vSAN ネットワーク設計

Update 3

VMware vSphere 8.0
VMware vSAN 8.0

vmware
by Broadcom

VMware by Broadcom の Web サイトで最新の技術ドキュメントを確認できます

https://docs.vmware.com/jp/

VMware by Broadcom
3401 Hillview Ave.
Palo Alto, CA 94304
www.vmware.com

目次

1 vSAN のネットワーク設計について 5

2 vSAN ネットワークについて 6

3 vSAN ネットワークについて 9
- vSAN ネットワークの特性 10
- ESXi トラフィック タイプ 11
- vSAN のネットワーク要件 12
  - 物理 NIC の要件 12
  - バンド幅と遅延の要件 13
  - レイヤー 2 とレイヤー 3 のサポート 14
  - ルーティングとスイッチの要件 14
  - vSAN ネットワーク ポートの要件 16
  - ネットワーク ファイアウォールの要件 16

4 vSAN ネットワークでのユニキャストの使用 18
- バージョン 5 より前のディスク グループの動作 18
- バージョン 5 のディスク グループの動作 19
- ユニキャスト ネットワークでの DHCP サポート 19
- ユニキャスト ネットワークでの IPv6 サポート 19
- ESXCLI を使用したユニキャストのクエリ 19
  - 通信モードの表示 20
  - vSAN クラスタ ホストの確認 20
  - vSAN ネットワーク情報の表示 21
- クラスタ内のトラフィック 21
  - 単一ラックでのクラスタ内のトラフィック 22
  - vSAN ストレッチ クラスタ内のクラスタ内トラフィック 22

5 IP ネットワーク転送の構成 24
- vSphere TCP/IP スタック 24
- Object Missing 26
- IPv6 サポート 26
- スタティック ルート 26
- ジャンボ フレーム 27

6 VMware NSX と vSAN の併用 28

7 輻輳制御とフロー制御の使用 29

8 基本的な NIC チーミング、フェイルオーバー、ロード バランシング 31
基本的な NIC チーミング 31
NIC チームのロード バランシングの設定 33

9 高度な NIC チーミング 35
リンク集約グループの概要 36
静的リンクと動的リンクの集約 36
静的 LACP と IP ハッシュに基づいたルート 37
ネットワークの air gap について 39
vSAN による air gap ネットワーク構成の長所と短所 39
NIC チーミングの構成例 40
構成 1：単一の vmknic、物理 NIC の負荷に基づいたルート 41
構成 2：複数の vmknic、発信元のポート ID に基づいたルート 42
構成 3：動的 LACP 44
構成 4：静的 LACP – IP ハッシュに基づいたルート 50

10 Network I/O Control 53
Network I/O Control の構成例 54

11 vSAN ネットワーク トポロジについて 56
標準の展開 56
vSAN ストレッチ クラスタの展開 59
2 ノード構成の vSAN の展開 64
データ サイトから監視ホストへのネットワークの構成 66
例外的な展開 68

12 vSAN ネットワークのトラブルシューティング 69

13 vSAN ネットワークでのマルチキャストの使用 79
Internet Group Management Protocol 79
プロトコルに依存しないマルチキャスト 80

14 vSAN ファイル サービスのネットワークに関する考慮事項 81

15 vSAN での iSCSI のネットワークに関する考慮事項 84
vSAN iSCSI ネットワークの特性 84

16 標準スイッチから Distributed vSwitch への移行 85

17 vSAN ネットワークのチェックリストのサマリ 90

vSAN のネットワーク設計について 1

『vSAN ネットワーク設計ガイド』では、可用性が高くスケーラブルな vSAN クラスタを展開するためのネットワーク要件、ネットワーク設計および構成作業について説明します。

vSAN は分散ストレージ ソリューションです。他の分散型ソリューションと同様に、システムを設計する上でネットワークは重要なコンポーネントとなります。不適切なネットワーク ハードウェアを使用したり、設計が適切でないと、好ましくない結果になる可能性があります。最適な結果を得るには、このドキュメントに記載されているガイドラインに従う必要があります。

VMware では、多様性の受け入れを尊重しています。ユーザー、パートナー、社内コミュニティ内でこの原則を促進するため、包括的な表現でコンテンツを作成します。

対象読者

このガイドは、vSAN クラスタの設計、展開、管理を行う方を対象としています。このガイドの情報は、ネットワークの設計と構成、仮想マシンの管理、仮想データセンターの運用に精通している経験豊富なネットワーク管理者を対象としています。また、VMware ESXi、vCenter Server、vSphere Client など、VMware vSphere に精通していることを前提としています。

関連ドキュメント

このガイドの他に、以下のガイドにも vSAN ネットワークの詳細が記載されています。

- 『vSAN のプランニングと展開』。vSAN クラスタの作成について詳しく説明しています。
- 『VMware vSAN の管理』。vSAN クラスタの構成、vSAN 機能の詳細について説明しています。
- 『vSAN の監視とトラブルシューティング ガイド』。vSAN クラスタの監視とトラブルシューティングについて説明しています。

vSAN ネットワークについて

2

vSAN を使用すると、vSphere 内で共有ストレージをプロビジョニングできます。vSAN はホスト クラスタのローカル デバイスまたは直接接続されたストレージ デバイスを統合し、vSAN クラスタのすべてのホストで共有される単一のストレージ プールを作成します。

vSAN は分散型の共有ストレージ ソリューションで、vSAN ストレージ トラフィック用に最適化された高可用性ネットワークが必要になります。高パフォーマンスで可用性の高いネットワークがなければ、vSAN の展開に成功することはできません。このガイドでは、vSAN ネットワークの設計と構成に関する推奨事項について説明します。

vSAN は、高パフォーマンスで回復力が高くスケーラブルなネットワークを前提とする分散アーキテクチャになっています。vSAN クラスタ内のすべてのホストノードは IP ネットワークを介して通信を行います。レイヤー 2 またはレイヤー 3 ネットワークを介した通信を可能にするため、すべてのホストが IP ユニキャスト接続を維持する必要があります。ユニキャスト通信の詳細については、「4 章 vSAN ネットワークでのユニキャストの使用」を参照してください。

vSAN ネットワークの用語および定義

vSAN では、独自の用語と定義が使用されており、これらを理解することが重要となります。vSAN ネットワークの設計を開始する前に、vSAN ネットワークの重要な用語と定義を確認してください。

用語	定義
CLOM	Cluster-Level Object Manager (CLOM) は、オブジェクトの構成がそのストレージ ポリシーと一致することを保証します。CLOM は、そのポリシーを満たすのに十分なフォルト ドメインが使用可能かどうかを確認します。クラスタ内のコンポーネントと監視の配置場所を決定します。
CMMDS	クラスタの監視、メンバーシップ、ディレクトリ サービス (CMMDS) は、ネットワーク化されたノード メンバーのクラスタのリカバリとメンテナンスを行います。ホスト ノード、デバイス、ネットワークなどのアイテムのインベントリを管理します。また、vSAN オブジェクトのポリシーや RAID 構成などのメタデータ情報も保存します。
DOM	Distributed Object Manager (DOM) は、コンポーネントの作成とクラスタ全体への配布を行います。DOM オブジェクトが作成されると、ノード（ホスト）の 1 つがそのオブジェクトの DOM 所有者に任命されます。このホストは、クラスタ全体でそれぞれの子コンポーネントを特定し、vSAN ネットワークを介して I/O を各コンポーネントにリダイレクトして、すべての IOPS をその DOM オブジェクトに渡します。DOM オブジェクトには、vdisk、スナップショット、vmnamespace、vmswap、vmem などがあります。

用語	定義
LSOM	Log-Structured Object Manager (LSOM) は、vSAN コンポーネントまたは LSOM オブジェクト（データ コンポーネントまたは監視コンポーネント）として vSAN ファイル システムにデータをローカルに格納します。
NIC チーミング	ネットワーク インターフェイス カード (NIC) チーミングは、高可用性とロード バランシングを実現するため、チームとして設定された 2 つ以上のネットワーク アダプタ (NIC) です。
NIOC	Network I/O Control (NIOC) は、vSphere Distributed Switch 上に複数のネットワーク トラフィック タイプがある場合に、そのバンド幅を判断します。バンド幅の分布はユーザーが設定可能なパラメータです。NIOC が有効になっている場合、Distributed Switch のトラフィックは、フォールト トレランス トラフィック、iSCSI トラフィック、vMotion トラフィック、管理トラフィック、vSphere Replication トラフィック、NFS トラフィック、仮想マシン トラフィックなど、事前定義済みのネットワーク リソース プールに分けられます。
オブジェクトとコンポーネント	各オブジェクトは、一連のコンポーネントで構成されます。これらは、仮想マシンのストレージ ポリシーが使用する機能に応じて決定されます。 vSAN データストアには、いくつかのオブジェクト タイプが含まれます。 ■ [仮想マシンの Home 名前空間] - 仮想マシンの Home 名前空間は、すべての仮想マシン構成ファイルが格納される仮想マシンのホーム ディレクトリです。ここには、.vmx、ログ ファイル、vmdks、スナップショットの差分記述ファイルなどが格納されます。 ■ [VMDK] - VMDK は、仮想マシンのハード ディスク ドライブの内容を格納する、仮想マシンのディスク ファイル（.vmdk ファイル）です。 ■ [仮想マシンのスワップ オブジェクト] - 仮想マシンのスワップ オブジェクトは、仮想マシンのパワーオン時に作成されます。 ■ [スナップショット差分 VMDK] - スナップショット差分 VMDK は、仮想マシンのスナップショットの作成時に生成されます。 ■ [メモリ オブジェクト] - メモリ オブジェクトは、仮想マシンの作成またはサスペンド時に、スナップショット メモリ オプションを選択すると作成されます。
RDT	Reliable Data Transport (RDT) プロトコルは、vSAN VMkernel ポートを介したホスト間の通信に使用されます。トランスポート レイヤーで TCP を使用し、必要に応じて TCP 接続（ソケット）の作成と破棄を行います。大きなファイルを送信するように最適化されています。
SPBM	ストレージ ポリシー ベースの管理 (SPBM) は、広範なデータ サービスおよびストレージ ソリューション間で単一の統合されたコントロール ペインとして機能するストレージ ポリシー フレームワークです。このフレームワークは、ストレージで仮想マシンのアプリケーションのニーズを満たすのに役立ちます。

用語	定義
VASA	vSphere Storage APIs for Storage Awareness (VASA) は、vCenter Server がストレージ アレイの機能を認識するための一連のアプリケーション プログラム インターフェイス (API) です。VASA プロバイダは、vCenter Server と通信を行い、ポリシーベースの管理、運用管理、DRS 機能をサポートするストレージ トポロジ、機能、状態情報を確認します。
VLAN	VLAN は、単一の物理 LAN セグメントをさらにセグメント化して、ポート グループが物理的に別々のセグメントにあるかのように、互いに分離できます。
監視コンポーネント	監視は、メタデータのみを含み、実際のアプリケーション データは何も含まないコンポーネントです。障害が発生した後、存続しているデータストアのコンポーネントの可用性に関して決定を下す場合のタイブレーカとして機能します。オンディスク フォーマット 1.0 を使用する場合、監視は vSAN データストアでメタデータにおよそ 2 MB の容量を使用し、バージョン 2.0 以降のオンディスク フォーマットでは 4 MB の容量を使用します。

vSAN ネットワークについて

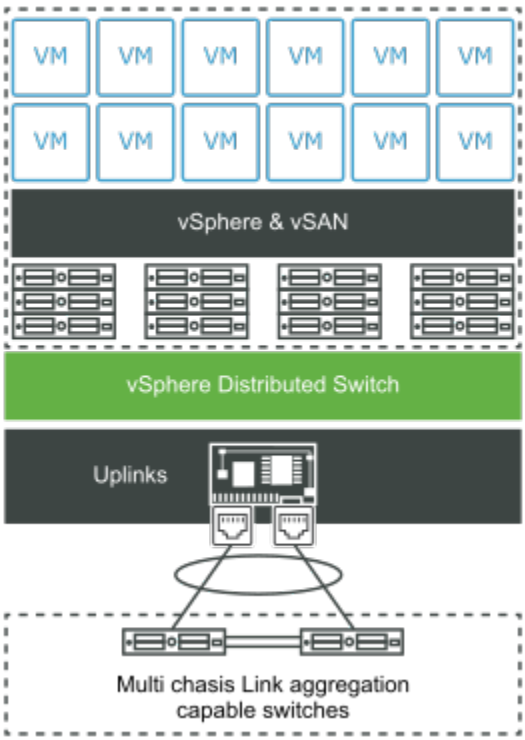
3

vSAN ネットワークは、クラスタ ホスト間の通信を容易にするため、高速なパフォーマンス、高可用性およびバンド幅を維持する必要があります。

vSAN は、ネットワークを介して ESXi ホスト間の通信と仮想マシンのディスク I/O を処理します。

vSAN データストア上の仮想マシン (VM) は一連のオブジェクトから構成され、各オブジェクトは 1 つ以上のコンポーネントから構成されます。これらのコンポーネントが複数のホストに分散し、ドライブやホストの障害に対する回復力を維持しています。vSAN は、vSAN ネットワークを使用して、これらのコンポーネントの保守と更新を行います。

次の図は、vSAN ネットワークの概要を表しています。



次のトピックを参照してください。

- vSAN ネットワークの特性
- ESXi トラフィック タイプ
- vSAN のネットワーク要件

vSAN ネットワークの特性

vSAN はネットワークに依存します。パフォーマンスや安定性の問題を回避するには、適切な vSAN ネットワーク設定が何かを理解し、そのような構成にすることが重要です。

信頼性の高い堅牢な vSAN ネットワークには、次のような特性があります。

[ユニキャスト]

vSAN 6.6 以降のリリースでは、ユニキャスト通信がサポートされます。ユニキャスト トラフィックとは、ネットワーク内のある地点から別の地点に 1 対 1 で IP パケットを送信することを意味します。ユニキャストは、1 秒ごとにプライマリ ホストから送信されたハートビートを他のすべてのホストに転送します。これにより、ホストがアクティブで、vSAN クラスタに参加していることが確認できます。vSAN 用にシンプルなユニキャスト ネットワークを設計できます。ユニキャスト通信の詳細については、「4 章 vSAN ネットワークでのユニキャストの使用」を参照してください。

注： 可能であれば、常に最新バージョンの vSAN を使用してください。

[レイヤー 2 とレイヤー 3 ネットワーク]

vSAN クラスタのすべてのホストは、レイヤー 2 またはレイヤー 3 ネットワークを介して接続されている必要があります。vSAN 6.0 より前の vSAN リリースでは、レイヤー 2 ネットワークのみがサポートされますが、それ以降のリリースではレイヤー 2 とレイヤー 3 の両方のプロトコルがサポートされます。データサイトと監視サイト間の通信を可能にするには、レイヤー 2 またはレイヤー 3 ネットワークを使用します。レイヤー 2 とレイヤー 3 ネットワーク トポロジーの詳細については、「標準の展開」を参照してください。

[VMkernel ネットワーク]

vSAN クラスタの各 ESXi ホストには、vSAN 通信用のネットワーク アダプタが必要です。クラスタ内のノード間の通信はすべて vSAN VMkernel ポート経由で行われます。VMkernel ポートは、各 vSAN ホストおよびホストされた仮想マシンにレイヤー 2 およびレイヤー 3 サービスを提供します。

[vSAN ネットワーク トラフィック]

vSAN ネットワークでは、ストレージ トラフィックやユニキャスト トラフィックなど、さまざまなトラフィック タイプを扱うことができます。仮想マシンのコンピューティングとストレージは、同じホストに配置することも、クラスタ内の異なるホストに配置することもできます。障害を許容するように構成されていない仮想マシンが 1 台のホスト上で実行され、別のホストに存在する仮想マシン オブジェクトやコンポーネントにアクセスしている可能性があります。これは、仮想マシンからのすべての I/O がネットワークを通過することを意味します。ストレージ トラフィックは、vSAN クラスタ内のほとんどのトラフィックを構成します。

すべての ESXi ホスト間のクラスタ関連の通信では、vSAN クラスタにトラフィックが作成されます。このユニキャスト トラフィックも vSAN ネットワーク トラフィックに影響します。

[仮想スイッチ]

vSAN は、次のタイプの仮想スイッチをサポートします。

- 標準仮想スイッチは、仮想マシンおよび VMkernel ポートから外部ネットワークへの接続を確保します。このスイッチは、それぞれの ESXi ホストに対してローカルに存在します。

- vSphere Distributed Switch は、複数の ESXi ホストに存在する仮想スイッチを一元的に管理できます。Distributed Switch は、vSphere または仮想ネットワークにサービス品質 (QoS) レベルの設定に役立つ Network I/O Control (NIOC) などのネットワーク機能を提供します。vCenter Server のバージョンに関係なく、vSAN には、vSphere Distributed Switch が含まれます。

[バンド幅]

vSAN トラフィックは、vSphere vMotion トラフィック、vSphere HA トラフィック、および仮想マシン トラフィックなどの他のシステムのトラフィック タイプと、物理ネットワーク アダプタを共有できます。また、vSAN、vSphere 管理、vSphere vMotion トラフィックなどが同じ物理ネットワーク上にある場合、より多くのバンド幅を共有ネットワーク構成に確保されます。vSAN に必要なバンド幅を確保するには、Distributed Switch で vSphere Network I/O Control を使用します。

vSphere Network I/O Control では、vSAN 送信トラフィックの予約とシェアを構成できます。

- vSAN の物理アダプタで使用できる最低のバンド幅が Network I/O Control で確保されるように予約を設定します。
- vSAN に割り当てられた物理アダプタが飽和状態になったときに特定のバンド幅を vSAN で使用できるように、シェア値を 100 に設定します。たとえば、チームの別の物理アダプタに障害が発生し、ポート グループのすべてのトラフィックがチーム内の別のアダプタに転送されると、物理アダプタが飽和状態になる可能性があります。

Network I/O Control を使用して vSAN トラフィックのバンド幅の割り当てを設定する詳細については、『vSphere ネットワーク』ドキュメントを参照してください。

ESXi トラフィック タイプ

ESXi ホストは、vSAN をサポートするために、さまざまなネットワーク トラフィック タイプを使用します。

vSAN 用に設定する必要があるトラフィック タイプは次のとおりです。

表 3-1. ネットワーク トラフィック タイプ

トラフィック タイプ	説明
管理ネットワーク	管理ネットワークは、VMkernel TCP/IP スタックを使用してホストの接続と管理を行うプライマリ ネットワーク インターフェイスです。また、vMotion、iSCSI、ネットワーク ファイル システム (NFS)、ファイバ チャネル オーバー イーサネット (FCoE)、フォルト トレランスなどのシステム トラフィックも処理できます。
仮想マシン ネットワーク	仮想ネットワークを使用すると、仮想マシンをネットワーク化し、単一の ESXi ホストまたは複数の ESXi ホスト間で複雑なネットワークを構築できます。

表 3-1. ネットワーク トラフィック タイプ （続き）

トラフィック タイプ	説明
vMotion ネットワーク	1 台のホストから別のホストへの仮想マシンの移行を容易にするトラフィック タイプ。vMotion による移行を行うには、ソース ホストとターゲット ホストに、正しく構成されたネットワーク インターフェイスが必要です。vMotion ネットワークは、vSAN ネットワークと異なります。
vSAN ネットワーク	vSAN クラスタでは、データ交換を行うため VMkernel ネットワークが必要です。vSAN クラスタの各 ESXi ホストには、vSAN トラフィック用の VMkernel ネットワーク アダプタが必要です。詳細については、『vSAN のプランニングとデプロイ』の「手動での vSAN の有効化」を参照してください。

vSAN のネットワーク要件

vSAN は、ホスト間の通信をネットワークに依存する分散型ストレージ ソリューションです。展開する前に、ご使用の vSAN 環境がすべてのネットワーク要件を満たしていることを確認します。

物理 NIC の要件

vSAN ホストのネットワーク インターフェイス カード (NIC) は、特定の要件を満たしている必要があります。vSAN は、10 Gbps、25 Gbps、40 Gbps、50 Gbps、および 100 Gbps ネットワークで動作します。

ホストが vSAN Original Storage Architecture (OSA) または vSAN Express Storage Architecture (ESA) の最小 NIC 要件を満たしていることを確認します。

表 3-2. vSAN OSA の最小 NIC 要件と推奨事項

トポロジまたは展開モード	アーキテクチャ	1 GbE NIC のサポート	10 GbE NIC のサポート	10 GbE を超える NIC のサポート	ノード間の遅延	サイト間リンクのバンド幅または遅延	ノードと vSAN 監視ホスト間の遅延	ノードと vSAN 監視ホスト間のバンド幅
単一サイトの vSAN クラスタ	ハイブリッド クラスタ	はい（最小）	はい（推奨）	はい	1 ミリ秒未満の RTT。	該当なし	該当なし	該当なし
	オールフラッシュ クラスタ	なし	はい	はい（推奨）				
vSAN ストレッチ クラスタ	ハイブリッドまたはオールフラッシュ クラスタ	なし	はい（最小）	はい	各サイト内で 1 ミリ秒未満の RTT	10 GbE（ワークロードに依存）と 5 ミリ秒以下の RTT を推奨します。	200 ミリ秒未満の RTT。サイトあたり最大 10 台のホスト。	1,000 コンポーネントごとに 2 Mbps（45k コンポーネントで 100 Mbps まで）。
							100 ミリ秒未満の RTT。サイトあたり 11 ～ 15 台のホスト。	

表 3-2. vSAN OSA の最小 NIC 要件と推奨事項 （続き）

トポロジまたは展開モード	アーキテクチャ	1 GbE NIC のサポート	10 GbE NIC のサポート	10 GbE を超える NIC のサポート	ノード間の遅延	サイト間リンクのバンド幅または遅延	ノードと vSAN 監視ホスト間の遅延	ノードと vSAN 監視ホスト間のバンド幅
2 ノード vSAN クラスタ	ハイブリッド クラスタ	はい（10 台までの仮想マシン）	はい（推奨）	はい	1 つのサイト内で 1 ミリ秒未満の RTT	10 GbE と 5 ミリ秒以下の RTT を推奨します。	500 ミリ秒未満の RTT。	1,000 コンポーネントごとに 2 Mbps（最大 1.5 Mbps）。
	オールフラッシュ クラスタ	なし	はい（最小）					

表 3-3. vSAN ESA の最小 NIC 要件と推奨事項

デプロイ タイプ	1 GbE NIC のサポート	10 GbE NIC のサポート	10 GbE を超える NIC のサポート	ノード間の遅延	サイト間リンクのバンド幅または遅延	ノードと vSAN 監視ホスト間の遅延	ノードと vSAN 監視ホスト間のバンド幅
単一サイトの vSAN クラスタ	なし	はい	はい	1 ミリ秒未満の RTT。	該当なし	該当なし	該当なし
vSAN ストレッチ クラスタ	なし	はい	はい	各サイト内で 1 ミリ秒未満の RTT	少なくとも 10 GbE（ワークロードによる）と 5 ミリ秒の RTT。	200 ミリ秒未満の RTT。サイトあたり最大 10 台のホスト。	1,000 コンポーネントごとに 2 Mbps（45k コンポーネントで 100 Mbps まで）。
						100 ミリ秒未満の RTT。サイトあたり 11 ～ 15 台のホスト。	
2 ノード vSAN クラスタ	なし	はい	はい	1 つのサイト内で 1 ミリ秒未満の RTT	25 GbE と 5 ミリ秒以下の RTT を推奨します。	500 ミリ秒未満の RTT。	1,000 コンポーネントごとに 2 Mbps（最大 1.5 Mbps）。

注： この NIC 要件は、ハイパーコンバージド環境でパケット ロスが 0.0001% を超えていないことを前提としています。超過した場合、vSAN のパフォーマンスに大きな影響を与える可能性があります。

vSAN ストレッチ クラスタ NIC の要件の詳細については、『vSAN ストレッチ クラスタ ガイド』を参照してください。

バンド幅と遅延の要件

高パフォーマンスと高可用性を維持するため、vSAN クラスタはバンド幅とネットワーク遅延について特定の要件を満たしている必要があります。

vSAN ストレッチ クラスタのプライマリ サイトとセカンダリ サイトのバンド幅の要件は、vSAN のワークロード、データ量、障害処理の方法によって異なります。詳細については、『VMware vSAN 設計とサイジング ガイド』を参照してください。

表 3-4. バンド幅と遅延の要件

サイト通信	バンド幅	遅延
サイトからサイト	vSAN OSA：10 Gbps 以上 vSAN ESA：10 Gbps 以上	遅延が 5 ミリ秒 RTT 未満。
サイトから監視	1,000 個の vSAN コンポーネントごとに 2 Mbps	■ サイトあたりのホスト数が 1 台の場合、遅延が 500 ミリ秒 RTT 未満。 ■ サイトあたりのホスト数が 10 台まで場合、遅延が 200 ミリ秒 RTT 未満。 ■ サイトあたりのホスト数が 11 ～ 15 台の場合、遅延が 100 ミリ秒 RTT 未満。

レイヤー 2 とレイヤー 3 のサポート

サブネットを共有するすべての vSAN ホスト間の接続に、レイヤー 2 接続を使用することを推奨します。

vSAN は、vSAN ホスト間のルーティングにレイヤー 3 接続を使用する環境もサポートします。トラフィックのルーティング中に発生するホップ数と追加の遅延を考慮する必要があります。

表 3-5. レイヤー 2 とレイヤー 3 のサポート

クラスタ タイプ	L2 のサポート	L3 のサポート	考慮事項
ハイブリッド クラスタ	はい	はい	L2 を推奨、L3 もサポート。
オールフラッシュ クラスタ	はい	はい	L2 を推奨、L3 もサポート。
vSAN ストレッチ クラスタ データ	はい	はい	データサイト間で L2 と L3 の両方をサポート。
vSAN ストレッチ クラスタ監視	なし	はい	L3 をサポート。データ サイトと監視サイト間の L2 はサポートされません。
2 ノード vSAN クラスタ	はい	はい	データサイト間で L2 と L3 の両方をサポート。

ルーティングとスイッチの要件

vSAN ストレッチ クラスタの 3 つのサイトはすべて、管理ネットワークと vSAN ネットワークを通じて通信を行います。すべてのデータ サイトにある仮想マシンは、共通の仮想マシン ネットワークを通じて通信します。

vSAN ストレッチ クラスタのルーティング要件は次のとおりです。

表 3-6. ルーティング要件

サイト通信	展開モデル	レイヤー	ルーティング
サイトからサイト	デフォルト	レイヤー 2	必須ではない
サイトからサイト	デフォルト	レイヤー 3	スタティック ルートまたはゲートウェイ オーバーライドを使用します。
サイトから監視	デフォルト	レイヤー 3	スタティック ルートまたはゲートウェイ オーバーライドを使用します。
サイトから監視	監視トラフィックの分離	レイヤー 3	管理 (vmk0) インターフェイス以外のインターフェイスを使用する場合は、スタティック ルートまたはゲートウェイ オーバーライドが必要です。
サイトから監視	監視トラフィックの分離	2 ホスト クラスタの場合はレイヤー 2	スタティック ルートは不要です。

仮想スイッチの要件

vSphere Standard Switch または vSphere Distributed Switch のいずれかを使用して、vSAN ネットワークを作成できます。vSAN トラフィックのバンド幅の優先順位を付けるには、Distributed Switch を使用します。vSAN は、すべての vCenter Server バージョンで Distributed Switch を使用します。

次の表に、標準スイッチを使用した場合と Distributed Switch を使用した場合の違いを示します。

表 3-7. 仮想スイッチのタイプ

設計要件	オプション 1：vSphere Distributed Switch	オプション 2：vSphere Standard Switch	説明
可用性	影響なし	影響なし	どちらのオプションも使用できます。
管理性	利点	欠点	Distributed Switch はすべてのホストで一元的に管理されますが、標準スイッチは各ホストで個別に管理されます。
パフォーマンス	利点	欠点	Distributed Switch には追加の制御機能があります。たとえば、Network I/O Control を使用すると、vSAN トラフィックのパフォーマンスを維持できます。

表 3-7. 仮想スイッチのタイプ （続き）

設計要件	オプション 1：vSphere Distributed Switch	オプション 2：vSphere Standard Switch	説明
復旧性	利点	欠点	Distributed Switch の構成はバックアップとリストアに対応していますが、標準スイッチでこの機能を使用することはできません。
セキュリティ	利点	欠点	Distributed Switch には、トラフィックを保護する追加のセキュリティ コントロールが組み込まれています。

vSAN ネットワーク ポートの要件

vSAN の展開には、アクセスとサービスを提供するための特定のネットワーク ポートと設定が必要です。

vSAN は、クラスタ内の各ホストの特定のポートでメッセージを送信します。ホストのファイアウォールがこれらのポートでのトラフィックを許可していることを確認します。サポート対象のすべての vSAN ポートとプロトコルのリストについては、の VMware Ports and Protocols ポータル (https://ports.vmware.com/) を参照してください。

