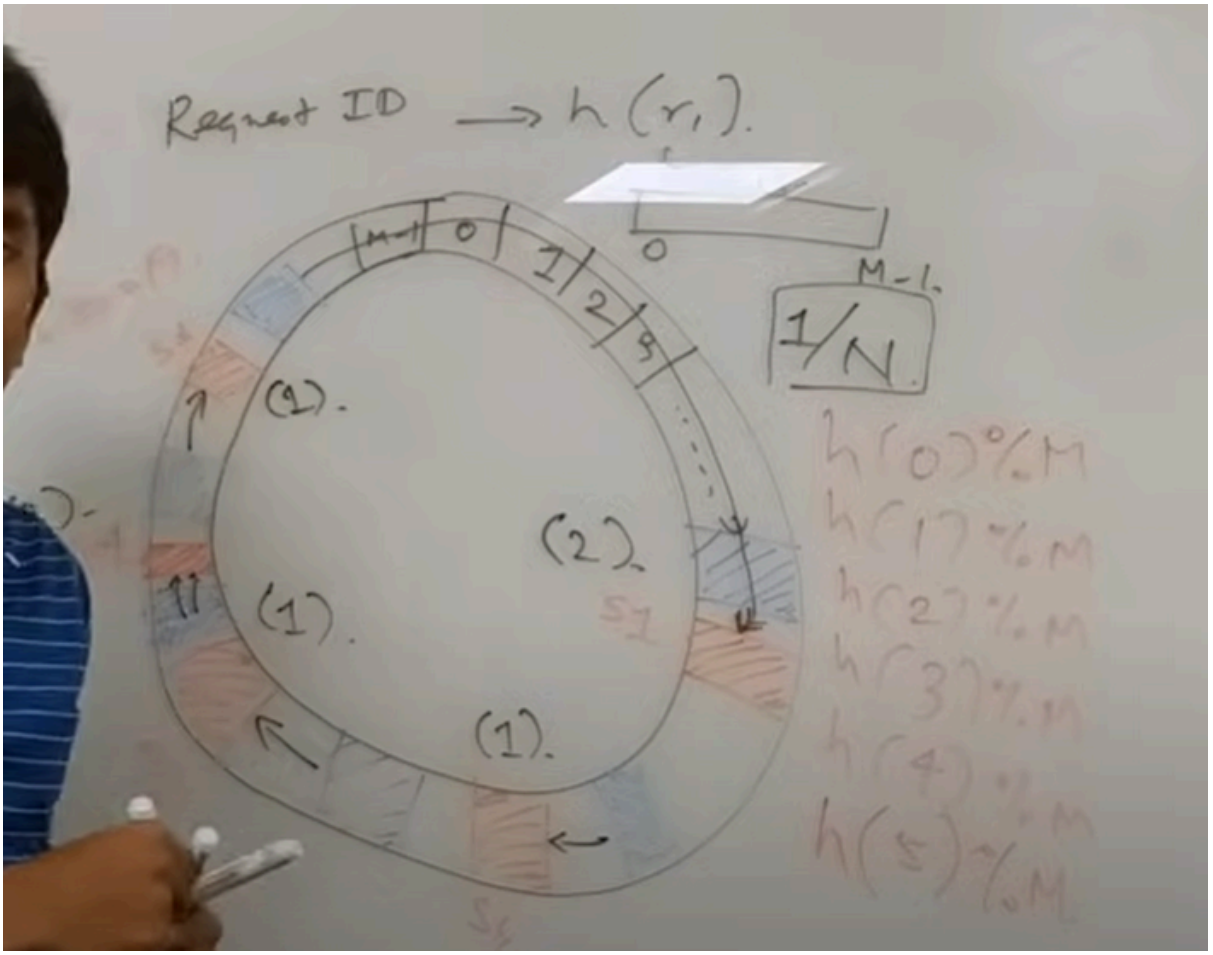
Load Balancing
If there is more than one server each server will carry the load while generating a response

Which server should be picked is to be decided
Therefore we need a Load Balancer

Now to decide where to send we use the Hash function to calculate which server should be choose
But Here is a major issue
If one more server is added then the Hash function works the same but to assign we mod it by the number of servers which changes therefore now previously assigned server numbers are different for some requests and if a request is more frequent it is obv cached

So We use consistent Hashing



In this architecture, we hash server ID as well

Now This would solve adding a server issue
But removing the server - is the issue (skewness)
The situation in which a single server might get a very much load

To solve this we use K hash functions for server id

If k is appropriate then skewness can be almost totally removed (k such as log(m) where m is no. of server),