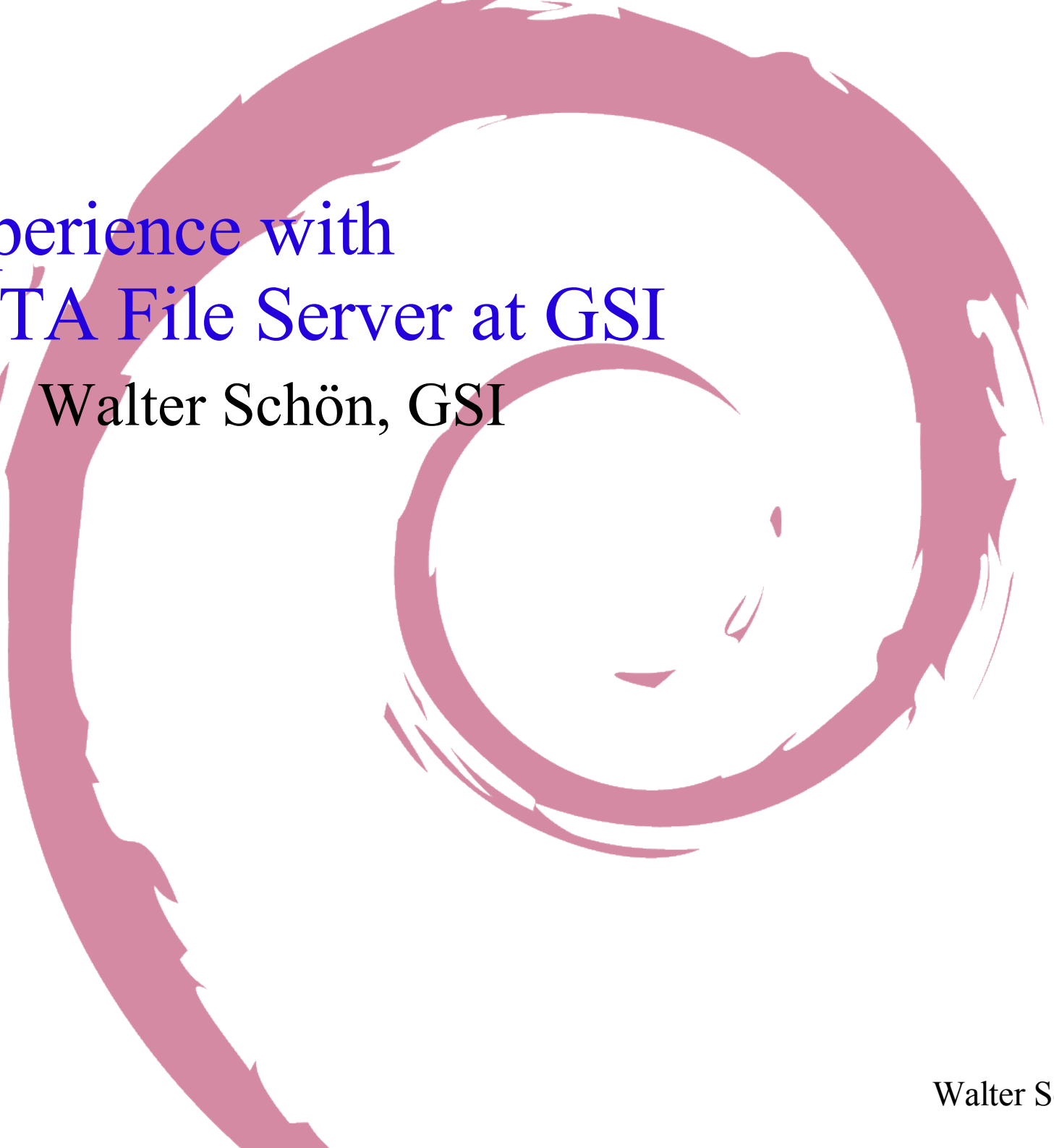
Experience with SATA File Server at GSI

Walter Schön, GSI



Topics

- File Server for Experiment Data
- New Hardware
  - 19¨ box
  - SATA Controller
- Performance Tests RAID 5
  - Kernel 2.6.8
  - Ext3
  - XFS
- Reliability
- Conclusion

File Server for Experiment Data at GSI

- typically units of about 1-2 TB
- large files
- one or few concurrent write processes
- a couple of concurrent read processes

History:

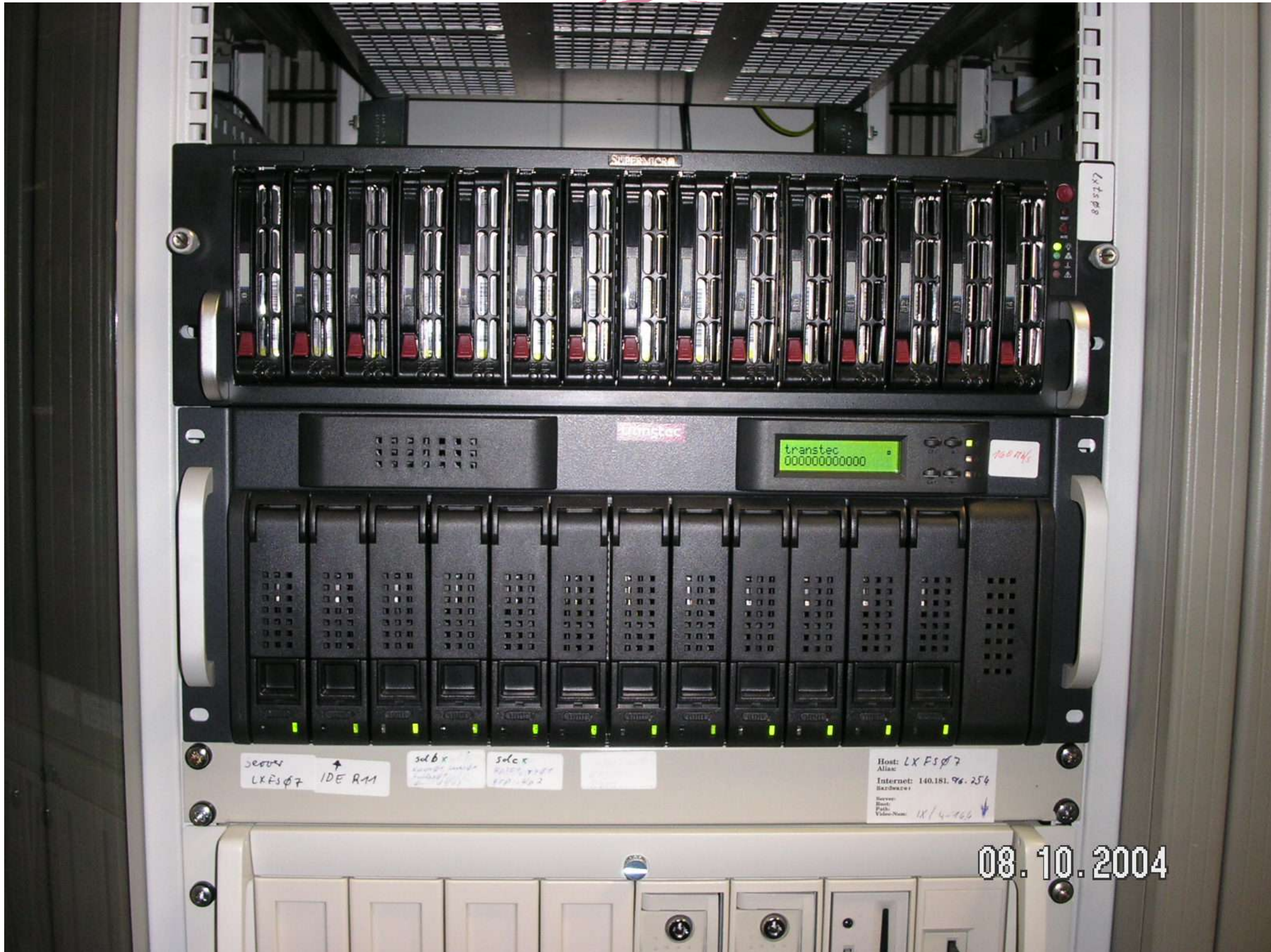
- some years ago: IDE-> SCSI technology
- Computer with scsi controller and
  a couple of attached (scsi cables) raid boxes
- EASY RAID 19¨ frames from TRANSTEC
  - Advantage:
    - cheap as compared to a scsi disks solution
    - >1 TB units possible
  - Disadvantage:
    - ATA disks not really reliable: not specified for 24x7 use
    - RAID controller not very performant:
      - About 24 MB/s for 1 Thread RAID 5
      - Very low performance for concurrent write processes
    - RAID controller not very reliable

New SATA file server

since 9 month testing, 6 month production

3 HU server with 15 ports (hot swap)

- Redundant cooler system, hot swap
- CPU with air system, no cooler on the CPUś
- Triple redundant power supplies, hot swap
- 24 x7 specified disks, 250 GB
- Two independent high performance SATA RAID controller
- With one system disk: 14 *250 = 3,5 TB capacity RAID 0
- With 4 GB Ram about 7 kEuro .....


transtec
000000000000
LXFS07
IDE R11
Host: LXFS07
Alias:
Internet: 140.181.96.254
08.10.2004

SATA Controller:

3Ware Escalade 8506-8, 8 channel

Bug: mixed up data under ¨rare conditions¨

3Ware Escalade 9500S-8, 8 channel

64 MByte cache

expandable to 512 MByte

RAID level: 0,1,5,10 JBOD, 50

Mainboard:

supermicro X5DAL-TG2

SATA controller onboard

64 Bit PCI-X

CPU:

2x 2,66 MHz Xeon

2 GB Ram

Disks:

