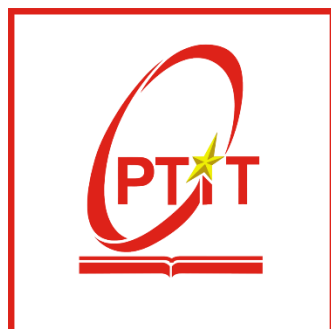


HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

KHOA AN TOÀN THÔNG TIN



BÁO CÁO BÀI TẬP LỚN

MÔN: AN TOÀN ỨNG DỤNG WEB VÀ CƠ SỞ DỮ LIỆU

**Đề tài: Tìm hiểu về chuỗi cân bằng tải các máy chủ web
(web server load balancing cluster).**

Họ và tên giảng viên	: Ninh Thị Thu Trang
Lớp/Nhóm	: 03/05
Họ và tên sinh viên	: Hoàng Thạch Hùng – B20DCAT074
	Ninh Chí Hướng – B20DCAT094

Hà Nội – 10/2023

Mục lục

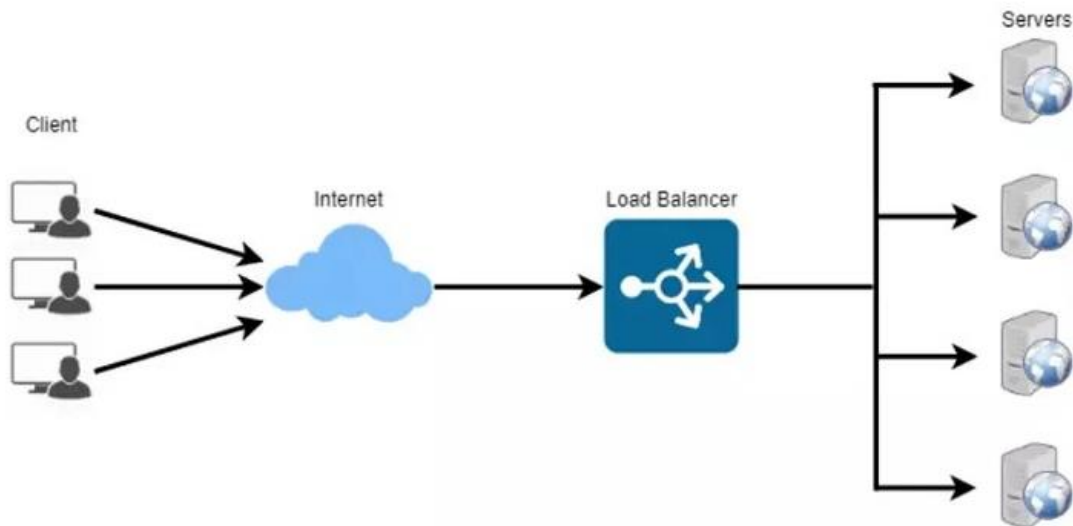
I. Khái quát về Load Balancing	3
1. Khái niệm về Load Balancing	3
2. Tầm quan trọng của Load Balancing	3
II. Cơ chế hoạt động của Load Balancing	4
III. Các loại Load Balancing	5
1. Cân bằng tải phần cứng	5
– Ưu điểm	6
– Nhược điểm	6
2. Cân bằng tải phần mềm	6
– Ưu điểm	6
– Nhược điểm	7
3. So sánh Load Balancing ở phần cứng và phần mềm	7
IV. Lợi ích của Load Balancing	7
1. Tăng tính sẵn sàng	7
2. Tính linh hoạt và mở rộng	7
3. Tăng tính bảo mật cho hệ thống, giám sát và phát hiện xâm nhập ...	8
4. Cải thiện khả năng chịu tải, tăng hiệu suất	8
V. Các giao thức được xử lý bởi Load Balancing	8
1. HTTP	8
2. HTTPS	9

3.	TCP	9
4.	UDP	9
VI.	Cách Load Balancing lựa chọn server backend	9
VII.	Thuật toán sử dụng trong Load Balancing	9
1.	Round Robin	9
2.	Weighted Round Robin.....	10
3.	Dynamic Round Robin.....	10
4.	Fastest Response Time	11
5.	Least Connections.....	11
6.	IP Hashing	11
VIII.	Load Balancing dự phòng.....	12
IX.	Demo.....	13
X.	Tài liệu tham khảo:	17

I. Khái quát về Load Balancing

1. Khái niệm về Load Balancing

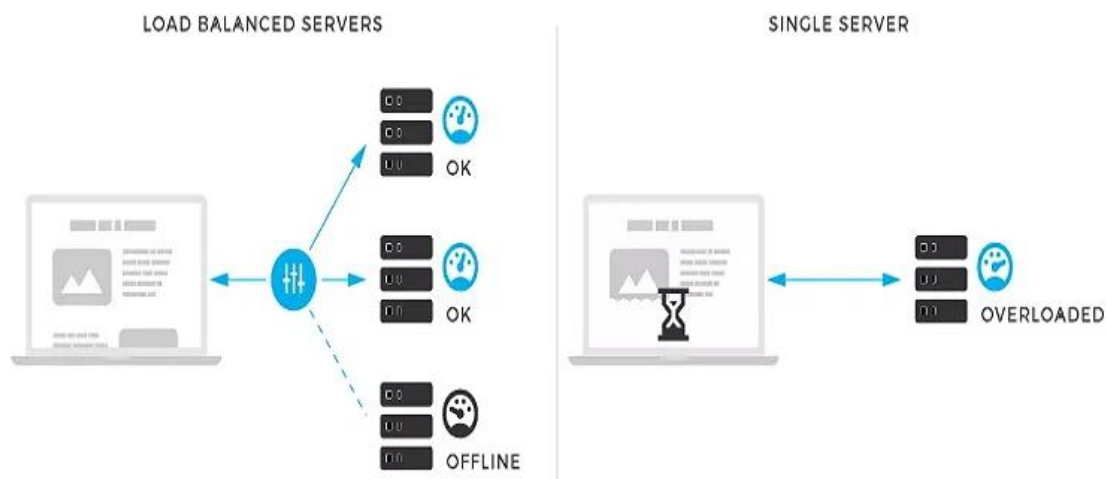
Load Balancing hay còn được biết đến với tên gọi trong tiếng Việt là cân bằng tải. Là quá trình của việc phân phối lưu lượng truy cập một cách hiệu quả thông qua nhiều server hay còn được gọi là server farm hay server pool. Load Balancing sẽ phân chia đều khối lượng công việc, truy cập của người dùng trên các máy chủ. Qua đó nâng cao khả năng chịu lỗi, đảm bảo tính sẵn sàng phục vụ của hệ thống.



