミングで、人手を介することなく迅速にサーバをプロビジョニングし、アプリケーションを実行するための環境を提供するのである。”

Computerworld DeepDrive のクラウド関連記事より抜粋

仮想化？

仮想化とは、コンピュータにおいてリソースの抽象化を指す用語である。「リソースの物理的特性を、そのリソースと相互作用するシステム/アプリケーション/エンドユーザーから隠蔽する技法。単一の物理リソース(サーバ、OS、アプリケーション、補助記憶装置など)を複数の論理リソースに見せかけたり、複数の物理リソース(複数の補助記憶装置群やサーバ群)を単一の論理リソースに見せかけたりできる」という実用的定義がある。また、単一の物理リソースを何らかの特性の異なる単一の論理リソースに見せかけることもできる。

ウィキペディア日本語版

仮想化？

.....単一の物理リソース（サーバ、OS、アプリケーション、補助記憶装置など）を複数の論理リソースに見せかけたり、複数の物理リソース（複数の補助記憶装置群やサーバ群）を単一の論理リソースに見せかけたりできる」という実用的定義がある。

コンソリデーション

単一の物理リソース（サーバ、OS、アプリケーション、補助記憶装置など）を複数の論理リソースとして提供

利用効率

(ディスクのパーティション)

柔軟性

(ネットワークのVLAN構成)

リソース活用

(サーバ仮想化)



Storage



Networking



Server

アグリゲーション

複数の物理リソース（複数の補助記憶装置群やサーバ群）を単一の論理リソースとして提供

柔軟性と拡張性

(RAIDとボリューム管理)

可用性と性能

(リンク・アグリゲーション)

リソース拡張と活用

(ハイエンド仮想化)

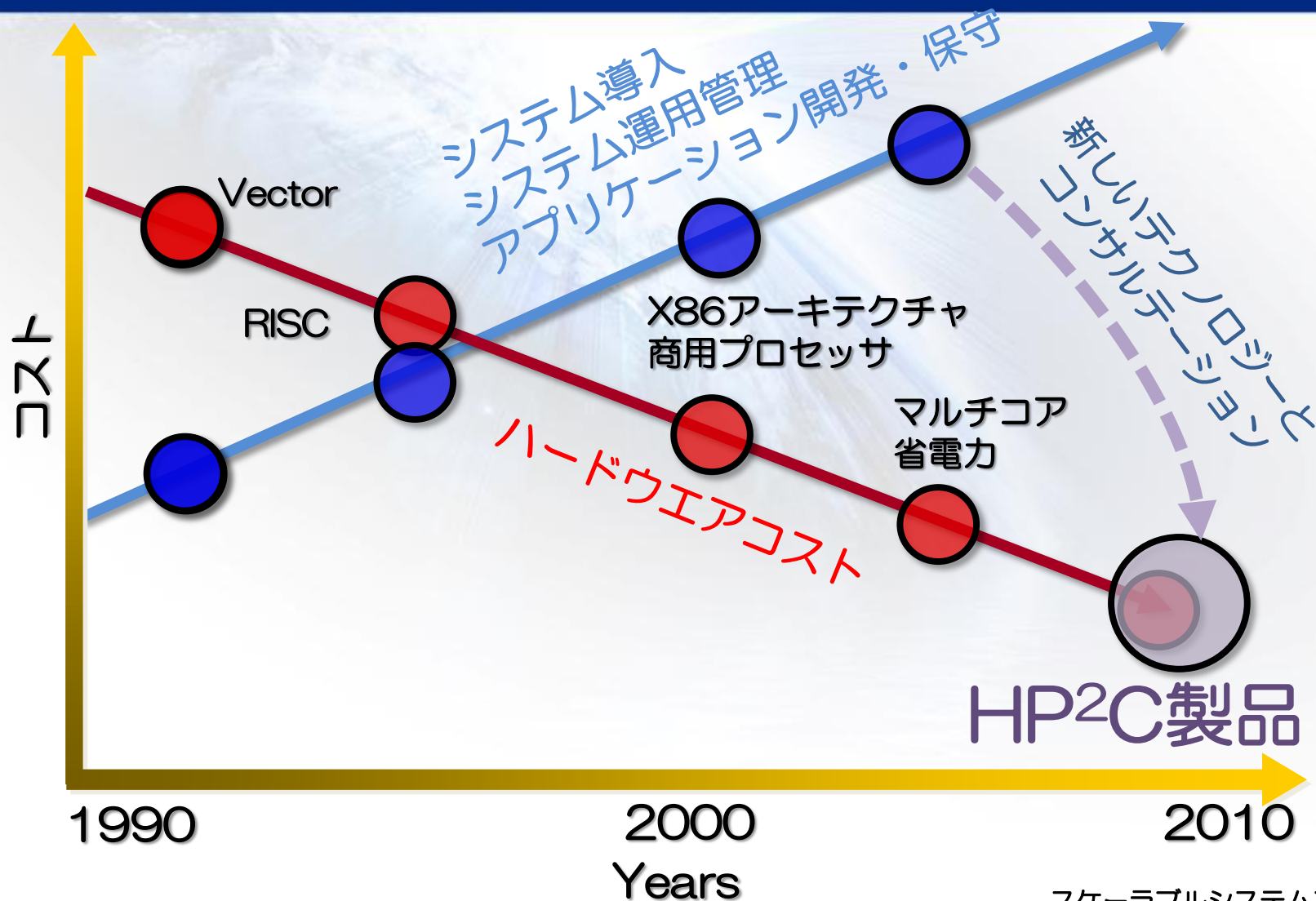
仮想化

- ・ 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とCAPEXの削減を目的



HPCシステムのTCOトレンド

TCO=ハードウェアコスト+導入運用管理コスト



HPCシステムのギャップ

SMP (Shared Memory Systems)

【利点】

運用管理が容易
身近で利用可能
容易なプログラミング

【課題と問題点】

スケーラビリティ
開発コスト
システムの互換性

HP²C製品:

SMPとクラスタのギャップを
埋める製品ソリューション

クラスタシステム

【利点】

高いスケーラビリティ
ハードウェアの導入コスト

【課題と問題点】

システム構築のスキルの要求
高い運用管理コスト
複雑なオペレーション環境
煩雑なインストール
やアップグレード

ワークステーション
サーバ

クラスタ

#Processors

2

4

8

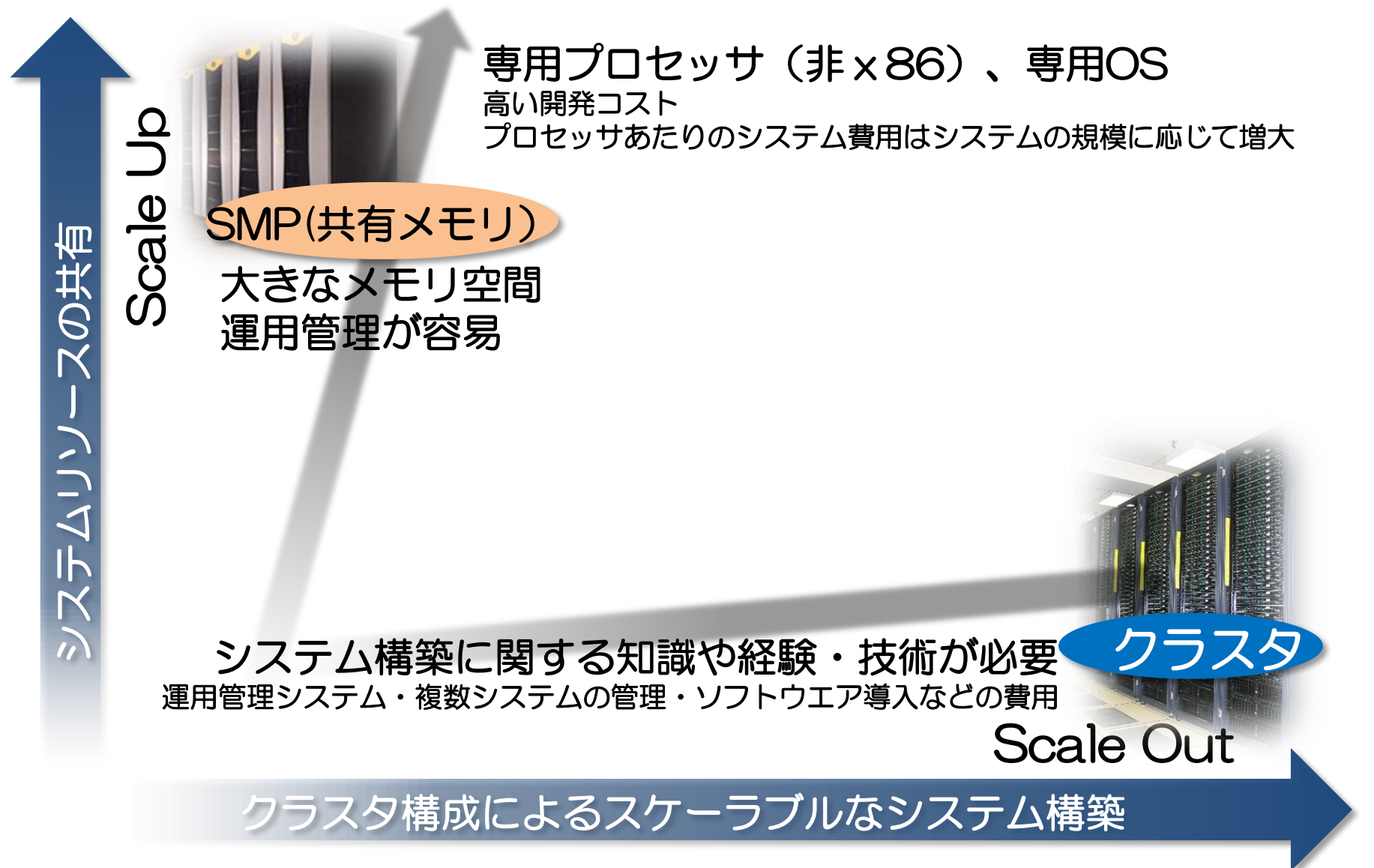
16

32

64

128

コンピューティングの課題



システムの複雑さ = f(システムの規模) ？

クラスタを利用すること、また、その管理運用を行うことは現在でも多くの課題がある

- ・ 電力、冷却、設置スペースは大きな問題
- ・ 全てのレベルでのインターコネクトの性能の問題
- ・ アプリケーション及び自作プログラムの並列化
- ・ 複数ノードを利用しての高い性能の実現
- ・ RAS (信頼性、可用性、サービス性) の重要度
- ・ ヘテロな計算環境に対する取り組み
- ・ ソフトウェアコストの問題
- ・ ストレージとデータマネージメントでのボトルネックが顕在化
- ・
- ・
- ・

