PowerVM and AIX v6.1 Technologies

Jaqui Lynch
Architect of Power
Mainline Information Systems
Jaqui.lynch@mainline.com
Ron.gordon@mainline.com
Larry.grubbs@mainline.com

Agenda

- Virtualization Level Set
  - Justification and Introduction
  - Live Partition Mobility
  - Workload Partitions
  - Shared Processor Pools
  - Shared Active Memory (Memory Pools)
- AIX v6.1

Introduction

VIRTUALIZATION

PowerVM Virtualization

New Branding, New Features, New Offerings
Previous name: Advanced POWER Virtualization option

Three PowerVM Offerings:

- Express Edition
  - Entry Systems, limited functionality
- Standard Edition
  - Base services, full hardware
- Enterprise Edition
  - Standard Edition with “Mobility” support
  - POWER6 Processor-based Systems only

Variety of Features

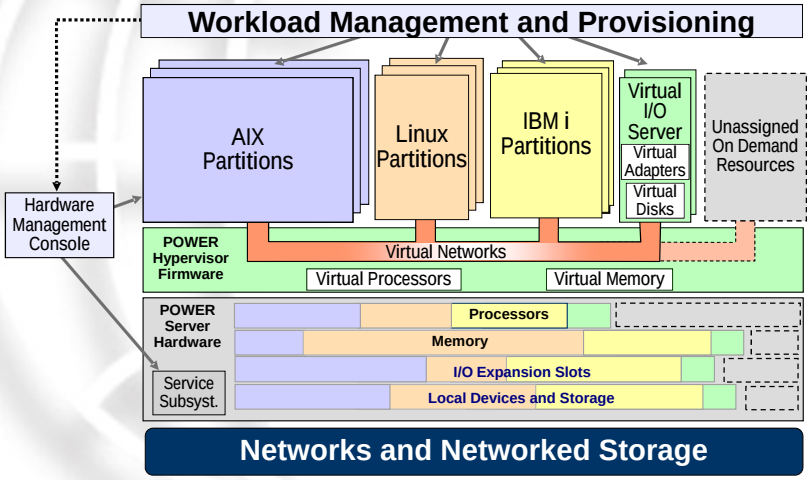
- Virtual Networks, Virtual Storage, Shared Processor Pools, etc.

Foundation for Server Consolidation

Reduce / Eliminate Planned Downtime

Green Computing Technology

PowerVM Virtualization Architecture



Virtualization of P5 & P6 servers is accomplished using two layers of firmware:

- **A thin core hypervisor that virtualizes processors, memory, and local networks**
- **One or more Virtual I/O Server partitions that virtualize I/O adapters and devices**

5

PowerVM Offerings

Feature/Function	Express Edition	Standard Edition	Enterprise Edition
Servers Supported	p520 / p550	P6 Blades, Power Systems	JS22/23/43, Power Systems (P6)
Max LPARs	2 DLPARS +1 VIOS per Server	10 / Core	10 / Core
Management	IVM	IVM & HMC	IVM & HMC
VIOS	Yes	Yes	Yes
Live Partition Mobility	No	No	Yes
Active Memory Sharing	No	No	Yes
Shared Processor Pools	No	Yes (P6 & HMC Required)	Yes (HMC Required)
Shared Dedicated Capacity	Yes	Yes (POWER6: Servers & Blades)	Yes
Operating Systems	AIX / Linux / i	AIX / Linux / i	AIX / Linux / i
PowerVM Lx86	Yes	Yes	Yes

6

Reasons to Virtualize

- Server Consolidation
  - Increase Utilization
- Hosting
- Migration or adding an application
- Application performance needs
- Production and Test on same hardware
- Multiple Operating Systems
- Consolidate Applications on different time zones
- Complying with license agreements or other compliance needs
- SOA application or SOI strategy decisions
- Flexibility and scalability
- Optimization of resources
- Reduce hardware, power and cooling and CO2 emissions

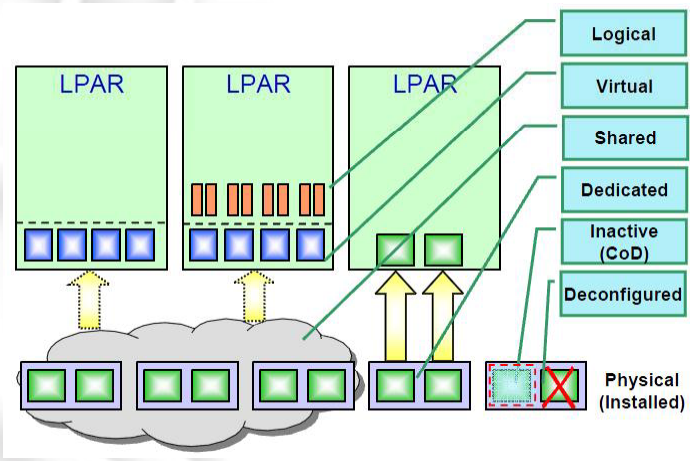
Justification

- Faster cores on servers mean you need to be able to micropartition them
- Fewer physical servers
  - Server management and administration costs
  - Fewer FTEs needed per server
  - Less complex asset management
- Power and cooling costs with increased utilization of existing servers
- Less floor space
- Use of technologies such as NIM allow rapid deployment of partitions
- Time to market, virtual resources can be deployed immediately

Virtualization Options

- Real Resources (unvirtualized)
  - Dedicated processors/cores
  - Dedicated fibre or SCSI
  - Dedicated Ethernet
- Virtual or Shared Resources
  - Shared processors/cores
  - Dedicated donating
  - Virtual ethernet
  - Shared ethernet adapter
    - Built on virtual ethernet
  - Shared SCSI (aka Virtual SCSI)
    - Can be SCSI or fibre
  - Ethernet and SCSI use a custom LPAR called a VIO server
    - Must include processor and memory resources in planning for that LPAR or LPARs
- Hybrids

Logical, Virtual, Physical



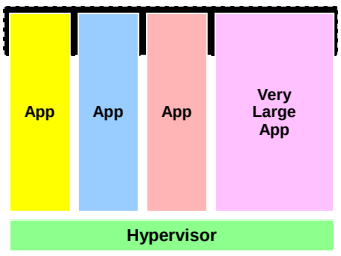
Courtesy Janel Barfield and IBM 2009

Dedicated Cores

- Allocated as whole cores to an LPAR
- Same physical cores used for that LPAR when it is running
- 1-1 relationship physical processor to LPAR proc? i.e. proc2
- Can still use SMT and have logical processors
- When LPAR stops cores may or may not go into shared pool depending on settings
- Uses processor affinity for best performance

Virtual (Shared) CPUs

- Potential Benefits
  - Increase CPU utilization
  - Actual deployment effort is modest
  - More applications and partitions per server
- Issues/Considerations
  - High utilization LPARs will be poor donors but might benefit from use of the uncapped pool
  - Most mainframes run in exclusively this mode
  - Understand entitlement, VPs, weight, capped/uncapped, weight, reserve capacity on demand, processor folding.
  - Software licensing - use of uncapped LPARs with unnecessary VPs may impact costs
  - Review performance management tools
  - Not every application likes sharing – depends on workload characteristics



MicroPartioning
Shared processor partitions

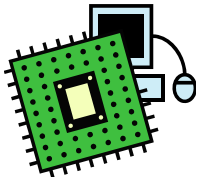
- Micro-Partitioning allows for multiple partitions to share one physical processor
- Up to 10 partitions per physical processor
  - Min is 0.1 processing units and increments of 0.01 after that
- Up to 254 partitions active at the same time
- One shared processor pool – more on the p6-570
- Partition's resource definition
  - Minimum, desired, and maximum values for each resource
  - Processor capacity (processor units)
  - Virtual processors
  - Capped or uncapped
    - Capacity weight
    - Uncapped can exceed entitled capacity up to number of virtual processors (VPs) or the size of the pool whichever is smaller
  - Dedicated memory
    - Minimum of 128 MB and 16 MB increments
  - Physical or virtual I/O resources
  - Some workloads hate the SPP – SAS is one

Applications and SPLPARs

- Applications do not need to be aware of Micro-Partitioning
- Not all applications benefit from SPLPARs
- Applications that may not benefit from Micro-Partitioning:
  - Applications with a strong response time requirements for transactions may find Micro-Partitioning detrimental:
    - Because virtual processors can be dispatched at various times during a timeslice
    - May result in longer response time with too many virtual processors:
      - Each virtual processor with a small entitled capacity is in effect a slower CPU
    - Compensate with more entitled capacity (2-5% PUs over plan)
  - Applications with polling behavior
  - CPU intensive application examples: DSS, HPC, SAS
- Applications that are good candidates for Micro-Partitioning:
  - Ones with low average CPU utilization, with high peaks:
    - Examples: OLTP, web applications, mail server, directory servers

