Paper: **#20240**

The Open Platform for Choice: Linux on Power Virtualization

Michael Perzl (mperzl@de.ibm.com)
Technical Sales Power Systems
IBM Deutschland GmbH

Trademarks

The following are trademarks of the International Business Machines Corporation in the United States, other countries, or both.

Not all common law marks used by IBM are listed on this page. Failure of a mark to appear does not mean that IBM does not use the mark nor does it mean that the product is not actively marketed or is not significant within its relevant market.

Those trademarks followed by ® are registered trademarks of IBM in the United States; all others are trademarks or common law marks of IBM in the United States.

For a complete list of IBM Trademarks, see www.ibm.com/legal/copytrade.shtml:

*, AS/400®, e business(logo)®, DBE, ESCO, eServer, FICON, IBM®, IBM (logo)®, iSeries®, MVS, OS/390®, pSeries®, RS/6000®, S/30, VM/ESA®, VSE/ESA, WebSphere®, xSeries®, z/OS®, zSeries®, z/VM®, System i, System i5, System p, System p5, System x, System z, System z9®, BladeCenter®

The following are trademarks or registered trademarks of other companies.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.
Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license therefrom.
Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.
Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.
Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.
UNIX is a registered trademark of The Open Group in the United States and other countries.
Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.
ITIL is a registered trademark, and a registered community trademark of the Office of Government Commerce, and is registered in the U.S. Patent and Trademark Office.
IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency, which is now part of the Office of Government Commerce.

* All other products may be trademarks or registered trademarks of their respective companies.

Notes:
Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios stated here.
IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.
All customer examples cited or described in this presentation are presented as illustrations of the manner in which some customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual customer configurations and conditions.
This publication was produced in the United States. IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area.
All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.
Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the performance, compatibility, or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.
Prices subject to change without notice. Contact your IBM representative or Business Partner for the most current pricing in your geography.

paper #20240: The Open Platform for Choice: Linux on Power Virtualization

Agenda

- Introduction / Recent announcement
- Power Virtualization Options – KVM
- Power Virtualization Options – PowerVM
- Linux on Power – Device and Virtualization Support
- PowerVM advantages over competitive virtualization technologies
- Summary

POWER = **P**erformance **O**ptimized **W**ith **E**nhanced **R**ISC

paper #20240: The Open Platform for Choice: Linux on Power Virtualization

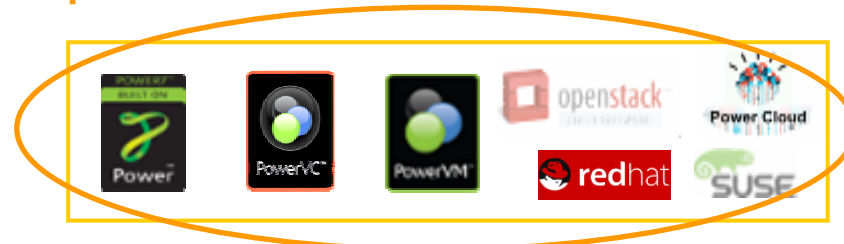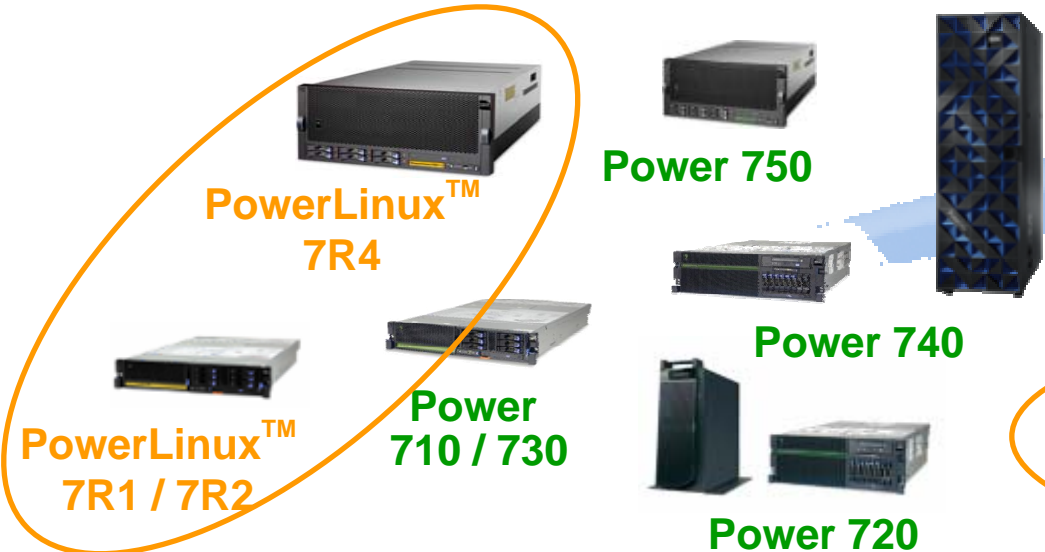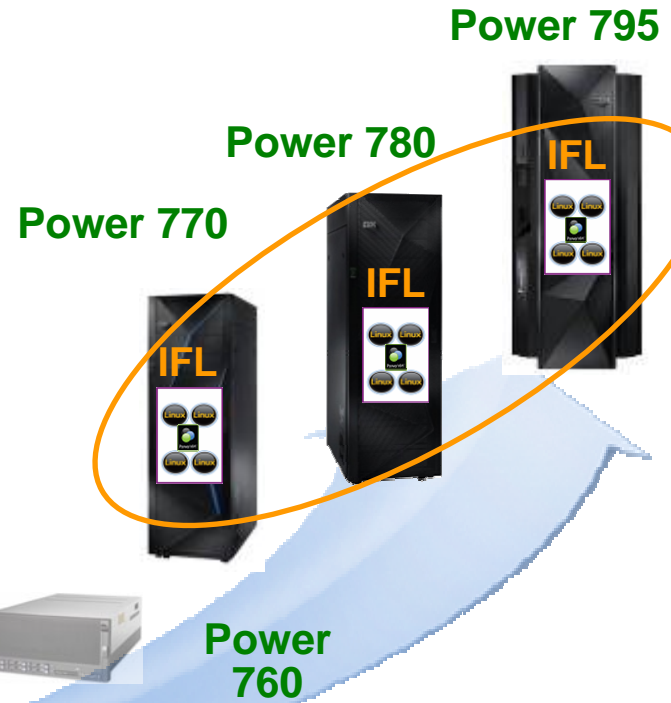# Linux supports all IBM Power Systems servers

Industry standard Linux
- Red Hat and SUSE versions consistent with x86_64
- Support available simultaneously with other platforms
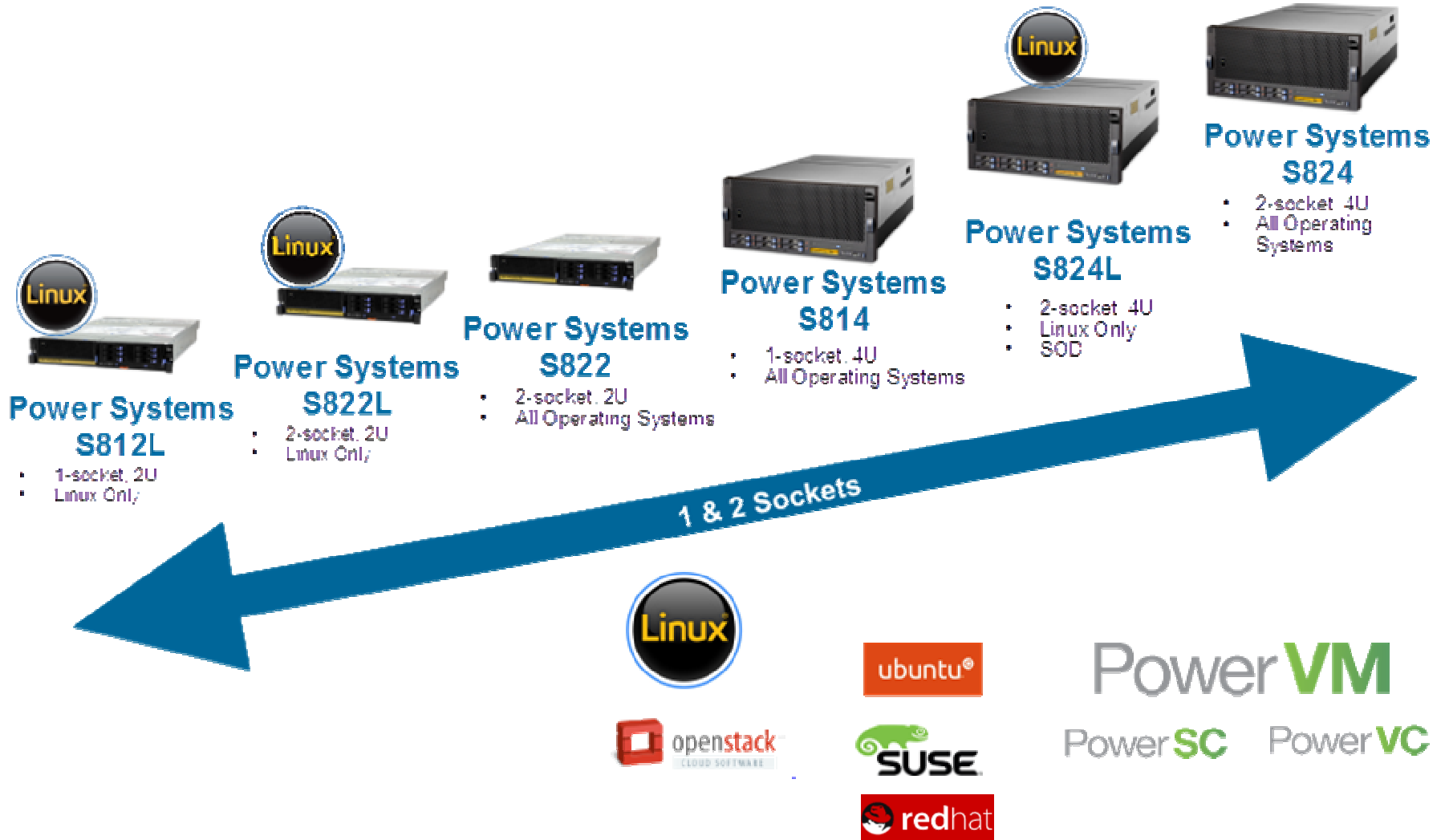
Optimized by IBM to exploit POWER7+ and PowerVM
- Virtualization, Performance, POWER7+ RAS

