Solaris and Xen

Stuart Maybee

stuart.maybee@sun.com

Solaris Kernel Engineering

Overview

- Hypervisors 101
- Paravirtualization vs unmodified OS
- About Xen
- Performance
- Migration
- Solaris on Xen
- Xen and Zones

Hypervisors 101

- Provides a "Virtual Machine"
- Not new – VM/370 over 30 yrs ago
- Controls hardware - memory/cpu/io devices
- Schedules cpus/memory/io rate
- May emulate real devices
- For x86/x86-64 multiple choices available:
  - > Xen
  - > VMware
  - > MSFT Virtual Server
  - > others

Para vs Full virtualization

- Full virtualization:
  - > Runs binary image of "metal" OS
  - > Must emulate real i/o devices
  - > Can be slow/need help from hardware
  - > May use trap and emulate or rewriting
- Para-virtualizattion:
  - > Runs OS ported to special arch
  - > Uses generic "virtual" device drivers
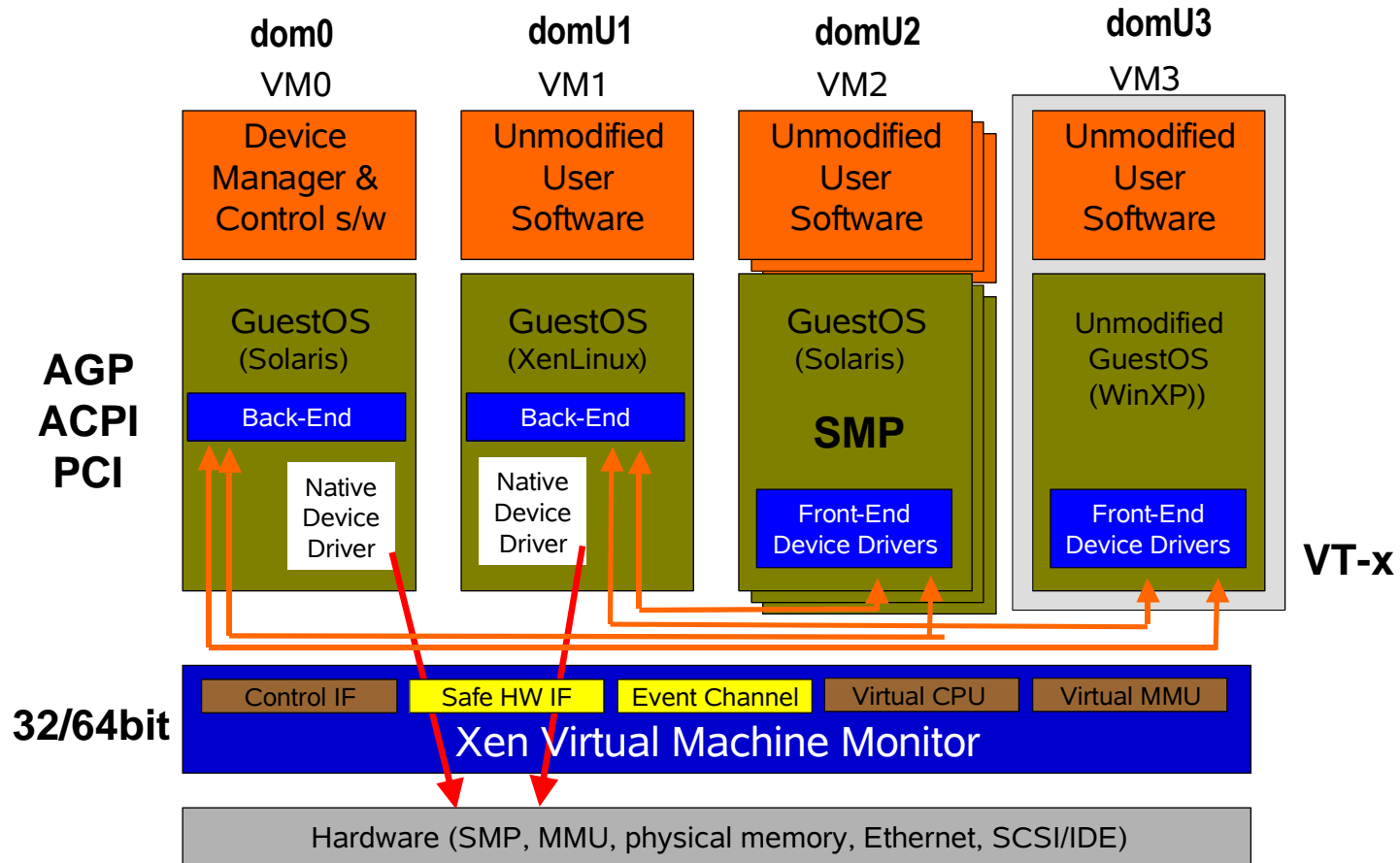 - > Can be more efficient since it is hypervisor aware

