

Server Load Balancer(SLB)とは

Server Load Balancer は、トラフィックを複数の ECS インスタンスに分散させるサービスです。トラフィックを分散することでアプリケーションシステムのサービス能力を向上すると共に、SPOF (単一障害点) を回避することによって可用性を高めることができます。

1 プロダクトの概要

1-1 プロダクト概要

- Server Load Balancer は、トラフィックを複数の ECS インスタンスに分散させるサービスです。トラフィックを分散することでアプリケーションシステムのサービス能力を向上すると共に、SPOF (単一障害点) を回避することによって可用性を高めることができます。
- Server Load Balancer は、同じリージョンにある複数の Elastic Compute Service (ECS) インスタンスを仮想化するための仮想 IP アドレスを高パフォーマンスと高可用性を備えたアプリケーションのプールに設定します。Server Load Balancer は、アプリケーションによって指定されたメソッドに基づいて、ネットワークリクエストをクライアントから ECS プールに配分します。
- Server Load Balancer によって、ECS プール内の ECS インスタンスのヘルスステータスがチェックされ、異常な ECS インスタンスが自動的に分離されるため、単一の ECS インスタンスの SPOF (単一障害点) が解決し、アプリケーションの全体的なサービス機能が向上します。標準のロードバランシング機能に加えて、Server Load Balancer は TCP および HTTP メソッドを使用して DDoS 攻撃を防御できるため、アプリケーションサーバーの防御機能が向上します。

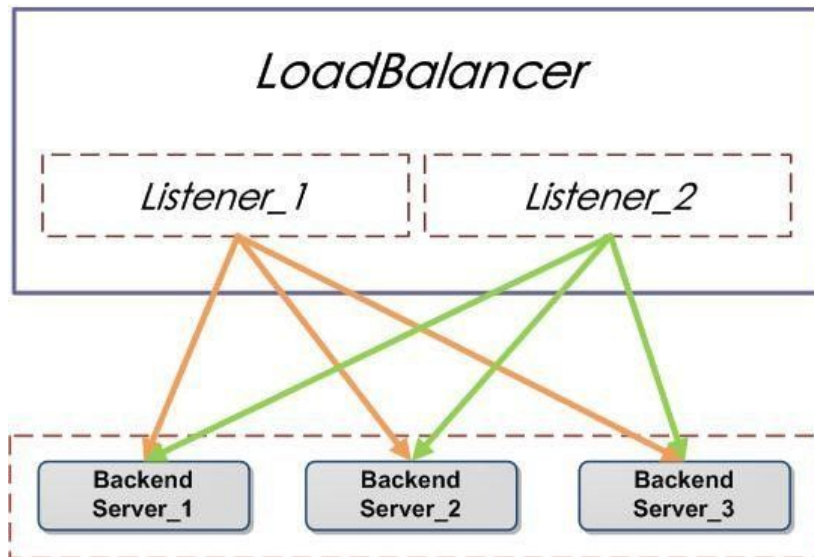
1-2 プロダクトコア概念

Server Load Balancer には、主に次の 3 つのコア概念があります。

- LoadBalancer は Server Load Balancer インスタンスを表します。
- Listener はカスタムロードバランシングポリシーおよび転送ルールを表します。
- BackendServer はバックエンド ECS インスタンスのグループを表します。

外部アクセスリクエストは、関連するポリシーと転送ルールに応じて、Server Load Balancer インスタンスを介してバックエンド ECS インスタンスに転送されて処理されます。次の図は、Server Load Balancer のコア概念を示しています。

