Case Study

Jan Weaver
Hewlett Packard

HP WORLD 2003
Solutions and Technology Conference & Expo

Case Study

Problem:

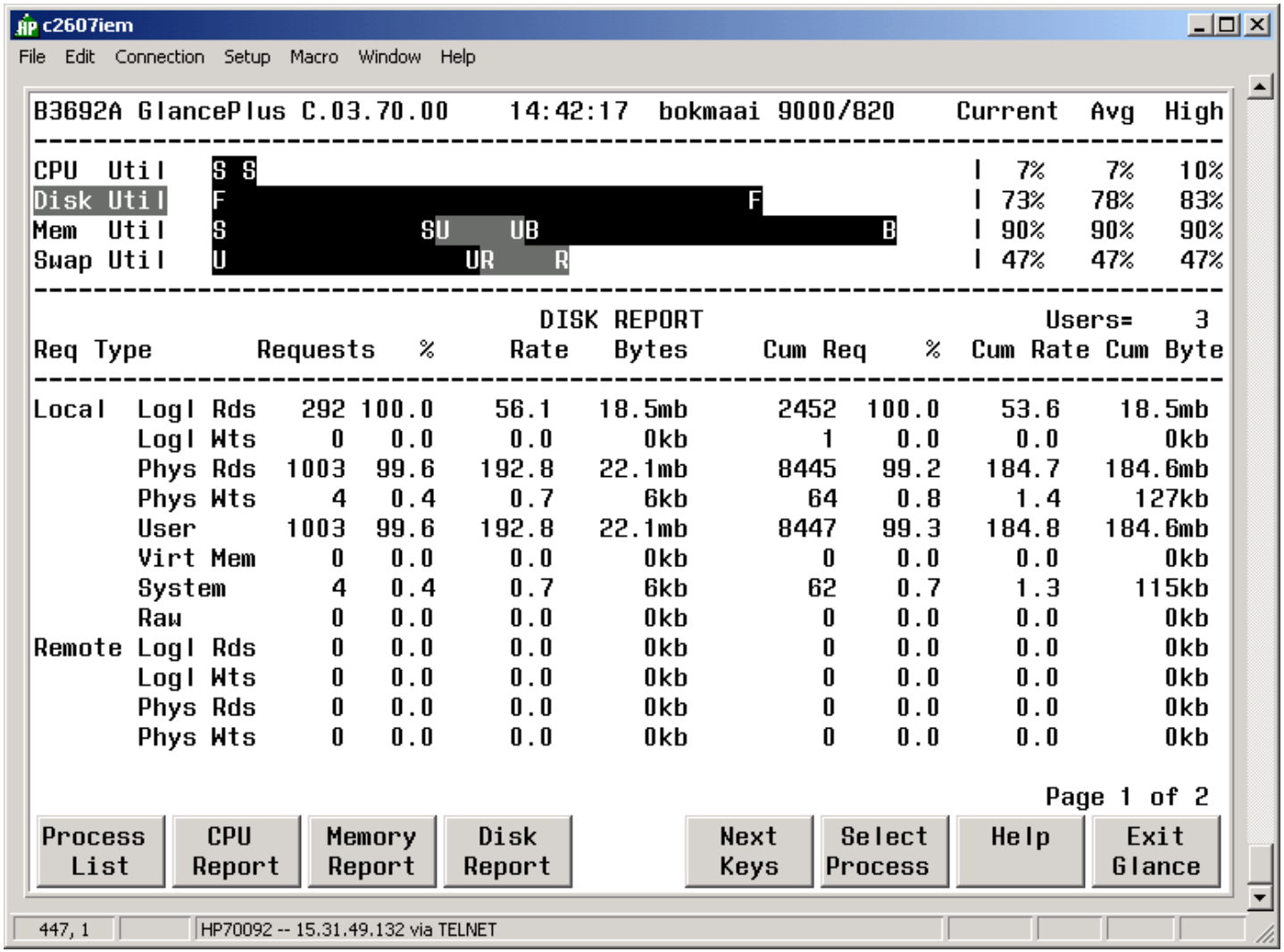
Customer upgraded from JFS 3.1 to JFS 3.3 or upgraded from HPUX 11.0 to HPUX 11i and he now has performance problems with his application and/or system.

He notices an increase in the disk activity.