ファイアウォールの考慮事項

クラスタで vSAN を有効にすると、必要なすべてのポートが ESXi ファイアウォール ルールに追加され、自動的に設定されます。管理者がファイアウォール ポートを開いたり、ファイアウォール サービスを手動で有効にしたりする必要はありません。

受信接続と送信接続用に開いているポートを確認できます。ESXi ホストを選択して、[構成] > [セキュリティ プロファイル] をクリックします。

ネットワーク ファイアウォールの要件

ネットワーク ファイアウォールを構成する場合は、展開する vSAN のバージョンを検討します。

クラスタで vSAN を有効にすると、必要なすべてのポートが ESXi ファイアウォール ルールに追加され、自動的に設定されます。ファイアウォール ポートを開いたり、ファイアウォール サービスを手動で有効にしたりする必要はありません。ESXi ホスト セキュリティプロファイル（[構成] > [セキュリティ プロファイル]） で、受信接続と送信接続用に開いているポートを確認できます。

[vsanEncryption ファイアウォール ルール]

クラスタで vSAN 暗号化を使用している場合は、ホストと KMS サーバ 間の通信を考慮する必要があります。

vSAN 暗号化を使用するには、外部のキー管理サーバ (KMS) が必要です。vCenter Server は KMS からキー ID を取得し、ESXi ホストに配布します。KMS サーバと ESXi ホストは相互に直接通信を行います。KMS サーバが異なるポート番号を使用している場合があるため、vsanEncryption ファイアウォール ルールを使用すると、各 vSAN ホストと KMS サーバ間の通信を簡素化できます。これにより、vSAN ホストは KMS サーバ上の任意のポート（TCP ポート 0 ～ 65535）と直接通信できます。

ホストが KMS サーバとの通信を確立すると、次の操作が実行されます。

- KMS サーバの IP アドレスが vsanEncryption ルールに追加され、ファイアウォール ルールが有効になります。
- 交換時に vSAN ノードと KMS サーバ間の通信が確立されます。
- vSAN ノードと KMS サーバ間の通信が行われた後、vsanEncryption ルールから IP アドレスが削除され、ファイアウォール ルールが再度無効になります。

vSAN ホストは、同じルールを使用して複数の KMS ホストと通信できます。

vSAN ネットワークでのユニキャストの使用

4

ユニキャスト トラフィックとは、ネットワーク内のある地点から別の地点に 1 対 1 で送信を行うことを意味します。ネットワークの設計と展開を簡素化するため、vSAN バージョン 6.6 以降ではユニキャストを使用しています。

すべての ESXi ホストがユニキャスト トラフィックを使用し、vCenter Server がクラスタ メンバーシップのソースになります。vSAN ノードは、vCenter Server から提供される最新のホスト メンバーシップ リストを使用して自動的に更新されます。vSAN は、CMMDS の更新でユニキャスト通信を行います。

vSAN バージョン 6.6 以前のリリースでは、ハートビートを有効にし、クラスタ内のホスト間でメタデータをやり取りするには、マルチキャストが必要です。vSAN クラスタにそれ以前のバージョンのソフトウェアを実行しているホストがある場合は、マルチキャスト ネットワークが必要です。マルチキャストからユニキャスト ネットワークへの切り替えにより、パフォーマンスとネットワークのサポートが向上します。マルチキャストの詳細については、13 章 vSAN ネットワークでのマルチキャストの使用を参照してください。

次のトピックを参照してください。

- バージョン 5 より前のディスク グループの動作
- バージョン 5 のディスク グループの動作
- ユニキャスト ネットワークでの DHCP サポート
- ユニキャスト ネットワークでの IPv6 サポート
- ESXCLI を使用したユニキャストのクエリ
- クラスタ内のトラフィック

バージョン 5 より前のディスク グループの動作

vSAN バージョン 6.6 ディスク グループ内でバージョン 5 のディスク グループが 1 つだけ使用可能な場合、クラスタは常にユニキャスト モードで通信を行います。

vSAN バージョン 6.6 のクラスタの場合、次の状況が発生すると、マルチキャスト通信に自動的に戻ります。

- すべてのクラスタ ホストがバージョン 6.5 以前の vSAN を実行している。
- すべてのディスク グループが、バージョン 3 以前のオンディスクを使用している。
- vSAN 6.2 や vSAN 6.5 など、vSAN 6.6 以外のホストがクラスタに追加されている。

たとえば、vSAN 6.5 以前を実行しているホストが既存の vSAN 6.6 クラスタに追加されると、クラスタはマルチキャスト モードに戻り、6.5 ホストが有効なノードとして追加されます。この動作を回避するには、ESXi ホストとオンディスク フォーマットの両方で最新バージョンを使用します。vSAN クラスタがユニキャスト モードでの通信を継続し、マルチキャストに戻らないようにするには、vSAN 6.6 ホストのディスク グループをオンディスク バージョン 5.0 にアップグレードします。

注： 同じクラスタで vSAN バージョン 6.5 以前と vSAN バージョン 6.6 以降を共存できる混合モードは使用しないでください。

バージョン 5 のディスク グループの動作

vSAN バージョン 6.6 クラスタ内にバージョン 5 のディスク グループが 1 つしか存在しない場合、クラスタは常にユニキャスト モードで通信を行います。

vSAN 6.6 クラスタがすでにオンディスク バージョン 5 を使用していて、vSAN 6.5 ノードがクラスタに追加されている環境では、次のイベントが発生します。

- vSAN 6.5 ノードは、独自のネットワーク パーティションを形成します。
- vSAN 6.5 ノードは引き続きマルチキャスト モードで通信を行いますが、ユニキャストモードを使用する vSAN 6.6 ノードと通信を行うことはできません。

オンディスク フォーマットについてクラスタのサマリ警告が表示され、1 つのノードが以前のバージョンであることが示されます。ノードを最新バージョンにアップグレードできます。クラスタが混合モードの場合、ディスク フォーマットのバージョンをアップグレードすることはできません。

ユニキャスト ネットワークでの DHCP サポート

vSAN 6.6 クラスタに展開された vCenter Server は、予約機能のない Dynamic Host Configuration Protocol (DHCP) からの IP アドレスを使用できます。

予約機能付きの DHCP を使用する理由は、割り当てられた IP アドレスが VMkernel ポートの MAC アドレスに関連付けられるためです。

ユニキャスト ネットワークでの IPv6 サポート

vSAN 6.6 は、ユニキャスト通信で IPv6 をサポートします。

IPv6 では、リンクローカル プリフィックスを使用して、任意のインターフェイスにリンクローカル アドレスが自動的に設定されます。デフォルトでは、vSAN はノードのリンク ローカル アドレスを他の隣接クラスタ ノードに追加しません。このため、vSAN 6.6 では、ユニキャスト通信に IPv6 リンク ローカル アドレスを使用できません。

ESXCLI を使用したユニキャストのクエリ

ESXCLI コマンドを実行して、ユニキャスト構成を特定できます。

通信モードの表示

`esxcli vsan cluster get` コマンドを使用すると、vSAN クラスタ ノードの CMMDS モード（ユニキャストまたはマルチキャスト）を表示できます。

手順

◆ `esxcli vsan cluster get` コマンドを実行します。

結果

```
Cluster Information
  Enabled: true
  Current Local Time: 2020-04-09T18:19:52Z
  Local Node UUID: 5e8e3dc3-43ab-5452-795b-a03d6f88f022
  Local Node Type: NORMAL
  Local Node State: AGENT
  Local Node Health State: HEALTHY
  Sub-Cluster Master UUID: 5e8e3d3f-3015-9075-49b6-a03d6f88d426
  Sub-Cluster Backup UUID: 5e8e3daf-e5e0-ddb6-a523-a03d6f88dd4a
  Sub-Cluster UUID: 5282f9f3-d892-3748-de48-e2408dc34f72
  Sub-Cluster Membership Entry Revision: 11
  Sub_cluster Member Count: 5
  Sub-Cluster Member UUIDs: 5e8e3d3f-3015-9075-49b6-a03d6f88d426, 5e8e3daf-e5e0-ddb6-a523-
a03d6f88dd4a,
  5e8e3d73-6d1c-0b81-1305-a03d6f888d22, 5e8e3d33-5825-ee5c-013c-a03d6f88ea4c,
5e8e3dc3-43ab-5452-795b-a03d6f88f022
  Sub-Cluster Member HostNames: testbed-1.vmware.com, testbed2.vmware.com,
  testbed3.vmware.com, testbed4.vmware.com, testbed5.vmware.com
  Sub-Cluster Membership UUID: 0f438e5e-d400-1bb2-f4d1-a03d6f88d426
  [ユニキャスト モード有効:true]
  Maintenance Mode State: OFF
  Config Generation: ed845022-5c08-48d0-aa1d-6b62c0022222 7 2020-04-08T22:44:14.889
```

vSAN クラスタ ホストの確認

`esxcli vsan cluster unicastagent list` コマンドを使用して、vSAN クラスタ ホストがユニキャスト モードで動作しているかどうかを確認します。

手順

◆ `esxcli vsan cluster unicastagent list` コマンドを実行します。

結果

```
NodeUuid                              IsWitness Supports Unicast IP Address  Port  Iface Name
Cert Thumbprint  SubClusterUuid
------------------------------------ --------- ---------------- ----------  ----- ----------
5e8e3d73-6d1c-0b81-1305-a03d6f888d22         0      true 10.198.95.10
12321
43:80:B7:A1:3F:D1:64:07:8C:58:01:2B:CE:A2:F5:DE:D6:B1:41:AB
5e8e3daf-e5e0-ddb6-a523-a03d6f88dd4a         0      true 10.198.94.240
12321
```

```
FE:39:D7:A5:EF:80:D6:41:CD:13:70:BD:88:2D:38:6C:A0:1D:36:69
5e8e3d3f-3015-9075-49b6-a03d6f88d426         0      true 10.198.94.244
12321
72:A3:80:36:F7:5D:8F:CE:B0:26:02:96:00:23:7D:8E:C5:8C:0B:E1
5e8e3d33-5825-ee5c-013c-a03d6f88ea4c         0      true 10.198.95.11
12321
5A:55:74:E8:5F:40:2F:2B:09:B5:42:29:FF:1C:95:41:AB:28:E0:57
```

出力には、vSAN ノード UUID、IPv4 アドレス、IPv6 アドレス、vSAN ノードが通信する UDP ポート、ノードがデータホスト (0) か Witness（監視）ホスト (1) かが表示されます。この出力を使用して、ユニキャスト モードで動作している vSAN クラスタ ノードを特定し、クラスタ内の他のホストを確認できます。vCenter Server は、出力リストを保持します。

vSAN ネットワーク情報の表示

`esxcli vsan network list` コマンドを使用して、vSAN が通信に使用する VMkernel インターフェイス、ユニキャストポート (12321)、vSAN インターフェイスに関連付けられたトラフィック タイプ（vSAN または監視）などの vSAN ネットワーク情報を表示します。

手順

- `esxcli vsan network list` コマンドを実行します。

結果

```
Interface
   VmkNic Name: vmk1
   IP Protocol: IP
   Interface UUID: e290be58-15fe-61e5-1043-246e962c24d0
   Agent Group Multicast Address: 224.2.3.4
   Agent Group IPv6 Multicast Address: ff19::2:3:4
   Agent Group Multicast Port: 23451
   Master Group Multicast Address: 224.1.2.3
   Master Group IPv6 Multicast Address: ff19::1:2:3
   Master Group Multicast Port: 12345
   [Host Unicast Channel Bound Port: 12321]
   Multicast TTL: 5
   Traffic Type: vsan
```

この出力には、マルチキャスト情報も表示されます。

クラスタ内のトラフィック

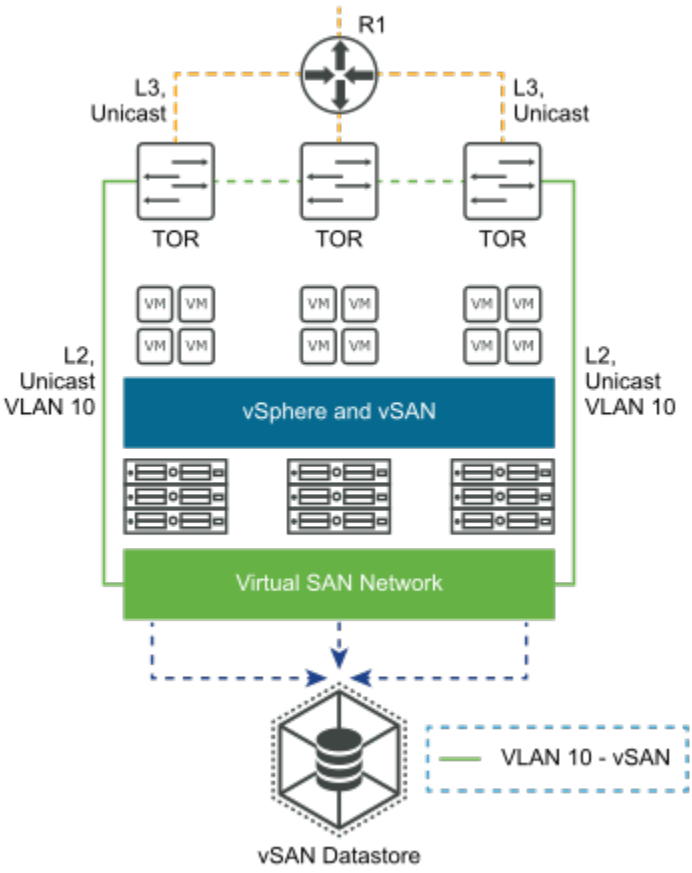
ユニキャスト モードの場合、プライマリ ノードはクラスタ内のすべての vSAN ノードに同じメッセージを送信するときに、すべてのクラスタ ノードにメッセージを送信します。

たとえば、vSAN ノードの数が N の場合、プライマリ ノードは N 回そのメッセージを送信します。その結果、vSAN CMMDS トラフィックがわずかに増加します。通常の安定した状態では、トラフィックがわずかに増加していることに気付かない可能性があります。

単一ラックでのクラスタ内のトラフィック

vSAN クラスタ内のすべてのノードが同じトップオブラック (TOR) のスイッチに接続されている場合、トラフィックの増加は、プライマリ ノードとスイッチの間でのみ発生します。

vSAN クラスタが複数の TOR スイッチにまたがる場合は、スイッチ間のトラフィックが増加します。クラスタが複数のラックにまたがる場合、ラックを認識するために複数の TOR でフォルト ドメイン (FD) が形成されます。プライマリ ノードは、ラックまたはフォルト ドメインに N 個のメッセージを送信します（N は各フォルト ドメインのホスト数）。

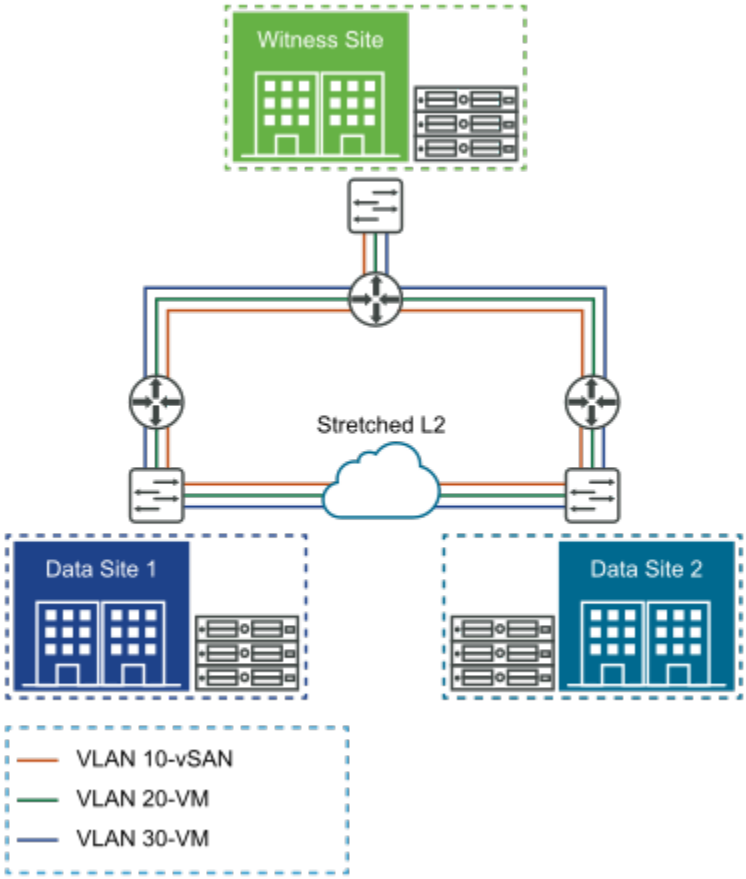


vSAN ストレッチ クラスタ内のクラスタ内トラフィック

vSAN ストレッチ クラスタでは、プライマリ ノードは優先サイトにあります。

フォルト ドメインでは、CMMDS データをセカンダリ サイトから優先サイトに送信する必要があります。vSAN ストレッチ クラスタ内のトラフィックを計算するには、セカンダリ サイトのノード数に CMMDS ノードのサイズ (MB) を掛け、その結果にセカンダリ サイトのノード数を掛ける必要があります。

vSAN ストレッチ クラスタ内のトラフィック = セカンダリ サイト内のノード数 * CMMDS ノード サイズ (MB) * セカンダリ サイトのノード数。



ユニキャスト トラフィックでは、監視サイトのトラフィック要件は変わりません。

IP ネットワーク転送の構成

5

トランスポート プロトコルはネットワーク全体に通信サービスを提供します。このサービスには、TCP/IP スタックやフロー制御が含まれます。

次のトピックを参照してください。

- vSphere TCP/IP スタック
- Object Missing
- IPv6 サポート
- スタティック ルート
- ジャンボ フレーム

vSphere TCP/IP スタック

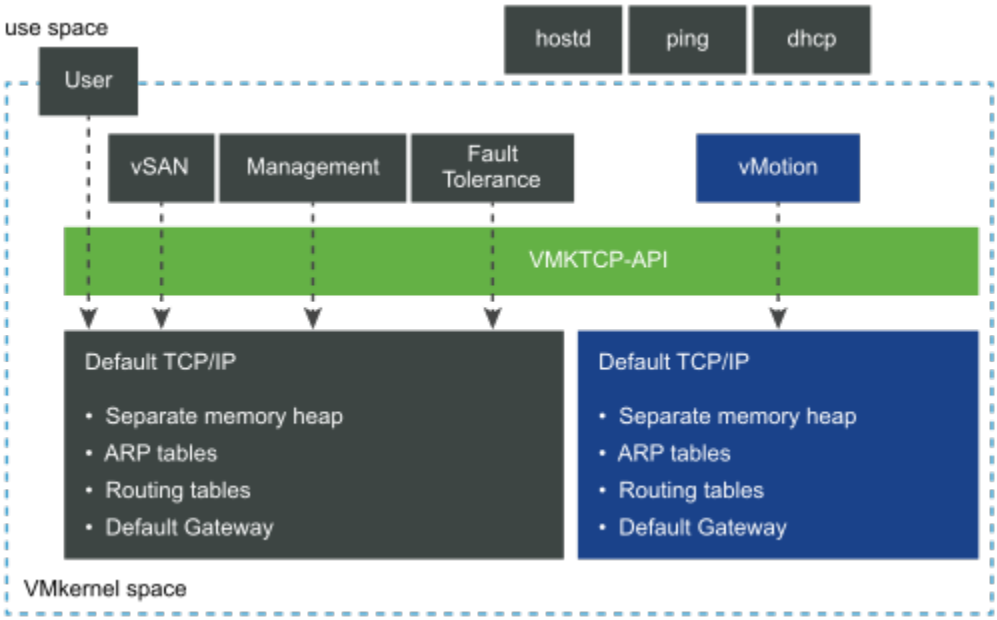
vSphere には、vSAN トラフィック サービス専用の TCP/IP スタックはありません。デフォルトの TCP/IP スタックに vSAN VMkernel ネットワーク インターフェイスを追加し、vSAN クラスタ内のすべてのホストにスタティック ルートを定義できます。

vSphere は、カスタム vSAN TCP/IP スタックの作成をサポートしていません。レイヤー 3 ネットワーク トポロジーの vSAN トラフィックが vSAN VMkernel ネットワーク インターフェイス上に残るようにすることができます。デフォルトの TCP/IP スタックに vSAN VMkernel ネットワーク インターフェイスを追加し、vSAN クラスタ内のすべてのホストにスタティック ルートを定義します。

注： vSAN に独自の TCP/IP スタックがありません。スタティック ルートを使用して、L3 ネットワーク間で vSAN トラフィックをルーティングします。

vSphere 6.0 では、新しい TCP/IP スタック アーキテクチャが導入されました。これにより、複数の TPC/IP スタックを使用して、さまざまな VMkernel ネットワーク インターフェイスを管理できます。このアーキテクチャでは、隔離された TCP/IP スタックに、vMotion、管理、フォルト トレランスなどのトラフィック サービスを設定し、複数のデフォルト ゲートウェイを使用することができます。

ネットワーク トラフィックの隔離とセキュリティの要件を満たすため、異なるネットワーク セグメント (VLAN) に異なるトラフィック サービスを展開します。これにより、異なるトラフィック サービスが同じデフォルト ゲートウェイを通過しないように設定できます。



個々の TCP/IP スタックでトラフィック サービスを設定する場合は、各トラフィック サービス タイプをそれぞれのネットワーク セグメントに展開します。ネットワーク セグメントは、VLAN セグメンテーションの物理ネットワーク アダプタを介してアクセスされます。各セグメントは、それぞれのトラフィック サービスが有効になっている異なる VMkernel ネットワーク インターフェイスにマッピングします。

vSphere で使用可能な TCP/IP スタック

vSphere は、vSAN トラフィック要件をサポートする TCP/IP スタックを提供します。

- [デフォルトの TCP/IP スタック]。ホスト関連のトラフィック サービスを管理します。このスタックは、設定されたすべてのネットワーク サービス間で単一のデフォルト ゲートウェイを共有します。
- [vMotion TCP/IP スタック]。vMotion トラフィックを専用スタックに分離します。このスタックを使用すると、デフォルトの TCP/IP スタックから vMotion トラフィックが完全に削除されるか、無効になります。
- [プロビジョニング TCP/IP スタック]。コールド移行、クローン作成、スナップショット、NFC トラフィックなど、一部の仮想マシン関連の操作を分離します。
- [ミラー TCP/IP スタック]。ポート ミラーリング トラフィックを管理トラフィックから分離します。このスタックがない場合、ミラー トラフィックはデフォルトの TCP/IP スタックにバインドされます。
- [ops TCP/IP スタック]。vSphere ネットワーク フロー データ収集のサポートを提供します。

VMkernel インターフェイスの作成時に、異なる TCP/IP スタックを選択できます。

vSphere トラフィック サービスに隔離されたネットワークの要件が存在する環境では、同じデフォルト ゲートウェイを使用してトラフィックを処理することはできません。異なる TCP/IP スタックを使用すると異なるデフォルト ゲートウェイを使用でき、スタティック ルートの追加を回避できるため、トラフィックの隔離管理が簡素化されます。デフォルト ゲートウェイを介してアクセスできない別のネットワークに vSAN トラフィックをルーティングする必要がある場合は、この方法を使用します。

Object Missing

This object is not available in the repository.

IPv6 サポート

vSAN 6.2 以降では、IPv6 がサポートされます。

vSAN では、次の IP バージョンがサポートされます。

- IPv4
- IPv6（vSAN 6.2 以降）
- IPv4/IPv6 の混合（vSAN 6.2 以降）

vSAN 6.2 より前のリリースでは、IPv4 のみがサポートされます。vSAN クラスタを IPv4 から IPv6 に移行する場合は、混合モードを使用します。

IPv6 マルチキャストもサポートされます。

IPv6 の使用方法については、ネットワークベンダーに問い合わせてください。

スタティック ルート

スタティック ルートを使用すると、vSAN ネットワーク インターフェイスを介して 1 つのサブネットのホストから別のネットワーク上のホストに接続することができます。

大半の組織は、vSAN ネットワークを管理ネットワークから分離しているため、vSAN ネットワークにデフォルト ゲートウェイはありません。L3 展開では、異なるサブネットまたは異なる L2 セグメントにあるホスト同士は、通常管理ネットワークに関連付けられているデフォルト ゲートウェイを介して通信を行うことができません。

vSAN ネットワーク インターフェイスを介して 1 つのサブネットのホストから別のネットワーク上のホストの vSAN ネットワークに接続する場合は、スタティック ルートを使用します。スタティック ルートは、デフォルト ゲートウェイではなく、インターフェイスを介して特定のネットワークに接続する方法をホストに指示します。

次の例は、ESXi ホストに IPv4 スタティック ルートを追加する方法を示しています。ゲートウェイ (-g) と、そのゲートウェイを介してアクセスするネットワーク (-n) を指定します。

```
esxcli network ip route ipv4 add –g 172.16.10.253 -n 192.168.10.0/24
```

スタティック ルートが追加されている場合、物理インフラストラクチャで許可されていれば、すべてのネットワークで vSAN トラフィック接続を使用できます。`vmkping` コマンドを実行して、リモート ネットワークの IP アドレスまたはデフォルト ゲートウェイに ping を実行し、異なるネットワーク間の通信をテストして確認します。また、異なるサイズのパケット (-s) をチェックし、パケットの断片化 (-d) を防ぐことができます。

```
vmkping –I vmk3 192.168.10.253
```

ジャンボ フレーム

vSAN は、vSAN ネットワーク上のジャンボ フレームを完全にサポートしています。

ジャンボ フレームは、1,500 バイトを超えるペイロードを持つイーサネット フレームです。通常、ジャンボ フレームは 9,000 バイトのペイロードを送信しますが、フレームのサイズはさまざまです。

ジャンボ フレームを使用すると、CPU 使用率を低減し、スループットを向上させることができます。

注： vSAN Max 環境でジャンボ フレームのサポートを有効にして、パフォーマンスを向上させます。

このようなメリットが、ネットワーク全体にジャンボ フレームを実装するオーバーヘッドを上回るかどうかを判断する必要があります。ネットワーク インフラストラクチャでジャンボ フレームが有効になっているデータセンターでは、vSAN でジャンボ フレームを使用できます。ネットワーク全体でジャンボ フレームを構成する場合の運用コストが、CPU とパフォーマンスのメリットを上回る可能性もあります。

VMware NSX と vSAN の併用

6

vSAN と VMware NSX は、同じ vSphere インフラストラクチャに展開し、共存させることができます。

NSX は、NSX で管理された VXLAN または Geneve オーバーレイを介した vSAN データ ネットワークの構成をサポートしていません。

vSAN と NSX には互換性があります。vSAN と NSX は、それぞれ独立して機能、リソース、サービスを提供します。

ただし、vSAN ネットワーク トラフィックを NSX 管理の VxLAN/ Geneve オーバーレイに配置することはできません。NSX は、NSX で管理された VxLAN/Geneve オーバーレイを介した vSAN データ ネットワーク トラフィックの構成をサポートしていません。

NSX で管理されている VxLAN オーバーレイを介して VMkernel トラフィックがサポートされない理由の 1 つは、VMkernel ネットワークとそれらがサポートする VxLAN オーバーレイ間の循環依存関係を回避するためです。NSX が管理する VxLAN オーバーレイとともに提供される論理ネットワークは仮想マシンで使用されます。これには、ネットワークのモビリティと柔軟性が必要です。

NSX で LACP/LAG を実装する場合、Cisco Nexus 環境は LAG を仮想ポート チャネル (vPC) として定義します。

輻輳制御とフロー制御の使用 7

フロー制御では、vSAN ネットワーク上の送信側と受信側間のデータ転送速度を管理します。輻輳制御では、ネットワーク内の輻輳を処理します。

フロー制御

フロー制御では、2 つのデバイス間のデータ転送速度を管理できます。

フロー制御は、物理的に接続された 2 つのデバイスがオートネゴシエーションを行うときに構成されます。

ネットワーク ノードが要求を処理しきれなくなると、一時停止フレームを送信し、送信側からの転送を一定期間停止することがあります。マルチキャスト宛先アドレスを含むフレームがスイッチに送信されると、このフレームはスイッチの他のすべてのポートを介して転送されます。一時停止フレームには、他のマルチキャスト トラフィックと区別するために特別なマルチキャスト宛先アドレスが設定されています。準拠しているスイッチは一時停止フレームを転送しません。この範囲に送信されたフレームはスイッチ内でのみ処理されます。一時停止フレームの持続期間は制限されており、一定の間隔が経過すると期限切れになります。スイッチを介して接続されている 2 台のコンピュータは、互いに一時停止フレームを送信しませんが、スイッチには一時停止フレームを送信できます。

