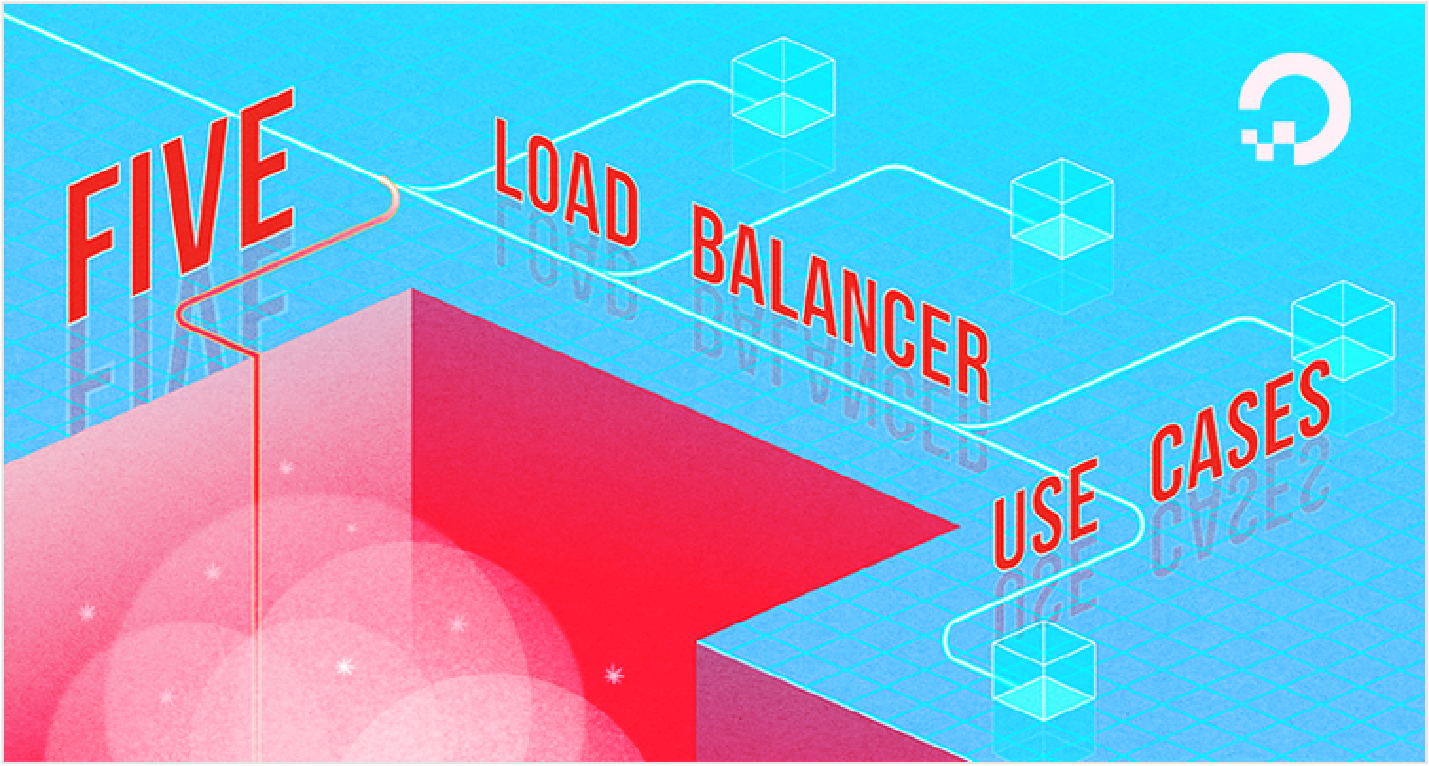
By: Brian Boucheron

Subscribe Share Contents



5 DigitalOcean Load Balancer Use Cases

Posted June 23, 2017 16.4k DIGITALOCEAN LOAD BALANCING 9

Introduction

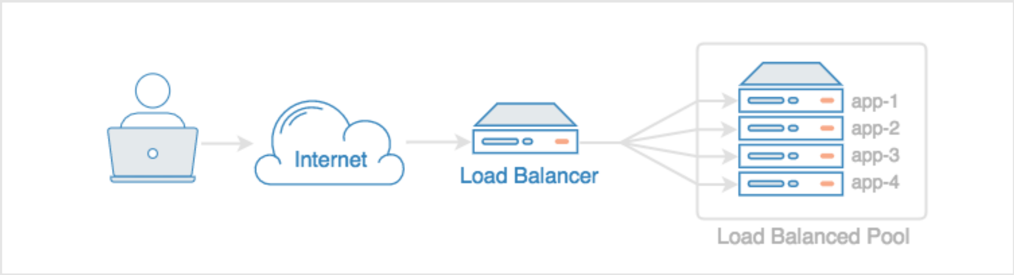
DigitalOcean Load Balancers allow you to split incoming traffic between multiple backend servers. Often this is used to distribute HTTP requests among a group of application servers to increase overall capacity. This is a common way to scale your application.

