Cloud Load Balancer

Product Introduction

Product Documentation
Copyright Notice

© 2013-2019 Tencent Cloud. All rights reserved.

Copyright in this document is exclusively owned by Tencent Cloud. You must not reproduce, modify, copy or distribute in any way, in whole or in part, the contents of this document without Tencent Cloud's the prior written consent.

Trademark Notice

© Tencent Cloud

All trademarks associated with Tencent Cloud and its services are owned by Tencent Cloud Computing (Beijing) Company Limited and its affiliated companies. Trademarks of third parties referred to in this document are owned by their respective proprietors.

Service Statement

This document is intended to provide users with general information about Tencent Cloud's products and services only and does not form part of Tencent Cloud's terms and conditions. Tencent Cloud’s products or services are subject to change. Specific products and services and the standards applicable to them are exclusively provided for in Tencent Cloud's applicable terms and conditions.
Contents

Product Introduction
  Overview
  Strengths
  Use Cases
  Principles
  Restrictions
  Instance Types
CLB Overview

Tencent Cloud Load Balancer (CLB) is a service that distributes traffic to multiple CVM instances so as to elevate service capabilities of an application system and eliminate single points of failure for a higher availability.

CLB virtualizes multiple CVM instances in the same region into a high-performance and high-availability application service pool by setting a virtual IP address (VIP) and then distributes the network requests from clients to the pool in the manner specified by the application.

CLB checks the health of the instances in the pool and automatically isolates unhealthy ones, solving the problem with single points of failure and improving the service capabilities of the application systematically.

CLB has advanced features such as self-service management, self-healing, and network attack prevention, making it ideal for various application scenarios ranging from business and community to ecommerce and gaming.

Components

A CLB cluster generally consists of the following components:

- CLB instance: CLB instance used for distributing traffic.
- Virtual IP (VIP): IP address through which CLB provides service to clients.
- Backend/Real server: backend CVM instance used for processing requests.
- VPC/Basic network: overall network environment.

Access requests from outside CLB will be distributed to real servers for processing by the CLB instance based on the configured policy and forwarding rule.

Glossary
<table>
<thead>
<tr>
<th>Term</th>
<th>Meaning</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLB</td>
<td>Cloud load balancer</td>
<td>It is a CLB service offered by Tencent Cloud, which provides TCP/UDP- and HTTP-based network load balancing services in conjunction with CVM instances.</td>
</tr>
<tr>
<td>CLB listener</td>
<td>Load balancer listener</td>
<td>It is the listener of a CLB service and consists of listening ports, load balancing policies, and health check configurations, where each listener corresponds to an application service on the backend.</td>
</tr>
<tr>
<td>RS</td>
<td>Real server</td>
<td>It is a CVM instance that accepts the requests distributed by CLB which forwards the access requests to a set of backend CVM instances for processing according to user-defined rules.</td>
</tr>
<tr>
<td>VIP</td>
<td>Virtual IP</td>
<td>It is a service address assigned by the system, which currently is an IP address. You can choose whether to open it to the internet so as to create a public or private network CLB instance.</td>
</tr>
</tbody>
</table>

**How It Works**

**Basic principle**
A CLB instance receives inbound traffic from clients and routes requests to real servers in one or multiple AZs for processing.

The load balancing service is mainly provided by a CLB listener which listens on the CLB instance for requests and executes corresponding policies to distribute requests to services such as real servers. You can configure the forwarding protocol and protocol port in **client - CLB and CLB - real server** dimensions to have CLB directly forward requests to real servers.

You are recommended to configure backend CVM instances in multiple AZs for a CLB instance. In this way, if one AZ becomes unavailable, the CLB instance will route the traffic to other normal instances in other AZs so as to avoid service interruption caused by AZ failure.

**Selecting request route**
A client request accesses the service through a domain name. Before a request is sent to a CLB instance, the DNS server will resolve the CLB domain name and return to the client the CLB IP address that receives the request. When the CLB listener receives requests, it will use different CLB algorithms to distribute them to real servers. Currently, Tencent Cloud supports multiple load balancing algorithms such as WRR, ip_hash, and weighted least-connection scheduling.

**Monitoring real server status**
A CLB instance can monitor the running status of real servers to ensure that traffic is routed to only healthy servers. When the CLB instance detects that a real server is exceptional, it will stop routing traffic to the server and resume routing after detecting that the server runs normally again.

