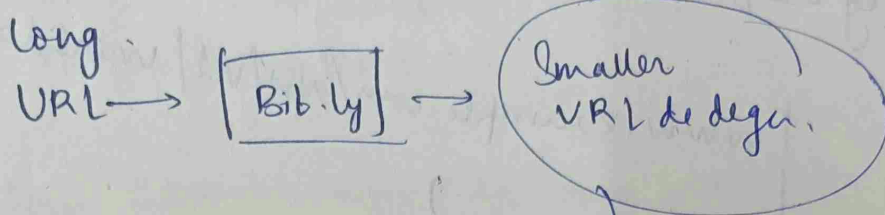


Lec 61 - URL Shortening Bitly

Design a URL Shortener

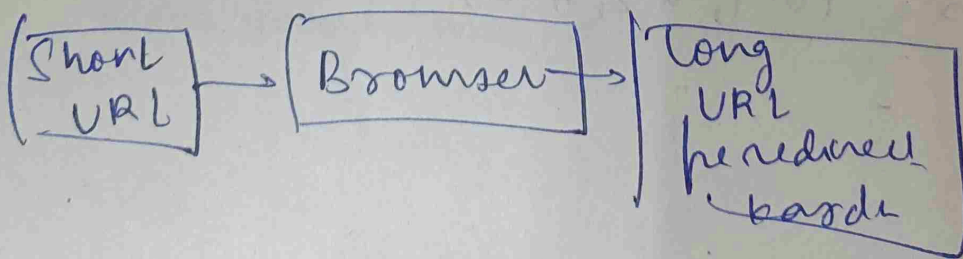
Q) URL Shortener kya hota h?



HLD ek open ended Question h.

→ matlab ki interviewers aapre expect karta h ki mai unse counterQ puchho, aur mai apni cheere (matlab kitne cheere per focus karna h) unse daba du.

→ To jaise ki caretke liye, aap use bahut saare question puchoge aur firka boundary set karoge ki mujhe kya karna, fir aap unse create karna start karenge.



Q1) How many people we are targeting?

→ 20 million URL's avg users per day

Q2) ex eg logs.

[www.example.com/api/v1/userId=1&username=Aditya]

↳ Ye long URL pan kye h.

↓
www.bit.ly/aQw12

Q3) How small ~~the~~ is the short URL?

→ As short as possible.

Q4) What characters can i use in short URL?

→ (0-9, a-z, A-Z)

Flow:-

- (i) User comes to our service by giving us a long URL.
- (ii) we will give him back a short URL.
- (iii) User put this short URL in browser & is redirected to the long URL.

Ek flow Sochna khud se aur us flow ko interviewer se confirm kar lena.

Note:

→ Haan ek HLD interview mai ye koshish karo ki aap "back of the envelope" calculations karo

(ye krna hⁿ).

→ Taki jab aap application bana rahi ho, to uske scale ka idea lagane ke liye, Query per sec, Storage sab

Note:

"Koi bhi interviewer ye ni kahega ki back of the envelope calculation karo, ye mere mind mai honi chahiye"

DPS (Query Per Second)

20 million URLs generated per day -

↓

Write Operation: 20 million / day

$$\begin{aligned}\text{per sec} &= 20 \times 2^{20} / 86400 \\ &= 231.48 \times 10^{-6} \times 2^{20} \\ &= \underline{231}^{\text{write}}, \text{ operation/sec.}\end{aligned}$$

Read operation. (Hamare main read operation jayda hoga)

Let's say 1:10 ratio (write: read) let's h.

$$= 231 + 10 \text{ read operation/sec.}$$

Storage Unit

Assume for 5 year.

20 million URLs daily.

$$= 20 \text{ million} \times 365 \times 5 \text{ year}$$

$$= 36,500 \times 2^{20} \text{ records.}$$

$$= 36.5 \times 10^3 \times 2^{20}$$

$$= 36.5 \times 2^{10} \times 2^{20}$$

$$= 36.5 \times 2^{30} \text{ records.}$$

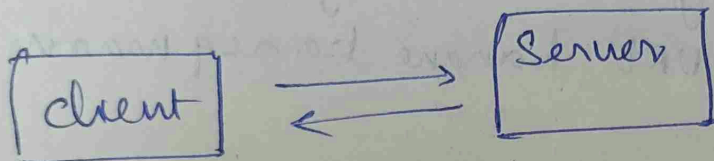
= 36.5 Billion

Assume 1 record takes $\rightarrow 100$ Byte

36.5 Billion record takes $\rightarrow 36.5 \times 2^{30} \times 0.1 \times 2^{10}$

$$= 3.65 \times 2^{40}$$

≈ 3.65 TeraByte



client aur server apas mai interact karte honge using some APIs.

