# 负载均衡

# 产品简介

腾讯云

版权所有：腾讯云计算（北京）有限责任公司

【版权声明】

©2013-2025 腾讯云版权所有

本文档（含所有文字、数据、图片等内容）完整的著作权归腾讯云计算（北京）有限责任公司单独所有，未经腾讯云事先明确书面许可，任何主体不得以任何形式复制、修改、使用、抄袭、传播本文档全部或部分内容。前述行为构成对腾讯云著作权的侵犯，腾讯云将依法采取措施追究法律责任。

【商标声明】

腾讯云

及其它腾讯云服务相关的商标均为腾讯云计算（北京）有限责任公司及其关联公司所有。本文档涉及的第三方主体的商标，依法由权利人所有。未经腾讯云及有关权利人书面许可，任何主体不得以任何方式对前述商标进行使用、复制、修改、传播、抄录等行为，否则将构成对腾讯云及有关权利人商标权的侵犯，腾讯云将依法采取措施追究法律责任。

【服务声明】

本文档意在向您介绍腾讯云全部或部分产品、服务的当时的相关概况，部分产品、服务的内容可能不时有所调整。您所购买的腾讯云产品、服务的种类、服务标准等应由您与腾讯云之间的商业合同约定，除非双方另有约定，否则，腾讯云对本文档内容不做任何明示或默示的承诺或保证。

【联系我们】

我们致力于为您提供个性化的售前购买咨询服务，及相应的技术售后服务，任何问题请联系 4009100100或95716。

# 文档目录

产品简介
- 产品概述
- 产品优势
- 使用场景
- 技术原理
- 产品对比
  - 实例类型对比
  - 实例规格对比
- 使用约束

版权所有：腾讯云计算（北京）有限责任公司

# 产品简介
# 产品概述

最近更新时间：2024-10-11 11:08:01

您可通过如下视频了解负载均衡的概念、组成以及工作原理。
观看视频

## 什么是负载均衡

负载均衡（Cloud Load Balancer，CLB）是对多台后端服务器进行流量分发的服务。负载均衡可以通过流量分发扩展应用系统对外的服务能力，通过消除单点故障提升应用系统的可用性。
负载均衡服务通过设置虚拟服务地址（virtual IP，VIP），将位于**同一地域**的多台后端服务器资源虚拟成一个高性能、高可用的应用服务池。根据应用指定的方式，将来自客户端的网络请求分发到服务器池中。
负载均衡服务会检查服务器池中后端服务器实例的健康状态，自动隔离异常状态的实例，从而解决了后端服务器的单点问题，同时提高了应用的整体服务能力。
腾讯云提供的负载均衡服务具备自助管理、自助故障修复，防网络攻击等高级功能，适用于企业、社区、电子商务、游戏等多种用户场景。

## 组成部分

一个提供服务的负载均衡组通常由以下部分组成：

- **负载均衡实例**：一个 CLB 实例是一个运行的负载均衡服务，用于流量分发。
- **监听器**：用来检查客户端请求并将请求转发给后端服务器。
- **后端服务器**：后端的一组服务器实例，用于接收前端的请求。

来自负载均衡外的访问请求，通过负载均衡实例并根据相关的策略和转发规则分发到后端服务器进行处理。

## 名词解释

| 术语 | 全称 | 说明 |
| --- | --- | --- |
| 负载均衡器 | Cloud Load Balancer | 腾讯云提供的一种负载均衡服务。用户可以使用 CNAME 或 A 记录解析对外提供访问。其中仅公网负载均衡可以选择 CNAME 方式。 |
| 负载均衡监听器 | Load Balance Listener | 负载均衡服务监听器，包括监听端口、负载均衡策略和健康检查配置等，每个监听项对应后端的一个应用服务。 |
| 后端服务器 | Real Server | 接受负载均衡分发请求的一组后端服务器实例，负载均衡服务将访问请求按照用户设定的规则转发到这一组后端服务器上进行处理。 |

## 工作原理

## 基本工作原理

负载均衡器接受来自客户端的传入流量，并将请求路由到一个或多个可用区的后端服务器实例上进行处理。
负载均衡服务主要由负载均衡监听器提供。监听器负责监听负载均衡实例上的请求、执行策略分发至后端服务器等服务，通过配置**客户端 – 负载均衡**和**负载均衡 – 后端服务器**两个维度的转发协议及协议端口，负载均衡可以将请求直接转发到后端服务器上。
建议您跨多个可用区配置负载均衡器的后端服务器实例。如果一个可用区变得不可用，负载均衡器会将流量路由到其他可用区正常运行的实例上去，从而避免可用区故障引起的服务中断问题。