Khi có một server trong pool bị down, không thể xử lý được, Load Balancer sẽ có chuyển hướng traffic đến các backend còn sống khác. Tương tự, khi một backend thêm vào pool, Load Balancer sẽ tự động đẩy traffic tới backend mới đó. Hoặc khi máy chủ chính hỏng hoặc nghẽn, không xử lý được thì một Load Balancer sẽ được bổ sung. Người dùng sẽ truy cập vào đó để được chuyển đến một máy chủ khác thực hiện tác vụ.

2. Tầm quan trọng của Load Balancing

Trong thời đại công nghệ bùng nổ, các website hiện đại có lưu lượng truy cập cao. Khi ấy, hệ thống sẽ phải tiếp nhận và xử lý tới hàng trăm nghìn, hàng triệu các requests đồng thời từ người dùng hoặc khách hàng. Sau đó, hệ thống lại cần phản hồi lại kết quả dưới dạng text, video, hình ảnh,... một cách chính xác, nhanh chóng.

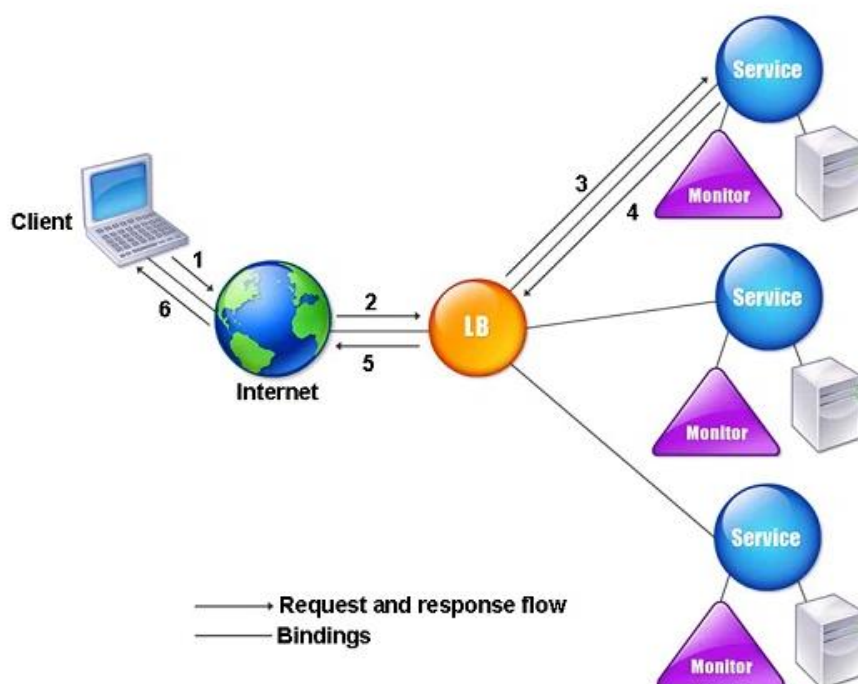


Nếu một website không có Load Balancing thì sẽ không thể kịp xử lý được khối lượng tác vụ khổng lồ trên. Thậm chí là không kết nối được khi máy chủ bị hỏng. Do đó, để giải quyết vấn đề này, mang lại trải nghiệm tuyệt vời của người dùng việc trang bị Load Balancing là rất cần thiết. Load Balancing sẽ giúp cân bằng tải hoàn hảo cho server.

II. Cơ chế hoạt động của Load Balancing

Một quá trình Load Balancing cơ bản sẽ diễn ra theo trình tự sau:

- Website tiếp nhận lượng truy cập của người dùng. Các yêu cầu này có thể đến máy chủ cùng một lúc thông qua mạng internet.
- Với lượng yêu cầu này, Load Balancing sẽ xử lý từng yêu cầu, điều hướng tới máy chủ thích hợp.
- Khi nhận được yêu cầu, máy chủ sẽ tiếp nhận yêu cầu, xử lý hoặc gửi lại phản hồi với Load Balancing trong trường hợp không thể thực hiện thêm.
- Quá trình xử lý kết thúc, máy chủ sẽ gửi phản hồi lại cho người dùng.



III. Các loại Load Balancing

Bộ cân bằng tải là một thiết bị nằm giữa người dùng và cụm máy chủ. Nó hoạt động như một người hỗ trợ vô hình đảm bảo rằng tất cả các tài nguyên máy chủ đều được sử dụng như nhau. Bộ cân bằng tải sẽ tồn tại dưới hai loại: Cân bằng tải phần cứng và phần mềm.

1. Cân bằng tải phần cứng

Cân bằng tải phần cứng là một thiết bị phần cứng có phần mềm tích hợp chuyên dụng và độc quyền được thiết kế để xử lý một lượng lớn lưu lượng ứng dụng. Các cân bằng tải này có khả năng ảo hóa tích hợp và cho phép sử dụng nhiều phiên bản cân bằng tải ảo trên một thiết bị. Theo truyền thống, các nhà cung cấp đã tải phần mềm độc quyền lên phần cứng chuyên dụng và bán nó cho người dùng dưới dạng thiết bị độc lập - thường theo cặp để cung cấp chuyển đổi dự phòng nếu một hệ thống gặp sự cố.



– Ưu điểm

Cân bằng tải phần cứng thường có hiệu suất cao. Nó phù hợp để xử lý tải lưu lượng truy cập cao điểm nên tốc độ cao và hiệu suất tốt hơn so với cân bằng tải phần mềm. Nó còn khả năng bảo mật tốt và tối ưu hóa mạng nâng cao vì được xử lý bởi tổ chức chứ không phải bởi bất kỳ bên thứ ba nào.

– Nhược điểm

- Khi số kết nối đạt đến giới hạn, các kết nối mới sẽ bị từ chối, bị rút hoặc xuống cấp. Cách duy nhất để khắc phục điều này là mua và cài đặt các máy bổ sung. Tuy nhiên, lại không thể mở rộng quy mô ngay lập tức.
- Khó quản lý và mở rộng. Đôi khi, sẽ phải cần thuê chuyên gia tư vấn để quản lý nó.
- Phải có chuyên môn để cấu hình và lập trình nó.
- Chi phí mua và bảo trì cao nên đắt hơn cân bằng tải phần mềm.

2. Cân bằng tải phần mềm.

Cân bằng tải phần mềm chạy trên máy ảo (VM) hoặc server hộp trắng, rất có thể là chức năng bộ điều khiển phân phối ứng dụng (ADC). ADC thường cung cấp các tính năng bổ sung, bao gồm bộ nhớ đệm, nén và định hình lưu lượng truy cập. Phổ biến trong môi trường đám mây, cân bằng tải ảo có thể mang lại mức độ linh hoạt cao. Ví dụ: nó cho phép người dùng tự động tăng hoặc giảm quy mô để phản ánh lưu lượng truy cập tăng đột biến hoặc giảm hoạt động mạng.

