



# ダイナミック ロード バランシングの構成

- ダイナミック ロード バランシング (1 ページ)
- CloudScale スイッチのダイナミック ロード バランシング (4 ページ)
- Silicon One スイッチのダイナミック ロード バランシング (14 ページ)
- ダイナミック ロード バランシングの設定 (26 ページ)
- ダイナミック ロード バランシングの構成の確認 (27 ページ)
- ダイナミック ロード バランシングのトラブルシュート (29 ページ)

## ダイナミック ロード バランシング

ダイナミック ロード バランシング (DLB) は、高度でインテリジェントなハッシュメカニズムであり、

- 従来の ECMP 転送を強化、
- リンクの負荷を考慮してトラフィック分散を最適化、そして、
- 十分に活用されていないリンクを介してトラフィックをダイナミックに誘導します。

これはIP レイヤ (OSI モデルのレイヤ3) で実行されるもので、最新のネットワーキングハードウェアの多くに実装されています。これには、Nexus 9000 シリーズ スイッチなども含まれます。

ECMP は、ネットワーク内の任意の 2 ポイント間を流れるトラフィックに、並列の複数のパスを許可することで、アプリケーションで使用可能な帯域幅を増やすために使用されます。複数の等コストパスを持つ接続先にパケットを転送する必要がある場合、ルータはハッシュアルゴリズムを使用して、そのパケットに使用する経路を決定します。このアルゴリズムでは、通常、送信元と宛先の IP アドレス、送信元と宛先のポート番号、場合によってはプロトコルタイプなどのパラメータが考慮されます。

従来のロードバランシングでは、ネットワークトポロジが変更されたり、ネットワーク管理者が手動で再構成したりしない限り、特定の IP フローに対して選択されたパスが時間の経過とともに変化することはありません。対照的に、レイヤ3 ECMP のダイナミック ロード バランシングでは、ネットワークの現在の状態に応じてパスの選択を変更できるようになっています。ルータまたはスイッチは、各パスのトラフィック負荷をモニターし、リンク使用率が最も低い

パスを選択して、使用可能なすべてのパスにトラフィックをより適切に分散できます。したがって、サポートされている Nexus 9000 スイッチのレイヤ 3 ECMP DLB 機能を使用すると、ネットワーク内の複数の等コストパスにトラフィックを効率的に分配することができます。

レイヤ 3 ECMP DLB は、人工知能と機械学習 (AI/ML) トレーニングネットワークのバックエンドで使用される、リーフ/スパイン型アーキテクチャを備えたイーサネット上の RDMA (RoCE) とともにサポートされます。DLB を備えたファブリックは、PFC とともに ECN と組み合わせると、使用率の向上、低遅延、および損失のないファブリックとして最適なネットワーク動作を実現します。

## 機能

ダイナミック ロード バランシングのいくつかの重要な機能は次のとおりです。

- 静的 ECMP ロード バランシングに関する従来のハッシュエントロピーの問題を回避し、
- 利用可能なネットワーク パスを最大限に使用し、
- すべてのパスにトラフィックを均等に展開させることで、輻輳を最小限に抑え、
- インフラストラクチャの追加や特殊なインフラストラクチャを必要とせずに、ネットワーク全体のパフォーマンスを向上させ、そして
- リンクまたはノード障害が発生した場合に高速コンバージェンスと冗長性を提供します

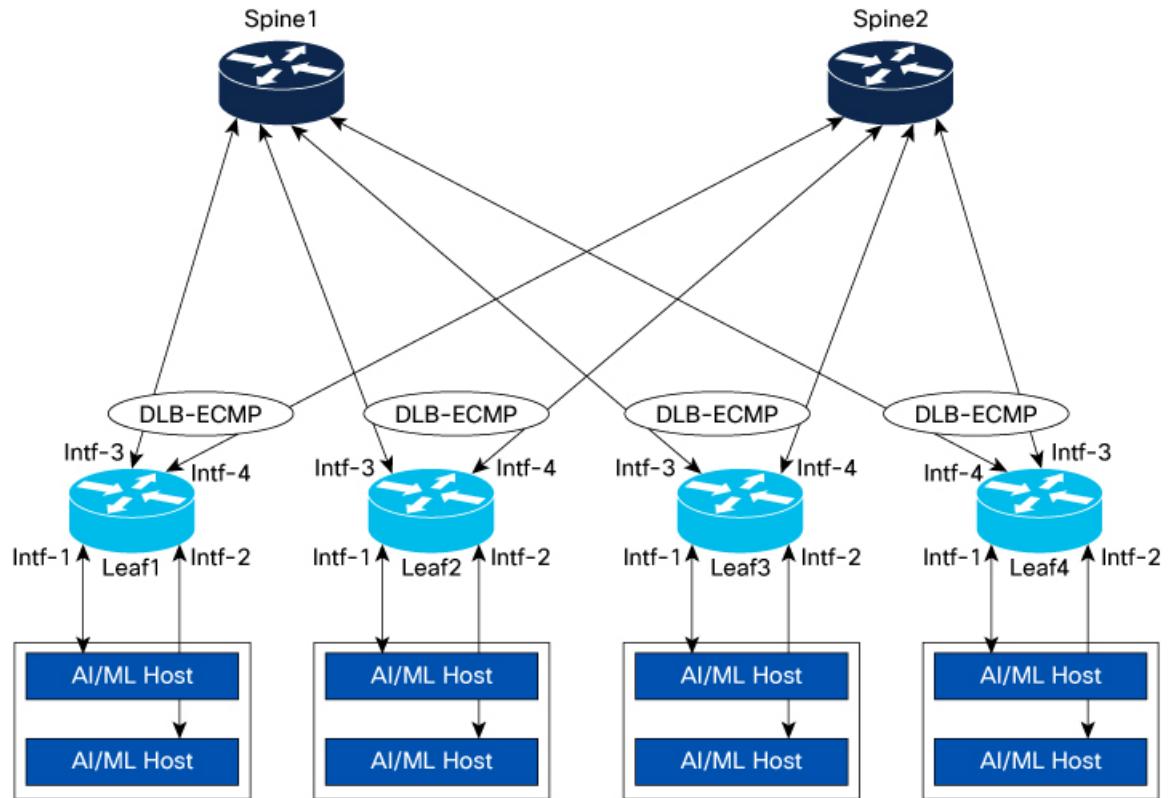
## AI/ML ネットワークでの DLB トポロジの仕組み

ダイナミック ロード バランシング (DLB) トポロジは、AI/ML トレーニング ネットワークで役立ちます。これらのネットワークは、図に示すスパインリーフアーキテクチャを使用しています。

このトポロジでは、AI および ML ホスト (サーバ) は、リーフスイッチのインターフェイス 1 (Intf-1) とインターフェイス 2 (Intf-2) に接続されています。リーフスイッチの Intf-3 と Intf-4 は、2 つのスパイン (スパイン 1 とスパイン 2) に接続されています。AI/ML ホスト間でデータ (トレーニングデータなど) を同期している間、トレーニングデータは、スパインリーフファブリックを介してすべてのホスト間で転送されます。

**process\_workflow**

図 1: DLB トポロジ

**result**

リーフスイッチは複数のリンクでスパインに接続されているため、ECMPは複数のリンク間でトラフィックを負荷分散するために使用されます。AI/MLトレーニングネットワークは、従来のネットワークと比較して、一意の5タプルIPフィールドを持つトラフィックフローが少ない。このようなフローの数が限られているため、従来のECMPでは偏りの問題が発生します。これは、冗長パスの使用が最適化されないという意味です。その結果、一部のリンクまたはインターフェイスでオーバーサブスクリプションが発生する可能性があります。これにより、ファブリック全体のスループットが低下する可能性があります。

ECMP DLB機能では、すべてのリンクの適切な使用を保証することにより、使用率がない、または不足しているなどのリンク使用率の問題を解決します。ECMPグループの一部であるすべてのポートでDLBを有効にする場合、新しいフローごとに、使用可能なリンクの中からTxリンク使用率が最低のリンクが選択されます。画像では、Intf-3とIntf-4でDLBが有効になっています。Intf-3が完全に使用されて、新しいフローが到着した場合、Intf-4が選択されます。従来のECMPでは、オーバーサブスクリプションが生じていても、Intf-3が選択される可能性があります。

ECMP DLBは静的ピン接続もサポートします。これにより、ユーザーは特定の送信元ポートからのトラフィックを常に特定のDLB対応出力ポートに送信できます。この画像では、Intf-3とIntf-4がメンバーであるDLB ECMPグループを取得するトラフィックの場合、ユーザーは常に

## NX-OS リリースのダイナミック ロード バランシング機能

Intf-1 からトラフィックをピン接続して常に Intf-3 を取得し、Intf-2 を使用して常に Intf-4 を取得できます。

## NX-OS リリースのダイナミック ロード バランシング機能

Cisco NX-OS リリース 10.5(1)F から、レイヤ3 ECMP ダイナミックロードバランシング (DLB) 機能は、発信リンクの現在の使用状況に応じて、トラフィックを効率的にロードバランシングするためのサポートを提供します。この機能のサポートは、表に示すように、さまざまな NX-OS リリースを介してさまざまな Nexus スイッチで提供されます。

