



Windows Server 2012 Community Day ～ Night Session, April 2013 ～

Windows Server 2012 Hyper-V の 本格採用に必要な エンタープライズ設計術



小川 大地

Microsoft MVP for Virtual Machine
日本ヒューレット・パッカー



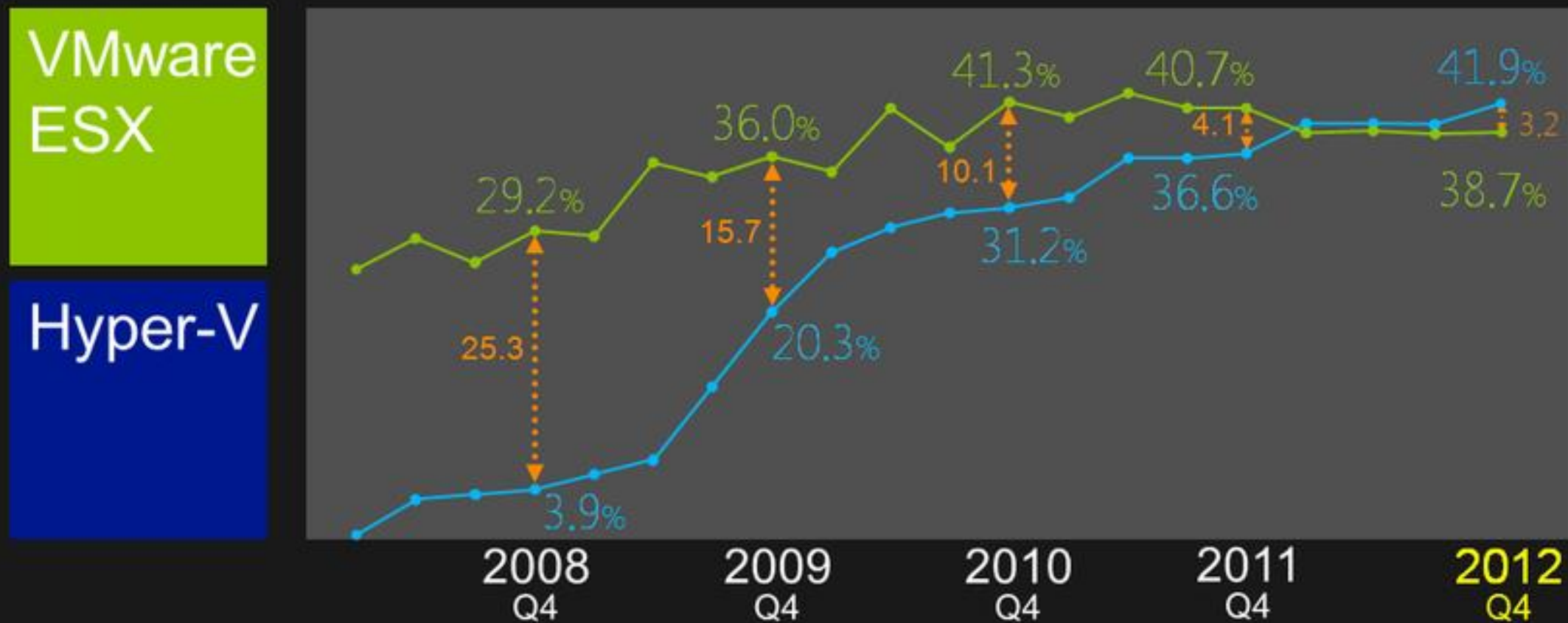
自己紹介

- 氏名： 小川 大地 (*Daichi Ogawa*)
- 仕事： サーバー・テクノロジーの啓蒙と、システム設計・実提案
(今日は Hyper-V の MVP として呼ばれたので、会社は無関係です)
- Blog： 仮想化でプリセールスしてる SE の一日
<http://d.hatena.ne.jp/ogawad/>

IDC のデータで Hyper-V が通年 No.1 になったそうです

2012年 仮想化プラットフォーム No.1を獲得

Windows Server 2012 Hyper-Vで導入が加速



注記：Virtualization Host 数のシェア、Virtualization Host とは新規出荷、既設サーバーにかかわらず、仮想化プラットフォーム (ESX, Hyper-Vなど) を導入したサーバーを表す。
出典：IDC Worldwide Quarterly Server Virtualization Tracker CY2012 Q4 承諾番号：IDCJ-13-0374-1

Microsoft 資料より

Hyper-V は本当に大規模システムでも伸びてる？

エンタープライズへの浸透まであと一歩

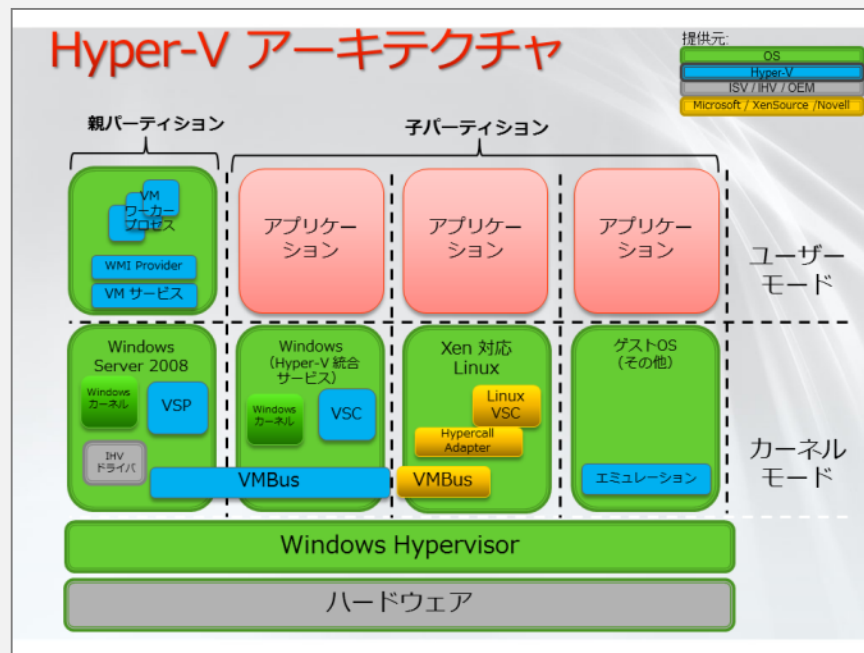
- 引き合いは着実に増えています。
- ROI を考えれば機能も性能も十分です。
- ただ1つだけ、、、
他社製品と比べて、エンタープライズ規模に対応するための設計や構築情報・ノウハウがあまりに少ない。。。
(ノウハウがないことで、SI 費用が変に高かったり ...)

本セッションでは、情報の少ないエンタープライズ設計のノウハウについて解説します

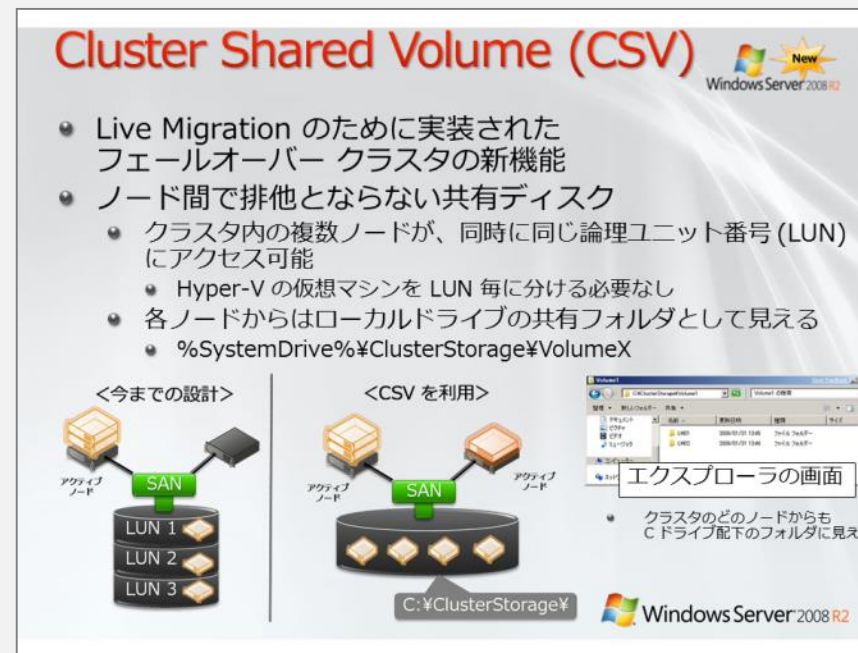
Hyper-V の基本アーキテクチャに変更はない

基本知識は従来バージョンと変わらない

マイクロカーネル型ハイパーバイザー

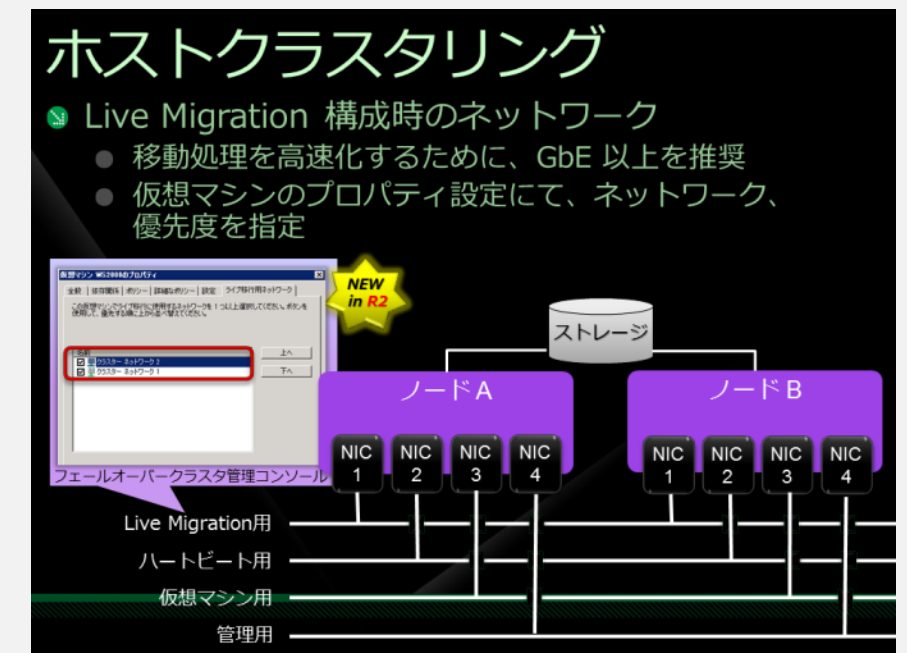


WSFC + CSV による高可用性



- CSV バージョン 2 になり、互換性が向上しました
- SMB ファイルサーバーを利用する場合、CSV は不要です

ネットワークは 4+1 系統が推奨



- 内訳は下記のとおり
「管理用」「仮想マシン用」「Heartbeat&CSV」
「Live Migration」 (+ IP ストレージ)

<http://blogs.technet.com/b/askpfplat/archive/2013/03/10/windows-server-2012-hyper-v-best-practices-in-easy-checklist-form.aspx>

“使える” 新機能と “使えない” 新機能を見定める

実際に “嬉しい” 新機能は何か？

Windows Server 2008 R2 Hyper-V		Windows Server2012 Hyper-V		VMware vSphere 5.1 Enterprise Plus
仮想化ホスト				
論理プロセッサ (LP)	64 LP	→ 5倍 320 LP		160 LP
メモリ	1 TB	→ 4倍 4 TB		2 TB
仮想マシン				
仮想プロセッサ (VP)	最大 4 VP	→ 16倍 最大 64 VP		最大 64 VP
仮想メモリ	64 GB	→ 15倍 1TB		1 TB
仮想ハードディスク	2 TB	→ 32倍 64 TB		2TB
仮想マシン数 (VM/ホスト)	384 VM	→ 2.7倍 1,024 VM		512 VM

有用なのは
2 つくらい

Microsoft 資料より抜粋

“使える” 新機能と “使えない” 新機能を見定める

実際に “嬉しい” 新機能は何か？

有用なのは
2 つくらい

仮想化ホスト		Windows Server 2008 R2 Hyper-V	→	Windows Server 2012 Hyper-V	→	VMware vSphere 5.1 Enterprise Plus
論理プロセッサ (LP)		64 LP		5倍 ↑ 320 LP		160 LP
メモリ		1 TB		4倍 ↑ 4 TB		2 TB
仮想マシン						
仮想プロセッサ (VP)		最大 4 VP		16倍 ↑ 最大 64 VP		最大 64 VP
仮想メモリ		64 GB		15倍 ↑ 1 TB		1 TB
仮想ハードディスク		2 TB		32倍 ↑ 64 TB		2 TB
仮想マシン数 (VM/ホスト)		384 VM		2.7倍 ↑ 1,024 VM		512 VM

➡ ゲストに CPU を
4 コア以上割当可能に

➡ 仮想ディスク形式で
2TB 以上を割当可能に

Microsoft 資料より抜粋

エンタープライズの仮想化設計のために重要なこと



可用性

- ハードウェアだけでなく、“インフラ”として死活監視・冗長化する



性能

- CPU から I/O まで全体バランスを保つ（オーバーヘッドの抑止）



運用管理（時間の都合で今日は扱いません）

- 仮想化対象の拡大に対して、管理者の負担をどこまで減らせるか？

Agenda

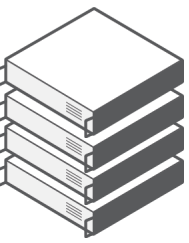
クラスタリング



ネットワーク

ストレージ
・
バックアップ

エンタープライズで求められる “可用性レベル”



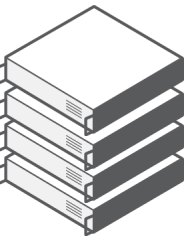
例えば... 「Web サーバーの冗長化」

- Windows Server には、かねてより NLB（ネットワーク負荷分散）がある。
- 小規模や重要度の低いシステムでは NLB が使われているが、本格的なシステムになると、Web サーバーが Windows（IIS）であってもハードウェア・ロードバランサーを導入するケースがほとんど。



