

Get unlimited access to the best of Medium for less than \$1/week. [Become a member](#)



The Essential Guide to Load Balancing Strategies and Techniques



Hayk Simonyan · [Follow](#)

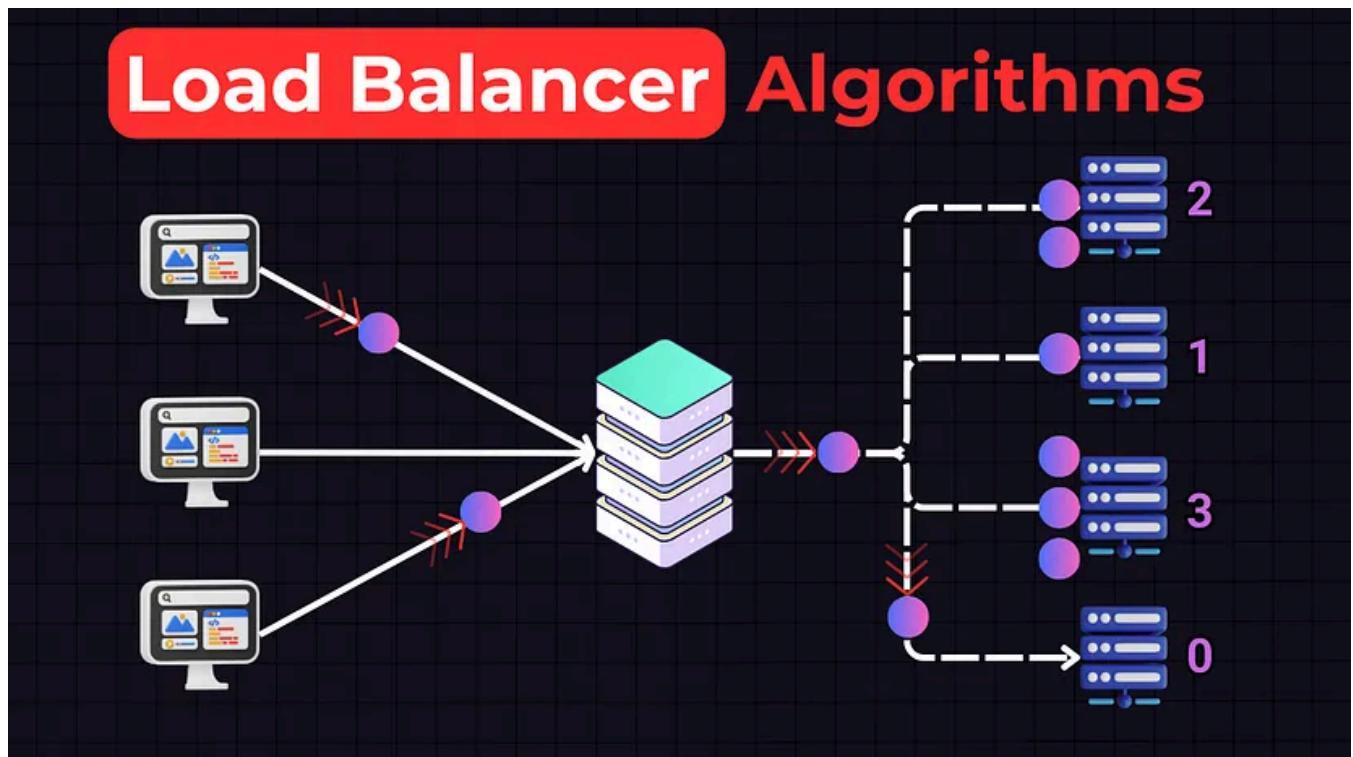
Published in Level Up Coding

5 min read · Feb 8, 2024

Listen

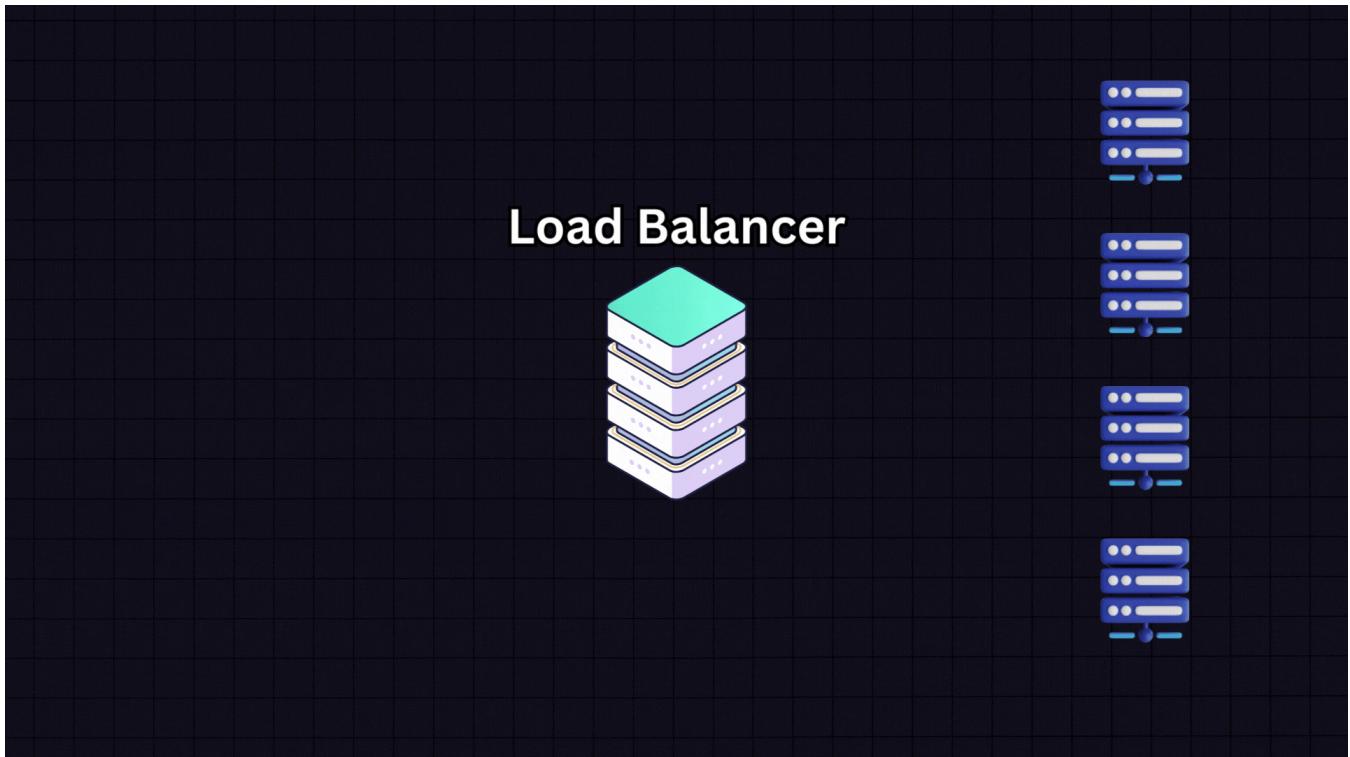
Share

More



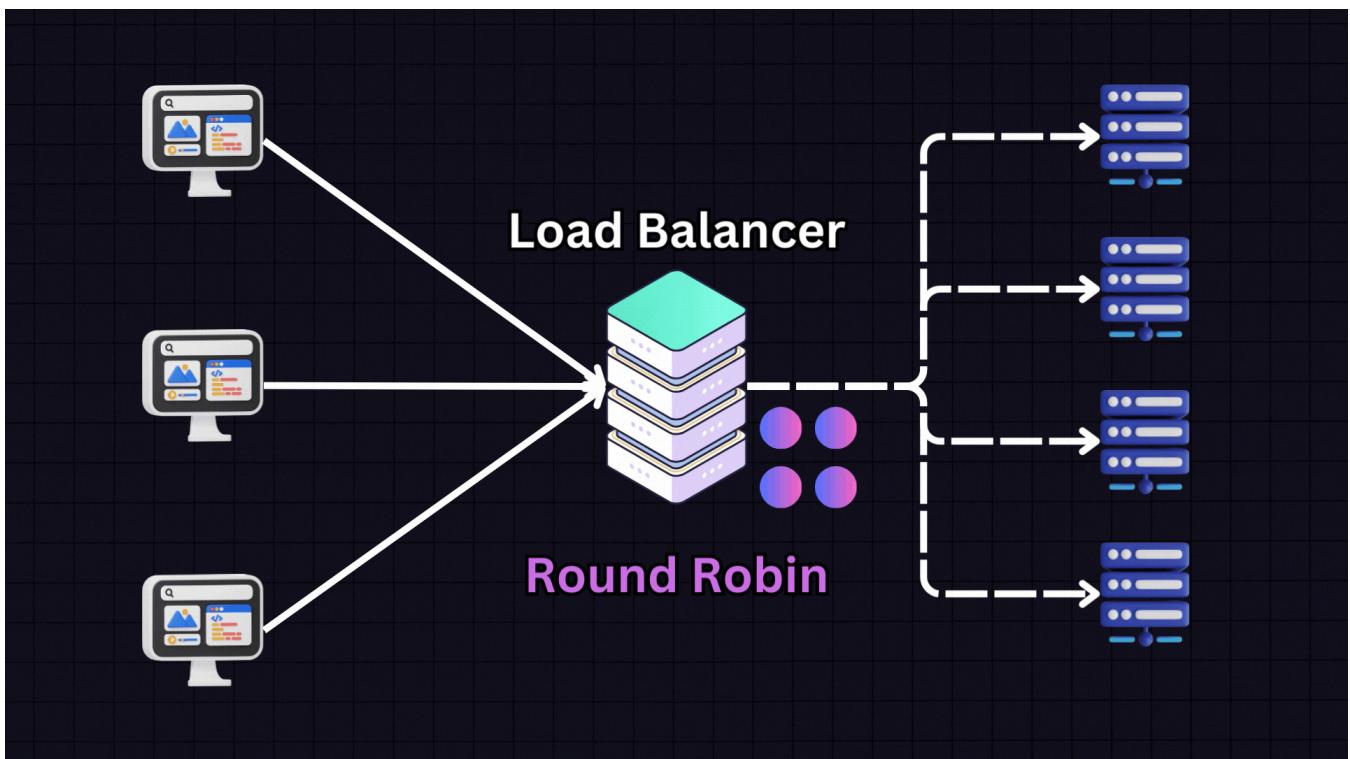
What are Load Balancers?

