ソリューションのアーキテクチャ

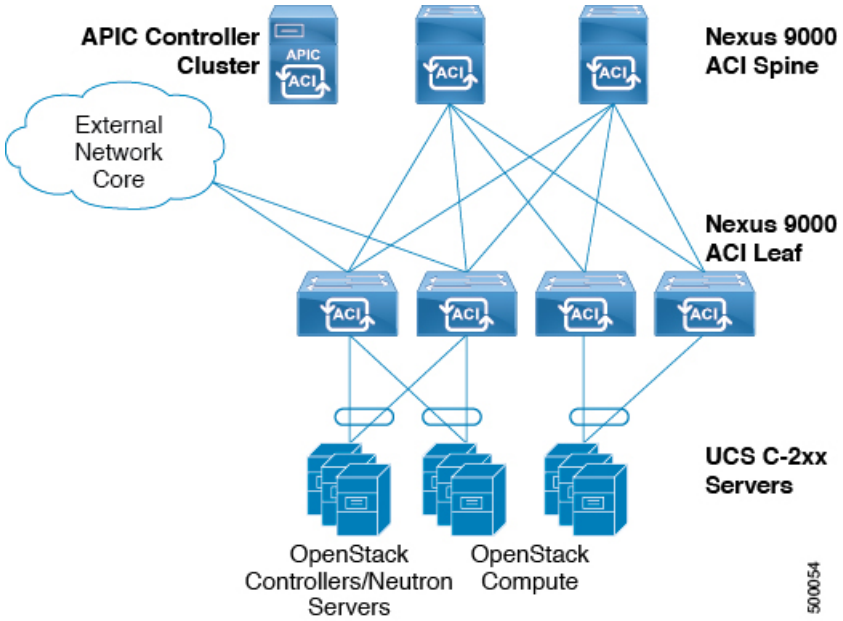
この章の内容は、次のとおりです。

OpenStack の物理アーキテクチャを備えた ACI

OpenStack を導入した通常の ACI ファブリックのアーキテクチャは、Nexus 9000 スパイン/リーフ トポロジ、APIC クラスタ、およびサーバ グループから構成され、OpenStack のさまざまな制御コンポーネントやコンピューティング コンポーネントを実行します。OpenStack はさまざまな方法で導入できますが、基本的なテスト アーキテクチャは、Neutron のネットワーク ノードとしても機能する少なくとも 1 つの OpenStack Controller サーバと、仮想マシン（VM）インスタンスをホストする 2 つ以上の OpenStack コンピューティング ノードから構成されます。ACI 外部ルーテッド ネットワーク接続をファブリック外のレイヤ 3 接続として使用して、OpenStack クラウド外の接続を提供することができます。

（注） この導入ガイドの検証済みの設定には、スタンドアロン モードの Cisco UCS C シリーズ ラック マウント サーバを使用しています。OpenStack を実行しているサードパーティ製のスタンドアロン ラック サーバにも対応できます。ACI ファブリックに接続されたファブリック インターコネクトで UCS Manager を実行するシステムについては、今後のリリースでサポートされる予定です。