Một số cân bằng tải phần mềm phổ biến: Nginx, HAProxy, AWS Elastic Load Balancer,...



– Ưu điểm

- Linh hoạt và dễ dàng tùy chỉnh: Chúng ta có thể linh hoạt tùy biến, mở rộng và quản lý tải dễ dàng hơn so với cân bằng tải phần cứng. Hơn thế nữa, còn có nhiều phiên bản phần mềm cho chúng ta lựa chọn khi có nhu cầu mở rộng quy mô vượt quá khả năng ban đầu.
- Chi phí rẻ hơn so với cân bằng tải phần cứng.

– Nhược điểm

Khi mở rộng quy mô vượt quá dung lượng có thể gây ra độ trễ ban đầu. Điều này thường xảy ra khi phần mềm cân bằng tải đang được cấu hình, phải xử lý thêm dữ liệu tại phần mềm. Cân bằng tải phần mềm có thể yêu cầu nhiều tài nguyên hệ thống và có thể không đủ mạnh để xử lý lượng lớn yêu cầu trong một số trường hợp.

3. So sánh Load Balancing ở phần cứng và phần mềm

Việc cân bằng tải có thể được thực hiện ở cả phần cứng lẫn phần mềm. Tất nhiên, mỗi cách đều có những ưu điểm và nhược điểm riêng. Bảng so sánh sau sẽ cung cấp sơ lược đặc điểm chung của việc cân bằng tải ở phần cứng và phần mềm.

	Phần cứng	Phần mềm
Hiệu năng	Cao hơn	Thấp hơn
Tính linh hoạt	Kém linh hoạt	Linh hoạt hơn
Khả năng ảo hoá	Tích hợp sẵn	Không tích hợp sẵn
Kiến trúc	Nhiều dung lượng vật lý	Dung lượng vật lý thấp
Giá thành	Cao, đặc biệt là chi phí bảo trì	Tương đối thấp
Khả năng cấu hình	Thấp	Cao

IV. Lợi ích của Load Balancing

Load Balancing hoạt động giống như một tấm lọc các lưu lượng truyền qua internet. Lợi ích của nó mang tới cho web không hề nhỏ. Một số lợi ích chính của Load Balancing có thể kể tới như:

1. Tăng tính sẵn sàng

Khi triển khai load balancing, nếu một máy chủ gặp sự cố, các yêu cầu có thể được tự động chuyển đến các máy chủ khác trong cụm. Điều này giúp giảm thiểu thời gian chết và đảm bảo rằng dịch vụ vẫn hoạt động ngay cả khi có sự cố xảy ra. Tính sẵn sàng cao hơn mang lại sự tin cậy và khả năng khôi phục cho hệ thống.

2. Tính linh hoạt và mở rộng

Load balancing cho phép mở rộng khả năng của hệ thống bằng cách thêm hoặc xóa các máy chủ mới vào cụm mà không ảnh hưởng đến khả năng hoạt động của hệ thống. Khi lưu lượng tăng lên, load balancer tự động phân phối công việc đến các máy chủ mới để giữ cho hệ thống hoạt động trơn tru, đáp ứng được nhu cầu người dùng và mở rộng lượng truy cập vào website. Điều này giúp tăng khả năng mở rộng

của hệ thống một cách linh hoạt và tiết kiệm chi phí so với việc phải nâng cấp một máy chủ duy nhất.

3. Tăng tính bảo mật cho hệ thống, giám sát và phát hiện xâm nhập

Khi sử dụng load balancing, toàn bộ máy chủ ứng dụng và cơ sở dữ liệu được nằm trong vùng DMZ (DeMilitary Zone). Vùng này như một mạng private cô lập, không public. Khi người dùng gửi yêu cầu đến hệ thống, yêu cầu đó sẽ được bộ cân bằng tải tiếp nhận. Sau đó, bộ cân bằng tải sẽ định tuyến từng yêu cầu đến máy chủ phù hợp nhất xử lý. Nó cũng khiếm luôn nhiệm vụ phản hồi kết quả cho người dùng. Vì thế sẽ ngăn người dùng giao tiếp trực tiếp với máy chủ, ẩn các thông tin và cấu trúc mạng nội bộ, ngăn ngừa các truy cập trái phép.

Load balancing cho phép giám sát hoạt động của các máy chủ trong cụm, bao gồm việc theo dõi lưu lượng, hiệu suất, tình trạng hoạt động và chặn nội dung độc hại. Điều này cung cấp khả năng phát hiện và phản ứng nhanh chóng đối với các hoạt động xâm nhập hoặc bất thường trong hệ thống. Tự động chuyển hướng lưu lượng tấn công đến nhiều máy chủ phụ trợ để giảm thiểu tác động.

4. Cải thiện khả năng chịu tải, tăng hiệu suất

Load balancing giúp phân phối tải trọng một cách cân bằng giữa các máy chủ, giảm thiểu tình trạng quá tải trên một máy chủ duy nhất. Điều này giúp tăng khả năng chịu tải của hệ thống và đảm bảo rằng nó có thể xử lý số lượng lớn yêu cầu từ người dùng mà không gặp vấn đề về hiệu suất.



V. Các giao thức được xử lý bởi Load Balancing

Quản trị viên của hệ thống cân bằng tải có thể tạo các quy định chuyển tiếp đối với bốn loại giao thức chính:

1. HTTP

Cách thức hoạt động dựa trên cơ chế HTTP chuẩn. HTTP Balancing sẽ đưa ra yêu cầu về tác vụ. Tiếp theo, Load Balancer đặt X-Forwarded-For, X-Forwarded-

Proto và tiêu đề X-Forwarded-Port cung cấp các thông tin Backends về những yêu cầu ban đầu.

2. HTTPS

Cân bằng HTTPS hoạt động tương tự như HTTP, với sự bổ sung mã hóa. HTTPS Balancing được bổ sung mã hóa và xử lý bằng 2 cách. Đầu tiên chính là passthrough SSL duy trì mã hóa tất cả con đường đến Backend. Kiểu khác là chấm dứt SSL, đặt gánh nặng giải mã vào Load Balancer và gửi lưu lượng được mã hóa đến Backend.

3. TCP

Khi không sử dụng được HTTP hay HTTPS, TCP sẽ là giải pháp cân bằng lưu lượng. Cụ thể, khi một lượng Traffic đi vào cụm cơ sở dữ liệu, giao thức sẽ giúp lan truyền lưu lượng đến tất cả máy chủ.

