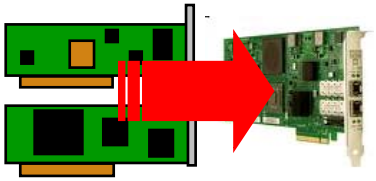
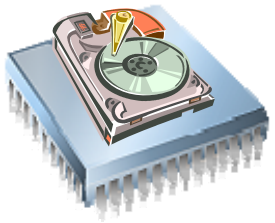
Ten Smarter Technologies coming to your … computer room & career

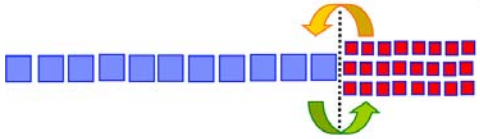
Systems Director
-Single Screen
- Update Mgr
- Automation Mgr

FCoCEE

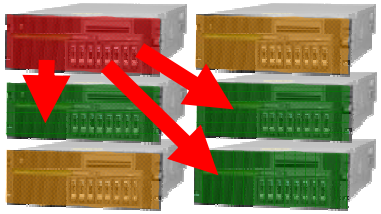


Solid State Drive





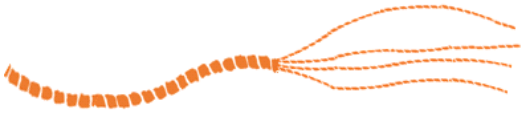
Cluster Automatic Balancing



Cluster Evacuation

Active Energy Manager

5 GHz barrier means multi-thread or die



Versioned WPAR

Pre-Built Solutions

This session makes no announcements what-so-ever

Any prices are quoted in an unknown currency !

The opinions expressed are Nigel's

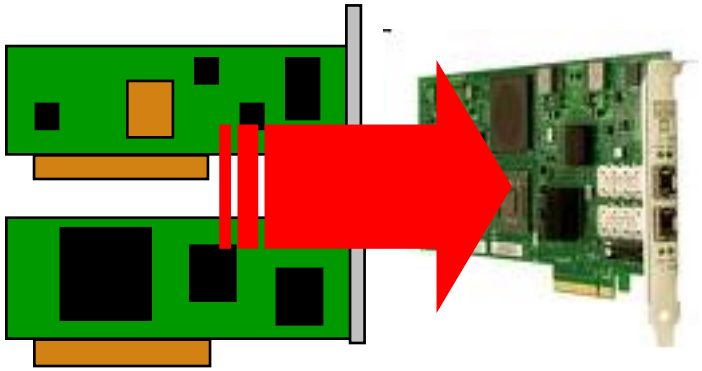
(Not IBM's)

Going to take for granted the "POWER Always"

- Reliability, Availability, Serviceability
  - From the guys that do mainframes
- Scaling for Consolidation of 100's of workloads
  - Powerful POWER processors
  - Power 795 – 256 cores with SMT=4
- Virtualisation
  - Virtual CPU – high utilisation
  - Virtual Disk – reduced costs, hardware independence
  - Virtual Network – elite
  - Virtual Optical & Tape

Already Delivered
What is Next?

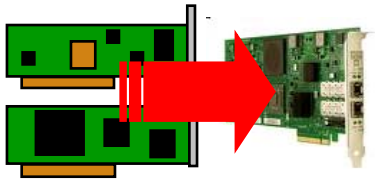
FCoCEE
Fibre Channel over Converged Enhanced Ethernet



Redpaper

An Introduction to FCoE and FCoCEE

www.redbooks.ibm.com/redpapers/pdfs/redp4493.pdf



FCoCEE
Fibre Channel over Converged Enhanced Ethernet

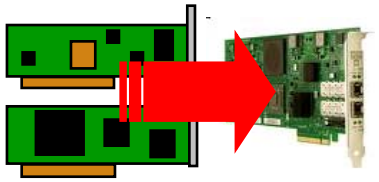
8 Gb FC/SAN

10 Gb Ethernet

Routing Information

Packet + Sum check

Data



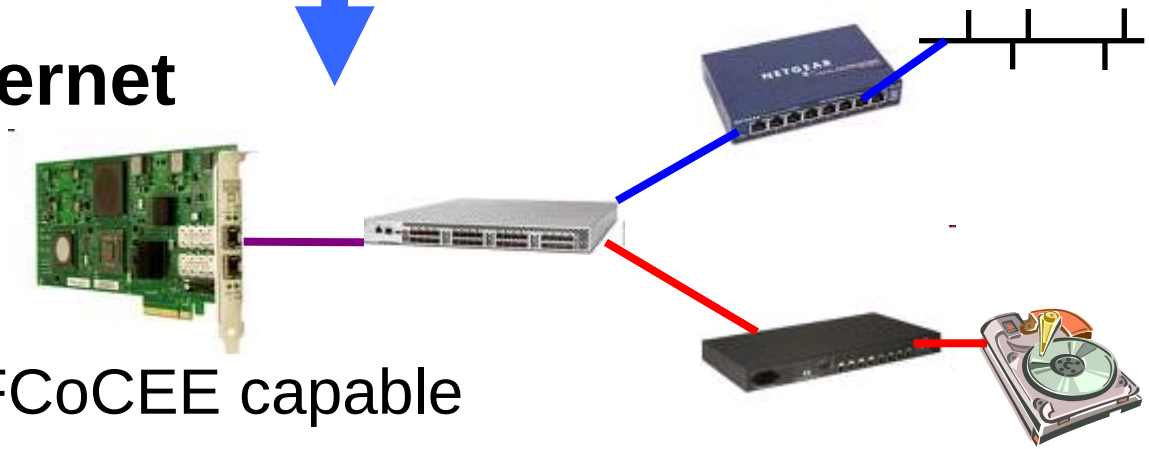
FCoCEE
Fibre Channel over Converged Enhanced Ethernet

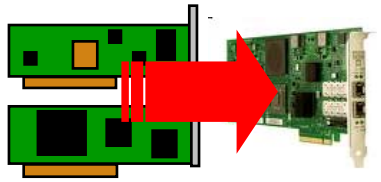
Ethernet 10 Gb

Fibre Channel 8 Gb

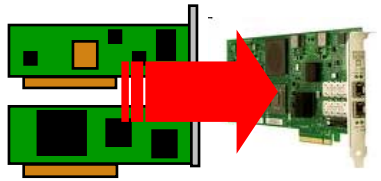
FCoCEE 10 Gb Ethernet

- Less PCIe slots
- Less Adapters
- Less cables
- Less switches but FCoCEE capable





But what about the cost?