## 请求路由选择

客户端请求通过域名访问服务，在请求发送到负载均衡器之前，DNS 服务器将会解析负载均衡域名，并将收到请求的负载均衡 IP 地址返回到客户端。当负载均衡监听器收到请求时，将会使用不同的负载均衡算法将请求分发到后端服务器中。目前腾讯云支持加权轮询和 ip_hash 加权最小连接数等多种均衡算法。

## 监控后端服务状态

负载均衡器还可以监控后端实例的运行状况，从而确保只将流量路由到正常运行的实例上去。当负载均衡器检测到运行不正常的实例时，它会停止向该实例路由流量，然后会在它再次检测到实例正常运行之后重新向其路由流量。

# 相关服务

负载均衡与以下服务一起使用，可以提高应用程序的可用性和可扩展性：

- CVM 实例：应用程序在云上运行的虚拟服务器。
- 弹性伸缩：弹性地控制实例数量。在弹性伸缩中启用负载均衡实例，则伸缩的实例将自动加入负载均衡组，同时终止的实例将自动被移出负载均衡组。
- 腾讯云可观测平台：帮助您监控负载均衡及所有后端实例的运行状况并执行所需操作。
- 域名注册 和 云解析 DNS：通过将您自定义的域名（如 `www.example.com` ）转换为网络通信所用的 IP 地址（如 `192.0.2.1` ），快速便捷地将请求路由至负载均衡实例。

# 产品优势

最近更新时间：2025-05-23 17:06:12

本文介绍负载均衡的产品优势。

## 高性能

CLB 单集群（非单个 CLB 实例）支持亿级并发连接数，每秒处理百万级的包数据量，轻松应对日访问量超过千万的电商网站、社交平台和游戏业务等。

## 高可用

CLB 采用集群化部署，可用性高达99.95%，域名化实例 SLA 从 99.95%提升至 99.99%，详情请参见 域名化实例上线公告 。在单台 CLB 物理服务器可用的极端情况下，仍可支撑千万级的并发连接数。同时，集群系统会及时剔除故障实例，筛选出健康实例，确保后端服务器业务正常运行。

## 高弹性

CLB 集群根据业务负载横向伸缩，弹性扩展应用系统对外的服务能力，同时借助弹性伸缩（Auto Scaling）的动态伸缩组自动创建和释放 CVM 实例，结合动态监控情况和秒级计费系统，您无需手工干预和预估资源，即可实现您的计算资源合理分配，防止资源浪费。

## 安全稳定

CLB 依靠大禹分布式防御系统能够防御绝大多数网络攻击（例如 DDoS、Web 入侵），针对流量攻击实现秒级清洗，极大避免 IP 被封、带宽被占满等情况发生。CLB 自带的 synproxy 防攻击机制，避免了大禹系统生效之前后端 CVM 被攻击压垮，保护数据更安全稳定。
CLB 对每个租户的流量进行严格隔离，提供主动 DDoS 防护能力，公网 CLB 默认支持 DDoS 基础防护 。CLB 还支持 DDoS 高防包 、DDoS 高防 IP 、Web 应用防火墙 等多种安全产品，保障业务安全性。

> **说明：**
> - 如果您对峰值流量有更高的要求，可以另行购买 DDoS 高防包，腾讯云提供不低于30 Gbps 的 DDoS 防护能力， 最高防护能力根据各个区域的实际网络情况动态调整。
> - 如果您对应用层安全防护有需求，可以另行购买 腾讯云 Web 应用防火墙（WAF） ，WAF 将为您提供应用层 Web 安全防护，抵御 Web 漏洞攻击、恶意爬虫和 CC 攻击等行为，保护网站和 Web 应用安全。

