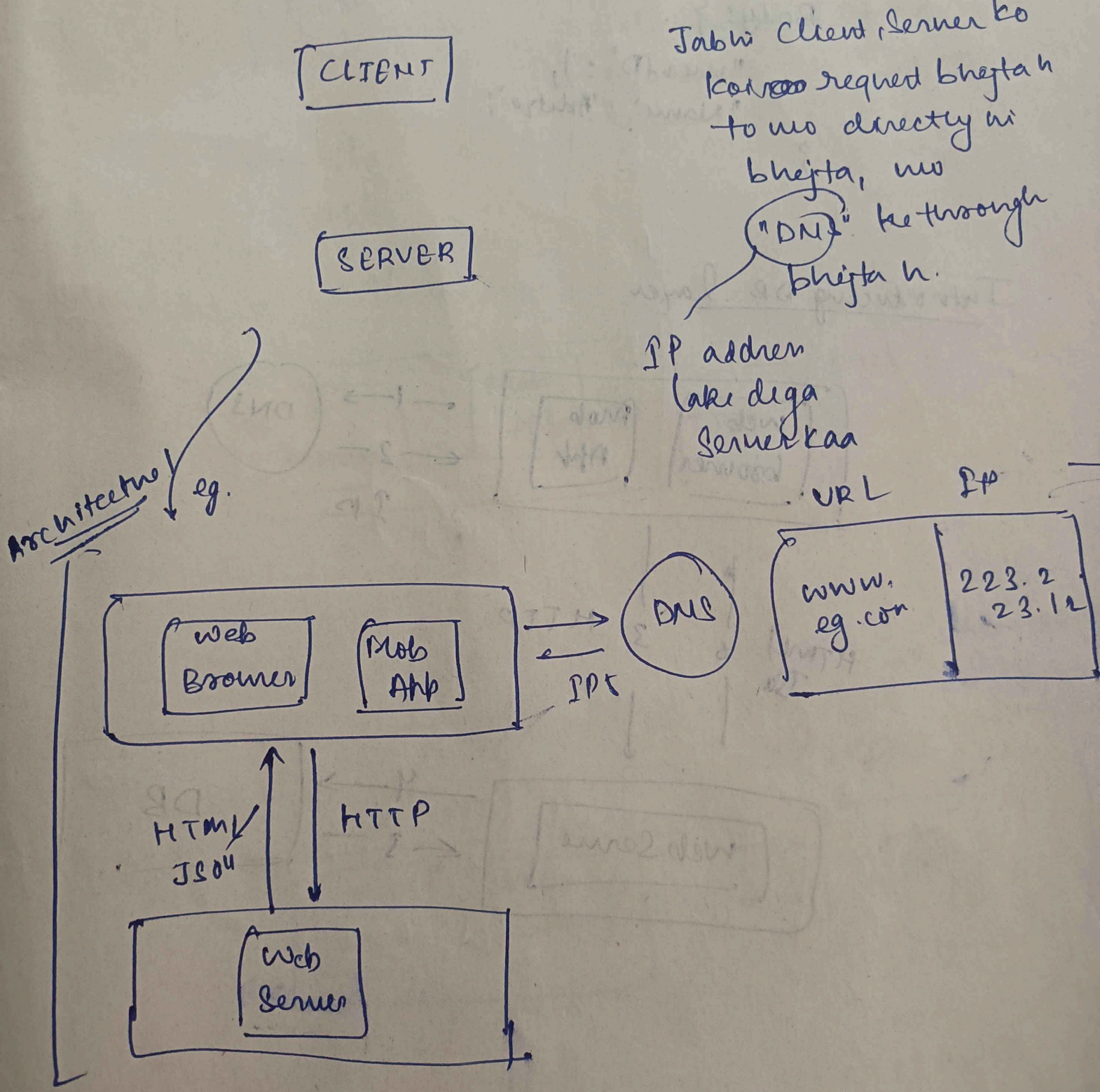
Lec 02

Anything that we could do in hld is to make our application Scale Better

Simple Client-Server Architecture



HTTP mai kya kya hota h?

(POST, GET, PUT etc.

URL: www.eg.com

Headers:

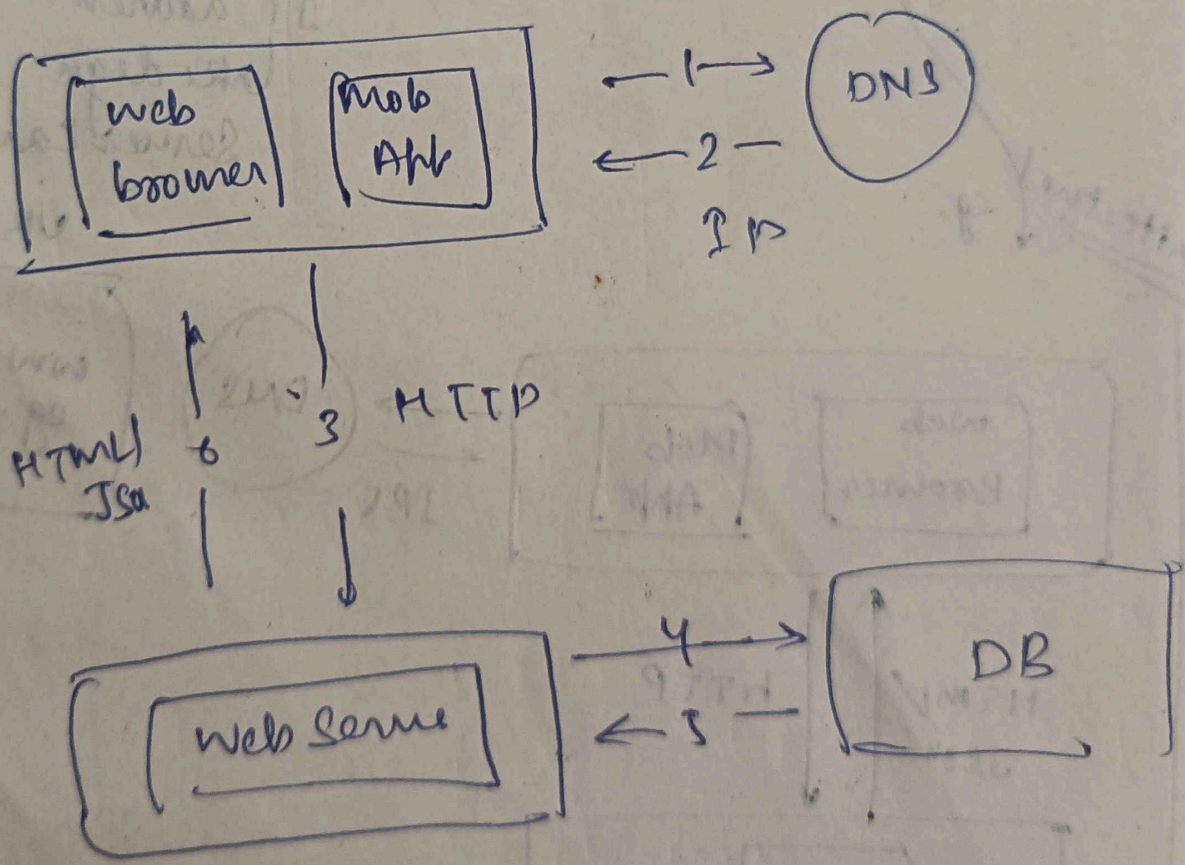
Body: {

"userID" : 1,

"Name" : "Aditya";

}

Introducing DB layer



Which DB to use?

