

はじめに：IT界に一大旋風を巻き起こす仮想化

データセンターで仮想化の利用が増えています。これを受けて、柔軟なリソース割り当て、別の場所へのワークロードの移行、実稼働での仮想ワークロードの利用に対応できる新たな形態のネットワークが必要となっており、最大限の可用性が求められています。物理サーバと仮想マシンの両方にまたがるネットワークを構築して一貫性のある機能を実現するためには、新しいアーキテクチャ手法でITインフラを設計および構築する必要があります。物理ネットワークと仮想ネットワークのインフラの管理に加えて、パフォーマンス、柔軟性、論理アドレス構造についても検討が必要です。仮想化に対応したネットワークを展開することで、共有インフラ上で画期的なサービスの数々を提供できます。

VMware、Citrix、Microsoft の仮想化技術では、既存のアプリケーションをカプセル化し、物理ハードウェアから切り分けて抽象化するという手法を用いています。物理マシンとは違って、仮想マシンはポータブルなソフトウェア・イメージであり、物理ハードウェア上で瞬時にインスタンス化できます。仮想化を利用すると、特定の物理サーバ上でアクティブに稼働する仮想マシンの数を調整するだけで、演算処理能力をオンデマンドで拡大/縮小でき、柔軟性が確保されます。しかも、稼働中の仮想マシンを別の物理サーバに移行することも可能です。大げさに言えば、仮想化によって場所の束縛から解放されるということです。遠く離れた場所どうしても仮想マシンをやり取りでき、その範囲はますます広がっています。クラウド・アーキテクチャやマルチテナント機能の進化が続くことで、スケールメリットも得られるようになりました。アプリケーションの違い、部署の違い、企業の違いを越えて、適切に分割された共通のインフラにリソースを集約することによって生まれるメリットです。

仮想マシンには、柔軟性、可搬性、自動化、高密度といった特徴があり、それに応じた新しいネットワーク・アーキテクチャが必要になっています。ハイ・パフォーマンスを重視し、可搬性に対応したアーキテクチャであることに加え、仮想マシンはデータセンターの新たな構成要素だということを前提としたアーキテクチャであることも求められています。また、仮想化インフラ内のさまざまな場所で作成や移動が行われる仮想マシンの状態と場所を明確に管理できるよう、仮想化に対応した一貫性のあるポリシーと制御を適用できるネットワークであることも必要です。

仮想マシンの氾濫

仮想サーバを利用した構成で生じる直接的な影響の1つが、仮想マシン (VM) の急増です。新しい処理アーキテクチャでは、サーバ1台につき10~20以上の仮想マシンを稼働させることが見込まれており、管理の対象要素が大幅に増えることとなります。このような要因は今後もさらに増えていきます。高速かつ高密度なマルチコアCPUを搭載した物理サーバの普及とも相まって、VMの氾濫はますます加速しています。これに比例して、それぞれの物理サーバ内でVMが接続する仮想スイッチも増えています。VMをホストする物理サーバはそれぞれ仮想スイッチを持つことから、管理対象のネットワーク要素の数は20~40倍の増加となります(図1を参照)。基本的に、仮想スイッチによって構成されるネットワーク・アクセス・レイヤは、現在では物理サーバ内に存在しています。

