



# PowerVM

## Session 4: Active Memory Sharing



**Nigel Griffiths**  
IBM Power Systems  
Advanced Technology Support  
EMEA

TLA overload!



© 2011 IBM  
2

Today  
AMS = Active Memory Sharing  
– PowerVM feature for Power6 & Power7

NOT

- AME = Active Memory Expansion – AIX only feature
- AEM = Active Energy Manager – part of Systems Director

## Why Active Memory Sharing (AMS) ?

Don't Set memory to an Virtual Machine (LPAR) and forget?

We monitor CPU use

- Shared CPU = good for moving cycles where they are needed
- Now the same for memory
- AMS moves Ram to where it does most good

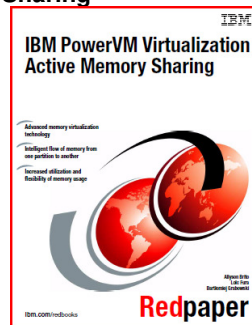
Like:

1. VM in different time zone
2. Day time users and night time batch
3. Many small LPAR but only a few really active

## Active Memory Sharing Reference

**Redbooks** <http://www.redbooks.ibm.com/portals/power>

**PowerVM Active Memory Sharing  
New Update June 2011 →**



Movies <http://tinyurl.com/AIXMovies> - 4 movies

Active Memory Sharing Regular Paging	19 mins
Active Memory Sharing Concepts	16 mins
Active Memory Sharing Setup	16 mins
Active Memory Sharing Simple Monitoring	11 mins
	~60 mins

## PowerVM Editions are tailored to client needs

PowerVM Editions offer a unified virtualization solution for all Power workloads

**PowerVM Express Edition**  
*Evaluations, pilots, PoCs  
Single-server projects*

**PowerVM Standard Edition**  
*Production deployments  
Server consolidation*

**PowerVM Enterprise Edition**  
*Multi-server deployments  
Advanced Functions*

PowerVM Editions	Express	Standard	Enterprise
Concurrent VMs	VIOS + 2 per VMs	10 per core (up to 1000)	10 per core (up to 1000)
Virtualization Management	IVM	IVM, HMC	IVM, HMC
Virtual I/O Server	✓	✓✓	✓✓
PowerVM Lx86	✓	✓	✓
Suspend/Resume		✓	✓
Shared Processor Pools		✓	✓
Shared Storage Pools		✓	✓
Thin Provisioning		✓	✓
Live Partition Mobility			✓
Active Memory Sharing			✓



\*IBM i supports shared storage but does not yet support Suspend & Resume nor LPM.



## Active Memory Sharing Media & Installing

There is none!



Note:

- First released in May 2009 on Power6
- At that time required new FW, VIOS & OS
- Now all current

## Active Memory Sharing Media & Installing



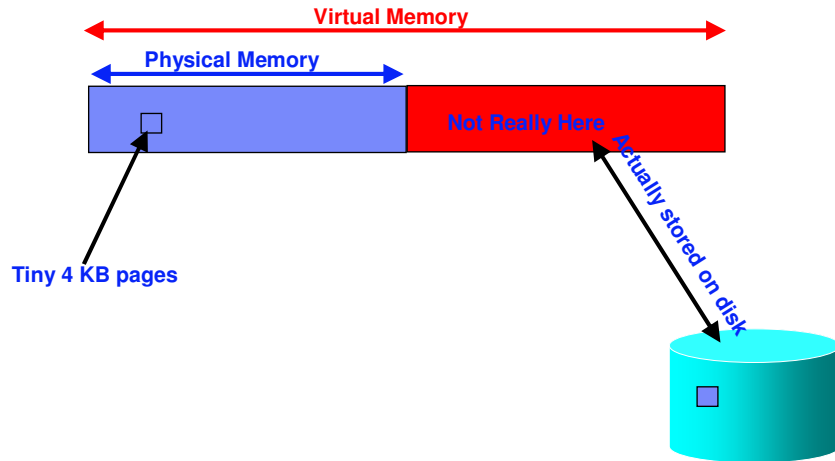
You need:

1. PowerVM AMS Activated → see HMC
2. VIOS → recent like 2.1.1 or 2.2
3. Operating system that supports AMS like current
  - AIX 6.1 TL3+ (not AIX 5.3)
  - IBM i 6.1.1+ Fixes
  - Linux – current
4. Uses 4 KB pages
5. Virtual Machines
  - Pure Virtual (shared CPU, network & disks)
  - Assumes differing workload pattern but friendly
  - Co-operate in sharing RAM across the system

## AMS pages memory between Virtual Machines

So let's talk about  
Paging ...

