cloudflare.com

What is cloud load balancing? | LBaaS

4-5 minutes

What is cloud load balancing?

Cloud load balancing, is a software-based load balancing service that distributes traffic between multiple cloud servers. Like hardware load balancers, cloud load balancers are designed to manage massive workloads so that no one server becomes overwhelmed by requests, which can increase latency and cause downtime.

Many cloud providers allow customers to rent load balancing services on an as-needed basis, rather than configuring and maintaining dedicated on-premise appliances to route their traffic themselves. This process is often referred to as 'load balancing-as-a-service' (LBaaS), though LBaaS can also balance workloads among on-premise servers.

How does cloud load balancing work?

Load balancing distributes traffic across multiple servers in order to reduce latency and improve server availability and reliability. When implemented successfully, the workload is shared among the servers to optimize performance and prevent server failure. Different load balancing techniques may be used to achieve this purpose: for example, a load balancer may evaluate server load or geographical distance before deciding where to direct traffic. In the event that a server goes offline, the load balancer redirects

incoming traffic to another available server in a process called failover.

Cloud-based load balancing, or LBaaS, operates in a similar way. However, instead of distributing traffic across a cluster of servers located in a single data center, it balances workloads across servers in a cloud environment, usually managed by a single cloud vendor. (Multi-cloud and hybrid cloud load balancers, by contrast, distribute traffic among multiple cloud providers.)

Why is load balancing necessary for cloud computing?

For cloud-hosted applications, load balancing is an essential service. Just as individual servers running in a data center can become overwhelmed and fail — causing significant latency and potential outages for end users — so can servers running in the cloud.

Hardware load balancing appliances not only are inefficient at managing traffic in the cloud, but are often prohibited from running in vendor-managed cloud environments as well. Software-based load balancers, meanwhile, can run in any environment and location, making them more suitable for cloud-hosted applications and infrastructure.

Regional cloud load balancing vs. global server load balancing

Cloud-based load balancers encompass both regional load balancers and global server load balancers (GSLB). As the name suggests, regional load balancers are designed to reduce strain on computing services within a specific region or localized network. On the other hand, global load balancers can balance workloads across servers in multiple locations worldwide, vastly cutting down

on latency for the end user.

What are the advantages of using LBaaS?

Reduced costs: LBaaS is often less costly than hardware appliances and requires less time, effort, and internal resources to maintain.

Scalability: LBaaS allows users to quickly and easily scale load balancing services to accommodate traffic spikes, rather than manually configuring additional physical load balancing infrastructure to do so.

Global availability: With GSLB, users can connect to the server that is geographically closest to them, minimizing latency and guaranteeing high availability even when a server is knocked offline.

Does Cloudflare offer cloud load balancing?

Cloudflare Load Balancing is a cloud-based load balancing solution that replaces on-premise load balancing hardware. It runs on a global network that spans 200+ cities worldwide, ensuring that traffic is quickly and efficiently routed. Learn more about Cloudflare Load Balancing.