Relational

SQL Based

mysql,
Postgresql,
oracle sql

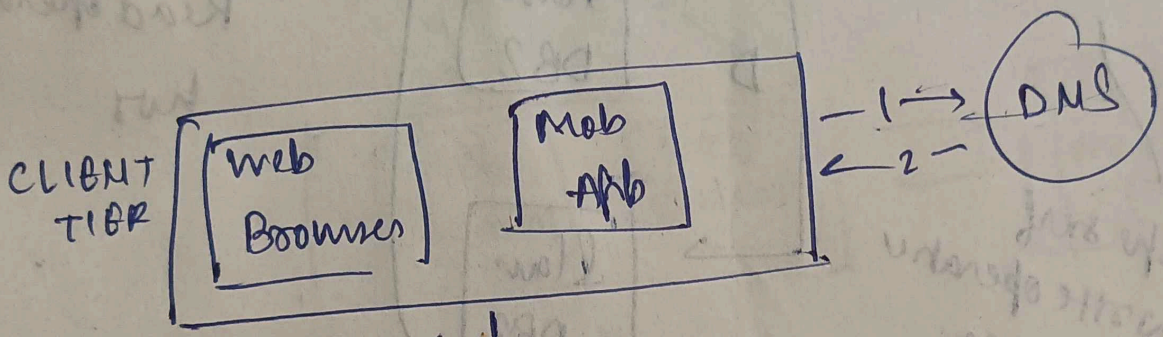
Non-Relational

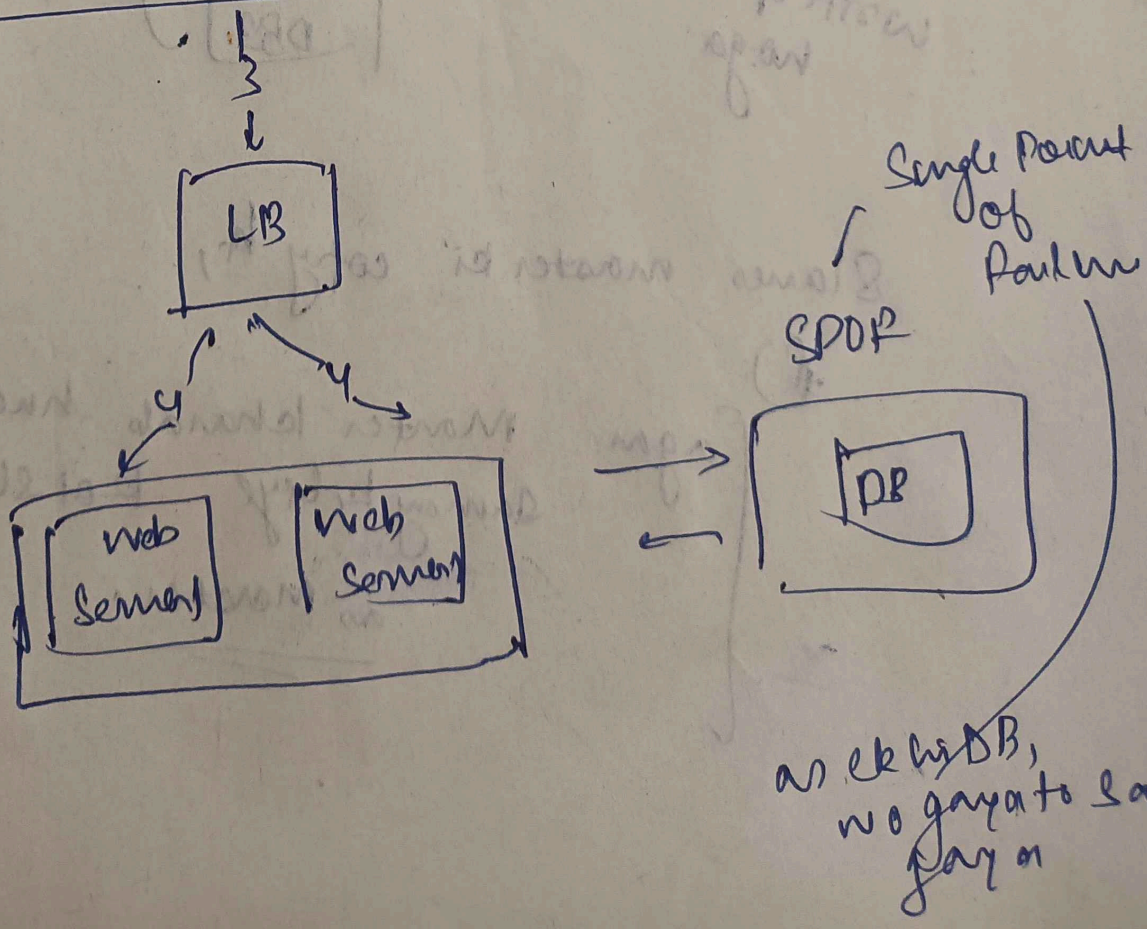
Key-value Store → eg:- Amazon DynamoDB