Broadest choice of Linux servers
- Linux supports Power 710 to 795 and new Power IFL
- Linux only one, two and four socket servers:
  - PowerLinux 7R1, 7R2, 7R4
  - Flex System p24L

Power 795

Power 780

Power 770

IFL

IFL

IFL

Power 760

PowerLinux™ 7R4

Power 750

IBM Flex System
- p460, p260, p24L

Power 740

Power 710 / 730

PowerLinux™ 7R1 / 7R2

Power 720

paper #20240: The Open Platform for Choice: Linux on Power Virtualization

New IBM Power Systems scale-out portfolio

Power Systems S812L
- 1-socket, 2U
- Linux Only

Power Systems S822L
- 2-socket, 2U
- Linux Only

Power Systems S822
- 2-socket, 2U
- All Operating Systems

Power Systems S814
- 1-socket, 4U
- All Operating Systems

Power Systems S824L
- 2-socket, 4U
- Linux Only
- SOD

Power Systems S824
- 2-socket, 4U
- All Operating Systems

1 & 2 Sockets

paper #20240: The Open Platform for Choice: Linux on Power Virtualization

Red Hat support for POWER

- **Red Hat Enterprise Linux 7**
  - Public beta available for existing RHEL customers
  - Expected full support for POWER8 (native mode) and POWER 7/7+ at operating system GA

- **Red Hat Enterprise Linux 6**
  - POWER8 supported with U5 (P7-compatibility mode)
  - Full support of POWER6 and POWER7 (native mode)
  - Last update: U5 GA December 2013

- **Fedora**
  - Fedora 16 was first release to re-launch POWER
  - Fedora 20 has POWER8 support
  - Fedora remains primary community for major innovation/collaboration

- **Supported add-ons**
  - JBoss
  - High Performance Network Add-on

> Built from the same source as x86
> Delivered on the same schedule as x86
> Supported at the same time as x86

SUSE support for POWER

- **SUSE Linux Enterprise Server 11**
  - POWER8 supported with SP3 (P7-compatibility mode)
  - POWER7+ encryption, RNG accelerators supported with SP3
  - Full support of POWER7 (native mode)
  - Earliest supported release: SLES 11 base
  - Last update: SP3 GA July 2013

- **SUSE Linux Enterprise Server 10***
  - POWER7 supported with SP3 (P6-compatibility mode)
  - Full support of POWER6 (native mode)
  - Last update: SP4 GA April 2011

- **openSUSE**
  - openSUSE 12.2 re-launched for IBM POWER
  - openSUSE 13.2 includes POWER8 support (native mode)

- **Supported add-ons**
  - SUSE Linux Enterprise High Availability Extension (included in base Power license)

> Built from the same source as x86
> Delivered on the same schedule as x86
> Supported at the same time as x86

* Not supported on POWER7+ and POWER8 systems

paper #20240: The Open Platform for Choice: Linux on Power Virtualization

Canonical support for POWER

- Ubuntu 14.04
  - POWER8 enabled (native mode)
  - No official support for POWER7+ and older systems
  - No support for 32-bit applications. 64-bit only.
  - Supported in KVM only at this time

- Supported add-ons
  - JuJu Charms
  - MaaS (Metal as a Service)
  - Landscape

- Debian
  - Community enablement in process

> Built from the same source as x86
> Delivered on the same schedule as x86
> Supported at the same time as x86

OpenPOWER™ **Foundation current Members**

Platinum	Gold	Silver	Academic
IBM	创和通讯	FUSION-io	JÜLICH FORSCHUNGSZENTRUM
Google	HITACHI Inspire the Next	XILINX ALL PROGRAMMABLE.	OSU Oregon State University
ubuntu Supported by Canonical	inspur 浪潮	SK hynix	UNIS
ALTERA.			
Mellanox TECHNOLOGIES	TEAMSUN 华胜天成	Servergy	**Individual**
Micron	VeriSilicon	EMULEX	JF YUAN
TYAN	ZTE中兴		
SAMSUNG			
NVIDIA			
PowerCore			

Technology — Fab
Chip — SoC, Dev, IP
I/O — Memory, Networking, Storage
Systems — ODM, OEM
Firmware — Open Source
Software — Linux, ISV, Open Source
Web 2.0 — Internet, Data Center, Cloud, Tech Computing

paper #20240: The Open Platform for Choice: Linux on Power Virtualization

Why should Power Processors become more relevant? OpenPower Foundation – Technical Direction

- OpenPower Foundation is an open development alliance based on IBM's POWER microprocessor architecture.

- In order to deliver more choice, control and flexibility to developers of next-generation, hyper-scale and cloud data centers the Consortium intends to build
  - advanced server,
  - networking,
  - storage,
  - and GPU-acceleration technology.

- To provide unprecedented customization for creating new styles of server hardware to address a variety of computing workloads the consortium will offer
  - OpenPOWER hardware architecture (i.a. ISA – Instruction Set Architecture),
  - open-source firmware (OPAL)
  - KVM virtualization with Linux software stack (e.g. new 64-bit little endian ABI).

OpenPower Target Software Stack

LE SDK	LE Apps	LE Apps	LE Apps
Linux LE	Linux LE	Linux LE	Linux LE

QEMU (Built LE – Supports BE & LE Guests)

KVM v2.0 (Host OS) (Built LE)

Device Drivers (Built LE)

Power Firmware Interface

Stand-alone Diagnostics

OPAL Firmware (Built BE)

SkiBoot

SkiRun (HAL Function) | OCC (Energy Mgmt)

Hostboot | SBE

BMC | Power 8 HW

paper #20240: The Open Platform for Choice: Linux on Power Virtualization

Why should Power Processors become more relevant? OpenPower Foundation – Technical Direction

- OpenPower Foundation is an open development alliance based on IBM's POWER microprocessor architecture.

- In order to deliver more choice, control and flexibility to developers of next-generation, hyper-scale and cloud data centers the Consortium intends to build
  - advanced server,
  - networking,
  - storage,
  - and GPU-acceleration technology.

- To provide unprecedented customization for creating new styles of server hardware to address a variety of computing workloads the consortium will offer
  - OpenPOWER hardware architecture (i.a. ISA – Instruction Set Architecture),
  - open-source firmware (OPAL)
  - KVM virtualization with Linux software stack (e.g. new 64-bit little endian ABI).

OpenPower Hardware Introduction

paper #20240: The Open Platform for Choice: Linux on Power Virtualization
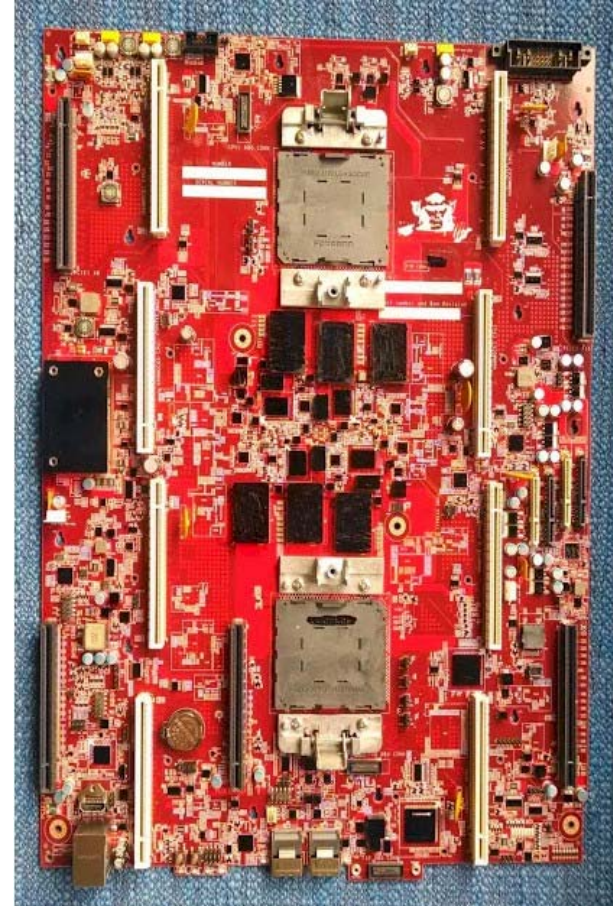
Power Virtualization Options

PowerKVM

Q2 2014
Initial Offering

PowerKVM provides an open source choice
for Power Virtualization for Linux workloads.
Best for clients that aren't familiar with Power
and Linux centric admins.