図 *1*：*OpenStack* の物理トポロジを使用した *ACI* の例



OpFlex ML2 のソフトウェア アーキテクチャ

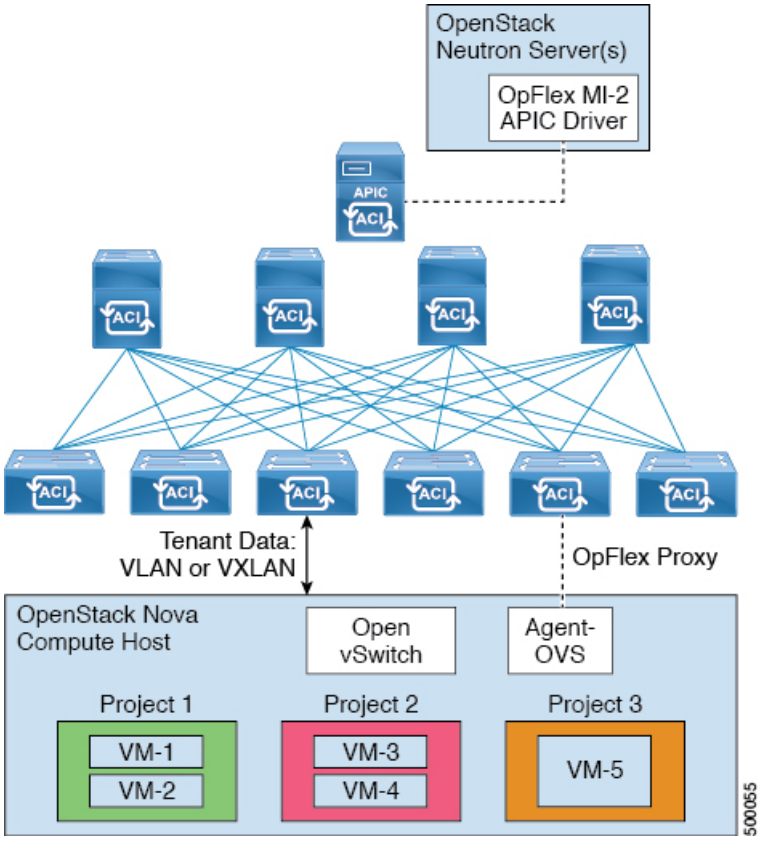
OpenStack のModular Layer 2 のフレームワークでは、TypeDriver と MechanismDriver に基づいてネットワーキング サービスを統合することができます。一般的なネットワーキング タイプのドライバには、ローカル、フラット、VLAN、VXLAN などがあります。OpFlex は、OpFlex の設定で定義した VXLAN か VLAN のいずれかの実際のパケット カプセル化により、ML2 を通じて新しいネットワークとして追加できます。メカニズム ドライバでは、ネットワーキングの要件をNeutron サーバから Cisco APIC クラスタへ伝えることができます。APIC メカニズム ドライバは、ネットワーク（セグメント）、サブネット、ルータ、または外部ネットワークなどの Neutron のネットワーキング要素を ACI ポリシー モデル内の APIC 構造に変換します。

現在は、OpFlex ML2 ソフトウェア スタックは修正後の Open vSwitch パッケージと、Neutron サーバおよびOVS と通信する各 OpenStack コンピューティング ホスト上のローカル ソフトウェア エージェントも利用しています。ACI リーフ スイッチからの OpFlex プロキシは、各コンピューティング ホストの Agent-OVS インスタンスとポリシー情報を交換して ACI スイッチ ファブリックとポリシー モデルを仮想スイッチまで効率的に拡張します。これにより、VM インスタンスがネッ

トワークに接続されるバーチャル ポートを起点とする仮想および物理スイッチング ファブリックの組み合わせのどこにでもネットワーク ポリシーを適用できる結束力のあるシステムがもたらされます。次の図に、OpFlex ML2 APIC ドライバおよびACI ファブリックの相互作用と、コンピューティング ホスト上の Agent-OVS サービスへの OpFlex プロキシの拡張を示します。

（注） Neutron への統合のための OpFlex ML2 APIC ドライバは、`neutron-server` サービスを実行しているサーバ上で実行します。このサーバは、他の OpenStack ソフトウェア要素を実行しているコントローラ ノードか、Neutron 機能専用のサーバである場合があります。複数の Neutron サーバによる高可用性設定もサポートされています。

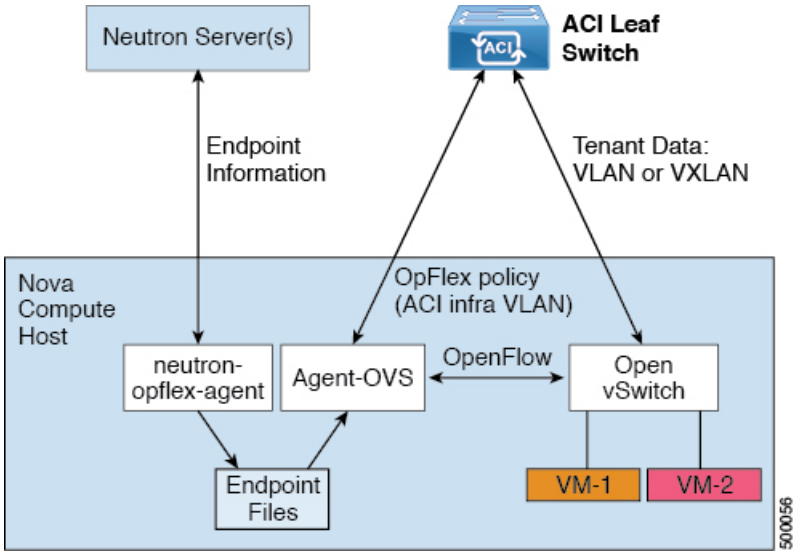
図 *2*：*OpFlex ML2* を持つ *ACI* アーキテクチャを備えた *OpenStack*