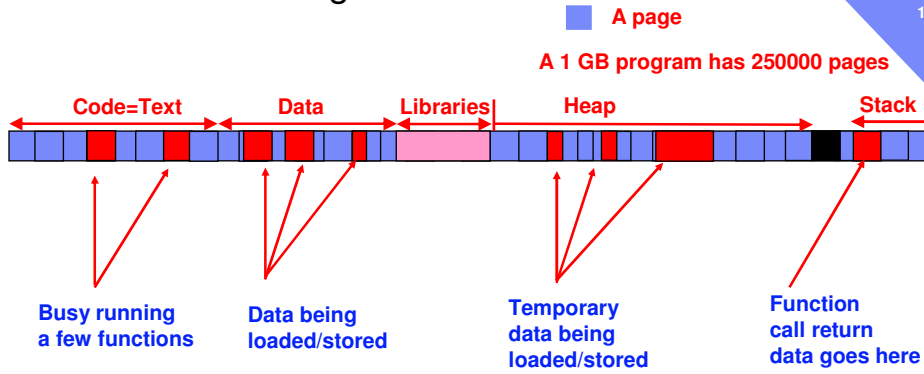
## Classic Virtual Memory (LPAR)



## Five Paging Golden Rules

1. Don't do it! → hurts performance
2. Don't panic! → 10+ pages/s per CPU=noise
3. Do it fast → use many disks
4. Always use Protection → mirror or RAID5
5. Never ever run out of paging space → mayhem!

## What is a "working set" ?

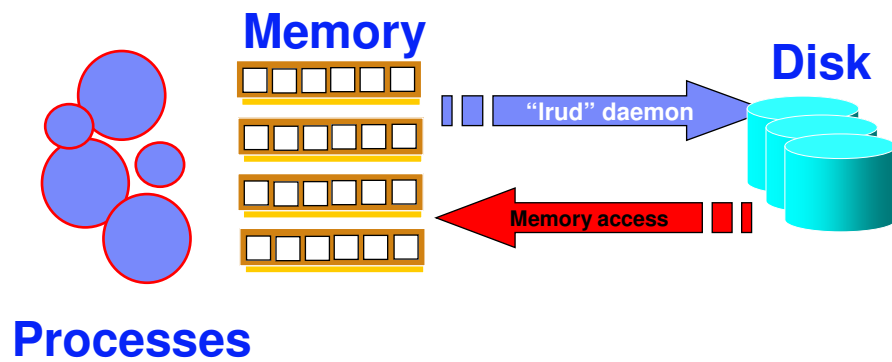


Working Set is the pages needed to run in the short term (seconds)

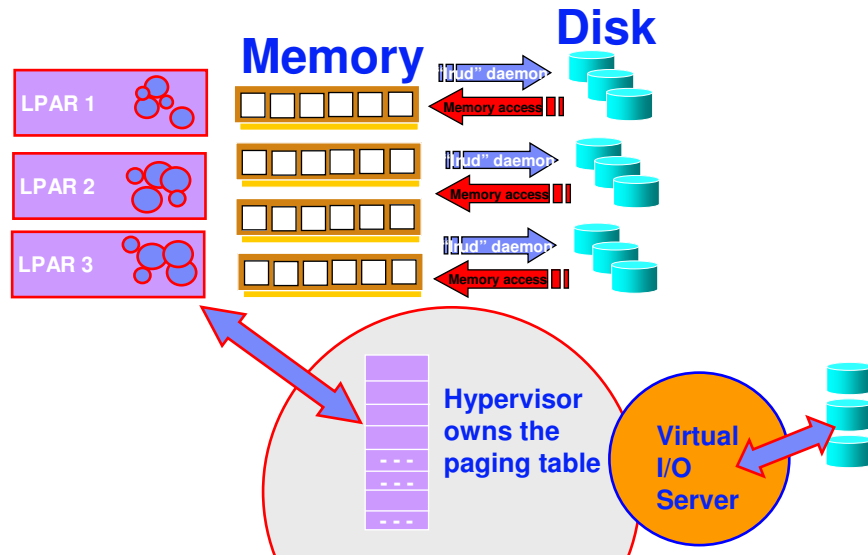
Also called Resident Set (resident in memory), see ps or nmon ResText & ResData

