データセンターでの N ポート バーチャライゼーション

概要

N ポート バーチャライゼーションは、データセンターにおいて重要性が高まりつつある機能です。このドキュメントでは、拡大するストレージ ニーズに対応可能な仮想環境を設計するための設計要件を扱います。3 つの展開シナリオ（ブレード サーバ展開、トップオブラック ファブリック設計、および VMware）と、これらの状況での N ポート バーチャライゼーションの利用について説明します。

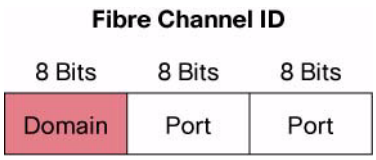
課題

データセンター設計における新たなトレンドは、サーバの仮想化です。これは、仮想マシン テクノロジーを使用して物理サーバの増加を抑えようとするものです。それぞれの仮想サーバは、ストレージ エリア ネットワーク上で一意なエンティティとして管理されるために、ファブリック上の個別のアドレスを必要とします。N ポート バーチャライゼーション機能は、仮想マシンを独立して管理するというこのニーズをサポートし、データセンターで使用される集約デバイスの増加に対応します。

イントロダクション

顧客のストレージ エリア ネットワーク（SAN）環境でブレード センター展開やトップオブラック集約デバイスの使用が増えるのに伴い、集約スイッチの展開と使用はますます拡大しつつあります。ファイバ チャネル テクノロジーの性質から、多数のエッジ スイッチを展開する際にはいくつかの懸案事項に対処する必要があります。ファイバ チャネル ベースの SAN を設計および構築する際の大きな懸案事項は、物理ファブリック内に存在可能なスイッチまたはドメインの総数です。エッジ スイッチの数が増えると、ドメイン ID の数が問題になります。ドメインは物理スイッチまたは論理仮想ファブリックのアドレスです。ドメイン ID は、エンドポイント ファイバ チャネル ID の中で先頭の 1 バイトです（図 1）。

図 1



スイッチは、このファイバ チャネル ID を使用して、SAN ファブリック内で所定の送信元（イニシエータ）から任意の送信先（ターゲット）にフレームを転送します。この 1 バイトで最大 256 のアドレスを表現できます。一部のドメイン アドレスは既知のアドレスで使用されており、残りは将来の拡張用に予約されています。ファイバ チャネル標準では、合計 239 のポート アドレスを使用できますが、そのようなファブリック サイズは非現実的です。

設計に関するもう一つの懸案事項は、サードパーティ製スイッチとの相互運用性です。従来、SAN ファブリック ベンダーは、それぞれ独自の方法でファイバ チャネル アドレッシング標準を解釈してきました。さらに、一部のベンダーに固有の属性として使用されているスイッ

チ間接続（拡張ポート [E-Port] 接続）のために、異なるベンダーのスイッチどうしを接続することが困難であるため、顧客はファブリック内のコア ディレクタの種類に合ったエッジ スイッチ テクノロジーを実装していました。

より小型のフォームファクタ デバイスを使用して複数のホスト接続を集約する場合は、この複雑なシステムの管理が懸案事項になります。通常は、この複雑さのために、プラットフォーム エンジニアリングまたはサーバ運用とファブリック運用との間で責任が共有されることになります。管理責任の線引きはあいまいです。

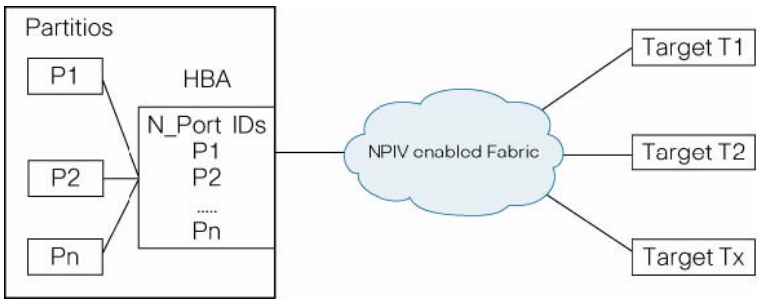
これらの懸案事項に対処するために、N ポート ID バーチャライゼーション（NPIV）および N ポート バーチャライザという 2 つの機能が開発されました。