## 低成本

使用腾讯云 CLB，您无需再投入额外的负载均衡硬件，无需投入过多繁琐的运维工作，为您节省硬件费用和人力支出。CLB 支持多种计费模式，可按需选择计费模式。

版权所有：腾讯云计算（北京）有限责任公司

# 使用场景

最近更新时间：2024-12-02 16:45:22

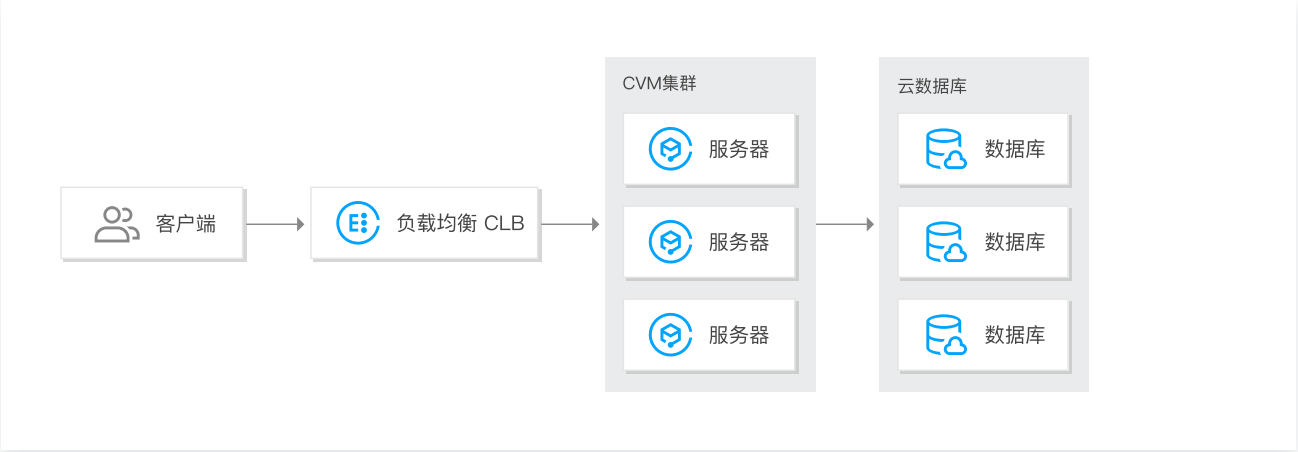
负载均衡主要适用于如下场景：

- 流量分发，将高访问量的业务通过负载均衡分发到多台后端服务器上。
- 消除单点故障，当其中一部分后端服务器不可用时，负载均衡可自动屏蔽故障的 CVM 实例，保障应用系统正常工作。
- 横向扩展，根据业务发展的需要，按需扩展应用系统的服务能力，适用于各种 Web Server和 App Server。
- 全局负载均衡，结合 云解析 DNS ，可支持全局多地域负载均衡，保障异地容灾。

## 流量分发和消除单点故障

您可以通过负载均衡，将业务流量分发到多台后端服务器上：

- 业务的客户端访问负载均衡。
- 多台后端服务器构成一个高性能、高可用的服务池，负载均衡将业务流量转发到这批后端服务器上。
- 当某台或某几台后端服务器不可用时，负载均衡可自动屏蔽故障的 CVM 实例，将请求分发给正常运行的 CVM 实例，保障应用系统正常工作。
- 若您的业务部署在多个可用区，建议您将一个负载均衡实例同时与多个可用区的 CVM 实例进行绑定，以便在后端服务器层保障多可用区容灾。
- 会话保持功能可将同一客户端的请求转发到同一台后端服务器，提高访问效率。

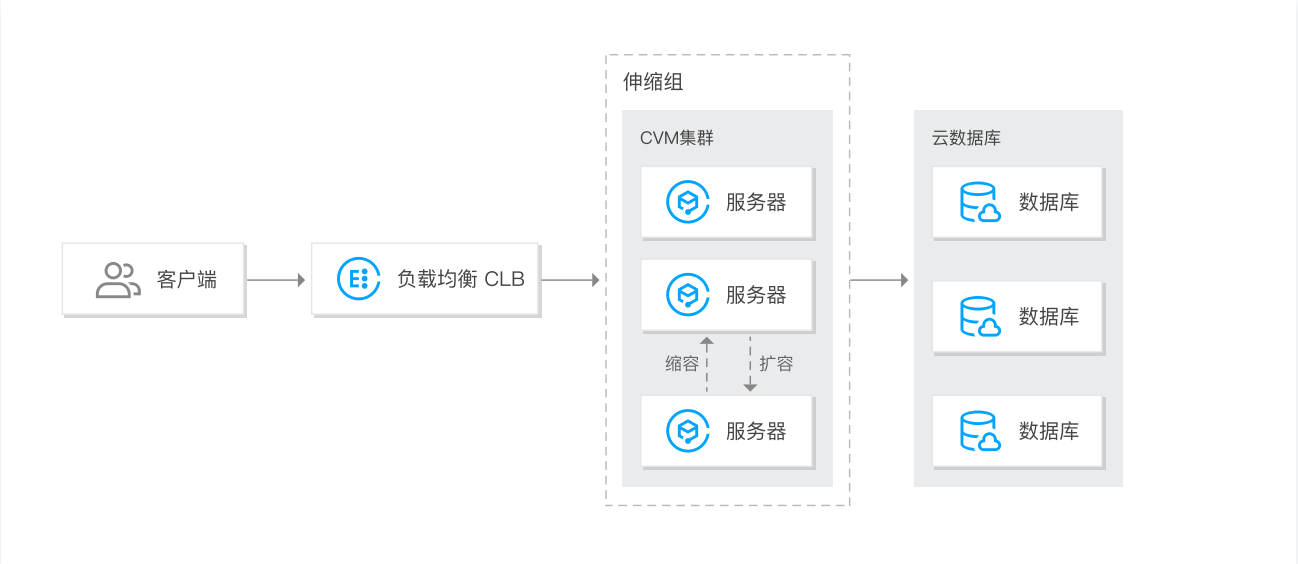



## 横向扩展

负载均衡结合 弹性伸缩 ，可为您按需创建和释放 CVM 实例。

- 您可以设定弹性伸缩策略来管理 CVM 实例数量，完成对实例的环境部署，并保证业务平稳顺利运行。在需求高峰时，自动增加 CVM 实例数量，以保证性能不受影响。当需求较少时，则会减少 CVM 实例数量以降低成本。
- 电商行业的“双11”、“6.18”等大促活动，Web 访问量可能瞬间陡增10倍，且只持续短暂的数小时。使用负载均衡及弹性伸缩能最大限度的节省 IT 成本。

