IBM CHAMPION

Belisama

Session: a111038

AIX / Linux on Power “Good” practice

Antony Steel

Power Systems

#include <std_disclaimer.h>

These notes have been prepared by an Australian, so beware of unusual spelling and pronunciation.

All comments regarding futures are probably nothing more than the imagination of the speaker and are IBM Confidential till after GA.

Abstract

Belisama

- An introduction to some guiding principles to maintaining your AIX / Linux environments on Power. New features of AIX and Linux on Power will also be covered
- Good practice – recommendations made in a vacuum for the ideal world
- Best practice – What can be implemented / is appropriate for each customer

© 2020 Belisama

Thanks to ….

Belisama

- Susan Schreitmueller
- Regina Moliff
- Steve Pittman
- Peter Nutt
- Bruce Spencer
- Mark Dixon
- Jerry Petru
- David Sinnot
- Shaun Mullen
- Alex Abderrazag
- Grover Davidson
- Nigel Griffiths
- Cesar D Maciel
- Maneesh Sharma
- Shawn Bodily
- Daryl Scott
- Ken Fleck
- Michael Sieber
- Chris Smith
- Scott (Tex) Nance
- Chris Gibson

Their ideas and suggestions made this presentation possible, but errors are all mine!

© 2020 Belisama

Agenda

Belisama

- What do Ineed if I have come back to Power?
- System maintenance refresher
- What does Power9 add
- What does AIX 7 bring
- Good AIX habits
- Monitoring Management tools
- AME (and AMS)
- NPIV / iSCSI / PowerVM
- AIX security expert
- AIX runtime expert
- WPAR - Kubernetees
- Conclusion / References

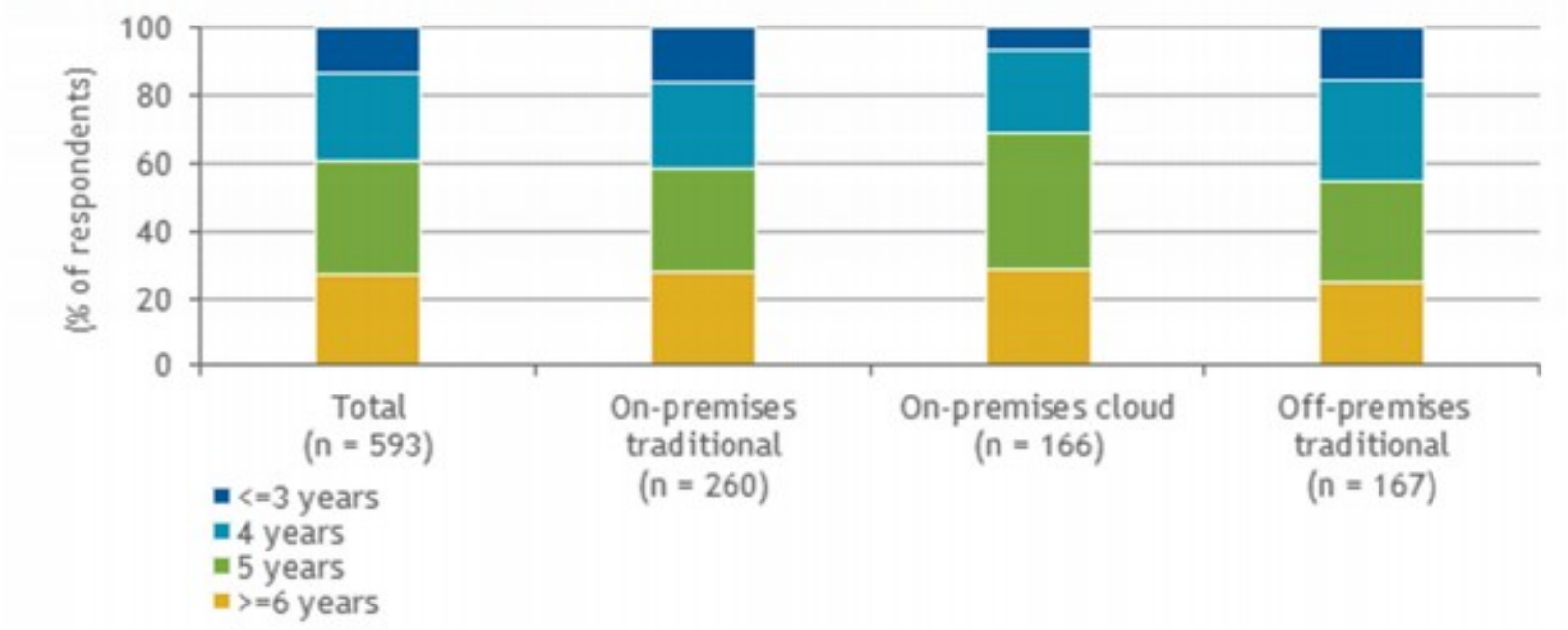
© 2020 Belisama

IDC PERSPECTIVE - Server Upgrade Cycles: Why Faster Is Better

Belisama

Server Life Cycle by Segment, 2016

Q. Approximately how many years do you estimate that your servers will be kept in operation?



Note: Mean life cycle = 5.15 years

Source: IDC's *Server and Storage Workloads Survey: Understanding 3rd Platform Usage*, May 2017

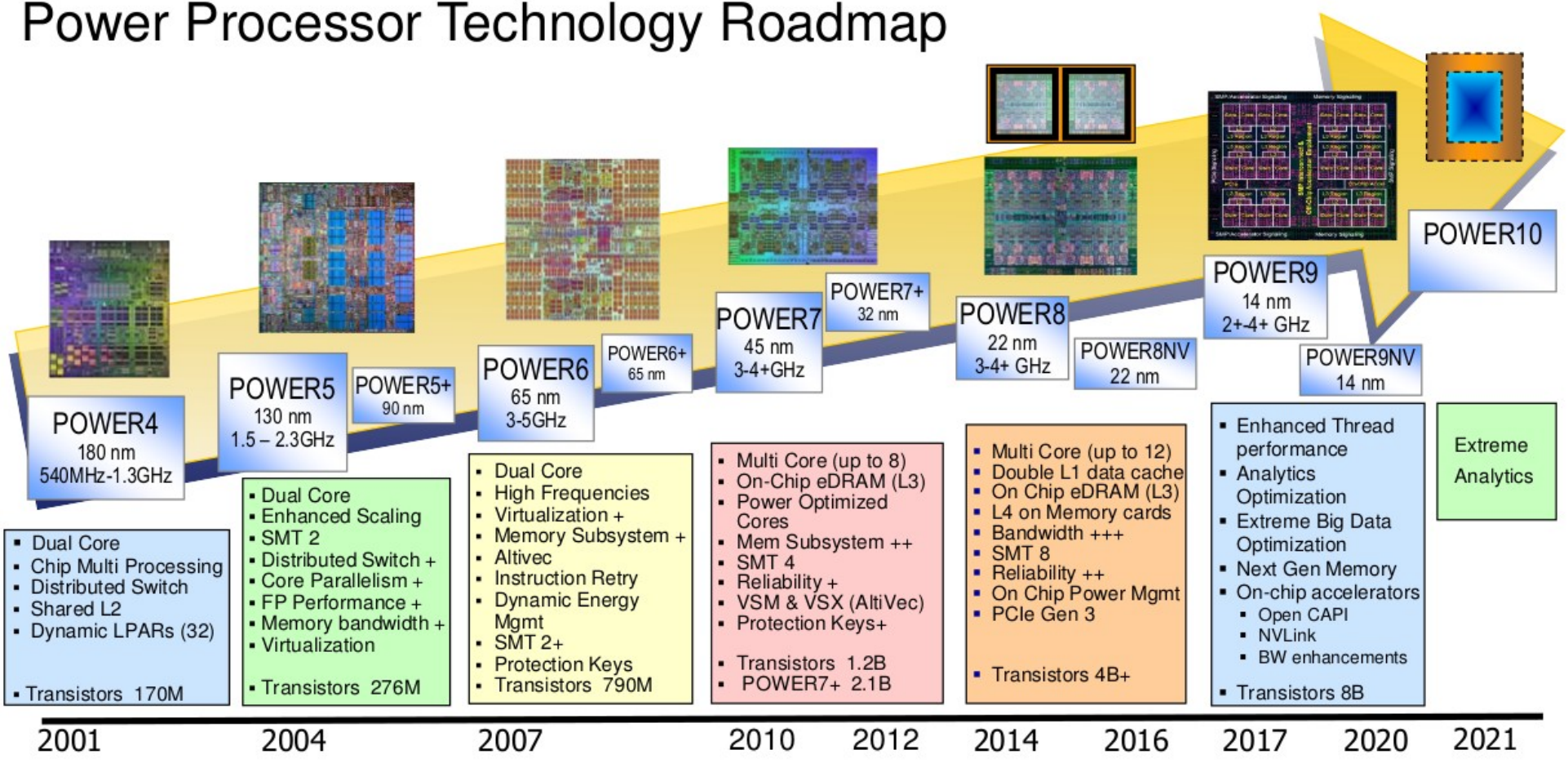
IDC PERSPECTIVE – Upgrade cycles

Belisama

- IDC Research finds organisations achieve broad benefits from upgrading servers in a regular cycle
- Organisations that upgrade servers more frequently can substantially reduce server-related costs, including server costs and maintenance, power, facilities and support costs more frequent server upgrades were also found to produce application performance improvements (eg reduce batch processing time)
- IT staff requirements and costs were significantly reduced through server upgrades enabling IT to focus resources on higher-value tasks

© 2020 Belisama

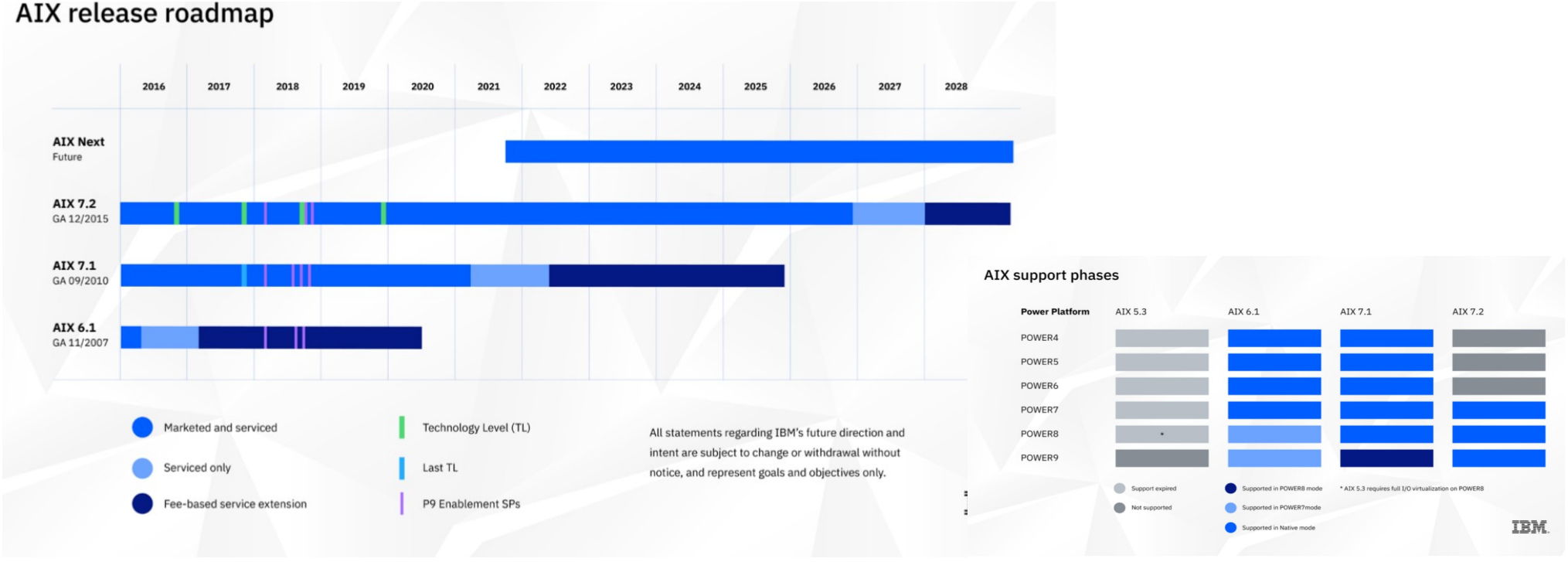
Power Processor Technology Roadmap



© 2020 Belisama

AIX?

- IBM has just released a white paper on the future of AIX.
- Rich sotfware stack:
  - PowerVM; PowerVC; PowerSC; PowerSC MFA; PowerHA; PowerHA / VRM DR
- Roadmap to 2030



IBM Best practice resources

Belisama

- IBM Best practices
  - https://www14.software.ibm.com/webapp/set2/sas/f/best/home.html
- Developer works
  - PowerVM: https://ibm.biz/powervmwiki
  - AIX: https://ibm.biz/aixwiki
  - PowerHA: https://ibm.biz/powerhawiki
  - PowerSC: https://ibm.biz/smcpowervc

Description	Instructions
Ensure firmware is current	Fix Central provides latest updates. Latest F/W levels as of this writing : FW910 for POWER9 models S914, S922 and S924 Use the FLRT tool to obtain the recommended levels for a given platform. **NOTE**: Ensure required HMC level is installed when updating F/W.
Memory DIMMs	For optimal performance on workloads that are memory bandwidth sensitive follow these recommendations: **S914/S922/S924**: • Assign minimum 4 DIMMs per socket • DIMMs on same memory channel must have the same size **All POWER** : Follow proper memory plug-in rules
Ensure OS level is current	Fix Central provides the latest updates for AIX, IBM i, VIOS, Linux, HMC and F/W. In addition to that, the FLRT tool provides the **recommended levels** for each H/W model. Use these tools to maintain your system up to date.
SMT8	In order to take full advantage of the improved performance of the POWER9 CPU, we recommend customers evaluate the use of SMT8 in their environment. Although a change between SMT modes is dynamic (via smtctl), we recommend when moving to SMT8 to reboot the given partition to get the best performance of this change.
40GbE adapter	RHEL7: For network bandwidth sensitive workloads, we recommend increase the receive queue size from 1024 to 8192.
Sizing a system	• When migrating to POWER9, we recommend considering using SMT8, and size the LPARs based on the SMT8 rPerf values; in many instances, this will likely reduce the number of VPs required. • Use Workload Estimator (WLE) for sizing LPARs for CPU consumption as it provides better sizing results.
Right-size your Shared LPARs	• Assign entitled capacity (EC) to sustained peak utilization for LPARs with critical SLA requirements • Assign EC to average utilization and number of virtual CPUs to peak utilization(physical core consumption) for LPARs with non-critical SLA • Ensure the average LPAR utilization is equal or less than 75% of the entitled capacity
Java	• IBM JDK8 SR5 is the minimum level to exploit POWER9 • Open JDK 1.8 provides partial support for P9 ISA • Use of 64k size pages normally increases application performance
Partition Placement	Current FW levels ensure optimal placement of the partitions. However, if constant DLPAR operations are executed on partitions on the CEC, it is recommended the use DPO to optimize placement.

Description	Instructions
Compilers	• IBM xlC for Linux : 13.1.5 & 15.1.6 support for P9 ISA • Advanced Toolchain : 11.0-3 and later • gcc: Version 7 of gcc is recommended for P9 ISA support. Also includes support for "-mtune=power9"
IBMi	Ensure Technology Updates are current (see link below)
AIX Tunables/ VIOS Tunables	• Tuning a VIOS is not recommended unless directed by VIOS/AIX support. • **Restricted** tunables **should not** be modified (unless directed by AIX/VIOS development) • Tunables **should not** be migrated across AIX levels.
AIX CPU utilization	The AIX OS system is optimized for best raw throughput at higher CPU usage. If the customer requires to reduce CPU usage (pc), use the **schedo** tunable vpm_throughput_mode to tune the workload and evaluate the benefits of raw throughput vs. CPU usage.
VIOS configuration	• If configured with shared processors : ➢ Assign total entitlement of all VIOS partitions to be 10-15% of cores in shared pool and assign CPU ratio of 2:1 (vCPUs:ec). Refer to the PowerVM Best Practices for additional recommendations ➢ Assign uncapped mode and set variable weight capacity of VIOS partition higher than all client LPARs serviced by VIOS • For performance and flexibility, it is recommended to use IBM i to virtualize internal storage to IBM i. If you must use VIOS, follow the wiki at the following link. • For vFC, ensure no more than 64 client connections total per physical fcs adapter on the VIOS. Also, ensure no more than 64 storage ports configured per vFC adapter on the client. These are physical limits; practical limits may differ based on workload. • For vSCSI disks, ensure the queue_depth for virtual disks is less than or equal the queue_depth of the physical disk in the VIOS. • For vSCSI adapters, ensure you configure VTDs based on the following formula: Max VTDs = (512 -2) / (virtual_q_depth + 3) • Only enable the largesend attribute on the SEA (physical adapter backing the SEA) if all LPARs serviced by the VIOS are AIX partitions.
Virtual Ethernet adapters on AIX	• Increase the virtual Ethernet (vETH) device driver buffers if the partition is dropping packets on the virtual interface even when running with entitled CPU capacity. e.g., chdev –l ent# -a max_buf_xxx=NNNN **NOTE**: For desired buffer size adjustments, refer to "AIX on Power – Performance FAQ" link below • Set largesend on vETH adapter to improve performance (AIX): chdev –l en# -a mtu_bypass=on (or) ifconfig en# largesend

© 2020 Belisama

Did you know?

Belisama

- Security iFixes
  - Security Alerts which publish iFixes to patch the vulnerability will now include iFixes for the most recent 3 Service Packs (the latest SP, n-1, and n-2) on each active Technology Level, whenever possible.

Live Kernel update

Belisama

- Updates and the need for rebooting
  - For a while now AIX, SuSE and RHEL have supported live application of hot fixes (concurrent updates)
  - With AIX 7.2 we have Live Kernel Update, which aims to achieve zero downtime while updating OS patches without the need to disrupt business-critical workloads. This can save cost and reduce risk not having to wait for downtime to install critical security patches.
  - Starting with AIX 7.2, the AIX OS has a built in live update function. Previous versions required a reboot after an OS patch was applied to a running system.
  - Steps
    1) If updates to a service pack or technology level are specified to be installed by using the Live Update function, the updates are applied and committed first on the original partition.
    2) If any interim fixes are specified along with the service pack and technology level updates, the interim fixes are installed on the original partition.
    3) The root volume group of the original partition (orig-rootvg) is cloned.
    4) If only interim fixes are specified for the Live Update operation, the interim fixes are applied on the cloned volume group that serves as the boot volume group for the surrogate partition (surr-boot-rootvg).
    5) After the surrogate partition is started and while the workloads are still running on the original partition, the root volume group of the surrogate partition is mirrored (surr-mir-rootvg).
    6) The workload processes are check pointed and moved to the surrogate partition.
    7) Workloads resume on the surrogate partition in a chrooted environment (changed root directory) on the original root volume group (orig-rootvg). During this process, the workloads continue to run without being stopped, although a short blackout time occurs when these workloads are suspended.

© 2020 Belisama

Live Kernel update (cont)

- If the process fails, it can be restarted after issue resolved.
- Limitations
  - Fully virtualised environment
  - Live update is planned for fixes that contain changes to the kernel or kernel extensions. However it may contain other changes to commands / libraries etc, and the Live update will not make any change here. For example a shared library may be changed in the file system, but running processes will continue to use the original version of the library. In this case, applications must be stopped and restarted. Since AIX 7.2 TL1 you can use the *genld -u* command to list the processes that are using an old version of a shared library or other object that hasn't been updated.
  - Below AIX 7.2 TL2 you need spare CPU / Memory, same or greater than the LPAR being updated.
  - After 7.2 TL2 you can move to a standby machine.
- Process can be launched through *geninstall -k* or via NIM.

https://www.ibm.com/support/knowledgecenter/ssw_aix_72/install/live_update_install.html

© 2020 Belisama

- Separate presentation on some specific examples around physical migrations
- AIX migrations
  - Perform via media (optical, usb, VIO media library...) or NIM
    - Presented with options New and complete overwrite; Preservation Install; Migration Install
  - But with NIM can dramatically reduce downtime (using *nimadm*)* and obviously must run on a NIM Client. The process does the following:
    - Create a copy of the rootvg on a spare client LUN (similar to alt disk install *alt_disk_copy*)
    - Migrate the newly created copy of rootvg to the new version of AIX, installing additional filesets as required – While the system is still running on the current version of AIX, ie no disruption
    - At convenient time schedule a reboot of the client – choosing the new rootvg as the target.
  - The advantages are:
    - Minimum disruption on the client (just a reboot) – no outage for the migration
    - As the process of the migration is run from NIM, it primarily uses NIM resources (and some network bandwidth)
    - Easy back-out and debugging. Can roll back to the original copy of the rootvg and investigate the updated rootvg to resolve problem – In a fully virtualised environment can even assign the rootvg to another LPAR!