システムの複雑さ

クラスタシステムの規模

クラスタ



プライベートクラウド 社内クラウドもしくは企業クラウド

- ・ホステドサービスをファイアウォール内の特定のグループに提供するコンピューティングアーキテクチャ
- ・仮想化、自動化、そして分散コンピューティングを活用してオンデマンドの柔軟なコンピューティング機能を社内ユーザーに提供
- ・クラウドスタイルのシステム導入に移行することで、既存の導入よりコストを削減することが可能

プライベートクラウドでの技術特性

仮想化

- 資産の高利用率

融通性

- 設備投資のないダイナミックなスケーリング

自動化

- 人の介入がない構築、導入、コンフィギュレーション、プロビジョニング、および移動

	Panasas ストレージクラスタ	vSMP Foundation
仮想化	● グローバルネームスペースによる異機種混在環境での、ストレージのロケーションに関係ないデータアクセス	● ScaleMのvSMP Foundation による仮想化サーバ ● 複数の物理的なサーバをひとつの論理サーバとして利用可能
融通性	● ユーザのアクセス方法や利用方法に影響を与えることなく、ストレージの拡張や移動が可能 ● システムの再構成などをオンライン中に実行することが可能 ● ダウンタイムを最小化することを可能としデータの管理業務における物理的な作業を大幅に低減	● マイクロプロセッサやチップセット、インターコネクトについて、常に最新技術を利用可能 ● シングルシステムとして運用利用が可能であり、そのオペレーションコストの削減が可能
自動化	● グローバルネームスペースによるシステム全体のデータ分析とそのレポート ● 高度な自己マネージメント機能と予防的な対障害対策の実施	● オペレーティングシステムやアプリケーションに変更を加えることなく利用可能 ● 2ソケットや4ソケットのサーバシステムをより多くのソケットが利用可能なシステムに拡張 ● シングルシステムイメージによるシステムの運用管理業務の簡便化とシステムのシンプル化

プライベートクラウドのための 仮想化ソリューション

仮想化(バーチャライゼーション)プロセッサやメモリ、ディスク、通信回線など、コンピュータシステムを構成する資源(および、それらの組み合わせ)を、物理的構成に拠らず柔軟に分割したり統合したりすること

「サーバ仮想化」

- ・ 1台のサーバコンピュータをあた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

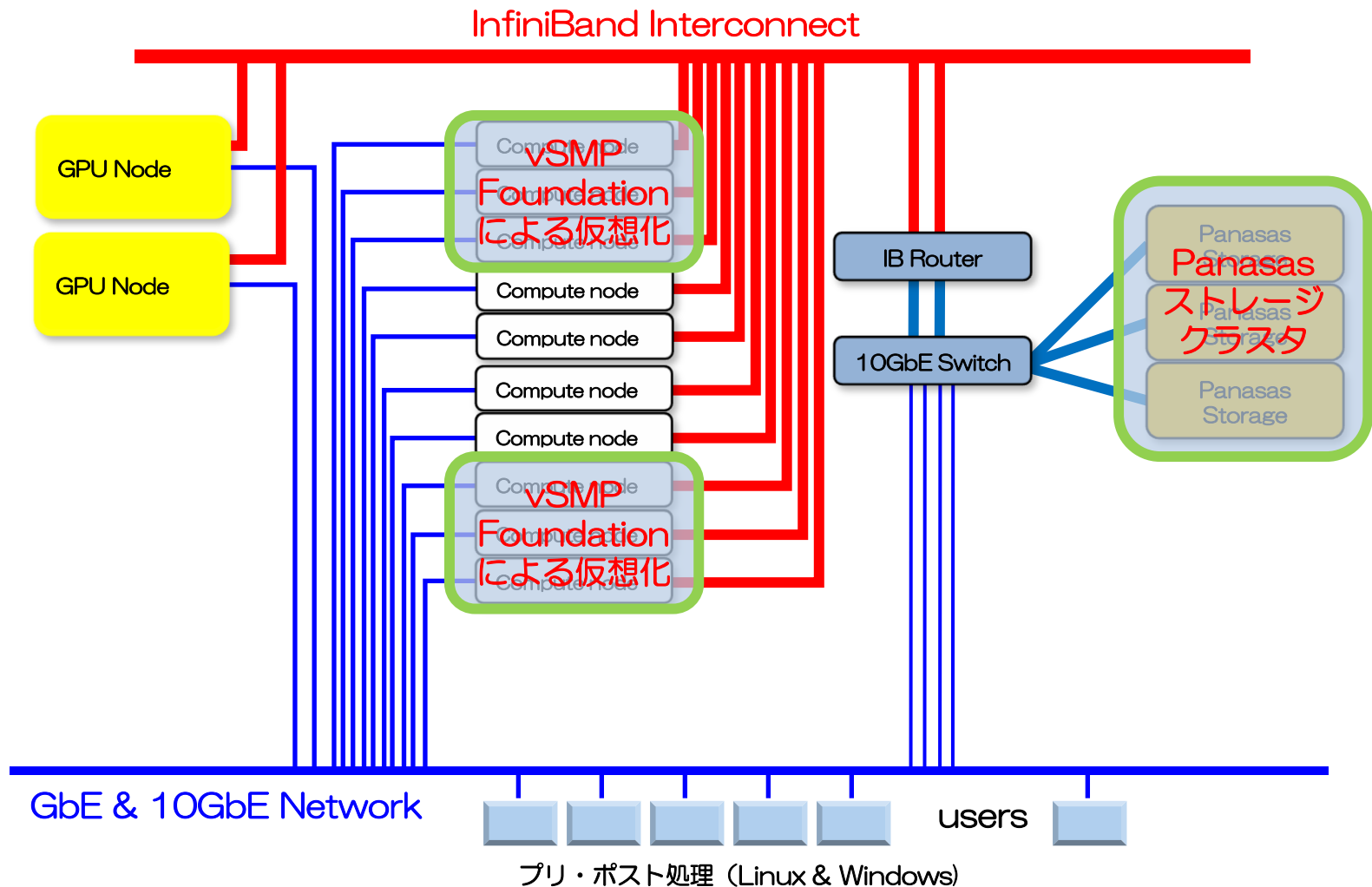
「ストレージ仮想化」

- ・ 複数のディスクをあたたかも1台のディスクであるかのように扱い、大容量のデータを一括して保存したり耐障害性を高めたりする技術

「ハイエンド仮想化」

- ・ 複数の物理的なサーバをひとつの論理サーバとして利用可能とするサーバ仮想化技術です。サーバ仮想化が1台のサーバで複数のサーバに仮想化によって分割するのと反対に複数システムを1台のシステムに統合する