Column Based Store → Cassandra DB

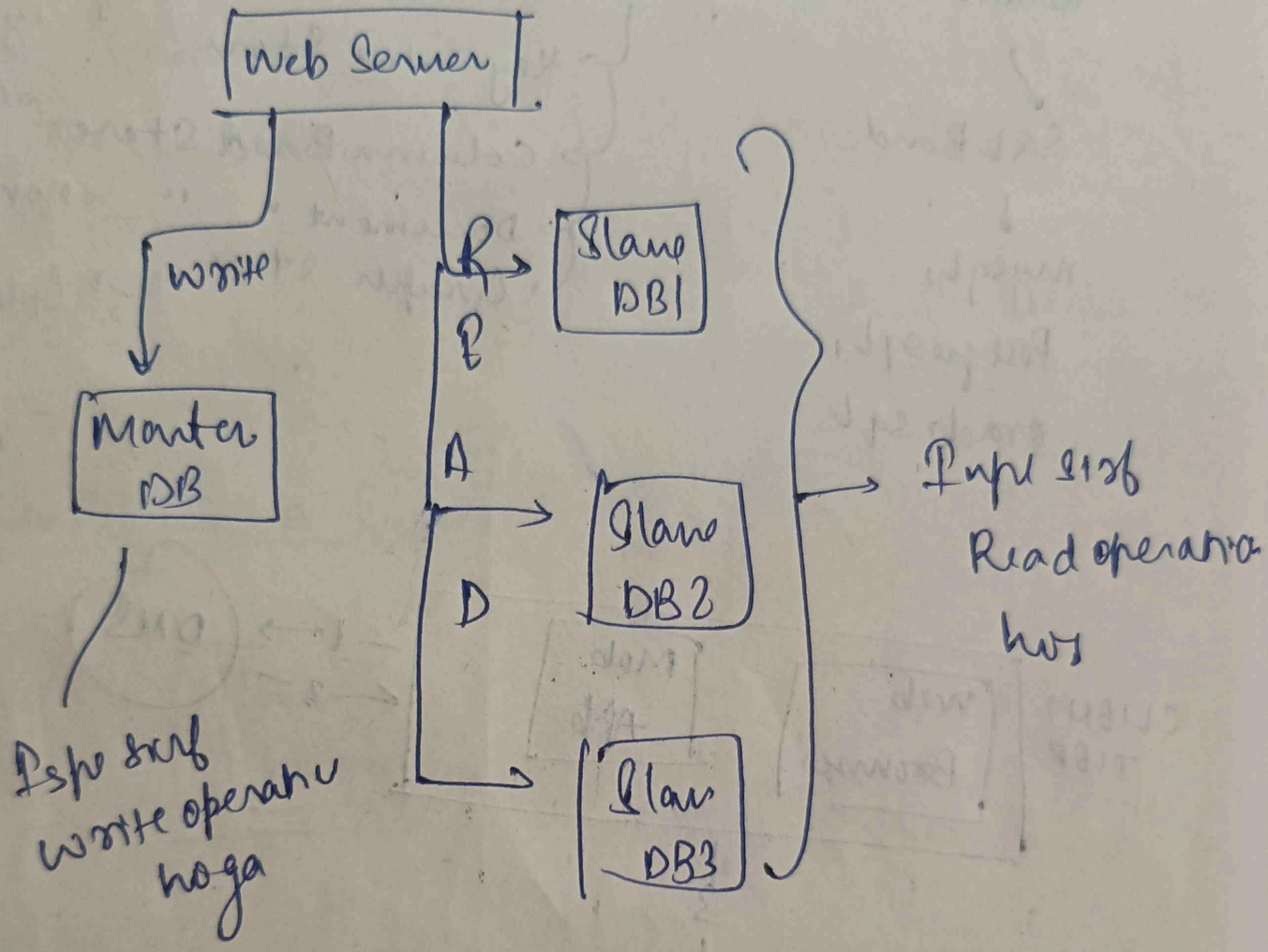
Document " " → mongo db

Graph Store → Graph QL



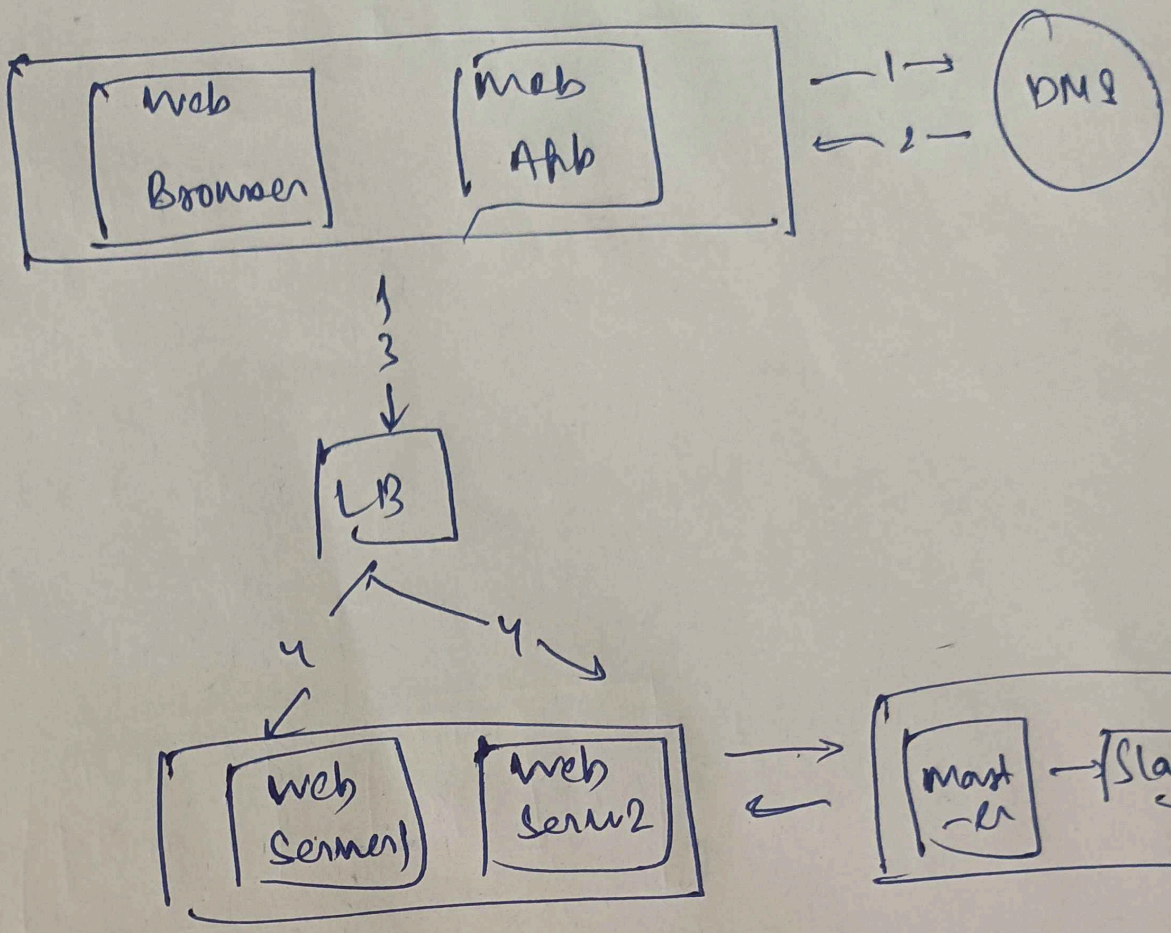


To SPOF se bachne ke liye.