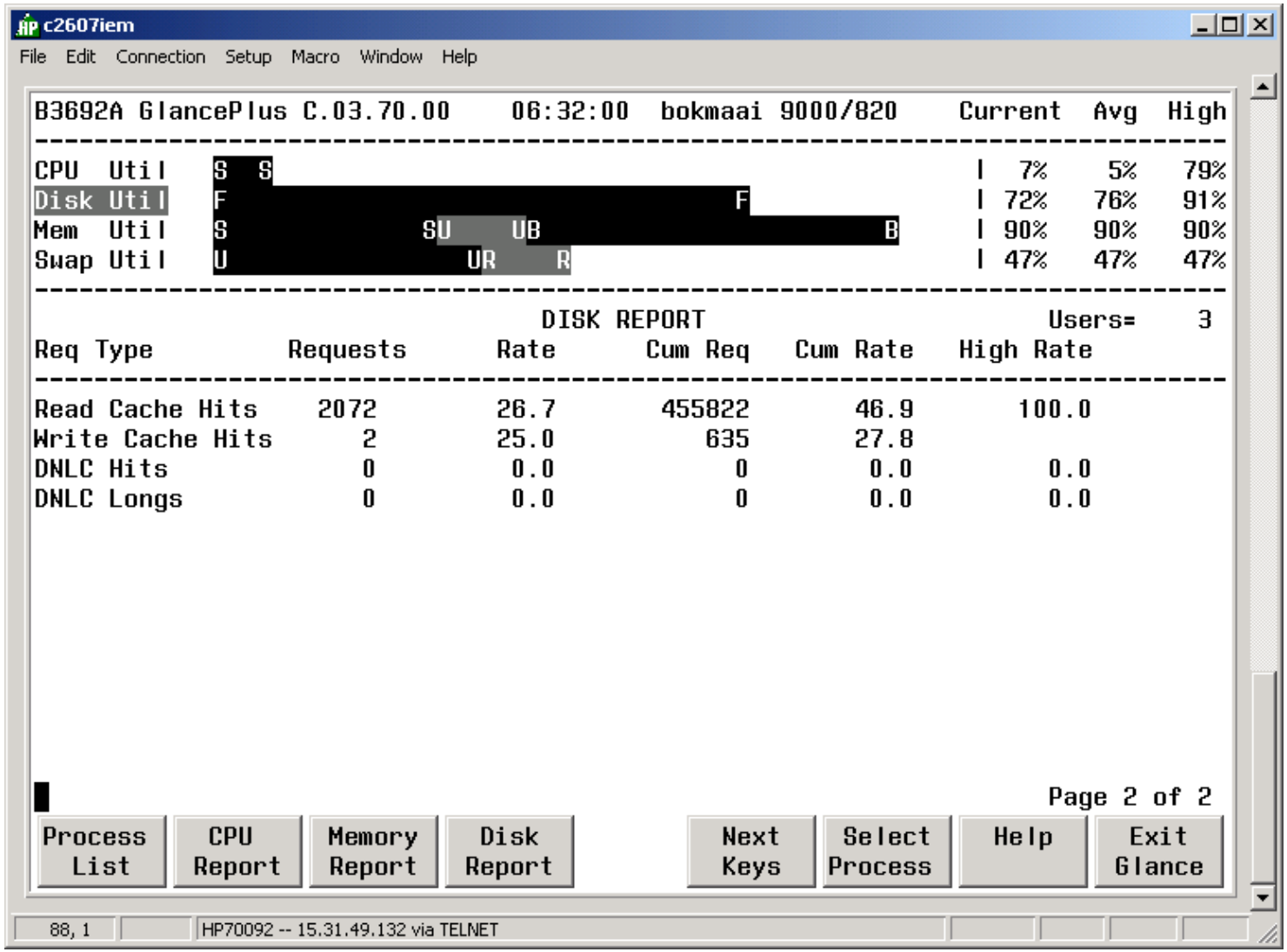
Case Study

Glance shows a high level of physical disk activity and a low buffer cache hit rate.

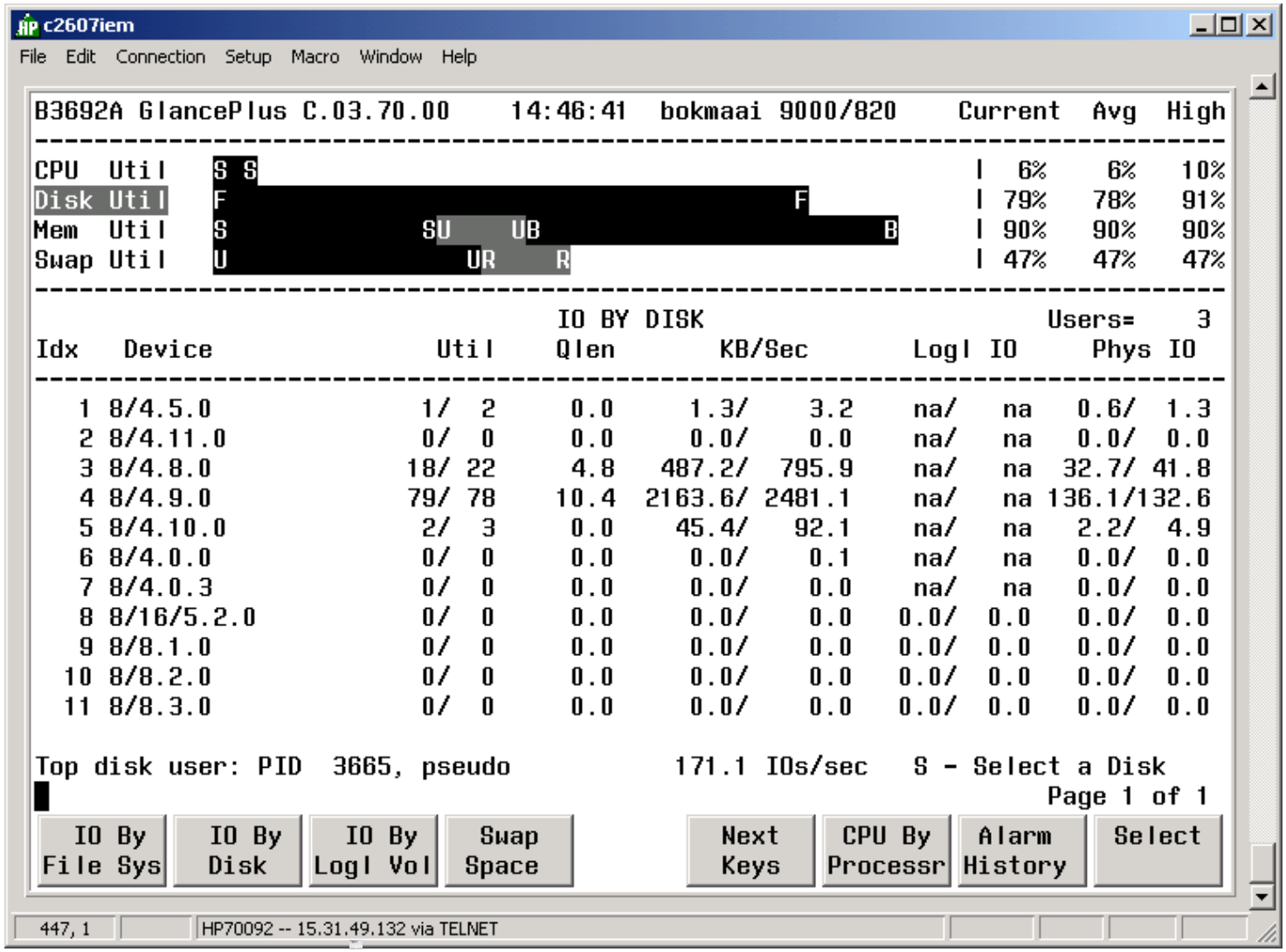
Case Study


c2607iem
File Edit Connection Setup Macro Window Help
B3692A GlancePlus C.03.70.00 14:42:17 bokmaai 9000/820 Current Avg High
CPU Util | 7% 7% 10%
Disk Util | 73% 78% 83%
Mem Util | 90% 90% 90%
Swap Util | 47% 47% 47%
DISK REPORT Users= 3
Req Type Requests % Rate Bytes Cum Req % Cum Rate Cum Byte
Local Logl Rds 292 100.0 56.1 18.5mb 2452 100.0 53.6 18.5mb
Logl Wts 0 0.0 0.0 0kb 1 0.0 0.0 0kb
Phys Rds 1003 99.6 192.8 22.1mb 8445 99.2 184.7 184.6mb
Phys Wts 4 0.4 0.7 6kb 64 0.8 1.4 127kb
User 1003 99.6 192.8 22.1mb 8447 99.3 184.8 184.6mb
Virt Mem 0 0.0 0.0 0kb 0 0.0 0.0 0kb
System 4 0.4 0.7 6kb 62 0.7 1.3 115kb
Raw 0 0.0 0.0 0kb 0 0.0 0.0 0kb
Remote Logl Rds 0 0.0 0.0 0kb 0 0.0 0.0 0kb
Logl Wts 0 0.0 0.0 0kb 0 0.0 0.0 0kb
Phys Rds 0 0.0 0.0 0kb 0 0.0 0.0 0kb
Phys Wts 0 0.0 0.0 0kb 0 0.0 0.0 0kb
Page 1 of 2
Process List
CPU Report
Memory Report
Disk Report
Next Keys
Select Process
Help
Exit Glance
447, 1
HP70092 -- 15.31.49.132 via TELNET

