Performance Tuning the OpenEdge Database in The Modern World

Gus Björklund, Progress

Mike Furgal, Bravepoint

PROGRESS EXCHANGE 2014

Performance tuning is not only about software configuration and turning knobs

Situation:

Your server is 5 years old

Vendor is raising support fees to get rid of old systems

What do you buy as a replacement ???

Hardware is cheap

Your new server will have:

Processors

Memory

Storage

Software

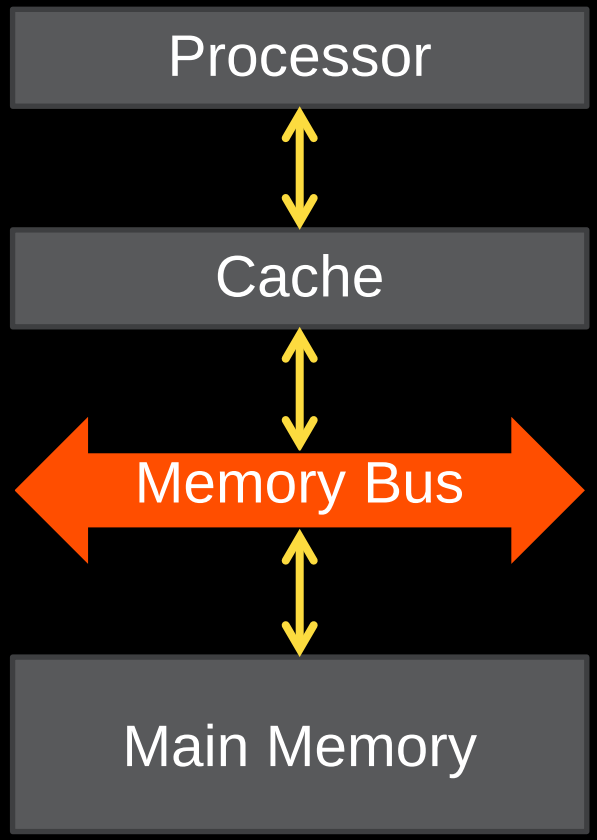
CPUs

Modern processors are very fast

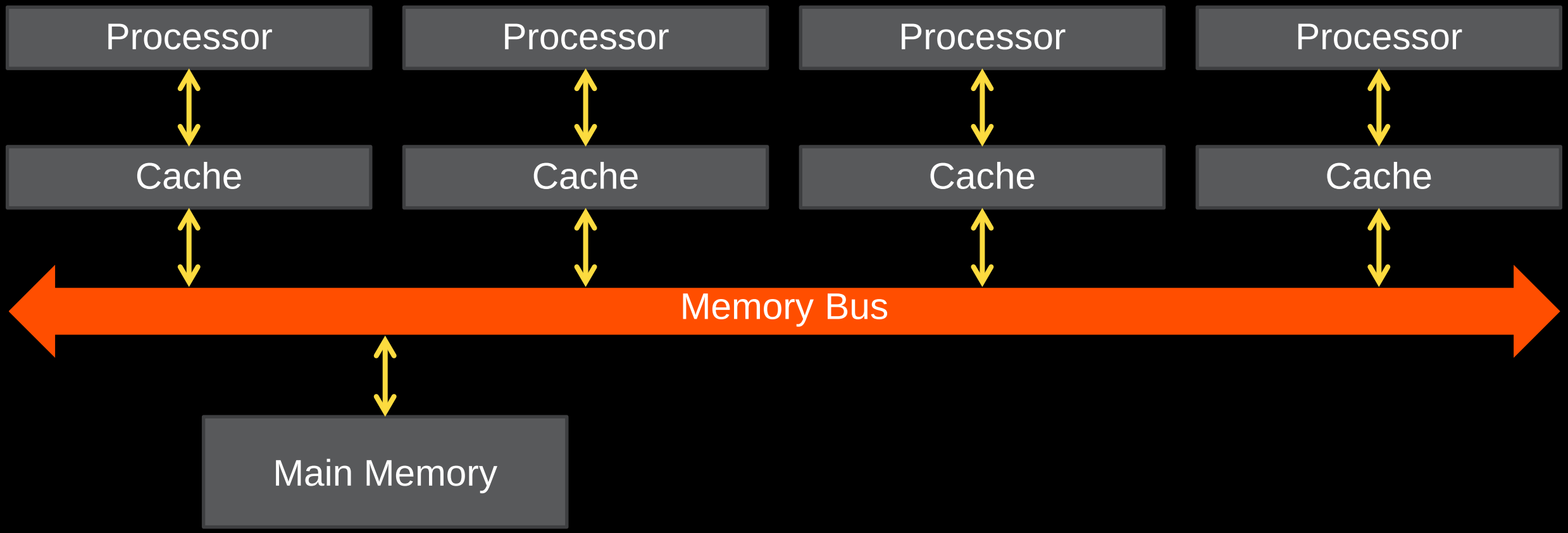
Single CPU machines
hardly exist anymore