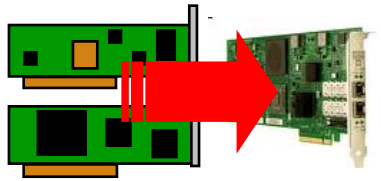


FCoCEE Adapters DON'T cost more

FCoCEE	List price in €$£
5708 10Gb FCoE PCIe Dual Port Adapter	4786
High Speed Ethernet	
5732 10 Gigabit Ethernet-CX4 PCIe	4492
5769 10 Gigabit Ethernet-SR PCIe	4960
5772 10 Gigabit Ethernet-LR PCIe	5549
High Speed Fibre Channel	
5735 8 Gigabit PCIe Dual Port Fibre Channel	3687

FCoCEE adapter config costs much less

2 x 8Gbit FC + 2 10 Gbit Ethernet	16358 = 70% more
2 x FCoCEE	9572



FCoCEE Adapters DON'T cost more

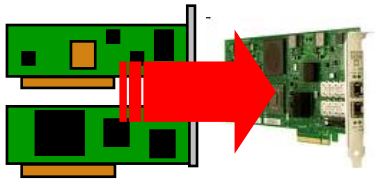
		List price in €$£
Low or Medium Speed		
5717	4-Port 10/100/1000 Base-TX PCIe	1013
5767	2-Port 10/100/1000 Base-TX PCIe	628
5768	2-Port Gigabit Ethernet-SX PCIe	1383
5774	4 Gigabit PCIe Dual Port Fibre Channel	2766

Even medium speed config costs roughly the same:

2 x 4Gbit FC + 4 x 1Gbit Ethernet	9588
(i.e. 2 x 2766 + 4 x 1013)	
2 x FCoCEE	9572

Bandwidth

Old style	12 GBit
FCoCEE	20 Gbit



FCoCEE Summary

For:

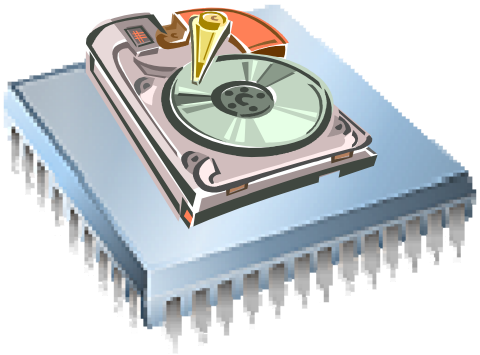
- Reduced costs
  - 50% less adapters, PCIe slots & remote I/O drawers
  - 50% less cables & switch ports
- Less floor space, less electricity & less heat
- Higher RAS - 50% less failures

Against

- Switches cost ~20% more than 10Gb + 8Gb
- Infrastructure changes happen slowly
- Politics & interdepartmental fighting!

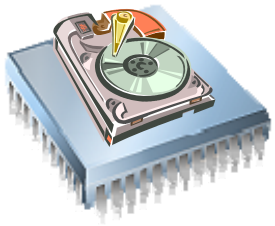
Solid State Disks

Disk storage on a chip



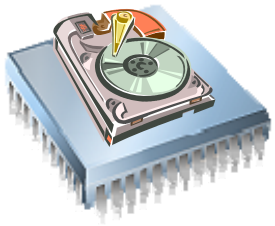
AIX SSD Wiki Page

http://www.ibm.com/developerworks/wikis/display/WikiPtype/Solid+State+Drives



Solid State Disks

- Extremely fast
  - 150 to 250 times faster than a hard disk
- Fairly expensive but price trend is going down
- Three ways of connecting it …

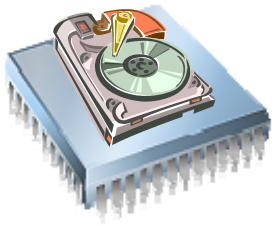


Three ways of connecting SSDs

1 Within SAN disk subsystem

- New option in the “fancy” SAN storage Sub-systems
- Speed & density fits between RAM cache & disks
- Main advantage = SAN aids availability

For more info talk to a Storage expert



Three ways of connecting SSDs

2 SAS 2.5 or 3.5 inch Hard Drive

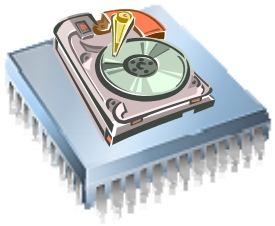
Internal to POWER System

I/O Drawer

SAS Drawer

- Main “issue” it is too fast! SAS controller saturation
- Size = 69 GB supported Power 6 & 7

Introduced in 2009



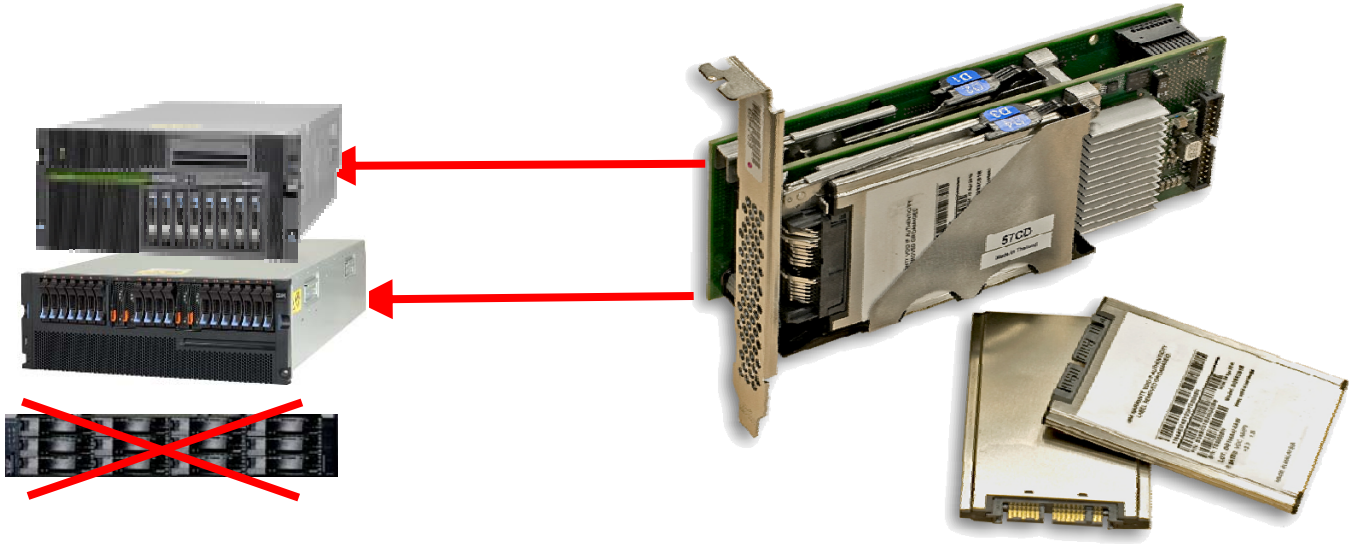
Three ways of connecting SSDs

3 PCIe Adapter with onboard storage

Internal to POWER System

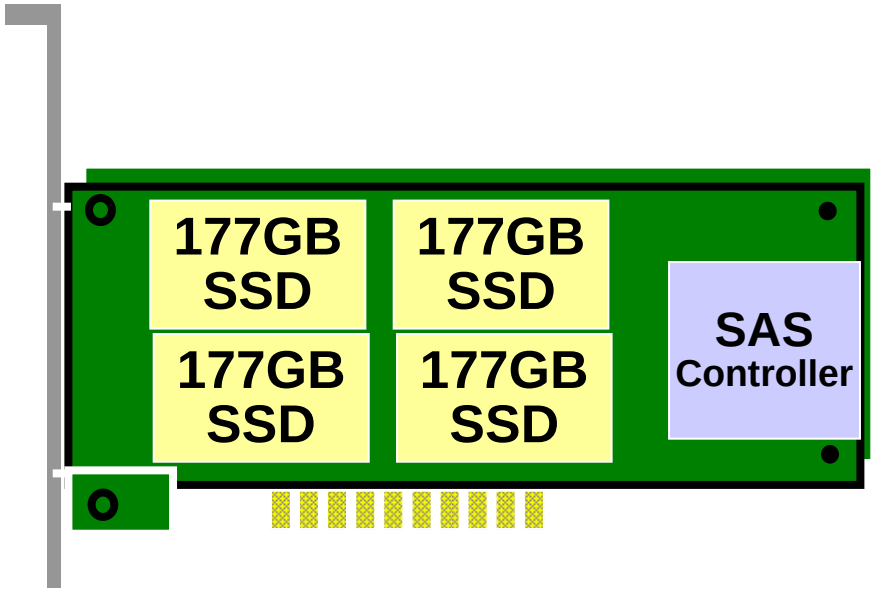
I/O Drawer

SAS Drawer



Introduced in August 2010

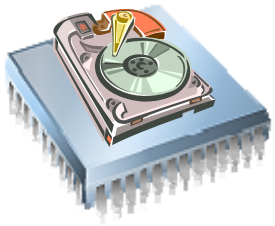
PCIe-Based SSD



Double-wide PCIe SAS Adapter

- **1, 2 or 4 SSD modules per adapter**
- **RAID 177 to 708 GB per card**
- **JBOD 200 to 800 GB per card**
  - **250 times faster than a hard disk & 32 times more expensive**
- **Power 7 machines (not 795)**