Load balancers distribute incoming network traffic across multiple servers to ensure no single server bears too much load. By spreading the requests efficiently, they increase the capacity and reliability of applications. Here are some common strategies and algorithms used in load balancing:



1. Round Robin

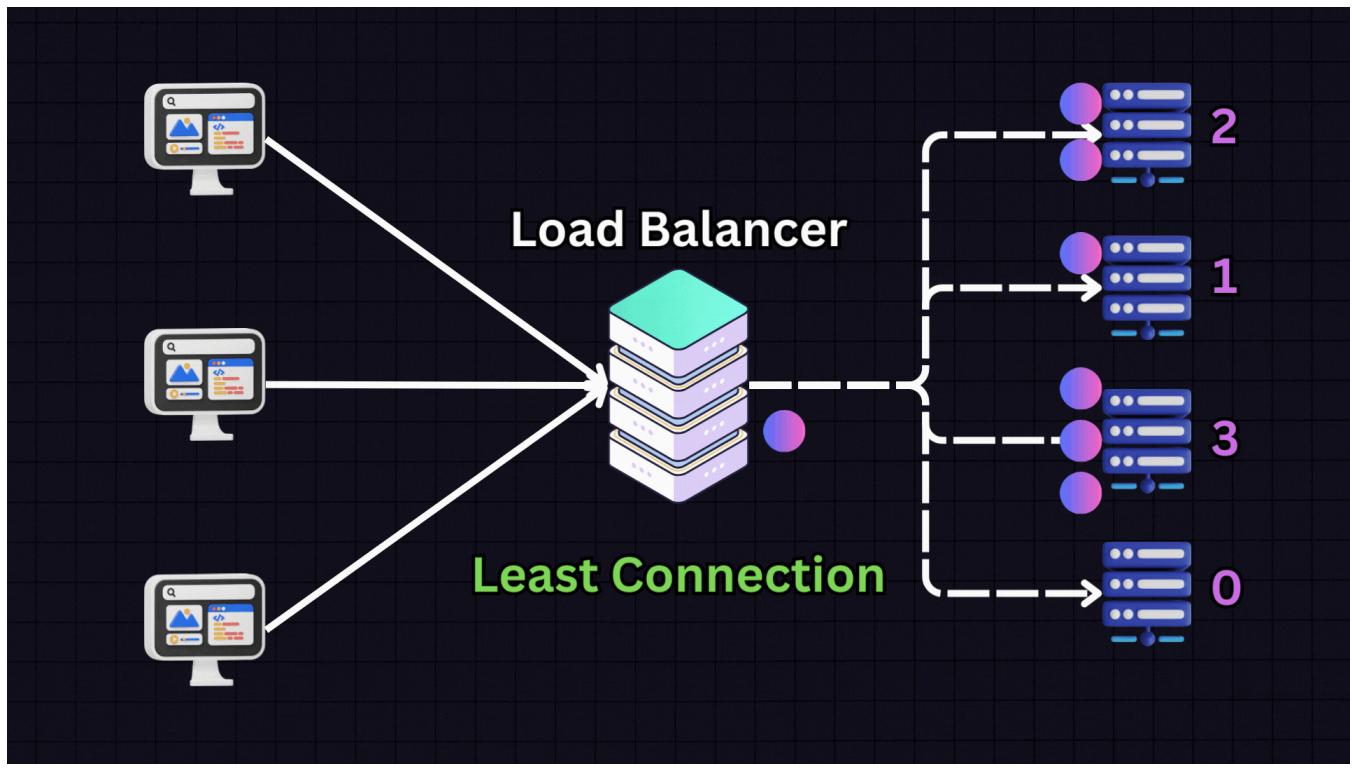
Round Robin is the simplest form of load balancing, where each server in the pool gets a request in a sequential, rotating order. When the last server is reached, it loops back to the first.



This works well for servers with similar specifications and when the load is uniformly distributable.

2. Least Connections

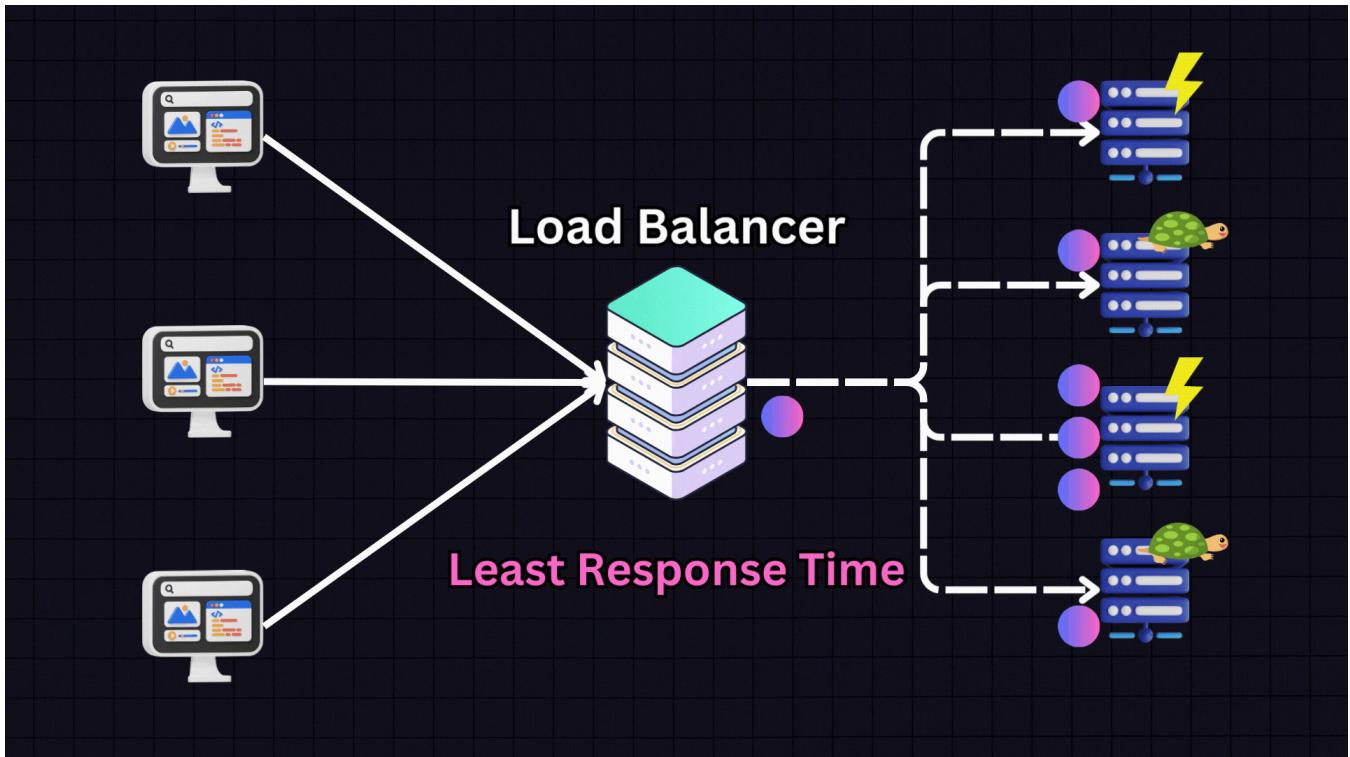
Least Connections algorithm directs traffic to the server with the **fewest active connections**. This is particularly useful when there are sessions of variable lengths and demands.



It is ideal for longer tasks or when the server load is not evenly distributed.

3. Least Response Time

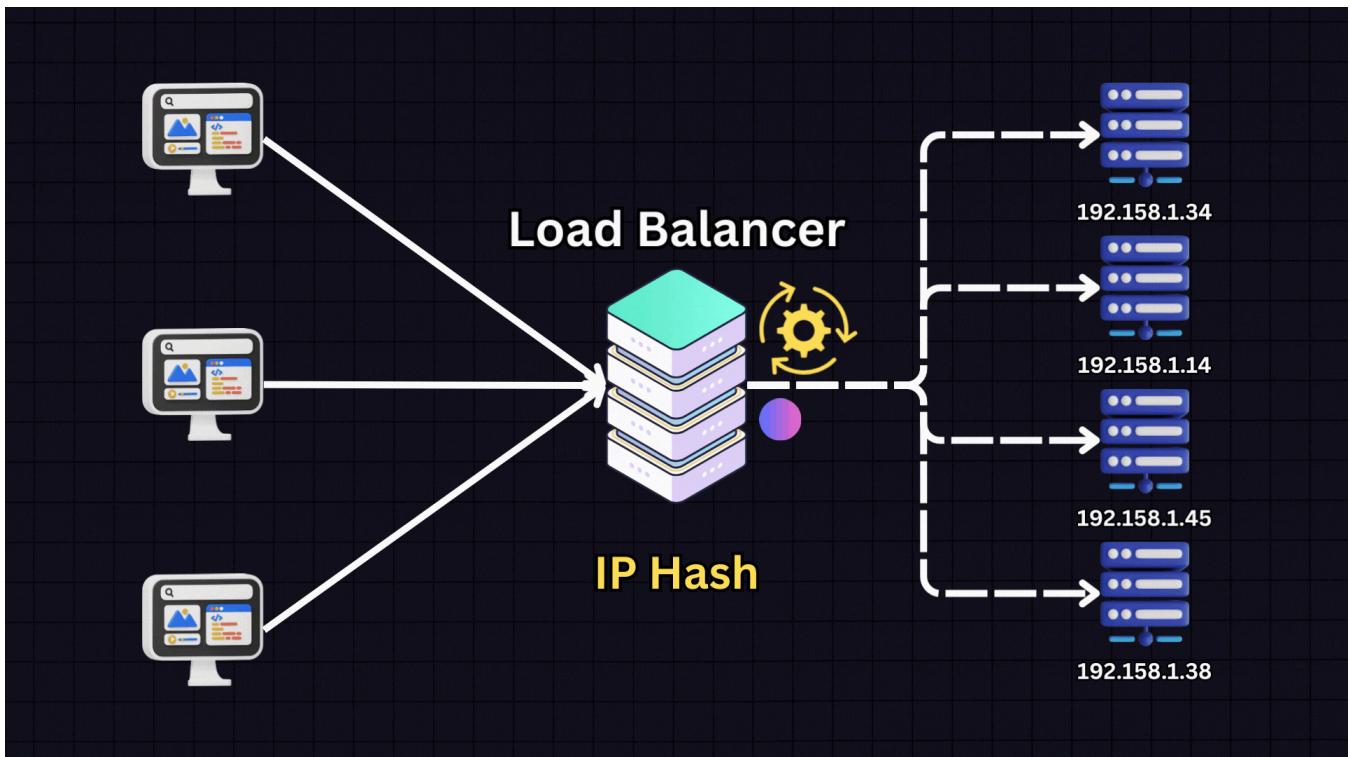
This responsiveness-focused algorithm chooses the server with the lowest response time and with the fewest active connections.



Effective when the goal is to provide the fastest response to requests.