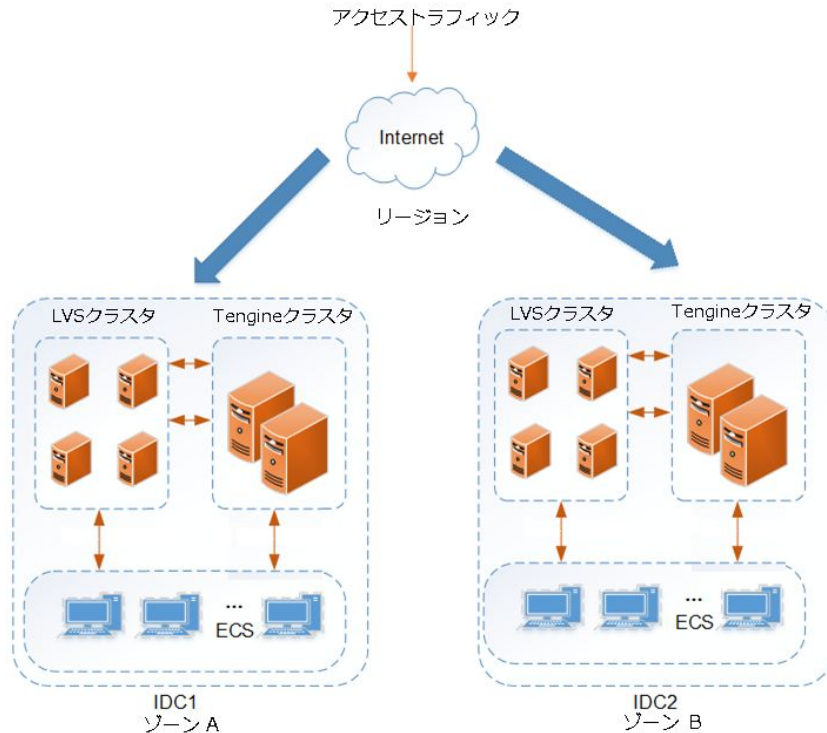
Server Load Balance



1-3 プロダクトアーキテクチャ

下図は、Server Load Balancer のアーキテクチャを示しています。その詳細は次のとおりです。

- 現在、Server Load Balancer はレイヤー 4 プロトコルとレイヤー 7 プロトコルの両方で使用できます。
- レイヤー 4 Server Load Balancer は、オープンソースソフトウェアの LVS と keepalived を基盤にしています。
- レイヤー 7 Server Load Balancer は、Taobao が開始した Web サーバープロジェクトである Tengine を利用します。Nginx を基盤とする Tengine には、トラフィックの多い Web サイトのニーズに合わせた幅広い高度な機能と特性が追加されています。
- Server Load Balancer をクラスターとしてデプロイすると、セッションを同期して、サーバーを SPOF (単一障害点) から保護することができるため、冗長性が向上し、サービスの安定性が確保されます。
- 特定のリージョンでは、ローカルな耐障害性を担保するために、2 つのゾーン (データセンター) がデプロイされます。



2 プロダクトの機能

2-1 インスタンスリスナー

2-1-1 サポートされるプロトコル

SLBのリスナーで利用できるプロトコルは、レイヤー4プロトコル(TCP および UDP)とレイヤー7プロトコル(HTTP および HTTPS)を使用することができます。ポート番号は任意に変更することが可能です。

2-1-2 ヘルスチェック

バックエンド ECS インスタンスでのヘルスチェックにより、Server Load Balancer は異常な ECS インスタンスを自動的にブロックし、再び正常に機能するようになったら復元できます。

2-1-3 セッション維持

レイヤー4およびレイヤー7のセッション維持をサポートして、レイヤー4 Server Load Balancer は IP アドレス、レイヤー7 Server Load Balancer はクッキーを使用して、セッション維持を実現します。セッションのライフサイクルの間、クライアントの要求を同じバックエンド ECS インスタンスに転送します。

レイヤー7(HTTP および HTTPS) Server Load Balancer では、クッキーベースのセッション維持がサポートされます。Server Load Balancer システムでは、クッキーベースの2つの処理方法が提供されます。

- クッキーシードとは、Server Load Balancer システムによってクライアント側のクッキーが直接割り当てられて管理されることを意味します。設定時にセッション維持のタイムアウト時間を指定する必要があります。

注意: セッション維持に対して“クッキーシード”モードが設定されていて、バックエンドRSから返されたHTTPステータスコードが4xxである場合、Server Load BalancerはSet-Cookieヘッダーをシード処理しません。シード処理すると、セッション維持に失敗する場合があります。バックエンドRSから返されたHTTPステータスコードが200、201、204、206、301、302、303、304、または307である場合、Server Load Balancerはセッション維持に必要なヘッダーをシード処理します。

- クッキー書き換えとは、Server Load Balancer システムにより、カスタムクッキー名に応じて、クライアント側のクッキーシード操作が割り当てられて管理されることを意味します。クッキー書き換えにより、カスタムクッキー名を特定して区別できるようになるため、バックエンドサーバーでさまざまなアプリケーションのセッション維持ルールを設定できます。設定時に対応するクッキーの名前を指定する必要があります。

レイヤー4 (TCP および UDP) Server Load Balancer では、IP アドレスベースのセッション維持がサポートされます。Server Load Balancer により、アクセスリクエストが IP アドレスから同じバックエンド ECS インスタンスに転送されて処理されます。

