

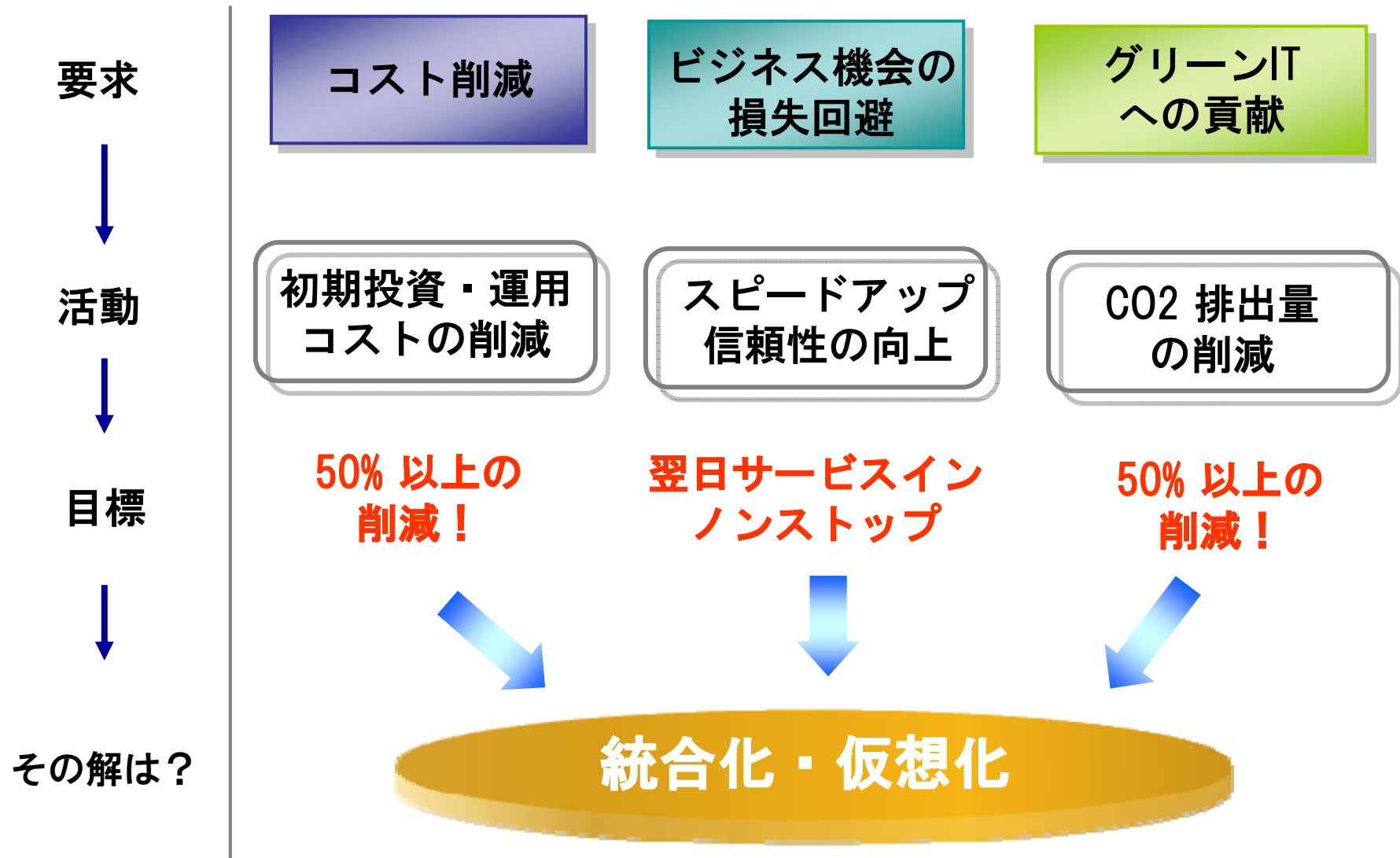
「急がれる I/O 統合、そして I/O 仮想化 そのわけとは？」

そして、絶間なく進化するインフラ技術。投資保護を実現するポイントとは？

シーゴシステムズ・ジャパン
I/O 仮想化 マーケティング
尾方 一成
(Issei Ogata)

xsigo
systems

経営層からの要求



- ・ サービスの立上げ
- ・ メンテナンス
- ・ 消費電力、熱対策、
冷却コスト
- ・ ラックスペース
- ・ データバックアップ
- ・ 災害対策
- ・ 障害対策
- ・ ドキュメント作成



そんな彼は、ある日
こんな夢をみました・・・

呪われた データセンタ . . .

xsiigo
systems



そうだ、統合化を始めよう！

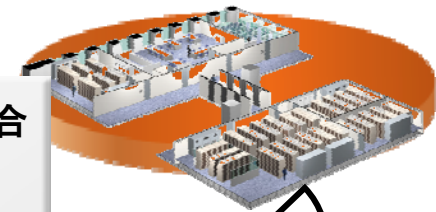
サーバ 統合

処理能力の高いサーバへの統合・移行



データセンタ 統合

設備管理費削減
運用効率の向上



運用・管理システム 統合

運用の標準化
効率の向上

I/O

ネットワーク 統合

アップリンク広帯域化
セキュリティ強化



ストレージ 統合

ディスク利用率向上
バックアップシステム
の整備



サーバのパフォーマンスは劇的に向上

CPU 数とメモリ容量の増加 → パフォーマンスの向上



サーバ仮想化の促進

複数のアプリケーションを同一サーバ上で稼働させることにより、サーバ性能を最大限有効活用。



広帯域 サーバ I/O のニーズが増加

サーバ性能の向上、サーバ仮想化の促進に加え、ディスクレスサーバの導入が始まり、広帯域サーバ I/O のニーズが向上。



ネットワークのアーキテクチャは絶間なく変化

I/Oに関連する新しい技術 (DCE, FCoE, 8Gbps FC 等) が登場。
データセンタ構築・運用の短期と長期の計画がますます複雑に。

今日のデータセンタニーズに応え、かつ将来においても既設のシステムと新しい技術を融合（統合）する形にてシステムの増設を可能にする『I/O 仮想化トータルソリューション』プロバイダ

- ・ US本社創設: 2004年8月
- ・ 日本法人設立: 2006年11月
- ・ 製品リリース: 2007年9月
- ・ 従業員: 110名
- ・ 本社:
 - サンノゼ、CA
- ・ 出資:
 - Kleiner Perkins,
 - Greylock Partners,
 - Khosla Ventures
 - Juniper Networks
- ・ ボードメンバー:
 - レイ・レーン (オラクル元 社長)
 - マーク・レスリー
(ベリタスソフトウェア創業メンバ)
 - ビノッド・コースラ
(サンマイクロシステムズ創業メンバ)
 - アショク・クリシュナマティ
(ジュニパーネットワークス創業メンバ)