AMS acts on Working Sets but at whole LPAR level

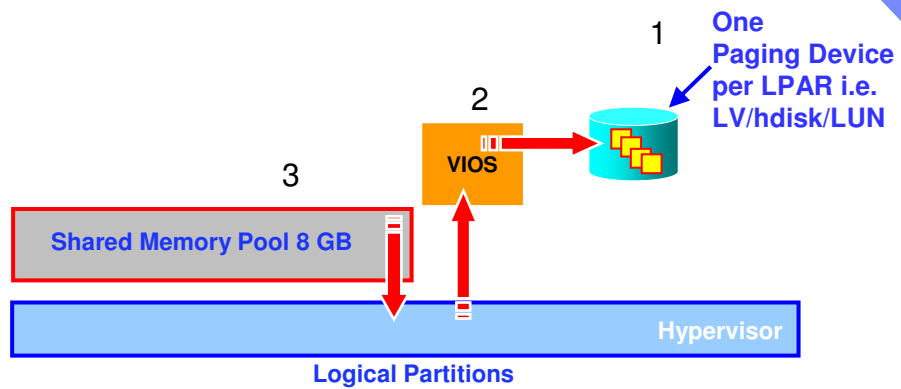
## OS Level Paging



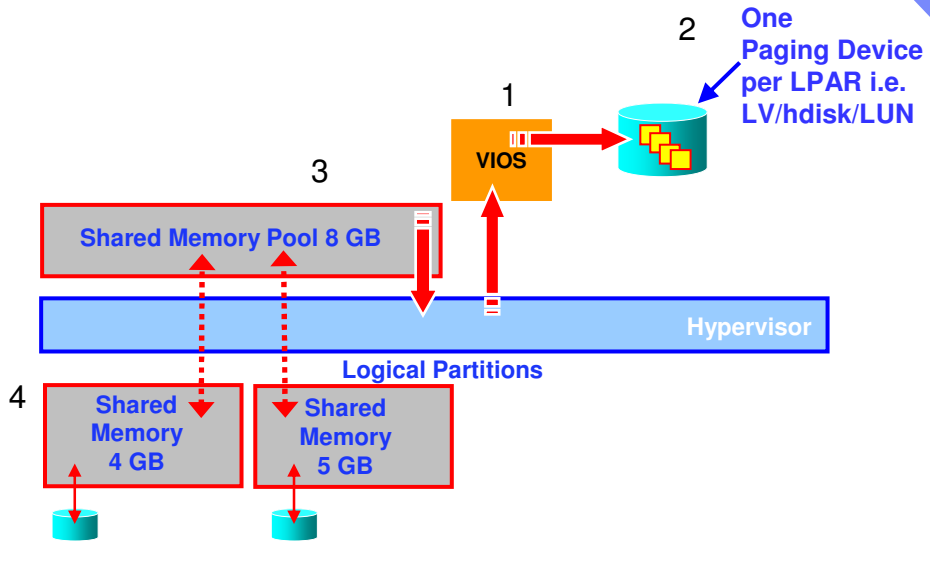
# VM Level Paging = AMS



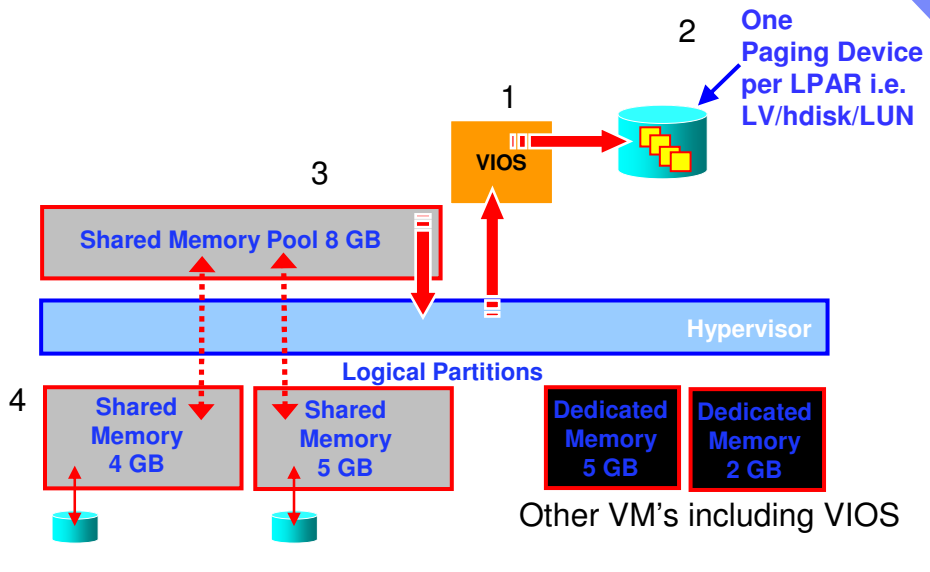
# How is it set up?



# How is it set up?



# How is it set up?



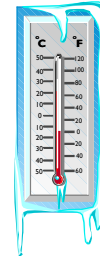


## AMS Algorithm 1 – It all fits

© 2011 IBM  
17

Assuming many VM sharing a pool  
& total VM logical memory > pool

Local paging at OS level  
Not an issue



**“Relaxed Mode”**



## AMS Algorithm 2 - If it nearly fits?

© 2011 IBM  
18

Hypervisor asks OS images for help

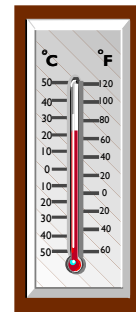
→ once a second

OS then frees memory, if necessary paging out

Loans pages to Hypervisor

Hypervisor gives pages to high demand VM

OS level AMS Tuning on how aggressive:  
none/File system cache/programs too



**“Co-operative Mode”**



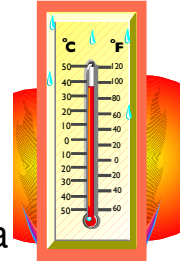
## AMS Algorithm 3 – Loans are not enough

© 2011 IBM  
19

VMs refuse to loan more memory

→ Hypervisor gets aggressive

1. Finds pages to steal
  - It can see the page tables
  - It avoids critical memory pages
  - Least Recently Used page table data
2. Asks VIOS to page out VM memory
3. Once the memory page is free
4. Gives pages to high demand LPAR



VM's are not aware of this happening

**“Aggressive Mode”**



## AMS Algorithm 3 – Loans are not enough

© 2011 IBM  
