

2005-2006

Joshua

11:35 - 40pm

clcl. ucio. us

- Create account
- Same urls
- Get all urls



RAM: 512 MB
HD: 40 GB
CS: 1.6 GHz

- ↔ Code
- ↔ db
- ↔ other apps

If 1 record takes 1 KB of memory.

1024 Bytes \approx 1000B

& if he gets 1 million same requests per day

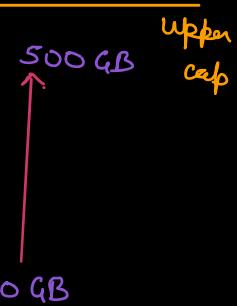
Amount of memory required to store URLs of a single day : $10^6 \times 10^3$ Bytes

$$\begin{aligned} &= 10^9 \text{ Bytes} \\ &\approx 1 \text{ GB} \end{aligned}$$

40 days $\triangle!$

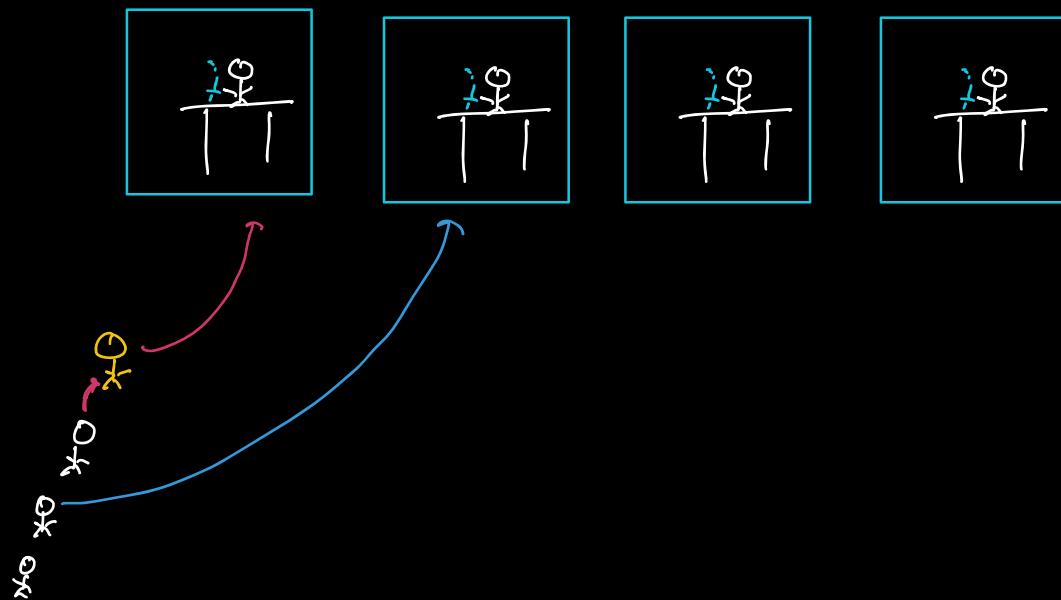
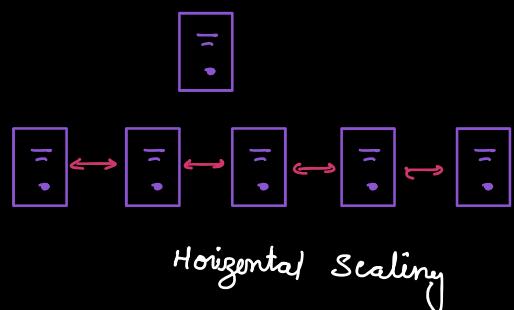
Solutions

