

An Edge DNS Global Server Load Balancing for Load Balancing in Edge Computing



P. Herbert Raj

Abstract The information technology era faces an inflection spot due to current changes in the network industry due to the surrogation of cloud application implementation instead of data centre-based application implementation. The fiery growth of portable devices, sensor, and IoT-based devices would cause a stern effect on cloud application development and deployment in the network industry. The edge computing model is developed to enhance cloud application deployment. Edge computing is defined as where the computed calculations and storage of data are located in close physical proximity to improve the response time, reduce the latency, and improve the network bandwidth. This article analyses the ways to achieve the goals of edge computing with the assistance of load balancing.

Keywords Edge computing · Data centre · Global server load balancing (GSLB) · Disaster recovery plan (DRP) · Distributed denial of service (DDoS) · Internet of Things (IoT) · 5G · Latency

1 Introduction

This is an epoch of cloud computing. Currently, many people are using personal computers and desktops to access many of the centralized server computers. At the same time, IBM, Amazon, Microsoft, and Google are providing cloud services to their users to improve scalability, efficiency, reduce IT expenditure, 24 × 7 connectivity, and much more [1]. The cloud computing business could surpass \$330 billion in 2020 due to the embracement of cloud computing [2]. This happened because users could access the resources from anywhere at any time. The upcoming 5G technology will enhance these capabilities because edge computing must provide high bandwidth to the connected users [3–5].

P. Herbert Raj (✉)
Bandar Seri Begawan, Brunei

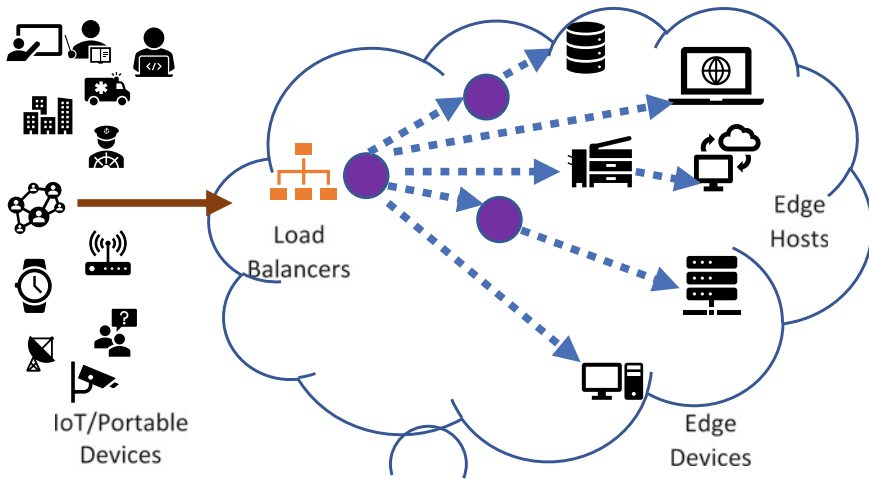


Fig. 1 Edge architecture with load balancers

1.1 Need for Load Balancing in Edge Computing

Edge computing tries to move the computations nearer to clients to minimize the delay and to reduce the bandwidth usage. The edge of a communication network is very close to the client; unlike the cloud network, the servers are placed in a faraway place. Cloud computing is a central network management system where applications are functioning in the data centres. Edge computing is also a central network management system but here applications are operating either in the device or in the network edge [6]. It intensifies the confidentiality and solitude of data being processed. Network instability or disruption will not affect the overall operations in edge computing [7]. Figure 1 depicts the edge computing architecture with load balancers.

1.2 Need for Load Balancing in Edge Computing

Scalability is the key issue in load balancing. Some services may not be available to clients due to the network or website or server failure. This may lead to disparity in the network traffic distribution. Load balancing algorithms take a lead to resolve the unbalanced traffic distribution. Even though numerous algorithms are available, achieving optimal load balancing is an ultimatum. The cloud bursting method is used in some servers to settle the load balancing issues in the edge environment [8]. The application of load balancing is greatly influenced by the tendencies of edge computing. The job of the load balancer is to migrate together with other apps. Data centres of edge have load balancers that utilize the global server load balancing

(GSLB) for linking the different edge data centres. This will maintain an effective reply time by spreading the load over multiple edge servers [9].