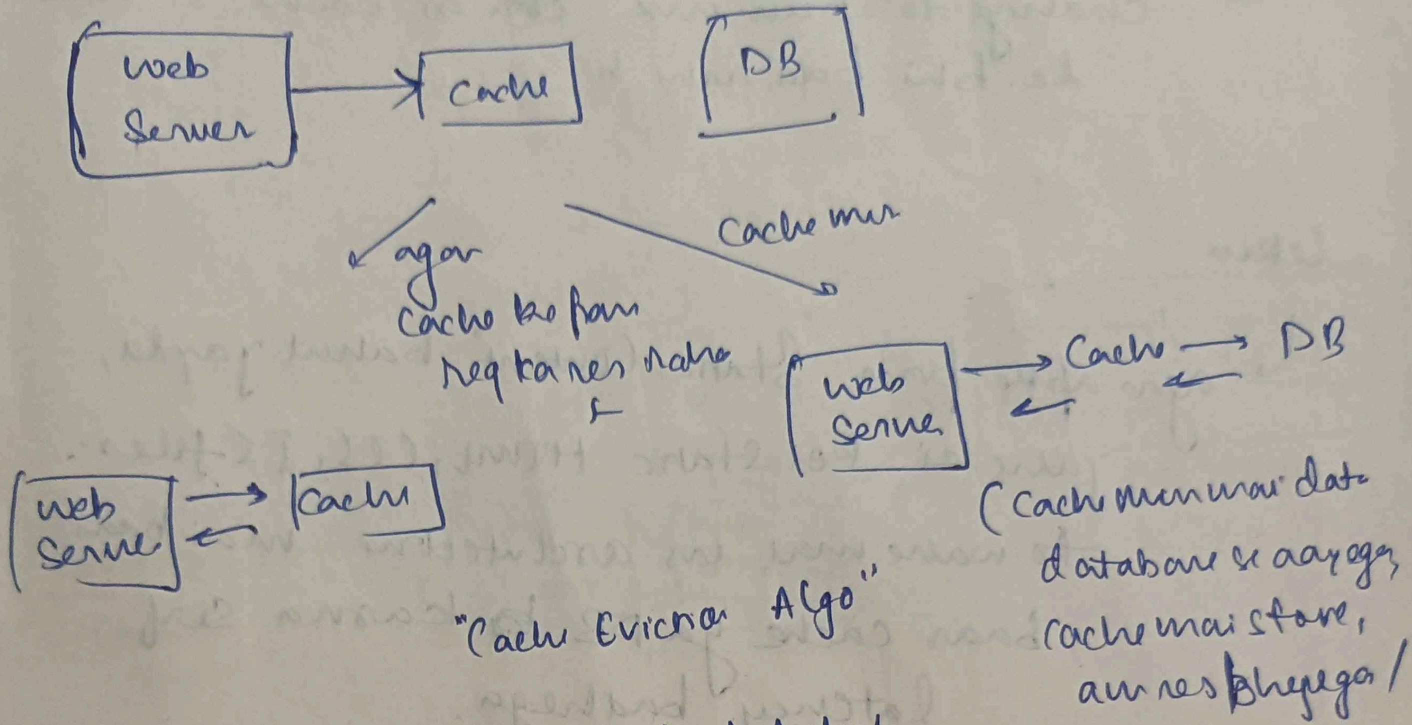
Slaves master ki copy hai,

f) agar master kharab hua to kuch samay ke liye koi slave aa jayega as master.