Load Balancers also offer other use cases. For example, they can increase the reliability of your site, or improve your deployment and testing processes. In this tutorial, we will review five Load Balancer use cases.

Before we begin, you should familiarize yourself with the basics of DigitalOcean's Load Balancers by reading our tutorial An Introduction to DigitalOcean Load Balancers.

1. Load Balancing for Scale

As mentioned above, scaling traffic is the most common use case for a Load Balancer. Often times scaling is discussed in *vertical* and *horizontal* terms. Vertical scaling is basically moving your application to a more powerful server to meet increasing performance demands. Horizontal scaling is distributing your traffic among multiple servers to share the load. Load Balancers facilitate horizontal scaling.



DigitalOcean Load Balancers allow you to distribute load via two different algorithms: round robin and least connections. Round robin will send requests to each available backend server in turn, whereas least connections will send requests to the server with the fewest connections. Round robin is by far the most frequently used scheme for load balancing, but if you have an application that keeps connections open for a long time, least connections may do a better job of preventing any one server from becoming overloaded.

A side benefit of horizontal scaling with load balancers is the chance to increase your service's reliability. We'll talk about that next.

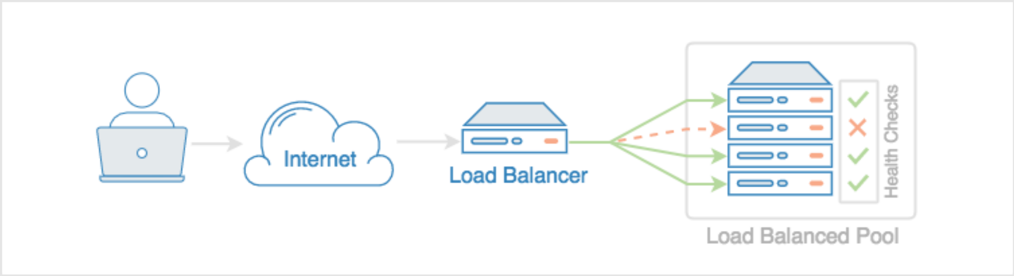
Related Tutorials:

- How To Create Your First DigitalOcean Load Balancer
- How To Balance TCP Traffic with DigitalOcean Load Balancers

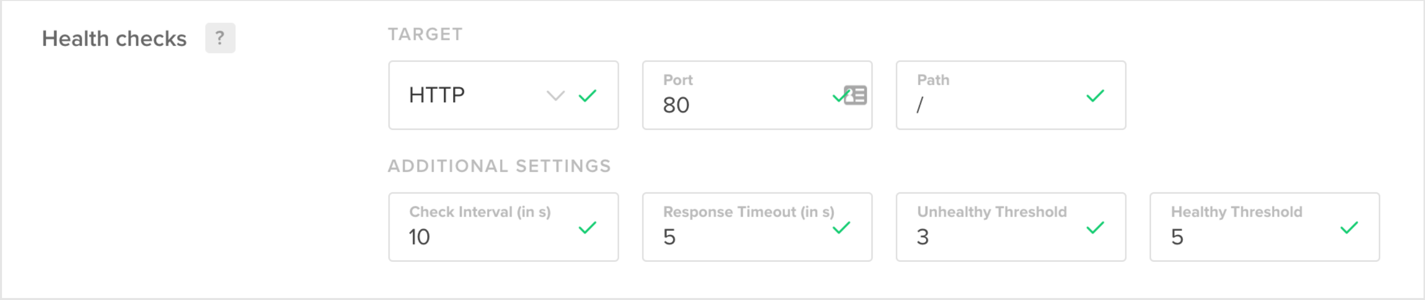
2. High Availability

High availability is a term that describes efforts to decrease downtime and increase system reliability. This is often addressed by improving performance and eliminating single points of failure.

A Load Balancer can increase availability by performing repeated health checks on your backend servers and automatically removing failed servers from the pool.



Health checks can be customized in the **Settings** area of the Load Balancer control panel:



Health checks ?