* Used for customer in Singapore AIX 5.3 on old h/w to POWER8

© 2020 Belisama

Changes in LVM – preferred read

Belisama

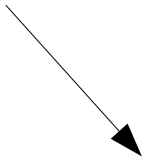
```
# lslv -L ulv32
LOGICAL VOLUME:     ulv32                  VOLUME GROUP:   mirrorvg
LV IDENTIFIER:      00fa77c000004c000000016874a96cb3.2 PERMISSION:     read/write
VG STATE:           active/complete        LV STATE:       opened/syncd
TYPE:               jfs2                   WRITE VERIFY:   off
MAX LPs:            512                    PP SIZE:        4 megabyte(s)
COPIES:             2                      SCHED POLICY:   parallel
LPs:                20                     PPs:            40
STALE PPs:          0                      BB POLICY:      relocatable
INTER-POLICY:       minimum                RELOCATABLE:    yes
INTRA-POLICY:       middle                 UPPER BOUND:    1
MOUNT POINT:        /data2                 LABEL:          /data2
DEVICE UID:         0                      DEVICE GID:     0
DEVICE PERMISSIONS: 432
MIRROR WRITE CONSISTENCY: on/ACTIVE
EACH LP COPY ON A SEPARATE PV ?: yes (superstrict)
Serialize IO ?:     NO
INFINITE RETRY:     no                     PREFERRED READ: 0
DEVICESUBTYPE:      DS_LVZ
COPY 1 MIRROR POOL: flash
COPY 2 MIRROR POOL: spin_rust
COPY 3 MIRROR POOL: None
```

- The parallel policy balances reads between the disks. If primary not busy, reads, if busy, checks secondary, if okay reads, else chooses shortest I/O queue. Writes concurrent
- The parallel/sequential policy always initiates reads on the primary copy. Writes are initiated concurrently.
- The parallel/round robin policy is similar to the parallel policy except it alternates between the copies. Writes are initiated concurrently.
- The sequential policy results in all reads being issued to the primary copy. Writes happen serially, and initiated when last write completed

Changes in LVM – preferred readl

Belisama

```
#chlv -R1 ulv32
```



```
# lslv -L ulv32
LOGICAL VOLUME:     ulv32                  VOLUME GROUP:   mirrorvg
LV IDENTIFIER:      00fa77c000004c000000016874a96cb3.2 PERMISSION:     read/write
VG STATE:           active/complete        LV STATE:       opened/syncd
TYPE:               jfs2                   WRITE VERIFY:   off
MAX LPs:            512                    PP SIZE:        4 megabyte(s)
COPIES:             2                      SCHED POLICY:   parallel
LPs:                20                     PPs:            40
STALE PPs:          0                      BB POLICY:      relocatable
INTER-POLICY:       minimum                RELOCATABLE:    yes
INTRA-POLICY:       middle                 UPPER BOUND:    1
MOUNT POINT:        /data2                 LABEL:          /data2
DEVICE UID:         0                      DEVICE GID:     0
DEVICE PERMISSIONS: 432
MIRROR WRITE CONSISTENCY: on/ACTIVE
EACH LP COPY ON A SEPARATE PV ?: yes (superstrict)
Serialize IO ?:     NO
INFINITE RETRY:     no                     PREFERRED READ: 1
DEVICESUBTYPE:      DS_LVZ
COPY 1 MIRROR POOL: flash
COPY 2 MIRROR POOL: spin_rust
COPY 3 MIRROR POOL: None
```

OpenPOWER Foundation

- The OpenPOWER Foundation is a collaboration around Power ISA-based products initiated by IBM and announced as the "OpenPOWER Consortium" on August 6, 2013.[1] IBM is opening up technology surrounding their Power Architecture offerings, such as processor specifications, firmware and software with a liberal license, and will be using a collaborative development model with their partners.[2][3]
- The goal stated is to enable the server vendor ecosystem to build their own customized server, networking and storage hardware for future data centers and cloud computing.[4]
- Power.org is still the governing body around the Power ISA instruction set but specific implementations are now free to use under a liberal license granted by IBM. Processors based on IBM's IP can now be fabricated on any foundry and mixed with other hardware products of the integrator's choice.

OpenPOWER Consortium
From Wikipedia, the free encyclopedia

© 2020 Belisama

- System tuning
  - https://developer.ibm.com/linuxonpower/docs/linux-on-power-system-tuning/
- Linux on Power Application tuning
  - https://developer.ibm.com/linuxonpower/docs/linux-on-power-application-tuning/

© 2020 Belisama

Linux support on Power

Belisama

Hardware	RHEL	SuSE	Ubuntu
POWER6	RHEL 4.5 and above* RHEL 5.1 and above RHEL 6 and above	SuSE LES 10 SP1 and above* SuSE LES 10 SP2 and above SUSE LES 11 and above	
POWER7	RHEL 5.7 and above* RHEL 6.2 and above RHEL7 and above	SuSE LES 10 SP4 and above* SUSE LES 11 SP1 and above	
POWER7+	RHEL 6.4 and above RHEL7 and above	SUSE LES 11 SP2 and above	
POWER8 (PowerVM)	LE: RHEL 8.0 for POWER8 RHEL 7.2 and above BE: RHEL 6.8 and above RHEL 7.2,and above	LE: SUSE LES 12 SP1 and above BE: SUSE LES 11 SP4	LE: Ubuntu 16.04.1, and above
POWER8 (KVM)	LE: RHEL 8.0 for POWER8 RHEL 7.2 and above Ubuntu KVM		LE: Ubuntu 14.04.5, and above Ubuntu 16.04.1, and above Ubuntu 18.04, and above
POWER8 (Bare Metal)	LE: RHEL 8.0 for POWER8 RHEL 7.2 and above		LE: Ubuntu 14.04.2 and above Ubuntu 16.04 and above Ubuntu 18.04 and above
POWER8 (Nutanix)		LE: SUSE LES 12 SP3, and above BE: SUSE LES 11 SP4	LE: Ubuntu 16.04

https://www.ibm.com/support/knowledgecenter/linuxonibm/liaam/liaamdistrospower7.html

* Not optimised

Linux support on Power (cont)

Belisama

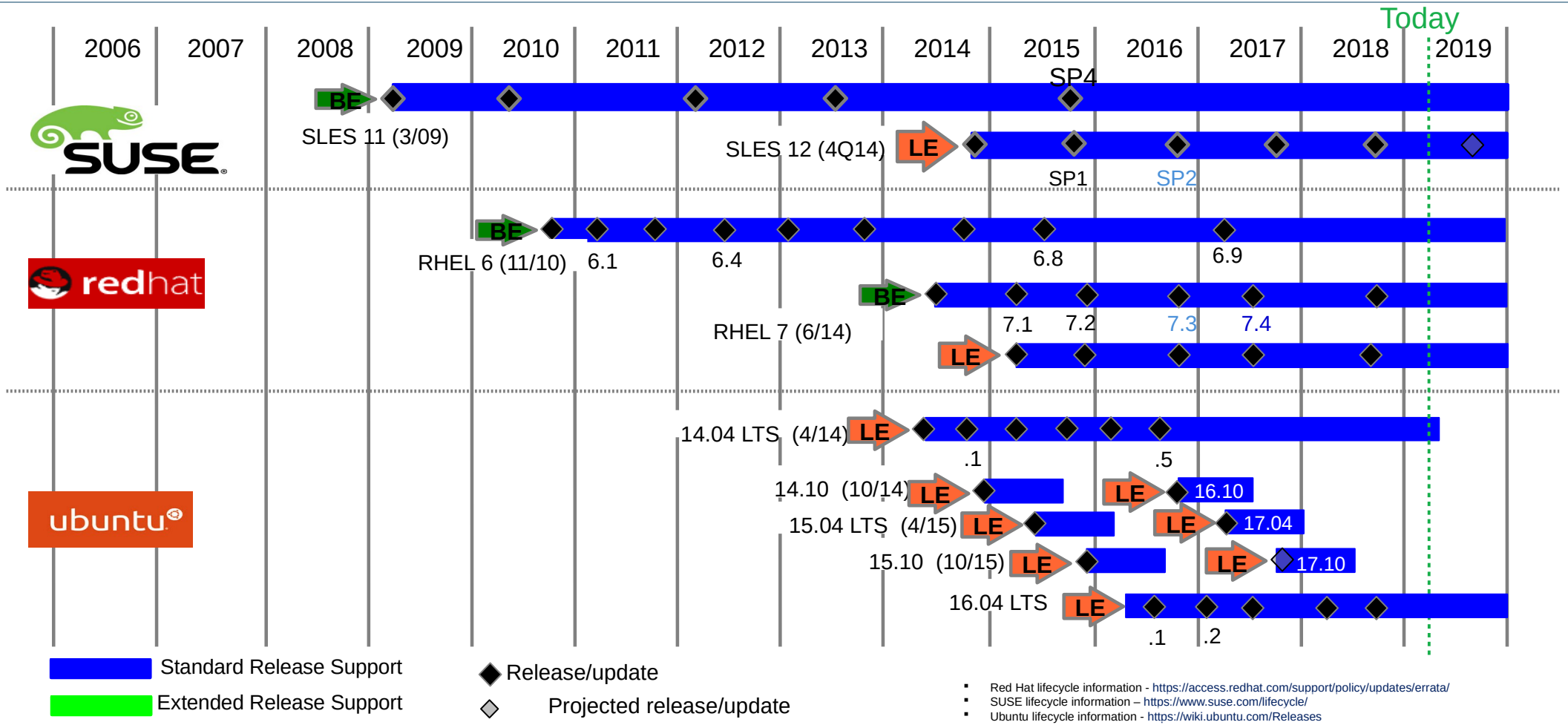
Hardware	RHEL	SuSE	Ubuntu
POWER9 (PowerVM)	LE: RHEL 8.0 for POWER9 RHEL 7.5 for POWER9 and above RHEL 7.4 and above (POWER8 Compatibility mode only)	LE: SUSE LES 15 and above SUSE LES 12 SP3 and above BE: SUSE LES 11 SP4 (POWER8 Compatibility mode only)	LE: Ubuntu 16.04.4 (POWER8 Compatibility mode only), and above
POWER9 (KVM)	LE: RHEL 8.0 for POWER9 RHEL 7.5 legacy and above RHEL 7.5 for POWER9 and above	LE: SUSE LES 12 SP3 and above SUSE LES 15 and above	LE: Ubuntu 18.04 and above
POWER9 (Bare Metal)	LE: RHEL 8.0 for POWER9 RHEL 7.5 for POWER9 and above		LE: Ubuntu 18.04 and above

https://www.ibm.com/support/knowledgecenter/linuxonibm/liaam/liaamdistrospower7.html

* Not optimised

Linux on Power transition to all little endian

Belisama



- Red Hat lifecycle information - https://access.redhat.com/support/policy/updates/errata/
- SUSE lifecycle information – https://www.suse.com/lifecycle/
- Ubuntu lifecycle information - https://wiki.ubuntu.com/Releases

POWER9 – Ecosystem Enablement

OpenPOWER™

Foundation

- Accelerating Open Innovation

Grown from 5 to over 200 members in less than 3 years

POWER9: Engineered for OpenPOWER Application

- Built for a Broad Range of Deployments and Platforms
- Open and Flexible Solutions
- Ideal for Developers

Linux on IBM Power Systems developer portal

Learn about developing on the Power architecture, find packages, get access to cloud resources, discover the tools and technologies you need to build applications, and connect with your peers and other developers, data scientists, and academics

IBM Cloud Private on Power

IBM Cloud Private (ICP) helps developers easily build and deploy new, containerized applications and enables administrators to provide top notch availability, performance, and security in an on-premises private cloud.

→ Learn more

Watson Machine Learning

The Watson Machine Learning family of products provides an enterprise-class software solution for quickly and easily building an end-to-end deep learning environment for your organization.

→ Learn more

Docker on Power

Docker is an open platform tool that you can use to deploy, execute, and manage containers on IBM Power Systems.

→ Learn more

https://developer.ibm.com/linuxonpower/

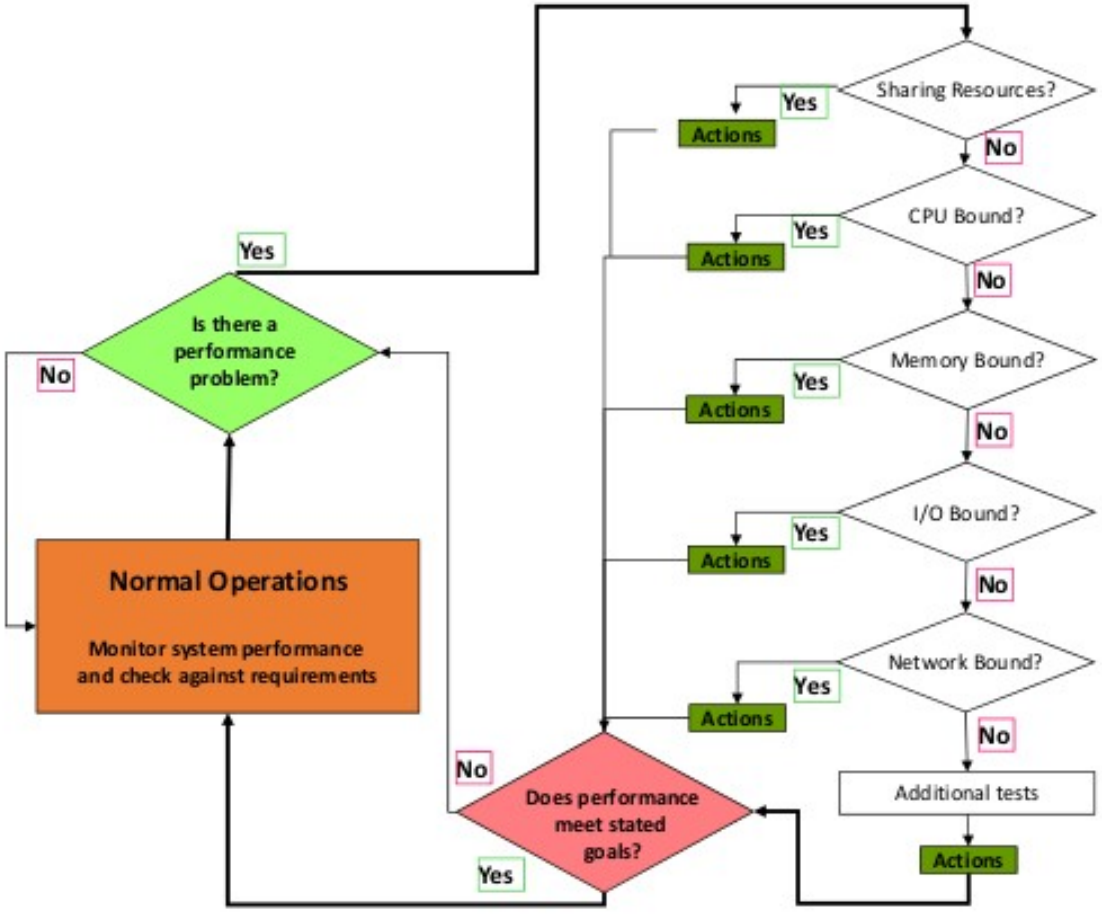
Linux tools: https://www.ibm.com/support/knowledgecenter/en/linuxonibm/liaae/liaaetoolsrepository.html

© 2020 Belisama

General recommendations

- Read IBM documentation around tuning and optimisation
  - Performance Optimization and Tuning Techniques for IBM Processors, including IBM POWER8
- Proceed slowly, change one variable at a time and document results
- Talk to the application owners and users – what is more important (eg for latency, mean, median, maximum or the standard deviation)
- Ensure that measurements taken are actually measuring the right of it, the whole of it and nothing but it.
- Ensure that the act of measuring is not impacting the result significantly
- Document baseline and settings
- Know the hardware and take full advantages of the features (ability to mix cores – some SMT=1 and some SMT=8)
- Isolate components to better locate bottlenecks

© 2020 Belisama


Sharing Resources?
Yes
Actions
No
CPU Bound?
Yes
Actions
No
Memory Bound?
Yes
Actions
No
I/O Bound?
Yes
Actions
No
Network Bound?
Yes
Actions
No
Additional tests
Actions
Does performance meet stated goals?
No
Yes
Is there a performance problem?
Yes
No
Normal Operations
Monitor system performance and check against requirements

Licensing

Belisama

Type of Activation	Static	Utility and Trial	Elastic On/Off	Integrated Facility for Linux	Mobile Power Enterprise Pools
Activation	Permanent	Temporary	Temporary	Permanent	Temporary and moveable
Supported OS	AIX, IBM i, Linux	AIX, IBM i, Linux	AIX, IBM i, Linux	Linux	AIX, IBM i, Linux
PowerVM Support	Standard or Enterprise	Standard or Enterprise	Standard or Enterprise	Standard or Enterprise	Standard or Enterprise
Supported Power	770+, 780+, 795, 850, 870, 880, 950, 980	770+, 780+, 795, 870, 880, 950, 980	770+, 780+, 795, 870, 880, 950, 980	770+, 780+, 795, 870, 880, 950, 980	770+, 780+, 795, 870, 880, 950, 980
Minimum Firmware Level	With All systems	720+	720+	P7(780), P8(820), P9(all)	P7(780), P8(820), P9(all)
Announced	With initial systems	With initial systems	With initial systems	P7(Oct 2013), P8&9(With initial systems)	P7(Oct 2013), P8&9(With initial systems)

Designed to eliminate planned and unplanned downtime for the most demanding workloads
Shift resources to support planned maintenance
Active – Active for efficient HA / DR
Seamless transition to new technology
Easily manage the changing workload demands of today's dynamic, real time business environment.
Move virtual processor and memory resources to address new demands without physically reconfiguring the data centre
Usage and utility based pricing
Minimize excess capacity required to manage availability and contingency for dynamic business environments
Ready for the Future Seamlessly integrate next generation technology in your Power Enterprise Pool
- PEP now pay as you go!
Inactive resources are used for processor and memory sparing
Inactive resources can be used for free trials of new applications
Inactive resources can be used with temporary activations for emergency backup

© 2020 Belisama

- AIX 7: (live kernel patching,
  - Runs on Power4-9, upgrade from 5 or 6 supported
  - Scaling to 256 Processor and 1024 threads
    - 5.2 and 5.3 WPARs in on Power7
    - 3 editions
      - Express: 4core, 8GB/core maximum per partition
      - Std: what we think of as AIX – 256core / 1024 thread, Cluster aware
      - Ent: Includes WPAR, Systems Director Enterprise Edn Cluster aware
  - IBM Licensing Metric Tool
    - Provides a data center view of IBM software licenses
    - Simplifies customer license tracking, compliance & audit reporting
    - Common IBM monitoring tool for SWG, Power Systems & Systems Director software products
    - Does NOT report back to IBM
  - Domain support for RBAC
  - Domains used to control access to VGs, Filesystems and Devices
  - AIX Security Expert and Compliance Expert updates

© 2020 Belisama

- AIX 7: (cont)
  - AIX binary compatibility
  - Improved performance using 1 TB segments – improved performance for 64 bit apps using shared memory regions (NB restricted tuneables)
  - Debugging improvements
    - Malloc debug fill; Core dump enhancements (unique names and control of what written); DBX and ProbVue enhancements
  - LVM enhancements
    - LVM enhanced support for solid-state disks and Hot files detection in JFS2
  - Continuous availability.
    - Firmware-assisted dump; Full memory dump options; Firmware-assisted dump support for non boot rootvg iSCSI device; Cluster data aggregation tool (cdat)
  - Cluster Aware AIX
    - Cluster wide: event management; device naming; command distribution and communication(mulitcast)
    - Kernel based heartbeat and monitoring
    - Cluster aware perfstat library interfaces

© 2020 Belisama

- AIX 7: (cont)
  - Distributed System Management
    - Commands: dpasswd; dkeyexch; dgetmacs; dconsole; dcp and dsh
    - Integrating with NIM (CSM removed)
    - Event management
  - AIX Profile manager (built on Runtime Expert)
  - Active Memory Expansion (AME)
  - PowerSC (Security and compliance)
  - Networking
    - Enhancement to IEEE 802.3ad link aggregation with LACP; Removal of BIND 8 application code; Support for Network Time Protocol version 4

© 2020 Belisama

Specific LVM improvements

Belisama

- AIX 7.1
  - LVM preferred read
- AIX 7.2
  - Space reclaim for supported storage

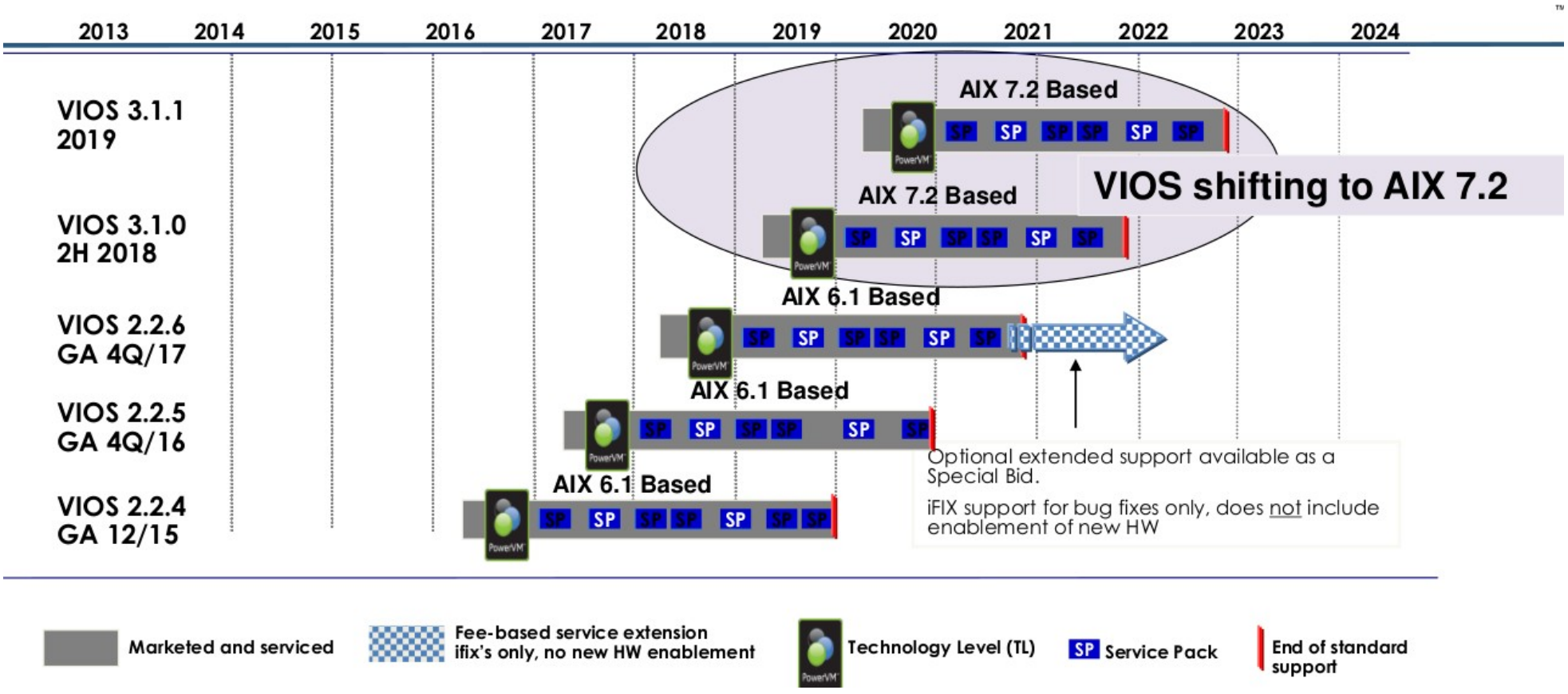
VIO upgrades

Belisama

- Security & Resiliency
  - Remove old software with a clean AIX 7.2 base
  - Storage multi-pathing enhancements
  - Improved CAA features for SSP clusters
- Flexible Infrastructure and Cloud
  - Virtualization of iSCSI storage
    - Includes iSCSI multipathing
- Features that have been removed
  - Integrated Virtualization Manager (IVM)
  - LPAR Hibernation
- Upgrade process
  - Save config, fresh install, restore config
  - NIM: bos and alt disk options supported
  - For SSP users (non disruptive) requires a few extra steps
- Modernization
  - Based on AIX 7.2 TL 3
  - Enabled for enhanced I/O scaling
  - Ready for P9 HW feature exploitation
    - XIVE*, SMT8, accelerators
- Platform Support
  - P7+, P8, P9
  - USB drive install support
- Usability
  - Improved version upgrade process options

* POWER9 processor closes this gap in the processor hardware virtualization support by introducing an eXternal Interrupt Virtualization Engine (XIVE) architecture