- 理由の 1 つ：

NLB はネットワークの死活監視だけで、HTTP 200 を返すかは見ていないから

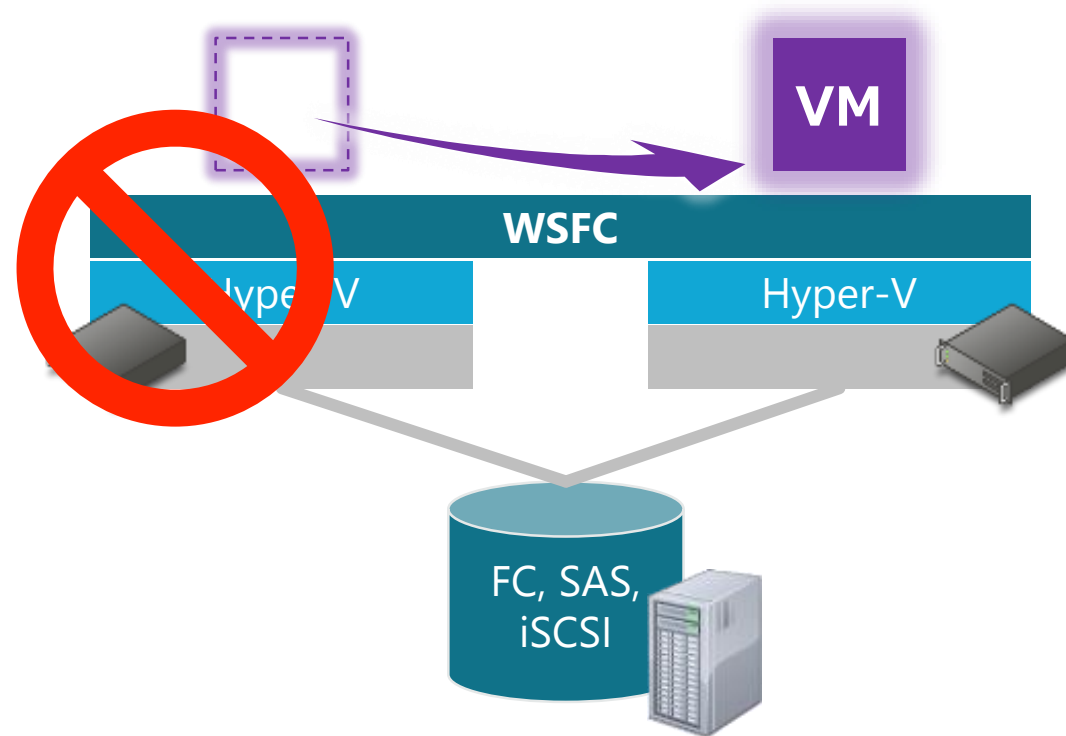


Hyper-V 仮想マシンの可用性を高める

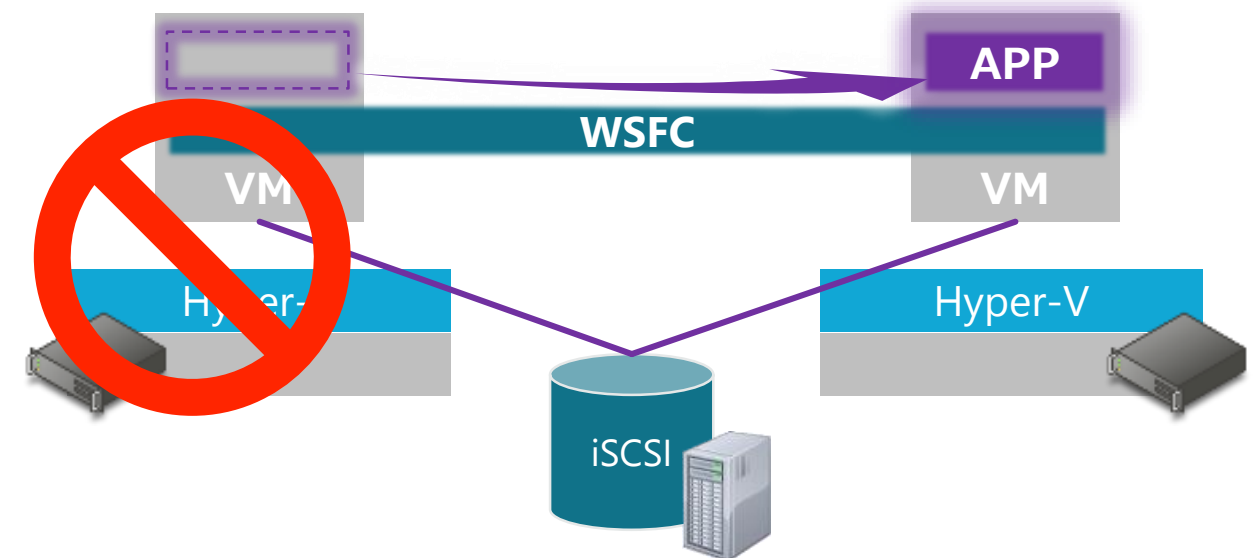
Hyper-V 仮想マシンにおける高可用性技術は？

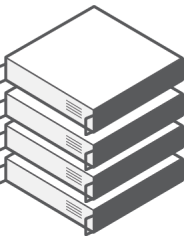
従来より、Hyper-V では WSFC ベースの 2 種類のクラスタリングを用意

ホストクラスタリング (HA : Clustered VM)



ゲストクラスタリング (In-Guest Clustering)





Hyper-V 仮想マシンの可用性を高める

HA（ホストクラスタリング）だけで本当に大丈夫？

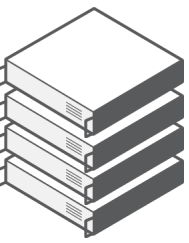
		ホストクラスタ（HA）	ゲストクラスタ（In-Guest）
		WS2008 R2	WS2008 R2
可用性	物理ホストのダウン	●	●
	仮想マシンネットワークの障害	×	●
	ゲスト OS のダウン・フリーズ	●	●
	ゲスト OS の内部障害	×	●
	ゲストのファイルシステム破	×	●
	アプリの全体障害（サービス）	×	●
	アプリの内部障害（プロセス）	×	●
性能	パフォーマンス	●	△
構成	対応する共有ストレージ	FC, SAS, iSCSI	iSCSI

OS 内部の障害や
プロセス障害は
監視すらしていない

チーミングされている
両パスが落ちても
フェイルオーバーしない

WSFC 組むくらい
重要なアプリなのに
性能悪くて良いの？

FC 対応して
ないの？？

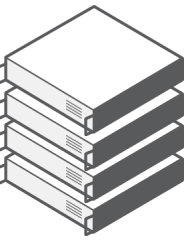


Hyper-V 仮想マシンの可用性を高める

WS2012 Hyper-V ではカバー範囲が広がり、より SLA の高いシステムにも対応

		ホストクラスタ (HA)		ゲストクラスタ (In-Guest)	
		WS2008 R2	WS2012	WS2008 R2	WS2012
可用性	物理ホストのダウン	●	●	●	●
	仮想マシンネットワークの障害	×	×	●	●
	ゲスト OS のダウン・フリーズ	●	●	●	●
	ゲスト OS の内部障害	×	×	●	●
	ゲストのファイルシステム破損	×	×	●	●
	アプリの全体障害 (サービス)	×	●	●	●
	アプリの内部障害 (プロセス)	×	×	●	●
性能	パフォーマンス	●	●	△	● ^{*1}
構成	対応する共有ストレージ	FC, SAS, iSCSI	FC, SAS, iSCSI, SMB3.0	iSCSI	iSCSI, FC, SMB3.0 ^{*2}

*1 In-Guest でハードウェア支援を受けられるのは、FC か SR-IOV 利用時のみ
*2 SMB3.0 に対応したクラスタウェア・アプリケーションは現状まだ少ない



Hyper-V 仮想マシンの可用性を高める

① ホストクラスタ：ゲストのサービス監視が可能に

ホスト側の WSFC でゲスト内のサービスを監視 (New!)

特徴

- **ゲスト側に手を入れる必要なし**
- Windows OS 標準以外のサービスにも対応
- VMware vSphere には無い機能 (3rd Party 製品が別途必要)

動作原理

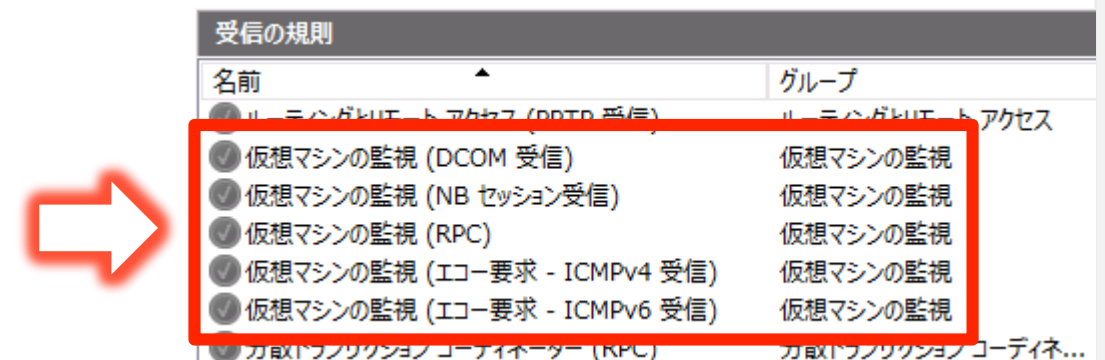
「ゲスト内でサービスが異常終了」すると...

- ゲスト内でサービス再起動を試行 (従来からの OS 標準機能)
- **同一ホストでゲストを再起動*** (WS2012 ホストクラスタ新機能)
- **別のホストでゲストの再起動** (WS2012 ホストクラスタ新機能)

* 自動再起動は行わず、ホストに通知だけ行うことも可能

要件

- ホストOS： Windows Server 2012 Hyper-V
- ゲストOS： Windows Server 2012 のみ
- ネットワーク：
 - ホストとゲストでネットワーク疎通できること
(統合サービスのバックドア通信ではなく、通常ネットワーク)
 - ゲスト上で対象のファイアウォールが開いていること

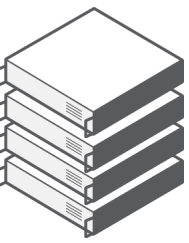




- ## ホスト側の WSFC でゲスト内のサービスを監視 (New!)

通知だけにするか
自動再起動を行うか

Demonstration



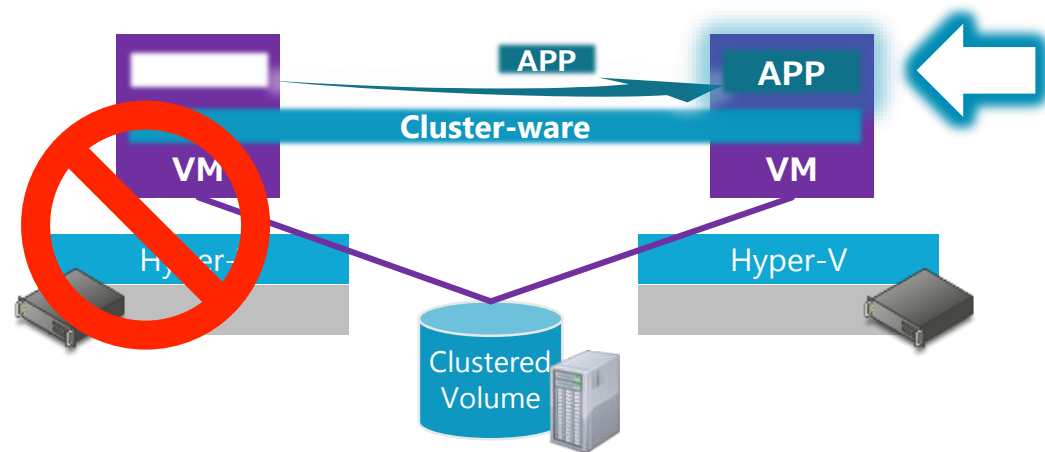
Hyper-V 仮想マシンの可用性を高める

② ゲストクラスタ： パフォーマンスの向上と対応ストレージの拡大

In-Guest Clustering

特徴

- ゲスト OS 上で WSFC などのクラスタリングを実行
- Windows Server 2012 以外の OS でも利用可能
- **アプリ・サービス単位での可用性**
 - ソフトによってはサービス内の正常性検知にも対応



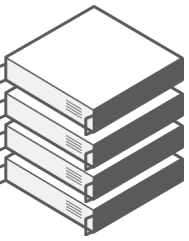
対応ストレージが拡大 (New!)

Hyper-V バージョン	WS2008 R2	WS2012
FibreChannel	×	○
Shared SAS	×	×
iSCSI	△	△
iSCSI + SR-IOV	×	○
SMB 3.0*	×	△
SMB 3.0* + SR-IOV	×	○
SMB 3.0* + Infiniband	×	×

○ 主にハードウェア処理 △ ソフトウェア処理（ゲストの vCPU） × 非対応

* SMB3.0 に対応したクラスタウェア・アプリケーションは現状まだ少ない

Hyper-V 仮想マシンの可用性を高める

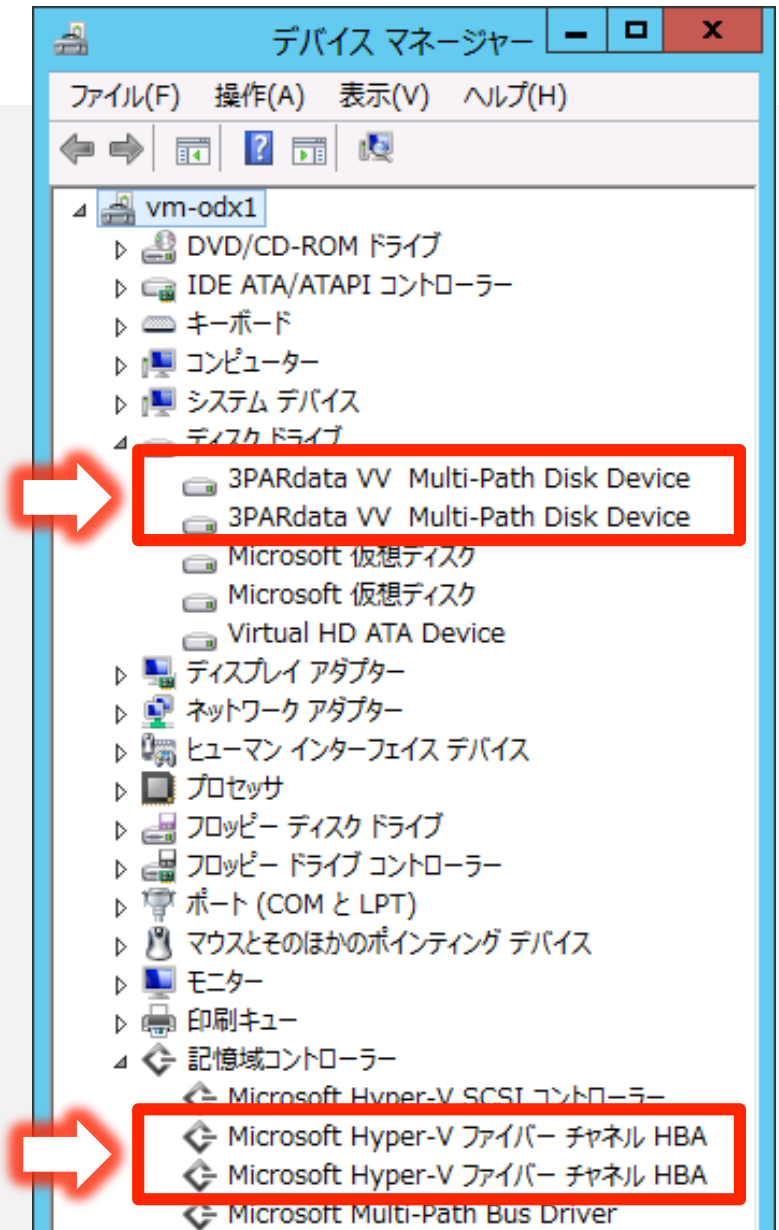
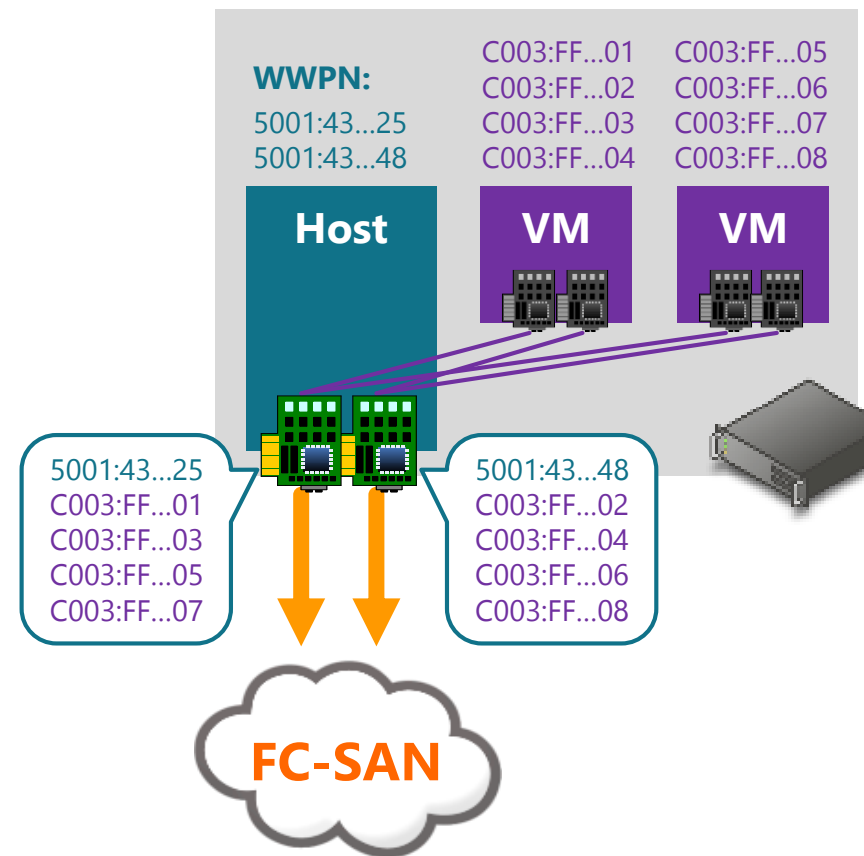