PowerVM

2004
Initial Offering

PowerVM is Power Virtualization that will continue to be enhanced
to support AIX, IBM i Workloads as well as Linux Workloads

Power Virtualization Options – KVM
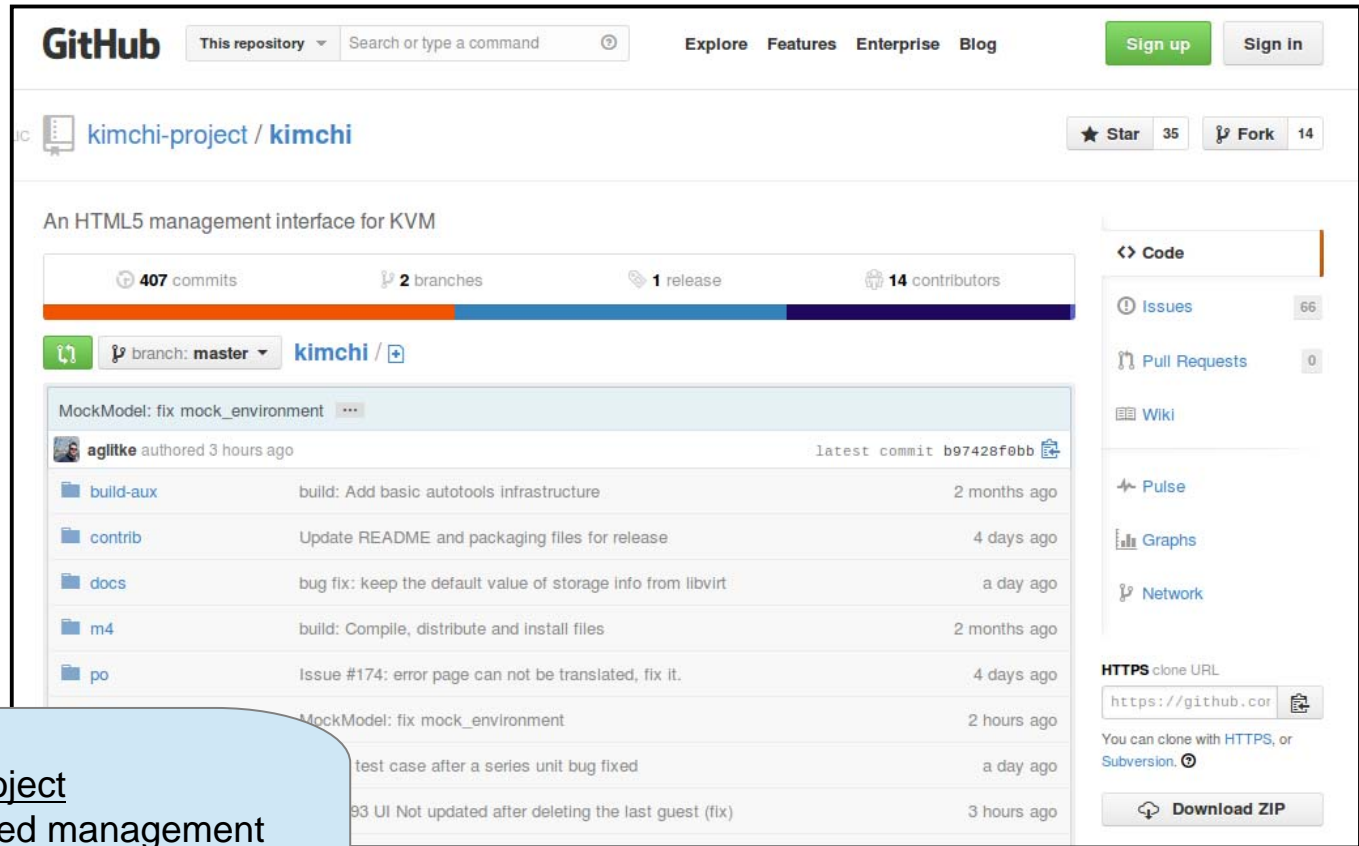
KVM on Power – PowerKVM

Standard interfaces

- Smart Cloud Entry
- Openstack controller
- PowerVC™
- Kimchi
- Openstack End-node components
- QEMU
- Kimchi
- Libvirt API & virsh CLI
- Browser Or Client
- SUSE Manager
- Console Shell — CLI
- Xcat Chef Puppet Custom scripts

Host OS (ppc64)

Linux Userspace
- VM1 RHEL
- VM2 SLES
- VM3 Fedora

Linux Kernel
- KVM

OPAL FW

Power8 Platform — FSP — CLI / IPMI
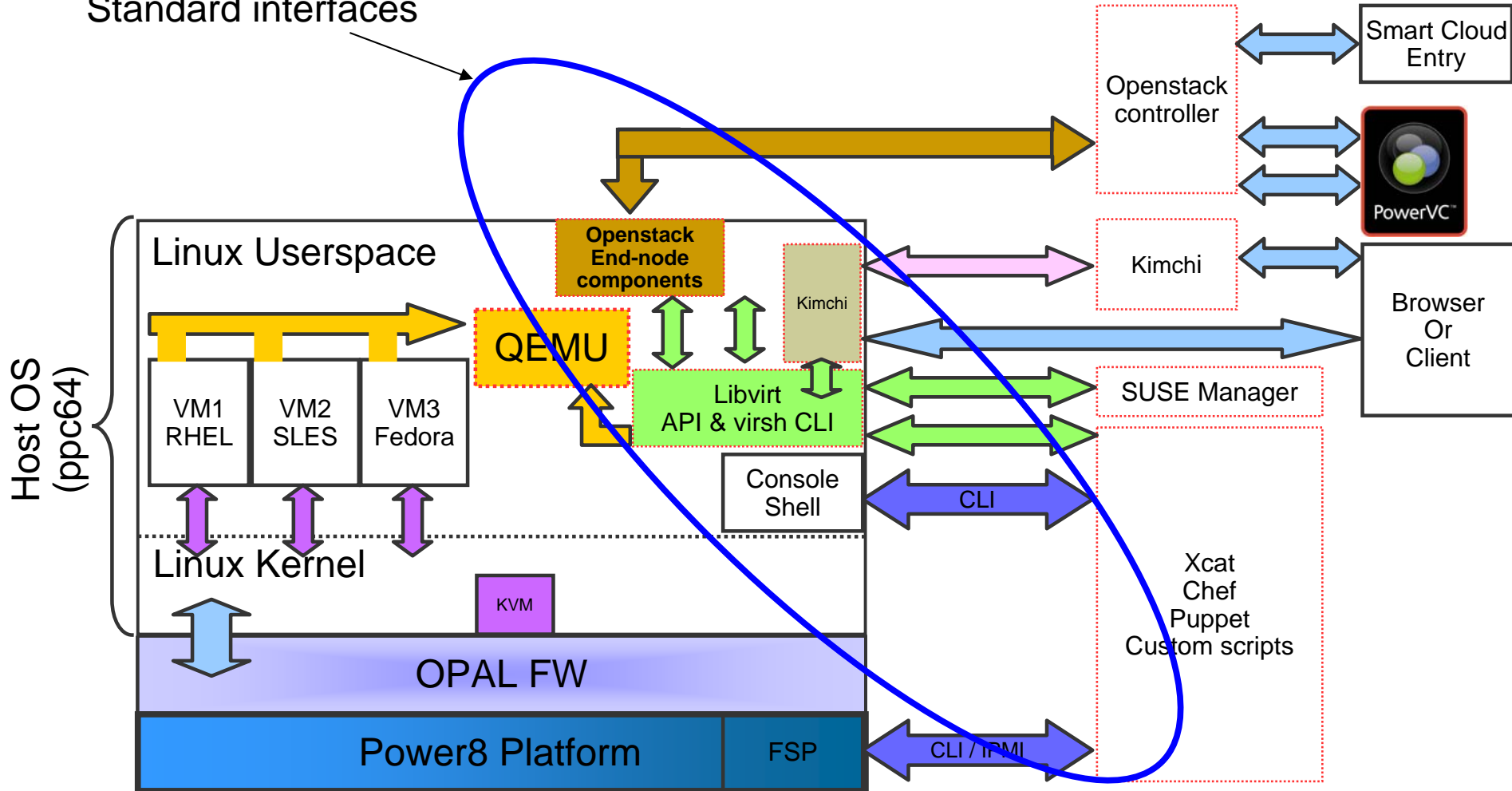
Project Kimchi – an emerging open source KVM management tool

Kimchi Project
Kimchi is an HTML5 based management tool for KVM. It is designed to make it as easy as possible to get started with KVM and create your first guest.

More information at https://github.com/kimchi-project/kimchi

KVM on POWER should behave identically to KVM on x86:

- KVM project wiki
  - http://www.linux-kvm.org/page/Main_Page
- IBM KVM overview
  - http://www.ibm.com/developerworks/cloud/library/cl-hypervisorcompare-kvm/
- Developments in KVM on Power
  - http://www.linux-kvm.org/wiki/images/7/70/Kvm-forum-2013-Mackerras.pdf
- KVM Forum 2013: Developments in KVM on Power by Paul Mackerras
  - http://www.youtube.com/watch?v=cLQI20LI6EQ

Power Virtualization Options – PowerVM

PowerVM Virtualization

Simplification through virtualization

- Micro-Partitioning (1/20th processor minimum)
- Multiple Shared processor pools
- Dynamic LPAR
- Virtual I/O
  - Storage
  - LAN

