
LOAD BALANCER - WHAT IS IT, HOW DOES LOAD BALANCER WORK

Kazakov Firdavs Farxod ugli

Tashkent State University of Economics., Uzbekistan

Abstract: To create reliable and secure operation of applications, a decentralized system of several hosts is now used. To set up their correct operation and eliminate failures and overloads, it is worth learning about what a hardware load balancer is and server traffic, how it works, whether the device is required or whether you can use an application. Let's consider all the main nuances that will help you understand and choose the right option for optimizing traffic. In addition to the main functions, they can also perform additional ones - for example, protect against hacker attacks, which are based on overloading web servers to gain access to data. Due to the distribution system, stability will be ensured and even millions of simultaneous requests will not be able to be processed by the hosts. Therefore, this method of protecting sites and applications from unauthorized access is relevant and in demand.

Key words: load balancer, web servers, Reliability.

With a large number of requests to the server, abnormal conditions are created under which the device may fail and require repair work. All this disrupts the performance of the system and makes it inaccessible to users. Load balancer is an automated component that allows you to evenly distribute requests between clusters, which significantly reduces the negative impact of increased load on one of the devices.

Thus, this technique allows you to create an "intermediary" between users and web servers, which will process part of the information and transmit it in such a way as to avoid overload on individual clusters.

The process in question is extremely important for maintaining optimal information flow between user devices that provide access to the site and server nodes. There are several main important features that make intermediaries indispensable:

- Reliability – providing a high UX even with high traffic.
- Availability - the automated system constantly checks between hosts in order to promptly recognize the lack of information processing and redirect new requests to other nodes.
- Removing inoperative parts from the circuit while they remain inaccessible.
- Security - protects against DDoS attacks by shifting attack traffic to the cloud provider, rather than the corporate site.
- Analytics - the system analyzes and finds communication weaknesses.

Let's look at some of these operating principles in more detail to understand why automation of request distribution is needed.

Availability

If an application or site fails, it may remain unavailable to users for a long time. This affects their effectiveness, as well as ranking by search robots. Performing network load balancing allows you to timely determine which nodes have stopped processing requests and redirect information flows to other hosts with which everything is fine.

This function is convenient to use not only for identifying problems, but also:

- to upgrade individual servers without shutting down all of them;

- automatically restore emergency sites, download backup files;
- prevent unwanted downtime and traffic churn.

Stable communication between users and host machines can be ensured.

Scaling

Network requests can be routed intelligently because load balancer allows you to:

- add power to the system;
- predicts the traffic that will be on a particular host device so that you can remove or add them to the system;
- prevents loss or, conversely, heavy-duty information flow in one section of the chain.

Therefore, you can make scope changes without stopping the site or application.

Safety

Often, attackers carry out DDoS attacks on hosts, flooding them with millions of parallel requests so that the hosts cannot cope with processing information. In such a case, they throw an unavailable error, which leads to the possibility of data leakage and results in users being unable to access the application. Restoration will require a lot of time and resources. Instead, you can install a load balancer that:

- will monitor and block harmful content;
- redistribute information waves, preventing overload in the system;
- You can set routing through firewalls, which will increase the level of security.

Performance

Taking into account the fact that now the threads will be optimally distributed, one or more host machines will not work, their efficiency in processing incoming data will be as high as possible. Besides:

- there will be a determination of the request to the geographically closest host, which will minimize the delay in data processing;
- computing resources, both physical and virtual, will be easier to cope with systematized signals than with chaotic signals.

Uniformity and the principle of queuing will help speed up the site's response, which will have a positive impact on its effectiveness and ranking.

Algorithmic processes involve a strict set of rules by which tasks will be performed. They will be aimed at ensuring that server load balancing identifies the best host devices for processing certain user requests. There are two categories:

- static;
- dynamic.

Let's take a closer look at how each of them works. How do the operating principles differ and where will the application be more relevant?

Static

The main advantage is that there is no connection to the server state. Strict algorithms are aimed at distributing the load as evenly as possible.

Round Robin Method

Balancers to the Round Robin - has many advantages:

- does not depend on the load;
- suitable for local and global balancing;
- does not require communication between hosts;
- have low sales costs.

In general terms, the round robin rule looks like passing tasks sequentially to each of the host devices. DNS servers store a "host name - IP address" pair, which is assigned to each device in the domain. One name can have several IPs. Thanks to circularity, it is not the first one in the list that takes the entire load, but each request will be distributed in turn to each of the list until it returns to the browser again.

There are also disadvantages to this technique. The algorithm does not take into account the workload, issuing in turn, one of the hosts can perform its tasks at 90%, and the other is loaded at 10-15%, and in turn, each will receive a request.

Weighted round robin method

Here the main distribution device will again be the server, storing the “host machine name - IP address” pair in memory. However, it is also possible to distribute the weight for each device. A large amount of incoming traffic will be directed exactly to those hosts where the highest priority is specified.

IP hashing method

IP-Hash is based on the user's IP; it is on this basis that mathematical calculations will be made and the determination of where exactly and to which host to send the request to be processed. It turns out that each individual client who requests a search from one address will be given a route to a specific server each time.

In this version of the algorithm, the load and performance of host machines is already assessed. Before distributing traffic, test requests will be given, and if the balancer detects problems or high load, it will redirect to more efficient devices.