Introduced in August 2010



Solid State Drive Summary

For

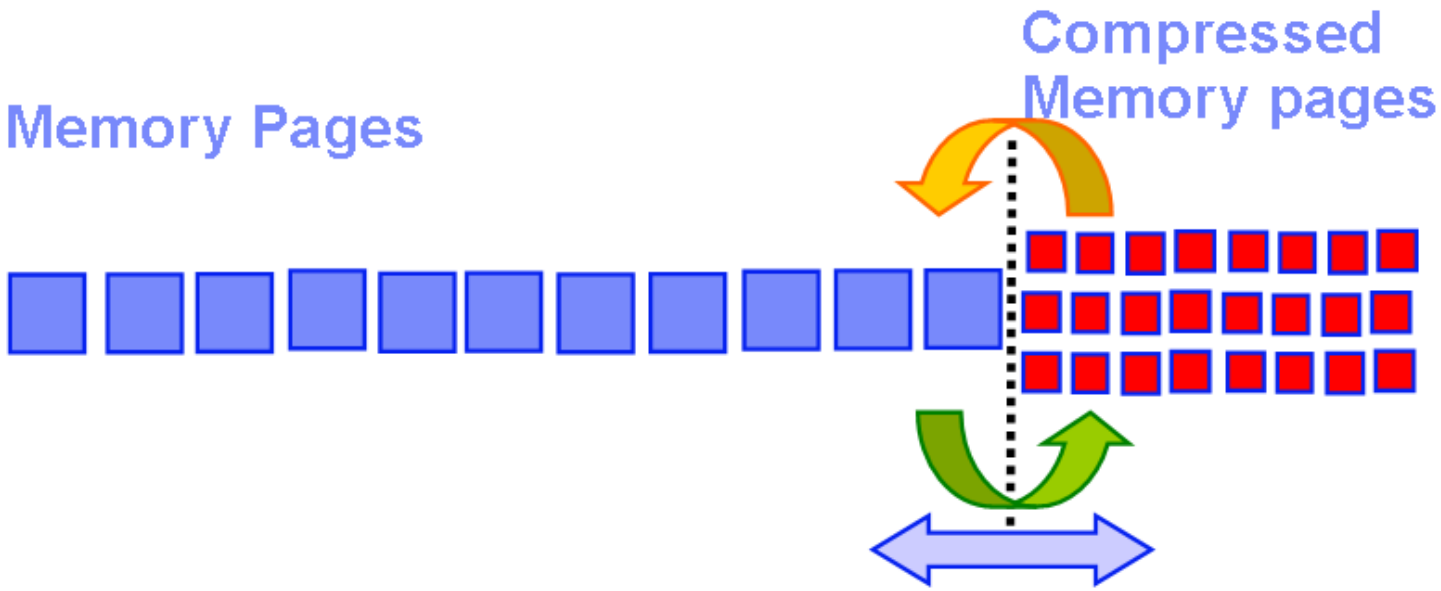
1. Extremely Fast I/O
2. Very low power & cold
3. Use for “hot” data like
   - critical database tables, sort area, temporary area
4. Use for “money is not an issue” problems
5. Will go round corners!

Against

- Expensive but already reducing as sizes rise

Active Memory Expansion (AME)

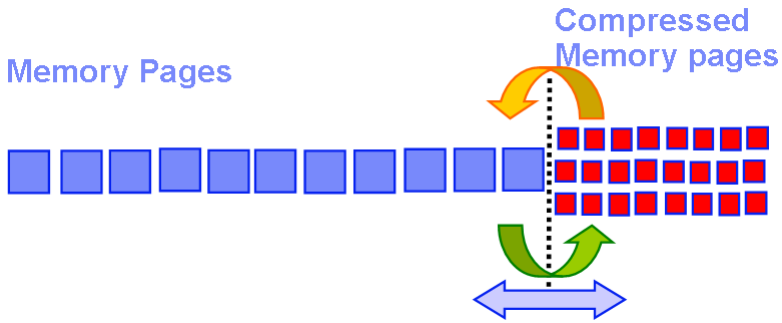
- Expansion by Compression



New AME external Wiki
http://www.ibm.com/developerworks/wikis/display/WikiPtype/IBM+Active+Memory+Expansion

Active Memory Expansion Highlights

1. POWER7 & AIX 6 TL04 SP2+
2. Activation Key by machine plus One-time Trail
3. Set at LPAR level & dynamic Expansion Factor
4. Memory shrink (release RAM) or
   Memory grow (performance)
5. Trade-off in more memory but some CPU use
6. Watch the movie Google: “AIX movies”



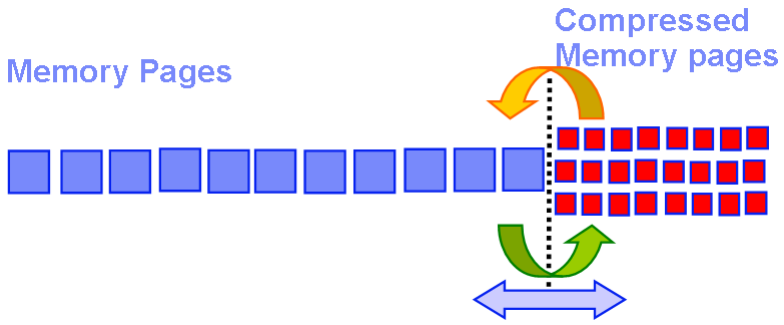
Active Memory Expansion Summary

For

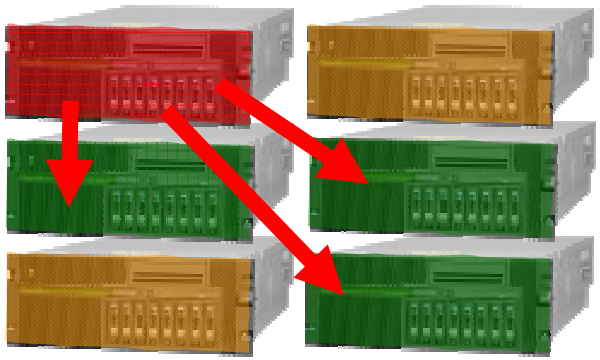
- Very simple to operate
- No-brainer performance improvements
- Fight the Java bloat
- Large benefits for RAM limited applications

Against

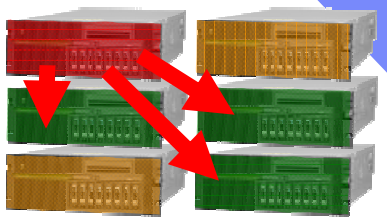
- Purchasing cost



Automatic Cluster Load Balancing



Automatic Cluster Load Balancing



I would argue

No one should buy 1 computer from IBM

Nor 2 computers for high availability

…

They should buy SIX !