一時停止フレームを使用する理由の 1 つは、最高速度での受信に対応できる十分なバッファのないネットワーク インターフェイス コントローラ (NIC) をサポートするためです。バス速度とメモリ サイズの進歩で、この問題はあまり発生しません。

輻輳制御

輻輳制御は、ネットワーク上のトラフィックを制御するのに役立ちます。

輻輳制御は主にパケット交換ネットワークで使用されます。スイッチ内のネットワークの輻輳は、スイッチ間のリンクにが過負荷状態になると発生する可能性があります。スイッチ間のリンクが物理レイヤーの処理能力を超えると、スイッチは自身を保護するために一時停止フレームを送信します。

物理フロー制御

Priority-based flow control (PFC) を使用すると、輻輳によるフレームの損失をなくすことができます。

Priority-based flow control (IEEE 802.1Qbb) は、フレームの一時停止と類似したメカニズムで実現されていますが、この制御は個々の優先順位に基づいて行われます。PFC は、Class-Based Flow Control (CBFC) または Per Priority Pause (PPP) ともいいます。

フロー制御と輻輳制御

フロー制御は、送信側と受信側の間のトラフィックを制御するエンドツーエンドのメカニズムです。フロー制御は、データ リンク レイヤーとトランスポート レイヤーで行われます。

輻輳制御は、ネットワークの輻輳を制御するためにネットワークで使用されます。バス速度とメモリ サイズが進歩した最近のネットワークで、この問題が発生することはまれです。ただし、スイッチ内でネットワーク輻輳が起きる可能性は考えられます。輻輳制御は、ネットワーク レイヤーとトランスポート レイヤーによって処理されます。

フロー制御を設計する際の考慮事項

デフォルトでは、ESXi ホストのすべてのネットワーク インターフェイスでフロー制御が有効になっています。

NIC のフロー制御の設定はドライバによって実行されます。NIC がネットワーク トラフィックを処理しきれなくなると、NIC は一時停止フレームを送信します。

一時停止フレームなどのフロー制御メカニズムでは、vSAN ネットワーク レイヤーでの遅延が増えるため、仮想マシンのゲスト I/O で全体的な遅延が発生する可能性があります。一部のネットワーク ドライバには、ドライバ内のフロー制御機能を構成するモジュール オプションが用意されています。ネットワーク ドライバによっては、ESXi ホストのコンソールで `ethtool` コマンドライン ユーティリティを使用して、構成オプションを変更できます。ドライバの実装の詳細に応じて、モジュール オプションまたは `ethtool` を使用します。

ESXi ホストでのフロー制御の構成については、VMware KB1013413 を参照してください。

1 Gbps の環境では、ESXi ネットワーク インターフェイスでフロー制御を有効にします（デフォルト）。一時停止フレームが問題になっている場合は、ハードウェア ベンダーのサポートまたは VMware グローバル サポート サービス (GSS) に連絡して、フロー制御の無効化を慎重に計画してください。

受信側から ESXi ホストに送信される一時停止フレームの存在を確認する方法については、12 章 vSAN ネットワークのトラブルシューティングを参照してください。通常、環境内で多くの一時停止フレームが発生している場合は、基盤となるネットワークまたは転送の問題を調査する必要があります。

基本的な NIC チーミング、フェイルオーバー、ロード バランシング

8

多くの vSAN 環境では、いくつかのレベルでネットワークの冗長性を実装する必要があります。

NIC チーミングを使用すると、ネットワークの冗長性を実現できます。2 つ以上のネットワーク アダプタ (NIC) をチームとして構成し、高可用性とロード バランシングに使用します。vSphere ネットワークでは基本的な NIC チーミングを使用できます。これらの技術は、vSAN の設計とアーキテクチャに影響を与える可能性があります。

いくつかの NIC チーミング オプションを使用できます。物理スイッチ構成を必要とする NIC チーミング ポリシー、またはリンク集約などのネットワークの概念を理解する必要がある NIC チーミング ポリシーは使用しないでください。最適な結果を得るには、基本的で、シンプルで、信頼性の高い設定を行います。

NIC チーミング オプションがよくわからない場合は、明示的なフェイルオーバー のアクティブ/スタンバイ構成を使用してください。

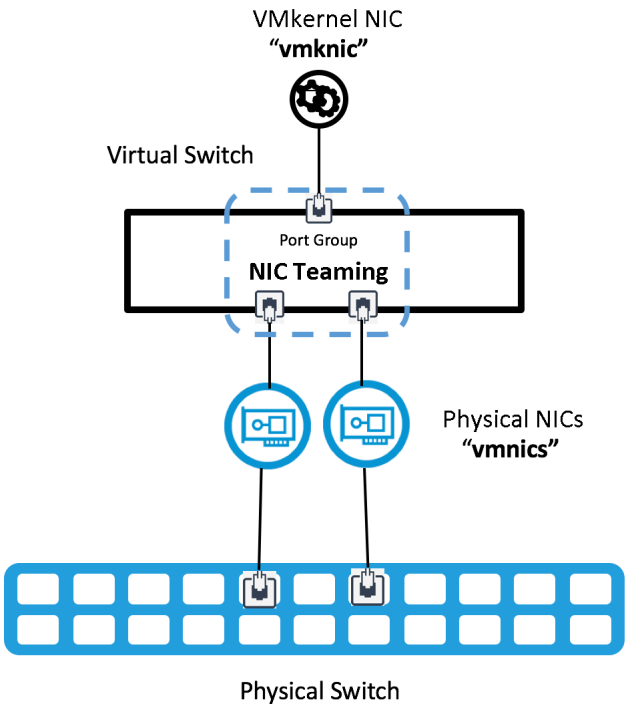
次のトピックを参照してください。

- 基本的な NIC チーミング

基本的な NIC チーミング

基本的な NIC チーミングでは、複数の物理アップリンク、1 つの vmknic、1 台のスイッチを使用します。

vSphere NIC チーミングでは、vmnic という名前の複数のアップリンク アダプタを 1 台の仮想スイッチに関連付け、1 つのチームを形成します。これは最も基本的なオプションで、vSphere 標準スイッチまたは vSphere Distributed Switch を使用して構成できます。



フェイルオーバーと冗長性

vSAN では、vSphere から提供される基本的な NIC チーミングおよびフェイルオーバー ポリシーを使用できます。

vSwitch の NIC チーミングは、複数のアクティブ アップリンク構成またはアクティブ/スタンバイ アップリンク構成にできます。基本的な NIC チーミングでは、物理スイッチ レイヤーで特別な構成を行う必要ありません。

注： vSAN では、ロード バランシングに NIC チーミングは使用されません。

標準的な NIC チーミングの構成では、次の設定を行います。Distributed Switch を使用している場合は、vSAN トラフィックに使用する分散ポート グループの設定を編集します。

- ロード バランシング：発信元の仮想ポートに基づいたルーティング
- ネットワーク障害の検出：リンク ステータスのみ
- スイッチへの通知：はい
- フェイルバック：はい

vSAN トラフィックをロード バランシングします。

- ロード バランシング：発信元の仮想ポートに基づいたルーティング
- ネットワーク障害の検出：リンク ステータスのみ
- スイッチへの通知：はい
- フェイルバック：はい

NIC チームのロード バランシングの設定

NIC チーミングにはいくつかのロード バランシング手法があり、それぞれに長所と短所があります。

発信元の仮想ポートに基づいたルート

アクティブ/アクティブ構成またはアクティブ/パッシブ構成では、基本的な NIC チーミングに [発信元の仮想ポートに基づいたルート] を使用します。このポリシーが有効な場合、VMkernel ポートごとに 1 つの物理 NIC が使用されます。

[長所]

- これは、最小の物理スイッチ構成を必要とする最もシンプルな NIC チーミング方法です。
- この方法では 1 つのポートで vSAN トラフィックを処理するので、トラブルシューティングが簡単になります。

[短所]

- 単一の VMkernel インターフェイスは、1 つの物理 NIC のバンド幅に制限されます。標準的な vSAN 環境では、1 つの VMkernel アダプタを使用するため、チーム内で使用する物理 NIC は 1 つのみになります。

物理 NIC 負荷に基づいたルート

[物理 NIC 負荷に基づいたルート]は、[発信元の仮想ポート]に基づいたルートをベースにしています。このルートでは、仮想スイッチが、アップリンクの実際の負荷を監視し、負荷がかかりすぎているアップリンクで負荷を低減するための処理を行います。このロード バランシングの方法は vSphere Distributed Switch でのみ使用できます。Standard Switch では使用できません。

Distributed Switch は、ポート ID と NIC チーム内のアップリンク数を使用して、VMkernel ポートごとのアップリンクの負荷を計算します。Distributed Switch は 30 秒ごとにアップリンクをチェックして、その負荷が 75 パーセントを超えると、I/O の使用率が最も高い VMkernel ポートのポート ID を別のアップリンクに移動します。

[長所]

- 物理スイッチ構成は必要ありません。
- vSAN には 1 つの VMkernel ポートがありますが、他の VMkernel ポートまたはネットワーク サービスで同じアップリンクを共有できます。vSAN では、vMotion や管理などの他の競合サービスとは異なるアップリンクを使用することで、メリットを得ることができます。

[短所]

- 標準的な vSAN には VMkernel ポートが 1 つしか構成されていないため、その効果は制限されています。
- ESXi VMkernel は、時間間隔ごとにトラフィックの負荷を再評価します。これにより、処理オーバーヘッドが発生する可能性があります。

設定：ネットワーク障害の検出

デフォルトの設定（[リンク ステータスのみ]）を使用します。リンク障害の検出にビーコンの検出を使用しないでください。スプリットブレインを回避するため、ビーコンの検出を使用するには 3 つ以上の物理 NIC が必要です。詳細については、VMware KB1005577 を参照してください。

設定：スイッチへの通知

デフォルトの設定（[はい]）を使用します。物理スイッチには、各 MAC アドレスを物理スイッチ ポートに関連付ける MAC アドレス フォワーディング テーブルがあります。フレームを受信すると、スイッチはこのテーブルを使用して宛先の MAC アドレスを特定し、正しい物理ポートを決定します。

NIC のフェイルオーバーが発生した場合、ESXi ホストは、何らかの変更があったことをネットワーク スイッチに通知する必要があります。通知を行わないと、物理スイッチは引き続き古い情報を使用して、誤ったポートにフレームを送信する可能性があります。

スイッチへの通知を [はい] に設定すると、1 つの物理 NIC に障害が発生し、トラフィックがチーム内の別の物理 NIC にルーティングされたときに、仮想スイッチはネットワークを介して通知を送信し、物理スイッチの検索テーブルを更新します。

この設定では、VLAN の誤構成や、ネットワークでさらにアップストリームが発生するアップリンクの損失は確認できません。これらの問題は、vSAN ネットワーク パーティションの健全性チェックで検出できます。

設定：フェイルバック

このオプションは、障害から復旧したあとで、物理アダプタをどのようにアクティブ モードに戻すかを決定します。フェイルオーバー イベントにより、ネットワーク トラフィックが 1 つの NIC から別の NIC に移動します。フェイルバックを [はい] に設定すると、元の NIC で [リンクアップ] 状態が検出されたときに、トラフィックが自動的に元のネットワーク アダプタに戻ります。フェイルバックを [いいえ] に設定した場合は、手動でフェイルバックを行う必要があります。

状況によっては、フェイルバックを [いいえ] に設定したほうが便利な場合もあります。たとえば、物理スイッチ ポートが障害から復旧した後で、ポートはアクティブになってからトラフィックの転送を開始するまでに数秒かかることがあります。スパニング ツリー プロトコルを使用する特定の環境では、自動フェイルバックが原因で問題が発生することが確認されています。スパニング ツリー プロトコル (STP) の詳細については、VMware KB1003804 を参照してください。

フェイルオーバーの順序の設定

フェイルオーバーの順序により、通常の操作中にアクティブにするリンクと、フェイルオーバーの発生時にアクティブにするリンクが決まります。vSAN ネットワークでは、さまざまな構成がサポートされます。

[アクティブ/スタンバイ アップリンク] : アクティブ/スタンバイの設定で障害が発生した場合、NIC ドライバはアップリンク 1 のリンクダウン イベントを vSphere に通知します。スタンバイ アップリンク 2 がアクティブになり、アップリンク 2 でトラフィックが再開されます。

[アクティブ/アクティブ アップリンク] : フェイルオーバーの順序をアクティブ/アクティブに設定した場合、vSAN トラフィックで使用される仮想ポートは両方の物理ポートを同時に使用できません。

アップリンク 1 とアップリンク 2 の両方の NIC チーミング構成がアクティブになっている場合、スタンバイ アップリンクをアクティブにする必要はありません。

注： アクティブ/アクティブ構成を使用する場合は、フェイルバックが [いいえ] に設定されていることを確認します。詳細については、VMware KB2072928 を参照してください。

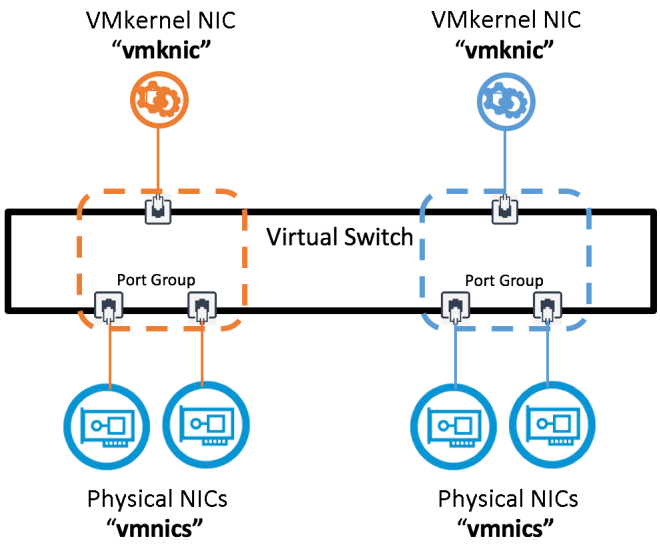
高度な NIC チーミング

9

vSAN ネットワークを構成するときに、複数の VMkernel アダプタを使用して高度な NIC チーミングを実装できます。リンク集約プロトコル (LAG/LACP) を使用すると、vSAN ネットワークを 1 つの VMkernel アダプタで構成できます。

高度な NIC チーミングを使用すると air gap ネットワークを実装できます。1 つのネットワーク パスで障害が発生しても、その影響が他のネットワーク パスに及ぶことはありません。1 つのネットワーク パスの一部で障害が発生した場合、他のネットワーク パスでトラフィックが処理されます。別の VLAN または物理ネットワーク ファブリックなどの異なるサブネットで、vSAN 用に複数の VMkernel NIC を構成します。

vSphere と vSAN では、同じサブネット上の複数の VMkernel アダプタ (vmknics) はサポートされません。詳細については、VMware KB 2010877 を参照してください。



次のトピックを参照してください。

- リンク集約グループの概要
- ネットワークの air gap について
- vSAN による air gap ネットワーク構成の長所と短所
- NIC チーミングの構成例

リンク集約グループの概要

LACP プロトコルを使用すると、ネットワーク デバイスはピアに LACP パケットを送信し、リンクを自動的に束ねるネゴシエーションを実行できます。

リンク集約グループ (LAG) は IEEE 802.1AX-2008 標準で定義されています。リンク集約によって 1 つ以上のリンクが集約され、リンク集約グループが形成されます。

LACP を使用して LAG 形成のネゴシエーションを行うことで、LAG を静的（手動）または動的として設定できます。LACP は次のように設定できます。

アクティブ

ポートが起動すると、デバイスはすぐに LACP メッセージを送信します。LACP が有効になっているエンド デバイス（たとえば、ESXi ホストや物理スイッチ）は、LACP メッセージというフレームを相互に送受信し、LAG を作成するネゴシエーションを行います。

パッシブ

デバイスは、ポートをパッシブ ネゴシエーション状態にします。この場合、ポートは受信した LACP メッセージにのみ応答しますが、ネゴシエーションは開始しません。

注： ホストとスイッチが両方ともパッシブ モードの場合、LAG は初期化されません。リンクをトリガするにはアクティブなパートが必要になります。少なくとも 1 つはアクティブにする必要があります。

vSphere 5.5 以降のリリースでは、この機能は [拡張 LACP] と呼ばれています。この機能は、vSphere Distributed Switch バージョン 5.5 以降でのみサポートされます。

vSphere Distributed Switch の LACP サポートの詳細については、『vSphere 6 ネットワーク』を参照してください。

注： 使用できる LAG の数は、基盤となる物理環境の機能性と、仮想ネットワークのトポロジによって異なります。

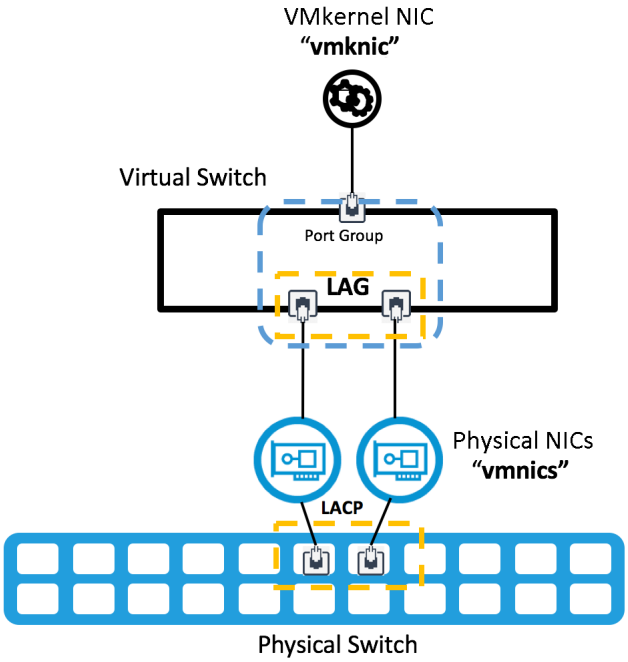
ロード バランシング オプションの詳細については、KB 2051826 を参照してください。

静的リンクと動的リンクの集約

LACP を使用すると、複数のネットワーク接続を結合または集約できます。

LACP が [アクティブ] または [動的] モードの場合、物理スイッチは ESXi ホストなどのネットワーク デバイスに LACP メッセージを送信し、リンク集約グループ (LAG) 作成のネゴシエーションを行います。

vSphere Standard Switch （および 5.5 以前の vSphere Distributed Switch）を使用してホストのリンク集約を設定するには、物理スイッチで静的チャネル グループを設定します。詳細については、ベンダーのドキュメントを参照してください。



動的リンク集約の長所と短所

動的リンク集約の使用にはトレードオフがあります。

[長所]

[パフォーマンスとバンド幅が向上します]。1 つの vSAN ホストまたは VMkernel ポートは、さまざまなロード バランシング オプションを使用して、他の多くの vSAN ホストと通信できます。

[ネットワーク アダプタの冗長性を提供します]。NIC で障害が発生し、リンク状態が停止すると、チーム内の残りの NIC は引き続きトラフィックを通過させます。

[トラフィック バランシングが向上します]。障害発生後のトラフィックのバランシングには、自動または高速があります。

[短所]

[柔軟性に劣ります]。物理スイッチ構成で、ポート チャネル構成に物理スイッチ ポートを設定する必要があります。

[より複雑になります]。複数のスイッチを使用して、物理的に完全な冗長性構成を実現するのは容易ではありません。ベンダー固有の実装は、よりいっそう複雑なものになります。

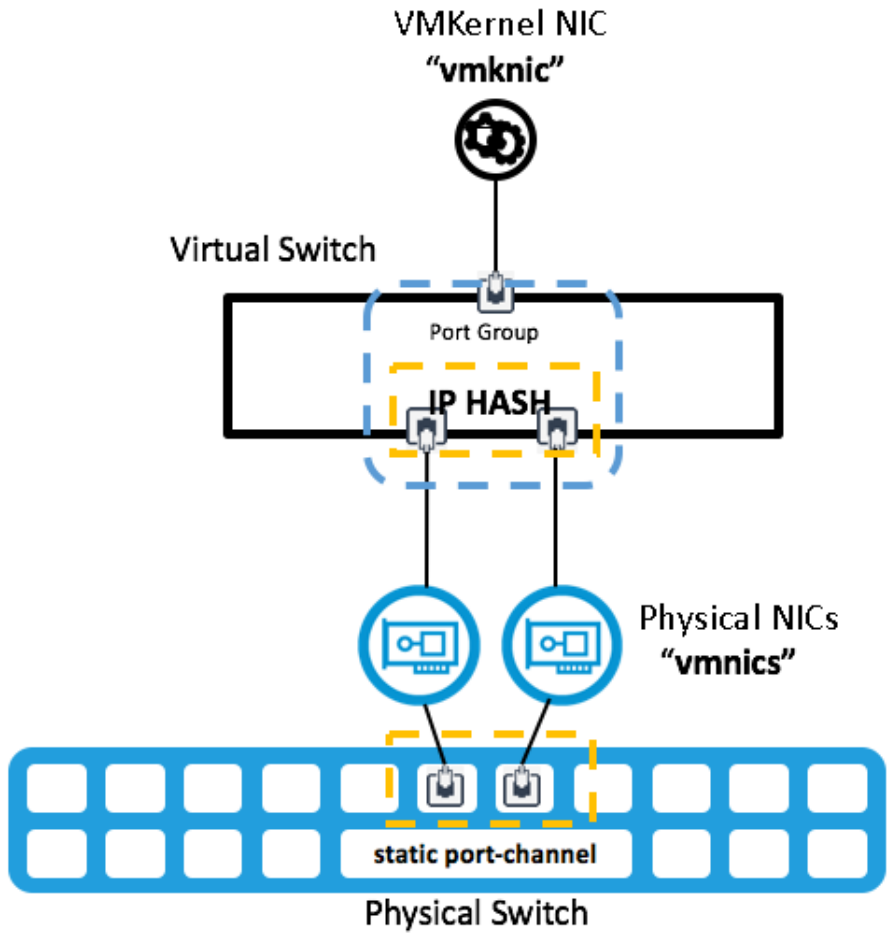
静的 LACP と IP ハッシュに基づいたルート

静的 LACP と IP ハッシュ ポリシーを使用して vSAN 6.6 クラスタを作成できます。このセクションでは、vSphere Standard Switch について説明しますが、vSphere Distributed Switch を使用することもできます。

「IP ハッシュに基づいたルート」ロード バランシング ポリシーを使用できます。

vSwitch またはポート グループ レベルで、[IP ハッシュに基づいたルート] ロード バランシング ポリシーを選択します。仮想スイッチまたはポート グループ レベルのチーミングおよびフェイルオーバー ポリシーで、静的チャネル グループに割り当てられたすべてのアップリンクをアクティブ アップリンクの位置に設定します。

IP ハッシュが vSphere ポート グループに設定されている場合、ポート グループは [IP ハッシュに基づいたルート] ポリシーを使用します。ポート チャネルのポート数は、チーム内のアップリンク数と同じにする必要があります。



静的 LACP と IP ハッシュを使用する場合の長所と短所

静的 LACP と IP ハッシュを使用する場合、次のトレードオフについて考慮してください。

[長所]

- [パフォーマンスとバンド幅が向上します]。1 つの vSAN ホストまたは VMkernel ポートは、IP ハッシュ アルゴリズムを使用して、他の多くの vSAN ホストと通信できます。
- [ネットワーク アダプタの冗長性を提供します]。NIC で障害が発生し、リンク状態が停止すると、チーム内の残りの NIC は引き続きトラフィックを通過させます。
- [柔軟性が向上します]。IP ハッシュは、vSphere Standard Switch と vSphere Distributed Switch の両方で使用できます。

[短所]

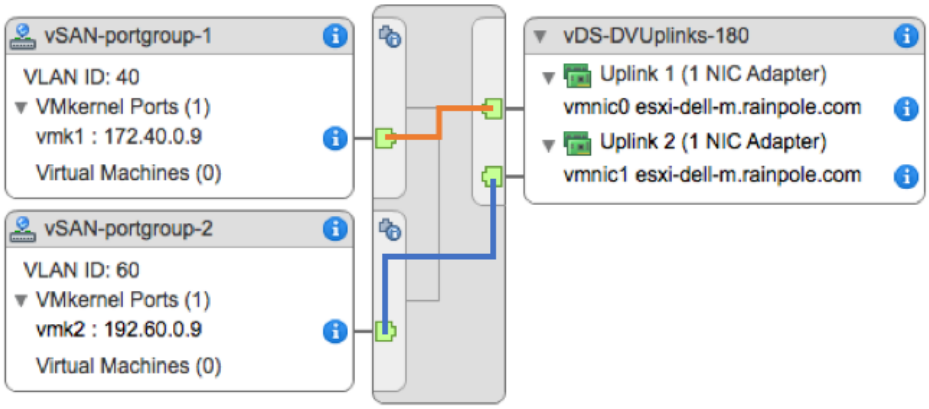
- [物理スイッチ構成の柔軟性は低くなります]。物理スイッチ ポートは、静的ポート チャネル構成で設定する必要があります。

- [誤構成が発生する機会が増えます]。LACP 動的ポート チャネルとは異なり、いずれかのエンドで検証を行わずに静的ポート チャネルが形成されます。
- [より複雑になります]。複数のスイッチを使用している場合、物理的に完全な冗長性構成を実現するのが難しくなります。実装はベンダー固有のものになる可能性があります。
- [ロード バランシングが制限されます]。使用環境の IP アドレス数が少ない場合、仮想スイッチはチーム内の 1 つのアップリンクで連続してトラフィックを通過させる可能性があります。これは特に、小規模な vSAN クラスタで発生する可能性があります。

ネットワークの air gap について

高度 NIC チーミング方法を使用すると、air gap ストレージ ファブリックを作成できます。2 つのストレージ ネットワークを使用して冗長ストレージ ネットワーク トポロジーを作成し、各ストレージ ネットワークを物理的かつ論理的に分離します。

vSphere 環境では、vSAN にネットワークの air gap を設定できます。vSAN ホストごとに複数の VMkernel ポートを構成します。1 つの vSwitch または複数の仮想スイッチ（vSphere Standard Switch や vSphere Distributed Switch など）を使用して、各 VMkernal ポートを専用の物理アップリンクに関連付けます。



通常、各アップリンクは、完全に冗長な物理インフラストラクチャに接続されている必要があります。

このトポロジは最適ではありません。同じネットワークに存在する異なるホスト上の NIC などのコンポーネントで障害が発生すると、ストレージ I/O が中断する可能性があります。この問題を回避するには、すべてのホストとすべてのネットワーク セグメントに物理 NIC の冗長性を実装します。このトポロジについては、構成例 2 で詳しく説明します。

これらの構成は L2 トポロジと L3 トポロジの両方に適用され、ユニキャスト構成とマルチキャスト構成の両方で使用できます。

vSAN による air gap ネットワーク構成の長所と短所

air gap ネットワークは、vSAN トラフィックの分離と隔離に役立ちます。このトポロジを構成する場合は注意してください。

[長所]

- vSAN トラフィックを物理的かつ論理的に分離できます。

[短所]

- vSAN では、同じサブネット上の複数の VMkernel アダプタ (vmknics) はサポートされません。詳細については、VMware KB2010877 を参照してください。
- 設定が複雑になり、エラーが発生しやすくなります。また、トラブルシューティングも複雑になります。
- 1 台のホストの 1 つの NIC で障害が発生し、別のホストの別の NIC に障害が発生するなど、複数の vmknic で非対称の障害が発生すると、ネットワークの可用性は保証されません。
- 物理 NIC 間での vSAN トラフィックのロードバランシングも保証されません。
- 複数の物理 NIC (vmnic) を保護するために複数の VMkernel アダプタ (vmknic) が必要になることもあり、vSAN ホストのコストが増加します。たとえば、2 つの vSAN vmknic に冗長性を提供するには、2 x 2 の vmnic が必要になる場合があります。
- VMkernel ポート、IP アドレス、VLAN など、必要な論理リソースも 2 倍になります。
- vSAN は、ポートのバインドを実装していません。これは、マルチパスなどの手法を使用できないことを意味します。
- レイヤー 3 トポロジは、複数の vmknic を使用する vSAN トラフィックには適していません。これらのトポロジは期待どおりに機能しない可能性があります。
- vSAN マルチキャスト アドレスを変更する場合、コマンドラインでのホストの構成が必要になる場合があります。

