OpenVMS Performance Tips & Tricks

Guy Peleg
President
Maklee Engineering
guy.peleg@maklee.com

MAKLEE
software engineering
solutions

Maklee Engineering

- › Consulting firm operating all over the world.
  - Team of "Top Gun" engineers.
  - Former members of various engineering groups at HP.
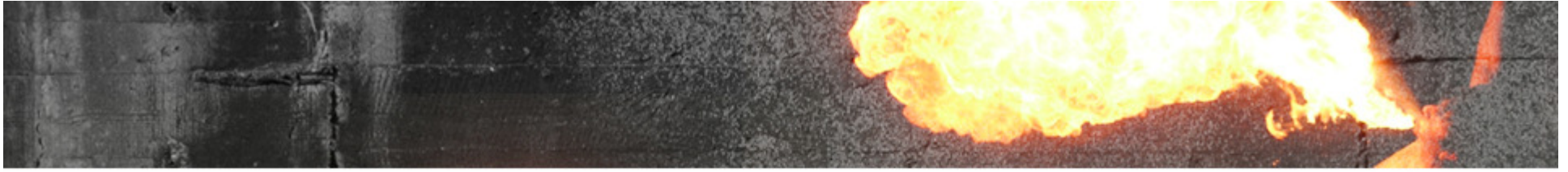 - Gold Oracle partner.

- › Specialize in performance tuning of:
  - OpenVMS
  - Oracle (HP-UX, Linux, VMS, Solaris, AIX, Windows)
  - Oracle Rdb
  - Java (HP-UX, Linux, VMS, Solaris, AIX, Windows)
  - Adabas

- › Also offers custom engineering services, custom training and on-going support contracts.

Maklee provides guarantee for success for all projects !!

Maklee Guarantees __doubling__ the performance of your Oracle database or our service is provided free of charge !
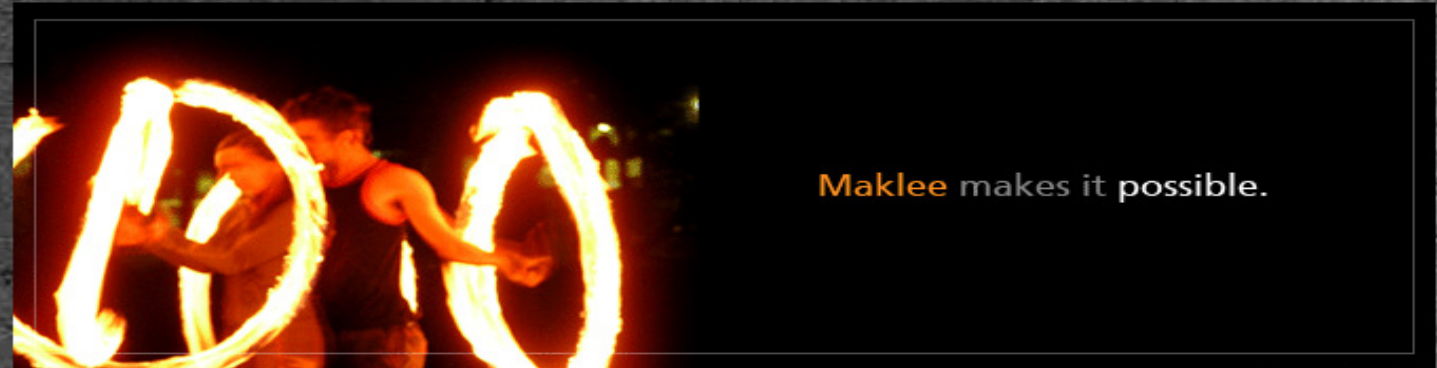
and….we speak German !!

http://www.maklee.com/indexDeutsch.html

MAKLEE

MAKLEE
software engineering solutions

Get Maklee/Oracle Brochure

Maklee makes it **possible.**

Oracle Services / SQL Tuning

ORACLE PARTNER

Maklee verfügt über umfassende Kompetenzen im Bereich Oracle Tuning mit spezialisierter Erfahrung bei der Arbeit am Tuning der anspruchsvollsten Workloads.

Der Vorteil von Maklee

Das Maklee-Team verfügt über ein tiefgreifendes Verständnis sowohl über Oracle als auch die darunter liegenden Betriebssysteme. Wir unterhalten enge Arbeitskontakte mit den Entwicklungsteams der führenden Hersteller von Betriebssystemen und mit den Entwicklungsgruppen der Oracle Corporation. In dem wir das Feedback des Kunden zu jeder Zeit berücksichtigen, erfüllen unsere Lösungen genau die Bedürfnisse des Kunden. Zusätzlich bleibt Maklee kontinuierlich bezüglich der aktuellsten technischen Entwicklungen und Veränderungen auf dem Laufenden.

Oracle Performance Tuning

Oracle Tuning birgt ein unendliches Potential zur Verbesserung der Performance. Die Standardeinstellungen von Oracle sind nicht immer optimal. Das Tuning ist ausschlaggebend, damit man das Beste aus einem System herausholen kann. Unser kreativer Ansatz resultiert in einem herausragenden Maß der Performance-Verbesserung. Kürzlich bei einem Einsatz für eine führende globale Bank konnte das Maklee-Team die Laufzeit einer Abfrage von 90 Minuten auf 4 Sekunden reduzieren – eine 1350-fache Steigerung der Performance konnte wiedergegeben werden. Unsere Spezialisierung beinhaltet das Monitoring und Tuning aller Oracle Datenbanken einschließlich RAC und Oracle Anwendungen, Oracle Instance Tuning und SQL Tuning. Um unsere Erfolgsgewährleistung realisieren zu können, führen wir während des gesamten Tuning-Prozesses Evaluationen durch. Diese Evaluationen berücksichtigen die Parameter des Betriebssystems und der Datenbank, die Execution-Pläne der Key SQL Statements und das Umschreiben der problematischen SQL Statements.

Kontakt

info@maklee.com
Telefon: 1-800-224-4513
Fax: 1-646-452-9402
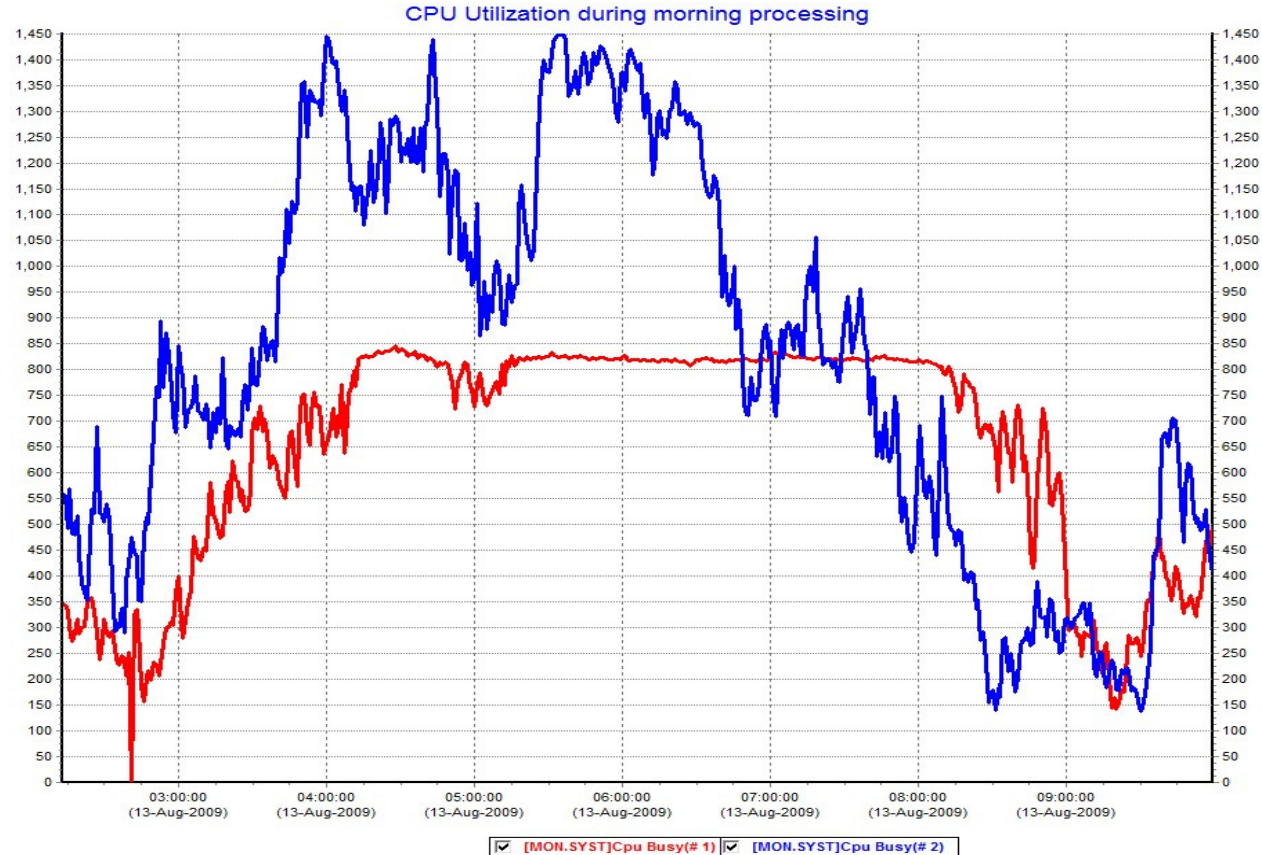
Corporate Headquarters:

