2012
IBM Power Systems Technical University
featuring IBM PureSystems

Enabling the infrastructure for smarter computing

Return to Basics I :

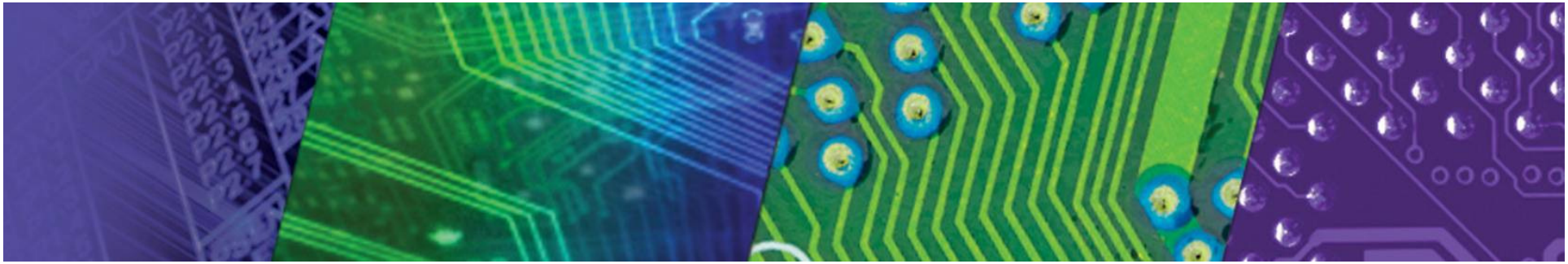
Understanding POWER7
Capacity Entitlement and Virtual Processors

VN211

Rosa Davidson

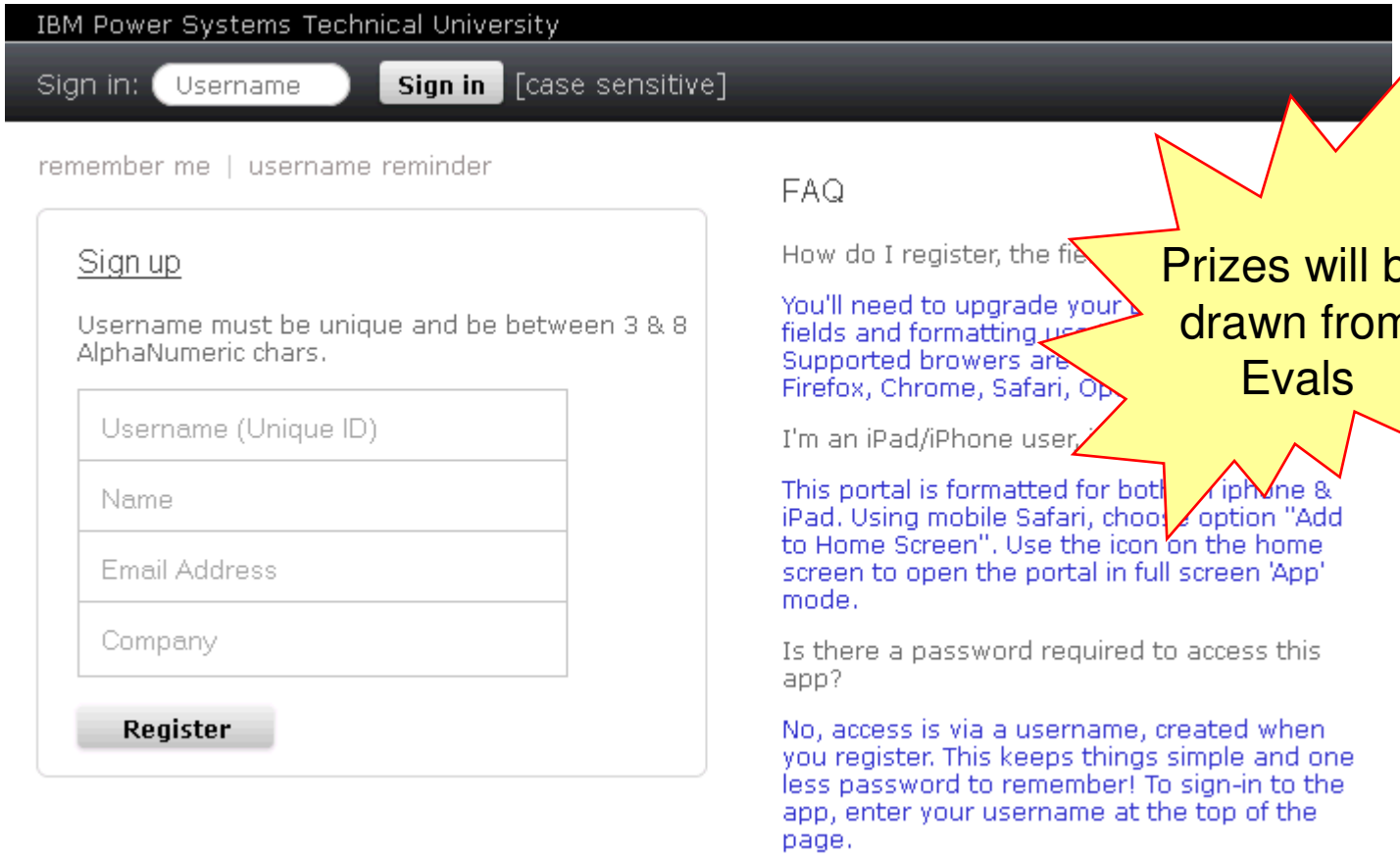
Advanced Technical Skills - Security and Performance - IBM

Advanced Technical Skills



Session Evaluations

- ibmtechu.com/vp

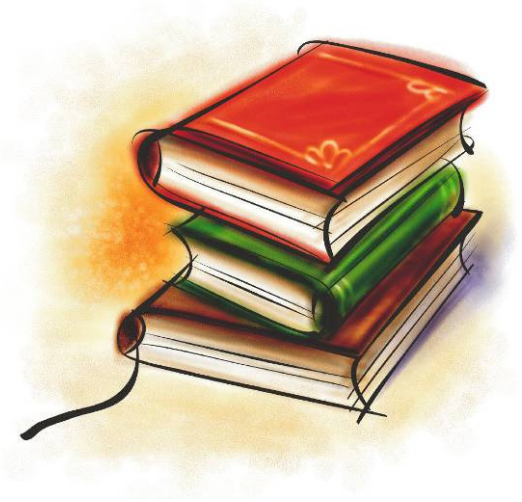


Outline – Part 1

- PowerVM Concepts: (20 slides)
  - Terminology : CPU, CPU within AIX.
  - Two Usages View : PowerVM vs OS
  - Types of Logical PARtitions (LPARs).
  - Virtual Processor Definition (VP).
  - Capacity Entitlement Definition (CE).
  - Dedicated LPAR Definition.
  - Dedicated Donating Definition.
  - VPs vs AIX SMT scheduling – VP Folding
  - Shared LPAR Definition.
    - Maximum VPs for each LPAR.
    - Shared: "*The cake & the Invitees*" *story*
  - Capacity Entitlement: Dispatch wheel.
    - "Limit the VPs, Core Access you receive"
    - The Uncapped: "*The False Hope*" story

Advanced Technical Skills

Bibliography - References

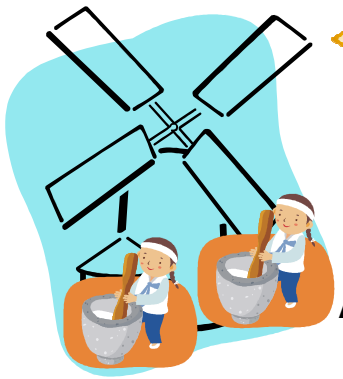


- **Beyond this presentation, read the White Paper from Mala Anand :**
  "***POWER7 Virtualization - Best Practice Guide***"
  http://www.ibm.com/developerworks/wikis/display/WikiPtype/Performance+Monitoring+Documentation
- ***Server virtualization with IBM PowerVM***
  http://www-03.ibm.com/systems/power/software/virtualization/resources.html
- ***IBM Systems Workload Estimator*** or http://www-912.ibm.com/estimator
- IBM System p Advanced POWER Virtualization Best Practices Redbook:
  http://www.redbooks.ibm.com/redpapers/pdfs/redp4194.pdf
- Virtualization Best Practice:
  http://www.ibm.com/developerworks/wikis/display/virtualization/Virtualization+Best+Practice
- Configuring Processor Resources for System p5 Shared-Processor Pool Micro-Partitions:
  http://www.ibmsystemsmag.com/aix/administrator/systemsmanagement/Configuring-Processor-Resources-for-System-p5-Shar/
- An LPAR Review:
  http://www.ibmsystemsmag.com/aix/administrator/lpar/An-LPAR-Review/
- Virtualization Tricks:
  http://www.ibmsystemsmag.com/aix/trends/whatsnew/Virtualization-Tricks/
- A Comparison of PowerVM and x86-Based Virtualization Performance:
  http://www-03.ibm.com/systems/power/software/virtualization/whitepapers/powervm_x86.html
- IBM Integrated Virtualization Manager:
  http://www-03.ibm.com/systems/power/hardware/whitepapers/ivm.html
- Achieving Technical and Business Benefits through Processor Virtualization:
  http://www.ibm.com/common/ssi/fcgi-bin/ssialias?infotype=SA&subtype=WH&appname=STGE_PO_PO_USEN&htmlfid=POL03027USEN&attachment=POL03027USEN.PDF
- Java Performance Advisor is available
  https://www.ibm.com/developerworks/wikis/display/WikiPtype/Java+Performance+Advisor
  ftp://ftp.software.ibm.com/aix/tools/perftools/JPA/AIX61/
- VIOS Performance Advisor is available
  http://www.ibm.com/developerworks/wikis/display/WikiPtype/VIOS+Advisor
- Virtualization Performance Advisor is in develeopment (expected Q1/2012)
  http://aixptools.austin.ibm.com/virt/virt_advisor/

Terminology : CPU

- The **Threads** of the core are the: **Logical Processor (LP) or Logical CPU (lcpu).**

- The **Core** becomes a "**Virtual Processor**" or "**Physical Processor**" or "**Phvsical CPU**".



100

THE WINNERS ARE:

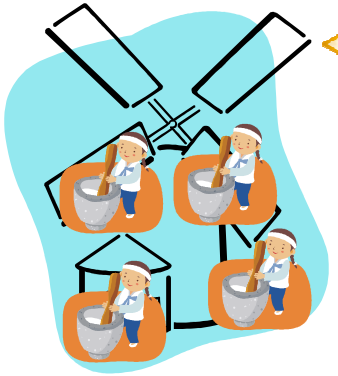
POWER5

POWER6

POWER7 ***Compatibility mode P6***

POWER7 SMT2

A core with 2 LOGICAL CPUS



150

THE WINNERS ARE:

POWER7 SMT4

A core with 4 LOGICAL CPUS

"Number ONE !" EFFECT

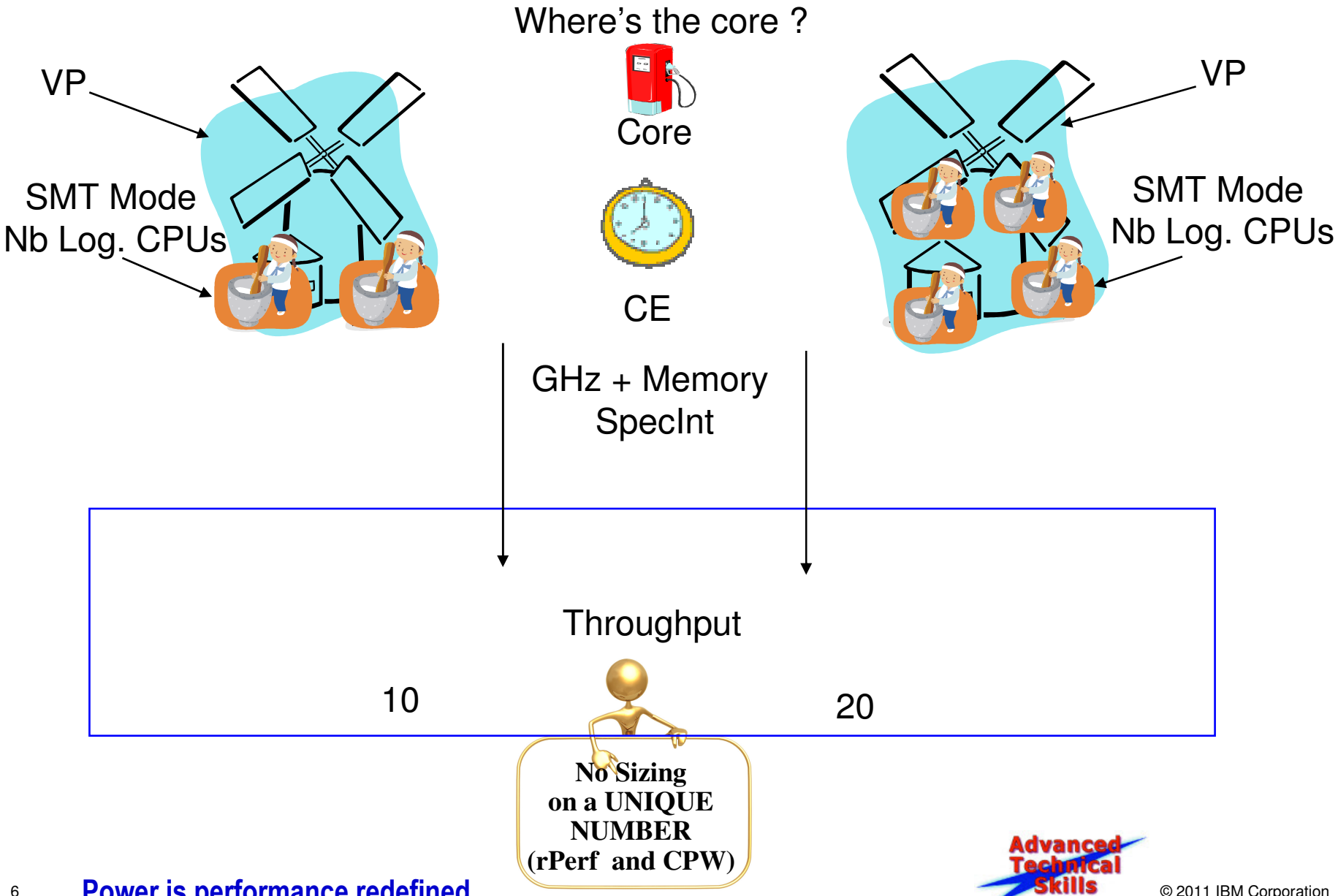
No Sizing on a UNIQUE NUMBER
(rPerf and CPW – ONE Number)