You can have way more CPU
power than you can ever use

Simple Single Processor Architecture One Level High Speed Cache Memory



Multi-Processor Architecture

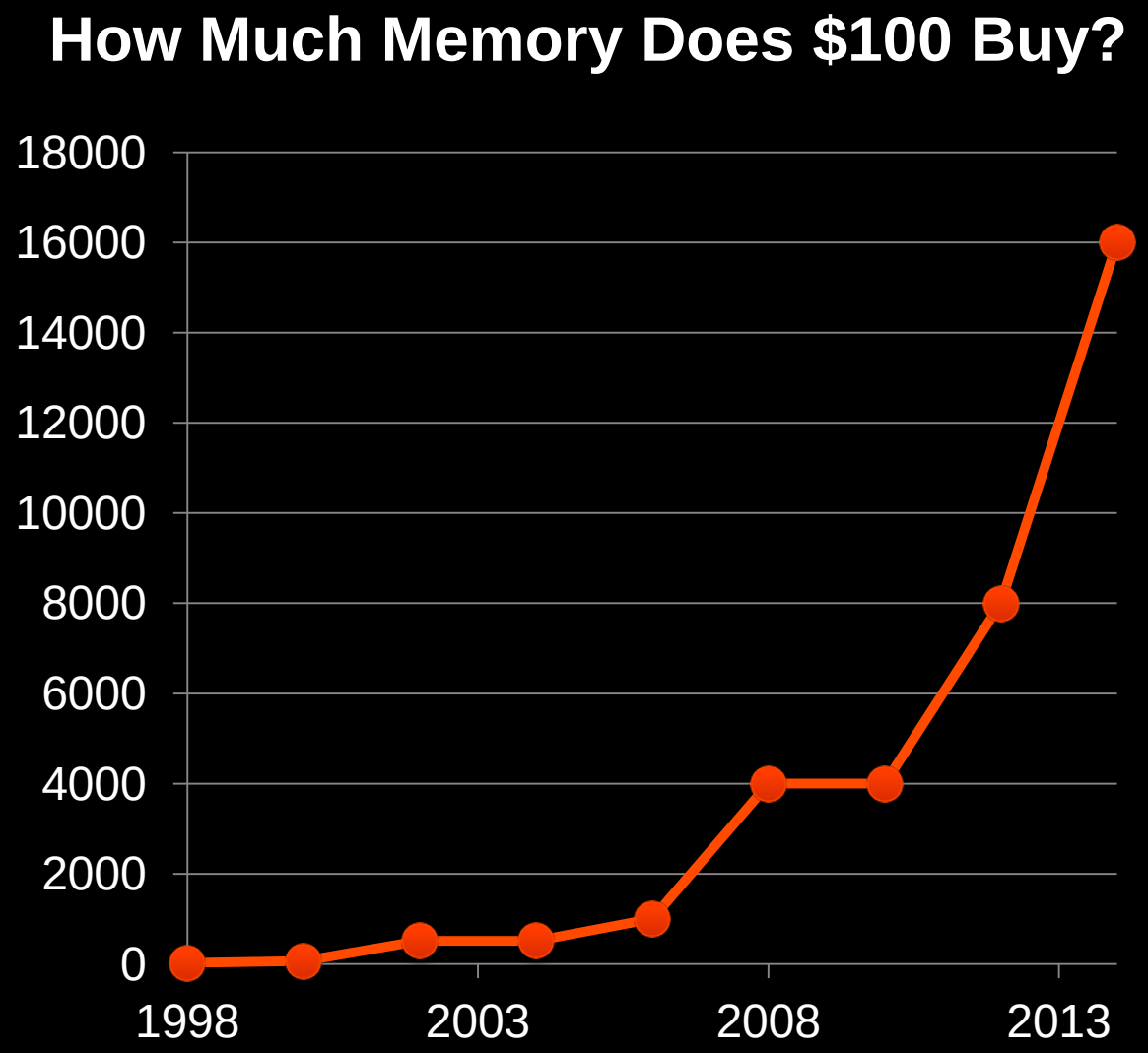


A technique to avoid Cache Coherency issues

Lessen the number of processes connected directly to shared memory

Main Memory

Memory prices have dropped significantly over the past years. For example in the year 2000, 64 MB of memory cost $100. In 2010 for $100 you could get 4 GB of memory. Today (2014) that same $100 gets you about 16 GB of memory



Main Memory

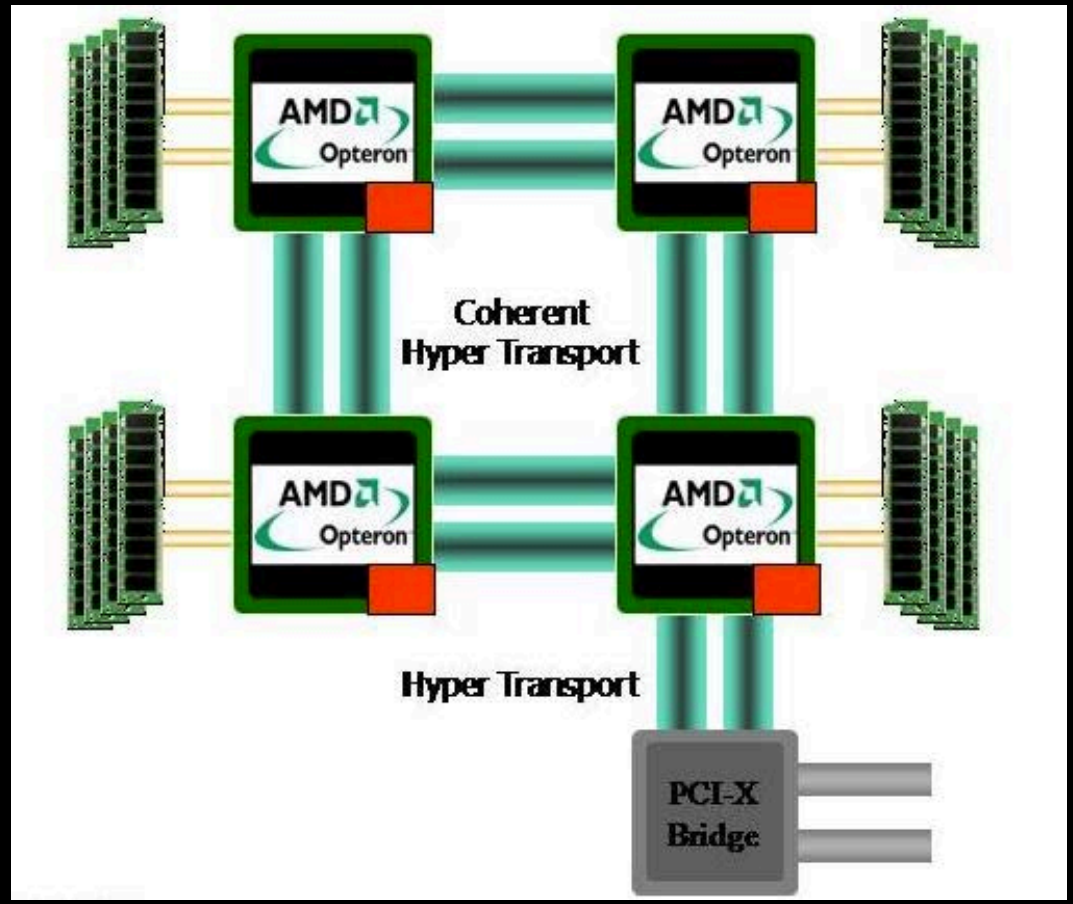
The least expensive way to enhance performance