TARGET

HTTP ✓ | Port 80 ✓ | Path / ✓

ADDITIONAL SETTINGS

Check Interval (in s) 10 ✓ | Response Timeout (in s) 5 ✓ | Unhealthy Threshold 3 ✓ | Healthy Threshold 5 ✓

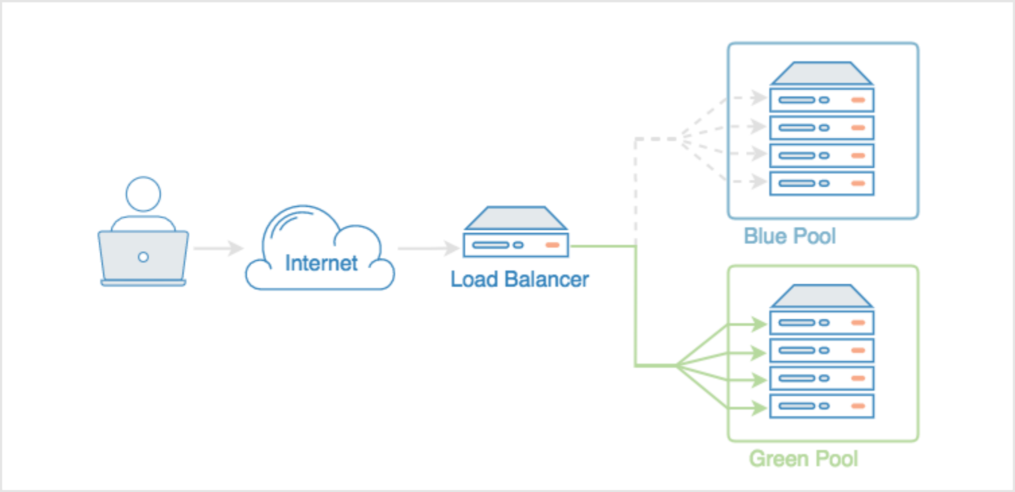
By default, the Load Balancer will fetch a web page every ten seconds to make sure the server is responding properly. If this fails three times in a row, the server will be removed until the problem is resolved.

Related Tutorials:

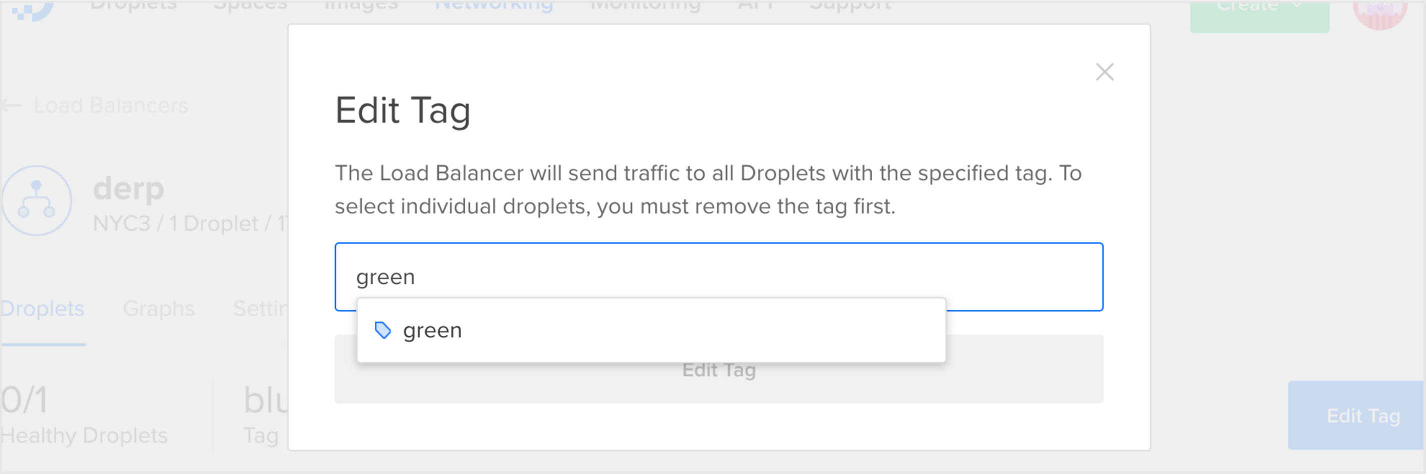
- What is High Availability?

3. Blue/Green Deployments

Blue/green deployments refer to a technique where you deploy your new software on production infrastructure, test it thoroughly, then switch traffic over to it only after verifying that everything is working as you expect. If the deploy ends up failing in new and unexpected ways, you can easily recover by switching the Load Balancer back to the old version.



DigitalOcean Load Balancers make blue/green deployments simple through their use of the Droplet tagging feature. Load Balancers can send traffic to a group of servers based on their tag, so you can have one set of Droplets tagged **blue** and the other **green**. When it's time to cut over, switch the tag in the Load Balancer control panel or through the API:



Edit Tag

The Load Balancer will send traffic to all Droplets with the specified tag. To select individual droplets, you must remove the tag first.

green

green

Edit Tag

After you save your changes, traffic will quickly switch over to the new set of Droplets.