4. IP Hash

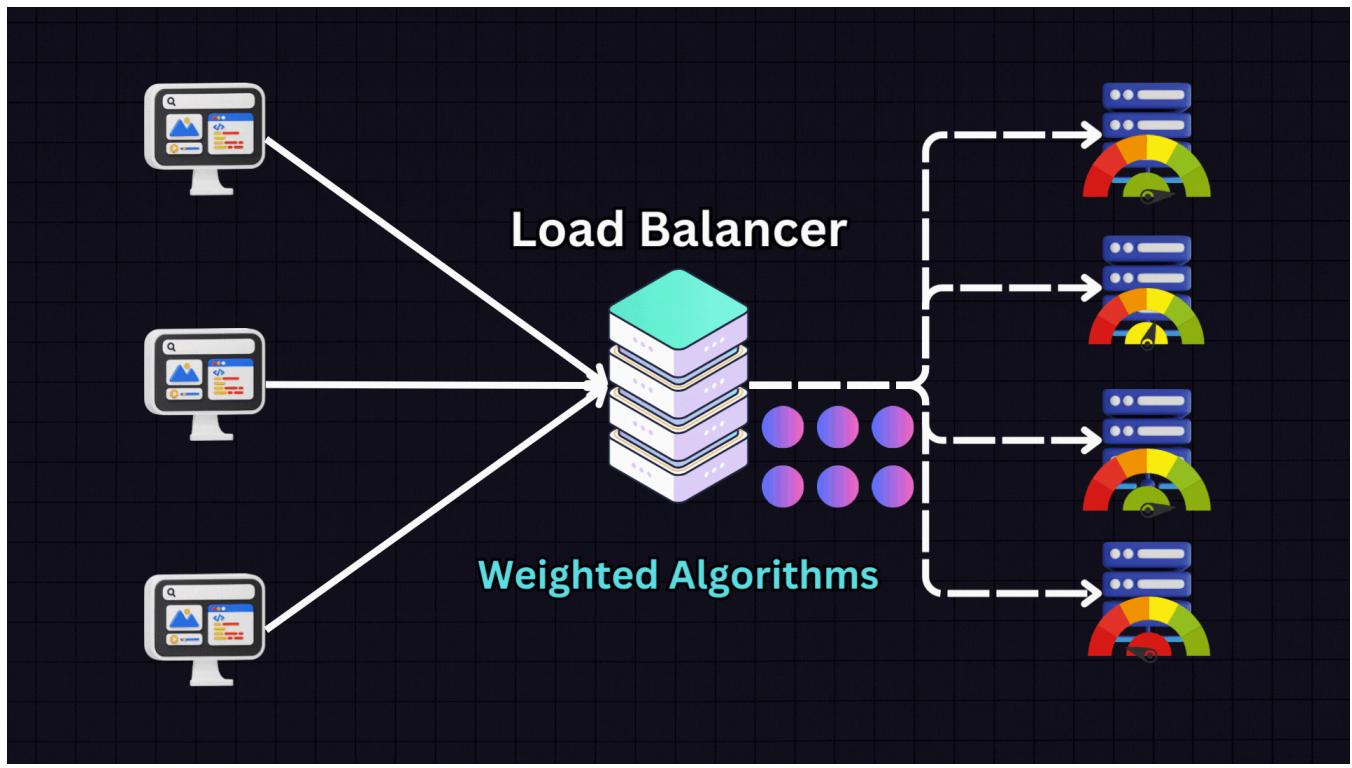
IP Hash determines which server receives the request based on the hash of the client's IP address. This ensures a client consistently connects to the same server.



Useful for session persistence in applications where it's important that a client consistently connects to the same server.

5. Weighted Algorithms

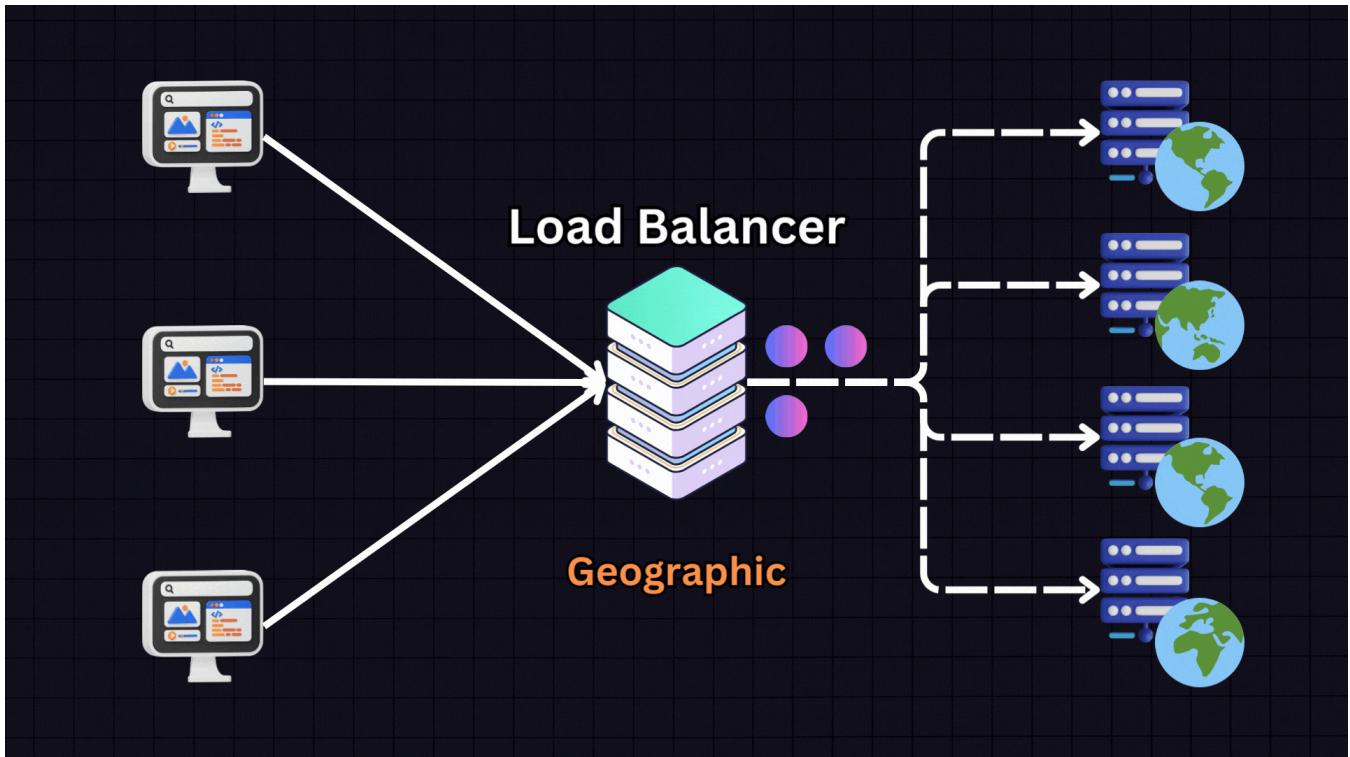
There are also variants of the above methods that can be weighted. For example, in Weighted Round Robin or Weighted Least Connections, servers are assigned weights typically based on their capacity or performance metrics.



Weighted algorithms are effective when servers in the pool have different capabilities (e.g., CPU, RAM).

6. Geographical Algorithms

These are location-based algorithms that direct requests to the server geographically closest to the user or based on specific regional requirements.

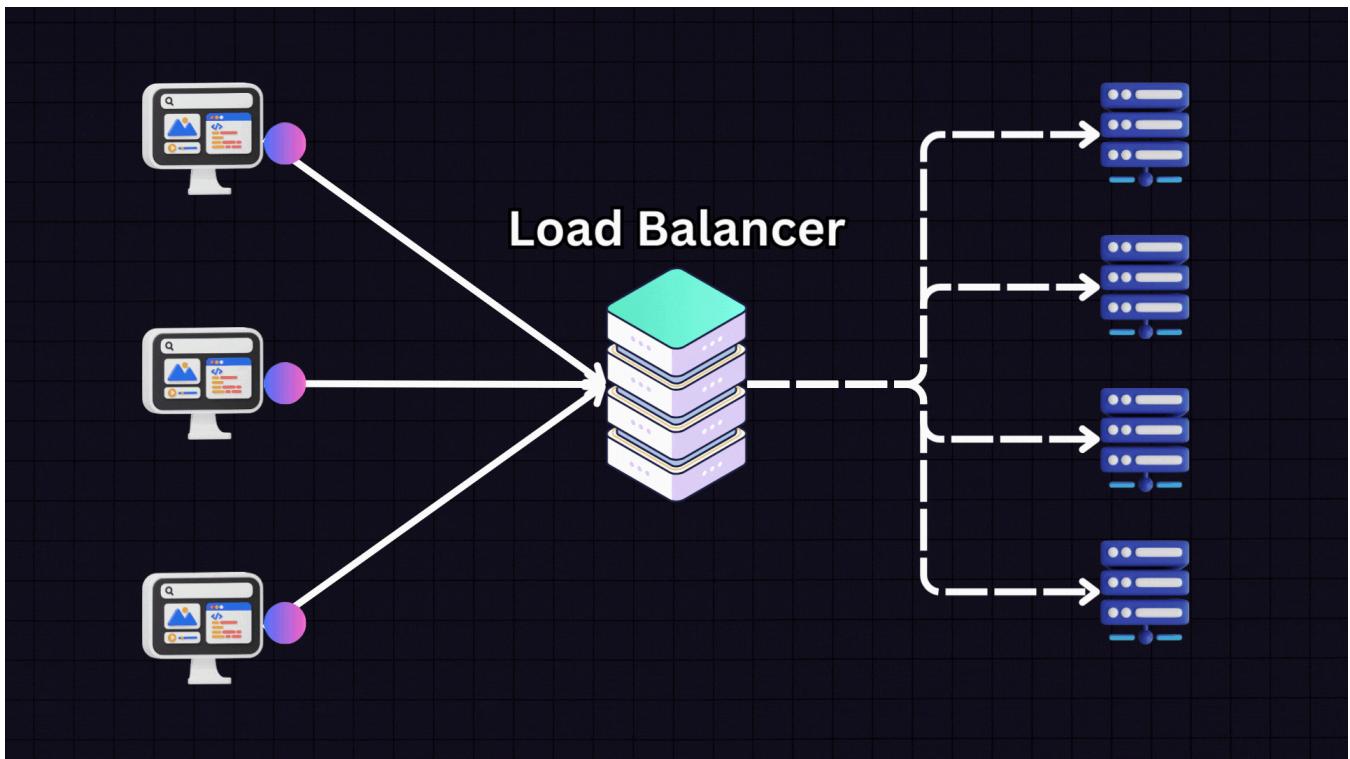


Useful for global services where latency reduction and local regulatory compliance are priorities.

7. Consistent Hashing

Consistent hashing uses a hash function to distribute data across various nodes.

