



DATA CENTER

イーサネット・ファブリックとは？

イーサネット・ファブリックは、従来の階層型アーキテクチャを凌ぐ、高水準の性能、リソース利用率、可用性、簡易性を実現します。これにより、データセンターは、現在そして将来のビジネス要件に対応することができます。

データセンターのネットワークは、イーサネットに依存しています。イーサネットは、数十年にわたって、新しいタイプのアプリケーションが出現するごとに進化を続けています。今日、データセンター・ネットワークは、クライアント/サーバ、Web サービス、ユニファイド・コミュニケーション、仮想マシン、ストレージなど、様々なアプリケーション群のトラフィックが転送され、それぞれに異なるトラフィック・パターンとネットワークサービス上の要件があります。また、サーバ・クラスタにホストされる仮想マシン上で展開されるアプリケーションの数も増加しています。共有ストレージプールの構築にもイーサネットが使用され、ロスレスの packets 伝送や、あらかじめ予測可能なレイテンシ、高帯域幅など、ネットワークに過酷な条件が要求されます。こうした変化を背景として、進化したイーサネットである“イーサネット・ファブリック”が必要とされています。

はじめに

データセンターは、デジタル資産の増加とアプリケーションのさらなる導入とともに成長を続けています。市場も競合相手もグローバル規模になるにつれて、企業の競争力維持のために月単位でなく分刻みの迅速なアプリケーション導入が期待されています。また、ラック・スペース、電力、冷却といったデータセンター・リソースはますます欠乏し、かつ高価になっています。ネットワーク、サーバ、ストレージの仮想化、そして IT オークストレーション・ツールは、高密度なマルチ・コア・サーバと効率的に組み合わせて展開することにより、IT リソースをプールするクラウド・アーキテクチャを実現することができます。クラウド・コンピューティングへの移行は、アプリケーションの統合を進めてリソース利用率を向上し、投資/運用の費用を削減するとともに、ビジネスのニーズに合わせて迅速にスケールアップすることのできるインフラを構築します。

サーバ仮想化によって実現される投資額の節減は、ビジネスに必須の“Do More With Less”の要件に応えるものです。しかし、従来の仮想化環境ツールやシステム管理ツールにある限界や、現在のネットワーク・アーキテクチャが持つ潜在的な制約のために、クラウド・コンピューティングが要求する性能、可用性、セキュリティ、移動性の要件に対応できないことがあります。システム管理ツールやオークストレーション・ツールは複雑なものが多く、詳細な設定が要求されることから、導入に費用がかかるうえ、効率的な運用が困難になるケースもあります。管理レイヤを簡素化して、仮想化の将来性を確実なものにしていくには、基本的なインフラ、特にネットワークを進化させなければなりません。例えば、物理ポート単位の管理ではなく、仮想サーバから仮想サーバへの通信、あるいは仮想サーバから仮想ストレージへの通信といった、フローを単位とする管理への移行が必要です。また、セットアップ、運用、拡張がシンプルで、柔軟性が高く、障害に強く、さらに VM に円滑に対応できるインフラが要求されます。ガートナー社は、「現在見られる変化は、ネットワーク・アーキテクチャを従来の階層ツリー型トポロジから、フラット・メッシュ型レイヤ 2 ネットワークのトポロジに進化させるものになる。」¹と述べています。これは、イーサネット・ファブリックを言いかえたものです。

本書では、従来のイーサネット・アーキテクチャを新しいデータセンターの要件という観点から見直し、従来型イーサネットとイーサネット・ファブリックの相違を概観し、データセンターで生じる課題に対してイーサネット・ファブリックをどのように利用することができるのかを説明します。

¹ “Eight Key Impacts on Your Data Center LAN Network”
(データセンターLANネットワークに及ぶ8つの大きな衝撃) (Munch, 4/27/11, ID G00211994)

