

RAID

RAID

RAID

3

%*Ž#

D 3;6

RAID

RAID

I/O

RAID

I/O

I/O

RAID

RAID

RAID

I/O

I/O

RAID

mirrored RAID

RAID

RAID

I/O

I/O

SCSI

SATA

RAID

RAID

DRAM

RAID

RAID

%*Ž\$

RAID

RAID

fail-stop

[S84]

RAID

RAID

%Ž%

D3;6

RAID

3RAIDcapacity

NRAIDN

N/2

reliability

44

performance

3RAIDRAID 0RAID 1RAID

4/5

PattersonGibsonKatz[P+88]

%Ž& D3;6 "

RAID	RAID	RAID 0	
striping			
	38.1	stripe	
4			
%* ž#	D3;6ž"		
0	1	2	3
0	1	2	3
4	5	6	7
8	9	10	11
12	13	14	15

38.1

0123

14KB

38.2

%\$

0	1	2	3	
0	2	4	6	
1	3	5	7	2
8	10	12	14	
9	11	13	15	

4KB

RAID

chunk size

8KB

4

32KB

RAID

RAID

the

mapping problem

RAID

RAID

$= 1 = 4\text{KB}$

A RAID

$= A\%$

$= A/$

$4/3 = 1.33333$

RAID 15 4

$14 \% 4 = 2$ 2 $14 / 4 = 3$ 3

2 0 3 0 14

[CL95]

4KB

64KB

D3;6Ž"

N

N

I/O

D3;6

RAID

I/O

RAID

RAID

I/O

RAID

sequential

random

1MB

B

B+1MB

10

4KB

550000

20100

DBMS

S MB/s

R MB/s

S *R*

S *R*

10MB

10KB

7ms

3ms

50MB/s

S

10MB

7ms

3ms

10MB @ 50MB/s

1/5s

200ms

10MB

210ms

S

$$S = \frac{10\text{MB}}{210\text{ms}} = 47.62\text{MB/s}$$

10KB @ 50MB/s 0.195ms

$$R = \frac{10\text{KB}}{10.195\text{ms}} = 0.981\text{MB/s}$$

$R = 1\text{MB/s}$ $S/R = 50$

D3;6Ž"

RAID-0

N

S
 $N \times R \text{ MB/s}$

I/O

RAID

%Ž D3;6 #

RAID

RAID 1

RAID

38.3

%Ž

D3;6Ž#

0	1	2	3
0	0	1	1
2	2	3	3
4	4	5	5
6	6	7	7

0

1

2

3

RAID-1

RAID-0

RAID-10

RAID 1+0

RAID-01

RAID 0+1

RAID-0

RAID-1

RAID

RAID

5

2

3

RAID

5

2

3

D3;6Ž#

RAID-1

RAID-1

=2

 N $N/2$

RAID-1

RAID-1

38.3

0

2

2

 $N/2$

RAID-1

RAID

RAID-1

RAID

consistent-update problem [DAA05]

