Aditya Sharma
@adyOnWeb

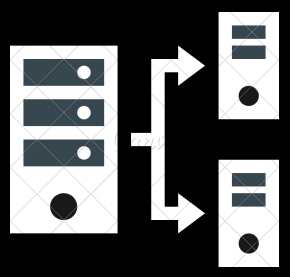
Be Interview Ready Series

How to scale your API to handle millions of calls?

Connect to learn a lot more around front-end and backend interviews
- Ady

Aditya Sharma
@adyOnWeb

Be Interview Ready Series

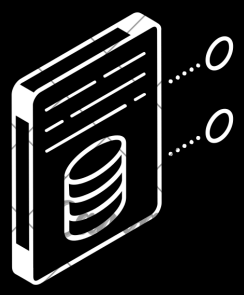


Load balancing

Distributing the incoming requests among multiple servers can help reduce the load on a single server and improve the overall performance of the API.

Aditya Sharma
@adyOnWeb

Be Interview Ready Series



Caching

Caching commonly used data in memory can significantly reduce the load on the API server, as well as reduce response time for clients.

Aditya Sharma

@adyOnWeb

Be Interview Ready Series

Content Delivery Network (CDN):

A CDN can help distribute the load of serving large files such as images and videos, freeing up the API server to focus on processing API requests.


Follow

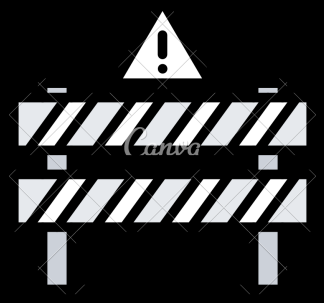

Like


Comment


Share

Aditya Sharma
@adyOnWeb

Be Interview Ready Series



API rate limiting

Limiting the number of requests that a client can make within a specified time frame can help prevent the API from being overwhelmed by too many requests.

Follow Like Comment Share

Sorry for interuption!

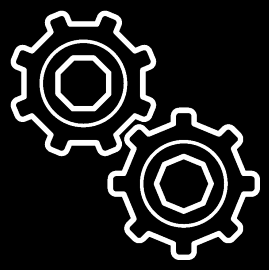
I'm a **newbie** creator!

If you like the content, pls help with a **repost** or shot out :)

Please continue!

Aditya Sharma
@adyOnWeb

Be Interview Ready Series



Asynchronous processing

Instead of blocking the API server while it processes a request, asynchronous processing can help process requests in the background, freeing up the API server to handle new requests.

 Follow

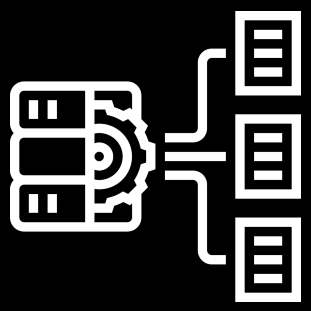
 Like

 Comment

 Share

Aditya Sharma
@adyOnWeb

Be Interview Ready Series



Database sharding

Distributing the database load across multiple servers can improve the performance and scalability of the API.

Aditya Sharma
@adyOnWeb

Be Interview Ready Series

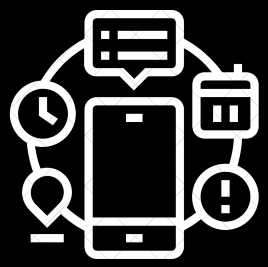
Monitoring & log analysis

Regular monitoring and analysis of API performance and logs can help identify bottlenecks and issues, enabling proactive measures to be taken to improve the API's performance and scalability.

 Follow  Like  Comment  Share

Aditya Sharma
@adyOnWeb

Be Interview Ready Series



Automated scaling

Automatically scaling the number of API servers based on demand can help ensure that the API is always able to handle the volume of requests it receives, without overloading the system.

Aditya Sharma
@adyOnWeb

Be Interview Ready Series

Keep in mind that the specific approach that works best for you will depend on the specific requirements and constraints of your API.

A combination of multiple approaches may be necessary to achieve the desired level of performance and scalability.

Follow Like Comment Share

Aditya Sharma
@adyOnWeb

Was it helpful?

Please take a moment to join me and say hello in DM :)

Like Comment Save