② ゲストクラスタ：パフォーマンスの向上と対応ストレージの拡大

仮想ファイバーチャネル (Synthetic FibreChannel)

NPIV* 技術の利用

- ゲスト OS に一意の FC WWN を付与して FC-SAN デバイスを直接マウント
- 非常に高い互換性
 - Live Migration に対応
 - vHBA あたり 2 つの WWPN を付与
 - パスの二重化 (MPIO) に対応
 - 仮想化の制約を解消
 - FC テープ装置の接続
 - SCSI コマンドの発行
 - 既存の NPIV 機器とのネストも OK

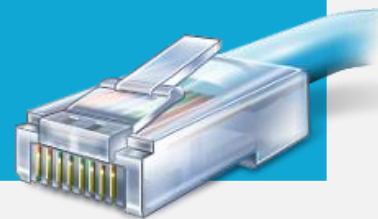


* NPIV = N_Port ID Virtualization

Agenda

クラスタリング

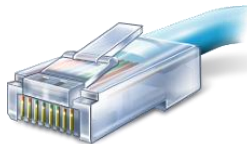
ネットワーク



ストレージ
・
バックアップ

NIC チーミングのベストプラクティス

誤った設定によるネットワークトラブルが非常に多い



WS2012 “LBFO” で利用できるモードのまとめ

^ 追加のプロパティ(A)

チーミング モード(T):

スウィッチに依存しない

負荷分散モード(L):

静的チーミング

スタンバイ アダプター(S):

スウィッチに依存しない

プライマリ チーム インターフェイス:

LACP

(自動的に生成された名前); 既定の VLAN

チーミングモード	対向ポートの要件	負荷分散	送信アルゴリズム
スウィッチに依存しない (既定)	なし	なし (Active-Standby)	
		送信のみ Active-Active	• Hyper-V Port • ハッシュ (MAC)
静的チーミング	Link Aggregation ネゴシエーション=なし ^{*1}	送受信 Active-Active	• Hyper-V Port • ハッシュ 以下より自動選択
LACP	Link Aggregation ネゴシエーション=lacp		• MAC Only • MAC + IP • MAC + IP + TCP Port#

他社での
呼び名^{*2}



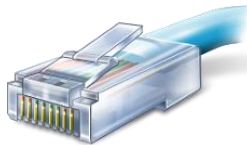
VMware vSS	HP NCU	Intel PROSet
明示的順序	NFT	AFT
VM Port ID MAC ハッシュ	TLB	ALB
IP ハッシュ	SLB	SLA
なし ^{*3}	802.3ad Dynamic	DLA

*1 Cisco Catalyst における EtherChannel “on”

*2 完全に同義でないものもあります

*3 VMware vSphere で LACP を利用するには vDS が必要です

NIC チーミングのベストプラクティス

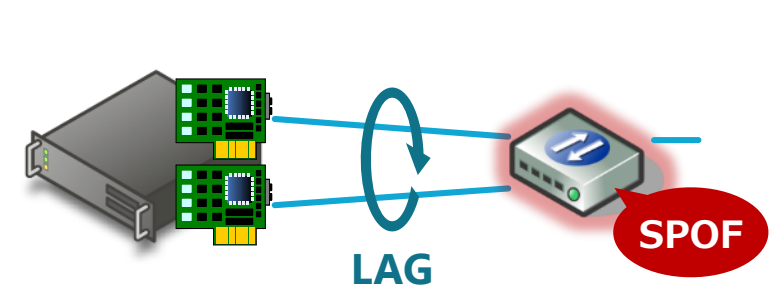


LACP の採用は、対向が機器をまたいで Link Aggregation が組めるスイッチ* の場合のみ！

モード選択のベストプラクティス

* Cisco StackWise, VSS, vPC や HP IRF など
(スタック対象が管理プレーンのみのスイッチは不可)

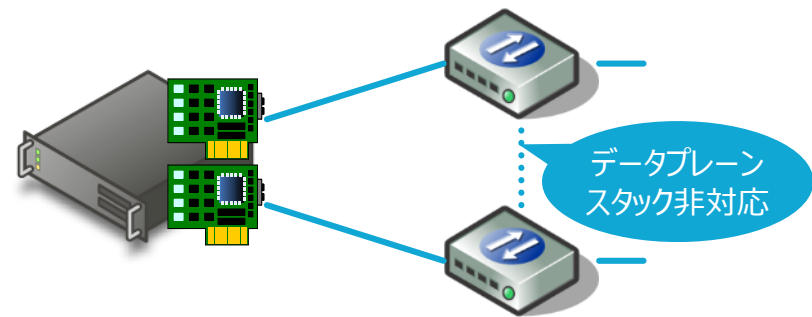
スイッチを冗長化しない場合



× スイッチが SPOF で障害耐性が無い

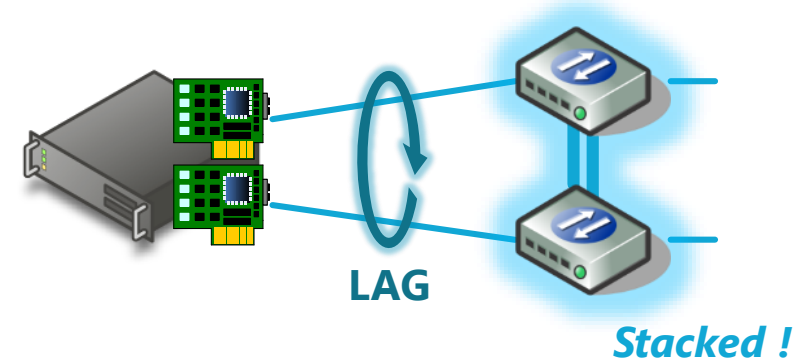
チーミングモード	利用可否
スイッチに依存しない	○
静的チーミング	○ 機能があれば
LACP	○ 機能があれば

スイッチがスタック非対応の場合



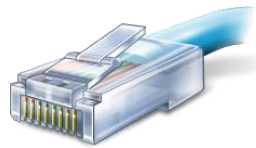
チーミングモード	利用可否
スイッチに依存しない	○
静的チーミング	×
LACP	×

スイッチがスタック可能な場合



チーミングモード	利用可否
スイッチに依存しない	○
静的チーミング	○
LACP	○ 推奨

Hyper-V ネットワーク設計のベストプラクティス

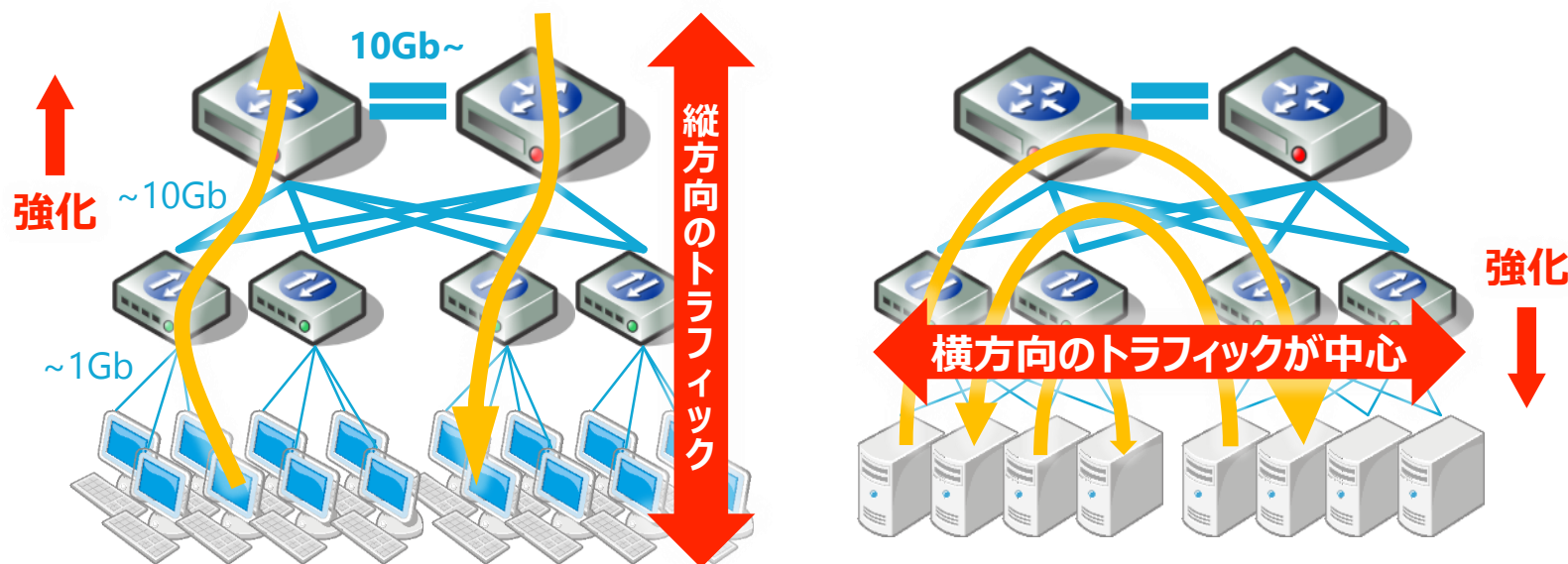


3つの視点での考慮が必要

① 仮想化基盤特有のトラフィックを理解

伝統的な“階層ツリー”設計に注意

- クライアントデバイス ... **L3** アクセスが中心 (Web, メール, クラサバ, ...)
- サーバー (特に仮想化) ... **L2** アクセスが中心 (DB アクセス, CSV, Migration...)



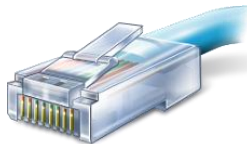
② ハードウェアトレンドの把握

10Gbps Ethernet の一般化

- 10GbE NIC を標準搭載するサーバーが増加
- イーサネットとストレージの統合 (CNA)
- 安価なRJ-45 ケーブルの使える **10GBase-T** への普及が加速



Hyper-V ネットワーク設計のベストプラクティス



3 つの視点での考慮が必要

③ Hyper-V のアーキテクチャを考慮

今回も、サーバーあたり 最低 4 + 1 系統が推奨

- HA 環境では最低 4 セグメント + ストレージ
 1. ホスト管理用
 2. CSV & クラスタハートビート用
 3. Live Migration 用
 4. IP ストレージ用 (iSCSI や SMB で接続する場合)
 5. 仮想マシン用 (複数の VLAN に接続する場合も多い)

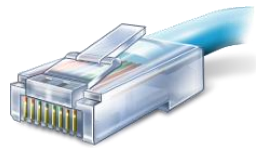
※ 10GbE 環境で親パーティションに接続する 1~4 を仮想スイッチ経由にするとスイッチング処理の増加でホストの CPU 負荷が上がることに注意 (→P.26)



	親パーティション	用途	帯域
NIC #1	接続する	ホスト管理	2Gbps
NIC #2			
NIC #3		CSV & クラスタハートビート	2Gbps
NIC #4			
NIC #5		Live Migration	2Gbps
NIC #6			
NIC #7	接続しない	(IP ストレージ)	2Gbps 帯域不足
NIC #8			
NIC #9		仮想マシン	2Gbps 帯域不足
NIC #10			

NIC を 10 枚挿してもまだ足りない ...

Hyper-V ネットワーク設計のベストプラクティス



Microsoft のベストプラクティス構成（10GbE + VLAN + QoS）を具現化

Microsoft のベストプラクティス

Windows Server 2012 Hyper-V Best Practices

Please note:

*we fully support and even recommend (in some cases) using the virtual switch to separate networks for Management, Live Migration, CSV/Heartbeat and even iSCSI. For example **two 10GB NIC's that are split out using VLANs and QoS.***

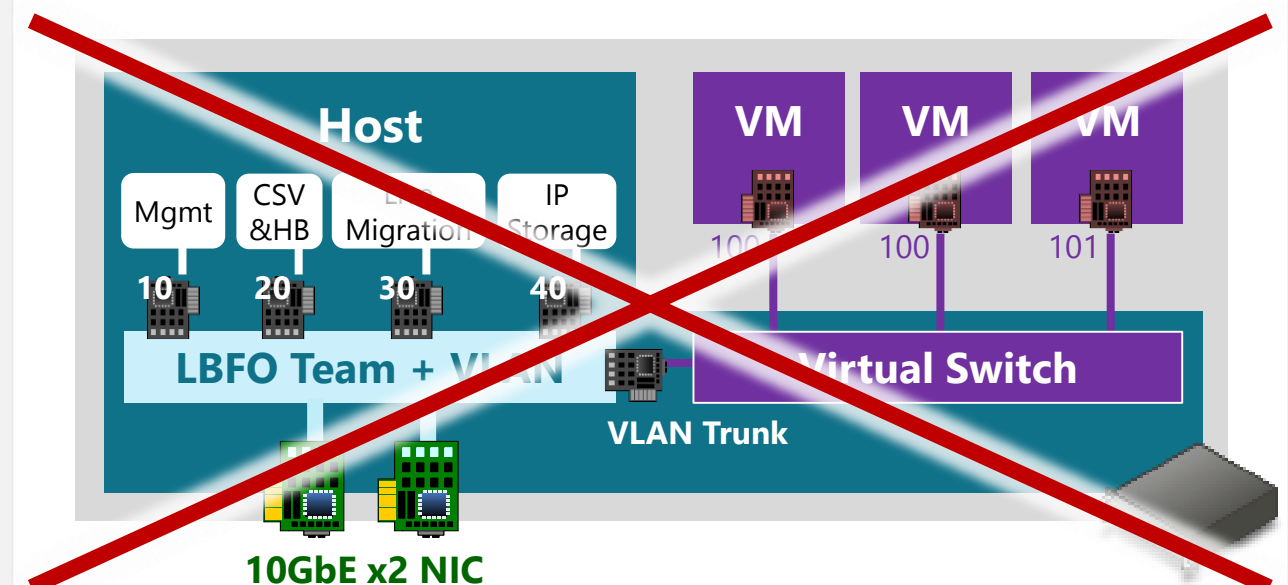
Min # of Networks on Host	Host Management	VM Network Access	CSV/Heartbeat	Live Migration	iSCSI
5	"Management"	"Production"	"CSV/Heartbeat"	"Live Migration"	"iSCSI"

<http://blogs.technet.com/b/askpfeplat/archive/2013/03/10/windows-server-2012-hyper-v-best-practices-in-easy-checklist-form.aspx>

優秀なアーキテクトほどやってしまうミス

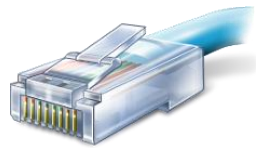
LBFO の VLAN 機能を利用した設計

- 各管理系ネットワーク → LBFO で VLAN 分割
- 各仮想マシン → 仮想スイッチ経由



LBFO の VLAN 分割と仮想スイッチの VLAN 機能は共存 NG
(WS2008R2 で一部話題になったのと同じ技術的理由)

Hyper-V ネットワーク設計のベストプラクティス



Microsoft のベストプラクティス構成（10GbE + VLAN + QoS）を具現化

ベストプラクティス ① - Hyper-V の機能だけでソフトウェア実装

two 10GB NIC's that are split out using VLANs and QoS.

