

# Alibaba Cloud Server Load Balancer

プロダクト紹介

Document Version20190524

# 目次

---

1 Server Load Balancer について.....	1
2 アーキテクチャ.....	3
3 機能.....	6
4 適用シナリオ.....	10
5 プロダクト用語.....	13

# 1 Server Load Balancer について

---

Server Load Balancer (SLB) は、設定された転送ルールに基づいて複数の ECS インスタンスにトラフィックを分配する、負荷分散制御サービスです。SLB はアプリケーションサービスの性能を拡張し、アプリケーションの可用性を向上させます。

## 概要

SLB は、仮想サービスアドレスを設定することによって、追加の ECS インスタンスを高性能で可用性の高いアプリケーションサービスプールに仮想化し、クライアントからのリクエストを、転送ルールに従ってサーバープール内の ECS インスタンスに分配します。

また、SLB は、追加されたバックエンドサーバーの状態をチェックし、異常状態の ECS インスタンスを自動的に分離します。そうすることで SPOF (単一障害点) 問題を除去し、アプリケーションの全体的なサービス性能を向上させます。それに加え、Alibaba Anti-DDoS と組み合わせることで、SLB は DDoS 攻撃を防御することができます。

## Server Load Balancer を構成するコンポーネント

SLB には次のコンポーネントが含まれています。

- ・ SLB インスタンス

SLB インスタンスは、実行中の負荷分散サービスで、着信トラフィックをバックエンドサーバーに分配します。負荷分散サービスを使用するには、SLB インスタンスを作成します。インスタンスには少なくとも 1 つのリスナーと 2 つのバックエンドサーバーを設定する必要があります。

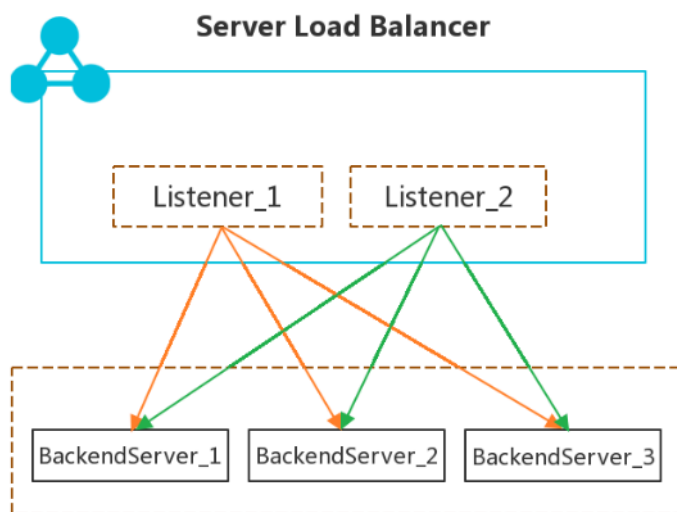
- ・ リスナー

リスナーはクライアントからのリクエストをチェックし、設定されたルールに基づいてバックエンドサーバーに転送します。また、バックエンドサーバーのヘルスチェックも実行します。

- ・ バックエンドサーバー

バックエンドサーバーは、分散リクエストを処理するために SLB インスタンスに追加された ECS インスタンスです。分散リクエストを処理する ECS インスタンスは、デフォルトサー

バーグループ、VServer グループ、アクティブ/スタンバイサーバーグループのいずれかに追加することができます。



## 特徴

- ・ 高可用性

Server Load Balancer は、SPOF なしで完全冗長モードで動作するように設計されています。Server Load Balancer はローカルおよびクロスリージョンでの耐障害性を提供しています。Server Load Balancer を DNS と併用すると、サービス可用性は最大 99.95% になります。