VIO Roadmap



© 2020 Belisama

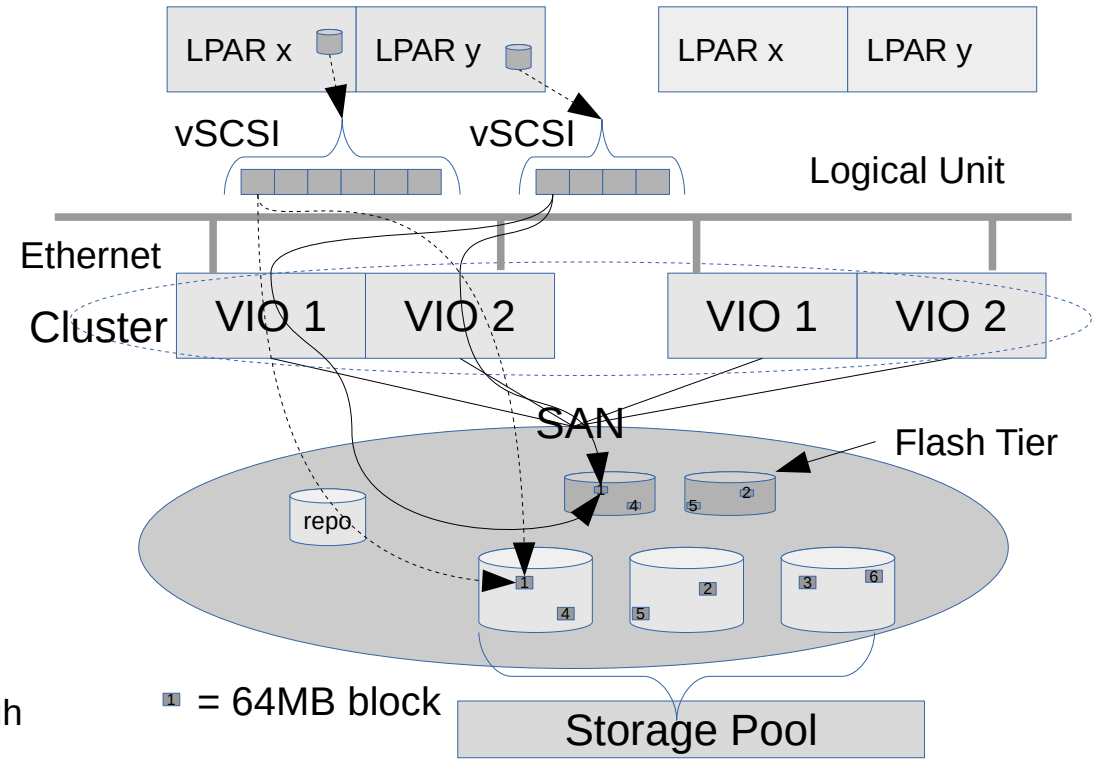
How Shared Storage Pools work

- Makes use of Cluster Aware AIX (Basis for PowerHA System Mirror 7.1)
  - Cluster wide: event management; device naming; command distribution and communication (mulitcast)
  - Kernel based heartbeat and monitoring
  - Cluster aware perfstat library interfaces
- Other 7.1 features.... RBAC, distributed commands...
- nmon/topaz part of VIOS code
- Storage team passes storage to AIX team, AIX team administers storage now
- Thin provisioning
- cfgassist or cli

© 2020 Belisama

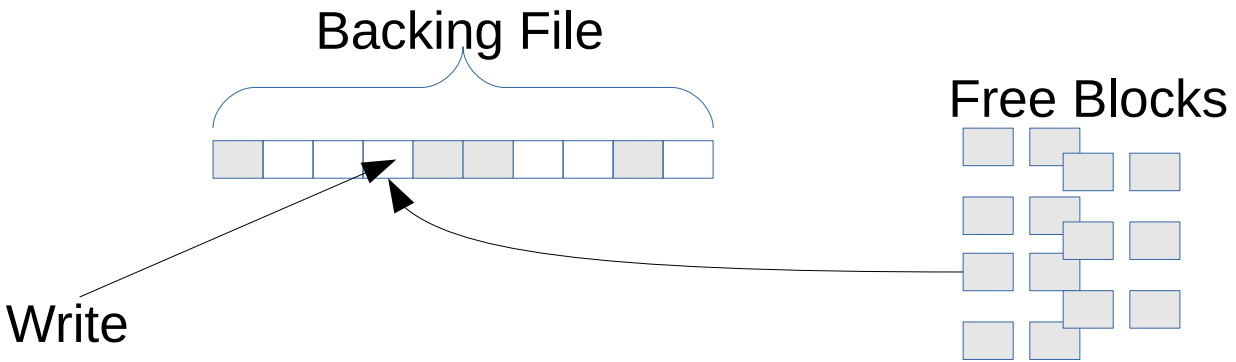
Shared Storage Pools operation

- How Shared Storage pools work
  - Makes use of Cluster Aware AIX (Basis for PowerHA System Mirror 7.1)
  - Cluster wide: event management; device naming; command distribution and communication (mulitcast)
  - Kernel based heartbeat and monitoring, cluster aware perfstat library interfaces
  - Other 7.1 features.... RBAC, distributed commands...
  - nmon/topaz part of VIOS code
  - Storage team passes storage to AIX team, AIX team administers storage now
  - Snapshots, Thick/Thin provisioning
  - cfgassist or cli
- Logical Units
  - Logical units are the file-backed backing devices that are provisioned to client partitions
    - Made up of a series of 64 MB blocks
    - Logical unit space is striped in 64MB blocks across the physical volumes in the shared storage pool
    - MPIO same as current vSCSI devices (map through dual VIOS)



Thin and thick provisioning

- There are two ways of provisioning storage space to clients:
  - Thick: allocate all the space that the client requests regardless if it is actually used
  - Thin: allocate space to the client as they use it, up to the amount of space that they request
  - Create a thin 16 GB logical unit and allocate to a vhost:
    - lu –clustername narcissus –sp reflection 16G –bd vdisk_lpar1 –vadapter vhost0
  - Create a thick 16 GB logical unit and allocate to a vhost:
    - lu –clustername narcissus –sp reflection 16G **–thick** –bd vdisk_lpar1 –vadapter vhost0



Limitations

- You cannot expand or reduce a backing device on the current versions

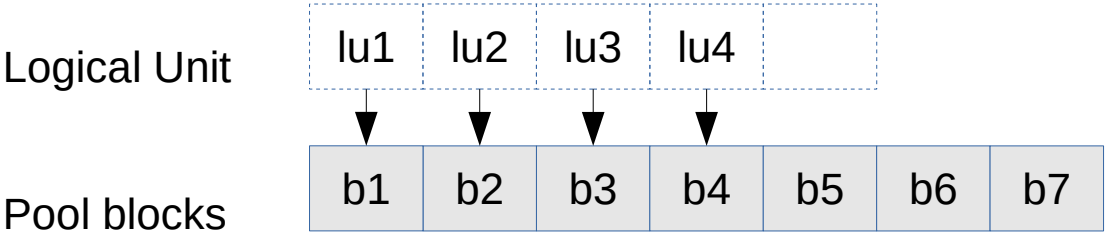
© 2020 Belisama

Snapshots

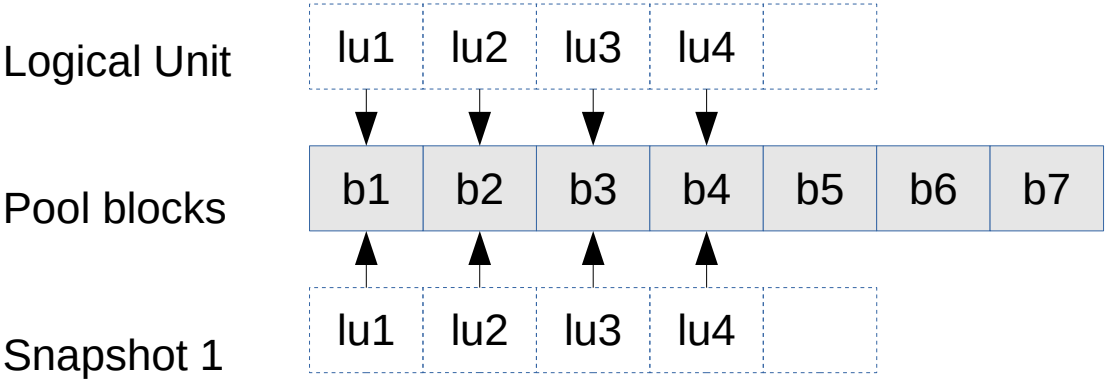
- Creating and managing snapshots
  - A snapshot is a quick point-in-time backup
  - Creating a snapshot involves copying the metadata (the list of blocks in the logical unit, not the blocks themselves)
  - When a client LPAR writes to a block of a logical unit that is part of a snapshot, the target block is copied, and the update is applied to the copy
    - This updates the current copy and the snapshot points to the original block
  - Snapshots are stored in the shared storage pool
- Snapshots can be used to roll-back a logical unit to a point-in-time or can be deleted
- Possible uses of roll-backs:
  - Practice software updates, reset images after testing or training, reset after benchmarks or troubleshooting

© 2020 Belisama

Original mapping



Create a snapshot

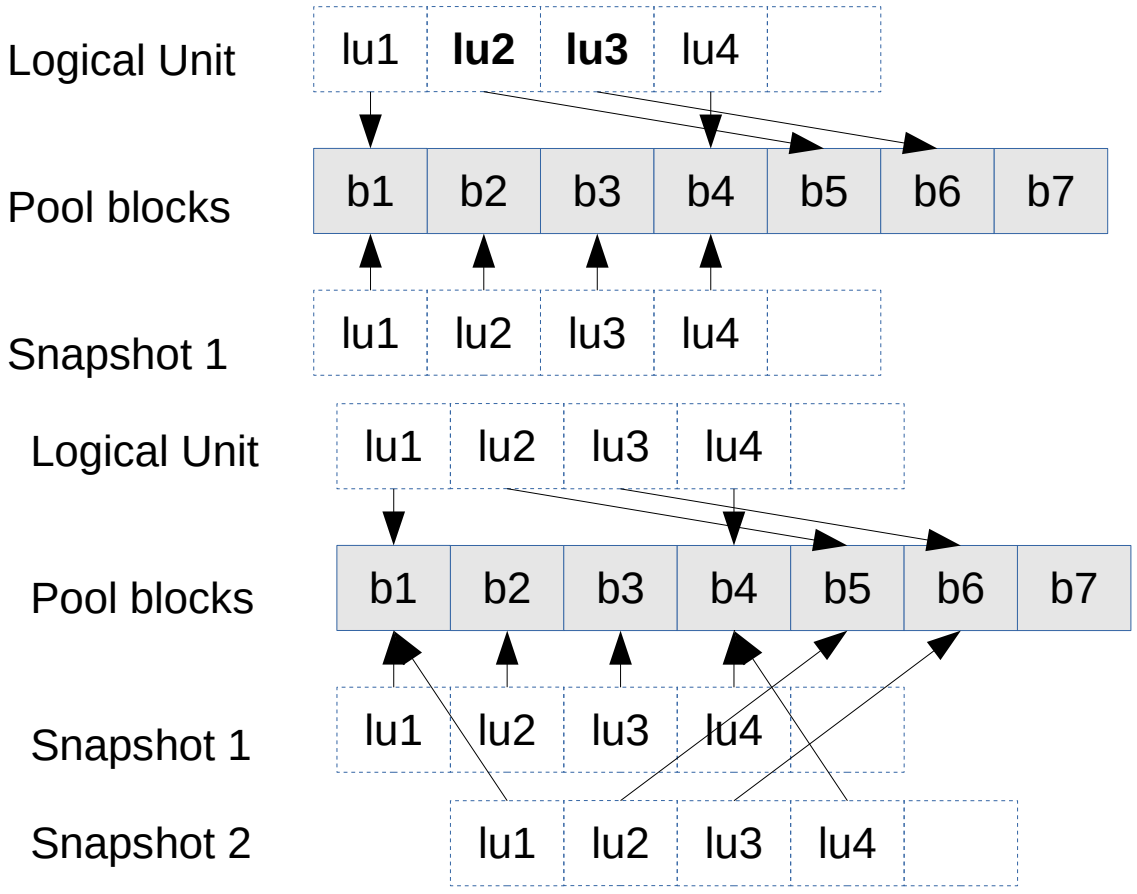


© 2020 Belisama

Snapshots

Modify some data in client

Take 2nd snapshot



© 2020 Belisama

Shared storage pools evolution

- Ease of Use
  - PowerVC, HMC, and Novalink
- Functionality Added in Recent Releases
  - Mirroring at pool or tier level
  - Grow LU
  - Storage tiers and move LU
  - Import PV
  - Improved resiliency (network related improvements, cluster-wide snapshot,automated log analysis of pool, viosbr auto-backup, DR with viosbr)
  - Flash acceleration
- Increased Scalability
  - 24 nodes (more than 16 nodes require all flash storage for system tier).
  - 2000 VMs / LPARs, 250 VMs/server
  - 512 TB pool capacity, 1024 pool disks

© 2020 Belisama

More recent updates to SSP

- Support for virtual IP address (VIPA) and multipath routing to allow 2 different networks for storage pool communication.
- Various resiliency improvements:
  - Cluster-wide outage avoidance (expel)
  - General hardening (Disk challenge; Meta-root replica set, etc)
  - Pool log retention: multiple log files and resolve log wrapping issue
- Dedicated, large scale SSP test cluster
  - Large # of nodes (24) and 100 TB+ SAN storage.
  - Mimic customer SSP usage (e.g. no cluster tear-down between tests, continuous I/O, etc.)

© 2020 Belisama

System maintenance - Introduction

Belisama

- Starting point
  - Best practices for implementing Oracle on AIX:
    - Oracle 9i & 10g on IBM AIX 5L: Tips & Considerations
    - Oracle Architecture and Tuning on AIX white paper
    - Tuning IBM AIX 5L for an Oracle Database white paper
    - Oracle DB & RAC 10gR2 on IBM AIX : Tips and Considerations
    - Oracle DB & RAC 11g on IBM AIX: Tips and Considerations
  - IBM System p Advanced POWER Virtualisation Best Practices Redbook
  - IBM eServer Security Planner and the Strengthening AIX Security: A System-Hardening Approach white paper
  - Learn 10 good UNIX usage habits web page
    - http://www.ibm.com/developerworks/aix/library/au-badunixhabits.html
  - Good AIX and Power system information
    - https://w3.tap.ibm.com/w3ki/display/wpSeriesFTSS/WR+System+p+FTSS+Wiki
  - PCI Adapter placement guide

© 2020 Belisama

System maintenance – Introduction (cont).

Belisama

- Starting point (cont)
  - Developerworks AIX installation wiki
    - http://www.ibm.com/developerworks/wikis/display/WikiPtype/Installation
  - Regularly visit Service and support best practices for UNIX servers
  - IBM Blogs
  - Nigel Griffiths youtube

Service and support best practices

For Power Systems

IBM provides many documents that recommend service and support strategies for IBM systems and software. These best practices documents describe system planning and support procedures to improve system administration operations. The documents also provide strategies for IBM Power Servers. You will find information on key components like firmware, the AIX operating system, cluster software and the Hardware Management Console.

AIX

January 2019	IBM AIX Operating System Service Strategy Details and Best Practices (729KB)
April 2018	POWER9 Performance Best Practices (202KB)
July 2017	AIX on Power - Performance FAQ (1691KB)
February 2015	POWER7 Performance Best Practices (336KB)
January 2015	POWER8 Performance Best Practices (202KB)
August 2014	Transitioning to POWER8: Migration Paths for AIX systems to POWER8 (539KB)
June 2013	Upgrading AIX from 5.3 to 7.1 (36KB)
January 2011	Managing AIX Updates using SUMA, NIM, and AIX Service Tools (120KB)

Hardware and firmware

May 2011	IBM Power 770/780 and 795 Servers CEC Hot Add & Repair Maintenance Technical Overview (1.5MB)
April 2009	IBM Power 595 and 570 Servers CEC Concurrent Maintenance Technical Overview (1.06MB)
December 2008	IBM Power Systems System Firmware (Microcode) Service Strategies and Best Practices (1.6MB)
April 2006	Architectural Considerations for Production Environments Incorporating System p Servers (404KB)

Hardware Management Console

January 2007	IBM Hardware Management Console (HMC) Best Practices p Servers (492KB)

Clustering

January 2008	High Availability Cluster Multiprocessing (HACMP) Best Practices (707KB)
April 2006	Considerations and Sample Architectures for High Availability on IBM eServer pSeries and System p Servers (1.1 MB)

http://www14.software.ibm.com/webapp/set2/sas/f/best/home.html

System maintenance – Introduction (cont).

Belisama

- Know SLA's for all applications, architect accordingly
- Record a view of your applications and SLA's by partition and by machine
- Build an adequate test environment. For EVERYTHING.
- Create and test your backup strategy, in its entirety, routinely. And when anything changes!
- Have a PLAN to apply maintenance. Yes even in a 24/7 52 weeks per year environment.
- MONITOR your environment
- TEST ALL CHANGES!
- Understand your baseline performance.
- Understand your peaks and how virtualisation features can help consolidate servers.
- Have a capacity plan in place for peaks. Test before needed.
- Review monitoring and escalation procedures.
- Run a 'test' problem if its been a while.
- Spend extra time on designing your I/O layout. Especially for databases!
- Know tools and monitoring techniques for problems and what they look like when the system is NORMAL…

© 2020 Belisama

System maintenance – Introduction (cont).

Belisama

- Know what tools are coming…
  - NIM is still your friend
  - Keep up with changes in management - HMC, PowerVC..
  - AIX tools for Linux, AIX toolbox…
  - PowerSC
- Develop a strategy for classifying servers (according to availability requirements and change/release tolerance)
- Develop the strategy for notification of fixes, commonly a mix of SUMA and subscription services.
- Review the tools you will use such as System Planning Tool, NIM, SUMA, FLRT
- Describe the different levels of code to disseminate
- Determine how to disseminate the changes (NIM or .... NIM?)
- Changes in AIX Security Expert and AIX Performance Expert

© 2020 Belisama

'Gold' code

- If you are maintaining a number of AIX servers, then a ‘golden code’ image should be maintained. A NIM server and a cloned image is an excellent way to maintain O/S consistency.
- Concept of 'Gold Code' or SOE
  - Creating the 'Gold' image
  - Maintaining the 'Gold' images
  - Actual image or script to create
- Categorise your images:
  - When creating an image that will be used on many servers, you may want to split out into two or three images that have unique characteristics
    - Database servers
    - Application servers
    - Web servers
  - Don’t Reinvent the wheel! – keep it consistent where possible

© 2020 Belisama

'Gold' code images

- Creating the images
  - Incorporate as many of the post-install tasks into the base image as you can and still maintain the ‘golden code’
  - You might consider sizing your filesystems (/usr / /tmp & dump) before you create a clone and creating any RCT monitoring that is standard along with utilising performance templates
  - Use NIM or some mechanism to track and organise images and fixes:
  - As code is moved into the environment, the golden code should be evaluated and kept current through the NIM process. You need to keep NIM up to date to be effective
  - Within a nim environment, include vg backups of the non rootvg volume groups structures.
  - Using *savevg -mrivf /usr/local/recovery/datavg datavg* makes sure there is a saved layout of the external disks prior to the mksysb running , and thus the info is also saved within the mksysb image on the nim server.
  - AIX ‘ghostdev’ systems flag
  - This flag is primarily used for duplicating AIX images to multiple environments.
    - When the AIX operation system comes up on this new VM, AIX detects that it is on another system and automatically removes the hostname, IP address configuration as it comes up on the new host. It also removes the volume group information as well as changing device configuration to their default values. As these are often tuned, this can be scripted around.

© 2020 Belisama

'Gold' code images (cont).

Belisama

- Creating the images (cont)
  - The most recent Technology Level and Service Pack should be researched and considered as a starting point if this is the initial setup of an environment. Get on the subscription list to get notified of hypers and critical fixes. Consider using SUMA to monitor and download fixes.
  - Get familiar with *niminv* and *getinv* commands as well as *compare_repor*t.
  - Firmware is just as important as software to keep current. Have a plan to allow currency of both – and check dependencies!