7200 RPM: Maxtor MaXLine Plus II

specified: 7 x 24, MTBF > 1* 10^6h

seek time < 9 ms

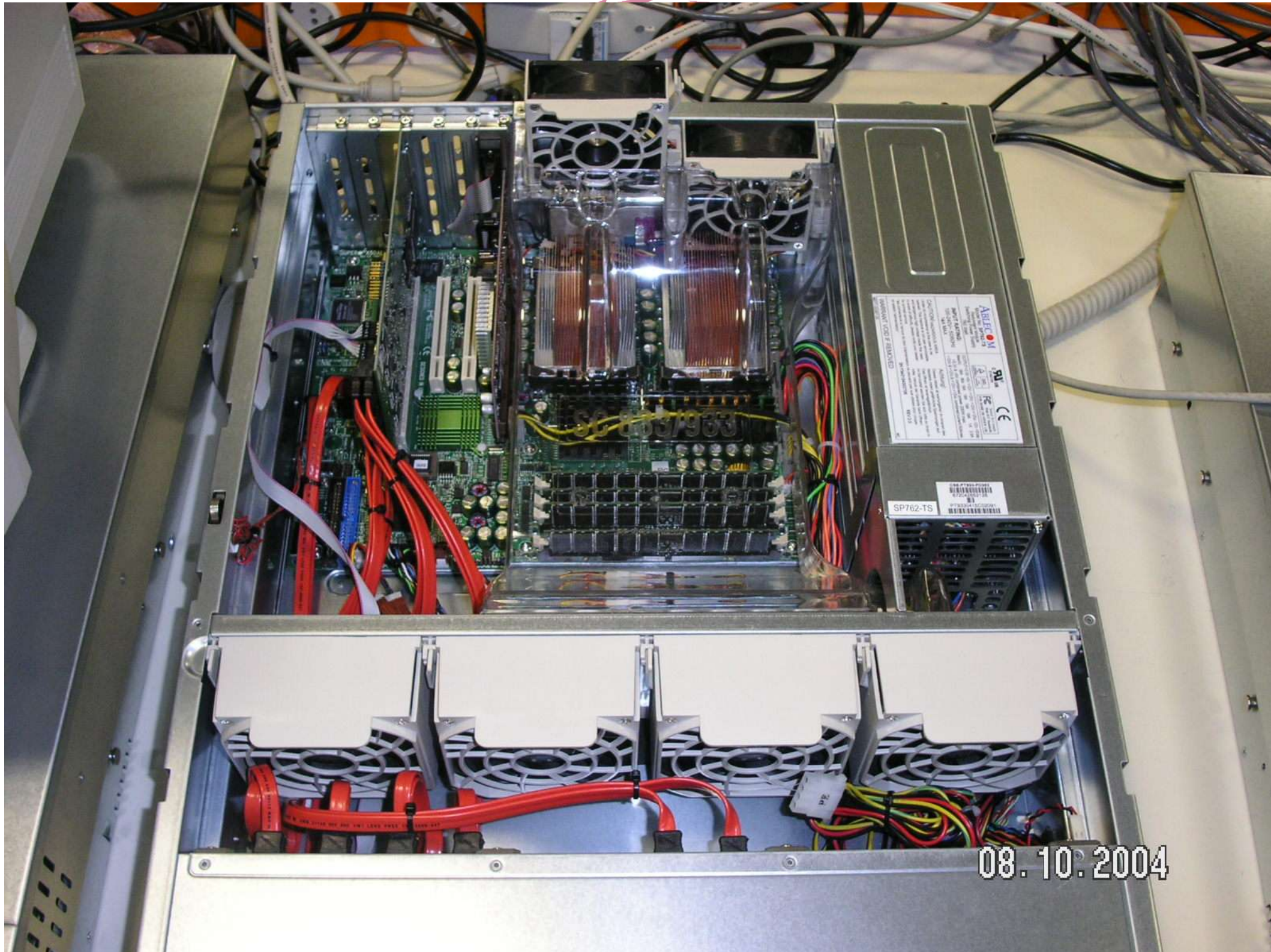
8 MB cache, 250 GB

10000 RPM: WD Raptor

specified 7x24, MTBF> 1,2 *10^6h

seek time 4,5 ms