322W 57th street
New York, NY 10019
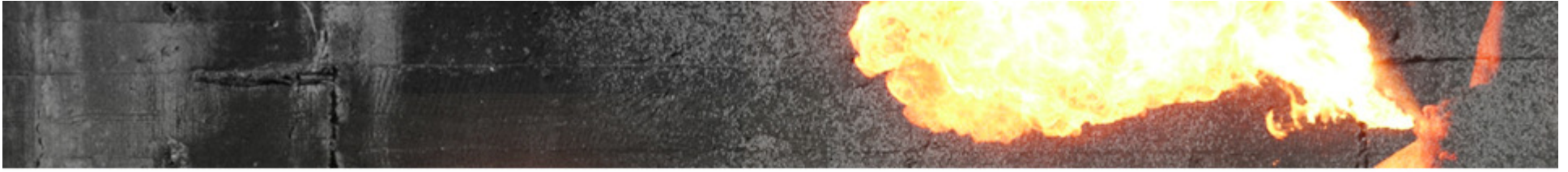
> Das deutsche Publikum ist nicht einfach

> Es fordert immer neues Material

(Mein Deutsch wird immer besser)

MAKLEE

The Easy Way…

> As mentioned MANY times in the past, the easiest way to improve performance is to throw more hardware on the problem.
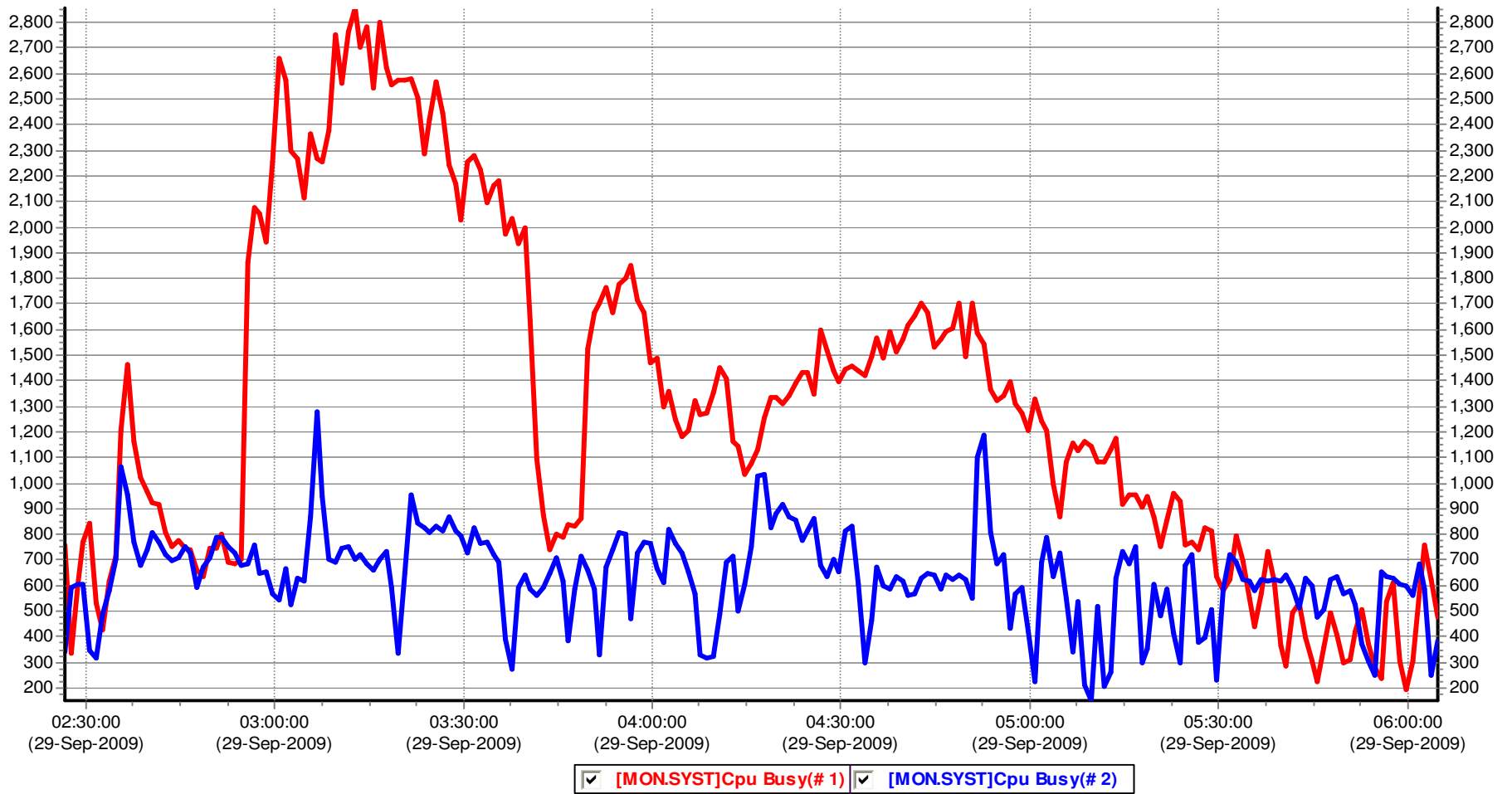

CPU Utilization during morning processing

The Hard Way…

> Problems start when:

- The latest and greatest hardware is already being used.
- Hardware solution is too expensive.
- Extra hardware degrades performance even more.
- Extra hardware improved the situation but is still not enough….

This is the time to seek professional help…..

MAKLEE

The Maklee Magic…

Alignment Faults

- Alignment faults are VERY expensive on Itanium.

- However…Alignment faults are not unique to Itanium and can be experienced on Alpha as well.

- On Alpha alignment faults impact the process generating the faults vs. the entire system.

- MONITOR ALIGN does not work on Alpha, but SDA PRF does.

- The following slide demonstrates the impact of eliminating alignment faults on Alpha.

  - 16 cores GS1280 (1.3Ghz).

  - OLTP application, for this particular workload every millisecond counts

  - One line code change to align one data structure.