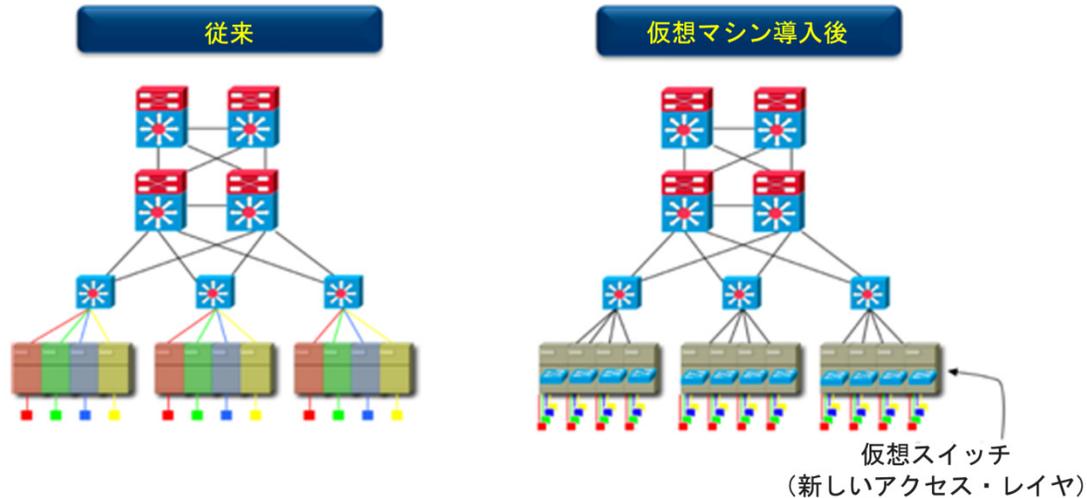


図 1: 仮想化に伴って仮想スイッチで形成される新しいアクセス・レイヤ

仮想化で求められるクラウド・ネットワーキングの一貫性

仮想マシンと仮想スイッチで構成されるこのような巨大なインフラでは、基盤となるネットワーク・ファブリックにおいて、シームレスなトランザクションに対応するための新たな要件が求められます。具体的には、ユーザーから VM、VM から VM、VM からデータ、VM からフォールト・トレラント・ピア、VM の可搬性といった、新しい種類のトランザクションが発生します。そこで、それらの要件を満たすことを目的として構築されたネットワーク・アーキテクチャが必要となります。それがクラウド・ネットワーク・アーキテクチャです。このアーキテクチャには次のような要件があります。

1. **VM の急増**: 仮想マシンは1台の物理サーバ上で多数インスタンス化できることから、物理 NIC の帯域幅の使用率もそれに比例して上昇します。そのため、NIC リンクは大幅なアンダーサブスクリプトではなくなります。従来型のオーバーサブスクリプトのネットワーク・トポロジは、仮想化とプライベート・クラウドに合わせた再構築が必要です。

ポータブルな VM イメージはサイズが数ギガバイトですので、VM の移動や移行の際には、ネットワーク上で大量のデータが伝送されることとなります。また、ワークロードの柔軟性を確保するために、演算リソースへの仮想マシンの割り当ては、さまざまな条件に基づいてプログラムでスケールアップ/スケールダウンされます。条件とはたとえば、負荷、時間帯、電力と冷却の可用性などです。このような仮想化環境向けに、ピーク時のパフォーマンスをふまえてネットワークを設計する必要があります。

2. **クラウド・アプリケーション**: 最近のクラウド・アプリケーションは、Web 2.0 技術やリッチ・メディア技術をマッシュアップなどの手法で組み合わせており、世界中で数百万人規模のユーザーが利用可能です。これに伴ってネットワーク上で発生する大量のトランザクションでは、下り (VM からユーザーへ) のトラフィックの方がはるかに多くなります。

クラウド・アプリケーションのワークロードは、演算処理のタスクを複数レイヤのワーカ・ノードとデータ・ノードに分散する設計のため、VM 間で発生するやり取りが従来とは桁

違いです。また、RESTful の考え方では、演算処理の状態をデータ・ノードにのみ保持しておくことになり、VM がネットワーク・ファブリックを介してバックエンド・データベースに絶えずアクセスする必要が生じます。

3. **VM の移行（マイグレーション）**：仮想マシンの可搬性はますます高まっています。仮想マシンを移行するタイミングは、サーバやオペレーティング・システムをリプレース/アップグレードするとき、使用率の低いサーバをシャットダウンして節電できるようにそのサーバのワークロードを移動するとき、アプリケーションのパフォーマンス管理を目的として使用率の高いサーバから VM を移動するとき、より低コストで演算処理を行える場所にワークロードを移行する機会を得たときなどです。VM の可搬性を確保するためには、大規模でフラットなレイヤ 2 ドメインのネットワークを構築し、VM の移動時にも IP アドレスや処理中のクライアント・トランザクションに支障が生じないようにする必要があります。
4. **仮想スイッチの管理**：仮想ネットワークの管理はサーバ管理者が行うことが一般的です。ネットワーク管理者は組み込みの仮想スイッチに直接アクセスできないためです。このことは、大規模な仮想化環境やプライベート・クラウド環境では厄介な問題となります。ネットワーク・インフラ全体で、膨大な数の仮想スイッチに対して、一貫性のあるポリシー適用、監視、診断を行う必要があるからです。
5. **クラウドの広がり**：インフラストラクチャ・アズ・ア・サービス（IaaS）、パブリック・クラウド、仮想プライベート・クラウドのサービスを提供するプロバイダが増える中、現時点では、プロバイダ間で仮想マシンを移動した場合にネットワーク・ポリシーと課金状態を維持しておく方法がありません。移動先の IaaS クラウド・プロバイダに対し、ネットワークの状態を安全かつ確実に伝達することで、利用企業自身の主導でセキュリティ・ポリシーを制御し、プロバイダ間での相互運用性を確立する必要があります。