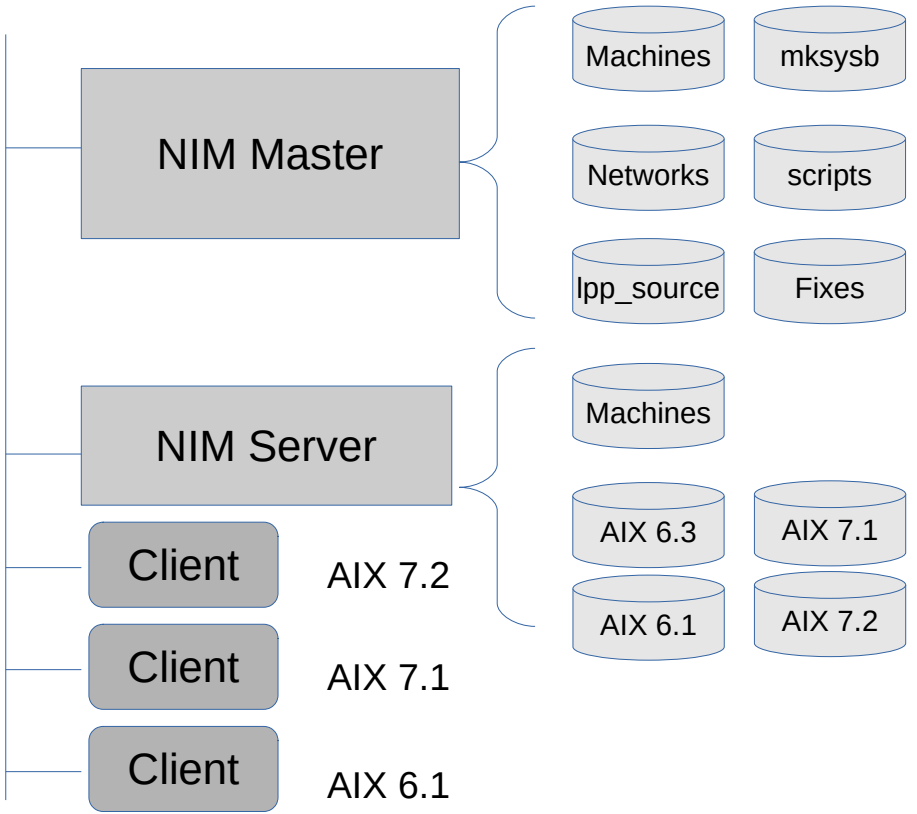
Service Update Management Assistant

- SUMA, included in the base AIX 5L™ Version 5.3 operating system, provides flexible, policy-based options to perform unattended downloads of AIX 5L™ updates from the Support Web site.
  - Notification of requester via email
  - SMIT or command line interface
  - TL's or SP's will be downloaded (no specific PTFs or APAR support after 10/08), but individual updates can be installed if desired after the download

```
                Create a New SUMA Task

Type or select values in entry fields.
Press Enter AFTER making all desired changes.

                                                        [Entry Fields]
  DISPLAY name for this task                          []
  ACTION                                               Download                  +
* Directory for item STORAGE                           [/usr/sys/inst.images]
  TYPE of item to request                              APAR                      +
  NAME of item to request                             []
  LEVEL of item to request                            []                         +
  Get PREREQUISITES/COREQUISITES?                      yes                       +
  Get IFREQUISITES?                                    yes                       +
  Get SUPERSEDING items?                               yes                       +
  Get items which fix REGRESSIONS/PEs?                 IfAvailable               +
  REPOSITORY to filter against                        [/usr/sys/inst.images]     +
* BASE ML to filter against                           [5300-01]                  +
* SYSTEM or lslpp path to filter against              [localhost]                +
* MAXIMUM total download size (MB)                    [256]                      #
  EXTEND file systems if space needed?                 yes                       +
* MAXIMUM file system size (MB)                       [2048]                     #

Scheduling Options:
*  NOTIFY email address                              []              +
*  Repeat FREQUENCY                                  []              #
*  Repeat Frequency UNITS                             hours +
*  Starting TIME                                     [1139]          +
*  Starting DAY                                      [15]            +
*  Starting MONTH                                     February       +
*  Starting YEAR                                     [2004]          +

F1=Help              F2=Refresh           F3=Cancel            F4=List
F5=Reset             F6=Command           F7=Edit              F8=Image
F9=Shell             F10=Exit             Enter=Do
```

© 2020 Belisama

- Install and maintain environment
  - Install servers (fresh, alt_disk_install, nim move_up)
  - Maintain servers (application code and pre-requisites)
  - Backup servers (pull mksysb images)
- But there is more.....
  - Gather, merge and download fixes based on inventory of the environment
  - Extend compare_report to the NIM resources
  - Generate an inventory of all defined systems
  - Manage Linux clients (need CSM)



compare_report

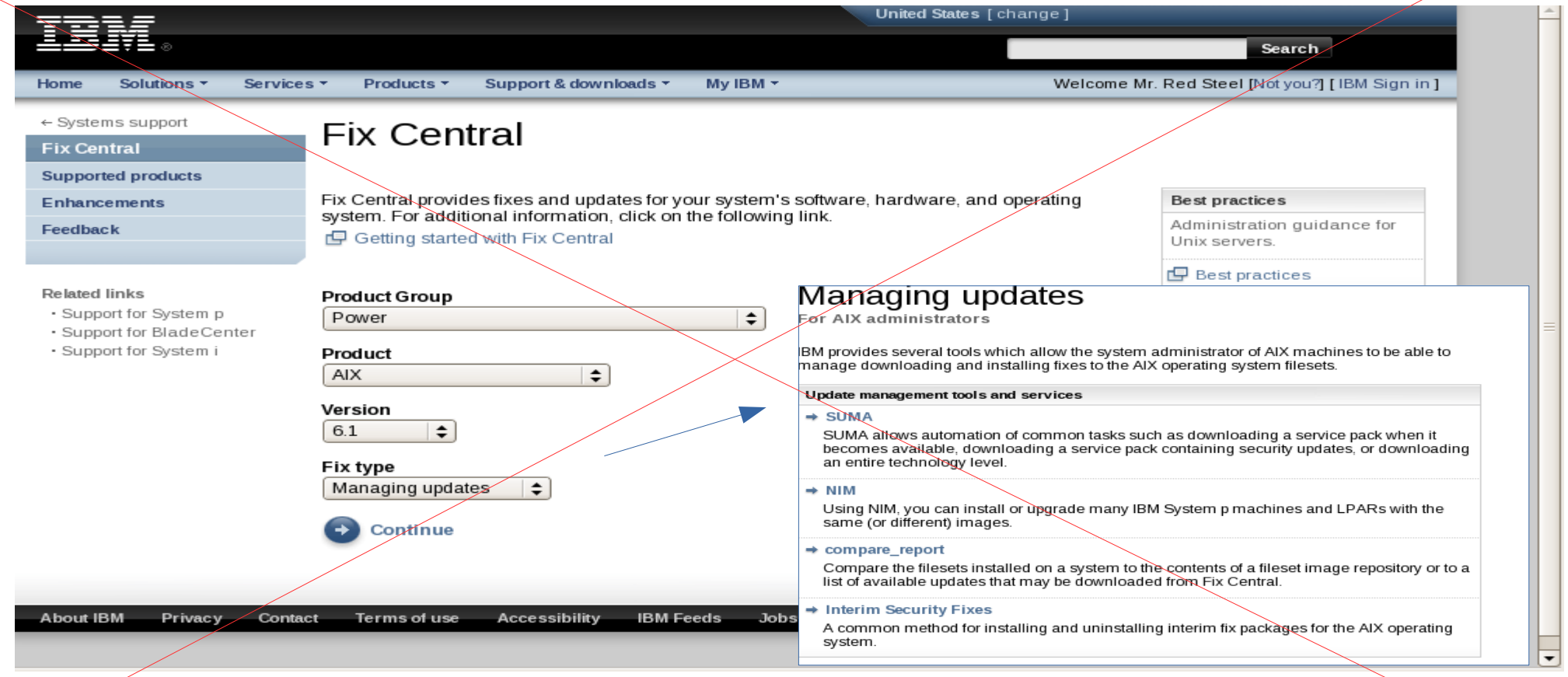
- The AIX compare_report command compares the filesets installed on a system to the contents of a fileset image repository or to a list of available updates that may be downloaded from Fix Central.
- It produces reports that simplify the process of determining the fixes to install to bring a system to the latest maintenance level or the latest level. Reports that are created using the list of available updates can be uploaded directly to Fix Central (5.1 uploads, 5.2 uploads, 5.3 uploads) to request the exact fixes needed for the system.

```
# compare_report -s -r 6100-04-05-1015.compare -l -t /tmp
#(lowerthanlatest1.rpt)
#Installed Software that is at a LOWER level
#PTF:Fileset_Name:Installed_Level:Available_Level
U830425:ICU4C.rte:6.1.4.0:6.1.4.1
U831671:X11.apps.xdm:6.1.4.0:6.1.4.1
U831697:X11.apps.xterm:6.1.4.0:6.1.4.1
U833911:X11.base.rte:6.1.4.0:6.1.4.2
U830377:X11.samples.apps.clients:6.1.4.0:6.1.4.1
..........
```

© 2020 Belisama

compare_report (cont)

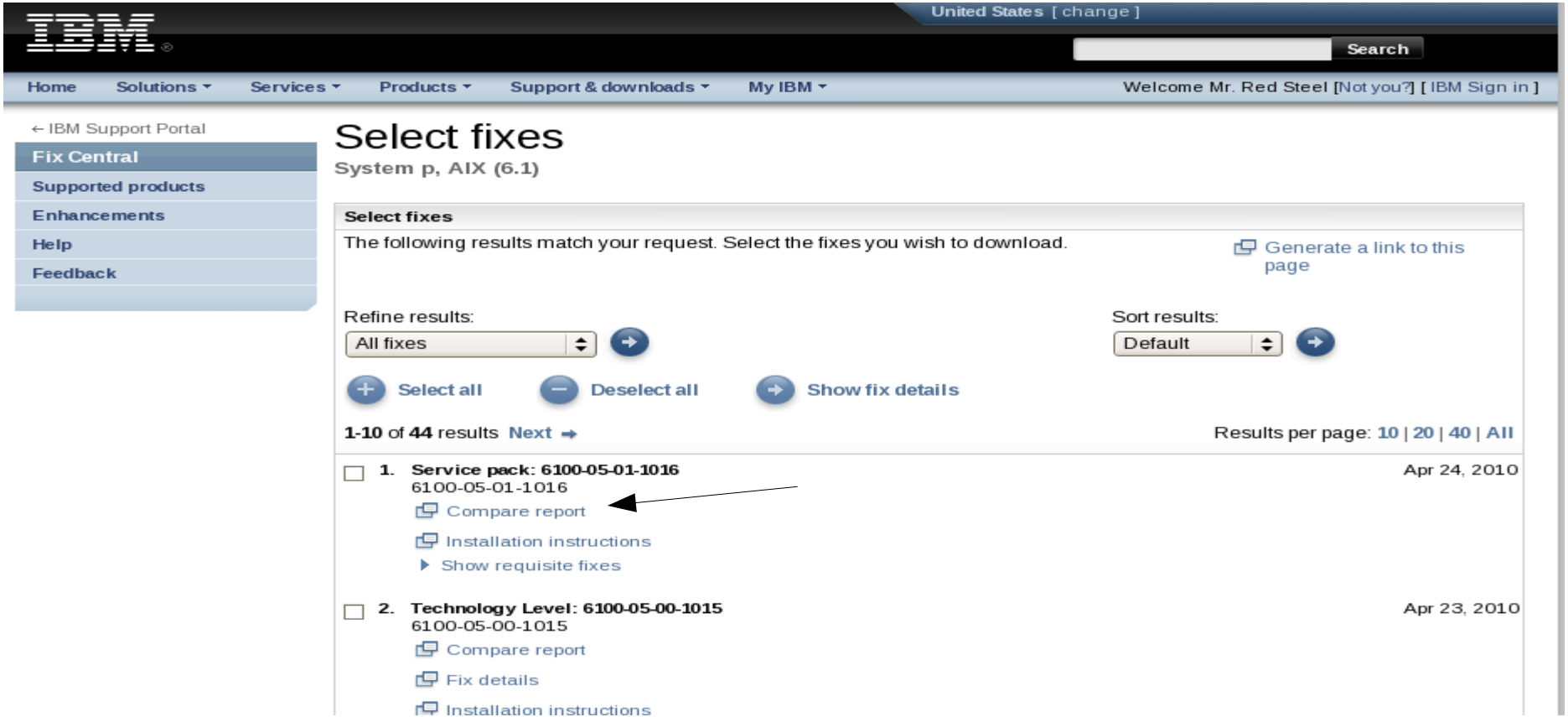
Belisama



© 2020 Belisama

compare_report (cont)

Belisama



Planning maintenance

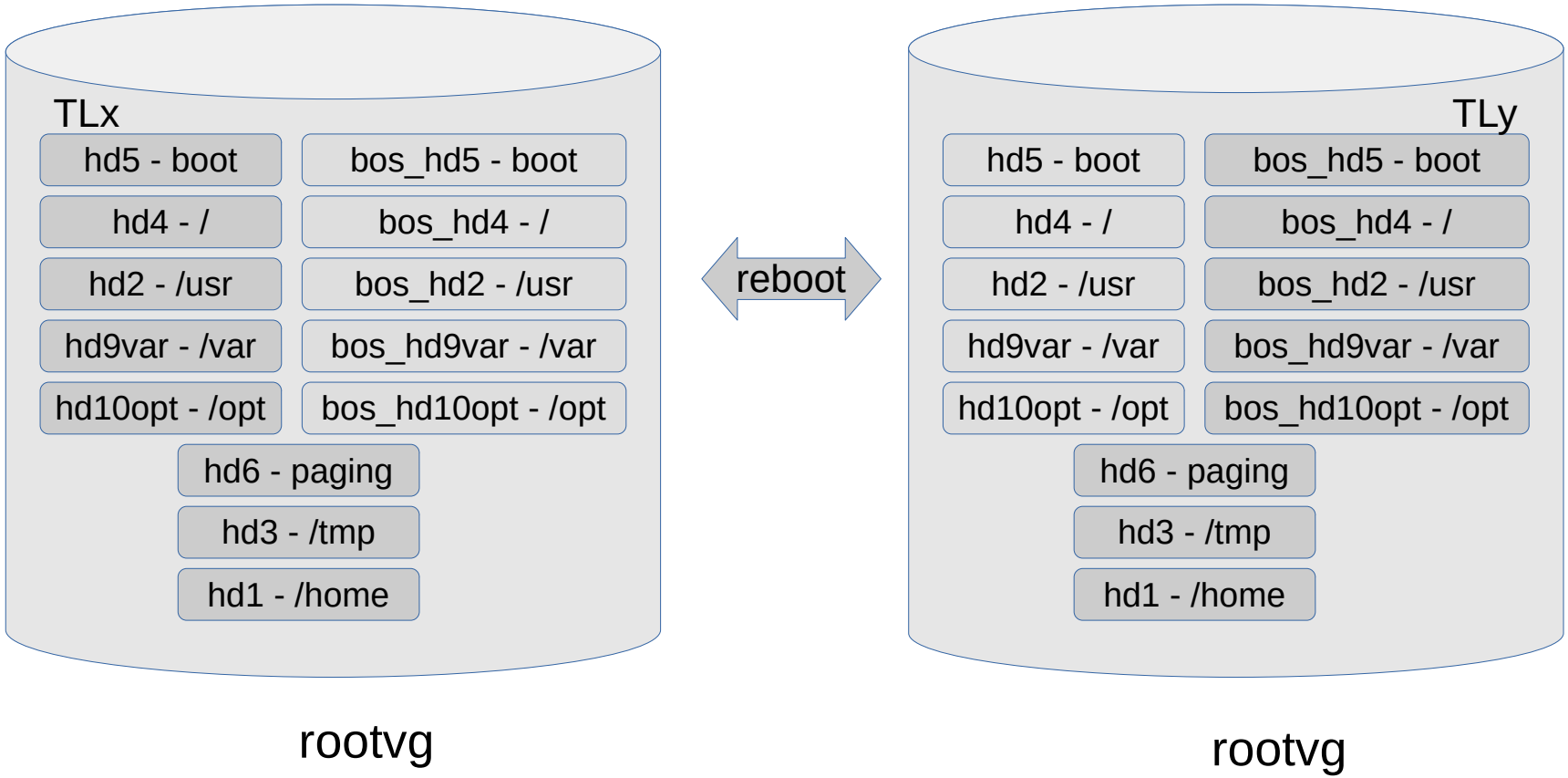
- Lobby for bi-monthly maintenance window (be happy if you get once a quarter), even if you don't use it every time…
- You MUST update your firmware once per year (firmware releases are supported for 1 year / 2yrs for POWER7/8/9 and should plan to embrace a Technology Level per year also. Good planning dictates an additional concurrent firmware upgrade during the year as well.
- Remember the high level vew. Have your SLA's defined and a view of them by machine and by partition. This will be useful in defining a maintenance policy.
- Check your machines ahead of time (such as readiness checker) to review your environment BEFORE you start the upgrade!

What expectation does management have of the system?

© 2020 Belisama

- Some recommendations
  - Use alt_disk_install or multi-bos to clone a running, optimally laid out system when beginning the mksysb setup.
  - rootvg should always be mirrored and quorum turned off in a single disk configuration (Storage RAID).
  - 2 dump spaces in mirrored environment; Dump always 'true'; regulary check size sufficient
  - If more than one paging space, keep same size. Don’t have more than one paging space / disk
  - A test environment should be maintained that can test the initial and periodic restore of the ‘golden code’ image. It must include like configuration (eg., HACMP).
  - Move code into and back out of test EXACTLY the same way before you do it in production.
  - Customers should install the latest AIX maintenance level on any system which will be tested extensively prior to deployment, unless the application vendor(s) strongly recommend some level other than the latest.
  - Create a new mksysb after the initial install is completed. Ensure that the prompt field is set to YES in bosinst.images. Label the tape and make it write protect.

© 2020 Belisama


TLx
hd5 - boot
bos_hd5 - boot
hd4 - /
bos_hd4 - /
hd2 - /usr
bos_hd2 - /usr
hd9var - /var
bos_hd9var - /var
hd10opt - /opt
bos_hd10opt - /opt
hd6 - paging
hd3 - /tmp
hd1 - /home
reboot
TLy
hd5 - boot
bos_hd5 - boot
hd4 - /
bos_hd4 - /
hd2 - /usr
bos_hd2 - /usr
hd9var - /var
bos_hd9var - /var
hd10opt - /opt
bos_hd10opt - /opt
hd6 - paging
hd3 - /tmp
hd1 - /home
rootvg
rootvg

Further tips....

Belisama

- Some other recommendations (cont)
  - Install bundles (tested together), build and test your own bundles.
  - Always re-apply current TL / SP after installing additional filesets
  - Plan to replace iFix with TL/SP as soon as possible. IBM does not test all combinations of iFixes.
  - Keep /usr clean, make /usr/local as separate file system to survive an upgrade
  - Have important hosts in /etc/hosts and netsvc.conf (hosts=local....) or use NSORDER variable to avoid DNS problems (particularly in clusters)
  - Don't use rsh/rcp if you can help it
  - Turn off all unnecessary services in inetd
  - Is your dump space big enough?
  - Editing /etc/filesystems?? <= will not survive an exportvg/importvg
  - Use Electronic Service Agent
  - Disable unused services in inetd.conf (AIX Security Expert?)
  - Have detailed system plan (particularly if virtualised) keep up to date!
  - Use VIO Server command line to update VIOS
  - Enable Stack execution disable feature (protects against buffer overflow security vulnerability) – do you check IBM’s security advisories??

© 2020 Belisama

Further tips....

- Some other recommendations (cont)
  - Do not boot clustered systems in maintenance mode from install media attached to storage.
  - If you don't use NIM (why?) make sure you have bootable media /image for the TL/SP level of your system
  - Regularly capture bootable backups – and test your restore process
  - REALLY test your DR plan
  - REALLY test backups – plan backup strategies to meet your return to service agreements. Focus on recovering data – not the backup process => what information do you need to fully recover a system... do you have it?
  - Remember aix only saves rootvg VG structures by default – don't get caught
  - AIX error notification can be customised to email alerts for particular events
  - You test environment should match your production environment (VIO etc)

© 2020 Belisama

AIX Security improvements

- Use different hashing algorithm (other crypt) for storing passwords.
  - Provides much better security for passwords, and supports passwords greater than 8 characters
- Trusted Execution feature in AIX 6.1 to enhance security. Protect against intrusion attacks.
  - Provides for integrity verification of the system
  - Lock down the system and protect against attacks from an intruder
- Segregate your important data and use AIX Encrypted file system to encrypt data
  - Protect important data against attacks
- RBAC
  - Increased granularity in controlling who can execute what command

© 2020 Belisama

- More to consider..
  - Consider creating a dummy empty LPAR with the same connectivity as primary LPARs to do upgrades into. Then use ALT_CLONE and upgrade the new software (i.e. Oracle, WAS, SAP, etc) into that and switch the addresses published externally when you are happy the upgrade has worked properly.
  - Work to standard infrastructure patterns for an estate. Make each machine adhere to the pattern for that role – even if there is something it requires less than other machines in that role. It makes builds and management easier.
  - Learn the HMC command line. Performance is better than the GUI.
  - Have a script that runs each night that packages up error report alerts for that day and Emails them. Sometimes systems management tools can get errors and not alert you appropriately when a problem occurs so a failsafe script (send me an email) is worth having.
  - Set up accounts to point to an LDAP / Kerberos environment off the LPARs – even if it uses AD. It makes managing passwords a whole order of magnitude easier because it is all done in one place.

© 2020 Belisama

Tips (cont).

- What we are trying to avoid...
  - If the machine goes down and is restarted, it will boot to the devices in the bootlist. If the bootlist is set incorrectly, the machine may not be able to boot.
  - If something happens to rootvg, the mirror in the hdisk can be used to recover rootvg.
  - If the mirrors are not synchronised and you lose one mirror, you may lose data.
  - People typically have multiple paths to a disk because they need to be able to access the disk even if all but one of the paths goes down. If all of the paths except one go down and you don't notice it, you are betting that the final path will not go down. It is better to try recover paths so you always have multiple paths to a disk and can risk losing one.
  - In case something happens to the server, rootvg can be recovered from the NIM server.

© 2020 Belisama

General recommendations

- And more ....
  - When installing AIX for the first time (or upgrading to a new release level), be sure to install available fixes. See AIX installation best practices for more information. When installing AIX on a LUN on a Storage Area Network (SAN), be aware of considerations unique to that environment. See AIX boot from SAN for more information.
  - Capture bootable backups periodically (monthly/quarterly?).
  - Store some bootable backups off site for recovery if the data centre is destroyed.
  - Test the restore process periodically (yearly?) by restoring the most recent bootable backup. It is best for the restore to be tested by someone other than the person who captured the backup, to confirm that restore procedure documentation is adequate.
  - Capture application Volume Group configurations (details of volume groups other than rootvg are not included in a mksysb).
  - Have you applied any code to your system? Are you likely to want to boot from the Installation Media?

© 2020 Belisama

System recovery

- Really test Disaster Recovery
- The support centre sees a wide variety of issues
  - Some of the issues require customers to restore files, entire/partial databases or even whole systems
  - A surprising number of customers have never had to restore anything and are unprepared or unable to complete the restoration successfully
  - As a result, customers suffer severe problems
- Understand there are several forms of backups, needs based on how they will be used – records retention, recovery, hot stand-by
- Evaluate the best backup options for a system considering time and cost
- Look at some of the details that prevent backups from successfully recovering
- **Focus on recovering data not just backing it up**

© 2020 Belisama

Backups

- Traditional backups
  - Saved for long periods, usually for historical purposes
  - Kept off site
  - Rarely used to recover a system
  - Rarely tested to ensure they are usable
  - May become unrecoverable due to software and environmental changes
  - Lower resources to run but take longer to complete
- mirror/standby copies
  - Usually made to another set of disks, can be local or remote but always accessible
  - Each copy requires additional disk space
  - Utilises the internal disk system or special links to transfer data at high speeds
  - Number of backups limited by copies of data and disk space and more expensive
  - Short time needed to make data available, not usually suitable for records retention

© 2020 Belisama

Backups (cont).

- Amount of downtime is becoming more critical and affect your business
  - Plan your backups from aim of restoring system
    - Backups should not be done just to complete a check box
    - Time is a major factor when backups are used to recover business critical system
    - The recovery process should be the main driving factor in backups
    - Without recovery, backups are wasted resources
      - Shift the focus to recovering data
      - Evaluate each system to understand the impact of downtime
      - Plan backup and recovery resources based on the impact of downtime
      - Do not expect one plan/strategy meets all needs
      - Know how long each recovery option takes
      - Identify super critical information and have a plan to deal with failures of that data – archive logs might need to be on 2 different backup media for example
      - Change the backup procedures to support recovery - generate the recovery process as part of the backups

© 2020 Belisama

Recovery (cont).

Belisama

- Check list for recovery
  - Create and save system plans for all LPARs
  - Create and save storage subsystem layout backups
  - Create a /recovery-$hostname file system in rootvg for each host – this way multiple recovery filesystems can be restored without a problem
  - Identify the unique data on each host and the best way to backup the data to the /recovery file system
  - Make sure to save structure information for all volume groups and the maps of all LVs
  - Save a snap with everything except a dump
  - Save configuration files for all other products – PowerHA, GPFS, db2, oracle, etc.
  - Licensing information available?
  - Test recovering a system on a regular basis
  - Be prepared to recover the backup/recover host and how to make it faster/easier
  - Documentation that can be read by anyone

© 2020 Belisama

Recovery (cont).