4. UDP

Load Balancer đã bổ sung thêm hỗ trợ cho cân bằng tải giao thức Internet lỗi như DNS và syslog sử dụng UDP.

VI. Cách Load Balancing lựa chọn server backend

Các hệ thống cân bằng tải sẽ chọn cloud server để chuyển tiếp yêu cầu dựa trên sự kết hợp của hai yếu tố. Đầu tiên là đảm bảo các cloud server đã chọn có thể phản hồi chính xác các yêu cầu, và sau đó là sử dụng một quy tắc được cấu hình sẵn để “nhặt” một hay nhiều cloud server trong số này.

Hệ thống cân bằng tải chỉ chuyển tiếp dữ liệu đến các server backend “khỏe mạnh”. Để kiểm tra sức khỏe của server backend, hệ thống sẽ thường xuyên kết nối đến backend server thông qua các giao thức và cổng thiết lập bởi quy tắc chuyển tiếp, từ đó đảm bảo server có phản hồi. Nếu server không vượt qua được bước kiểm tra này, chúng sẽ tự động bị loại bỏ khỏi hệ thống cho đến khi có thể phản hồi lại bước kiểm tra sức khỏe đó.

VII. Thuật toán sử dụng trong Load Balancing

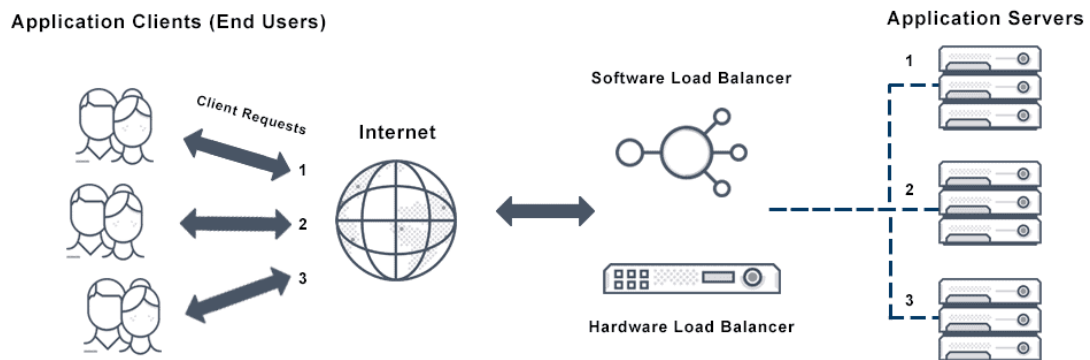
Tùy thuộc vào công nghệ Load Balancing sẽ ứng dụng thuật toán khác nhau. Dưới đây là phân tích về cách thức hoạt động của những loại thường thấy nhất.

1. Round Robin

Round Robin được biết đến là thuật toán lựa chọn các máy chủ theo trình tự. Theo đó, Load Balancer sẽ bắt đầu đi từ Server số 1 trong danh sách ứng với yêu cầu đầu tiên. Tiếp theo, cân bằng tải di chuyển dần xuống các đối tượng phía dưới. Khi đã đi tới máy chủ cuối cùng sẽ quay lại ở đầu trang. Đây là phương pháp đơn giản và công bằng, đảm bảo cân bằng tải trọng giữa các máy chủ.

Tuy nhiên, thuật toán này cũng tồn tại nhược điểm nhất định. Điều này xảy ra khi 2 yêu cầu từ người dùng được gửi đến các Server khác nhau. Hoạt động này làm tốn thời gian kết nối với máy chủ thứ hai. Trong khi đó, ngay tại Server đầu tiên đã giải

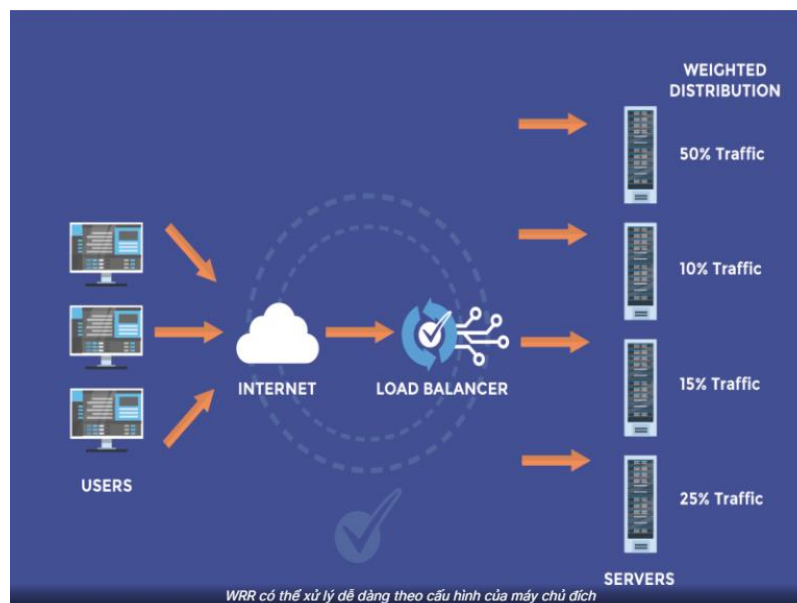
quyết được thông tin cần đáp ứng. Để khắc phục, Round Robin thường cài đặt cùng phương pháp duy trì session như sử dụng cookie.



2. Weighted Round Robin

Ngoài những điều có ở thuật toán Round Robin, WRR còn có khả năng xử lý theo cấu hình của từng Server đích. Theo đó, mỗi máy chủ được đánh giá trị bằng một trọng số. Trọng số này thể hiện khả năng xử lý của từng máy chủ.

Một Server có khả năng xử lý gấp đôi Server khác sẽ được nhận con số lớn hơn. Kéo theo đó là số lượng Request gấp hai từ bộ cân bằng tải. Điều này cho phép điều chỉnh tải trọng một cách linh hoạt dựa trên hiệu suất và khả năng của từng máy chủ. Nhược điểm là làm suy giảm Load Balancing nếu yêu cầu thay đổi liên tục trong một khoảng thời gian.



3. Dynamic Round Robin

Thuật toán này hoạt động gần giống với Weighted Round Robin. Thế nhưng, điểm khác biệt nằm ở trọng số dựa trên sự kiểm tra Server liên tục. Do đó, trọng số liên tục thay đổi.