N ポート ID バーチャライゼーション

NPIV を使用すると、1 つのリンク上のファイバ チャネル ホスト接続つまり N ポートに対して、複数の N ポート ID つまりファイバ チャネル ID（FCID）を割り当てることができます。割り当てられたすべての FCID は、ファイバ チャネル ファブリック上では同じ物理ホスト上の一意なエンティティとして管理できます。複数のアプリケーションを NPIV と組み合わせて使用することができます。多数のホスト オペレーティング システムやアプリケーションが物理ホスト上で稼働している仮想マシン環境では、それぞれの仮想マシンをゾーニング、エイリアス、およびセキュリティ上の観点から独立して管理できます。

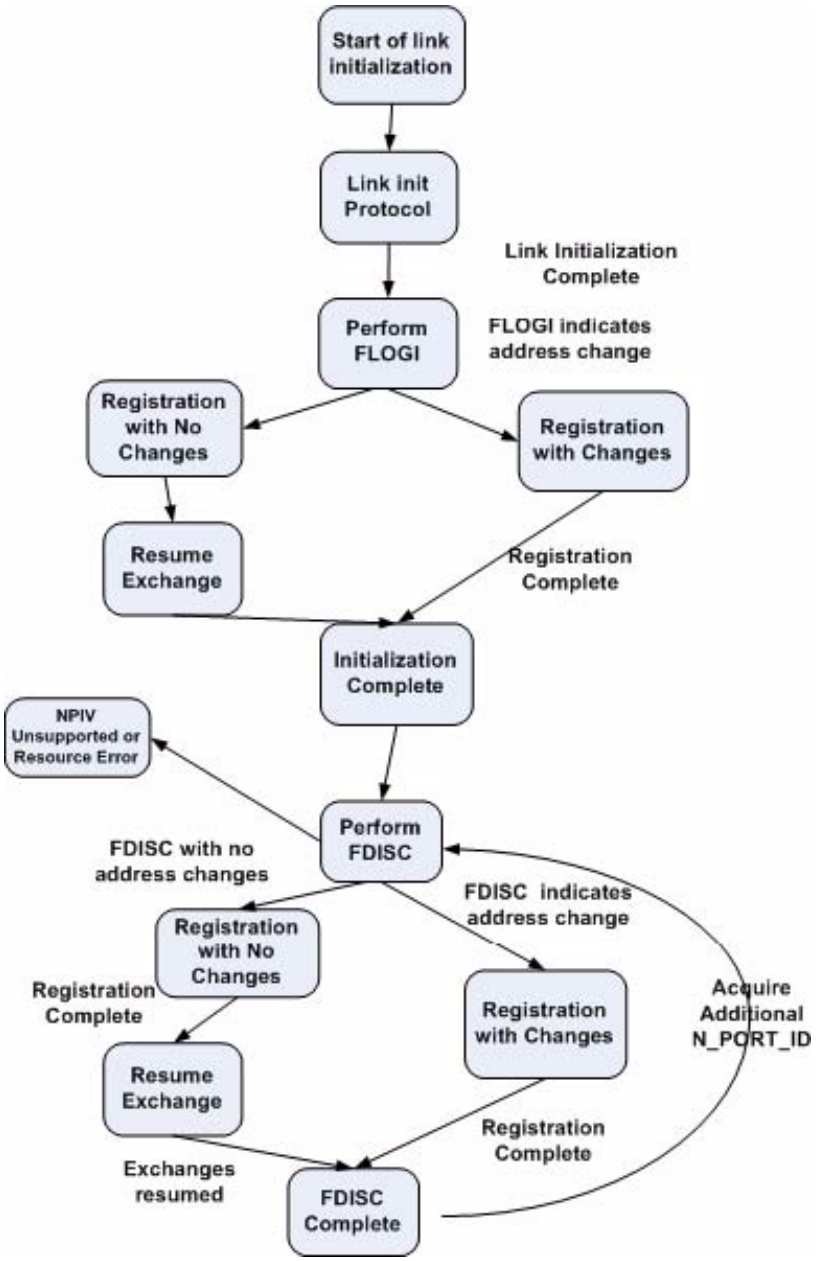
図 2 は、NPIV に対応したホスト接続の例です。

図 2　NPIV 対応型サーバ ホスト接続



NPIV 機能をサポートするホスト バス アダプタ（HBA）は、標準のログイン プロセスに従います。ファブリックへの最初の接続とログインは、標準の F ポート ログイン（FLOGI）プロセスを介して実行されます。仮想マシンまたはメインフレームの論理パーティションへの 2 回目以降のログインは、すべて FDISC ログイン コマンドに変換されます。FDISC ログインは、同じ標準のプロセスに従います。図 3 に、NPIV アップリンクのログイン プロセスと NPIV 対応アダプタへのローカル ログインの手順を 1 つずつ示します。

図 3　N ポート バーチャライゼーションのログイン プロセス

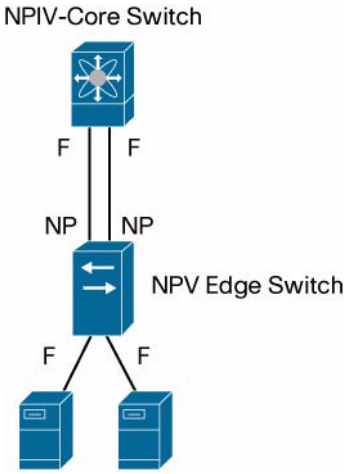


Cisco® MDS 9000 ファミリ環境では、それぞれのホスト接続を 1 つの仮想 SAN（VSAN）としてログインできます。VSAN は、管理ドメインからのファブリック サービス セットを共有する物理ポートをまとめた大きなグループの論理パーティションです。VSAN アーキテクチャは、イーサネット ネットワークにおける VLAN 展開と似ています。この環境で NPIV を使用する場合、その後の各 FDISC ログインは、元のファブリック ログインと同じ VSAN に属することになります。

N ポート バーチャライザ

