

[Open in app](#)

System Design: Load Balancers



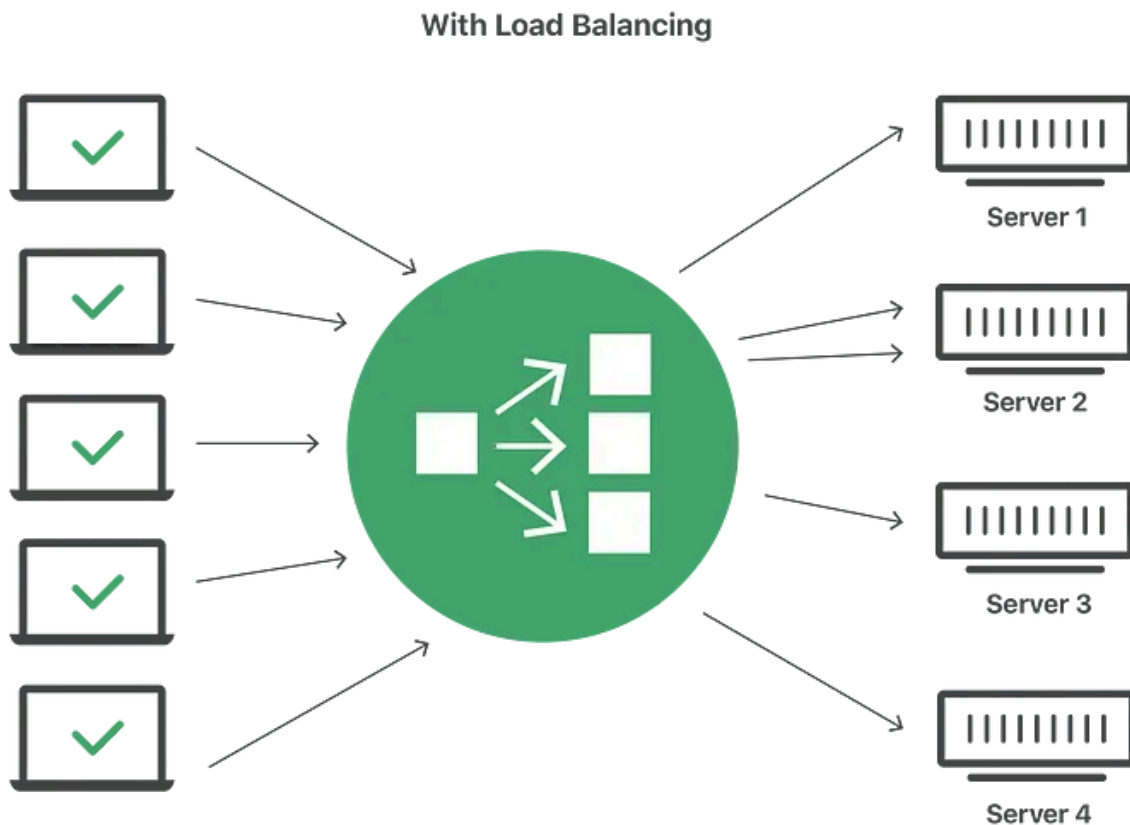
Darshan Trivedi · [Follow](#)

3 min read · 9 hours ago

Listen

Share

More



A load balancer distributes incoming requests across computing resources or app servers.

Advantages of Load Balancers:

- If we use auto-scaling, then it will be the responsibility of the load balancer to attach a new app server or detach and old stopped app server.

- It helps to detect over utilisation and under utilisation of app servers, so we can optimise out resources effectively.

How load balancers registers or de-registers an app server?

There are 2 models to handle this -

1. Pull (Heart beat model)
2. Push

Pull Model

- Also called as heart beat model.
- LB server regularly keeps on sending a request to app servers, registered in a REGISTRY, if the app server responds back then it is alive.
- If not then we assume the app server died.

But what about when we turn up a new app server?

- In this case the new app server registers itself in the REGISTRY, and because LB regularly reads this, it will automatically start sending heartbeat signal to a new server.

