



シスコ ネットワーク環境における VMware Infrastructure 3

May 28, 2008

【注意】 シスコ製品をご使用になる前に、安全上の注意
(www.cisco.com/jp/go/safety_warning/) をご確認ください。

本書は、米国シスコシステムズ発行ドキュメントの参考和訳です。
米国サイト掲載ドキュメントとの差異が生じる場合があるため、正式な内容については米国サイトのドキュメントを参照ください。
また、契約等の記述については、弊社販売パートナー、または、弊社担当者にご確認ください。

このマニュアルのすべての設計、仕様、告示、情報、および推奨事項（集成的に「デザイン」）は、障害を含めて「現状のまま」として提供されます。シスコシステムズおよびその代理店は、商品性や特定の目的への準拠性、権利を侵害しないことに関する、または取り扱い、使用、または取引によって発生する、一切の保証の責任を負わないものとします。いかなる場合においても、シスコシステムズおよびその代理店は、このデザインの使用またはこのデザインを使用できないことによって起こる制約、利益の損失、データの損傷など間接的で偶発的に起こる特殊な損害のあらゆる可能性がシスコシステムズまたは代理店に知らされていても、それらに対する責任を一切負いかねます。

デザインは予告なしに変更されることがあります。このデザインの使用は、すべてユーザ側の責任になります。これらのデザインは、シスコ、その代理店、またはパートナーの技術的またはその他の専門的アドバイスではありません。これらのデザインを実装する前に、ユーザ企業の技術アドバイザーに相談する必要があります。結果はシスコでテストしていない要因によって異なる場合があります。

CCDE, CCVP, Cisco Eos, Cisco StadiumVision, the Cisco logo, DCE, and Welcome to the Human Network are trademarks; Changing the Way We Work, Live, Play, and Learn is a service mark; and Access Registrar, Aironet, AsyncOS, Bringing the Meeting To You, Catalyst, CCDA, CCDP, CCIE, CCIP, CCNA, CCNP, CCSP, Cisco, the Cisco Certified Internetwork Expert logo, Cisco IOS, Cisco Press, Cisco Systems, Cisco Systems Capital, the Cisco Systems logo, Cisco Unity, Collaboration Without Limitation, Enterprise/Solver, EtherChannel, EtherFast, EtherSwitch, Event Center, Fast Step, Follow Me Browsing, FormShare, GigaDrive, HomeLink, Internet Quotient, IOS, iPhone, IP/TV, iQ Expertise, the iQ logo, iQ Net Readiness Scorecard, iQuick Study, IronPort, the IronPort logo, LightStream, Linksys, MediaTone, MeetingPlace, MGX, Networkers, Networking Academy, Network Registrar, PCNow, PIX, PowerPanels, ProConnect, ScriptShare, SenderBase, SMARTnet, Spectrum Expert, StackWise, The Fastest Way to Increase Your Internet Quotient, TransPath, WebEx, and the WebEx logo are registered trademarks of Cisco Systems, Inc. and/or its affiliates in the United States and certain other countries.

All other trademarks mentioned in this document or Website are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (0801R)

シスコ ネットワーク環境における VMware Infrastructure 3
Copyright © 2008 Cisco Systems, Inc. All rights reserved.

Copyright © 2008, シスコシステムズ合同会社.
All rights reserved.



CONTENTS

はじめに	1
概要	1
ESX Server ネットワークおよびストレージの接続	2
ESX Server のネットワーキング コンポーネント	2
vmnic、vNIC、およびバーチャル ポート	3
ESX バーチャル スイッチ	5
バーチャル スイッチの概要	5
ポート グループ	8
レイヤ 2 セキュリティ 機能	10
管理	10
vSwitch のスケーラビリティ	11
vSwitch で無効な設定	11
ESX LAN ネットワーキング	12
vSwitch の転送特性	12
VLAN タギング	15
NIC チーミングによる接続の冗長化	18
vSwitch の設定	23
ESX 内部のネットワーキング	30
ESX Server Storage ネットワーキング	33
VMware ESX Server のストレージ コンポーネント	35
ファイル システムのフォーマット	37
マルチパス化およびパス フェールオーバー	40
ESX Server の接続およびネットワーキング設計に関する考慮事項	43
LAN 接続	43
予備設計の考慮事項	44
2 つの NIC を使用する ESX ホスト	50
4 つの NIC を使用する ESX ホスト	57
SAN 接続	64
FibreChannel の実装に関する考慮事項	65
NPIV	66
パフォーマンスの考慮事項	71
iSCSI の実装に関する考慮事項	73
VMotion ネットワーキング	75
同一サブネット上での VMotion 移行（フラット ネットワーク）	77
ESX HA クラスタ	79
その他のリソース	82



シスコ ネットワーク環境における VMware Infrastructure 3

はじめに

このマニュアルは、シスコおよび VMware によって共同制作されています。このマニュアルでは、シスコ ネットワーク環境で VMware Infrastructure (VI) 3.x および VMware ESX Server 3.x を使用するための推奨ベスト プラクティスについて説明します。また、ESX Server の内部構造に関する詳細、さらに ESX Server と外部のシスコ ネットワーク デバイスの関係について説明します。

このマニュアルは、シスコ データセンター環境における VMware ESX Server 3.x ホストを理解して配置する必要のあるネットワーク設計者、ネットワーク技術者、およびサーバ管理者が対象です。

概要

現在、企業のデータセンターを構成するハードウェア プラットフォームおよびソフトウェア プラットフォームを統合して標準化する作業が進められています。IT グループはデータセンター ファシリティ、そこに収容されたサーバ、およびネットワーク コンポーネントを、個々の業務要件を解決するための「サイロ」に格納された関連性のない資産としてではなく、リソースのプールと考えています。サーバ バーチャライゼーションは、サーバリソースの抽象化によって柔軟性を確保し、標準化されたインフラストラクチャ上でそれらを最適に利用するための技法です。これにより、データセンター アプリケーションは特定のハードウェア リソースの制約がなくなるので、アプリケーション側で基盤ハードウェアを意識しなくてすむようになります。さらに、CPU、メモリ、およびネットワーク インフラストラクチャを、サーバで使用可能な共有リソース プールとして認識できるようになります。

ネットワーク、ストレージ、およびサーバプラットフォームのバーチャライゼーション技術は成熟しつつあります。VLAN (バーチャル LAN)、VSAN (バーチャル ストレージエリア ネットワーク)、バーチャル ネットワーク デバイスなどのテクノロジーは今日の企業データセンターで広く使用されています。メインフレームのレガシー システムは、すでに長年にわたって「仮想化」されており、論理パーティション (logical partition; LPAR) の採用によってリソースの使用率が向上しています。

サーバ バーチャライゼーションを利用すると、オペレーティング システムから物理ハードウェア (CPU、メモリ、ディスクなど) 間のリンクを切り離すことができるので、物理レベルを超えて統合を図り、リソースの利用とアプリケーションのパフォーマンスを最適化する新しい可能性が生まれます。この革新を推し進めた結果、仮想環境をサポートする、これまで以上に強力な x86 プラットフォームが登場しました。具体的には、マルチコア CPU が使用可能になり、AMD Virtualization (AMD-V) および Intel Virtualization Technology (IVT) が利用できるようになりました。



(注)

このテクノロジーをサポートする AMD プロセッサの詳細については、次の URL を参照してください。
http://www.amd.com/us-en/Processors/ProductInformation/0,,30_118_8796,00.html

このテクノロジーをサポートする Intel プロセッサの詳細については、次の URL を参照してください。
http://www.intel.com/business/technologies/virtualization.htm?iid=servproc+rhc_virtualization

VMware インフラストラクチャは、高度な企業ネットワークと統合するためのネットワーキング機能を豊富に提供します。これらのネットワーキング機能は、VMware ESX Server により提供され、VMware VirtualCenter によって管理されます。バーチャル ネットワーキングでは、バーチャルマシンと物理マシンの両方を一貫性のある方法でネットワークに組み込むことができます。さらに、単一の ESX Server ホスト内、または複数の ESX Server ホストにまたがって複雑なネットワークを構築できます。このネットワーク上でバーチャルスイッチを活用すれば、物理スイッチ上で使用すると同じネットワークプロトコルを使用して、同一 ESX Server ホスト上のバーチャルマシン同士を通信させることができます。ネットワーキングハードウェアを追加する必要はありません。ESX Server バーチャルスイッチは、標準 VLAN と互換性がある VLAN であれば、他のベンダー製のものもサポートします。

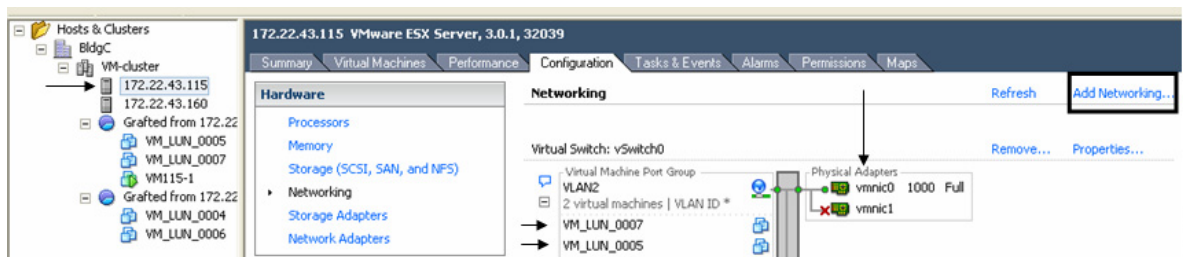
バーチャルマシンは、専用の IP アドレスと MAC アドレスが割り当てられた 1 つまたは複数のバーチャルイーサネットアダプタで構成できます。このため、バーチャルマシンには、物理マシンとの間に一貫性のあるネットワーキングプロパティが与えられます。

ESX Server ネットワークおよびストレージの接続

VMware ネットワーキングは ESX ホストごとに定義し、バーチャルインフラストラクチャの実装全体を管理するためのツールである VMware VirtualCenter Management Server を使用して設定します。ESX Server ホストは、複数のバーチャルマシン (VM) を実行できます。また、物理 LAN スイッチングネットワークにトラフィックを送出する前に、ホストのバーチャルネットワーク内部のスイッチングを実行できます。

ESX Server のネットワーキング コンポーネント

図 1 ESX ホストごとに定義する VMware ネットワーキング



vmnic、vNIC、およびバーチャルポート

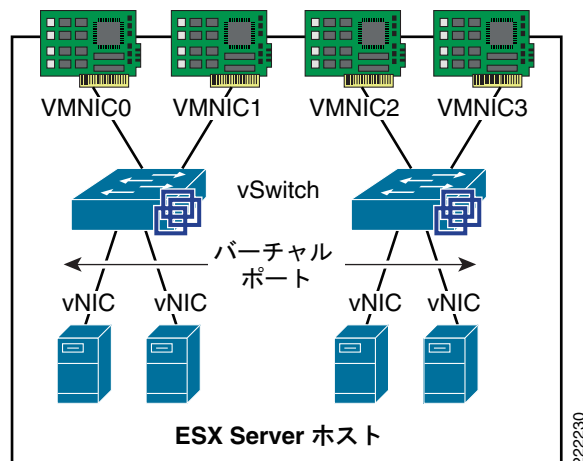
VMware 仮想化環境において、「NIC」には 2 つの意味があります。ホストサーバハードウェアの物理ネットワークアダプタ (vmnic) を表すこともあれば、VMware のハードウェア抽象化レイヤが仮想マシンに提供する仮想ハードウェアデバイスとしての仮想 NIC (vNIC) を表すこともあります。vNIC は単なる仮想デバイスですが、物理 NIC が提供するハードウェアアクセラレーション機能を活用できます。

VirtualCenter を使用し、必要な ESX ホストを選択すると (インターフェイスの左側、[図 1](#) を参照)、ネットワーキングの構成を表示できます。**Configuration** タブ (インターフェイスの右側) で、VM の vNIC ([図 1](#) の VM_LUN_0007 および VM_LUN_0005) と物理 NIC (vmnic0 および vmnic1) 間のアソシエーションを確認できます。仮想 NIC と物理 NIC は仮想スイッチ (vSwitch) によって接続されます。vSwitch は vNIC と vmnic 間でトラフィックを転送します。vNIC と vSwitch 間の接続ポイントを仮想ポートといいます。

Add Networking ボタンをクリックすると、*Add Network Wizard* が開くので、ウィザードに従って新しい vSwitch を作成したり、または既存の vSwitch を分割するための機能である **ポートグループ** を新しく作成したりできます。

[図 2](#) に、ESX ホストにおける物理アダプタと VM アダプタのプロビジョニングを示します。

図 2 ESX Server のインターフェイス



[図 2](#) では、物理ホスト上に 4 つの vmnic があります。サーバ管理者は VM トラフィックを伝送する vmnic を指定できます。この ESX Server は 2 つの vSwitch で構成されています。4 つの VM があり、それぞれに vNIC が 1 つずつあります。vNIC は一方の vSwitch の仮想ポートに接続されています。

vNIC の MAC アドレス、ブートアップ、VMotion 移行

VM は最大 4 つの vNIC で構成できます。vNIC の MAC アドレスは、ESX Server が自動的に生成しますが（プロセスについては、次の項で説明）、管理者が指定することもできます。この機能は、DHCP ベースのサーバアドレッシングを使用する環境で VM を使用する場合に便利です。指定 MAC アドレスを使用することによって、VM に必ず同じ IP アドレスが与えられるようにできるからです。



(注) 標準の NIC と異なり、vNIC の「チーミング」は通常、不要であり、有用とはいえません。VMware 環境において、NIC チーミング (NIC Teaming) は複数の *vmnic* を vSwitch に接続し、ネットワークのロードシェアリングまたは冗長性を実現することを意味します。

vNIC の MAC アドレスには、IEEE が VMware に割り当てた OUI (組織固有識別子) が含まれます。vNIC の MAC アドレスは、ESX ホストおよびコンフィギュレーションファイル名の情報を使用して作成されます。VMware が使用する OUI は 00-50-56 および 00-0c-29 です。MAC アドレスの生成に使用するアルゴリズムによって、MAC アドレスが衝突する可能性は少なくなります。プロセスで MAC アドレスの一意性を保証することはできません。生成 MAC アドレスは、次の 3 つの部分を使用して作成されます。

- VMware の OUI
- 物理 ESX Server マシンの SMBIOS UUID
- MAC アドレスの生成対象となるエンティティの名前に基づくハッシュ

ESX ホストは VM 間における MAC の衝突を検出し、必要に応じて解消できます。スタティックに割り当てる VM MAC アドレス用として、VMware は範囲 00:50:56:00:00:00 → 00:50:56:3F:FF:FF を予約しています。管理者が VM にスタティック MAC アドレスを割り当てる場合は、この範囲内のアドレスを使用する必要があります。

VM ごとに固有の「.vmx」ファイルがあります。これは VM コンフィギュレーション情報のファイルです。ダイナミックに生成された MAC アドレスは、このファイルに保存されます。このファイルを削除すると、VM の MAC アドレスが変更される可能性があります。ファイルのロケーション情報がアドレス生成アルゴリズムに組み込まれているからです。



(注) VMotion は、ESX Server ファーム内において、電源の入った VM を物理 ESX ホスト間で移行させる場合に ESX Server が使用する方式です。VMotion 移行によって VM MAC が変更されることはありません。VMotion 移行によって、ある ESX ホストから別の ESX ホストに VM を移動させても、VM の MAC アドレスは変わりません。VMware Virtual Machine File System (VMFS) ボリュームは SAN 上にあり、起点 ESX ホストとターゲット ESX ホストの両方にアクセスできるからです。したがって、.vmx コンフィギュレーションファイルおよび VM ディスクを別の場所にコピーする必要はありません。このようなコピーを行うと、新しい MAC が生成される可能性があります。



(注) ただし、これは電源がオフである VM を移行させる場合は（非 VMotion）、必ずしも当てはまりません。この場合、VM のロケーションを決定することもできますが、VM の MAC アドレスが変更される可能性があります。

ESX バーチャル スイッチ

ESX ホストは、カーネルのコンテキストで動作するソフトウェア バーチャル スイッチを使用して、ローカル VM を相互に、または外部企業ネットワークにリンクさせます。

バーチャル スイッチの概要

バーチャル スイッチは VMware Infrastructure 3 の重要なネットワークング コンポーネントです。各 ESX Server 3 ホスト上で最大 248 のバーチャル スイッチを同時に作成できます。バーチャル スイッチは小さい機能ユニットの集合から、実行時に「受注生産」されます。

主要な機能ユニットは、次のとおりです。

- コア レイヤ フォワーディング エンジン — このエンジンはパフォーマンスと正確性の両面で、システムの重要な部分です。バーチャル インフラストラクチャでは簡素化され、レイヤ 2 イーサネット ヘッダーを処理するだけです。物理イーサネット アダプタにおける相違、バーチャルイーサネット アダプタにおけるエミュレーションの相違といった、他の実装の詳細からは完全に切り離されています。
- VLAN タギング、ストリッピング、およびフィルタリング ユニット
- レイヤ 2 セキュリティ、チェックサム、およびセグメンテーション オフロード ユニット

バーチャル スイッチを実行時に作成する場合、ESX Server は必要なコンポーネントだけをロードします。構成に使用される特定の物理およびバーチャル イーサネット アダプタ タイプをサポートするために、実際に必要なものだけをインストールして実行します。したがって、複雑さとシステムパフォーマンスに対する要求に関して、システムのコストは最小限で済みます。

ESX Server は設計上、その場で特定のコンポーネントを一時的にロードできます。この機能を使用すると、適切な設計の診断ユーティリティを実行したりできます。モジュラ設計のもう 1 つの利点として、VMware およびサードパーティの開発者は将来、システムを拡張するためにモジュールを容易に組み込むことができます。

ESX Server バーチャル スイッチはさまざまな点で、物理スイッチに類似しています。また、明らかな相違点もいくつかあります。このような類似性と相違性を理解しておく、バーチャル ネットワークの構成、バーチャル ネットワークと物理ネットワークの接続をプランニングするとき役立ちます。

バーチャル スイッチと物理スイッチの類似点

ESX Server 3 で実装されたバーチャル スイッチは、最近のイーサネット スイッチとほとんど同様に動作します。バーチャル スイッチは MAC アドレスおよびポート フォワーディング テーブルを維持し、次の機能を実行します。

- 着信した各フレームの宛先 MAC を調べます。
- 1 つまたは複数のポートにフレームを転送して送信できるようにします。
- 不必要な配信を回避します (ハブではないということです)。

ESX Server 3 バーチャル スイッチは、ポート レベルで VLAN セグメンテーションをサポートします。したがって、次のどちらかに各ポートを設定できます。

- シングル VLAN アクセス。物理スイッチにおける、またはバーチャル スイッチ タギングを使用する ESX Server 用語でのアクセス ポートです。
- マルチ VLAN アクセス。物理スイッチにおける、またはバーチャル ゲスト タギングを使用する ESX Server 用語でのトランク ポートです。

さらに、管理者は Virtual Infrastructure Client を使用し、スイッチ全体および個々のポートに対応する多数の設定オプションを管理できます。

バーチャルスイッチと物理スイッチの相違点

ESX Server は正規 MAC フィルタ アップデートなどの設定情報得るために、バーチャルイーサネットアダプタからのダイレクトチャンネルを提供します。したがって、ユニキャストアドレスを学習したり、IGMP スヌーピングを実行してマルチキャストグループメンバシップを学習したりする必要はありません。

バーチャルスイッチでの STP の不使用

VMware インフラストラクチャでは、ESX Server 内で強制的に単一階層ネットワークトポロジを使用します。言い換えると、複数のバーチャルスイッチを相互接続できないので、ESX ネットワークはループが発生する構成にはできません。したがって、ESX ホスト上の vSwitch は STP (スパンニングツリープロトコル) を実行しません。



(注)

実際には、ある種の設定を行えば、バーチャルスイッチでループを発生させることは可能です。しかし、それには、ゲスト上で2つのバーチャルイーサネットアダプタを同じサブネットに接続し、レイヤ2ブリッジングソフトウェアを実行する必要があります。偶発的にこのようなシステム構成が行われることはなく、一般的な構成ではわざわざそうする理由もありません。

バーチャルスイッチの分離

同じホスト内で、あるバーチャルスイッチから別のバーチャルスイッチにネットワークトラフィックを直接流すことはできません。バーチャルスイッチは、1つのスイッチに必要なすべてのポートを提供し、次の利点をもたらします。

- バーチャルスイッチのカスケードが不要なので、バーチャルインフラストラクチャはバーチャルスイッチの接続機能を提供しません。
- バーチャルスイッチの接続手段がないので、バーチャルスイッチの接続不良を防止する必要がありません。
- バーチャルスイッチは物理イーサネットアダプタを共有できないので、イーサネットアダプタをループバックさせるなど、バーチャルスイッチ間でリークが発生するような構成にはなりません。

さらに、バーチャルスイッチごとに専用のフォワーディングテーブルが与えられます。あるテーブルのエントリが別のバーチャルスイッチ上のポートを示すというメカニズムはありません。したがって、他のバーチャルスイッチのルックアップテーブルにスイッチのアドレスのエントリが含まれている場合でも、スイッチが検索する各宛先が一致するのはフレームの起点ポートと同じバーチャルスイッチ上のポートに限られます。

この分離には制限があります。2つのバーチャルスイッチのアップリンクを同時に接続する場合、またはバーチャルマシンで動作しているソフトウェアを使用して2つのバーチャルスイッチをブリッジする場合です。

アップリンク ポート

アップリンク ポートは物理アダプタと関連付けられたポートであり、バーチャル ネットワークと物理ネットワーク間を接続します。物理アダプタがアップリンク ポートに接続するのは、デバイスドライバによって物理アダプタが初期化されたとき、またはバーチャル スイッチのチーミング ポリシーが設定変更されたときです。一部のバーチャル スイッチは物理ネットワークに接続しないので、アップリンク ポートがありません。これが当てはまるのは、ファイアウォールバーチャルマシンとファイアウォールで保護されたバーチャル マシン間を接続するバーチャル スイッチの場合などです。

バーチャル イーサネット アダプタがバーチャル ポートに接続するのは、アダプタが設定されたバーチャル マシンの電源投入時またはレジューム時、デバイスを接続する明示的なアクションを実行したとき、または VMotion を使用してバーチャル マシンを移行させたときです。バーチャル イーサネット アダプタは、初期化時および MAC フィルタリング情報が変更されるたびに、MAC フィルタリング情報でバーチャル スイッチをアップデートします。バーチャル ポートは、バーチャル イーサネット アダプタから要求があっても、そのポートで有効なレイヤ 2 セキュリティ ポリシーに違反する場合は、その要求を無視することがあります。たとえば、MAC スプーフィングがブロックされている場合、この規則に違反するパケットはポートで廃棄されます。

バーチャル スイッチの正確性

正確性に関しては、2つの問題が特に重要です。バーチャル マシンまたはネットワーク上のその他のノードによってバーチャル スイッチの動作が左右されないようにすることが重要です。ESX Server は次の方法で、このような影響を受けないようにします。

- バーチャル スイッチは、それぞれのフォワーディング テーブルに入力する目的で、ネットワークから MAC アドレスを学習することはありません。したがって、(直接的な DoS 攻撃 [サービス拒絶攻撃] の場合やワームやウイルスが脆弱なホストをスキャンした場合のような) 他の攻撃の副作用として DoS 攻撃または漏洩攻撃が引き起こされる可能性がなくなります。
- バーチャル スイッチは、フォワーディングまたはフィルタリングの決定に使用するフレームデータのプライベート コピーを作成します。これはバーチャル スイッチの重要な機能であるとともに、バーチャル スイッチ固有の機能です。効率面の問題から、バーチャル スイッチはフレーム全体をコピーするわけではありません。ただし、フレームがバーチャル スイッチに到達した場合はゲスト オペレーティング システムから機密データにアクセスできないように ESX Server を設定する必要があります。

ESX Server は、フレームがバーチャル スイッチ上の適切な VLAN 内に格納されるようにします。その方法は次のとおりです。

- VLAN データは、バーチャル スイッチを通過する際に、フレームの外部で伝送されます。フィルタリングは単純な整数比較です。これは、ユーザがアクセスできるデータを信用しないというシステムの一般原則から見ると、特例といえます。
- バーチャル スイッチはダイナミック トランッキングをサポートしません。

VMware インフラストラクチャの VLAN

VLAN は、ステーションまたはスイッチ ポートを論理グループに分割することで、ステーションまたはポートが同じ物理 LAN セグメントにある場合と同様に通信できるようにします。ブロードキャスト トラフィックをスイッチ ポートまたはエンドユーザのサブセットに限定すると、かなりのネットワーク帯域幅とプロセッサ時間を節約できます。

VMware インフラストラクチャ ユーザに対して VLAN をサポートするには、バーチャル ネットワークまたは物理ネットワーク上の要素の 1 つで、イーサネット フレームに 802.1Q タグを付ける必要があります。バーチャル マシン フレームのパケットにタグを付ける (またはタグを外す) 設定モードは 3 種類あります。

- バーチャル スイッチ タギング (VST モード) — これが最も一般的な設定です。このモードでは、バーチャル マシンのバーチャル アダプタをバーチャル スイッチに直接接続するのではなく、バーチャル スイッチの VLAN ごとにポート グループを1つずつプロビジョニングすることによりポート グループとの接続を行います。バーチャル スイッチのポート グループは、すべての発信フレームにタグを付け、すべての着信フレームからタグを外します。さらに、ある VLAN のフレームが別の VLAN に漏れないようにします。このモードを使用するには、物理スイッチでトランクを提供する必要があります。
- バーチャル マシン ゲスト タギング (VGT モード) — バーチャル マシン内部に 802.1Q VLAN トランキング ドライバをインストールできます。バーチャル スイッチとの間でフレームを受け渡すときに、バーチャル マシン ネットワーキング スタックと外部スイッチの間でタグが維持されます。このモードを使用するには、物理スイッチでトランクを提供する必要があります。
- 外部スイッチ タギング (EST モード) — VLAN タギングに外部スイッチを使用できます。これは物理ネットワークの場合と同様であり、VLAN 設定は通常、個々の物理サーバに対して透過的です。この環境では、トランクを提供する必要はありません。

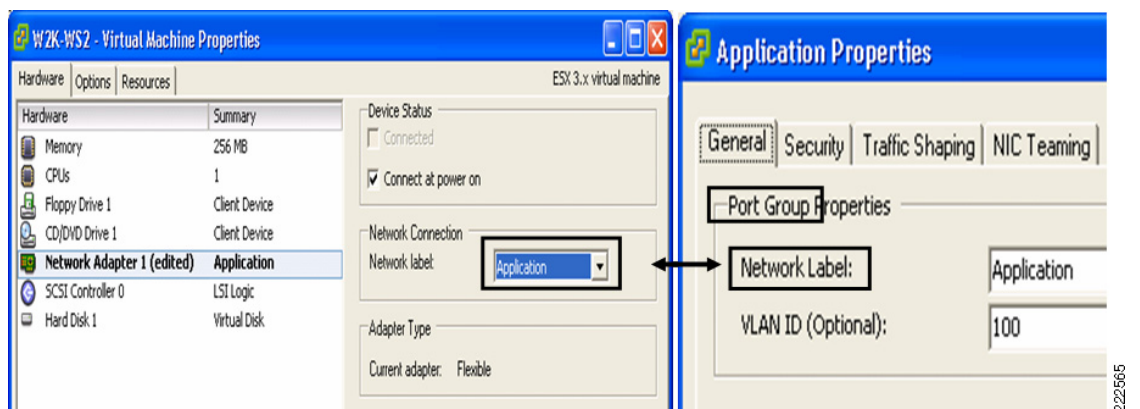
ポート グループ

バーチャル マシンは vNIC を介して vSwitch に接続します。ESX Server 上のネットワーキング設定で、vNIC (別名、VM ネットワーク アダプタ) とネットワーク ラベルを関連付け、それによってポート グループを識別します。したがって、VM と vSwitch を関連付けるには、ポート グループに vNIC を割り当てる必要があります。

VLAN への VM の割り当て

図 3 に、ポート グループと VLAN の関係を示します。図 3 の左側に VM Network Adapter の設定があり、Network Connection の設定値で使用可能なポート グループの 1 つを参照しています。右側に、そのポート グループに対応するバーチャル スイッチ プロパティがあり、ネットワーク ラベルと VLAN が関連付けられています。

図 3 ポート グループと VLAN の関係



VM ネットワーク アダプタ (vNIC) とポート グループ (すなわち、ネットワーク ラベル) の関連付けでは、次の作業を行います。

- vSwitch に vNIC を割り当てます。
- 特定の VLAN に vNIC を割り当てます。
- 特定の「NIC チーミング」ポリシーに vNIC を割り当てます。これは vNIC チーミングではなく、また、従来の意味の NIC チーミングでもないので、詳細は後述します。VM からのトラフィックは、ポート グループの「NIC チーミング」設定で定義されたトラフィック ロード バランシング ポリシーに基づいて、vSwitch 経由で LAN スイッチング ネットワークに伝送されるという方がより適切でしょう。

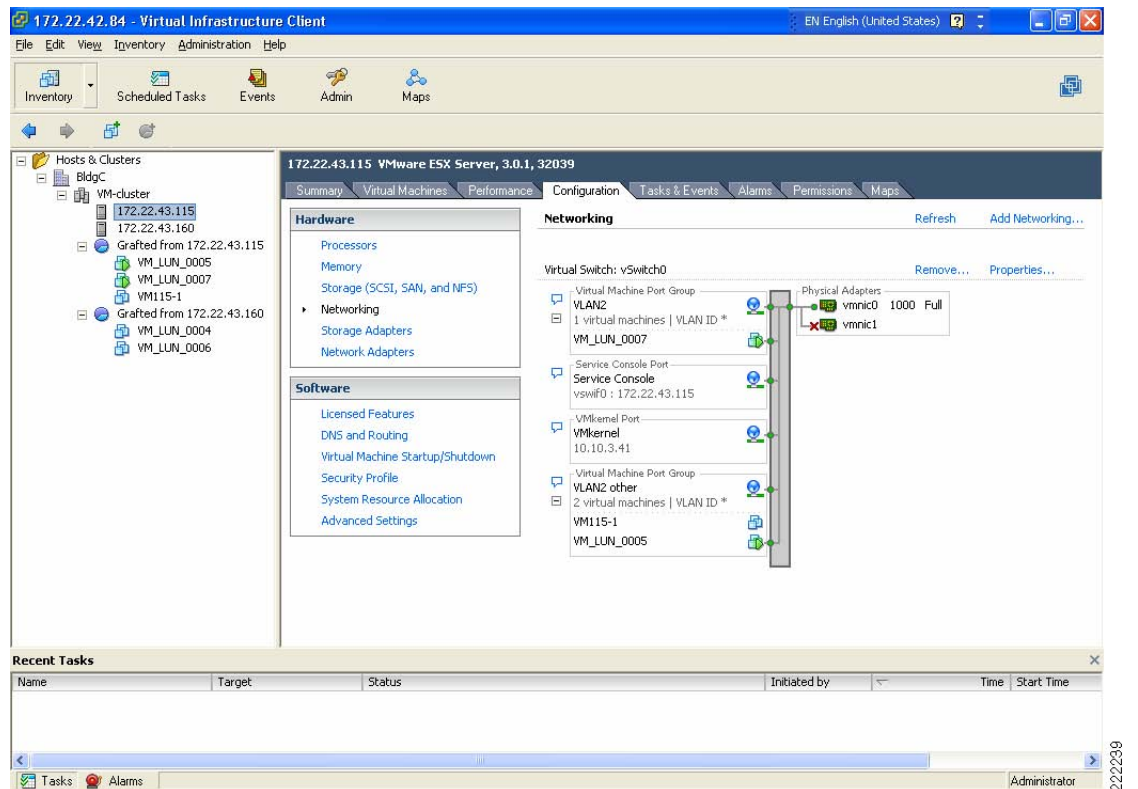
ポート グループは VLAN ではない

ポート グループは vSwitch 上の vNIC ポート用コンフィギュレーションテンプレートです。ポート グループを使用すると、管理者は複数の VM の vNIC をグループとしてまとめ、同時に設定できます。管理者はポート グループ設定を変更することによって、特定の QoS、セキュリティ ポリシー、および VLAN を設定できます。

ポート グループで VLAN に vNIC (つまり VM) を割り当てた場合でも、ポート グループと VLAN 間に 1 対 1 のマッピングは存在しません。実際は、任意の数の異なるポート グループに同じ VLAN を使用させることができます。

図 4 に例を示します。VM_LUN_0007 および VM_LUN_0005 は 2 つの異なるポート グループ上にあり、前者を **VLAN2**、後者を **VLAN2 other** といいます。どちらのポート グループも VLAN2 を使用し、事実 VM_LUN_0007 は VM_LUN_0005 と通信できます。構成として考えた場合、ポート グループはこのようなスイッチ ポートの分離によりスイッチ ポートを分割するのではなく、単にスイッチ ポートをグループ化することによって分割します。

図 4 vSwitch およびポート グループ



要約

ネットワーキングの専門家でも、ポート グループの概念を誤解していることがあります。ポート グループに関して、把握しておくべき重要な概念を要約すると、次のようになります。

- ポート グループは構成管理メカニズムです。
- ポート グループは VLAN ではありません。
- ポート グループはポート チャンネルではありません。
- VM と vSwitch 間のアソシエーションは、vNIC (VM ネットワーク アダプタ) 設定画面からポート グループ (ネットワーク ラベルという) を選択することによって定義します。
- ポート グループでは、所属する vNIC ポートに対して次の設定パラメータを定義します。VLAN 番号、レイヤ 2 セキュリティ ポリシー、QoS (Quality of Service)、および NIC チーミングと呼ばれるトラフィック ロード バランシング ポリシーです。

レイヤ 2 セキュリティ機能

バーチャル スイッチはセキュリティ ポリシーを適用して、バーチャル マシンがネットワーク上の他のノードを偽装しないようにします。この機能には 3 種類のコンポーネントがあります。

- すべてのバーチャル マシンに関して、**Promiscuous mode** (無差別モード) はデフォルトでディセーブルです。したがって、ネットワーク上の他のノードへのユニキャスト トラフィックを見ることはできません。
- MAC アドレス変更ロックダウンによって、バーチャル マシンはそれぞれのユニキャスト アドレスを変更できなくなります。さらに、ネットワーク上の他のノードへのユニキャスト トラフィックを見ることはできません。そのため、**Promiscuous mode** (無差別モード) より狭いとはいえ、潜在的なセキュリティの脆弱性を保護できます。
- 偽造送信ブロック機能をイネーブルにすると、バーチャル マシンは自身のネットワーク ノードから着信したトラフィックを送信できなくなります。

管理

ESX Server を管理する手段は、3 種類あります。

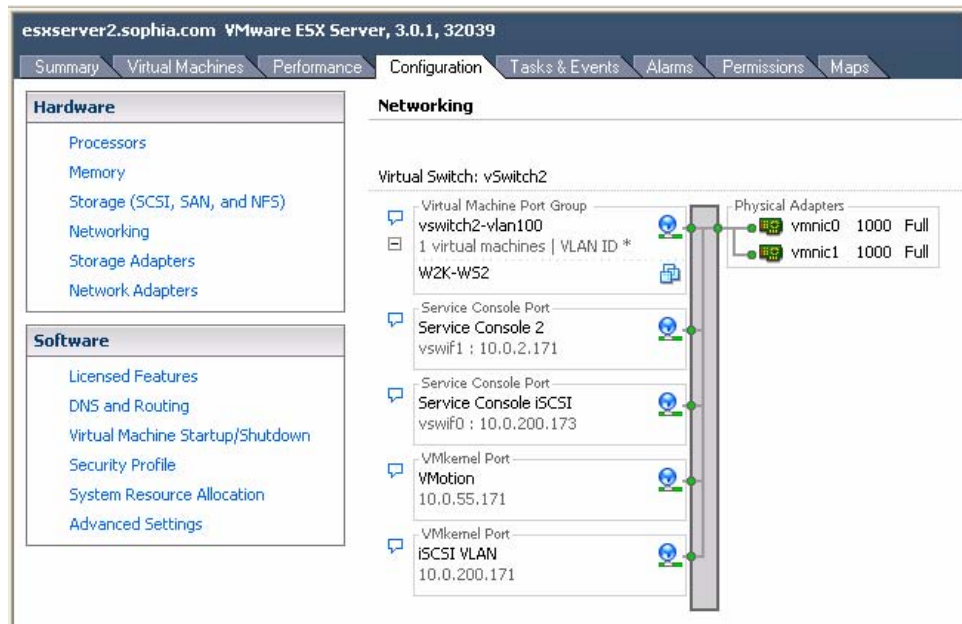
- Service Console
- Web ベース ユーザ インターフェイス
- VMware VirtualCenter などの管理アプリケーション