- Special considerations for VIO
  - Special Considerations: VIO Servers
  - Documented at:
    - https://www-947.ibm.com/support/entry/myportal/Overview/Software/Virtualization_software/PowerVM_Virtual_I~O_Server
  - Have you taken a configuration backup from the HMC? From the VIO Server itself (viosbr)
  - Should also perform a snap command from the padmin login and before normal backup process outlined above
  - Look at ‘Methods for backup and restore’ near the bottom in the “Problem Solving” section
  - Need to keep current with HMC/VIO/Sysplan code to avoid issues
  - Needs to be pulled off the VIO server and saved someplace
  - Each VIO Server needs to be backed up individually
  - Remember that the LPAR/CEC configuration needs to be in place for the VIO server restoration to work correctly
  - Virtual device configuration will be lost if recovery is on different hardware

© 2020 Belisama

Recovery (cont).

Belisama

- Special considerations for VGs
  - Volume group structure data is only saved for rootvg by default
  - Having volume group structure data allows for quick recovery of the volume group, logical volumes and file systems including customised attributes
  - Damaged volume groups may be able to be reconstructed if maps are available
  - Volume group structures can also be modified by changing the data file before restoring
  - Consider using *'savevg –i –p –r –f OUTFILE VGNAME'*
    - NOTE: this does not save any data!
    - Creates a backup/restore format file
    - Attribute of the VG can then be edited before running restvg
    - restvg -f date_file -r hdiskx

General Recommendations

- Summary
  - Test application data restore process periodically (yearly?) by restoring the most recent backup. It is best for the restore to be tested by someone other than the person who captured the backup, to confirm that restore procedure documentation is adequate.
  - Monitor the system for errors. Attempt to discover the root cause of every error (apply latest fixes is not the answer.... may be an answer) and to address the cause to minimise the number of errors which occur, while acknowledging that getting a failed system back in operation must sometimes take precedence over collecting the diagnostic information required to determine failure root cause.
  - The primary AIX error log can be displayed using the errpt command.
  - Please note that an AIX Error Notification exit can be used to take action (eg, send an email) if a particular error occurs.
  - The primary HMC error log can be displayed from the HMC GUI using Service Applications.
  - The HMC can be configured to send an email when a new serviceable event is logged.

© 2020 Belisama

Post incident

- Conduct a post mortem after each application outage. Attempt to answer the following questions and then act upon the answers:
  - Was there any warning of this outage? If so, why was the warning not acted upon in time to prevent the outage?
  - What changes can be made to prevent the outage in the future?
  - Are other servers exposed to similar outages?

© 2020 Belisama

Don't Testing In Production....

Belisama

- Don't Test In Production
  - Test environment should mirror production
    - Same hardware, aim to achieve similar load for testing
  - Don't use a script for the first time in production.
  - If using PowerHA, test after every change (test tool)
  - Always apply fixes in production the same way as test
  - Test your back-out plan
  - Always complete changes (did you test the incomplete change?)
  - Best practice – have a staging environment to properly test changes.
  - Use IBM's tools to check
    - compare_report to compare test to staging to production
    - Use commit and apply
    - Use alt_disk_install

© 2020 Belisama

General firmware strategies

- IBM releases new firmware for the following reasons:
  - The addition of new system function.
  - To correct or avoid a problem.
- There are some natural points at which firmware should be evaluated for potential updates:
  - When a subscription notice advises of a critical or HIPER (highly pervasive) fix, the environment should be reviewed to determine if the fix should be applied.
  - When one of the twice-yearly updates is released.
  - Whenever new hardware is introduced into the environment the firmware pre-reqs and co-reqs should be evaluated.
  - Any time HMC firmware levels are adjusted.
  - Whenever an outage is scheduled for a system which otherwise has limited opportunity to update or upgrade.
  - When the firmware level your system is on is approaching end-of-service.
  - If other similar hardware systems are being upgraded and firmware consistency can be maximised by a more homogeneous firmware level.
  - On a yearly cycle if firmware has not been updated or upgraded within the last year.

© 2020 Belisama

General OS strategies

- A good fix maintenance strategy is an important part of maintaining and managing your server. Regular maintenance of your server, and application of the latest fixes help to maximise server
- IBM recommends that all servers be kept on a supported release and current with latest available fix packages for HMC and server firmware fixes.
- The most important scenario to avoid is remaining on a release so long that all subsequent releases that support a single-step upgrade are withdrawn from marketing. Without a single-step upgrade available, there are no supported ways for you to upgrade your server.
- IBM recommends at minimum you apply a TL / SP per year, preferably twice...
  - Release strategy
    - Technology Levels
      - From 6.1 TL 6 and for 7.1 / 7.2 once a year.
      - Technology Levels contain fixes for defects found by customers and internal testing, exploitation of new hardware, and may contain software enhancements.
    - Service Packs
      - Generally released approximately every three months
      - Can be released any time as needed if important fixes are available
      - Service Packs contain fixes for defects impacting customers, fixes for critical defects found in internal testing, and may contain enablement for new hardware. In general, changes that are allowed in a Service Pack are minimal defect fixes that do not change default behaviour nor add new functionality

6.1 EOS!

- Service Pack (cont)
  - Fixes that cannot wait till next TL
  - recommend it be applied entirely
  - released 8-12 weeks
- Both TL and SP will require a reboot (did!)

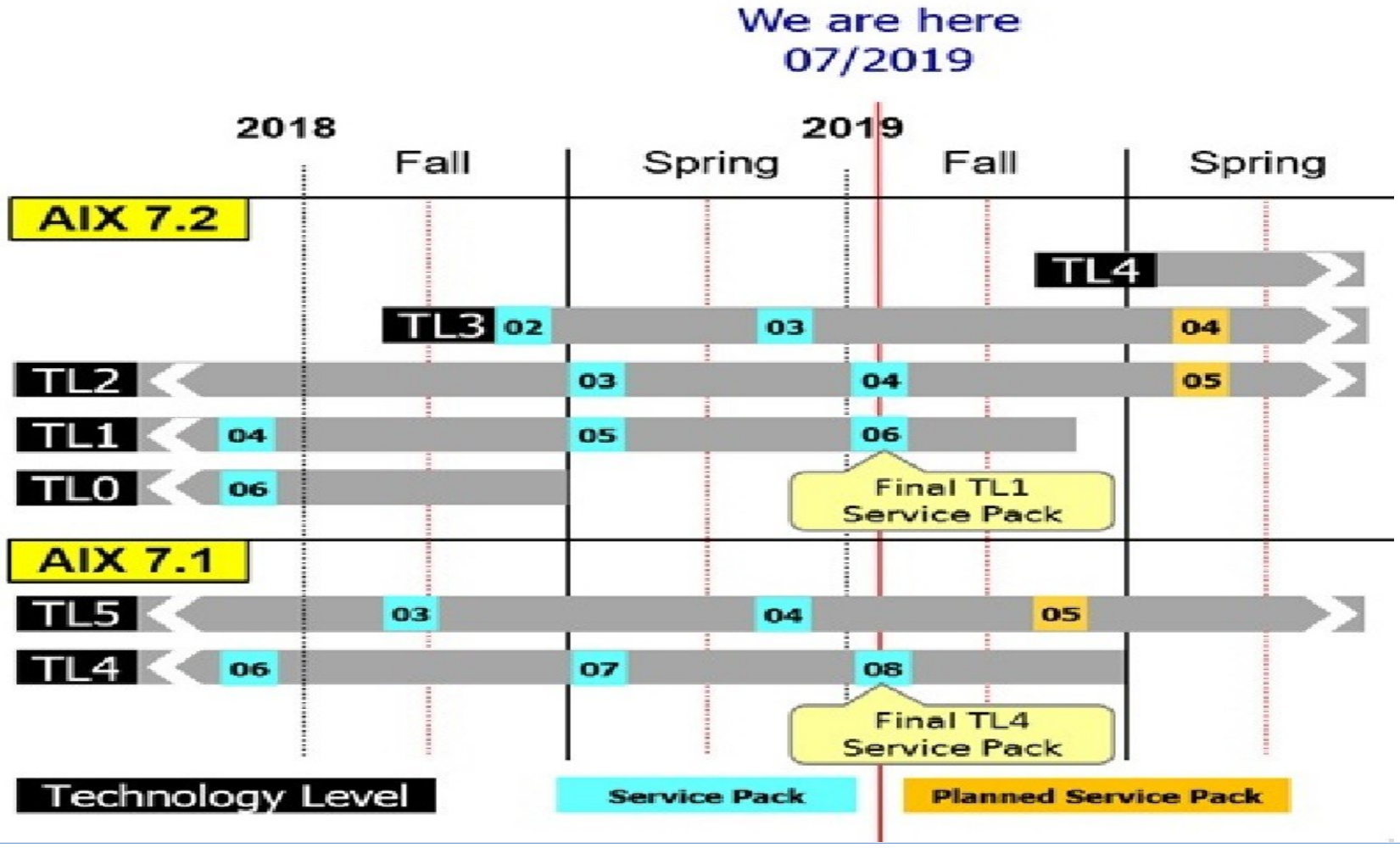
Remember that you may require a new boot image after applying a SP or TL – these can also be downloaded

- Interim fixes
  - Client at latest SP finds a problem
    - If problem resolved in existing SP, need to install that
    - If client cannot install SP, will get interim fix with recommendation to move to SP at next maintenance window
    - If client finds problem in supported level (>36 months) will get interim fix on their level if possible
    - If client has multiple interim fixes on a prior SP, will have to move up to later SP if those fixes are included
    - If client has multiple interim fixes on a prior SP, only some included in later SP, will have to move up and get new interim fix at that level

© 2020 Belisama

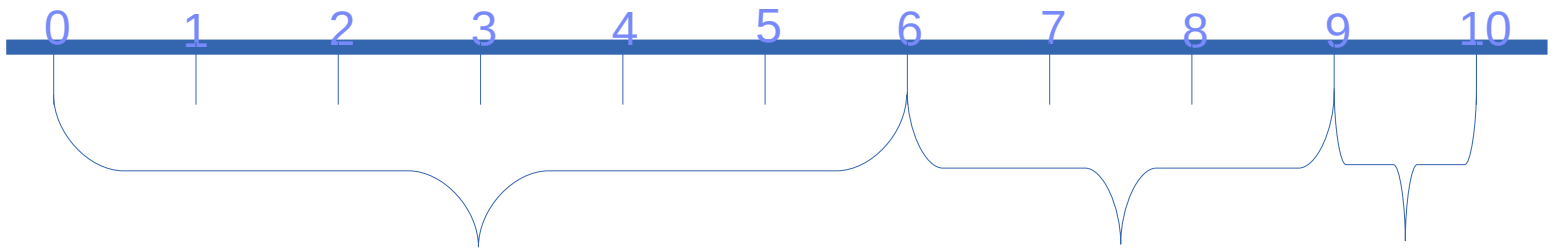
General OS strategies (cont)

Belisama



General OS strategies (cont)

Belisama



- Usage and new defects
- Phone and electronic service
- Standard SWMA required

- Usage and new defect
- Phone and electronic service
- Standard SWMA and extension contract required

- Via internet:
  - Existing defects
  - Online tech support data
  - Product publications
- No new defect or phone or electronic service available

© 2020 Belisama

General OS strategies (cont)

Belisama

	Old practice	New Method
length of Service for TL	12 months	36 months
Extended service via	CSP and IF	PTF, IF or SP
TL every	6 months	yearly
SP includes	only fixes	fixes and new h/w and s/w
SP shipped every	4-6 weeks	8-12 weeks
H/W support via	latest TL only	latest TL and some SP
VRMF	6.1.0.x	7.1.1.x

© 2020 Belisama

General OS strategies (cont)

Belisama

- Apply all of TL / SP
- Naming convention
  - 6100-06-03-1048
  - AIX_Rel – TL – SP – year week
- Technology Levels
  - Support for new hardware including exploitation of new hardware features
  - Support for new software enhancements
  - CSM (except for AIX 71) and RSCT fileset updates
  - Customer reported problems (APARs)
  - Critical problems found by development or test teams
- If new feature introduced, configured for previous behaviour

© 2020 Belisama

Is there an order when updating?

Belisama

- What is the order I should update my system
  - HMC
  - System microcode
  - VIO servers
  - AIX
    - (may have application dependencies)
  - GPFS, PowerHA etc

Check readme's and dependencies for each step

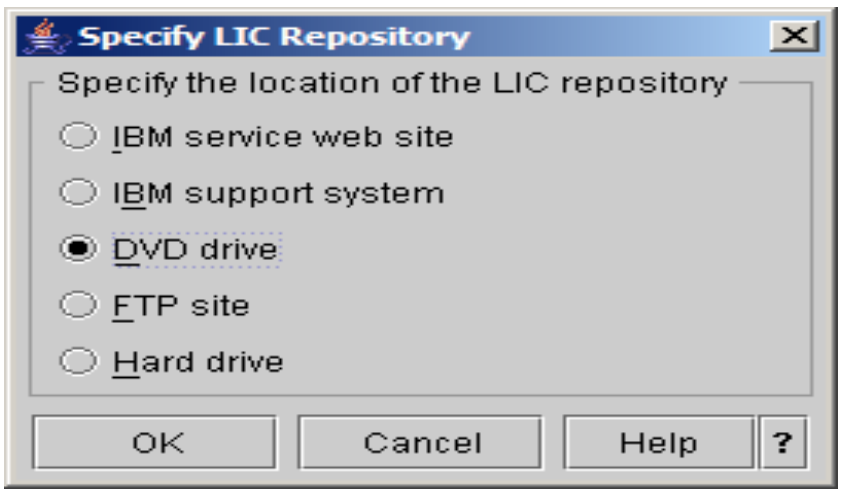
Code update readiness checker

Belisama

- Check for anything that could cause the code update to fail - before attempting!
  - Network connections
  - Pending serviceable events
- Code Update Readiness Check function in HMC
  - Analyse system for problems that will prevent success
  - Inform operator of problems to be corrected
  - Many of these conditions will not inhibit normal system operation, but will prevent a successful code update
- Run Code Update Readiness Check in Advance
  - We recommend to run readiness checker one week in advance of code update to allow time to resolve errors if any are found
  - These must be resolved before code update

Code update readiness checker (cont).

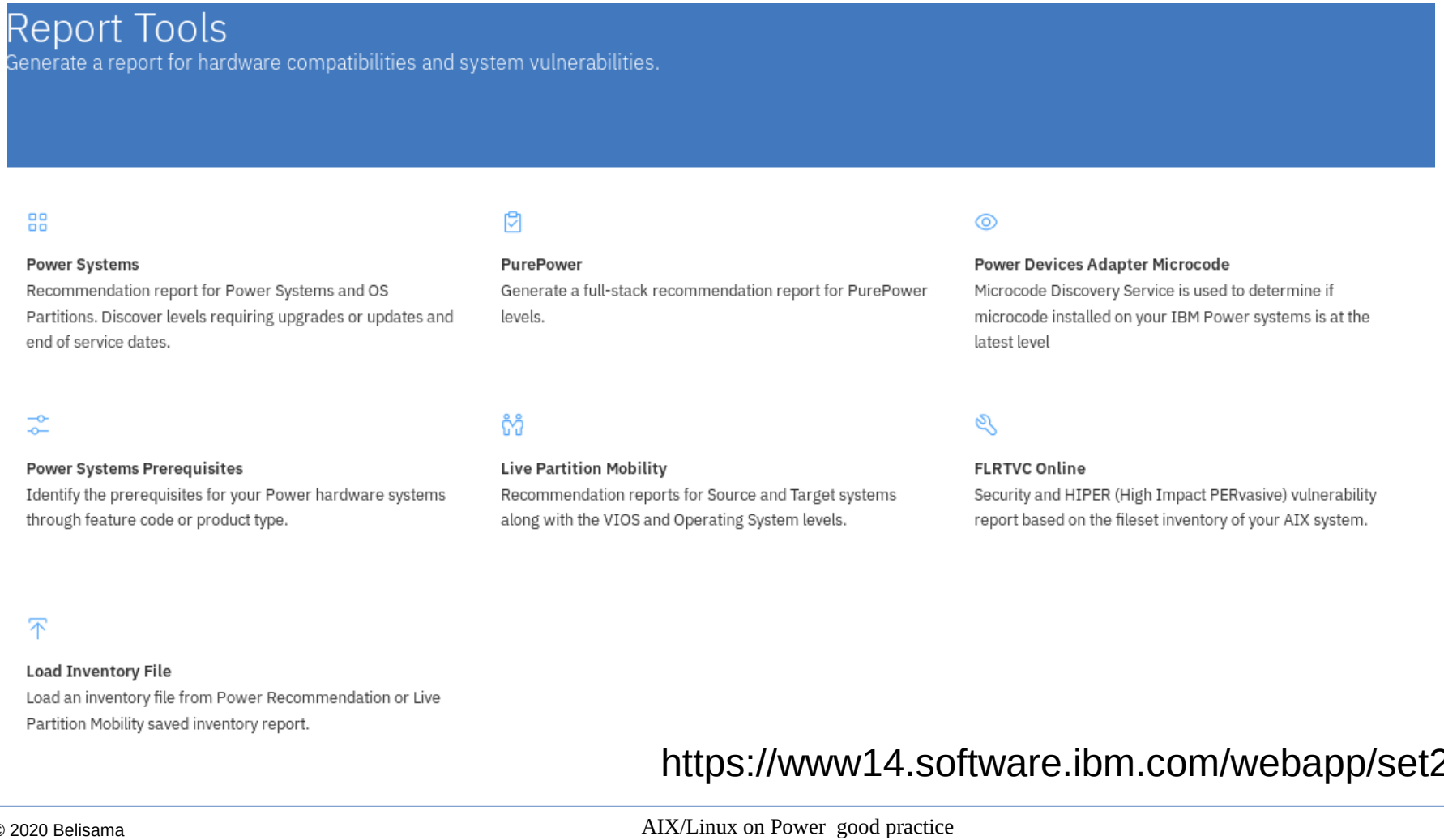
- How to Run Readiness Check in Advance
  - Change Licensed Internal Code for Current Release
  - Select target
  - Start Change Licensed Internal Code Wizard
  - If you reach “Specify LIC Repository” panel, the readiness checker has passed – select Cancel



© 2020 Belisama

IBM Report tools

Belisama



https://www14.software.ibm.com/webapp/set2/flrt/#reports

Power code matrix

POWER Code Matrix

Supported code combinations for IBM Power Systems

IBM Fix Central provides access to product updates, fixes and readme files. Readme files contain support information. Please update your bookmark to Fix central for product downloads and readme information.

→ IBM Fix Central

For AIX, VIOS and IBM i prerequisite information by machine type model, please visit the Systems Software Maps page.

Verify HMC and server firmware combinations for IBM Systems that use POWER processors.

Supported code combinations for HMC and server firmware
View a matrix of POWER HMC and system firmware levels. Use these tables to find supported combinations of HMC and system firmware levels. See the Terminology link below for a key to reading the matrix tables and a glossary of terms.

→ View Hardware Machine Console (HMC) HW Listing

→ View Terminology

POWER9 | POWER8 | POWER7 | POWER6 | POWER5 | PureFlex | Blades

Search:

Machine-Type-Model	Family Name	Power Code Matrix
9008-22L	IBM Power System L922	→ View
9009-22A	IBM Power System S922	→ View
9009-41A	IBM Power System S914	→ View
9009-42A	IBM Power System S924	→ View
9040-MR9	Power Systems E950	→ View
9080-M9S	Power Systems E980	→ View
9223-22H	IBM Power System H922	→ View
9223-42H	IBM Power System H924	→ View

https://www14.software.ibm.com/webapp/set2/flrt/mtm

© 2020 Belisama

Power code matrix (cont).

Belisama

FLRT > Power Code Matrix >

Power systems HMC and firmware

For Server MTM 9009-22A

Server MTM 9009-22A	System firmware
HMC and HMC Virtual Appliance	VL910_073 VL910_089 VL910_107 VL910_115 VL910_122 VL910_127 VL910_135
V9 R1 M910 V9 R1 M911 V9 R1 M920 V9 R1 M921 V9 R1 M930 (1) (2)	Supported

Matrix key

- Within six months of End of Service Pack Support
- End of Service Pack Support
- Reduced Fix support

1. **V9 R1 M910, V9 R1 M911, V9 R1 M920, V9 R1 M921, V9 R1 M930 -** Note: Starting with HMC R870 additional options are available to run HMC on Power machines either directly or as a virtual appliance. To download HMC to run on Power 7063 machine types, visit IBM Fix Central here.
2. **V9 R1 M920 -** Note: POWER6 servers or HMC hardware models may impact upgrade, check readme documentation for more information.