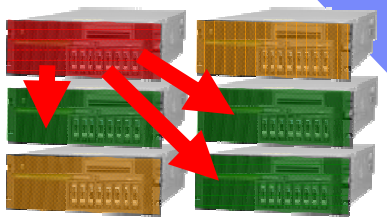
OK perhaps 4 growing up to 6 or multiple 6 packs

Think CLUSTER

1. Zero single point of failure
2. Smaller boxes are cheaper
3. Plenty HA options
   - 3 + 3
   - 2 + 2 +2
   - 5 + 1
4. Islands of compute cycles
   - Good for isolation
   - Bad for hot spots
   - Ugly if hot spots move around

   So “Flow” workload across the cluster
   - Need to flow to new machines
5. You already have this technology…

Automatic Cluster Load Balancing



Live Partition Mobility since 2007

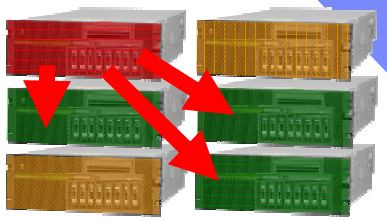
- Field tested and approved
- Dumb to manually monitor & move
  = man-power intensive

Systems Director + VMControl Systems Pools*

- System Pool fancy name for cluster!
- Monitor across the pool
- Automated LPM to level workload
- Option ”Recommend & ask permission mode”

* Combined package best purchased as Systems Director Enterprise Edition

Automatic Cluster Load Balancing



Summary

For:

- Think Clusters
  - For Cost effective HW purchase
  - We have the proven tools to load balance
- Maximising investment & flexibility from day 1

Against

- LPM pre-reqs
- SW costs but are these relatively low costs

Machine Evacuation for Maintenance

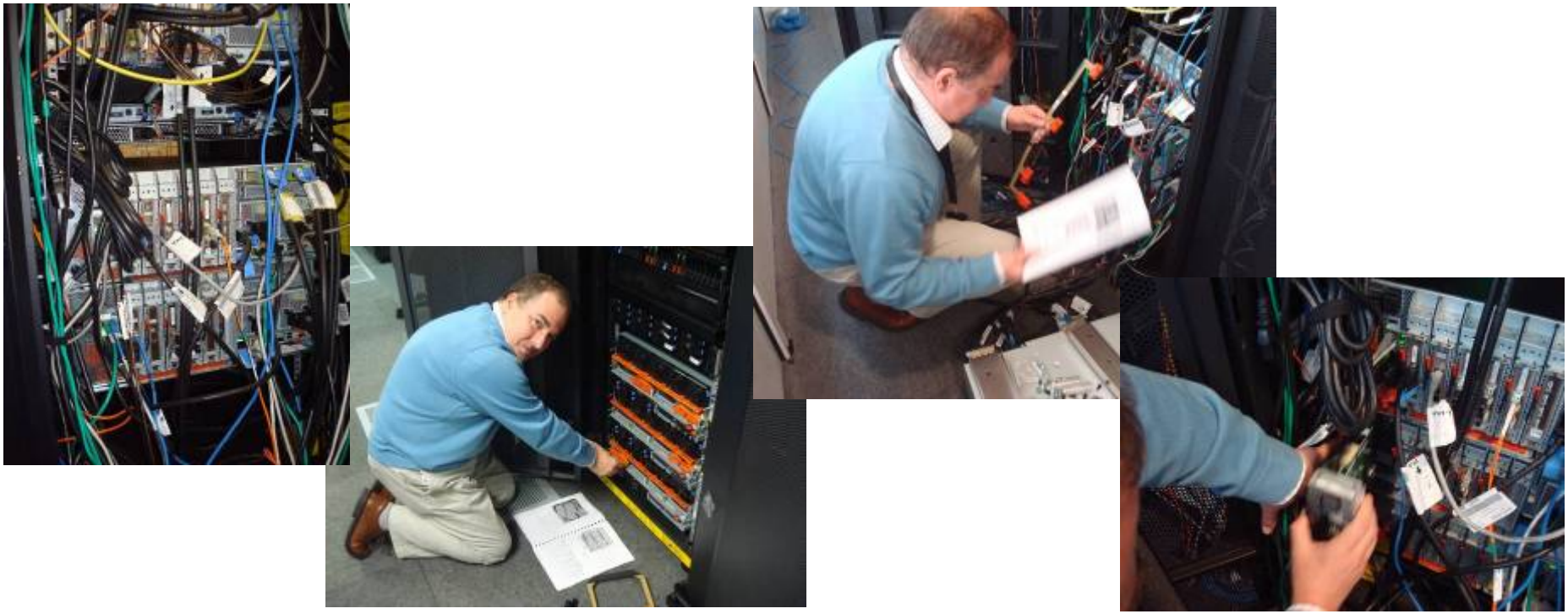
Maintenance like:

- Adding CPUs or memory
- Replacing parts
- Updating system firmware

Concurrent Maintenance
now called CHARM

Concurrent Maintenance

I have done “hot node add” to my Power 770 with near zero training & a PE watching



FSP Replace

- It works fine but makes assumption …

Concurrent Maintenance

But ….

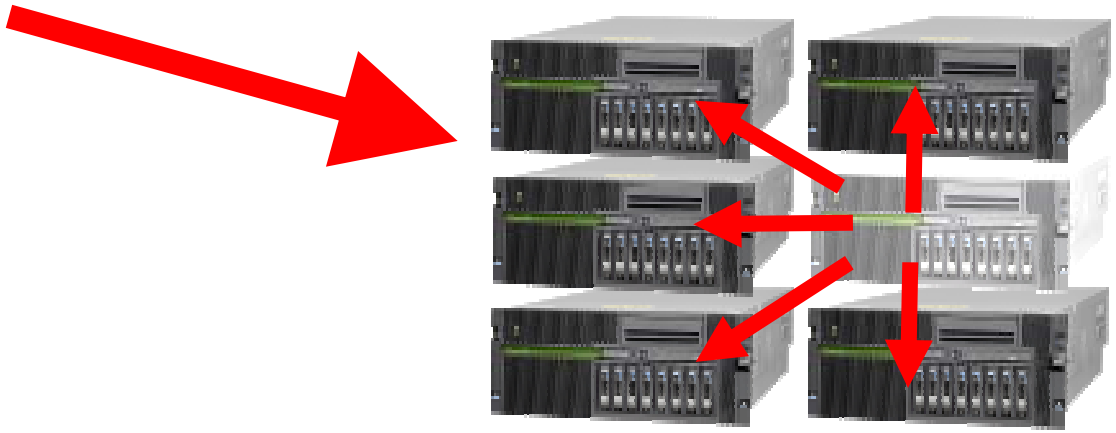
I see flash backs as a child to a TV program showing changing a car tyre - while driving along a motorway!

http://www.mothershiprv.com/2010/04/san-diego-automotive-museum.html

Machine Evacuation for Maintenance

Is there a better way to do CM/CHARM?
→ No risks of hardware accident
→No risks of config stopping the power off
→ No need for disruptive Firmware upgrade before the non-disruptive upgrade.