20

Now VM accesses a page that is not present

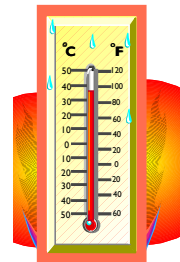
→ Causes page fault

Normally, Hypervisor hands interrupts  
to the VM to handle

Checks: if it's a Hypervisor paged pages

If yes, it recovers the page and  
restarts the instruction

If no, it passes the page fault onto  
OS to handle as normal



**“Aggressive Mode”**



# Creating the AMS Pool

## Machine Level - Memory Pool on HMC

Hardware Management Console

Systems Management > Servers

Select	Name	Status	Available Processors	Available Memory (C)	Referen Code	Configural Memory (C)	Serial Numbr	It	Processi Units	Memory (C)	
<input checked="" type="checkbox"/>	p520-bronze-SN10E0A21			3,1875		16	10E0A21		2	0.5	2
<input type="checkbox"/>	bronze_lpar2								3	0.5	2
<input type="checkbox"/>	bronze_lpar3								4	0.5	2
<input type="checkbox"/>	bronze_lpar4								5	1.75	2
<input type="checkbox"/>	bronze_lpar5								6	0.25	2
<input type="checkbox"/>	bronze_lpar6								1	0.5	2
<input type="checkbox"/>	bronze_vios1										
<input type="checkbox"/>	p520-gold-SN10E0A11	Operating	1.1								
<input type="checkbox"/>	p520-silver-SN10E0A31	Operating	0								
<input type="checkbox"/>	p550-8way	Operating	0.5	7							
<input type="checkbox"/>	p570-8F	Operating	0.6	0							
<input type="checkbox"/>	Power5-p550Q	Operating	0	0			16	85DCCB0			