サービスの継続性を妨げることなく、アプリケーションの負荷状況に応じてサービスを拡張することができます。

- ・ 拡張性

必要に応じてバックエンドサーバーの数を増減し、アプリケーションのサービス性能を拡張することができます。

- ・ 高い費用対効果

従来のハードウェア負荷分散システムと比較して、Server Load Balancer はコストを 60% 削減しています。

- ・ セキュリティ

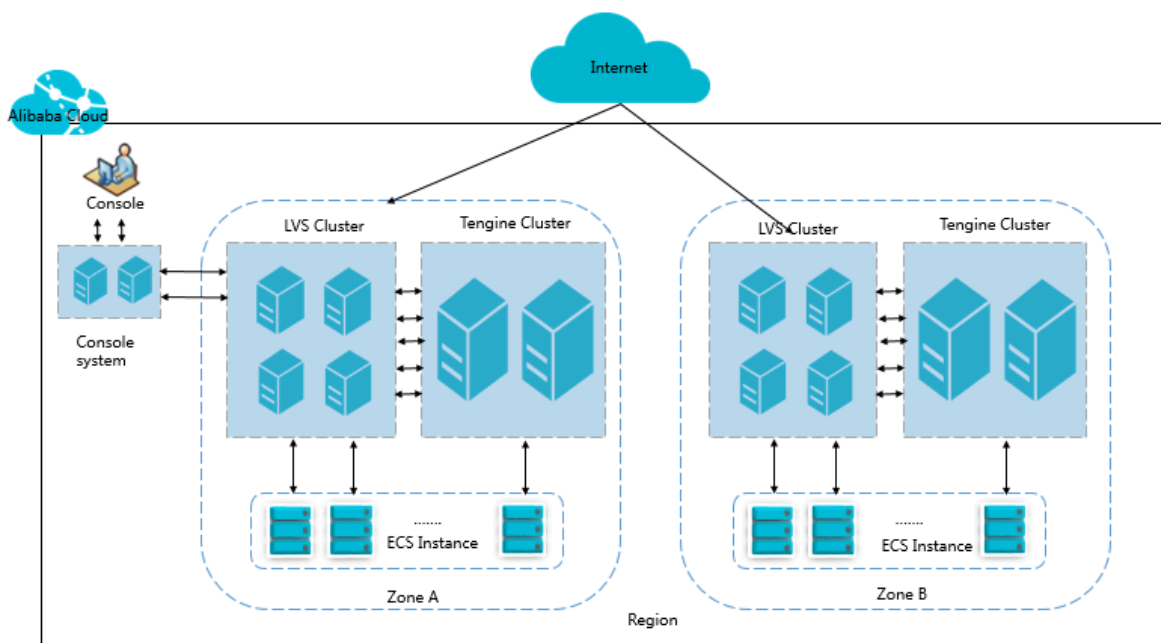
Alibaba Cloud Security と併用することで、Server Load Balancer は、HTTP フラッド攻撃や SYN フラッド攻撃など、最大 5 Gbps の DDoS 攻撃を防御できます。

