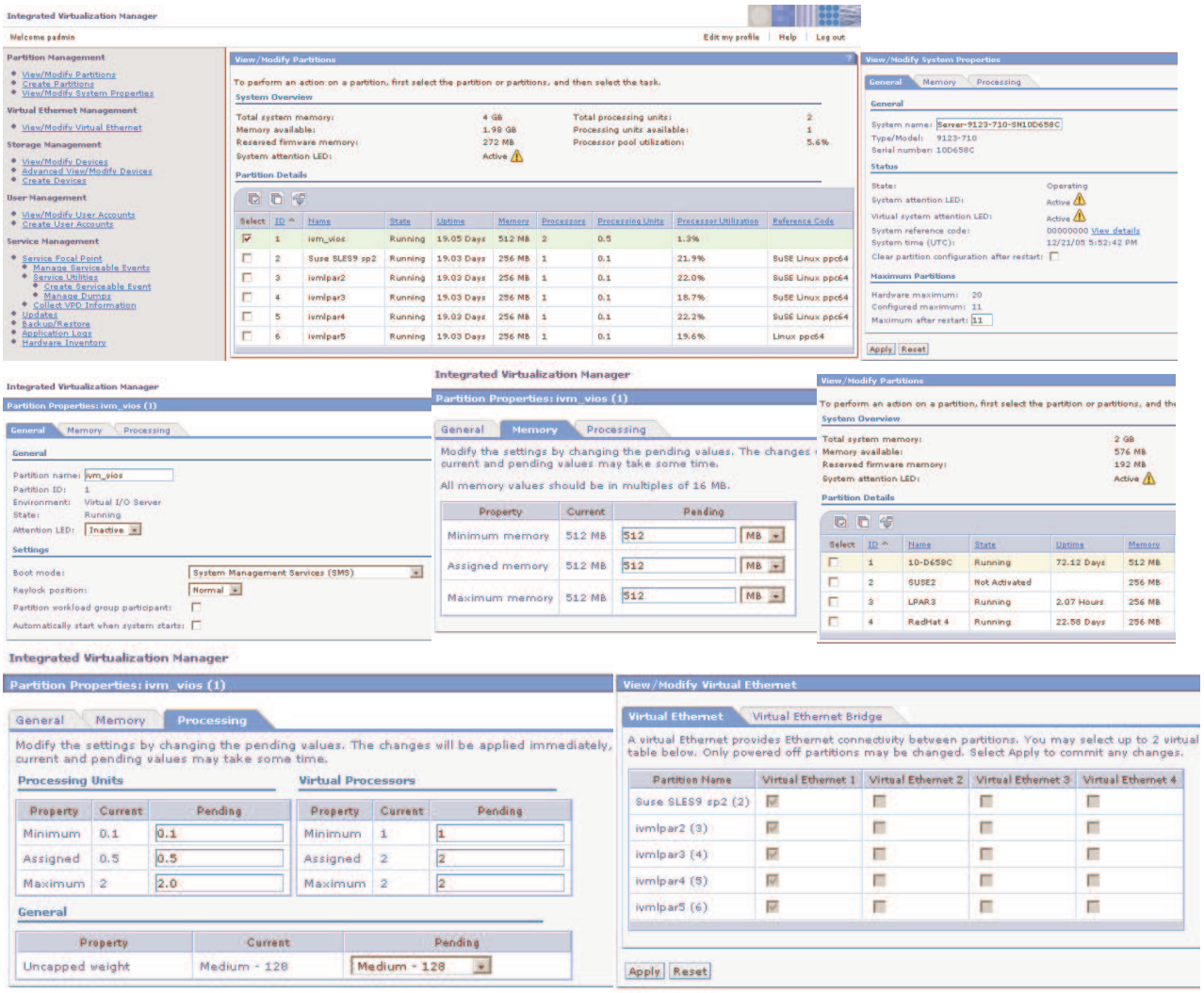
Integrated Virtualisation Manager (IVM)

Advanced Hands-On Topics



Author: Nigel Griffiths
pSeries Technical Sales Support, UK.
Virtualisation Technical Focus Group Leader (Europe)
Contact: nag@uk.ibm.com,

Document: IVM_FAQ.doc
Version: 1.6 - Monday, 09 January 2006

Topics Covered are:

1. Assumptions and Goals of this document.
2. Key places to get more information on Integrated Virtualisation Manager (IVM):
3. Can I set up the VIO Server without using the Flexible Service Processor (FSP) IP addresses setup and ASMI via a network?
4. Performance - Explain the performance numbers on the IVM administration web interface?
5. The Performance Numbers are not quite what you might expect.
6. Apparent Processor Utilization time lag issues?
7. Why aren't these Processor Utilization numbers automatically updated?
8. If I am mirroring disks on the VIO Server, how can I ensure they are on different disks?
9. What sort of disks are recommended with IVM for data protection?
10. Should I use rootvg for VIO client virtual disks?
11. How do I stop/halt a partition that is completely stuck?
12. What appears in the "Reference code" on View Partitions panel when there is no problem?
13. Processor Utilization numbers are missing on the View Partitions Panel?
14. How do I get out of the mkvt command?
15. User Management – what is the point?
16. How do I remove a storage pool?
17. Why can't I delete a partition and get the error message as below?
18. Will the CD be the default boot device?
19. Shutting down – what is the best way?
20. What is the difference between View/Modify Devices and Advanced View/Modify Devices?
21. If you give a whole physical disk to a partition, is it a physical disk or a virtual physical disk?
22. Can I remove a disk from a storage pool?
23. Is moving the optical device (CD or DVD) safe?
24. Why won't Linux shutdown when using the IVM web interface "shutdown" option?
25. How to avoid this long Linux shut down wait and the subsequent error?
26. Why telnet to the VIO Server partition, when ssh is much more secure?
27. How do I get the System Attention amber LED switched off?
28. Removing the Attention LED setting for a partition?
29. Where does all the memory go?
30. Can I control the "Reserved firmware memory"?
31. What should I do about the little bit of free memory left over?
32. Is the memory release when I stop the partition?
33. Is the CPU released when I stop the LPAR?
34. Performance - How can I ensure the VIO Server is not staved of CPU?
35. Performance – How much memory is right for the VIO Server?
36. What is the point of the Processing Units Minimum and Maximum do?
37. What is the point of the Memory Minimum and Maximum do?
38. What is the point of the Virtual Processor Minimum and Maximum do?
39. Performance - To cap or not to cap, that is the question?
40. What does the Virtual Processor (VP) really mean?
41. Performance - Is virtual SCSI or virtual Ethernet harder on the VIO Server?
42. Performance - What is best: balancing CPU across LPARs using Weight Factor or Entitlement?

Assumptions
This document assumes that you have already done the following and understood what happened:

- Read the IVM Redbook
- Installed the VIO Server with IVM on the machine
- Have access to the VIO Server by telnet or a Dumb ASCII terminal and RS232/serial cable
- Access via a web browser to the IVM administration website
- Created at least one VIO Server client logical partitions (LPAR)
- Installed an operating system (AIX or Linux) into the client LPAR

If not please consult the IVM Article, IVM Redbook in that order and the install and use IVM before continuing with this document.