コンピューティング ノードでは、neutron-opflex-agent サービスが OpenStack のエンドポイントに関する情報を Neutron サーバ上の ML2 ドライバ ソフトウェアから受信します。この情報は、`/var/lib/opflex-agent-ovs/endpoints` にあるエンドポイント ファイルにローカルに保存されます。`agent-ovs` サービスはエンドポイント情報を使用して、接続された ACI リーフ スイッチ上の OpFlex プロキシを通じてエンドポイントのポリシーを解決します。次に、`agent-ovs` が、ローカルに適用可能なポリシーに OpenFlow を使用して OVS 上でポリシーをプログラミング

します。非ローカルポリシーは、アップストリーム リーフ スイッチで適用されます。次の図に、コンピューティング ノードで実行している OpFlex モジュールと OVS 間の相互作用を示します。

図 *3*：コンピューティング ホスト上の *OpFlex* エージェントのアーキテクチャ



論理 OpenStack トポロジ

OpenStack は、クラウド サービスを提供するサーバ ノードに関する複数のネットワーク接続要件を定義します。さまざまな OpenStack サービス間の API 通信のほかに、管理トラフィック、テナント データ、および外部ネットワーキング要件について、通信パスを定義し、提供する必要があります。また、導入でストレージ トラフィックやその他の特定のニーズ専用のネットワーク セグメントを指定する場合もあります。ACI スイッチング ファブリックは、これらのすべての要件を満たすネットワーク サービスを提供できます。サーバ接続は、別個の物理インターフェイスか、Cisco VIC などの仮想化ネットワーク アダプタ、または Cisco UCS B シリーズなどの管理型のブレード サーバ システムのいずれかから構成できます。

- 管理および API ネットワーク：このネットワーク セグメントは、サービスへの API 直接通信および OpenStack 機能間での API 通信とともに、OpenStack サーバへの管理用セキュア シェル アクセスを行えるようにするためのものです。また、管理および API 機能はさまざまなネットワークセグメントにさらに分割できます。このガイドでは、単一ネットワーク セグメントを両方の設定例に使用します。
- 外部ネットワーク：OpFlex と統合された ACI ファブリックでは、APIC の外部ルーテッド ネットワーク設定によって、外部ネットワーク パスが提供されます。Neutron L3 エージェントを実行しているシステム内の外部ネットワークが、ソフトウェア ベースのルーティング機能の外部にあるネットワークです。外部ネットワークは NAT サービスを利用して、隠れているか、または重複している IPv4 アドレス空間をテナントで使用できるようにします。

- テナントのデータ ネットワーク：OpenStack 内のテナント ネットワークはテナントによって動的に作成され、クラウド内の VM インスタンス間の接続を提供するほか、クラウドベースのルーティング サービスを他のテナント ネットワークまたは外部ネットワークに接続します。OpFlex でテナント ネットワークに割り当てられたセグメント ID は ACI ファブリックによって追跡され、リーフ スイッチ間の VXLAN と、リーフ スイッチとサーバ間の VXLAN または VLAN で構成されます。

分散 Neutron サービス

OpenStack Neutron は、クラウド環境で動作する VM インスタンスに必要な共通のネットワーキング構造とサービスを定義します。これらの機能のすべてを単一サーバ上、または小規模なサーバのクラスタ上に実装する場合、Neutron サービスの可用性と拡張性の両方が関心事項となる場合があります。OpFlex ML2 Driver ソフトウェアは、サービスの可用性を高めながらも単一インスタンスのサービス負荷を軽減するスケールアウトのアプローチを使用して、これらのネットワーク サービスをクラスタ内のコンピューティング ノードに分散する機能を提供します。

次の OpFlex OpenStack Neutron ML2 サービスは、OpFlex ML2 Driver ソフトウェアの使用時にコンピューティング ノードに分散させることができます。

- 外部ネットワーク用の NAT：外部ネットワークをサポートするための Opflex ML2 Driver のアプローチでは、OpenStack の送信元 NAT 機能とフローティング IP 機能をコンピューティング ホストの Open vSwitch に分散します。プライベート OpenStack 空間で定義されていない IP アドレス宛のパケットは、コンピューティング ホストから出力される前に自動的に NAT で変換されます。次に、変換されたパケットは、APIC に定義されている外部ルーテッド ネットワークにルーティングされます。分散 NAT サービスはソリューションに組み込まれます。