Buy as much as you can

NUMA

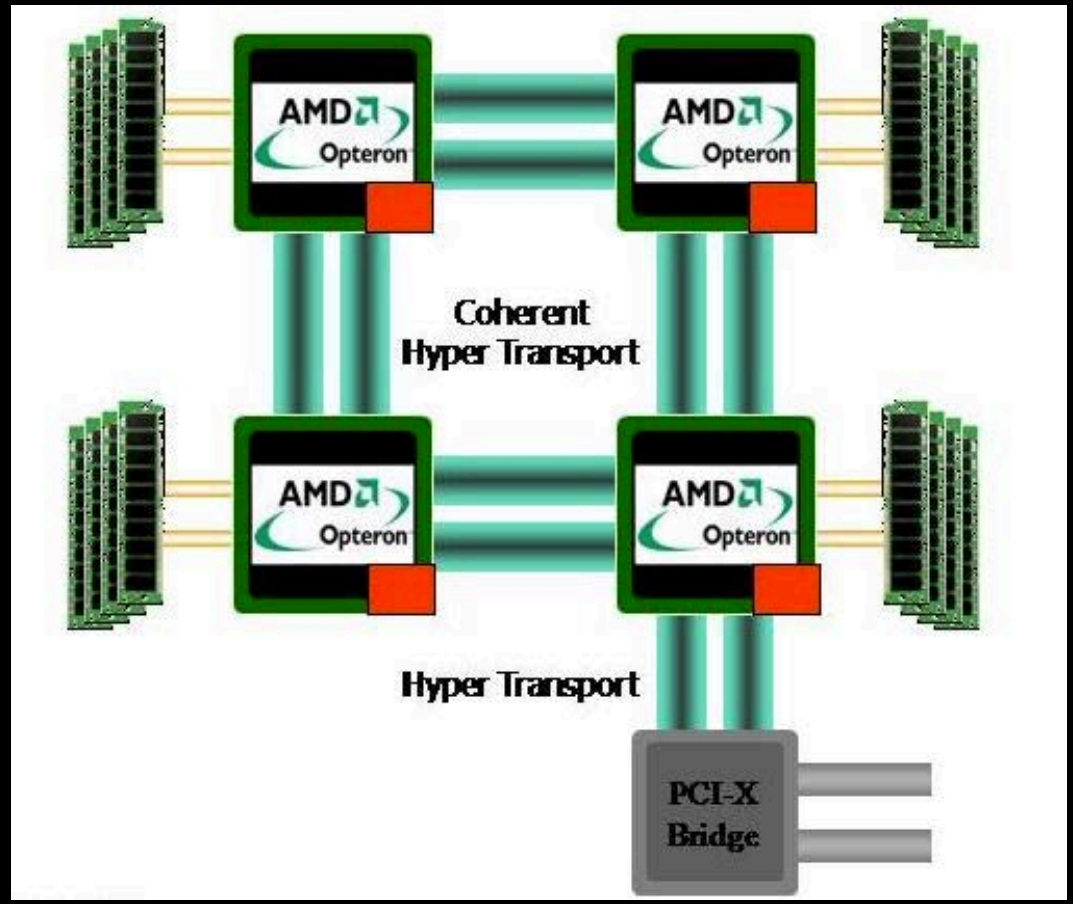
NUMA Stands for Non-Uniform Memory Access

In layman's terms, a NUMA machine is the coupling of several machines in a single physical unit, running a single Operating System. Like a "cluster" (if you squint)