Xen

- Open source hypervisor technology developed at the University of Cambridge
  http://www.cl.cam.ac.uk/Research/SRG/netos/xen/
  http://www.opensolaris.org/os/community/xen

- 2006: Hardware Virtualization Everywhere
  - x64 cpu capabilities (VT-x, Pacifica)
  - Workload consolidation
  - Community software wanted!

  "Every grad student will have their own hypervisor"

Xen's Design Principals and Goals

- Existing applications and binaries must run unmodified

- Support for multi-process, multi-application application environments
  - > Permit complex server configurations to be virtualized within a single guest OS instance

- Paravirtualization enables high performance and strong isolation between domains
  - > Particularly on uncooperative architectures (x86)

- Support up to 100 active VM instances on modern servers

- Live migration of VM instances between servers

Xen 3.x Architecture

dom0
VM0

domU1
VM1

domU2
VM2

domU3
VM3

Device Manager & Control s/w

Unmodified User Software

Unmodified User Software

Unmodified User Software

**AGP
ACPI
PCI**

GuestOS (Solaris)

GuestOS (XenLinux)

GuestOS (Solaris)

Unmodified GuestOS (WinXP))

Back-End

Back-End

SMP

Native Device Driver

Native Device Driver

Front-End Device Drivers

Front-End Device Drivers

VT-x

32/64bit

| Control IF | Safe HW IF | Event Channel | Virtual CPU | Virtual MMU |

Xen Virtual Machine Monitor

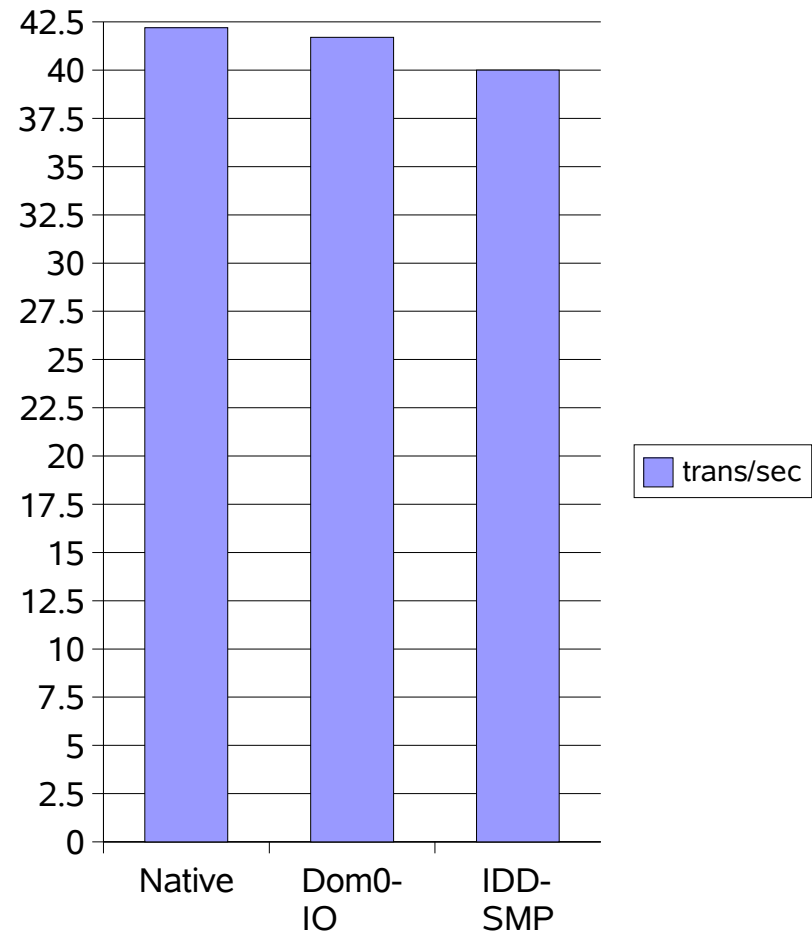Hardware (SMP, MMU, physical memory, Ethernet, SCSI/IDE)