  - 25% performance improvement

MAKLEE

CPU Utilization

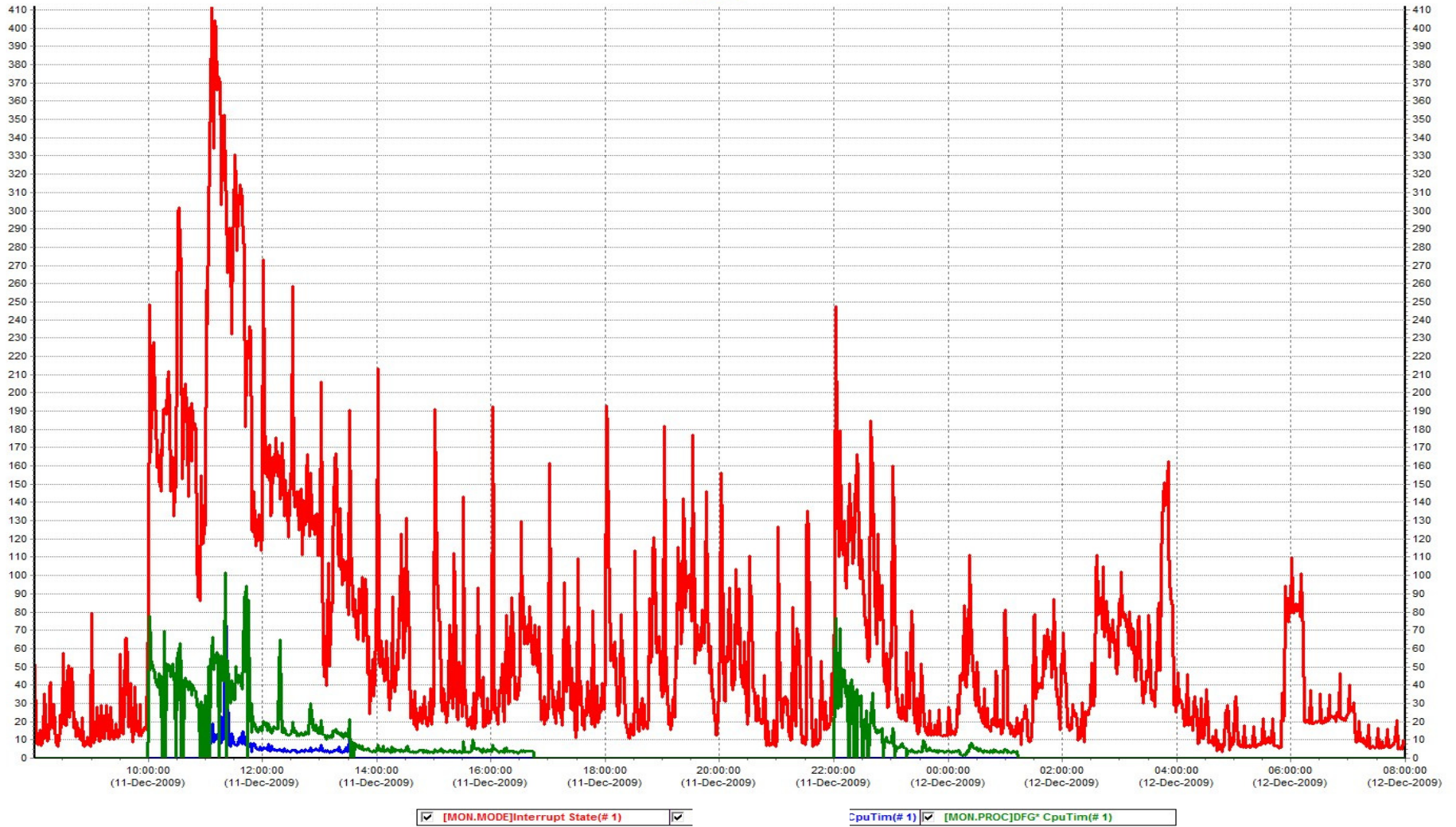[MON.SYST]Cpu Busy(# 1)   [MON.SYST]Cpu Busy(# 2)

10

DFG
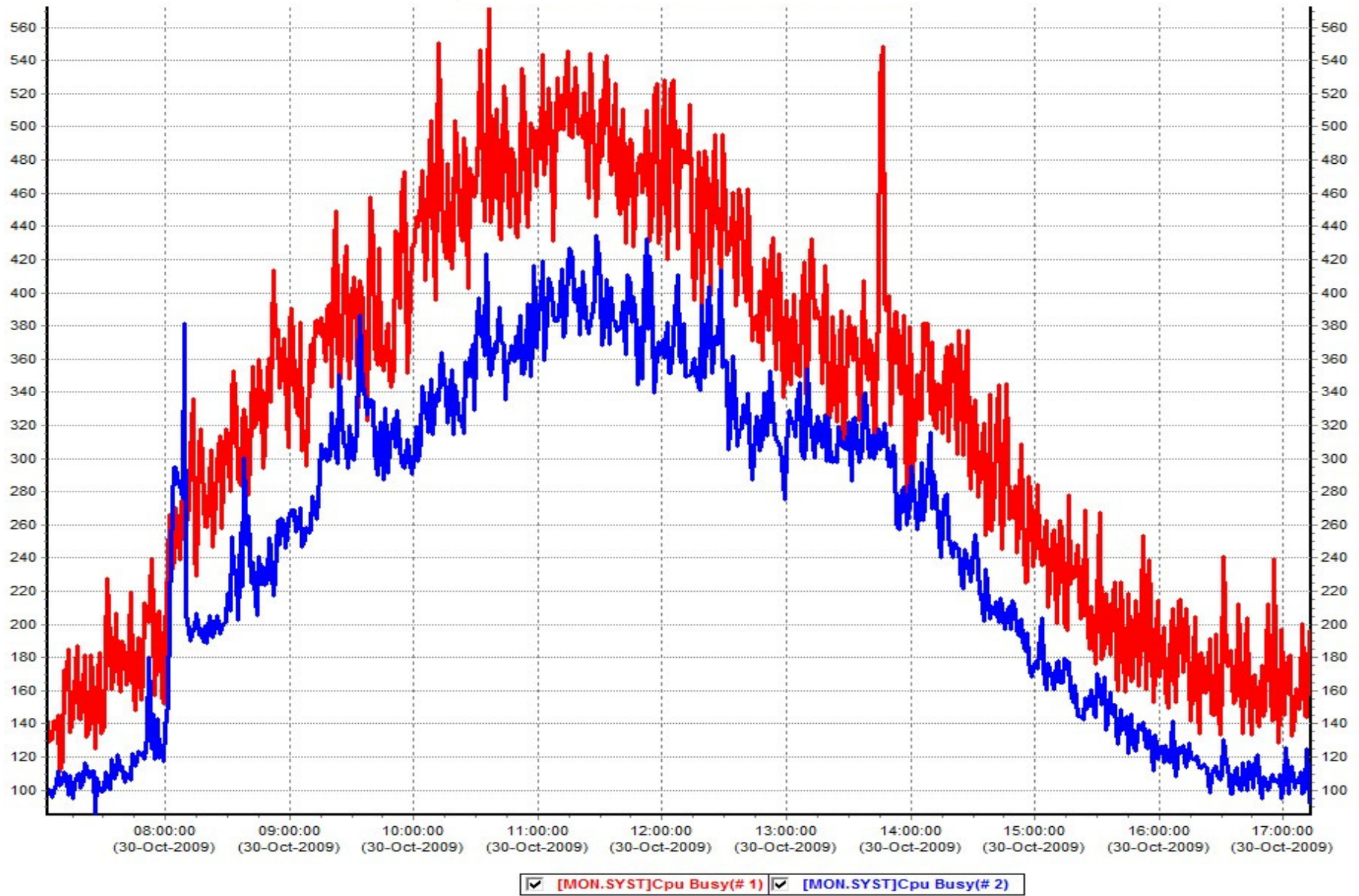
> Disk defragger

- The defragmentor puts significant load on the I/O subsystem.

- Avoid defragging multiple disks at the same time.

- Stagger them and start them maybe 5 min after each other to reduce contention

MAKLEE

Defrag Footprint

Interrupt State Vs. DFG CPU Utilization

CONVERT

> Improving performance of CONVERT operations is always challenging.

> Convert is using two types of temporary work files:
  - Sort work files (used when /sort is specified)
  - Convert work files

> To speed up convert operations, we evaluated the impact of moving temporary work files to other disks.
  - VMS allows controlling the location of temporary work files using logical names.

> On Itanium, the MCOE & EOE packages, include the DECRAM license.
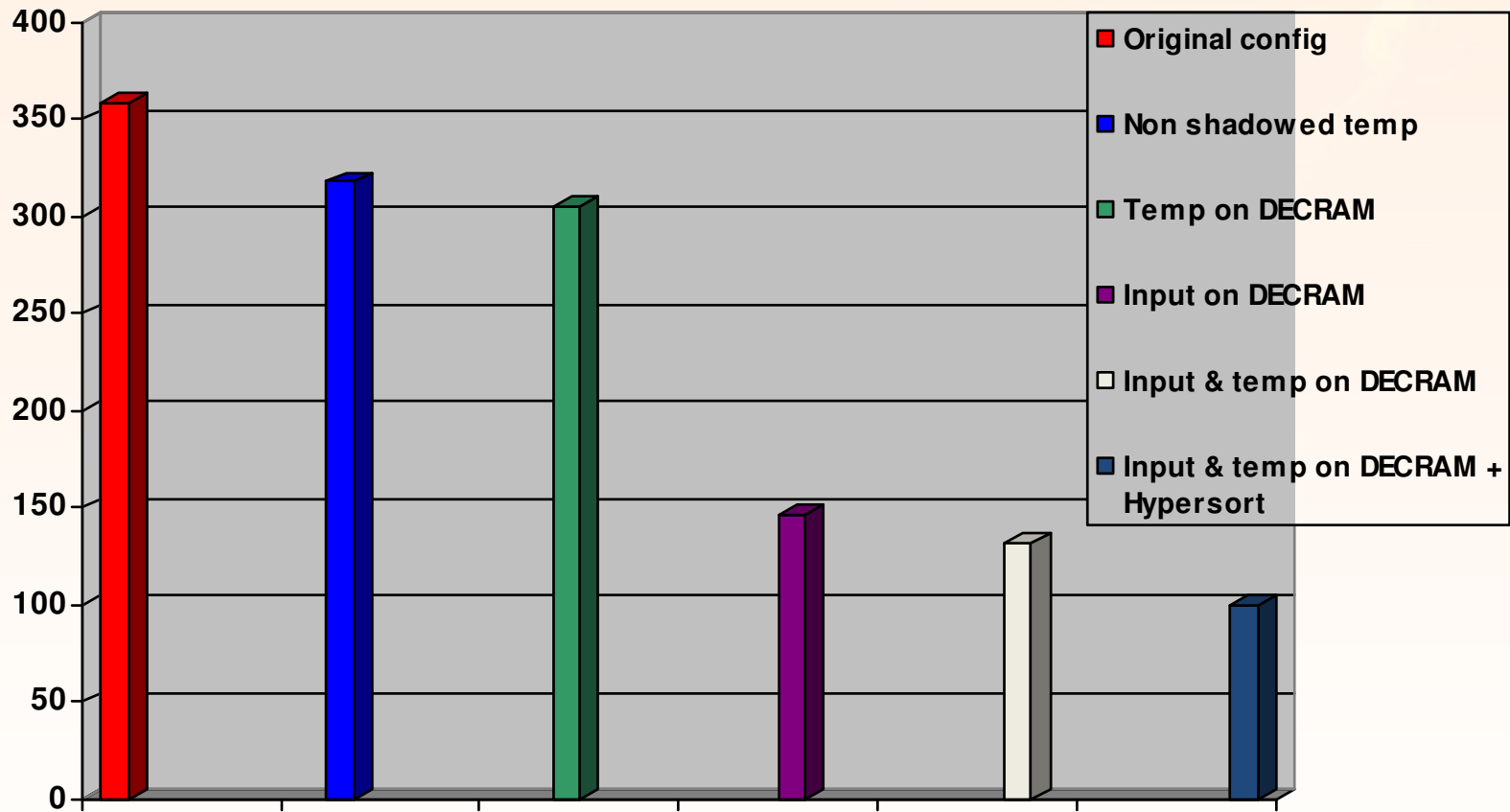
MAKLEE

