DATA CENTER FABRIC

Brocade DCX バックボーンによる サーバ仮想化環境の実現

サーバ仮想化の普及に伴い、データセンターでは「拡張性」、「性能」、および「管理性」の 3 つの側面において、新たなレベルの要件に対応することが求められています。本ホワイトペーパーでは、Brocade DCX バックボーンと他の Brocade 製品およびテクノロジを組み合わせることで、これらの課題をどのように解決できるのか、ケーススタディを含めてわかりやすく解説します。

BROCADE

目　次

はじめに ・・・

はじめに

データセンターでは今、抜本的な変革が進行しつつあります。企業はデータを戦略的資産として捉えはじめ、そのために企業が管理しなければならないデジタル・データの量は急激に増大しています。そしてこのデータの増加によって、データセンターのアーキテクチャに変革がもたらされようとしています。そしてこの変革を導く原理原則には、統合、仮想化、サービス・レベルに基づくリソース割り当て、およびポリシーベースのデータ管理の自動化などがあります。

Brocade® DCF（Data Center Fabric）アーキテクチャは、これらの要素を念頭におき、実際のデータセンター環境における動向に基づいて構想されました。この戦略的アーキテクチャは、既存の SAN（ストレージ・エリア・ネットワーク）ファブリックを、仮想インフラの管理を簡素化する、統合されたコンバージド・マルチプロトコル・ファブリックへと拡張します。Brocade DCF の中核を成す Brocade DCX® バックボーンは、次世代のファブリック・インフラで、卓越した性能（帯域幅、収容ポート数、電力効率）、運用を中断しない拡張性、優れた可用性、そしてエンド・トゥ・エンドのセキュリティといった DCF アーキテクチャに求められる厳しい要件に対応しています。また Brocade DCX は、FCoE（Fibre Channel over Ethernet）、CEE（Converged Enhanced Ethernet）、暗号化サービス、およびポリシー制御によるデータモビリティなど、先進のテクノロジにも対応するよう予め設計されています。Brocade DCX は、データの制御のためにアプリケーションと競合するのではなく、ファブリックのインテリジェンスをアプリケーションにまで拡張し、ファブリックを論理的にアプリケーション・サービス・レベルとデータ管理ポリシーに関連させます。これによって、インフラの利用率が向上し、プロビジョニング、キャパシティ・プランニング、および管理が容易になり、その結果インフラのコストが減少し、ビジネス拡大のための投資により多額の予算を割り当てることが可能になります。

Brocade DCX は、仮想サーバ導入のための必須条件となる「バックボーン・クラス」の性能を提供します。

- 8 ギガビット / 秒でシャーシあたり最大 384 ポートまでサポート
- 4 ギガビット / 秒の SAN ダイレクタに比べ、スイッチング・性能を 4 倍以上向上
- 革新的な 1 テラビット / 秒の ICL（Inter-Chassis Link）を実現

サーバリソースの利用率が向上するということは、帯域幅、特にファブリック・コアで必要となる帯域幅が増加するということを意味します。サーバ仮想化に対する需要の高まりと、8 コアまたはそれ以上の CPU 用を搭載した新しいブレードサーバの登場により、1 サーバ上で稼働する仮想マシンの数が増え、こうした要件に対応するため、ファブリック・コアには運用を中断することなく規模を拡張できる柔軟な拡張性が必要となります。

Brocade DCF テクノロジ

Brocade DCF アーキテクチャは、仮想サーバ環境の性能、拡張性、および管理性を向上するための製品とテクノロジで構成されています。次に示す製品とテクノロジは、サーバ仮想化のためのコスト効率性の高いデータセンター・ファブリックを構築する上で必要不可欠な要素です。