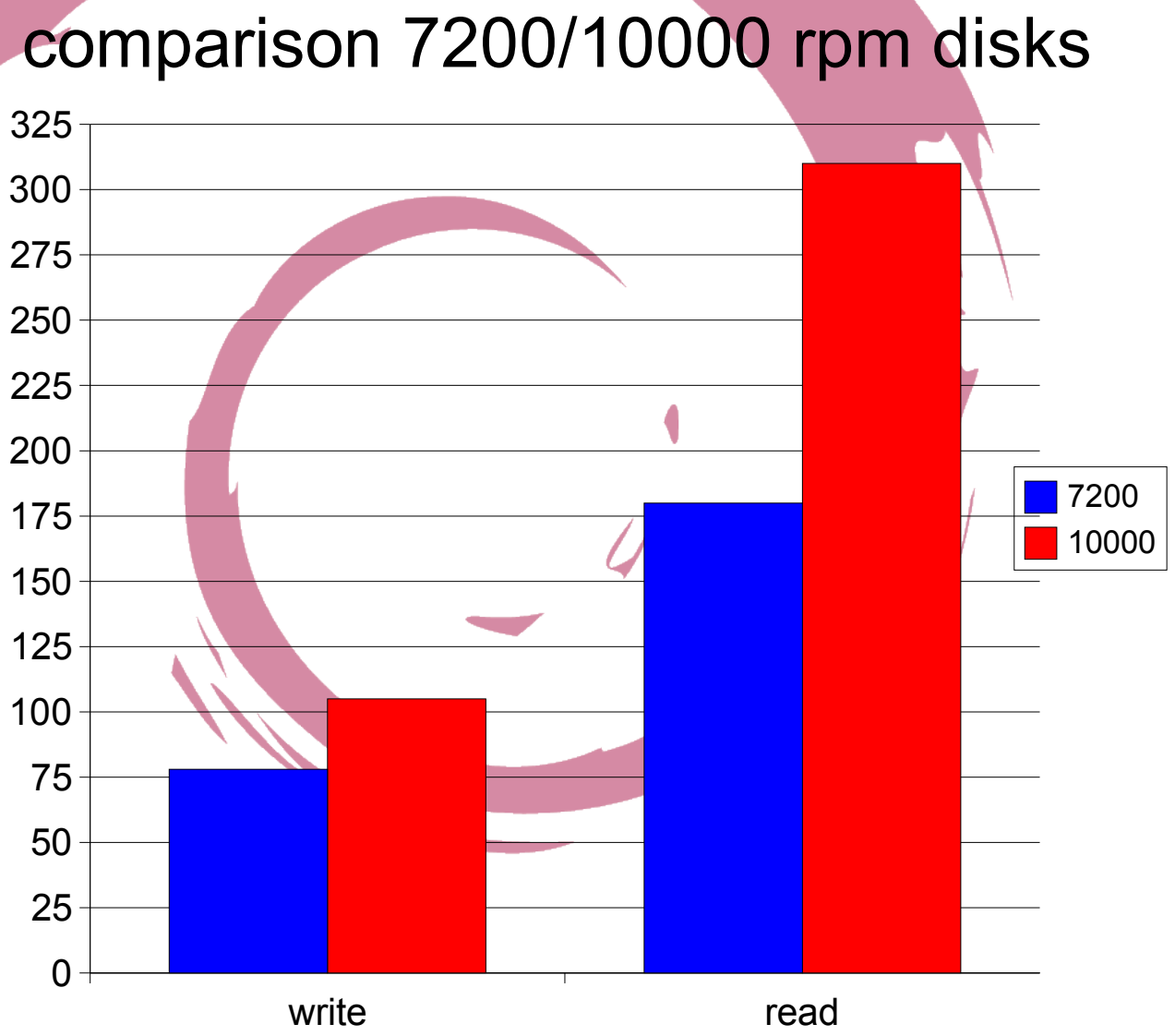
8 MB cache, 73 GB


SP762-TS
08.10.2004


08.10.2004

Performance Tests RAID 5

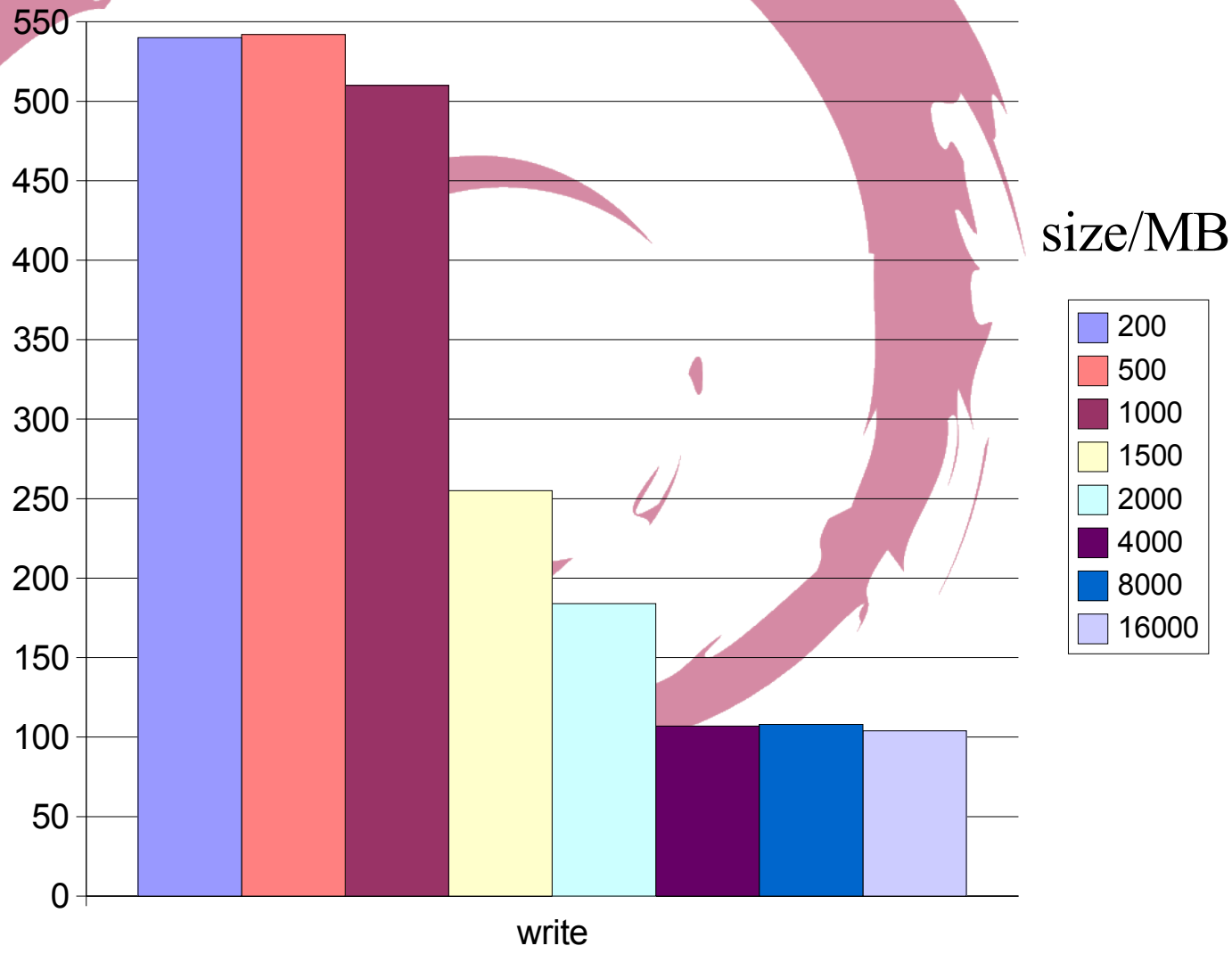
System:
Debian „Sarge“
Kernel 2.6.8, RAID 5
File System: XFS
Read-Ahead Cache:16384