- レイヤ 3 フォワーディング：レイヤ 3 エージェントの Neutron リファレンス ソフトウェアの実装は、ACI ファブリック内のレイヤ 3 フォワーディングと、コンピューティング ノード内のローカル フォワーディングによって置き換えられます。同じ OpenStack テナント ルータに接続する 2 つの VM が同じコンピューティング ノードに存在する場合、それらの間のレイヤ 3 トラフィックは OVS によって転送され、その物理サーバにローカルで維持されます。コンピューティング ノードにローカルなトラフィックの分散型レイヤ 3 はソリューションに固有のものです。

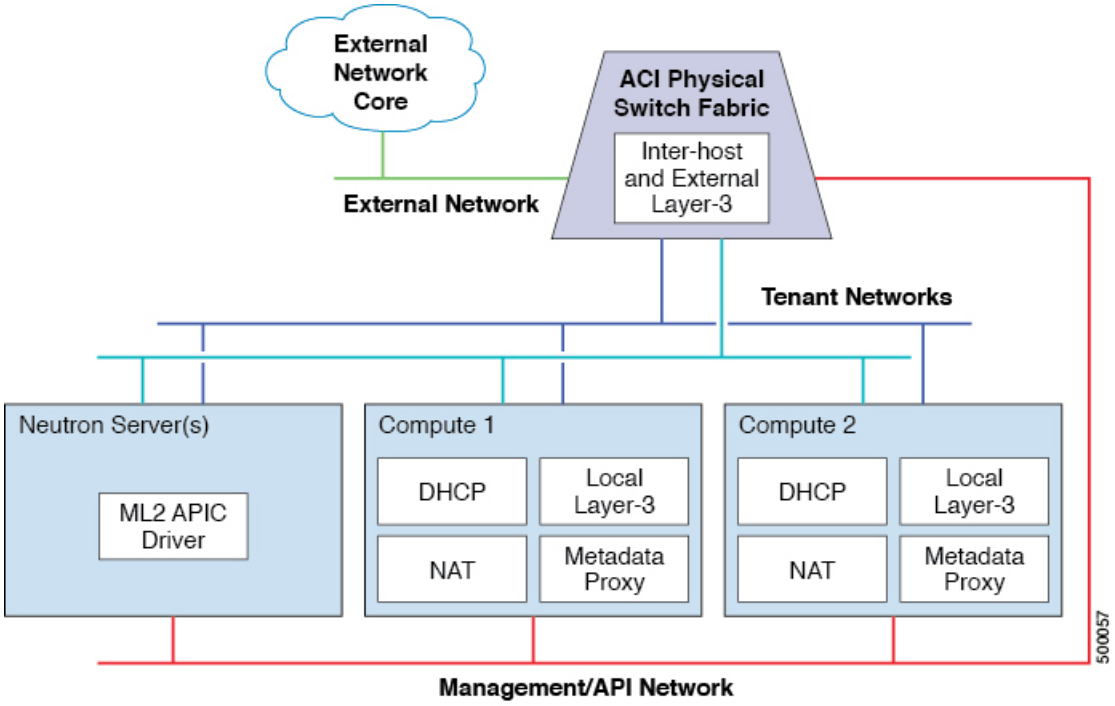
- DHCP：リファレンス Neutron ソフトウェアの実装には、Neutron サーバに集中化された DHCP エージェント サービスがあります。OpFlex ML2 ドライバ ソフトウェアでは、agent-ovs サービスを使用した分散 DHCP アプローチが有効です。DHCP 機能をコンピューティング ノード全体に分散することで、DHCP ディスカバリ、オファー、リクエスト、および確認応答（DORA）のトラフィックをホストに対してローカルに保つことで、VM インスタンスへの IP アドレッシングの割り当てに信頼がおけるようにします。集中型の Neutron アドレス管理機能は、DHCP アドレッシングおよびオプションを管理ネットワークを通じてローカルの agent-OVS に通知します。この最適化された DHCP のアプローチは、このソリューションでデフォルトによって有効になりますが、必要に応じて従来の集中型モードに戻すこともできます。

- メタデータ プロキシ：OpenStack VM は、Nova メタデータ サービスからのインスタンス ID、ホスト名、および SSH キーなどのインスタンス固有の情報を受信することができます。この

サービスには、通常、OpenStack VM インスタンスの代わりのプロキシとして機能する Neutron サービスを通じて到達します。OpFlex ML2 ソフトウェアでは、このプロキシ機能をコンピューティング ノードのそれぞれに分散することができます。この最適化されたメタデータ プロキシはデフォルトで無効になっています。また、従来の集中型または分散型のアプローチのいずれかを設定できます。

次の図の論理トポロジは、Neutron サーバと分散 Neutron サービスを含むコンピューティングホストからの OpenStack ネットワーク セグメントへの接続を示します。