Reduced resources

- Fewer processors & I/O adapters
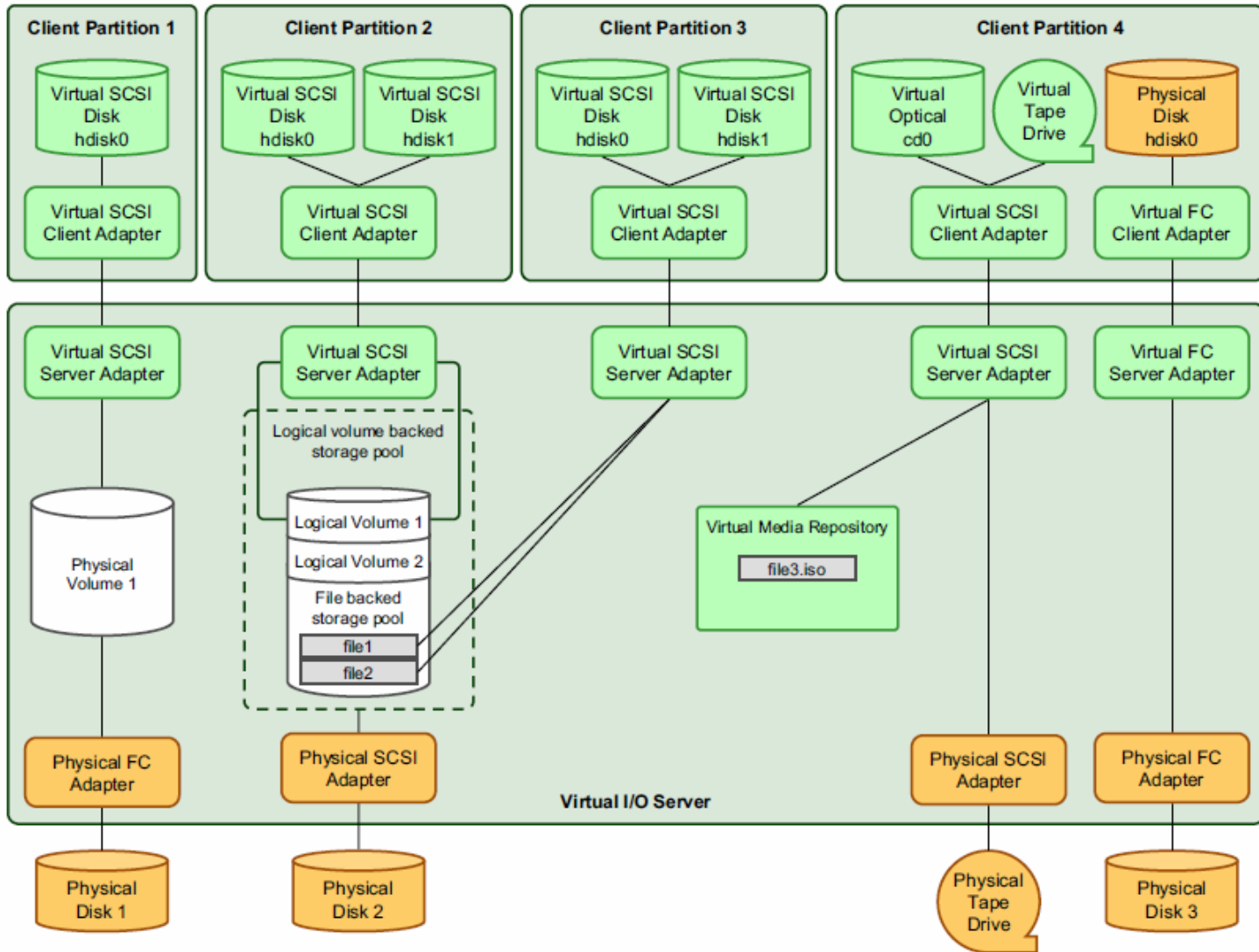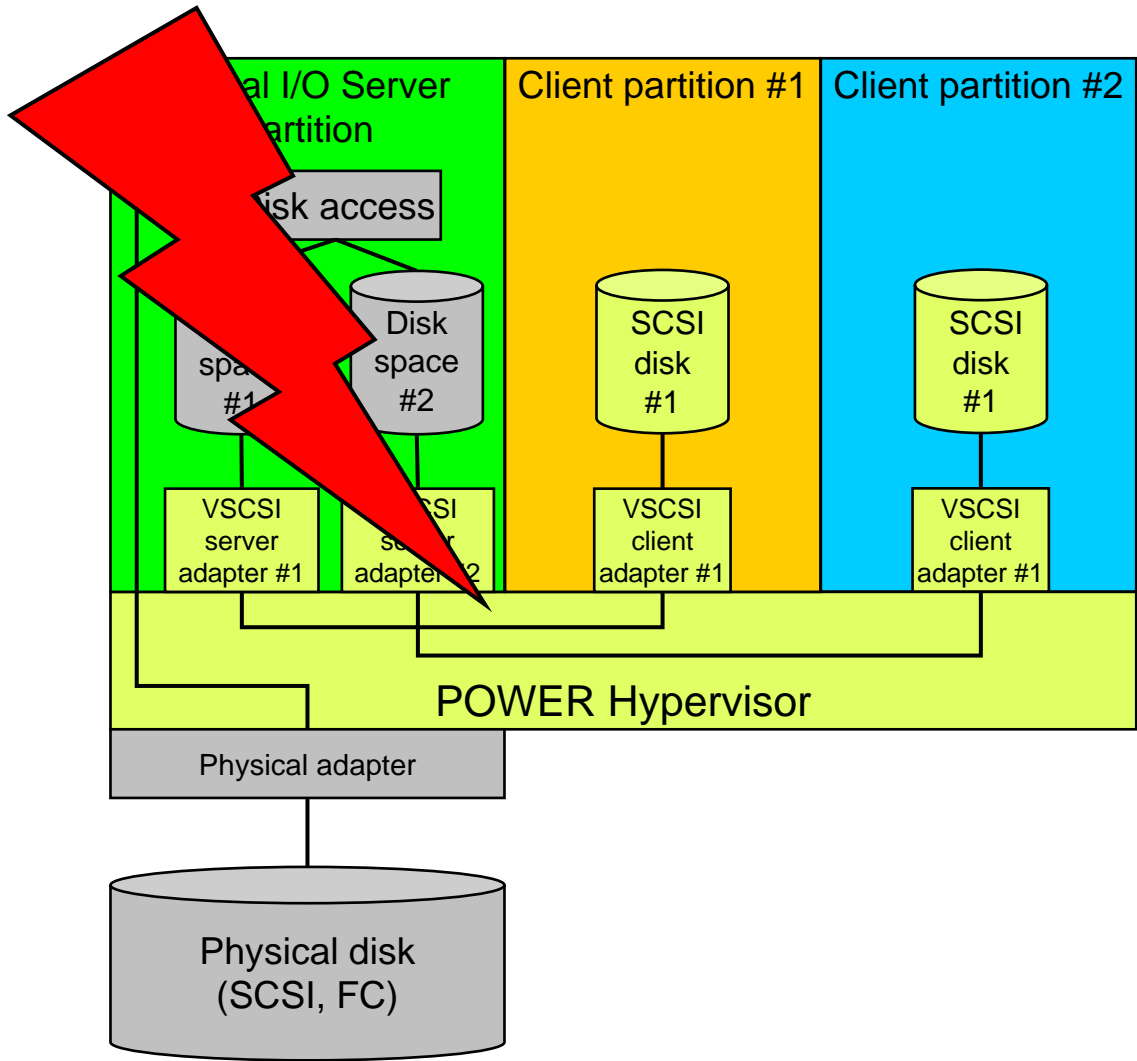- Increased overall system utilization and performance

Dynamically Resizable

Virtual I/O Server Partition

CPUs

Micro-partitioning

Storage Sharing

Ethernet Sharing

Linux Linux Linux Linux Linux Linux Linux

Virtual I/O paths

POWER Hypervisor
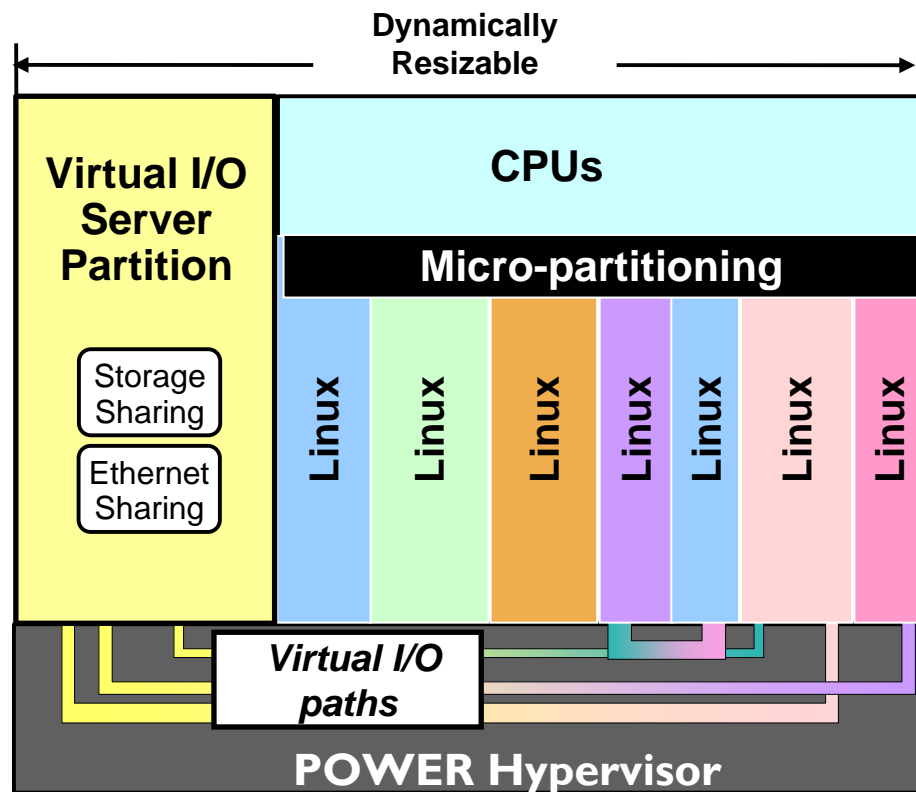
Hardware Management Console

Virtual I/O Server Storage Virtualization (1/2)