NPIV の拡張の 1 つに、N ポート バーチャライザ機能があります。N ポート バーチャライザ機能を使用すると、ブレード スイッチまたはトップオブラック ファブリック デバイスは、コア ファイバ チャネル ディレクタに対して NPIV ベースの HBA として動作することができます。このデバイスは、ローカル接続されたホスト ポートつまり N ポートを、コア スイッチに対する 1 つ以上のアップリンク（疑似スイッチ間リンク）に集約します。

図 4　NPV によってエッジ スイッチはコア スイッチに対して HBA として動作できるようになる



N ポート アップリンク（NP）のログイン プロセスは、NPIV に対応した HBA と同じです。この場合のコア ディレクタの唯一の要件は、NPIV 機能をサポートすることです。エンド デバイスが NPV 対応エッジ スイッチにログインすると、FCID アドレスが割り当てられますが、このアドレスにはコア ディレクタのドメインが使用されます。この接続は N ポートとして扱われ、コア ディレクタへの E ポートとしては扱われないので、エッジ スイッチは、FCID が割り当てられているコア スイッチのドメイン ID を共有します。NPV 対応エッジ スイッチは、ファブリックとの接続を受信するための独自のドメイン ID を持つ必要がなくなりました。トップオブラックまたはブレード センター テクノロジーを使用して SAN を設計する場合にドメインに関して発生する副産物は、NPV を使用することで解消されます。

アップリンク ポートが N ポート バーチャライザ デバイス上で有効になっている場合は、最初にファブリック ログインを実行し、コア ディレクタへのログインを処理して、ファイバ チャネル ネーム サーバに登録する必要があります（図 3）。物理アップリンクでのログイン プロセスが完了したら、エンド デバイスは固有のログイン プロセスを開始できます。Cisco NPV プロセスでは、エンド デバイスが固有のログイン プロセスを開始するためには、まずアップリンク ポートがコア ディレクタへのログインを実行する必要があります。エンド デバイスがエッジ NPV スイッチへのログイン プロセスを開始する場合は、FLOGI による NPV エッジ スイッチへの標準ログイン プロセスの後に行います。その後、エッジ スイッチによって FLOGI ログインが FDISC ログインに変換され、NPV アップリンク経由でログインを進めることができます。複数の NPV アップリンクを 1 つ以上のコア ディレクタに接続できます。現在、シスコの実装では、各エッジ スイッチから複数のコア ディレクタへの接続が可能です。コア NPIV ディレクタは、同じファブリックに含まれている必要があります。