## 2 アーキテクチャ

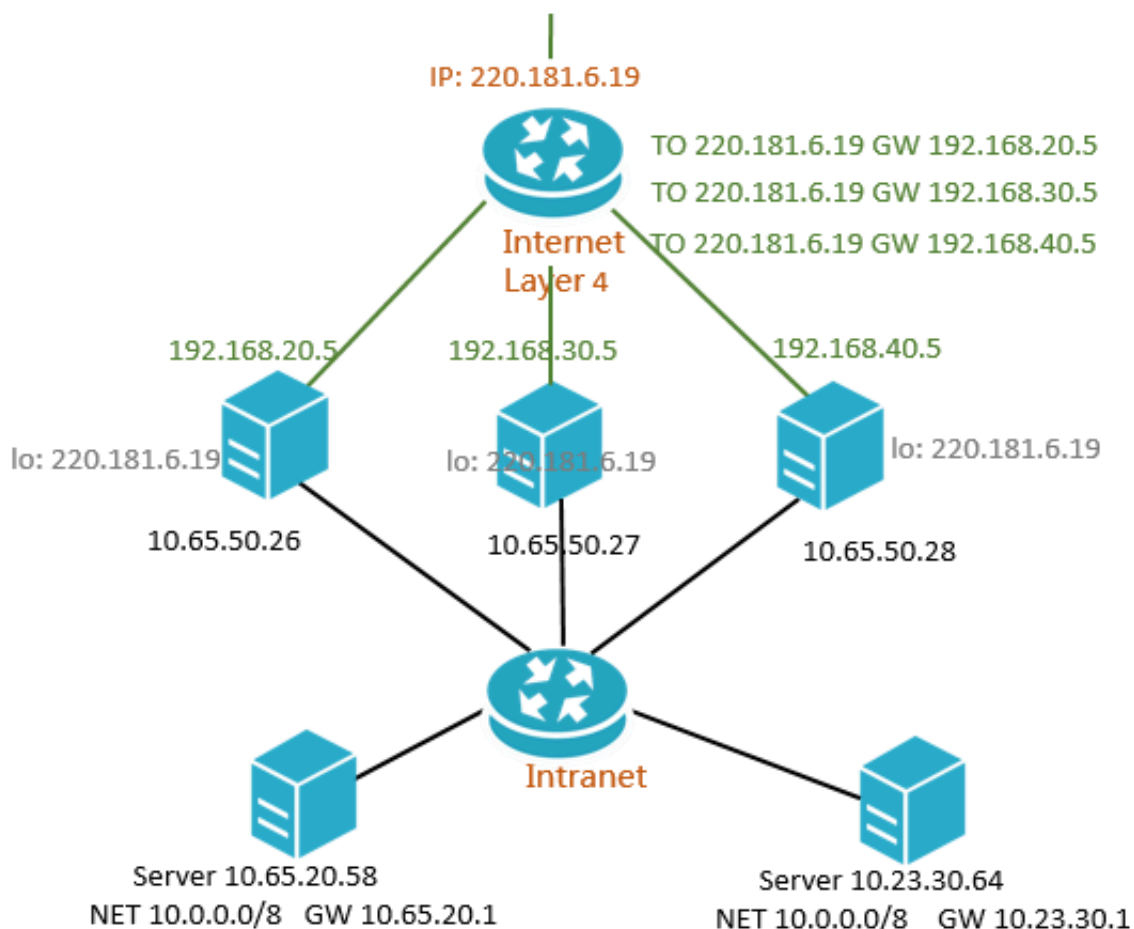
Server Load Balancer は、クラスターにデプロイされています。クラスターデプロイメントモデルにより、バックエンドサーバーの単一障害点が排除され、冗長性とサービスの安定性が向上します。

Alibaba Cloud は、レイヤー 4 (TCP プロトコルと UDP プロトコル) とレイヤー 7 (HTTP プロトコルと HTTPS プロトコル) のロードバランシングサービスを提供します。

- ・ レイヤー 4 は、オープンソースソフトウェアの LVS (Linux Virtual Server) と keepalived を利用してロードバランシングを実現し、クラウドコンピューティングの要件に基づいてカスタマイズを行っています。
- ・ レイヤー 7 SLB は Tengine を使用してロードバランシングを実現します。Tengine はタオバオが立ち上げた Web サーバープロジェクトです。Nginx を基盤とし、トラフィックの多い Web サイト向けに幅広い高度な機能を追加しています。



下図のように、各リージョンのレイヤー 4 SLB は実際のところ複数の LVS マシンでデプロイした LVS クラスター内で実行されます。クラスターデプロイメントモデルは、異常な状態でロードバランササービスの可用性、安定性および拡張性を強化します。