NPIV

仮想サーバからストレージへのアクセスには、物理サーバ環境と同等の安全性が求められます。しかし、単一の物理的な接続では、個々の仮想サーバに対して独立したストレージ・アクセスを提供することはできません。それどころか、すべてのストレージ・ポートおよび論理ユニット（LUN またはストレージ）がすべての仮想マシンに公開されることになり、セキュリティと管理性が低下してしまいます。

NPIV（N_Port ID Virtualization）は、この問題を解決するために設計された ANSI の標準規格です。NPIV によって、1 台の物理サーバにインストールされた 1 つのホスト・バス・アダプタ（HBA）で、仮想サーバ用の固有の WWPN（World Wide Port Name）を 255 個まで提供することができるようになります。また、NPIV をサポートするファブリック・スイッチは、仮想サーバがファブリックへログインする際にそれぞれの仮想サーバに固有のファブリック ID を割り当てることができます。NPIV サポートでは、標準的なファブリック・ゾーニングおよびストレージ LUN マスキングを仮想マシン（VM）で使用できるようになり、物理サーバ環境と同様にストレージ・ポートと LUN を適切な仮想サーバに分離することでセキュリティと管理性を維持します。

※ NPIV は、Brocade DCX バックボーンでサポートされています。

Access Gateway

Brocade Access Gateway オプションを使用すると、数の限られたスイッチ・アドレス（ドメイン ID）を使わずに、ブレードサーバ・シャーシのすべてのブレードサーバからファイバーチャネル（FC）トラフィックを集約できます。これによって、収容ポート数の少ないスイッチを大量に使用する必要がなくなるため、ブレードサーバ環境の拡張性が大きく向上します。また、サーバ管理と SAN 管理を明確に分離できるため管理性が向上し、サーバ管理者は、ファブリックに影響をまったく与えずに Access Gateway を構成することができます。F ポート・トランキング、帯域幅利用率、および自動リンク・フェイルオーバにより、管理の簡素化とコスト削減を実現できます。さらに NPIV サポートを使用することで、ストレージ管理の対象は物理サーバ HBA からさらにその先の仮想マシンにまで拡張されます。

※ Access Gateway は、Brocade 200E、Brocade 300、および Brocade 製各種ブレードサーバ内蔵スイッチでサポートされています。

Adaptive Networking

サーバ仮想化とストレージ仮想化の普及が進むことにより、データセンター・ファブリックには動的ワークロードを迅速に処理し、予想外のリソース不足による輻輳を回避しなければならないという新たなレベルの要件が課せられています。Brocade アダプティブ・ネットワーキング・サービスは、単一の物理リンクを介して複数の論理接続を作成し、これにより、仮想サーバから仮想ストレージに接続する際、QoS（クオリティ・オブ・サービス）の優先順位（高、中、低）に基づいて効率よくトラフィックを分離し、ファブリック・リソースを効率的に割り当てることができるようになります。アダプティブ・ネットワーキング機能には、Ingress Rate Limiting、Congestion Detection and Management、Traffic Isolation、および Fabric Dynamic Profiling（「Top Talkers」など）が含まれます。

これらのアダプティブ・ネットワーキング・サービスを使用することは、管理を簡素化し、ファブリック・リソースの効率利用を可能にします。物理リンクに複数の仮想チャネルをのせることにより、サーバの SAN ブート、ストレージ I/O、サーバ・クラスタのトラフィックを同時に扱うことができると共に、単一の物理接続を共有する場合に発生する動的な負荷の変動を分離して、互いに干渉しないようにすることができます。また、アダプティブ・ネットワーキング・サービスはファブリックの俊敏性を高め、例えばアプリケーションが物理サーバをまたいで移動したとしても、サービス・レベルに基づいて確実にアプリケーション・データが継続的に流れるようにします。

Brocade DCX バックボーン

Brocade DCF のコアとなるのが Brocade DCX バックボーンです。Brocade DCX は、サーバ統合により必要となる性能への厳しい要件に対応し、「妥協のない」仮想サーバの導入を実現します。