For Server MTM 8408-44E

Server MTM 8408-44E	System firmware
HMC and HMC Virtual Appliance	SV860_056 SV860_063 SV860_070 SV860_082 SV860_103 SV860_109 SV860_118 SV860_127 SV860_138 SV860_160 SV860_165 SV860_180 SV860_205
V9 R1 M910 V9 R1 M911 V9 R1 M920 V9 R1 M921 V9 R1 M930 (1) (2)	Supported
V8 R870 V8 R870 SP1 V8 R870 SP2 V8 R870 SP3 (1) (3)	Supported
V8 R860 V8 R860 SP1 V8 R860 SP2 V8 R860 SP3 (3)	Supported

IBM Pre-requisite tool

Download Prerequisites

Choose a machine type and model

9179-MHC

Select one, or some, or all Feature Codes

All Feature Codes

Choose an Operating System (optional)

AIX

SUBMIT CLEAR

Date: 01 Aug 2019
Machine Type Model: 9179-MHC
Operating System: AIX
Prerequisite(s): 1141 found
Feature codes (188): 0265, 0551, 0553, 0555, 1104, 1106, 1107, 1738, 1751, 1752, 1768, 1769, 1775, 1790, 1808, 1815, 1817, 1818, 1819, 1844, 1866, 1868, 1869, 1880, 1882, 1883, 1884, 1885, 1886, 1890, 1891, 1899, 1912, 1914, 1917, 1925, 1927, 1928, 1929, 1947, 1948, 1953, 1956, 1962, 1964, 1995, 2055, 2146, 2728, 2738, 2849, 2893, 2894, 2943, 2947, 2962, 3278, 3279, 3287, 3288, 3289, 3290, 3293, 3450, 3451, 3452, 3453, 3454, 3455, 3456, 3585, 3586, 3646, 3647, 3648, 3649, 3689, 4764, 4791, 4808, 4809, 5003, 5285, 5287, 5288, 5289, 5564, 5600, 5601, 5602, 5646, 5647, 5652, 5662, 5665, 5700, 5701, 5706, 5708, 5713, 5714, 5716, 5717, 5721, 5722, 5723, 5729, 5732, 5735, 5736, 5740, 5741, 5742, 5744, 5745, 5748, 5749, 5758, 5759, 5762, 5767, 5768, 5769, 5771, 5772, 5773, …08, 5912, 5913, …35, 7545, 7547, …J28, EN01, EN02,

Prerequisites

Search:

Feature Code	Product Name	Release	Fix PTF	External Info
0265	AIX 5L Version 5.3	Base AIX 5.3 Levels for 8233-E8B; 8236-E8C; 9117-MMB; 9179-MHB: 5300-11 SP2; 5300-10 SP4; 5300-09 SP7		
0265	AIX Version 6.1			
0265	AIX Version 6.1	Base AIX 61 Levels for 8233-E8B; 8236-E8C: 6100-04 SP2; 6100-03 SP5; 6100-02 SP8.		
0551	No prerequisites required	N/A		
1104	AIX Version 7.1	7100-00		
1104	Server Firmware	Base Server Firmware Level for 8233-E8B; 8236-E8C: FW 7.1.0	AL710_043	
1104	Server Firmware	FW 7.2.0	AL720_066, AM720_064, AH720_064	
1104	HMC Firmware	Base HMC Firmware Level for 8231-E2B; 8202-E4B; 8205-E6B and 9119-FHB: V7R7.2.0M0	MH01235	

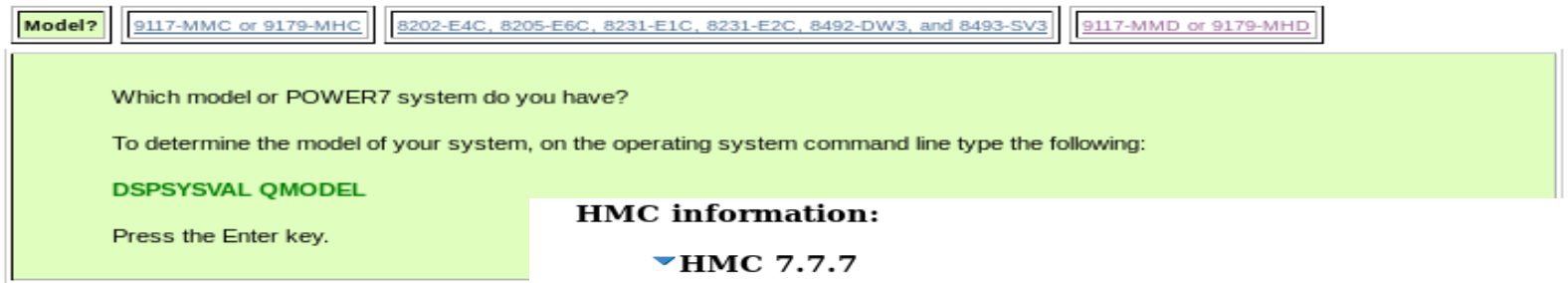
https://www14.software.ibm.com/webapp/set2/iprt/home

© 2020 Belisama

Another useful tool (thanks IBM I team)

Belisama

- IBM Server Firmware and HMC Code Wizards
  - http://www-912.ibm.com/s_dir/slkbase.NSF/docNumber/408316083



HMC information:

HMC 7.7.7

HMC 7.7.7
For HMCs at 7.7.7 the following document provides a list of suggested updates. To apply the update(s) on the HMC, refer to Rochester Support Center knowledgebase document 661704005, **Update(s) for HMC 7.7.7, To link to document 661704005 immediately, click here**,

For HMCs at 7.3.x or 7.7.x upgrading to 7.7.7
To upgrade the HMC to 7.7.7, refer to Rochester Support Center knowledgebase document 661742145, **Upgrading the HMC from Version 7.3.x or Version 7.7.x to Version 7.7.7. To link to document 661742145 immediately, click here**.

For HMCs at 6.1.3 and Earlier upgrading to 7.7.7
To upgrade the HMC to 7.7.7, refer to Rochester Support Center knowledgebase document 661702731, **Upgrading the HMC from Version 6.1.3 and Earlier to Version 7.7.7. To link to document 661702731 immediately, click here**.

© 2020 Belisama

Electronic Service Agent (ESA) Call home

Belisama

- The Electronic Service Agent™ is a "no-charge" software tool that resides on your system p servers to monitor events.
  - ESA is able to automatically report hardware problems.
  - This proactive tool enables support to arrive on-site with the knowledge and parts required to resolve issues quickly.
  - We recommend our clients utilise this “Phone home” capability
  - HTTP Proxy Support

Fix Level Recommendation Tool (FLRT)

- Initial release enabled customers to obtain recommended minimum fix levels for key components of IBM System p5 servers.
  - System Firmware
  - Hardware Management Console
  - Virtual I/O Server virtualisation partition
  - AIX 5L operating system
- We are consistently expanding this tool to support more IBM products.
  - High Availability Cluster Multi Processor (HACMP)
  - Customer Systems Management (CSM)
  - Parallel Environment
  - General Parallel File System
  - Others

A report providing customers with a quick reference to a minimum set of IBM recommendations to better prevent system outages.

© 2020 Belisama

Fix Level Recommendation Tool (cont).

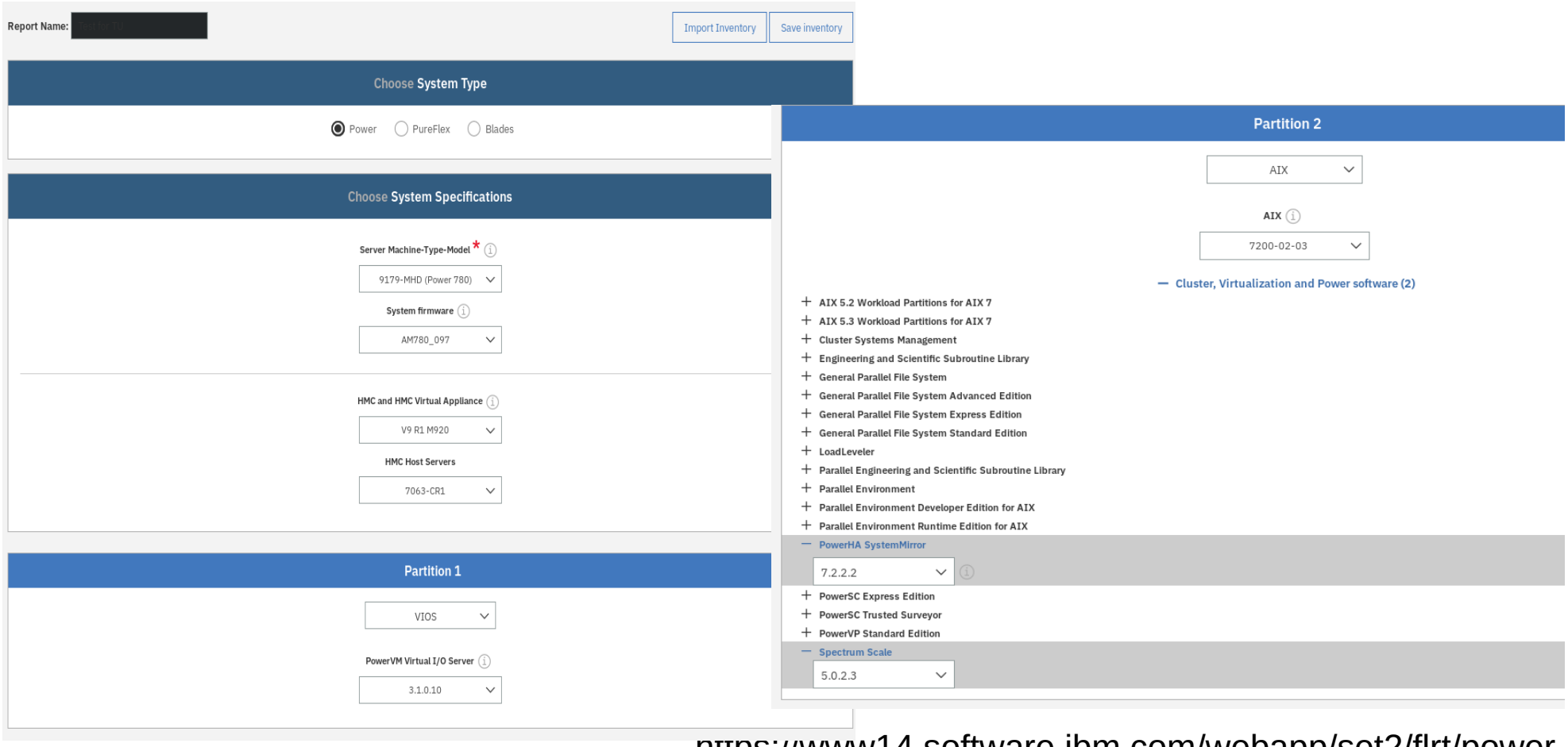
Belisama

- Highlights of FLRT
  - Scripting enabled to evaluate current fix levels
  - Easy to create and understand reports
  - Useful for “what if” planning needs
  - Links to fix distribution sites
  - Print friendly view provides printable report for maintenance planning
  - Option to manually determine fix levels for all support products for clients who do not wish to use automated determination
  - Easily obtainable tool from all fix distribution sites

© 2020 Belisama

Fix Level Recommendation Tool (cont).

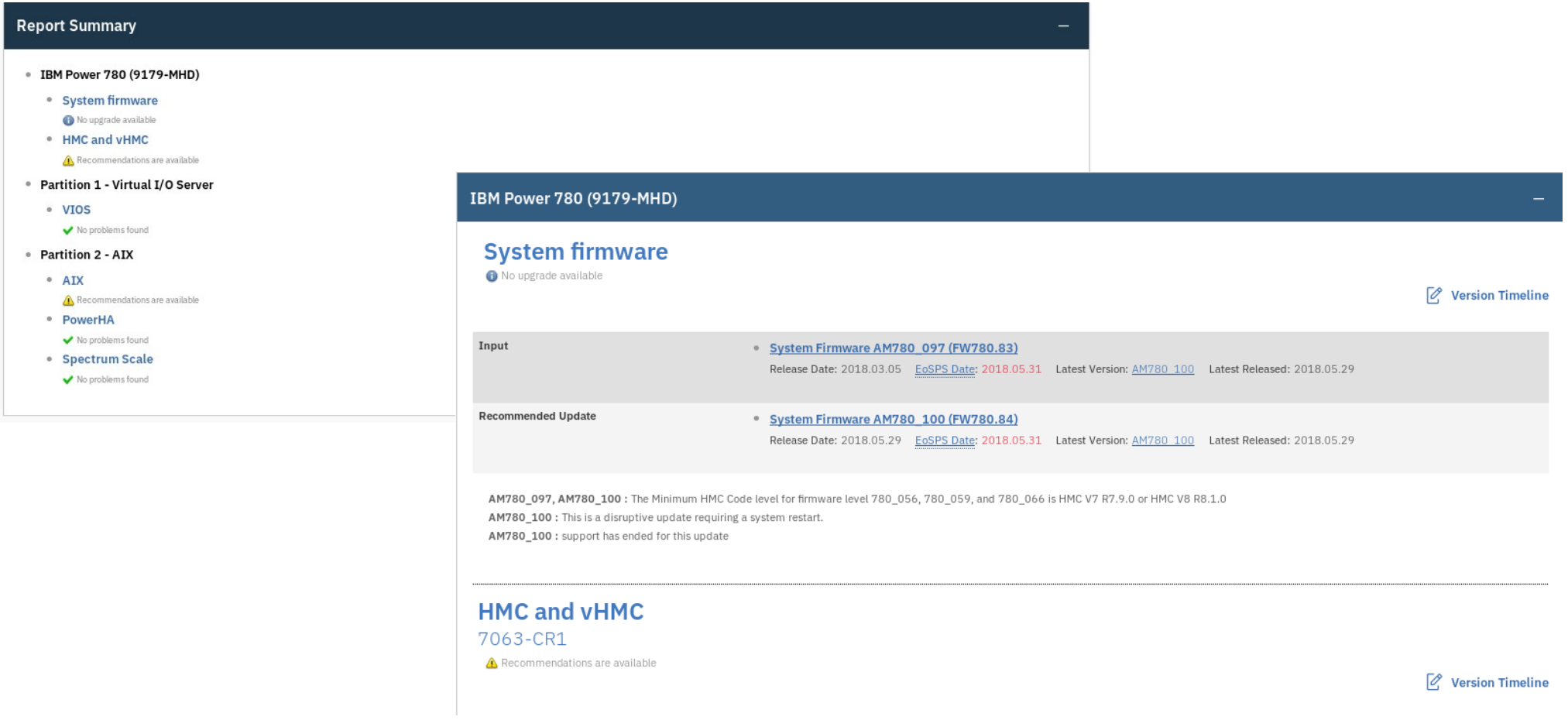
Belisama



https://www14.software.ibm.com/webapp/set2/flrt/power

Fix Level Recommendation Tool (cont).

Belisama



Use the performance Advisor tools

Belisama

- PowerVM Virtualization Performance LPAR Advisor
  - https://www.ibm.com/developerworks/community/wikis/home?lang=en#!/wiki/Power%20Systems/page/PowerVM%20Virtualization%20Performance%20Advisor

The PowerVM Virtualization Performance Advisor assists diagnosis of performance issues and provide recommendations - can be run on any AIX partition and supports POWER8 and has a new Network Advisor. The Advisor runs on the AIX partition reports on system configuration details, partition configuration details and recommendations to improve performance based on the analysis done

- VIOS Advisor
  - https://www.ibm.com/developerworks/mydeveloperworks/wikis/home?lang=en#/wiki/Power%20Systems/page/VIOS%20Advisor
- Java Performance Advisor
  - https://www.ibm.com/developerworks/mydeveloperworks/wikis/home?lang=en#/wiki/Power%20Systems/page/Java%20Performance%20Advisor%20%28JPA%29

No longer supported
– however works on AIX 6.1/7.1

part -i 60

VIOS - Processor

Risk/Impact 1=lowest 5=highest

Name	Measured Value	Suggested Value	First Observed	Last Observed	Risk	Impact
CPU Capacity	2.0 ent		03/09/2016 12:04 AM			
CPU consumption	avg:0.3% high:2.0%		03/09/2016 12:04 AM	03/09/2016 12:15 AM		
Processing Mode	Dedicated CPU		03/09/2016 12:04 AM			
Processor Donation	off	on	03/09/2016 12:04 AM	03/09/2016 12:15 AM	3	3
SMT Mode	SMT4		03/09/2016 12:04 AM			

VIOS - Memory

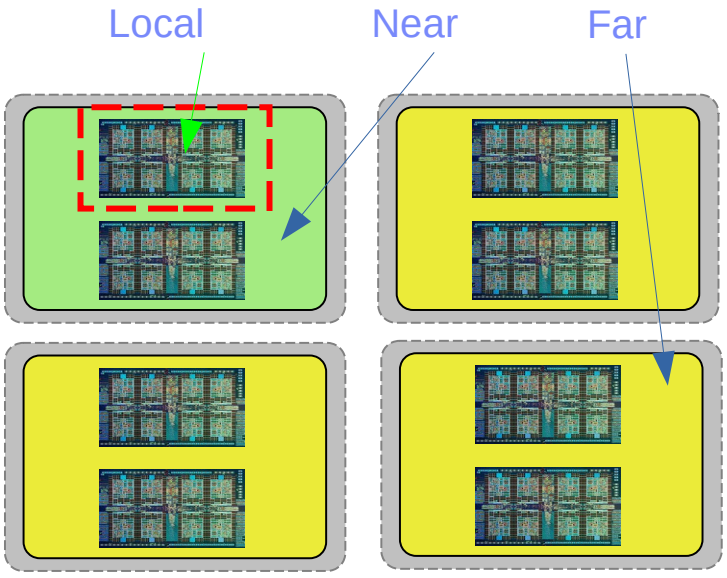
Risk/Impact 1=lowest 5=highest

Name	Measured Value	Suggested Value	First Observed	Last Observed	Risk	Impact
Real Memory	6.000 GB		03/09/2016 12:04 AM			
Available Memory	3.463 GB		03/09/2016 12:04 AM			
Paging Rate	0.0 MBps Paging Rate		03/09/2016 12:04 AM	03/09/2016 12:15 AM		
Paging Space Size	1.500 GB		03/09/2016 12:04 AM	03/09/2016 12:15 AM		
Free Paging Space	1.486 GB free		03/09/2016 12:04 AM			
Pinned Memory	1.086 GB pinned		03/09/2016 12:04 AM			

© 2020 Belisama

What / Why / Wherefore DPO (Dynamic Platform Optimizer)

- During a platform IPL of a System, the Hypervisor determines an optimal resource placement strategy for the server based on the partition configuration and hardware topology on the system. (POWER7 & POWER8)
- Partition placement can become sub-optimal because of:
  - Dynamic creation and deletion of partitions
  - DLPAR operations that add and remove processors or memory
  - Making changes to a partition (processor or memory configuration changes)
  - Hibernation (Suspend or Resume)
  - LPM – Live Partition Mobility
  - CHARM – CEC Hot add and Repair Maintenance (This is a Node operation)



https://www.ibm.com/support/knowledgecenter/TI0003N/p8hat/p8hat_dpoovw.htm

© 2020 Belisama

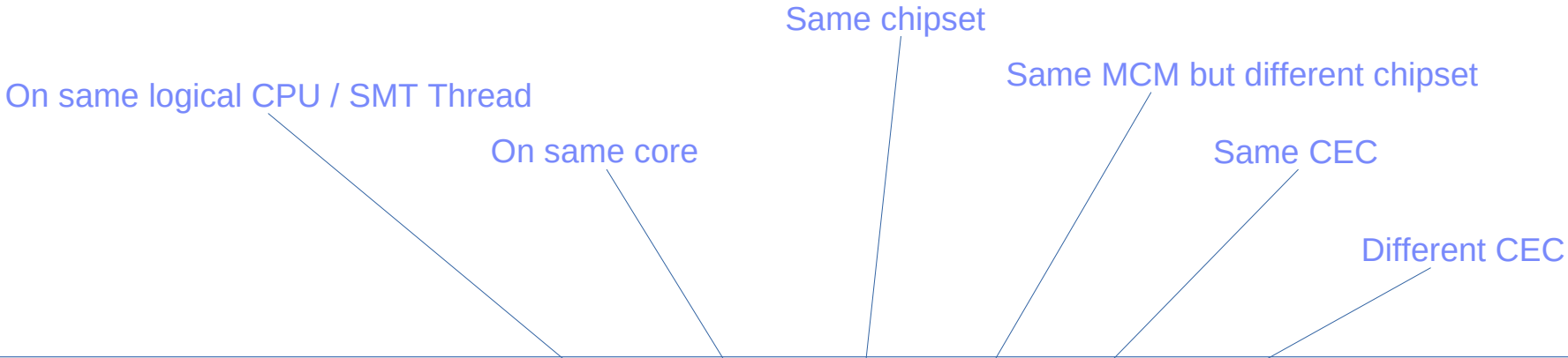
DPO (cont.)

- DPO is a PowerVM virtualization feature that enables users to improve partition memory and processor placement (affinity) on Power Servers after they have been up and running.
- DPO performs a sequence of memory and processor relocations to transform the existing server layout to the optimal layout based on the server topology.
- Client Benefits
  - Ability to run without a platform IPL
  - Improved performance in a cloud or highly virtualized environments
  - Dynamically adjust topology after mobility
- Commands
  - *lsmemopt*: An affinity score is a measure of the processor-memory affinity on the system or for a partition in the range 0 - 100
  - *optmen*: you can schedule DPO operations on POWER7 or POWER8 (HMC > 7.8.0)

© 2020 Belisama

Managing affinity - mpstat

- mpstat output explained




```
 mpstat -d

System configuration: lcpu=8 ent=0.2 mode=Uncapped

 cpu     cs    ics       bound    rq   push S3pull  S3grd  S0rd  S1rd  S2rd  S3rd  S4rd  S5rd    ilcs      vlcs   S3hrd S4hrd S5hrd
  0 230011156 1773410      1      3      0      0      0  99.9   0.0   0.0   0.1   0.0   0.0 375287   361488328 100.0   0.0   0.0
  1 303083     241957      0      0      0      0      0  28.8  71.2   0.0   0.0   0.0   0.0 123188    56199371  100.0   0.0   0.0
  2 206811     204996      0      0      0      0      0   0.8  99.2   0.0   0.0   0.0   0.0 111429    55916627  100.0   0.0   0.0
  3 206311     205009      0      0      0      0      0   0.7  99.3   0.0   0.0   0.0   0.0  13769    55957608  100.0   0.0   0.0
  4 284988     215902      0      0      0      0      0  25.4   0.9   0.0  73.7   0.0   0.0   9872    67933806  100.0   0.0   0.0
  5 233186     208933      0      0      0      0      0  11.7  88.3   0.0   0.0   0.0   0.0 33038945  193455    100.0   0.0   0.0
  6 205672     205070      0      0      0      0      0   0.6  99.4   0.0   0.0   0.0   0.0 33038032  185484    100.0   0.0   0.0
  7 205306     205030      0      0      0      0      0   0.4  99.6   0.0   0.0   0.0   0.0 33044293  178363    100.0   0.0   0.0
ALL 231656513 3260307      1      3      0      0      0  99.3   0.5   0.0   0.2   0.0   0.0 99754815  598053042 100.0   0.0   0.0
```

© 2020 Belisama

Managing VP and SMT - mpstat

- Display Virtual Processor and SMT activity on an existing workload with -v flag
  - This option displays the actual Virtual Time Base (VTB) – the dispatch time for each Virtual Processor at the physical layer, physical consumption (pc) and the activity of the SMT

vcpu	lcpu	us	sy	wa	id	pbusy	pc	VTB(ms)
0		2.68	18.80	0.00	78.52	0.00[21.5%]	0.00[0.3%]	19
	0	2.68	16.28	0.00	20.92	0.00[19.0%]	0.00[39.9%]	-
	...							
1		58.97	0.02	0.00	41.01	0.59[59.0%]	1.00[99.9%]	4995
	4	58.97	0.01	0.00	0.00	0.59[59.0%]	0.59[59.0%]	-
	5	0.00	0.00	0.00	13.67	0.00[0.0%]	0.14[13.7%]	-
	6	0.00	0.00	0.00	13.67	0.00[0.0%]	0.14[13.7%]	-
	7	0.00	0.00	0.00	13.67	0.00[0.0%]	0.14[13.7%]	-