RAID

RAID

RAID

0

1

RAID

0

RAID

1

0

1

inconsistent

0

1

atomically

write-ahead log

RAID

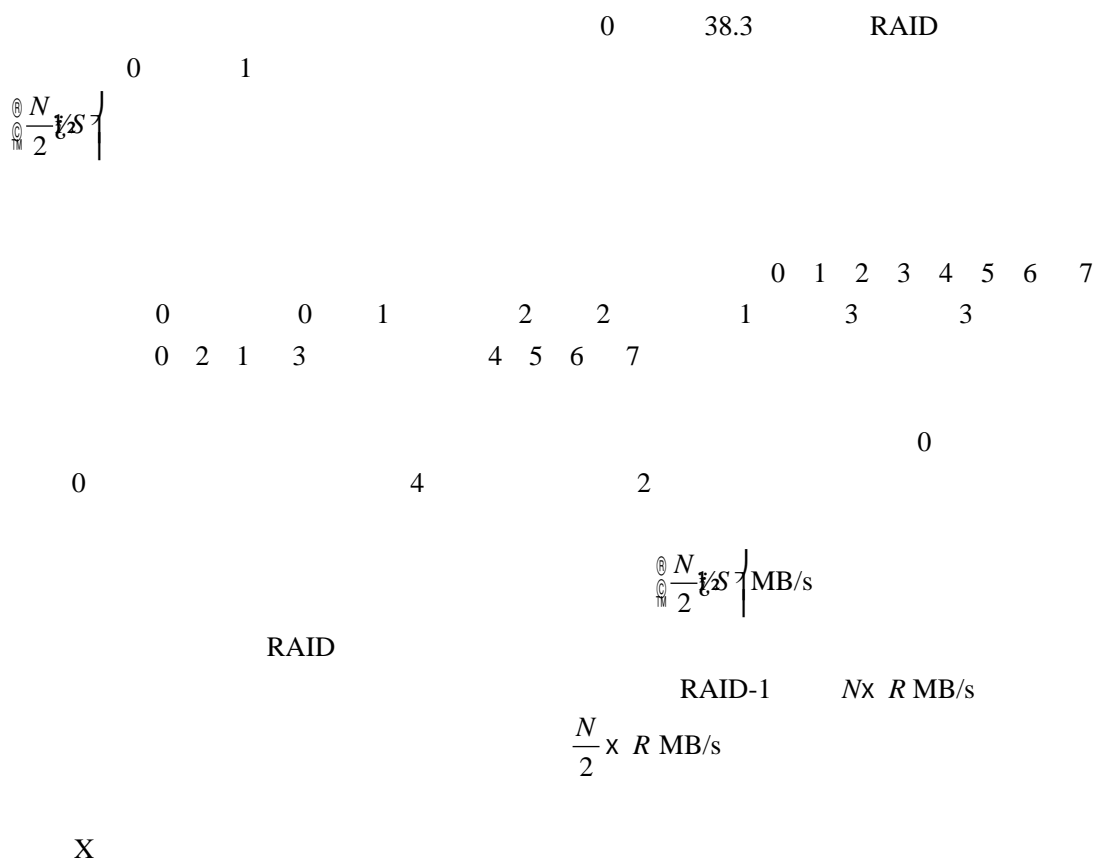
recovery

RAID

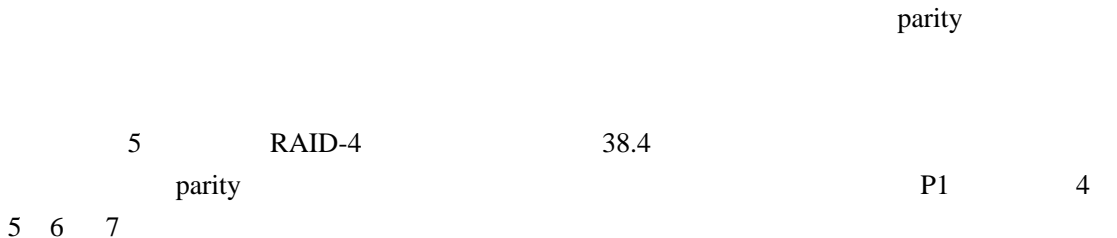
RAID-1

RAID

RAM



%*(D3;6 &



D3;6Ž&				
0	1	2	3	4
0	1	2	3	P0
4	5	6	7	P1
8	9	10	11	P2
12	13	14	15	P3

XOR

1 XOR 0 1 1 38.5

%Z

C0	C1	C2	C3	P
0	0	1	1	XOR(0,0,1,1) = 0
0	1	0	0	XOR(0,1,0,0) = 1

0 0 1 1 1 C2 C3 0 P

1 C1 XOR 1 P

1 RAID

invariant

C2

XOR reconstruct C2

1 C0 0 C1 0 C3 1

P 0 0 0 1 0 XOR 1

1

RAID

4KB XOR XOR XOR 4KB

4

38.6

%Z

JAD

Block0	Block1	Block2	Block3	Parity
00	10	11	10	11
10	01	00	01	10

D3;6Z&

RAID-4 RAID-4 1

RAID N 1

RAID-4 1

N 1 x S MB/s

RAID-4 full-stripe write 0

1 2 3 RAID 38.7

% ž D3;6Ž&

0	1	2	3	4
0	1	2	3	P0
4	5	6	7	P1
8	9	10	11	P2
12	13	14	15	P3

RAID P0 0 1 2 3 XOR
5

RAID-4

RAID-4

$N-1 \times S$ MB/s

38.7

1

$N-1 \times R$ MB/s

RAID-4

1

P0

P0

additive parity

0 2 3

1

RAID

subtractive parity

4

C0	C1	C2	C3	P
0	0	1	1	XOR(0,0,1,1)=0

$C2$
 $C2_{old} = 1$
 $P_{new} = P_{old}$

$C2$
 $P_{old} = 0$
 $C2_{new} = C2_{old}$

P_{new} 1 XOR XOR $P_{old} = 0$ P_{new} 0 $P_{old} = 0$
 $P_{new} = (C_{old} \oplus C_{new}) \oplus P_{old}$ 38.1

4096

8

I/O

RAID

RAID-4

38.8

I/O

RAID-4

RAID-4

38.8

%* ž

& #%

0	1	2	3	4
0	1	2	3	P0
*4	5	6	7	+P1
8	9	10	11	P2
12	*13	14	15	+P3

RAID-4

2

4

13

38.8

0

1

4

13

1

3

+

RAID

small-write problem

I/O

I/O

I/O

RAID-4

R / 2 MB/s

RAID-4

RAID-4

I/O

%* ž D3;6 '