CONVERT

- CONVERT
  - 32 cores RX8640
  - 96GB RAM
  - OpenVMS V8.3-1H1
  - Input file - 1.3GB RMS indexed file
  - Temporary work file peaked at 7GB
  - Convert performed using the /NOSORT qualifier.

- The convert operation required 6 minutes to complete. The customer needed to complete the operation as quickly as possible

MAKLEE

CONVERT

Legend:
- Original config
- Non shadowed temp
- Temp on DECRAM
- Input on DECRAM
- Input & temp on DECRAM
- Input & temp on DECRAM + Hypersort

Elapsed time (seconds) to sort 1.3GB
Less is better

DECRAM Vs. SSD

> Solid state disks are VERY fast.

> SSD provide very low read latency for random I/O operations.

> DECRAM would ALWAYS be faster.
- DECRAM access does not leave the host.
- In the x86 world Fusion I/O cards provide similar functionality to DECRAM

MAKLEE

Disk I/O Performance

	Sequential Reads (128KB transfers)	Random Reads (4KB transfers)	Random Writes (4KB transfers)
RamSan (local)	300	46	(*)
RamSan (remote)	125	6.4	(*)
EVA 8100 (local)	240	1.6	22
EVA 8100 (remote)	120	1.3	3
DECram	1180	250	250

Rates are in MB/sec

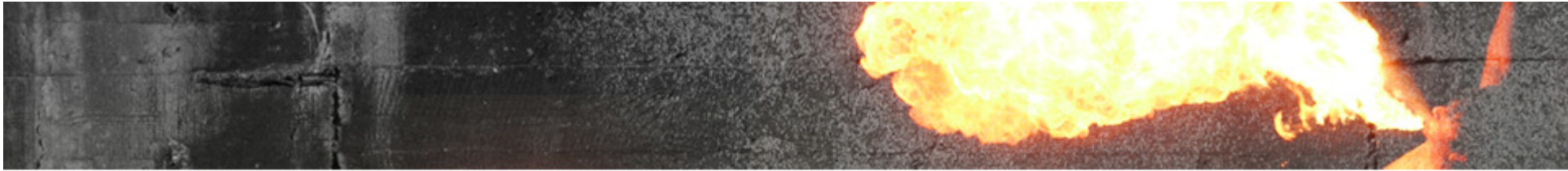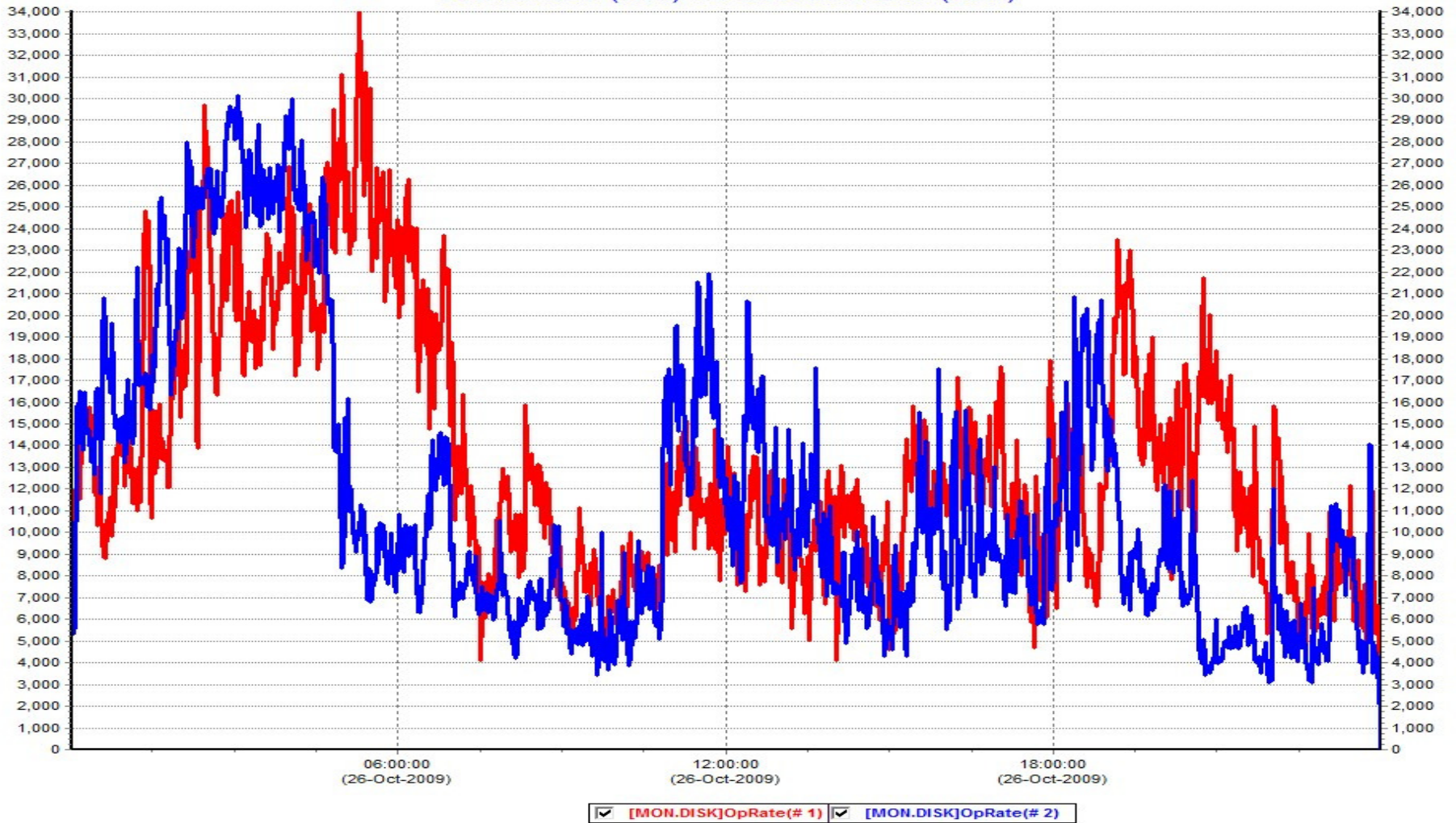(*) RamSan writes could not be tested since no scratch disks were available

MAKLEE

FC Queue Full

> A customer reported degradation in I/O performance.

> BACKUP to the null device was able to read 20-25MB per second.