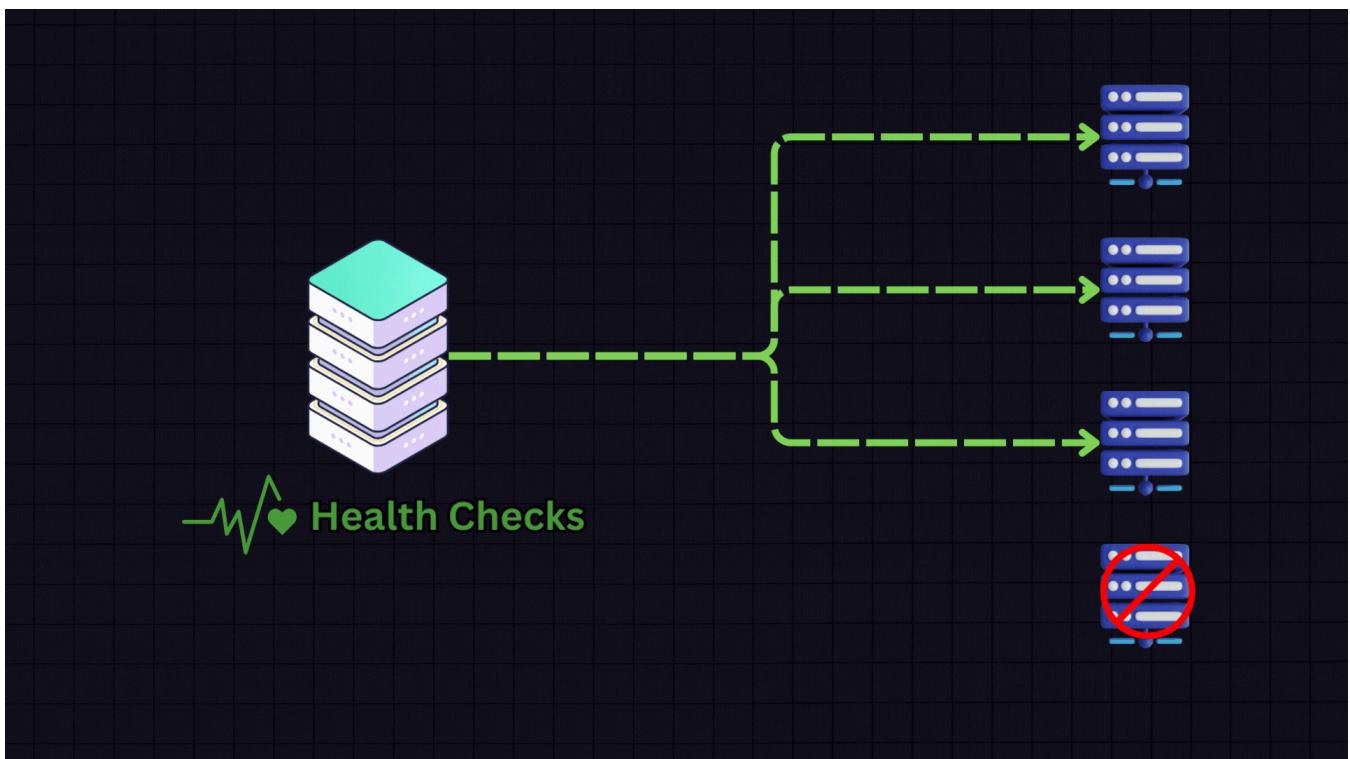
Imagine a hash space that forms a circle where the end wraps around to the beginning, often referred to as a “hash ring”. Both the nodes (servers) and the data (like keys of stored values) are hashed onto this ring.



This also ensures that the same client consistently connects to the same server.

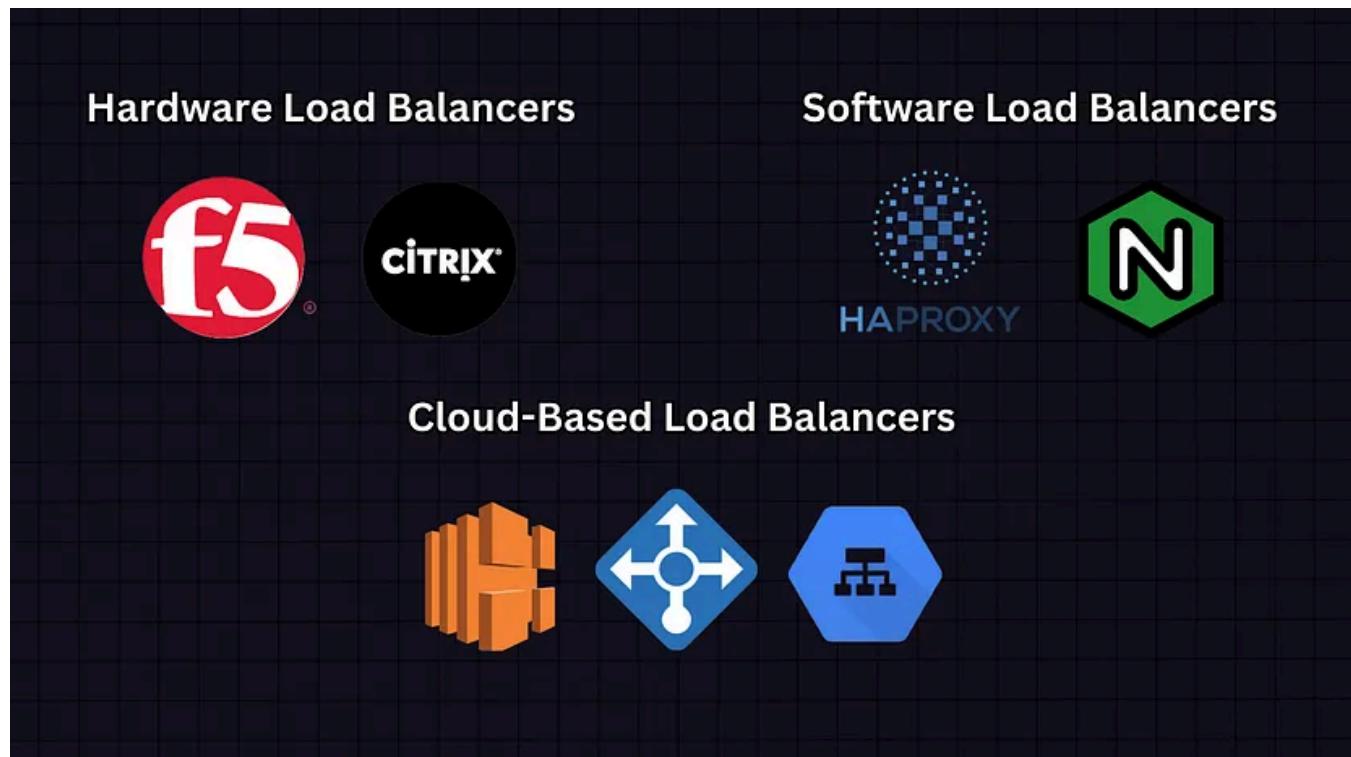
Health Checks

An essential feature of load balancers is continuous health checking of servers to ensure traffic is only directed to servers that are online and responsive. If a server fails, the load balancer will stop sending traffic to it until it is back online.



Load Balancer Examples

Load balancers come in various forms, including hardware appliances, software solutions, and cloud-based services. Here are some examples:



Hardware Load Balancers

1. **F5 BIG-IP:** A widely used hardware load balancer known for its high performance and extensive feature set, offering local traffic management, global server load balancing, and application security.
2. **Citrix ADC:** Formerly known as NetScaler, it provides load balancing, content switching, and application acceleration.

Software Load Balancers

1. **HAProxy:** A popular open-source software load balancer and proxy server for TCP and HTTP-based applications. It's known for its efficiency, reliability, and low memory footprint.
2. **NGINX:** Often used as a web server, NGINX also functions as a load balancer and reverse proxy for HTTP and other network protocols.

Cloud-Based Load Balancers

1. **AWS Elastic Load Balancing (ELB):** Offers several types of load balancers as part of the AWS cloud services, including the Application Load Balancer, Network Load Balancer, and Classic Load Balancer.

2. Microsoft Azure Load Balancer: Provides high availability and network performance to applications running in Azure. It supports inbound and outbound scenarios.

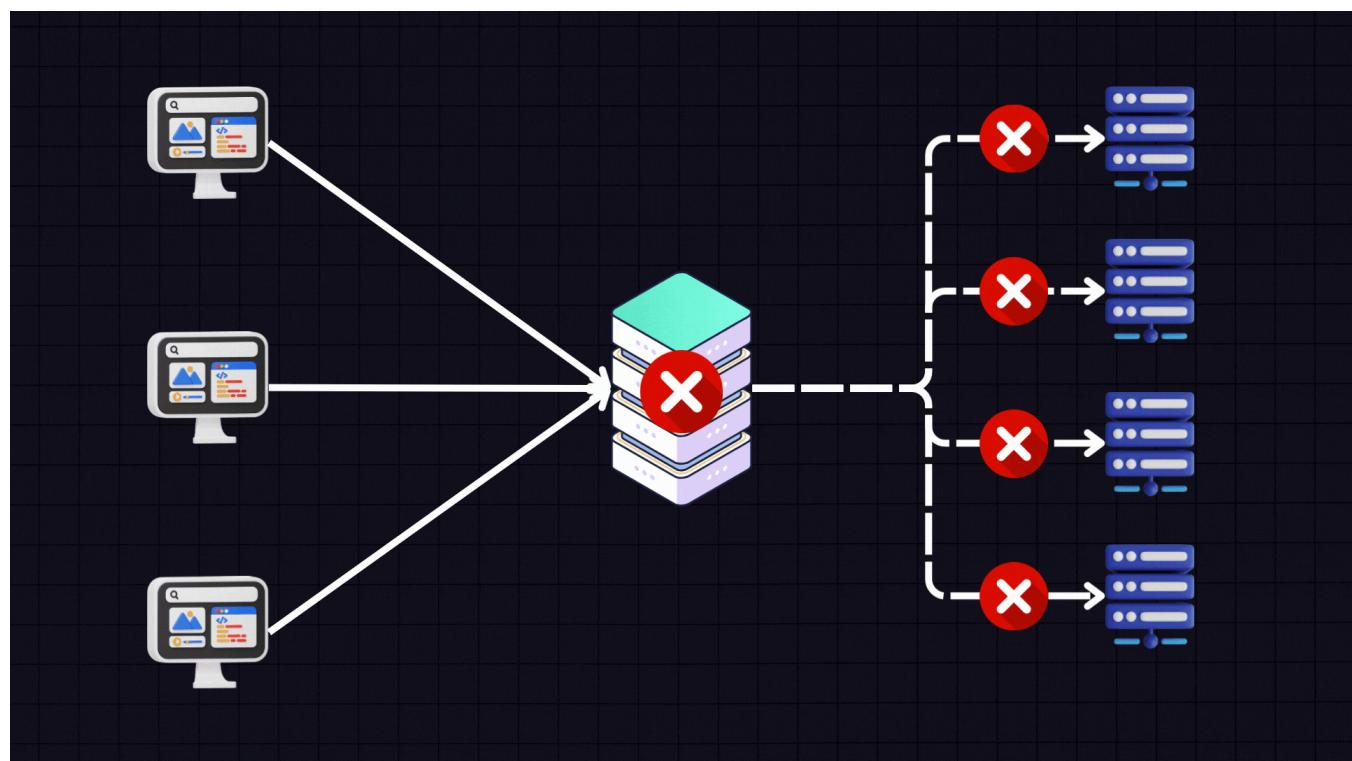
3. Google Cloud Load Balancing: A fully distributed, software-defined, managed service for all your traffic. It offers various types, including HTTP(S), TCP/SSL, and UDP load balancing.