従来のイーサネット・ネットワーク

イーサネット・ファブリックをより理解するために、最初に従来のイーサネット・ネットワークについて考えてみます。データセンターでは、必要なポート数が1台のイーサネット・スイッチで利用できる量を超えているため、複数台のスイッチを接続してネットワークを形成し、接続性を向上しています。例えば、サーバ・ラックは、トップ・オブ・ラック (ToR) にスイッチが置かれる場合が多いですが、複数のラックに収容したサーバを、ミドル・オブ・ロー (MoR) またはエンド・オブ・ロー (EoR) のスイッチに接続することもあります。これらのイーサネット・スイッチは、すべて接続され、図1に示すような階層型の“イーサネット・ツリー”トポロジを形成します。

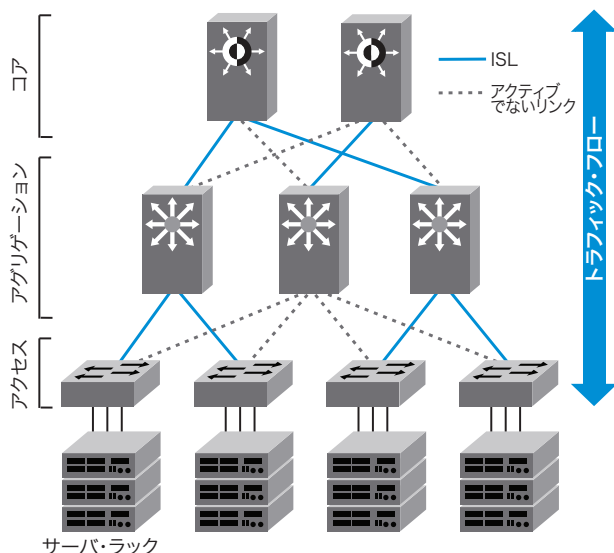


図1.
従来のイーサネット・ネットワーク

従来のイーサネットの制約

従来型のイーサネットでは、スイッチ間を接続するISL (Inter-Switch Link : 図1では青色の実線) がループを形成することは許可されず、ループがあるとフレームが転送されません。スパンニング・ツリー・プロトコル (STP) は、ツリー型トポロジを作成し、いずれの2台のスイッチ間も経路を1本だけアクティブにすることによりループを回避します (図1の点線で示したアクティブでない経路)。これにより、スイッチ間において複数の経路が禁止され、ISL帯域幅が単一の接続に限定されることとなります。この制約を解消するためにイーサネットが改良され、スイッチ間を接続する複数のリンクを1本の接続として扱ってループを回避する、LAG (Link Aggregation Group) が規定されています。しかし、LAGは、ポートごとに手作業でコンフィグレーションを行う必要があり、柔軟性もあまり高くありません。

ツリー型トポロジでは、別のラックへ向かうトラフィックは、ツリーを上下すなわち“南北”に移動する必要があります。アクセス・トラフィックの大半が同一ラック内のサーバ間にある場合は、問題になりません。しかし、クラスタ化アプリケーションやサーバ仮想化で要求されるようなサーバ・クラスタの場合、別のラックにあるサーバの間で“東西”に転送されるトラフィックがあるため、ツリー型トポロジではマルチホップのレイテンシが増大するとともに、スイッチ間が単一リンクであることで帯域幅が制限されます。

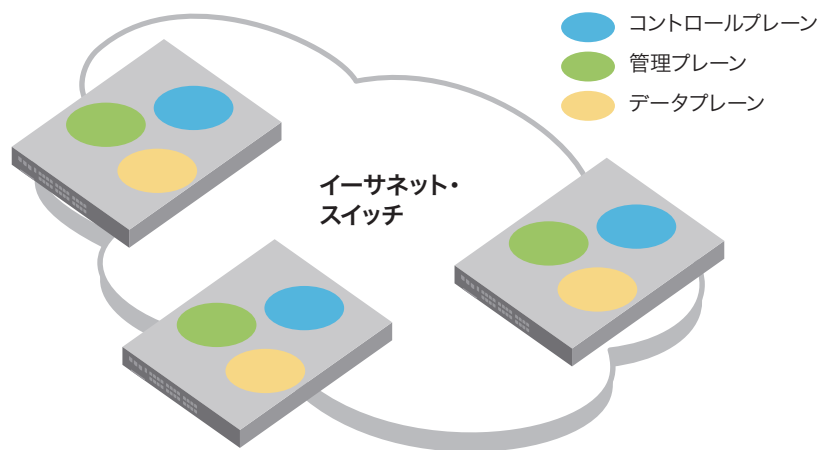
リンクに異常が発生した場合、STPは自動復旧します。ただし、STPは、ネットワークを通過するトラフィックをすべて中断し、ネットワークにあるすべてのスイッチ間で単一のパスを再構成しなければ、トラフィック・フローを再開することができません。数十秒から数分にわたって、全リンクのトラフィックをすべて停止するため、スケーラビリティが制限されるとともに、トラフィックは、リンク復元力を確保するためにデータバスをブロックすることが許容されるようなアプリケーションに限定されます。過去には、トラフィックはTCPに依存し、このようなサービス中断を処理していました。今日では、データセンターのほぼすべてのアプリケーションが、年中無休の高可用性モードで稼働しているほか、イーサネット・ネットワークを通過するストレージ・トラフィックが増加する中、たとえ数秒でもデータバスの接続が中断することは許されません。