Brocade DCX バックボーンは、1 台のサーバで稼働する仮想マシン数がかつてないほど増加することによって生じる、要求される帯域幅の大幅な増大に合わせて拡張することができます。将来においては、システムの運用を中断せずに Brocade DCX バックボーン・アーキテクチャを FCoE および CEE といった新たなテクノロジに統合でき、VMware クラスタ接続に対応することができます。

CASE STUDY

以下の各ケースでは、仮想サーバ環境の 3 つの課題である拡張性、性能、管理性に対応するために、Brocade DCX バックボーンと他の Brocade 製品をどのように活用すればよいかを解説します。

Brocade DCX バックボーンは、8 ギガビット / 秒ファイバーチャネルのフル性能 / ノンブロッキング・ポートを、コア側で最大 384 個収容でき、各 Brocade DCX ポートは最大 255 個の NPIV アドレス（接続）をサポートします。この管理が容易な「シングル・ドメイン」ファブリックによって、膨大な数の物理サーバを統合することができます。また、仮想サーバを統合するための管理の容易なコンポーネントの一つとなります。

ケース 1: エッジ側での物理サーバ統合

物理サーバの統合には、コスト削減の点で大きなメリットがあります。例えば、ブレードサーバ・シャーシの場合、最大 14 ものサーバを 9U のラック・スペースにイーサネットおよびファイバーチャネル・スイッチのスペースも含めて統合することができます。イーサネットおよびファイバーチャネル・スイッチは、クラスタ間通信や SAN ストレージへの接続性を提供します。

Brocade がかなり早い時期からブレードサーバに対応していたのは、ファブリックに対するブレードサーバの影響を早くから認識していたからです。問題は、ファイバーチャネル・スイッチ（またはドメイン ID）アドレスには 239 個という制限があるため、収容ポート数の少ないスイッチをたくさん導入すると、これらのアドレスをすぐに使い尽くしてしまうということでした。Brocade は、ブレードサーバ・エッジ側での性能の向上を

実現し、管理の複雑さを軽減し、ファブリックの拡張性を向上するために、Access Gateway 機能を Fabric OS®（FOS）に組み込みました。図 1 に示すように、Access Gateway は、ブレードサーバの既存の 4 ギガビット / 秒の内蔵ファイバーチャネル・スイッチで使用でき、これらのスイッチを「N ポート・クラスタ」に変換します。Access Gateway のバックエンドでは、F ポートでシャーシ内の各ブレードサーバに接続します。フロントエンドでは、Access Gateway は複数の N ポート（モデルによって 4 から 6 個）を提供します。これらの N ポートは、負荷分散された 1 つの F ポート・トランクに集約してコアに接続することができます。Brocade DCX バックボーンの NPIV サポートにより、複数のブレードサーバの N ポートが 1 つの Brocade DCX スイッチ・ポートを介してログインすることができます。