Push Model

- In push model, app servers themselves send request to LB denoting they are alive.
- LB doesn't initiate any request to app servers here.

More use cases of load balancer

- We can extract out the encryption/decryption logic to LB instead of app servers.
- -We can make LB support https request and internal app servers can be on HTTP without being exposed to everyone.

How will the load balancer route the request to app servers?

- Round Robin
- Least active connection first (can be clubbed with Round Robin to make it weighted Round Robin)
- Least response time first (can be clubbed with Round Robin to make it weighted Round Robin).

More use cases of load balancer

- We can extract out the encryption/decryption logic to LB instead of app servers.
- We can make LB support https request and internal app servers can be on HTTP without being exposed to everyone.

Types of load balancer

- Layer 7 (application layer in OSI)
- Layer 4 (transport layer in OSI)

Layer 4 Load Balancer

- **Functionality:** Distributes incoming traffic based on data like IP address and port numbers.
- **Protocols:** Typically works with TCP and UDP.
- **Decision Making:** Makes routing decisions based on the source and destination IP addresses and ports. It doesn't inspect the content of the packet.
- **Performance:** Generally faster than Layer 7 load balancer because it doesn't have to inspect packet content.
- **Use Cases:** Suitable for routing traffic where application-level content inspection is not required, such as VPN traffic or some types of VoIP.
- NLB in AWS is layer 4.

Layer 7 Load balancer

- **Functionality:** Distributes incoming traffic based on content type, URL, HTTP header, etc.
- **Protocols:** Works primarily with HTTP/HTTPS but can also handle other application layer protocols.
- **Decision Making:** Can make routing decisions based on the content of the message. For example, it can route a request to a specific server based on the URL path or the type of content requested (e.g., video vs. text).
- **Performance:** Can be slower than Layer 4 load balancers because it inspects packet content, but modern hardware and software optimizations have minimized this difference.
- **Advanced Features:** Supports features like SSL termination, cookie-based session persistence, content caching, and application firewall capabilities.

- **Use Cases:** Suitable for web applications where routing decisions need to be made based on content or when advanced features like SSL termination are required. More useful for basic microservice based architecture.
- ALB or ELB is aws is Layer 7.

Happy learning! 📚

Let's connect on X : [Darshan \(@dbtrivedi10\) / X](#)

System Design Interview

Load Balancer

Computer Science

Backend Engineering

Software Engineering



Follow

Written by Darshan Trivedi

7 Followers

Associate Software Engineer - <http://Hopscotch.in> 🧑💻 | Previous Intern @Hopscotchindia , @allevents_in 🖥️
| <https://twitter.com/dbtrivedi10>

More from Darshan Trivedi

ools

Devices

;

☒ Discover USB devices

Port forwarding.

ons

☒ Discover network targets


Configure...