Case Study


c2607iem
File Edit Connection Setup Macro Window Help
B3692A GlancePlus C.03.70.00 06:32:00 bokmaai 9000/820 Current Avg High
CPU Util | 7% 5% 79%
Disk Util | 72% 76% 91%
Mem Util | 90% 90% 90%
Swap Util | 47% 47% 47%
DISK REPORT Users= 3
Req Type Requests Rate Cum Req Cum Rate High Rate
Read Cache Hits 2072 26.7 455822 46.9 100.0
Write Cache Hits 2 25.0 635 27.8
DNLC Hits 0 0.0 0 0.0 0.0
DNLC Longs 0 0.0 0 0.0 0.0
Page 2 of 2
Process List
CPU Report
Memory Report
Disk Report
Next Keys
Select Process
Help
Exit Glance
88, 1
HP70092 -- 15.31.49.132 via TELNET

Case Study


c2607iem
File Edit Connection Setup Macro Window Help
B3692A GlancePlus C.03.70.00 14:46:41 bokmaai 9000/820 Current Avg High
CPU Util | 6% 6% 10%
Disk Util | 79% 78% 91%
Mem Util | 90% 90% 90%
Swap Util | 47% 47% 47%
IO BY DISK Users= 3
Idx Device Util Qlen KB/Sec Logl IO Phys IO
1 8/4.5.0 1/ 2 0.0 1.3/ 3.2 na/ na 0.6/ 1.3
2 8/4.11.0 0/ 0 0.0 0.0/ 0.0 na/ na 0.0/ 0.0
3 8/4.8.0 18/ 22 4.8 487.2/ 795.9 na/ na 32.7/ 41.8
4 8/4.9.0 79/ 78 10.4 2163.6/ 2481.1 na/ na 136.1/132.6
5 8/4.10.0 2/ 3 0.0 45.4/ 92.1 na/ na 2.2/ 4.9
6 8/4.0.0 0/ 0 0.0 0.0/ 0.1 na/ na 0.0/ 0.0
7 8/4.0.3 0/ 0 0.0 0.0/ 0.0 na/ na 0.0/ 0.0
8 8/16/5.2.0 0/ 0 0.0 0.0/ 0.0 0.0/ 0.0 0.0/ 0.0
9 8/8.1.0 0/ 0 0.0 0.0/ 0.0 0.0/ 0.0 0.0/ 0.0
10 8/8.2.0 0/ 0 0.0 0.0/ 0.0 0.0/ 0.0 0.0/ 0.0
11 8/8.3.0 0/ 0 0.0 0.0/ 0.0 0.0/ 0.0 0.0/ 0.0
Top disk user: PID 3665, pseudo 171.1 IOs/sec S - Select a Disk
Page 1 of 1
IO By File Sys
IO By Disk
IO By Logl Vol
Swap Space
Next Keys
CPU By Processr
Alarm History
Select
447, 1
HP70092 -- 15.31.49.132 via TELNET

Case Study

We need to focus on the IO – who is doing it and why.

Kitrace can be used to look at the individual IO's and the system calls made by the process

In this case kitrace shows mostly random IO – lseek, read,lseek,read

However, occasionally we see sequential IO – lseek, read, read

Case Study

```
pid=3665 read ret1=8192
pid=3665 lseek ret1=365633536
pid=3665 read ret1=8192
pid=3665 lseek ret1=284893184
pid=3665 read ret1=8192
pid=3665 lseek ret1=466845696
pid=3665 read ret1=8192
pid=3665 lseek ret1=262332416
pid=3665 read ret1=8192
pid=3665 lseek ret1=118677504
pid=3665 read ret1=8192
pid=3665 read ret1=8192
pid=3665 lseek ret1=204439552
pid=3665 read ret1=8192
pid=3665 lseek ret1=229343232
```

Case Study

When the sequential reads occur, we see lots of physical IO being launched to the disks.

Case Study


```
pid=3665 lseek ret1=118677504
ENQUEUE pid=3665 wr=read len=8192
ENQUEUE pid=3665 wr=read len=8192
pid=3665 read ret1=8192
ENQUEUE pid=3665 wr=read len=8192
ENQUEUE pid=3665 wr=read len=24576
ENQUEUE pid=3665 wr=read len=16384
ENQUEUE pid=3665 wr=read len=8192
ENQUEUE pid=3665 wr=read len=32768
ENQUEUE pid=3665 wr=read len=8192
ENQUEUE pid=3665 wr=read len=32768
ENQUEUE pid=3665 wr=read len=8192
.
.
.
ENQUEUE pid=3665 wr=read len=24576
ENQUEUE pid=3665 wr=read len=65536
ENQUEUE pid=3665 wr=read len=40960
ENQUEUE pid=3665 wr=read len=8192
ENQUEUE pid=3665 wr=read len=8192
ENQUEUE pid=3665 wr=read len=57344
ENQUEUE pid=3665 wr=read len=32768
ENQUEUE pid=3665 wr=read len=8192
ENQUEUE pid=3665 wr=read len=8192
pid=3665 read ret1=8192
```

Case Study

System is doing read ahead when the sequential IO is detected.

Read ahead is more aggressive on JFS 3.3 than it was on JFS 3.1.

It is controlled by the vxtunefs parameters read_nstream and read_pref_io.

Case Study

```
>vxtunefs /data
Filesystem i/o parameters for /data
read_pref_io = 65536
read_nstream = 10
read_unit_io = 65536
write_pref_io = 65536
write_nstream = 1
write_unit_io = 65536
pref_strength = 10
buf_breakup_size = 131072
discovered_direct_iosz = 262144
max_direct_iosz = 655360
default_indir_size = 8192
qio_cache_enable = 0
max_diskq = 1048576
initial_extent_size = 4
max_seqio_extent_size = 2048
max_buf_data_size = 8192
```