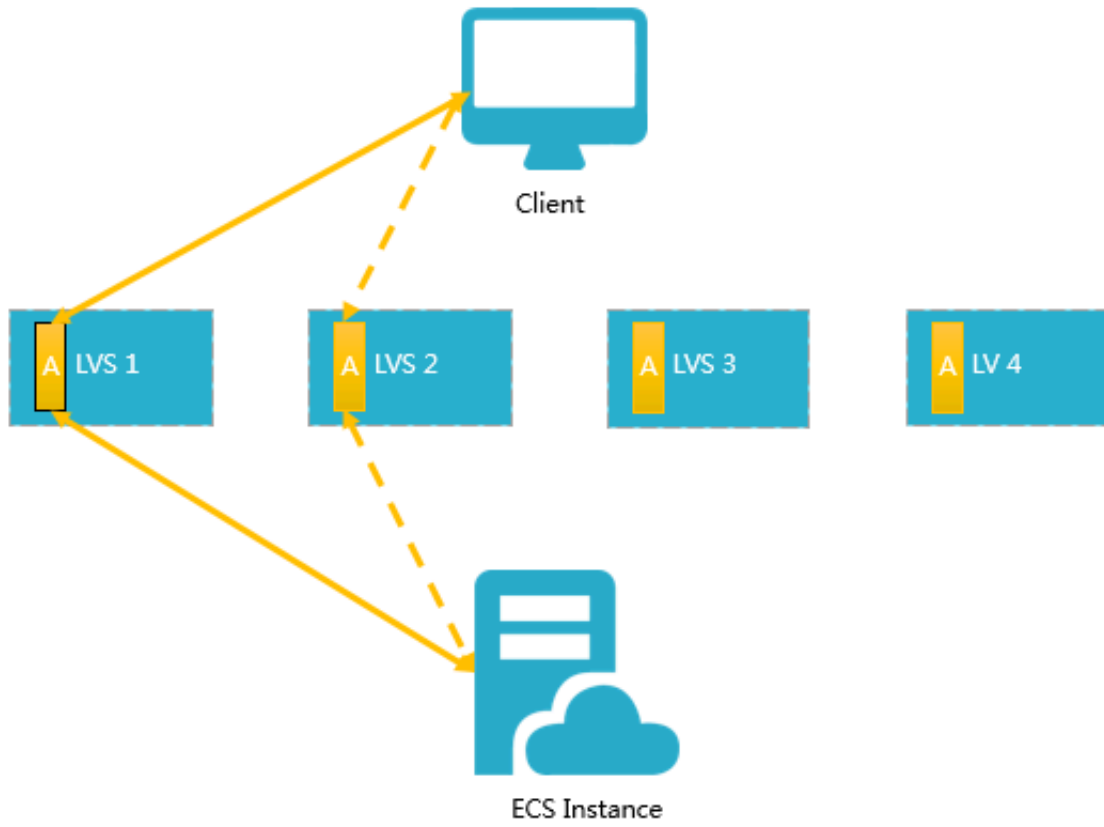


さらに、LVS クラスタ内の LVS マシンは、マルチキャストパケットを使用してセッションを他の LVS マシンに同期させることが可能です。下図のように、LVS1 で確立されたセッション A は、3つのパケットが転送された後、他の LVS マシンと同期されます。通常の状況では、セッションのリクエストは下図に示す実線のように LVS1 に送信されます。LVS1 が異常であるかメンテナンスされている場合、セッションのリクエストは下図に示す点線のように他の正常に動作しているマシンに送信されます。それにより、SLB クラスタはホットアップグレードに対応できるようになり、マシン障害またはシステムメンテナンスなどの状況によりビジネスに与える影響を防ぐことができます。



注：

接続が確立されていない場合 (3 ウェイハンドシェイクができていない場合)、または接続は確立されているがセッションの同期は開始されない場合、ホットアップグレードは接続が中断されないという保証はできません。クライアントが接続を再開する必要があります。



## 3 機能

Alibaba Cloud は、レイヤー 4 とレイヤー 7 の負荷分散サービス以外に、ヘルスチェック、セッション維持、ドメイン転送などの機能を提供し、アプリケーションの高可用性を保証します。

機能	レイヤー 4 Server Load Balancer	レイヤー 7 Server Load Balancer
<p>スケジューリングアルゴリズム</p> <p>Server Load Balancer はラウンドロビン、重み付きラウンドロビン (WRR)、重み付き最小接続数 (WLC) の各方式と、コンシステントハッシュ法をサポートしています。</p>	#	#
<p>ヘルスチェック</p> <p>Server Load Balancer はバックエンドサーバーの状態をチェックします。あるバックエンドサーバーが異常状態であると判明すると、Server Load Balancer はそのサーバーへの分配を停止し、他の正常なバックエンドサーバーへリクエストを分配します。</p>	#	#
<p>セッション維持</p> <p>Server Load Balancer はセッション維持機能に対応しています。Server Load Balancer はセッション期間中、同一クライアントからのリクエストを常に同じバックエンドサーバーへ転送できます。</p>	#	#
<p>アクセス制御</p> <p>Server Load Balancer はアプリケーションへのアクセス制御として、ホワイトリスト機能とブラックリスト機能をサポートしています。</p>	#	#