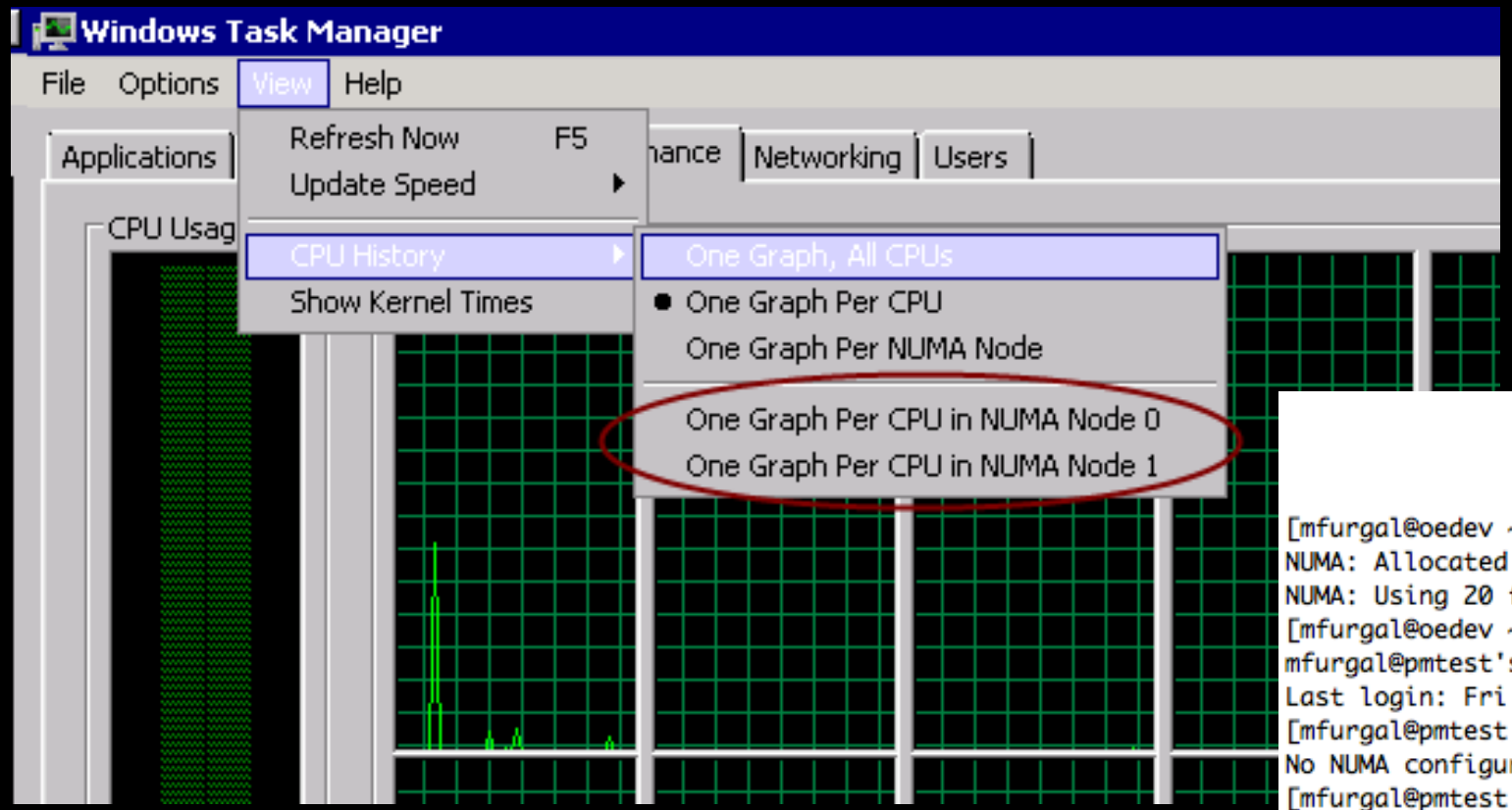


The NUMA Quotient

This is the time it takes for a CPU to read memory on a remote node as compared to reading memory locally



How Do You Know if You Have a NUMA Machine?




```
[mfurgal@oedev ~]$ dmesg | grep -i numa
NUMA: Allocated memnodemap from 11000 - 12040
NUMA: Using 20 for the hash shift.
[mfurgal@oedev ~]$ ssh pmtest
mfurgal@pmtest's password:
Last login: Fri Aug 29 13:08:32 2014 from 172.16.61.127
[mfurgal@pmtest ~]$ dmesg | grep -i numa
No NUMA configuration found
[mfurgal@pmtest ~]$
```

So now you know you have a NUMA machine

Is all hope lost?

On some machines you can pin memory and processes to a particular node

On some you can disable nodes but may lose memory too

Bottom line – don’t buy a NUMA machine

Storage

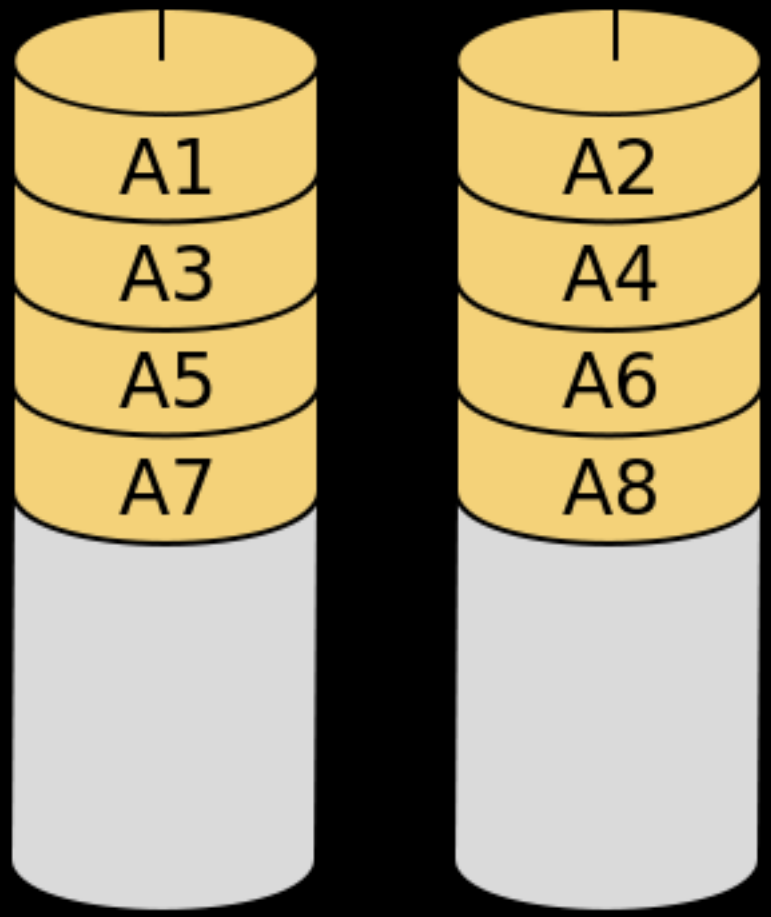
RAID

RAID

Why?