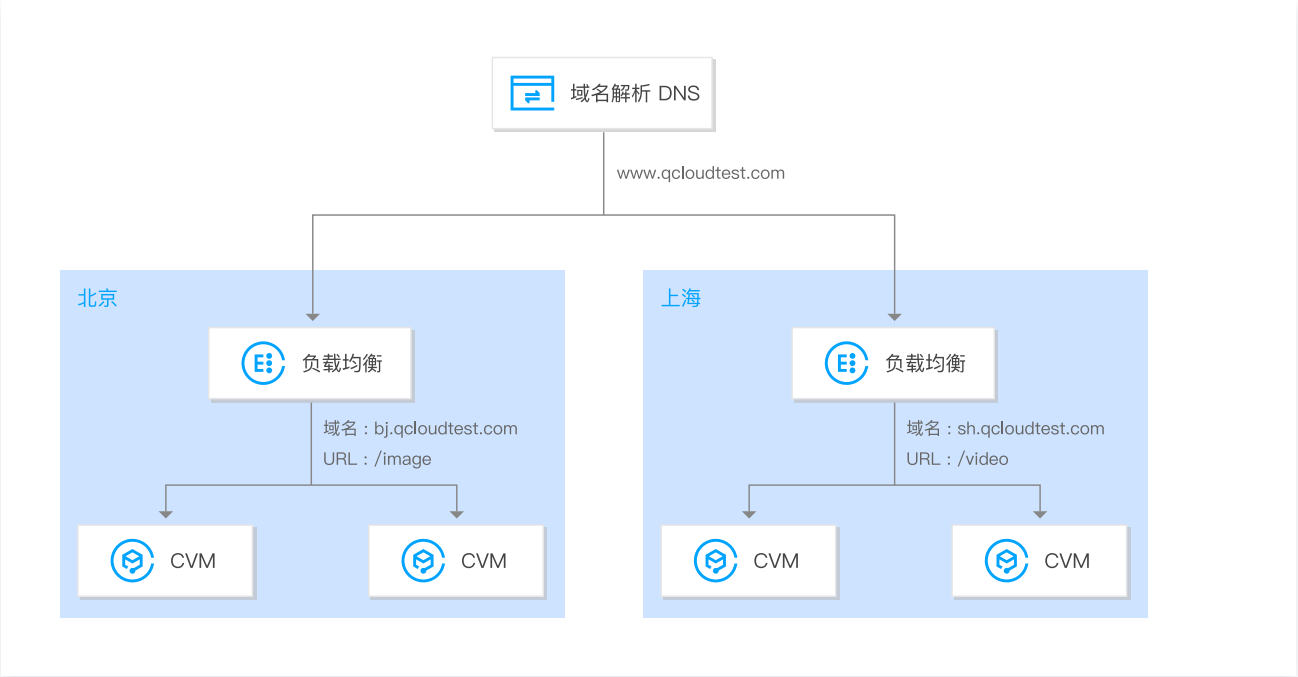
版权所有：腾讯云计算（北京）有限责任公司




## 全局负载均衡

结合云解析 DNS，您可以将业务流量解析到全局各个地域的负载均衡，保障异地多活和容灾。

- 您可以在不同地域部署负载均衡实例，并分别绑定对应地域的后端服务器。
- 使用云解析 DNS 将域名解析到各个地域的负载均衡 VIP 下。
- 业务流量会通过域名解析和负载均衡转发到多个地域的多个后端服务器上，以此实现全局负载均衡。
- 当某个地域不可用时，暂停对应地域负载均衡 VIP 的解析即可保障业务不受影响。