Việc chọn máy chủ sẽ tính toán từ nhiều khía cạnh trong phân tích hiệu năng máy chủ tại thời gian thực. Thuật toán thường không cài đặt trong bộ cân bằng tải đơn giản. Thay thế vào đó là những ứng dụng trong Load Balancing của F5 Network

4. Fastest Response Time

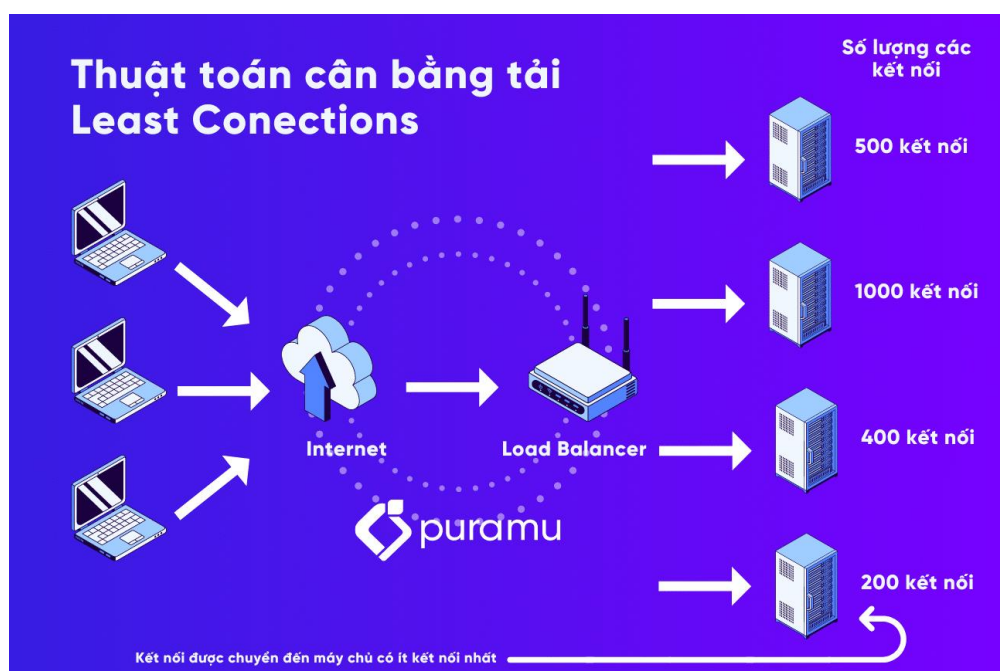
Fastest hoạt động dựa trên tính toán thời gian đáp ứng của mỗi Server. Sau đó, hệ thống sẽ chọn ra máy chủ có khả năng phản hồi nhanh nhất. Thời gian được xác định bằng hai mốc là thời điểm gửi một gói tin đến Server và nhận gói trả lời. Quy trình đến và đi này sẽ được bộ Load Balancing đảm nhiệm. Theo đó, Load Balancing tự đánh giá được cần chuyển yêu cầu tiếp theo đến đâu.

Thuật toán này áp dụng cho các máy chủ nằm tại khu vực địa lý khác nhau. Vì thế, người dùng ở gần vị trí nào sẽ được đáp ứng từ Server gần nhất.

5. Least Connections

Nguyên lý là Request sẽ chuyển đến Server có ít kết nối nhất hệ thống, cách này được khuyên dùng khi tốc độ truy cập bị chậm. Điều này giúp phân phối tải trọng một cách tự động và linh hoạt, đồng thời tránh quá tải trên các máy chủ. Đây được coi là thuật toán động vì phải đếm số Connection đang hoạt động của máy chủ.

Trong thực tế, Least Connections hoạt động rất tốt. Điều này vẫn đúng ngay cả khi các kết nối biến thiên trong khoảng thời gian lớn.



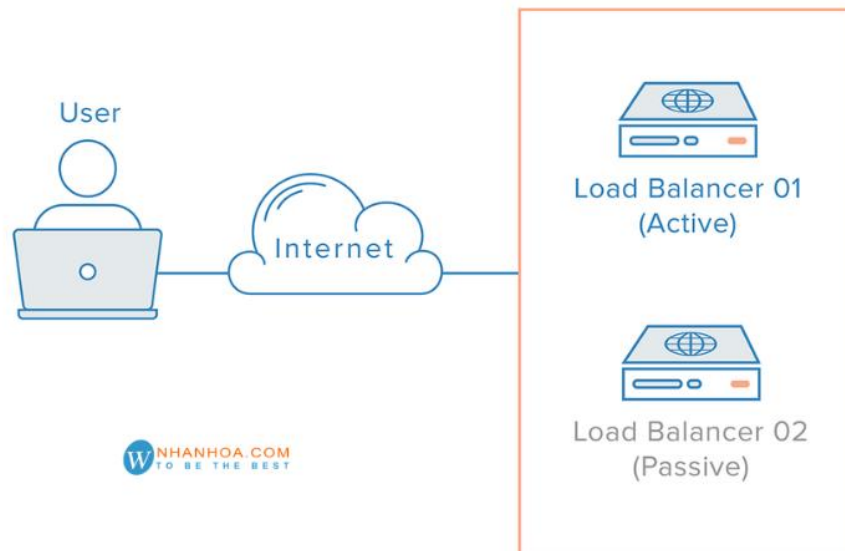
6. IP Hashing

Phương pháp IP Hashing sử dụng địa chỉ IP nguồn của yêu cầu để xác định máy chủ mục tiêu. Một thuật toán băm được áp dụng cho địa chỉ IP và kết quả băm được sử dụng để xác định máy chủ. Điều này đảm bảo rằng các yêu cầu từ cùng một địa

chỉ IP sẽ luôn được chuyển đến cùng một máy chủ, giúp duy trì tính nhất quán của các phiên làm việc.

VIII. Load Balancing dự phòng

Để ngăn chặn khả năng ảnh hưởng đến toàn hệ thống nếu bộ cân bằng tải gặp trục trặc, một bộ cân bằng tải dự phòng có thể kết nối vào để tạo thành một cụm cân bằng tải. Mỗi bộ cân bằng tải đều có thể phát hiện lỗi và phục hồi.