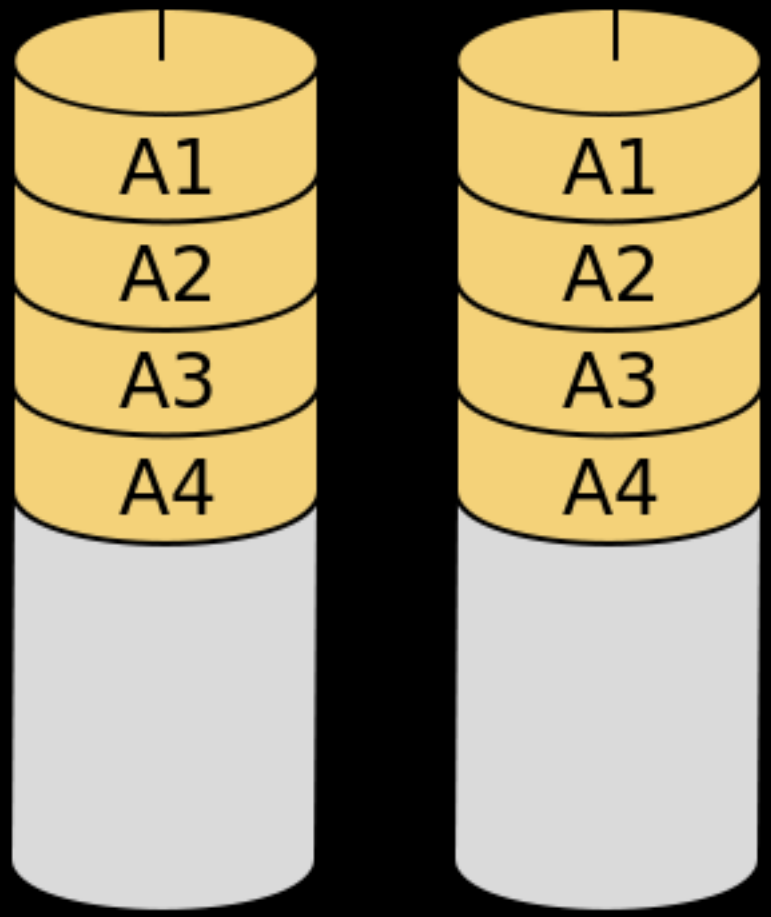
RAID 0: disk striping

performance but NO reliability



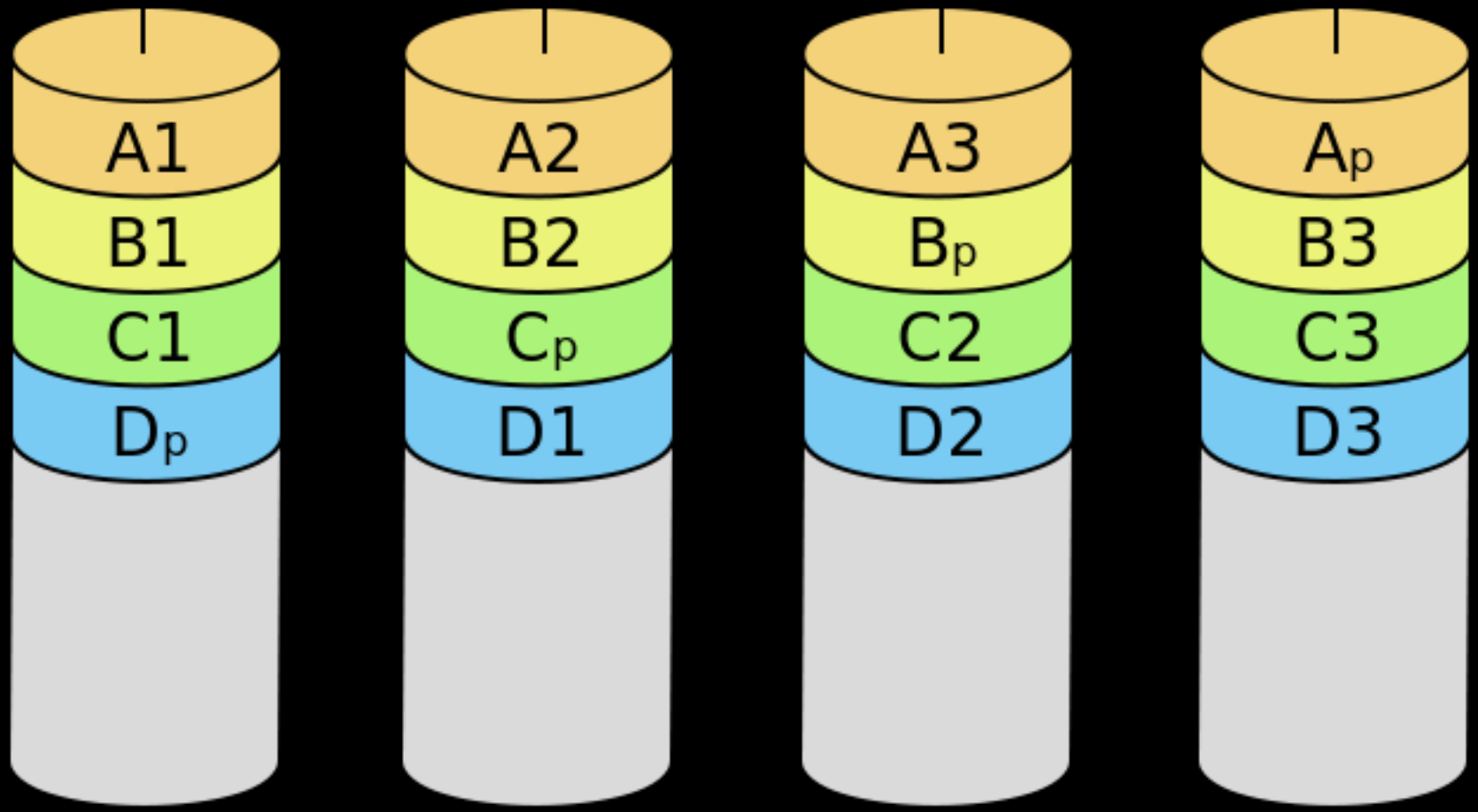
RAID 1: disk mirroring

reliability – two copies



RAID 5: disk striping with parity

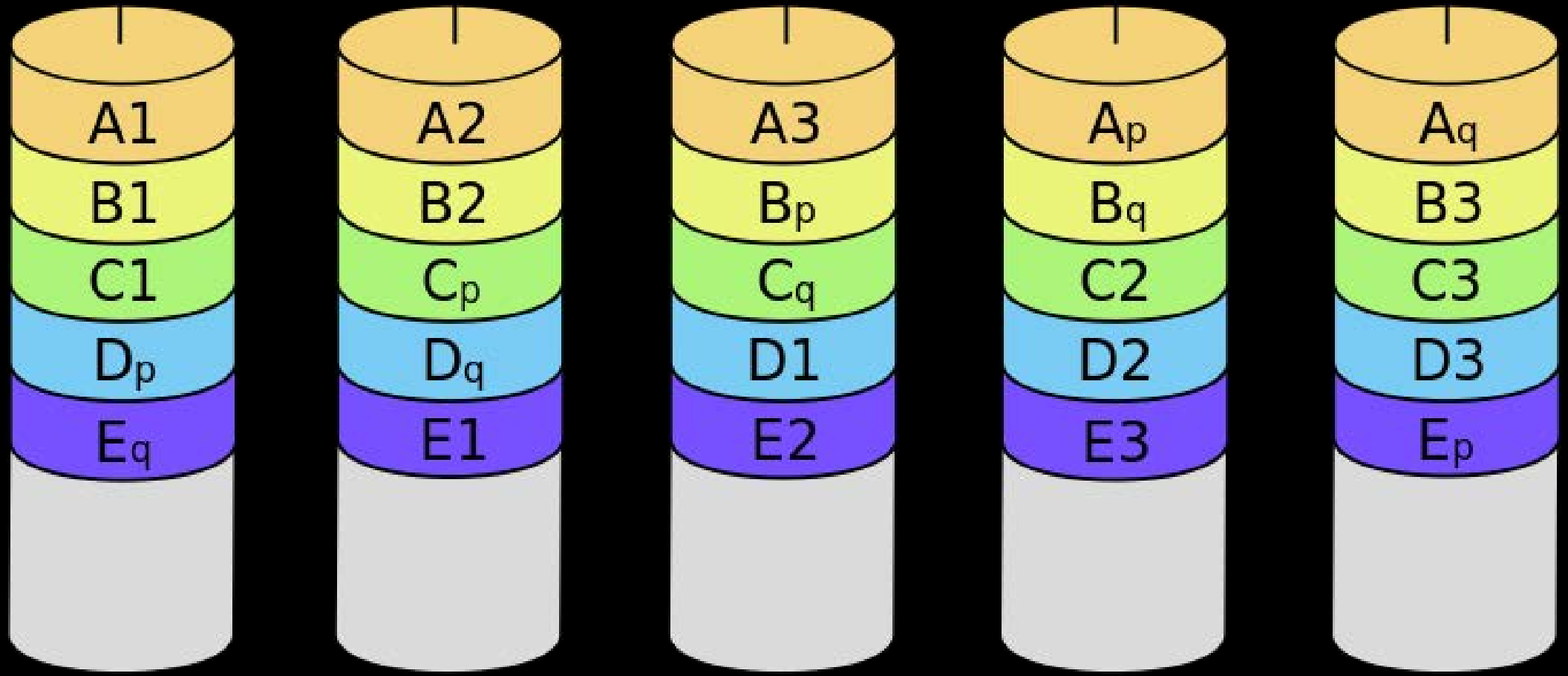
reliability and bad performance



all writes update 2 drives

RAID 6: disk striping with two parity disks

reliability and worse performance



all writes update 3 drives

RAID (Redundant Arrays of Inexpensive Disks)

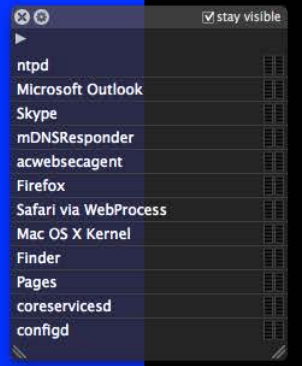
Type	Description	Use?
RAID 0	Block striping (no redundancy at all)	Bad
RAID 1	Mirroring	OK
RAID 10	Block striping + mirroring	Excellent
RAID 2	Bit level striping, dedicated parity	Bad
RAID 3	Byte level striping, dedicated parity	Bad
RAID 4	Block striping, dedicated parity	Bad
RAID 5	Block striping with striped parity	Poor
RAID 6	Block striping with dual striped parity	Poor
RAID 60, 6+, DP, etc.	Marketing	Poor

RAID (Redundant Arrays of Inexpensive Disks)

Type	Description	Use?
~~RAID 0~~	~~Block striping (no redundancy at all)~~	~~Bad~~
~~RAID 1~~	~~Mirroring~~	~~OK~~
RAID 10	**Block striping + mirroring**	**Excellent**