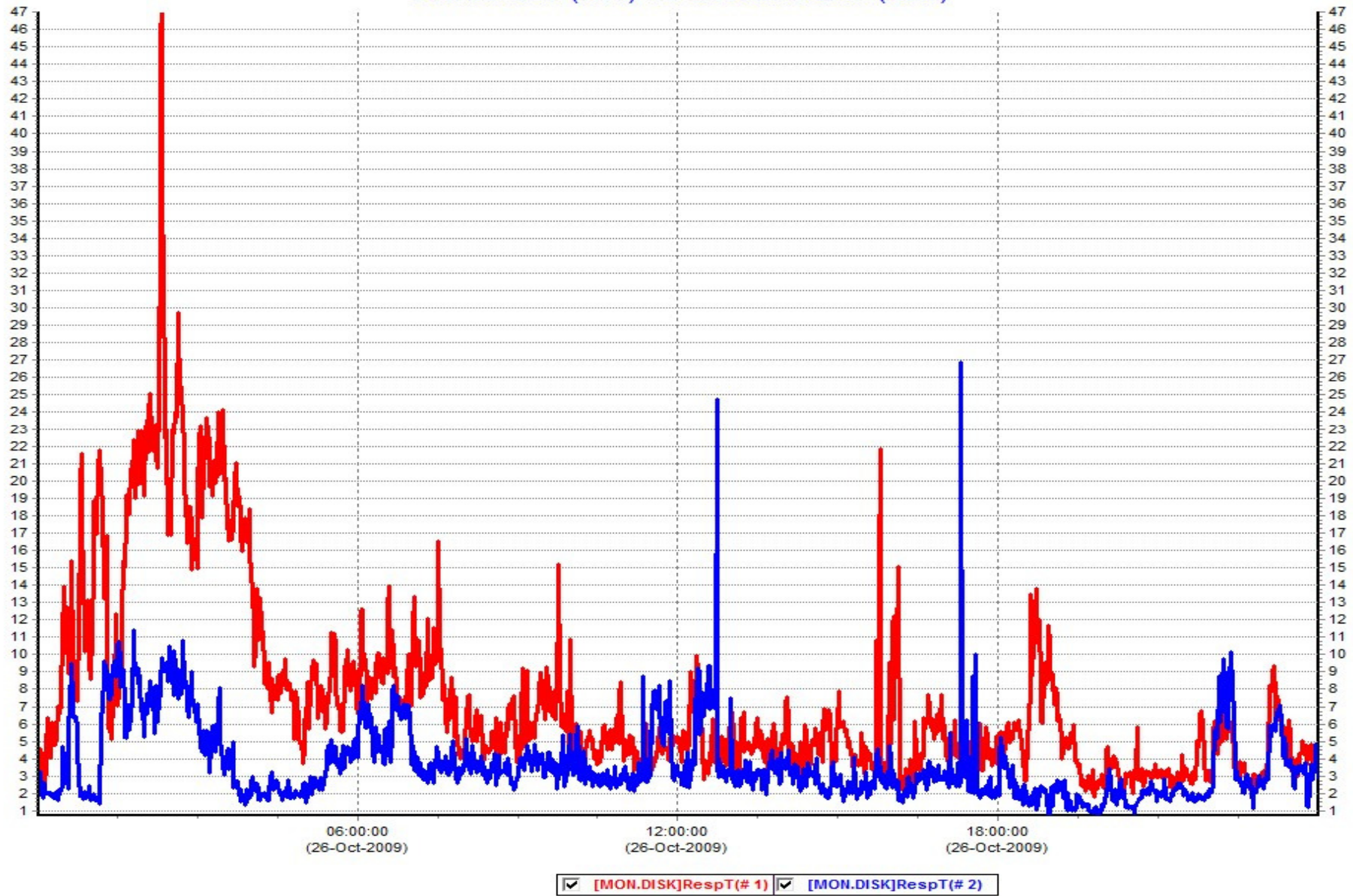> Expected throughput is in the range of 80-100MB/sec

Ave. Disk Operation Rate
October 26th (Red) Vs. December 24th (Blue)

Ave. Disk Response time

October 26th (Red) Vs. December 24th (Blue)

[MON.DISK]RespT(# 1) [MON.DISK]RespT(# 2)

20

FC Queue Full

› Use SANCP utility to check for "queue full" and reduced "credits available"

- Dev I/Os   - number concurrent I/Os in flight
- Cred I/Os   - number of pending I/Os because no credit available
- QF seen   - number of queue full event seen
- Cred Avl   - number of credits available (resources available)

› After a queue full event, OpenVMS will throttle, but should reset the credits once things are back to normal

- This is a VMS bug which re-appeared, need to log a call with HP
- Typically, will be seen with very busy storage arrays servicing multiple hosts.

› Use SANCP to reset credits available if too low

- $ mc sancp set wwid /pro=hsv200 /port=5000.1FE1.500B.3358 /default

MAKLEE

FC Queue Full

```
$ mc sancp show wwid

                                               | Conn  Port  Dev   Cred  Appr | QF   Cred  User  Current
Product ID  Node WWID            Port WWID      | Cnt   I/Os  I/Os  I/Os  I/Os | Seen Avl   Cap   QF Alg
----------  -------------------  -------------------  + ----  ----  ----  ----  ---- + ----  ----  ----  -------
HSV111 (C)  5000.1FE1.5000.DAA0  5000.1FE1.5000.DAA8  | 0001  0000  0000  0000  0000 | 0000 8000  0000  T/5000
HSV111 (C)  5000.1FE1.5000.DAA0  5000.1FE1.5000.DAA9  | 0001  0000  0000  0000  0000 | 0000 8000  0000  T/5000
HSV111 (C)  5000.1FE1.5000.DAA0  5000.1FE1.5000.DAAC  | 0001  0000  0000  0000  0000 | 0000 8000  0000  T/5000
HSV111 (C)  5000.1FE1.5000.DAA0  5000.1FE1.5000.DAAD  | 0001  0000  0000  0000  0000 | 0000 8000  0000  T/5000
HSV200      5000.1FE1.500B.3350  5000.1FE1.500B.3358  | 0001  0000  0001  0000  0000 | 0003 0034  0000  T/5000
HSV200      5000.1FE1.500B.3350  5000.1FE1.500B.3359  | 0001  0000  0001  0000  0000 | 0001 002C  0000  T/5000
HSV200      5000.1FE1.500B.3350  5000.1FE1.500B.335C  | 0001  0000  0003  0000  0000 | 0000 7FFD  0000  T/5000
HSV200      5000.1FE1.500B.3350  5000.1FE1.500B.335D  | 0001  0000  0001  0000  0000 | 0000 7FFF  0000  T/5000
```

```
$ mc sancp show wwid

                                               | Conn  Port  Dev   Cred  Appr | QF   Cred  User  Current
Product ID  Node WWID            Port WWID      | Cnt   I/Os  I/Os  I/Os  I/Os | Seen Avl   Cap   QF Alg
----------  -------------------  -------------------  + ----  ----  ----  ----  ---- + ----  ----  ----  -------
HSV111 (C)  5000.1FE1.5000.DAA0  5000.1FE1.5000.DAA8  | 0001  0000  0000  0000  0000 | 0000 8000  0000  T/5000
HSV111 (C)  5000.1FE1.5000.DAA0  5000.1FE1.5000.DAA9  | 0001  0000  0000  0000  0000 | 0000 8000  0000  T/5000
HSV111 (C)  5000.1FE1.5000.DAA0  5000.1FE1.5000.DAAC  | 0001  0000  0000  0000  0000 | 0000 8000  0000  T/5000
HSV111 (C)  5000.1FE1.5000.DAA0  5000.1FE1.5000.DAAD  | 0001  0000  0000  0000  0000 | 0000 8000  0000  T/5000
HSV200      5000.1FE1.500B.3350  5000.1FE1.500B.3358  | 0001  0000  005F  0004  0000 | 0002 0000  0000  T/5000
HSV200      5000.1FE1.500B.3350  5000.1FE1.500B.3359  | 0001  0000  0001  0000  0000 | 0000 7FFF  0000  T/5000
HSV200      5000.1FE1.500B.3350  5000.1FE1.500B.335C  | 0001  0000  0006  0000  0000 | 0000 7FFA  0000  T/5000
HSV200      5000.1FE1.500B.3350  5000.1FE1.500B.335D  | 0001  0000  0000  0000  0000 | 0000 8000  0000  T/5000
```
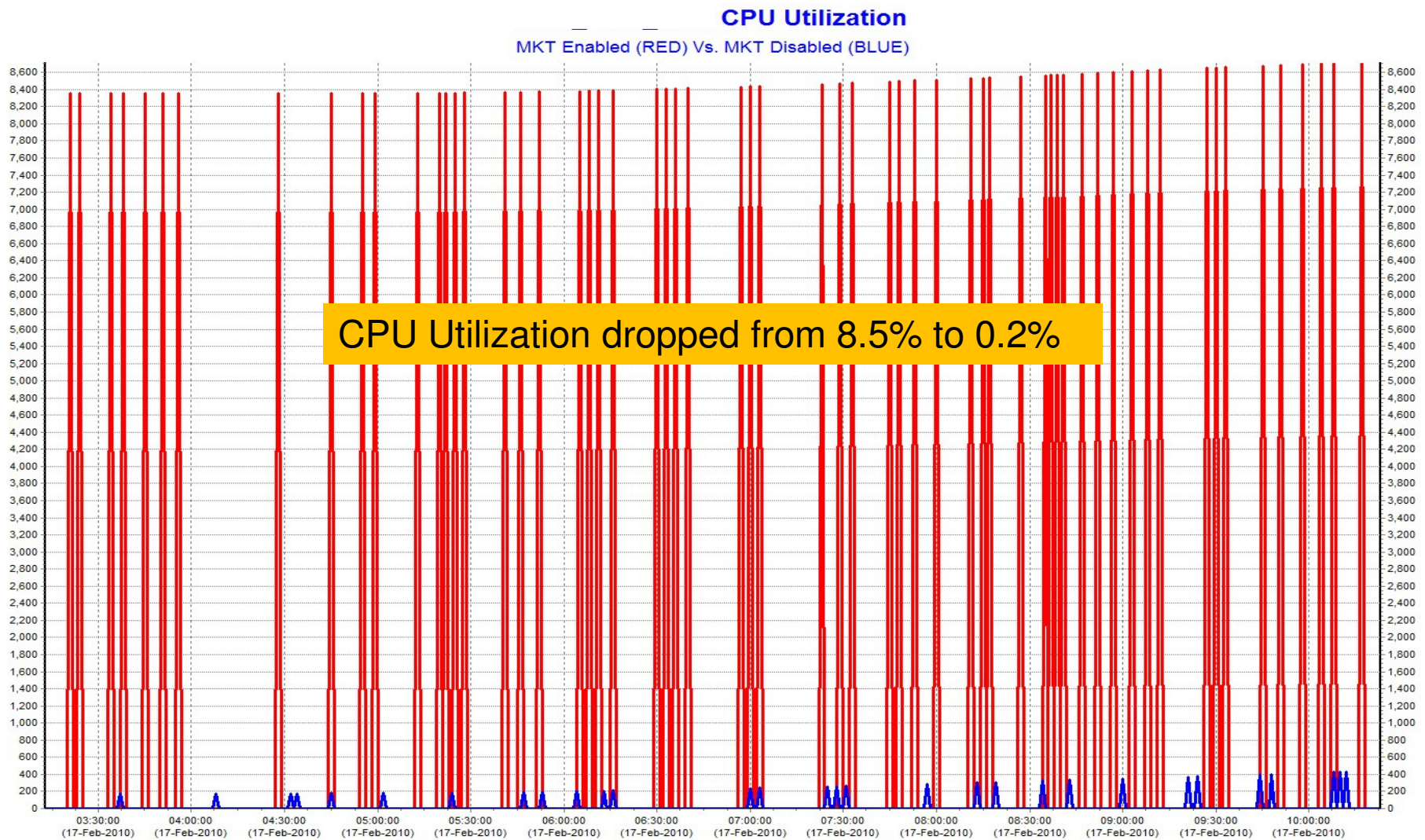
MAKLEE