ESX Server の Service Console には、SSH (セキュア シェル)、Telnet、HTTP (ハイパーテキスト転送プロトコル)、および FTP (ファイル転送プロトコル) を使用してアクセスできます。Service Console はさらに、認証およびシステム モニタリング サービスを提供します。組み込みバージョンの ESX Server である ESX 3i には、ユーザがアクセスできる Service Console はありませんが、管理用の Web インターフェイスをサポートします。ESX ホストを 1 つだけ管理するには、Service Console および Web ベース ユーザ インターフェイスで十分です。

VMware VirtualCenter は、VC プラットフォームに応じて拡張され、多数のクライアント、ESX ホスト、および VM をサポートする中央管理ソリューションです。VirtualCenter は、バーチャル ネットワーク インフラストラクチャを構築して維持するための各種ツールを提供します。VirtualCenter を使用すると、バーチャル スイッチを追加、削除、変更し、VLAN およびチーミングを指定してポート ブループを設定できます。

VMware ESX Server/Configuration/Networking タブから設定例を表示できます。図 5 を参照してください。この例では、W2K-WS2 という VM が VLAN 100 上の vSwitch2 に接続します (図 5 では、省略されて VLAN * になっています)。

図 5 vSwitch の最終的な設定



vSwitch の右側にある **Properties** ボタンを選択すると、vSwitch の特性をさらに変更できます。ポートグループの追加、NIC Teaming プロパティの変更、トラフィック レート制限の設定などが可能です。

VirtualCenter のロール機能を使用すると、バーチャルネットワークの管理に必要な権限をネットワーク管理者に割り当てることができます。詳細については、<http://www.vmware.com/vmtn/resources/826> にアクセスし、『*Managing VMware VirtualCenter Roles and Permissions*』を参照してください。

vSwitch のスケーラビリティ

ESX Server に複数の vSwitch を組み込み、そのそれぞれに最大 1016 の「内部」バーチャルポートを VM 用に設定できます。vSwitch に割り当てられた vNIC ごとにバーチャルポートを 1 つずつ使用するので、理論上、各 vSwitch の最大 VM 数は 1016 です。バーチャルスイッチは発信 vmnic アダプタを介して企業ネットワークに接続します。バーチャルスイッチが外部接続に使用できる vmnic は最大 32 です。

vSwitch で無効な設定

vSwitch ではある種の設定が認められません。

- vSwitch 相互間の直接接続はできません。つまり、vSwitch に接続できるのは vNIC および vmnic だけです。2 つの vNIC を備えた VM を使用し、Microsoft Windows などのブリッジ機能を活用して、ある vSwitch から別の vSwitch にトラフィックを流すことは可能です。しかし、レイヤ 2 ループが生じるリスクがあるので、この手法は避けるべきです。
- vmnic および vmnic に関連付けられた物理 NIC を複数の vSwitch に割り当てることはできません。
- vSwitch を LAN スwitチング ネットワークのトランジットパスにすべきではありません。実際、トランジットパスにはできません（詳細については、「[vSwitch の転送特性](#)」[p.12] を参照）。

ESX LAN ネットワーキング

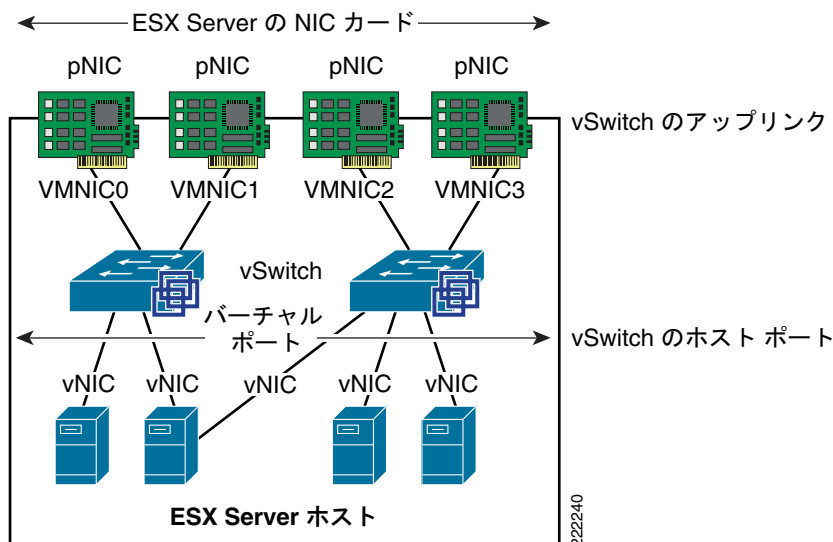
バーチャルスイッチでは、物理サーバ上で1つ以上の vmnic を使用して、外部ネットワークに VM を接続します。VMkernel を使用すると、次の機能をはじめ、物理 NIC 上で使用できるハードウェア アクセラレーション機能を vSwitch ソフトウェア上で使用できるようになります。

- TCP セグメンテーション オフロード
- VLAN タギング
- チェックサム計算オフロード

vSwitch の転送特性

vSwitch は標準のレイヤ 2 イーサネット スイッチと同様に動作します。vSwitch は VM 間、VM と LAN スイッチング インフラストラクチャ間でトラフィックを転送します。ESX Server の vmnic は vSwitch のアップリンクです。図 6 を参照してください。

図 6 vSwitch のコンポーネント



vSwitch が標準のイーサネット スイッチと類似している点は、次のとおりです。

- 転送は MAC アドレスに基づいて行われます。
- 同じ vSwitch および VLAN 内の VM 間トラフィックは、ローカルなままです。
- vSwitch は VLAN ID を使用してトラフィックにタグ付けできます。
- vSwitch はトランッキングに対応できます (ネゴシエーションプロトコルを使用しない 802.1Q トランク)。
- vSwitch は一部の QoS 機能 (レート制限) を実行できます。
- vSwitch は一部のレイヤ 2 セキュリティ機能を実装します。
- vSwitch はポート チャンネルを確立できます (ネゴシエーションプロトコルは使用しない)。

vSwitch が標準のイーサネット スイッチと異なる点は、次のとおりです。

- vSwitch のフォワーディング テーブルは、VM と vSwitch 間の通知メカニズムによって設定されます。vSwitch はネットワークから MAC アドレスを学習しません。
- vSwitch は STP を実行しません。また、STP を必要としません。アップリンクで受信したトラフィックが別のアップリンクに転送されることがないからです。
- vSwitch は IGMP スヌーピングを実行しませんが、vSwitch ですべての vNIC の関連マルチキャストを認識するので、マルチキャストトラフィックのフラッドは行われません。
- vSwitch のポート ミラーリング機能は、SPAN 機能のサブセットです。



(注)

VMware 情報については、http://www.vmware.com/files/pdf/virtual_networking_concepts.pdf を参照してください。

vSwitch のフォワーディング テーブル

vSwitch はレイヤ 2 フォワーディング テーブルを使用して、宛先 MAC アドレスに基づいてトラフィックを転送します。vSwitch のフォワーディング テーブルには、VM の MAC アドレスおよび VM に関連付けられているバーチャル ポートが指定されます。フレームの宛先が VM の場合、vSwitch はその VM にフレームを直接送信します。宛先 MAC アドレスが VM に存在しない場合、あるいはマルチキャストまたはブロードキャストの場合、vSwitch は vmnic (すなわちサーバ NIC ポート) にトラフィックを送信します。

複数の vmnic (物理 NIC) が存在する場合、マルチキャストおよびブロードキャストトラフィックが同じ VLAN 内のすべての vmnic にフラッドされることはありません。これは NIC チーミングが設定されていてもいなくても同じです。アクティブ / スタンバイ構成の場合は自明ですが、アクティブ / アクティブ構成でも同じです。任意の時点で、各 VM がトラフィック転送に使用する vmnic は 1 つだけだからです。

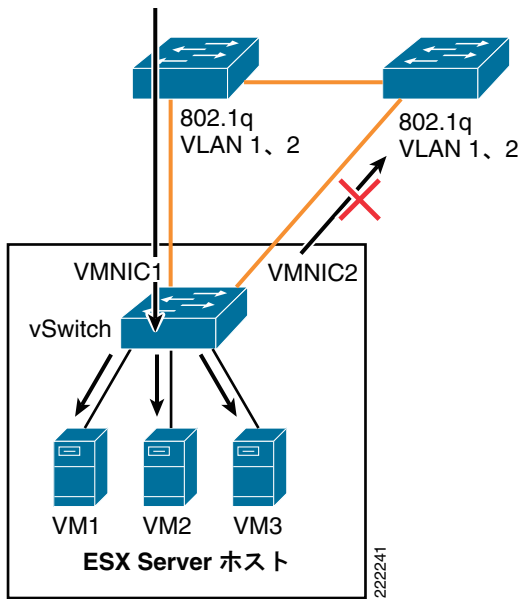
要約すると、標準イーサネット スイッチは、ポートで認識したトラフィックに基づいてフォワーディング テーブルを学習します。一方、vSwitch のフォワーディング テーブルに含まれるのは、VM の MAC アドレスだけであり、VM エントリと一致しないものは、ブロードキャストおよびマルチキャストトラフィックを含めてすべて、サーバの NIC カードに転送されます。

vSwitch のループ防止

冗長 vSwitch アップリンクの設定については、「NIC チーミングによる接続の冗長化」(p.18) で扱います。ここでは、vSwitch では STP 以外のループ防止メカニズムを実装するので、STP を実行しない、また、必要としないということを理解してください。これらのループ防止メカニズムには、戻りフレームの可能性がある発信トラフィックを廃棄するディスタンス ベクトルや、ある NIC (アップリンク) に着信したフレームは、別の NIC カードを通じて ESX Server の外部に送信されることはない (つまり、ブロードキャストなどとは異なる) というディスタンス ベクトルが含まれます。

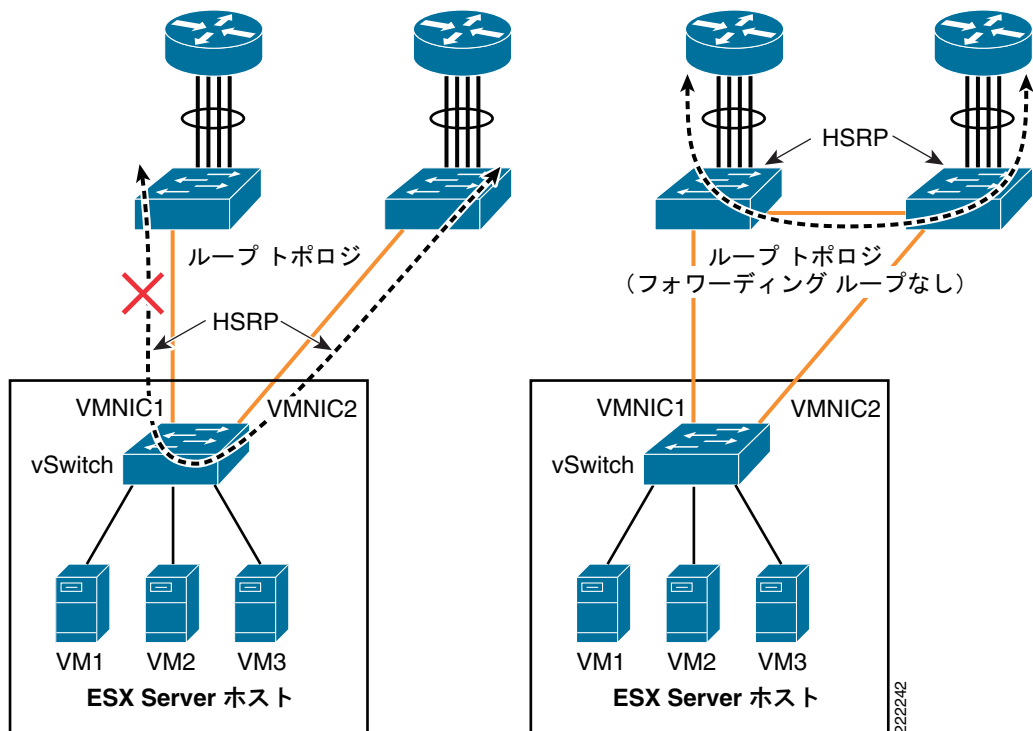
例を使用して動作を説明します。図 7 に典型的な「ループ」設計を示します。つまり、レイヤ 2 ループが存在するので、STP なしでは機能しないトポロジです。図 7 には、ESX NIC1 にブロードキャストが着信した場合の vSwitch の動作を示します。vSwitch でスパニング ツリーを実行していた場合、NIC2 がブロッキング ステートになり、NIC2 からブロードキャストを送信できなくなります。標準スイッチの場合と異なり、vSwitch はスパニング ツリーを実行しませんが、やはり NIC2 からブロードキャストが転送されることはありません。これは望ましい状況です。

図 7 vSwitch およびブロードキャスト



次に、ループフリー トポロジ (図 8 の左側) について検討します。これは、固有のループが存在しないトポロジです。このトポロジの場合、ループ防止メカニズムとして以外、スパニング ツリーは必要ありません。このトポロジでは、冗長ゲートウェイを確保するために、アップストリーム ルータで HSRP hello を交換する必要があります。vSwitch が標準イーサネット スイッチだった場合、HSRP hello によって、2 つのルータのうち的一方がゲートウェイ機能のアクティブ側になり、他方がスタンバイ側になることで合意が得られます。HSRP アドバタイズメントは vSwitch を通過しなければならないので、この場合、2 つのルータは収束できません。どちらも、自分が HSRP ゲートウェイ機能のアクティブ側であると認識します。vSwitch が HSRP データグラムを渡さないことがその理由です。図 8 の右側に、この固有の問題を解決するトポロジの例 (必ずしも最適な設計というわけではありません) を示します。

図 8 vSwitch および HSRP トラフィック



(注)

これらの例は、推奨設計ではありません。標準イーサネットスイッチとの比較において、vSwitchの転送およびループ防止特性を説明するために使用しているだけです。推奨設計については、「ESX Server ネットワークおよびストレージの接続」(p.2) を参照してください。

VLAN タギング

データセンターの物理アクセススイッチは、VLAN タギング機能を提供し、1つのネットワークインフラストラクチャで複数のVLANをサポートできるようにします。データセンターにESX Serverを導入すると、従来のVLAN タギング方式が唯一のオプションではなくなります。

vSwitchはVLAN タギングをサポートします。VMからのトラフィックをVLAN タグを付けずにそのまま、ESX Server NICに接続されたアップストリームのシスコスイッチに渡すように、vSwitchを設定できます。VMwareではこの方式をExternal Switch Tagging (EST; 外部スイッチ タギング) といいます。ESX Server NICに接続されたアップストリームのシスコスイッチにトラフィックを渡すときに、VM Guest OSから割り当てられたVLAN タグを維持するようにvSwitchを設定することもできます。VMwareではこの方式をVirtual Guest Tagging (VGT; バーチャルゲスト タギング) といいます。

最も一般的であり、なおかつ優先すべきオプションは、VLAN タグでVMからのトラフィックを色分けし、ESX Server NICに接続されたシスコスイッチに対して802.1q トランクを確立するようにvSwitchを設定することです。VMwareではこの方式をVirtual Switch Tagging (VST; バーチャルスイッチ タギング) といいます。



(注)

VMware 情報については、http://www.vmware.com/pdf/esx3_vlan_wp.pdf を参照してください。

ここでは、各方式 (EST、VTG、および VST) の長所と短所について説明します。

EST

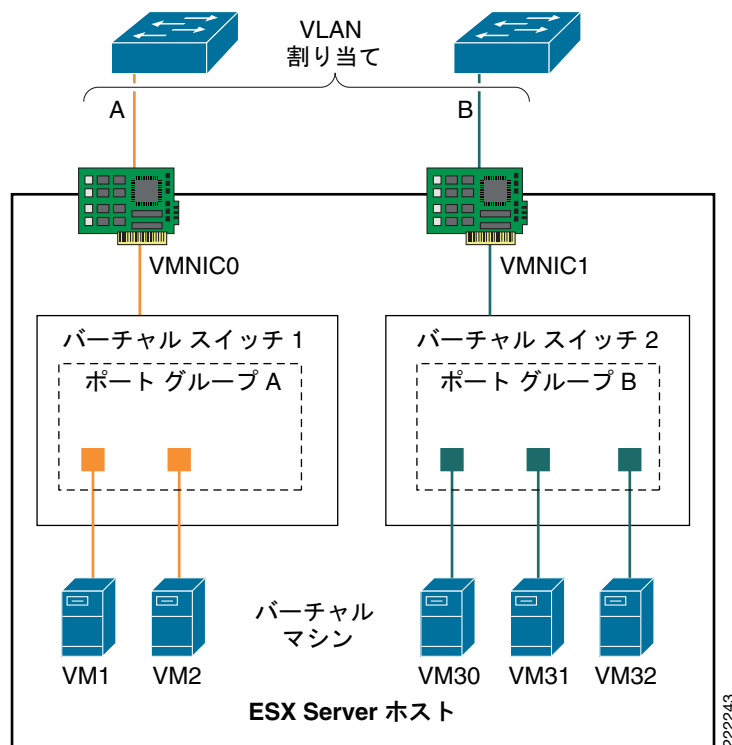
EST では、ESX ホストのアクセス ポートで VLAN タギングを定義します。VMware の設定に関しては、ポート グループ設定の VLAN ID フィールドで VLAN 0 を指定するか、または VLAN ID を空のままにしておくことにより、EST を実行できます。

予想に反するかもしれませんが、vNIC と vmnic の間に 1 対 1 のマッピングはありません。vSwitch では引き続き、ローカル スイッチングが行われます。VM からのトラフィックを Cisco Catalyst スイッチに流すときに、vSwitch は 802.1q VLAN ラベルをプリペンドしません。

図 9 では、バーチャル スイッチごとに 1 つの VLAN が関連付けられています。VLAN A および B です。外部ネットワークでは、バーチャル スイッチへの vmnic リンクを、ポートごとに 1 つの VLAN をサポートするアクセス ポートとして定義します。vSwitch は VLAN タグ機能を実行しません。VM1 から VM2 へのトラフィックは、vmnic0 から送出されず、バーチャル スイッチ 1 上でスイッチングされます。VM1 (または VM2) のトラフィックが VM2 (または VM1) 宛てではない場合は、vmnic0 から外部へ送られ、Cisco Catalyst スイッチによって VLAN A に割り当てられます。

VM30、VM31、VM32 間のトラフィックも同様に、vmnic1 から送出されず、vSwitch2 上でスイッチングされます。VM30、VM31、または VM32 からバーチャル スイッチ 2 上の VM 以外を宛先とするトラフィックが送信された場合、そのトラフィックは vmnic1 から外部へ送られ、Cisco Catalyst スイッチによって VLAN B に割り当てられます。

図 9 EST



VGT

VGT の場合、VM ゲスト オペレーティング システムを使用した 802.1q タグのサポートおよび管理が必須条件です。VM が vNIC を管理し、タギングに関するあらゆる役割をバーチャル スイッチが行わなくて済むようにします。vSwitch の 802.1q タグ サポートをディセーブルにするには、ポートグループ設定で VLAN フィールドを 4095 に設定します。さらに、vNIC ドライバを e1000 に設定しなければならない場合があります。

VGT を設定した場合、各 VM に求められる処理能力が大きくなるので、VM および ESX ホスト全体の効率が下がります。VGT 設定は一般的ではありません。ただし、1 つの VM で 5 つ以上の VLAN をサポートしなければならない場合、VGT を使用する必要があります。



(注)

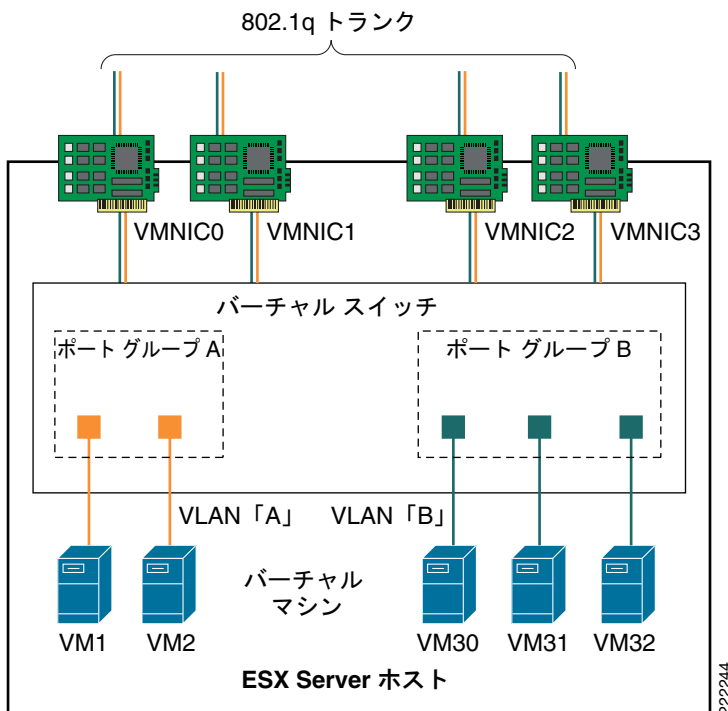
1 つの VM に最大 4 つの独立した vNIC を与え、それぞれ異なる VLAN に配置することができます。したがって、異なる VLAN に VM を配置しなければならない場合は、VGT モードを使用する必要はありません。vNIC ドライバ vlance および vmxnet は VGT をサポートしません。VGT を実行しなければならない場合は、VM vNIC ドライバを e1000 に設定する必要があります。

VST

VST を使用すると、バーチャル スイッチで 802.1q タグ プロセスを実行できます。VMkernel を使用すると、物理アダプタで VLAN タグ処理を実行できるようになるので、VMkernel の負担が軽くなり、システム全体のパフォーマンスが向上します。VST を使用するには、vSwitch に接続された vmnic を 802.1q トランクにする必要があります。ただし、ESX ホスト上で特別な設定は不要です。また、外部ネットワーク ポートも 802.1q トランクとして設定する必要があります。

図 10 に、VST の論理構成図を示します。

図 10 VST



VM の vNIC は、特定の VLAN（この場合は VLAN 「A」 および VLAN 「B」 に対応付けられたポートグループ）に割り当てられます。vSwitch では、スイッチ内部のすべての VLAN をサポートするポート、すなわちトランクとして vnic を定義します。



(注)

DTP（ダイナミック トランッキング プロトコル）を使用すると、2つのスイッチ間でトランク作成に関するネゴシエーションを行えるようになります。ESX バーチャルスイッチは DTP をサポートしません。したがって、vSwitch に接続する Cisco Catalyst スイッチは、スタティック トランッキング用に設定する必要があります。

VST モードでは、vSwitch が限られた数のポート上で多数の VLAN をサポートできます。したがって、サーバ管理者は物理アダプタより多くの VLAN を定義できます。

ネイティブ VLAN

Cisco Catalyst スイッチは、デフォルトで、タグ付けされていないネイティブ VLAN 上のトラフィックに VLAN タグを割り当てることができます。Cisco Catalyst スイッチ上の設定オプション `vlan dot1q tag native` を使用すると、ポートに送出されるトラフィック上の ネイティブ VLAN 対応タグの有無をスイッチが想定できるようになります。vSwitch 上のポートグループが VST 設定方式に基づいて「ネイティブ VLAN」を使用する場合、ネイティブ VLAN 上のトラフィックは 802.1q VLAN タグ付きで転送されます。vSwitch はさらに、ネイティブ VLAN を介してシスコ スイッチから着信するトラフィックにタグが付いていることを想定します。アップストリームのシスコ スイッチポート トランクをこの設定と適合させるには、シスコ スイッチ上で `vlan dot1q tag native` コマンドを設定する必要があります。

次に、VST モードと EST モードが混在する場合について検討します。この場合、vSwitch 上の一部のポートグループでは VLAN ID を定義しますが、その他のポートグループでは EST 方式に従って VLAN ID を使用しません。このようなポートグループからのトラフィックは同じ VLAN に割り当てられ（すなわち、vSwitch 上の VLAN ではない）、スイッチポートで設定されたネイティブ VLAN を使用し、Cisco Catalyst スイッチによって色分けされます。この場合、Cisco Catalyst スイッチ上で `vlan dot1q tag native` コマンドをディセーブルにする必要があります。

NIC チーミングによる接続の冗長化

VMware ネットワーキングのコンテキストにおいて、NIC チーミングとは vSwitch のアップリンクとして使用する ESX Server NIC カードの冗長構成を意味します。bonding ともいいます。図 11 に示したように、NIC チーミングは冗長 vnic 構成を作成することであり、冗長 vNIC 構成を作成することではありません。ESX NIC チーミングは、ポートグループ レベルで設定します。ESX Server 上で冗長構成にするには、アクセス レイヤへの冗長接続を指定して、vSwitch、ポートグループ、またはその両方を設定する必要があります。



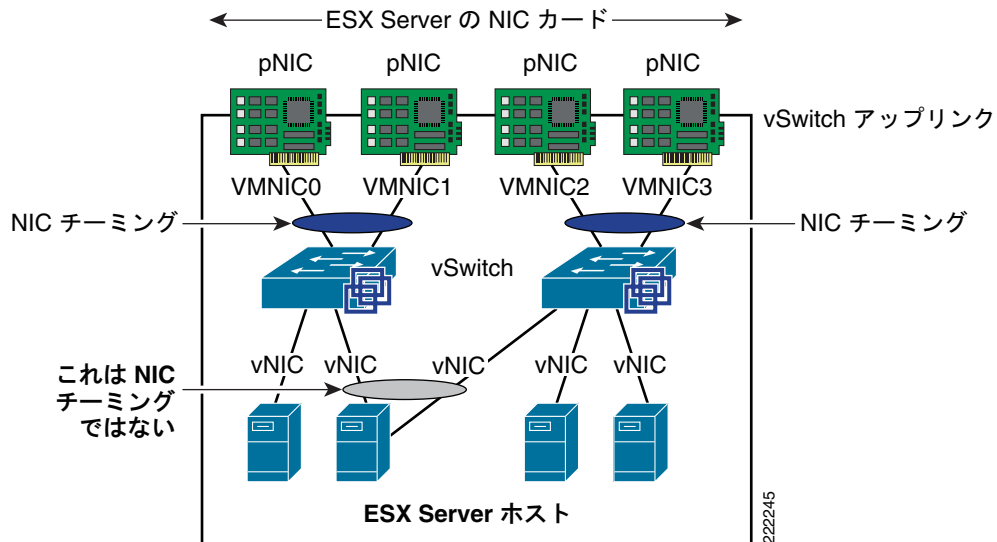
(注)

NIC ベンダー固有の NIC チーミング ソフトウェアで ESX Server の NIC チーミングを設定しないでください。ESX NIC チーミングは、vSwitch の「アップリンク」(vnic) の設定を意味します。

複数の vNIC で VM を構成することは可能ですが、この構成によって冗長性やパフォーマンスが向上することはありません。vSwitch および VM は VMkernel 内部で動作するソフトウェアだからです。異なる物理ネットワーク パスおよび複数の物理 NIC カードまたはポートを利用して冗長 vSwitch アップリンクを設定すると、冗長性が増し、パフォーマンスの向上も望めます。

NIC チーミングを使用すると、PCI カード障害によってネットワーク接続が切断されることがないように異種 NIC カードをバンドルできます。サーバ上での通常の NIC チーミングと同様、NIC を同じレイヤ 2 ドメインに含める必要があります。

図 11 ESX Server における NIC チーミングの意味



NIC チーミングでは、複数の設定オプションを使用できます。いずれも vSwitch 単位またはポートグループ単位で設定できます。

- アクティブ/スタンバイ
- アクティブ/アクティブ、VM ポート ID に基づくロードバランシング
- アクティブ/アクティブ、VM MAC アドレス ハッシュに基づくロードバランシング
- アクティブ/アクティブ、送信元および宛先 IP アドレスのハッシュに基づくロードバランシング。VMware ではこれを IP ベース ハッシュとといいます。シスコではこの設定をポート チャネリングとといいます。

アクティブ/アクティブ (ポートベースおよび MAC ベース)

アクティブ/アクティブモードでは、チーム内のすべての NIC (vmnic) がトラフィックを送受信します。NIC は複数の異なる Cisco Catalyst スイッチに接続することも、1 台のスイッチに接続することもできますが、別々のスイッチを使用して冗長性を確保する方が一般的です。VMware バーチャルスイッチは、送信元 vNIC MAC アドレスを使用して (MAC ベース モード)、またはバーチャルポート ID に基づいて (ポートベース)、チーム内の vmnic 間で出力トラフィックの負荷分散を行います。バーチャルスイッチはチーム内のすべての vmnic を使用します。リンク障害が発生すると、チームで定義されている残りの動作可能インターフェイスに、vSwitch が VM トラフィックを再割り当てします。

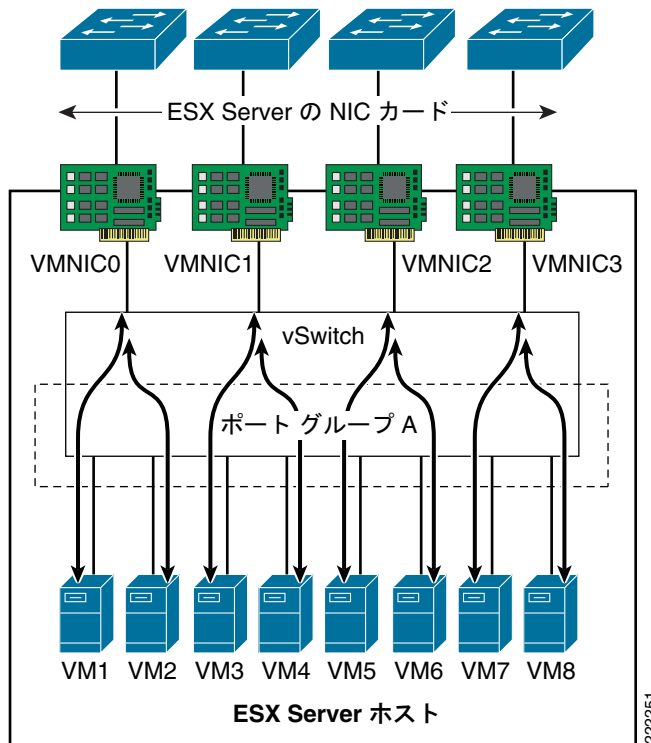
どちらのモードでも、所定の vmnic が動作しているかぎり、VM からのネットワークトラフィックはその特定の vmnic に割り当てられます。VM からのトラフィックは、使用可能なすべての vmnic に均等に振り分けられます。たとえば、チームに 4 つの NIC があり、VM が 8 つの場合、2 つの VM で 1 つの NIC を使用します。

図 12 を参照してください。この例では、VM1 および VM2 が vmnic0 を使用し、VM3 および VM4 が vmnic1 を使用します (以下同様)。vmnic0 をあるレイヤ 2 スイッチに接続した場合、vmnic1 はその同じレイヤ 2 スイッチに接続することも、別のレイヤ 2 スイッチに接続することもできます。

また、vmnic3 についても、同じスイッチに接続することも、異なるスイッチに接続することもできます（以下同様）。vmnic を 1 台のスイッチに接続するか、複数のスイッチに接続するかについて、特別な要件がないのと同様、vmnic または VM を同じ VLAN に配置するか、別の VLAN に配置するかについても、要件はありません。

アクティブ/アクティブ チーミング メカニズムによって、着信トラフィックと発信トラフィックの両方について、VM の MAC と シスコの LAN スイッチ ポート間で一貫性のあるマッピングが保証されます。MAC が別の vmnic（すなわち別の Cisco LAN スイッチポート）に移動するのは、vmnic の 1 つで障害が発生した場合、または管理者が VM を（別の ESX Server などに）移動させた場合だけです。

図 12 vSwitch NIC チーミングでアクティブ/アクティブに設定した場合のトラフィックの分散



アクティブ/アクティブ (IP ベース) (ポート チャネリング)

EtherChannel (別名、802.3ad リンク アグリゲーション) は、個々のイーサネットリンクを集約して 1 つの論理リンクにします。その論理リンクは、最大 8 つの物理リンクからなる集約帯域幅を提供します。VMware 用語では、これを IP ベース ロード バランシングといい、ESX ホストの NIC チーミング設定で使用します。IP ベース ロード バランシングを設定すると、送信元および宛先 IP アドレスのハッシュに基づいて、VM からの発信トラフィックが分散されます。このロード バランシング方式を機能させるには、vmnic の接続先となる Cisco LAN スイッチ上で EtherChannel (すなわち 802.3ad リンク アグリゲーション) を設定する必要があります。

シスコ スイッチ上では、手動で EtherChannel を設定することも、Port Aggregation Control Protocol (PAgP) または 802.3ad Link Aggregation Control Protocol (LACP) を使用して EtherChannel を形成することもできます。EtherChannel プロトコルを使用すると、接続ネットワーク デバイスとのダイナミック ネゴシエーションによって、特性が類似しているポートで EtherChannel が形成されます。PAgP はシスコ独自のプロトコルであり、LACP は IEEE 802.3ad で定義されています。シスコのスイッチは両方のプロトコルをサポートします。

ESX ホスト上では、vSwitch IP ベース ロード バランシングを行っても 802.3ad LACP は実行されません。したがって、Cisco Catalyst スイッチ上の EtherChannel 設定ではダイナミック ネゴシエーションを使用できません。この場合、channel-group を ON に設定します。このような設定はスタティックであり、バンドルの反対側がチャネリング対応として設定されているかどうかに関係なく、Cisco Catalyst スイッチと vSwitch の両方がそれぞれのロード バランシング アルゴリズムを使用して、バンドル リンクのいずれかの側で負荷を分散させます。

図 13 に、この方式の NIC チーミングを示します。すべての NIC は 1 つのスイッチ（または適切なテクノロジーで「クラスタ」化された複数のスイッチ）に接続します。同じ IP ベース NIC チーミング構成に所属するすべての NIC ポートは、Cisco LAN スイッチ上の同じポートチャネルのメンバーでなければなりません。

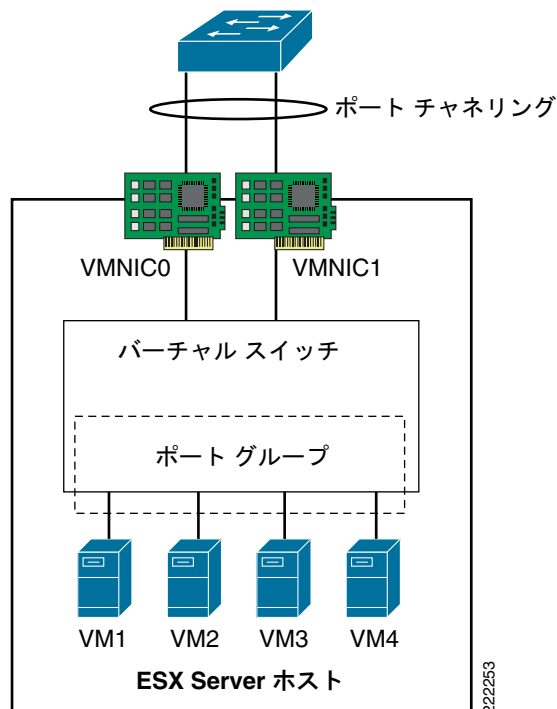


(注) Cisco Catalyst スイッチ上での EtherChannel 設定の詳細については、次の URL を参照してください。
<http://www.cisco.com/en/US/partner/docs/switches/lan/catalyst6500/ios/12.2SX/configuration/guide/channel.html>