~~RAID 2~~	~~Bit level striping, dedicated parity~~	~~Bad~~
~~RAID 3~~	~~Byte level striping, dedicated parity~~	~~Bad~~
~~RAID 4~~	~~Block striping, dedicated parity~~	~~Bad~~
~~RAID 5~~	~~Block striping with striped parity~~	~~Poor~~
~~RAID 6~~	~~Block striping with dual striped parity~~	~~Poor~~
~~RAID 60, 6+, DP, etc.~~	~~Marketing~~	~~Poor~~

- Advancements in technology can never make a silk purse from the RAID 5 sow's ear
- Local disks beat SAN storage

SSD



Windows

A fatal exception 0E has occurred at 0028:C00068F8 in VxD VMM(01) + 000059F8. The current application will be terminated.

* Press any key to terminate the application.
* Press CTRL+ALT+DEL to restart your computer. You will lose any unsaved information in all applications.

Press any key to continue

SSD

- Prices have dropped – a LOT. Low end is $0.50 per gigabyte
- Reliability is now very good – better than spinning rust
- SSD devices are fast, and getting faster
- Use Mirrorred pairs – NO RAID 5
- When you need to replace one, you may not be able to get matching units anymore
- Fetching a record that is already in the database buffer pool is 75 times faster than SSD !!!!