Doh!



Go Five-Pack

If you remove one node from a six pack, you still have 83% performance

1. Live Partition Mobility work to other nodes
2. Remove complete node

Go Five-Pack

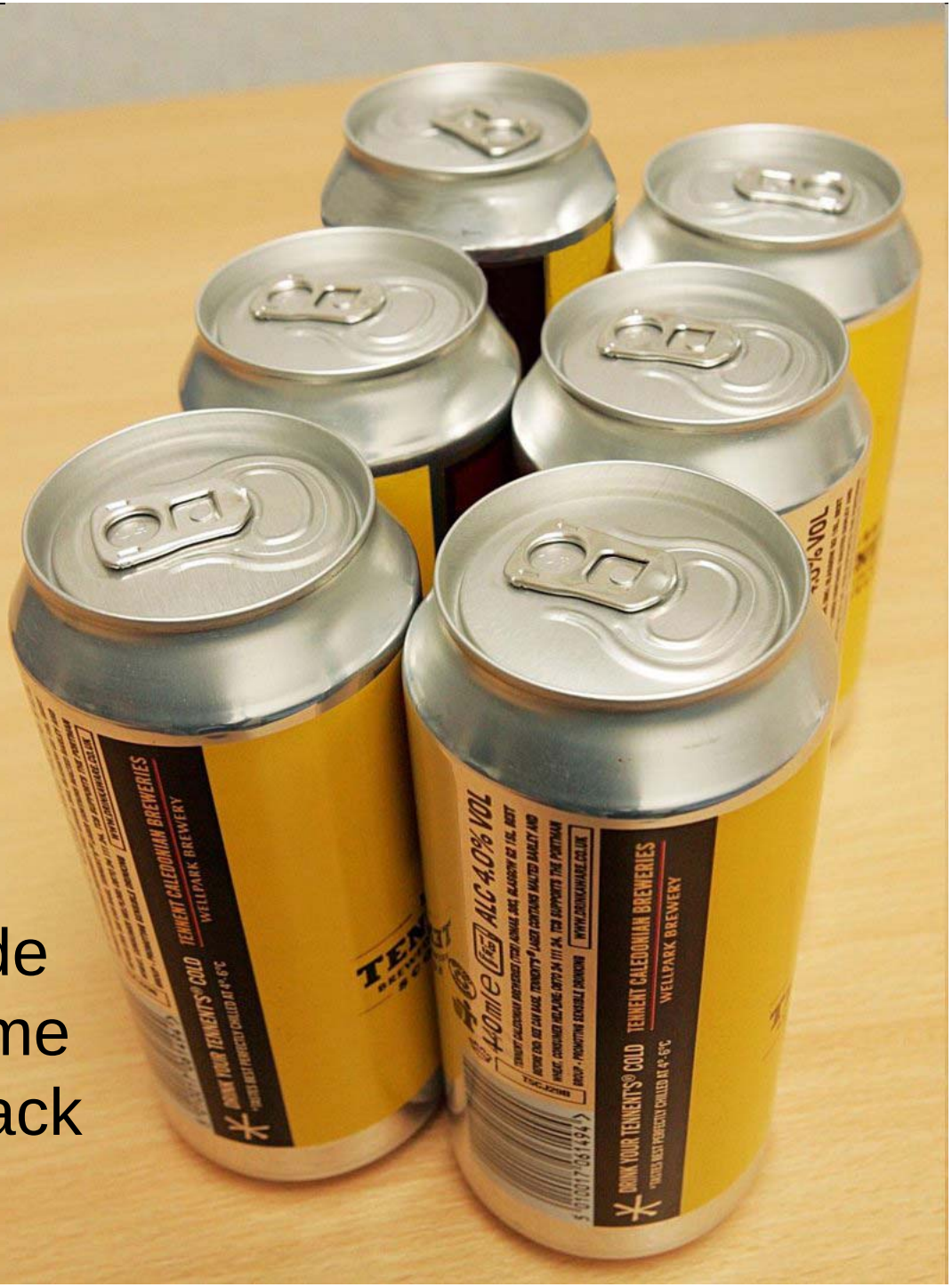
If you remove one node from a six pack, you still have 83% performance

1. Live Partition Mobility work to other nodes
2. Remove complete node
3. CE can take his/her time

Go Five-Pack

If you remove one node from a six pack, you still have 83% performance

1. Live Partition Mobility work to other nodes
2. Remove complete node
3. CE can take his/her time
4. Add improved node back



Machine Evacuation for Maintenance

Live Partition Mobility since 2005

- Field tested and approved

Systems Director + VMControl Systems Pools*

- Define “empty node X” plan
- Execute the plan – it automatically decides where
- →Do hassle free maintenance here
- Finally VMControl decides load balancing reuse

- Also can be use for weekend power offs

* Combined package best purchased as Systems Director Enterprise Edition

Machine Evacuation for Maintenance

Summary

For:

- Simple and Zero risk maintenance

Against

- LPM pre-reqs
- SW costs but these are relatively low costs

Active Energy Manager (AEM)

A Systems Director Plug-in
The oldest and cheapest!

Save Electricity

POWER7 Over-clocking

Active Energy Manager (AEM)

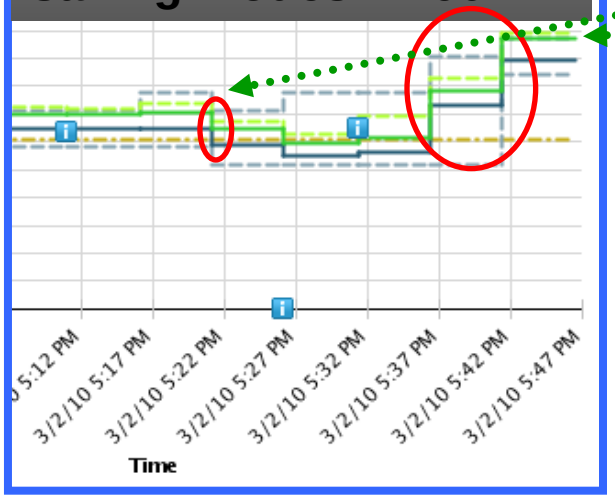
Four reasons

1. Less electricity - Near your building maximum?
2. Reduced costs - €$£
3. Green credentials - Save the planet + Annual Report
4. POWER7 Over-clocking
   - Serious “Street credibility”
   - Very cheap extra CPU cycles
   - Payback in 6 months (depends on the Power model)