Influence of Caching on Performance

Disk: ¨raptor¨
RAID 5, 8 disks
Kernel: 2.6.8
Read-Ahead Cache:16384

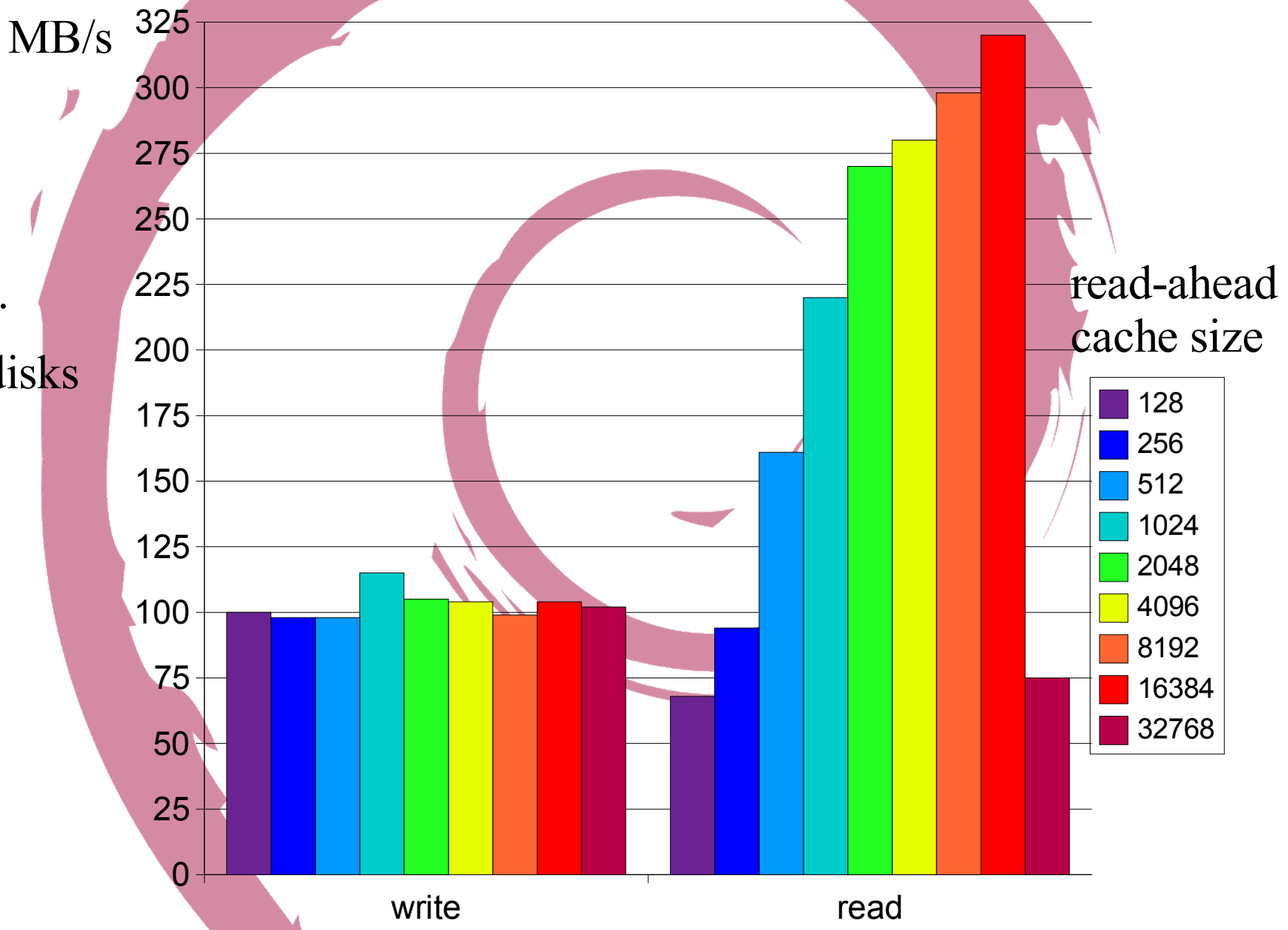
CPU usage: low



=> all test performed with file size > 10 GB >> RAM = 2 GB

Tuning with Read Ahead Cache Size

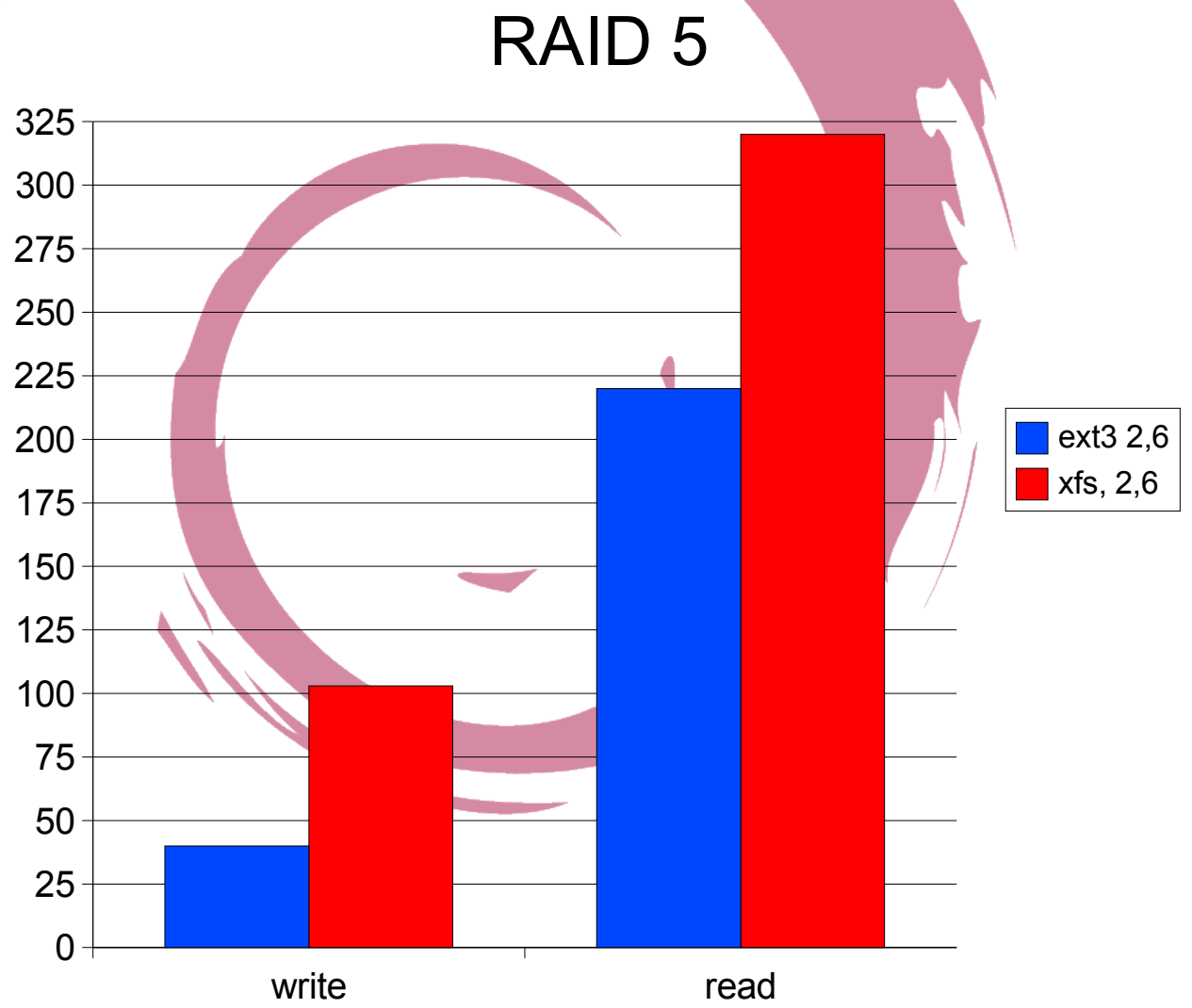
Disk: ¨raptor¨
RAID 5, 8 disks
Kernel: 2.68