Goals
This document is intended to take you to the next step and develop you into an Advanced IVM Administrator by covering many more advanced subjects in a Frequently Asked Questions (FAQ) format. Hopefully by break them into different topics, you can find what you need and when you need it, as well as read it from front to back for a more comprehensive understanding.

Note: Performance related topics are highlighted by starting with the work "Performance".

Integrated Virtualisation Manager (IVM) - A Brief Overview
IVM allows you to use the POWER5 Virtualisation features on pSeries p5 (505, 510, 520, 550) and OpenPower (710 and 720) without using a Hardware Management Console (HMC).

The main IVM points are:

- IVM makes a lot of sense on these lower end machines, on systems where full HMC functions are not required and where the HMC can seem expensive (for example, when managing just one machine).
- The many of HMC functions are actually performed by the IVM running on the first logical partition (LPAR) - which is currently a VIO Server version 1.2.
- This VIO Server partition owns all the physical adapters.
- All the other partitions use virtual disks, virtual Ethernet, virtual console and virtual CD/DVD.
- The IVM functions are accessed mainly by the IVM web interface in a browser communicating to the web server on the VIO Server partition but there are a few commands which you access by telnet to the VIO Server (ssh is an option).
- Only four simple commands are used for the initial IVM setup
- One further VIO Server command is used to start the virtual console to the other partitions to install the partition Operating System (AIX, Red Hat Linux or SUSE Linux).
- There are more commands but these are rarely used.
- Everything else is done from the good looking and simple to use IVM web interface.

For people without HMC experience, this IVM web interface is a much simpler and friendly way to manage Virtualisation functions but there are a few limitations that you should understand before deciding to use the IVM. For people with HMC and VIO Server experience using the IVM will be easy but bare in mind that the VIO Server "owns" all the adapters – most of your questions about limitations will be answered by this one fact.

Key places to get more information on Integrated Virtualisation Manager (IVM):

1. The IVM **Redbook** http://www.redbooks.ibm.com/abstracts/redp4061.html

2. An excellent IVM web based **article** on Installation IVM and setting up Linux LPARs http://www-128.ibm.com/developerworks/linux/library/l-pow-ivm/

3. If you have it installed then IVM then its website has **built in help** information

4. The **Info Center which is IBM's Web based information database** http://publib.boulder.ibm.com/infocenter/eserver/v1r3s/index.jsp?topic=/iphcn/iphcnkickoff.htm

The below two Redbooks mainly cover HMC based Virtualisation but a lot of the hardware details, material and concepts are relevant to IVM.

5. Advanced POWER Virtualization on IBM System p5 - Introduction and Basic Configuration http://www.redbooks.ibm.com/Redbooks.nsf/RedbookAbstracts/sg247940.html

6. Advanced POWER Virtualization on IBM System p5 - Architecture and Performance Considerations http://publib-b.boulder.ibm.com/abstracts/sg245768.html?Open

Can I set up the VIO Server without using the Flexible Service Processor (FSP) IP addresses setup and ASMI via a network?

Yes, you can. If you have ready access to dumb terminal (or equivalent) connected to the RS232 serial port of the POWER5 machine then there is no need to setup the network interface for the FSP as you can use this interface instead but there are some drawbacks.

Two classic problems are:

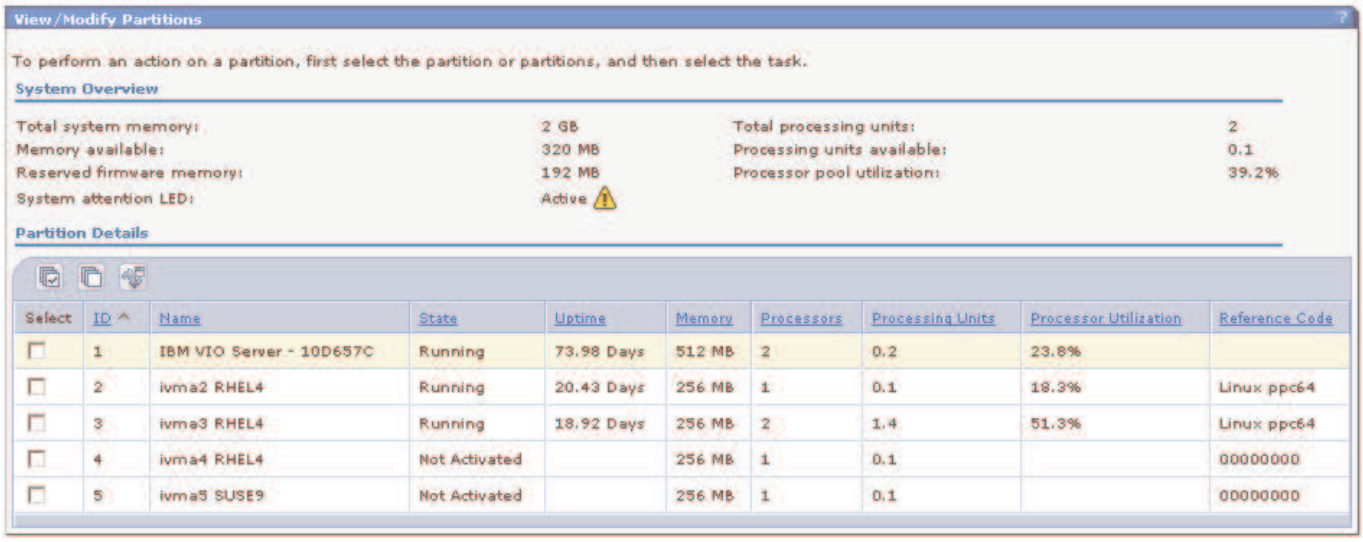
- you close down the VIO Server running IVM and forget the automatic reboot option or
- a power failure and the machine does not automatically restart.

In these cases you have to restart the machine using the FSP. You can use either the network and ASMI or you can use the serial interface. Unless you have set up some remote means of using the serial interface like a terminal server then this will require going to the machine to use the dumb terminal. The advantage of having the FSP on the network via ASMI is that you can do this remotely.

This is why the network access to the FSP is recommended and covered in the IVM Redbook.

Performance - Explain the performance numbers on the IVM administration web interface?

The IVM administrators View Partitions panel brings together a lot of useful system information into one place and a lot of this is performance related.



View/Modify Partitions

To perform an action on a partition, first select the partition or partitions, and then select the task.