- LBFO からのダウンリンクは、一旦すべて仮想スイッチに接続し、仮想スイッチの親パーティションポートを増やしてマップする

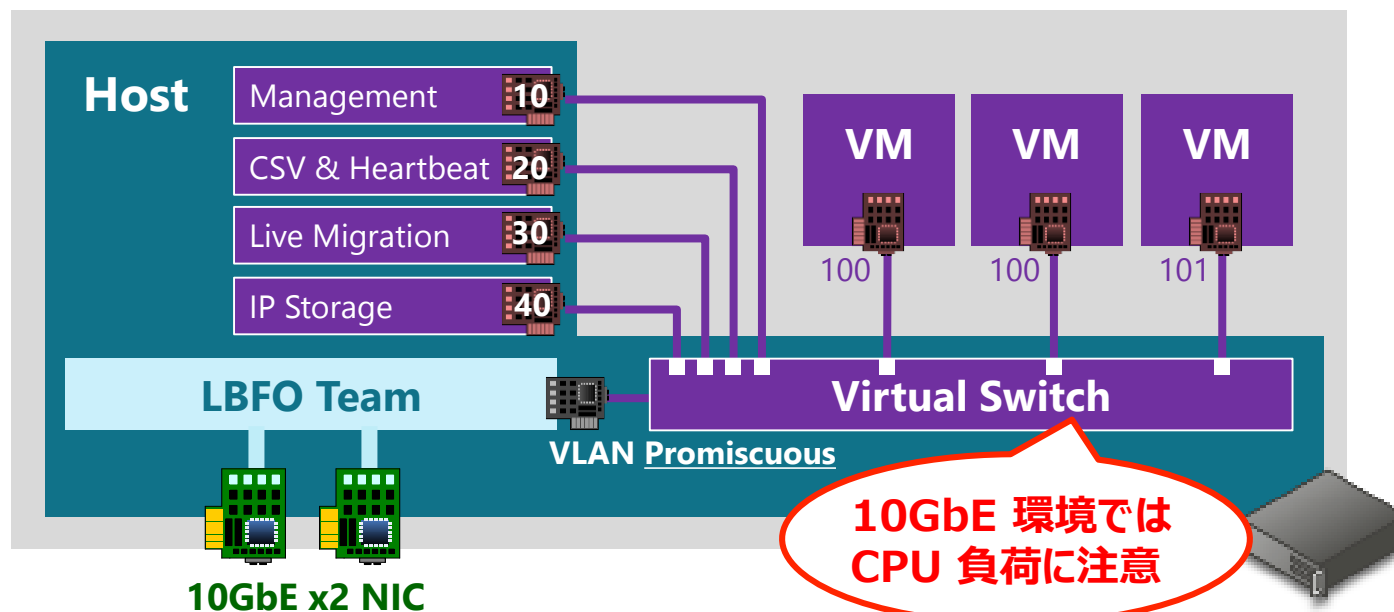
- Hyper-V マネージャーでは親パーティションポートを 1 個しか作れないため、PowerShell での設定が必要

```
PS C:\> Add-VMNetworkAdapter -ManagementOS  
-Name "Live Migration" -SwitchName "vSwitch1"
```

```
PS C:\> Set-VMNetworkAdapterVlan -ManagementOS  
-VMNetworkAdapterName "Live Migration" -Access -VlanId 30
```

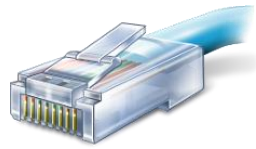
- 作成したポートに対して QoS（帯域制御）を設定できるが、ルールを正確に理解しておかないと性能低下や起動不可になる

```
PS C:\> Set-VMNetworkAdapter -ManagementOS  
-Name "Live Migration" -MinimumBandwidthWeight 20
```



Hyper-V ネットワーク設計のベストプラクティス

Microsoft のベストプラクティス構成（10GbE + VLAN + QoS）を具現化

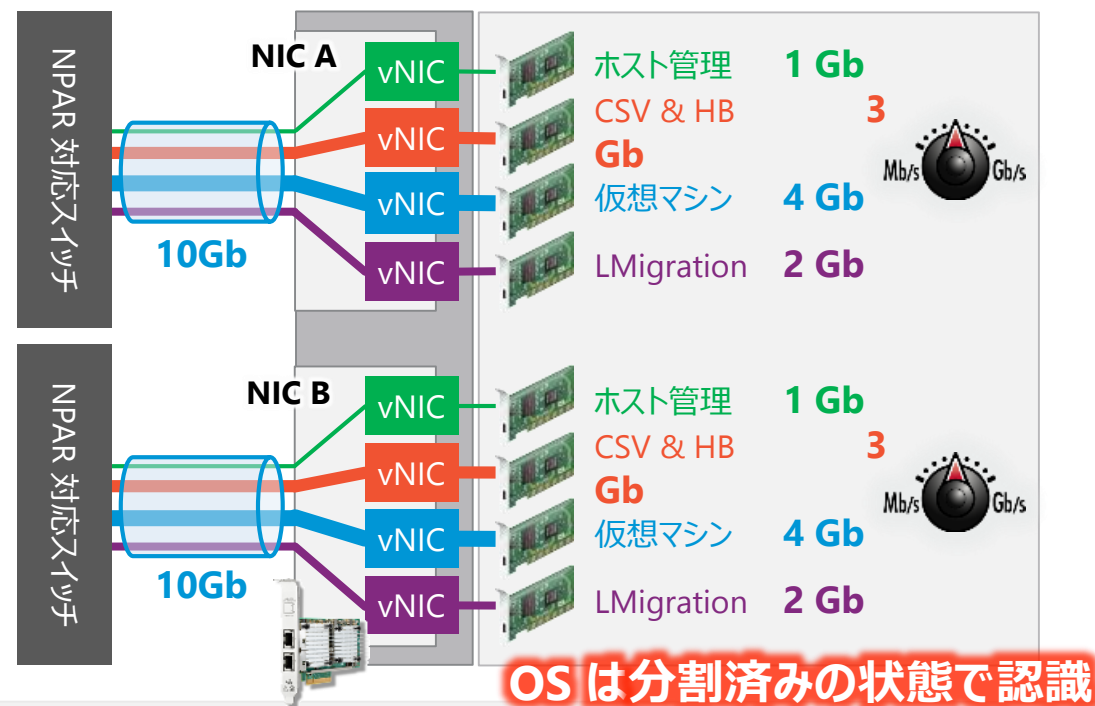


ベストプラクティス ② - NIC のパーティショニング機能* を利用

* Cisco VM-FEX, Dell NPAR, IBM Virtual Fabric, HP VirtualConnect Flex-10/FlexFabric

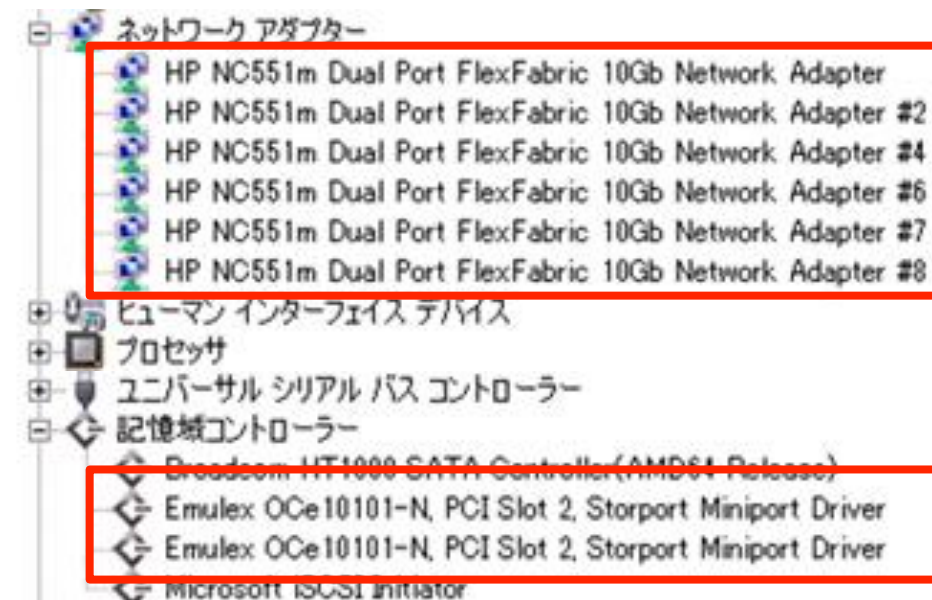
ハードウェアレベルで NIC 分割

- ハードウェアで処理するため CPU に負担が掛からず、設定もシンプル

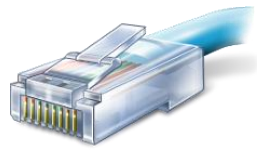


コストメリット

- GbE NIC を 10 枚近く増設して、束ねて使うよりも安価
- ストレージ HBA まで統合できる製品もあり（FC, HW-iSCSI）



SR-IOV の原理と効果

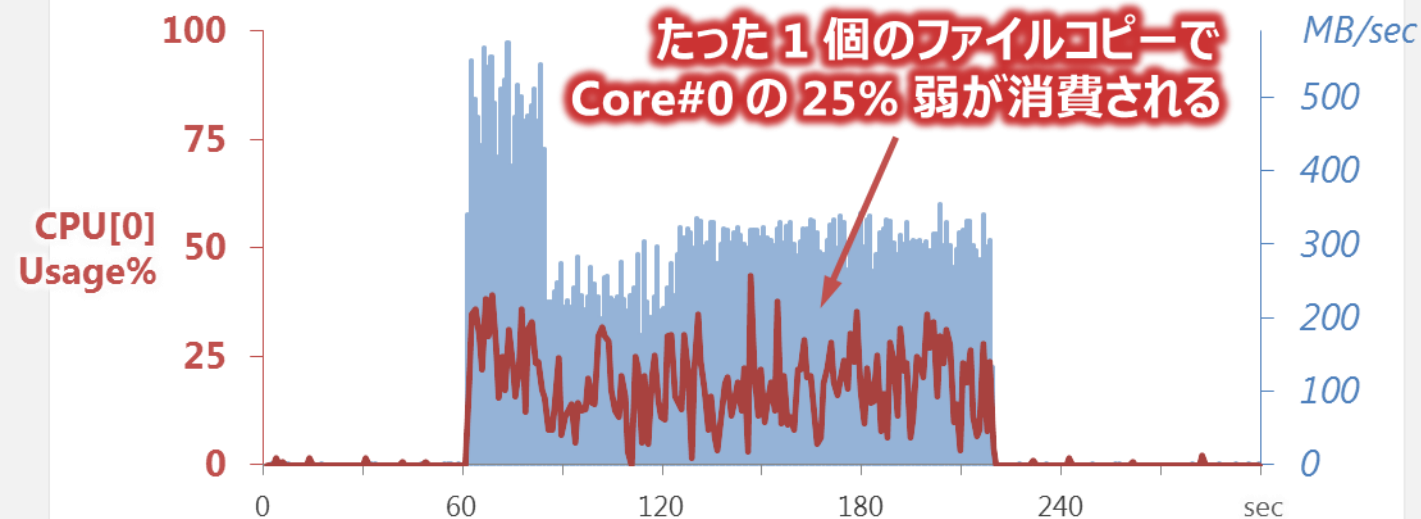
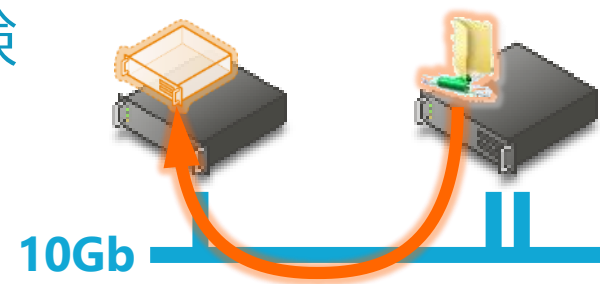


10GbE 環境で課題となる、仮想スイッチの過負荷と遅延問題を改善

10Gbps クラスの速度でデータを転送すると、ホストの CPU を簡単に振り切ってしまう

最新機器* で簡単な実験

- 仮想マシンを 1 台だけ動かし、SMB 3.0 ファイルサーバーから 50 GB のファイルをコピー

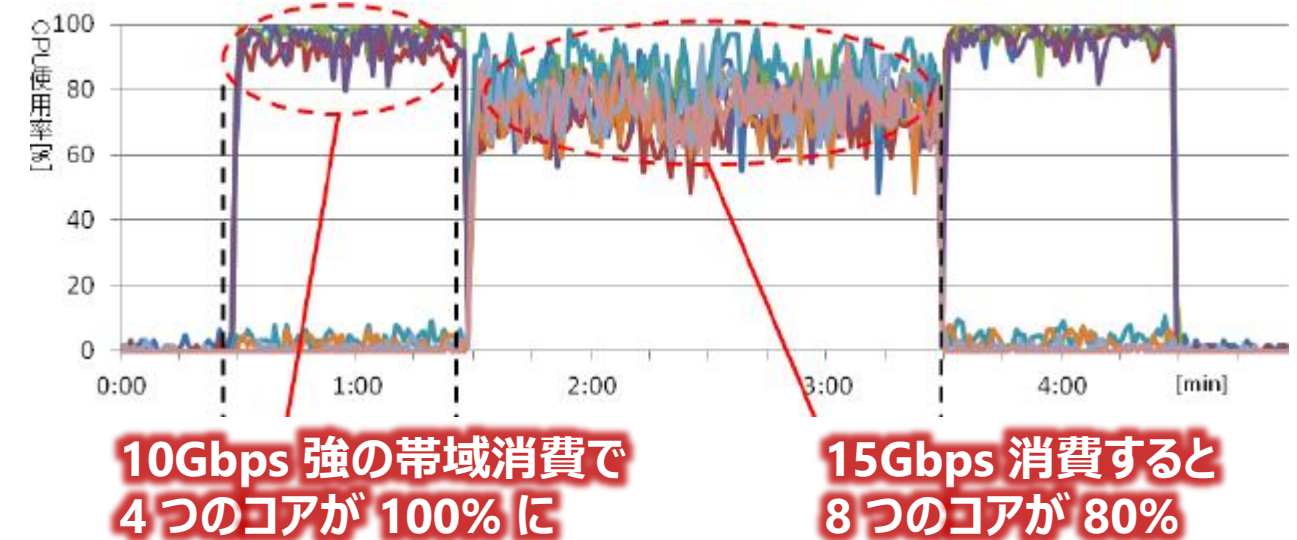


* CPU: Intel Xeon E5-2670@2.60GHz 16-Cores NIC: Intel X540-AT2 (82599)
VM: 2 vCPU, 8GB vRAM, VMQ on, VLAN なし, Windows Server 2012

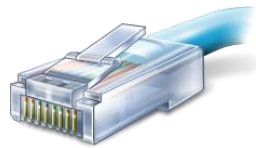
本格的な実験 (富士通さんのホワイトペーパー*より)

- 仮想マシンを 8 台動かし、ベンチマークツールを用いて性能検証

Intel Xeon X5670@2.93GHz 12-Cores



* <http://jp.fujitsu.com/platform/server/primergy/software/windows/technical/tips/win8-beta/pdf/01-03.pdf>