# 技术原理

最近更新时间：2024-12-02 16:45:22

四层负载均衡原理及应用

七层负载均衡原理及应用

负载均衡的轮询方式和高可靠性

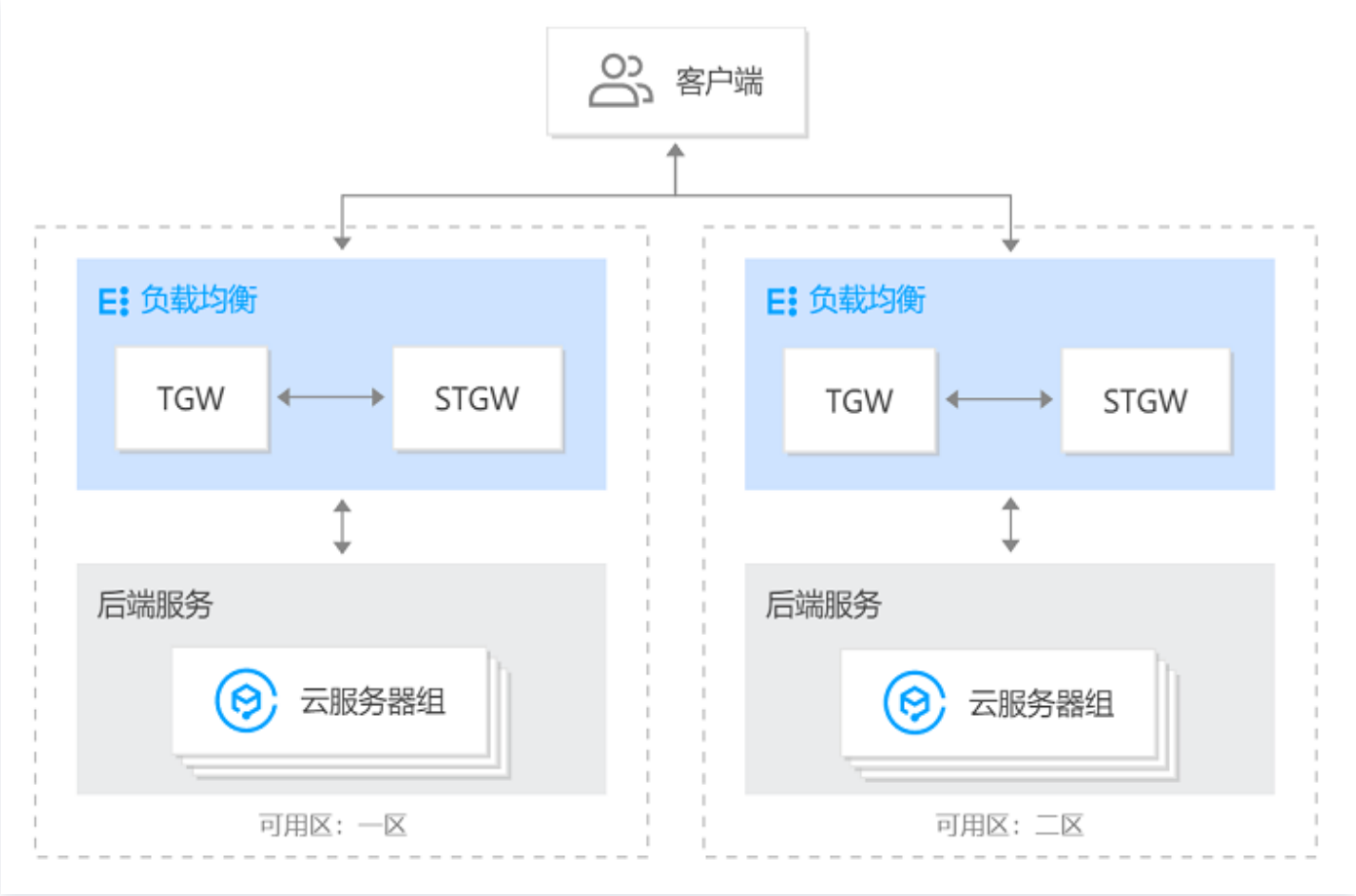
负载均衡 CLB 提供四层（TCP 协议/UDP 协议/TCP SSL 协议）和七层（HTTP 协议/HTTPS 协议）负载均衡服务。您可以通过负载均衡 CLB 将业务流量分发到多个后端服务器上，消除单点故障并保障业务可用性。负载均衡 CLB 自身采用集群部署，可实现会话同步，消除服务器单点，提升系统冗余，保证服务稳定，可在同一个地域部署多个机房，实现同城容灾。