NO BOX SIZING
(VIO ? Which VIO?)
BOX = ONE LPAR

Power is performance redefined

Advanced Technical Skills

Mills crunching data or My Performance Job

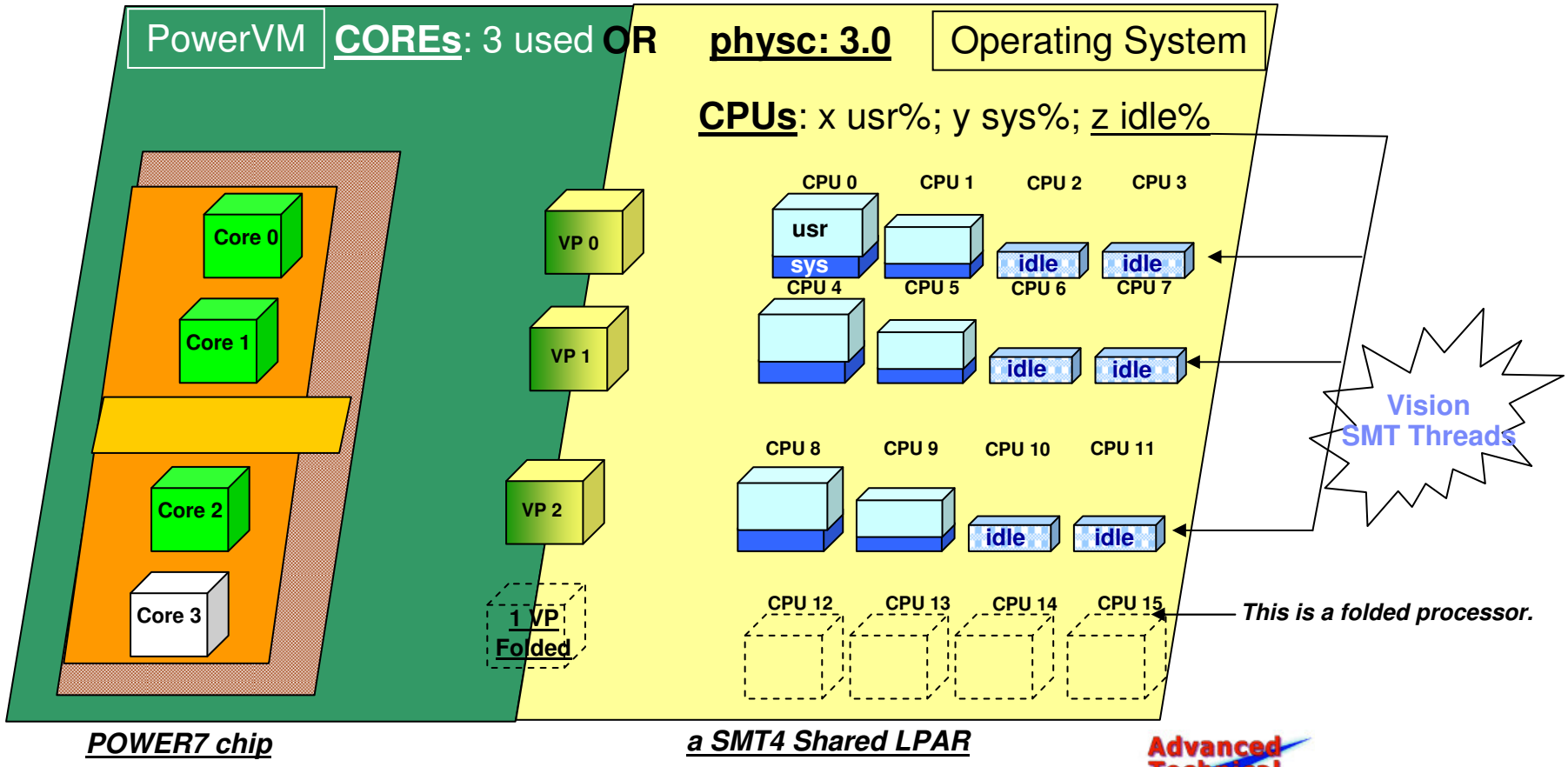


Power is performance redefined

Advanced Technical Skills

Two Usage views: PowerVM and OS

- **IDLE CAPACITY:**
  - PowerVM does report **cores' usage (idle core).**
  - Operating Systems (**OS**) do report the **idle capacity (%idle) of Logical CPUs (SMT threads).**
    - To report how many cores are used, OS gives the **physical consumed**: **physc**



Power is performance redefined

Advanced Technical Skills

Terminology : CPU within AIX

- To see the **Logical CPUs - Physical CPU association**: AIX command **'smtctl'**
- AIX V3 is born 1990 where 1 CPU = 1 Processor.
  - Some historic Unix commands are naming logical cpus as processor.
  - New AIX Commands are showing logical cpus with the right denomination.

On a POWER5 SERVER – With AIX

PHYSICAL CPUs or VIRTUAL PROCESSORS "PROC"

```
/=>lsdev -C | grep proc
proc0      Available 00-00          Processor (VP 1)
proc2      Available 00-02          Processor (VP 2)
proc4      Available 00-04          Processor (VP 3)
proc6      Available 00-06          Processor (VP 4)
```

LOGICAL CPUs "processor"

```
/=>bindprocessor -q
The available processors are:  0 1 2 3 4 5 6 7
/=>lparstat
…type=Shared mode=Uncapped smt=On lcpu=8 mem=2048MB…
```

ASSOCIATION LOGICAL CPUs–PHYSICAL CPU

```
/=>smtctl
This system supports up to 2 SMT threads per processor.
SMT threads are bound to the same virtual processor.
proc0 has 2 SMT threads.
Bind processor 0 (lcpu0) is bound with proc0 (VP 1)
Bind processor 1 (lcpu1) is bound with proc0 (VP 1)
```

On a POWER7 SERVER – With AIX

PHYSICAL CPUs or VIRTUAL PROCESSORS "PROC"

```
# lsdev -C | grep proc
proc0      Available 00-00          Processor (VP 1)
proc4      Available 00-04          Processor (VP 2)
```

LOGICAL CPUs "processor"

```
# bindprocessor -q
The available processors are:  0 1 2 3 4 5 6 7
# lparstat
…type=Shared mode=Uncapped smt=4 lcpu=8 mem=2048MB …
```

ASSOCIATION LOGICAL CPUs–PHYSICAL CPU

```
# smtctl
This system supports up to 4 SMT threads per processor.
SMT threads are bound to the same virtual processor.
proc0 has 4 SMT threads.
Bind processor 0 (lcpu0) is bound with proc0 (VP 1)
Bind processor 1 (lcpu1) is bound with proc0 (VP 1)
Bind processor 2 (lcpu2) is bound with proc0 (VP 1)
Bind processor 3 (lcpu3) is bound with proc0 (VP 1)
```


Advanced Technical Skills

Questions / Answers

```
root@davidson /=>lsdev -Cc processor
proc0  Available 00-00 Processor
proc2  Available 00-02 Processor
proc4  Available 00-04 Processor
proc6  Available 00-06 Processor
proc8  Defined   00-08 Processor
proc10 Defined   00-10 Processor
proc12 Defined   00-12 Processor
proc14 Defined   00-14 Processor
proc16 Defined   00-16 Processor
proc18 Defined   00-18 Processor
proc20 Defined   00-20 Processor
proc22 Defined   00-22 Processor
proc24 Defined   00-24 Processor
proc26 Defined   00-26 Processor
proc28 Defined   00-28 Processor
proc30 Defined   00-30 Processor
proc32 Defined   00-32 Processor
proc34 Defined   00-34 Processor
proc36 Defined   00-36 Processor
proc38 Defined   00-38 Processor
```

4

16

Virtual Processors or Logical CPUs ?

How many XX do we have ?

Why 4 available / 16 defined ?

ANSWER

```
root@davidson /=>lparstat -i
Node Name                  : davidson
Partition Name             : aix61
Partition Number           : 1
Type                       : Shared-SMT
Mode                       : Uncapped
Entitled Capacity          : 4.00
Partition Group-ID         : 32769
Shared Pool ID             : 0
Online Virtual CPUs*       : 4
Maximum Virtual CPUs*      : 10
Minimum Virtual CPUs       : 1
```

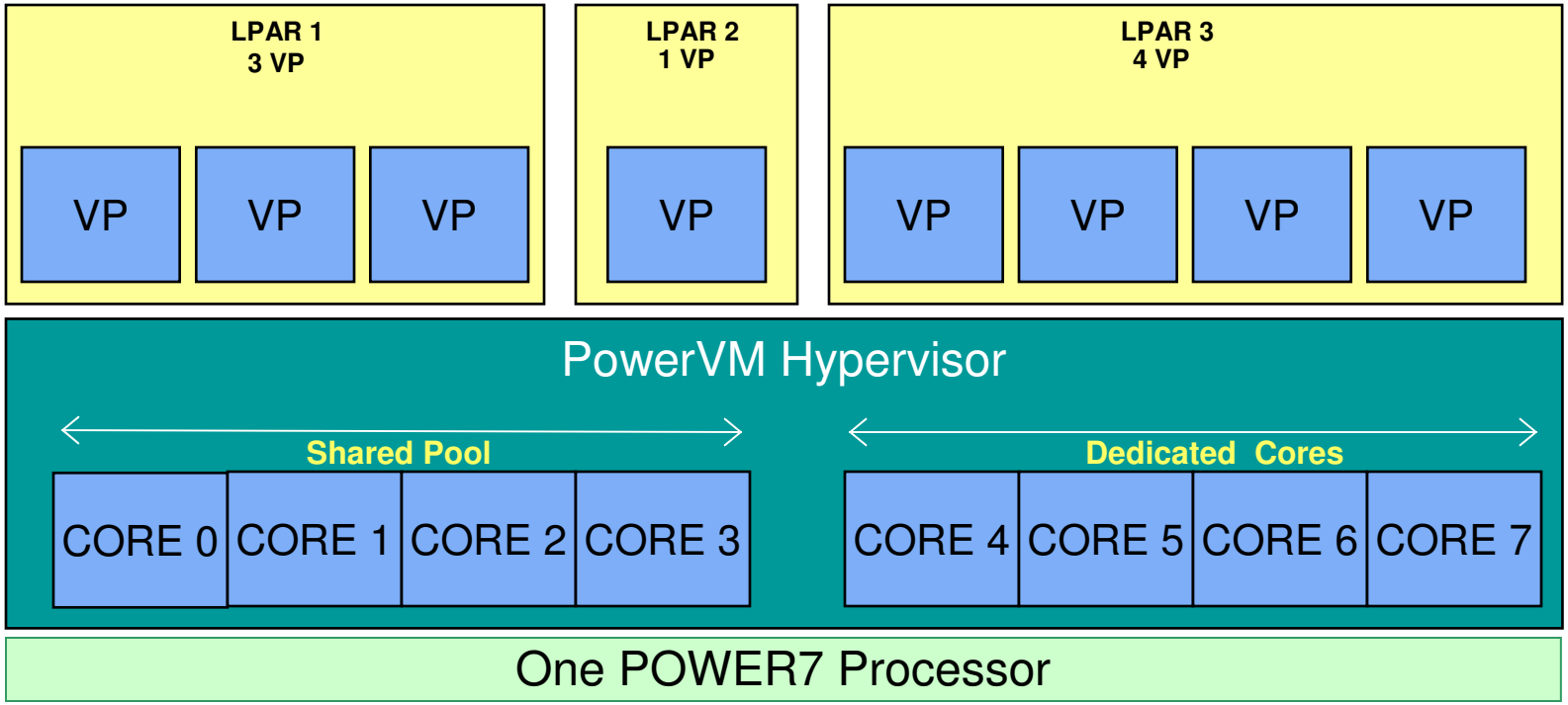
Why 16 defined and not 6 defined ?

*** "Virtual CPUs" means Virtual Processors**

Types of **L**ogical **PAR**tition (**LPAR**)

- **Dedicated Partition:**
  - One partition ***has a whole number of cores*** (from one to the maximum supported by the server).
  - Can use virtual adapters and virtual disks.
  - Can share its non-used cores (**Dedicated Donating**) or Do not share them (**Dedicated**).
  - Can not borrow any cores if needed.
- **Shared Partition in one Shared Pool:**
  - One partition ***has a fractional number of cores*** (from 0.1 to the maximum supported by the server).
  - Can use virtual adapters and virtual disks.
  - Share its non-used cores, belonging shared pool.
  - Can borrow a fractional number of any cores if needed.
    - Have ***a borrowing limit (Capped)*** or ***no borrowing limit (Uncapped).***
  - Is assigned to a pool of shared cores: **Processors Pool or Shared Pool**
    - Group the cores which are shared by Shared Partitions.
- **VIO Server:**
  - ***An Appliance Partition*** used to virtualize physical network adapters, physical storage adapters and CD devices.
  - A server can have one to many VIO servers.