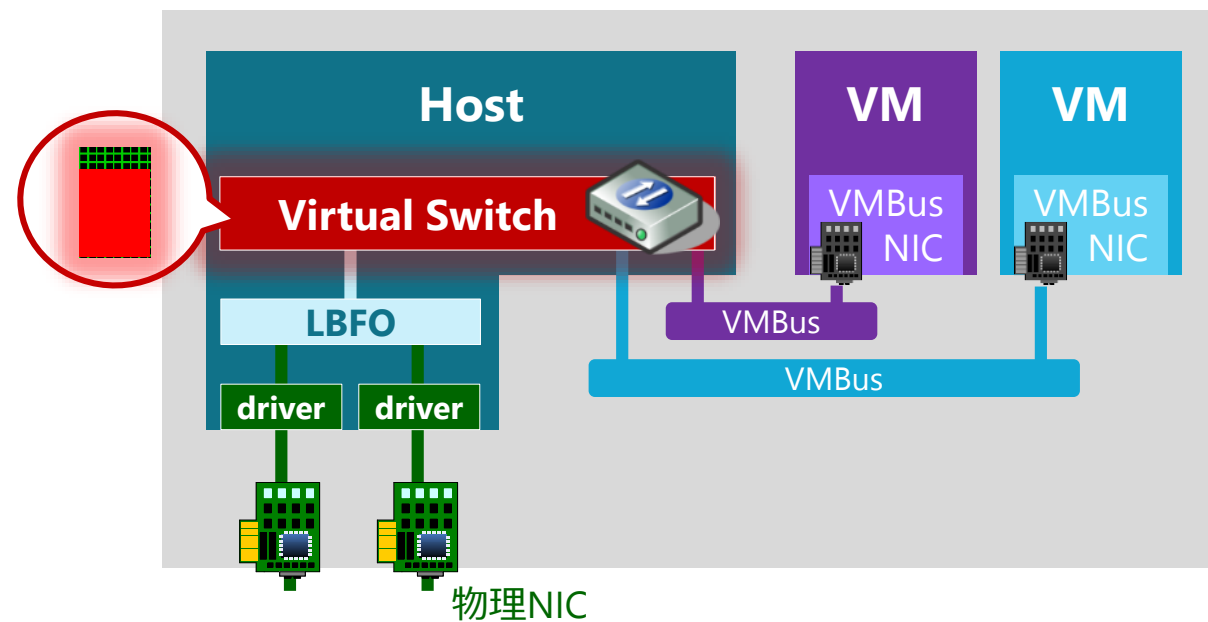
SR-IOV の原理と効果

10GbE 環境で課題となる、仮想スイッチの過負荷と遅延問題を改善

CPU 過負荷の理由：仮想スイッチのソフトウェア処理

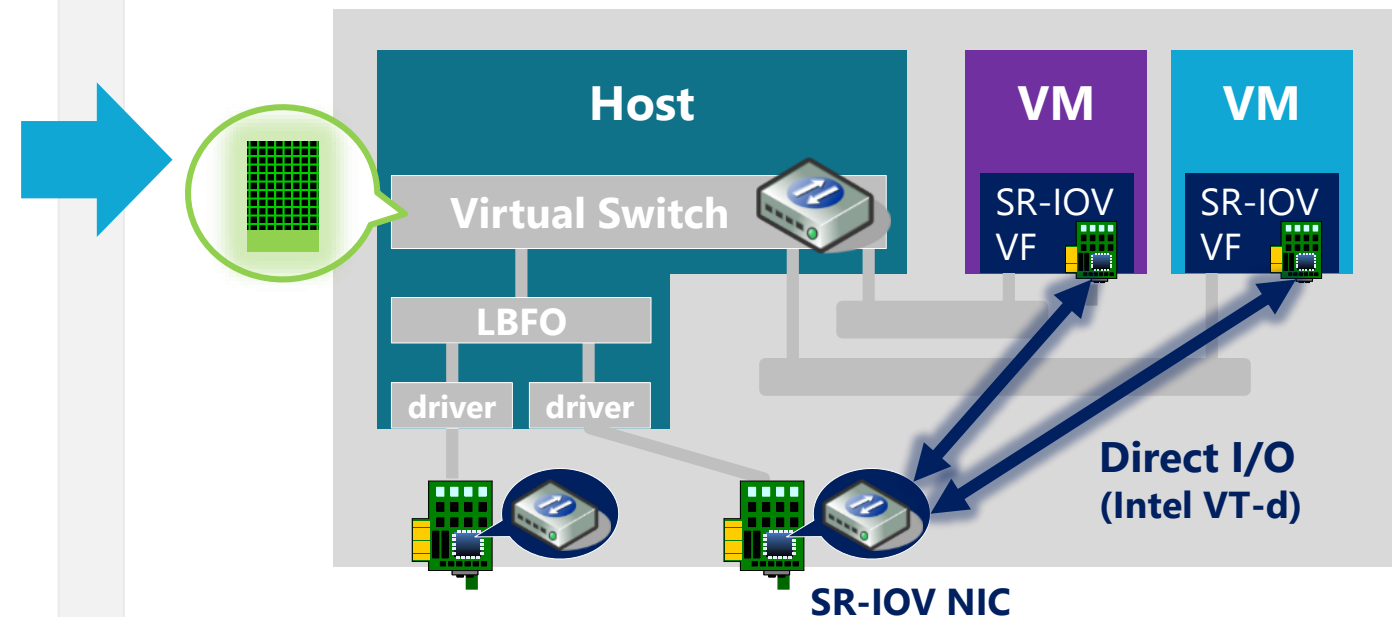
Hyper-V の仮想スイッチ

- サーバーの CPU パワーでスイッチングと転送を担うため、帯域が広いほど負荷が掛かる

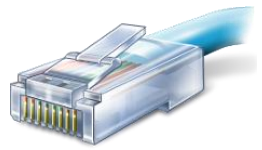


SR-IOV ハードウェアスイッチング

- 仮想スイッチ機能を内蔵した NIC を利用することでサーバーの CPU に負担を掛けない



SR-IOV の原理と効果

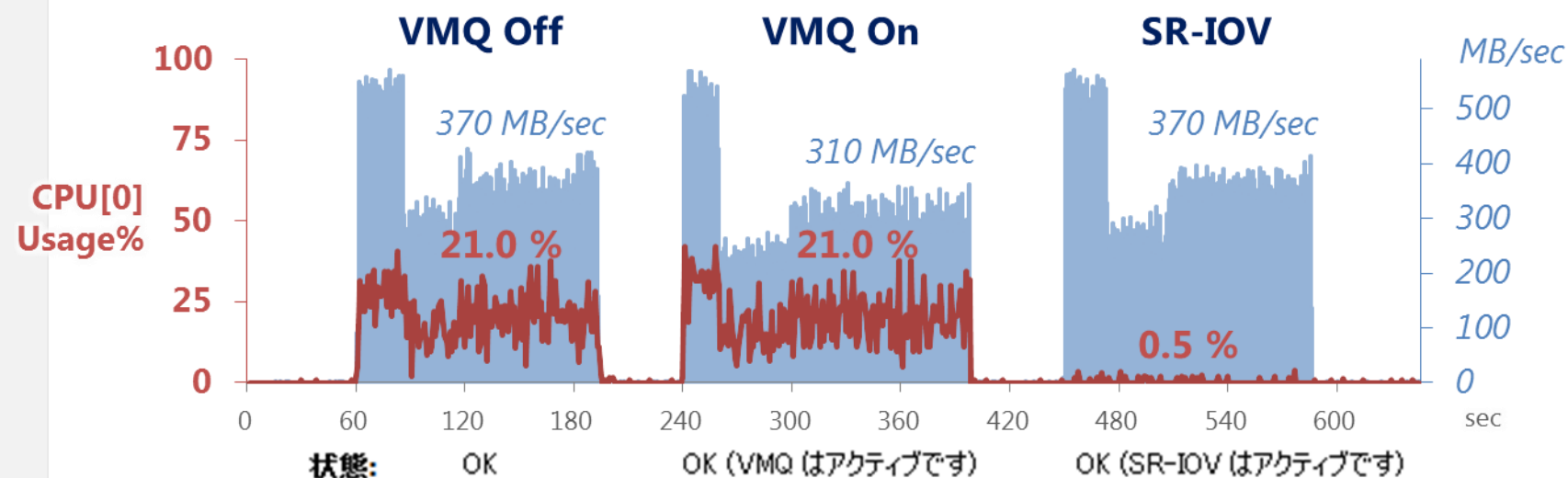


10GbE 環境で課題となる、仮想スイッチの過負荷と遅延問題を改善

SR-IOV の導入効果

SR-IOV の特長

- 仮想スイッチのソフトウェア処理による **CPU 負荷** と **転送遅延** を改善
- スループットや所要時間はさほど変わらない (VMQ=on よりは速い)
- ライブマイグレーションや VM スナップショットに対応 (Hyper-V のみの特殊実装)



SR-IOV の利用要件



- SR-IOV 対応サーバー
- SR-IOV 機能のある NIC
 - 「WS2012 SR-IOV」への対応が必要
 - NICによって内蔵スイッチのポート数 (=VF数) は様々であり、仮想マシンの統合数に影響する
 - 100 VMs 以上を稼働させる VDI には向かない
- ゲスト OS 用の SR-IOV VF ドライバ
 - WS2012 には数種類が INBOX 済み
- Teaming, Jumbo Frame 設定はゲスト側で
 - ホストの仮想スイッチや LBFO を経由しないため
 - WS2012 の LBFO はゲスト上での利用もサポート

CPU: Intel Xeon E5-2670@2.60GHz 16-Cores NIC: Intel X540-AT2 (82599)
VM: 2 vCPU, 8GB vRAM, VLAN なし, Windows Server 2012

Agenda

クラスタリング

ネットワーク

ストレージ
・
バックアップ



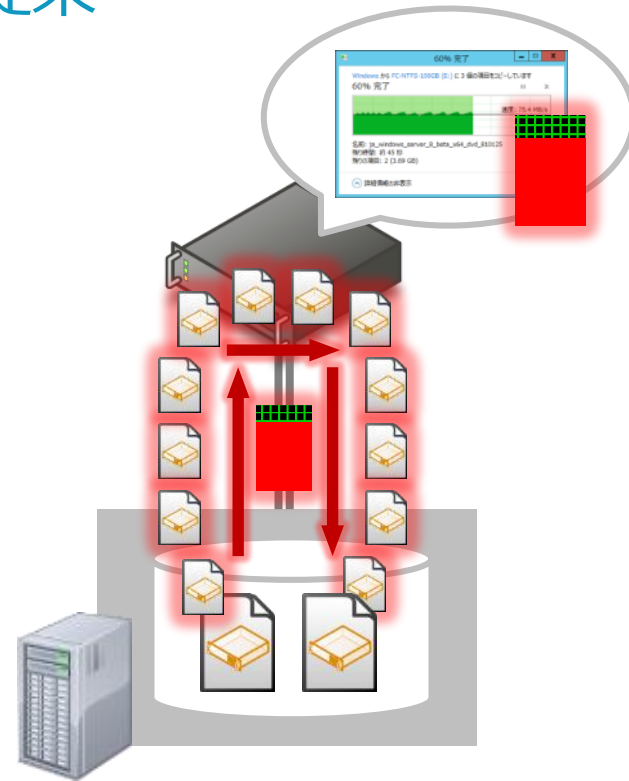
Offloaded Data Transfers (ODX)



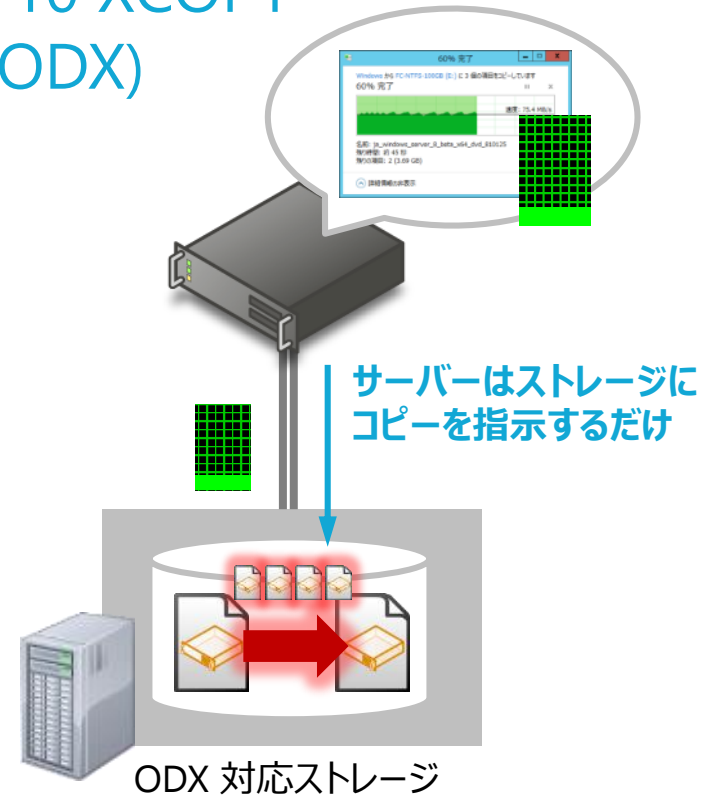
VMware の “VAAI Full Copy” と同じ T10 XCOPY に対応しただけ？

ODX の動作イメージ

従来



T10 XCOPY
(ODX)



ODX ⇒ T10 SCSI XCOPY

- SCSI を策定している T10 委員会が定義した「**SCSI XCOPY**」コマンドをサポート
- VMware は “VAAI Full Copy” として実装済み
 - 仮想マシンクローンやストレージ移行の超高速化
 - しかもホストやストレージパスに負荷を掛けない