動的 LACP は、同時に複数のネットワーク接続を結合または集約してスループットを向上させ、冗長性を提供します。NIC チーミングが LACP で構成されている場合、複数のアップリンク間で vSAN ネットワークのロード バランシングが発生します。このロード バランシングはネットワーク レイヤーで行われ、vSAN を介して実行されることはありません。

注： リンク集約の説明では、ポート トランク、リンク バンドル、イーサネット/ネットワーク/NIC の結合、EtherChannel などの用語も使用されます。

このセクションでは、リンク集約制御プロトコル (LACP) を中心に説明します。IEEE 標準 は 802.3ad ですが、一部のベンダーは、PAgP（ポート集約プロトコル）などの独自の LACP 機能を提供しています。ベンダー推奨のベスト プラクティスに従ってください。

注： vSphere Distributed Switch 5.1 で導入された LACP サポートは、IP ハッシュのロード バランシングのみをサポートします。vSphere Distributed Switch 5.5 以降では、LACP が完全にサポートされています。

LACP はポートチャネルを使用する業界標準です。多くのハッシュ アルゴリズムを使用できます。vSwitch ポート グループ ポリシーとポート チャネル構成は一致している必要があります。

NIC チーミングの構成例

次の NIC チーミング構成は、標準的な vSAN ネットワークのシナリオを示しています。

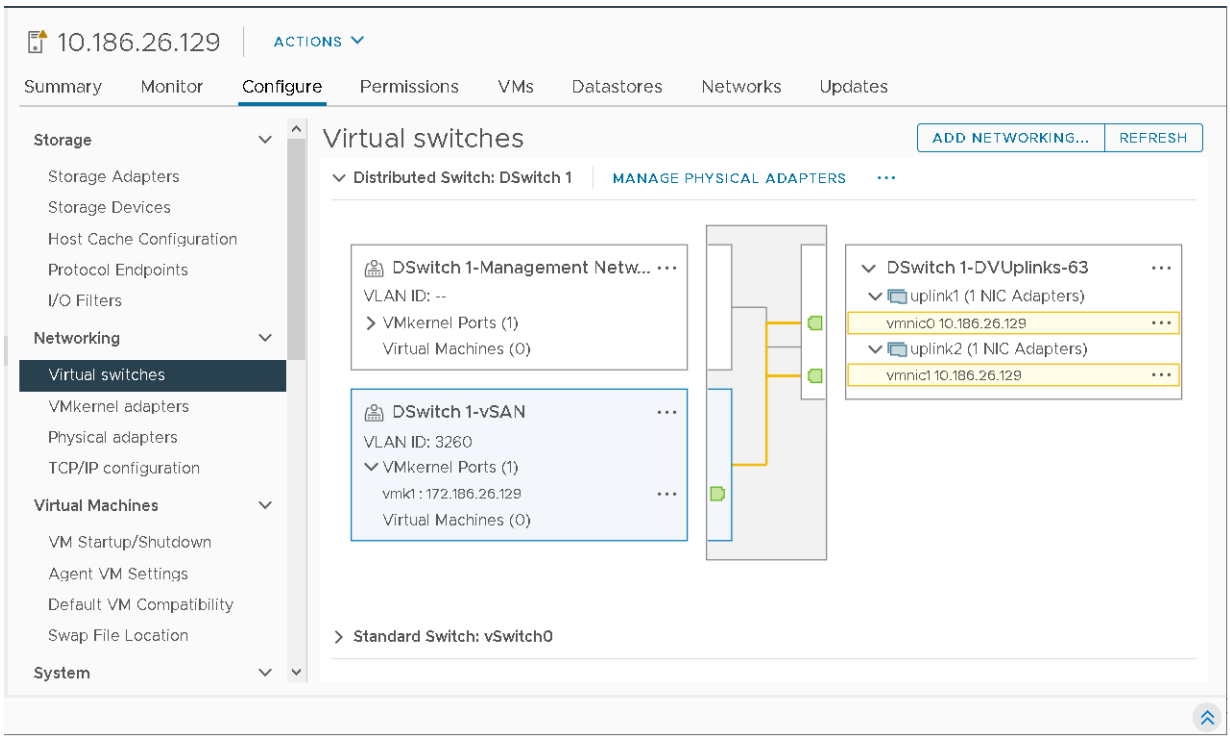
構成 1：単一の vmknic、物理 NIC の負荷に基づいたルート

vSAN ホストの [物理 NIC 負荷に基づいたルート] ポリシーを使用すると、アクティブ/アクティブの基本的な NIC チーミングを構成できます。vSphere Distributed Switch (vDS) を使用します。

この例の場合、vDS で各ホストに 2 つのアップリンクが構成される必要があります。分散ポート グループが vSAN トラフィック用に指定され、特定の VLAN に隔離されています。vDS でジャンボ フレームが有効になっており、MTU 値が 9,000 に設定されています。

次のように、vSAN トラフィックの分散ポート グループにチーミングとフェイルオーバーを設定します。

- ロード バランシング ポリシーに [物理 NIC 負荷に基づいたルート] を設定します。
- ネットワーク障害の検出を [リンク ステータスのみ] に設定します。
- スイッチへの通知を [はい] に設定します。
- フェイルバックを [いいえ] に設定します。フェイルバックを [はい] に設定できますが、この例では設定しません。
- 両方のアップリンクが [アクティブ アップリンク] の位置にあることを確認します。



ネットワーク アップリンクの冗長性の消失

リンクダウン状態が検出されると、ワークロードのアップリンクが切り替わります。vSAN クラスタと仮想マシンのワークロードに対して大きな影響はありません。

リカバリとフェイルバック

[フェイルバック] を [いいえ] に設定すると、トラフィックは元の vmnic に戻りません。[フェイルバック] を [はい] に設定すると、リカバリ時にトラフィックが元の vmnic に戻ります。

ロード バランシング

VMkernel NIC が 1 つしかないため、[物理 NIC 負荷に基づいたルート] を使用してもパフォーマンスは向上しません。

使用される物理 NIC は常に 1 つだけです。他の物理 NIC はアイドル状態になります。

構成 2：複数の vmknic、発信元のポート ID に基づいたルート

論理的かつ物理的に分離されたルーティング不可の 2 つの VLAN を使用すると、air-gap トポロジを作成できます。

この例では、vSphere Distributed Switch の構成手順について説明しますが、vSphere Standard Switch を使用することもできます。その場合は、10 Gb の物理 NIC を 2 つ使用し、それらを vSphere ネットワーク レイヤーで論理的に分離します。

vSAN VMkernel vmknic ごとに 2 つの分散ポート グループを作成します。ポート グループごとに個別の VLAN タグがあります。vSAN VMkernel 構成の場合、両方の VLAN で vSAN トラフィックに 2 つの IP アドレスが必要になります。

注： 通常の実装では、完全な冗長性を実現するために 4 つの物理アップリンクを使用します。

各ポート グループに、チーミングおよびフェイルオーバー ポリシーはデフォルトの設定を使用します。

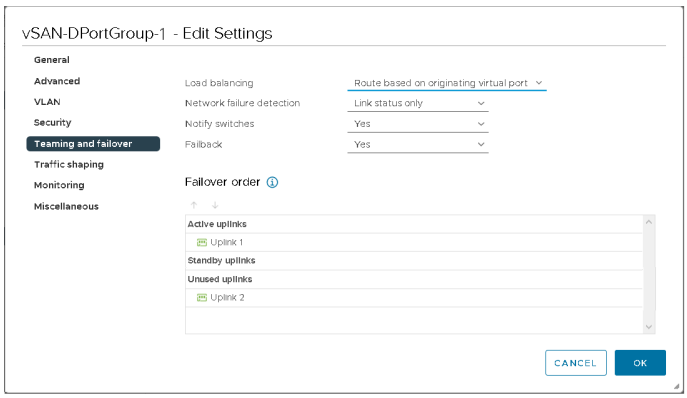
- ロード バランシングを [発信元のポート ID に基づいたルート] に設定
- ネットワーク障害の検出を [リンク ステータスのみ] に設定
- スイッチへの通知をデフォルト値の [はい] に設定
- フェイルバックをデフォルト値の [はい] に設定
- アップリンク構成で、1 つのアップリンクが [アクティブ] の位置にあり、1 つのアップリンクが [未使用] の位置にあります。

1 つのネットワークは、他のネットワークから完全に隔離されています。

vSAN ポート グループ 1

この例では、[vSAN-DPortGroup-1] という分散ポート グループを使用しています。このポート グループには [VLAN 3266] というタグが付き、次のチーミングおよびフェイルオーバー ポリシーが設定されています。

- VLAN 3266 タグが付いたポート グループのトラフィック
- ロード バランシングを [発信元のポート ID に基づいたルート] に設定
- ネットワーク障害の検出を [リンク ステータスのみ] に設定
- スイッチへの通知をデフォルト値の [はい] に設定
- フェイルバックをデフォルト値の [はい] に設定
- アップリンク構成で、[アップリンク 1] が [アクティブ] の位置にあり、[アップリンク 2] が [未使用] の位置にあります。



vSAN ポート グループ 2

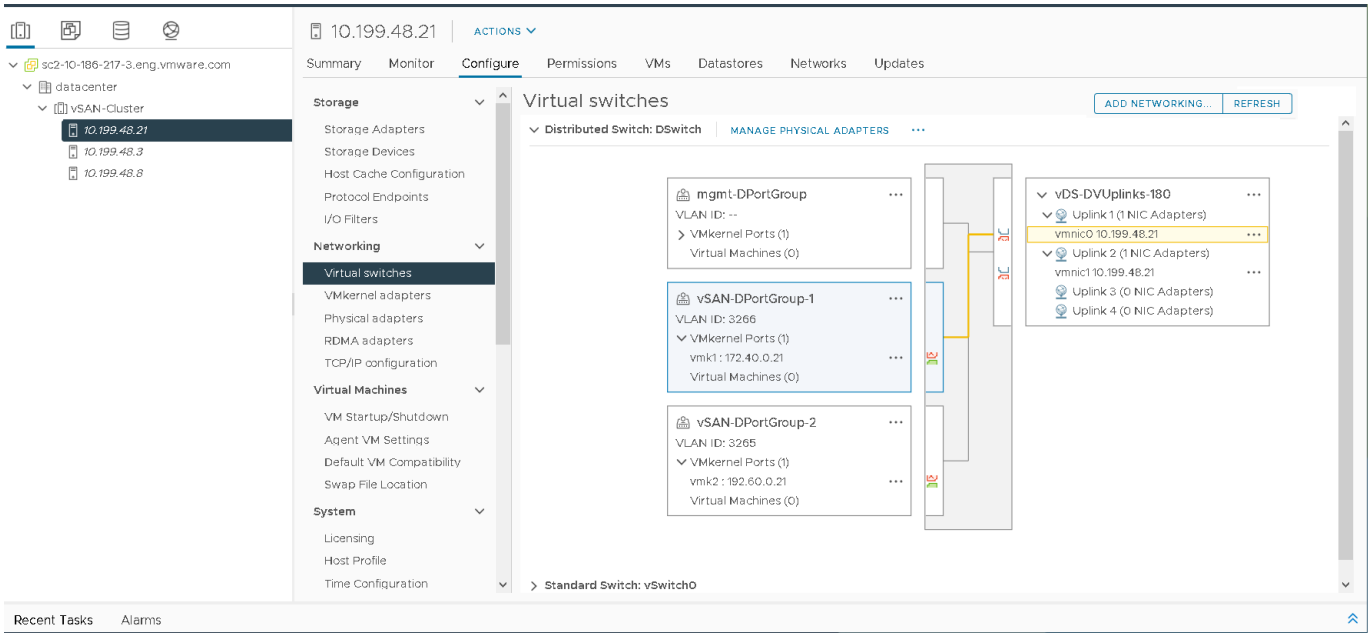
vSAN ポート グループ 1 を補完するため、[vSAN-portgroup-2] という 2 つ目の分散ポート グループを構成します。次のような違いがあります。

- VLAN 3265 タグが付いたポート グループのトラフィック
- アップリンク構成で、[アップリンク 2] が [アクティブ] の位置にあり、[アップリンク 1] が [未使用] の位置にあります。

vSAN VMkernel ポート構成

両方のポート グループに 2 つの vSAN VMkernel インターフェイスを作成します。この例では、ポート グループに [vmk1] と [vmk2] という名前を使用しています。

- [vmk1] は VLAN 3266 (172.40.0.xx) に関連付けられ、結果としてポート グループ [vSAN-DPortGroup-1] に関連付けられます。
- [vmk2] は VLAN 3265 (192.60.0.xx) に関連付けられ、結果としてポート グループ [vSAN-DPortGroup-2] に関連付けられます。



ロード バランシング

vSAN には複数の vmknic を区別するロード バランシング メカニズムがありません。そのため、選択された vSAN I/O パスを物理 NIC 間で特定できません。vSphere のパフォーマンス チャートを見ると、1 つの物理 NIC が他の物理 NIC よりも使用率が高いことが分かります。ラボでシンプルな I/O テストを実行しました。120 台の仮想マシンを使用し、4 ホスト構成のオール フラッシュ vSAN クラスタで 64K ブロックの読み取り/書き込みを 70:30 の割合で行った結果、NIC 間で負荷が分散されていないことが明らかになりました。

vSphere パフォーマンス グラフには、NIC 間で負荷が分散されていないことが表示されます。

ネットワーク アップリンクの冗長性の消失

この構成で発生するネットワーク障害について考えてみましょう。特定の vSAN ホストで vmnic1 は有効になっていません。その結果、ポート [vmk2] が影響を受けます。NIC に障害が発生すると、ネットワーク接続アラームと冗長性アラームの両方がトリガされます。

vSAN の場合、CMMDS（クラスタ監視およびメンバーシップ ディレクトリ サービス）が障害を検出してから約 [10] 秒後に、このフェイルオーバー プロセスがトリガされます。フェイルオーバーとリカバリの実行中、vSAN は、障害が発生したネットワークでアクティブな接続をすべて停止し、機能している残りのネットワーク上で接続の再確立を試みます。

2 つの個別の vSAN VMkernel ポートは隔離された VLAN で通信を行うため、vSAN 健全性チェックでエラーが発生することがあります。VLAN 3265 で [vmk2] がピアと通信できなくなるため、これは想定内の状況です。

パフォーマンス チャートには、vmnic1 に障害が発生したため、影響を受けたワークロードが vmnic0 で再起動したことが表示されます。このテストでは、vSphere NIC チーミングとトポロジの重要な違いを表しています。vSAN は、残りのネットワークで接続を再確立または再開しようとします。

ただし、障害の状況によっては、ESXi TCP 接続タイムアウトが原因で、影響を受けた接続のリカバリが完了するまでに [90 秒] ほどかかる場合があります。以降の接続も失敗する可能性がありますが、接続の試行は 5 秒でタイムアウトし、使用可能なすべての IP アドレスを順番に試していきます。この動作は、仮想マシンのゲスト I/O に影響を与えることがあるため、アプリケーションと仮想マシンの I/O の再試行が必要になる可能性があります。

たとえば、Windows Server 2012 仮想マシンで、フェイルオーバーとリカバリの処理中に、イベント ID 153（デバイス リセット）や 129（再試行イベント）が記録される場合があります。この例では、I/O がリカバリされるまで、イベント ID 129 は約 90 秒間でログに記録されていました。

深刻な影響を受ける前に、一部のゲスト OS のディスク タイムアウト設定を変更する必要がある場合があります。ディスク タイムアウト値は、VMware Tools の有無、特定のゲスト OS のタイプ、バージョンによって異なる場合があります。ゲスト OS のディスク タイムアウト値の変更については、VMware KB1009465 を参照してください。

リカバリとフェイルバック

別の障害が原因で新たな障害が発生した場合を除き、ネットワークの修復後にワークロードは自動的に再分散されません。影響を受けたネットワークがリカバリされるとすぐに、新しい TCP 接続で使用できるようになります。

構成 3：動的 LACP

スイッチに 2 ポートの LACP ポート チャネルを構成し、vSphere Distributed Switch に 2 つのアップリンクのリンク集約グループを構成できます。

この例では、サーバごとに 2 つの物理アップリンクを設定し、10Gb のネットワークを使用します。

注： vSAN over RDMA を使用する場合、この構成はサポートされません。

ネットワーク スイッチの構成

次の設定を使用して、vSphere Distributed Switch を構成します。

- vSAN ホストが接続するポートを特定します。
- ポート チャネルを作成します。
- VLAN を使用する場合は、正しい VLAN をポート チャネルにトランクします。
- 必要なディストリビューションまたはロード バランシング オプション（ハッシュ）を構成します。
- LACP モードをアクティブ/動的に設定します。
- MTU の設定を確認します。

vSphere の構成

次の設定を使用して、vSphere ネットワークを構成します。

- 適切な MTU を使用して vDS を構成します。
- ホストを vDS に追加します。
- 正しい数のアップリンクとポート チャネルに一致する属性を指定して LAG を作成します。
- 物理アップリンクを LAG に割り当てます。
- vSAN トラフィックの分散ポート グループを作成し、正しい VLAN を割り当てます。
- 適切な MTU を使用して、vSAN に VMkernel ポートを設定します。

物理スイッチの設定

次の設定を使用して、物理スイッチを構成します。この構成を Dell サーバで設定する方法のガイダンスについては、http://www.dell.com/Support/Article/jp/ja/19/HOW10364 を参照してください。

2 つのアップリンク LAG を設定します。

- スイッチ ポート 36 と 18 を使用します。
- この構成では VLAN トランクを使用します。ポート チャネルは VLAN トランク モードで、適切な VLAN がトランクされます。
- ロード バランシングまたは負荷分散に次の方法を使用します。[送信元と宛先の IP アドレス、TCP/UDP ポートと VLAN]
- LACP モードが [アクティブ]（動的）になっていることを確認します。

次のコマンドを使用して、Dell スイッチの個々のポート チャネルを構成します。

- ポート チャネルを作成します。

```
#interface port-channel 1
```

- ポート チャネルを VLAN トランク モードに設定します。

```
#switchport mode trunk
```

- VLAN アクセスを許可します。

```
#switchport trunk allowed vlan 3262
```

- ロード バランシング オプションを構成します。

```
#hashing-mode 6
```

- 正しいポートをポート チャネルに割り当て、モードをアクティブに設定します。
- ポート チャネルが正しく構成されていることを確認します。

```
   #show interfaces port-channel 1


   Channel Ports Ch-Type Hash Type Min-links Local Prf
   ------- ------------------------------ -------- --------- --------- ---------
   Po1 Active: Te1/0/36, Te1/0/18 Dynamic 6 1 Disabled
   Hash Algorithm Type
   1 - Source MAC, VLAN, EtherType, source module and port Id
   2 - Destination MAC, VLAN, EtherType, source module and port Id
   3 - Source IP and source TCP/UDP port
   4 - Destination IP and destination TCP/UDP port
   5 - Source/Destination MAC, VLAN, EtherType, source MODID/port
   6 - Source/Destination IP and source/destination TCP/UDP port
   7 - Enhanced hashing mode
#interface range Te1/0/36, Te1/0/18
#channel-group 1 mode active
```

完全な構成：

```
#interface port-channel 1
#switchport mode trunk
#switchport trunk allowed vlan 3262
#hashing-mode 6
#exit
#interface range Te1/0/36,Te1/018
```

```
#channel-group 1 mode active

#show interfaces port-channel 1
```

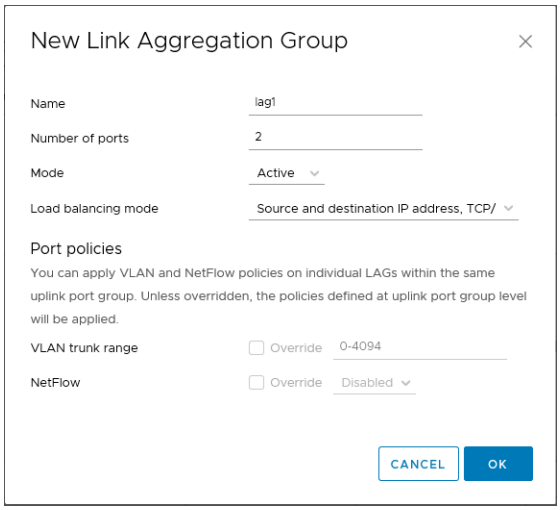
注： vSAN ホストに接続している参加中のすべてのスイッチ ポートで、この手順を繰り返します。

vSphere Distributed Switch の設定

開始する前に、vDS が LACP 対応のバージョンにアップグレードされていることを確認してください。確認するには、vDS を右クリックし、アップグレード オプションが利用可能かどうか確認します。必要であれば、LACP 対応のバージョンに vDS をアップグレードします。

vDS での LAG の作成

Distributed Switch に LAG を作成するには、その vDS を選択して [設定] タブをクリックし、[LACP] を選択します。新しい LAG を追加します。



次のプロパティを使用して LAG を設定します。

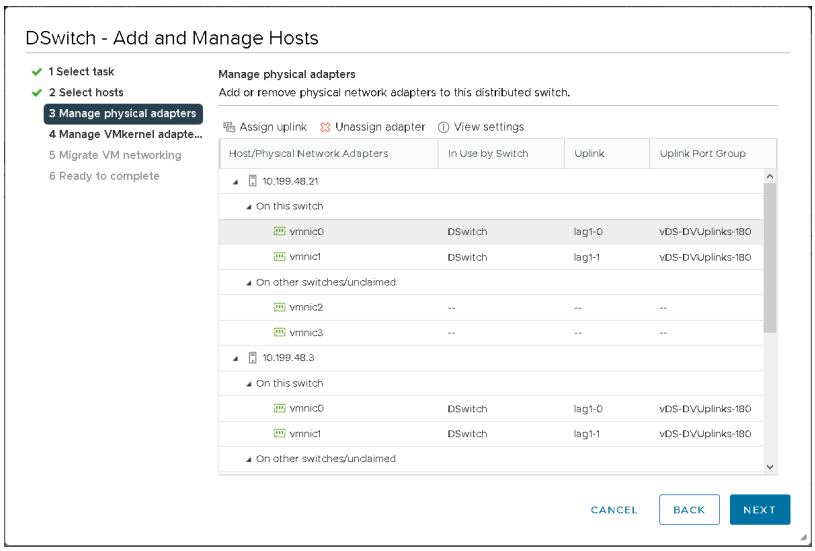
- LAG 名：[lag1]
- ポート数：[2]（スイッチのポート チャネルと一致させます）
- モード：[アクティブ]（物理スイッチと一致させます）。
- ロード バランシング モード：[送信元と宛先の IP アドレス、TCP/UDP ポートと VLAN]

LAG への物理アップリンクの追加

vSAN ホストが vDS に追加されました。個々の vmnic を適切な LAG ポートに割り当てます。

- vDS を右クリックし、[ホストの追加と管理...] を選択します。
- [ホスト ネットワークの管理] を選択して、接続されているホストを追加します。
- [物理アダプタの管理] で、必要なアダプタを選択し、LAG ポートに割り当てます。
- LAG1 のアップリンク 1 の位置からポート 0 に vmnic0 を移行します。

この手順を vmnic1 に繰り返し、2 つ目の LAG ポート位置 lag1-1 に割り当てます。



分散ポート グループのチーミングおよびフェイルオーバー ポリシーの構成

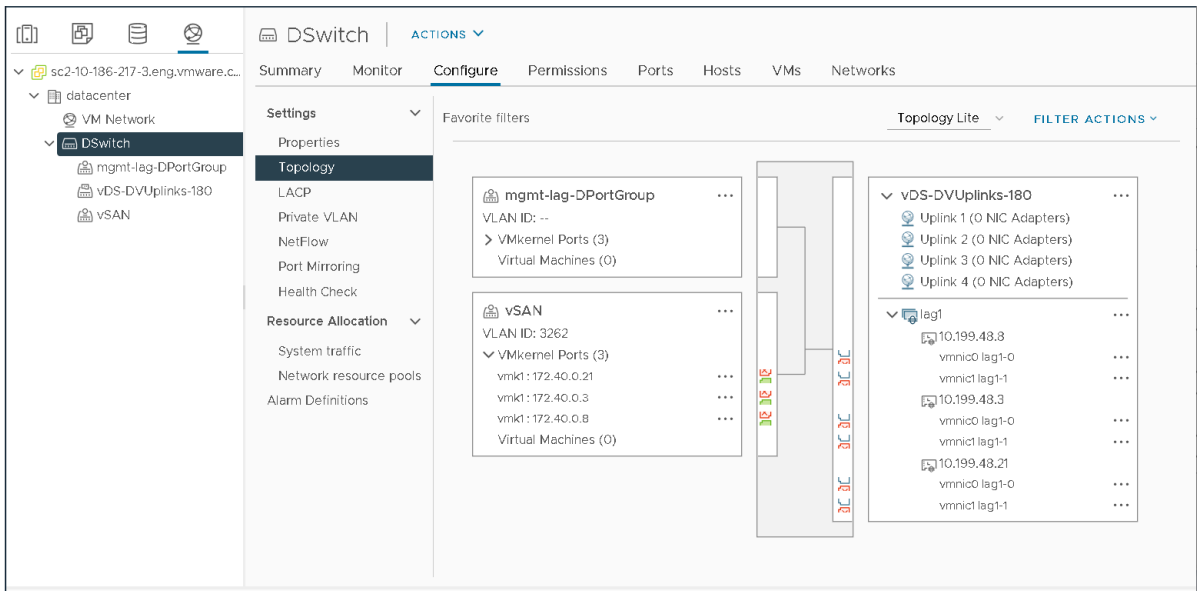
分散ポート グループのチーミングおよびフェイルオーバー ポリシーで、[アクティブ アップリンク]として LAG グループを割り当てます。vSAN トラフィック用に指定された分散ポート グループを選択または作成します。この構成で使用する vSAN ポート グループは [vSAN] という名前で、VLAN ID 3262 というタグが付いています。ポート グループを編集して、新しい LAG 構成を反映するようにチーミングおよびフェイルオーバー ポリシーを設定します。

LAG グループ [lag1] がアクティブ アップリンクの位置にあり、残りのアップリンクが [未使用] の位置にあることを確認します。

注： リンク集約グループ (LAG) が唯一のアクティブ アップリンクとして選択されている場合、LAG のロードバランシング モードにより、ポート グループのロードバランシング モードがオーバーライドされます。したがって、[発信元の仮想ポートに基づいたルート] ポリシーは機能しません。

VMkernel インターフェイスの作成

最後に、新しい分散ポート グループを使用する VMkernel インターフェイスを作成します。これらのインターフェイスには、vSAN トラフィックのタグを付けます。各 vSAN vmknic が LAG グループの vmnic0 と vmnic1 を介して通信を行い、ロード バランシングとフェイルオーバーが実行されていることを確認します。



ロード バランシングの設定

ロード バランシングの点から見ると、この LAG 設定のすべての vmnic のホストでトラフィックが一貫した方法で分散されているわけではありませんが、構成 1 の[物理 NIC 負荷に基づいたルート]や、構成 2 で使用した air-gap/マルチ vmknic に比べると一貫性は増しています。

個々のホストの vSphere パフォーマンス グラフには、ロード バランシングの向上が示されます。

ネットワーク アップリンクの冗長性の消失

特定の vSAN ホストで vmnic1 が有効になっていない場合、ネットワーク冗長性アラームがトリガされます。

vSAN 健全性アラームはトリガされません。air-gap、マルチ vmknic 構成に比べると、ゲスト I/O への影響は最小限になります。この構成では、LACP が構成された TCP セッションを停止する必要はありません。

リカバリとフェイルバック

フェイルバックシナリオでは、負荷ベースのチーミング、マルチ vmknic、および vSAN 環境の LACP の動作が異なります。vmnic1 がリカバリされると、両方のアクティブ アップリンク間でトラフィックが自動的に分散されます。この動作は、vSAN トラフィックに対有効な場合があります。

フェイルバックの設定

LAG ロード バランシング ポリシーは、vSphere 分散ポート グループのチーミング ポリシーとフェイルオーバー ポリシーをオーバーライドします。また、フェイルバック値のガイダンスについても考慮する必要があります。ラボテストでは、LACP でフェイルバックが [はい] に設定されても、[いいえ] に設定されていも、動作に大きな違いはありません。LAG の設定は、ポート グループの設定よりも優先されます。

注： LACP ではビーコンの検知がサポートされていないため、ネットワーク障害の検出値は [リンク ステータス] のままです。VMware KB Understanding IP Hash load balancing (KB2006129)を参照してください。