Related Services

CLB can be used together with the following services to improve the availability and scalability of applications:

- **CVM instance**: it is a virtual server for applications to run in the cloud.
- **Auto Scaling**: it automatically controls the instance quantity. If CLB instances are enabled in Auto Scaling, instances scaled out will be automatically added to the CLB cluster, while instances scaled in will be automatically removed from the cluster.
- **Cloud Monitor**: it helps you monitor running status of CLB instances and all real servers and allows you to perform relevant operations.
Strengths

CLB performance is evaluated mainly based on the following metrics:

- **TPS (connection per second):** the number of TCP connections created by a CLB instance per second.
- **Maximum number of concurrent connections:** the maximum number of connections that a CLB instance can sustain at the same time.
- **QPS (query per second):** also known as request per second (RPS). It is the number of GET, POST, and HEAD packets sent by the client to the HTTP service after a connection is established.
- **Throughput:** the total traffic/bandwidth supported by a CLB instance.

Tencent Cloud provides high-performance CLB services:

- A single CLB cluster can provide over 120 million concurrent connections and respond to hundreds of millions of web access requests.
- A single CLB cluster can handle a peak traffic of 40 GB/s with up to 6 million PPS (packets per second).
- CLB strictly isolates the traffic of each tenant and provides active protection against DDoS attacks.

If you have higher protection requirements, you can purchase Anti-DDoS Pro, which helps defend against up to 300 GB of attack traffic.

To protect the application layer, you can purchase Tencent Cloud Web Application Firewall (WAF). It protects web security at the application layer against web vulnerability attacks, malicious crawlers, and CC attacks.
Use Cases

Last updated: 2020-03-09 15:04:07

CLB is mainly suitable for the following scenarios:

- **Traffic distribution.** CLB can distribute the traffic of business with a large number of access requests to multiple CVM instances.
- **Elimination of single point of failure.** When some CVM instances become unavailable, CLB can automatically block them to ensure the normal operation of the application system.
- **Horizontal scalability.** You can scale out the service capability of application systems as needed, which is suitable for web and app servers.
- **Global load balancing.** With Tencent Cloud DNS, CLB can support global and multi-regional load balancing for remote disaster recovery.

Traffic distribution and elimination of single point of failure

You can use CLB to distribute business traffic to multiple CVM instances.

- **Business client accesses CLB.**
- **Multiple CVM instances form a high-performance and high-availability service pool,** to which CLB forwards the business traffic.
- **When one or more CVM instances become unavailable,** CLB can automatically block them and distribute the requests to normal CVM instances, ensuring the operation of the application system.
- **The session persistence feature** can forward requests from the same client to the same real server, improving access efficiency.

Horizontal Scalability

With **Auto Scaling (AS)**, CLB can automatically create and release CVM instances based on your business needs.

- **You can configure auto scaling policies** to manage the number of CVM instances, deploy the instance environment, and ensure the operation of your business. CLB can automatically add CVM instances when demands peak to keep high performance, while removing CVM instances when demands drop to reduce costs.
- **For example, during major sales campaigns in ecommerce such as Black Friday,** web traffic may suddenly increase by 10 times and lasts only for a few hours. In this case, CLB and AS can be used.
to maximize IT cost savings.

Global Load Balancing

With Tencent Cloud DNS, you can resolve your business traffic to global and multi-regional CLB instances to implement active-active and remote disaster recovery across regions.

- You can deploy CLB instances in different regions and bind them to CVM instances in corresponding regions.
- You can use Tencent Cloud DNS to resolve domain names to the CLB VIP in each region.
- Business traffic will be forwarded to multiple CVM instances in multiple regions via DNS and CLB, achieving global load balancing.
- When a region becomes unavailable, you can suspend resolution of the CLB VIP in that region to ensure that your business is not affected.
Principles

CLB provides Layer-4 (TCP, UDP, and TCP SSL protocols) and Layer-7 (HTTP and HTTPS protocols) load balancing. You can use CLB to distribute business traffic to multiple real servers to eliminate single point of failure and guarantee business availability. CLB adopts cluster deployment to achieve session synchronization, eliminating server’s single point of failure and improving system redundancy to ensure service stability. CLB can be deployed in multiple data centers in the same region to implement intra-city disaster recovery.

Infrastructure

Currently, Tencent Cloud CLB provides Layer-4 and Layer-7 load balancing services:

- At Layer-4, load balancing is implemented based on the unified Tencent Gateway (TGW). TGW has features such as high availability, high scalability, high performance, and strong anti-attack capability. It supports high-performance forwarding based on Data Plane Development Kit (DPDK). With TGW, a single cluster can support hundreds of millions of concurrent requests and tens of millions of packets per second (PPS). Many Tencent businesses, such as Tencent Games, Tencent Video, WeChat, and QQ, use TGW for service access.

- At Layer-7, load balancing is implemented based on Secure Tencent Gateway (STGW). It is a load balancing service developed by Tencent based on Nginx that supports large-scale concurrence. It carries a large amount of Tencent’s Layer-7 business traffic, such as Tencent News, Licaitong,
Forwarding Path

CLB forwards business traffic and real servers process business requests. CLB communicates with real servers via Tencent Cloud private network. Both TGW and STGW are deployed on multiple servers, and provide load balancing services through clusters. The forwarding path of CLB is as
1. TCP and UDP protocol:
   - The forwarding logic of TCP/UDP protocol is processed by TGW cluster.
   - After receiving the business traffic, TGW forwards it to real servers via Tencent Cloud's private network. The return packets from real servers are also returned to the client via TGW.