First shipment : January 2007

Award と各種メディアからの注目度

xsigo
systems



ByteandSwitch
The Storage Networking Site

eWEEK
CHANNEL INSIDER

COMPUTERWORLD

InfoWorld

eWEEK.COM

INFOSTOR

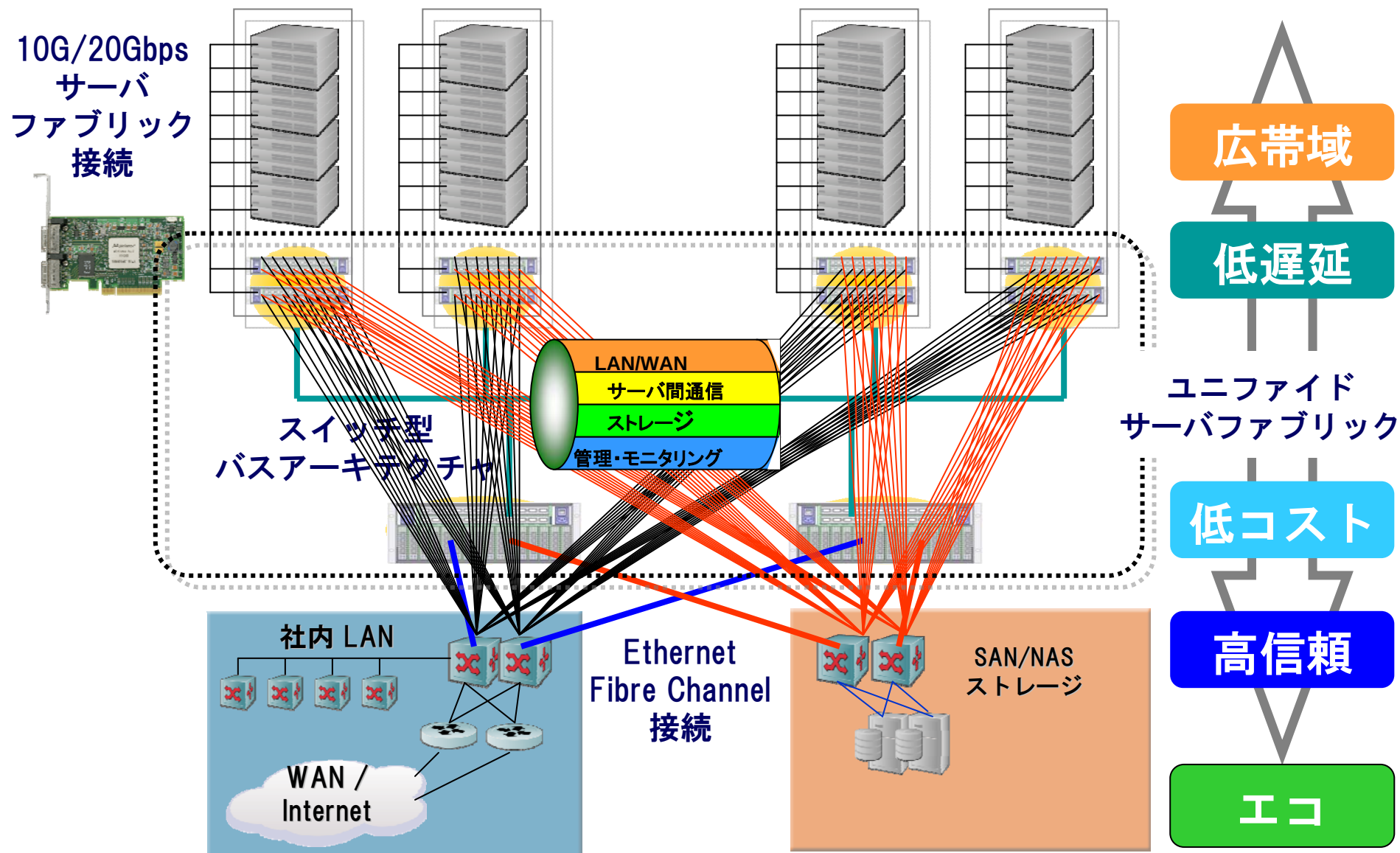
Network
Computing

TechTarget

BusinessWeek

NETWORKWORLD

サーバ I/O 統合に必要なもの？



高速サーバ I/O = InfiniBand



InfiniBandは、広い帯域幅と非常に高いRAS（信頼性・可用性・保守性）、そして低遅延の特徴を持つスイッチ型のサーバ用高速 I/O バスアーキテクチャであり、コンピュータ同士、コンピュータとストレージ、もしくはネットワーク機器を接続。

HCAは、Host Channel Adapter の略で、InfiniBand機器がホスト側のCPUやメモリにアクセスできるようにするもので、ホストにとってInfiniBandの出入口となる。

圧倒的なコストパフォーマンス！

1ポート 10Gbps	¥86,250	1ポート 20Gbps	¥112,500
2ポート 10Gbps	¥122,500	2ポート 20Gbps	¥132,750

【特徴】

- ・フロー制御を備えた信頼性の高い伝送
- ・ハイパフォーマンス（10,20,40Gbps）
- ・効率的なデータ転送（RDMAサポート）
- ・各トラフィック毎の専用リソース
- ・拡張性（数千台のノード管理）

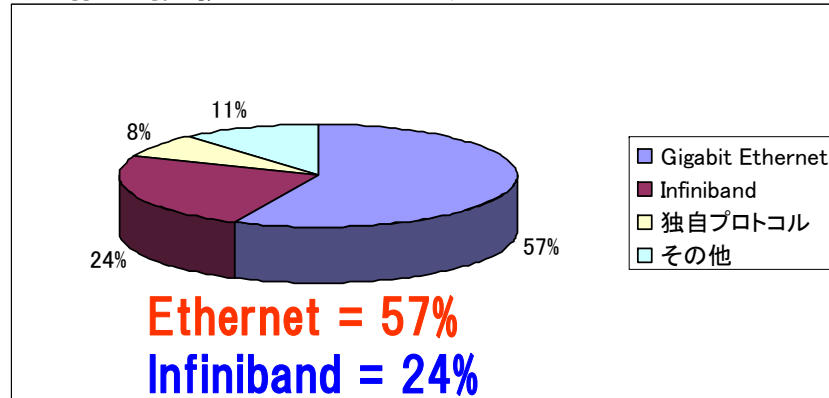
標準 4Gb FC HBA
¥128,000
標準 10GE NIC
¥240,000

40Gbps QDR（Quad Data Rate）
Mellanox社 4月リリース

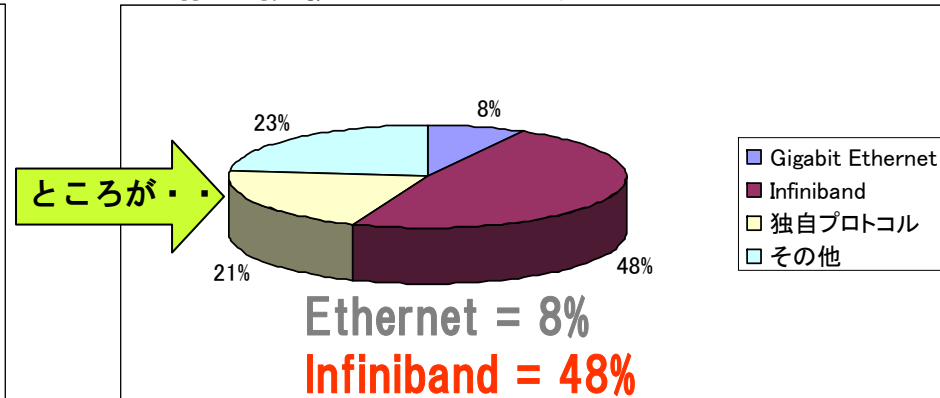
Infiniband の真の実力とは？(Top500)

Top500：スーパーコンピュータの性能を計測し、ランキングを公表するプロジェクト

◇ 相互接続プロトコル別（1位～500位）



◇ 相互接続プロトコル別（1位～100位）



◇ 相互接続プロトコル別（1位～10位）

順位	システム	導入先	ベンダー	スコア (TFLOP/s)	相互接続
1	Roadrunner	DOE/NNSA/LANL(米)	IBM	1,026.0	Infiniband
2	BlueGene/L	DOE/NNSA/LANL(米)	IBM	478.2	BuleGene/L独自
3	Blue Gene/P	Argonne National Laboratory(米)	IBM	450.3	BuleGene/L独自
4	Ranger	Texas Advanced Computing Center(米)	Sun	326.0	Infiniband
5	Jaguar	Oak Ridge National Laboratory(米)	Cray	205.0	Cray XT4 独自
6	JUGENE	Forschungszentrum Juelich(独)	IBM	180.0	BuleGene/P独自
7	Encanto	New Mexico Computing Applications Center	SGI	133.2	Infiniband
8	EKA	Computational Research Laboratories(印)	HP	132.8	Infiniband
9	Blue Gene/P	IDRIS(仏)	IBM	112.5	BuleGene/P独自
10	SGI Altix ICE 8200EX	Total Exploration Production(仏)	SGI	106.1	Infiniband

*** 10サイト中、5サイトが Infiniband を採用。Ethernet の採用は0。**

ネットワーク運用の統合化 - FCoE

Fibre Channel over Ethernet (FCoE) は、ANSI INCITS T11で標準化がすすめられている新しいプロトコル。ファイバチャネル (FC) のプロトコルをイーサネット フレームとして運ぶことにより、LAN (イーサネット) とSAN (ファイバチャネル) のI/O 統合が実現。

(課題) FCoE をサポートするイーサネット ネットワークは、ネットワークを通過するパケットがドロップすることなく転送されることが必須。そのためには、FCoE スイッチ設計に、ドロップの発生しないパケット転送とネットワーク フロー制御のメカニズムが組み込まれている必要がある。

2008年：標準化作業 → 2009年：テスト導入期間 → 2010年：実運用開始

FCoE 対応 10GE CNA (Converged Network Adapter Card)