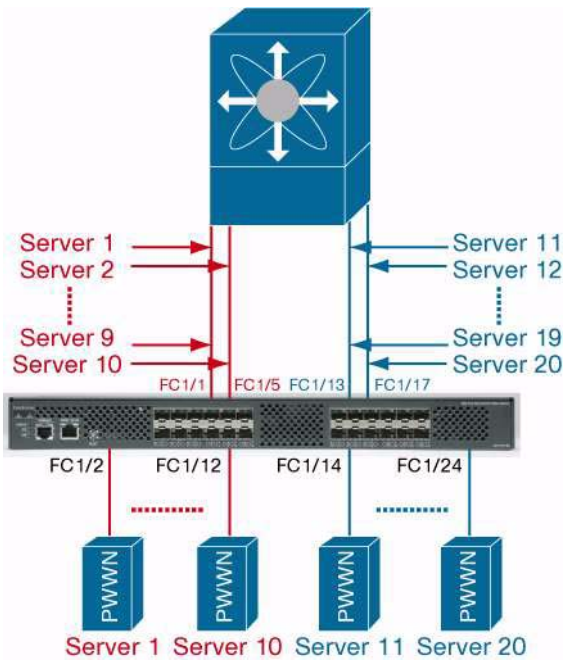
NPV 対応デバイスはプロキシ ログインを実行すると同時に、複数の N ポート アップリンクを介して FDISC ログインをコア NPIV ディレクタに送信するメカニズムも備えています。ログインを複数のリンクに送信するには、手動セットアップ、動的なロード バランシング ア

ルゴリズム、および VSAN ベースのロード バランシングという 3 つの方法があります。現在、シスコの実装は、動的なロード バランシングと VSAN ベースのロード バランシングを使用しています。

動的なロード バランシングでは（すべてのアップリンクは同じ VSAN 内にある）、ホストが NPV デバイスにログインすると、FDISC ログインはラウンドロビン方式でリンクを介して送信されます。エンド デバイスからの I/O は、常に、FDISC ログインの送信先と同じコア スイッチまでの NPV パスが使用されます。安定したファブリックでは、それぞれのリンクが均等に利用されることになります。たとえば、ホストが 8 台、コア スイッチへのアップリンクが 2 本ある場合、各リンクは 4 つのイニシエータを持つことになります。1 本のアップリンクに障害が発生すると、ログインを受けてこのアップリンクをコア スイッチへのアップリンクとして使用しているホストは、ファブリックからログアウトされ、残っているリンクから FDISC ログイン手順を実行することになります。N ポート バーチャライザ機能の初期リリースでは、ホストを動的に元のリンクに戻す優先アップリンク設定はありません。回復したパスをホストが使用するには、ホストを再初期化する必要があります。ホストの優先アップリンクへの手動割り当ては、後日リリースされるファームウェアで実装される予定です。

NPV エッジ デバイスは、複数の VSAN 内のコア Cisco MDS 9000 ファミリ スイッチに接続できます。各 N ポートはそれぞれの VSAN のファイバ チャネル ネーム サーバに対して、ログイン手順を実行します。スイッチ上のホスト ポートは、そのアップリンクを含む VSAN に合わせて設定されます。すべてのホストの FDISC ログインは、アップリンク ポートの適合する VSAN タグに従って送信されます。同じ VSAN 内に複数のアップリンクが存在する場合、FDISC は VSAN ID と一致するリンクにラウンドロビン方式で分散されます。