(注) Virtual Switching System (VSS) 対応として設定された複数の Cisco Catalyst 6500 スイッチ、または Virtual Blade Switching (VBS) 対応として設定された複数の Cisco Blade Switch (CBS) に ESX Server を接続し、IP ベース ハッシュの NIC チーミングを設定できます。

図 13 vSwitch の EtherChannel 設定



チーミング フェールバック

Teaming Failback (チーミング フェールバック) 機能は、NIC チーミングのプリエンプト動作を制御します。たとえば、2 つの *vmnic*、*vmnic0* および *vmnic1* があるとします。*vmnic0* で障害が発生すると、トラフィックは *vmnic1* に送られます。*vmnic0* が使用可能になると、デフォルトの動作 (*Teaming Failback* を *ON* に設定) として、トラフィックは再び *vmnic0* に割り当てられます。

この動作では、LAN スイッチング側がリンクアップし、ポートがただちにフォワーディングモードにならなかった場合、トラフィックのブラックホールが生じるリスクがあります。この問題は、Cisco Catalyst スイッチ側で *trunkfast* を使用し、トランクモードを *ON* に設定することによって、容易に対処できます。ESX ホスト側から、チーミング フェールバック機能をディセーブルにすることによって、この問題に対処することもできます。この場合、*vmnic0* が再びリンクアップになっても、NIC は現在アクティブな *vmnic1* で障害が発生するまで、非アクティブアップの状態を維持します。

ESX 3.5 より前のリリースでは、*Rolling Failover* をディセーブルにすると、チーミング フェールバックモードがイネーブルになります。*Failback = No* (ESX 3.5) は *Rolling Failover = Yes* (ESX 3.5 より前のリリース) と同じです。逆も真です。

ビーコン機能

ビーコン機能は、チーム内の *vmnic* の状況を ESX ホストに監視させるブローブ機能です。ビーコン機能を使用するには、*vmnic* を同じブロードキャストドメインに配置する必要があります。ビーコンは、複数の外部スイッチに接続されたチームで使用します。ESX Server は、ビーコンブローブの損失を監視して、外部ネットワーク障害を判別します。障害条件が存在する (つまり、ビーコンの開始側から *x* の数のビーコンを受信したことを *vmnic* が報告しなかった) 場合、ESX Server はアダプタを切り替え、プライマリアダプタのダウンを宣言します。

ビーコンフレームはレイヤ 2 フレームで、*EtherType* は *0x05ff*、送信元 MAC アドレスは (VMware の MAC アドレスではなく) NIC カードに焼き付けられたアドレスおよびブロードキャスト宛先アドレスです。フレームは *vSwitch* が配置されているあらゆる VLAN 上で送信されます。



(注) ビーコン機能は、ポートグループ単位で設定します。

外部ネットワーク障害検出の形式として、ビーコン機能を使用することは推奨できません。False Positive である可能性があり、また、アップストリーム障害を検出できないからです。可用性の高い外部ネットワークインフラストラクチャを提供するには、冗長パス、ネットワークベースのロードバランシングを実行するプロトコル、またはその両方を使用して、ハイアベイラビリティを実現してください。

Link State Tracking (リンクステートトラッキング) 機能は、ESX Server のアップストリームリンクとダウンストリームリンクを関連付けます。アップストリームリンク障害がダウンストリームリンク障害のトリガーになり、ESX Server は *Network Failover Detection* の *Link State Only* を使用して、このダウンストリームリンク障害を検出できます。*Link State Tracking* 機能は、Cisco Catalyst ブレードスイッチ、Catalyst 3750、Catalyst 2960、および Catalyst 3560 で使用できます。ご使用の Cisco Catalyst スイッチがこの機能をサポートするかどうかについては、シスコの Web サイトで確認してください。*Link State Tracking* (リンクステートトラッキング) 機能は、ESX Server のアップストリームリンクとダウンストリームリンクを関連付けます。アップストリームリンク障害がダウンストリームリンク障害のトリガーになり、ESX Server は *Network Failover Detection* の *Link State Only* を使用して、このダウンストリームリンク障害を検出できます。

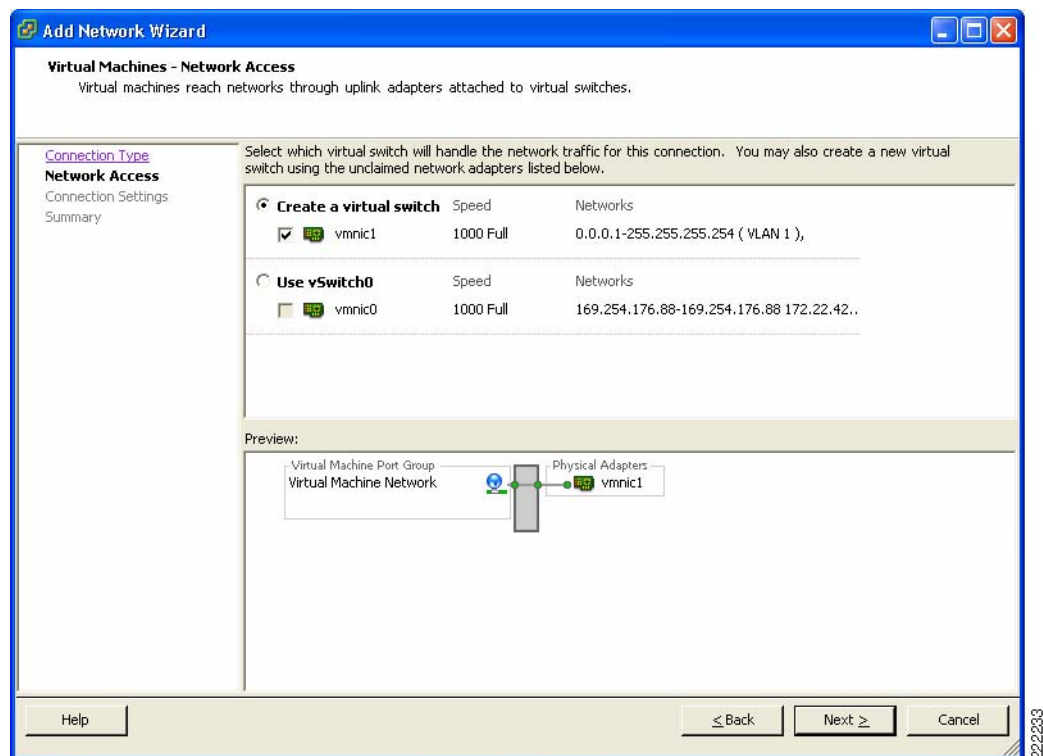
vSwitch の設定

ESX ホストは、カーネルのコンテキストで動作する、vSwitch というソフトウェア構造を使用して、ローカル VM を相互に、または外部企業ネットワークにリンクさせます。vSwitch は、従来の物理イーサネット ネットワーク スイッチをエミュレートすることにより、データ リンク レイヤ（レイヤ 2）でフレームを転送します。

vSwitch の作成

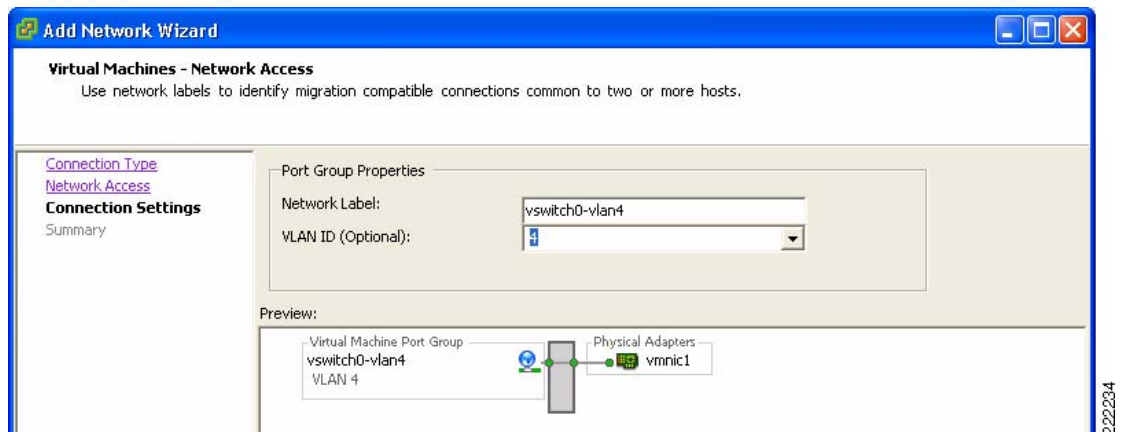
VM を作成するたびに、2 つのオプションのどちらかを選択して、ネットワーク アクセスを設定する必要があります。既存の vSwitch を使用するか、新しい vSwitch を作成するかです。図 14 を参照してください。

図 14 VirtualCenter の GUI から新しい vSwitch を作成する方法



新しい vSwitch を作成することにした場合は、さらにポート グループを作成できます。ポート グループは、ネットワーク ラベル、VLAN 番号、およびその他の特性（このマニュアルの他の項で説明）によって定義します。図 14 および図 15 に vSwitch0 を作成し、vSwitch0 上で VLAN 4 を使用するポート グループを定義する例を示します。このポート グループでは、ネットワーク ラベル **vSwitch0-vlan4** を使用します。

図 15 VirtualCenter の GUI から VLAN を作成する方法



(注) vSwitch 内部でのポートグループのネーミング、つまりラベリングは、ESX 環境を作成して維持するうえで重要な基準です。ポートグループのネットワーク ラベルを VLAN にちなんだ名前にすることも、vSwitch 名および VLAN を指定することもできます。また、このポートグループに接続するアプリケーションの名前を使用することもできます。



(注) ここでは、使用する vSwitch と VLAN が反映されたネットワーク ラベル名を選択します。vSwitch 名にはローカルな意味しかないうえに、自動生成されるので、これがポートグループのネーミング方式として最良だというわけではありません。VM のモビリティを得るために、LAN と同様、起点ネットワーク ラベルと宛先ネットワーク ラベルを同じにする必要があります。そのため、VM をどの vSwitch に移行させるかに関係なく、別の ESX Server 上で使用できるネットワーク ラベル名にする必要があります。

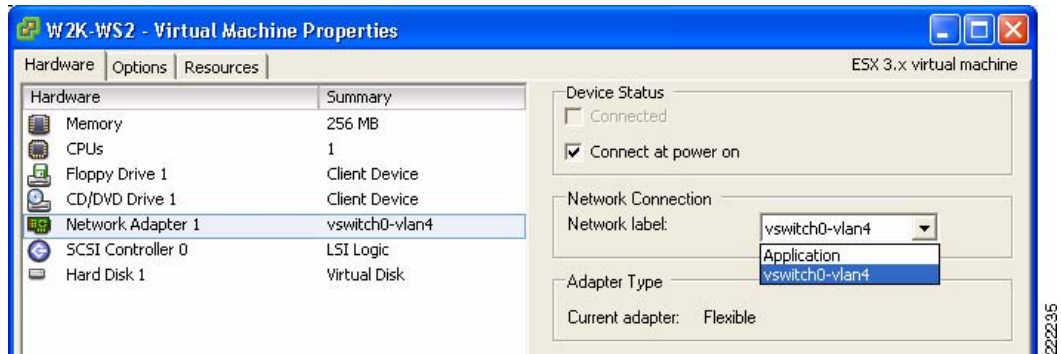


(注) ポートグループ ラベルは、同じ ESX ホスト上の vSwitch 間で重複しないようにする必要があります。

VM の **Edit Settings** ウィンドウでは、VM の vNIC を *Network Adapter* (ネットワーク アダプタ) といいます。VM ネットワーク アダプタ設定 (すなわち vNIC 設定) で、対応するネットワーク ラベルを参照することによって、VM の接続先 vSwitch 内のポートグループを選択します。ネットワーク アダプタが使用するポートグループを選択すると、vSwitch、および vSwitch 内の特定の VLAN に暗黙的に VM を割り当てることとなります。

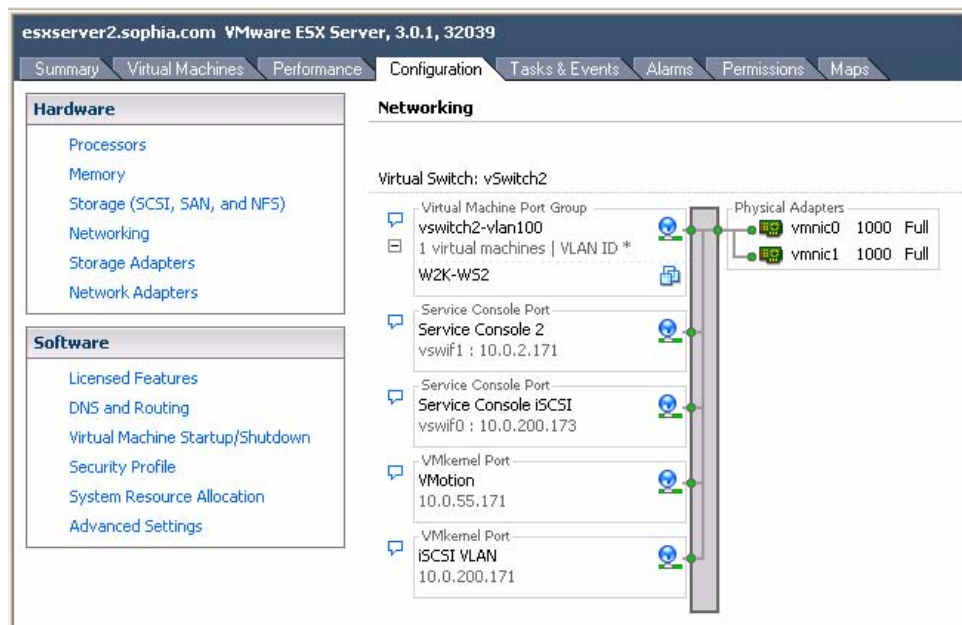
図 16 で、ポート グループに与えられているラベルは、トラフィックの色分けに使用する VLAN、およびそのポート グループが定義されている vSwitch にちなんだものです。

図 16 vNIC および VLAN の追加



VMware ESX Server -> Configuration -> Networking タブから設定例を表示できます。図 17 を参照してください。この例では、W2K-WS2 という VM が VLAN 100 上の vSwitch2 に接続されます。

図 17 vSwitch の最終的な設定



vSwitch の右側にある **Properties** ボタンを選択すると、vSwitch の特性をさらに変更できます。ポート グループの追加、NIC Teaming プロパティの変更、トラフィック レート制限の設定などが可能です。

VM、VMkernel、および Service Console のネットワーク設定

図 17 が示すとおり、vSwitch2 は複数の VM (W2K-WS2)、Service Console、および VMkernel へのネットワーク接続を提供します。

Service Console は、ESX Server の管理用として VirtualCenter または Virtual Infrastructure Client が使用するため、Service Console の設定を変更する場合は慎重に見直し、ESX Server への管理アクセスが失われないようにする必要があります。VMkernel のネットワーク設定は、NAS、iSCSI アクセス、および VMotion に使用されます。

Service Console および VMkernel の設定では、IP アドレス、デフォルト ゲートウェイ (Service Console の IP およびゲートウェイとの一致は不要)、VLAN 番号などを設定できます。図 17 に、これらのアドレスを示しています。CLI コマンドを使用すると、Service Console および VMkernel の接続を確認できます。Service Console には **ping** コマンド、VMkernel には **vmkping** コマンドを使用します。

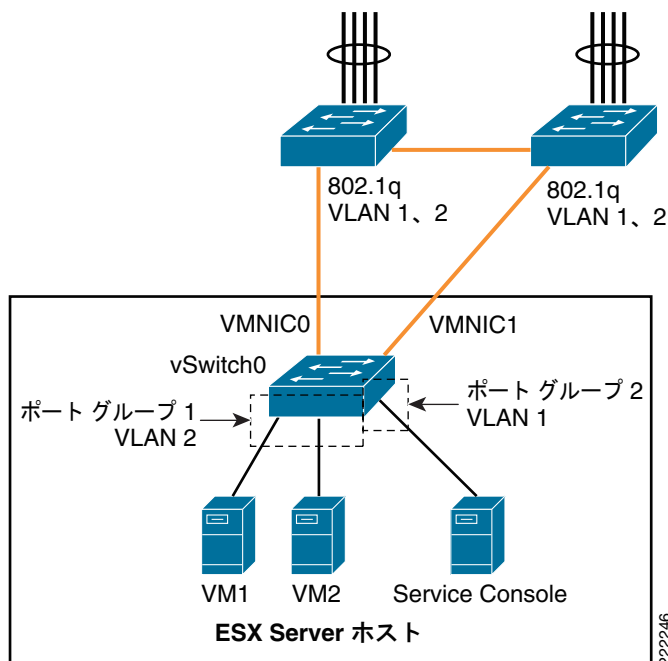
Service Console および VMkernel に、それぞれ専用の NIC を与えることが望ましいのですが、実際には vmnic を共有する形が多くなります。この場合、同じ vSwitch を共有しつつ、Service Console および VM カーネル ポートに専用の VLAN に配置し、VM はこの 2 つとは異なる VLAN に配置することを推奨します。

このように構成すると、後述するように、vmnic が 802.1q トランッキング対応として設定されます。

vSwitch NIC チーミングの設定

NIC チーミングは、vSwitch および vmnic の設定であり、vNIC の設定ではありません。VMware ネットワーキングのコンテキストにおいて、NIC チーミングは vSwitch アップリンクの冗長構成を意味します。NIC は複数の異なる Cisco Catalyst スイッチに接続することも、1 台のスイッチに接続することもできますが、別々のスイッチを使用して冗長性を確保する方が一般的です。図 18 に、基本設定を示します。

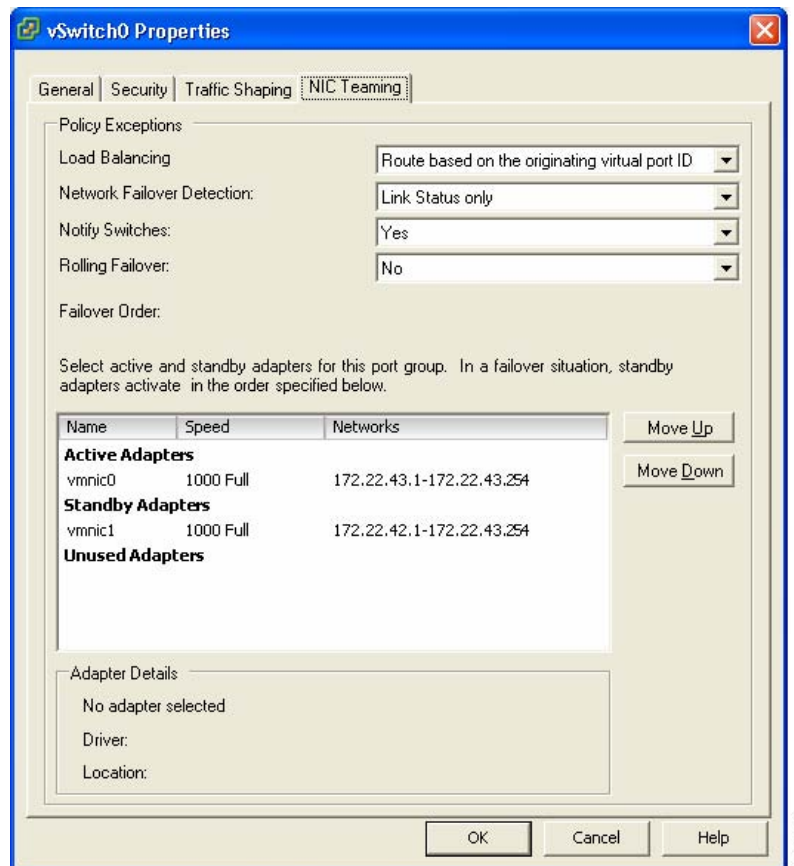
図 18 アクティブ/スタンバイ NIC チーミング



この例では、vmnic0 をアクティブアップリンク、vmnic1 をスタンバイアップリンクとして、vSwitch0 を設定しています。どちらのリンクも VLAN1 および VLAN2 を伝送します。VLAN 1 は Service Console のトラフィックを、VLAN 2 は VM1 および VM2 のトラフィックを伝送します。

vSwitch0 を設定するには、ESX Host Network Configuration を表示し、vSwitch0 Properties を選択して、図 19 のように vSwitch の設定を編集します。

図 19 vSwitch NIC チーミングのプロパティ



これは vSwitch 全体の設定になります。

- この NIC チーミング設定は、vSwitch 上で設定したすべてのポートグループ/VLAN に適用されます。
- vmnic は設定されているすべての VLAN を伝送します。
- アクティブ vmnic (vmnic0) で障害が発生した場合は、すべてのトラフィックがスタンバイ vmnic (vmnic1) に割り当てられます。
- vSwitch 全体の設定を各ポートグループから変更できます。

ポートグループの設定

アクティブ / スタンバイ方式の NIC チーミング設定には、NIC のうちの 1 つ（例ではスタンバイ vmnic1）が使用されないという短所があります。NIC チーミングはほとんどの場合、vSwitch 全体の設定ですが、ポートグループレベルでグローバルな設定を変更することもできるので、ポートグループ単位でアクティブ / スタンバイを設定すると、NIC をフルに活用できます。

例として、図 20 に ESX Server の論理構成図を示します。チーミングが設定された vSwitch が 1 つあります。2 つの ESX Server NIC がアップリンクとして vSwitch に接続されています。VM5 および VM7 は vSwitch1 のポートグループ 1 に接続し、vSwitch から送出されるトラフィックに vmnic0 を使用するよう設定されています。vmnic1 はスタンバイであり、vmnic0 で障害が発生したときに引き継ぎます。VM4 および VM6 は vSwitch のポートグループ 2 に接続し、優先アップリンクとして vmnic1 を使用します。vmnic0 はスタンバイであり、vmnic1 で障害が発生したときに引き継ぎます。

VM4、VM5、VM6、および VM7 を別々の LAN に配置する必要はありません。同じ VLAN に含めることができますが、単に 2 種類のアップリンクにトラフィックを分散させる目的で、2 つの異なるポートグループに配置しています。



(注)

VM4、VM5、VM6、および VM7 が同じ VLAN 上の 2 つの異なるポートグループにある場合、それらは同じブロードキャストドメインに含まれます。

図 20 vSwitch ポートグループの NIC チーミング設定

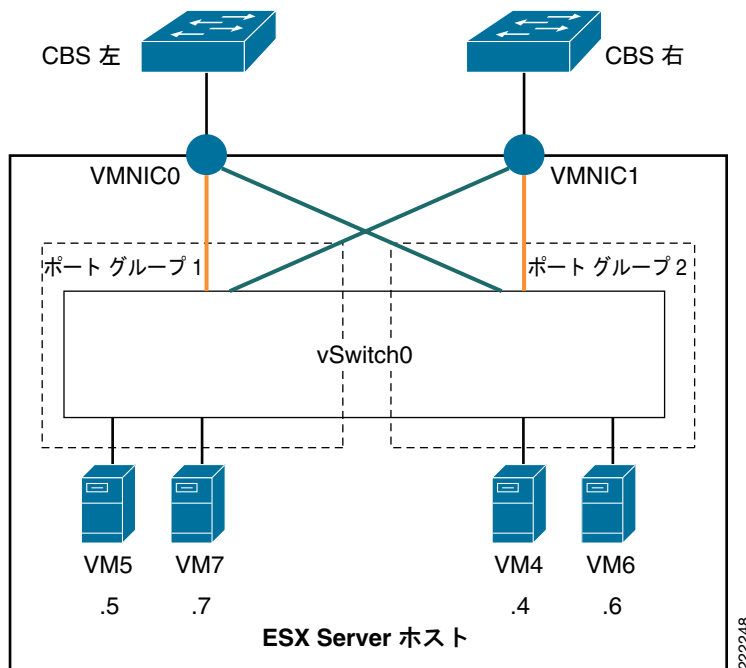
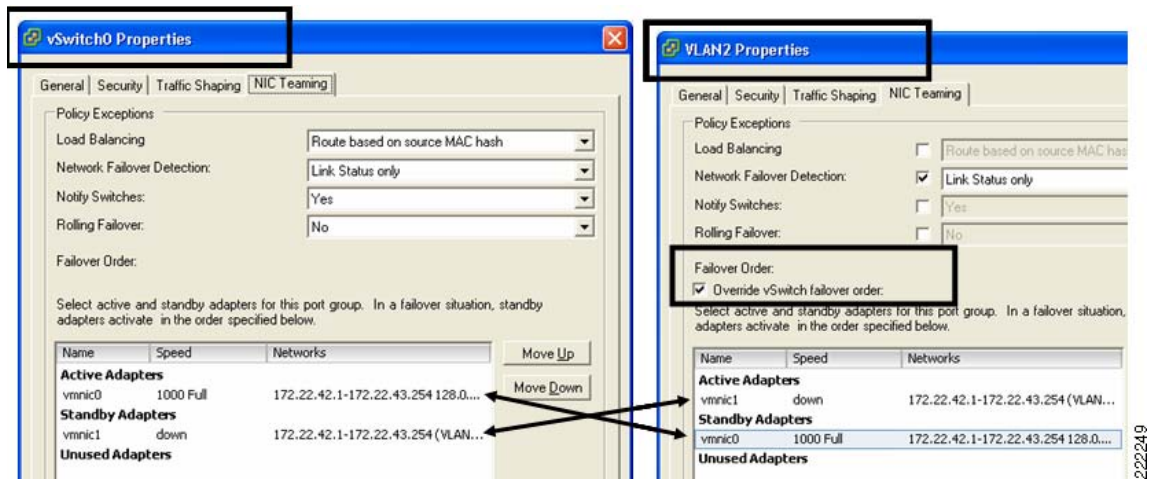


図 21 に、VirtualCenter でこれを設定する方法を示します。ESX ホスト設定、ネットワーキングから、vSwitch0 のプロパティを選択します。「ポート タブ」内で該当するポート グループを選択すると（たとえば、VLAN 番号などのネットワーク ラベルで識別）、NIC Teaming (NIC チーミング) の設定を含めてポート グループのプロパティを変更できます。図 21 で、vSwitch のプロパティとポート グループのプロパティを比較します。ポート グループの設定によって、vSwitch 全体の設定を上書きできます。2つの設定で vmnic の順序が逆になっていることを確認してください。

図 21 vSwitch ポート グループの NIC チーミングによる vSwitch NIC チーミングの更新



この例が示しているものは、次のとおりです。

- アクティブ / スタンバイ NIC チーミング
- 同じ vSwitch 上のポート グループごとに、異なるチーミング / NIC Teaming 機能を設定 (VLAN が両方のポート グループで同じ場合も同様)

NIC チーミングの設定変更

同じ NIC チーミング ポリシーを共有するすべての VM は、同じポート グループに含まれます。vSwitch のプロパティから、該当するポート グループ (すなわち ネットワーク ラベル) および NIC Teaming プロパティを選択します。ロードバランシングのスクロールバーで、ポートベース、MAC ベース、または IP ベースのロードバランシングを選択できます。

使用可能な vmnic の一覧を大型のウィンドウで表示できます。アクティブ リスト、スタンバイ リスト、または未使用リストを上下に移動できます。そのため、有効な vmnic のプールを、さまざまなポート グループから柔軟に使用できるようになります。

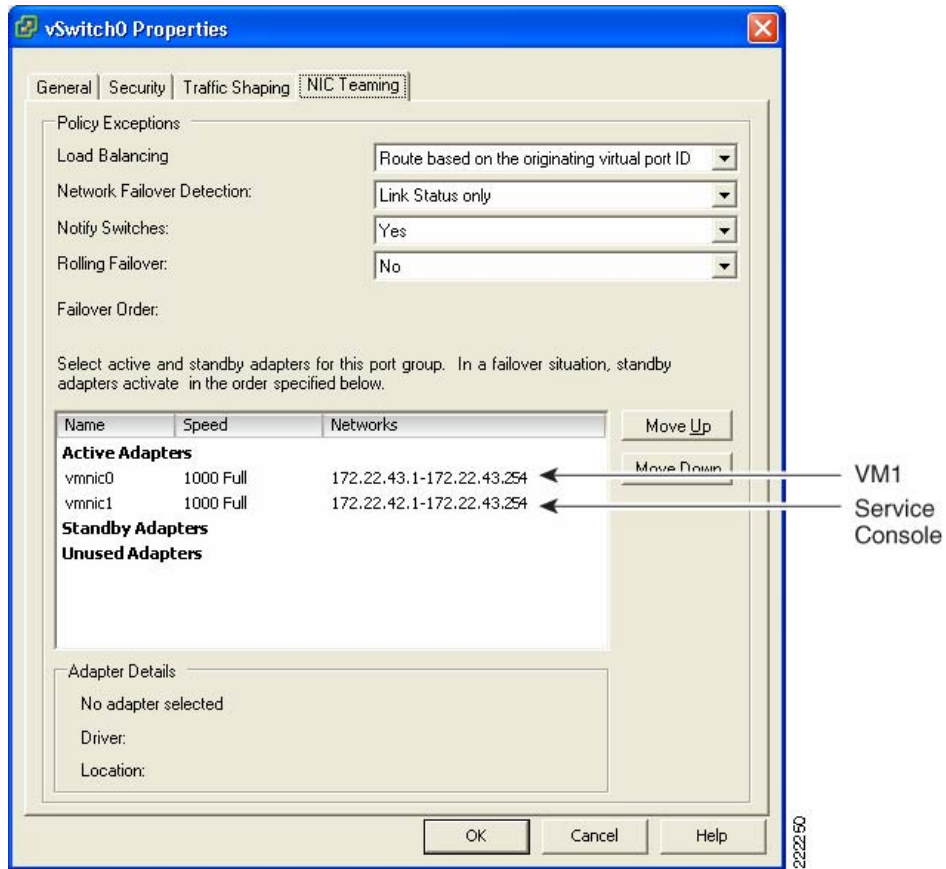
- あるポート グループ (すなわち、1 つの VM セット) では vmnic0 だけを使用していて、vmnic1 はスタンバイのままになっているということもあります。
- 別のポート グループでは、vmnic0 と vmnic1 の両方を使用して、ポートベースのロードバランシングを行っているということもあります。
- さらに別のポート グループでは、逆の順序で vmnic0 と vmnic1 の両方を使用していて、vmnic 間の VM 展開を向上させているということもあります。
- あるポート グループでは vmnic0 だけを使用していて、vmnic1 は Unused (未使用) リストに残っているということもあります (この場合、vmnic0 で障害が発生すると、VM が切り離されます)。



(注)

複雑化を避ける場合は、vSwitch レベルでアクティブ / アクティブ ポートベース ロードバランシング設定だけを選択し、他のポート グループ NIC チーミング設定には触れないでおきます。

図 22 vSwitch NIC チーミングのオプション

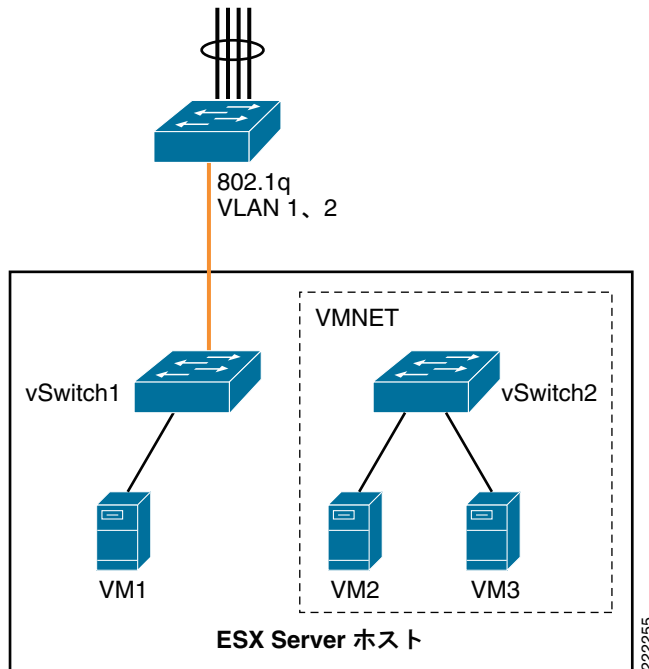


ESX 内部のネットワーク

VM の内部ネットワークは、*vmnet* またはプライベート ネットワークともいいます。内部ネットワークは ESX ホストに対してローカルであり、LAN スイッチング環境には接続できません。vmnet では仮想スイッチを使用して、ESX Server に VM を内部的にリンクします。内部 vSwitch には vmnic が割り当てられないことを除き、設定は外部ネットワーク設定と特に変わりません。システムバスがトランスポートを提供し、CPU がトラフィックを管理します。vmnet は通常、テスト環境および開発環境で使用します。

図 23 に、内部ネットワーク設計の使用例を示しますが、それほど便利なものではありません。この例では、VM2 および VM3 は vSwitch2 のメンバです。vSwitch2 上のポートグループ/VLAN は、完全に ESX Server の内部にあります。VM2 および VM3 は外部と通信しません。

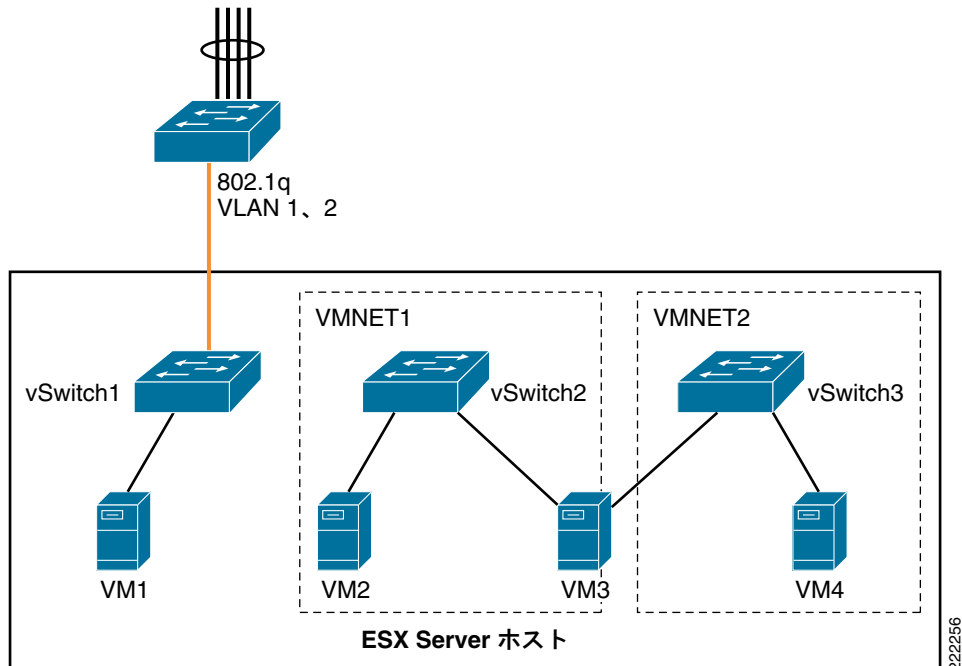
図 23 ESX Server のプライベート バーチャル スイッチ



ルーティングおよびブリッジング

図 24 のように、ルーティングまたはブリッジング対応として設定された VM を使用すると、複数のプライベート ネットワークを作成して相互接続することができます。図 24 では、VMNET1 (vSwitch2) および VMNET2 (vSwitch3) が VM3 によって相互接続されています。VM3 には VNET ごとに 1 つずつ、2 つの vNIC があります。

図 24 ルーティングまたはブリッジングを使用する ESX Server のプライベート バーチャル スイッチ



使用例

プライベート vSwitch はテスト目的で使用しますが、VM 間にプライベートな通信チャンネルを作成しなければならない場合にも便利です。図 25 に例を示します（推奨例ではありません）。
 図 25 では、VM4 と VM6 に vNIC が 2 つあります。パブリック vNIC は、vSwitch0 を介して、外部の Cisco LAN スイッチング ネットワークに接続します。プライベート vNIC は、プライベート vSwitch (vSwitch1) に接続します。この構成では、必要に応じて 2 つの VM がサーバ上でローカルにハートビートまたはステータス情報を交換できます。