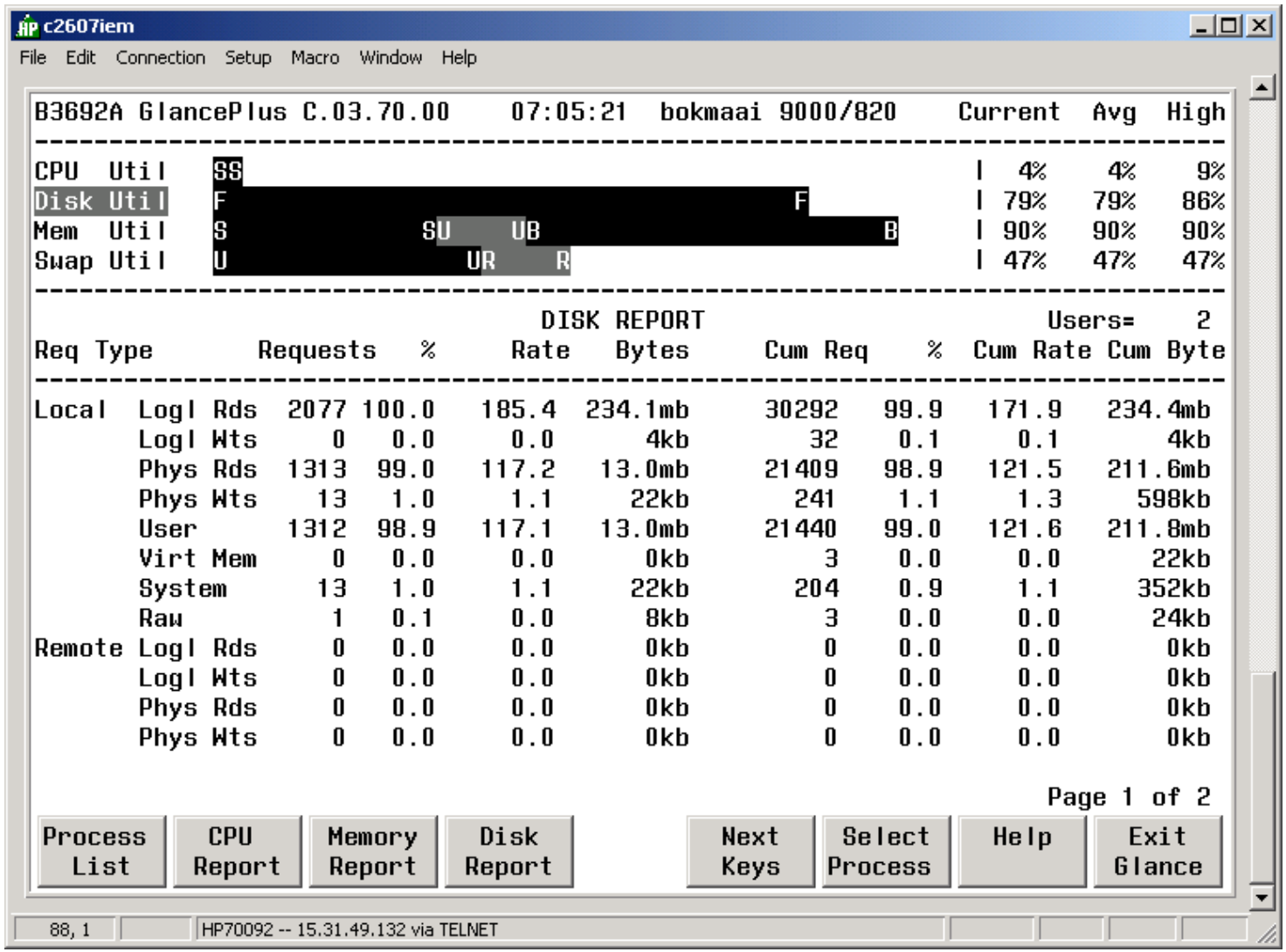
Case Study

Due to the generally random IO of the application the read ahead was unnecessary and in fact was likely harmful.

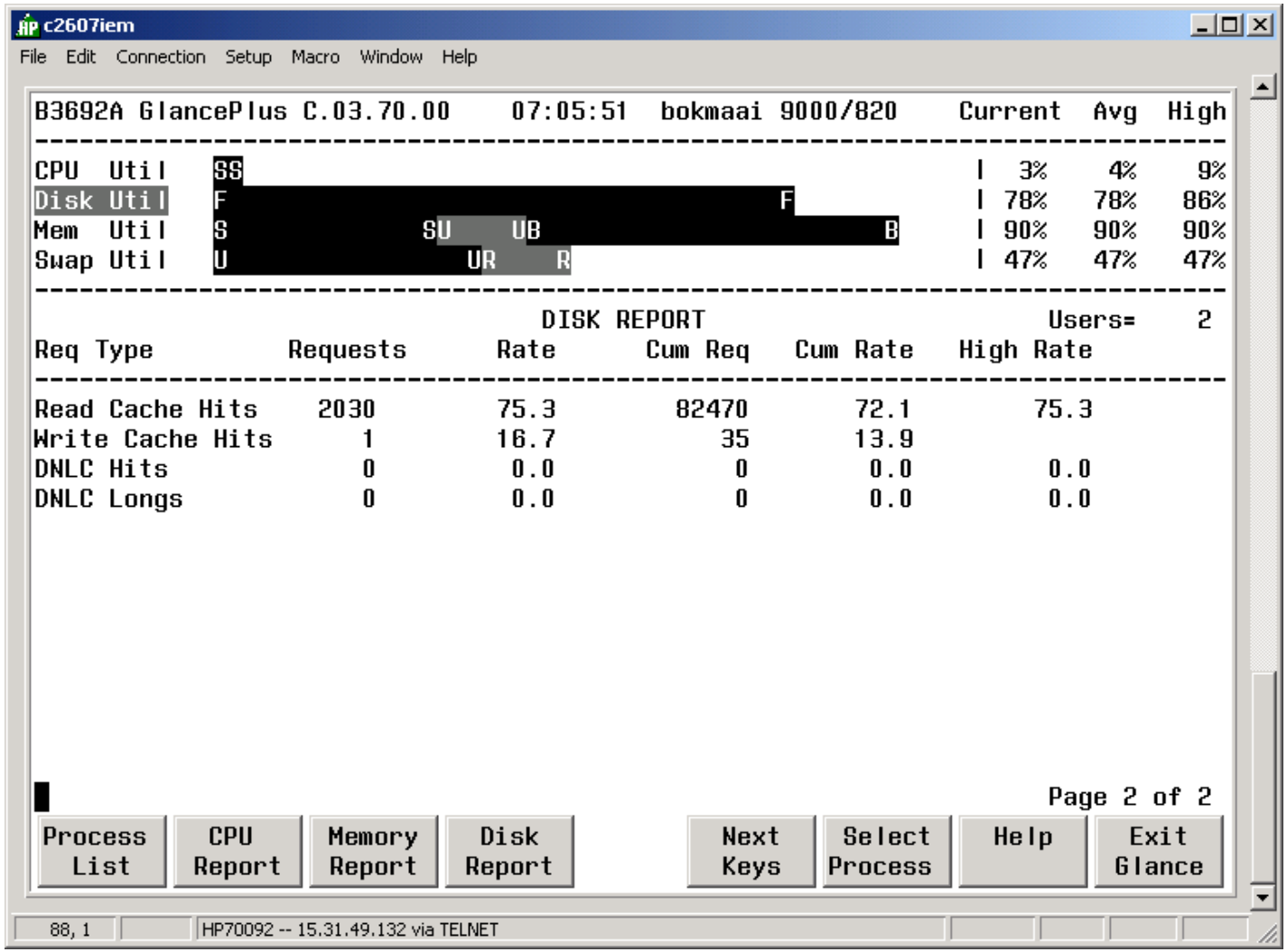
Filesystem parameters read_nstream and/or read_pref_io can be tuned to reduce the amount of read ahead that is performed.

Note that the application could also be changed to include code to advise the filesystem that the IO is random.