従来のイーサネット・スイッチ・アーキテクチャには、ほかにも制約があります。スイッチにはそれぞれがコントロールプレーンと管理プレーンがあります。すべてのスイッチでは、入力ポートにフレームが着信する度に、プロトコルの検出と処理が実行されます。したがって、スイッチの台数を追加するにつれて、プロトコル処理の時間が増えることにより、レイテンシが増加します。また、スイッチ間で共通のコンフィグレーションやポリシーの情報を共有することがないため、各スイッチとスイッチ内のポートごとに個々に構成しなければなりません。デバイス数が増えるとますます複雑になり、コンフィグレーション作業のミスが増加し、運用・管理の総コストを計算することも難しくなります。

イーサネット・ファブリック・アーキテクチャ

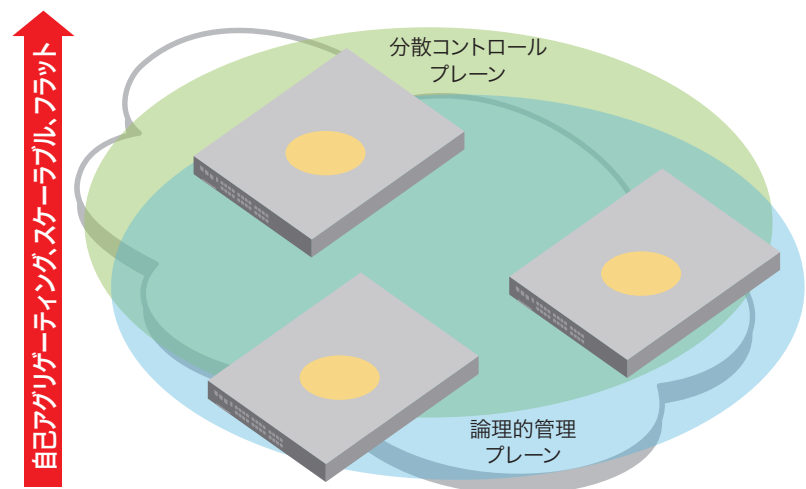
図2に、従来のイーサネット・スイッチのアーキテクチャを示しています。コントロールプレーンと、データプレーン、管理プレーンが、バックプレーンを介して全ポートに論理的に接続されています。コントロールプレーンと管理プレーンは、ネットワークのレベルではなく、スイッチレベルで運用します。

図2.
イーサネット・スイッチの
アーキテクチャ



イーサネット・ファブリックは、コントロールプレーンと管理プレーンとを、物理スイッチを超えてファブリックにまで拡張するものと考えることができます。図3に示すように、スイッチのレベルではなく、ファブリックレベルでコントロールプレーンと管理プレーンを運用するようになります。

図3.
イーサネット・ファブリックの
アーキテクチャ



イーサネット・ファブリックでは、コントロール・パスによって、STP はリンク・ステート・ルーティングと置き換えられる一方、データパスはレイヤ 2 の等価マルチパス転送を可能にし、データは複数の ISL 接続を使用しつつ、ループを起こさずに、つねに最短経路が選択されます。ファブリックのコントロールプレーンと組み合わせられ、帯域幅のスケーリングはシンプルになります。例えば、新しいスイッチをファブリックのほかのスイッチに接続する場合、自動的に新しいトランクが形成されます。トランクの中のリンクが中断した場合、または削除された場合には、既存のリンク間でトラフィックが再度負荷分散され、その際にも停止は起こりません。さらに、ファブリックのいずれかの場所で ISL が増設または削除された場合にも、ほかの ISL のトラフィックは、STP の場合と同様、停止することなく転送が継続されます。