1.3 Load Distribution

Load in the network will be distributed evenly between clients and servers by using various load balancing algorithms. The algorithm will be selected based on what sort of facility provided, and the present stage of the network will be taken into consideration. There are few load balancing algorithms used to balance the load, namely round-robin, weighted least connection, and resource-based.

1.4 Edge Computing and Load Balancing

Internet of Things can able to nexus with numerous portable and smart gadgets and automatically turn out to be a component of varying cutting-edge applications. This produces a huge volume of data, and it may increase in future [10, 11]. The data centres of clouds are positioned very in far-flung locations from these gadget users, and the bandwidth is limited [12]. So, latency in request-reply is inevitable. Edge computing assists to reduce this distance and moving closely from federal services to the boundary of the network. These edge devices may handle a huge sum of data and concurrent tasks without any latency [13]. Nevertheless, this might lead to unbalanced traffic distribution due to the huge amount of data to be processed. Multiple data centres in the edge can be combined as a single virtual one. So, load balancing can be done here by using GSLB. Nowadays, edge computing load balancing is drawn the attention of scholars because of its physical proximity to the users and latency reduction [1]. In this article, the process of GSLB comfort edge load balancing is discussed in detail.

2 GSLB Load Balancing

In GSLB, the network traffic is disseminated to the numerous linked servers in this universe. This is one of the most credible load balancing methods for industries, and it greatly minimizes latency. GSLB empowers multi-edge data centres to control and manipulate the availability of the resources placed across the world. This can be achieved by incorporating virtual multi-edge data centres, several servers, and a cloud environment. It has optimized traffic redirection methods in case of any failure or service intrusion over the network distribution. The user's request is sent to a location where it can obtain the best service and also examines the availability of the resources.

2.1 *Need for Edge DNS GSLB*

- (a) **Load Balancers:** Load balancers function well on a smaller scale, which means when the end users' devices and servers are limited. LBs may not function well in the multi-cloud with numerous data centre architectures due to their complicated topology. For the LBs, there is a need for definite design and bandwidth to link the data centres and user devices. It is not possible to provide these facilities for all the services. To alleviate this problem, DNS can be used [14]. Here, user requests can select one of the offered IP address at random and send their traffic to a particular destination. This will disseminate the load amid data centres. The LBs mounted in the local data centres lever the remaining traffic.
- (b) **DNS:** DNS disseminates the network load to numerous addresses with a similar application label. Setting up a DNS connection is very straightforward, and the service provided by DNS can be easily used by users. In load balancing, DNS redundancy cannot be updated automatically when a server is added or removed. Some automatic updates or manual updates are needed at this point. This may lead to application unavailability for some clients. Global DNS system may provide a similar reply to all the connected users. Some may need a different reply according to their site location. So vibrant updates and site-specific replies are mandatory here. To alleviate this sort of problem, DNS global is hosted [14].
- (c) **DNS GSLB:** DNS GSLB is able to reply vibrantly to the client requests and robotically updates the server. The disadvantage of this method is that those who need to use this service have to be in this geographical zone. This system is highly centralized. Always on service availability may not be possible due to the capability to find out the available servers. This problem leads to find another better method to route the traffic smartly and directly.
- (d) **Edge DNS GSLB:** Edge DNS GSLB is considered to be the best solution for traffic routing because it offers the best routing decision to reach the destination for a requested client. It can answer any client's request without any centralized committed zone. The traffic routing of edge DNS GSLB can be integrated with characteristics of a recursive DNS server which offers a lot of benefits to the clients. The benefits of this method are scalability, swift responsiveness, obtainability of local data, proficient caching, security, optimized DNS traffic, and augments multicentre robustness [14].

3 **Edge DNS GSLB and Multi-cloud**

Edge DNS GSLB produces a professional and modest way to load balance the network load. The application routing choice will be made at the location of users which is closer to the edge network. The forthcoming sessions describe the advantages of executing Edge DNS GSLB.