PowerVM: Virtual Processor Definition

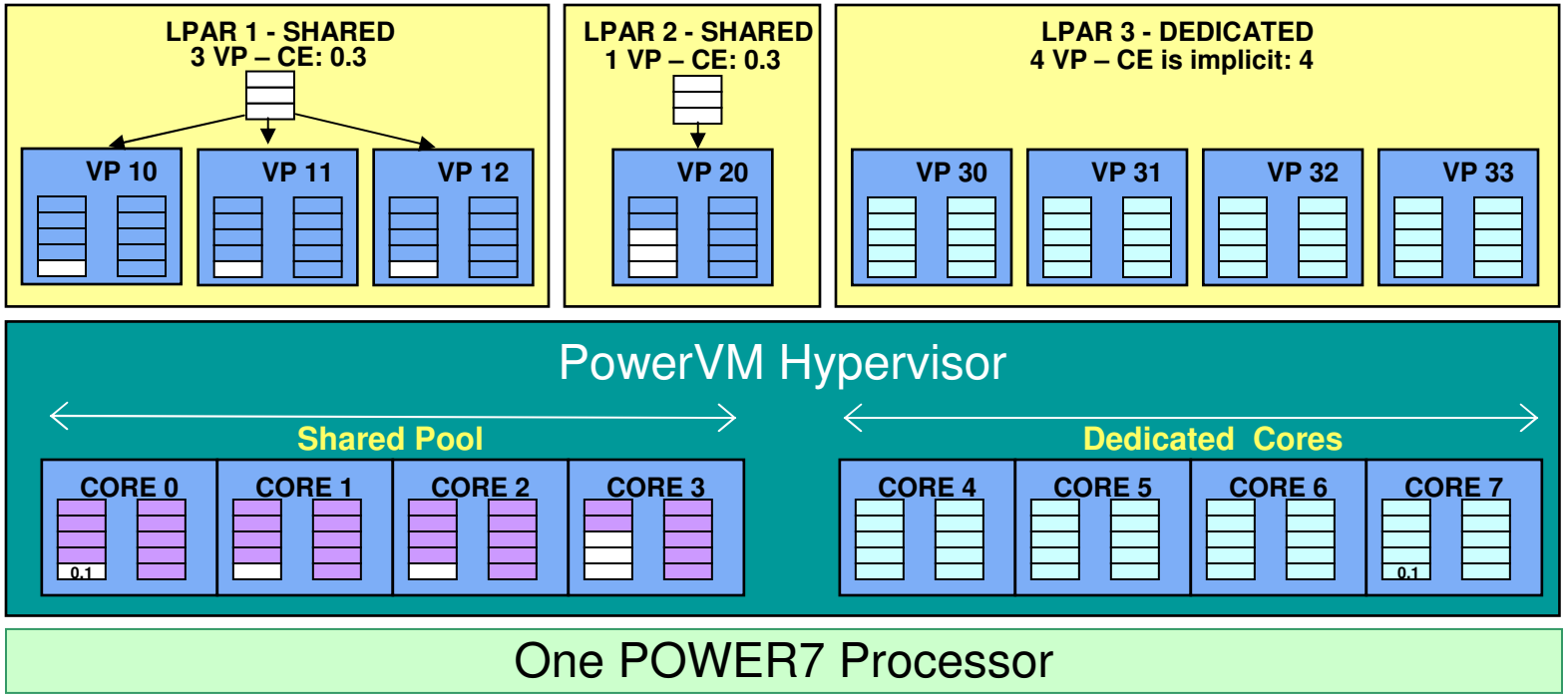


- Cores are presented to LPARS' Operating Systems as ***Virtual Processor (VP).***
- PowerVM virtualizes the physical cores, delivering a portion of a them to each LPAR.

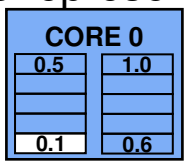
Power is performance redefined

Advanced Technical Skills

PowerVM: Capacity Entitlement Definition



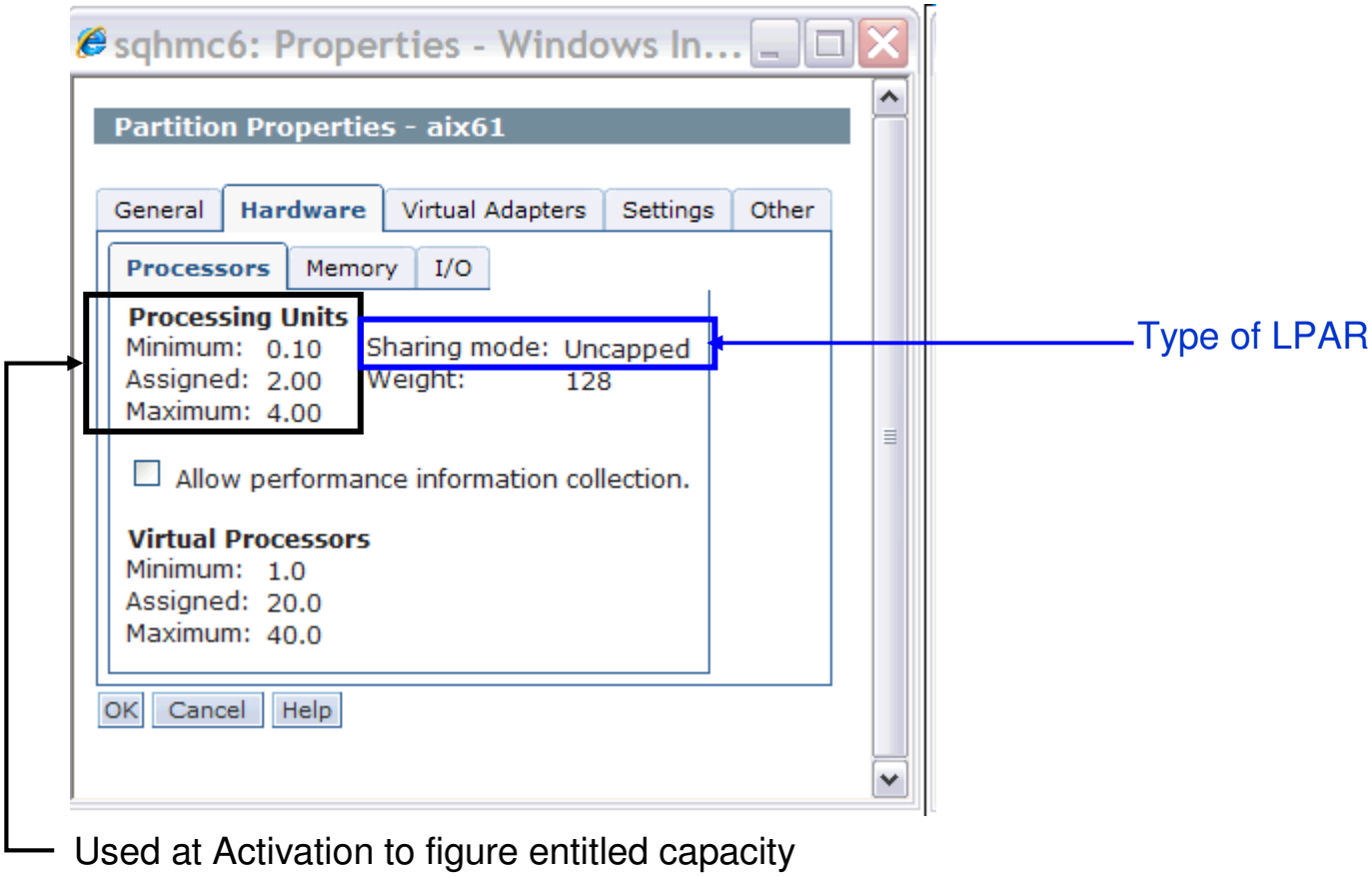
- The whole or fractional number of cores is represented by the : ***Capacity Entitlement (CE).***
  - The initial fraction is 1/10th of a core.



  - Additional fraction can be in order of 1/100th of a core
    - This presentation does represent only 1/10th.
- Within HMC, ***Capacity Entitlement (CE)*** is called ***Processing Unit.***

Advanced Technical Skills

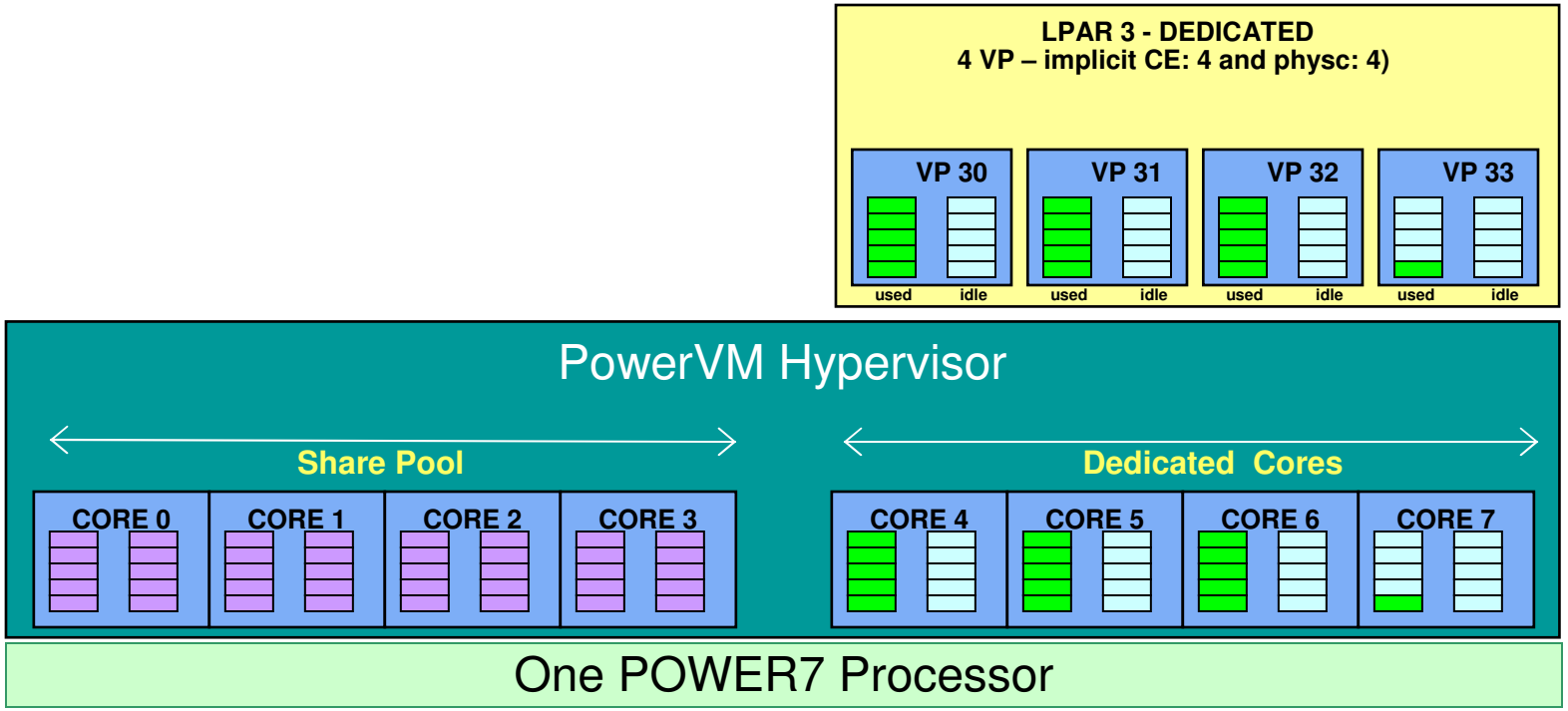
HMC: Processing Unit for Capacity Entitlement



Power is performance redefined

Advanced Technical Skills

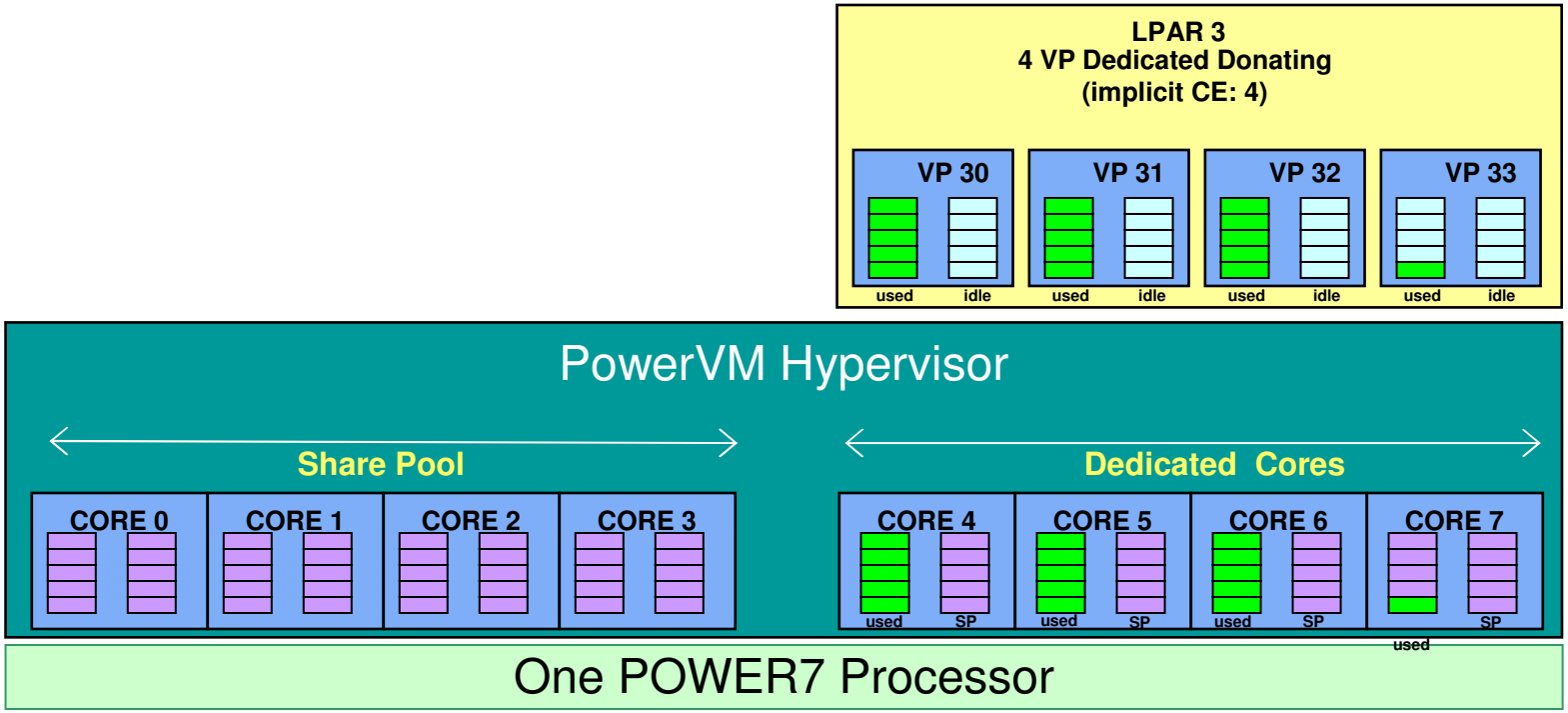
PowerVM: Dedicated LPAR Definition



- ***Dedicated / Dedicated Donating LPARs:*** *receive all fractions of cores (the whole core).*
- ***Dedicated LPAR:***
  - *Idle cycles « of the core » are not ceded: % idle is present.*
  - *You have dedicated cores : relationship 1:1*
  - *Your cores are there for you and you only: IMMEDIATELY available.*
  - *As there is no sharing, there is the best processor affinity : you're the only one to use these cores.*