機能	レイヤー 4 Server Load Balancer	レイヤー 7 Server Load Balancer
<p>高可用性</p> <p>Server Load Balancer は、着信トラフィックを別のゾーン内のバックエンドサーバーに転送できます。また、Server Load Balancer は、ほとんどのリージョンにアクティブ/スタンバイモードでデプロイされています。プライマリゾーンが使用できない場合、Server Load Balancer は自動的にスタンバイゾーンに切り替えて負荷分散サービスを提供します。</p>	#	#
<p>セキュリティ</p> <p>Alibaba Cloud Security と組み合わせると、Server Load Balancer は最大 5 Gbps の DDoS 攻撃を防御できます。</p>	#	#
<p>インターネットとイントラネットの負荷分散</p> <p>Server Load Balancer は、インターネット、イントラネット両方において負荷分散サービスを提供します。VPC ネットワーク上のトラフィックの負荷分散にはイントラネット SLB インスタンスを、インターネットからのトラフィックを負荷分散するにはインターネット SLB インスタンスを作成します。</p>	#	#
<p>モニタリング</p> <p>CloudMonitor サービスを使用して、接続数やトラフィック、その他の SLB インスタンスの情報を表示できます。</p>	#	#
<p>IPv6 のサポート</p> <p>Server Load Balancer は IPv6 クライアントからのリクエスト転送に対応しています。</p>	#	#

機能	レイヤー 4 Server Load Balancer	レイヤー 7 Server Load Balancer
<p>アクセスログ</p> <p>Log Service を使用すると、SLB インスタンスのアクセスログを分析して、ユーザーの操作や地理的分布を把握したり、問題のトラブルシューティングを行ったりできます。</p>	—	#
<p>ヘルスチェックログ</p> <p>Server Load Balancer はバックエンドサーバーのヘルスチェックログをデフォルトで 3 日分保存しています。トラブルシューティングのために OSS 内にすべてのヘルスチェックログを保存することもできます。</p>	#	#
<p>ドメイン名や URL による振り分け</p> <p>レイヤー 7 の Server Load Balancer ではドメイン名や URL を用いた転送ルールの設定ができ、さまざまなドメインまたは URL へのリクエストをそれぞれ別のバックエンドサーバーに転送することができます。</p>	—	#
<p>証明書管理</p> <p>Server Load Balancer は、HTTPS プロトコルを使用するアプリケーション用に証明書の一元管理サービスを提供します。バックエンドサーバーに証明書をアップロードする必要はありません。復号化は Server Load Balancer で実行されるため、バックエンドサーバーの CPU 使用率を減らすことができます。</p>	—	#
<p>SNI のサポート</p> <p>Server Load Balancer は、1 つの HTTPS リスナーに複数の証明書を設定することができ、異なるドメインへのリクエストをそれぞれ別のバックエンドサーバーへ割り当てられます。</p>	—	#

機能	レイヤー 4 Server Load Balancer	レイヤー 7 Server Load Balancer
リダイレクト  Server Load Balancer は、HTTP リクエストを HTTPS リクエストにリダイレクトすることができます。	—	#
WS と WSS のサポート  WebSockets は HTML の新しいプロトコルです。クライアントとサーバーの間の双方向通信チャンネルを提供して、サーバーのリソースと帯域幅を節約し、リアルタイム通信を実現します。	—	#
HTTP/2 のサポート  HTTP/2 は、Hypertext Transfer Protocol の第 2 のバージョンです。HTTP/1.X と下位互換性があり、パフォーマンスが著しく向上しています。	—	#

## 4 適用シナリオ

---

Server Load Balancerは、高トラフィックのアプリケーションに適しており、可用性と信頼性が向上します

### アプリケーションの負荷分散

Server Load Balancerは、着信トラフィックを複数のバックエンドサーバー（ECSインスタンス）に自動的に配信できます。また、セッション維持を構成することによって、同一クライアントからのリクエストを同一バックエンドサーバーに分散することができます。