仮想化向けに最適化したクラウド・ネットワークの設計

仮想ネットワークと物理ネットワークを組み合わせ、物理環境、仮想環境、クラウド環境に対応したネットワークを構築する作業は、一筋縄では行きません。パフォーマンス、耐障害性、ポリシー制御、管理のしやすさについて、設計段階で考慮しておく必要があります。先ほど説明した、仮想化とクラウド・コンピューティングに対応したネットワークの特性をふまえると、旧来のネットワーク手法とは発想を変えて、最新のクラウド・ネットワーキング・アーキテクチャを採用することが求められます。そんなアーキテクチャを実現できるのが、Arista 7000 ファミリの 1Gb/10Gb イーサネット・スイッチと Extensible Operating System (EOS) です。EOS には次のような特長があります。

1. **10Gb イーサネット**：高性能な 10GbE ネットワークは、将来の 40/100GbE への拡張を見据えた コア・ネットワークには不可欠であり、エッジ・ネットワークの多く（特にブレード・サーバの場合）にも必要です。VM によってネットワークの利用帯域が増えたり、数百万人のユーザーが同じクラウド・アプリケーションを同時に利用したり、ネットワーク・ファブリック内でギガバイト級のサイズの VM イメージを移動したりという状況に対応するには、10Gb イーサネット・ネットワークが現時点で唯一の選択肢です。高性能で反応が速く、クラウドと VM のワークロードで生じるピーク時の利用帯域にも対応できるネットワークを構築するためのトランスポート技術として、10Gb イーサネットは最適です。このようなハイ・パフォーマンスが求められる環境には、Arista のワイヤスピード 10Gb イーサネット・スイッチがぴったりマッチします。
2. **対称なクロスセクション帯域幅**：現在のネットワーク・アーキテクチャの大半は、電子メールという一大アプリケーションへの対応を念頭に置いて設計されています。しかし、仮想化や最近のアプリケーション・ワークロードの出現によって、従来のネットワーク設計の前提だったオーバーサブスクリプトの原則は変わりつつあります。NIC リンクの使用率が上がったことと、ユーザーから VM へのトラフィックや VM 間トラフィックの対称性によって、入力と出力のスイッチング帯域幅のバランスを高めることが必要になっています。入力と出力の帯域比は 1:1 または 2:1 です。従来の設計で 20:1 や 40:1 の場合もあったのとは対照的です。たとえば、トップオブラックの Arista 7048 リーフ（葉）スイッチで、1GbE のサーバ・アクセス・ポートが 48 個の場合、1:1 の設計規則では、10GbE のアップリンクが 4 個必要です。10GbE ポートが 48 個のスイッチ（ブレード・サーバのアクセス・スイッチとして、または複数のリーフ・スイッチを集約するスパイン（幹）スイッチとして使用）を 2:1 のオーバーサブスクリプトで稼働する場合、入力には 10GbE ポート 32 個、出力には 10GbE ポート 16 個を割り当てる必要があります。ワイヤスピード・アーキテクチャを備えた Arista 7000 ファミリの 1GbE/10GbE スイッチなら、バランスの取れたトラフィック・パターンでクラウド・ネットワークを展開できます。オーバーサブスクリプトがなければ、その管理も必要ありません。
3. **リーフ/スパイン型アーキテクチャ**：VM 間の通信や VM の可搬性が常に必要なことから、レイヤ 2 ドメインが拡大し、従来の企業で見られた 3 層設計ではなく、フラットな 2 層のリーフ/スパイン型アーキテクチャが増えています。このアーキテクチャでは、VM から別の VM への通信の物理ホップは 2 回以下で済みます。図 2 は、Arista 7000 ファミリの 1GbE/10GbE スイッチでこのアーキテクチャを構成した例です。