System Overview

Total system memory:	2 GB	Total processing units:	2
Memory available:	320 MB	Processing units available:	0.1
Reserved firmware memory:	192 MB	Processor pool utilization:	39.2%
System attention LED:	Active		

Partition Details

Select	ID	Name	State	Uptime	Memory	Processors	Processing Units	Processor Utilization	Reference Code
	1	IBM VIO Server - 10D657C	Running	73.98 Days	512 MB	2	0.2	23.8%	
	2	ivma2 RHEL4	Running	20.43 Days	256 MB	1	0.1	18.3%	Linux ppc64
	3	ivma3 RHEL4	Running	18.92 Days	256 MB	2	1.4	51.3%	Linux ppc64
	4	ivma4 RHEL4	Not Activated		256 MB	1	0.1		00000000
	5	ivma5 SUSE9	Not Activated		256 MB	1	0.1		00000000

There are two sorts of numbers here

- Configuration Numbers are settings that are static (unless you change them):
  - Total System Memory – the memory in the machine
  - Memory Available – the memory unassigned to a partition or firmware
  - Memory column – the memory assigned to each partition
  - Total processing units - the number of CPUs in the machine and online,
  - Processing units available – unassigned CPU power. This can be used if partitions are uncapped, you add a further partition and change a partition to have more CPU
  - Processors column – the Virtual Processors (VP) for each partition
  - Processing Units column – the CPU assigned to each partition (the Entitlement)
- Performance Numbers – which are dynamic (i.e. depend on the workload in the partitions):
  - Processor pool utilisation - this is the percentage of the time spend running the partitions but excludes CPUs assigned to dedicated CPU partitions.
  - Processor Utilization column – this is the CPU time actually used in each partition but compared to its Processing units (i.e. Entitlement). This means if the partition is uncapped the utilization can be over 100%. The absolute maximum is 1000%.

Warning: these numbers are not automatically updated or refreshed in the web browser. You have to hit the "Refresh" button to get these numbers updated. It is the button above the Partition ID column.

The button looks like this:

A reminder:

- The Processors per partition column (this is the Virtual Processors (VP)) is the number of CPUs that this partition can spread across. In the example, the first and third partitions can use both CPUs and the others can use a maximum of one whole CPU.
- The Processing Units per partition column is often called the CPU Entitlement for a partition. This is the amount of CPU time the partition is guaranteed to have on the processor, no matter what the other partitions are doing or demanding. If the partition is "capped" then this Entitlement is the maximum CPU time it regardless of the workloads demand. If the partition is "uncapped" this Entitlement can be exceeded but only if thee is unused CPU time available.

The Performance Numbers are not quite what you might expect. When you hit the Refresh button they are updated on your screen but these are not instantaneous details (i.e. not the performance numbers from the last second or so) as it could be very misleading. Many workloads can be very peaky and each successive refresh could yield a seemingly random result and the workload fluctuates wildly. So:

a) The IVM **only queries the hardware every 30 seconds** – so refreshing say every 5 seconds will usually result in the numbers changing only each sixth time.
b) The performance numbers are only roughly **a five minute rolling average**.

The peaks and troughs in workload utilization are smoothed out by the averaging.

How do the numbers in the above screen shot relate to each other?

- Processor pool utilisation = 39.2%
  - This means that in the last five minutes, the two CPUs have on average been 39.2% busy.
- Processor Utilisation
  - Partition 1 is the VIO Server and has used 23.8% of its 0.2 Entitled CPU time.
  - Partition 2 running Linux has used 18.3% of its 0.1 CPU time.
  - Partition 3 running Linux has used 51.3% of its 1.4 CPU time.
  - The other partitions are currently off so they are not using any CPU time

So how do these numbers add up?
First remember these are some other factors to take in to account like rolling these averages inaccuracies, there is Hypervisor CPU time and processor Idle and Waiting for I/O time, so they are not going to be 100% precise.

- Partition 1 → 23.8% of 0.2 CPU time = 4.8% of a CPU
- Partition 2 → 18.3% of 0.1 CPU time = 1.8% of a CPU
- Partition 3 → 51.3% of 1.4 CPU time = 71.8% of a CPU
- The total comes to 78.4% of a CPU.

As we have 2 CPU this is 78.4 / 2 percent of both CPUs = 39.205% CPU used. This 39.2% compares to the Processor Pool Utilisation of 39.2.

Apparent Processor Utilization time lag issues?
The roughly 5 minute averaging of the Processor Utilization numbers can cause some interesting side effects.

- If you have a have an idle partition with low CPU use and suddenly start a CPU bound process or workload ("yes >/dev/null" is very effective at this) and quickly update the IVM web interface and View Partitions panel, you may find the Processor Utilization number has not changed. This is because the high CPU use is only a small fraction of the last five minutes. If you wait, say a further 10 seconds, and update the panel again then you should see the Processor Utilization starting rise but it can take a few minutes for the numbers to reach their maximum. This highlights the need for patience.
- The same happens when a workload stops, again the Processor Utilization numbers lag behind what you know has happened..
- If you add a CPU workload and before the Processor Utilization reaches the maximum, you stop the workload, you can even find the numbers still rise for a little while due to this lag.

This means you have to be patient with the utilisation numbers, when you have made changes that will drastically affect CPU use but on the good side the numbers will not wildly fluctuate.

Why aren't these Processor Utilization numbers automatically updated?
It was decided by the IVM designers that, as the web server is running on the POWER5 machine, automatic updating would be a waste of important CPU cycles that would best be used to run server workloads. Each person viewing the IVM web interface would cause this waste.

If I am mirroring disks on the VIO Server, how can I ensure they are on different disks?
This assumes you have more than disk. It is important to ensure the data is mirrored across disks to ensure data redundancy and security. Mirroring the data via the AIX LVM on the VIO Server is not allowed or a supported feature. The mirroring has to be performed on the VIO Client partition side. But it is important that virtual disks are physically on different disks or a single disk failure would cause the loss of all data.
There are two approaches:

1. If you have enough disks create two storage pools containing different disks. Then you can assign the virtual disks from the two different pools – this is a simple and clean way to ensure the mirrors are configured safely. This can all be done using the IVM Web interface.
2. The alternative manually creating the logical volumes on different disks using VIO Server command but using this approach you must be much more careful. Then you have to map the logical volumes to virtual disks. Then you have to assign them to the partition and unsure they are on different physical disks. This requires the use of many commands at the VIO Server, is more complex and so error prone.

As protecting your data is vital, please double check you know where the data is and it is separated. See the IVM Redbook for more information.

What sort of disks are recommended with IVM for data protection?
It is recommended that you get the hardware to do the hard work and not the people. So the author recommended in the following order:

1. SAN disk sub-system – Why? Fast inter-connect, data protection and caching for higher performance.
2. SCSI RAID5 adapter and disks – RAID5 offers good data protection at low cost and caching without operating system overheads for good performance.
3. Storage Pool based VIO Server mirroring implemented using the IBM web interface – as it is simple.
   - AIX on the client side is easy to configure for mirroring but Linux mirroring is much more challenging and difficult to test.
4. Storage Pool based VIO Server mirroring using manual assignment based VIO Server mirroring using the VIO Server command line
5. NOT supported - Mirroring using the VIO Server using the AIX LVM – it is easy to accidentally end up with the if you use the "mirrorios" command and there is a virtual disk in rootvg – so be careful

Should I use rootvg for VIO client virtual disks?
The manual and IVM Redbook suggests it is a good idea to avoid the rootvg Volume Group and matching Storage group for any virtual disks. This is a good for a number of reasons.

1. VIO Server performance – the VIO Server performance is important for the performance of the whole IVM machine and not overloading rootvg with data will help achieve this goal.
2. Backup of the VIO Server will be quicker – and recovered more quickly in the case of a disk loss.
3. KISS – Keep It Simple Stupid – mistakes are less likely with VIO Server and virtual disks on separate disks
4. Mirroring concerns – you can mirror the rootvg logical volumes with the "mirrorios" command but this should not be used to mirror VIO client virtual disks as it is not supported.