2-1-4 スケジューリングアルゴリズム

Server Load Balancer は、重み付きラウンドロビンと最小接続アルゴリズムとポーリングモードをサポートします。重み付きラウンドロビンは、外部要求をバックエンド ECS インスタンスに順番に分配します、重みが高いバックエンド ECS インスタンスほど、要求を受け取る可能性が高くなります。最小接続アルゴリズムは、現在接続数が最も少ないバックエンド ECS インスタンスに外部要求を分配します。ポーリングモードは順番に ECS インスタンスに外部要求を分配します。

2-1-5 帯域幅制限

インターネットとの接続ネットワークの帯域幅を設定することができます。デフォルト設定では帯域幅制限はありませんが、1~1,000Mbpsで制限することができます。本機能は、帯域幅を制限する機能であり、帯域幅を保証する機能ではありません。

2-2 バックエンドサーバ

2-2-1 ECSの追加

Server Load Balancer は、トラフィックを複数の ECS インスタンスに分散させるサービスです。ECSの追加、削除、および重みを設定し、Server Load Balancer インスタンスのリスナーのトラフィック転送を制御することができます。

2-2-2 仮想サーバーグループの追加

仮想サーバーグループはリスナーで設定されているリージョンにある ECS のみで構成されるものである。デフォルトのロードバランシングは、ディメンションインスタンス上のバックエンド

サーバーグループを管理して、モニター配下のバックエンドサーバーグループのすべてのインスタンスは同じバックエンドサーバーグループを使用します。仮想サーバーグループによって、リスニングディメンションが定義されたサーバーグループをカスタマイズできます。つまり、異なるリスナーのインスタンスが、個々のニーズに合わせて異なるセットのバックエンドサーバー、ドメイン名および URL 転送を使用できるということです。

2-3 転送ルール

2-3-1 ドメイン名/URL ベースの転送

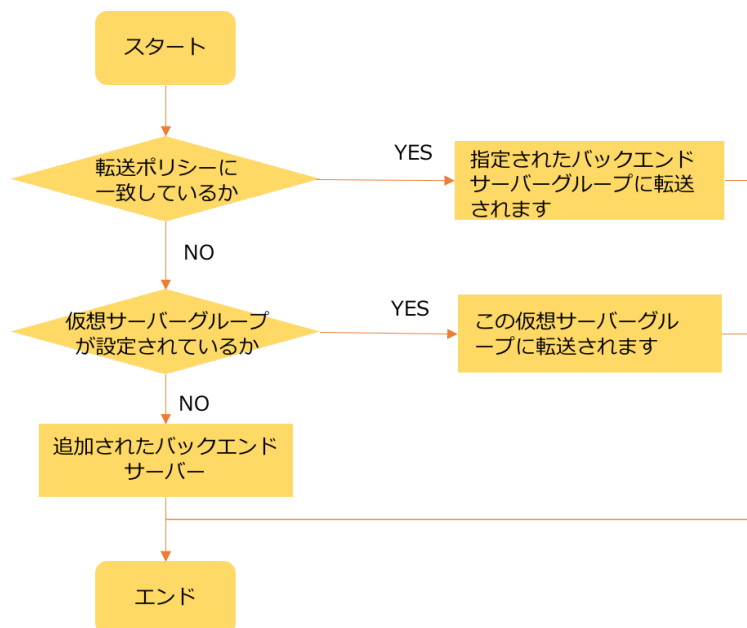
レイヤー 7 (HTTP および HTTPS) プロトコルの場合、Server Load Balancer はユーザーのドメイン名または URL に基づいて異なる仮想サーバーグループにトラフィックを転送します。Server Load Balancer では、次の 2 つの方法でドメイン名/URL ベースの転送機能が提供されます。

(1) リスナーレベルに基づいてすべてのリスナー (TCP/UDP/HTTP/HTTPS) のバックエンドサーバーグループをカスタマイズできます。また、異なるサーバーグループへの転送を行うことができます。

(2) HTTP および HTTPS リスナーの場合、ドメイン名および URL に基づいてリスナーに転送ルールを設定し、それぞれのバックエンドサーバーグループにリクエストを転送できます。

2-3-2 転送策略

ユーザートラフィックが Server Load Balancer ポートを通過するとき、まず“転送ルール”に一致しているかどうかを確認されます。一致している場合、トラフィックはこのルールによって指定されたバックエンドサーバーグループに転送されます。一致していない場合、およびこのリスナーに仮想サーバーグループが設定されている場合、トラフィックはこの仮想サーバーグループに転送されます。このリスナーに仮想サーバーグループが設定されていない場合、トラフィックはインスタンスレベルで追加されたバックエンドサーバーに転送されます。