構成 4：静的 LACP – IP ハッシュに基づいたルート

スイッチの 2 ポート構成の LACP 静的ポートチャネルと、vSphere Standard Switch の 2 つのアクティブ アップリンクを使用できます。

この構成では、サーバごとに 2 つの物理アップリンクを設定し、10Gb のネットワークを使用します。各ホストには vSAN 用に 1 つの VMkernel インターフェイス (vmknic) があります。

ホスト要件の詳細と構成例については、次の VMware ナレッジベースの記事を参照してください。

- Host requirements for link aggregation for ESXi and ESX (KB1001938)
- Sample configuration of EtherChannel / Link Aggregation Control Protocol (LACP) with ESXi/ESX and Cisco/HP switches (KB 1004048)

注： vSAN over RDMA を使用する場合、この構成はサポートされません。

物理スイッチの構成

次のように、2 つのアップリンク静的ポートチャネルを設定します。

- スイッチ ポート 43 および 44
- VLAN トランク。ポートチャネルは VLAN トランク モードで、適切な VLAN がトランクされてます。
- ポート チャネル グループにロード バランシング ポリシーを指定しないでください。

次の手順に従ってスイッチの個々のポート チャネルを設定します。

手順 1：ポート チャネルを作成します。

```
#interface port-channel 13
```

手順 2：ポート チャネルを VLAN トランク モードに設定します。

```
#switchport mode trunk
```

手順 3：適切な VLAN を許可します。

```
#switchport trunk allowed vlan 3266
```

手順 4：正しいポートをポート チャネルに割り当て、モードをアクティブに設定します。

```
#interface range Te1/0/43, Te1/0/44

#channel-group 1 mode on
```

手順 5：ポート チャネルが静的ポート チャネルとして構成されていることを確認します。

```
#show interfaces port-channel 13


Channel Ports Ch-Type Hash Type Min-links Local Prf

------- ------------------------------ -------- --------- --------- --

Po13 Active: Te1/0/43, Te1/0/44 Static 7 1 Disabled
```

```
Hash Algorithm Type

1 - Source MAC, VLAN, EtherType, source module and port Id

2 - Destination MAC, VLAN, EtherType, source module and port Id

3 - Source IP and source TCP/UDP port

4 - Destination IP and destination TCP/UDP port

5 - Source/Destination MAC, VLAN, EtherType, source MODID/port

6 - Source/Destination IP and source/destination TCP/UDP port

7 - Enhanced hashing mode
```

vSphere Standard Switch の構成

この例は、vSphere Standard Switch の構成と作成について理解していることを前提としています。

この例では、次の構成を使用します。

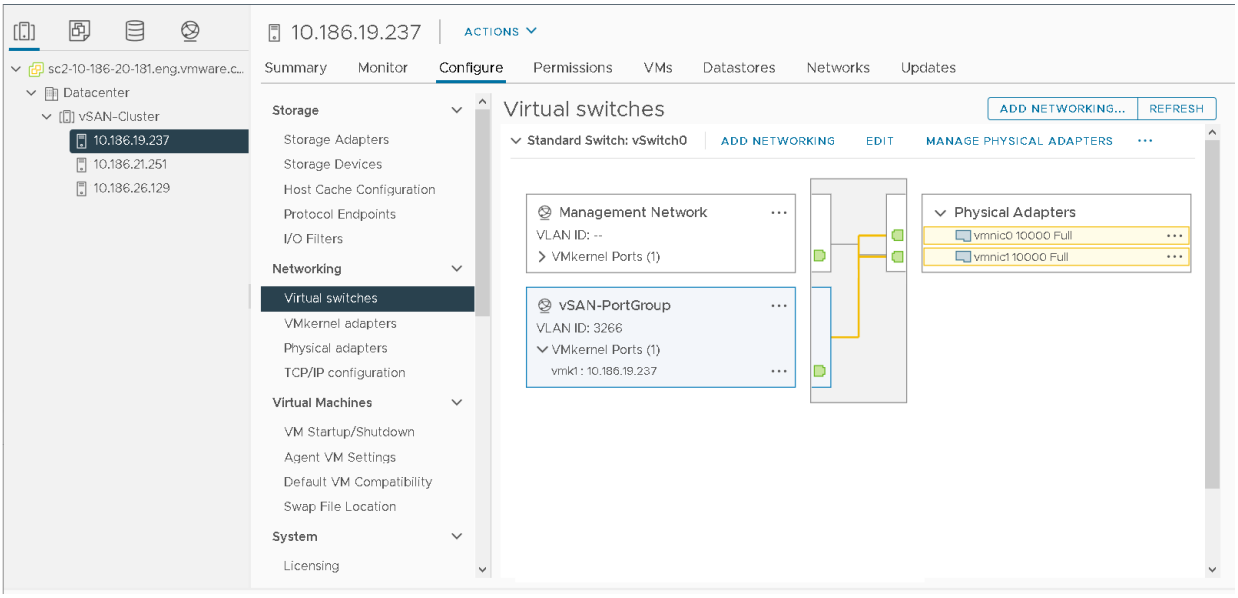
- 同一 vSAN ホスト
- vmnic0 および vmnic1 という名前のアップリンク
- スイッチ ポートおよびポート チャネルにトランクされる VLAN 3266
- ジャンボ フレーム

各ホストで、MTU が 9,000 に設定された [vSwitch1] を作成し、vmnic0 と vmnic1 を vSwitch に追加します。チーミングおよびフェイルオーバー ポリシーで、両方のアダプタを [アクティブ] の位置に設定します。ロード バランシング ポリシーを [IP ハッシュに基づいたルート] に設定します。

次のように、vSAN トラフィックの分散ポート グループにチーミングとフェイルオーバーを設定します。

- ロード バランシング ポリシーに [IP ハッシュに基づいたルート] を設定します。
- ネットワーク障害の検出を [リンク ステータスのみ] に設定します。
- スイッチへの通知を [はい] に設定します。
- フェイルバックを [はい] に設定します。
- 両方のアップリンクが [アクティブ アップリンク] の位置にあることを確認します。

ネットワーク検出、スイッチへの通知、フェイルバックにデフォルトを使用します。すべてのポート グループは、vSwitch レベルで設定されたチーミングおよびフェイルオーバー ポリシーを継承します。個々のポート グループのチーミングおよびフェイルオーバー ポリシーをオーバーライドして、親 vSwitch と異なる値を設定できますが、すべてのポート グループの IP ハッシュ ロード バランシングには同じアップリンクのセットを使用してください。



ロード バランシングの設定

両方の物理アップリンクが使用されていますが、すべての物理 vmnic でトラフィックが均等に分散されていません。この図では、アクティブ トラフィックのみが vSAN トラフィックになっていますが、実際は 4 つの vmknic または IP アドレスです。この動作は、IP アドレスと使用可能なハッシュの数が少ない場合に発生する可能性があります。ただし、状況によっては、仮想スイッチがチーム内の 1 つのアップリンクを介してトラフィックを通過させることがあります。IP ハッシュ アルゴリズムの詳細については、IP ハッシュに基づいたルートに関する『vSphere のドキュメント』を参照してください。

ネットワークの冗長性

この例では、障害や冗長性の動作に重点を置くため、vmnic1 はスイッチから切断されたポートに接続しています。ネットワーク アップリンクの冗長性アラームはトリガされています。

vSAN 健全性アラームはトリガされていません。クラスタおよび仮想マシンのコンポーネントに影響はありません。この障害でゲスト ストレージ I/O は中断しません。

リカバリとフェイルバック

vmnic1 がリカバリされると、両方のアクティブ アップリンク間でトラフィックが自動的に分散されます。

Network I/O Control

10

vSphere Network I/O Control を使用して、ネットワーク トラフィックにサービス品質 (QoS) レベルを設定します。

vSphere Network I/O Control は、vSphere Distributed Switch で使用可能な機能です。これにより、ネットワーク トラフィックにサービス品質 (QoS) を実装します。これは、vSAN で vSAN トラフィックと vMotion、管理、仮想マシンなどの他のトラフィック タイプが物理 NIC を共有する場合に役立ちます。

予約、シェアおよび制限

vSAN の物理アダプタで使用できる最低のバンド幅が Network I/O Control で確保されるように [予約] を設定できます。

vMotion や完全なホスト退避などの 突発的なトラフィックが vSAN トラフィックに影響する可能性がある場合、予約は有効な手段となります。予約は、ネットワークバンド幅に競合がある場合にのみ呼び出されます。Network I/O Control の予約の短所としては、未使用の予約バンド幅を仮想マシンのトラフィックに割り当てることができない点があります。すべてのシステム トラフィック タイプで予約される合計バンド幅は、最低キャパシティを備えた物理ネットワーク アダプタが提供できるバンド幅の 75 パーセントを超過することはできません。

[予約に関する vSAN のベスト プラクティス]。vSAN 用に予約されたトラフィックは仮想マシン トラフィックに割り当てることができません。vSAN 環境では、NIOC の予約を使用しないでください。

[シェア] を設定すると、vSAN に割り当てられた物理アダプタが飽和状態になったときに、特定のバンド幅を vSAN で使用できるようになります。これにより、再構築および同期の操作中に、vSAN が物理アダプタのキャパシティ全体を消費するのを防ぎます。たとえば、チームの別の物理アダプタに障害が発生し、ポート グループのすべてのトラフィックがチーム内の残りのアダプタに転送されると、物理アダプタが飽和状態になる可能性があります。[シェア] オプションをすると、他のトラフィックが vSAN ネットワークに影響を与えることがなくなります。

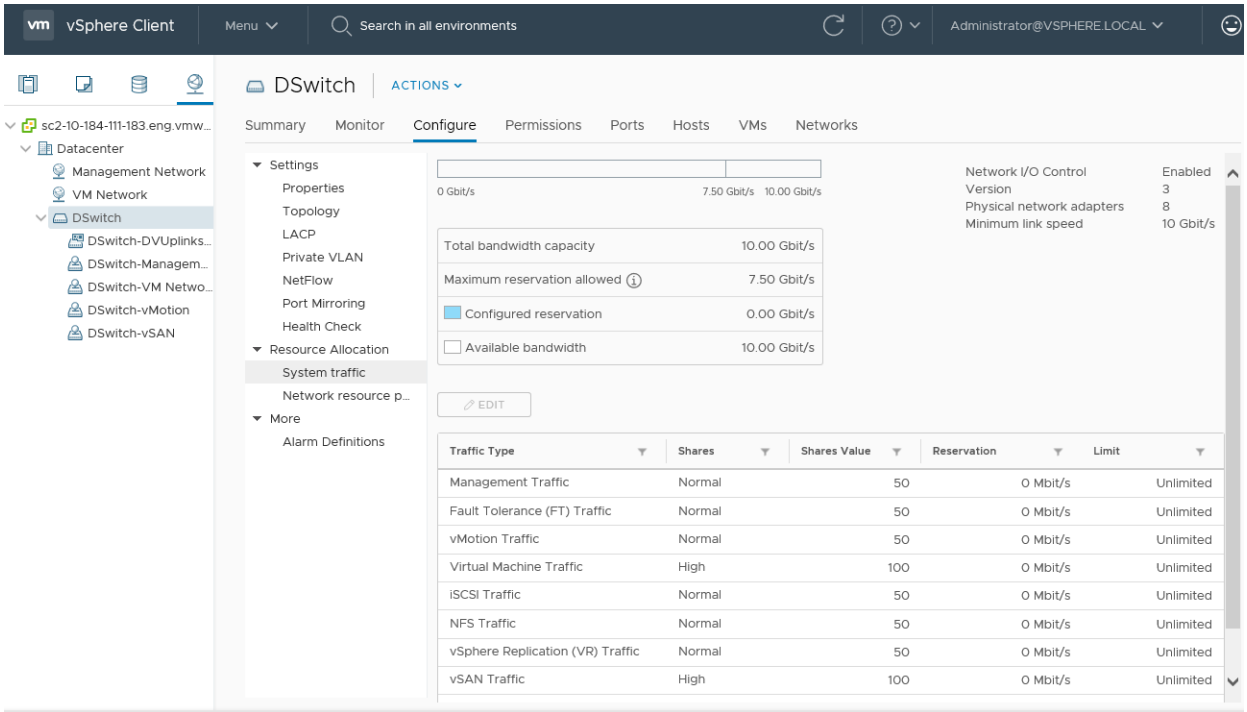
[シェアに関する vSAN の推奨事項]。これは NIOC で最も公平なバンド幅割り当て方法であり、vSAN 環境で使用する場合に推奨されます。

[制限] を設定すると、特定のトラフィック タイプがアダプタで使用できる最大バンド幅を定義できます。追加のバンド幅を使用しているユーザーがいない場合は、制限付きのトラフィック タイプも使用できません。

[制限に関する vSAN の推奨事項]。制限付きのトラフィック タイプでは追加のバンド幅を使用できないため、vSAN 環境では NIOC の制限を使用しないでください。

ネットワーク リソース プール

Network I/O Control で制御できるすべてのシステム トラフィック タイプを表示できます。複数の仮想マシン ネットワークがある場合は、仮想マシンのトラフィックに一定のバンド幅を割り当てることができます。ネットワーク リソース プールを使用して、仮想マシンのポート グループに基づいてバンド幅の一部を消費します。



Network I/O Control の有効化

vDS の構成プロパティで Network I/O Control を有効にできます。vSphere Client で vDS を右クリックし、[設定] > [設定の編集] を選択します。

注： Network I/O Control は、vSphere Distributed Switch でのみ使用できます。vSphere Standard Switch では使用できません。

Network I/O Control を使用すると、ホストの物理アダプタのキャパシティに基づいて、ネットワーク トラフィックのバンド幅を予約できます。たとえば、vSAN トラフィックが 10 GbE の物理ネットワーク アダプタを使用し、これらのアダプタが他のシステム トラフィック タイプと共有されている場合、vSphere Network I/O Control を使用して、vSAN に一定量のバンド幅を確保できます。これは、vSphere vMotion や vSphere HA などのトラフィックと仮想マシンのトラフィックが vSAN ネットワークと同じ物理 NIC を共有する場合に便利です。

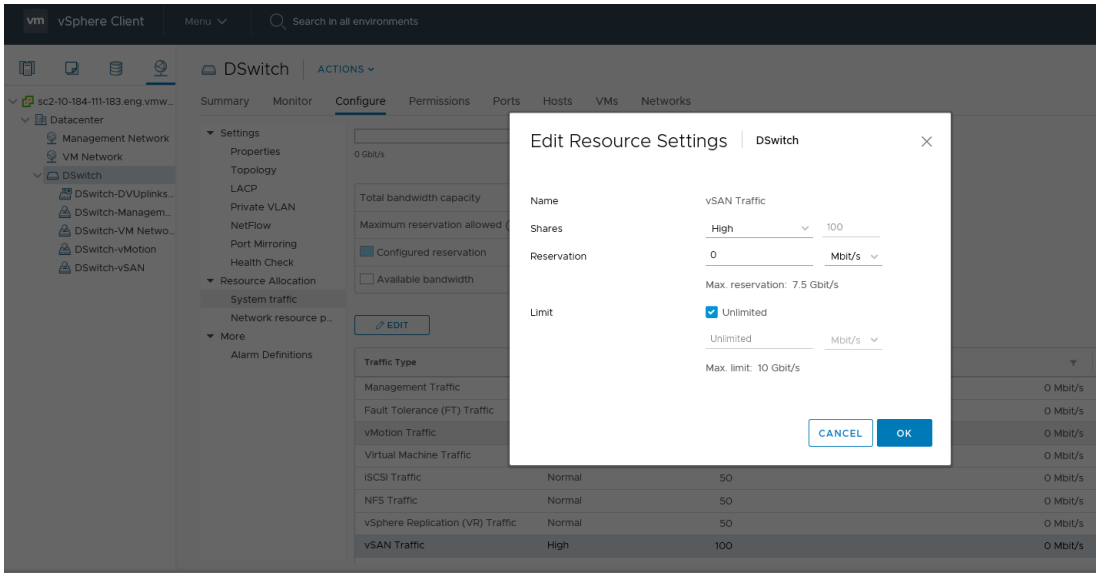
次のトピックを参照してください。

- Network I/O Control の構成例

Network I/O Control の構成例

vSAN クラスタに Network I/O Control を設定できます。

1 つの 10 GbE の物理アダプタを含む vSAN クラスタについて考えてみましょう。この NIC は、vSAN、vSphere vMotion、仮想マシンのトラフィックを処理します。トラフィック タイプのシェア値を変更するには、[システム トラフィック] ビューからトラフィック タイプを選択して([VDS] > [構成] > [リソース割り当て] > [システム トラフィック] の順に移動)、[編集] をクリックします。vSAN トラフィックのシェア値がデフォルトの「正常/50」から「高/100」に変更されました。



テーブルに表示されたシェア値と一致するように、他のトラフィック タイプを編集します。

表 10-1. NIOC の設定例

[トラフィック タイプ]	[シェア]	[値]
[vSAN]	高	100
[vSphere vMotion]	低	25
[仮想マシン]	正常	50
[iSCSI/NFS]	低	25

10 GbE アダプタが飽和状態になると、Network I/O Control は、物理アダプタの vSAN に 5 Gbps、仮想マシン トラフィックに 3.5 Gbps、vMotion に 1.5 Gbps を割り当てます。これらの値を開始点として使用して、vSAN ネットワークの NIOC 構成を設定します。すべてのプロトコルで vSAN が最優先になっていることを確認します。

バンド幅の割り当てパラメータの詳細については、『vSphere ネットワーク』を参照してください。

vSAN の各 vSphere エディションで、エディションの一部として vSphere Distributed Switch が提供されます。Network I/O Control は、どの vSAN エディションでも構成できます。

vSAN ネットワーク トポロジについて

11

vSAN アーキテクチャは、異なるネットワーク トポロジをサポートします。これらのトポロジは、vSAN の展開と管理全体に影響します。

vSAN 6.6 でユニキャスト サポートが導入されました。これにより、ネットワークの設計を簡単に行うことができます。

次のトピックを参照してください。

- 標準の展開
- vSAN ストレッチ クラスタの展開
- 2 ノード構成の vSAN の展開
- データ サイトから監視ホストへのネットワークの構成
- 例外的な展開

標準の展開

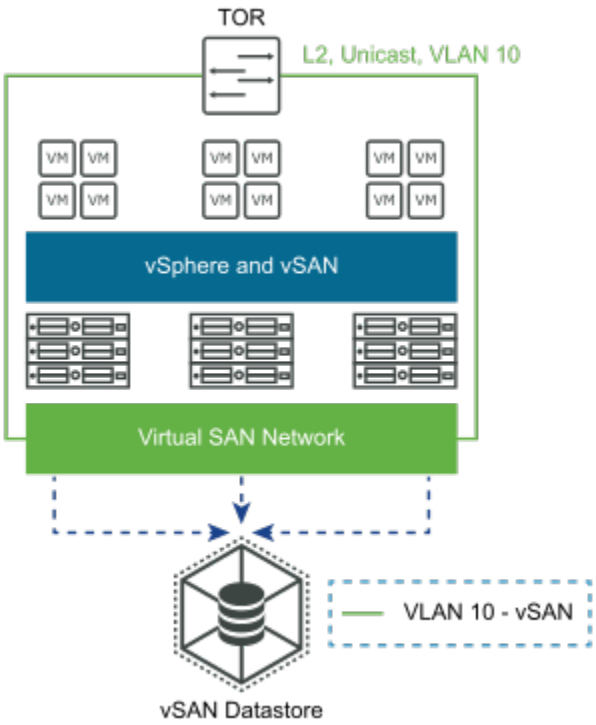
vSAN は、複数の単一サイト デプロイ タイプをサポートします。

レイヤー 2、単一サイト、単一ラック

このネットワーク トポロジは、ホスト、ブリッジ、スイッチなどのレイヤー 2 の中間デバイスを介してパケット転送を行う役割を担います。

レイヤー 2 ネットワーク トポロジは、vSAN の最もシンプルな実装と管理を実現します。ネットワーク上で不要なマルチキャスト トラフィックを送信しないようにするには、IGMP スヌーピングを構成することをお勧めします。最初の例では、単一サイトと vSAN 6.5 以前を使用したサーバの単一ラックについて考えてみます。このバージョンはマルチキャストを使用するため、IGMP スヌーピングを有効にします。すべてが同じ L2 にあるため、マルチキャスト トラフィックのルーティングを設定する必要はありません。

vSAN 6.6 以降ではユニキャスト サポートが導入され、レイヤー 2 の実装がさらに簡素化されています。IGMP スヌーピングは必要ありません。



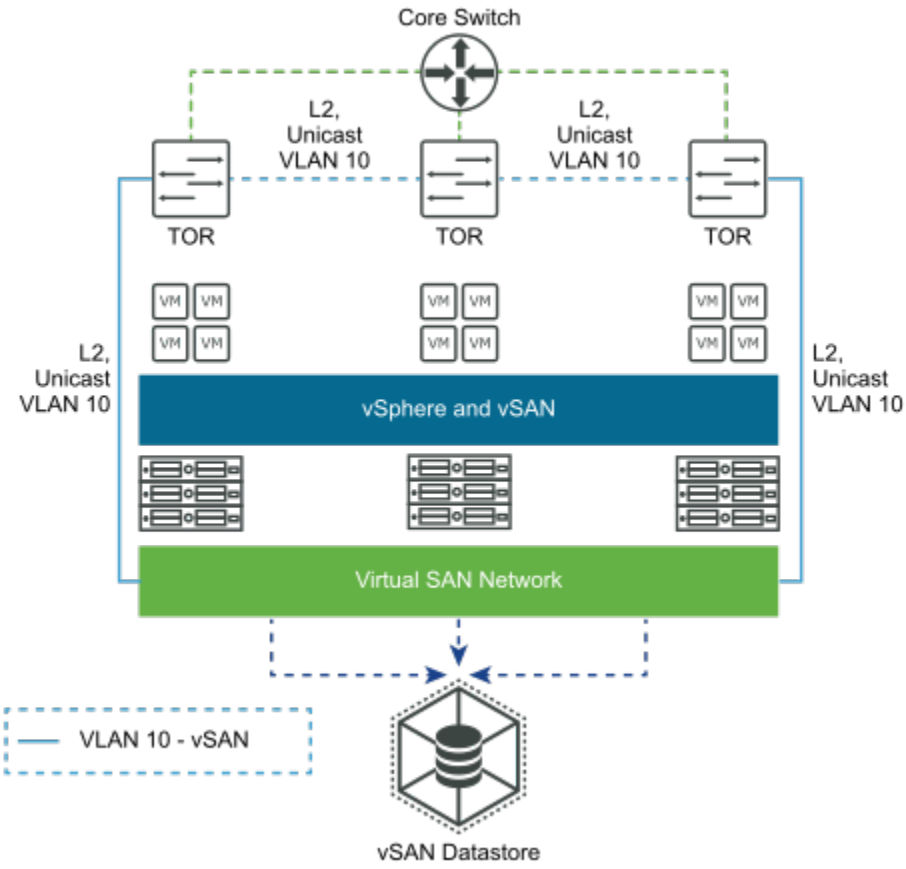
レイヤー 2、単一サイト、複数のラック

このネットワーク トポロジーは、複数のラックが存在し、複数のトップオブラック (TOR) のスイッチが 1 つのコアスイッチに接続さしているレイヤー 2 の実装で機能します。

次の図で TOR 間の青い点線は、vSAN ネットワークが使用可能で、vSAN クラスタのすべてのホストからアクセスできることを表しています。ただし、異なるラックのホストはレイヤー 3 で相互に通信するため、PIM を使用してホスト間でマルチキャスト トラフィックをルーティングします。物理的には TOR は相互に接続されていません。

ネットワーク上の不要なマルチキャスト トラフィックが発生しないように、すべての TOR で IGMP スヌーピングを設定することをお勧めします。トラフィックがルーティングされないため、マルチキャスト トラフィックをルーティングするように PIM を設定する必要はありません。

vSAN トラフィックはユニキャストのため、vSAN 6.6 以降ではこの実装がさらに簡単になります。ユニキャスト トラフィックでは、スイッチに IGMP スヌーピングを設定する必要はありません。



レイヤー 3、単一サイト、複数のラック

このネットワーク トポロジは、vSAN トラフィックのルーティングにレイヤー 3 が使用される vSAN 環境で機能します。

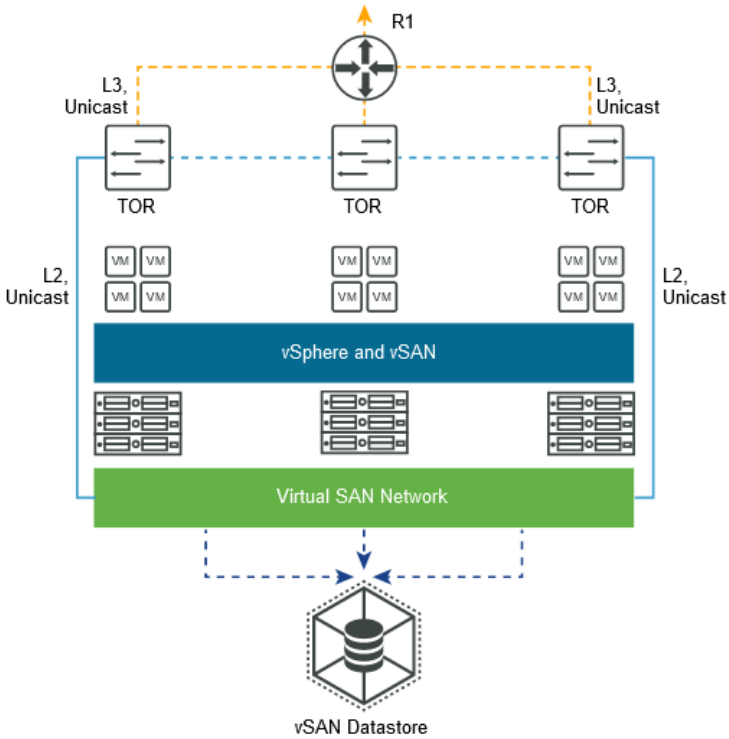
このシンプルなレイヤー 3 ネットワーク トポロジは、同じデータセンター内の複数のラックを使用します。それぞれのラックに TOR スイッチがあります。vSAN クラスタ内のすべてのホストが通信できるように、L3 を介して異なるラック間で vSAN ネットワークをルーティングします。vSAN VMkernel を別のサブネットまたは VLAN に配置し、ラックごとに個別のサブネットまたは VLAN を使用します。

このネットワーク トポロジーは、ルーターやレイヤー 3 対応スイッチなど、レイヤー 3 対応の中間デバイスを介してパケットをルーティングします。異なるレイヤー 3 ネットワーク セグメントにまたがってホストを展開すると、ルーティングされたネットワーク トポロジーが形成されます。

vSAN 6.5 以前ではマルチキャストを必要とするため、IGMP スヌーピングの使用と設定をお勧めします。物理スイッチで PIM を設定して、マルチキャスト トラフィックのルーティングを容易にします。

vSAN 6.6 以降では、このトポロジが簡素化されています。マルチキャスト トラフィックがないため、IGMP スヌーピングを設定する必要はありません。マルチキャスト トラフィックをルーティングするように PIM を設定する必要はありません。

ここでは、L3 での vSAN 6.6 環境の概要について説明します。マルチキャスト トラフィックがないため、IGMP スヌーピングまたは PIM の要件はありません。



vSAN ストレッチ クラスタの展開

vSAN では、2 つの場所にまたがるストレッチ クラスタを展開できます。

vSAN 6.5 以前では、データ サイト間の vSAN トラフィックは [マルチキャスト]（メタデータ）と [ユニキャスト] (I/O) になります。

vSAN 6.6 以降では、すべてのトラフィックが [ユニキャスト] になります。データ サイトと Witness（監視）ホスト間の監視トラフィックは、vSAN のすべてのバージョンでユニキャストになります。

レイヤー 2 のすべての場所

vSAN ストレッチ クラスタはレイヤー 2 ネットワーク内で構成できますが、この構成はおすすめしません。

vSAN ストレッチ クラスタが 1 つの大きなレイヤー 2 設計で構成されている場合について考えてみましょう。データ サイト 1 とサイト 2 は、仮想マシンが展開される場所です。サイト 3 には Witness（監視）ホストが存在します。