Emulex LP21000



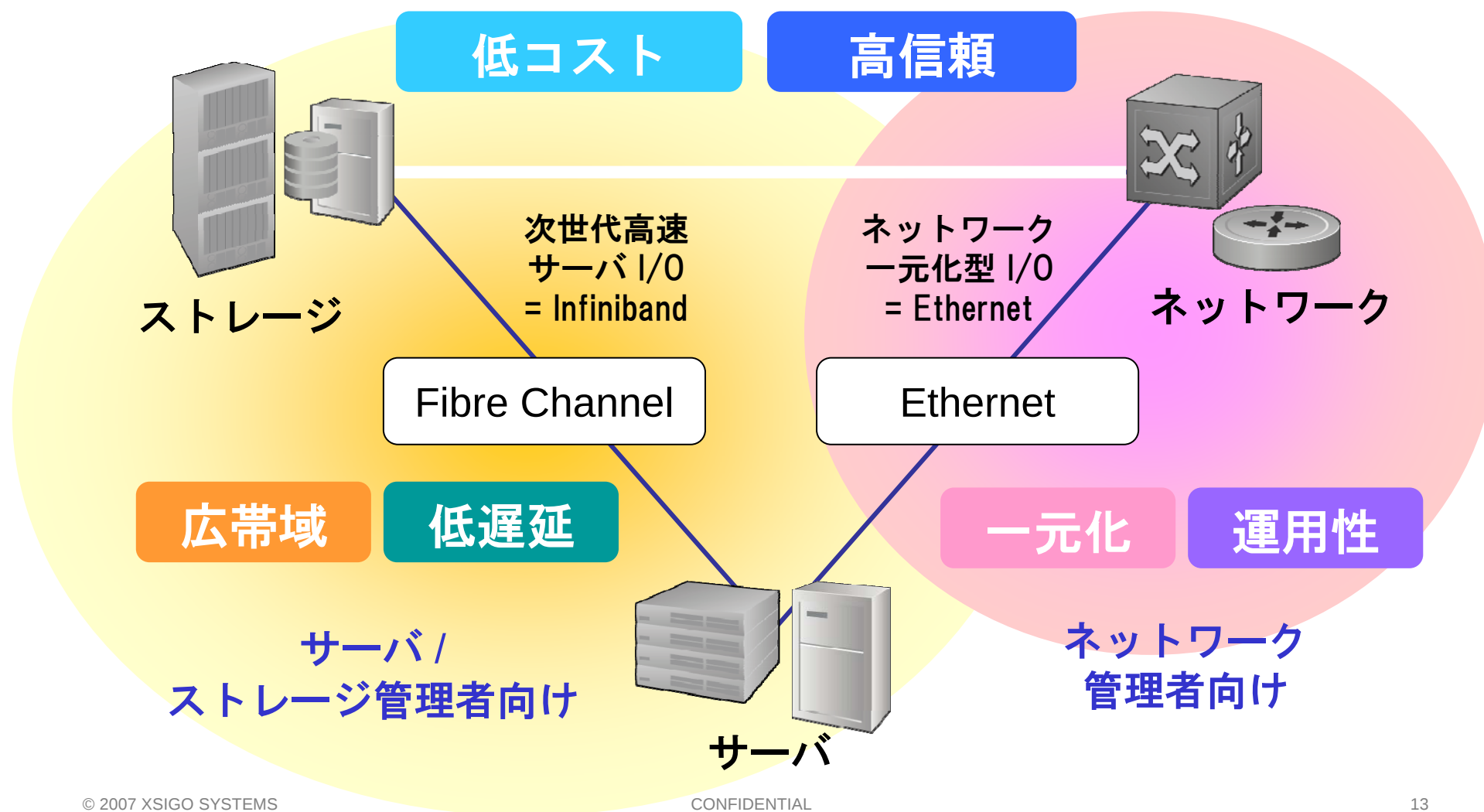
Qlogic QLE8042

＜フロー制御とロスレスイーサネット＞

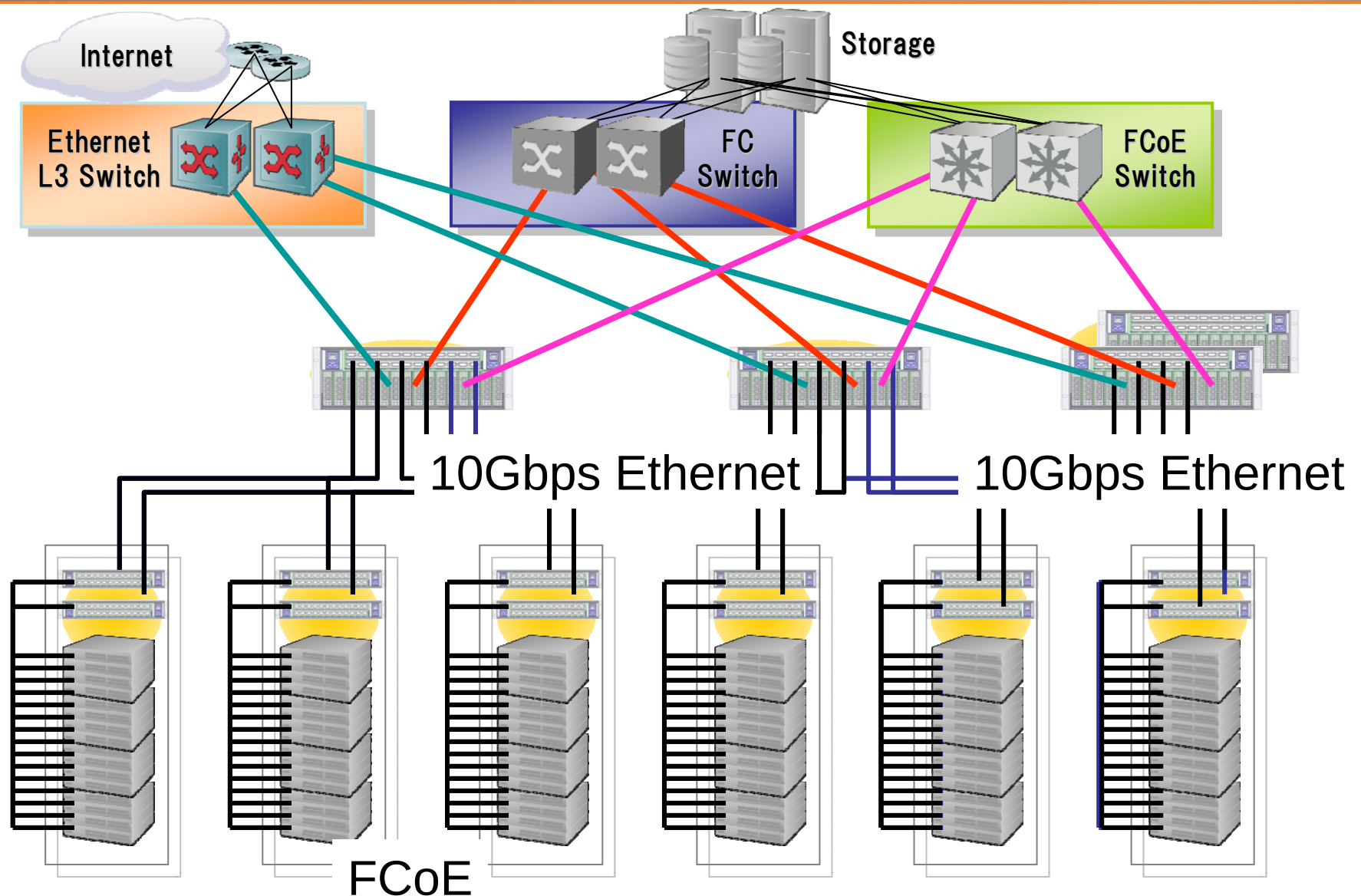
ファイバチャネルでは、データが途中でドロップされずに送信先に届くことを保障しているが、イーサネットでは送信先に確実に届くことは保障しておらず、パケットロスは上位階層の TCP/IP の再送機能で対応する取り決めとなっている。イーサネットのようなパケットドロップ方式の輻輳管理は、ストレージネットワーク環境には使用できないため、FCoE には新たな機能拡張が必要となる。

そこで FCoE では、輻輳管理にイーサネットの PAUSE 機能を転用。これは、ある受信ポートのトラフィックが多くなると、その受信ポートから送信元ポートに対して PAUSE 信号を送り、送信を一時的にストップさせる機能。