2. TCP SSL protocol
   - When TCP SSL protocol is processed, business traffic passes through the TGW cluster and then STGW cluster, which forwards the traffic to real servers.
   - Before a new session is established, it must pass through the accelerator card cluster for certificate verification, encryption, decryption and other operations.
   - When business traffic arrives, it passes through TGW, STGW, and real servers in sequence via Tencent Cloud's private network. The return packets are sent to the client in reverse sequence.

3. HTTP and HTTPS protocols
   - When HTTP or HTTPS protocol is processed, business traffic passes through the TGW cluster and then STGW cluster, which identifies the HTTP protocol and forwards the traffic to real servers.
   - Before a new HTTPS session is established, it must pass through the accelerator card cluster for certificate verification, encryption, decryption and other operations. HTTPS will be converted to HTTP protocol and then forwarded to real servers.
   - When business traffic arrives, it passes through TGW, STGW, and real servers in sequence via Tencent Cloud's private network. The return packets are sent to the client in reverse sequence.
The use of Tencent Cloud CLB has certain restrictions, and different types of CLB instances have their own use limits. For more information on CLB instance types, please see [Instance Types](#).

<table>
<thead>
<tr>
<th>Instance Type</th>
<th>Resource</th>
<th>Default Restriction</th>
</tr>
</thead>
<tbody>
<tr>
<td>General restrictions for all instances</td>
<td>Number of public network instances can be created under one account in a single region</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Number of private network instances can be created under one account in a single region</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Number of listeners can be added to an instance</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>Ports that can be selected by a listener in an instance</td>
<td>An integer between 1 - 65535</td>
</tr>
<tr>
<td>CLB (formerly &quot;application CLB&quot;)</td>
<td>Number of domain name and URL forwarding rules can be configured for an HTTP/HTTPS listener in a CLB instance</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>Number of servers can be bound to a forwarding rule in a CLB instance</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Number of backend ports that can correspond to a frontend port of a CLB instance</td>
<td>Multiple ports</td>
</tr>
<tr>
<td>Classic CLB</td>
<td>Number of servers can be bound to a listener in a classic CLB instance</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Number of backend ports that can correspond to a frontend port of a classic CLB instance</td>
<td>1 port</td>
</tr>
</tbody>
</table>

A CLB instance **will not unbind itself** from the CVM instance. After a CVM instance becomes isolated (pay-as-you-go CVM instance has been in arrears for more than 2 hours), it **will not unbind itself** from the CLB instance either.
CLB offers two types of instances: CLB (formerly "application CLB") and Classic CLB.

- CLB supports TCP/UDP/HTTP/HTTPS protocols to provide flexible forwarding capabilities based on domain names and URL paths.
- Classic CLB does not support HTTP/HTTPS protocols on the private network but is easy to configure.

CLB has all features of Classic CLB. Given their product features and performance, we recommend that you use CLB. For a detailed comparison, please see below:

<table>
<thead>
<tr>
<th>Product Type</th>
<th>CLB</th>
<th>Classic CLB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Public network</td>
<td>Private network</td>
</tr>
<tr>
<td>Layer-7 forwarding (HTTP/HTTPS)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Layer-4 forwarding (TCP/UDP)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Encrypted Layer-4 forwarding (TCP SSL)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>HTTP/2 and WebSocket (Secure) support</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Load balancing policy</td>
<td>IP hash (Layer-7) weighted round-robin</td>
<td>IP hash (Layer-7) weighted round-robin</td>
</tr>
<tr>
<td></td>
<td>IP hash (Layer-7) weighted least-connection scheduling</td>
<td>IP hash (Layer-7) weighted least-connection scheduling</td>
</tr>
<tr>
<td>Session</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Feature</td>
<td>Classic CLB Instance</td>
<td>Custom CLB Instance</td>
</tr>
<tr>
<td>------------------------------------------------</td>
<td>----------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Persistence</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Health check</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Custom forwarding rule (domain name/URL)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>SNI multi-certificate support</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Forwarding to different real ports</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Custom Layer-7 configuration</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Layer-7 redirect (rewrite)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Cross-region binding support</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Storing Layer-7 access logs in CLS</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

**Note:**
- CLB instance: a CLB instance supports enabling or disabling the HTTP/2 protocol. For more information, see [Configuring an HTTPS Listener](#).
- Classic CLB instance: HTTPS listeners created for Classic CLB before April 2018 do not support the HTTP/2 protocol. HTTPS listeners created after April 2018 support but cannot disable the HTTP/2 protocol.