Total: 12 Filtered: 12 Selected: 1

Context menu for p520-bronze-SN10E0A21:

- Properties
- Operations
- Configuration
  - Create Logical Partition
  - System Plans
  - Partition Availability Priority
  - View Workload Management Groups
  - Manage Custom Groups
  - Manage Partition Data
  - Manage System Profiles
- Virtual Resources
  - Shared Processor Pool Management
  - Shared Memory Pool Management
  - Virtual Storage Management
  - Virtual Network Management

**Shared Memory Pool Management**

## Machine Level - Memory Pool - skipping selection panels

Create Shared Memory Pool - p520-bronze-SN10E0A21

**Summary**

Here is summary of all the selections that have been made. Once you make sure all the entries are correct, select 'Finish' button to create the pool with these entries.

Maximum pool size: 3.0 GB  
Pool size: 2.0 GB  
Paging VIOS 1: bronze\_vios1  
Paging VIOS 2:

Paging space device(s):

VIOS	Device Name	Device Size (GBs)	Device Status	Redundancy Capable	Physical Location Code
bronze_vios1	ams100	4.0		No	
bronze_vios1	ams101	4.0		No	
bronze_vios1	ams102	4.0		No	
bronze_vios1	ams103	4.0		No	
bronze_vios1	ams104	4.0		No	

< Back   Next >   **Finish**   Cancel   Help

ibm.com https://hmc8.aixncc.uk.ibm.com/

Creating Memory Pool, please wait...

**Annotations:**

- Pool Size (points to 3.0 GB)
- Decide VIOS(s) (points to bronze\_vios1)
- My naming convention (points to bronze\_vios1 in table)
- Decide the AMS paging devices (points to Device Name column)
- Single VIOS (points to Redundancy Capable column)
- Note (points to Physical Location Code column)

## Making the VM use the Pool

## Modify VM to use Shared Memory

**Managed Profiles -- silver\_lpar3**

Actions  
New...  
Edit... contains the resources for this profile. You can modify the profile by editing the profile.  
Copy...  
Delete  
Activate... Default Profile

**Logical Partition Profile Properties: normal @ silver\_SN10E0A31 - silver\_lpar3**

General Processors Memory I/O Virtual Adapters Power Controlling

Memory mode  
 Dedicated  
 Shared

**Dedicated Memory**  
Installed memory (MB): 16  
Current memory available for partition usage:  
Minimum memory : 1  
Desired memory : 2  
Maximum memory : 4

**Shared Memory Warning - silver\_lpar3**  
Switching from Dedicated Memory Mode to Shared Memory Mode will remove all Physical I/O Devices.  
Are you sure you want to switch to Shared Memory Mode?  
Yes No

**Logical Partition Profile Properties: normal @ silver\_lpar3 @ p520-silver-SN10E0A31 - silver\_lpar3**

General Processors Memory I/O Virtual Adapters Power Controlling Settings Logical Host Ethernet Adapters (LHEA)

Memory mode  
 Dedicated  
 Shared

**Logical Memory**  
Shared memory pool size (MB): 16384  
Total assigned logical memory (MB) : 15552  
Minimum memory : 1 GB  
Desired memory : 2 GB  
Maximum memory : 4 GB

**Shared Memory Options**  
Memory Weight (0-255) 0

**Can't be both**

**So copy the profile 1st ☺**

**Now logical memory i.e. what we would "like"**

Now cold stop & restart the Virtual Machine (LPAR) using the new AMS profile

Memory given to the VM on demand so it starts small and grows

## Each VM uses 1 AMS paging space → So RAID5 LUNS using lots of underlying spindles is good

Systems Management > Servers

Select	Name	Status	ID	Available Processing Units	Processing Units	Processor	Available Memory (GB)	Memory (GB)
<input checked="" type="checkbox"/>	p520-silver-SN10E0A31	Operating		2			975	
<input type="checkbox"/>	silver_lpar3	Running	3		0.5	1		2
<input type="checkbox"/>	silver_lpar4	Running	4		0.5	1		2
<input type="checkbox"/>	silver_lpar5	Not Activated	5		0	0		0
<input type="checkbox"/>	silver_vios	Running	2		0.5	1		1
<input type="checkbox"/>	silver_lpar2	Running	1		0.5	1		2

In use

Pool Properties - p520-silver-SN10E0A31  
Virtual I/O Server

The table below shows the paging devices and their assigned partitions. To add or remove paging devices or to change the Virtual I/O Server, select Add/Remove Devices.