The counter arguments are:

1. If you have one disk then you don't have any option at all.
2. If you have two and decide to mirror rootvg or VIO client virtual disks you don't have an option as both disks will have to be in the rootvg.
3. The rootvg disk is 73GB and the VIO Server only takes up 25% of the disk, so not using this disk space it is a large wasted resource and the VIO Server itself does not use the rootvg very much once it is started up.
4. The rootvg is actually a SCSI RAID 5 pack or a large SAN sub-system disk – see the above point.

You will have to make up your own mind based on your priorities.

How do I stop/halt a partition that is completely stuck?
If you have a partition that is running in any of the below modes

- at the firmware level prompt
- in the SMS menu but you can't start the partition
- stuck in boot or installation code with a problem
- just stuck

then it is not yet running an operating system. In this case you cannot stop the partitions by login on to it and using the shutdown command. Nor can you stop it by using:

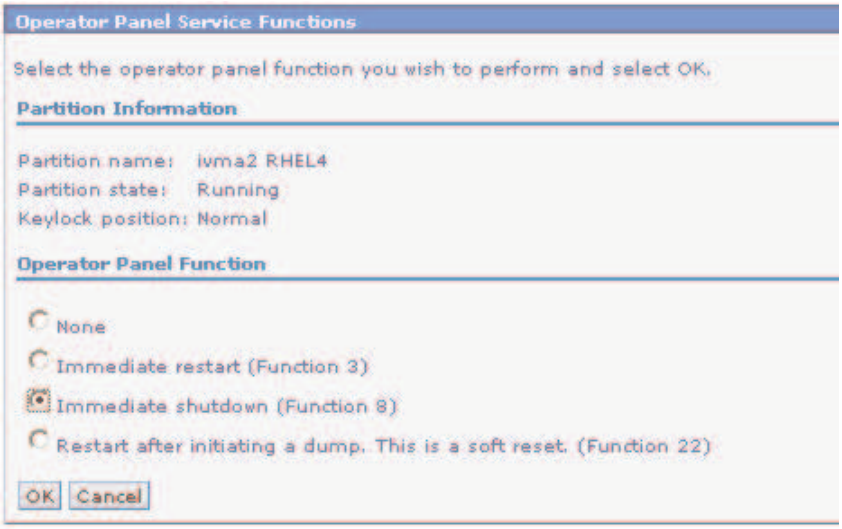
View Partitions → Select a partition → Shutdown

Both of these options assume that there is an operating system that will shutdown. If you try the IVM web interface shutdown, you will find that it will attempt to communicate with the operating system for something like 45 to 50 minutes before halting the partition in an ugly way.

The answer is to use:

View Partitions → Select a partition → Operator panel service functions

Then on the below panel:



Select the Immediate Shutdown option and OK. This will halt the partition in just a second or two. Note: this halts the processor within the partition, so it is **not** recommended for a regular system shutdown as it can cause corruption of filesystems and disk data.

What appears in the "Reference code" on View Partitions panel when there is no problem?
The Reference Codes are what would appear on the LED panel of the machine. When the partition is running it will display one of the following:

- Blank – this means it is running AIX
- "SuSE Linux ppc64" – this means is it SUSE Linux
    - Note: SuSE became SUSE when Novell bought them
- "Linux ppc64" – Red Hat Linux

Processor Utilization numbers are missing on the View Partitions Panel?
This is a common problem on many systems and frustrating when it occurs as there is no obvious reason or solution. The problem is caused by the partitions and the VIO Server not having matching:

- time,
- date and
- zone.

If you set the time, data, zone on each partition correctly on the partitions without blank Processor Utilization and wait a couple of minutes, the data should appear. What commands to use?

- On the VIO Server use the chdate command for all three details.
- On AIX partitions use the "date" and chtz" commands.
- On Linux partitions use the "date" command and a link file in the /etc/localtime directory.

How do I get out of the mkvt command?
The mkvt command is used to access a virtual console from the VIO Server to the VIO client partitions. You have to telnet to the VIO Server and then used, for example to access partition 2, the command: mkvt –id 2
But now you want to exit the console and get back to the VIO Server command line.
To do this, use the two characters: ~.
Spelt out this is a tilde and a full stop.

Alternatively, running the “rmvt” command is a different console will stop this one

User Management – what is the point?

The User Management options on the IVM web interface allows the creation, viewing and modifying user accounts but, what is the point for having more user accounts as we can do every thing using “padmin”? The following are useful options:

- For increased password security - you can set advanced password features like
  - limiting login retries,
  - password aging and
  - password choice restrictions.

- It is recommended that a **read-only user account** is created. This allows you to let other people to take a look at the partition details and machine setup but they cannot alter it in any way. This allows investigation by less skilled staff with no risks or you can leave a session open on your screen and not have to manually disconnect before moving away from your workstation/PC.

- Note that only the “padmin” user can use the “oem_setup_env” command to get to the root super-user account. Other user accounts with View/Modify privilege can perform all the other commands except “oem_setup_env”.

- It is recommended that a user similar to “padmin” (i.e. with View/Modify privileges) is created and this user is used for day to day IVM and Virtual I/O Server administration. This non “padmin” account should be used most of the time by administrators to avoid the temptations and risks of the root user. This is the same as regular UNIX machines and avoiding logging in as root. This user account can also be used in emergency circumstances i.e. where there is a problem with the logging in as “padmin” or with the password being changed, aged, damaged or forgotten.

- Also “padmin” is a known user account – this makes a good hacker target.

How do I remove a storage pool?

Most tasks can be performed from using the IVM web interface but others have to be performed on the VIO Server at the command line interface. This is how you can remove a storage pool:

1. First make sure you have no virtual disks on the disks in this storage pool. This can be done on the IVM web interface using Storage Management → View/Modify → Virtual Disks tab.
2. Second, reduce the disks in the storage pool to just one disk. This can be done on the IVM web interface using Storage Management → Advanced View → Storage Pool tab → Select a storage pool → Reduce → select a disk → OK.
3. Third, if this storage pool is the default storage pool then this needs to be changed. This can be done on the IVM web interface using Storage Management → Advanced View → Storage Pool tab → select an alternated storage pool → Assign as default storage pool → OK.

   Or 3 can be performance from the VIO Server command line. For example, change the default storage pool to the rootvg pool which always exists: chsp –default rootvg

4. Fourth, remove the storage pool. Using the VIO Server command line, for example, if the storage pool to be removed is called pool1 and the last disk is hdisk1: chsp –rm –f –sp pool1 hdisk1

During the review of this document I have been informed that you can indeed remove a Storage Pool entirely from the IVM web interface as follows (but I have not tried this personally) by select all the disks in the Storage Pool in the Reduce option.