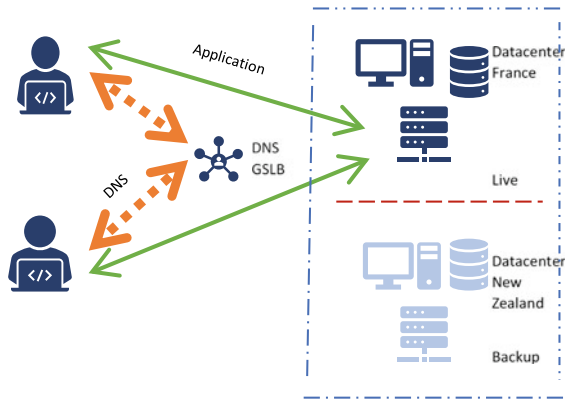


Fig. 2 Live connection. Source <https://www.efficientip.com/wp-content/uploads/2020/10/sp-GSLB-Use-Cases-EN.pdf>

3.1 Advanced Disaster Recovery

Industries host significant applications in a particular server that will be available in the disaster recovery plan sites. These kinds of applications will be accessed and used by numerous users. IP address maintenance will be handled by DNS. In case of any failure arises, these applications will not be accessible. The new location of the network must be reflected in the IP address. Without blunders, changing the DNS information is impossible. Moreover, it is a time laborious task. All users switching over to the DRP site would complicate things furthermore.

At the application level, Edge DNS GSLB permits amalgamating manifold design facilities and switching over.

- (a) **Failovers:** Here, each application is established with two nodes, one is to link the key site and another one is to link the reserve site. If the key site is incapacitated, then all the applications are connected to the DR site. This will be done automatically [14]. Figure 2 shows the live connection between the user and the application data centres.

3.2 Sharing of Network Load Between Multiple Data Centres

Consider a scenario, an application is accommodated in several data centres for distributing the network load. The users' requests will be directed according to their own site to the desired data centres. The challenge is how to make this possible for a user-specific site. Figure 3 describes the traffic routing with a health check.

- (a) **Dynamic Distribution:** The level of the application server link will be checked on an ordered basis by Edge DNS GSLB to determine whether the users can be pointed to it. This will decide on which is the best link for the user [14].

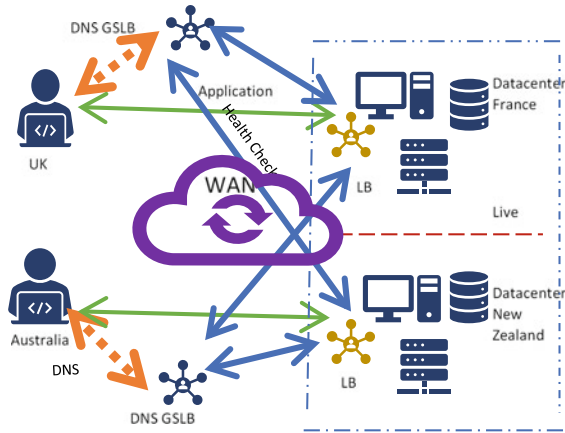
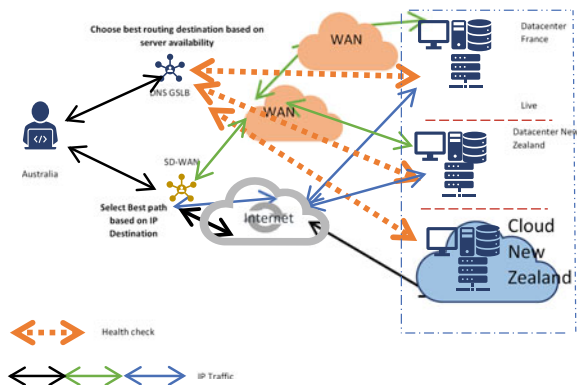


Fig. 3 Application traffic routing with a health check. Source <https://www.efficientip.com/wp-content/uploads/2020/10/sp-GSLB-Use-Cases-EN.pdf>

3.3 WAN Failure Detection

In the case of WAN failure, clients' requests should be led to the application server which is accessible and the top pick. The challenge is that accessing from a middle of a server to a client placed in a far-flung site may not be possible all the time. Edge DNS GSLB identifies any differences in the network among the far-flung location and the application server and gives the best solution for routing [15]. Figure 4 shows the selection of the best destination routing.