プライベートクラウドのための 仮想化ソリューション



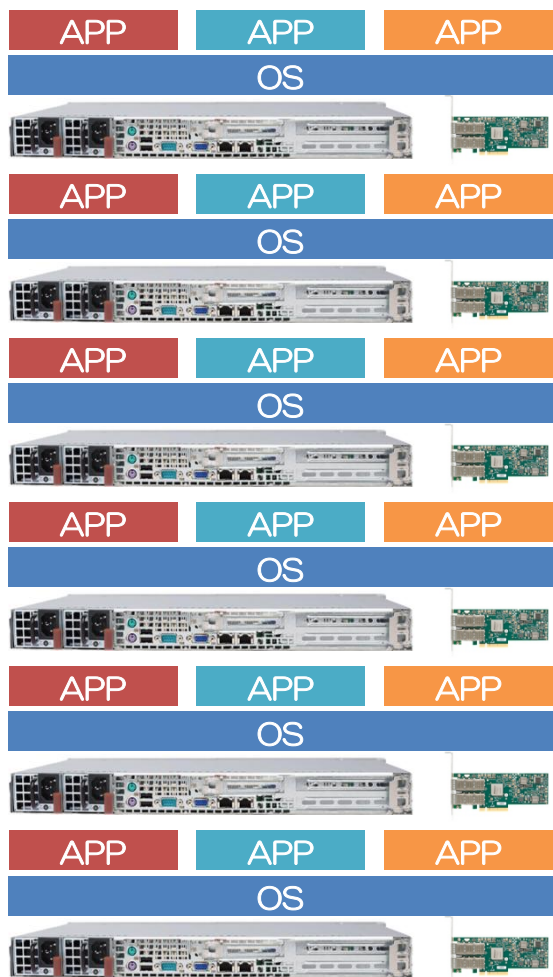


ハイエンド仮想化ソフトウェア

vSMP Foundation

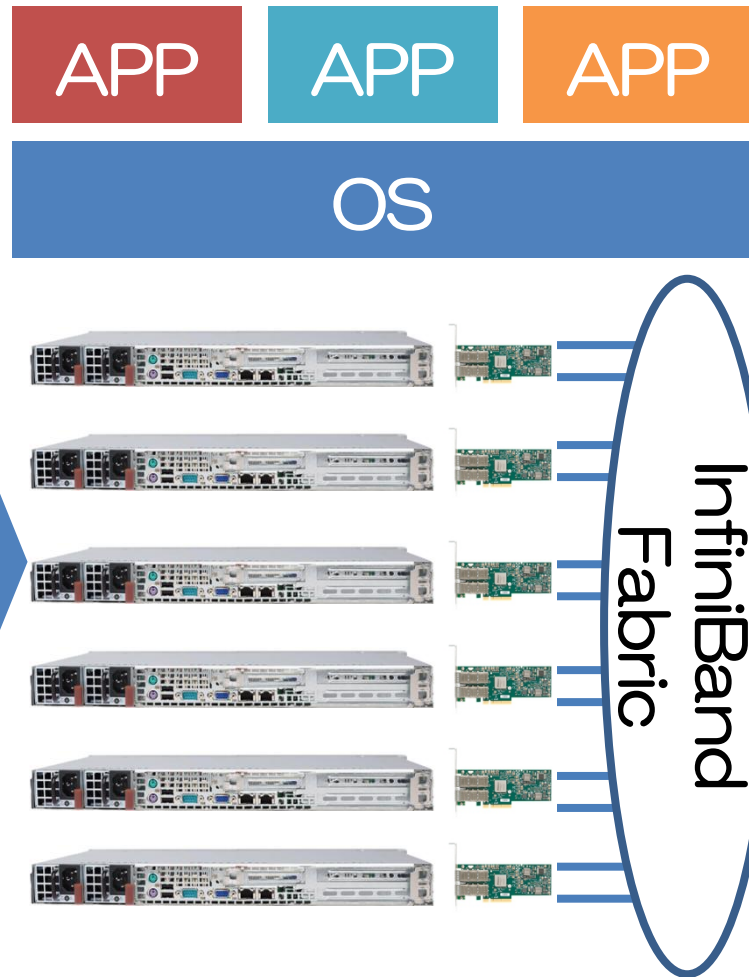
vSMP Foundation 仮想化

複数のコンピュータ
複数のオペレーティングシステム



vSMP
Foundation

複数のコンピュータ
シングルオペレーティングシステム



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
- ソフトウェ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



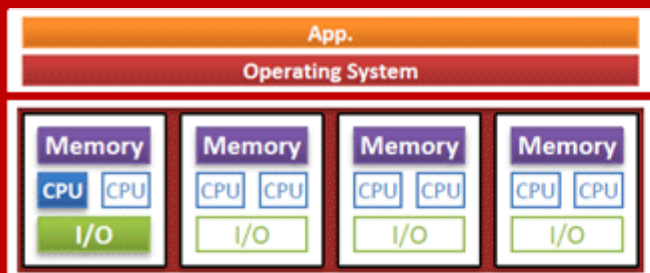
- すべてのIOリソースに対して、シングルOSがアクセス可能
- シングルファイルシステムを実現

共有IO

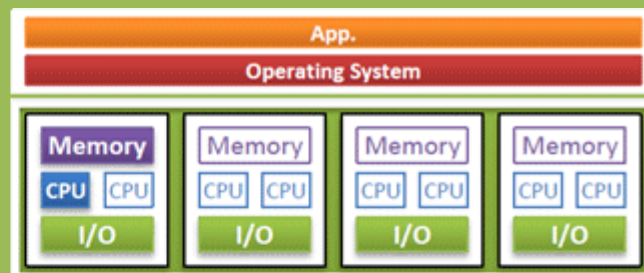


ハイエンド仮想化の目的は？

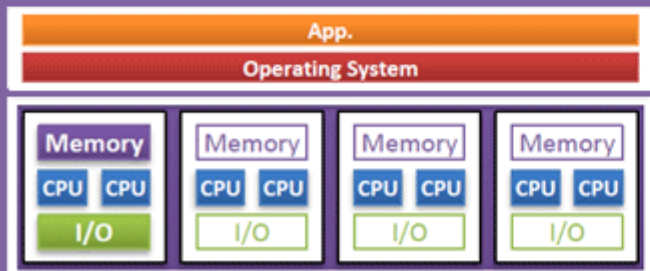
大規模メモリ空間



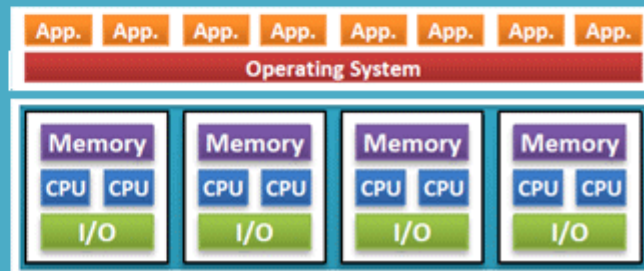
I/O処理能力



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

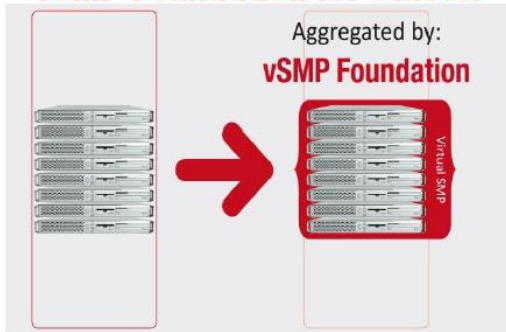
システム統合

vSMP 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

vSMP Foundation for Cluster



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

vSMP Foundation for SMP

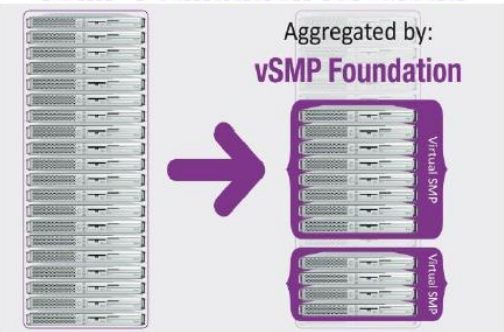


CAPEXの削減

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)

vSMP Foundation for Cloud



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)

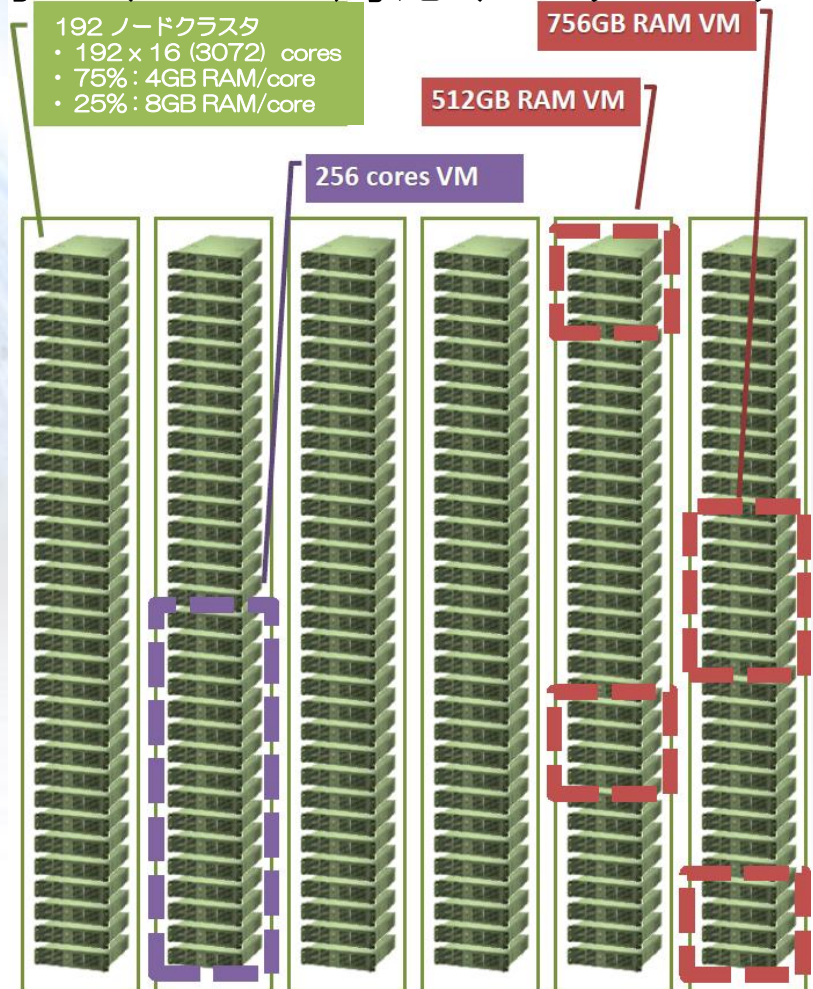
スケーラブルシステムズ株式会社

シングル インフラストラクチャ

OPEX&CAPEXの削減

- ・ 同じハードウェア基盤とソフトウェアを利用して、
- ・ 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