Walter Schön, GSI

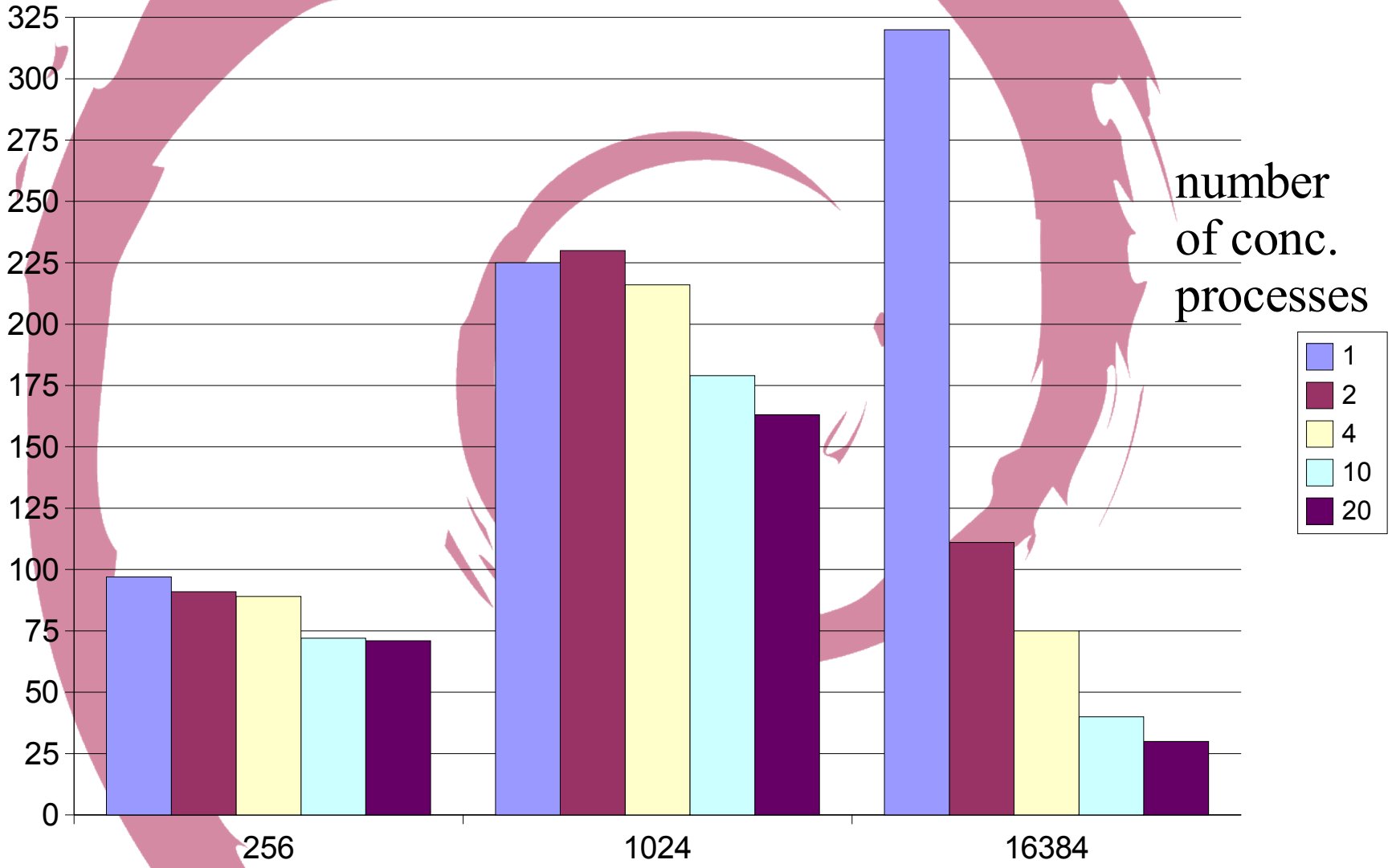
Comparison ext3 with XFS

Disk: ¨raptor¨
RAID 5, 8 disks
Kernel: 2.6.8
Read-Ahead Cache:16384



=> poor write performance
of ext3 !