## 基础架构

腾讯云负载均衡 CLB 当前提供四层和七层的负载均衡服务：

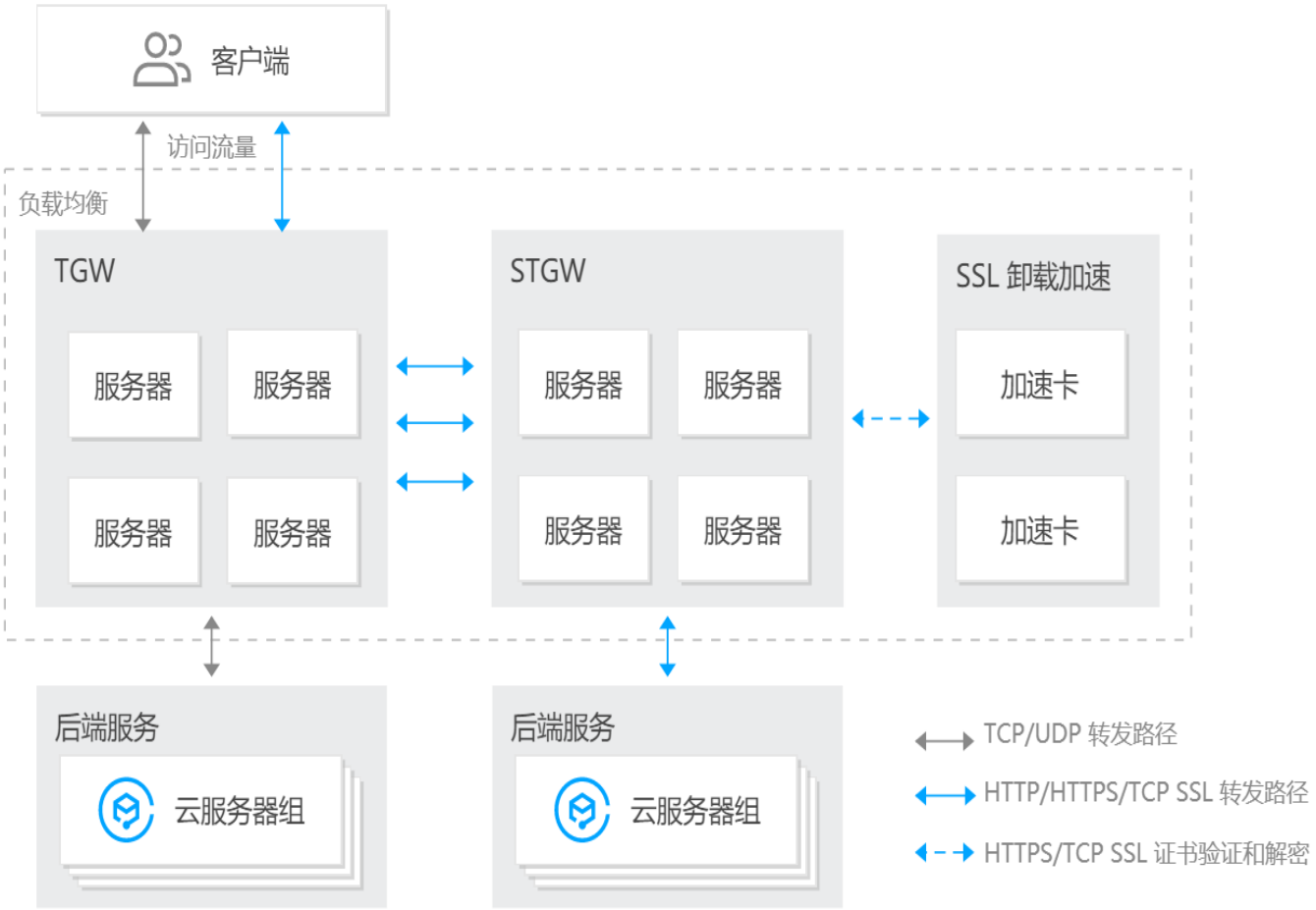
- 四层主要基于腾讯自研的统一接入网关（Tencent Gateway，TGW）来实现负载均衡，TGW 具有可靠性高、扩展性强、性能高、抗攻击能力强等特点，支持 Data Plane Development Kit（DPDK）高性能转发，单集群可支持亿级并发、千万级 PPS（Packets Per Second）。腾讯内部诸多业务均通过 TGW 接入服务，包括腾讯游戏、腾讯视频、微信、QQ 等。
- 七层主要基于 Secure Tencent Gateway（STGW）实现负载均衡，STGW 是腾讯基于 Nginx 自研的支持大规模并发的七层负载均衡服务，承载了腾讯内大量的七层业务流量，包括腾讯新闻、理财通、腾讯游戏、微信等。

版权所有：腾讯云计算（北京）有限责任公司




## 转发路径

负载均衡 CLB 负责转发业务流量，由后端服务实际处理业务请求。负载均衡 CLB 与后端 CVM 之间是通过腾讯云内网进行通信的。TGW 和 STGW 均由多台服务器部署，通过集群来提供负载均衡服务。负载均衡 CLB 的转发路径如下图所示：




1. TCP/UDP协议：
   - TCP/UDP 协议由 TGW 集群处理转发逻辑。
   - 业务流量到达后，由 TGW 通过腾讯云内网转发给后端 CVM，后端 CVM 的回包也是通过 TGW 返回给客户端。
2. TCP SSL 协议
   - 处理 TCP SSL 协议时，业务流量会先经过 TGW 集群而后由 STGW 集群来转发给后端 CVM。
   - 新建会话时需经过加速卡集群来进行证书的验证和加解密等前置操作。
   - 业务流量到达后，在腾讯云内网中依次通过 TGW、STGW、后端 CVM，回包也依次逆向返回给客户端。
3. HTTP/HTTPS 协议
   - 处理 HTTP/HTTPS 协议时，业务流量会先经过 TGW 集群，而后由 STGW 识别 HTTP 协议并转发给后端 CVM。
   - 新建 HTTPS 会话时需经过加速卡集群来进行证书验证和加解密等前置操作，将 HTTPS 转换成 HTTP 协议，再转发给后端 CVM。
   - 业务流量到达后，在腾讯云内网中依次通过 TGW、STGW、后端 CVM，回包也依次逆向返回给客户端。