- Virtual I/O Server allows virtualization of physical storage resources.
- Virtualized storage devices are accessed by client partitions by one of these methods:
  - Virtual SCSI
    - Provides standard SCSI compliant access by client partitions to disk devices, optical devices and tape devices.
  - Virtual Fibre Channel
    - Provides access by Virtual Fibre Channel (VFC) devices to Fibre Channel attached disk and tape libraries.
- The following logical storage devices can be used to back virtualized storage devices:
  - Logical volumes
  - Logical volume storage pools
  - File storage pools
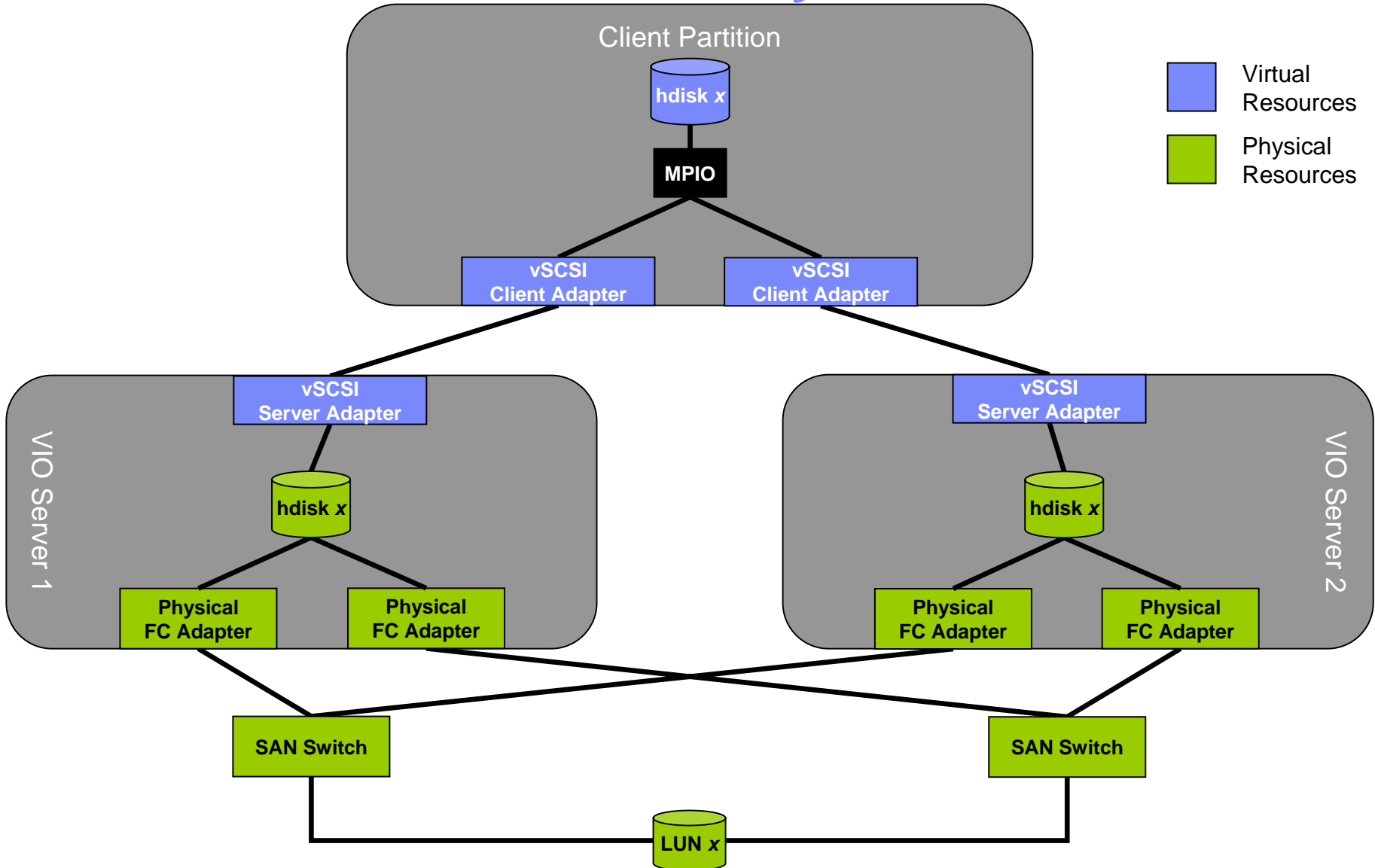 - Shared storage pools
  - Virtual media repository

paper #20240: The Open Platform for Choice: Linux on Power Virtualization

Virtual I/O Server Storage Virtualization (2/2)

paper #20240: The Open Platform for Choice: Linux on Power Virtualization © 2014 IBM Corporation

Virtual I/O Server: Virtual SCSI – Simple Setup

Virtual I/O Server partition

Disk access

Disk space #1

Disk space #2

VSCSI server adapter #1

VSCSI server adapter #2

Client partition #1

SCSI disk #1

VSCSI client adapter #1

Client partition #2

SCSI disk #1

VSCSI client adapter #1

POWER Hypervisor

Physical adapter

Physical disk (SCSI, FC)

paper #20240: The Open Platform for Choice: Linux on Power Virtualization

Dual VIOS Server Redundancy: VSCSI

Client Partition

hdisk *x*

MPIO

vSCSI Client Adapter
vSCSI Client Adapter

Virtual Resources

Physical Resources

VIO Server 1
vSCSI Server Adapter
hdisk *x*
Physical FC Adapter
Physical FC Adapter

VIO Server 2
vSCSI Server Adapter
hdisk *x*
Physical FC Adapter
Physical FC Adapter

SAN Switch
SAN Switch

LUN *x*

paper #20240: The Open Platform for Choice: Linux on Power Virtualization

Dual VIOS Server Redundancy: NPIV/VFC

Power Systems

IBM

Client Partition

hdisk *x*

MPIO

| vFC Client Adapter | vFC Client Adapter | vFC Client Adapter | vFC Client Adapter |

Virtual Resources

Physical Resources

VIO Server 1

| vFC Server Adapter | vFC Server Adapter |

| Physical FC Adapter | Physical FC Adapter |

VIO Server 2

| vFC Server Adapter | vFC Server Adapter |

| Physical FC Adapter | Physical FC Adapter |

SAN Switch

SAN Switch

LUN *x*

paper #20240: The Open Platform for Choice: Linux on Power Virtualization

Virtual Ethernet – Overview

- Memory based inter-partition LAN
  - Packets copied between LPARs
- Physical network adapters are not needed for inter-partition communication
- Virtual LAN adapters appear to the OS as physical adapters
  - HMC generates MAC addresses
  - Supports BOOTP, DHCP…
  - VLAN support

Hosting partition | **Linux** | **Linux** | **Linux**

Packet forwarder | Virtual Ethernet driver | Virtual Ethernet driver | Virtual Ethernet driver