Why can't I delete a partition and get the error message as below?
While trying to remove a partition using the IVM web interface (in this example partition 6) the follow message appeared:

> **Problems occurred while processing the data. A summary of all problems for this page are listed below. Additional details for each problem may be located next to the field causing the problem.**
> **Failed to delete partition 6.**
> **Command issued: rmsyscfg -r lpar --id 6**
> **Command response: [VIOSE03000000-0148] Unhandled firmware extended error - Category: 0x3, Code: 0x40000147**

This means that you have an open virtual console to the partition and you need to close or remove the virtual console session before removing the partition.
Either:

1. If you can find the VIO Server telnet session or physical console session that is running the mkvt command and use ~. to stop the console

or

2. You need to telnet to the VIO Server involved with this partition and remove the console. Force the console closed but issuing the "rmvt" command. For example, for partition 6 (the 6 is listed in the first column of the View Partitions on the IVM web interface) on the VIO Server use:

rmvt –id 6

Then you can remove the partition.

Good News: I am told that a future IVM version will automatically remove the open console, so this problem will go away once everyone have updated to that version.

Will the CD be the default boot device?
The answer is yes and no.

- **Yes** - If this is the first time the Partition ID has been used and the partition includes a CD drive. This happens when it is new IVM setup and you add a further partition. In this case a new boot list is created, when you create the partition and the CD or DVD will be the default boot device (if present in the partition). This can be very useful. For example, you can leave a Linux CD in the CD drive – then with each new LPAR, you can simply boot off this CD without having to use SMS and then proceed to do a network install of Linux. The result is you don't have to change CD's (or even go near the machine).
- **No** – If you have previously used the Partition ID then the boot list will still include the previously used devices. This may be the case where you have removed a partition or you are reinstalling it. This old boot list can have been set in a number of ways:
    - You may have manually set the boot list with the SMS menus.
    - The operating system has set the boot list as part of the installation process.
    - You have previously used operating system level commands to set the boot list.

Shutting down – what is the best way?
This is what I would recommend.

1 Halt all partition that have hung or are in SMS – see the question:
How do I stop/halt a partition that is completely stuck?
2 Go to each client partition and using the command line shut it down. Get to the command line use telnet for AIX or ssh for Linux or via the VIO Server and the remote console session by using the mkvt –id <ID> command. When at the command line, shut the partition down with, either:
   a. For AIX use: shutdown –Fh
   b. For Linux use: shutdown –fh now
2. The alternative is with AIX partitions only using the IVM web interface
View Partitions → select the AIX partitions → Shutdown (the command at the bottom)→ OK
3. Then shutdown the VIO Server by the command line:
   a. shutdown
   b. shutdown –restart

Warning: it is a classic mistake to shutdown the VIO Server remotely without thinking how to restart it. If you are a previous HMC user then you would expect to use the HMC to restart the machine but, of course, the IVM controlled machine does not have a HMC. If you did not use the "–restart" option then the machine will shutdown and stay down. You typically have to use the dumb terminal on the serial port to reboot or if you have the Service Processor (FSP) setup to use the network ASMI interface then it can be used to restart the machine.

What is the difference between View/Modify Devices and Advanced View/Modify Devices?
There are a few differences, some are very subtle (or even look like a mistake) but it is confusing.

- The large difference is the Advanced View/Modify Devices has the Storage Pool tab and options. This allows a storage pool to have more or less disks, choose the Default Storage Pool and viewing the storage pool properties.
- Subtle difference: that View/Modify Devices has the Physical Volumes tab that has a command to add to/delete from the Default Storage Pool. But and Advanced View/Modify Devices has the Physical Volumes tab but the command is **not** restricted to the Default Storage Pool.
- Trivial Difference (mistake): that View/Modify Devices has Virtual Disks but Advanced View/Modify Devices has Logical Volumes but they refer to the same thing and have the same options.

If you give a whole physical disk to a partition, is it a physical disk or a virtual physical disk?
The VIO Server partition owns all the hardware adapters in the machine. This means that all disks are connected to the VIO Server partition. When you assign a whole disk to a client partition, the client partition still has to use the virtual device drivers to the VIO Server to access the disks. So this is virtual access to a physical disk.

Can I remove a disk from a storage pool?
Yes you can from the IVM web interface but you need to ensure the disk does not have any virtual disks on it. To do this you need to use the command line on the VIO Server and the "migrate" to move the partitions or "reducevg" with the "-rmlv" option (if you not longer need the logical volume). These commands are familiar to AIX administrator and you can find out detail sin the manuals.

Is moving the optical device (CD or DVD) safe?

If you move the optical device between two partitions you will get a warning message like this:

> "*One of more of the selected devices is currently assigned to a running partition. It is recommended that you power off the partition prior to removing assigned devices. If you would like to proceed anyway, select the checkbox below, and select OK again.*"

To be very safe you can power off the partition but this is really unnecessary.
Ask yourself "What is the worst that can happen?"

- If the partition that currently has the device is not using it, then it will not notice the device has gone missing. If it later tries to access the device it will report errors to the user and log files but no harm is done. If you return the optical device again to the partitions before the partition attempts to use it, then it will work fine and no errors are reported.
- If the partition is currently using the device then it is much more harmful, as the next device access will report errors and the application will have to handle the problem – usually they will halt with an error message.
- If you are using AIX and want to be safe, then in the partition with the device, remove the device from AIX with the "rmdev" command before removing it from the partition. This ensures AIX can not even attempt to use the device but you have to remember to run "cfgmgr" again to find the device once is reconnected.
- On Linux, I would not bother removing the device.

There are three typical uses of optical devices:

Installation

Removing the devices during installation is likely to be fatal for the partition but presumably you as administrator should be aware of any installations taking place on your machine. So this is unlikely to happen.

Installing extra software

In this case you will have the device mounted as a filesystem and be reading the device as part of the loading data or applications. Again you should be aware of installations taking place but it would be a good idea before removing the device to go to the owning partitions and check the device is not mounted with the "df" command. If it is mounted, work out why, unmount (umount on Linux) the device and you are ready to go.

Backing up

If this is a writable device then it can be used to back up data or the entire system (if DVD). Again you should, as administrator, be aware of backup schedules and other such activity, but you could check the device is not mounted or the write LED on the device is not flashing.

One final trick: If you are near the machine, eject the CD/DVD drive tray with the button. Once it is ejected you are very sure the partition can not be using the drive or have it mounted.☺

Why won't Linux shutdown when using the IVM web interface "shutdown" option?
With a Linux partition you might try the following on the IVM web interface:

View Partitions → Select a partition → Shutdown

This will attempt to communicate with the operating system for something like 45 to 50 minutes. It will then halt the partition the hard way and produce an error message (on the IVM web interface View Partitions), as follows in the partition State field "Error Attention" and in the Reference Codes field "B200F005 LP=002" as in the below example:

This partition is now in the Error State and can't be restarted. To get out of this error state, on the IVM web interface: View Partitions → select the partition in error state → Operator panel service functions

→ Immediate shutdown → OK

How to avoid this long Linux shut down wait and the subsequent error?
What you are missing is the IBM supplied RPMs for the POWER5 advanced functions. If you install these then the Linux partitions will close down cleanly as the VIO Server partition will communicate with the Linux partition to inform it that a shutdown is necessary. You will then get a message on all terminal sessions to the partition (via mkvt or ssh) like this one below:

Broadcast Message from root@ivmc3
/etc/rc.powerfail: A normal shutdown with no delay has been requested.
A system shutdown is being initiated.
The system is going down for system halt NOW!