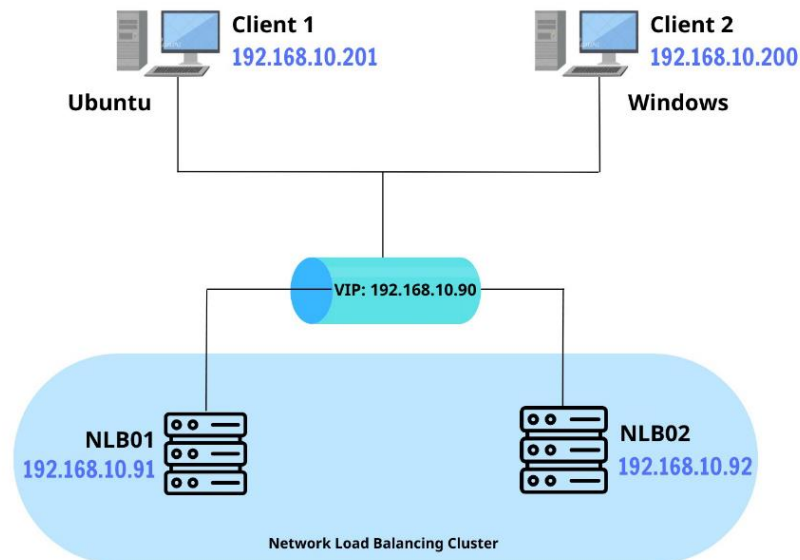


Trong trường hợp bộ cân bằng tải chính bị lỗi, DNS sẽ đưa người dùng đến với bộ cân bằng tải dự phòng. Để tiết kiệm thời gian khi quá trình thay đổi DNS có thể khá lâu mới thông báo được lên Internet, ngoài ra để việc chống lỗi này hoạt động tự động thì nhiều quản trị viên sử dụng hệ thống cho phép linh hoạt thay đổi IP, chẳng hạn như Floating IP.

Bằng cách cung cấp một địa chỉ IP tĩnh tùy chỉnh được khi cần thiết, các vấn đề về mất thời gian thông báo lên Internet và lưu bộ nhớ đệm khi thay đổi DNS có thể loại bỏ. Tên miền có thể duy trì liên kết với cùng một địa chỉ IP, trong khi địa chỉ IP này được di chuyển giữa các cloud server.

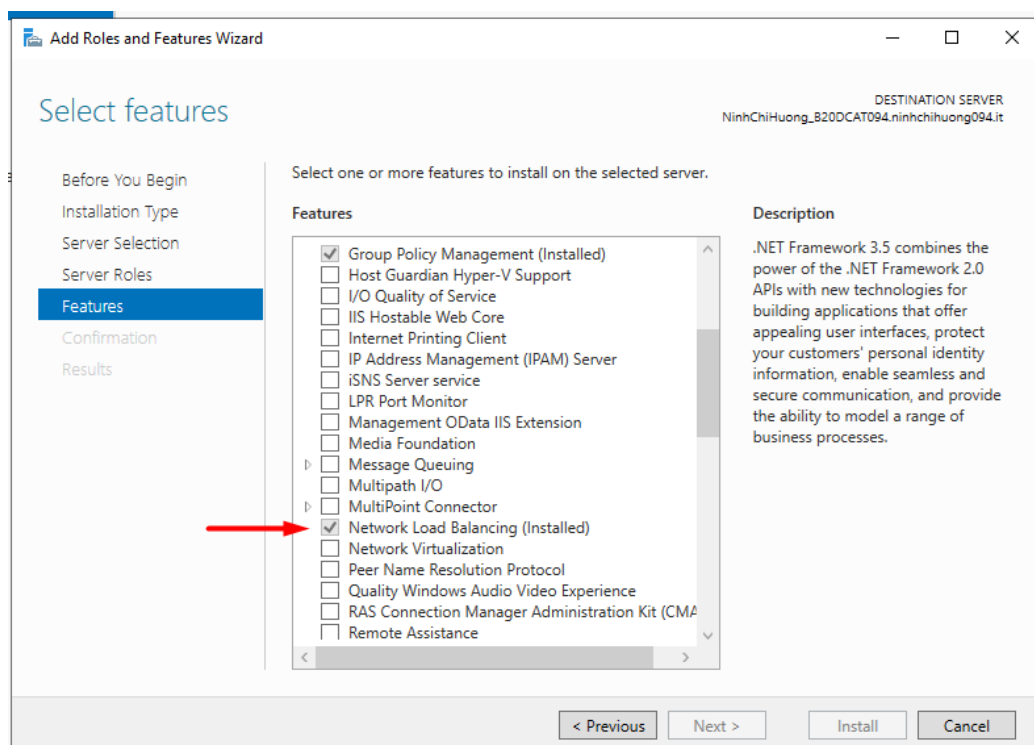
IX. Demo

Ở phần demo này, nhóm em sẽ tạo một web cluster gồm 2 node đều là Windows server 2019 và sử dụng tính năng Network Load Balancing của MS Windows server để cân bằng tải. Cấu trúc của mạng của phần demo sẽ được thực hiện theo mô hình sau :

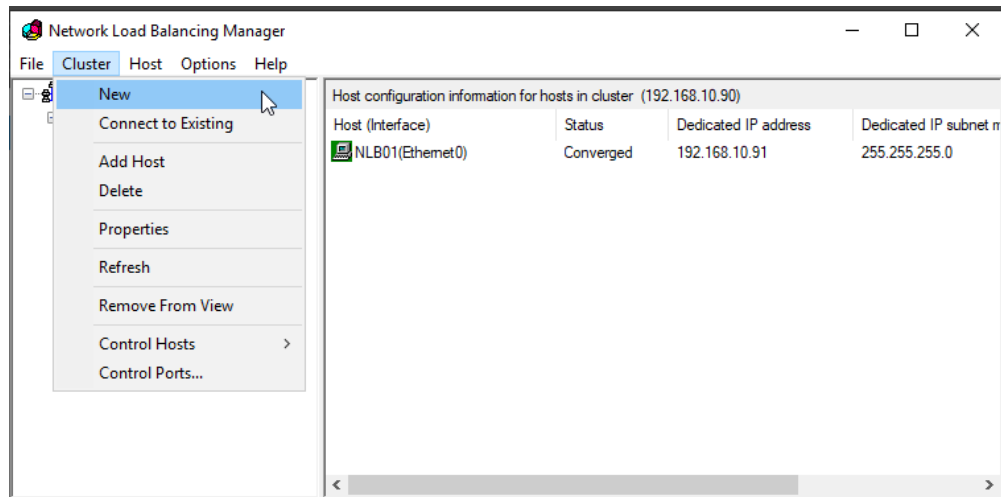


Quá trình thực hiện :