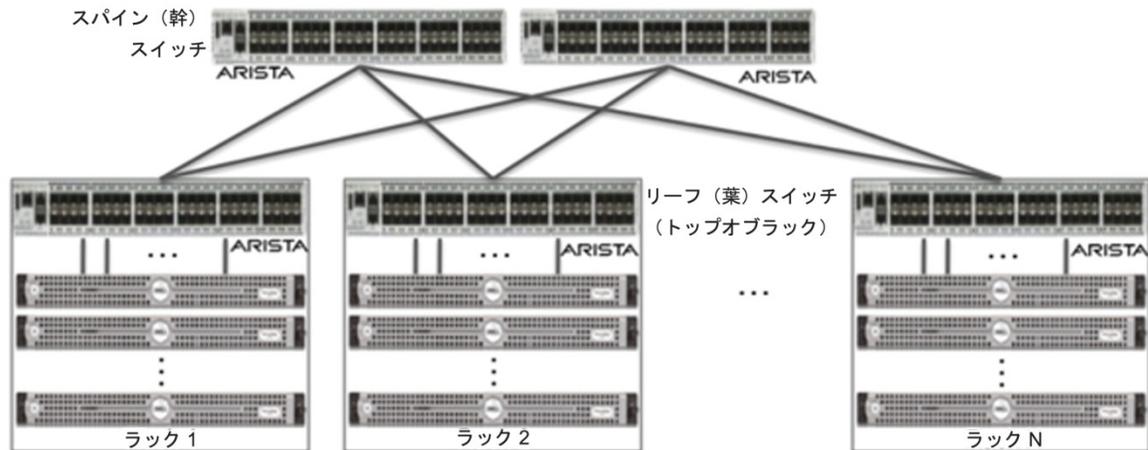


図 2 : リーフ/スパイン型のクラウド・ネットワーキング・アーキテクチャ

- 低レイテンシーでのスイッチング** : アプリケーションのレスポンス・タイムを向上するためには、レイテンシーを抑えることと、適切な帯域幅をプロビジョニングすることがきわめて重要です。カットスルー方式の packet 転送モードを利用したスイッチは、ストア・アンド・フォワード方式の転送モードの場合と比べて、スイッチあたりのレイテンシーを 50~90% 抑えることができます。また、帯域幅のプロビジョニングによってシステム全体でレイテンシーを抑えるためには、アーキテクチャ面からの対策が欠かせません。ここでも、10GbE の伝送基盤と、2 層のネットワーク設計が必要です。レイテンシーを抑えることで、演算処理の効率が上がり、短時間、省電力、低コストでビジネスの成果を上げられるようになります。
- 耐障害性** : フォールト・トレラント・コンピューティングの原則に従えば、物理ノードにせよ仮想ノードにせよ、一部の演算処理ノードで障害が発生しても、ワークロードは影響を受けません。現代のクラウド・ネットワークでは、耐障害性をサービス・レベルで実現する必要があります。スイッチで発生する障害の大きな原因は、旧式化したオペレーティング・システムやソフトウェア・アーキテクチャにあります。フォールト・トレラント・コンピューティングの原則と同様に、ネットワーク・オペレーティング・システムも、フォールト・トレラント性と拡張性を中核に取り入れたオペレーティング・システム設計原則にのっとり構築しておく必要があります。さらに、クラウド・ネットワークの耐障害性を実現するうえでは、コントロール・プレーンとデータ・プレーンを分離し、高速なネットワーク・コントロール・プレーンを備えておくことも前提になっています。Arista の EOS は、このような機能が設計の根底からネイティブに組み込まれており、最高レベルのネットワーク耐障害性を実現できます。
- 仮想化** : クラウド間にまたがって VM を簡単に移動できれば、戦略的に見て大きなメリットが得られるはずですが、しかし、手動で管理する現在のネットワークは、ポート割り当てやポリシー定義を静的に行うことが前提となっていて、物理ネットワークと仮想ネットワークの間につながりがなく、ネットワーク・オペレーティング・システムで一貫したポリシーを適用するしくみもありません。仮想スイッチと物理スイッチの間で一貫性のあるシームレスな対応を実現するために、仮想スイッチでは、標準ベースの各種技術を利用した透過的なリダイレクトを行います。たとえば、IEEE 802.1Q の VLAN タグ、MAC アドレス、トンネリングなどを利用して、物理スイッチに対して透過的な処理を行います。独自規格のタグは、ベンダー選択の幅が狭まり、相互運用性にも影響するため、避けるのが賢明です。