スイッチ	リリース
9300-FX3、-GX、-GX2、-H1、-H2R TOR などの Nexus CloudScale スイッチ	Cisco NX-OS リリース 10.5(1)F
Silicon One スイッチ : N93C64E-SG2-Q および N9364E-SG2-O スイッチのみ	Cisco NX-OS リリース 10.5(3)F

## Nexus スイッチのダイナミック ロード バランシング

DLB 機能の概念と構成は、CloudScale スイッチと Silicon One スイッチで異なります。詳細については、次の 2 つのセクションを参照してください：

- [CloudScale スイッチのダイナミック ロード バランシング](#)
- [Silicon One スイッチのダイナミック ロード バランシング](#)

## CloudScale スイッチのダイナミック ロード バランシング

CloudScale スイッチの概念、ガイドライン、制限事項、および構成セクションは、Silicon One スイッチとは異なるため、このセクションで説明します。

## CloudScale スイッチのダイナミック ロード バランシングの主なコンセプト

このセクションでは、CloudScale スイッチで DLB を構成する前に知っておく必要がある主な概念について説明します。

### 高速リンク フェールオーバー

Nexus 9000 スイッチの DLB ECMP ロード バランシングの背景における高速リンク フェールオーバーとは、ネットワークが物理リンク障害に迅速に対応し、回復できるようにする機能のことです。ECMP グループで使用されるリンクに障害が発生した場合、高速リンク フェールオーバーにより、トラフィックは残りの動作可能なリンクにすぐにリダイレクトされます。リ

リンクのフェールオーバーはハードウェアによって検出され、残りのリンクから新しいリンクが自動的に選択されます。これはハードウェア層で実行されるため、コンバージェンスが高速化されます。

### ダイナミック使用率評価機能

Dynamic Rate Estimator (DRE) は、現在のリンク使用率を測定するためにハードウェアに実装されています。DLBにおけるDREの役割は、さまざまなリンクのトラフィックレートをリアルタイムで推定することです。このリアルタイム分析により、スイッチはトラフィックを分散する際に、より多くの情報に基づいた決定を行うことができ、単一の経路が過飽和にならないようにします。新しいフローを開始するとき、DLBはDREメトリックを使用して、DLB ECMP グループ内の複数の経路内で使用率が最も低いパスを決定します。

任意の時点で、DLB 対応インターフェイスは、リンクの使用率と構成された DRE しきい値に基づいて、レベル 1 からレベル 7 の DRE レベルのいずれかになります。レベル 1 は最も低い使用率を示し、レベル 7 は最も高い使用率を示します。DLB は判断する際、常に最も低い DRE レベルのリンクを選択します。複数のリンクが、同じ最も低い DRE レベルである場合、それらの中のリンクの 1 つがランダムに選択されます。詳細については、「[リンク使用率レベルの計算](#)」を参照してください。

### モード

レイヤ 3 ECMP ダイナミック ロード バランシングは、グローバル構成で次のいずれかのモードをサポートします。

- フローレット ロード バランシング (FLB) : このモードでは、DRE メトリックに基づいてフローレット レベルでロード バランシングを実行されます。これがデフォルトのモードです。
- パケット単位のロード バランシング (PLB) : このモードでは、フローレット レベルではなくパケット単位でロード バランシングの判断が行われます。

### Flowlet ロード バランシング

フローレットは、フローからのパケットのバーストであり、5 タプル (つまりパケットから選択されたフィールド) で識別されます。並べ替えを発生させることなく個別にルーティングできるように、十分に大きなギャップで区切られています。

フローレットは、DLB がフローレット モードで動作するときに使用されるトラフィックの単位です。フローレットごとに、ポート単位の DRE によって示される、Tx 使用率が最小のものが、最適な発信ポートとしてハードウェアによって選択されます。使用率がすべてのポートで同じ場合、ポートの 1 つがランダムに選択されます。

あるフローレット用にポートが選択されると、そのフローからの後続のすべてのパケットに同じポートが使用されます。新しいポートは、構成されたフローレットエージング時間を超えるパケット間ギャップがフローレットにある場合、または現在使用されているポートがダウンした場合にのみ、選択されます。

### パケット単位のロード バランシング

パケット単位のロード バランシング (PLB) は、エンドポイント (スマート NIC など) でパケットの並べ替えが可能なシナリオで使用できます。このモードは、DLB ECMP 内の使用可能なリンク全体にトラフィックを分散させることにより、ネットワークの輻輳を軽減します。フロー内のパケットごとに、新しい出力ポートが選択されます。そのため、同じフローからのパケットが複数のパスを介して送信され、パケットの順序が変更される可能性があります。ポート選択プロセスでは DRE が使用されます。つまり、どのパケットでも、DRE メトリックが最小のポートが選択されます。DRE がすべてのポートで同じである場合、ポートの1つがランダムに選択されます。

### 適応型ルーティング

Nexus スイッチが適応型ルーティングをサポートする NVIDIA NIC とピアリングし、NX-OS スイッチで **hardware profile spectrum-x** を有効にすると、適応型ルーティング機能が NIC で有効になります。

**spectrum-x** コマンドは、NVIDIA の適応型ルーティング機能を有効にします。これには、トラフィックの状態、輻輳、またはネットワクト プロジの変更に基づいてネットワークルーティング パスを動的に調整することが含まれます。out-of-order パケットの処理は、このモードで有効化される NIC の重要な機能の1つです。NX-OS スイッチのパケット単位のロード バランシング (PLB) をこれと組み合わせると、パフォーマンスが向上します。

### 静的ピン接続

静的ピン接続は、DLB でサポートされています。静的ピン接続では、送信元ポートは、DLB 対応の ECMP グループの一部である宛先ポートにピン止めされます。このポートがこのフローに使用される DLB ECMP グループの一部である場合、この送信元ポートからのすべてのトラフィックは、ピン止めされた宛先ポートに送信されます。前面パネルのポート（ブレークアウトポートを含む）は、静的ピン接続の送信元インターフェイスとして使用できます。宛先インターフェイスは、DLB インターフェイスリストに含まれている必要があります。

静的ピン接続が有効になっている場合、静的ピン接続は DLB DRE ベースのポート選択を上書きします。

DLB ポートが静的ピン接続の宛先ポートとして使用されている場合、そのポートの静的ピン接続構成が削除されない限り、このポートを DLB インターフェイスリストから削除することはできません。



(注) 静的ピン接続または PLB モードについては、どちらか一方のみを有効にできます。両方を同時にイネーブルにはできません。



(注) Cisco NX-OS リリース 10.5(2)F 以降、システムは最大 512 の静的ピン接続ペアをサポートします。

# CloudScale スイッチのダイナミック ロード バランシングに関するガイドラインと制限事項

CloudScale スイッチのレイヤ3 ダイナミック ロード バランシングのガイドラインと制限事項は、次の4つのサブセクションに分類されます。

- ECMP グループ、
- 機能サポート、
- ポート、および
- DLB パラメータです。

## ECMP グループ

- DLB を有効にするかどうかの判断は、次の3つの条件に基づき、ECMP グループの作成時に行う必要があります。
  - ECMP グループのすべてのメンバーが、DLB 対応インターフェイスリストに含まれています。ECMP グループに DLB インターフェイスリストにないメンバーが1つ以上ある場合、その ECMP グループには通常の ECMP を使用されます。
  - ECMP グループのメンバーは、レイヤー3 インターフェイスのみです。ブレークアウトポート、サブインターフェイス、SVI、およびポートチャネルを DLB ECMP グループのメンバーにすることはできません。
  - ECMP は重み付け ECMP またはレジリエント ECMP ではありません。
- レジリエント ECMP 機能と DLB 機能を同時に有効にすることはできません。レジリエント ECMP の詳細については、[Cisco Nexus 9000 シリーズ NX-OS インターフェイス構成ガイド](#)を参照してください。
- 重み付け ECMP グループの場合、DLB は適用されません。加重 ECMP または、UCMP の詳細については、[BGP 経由不等コストマルチパス \(UCMP\)](#) を参照します。
- show routing hash コマンドは、DLB ECMP グループを使用するルートでは機能しません。これは、DLB が有効になっている場合、ポートの選択は静的ハッシュではなくリンク使用率に基づいて動的に行われるためです。
- DLB ECMP スケールに達した場合、または何らかの条件により DLB を有効にできない場合は、DLB なしの通常の ECMP が使用されます。



(注)

DLB 対応 ECMP グループのメンバーポートの1つがダウンすると、ハードウェアは、ポートを通してのトラフィックの送信をすぐに停止します。これにより、リンク障害時のトラフィック損失を最小限に抑えることができます。