Case Study


c2607iem
File Edit Connection Setup Macro Window Help
B3692A GlancePlus C.03.70.00 07:05:21 bokmaai 9000/820 Current Avg High
CPU Util SS | 4% 4% 9%
Disk Util F F | 79% 79% 86%
Mem Util S SU UB B | 90% 90% 90%
Swap Util U UR R | 47% 47% 47%
DISK REPORT Users= 2
Req Type Requests % Rate Bytes Cum Req % Cum Rate Cum Byte
Local Logl Rds 2077 100.0 185.4 234.1mb 30292 99.9 171.9 234.4mb
Logl Wts 0 0.0 0.0 4kb 32 0.1 0.1 4kb
Phys Rds 1313 99.0 117.2 13.0mb 21409 98.9 121.5 211.6mb
Phys Wts 13 1.0 1.1 22kb 241 1.1 1.3 598kb
User 1312 98.9 117.1 13.0mb 21440 99.0 121.6 211.8mb
Virt Mem 0 0.0 0.0 0kb 3 0.0 0.0 22kb
System 13 1.0 1.1 22kb 204 0.9 1.1 352kb
Raw 1 0.1 0.0 8kb 3 0.0 0.0 24kb
Remote Logl Rds 0 0.0 0.0 0kb 0 0.0 0.0 0kb
Logl Wts 0 0.0 0.0 0kb 0 0.0 0.0 0kb
Phys Rds 0 0.0 0.0 0kb 0 0.0 0.0 0kb
Phys Wts 0 0.0 0.0 0kb 0 0.0 0.0 0kb
Page 1 of 2
Process List
CPU Report
Memory Report
Disk Report
Next Keys
Select Process
Help
Exit Glance
88, 1
HP70092 -- 15.31.49.132 via TELNET

Case Study

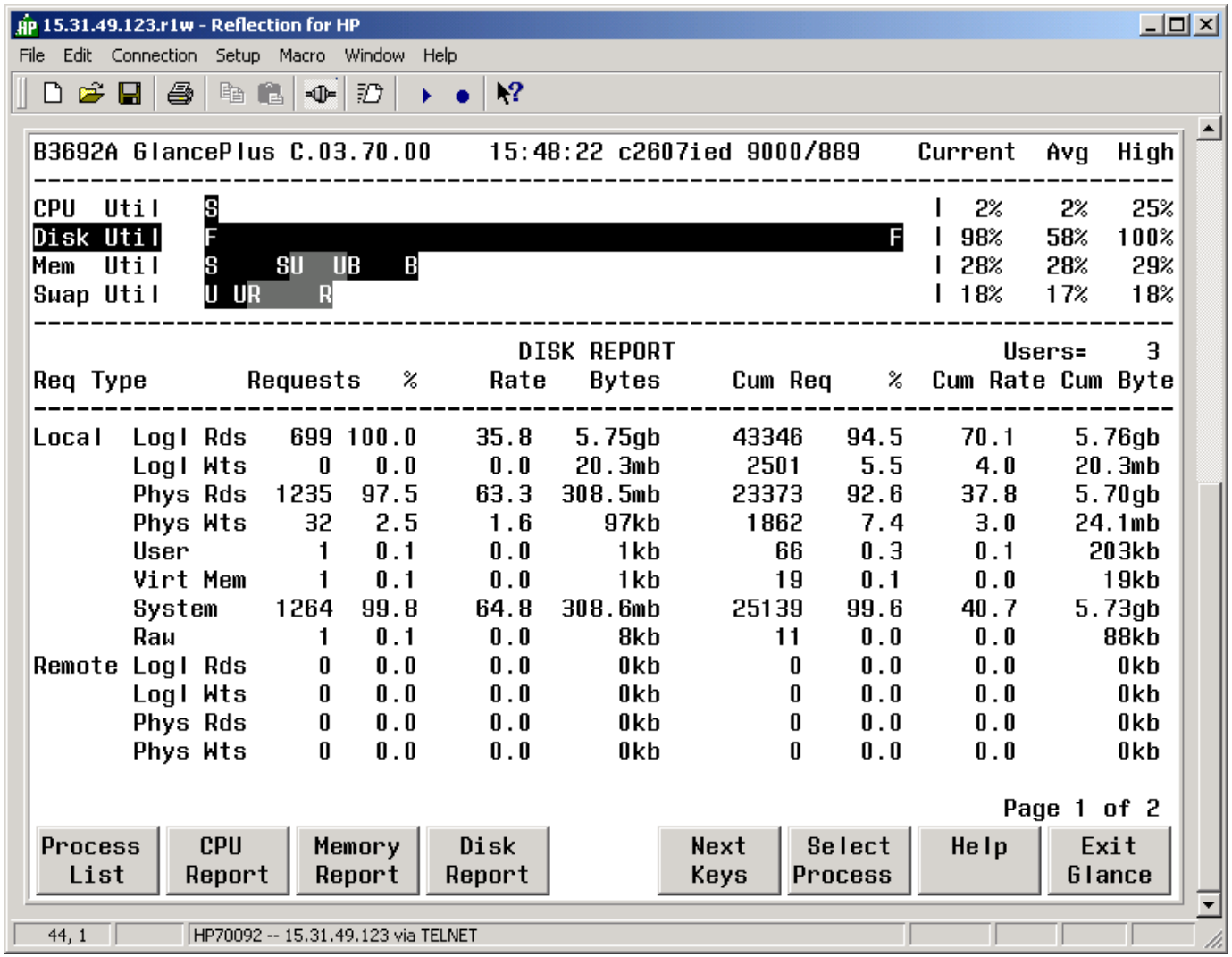

c2607iem
File Edit Connection Setup Macro Window Help
B3692A GlancePlus C.03.70.00 07:05:51 bokmaai 9000/820 Current Avg High
CPU Util SS | 3% 4% 9%
Disk Util F F | 78% 78% 86%
Mem Util S SU UB B | 90% 90% 90%
Swap Util U UR R | 47% 47% 47%
DISK REPORT Users= 2
Req Type Requests Rate Cum Req Cum Rate High Rate
Read Cache Hits 2030 75.3 82470 72.1 75.3
Write Cache Hits 1 16.7 35 13.9
DNLC Hits 0 0.0 0 0.0 0.0
DNLC Longs 0 0.0 0 0.0 0.0
Page 2 of 2
Process List
CPU Report
Memory Report
Disk Report
Next Keys
Select Process
Help
Exit Glance
88, 1
HP70092 -- 15.31.49.132 via TELNET

