

Products	>
Solutions	>
Developers	>
Partners	>
Pricing	

[Log in](#)[Sign up](#)[Blog](#)[Docs](#)[Get Support](#)[Contact Sales](#)

Article

Load balancing 101: Increasing the availability and resilience of your applications

[← Back to All Articles](#)

Share



Cloud Education

As a business, your website or app can't afford downtime or slow loading times. These issues not only frustrate users but can also hinder business growth. Load balancing tackles this problem head-on by intelligently distributing traffic across multiple servers, ensuring optimal performance and uptime. This article provides a deep-dive on load balancing, helping you implement and manage a robust load-balancing strategy to drive business success.

What is load balancing?

[Load balancing](#) is a networking technique that efficiently distributes network traffic among a group of servers known as a server farm. Its primary purpose is to enhance network performance, reliability, and capacity while minimizing latency. This is achieved by evenly distributing the demand across multiple servers and compute resources.

To accomplish this, load balancing employs an appliance, which can be physical or virtual, to assess which server in the pool is best suited to handle a specific client request in real time. This ensures that no single server becomes overwhelmed by heavy network traffic.

Furthermore, load balancing offers failover capabilities. In the event that one server fails, the load balancer swiftly redirects the workloads to a backup server, minimizing disruptions for end-users.

Load balancers play a crucial role in managing incoming requests from users for various services. They act as intermediaries between the servers handling these requests and the internet. When a request is received, the load balancer identifies an available online server within the pool and directs the request to that server. During periods of high demand, load balancers can dynamically add servers to handle traffic spikes, and conversely, they can remove servers during low-demand periods.

Why do companies choose to use load balancing?

Companies opt for load balancing to enhance the availability, reliability, and performance of their applications. Some of the benefits of using load balancers include:

Decreasing downtime

Almost every tech company needs to schedule maintenance or plan downtime at some point. Typically, they schedule these activities during off-peak hours, like early Sunday mornings, to minimize disruption. However, for global businesses with users in different time zones, someone is bound to be inconvenienced.

Load balancing solves this issue by [minimizing downtime](#) and increasing efficiency. When you need to perform maintenance on a server, you can simply shut it down and direct traffic to other resources without causing disruptions.

Facilitating peak performance

Load balancing offers another valuable benefit: the flexibility to add or remove resources, such as virtual machines or servers, without any disruption to incoming traffic. For businesses like online retailers or media publishers, events like Black Friday or holiday sales surges, as well as any industry-related news, can lead to heavy server loads. In

these cases, the ability to seamlessly and [rapidly scale resources](#) while balancing the load is crucial for maintaining a smooth user experience.

Increasing scalability

Load balancing leverages the scalability and agility of the cloud to manage website traffic effectively. Using efficient load balancers, you can seamlessly handle surges in user traffic by distributing it across multiple servers or network devices. This is particularly crucial for e-commerce websites, which contend with a high volume of visitors, especially during sales or promotional events. Effective load balancers are essential for efficiently distributing workloads in these scenarios.

Improving customer experience

Enhancing the customer experience is perhaps the most evident motive for implementing load balancing. During high-traffic periods, such as seasonal sales and marketing campaigns, load balancing prevents server crashes. This automated mechanism efficiently manages all incoming traffic, regardless of the influx of new visitors. Proper load balancing strategy safeguards against poor user experiences, ensuring that customers enjoy a consistently satisfying interaction every time.

Types of load balancing

Numerous load balancing options are available, each offering unique advantages while remaining cost-effective. Choose from DNS Failover, Round Robin, Weighted Round Robin, or Latency Routing to tailor the solution to your organization's specific requirements.

DNS failover

One of the most commonly employed load balancing methods is [Failover](#). This robust DNS solution ensures the continuous operation of your domains, even during outages or when your primary server encounters problems. If one of your resources becomes unavailable, traffic is seamlessly redirected to a healthy resource, offering straightforward yet effective redundancy.

Round robin and round robin with failover

[Round robin](#) is a widely used load balancing method. It evenly distributes traffic among redundant internet connections or web servers. This works because it operates with multiple active endpoints. When combined with Failover, it provides an extra advantage: if one resource fails or becomes unhealthy, the remaining healthy resource(s) will handle queries.

Weighted round robin

A more sophisticated load balancing setup is the Weighted Round Robin. It lets you distribute varying traffic loads across your network, allowing you to prioritize resources based on capacity or responsiveness. Weighted Round Robin is also handy for A/B testing, timed rollouts, or targeted user deliveries.