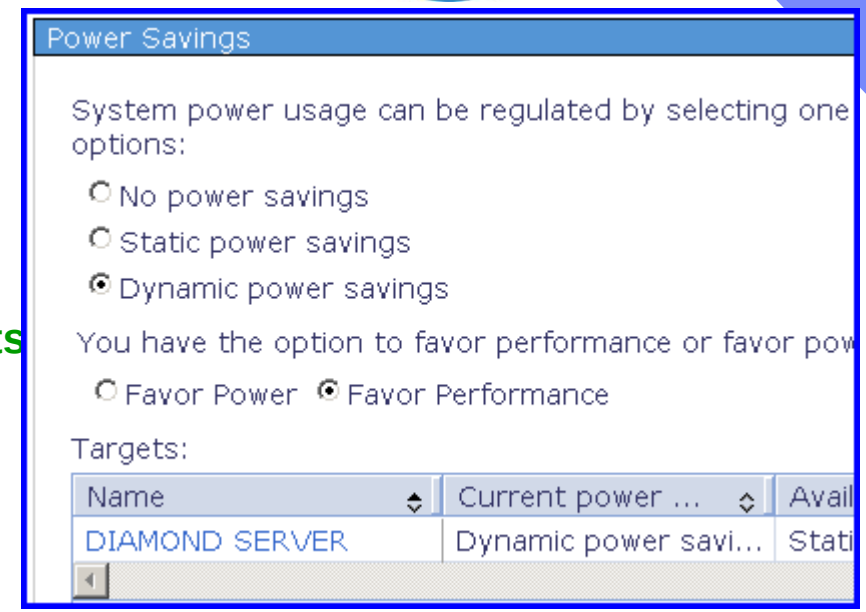
Active Energy Manager

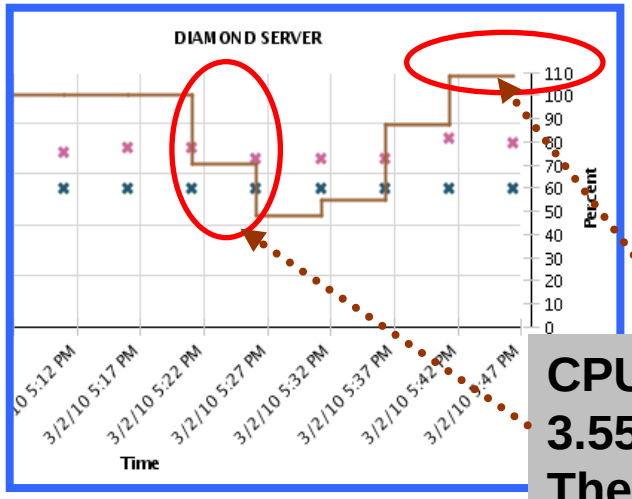
- Power 750→£22+£89/core or Director Std Edition

50 Watts saved.
Idle CPU already using intelligent power saving modes = 190 W



Power Input Watts
350 to 300 to 490





190W for 8 core
~800W for 32 core

CPUs drop buy 50% GHz
3.55 GHz → 1.65 GHz
Then up to 3.86 GHz
= 9% over clocking

Active Energy Manager (AEM)

Summary

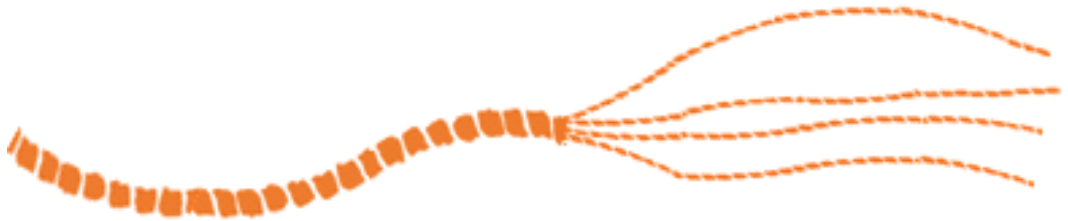
For:

- No brown out, €$£, Green, Over-clock

Against:

- Small initial costs

5 GHz barrier



5 GHz Barrier

Different chips have hit different Barriers

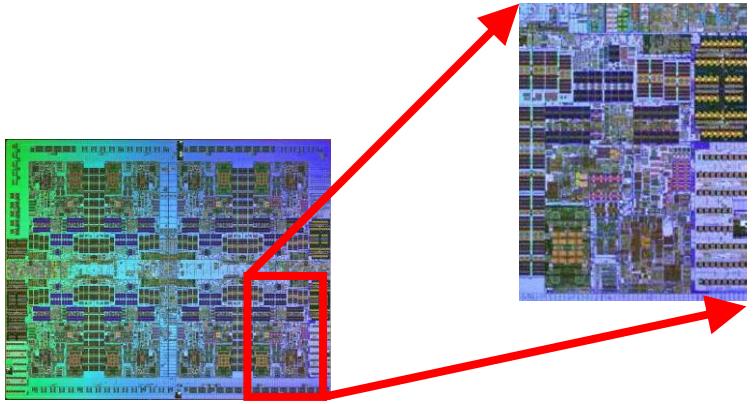
IBM	**POWER**	5 GHz
Intel	x86_64	4 GHz
Sun/Oracle	Sparc	3 GHz
HP	Itanium	2 GHz

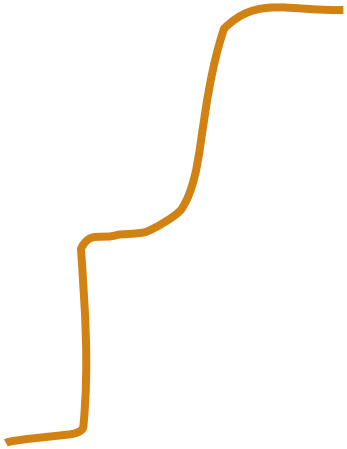
Bottom line is performance gains via

- Having lots of cores & threads
- Not more GHz

Faster: yes, yes, yes-ish, maybe!

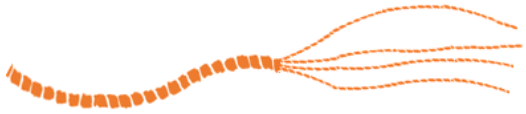
Faster box=YES, chip=YES, core=yes thread=?





Core speeds up only a little (even with slower GHz)
Thread speed for single threaded application

- High POWER6 GHz to low POWER7 GHz = borderline
- Intelligent threads mean better single change to single thread
- Get the application vendor fixed or changed



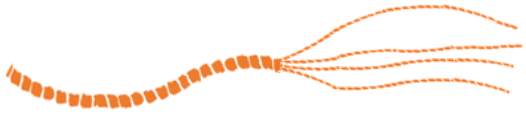
5 GHz Barrier

Worst case is the “single thread application”

If you have one …
start planning a change of vendor
if they haven’t fixed this by now
… they are going bust!

And you can quote me on that.

Symmetric multiprocessing arrived in 1994



5 GHz Barrier Summary

For:

- It has already happened
- Its not optional

Against:

- Stuck with old 1 thread apps
- No more single threaded application performance fixed via GHz

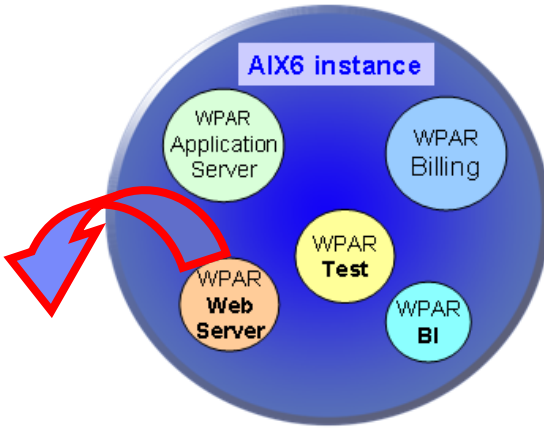
Versioned WPARs for AIX 5.2

We are all experts in WPARs already … Right!