Defining Processors

- Minimum, desired, maximum
- Maximum is used for DLPAR
  - Max can be used for licensing
- Shared or dedicated
- For shared:
  - Desired is also called entitlement
  - Partition's guaranteed amount of core is its Entitled Capacity
  - Capped
  - Uncapped
    - Variable capacity weight (0-255 – 128 is default)
    - Weight of 0 is capped
    - Weight is share based
    - Can exceed entitled capacity (desired PUs)
    - Cannot exceed desired VPs without a DR operation
  - Minimum, desired and maximum Virtual Processors
    - Max VPs can be used for licensing



More on shared processor pool

- Dispatch time for a core is a 10ms timeslice
- A .3 entitlement means 30% of a core or a 3ms timeslice
- A 1.4 entitlement means the LPAR is entitled to 14ms of processing time for each 10ms timeslice (obviously across multiple cores)
- Hypervisor dispatches excess idle time back to the pool (called a cede)
- Processor affinity tried to take into account hot cache
- LPAR may run on multiple cores depending on entitled capacity, virtual processors and interrupts

Uncapped vs Capped

- Capped LPARs can cede unused cycles back but can never exceed entitlement
- Uncapped LPARs can exceed entitlement up to the size of the pool or the total virtual processors, whichever is smaller
- Unused capacity is ceded back
- User defined weighting (0 to 255) is used to resolve competing requests

Virtual Processors

- Used to tell the operating system how many physical processors it thinks it has
- Partitions are assigned PUs (processor units)
- VPs are the whole number of concurrent operations
  - Do I want my .5 as one big processor or 5 x .1 (can run 5 threads then)?
- VPs round up from the PU by default
  - For every 1.00 or part thereof of a processor unit there will be at least 1 VP allocated
  - .5 Pus will be 1 VP
  - 2.25 Pus will be 3 VPs
  - You can define more and may want to
  - Basically, how many physical processors do you want to spread your allocation across?
- VPs put a cap on the partition if not used correctly
  - i.e. define .5 PU and 1 VP you can never have more than one PU even if you are uncapped
- Cannot exceed 10x entitlement
- VPs are dispatched to real processors
- Dispatch latency – minimum is 1 millisec and max is 18 millisecs
- VP Folding
- Maximum is used by DLPAR
- Use commonsense when setting max VPs!!!
- In a single LPAR VPs should never exceed Real Processors
- Maximum VPs per partition is 64 or 10 x entitlement (desired PUs)

Virtual Processors 2

- Used to tell the operating system how many physical processors it thinks it has
- Operating system does not see entitlement – it sees configured virtual processors
  - EC=1.4
  - VP = 2
  - Operating system sees 2 processors (proc0 and proc2)
  - If SMT is on then you will see those broken down into 4 logical cpus (cpu0, cpu1, cpu2, cpu3)
- Both entitled capacity and number of virtual processors can be changed dynamically for tuning
- Virtual processor count does not change entitlement – it is about how the entitlement will be delivered
- Capped
  - Entitlement = 1.4 and 2 x VPs
  - For each 10ms timeslice the LPAR is entitled to 14ms of processing time
  - For 2 VPs that equates to 7ms on each of 2 physical processors
- Uncapped
  - Entitlement = 1.4 and 2 x VPs
  - Each VP can get up to 10ms of processor time
  - For 2 VPs that equates to 20ms across 2 physical processors
  - Can't grow beyond 2 VPs even if more are available in the pool
  - VPs becomes a soft cap on the LPAR
  - Be sure to allocate enough VPs to grow for uncapped LPARs

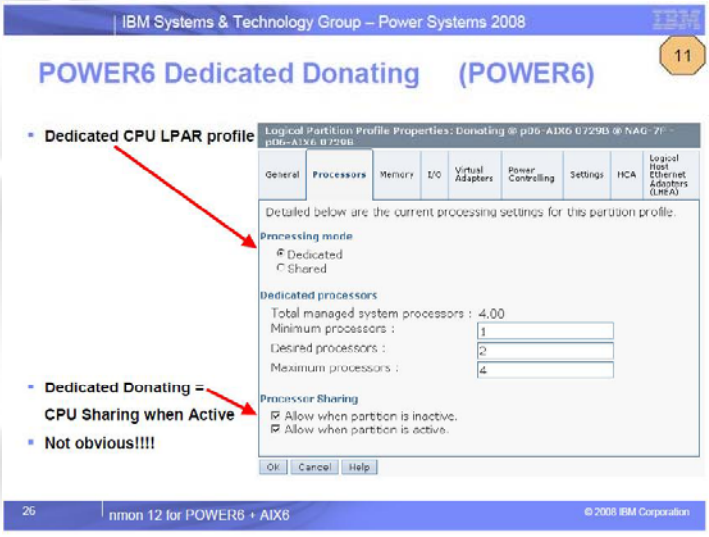
Examples

- LPAR 1 - uncapped
  - Ent = 2.0
  - Max = 6.0
  - VPs = 4.0
  - Can grow to 4 processor units
  - VPs cap this
- LPAR 2 - uncapped
  - Ent = 2.0
  - Max = 6.0
  - VPs = 6.0
  - Can grow to 6 processor units
- LPAR 3 – Set as Capped
  - Ent = 2.0
  - Max = 6.0
  - VPs = 4.0
  - Can't grow at all beyond 2 processor units

SMT

- Simultaneous Multi-Threading
- Each virtual processor (or processor if dedicated) runs two threads
- Each thread is called a logical processor
- lcpu=4 in vmstat, etc
  - If shared then you either have a PU of 2 plus SMT or you have a PU of 4 and no SMT
  - If dedicated then you either have 2 cores with SMT or 4 cores and no SMT

Dedicated Donating



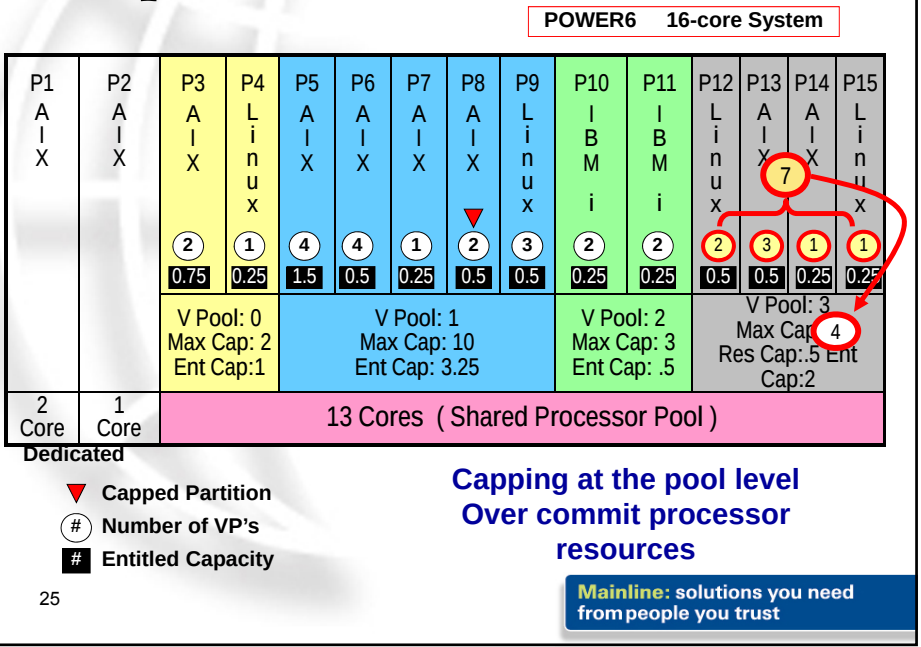
Introduction

MULTIPLE SHARED PROCESSOR POOLS

Multiple Shared Processor Pools

- POWER6 processor-based systems
- Grouping of partitions into subsets called "Pools"
- Can manage processors resources at the subset
  - AIX or Linux partitions
  - Can assign caps at the group level
- Provides the ability to balance processing resources between partitions assigned to the shared pools.
- Helps optimize use of processor cycles
- Segment production / development / test / etc.
- Mobility of partitions is supported
- Maximum: 64 Pools