## 機能のサポート

- この機能は、以下ののみサポートされています。
  - レイヤ 3 物理インターフェイス
  - IP ルーティングファブリックおよび VXLAN ファブリック
  - 9300-FX3、-GX、-GX2、H1 および H2R TOR プラットフォーム。
- この機能は、ラインカード拡張モジュール (LEM) および N9K-C9408 を搭載した TOR ではサポートされません。
- 出力インターフェイスで構成された出力アクセリストポリシー、出力 QoS ポリシー、および TX SPAN は、ECMP DLB を使用するフローには適用されません。
- MPLS/GRE トンネルは DLB ECMP を使用せず、通常の ECMP にフォールバックします。
- この機能を使用する場合、DLB ECMP スケールが大きい場合は、システム pic-core オプションを使用することを推奨します。システム pic-core オプションを使用すると、スイッチのリロードが必要です。詳細については、「[BGP PIC コアの構成](#)」を参照します。
- DLB は、ポリシーベースルーティング (PBR) ロジックが適用されるトラフィックフローには適用されません。これらのフローは、通常の ECMP 機能を使用します。
- MTU は、DLB フローで使用されるパケットの最大サイズに基づいて、すべての DLB 対応インターフェイスで構成する必要があります。そうしなかった場合、出力がドロップするとトラフィックは出力インターフェイスでドロップされます。
- Only IPv4 と IPv6 ユニキャストトラフィックのみがサポートされています。
- Cisco NX-OS リリース 10.6(6)F 以降、AI/ML ライセンスの機能拡張の一環として、適応型ルーティング機能は CloudScale スイッチでサポートされます。関連するライセンスが必要です。さらに、DLB 機能には、10.6(1)F 以降での Advantage ライセンスが必要です。DLB が有効になっていて、有効なインターフェイスリストを持っている場合、ライセンスを消費します。ライセンスの詳細については、「[Cisco NX-OS ライセンスオプションガイド](#)」を参照してください。

## ポート

- ブレークアウトポート、ポートチャネル、SVI、ポートチャネルメンバー、またはサブインターフェイスを DLB 対応インターフェイスリストに含めることはできません。
- 最大 63 個の物理ポートを DLB インターフェイスリストに含めることができます。

## DLB パラメータ

このセクションでは、MAC、エージング、DRE しきい値、モード、静的ピン接続などの DLB 関連パラメータのガイドラインを示します。

- DLB インターフェイスリストを初めて構成する場合、または変更した場合には、スイッチをリロードして構成を有効にする必要があります。
- フローレットエージングタイムは、ファブリックでのラウンドトリップ時間に基づいて選択します。そうしなかった場合、フローの順序が変わる可能性があります。
- DLB 構成のすべての DLB 関連パラメータは、有効な適用済み DLB インターフェイスリストがある場合にのみ、ハードウェアでプログラムされます。
- MAC、エージング、モード、または DRE しきい値構成のいずれかを削除すると、すべてのパラメータがデフォルト値に設定されます。
- ポートを DLB インターフェイスリストに追加した後、ポートがブレークアウトポートに変更されるか、PO の一部になるように追加されるか、このインターフェイスでサブインターフェイスが作成されるかすると、DLB は、ポートを含む ECMP グループに対して有効ではなくなります。ユーザーは、DLB インターフェイスリストからポートを削除する必要があります。
- DRE しきい値を変更すると、DLB 対応フローのトラフィックに一時的な影響が及ぶ可能性があります。これは中断を伴うトリガーです。
- DLB MAC 構成は、ファブリック内のすべてのノードで同一である必要があります。これらのフローを受信するファブリック内の接続ノードで DLB MAC 構成を変更せずに、スイッチでその構成を変更すると、トラフィックはドロップされます。
- 静的ピン接続モードと、パケットごとの DLB モードは、同時にサポートできません。
- ブレークアウトポートは、静的ピン接続の送信元インターフェイスの一部にすることができます。ただし、サブインターフェイスとポートチャネルを、静的ピン接続の送信元インターフェイスの一部にすることはできません。

## CloudScale スイッチのダイナミック ロード バランシングの構成

CloudScale スイッチでレイヤ3 ダイナミック ロード バランシングを構成するには、このセクションに記載されているコマンドを **hardware profile dlb** サブモードで実行します。

### 始める前に

**configure terminal** コマンドを使用して、グローバル コンフィギュレーション モードになっていることを確認します。

### 手順

**ステップ1** **hardware profile dlb** コマンドを使用して、ハードウェア プロファイルのダイナミック ロード バランシング モードを開始します。このコマンドでは、リロードする必要があります。

例：

```
switch(config)# hardware profile dlb
```

**ステップ2** **dlb-interface <interface\_range>** コマンドを使用して、DLB を有効にするインターフェイスのリストを指定します。カンマ区切りで、インターフェイスを追加します。このリストは動的には変更できません。インターフェイスリストを有効にするには、スイッチのリロードが必要です。

(注)

- ・インターフェイス リストを変更した場合は、構成を有効にするためにリロードが必要です。
- ・現在の適用済みリストを確認するには、**show hardware profile dlb** コマンドを実行します。
- ・インターフェイス リストへの部分的な追加や削除はサポートされていません。構成は、新しく提供されたインターフェイスに置き換えられます。

例 :

```
switch(config-dlb) # dlb-interface Eth1/5,Eth1/7,Eth1/17,Eth1/21,Eth1/26
```

**ステップ3** (任意) **dre-thresholds [level-1 percentage\_1 | level-2 percentage\_2 | level-3 percentage\_3 | level-4 percentage\_4 | level-5 percentage\_5 | level-6 percentage\_6 | level-7 percentage\_7]** コマンドを DLB モードに入力して、DRE レベルをレベル 1 からレベル 7 に定義します。レベルごとに構成される値は、前のレベルからのポート帯域幅の使用率の範囲です。指定するすべてのレベルの合計は、100 にする必要があります。DRE しきい値レベルを構成しなかった場合は、デフォルト値として、30、20、15、10、10、10、および5が使用されます。

DRE レベルのリンク使用率の詳細については、[リンク使用率レベルの計算](#)を参照してください。

例 :

```
switch(config-dlb) # dre-thresholds level-1 15 level-2 20 level-3 30 level-4 15 level-5 10 level-6 5 level-7 5
```

**ステップ4** (任意) DLB モードで **flowlet-aging usec** コマンドを使用して、usec でフローレットエージングタイムを構成します。デフォルトは 500 マイクロ秒で、最大値は 2 秒つまり 2000000 マイクロ秒です。

(注)

フローレットのエージング時間は慎重に選択してください。不用意に選択すると、フローの順序が変わってしまう可能性があります。

例 :

```
switch(config-dlb) # flowlet-aging 600
```

**ステップ5** (任意) DLB MAC アドレスを構成するには、DLB モードで **mac-address macaddr** コマンドを使用します。このアドレスは、DLB を使用するすべてのフローのネクストホップ MAC アドレスとして使用されます。この DLB MAC は、出力インターフェイスの学習されたネクストホップ MAC アドレスを変更するものとして、DLB フローの宛先 MAC を書き換えるために使用されます。

ガイドラインおよび制約事項は次のとおりです :

- ・このコマンドを構成しなかった場合、機能の初期化中に使用されるデフォルトの DLB MAC アドレス (00:CC:CC:CC:CC:CC) が、デフォルトの DMAC として使用されます。
- ・DLB のMAC アドレスを構成すると、デフォルトの MAC は、新しい構成可能 MAC アドレスで置き換えられます。
- ・このスイッチで、この DLB MAC を宛先 MAC として受信されたすべてのパケットは、ルーテッド パケットとして扱われます。

- この設定を適用する場合は、ファブリック内の他のすべてのノードが同じ DLB MAC で構成されていることを確認してください。
- db-interface リストが適用されていない場合、DLB MAC を追加のルータ MAC として使用することはできません。
- ブロードキャストおよびマルチキャスト MAC アドレスを DLB MAC アドレスとして設定することはできません。

例：

```
switch(config-dlb)# mac-address aa:bb:cc:dd:ee:ff
```

**ステップ6** (任意) **mode [flowlet | per-packet]** コマンドを DLB モードで使用して、フローレットまたはパケットごとの DLB モードを有効にします。デフォルトのモードはフローレットです。

(注)

パケット単位モードの場合、静的ピン接続を有効にすることはできません。

例：

```
switch(config-dlb)# mode flowlet
```

**ステップ7** (任意) **static-pinning** コマンドを DLB モードで使用して、静的ピン接続機能を設定します。

(注)

静的ピン接続の場合、パケット単位モードを有効にすることはできません。

例：

```
switch(config-dlb)# static-pinning
```

**ステップ8** (任意) 静的ピン接続用に送信元インターフェイスと宛先インターフェイスを構成するには **source source physical interface destination destination physical interface** コマンドを使用します。ただし、これらは物理インターフェイスのみでなければなりません。つまり、前面パネルのイーサネットインターフェイスです。SVI、ポートチャネル、またはサブインターフェイスを送信元または接続先インターフェイスにすることはできません。