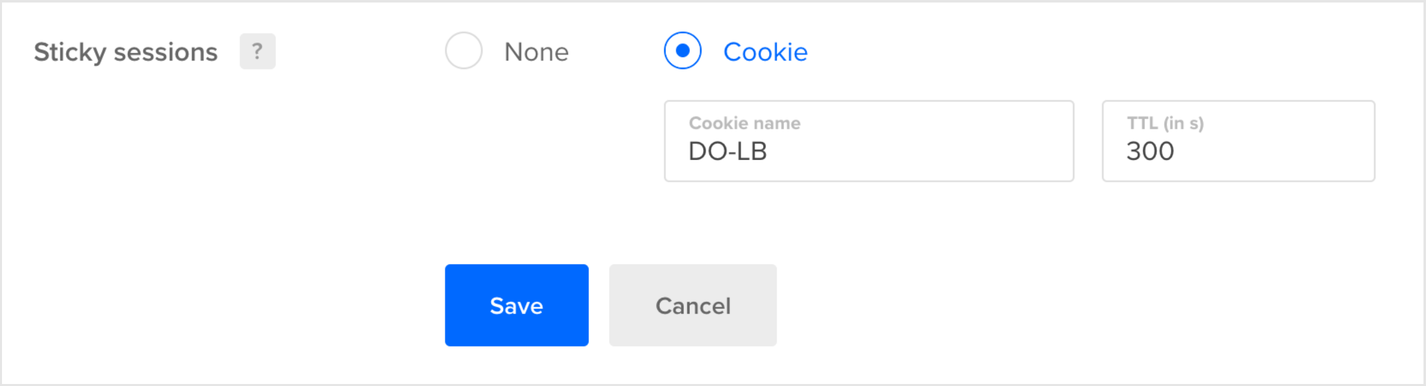
Related Tutorials:

- How To Use Blue-Green Deployments to Release Software Safely

4. Canary Deployments

Canary deployments are a way of testing a new version of your application on a subset of users before updating your entire pool of application servers. With DigitalOcean Load Balancers you could do this by, for instance, adding just one canary server to your Load Balancer's pool. If you don't see any increase in errors or other undesirable results through your logging and monitoring infrastructure, you can then proceed to deploy updates to the rest of the pool.

You'll want to turn on *sticky sessions* for this use case, so that your users aren't bounced between different versions of your application when making new connections through the Load Balancer:



Sticky sessions ?

○ None ● Cookie

Cookie name: DO-LB

TTL (in s): 300

Sticky sessions will use a cookie to ensure that future connections from a particular browser will continue to be routed to the same server. You can access this feature in the **Advanced settings** area of the Load Balancer's control panel.

5. A/B Deployment

A/B deployments are functionally similar to canary deployments, but the purpose is different. A/B deployments test a new feature on a portion of your users in order to gather information that will inform your marketing and development efforts. You'll need to do this in conjunction with your existing monitoring and logging infrastructure to get back meaningful results.

On the server side, you'll add one or more **B** servers to your existing pool of **A** servers. If you need to launch multiple **B** servers to gather enough data, you could organize this with tags as we did for blue/green deployments.

Conclusion

Although Load Balancers are most often considered when scale is needed, we've shown that there are many other cases where it's useful to have the ability to distribute or shuffle traffic among various backend servers. Whether it's for high availability or leveraging various deployment techniques, Load Balancers are a flexible and powerful tool in your production infrastructure.

For more in-depth and specialized information on DigitalOcean Load Balancers, take a look at the following tutorials:

- How To Balance TCP Traffic with DigitalOcean Load Balancers
- How To Configure SSL Termination on DigitalOcean Load Balancers
- How To Work with DigitalOcean Load Balancers Using Doctl
- DigitalOcean API v2 Load Balancer Documentation

By: Brian Boucheron

♡ Upvote (9) Subscribe Share

Related Tutorials

How To Work with DigitalOcean Load Balancers Using Doctl

How To Use Python Decouple with Spaces in Django

DigitalOcean Private Networking FAQ

How To Set Up Object Storage with Django

How To Automate the Scaling of Your Web Application on DigitalOcean Ubuntu 16.04 Droplets

3 Comments

Leave a comment...

Log In to Comment

djimnz *July 19, 2017*

0 Nice to read :)

CD2TeamJJB *September 11, 2017*

1 Would love to see an unattended API only version of this. Is the load balancer available via API like Amazon?

raviu *October 23, 2017*

0 Are the deployment diagram graphics available on some kind of diagram tool ? It would be nice to use DO specific icons when creating deployment diagrams.