Multiple Shared Processor Pools



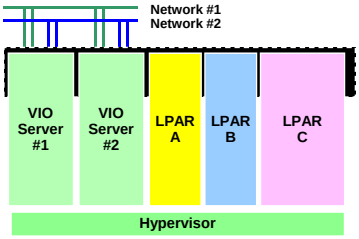
Capping at the pool level
Over commit processor resources

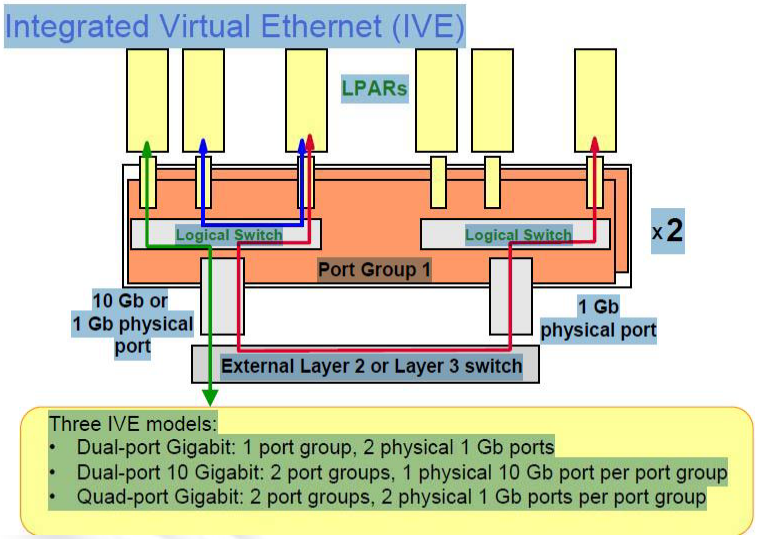
Multiple Shared Processor Pool Details

- Default, there is one virtual shared processor pool (pool identifier of zero) defined on the system.
  - ✓Maximum capacity of the default pool is the number of processors in the physical shared processor pool
- Virtual shared pools are created from the HMC interface
- Maximum capacity of virtual shared processor pool can be adjusted dynamically
- Partitions can be moved between virtual shared processor pools
- Partition Mobility of a partition in a pool is supported
- When adding / removing a partition to a pool will increase / decrease the entitled capacity of the pool by the amount of the entitled capacity of the partition

Sharing Ethernet

- Potential Benefits
  - Reduce the number Ethernet adapters, ports
    - Better resource utilization
  - Reduce cabling efforts and cables in frames
  - Reduce number of I/O drawers and/or frames
- Issues/Considerations
  - Understand Ethernet adapter/port utilization
  - Understand high availability cluster support requirements
  - Understand implications on backup architecture
  - Understand virtual I/O sizing and large send capabilities
  - Understand use of link aggregation and/or VLANS
  - Understand VIO high availability Ethernet options
  - Simplicity!!
  - Need to plan as it is not an either/or choice – you can have both dedicated and shared resources





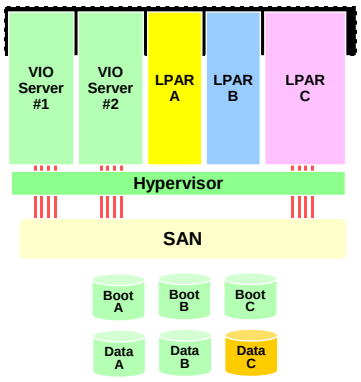
Courtesy IBM - Configuration of Virtual I/O on POWER6 – Janel Barfield - Session ID: pVI08 Power Technical University – September 2008

IVE Notes (Power6 only)

- Which adapters do you want? Each CEC requires one.
  - Dual 10/100/1000 TX (copper)
  - Quad 10/100/1000 TX (copper)
  - Dual 10/100/1000 SX (fiber)
- Adapter ties directly into GX Bus
  - No Hot Swap
  - No Swap Out for Different Port Types (10GbE, etc.)
- Not Supported for Partition Mobility, except when assigned to VIOS
- Partition performance is at least the same as a real adapter
  - No VIOS Overhead
  - Intra-partition performance may be better than using Virtual Ethernet
- Usage of serial ports on IVE
  - Same restrictions as use of serial ports that were on planar on p5
  - Once an HMC is attached these become unusable
- Naming
  - Integrated Virtual Ethernet – Name used by marketing
  - Host Ethernet Adapter (HEA) Name used on user interfaces and documentation

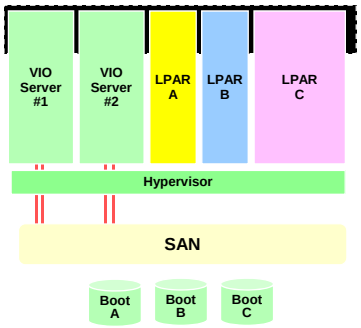
Sharing Disk Adapters

- Potential Benefits
  - Reduce the number FC adapters and ports
  - Reduce cabling efforts and cables in frames
  - Reduce number of I/O drawers and frames.
- Issues/Considerations
  - Understand current SAN adapter / port utilization
  - Investigate high availability cluster support for virtual I/O
  - Understand implications on backup architecture
  - Understand virtual I/O server sizing
  - Understand availability choices such as dual VIOS, number of HBAs, O/S mirroring, etc
  - Need to plan as it is not an either/or choice – you can have both dedicated and shared resources

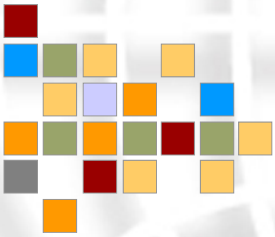


Boot from SAN

- Potential Benefits
  - Reduce the number of I/O drawers
  - Reduce number of frames
- Issues/Considerations
  - Use internal disk for VIO servers
  - Need robust, available SAN
  - Understand and size VIOS LPARs
  - Understand availability choices such as dual VIOS, multi-path I/O, O/S mirroring, etc.



Note: LPARs could boot through the VIOS and have dedicated HBAs for data access.



AIX

20+ Years of AIX Progress

AIX V2 & V3
Establishment in the market:
- RISC Support
- UNIX credibility
- Open Sys. Stds..
- Dynamic Kernel
- JFS and LVM
- SMIT

AIX V3.2.5
Maturity:
- Stability
- Quality

AIX V4.1/4.2
SMP Scalability:
- POWERPC spt.
- 4-8 way SMP
- Kernel Threads
- Client/Server pkg
- NFS V3
- CDE
- UNIX95 branded
- NIM
- > 2GB filesystems
-HACMP Clustering
- POSIX 1003.1, 1003.2, XPG4
- Runtime Linking
- Java 1.1.2

AIX V4.3
Higher levels of scalability:
- 24-way SMP
- 64-bit HW support
- 96 GB memory
- UNIX98 branded
- TCP/IP V6
- IPsec
- Web Sys. Mgr.
- LDAP Dir. Server.
- Workload Mgr
- Java JDT/JIT
- Direct I/O
- Alt. Disk Install
- Exp/Bonus CDs

AIX V5.1
Industry Leading Performance:
- POWER4 support
- Static LPAR
- New 64bit kernel
- 32-way SMP
- 256GB mem
- JFS2
- Networking enh.
- Java 2 support
- Dynamic CPU Deallocation
- Cluster Mgt (CSM)
- GRID Toolkit

AIX V5.2
Flexible Resource Management:
- POWER4+ spt.
- Dynamic LPAR
- Dynamic CUoD
- Dyn. CPU Sparing
- 512GB mem
- 16 TB filesystems
- UNIX03 branded
- Concurrent I/O
- MultiPath I/O
- Mobile IP V6
- System UE Gard
- Flex LDAP Client
- XSSO PAM spt

AIX V5.3
Advanced Virtualization:
- POWER5 spt.
- 64-way SMP
- SMT
- MicroPartitions™
- Virt I/O Server
- Partition Load Mgr
- NFS Version 4
- Adv. Accounting
- Scaleable VG
- JFS2 Shrink
- SUMA
- SW RAS features
- POSIX Realtime

AIX V6.1
PowerVM Virtualization:
- POWER6 spt.
- 64-way SMP
- Enhanced SMT
- MicroPartitions™
- Virt I/O Server
- Workload Partitions
- Variable Pages
- Storage Keys
- App Mobility
- Hot Patch
- Tracing Facilities
- Dec Floating Point
- Improved Dumps

Open Systems Workstations	*Distributed Client-Server*	*Network Centric Computing*	*e-Business Computing*	*On Demand Business*	*POWER of 6*
Uni-processor	4-8 way SMP	24-way SMP	32-way SMP	32-way SMP	64/way SMP