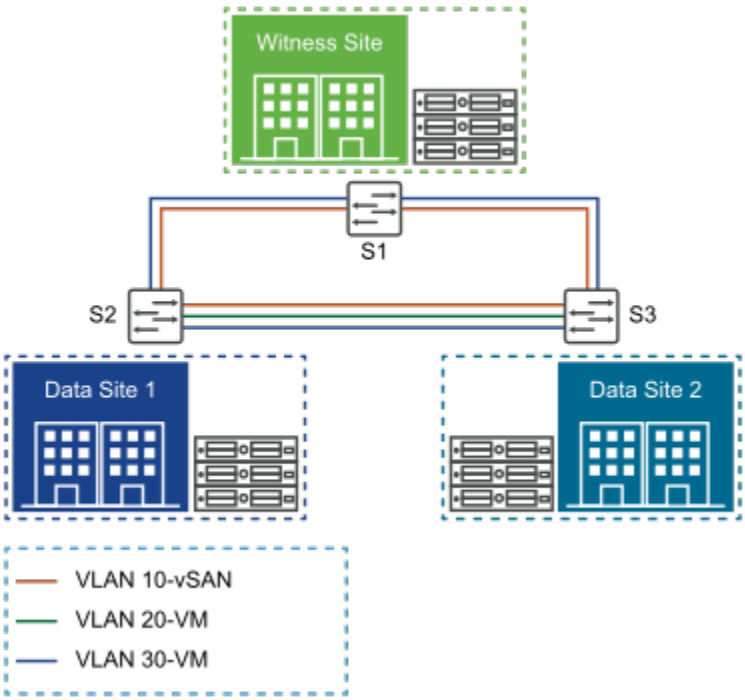
注： 最適な結果を得るには、すべてのサイトにまたがる拡張レイヤー 2 ネットワークを使用しないでください。

レイヤー 2 の場所を分かりやすくするため、トポロジではルーターではなくスイッチを使用します。

レイヤー 2 ネットワークにはループ（複数のパス）を含めることはできません。このため、サイト 1 とサイト 2 間の接続の 1 つをブロックするには、スパニング ツリー プロトコル (STP) などの機能が必要になります。ここで、サイト 2 とサイト 3 間のリンク（サイト 1 とサイト 2 間のリンク）が切断されている状況について考えてみます。ネットワーク トラフィックは、サイト 3 の Witness（監視）ホストを介してサイト 1 からサイト 2 にスイッチングされます。Witness（監視）ホストでは非常に狭いバンド幅と長い遅延が許容されるため、データ ネットワーク トラフィックが低い仕様の監視サイトを通過すると、パフォーマンスが大幅に低下します。

監視サイトを介してデータ サイト間のトラフィックをスイッチングしても、アプリケーションの遅延には影響はなく、バンド幅が許容できる範囲内であれば、サイト間に拡張 L2 構成を使用できる場合があります。ほとんどの場合、このような構成は現実的でなく、ネットワーク要件も複雑になります。

マルチキャスト トラフィックを使用する vSAN 6.5 以前では、スイッチで IGMP スヌーピングを構成する必要があります。vSAN 6.6 以降では、このような操作は必要ありません。マルチキャスト トラフィックのルーティングが行われないため、PIM は必要ありません。



サポートされる vSAN ストレッチ クラスタ構成

vSAN は、ストレッチ クラスタ構成をサポートします。

以下の構成では、データ サイトのネットワークのいずれかで障害が発生した場合、サイト 1 からサイト 2 へのトラフィックが Witness（監視）ホスト経由でルーティングされなくなります。この構成は、パフォーマンスの低下を防ぐことができます。データ トラフィックが Witness（監視）ホスト経由でスイッチングされないようにするには、次のネットワーク トポロジを使用します。

サイト 1 とサイト 2 の間で、拡張レイヤー 2 のスイッチング構成またはレイヤー 3 のルーティング構成を実装します。両方の構成がサポートされます。

サイト 1 と Witness（監視）ホストの間で、レイヤー 3 のルーティング構成を実装します。

サイト 2 と Witness（監視）ホストの間で、レイヤー 3 のルーティング構成を実装します。

これらの構成（L2 + L3 と L3 のすべての場所）と一緒に、マルチキャスト（vSAN 6.5 以前）、ユニキャストのみ（vSAN 6.6 で使用可能）に関する注意事項を示します。マルチキャスト トラフィックの場合、IGMP スヌーピング用に追加の構成手順が必要になります。また、マルチキャスト トラフィックのルーティングで PIM を使用されます。

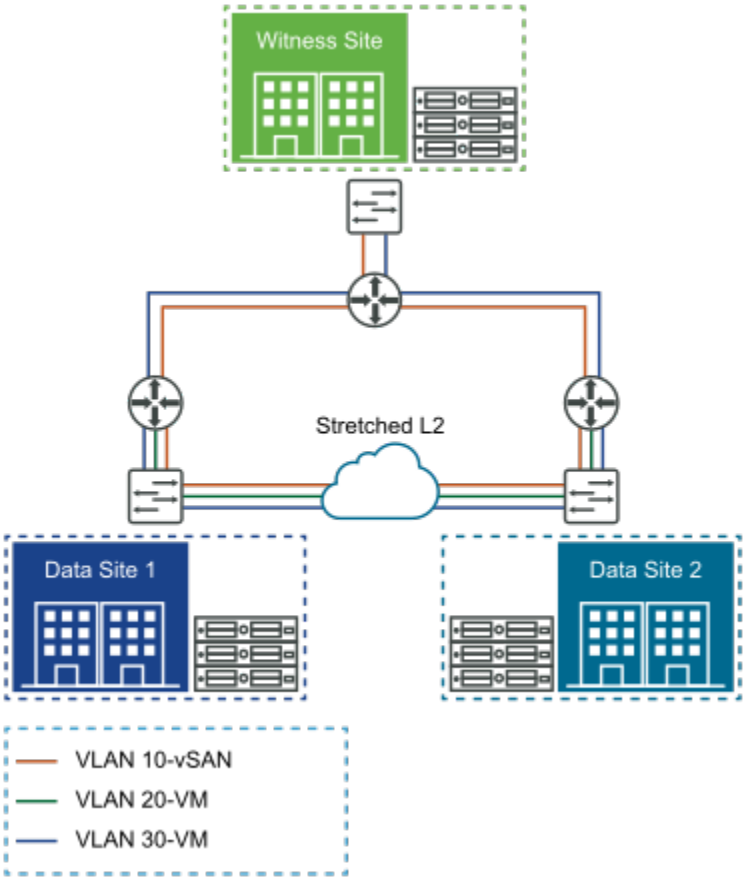
データ サイト間の拡張レイヤー 2 ネットワークと、監視サイトにルーティングされているレイヤー 3 ネットワークが確認されます。レイヤー 2 とレイヤー 3 の組み合わせをできるだけ単純に表すため、トポロジではスイッチとルーターの組み合わせを使用しています。

データ サイト間の拡張レイヤー 2、Witness（監視）ホストへのレイヤー 3

vSAN は、データ サイト間の拡張レイヤー 2 構成をサポートします。

この場合、ルーティングされるのは監視トラフィックだけです。マルチキャストを使用する vSAN 6.5 以前では、データ サイト間の拡張 L2 vSAN でマルチキャスト トラフィックに IGMP スヌーピングを使用します。ただし、監視トラフィックはユニキャストのため、レイヤー 3 のセグメントに PIM を実装する必要はありません。

ユニキャストを使用する vSAN 6.6 では、IGMP スヌーピングまたは PIM を考慮する必要はありません。



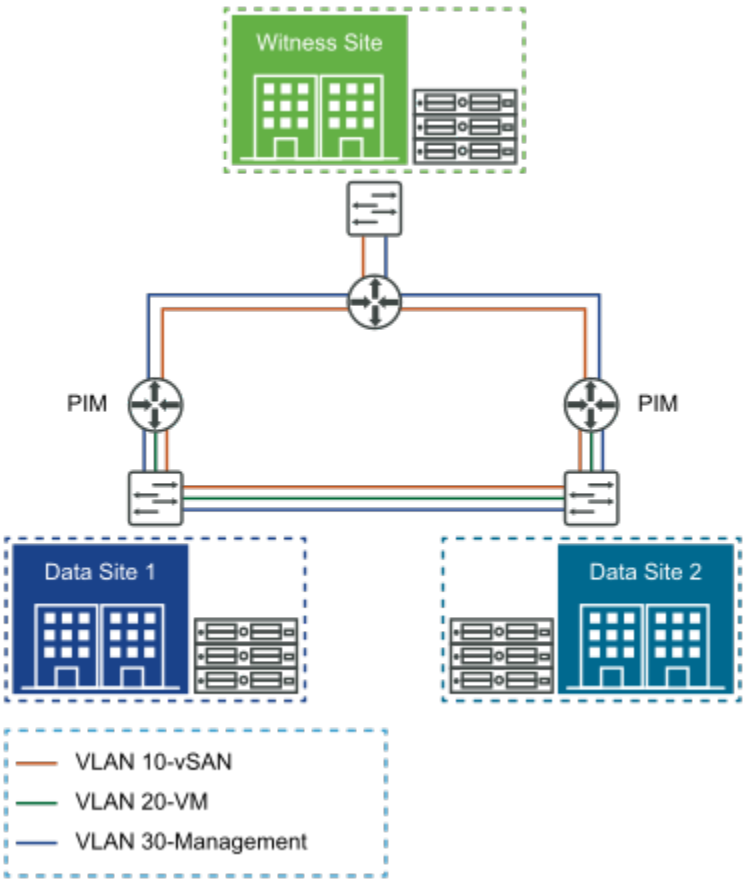
レイヤー 3 のすべての場所

この vSAN ストレッチ クラスタ構成では、データ トラフィックはデータ サイトと Witness（監視）ホスト間でルーティングされます。

レイヤー 3 を分かりやすい場所に実装するため、トポロジではルーターまたはルーティング スイッチを使用します。

たとえば、マルチキャスト トラフィックを使用する vSAN 6.5 以前の環境について考えてみましょう。この場合、ネットワーク上のマルチキャスト トラフィックの量を管理するため、データ サイトのスイッチで IGMP スヌーピングを設定します。監視トラフィックはユニキャストのため、この操作を Witness（監視）ホストで行う必要はありません。このマルチキャスト トラフィックはデータ サイト間でルーティングされるため、マルチキャスト ルーティングを許可するように PIM を設定します。

vSAN 6.6 以降では、ルーティングされたすべてのトラフィックがユニキャストのため、IGMP スヌーピングも PIM も必要ありません。



vSAN ストレッチ クラスタでの監視トラフィックの分離

vSAN は、ストレッチ クラスタでの監視トラフィックの分離をサポートします。

vSAN 6.5 以降のリリースでは、2 ノード構成で vSAN トラフィックから監視トラフィックを分離できます。つまり、10 Gb スイッチを介さずに 2 台の vSAN ホストを直接接続できます。

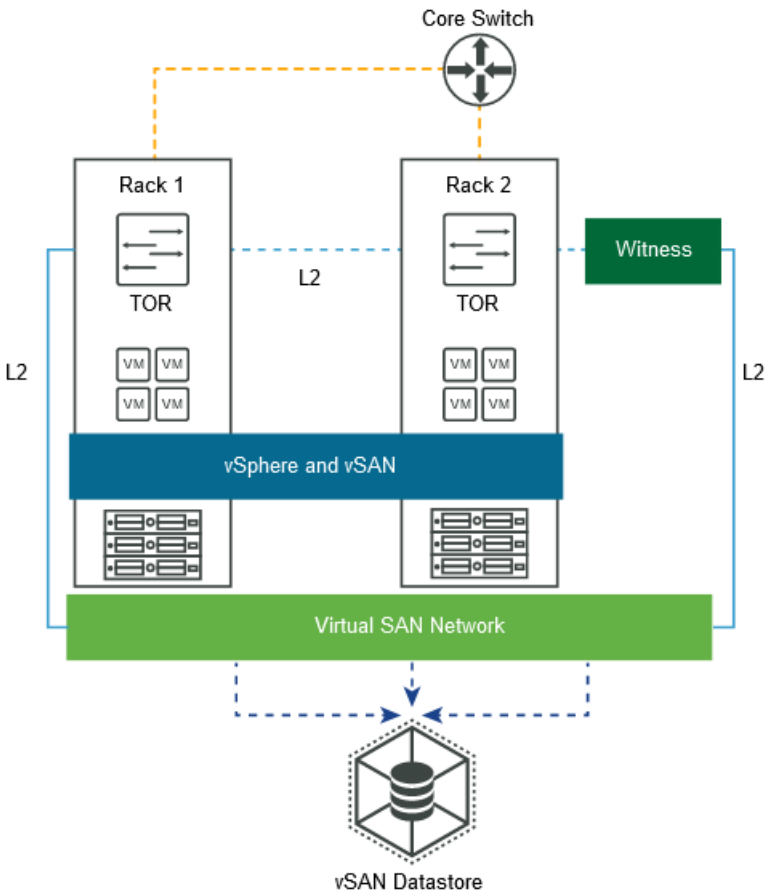
この監視トラフィックの分離は、vSAN 6.6 の 2 ノード構成の展開でのみサポートされます。vSAN ストレッチ クラスタでの監視トラフィックの分離は、vSAN 6.7 以降でサポートされています。

vSAN ストレッチ クラスタでのラック認識

vSAN ストレッチ クラスタを使用すると、vSAN は単一サイトでラックを認識できます。

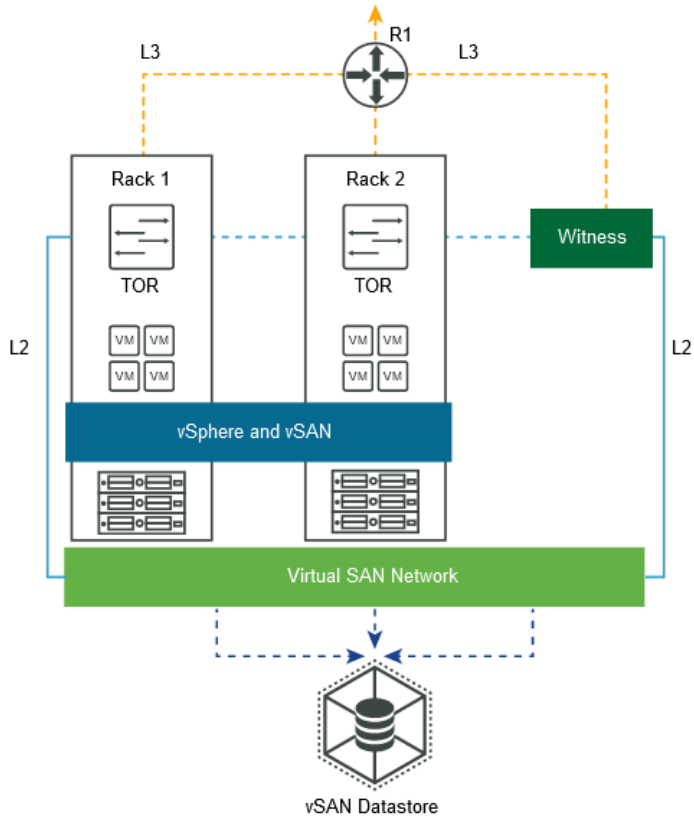
vSAN ホストが 2 つのラックに格納されている場合、1 つのラックで障害が発生しても vSAN クラスタの実行を継続できます。この場合、残りのラックとリモート Witness（監視）ホストにより、仮想マシン ワークロードの可用性が提供されます。

注： この構成を使用する場合は、vSAN ホストが格納されている 2 つのラック内に Witness（監視）ホストを配置しないでください。



この例では、ラック 1 に障害が発生した場合、ラック 2 と Witness（監視）ホストにより仮想マシンの可用性が提供されます。この構成は vSAN 6.6 より前の環境で、ネットワークでマルチキャストが構成されている必要があります。Witness（監視）ホストは、vSAN ネットワーク上にある必要があります。監視トラフィックはユニキャストです。vSAN 6.6 以降では、すべてのトラフィックがユニキャストになります。

このトポロジは L3 経由でもサポートされます。vSAN VMkernel を別のサブネットまたは VLAN に配置し、ラックごとに個別のサブネットまたは VLAN を使用します。



このトポロジは、2 つのラックを使用する展開をサポートしており、vSAN ストレッチ クラスタでラック認識（フォルト ドメイン）を実現しています。このソリューションでは、クラスタの外部にある Witness（監視）ホストを使用します。

2 ノード構成の vSAN の展開

vSAN は、2 ノード構成の展開をサポートします。2 ノード構成の vSAN 展開は、高可用性が必要な少数のワークロードを実行するリモート オフィスや支社 (ROBO) でよく使用されます。

2 ノード構成の vSAN の展開では 3 台目の監視ホストを使用します。このホストは、支社から離れた場所に設置できます。多くの場合、監視機能は vCenter Server などの管理コンポーネントと一緒に支社に維持されます。

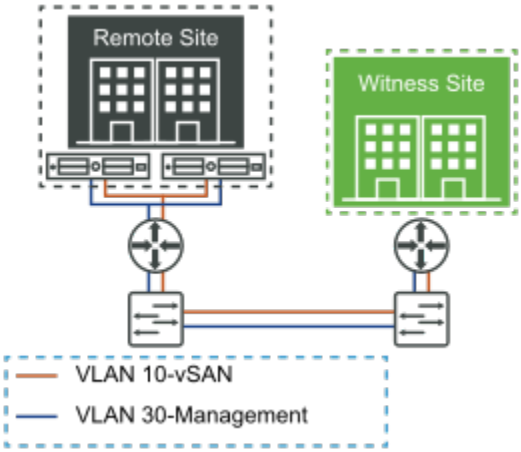
vSAN 6.5 より前での 2 ノード構成の vSAN の展開

6.5 より前の vSAN で 2 ノード構成の展開をサポートするには、リモート サイトに物理スイッチが必要です。

初期の 2 ノード構成の vSAN では、物理 10 Gb スイッチをリモート サイトに配置する必要があります。このリモート サイトのサーバのみが vSAN ホストの場合、これは非効率なソリューションになる可能性があります。

この展開で 10 Gb スイッチを使用しているデバイスが他にない場合は、IGMP スヌーピングを考慮する必要はありません。リモート サイトの他のデバイスと 10 Gb スイッチを共有している場合は、IGMP スヌーピングを行うと過剰なマルチキャスト トラフィックを防止できます。

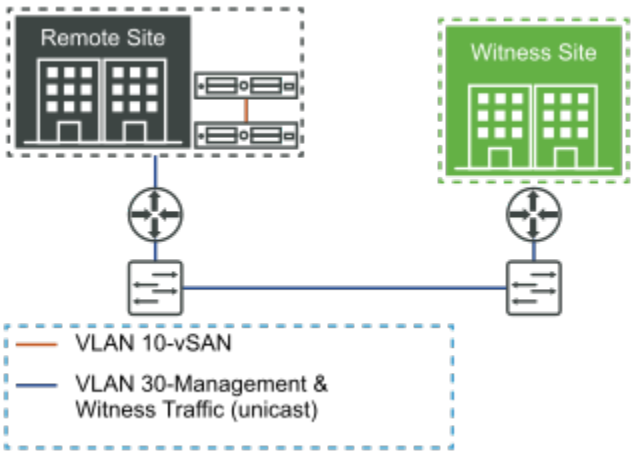
ルーティングされたトラフィックのみが監視トラフィックになり、これはユニキャストのため、PIM は必要ありません。



vSAN 6.5 以降での 2 ノード構成の展開

vSAN 6.5 以降は、2 ノード構成の展開をサポートします。

vSAN バージョン 6.5 以降の場合、この 2 ノード構成の vSAN は非常に簡単に実装できます。vSAN 6.5 以降では、データサイトの 2 台のホストを直接接続できます。

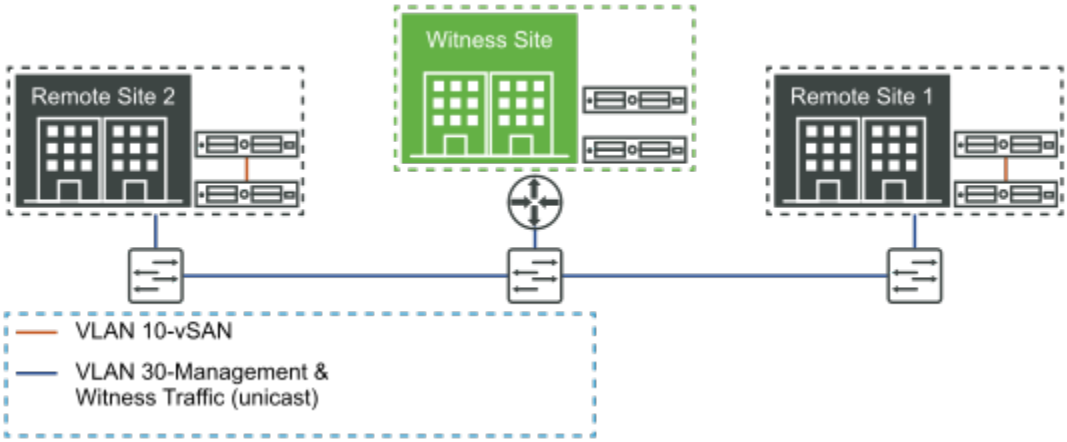


この機能を有効にするため、監視トラフィックが vSAN データ トラフィックから完全に分離されています。これにより、vSAN のデータ トラフィックを直接接続された 2 つのノード間で送受信し、監視トラフィックを管理ネットワーク経由で監視サイトにルーティングできます。

監視アプライアンスは、支店から離れた場所に配置できます。たとえば、管理インフラストラクチャ（vCenter Server、vROps、Log Insight など）と一緒にメインのデータセンターで監視を実行することもできます。監視を支店から離れた場所で実行できる場所としては vCloud Air もあります。

この構成では、リモート サイトにスイッチがありません。そのため、vSAN バックツーバック ネットワークのマルチキャスト トラフィックをサポートするように構成する必要はありません。監視トラフィックがすべてユニキャストのため、管理ネットワークでマルチキャストを考慮する必要もありません。

vSAN 6.6 以降では、すべてユニキャストを使用するので、マルチキャストに関する考慮事項はありません。それぞれ独自に監視を行う限り、複数のリモート オフィス/支店で 2 ノード構成を展開することも可能です。



2 ノード構成の vSAN の展開に関する一般的な考慮事項

2 ノード構成の vSAN の展開では、他のトポロジもサポートされます。このセクションでは、一般的な構成について説明します。

2 ノード構成の詳細と、ネットワーク外に展開する場合の考慮事項については、「vSAN のコア ドキュメント」を参照してください。

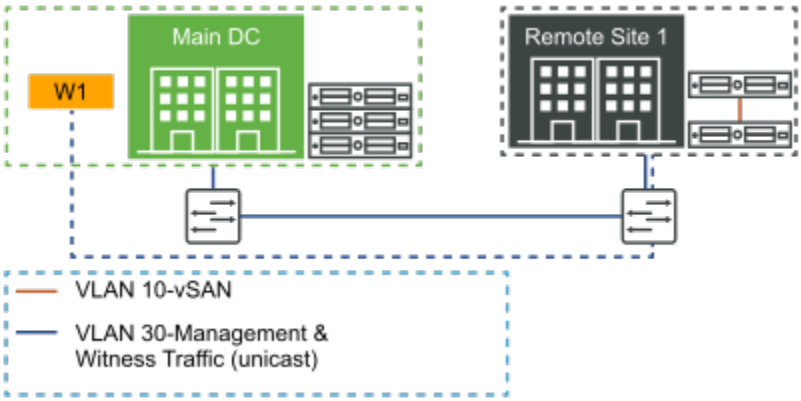
別の 2 ノード vSAN クラスタでの監視の実行

vSAN は、別の 2 ノード構成クラスタでの監視の実行をサポートしていません。

別の標準 vSAN 展開での監視の実行

vSAN は、別の標準 vSAN 展開での監視の実行をサポートします。

この構成はサポートされています。リモート サイトにある 2 ノード構成の vSAN で障害が発生しても、メインのデータセンターにある標準の vSAN 環境の可用性には影響ありません。



データ サイトから監視ホストへのネットワークの構成

データ サイトのホスト インターフェイスは、vSAN ネットワークを介して監視ホストと通信します。いくつかの構成オプションがあります。

このトピックでは、このような構成を実装する方法について説明します。ここでは、データ サイト内で vSAN ネットワークを介して相互に通信するホストのインターフェイスが監視ホストと通信する方法について説明します。

オプション 1：スタティックルートを使用して L3 経由で物理 ESXi 監視に接続する

データ サイトにはストレッチ L2 ネットワーク経由で接続できます。これは、データ サイトの管理ネットワーク、vSAN ネットワーク、vMotion ネットワーク、仮想マシン ネットワークの場合にも使用します。

このネットワーク インフラストラクチャの物理ネットワーク ルーターは、データ サイト（サイト 1 とサイト 2）内のホストから監視サイト（サイト 3）のホストにトラフィックを自動的に転送しません。vSAN ストレッチ クラスタを正常に構成するには、クラスタ内のすべてのホストが相互に通信を行う必要があります。この環境には vSAN ストレッチ クラスタを展開できます。

この解決策では、サイト 1 とサイト 2 からの vSAN トラフィックがサイト 3 の監視ホストに到達できるように、ESXi ホストで構成されているスタティック ルートを使用します。データ サイトの ESXi ホストの場合は、vSAN インターフェイスにスタティック ルートを追加します。これにより、そのネットワークに対して指定されたゲートウェイを経由してサイト 3 の監視ホストにトラフィックがリダイレクトされます。監視ホストの場合は、vSAN インターフェイスにスタティックルートが追加されている必要があります。これにより、データサイト内のホストに送信されるトラフィック vSAN トラフィックがリダイレクトされます。vSAN ストレッチ クラスタ内の各 ESXi ホストにスタティック ルートを追加するには、次のコマンドを使用します。

esxcli network ip route ipv4 add -g <gateway> **-n** <network>

注： vCenter Server では、データ サイトと監視サイトの両方の ESXi ホストを管理する必要があります。監視ホストから vCenter Server への直接接続が確立されていれば、管理ネットワークについてさらに考慮すべき点はありません。

vMotion ネットワークや仮想マシン ネットワークを構成する必要はありません。vSAN ストレッチ クラスタのコンテキストで、これらのネットワークにスタティック ルートを追加する必要はありません。仮想マシンは、vSAN 監視ホストに移行または展開されません。これは、監視オブジェクトのみを保持することを目的としています。このタスクでは、これらのネットワークは必要ありません。

オプション 2：スタティック ルートを使用して L3 経由で仮想 ESXi 監視アプライアンスに接続する

監視ホストは、vSAN クラスタの一部ではない物理 ESXi ホストに展開される仮想マシンです。このため、物理 ESXi ホストには少なくとも 1 台の仮想マシン ネットワークが事前に構成されている必要があります。この仮想マシン ネットワークは、管理ネットワークと、データ サイトの ESXi ホストによって共有されている vSAN ネットワークの両方にアクセスする必要があります。

注： 監視ホストは、専用ホストである必要はありません。監視をホストしながら同時に他の多くの仮想マシン ワークロードに使用できます。

また、基盤となる物理 ESXi ホスト上に 2 つの仮想マシン ネットワークを事前に構成し、1 つを管理ネットワーク用、もう 1 台を vSAN ネットワーク用に使用することもできます。この物理 ESXi ホストに仮想 ESXi 監視が展開されている場合は、ネットワークを適切に接続して構成する必要があります。

仮想 ESXi 監視ホストを展開したら、スタティック ルートを構成します。たとえば、データ サイトがストレッチ L2 ネットワークを介して接続されているとします。これは、データ サイトの管理ネットワーク、vSAN ネットワーク、vMotion ネットワーク、仮想マシン ネットワークの場合にも使用します。デフォルト ゲートウェイを介してデータ サイト（サイト 1 とサイト 2）内のホストから監視サイト（サイト 3）のホストに vSAN トラフィックがルーティングされません。vSAN ストレッチ クラスタを正常に構成するには、クラスタ内のすべてのホストにスタティック ルートが必要です。これにより、サイト 1 とサイト 2 の vSAN トラフィックがサイト 3 の監視ホストに到達できるようになります。**`esxcli network ip route`** コマンドを使用して、各 ESXi ホストにスタティック ルートを追加します。

例外的な展開

vSAN は、通常でない例外的な構成にも展開できます。

このような例外的なトポロジには特別な考慮事項があります。

3 つの場所、vSAN ストレッチ クラスタなし、分散監視ホスト

ストレッチ クラスタ構成を展開するのではなく、複数の部屋、建物、またはサイトに vSAN を展開できます。

この構成はサポートされています。1 つの要件として、サイト間の遅延は、同じデータセンター内の通常の vSAN 環境で想定されている遅延と同じレベルにする必要があります。遅延は、すべてのホスト間で [1 ミリ秒未満] にする必要があります。遅延がこの値を超える場合は、5 ミリ秒の遅延が許容される vSAN ストレッチ クラスタを検討してください。vSAN 6.5 以前の場合は、マルチキャストについて別の事項も考慮する必要があります。