### アプリケーション機能の拡張

ビジネスニーズに応じて、いつでもバックエンド ECS インスタンスを追加または削除してサービス機能を拡張できます。自動スケーリング機能と一体化することで、バックエンドサーバーに事前設定されたしきい値が超えられると、アプリケーションはいつでも拡張可能です。

### 単一障害点からアプリケーションを保護

SLBインスタンスに複数のバックエンドサーバーを追加可能です。一部のバックエンドサーバーが不具合になった場合、Server Load Balancerはそのサーバーへの着信トラフィックの配信を停止し、正常なサーバーにトラフィックを配信します。バックエンドサーバーが復帰すると、Server Load Balancerにより自動的にトラフィックの配信が再開されます。

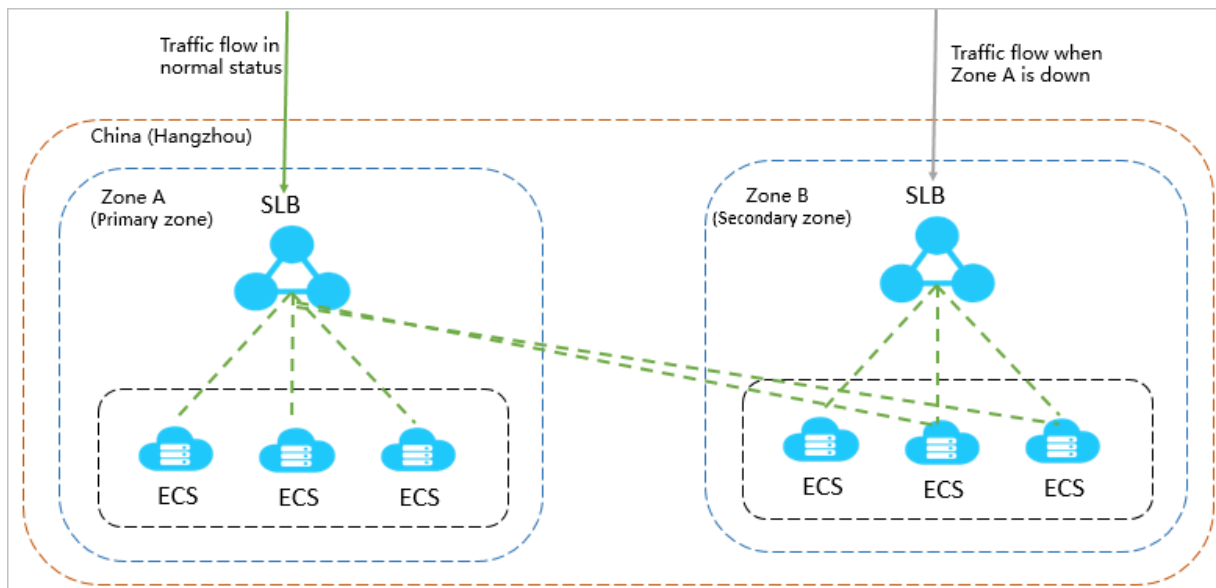
### 複数のゾーンでより優れた耐災害性を実現

より信頼性の高いサービスを提供するために、Server Load Balancerはほとんどのリージョンでの複数のゾーンにデプロイされています。プライマリゾーンが使用できなくなる場合、スタンバイゾーンのServer Load Balancerは30秒以内に負荷分散サービスを引き継ぐことができます。プライマリゾーンが使用可能になると、Server Load Balancerは自動的にプライマリゾーンに戻ります。

