2005-2006

Joshua

11:35 - 40pm

del.icio.us

→ Create account
→ Save 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 ≈ 1000B

& if he gets 1 million save requests per day

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

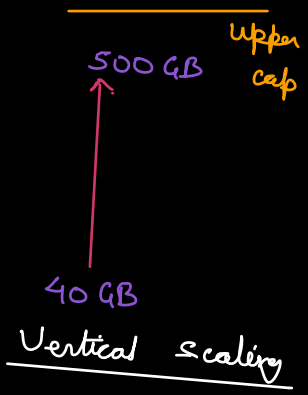
$= 10^9$ Bytes

≈ 1 GB

40 days ⚠

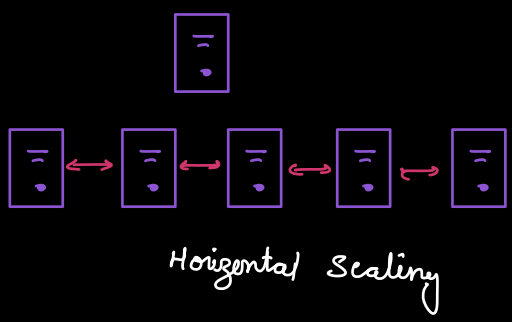
Solutions

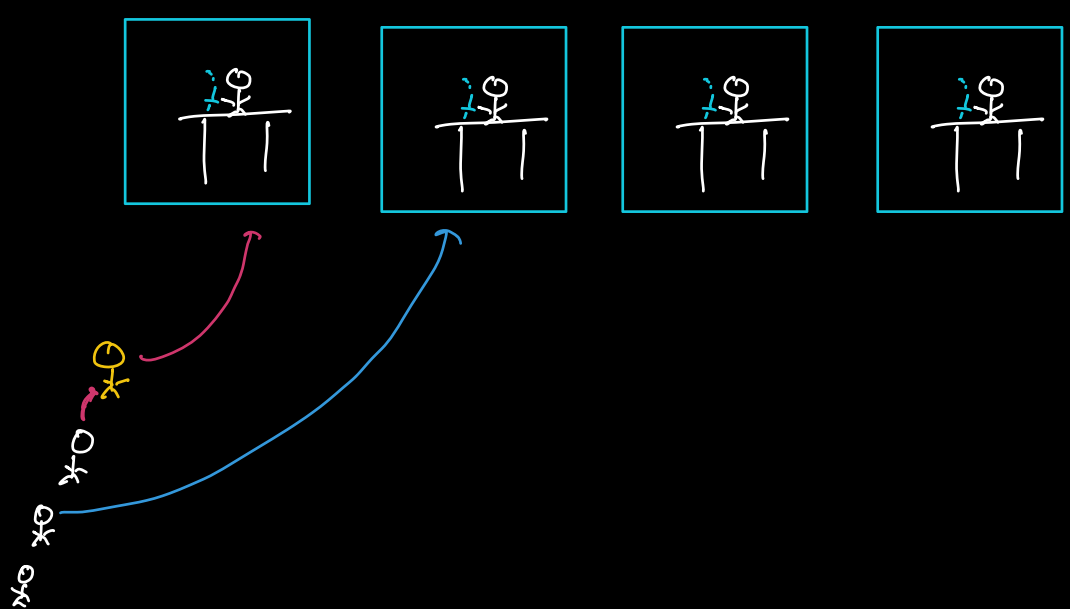
1. Buy a better system ⟶ more memory

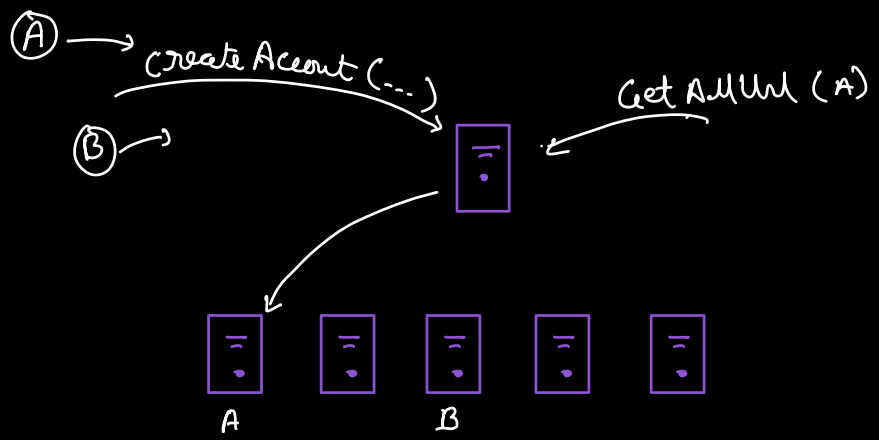


2. Buy more systems

Data partitioning / Data Sharding

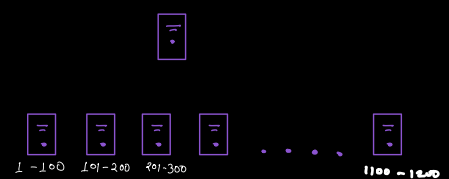






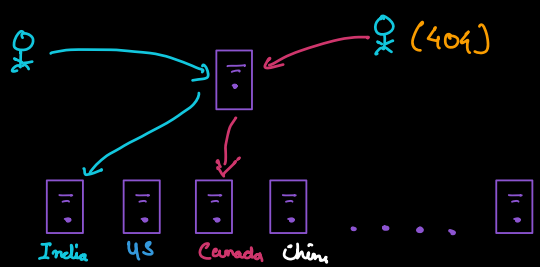
① Range base (userId / Name)