RAID-4

Patterson Gibson Katz

RAID-5

RAID-5

38.9

%* ž+

D3;6 ž'

0	1	2	3	4
0	1	2	3	P0
5	6	7	P1	4
10	11	P2	8	9
15	P3	12	13	14
P4	16	17	18	19

RAID-4

D3;6Ž'

RAID-5

RAID-4

RAID-4

RAID-4

1	4	1	1	10	10
0	2				

$$\frac{N}{4} \times R \text{ MB/s} \quad 4$$

RAID-5

4 I/O

RAID

RAID-5

RAID-4

RAID-4

[HLM94]

RAID-4

%*Ž' D3;6

38.10

RAID

RAID-4/5

%*Ž''

D3;6

	RAID-0	RAID-1	RAID-4	RAID-5
	N	$N/2$	$N-1$	$N-1$
	0	1		
		$N/2$		
	$N \frac{1}{2} S$	$(N/2) \frac{1}{2} S$	$(N-1) \frac{1}{2} S$	$(N-1) \frac{1}{2} S$
	$N \frac{1}{2} S$	$(N/2) \frac{1}{2} S$	$(N-1) \frac{1}{2} S$	$(N-1) \frac{1}{2} S$
	$N \frac{1}{2} R$	$N \frac{1}{2} R$	$(N-1) \frac{1}{2} R$	$N \frac{1}{2} R$
	$N \frac{1}{2} R$	$(N/2) \frac{1}{2} R$	$1/2 \frac{1}{2} R$	$N/4 \frac{1}{2} R$

	RAID-0	RAID-1	RAID-4	RAID-5
	T	T	T	T
	T	T	2T	2T

38.10

RAID

T

I/O

RAID-5

I/O

RAID-5

%* ž+

D3;6

RAID

RAID

236

[C+04]

RAID

hot

spare

latent sector error

block corruption [B+08]

RAID

RAID

[DAA05]

%* ž#'

RAID RAID

RAID

RAID

RAID-5

RAID

[B+08] An Analysis of Data Corruption in the Storage Stack

Lakshmi N. Bairavasundaram, Garth R. Goodson, Bianca Schroeder, Andrea C. Arpaci-Dusseau, Remzi H. Arpaci-Dusseau

FAST 2008, San Jose, CA, February 2008

[BJ88] Disk Shadowing

D. Bitton and J. Gray VLDB1988

[CL95] Striping in a RAID level 5 disk array Peter M. Chen, Edward K. Lee

SIGMETRICS 1995

RAID-5

[C+04] Row-Diagonal Parity for Double Disk Failure Correction

P. Corbett, B. English, A. Goel, T. Grcanac, S. Kleiman, J. Leong, S. Sankar FAST 2004, February 2004

RAID

[DAA05] Journal-guided Resynchronization for Software RAID Timothy E. Denehy, A. Arpaci-Dusseau, R.

Arpaci-Dusseau

FAST 2005

RAID

RAID

[HLM94] File System Design for an NFS File Server Appliance Dave Hitz, James Lau, Michael Malcolm

USENIX Winter 1994, San Francisco, California, 1994

WAFL

NetApp

[K86] Synchronized Disk Interleaving

M.Y. Kim.

IEEE Transactions on Computers, Volume C-35: 11, November 1986

RAID

[K88] Small Disk Arrays - The Emerging Approach to High Performance

F. Kurzweil.

Presentation at Spring COMPCON 88, March 1, 1988, San Francisco, California

RAID

[P+88] Redundant Arrays of Inexpensive Disks

D Patterson, G. Gibson, R. Katz. SIGMOD 1988

Patterson Gibson Katz

RAID

RAID

[PB86] Providing Fault Tolerance in Parallel Secondary Storage Systems

A Park and K. Balasubramaniam

Department of Computer Science, Princeton, CS-TR-O57-86, November 1986

RAID

[SG86] Disk Striping

K. Salem and H. Garcia-Molina.

IEEE International Conference on Data Engineering, 1986

RAID

RAID

SIGMOD

[S84] Byzantine Generals in Action: Implementing Fail-Stop Processors

F.B. Schneider.

ACM Transactions on Computer Systems, 2(2):145154, May 1984

RAID

raid.py

RAID

RAID

README

1

RAID

0 1 4 5

RAID-5

left-symmetric

left-asymmetric

2

-C

3

-r

4

-S

8KB 12KB 16KB

RAID

I/O

		-W sequential		RAID-4	RAID-5	I /
O						
5		-t	100	RAID		RAID
	4					
6					RAID	
7			-w 100			
	RAID			100		
8				-W sequential		
RAID					RAID-4	
RAID-5	RAID					