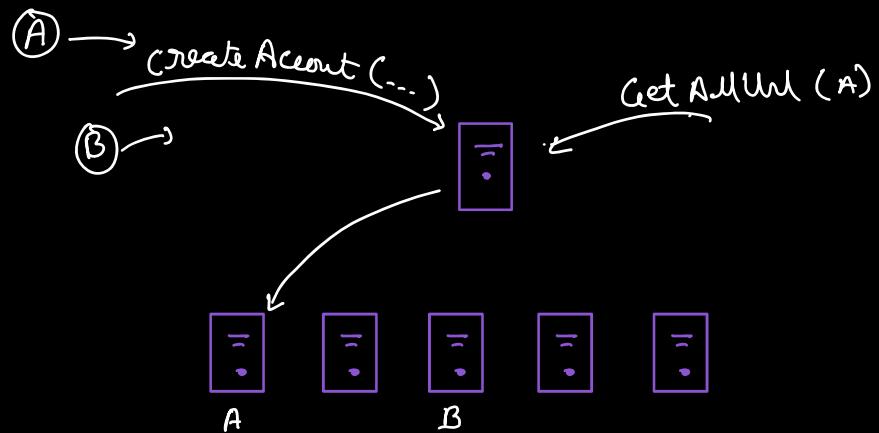
① Buy a better system \longrightarrow More memory



② Buy more systems

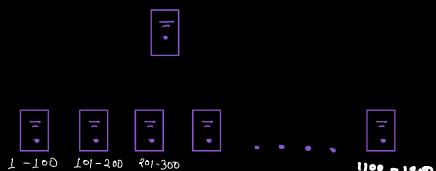
Data partitioning / Data
Sharding





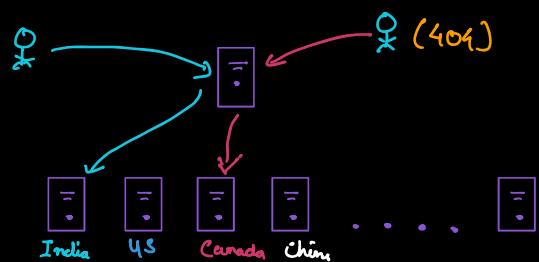
① Range based (userId / Name)

- Uneven distribution



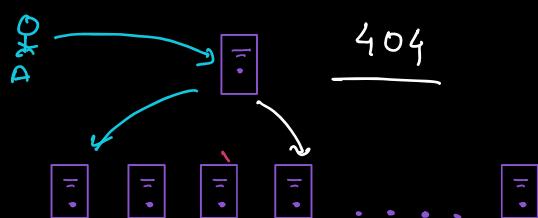
② Based on regions

- Inconsistent
- Uneven distribution



③ Load based

- Inconsistent



④ Round Robin

- Inconsistent

⑤ Hashing / Hash function

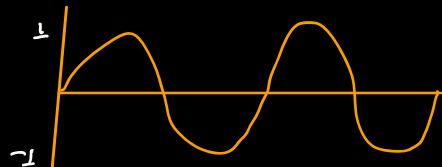
Hash Function

$$\left. \begin{array}{l} f(x) = x^2 \\ f(10) = 100 \\ f(5) = 25 \\ f(100) = 10^4 \end{array} \right\}$$

Range of output

is not limited

$$f(x) = \sin x$$



$$f(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0 \end{cases}$$

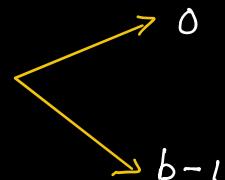
Mod %

$a \% b \longrightarrow$ Remainder of $a \div b$

$$112 \% 10 \longrightarrow 2$$

$$35 \% 7 \longrightarrow 0$$

$$f(a, b) = a \% b$$



$$\text{userId \% 4} \longrightarrow [0-3]$$

userId	userId \% 4	userId \% 5	\vdash	\vdash	\vdash	\vdash
15	$15 \% 4 \Rightarrow 3$	$15 \% 5 \Rightarrow 0$	\vdash	\vdash	\vdash	\vdash
41	$41 \% 4 \Rightarrow 1$	$41 \% 5 \Rightarrow 1$	\vdash	\vdash	\vdash	\vdash
7	$7 \% 4 \Rightarrow 3$	$7 \% 5 \Rightarrow 2$	\vdash	\vdash	\vdash	\vdash
8	$8 \% 4 \Rightarrow 0$	$8 \% 5 \Rightarrow 3$	\vdash	\vdash	\vdash	\vdash