PTHREAD API

› A program was using the PTHREAD run-time library (pthread_X API) but was not a multi-threaded program.

› Even though the program was NOT multi-threaded, OpenVMS created 16 kernel threads when activating the image (rx6600, HT enabled).

› The OpenVMS threads manager has to keep all the kernel threads synchronized.

› Profiling of the program uncovered that significant CPU time is spent synchronizing the kernel threads.

› As the application was not multi-threaded, multiple kernel threads has been disabled for the image, resulting in SIGNIFICANT reduction of CPU time consumed by the process.

MAKLEE

CPU Utilization
MKT Enabled (RED) Vs. MKT Disabled (BLUE)

CPU Utilization dropped from 8.5% to 0.2%

Reduce Locking

> Reducing locking footprint of single process operating on a file.

> Methods to use No Query Locking (NQL)
- Open files with NQL bit set or perform $GET operations with NQL bit set
- Open files with NLK+RRL bits set or perform $GET operations with NLK+RRL bits set, plus RMS switch to turn NLK+RRL into NQL

> RMS still locks the record eventhough NLK+RRL bits are set
- $ SET RMS /QUERY=DISABLE /SYSTEM
- will not lock data record for any $GET operation with NLK+RRL bits set

MAKLEE

Reduce Locking (cont'd)

› **Risk**

- Very low risk to disable query locking systemwide via
- $ SET RMS /QUERY=DISABLE /SYSTEM
- Only affects $GET operations with NLK+RRL bits set

› **Performance benefit**

- Anything between 0% and a positive percent number
- Depends on how many $GETs have NLK+RRL bits set
- Depends on how busy the dedicated lock manager is

MAKLEE

No Query Locking

> ## Before

```
Accounting information:
Buffered I/O count:                    169    Peak working set size:        19552
Direct I/O count:                   241835    Peak virtual size:           186144
Page faults:                          1320    Mounted volumes:                  0
Charged CPU time:         0 00:03:06.73       Elapsed time:         0 00:21:55.09
```
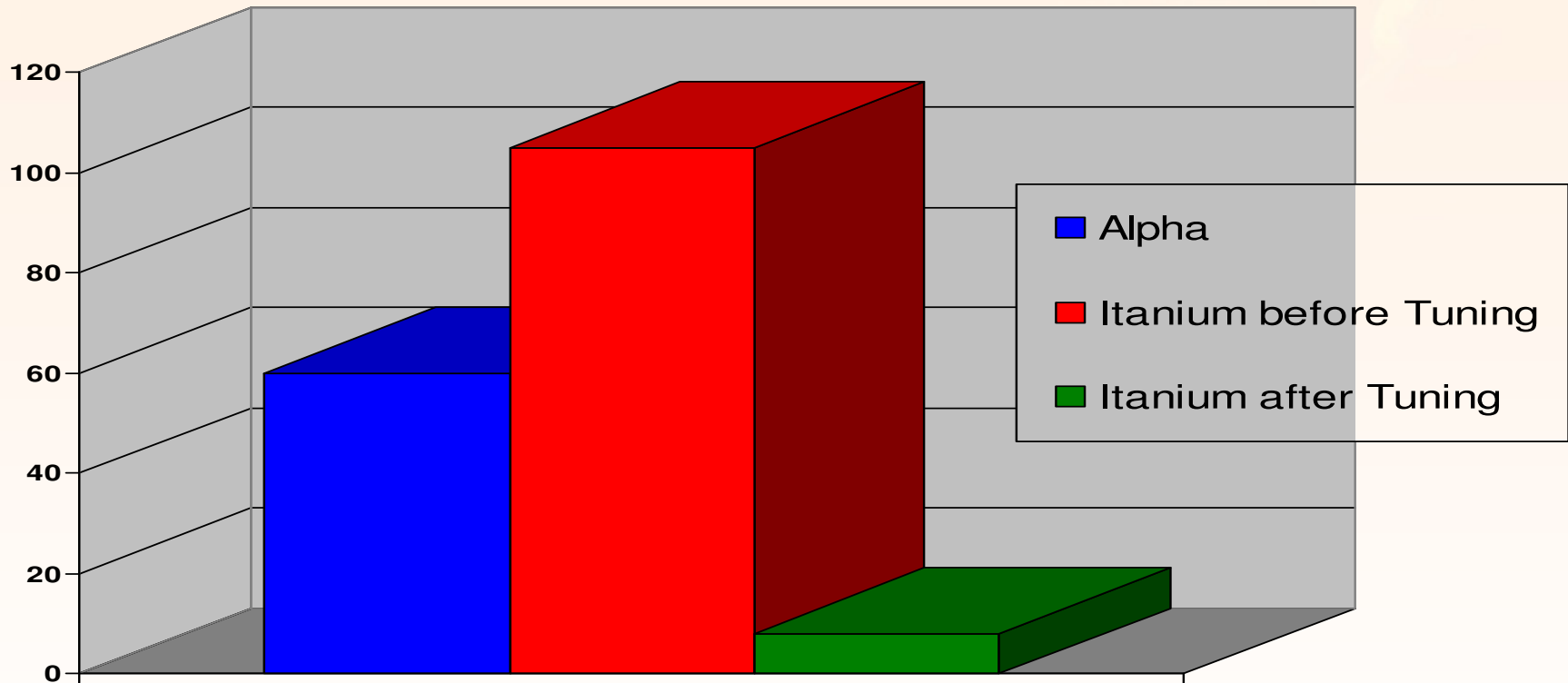
> ## After

```
Accounting information:
Buffered I/O count:                    173    Peak working set size:        19536
Direct I/O count:                   241358    Peak virtual size:           186144
Page faults:                          1367    Mounted volumes:                  0
Charged CPU time:         0 00:01:49.36       Elapsed time:         0 00:14:44.91
```