最適な結果を得るには、トポロジ内のすべてのサイトで統一された構成を維持する必要があります。仮想マシンの可用性を維持するため、それぞれの部屋、建物またはサイトのホストが同じフォルト ドメインに配置されるようにフォルト ドメインを構成します。クラスタの非対称パーティショニングは避けてください。この場合、ホスト A からホスト B に接続できなくても、ホスト B からホスト A への接続は可能になります。

2 ノード構成を 1+1+W ストレッチ クラスタとして展開

2 ノード構成を vSAN ストレッチ クラスタ構成として展開し、各ホストを異なる部屋、建物またはサイトに配置することができます。

各サイトでホスト数を増やすと、ライセンス関連のエラーが発生します。クラスタに 2 台以上のホストが存在し、専用監視アプライアンス/ホスト機能を使用する場合（N+N+W、N > 1）、この構成は vSAN ストレッチ クラスタと見なされます。

vSAN ネットワークのトラブルシューティング

12

vSAN では、正しく設定されていない vSAN ネットワークに起因するさまざまな問題を確認し、トラブルシューティングを行うことができます。

vSAN の処理はネットワークの構成、信頼性、パフォーマンスに依存します。サポート リクエストの多くは、ネットワークの構成に問題があるか、ネットワークのパフォーマンスの低下が原因で発生しています。

vSAN 健全性サービスを使用して、ネットワークの問題を解決してください。ネットワーク健全性チェックでは、健全性チェックの結果に応じて、適切なナレッジベースの記事が表示されます。ナレッジベースの記事には、ネットワークの問題を解決する手順が記載されています。

ネットワーク健全性チェック

健全性サービスでは、カテゴリ別にネットワーク健全性チェックが実行されます。

各健全性チェックには、[AskVMware] リンクがあります。健全性チェックが失敗した場合は、[AskVMware] をクリックして、関連する VMware のナレッジベースの記事を参照します。ここで問題の詳細や問題の解決方法を確認します。

次のネットワーク健全性チェックを使用すると、vSAN 環境に関する有用な情報を取得できます。

- vSAN[：基本（ユニキャスト）接続チェック]。このチェックでは、vSAN ネットワーク上の ESXi ホスト間で ping を送信することにより、vSAN クラスタ内のすべての ESXi ホストに IP 接続が存在することを確認します。
- [vMotion：基本（ユニキャスト）接続チェック]。このチェックでは、vMotion が構成されている vSAN クラスタ内のすべての ESXi ホスト間に IP 接続が存在することを確認します。vMotion ネットワーク上の各 ESXi ホストが他のすべての ESXi ホストに ping を実行します。
- [すべてのホストで vSAN vmknic が構成済み]。このチェックでは、vSAN クラスタの各 ESXi ホストで、vSAN トラフィック用の VMkernel NIC が構成されていることを確認します。
- すべてのホストでマルチキャスト設定が一致。このチェックでは、各ホストに適切なマルチキャスト アドレスが設定されていることを確認します。
- [すべてのホストでサブネットが一致]。このチェックでは、すべての vSAN VMkernel NIC が同じ IP サブネットに存在するように、vSAN クラスタ内のすべての ESXi ホストが構成されていることをテストします。
- [vCenter Server から切断されたホスト]。このチェックでは、vCenter Server から vSAN クラスタ内のすべての ESXi ホストに対してアクティブな接続が存在することを確認します。

- [接続に問題のあるホスト]。このチェックでは、vCenter Server でホストが接続済みになっていても、vCenter Server からホストへの API 呼び出しが失敗している状況を確認します。ホストと vCenter Server 間の接続の問題が強調表示されます。
- [ネットワーク遅延]。このチェックでは、vSAN ホストのネットワーク遅延チェックを実行します。しきい値が 5 ミリ秒を超えると、警告が表示されます。
- [vMotion：MTU チェック（パケット サイズの大きい ping)]。このチェックは、vMotion の基本的な ping 接続チェックを補完します。ネットワーク パフォーマンスを向上させるため、最大転送ユニットのサイズが大きくなっています。誤って構成された MTU は、ネットワーク構成の問題として表示されていなくても、パフォーマンスに関する問題が発生する可能性があります。
- [vSAN クラスタ パーティション]。この健全性チェックは、クラスタに存在するパーティションの数を確認します。vSAN クラスタに複数のパーティションが存在すると、エラーが表示されます。
- [その他のチェックに基づくマルチキャスト評価]。この健全性チェックは、すべてのネットワーク健全性チェックのデータを集計します。このチェックに失敗した場合、ネットワーク パーティションの問題の根本原因がマルチキャストに可能性があります。

ネットワークを確認するコマンド

vSAN ネットワークが構成されている場合は、次のコマンドを使用してその状態を確認します。vSAN に使用されている VMkernel アダプタ (vmknic) とその属性を確認できます。

ESXCLI コマンドと RVC コマンドを使用して、ネットワークが完全に機能していることを確認し、vSAN のネットワークに関する問題を解決します。

vSAN ネットワークに使用されている vmknic がすべてのホストで正しく構成されていることを確認し、マルチキャストが機能していることを確認して、vSAN クラスタに参加しているホスト同士が正常に通信できることを確認できます。

esxcli vsan network list

このコマンドを使用すると、vSAN ネットワークで使用されている VMkernel インターフェイスを識別できます。

次の出力は、vSAN ネットワークで vmk2 が使用されていることを示しています。vSAN がオフになっていて、ホストが vSAN に参加していない場合でも、このコマンドは引き続き機能します。

エージェント グループのマルチキャストとマスター グループのマルチキャストも確認する必要があります。

```
[root@esxi-dell-m:~] esxcli vsan network list
Interface
   VmkNic Name: vmk1
   IP Protocol: IP
   Interface UUID: 32efc758-9ca0-57b9-c7e3-246e962c24d0
   Agent Group Multicast Address: 224.2.3.4
   Agent Group IPv6 Multicast Address: ff19::2:3:4
   Agent Group Multicast Port: 23451
   Master Group Multicast Address: 224.1.2.3
   Master Group IPv6 Multicast Address: ff19::1:2:3
   Master Group Multicast Port: 12345
```

```
    Host Unicast Channel Bound Port: 12321
    Multicast TTL: 5
    Traffic Type: vsan
```

これは、vSAN トラフィックに使用されている VMkernel インターフェイスなど、有用な情報を提供します。この場合は [vmk1] です。ただし、マルチキャスト アドレスも表示されます。クラスタがユニキャスト モードで実行されている場合でも、この情報が表示されることがあります。グループのマルチキャスト アドレスとポートも表示されています。ポート 23451 は、プライマリが毎秒送信するハートビートに使用され、クラスタ内の他のすべてのホストに表示されます。ポート 12345 は、プライマリとバックアップ間の CMMDS の更新に使用されます。

esxcli network ip interface list

このコマンドを使用すると、vSwitch または Distributed Switch などのアイテムを確認できます。

このコマンドを使用して、接続している vSwitch または Distributed Switch、MTU サイズを確認します。この情報は、環境内にジャンボ フレームが構成されている場合に役立ちます。この例では、MTU はデフォルトの 1,500 になっています。

```
[root@esxi-dell-m:~] esxcli network ip interface list
vmk0
   Name: vmk0
   <<truncated>>
vmk1
   Name: vmk1
   MAC Address: 00:50:56:69:96:f0
   Enabled: true
   Portset: DvsPortset-0
   Portgroup: N/A
   Netstack Instance: defaultTcpipStack
   VDS Name: vDS
   VDS UUID: 50 1e 5b ad e3 b4 af 25-18 f3 1c 4c fa 98 3d bb
   VDS Port: 16
   VDS Connection: 1123658315
   Opaque Network ID: N/A
   Opaque Network Type: N/A
   External ID: N/A
   MTU: 9000
   TSO MSS: 65535
   Port ID: 50331814
```

最大転送ユニット サイズは 9,000 と表示されているので、この VMkernel ポートはジャンボ フレーム用に構成され、約 9,000 の MTU が必要になります。ジャンボ フレームの使用に関しては特に推奨事項はありません。ただし、ジャンボ フレームは vSAN で使用できます。

esxcli network ip interface ipv4 get –i vmk2

このコマンドを実行すると、vSAN VMkernel インターフェイスの IP アドレスやネットマスクなどの情報が表示されます。

この情報を使用すると、管理者はコマンドラインで使用可能な他のコマンドを実行して、vSAN ネットワークが正常に動作しているかどうか確認できます。

```
[root@esxi-dell-m:~] esxcli network ip interface ipv4 get -i vmk1
Name  IPv4 Address  IPv4 Netmask   IPv4 Broadcast  Address Type  Gateway  DHCP DNS
----  ------------  -------------  --------------  ------------  -------  --------
vmk1  172.40.0.9   255.255.255.0   172.40.0.255   STATIC         0.0.0.0 false
```

vmkping

`vmkping` コマンドは、ネットワーク上の他のすべての ESXi ホストが ping 要求に応答しているかどうか確認します。

```
~ # vmkping -I vmk2 172.32.0.3 -s 1472 -d
 PING 172.32.0.3 (172.32.0.3): 56 data bytes
 64 bytes from 172.32.0.3: icmp_seq=0 ttl=64 time=0.186 ms
 64 bytes from 172.32.0.3: icmp_seq=1 ttl=64 time=2.690 ms
 64 bytes from 172.32.0.3: icmp_seq=2 ttl=64 time=0.139 ms

 --- 172.32.0.3 ping statistics ---
 3 packets transmitted, 3 packets received, 0% packet loss
 round-trip min/avg/max = 0.139/1.005/2.690 ms
```

マルチキャスト機能は確認されませんが、ネットワーク問題が発生している ESXi ホストの特定に役立ちます。また、応答時間を調べることで、vSAN ネットワークで異常遅延が発生していないかどうか確認することもできます。

ジャンボ フレームが構成され、ジャンボ フレームの MTU サイズが正しくない場合、このコマンドで問題を報告することはできません。デフォルトでは、このコマンドは 1,500 の MTU サイズを使用します。ジャンボ フレームがエンドツーエンドで正常に動作しているかどうか確認する場合は、次のように、より大きいパケット サイズ (-s) オプションを使用して vmkping を実行します。

```
 ~ # vmkping -I vmk2 172.32.0.3 -s 8972 -d
 PING 172.32.0.3 (172.32.0.3): 8972 data bytes
 9008 bytes from 172.32.0.3: icmp_seq=0 ttl=64 time=0.554 ms
 9008 bytes from 172.32.0.3: icmp_seq=1 ttl=64 time=0.638 ms
 9008 bytes from 172.32.0.3: icmp_seq=2 ttl=64 time=0.533 ms

 --- 172.32.0.3 ping statistics ---
 3 packets transmitted, 3 packets received, 0% packet loss
 round-trip min/avg/max = 0.533/0.575/0.638 ms
 ~ #
```

vmkping コマンドに-d を追加し、フラグメンテーションなしでパケットを送信できるかどうかテストすることも検討してください。

esxcli network ip neighbor list

このコマンドは、すべての vSAN ホストが同じネットワーク セグメント上にあるかどうかを確認するのに役立ちます。

この構成では 4 ホスト構成のクラスタを使用しています。このコマンドを実行すると、他の 3 台のホストの ARP（アドレス解決プロトコル）エントリが返されます。この中には、IP アドレスや vmknic（vSAN がこのクラスタ内のすべてのホストで vmk1 を使用するように構成した場合）も含まれます。

```
[root@esxi-dell-m:~] esxcli network ip neighbor list -i vmk1
Neighbor     Mac Address        Vmknic   Expiry  State  Type
-----------  -----------------  ------  -------  -----  -------
172.40.0.12  00:50:56:61:ce:22  vmk1    164 sec         Unknown
172.40.0.10  00:50:56:67:1d:b2  vmk1    338 sec         Unknown
172.40.0.11  00:50:56:6c:fe:c5  vmk1    162 sec         Unknown
[root@esxi-dell-m:~]
```

esxcli network diag ping

このコマンドでは、ネットワーク内の重複や往復時間を確認します。

さまざまなホスト間の vSAN ネットワーク接続の詳細については、ESXCLI が提供する強力なネットワーク診断コマンドを使用します。以下に、このような出力の例を示します。ここで VMkernel インターフェイスが vmk1 上にあり、ネットワーク上の別のホストのリモート vSAN ネットワーク IP アドレスが 172.40.0.10 です。

```
[root@esxi-dell-m:~] esxcli network diag ping -I vmk1 -H 172.40.0.10
   Trace:
         Received Bytes: 64
         Host: 172.40.0.10
         ICMP Seq: 0
         TTL: 64
         Round-trip Time: 1864 us
         Dup: false
         Detail:

         Received Bytes: 64
         Host: 172.40.0.10
         ICMP Seq: 1
         TTL: 64
         Round-trip Time: 1834 us
         Dup: false
         Detail:

         Received Bytes: 64
         Host: 172.40.0.10
         ICMP Seq: 2
         TTL: 64
         Round-trip Time: 1824 us
         Dup: false
         Detail:
   Summary:
         Host Addr: 172.40.0.10
         Transmitted: 3
         Recieved: 3
         Duplicated: 0
         Packet Lost: 0
         Round-trip Min: 1824 us
         Round-trip Avg: 1840 us
```

```
          Round-trip Max: 1864 us
[root@esxi-dell-m:~]
```

vsan.lldpnetmap

この RVC コマンドは、アップリンク ポートの情報を表示します。

環境でリンク層探索プロトコル (LLDP) が有効になっている Cisco 以外のスイッチが存在する場合、RVC コマンドを使用すると、アップリンク <-> スイッチ <-> スイッチ ポート情報を表示できます。RVC の詳細については、RVC コマンド ガイドを参照してください。

これは、vSAN クラスタが複数のスイッチにまたがる場合にどのホストがどのスイッチに接続するのかを判断する際に役立ちます。クラスタ内のホストのサブセットのみが影響を受ける場合は、特定のスイッチに対する問題を隔離するのに役立ちます。

```
> vsan.lldpnetmap 02013-08-15 19:34:18 -0700: This operation will take 30-60
seconds ...+---------------+---------------------------+| Host          | LLDP
info                 |+---------------+---------------------------+| 10.143.188.54 | w2r13-
vsan-x650-2: vmnic7 ||               | w2r13-vsan-x650-1: vmnic5 |+---------------
+---------------------------+
```

これは、LLDP をサポートするスイッチでのみ使用できます。これを構成するには、スイッチにログインし、次のように実行します。

```
switch# config t
Switch(Config)# feature lldp
```

LLDP が有効になっていることを確認するには:

```
switch(config)#do show running-config lldp
```

注： []デフォルトでは、LLDP は送信モードと受信モードの両方で動作します。物理スイッチ情報が見つからない場合は、vDS プロパティの設定を確認します。デフォルトでは、vDS は CDP（Cisco 検出プロトコル）に設定された検出プロトコルで作成されます。これを解決するには、検出プロトコルを LLDP に設定し、vDS で操作を [両方] に設定します。[]

マルチキャスト通信の確認

マルチキャスト構成では、vSAN の最初の展開で問題が発生する可能性があります。

vSAN 環境でマルチキャストが正常に機能しているかどうかを確認する最も簡単な方法の 1 つは、`tcpdump-uw` コマンドを使用することです。このコマンドは、ESXi ホストのコマンドラインから実行できます。

この `tcpdump-uw` コマンドを実行すると、プライマリがマルチキャスト パケット（ポートと IP 情報）を正しく送信し、クラスタ内の他のすべてのホストがそれらを受信しているかどうかを確認できます。

このコマンドを実行すると、プライマリではマルチキャスト アドレスに送信されたパケットが表示されます。その他のすべてのホストでは、同じパケットが表示されます（プライマリからマルチキャスト アドレスまで）。表示されない場合、マルチキャストは正常に動作していません。ここに示した `tcpdump-uw` コマンドをクラスタ内の任意のホストで実行し、プライマリからのハートビートを表示します。この場合、プライマリは IP アドレス 172.32.0.2 にあります。詳細表示の [-v] はオプションです。

```
[root@esxi-hp-02:~] tcpdump-uw -i vmk2 multicast -v
tcpdump-uw: listening on vmk2, link-type EN10MB (Ethernet), capture size 96 bytes
11:04:21.800575 IP truncated-ip - 146 bytes missing! (tos 0x0, ttl 5, id 34917, offset 0,
flags [none], proto UDP (17), length 228)
    172.32.0.4.44824 > 224.1.2.3.12345: UDP, length 200
11:04:22.252369 IP truncated-ip - 234 bytes missing! (tos 0x0, ttl 5, id 15011, offset 0,
flags [none], proto UDP (17), length 316)
    172.32.0.2.38170 > 224.2.3.4.23451: UDP, length 288
11:04:22.262099 IP truncated-ip - 146 bytes missing! (tos 0x0, ttl 5, id 3359, offset 0,
flags [none], proto UDP (17), length 228)
    172.32.0.3.41220 > 224.2.3.4.23451: UDP, length 200
11:04:22.324496 IP truncated-ip - 146 bytes missing! (tos 0x0, ttl 5, id 20914, offset 0,
flags [none], proto UDP (17), length 228)
    172.32.0.5.60460 > 224.1.2.3.12345: UDP, length 200
11:04:22.800782 IP truncated-ip - 146 bytes missing! (tos 0x0, ttl 5, id 35010, offset 0,
flags [none], proto UDP (17), length 228)
    172.32.0.4.44824 > 224.1.2.3.12345: UDP, length 200
11:04:23.252390 IP truncated-ip - 234 bytes missing! (tos 0x0, ttl 5, id 15083, offset 0,
flags [none], proto UDP (17), length 316)
    172.32.0.2.38170 > 224.2.3.4.23451: UDP, length 288
11:04:23.262141 IP truncated-ip - 146 bytes missing! (tos 0x0, ttl 5, id 3442, offset 0,
flags [none], proto UDP (17), length 228)
    172.32.0.3.41220 > 224.2.3.4.23451: UDP, length 200
```

この出力では少し分かりづらくなっていますが、クラスタ内の 4 台のホストがプライマリからハートビートを取得しています。すべてのホストでハートビートを受信していることを確認するには、この **`tcpdump-uw`** コマンドをすべてのホストで実行する必要があります。これにより、プライマリがハートビートを送信していて、クラスタ内の他のすべてのホストがハートビートを受信していることを確認できます。これらの条件を満たしている場合、マルチキャストは機能しています。

一部の vSAN ホストがプライマリから毎秒ハートビートを取得していない場合、ネットワーク管理者はスイッチのマルチキャスト構成を確認する必要があります。

煩わしい「**`truncated-ip - 146 bytes missing!`** 」メッセージが表示されないようにするには、同じコマンドに [–s0] オプションを使用して、パケットの切り捨てを停止する必要があります。

```
[root@esxi-hp-02:~] tcpdump-uw -i vmk2 multicast -v -s0
tcpdump-uw: listening on vmk2, link-type EN10MB (Ethernet), capture size 65535 bytes
11:18:29.823622 IP (tos 0x0, ttl 5, id 56621, offset 0, flags [none], proto UDP (17), length
228)
    172.32.0.4.44824 > 224.1.2.3.12345: UDP, length 200
11:18:30.251078 IP (tos 0x0, ttl 5, id 52095, offset 0, flags [none], proto UDP (17), length
228)
    172.32.0.3.41220 > 224.2.3.4.23451: UDP, length 200
11:18:30.267177 IP (tos 0x0, ttl 5, id 8228, offset 0, flags [none], proto UDP (17), length
316)
```

```
    172.32.0.2.38170 > 224.2.3.4.23451: UDP, length 288
11:18:30.336480 IP (tos 0x0, ttl 5, id 28606, offset 0, flags [none], proto UDP (17), length
228)
    172.32.0.5.60460 > 224.1.2.3.12345: UDP, length 200
11:18:30.823669 IP (tos 0x0, ttl 5, id 56679, offset 0, flags [none], proto UDP (17), length
228)
    172.32.0.4.44824 > 224.1.2.3.12345: UDP, length 200
```

tcpdump コマンドは、IGMP (Internet Group Management Protocol) のメンバーシップに関連しています。ホスト（とネットワーク デバイス）は、IGMP を使用してマルチキャスト グループ メンバーシップを確立します。

vSAN クラスタ内の各 ESXi ホストは、通常の IGMP メンバーシップ レポート (参加) を送信します。

tcpdump コマンドを実行すると、ホストからの IGMP メンバー レポートが表示されます。

```
[root@esxi-dell-m:~] tcpdump-uw -i vmk1 igmp
tcpdump-uw: verbose output suppressed, use -v or -vv for full protocol decode
listening on vmk1, link-type EN10MB (Ethernet), capture size 262144 bytes
15:49:23.134458 IP 172.40.0.9 > igmp.mcast.net: igmp v3 report, 1 group record(s)
15:50:22.994461 IP 172.40.0.9 > igmp.mcast.net: igmp v3 report, 1 group record(s)
```

この出力には IGMP v3 レポートの生成が表示されています。これは、ESXi ホストがそのメンバーシップを定期的に更新していることを意味します。vSAN ESXi ホストが IGMP を正しく実行しているかどうかネットワーク管理者が疑っている場合は、このコマンドをクラスタ内の各 ESXi ホストで実行して、このトレースを表示して検証できます。

マルチキャスト通信を使用している場合は、IGMP v3 を使用します。

実際には、次のコマンドを使用すると、マルチキャスト トラフィックと IGMP トラフィックを同時に確認できます。

```
[root@esxi-hp-02:~] tcpdump-uw -i vmk2 multicast or igmp -v -s0
```

よくある問題としては、vSAN クラスタが複数の物理スイッチ間で構成されているときに、1 台のスイッチでマルチキャストが有効になっていて、スイッチ間で有効になっていないことがあります。この状況では、クラスタが 1 つのパーティションにある 2 台の ESXi ホストから構成され、別の ESXi ホスト（他のスイッチに接続されているホスト）がこのクラスタに参加できません。代わりに、別のパーティションに独自の vSAN クラスタを形成します。前述の `vsan.lldpnetmap` コマンドを使用すると、ネットワーク構成を特定し、どのホストがどのスイッチに接続しているのかを確認できます。

vSAN クラスタが形成されている場合、マルチキャストの問題かどうか確認できる指標がいくつかあります。

たとえば、サブネット、VLAN、MTU のチェックリストを実行し、`vmkping` を実行してクラスタ内の各ホストがクラスタ内の他のすべてのホストと接続可能な状態であることが確認されているとします。

クラスタの作成時にマルチキャストの問題が発生した場合、それぞれの ESXi ホストが独自の vSAN クラスタを形成し、自身がプライマリとして機能していることがよくあります。各ホストに一意のネットワーク パーティション ID がある場合、この状況はホスト間にマルチキャストがないことを示しています。

ただし、ESXi ホストのサブセットがクラスタを形成し、別のサブセットが別のクラスタを形成して、それぞれが独自のプライマリ、バックアップ、エージェント ホストを持つ一意のパーティションを使用している場合、スイッチではマルチキャストが有効になっていますが、スイッチ間では有効になっていません。vSAN には、独自のクラスタ パーティションを形成している最初の物理スイッチのホストと、独自のクラスタ パーティションを形成している 2 番目の物理スイッチのホストが表示され、それぞれが独自のプライマリを使用しています。クラスタ内のホストが接続しているスイッチと、クラスタ内のホストが同じスイッチに接続していることを確認できる場合は、この問題が発生している可能性があります。

vSAN ネットワークのパフォーマンスの確認

ESXi ホスト間に十分なバンド幅があることを確認します。このツールは、vSAN ネットワークのパフォーマンスが最適かどうかをテストする際に役立ちます。

vSAN ネットワークのパフォーマンスを確認するには、`iperf` ツールを使用して TCP の最大バンド幅と遅延を測定します。このツールは `/usr/lib/vmware/vsan/bin/iperf.copy` にあります。`--help` を指定して実行すると、さまざまなオプションを確認できます。このツールを使用して、vSAN クラスタに参加している ESXi ホスト間のネットワークバンド幅と遅延を確認します。

設定とテストを行う方法については、VMware KB「2001003」を参照してください。

vSAN クラスタを使用している場合、これが最も便利なツールです。クラスタがすでに本番環境にある場合、vSAN ネットワークで [iperf] テストを実行すると、クラスタ上で実行されている仮想マシンのパフォーマンスに影響を及ぼす可能性があります。

vSAN ネットワークの制限の確認

`vsan.check.limits` コマンドは、違反している vSAN しきい値がないことを確認します。

```
> ls
0 /
1 vcsa-04.rainpole.com/
> cd 1
/vcsa-04.rainpole.com> ls
0 Datacenter (datacenter)
/vcsa-04.rainpole.com> cd 0
/vcsa-04.rainpole.com/Datacenter> ls
0 storage/
1 computers [host]/
2 networks [network]/
3 datastores [datastore]/
4 vms [vm]/
/vcsa-04.rainpole.com/Datacenter> cd 1
/vcsa-04.rainpole.com/Datacenter/computers> ls
0 Cluster (cluster): cpu 155 GHz, memory 400 GB
1 esxi-dell-e.rainpole.com (standalone): cpu 38 GHz, memory 123 GB
2 esxi-dell-f.rainpole.com (standalone): cpu 38 GHz, memory 123 GB
3 esxi-dell-g.rainpole.com (standalone): cpu 38 GHz, memory 123 GB
4 esxi-dell-h.rainpole.com (standalone): cpu 38 GHz, memory 123 GB
/vcsa-04.rainpole.com/Datacenter/computers> vsan.check_limits 0
2017-03-14 16:09:32 +0000: Querying limit stats from all hosts ...
```

```
2017-03-14 16:09:34 +0000: Fetching vSAN disk info from esxi-dell-m.rainpole.com (may take a
moment) ...
2017-03-14 16:09:34 +0000: Fetching vSAN disk info from esxi-dell-n.rainpole.com (may take a
moment) ...
2017-03-14 16:09:34 +0000: Fetching vSAN disk info from esxi-dell-o.rainpole.com (may take a
moment) ...
2017-03-14 16:09:34 +0000: Fetching vSAN disk info from esxi-dell-p.rainpole.com (may take a
moment) ...
2017-03-14 16:09:39 +0000: Done fetching vSAN disk infos
+--------------------------+--------------------
+------------------------------------------------------------------+
| Host                     | RDT                |
Disks                                                      |
+--------------------------+--------------------
+------------------------------------------------------------------+
| esxi-dell-m.rainpole.com | Assocs: 1309/45000 | Components:
485/9000                                             |
|                          | Sockets: 89/10000  | naa.500a075113019b33: 0% Components:
0/0                      |
|                          | Clients: 136       | naa.500a075113019b37: 40% Components:
81/47661                 |
|                          | Owners: 138        |
t10.ATA_____Micron_P420m2DMTFDGAR1T4MAX_____ 0% Components: 0/0 |
|                          |                    | naa.500a075113019b41: 37% Components:
80/47661                 |
|                          |                    | naa.500a07511301a1eb: 38% Components:
81/47661                 |
|                          |                    | naa.500a075113019b39: 39% Components:
79/47661                 |
|                          |                    | naa.500a07511301a1ec: 41% Components:
79/47661                 |
<<truncated>>
```

ネットワークの観点から見ると、重要なものは RDT の関連付け (Assocs) とソケット数です。vSAN 6.0 以降では、ホストごとに 45,000 の関連付けがあります。RDT の関連付けは、vSAN 内でピアツーピア ネットワークの状態を追跡するために使用されます。RDT の関連付けが不足しないように、vSAN のサイズが調整されます。vSAN では、使用を許可する TCP ソケットの数も制限され、TCP ソケットの割り当てが不足しないように vSAN のサイズが調整されます。ホストあたりのソケット数の上限は 10,000 です。

vSAN [クライアント] は、vSAN クラスタでのオブジェクトのアクセスを表します。クライアントは通常、ホストで実行されている仮想マシンを表します。クライアントとオブジェクトが同じホスト上にない場合もあります。ハード定義の制限はありませんが、ホスト間でのクライアントのバランスを理解するために、このメトリックが表示されます。