Virtual Load Balancers

1. VMware NSX Advanced Load Balancer (Avi Networks): Offers a software-defined application delivery controller that can be deployed on-premises or in the cloud.

What happens when the load balancer goes down?

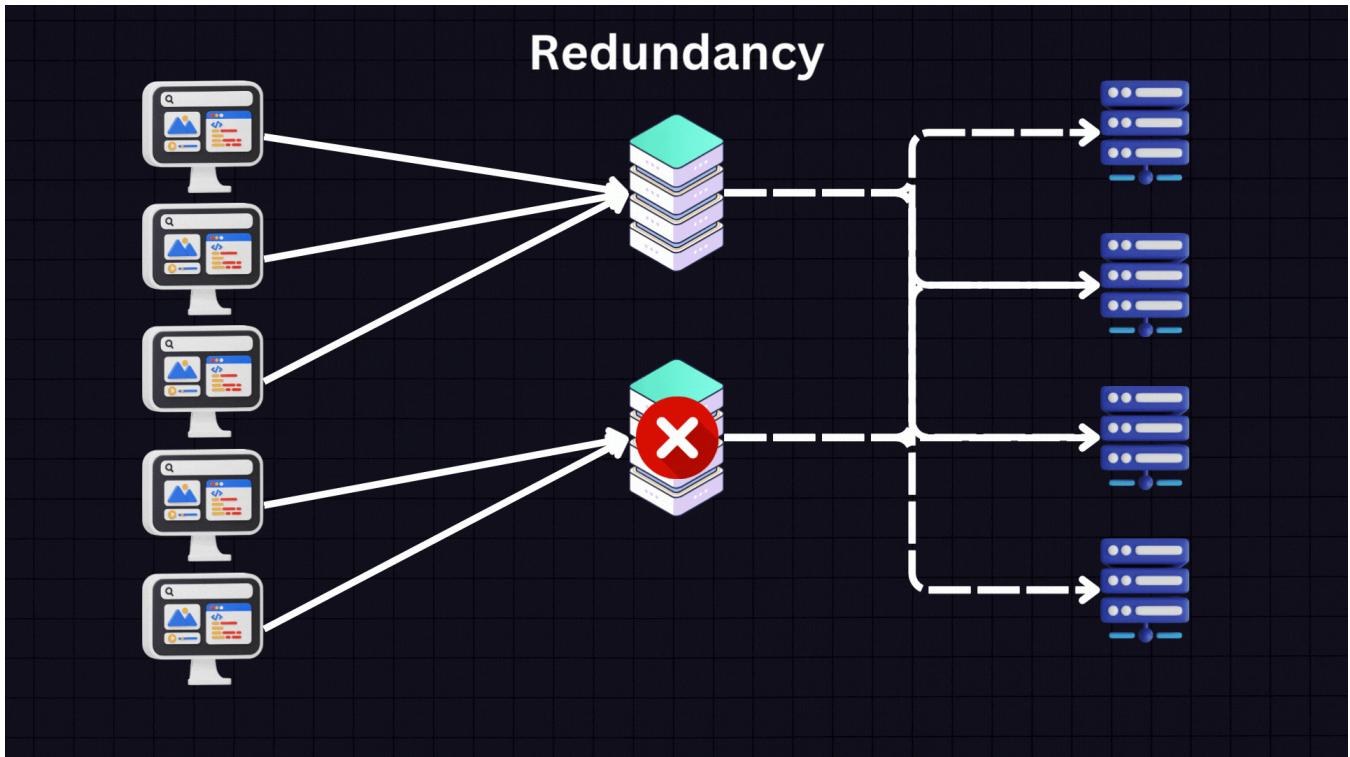
The load balancer is basically a single point of failure, and in case it goes down, all of the servers become unavailable for the clients.



To avoid or minimize the impact of a load balancer failure, several strategies are typically employed:

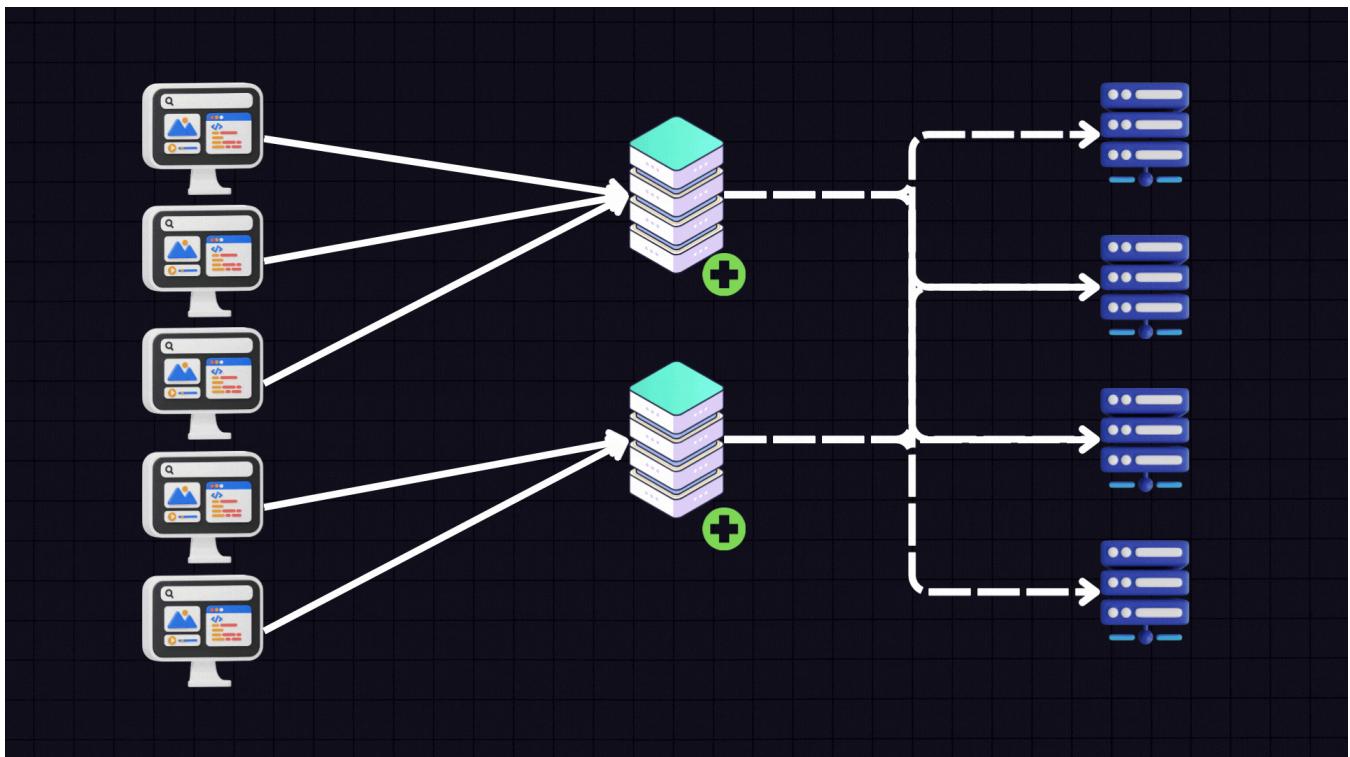
Redundancy

Implementing redundant load balancing by using more than one load balancer, often in pairs, is a common approach. If one fails, the other takes over, a method known as failover.



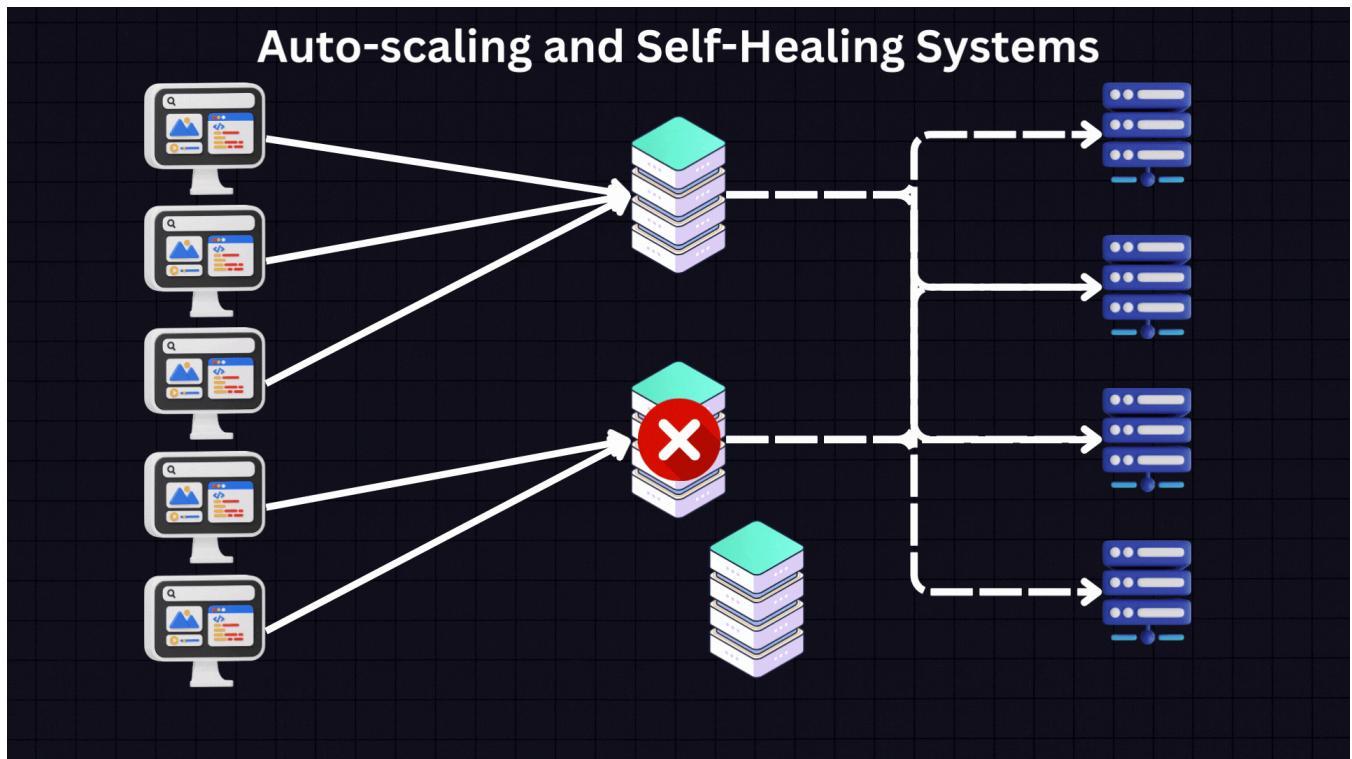
Health Checks and Monitoring

Continuous monitoring and health checks of the load balancer itself can ensure that any issues are detected early and can be addressed before causing significant disruption.