Key Capabilities

- Checkpoint/Restart and Live Migration
  - N1 provisioning
  - Grid operations: virtual platform
- Multiple OSes running simultaneously
  - Linux, Solaris, Windows XP
  - No longer a boot-time decision
- Special purpose kernels
 - Drivers, filesystems

Performance - Postmark

- Emulates workload of mail server

- Number of transactions/second.

- 1% - Dom0 I/O

- 5% - IDD

Performance data from "Xen and the Art of Virtualisation", Pratt et al, University of Cambridge, UK
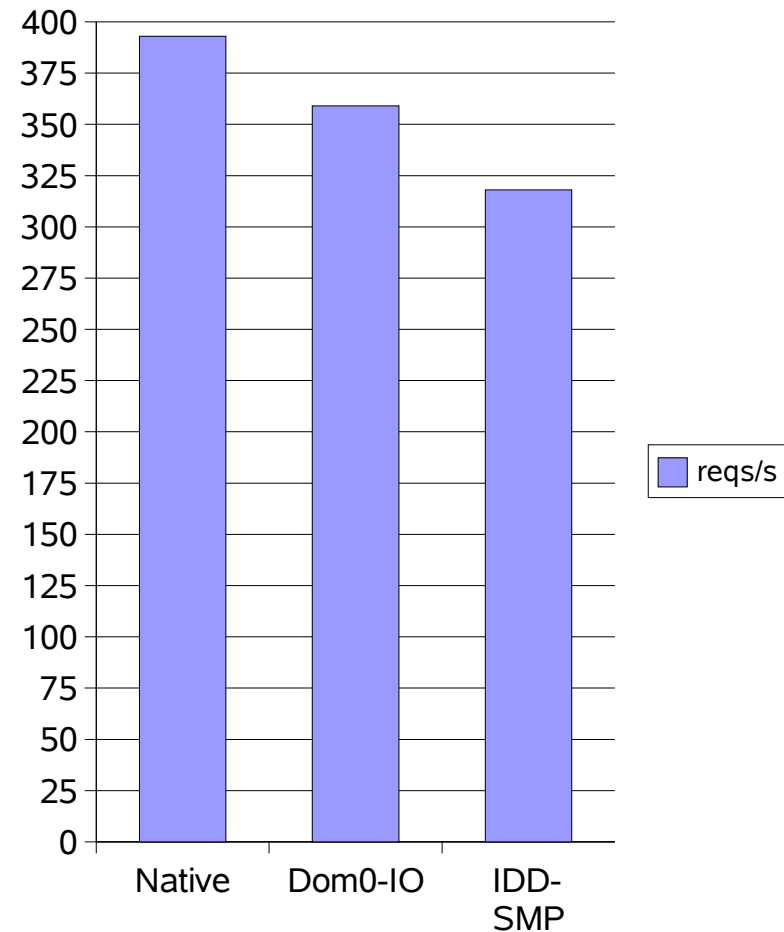
Performance - OLTP

OSDB/OLTP

- PostgresSQL 7.3.2

- Both Data Mining & OLTP access patterns

- No measurable impact on DM access (read-only)