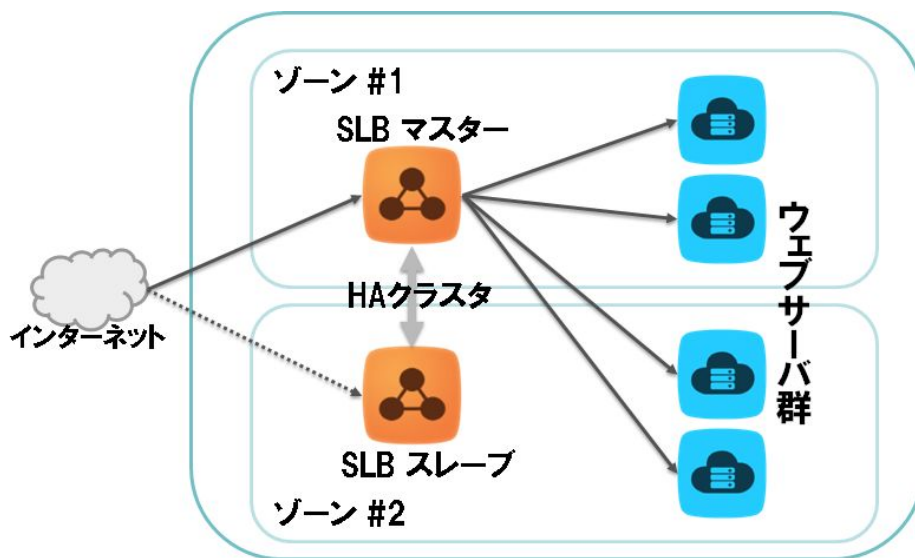


2-4 モニタリング

Server Load Balancer の実行ステータスがリアルタイムでわかるように、豊富なモニタリングデータが提供されます。Server Load Balancer は、次の 7 つのモニターとアラームを提供します。インバウンドトラフィック、アウトバウンドトラフィック、新しい接続の数、受信パケットの数、送信パケットの数、有効な接続の数、無効な接続の数。

2-5 マルチゾーン

指定されたゾーンの Server Load Balancer インスタンスの作成をサポートします。マルチゾーンリージョンにはマスターゾーンとスレーブゾーンをデプロイできます。マスターゾーンが障害のときは、スレーブゾーンが障害のあるゾーンからサービスを自動的に引き継ぎます。



2-6 ホワイトリスト機能

ホワイトリストは、Server Load Balancer モニタリングにアクセスできる IP アドレスを決定するために使用されるアクセス制御方法です。ユーザーアプリケーションが特定の IP アドレスだけを許可する場合に適用されます。

2-7 セキュリティ

アプリケーションファイアウォールと CC 保護をサポートします。クラスターの WAF モジュールを使用することで、CNAME を変更することなく WAF 保護を有効にできます。Anti-DDoS サービスと組み合わせて使用することで、システムは 5 Gbps 以下の DDoS 攻撃を防ぐことができます。

2-8 証明書管理

これは HTTPS プロトコルでの統合証明書管理サービスであり、バックエンド ECS インスタンスに証明書をアップロードする必要がなくなります。復号化は Server Load Balancer で実行されるので、バックエンド ECS インスタンスの CPU オーバーヘッドが減ります。

2-9 サポートするインスタンスタイプ

Server Load Balancer はパブリックネットワークまたはプライベートネットワークでサポートします。プライベートネットワーク (VPC) 内に Server Load Balancer を設置し内部ロードバランサーとして利用することができます。

ご利用上の注意事項

この資料は、Alibaba Cloudの提供するクラウドサービスの機能について説明したもので、サービスのご利用を検討する際の参考となる技術的情報を提供するものです。

今後、本資料はクラウドサービスの機能追加・変更等にに合わせて、予告なく変更される場合があります。閲覧された情報は最新のものではない場合がありますので、予めご了承下さい。

改版履歴

| 日付 | 版数 | 変更内容 |
|-----------|-----|---------------|
| 2017/1/23 | 1.0 | 初版作成 |
| 2017/2/14 | 1.1 | 帯域幅の仕様を追加 |
| 2017/2/15 | 1.2 | 帯域幅制限機能の説明を修正 |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |

本文書中に記載されている社名・商品名等は各社の商標または登録商標です。