- Uneven distribution



② Based on regions

- Inconsistent
- Uneven distribution

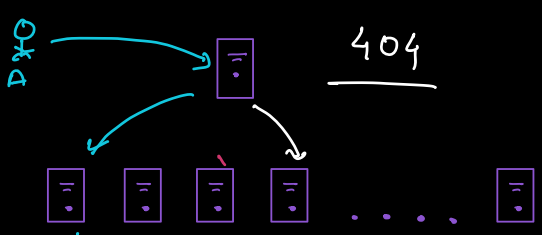


③ Load based

- Inconsistent

④ Round Robin

- Inconsistent



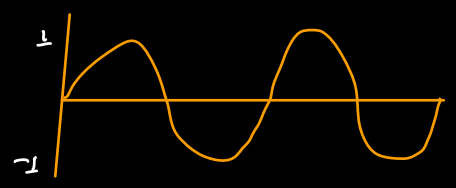
⑤ Hashing / Hash function

Hash Function

$f(x) = x^2$
$f(10) = 100$
$f(5) = 25$
$f(1000) = 10^6$

Range of out put 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 ⟶ Remainder of $a \div b$

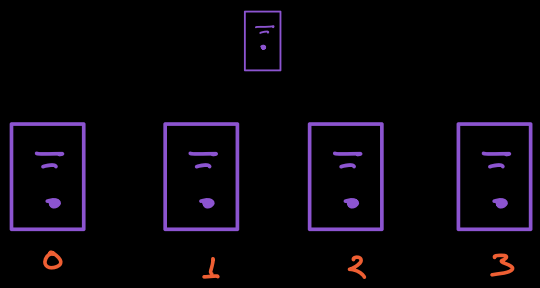
112 % 10 ⟶ 2

35 % 7 ⟶ 0

$f(a, b) = a \% b$ ⟶ 0 to $b-1$

UserId % 4 ⟶ [0 - 3]

UserId	UserId % 4	UserId % 5
15	15 % 4 ⇒ 3	15 % 5 ⇒ 0
41	41 % 4 ⇒ 1	41 % 5 ⇒ 1
7	7 % 4 ⇒ 3	7 % 5 ⇒ 2
8	8 % 4 ⇒ 0	8 % 5 ⇒ 3