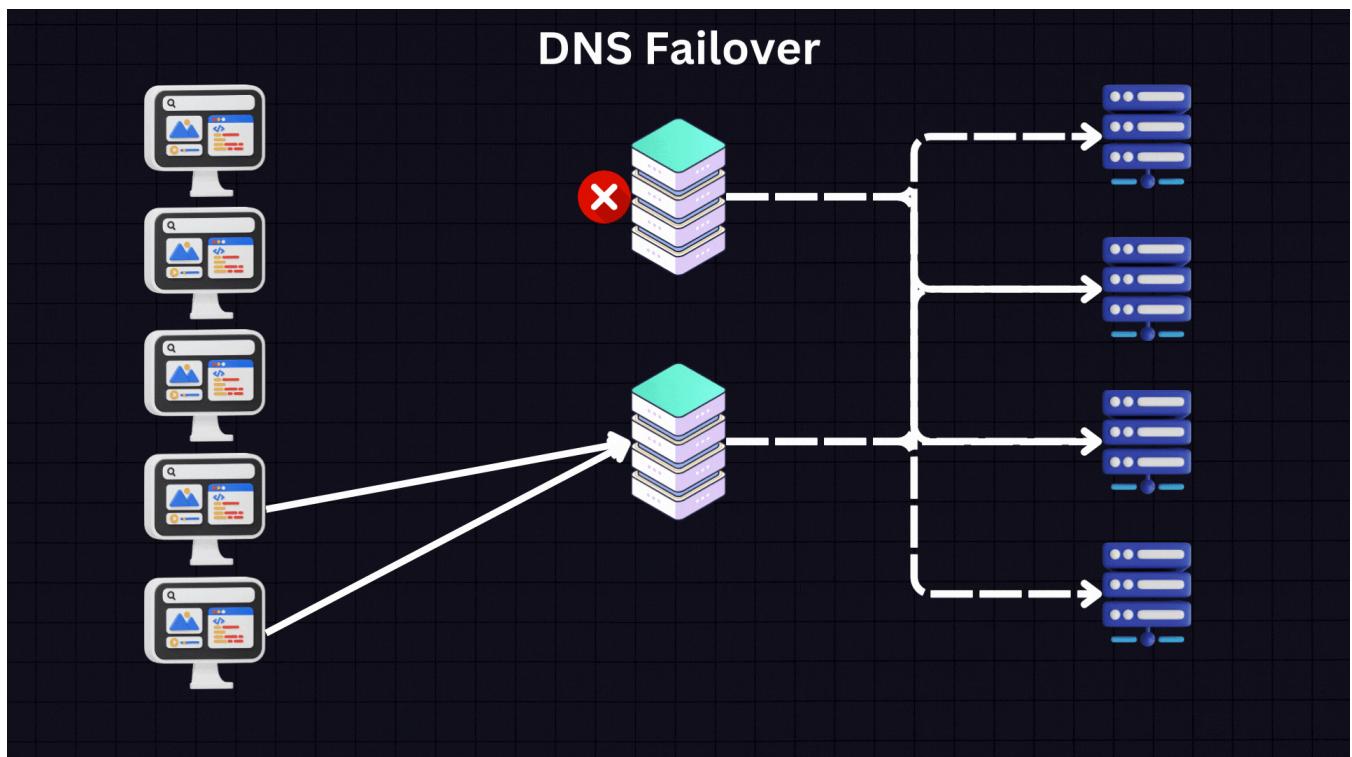
Auto-scaling and Self-Healing Systems

Some modern infrastructures are designed to automatically detect the failure of a load balancer and replace it with a new instance without manual intervention.



DNS Failover

In some configurations, DNS failover can reroute traffic away from an IP address that's no longer accepting connections (like a failed load balancer) to a preconfigured standby IP.



If you're new here, I'm Hayk. I help web developers secure their first tech jobs or advance to senior roles at the [Web Dev Mastery](#) community.

Web Dev Mastery

Master Essential Web Development Skills, Get Hired, and Advance Your Career.

www.skool.com

Thanks for the read! For weekly insights on web development that you won't want to miss, subscribe to [My Newsletter](#).

Hayk Simonyan | Substack

Learn in-demand Web Development skills, get hired, and advance your career.

hayksimonyan.substack.com

Load Balancing

Load Balancer

Load Balancing Algorithm

Nginx



Follow

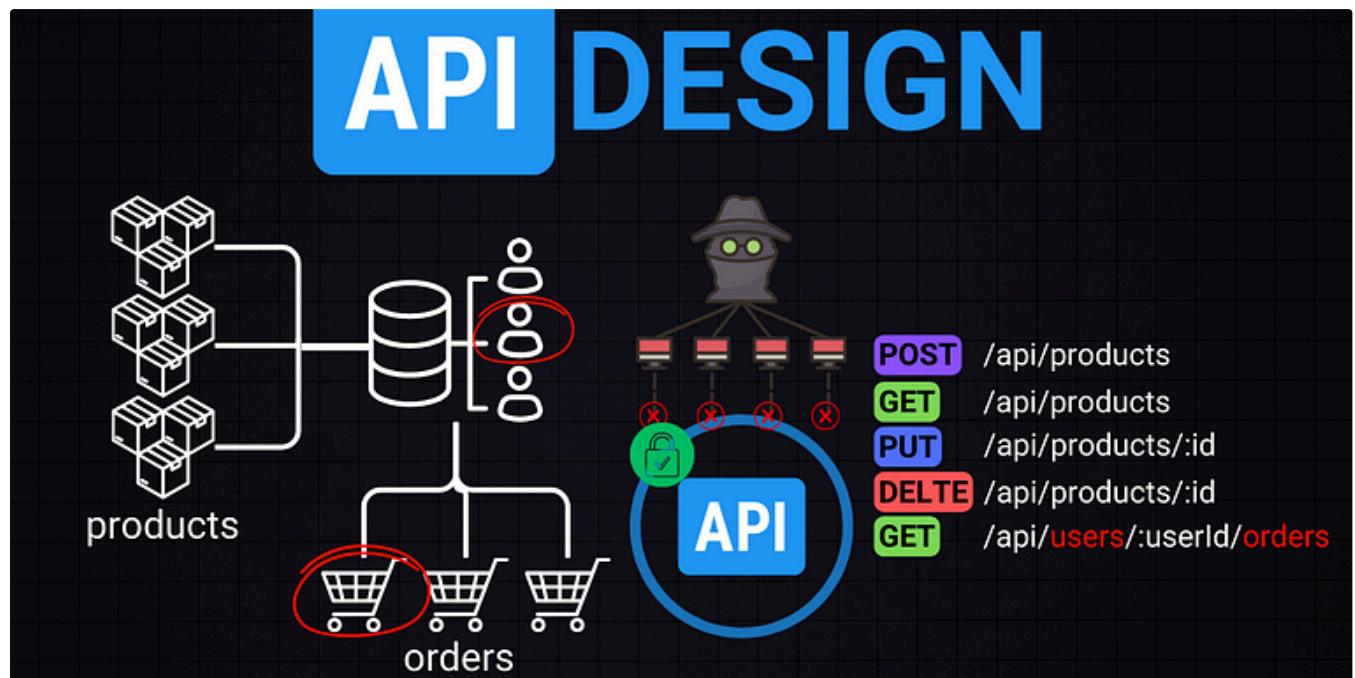


Written by Hayk Simonyan

8.7K Followers · Writer for Level Up Coding

Senior Software Engineer  <https://www.skool.com/web-dev-mastery>

More from Hayk Simonyan and Level Up Coding



Hayk Simonyan in Level Up Coding

API Design 101: From Basics to Best Practices

API design, from the basics to best practices.

5 min read · Dec 28, 2023

2.7K 15



...





Tushar Aggarwal in Level Up Coding

101 Python Automation Scripts: Streamlining Tasks and Boosting Productivity(Part 1)

#1 of 101-Awesome Python Guides with Tushar Aggarwal

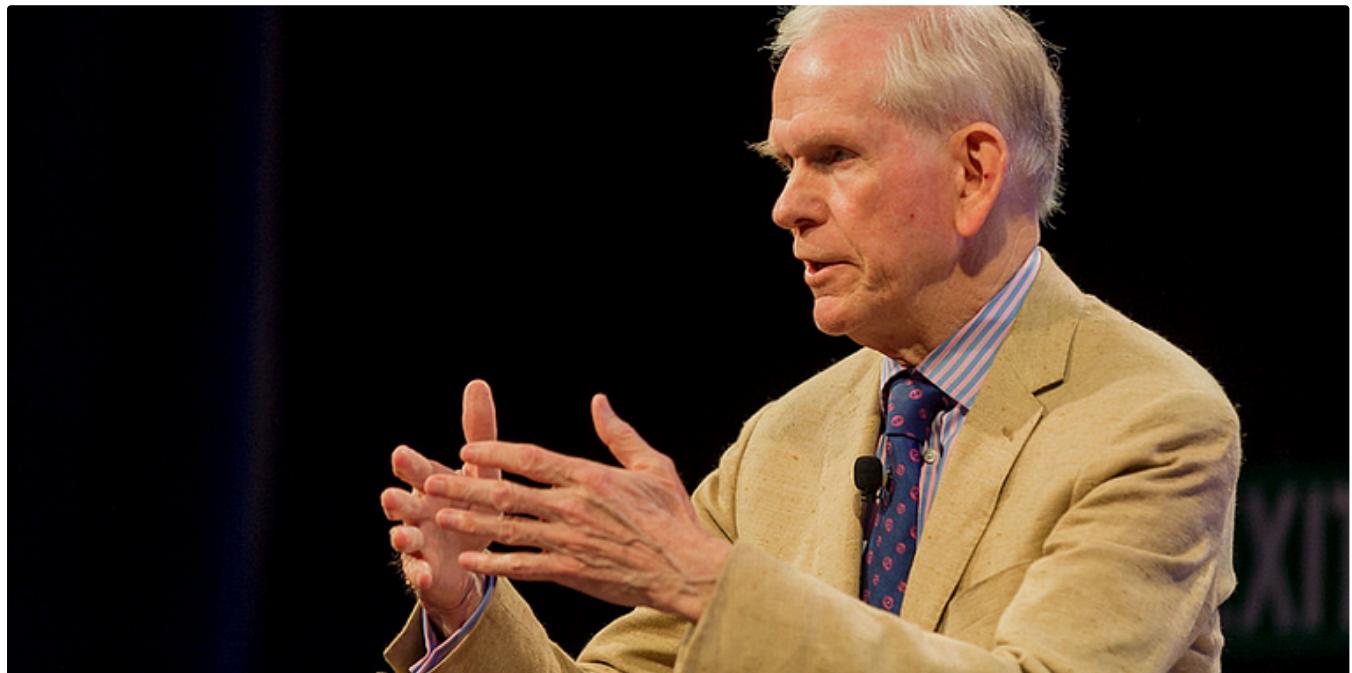
52 min read · Apr 13, 2024

👏 1.8K

💬 14



...