ガイドラインおよび制約事項は次のとおりです：

- 接続先インターフェイスは、DLB が適用または構成されたインターフェイスリストの一部である必要がありますが、送信元インターフェイスをこのリストの一部にすることはできません。
- Cisco NX-OS リリース 10.5(2)F 以降では、DLB が適用または構成された interface-list のインターフェイスは、静的ピン接続の送信元インターフェイスとしても使用できます。
- 2 つの構成で同じ送信元インターフェイスが使用されている場合、最初の接続先インターフェイスは 2 番目の接続先インターフェイスで置き換えられます。送信元インターフェイスは同じだからです。
- ブレークアウトポートは、送信元インターフェイスとして構成できます。
- ポートでブレークアウトまたはブレークアウト操作が実行されていない場合、ユーザーは DLB または静的ピン接続構成を更新する必要があります。

## リンク使用率レベルの計算

- 静的ピン接続の接続先インターフェイスとして構成されているインターフェイスは、DLB インターフェイスリストから削除できません。削除するには、最初に静的ピン接続構成を削除して、その後 DLB インターフェイスリストからインターフェイスを削除します。

例：

```
switch(config-dlb-static-pinning)# source ethernet 1/1 destination ethernet 1/2
```

**ステップ 9** (任意) **hardware profile spectrum-x** コマンドを使用して、NVIDIA spectrum-x 適応型ルーティング機能を有効にします。このコマンドを実行すると、DLB 機能は**hardware profile db** での設定に基づいて実行されます。

(注)

**feature lldp** ブリッジを有効にしてください。詳細については、『Cisco Nexus 9000 シリーズ NX-OS システム管理コンフィギュレーションガイド』を参照してください。spectrum-x コマンドでは、リロードは必要ありません。

例：

```
switch(config)# hardware profile spectrum-x
```

---

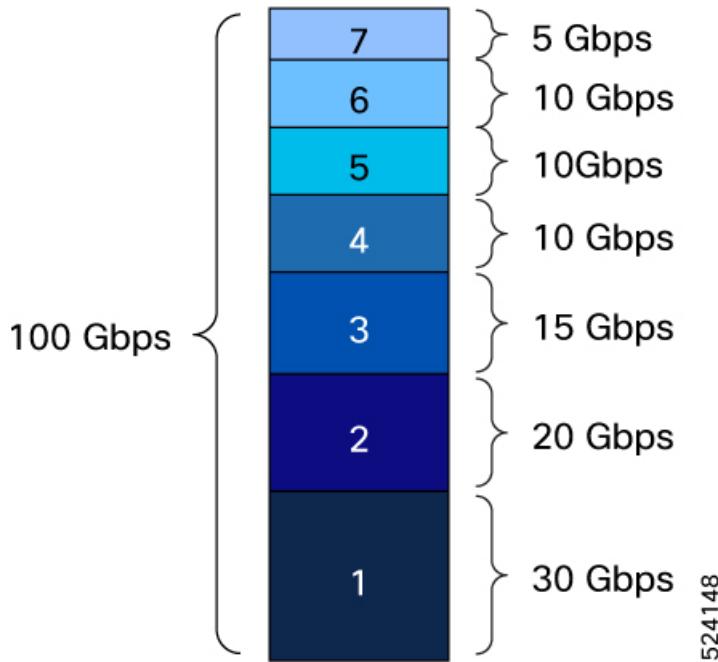
## リンク使用率レベルの計算

各レベルのリンク使用率は、次のように計算されます。

- 現在のレベル範囲の開始：以前のすべてのレベルに指定された % 値の合計。
- 現在のレベル範囲の終了：現在のレベル範囲の開始 + 現在のレベルの値。

次に例を示します。

Level-1 : 30、Level-2 : 20、Level-3 : 15、Level-4 : 10、Level-5 : 10、Level-6 : 10、Level-7 : 5。



- レベル 5 範囲の開始 : 75% ( $30 + 20 + 15 + 10$ )
- レベル 5 範囲の終了 : 85% ( $75 + 10$ )

#### DRE しきい値レベルのガイドライン

Dynamic Rate Estimator (DRE) のしきい値レベルに関するガイドラインと制限事項を次に示します。

- 上記のロジックは、いずれかのレベルがゼロ (0) のシナリオに適用されます。レベルを追加すると、そのレベルは、ゼロ以外の値を持つ前のレベルと同じ範囲になります。

次の例を参考してください。

Level-1 : 30、Level-2 : 0、Level-3 : 35、Level-4 : 10、Level-5 : 10、Level-6 : 10、Level-7 : 5。

- レベル 2 の範囲 : レベル 2 はレベル 1 と同じ、つまり 30% になります。

- 以前の値がすべてゼロの場合、現在のレベルは最初に指定されたゼロ以外のレベルになります。

次の例を参考してください。

Level-1 : 0、Level-2 : 0、Level-3 : 0、Level-4 : 0、Level-5 : 50、Level-6 : 30、Level-7 : 20。

- この場合、リンク使用率の開始レベルは Level-5 になります。

- すべてのレベルを指定してください。いくつかのレベルが指定されていない場合、それらはゼロと見なされます。

次の例を参考にしてください。

Level-1 : 50、Level-2 : 0、Level-3 : 0、Level-4 : 0、Level-5 : 0、Level-6 : 30、Level-7 : 20

- レベル 2、3、4、および 5 はゼロと見なされます。

## Silicon One スイッチのダイナミック ロード バランシング

Silicon One スイッチの概念、ガイドライン、制限事項、および構成セクションは CloudScale スイッチとは異なるため、このセクションで説明します。

## Silicon One スイッチのダイナミック ロード バランシングの主なコンセプト

このセクションでは、Silicon One スイッチで DLB を構成する前に知っておく必要がある主な概念について説明します。

### フローレット ロード バランシング

フローレットは、フローからのパケットのバーストであり、5 タプル（つまりパケットから選択されたフィールド）で識別されます。並べ替えを発生させることなく個別にルーティングできるように、十分に大きなギャップで区切られています。

DLB がフローレットモードで機能して、フローレット用にポートが選択されると、そのフローからの後続のすべてのパケットに同じポートが使用されます。これは、フローレット ロード バランシング (FLB) と呼ばれます。

新しいポートは、構成されたフローレットエージング時間を超えるパケット間ギャップがフローレットにある場合、または現在使用されているポートがダウンした場合にのみ、選択されます。新しいポートの選択は、各フローレット間隔の開始時に行われ、DLB ECMP グループのすべてのリンクでトラフィックを均等に分配できます。ポート選択プロセスでは、フローレット境界で使用可能なすべてのポートにパケットを均等に分散しようとします。ポートの選択は、リンク使用率に基づいているか、またはランダムに行われます。

### パケット別 ロード バランシング

パケット単位のロード バランシング (PLB) は、エンド ポイント（スマート NIC など）でパケットの並べ替えが可能なシナリオで使用できます。このモードは、DLB ECMP グループ内の利用可能なすべてのリンク全体にトラフィック パケットごとに分散し、トラフィックの分散に協力し、ネットワーク 輻輳を減少させます。このモードでは、ネットワーク 帯域幅と迅速なエンドツーエンド 遅延を最適に使用できます。フロー内のパケットごとに、新しい出力ポートが選択されます。そのため、同じフローからのパケットが複数のパスを介して送信され、それによりパケットの順序が変更される可能性があります。

ポート選択プロセスでは、使用可能なすべてのポートにパケットを均等に分散しようとします。ポートの選択は、リンク 使用率に基づいているか、またはランダムに行われます。

## ポリシー主導 DLB

ポリシー主導型ダイナミック ロード バランシングは、事前定義されたインターフェイス QoS ポリシーに基づいて入力トラフィック フローを動的に分類し、フローのロード バランシング モードを決定することにより、ネットワーク トラフィックを管理するために使用される手法です。

ポリシーは、ダイナミック ロード バランシングがいつ、どのように行われるかを定義するルールで構成されます。これらのルールは有効または無効にできるため、ネットワーク トラフィック の転送を柔軟に管理できます。トラフィックがフローレットまたはパケットごとのアクションで QoS ポリシールールと一致しない場合、ポリシーのオーバーライドが行われ、トラフィックは通常の ECMP を使用します。

ポリシー主導型 DLB モードは QoS ポリシーの一一致に依存します。サポートされるポリシー主導型 DLB には、次の 2 種類があります。

- **シンプルなポリシー主導 DLB :**

ポリシーは、フローレットまたはパケットごとの DLB 動作のいずれかのみを駆動します。QoS ポリシーの dlb モードは、**hardware profile dlb** 設定のモードと一致している必要があります。

- **混合モードでのポリシー主導型 DLB :**

この柔軟な動作モードでは、システムはフローレット転送とパケット単位の転送の両方を同時にサポートできます。このプロファイルの ECMP スケールは、512 ECMP グループから 256 ECMP グループに削減されます。