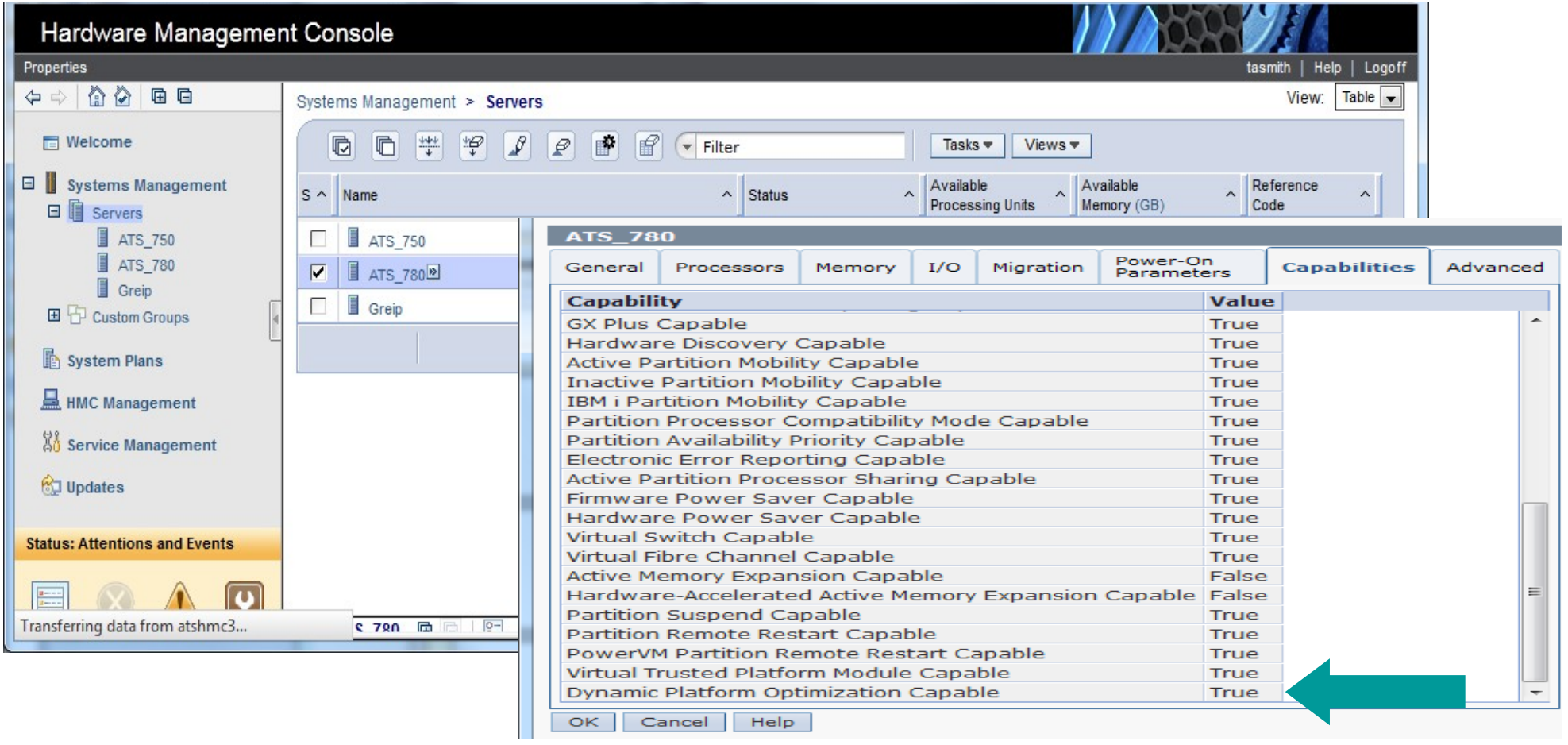
How many VPs are actually dispatching

Dispatch time in milliseconds

© 2020 Belisama

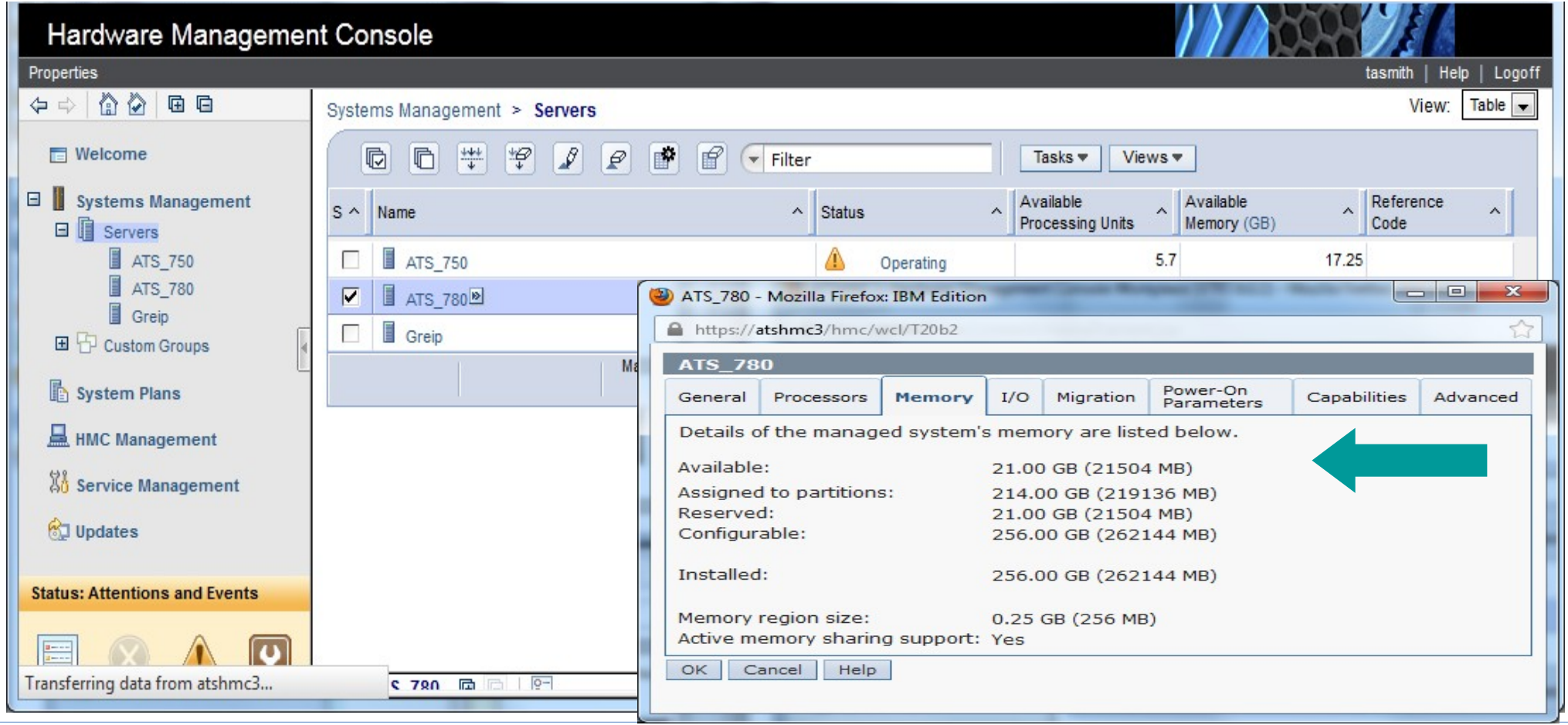
System supports DPO

Belisama



HMC view of free memory blocks (System properties)

Belisama



© 2020 Belisama

Managing affinity - lssrad

- # lssrad -av

```
REF1   SRAD       MEM      CPU
0
          0   12363.94      0-7
          2    4589.00     12-15
1
          1    5104.50      8-11
          3    3486.00     16-19
```

- REF‘s Sockets/Nodes/Dual Chip Module (DCM)
- SRAD Scheduler Resource Allocation Domain
  - 0 & 2 are two chips in the first DCM
  - 1 & 3 belong to other DCM
- If a thread’s ‘home’ node was SRAD 0
  - SRAD 2 would be ‘near’
  - SRAD 1 & 3 would be ‘far’

When may I have a problem?
- . SRAD has CPUs but no memory or vice-versa
- . When CPU / Memory is very unbalanced

But how do I really know?
- . Tools tell you!
- . Users complain
- . Disparity in performance between equivalent systems

In the real world, SRADs will never be perfectly balanced and many workloads do not care

© 2020 Belisama

Service and support

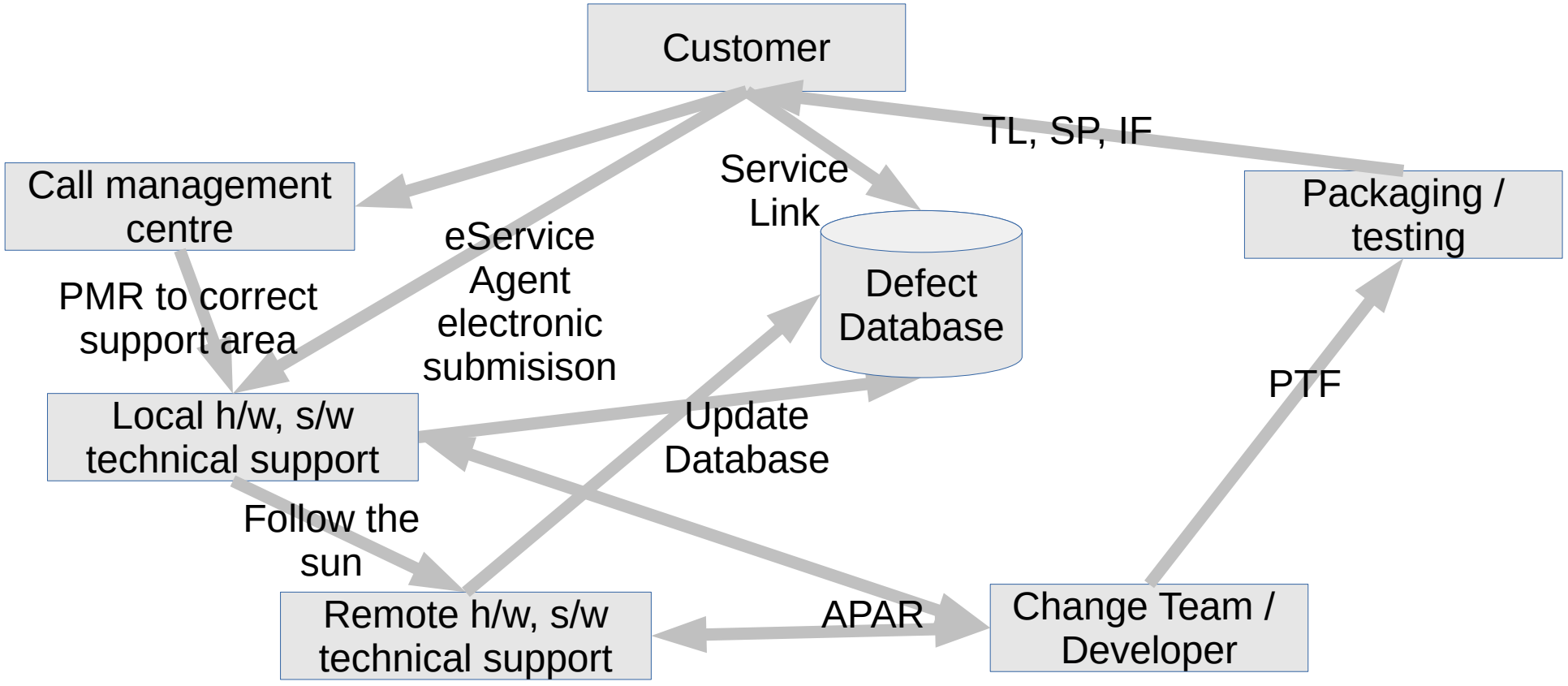
- Support database shows:
  - A large percentage of problems that cause outages are re-occurrences of a problem that a fix has already been developed.
  - A significant number of calls are closed due to failure to contact
  - PMR analysis shows that significant amount of time is lost due to:
    - inability to contact decision maker
    - inability to get access to system (physically or logically)
    - Inaccurate capture/reporting of error
  - A large percentage of calls are delayed because the testcase is invalid or in the wrong place.
  - Try not to lead support staff to your conclusions
  - Have you tested the support process (are your details in the IBM database correct?)

Place your call electronically:
https://www-946.ibm.com/support/servicerequest/Home.action
(linkadmn@au1.ibm.com)

© 2020 Belisama

Service and support (cont)

Belisama



Service and support (cont)

Belisama

- What to do before calling support (or electronic call creation)
  - Check what has changed
  - Are there updated firmware, BIOS, fixes for your problem?
  - Where does the problem most likely lie?
    - System p h/w or s/w ; storage; communications; application; peripheral?
    - Check logs, perform initial PD
  - Step back.... what data required ? – start collecting
  - Do you have the authority to collect and send required data?
  - Is your system documentation available
  - Your customer number, Serial number for affected systems
  - Other vendor details (Storage, switches)
  - Internal processes, timelines
  - Start record of events and times, confirm with business or application owners the 'must be back' time

Test the support process!

Summary

- Although this is by no means a complete and comprehensive list, this should give you lots of first thoughts about how to manage AIX.
- The most important thing to remember on monitoring – set your system up to warn you of impending problems before your customers tell you about them.
- Remember that if you manage a UNIX box like a mainframe, you work towards mainframe-like availability. If you manage it like a PC …….

© 2020 Belisama

Some other hints

- Create a SYSTEMBOOK to document your system and your procedures for startup and shutdown (OS, Applications, Virtualised environment)
- Run FSCK periodically when you have a maintenance window
- Failover HACMP (forward and back) during a maintenance window
- Run a checksum (lppchk) after your installs and also
  - Check (instfix –icqk 5300-05_AIX_ML | grep “:-:”)
  - oslevel -l | g <OS_Level> -s

© 2020 Belisama

Components	Description	Editions	Prerequisites
Security and Compliance Automation *	Automates the setting, monitoring, and auditing of security and compliance configuration for the following standards: - Payment Card Industry Data Security Standard v1.2 (PCI DSS) - Sarbanes-Oxley Act and Control Objectives for for Information and related Technologies compliance (SOX/COBIT) - U.S. Department of Defense (DoD) Security Technical Implementation Guide for Unix (STIG) - Health Insurance Portability and Accountability Act (HIPAA)	PowerSC Express Edition PowerSC Standard Edition	POWER5, 6, 7, 8 & 9
Real Time Compliance	Monitors an enabled AIX system to maintain security and provides alerts when a change to the system violates a rule that is identified in the configuration policy.	PowerSC Express Edition PowerSC Standard Edition	There are no specific hardware requirements
Tusted Boot	Measures the boot image, operating system, and applications, and attests their trust by using the virtual trusted platform module (TPM) technology.	PowerSC Standard Edition	POWER7 firmware eFW7.4, or later
Trusted Firewall	Saves time and resources by enabling direct routing across specified virtual LANs (VLANs) that are controlled by the same Virtual I/O Server.	PowerSC Standard Edition	POWER6, 7, 8 & 9 Virtual I/O Server Version 2.2.1.4, or later

* Version 1.2 introduces monitoring of RHEL and SuSE endpoints

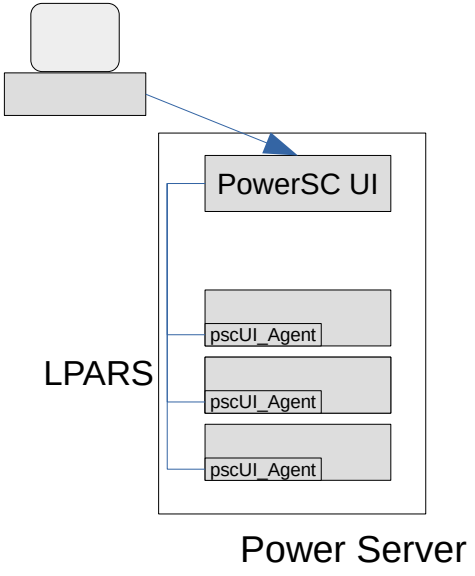
PCI document library: https://www.pcisecuritystandards.org/security_standards/documents.php
DoD directives library: http://www.dtic.mil/whs/directives/corres/dir.html
COBIT documentation: https://www.isaca.org/cobit/pages/default.aspx
HIPAA overview: http://www.hhs.gov/ocr/privacy/

Components	Description	Editions	Prerequisites
Trusted Logging	The logs of AIX are centrally located on the Virtual I/O Server (VIOS) in real time. This feature provides tamperproof logging and convenient log backup and management.	PowerSC Standard Edition	POWER5, 6, 7, 8 & 9 Virtual I/O Server Version 2.2.1.0, or later
Trusted Network Connect and patch management	Verifies that all AIX systems in the virtual environment are at the specified software and patch level and provides management tools to ensure that all AIX systems are at the specified software level. Provides alerts if a down-level virtual system is added to the network or if a security patch is issued that affects the systems.	PowerSC Standard Edition	POWER5, 6, 7, 8 & 9 Virtual I/O Server Version 2.2.1.0, or later
Trusted Surveyor	Monitors virtual network segregation policy compliance.	PowerSC Trusted Surveyor	There are no specific hardware requirements
MFA	PowerSC Multi-Factor Authentication (MFA) provides a method of controlling computer access in which a user is granted access only after successfully presenting several separate pieces of evidence to an authentication mechanism. Two factors options: • PIN-protected certificates on PIV/CAC smart cards • RSA SecurID	PowerSC MFA	IBM POWER7, 8 & 9 AIX 7.1 TL5 and above AIX 7.2 TL2 and above RHEL 7.4 and above SLES 12.3 and above

© 2020 Belisama

AIX runtime expert

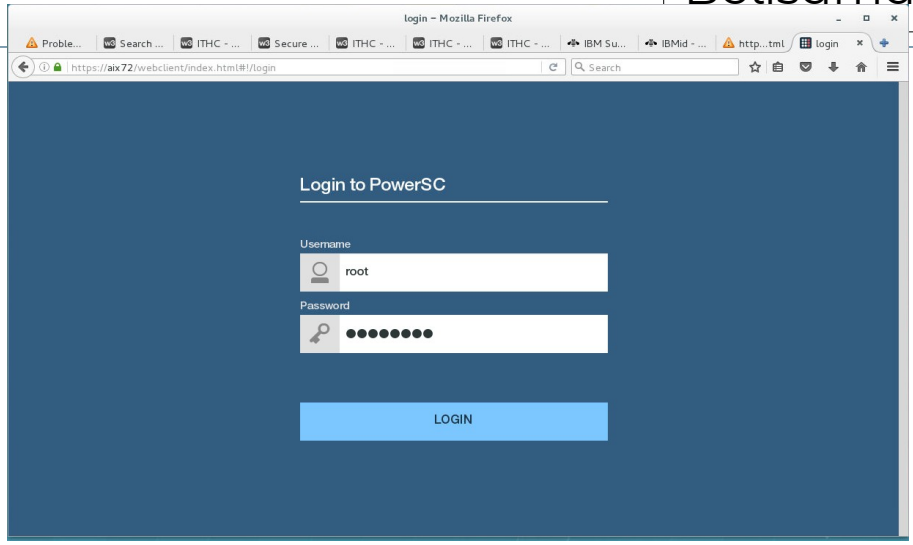
- The AIX Runtime Expert is intended to provide a simplified “one button” solution for managing runtime properties of one or more AIX instances.
- The approach of AIX Runtime Expert is to provide a framework with a standard data representation for runtime controls and a simplified set of actions that can be used against the representation for collecting, applying, and verifying the runtime environment.
- This approach avoids diversity of control points to manage runtime behaviour of AIX components and subsystems, which include configuration files, command lines, and environment variables.
- Standards
  - HIPAA (Healthcare); PCI (Financial & Retail); NERC (Utilities); DoD STIG (Federal); SOX-COBIT (General); and Custom Low, Medium and High Profiles for AIX (1.2 introduced GDPR)
- Compliance GUI
  - UI Server
    - AIX LPAR as a dedicated appliance server partition
  - UI Endpoint Agent
    - Monitoring
    - Command Execution
  - Browser
    - User interaction

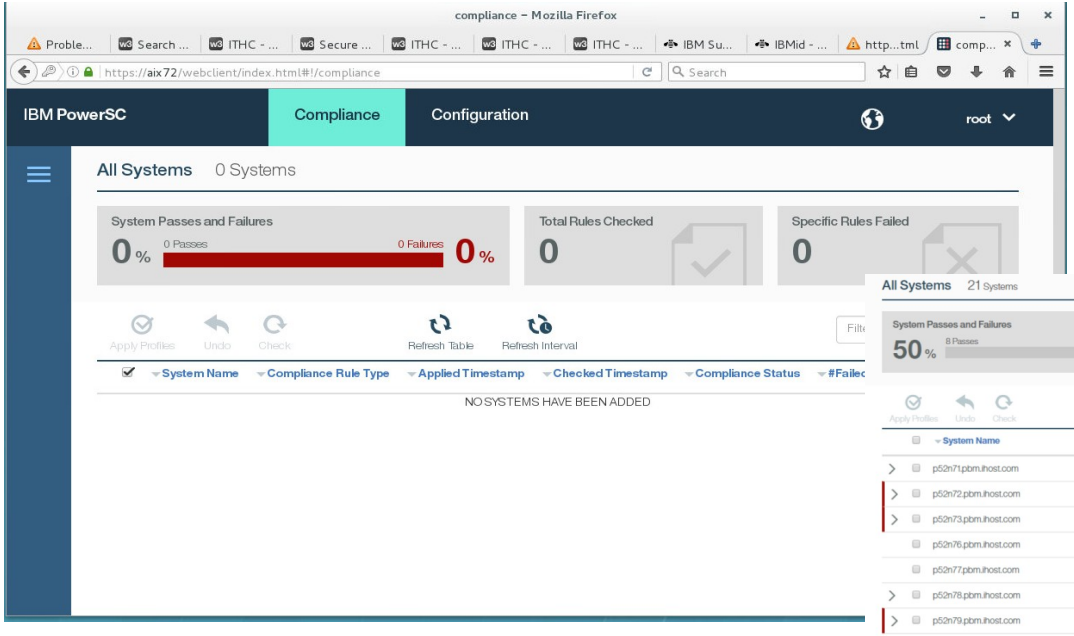


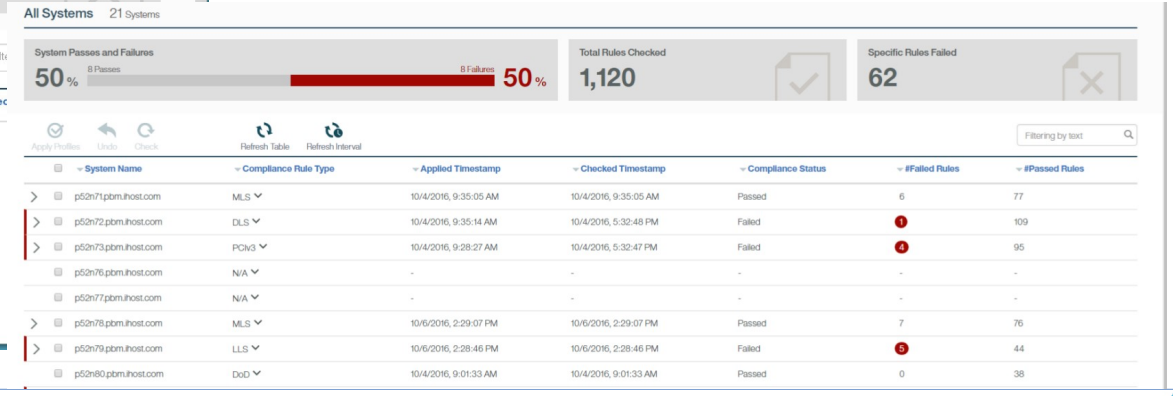
© 2020 Belisama

Compliance GUI - Security

- Requires AIX 7.2 or later
- SSL certificates used to verify connections
- Traffic between Server and endpoints is encrypted
- User authorisation local or LDAP
- Regular heartbeat between Server and all endpoints to keep
- Server up to date