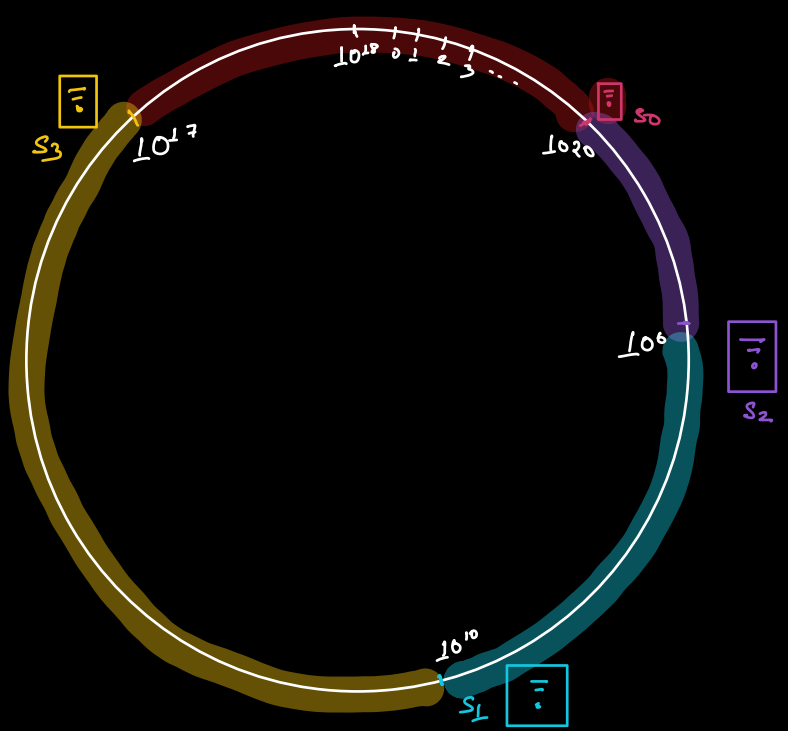


Inconsistent Hashing

Consistent Hashing

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

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



$h_s(S_0) = 10^{20}$

$h_s(S_1) = 10^{10}$

$h_s(S_2) = 10^6$

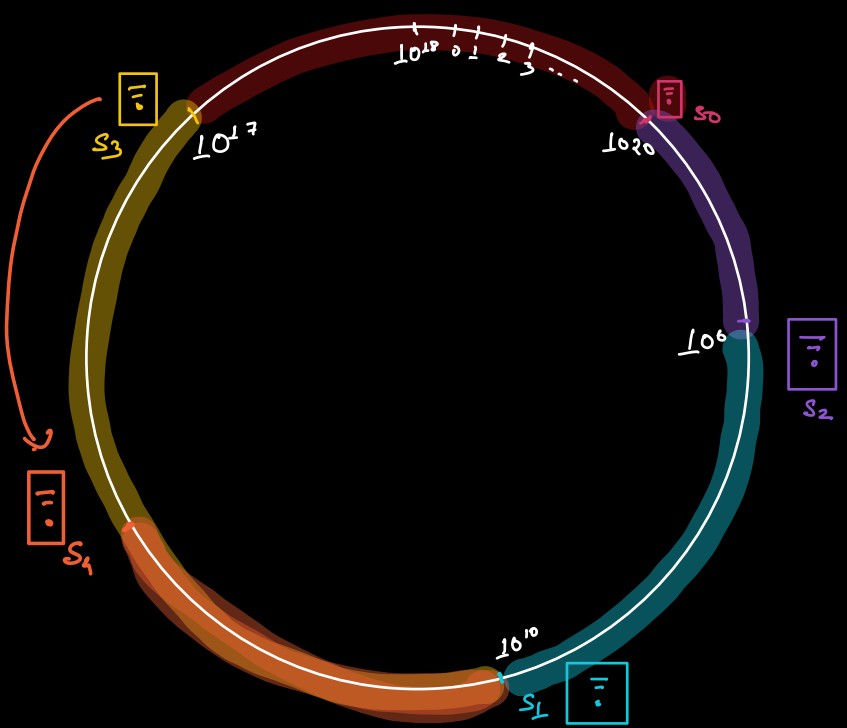
$h_s(S_3) = 10^{17}$

$h_4(101) = 514$

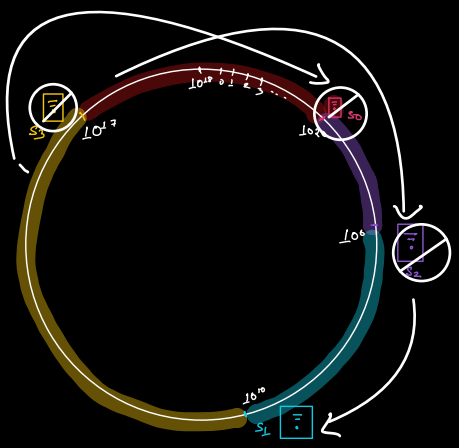
$h_4(123) = 1000467$

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

1020	10^6	10^{10}	10^{17}
s_0	s_2	s_1	s_3



$h_3(s_4) = 10^{15}$



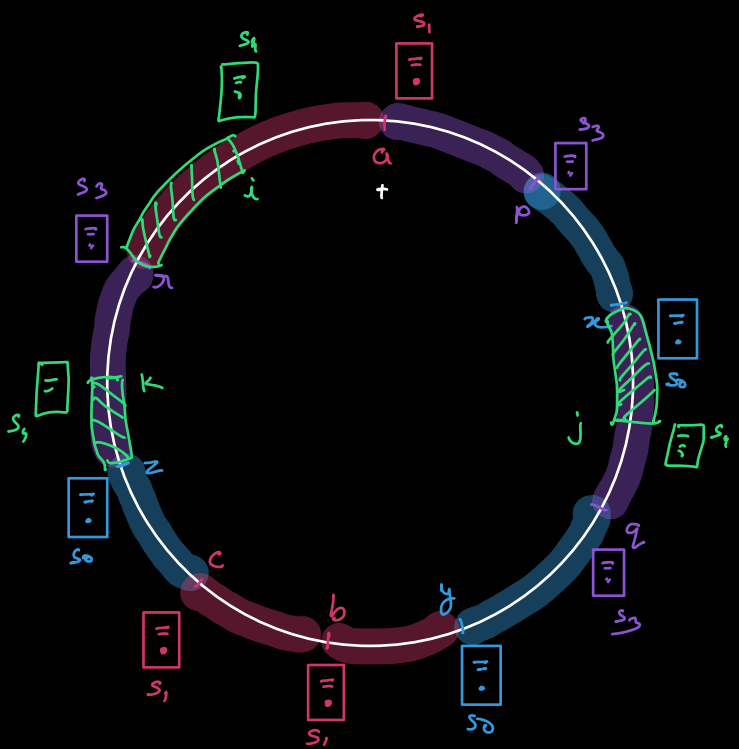
Ideally

When we added a new server ⟶ load of all servers should be distributed equally.

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

$PV = nRT$

	$h_1()$	$h_2()$	$h_3()$
S_0	x	y	z
S_1	a	b	c
S_3	p	q	r
S_4	i	j	k

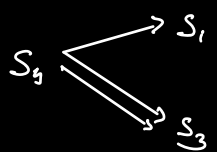


S_0

$[p+1, x]$

$[q+1, y]$

$[c+1, z]$



Yahoo

a	p	x	j	q	y	b	c	z	k	r	i