Case Study

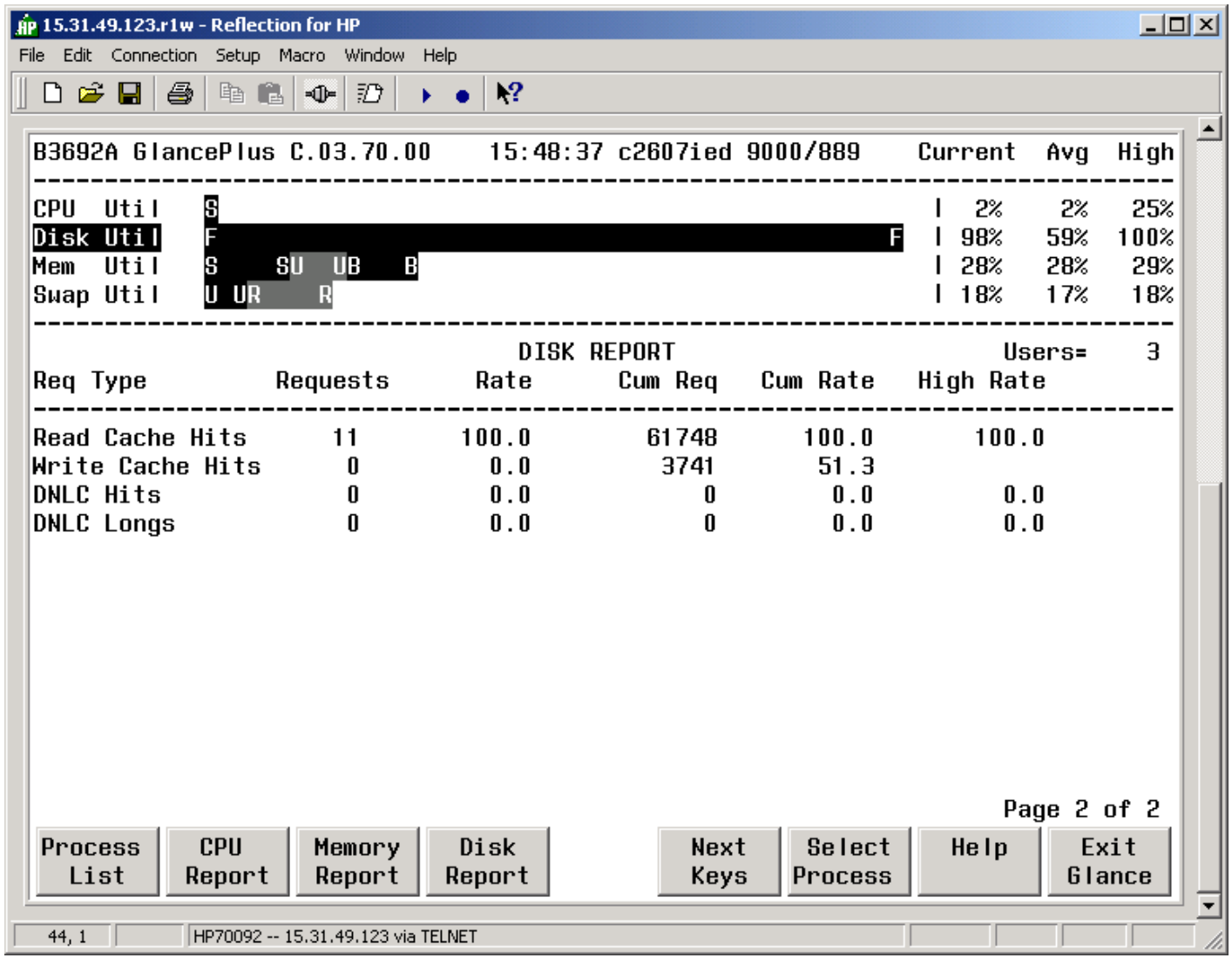
Customer sees a similar performance slowdown after adding online JFS.

Applications run slower and there is more physical IO than seen previously.