## モード

レイヤ 3 ECMP ダイナミック ロード バランシングは、Silicon One スイッチ上のグローバル構成でいずれかの 7 モードをサポートします。

- **フローレット ロード バランシング (FLB) :**

このモードでは、ポート ロードに基づいてフローレット レベルでロード バランシングを実行されます。DLB がイネーブルの場合、これがデフォルト モードです。

- **パケット単位 ロード バランシング (PLB) :**

このモードでは、フローレット レベルではなくパケット単位でロード バランシングの判断が行われます。このモードでは、すべてのメンバーポートで均等な使用率で最大のスループットが得られますが、パケットの順序が変更される可能性があります。

- **ポリシー主導型 フローレット :**

QoS **set dlb mode** オプションはフローレットである必要があります。モードがポリシーと一致しない場合、適用されたインターフェイスは err-disabled 状態になります。

- **ポリシー主導型 パケット単位 :**

QoS **set dlb mode** オプションはパケット単位にする必要があります。モードがポリシーと一致しない場合、適用されたインターフェイスは err-disabled 状態になります。

- **ポリシー主導型混合（デフォルトのecmp）：**

：QoS ポリシーがない場合、通常の ECMP ベースの転送が行われます。フローレットとパケットごとの動作の両方を推進するさまざまな QoS ポリシーを持つことができます。

- **ポリシー主導型混合デフォルトフローレット：**

デフォルトでは、トラフィックはフローレット DB 転送を使用します。QoS ポリシーは、パケット単位の転送をサポートするように動作を上書きできます。

- **ポリシー主導型混合デフォルト パケット単位：**

デフォルトでは、トラフィックはパケットごとの DB 転送を行います。QoS ポリシーは、フローレットベースの DB 転送をサポートするように動作をオーバーライドできます。

トラフィックとスケールの要件に基づいてモードを選択して設定します。すべてのポートで着信するすべてのトラフィックを DB にするデフォルトの動作モードでは、FLB または PLB を使用します。選択的なトラフィック、たとえば、RDMA-over Converged Ethernet (RoCE) ベースのトラフィックに DB が必要で、通常のトラフィックではない場合は、混合ポリシー駆動型 DB モードを使用します。

### 適応型ルーティング

Nexus スイッチが適応型ルーティングをサポートする NVIDIA NIC とピアリングし、NX-OS スイッチで **hardware profile spectrum-x** を有効にすると、適応型ルーティング機能が NIC で有効になります。

spectrum-x コマンドは、NVIDIA の適応型ルーティング機能を有効にします。これには、トラフィックの状態、輻輳、またはネットワークトポジの変更に基づいてネットワークルーティングパスを動的に調整することが含まれます。out-of-order パケットの処理は、このモードで有効化される NIC の重要な機能の 1 つです。NX-OS スイッチのパケット単位のロード バランシング (PLB) をこれと組み合わせると、パフォーマンスが向上します。

### 静的ピン接続

静的ピン接続は、DB でサポートされています。静的ピン接続では、送信元ポートは、DB 対応の ECMP グループの一部である宛先ポートにピン止めされます。このポートがこのフローに使用される DB ECMP グループの一部である場合、この送信元ポートからのすべてのトラフィックは、ピン止めされた宛先ポートに送信されます。前面パネルのポート（ブレークアウトポートを含む）は、静的ピン接続の送信元インターフェイスとして使用できます。宛先インターフェイスは、DB インターフェイスリストに含まれている必要があります。システムでは、最大 256 の静的ピン接続のペアがサポートされます。

DB ポートが静的ピン接続の宛先ポートとして使用されている場合、そのポートの静的ピン接続構成が削除されない限り、このポートを DB インターフェイスリストから削除することはできません。



(注) 静的ピン接続または PLB モードについては、どちらか一方のみを有効にできます。両方を同時にイネーブルにはできません。

## 高速リンク フェールオーバー

Silicon One スイッチの DLB ECMP ロード バランシングの背景における高速リンク フェールオーバーとは、ネットワークが物理リンク障害に迅速に対応し、回復できるようとする機能のことです。ECMP グループで使用されるリンクに障害が発生した場合、高速リンク フェールオーバーにより、トラフィックは残りの動作可能なリンクにすぐにリダイレクトされます。リンクのフェールオーバーはハードウェアによって検出され、残りのリンクから新しいリンクが自動的に選択されます。これはハードウェア層で行われるため、コンバージェンスが高速化されます。

## DLB を使用した階層型重み付け ECMP

スパインリーフトポロジでは、通常、トラフィックは両方のスパインに均等に分散されます。ただし、ファブリック内にリンク障害がある場合、入力リーフで対処しないと、スパインの1つで輻輳が発生する可能性があります。この問題は、重み付け ECMP (WCMP) を使用して入力リーフのスパインに到達するトラフィックを考慮してリンク障害を考慮することで解決できます。

階層型 WCMP は、1つの DLB ECMP グループへのリーフへの全帯域幅を持つスパインと、別の DLB ECMP グループへの部分的な帯域幅を持つスパインを選択する機能を提供します。DLB は、個々の DLB ECMP グループのメンバー間でのみ実行されます。

DLB 機能を使用した階層型加重 ECMP は、問題を解決し、重み付けされた分散後にスパインへのすべてのリンク全体に均等な分散を確保して、リンク使用率を最大化します。

DLB 静的ピン接続は、ECMP のレベル 2 でのみ発生します。WCMP がイネーブルで、すべての重みが類似している場合、静的ピン接続は通常どおりに機能します。これは、WCMP がこのレベルのメンバーグループを1つしか持っていないためです。たとえば、ネットワーク内の他の場所でのリンク障害のために、重みが異なる場合も、静的ピン接続はレベル 2 ECMP でのみ動作します。これは、特定の重みに一致するトラフィックだけがポートにピン留めされることを意味します。

UCMP とも呼ばれる加重 ECMP の詳細については、Cisco.com の『Cisco Nexus 9000 シリーズ NX-OS ユニキャストルーティング構成ガイド』の「BGP 上の不等コストマルチパス (UCMP)」の項を参照してください。

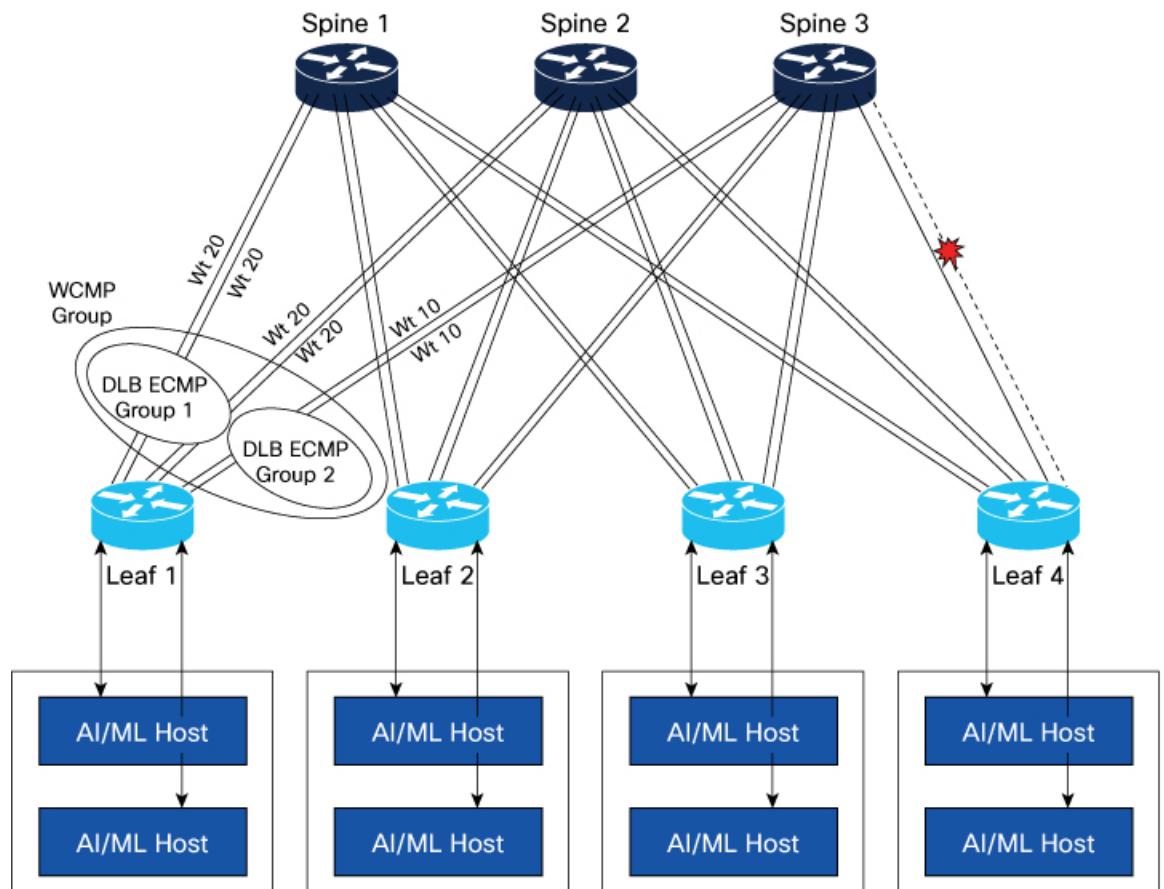
参考のために、階層型重み付け ECMP の例をここに示します。

この画像は、3つのスパインが4つのリーフに接続され、それぞれが個々の AI/ML ホストに接続されているトポロジを示しています。eBGP は、リーフをスパインに接続するすべてのリンクで構成されます。リーフはスパインに重みを持つルートをアドバタイズし、スパインは累積重みをアドバタイズします。