Xsigo I/O 統合 ソリューション



あらゆる I/O ニーズへ対応



VP780 : I/O 仮想化コントローラ

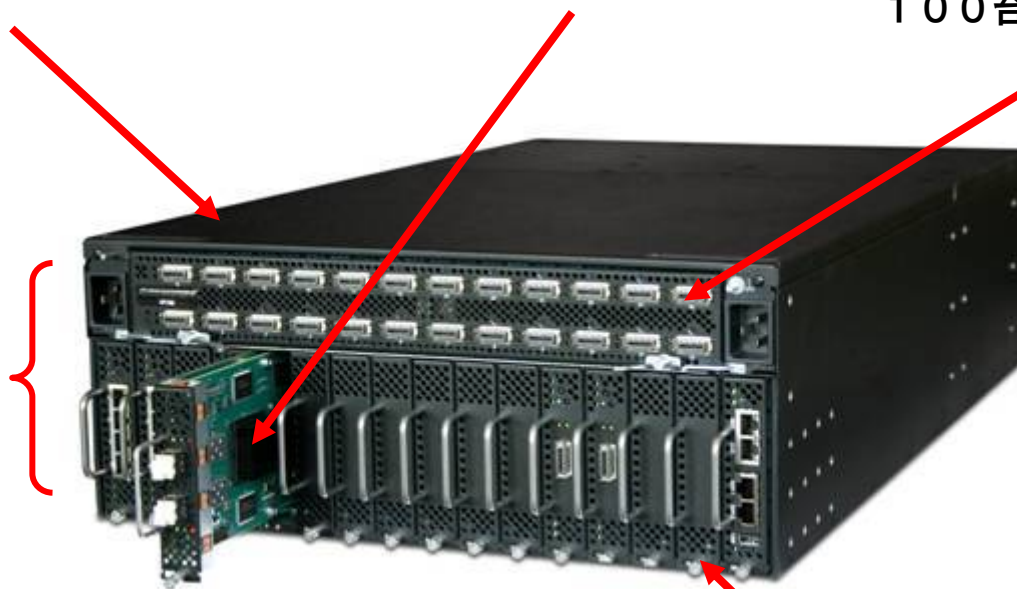


ハードウェアベースのアーキテクチャ
・ノンブロッキング ファブリック

Xsigo 独自の ASIC
ラインレートのスループット

24 サーバポート
サーバエッジスイッチの利用により、
100台を超えるサーバも接続可能

4U サイズ



SCP (システムコントロール
プロセッサ)

冗長化ファン

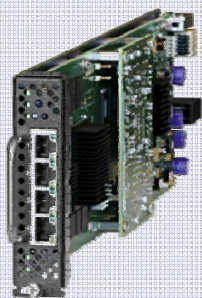


冗長化ホットスワップ対応電源

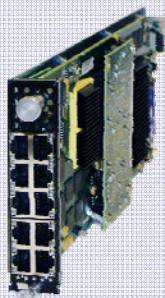
15 x I/O モジュールスロット

- ・ 4ポート x 1GbE
- ・ 10ポート x 1GbE
- ・ 1ポート x 10GbE
- ・ 2ポート x 4Gb Fibre Channel
- ・ 1ポート x 10Gb InfiniBand

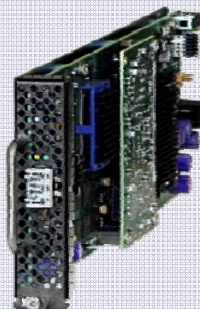
I/O モジュール



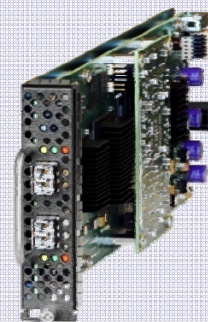
4ポート
1Gigabit
Ethernet



10ポート
1Gigabit
Ethernet



1 ポート
10 Gigabit
Ethernet



2ポート
4 Gigabit
Fibre Channel

ラックスイッチ : IS24



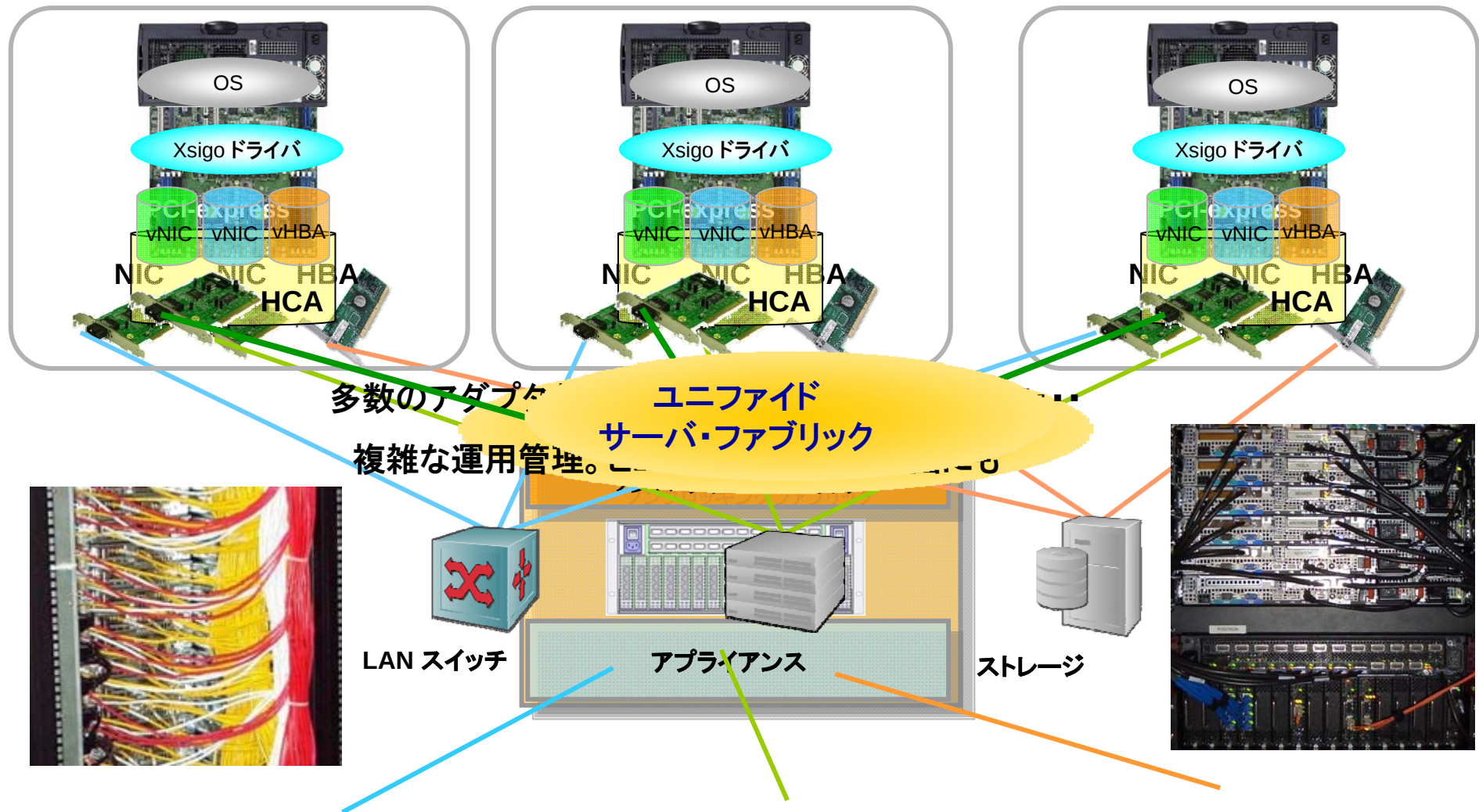
24 ポート Infiniband スイッチ

- 20Gbps / Double Data Rate (DDR)
- 10Gbps / Single Data Rate (SDR)