- *IDD outperforms native (non-Xen) OS*
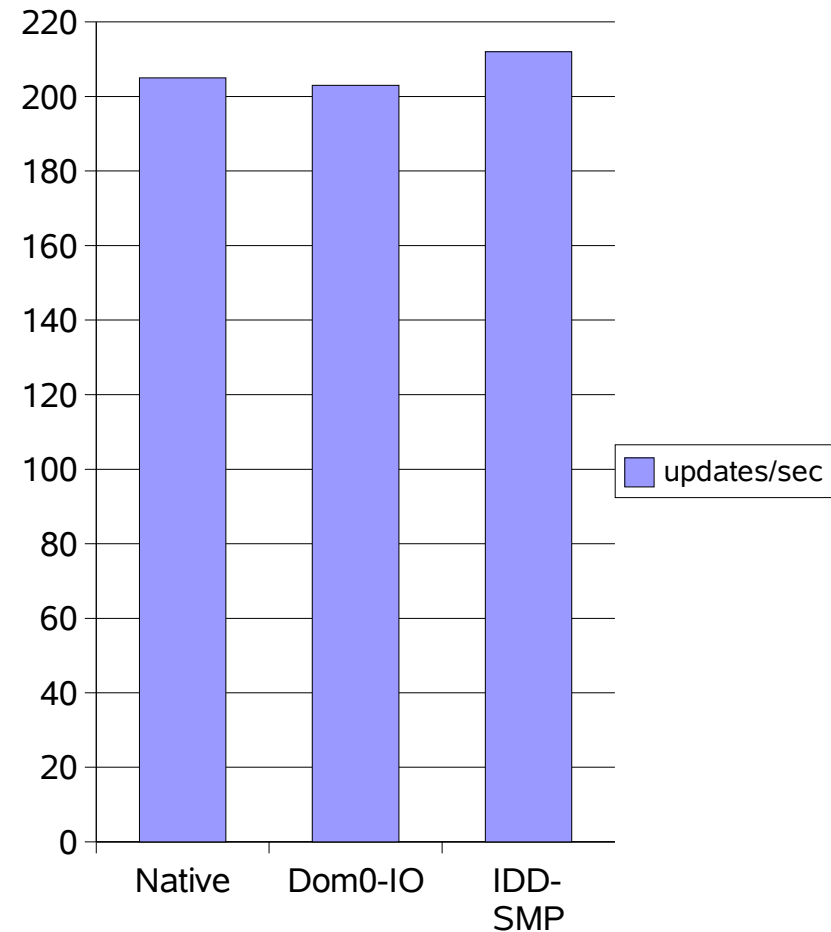