図 *4*：分散 *Neutron* サービスによる論理 *OpenStack* ネットワークの接続



（注） OpenStack 用の管理/API ネットワークは、共通アップリンク上の追加の仮想 NIC と ACI ファブリックへのテナント ネットワーキングを使用するか、または別途の物理インターフェイスを介してサーバに接続することができます。

OpenStack と ACI 構造のマッピング

Cisco ACI はポリシー モデルを使用して、ファブリックに接続されたエンドポイント間のネットワーク接続を可能にします。OpenStack Neutron は従来型のレイヤ 2 とレイヤ 3 のネットワーキングの概念を使用して、ネットワーキング接続を定義します。OpFlex ML2 ドライバは必要な ACI ポリシー モデル構造に Neutron ネットワーキング要件を変換して、必要な接続を実現します。次

の表に、OpenStack Neutron 構造と、それらの作成時に設定される対応 APIC ポリシー オブジェクトを示します。

OpenStack オブジェクト	APIC オブジェクト
Neutron インスタンス	ACI テナント、VMM ドメイン
テナント/プロジェクト	アプリケーション プロファイルまたは個別の ACI テナント
テナント ネットワーク	エンドポイント グループ + ブリッジ ドメイン
Subnet	Subnet
セキュリティ グループ/ルール	対象外（Linux iptables ルールはホスト単位で維持される）
ルータ	コントラクト + EPG + ブリッジ ドメイン
外部ネットワーク	レイヤ 3 Out/外部 EPG

デフォルトでは、OpFlex ML2 ドライバは OpenStack Neutron のインスタンス全体を単一の ACI テナントに関連付け、`/etc/neutron/plugins/ml2/ml2_conf_cisco_apic.ini` ファイルの `apic_system_id` 設定に従ってこのテナントに名前を付けます。これにより、ACI 管理者はファブリックに接続されたクラウド インスタンスそれぞれを単一のエンティティとして管理することができ、複数のシステムに使用するファブリックの APIC に多くの ACI テナントを生成しません。このモードでは、異なるアプリケーション プロファイルとして APIC に個別の OpenStack テナントが定義されます。

また、新しい ACI テナントを各 OpenStack テナントに作成するようにシステムに通知する `single_tenant_mode = False` 設定を使用して、インストール時に `ml2_conf_cisco_apic.ini` ファイルに設定できる代替オプションもあります。これにより、OpenStack テナントと ACI テナントは 1:1 の関係になり、各 OpenStack テナントに `convention_<apic_system_id>_<openstack tenant name>` に従って名前が付けられた ACI テナントが生成されます。マルチテナント モードを使用する場合は、システムが正しく機能するように、値 `apic_name_mapping = use_uuid` も `ml2_conf_cisco_apic.ini` ファイル内に設定する必要があります。

OpFlex NAT の動作

OpFlex ML2 ドライバ ソフトウェアは、OpenStack の各コンピューティング ノードでローカル OVS インスタンスを使用し、ネットワーク アドレス変換（NAT）機能を分散方式でサポートできるようにします。この分散方式のアプローチによってソリューション全体の可用性が向上し、リファレンス実装で使用される Neutron サーバ L3 エージェントからの NAT の中央処理が軽減されます。

NAT に必要な IP サブネット

OpFlex ML2 ドライバで外部ネットワークの機能をフルに活用するには、3 つの異なる IP サブネットが必要です。これは、これらの機能に通常は単一の外部サブネットを使用するデフォルトの Neutron 外部ネットワークの動作とは異なるアプローチです。

- リンク サブネット：このサブネットは、ファブリック外の外部ネクストホップ ルータへの実際の物理接続を表します。設定に応じて、これはルーテッドインターフェイス、サブインターフェイス、または SVI に割り当てられます。
- 送信元 NAT サブネット：OpenStack の送信元 NAT または SNAT という用語は、外部ネットワークのアドレスを共有することによって VM インスタンスにクラウド外の宛先との接続を許可することを説明するために使用されています。複数の VM による外部のルーティング可能な IP アドレスの共有を許可するポート アドレス変換（PAT）にこのサブネットが使用されます。単一の IP アドレスを各コンピューティング ノードに割り当て、一意のセッション トラフィックの維持にレイヤ 4 のポート番号操作を使用します。
- フローティング IP サブネット：OpenStack でのフローティング IP という用語は、VM インスタンスが異なるスタティック NAT アドレスを要求して、クラウド外からの VM へのインバウンド接続をサポートできるときに使用されます。フローティング IP サブネットは、OpenStack 内で Neutron 外部ネットワーク エンティティに割り当てられるサブネットです。