Open dedicated DevTools for Node

workers

workers


Remote Target #127.0.0.1

 Darshan Trivedi

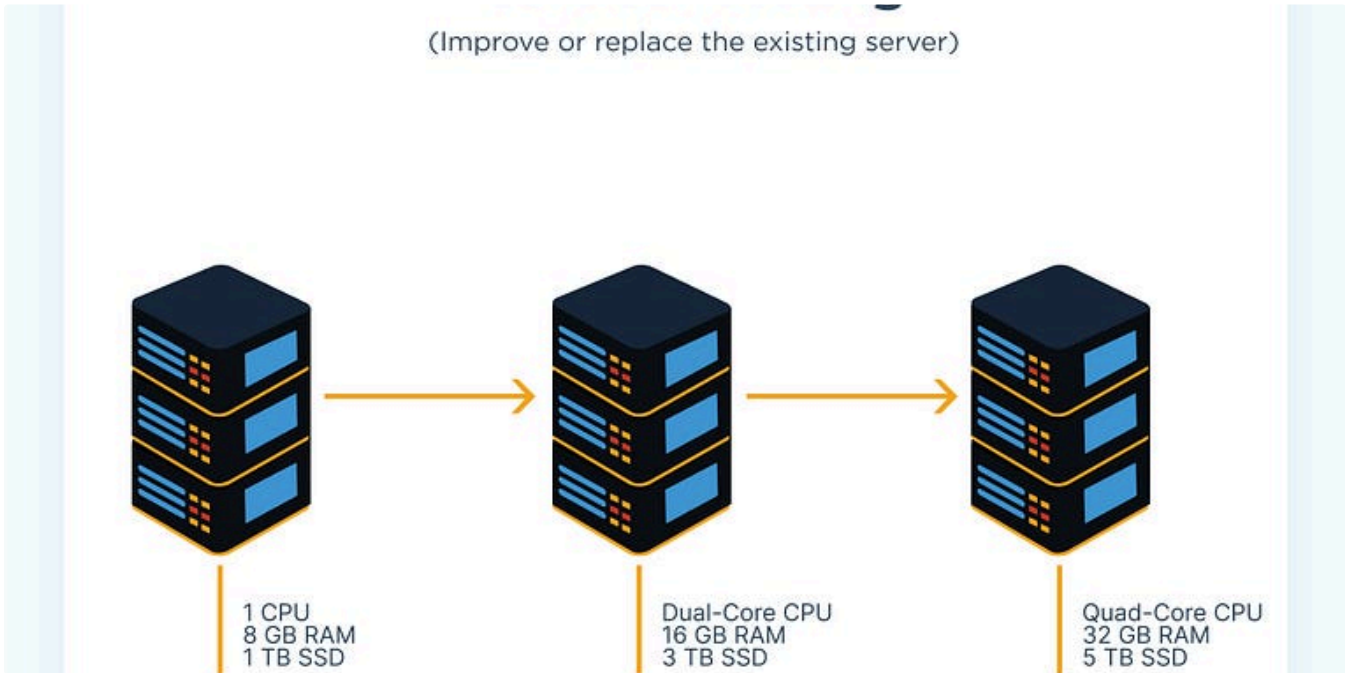
Remote Debugging in Puppeteer to Debug Browser Automation


Puppeteer is a Node.js library that provides a high-level API to control headless Chrome or Chromium over the DevTools Protocol.

2 min read · Feb 15, 2023

 10 

 ...



 Darshan Trivedi

Vertical Scaling—Scale UP

In earlier days, people used to upgrade to higher machine config.

2 min read · May 25, 2024



Darshan Trivedi

Understanding the Layers of MVC Architecture: Key to Building Scalable and Maintainable Web...

MVC, or Model-View-Controller, is a widely used design pattern for building web applications. It provides a clear separation of concerns...