図 25 VM 間のプライベート通信に使用するプライベート vSwitch

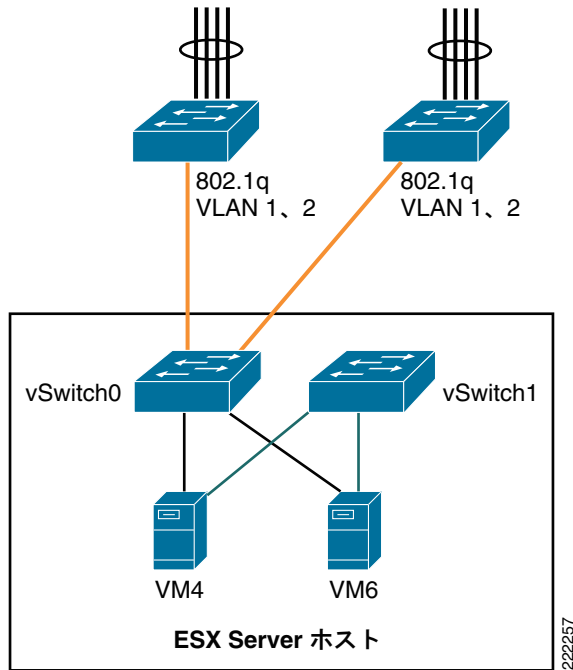
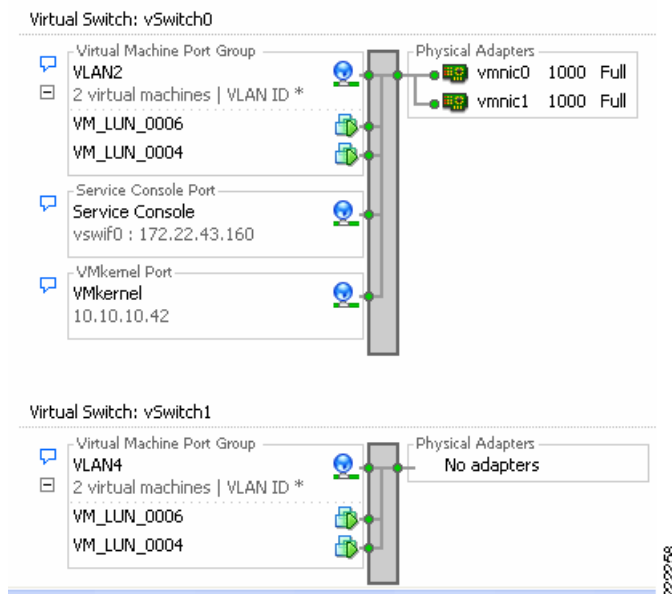


図 25 の設定は、VirtualCenter では図 26 のように表示されます。

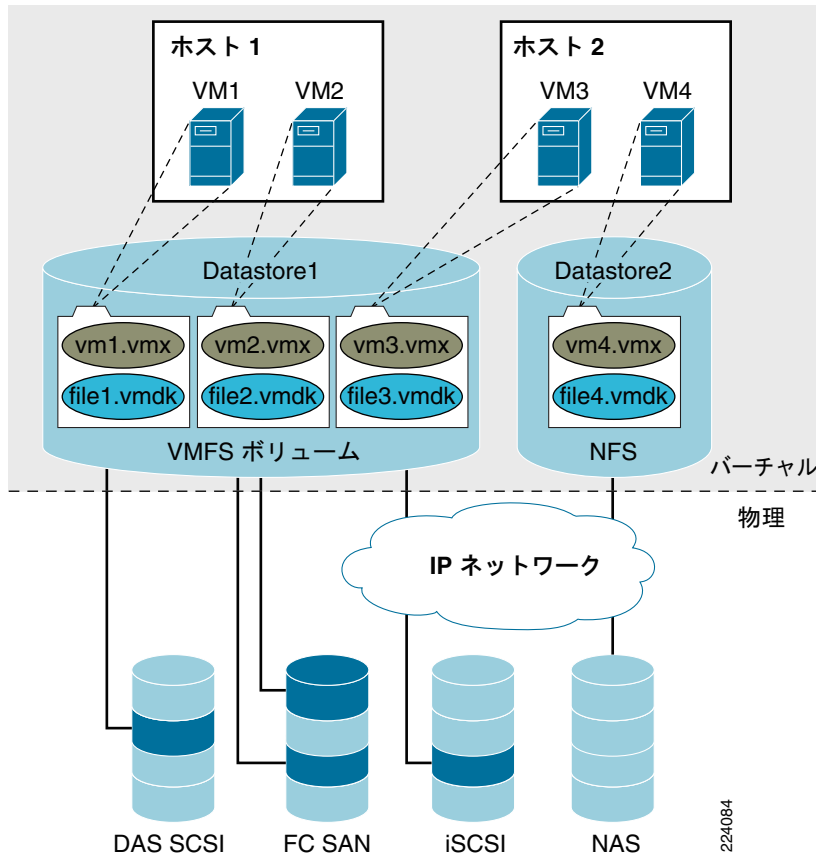
図 26 VirtualCenter でのプライベート ネットワークの表示



ESX Server Storage ネットワーキング

VMware Infrastructure Storage アーキテクチャ（図 27 を参照）は、抽象化レイヤを提供します。このレイヤでは、物理ストレージサブシステムの複雑性および物理ストレージサブシステム間の相違を隠したり管理したりします。各仮想マシン内部のアプリケーションおよびゲストオペレーティングシステムに対して、ストレージは単に仮想 BusLogic として表示されるか、または LSI SCSI HBA に接続された SCSI ディスクとして表示されます。

図 27 VMware Infrastructure Storage アーキテクチャ



仮想マシン内部の仮想 SCSI ディスクは、データセンターのデータストア要素からプロビジョニングを行います（図 27 を参照）。データストアはストレージアプライアンスに類似しており、仮想マシン内部の仮想ディスクにストレージスペースを提供し、仮想マシン定義そのものを保管します。図 27 に示したとおり、仮想マシンはデータストアの専用ディレクトリに、一組のファイルとして格納されます。

仮想ディスク (vmdk) は、データストア内のファイルで、ESX が管理します。データストアは、ブロックベースストレージ用の VMFS ボリュームまたは NFS ストレージ用のマウントポイントに配置します。VMFS ボリュームは通常、単一 LUN で構成されますが、複数の LUN にまたがるようにすることもできます。仮想ディスクはファイルとまったく同じように、容易に操作（コピー、移動、バックアップなど）できます。ゲスト OS が新しいディスクのホットアドをサポートする場合、VM をシャットダウンしなくても、新しい仮想ディスクを追加できます。

データストアを使用すると、下記をはじめ、各種物理ストレージテクノロジーの複雑性を気にせずに、バーチャルマシンにストレージスペースを割り当てられる単純なモデルを構成できます。

- **ファイバチャネル SAN** — 最も一般的な稼働オプションです。VMotion、SAN からの ESX ブート、raw デバイスマッピングのサポート、ハイアベイラビリティクラスタ (Microsoft MSCS など) のサポートが使用できます。Virtual Machine File System (VMFS; バーチャルマシンファイルシステム) をサポートします。
- **iSCSI SAN** — ハードウェアベースのアクセラレーションと関連付けた場合に、ファイバチャネル SAN と同様の機能が使用できるようになります。VMotion 移行、SAN からの ESX ブート、raw デバイスマッピングのサポートが含まれます。ソフトウェアベースの iSCSI の場合、SAN からのブートはサポートされません。VMFS をサポートします。
- **直接接続ストレージ** — 共有されないため、一般的ではありません。ESX Server 3.5 は、ローカル (DAS) ストレージ上のスワップファイルを使用する VMotion をサポートします。
- **NAS** — NFS ベースのストレージは、raw デバイスマッピングもハイアベイラビリティクラスタもサポートしません。また、VMFS を使用しませんが、VMotion 移行および NFS からのブートは可能です。

データストアは、物理的には単なる VMFS ボリュームまたは NFS マウントディレクトリです。各データストアが複数の物理ストレージサブシステムにまたがるようにできます。

図 27 に示したとおり、1 つの VMFS ボリュームに、物理サーバ上の直接接続 SCSI ディスクアレイ、ファイバチャネル SAN ディスクファーム、または iSCSI SAN ディスクファームの小規模なボリュームを 1 つまたは複数含めることができます。物理ストレージサブシステムに追加された新しいボリューム、またはストレージサブシステム内で展開される、ESX Server が認識済みの LUN は、VirtualCenter 管理インターフェイスを通じて再スキャン要求が出された場合に、ESX Server によって検出されます。これにより、物理サーバまたはストレージサブシステムの電源を切断しなくても、作成済みのデータストアを拡張できます。

VMware ESX Server のストレージ コンポーネント

ここでは、ESX Server の内部コンポーネントとその動作について、詳しく説明します。図 28 に、ESX Server アーキテクチャと VMware ストレージ動作を実行する個々のコンポーネントの詳細を示します。

図 28 ストレージ アーキテクチャ コンポーネント

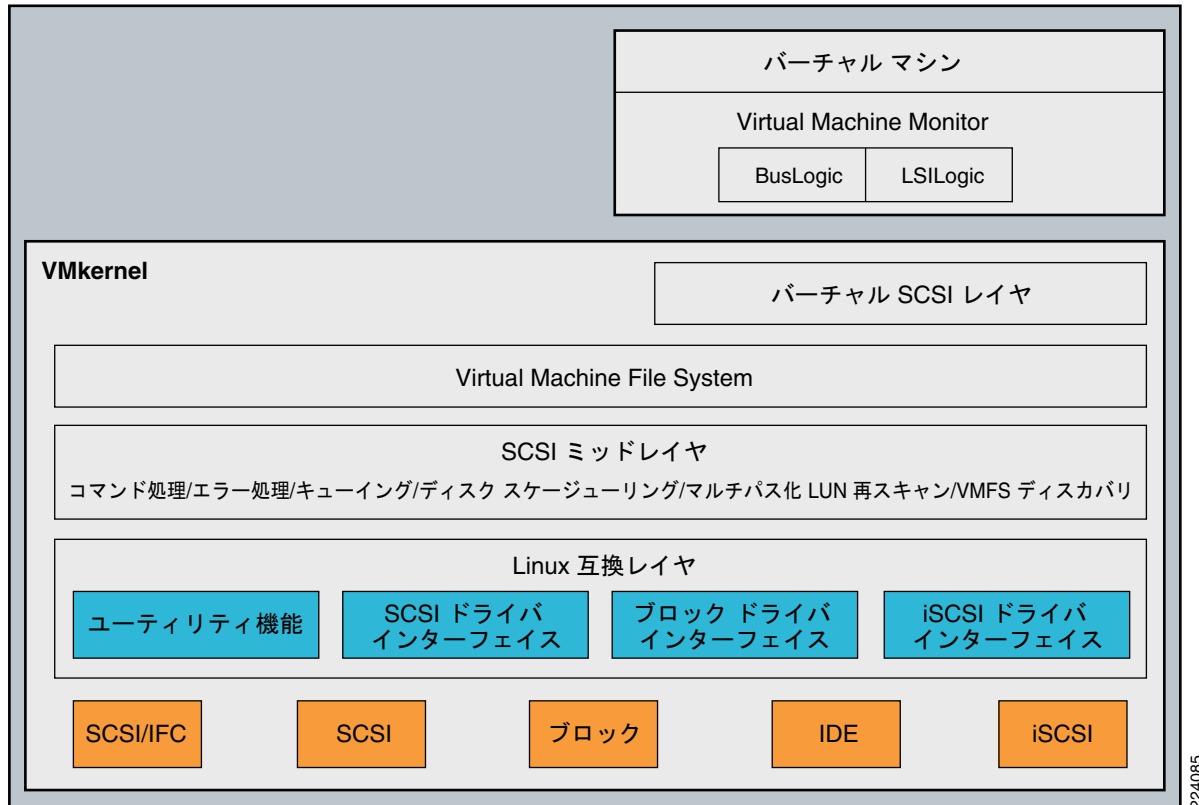


図 28 に示した主要なコンポーネントは、次のとおりです。

- Virtual Machine Monitor (VMM)
- バーチャル SCSI レイヤ
- VMFS
- SCSI ミッドレイヤ
- Host Bus Adapter (HBA) デバイス ドライバ

Virtual Machine Monitor (VMM)

VMM モジュールの主な役割は、バーチャル マシンのアクティビティをあらゆるレベル (CPU、メモリ、入出力、およびその他のゲスト オペレーティング システム機能および VMkernel との対話) で監視することです。VMM モジュールには、バーチャル マシン内部で SCSI デバイスをエミュレートするレイヤがあります。バーチャル マシンのオペレーティング システムは、ファイバチャネル デバイスに直接アクセスすることはできません。VMware インフラストラクチャはストレージを仮想化して、オペレーティング システムに SCSI インターフェイスを提供するだけだからです。したがって、バーチャル マシンのタイプを問わず (オペレーティング システムに関係なく)、アプリケーションは SCSI ドライバのみを使用してストレージ サブシステムにアクセスするだけです。バーチャル マシンでは、BusLogic または LSI Logic の SCSI ドライバをどちらでも使用できます。これらの SCSI ドライバによって、バーチャル マシン内部でバーチャル SCSI HBA を使用できるようになります。



(注)

Windows バーチャル マシン内部で、Windows コントロールパネルから **Computer Management > Device Manager > SCSI and RAID Controllers** の順に選択すると、**BusLogic** ドライバまたは **LSI Logic** ドライバが表示されます。**BusLogic** は、Mylex BusLogic BT-958 エミュレーションが使用されていることを意味します。BT-958 は、40 Mbps の Ultra SCSI (Fast-40) 転送速度を提供する、SCSI-3 プロトコルです。ドライバエミュレーションは、「SCSI Configured Automatically」(SCAM) の機能をサポートします。SCAM を使用すると、ID 番号を自動的に指定して SCSI デバイスが設定されるので、手動で ID を割り当てる必要がありません。

バーチャル SCSI HBA

ESX Server 環境では、各バーチャル マシンに 1 ~ 4 のバーチャル SCSI HBA が含まれます。バーチャル SCSI HBA によって、バーチャル マシンは論理 SCSI デバイスにアクセスできます。物理 HBA によって物理ストレージデバイスにアクセスできるのとまったく同様です。ただし、物理 HBA とは異なり、バーチャル SCSI HBA でストレージ管理者 (SAN 管理者など) が物理マシンにアクセスすることはできません。

バーチャル SCSI レイヤ

バーチャル SCSI レイヤの主な役割は、SCSI コマンドを管理し、VMM、VMFS、および SCSI ミッドレイヤ以下の相互通信を管理することです。バーチャル マシンから実行する SCSI コマンドは必ず、バーチャル SCSI レイヤを通過しなければなりません。入出力の打ち切り、リセットなどの操作も、このレイヤで管理します。バーチャル マシンからの入出力または SCSI コマンドは、バーチャル SCSI レイヤから下位レイヤに VMFS または デバイス マッピング (RDM) を使用して渡されます。この場合、2 種類のモードがサポートされます。パススルーおよび非パススルーです。RDM パススルー モードでは、すべての SCSI コマンドがトラップなしでパススルーされます。

Virtual Machine File System (VMFS)

VMFS はクラスタ型ファイル システムです。VMFS を使用すると、複数の物理サーバが共有ストレージを活用して、同じストレージに読み取りおよび書き込みを同時に実行できるようになります。VMFS では、オンディスク分散ロックングを使用して、複数のサーバが同時に同じバーチャルマシンをオンにしないようにします。

単純な構成では、バーチャル マシンのディスクは VMFS 内にファイルとして保管されます。ゲスト オペレーティング システムが対応するバーチャル ディスクに SCSI コマンドを発行すると、バーチャライゼーション レイヤでそれらのコマンドが VMFS ファイル操作に変換されます。VMFS の詳細については、「[ファイル システムのフォーマット](#)」(p.37) を参照してください。

SCSI ミッドレイヤ

SCSI ミッドレイヤは、VMkernel のストレージ動作に最も重要なレイヤであり、ESX Server ホスト上の物理 HBA を管理し、要求をキューに格納し、SCSI エラーを処理します。さらに、このレイヤには ESX Server ホストに割り当てられた LUN マッピングの変更を検出する、自動再スキャン ロジックが組み込まれています。自動パス選択、パス折りたたみ、フェールオーバー、特定のボリュームへのフェールバックなどのパス管理も、SCSI ミッドレイヤで処理されます。

SCSI ミッドレイヤは HBA、スイッチ、およびストレージポート プロセッサから情報を集めて、ESX Server ホストとストレージアレイ上の物理ボリューム間のパス構造を識別します。再スキャン時に、ESX Server は Network Address Authority (NAA) 識別子、シリアル番号などのデバイス情報を調べます。ESX Server はストレージアレイへの使用可能なすべてのパスを識別し、使用できるパスの数に関係なく、折りたたんで 1 つのアクティブ パスにします。他のすべての使用可能なパスは、スタンバイとして表示されます。パス変更検出は、自動的に行われます。TEST_UNIT_READY SCSI コマンドに対するストレージ デバイスの応答に応じて、ESX Server はパスをオン、アクティブ、スタンバイ、またはデッドとして表示します。

ファイル システムのフォーマット

次のファイル システム フォーマットのデータストアを使用できます。

- VMware Virtual Machine File System (VMFS) — ESX Server は、ローカル SCSI ディスク、iSCSI ボリューム、またはファイバ チャンネル ボリューム上でこのタイプのファイル システムを使用し、バーチャルマシンごとに 1 つのディレクトリを作成します。
- Raw Device Mapping (RDM) — ボリューム上の既存ファイル システムをサポートできます。バーチャルマシンは VMFS ベースのデータストアを使用する代わりに、RDM をプロキシとして使用し、raw デバイスに直接アクセスできます。
- NFS (ネットワーク ファイル システム) — ESX Server は NFS サーバ上の指定 NFS ボリュームを使用できます (ESX Server は NFS Version 3 をサポート)。ESX Server は NFS ボリュームをマウントし、バーチャルマシンごとに 1 つのディレクトリを作成します。クライアント コンピュータ上のユーザにとって、マウントされたファイルとローカルファイルの区別はつきません。

このマニュアルでは、最初の 2 つのファイル システム タイプについて説明します。VMFS および RDM です。

VMFS

VMFS はクラスタ型ファイル システムです。VMFS を使用すると、複数の物理サーバが共有ストレージを活用して、同じストレージに読み取りおよび書き込みを同時に実行できるようになります。VMFS では、オンディスク分散ロックングを使用して、複数のサーバが同時に同じバーチャルマシンをオンにしないようにします。物理サーバで障害が発生すると、各バーチャルマシンのオンディスク ロックが解除されるので、他の物理サーバ上でバーチャルマシンを再起動できます。

VMFS ボリュームは、SAN ボリュームおよびローカル ストレージを含め、32 の物理ストレージ上に展開できます。したがって、ストレージをプールして、バーチャルマシンに必要なストレージ ボリュームを柔軟に作成できるようになります。新しい ESX Server 3 LVM では、ボリュームを拡張しながら、同じボリューム上でバーチャルマシンを実行できます。これにより、バーチャルマシンでの必要性に応じて、VMFS ボリュームに新しいスペースを追加できます。

VMFS は ESX Server のインストール時に最初に設定します。VMFS 設定の詳細については VMware の『*Installation and Upgrade Guide*』および『*Server Configuration Guide*』を参照してください。



(注)

ESX Server Version 3 がサポートするのは VMFS Version 3 (VMFSv3) だけです。VMFS-2 のデータストアを使用している場合、そのデータストアは読み取り専用になります。VMFSv3 には、ESX Server Version 3 より前の ESX Server バージョンとの下位互換性はありません。

『*VI3 SAN System Design and Deployment Guide*』の「Upgrading Datastores」に記載されている手順に従うと、無停止型の方式で VMFS-2 データストアを VMFS-3 にアップグレードできます。次の URL にアクセスしてください。

http://www.vmware.com/pdf/vi3_san_design_deploy.pdf

VMFS の詳細については、『*VMware Virtual Machine File System: Technical Overview and Best Practices*』を参照してください。URL は次のとおりです。

<http://www.vmware.com/pdf/vmfs-best-practices-wp.pdf>

VMFS を使用する主な利点は、次のとおりです。

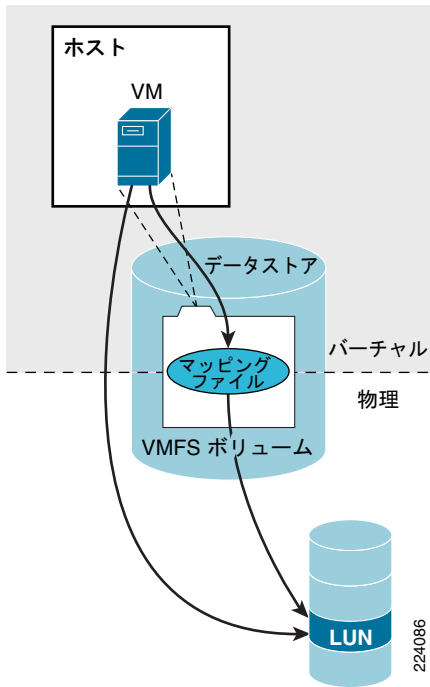
- VMFS は VMware ESX Server 独自の仕様であり、大型ファイルの保管およびアクセス用に最適化されています。大きいブロックサイズを使用しても、バーチャルマシンでネイティブ SCSI ディスクに近いディスクパフォーマンスが維持されます。単純なアルゴリズムでディスクをフォーマットできます。さらに、VMFS フォーマットのボリュームはオーバーヘッドが小さいので、VMFS ディスクが大きくなるほど、メタデータの保管に使用するスペースの割合が下がります。
- LUN の変更を自動的に検出する、再スキャン用組み込みロジックがあります。
- VMFS は分散ジャーナル処理、クラッシュ整合性バーチャルマシン入出力パス、マシンステートスナップショットなど、クラッシュに対するエンタープライズクラスの整合性および回復メカニズムも備えています。これらのメカニズムは、迅速に根本原因を分析し、バーチャルマシン、物理サーバ、およびストレージサブシステム障害から回復するうえで有効です。

Raw Device Mapping (DRM)

VMFS は RDM もサポートします。RDM はバーチャルマシンから物理ストレージサブシステム上のボリュームに (ファイバチャネルまたは iSCSI 限定で) 直接アクセスできるメカニズムを提供します。これは、Guest OS 上で動作するハイアベイラビリティクラスタ、SAN ベースのバックアップなどのアプリケーションに使用します。

RDM は、VMFS ボリュームから raw ボリュームへのシンボリックリンクを提供するものとして考えることができます (図 29 を参照)。マッピングによって、VMFS ボリュームではボリュームがファイルとして認識されます。バーチャルマシンコンフィギュレーションでは、raw ボリュームではなく、マッピングファイルを参照します。

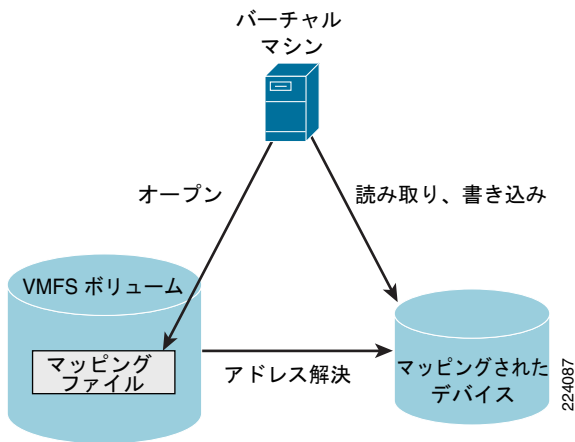
図 29 VMware DRM



ボリュームを開いてアクセスするときに、VMFS は RDM ファイルを適切な物理デバイスに割り当て、適切なアクセス チェックとロックを実行してから、ボリュームにアクセスします。したがって、読み取りおよび書き込みはマッピングファイルを経由するのではなく、raw ボリュームに直接送られます。

RDM ファイルにはディスク アクセスを管理し、物理デバイスにリダイレクトするために使用するメタデータが指定されます。RDM を使用すると、VMFS のバーチャルディスクの利点のある程度維持しながら、物理デバイスに直接アクセスする利点が得られます。実質的に、管理の容易な VMFS と raw デバイス アクセスを結合できます。図 30 を参照してください。

図 30 RDM によるデータ転送のリダイレクト



RDM を使用すると、次の作業が可能になります。

- VMotion を使用して raw ボリュームを使用しているバーチャル マシンを移行させる
- バーチャル インフラストラクチャ クライアントを使用しているバーチャル マシンに raw ボリュームを追加する
- 分散ファイル ロック、アクセス権、ネーミングなどのファイル システム機能を使用する。

RDM には 2 種類の互換モードを使用できます。

- **物理互換性** — ゲスト オペレーティング システムからハードウェアに直接アクセスできます。物理互換性は、バーチャル マシンで SAN 認識アプリケーションを使用している場合に便利です。ただし、物理互換性モードの RDM が設定されたバーチャル マシンをクローン化する場合、テンプレートにする場合、または移行にディスクのコピーが伴う場合に移行させることはできません。
- **バーチャル互換性** — RDM をバーチャル ディスクとして動作させることで、スナップショット、クローニングなどの機能を使用できるようになります。どちらを選択するかによって、その後、画面に表示されるオプションが異なります。



(注) VMware VMotion、VMware Dynamic Resource Scheduler、および VMware HA はいずれも、RDM 物理およびバーチャル互換性モードの両方でサポートされます。

大部分のバーチャル ディスク ストレージには VMFS を推奨しますが、場合によっては raw ディスクが必要です。バーチャル マシン間または物理マシンとバーチャル マシン間でクラスタを使用する、Microsoft Cluster Service (MSCS) 構成のデータ ドライブとして使用するのが、最も一般的です。クラスタ データおよび定数ディスクは、共有 VMFS 上で個々のファイルとしてではなく、RDM として設定する必要があります。



(注) VMware インフラストラクチャでサポートされる MSCS の詳細については、VMware の『*Setup for Microsoft Cluster Service*』を参照してください。URL は次のとおりです。
<http://www.vmware.com/support/pubs>

マルチパス化およびパス フェールオーバー

ファイバチャネルパスは、ルートを次のように記述します。

1. ホストの特定の HBA ポートから
2. ファブリックのスイッチを経由して
3. ストレージアレイの特定のストレージポートまで

1つのホストが複数のパスをたどって、ストレージアレイ上のボリュームにアクセスできる場合があります。ホストからボリュームへのパスを複数用意することをマルチパス化といいます。

デフォルトでは、VMware ESX Server システムが任意の一時点で、ホストから特定のボリュームに使用するパスは 1 つだけです。VMware ESX Server システムがアクティブに使用しているパスで障害が発生した場合は、使用可能な別のパスをサーバが選択します。組み込み ESX Server マルチパス化メカニズムで障害パスを検出し、別のパスに切り替えるプロセスをパス フェールオーバーといいます。パス上のコンポーネントのいずれかで障害が発生すると、パス障害になります。パス上のコンポーネントとしては、HBA、ケーブル、スイッチポート、ストレージプロセッサなどがあります。このサーバベースのマルチパス化方式は、SAN コンポーネント（すなわち、SAN アレイ ハードウェア コンポーネント）が使用する回復メカニズムによって、処理が完了するまでに最大で 1 分かかることがあります。

ESX のマルチパス化には、主要な動作モードとして次の 2 種類があります。

- Fixed (固定) — ユーザが優先パスを指定します。パスが失われた場合は、セカンダリ パスが使用されます。プライマリ パスが再び使用可能になると、プライマリ パスにトラフィックが切り替えられます。
- Most Recently Used (最終使用) (MRU) — 現在のパスで障害が発生した場合、ESX はセカンダリ パスを選択します。以前のパスが再び使用可能になっても、トラフィックはセカンダリ パスに切り替えられたままです。

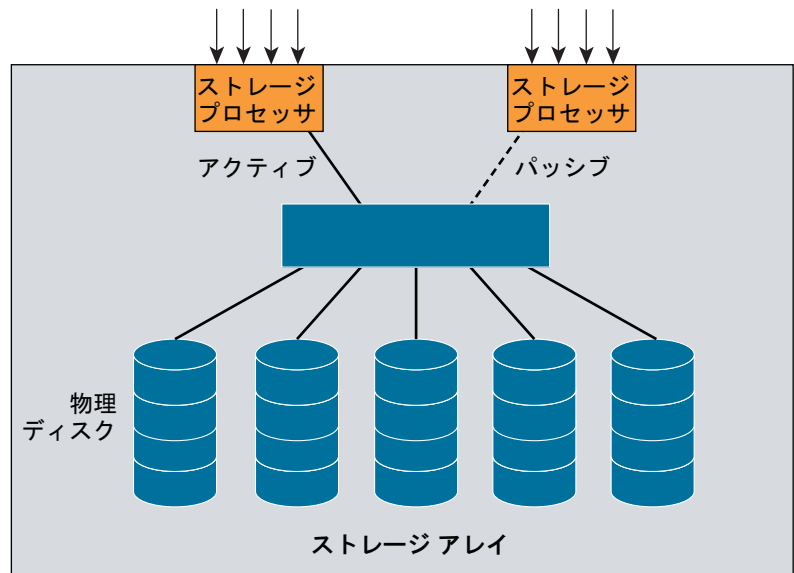
アクティブ/アクティブおよびアクティブ/パッシブのディスク アレイ

アクティブ/アクティブのディスク アレイとアクティブ/パッシブのディスク アレイを区別すると便利です。

- アクティブ/アクティブ ディスク アレイでは、使用可能なすべてのストレージ パスを使用して、ボリュームに同時にアクセスできます。パフォーマンスが大幅に低下することはありません。パス障害が発生しないかぎり、すべてのパスが常時アクティブです。
- アクティブ/パッシブ ディスク アレイでは、1つのストレージパスが特定のボリュームにアクティブに対応します。他方のストレージパスは、ボリュームのバックアップとして機能し、他方のボリューム入出力をアクティブに処理する場合があります。入出力の送信先はアクティブ プロセッサに限られます。プライマリ ストレージ プロセッサで障害が発生すると、自動的に、または管理者の介入によって、セカンダリ ストレージ プロセッサの 1 つがアクティブになります。

図 31 では、一方のストレージ プロセッサがアクティブで、他方がパッシブです。データが届くのは、アクティブ アレイからだけです。

図 31 アクティブ/パッシブストレージアレイ



224088

パス障害の検出

ブートまたは再スキャン処理時に、ESX Server はすべてのアクティブ/アクティブ ストレージアレイタイプに、*fixed* (固定) のパス ポリシーを割り当てます。固定パス ポリシーが設定されていると、パスが *ON* ステートの場合に、優先パスが選択されます。

ESX Server のマルチパス化ソフトウェアは、入出力要求の中止をバーチャル マシンに積極的に伝えることはありません。マルチパス化メカニズムで、現在のパスが機能しなくなっていることが検出されると、(バーチャル マシンにただちに入出力障害を戻すのではなく) ボリュームへの別のパスをアクティブにして、新しいパスにバーチャル マシンの入出力要求を再発行するプロセスを ESX Server が開始します。

アクティブ/アクティブ ストレージアレイタイプでは、ESX Server がパス フェールオーバーを実行するのは、FC 接続損失を表す `NO_CONNECT` の FC ドライバステータスで、SCSI 入出力要求が失敗した場合だけです。失敗したコマンドはチェック条件とともに、ゲスト オペレーティング システムに戻されます。パス フェールオーバーが完了すると、ESX Server が次の *ON* ステートになっているパスにコマンドを発行します。

アクティブ/パッシブ ストレージアレイタイプでは、ESX Server が MRU (最終使用) のパス ポリシーを自動的に割り当てます。 `NO_CONNECT` の `TEST_UNIT_READY` に対するデバイス応答および特定の SCSI チェック条件によって、ESX Server はすべての使用可能パスのテストを開始し、それらのパスが *ON* ステートかどうかを調べます。



(注)

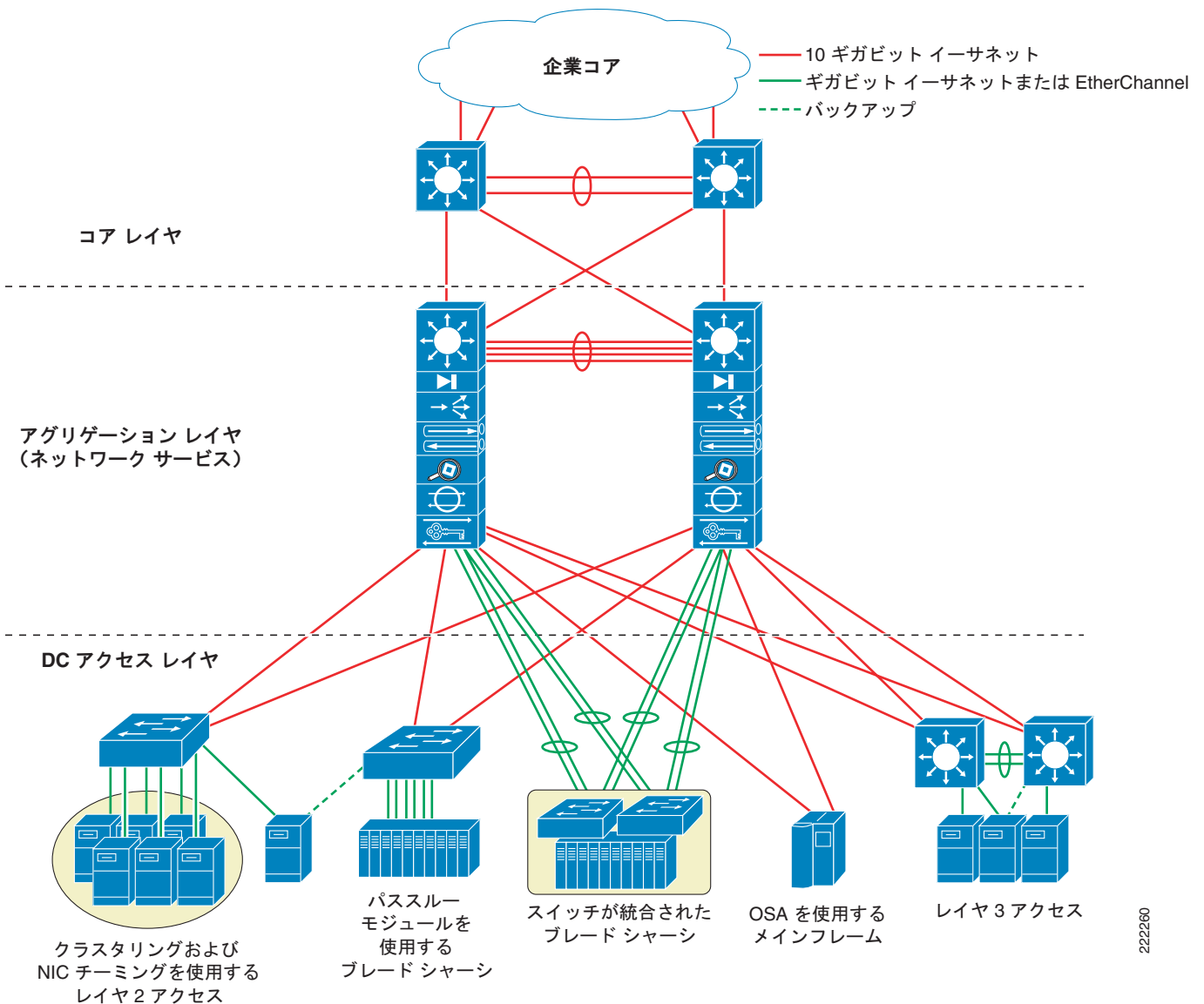
VMware SAN Compatibility リストに含まれていないアクティブ/パッシブ ストレージアレイの場合、アレイと ESX Server の完全な相互運用を実現するには、アクティブ/パッシブ アレイを手動で変更して MRU ポリシーを使用できるようにするだけでは十分ではありません。新しいストレージアレイが VMware の承認を受け、VMware の *SAN Compatibility Guide* に記載される必要があります。