## Silicon One スイッチのダイナミック ロード バランシングの主なコンセプト

リーフ 1 から出力され、スパン 1、2、および 3 を介してリーフ 4 に入力されることになっている 100 ギガのトラフィックについて考えてみます。従来の方法では、スパン 1、2、3 へのすべてのリンクの 6 ウェイ ECMP を経由します。1 つのリンクがスパン 3 からリーフ 4 にダウンしているシナリオでは、そのリンクで使用可能な帯域幅は、他のスパンで使用可能な帯域幅まで減少します。これにより、スパン 3 で輻輳が発生する可能性があります。

図 2: 階層型 WCMP



524877

この問題は、DLB ECMP グループの階層型 WCMP を使用して解決されます。これにより、リーフへの全帯域幅を持つスパンを 1 つのグループ (DLB ECMP Group1) に、部分的な帯域幅を持つスパンを別のグループ (DLB ECMP Group2) にピックアップできます。これらが一緒になって重み付け ECMP グループを形成し、メンバー間で FLB が実行されますをサポートします。WCMP は、帯域幅に基づいて 80% のトラフィックに対してスパン 1 と 2 を動的に選択し、Leaf4 宛てのトラフィックの 20% に対してスパン 3 を選択します。したがって、重み 20 のトラフィックはスパン 1 と 2 のそれぞれ 2 つのリンクを介してリーフ 4 に伝送され、重み 10 のトラフィックはそれぞれスパン 3 の 2 つのリンクに伝送され、重み 20 のトラフィックをリーフ 4 に伝送する 1 つのアクティブリンクを介して配信されます。

つまり、DLB ECMP 機能を備えた WCMP は、ダイナミック ロード バランシングを使用して、リンク使用率に基づいて最適なリンクを選択します。

### IP ロード シェアリング

フローレットは、**ip load-sharing** コマンドで構成されたパラメータを使用して一意に識別されます。デフォルトで使用されるパラメータは、送信元 IP、宛先 IP、宛先ポート、送信元ポート、および IPv4 プロトコルです。

Silicon One スイッチを含む「ECMP ロード バランシングの構成」の詳細については、Cisco.com の『インターフェイス コンフィギュレーション ガイド』を参照してください。

## Silicon One スイッチのダイナミック ロード バランシングに関するガイドラインと制限事項

Silicon One スイッチにおけるレイヤ 3 ダイナミック ロード バランシングの注意事項と制約事項は、4 つのサブセクションに分かれています。

- [機能サポート](#)、
- [ポート](#)、
- [ECMP](#)、および
- [DLB パラメータ](#)です。

### 機能のサポート

- レイヤ 3 ECMP DLB は、次でのみサポートされます：
  - レイヤ 3 物理インターフェイス
  - IP ルーティングファブリックおよび VXLAN ファブリック
  - および N9364E-SG2 スイッチの 10.5 (3) F からです。
- DLB ECMP スケールが大きい場合は、システム pic-core オプションを使用してから、スイッチをリロードします。詳細については、「[BGP PIC コアの構成](#)」を参照してください。
- DLB フローで使用される最大パケット サイズに基づいて、すべての DLB 対応インターフェイスで MTU を構成します。そうでない場合、出力インターフェイスで出力がドロップされるとトラフィックがドロップされます。
- Only IPv4 と IPv6 ユニキャスト トラフィックのみがサポートされています。
- Cisco NX-OS リリース 10.6(1)F 以降、AI/ML ライセンスの機能拡張の一環として、適応型ルーティング機能は Silicon One スイッチでサポートされます。関連するライセンスが必要です。さらに、DLB 機能には、10.6(1)F 以降での Advantage ライセンスが必要です。DLB が有効になっていて、有効なインターフェイスリストを持っている場合、ライセンスを消費します。ライセンスの詳細については、「[Cisco NX-OS ライセンス オプション ガイド](#)」を参照してください。

- リリース 10.6(2)F 以降、静的ピン接続はサポート。

## ポート

- ポートチャネル、SVI、ポートチャネルメンバー、またはサブインターフェイスを DLB インターフェイスリストに含めることはできません。
- DLB インターフェイスリストには、NX-OS リリース 10.6(1)F からのブレークアウトポートを含む、最大 254 個の物理ポートを含めることができます。 **dlb interface all** オプションを使用すると、システム内のすべての物理ポートで DLB をサポートできます。
- DLB インターフェイスでは、1 つの IPv4、1 つの IPv6 ネクストホップ、および 1 つの IPv6 リンクローカル ネクストホップのみがサポートされます。

## ECMP

- DLB ECMP のメンバーの最大サポートは 127-way ECMP です。
- DLB ポリシー主導型混合モードでは、サポートされる範囲が 512 グループから 256 グループに減少します。
- DLB は、グループの作成中にハードウェアの ECMP グループで有効になり、次を含むリストされた条件が満たされます：
  - ECMP グループのすべてのメンバーが、DLB 対応インターフェイスリストに含まれています。 ECMP グループに DLB インターフェイスリストにないメンバーが 1 つ以上ある場合、その ECMP グループには通常の ECMP を使用されます。
  - interface-list all** オプションを選択すると、スイッチ上のブレークアウトインターフェイスを含むすべてのレイヤ3物理インターフェイスを DLBECMP グループのメンバーにできます。
  - サブインターフェイス、 SVI、およびポートチャネルを DLB ECMP グループのメンバーにすることはできません。これらのいずれかが ECMP グループのメンバーである場合、そのグループに対して DLB は有効になりません。
  - ECMP は、レジリエント ECMP ではありません。
- レジリエント ECMP 機能と DLB 機能を同時に有効にすることはできません。 レジリエント ECMP の詳細については、[Cisco Nexus 9000 シリーズ NX-OS インターフェイス構成ガイド](#)を参照してください。
- 従来の重み付け ECMP グループの場合、DLB は適用されません。重み付け ECMP または UCMP の詳細については、[BGP 経由不等コストマルチパス \(UCMP\)](#) を参照してください。
- show routing hash** コマンドは、DLB ECMP グループを使用するルートでは機能しません。これは、DLB が有効になっている場合、ポートの選択は静的ハッシュではなくリンク使用率に基づいてダイナミックに行われるためです。



(注) DLB 対応 ECMP グループのメンバーポートの 1 つがダウンすると、ハードウェアは、ポートを通してのトラフィックの送信をすぐに停止します。これにより、リンク障害時のトラフィック損失を最小限に抑えることができます。

### DLB パラメータ

DLB パラメータには、エージングとモードなどのパラメータが含まれます。

- DLB インターフェイスを初めて構成または変更する場合、構成を有効にするためにスイッチをリロードする必要があります。
- フローレットエージングタイムは、ファブリックでのラウンドトリップ時間に基づいて選択します。そうしなかった場合、フローの順序が変わる可能性があります。
- DLB モード間の変更は可能ですが、スイッチのリロードが必要です。インターフェイスに適用される DLB ポリシー ベースのアクションについては、QoS 設定が新しい **dlb mode** と一致することを確認します。
- エージング、モード、減衰係数、サンプリング間隔、負荷認識などのパラメータを削除すると、デフォルト値にリセットされます。
- ポートを DLB インターフェイスリストに追加した後、ポートがブレークアウトポートに変更される場合、PO の一部になるように追加される場合、このインターフェイスでサブインターフェイスが作成される場合、DLB は、ポートを含む ECMP グループに対して有効ではなくなります。DLB インターフェイスリストからポートを削除します。

## Silicon One スイッチのダイナミック ロード バランシングの構成

この手順では、Silicon One スイッチでダイナミック ロード バランシング (DLB) を設定するステップの概要を示します。設定変更の多くは、変更を有効にするためにスイッチのリロードが必要であることに注意してください。



(注) ポリシー主導型 DLB を機能させるには、このセクションで説明した DLB コンフィギュレーションを実行してから、`set dlb mode` で QoS ポリシーを設定します。[Cisco.com の「Cisco Nexus 9000 シリーズ NX-OS Quality of Service 構成ガイド」の「ポリシー主導型ダイナミック ロード バランシングの構成」](#) の項を参照してください。