POWER6 Delivers with Your Choice of AIX or Linux

- Broad application selection
- Wide range of workloads
- Reduced Complexity
- Potential cost savings with consolidation
- Live Partition Mobility
  - *Linux, AIX V5.3 and AIX V6.1*

AIX 6*
- Virtualization
  - Workload Partitions
  - Live Application Mobility
- Security
- Availability
- Manageability
- Binary compatible**

AIX 5L V5.2/5.3
- Binary compatible with existing applications on POWER6*
- Micro-Partitioning
- Mainframe-inspired RAS features hardware and operating system
- Scalability up to 128 threads

Linux on POWER
- POWER and x86 apps [2H07]
- PowerVM
- Reliability, Availability, Serviceability features
- Scalability to 128 threads

*Complete details on AIX binary compatibility can be found at http://www.ibm.com/servers/aix/os/compatibility/

AIX Binary Compatibility Guarantee

*Complete details on AIX binary compatibility can be found at http://www.ibm.com/servers/aix/os/compatibility/

IBM

AIX 6 Binary Compatibility

Dear System p and System i clients:

We listened and we have delivered for you, and in fact we hope you have been ecstatic with the recent announcements on the AIX® 6 and POWER6™ products. Not only can the new POWER6 servers run AIX V5.2, V5.3 and AIX V6.1–with binary compatibility for applications currently running on AIX V5.1, V5.2 and V5.3–but AIX 6 will even run on older hardware, based on POWER5™, and POWER4™ processors. This broad support for multiple levels of the AIX operating system on multiple generations of POWER systems is the strongest that we have ever had.

But some clients have said that they want to hear it from me. We've said we will offer binary compatibility and we mean it. We are offering a guarantee* that your applications, whether written in house or supplied by an application provider, will run on AIX 6 if they currently run on AIX V5.1, V5.2 or V5.3–without recompilations or modification. Even well-behaved 32-bit applications from AIX V4.1, V4.2, and V4.3 will run without recompilation.

Take us up on that challenge. We assume (and require) that these applications comply with reasonable programming standards (see ibm.com/systems/p/os/aix/compatibility/conditions), but if they do and the applications will not run on AIX 6, contact us. We will investigate and assign our developers to work on the binary compatibility problem. I don't anticipate anyone will call but I wanted to assure you that we are committed to the binary compatibility of AIX 6.

The qualities of the AIX operating system—virtualization, security, performance, and quality—have won many new clients to AIX. AIX 6 will be the next step forward in the evolution of UNIX, while allowing existing AIX V5.1, V5.2 and V5.3 applications to continue to run. AIX is and will remain the strategic UNIX operating system for IBM.

Thank you for your continued confidence in IBM System p and System i servers and in the AIX operating system. Keeping your applications up and running is one of our primary goals. I want you to rest assured that we are taking great care to insure that when you upgrade to AIX 6, your AIX V5.1 V5.2 and V5.3 applications will not only run unmodified, but you will also be able to take advantage of the new innovations in AIX 6.

Sincerely,

Ross A. Mauri
General Manager
IBM System p

POWER6 and AIX 6 new function

Feature	Licensed Via		Supported OS			Supported Hardware			GA Date
	APV	AIX v6.1	AIX v5.3	AIX v6.1	Linux	POWER4	POWER5	POWER6	
Dedicated processor sharing	✓		✓	✓	✓			✓	6/07
Hardware Decimal FP			✓	✓	✓			✓	6/07
Integrated Virtual Ethernet			✓	✓	✓			✓	6/07
Storage keys - application			✓	✓				✓	6/07
Storage keys – kernel				✓				✓	4Q07
Live Partition Mobility	✓		✓	✓	✓			✓	4Q07
Multiple virtual shared pools	✓		✓	✓	✓			✓	4Q07
WPARs		✓		✓		✓	✓	✓	4Q07
Live Application Mobility		✓		✓		✓	✓	✓	4Q07

Introducing AIX 6

AIX 6 is binary compatible* **with AIX 5L™**
It is *named* to reflect its unity with POWER6

The Power of SIX – AIX 6 and POWER6

- **Workload Partitions**
- **Live Application Mobility**
- **Firmware assisted Dump**
- **Dynamic, variable page size**
- **Kernel Storage Keys**

AIX 6

- **Live Partition Mobility**
- **Application Storage Keys**
- **Hardware Decimal Floating-Point**
- **Shared Dedicated processor**
- **Improved SMT**
- **Energy Management**

POWER6

IBM Power Systems Innovation and *PowerVM*
Provide Unique Features for ISV and Client Exploitation

*Complete details on AIX binary compatibility can be found at http://www.ibm.com/servers/aix/os/compatibility/

37

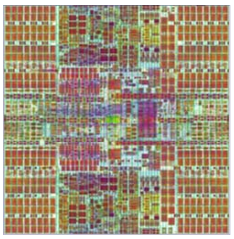
AIX 6 Features….

- POWER6 Exploitation
- Software Reliability Availability Serviceability
- Enhancements to existing Virtualization Technologies
- Workload Partitions (Software based Virtualization)
- Application Mobility (Cross system Workload Mobility)
- 64-bit Kernel only
- Integrated Multilevel Security
- Role Base Access Control (Partial Root base)
- Encrypted File system
- CAPP EAL4+ and LSPP Security Certification
- Solution Performance Tuning
- AIX Kernel Hot-Patching
- Dynamic Tracing for AIX
- Ease of Use
  - –Portal base SMT, LPAR Simplification

38

POWER6 Support

- POWER6 processors are Binary Compatible* with previous POWER processors
- Exploitation of some features of POWER6 will require AIX 6
  - **Dynamic, variable page size**
  - **Firmware assisted Dump**
  - **Kernel Storage Keys**
- But most features of POWER6 supported by AIX V5.3
  - **Live Partition Mobility**
  - **Shared Dedicated processor**
  - **Improved Simultaneous Multithreading**
  - **Hardware Decimal Floating-Point**
  - **Application Storage Keys**
  - **POWER6 Energy Management**



POWER6

*Complete details on AIX binary compatibility can be found at http://www.ibm.com/servers/aix/os/compatibility/

Smooth Upgrade to AIX 6

- **AIX 6 is binary compatible with AIX 5L**
  - Current applications will continue to run
  - Runs on POWER4, POWER5, POWER6
  - Open beta will provide early access to AIX 6
  - Binary compatibility letter from Ross Mauri – General Manager Power Systems
  - Other compatibility activity planned
- **No charge upgrade for current AIX 5L clients with SWMA**
  - No additional out of pocket expense for clients
- **Upgrade process**
  - Tools like alt disk installation and NIM minimize client risk

AIX 6 Processor Support….

AIX 6 will not support 32 bit hardware and older 64 bit hardware

Processors no longer supported:

- All 32-bit processors — i.e. F50, E30, etc.
- RS64 family of servers — i.e. S80, H80, M80, etc.
- POWER3™ servers — i.e. p610, p640, etc.

Processors supported:

- PPC970MP
- POWER4
- POWER5
- POWER6

AIX 6 will support both 32-bit & 64-bit libraries & applications

- Full support for 32-bit applications

41

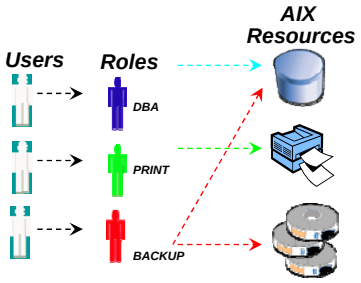
Mainline: solutions you need from people you trust

AIX 6.1 Role Based Access Control

- Improved Administrative Security
  - Improved security by reducing the need for many root users
  - Reduced administration cost thought delegation

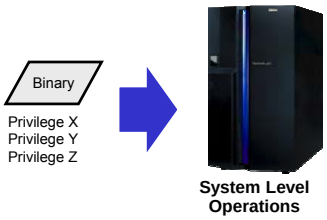
aix	device	boot	create	"create boot image"
	fs	config	Halt	"halt the system"
	network	install	Info	"display boot information"
	proc	stat	Reboot	"reboot the system"
	ras		Shutdown	"shutdown the system"
	security			
	system			
	wpar			

auth = aix.system.boot.create



Improved Program Security

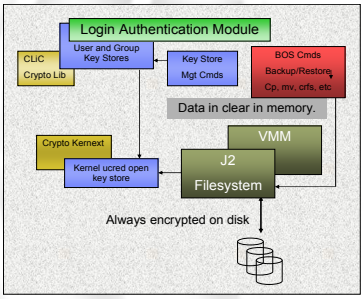
- Allows programs to do system level operations without running as root or having setuid root capability
- Only allow program to perform restricted set of needed operations