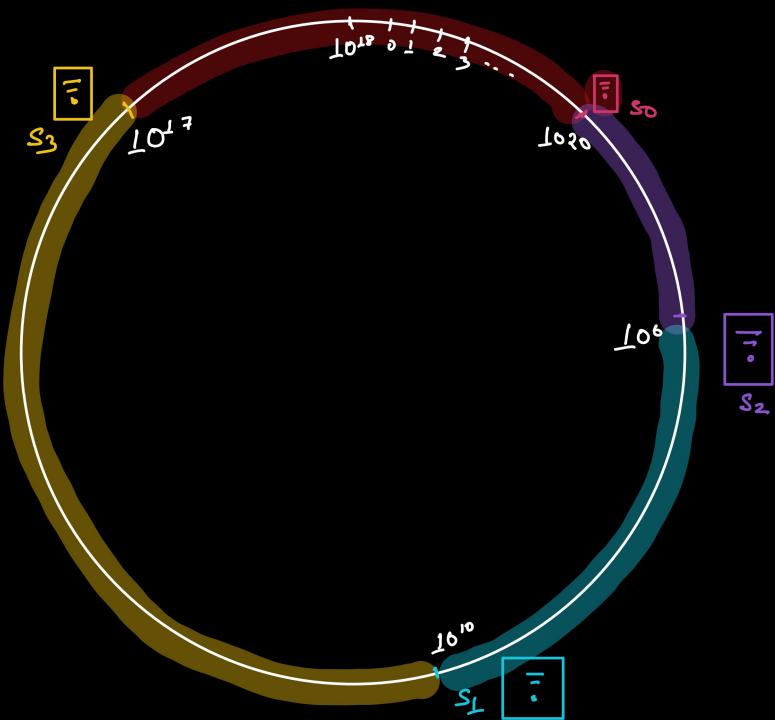
Inconsistent Hashing

Consistent Hashing

$$h_u(\text{userId}) = [0, 10^{18}]$$

$$h_s(\text{ServerId}) = [0, 10^{18}]$$

\uparrow
 Range No



$$h_s(s_0) = 10^{20}$$

$$h_s(s_1) = 10^{10}$$

$$h_s(s_2) = 10^6$$

$$h_s(s_3) = 10^{-3}$$

$$h_4(101) = 514$$

$$h_4(123) = 1000467$$

$$h_4(12) = 10^{17} + 20$$

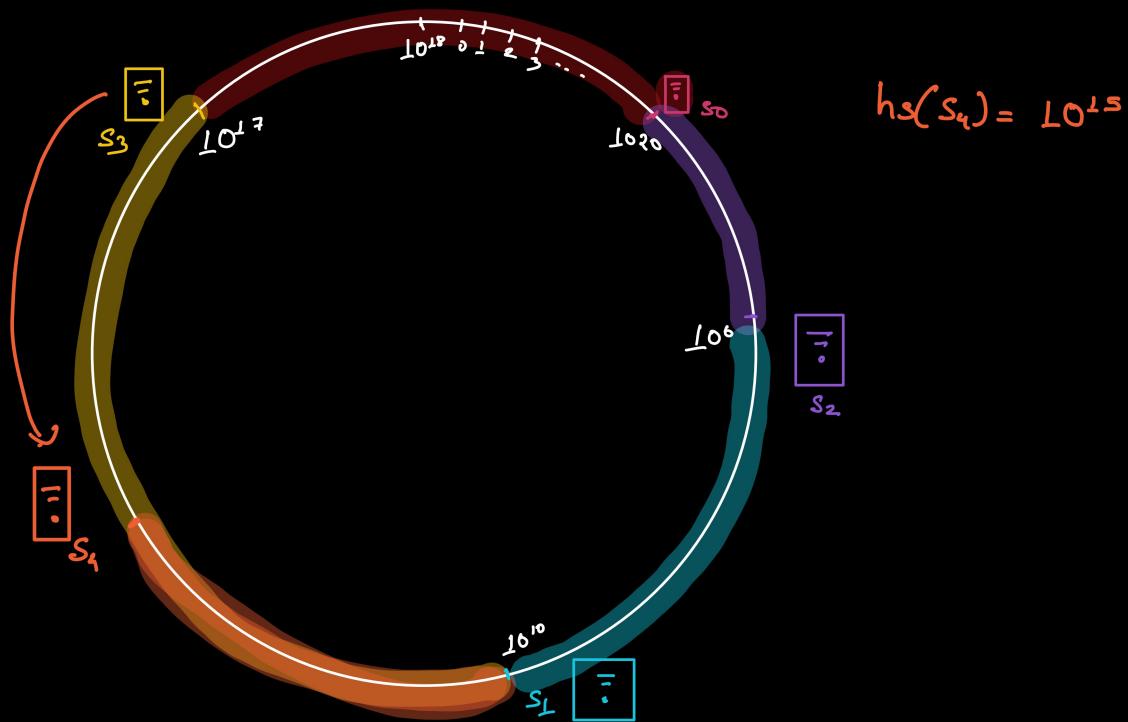
1020	10^6	10^{10}	10^{17}
------	--------	-----------	-----------