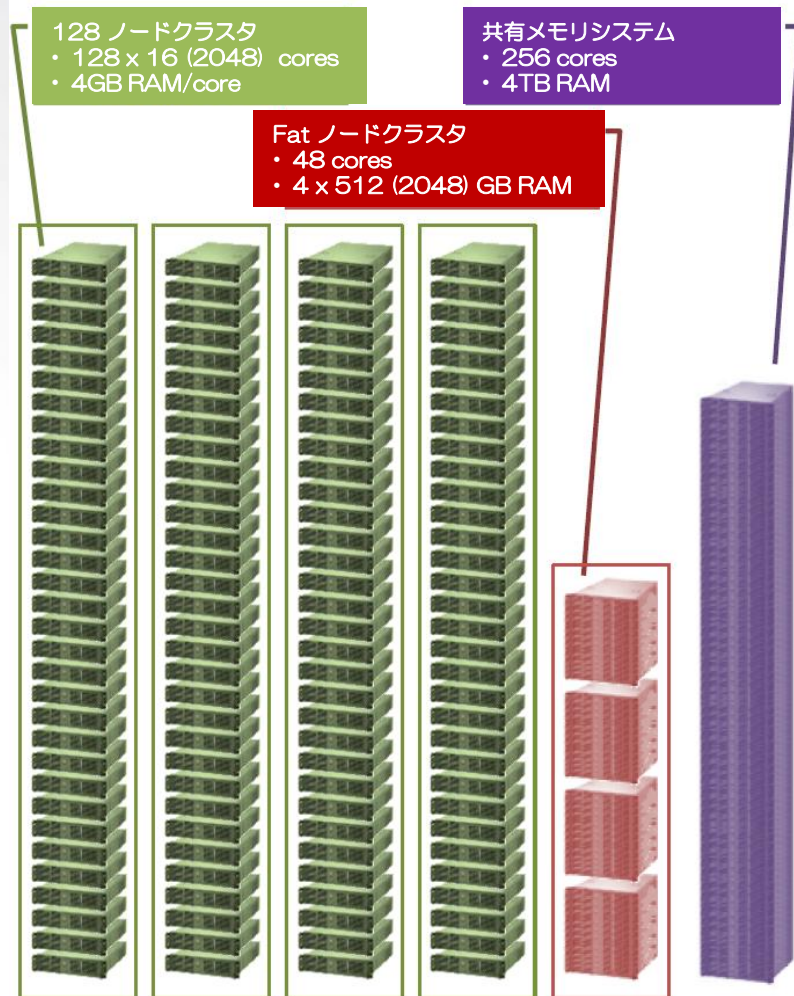
オンデマンド対応データセンター



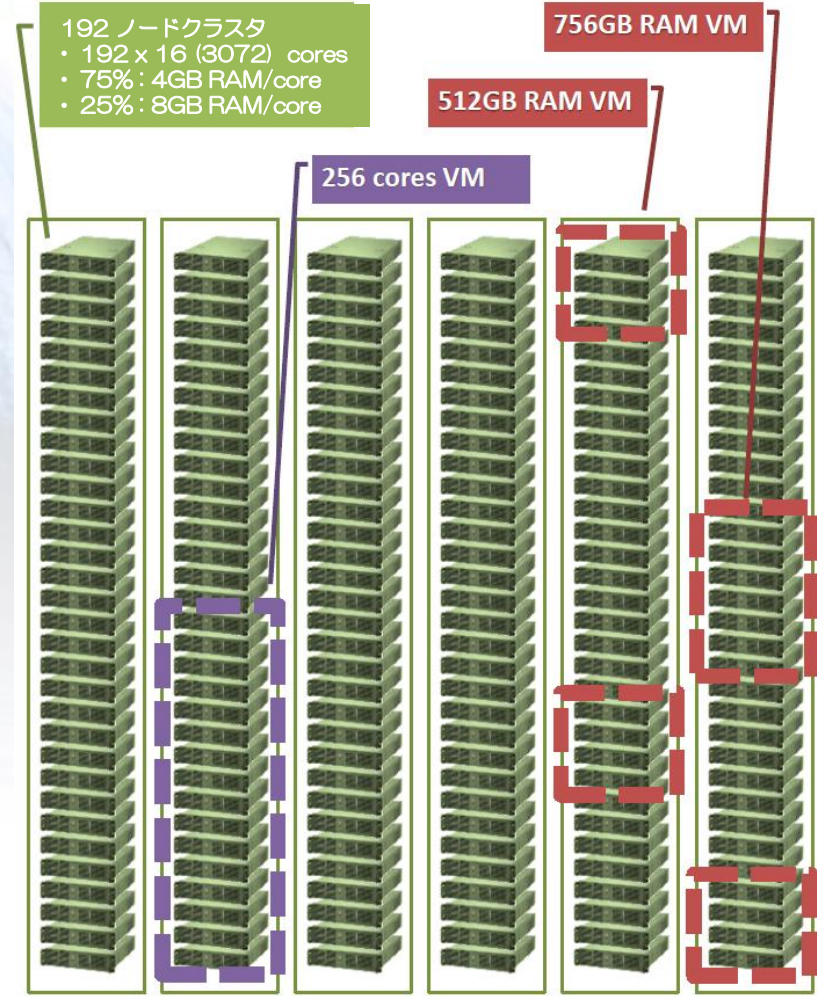
- Major deployment at STFC - Daresbury Lab, UK
 - 512 IBM iDataPlex M4 nodes:
 - 2 x Intel E5-2670, 128GB, InfiniBand
- Maximum VM capability: 384 nodes, 48TB RAM
- On-demand SMP by Platform Computing's provisioning and scheduling tools

