

Cloud load balancing

Cloud load balancing is a type of [load balancing](#) that is performed in [cloud computing](#).^[1] Cloud load balancing is the process of distributing workloads across multiple computing resources. Cloud load balancing reduces costs associated with document management systems and maximizes availability of resources. It is a type of [load balancing](#) and not to be confused with [Domain Name System \(DNS\) load balancing](#). While DNS load balancing uses software or hardware to perform the function,^[2] cloud load balancing uses services offered by various [computer network](#) companies.^[3]

Comparison With DNS load balancing

Cloud load balancing has an advantage over DNS load balancing as it can transfer loads to servers globally as opposed to distributing it across local servers.^[3] In the event of a local server outage, cloud load balancing delivers users to the closest regional server without interruption for the user.

Cloud load balancing addresses issues relating to [TTL reliance](#) present during DNS load balancing.^[4] DNS directives can only be enforced once in every TTL cycle and can take several hours if switching between servers during a lag or server failure. Incoming server traffic will continue to route to the original server until the TTL expires and can create an uneven

performance as different internet service providers may reach the new server before other internet service providers.^[4] Another advantage is that cloud load balancing improves response time by routing remote sessions to the best performing data centers.^{[1][5]}

Importance of Load Balancing

Cloud computing brings advantages in "cost, flexibility and availability of service users."^[6] Those advantages drive the demand for Cloud services. The demand raises technical issues in Service Oriented Architectures and Internet of Services (IoS)-style applications, such as high availability and scalability. As a major concern in these issues, load balancing allows cloud computing to "scale up to increasing demands" ^[6] by efficiently allocating dynamic local workload evenly across all nodes.^[7]

Load Balancing Techniques

Scheduling Algorithms

Opportunistic Load Balancing (OLB) is the algorithm that assigns workloads to nodes in free order. It is simple but does not consider the expected execution time of each node. ^[8] Load balance Min-Min (LBMM) assigns sub-tasks to the node which requires minimum execution time.^[8]

Load Balancing Policies

Workload and Client Aware Policy (WCAP) is "implemented in a dis-centralized manner with low overhead."^[9] It specifies the unique and special property (USP) of requests and computing nodes. With the information of USP, the schedule can decide the most suitable node to complete a request. WCAP makes the most of computing nodes by reducing their idle time. Also, it reduces performance time through searches based on content information.

A Comparative Study of Algorithms

Biased Random Sampling bases its job allocation on the network represented by a directed graph. For each execution node in this graph, in-degree means available resources and out-

degree means allocated jobs. In-degree will decrease during job execution while out-degree will increase after job allocation.

Active Clustering is a self-aggregation algorithm to rewire the network.

The experiment result is that "Active Clustering and Random Sampling Walk predictably perform better as the number of processing nodes is increased"^[6] while the Honeyhive algorithm does not show the increasing pattern.

Client-side Load Balancer Using Cloud Computing

Load balancer forwards packets to web servers according to different workloads on servers. However, it is hard to implement a scalable load balancer because of both the "cloud's commodity business model and the limited infrastructure control allowed by cloud providers."^[10] Client-side Load Balancer (CLB) solve this problem by using a scalable cloud storage service. CLB allows clients to choose back-end web servers for dynamic content although it delivers static content.

References

1. Chee, Brian J.S. (2010). *Cloud Computing: Technologies and Strategies of the Ubiquitous Data Center*. CRC Press. ISBN 9781439806173.
2. Xu, Cheng-Zhong (2005). *Scalable and Secure Internet Services and Architecture*. CRC Press. ISBN 9781420035209.
3. "Research Report - In Demand – The Culture of Online Service Provision" (<https://web.archive.org/web/20140123013059/http://www.citrix.com/news/market-research/oct-2013/quocirca-in-demand-the-culture-of-online-service-provision.html>) . Citrix. 14 October 2013. Archived from the original (<http://www.citrix.com/news/market-research/oct-2013/quocirca-in-demand-the-culture-of-online-service-provision.html>) on 23 January 2014. Retrieved 30 January 2014.
4. Furht, Borko (2010). *Handbook of Cloud Computing*. Springer. ISBN 9781441965240.
5. Nolle, Tom. "Designing public cloud applications for a hybrid cloud future" (<http://searchcloudcomputing.techtarget.com/tip/Designing-public-cloud-applications-for-a-hybrid-cloud-future>) . Tech Target. Retrieved 30 January 2014.
6. Randles, Martin, David Lamb, and A. Taleb-Bendiab. "A comparative study into distributed load balancing algorithms for cloud computing." *Advanced Information Networking and Applications Workshops (WAINA), 2010 IEEE 24th International Conference on*. IEEE, 2010.

7. Ferris, James Michael. "Methods and systems for load balancing in cloud-based networks." U.S. Patent Application 12/127,926.
8. Wang, S. C.; Yan, K. Q.; Liao, W. P.; Wang, S. S. (2010), "Towards a load balancing in a three-level cloud computing network", *Proceedings of the 3rd International Conference on Computer Science and Information Technology (ICCSIT)*, IEEE: 108–113, ISBN 978-1-4244-5537-9
9. Kansal, Nidhi Jain, and Inderveer Chana. "Cloud load balancing techniques: A step towards green computing." *IJCSI International Journal of Computer Science Issues* 9.1 (2012): 1694-0814.
10. Wee, Sewook, and Huan Liu. "Client-side load balancer using cloud." *Proceedings of the 2010 ACM Symposium on Applied Computing*. ACM, 2010.

Retrieved from

["https://en.wikipedia.org/w/index.php?](https://en.wikipedia.org/w/index.php?title=Cloud_load_balancing&oldid=1088513485)

[title=Cloud_load_balancing&oldid=1088513485"](https://en.wikipedia.org/w/index.php?title=Cloud_load_balancing&oldid=1088513485)

Last edited 2 months ago by VulcanSphere