# 产品对比
# 实例类型对比

最近更新时间：2024-11-22 16:19:44

负载均衡有两种实例类型：负载均衡（此前亦被称为“应用型负载均衡”）和传统型负载均衡。

- 负载均衡：支持 TCP/UDP/HTTP/HTTPS 协议，提供基于域名和 URL 路径的均衡能力，支持灵活转发。
- 传统型负载均衡：内网不支持 HTTP/HTTPS 协议，配置方法简单。

负载均衡可覆盖传统型负载均衡的所有功能。从产品功能、产品性能等多方面考虑，建议您使用的实例类型是负载均衡。二者的详细对比如下：

| 产品类型 | 负载均衡 | | 传统型负载均衡 | |
|---|---|---|---|---|
| | 公网 | 内网 | 公网 | 内网 |
| 七层转发（HTTP/HTTPS） | ✓ | ✓ | ✓ | × |
| 四层转发（TCP / UDP） | ✓ | ✓ | ✓ | ✓ |
| 四层加密转发（TCP SSL） | ✓ | ✓ | × | × |
| 支持 HTTP/2 及 websocket（secure） | ✓ | ✓ | ✓ | × |
| 负载均衡策略 | IP hash（七层）加权轮询加权最小连接数 | IP hash（七层）加权轮询加权最小连接数 | IP hash（七层）加权轮询加权最小连接数 | 加权轮询 |
| 会话保持 | ✓ | ✓ | ✓ | ✓ |
| 健康检查 | ✓ | ✓ | ✓ | ✓ |
| 自定义转发规则（域名/URL） | ✓ | ✓ | × | × |
| 支持 SNI 多证书特性 | ✓ | ✓ | × | × |

版权所有：腾讯云计算（北京）有限责任公司

| | | | | |
|---|---|---|---|---|
| 转发到不同的后端端口 | ✓ | ✓ | × | × |
| 七层个性化配置 | ✓ | ✓ | × | × |
| 七层重定向功能（rewrite） | ✓ | ✓ | × | × |
| 支持跨地域绑定功能 | ✓ | ✓ | × | × |
| 配置七层访问日志到 CLS | ✓ | ✓ | ✓ | × |

**说明：**

- 负载均衡型实例：您可以选择开启或关闭 CLB 对 HTTP/2 的支持，详情请参考 配置 HTTPS 监听器。
- 传统型负载均衡型实例：2018年4月之前创建的 HTTPS 监听器无法启用 HTTP/2，2018年4月后创建的 HTTPS 监听器可以启用 HTTP/2。传统型负载均衡不支持修改 HTTP/2 的开关。

# 实例规格对比

最近更新时间：2025-05-29 10:27:41

CLB 实例规格分为：共享型和性能容量型。

> **说明：**
> 为保证 CLB 的性能和稳定性，CLB 采用分布式高可靠集群架构，因此限速也为分布式限速。
> CLB 集群内的多台服务器同时承载业务，共享同一个上限，集群内多台服务器的限速配置会随业务负载情况实时调整，以此平衡资源分配、保障服务稳定。各限速指标在接近上限时可能触发限速，实际指标会在限速值上下轻微浮动。建议您按需规划资源使用，若您的业务对限速丢包敏感度较高，建议限速值往上预留一定的空间（如上浮10%）。

| 对比项 | 共享型 | 性能容量型 | |
|---|---|---|---|
| 规格上限 | 每分钟并发连接数：50,000，每秒新建连接数：5,000，每秒查询数：5,000 | 标准型 | 每分钟并发连接数：100,000，每秒新建连接数：10,000，每秒查询数：10,000，带宽上限：2Gbps |
| | | 高阶型1 | 每分钟并发连接数：200,000，每秒新建连接数：20,000，每秒查询数：20,000，带宽上限：4Gbps |
| | | 高阶型2 | 每分钟并发连接数：500,000，每秒新建连接数：50,000，每秒查询数：30,000，带宽上限：6Gbps |
| | | 超强型1 | 每分钟并发连接数：1,000,000，每秒新建连接数：100,000，每秒查询数：50,000，带宽上限：10Gbps |
| | | 超强型2 | 每分钟并发连接数：2,000,000，每秒新建连接数：200,000，每秒查询数：100,000，带宽上限：20Gbps |
| | | 超强型3 | 每分钟并发连接数：5,000,000，每秒新建连接数：500,000，每秒查询数：200,000，带宽上限：40Gbps |
| | | 超强型4 | 每分钟并发连接数：10,000,000，每秒新建连接数：1,000,000，每秒查询数：300,000，带宽上限：60Gbps |

