**Chương II: Tổng quan về cân bằng tải**

**1. Khái niệm cân bằng tải**

Cân bằng tải là một phương pháp phân phối khối lượng tải trên nhiều máy tính hoặc một cụm máy tính để có thể sử dụng tối ưu các nguồn lực, tối đa hóa thông lượng , giảm thời gian đáp ứng và tránh tình trạng quá tải trên máy chủ. [1]



**Hình 1.** Khi người dùng gửi requests đến một web server thì proxy server sẽ có nhiệm vụ tiếp nhận rồi phân phối lượng requests đó cho server 1 và server 2 .

**2. Ưu nhược điểm của cân bằng tải**

**a. Ưu điểm**

\* Tăng độ tin cậy và khả năng dự phòng cho hệ thống: Sử dụng cân bằng tải giúp tăng tính sẵn sàng cao HA ( High Availability )[2] cho hệ thống, đồng thời đảm bảo cho người dùng không bị gián đoạn dịch vụ khi xảy ra lỗi sự cố lỗi tại một thời điểm cung cấp dịch vụ.

\* Tăng tính bảo mật cho hệ thống:



Thông thường khi người dùng gửi yêu cầu đến webserver thì webserver sẽ truy vấn vào database rồi trả về yêu cầu cho người dùng (Hình 2).Khi chúng ta xây dựng hệ thống load balancer (Hình 1) yêu cầu đó sẽ được xử lý trên bộ cân bằng tải (proxy server ), sau đó thành phần cân bằng tải mới chuyển tiếp các yêu cầu cho các máy chủ bên trong. Quá trình trả lời cho người dùng thông qua thành phần cân bằng tải , vì vậy mà người dùng không thể biết được chính xác các máy chủ bên trong cũng như phương pháp phân tải được sử dụng. Bằng cách này có thể ngăn chặn người dùng giao tiếp trực tiếp với các máy chủ , ẩn các thông tin và cấu trúc mạng nội bộ, ngăn ngừa các cuộc tấn công trên mạng hoặc các dịch vụ không liên quan đang hoạt động trên các cổng khác.

**b. Nhược điểm của cân bằng tải :**

Tuy nhiên đối với những trung tâm dữ liệu lớn như là của facebook, google hay youtube ... việc quá tải vẫn có thể xảy ra . Đối mặt với điều này , giải pháp truyền thống là mua thêm nhiều server , tăng băng thông truy cập cho server farm hoặc bổ sung các module cần thiết nhưng những giải pháp này có nhiều điểm không thích hợp cho phát triển tương lai vì nó kéo theo nhiều chi phí như mua nhiều server , chi phí cho băng thông kết nối , duy trì và quản trị , việc quản lý sẽ trở nên phức tạp hơn. [3]

**3 Các thuật toán cân bằng tải**

**a. Thuật toán Round Robin**



Hình 2.Đây gọi là thuật toán luân chuyển vòng, các máy chủ sẽ được xem ngang hàng và sắp xếp theo một vòng quay. Các truy vấn dịch vụ sẽ lần lượt được gửi tới các máy chủ theo thứ tự sắp xếp.

**b. Thuật toán Weighted Round Robin:**

Bản chất giống như thuật toán Round Robin, tuy nhiên chúng ta có thể cấu hình cho một máy chủ đó thường xuyên được sử dụng hơn.

**c. Thuật toán Least Connection**

Đây là thuật toán dựa trên tính toán số lượng kết nối để thực hiện cân bằng tải cho máy chủ, nó sẽ tự động lựa chọn máy chủ với số lượng kết nối đang hoạt động là nhỏ nhất.

**d. Thuật toán Weights Least Connection:**

Bản chất giống thuật toán Least Connection , nhưng chúng ta có thể cấu hình ưu tiên cho một máy chủ trong cụm máy chủ hoạt động.

**e. Thuật toán Least Response Time:**

Đây là thuật toán sử dụng phương pháp thời gian đáp ứng ít nhất, lựa chọn dịch vụ trên máy chủ với thời gian đáp ứng là thấp nhất.

Ngoài ra còn có rất nhiều thuật toán cân bằng tải khác tùy theo phần mềm hoặc phần cứng cân bằng tải được sử dụng.

**f. Hàm băm:**

Hàm băm là giải thuật nhằm sinh ra các giá trị băm tương ứng với mỗi database server. Giá trị băm đóng vai gần như một khóa để phân biệt các database server cũng giống với giải thuật Round-robin, hàm băm chỉ định server trả lời truy vấn cho client mà không quan tâm đến trạng thái hiện tại của server đó ( có ít hay nhiều truy vấn ).

**g. Giải thuật xác định tổng số kết nối nhỏ nhất :**

Giải thuật này xác định tổng số kết nối hiện tại trên các database server, nếu server nào có tổng kết nối nhỏ nhất thì server đó sẽ được chỉ định là trả lời truy vấn tiếp theo của client. Trong giải thuật này có quan tâm đến trạng thái của server. Đây là một giải thuật tối ưu để cân bằng tải cho các database server. Tuy nhiên, thời gian thực hiện bài toán là lâu hơn hai giải thuật trên, bởi vì nó phải ghi lại và tính số kết nối đến các server hiện thời.

*Trích Dẫn-*

[1] Tony Bourke, ”*Server Load Balancing* ”, O'Reilly Media , 14-15 , 2001

[2] <https://en.wikipedia.org/wiki/High_availability>

[3] <https://myhost.vn/load-balancing-va-cac-ky-thuat-can-bang-tai/>;<http://baophat.com/post/can-bang-tai-may-chu-load-balance-server-id21.html>