> ## More than 30% reduction in elapsed time

> ## More than 40% reduction in CPU time

MAKLEE

Oracle Index Creation on Large Table

Elapsed Time (in min) - less is better

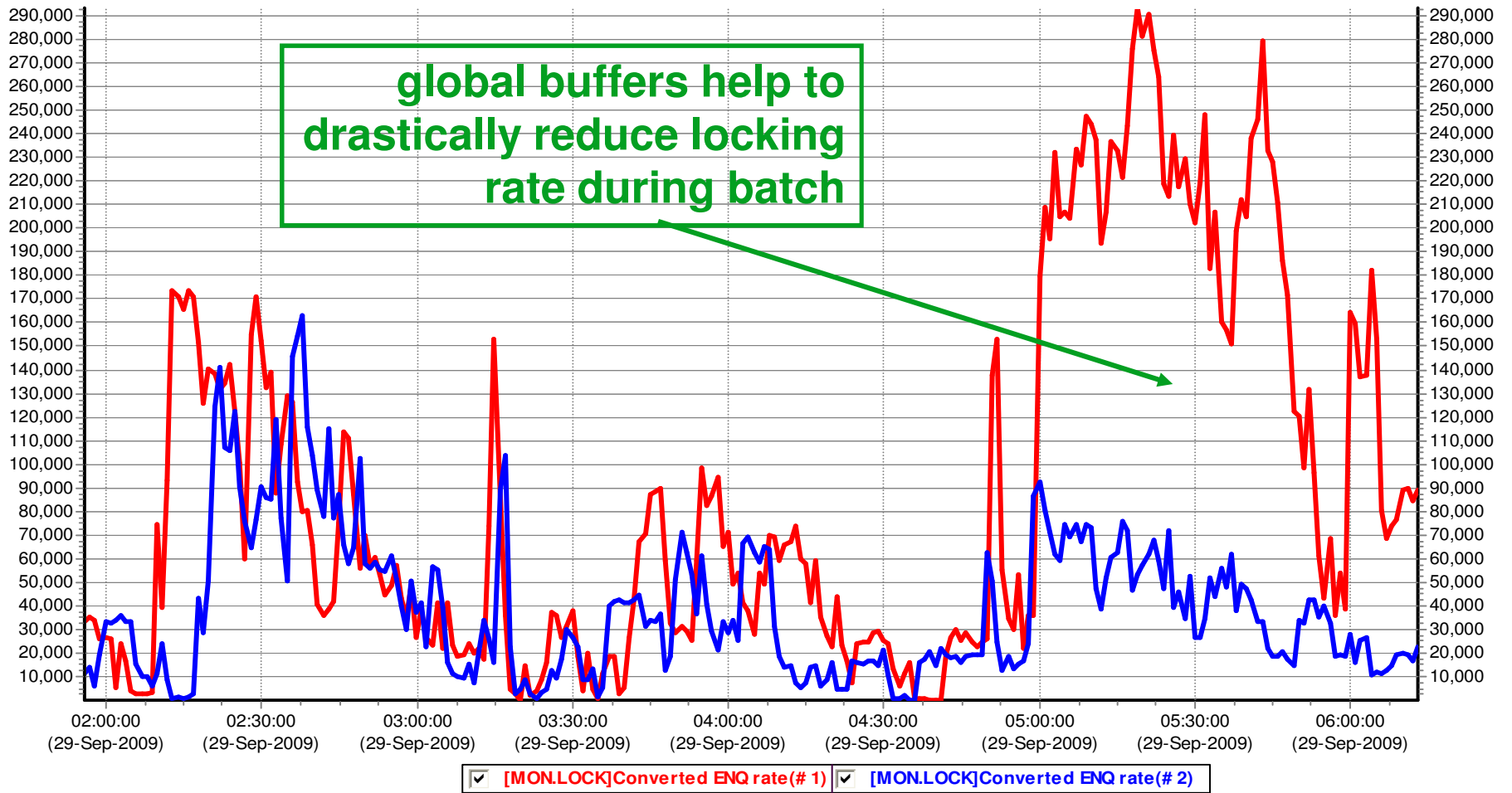MAKLEE

RMS Global Buffers

- To improve performance on hot RMS files, global buffers can help to reduce disk I/O activity

- SYSUAF and RIGHTSLIST are good candidates
  - XFC is disabled for files opened for write sharing on more than one node in a cluster.

- As a rule of thumb, single digit global buffers don't do good.

MAKLEE

Global Buffers in Action

Lock Conversion Rate (per sec)

Red = 28-Sep | Blue = 29-Sep

global buffers help to drastically reduce locking rate during batch

[MON.LOCK]Converted ENQ rate(# 1) [MON.LOCK]Converted ENQ rate(# 2)

Dedicated Lock Manager Busy

Dedicated Lock Manager Busy

Red = 28-Sep | Blue = 30-Sep
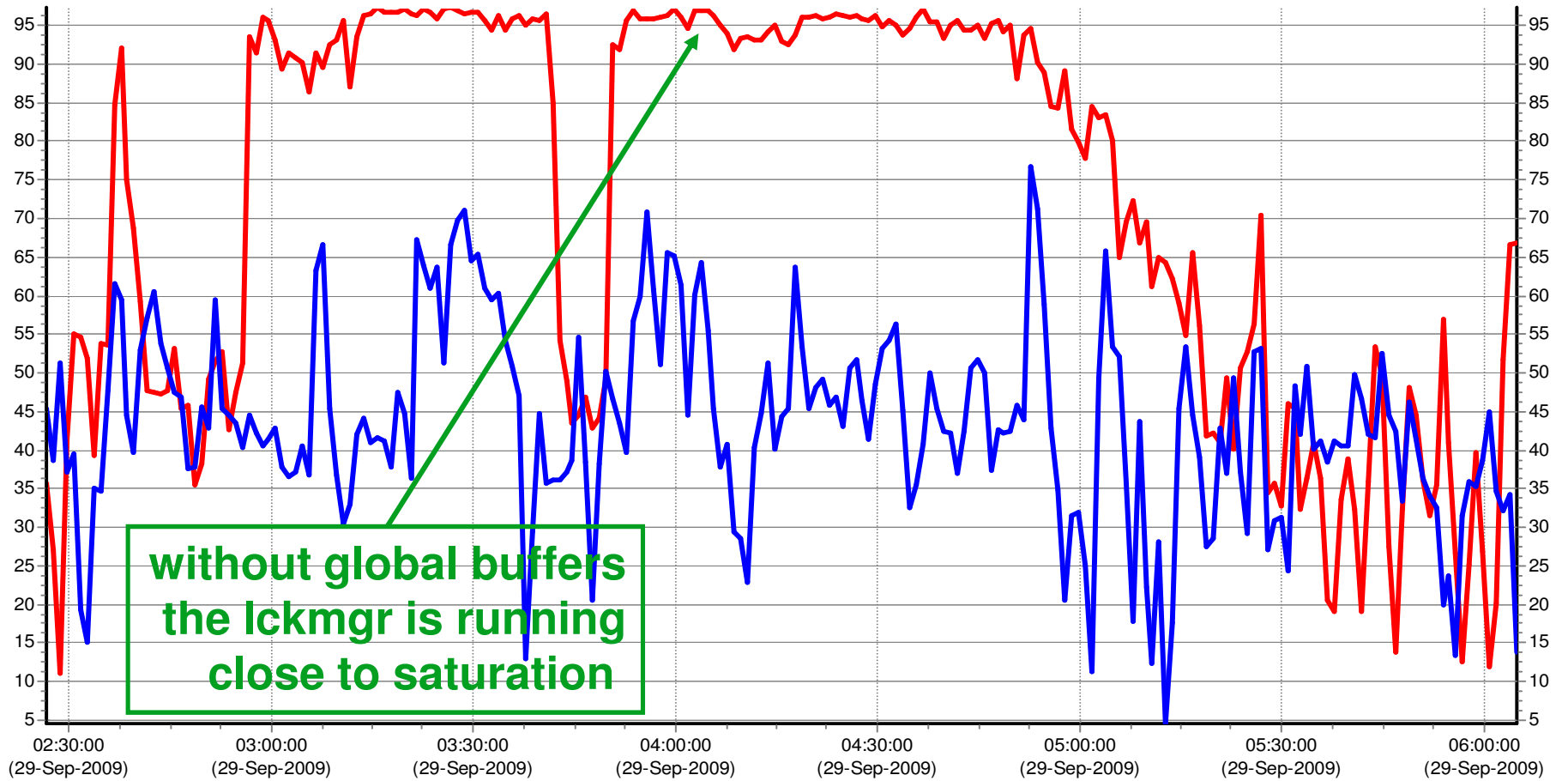
without global buffers
the lckmgr is running
close to saturation
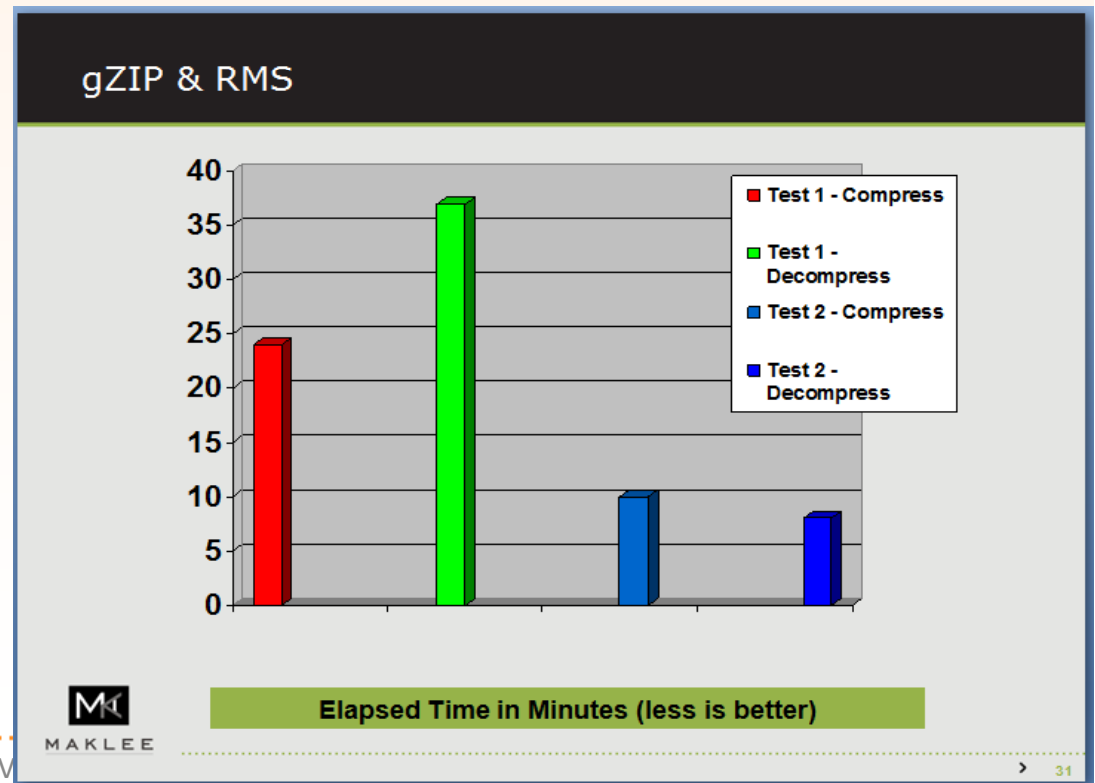
[LCK73]Busy %(# 1)   [LCK73]Busy %(# 2)