### 手順の概要

1. ハードウェアプロファイル **dlb** モードを開始します。
2. ダイナミック ロード バランシング (DLB) のインターフェイスを指定します。**dlb-interface [interface\_range|all]** コマンドを使用して、DLB を有効にするインターフェイスのリストを指定します。カンマ区切りで、インターフェイスを追加します。このリストは動的には変

## Silicon One スイッチのダイナミック ロード バランシングの構成

更できません。インターフェイスリストを有効にするには、スイッチのリロードが必要です。

3. (任意) フローレットのエージング時間を設定します。DLB モードで **flowlet-aging usec** コマンドを入力して、フローレットのエージング時間をマイクロ秒単位 (*usec*) で設定します。デフォルト値は 255 マイクロ秒、最大値は 1024 マイクロ秒です。
4. (任意) ダイナミック ロード バランシングの構成
5. (任意) DLB のロード バランシング動作を調整します。DLB がトラフィックの変更にどのように対応して、サンプルを収集するかを微調整します。
6. (任意) (オプション) 静的ピン接続を設定します。
7. (任意) NVIDIA Spectrum-x 適応型ルーティングを有効にします。NVIDIA の spectrum-x 適応型ルーティング機能を有効にするには、**hardware profile spectrum-x** コマンドを使用します。このコマンドを実行すると、DLB 機能は **hardware profile dlb** での設定に基づいて実行されます。

## 手順の詳細

### 手順

---

#### ステップ1 ハードウェアプロファイル dlb モードを開始します。

- a) **configure terminal** コマンドを使用して、グローバルコンフィギュレーションモードになっていることを確認します。

例 :

```
switch# configure terminal
```

- b) **hardware profile dlb** コマンドを使用して、ハードウェア プロファイルのダイナミック ロード バランシングモードを開始します。このコマンドでは、リロードする必要があります。

例 :

```
switch(config)# hardware profile dlb
```

(注)

**hardware profile dlb** サブ モードでこのセクションに記載されているコマンドを実行して、Silicon One スイッチでレイヤ 3 ダイナミック ロード バランシングを設定します。

#### ステップ2 ダイナミック ロード バランシング (DLB) のインターフェイスを指定します。**dlb-interface [interface\_range | all]** コマンドを使用して、DLB を有効にするインターフェイスのリストを指定します。カンマ区切りで、インターフェイスを追加します。このリストは動的には変更できません。インターフェイスリストを有効にするには、スイッチのリロードが必要です。

(注)

- インターフェイス リストを変更した場合は、構成を有効にするためにリロードが必要です。
- インターフェイス リストへの部分的な追加や削除はサポートされていません。構成は、新しく提供されたインターフェイスに置き換えられます。

**all** オプションを使用する場合、スイッチ上のすべてのレイヤ3物理およびレイヤ3ブレークアウトインターフェイスでDLBを有効にする必要があります。これを有効にするには、スイッチのリロードが必要です。

有効な **dlb-interface all** モードでは、リロードは必要ありません。これは新しく追加された DLB 対応インターフェイス用です。

例：

```
switch(config-dlb)# dlb-interface Eth1/5,Eth1/7,Eth1/17,Eth1/21,Eth1/26
switch(config-dlb)# dlb-interface all
```

リリース 10.6(1)F から、以下に含まれる 1 つ以上のアクションを実行できます。

- DLBインターフェイスリストにブレークアウトポートを追加します。たとえば、最初の 64 個の 4 ウェイポートをブレークアウトします。
- インターフェイスの範囲を追加するための **range** コマンドを使用します。
- インターフェイスとインターフェイスの範囲の組み合わせを使用します。

(注)

- 最大で 64 個のインターフェイスまたはインターフェイスの範囲、または両方の組み合わせを追加できます。
- **no interface breakout...** コマンドを実行して dlbg インターフェイス設定を削除しできます。リロードの警告に従ってシステム リロードを実行します。

ポートを再度ブレークアウトした場合は、**dlb-interface** 設定を再適用してください。ただし、適用された **dlb** インターフェイスのリストにすでに存在するブレークアウトポートの新しい設定は、リロードしなくとも FLB で受け入れられます。

例：

```
switch(config-dlb)# dlb-interface eth1/1/1-4, eth1/2/1-4, eth1/3/1-4 ... eth1/64/1-2
switch(config-dlb)# dlb-interface eth1/1, eth1/2/1-4, eth1/3/1-4
switch(config-dlb)# no interface breakout module 1 port 1 map 100g-4x
```

a) インターフェイスリストと保留中のリロードを確認します。

- 現在の適用済みリストを確認するには、**show hardware profile db** コマンドを実行します。
- 保留中のリロードについて確認するには、**show system config reload-pending** コマンドを実行します。

**ステップ3** (任意) フローレットのエージング時間を設定します。DLB モードで **flowlet-aging usec** コマンドを入力して、フローレットのエージング時間をマイクロ秒単位 (*usec*) で設定します。デフォルト値は 255 マイクロ秒、最大値は 1024 マイクロ秒です。

(注)

フローの再順序付けを防ぐために、フローレット エージング タイムを慎重に選択してください。

例：

## Silicon One スイッチのダイナミック ロード バランシングの構成

```
switch(config-dlb) # flowlet-aging 600
```

### ステップ4 (任意) ダイナミック ロード バランシングの構成

- a) **mode [flowlet | per-packet | policy-driven]** コマンドを DLB モードで入力して、フローレット、パケットごと、またはポリシー駆動型 DLB モードを有効にします。デフォルトのモードはフローレットです。

(注)

モードを変更した場合には、スイッチをリロードして変更を適用します。

例 :

```
switch(config-dlb) # mode flowlet
```

- b) **mode policy-driven [flowlet | per-packet | mixed]** コマンドを DLB モードで入力して QoS ポリシーを有効にし、フローレット、パケットごと、または混合ポリシー駆動型 DLB モードに変更します。QoS ポリシーで設定されたモードは、ポリシー主導の DLB モードと一致している必要があります。

混合モードでは、Flowlet と Per-packet の両方がサポートされ、ECMP グループのスケールは 512 から 256 に縮小されます。いずれかのモードをデフォルトとして設定できます。

フローレットまたはパケットごとのアクションでトラフィックが QoS ポリシールールと一致しない場合、デフォルトで通常の ECMP が使用されます。

混合モードでは、[ステップ6](#)で説明されている構成を使用して、デフォルトの動作を変更できます。

QoS ポリシーの一一致の詳細については、Cisco.com の「Cisco Nexus 9000 シリーズ NX-OS Quality of Service 構成ガイド」の「ポリシー主導型ダイナミック ロード バランシングの構成」の章を参照してください。

(注)

モードを変更する場合は、リロードが必要です。

例 :

```
switch(config-dlb) # mode policy-driven mixed default ecmp
```

- c) **mode policy-driven mixed default [ecmp | flowlet | per-packet]** コマンドを DLB モードで入力し、混合モードでのデフォルトの動作を変更します。トラフィックがフローレットまたはパケットごとのアクションで QoS ポリシールールに一致しない場合、トラフィックはこのコマンドを使用して設定されたデフォルトモードを選択します。モードが設定されていない場合、設定はデフォルトで通常の ECMP になります。

(注)

モードを変更する場合は、リロードが必要です。

例 :

```
switch(config-dlb) # mode policy-driven mixed default ecmp
```

### ステップ5 (任意) DLB のロード バランシング動作を調整します。DLB がトラフィックの変更にどのように対応して、サンプルを収集するかを微調整します。

- a) DLB モードでは、**decay-factor** コマンドを使用して、ロード バランシングが過去のトラフィック サンプルからどのような影響を受けるかを制御します。小さい値を使用すると、最近のデータが考慮され

るため、変更への対応が速くなります。値を大きくすると、履歴データに重点が生じるため、変更への対応が遅くなります。デフォルト値は 2 です。減衰係数値の範囲は 0 ~ 15 です。

例：

```
switch(config-dlb)# decay-factor 2
```

- b) DLB モードで **sampling-interval** コマンドを使用して、各ポート負荷サンプル収集の時間間隔をナノ秒 (nsecs) 単位で設定します。デフォルトは 32000 ナノ秒です。指定できる範囲は、512 ナノ秒から 16,384,000 ナノ秒までです。サポートされる値は 512、1000、2000、4000、8000、16,000、32,000、64,000、128,000、256,000、512,000、1,024,000、2,048,000、4,096,000、8,192,000、および 16,384,000 です。

例：

```
switch(config-dlb)# sampling-interval 32000 nsecs
```

- c) ポートの負荷に基づいてポートを選択するには、DLB モードで [**no**] **load-awareness** コマンドを入力します。デフォルト値は **load-awareness** です。