ESX Server の接続およびネットワーク設計に関する考慮事項

LAN 接続

シスコのデータセンターアーキテクチャは、企業のサーバファームにスケーラビリティ、可用性、およびネットワーク サービスをもたらします。図 32 に、シスコデータセンター ネットワークの設計を示します。

図 32 シスコ データセンター アーキテクチャ



この設計は、主要な 3 つの機能レイヤ（コア、アグリゲーション、およびアクセス）からなり、次の機能を提供します。

- レイヤ 2/3 の要件（HSRP、STP によるハイ アベイラビリティ）サポート
- ハイ パフォーマンス マルチレイヤ スイッチング
- 複数のアップリンク オプション
- 統合物理インフラストラクチャ
- ネットワーク サービス（セキュリティ、ロード バランシング、アプリケーション最適化）
- スケーラブルなモジュラ設計

シスコ データセンター ネットワークの詳細については、次の URL を参照してください。
<http://www.cisco.com/go/datacenter/>

予備設計の考慮事項

ESX Server を LAN インフラストラクチャに接続するのは、主に LAN インフラストラクチャへの vSwitch の接続を設計する練習ともいえます。vSwitch の転送動作を考慮する必要があります。また、理解に不安がある場合は、（ESX ホスト構成に応じて）アクティブ / スタンバイ TIC チーミングまたはポートチャネリングを設定したサーバを前提に、LAN スイッチング インフラストラクチャを設計することを推奨します。

もう 1 つ、設計で重要なのは、Service Console および VMkernel による ESX ホストのネットワーク管理です。ネットワーク リンク障害が発生した場合でも、パフォーマンスとハイ アベイラビリティの両方が得られる構成にする必要があります。ESX ホスト ネットワーキング構成と Cisco LAN スイッチング テクノロジーを組み合わせることによって、この両方を実現できます。

vSwitch と LAN スイッチの比較

これまで、vSwitch が標準 LAN スイッチとどのように異なるかについて説明してきました。「ESX バーチャル スイッチ」(p.5) で、これらの特性を示しています。ここでは、vSwitch の最も重要な動作を示します。

- VM の宛先 MAC アドレスを使用し、VM によって生成されたトラフィックは、ローカル VM に送信されます。それ以外はすべて vmnic に送信されます。
- LAN スイッチング インフラストラクチャからのトラフィックに VM の宛先 MAC アドレスが指定されている場合は、VM に送信されます。それ以外は廃棄されます。
- 外部ブロードキャストおよびマルチキャスト トラフィックは VM にフラッディングされますが、他の vmnic にはフラッディングされません。
- VM で生成されたブロードキャストおよびマルチキャスト トラフィックは、単一の vmnic から送信されます。
- 同じ vSwitch および VLAN 内の VM 間トラフィックは、ローカルなままです。
- vSwitch はトランッキングに対応できます（ネゴシエーションプロトコルを使用しない 802.1q トランク）。
- vSwitch では、ポート グループ設定で指定された VLAN ID を使用して、VM からのトラフィックを色分けします。

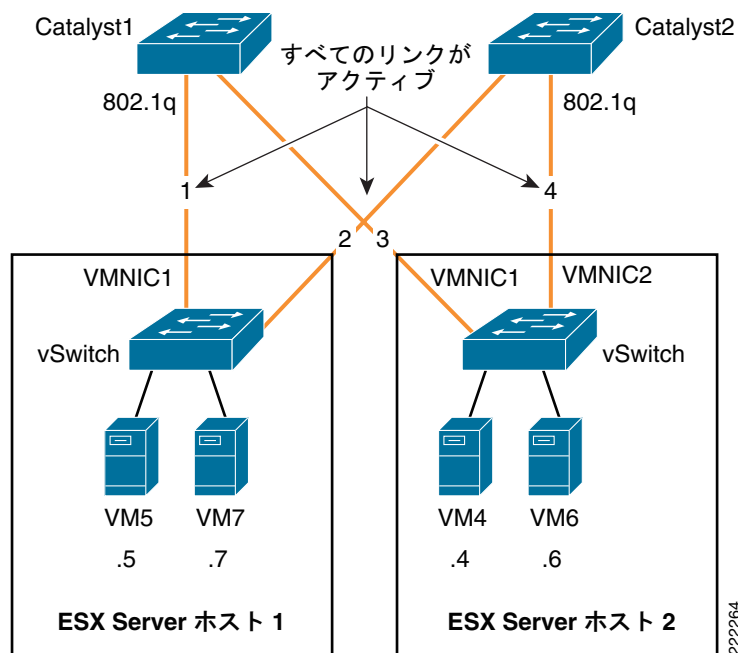
VLAN プロビジョニング

冗長構成で LAN に接続する vSwitch のネットワーク設計では、ある vSwitch NIC から同じレイヤ 2 ドメイン内の他の vSwitch へのレイヤ 2 パスを必ず確保する必要があります。このレイヤ 2 パスは、Cisco LAN ネットワークが提供する必要があります。vSwitch が提供することはできません。

vSwitch は LAN スイッチとよく似ているように見えますが、一部の転送特性およびループ回避に使用する技術について、可能な設計とそうでないものがあります。

vSwitch および VM に関して、ESX Server の内部を理解していない場合は、NIC チューニングを使用する標準サーバと同様に、LAN 接続を設計してください。2 つの NIC 間にレイヤ 2 の隣接関係が不可欠だという前提に立つ必要があります。図 33 の例で、その理由を示します。レイヤ 2 設計に関して、図 33 は「ループ」トポロジを示しています。つまり、各宛先に冗長レイヤ 2 パスがあります。vSwitch は、送信元 NIC にトラフィックが戻らないようにするなど、ディスタンス ベクトルテクノロジーの一部に関連した VMware 固有のメカニズムを使用して、ループフリーのトポロジを維持します。VM からのトラフィックは、すでに説明した NIC チューニングロード バランシング アルゴリズムに基づいて、両方の vmnic に負荷分散されます。

図 33 VLAN 設計



vSwitch および Cisco Catalyst スイッチによって作成されるネットワークを検討する場合、このトポロジでは、すべてがレイヤ 2 の隣接関係です。リンクはすべて、VM の VLAN を伝送する 802.1q トランクです。すべての VM が同じ VLAN にあるのはこのためです。

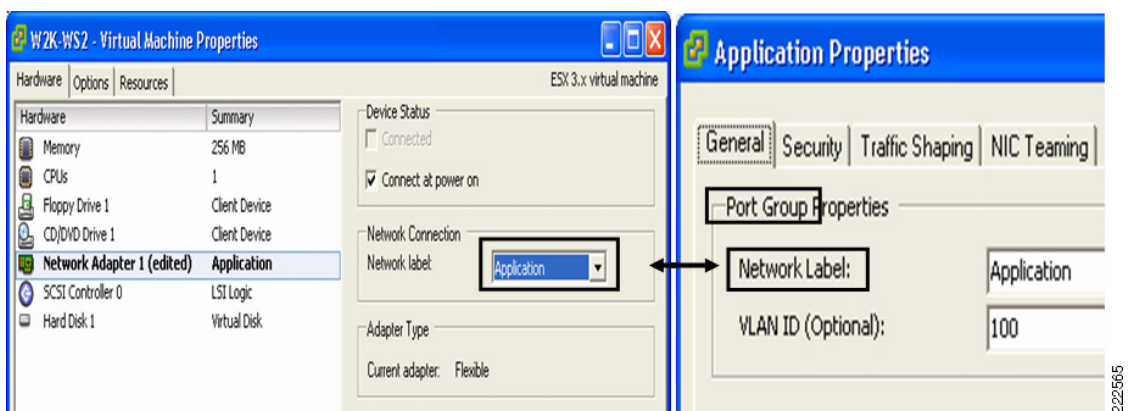
VM5 が VM7 と通信できるのに、VM4 とは通信できないこと納得できなかもしれません。VM5 と VM7 がどちらも vmnic1 にハッシュし、VM4 と VM6 がどちらも vmnic2 にハッシュすると考えてみてください。Catalyst1 は VM5 および VM7 の MAC アドレスをリンク 1 で学習し、Catalyst2 は VM4 および VM6 の MAC アドレスをリンク 4 で学習します。

VM5 が VM4 にトラフィックを送信する場合、Catalyst 1 はリンク 3 から、VM5 から VM4 へのトラフィックをフラディングする必要があります。ESX Server ホスト 2 の vSwitch は、ループ防止のために、このトラフィックを採用しません。この vSwitch はリンク 4 から届く、VM4 宛てのトラフィックを待ちます。この問題を解決するために必要なのは、VM の VLAN をトランクにするレイヤ 2 リンクで、Catalyst1 を Catalyst2 に接続することだけです。

図 37 の VLAN 設計は、問題を解消する LAN スイッチング トポロジ (U 型、後述) 全体の中の一部にすぎません。図 34 では、ネットワーク管理者が Cisco Catalyst スイッチを使用して、NIC カード間のレイヤ 2 隣接関係をプロビジョニングしました。

まとめると、VMware をサポートするネットワークを設計する場合、異なる vSwitch の VM が相互に通信できるように、vSwitch に依存しないレイヤ 2 冗長パスをプロビジョニングする必要があります。

図 34 VM の接続を保証する VLAN 設計



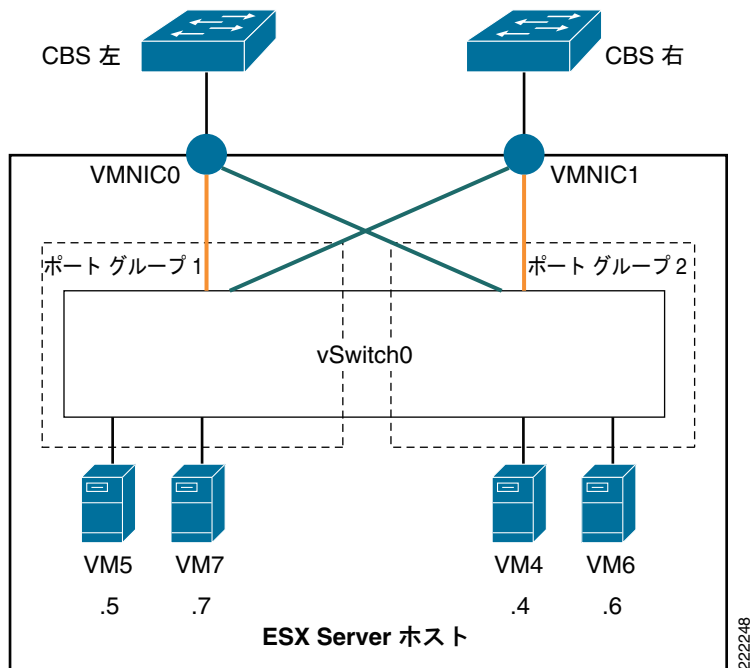
トラフィック ロード バランシング

ネットワーク接続を設計する場合に共通する要件の 1 つは、サーバからネットワークへの使用可能なリンクをすべて使用することです。一部の NIC チューニング設定では、これが本質的に可能です。

- アクティブ/アクティブ、MAC アドレスのハッシュに基づくロード バランシング
- アクティブ/アクティブ、起点バーチャル ポート ID のハッシュに基づくロード バランシング (推奨オプション)
- EtherChannel 化

いずれも、すでに説明しました。これらのメカニズムのほかに、ここで説明するように、アクティブ/スタンバイを使用することによって、発信および着信トラフィックのロード バランシングを実行することもできます。VM は、ポート グループ 1 に所属するものとポート グループ 2 に所属するものの、2 つのグループに分けられます。どちらのポート グループも同じ vSwitch を使用しますが、ポート グループ 1 はメイン NIC として vmnic0 を、スタンバイ NIC として vmnic1 を使用します。一方、ポート グループ 2 はメイン NIC として vmnic1 を、スタンバイ NIC として vmnic0 を使用します。図 35 に、この構成を示します。

図 35 トラフィック ロード バランシングを使用する ESX Server の冗長構成



この設計の利点は、任意の時点で障害が発生しなかった場合、VM5 と VM7 が vmnic0 を使用し、VM4 と VM6 が vmnic1 を使用することがネットワーク管理者に分かることです。



(注)

vSwitch から Cisco LAN スイッチング インフラストラクチャへのトラフィック ロード バランシングには、VLAN を 2 組使用する必要はありません。つまり、VM4、VM5、VM6、VM7 をすべて同じ VLAN 上に配置できます。必要なのは、2 種類のポート グループに分けることです。VM を 2 つのポート グループに分けても、VM5 および VM7 が VM4 および VM6 と通信する能力に影響はありません。



(注)

この設計は、Service Console および VMkernel の接続をプロビジョニングする場合に、検討に値するオプションです。2 つの NIC を使用すると、一方の NIC の帯域幅全体を Service Console および VMkernel に与え、他方の NIC の帯域幅全体を VM に与え、なおかつ冗長性を確保できます。一方の NIC で障害が発生した場合は、残った方がすべてのポート グループの接続を提供します。すなわち、Service Console、VMkernel、および VM です。

VMware 管理インターフェイスの割り当て

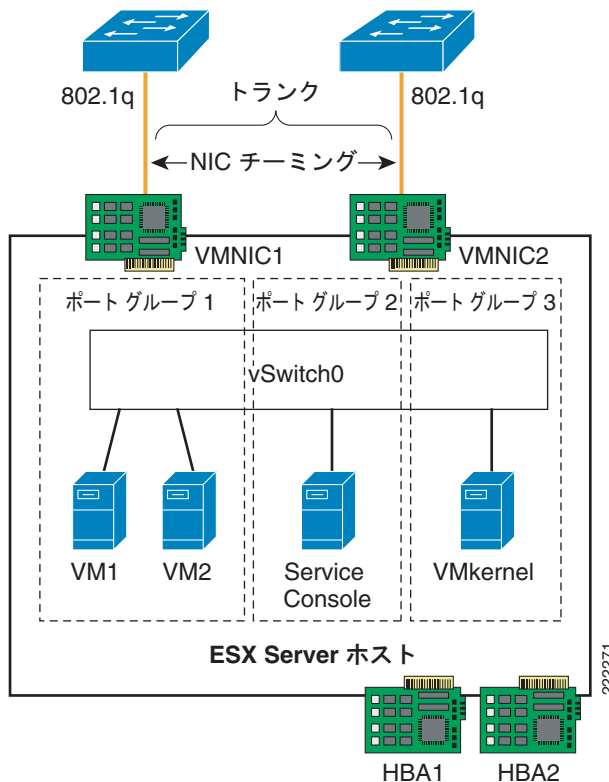
多くのサーバプラットフォームがありながら、使用できる NIC カードの数が限られていて（ブレードサーバで顕著）、なおかつ Service Console および VMkernel に冗長パスを提供しなければならない場合は、管理トラフィックと本稼働トラフィックで NIC を共有するようにサーバプラットフォーム上の ESX ホストを設定する必要があります。



(注) ブレードサーバベンダーによっては、使用できるイーサネットアダプタが2つまたは3つだけという場合もあります。したがって、ESX は 802.1q トランクを介した物理リソースの共有によって、ブレードサーバの配置をサポートします。

図 36 に、Service Console ポートグループ、VMkernel ポートグループ、および実稼働トラフィックが vmnic1 および vmnic2 を共有するネットワークの構成例を示します。

図 36 アダプタ リソースの共有



バーチャルスイッチは、VM を起点とするトラフィックのほかに、VMotion トラnsポートをサポートします。Service Console および VMkernel は、VM とは異なるポートグループで設定し、それぞれに対応する VLAN を与えます。VM VLAN、Service Console VLAN、および VMkernel VLAN で ESX Server の NIC ポートを共有します。

ネットワークリソースを共有すると、限られたリソースをめぐる競争が発生する場合がありますので、あらゆるトラフィックタイプにおいてパフォーマンスが影響を受ける可能性があります。たとえば、インターフェイスが実稼働トラフィックおよび管理トラフィックに割り当てられている場合、VMotion プロセスの所要時間が長引くことがあります。

vSwitch の設定に関する注意事項

次の注意事項は、ESX ホストで vSwitch を設定する場合に当てはまります。

- 複数の vSwitch を作成して VM トラフィックまたは Service Console/VMkernel トラフィックを分割する必要はありません。異なる VLAN ID を指定したポート グループを使用し、グローバル チーミング設定を上書きして、ポート グループ固有のトラフィック ロード バランシングを実装すればよいだけです。
- 推奨する 802.1q VLAN タギング メカニズムは VST (すなわち、ポート グループに特定の VLAN ID を割り当てる) です。
- ネイティブ VLAN を明示的 VLAN ID としては使用しないでください。ネイティブ VLAN を使用しなければならない場合は、関連ポート グループに対して VLAN ID = 0 を指定します。
- 推奨するトラフィック ロード バランシング メカニズムは、アクティブ / アクティブのバーチャル ポート ID ベースです。
- 可能な場合は、複数の異なる NIC チップセットにまたがる NIC チーミング構成を作成してください。
- ビーコン機能は推奨しません。
- Failback = Yes (ESX 3.5) または Rolling Failover = No (ESX 3.0.x) を推奨するのは、ブラックホール トラフィックの可能性がない場合に限られます。これはアグリゲーション スイッチ上の trunkfast (トランクファースト)、リンク ステート トラッキング、またはその両方を組み合わせて使用する必要があります。「[チーミング フェールバック](#)」(p.22) を参照してください。

アクセス ポートの設定に関する注意事項

このマニュアルでは、アクセス ポートは ESX Server の NIC カードに接続するポートを意味します。このポートは、スイッチポート アクセス タイプの設定にすることも、スイッチポート トランクの設定にすることもできますが、実際はほとんどの場合、スイッチポート トランクです。これは ESX Server が内部で複数の VLAN を使用するからです。

最も一般的な ESX Server 設計では、Virtual Switch Tagging (VST) 設定を使用します。それには Cisco LAN スイッチ ポートをトランクとして設定する必要があります。

VMware 固有の設計に関する考慮事項は、次のとおりです。

- スイッチポート トランクをトランクファーストとして設定します (すなわち、ポートがスパニング ツリー タイマーの満了を待たずに、ただちにフォワーディングになるようにします)。
- ESX では、VLAN ID = 0 (EST) を指定してネイティブ VLAN を使用する VM ポート グループを設定した場合を除き、スイッチポートのネイティブ VLAN に必ずタグが付きます。設定を簡素化し、トラフィックの孤立を回避するために、VM ではネイティブ VLAN を使用しないでください。VST コンフィギュレーションでネイティブ VLAN を使用しなければならない場合は、**vlan dot1q tag native** コマンドで Cisco LAN スイッチのネイティブ VLAN タギングをイネーブルにしてください。EST モード (VLAN ID = 0) で VM ポート グループを使用する場合は、Cisco Catalyst スイッチ上で特別な設定は不要です (**no vlan dot1q tag native** コマンドがイネーブルになります)。
- ポート セキュリティは推奨しません。あるスイッチポートから、同じスイッチまたは別のスイッチ上、さらに同じ VLAN 上の別のスイッチポートに、ポートを物理的に停止させずに VM MAC アドレスを移動させなければならないからです。

VST を使用するポート設定

VST によって、vSwitch はすべての出力トラフィックにタグを付け、逆にすべての入力トラフィックからタグを外すことができます。VST モードでは、802.1q トランクを使用する必要があります。次に示すスイッチポート設定では、カプセル化を 802.1q に設定し、ネイティブ VLAN および許可 VLAN を指定し、モードを（アクセスではなく）トランクに設定し、DTP（ダイナミック トランキング ネゴシエーション プロトコル）がイネーブルではないことを指定します。さらに、リンクアップと同時に、ポートがフォワーディング モードになることも指定します。

ESX では CDP（シスコ検出プロトコル）がサポートされ、どの物理スイッチ ポートが vSwitch vmnic に接続しているかが管理者に分かります。

```
spanning-tree portfast bpduguard default
!
interface GigabitEthernetX/X
  description <<** VM Port **>>
  no ip address
  switchport
  switchport trunk encapsulation dot1q
  switchport trunk native vlan <id>
  switchport trunk allowed vlan xx,yy-zz
  switchport mode trunk
  switchport nonegotiate
  no cdp enable
  spanning-tree portfast trunk
!
```

2 つの NIC を使用する ESX ホスト

アクティブ/アクティブ、すべての NIC を共有

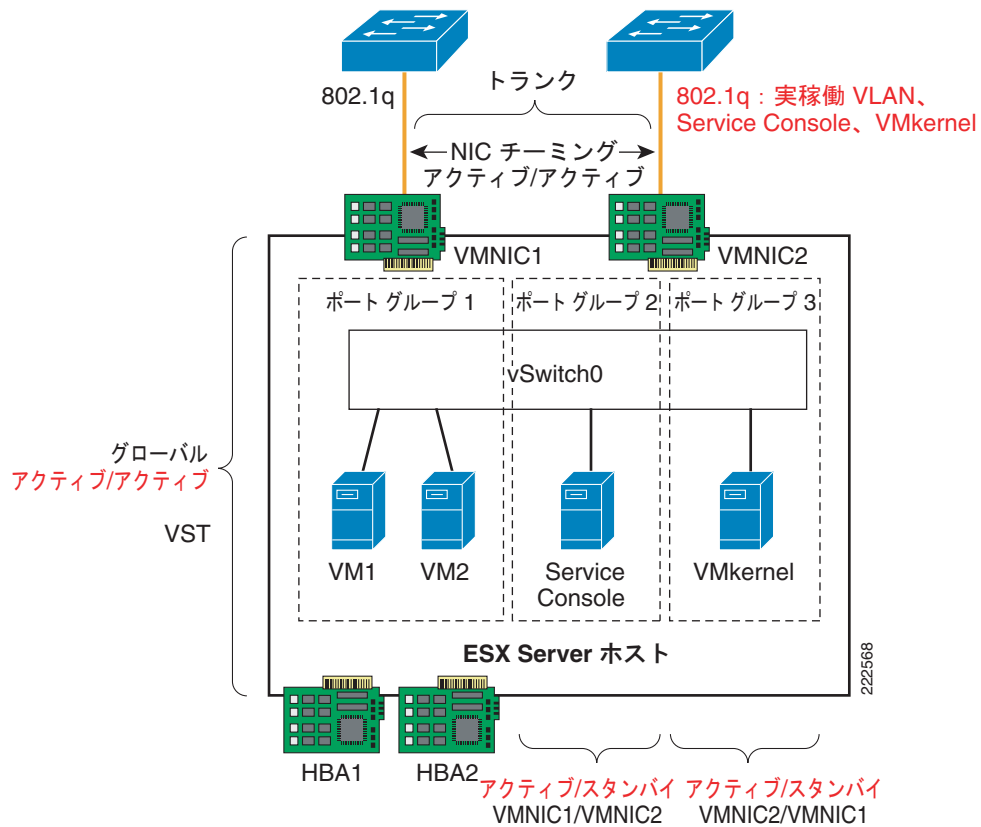
VMware の最も単純な設計では、ESX Server の物理 NIC が 2 つのアクセス レイヤ スイッチにまたがるように設定します。また、Service Console および VMkernel 用の NIC は分割し、ゲスト VM 実稼働トラフィック用の NIC は共有します。この設計では、NIC が ESX NIC チューニング構成に含まれます。図 37 に、この設計を示します。この設計では、アクティブ/アクティブの VM ポート ID ベースのロード バランシングを使用し、VM が 2 つの vmnic を共有します。802.1q VLAN 割り当て方式としては、VST を選択しています。Service Console ポート グループは、専用の VLAN 上にあり、vmnic1 を使用し、vmnic1 で障害が発生した場合は vmnic2 を使用するように設定されます。VMkernel ポート グループも同様に、専用の VLAN 上にあり、vmnic2 を使用し、vmnic2 で障害が発生した場合は vmnic1 を使用するように設定されます。

この方式では、すべての vmnic が使用され、完全な冗長接続が得られ、管理トラフィックは 2 つのアップストリーム vmnic に分散されます。Service Console および VMkernel のポート グループ チューニング設定によって、グローバルな vSwitch NIC チューニング設定が上書きされることに注意してください。

この設計の主な利点は、実稼働トラフィックが両方のアップストリーム vmnic に分散されることです。主な欠点は、実稼働トラフィックとのトラフィック共有によって、VMotion による移行が低速になる可能性があることです。

この設定の ESX Server は、アクセス レイヤ スイッチへのデュアル ホーミングです。物理スイッチ ポートを 802.1q トランクとして設定すると、vSwitch のさまざまな VLAN 上の VM でトラフィック 転送が可能になります。これらのポートをスパンニング ツリーに対するエッジ ポート（トランク ファースト）として設定すると、障害条件の発生時に、ネットワーク コンバージェンスが高速になります。VMotion プロセスは、VLAN によって仮想化された同じ物理インフラストラクチャを使用して、アクティブ VM を移行させます。

図 37 アクティブ/アクティブの NIC チーミングを使用する ESX Server vSwitch の設計

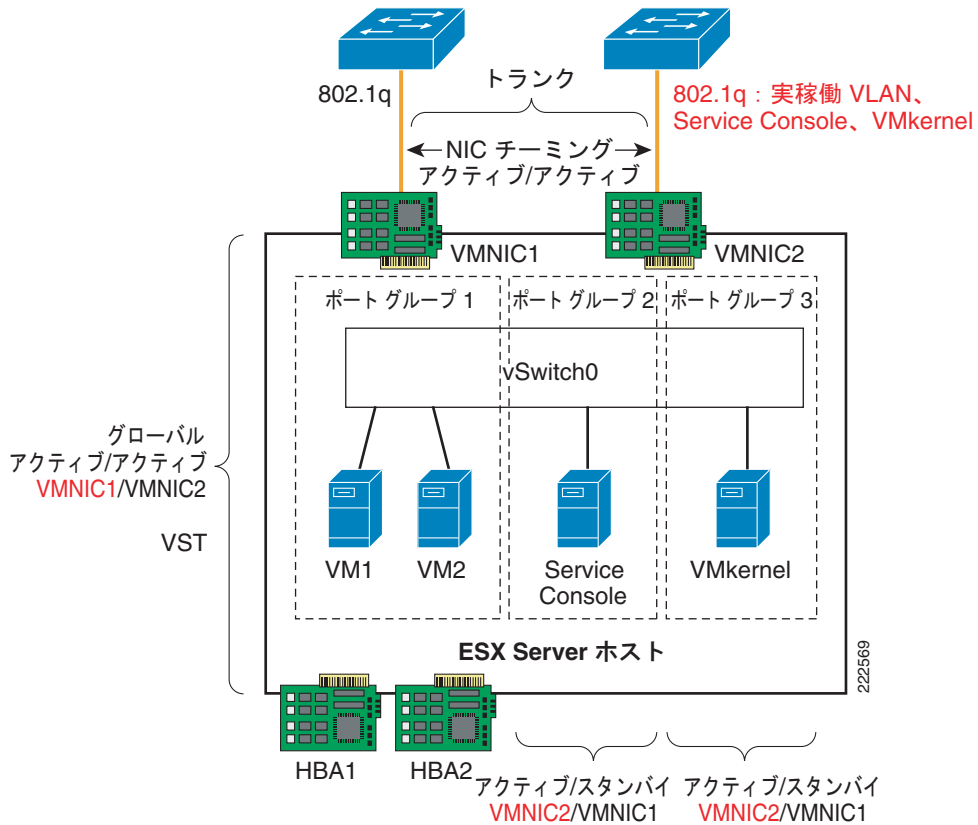


実稼働用および VMkernel/Service Console 用の冗長専用 NIC

VMkernel に十分な帯域幅を確保しなければならない場合は、前の項で説明したのとは多少異なる設計を使用できます。図 38 に、もう 1 つの設計を示します。この設計では、すべての VM がアクティブ NIC として vmnic1 を使用し、スタンバイ NIC として vmnic2 を使用します。同時に、Service Console ポートグループおよび VMkernel1 ポートグループは、アクティブ NIC として vmnic1 を使用し、スタンバイ NIC として vmnic2 を使用します。この方式では、VMotion 用の十分な帯域幅を VMkernel が必ず利用でき、vmnic2 で障害が発生した場合は、vmnic1 で実稼働トラフィックと VMkernel トラフィックの両方をサポートできます。前の設計と同様、vSwitch が Service Console トラフィック、VMkernel トラフィック、および VM トラフィックに VST 方式で VLAN を割り当てます。

この設計の主な利点は、VMkernel トラフィックに専用の vmnic が与えられ、この vmnic には実稼働トラフィックが存在しないことです。実稼働トラフィックは、アップストリーム vmnic を 1 つだけ使用します。

図 38 VMkernel および Service Console 用の専用アクティブ NIC



この設定の ESX Server は、アクセス レイヤ スイッチへのデュアル ホーミングです。アクセス ポートを 802.1q トランクとして設定すると、vSwitch のさまざまな VLAN 上の VM でトラフィック転送が可能になります。これらのポートをスパンニング ツリーに対するエッジ ポート（トランクファースト）として設定すると、障害条件の発生時に、ネットワーク コンバージェンスが高速になります。VMotion プロセスは、VLAN によって仮想化された同じ物理インフラストラクチャを使用して、アクティブ VM を移行させます。

クラシック アクセス レイヤ設計を使用するアクセス レイヤ接続

これまで説明した設計（図 37 および図 38 を参照）はどちらも、図 39 の設計タイプの Cisco LAN スイッチング インフラストラクチャに接続できます。ループフリーの U 型設計（スパンニング ツリー ブロッキング ポートなし）および V 型設計（アクセス スイッチごとにフォワーディング リンクが 1 つずつ、ブロッキング モードの 1 つまたは複数の冗長リンク）です。

図 40 を使用して、設計についてさらに説明します。この図は、U 型アクセス スイッチ Catalyst1 および Catalyst2 がどのように相互接続し、どのようにアグリゲーション レイヤに接続するかを示しています。2 つのアクセス スイッチを接続するダイレクト リンクが、vmnic に必要なレイヤ 2 の隣接関係を提供します。

Rapid PVST+ は、ポートがブロッキング ステートになっていなくても、高速スパンニング ツリー コンバージェンスを行います。ESX ホストはスパンニング ツリーを実行しないで、冗長方式で両方のアクセス スイッチに接続します。ESX ホストに接続するレイヤ スイッチ ポートは、すべて トランクであり、Service Console VLAN、VMkernel VLAN、および VM 実稼働 VLAN を伝送するように設定されています。

図 39 ESX Server の有効な VLAN 設計

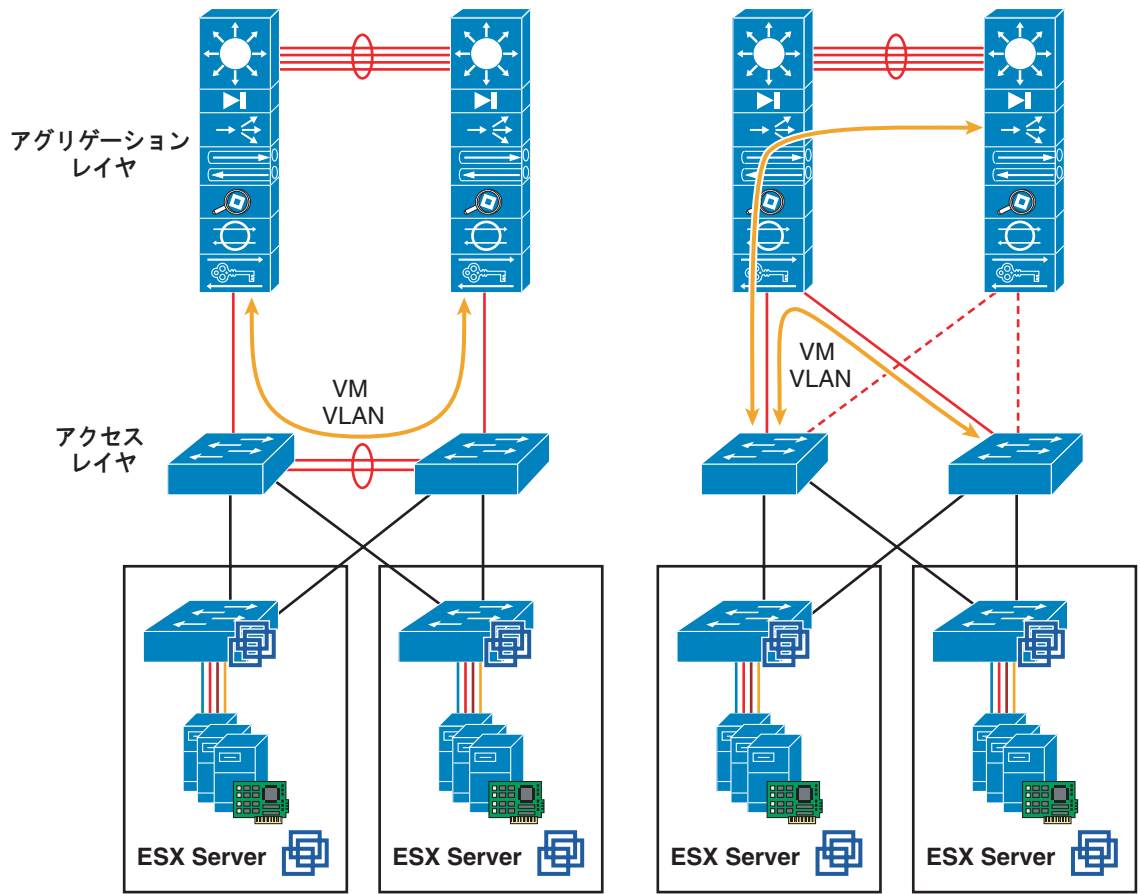


図 40 シスコ スイッチ U 型設計を使用する ESX Server

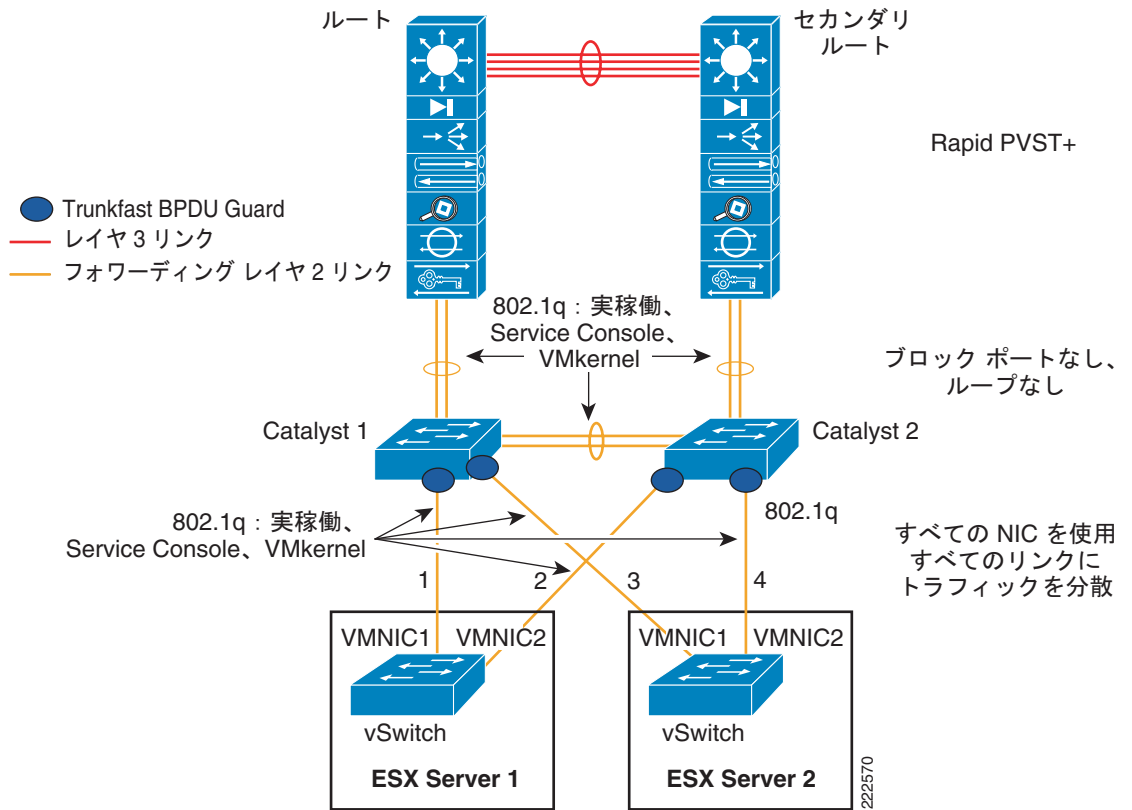
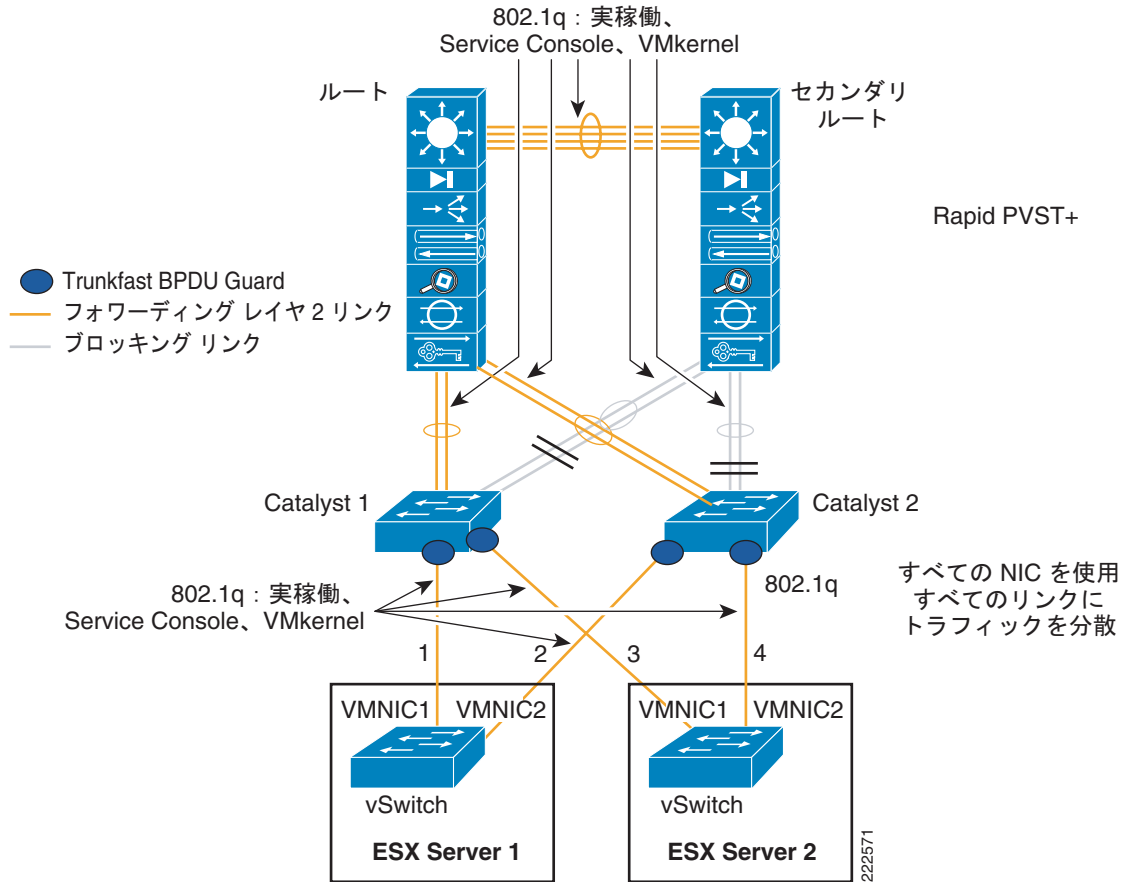