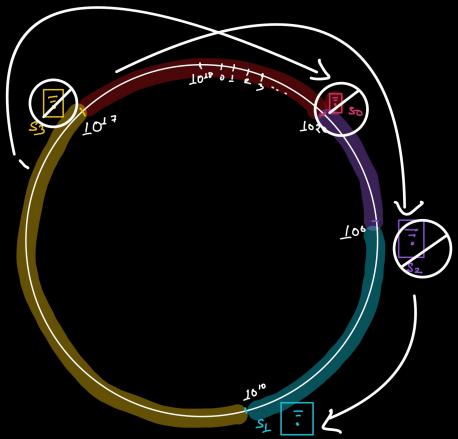
s_0

s_2

s_1

s_3





Ideally

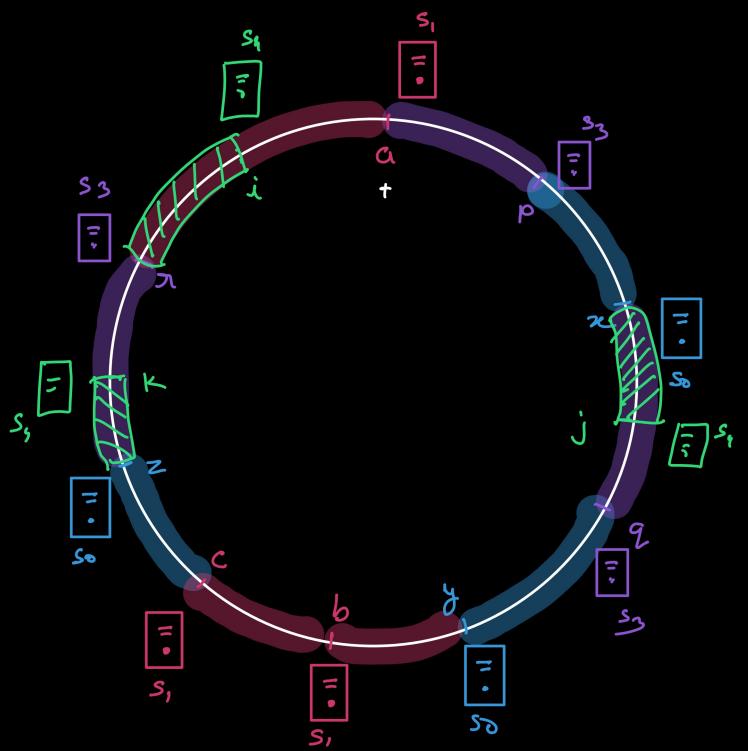
When we add a new server \rightarrow load of all

servers should be
distributed equally

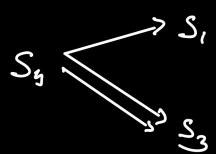
When a server goes down \rightarrow load of that server should
get distributed equally across
all servers.

$$PV = nRT$$

$h_1()$	$h_2()$	$h_3()$
s_0	x	y
s_1	a	b
s_3	p	q
s_4	i	j



$[p+1, x]$
 $[q+1, y]$
 $(c+1, z_j)$



Yahooo

$\boxed{a \mid p \mid x \mid j \mid a \mid y \mid b \mid c \mid z \mid k \mid r \mid i}$