Time to Grow a 96 MB File		
Disk Type	**Duration**	**Speed**
Spinning Disk	7–10	9–13 MB/Sec
SSD	1–2	43–96 MB/Sec

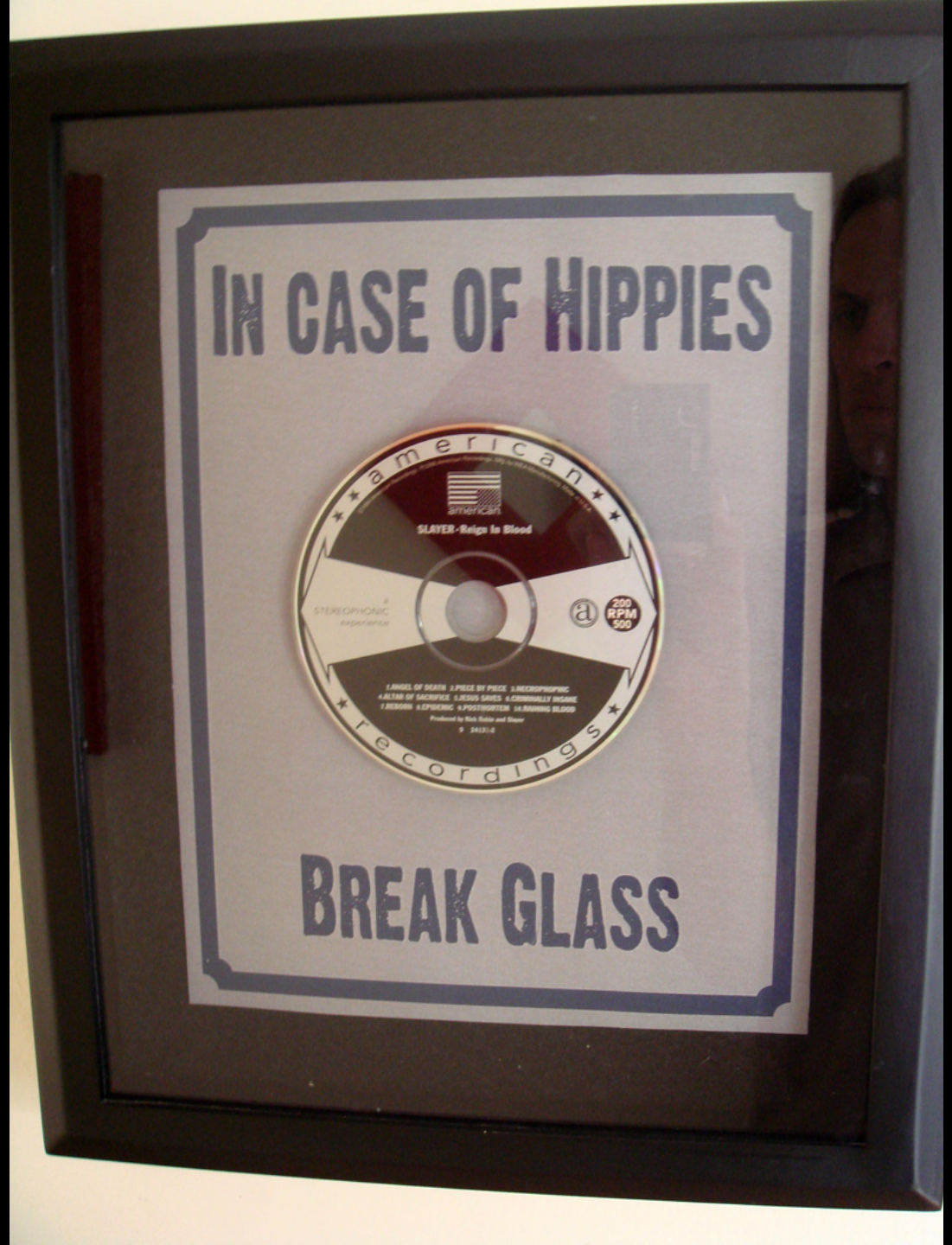
… in Big B You Should Trust!

Layer	Time	# of Recs	# of Ops	Cost per Op	Relative
Progress to –B	0.96	100,000	203,473	0.000005	1
-B to FS Cache	10.24	100,000	26,711	0.000383	75
FS Cache to SAN	5.93	100,000	26,711	0.000222	45
-B to SAN Cache*	11.17	100,000	26,711	0.000605	120
SAN Cache to Disk	200.35	100,000	26,711	0.007500	1500
-B to Disk	211.52	100,000	26,711	0.007919	1585

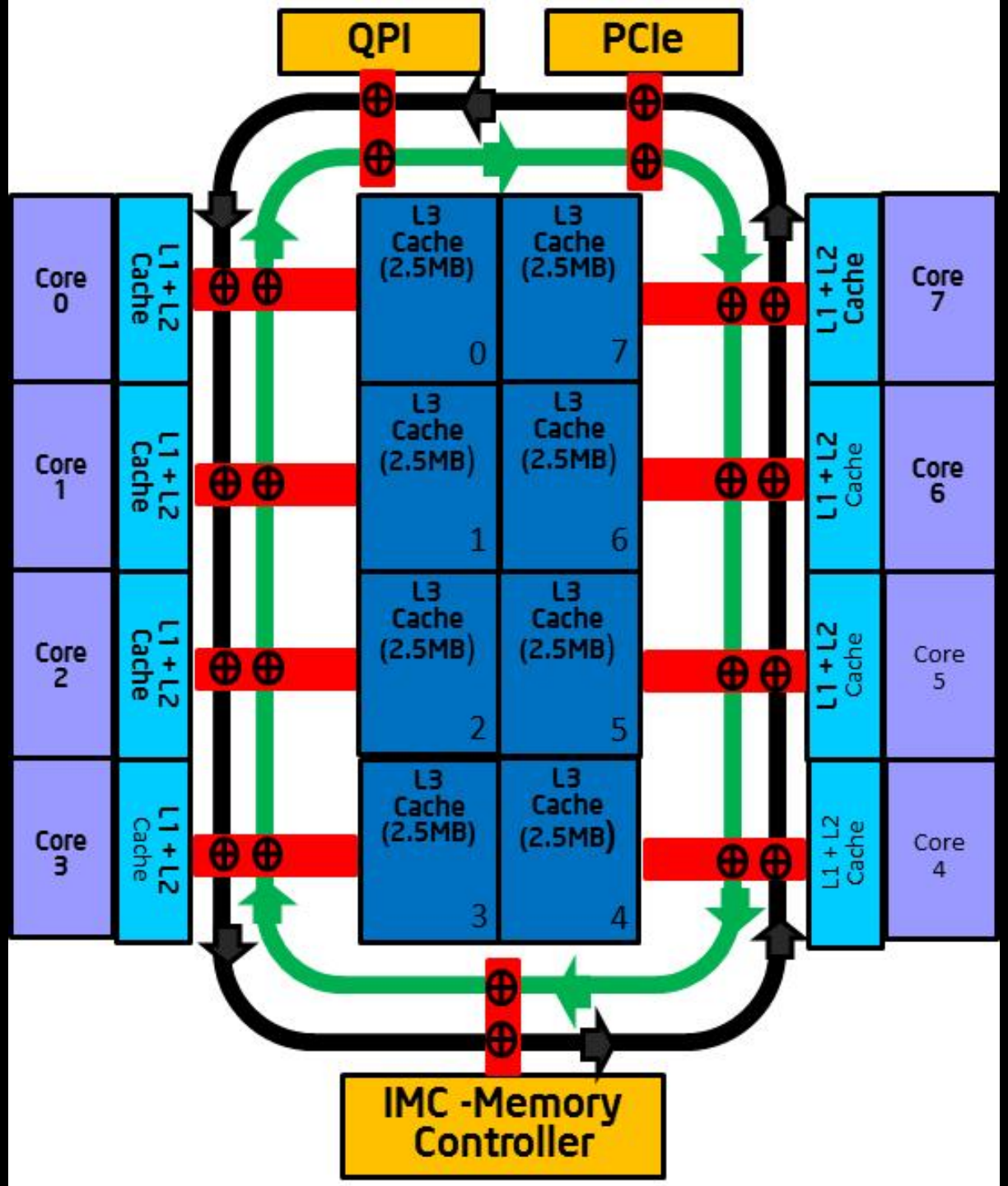
* Used concurrent IO to eliminate FS cache