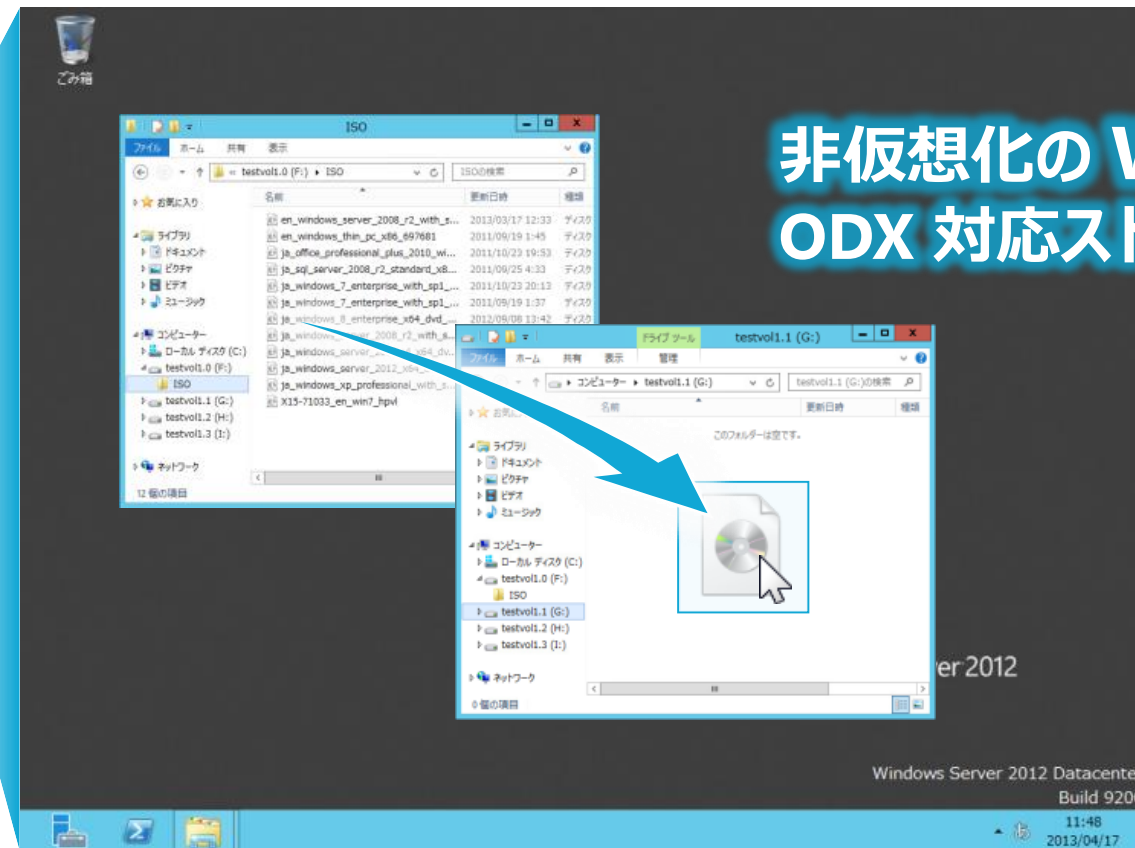
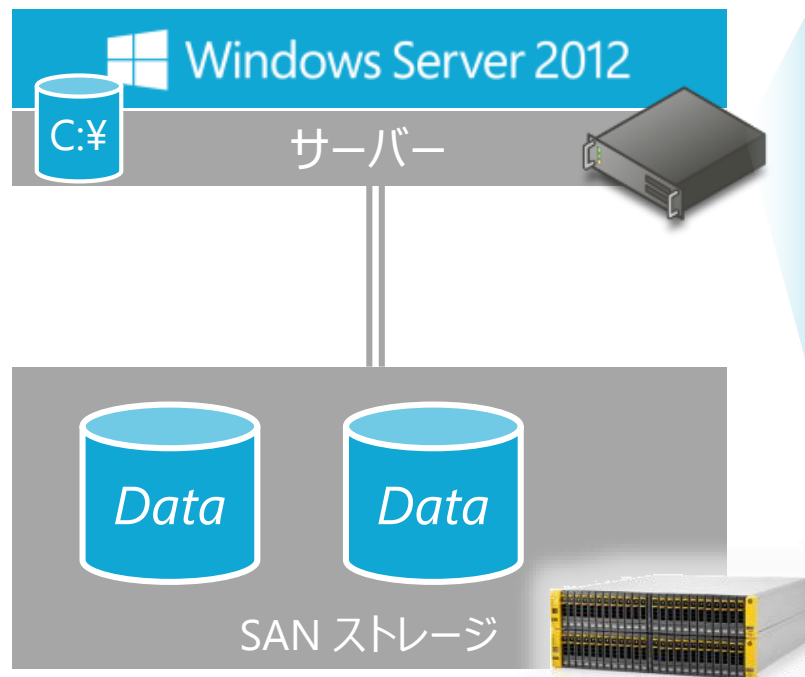
VMware VAAI との違いは、
**ハイパーバイザーではなく
Windows OS として実装**

Offloaded Data Transfers (ODX)



Windows OS レベルで対応したということは、つまり ...

Hyper-V 環境でなくても、...



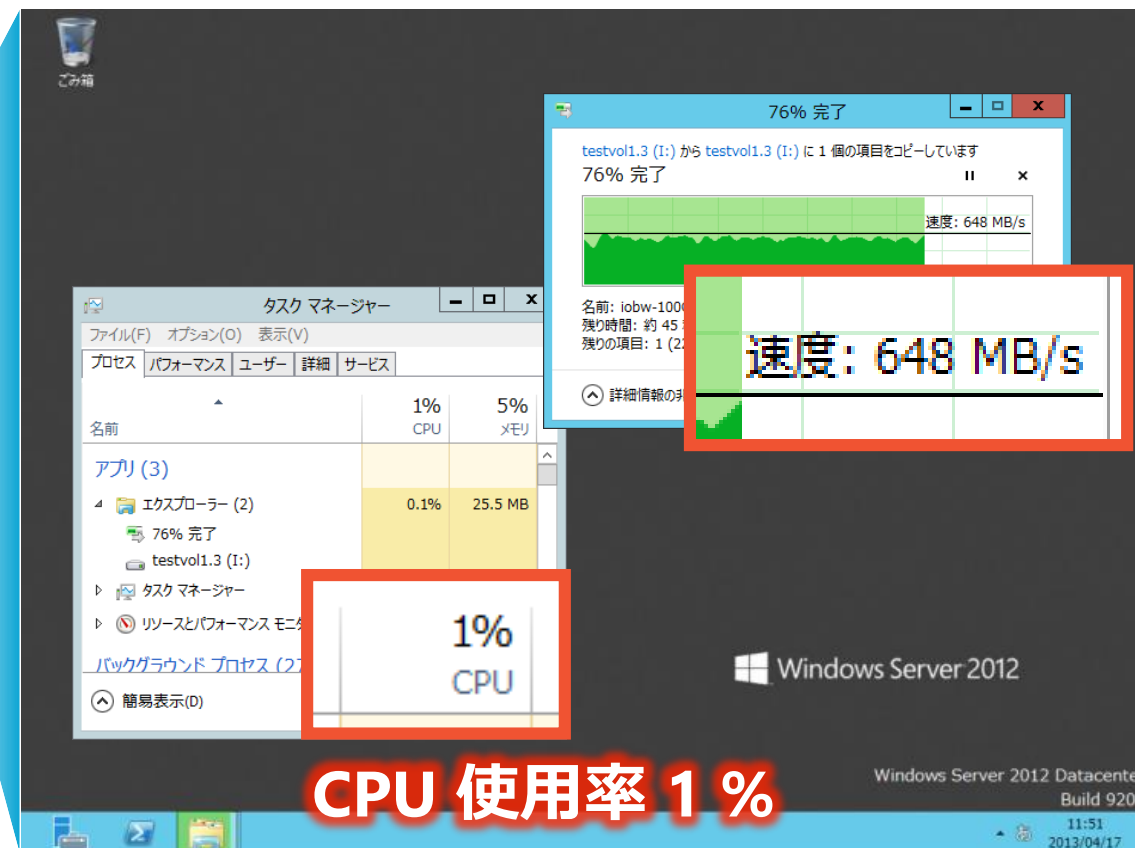
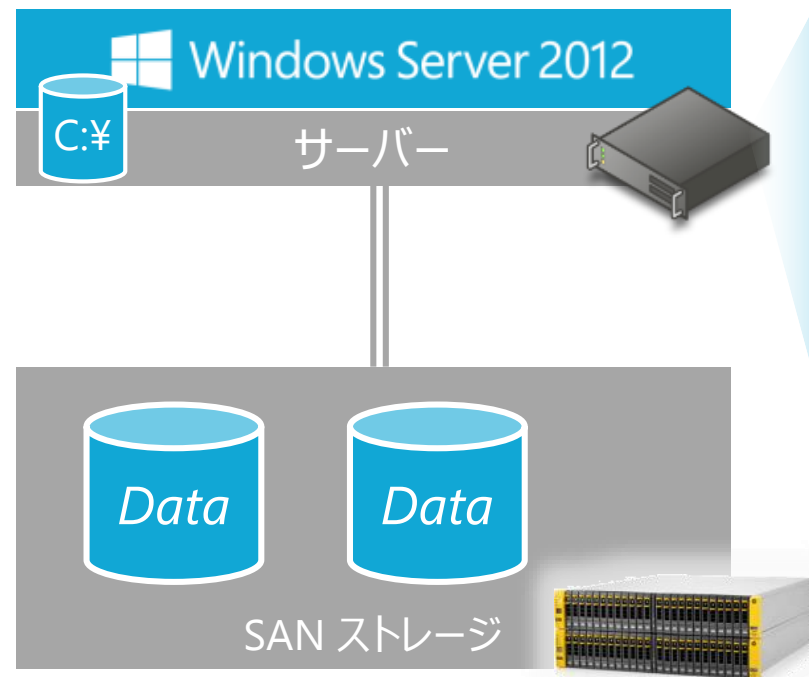
非仮想化の WS2012 物理マシンで
ODX 対応ストレージにファイルコピー

Offloaded Data Transfers (ODX)



Windows OS レベルで対応したということは、つまり ...

日常的なファイルコピーでも ODX の恩恵を受けられる！！

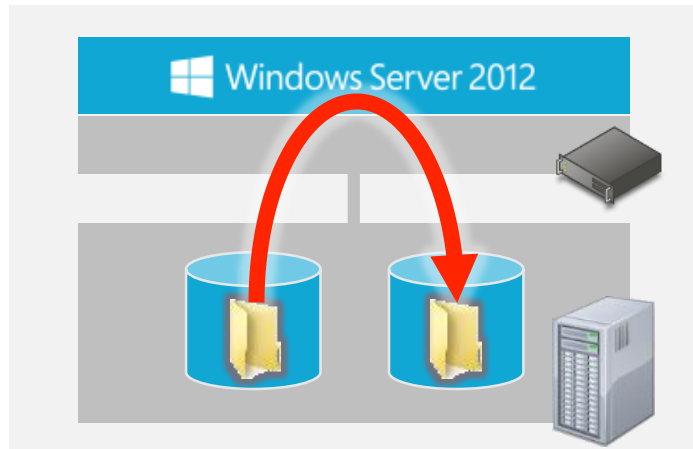


**648 MB/sec
= 5~6 Gbps**

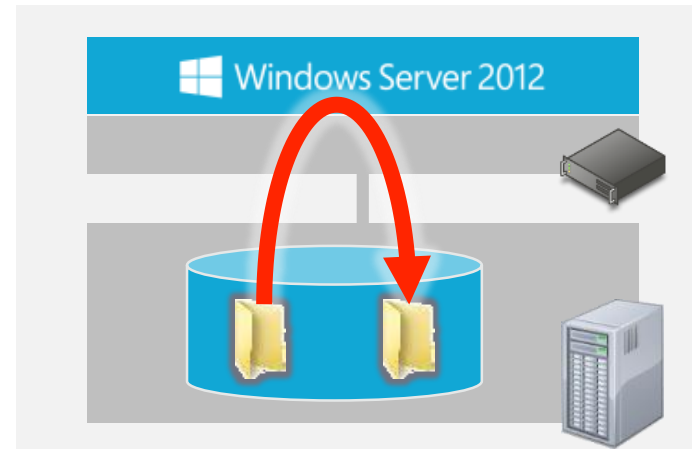
Q) どこまでが ODX でオフロードできる??