物理ネットワークと仮想ネットワークにまたがって一貫したネットワーク管理を実現するために必要なのは、ネットワーク管理者が使い慣れたコマンド・ライン・インターフェイスで多種多様な仮想スイッチを管理できるようにして円滑な導入を実現することと、SNMP、XML、XMPP など、プログラムとの親和性の高い抽象化の手法により、クラウド OS がネットワーク・インフラを API ベースで管理できるようにすることです。仮想ネットワークと物理ネットワークの違いを越えて、VM の移行時も含めて一貫性のある設定と管理を実現するためには、物理ネットワーク、仮想マシン、クラウド実装のすべてで統一された管理が重要になります。

Arista の vEOS ソリューション

仮想ネットワークと物理ネットワークにまたがってシームレスな運用管理を実現するには、以下の点が鍵となります。

- ネットワーク管理者が仮想スイッチを物理ネットワークの延長線上でとらえられるようにし、現在の物理スイッチと同様に管理できるようにすること。
- VM のホストの管理者はネットワーク設定を行わずに済むこと。
- ネットワーク管理者にとってわかりやすいトラブルシューティングと自動検出を実現すること。
- 物理スイッチと仮想スイッチの違いを越えて、一貫性のあるポリシーの導入と適用を全体に対して一元的に行えること。

Arista では、物理ネットワークと仮想ネットワークの両方を管理するための業界標準のインターフェイスを用意しています。Arista 7000 ファミリのネットワーク・プラットフォームでは、Extensible Operating System (EOS) で物理クラウド・ネットワークの管理が可能です。仮想ネットワークもシームレスに一元管理できるよう、EOS ソフトウェアの実装には、vEOS、仮想アプライアンス、スイッチが組み込んであります。vEOS によって、物理ネットワークで使用してきた業界標準のコマンド・ライン・インターフェイス (CLI) や標準ベースの管理プロトコルが、仮想ネットワークにも拡張されます (図 3 を参照)。