図 41 に、V 型の Cisco LAN スイッチング トポロジを使用する設計を示します。アクセス スイッチ Catalyst1 および Catalyst2 は、アグリゲーション レイヤ スイッチに二重接続されています。アグリゲーション レイヤは、ルート機能およびセカンダリ ルート機能を提供します。アクセス レイヤからの 2 組のリンクのうち、セカンダリ ルートに接続する方は、スパンニング ツリー ブロッキング ステートです。

前の設計と同様、ESX ホストに接続するアクセス ポートは、802.11 トランクとして設定されていて、Service Console VLAN、VMkernel VLAN、および VM 実稼働 VLAN を伝送します。

図 41 シスコ スイッチ V 型設計を使用する ESX Server



EtherChannel を使用するアクセス レイヤ接続

もう 1 つの設計では、vSwitch の両方の vmnic を EtherChannel で Cisco Catalyst スイッチに接続します。この EtherChannel がすべての VLAN をトランクにします。VM 実稼働 VLAN、Service Console VLAN、および VMkernel VLAN です。

Catalyst スイッチ側には、3 つのオプションがあります。

- 1 つのスイッチに、ステートフル スイッチオーバー (SSO) が可能なデュアル スーパーバイザを搭載して使用します。
- 2 つの Catalyst システムを仮想化 (クラスタ化) して単一スイッチとしてアクセスできるように設定した、Cisco Catalyst 6500 Virtual Switching System (VSS) を使用します。
- 仮想化 (クラスタ化) によって最大 8 つのブレード スイッチに単一スイッチとしてアクセスできるように設定された、Virtual Blade Switch (VBS) モードで Cisco Blade Switch (CBS) を使用します。
- 最大 8 つのスイッチを単一論理スイッチとして仮想化する、Cross Stack EtherChannel 構成で Catalyst 3750 を使用します。

SSO はアクティブ / スタンバイ方式で冗長スーパーバイザを使用し、レイヤ 2 ハイ アベイラビリティを実現します。スーパーバイザのスイッチオーバー時に、0 ~ 3 秒程度のパケット損失が発生します。スーパーバイザの冗長性を備えた単一アクセス レイヤ スイッチに ESX Server を接続すると、企業によっては十分なレベルの冗長性が得られます。

ESX Server のリンクをアクセス レイヤ スイッチに集約すると、サーバリソースの使用率向上を図ることができます。ESX 管理者は、送信元および宛先 IP アドレス情報に基づいて出力トラフィックのロードバランスを図るように、チームを設定できます。このアルゴリズムは、集約リンク間にバランスよく分散させることによって、ESX システム全体のリンク使用状況を改善します。

802.3ad リンクを使用すると、サーバアップリンクに関して、シングルポイント障害を取り除けるため、VM トラフィックがブラックホール化する可能性が低くなります。一方、アップストリームのスイッチが使用できなくなると、管理面と実稼働 VLAN の両方から ESX ホストが切り離されます。

Cisco 3750 プラットフォームなどのスタック可能なスイッチからなる、集約リンク EtherChannel 構成を使用することもできます。スイッチ スタックは論理上、単一スイッチであり、単一スイッチアクセス設計と同様の集約ポートおよび送信元 / 宛先 IP ロード バランシングを使用できます。

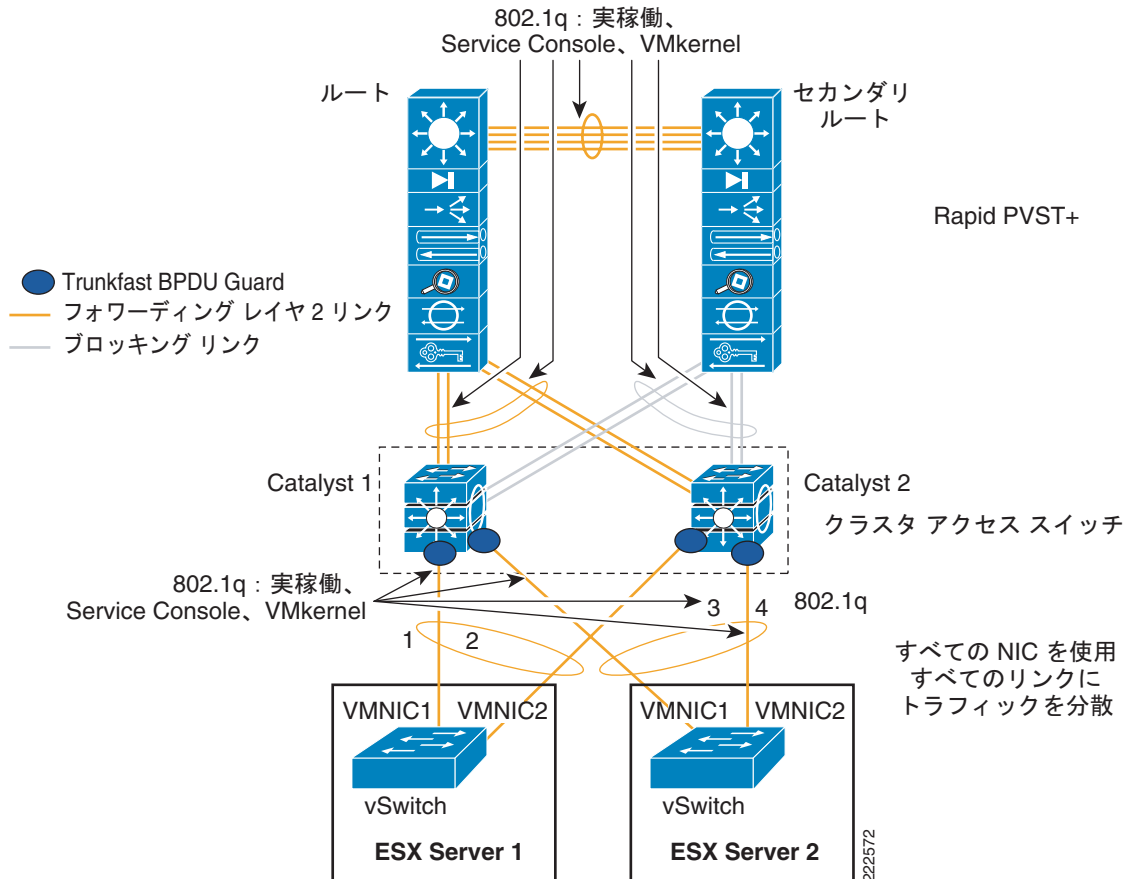
vmnic0 および vmnic1 へのチャンネルが設定された、2つのシスコ スイッチ アクセス ポートの設定例を示します。

```
interface GigabitEthernet1/0/14
  description Link to ESX vmnic0
  switchport trunk encapsulation dot1q
  switchport trunk allowed vlan 55,100,200,511
  switchport mode trunk
  switchport nonegotiate
  no mdix auto
  channel-group 5 mode on
  spanning-tree portfast trunk
end
!
interface GigabitEthernet2/0/14
  description to_ESX_vmnic1
  switchport trunk encapsulation dot1q
  switchport trunk allowed vlan 55,100,200,511
  switchport mode trunk
  switchport nonegotiate
  no mdix auto
  channel-group 5 mode on
  spanning-tree portfast trunk
end
```

channel-group <number> mode on コマンドを設定すると、ポートは LACP とのネゴシエーションを行わずに、強制的に EtherChannel を形成することになります。

図 42 に、スイッチ 1 上の 1 つのポート (GigabitEthernet1/0/x) とスイッチ 2 上の 1 つのポート (GigabitEthernet2/0/x) でチャンネルが設定された、クラスタ構成のアクセススイッチからなるトポロジを示します。アクセス スイッチへのすべてのリンクは 802.1q トランクとして設定されていて、VM 実稼働 VLAN、Service Console トラフィック、および VMkernel トラフィックを伝送します。

図 42 シスコのクラスタ スイッチを使用する ESX Server EtherChannel



アグリゲーション スイッチは EtherChannel でアクセス スイッチを接続し、1 つの EtherChannel (セカンダリ ルートへの EtherChannel) がスパニング ツリーのブロッキング モードです。

4 つの NIC を使用する ESX ホスト

4 つの NIC を使用する ESX ホストでは、2 つの NIC を使用する ESX ホストより接続オプションが増えます。Service Console に専用 NIC を 1 つ、VMkernel にも 1 つ割り当て、残り 2 つを VM 実稼働トラフィックに割り当てるのが自然です。このような構成は有効ですが、最適化が必要です。

実際のところ、冗長 Service Console 構成では、ESX ホストが管理から切り離される可能性があります。VMware HA を使用している場合は、関連 ESX ホスト上で VM がオフになり、別のホストで起動することになりますが、ESX ホストが VM 実稼働ネットワークに接続されたままの可能性もあるので、これは望ましくありません。

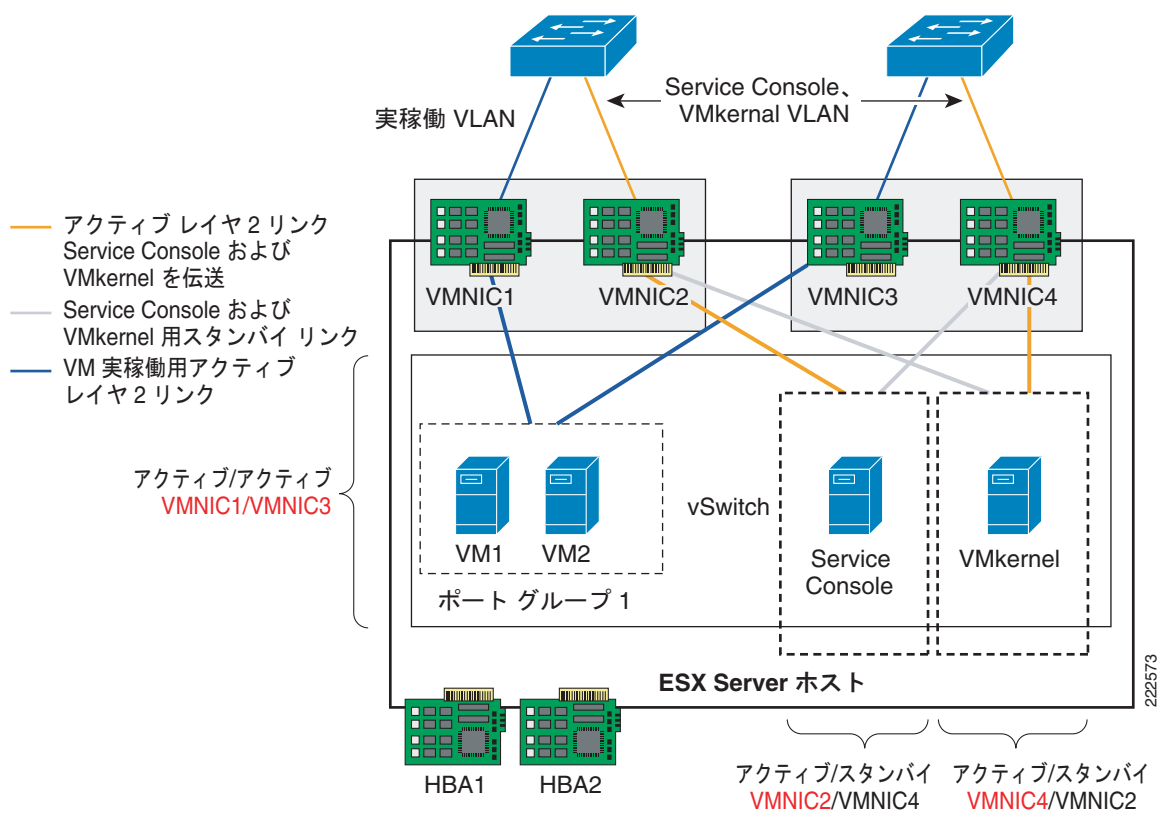
同様に、冗長構成にしないで 1 つの NIC を VMkernel 専用にとすると、iSCSI を使用している場合に回復が非常に困難になり、また、VM の VMotion 移行が不可能なので、問題が起きることがあります。

冗長性のために最適化された構成

ここで説明する構成では、管理から切り離されたり、VMware HA の偽アラームを引き起こしたりするシングルポイント障害が発生しません。この構成では、複数の異なるチップセットにまたがって NIC カードをチーミングすることによって、チップセットの問題に対する耐障害性も得られます。図 43 に、このような構成の例を示します。

VM ポート グループは、異なるチップセット上に配置された vmnic1 と vmnic3 の両方に接続されています。どちらの vmnic も、アクティブ/アクティブ バーチャルポート ID ベースロードバランシング対応として設定されています。Service Console ポートグループは、vmnic2 をアクティブ vmnic、vmnic4 をスタンバイ vmnic として使用するように設定されています。VMkernel ポートグループは、vmnic4 をアクティブ vmnic、vmnic2 をスタンバイ vmnic として使用するように設定されています。

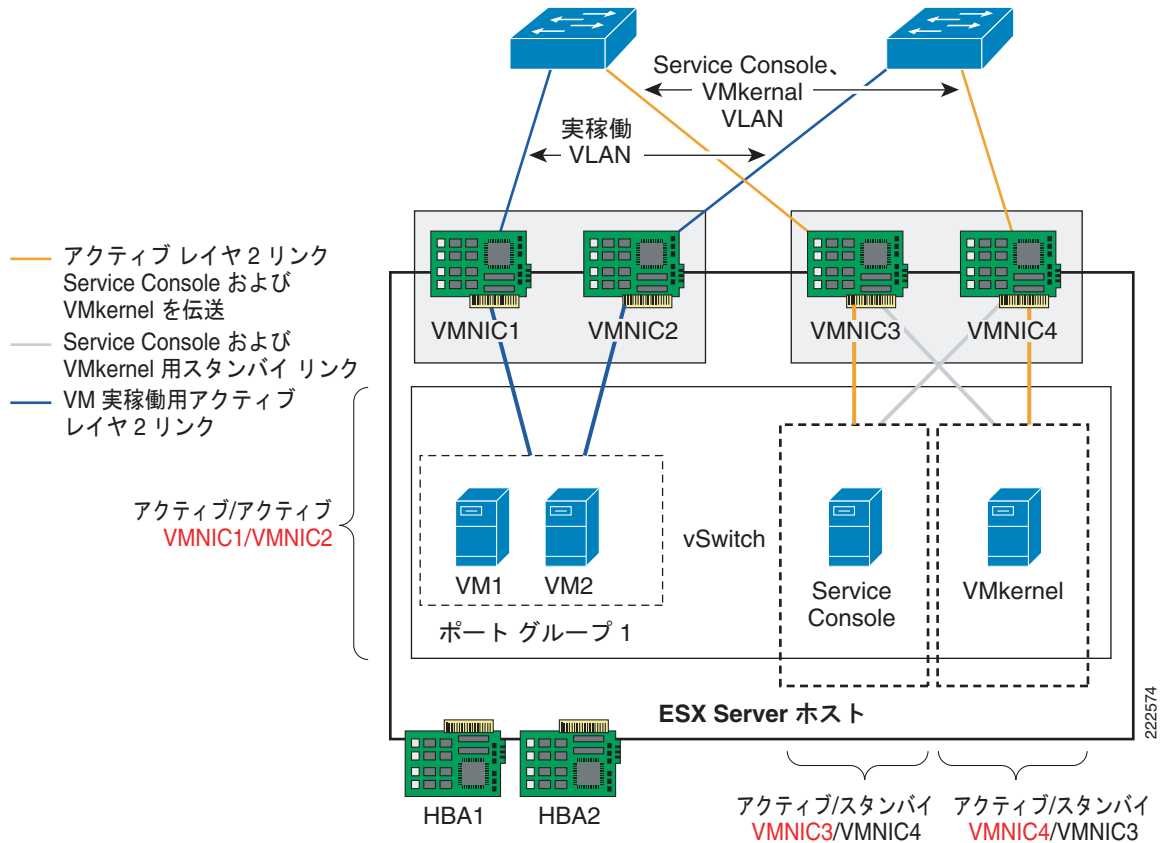
図 43 完全冗長構成の 4 NIC ESX Server



この構成では、vmnic1 および vmnic3 に接続する Catalyst スイッチのアクセスポートを VM 実稼働 VLAN の 802.1q トランッキング用に設定する必要があります。vmnic2 および vmnic4 に接続する Catalyst スイッチのアクセスポートは、Service Console および VMkernel VLAN の 802.1q トランッキング用として設定する必要があります。

サーバアーキテクチャのパフォーマンス上の理由から、単一 NIC チップセットによる NIC チーミング構成でのトラフィックロードバランシングを優先すべき場合があります。この場合は、[図 43](#) の構成を [図 44](#) のように変更できます。2 つの NIC のチップセットの 1 つを VM 実稼働トラフィックが使用し、2 つの NIC の他方のチップセットを Service Console および VMkernel トラフィックが使用します。

図 44 4 つの NIC およびチップセット単位の NIC チーミングを使用する ESX Server



クラシック アクセス レイヤ設計を使用するアクセス レイヤ接続

これまで説明した設計 ([図 43](#) および [図 44](#) を参照) はどちらも、[図 39](#) の設計タイプの Cisco LAN スイッチング インフラストラクチャに接続できます。ループフリーの U 型設計 (スパンニング ツリー ブロッキング ポートなし) および V 型設計 (アクセス スイッチごとにフォワーディング リンクが 1 つずつ、ブロッキング モードの 1 つまたは複数の冗長リンク) です。

[図 45](#) を使用して、設計についてさらに説明します。この図は、U 型アクセス スイッチ Catalyst1 および Catalyst2 がどのように相互接続し、どのようにアグリゲーション レイヤに接続するかを示しています。2 つのアクセス スイッチを接続するダイレクト リンクが、vmnic に必要なレイヤ 2 の隣接関係を提供します。

Rapid PVST+ は、ポートがブロッキング ステートになっていなくても、高速スパンニング ツリー コンバージェンスを行います。ESX ホストはスパンニング ツリーを実行しないで、冗長方式で両方のアクセス スイッチに接続します。VM 実稼働 VLAN 用として ESX ホストに接続するアクセス レイヤ スイッチ ポートはトランクであり、関連付けられた VLAN を伝送するように設定されています。管理ポートに接続するアクセス レイヤ スイッチ ポートは、Service Console VLAN および VMkernel VLAN のトランクとして設定されています。

図 45 シスコ スイッチ U 型設計を使用する ESX Server

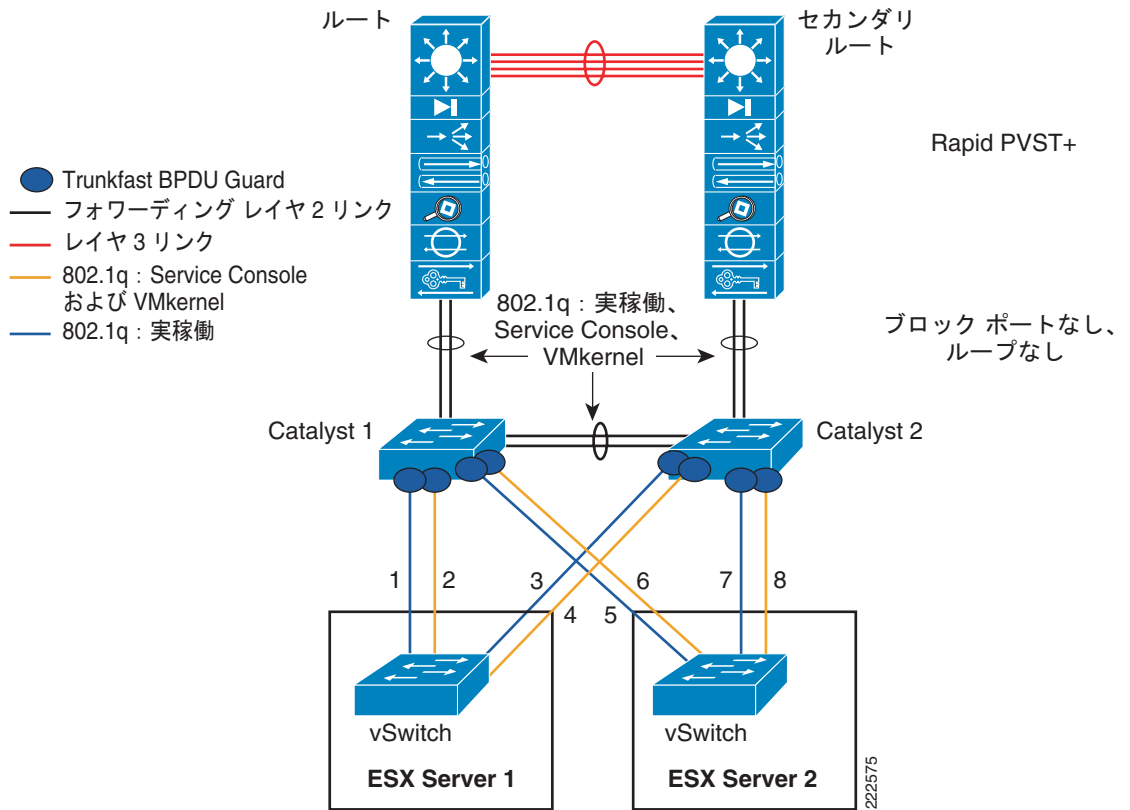
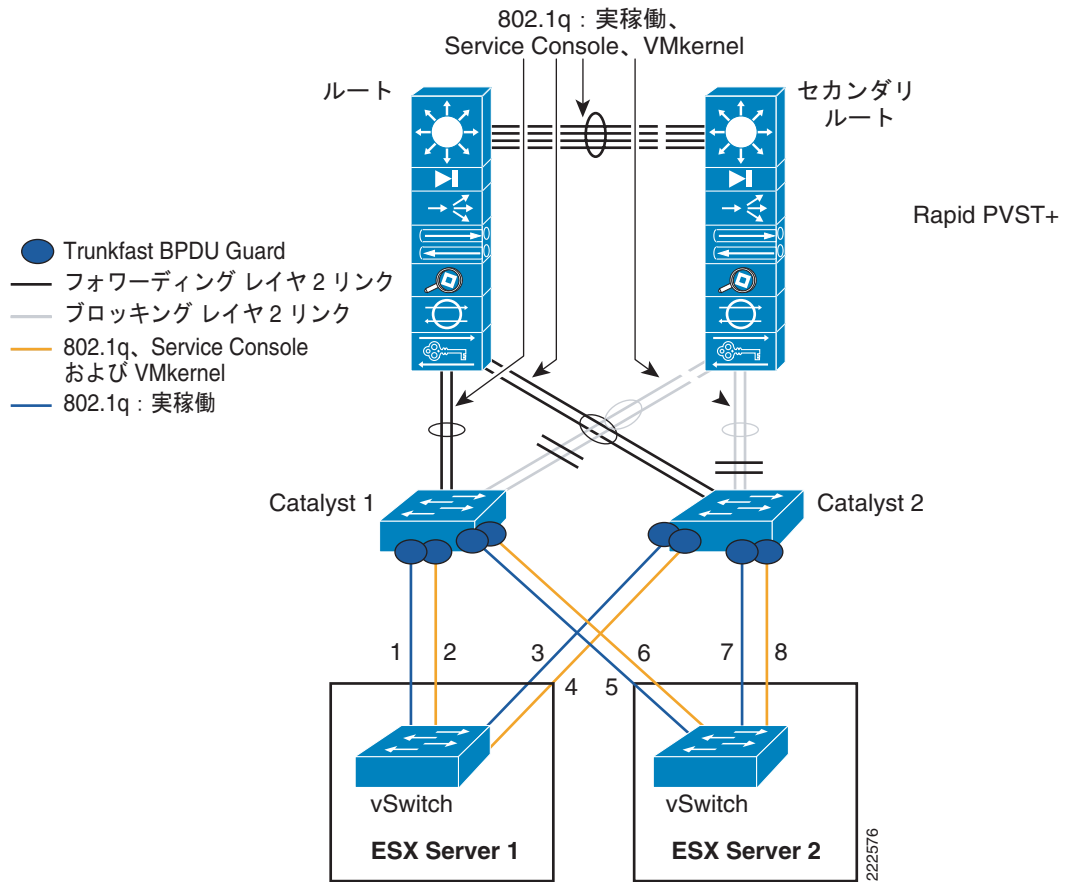


図 46 に、V 型の Cisco LAN スイッチング トポロジを使用する設計を示します。アクセス スイッチ Catalyst1 および Catalyst2 は、アグリゲーション レイヤ スイッチに二重接続されています。アグリゲーション レイヤは、ルート機能およびセカンダリ ルート機能を提供します。アクセス レイヤからの 2 組のリンクのうち、セカンダリ ルートに接続する方は、スパンニング ツリー ブロッキング ステートです。

前の設計と同様、ESX ホストに接続するアクセス ポートは、802.11 トランクとして設定されていて、Service Console VLAN および VMkernel VLAN または VM 実稼働 VLAN を伝送します。

図 46 シスコ スイッチ V 型設計を使用する ESX Server



EtherChannel を使用するアクセス レイヤ接続

ESX ホストの 4 つの NIC を使用できる場合、シスコ アクセス レイヤ スイッチで使用しているテクノロジーにもよりますが、さまざまな形で EtherChannel としてそれらの NIC を設定できます。図 47 に、次善のトポロジ例を示します。実稼働トラフィック用の 2 つの vmnic は EtherChannel で Catalyst1 に接続され、Service Console および VMkernel トラフィック用の 2 つの vmnic は EtherChannel で Catalyst2 に接続されます。このトポロジの主な問題は、1 台の Catalyst スイッチで障害が発生すると、両方の ESX ホストの動作が影響を受けることです。この例では、Catalyst 1 スイッチで障害が発生した場合、ESX server 1 が実稼働 VLAN 上で通信できないので、VM が孤立します。同時に、ESX2 が管理接続を失います。これは、VMware HA クラスタに含まれている場合は、ESX Server 2 上の VM がオフである可能性を示すことがあります。

図 47 サブオプションとしての ESX EtherChannel 構成

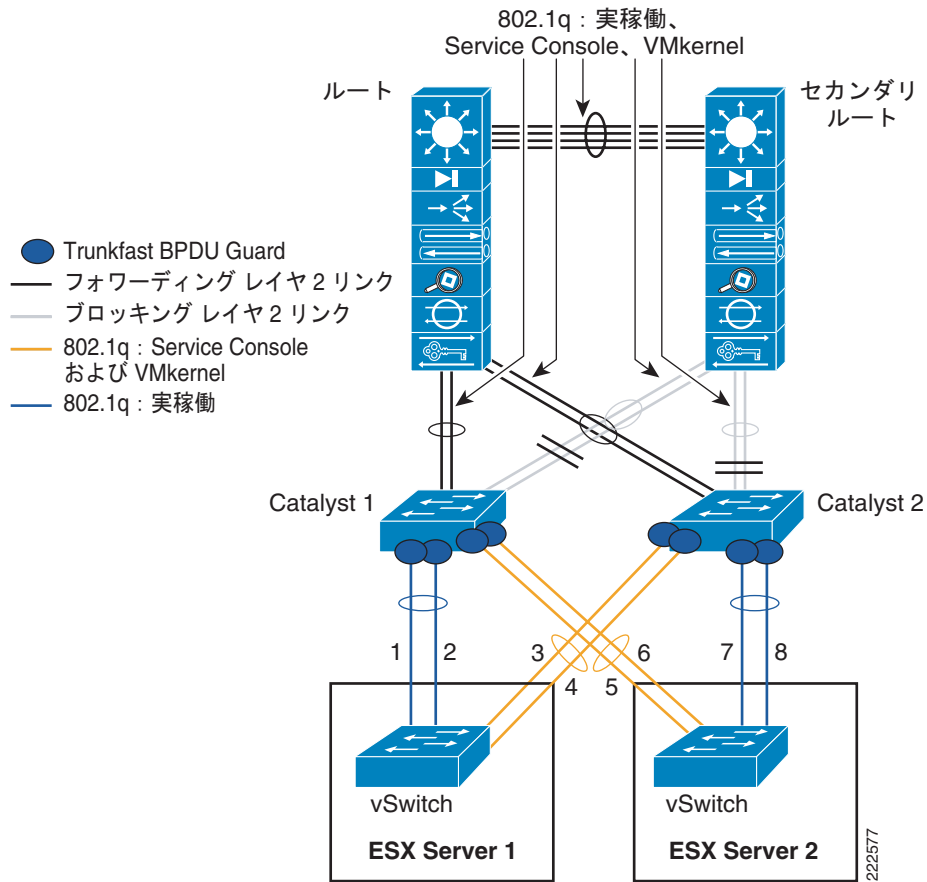
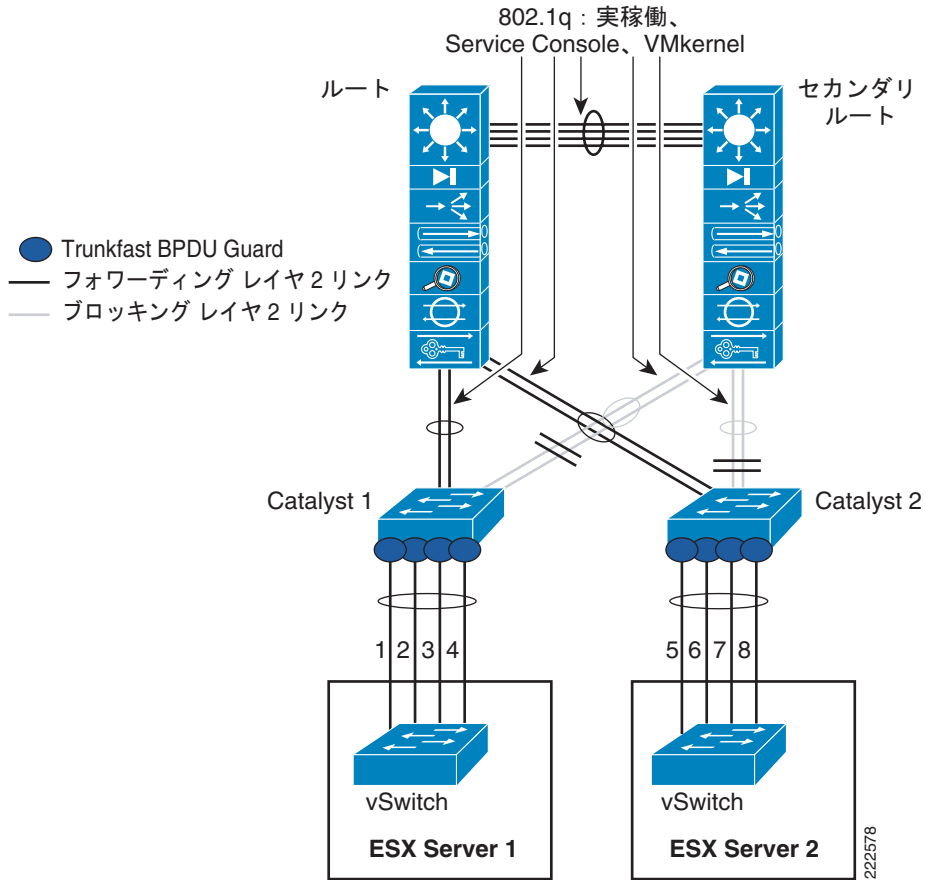


図 48 に、さらに適切なトポロジを示します。このトポロジでは、各 ESX ホストの接続先が 1 つの Catalyst スイッチ (SSO モードで 2 つのスーパーバイザを使用して動作可能) だけです。SSO はアクティブ/スタンバイ方式で冗長スーパーバイザを使用し、レイヤ 2 ハイアベイラビリティを実現します。スーパーバイザのスイッチオーバー時に、0 ~ 3 秒程度のパケット損失が発生します。スーパーバイザの冗長性を備えた単一アクセス レイヤ スイッチに ESX Server を接続すると、十分なレベルの冗長性が得られます。

ESX ホストと Catalyst スイッチ間の EtherChannel が VM 実稼働 VLAN、Service Console、および VMkernel VLAN を伝送します。2 種類の構成が可能です。すべてのリンク (1、2、3、および 4) で実稼働トラフィックと管理トラフィックを一緒に伝送するか、または 2 つの独立した EtherChannel を作成して、1 および 2 をバンドルして 1 つを実稼働用とし、3 および 4 をバンドルしてもう 1 つを管理用にするかです。