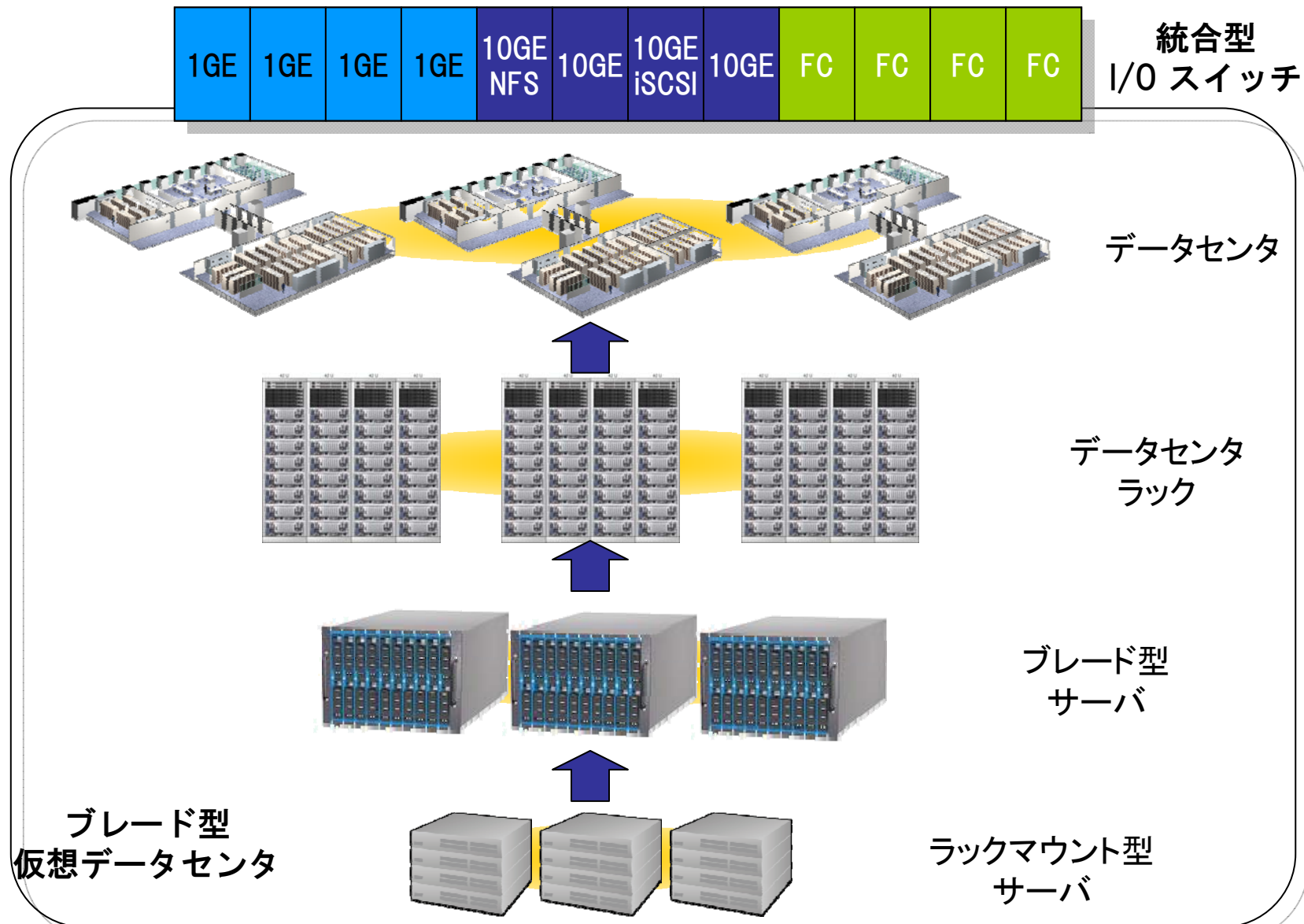
ノンブロッキングファブリック

- 480 ギガビットの帯域

シーゴによる I/O 仮想化のイメージ図



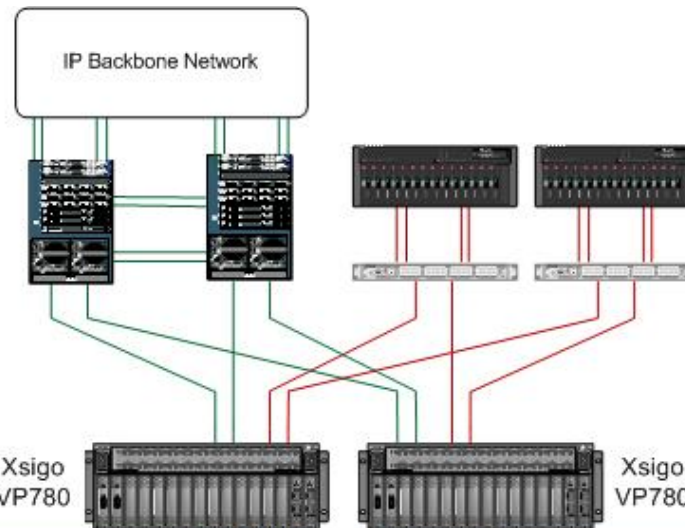
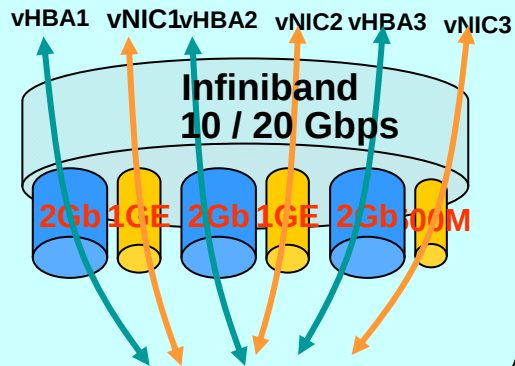
Xsigo の I/O 統合コンセプト



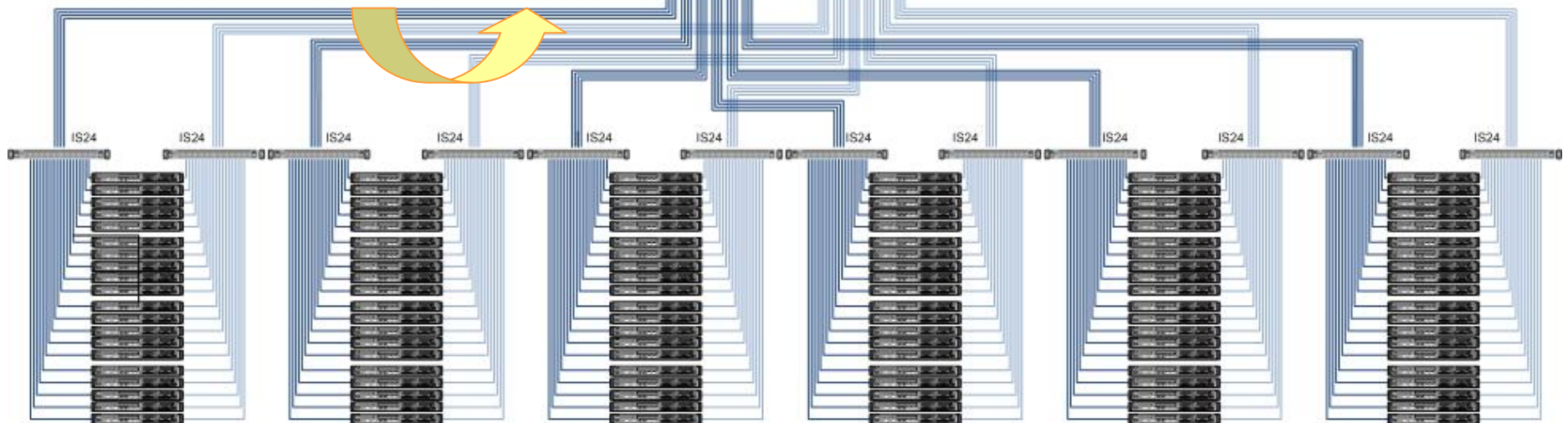
例) 120 ラックマウント型サーバ構成

xSigo
systems

10/20 Gbps の物理 IB リンクを
Xsigo の帯域保証機能により、複
数の仮想リンクに分けて使用。



経営層からの要求
CO2 排出量削減
クリア!



初期投資コストの削減

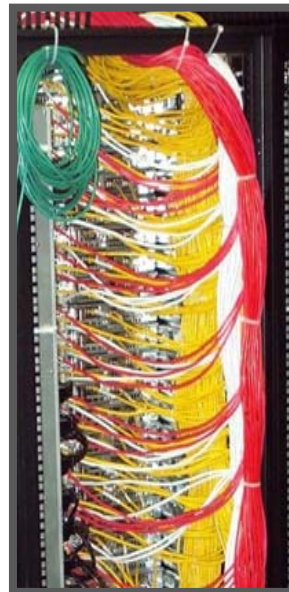
例) サーバ120台の I/O 冗長化構成 (定価)

経営層からの要求
初期コスト削減
クリア!