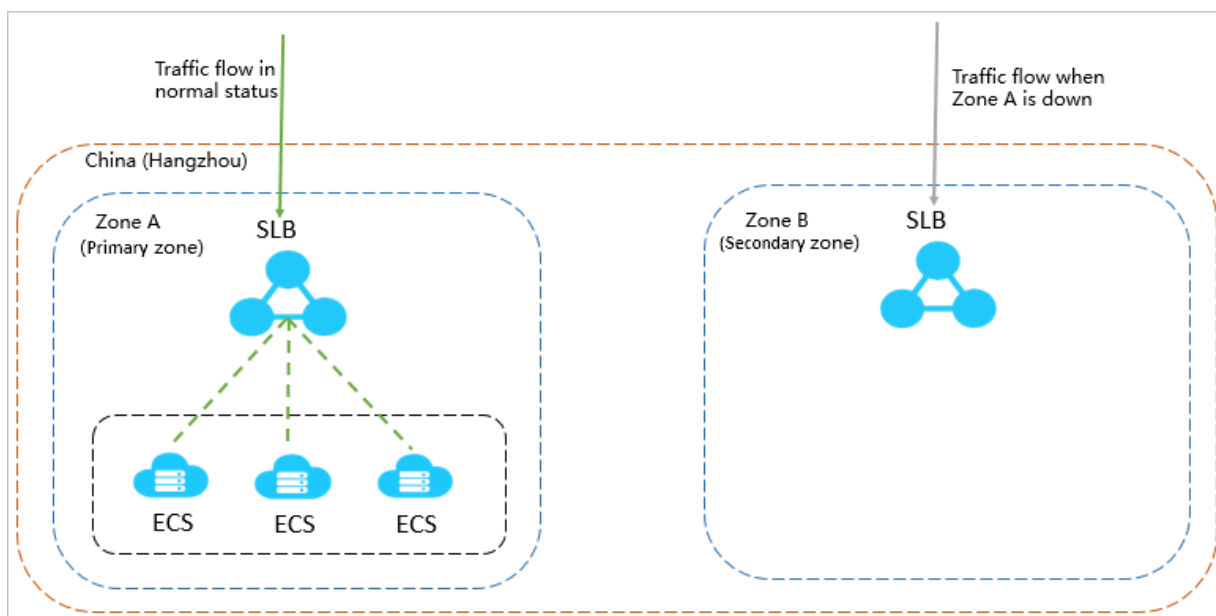
Server Load Balancerインスタンスを作成してリージョン内の複数のゾーンにデプロイすることをお勧めします。また、バックエンドサーバーのデプロイメントにも考慮する必要があります。効率向上のため、各ゾーンに少なくとも1つのバックエンドサーバーを追加することをお勧めします。

次の図に示すように、プライマリゾーン A が正常に動作している場合、青の線のように各ゾーンの ECS インスタンスにトラフィックが分散されます。プライマリゾーン A が使用できなくなると、黒の点線のようにバックアップゾーンの ECS インスタンスにトラフィックが分散されます。

これにより、1つのゾーンの障害によるビジネスの中断が回避されると共に、さまざまなゾーンのプロダクト間のレイテンシが減ります。



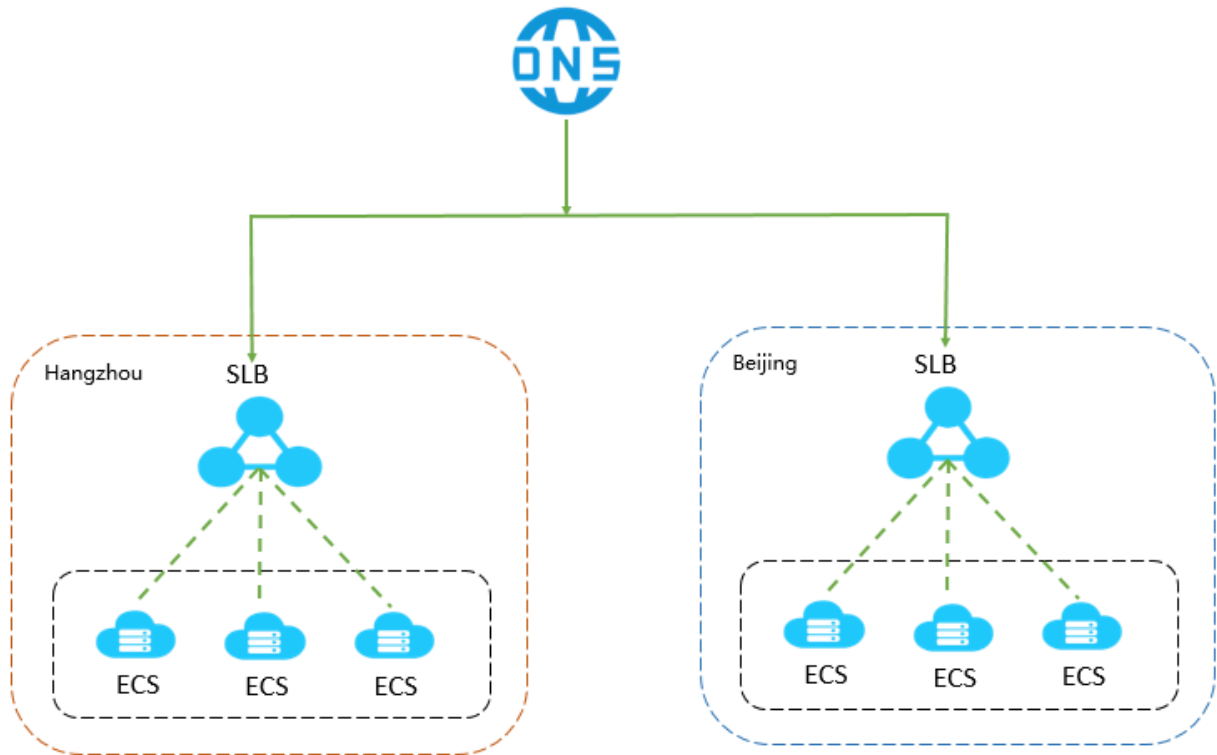
ただし、すべてのECSインスタンスをプライマリゾーンにデプロイし、スタンバイゾーンにはECSインスタンスがない場合は、スタンバイゾーンに分散化されたリクエストを処理するECSインスタンスがないため、プライマリゾーンが利用できない場合はサービスが中断されます。このようなデプロイメントは、高可用性を犠牲にして低遅延が達成されます。