Paging Devices:

Partition ID	PSP Devices	Device Name	Device Size	Device Status	Location Code
4	silver_vios_ams1	silver_vios_ams1	32768	Active	
3	silver_vios_ams2	silver_vios_ams2	32768	Active	
1	silver_vios_ams3	silver_vios_ams3	32768	Active	
	silver_vios_ams4	silver_vios_ams4	32768	Inactive	

Named VIOS

Logical Volume on the VIOS

32 GB

In use

Only 1 more AMS VM can be started

## AIX Level: vmstat -h

Logical Memory

Memory Pool Size

```
# vmstat -h 10
System configuration: lcpu=2 mem=2048MB ent=0.50 mmode=shared mpsz=4.00GB
```

kthr	memory	page	faults	cpu	hypv-page
r b	avm fre	re pi po fr sr cy in sy cs	us sy id wa pc ec	hpi hpit pmem Loan	
0 0	190419 173073	0 0 0 0 0 0 0 2 149 159	0 1 99 0 0.01 1.6	0 0 1.20 0.80	
0 0	190419 173073	0 0 0 0 0 0 0 2 24 152	0 0 99 0 0.00 0.8	0 0 1.20 0.80	
0 0	190419 173073	0 0 0 0 0 0 0 1 19 166	0 0 99 0 0.00 0.8	0 0 1.20 0.80	
0 0	207225 189696	0 0 0 0 0 0 0 6 334 196	35 2 64 0 0.18 36.9	25 53 1.33 0.67	
0 0	207227 189694	0 0 0 0 0 0 0 2 39 170	50 1 50 0 0.25 50.8	7 15 1.33 0.67	
0 0	207227 189694	0 0 0 0 0 0 0 5 20 164	50 1 50 0 0.25 50.8	0 0 1.33 0.67	

Logical Memory Statistics

Watching AMS on a single LPAR is ... "insane"

Hypervisor Page-ins Faults/s

Time waiting for hypervisor page-ins (in milliseconds)

Physical Memory → pmem

Loaned Memory

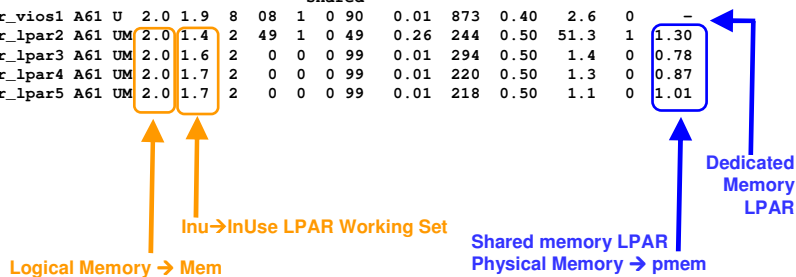
If AIX\_ams\_loan\_policy=0 (off) this will be zero