今までの I/O

480 NIC ポート
240 HBA ポート
480 Ethernet ケーブル
240 Fibre ケーブル
480 Ethernet スイッチポート
240 FC スイッチポート

総合計: 約 \$700,000



Xsgo の I/O

240 HCA ポート
240 + 48ケーブル
2 Xsgo 筐体 + OS
1GE + 4Gb FC モジュール
12 Xsgo サーバスイッチ

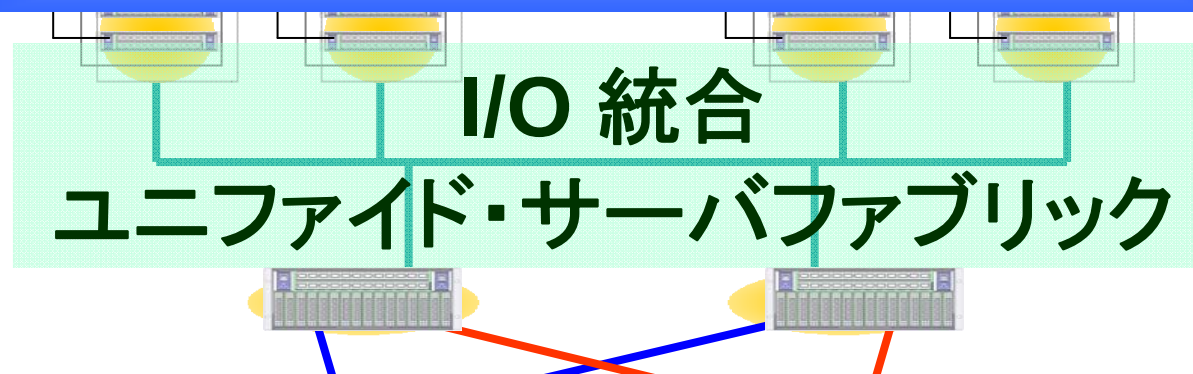
総合計: 約 \$450,000



I/O だけで 35% の初期投資コスト削減が可能に!

I/O 統合から I/O 仮想化へ

いかに**柔軟性・運用性・冗長性**が高く、**サービス品質の保証**を実現できるサーバI/Oを提供できるか？



I/O 仮想化

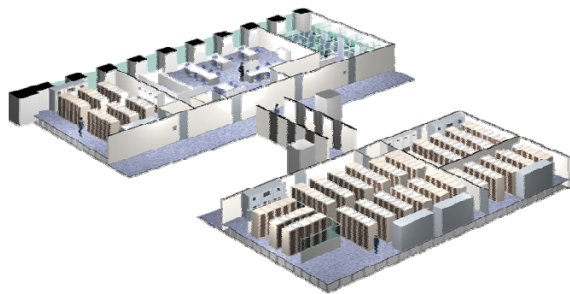
柔軟性
自由自在に仮想I/Oを制御

運用性
サーバI/Oを一元的に集中管理

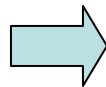
冗長性
I/Oの高速切替を完全サポート

帯域保証(QoS)
仮想サーバ環境を最適化

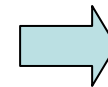
- 新製品/サービスの発表により、Webサーバとアプリケーションサーバのネットワーク増強が必要に



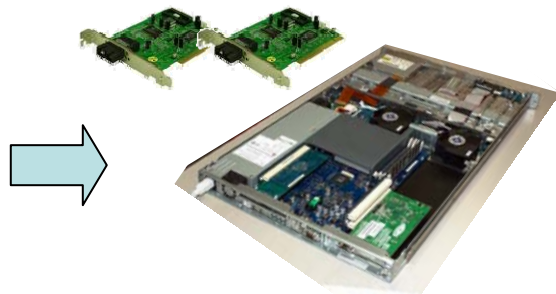
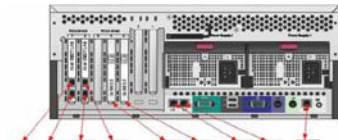
データセンター内へ
エンジニアを派遣



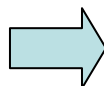
物理ネットワーク接続を確認



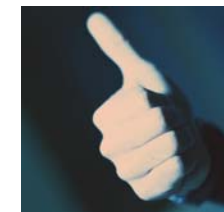
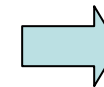
サーバのシャットダウン
ラックからの取り外し作業