42

Mainline: solutions you need from people you trust

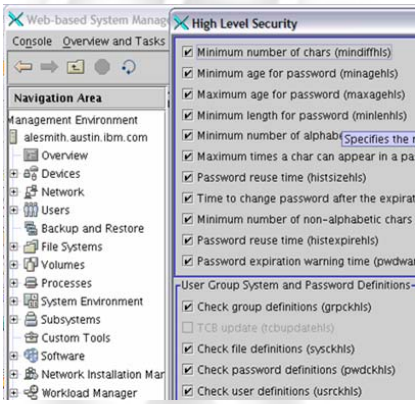
AIX 6.1 Encrypted File System



- Enables improved security by reducing unauthorized access to data, even by privileged users
- Secure backups reduces the exposure of data compromised when backup media is taken outside of secure facilities
- Automatic management of protection keys can reduce the administrative effort of using encrypted data

- The capability to automatically encrypt data in a JFS2 filesystem
- Data can be protected from access by privileged users
- Backup in encrypted or clear formats
- Automated key management - key store open on login, integrated into AIX security authentication
- Each file encrypted with a unique key
- No keys stored in clear in kernel memory
- A variety of AES, and RSA cryptography keys supported

AIX V6.1 Security Expert



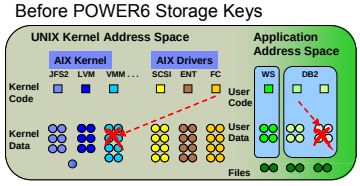
- A centralized security management tool that can control over 300 security settings from a single console
- Administrators can start from a “Low”, “Medium”, “High” or “Sarbanes-Oxley” security template and customize settings to met business requirements
- Security settings can be exported and imported as a security profile to multiple systems
- On AIX V6.1, security profiles can be stored in an LDAP directory for ease of distribution
- AIX Security Expert was first included in AIX V5.3 TL5

- Can reduce the cost and complexity of security administration by allowing federated management of security profiles across multiple servers
- Enables a more secure IT infrastructure by reducing the effort of maintaining system security
- “Check” functionality can provide additional security by validating that the security profile for each system matches the actual security settings

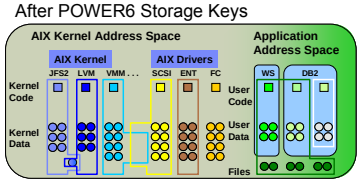
AIX V6.1 POWER6 Storage Keys

- Exploitation of a POWER6 processor hardware feature to provide additional isolation of kernel and application data
- Storage keys can prevent invalid changes to memory cause by programming errors
- Application use of POWER6 storage keys is enabled in AIX V5.3
- AIX kernel exploitation of POWER6 storage keys is included in AIX V6.1

Before POWER6 Storage Keys

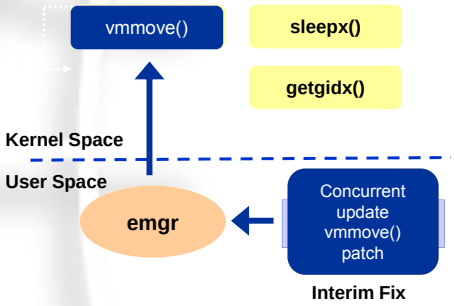


After POWER6 Storage Keys



Can provide for higher AIX availability by reducing the number of unplanned outages due to intermittent memory overlay

AIX 6.1 Concurrent Maintenance



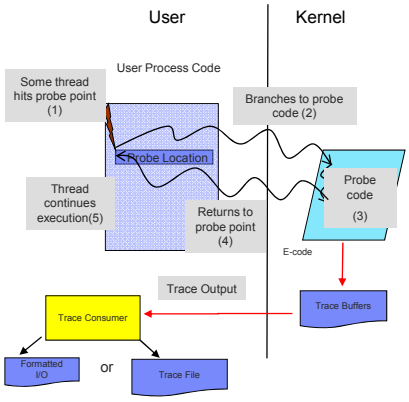
Fix selected AIX kernel problems without a service outage

Non-disruptive fixes to executable code in a running AIX kernel
- Base AIX Kernel (/unix), kernel extension, or device driver

No downtime (reboot) required to apply fix and make it active
Concurrent updates will be packaged as Interim Fixes
Maintenance can be backed off without an outage

AIX 6.1 Dynamic Tracing With *probevue*

- Dynamically extract information about a program as it is running
- Trace existing programs without recompiling
- Dynamic placement of trace probes without restarting the program
- For debugging and performance analysis
- AIX system calls, application functions, and application calls to library functions traceable
- Dynamic tracing language called Vue
- Initial support for "C" programs



"Vue" probe code example

```
•#!/usr/bin/probevue
/* countreads.v */

@@syscall.$1.read.entry
{
  count++;
}
@@interval.*.clock.100
{
  printf("Number of reads = %d\n", count);
  count = 0;
}
```

```
# countreads.v 404
Number of reads = 22
Number of reads = 0
Number of reads = 1
Number of reads = 17
.....
```

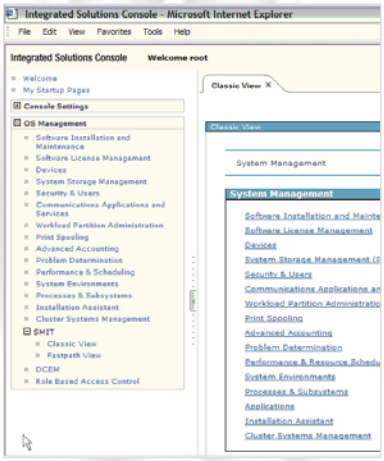

AIX V6.1 Non-intrusive Service Aids

▪A number of new reliability, availability and serviceability features that are designed to improve system and application reliability

- New features include:
  - Live Dump
  - Firmware Assisted Dump
  - Enhanced first failure data capture (FFDC) for AIX

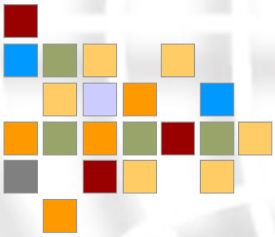
▪Enhanced features:
- Lightweight malloc debug
- Lightweight memory trace
- Consistency checkers
- Component trace

AIX V6.1 Systems Director Console for AIX



- New web based management tool that provides easy access to common system administration tasks
- Administrators can access Systems Management Interface Tool (SMIT) menus from a browser
- Graphical user interface is fast and consistent with IBM Systems Director look and feel
- All necessary components for the Console are included in AIX
- Distributed Command Execution Manager (DCEM) feature of the Console allows an administrative task to run on multiple systems at once

49



AIX Editions

50

AIX Editions…..

- **AIX 5.3 Management Edition bundle consisting of**
  - –AIX V5.3
  - –Tivoli® Application Dependency Discovery Manager
  - –IBM Tivoli Monitoring
  - –IBM Usage & Accounting Mgr Virtualization Edition for Power Systems

AIX 6.1 Enterprise Management bundle consisting of

- AIX V6.1
- **PowerVM AIX Workload Partitions Manager**
- Tivoli® Application Dependency Discovery Manager
- IBM Tivoli Monitoring
- IBM Usage & Accounting Mgr Virtualization Edition for Power Systems

51

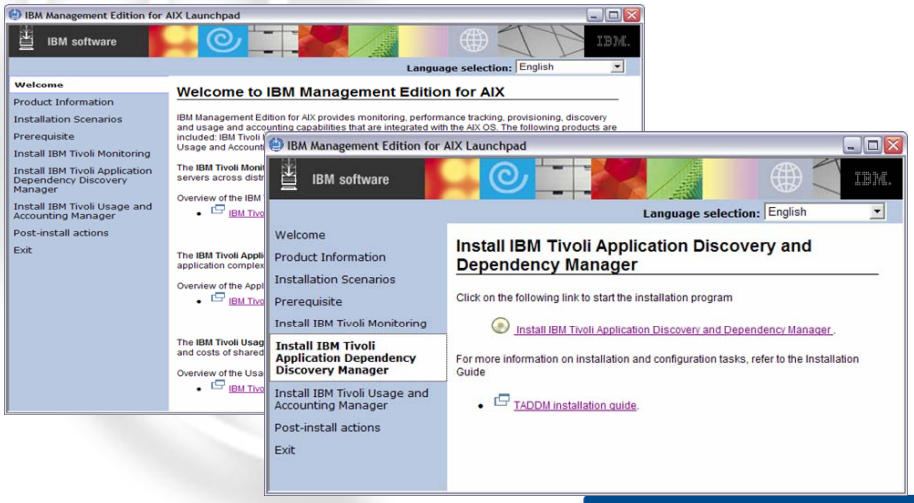
AIX Editions Solution Components Overview