2 min read · Feb 11, 2023





Darshan Trivedi

Authorization vs Authentication

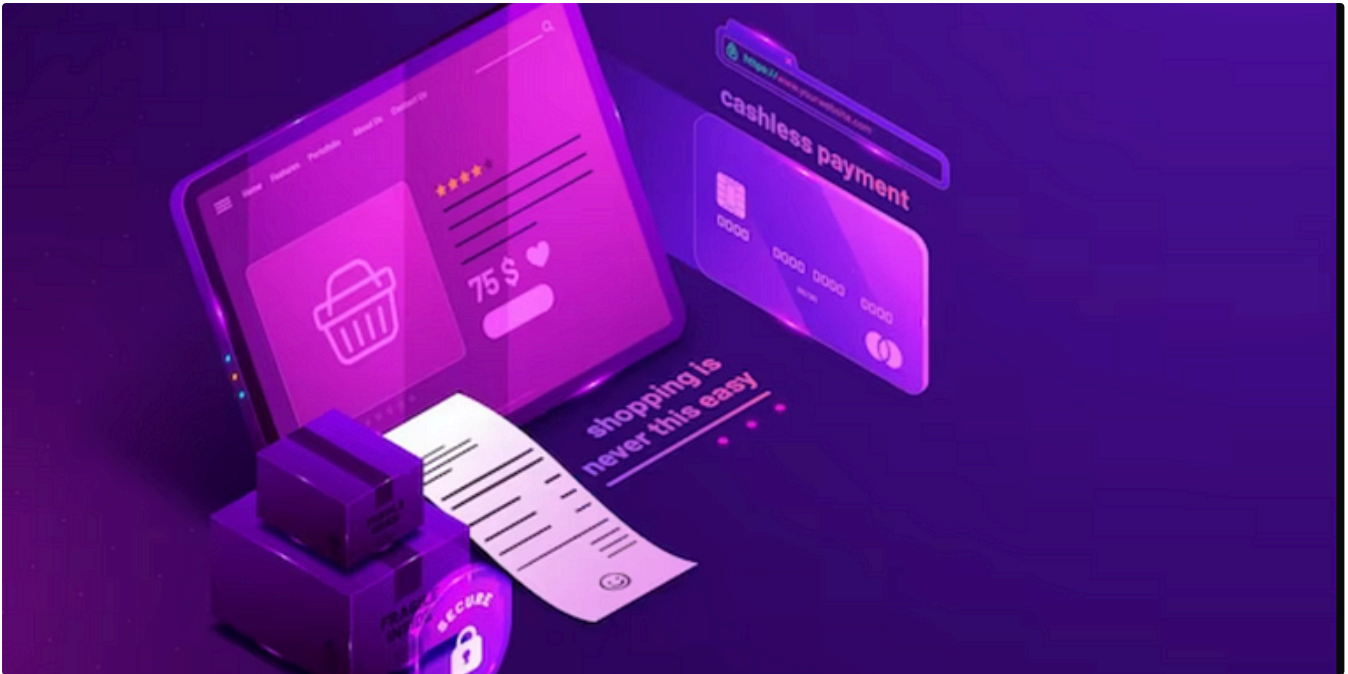
The Importance of Understanding Authentication and Authorization in Computer Security

3 min read · Feb 14, 2023



See all from Darshan Trivedi

Recommended from Medium



 sourav mansingh

Designing a Payment System: Part 2

In an attempt to design a payment system, we would moving forward to understand different components that part of a payment system...