サーバのふたを取り外し、
アダプタカードのインストール



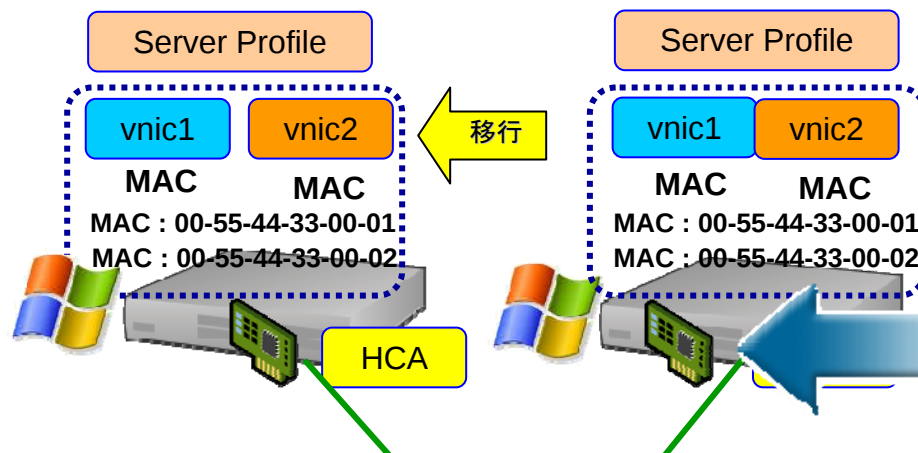
再度、サーバをラックマウント後、
サーバ再起動。最終動作確認



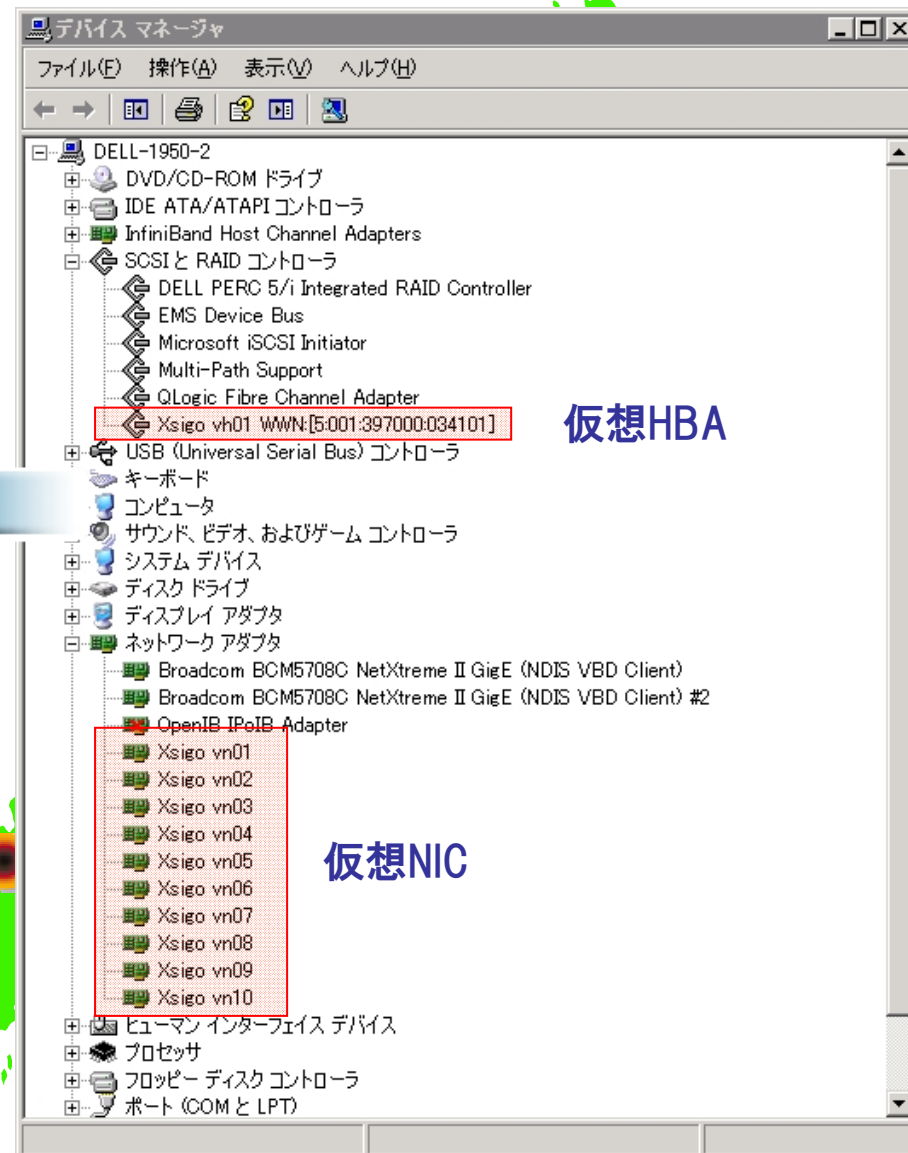
お疲れ様でした

Xsigo リモート I/O 管理イメージ

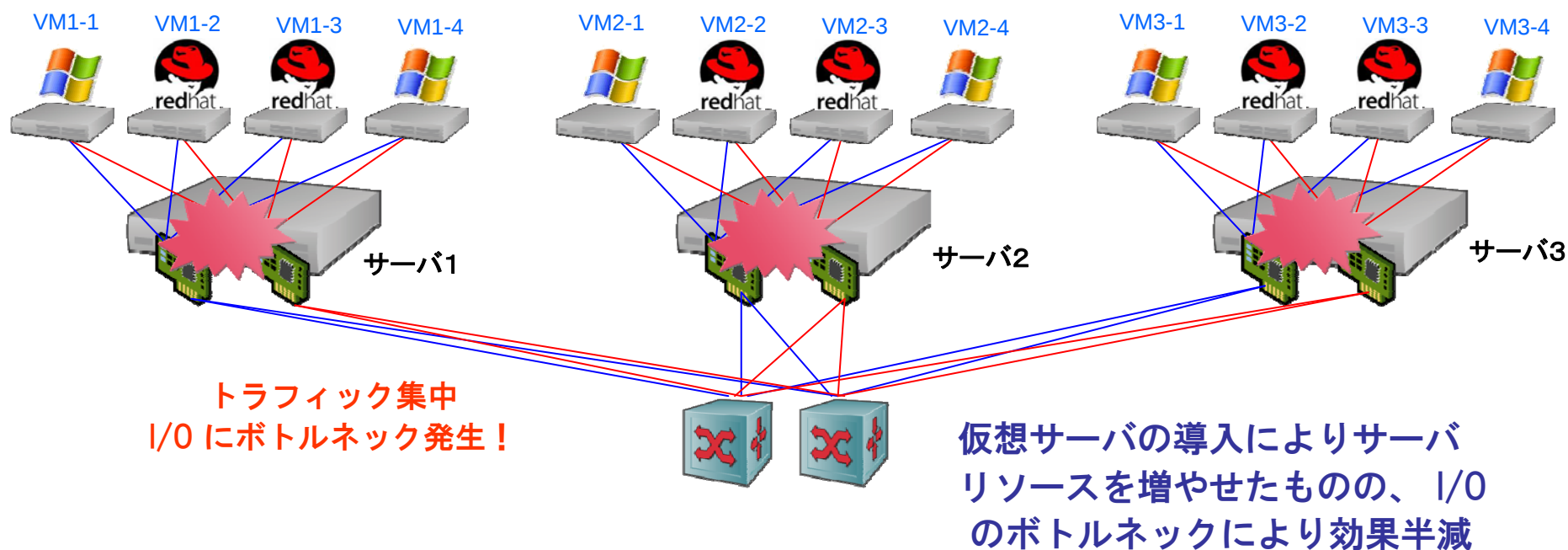
サーバの再起動をすること
なく作業完了！



```
add server-profile server1 HP_DL360@ExtSw-b8cfff004401-Port16")
add vnic vn01.server1 10/1")
set vnic vn01.server1 -addr-type=static -ip-addr=192.168.10.201 -netmask=255.255.255.0
add vnic vn02.server1 10/1")
set vnic vn02.server1 -addr-type=static -ip-addr=192.168.10.202 -netmask=255.255.255.0
add vnic vn03.server1 10/1")
set vnic vn03.server1 -addr-type=static -ip-addr=192.168.10.203 -netmask=255.255.255.0
add vnic vn04.server1 10/1")
set vnic vn04.server1 -addr-type=static -ip-addr=192.168.10.204 -netmask=255.255.255.0
add vnic vn05.server1 10/1")
set vnic vn05.server1 -addr-type=static -ip-addr=192.168.10.205 -netmask=255.255.255.0
add vnic vn06.server1 10/1")
set vnic vn06.server1 -addr-type=static -ip-addr=192.168.10.206 -netmask=255.255.255.0
add vnic vn07.server1 10/1")
set vnic vn07.server1 -addr-type=static -ip-addr=192.168.10.207 -netmask=255.255.255.0
add vnic vn08.server1 10/1")
set vnic vn08.server1 -addr-type=static -ip-addr=192.168.10.208 -netmask=255.255.255.0
add vnic vn09.server1 10/1")
set vnic vn09.server1 -addr-type=static -ip-addr=192.168.10.209 -netmask=255.255.255.0
add vnic vn10.server1 10/1")
set vnic vn10.server1 -addr-type=static -ip-addr=192.168.10.210 -netmask=255.255.255.0
```



仮想サーバ環境のサーバ I/O ボトルネック = 複数の仮想サーバが 物理 NIC/HBA を共有

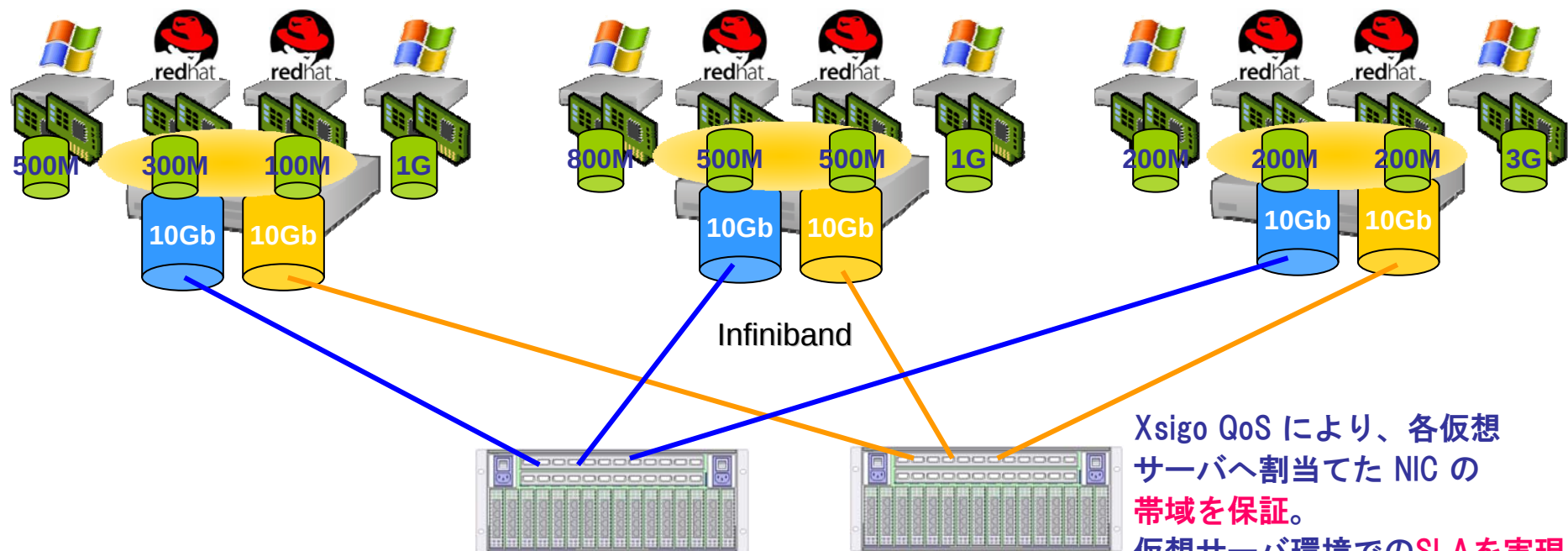


仮想サーバ環境での帯域保証

各仮想サーバへの個別 I/O の割当て =

NIC共有を回避することによる負荷分散

仮想サーバ（アプリケーション）毎の帯域保証



Xsigo 仮想 I/O 運用・管理ツール

localhost - VMware Infrastructure Client

File Edit View Inventory Administration Plugins Help

Inventory Scheduled Tasks Events Administration Maps Consolidation

xsgo Virtual I/O

Xsigo アイコン

phantom.poc.xsigo.com VMware ESX Server, 3.5.0, 64283 | Evaluation (11 day(s) remaining)