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

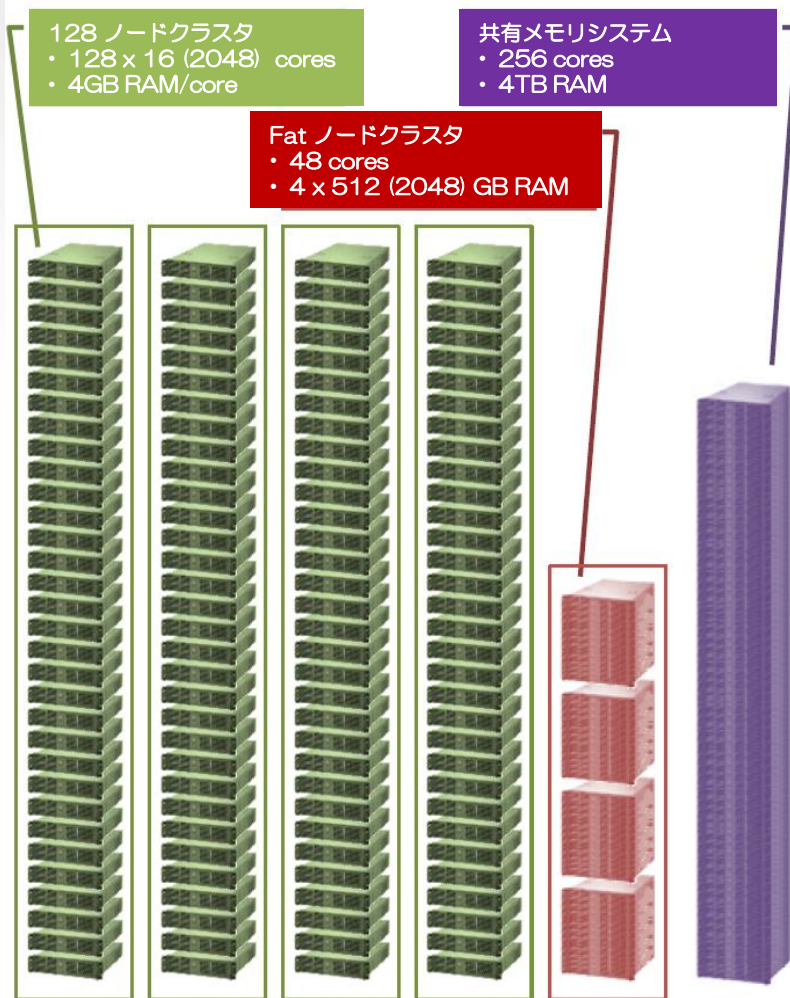


オンデマンド対応データセンター



従来型データセンター

従来型データセンター

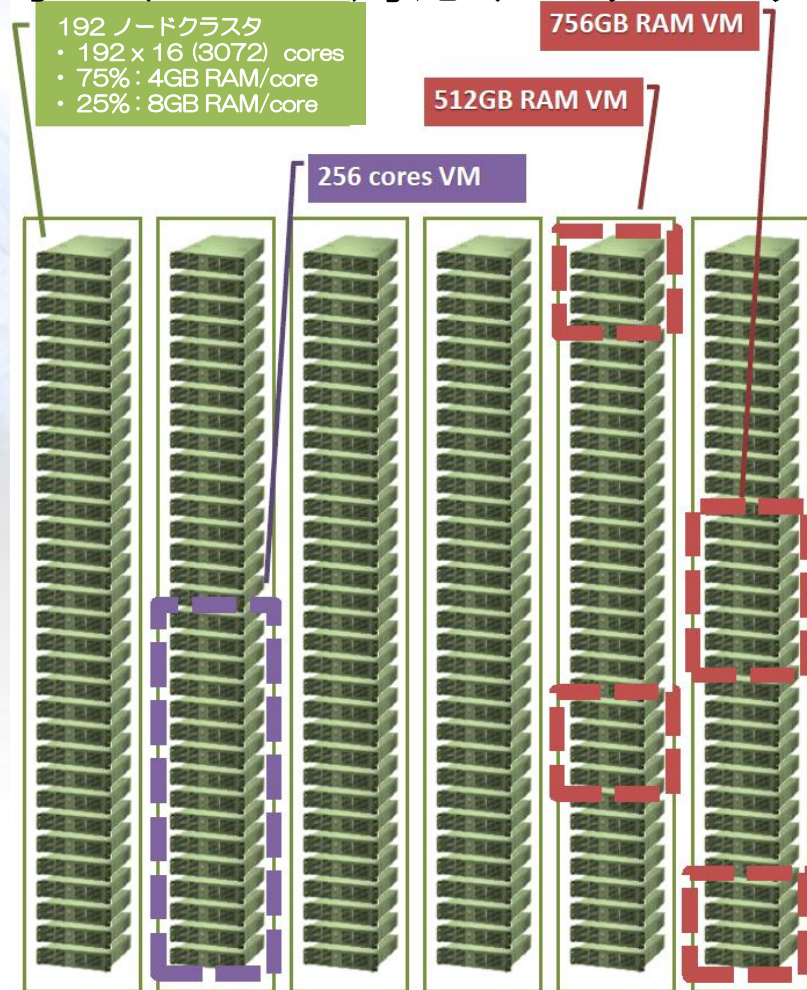


- ・ 利用環境
 - アプリケーション毎の個別システム → OSやコンパ
イラを別個に用意
- ・ 運用効率
 - 利用状況の変化に柔軟に対
応出来ない → 必要とするリ
ソースの偏在
- ・ 費用
 - クラスタシステムと比較し
て、SMPシステムの導入
コストが高い → クラスタ
システムの規模が小さくな
る

vSMP Foundationによる オンデマンド対応データセンター

- ・ 利用環境
 - 単一のハードウェアインフラ
 - 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
- ・ 運用管理
 - クラスタ環境での保守サービス
 - 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
 - 必要なハードウェアはクラスタシステム構築に必要なもので十分

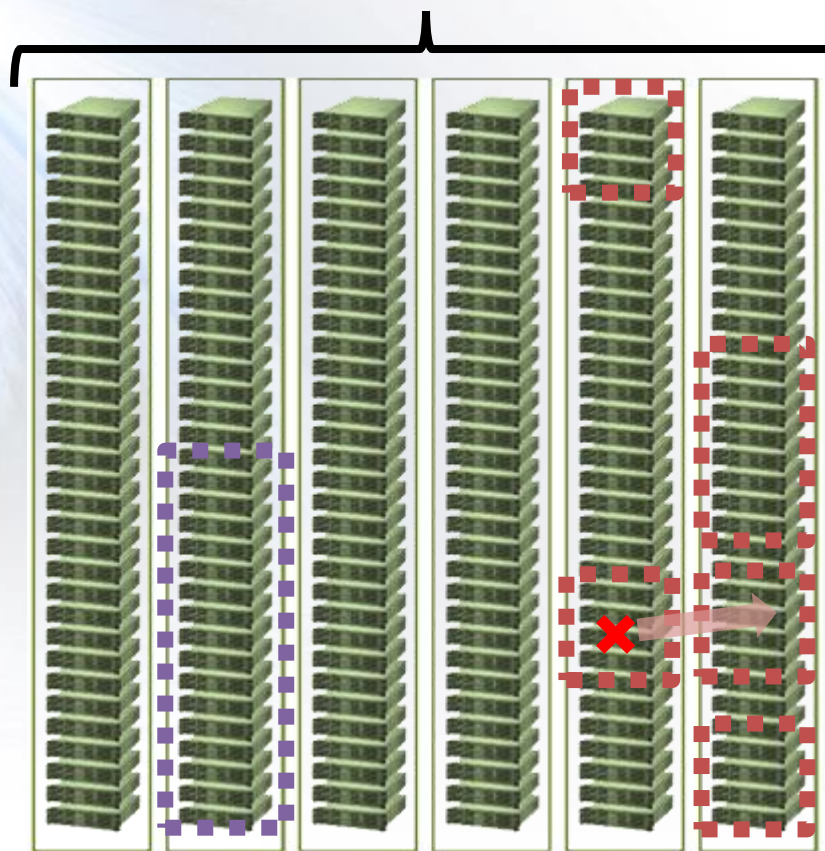
オンデマンド対応データセンター



保守サービスでの利点

- ・ クラスタシステムとの共通ハードウェアコンポーネント
 - ハードウェアについて、共通のサポート体制
 - 共通の管理ツールや診断ツールの利用が可能
 - パーツなどの共用
- ・ リソースプールからSMP構築
 - 特定ノードのハードウェア障害によるサービス停止の回避
 - 常に必要とするSMP計算リソースの確保

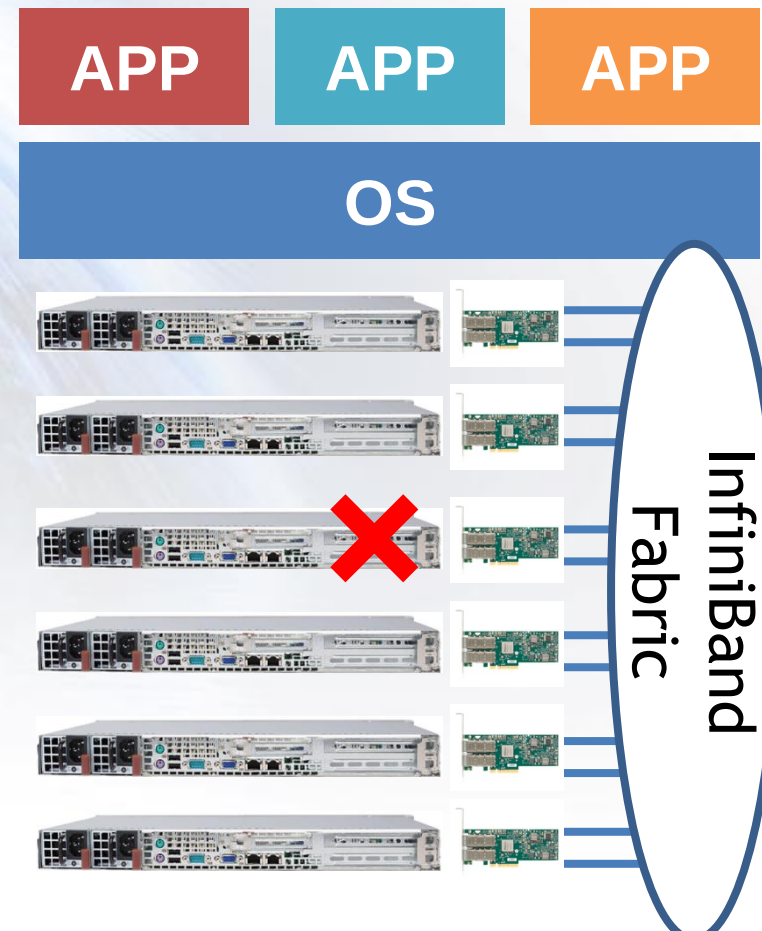
同一HW/サポート



可用性・信頼性機能

- ハードウェア（ノードに依存）
 - 冗長電源・冗長冷却ファン・メモリミラー
- vSMP Foundation アーキテクチャ
 - アクティブ-アクティブバックプレーン
 - フォールトトレランス - 故障コンポーネントを排除して即時リスタート可能
 - パーティション / コンテナサポート
- リモートマネージメント（ハードウェアベンダ+ScaleMP)
 - ハードウェア筐体マネージメント
 - vSMP Foundation SOL (Serial-over-LAN) サポート