\rightarrow To jabhi hum High Level Design banate h to hum ye define karte h ki kaun kaun si APIs hongi jinkethrough hamara client hamari service se baat kar raha hoga.

To abhi filhaal two main ^{API} endpoints consider kar rahi.

① Generate a short URL

② Redirect to long URL

matlab jab us short URL ko browser pe daaloge to phle us req hamari website pe aayegi jaha se hum long URL de dengi wapas.

① Generate a short URL

Ab alag alag tariki ke jisme mai Short URL banane ka req maaru

Case ①:-

POST: `www.bit.ly/api/v1/generate`.

QP: `longURL: value`.

"Shayad Query Part"

OR

Case ②:-

`www.bit.ly/api/v1/generate/`
`??longURL??`

OR

Case ③:- (Body mai ban karu)

POST: `www.bit.ly/api/v1/generate`

Body: {

longURL: value

}

→ Is sab me ye method sabse kharab h kyunki

agar mai longURL sath mai daalunga to mere browser URL ki link bahut
badhi ho sakti, aur kahi bad

browser link ke size pe limit
bhi laga sakta h.

Case (II) :- Ye best Practice h to
fahaal ke liye Case (III) ke
sath jayenge.

Main task and their method

① Generate a Short URL

POST: www.bit.ly/api/v1/generate

Body: {longURL?}

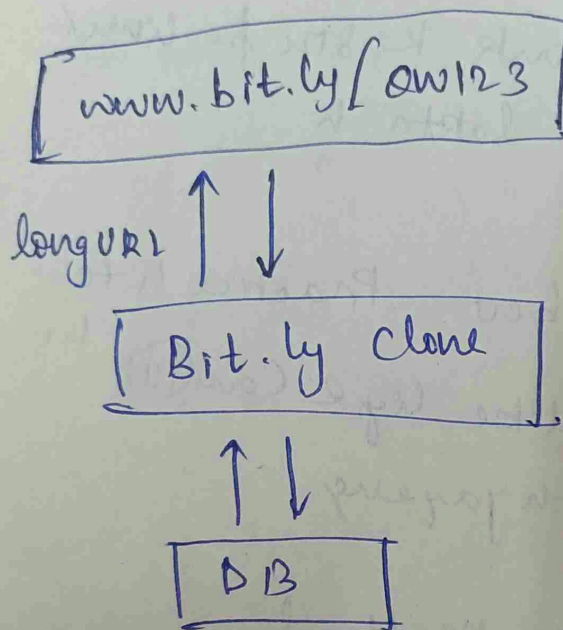
Return: ShortURL

(1) Redirect to long URL

Get: `www.bit.ly/{shortURL}`

Response: Redirect to long URL

How to Redirect?



Status Code: → Same 200 code hote h

In sab code ko me kiya mai to mai

browser ko link to de dunge

per user ko fir se me link to

ctrl+C & ctrl+V karna padega

lekar mai chahata hu ki mai URL bheja to
browser khud ba khud us URL ko handle

↳ to unke liye h → 301 & 302

↳ dono hi Redirection
ke liye, bas ek
antar h.

301: Redirection: Permanent
Redirection.

302: Redirection: Temporary redirection.

↳ Isme apni browser ko kehnahe ki ye data
cache mai store karlo, ye ye link to link
to map krige,
to agli baar req mere server pe na aayegi,
browser cache se hi map kar lega
(Client Side Caching)

↳ Isme haan req mere server se hoke jayegi,
(kise koi Client Side Caching ni
hoti h)

~~www.bit.ly/Qw123~~

www.bit.ly/Qw123 $\xrightarrow{301}$ `{{longURL?}}`
Client Side Caching

www.bit.ly/Qw123 $\xrightarrow{302}$ `{{longURL?}}`

How do we decide ki koun sa status code bhagna h?