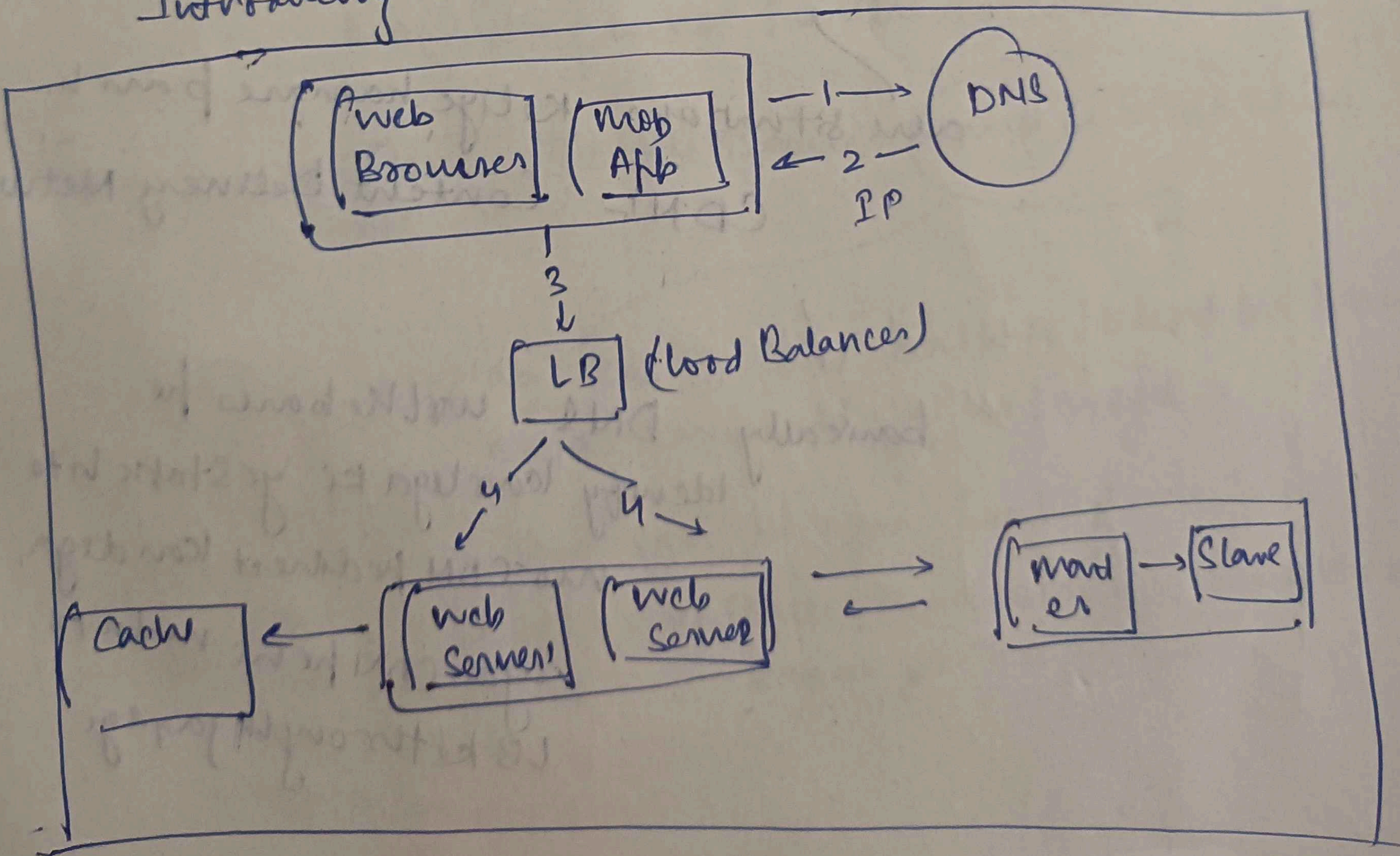

Web Browser
Web Apb
1
2
DMS
1
3
LB
4
4
web server1
web serv2
mast -er
Slave

Cache

- faster than DB



Introducing Cache in Architecture



Abhi dikkat to koi ni aa rahi h, hamari application
aache se chal rahi h, bahut saare users ko handle
bhi kar paa rahi h. , aur agar user ko jaldi data
chahiye to hum unke cache se bhi
de bhi paa rahe h.

Lekin

↳ agar apke pass 'Static Content' bahut jayda,
jaise ki koi static HTML, CSS, JS files.
to wahe mai is architecture mai baar
baar cache ya DB ko karna sirf
latency badhega.

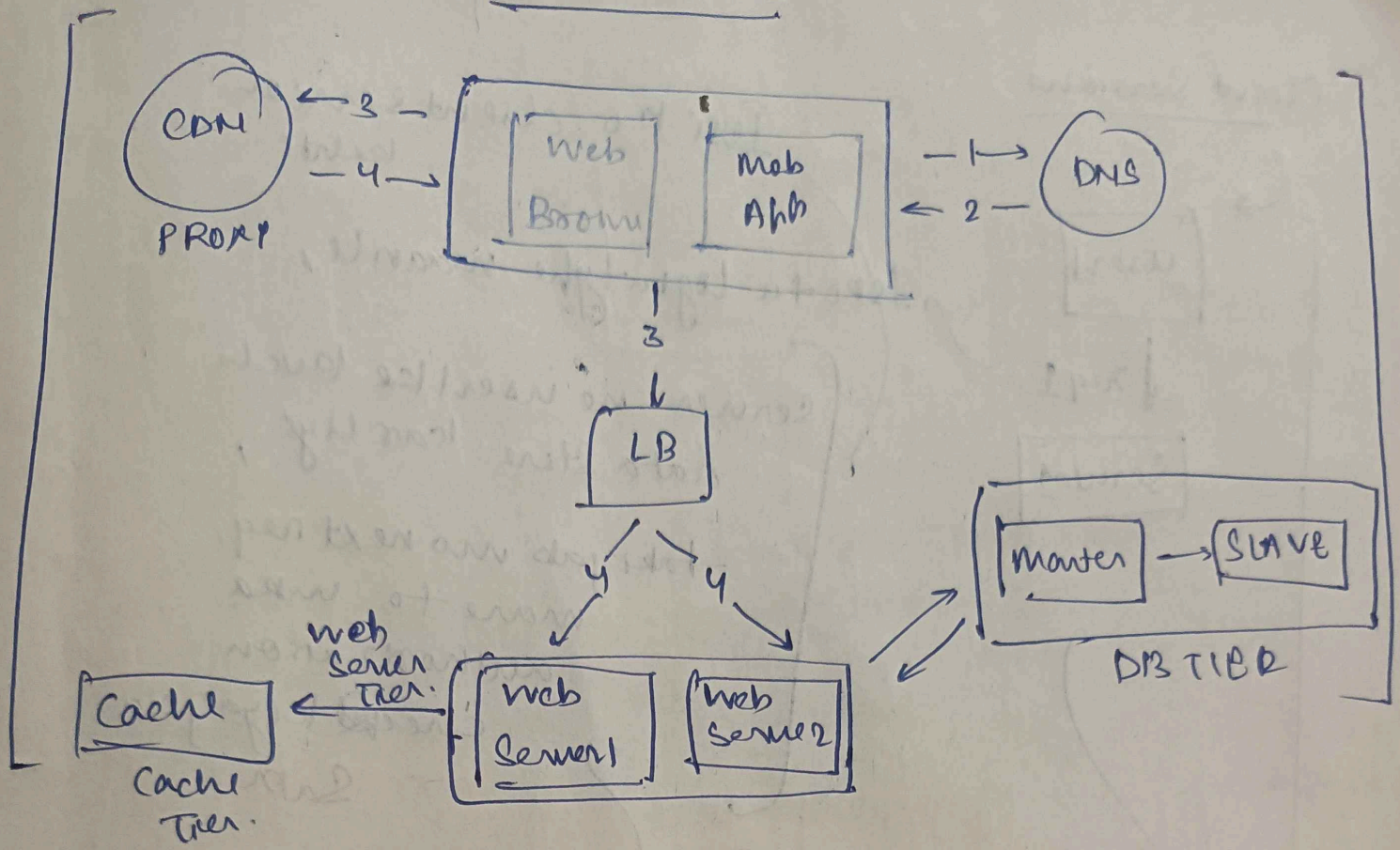
↓

aise situations ke liye hamare pass h.
CDN:- Content Delivery Netwo

↓

basically DNS work karega ke
identify karlega ki ye static h to
uss CDN pe redirect kar dega,
agar CDN pe nhi mila to
LB ke through jayega

CDN Introduced



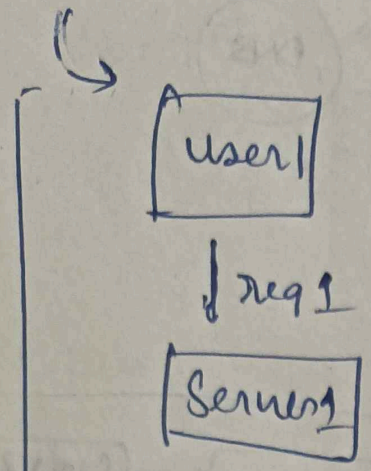
AutoScaling is not possible for server.

