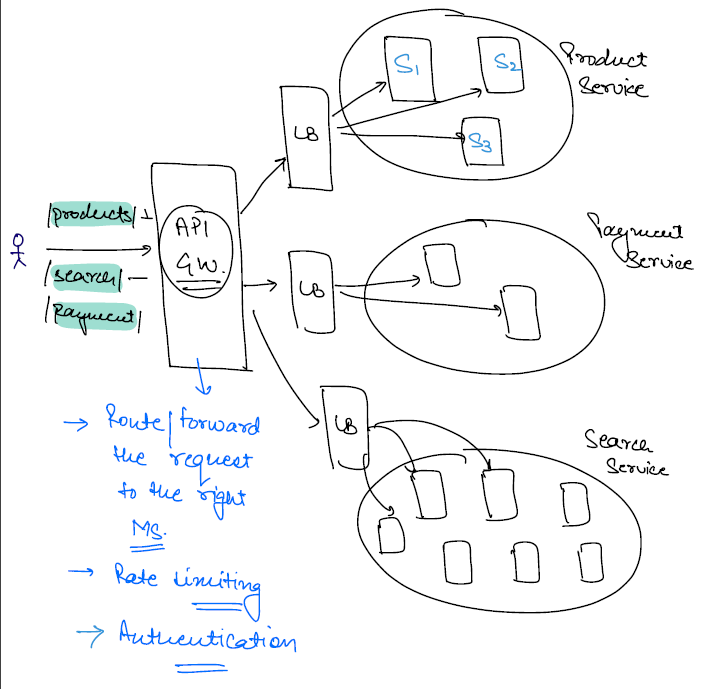
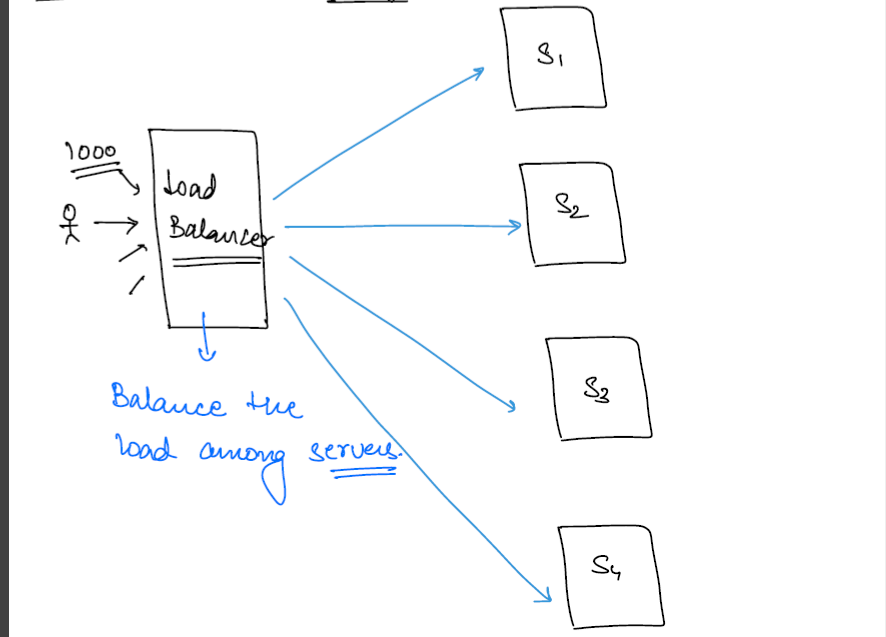
API Gateway

* It is used to forward the request to the right services
* It is also used for rate limiting (e.g if lets say, thousands of request is coming within the seconds from the same ip address, then it should block that ip address… somebody is trying to hack it ☹)
* Used for authentication (lets say for amazon.in, if the request is coming to search the product, then request should be directly transfer to the particular service. But lets say if the request is coming for payment then it will check the authentication first. Will first call the user services to check the user is logged-in or not? If yes, then it will further forward that request to the payment service.



Load Balancer

* The task of Load balancer is to balance the load among multiple servers.
* e.g lets say the load balancer getting 1000 requests, and lets say we have 4 servers, then ideally every server should get 250 requests(i.e 25% of request to each server). No server should say that why I am getting more request 😉
* Consistent hashing algorithm is used behind it.



In practical scenarios, you will have same machine on which API Gateway and Load Balancer is running.