- **Workload Manager (AIX 6.1 Enterprise Edition)**
- **Systems Management offering that provides:**
  - –Discovery
  - –Monitoring
  - –Performance Tracking and Usage Accounting
- **Integrated with the AIX® operating systems:**
  - –AIX 5.3 and AIX 6.1
  - –Provides tools to effectively and efficiently manage their enterprise IT infrastructure.
- **The following are the underlying products that are included:**
  - –IBM Tivoli Monitoring (ITM), version 6.2 (Virtualization enabled)
  - –IBM Tivoli Application Dependency Discovery Manager (TADDM) 7.1
  - –IBM Usage and Accounting Manager (IUAM) Virtualization Edition 7.1
- **Prerequisite: Limited license of DB2® 9.1**
  - – Included in the offering

52

ME for AIX Solution Components

Common Installation Console



AIX Editions Solution Components: Discovery

IBM Tivoli Application Dependency Discovery Manager (TADDM)

IBM Tivoli Application Dependency Discovery Manager initiates and assists planning for consolidation by providing best-of-breed discovery capabilities

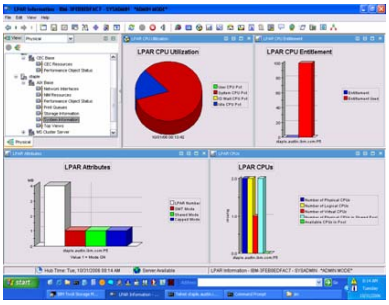
- Discovers the **COMPONENTS** in a Data Center Environment
- **CENTRALIZES** and **VISUALIZES** the **CONFIGURATION** of the Components in a Data Center Environment
- Discovers the **RELATIONSHIP** of the Components in a Data Center Environment
- **DISCOVERS AND TRACKS THE CHANGES** in a Data Center Environment

Can Feed this Information to other IBM Tivoli® Products

AIX Editions– Monitoring
IBM Tivoli Monitoring

•IBM Tivoli Monitoring helps prioritize consolidation decisions by visualizing the actual virtual server utilization against historical trends. It automates a clients best practices in response to system events

- Improves mean-time-to-recovery by **visualizing the virtual world** to solve "virtual performance problems"
- **Side-by-side real-time and historical** data assists in separating intermittent problems from reoccurring problems from peak workloads
- **Out-of-the-box reporting** allows clients to quickly provide executive level reports and identify resource bottlenecks

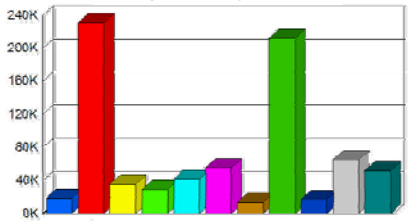


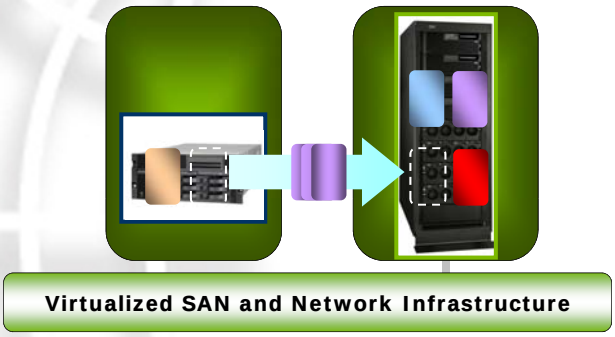
IBM Tivoli Monitoring

Usage
IBM Usage & Accounting Manager

- **AIX Editions Solution Component – Virtualization Edition**
- Apportion usage by account, department or organization
  - **Accountability and usage tracking ensures optimized usage by each department**
  - **Easily forecast growth by department to justify year-to-year budget changes**
- Single hardware system metrics and reports
- Data collectors
  - **AIX, Linux® and AIX Advanced Accounting –**
    - **Processor, server, LPAR, I/O, and VIO**
  - **OS File System – allocated and used**
- Usability – Power System tailored:
  - **Administration Console**
  - **JobRunner GUI**
- Reporting
  - **Business Intelligence Reporting Tool Reports**
  - **Reports will be provided, with aggregation by userid within a given server**
  - **"Pre-Defined" Accounting Schema**
  - **Export to spreadsheet, comma delimited, and CSB**

IT Expenses by Account





Live Partition Mobility

LPM

LPM Introduction

- Uses
  - Server Consolidation
  - Workload Balancing
  - Can use to allow power off of unused servers to assist in Green IT efforts
  - Planned maintenance and upgrades
- Inactive partition migration moves a powered-off partition
  - Does not require a mover service partition or VASI or TOD sync
- Partitions cannot be migrated from failed machines
- IT IS NOT A REPLACEMENT FOR HACMP OR OTHER HA

Requirements for LPM

- **PLANNING IS CRITICAL**
- http://www14.software.ibm.com/webapp/set2/sas/f/pm/component.html
- Hardware POWER6 Only
- HMC v7.3.2 with MH01062
- Firmware E*340_039 min
- AIX v5.3 5300-07-01
- AIX v6.1 6100-00-01
- VIO Server 1.5.2.1-FP-11.1 or v2.1
- Two Power6 systems managed by HMC or IVM (no mixing)
- PowerVM Enterprise Edition
- Virtualized SAN storage (rootvg and all other vgs)
- Virtualized Ethernet (SEA)
- LPAR being moved cannot be using the HEA/IVE (VIO can though)
- RHEL5 Update 1 and SLES10 Update 1 supported (or later)

LPM

- Check LPAR on HMC under Capabilities
  - Look for Active and Inactive Partition Mobility Capable=True
  - Ensure VIO server is set up as a Mover Service Partition (MSP) under the general tab on the VIO server
  - By default MSP is set to no on a VIO server
  - Mover partition must have a VASI (Virtual Asynchronous Services Interface) device defined and configured (done automatically by HMC)
- The pHypervisor will automatically manage migration of CPU and memory
- The target must have enough spare memory and cores so the workloads can migrate
- Dedicated IO adapters must be de-allocated before migration
- cd0 in VIO may not be attached to mobile LPAR as virtual optical device
- Time of Day clocks for VIO servers should be synchronized
- The operating system and applications must be migration-aware or migration-enabled
  - As of today Oracle 10G supports LPM

Preparation

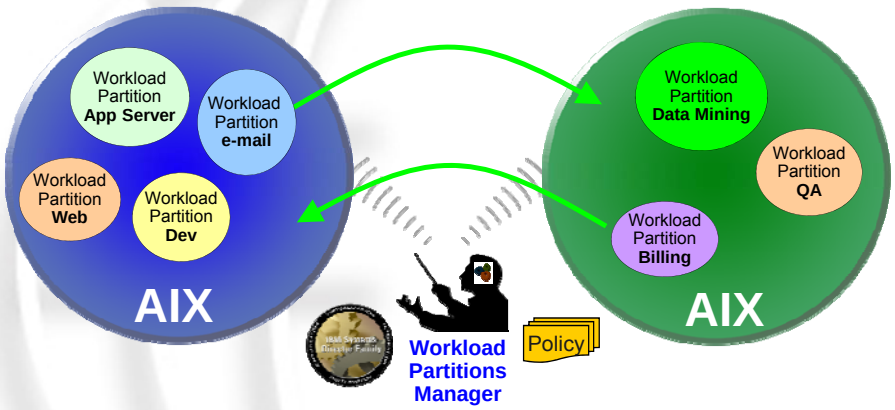
- External storage
  - SAN disks must be assigned to both source and target VIO servers
  - Hdisks must have reserve_policy=no_reserve
  - See section 3.7 of the LPM red book SG24-7460
- Network
  - Mobile partitions must use SEAs
  - See section 3.8

During Migration

- HMC configures MSPs (Mover Service Partitions)
- MSPs set up private, full-duplex channel to transfer state data
- HMC creates virtual target devices and adapters in destination MSP
- MSP on source starts sending partition state data
- Once memory pages, etc copied MSP tells PHYP to suspend the mobile partition
- Last few memory pages and state data are copied and execution is resumed on destination server
- Mobile partition recovers I/O on destination server and retries any uncompleted I/O