→ Dikkat ye h ki Sakte. Ye WebServer khud se Scale ni ho

matlab kal ko load badh to khud ba khud naya server add ni ho jayega,

aur sath hi agar load decrease to webServer ka count khud ba khud kaam ni ho jayega

Client Sessions



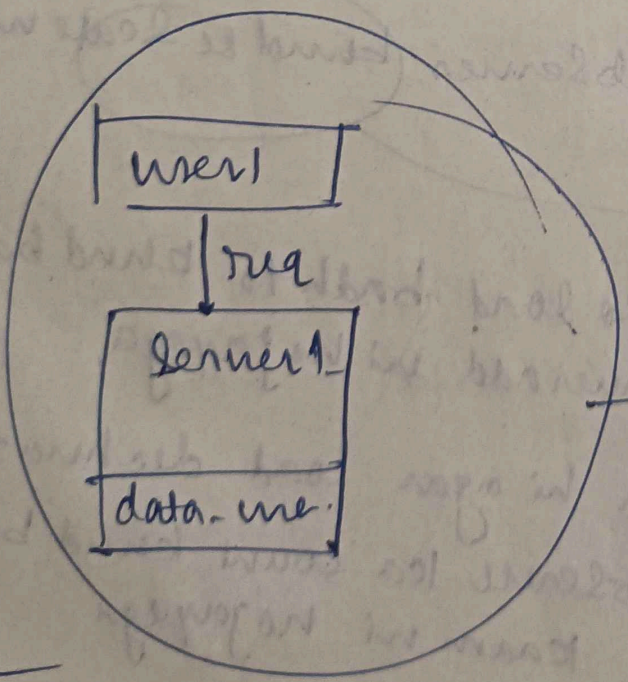
Jese ko client server kehte.

Jeko tu login type maanle,

server ne user1 ke kuch data store kar liye,

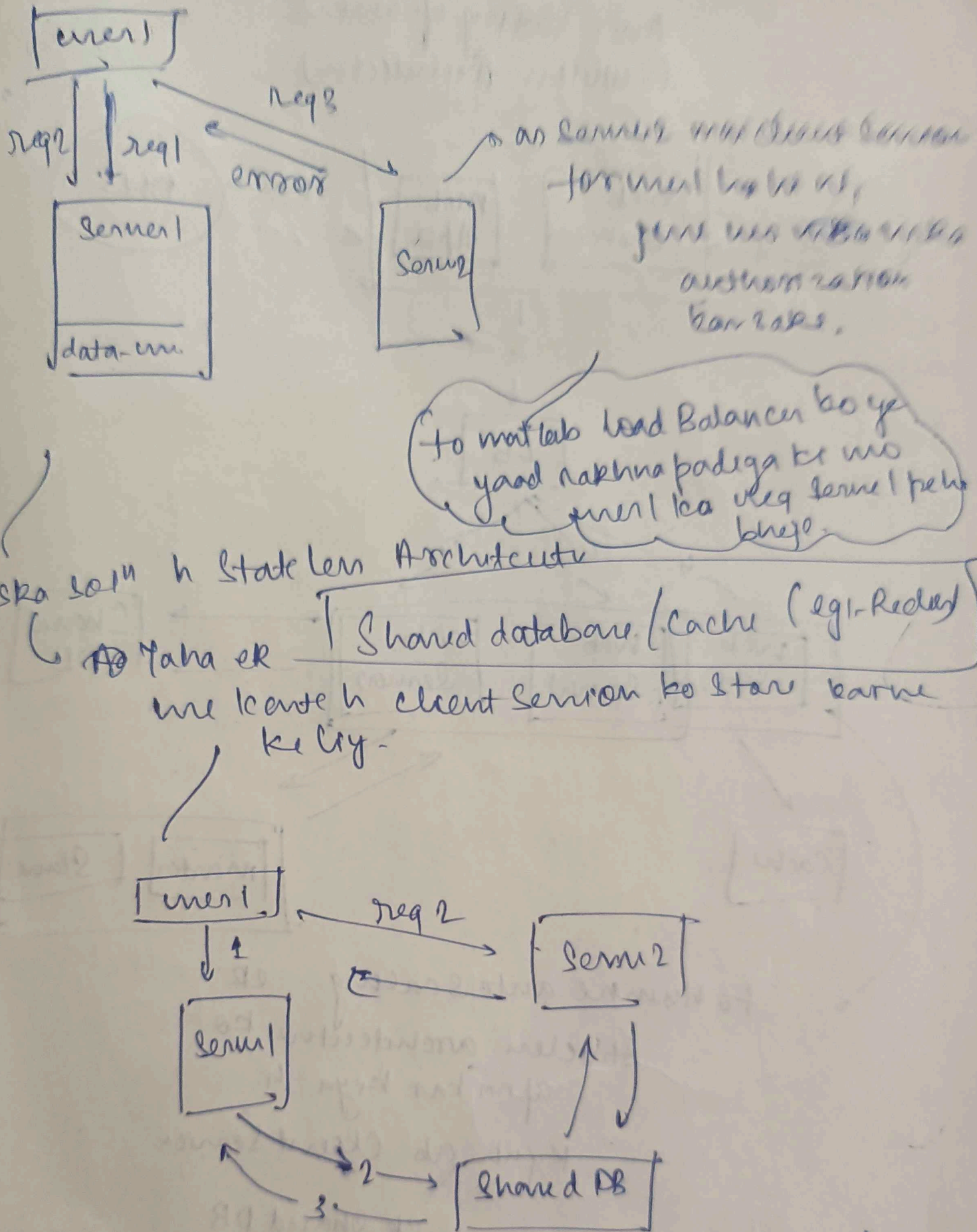
taki jab wo next req. mare to uska authorization check kiya jaa sake

let say us data to "data-user" keh rahe.