Courtesy of Tom Bascom

Mid-range server replacement example


IN CASE OF HIPPIES
american
SLAYER - Reign In Blood
STEREOPHONIC
200 RPM 500
1.ANGEL OF DEATH 2.PIECE BY PIECE 3.NECROPHOBIC
4.ALTAR OF SACRIFICE 5.JESUS SAVES 6.CRIMINALLY INSANE
7.REIGN 8.EPIDEMIC 9.POSTMORTEM 10.RAINING BLOOD
Produced by Rick Rubin and Slayer
9 24131-4
american recordings
BREAK GLASS

Name	Quantity	Value
CPU	4	Intel Xeon E5 4603, 8 cores
RAM	8	1866MT/s 4 GB RDIMM
Ether	1	Intel GB Ethernet Card
Disk Controller	1	PERC H10
Storage, hot plug	8	146 GB 15,000 rpm SAS
Stuff	?	dual psu. case, etc.
Operating system	1	Linux, not included


QPI
PCIe
Core 0
L1 + L2 Cache
L3 Cache (2.5MB) 0
L3 Cache (2.5MB) 7
L1 + L2 Cache
Core 7
Core 1
L1 + L2 Cache
L3 Cache (2.5MB) 1
L3 Cache (2.5MB) 6
L1 + L2 Cache
Core 6
Core 2
L1 + L2 Cache
L3 Cache (2.5MB) 2
L3 Cache (2.5MB) 5
L1 + L2 Cache
Core 5
Core 3
L1 + L2 Cache
L3 Cache (2.5MB) 3
L3 Cache (2.5MB) 4
L1 + L2 Cache
Core 4
IMC -Memory Controller

Select Components

1. COMPONENTS | 2. SERVICES & ACCESSORIES

PowerEdge R820

Starting Price	$12,962.00
Instant Savings	$3,637.09
Subtotal	**$9,324.91**

Modern OpenEdge RDBMS

Advanced Tuning Techniques

PROGRAMMING

Nothing else matters if you're good at it.

Get Current

Better be on 10.2B08 or later

-lruskips

-B2

Client Database-Request Statement Caching

Procedure Call Stack

- Top is last procedure executed
- Bottom is first procedure executed
- Top down, newest to oldest
- One time full stack
- Continuous full stack
- Continuous current location

Top

Newest

#	Procedure Name	File Name
19	: reallyLongNamedInternalProcedure3	proctestb.r
12	: reallyLongNamedInternalProcedure2	proctestb.r
5	: reallyLongNamedInternalProcedure1	proctesta.r
445	: reallyLongNamedInternalProcedure0	proctesta.r
1	: /usr1/stmtest/p72340_Untitled1.ped	

Oldest

Bottom

table partitioning

index rebuild

Index Rebuild Performance (OE 10.2B06, OE 11.2)

-TB	sort block size (8K – 64K, note new limit)	64
-datascanthreads	# threads for data scan phase	1.5 X #CPUs
-TMB	merge block size (default -TB)	64
-TF	merge pool fraction of system memory (in %)	80%
-mergethreads	# threads per concurrent sort group merging	X -threadnum = 1.5 X #CPUs
-threadnum	# concurrent sort group merging	2 or 4
-TM	# merge buffers to merge each merge pass	32
-rusage	report system usage statistics	-rusage
-silent	a bit quieter than before	-silent

Index Rebuild Performance (OE 10.2B06, OE 11.2)

-TB	sort block size (8K – 64K, note new limit)	64
-datascanthreads	# threads for data scan phase	PUs
-TMB	merge block size (def	
-TF	mer	
		1.5 X #CPUs
		2 or 4
	ge each merge pass	32
	report system usage statistics	-rusage
-silent	a bit quieter than before	-silent

12.5 hours ➔ 2.5 hours
5x improvement!

-omsize

How to Manage Object Mapping Cache

- Do I have a problem?

```
define variable prev-latches as integer.
repeat:
  find _latch where _latch-name = "MTL_OM".
  display _Latch-Name
          _Latch-Lock   /* # times latch acquired */
          _Latch-Wait   /* # time conflict occurred */
          _Latch-Lock - prev-latches label "latch/sec".
          prev-latches = _Latch-Lock.
  pause 1.
end.
```

The dawn rises only when the rooster crows.

– Burmese proverb

Answers

email:

gus@progress.com

mfurgal@bravepoint.com

PROGRESS EXCHANGE 2014

Visit the Resource Portal

- Get **session details** & presentation **downloads**
- Complete a **survey**
- Access the latest Progress **product literature**

www.progress.com/exchange2014

PROGRESS