Virtual Ethernet switch

Hypervisor

Network adapters

LAN, WAN, …
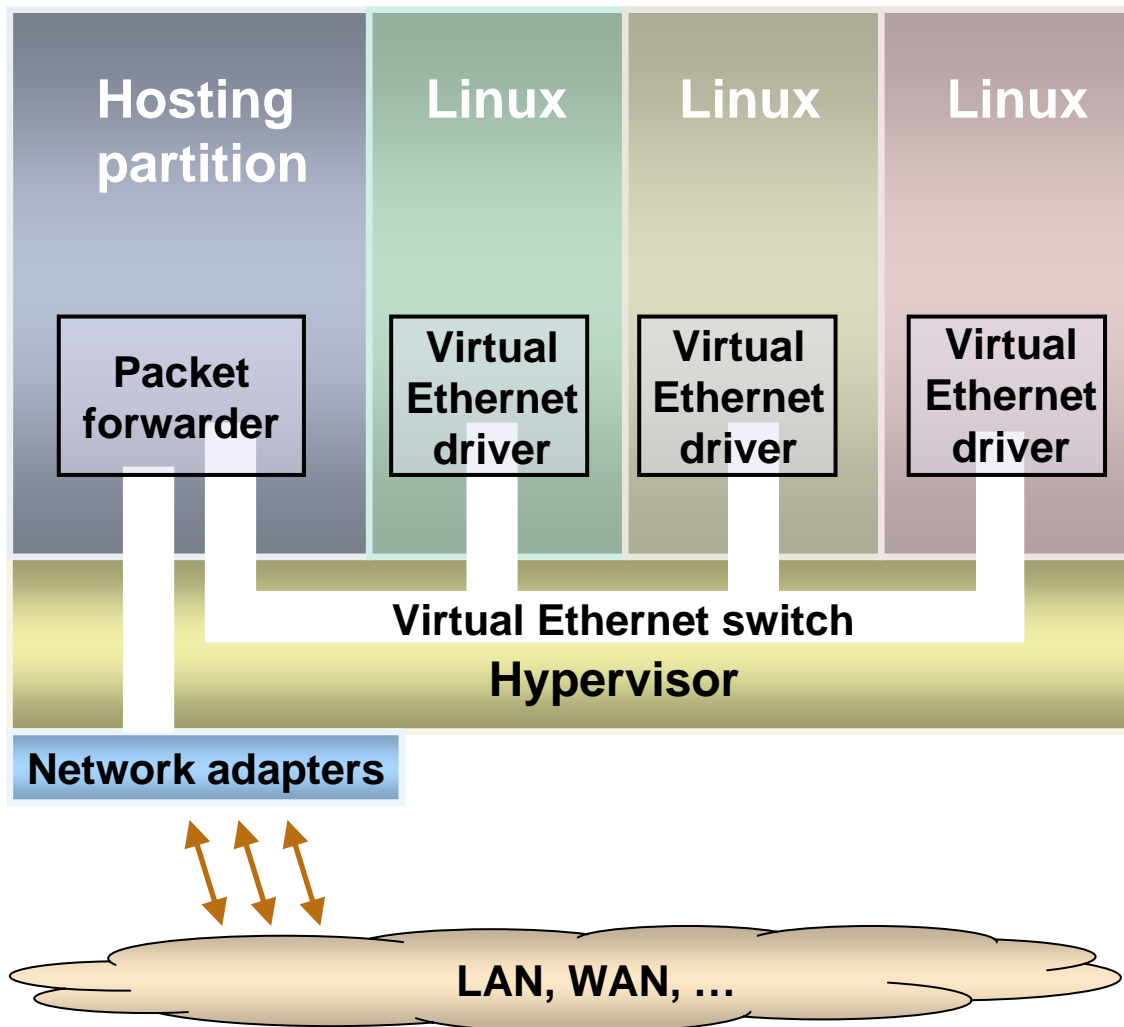
How does Shared Ethernet Adapter (SEA) failover work ?
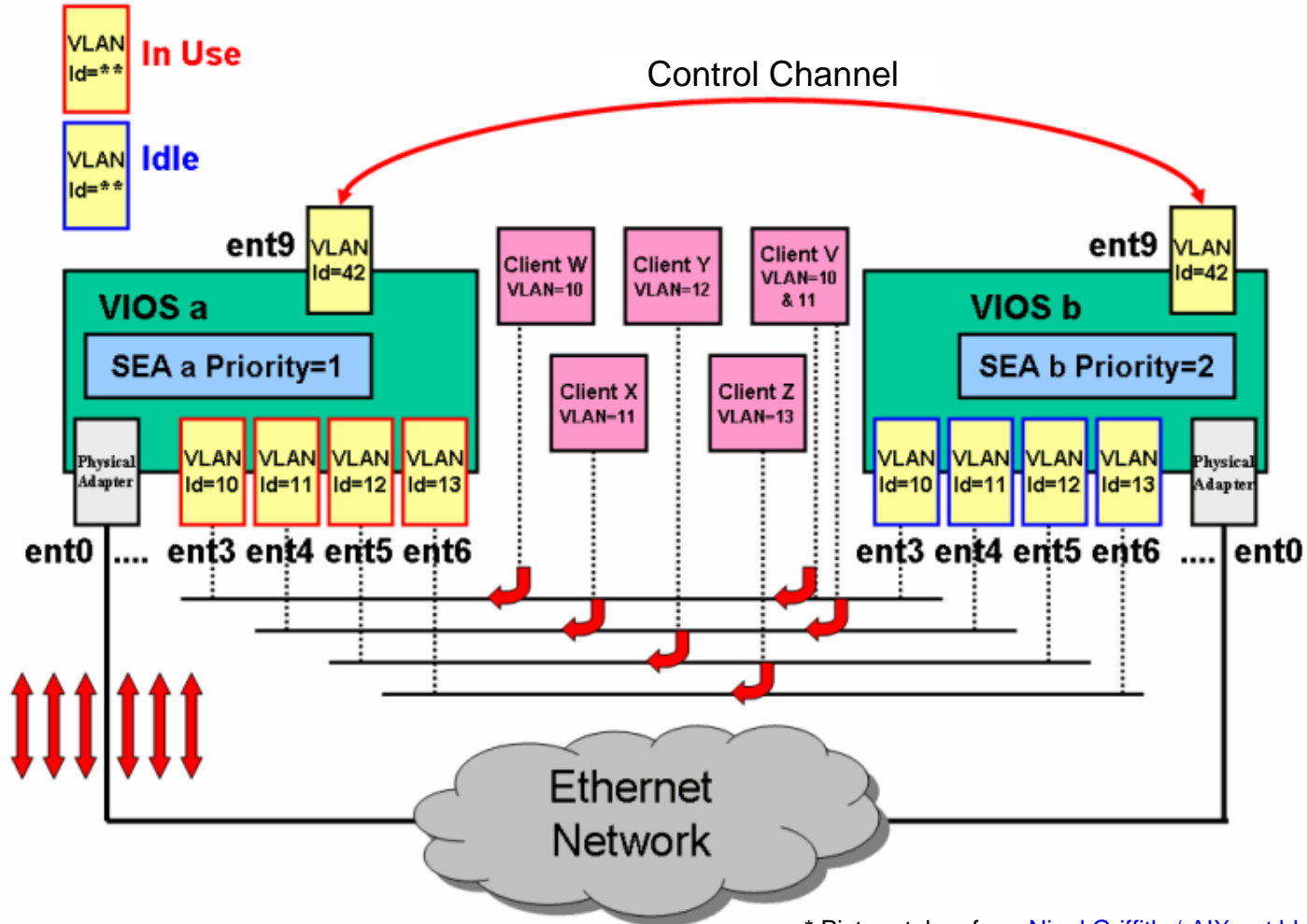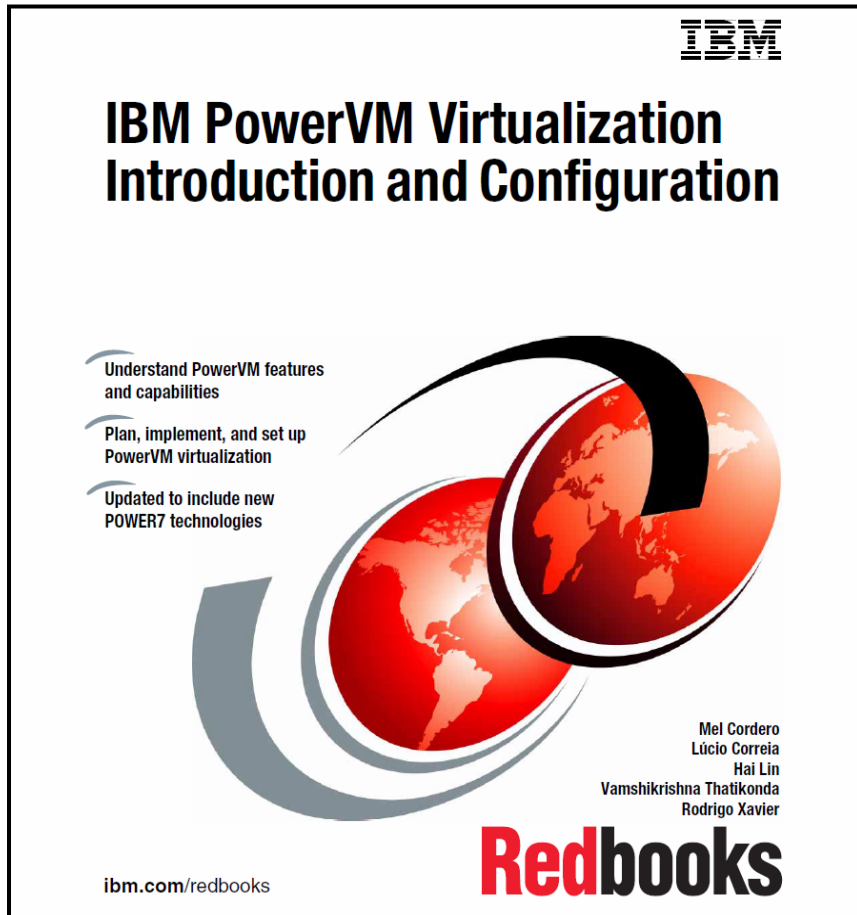
paper #20240: The Open Platform for Choice: Linux on Power Virtualization

Shared Ethernet Adapter failover

* Picture taken from Nigel Griffiths' AIXpert blog on this topic.

paper #20240: The Open Platform for Choice: Linux on Power Virtualization

Shared Ethernet Adapter failover with load sharing
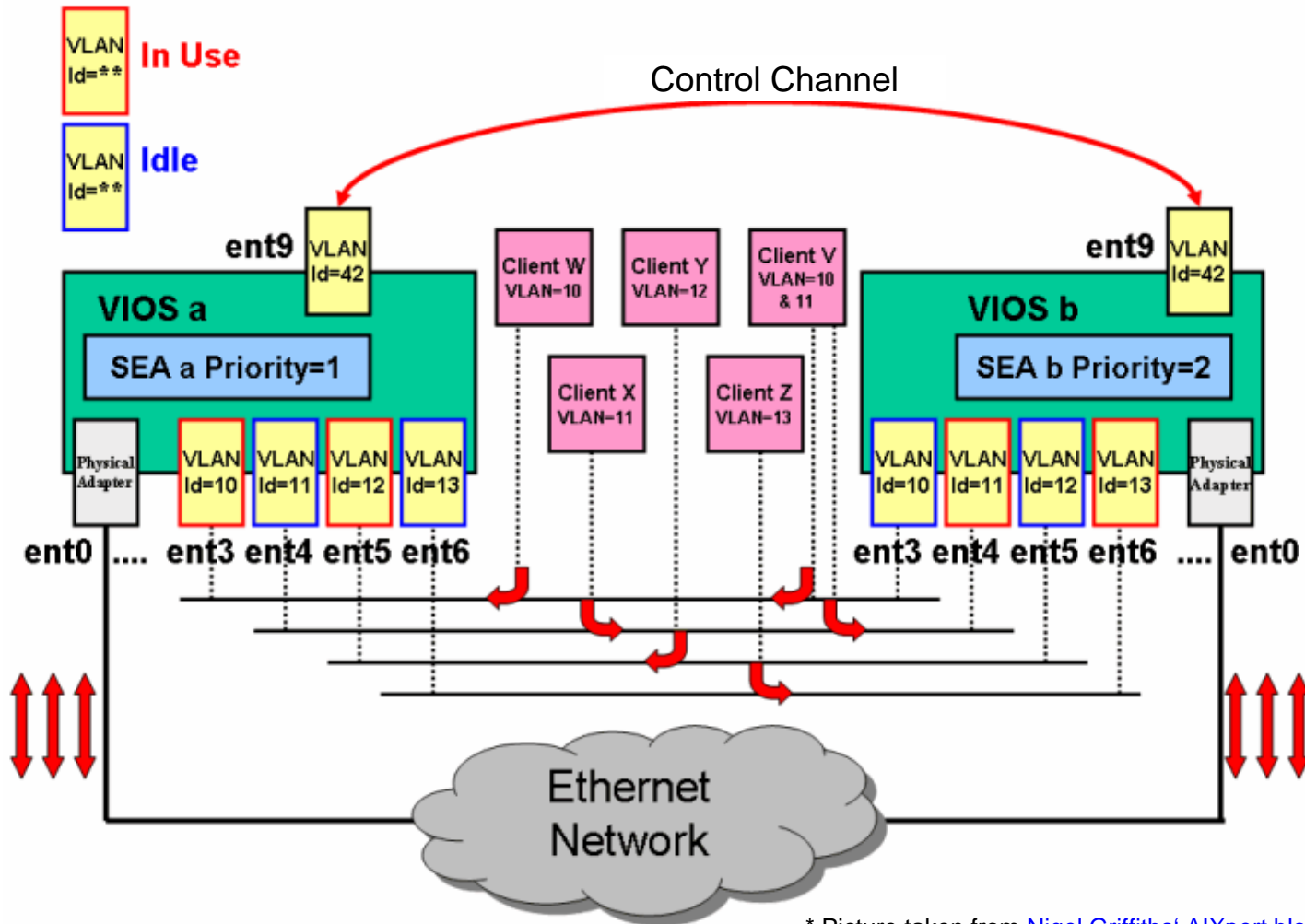
* Picture taken from Nigel Griffiths' AIXpert blog on this topic.