Jayden Levitt in Level Up Coding

Billionaire Investor Warns: What's Ahead Is Worse Than a Recession.

See the world as it is and not how you'd like it to be.

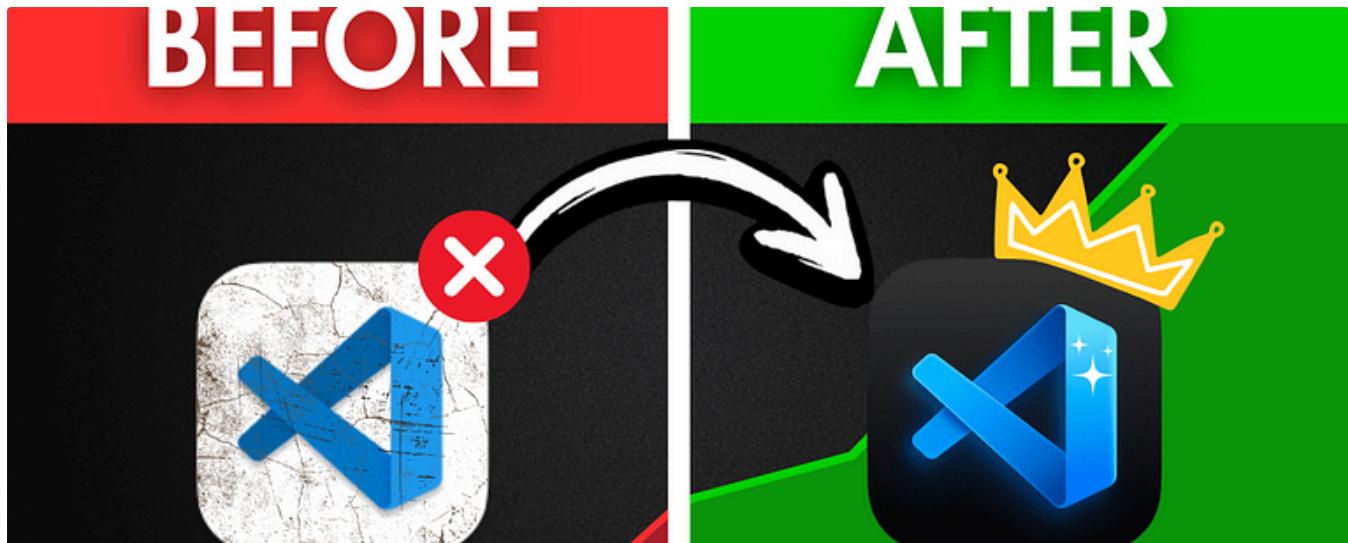
⭐ · 8 min read · May 8, 2024

👏 4.7K

💬 94



...

[Open in app ↗](#)

Search



Hayk Simonyan in Level Up Coding

20 VS Code Extensions to INCREASE Productivity 2024 + Themes, Icons, & Shortcuts

My top 20 vscode extensions for productivity, plus essential tips on themes, icons and shortcuts.

5 min read · May 22, 2024

417

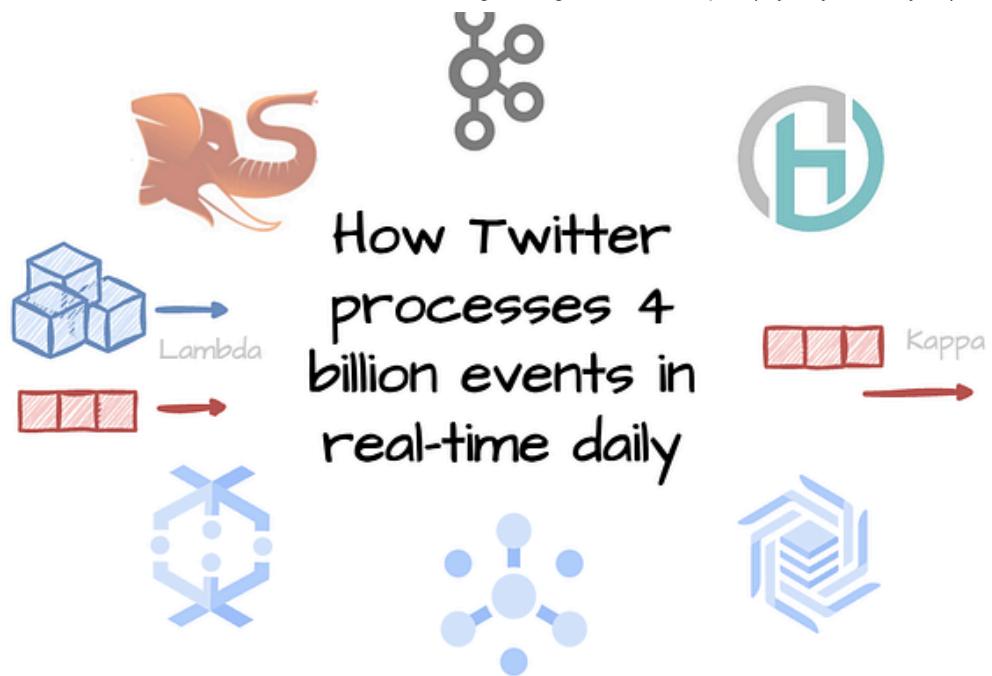
4



...

[See all from Hayk Simonyan](#)[See all from Level Up Coding](#)

Recommended from Medium



Vu Trinh in Data Engineer Things

How Twitter processes 4 billion events in real-time daily

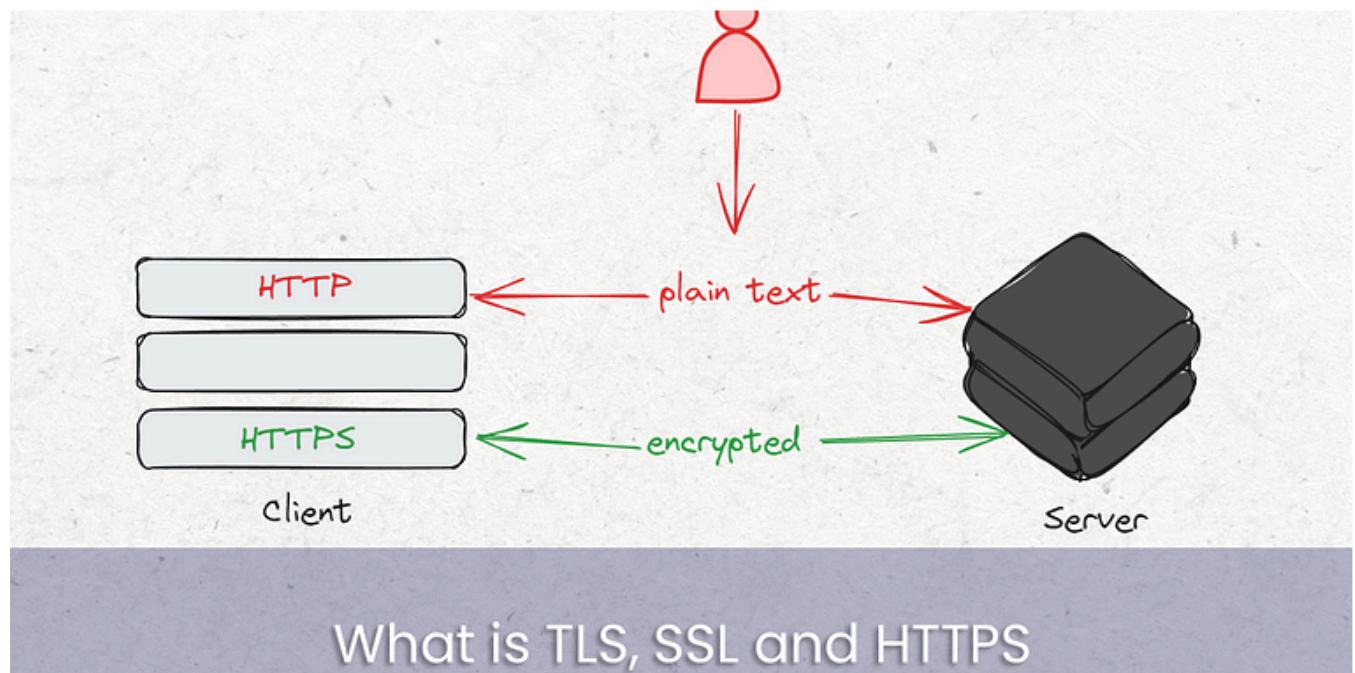
From Lambda to Kappa

6 min read · May 25, 2024

236 1



...



Arton Demaku

What is TLS, SSL and HTTPS

If you go to a website that does not use https, you will now see a warning and your browser will warn you. Why is this? Simply put, without...

◆ · 3 min read · May 11, 2024

507 8



...