This shows the IVM is informing the Linux partition via a "power failure" signal that a shutdown is required and Linux then warns the users. You can download the IBM RPMs for Linux on POWER from:

http://techsupport.services.ibm.com/server/lopdiags

Make sure you install the right versions of the RPMs (different ones for Red Hat and SUSE and for HMC or stand alone) and be careful to fully obey the pre-requisites before installing these RPMs. While reviewing this document I was told that the above website will be modified to make it clear which RPMs are required for IVM operation rather than the HMC.

Why telnet to the VIO Server partition, when ssh is much more secure?
Telnet is available as standard but that is known as one of the most insecure protocol and is hackers dream. Secure Shell (ssh) would be much better for higher security for day to day management. Fortunately, ssh is available but not installed as standard on the VIO Server due to open source licensing limitations. OpenSSL and Portable OpenSSH allow you to connect securely to the Virtual I/O Server from a remote location. These must be downloaded, installed and configured properly to work securely.
For more information on OpenSSL and Portable OpenSSH go to:

- http://www.openssl.org/
- http://openssh.org/portable.html.

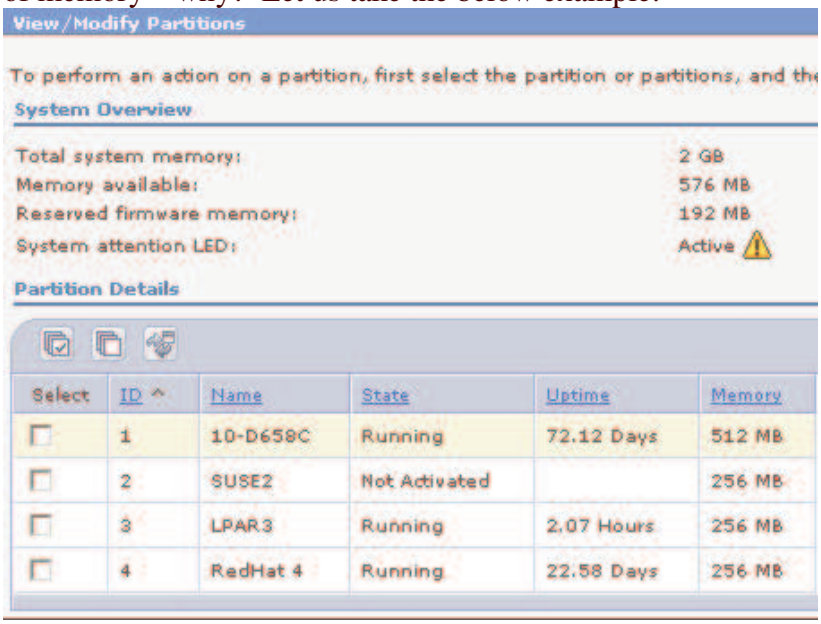
How do I get the System Attention amber LED switched off?
Two methods:

- Assuming you have the Flexible Service Processor (FSP) is setup on the network, then you can use the FSP ASMI web pages to switch this System Attention indicator off. Look under the System Information menu.

- This method requires the machines to be shutdown. Via a dumb terminal (or equivalent) and RS232 connection to the FSP menus and follow the menus to switch it off – unfortunately this requires all the partitions and the VIO Server to be stopped.

Removing the Attention LED setting for a partition?
This annoying yellow triangle gets switched on as the result of an error condition and can be removed as follows: View Partitions → select the partition → Properties → change Attention LED to Inactive → OK

Where does all the memory go?
For example, I have 2 GB of memory in the machine and I can only start three LPAR with 512 GB of memory – why? Let us take the below example:



The machine has 2 GB (2048MB) of memory, where is this being used?

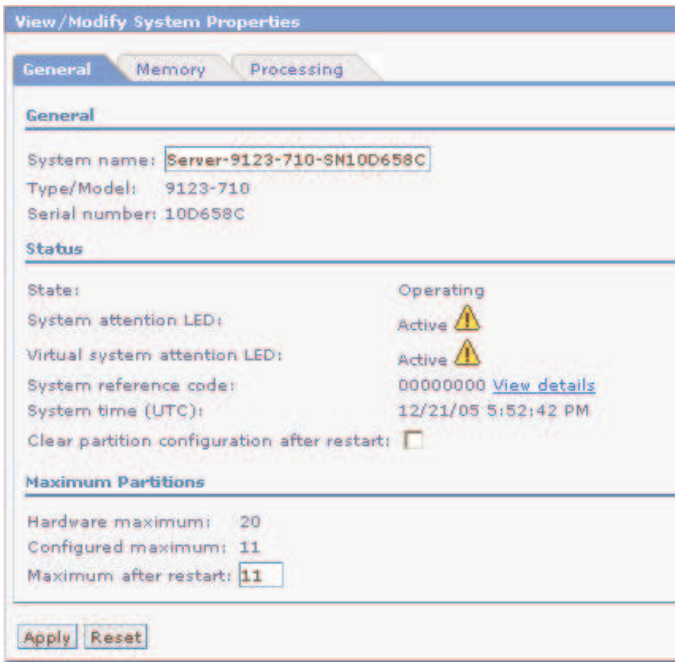
- 512 MB - The VIO Server (partition 1) has 512MB - which is the default on a 2 GB machines
- 768 MB - This is being used by the three partitions - 256 MB each.
- 576 MB – Is the "Memory available" - i.e. not allocated at the moment
- 192 MB – This Reserved firmware memory – being used by the Hypervisor. This is used for a number of purposes:
  - Hypervisor control tables.
  - The Hypervisor functions
  - The Hypervisor manages the virtual memory page tables. This is roughly 1/64th on memory, so on this 2 GB machine this comes to 32 MB.
  - Special Translation Control Entry (TCE) area for device driver control
  - Running virtual disks and virtual network buffers

This adds up to the full 2048 MB of memory.

Can I control the "Reserved firmware memory"?

Yes, you can affect this memory use. There are four uses of this memory and some can be controlled:

1. The Hypervisor needs some memory – you cannot control this
2. Each Virtual Ethernet you use. When you set this in the LPAR Virtual Ethernet details for the zero, 1 or 2 virtual Ethernets then memory is allocated. You cannot really control this as you are unlikely to needlessly set up unneeded virtual Ethernets.
3. By the Maximum Memory setting of each logical partition. When you set a Maximum memory for a logical partition that is greater than the Assign memory then the Hypervisor has to reserve enough memory to accommodate the virtual memory page tables (even if they are not currently used) in case the memory assigned is enlarged later on. By default the Assigned and Maximum memory is the same. If they are different then some memory can be wasted. For an extreme example, on a 64 GB of memory machine, setting the Assigned to 256 MB and the Maximum to 64 GB would result in roughly 1 GB of memory being wasted.
4. Setting the maximum number of logical partitions to a lower value. This has a small impact as explained below: The each partition that is created takes up memory in this area – in addition to the memory used by the partition. On the above diagram system example, the View System Properties looks like the below:



If you know that you do not want this "Configured maximum" number of partitions (in the above picture this is set to 11) then you can reduce the number of partitions and free up some memory. Put the new value in the "Maximum after restart" field, and then click on Apply. On the next reboot the systems (after shutting down the partitions cleanly, of course), you will have freed up or use extra memory. Well that is the theory – I have changed the number of partitions from and it made little difference. At 20 partitions the "Reserved firmware memory" when up to 224MB – from 192 for 11 partitions. Going down to 4 partitions had no effect. No real rule of thumb can be extracted from this but reducing the maximum partitions on machines with larger numbers of CPU might save a few MB of memory. This may be different on large memory configurations.