More documentation about PowerVM...

http://www.redbooks.ibm.com/abstracts/sg247940.html

http://www.redbooks.ibm.com/abstracts/sg247590.html

paper #20240: The Open Platform for Choice: Linux on Power Virtualization

© 2014 IBM Corporation

Linux on Power –
Device and Virtualization Support

Linux Kernel Virtualization Support

- Virtual device support implemented with Linux kernel modules
  - ibmveth - virtual ethernet device driver
  - ibmvscsic - virtual SCSI client device driver
  - ibmvfc - virtual Fibre Channel client device driver
  - ibmvstgt - virtual SCSI target device driver

```
# find /lib/modules -name "ibmv*ko" -print
/lib/modules/x.x.xx/kernel/drivers/net/ibmveth.ko
/lib/modules/x.x.xx/kernel/drivers/scsi/ibmvscsi/ibmvscsic.ko
/lib/modules/x.x.xx/kernel/drivers/scsi/ibmvscsi/ibmvfc.ko
/lib/modules/x.x.xx/kernel/drivers/scsi/ibmvscsi/ibmvstgt.ko
```

- No closed source device drivers for Linux on Power, all Linux on Power device drivers are open source.

- All contained in the standard "vanilla" Linux kernel (from http://kernel.org) for a long time!

paper #20240: The Open Platform for Choice: Linux on Power Virtualization

PowerVM advantages over competitive virtualization technologies

Power Systems RAS vs x86

RAS Feature	POWER7	x86
System RAS		
OS independent First Failure Data Capture	Yes	No
Memory Keys (including OS exploitation)	Yes	No
Processor RAS		
Processor Instruction Retry	Yes	No
Alternate Processor Recovery	Yes	No
Dynamic Processor Deallocation	Yes	No
Dynamic Processor Sparing	Yes	No
Memory RAS		
Chipkill™	Yes	Yes
Survives Double Memory Failures	Yes	No
Selective Memory Mirroring	Yes	No
Redundant Memory	Yes	Yes
I/O RAS		
Extended Error Handling	Yes	No
I/O Adapter Isolation (PI-Bus and TCEs)	Yes	No

PowerVM is the only platform that demonstrated over 98% availability across all virtualization deployments

- PowerVM versus competitive virtualization study with over 61,000 clients analyzed

- PowerVM virtualization contributes to both stability and reliability of an organization'simplementation

- **Virtualized x86 system availability decreases as an organization's size increases**

- Key factors
  - Reliability and availability that meet today's business requirements
  - PowerVM on Power demonstrates superior reliability and availability over all other virtualization platforms
  - Detailed claims and discussion
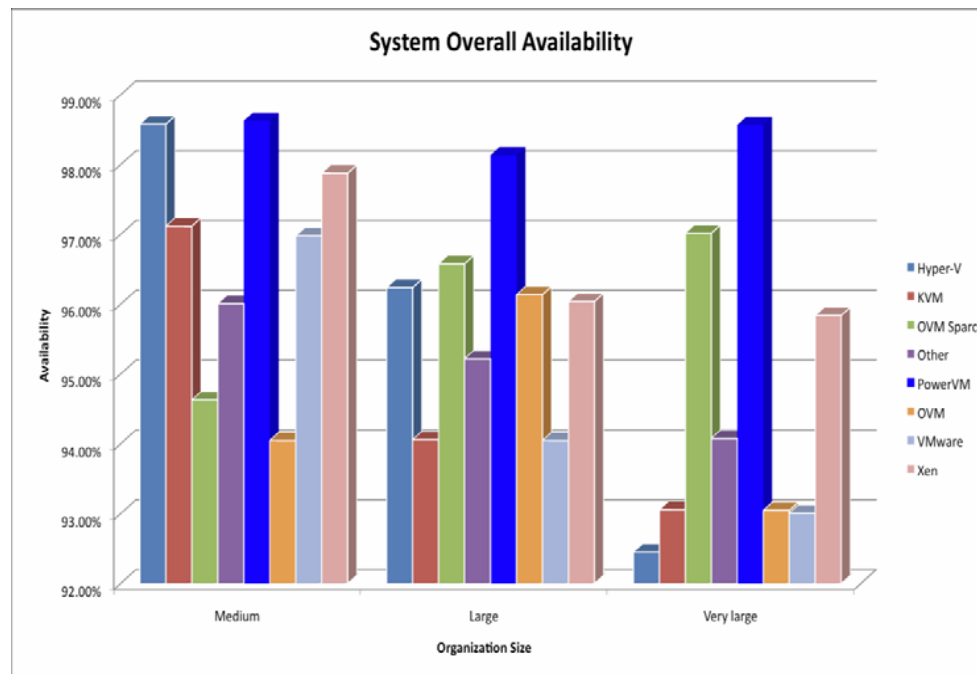
Source: Does Your Virtualization Platform Matter? Getting the Most Out of Your IT Platforms with Virtualization; Solitaire Interglobal Ltd (All rights reserved); April 2012.

PowerVM, with its optimized "dense virtualization capability", allows clients to leverage virtualization technology efficiently for a maximum number workload deployment on a single system in the most cost-effective manner.

- **x86 server virtualization solutions like VMware vSphere, Oracle VM, and Microsoft Hyper-V**
  - are engineered for less VM density per system
  - have a VM density per system of up to 6X less than PowerVM
- **PowerVM leads all server virtualization solutions in VM density**
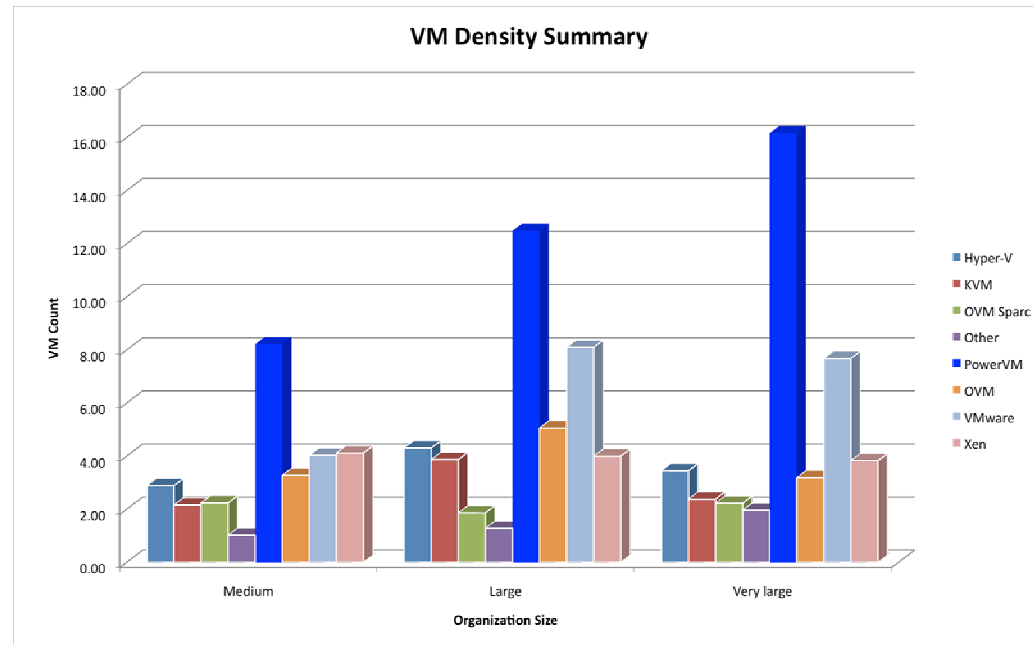
Source: Does Your Virtualization Platform Matter? Getting the Most Out of Your IT Platforms with Virtualization; Solitaire Interglobal Ltd (All rights reserved); April 2012.
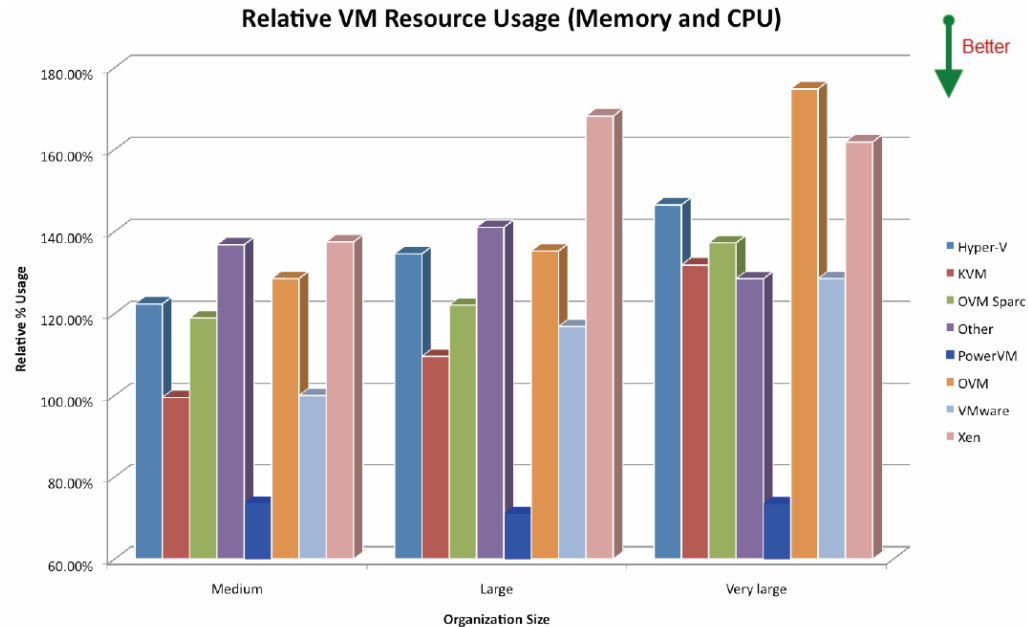
paper #20240: The Open Platform for Choice: Linux on Power Virtualization
© 2014 IBM Corporation

Optimized for efficiency
PowerVM resource usage is much lower than other competitive virtualization solutions

PowerVM versus competitive virtualization study
61,000 customers surveyed

- Compared to a baseline of a mid-sized VMware deployment

- PowerVM is up to 105% more efficient in VM resource usage over competitive virtualization offers

Relative VM Resource Usage (Memory and CPU)

Better

Legend: Hyper-V, KVM, OVM Sparc, Other, PowerVM, OVM, VMware, Xen

Source: Solitaire Interglobal Ltd (All rights reserved); April 2012.