このアーキテクチャでは、スイッチのグループを“論理シャーシ”の一部として定義することができ、スイッチをシャーシ型スイッチのポート・カードのように取り扱うことができます。論理シャーシの中では、ポリシーやセキュリティのコンフィグレーション・パラメータを全台のスイッチで簡単に共有することができるため、管理、モニター、運用が簡素化されます。さらに、ファブリックでは、物理/仮想サーバとストレージへの接続情報はすべてのスイッチで認識されているため、どの仮想マシンが移動しても、またどこに存在していても、ネットワーク・ポリシーとセキュリティ設定を適用し続けることができます。

イーサネット・ファブリックによるデータセンターの課題への対応

現在の仮想化データセンターの課題として、仮想サーバ環境のスケーリング、アプリケーション・モビリティの実現、インフラの複雑さと管理オーバーヘッドへの対応があります。

仮想サーバ環境のスケーリング

仮想サーバ環境をスケーリングしようとするとき、ネットワークにはスパンニング・ツリー・プロトコル (STP、図 2 を参照) の制約や、サーバあたりの GbE 接続数の増加、低い利用率、リンク障害の回復など、さまざまな課題があります。

仮想マシン (VM) モビリティなどの仮想化機能を使用する場合、VM は 1 つのレイヤ 2 ネットワークの内側でしか移動できず、レイヤ 3 プロトコルを使用する仮想 LAN (VLAN) をまたいだ VM の移動を無停止で実行することは、仮想化ハイパーバイザではサポートされていません。従来のレイヤ 2 イーサネット・ネットワークでは、ネットワークの可用性を向上するうえで、STP を使用してネットワーク内のパスをアクティブかスタンバイかに指定します。これによって代替パスが提供されますが、一度に 1 つのパスしか利用できないため、ネットワーク帯域幅は十分に活用されません。サーバ仮想化の目標の 1 つは、物理サーバの利用率向上にあるため、ネットワーク帯域幅の利用率向上も当然期待されています。

ネットワーク利用率を向上するためには、MSTP (Multiple Spanning Tree Protocol) や同種のプロトコルを使用し、VLAN ごとにスパンニング・ツリーを分離することができます。しかし、これにより帯域幅利用率は向上しますが、スイッチ間につきアクティブなパスが 1 つのみという STP の制約は解決されません。また、MSTP では手作業でトラフィック経路を構成するため、より複雑になるという問題もあります。

STP のもう 1 つの課題は、リンクに異常が発生した際のネットワークの動作です。障害が発生すると、スパンニング・ツリーの再定義が必要となります。これには RSTP (Rapid Spanning Tree) で最短 5 秒間、STP では最長数分間かかり、小さなトポロジ変化でも予測のつかない変動をもたらします。無停止のトラフィック・フローへの要求はサーバ仮想化とともに高まっていることから、ネットワークの収束時間は短縮しなければなりません。STP には、この要件に対応する優れた解決策は備えられていません。さらに、スパンニング・ツリーが再収束する際に、ブロードキャストストームが発生して、ネットワークをスローダウンさせる場合もあります。これらの STP の制約が原因となり、データセンターのレイヤ 2 ネットワークは、概して小規模なものとなっています。

イーサネット・ファブリック

イーサネット・ファブリックは、従来の階層型イーサネット・アーキテクチャと比較して、高水準の性能、利用率、可用性、シンプルさを実現します。イーサネット・ファブリックには、次のような特長があります。

フラット: イーサネット・ファブリックでは、スパンニング・ツリー・プロトコルの必要性がなくなりますが、従来のイーサネット・ネットワークと完全に相互運用が可能です。