### クロスリージョンによるより優れた耐障害性を実現

複数のゾーンに複数のSLBインスタンスを作成し、対応したリージョンにECSインスタンスをバックエンドサーバーとして追加することができます。そしてAlibaba Cloud DNSを使用し

て、ドメインを作成されたSLBインスタンスのIPアドレスに解決します。リージョンが使用できなくなる場合、DNSは自動的にドメインをスタンバイServer Load Balancerに解決します。



## 5 プロダクト用語

用語	説明
Server Load Balancer	Alibaba Cloud Server Load Balancerとはトラフィックの負荷分散制御機能で、設定されたスケジューリングアルゴリズムとリスニングルールに基づき複数のECS間における受信トラフィックの負荷分散を行います。
Server Load Balancer instance	Server Load BalancerインスタンスとはServer Load Balancerのサービスとして起動しているインスタンスのことです。Server Load Balancerを使用するには初めにServer Load Balancerを作成する必要があります。
サービスアドレス	SLBインスタンスに割り当てられたIPアドレスのことです。インスタンスタイプの違いによっては、IPアドレスはパブリックIP、プライベートIPのいずれかになります。外部サービスを提供するためにパブリックIPの名前解決を行うことができます。
リスナー	リスナーとは受信リクエストがどのように配信されるかを定義付けるものです。ユーザは少なくとも一つのリスナーをServer Load Balancerインスタンスに追加しなければなりません。
バックエンドサーバー	バックエンドサーバとは配信されたリクエストを処理するECSインスタンスです。
デフォルトサーバーグループ	配信されたリクエストを受信するECSインスタンスグループです。 リスナーがVServerグループまたはアクティブ/スタンバイサーバーグループを設定しない場合は、デフォルトサーバーグループが使用されます。着信トラフィックは、デフォルトサーバーグループのECSインスタンスに配信されます。
VServerグループ	配信されたリクエストを受信するECSインスタンスグループです。 他のリスナーまたは転送ルールが、複数のVServerグループを構成して、様々なリクエストを複数バックエンドサーバに配信できます。
アクティブ/スタンバイサーバーグループ	アクティブ/スタンバイサーバーグループには2つだけのECSインスタンスが含まれています。そのうちの1つはアクティブサーバーで、もう1つはスタンバイサーバーです。アクティブ/スタンバイサーバーグループはTCPとUDPリスナーにのみ使用できます。アクティブサーバーのヘルスチェックが失敗した場合、トラフィックがServer Load Balancerにより自動的にスタンバイサーバーに切り替わります。