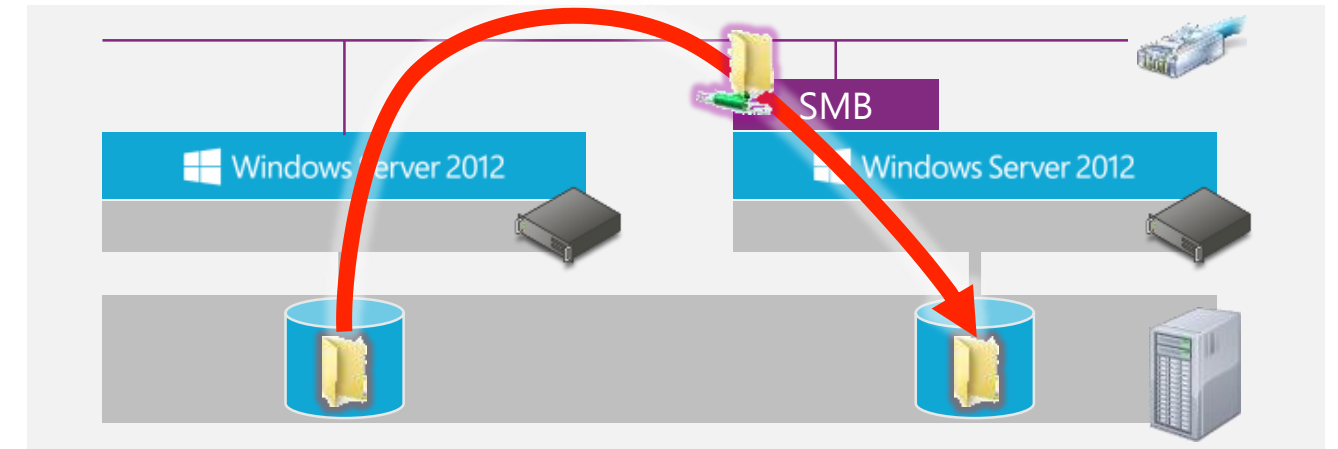
① LUN 間のコピー



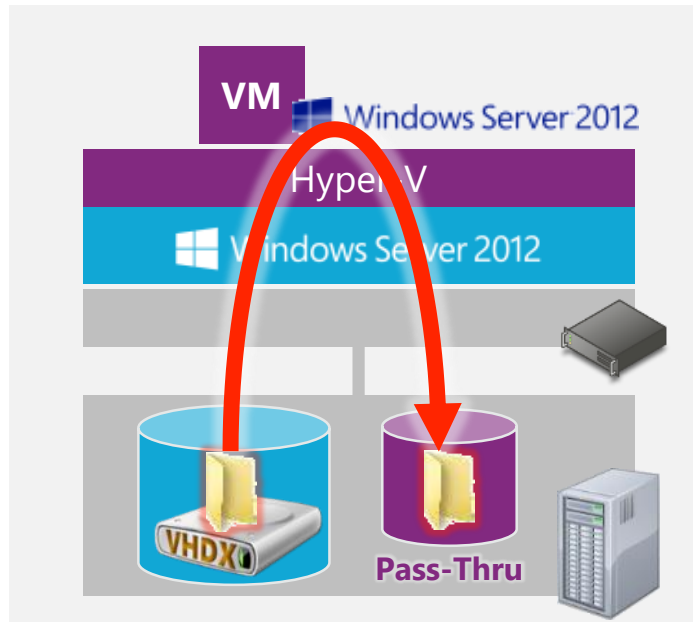
② LUN 内のコピー



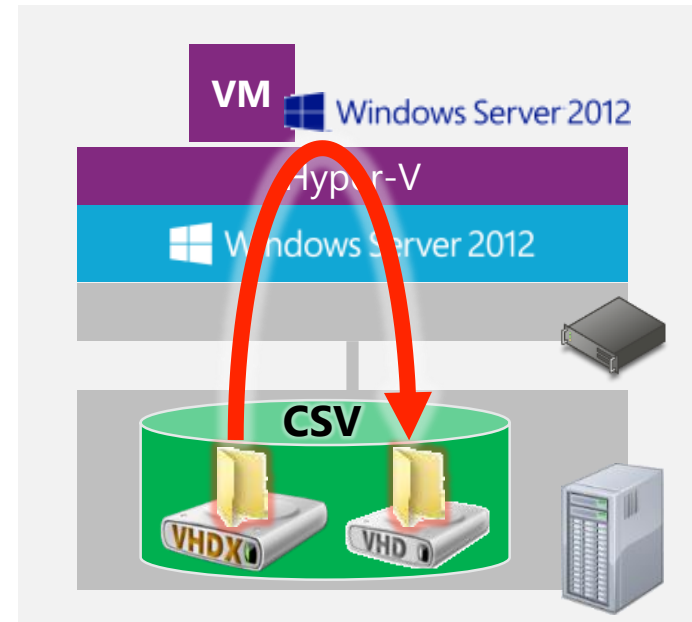
⑤ ファイルサーバーへのネットワークコピー



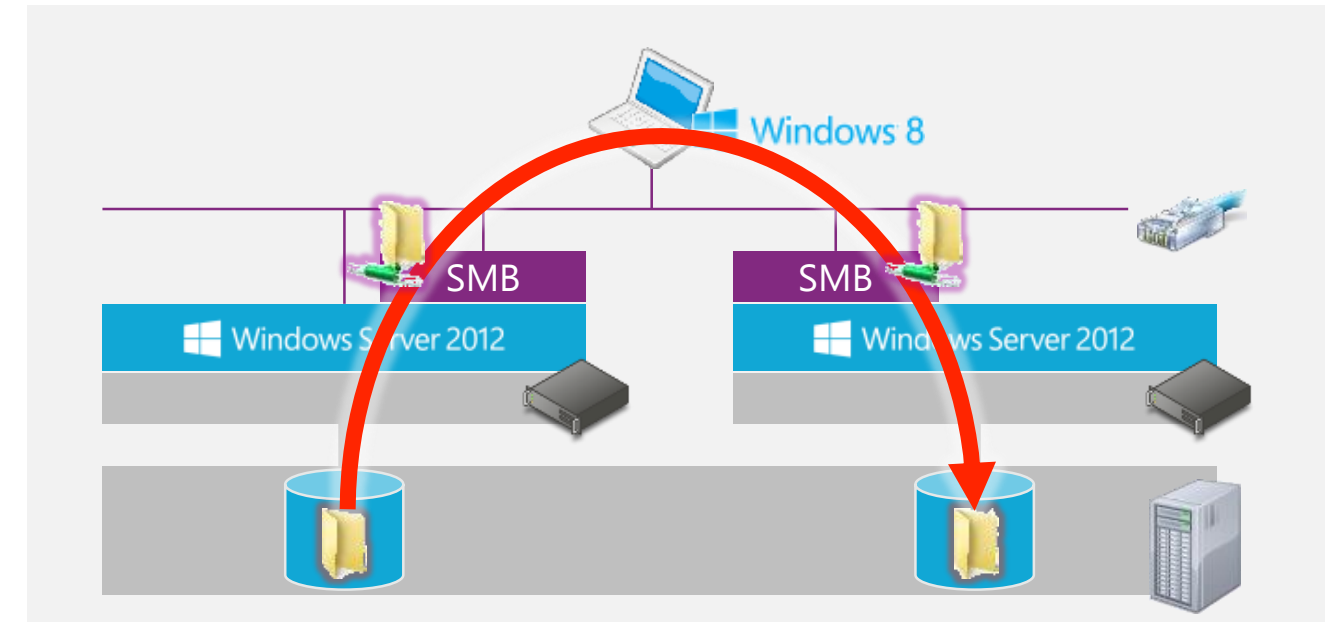
③ 仮想マシン + VHDX, Path-Thru



④ 仮想マシン + CSV + VHD



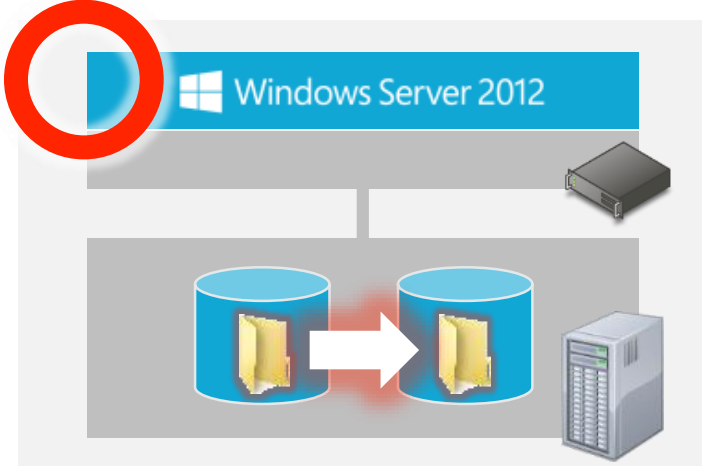
④ クライアント PC より、ファイルサーバー間のコピー



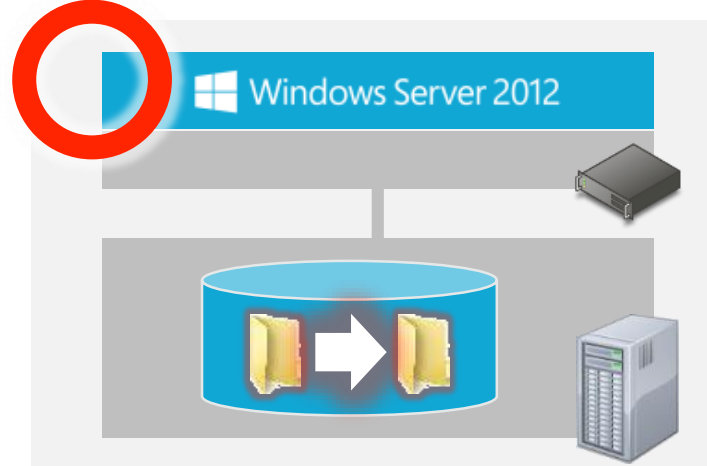
A) どのパターンも ODX オフロード対象



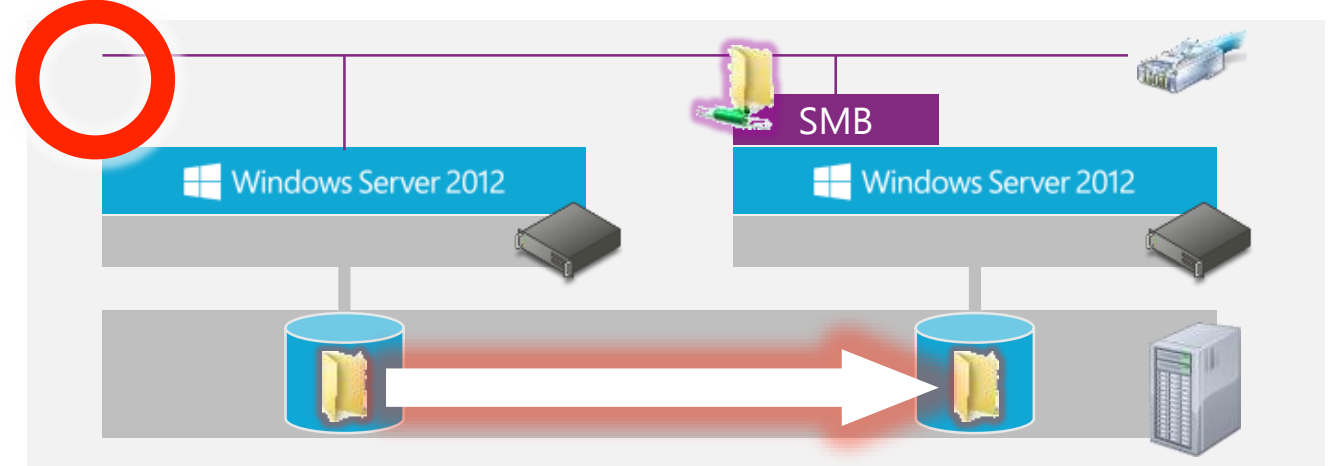
① LUN 間のコピー



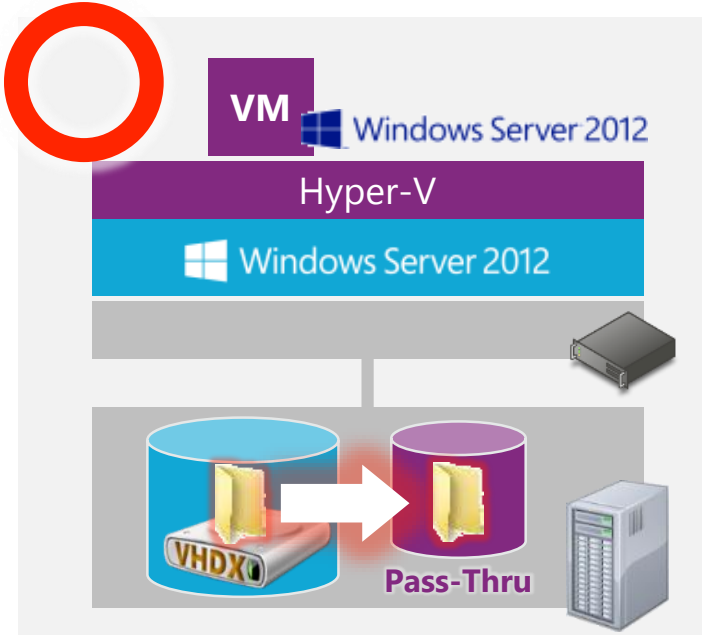
② LUN 内のコピー



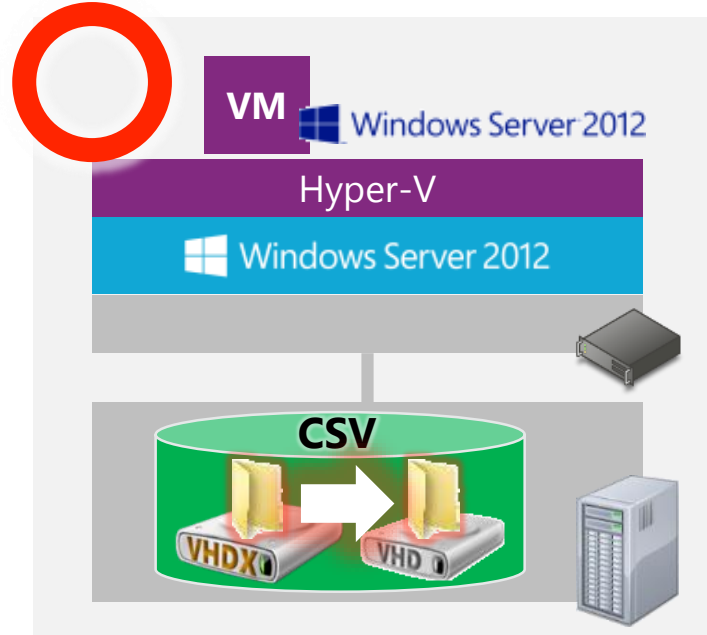
⑤ ファイルサーバーへのネットワークコピー



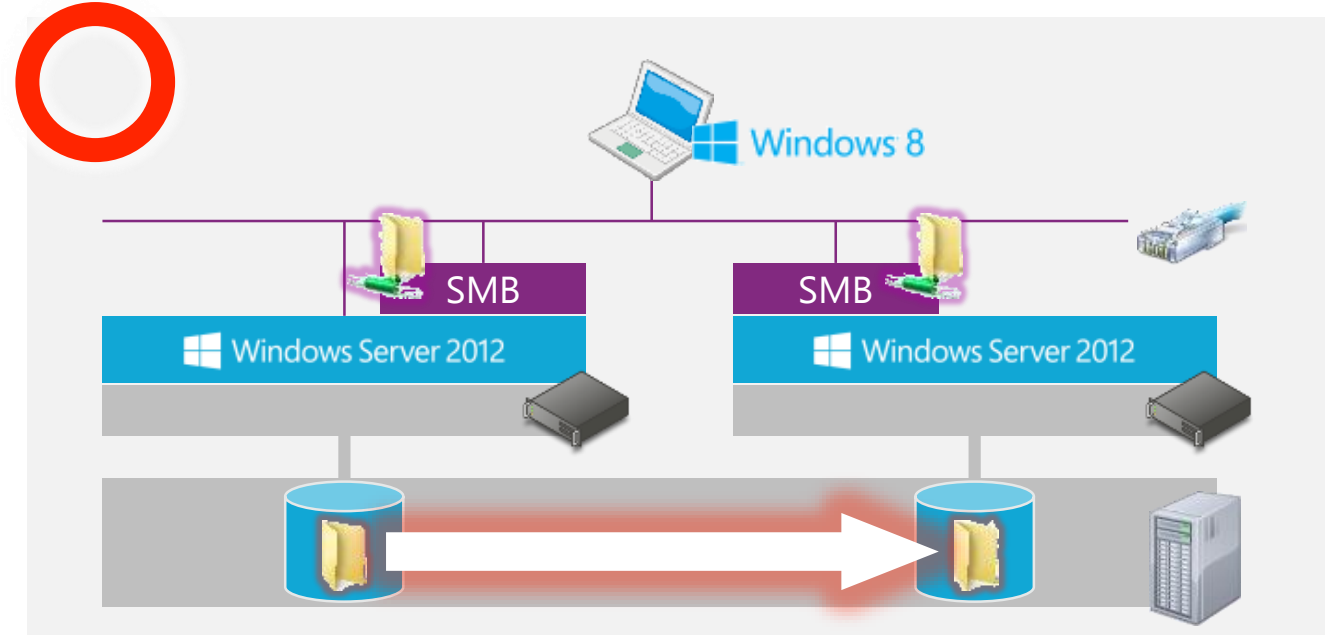
③ 仮想マシン + VHDX, Path-Thru



④ 仮想マシン + CSV + VHD



④ クライアント PC より、ファイルサーバー間のコピー



適用範囲の幅広い ODX

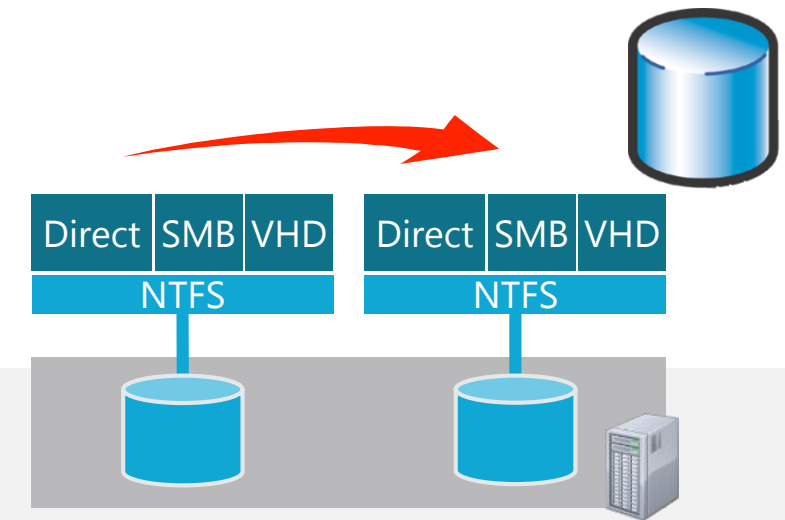
Windows OS レベルで対応したことで “夢が広がる”

ODX の要件

対象ボリューム（src, dst）

- ✓ 物理ボリューム
 - CSV ボリュームにも対応
- ✓ 仮想マシン上のボリューム^{*1}
 - 仮想 SCSI コントローラー配下の仮想ディスク (*.vhdx, *.vhd) および パススルーディスク^{*2}
 - 仮想ファイバーチャネル配下の SAN ボリューム
 - ゲスト内で iSCSI イニシエーターから直接接続した LUN
- ✓ SMB 3.0 ファイルサーバー上のボリューム

いずれも、実体は ODX 対応ストレージ上にあり
NTFS でフォーマットされていること



オフロード対象となる操作

- Windows Server 2012 or Windows 8 上で下記の API を用いたオブジェクトのコピー・移動処理
 - ✓ CopyFile, CopyFileEx, MoveFile, MoveFileEx, CopyFile2
 - ✓ 例えば ... エクスプローラー, copy/robocopy コマンド, PowerShell