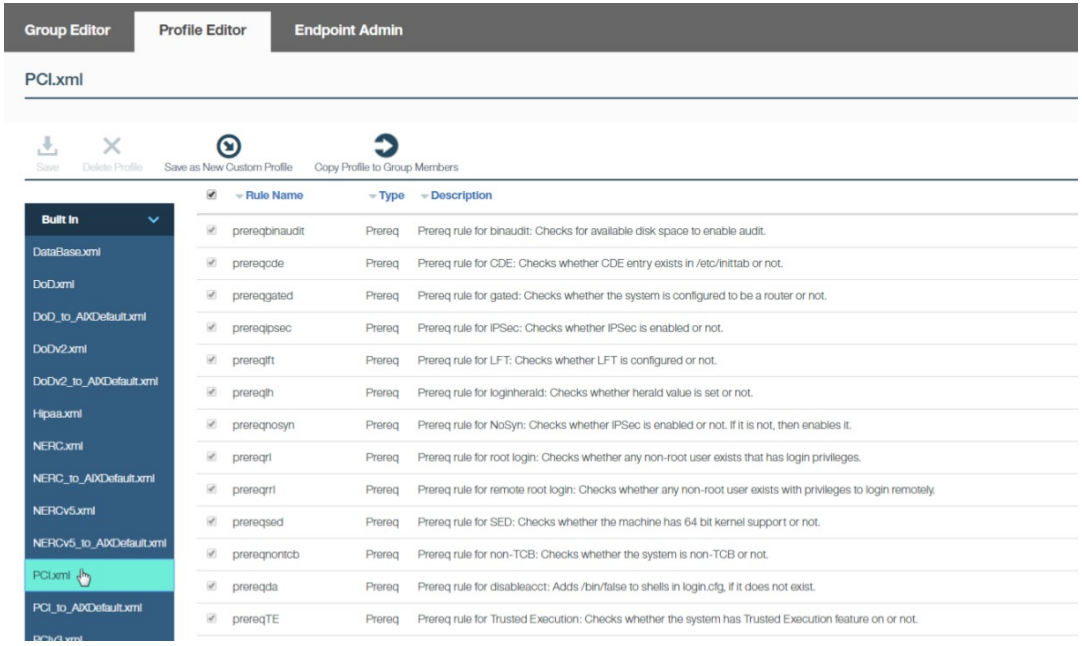






© 2020 Belisama

- Groups
  - Create group; select elements after sorting list by any status column
  - Edit group (add to, remove from, delete group)
- Profiles
  - Load an existing profile
  - De-select (remove) one or more rules
  - Save as a New Custom Profile
  - Copy a Custom Profile to all members of a Group
- Endpoint Administration
  - Delete one or more existing endpoints
  - Verify connectivity and security of one or more
  - endpoints
  - View AIX Group associations



© 2020 Belisama

PowerSC Multi Factor Authentication

Belisama

- Features
  - Two factors:
    - PIN-protected certificates on PIV/CAC smart cards
    - RSA SecurID
- Out-of-band MFA through a web interface for end users
  - End users go to the MFA web interface, provide factors required for their account, and receive a Login Token that they enter on their workstation.
- MFA through Pluggable Authentication Module (PAM)
  - End users plug their PIV/CAC card reader into the IBM Power server.
- Multiple concurrent logins (hundreds of users).
- Fast path for subsequent IBM AIX logins shortly after the previous successful full-fledged MFA authentication (configurable).
- All client-server communication encrypted (HTTPS/TLS).
- Centrally administer different factors for different user populations through a web interface:
  - Configuring MFA-wide settings
  - Configuring factor-specific settings
  - Enrolling users for MFA
  - Reviewing/approving user self-service enrollment events where applicable
  - Supports user self-service enrollment of certificates or smart cards through a web interface

© 2020 Belisama

Monitoring goal

- Set up monitoring on your system to ensure that you, the system's administrator are warning of impending problems and slow-downs BEFORE your customers tell you about them.
- Monitoring should be proactive and exception-based. When something is out of spec or out of norm, an alert should be sent rather than relying on review of logs or reports to assess after the fact.
- Tools
  - Tivoli
  - lpar2rrd
  - Ganglia
  - nmon and rrdt tool
  - njmon

© 2020 Belisama

Monitoring goal (cont).

- At the very least:
  - Ensure that your application logs are being passed to your monitoring automated alert system as well as the AIX error log
  - Look at the general monitoring recommendations in the RMC and ensure that your monitor of choice includes these things
  - Regularly check the Service Agent logs for anomolies
  - When you DO have a problem – ask yourself five questions:
    - Did my monitoring system catch this problem?
    - If it didn't why not? If it did – why didn't it warn me soon enough to prevent a catastrophic problem?
    - Have I changed my monitoring to catch this in the future?
    - Do I have OTHER systems that could potentially be hit by this error? (consider your other departments / geos too!)
      - Review the Problem determination / support process – was sufficient information captured in the logs to catch the problem.

© 2020 Belisama

Monitoring goal (cont).

- Something to look at
  - lscfg –v
  - lsconf
  - lsdev –CH
  - nmon
  - errpt for last 24 hours
  - vmtune | vmo|ioo
  - no –a
  - netstat –rn
  - df –k
- After each incident review monitoring – what do you need to add to be warned in the future
- Update after each change!

© 2020 Belisama

Things really worth doing

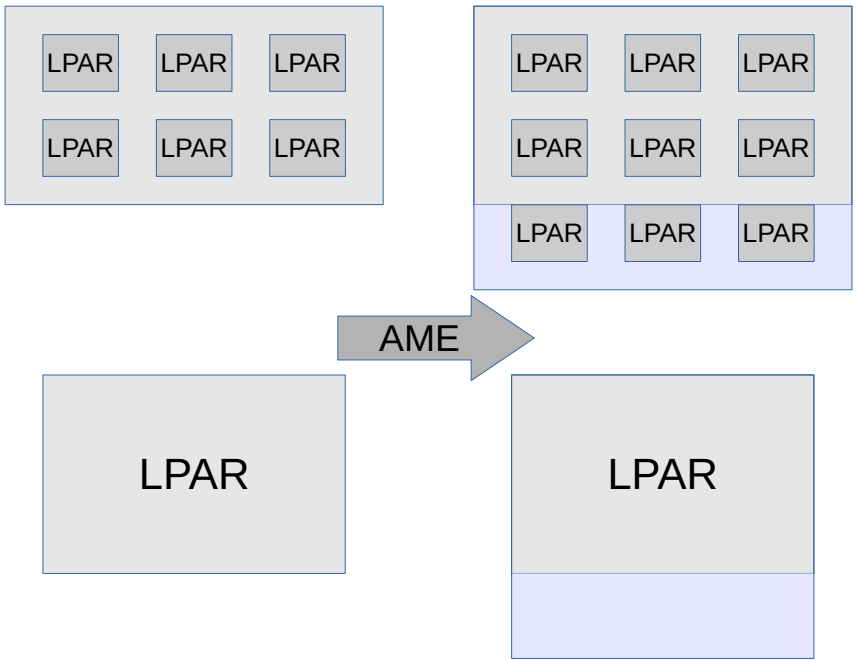
Belisama

- Sign up for cert advisories
- Turn on Service Agent
- Sign up for hyper-fix alerts
- Routinely review RML
- Use FLRT
- Use pre-req
- UPDATE YOUR FIRMWARE!!

© 2020 Belisama

Active Memory Expansion (AME)

- Compress a percentage of the LPAR memory
  - Uses
    - Increase consolidation – fit more lpars into the system by more effectively using the memory
    - LPAR increase – increase the effective memory size of an LPAR
    - One feature per server
    - One time, 60 day trial – no charge
      - www.ibm.com/systems/power/hardware/cod/
  - amepat tool to determine appropriateness for your environment



© 2020 Belisama

Active Memory Expansion (AME) (cont).

Belisama

- AMS relies on the compression of in memory data
  - Managed by the OS, so transparent to Users and Applications
  - Can be activated for individual LPARs
  - When enabled, memory is broken into two pools – 1 compressed and 1 uncompressed
  - The OS will uncompress and move data into the uncompressed pool when required.
  - When uncompressed pool is full, data will be compressed and moved into the compressed pool
  - The relative sizes of the pools is maintained automatically by the OS
  - One option used to configure – the memory compression factor
    - LPAR expanded memory size = LPAR true memory size * memory compression factor
    - For example, MCF = 1.5 and a 20GB system, the OS will attempt to compress enough in memory data to fit 30GB of data into the 20GB of memory.

© 2020 Belisama

Active Memory Sharing (AMS)

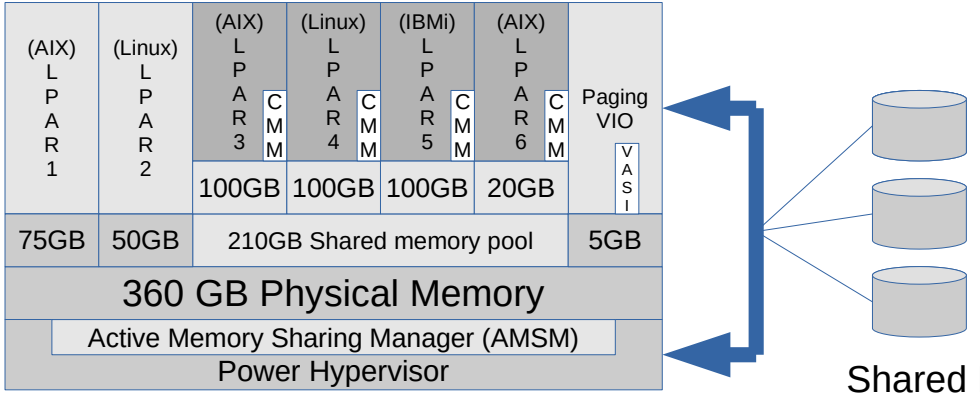
- Active Memory Sharing intelligently flows memory from one partition to another for increased utilisation and flexibility of memory usage
  - Memory virtualisation enhancement for Power Systems
    - Next innovation in PowerVM virtualisation: extends resource optimisation to include memory
    - A pool of physical memory is dynamically allocated amongst logical partitions as needed to optimise overall memory usage in the pool
    - Can improve overall memory utilisation similar to the way micro-partitioning improves CPU utilisation
  - Blends Power Systems hardware, firmware and software enhancements to optimise resources
    - Supports over-commitment of logical memory with overflow going to VIOS managed paging devices
    - Two paging VIOS partitions can be used for redundancy
    - Compatible with Live Partition Mobility
  - More efficient utilisation of memory through collaboration with Operating System
    - Enables fine-grained sharing of physical memory and automated expansion and contraction of a partition's physical memory footprint
    - Supports OS collaborative memory management to reduce hypervisor paging

© 2020 Belisama

Active Memory Sharing (AMS) (cont).

Belisama

- Shared Memory Pool
  - Specify desired and maximum pool size
  - Assign paging devices and paging VIOS
    - Single or Redundant Paging VIOS(s)
  - Dynamically change pool size



(Total defined memory = 450 GB)

Shared Memory Pool paging devices

- Shared Memory Partition
  - Partition Attributes
  - Min, Max, Assigned Memory refer to logical memory.
  - I/O Entitled Memory: maximum amount of physical memory available for I/O mapping.
  - Memory Weight: partition's priority to get physical pages
  - Paging VIOSs: single or redundant; primary and secondary paging VIOS (optional). With IVM only 1 VIO

© 2020 Belisama

AMS and memory de-duplication

- How does it work?
  - The hypervisor in the background uses CPU cycles to compute a quick fingerprint of all pages in the shared memory pool (uses fraction of CPU that would be required for full checksum) – kept in cache
  - Any matches are then checked to confirm if identical, if so, one goes on the free list, and the other has two pointers to it.
  - If updated by one LPAR, then memory is temporarily marked read-only while a copy is made before modification takes place (Copy on Write)
  - Once turned on, in the HMC, you can then modify the default active memory deduplication cache that is created.

© 2020 Belisama

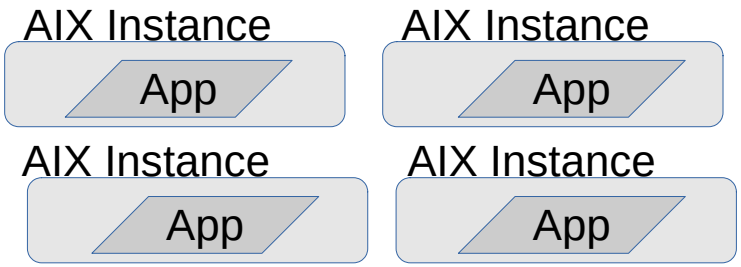
AMS and AME

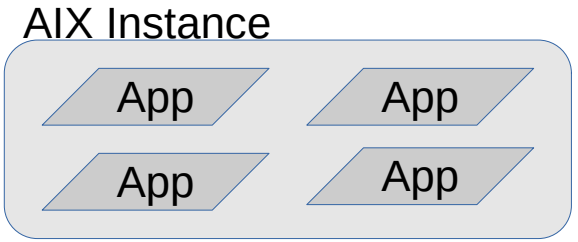
Belisama

- AMS and AME are complementary memory virtualisation technologies
  - AMS allows for great use of a systems memory by improving the utilisation of memory between multiple LPARs.
  - AME allows one LPAR to pack more data into it's memory
- AMS and AME can be used together
  - Some LPARs can have AME enabled so that they better use their memory, and if some of these LPARs have workloads that peak at different times, then AMS can be enabled.

Workload Partitions (WPAR)

- Consolidation of isolated workloads in a single AIX instance (AIX 6 & 7)
- Application inside a WPAR
  - Isolated users, processes, IPC, networks, IP address
  - Isolated filesystems: root, /tmp, /var & /home
  - Optional separate /usr & /opt filesystems
  - Full resource control – CPU, memory, paging space
  - Common AIX kernel services
  - Mobility between LPARs (machines)
  - 2 types
    - System – comprehensive
    - Application – 1 application, quick and simple, dies when app stops





© 2020 Belisama

WPAR (cont)

- Why?
  - Reduces administration overhead
  - Application encapsulation, monitoring and control
  - Rapid deployment of a new application
  - Separate system security and administration at the application level
  - Simple to move an application to another machine.
  - For a development environment?

Power7 / AIX 7
AIX 5.2 WPAR

Power7 / AIX 7
AIX 5.3 WPAR

© 2020 Belisama

Summary

- Simplify management
  - Use tested OS (AIX/Linux) images
  - Use a small number of tested OS images in the environment – include S/W and F/W. Keep consistent – easier to manage
  - Monitor IBM/Linux alerts for updates / warnings, use IBM tools:
    - SUMA, FLRT, NIM, alt_disk_install, multi-bos
  - Keep S/W and F/W up to date
- Don't TIP (Test In Production):
  - Build a realistic test environment for everything
  - Test ALL changes
  - Plan regular outages for maintenance
- Backup and recovery
  - Create and test your entire backup strategy – and DR strategy
  - Have an up to date copy of your mmsdrfs file

© 2020 Belisama

References

Belisama

- White papers:
  - IBM AIX Operating System Service Strategy Details and Best Practices
    - https://www14.software.ibm.com/webapp/set2/sas/f/best/aix_service_strategy.pdf
  - IBM AIX - An executive guide to the strategy and roadmap for the AIX Operating System for IBM Power Systems 2019
    - https://www.ibm.com/it-infrastructure/power/os/aix
  - Software support handbook
    - http://techsupport.services.ibm.com/guides/beforecontacting.html
- Web links, good start:
  - http://www-03.ibm.com/servers/eserver/support/unixservers
- VIO
  - http://www-941.ibm.com/collaboration/wiki/display/virtualization/Updating+VIO+Server
  - http://www.ibm.com/developerworks/aix/library/au-aix-vioserver-v2/index.html
  - InfoCenter
    - Link: http://publib.boulder.ibm.com/infocenter/systems/scope/hw/index.jsp
- IBM Developer Wiki
- IBM Power Performance Redbook
  - http://www.redbooks.ibm.com/abstracts/sg248080.html

© 2020 Belisama

References

- Performance Optimization and Tuning Techniques for IBM Power Systems Processors Including IBM POWER8 (Redbook)
  - https://www.ibm.com/account/reg/us-en/signup?formid=urx-18226
- AIX
  - IBM AIX - An executive guide to the strategy and roadmap for the AIX Operating System for IBM Power Systems 2019
- POWER9 Performance Best Practices
  - https://www14.software.ibm.com/webapp/set2/sas/f/best/power9_performance_best_practices.pdf
- AIX on Power - Performance FAQ
  - https://www14.software.ibm.com/webapp/set2/sas/f/best/aix_perf_FAQ.pdf
- Security
  - PowerSC details (a bit long in the tooth now): https://www.ibm.com/developerworks/aix/library/au-powersc/index.html
- Network tuning on AIX
  - http://www.ibmsystemsmag.com/aix/administrator/networks/network_tuning/

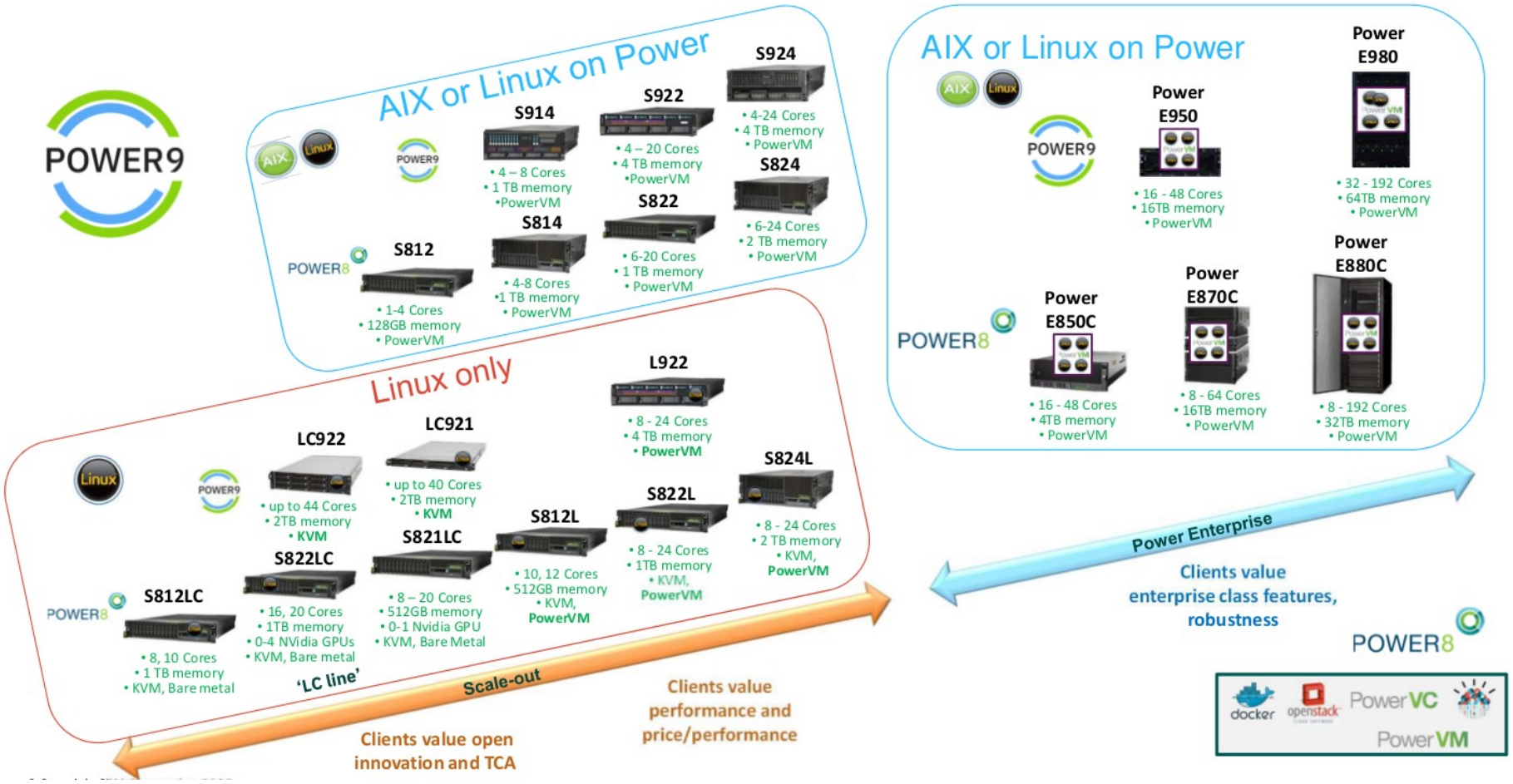
© 2020 Belisama

- IBM Techdocs - the Technical Sales Library
  - http://www-03.ibm.com/support/techdocs/atsmastr.nsf/Web/TechDocs
- Oracle 12.1.x and 11.2.0.4 Database Performance Considerations with AIX on POWER8
  - http://www-03.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP102608
- IBM Power System, AIX and Oracle Database Performance Considerations (for Oracle versions up to 11.2.0.3)
  - https://www-03.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP102171
- Oracle Database 11g and 12c on IBM Power Systems S924, S922 and S914 with POWER9 processors
  - http://www-03.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP102750
- Oracle on IBM Power Technology: Adoption Roadmap
  - http://www-03.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/PRS4711
- IBM White papers
  - https://www-03.ibm.com/support/techdocs/atsmastr.nsf/Web/WP-ByDate
- Getting started with VIO Shared storage pools
  - https://www.ibm.com/support/knowledgecenter/en/9080-M9S/p9hb1/p9hb1_clustervioscli.html

© 2020 Belisama

Oracle is supported on

Belisama



Session: a111038

AIX / Linux on Power “Good” practice

¿ Questions ?

Thanks!

Your feedback about this session is very important to us.

Please remember to submit a survey

For further information….
Contact:

Antony (Red) Steel antony.steel@belisama.com.sg
+65 9789 6663

Power Systems

© 2020 Belisama