What should I do about the little bit of free memory left over?
For example, I have 384 MB of free memory and was planning to use 256 MB – what should I do with the 128MB left over?
AIX 5.3, Red Hat EL 4 Linux and SUSE SLES 9 Linux all recommend 256 MB of memory as a minimum and some will refuse to install with less than this. This means 128 MB of memory (or less) is not enough to run an extra partition. You have the following options:

- Give the extra memory to the last partition to be created as a little extra never hurts.
- Decide which partition is more important and assign the memory to it. Note: you will have to stop the partition to give it the extra memory and Dynamic partition changes are not possible.
- Give it to the VIO Server partition – where dynamic changes like adding memory is possible.

You will have to decide your own priorities.

Is the memory release when I stop the partition?
No. Once you create a partition, the memory is assigned to it regardless of whether the partition is running or not. This is to keep things simple. As a contrast, the HMC allows you to create many logical partition profiles with overlapping and over committing resources by which I mean, if you add up all the memory in all the profiles this can be a lot more then the physical memory. You can't, of course, actually start all the HMC based partitions at the same time – memory is allocated on a first started basis until it is all used. With IVM it was decided this is unnecessarily complicated, confusing and error prone.

If you have a partition that is currently stopped and you want to temporarily use its memory in another partition you have to remove the memory from the stopped partition first using
View partitions → select the stopped partition → Properties → Memory tab → change the minimum, assigned and maximum memory to zero → OK
Then the freed up memory can be assigned to a new partition or to an existing partition.

Is the CPU released when I stop the LPAR?
Yes and No.

- Yes: if you are using the shared processing pool, these proc resources can be utilized by other LPARs, assuming they are not 'capped'. The proc units assignment merely indicates the amount of processing power to which the LPAR is entitled - if it uses less than this amount (which is 0 when powered off), the remaining resources may be used elsewhere.
- No: in terms of the processing power being assigned to create another LPAR. See the answer to the memory question above.

To reuse the CPU assigned to a stopped partition, set all of the below properties to zero for the stopped partition using
View partitions → select the stopped partition → Properties → Processing tab → set the below to zero

1. Processing Units minimum
2. Processing Units assigned
3. Processing Units maximum
4. Virtual processors minimum
5. Virtual processors assigned
6. Virtual processors maximum

Then the freed up CPU time can be assigned to a new partition or to an existing partition.

Performance - How can I ensure the VIO Server is not staved of CPU?
If you don't have performance information available in advance for detailed VIO Server planning and sizing then you have to make a first guess and monitor the machine for further tuning. Here is a suggestion to get you started:

10% of available CPUs

- 10% is about right from reading the recommendations in both Virtualisation Redbooks.

High priority = 255

- So the VIO Server gets first call on any spare CPU cycles if it gets very busy. This will avoid the VIO Server slowing down other partitions.

Uncapped

- The VIO Server will not be a run away partition as it only works on demand driven by other partitions. Capping the VIO Server can be harmful.

Virtual Processors = the greater of 2 CPUs or one more than necessary for the 10% entitlement

- This will give the VIO Server head room to use extra CPUs if demand grows or peaks. As the VIO Server is AIX 5.3, it will "fold" spare Virtual Processors and so not waste resources if the VP number is too high.

In actual fact, partitions and the VIO Server will naturally balance out (provided the VIO Server is uncapped and has a VP head room). To explain his let's take and example. The client partitions are demanding high volumes of disk and network activity to be driven through the VIO Server. Eventually, each of the partitions will have to wait for further disk I/O or further network I/O for the now slow VIO Server to complete its tasks and so they will start blocking and waiting for I/O to complete – at this time the only partitions running is the over worked VIO Server, so it can use all the CPU power in needs. This is a natural load balancing provided you have not capped or set the Virtual Processor number too low for the VIO Server.

Performance – How much memory is right for the VIO Server?
On my smaller machines with 2 GB of memory the default is 512MB for the VIO Server. More for the VIO Server would be ideal but there is not a lot of memory in the first place.
For machines with 4 GB or more memory, I suggest 1 GB of memory for the VIO Server. If you have lots more memory and more then 2 fast Ethernet connections then an extra GB of memory would be a good idea. For example, above 16GB of memory and with 4 fast GBit Ethernet adapters (perhaps running Ether-Channel) then 2 GB of memory for the VIO Server.

What is the point of the Processing Units Minimum and Maximum do?

VIO Server

For the VIO Server this sets limits on the Dynamic CPU changes that can be performed while the machine is running. For example,

- Processing Units Minimum = .05
- Processing Units Assigned = 1
- Processing Units maximum = 2

When the VIO Server starts it will be assigned 1 CPU of Entitled CPU power. Later the systems administrator can make Dynamic changes to the entitlement down to 0.2 CPU or up to 2 CPU but not outside these ranges. However, the Minimum and Maximum can be changed, if it is found that the Minimum and Maximum are unsatisfactory!

Client Partitions

For client partitions these have no meaning what-so-ever. They can be changed but requires the partition to be stopped and restarted. A future release of IVM, may allow dynamic changes to client partitions and then these would be used but this is a wild guess and not an announcement. There is also the possibility that a Workload Management application (in the future) could use these to finely manage system resources.

What is the point of the Memory Minimum and Maximum do?

See the answer to the above question but for Processing Units.

What is the point of the Virtual Processor Minimum and Maximum do?

See the answer to the above question but for Processing Units.

Performance - To cap or not to cap, that is the question?

There are three answers to this question. You will have to decide which you like:

Old Style

Many systems administrators like to be in control of their machines. They will perform detailed analyse of the workloads and decide the size of the partition in CPU terms (processing Units). They will then cap the partitions and actively monitor the partitions CPU use and make adjustments to the partitions based on experience and performance priorities.