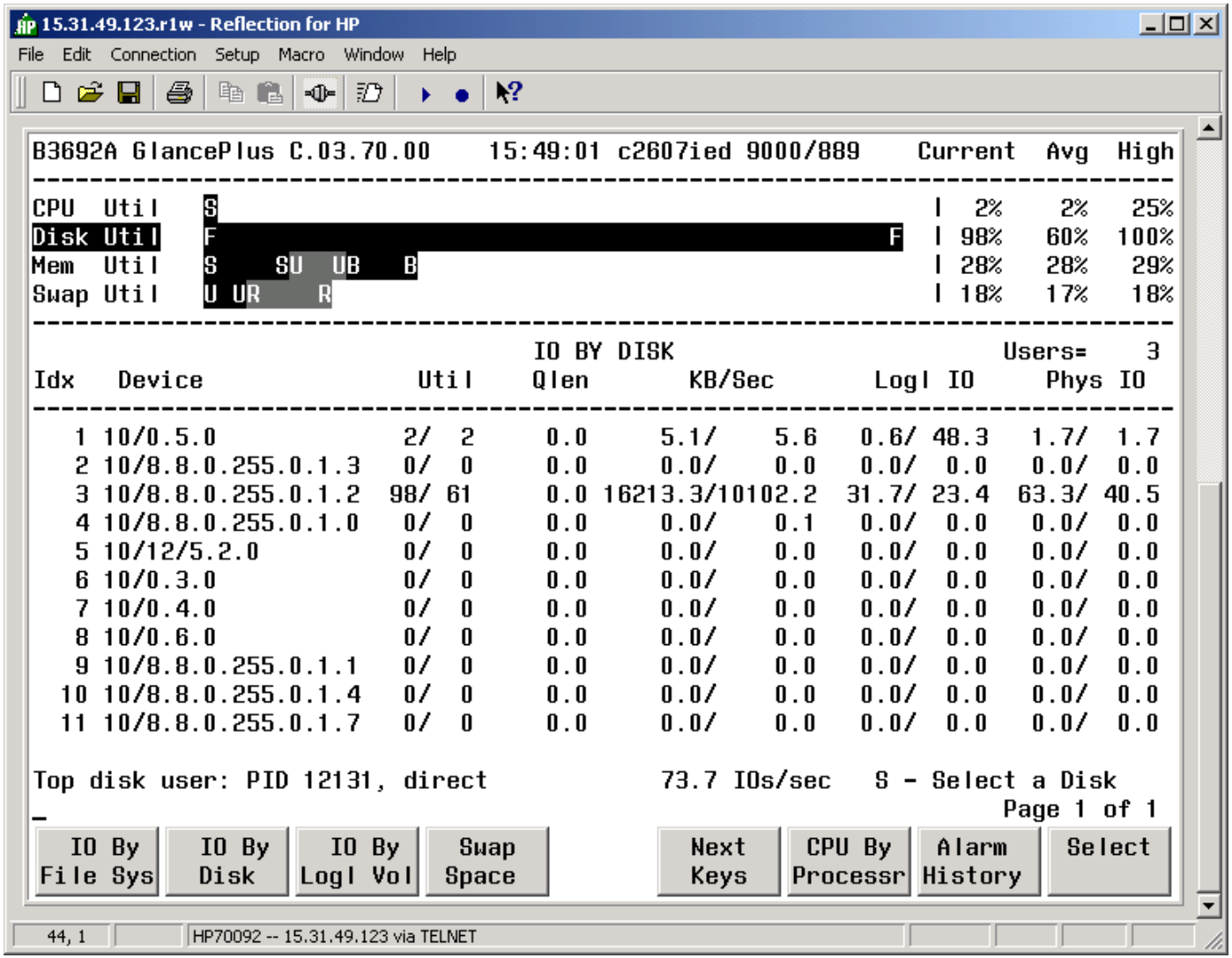
Case Study


15.31.49.123.r1w - Reflection for HP
File Edit Connection Setup Macro Window Help
B3692A GlancePlus C.03.70.00 15:48:22 c2607ied 9000/889 Current Avg High
CPU Util | 2% 2% 25%
Disk Util | 98% 58% 100%
Mem Util | 28% 28% 29%
Swap Util | 18% 17% 18%
DISK REPORT
Users= 3
Req Type Requests % Rate Bytes Cum Req % Cum Rate Cum Byte
Local Logl Rds 699 100.0 35.8 5.75gb 43346 94.5 70.1 5.76gb
Logl Wts 0 0.0 0.0 20.3mb 2501 5.5 4.0 20.3mb
Phys Rds 1235 97.5 63.3 308.5mb 23373 92.6 37.8 5.70gb
Phys Wts 32 2.5 1.6 97kb 1862 7.4 3.0 24.1mb
User 1 0.1 0.0 1kb 66 0.3 0.1 203kb
Virt Mem 1 0.1 0.0 1kb 19 0.1 0.0 19kb
System 1264 99.8 64.8 308.6mb 25139 99.6 40.7 5.73gb
Raw 1 0.1 0.0 8kb 11 0.0 0.0 88kb
Remote Logl Rds 0 0.0 0.0 0kb 0 0.0 0.0 0kb
Logl Wts 0 0.0 0.0 0kb 0 0.0 0.0 0kb
Phys Rds 0 0.0 0.0 0kb 0 0.0 0.0 0kb
Phys Wts 0 0.0 0.0 0kb 0 0.0 0.0 0kb
Page 1 of 2
Process List
CPU Report
Memory Report
Disk Report
Next Keys
Select Process
Help
Exit Glance
44, 1
HP70092 -- 15.31.49.123 via TELNET

Case Study


15.31.49.123.r1w - Reflection for HP
File Edit Connection Setup Macro Window Help
B3692A GlancePlus C.03.70.00 15:48:37 c2607ied 9000/889 Current Avg High
CPU Util S | 2% 2% 25%
Disk Util F F | 98% 59% 100%
Mem Util S SU UB B | 28% 28% 29%
Swap Util U UR R | 18% 17% 18%
DISK REPORT Users= 3
Req Type Requests Rate Cum Req Cum Rate High Rate
Read Cache Hits 11 100.0 61748 100.0 100.0
Write Cache Hits 0 0.0 3741 51.3
DNLC Hits 0 0.0 0 0.0 0.0
DNLC Longs 0 0.0 0 0.0 0.0
Page 2 of 2
Process List
CPU Report
Memory Report
Disk Report
Next Keys
Select Process
Help
Exit Glance
44, 1
HP70092 -- 15.31.49.123 via TELNET

Case Study


15.31.49.123.r1w - Reflection for HP
File Edit Connection Setup Macro Window Help
B3692A GlancePlus C.03.70.00 15:49:01 c2607ied 9000/889 Current Avg High
CPU Util S | 2% 2% 25%
Disk Util F F | 98% 60% 100%
Mem Util S SU UB B | 28% 28% 29%
Swap Util U UR R | 18% 17% 18%
IO BY DISK Users= 3
Idx Device Util Qlen KB/Sec Logl IO Phys IO
1 10/0.5.0 2/ 2 0.0 5.1/ 5.6 0.6/ 48.3 1.7/ 1.7
2 10/8.8.0.255.0.1.3 0/ 0 0.0 0.0/ 0.0 0.0/ 0.0 0.0/ 0.0
3 10/8.8.0.255.0.1.2 98/ 61 0.0 16213.3/10102.2 31.7/ 23.4 63.3/ 40.5
4 10/8.8.0.255.0.1.0 0/ 0 0.0 0.0/ 0.1 0.0/ 0.0 0.0/ 0.0
5 10/12/5.2.0 0/ 0 0.0 0.0/ 0.0 0.0/ 0.0 0.0/ 0.0
6 10/0.3.0 0/ 0 0.0 0.0/ 0.0 0.0/ 0.0 0.0/ 0.0
7 10/0.4.0 0/ 0 0.0 0.0/ 0.0 0.0/ 0.0 0.0/ 0.0
8 10/0.6.0 0/ 0 0.0 0.0/ 0.0 0.0/ 0.0 0.0/ 0.0
9 10/8.8.0.255.0.1.1 0/ 0 0.0 0.0/ 0.0 0.0/ 0.0 0.0/ 0.0
10 10/8.8.0.255.0.1.4 0/ 0 0.0 0.0/ 0.0 0.0/ 0.0 0.0/ 0.0
11 10/8.8.0.255.0.1.7 0/ 0 0.0 0.0/ 0.0 0.0/ 0.0 0.0/ 0.0
Top disk user: PID 12131, direct 73.7 IOs/sec S - Select a Disk
Page 1 of 1
IO By File Sys
IO By Disk
IO By Logl Vol
Swap Space
Next Keys
CPU By Processr
Alarm History
Select
44, 1
HP70092 -- 15.31.49.123 via TELNET

Case Study

Again we can use KItrace to see the characteristics of the IO.

Kparse will take the KItrace output and extract such things as disk service times, queue lengths and disk block frequency.

Case Study

From the Kparse report:

```
Disk block frequency...
Freq    Dev                     Block
597     dev_t=31/0x031200       blkno=0xb37340  wr=read
597     dev_t=31/0x031200       blkno=0xb37240  wr=read
4       dev_t=31/0x031200       blkno=0x538     wr=write
2       dev_t=31/0x031200       blkno=0xbee378  wr=write
2       dev_t=31/0x031200       blkno=0xb4f2ec  wr=write
2       dev_t=31/0x031200       blkno=0xb4f2cc  wr=write
2       dev_t=31/0x025000       blkno=0x3fb2b4  wr=write
2       dev_t=31/0x025000       blkno=0x30badc  wr=write
```

Case Study

We see the same physical blocks being read from the disk multiple times during the short (20 second) data collection.

Why are these blocks being continuously read from the disk when the file system should be using the buffer cache and therefore the block should be available in the buffer cache?

Case Study

If we look at a particular pid doing IO we can see what the IO looks like:

```
 pid=12131 ktid=13338 lseek err=0 ret1=0
ENQUEUE dev_t=31/0x031200  pid-u/a=12131/12131 wr=read blkno=0xb37240
b_flags=call/ndelay/busy/read/pftimeout/phys/
ENQUEUE dev_t=31/0x031200  pid-u/a=12131/12131 wr=read blkno=0xb37340
b_flags=call/ndelay/busy/read/pftimeout/phys/
  pid=12131 ktid=13338 read err=0 ret1=524288
  pid=12131 ktid=13338 lseek err=0 ret1=0
ENQUEUE dev_t=31/0x031200  pid-u/a=12131/12131 wr=read blkno=0xb37240
b_flags=call/ndelay/busy/read/pftimeout/phys/
ENQUEUE dev_t=31/0x031200  pid-u/a=12131/12131 wr=read blkno=0xb37340
b_flags=call/ndelay/busy/read/pftimeout/phys/
  pid=12131 ktid=13338 read err=0 ret1=524288
  pid=12131 ktid=13338 lseek err=0 ret1=0
ENQUEUE dev_t=31/0x031200  pid-u/a=12131/12131 wr=read blkno=0xb37240
b_flags=call/ndelay/busy/read/pftimeout/phys/
ENQUEUE dev_t=31/0x031200  pid-u/a=12131/12131 wr=read blkno=0xb37340
b_flags=call/ndelay/busy/read/pftimeout/phys/
  pid=12131 ktid=13338 read err=0 ret1=524288
```

Case Study

Here we see the same blocks being read repeatedly by the application (lseek to position 0, read), the reads rather large (524288 bytes) and the IO bypassing the buffer cache (b_flags=phys)

```
 pid=12131 ktid=13338 lseek err=0 ret1=0
ENQUEUE dev_t=31/0x031200  pid-u/a=12131/12131 wr=read blkno=0xb37240
b_flags=call/ndelay/busy/read/pftimeout/phys/
ENQUEUE dev_t=31/0x031200  pid-u/a=12131/12131 wr=read blkno=0xb37340
b_flags=call/ndelay/busy/read/pftimeout/phys/
  pid=12131 ktid=13338 read err=0 ret1=524288
```

Case Study

This is the discovered_direct_io feature of Online JFS.