Effects of Migration

- Server Properties
  - Logical Memory Block affinity characteristics may change
    - Helps if you ensure LMB is the same on both servers (Modify via ASMI)
  - Maximum potential and installed physical cores may change
  - L1 and L2 cache size and association may change
- Console
  - Active console sessions are closed when the partition is migrated
  - Consoles sessions must be re-opened on target system after migration
- LPAR
  - Uname will change (new serial, etc)
  - Partition ID may change
  - IP address and MAC address stay the same
- May see a temporary network outage of seconds during suspend
- Necessary VSCSI controllers and target devices will be created on target VIO server ONLY if there are unused virtual slots



Workload Partitions

WPARS

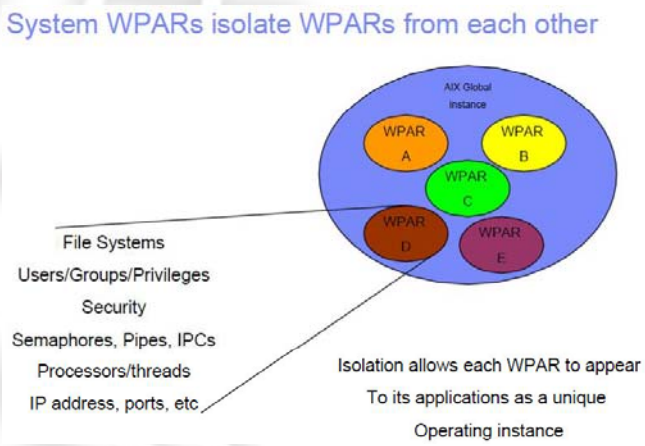
WPAR Intro

- Aim
  - Reduce the number of images to maintain
  - Reduce install and updating of AIX, backup, recovery, etc
  - Encapsulate applications as units of control
  - Rapid creation of new application environments
  - Reduces memory needs as only one copy of AIX
  - Mobility – for performance balancing as well as planned maintenance
- Requires
  - AIX v6 on Power4, Power5 or Power6 hardware
- Common operating system running a group of WPARs
  - Each WPAR gets a regulated share of processor and memory resources
  - Each WPAR has separate network and file systems
  - Each WPAR is a separate administrative and security domain
  - Shared resources are I/O devices, Processor, operating system and shared library and text
  - Allows for automatic, policy based relocation of workloads

System WPARs

- Complete virtualized OS environment
- Needs to be created and removed explicitly
  - Builds its own file systems
- Has its own root user, services, etc and can be stopped and started
- Allows you to delegate root authority without exposing other WPARs or global instance
- Does not share writable file systems with other WPARs or the global image
- Used for Namespace Isolation
- Integrated with AIX RBAC (granular security controls)
- Resource Management

System WPARs



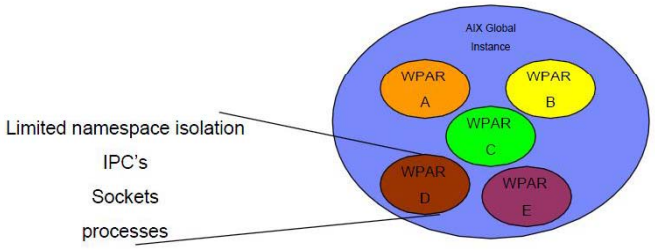
Source: PAI10 Steve Nasypany, 2008 IBM Power Technical University

Application WPARs

- Used to isolate an individual application
- Lightweight
  - One process but can start further processes
  - Basically a wrapper around processes associated with an application
- Starts when process starts and ends when process ends
- Shares the global file system
- Good for HPC long running applications
- Resource Management
- No separate security context or root level user
- Has it's own IP address

Application WPARs

Application WPARs work as a way to manage the resources of a set of processes



Very different than system WPARs

An application WPAR is collection of processes under a virtualization layer

Manage resources as a unit with resource controls

Live Application Mobility

Source: PAI10 Steve Nasypany, 2008 IBM Power Technical University

LAM (Live Application Mobility)

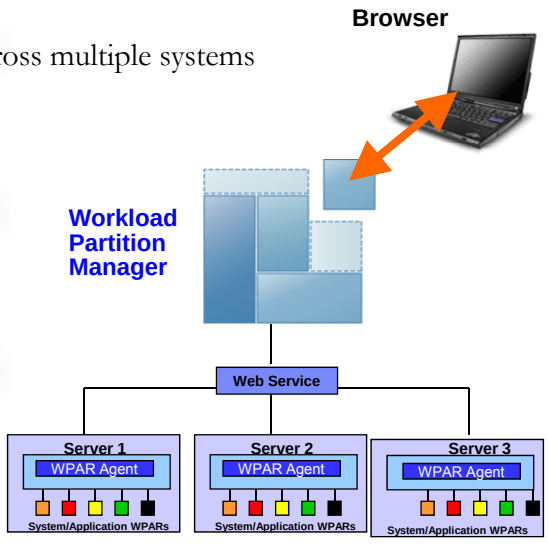
- Move a running WPAR to another machine or LPAR
- Use to install new machine
- Use for multi-system workload balancing
- Use to empty a machine for green computing, firmware or AIX upgrades or hardware maintenance

Notes

- No per-WPAR tuning options
- WPARs are built on top of AIX WLM
  - Around since AIX v4.3.3
- CPU managed 3 ways
  - Percentage, share and RSET (resource sets of virtual processors)
- All WPARs share processor, memory and I/O resources from the Global Instance
  - No devices directly attached to a WPAR
- Commands updated
  - iostat, vmstat, wlmstat, topas
  - Tivoli ITM can monitor WPARs

Workload Partitions Manager

- Management of WPARS across multiple systems
- Lifecycle operations
- Single Console for:
  - Graphical Interface
  - Create & Remove
  - Start & Stop
  - Checkpoint & Restart
  - Monitoring & Reporting
  - Manual Relocation
  - Automated Relocation
  - Policy Driven Change
- Infrastructure Optimization
- Load Balancing



Introduction

ACTIVE MEMORY SHARING

What is Active Memory Sharing ?

- Pools of memory that can be shared by partitions
  - Similar to shared processor partitions
  - Pool of processor resources
- OS Support
  - AIX 6.1, IBM i and Linux
- Features
  - Allows for the dynamic sharing of memory
  - Provides the ability to "Over-Commit" physical memory
  - Overflow of memory request paged to system disk.
  - Fine-grained sharing of physical memory
  - Automated ballooning (expansion and contraction) of a partition's physical memory footprint based on workload demands.
  - Sharing of common code pages between partitions (Future)
  - Reduces the memory and cache footprints
  - Partitions with the same OS and application code.

Active Memory Sharing Overview

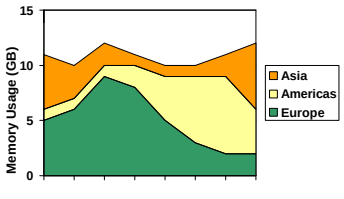
• *Active Memory Sharing intelligently flows memory from one partition to another for increased utilization and flexibility of memory usage*

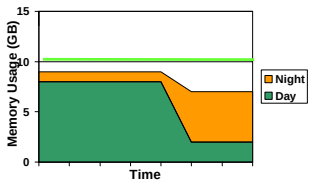
- Memory virtualization enhancement for Power Systems
  - Memory dynamically allocated based on partition's workload demands
  - Contents of memory written to a paging device
  - Improves memory utilization
- Designed for partitions with "Variable Memory" requirements
  - Low average memory requirements
  - Active / Inactive environments
  - Workloads that peak at different times across partitions
- Available with PowerVM Enterprise Edition
  - AIX 6.1, Linux, and IBM i 6.1 partitions that use VIOS and shared processors
  - POWER6 processor-based systems
  - Must use Shared Processor and have Virtual IO (VIOS managed)

Dynamically optimize memory utilization across virtual images

PowerVM Active Memory Sharing

- Dynamically adjusts memory available on a physical system for multiple virtual images based on their workload activity levels:
  - Different workload peaks due to time zones
  - Mixed workloads with different time of day peaks (e.g. CRM by day, batch at night)
  - Ideal for highly-consolidated workloads with low or sporadic memory requirements
- Available with PowerVM Enterprise Edition
  - Supports AIX, IBM i and Linux workloads
- Blends Power Systems hardware, firmware and software enhancements to optimize resources
  - Supports over-commitment of logical memory
  - Overflow managed by VIOS paging device
  - Two VIOS partitions can be used for redundancy
  - Compatible with Live Partition Mobility