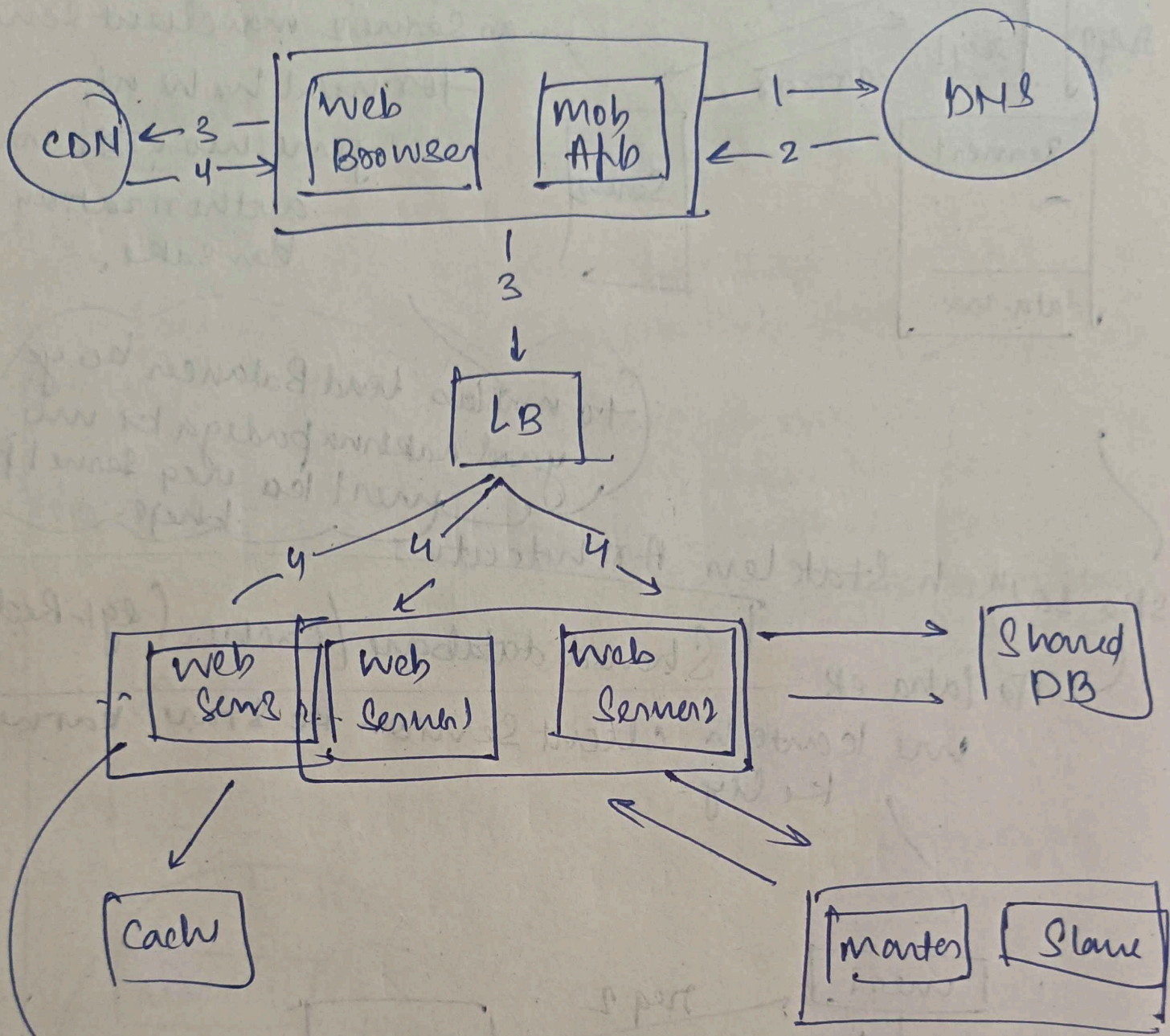


ab to me sikkar ye h ki, agar alga req user1 he Server2 pe maarato usse error milega as Server2 mai "data-user" h hi ni

"Statefull" architect


user1
req2
req1
req3
error
Server1
data-cm
Serv2
to matlab load Balancer ko ye yaad rakhna padega ki wo user1 ka req Server1 pe hi bheje.
Iska soln h Stateless Architecture
Shared database / cache (eg:- Redis)
Yaha ek user k create h client session ko store karne ke liy.
user1
req 2
Serv2
Server1
1
2
3
Shared DB
Yaha par client session store honge.

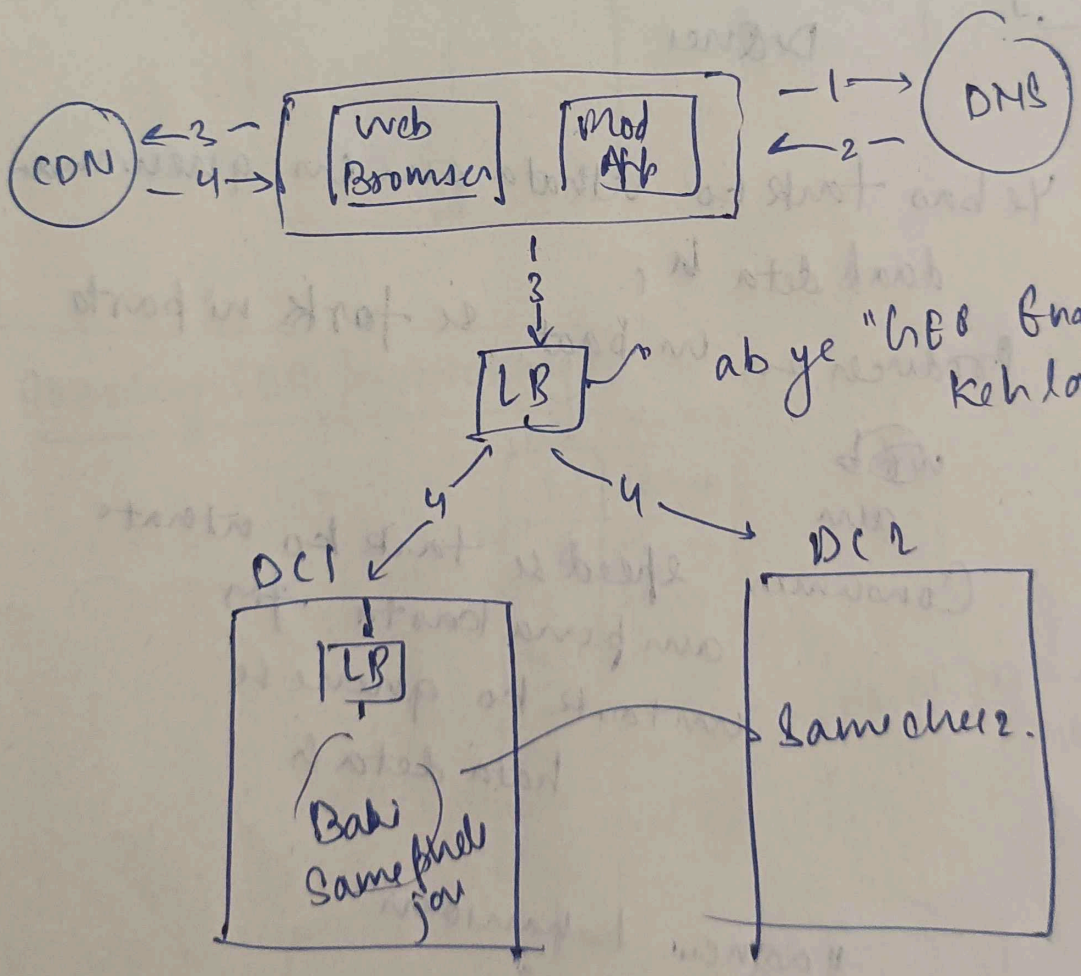
Auto Scaling of Server.
(Stateless Architecture)



to humne auto scaling ek
Stateless architecture ko
apna kar liya h.
kyunki ab Client Server
ek Shared DB
mai store hoga
naa ki ki ek
Server mai

Stateless ka matlab hi hai DB ko bhi matlab ki
kaun mera aaaya h, mujhko aur
authenticate karde raha h.