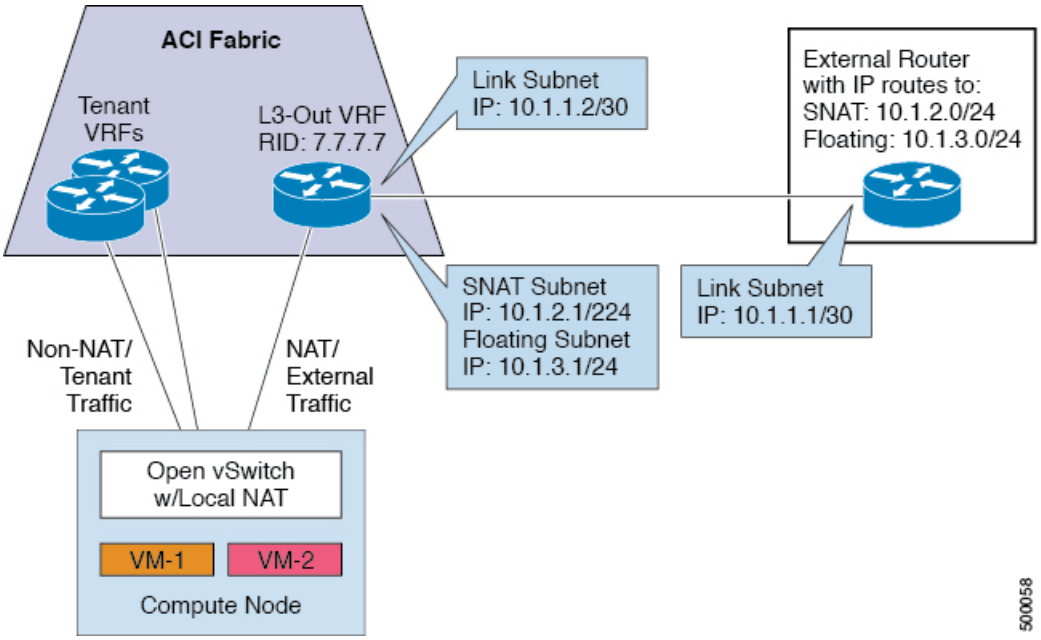
  クラウドから出力されるトラフィックは、SNAT サブネットかフローティング サブネットのいずれかの送信元 IP アドレスを伝送します。リターン トラフィックが OpenStack へ戻る経路を見つけられるように、動的にルーティングされたプロトコルかスタティック設定のいずれかを通じて、これらのサブネットに戻るルートが ACI の外部のルーティング ホップに必要です。

OVS NAT と外部ルーティング

コンピューティング ノードのローカル OVS で実行されている NAT 機能自体では、ACI ファブリック内の物理スイッチで外部ネクストホップ ルータとの間の外部トラフィックのルーティングのみが必要になります。この外部ルーティングは、レイヤ 3 Out に関連付けられた Virtual Routing and Forwarding（VRF）インスタンスを通じて処理されます。この L3-Out VRF には、外部ネクストホップ ルータへの物理リンクに関連付けられたインターフェイスがあります。また、この同じ VRF には、割り当てられた送信元 NAT サブネットの IP アドレスのインターフェイスと、フローティング IP サブネットもあります。さらに、この VRF には、ルーティング プロトコルの相互作

用のためのループバック インターフェイスも存在します。次の図に、この NAT アプローチをサポートするサブネット アーキテクチャを示します。

図 *5*： *OpFlex* ローカル *OVS* の *NAT* サブネットのアーキテクチャ



OpenStack Neutron の外部ネットワークに関連付けられた L3-Out VRF はコンピューティング ホスト上で OVS を出力する NAT トラフィックを処理します。非 NAT トラフィックは、VM インスタンスの OpenStack テナントとプロジェクトの関連付けに基づいてテナント VRF によって処理されます。