ポートの負荷を無視するには、**no load-awareness** コマンドを使用します。

例：

```
switch(config-dlb)# load-awareness
```

## ステップ 6 (任意) (オプション) 静的ピン接続を設定します。

- a) 静的ピン接続機能を設定するには、DLB モードで **static-pinning** コマンドを使用します。

例：

```
switch(config-dlb)# static-pinning
```

- b) 静的ピン接続用に送信元インターフェイスと宛先インターフェイスを構成するには、**source source physical interface destination destination physical interface** コマンドを使用します。ただし、これらは物理インターフェイスのみでなければなりません。つまり、前面パネルのイーサネットインターフェイスです。SVI、ポートチャネル、またはサブインターフェイスを送信元または接続先インターフェイスにすることはできません。

例：

```
switch(config-dlb-static-pinning)# source ethernet 1/9 destination ethernet 1/31
```

このコマンドのガイドラインおよび制約事項は次のとおりです：

- 接続先インターフェイスは、DLB が適用または構成されたインターフェイスリストの一部である必要があります。
- 送信元インターフェイスは、スパインレイヤ上にある場合はDLB リストの一部になりますが、アクセス レイヤ上にある場合は DLB リストの一部にはなりません。
- DLB が適用または構成されたインターフェイスリストのインターフェイスは、静的ピン接続の送信元インターフェイスとしても使用できます。
- 2つの構成で同じ送信元インターフェイスが使用されている場合、最初の接続先インターフェイスは 2 番目の接続先インターフェイスで置き換えられます。送信元インターフェイスは同じだからです。

## ■ ダイナミック ロード バランシングの設定

- ・ブレークアウト ポートは、送信元インターフェイス、接続先インターフェイス、またはその両方として構成できます。
- ・ポートでブレークアウトまたはブレークアウト操作が実行されていない場合、ユーザーはDLB または静的ピン接続構成を更新する必要があります。
- ・静的ピン接続の接続先インターフェイスとして構成されているインターフェイスは、DLB インターフェイスリストから削除できません。削除するには、最初に静的ピン接続構成を削除して、その後 DLB インターフェイスリストからインターフェイスを削除します。
- ・ポリシー主導モードで静的ピン接続が有効になっている場合、ポリシーは適用されません。静的ピン接続はポリシーよりも優先されます。
- ・接続先インターフェイスがダウンすると、トラフィックはデフォルトのDLB モードを使用して転送されます。

**ステップ7** (任意) NVIDIA Spectrum-x 適応型ルーティングを有効にします。NVIDIA の spectrum-x 適応型ルーティング機能を有効にするには、**hardware profile spectrum-x** コマンドを使用します。このコマンドを実行すると、DLB 機能は **hardware profile dlb** での設定に基づいて実行されます。

(注)

**feature lldp** を有効にしていることを確認します。詳細については、『Cisco Nexus 9000 シリーズ NX-OS システム管理コンフィギュレーションガイド』を参照してください。spectrum-x コマンドでは、リロードは必要ありません。

例：

```
switch(config)# hardware profile spectrum-x
```

---

# ダイナミック ロード バランシングの設定

Nexus 9000 シリーズ スイッチで DLB を有効にするには、設定例を参考にしてください。

CloudScale および Silicon One ハードウェア プラットフォーム向けに個別の設定例が提供されています。

CloudScale スイッチには、この設定を使用してください。

```
switch# configure terminal switch(config)# hardware profile dlb switch(config-dlb)#
dlb-interface Eth1/5,Eth1/7,Eth1/17,Eth1/21,Eth1/26 switch(config-dlb)#
dre-thresholds level-1 15 level-2 20 level-3 30 level-4 15 level-5 10 level-6 5 level-7 5
switch(config-dlb)#
flowlet-aging 600 switch(config-dlb)#
mac-address aa:bb:cc:dd:ee:ff
switch(config-dlb)#
mode flowlet switch(config-dlb)#
static-pinning
switch(config-dlb-static-pinning)#
source ethernet 1/1 destination ethernet 1/2
```

Silicon One スイッチには、この設定を使用してください。

```
switch# configure terminal switch(config)# hardware profile dlb switch(config-dlb)#
dlb-interface all switch(config-dlb)#
flowlet-aging 600 switch(config-dlb)#
mode flowlet
switch(config-dlb)#
decay-factor 2 switch(config-dlb)#
sampling-interval 32000 nsecs
switch(config-dlb)#
load-awareness switch(config-dlb)#
static-pinning
switch(config-dlb-static-pinning)#
source ethernet 1/9 destination ethernet 1/31
```

# ダイナミック ロード バランシングの構成の確認

ダイナミック ロード バランシング (DLB) の設定に関する情報を表示するには、テーブルからいざれかのコマンドを選択します。次の表に、DLB 用の show コマンドとその目的を示します。

コマンド	目的
<b>show hardware profile dlb</b>	<p>DLB の構成を表示します。</p> <p>(注)</p> <ul style="list-style-type: none"> <li>[構成済みインターフェイスリスト (Configured Interface-list) ] : コマンドラインインターフェイスを使用して構成されている現在のインターフェイスのリストアップします。リロード後このリストは、適用済みインターフェイスリストを入力します。</li> <li>[適用済みインターフェイスリスト (Applied Interface-list) ] : DLB に現在使用されているインターフェイスのリストを提供します。</li> <li>[設定済み DLB モード (Configured DLB mode) ] : 設定済みの DLB モードを指定します。スイッチをリロードすると、[適用済み DLB (Applied DLB) ] モードになります。</li> <li>[適用済み DLB モード (Applied DLB mode) ] : 現在使用されている DLB モードを指定します。</li> </ul>
<b>show hardware spectrum-x</b>	spectrum-x 適応型ルーティング機能が有効になっているかどうかを表示します。
<b>show system config reload-pending</b>	リロード保留中の構成を表示します。DLB の場合、インターフェイスリスト構成に変更が加えられた場合、アプリケーションを保留中のインターフェイスリストが表示されます。



(注) **show routing hash** コマンドは、DLB ECMP グループを使用するルートでは機能しません。

## ■ ダイナミック ロード バランシングの構成の確認

### 表示コマンドの出力

CloudScale プラットフォームでは、**show hardware profile dlb** コマンドの出力例を表示できます。

```
switch# show hardware profile dlb
DLB Configurations:
=====
Enabled: yes
Mode: flowlet
Mac-address: aa:bb:cc:dd:ee:ff
Flowlet aging time: 600 usec(s)
DRE-thresholds:
  Level-1:15
  Level-2:20
  Level-3:30
  Level-4:15
  Level-5:10
  Level-6:5
  Level-7:5
DLB interface list:
-----
Configured Interface-list (size: 5):
  Eth1/5,Eth1/7,Eth1/17,Eth1/21,Eth1/26

Applied interface-list (size: 5):
  Eth1/5,Eth1/7,Eth1/17,Eth1/21,Eth1/26

Static-pinning enabled: yes

DLB static-pinning pairs:
-----
static-pinning pairs (1):

  source: Eth1/1    dest: Eth1/5
```

Silicon One プラットフォームでは、**show hardware profile dlb** コマンドの出力例を表示できます。

```
switch# show hardware profile dlb
DLB Configurations:
=====
Enabled: yes
Configured DLB Mode: flowlet
Applied DLB Mode: flowlet
Flowlet aging time: 512 usec(s)
Decay Factor: 2
Sampling Interval: 32000 nsec(s)
Load Awareness: True
DLB interface list:
-----
Configured Interface-list : all

Applied Interface-list : all

Static-pinning enabled: yes

DLB static-pinning pairs:
```

```
-----  
static-pinning pairs (1):  
source: Eth1/9 dest: Eth1/31  
switch#
```

サポートされているすべての Nexus 9000 シリーズ スイッチ（CloudScale と Silicon One の両方）で、**show hardware spectrum-x** コマンドの出力例を表示できます。

```
switch# show hardware profile spectrum-x  
Spectrum-x Configurations:  
=====  
Enabled: yes  
spine1(config)#
```

Nexus 9000 シリーズ スイッチ（CloudScale と Silicon One の両方）で、DLB の **show system config reload-pending** コマンドの出力例を表示できます。

```
switch# show system config reload-pending  
  
Following config commands require copy r s + reload :  
=====  
0      hardware profile dlb ; dlb-interface Eth1/5,Eth1/7,Eth1/17,Eth1/21,Eth1/26  
=====
```

## ダイナミック ロード バランシングのトラブルシュート

すべての Nexus スイッチの DLB のトラブルシューティングの詳細については、Cisco.com の『Cisco Nexus 9000 シリーズ NX-OS トラブルシューティング ガイド』の「ダイナミック ロード バランシングのトラブルシューティング」の項を参照してください。

## ■ ダイナミックロードバランシングのトラブルシュート

## 翻訳について

このドキュメントは、米国シスコ発行ドキュメントの参考和訳です。リンク情報につきましては、日本語版掲載時点で、英語版にアップデートがあり、リンク先のページが移動/変更されている場合がありますことをご了承ください。あくまでも参考和訳となりますので、正式な内容については米国サイトのドキュメントを参照ください。