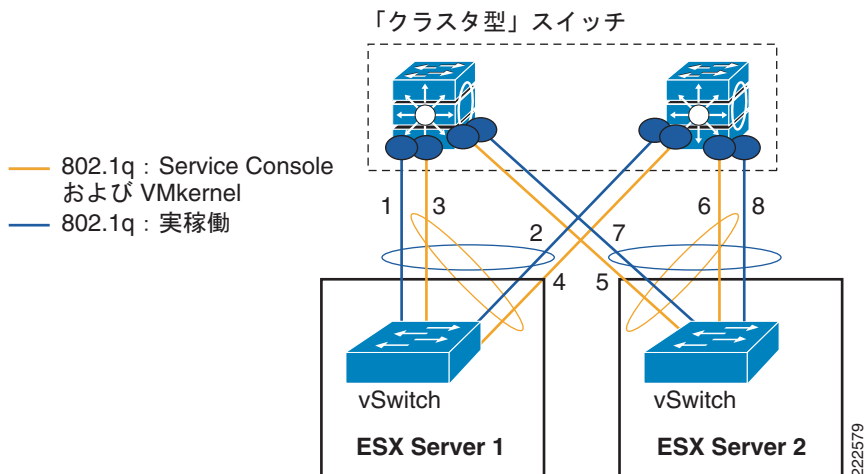
ESX Server1 と ESX Server2 間の VMware HA クラスタ構成では、Catalyst 1 または Catalyst 2 スイッチで障害が発生すると、2 つのサーバ間の HA ハートビート接続が切断されます。その結果、ESX Server1 は自分のバーチャルマシンを停止させ、ESX Server2 が自分のバーチャルマシンを立ち上げます。

図 48 Catalyst スイッチを使用する ESX EtherChannel 構成



最後に、図 49 に示したように、クラスタ型 Catalyst スイッチ間で EtherChannel を実行するように、ESX Server を設定することもできます。この図で示しているのは、構成のアクセススイッチの部分です。ESX Server 1 および 2 がクラスタの両方のスイッチに接続されています。

図 49 クラスタ型スイッチ（Cisco VSS、VBS テクノロジーなど）を使用する ESX EtherChannel 構成



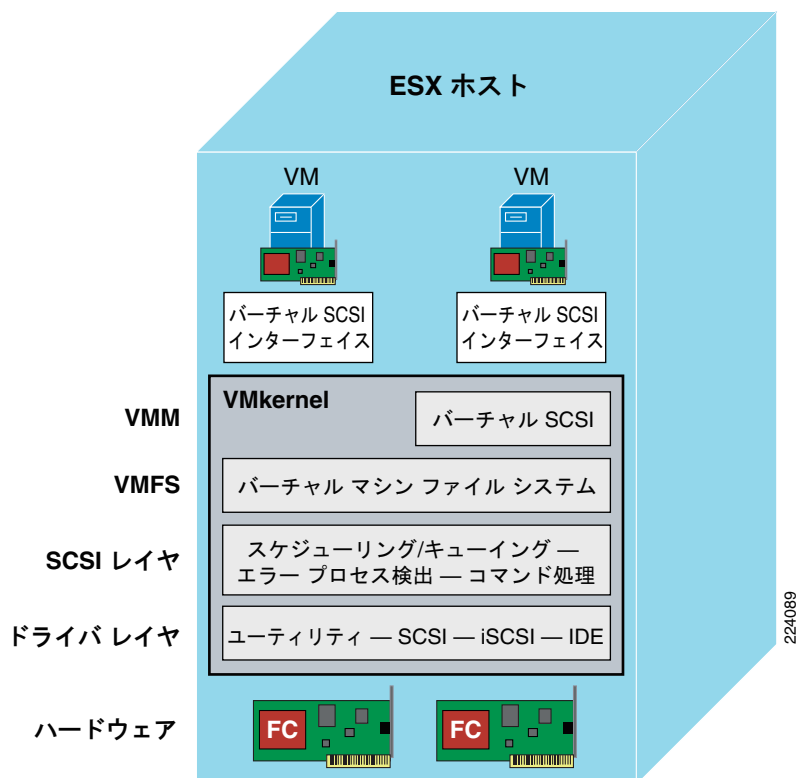
2 つの EtherChannel 構成が可能です。リンク 1、2、3、および 4 をバンドルした実稼働トラフィックと管理トラフィックの両方を伝送するか、または (図 49 に示したように) ワイヤ 1 および 2 のリンクを実現する 1 つの EtherChannel で実稼働トラフィックを伝送し、インターフェイス 3 および 4 のリンクを実現するもう 1 つの EtherChannel で管理トラフィックを伝送するかです。

SAN 接続

大規模な VMware 環境では、大部分でファイバ チャンネル接続を使用します。この場合、ESX ホストは SAN からブート可能であり、ゲストオペレーティングシステムはバーチャル SCSI コントローラを介してストレージにアクセスします。パケットフローは次のようになります (図 50 を参照)。

- バーチャルマシンからディスクに、読み取り / 書き込みコマンドを発行します。
- ゲストオペレーティングシステムドライバがバーチャル SCSI コントローラに要求を送信します。
- バーチャル SCSI コントローラが VMkernel にコマンドを送信します。
- VMkernel が VMFS 上の VM ファイルを検索して、物理ブロックにバーチャルブロックをマッピングして、物理 HBA ドライバに要求を送信します。
- HBA がワイヤ上に FCP 処理を送信します。

図 50 ESX ホストストレージアーキテクチャ



VMFS レイヤは、基礎のファイバチャンネルを形成し、ディスクアレイを同種のストレージスペースからなるプールに見えるようにします。VMFS は管理の軽減、LUN 集約、クラスタリングのための VM ロックなど、多数の有用なサービスを提供します。

FibreChannel の実装に関する考慮事項

ファイバチャネル環境に VMware を組み込む場合、設計面で 2 つの重要事項があります。1 つはゾーニングに関するものであり、もう 1 つはマルチパス化に関するものです。単一 ESX Server 上で動作するすべての VM は、同じ物理 HBA を使用してストレージにアクセスするので、ファイバチャネルネットワーク上では、唯一のポート ワールドワイド ネームを考慮したシングル ログインおよびゾーン分割が必要です。

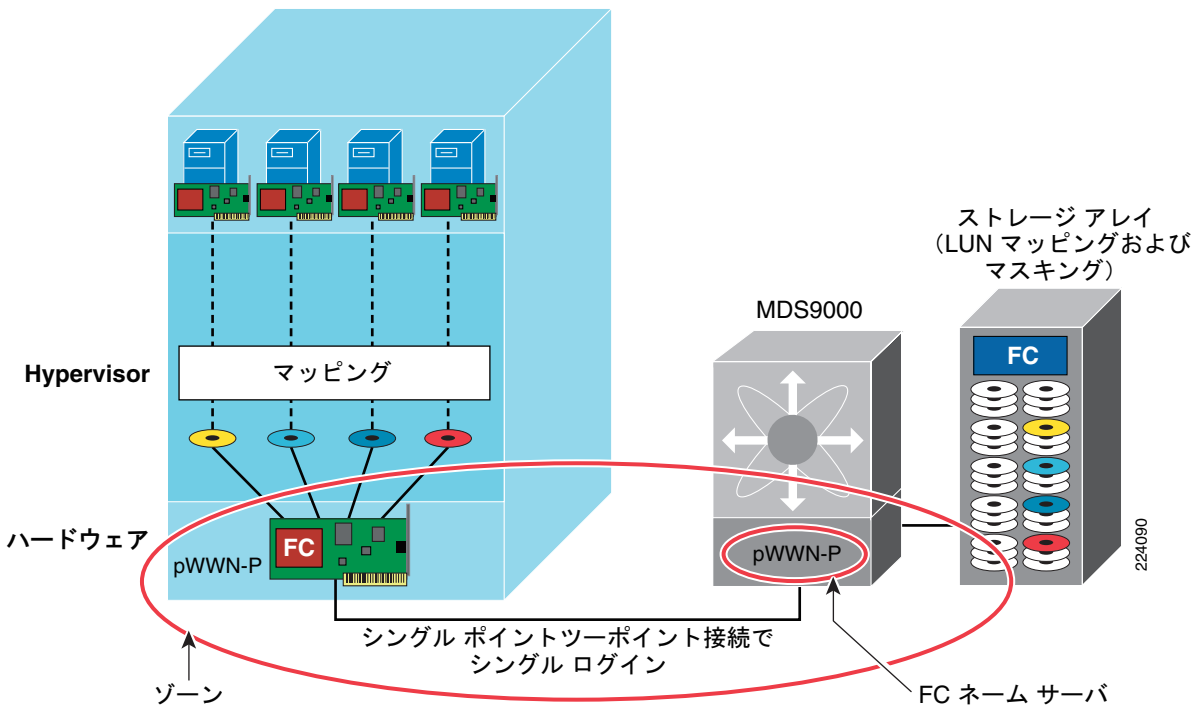
ゾーン分割および LUN マスキング

SAN のベストプラクティスは、ファイバチャネルで接続されたストレージエリア ネットワークにおいて、最良のデータ管理、保護、およびセキュリティが得られる共通モデルの使用を示しています。主要なプラクティスは次のとおりです。

- LUN (論理ユニット番号) マスキングは、ホストに応じて LUN 使用の可否を決定するプロセスです。LUN マスキングは通常、サーバの誤作動によるデータ破損からの保護に使用します。
- LUN マッピングは、物理 SCSI デバイス (ストレージ) の LUN と、オペレーティング環境に提示されるストレージの LUN 間の変換を意味します。

これらの SAN ベストプラクティスを実装するには通常、サーバとストレージ LUN 間のデータパスを指定する手段として、HBA WWPN (ワールドワイドポートネーム) を使用する必要があります。これを作成すると、このデータパスがサーバに組み込まれた HBA の物理 WWPN に記述されます。物理 HBA ポートのワールドワイドネームは、すべてのバーチャルマシンで同じです。図 51 にこれを示します。

図 51 ESX ファイバチャネル構成



ESX Server をディスク アレイにゾーン分割し、LUN マスキングを適切に設定する以外に、どの ESX ホストを同じクラスタに含めるかを検討する必要があります。VMotion 移行を行うことができるのは、ある ESX ホストから同じクラスタ内の別の ESX ホストに対してです。これには、両方のホストをゾーン分割して、同じストレージを見せる必要があります。ファイバチャネルに関しては、このゾーン分割構成は、同じクラスタ内のすべての ESX ホストがすべての LUN を認識できるという点でオープンであり、VMFS がデータ破損を防止するためのオンディスク分散ロック メカニズムを提供します。



(注)

VM が認識できるものについては、NPIV (N-Port ID Virtualization) でさらにきめ細かく制御します。

マルチパス化

EMC Powerpath のような製品を VM にインストールすることはできません。マルチパス化がゲストオペレーティング システムによって行われるのではなく、SCSI ミッドレイヤで行われるからです。ESX Server はストレージアレイへの使用可能なすべてのパスを自動的に識別し、使用できるパスの数に関係なく、折りたたんで 1 つのアクティブ パスにします。その結果、VMware ESX ホストは唯一のアクティブ パスとプライマリ パスが消失した場合のフェールオーバー メカニズムを提供します。

NPIV

VMware ESX Server 3.5 で使用できるようになった新しいテクノロジーの 1 つは、NPIV 規格のサポートです。NPIV は Emulex と IBM によって作成された T11 ANSI 規格であり、ファブリック スイッチが同一物理 HBA ポート上で複数の WWPN を登録できるようにします。

NPIV サポートによって、VMware ESX Server 上の各バーチャル マシン (VM) は、固有のファイバチャネル WWPN を使用し、バーチャル HBA ポートから SAN への独立したデータ パスを提供できるようになります。固有のバーチャル HBA ポートを提供することによって、ストレージ管理者は個々のバーチャル マシンに対して、LUN マスキング、ゾーン分割などの SAN ベスト プラクティスを実装できます。

機能

バーチャル マシンに WWN が割り当てられると、そのバーチャル マシンのコンフィギュレーション ファイル (.vmx) がアップデートされ、WWPN と WWNN (ワールドワイド ノードネーム) からなる WWN ペアが組み込まれます。そのバーチャル マシンがオンになると、VMkernel が LUN アクセスに使用する物理 HBA 上のバーチャル ポート (VPORT) をインスタンス化します。VPORT はバーチャル HBA ですが、ファイバチャネル ファブリックには物理 HBA として認識されます。したがって、バーチャル マシンに割り当てられた固有の識別情報である WWN ペアを使用します。各 VPORT はバーチャル マシンに固有であり、バーチャル マシンがオフになると、その VPORT はホスト上で破棄され、ファイバチャネル ファブリックで認識されなくなります。

要件

NPIV には次の要件があります。

- NPIV を使用できるのは、RDM ディスクのあるバーチャル マシンだけです。標準バーチャル ディスクが設定されたバーチャル マシンでは、ホストの物理 HBA の WWN を使用します。RDM の詳細については、『*ESX Server 3 Configuration Guide*』または『*ESX Server 3i Configuration Guide*』を参照してください。
- ESX Server ホスト上の物理 HBA は、ホスト上で動作するバーチャル マシンのアクセス先となるすべての LUN に対して、アクセス権が必要です。
- ESX Server ホストの物理 HBA は、NPIV をサポートする必要があります。現在、NPIV をサポートする HBA ベンダーおよび HBA タイプは、次のとおりです。
QLogic — あらゆる 4 GB HBA、Emulex — NPIV 対応ファームウェアが組み込まれた 4 GB HBA
- WWN が割り当てられたバーチャル マシンまたはテンプレートをクローン化した場合、クローンでは WWN が維持されません。
- ファイバチャネルスイッチが NPI を認識する必要があります。
- WWN が割り当てられたバーチャル マシンを操作するには、必ず、VI クライアントを使用します。

WWN の割り当て

RDM ディスクを使用する新しいバーチャル マシンを作成するときに、そのバーチャル マシンに WWN を割り当てることができます。または、一時的にオフにできるバーチャル マシンであれば、既存のバーチャル マシンに WWN を割り当てることができます。

RDM ディスクを使用するバーチャル マシンを作成するには、Virtual Machine 設定を表示して、SCSI アダプタのタイプを選択するだけです。そこで、**Raw Device Mapping** を選択してください。SAN ディスクまたは LUN のリストから、バーチャル マシンが直接アクセスできる raw LUN を選択します。さらに、既存のデータストアを選択するか、新しいデータストアを指定します。

compatibility mode メニューで、物理またはバーチャルのどちらでも選択できます。

- **物理互換性** — ゲスト オペレーティング システムからハードウェアに直接アクセスできます。物理互換性は、バーチャル マシンで SAN 認識アプリケーションを使用している場合に便利です。ただし、物理互換性モードの RDM が設定されたバーチャル マシンをクローン化する場合、テンプレートにする場合、または移行にディスクのコピーが伴う場合に移行させることはできません。
- **バーチャル互換性** — RDM をバーチャル ディスクとして動作させることで、スナップショット、クローニングなどの機能を使用できるようになります。どちらを選択するかによって、その後、画面に表示されるオプションが異なります。

WWN は *Specify Advanced Options* ページから割り当てることができます。このページのメニュー *To assign or modify WWNs* で、バーチャル デバイス ノードを変更できます。

ゾーン分割

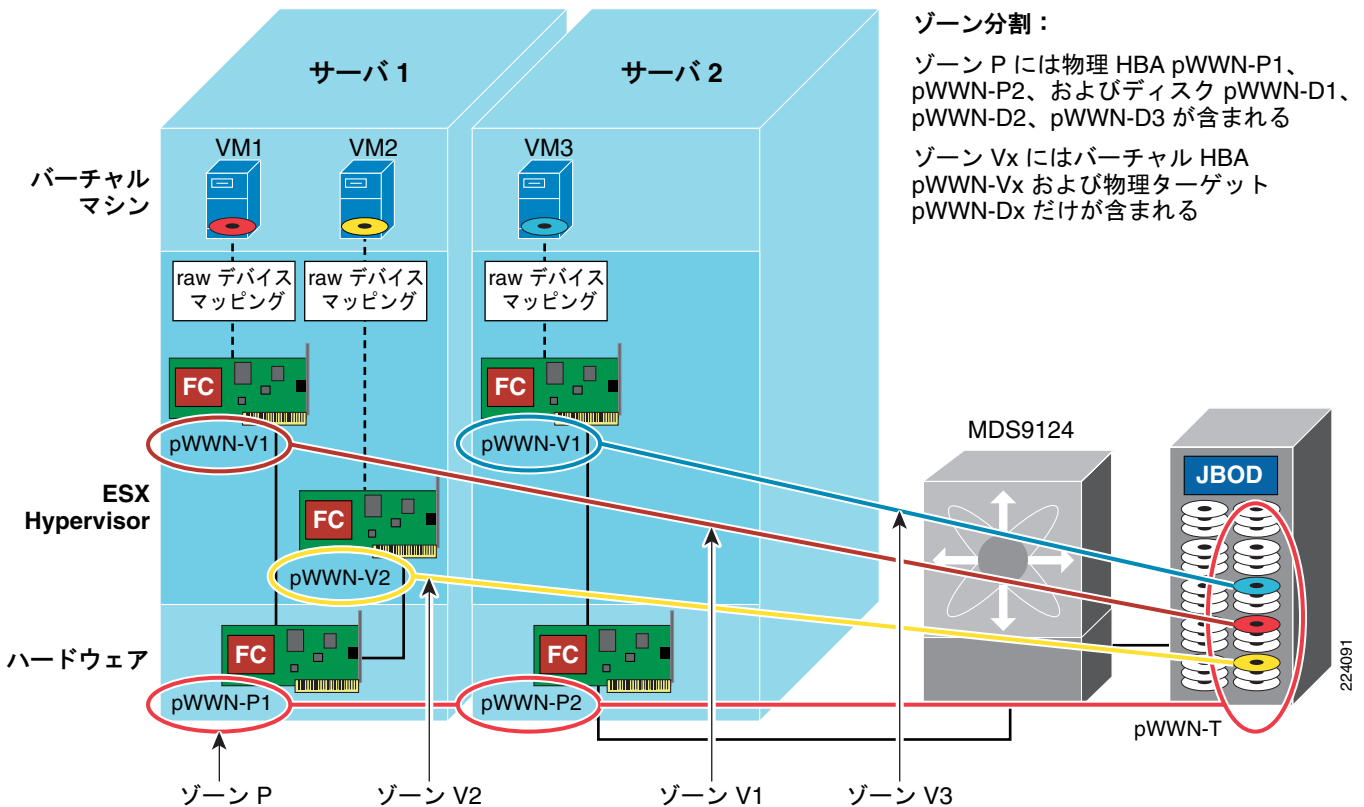
このマニュアルの作成時点では、ゾーン分割を行うには、HBA WWPN (ワールドワイド ポートネーム) を使用してデータ パスを指定する必要があります。ESX 3.5 では、ゾーン分割を使用して、同じ ESX Server 上で動作している VM を相互に切り離すことができます。すべての物理 HBA およびアレイ コントローラを VMkernel に認識させなければならないので、アレイ ポートには次の 2 つのゾーンが必要です。

- 現用ゾーン。VM にリンクされたバーチャル ポートおよび VM が使用する LUN のアレイ ポートが含まれます。
- コントロールゾーン。ESX Server 上のすべての物理 HBA およびその ESX Server に接続されたすべてのアレイ ポートが含まれます。

さらに、ゾーン分割および VMotion では、コントロールゾーンに移行先になる ESX Server のすべての物理 HBA ポートを含める必要があります。

図 52 に、2 つのサーバ上の 3 つの VM を示します。ゾーン分割を使用して、ユーザ、アプリケーション、または部門別に分離します。この場合、個々のディスクは関連仮想イニシエータとともに、個々のゾーンに配置します。ゾーン P には、物理デバイスが含まれます。物理 HBA (pWWN-P1、pWWN-P2) およびディスク (pWWN-D1、pWWN-D2、pWWN-D3) です。各ゾーン Vx には仮想 HBA (pWWN-Vx) および物理ディスク Dx (pWWN-Dx) だけが含まれます。

図 52 NPIV を使用する仮想マシン



(注) VMkernel は、システム上のすべての仮想マシンが使用するすべての LUN に対して、可視性が必要です。実用上の理由で、物理 HBA WWPN、Vmkernel、および LUN を使用する VM に関連付けられた仮想 WWPN に認識されるように各 LUN を設定してマスキングし、他の仮想 WWPN および VM に対してはブロックすることを推奨します。

NPIV 環境における VMotion

NPIV を利用する ESX Server では、VM ごとにバーチャル HBA または VPORT を作成できます。図 53 および図 54 に、NPIV を認識し、VMotion に対応する、将来の ESX Server 環境で有効な LUN マッピングを示します。LUN マスキングの要件および動作は、アレイによって異なります。

図 53 では、各 LUN_x を参照できるのは、物理イニシエータ (pWWN-P1、pWWN-P2) およびバーチャルマシン VM_x (pWWN-V_x) に限定されます。ゾーン P には、物理デバイスが含まれます。物理 HBA (pWWN-P1、pWWN-P2) およびストレージアレイポート (pWWN-T) です。各ゾーン V_x には VM_x のバーチャル HBA (pWWN-V_x) およびストレージアレイポート (pWWN-T) だけが含まれます。

図 54 は、SRV-1 から SRV-2 への VM-2 の VMotion 移行を示しています。SAN 構成は変わりません。ゾーン V2 は変わりませんが、VM2 は SRV-2 に配置されています。LUN マッピングおよびマスキングは変わりません。

NPIV を使用すると、ある物理 ESX Server から別の物理 ESX Server にバーチャルマシンを移動させることができます。このとき、ストレージやファブリックの管理者が、ゾーン分割または LUN マッピングおよびマスキングの設定変更を要求する必要はありません。同時に、ゾーン分割および LUN マスキング/マッピングの両方によって、他のバーチャルサーバに対してストレージアクセスが保護されます。これは、物理サーバでは一般的な手法です。

図 53 NPIV を使用するバーチャルマシン (ゾーン分割、LUN マスキング)

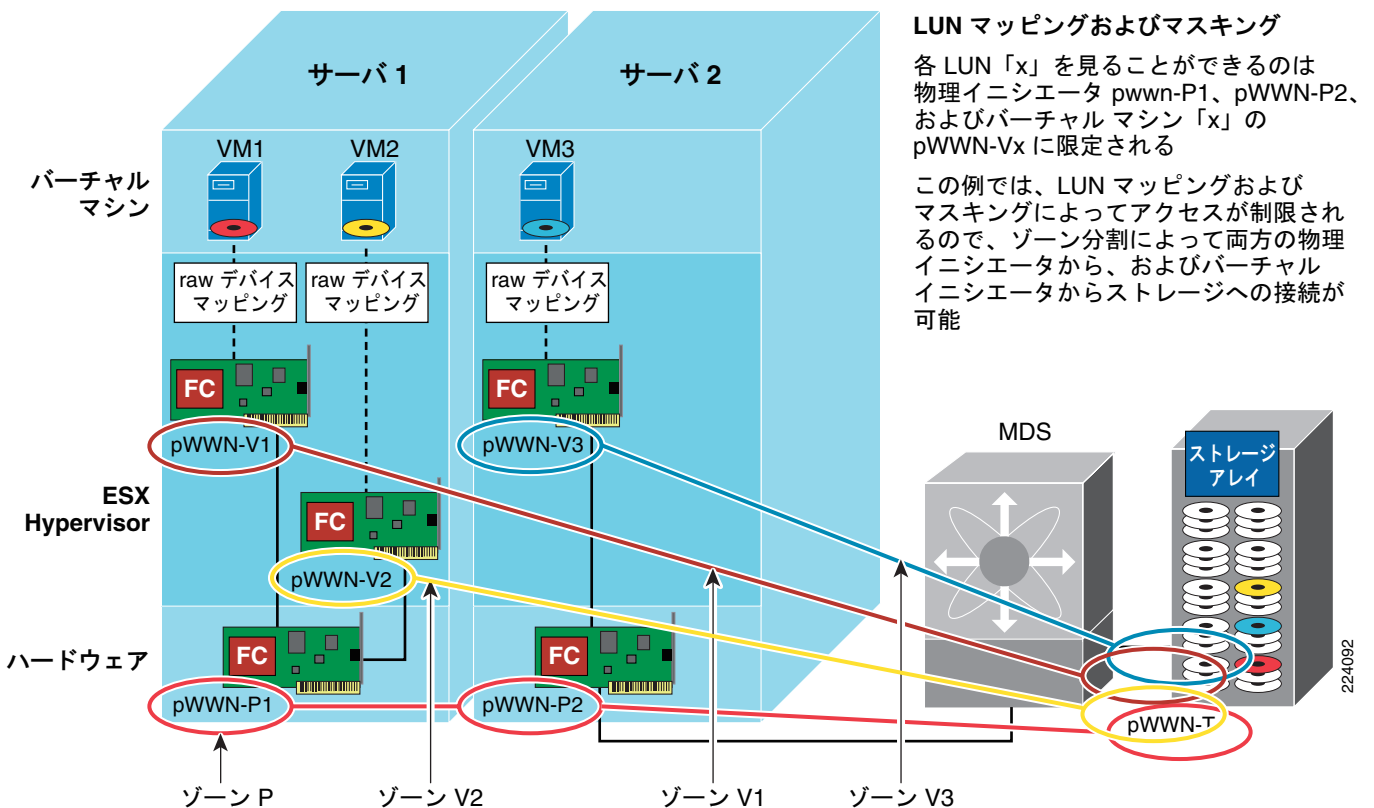
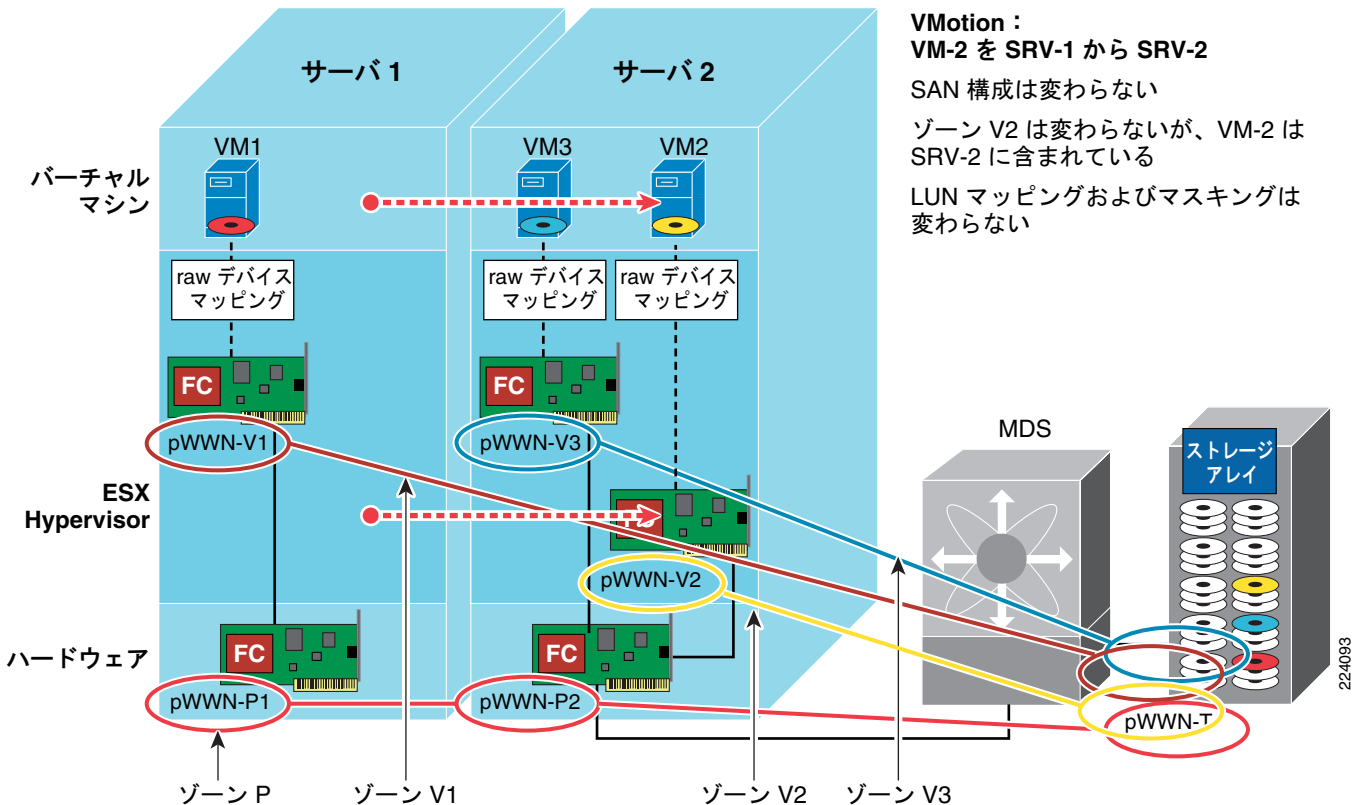


図 54 NPIV を使用する仮想マシン（サーバ 1 からサーバ 2 への VM2 の VMotion 後にゾーン分割、LUN マスキングおよびマッピング）



マルチパス化

NPIV がイネーブルの場合、各仮想マシンの作成時に 4 つの WWN ペア (WWPN および WWNN) が指定されます。NPIV を使用する仮想マシンがオンになると、その仮想マシンはこれらの各 WWN ペアを順番どおりに使用して、ストレージへのアクセスパスを発見しようとします。インスタンス化される VPORT の数とホスト上に存在している物理 HBA (最大 4) の数は同じです。物理パスが見つかった物理 HBA ごとに、VPORT が 1 つずつ作成されます。各物理パスを使用して、LUN アクセスに使用する仮想パスが決定されます。NPIV を使用しない場合と同様、すべてのファイバチャネルマルチパス化はアクティブ/パッシブです。



(注) NPIV を認識しない HBA については、このディスカバリ プロセスは省略されます。NPIV を認識しない HBA 上では、VPORT をインスタンス化できないからです。

トラブルシューティング

fcping を使用して、ファイバチャネル ネットワーク ポイント間の基本接続が判別され、同時にネットワーク遅延の監視と計測が行われます。**traceroute** がノードホップ、待ち時間データを含め、SAN パスについて報告します。物理、または NPIV を使用し、VM と関連付けられた仮想 HBA からの接続には、**fcping** および **fctraceroute** の両方を使用できます。

利点

次に、VMware ESX Server 内部で NPIV を使用する利点を示します。IT コミュニティにとってメリットが大きい順になっています。

- SAN 上での各 NPIV エンティティが固有のものとして識別されるので、バーチャル サーバの個別の SAN 利用を追跡できます。NPIV が開発される以前は、SAN および ESX Server で調べることができたのは、そのシステムで動作しているすべての バーチャル マシンを合わせた、物理 FC ポートの総合利用率だけでした。
- バーチャル マシンを RDM 環境でマッピングされたデバイスと関連付けることによって、アプリケーションのニーズに応じた LUN トラッキングおよびカスタマイズが可能になります。個々の FCID および WWPN を追跡する各種 SAN ツールで、バーチャル マシン固有のパフォーマンスまたは診断データを報告できます。SAN 上で各 NPIV が固有のものとして識別されるので、スイッチ側のレポート ツールおよびアレイ側のツールで、バーチャル マシンごとの診断およびパフォーマンス関連データを報告できます。
- バーチャル マシンとストレージの双方向アソシエーションによって、管理者はバーチャル マシンから SAN 上でプロビジョニングされた LUN へのトレースとともに、SAN 上でプロビジョニングされた LUN から VM へのトレース バックも可能です (NPIV サポートによる重要な拡張)。
- ESX Server をホストとするバーチャル マシンのストレージ プロビジョニングでは、物理サーバと同じ方式、ツール、および専門知識を使用できます。この場合も、バーチャル マシンが一意に WWPN に関連付けられているので、従来のゾーン分割および LUN マスキング方式を引き続き使用できます。
- ファブリック ゾーンで、ターゲットの可視性を特定のアプリケーションに制限できます。アプリケーションに基づく固有の物理アダプタが必要な構成を ESX Server 上の固有の NPIV インスタンスに再マッピングできます。
- バーチャル マシンの移行で、ストレージの可視性の移行をサポートします。ストレージ アクセスを、アクティブにバーチャル マシンを実行している ESX Server に限定できます。新しい ESX Server にバーチャル マシンを移行させる場合に、別の物理ファイバ チャンネル ポートを使用できるように SAN の設定を変更して調整する必要はありません。バーチャル マシンに割り当てられた WWN が維持されるからです。
- HBA のアップグレード、拡張、および交換がシームレスになりました。物理 HBA の WWPN が SAN ゾーン分割および LUN マスキングの準拠エンティティではなくなったので、SAN の設定を変更しなくても、物理アダプタを交換したりアップグレードしたりできます。

パフォーマンスの考慮事項

VMware の配置には、同じ物理マシンにアプリケーションを集約させて入出力を少なくすることを目標とした、統合プロジェクトが関連することがよくあります。その結果、使用中のファイバ チャンネル ポートが少なくなりますが、残りのポートでの帯域幅使用率が高くなります。Cisco MDS9000 には、ESX Server の帯域幅使用率を最適化するツールがいくつもあります。

MDS ポート グループの選択

帯域幅を柔軟に割り当てることのできる Cisco MDS ラインカードを再構成することによって、より多くの帯域幅要求に対処できます。24 ポートのモジュールを例にとります。このラインカードは、4 つの Cisco MDS ポート グループを提供します (ポート グループの説明については、後ろの注を参照)。ESX ホストの HBA には、必要に応じて最大 4 Gbps の帯域幅を割り当てることができます (ポート グループごとに最大 3)。残りのポートはシャットダウンするか、または非 ESX Server の共有モードで使用します。ポート グループをすべて使用している場合 (4 Gbps の専用ポート × 3)、別のポート グループのポートを使用し始める必要があります。



(注)

ポート グループは、一定レベルの帯域幅を共有するポートのグループ化を意味するシスコの用語です。帯域幅割り当てに応じて、使用可能な帯域幅の合計が 12 Gbps を超えないようにしながら、一組のポートに 4 Gbps または 2 Gbps の専用帯域幅を与え、残りのポートに 4 Gbps または 2 Gbps の共有帯域幅を割り当てることができます。

FCC

FCC (ファイバ チャンネル輻輳制御) は、ファイバ チャンネル ネットワーク上の輻輳を緩和する、シスコ独自のフロー制御メカニズムです。詳細については、次の URL を参照してください。

http://www.cisco.com/en/US/products/ps5989/products_configuration_guide_chapter09186a0080663141.html

FCC は、送信元に接続している F ポートの Receiver Ready フレーム生成をペーシングすることによって、送信元を低速にします。VMware 環境では、この機能は十分に注意して使用する必要があります。1 つの HB の背後に複数の VM があり、1 つの VM が輻輳を引き起こすと、同じマシン上のすべての VM が低速になるからです。