最適化された DHCP とメタデータ プロキシの動作

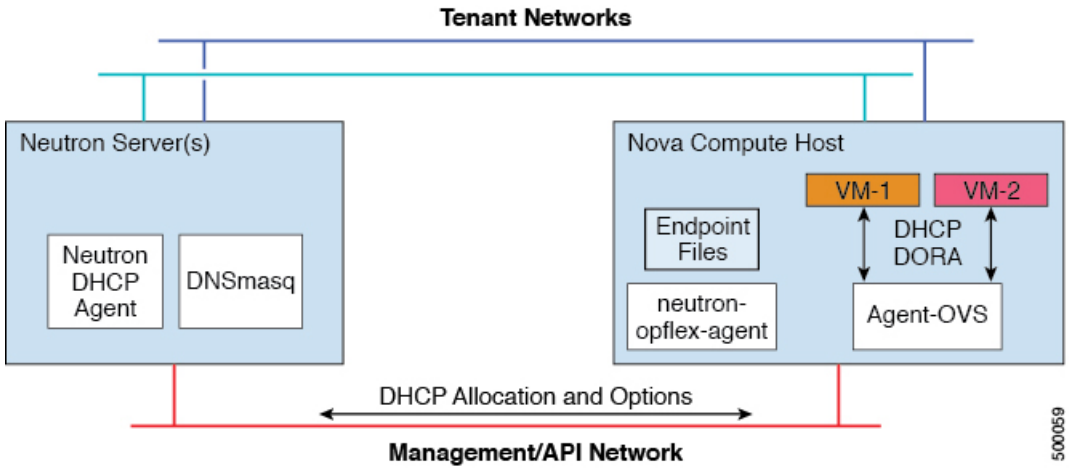
OpFlex ML2 ドライバ ソフトウェア スタックは最適化されたトラフィック フローと分散処理を実現し、DHCP とメタデータ プロキシ サービスを VM インスタンスに提供します。これらのサービスは、可能な限り多くの処理とパケット トラフィックをコンピューティング ホストにローカルに保持するように設計されています。分散要素は集中型の機能と通信し、システムの一貫性を確保します。

最適化された DHCP サービス

OpenStack Neutron のリファレンス アーキテクチャでは、Neutron サーバ 上で実行する `neutron-dhcp-agent` サービスを利用して、OpenStack テナント ネットワーク上で DM インスタンスへのすべての DHCP 通信を実現します。`neutron-dhcp-agent` は、IP アドレス管理を集中的に実行するとともに、DHCPディスカバリ、オファー、リクエスト、および確認応答（DORA）機能の各 VM インスタンスと通信します。

一方、OpFlex の最適化された DHCP アプローチでは、agent-ovs サービスを介してすべての DORA サービスをコンピュータ上でローカルに提供します。分散サービスは管理ネットワークで Neutron サーバへの通信を行い、IP アドレッシングと DHCP オプションを割り当てます。このアーキテクチャは、コンピューティング ホスト自体に DHCP リリースをローカルに発行するために必要な大量のパケット トラフィックを保持する一方で、Neutron サーバからのこの相互作用の処理も軽減します。次の図に、この DHCP アーキテクチャを示します。

図 *6*： *OpFlex* ベースの *DHCP* アーキテクチャ



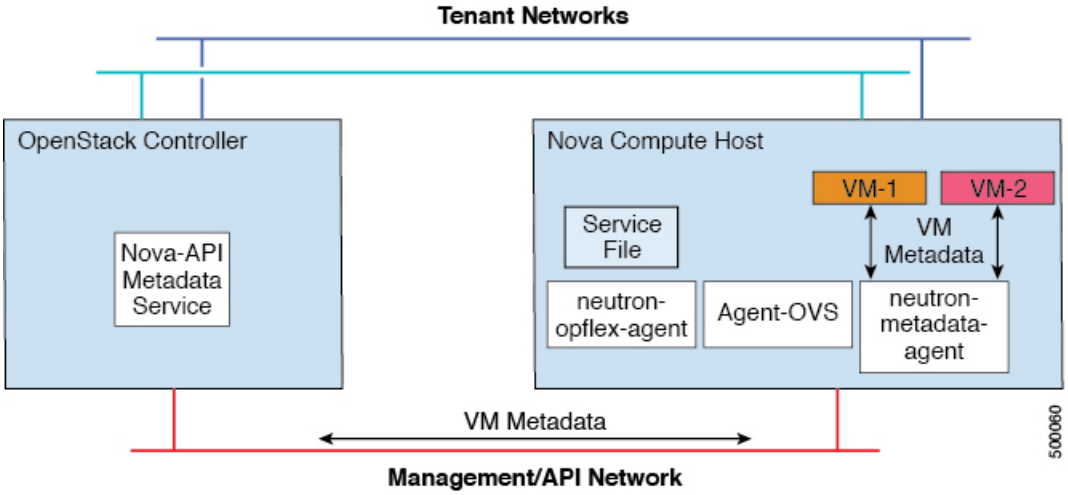
最適化された メタデータ サービス

VM インスタンスへのメタデータ配信用の OpenStack Neutron のリファレンス アーキテクチャでは、プロキシ サービスを Neutron サーバ上で集中的に実行します。このプロキシ サービスでは Nova API のインスタンス情報を検索して HTTP ヘッダーを追加し、メタデータ要求を Nova メタデータ サービスにリダイレクトします。VM インスタンスからのメタデータ要求は OpenStack テナント ネットワーク上で送信されます。