柔軟性: あらゆるワークロードの要件に最適なトポロジを構成することができます。

耐障害性: 複数の“最小コスト”経路を使用して、高い性能と信頼性を実現します。

弾力性: ニーズに応じたスケーリングの拡張・縮小が容易にできます。

さらに、ファイバーチャネル・ファブリックから継承された、次のような高度な機能があります。

- 自己形成機能を備え、論理的に単一のエントリとして機能します。スイッチはすべて相互に認識され、接続されたすべての物理/論理デバイスを認識します。
- 管理は、デバイスごとではなく、ドメイン・ベースで行います。繰り返し作業を省いたポリシーによる設定が可能です。
- これらの機能は、仮想化に対応した拡張機能とともに、VM オートメーションの課題への対応、IT 自動化を容易にします。
- FCOE (Fibre Channel over Ethernet) などのプロトコル統合は、LAN トラフィックと SAN トラフィックのブリッジングに向けた機能として備えられています。

そこで要求されるのは、次のような特長を備えたレイヤ 2 ネットワークです。

- 可用性が高い
- コストが同等なバスを介して高い帯域幅利用率を保証
- ネットワーク障害や再構成時のリンクの追加や削除の際も、トラフィックを継続
- 遅延が確定的で、しかもロスレス
- IP トラフィックとミッションクリティカルなストレージ・トラフィックを同一ワイヤ上で伝送可能

これらは、いずれもイーサネット・ファブリックが備える機能です。これによって、VM モビリティの制約や潜在的なネットワーク・ダウンタイムを排除し、仮想サーバ環境を効率的にスケールアップすることができます。

アプリケーション・モビリティ

アプリケーションが物理サーバ上ではなく VM 上で稼働すると、アプリケーションは特定の物理サーバに拘束されなくなります。これにより、アプリケーションの要求が変化した場合、サーバにメンテナンスを実行する場合、あるいは拠点に災害があって迅速な回復が必要な場合などに、VM は物理サーバ間を移動することができます。

VM の移動は、同じ IP サブネットおよび VLAN の中にある物理サーバのクラスタ内で可能となります。これは、マイグレーションによってクライアントトラフィックを停止させないための条件です。というのは、IP サブネットの変更は、運用停止を伴うものであってはならないからです。STP による制約について先に述べたように、VM マイグレーションの領域は、他にも制約を受けることがあります。柔軟性の高い VM モビリティのための解決策は、ネットワーク帯域幅の利用率を向上し、よりスケールアップで可用性に優れたレイヤ 2 ネットワークです。

VM が 1 台のサーバから別のサーバに移動する場合、多数のサーバ属性が移動元と移動先のサーバで同一に設定されなければなりません。これはネットワークも同様で、VLAN、ACL (Access Control List)、QoS (Quality of Service)、セキュリティプロファイルは、移動元と移動先のアクセス・スイッチ・ポートの 2 つで同一である必要があります。スイッチポート構成に差異があると、マイグレーションの事前準備が失敗するか、VM のネットワークアクセスが途絶えることとなります(図 3 を参照)。すべての設定をすべてのネットワークポートにマップしておくことが考えられますが、それはネットワークとセキュリティにおける多くのベスト・プラクティスに反します。VMware vSphere 4 の分散仮想スイッチは、これらいくつかの問題に対処していますが、代償としてスイッチングのために物理サーバのリソースを消費することや、多層化したスイッチ構成によりネットワークポリシーの管理が複雑になること、VM 間トラフィックで一貫したセキュリティが実行できないことなどの問題があります。

自動化された VM マイグレーションの場合には、ネットワーク管理者はアプリケーションの所在を限定的にしか見ることができません。そのためトラブルシューティングが困難になり、特定の VM の中で問題の原因を正確に特定することはできません。

ここでまた、次のようなレイヤ 2 ネットワークを考えてみましょう。

- VM マイグレーションの物理的障壁を排除
- VM の所在を認識し、ネットワークポリシーを一貫して適用
- VM が移動した場合の手作業での再設定などの必要性を排除
- ハイパーバイザからスイッチング・トラフィックのオーバーヘッドを抑制し、最大限の効率性と機能性を実現
- 同一ネットワークで異機種混在のサーバ仮想化に対応