Advanced Technical Skills

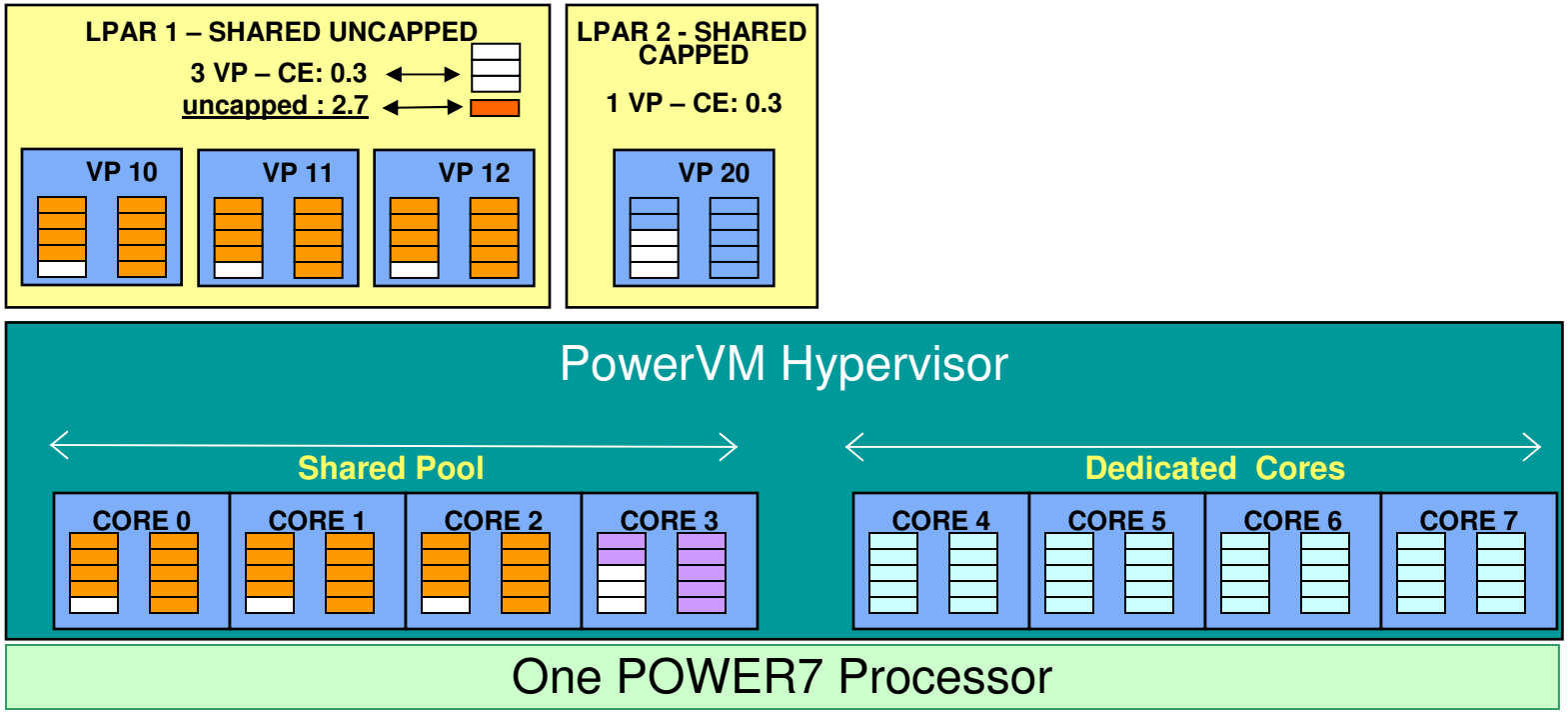
PowerVM: Dedicated Donating LPAR Definition



- ***Dedicated / Dedicated Donating LPARs:*** *receive all fractions of cores (the whole core).*
- ***Dedicated Donating LPAR:***
  - *Idle cycles « of the core » are ceded to the share pool : share pool is extended.*
    - *%idle can be present. It does represent the %idle 4 Logical CPUs of the used core (SMT 4).*
  - *Your cores are there for you and you only: IMMEDIATELY available.*
  - *The processor affinity is restored after shared LPARs have been invited to use your idle cycles.*

Advanced Technical Skills

PowerVM: Shared LPAR Definition



- ***Shared LPARs:*** *receive a fractional number of cores.*
  - ***Capped :*** *the fractional number of cores goes up to the value of this fractional number (**CE**).*
    - *LPAR 2 can run up to 0.3 only because CE is 0.3 AND LPAR 2 is capped.*
  - ***Uncapped:*** *the fractional number of cores goes beyond the CE up to the whole number of **VP** (cores).*
    - *LPAR 1 can run up to 3.0 because VP is 3 AND LPAR 1 is uncapped.*

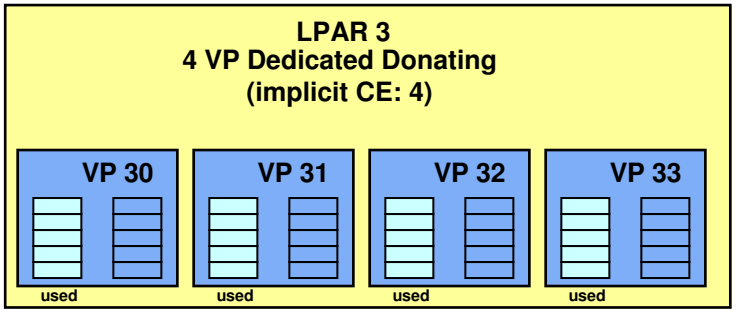
Advanced Technical Skills

The “Virtual Processors”

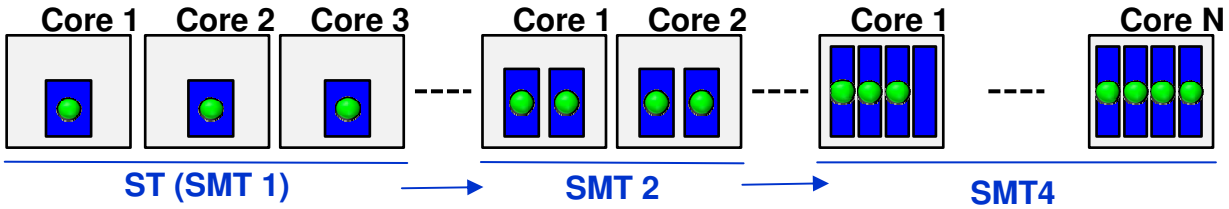
THE «VPs»

Advanced Technical Skills

VPs vs AIX SMT Scheduling



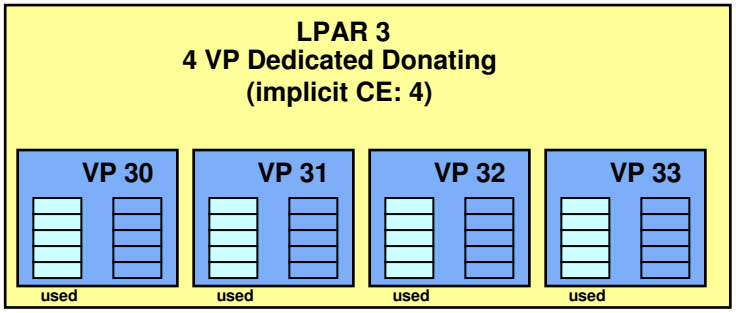
Default AIX Folding / SMT Approach



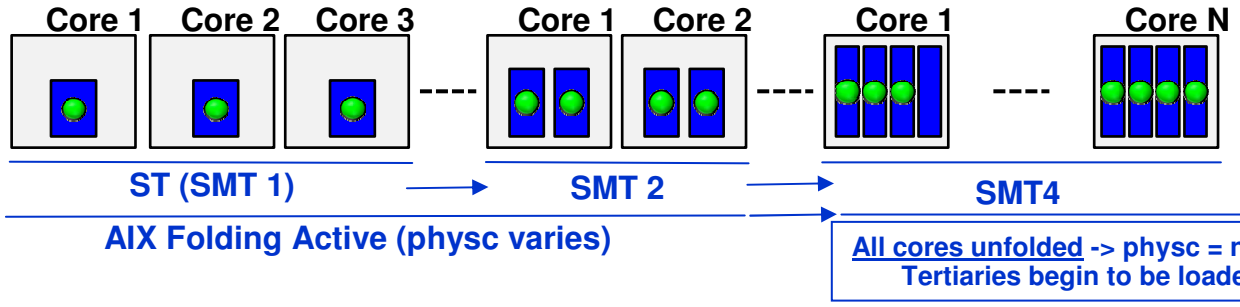
- ***Current AIX folding / SMT scheduling approach favors :***
  - **single threaded performance, faster response time, and total LPAR level throughput**
    - As load increases, **cores are unfolded to keep them Single Threaded (ST mode)** until all available LPAR cores are unfolded.
    - Some spare core capacity above what is minimally needed is held for workload spikes
    - Further load increases will start switching cores (VP) to SMT2 then to SMT4

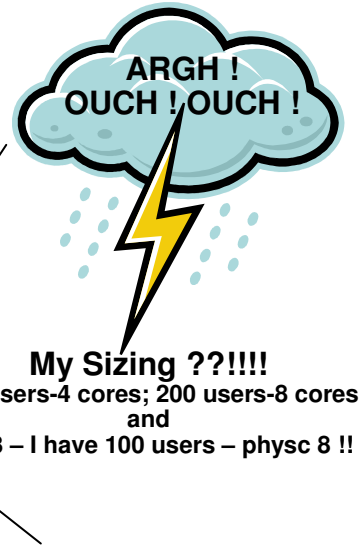
Advanced Technical Skills

VPs vs AIX Folding



Default AIX Folding / SMT Approach



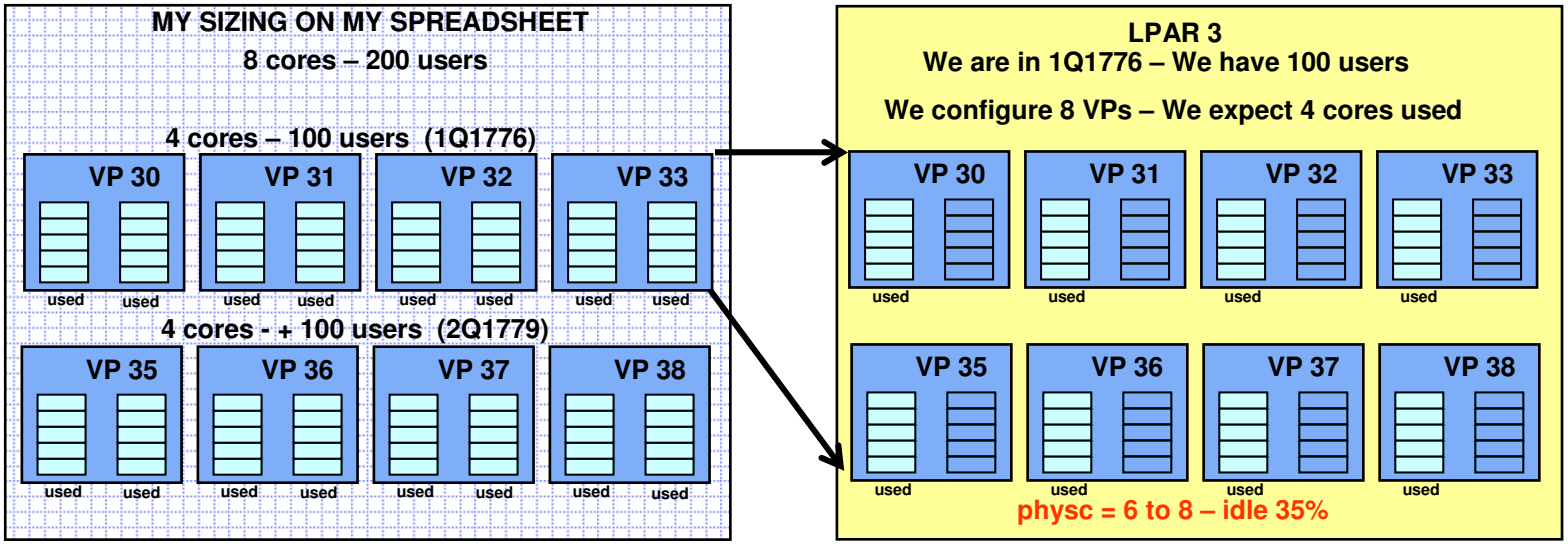


- Current AIX folding / SMT scheduling approach favors :
  - single threaded performance, faster response time, and total LPAR level throughput
    - As load increases,
      cores are unfolded to keep them single threaded **until all available LPAR cores are unfolded (VP)**

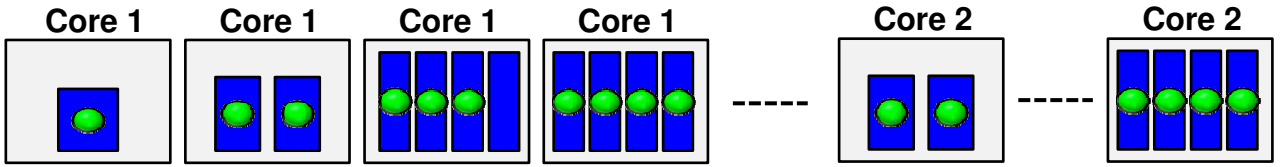
Power is performance redefined

Advanced Technical Skills

What your sizing did not tell you…



- ***Sizing does compute an equivalent of "processing" capacity***
  - It's an "in-depth" scheduling: it is not true at all.



  - "sizing are linear" : it is not true at all.