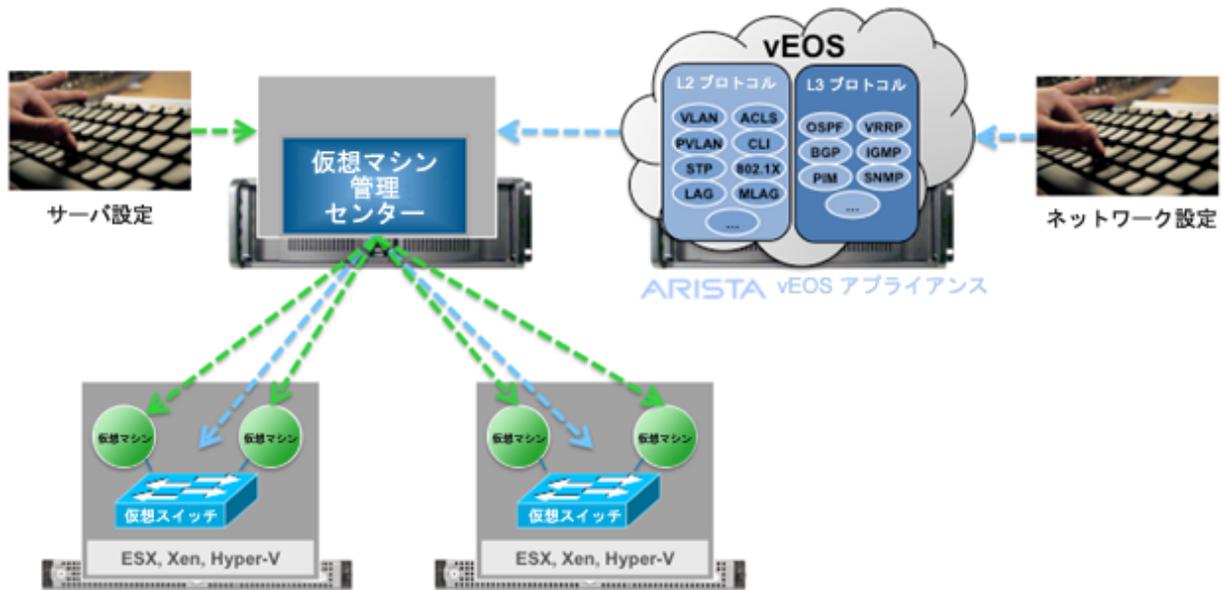


図 3：物理ネットワーク、仮想ネットワーク、クラウド・ネットワークを管理する vEOS

ネットワーク管理者にとっては、vEOS の場合でも、物理スイッチと同様の管理インターフェイスを利用できます。たとえば、SSH/Telnet によるアクセスや、業界標準の CLI などです。HP Openview などの SNMP 管理ツールに統合して、VM の自動検出も可能です。さらに、次のようなネットワーク管理作業も進めやすくなります。

- ポート・グループの設定
- アップリンクの設定
- 仮想スイッチの監視、運用監視、状態監視
- 分散ポート・プロファイルの作成による物理ネットワークと仮想ネットワークの一元管理

vEOS では、サーバ管理者とネットワーク管理者の双方が、仮想化されたクラウド・インフラの管理を並行して進めることができます。そのとき使用するワークフローも、お互いが使い慣れた個別のワークフローです。

vEOS のもう 1 つの特長が、VM 移行時のネットワーク・プロビジョニングを自動化できることです。P2V（物理から仮想）、V2V（仮想から仮想）、V2C（仮想からクラウド）のそれぞれの移行に対応しています。物理ポートから仮想ポートに、あるいは仮想ホストから別の仮想ホストにネットワーク経由で VM を移行するときに、vEOS では、該当するネットワーク設定も移行されるため、エンド・ツー・エンドの一貫性が確保されます。他社のソリューションでは、既存の仮想スイッチを当該ベンダーの製品に置き換えることが必要で、その間の中断が発生します。これに対し、Arista の EOS では、既存の仮想スイッチを利用したままで一元的な管理を実現できます。

まとめ

仮想化とクラウド・コンピューティングを組み合わせた新発想のコンピューティング環境では、ネットワークに関する入念な検討が欠かせません。クラウド・ネットワークは、仮想サービスを特定の物理サーバから切り分けて抽象化する処理に対応する必要があります。具体的には、10GbE ネットワークと、協調動作のしくみを使って、ネットワーク帯域幅と仮想マシンのリソースをオンデマンドで確保する機能や、複数のワークロードや顧客を分離する機能が必要です。

サイロ型のネットワークから、仮想化とクラウドに移行するときには、大幅なオーバーサブスクライプのスイッチを、仮想化とクラウド向けに設計されたネットワーク・インフラに置き換えることになります。そんなインフラの例が、リーフとスパインの 2 層構成で展開した Arista 7000 ファミリです。Arista 7000 ファミリと EOS/vEOS という組み合わせで、VMware の vSphere/vNetwork Distributed Switch、Citrix の Xen、Kernel-based Virtual Machine (KVM)、Microsoft の Hyper-V のハイパーバイザへのスムーズな移行が期待できます。協調動作は、業界標準の探索プロトコル、CLI、拡張可能な API によって実現されます。もちろん、物理ネットワーク、仮想ネットワーク、クラウド・ネットワークは、単に共存するだけでなく、シームレスな管理でスケーラブルな展開が可能でなくてはなりません。

Arista 7000 ファミリと vEOS という組み合わせは、物理ネットワークと仮想化サービスを運営する事業者や、プライベート・クラウドやパブリック・クラウドのシステムにとって、大きな支えになります。ネットワーク、サーバ、組織の垣根を越えて、仮想マシンをシームレスに移行できるようになり、一貫性のあるポリシーや課金処理を適用できます。物理ネットワークから仮想ネットワークへ、さらにはクラウド・ネットワークへとワークロードを移行できるのは、現在のサイロ型ネットワークからの大きな飛躍です。

本ドキュメントに記載の情報は、Arista Networks の製品に関連するものです。詳しくは、<http://www.aristanetworks.com> をご覧いただくか、sales@aristanetworks.com までお問い合わせください。