Total Read Performance as Function of Concurrent Processes and Readahead

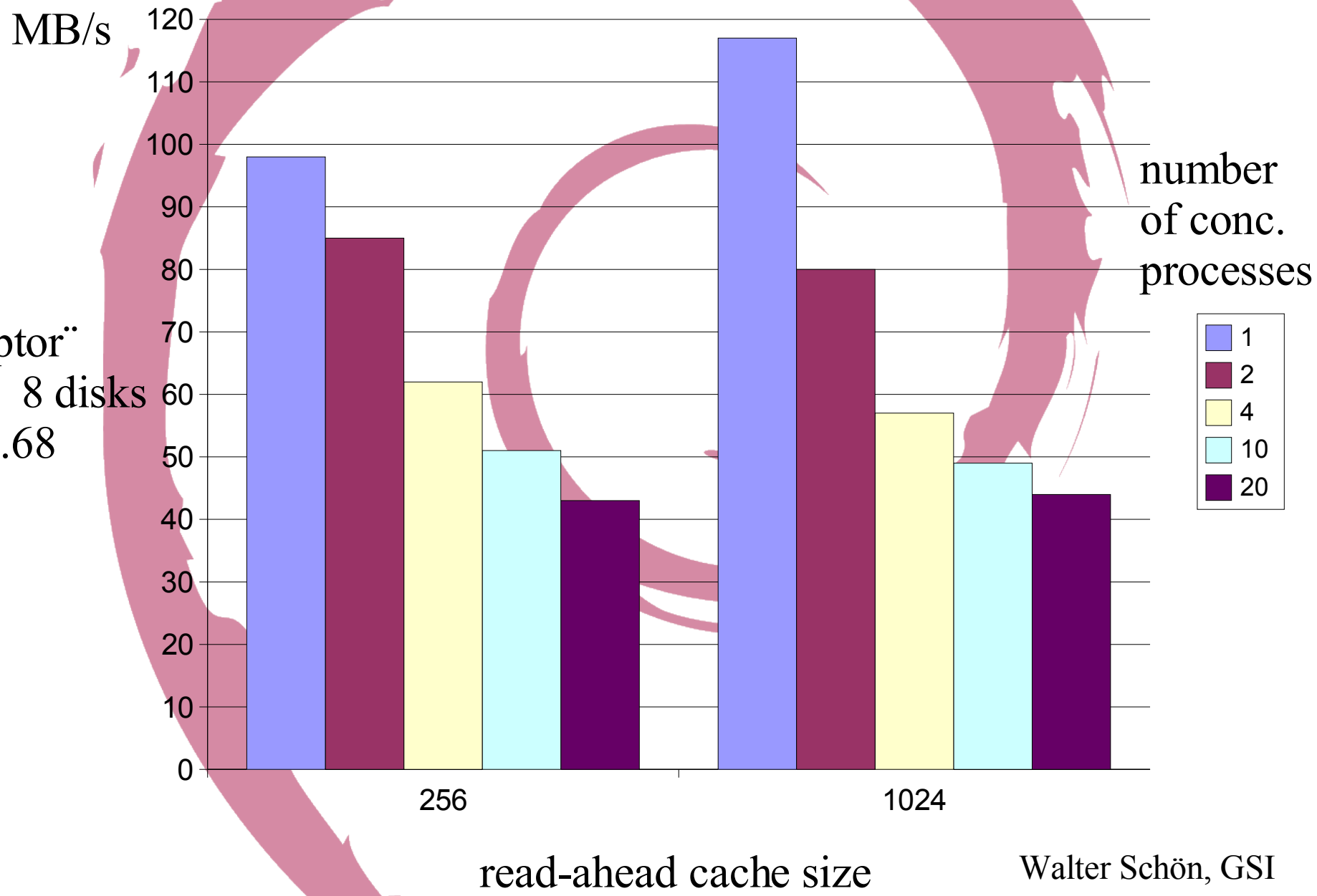


Disk: ¨raptor¨
RAID 5,
8 disks
Kernel: 2.6.8

!read-ahead
16384 not optimal!