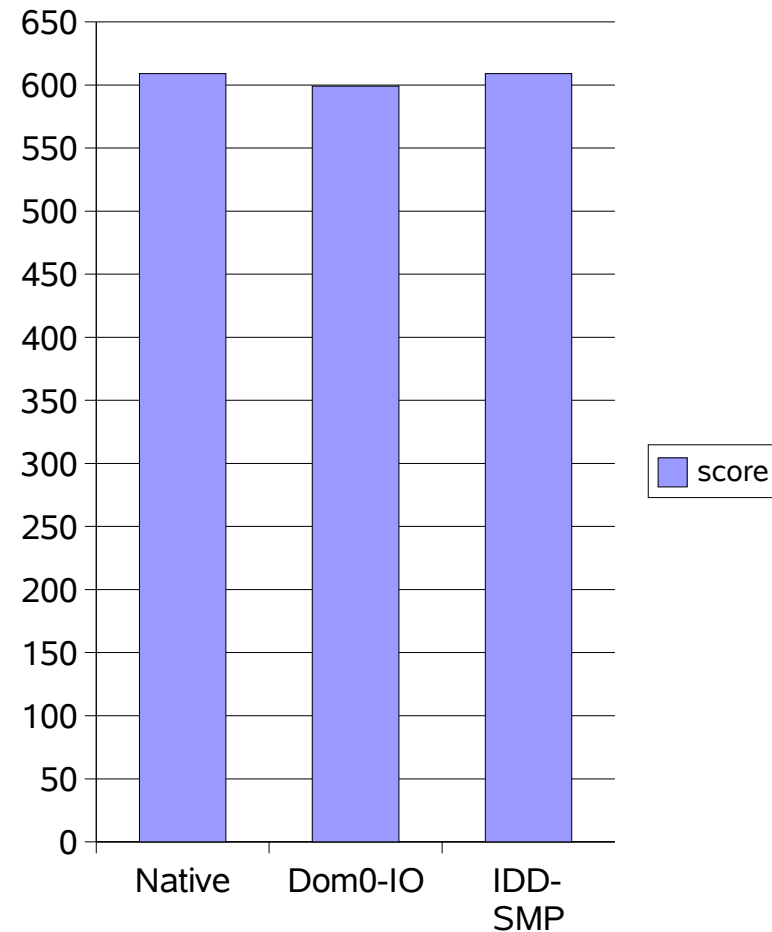
Performance - httperf
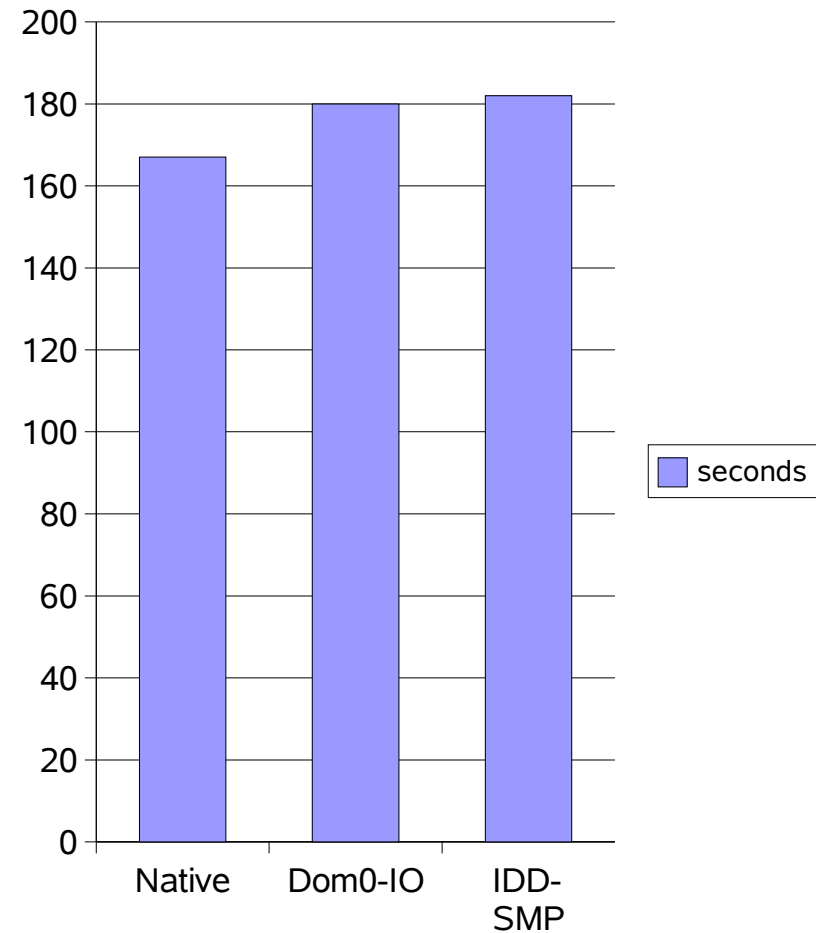
httperf-0.8