6 min read · Mar 20, 2024



56



1



 Tahir Rauf

System Design—Web Crawler

Functional Requirements

3 min read · Apr 23, 2024



2



Lists



General Coding Knowledge

20 stories · 1270 saves



Stories to Help You Grow as a Software Developer

19 stories · 1100 saves



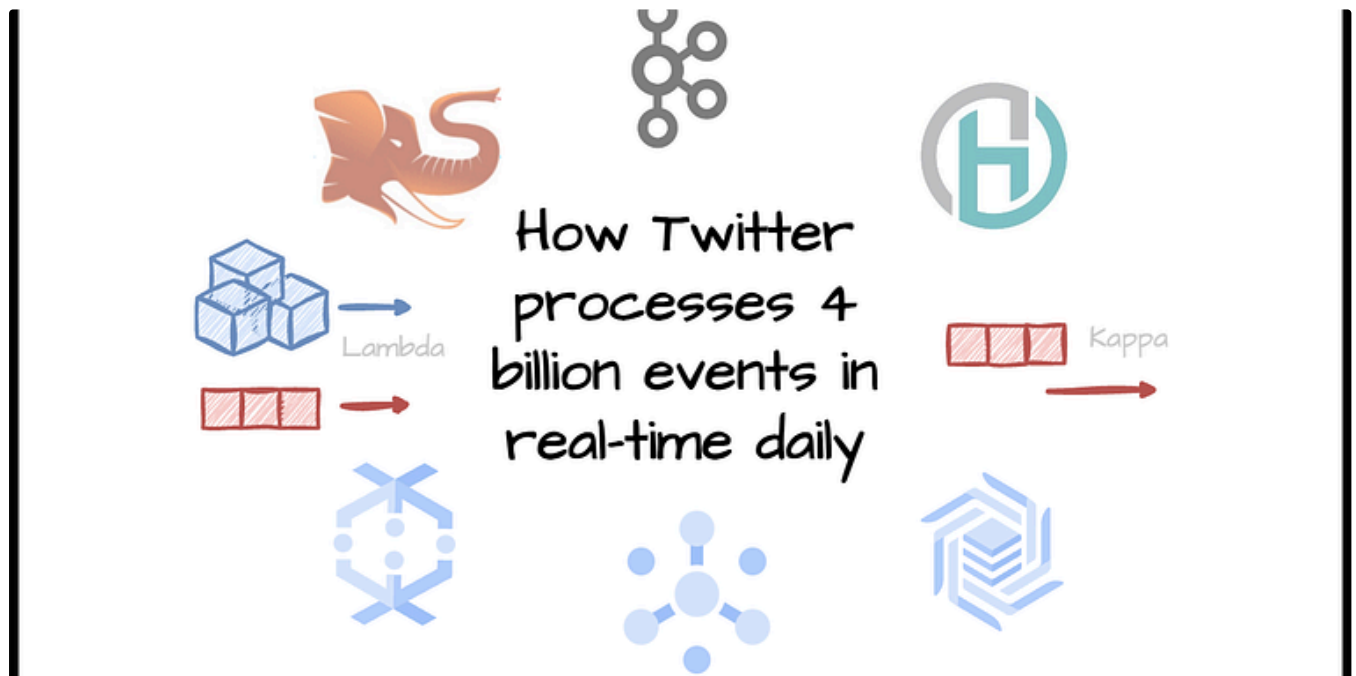
Leadership

50 stories · 342 saves



Good Product Thinking

11 stories · 586 saves

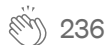


Vu Trinh in Data Engineer Things

How Twitter processes 4 billion events in real-time daily

From Lambda to Kappa

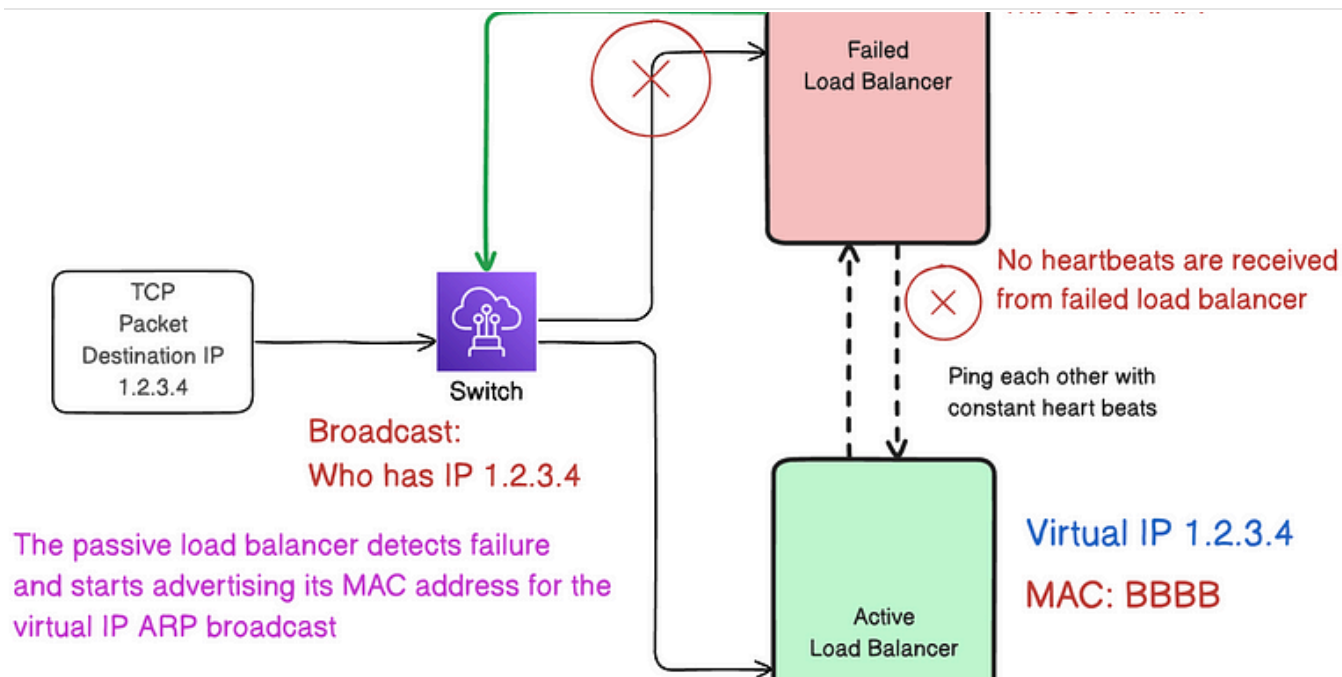
6 min read · May 25, 2024



236



1



Dayem Siddiqui

How to prevent load balancers from being a single point of failure

Lets dig down into the networking layer to understand how can we prevent load balancers from being a single point of failure in our systems