Walter Schön, GSI

Total Write Performance as Function of Concurrent Processes



Disk: ¨raptor¨
RAID 5, 8 disks
Kernel: 2.68

Reliability

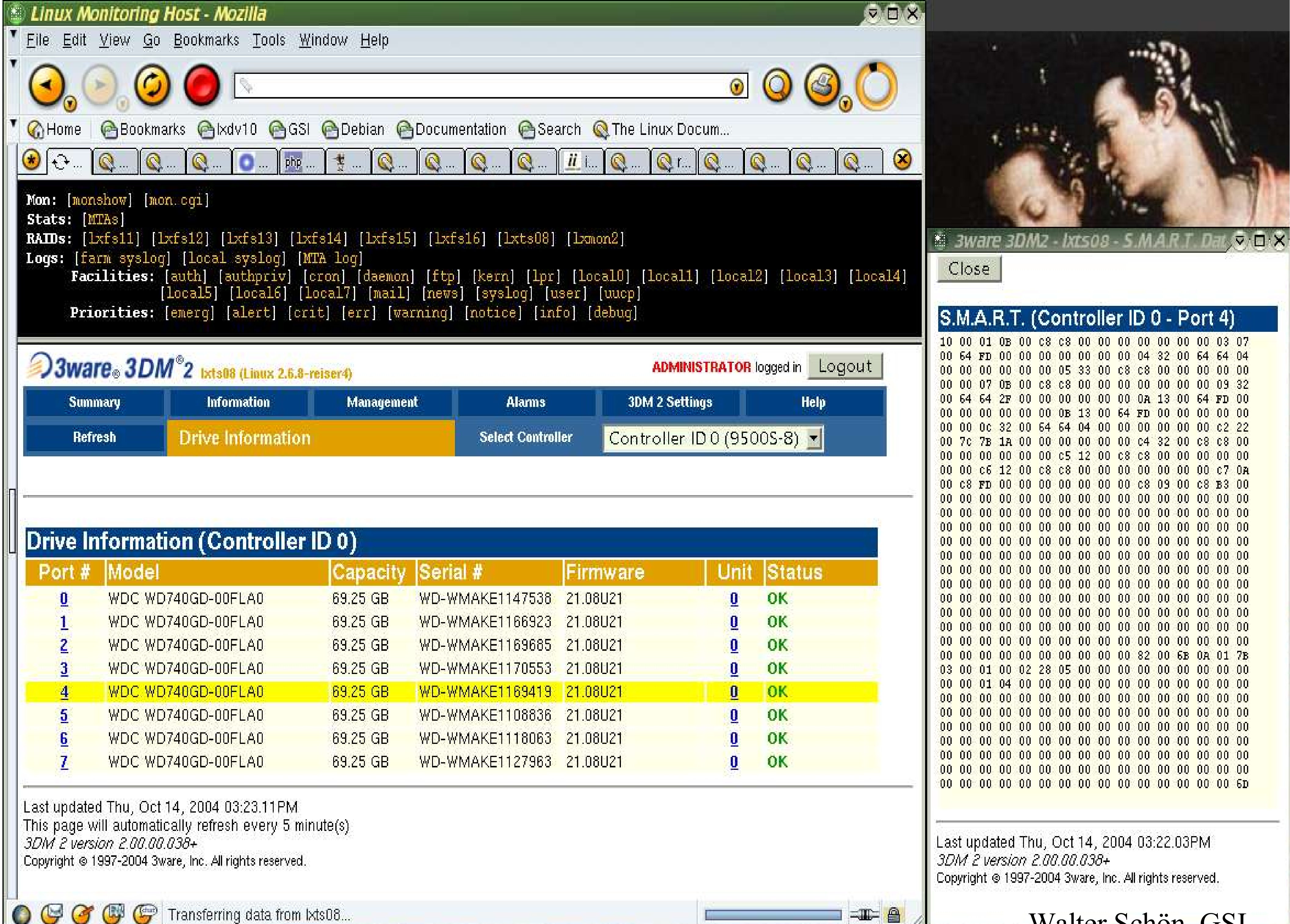
~ 15 servers , 200 disks, 6 months

=> 864000 h heavy load 24 x 7
no disk failure,
no hardware failure at all ....
„experimental MTBF: > 0.9*10^6h“

Reliability: very good!
.... at least up to now ... ;-)

Konfiguration, Monitoring

- At BIOS level
- Command line interface (very powerfull)
- Web browser (very easy)



Linux Monitoring Host - Mozilla

File Edit View Go Bookmarks Tools Window Help

Home | Bookmarks | lxdv10 | GSI | Debian | Documentation | Search | The Linux Docum...

Mon: [monshow] [mon.cgi]
Stats: [MTAs]
RAIDs: [lxfs11] [lxfs12] [lxfs13] [lxfs14] [lxfs15] [lxfs16] [lxts08] [lxmon2]
Logs: [farm syslog] [local syslog] [MTA log]
Facilities: [auth] [authpriv] [cron] [daemon] [ftp] [kern] [lpr] [local0] [local1] [local2] [local3] [local4] [local5] [local6] [local7] [mail] [news] [syslog] [user] [uucp]
Priorities: [emerg] [alert] [crit] [err] [warning] [notice] [info] [debug]

3ware® 3DM®2 lxts08 (Linux 2.6.8-reiser4)

ADMINISTRATOR logged in Logout

Summary | Information | Management | Alarms | 3DM 2 Settings | Help

Refresh | Drive Information | Select Controller | Controller ID 0 (9500S-8)

Drive Information (Controller ID 0)

Port #	Model	Capacity	Serial #	Firmware	Unit	Status
0	WDC WD740GD-00FLA0	69.25 GB	WD-WMAKE1147538	21.08U21	0	OK
1	WDC WD740GD-00FLA0	69.25 GB	WD-WMAKE1166923	21.08U21	0	OK
2	WDC WD740GD-00FLA0	69.25 GB	WD-WMAKE1169665	21.08U21	0	OK
3	WDC WD740GD-00FLA0	69.25 GB	WD-WMAKE1170553	21.08U21	0	OK
4	WDC WD740GD-00FLA0	69.25 GB	WD-WMAKE1169419	21.08U21	0	OK
5	WDC WD740GD-00FLA0	69.25 GB	WD-WMAKE1108836	21.08U21	0	OK
6	WDC WD740GD-00FLA0	69.25 GB	WD-WMAKE1118063	21.08U21	0	OK
7	WDC WD740GD-00FLA0	69.25 GB	WD-WMAKE1127963	21.08U21	0	OK

Last updated Thu, Oct 14, 2004 03:23.11PM
This page will automatically refresh every 5 minute(s)
3DM 2 version 2.00.00.038+

Transferring data from lxts08...

3ware 3DM2 - lxts08 - S.M.A.R.T. Dat

Close

S.M.A.R.T. (Controller ID 0 - Port 4)


```
10 00 01 0B 00 c8 c8 00 00 00 00 00 00 00 03 07
00 64 FD 00 00 00 00 00 00 00 04 32 00 64 64 04
00 00 00 00 00 00 05 33 00 c8 c8 00 00 00 00 00
00 00 07 0B 00 c8 c8 00 00 00 00 00 00 00 09 32
00 64 64 2F 00 00 00 00 00 00 0A 13 00 64 FD 00
00 00 00 00 00 00 0B 13 00 64 FD 00 00 00 00 00
00 00 0c 32 00 64 64 04 00 00 00 00 00 00 c2 22
00 7c 7B 1A 00 00 00 00 00 00 c4 32 00 c8 c8 00
00 00 00 00 00 00 c5 12 00 c8 c8 00 00 00 00 00
00 00 c6 12 00 c8 c8 00 00 00 00 00 00 00 c7 0A
00 c8 FD 00 00 00 00 00 00 00 c8 09 00 c8 B3 00
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00 00 00 00 00 00 00 00 00 00 82 00 6B 0A 01 7B
03 00 01 00 02 28 05 00 00 00 00 00 00 00 00 00
00 00 01 04 00 00 00 00 00 00 00 00 00 00 00 00
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 6D
```


Last updated Thu, Oct 14, 2004 03:22.03PM
3DM 2 version 2.00.00.038+
Copyright © 1997-2004 3ware, Inc. All rights reserved.

SATA file servers - Conclusion:

- very good reliability
- good performance in combination with
  - high performance controllers
  - xfs file system
  - Kernel 2.6 + „tuning/optimisation“

however:

- performance decreases for concurrent access
  - => needs optimisation