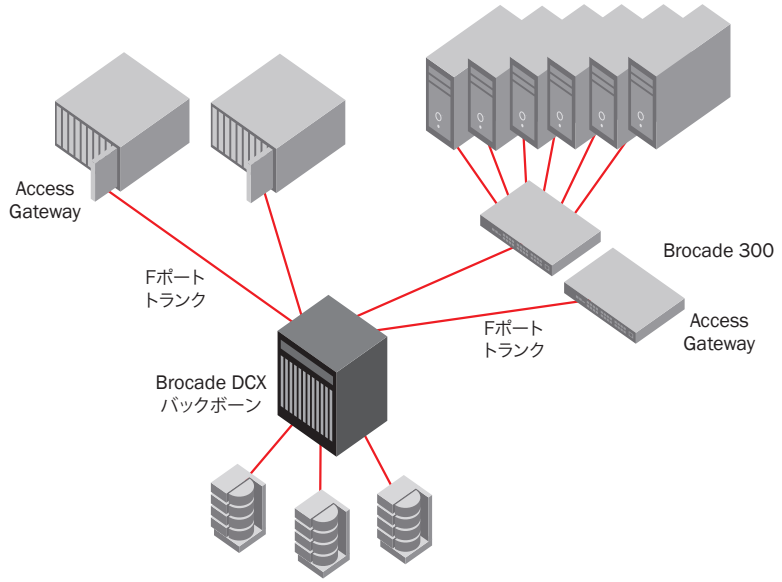

図 1. エッジ側でのサーバ統合

Access Gateway モードを使用してファブリック・エッジをまとめることには、いくつかのメリットがあります。第一に、ファイバーチャネル・アドレス・スペースを効率よく使うことができるため、拡張性が格段に向上します。8 台から 14 台のサーバごとに小さいドメイン・アドレス・スペースからアドレスを使用するより、はるかに大きいアレイとポート・アドレス・スペース（約 65,000 個のアドレス）を使うことができるため、拡張性が大幅に向上します。第二に、サーバ管理者は、Access Gateway をストレージ管理者側の SAN 管理から切り離して、独立して管理することができます。第三に、Access Gateway では、F ポート・トランキングを使用してバックエンド・ポートとフロントエンド・ポートの間のロード・バランシングを提供します。これにより物理リンク数をオーバープロビジョン（過剰投資）することなく帯域幅の利用率を向上させることができます。F ポート・トランク内のリンクの 1 つに障害が発生しても、トラフィックはトランク内の残りのリンク間に自動的に分散され、影響を最小限に抑えることができます。

ケース 2: エッジ側での仮想サーバ統合

サーバ統合の成果を上げるための次のステップとしては、当然サーバ仮想化の導入が検討されるでしょう。サーバ仮想化では、1 台の物理サーバ内に複数の仮想サーバを稼働させることで、サーバリソースの利用率を向上させます。つまり、複数のアプリケーション・ワークロードを単一のサーバ上で組み合わせてサーバ利用率を向上させ、必要な物理サーバ数を削減します。また、管理を簡素化するために、物理サーバのプールを作り、ワークロードの管理ソフトを使ってアプリケーションの負荷の変動にあわせて仮想サーバを物理サーバ間で自動的に移動する、といったことが可能です。ただし、仮想サーバによってサーバ利用率は大幅に向上しコストは削減されますが、管理者にとっては新しい管理上の課題が増えることになります。

物理サーバは、現在では複数のアプリケーションをサポートするため、1 物理サーバ当たりのファブリック帯域幅が増加し、それがネットワーク性能に（特にコア側に）影響を与えます。ゾーニングに関する問題もあります。ゾーニングでは物理サーバに接続されたすべてのストレージ LUN がすべての仮想マシンに対して公開されてしまうため、リスクが増大し、ファブリックの性能が低下する可能性があります。物理サーバ・クラスタを作成した場合、仮想サーバはプール内のどの物理サーバにでも移行することができます。また、すべての物理サーバに使用可能なすべてのストレージ LUN は、ファブリック内のすべての仮想マシンから参照できる状態になります。したがって、アクセス制御とセキュリティはさらに低下してしまいます。