New Style

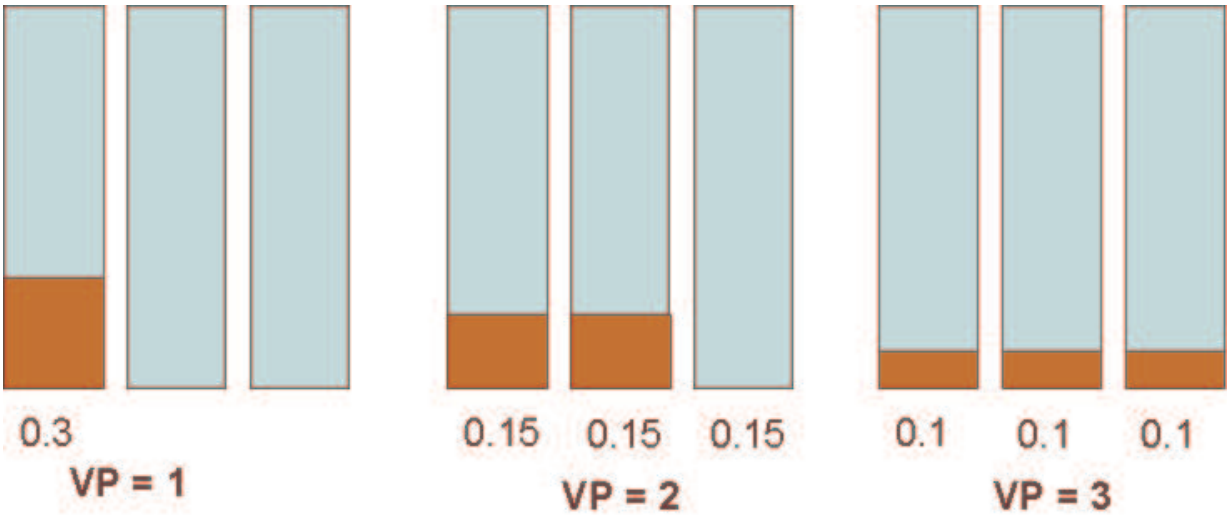
This type of administrator decides to make the machine do the hard work. They too set the Processing Units assigned to their best guess and decide to go uncapped. The Hypervisor will then assign free CPU time to which ever partitions need the extra CPU every 10 milliseconds. This means the partitions do not need to be closely monitored nor adjusted further.

Charge Back Environments

In the charge back ennvironemnt companies like to charge their user departments for computer resources used. The idea of letting the users get unused resources for free is not tolerated. In this environment, the user department has to agree to pay more cash before the CPU resources are adjusted.

Unfortunately, with the current IVM adjustments to the Processing Units for partitions (other than the VIO Server) requires stopping and restarting the partitions which is not usually possible with production workloads. This means New Style is recommended.

What does the Virtual Processor (VP) really mean?
This term lead to a lot of confusion especially as we have physical processors and logical processors already in POWER5 based machines. The Virtual Processors (VP) is the number of CPUs that this partition can spread across. For example, a partition has 0.3 Processing Units (also known as its Entitlement while it is running) this could be assigned to the CPUs in different ways:



There are three ways:

- With a VP of 1 all of the 0.3 CPU power is run on one CPU.
- With a VP of 2 the CPU power is split between 2 CPUs.
- With a VP of 3 the CPU power is split between 2 CPUs.

Lower VP numbers are slightly more efficient as lower numbers of CPU results in higher memory cache hits and higher performance. Higher VP numbers means that the workload can peak to higher numbers of CPU involved in working through the peak. So it is a matter of balance. Each VP has to have a minimum of 0.1 CPU. so in our example of 0.3 CPU power, a VP of 4 is not allowed.

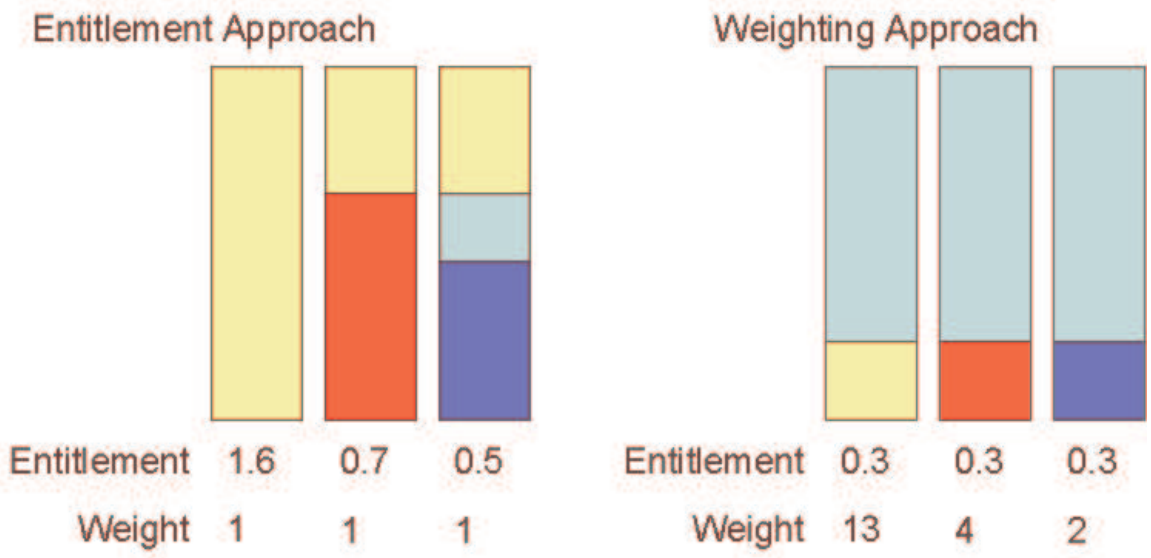
Many people prefer the term “spreading factor” rather than Virtual Processor.

Performance - Is virtual SCSI or virtual Ethernet harder on the VIO Server?
From the Virtualisation Redbooks is can be seen that the virtual Ethernet is much more CPU intensive then virtual disks. Virtual disks on the VIO Server is a simple device driver and it changes virtual memory so that the adapter does the I/o directly in to the client partitions memory so the disks blocks are not copied around the systems i.e. this is off loading the work to the adapter. Virtual Ethernet on the other hand has to use the main POWER5 CPUs to run the TCPIP stack. Although, this is efficient is requires more CPU cycles that disks.

So for sizing partitions it is more important to get the network I/O correct.

Performance - What is best: balancing CPU across LPARs using Weight Factor or Entitlement?

There are two approaches to assigning CPU power to partitions as shown in the below diagram:



Entitlement Approach

Here the importance of the partition is used to decide the Processing Units i.e. the Entitlement while it is running. This leaves a small amount of free CPU and the weighting factor will decide in the case where all partitions are very busy which partitions gets the extra CPU time. In many systems all of the CPU time is assigned. Depending on the workloads it might be rare for them all to peak at the same time – this means one of the partitions will not use all its entitlement thus makes some free CPU time for the other partitions.

Weighting Approach

In this approach every partition is given the minimum Entitlement to allow the partitions to spread across the available CPUs. In this case an Entitlement of 0.3 = three times the minimum per CPU of 0.1 and a VP of 3. Then the weighting factors are set to the relative importance of the workloads for the rest of the CPU time (i.e. allowing for the 0.3 Entitlement). Most of the time each partition will be over its Entitlement and the large free CPU time will be allocated according to the weight factor.

So which approach is best? Neither as they will result in the same or very similar behaviour.

Most people seem to naturally think along the lines of the Entitlement approach. If you need to add a further partition then the Weighting factor approach has one advantage. For example, you need a further partition and decide to give it half a CPU:

- With the Entitlement Approach you will have to free up 0.5 CPU by changing each on the existing partitions, only then can you start the partition. With the IVM this would mean restarting every partition.
- With the Weighting Approach you can immediately create and start the partition with 0.3 CPU Entitlement and a weighting factor of 5.

--- The End ---