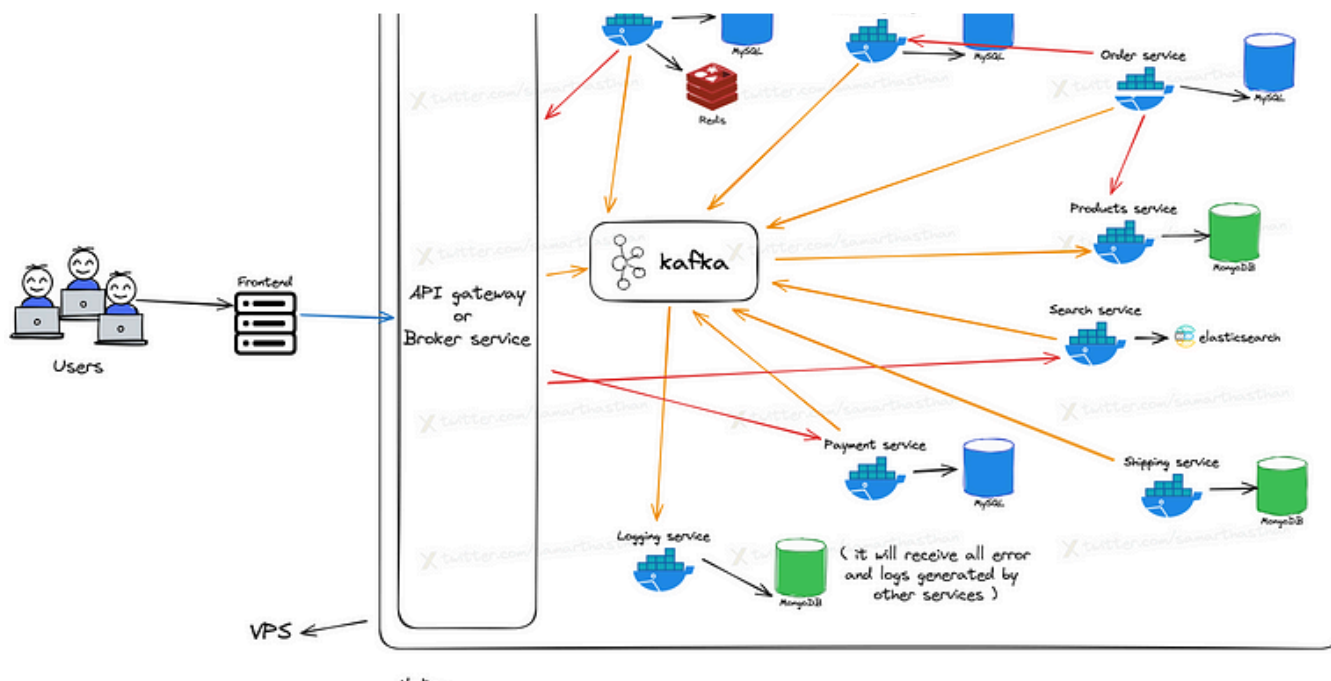
5 min read · Apr 28, 2024



78



1





Samarth Asthan

Building a Scalable E-commerce Empire: A Micro-services System Design Approach

Hey everyone! I'm Samarth Asthan, a 3rd-year computer science student, and I'm new to the world of system design. But, I'm excited to share...

3 min read · Dec 11, 2023



8



Anoop Nayak

Mastering System Design: Caching (Part 5)

Caching is the process of storing frequently accessed data in a temporary storage location, called a cache, in order to quickly retrieve it...

6 min read · Apr 2, 2024



14



See more recommendations