Lists



Staff Picks

655 stories · 1016 saves



Stories to Help You Level-Up at Work

19 stories · 638 saves



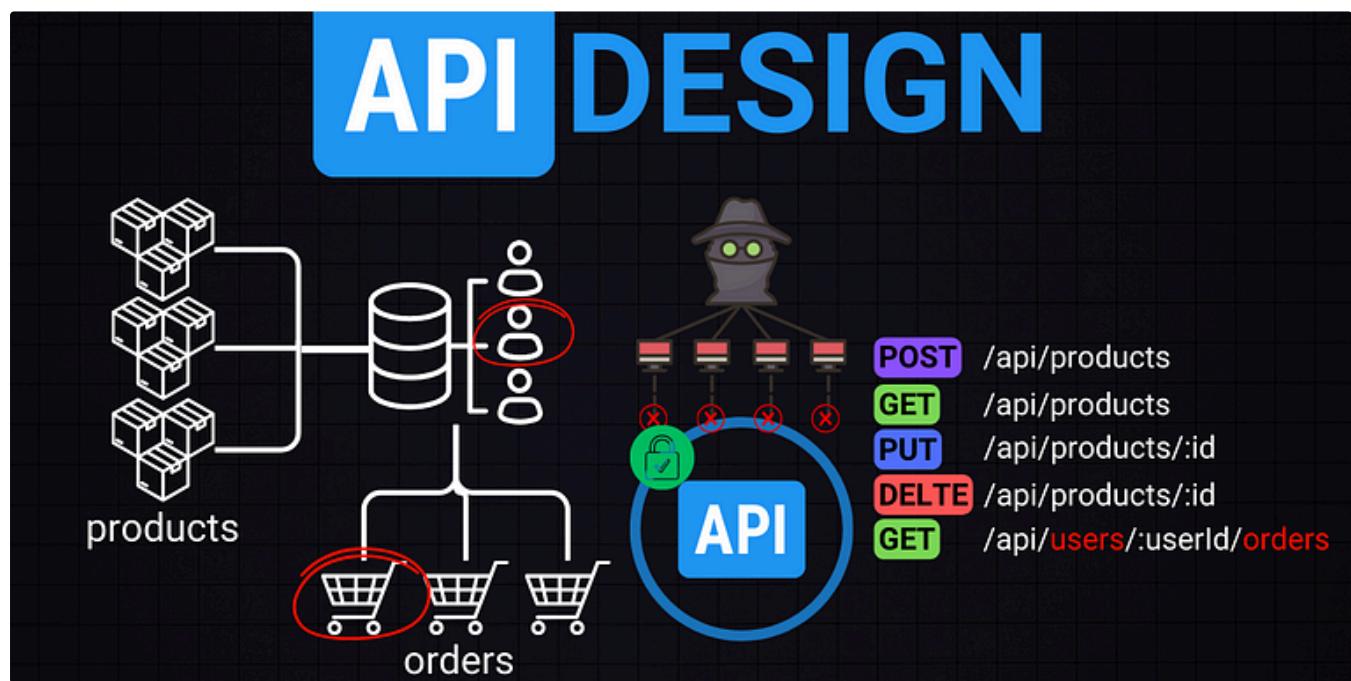
Self-Improvement 101

20 stories · 1971 saves



Productivity 101

20 stories · 1778 saves



Hayk Simonyan in Level Up Coding

API Design 101: From Basics to Best Practices

API design, from the basics to best practices.

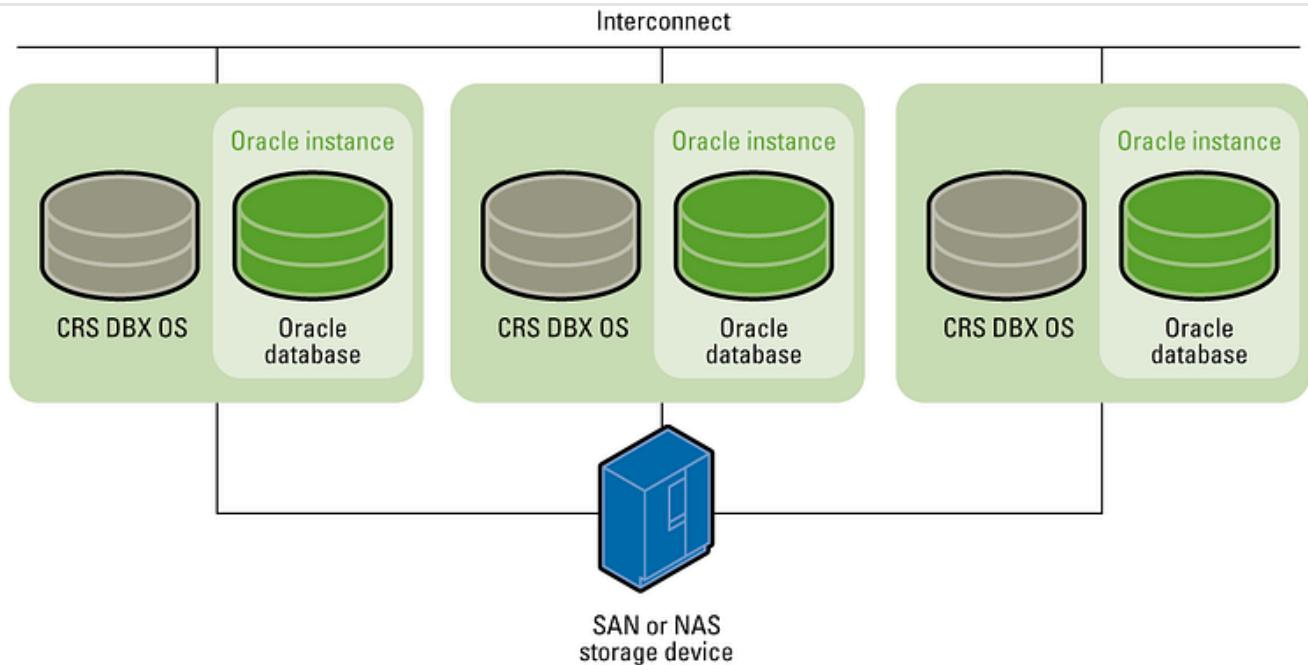
5 min read · Dec 28, 2023

2.7K

15



...



Devansh in CodeX

How Pinterest Scaled to 11 Million Users With Only 6 Engineers

Scaling Pinterest—From 0 to 10s of Billions of Page Views a Month in Two Years

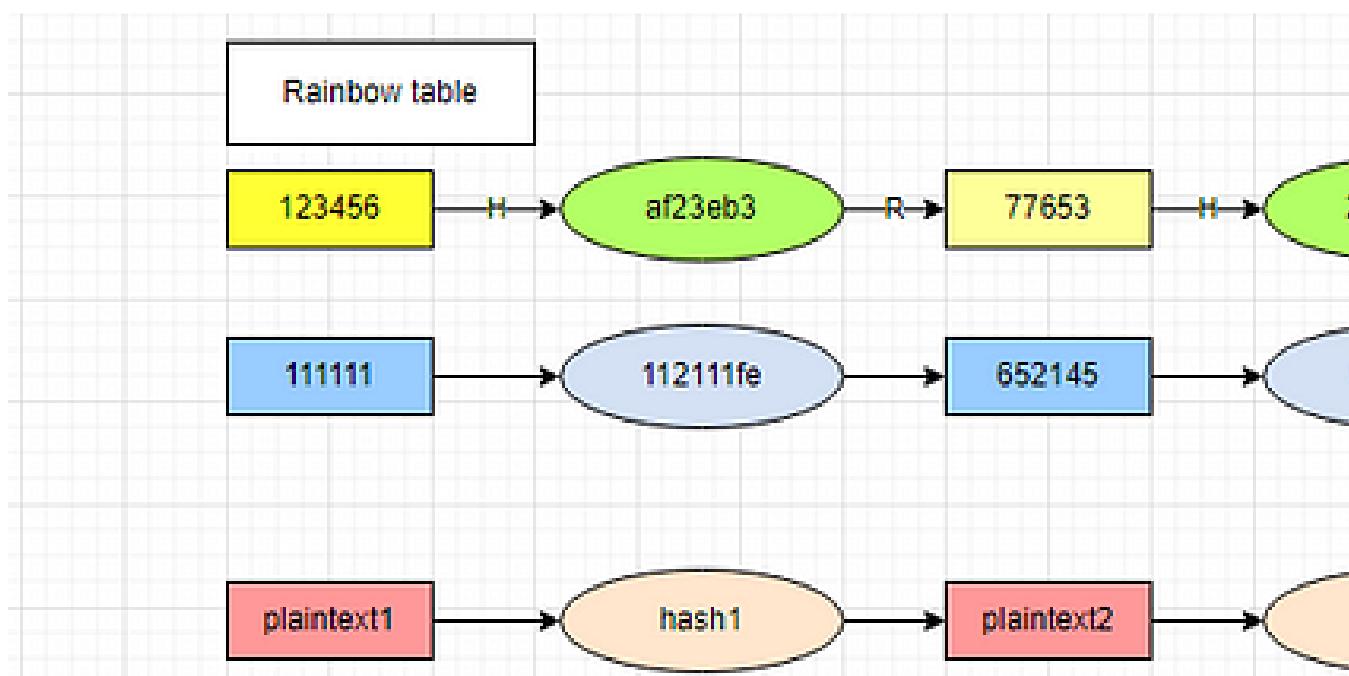
8 min read · May 13, 2024

383

2



...



I LORY

A basic question in security Interview: How do you store passwords in the database?

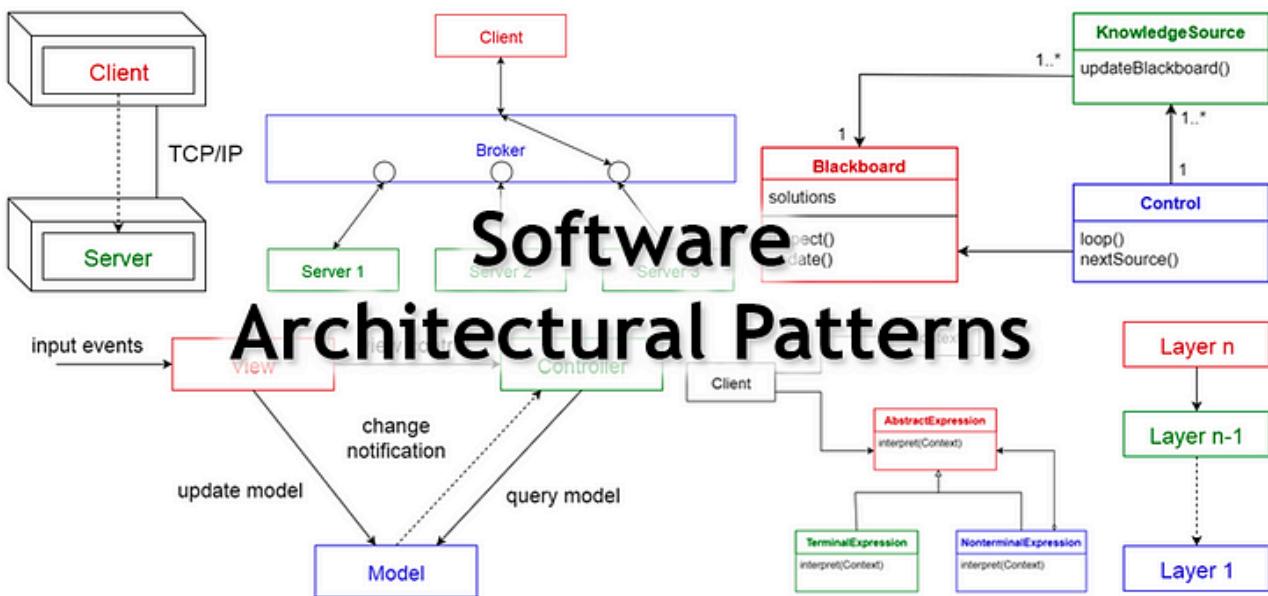
Explained in 3 mins.

◆ · 7 min read · May 12, 2024

👏 2.8K ⚡ 38



...



hubian

12 common software architecture styles, essential for architects

What is software architecture?

14 min read · Jan 8, 2024

👏 53 ⚡ 1



...

See more recommendations