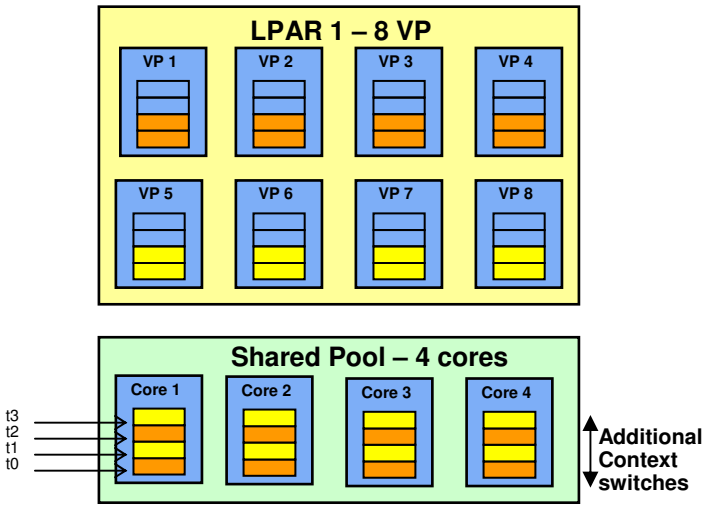
rPerf

	GHz	cores	rPerf	rPerf/core
595	5.00	64	553.01	8.64
595	5.00	48	430.53	8.97
595	5.00	32	307.12	9.60
595	5.00	16	164.67	10.29
595	5.00	8	87.10	10.89

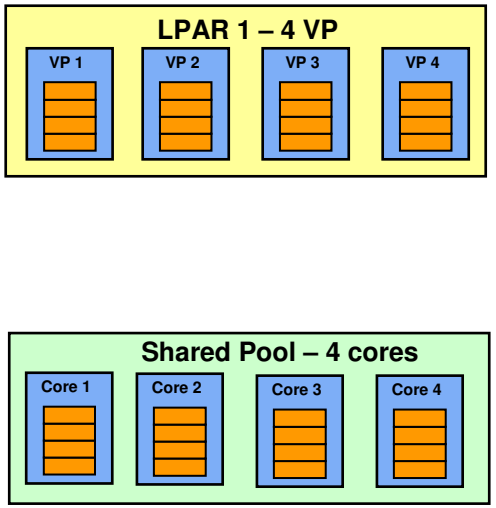
Power is performance redefined

Advanced Technical Skills

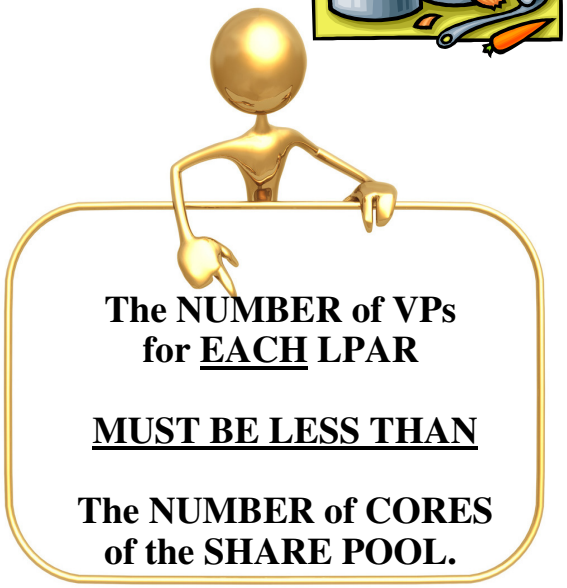
MAXIMUM VPs FOR EACH LPAR



Configuration A
POOR PERFORMER - 8 VP > 4 CORES



Configuration B
GOOD PERFORMER 4 VP < 4 CORES
For my example it is equal : it should not.



- The VPs exceeding the number of cores will be **dispatched sequentially**:
  - Defining 8 VPs (Configuration A) gives a wrong assumption on the “real” parallelism.
  - Only 4 VPs are able to run physically at the same time because we have only 4 Cores
  - ***The maximum parallelism degree (nb of VPs) for any LPAR is nb of CORES.***
- Configuration A does lose performance:
  - It can generate up to 3 VP Context Switches PLUS a hurting and useless loss of Processor affinity.

Advanced Technical Skills

Configuration A under Micro…second (< 1 s)

Shared Pool – 4 cores

t0

Core 1 | Core 2 | Core 3 | Core 4

Launch/Stop | Launch/Stop | Launch/Stop | Launch/Stop

t2

Core 1 | Core 2 | Core 3 | Core 4

Launch/Stop | Launch/Stop | Launch/Stop | Launch/Stop

CLEANING
(Save VP Context / Restore VP Context)

CLEANING
(Save VP Context / Restore VP Context)

CLEANING
(Save VP Context / Restore VP Context)

t1

Core 1 | Core 2 | Core 3 | Core 4

Launch/Stop | Launch/Stop | Launch/Stop | Launch/Stop

t3

Core 1 | Core 2 | Core 3 | Core 4

Launch/Stop | Launch/Stop | Launch/Stop | Launch/Stop

LPAR 1 – 8 VP – 32 CPUs

CE: 4 due to 4 cores in Share Pool

Ratio CE/VP : 0.5

VP 1 | VP 2 | VP 3 | VP 4

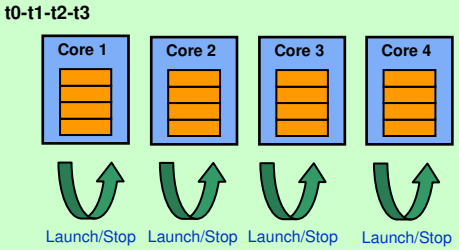
VP 5 | VP 6 | VP 7 | VP 8

0-3	4-7	8-11	12-15	run	**t0**: 16 Effective CPUs running
0-3	**4-7**	**8-11**	**12-15**	stop	**t1**: 16 Effective CPUs running
0-3	4-7	8-11	12-15	run	**t2**: 16 Effective CPUs running
0-3	**4-7**	**8-11**	**12-15**	stop	**t3**: 16 Effective CPUs running

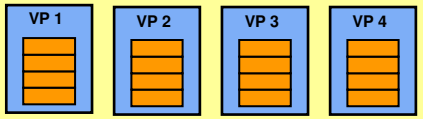
16-19	**20-23**	**24-27**	**28-31**	stop
16-19	20-23	24-27	28-31	run
16-19	**20-23**	**24-27**	**28-31**	stop
16-19	20-23	24-27	28-31	run

Let's see Configuration B under microscope

Shared Pool – 4 cores



- Less work for PowerVM Hypervisor
  - ✓ More performance for the LPAR
- Processor Affinity (No extra. Cleaning work).
  - ✓ More performance for the LPAR



0-3	4-7	8-11	12-15
0-3	4-7	8-11	12-15
0-3	4-7	8-11	12-15
0-3	4-7	8-11	12-15

LPAR 1 – 4 VP – 16 CPUs

CE: 4 due to 4 cores in Share Pool

Ratio CE/VP : 1.0

t0: 16 Logical CPUs running

t1: 16 Logical CPUs running

t2: 16 Logical CPUs running

t3: 16 Logical CPUs running

- Always 16 "Effective" CPUs running
  - Continuously vs Disruptive

Power is performance redefined

PowerVM Published Informations

- At http://www.ibm.com/developerworks/wikis/display/WikiPtype/Performance+Monitoring+Documentation

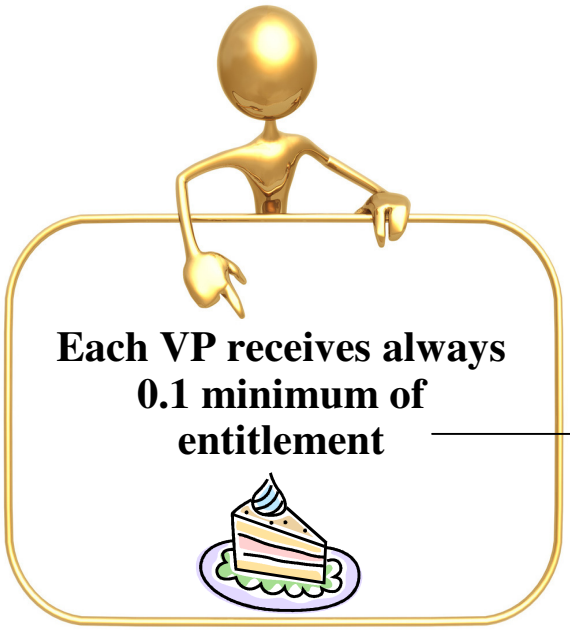
AIX Online Manual	The AIX Performance Management Guide	Very good base information - good luck finding it in Info Center but it is there somewhere Here is an early version of the Guide (for AIX V5.1.)
Web Article	Setting up AIX Workload Manager in 30 minutes	Excellent starter pack for WLM
Web Article	Methods for Detecting Memory Leaks in AIX Systems	Provides tools and techniques that can be used to isolate a memory leak in AIX process/application.
Web Article	VMM Tuning Tip: Protecting Computational Memory	This document provides guidance to tune the VMM to protect computational memory. How to tune maxclient%, maxperm%, lru_file_repage, strict_maxperm and strict_maxclient.
Web Article	Understanding DIO & CIO	This document provides information on DIO and CIO, and provides tools to assistance in detecting problems with CIO and DIO.
Web Article	Understanding Processor Utilization on POWER Systems - AIX	Article explains how the processor utilization is calculated and need to be interpreted on POWER 5, 6 and 7 architectures
Word document	P7 Virtualization Best Practice	This document is intended to address POWER7 PowerVM best practices to attain best LPAR performance. This document should be used in conjunction with other PowerVM documents.

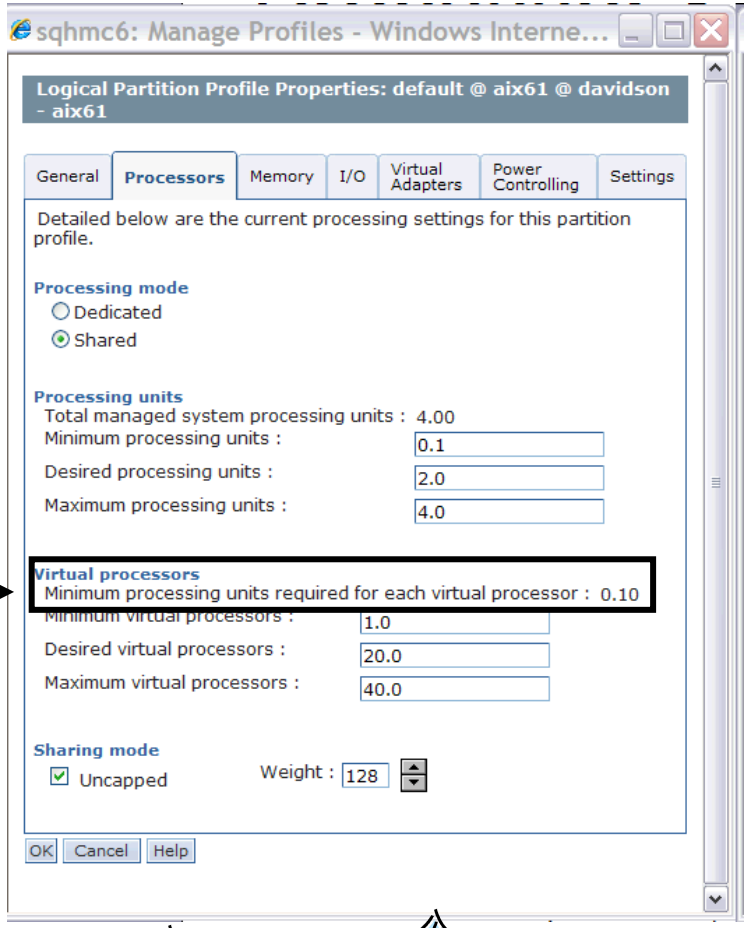
- "P7 Virtualization Best Practice" – Page 5

" 1.1.1 Sizing/configuring virtual processors:

The number of virtual processors in each LPAR in the system should not "exceed" the number of cores available in the system (CEC/framework) or if the partition is defined to run in specific virtual shared processor pool, the number of virtual processors should not exceed the maximum defined for the specific virtual shared processor pool. Having more virtual processors configured than can be running at a single point in time does not provide any additional performance benefit and can actually cause additional context switches of the virtual processors reducing performance.

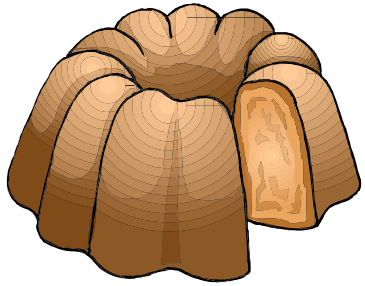
PowerVM : Minimum of entitlement per VP





THE «VPs»

0.x CE PER VP
(Desired CE)



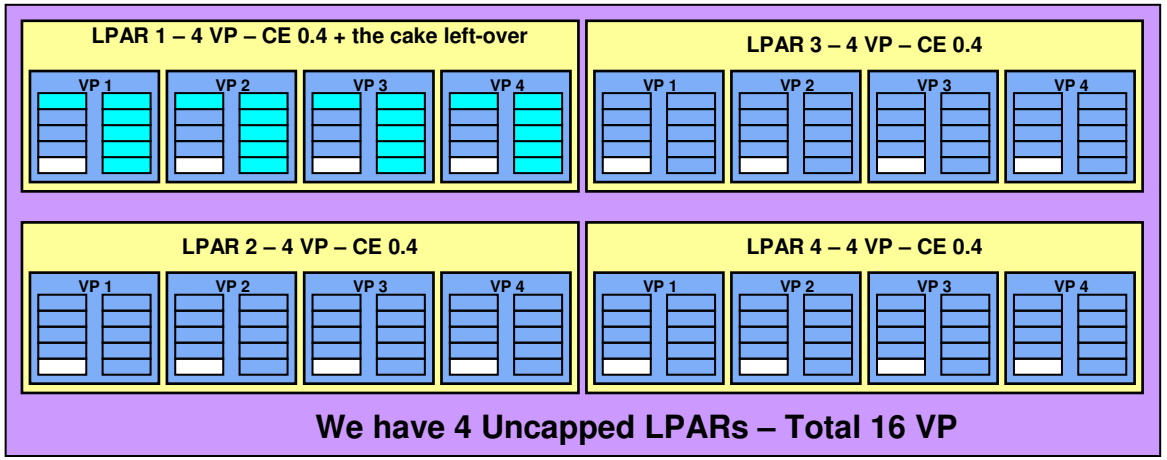
SHARE POOL
SLICES = NB. CORES x 10

Power is performance redefined

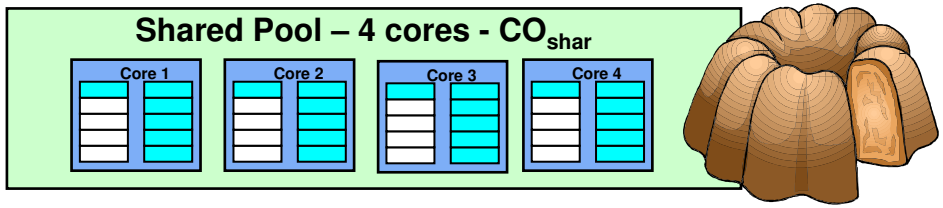
Advanced Technical Skills

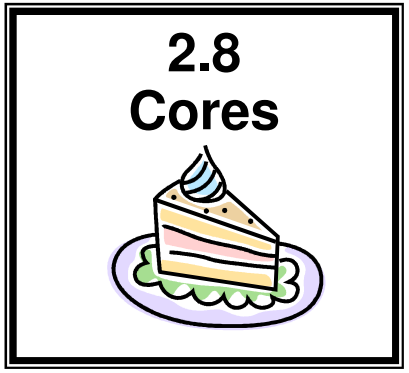
The CAKE: Who's eating the cake tonight ?