Getting Started Summary Virtual Machines Performance Configuration Tasks & Events

xsgo Version

Current Alarms: 10 3 0 0 0 0 0 0

Xsigo Management System > vNIC Summary

vNIC Summary (6)

Name	Termination	State (Admin/Oper)
☐ vnic2.wantest_B	enzo/2/2	up/up
☐ vnic3.wantest_A	enzo/2/3	up/up
☐ vnic1.wantest_A	enzo/5/1	up/up
☐ vnic3.wantest_B	enzo/2/3	up/up
☐ vnic2.wantest_A	enzo/2/2	up/up
☐ vnic1.wantest_B	enzo/5/1	up/up

Virtual Resources

- vNICs
- vHBAs
- Server Profiles
- Virtual Machines
- Templates
- Network QoS
- SAN QoS
- Termination Groups
- LUN Mask Profiles
- Persistent Mapping Profiles
- SAN Boot Configurations

Recent Tasks

Name	Target	Status
Migrate Virtual Machine	W2K3_VM1	Completed

Tasks Alarms

administrator

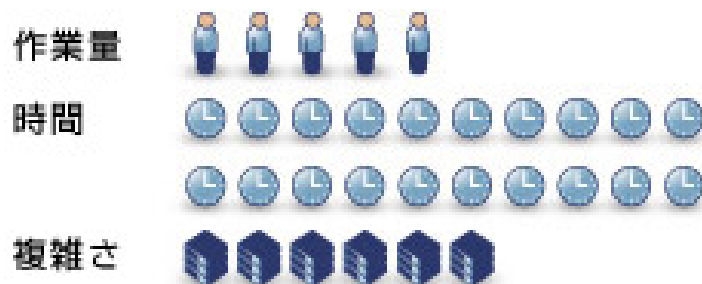
localhost - VMware In...

3:15 PM

運用負担比較 (例: 新規サーバの導入) xsigno systems

既存 固定I/O環境

- 先を見据えたI/O構成の計画が必要となり、サーバへは物理的にアダプタカードを追加。
- 関連するケーブルの配置計画も必要となり、**全て物理的な取り付け作業。**
- 各サーバには**アプリケーション種別により、個別のI/O構成**があり、それぞれ**ストレージおよびネットワーク機器側の設定に反映**させる必要がある。
- 通常の導入時間：**数週間**



次世代 仮想I/O環境

- サーバとの**接続は仮想**であり、**必要に応じてサーバに実装**できる。
- シンプルな配線計画。**物理ケーブル配線は、アプリケーションに関わらず全サーバで統一可能。**
- ストレージの事前割り当てに加え、ネットワークリソースを一度に複数のサーバに割り当て可能。サーバマネージャは、**チーム全体を巻き込むことなく必要に応じてサーバを導入**できる。
- 通常の導入時間：**数時間**

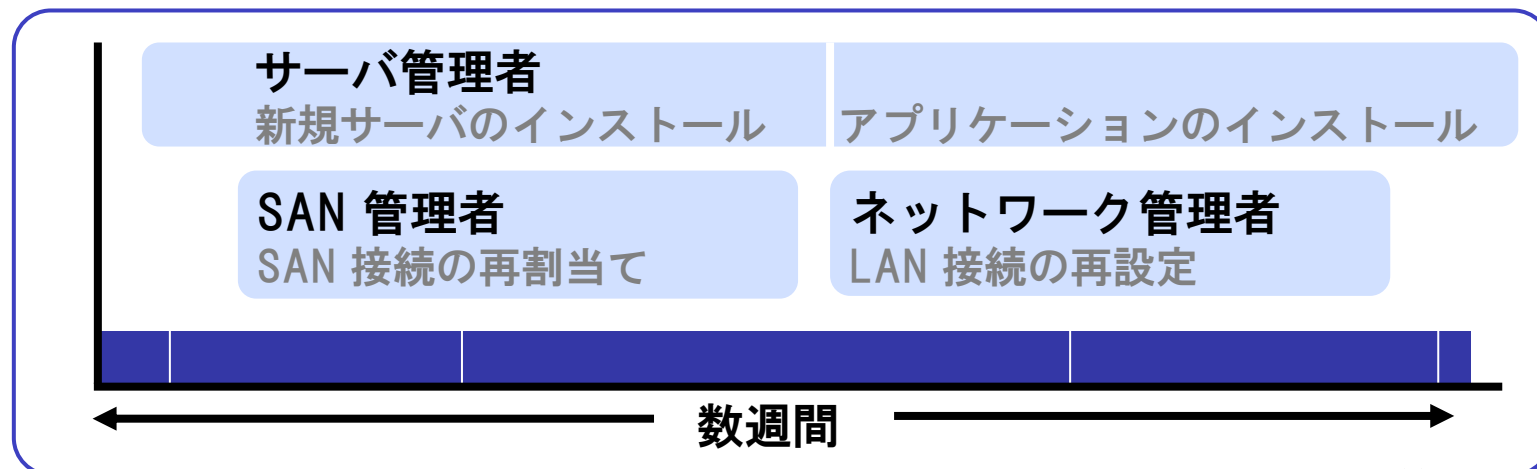


経営層からの要求
運用コスト削減
クリア！

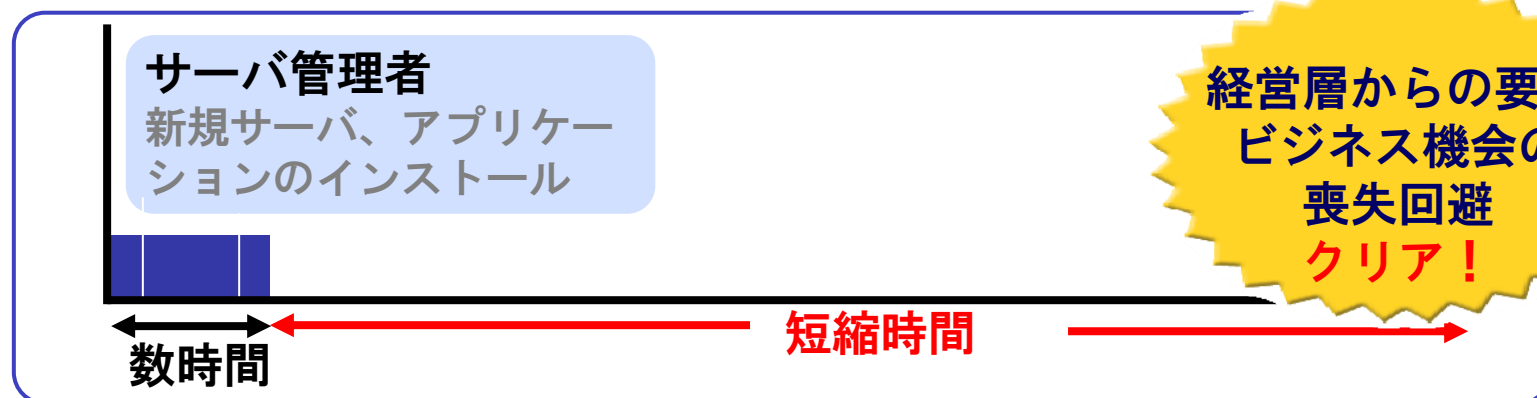
サービスをアップを急げ！

システムの設定変更や再構築を完了までに要する時間

現在



Xsigo



数百倍の速さで システムの再構築が可能！

まとめ：仮想I/Oの導入効果

複雑化するインフラには・・・

I/O統合によりケーブルを削減し、システム構成を簡素化。消費電力の削減によりCO2排出量を削減！

運用は火の車、そこで・・・

便利な仮想 I/O運用管理ツールやサーバ I/Oの集中統合管理により、日々のシステム運用をもっと楽に

柔軟性のないインフラには・・・

広帯域 I/Oにより I/Oボトルネックを排除し、I/O仮想化により自由自在にリモートから、I/O設計・構築・拡張を可能に！

設備はパンク寸前、そこで・・・

様々な物理 I/O 制限を取り除き、ラックスペースの有効活用と、仮想化環境を最適化！

I/O 仮想化といえば . . .

Issei Ogata

「イツセー おがた」まで

iogata@xsigo.com