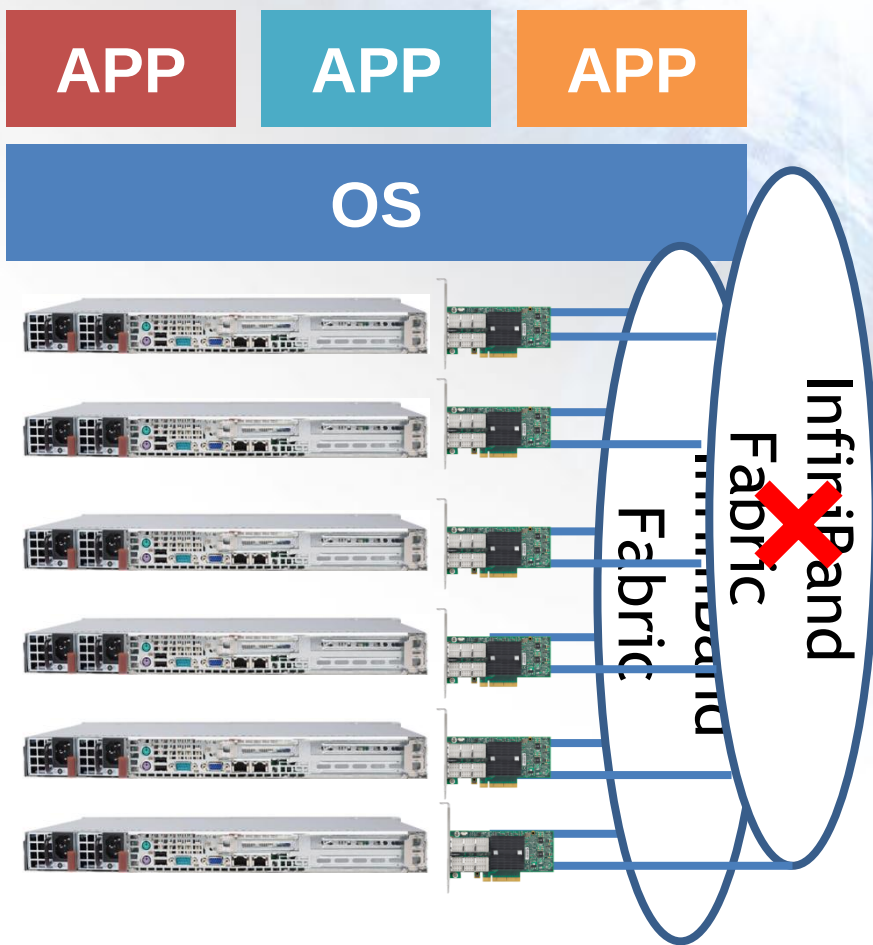
複数のコンピュータ
シングルオペレーティングシステム



InfiniBand Fabricの多重化

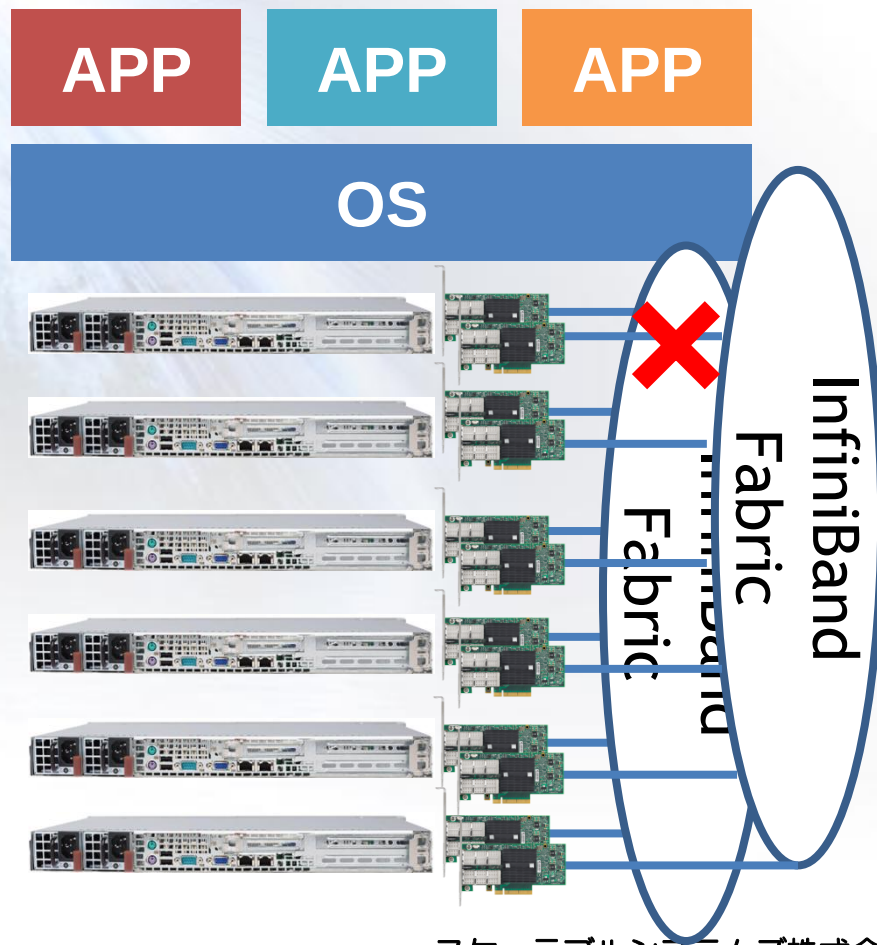
InfiniBand Fabricの2重化

障害時にも継続オペレーションが可能



InfiniBand HCAの多重化

性能向上と可用性の向上



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

- ・ 独自規格でのシステム開発による最新技術への追従の遅れ
- ・ システム開発期間とシステムコストの問題

vSMPでのソリューション

- ・ サーバプラットフォームの技術革新を直ぐに活用可能
- ・ 導入後もソフトウェ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