VM に最適なイーサネット・ファブリックの高度な機能によって、アプリケーション・モビリティの領域を拡大し、VM アウェアネスと、アプリケーションのためのサーバリソース最適化を実現することができます。

ネットワーク管理

今日のデータセンターのLANと同様に、マルチティア・アーキテクチャはきわめて複雑で(図4を参照)、その上管理者は、多くのレイヤ2、レイヤ3プロトコルに精通していなければなりません。そして、他のドメインとの境界を含めたネットワークの管理は一層複雑になります。アクセス・レイヤでも、単一のスイッチの管理ではなく、現在ではハイパーバイザ内のソフトウェアスイッチ(“ソフトスイッチ”と呼ばれる)から、トップ・オブ・ラックまたはエンド・オブ・ローのアクセススイッチまで広がる多段型のスイッチングまで含まれます。VMをホストするサーバ・ラックを増設するたびに、スイッチングレイヤごとに構成しなければならず、これによりコストと複雑さが増大しています。

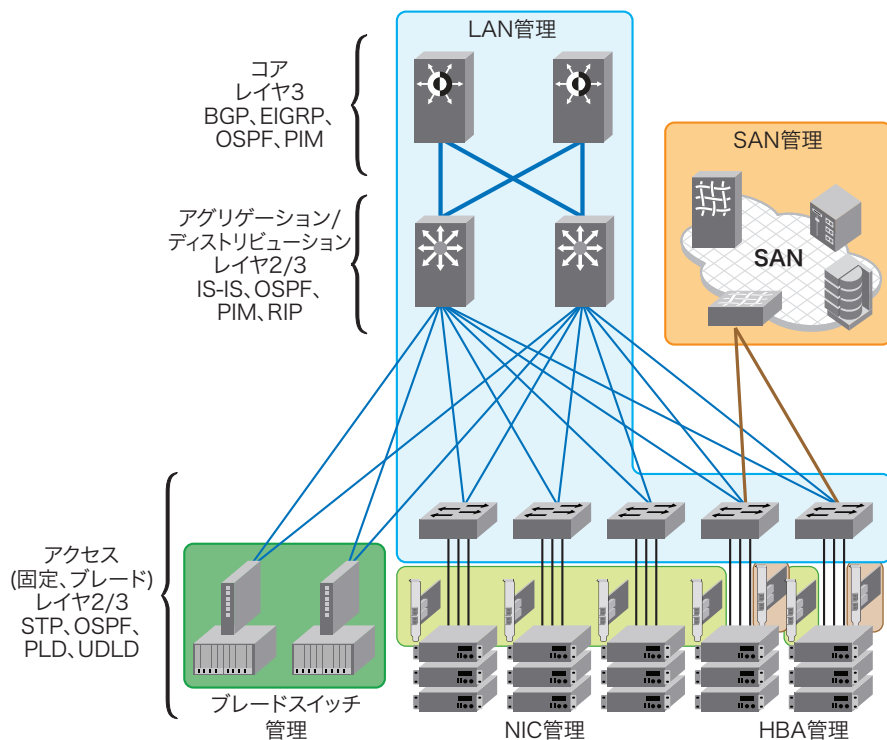


図4.

マルチティアのネットワーク・アーキテクチャと多数のレイヤ2/3プロトコルによって複雑性と管理コストが増大

管理をさらに複雑にするのは、それぞれの管理に個別のツールが使用されることです。LAN、SAN、ブレードサーバの接続、NIC (Network Interface Card)、HBA (Host Bus Adapter) のそれぞれに異なるツールが使用されます。管理者は、直接担当する範囲しか確認できない場合が多く、ネットワーク環境全体を把握することは困難です。この問題は、データセンター管理ツールやオーケストレーション・ツールのみでは解決できず、また仮想化だけに頼って解決することもできません。インフラ自体を、ずっとシンプルにすると同時に、管理スタックの上方向だけでなく、横方向のデバイスとも自動的に情報を共有できるようにすることが必要です。