図 5　NPV によってコア スイッチとの透過的な接続が可能になる



N ポート バーチャライザ機能を使用すると、NPIV 機能をサポートする任意のコア スイッチと透過的に接続できます。そのコア スイッチが標準の NPIV 実装に準拠している場合、スイッチ ベンダー間の相互運用性を考慮する必要はなくなります。

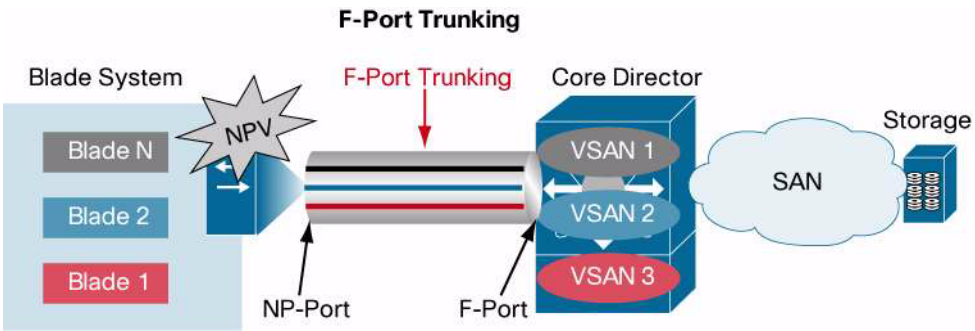
将来における拡張

近い将来、NPV をより詳細に管理するための 2 つの機能が利用できるようになります。ファブリック ポート（F ポート）トランキングと F ポート チャネリングです。PortChannel は、複数の物理インターフェイスを 1 つの論理高帯域幅リンクにまとめる機能です。PortChannel を使用すると、高帯域幅を利用できるようになり、リンクの冗長性が増し、2 つのスイッチ間でのロード バランシングが可能になります。シスコの実装では、スイッチ上の場所に関係なく任意のモジュール上のポートを PortChannel に含めることができます。PortChannel インターフェイスの 1 つのリンクに障害が発生しても、ルート テーブルの変更は発生しません。

シスコの仮想ファブリック実装では、1 つの F ポートが属することができる VSAN は常に 1 つだけです。ただし、トランク モードに設定されているスイッチ間リンク（ISL）は、1 つの ISL または PortChannel で複数の VLAN を伝送することができます。その場合も VSAN 間のファブリックの分離状態は維持されます。エッジ デバイスを NPV モードで設定する場合の欠点の 1 つは、F ポート アップリンクを単一の VSAN にしか含めることができず、リンクで単一の VSAN しか伝送できないことです。これは、スイッチ間のリンクが F ポートとしてファブリックにログインするように設定されているためです。NPV デバイスは、複数のアップリンクに対して設定して、それぞれを異なる VSAN に配置できます。この NPV デバイス上のホストは、それらの VSAN のいずれかに配置できるので、適切な NP アップリンク ポートを介してログインすることになります。

F ポート トランキングを使用すると、単一の F ポートを複数の VSAN に含めることが可能です。F ポート トランキングは、シスコが現在 ISL で実行しているトランキング プロトコルに準拠します。この機能により、VSAN 接続を NP デバイスにまで拡張するために必要なアップリンク ポートを統合できます。図 6 に、F ポート トランキングによる接続と、ネットワークの概要を示します。