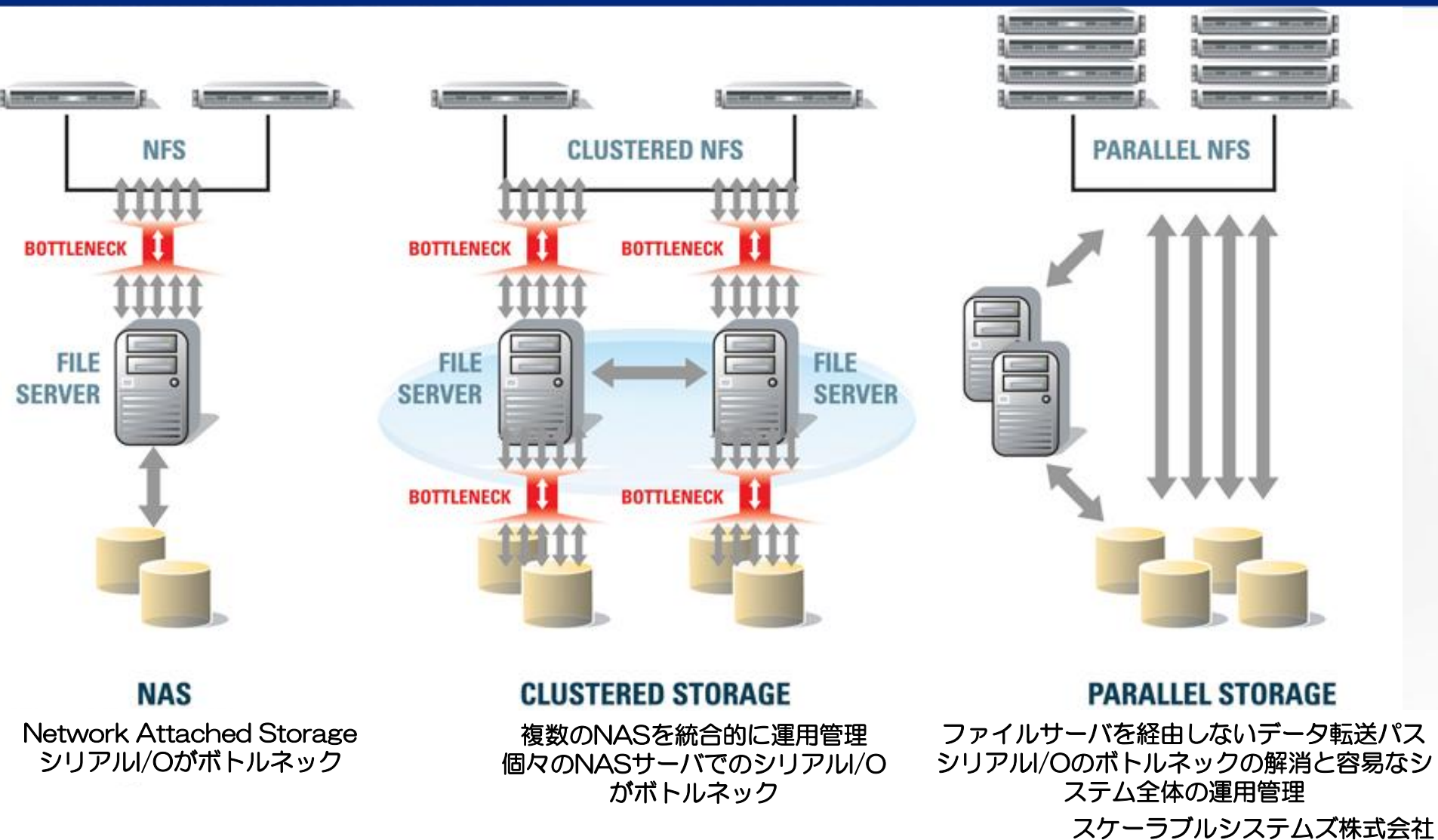


ストレージクラスタ

Panasas ActiveStore

Panasasによるソリューション

パラレルNFSの実装



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

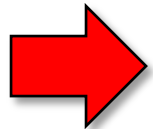


データセンターの課題

- ・ 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

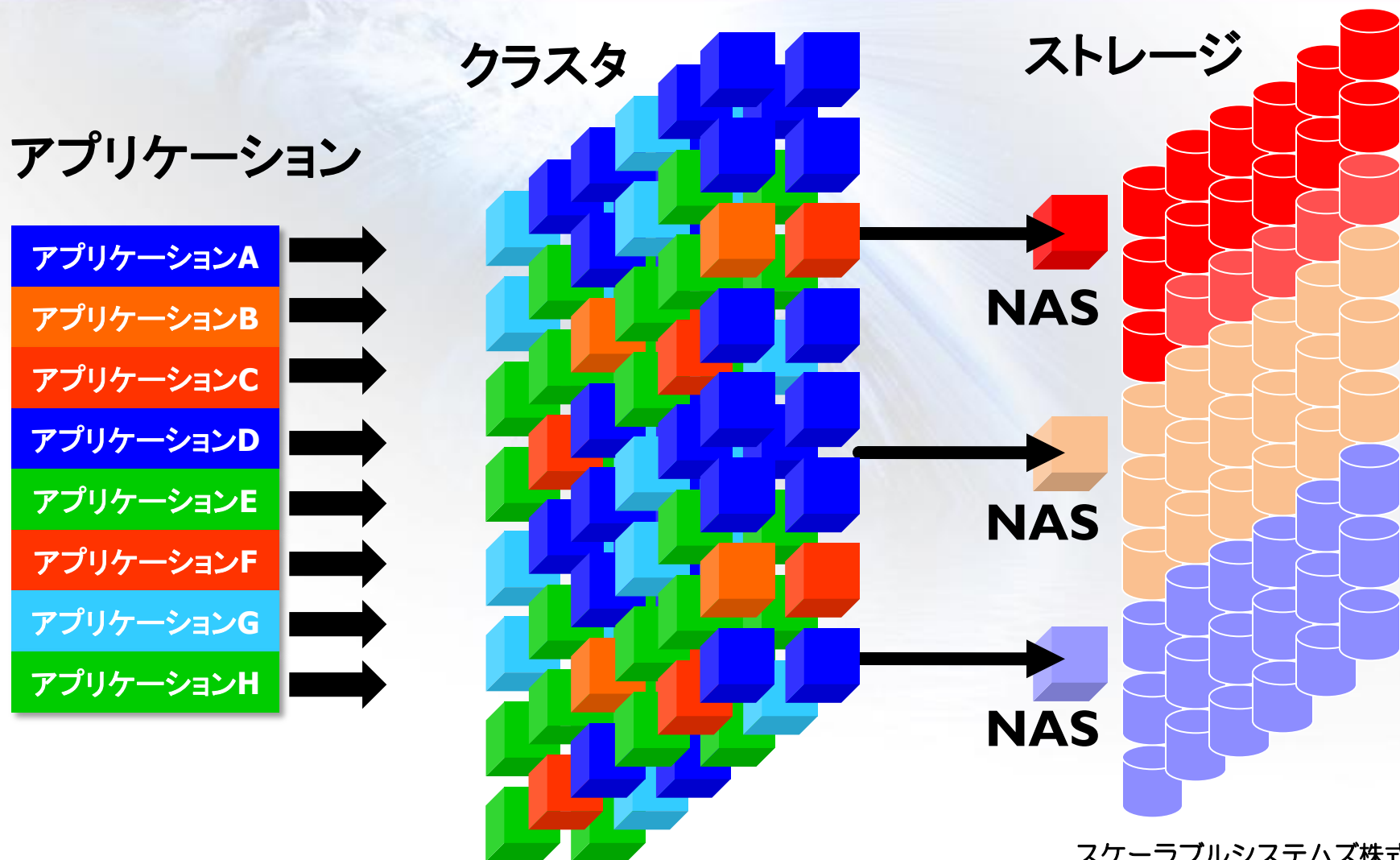
ストレージシステムの課題

- ・ データアクセスの問題
 - 複数のアーキテクチャ（ハードウェア、ソフトウェア）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

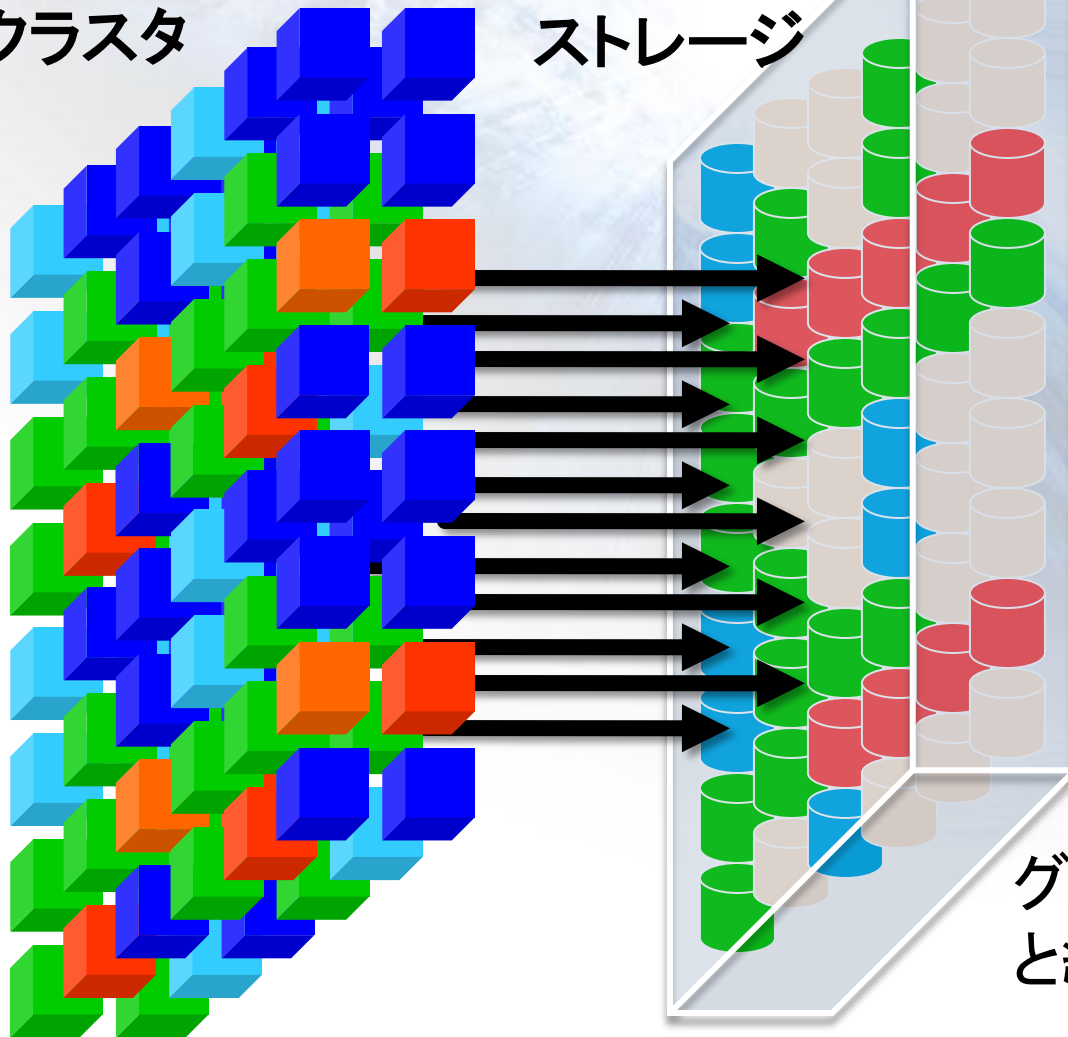
NASストレージでのシステム構成



グローバルネームスペースによる ソリューション

クラスタ

ストレージ



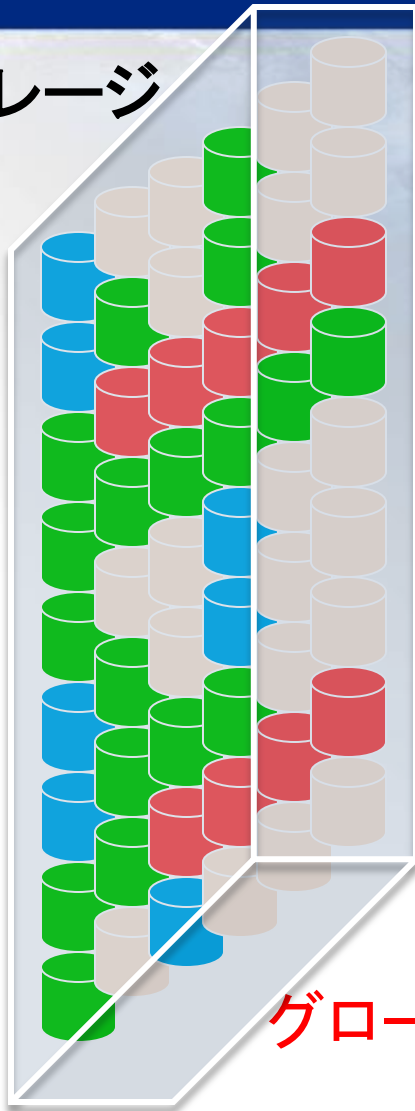
ストレージプール

- ・ 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

ストレージ

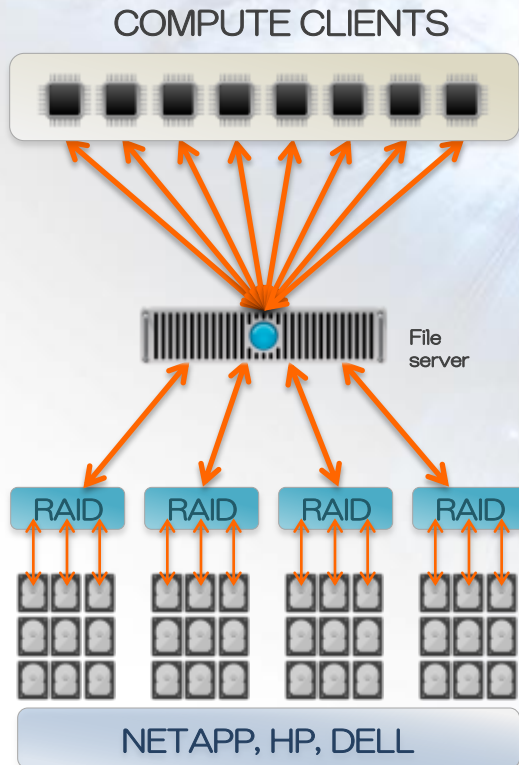


- ・ 利用の簡便さ
 - 全クライアントが全データを見ることが可能
 - マウント・ポイント管理が不要
 - クライアント側の変更が不要
- ・ 透過性
 - 容易な拡張
 - Failover
- ・ スケーラビリティ
 - ネームスペースをペタバイトにまで拡張可能
 - 大規模ボリュームの容易な管理