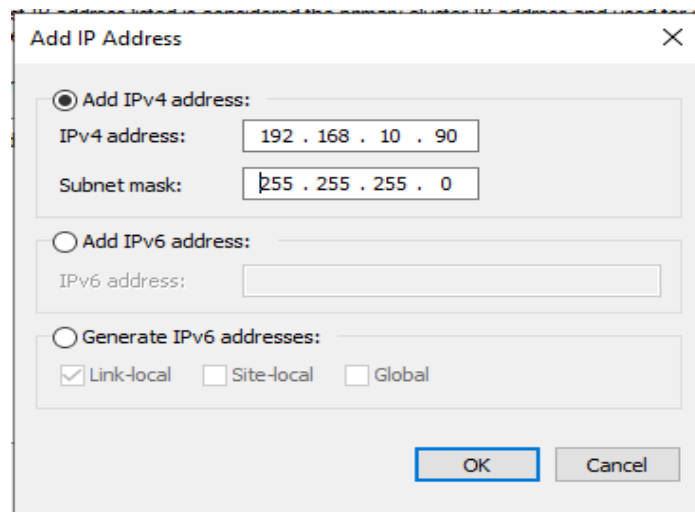
- Trước tiên, trên cả 2 máy Server, ta cài đặt dịch vụ Network Load Balancing



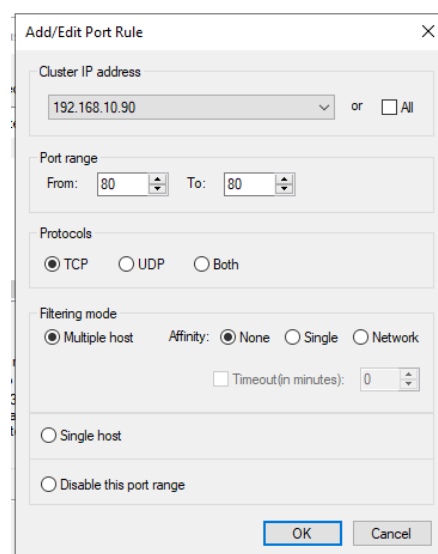
Trong cửa sổ NLB manager ta tạo một cluster . Sau đó, thêm các host vào cluster đó.



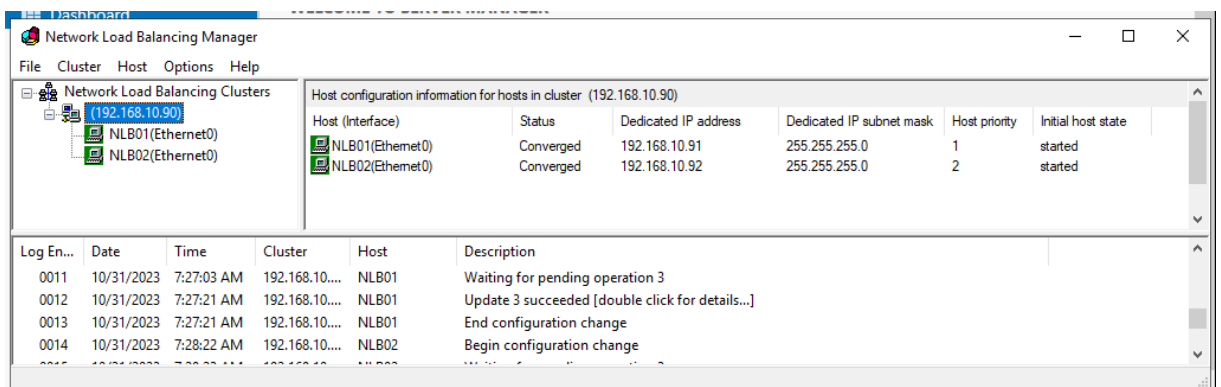
Bây giờ, ta sẽ gán một virtual IP cho cluster, Client sẽ dùng địa chỉ ip này để truy cập vào dịch vụ.



Ta sẽ chọn tùy chọn “None” để có thể phân bổ đều request vào các node .

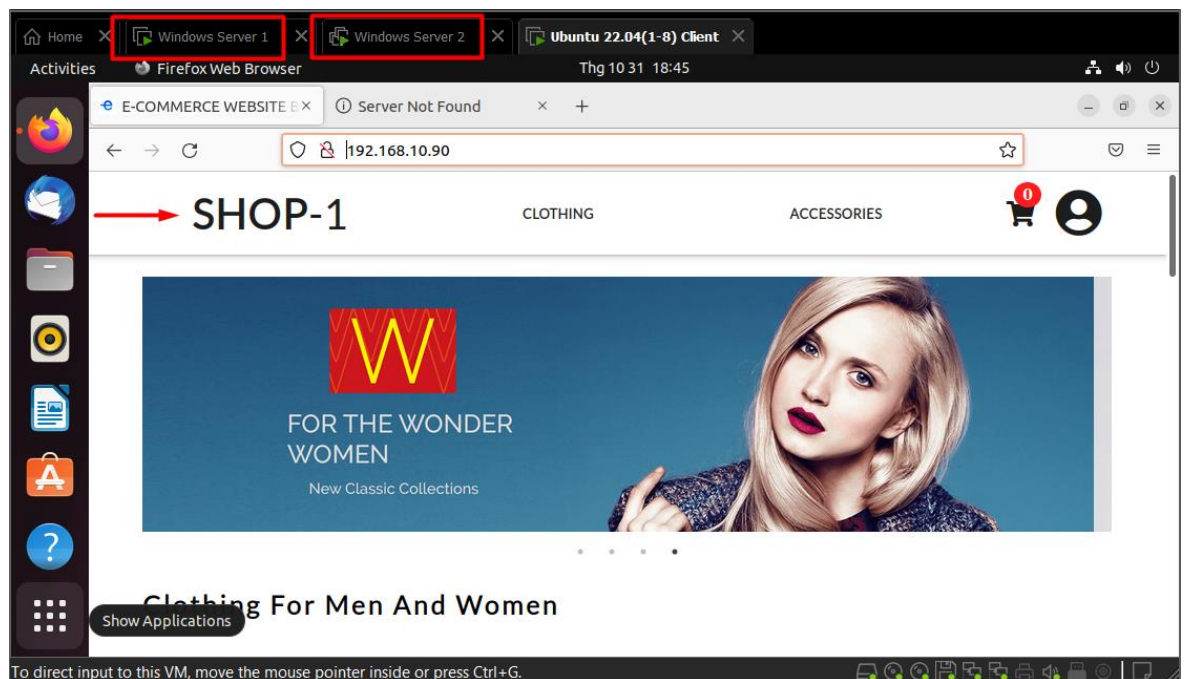


Sau đó, ta thêm lần lượt 2 server NLB01 và NLB02 vào cluster, và đồng thời gán giá trị priority cho từng node. Node nào có priority thấp hơn thì sẽ ưu tiên xử lý các request từ user.