Micro AIX running inside AIX – since 2008

Standard WPAR Reminder:

- Application encapsulated in each WPAR
- Share AIX6+AIX7 kernel – default memory is just 60 MB
- Private file systems /, /tmp, /home & /var plus read-only/private /usr & /opt
- Network alias for each WPAR
- Resource constraints each WPAR
- Live Application Mobility jumps a WPAR to other global AIX



Versioned WPARs for AIX 5.2

- AIX 5.2
  - Came out in 2002
  - Functionally stabilised 2007
  - Last fix in 2008
- Runs on AIX 7 with POWER7
- Separate LPP = £$€

Versioned WPARs for AIX 5.2

Power3/RS64/POWER4

- 24 CPU, 128GB RAM,
- 8 to 18 GB SCSI disk,
- 10 to 100 Mb network
- CD [S80=~6 rPerf]

Area	Support/Maintenance
Old application	maybe
AIX 5.2	No
Old firmware	No
Old hardware	Costly
Virtual	None
Performance	Slow
Electricity	High
Footprint	Large

VWPAR + POWER7 + SMT4

- 256 CPU, 1 TB RAM,
- 400 GB SAS/FC disks+SSD
- 10 Gbit Net
- DVD [710=45 rPerf]

Area	Support/Maintenance
Old application	maybe
AIX 5.2/7.1	YES
Firmware	yes
Hardware	cheap
Fully virtual	completely
Performance	fast
Electricity	low
Footprint	tiny

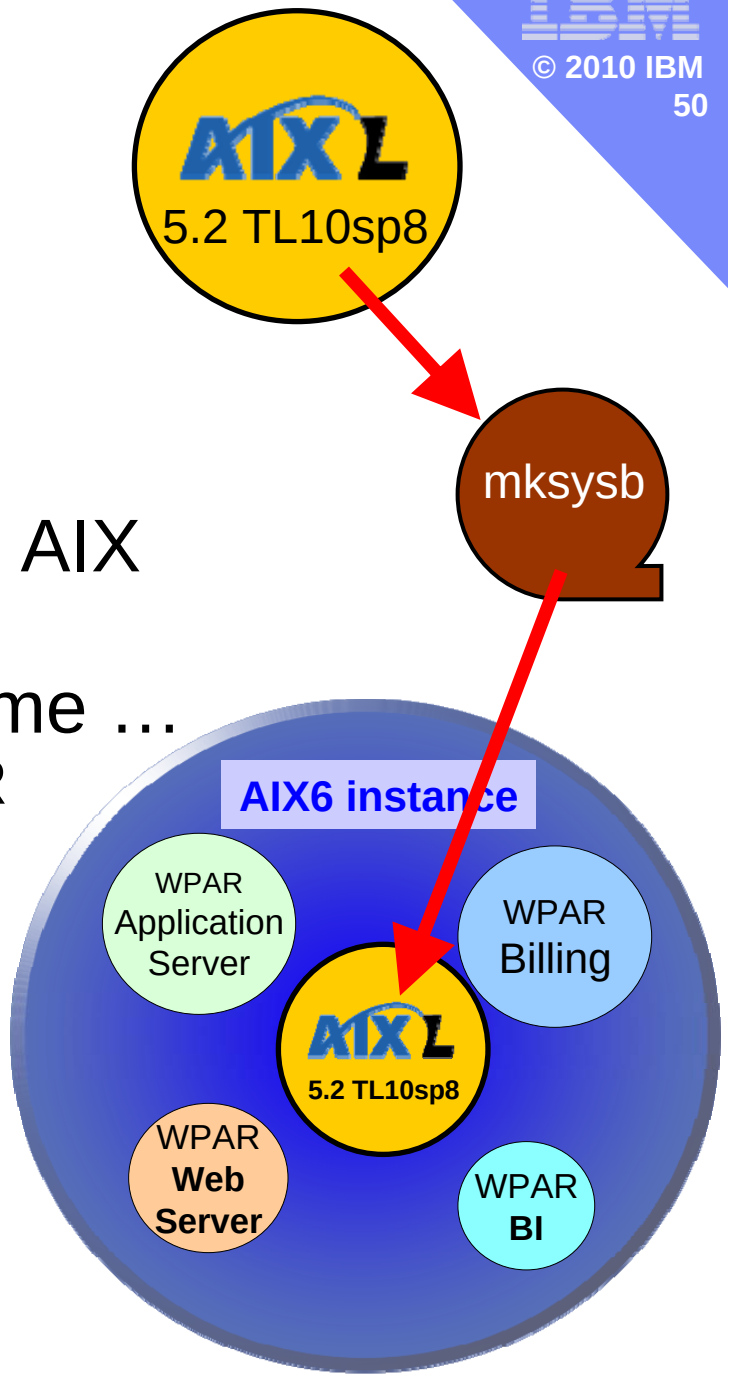
Creating one is easy

Prepare

- AIX 5.2 TL10 service pack 8
- mksysb of AIX 5.2
- Add vWPAR packages to global AIX

Create

- mkwpar -C -B mksysb-file -n name …
  -C pulls in extra packages into WPAR
  -B = AIX 5.2 mksysb file
- Then runs as any other WPAR



Versioned WPARs for AIX 5.2

Summary

For:

- Get AIX 5.2 supported
- Ditch ancient hardware, free up space
- Lower HW maintenance
- Move up to the virtual shared everything world
- High performance

Against:

- Small Cost
- Application support – if vendor supporting at all !

Prebuilt Solutions

Reason:

- IBM does the integration & not 1000's of customers
- Sold as whole package – not parts
- Services as a whole – part parts
- Upgradable as workloads grow

Prebuilt Solutions

Examples:

Cloud on Power

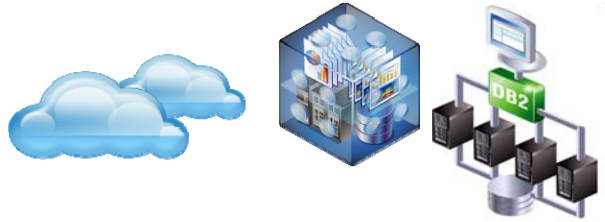
Web orderable automated solution

Analytics

Scalable Data Warehouse

DB2 PureScale

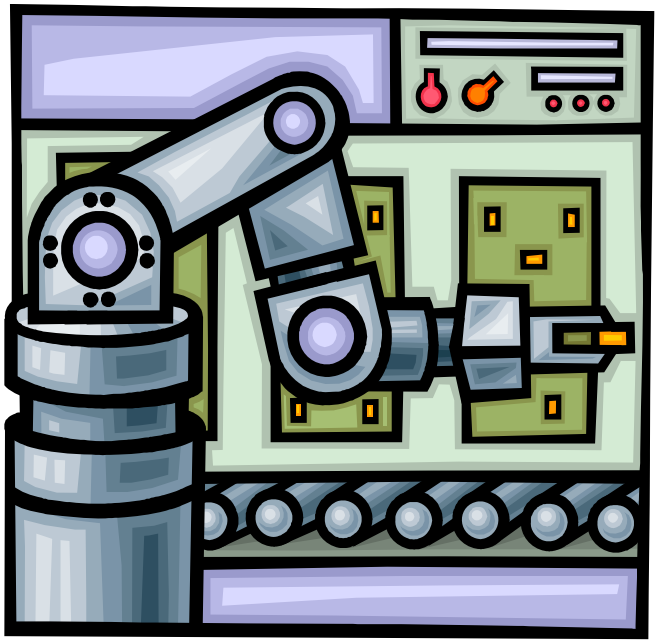
Scalable OLTP

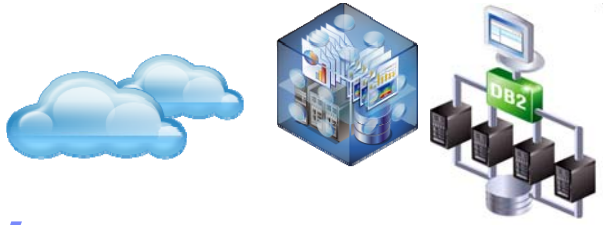


Prebuilt Solutions

Why?
To stop “cottage industry” of unique hand-made LPARs

“Drive down costs by making it simpler, use standard parts & fully automated.”