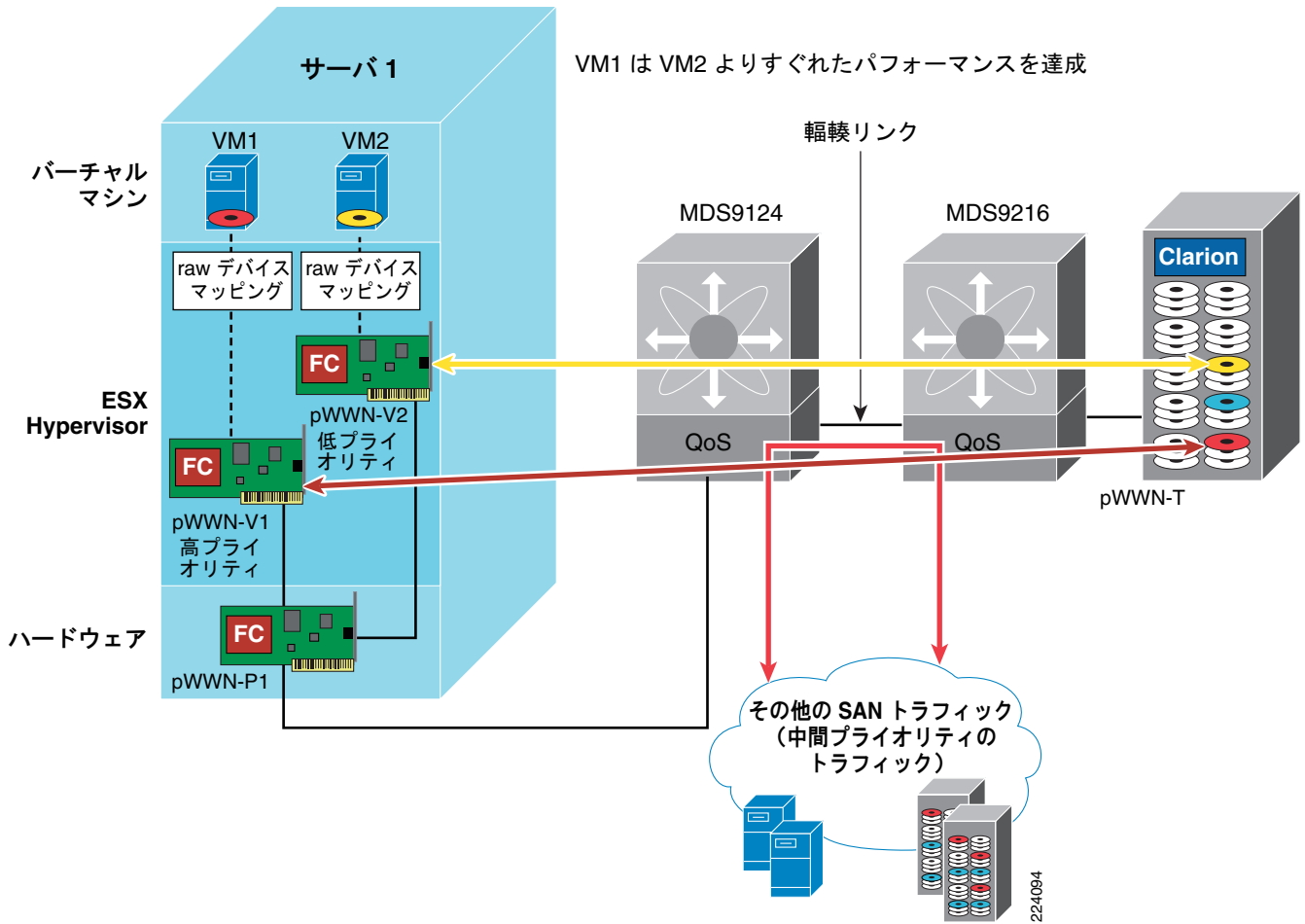
QoS (NPIV を使用する場合)

Cisco MDS Fabric Manager コンソールを使用すると、管理者は任意の WWPN (物理またはバーチャル) に高、中、低のトラフィック プライオリティ レベルを割り当てることができます。輻輳の問題が一般に最小限のローカル SAN では、ポートごとに異なるプライオリティ レベルを設定していても、目立った結果は得られません。NPIV を使用しない場合、QoS (Quality of Service) は VMware 環境で有用ではありません。すべての VM が同じプライオリティになるからです。

NPIV を使用する場合は、管理者が VM ごとに個別に異なる QoS プライオリティを割り当てることができます。たとえば、同じサーバ上のある VM にはリモート ミラーリングを使用し、応答時間をきわめて短くする必要があり、別の VM はリモートバックアップサーバまたはバーチャルテープ アプライアンスにバックアップ ファイルを送信するように設定するという場合があります。この場合、リモート ミラーリングに関するバーチャル WWPN に高いプライオリティを割り当て、バックアップ用の WWPN には低いプライオリティを割り当てることとなります。

図 55 に、IP または DWDM ゲートウェイによるリモート SAN アクセスなど、帯域幅を制限する場合に、QoS を使用して帯域幅割り当てにプライオリティを設定する構成を示します。この場合、VM1 で発生したトラフィックには、VM2 または共通のインフラストラクチャを共有するその他のデバイスで発生したトラフィックより高いプライオリティが与えられます。IOmeter などの SCSI トラフィック ジェネレータを使用すると、所定の構成のパフォーマンス要件が満たされているかどうかを評価し、特定の VM に適切な QoS レベルを設定できます。

図 55 NPIV を使用する VM の QoS



iSCSI の実装に関する考慮事項

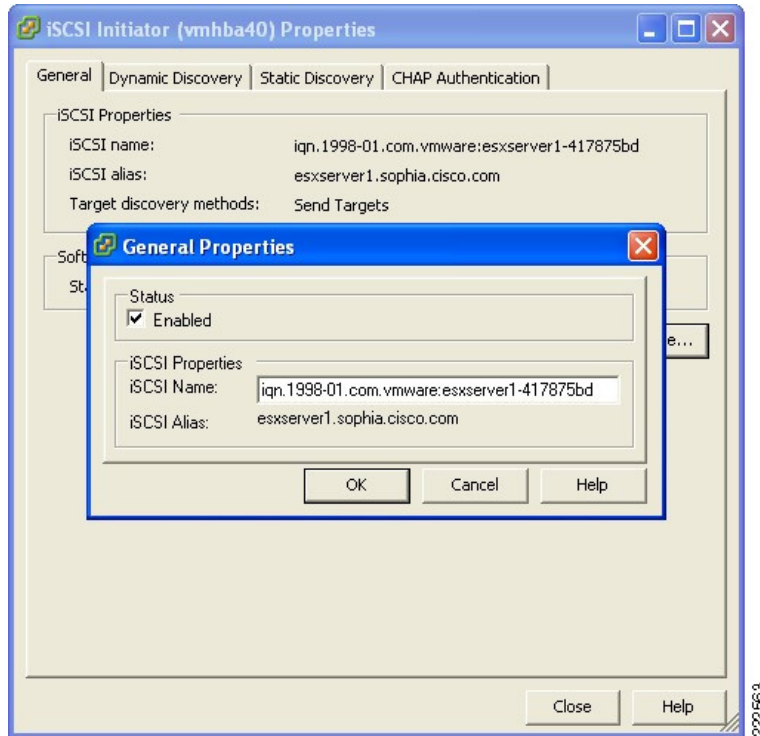
iSCSI を使用すると、ESX ホストからストレージにアクセスできます。iSCSI を機能させるには、VMkernel インターフェイスと Service Console インターフェイスの両方に iSCSI ターゲットへのアクセス権を与え、同じ vSwitch 上で設定する必要があります。VMkernel が iSCSI トラフィックをどのように扱うかについては、すでに説明しました。

図 17 に、有効な構成の例を示します。Service Console 2 は、IP アドレスが 10.0.2.171 で、VLAN 511 上にあり、ESX ホストへのアクセス管理に使用されます。VLAN 200 上の別の Service Console インスタンスでは、IP アドレスは 10.0.200.173 です。VLAN 200 上にあり、IP アドレスが 10.0.200.171 の VMkernel インスタンスは、iSCSI ソフトウェア イニシエータとして使用されます。VLAN 100 上にあり、IP アドレスが 10.0.55.171 の VMkernel インスタンスは、VMotion 用です。

この設定を完了するには、Firewall Properties 設定を表示し (ESX Host、**Configuration** タブ、Software ボックス、Security Profile、Properties の順に選択)、iSCSI ソフトウェア イニシエータ トラフィックを許可します。

図 56 に、iSCSI イニシエータの設定を示します。Configuration タブで ESX ホストを選択してから、**Storage Adapters** を選択します。ここに iSCSI Software Adapter があるので、イニシエータをイネーブルにできます (図 56 を参照)。

図 56 iSCSI イニシエータのイネーブル設定



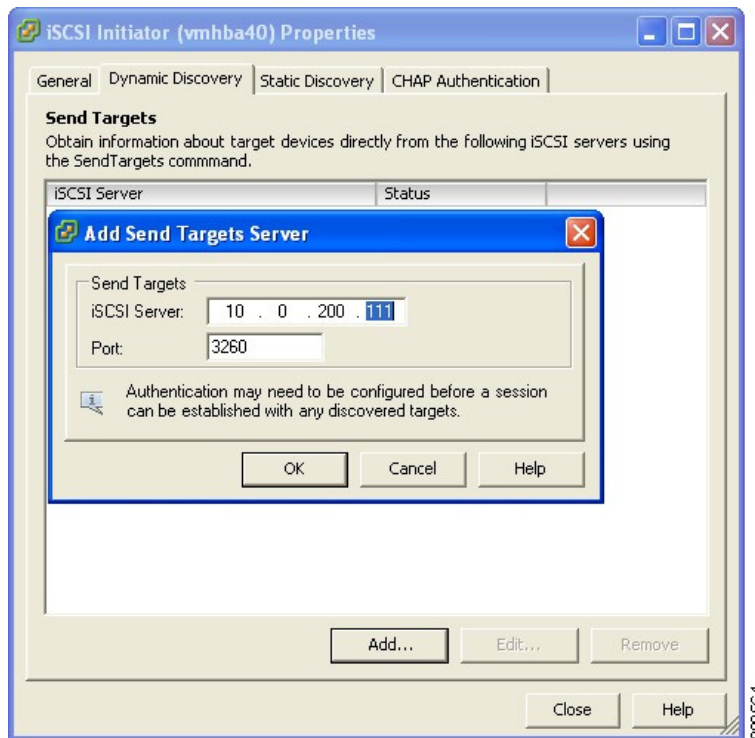
Dynamic Discovery タブから **Add Send Target Server** を選択し、iSCSI ターゲット IP アドレスを指定します。図 57 を参照してください。



(注)

IPS (IP サービス モジュール) ブレードを装備した MDS9000 ファミリーファイバチャネルスイッチまたはダイレクタで iSCSI ターゲットを指定する場合は、VRRP (バーチャルルータ) を使用すると、iSCSI サーバアドレスの可用性を高めることができます。バーチャルルータアドレスは、アクティブ/パッシブ方式 (ピュア VRRP グループ) または iSLB (iSCSI サーバロード バランシング アクティブ VRRP グループ) を使用して、物理 MDS9000 ギガビットイーサネットインターフェイスにマッピングされます。

図 57 iSCSI ターゲットの検出



222564

VMotion ネットワーキング

VMotion は、ESX ファーム/データセンター (VMware VirtualCenter の用語) 内で、ある物理 ESX ホストから別の物理 ESX ホストへオン状態の VM を移行させるために、ESX Server が使用する方式です。VMotion は ESX バーチャル環境で最も強力な機能とも言え、アクティブ VM を最小限の停止時間で移動させることができます。サーバ管理者は VMware VirtualCenter の管理ツールを使用して、VMotion プロセスを手動でスケジューリングしたり開始したりできます。

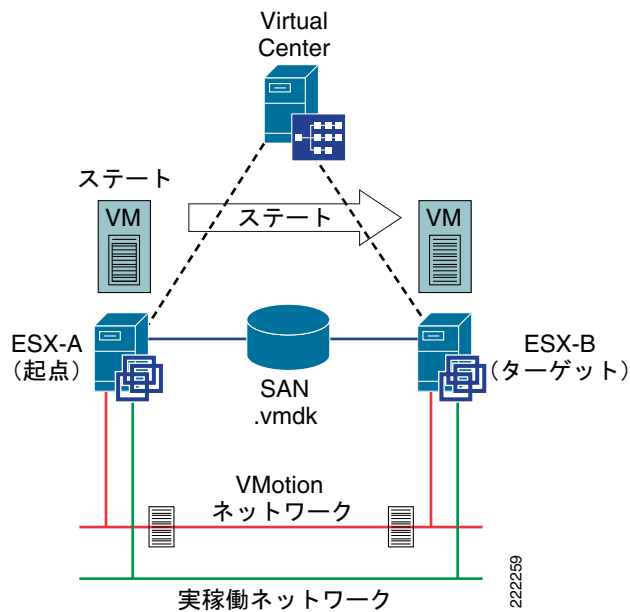
VMotion プロセスは次の手順で実行されます。

- ステップ 1** VirtualCenter が VM およびターゲット ESX ホストの状態を確認します。VirtualCenter は、ターゲット ホスト上で VM をサポートするために必要なリソースの可用性を判別します。
- ステップ 2** ターゲット ホストに互換性がある場合は (同じベンダーまたはファミリの CPU など)、アクティブ VM ステートのコピーが起点 ESX ホストからターゲット ESX ホストに送信されます。ステート情報にはメモリ、レジスタ、ネットワーク接続、および設定情報が含まれます。プレコピー ステートの間に、メモリ ステートがコピーされ、VM のスタン後に初めて、すべてのアクティブ デバイス ステートが移行します。
- ステップ 3** 起点 ESX Server の VM は一時停止状態になります。
- ステップ 4** .vmdk ファイル (バーチャルディスク) のロックが起点 ESX ホストによって解除されます。
- ステップ 5** ステート情報の残りのコピーがターゲット ESX ホストに送信されます。
- ステップ 6** ターゲット ESX ホストが新しい常駐 VM をアクティブにして、同時に関連 .vmdk ファイルをロックします。

ステップ 7 ターゲット ESX ホスト上の vSwitch が通知を受けて、VM の MAC アドレスに対応する RARP を作成します。これによって、Cisco Catalyst スイッチ上のレイヤ 2 フォワーディング テーブルがアップデートされます。VMotion プロセスの間に VM の MAC アドレスが変更されることはないので、gratuitous ARP は不要です。

図 58 に、VMotion プロセスおよびシステムの主要コンポーネントを示します。ESX ホストと VLAN セグメントの両方からアクセスできる SAN ベース VMFS ボリューム、メモリ情報の同期に使用される VMotion ネットワーク (VMkernel に接続するネットワーク)、および VM 上で動作しているアプリケーションへのクライアント アクセスに使用される実稼働ネットワークです。

図 58 VMotion プロセス



VMotion はある ESX ホストから別の ESX ホストへの、バーチャルディスクの完全なコピーではなく、「ステート」のコピーです。 .vmdk ファイルは、SAN の VMFS パーティションにあり、固定です。ESX 起点およびターゲット サーバは、VM ステート情報の同期後に、ファイル ロックの制御を単純に交換します。

VMotion 対応 ESX Server ファームを展開するために必要なものは、次のとおりです。

- VMotion モジュールの組み込まれた VirtualCenter 管理ソフトウェア。
- ESX ファーム/データセンター (VMotion が動作するのは、VirtualCenter コンフィギュレーションの同じデータセンターに含まれている ESX ホストと組み合わせた場合だけです)。ファームの各ホストは、移行後のエラーを避けるために、ほぼ同じハードウェア プロセッサを備えている必要があります (VMware の互換性情報を確認)。
- 共有 SAN。起点 ESX ホストとターゲット ESX ホストで、同じ VMFS ボリューム (.vmdk ファイル) へのアクセス権を与えます。
- ESX ホスト間での WWN の問題を回避するために、VMFS ボリュームを参照する場合のボリューム名。
- VMotion の「接続性」(すなわち、起点 ESX からターゲット ESX、またはその逆の VMkernel の到達可能性)。VMotion は VLAN 上で十分に機能しますが、ステート情報を交換できるように、ギガビット イーサネット ネットワークの使用が望まれます。
- ESX 起点ホストと ESX ターゲット ホストに、同じセキュリティ ポリシーを使用する、同じネットワーク ラベルを設定しておく必要があります。



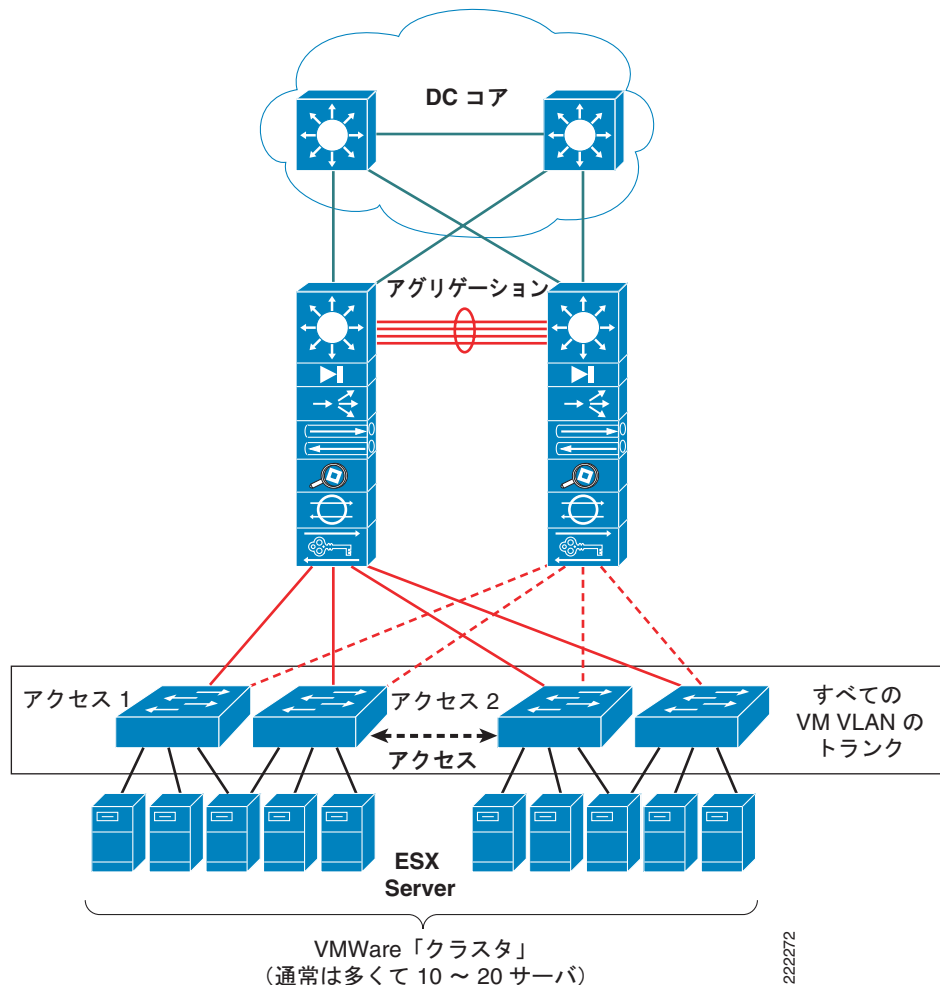
(注) 標準移行（すなわち、非 VMotion 移行）は、電源オフの VM の移行です。このタイプの移行では、メモリのコピーがないので、特に難しいことはありません。また、SAN が存在する場合、VMFS ボリュームは同じデータセンター内の ESX ホストで参照できます。リロケーションを伴う場合（すなわち、.vmdk ファイルを別のデータストアに移動させる必要がある場合）は、電源オンの VM の MAC アドレスが変更される可能性があります。

同一サブネット上での VMotion 移行（フラット ネットワーク）

最も一般的な VM 移行構成の場合、関係マシン間にレイヤ 2 の隣接関係が必要です (図 59 を参照)。レイヤ 2 ドメインの範囲は大部分、アグリゲーション スイッチの同じペアからのアクセス レイヤに限定されます。言い換えると、同じ施設（建物内部）またはキャンパス内の建物にまたがる程度が一般的です。通常、移行を考えると、ホストを同じデータセンターに配置する必要があり、また、DRS のためには同じクラスタに含める必要があるため、関係する ESX ホストの数は多くて 10 ~ 20 です。

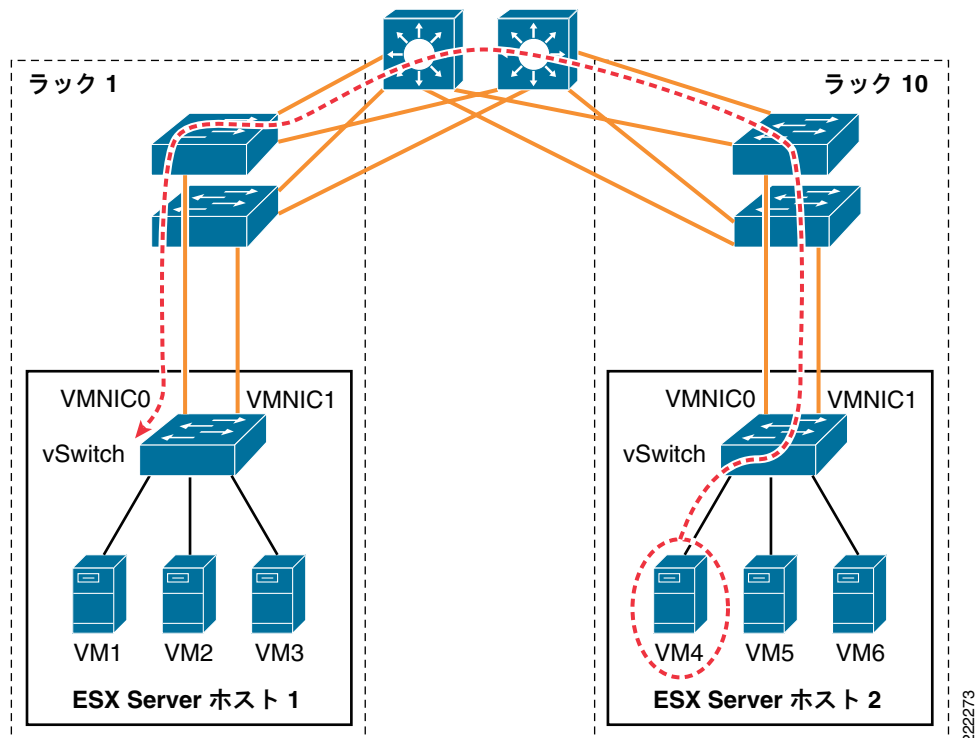
VMware クラスタのレイヤ 2 ソリューションは、クラスタ内のどこであってもマシンをオンにでき、ユーザに気づかれるほどの中断を伴わずに、ある ESX Server から別の ESX Server にアクティブ マシンを移行できるという要件を満たします。

図 59 VMware レイヤ 2 ドメインの要件



VMotion を分かりやすく説明するために、実際の例を使用します。図 60 のようなサーバファーム構成を考えてみてください。ESX Server ホスト 1 はデータセンターのラック 1 にあります。ESX Server ホスト 2 は同じデータセンターのラック 10 にあります。各ラックはサーバに対して、レイヤ 2 接続を提供します（ラック設計の上に位置する設計方法）。レイヤ 3 スイッチのペアがラックを相互接続しますが、互いに何列か離れていることが十分考えられます。実装の目標は、ラック 10 の ESX Server ホスト 2 からラック 1 の ESX Server ホスト 1 に VM4 を移動できるようにすることです。

図 60 レイヤ 2 ネットワークにおける VMotion 移行



これを実現するには、ESX Server ホスト 2 から ESX Server ホスト 1 に VMkernel トラフィックを伝送するようにネットワークのプロビジョニングを行い、ESX Server ホスト 1 で動作しているときに、クライアントが VM4 に到達できるようにする必要があります。

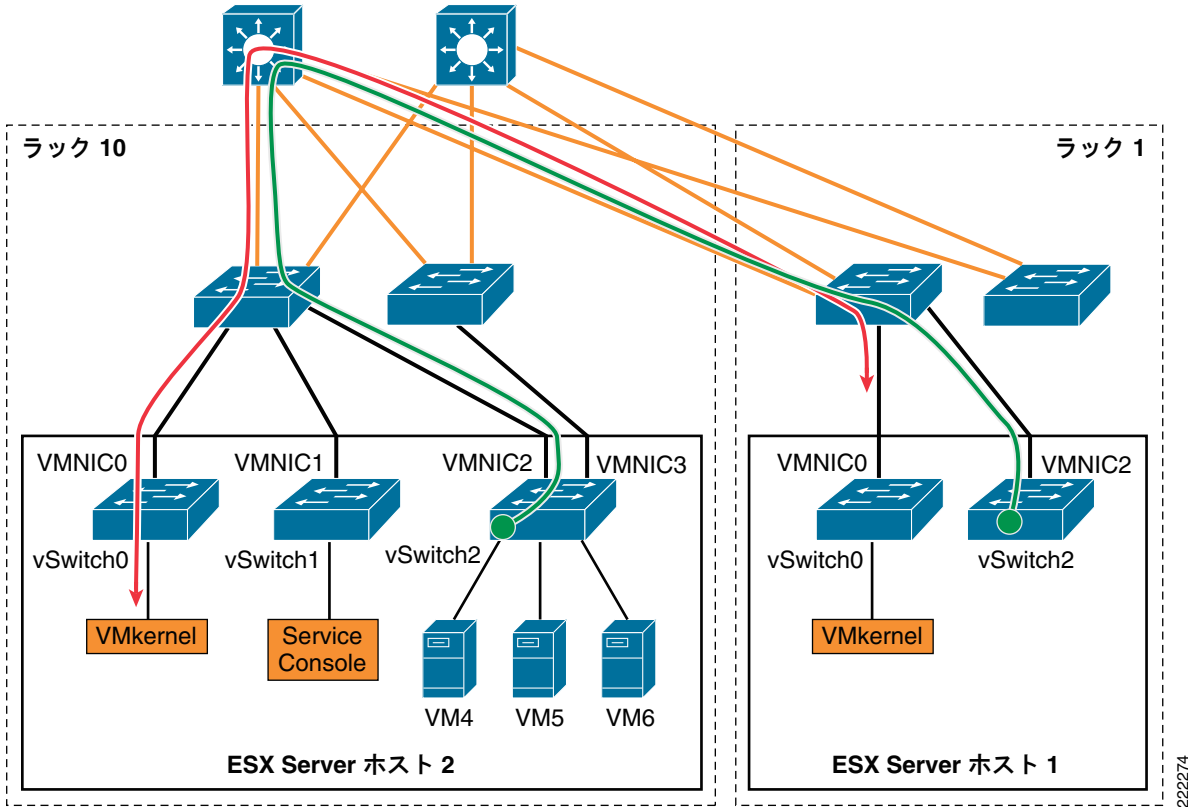
この要件を満たすソリューションは、次のようになります。

- VMkernel に関して VLAN をプロビジョニングします。
- この VLAN を ESX Server ホスト 2 から LAN ネットワーク経由で ESX Server ホスト 1 にトランッキングします。
- VM パブリック アクセスに関して、VLAN をプロビジョニングします。
- この VLAN を ESX Server ホスト 2 から LAN ネットワーク経由で ESX Server ホスト 1 にトランッキングします。
- VMkernel VLAN と VM VLAN を（同じ物理リンクを共有する可能性はあっても）分離しておく必要があります。

ESX ホストの構成は、図 61 のようになります。ESX ホストは、VMkernel 専用 NIC を備えた vSwitch を使用します。VMkernel VLAN のトランクは、ラック 1 のアグリゲーション スイッチからアクセス スイッチを介して ESX Server ホスト 2 の vSwitch までになります。

同様に、VM4 VLAN およびネットワーク ラベルは、ESX Server ホスト 2 上の vSwitch2 でも、ESX Server ホスト 1 上の vSwitch2 と同じになります。

図 61 VM モビリティおよび VLAN 割り当て



222274

ESX HA クラスタ

すでに説明したとおり、VMware ESX HA クラスタは標準 HA クラスタと異なり、アプリケーションの可用性は提供しませんが、故障した ESX ホストで動作していた VM を別の ESX ホストで再起動する機能を提供します。

ESX ホストごとに 1 つずつ HA エージェントが動作します。HA エージェントを使用して、同じ VMware HA クラスタに含まれている他の ESX ホストの可用性を監視します。ESX ホスト ネットワークのモニタリングは、Service Console の VLAN 上で、または複数の Service Console が構成されている場合は複数の VLAN 上でユニキャスト UDP フレームを交換して行います。エージェントは ~8042 など、4 種類の UDP ポートを使用します。毎秒、UDP ハートビートが送信されます。実稼働 VLAN 上ではハートビートの交換は行われません。ESX HA の初期設定には、DNS (ドメインネーム サービス) が必要です。

他の ESX ホストとのハートビートによる接続を失った ESX ホストは、ゲートウェイへの ping を開始して、ネットワークにまだアクセスできるのか、それとも孤立したのかを調べます。原則として、孤立していると認識した ESX ホストは、`.vmdk` ファイルのロックが解除されるように VM を停止して、他の ESX ホストで VM を再起動できるようにします。

分離を回避するために、デフォルトの設定では VM がシャットダウンされます。ESX はネットワークアイソレーションがきわめてまれなので、可用性と冗長性の高いネットワークとされています。HA に関する推奨事項は、次のとおりです。

- ステップ 1** 2 つの Service Console を別々のネットワーク上で設定します。
- ステップ 2** Rolling Failover = Yes (ESX 3.0.x) または Failback = No (ESX 3.5) とし、2 つのチーミング vmnic を指定して各 Service Console を設定します。
- ステップ 3** チーミング vmnic を必ず、別々の物理スイッチに接続します。
- ステップ 4** ESX ホスト間に完全な冗長性のある物理パスが存在し、分離状況を引き起こす可能性のあるシングルポイント障害がないようにします。

VMware の VM を別の ESX ホストで再起動する場合と同様、電源オン時に VM が接続するネットワーク ラベルがなければなりません。後述する障害状況で、VMware HA クラスタの動作を理解してください。

VMware ESX HA の詳細については、『VMware HA: Concepts and Best Practices』を参照してください。URL は次のとおりです。

<http://www.vmware.com/resources/techresources/402>

メンテナンス モード

最初の例では、**esxserver1** をメンテナンス モードにします。VM は図 62 のように、VMotion で **esxserver2** に移行します。ネットワークに関して、VM が **esxserver2** に移行する唯一の条件は、同じネットワーク ラベルが存在していることです。クライアントが移行した VM で動作を続けるには、ネットワーク ラベルが **esxserver1** と同じ VLAN でなければなりません。また、この VLAN と Cisco Catalyst スイッチ間に vmnic トランクが必要です。



(注)

ネットワーク ラベルの名称を別にする、VirtualCenter および HA クラスタ ソフトウェアは、VLAN 番号も vSwitch 上での vmnic の存在も検証しません。VLAN 構成および vmnic が正しく設定されているかどうかの事前確認は、管理者に委ねられます。

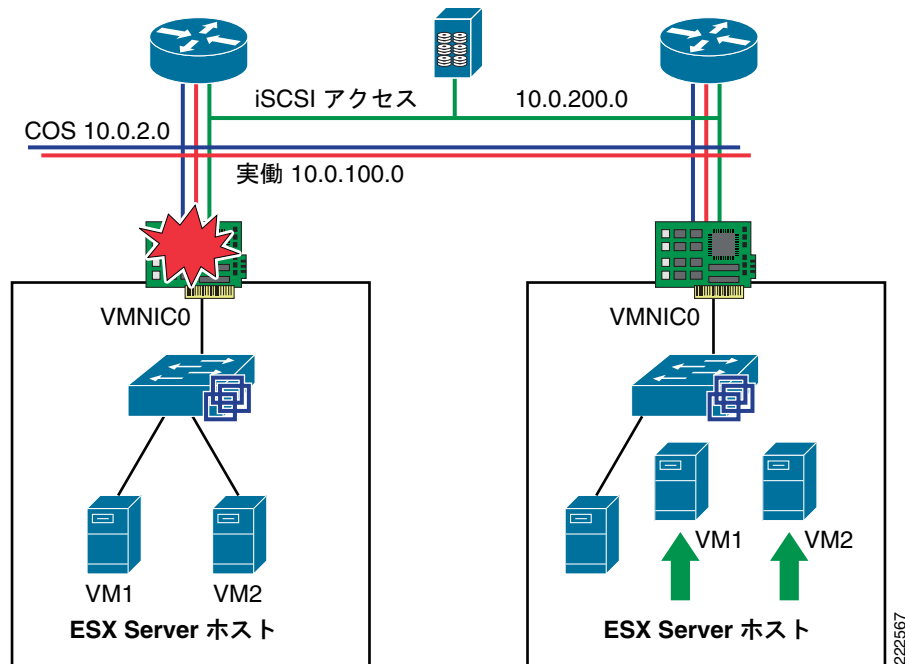
図 62 HA クラスタおよび ESX ホストのメンテナンス モード



実稼働、管理、およびストレージからの ESX ホストの切り離し

この例では、ESX Server ホスト 1 は 1 つの vmnic だけで LAN スイッチング ネットワークに接続されています (図 63 を参照)。ある 1 つの VLAN がネットワーク管理アクセスを提供し (10.0.2.0)、1 つの VLAN が実稼働ネットワークへのアクセスを提供し (10.0.100.0)、さらに 1 つの VLAN が iSCSI による SAN へのアクセスを提供します (10.0.200.0)。

図 63 HA クラスタおよび NIC 障害



NIC vmnic0 が切断されると、ESX1 の接続がすべての VLAN で失われます。ESX2 の HA エージェントが Service Console ネットワークから ESX1 に到達できないことを判別し、VM1 および VM2 を起動できるように、VMFS ボリュームの確保を試みます。ESX1 は切り離されているだけでなく、iSCSI ディスクのコントロールも失っているため、ロックがやがてタイムアウトし、ESX2 がディスクを確保して VM1 および VM2 を再起動できます。VM が再起動するまでの合計時間は、18 ~ 20 秒程度です。



(注) この間、ESX2 ESX1 への接続を数回試みてから、障害を宣言し、さらにゲートウェイに到達できるかどうかを確認します。



(注) たまたま、VirtualCenter から ESX1 の障害を確認したとしても、vmnic0 が障害として表示されるわけではありません。vmnic0 を失うということは、ネットワーク管理接続を失うことでもあるので、VirtualCenter は ESX1 からモジュールステータス情報を収集できません。

全体として考えると、2 つの vmnic を NIC チーミングとして設定する方が、別の ESX ホストで VM を再起動するより、コンバージェンス タイムが短縮されます。

Server Console 単独での接続損失

次に、[図 63](#) と同様ですが、Service Console 用の vmnic が実稼働、VMkernel、および iSCSI トラフィック用の他の vmnic と切り離されている点が異なるネットワークについて検討します。ESX ホスト 1 上で Service Console vmnic 障害が発生すると、管理通信ができなくなるので、ESX ホスト 1 は見かけ上、VirtualCenter から切り離されます。しかし、ESX ホスト 1 は引き続きアップで動作しており、VM もオンの状態です。ESX ホスト 2 が冗長 Service Console（ルーティングされていない可能性あり）を介して ESX ホスト 1 と引き続き通信できる場合は、VM をオフにして ESX ホスト 2 上で再起動する必要はありません。

冗長 Service Console を推奨します。ソフトウェア iSCSI イニシエータは、VMkernel アドレスを使用して実装します。iSCSI では Service Console を使用して認証を行うので、iSCSI VLAN またはネットワークとのレイヤ 3 またはレイヤ 2 接続が可能でなければなりません。

実稼働単独での接続損失

[図 63](#) で、ESX1 上の NIC は切り離されていないが、なんらかの理由で、実稼働ネットワーク上のパスが使用できない状況（通常は誤設定による）について考えてみましょう。HA クラスタが ESX1 上の VM をシャットダウンして、ESX2 上で再起動することはありません。Service Console ネットワークで ESX1 と ESX2 が引き続き通信できるからです。その結果、ESX1 上の VM が孤立します。

同様に、ESX1 に 2 つの vmnic があり、一方を実稼働ネットワークに、他方を Service Console、VMkernel、および iSCSI に使用するとします。最初の vmnic で障害が発生すると、ESX2 上での VM 再起動に至ります。その結果、ESX1 上の VM が孤立します。

この状況を回避するには、1 つの vmnic を実稼働トラフィック単独の専用にするのではなく、NIC チーミング構成で複数の vmnic を使用するのが最良です。

その他のリソース

ESX Server リリースの詳細については、VMware テクノロジー ネットワーク サイトにアクセスしてください。URL は次のとおりです。

http://www.vmware.com/support/pubs/vi_pubs.html

<http://www.vmware.com/products/vc/>

http://www.vmware.com/pdf/vi3_301_201_admin_guide.pdf

Cisco Catalyst 6500 :

<http://www.cisco.com/en/US/products/hw/switches/ps708/>

Cisco ブレード スイッチ :

http://www.cisco.com/en/US/prod/collateral/switches/ps6746/ps8742/Brochure_Cat_Blade_Switch_3100_Next-Gen_Support.html

http://www.cisco.com/en/US/prod/collateral/switches/ps6746/ps8742/White_Paper_Cat_Blade_Switch_3100_Design.html

<http://www.cisco.com/en/US/products/ps6748/index.html>

<http://www.cisco.com/en/US/products/ps6294/index.html>

<http://www.cisco.com/en/US/products/ps6982/index.html>

<http://www.vmware.com/resources/techresources/>

<http://pubs.vmware.com/vi35/wwhelp/wwhimpl/js/html/wwhelp.htm>