イーサネット・ファブリックは、次のような機能によってこの課題に対処します。

- 多層に分かれたスイッチングの管理を論理的に省略する。
- 多数の物理スイッチに対するポリシー適用とトラフィック管理を1台のスイッチのように実行する。
- ネットワーク帯域幅のスケーリングにあたって、スイッチポートやネットワークポリシーの構成の変更に手作業を必要としない。
- サーバ管理者、ネットワーク管理者、およびストレージ管理者に対して、単一のカスタマイズされたネットワークステータスのビューを提供する。

BROCADE VCS イーサネット・ファブリック・テクノロジー

Brocade VCS™ ファブリック・テクノロジーにより、確実に稼働する、効率性の高いデータセンター・ネットワークを構築することができます。Brocade VCS テクノロジーに基づいて構築されるイーサネット・ファブリック・アーキテクチャは、情報をノード全体で共有することにより、単一の論理シャーシとして管理し、管理を簡素化すると同時に、運用のオーバーヘッドを低減します。Brocade VCS テクノロジーは、従来のアーキテクチャや他社のファブリック・ソリューションにはない VM 対応機能とオートメーションを備え、統合されたファブリック上でのストレージをサポートします。

実績あるファブリック開発の歴史を背景にした Brocade VCS テクノロジーは、IT の俊敏性によって高い信頼性を保証し、初期投資において高いコスト効率性をもたらします。これにより、仮想化データセンターにおいて、弾力性が高く、高度に自動化された、ミッションクリティカルなネットワークへと円滑に移行していくことができます。

VCS テクノロジーは、Brocade VDX™ データセンター・スイッチ製品に組み込まれています。Brocade VDX データセンター・スイッチを利用することにより、クラウドに最適化されたネットワークングに対応したイーサネット・ファブリックを構築し、事業の俊敏性を向上することができます。Brocade VCS テクノロジーの詳細については、以下の Web サイトをご覧ください。

<http://brocade.com/solutions-technology/technology/vcs-technology>

ブロードについて

ブロードのネットワークング・ソリューションは、世界のトップクラスの企業に対して、アプリケーションと情報が各所に偏在する仮想化世界へのスムーズな移行を支援しています。これは Brocade One™ 統合ネットワーク戦略に基づいたアプローチであり、広い範囲にわたるコンソリデーション、コンバージェンス、仮想化、クラウド・コンピューティングの取り組みを実現しています。

業界をリードする、イーサネット、ストレージ、およびコンバージド・ネットワークングのソリューションは、これまでにないシンプルさ、ノンストップ・ネットワークング、アプリケーションの最適化、そして投資保護を実現することによって、最重要のビジネス目標の達成を目指す企業を支援します。ブロードでは、世界有数の IT 企業と提携して、教育、サポート、プロフェッショナル・サービスなど、広範囲にわたる総合的なソリューションを提供しています。詳細については、以下の Web サイトをご覧ください。

www.brocadejapan.com



BROCADE

ブロード コミュニケーションズ システムズ株式会社

〒100-0013 東京都千代田区霞ヶ関1-4-2 大同生命霞ヶ関ビル
TEL.03-6203-9100 FAX.03-6203-9101 Email:japan-info@brocade.com

BROCADEに関するより詳しい情報は、以下のWebサイトをご覧ください。

<http://www.brocadejapan.com>

©2012 Brocade Communications Systems, Inc. All Rights Reserved. 01/12 GA-WP-1550-03-J

Brocade、B-wing シンボル、DCX、Fabric OS、および SAN Health は、登録商標であり、Brocade Assurance、Brocade NET Health、Brocade One、CloudPlex、MLX、VCS、VDX、および "When the Mission Is Critical, the Network Is Brocade" は、米国またはその他の国における Brocade Communications Systems Inc. の商標です。その他のブランド、製品名、サービス名は各所有者の製品またはサービスを示す商標またはサービスマークである場合があります。

注意：本ドキュメントは情報提供のみを目的としており、Brocade が提供しているか、今後提供する機器、機器の機能、サービスに関する明示的、暗示的な保証を行うものではありません。Brocade は、本ドキュメントをいつでも予告なく変更する権利を留保します。また、本ドキュメントの使用に関しては一切責任を負いません。本ドキュメントには、現在利用することのできない機能についての説明が含まれている可能性があります。機能や製品の販売/サポート状況については、Brocade までお問い合わせください。