## CEC Level on VIOS or AIX: topas -C (hit "g" for the extra top info)

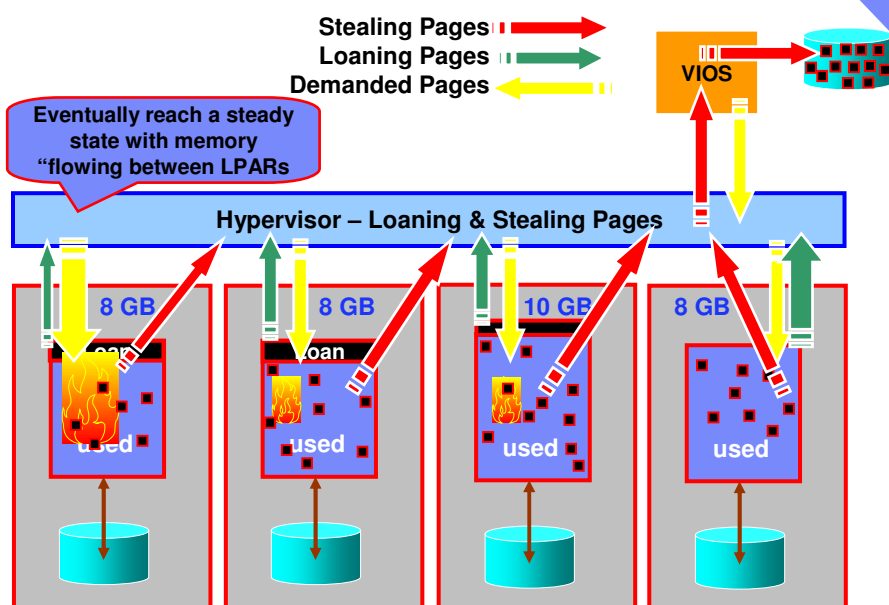
```

Topas CEC Monitor          Interval: 10          Wed Dec 3 10:15:06 2008
Partition Info  Memory (GB)  Processor  Virtual Pools : 0
Monitored : 4  Monitored : 8.0  Monitored : 2.0  Avail Pool Proc: 3.7
UnMonitored: -  UnMonitored: -  UnMonitored: -  Shr Physical Busy: 0.28
Shared : 4  Available : -  Available : -  Ded Physical Busy: 0.00
Uncapped : 4  UnAllocated: -  UnAllocated: -  Donated Phys. CPUs: 0.00
Capped : 0  Consumed : 6.5  Shared : 2  Stolen Phys. CPUs: 0.00
Dedicated : 0  Donated : 0  Hypervisor
Donating : 0  Pool Size : 4  Phantom Interrupts : 1
  
```

Host	OS	M	Mem	InU	Lp	Us	Sy	Wa	Id	PhysB	Vcsw	Ent	%EntC	PhI	pmem
-----shared-----															
silver_vios1	A61	U	2.0	1.9	8	08	1	0	90	0.01	873	0.40	2.6	0	-
silver_lpar2	A61	UM	2.0	1.4	2	49	1	0	49	0.26	244	0.50	51.3	1	1.30
silver_lpar3	A61	UM	2.0	1.6	2	0	0	0	99	0.01	294	0.50	1.4	0	0.78
silver_lpar4	A61	UM	2.0	1.7	2	0	0	0	99	0.01	220	0.50	1.3	0	0.87
silver_lpar5	A61	UM	2.0	1.7	2	0	0	0	99	0.01	218	0.50	1.1	0	1.01



## Active Memory Sharing in Action



## Active Memory Sharing - Expectations

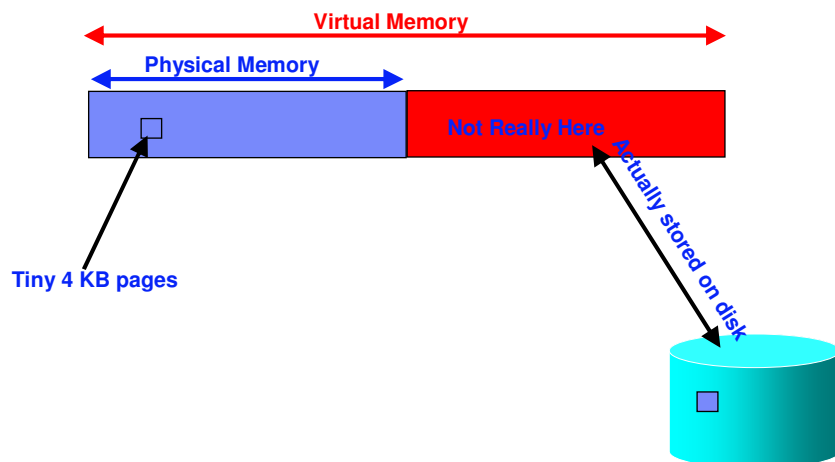
Memory pages flow between Virtual Machines



Lava Wave Machine

1. Sudden GB's of memory moved would require massive paging & a large performance hit.
2. Don't want this for transitory peak.
3. So few MB/s arrive until demand eases off

## Classic Virtual Memory (VM level)





## Active Shared Virtual Memory (VM Level)