図 6　F ポート トランキング

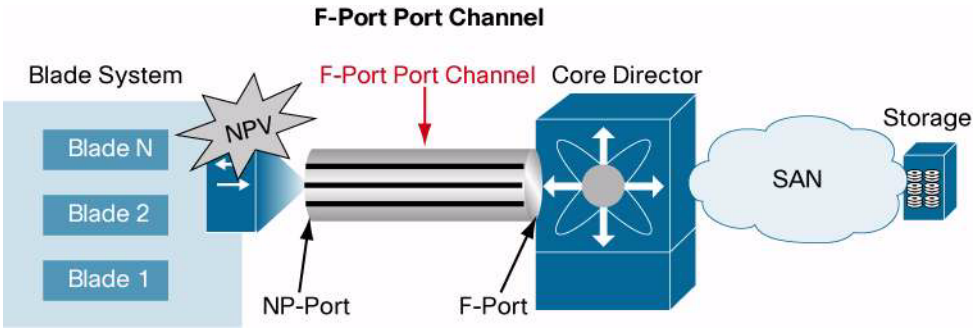


F ポート トランキングを有効にすると、VSAN がトランクされた任意の NP アップリンクを介して、すべてのブレード サーバとの接続ができるようになります。HBA ベンダーがそのファームウェア内で VSAN トランキングを有効にすると、F ポート トランキングを HBA 接続にまで拡張できます。F ポート トランキングを使用すると、物理サーバ上の仮想マシンがアプリケーションや業務ごとに別の VSAN にログインし、参加できるようになります。

近く登場するもう 1 つの重要な機能は、F ポート チャネリングです（図 7）。前述したように、ホストがローカル スイッチにログインすると、FDISC メッセージはラウンド ロビン方式でロード バランシングされます。ホストがセッションを実行しているリンクに障害が発生

した場合、ホストはファブリックにログインし直して、ログイン プロセスを実行し直す必要があります。NPV デバイスからの N ポート アップリンクは、ハイアベイラビリティが適用されるようになっていません。

図 7　F ポート チャネリング



F ポート チャネリングを使用すると、現在の ISL PortChannel と同様の復元力とアベイラビリティがスイッチ間で実現されるようになります。この設定では、複数の NP アップリンクを束ねて（チャネリング）、単一の論理リンクを形成することができます。チャネルの 1 つのリンクに障害が発生した場合、そのリンクを通過中のデータ フレームは失われ、一般的なアプリケーションやネットワーク エラー回復処理が実行されます。

F ポート チャネリングによって解決されるもう 1つの問題も、リンク障害に関係しています。F ポート チャネリングを設定すると、最初に NPV エッジ デバイスからコアへとログイン プロセスを伝送したリンクに障害が発生しても、ホストはファブリックへのログインを最初から実行し直す必要はなくなります。リンクに障害が発生しても、ログイン状態はそのまま残ります。チャネル内のリンクが 1 つでも稼働している限り、ホストはログイン プロセスを実行しません。

F ポート チャネリングの論理リンクでは、チャネル内の 1 つのリンクに障害が発生しても、ホストはファブリックに接続するためにログインし直す必要はありません。ホストはログイン状態が維持されます。障害が発生したリンク上のデータ フローは回復モードを実行する必要がありますが、ホストはネットワークにログインしたままになっています。データ トラフィックは束ねられたアップリンクを介して送信されますが、その際も、src/dst ハッシュや src/dst/oxid ハッシュを使用してロード バランシングが行われます。この機能により、障害発生中や回復中も NP アップリンク間のロード バランシングを手動で行う必要がなくなります。

まとめ

顧客がデータセンター アーキテクチャを変更してデータセンター リソースの統合をさらに進めようとする中で、ネットワークベースのテクノロジーは、ファブリック内のデバイス数を設計および管理できるように進化し続ける必要があります。N ポート バーチャライゼーションは、データセンターにおいて重要性が高まりつつある機能です。このドキュメントでは、ブレード サーバ展開、トップオブラック ファブリック設計、およびVMware という 3 つの展開シナリオについて説明してきました。これらすべてのシナリオで、NPIV と NPV がさまざまな方法で利用されます。シスコは、顧客の成長と進化に応じたファブリックの設計、展開、および管理に役立つ NPIV の拡張機能を継続的に開発しています。

関連情報

Cisco MDS 9000 ファミリ製品の詳細については、http://www.cisco.com/jp/go/storage/ を参照してください。

ソリューションの概要については、http://www.cisco.com/jp/go/dc/を参照してください。

シスコシステムズ合同会社
〒 107-6227 東京都港区赤坂 9-7-1 ミッドタウン・タワー
http://www.cisco.com/jp
お問い合わせ先：シスコ コンタクトセンター
0120-092-255（フリーコール、携帯・PHS 含む）
電話受付時間：平日 10:00 ～ 12:00、13:00 ～ 17:00
http://www.cisco.com/jp/go/contactcenter/

お問い合せ先

08.12