It's LPAR 1 !



You are LPAR1





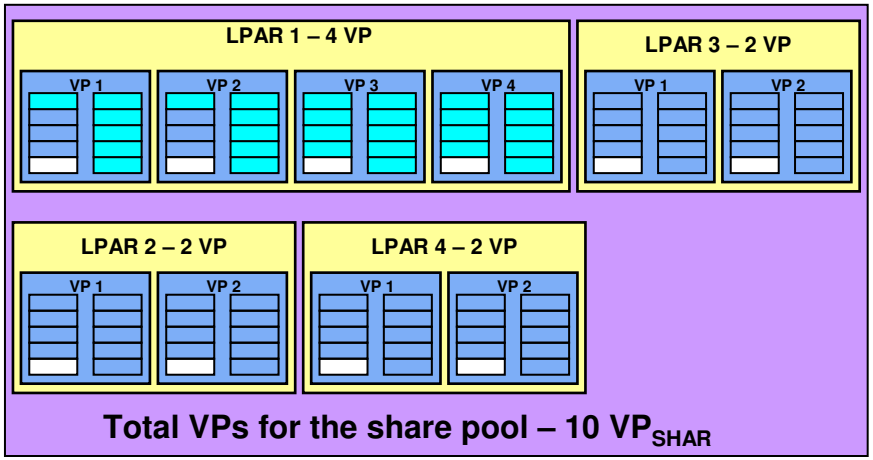
- ***Minimum to give to "the others"***
  - The total of LPAR 2,3,4 deliver with the minimum each: ***(VP_{total} - VP_{LPAR1})* 0.1***

    Here, you can not eat and you must give : (16 – 4) * 0.1 = 1.2 cores

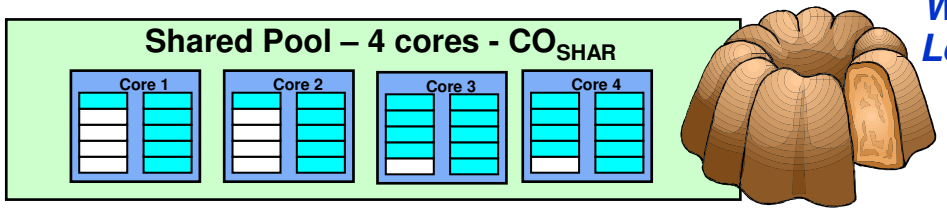
- ***LPAR 1 can consume at maximum = Max physc***
  - LPAR 1 can only consume a maximum of: ***CO_{shar} – Minimum to Give***
    Here, 4 – 1.2 = **2.8** as **Physical Consumed for LPAR 1**

Advanced Technical Skills

The CAKE: How to eat MORE cake tonight ?



You are LPAR1



We reduce the "others" VPs !!
Less Guests, More Cake (sic).

Who's in charge of your Physc ?
"The others VPs" (ouch, ouch).

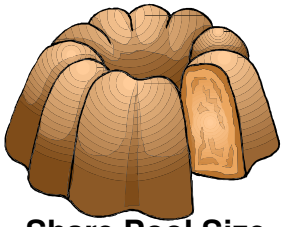
- ***Minimum to give to "the others"***
  - The total of LPAR 2,3,4 deliver with the minimum each: $VP_{total} - VP_{LPAR1} * 0.1$

    Here, you can not eat and you must give : (10– 4) * 0.1 = 0.6 cores
- ***LPAR 1 can consume at maximum = Max physc***
  - LPAR 1 can only consume a maximum of: CO_{shar} ***– Minimum to give***
    Here, 4 – 0.6 = **3.4** as **Physical Consumed for LPAR 1**

Advanced Technical Skills

The CAKE: What we have seen is so far

THE «others VPs»

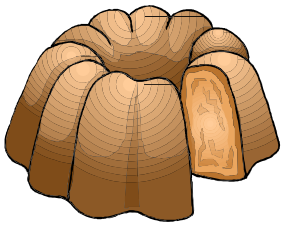


Share Pool Size
(nb of cores)

- ***"Reduce the others" means there's a ratio between nb of Cores and the nb of VPs***
  - Example 1 : ratio = 4 (16 VPs / 4 cores) – LPAR 1: max. physc is 2.8
  - Example 2 : ratio = 2.5 (10 VPs / 4 cores) – LPAR 1: max. physc is 3.4
- ***People increase the number of VPs for an LPAR…***
  - Thinking it guarantees a growth capacity of the LPAR.
    - For real, it increases automatically the physc with idle%
    - It activates folding activity.
    - It decreases the uncapped capacity of all others LPARs.
    - People doing server sizing become stressed or angry
- ***Large Share Pool with large uncapped are less predictable.***

Advanced Technical Skills

What your sizing is not telling you clearly…

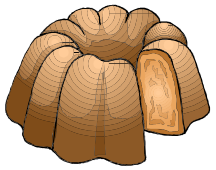


Share Pool Size
60 cores Techno X

180 VPs defined : ratio 3

Sizing by rPerf: Techno Y = 1.5 x Techno X

Reducing the number of cores



Share Pool Size
40 cores Techno Y

You reduce the number of cores

You reduce your nb of VPs!

180 VPs defined : ratio 4.5
should become
120 VPs defined : ratio 3

- ***As rPerf tells you :***
  - Because the core is more powerful (4 arms)….you have sized with less cores.
  - Be fair, you buy less cores. Thus, use less VPs for your LPARs to keep at least the original ratio.
- ***Mostly doing migration to P7 servers,***
  - People do not readjust the number of VPs.
  - BUT the number of cores has sometimes drastically decreased and thus, the ratio drastically increased.
  - ***Avoid Global Frame computation***

-> Thus, Re-Adjust your number of VPs per LPAR.

Advanced Technical Skills

VP are always dispatched…except if folded

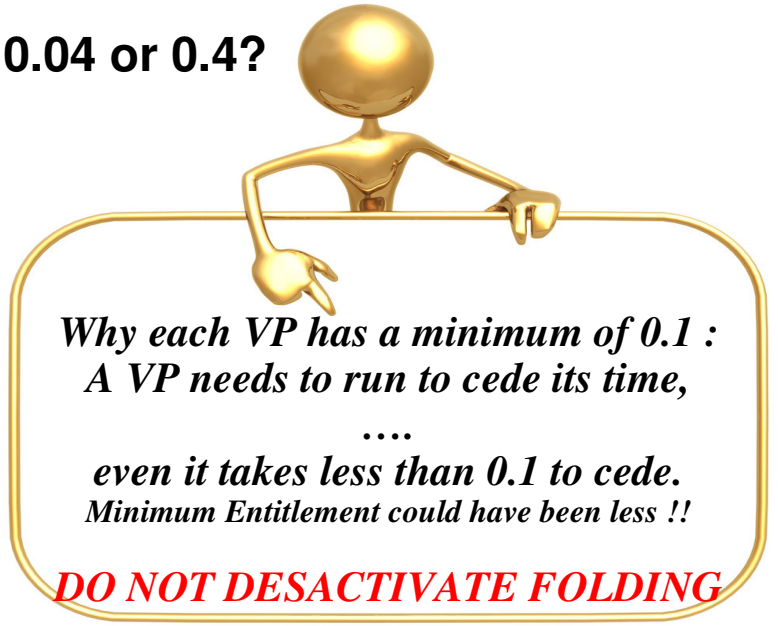
My machine is a POWER5 SMT2 with 4 VPs. My Machine is idle : physc is 0.01. Why not 0.04 or 0.4?

```
System configuration: lcpu=8 mem=2048MB ent=4.00

thr    memory                 page                    faults                cpu
---- ----------- ------------------------ ------------ -----------------------
r  b    avm   fre  re  pi  po  fr   sr  cy  in   sy  cs us sy id wa    pc    ec
4  0 420690 55828   0   0   0   0    0   0   2  138 182  0  0 99  0  0.01   0.2
4  0 420690 55828   0   0   0   0    0   0   2  314 372  0  0 99  0  0.01   0.2
4  0 420690 55828   0   0   0   0    0   0   3  101 180  0  0 99  0  0.01   0.2
5  0 420690 55828   0   0   0   0    0   0   5  511 370  0  0 99  0  0.01   0.2
4  0 420690 55828   0   0   0   0    0   0   3  355 398  0  0 99  0  0.01   0.2
```

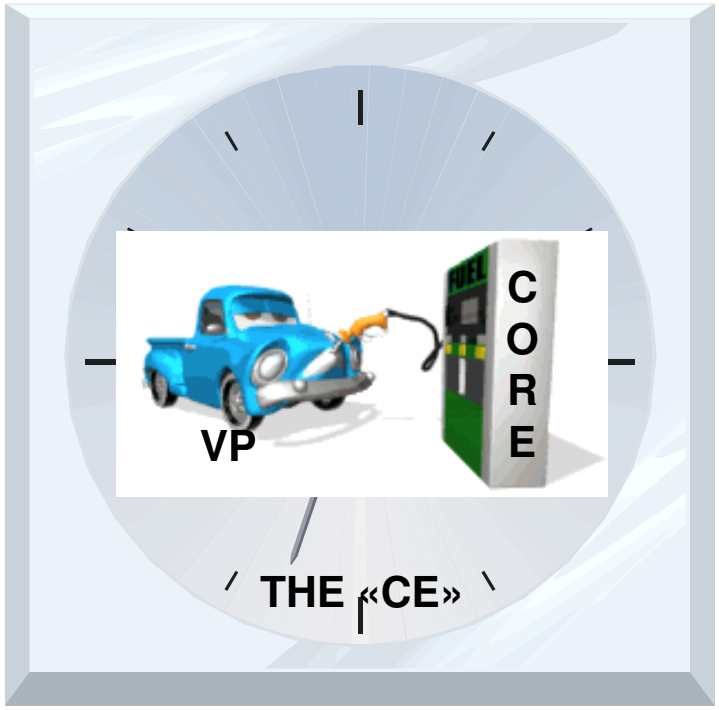
Why my physc is 0.01 and not 0.04 or 0.4?

physc is 0.01 because only 1 VP is unfolded (principle to be idle).

```
VSD Thread State.
CPU VP_STATE    SLEEP_STATE

  0  ACTIVE     AWAKE
  1  ACTIVE     AWAKE
  2  DISABLED   SLEEPING
  3  DISABLED   SLEEPING
  4  DISABLED   SLEEPING
  5  DISABLED   SLEEPING
  6  DISABLED   SLEEPING
  7  DISABLED   AWAKE
```



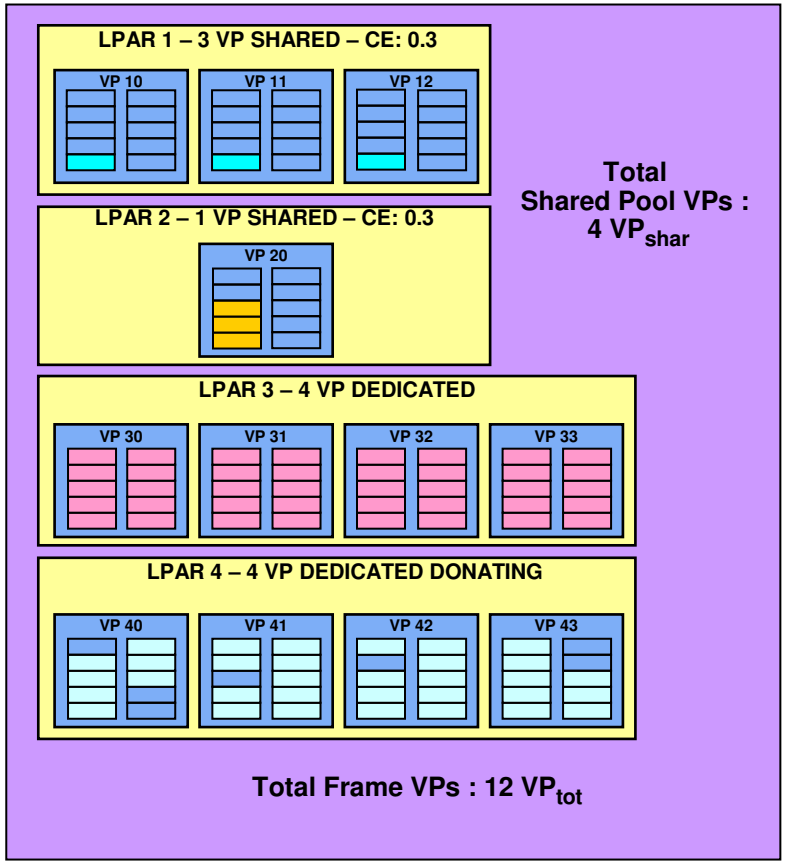
The “Capacity Entitlement”