Security of critical workload (SAP) deployments on Power is beyond reproach

0
reported security breaches with SAP and IBM DB2 or Oracle on Power

- SAP on Power versus competitive SAP deployments study with over 54,150 clients analyzed.

- The security for ERP systems, including SAP, can be very challenging – by nature, the mixture of application modules, user profiles, plug-in components and so on, provide many avenues for security breaches.
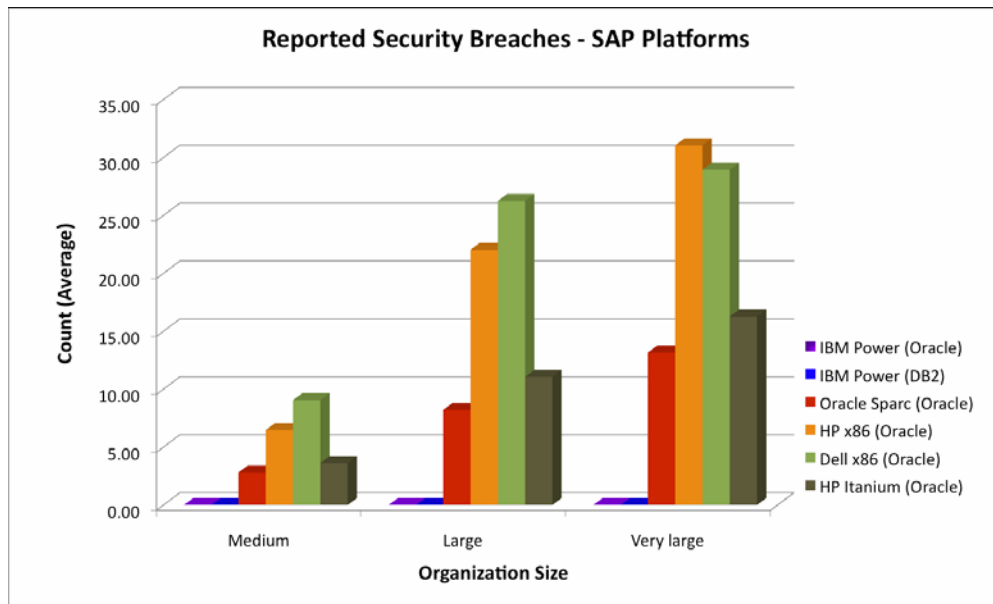
Source: Business Impacts on SAP Deployments; Solitaire Interglobal Ltd (All rights reserved); January 2013.

Server virtualization security is critical

0
reported security breaches on the PowerVM hypervisor

- The PowerVM hypervisor has never had a reported security vulnerability and provides the bullet-proof security that customers demand for mission-critical workloads.

- The VIOS, which is part of the overall virtualization has had **0** reported security vulnerabilities.

- **Dare to compare – search any security tracking DB and compare Power against x86.**

Search term or Hypervisor (unfiltered)	NIST NVD Results	Processor Architecture
VMware	640	x86
Xen	153	x86
VMware ESX	95	x86
KVM	58	x86
VMware vSphere	48	x86
Windows Server 2012	43	x86
Oracle VM	24	x86
Hyper-V	3	x86
PowerVM	**0**	**POWER**

Source: National Vulnerability Database, http://nvd.nist.gov/home.cfm, July 2013.

NVD is the U.S. government repository of standards based vulnerability management data. This data enables automation of vulnerability management, security measurement, and compliance. NVD includes databases of security checklists, security related software flaws, misconfigurations, product names, and impact metrics. NVD is a product of the NIST Computer Security Division, Information Technology Laboratory and is sponsored by the Department of Homeland Security's National Cyber Security Division.

CVE is co-sponsored by the National Cyber Security Division of the U.S. Department of Homeland Security. Copyright © 1999–2012, The MITRE Corporation. CVE and the CVE logo are registered trademarks and CVE-Compatible is a trademark of The MITRE Corporation. This Web site is sponsored and managed by The MITRE Corporation to enable stakeholder collaboration.

Summary

Power Systems advantages over x86

- Performance
  - Highest single per core performance
  - Up to eightfold hardware multithreading (SMT=1,2,4,8) with Power8
  - Power Systems delivers more performance compared to x86 systems at a N-1 semiconductor manufacturing level.
- Scalability
  - Up to 256 cores
  - Up to 16 TB of memory
- Higher utilization of systems (>= 65% IBM guarantee)
- RAS (Reliability, Availability, Serviceability) capabilities
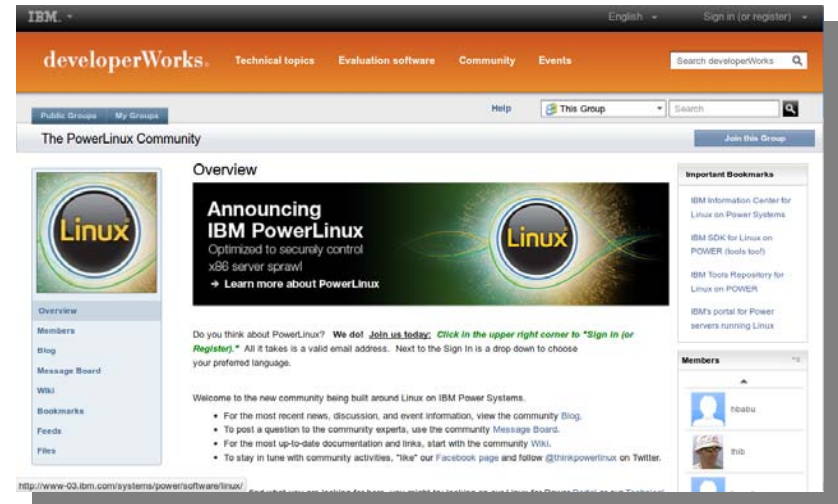
With PowerVM only:

- Redundancy of virtualization engine (VIOS) possible
- Secure by design: 'bare metal' hypervisor
  - PowerVM hypervisor is digitally-signed firmware with strong cryptography
  - Impossible to remotely install a modified fileset into the EPROMs of Power Systems

- Positioning
  - PowerVM → more towards enterprise environments (DBs, ERP, etc.)
  - PowerKVM → new applications (cloud (OpenStack), social, etc.)

Learn more about PowerLinux

Power Systems Linux Portal
(Product Information)

www.ibm.com/systems/power/software/linux/

The PowerLinux Community
(developerWorks)

www.ibm.com/developerworks/group/tpl/

@thinkpowerlinux

plus.google.com/communities/100156952249293416679

Questions ?

Thank you for your attention !
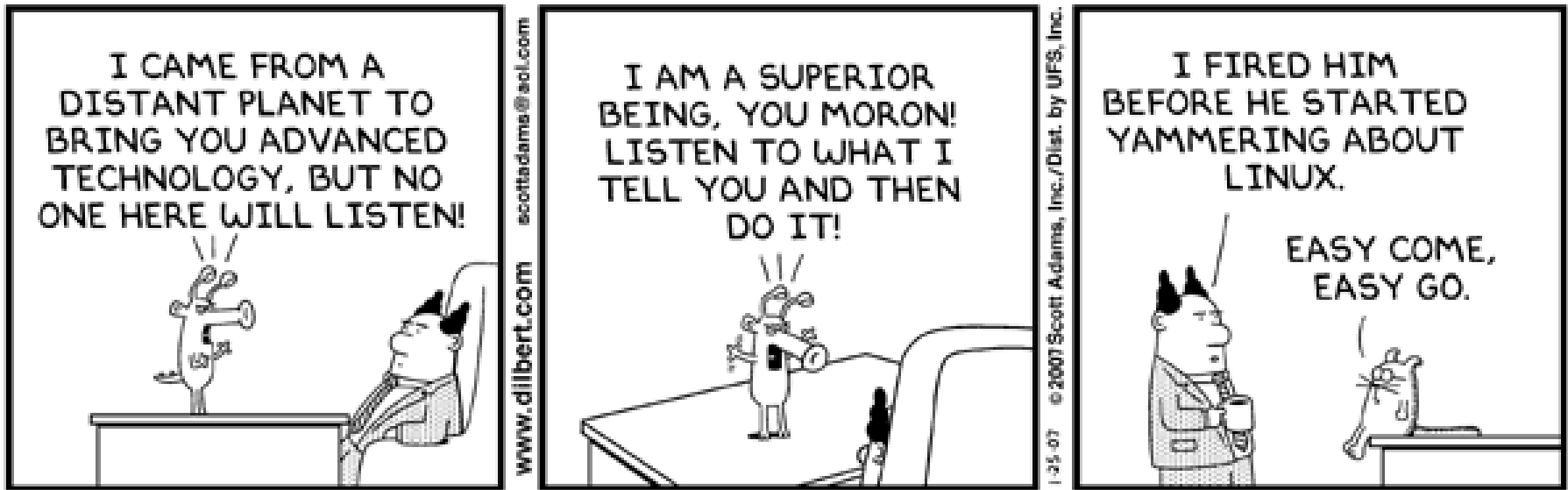
© Scott Adams, Inc./Dist. by UFS, Inc.

paper #20240: The Open Platform for Choice: Linux on Power Virtualization