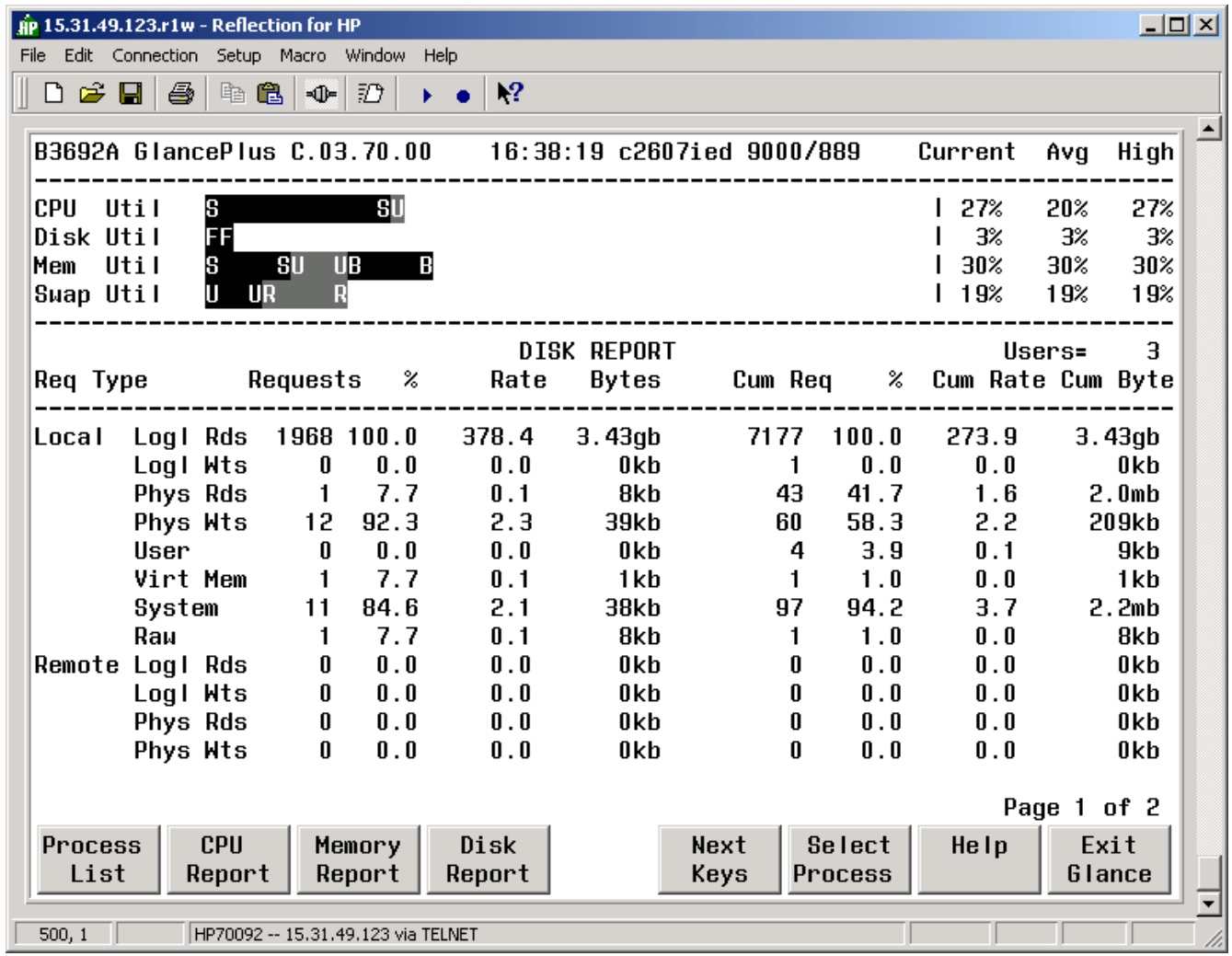
Large reads typically are done once (backups or copies) and do not need to be kept in the buffer cache.

However, in this case the reads were repeated. The discovered_direct_io parameter should be tuned for this application.

Case Study

```
# vxtunefs /home/jan
Filesystem i/o parameters for /home/jan
read_pref_io = 65536
read_nstream = 1
read_unit_io = 65536
write_pref_io = 65536
write_nstream = 1
write_unit_io = 65536
pref_strength = 10
buf_breakup_size = 262144
discovered_direct_iosz = 262144
max_direct_iosz = 1048576
default_indir_size = 8192
qio_cache_enable = 0
max_diskq = 1048576
initial_extent_size = 2
max_seqio_extent_size = 2048
max_buf_data_size = 8192
```

Case Study


15.31.49.123.r1w - Reflection for HP
File Edit Connection Setup Macro Window Help
B3692A GlancePlus C.03.70.00 16:38:19 c2607ied 9000/889 Current Avg High
CPU Util S SU | 27% 20% 27%
Disk Util FF | 3% 3% 3%
Mem Util S SU UB B | 30% 30% 30%
Swap Util U UR R | 19% 19% 19%
DISK REPORT Users= 3
Req Type Requests % Rate Bytes Cum Req % Cum Rate Cum Byte
Local Logl Rds 1968 100.0 378.4 3.43gb 7177 100.0 273.9 3.43gb
Logl Wts 0 0.0 0.0 0kb 1 0.0 0.0 0kb
Phys Rds 1 7.7 0.1 8kb 43 41.7 1.6 2.0mb
Phys Wts 12 92.3 2.3 39kb 60 58.3 2.2 209kb
User 0 0.0 0.0 0kb 4 3.9 0.1 9kb
Virt Mem 1 7.7 0.1 1kb 1 1.0 0.0 1kb
System 11 84.6 2.1 38kb 97 94.2 3.7 2.2mb
Raw 1 7.7 0.1 8kb 1 1.0 0.0 8kb
Remote Logl Rds 0 0.0 0.0 0kb 0 0.0 0.0 0kb
Logl Wts 0 0.0 0.0 0kb 0 0.0 0.0 0kb
Phys Rds 0 0.0 0.0 0kb 0 0.0 0.0 0kb
Phys Wts 0 0.0 0.0 0kb 0 0.0 0.0 0kb
Page 1 of 2
Process List
CPU Report
Memory Report
Disk Report
Next Keys
Select Process
Help
Exit Glance
500, 1
HP70092 -- 15.31.49.123 via TELNET

Interex, Encompass and HP bring you a powerful new HP World.