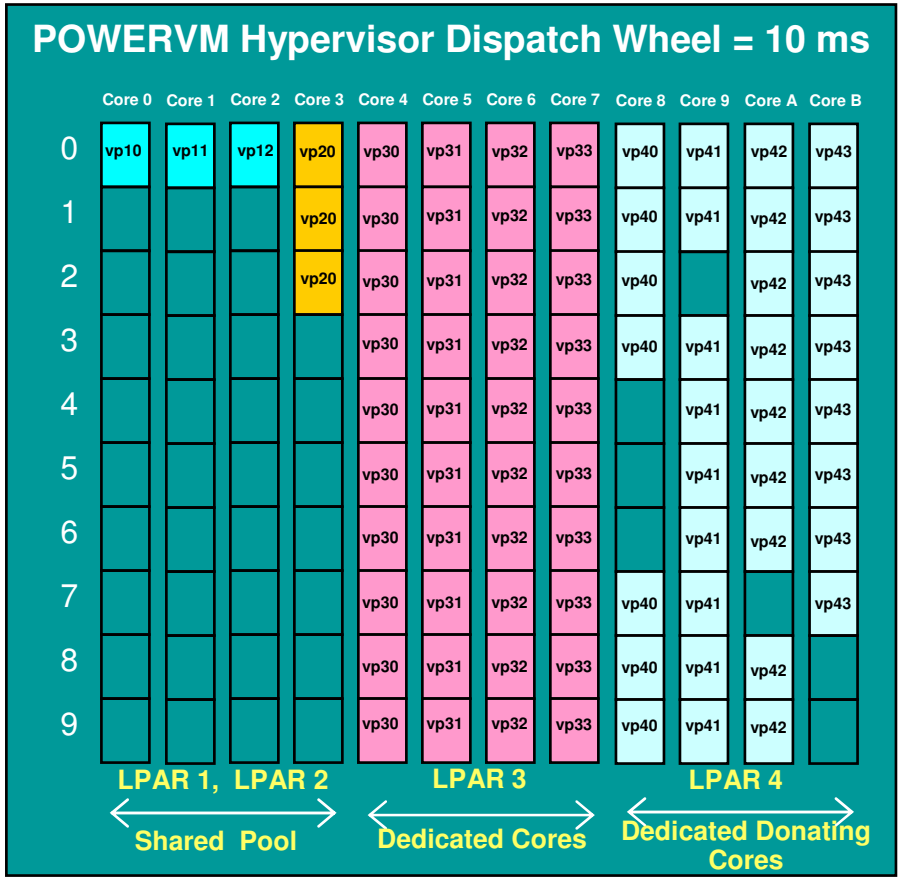


Power is performance redefined

Advanced Technical Skills

PowerVM: Capacity Entitlement – Dispatch Wheel



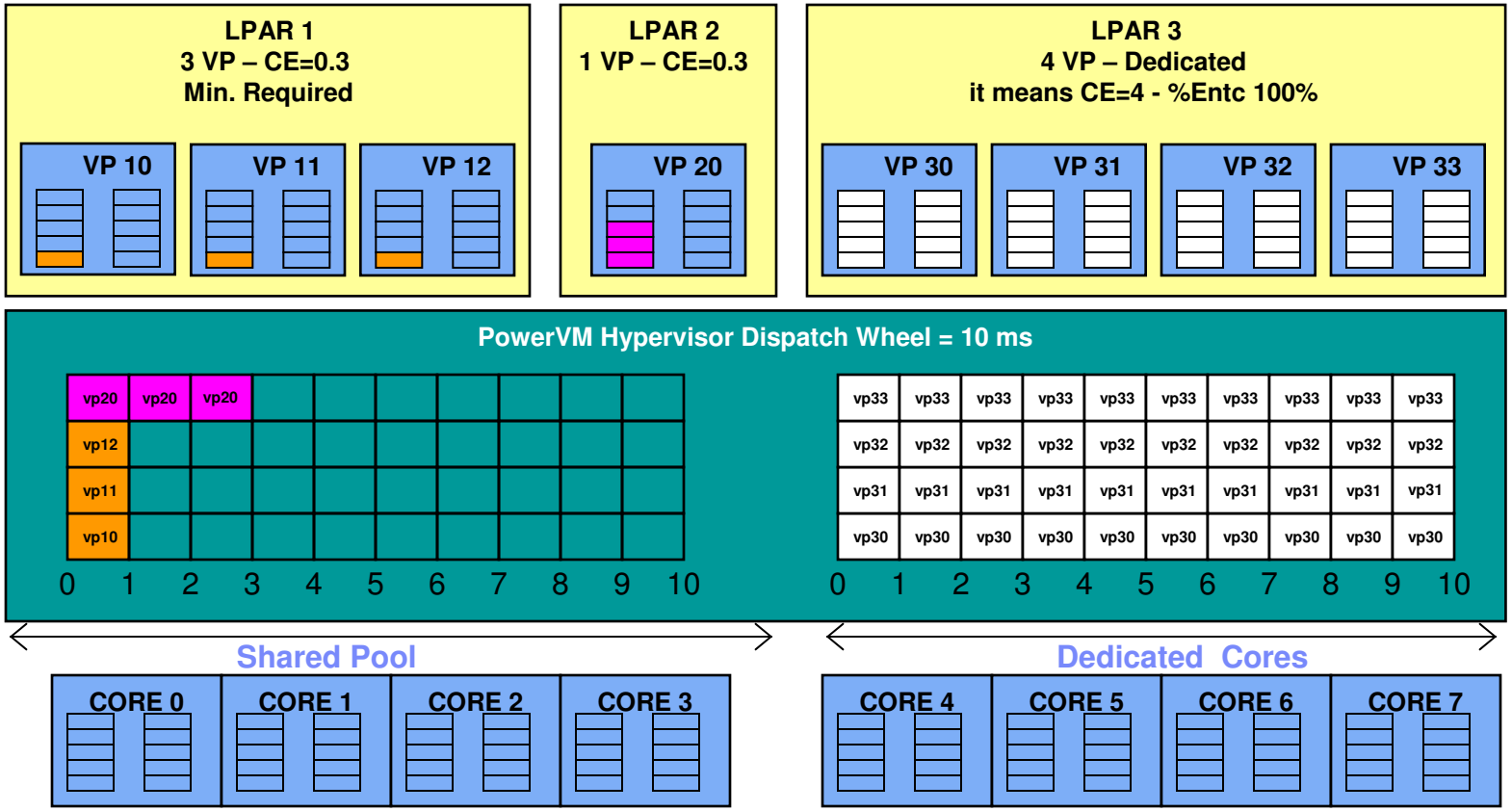


- ***Each LPAR will receive its CAPACITY Entitlement (CE) every 10 ms (dispatch wheel).***
  - Major difference between «Shared», «Dedicated», «Dedicated Donating»: **IDLE CYCLES of Core**.
- Shared Pool: Processor Affinity is optimum due a fabulous ratio of $4VP_{SHAR} = 4\ CO_{SHAR}$.
- Shared Pool: Processor Affinity determined by pHyp based on the CE of each Shared LPAR.

Power is performance redefined

Advanced Technical Skills

PowerVM: The ratio CE/VP - Access to the core

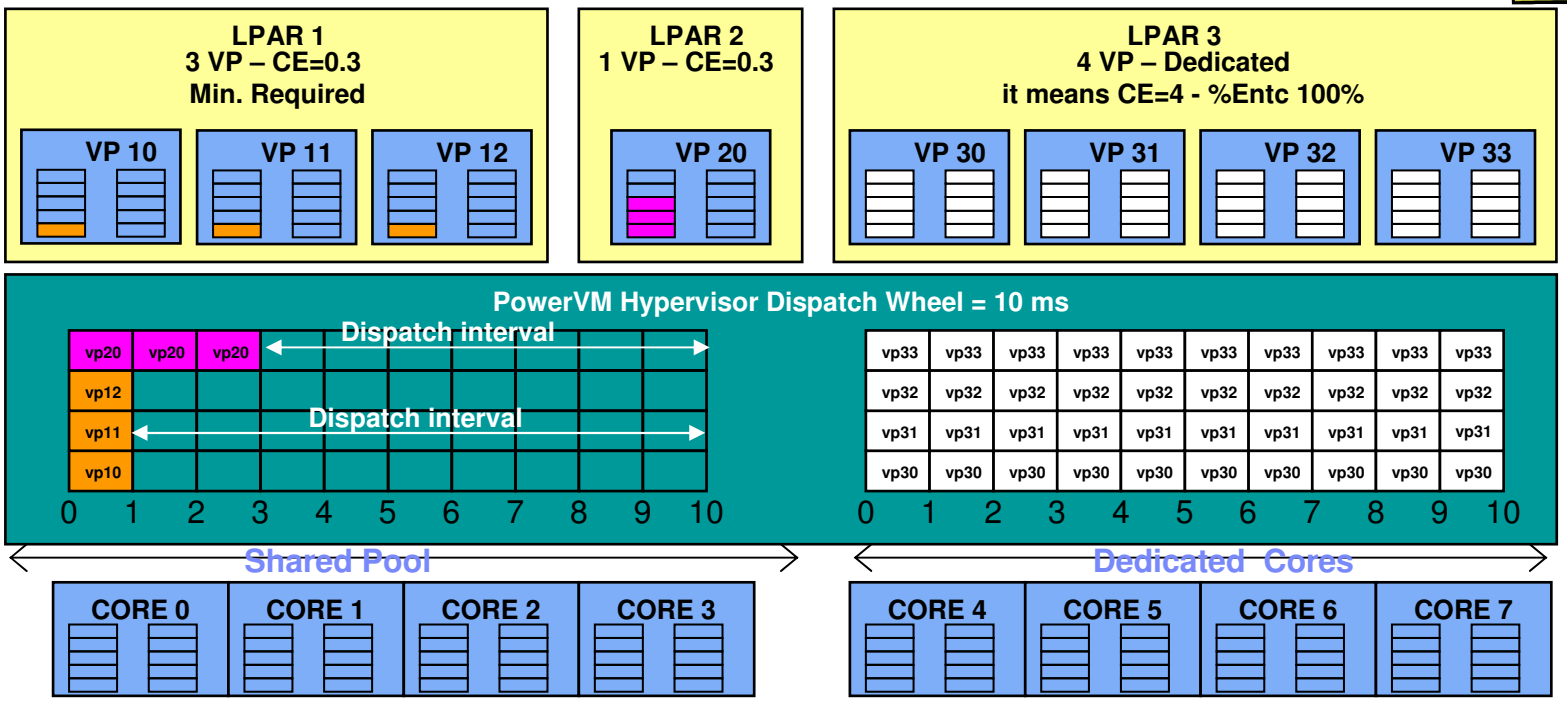


- ***Each LPAR will receive its CAPACITY Entitlement every 10 ms (dispatch wheel).***
  - CE_{VP} is a ratio (CE_{LPAR}/ VP_{LPAR}).
- With the minimum of 0.1 of CE/VP,
  - the Virtual processor has accessed **during 1 ms only to the core.**

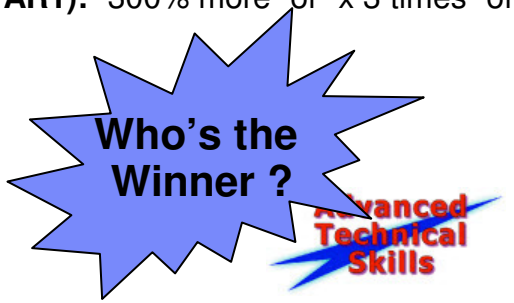
$$\textbf{CORE}_{\textbf{Access}}\ \textbf{in ms} = (CE_{LPAR}/VP_{LPAR}) * 10$$

Advanced Technical Skills

"Limit the VPs, Core Access you receive"

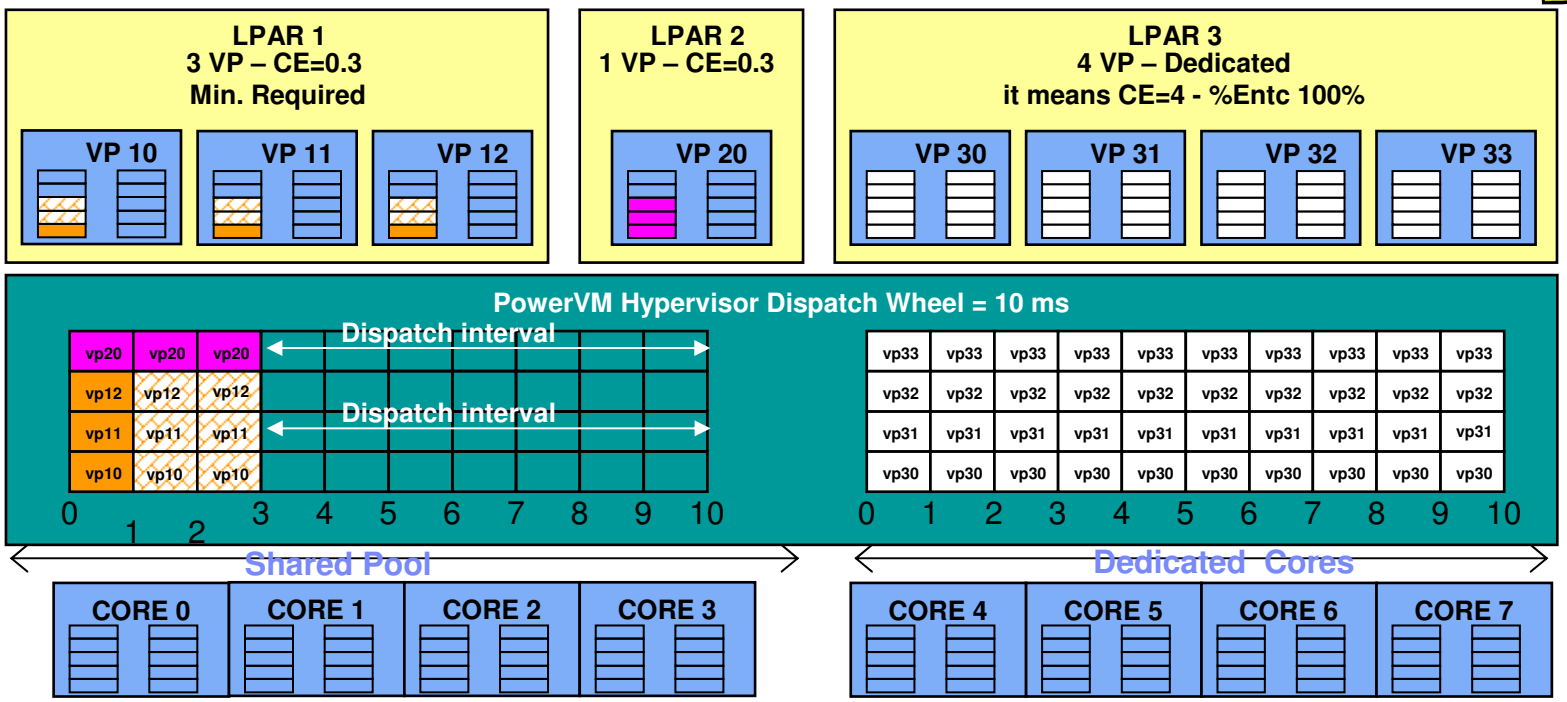


- **For the SAME CE_{LPAR}, let's compare Performance of LPAR 1 with Performance of LPAR 2:**
  - **3 ms of core access (LPAR2)** vs **1 ms of core access (LPAR1):** "300% more" or "x 3 times" of core access !!
  - **Dispatch Interval**:
    - LPAR 1 : dispatch between 9 ms to 18 ms.
    - LPAR 2 : dispatch between 7 ms to 14 ms



Power is performance redefined

Welcome to "Adventures in Uncapland"

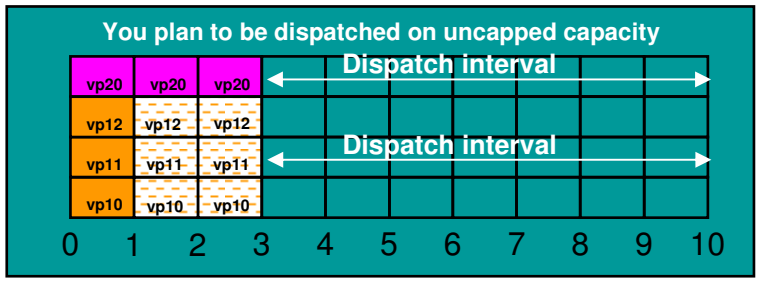


- ***My performance of LPAR 1 can be rescued by the UNCAPPED feature.***
  - Same time of core access: 3 ms of core access.
  - Same dispatch Interval: LPAR 1 : between 7 ms to 14 ms

BUT,
"Who's eating the cake tonight ?"