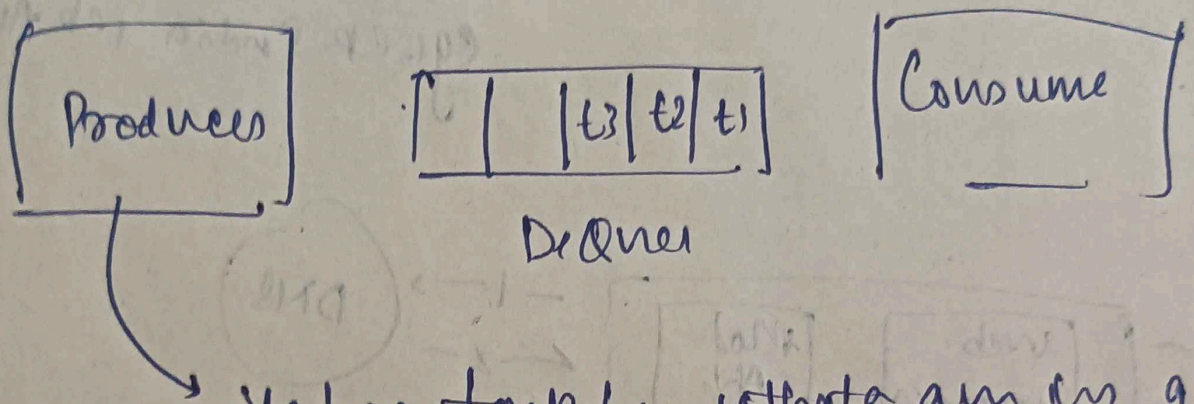
Lekin ye international ko server karne mai kafi time
lagayega, isliye hum data regions bnayenge.
(
or
data center.
Eg. CR India ,eb America



Managing Queue (or PUB-SUB Model)

eg:- Kafka, RabbitMQ

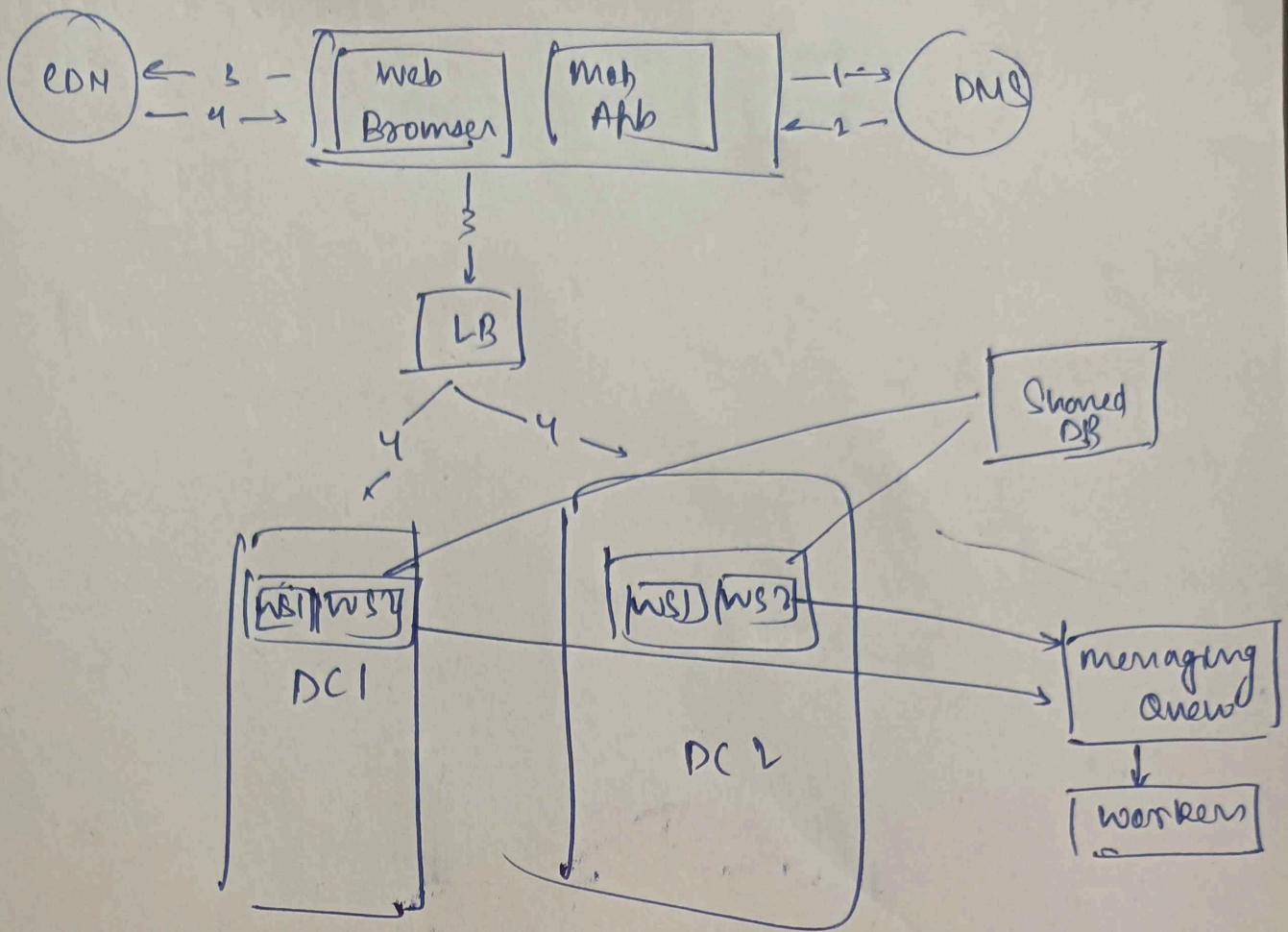
Isme do parts hote h → Producer
→ Consumer.



Ye bas task ko utha ta aur us queue mai
daal deta h,
Producer ko us baat se fark ni parta

aur
Consumer speed se task ko utha te
aur pura karta fir
us task ko queue se
hata deta h

-ve async behaviour.



Sharding (DB partitioning)

uid	name
1	adityo
2	abhay
3	erme
4	das
5	sda

no. DB jisme partitioning hoke jaa sakta h.

mod % 3

erme

adity das

abhay sda

DB0 DB1 DB2.

Ye ye master h, unke kaun ke slaves bhi honge.