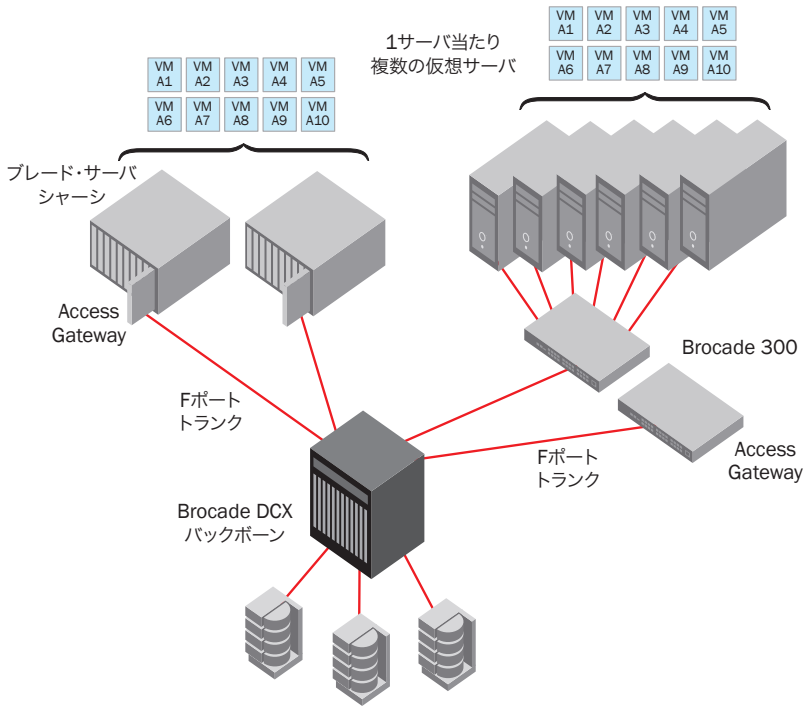

図 2. エッジ側でのサーバ統合

これらの問題は、NPIV（N_Port ID Virtualization）により解決することができます。 NPIV では、各仮想マシンに専用の仮想 N ポート（V ポート）を割り当てます。これにより、ファブリック・ゾーニングで、各仮想サーバが自分が使うストレージ LUN に割り当てられたストレージ・ポートしか認識できない状況を確保することができます。これには、サーバ HBA とファブリック・スイッチ・ポートの両方が NPIV をサポートしていなければならないことに注意が必要です。

仮想サーバ管理ソフトウェアは、多くの場合、仮想マシンの別の物理サーバへの移行をサポートしています。この移行プロセスでは、割り当てられている WWPN を仮想マシンと一緒に移動する必要があります。多くの仮想サーバでは、「パーソナリティ・プロファイル」と呼ばれる構成ファイルを使用して、そこに仮想マシンの構成情報を保存しています。この構成ファイルは、現在どの物理サーバが仮想マシンをホスティングしているかに関係なく、仮想マシンからいつでもアクセスが可能です。この例では、仮想マシンはオリジナルの構成ファイルの再読み込みを行い、オリジナルの WWPN を保持しているので、ファブリック・ゾーニングを変更する必要はありません。

ケース 3: エッジ側での拡張性に優れた性能

今日、ブレードサーバでは 4 コア CPU テクノロジと 4 ギガビット / 秒のファイバーチャネル接続が主流です。今後は、8 コア CPU と 8 ギガビット / 秒のファイバーチャネル・テクノロジを連携させることで、ブレードサーバの性能が向上することが予想されます。サーバの性能が飛躍的に向上すると、1 つの物理サーバで稼働する仮想マシンの数の増加に拍車がかかり、その結果ファブリックの性能に対するさらなる要求が生まれます。

次世代のブレードサーバ・シャーシでは、現行レベルの 2 倍から 3 倍、あるいはそれ以上に仮想マシン密度が増加することが予想されます。8 ギガビット / 秒テクノロジを搭載し、Fabric OS 6.0 以上が動作する Brocade DCX バックボーンであれば、この要求に応じた性能の拡張も容易です。図 3 に示すのは、初期時点では今日の 4 ギガビット / 秒、4 コア CPU のブレードサーバに接続されている Brocade DCX です。Brocade DCX は 1 つの物理スイッチ・ポート当たり最大 255 個の NPIV アドレスをサポートする独自の設計により、1 ブレードサーバ当たり 20 から 30 の仮想マシンという現在の比率から、8 コア CPU と 8 ギガビット / 秒の接続がもたらすはるかに高い集約率へと拡張することができます。