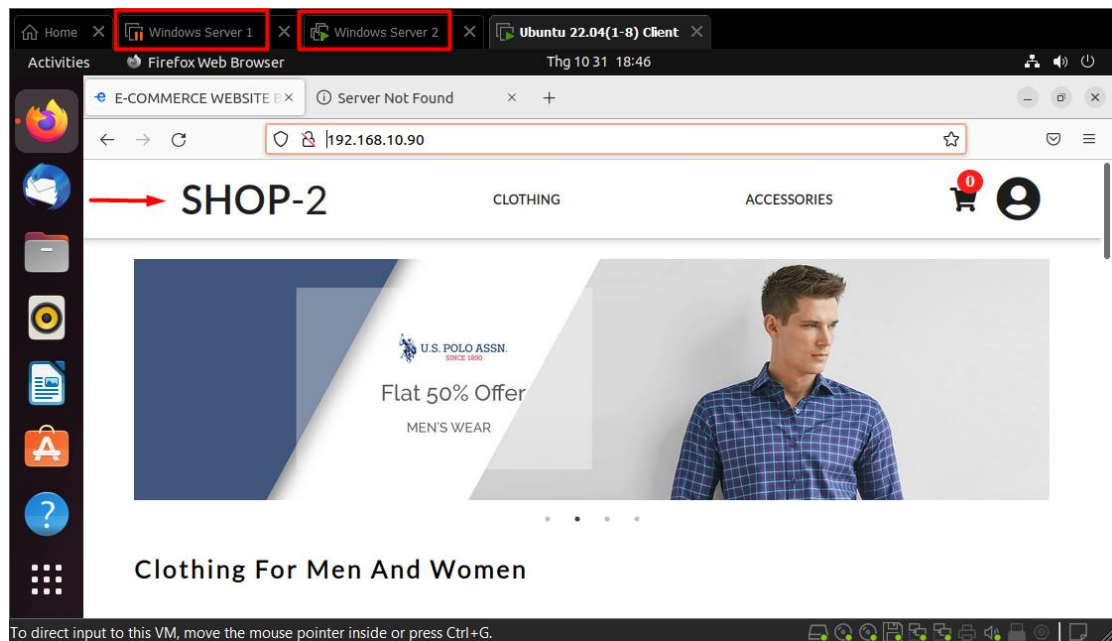


Cuối cùng, ta sẽ kiểm tra hệ thống có hoạt động không bằng cách vào máy client và truy cập vào địa chỉ 192.168.10.90 trên trình duyệt, ở đây ta thử nghiệm với 3 trường hợp :

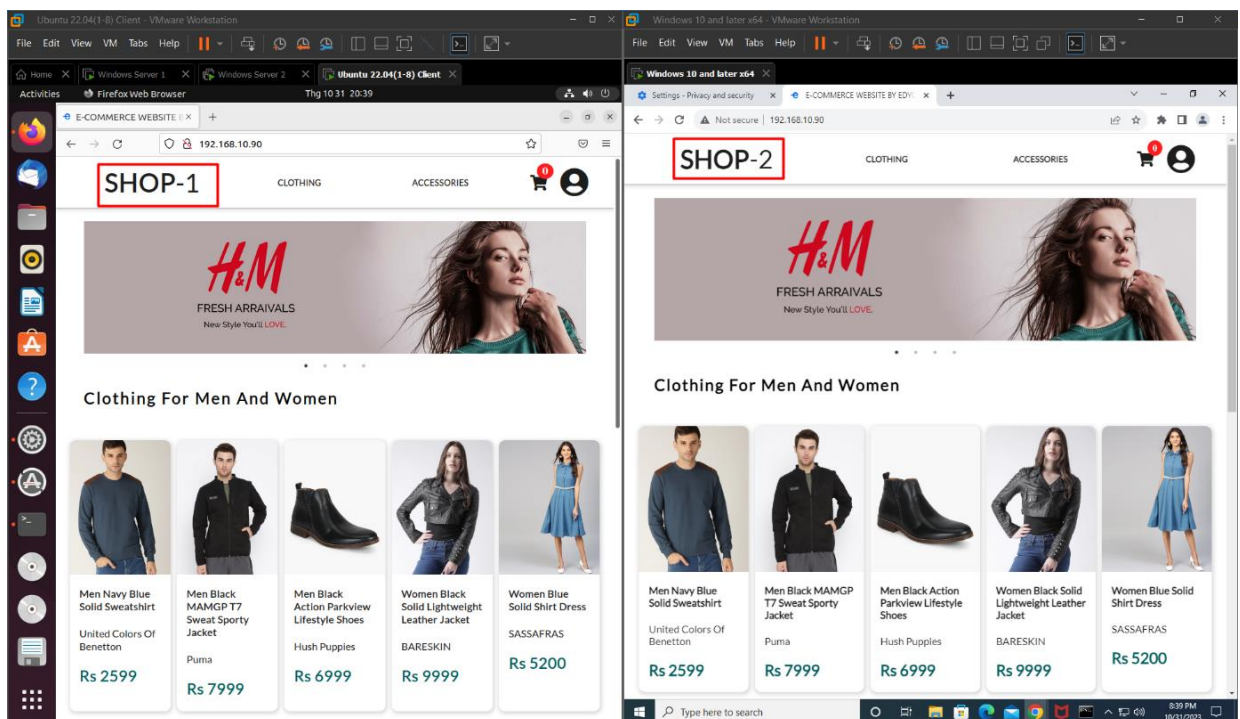
+) **Trường hợp 1:** cả 2 server đều đang hoạt động. Kết quả là hệ thống trả về web của server 1



+) **Trường hợp 2 :** Tắt server 1 đi. Kết quả là hệ thống trả về web của server 2.



+)**Trường hợp 3:** Cả 2 server đều đang hoạt động và lần này ta sẽ dùng 2 client để truy cập vào. Kết quả là nếu client-1 đang truy cập web của server 1 thì client-2 sẽ truy cập vào web của server 2 và ngược lại.



X. Tài liệu tham khảo:

- Tìm hiểu về cơ chế load balancing: <https://viblo.asia/p/tim-hieu-ve-co-che-load-balancing-GrLZD0X2Zk0?fbclid=IwAR1JSBkBBc4TsmzuFARzSDKa5UmDo7LYTWW3vIhMZRIw348gGBdEzxob0A8>
- Các giải pháp cân bằng tải :<https://viettelidc.com.vn/tin-tuc/can-bang-tai-load-balancing-la-gi-va-tai-sao-phai-dung-can-bang-tai>
- Tìm hiểu về load balancing: <https://fptcloud.com/load-balancing/>
- Network Load Balancing : <https://learn.microsoft.com/en-us/windows-server/networking/technologies/network-load-balancing>
- Configuring Network Load Balancing (NLB) for a Windows Server cluster : <https://4sysops.com/archives/configuring-network-load-balancing-nlb-for-a-windows-server-cluster/>