版权所有：腾讯云计算（北京）有限责任公司

| | | |
|---|---|---|
| 限速能力 | • 自2023年07月25日12:00:00起，对于新用户创建的共享型 CLB 实例，将按照实例规格进行限速。<br>• 对于2023年07月25日12:00:00前已经在使用共享型实例的用户，平台侧已于2023年08月25日12:00:00起，对存量用户所使用的共享型 CLB 实例按照实例性能上限进行限速。<br>**说明：**<br>存量实例在执行按规格限速前，超过性能上限保障范围的部分会与其它实例共享集群资源，可能存在性能抢占现象。 | 按照上限进行限速 |
| 计费项 | CLB 实例费、网络费、跨域绑定费。<br>其中：<br>• 仅公网 CLB 实例收取网络费用，内网 CLB 实例不收取网络费用<br>• 跨域绑定费仅开启跨地域绑定功能后会收取，不开启则不收取 | CLB 实例费、网络费、性能容量单位 LCU 费用、跨域绑定费。<br>其中：<br>• 仅公网 CLB 实例收取网络费用，内网 CLB 实例不收取网络费用<br>• 跨域绑定费仅开启跨地域绑定功能后会收取，不开启则不收取 |

**说明：**
若您有更高规格负载均衡需求，请联系您的商务经理申请使用独占集群。独占集群3个月起购。

# 使用约束

最近更新时间：2025-05-13 09:27:22

本文介绍负载均衡的使用限制。

## 通用限制

腾讯云负载均衡使用时有一些通用限制，不同类型的负载均衡实例也有该类型特定的使用限制。有关负载均衡类型的更多内容，请参见 实例类型 。

| 实例类型 | 资源 | 默认限制 |
| --- | --- | --- |
| 全部实例的通用限制 | 一个账号在单地域可创建的公网实例数量 | 常规 IP 的公网实例数量限制为100个。<br>静态单线 IP 中，个人账户的公网实例数量限制为3个，企业账户为15个。 |
| | 一个账号在单地域可创建的内网实例数量 | 100 |
| | 一个实例可添加的监听器数量 | 50 |
| | 一个实例中的监听器可选择的端口 | 端口为1 - 65535的整数 |
| 负载均衡（原“应用型负载均衡”） | 一个负载均衡实例中，HTTP/HTTPS 监听可配置的域名和 URL 转发规则数量 | 50 |
| | 一个负载均衡实例的监听器或转发规则可绑定的服务器数量 | 100 |
| | 一个负载均衡实例的前端端口可对应的后端端口数量 | 暂无限制 |
| 传统型负载均衡 | 一个传统型负载均衡实例的监听器可绑定的服务器数量 | 100 |
| | 一个传统型负载均衡实例的前端端口可对应的后端端口数量 | 1个端口 |

负载均衡**不会主动解除**与 CVM 的绑定关系，当云服务器进入隔离状态（包年包月云服务器进入回收站、按量计费云服务器欠费两小时以上）时，也**不会解除**与 CLB 的绑定关系。

## 带宽峰值

不同网络计费类型的带宽峰值含义有所不同，具体区别如下：

> **说明：**

所有的网络计费模式：

- 用户购买的出带宽小于等于10Mbps 时，腾讯云分配10Mbps 外网入方向带宽。
- 用户购买的出带宽大于10Mbps 时，腾讯云会分配与购买的出带宽相等的外网入方向带宽。

| 计费模式 | 带宽峰值区别 | 说明 |
| --- | --- | --- |
| 按固定带宽计费（包括按小时带宽和包月带宽） | 带宽峰值作为承诺指标。当出现带宽资源争抢时，带宽峰值有保证，不会受到限制。 | 单个地域中，所有按固定带宽计费（包括按小时带宽、包月带宽计费）的实例，购买和实际运行的总带宽峰值不大于50Gbps。若您有更大带宽需求，请联系您的商务经理申请调整。 |
| 按流量计费 | 带宽峰值仅作为带宽最高上限峰值，不作为承诺指标。当出现带宽资源争抢时，带宽峰值可能会受到限制。 | 单个地域中，所有按流量计费的实例，实际运行的总带宽峰值不大于5Gbps。若您的业务要求带宽保障或需更大带宽峰值，请购买按固定带宽计费的公网带宽。 |
| 共享带宽包 | | 单个地域中，所有按共享带宽包计费的实例，购买和实际运行的总带宽峰值不大于50Gbps。若您有更大带宽需求，请联系您的商务经理申请调整。 |

版权所有：腾讯云计算（北京）有限责任公司