8 ギガビット / 秒サービスへのアップグレードは、4 ギガビット / 秒の SFP 光モジュールを 8 ギガビット / 秒の SFP+ 光モジュールに交換するだけで、ポート・レベルで行うことができます。ポートブレード単位でなくポート単位のアップグレードであるため実装が容易で、変更による影響も最小限に抑えることができます。アップグレードは、ケーブルと 4 ギガビット / 秒の光モジュールのプラグを抜いて、新しいケーブルと 8 ギガビット / 秒の SFP+ 光モジュールのプラグを差すだけという、非常に簡単な手順で行うことができます。

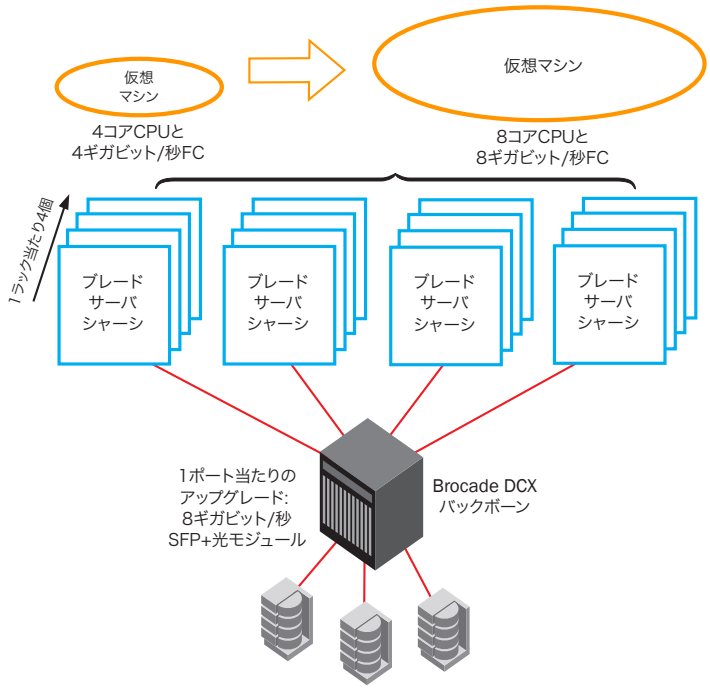

図 3. Brocade DCX の 8 ギガビット / 秒テクノロジによる仮想サーバの拡張

ケース 4: アダプティブ・ネットワーキング・サービス

性能要件はビジネス要件によって左右されるため、仮想マシンのワークロードには、アプリケーション・サービス・レベルとは異なる性能要件がある場合があります。例えば、ワークロードは高いがビジネス上の重要性は中程度のアプリケーションがある一方、別のアプリケーションは、ワークロードは低いがビジネス上の重要性はもっと高いということがあります。このような場合、アダプティブ・ネットワーキング・サービスを使用することにより、リソース要件（ワークロード）をサービス・レベル（QoS）と切り離すことが効果的です。

例えば、図 4 では、ブレードサーバと 1U サーバにさまざまなワークロードの仮想サーバが配置されています。ブレードサーバでは、1 物理サーバ当たり 5 つの仮想サーバが稼働しています。これに対して、1U サーバでは、アプリケーション・ワークロードが低いので、1 物理サーバ当たり 10 個の仮想サーバが稼働しています。ただし、ブレードサーバで実行されるアプリケーションにはサービス・レベル「中」が割り当てられているのに対し、1U サーバのアプリケーションにはサービス・レベル「高」が割り当てられています。この時、アダプティブ・ネットワーキングを使用することで、サービス・レベルとは切り離した中レベルの QoS をブレードサーバ・アプリケーションに割り当て、高レベルの QoS を 1U サーバ・アプリケーションに割り当てることができるようになります。