Fig. 4 Selecting the best destination. Source <https://www.efficientip.com/wp-content/uploads/2020/10/sp-GSLB-Use-Cases-EN.pdf>



4 Conclusion

Until now, IP address resolution is done by DNS service. This is a significant procedure in IP networks because it permits users to access the application server. The operational requirements made the network engineers devise new protocols and devices to enable the users to access everything smoothly and securely. When data centres provide services to the users to store or retrieve data and a lot of security measures are devised to access the data securely. The design complications are diminished with the help of Edge DNS GSLB. Load balancing and multi-cloud management can be designed undoubtedly with the help of Edge DNS GSLB. The significant benefits of Edge DNS GSLB have increased scalability, dexterity, robust multi-cloud management, user experience, and simple DR plans. Most importantly, this load balancing technique minimizes the energy consumptions of the processor without lowering the edge network functioning. Finally, this method guarantees the application accessibility for the clients, if any WAN breakdown happens. The implementation benefits of Edge DNS GSLB perfect match for load balancers, SD-WAN, and application delivery controllers to make correct routing decisions. This is very straightforward to implement and expands huge benefits.

References

1. Li G, Yao Y, Wu J, Liu X, Sheng X, Lin Q (2020) A new load balancing strategy by task allocation in edge computing based on intermediary nodes. *EURASIP J Wireless Commun Networking* 2020(3). <https://doi.org/10.1186/s13638-019-1624-9>
2. Nick Galov (2020) 25 must-know cloud computing statistics in 2020, hosting tribunal. <https://hostingtribunal.com/blog/cloud-computing-statistics/>
3. Ravi Kumar P, Herbert Raj P, Jelciana P (2017) Exploring security issues and solutions in cloud computing services. *Cybern Inf Technol* 17(4):29. Print ISSN: 1311-9702; Online ISSN: 1314-4081
4. Abdulmohson A, Pelluri S, Sirandas R (2015) Energy efficient load balancing of virtual machines in cloud environments. *Int J Cloud-Comput Super-Comput* 2(1):21–34. <http://dx.doi.org/10.21742/ijcs.2015.2.1.03>
5. Efficient IP (2019) How edge DNS GSLB ensures app availability during WAN failure. <https://www.efficientip.com/dns-gslb-wan-failure/>
6. Cloud Fare (2020) What is edge computing. <https://www.cloudflare.com/learning/serverless/glossary/what-is-edge-computing/>
7. Gonzalez J, Hunt J, Thomas M, Anderson R, Mangla U (2020) Edge computing architecture and use cases. LF Edge Premium Member Company IBM. <https://www.lfedge.org/2020/03/05/edge-computing-architecture-and-use-cases/>
8. Nugara A (2019) Load balancing in Microsoft Azure. O'Reilly Publisher(s) Inc., ISBN: 9781492053927
9. Yue F (2020) Edge computing and load balancing. Kemp technologies. <https://kemptechnologies.com/blog/edge-computing-and-load-balancing/>
10. Wan S, Zhao Y, Wang T, Gu Z, Abbasi QH, Choo KKR (2019) Multi-dimensional data indexing and range query processing via Voronoi diagram for Internet of things. *Future Gener Comput Syst* 91:382–391

11. Zaslavsky A, Perera C, Georgakopoulos D (2013) Sensing as a service and big data. arXiv preprint [arXiv:1301.0159](https://arxiv.org/abs/1301.0159)
12. Herbert Raj P, Ravi Kumar P, Jelciana P (2016) Mobile cloud computing: a survey on challenges and issues. *Int J Comput Sci Inf Secur (IJCSIS)* 14(12). ISSN 1947–5500
13. Zhu T, Shi T, Li J, Cai Z, Zhou X (2018) Task scheduling in deadline-aware mobile edge computing systems. *IEEE Internet Things J* 1. <https://doi.org/10.1109/jiot.2018.2874954>
14. Efficient IP (2020) Edge DNS GSLB complement your load balancing and multi-cloud strategy. EfficientIP, SAS
15. Efficient IP (2020) Edge DNS GSLB use cases improving UX, DRP and datacenter agility. EfficientIP, SAS. <https://www.efficientip.com/wp-content/uploads/2020/10/sp-GSLB-Use-Cases-EN.pdf>