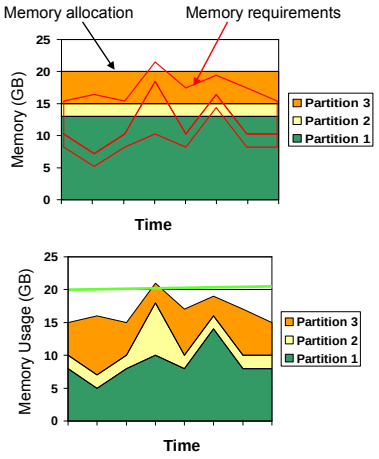


Innovative solution for more efficient utilization of memory resources

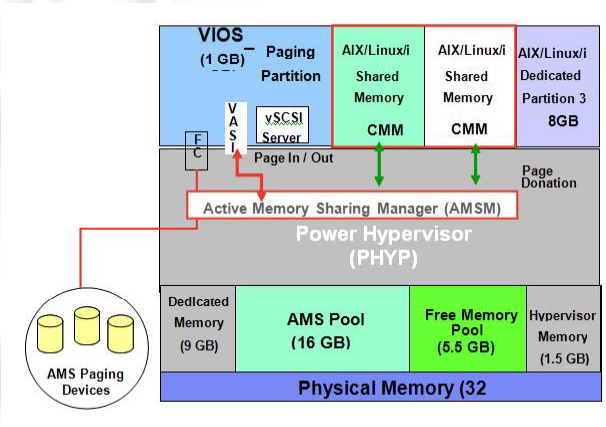
PowerVM Active Memory Sharing

Partitions with dedicated memory
- Memory is allocated to partitions
- As workload demands change, memory remains dedicated
- Memory allocation is not optimized to workload

Partitions with shared memory
- Memory is allocated to shared pool
- Memory is used by partition that needs it enabling more throughput
- Higher memory utilization



Active Memory Sharing



Definitions

- **Physical Over-commit:**
  - A shared memory pool is overcommitted when the combined resident memory of all shared memory partitions exceeds the physical memory in the pool. In this situation, some of the shared partitions' actively-used memory will need to be resident on a paging device assigned to the shared memory pool.
- **Logical Over-commit:**
  - A shared memory pool is said to be logically overcommitted when the actively referenced memory pages of all shared memory partitions can be backed by the physical memory allocated in the AMS Shared Memory Pool, but the total logical memory exceeds the amount of physical memory in the shared memory pool. In such an environment, the partitions' actively used memory will reside in physical memory, with the remainder of the partitions' logical memory residing on a paging device assigned to the shared memory pool.

From Mala Anand, Active Memory Sharing Performance, May 20, 2009

AMS Useage

- AMS physical overcommit is appropriate for workloads that have the following characteristics:
  - Workloads that currently run on the AIX operating system and use AIX file cache. AIX operating system tends to use all the available free memory for file cache, in such a situation, AMS CMM will release these unreferenced file pages, which can be used for other shared memory partitions.
  - Workloads that are less sensitive to I/O latency such as file servers, print servers, and network applications.
  - Partitions that are inactive most of the time.
-
- AMS logical overcommit is favored for workloads that have the following characteristics:
  - Workloads that time multiplex, e.g., AM/PM scenarios, peaks and valleys of multiple workloads overlap leading to logical overcommit levels.
  - Workloads that have low average memory residency requirements. Consolidating them would lead to logical overcommit.
  - Workloads that do not have sustained load such as retail headquarters and university environments.
  - Fail over and backup server partitions if configured on the same server as its primary server partitions then it could share resources since these fail-over servers will need resources only when the primary servers go down. Resources do not have to be dedicated to these redundant servers.
  - Test and development environments are also suitable.
-
- It is recommended that dedicated memory partitions be used for the following workloads:
  - Workloads that have high quality of service criteria.
  - Workloads that have high sustained memory consumption due to sustained peak load.
  - Workloads that mandate predictable performance.
  - Workloads with high degree of frequent load changes.
  - High Performance Computing workloads such as scientific computation-intensive workloads which tend to exhibit sustained high CPU utilization.
  - If the above category workloads are deployed in AMS environment, ensure memory over subscription is minimal or low.

From Mala Anand, Active Memory Sharing Performance, May 20, 2009

Security in AIX v6

SECURITY

AIX v6 Security 1/3

- Trusted Execution
  - trustchk command compares AIX binaries to the TSD (trusted signature database) which includes all AIX system files
  - Can add other files
- Secure by Default (SBD)
  - Enabled at install
  - Installs minimal set of packages (about 100)
  - Enforces AIX security expert level of high
    - Ensure you change root password just before doing this as password aging is enforced
  - If you need telnet, ftp or r* then this is not for you

AIX v6 Security 2/3

- Trusted AIX
  - Enabled at install
  - Removes concept of root
  - Mandatory access controls and auditing
  - Provides mandatory integrity labels for data
- RBAC (Role Based Access Control)
  - Allocates specific roles to users and negates need for sudo
  - Privileges can be set for processes, files and devices
  - User must swrole to get any privilege
  - Provides tiered security levels
  - 3 key elements – authorizations, roles and privileges
  - Info can be stored in an LDAP database

AIX v6 Security 3/3

- File Permission Manager (FPM)
  - Reduces programs with setuid bits
  - Can be run at various levels to remove setuid bits or to set back to default
- Encrypted File system (EFS)
  - Only available for JFS2
  - New file system type of efs
  - File system is encrypted on a per file basis
  - Requires CLiC libraries
  - Uses keys and key stores
- Secure FTP
  - Encrypts both the data and command channels
  - Built on OpenSSL
- Updates to AIX Security Expert

System Hardening

- Traditionally done by scripts, commands or packages
- SOX – Sarbanes Oxley Act of 2002
- COBIT – Control Objectives for Information related Technology
- AIX Security Expert is IBM's implementation of those recommendations for AIX
  - Available as of AIX v5.3 tl05 and enhanced in v6
- 5 ways to access it
  - smitty, pconsole, Director 6, Websm, command line
  - aixpert -?

AIX Security Expert Levels

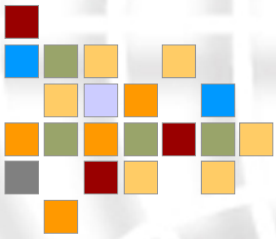
High Security
- Direct internet running web server with important data
- Banned are Telnet, FTP, rlogin
- Start services and then go High Security – in use ports not blocked

Medium Security
- Corporate network Firewall protected
- Telnet, FTP are in use
- Wants port scanning and user account protection

Low Security
- Been running for a long time on isolated secure network
- Need to keep all services available

Default
- As comes with AIX standard install

SOX-COBIT
- The setting recommended for compliance



Linux

Linux on POWER*

- Enterprise Class Computing with Linux on POWER (LoP)
  - Flexibility with LPAR, dynamic LPAR, and virtualization features
  - Reliability with built in self-healing capabilities
  - Power Architecture/Servers include POWER5, POWER6 and PowerPC970 (JS20, JS21 blades)
- Linux distributions available for LoP:
  - SUSE LINUX Enterprise Server 9 or 10 for POWER (SLES 9, SLES 10)
  - Red Hat Enterprise Linux 4, 5 for POWER (RHEL)
- Technical support available through IBM SupportLine contract.
  - SUSE LINUX and Red Hat, Inc. also provide support, upgrades and maintenance
- Orderable from IBM or directly from Linux distributors
- For more information about Linux running on IBM Power Systems servers:
  - http://www.ibm.com/systems/p/linux/
  - http://www.redhat.com/rhel/server/
  - http://www.novell.com/products/server/

References

- **IBM Movies**
  - http://www.ibm.com/developerworks/wikis/display/WikiPtype/Movies
- **Active Memory Sharing RedBook**
  - http://www.redbooks.ibm.com/redpapers/pdfs/redp4470.pdf
- **PowerVM Best Practices**
  - http://www.redbooks.ibm.com/redpapers/pdfs/redp4194.pdf
- **PowerVM Virtualization - Introduction and Configuration**
  - http://www.redbooks.ibm.com/redbooks/pdfs/sg247940.pdf
- **PowerVM Live Partition Mobility**
  - http://www.redbooks.ibm.com/redbooks/pdfs/sg247460.pdf

Questions???