ODX 適用対象外となる主な例

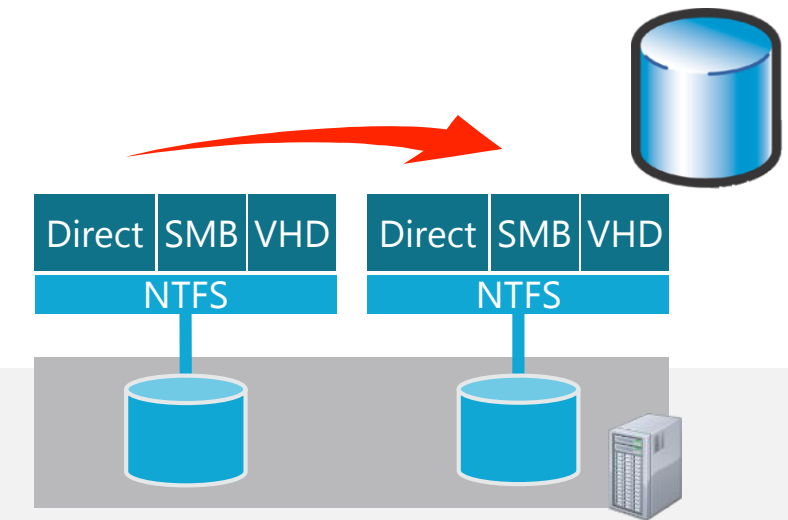
- ReFS, BitLocker, EFS, シャドウコピー有効時
- スパースファイル, 重複除去, NTFS 圧縮済みファイル
- SMB 3.0 以外のリモートコピープロトコル (FTP, WebDAV, **BITS**, ...)
- 仮想マシンの C ドライブ (IDE のため SCSI コマンドを送信できない)

^{*1} Hyper-V のホスト OS は Windows Server 2012 である必要があります

^{*2} パススルーディスクがコピー元の場合、オフロードが掛からないかもしれません

適用範囲の幅広い ODX

Windows OS レベルで対応したことで “夢が広がる”



ストレージの対応状況

現在対応済みのストレージ

Dell
EqualLogic PS
Firmware 6 ~



HP
3PAR / SS
OS 3.1.2 ~



NetApp
FAS
Data OnTap 8.2 ~



Planned

- EMC (VNX, VMAX), Fujitsu, HP (P4000), IBM, NEC

主な導入効果

ストレージ帯域が **1Gbps** クラスの場合:

- ファイルコピー時間の短縮（ファイルサーバーなどで特に効果的）
 - 速度はスピンドル状況などに依存するが、各社最大は **1GB/sec 程度** の様子
- ストレージパスの負荷改善

ストレージ帯域が **8Gbps** 以上の場合:

- サーバーの負荷の改善
 - 特に、仮想マシンでは非常に効果が高い（I/O エミュレーションから解放されるため）

名前	▼ 97% CPU
システムの割り込み	75.0%
4 エクスプローラー	20.7%
46% 完了	
名前	0% CPU
アプリ (3)	
4 エクスプローラー	0.1%
73% 完了	

注意: 逆にストレージは非常に高負荷になります。RAID 処理専用チップを搭載しているものが推奨です。

参考情報



各ハードウェアコンポーネントにおける、転送速度の理論値

NIC	Throughput
1Gb Ethernet	~ 0.1 GB/sec
10Gb Ethernet	~ 1.1 GB/sec
40Gb Ethernet	~ 4.5 GB/sec
32Gb InfiniBand (QDR)	~ 3.8 GB/sec
56Gb InfiniBand (FDR)	~ 6.5 GB/sec

Bus Slot	Throughput
PCIe Gen2 x4	~ 1.7 GB/sec
PCIe Gen2 x8	~ 3.4 GB/sec
PCIe Gen2 x16	~ 6.8 GB/sec
PCIe Gen3 x4	~ 3.3 GB/sec
PCIe Gen3 x8	~ 6.7 GB/sec
PCIe Gen3 x16	~ 13.5 GB/sec

Intel QPI	Throughput
4.8 GT/s	~ 9.8 GB/sec
5.8 GT/s	~ 12.0 GB/sec
6.4 GT/s	~ 13.0 GB/sec
7.2 GT/s	~ 14.7 GB/sec
8.0 GT/s	~ 16.4 GB/sec

HBA	Throughput
SAS 3Gb x4	~ 1.1 GB/sec
SAS 6Gb x4	~ 2.2 GB/sec
SAS 12Gb x4	~ 4.4 GB/sec
FC 4Gb	~ 0.4 GB/sec
FC 8Gb	~ 0.8 GB/sec
FC 16Gb	~ 1.5 GB/sec

Memory	Throughput
DDR2-400 (PC2-3200)	~ 3.4 GB/sec
DDR2-667 (PC2-5300)	~ 5.7 GB/sec
DDR2-1066 (PC2-8500)	~ 9.1 GB/sec
DDR3-800 (PC3-6400)	~ 6.8 GB/sec
DDR3-1333 (PC3-10600)	~ 11.4 GB/sec
DDR3-1600 (PC3-12800)	~ 13.7 GB/sec
DDR3-2133 (PC3-17000)	~ 18.3 GB/sec

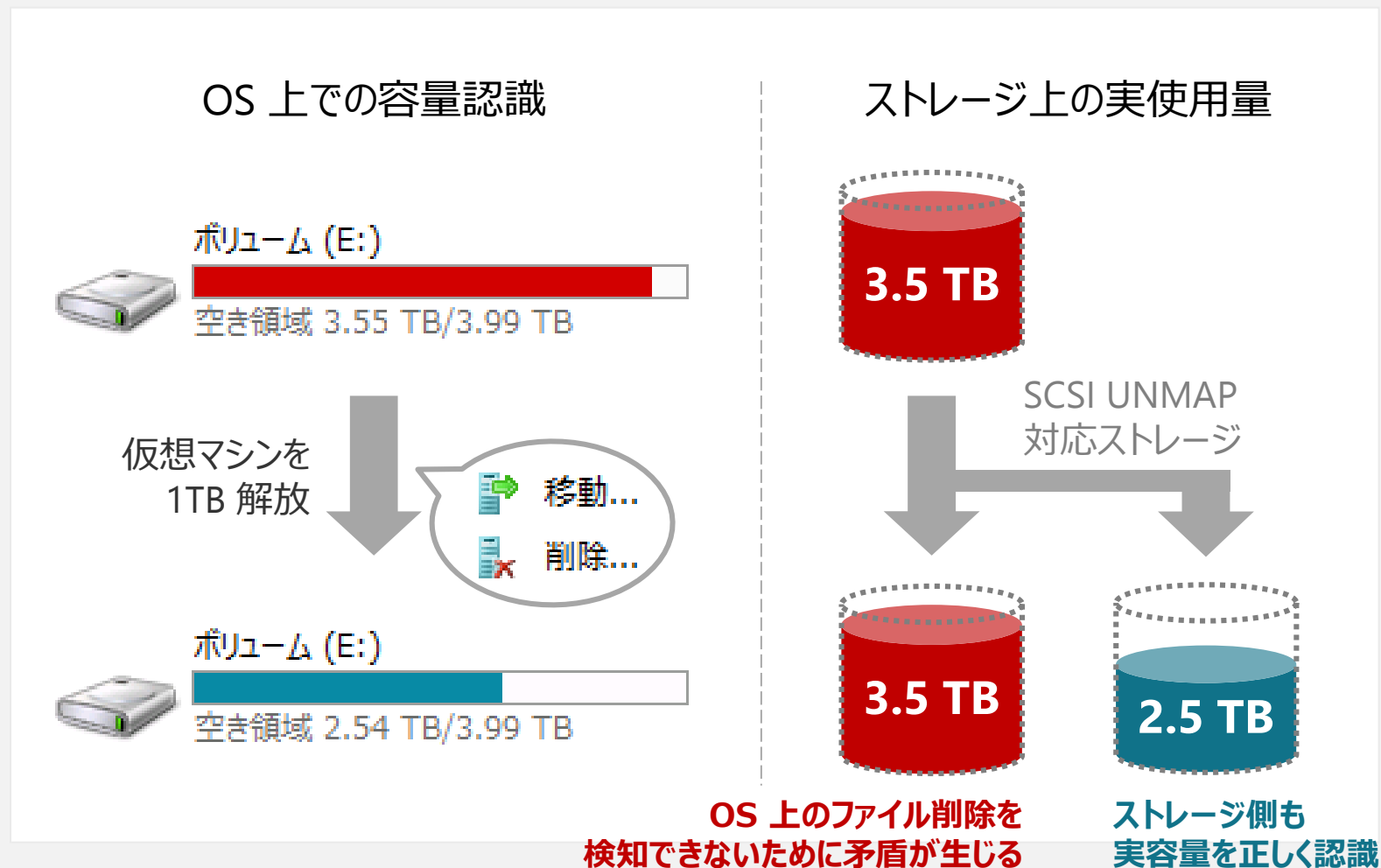
Only a few common configurations listed. Numbers are rough approximations. Actual throughput in real life will be lower than these theoretical maximums. Numbers provided are for one way traffic only (double for full duplex). One interface/port only. Numbers use base 10 (1GB/sec = 1,000,000,000 bytes per second)



T10 SCSI UNMAP

開発基盤やクラウドといった「スクラップ&ビルド」システムで重要となる、もう1つの T10 対応

VHD ファイルの削除や移動で発生した領域解放をストレージ側に伝達



特に推奨されるケース

- スモールスタートやフェーズドアプローチなど、ストレージのシンプロビジョニング機能を利用する場合
- 仮想マシンのスクラップ&ビルド (=作成&削除) を繰り返す場合

つまり、、、**「開発向け基盤」**や**「クラウド」**

利用要件

- Windows Server 2012 は既定で UNMAP を発行する
<http://technet.microsoft.com/en-us/library/jj674351.aspx>
- ストレージアレイ側が UNMAP を受信できる必要あり

Hyper-V ストレージ設計のベストプラクティス



エンタープライズ向けのベストプラクティスを 1 ページで

仮想マシンのディスク設計



- ODX を活かすため、ゲストの C ドライブはできるだけスリムに
- 新形式「VHDX」により、ゲスト上で 2TB 超えが可能
 - VMware VMDK は現在 2TB を超えられない
- 容量可変はできるだけ使わない
 - I/O 性能としては容量固定とほとんど変わらないが、ファイルが拡張されるたびに CSV 通信が発生し、I/O が遅延する
- **RAID を RAID コントローラーで処理させるのと同様に、シンプロビジョニングもハードウェア側に任せるのが一般的**
- パススルーディスク？
 - 今回も VHD, VHDX より 10~20% くらい I/O 性能が高い（実測値）
 - ODX 時も 20% 程度スループットが高いが、動作が若干怪しい？
 - ライブマイグレーションでの瞬停時間が長引くので注意

ストレージアレイ



- ODX に対応したストレージ装置を推奨
 - シンプロビジョニングを使う場合は SCSI UNMAP 対応をチェック
 - HA できる接続方式
 - FibreChannel,
 - Shared SAS
 - iSCSI *
 - SMB 3.0 *
- * Hyper-V ホストから 10Gbps 以上を敷く場合
iSCSI: ソフトウェアイニシエータは避けて CNA に任せる
SMB: Infiniband HCA を利用する

その他 Tips

- Hyper-V では SCSI Reservation による LUN の I/O Lock は起きない
- SMB ストレージを使う場合、ドメインコントローラーの配置を考慮

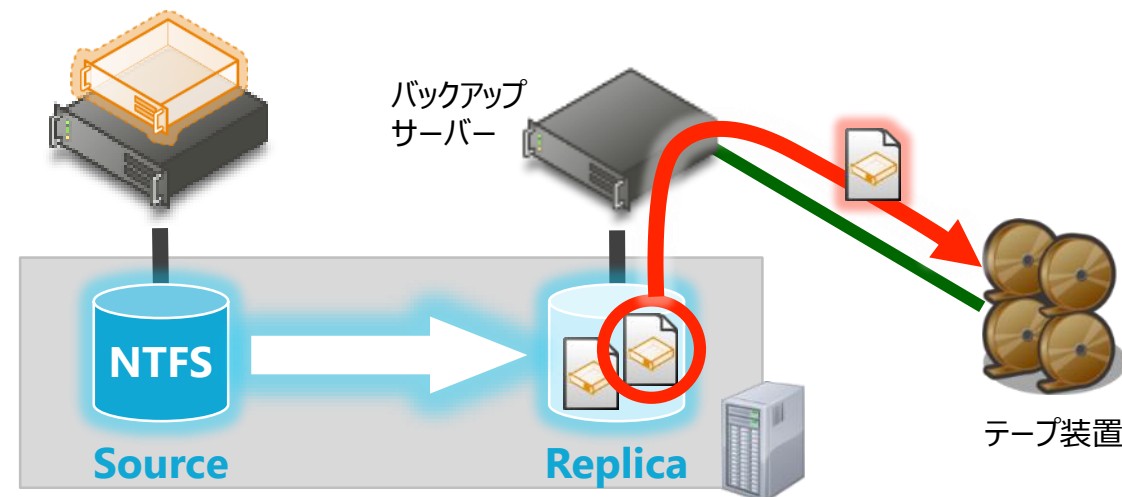
バックアップ



Hyper-V はアレイベースバックアップとの親和性が高い

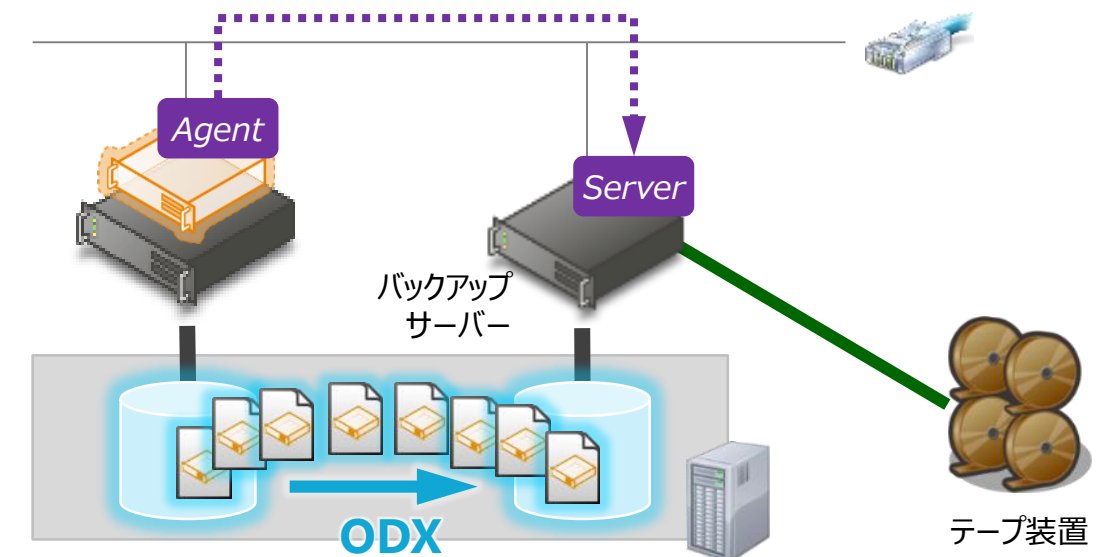
Hyper-V はアレイベースのテープバックアップが簡単！

- バックアップ容量によっては、夜間にバックアップを終わらせるために転送速度に優れた「アレイベースのバックアップ」が必要
- Hyper-V はファイルシステムが **NTFS** なので、バックアップサーバーはレプリカの中身を読み取れる（RAW バックアップ不要）
- 仮想ファイバーチャネルを使えば、バックアップサーバーの仮想化も



近い将来、バックアップ設計は大きく変わる

- バックアップソフトが **ODX** に対応すると、LAN 経由の Agent バックアップでも、転送速度はアレイベースになる
- Agent バックアップのアプリ整合性を保ちつつ、アレイベースの速度・サーバー負荷でバックアップ可能に



まとめ

- Windows Server 2012 Hyper-V は、エンタープライズへの本格採用に耐えうるハイパーバイザーです。
- 今回の Hyper-V は高負荷時の性能低下を防ぐためにハードウェアの最新テクノロジーを積極的に採用しています。せっかくの新機能をフルに活かすためにも、サーバーだけでなく、ネットワークやストレージについても選定に気を遣いましょう。
- インフラ設計を終えたら、次は「管理者がラクできる運用設計」です。System Center などの自動化製品をうまく活用して負担を軽減してください！



Thank You!