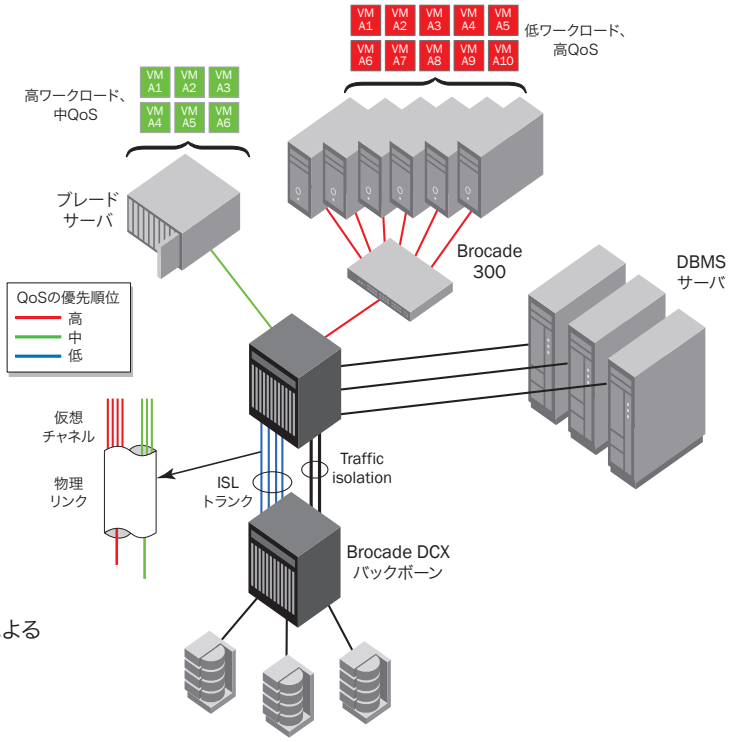

図 4. アダプティブ・ネットワーキング・サービスによるファブリック管理の簡素化

それぞれの QoS 優先順位が割り当てられたトラフィックは、すべてのスイッチで特定の仮想チャネルに割り当てられるため、ファブリック全体として QoS 優先順位が維持されます。つまり、QoS により、各仮想サーバからのデータ・フローは必ずファブリック全体にわたって同じ優先順位を持つことになるため、万一物理サーバ間で仮想サーバの移行が行われても、ファブリック内で予期せぬ輻輳が発生することを防ぐことができます。

もう 1 つのアダプティブ・ネットワーキング・サービスである Traffic Isolation を使用することにより、ISL（Inter-Switch Link）パスを、大きな帯域幅を持続的に必要とするアプリケーション専用に割り当てることができます。例えば、DBMS サーバからのワークロードが原因でピークが大きく、持続的に広い帯域が必要になることがあります。このトラフィックが仮想サーバ・アプリケーションのトラフィックと干渉しないことを確実にするためには、特定の ISL 接続に Traffic Isolation を導入します。Traffic Isolation は、DBMS トラフィックを特定の ISL に制限し、これによって DBMS トラフィックにバーストが生じてもスムーズにデータが流れ、予想外のボトルネックが発生する危険性を排除します。仮想サーバのリソースのボトルネックを抑制するために動的なワークロードの移行を行っている場合には、DBMS のデータ・フローが原因でボトルネックが発生すると、不必要な仮想マシンの移行を引き起こすことがあります。このような場合には、Traffic Isolation を使用することで、仮想サーバの不必要なスラッシングを回避できます。

アダプティブ・ネットワーキング・サービスは、Brocade DCX バックボーンと共に利用することによってリソース利用率を向上し、例えば仮想サーバの物理サーバ間の移行や、不必要な仮想サーバの移行を行わないようにするといった点においてシンプルな管理を実現します。

BROCADE

ブロケード 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11 階
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

©2009 Brocade Communications Systems, Inc.　All Rights Reserved. 02/09 GA-TB-079-01-J

Brocade、B-wing ロゴマーク、DCX、Fabric OS、File Lifecycle Manager、MyView、StorageX は、米国またはその他の国における Brocade Communications Systems Inc. の登録商標です。DCFM、SAN Health は商標です。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。