グローバルネームスペースと統合された運用管理環境

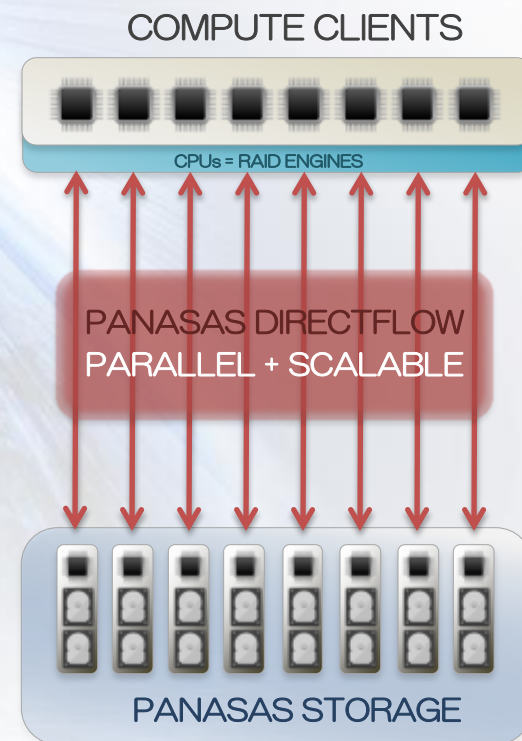
Panasas ActiveStor ストレージアーキテクチャ

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ェア

- 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

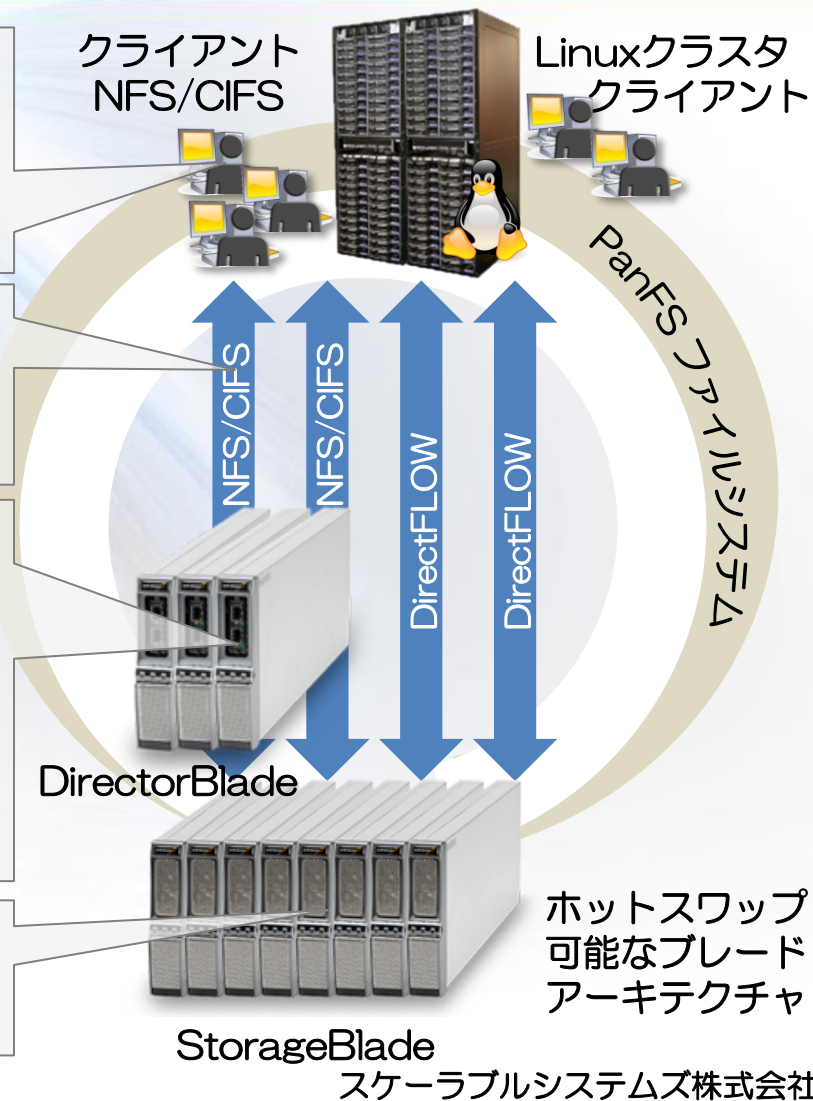
- 負荷を自動的にストレージクラスタ全体に分散
- クライアント数の増加に合わせてスケーラブルな性能拡張
- 全てのDirectorBladeが全てのファイルにアクセス可能

シングルネームスペース

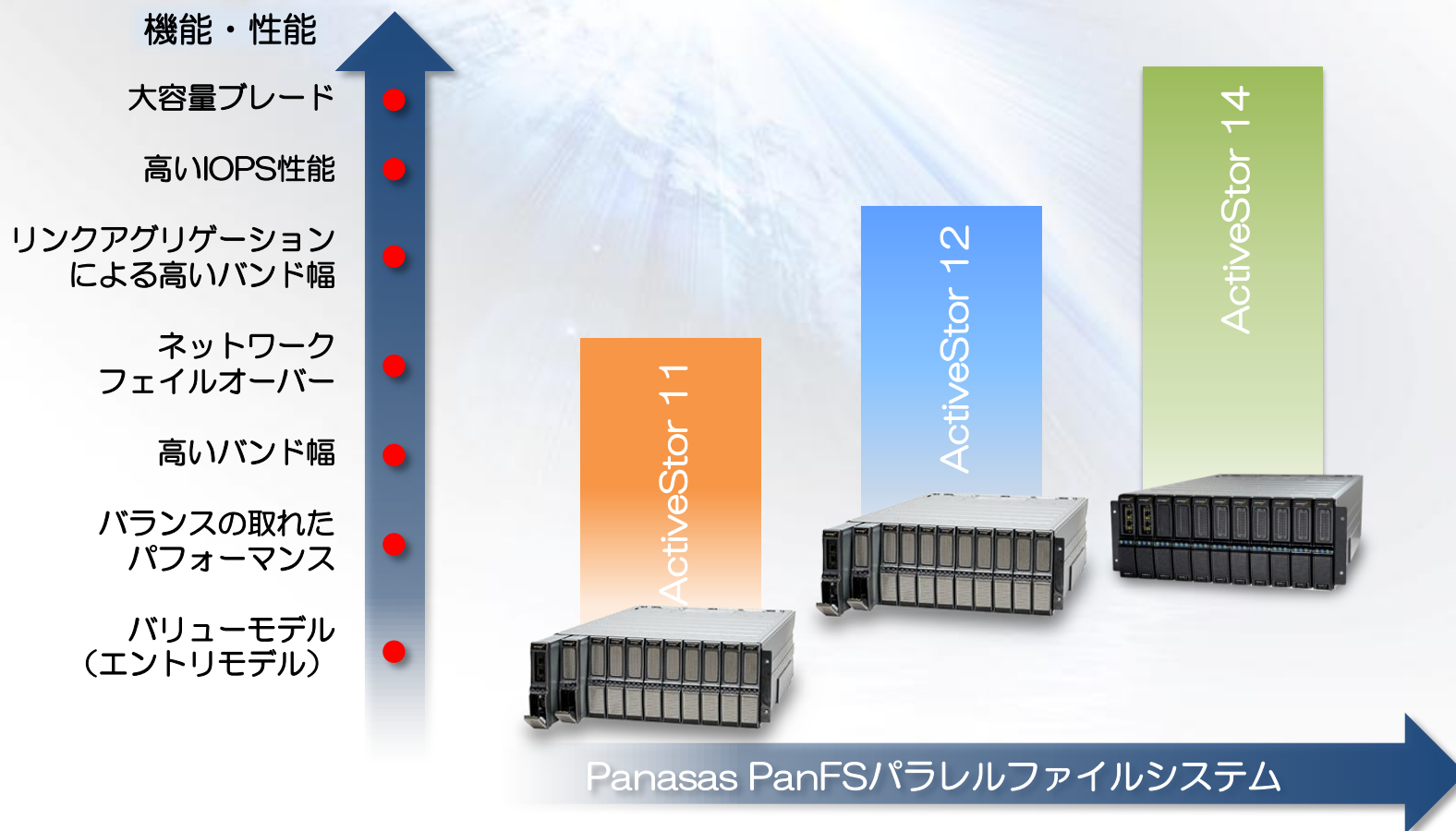
- 同一データへのいずれのプロトコルでのアクセスも可能
- シングルファイルシステム
- DirectFLOW/NFS/CIFS/NDMP間の完全なコヒレンシの実現
- 非Linuxのデバイスをシステムに統合
- グローバルネームスペースによるシステムの容易な拡張と運用の容易さ

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ー	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のインテリジェントな活用
- 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

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. , 2005-2013, Unauthorized use is strictly forbidden.