一方、OpFlex の最適化されたメタデータ プロキシのアプローチでは、各コンピューティング ホスト上で実行する分散型のメタデータ プロキシ インスタンスを使用してメタデータを配信します。`agent-ovs` サービスは OpFlex サービス ファイルを読み取り、メタデータ サービス要求をローカルの `neutron-metadata-agent` に送信するように OVS でフローをプログラミングします。このローカル エージェントはコンピューティング ホスト上の個別の Linux ネームスペースで動作します。次にメタデータ プロキシ機能は OpenStack コントローラ上で管理ネットワークを介して実行する Nova-API と Nova メタデータ サービスにアクセスして、VM 固有のメタデータを各

VM インスタンスに配信します。次の図に、このメタデータ プロキシ アーキテクチャを示します。

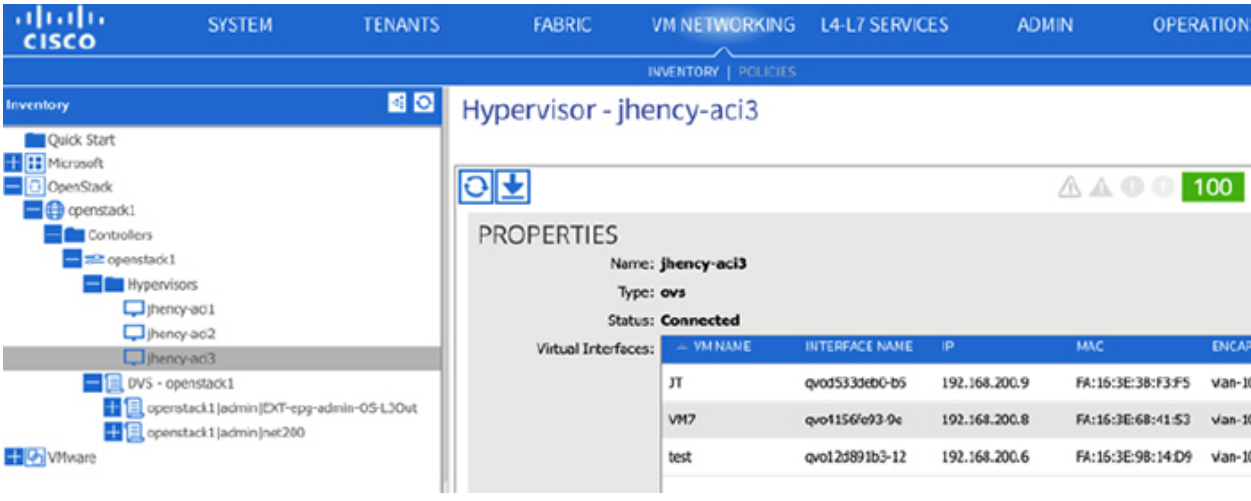
図 *7* : ***OpFlex*** ベースのメタデータ プロキシ アーキテクチャ



APIC OpenStack VMM の統合

Cisco ACI は、OpenStack などの複数の Virtual Machine Manager（VMM）システムとの統合をサポートします。この統合により、各ノードのすべての VM インスタンスの詳細なリストと学習されている各ポートの仮想インターフェイス情報を含めて、OpenStack のコンピューティング ノードから APIC を直接確認できるようになります。次の図に、OpenStack のハイパーバイザの VM ネットワーキングのビューを示します。

図 *8* : ***APIC VM*** ネットワークのハイパーバイザのビュー



また、APIC Webインターフェイスの VM ネットワーク セクションも、分散型仮想スイッチ（DVS）別のビューを提供します。各 DVS は、複数のコンピューティング ノードにわたって分散している可能性がある OpenStack ネットワークに対応します。このリストには、各コンピューティング ノードと ACI リーフのどこで VM インスタンスが接続されているかの詳細が含まれます。このリストには並べ替え機能とフィルタリング機能が備わっており、IP または MAC アドレスによって VM を検出できます。次の表に OpenStack DVS インスタンスの VM ネットワーキングのビューの例を示します。

図 *9*： *APIC VM* ネットワーキング *DVS* のビュー