"Black Hole"

TCPIP Hostname Translation

› When TCPIP needs to translate a hostname to an IP address the order is

- SYS$SYSTEM:TCPIP$HOST.DAT
- TCPIP$ETC:IPNODES.DAT
- Bind Resolver

› For real time processes, verify all addresses are defined in the local host database.

MAKLEE

V8.4 – Performance Features

> RAD awareness (discussed separately).

> BACKUP compression is now documented.

> SET RMS allows increasing MBC up to 256 blocks.

gZIP & RMS

- Test 1 - Compress
- Test 1 - Decompress
- Test 2 - Compress
- Test 2 - Decompress

Elapsed Time in Minutes (less is better)

MAKLEE

SET RMS/BLOCK=250

```
$ show rms
         MULTI-   |                 MULTIBUFFER COUNTS                  | NETWORK
         BLOCK    | Indexed  Relative              Sequential          | BLOCK
         COUNT    |                       Disk   Magtape  Unit Record |  COUNT
Process  250      |    0         0          0       0         0        |    0
System   127      |    8         8          8       8         8        |   32


         Prolog    Extend Quantity    PATH_TIMEOUT
Process    0             0                  0
System     0             0                 10


         QUERY_LOCK        CONTENTION_POLICY
Process   System              System
System    Enabled             Never
```
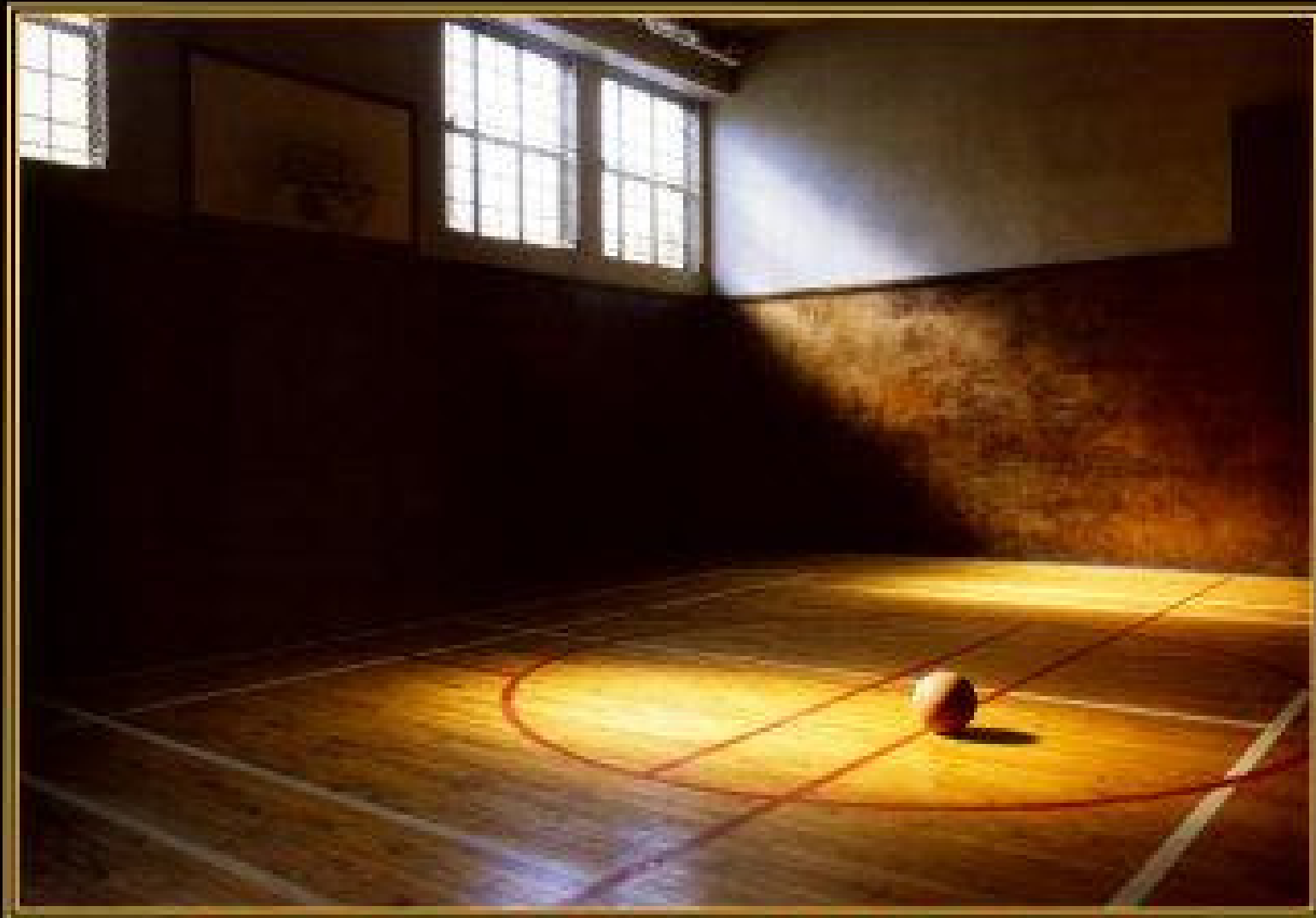
V8.4 UPDATE V4.0

- VMS84I_UPDATE V4.0 changes the default value of the SYSTEM_CHECKS parameter to 1.

- Running AUTOGEN using NOFEEDBACK will result in SYSTEM_CHECKS being set to 1.

- Significantly degrades performance.

- Make sure to set SYSTEM_CHECKS to 0.

MAKLEE

OPPORTUNITY

You will always miss 100% of the shots you do not take

37