Latency routing (ITO)

ITO stands out as a regionally-based load balancing solution. It employs health checks to determine the most suitable resources for managing your web traffic. Unlike Round Robin with Failover, ITO takes into account the round trip time (RTT) of resources in your setup, ensuring that traffic is consistently directed to the fastest resource.

Regional/Multi-location load balancing

Enhance your website's traffic management by employing [multi-location load balancing](#) with GeoDNS. This feature enables the creation of customized routing rules according to end-user locations, effectively allowing you to construct your own region-specific content delivery network (CDN).

How to create DigitalOcean load balancer

DigitalOcean's load balancers are a fully-managed, exceptionally reliable network load balancing service. They efficiently distribute traffic to clusters of Droplets, effectively separating the backend service's overall health from that of a single server. This ensures your services remain consistently available, assuring their uninterrupted online presence.

Setting up a load balancer is a two step process: [creating the load balancer](#) and then adding Droplets or Kubernetes nodes to its backend pool.

Also learn how to [manage load balancers](#), [destroy load balancers](#) and [scale load balancers](#) with DigitalOcean [documentation](#) and [resources](#).

DigitalOcean load balancer use cases

[DigitalOcean Load Balancers](#) effectively divide incoming traffic across multiple backend servers. Typically, this is employed for distributing HTTP requests among application servers to boost overall capacity, a common method for scaling applications. Moreover, load balancers can enhance site reliability and streamline deployment and testing procedures.

Five such [DigitalOcean load balancer use cases](#) include:

Load balancing for scalability

DigitalOcean Load Balancers provide two load distribution algorithms: round robin and least connections. Round robin cycles requests through available backend servers, while least connections directs requests to the server with the fewest connections. Round robin is the most commonly used scheme for load balancing, but for applications with long-lasting connections, least connections might be better at preventing server overload.

Load balancing for higher availability

DigitalOcean Load Balancers contribute to higher availability by conducting regular health checks on your backend servers and automatically excluding failed servers from the pool. These health checks are customizable in the load balancer control panel's settings section. By default, the load balancer checks the server's response by fetching a web page every ten seconds. If this check fails three times consecutively, the server is temporarily removed until the issue is resolved.

Blue/green deployments

Blue/green deployments involve deploying new software on production infrastructure, rigorously testing it, and only redirecting traffic to it after confirming everything works correctly. In case of unexpected issues, DigitalOcean Load Balancers simplify recovery by switching back to the old version using Droplet tagging, where blue and green tags distinguish sets of Droplets.

Canary deployments

Canary deployments involve testing a new application version on a subset of users before rolling it out to all servers. With DigitalOcean Load Balancers, you can achieve this by adding a single canary server to the Load Balancer's

pool. Monitoring for errors or issues through logging, if everything looks good, you can deploy updates to the entire pool.

A/B deployment

A/B deployments share similarities with canary deployments, yet their purpose differs. A/B deployments assess a new feature's performance with a subset of users to collect data for informing marketing and development strategies. This process should be carried out alongside your current monitoring and logging systems to yield meaningful insights.

Improve application scalability and resilience by using DigitalOcean Load Balancers

Enhance your application's scalability and resilience with DigitalOcean Load Balancers. Our fully-managed solution intelligently distributes traffic across multiple servers, ensuring high availability and reducing the risk of performance issues.

In addition to the flexibility of selecting the appropriate load balancer size, we have retained numerous features, such as seamless integration with [Let's Encrypt](#) for hassle-free SSL certificate creation and automatic renewal. We have also introduced the capability to generate wildcard certificates, ensuring coverage for all subdomains. You can find detailed information on configuring SSL termination in our [official documentation](#).

Load Balancers are compatible with [Droplets \(Linux-based virtual machines\)](#) and [DigitalOcean Managed Kubernetes](#).

Boost your application's performance and reliability with [DigitalOcean Load Balancers](#).

To start using our load balancers, [sign up for DigitalOcean](#) today.

Related Resources



Articles

What Are Vector Databases? Why They Are So Important?

Read more →



Articles

AWS Hidden Costs: Identifying and Avoiding Them

[Read more →](#)



Articles

What is Cloud Encryption?

[Read more →](#)

Get started for free

Sign up and get \$200 in credit for your first 60 days with DigitalOcean.*

[Get started](#)

*This promotional offer applies to new accounts only.

Company	▼
Products	▼
Resources	▼
Solutions	▼
Contact	▼