301 → bit.ly/QW123 → {{long URL}}

301 bhjunga to req mere site pe ni aayenge
browser apni cache pe data rakh lega,
wahi se map kar dega.

↳ Ye tab aachra - jab.

Aapka server small h.

You want minimum DB Read Query

→ Par me mujhe pata ni chaliga ki mera user koun koun

302 → bit.ly/QW123 → Service → {{long URL}} } bana h

↳ Iss me saare req

mere ~~site~~ site se ~~map~~ hoke jayenge

↳ To isme mujhe apni
user bare karde rakhna.


Ye tab shi h jab:

→ Server is Scalable.

→ DB Read Query is Optimized

→ We can have user metrics

Interviewer se discussion ke baad humne ye decide
krega ki

"302"  "backed by the fact ki
1:10 (write:
read.
operation h)

uri karenge

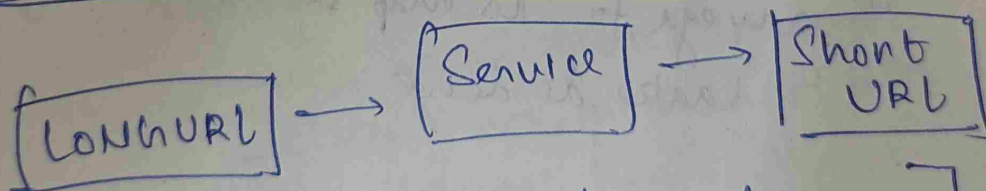
Design an Initial Flow

Two main Task h:

① URL Shortening (POST)

② URL Redirect (GET)

① URL Shortening



Post: `www.bit.ly/api/v1/generate`
Body: `{longURL}`
Return: Short URL

To matlab ek map ke tarah h kuch.

mapping, long URL \rightarrow Short URL

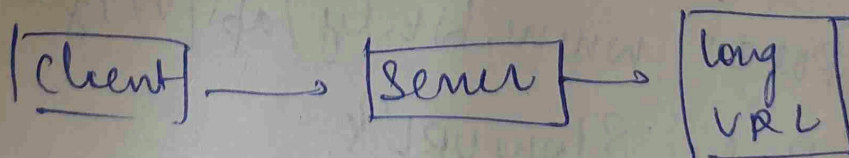
long URL \rightarrow mapped to \rightarrow Short URL

✓
sabse phle demag mai

HashMap/Hash Table data hai

[Key: Short URL
Value: long URL]

Agli baar jab client Short URL ki sath
ayega to is map se hum long URL
laake de deng.



How to generate Short URL from long URL?
→ Hash fn

Long URL → Hash fn → Short URL

↳ bahut sare h jann SHA-1
MD5

Ques

aur fir is Short URL aur long URL ko store
kar lenge Database.

hashmap

key	
Short URL	long URL

Storing:-

→ user comes with a long URL.

→ we generate a short URL for the same
using hash fn.

→ Store the short URL as key and
long URL as value in our hashmap

→ Return the short URL.

Redirecting:

(i) User comes with a short URL

(ii) we find the equivalent long URL by doing map.get(shortURL).

(iii) Redirect the user to long URL.
(302 redirect)

Q1) Apne kaha apne hashmap me krige, lekin hashmap to apke application ke andar hota h.

to jaise hi apki application restart hogi.
(matlab RAM ka. sara data gaya,
to apka ye hashmap kisi kaam ka nikahe)

→ To apko na kisi tarah se use kisi database mei store karna padega

You have to store the value in DB

To Key: Value ka alternative for DB is.

DB: Hashmap → NoSQL (Amazon Dynamo)
MongoDB

Dekho hum HLD mai jo bhr application
ki baad kar rahi hote h un

"Distributed Application" hoti h

↙
Aur hame pata h ki Distributed Application
mai hamare paas multiple "servers" hote h,
aur multiple "DB Server" bhi hoti h
(matlab Db Replication and Scaling)
Horizontal Scaling → Sharding

To agar aap sara kuch ek hi mai store karoge to
dikkat ho jayegi, to aapko interviewer ko
batana padega, ki hamara application
kal ko scale karega to

iske liye hum

"Consistent Hashing" use karenge.

↙
Ab iske aane se hamare paas
2 Hous fn ho gaye.