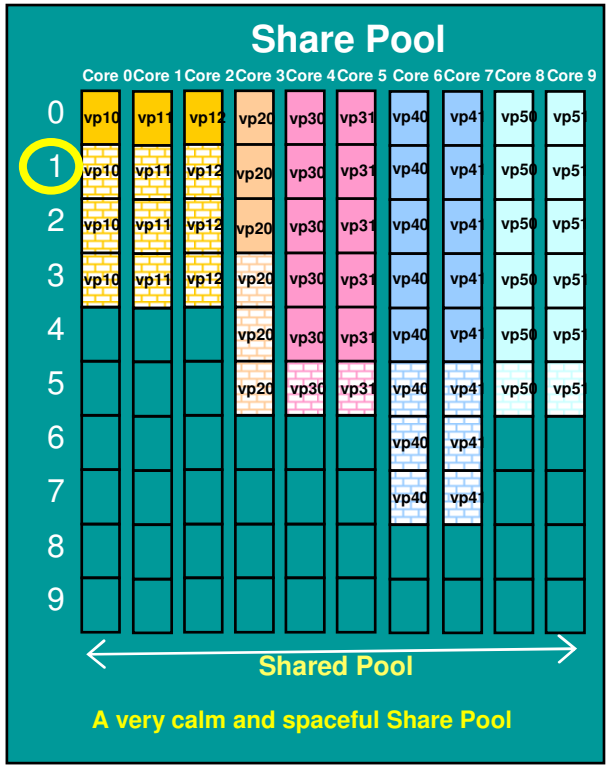
Advanced Technical Skills

"Adventures in Uncapland" is Adventures first



CAKE: 10 cores
VPs : 10 VPs

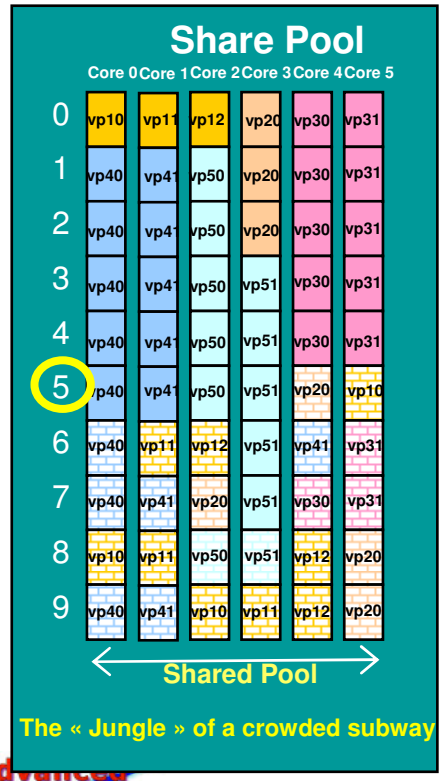
Your «VPs» can be dispatched at 0.1ms on SAME core.



GOOD FOR PERF

CAKE: 6 cores
VPs : 10 VPs

Your «VPs» can be dispatched only at 0.5 ms on ANOTHER core (core5).



BAD FOR PERF

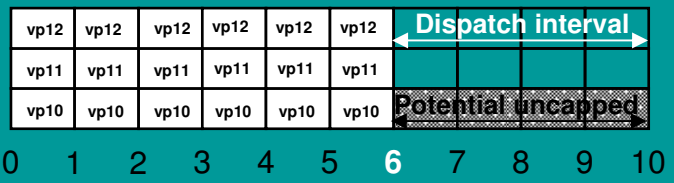
Story of the "Interrupt at 3ms"

Advanced Technical Skills

Power is performance redefined

Shared LPAR: Estimate Uncapped Exposure

PowerVM Hypervisor Dispatch Wheel = 10 ms



- Try to use the cores you bought !
- Not to spend your time in dispatch wait.

CORE Access in ms = (CE_{LPAR}/VP_{LPAR}) * 10

6ms = (CE_{LPAR}/VP_{LPAR}) * 10

0.6 = CE_{LPAR} / VP_{LPAR}

Or

VP_{LPAR} = 1.67 CE_{LPAR}

(CE/VP)
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Performance for growing uncapped is not safe here.

LPAR1 VP 1 - CE 0.1 is OK if physc 0.1 or 0.2.

LPAR2 VP 1 - CE 0.1 to physc 1.0 is NOT OK
0.9 to gain "Adventures in Uncapland"

Performance for uncapped is better : from acceptable to good.
LPAR2 VP 2- CE 1.2 capped is OK

LPAR2 VP 2- CE 1.2 physc 2.0 is better
"Adventures in UncapLand" is more limited.

Power is performance redefined

** CE is based on the expected number of sized cores.*

Advanced Technical Skills

Conclusions

We have reviewed :

- Part I : Basic Concepts of PowerVM.
- Part II : Technical Insights on the balance on CE / VP for shared LPARs.

We hope this will help you to size, configure and implement Power7 Servers.

Advanced Technical Skills

IBM Systems Lab Services and Training

Helping you gain the IBM Systems skills needed for smarter computing

- Comprehensive education, training and service offerings
- Expert instructors and consultants, world-class content and skills
- Multiple delivery options for training and services
- Conferences explore emerging trends and product strategies

www.ibm.com/training

Special Programs:

- IBM Systems 'Guaranteed to Run' Classes -- ***Make your education plans for classes with confidence!***
- Instructor-led online (ILO) training ***The classroom comes to you.***
- Customized, private training
- Lab-based services assisting in high tech solutions

Notes on benchmarks and values

The IBM benchmarks results shown herein were derived using particular, well configured, development-level and generally-available computer systems. Buyers should consult other sources of information to evaluate the performance of systems they are considering buying and should consider conducting application oriented testing. For additional information about the benchmarks, values and systems tested, contact your local IBM office or IBM authorized reseller or access the Web site of the benchmark consortium or benchmark vendor.

IBM benchmark results can be found in the IBM Power Systems Performance Report at http://www.ibm.com/systems/p/hardware/system_perf.html .

All performance measurements were made with AIX or AIX 5L operating systems unless otherwise indicated to have used Linux. For new and upgraded systems, AIX Version 4.3, AIX 5L or AIX 6 were used. All other systems used previous versions of AIX. The SPEC CPU2006, SPEC2000, LINPACK, and Technical Computing benchmarks were compiled using IBM's high performance C, C++, and FORTRAN compilers for AIX 5L and Linux. For new and upgraded systems, the latest versions of these compilers were used: XL C Enterprise Edition V7.0 for AIX, XL C/C++ Enterprise Edition V7.0 for AIX, XL FORTRAN Enterprise Edition V9.1 for AIX, XL C/C++ Advanced Edition V7.0 for Linux, and XL FORTRAN Advanced Edition V9.1 for Linux. The SPEC CPU95 (retired in 2000) tests used preprocessors, KAP 3.2 for FORTRAN and KAP/C 1.4.2 from Kuck & Associates and VAST-2 v4.01X8 from Pacific-Sierra Research. The preprocessors were purchased separately from these vendors. Other software packages like IBM ESSL for AIX, MASS for AIX and Kazushige Goto's BLAS Library for Linux were also used in some benchmarks.

For a definition/explanation of each benchmark and the full list of detailed results, visit the Web site of the benchmark consortium or benchmark vendor.

TPC	http://www.tpc.org
SPEC	http://www.spec.org
LINPACK	http://www.netlib.org/benchmark/performance.pdf
Pro/E	http://www.proe.com
GPC	http://www.spec.org/gpc
VolanoMark	http://www.volano.com
STREAM	http://www.cs.virginia.edu/stream/
SAP	http://www.sap.com/benchmark/
Oracle Applications	http://www.oracle.com/apps_benchmark/
PeopleSoft - To get information on PeopleSoft benchmarks, contact PeopleSoft directly	
Siebel	http://www.siebel.com/crm/performance_benchmark/index.shtm
Baan	http://www.ssaglobal.com
Fluent	http://www.fluent.com/software/fluent/index.htm
TOP500 Supercomputers	http://www.top500.org/
Ideas International	http://www.ideasinternational.com/benchmark/bench.html
Storage Performance Council	http://www.storageperformance.org/results

Revised March 12, 2009

Notes on HPC benchmarks and values

The IBM benchmarks results shown herein were derived using particular, well configured, development-level and generally-available computer systems. Buyers should consult other sources of information to evaluate the performance of systems they are considering buying and should consider conducting application oriented testing. For additional information about the benchmarks, values and systems tested, contact your local IBM office or IBM authorized reseller or access the Web site of the benchmark consortium or benchmark vendor.

IBM benchmark results can be found in the IBM Power Systems Performance Report at http://www.ibm.com/systems/p/hardware/system_perf.html .

All performance measurements were made with AIX or AIX 5L operating systems unless otherwise indicated to have used Linux. For new and upgraded systems, AIX Version 4.3 or AIX 5L were used. All other systems used previous versions of AIX. The SPEC CPU2000, LINPACK, and Technical Computing benchmarks were compiled using IBM's high performance C, C++, and FORTRAN compilers for AIX 5L and Linux. For new and upgraded systems, the latest versions of these compilers were used: XL C Enterprise Edition V7.0 for AIX, XL C/C++ Enterprise Edition V7.0 for AIX, XL FORTRAN Enterprise Edition V9.1 for AIX, XL C/C++ Advanced Edition V7.0 for Linux, and XL FORTRAN Advanced Edition V9.1 for Linux. The SPEC CPU95 (retired in 2000) tests used preprocessors, KAP 3.2 for FORTRAN and KAP/C 1.4.2 from Kuck & Associates and VAST-2 v4.01X8 from Pacific-Sierra Research. The preprocessors were purchased separately from these vendors. Other software packages like IBM ESSL for AIX, MASS for AIX and Kazushige Goto's BLAS Library for Linux were also used in some benchmarks.

For a definition/explanation of each benchmark and the full list of detailed results, visit the Web site of the benchmark consortium or benchmark vendor.

SPEC	http://www.spec.org
LINPACK	http://www.netlib.org/benchmark/performance.pdf
Pro/E	http://www.proe.com
GPC	http://www.spec.org/gpc
STREAM	http://www.cs.virginia.edu/stream/
Fluent	http://www.fluent.com/software/fluent/index.htm
TOP500 Supercomputers	http://www.top500.org/
AMBER	http://amber.scripps.edu/
FLUENT	http://www.fluent.com/software/fluent/fl5bench/index.htm
GAMESS	http://www.msg.chem.iastate.edu/gamess
GAUSSIAN	http://www.gaussian.com
ANSYS	http://www.ansys.com/services/hardware-support-db.htm Click on the "Benchmarks" icon on the left hand side frame to expand. Click on "Benchmark Results in a Table" icon for benchmark results.
ABAQUS	http://www.simulia.com/support/v68/v68_performance.php
ECLIPSE	http://www.sis.slb.com/content/software/simulation/index.asp?seg=geoquest&
MM5	http://www.mmm.ucar.edu/mm5/
MSC.NASTRAN	http://www.mscsoftware.com/support/prod%5Fsupport/nastran/performance/v04_sngl.cfm
STAR-CD	www.cd-adapco.com/products/STAR-CD/performance/320/index/html
NAMD	http://www.ks.uiuc.edu/Research/namd
HMMER	**http://hmmer.janelia.org/** http://powerdev.osuosl.org/project/hmmerAltivecGen2mod

Revised March 12, 2009

Notes on performance estimates

rPerf for AIX

rPerf (Relative Performance) is an estimate of commercial processing performance relative to other IBM UNIX systems. It is derived from an IBM analytical model which uses characteristics from IBM internal workloads, TPC and SPEC benchmarks. The rPerf model is not intended to represent any specific public benchmark results and should not be reasonably used in that way. The model simulates some of the system operations such as CPU, cache and memory. However, the model does not simulate disk or network I/O operations.

- rPerf estimates are calculated based on systems with the latest levels of AIX and other pertinent software at the time of system announcement. Actual performance will vary based on application and configuration specifics. The IBM eServer pSeries 640 is the baseline reference system and has a value of 1.0. Although rPerf may be used to approximate relative IBM UNIX commercial processing performance, actual system performance may vary and is dependent upon many factors including system hardware configuration and software design and configuration. Note that the rPerf methodology used for the POWER6 systems is identical to that used for the POWER5 systems. Variations in incremental system performance may be observed in commercial workloads due to changes in the underlying system architecture.

All performance estimates are provided "AS IS" and no warranties or guarantees are expressed or implied by IBM. Buyers should consult other sources of information, including system benchmarks, and application sizing guides to evaluate the performance of a system they are considering buying. For additional information about rPerf, contact your local IBM office or IBM authorized reseller.

==========

CPW for IBM i

Commercial Processing Workload (CPW) is a relative measure of performance of processors running the IBM i operating system. Performance in customer environments may vary. The value is based on maximum configurations. More performance information is available in the Performance Capabilities Reference at: www.ibm.com/systems/i/solutions/perfmgmt/resource.html.

Revised April 2, 2007