特定の vSAN オブジェクトには 1 つの vSAN [所有者] があり、通常、このオブジェクトにアクセスする vSAN クライアントと一緒に存在します。vSAN 所有者は、vSAN オブジェクトへのすべてのアクセスを調整し、ミラーリングやストライピングなどの機能を実装します。ハード定義の制限はありませんが、ホスト間での所有者のバランスを理解するために、このメトリックが表示されます。

vSAN ネットワークでのマルチキャストの使用

13

マルチキャストは、IP ネットワーク経由で宛先のグループに情報パケットを送信するネットワーク通信技術です。

vSAN バージョン 6.6 より前のリリースは IP マルチキャストをサポートしています。検出プロトコルとして IP マルチキャスト通信を使用し、vSAN クラスタに参加を試みているノードを識別します。vSAN バージョン 6.6 より前のリリースでは、IP マルチキャスト通信を使用して、クラスタ グループの参加や離脱、クラスタ内の他の通信処理を行います。vSAN トラフィック サービスを実行するには、IP ネットワーク セグメントで IP マルチキャストを有効にして構成します。

IP マルチキャスト アドレスは、マルチキャスト グループ (MG) といいます。IP マルチキャストは、グループ転送として複数の受信者にソース パケットを送信します。IP マルチキャストは、ホスト、クライアント、ネットワーク デバイスがマルチキャストベースの通信に参加するために使用する通信プロトコルに依存します。Internet Group Management Protocol (IGMP) などの通信プロトコルとプロトコルに依存しないマルチキャスト (PIM) は、IP マルチキャスト通信を使用する際の主要なコンポーネントと依存関係です。

vSAN クラスタの作成中に、各 vSAN クラスタにデフォルトのマルチキャスト アドレスが割り当てられます。vSAN トラフィック サービスは、デフォルトのマルチキャスト アドレスの設定を各ホストに自動的に割り当てます。このマルチキャスト アドレスは、デフォルトのマルチキャスト グループとマルチキャスト グループのエージェントにフレームを送信します。

複数の vSAN クラスタが同じレイヤー 2 ネットワークに存在する場合、追加の vSAN クラスタ内でデフォルトのマルチキャスト アドレスを変更することをお勧めします。これにより、複数のクラスタがすべてのマルチキャスト ストリームを受信するのを防ぐことができます。デフォルトの vSAN マルチキャスト アドレスの変更方法については、VMware KB2075451 を参照してください。

次のトピックを参照してください。

- Internet Group Management Protocol
- プロトコルに依存しないマルチキャスト

Internet Group Management Protocol

Internet Group Management Protocol (IGMP) を使用すると、レイヤー 2 ドメイン内の IP マルチキャスト グループ メンバーシップに受信側を追加できます。

IGMP を使用すると、受信側は参加するマルチキャスト グループに要求を送信できるようになります。マルチキャスト グループのメンバーになると、ルーターはマルチキャスト グループのトラフィックをレイヤー 3 セグメントで転送し、受信側はこのセグメントのスイッチ ポートに接続します。

IGMP スヌーピングを使用すると、マルチキャスト グループに参加している物理スイッチ ポートを vSAN VMkernel ポート アップリンクに制限できます。IGMP スヌーピングは IGMP スヌーピング クエリで設定されます。IGMP スヌーピングをサポートするよう IGMP スヌーピングクエリを設定する必要があるかどうかは、スイッチ ベンダーによって異なります。IGMP スヌーピングの設定については、スイッチ ベンダーに問い合わせてください。

vSAN は、IGMP バージョン 2 と IGMP バージョン 3 の両方をサポートしています。レイヤー 3 ネットワーク セグメントにわたって vSAN を展開すると、同じレイヤー 3 ネットワーク セグメントに接続し、アクセスできるルーターやスイッチなどのレイヤー 3 対応デバイスを構成できます。

vSAN ネットワーク上のすべての VMkernel ポートは、IGMP を使用してマルチキャスト グループにサブスクライブし、すべてのネットワーク ポートでマルチキャスト フラッディングを回避します。

注： vSAN が、ルーティングまたはトランクされていない VLAN にあり、クラスタ内のすべてのホストの vSAN ポートに拡張可能な場合は、IGMP スヌーピングを無効にできます。

プロトコルに依存しないマルチキャスト

プロトコルに依存しないマルチキャスト (PIM) は、レイヤー 3 マルチキャスト ルーティング プロトコルで構成されます。

マルチキャスト グループ ソースと異なるレイヤー 3 セグメントにある受信側に到達するため、IP マルチキャスト トラフィックに対して別の通信技術を提供します。vSAN バージョン 6.6 より前のクラスタで、異なるサブネット間でのマルチキャスト トラフィックのフローを有効にするは、PIM を使用する必要があります。PIM の実装については、ネットワーク ベンダーに問い合わせてください。

vSAN ファイル サービスのネットワークに関する考慮事項

14

vSAN ファイル サービスは vSAN の上にあるレイヤーで、ファイル共有を提供します。現在、SMB、NFSv3、NFSv4.1 ファイル共有をサポートしています。

vSAN ファイル サービスのネットワークに関する考慮事項は次のとおりです。

- vSAN ファイル サービス ネットワークからファイル サーバの IP アドレスとして固定 IP アドレスを割り当てる必要があります。各 IP アドレスは、vSAN ファイル共有へのアクセス ポイントになります。
  - 最適なパフォーマンスを実現するには、IP アドレスの数は vSAN クラスタ内のホスト数と同じにする必要があります。
  - すべての固定 IP アドレスは、同じサブネットのアドレスにする必要があります。
  - 各固定 IP アドレスには FQDN が対応しています。これは、DNS サーバの正引き参照ゾーンと逆引きゾーンの一部にする必要があります。
- ネットワークを vSAN ファイル サービス ネットワークとして準備する必要があります。
  - 標準スイッチ ベースのネットワークを使用している場合、vSAN ファイル サービス有効化プロセスで無作為検出モードと偽装転送が有効になります。
  - DVS ベースのネットワークを使用している場合、vSAN ファイル サービスは DVS バージョン 6.6.0 以降でサポートされています。DVS で vSAN ファイル サービス用の専用ポート グループを作成します。MacLearning と偽装転送は、指定された DVS ポート グループの vSAN ファイル サービス有効化プロセスで有効になります。

    注： NSX ベースのネットワークを使用している場合は、NSX 管理コンソールで指定のネットワーク エンティティで MacLearning が有効になっており、すべてのホストとファイル サービス ノードが目的の NSX-T ネットワークに接続していることを確認します。
- Kerberos セキュリティを使用する SMB 共有と NFS 共有の場合は、Active Directory ドメインと組織単位（オプション）に関する情報を指定する必要があります。また、オブジェクトを作成および削除するのに適切な権限を持つユーザー アカウントが必要です。
- ファイル サーバが Active Directory サーバと DNS サーバにアクセスできることを確認します。ファイル サーバは、Active Directory サービスで必要なすべてのポートにアクセスできる必要があります。

vSAN ファイル サービスがネットワーク接続に使用するポートは次のとおりです。これらのポートがファイアウォールによってブロックされていないことを確認します。

サービス	ポート番号	エンティティ	接続の要件
サーバ メッセージ ブロック (Server Message Block, SMB)	TCP ポート 445	ファイル サーバ	外部ネットワークからファイル サーバ
ローカル ファイル システムのユーザーの割り当て容量 (RQUOTA)	TCP ポート 875	ファイル サーバ	外部ネットワークからファイル サーバ
ネットワーク ファイル システム (Network File System, NFS)	TCP および UDP ポート 2049	ファイル サーバ	外部ネットワークからファイル サーバ。 NFSv3 は TCP ポートと UDP ポートの両方を使用できますが、NFSv4.1 は TCP のみを使用します。
NFS マウント	TCP および UDP ポート 20048	ファイル サーバ	外部ネットワークからファイル サーバ
ネットワーク ステータス モニター (Network Status Monitor, NSM) サーバ デーモン	TCP および UDP ポート 27689	ファイル サーバ	外部ネットワークからファイル サーバ。 内向きと外向きの両方の通信を許可する必要があります。
ネットワーク ロック マネージャ (Network Lock Manager, NLM)	TCP および UDP ポート 32803	ファイル サーバ	外部ネットワークからファイル サーバ。 ファイル サーバからクライアントへの接続を許可します。ファイアウォールで受信接続と送信接続を許可する必要があります。デフォルトのポートは UDP です。
Sun リモート プロシージャ コール (sunrpc)	TCP および UDP ポート 111	ファイル サーバ	外部ネットワークからファイル サーバ
LDAP	TCP ポート 389	Active Directory (AD) サーバ（Active Directory ドメインが構成されている場合）	ファイル サーバから Active Directory サーバ
LDAP からグローバル カタログ	TCP ポート 3268	Active Directory サーバ（Active Directory ドメインが構成されている場合）	ファイル サーバから Active Directory サーバ
Kerberos	TCP ポート 88	Active Directory サーバ（Active Directory ドメインが構成されている場合）	ファイル サーバから Active Directory サーバ
Kerberos パスワードの変更	TCP ポート 464	Active Directory サーバ（Active Directory ドメインが構成されている場合）	ファイル サーバから Active Directory サーバ
ドメイン ネーム サーバ (Domain Name Server, DNS)	TCP および UDP ポート 53	DNS サーバ	ファイル サーバから DNS サーバ

サービス	ポート番号	エンティティ	接続の要件
vSAN 分散ファイル システム (VDFS) サーバ	TCP ポート 1564	ESXi ホスト	vSAN ネットワーク内
リモート プロシージャ コール	TCP ポート 135	Active Directory サーバ（Active Directory ドメインが構成されている場合）	ファイル サーバから Active Directory サーバ
NetBIOS Session Service	TCP ポート 139	Active Directory サーバ（Active Directory ドメインが構成されている場合）	ファイル サーバから Active Directory サーバ
DNS	UDP ポート 53	Active Directory サーバ（Active Directory ドメインが構成されている場合）	ファイル サーバから Active Directory サーバ
LDAP、DC ロケータ、ネットログイン	UDP ポート 389	Active Directory サーバ（Active Directory ドメインが構成されている場合）	ファイル サーバから Active Directory サーバ
ランダムに割り当てられた大きい TCP ポート番号	TCP 49152 - 65535	Active Directory サーバ（Active Directory ドメインが構成されている場合）	ファイル サーバから Active Directory サーバ

vSAN での iSCSI のネットワークに関する考慮事項

15

vSAN iSCSI ターゲット サービスを使用すると、vSAN クラスタ外のホストおよび物理ワークロードが vSAN データストアにアクセスできます。この機能を使用すると、リモート ホスト上の iSCSI イニシエータが、ブロックレベルのデータを vSAN クラスタ内のストレージ デバイス上の iSCSI ターゲットに転送できます。

vSAN の iSCSI ターゲットは、他の vSAN オブジェクトと同様に、ストレージ ポリシー ベース管理 (SPBM) を使用して管理されます。iSCSI LUN の場合、この領域の容量はデデュープと圧縮によって節約されます。また、暗号化によるセキュリティも提供されます。セキュリティを強化するため、vSAN iSCSI ターゲット サービスは、CHAP（チャレンジ ハンドシェイク認証プロトコル）と相互 CHAP 認証を使用します。

vSAN は、一意の iSCSI 修飾名 (IQN) で各 iSCSI ターゲットを識別します。iSCSI ターゲットは、IQN を使用してリモート iSCSI イニシエータに提供されるため、イニシエータはターゲットの LUN にアクセスできます。vSAN iSCSI ターゲット サービスにより、iSCSI イニシエータ グループを作成できます。iSCSI イニシエータ グループは、グループのメンバーであるイニシエータのみにアクセスを制限します。

次のトピックを参照してください。

- vSAN iSCSI ネットワークの特性

vSAN iSCSI ネットワークの特性

vSAN iSCSI ネットワークには次のような特性があります。

- iSCSI ルーティング：iSCSI イニシエータは、L3 ネットワーク経由でルーティングされた接続を vSAN iSCSI ターゲットと確立します。
- IPv4 および IPv6：vSAN iSCSI ネットワークは IPv4 と IPv6 の両方をサポートします。
- IP セキュリティ：vSAN iSCSI ネットワークの IPSec により、セキュリティが強化されます。

  注： ESXi ホストは、IPv6 のみを使用した IPsec をサポートします。

- ジャンボ フレーム：vSAN iSCSI ネットワークではジャンボ フレームがサポートされます。
- NIC チーミング：vSAN iSCSI ネットワークでは、すべての NIC チーミング構成がサポートされます。
- Multiple Connections per Session (MCS)：vSAN iSCSI 実装は MCS をサポートしていません。

標準スイッチから Distributed vSwitch への移行

16

vSphere Standard Switch から vSphere Distributed Switch に移行し、Network I/O Control を使用できます。これにより vSAN トラフィックに QoS (Quality of Service) の優先順位を設定できます。

注意: ESXi ホストへのアクセスをお勧めしますが、必要でない場合もあります。問題が発生した場合は、ESXi ホストのコンソールにアクセスできます。

元の vSwitch 設定をメモしておきます。特に、ソースのロード バランシングと NIC チーミングの設定は必ず記録してください。ターゲットの構成がソースと一致していることを確認します。

Distributed Switch の作成

Distributed vSwitch を作成し、名前を付けます。

1 [vSphere Client ホストおよびクラスタ] ビューで、データセンターを右クリックし、[新しい Distributed Switch] メニューを選択します。

2 名前を入力します。

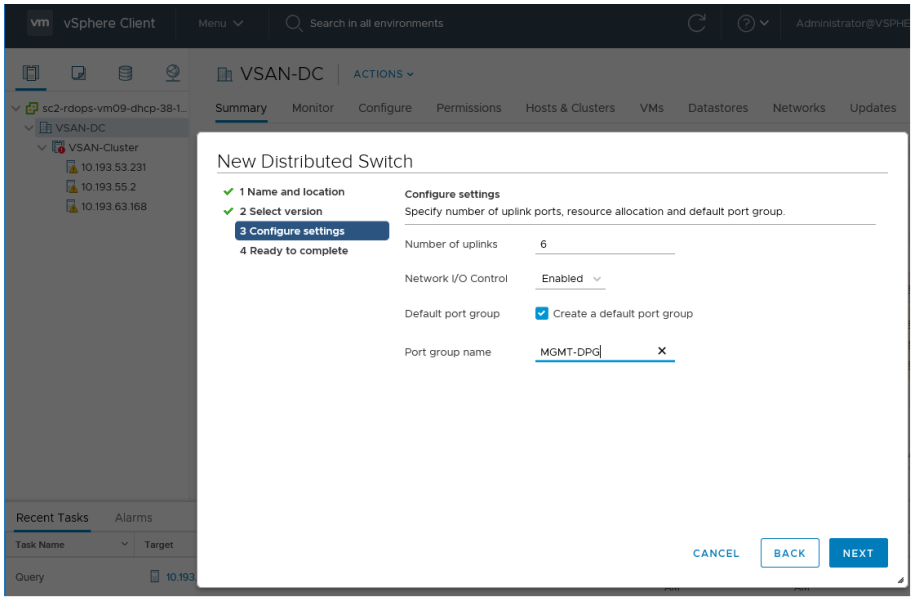
3 vSphere Distributed Switch のバージョンを選択します。この例では、移行にバージョン 6.6.0 が使用されています。

4 設定を追加します。現在ネットワークで使用しているアップリンクの数を確認します。この例では、管理、vMotion、仮想マシンに 6 個、vSAN（LAG 構成）に 3 個使用されています。アップリンクの数として 6 と入力します。実際の環境はこれと異なる場合があります。この値は後で編集できます。

   この時点で、デフォルトのポート グループを作成できますが、追加のポート グループが必要になります。

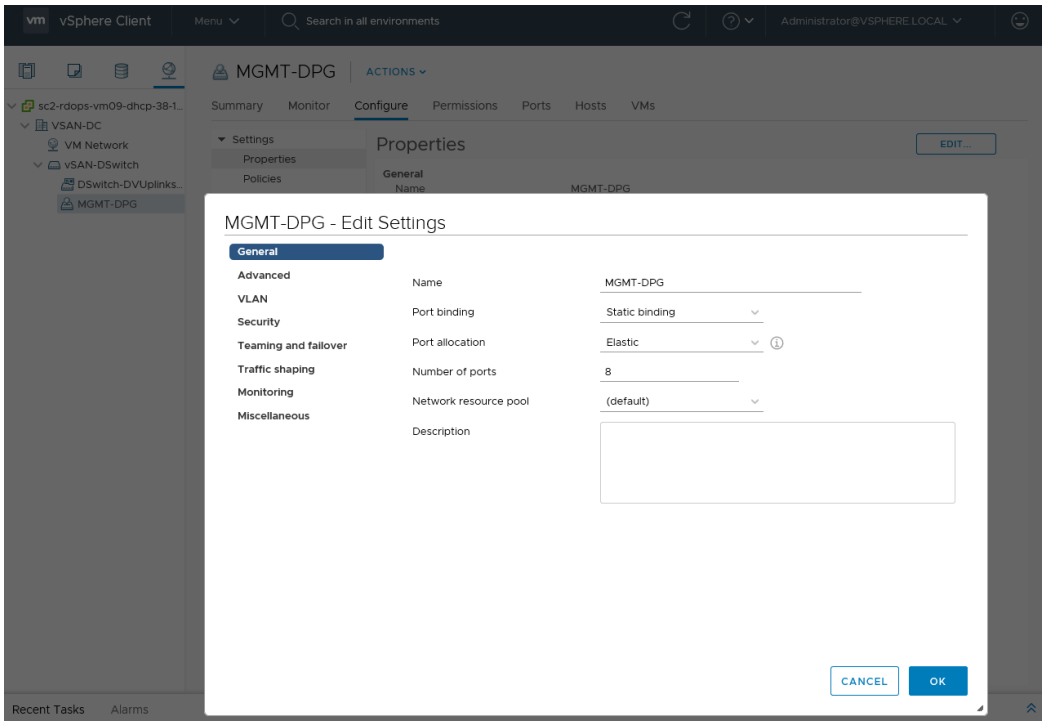
5 Distributed vSwitch の設定を完了します。

次に、追加のポート グループを設定して作成します。



ポート グループの作成

管理ネットワークに 1 つのデフォルト ポート グループが作成されました。このポート グループを編集して、VLAN、NIC チーミング、フェイルオーバー設定など、標準 vSwitch の管理ポート グループのすべての特性を設定します。



管理ポート グループを設定します。

1 vSphere Client の [ネットワーク] ビューで、分散ポート グループを選択して [編集] をクリックします。

2 ポート グループによっては、VLAN の変更が必要になります。VLAN 51 が管理 VLAN であるため、それに応じて分散ポート グループにタグを付けます。

3 [OK] をクリックします。

vMotion、仮想マシン ネットワーク、vSAN ネットワークの分散ポート グループを作成します。

1 vSphere Distributed Switch を右クリックして、[分散ポート グループ] > [新しい分散ポート グループ] を選択します。

2 この例では、vMotion ネットワークにポート グループを作成します。

Distributed vSwitch で分散ポート グループをすべて作成します。次に、アップリンク、VMkernal ネットワーク、仮想マシン ネットワークを Distributed vSwitch と関連する分散ポート グループに移行します。

注意: 作業をスムーズかつ慎重に行うため、アップリンクとネットワークを段階的に移行します。

管理ネットワークの移行

管理ネットワーク (vmk0) および関連するアップリンク (vmnic0) を標準スイッチから Distributed vSwitch (vDS) に移行します。

1 ホストを vDS に追加します。

   a vDS を右クリックし、[ホストの追加と管理] メニューを選択します。

   b ホストを vDS に追加します。緑色の追加アイコン (+) をクリックし、クラスタからすべてのホストを追加します。

2 物理アダプタと VMkernel アダプタを構成します。

   a [物理アダプタの管理] をクリックして、物理アダプタと VMkernel アダプタ (vmnic0 と vmk0) を vDS に移行します。

   b vDS で物理アダプタ vmnic0 に適切なアップリンクを選択します。この例では、Uplink1 を使用します。物理アダプタが選択され、アップリンクが選択されます。

3 vmk0 上の管理ネットワークを標準 vSwitch から Distributed vSwitch に移行します。各ホストで次の手順を実行します。

   a vmk0 を選択し、[ポート グループの割り当て] をクリックします。

   b 前に管理ネットワーク用に作成した分散ポート グループを割り当てます。

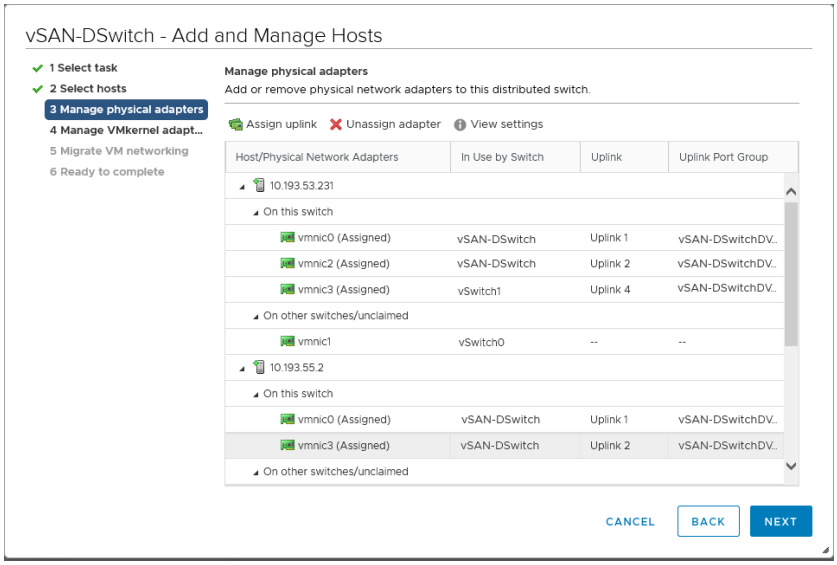
4 構成を完了します。

   a 変更を確認します。4 つのアップリンク（各ホストの vmnic0）と 4 つの VMkernel アダプタ（各ホストの vmk0）が追加されていることを確認します。

   b [終了] をクリックします。

各ホストのネットワーク構成については、スイッチの設定を確認します。各ホストに 1 つのアップリンク (vmnic0) と vmk0 管理ポートがあることを確認します。

このプロセスを他のネットワークに繰り返します。



vMotion の移行

vMotion ネットワークを移行するには、管理ネットワークの場合と同じ手順を実行します。

開始する前に、vMotion ネットワークの分散ポート グループの属性が標準 vSwitch のポート グループと同じであることを確認します。次に、vMotion (vmnic1) に使用されているアップリンクを VMkernel アダプタ (vmk1) と一緒に移行します。

vSAN ネットワークの移行

vSAN ネットワークにアップリンクが 1 つしかない場合は、前と同じプロセスを使用します。ただし、複数のアップリンクを使用している場合は、追加の手順を行います。

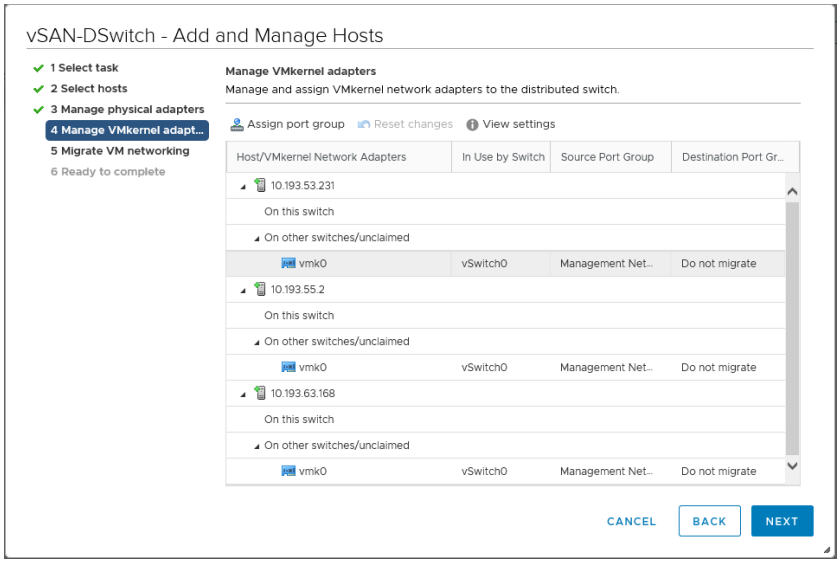
vSAN ネットワークがリンク集約 (LACP) を使用している場合、または他の VMkernel ネットワークと異なる VLAN にある場合は、特定の VMkernel アダプタで一部のアップリンクを未使用状態にします。

たとえば、VMkernel アダプタ vmk2 は vSAN に使用されています。ただし、アップリンク vmnic3、4、5 は vSAN に使用され、LACP を構成しています。したがって、vmk2 の場合、他のすべての vmnic（0、1、2）は未使用状態にする必要があります。同様に、管理アダプタ (vmk0) と vMotion アダプタ (vmk0) の場合は、vSAN アップリンク/vmnic を未使用状態にします。

分散ポート グループの設定を変更し、パス ポリシーとフェイルオーバーの設定を変更します。[物理ネットワーク アダプタの管理] ページで、複数のアダプタに操作を行います。

vSAN の分散ポート グループに vSAN VMkernel アダプタ (vmk2) を割り当てます。

注： vSAN ネットワークのアップリンクを移行中の場合は、移行が完了するまで分散ポート グループの設定を変更できないことがあります。この間、vSAN で通信の問題が発生する可能性があります。移行後、分散ポート グループの設定に移動し、すべてのポリシー変更を行い、すべてのアップリンクを未使用としてマークします。このタスクを終了すると、vSAN ネットワークが通常の状態に戻ります。vSAN 健全性サービスで、すべてが正常に機能していることを確認します。



仮想マシン ネットワークの移行

標準 vSwitch から Distributed vSwitch にネットワークを移行するための最後のタスクは、仮想マシン ネットワークの移行です。

ホスト ネットワークを管理します。

1 vDS を右クリックし、[ホストの追加と管理] メニューを選択します。

2 クラスタ内のすべてのホストを選択し、すべてのホストの仮想マシン ネットワークを Distributed vSwitch に移行します。

アップリンクは移行しないでください。ただし、ホストの仮想マシン ネットワークが別のアップリンクを使用していた場合は、標準 vSwitch からアップリンクを移行します。

3 標準 vSwitch の仮想マシン ネットワークから Distributed vSwitch の仮想マシン分散ポート グループに移行する仮想マシンを選択します。[ポート グループの割り当て] をクリックし、分散ポート グループを選択します。

4 変更内容を確認して、[終了] をクリックします。この例では、仮想マシンに移動します。元の標準 vSwitch の仮想マシン ネットワークを使用しているテンプレートは、仮想マシンに変換し、編集する必要があります。仮想マシンの新しい分散ポート グループをネットワークとして選択する必要があります。この手順を移行ウィザードで行うことはできません。

標準 vSwitch にアップリンクまたはポート グループがなくなったため、vSwitch を安全に削除できます。

これで、vSphere Standard Switch から vSphere Distributed Switch への移行は完了です。

vSAN ネットワークのチェックリストのサマリ

17

チェックリストのサマリで、vSAN ネットワークの要件を確認してください。

- 共有の 10Gb NIC を使用するのか、専用の 1Gb NIC を使用するかを確認します。オールフラッシュ クラスタには、10 Gb の NIC が必要です。
- 冗長な NIC チーミング接続が構成されていることを確認します。
- ESXi ホストの NIC でフロー制御が有効になっているかどうか確認します。
- 各ホストで vSAN ネットワーク トラフィック用の VMkernel ポートが構成されていることを確認します。
- すべてのインターフェイスで同一の VLAN、MTU、サブネットが設定されていることを確認します。
- すべてのホスト間で **`vmkping`** が正常に実行されることを確認します。健全性サービスで検証を行います。
- ジャンボ フレームを使用する場合は、すべてのホスト間で **`vmkping`** が正常に実行され、パケット サイズ 9,000 が返されることを確認します。健全性サービスで検証を行います。
- vSAN のバージョンが v6.6 よりも前の場合は、ネットワークでマルチキャストが有効かどうかを確認します。
- vSAN のバージョンが v6.6 よりも前で、同じネットワーク上に複数の vSAN クラスタがある場合は、一意のマルチキャスト アドレスを使用するようにマルチキャストを構成します。
- vSAN のバージョンが v6.6 よりも前で、複数のスイッチにわたっている場合は、スイッチ間でマルチキャストが構成されているかどうかを確認します。
- vSAN のバージョンが v6.6 よりも前で、ルーティングされている場合は、マルチキャスト ルーティングを許可するように PIM が構成されているかどうかを確認します。
- 物理スイッチが vSAN の要件（マルチキャスト、フロー制御、機能の相互運用性）を満たしていることを確認します。
- 大量のパケット ドロップや一時停止フレームなど、ネットワークのパフォーマンスに問題がないことを確認します。
- ネットワーク制限が許容可能な範囲内にあることを確認します。
- **`iperf`** を使用して vSAN ネットワークのパフォーマンスをテストし、期待値を満たしていることを確認します。