Least connection method

Mathematically, all hosts should have the same processing power. After checking, the load balancer will direct the user's request to the device that has the fewest active connections at a particular time, which accordingly has a low load and higher productivity.

Least connections method

Some larger hosts can process many connected addresses, so some will have more active tasks, while others will have fewer. Accordingly, those that are less busy can process incoming requests more productively. Therefore, the balancer redirects traffic to those with less capacity.

Shortest response time method

In this case, the load balancer will analyze how long it takes for each host to respond, and it will redistribute the traffic so that it is received by the host device that can process the incoming data the fastest.

Resource based method

A special software agent is installed on each host machine, which will constantly monitor power and memory used. The installed balancer will read this data before sending traffic; if there are enough resources, then the request will be defined there. If not, go to a freer device.

Large companies create farms to handle application needs. These are server rooms where several devices are located. There may be two or a hundred. Accordingly, it is necessary to provide work for all of them, avoiding a situation where one is overloaded and the others are idle. This is precisely the main functionality of the balancer. It redirects traffic to those hosts that are best suited for this.

There are several main categories into which intermediary devices can be divided. The main criterion will be the verification of the client request and redirection algorithms, depending on what exactly was verified.

Most modern utilities have a complex structure. Each section of which is stored on a separate server, ensuring stable operation. In this case, load balancers are needed to view HTTP or SSL, and in accordance with specified user requests, redirect to specific hosts.

These types of devices will examine IP and network data to select the optimal host machine to process the information. They can use both static and dynamic methods, which were discussed above. According to the specified algorithms, based on the results of traffic analysis, send requests to the host's IP address.

There are very large projects that work all over the world. To maintain their functionality, server farms are located in each country, zone or individual region. Load balancers should also be positioned accordingly to forward incoming requests to the closest geographically located servers. If they do not work, malfunction or are under maintenance, go to other regional or zonal farms.

The domain name system will act as a balancer, that is, it determines traffic and distributes it to other servers. This technique can be dynamic with checking the availability of hosts. If the answer was no, then incoming requests will be pushed to another host machine that is in working order.

Let's move on to the process of implementing this system. First you need to figure out exactly what technical means can be used for this. Both individual devices and certain software installed on the host can be used as a load balancer.

These are devices that perform all the same functions described above. At the same time, they are quite tangible, they can be stored, and virtual centers can be created to control several devices. Some consider them obsolete, preferring cloud resources that do not require reconfiguring hardware or looking for a place to host them. In addition, they have one more disadvantage - a purely physical one, they can be mechanically damaged or fail.

Everything is simple here - you just need to install the software, it can be managed like a separate centralized resource. High-quality automated systems can be ordered from Cleverens specialists, who will offer solutions for specific business problems.

Virtual ones with cloud storage are now considered more relevant for many reasons:

- devices, like all other material resources, can break down;
- the hardware has a limited resource; if more is needed, when the traffic exceeds the expected power threshold, additional equipment will need to be purchased;
- initial implementation will require considerable capital investment to purchase everything necessary for connection;
- as long as new balancers are purchased and installed, users will experience shortcomings and overloaded servers, which will affect the stability of the services provided, loss of reputation and profit.

Therefore, it is best to use virtual software that will not cause problems; it will take several hours to install and configure using NGINX or similar tools. As well as expanding the resource if necessary.

Now you know what network load balancing is, how this system works, how traffic is distributed between servers. In connection with the information discussed, it becomes clear that "balancers of chaos and order," as the analogy is given, really helps to balance chaos and order, protecting the project from destabilization by a large number of requests. Therefore, its use is relevant both for the backend of online stores and for more voluminous utilities.

REFERENCES:

1. Babich, A.M. State and municipal finance: textbook. for universities [Text]/ A.M. Babich, L.N. Pavlova. – M.: UNITY, 2009. – 688 p.
2. Budgetary system of Russia: textbook. [Text]/ Ed. G. B. Polyak. – M.: UNITY-DANA, 2012. – 212 p.
3. Zharkovskaya, E.P. Finance: textbook. manual [Text]/E.P. Zharkovskaya, I.O. Arends.-M.: Omega-L, 2011. – 400 p.
4. Konishcheva, M.A. Finance of organizations: textbook / M.A. Konishcheva, O.I. Kurgan, Yu.I. Cherkasova; Ministry of Education and Science of the Russian Federation, Siberian Federal University. - Krasnoyarsk : Siberian Federal University, 2015. - 383 p. : table, diagrams. - Bibliography: p. 257-261. - ISBN 978-5-7638-3231-0; The same [Electronic resource]. - URL: <http://biblioclub.ru/index.php?page=book&id=435687> (06/23/2016).
5. Leontiev, V.E. Finance, money, credit and banks: textbook. manual [Text] /V.E.Leontyev, N.P. Radkovskaya. – St. Petersburg: IVESEP, Znanie, 2011. – 384p.
6. Podyablonskaya, L.M. State and municipal finance: textbook / L.M. Podyablonskaya. - M.: Unity-Dana, 2015. - 559 p. : table, diagrams - Bibliography: p. 552-553. - ISBN 978-5-238-01488-3; The same [Electronic resource].-URL: <http://biblioclub.ru/index.php?page=book&id=114698> (06/23/2016).
7. Finance: textbook. [Text] / Ed. A.G. Gryaznova, E.V. Markina. – M.: Finance and Statistics, 2011. – 504 p.