The 2 Hark fu:

(1) To find out Shard, (DB) (SHA-1)

(2) Find out the shortURL from the longURL (MOD-5)

lekin itne mai further round theek h, lekin
mai to "Silent killer" hu to le

Better Design dekh

Ye jo data aap NoSQL mai store kar sakte
uske jagah

aap SQL use karso.

Currently hum an key value pair NoSQL
mai store kar sakte.

Dynamo DB

Key	value
Short	Long
URL	URL

lekin agar kal to same url again
data h to, mujhe check
karne ke liye ye volphile se
h ya ni, mujhe $O(N)$ lagenge.

Reason

② Agar complex query karni h

joane \rightarrow

"for a particular user how many

URLs are generated till now".

C to no is key, value se hi data kero paoge

SQL \rightarrow Support Complex Queries
efficiently.

Reason

③ write operations are faster in SQL than
in NoSQL.

To SQL mai kuch area bana kr

To SQL mai kuch area bana kr		
ID (Primary key)	Short URL	Long URL

To convert a long url to a short URL
we have many algorithms like.

(MD5, SHA-1, SHA-256 etc)

"Starting mai pucha tha humne ki generated
URL ki length kitni honichaligi?"

↳ kaam se kaam possible.
(How small?)

Possible symbols: (0-9, a-z, A-Z)

Total = $10 + 26 + 26 = 62$ characters.

(matlabek place ke liye 62 options)

↳ lekin ye saare kaafi bare hash value
generate karte h

Wrong Permutation:

www.b.t.ly/

(
If we take just 1 character $\rightarrow 62$ possible url
" " " 2 " $\rightarrow 62^2$
" " " 3 " $\rightarrow 62^3$

How much char I should take $\rightarrow n \rightarrow 62^n$

Hamare paas 36.5 Billion to store
to n s.t $62^n > 36.5$ Billion.

$n=6$ ~~042~~

7

8 6 Bala

www.bit.ly/-----

to ab name pata chal gaya h kr

$n=6$

lekin

SHA-1, MD5, SHA 256

to kaafi badi character generate krte h.

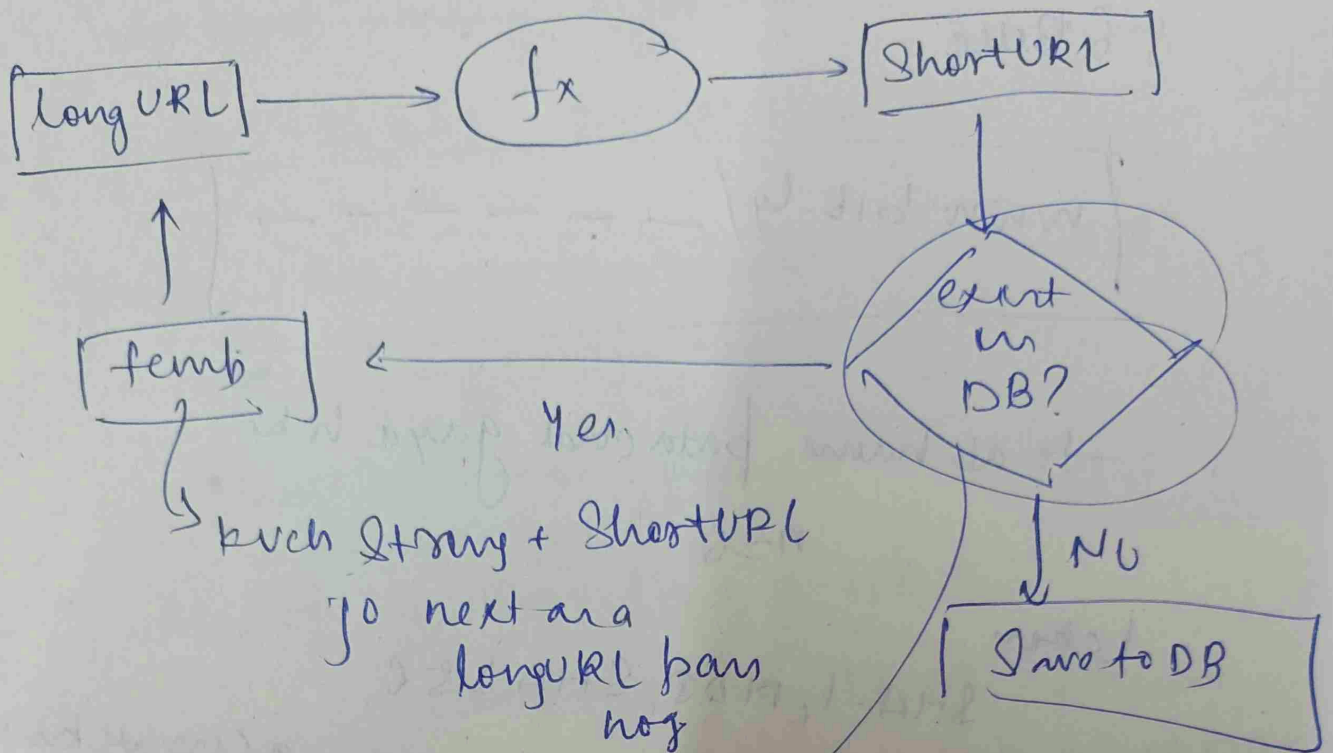
To reduce we can:

(1) Exerting hash value

se first 6 character lein

→ lekin isme collision ke chances badh jayega.

to Hash + collision ko handle krna padega.



Disadvantage:

(i) DB hits increases:

(ii) Slow

(iii)

BASE X Approach

Base 10 \rightarrow Base 62

62 Numbers (0-9, a-z, A-Z)

0 - 0
1 - 1
:
9 - 9
10 - a
11 - b
:
35 - z
36 - A
61 - Z

eg convert

14320 into Base 62

62	14320	
	230	60
	3	40
	0	3

3HY

to jo hamare SQL db banayi hai

Id	ShortURL	LongURL
14320		www.example.com

Ab rather than LongURL ko hum kaha
store karne ke jagah

hum Id ko Base 10 se Base 62 mai

convert karenge aur ShortURL
mai store kar dengi.

Base 10

14320

Base 62

3IY

Id	ShortURL	LongURL
14320	3IY	www.example.com

www.bit.ly/3IY

Advantages:

→ As Id unique hogi to ShortURL

is unique

to collision impossible hogaya

Disadvantages:

(i) Need a unique ID generator

(Although SQL khud Row ID generate karta h, par use use kr kar sakte hyn; no (batayenge baad mai))

(ii) URL length is not fixed

(iii) It is easy to figure out the next generated short URL

14320 → 3IY

n to

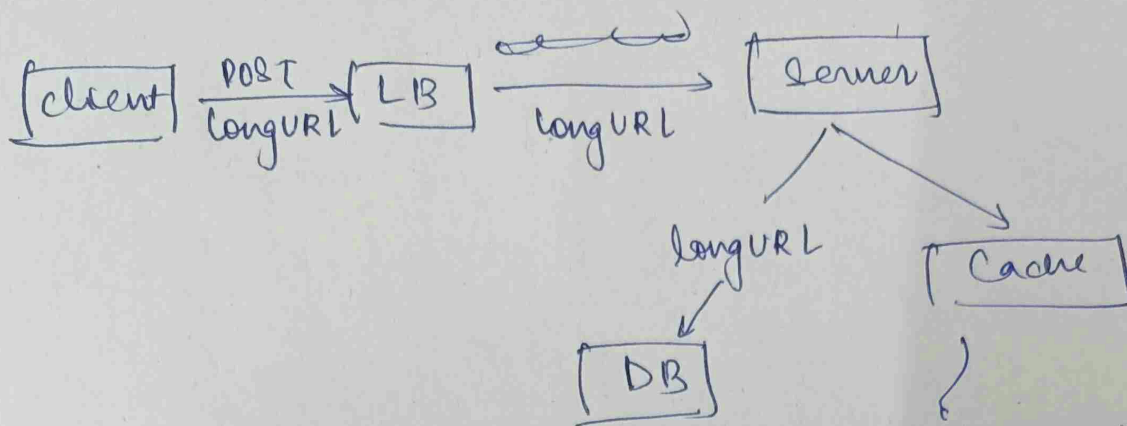
14321 → 3IZ

kuch tyhe hi hoga

Kyunki hum distributed system mai h, an sharding time, do primary key same krka sakte h, are mai hume

centralized ID generator ke taraf jana padega

Storing



(Agar koi URL jayda frequently access ho rahi ho to uske cache mein store kar denge)

Redirection