Prebuilt Solutions Summary

For:

- **Cost by reducing man power**
- Speed of implementing solution to days
- Higher utilisation of resources
- Managed Solution (not sum of the parts)

Against:

- Politics

Systems Director

“One tool to rule them all”

Systems Director

Higher function Plug-ins covered in other items

- Active Energy Manager (AEM)
- VMControl – Cluster load Balancing & Evacuation
- Workload partition Manager

Systems Director Base

- Inventory & Topology views for relationships
- Automation Plans for self healing
- Automated Updates for HMC, Firmware, AIX, VIOS
- Future HMC !

AIX is supplied as .ova [VMControl Image Mgr file]

Now its your turn …

Vote for 4 technologies

“mostly likely to be normal in 5 years time”

Then I show you mine

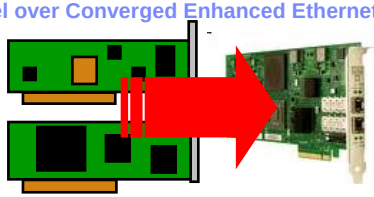
Ten Smarter Technologies coming to your … computer room & career

Systems Director

- Single Screen
- Topology Views
- Update Mgr
- Automation Mgr

FCoCEE

Fibre Channel over Converged Enhanced Ethernet

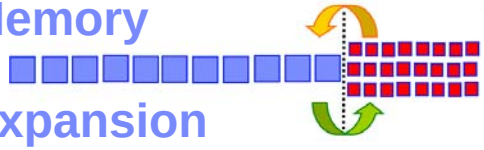


Halve the number of adapters & PCIe slots

Solid State Drive

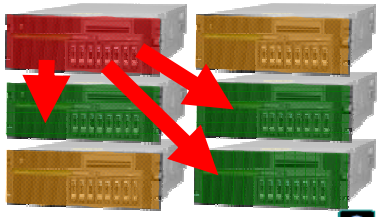
Already 150 times faster and already seeing higher densities/low prices

Active Memory Expansion



Compressed RAM – reduced apps bloat

VMControl System Pool Cluster Automatic Balancing



VMControl System Pool replacement for CM Cluster Evacuation & Fix

Active Energy Manager

Reduce electricity
+ POWER7 Over-clocking

CPU GHz barrier means multi-thread or die

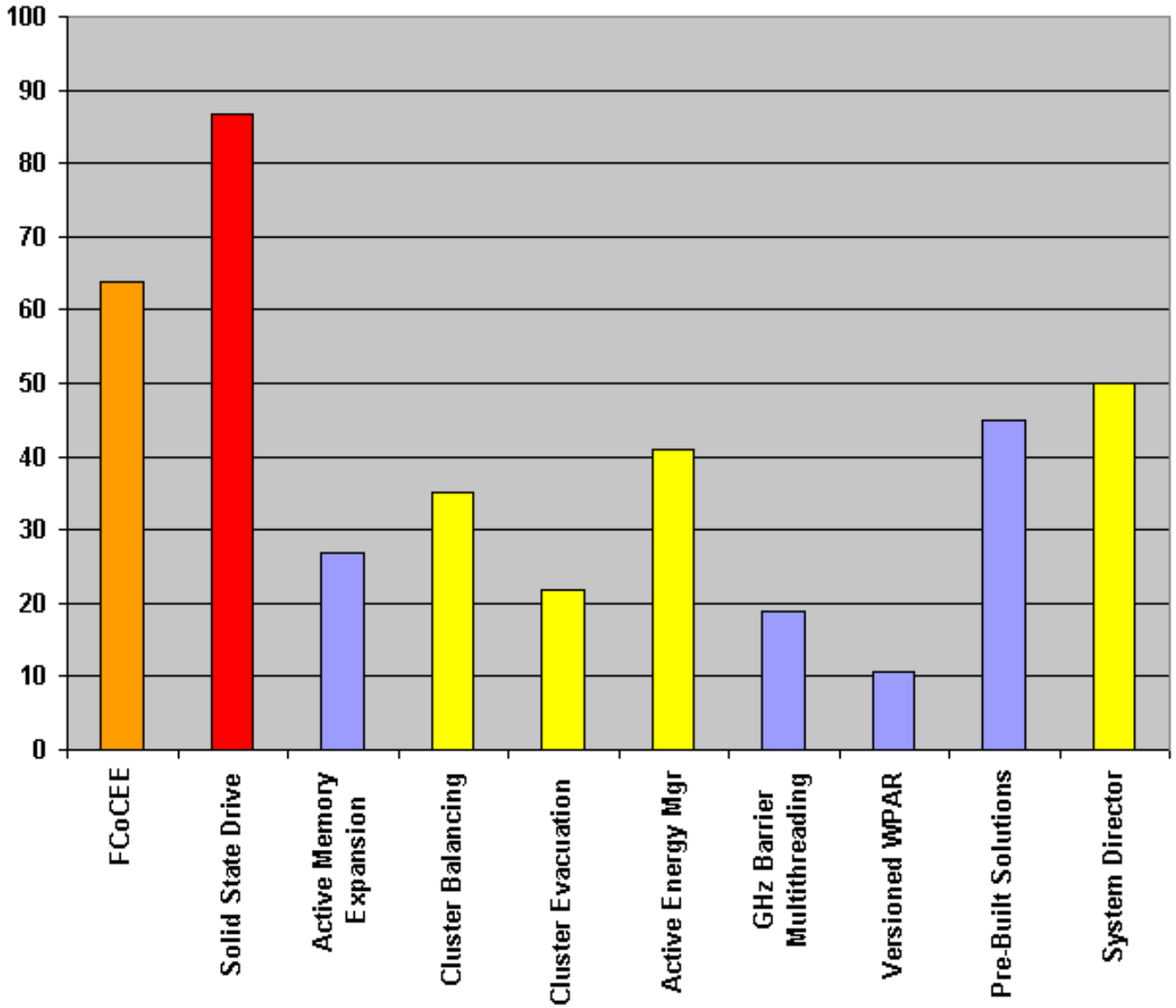
ISV offering single threaded apps are dead
Size apps in terms of dozens of cores

Versioned WPAR

Running AIX 5.2 in a WPAR

Pre-Built Solutions

Cloud Analytics pureScale
No more hand-made software stacks



Techie IBMers
- 40

Customer Technical Universities
- 300 USA
- 90 France

They all got votes = they are all winners
SSD, FCoCEE, Systems Director (+ plug-ins=yellow), Pre-built Solutions
Ignore GHz barrier – not optional + already happened
See AIXpert blog for other comments

Got an opinion? – find me at …

AIXpert blog
- https://www.ibm.com/developerworks/mydeveloperworks/blogs/aixpert

AIX & POWER Movies (now 84 of them!)
- http://www.ibm.com/developerworks/wikis/display/WikiPtype/Movies

Performance Tools Forum
- http://www.ibm.com/developerworks/forums/forum.jspa?forumID=749

AIX Wiki
- https://www.ibm.com/developerworks/wikis/display/WikiPtype/AIX+Wiki

Email nag@uk.ibm.com

Twitter mr_nmon