- Apache server
- 200Mb/s server

Performance – Spec WEB 99

- Benchmark score reflects QoS supplied by webserver

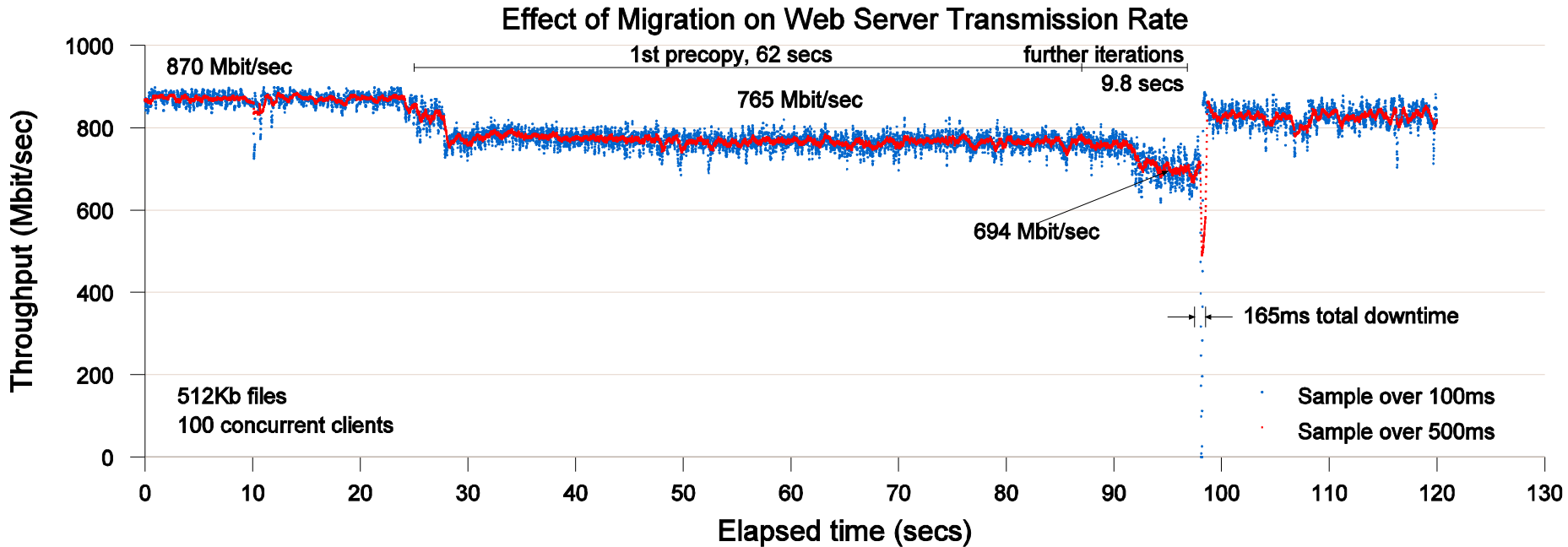# Impact On Performance

OS Build Time

OS Migration Experiment, 2004

- Two machines running Xen 2.0
  - 2GHz Hyperthreaded CPUs
  - 1Gbit Ethernet
  - Remote storage
  - XenoLinux
- SPECweb99 benchmark
  - 800Mbyte domU, 90% CPU utilization

Now, move workload from machine A to machine B ...

SPECweb99 Migration Experiment

Effect of Migration on Web Server Transmission Rate

From LinuxWorld 2005 Virtualization BoF

Solaris Xen Port

- "Platform" rather than "arch" port
- Wrote generic virtual device drivers
  - Net
  - Disk
  - Console
- Support Xen "event" interrupt model
- HAT layer changes for TLB/page table management
- Replace priviliged instructions with hypervisor calls

Xen and Sun - Opportunities

- "Deep-frozen" solution stacks
  - Preconfigured, suspended, ready-to-resume
- Embedded Stacks
  - Driver domains
  - Network switch
  - domU JavaOS
- Grid computing
  - "Internet Suspend/Resume"
- Massive Vertical Scale
  - Reliability

Xen and Sun – How to Play?

- Participate in Xen community
  - Fix bugs, improve stability and portability
- Innovate in Solaris-on-Xen
  - Fault Management, Observability
  - Performance, Scaling
  - Networking, Security, Manageability
- Manage Xen-based domains
  - N1P1

Solaris on Xen - Status

April 20, 2006

- In the lab in various states of functionality:
  - > 32-bit Solaris dom0, mp domU
  - > 64-bit versions of the same
  - > Currently Based on Xen 3.0.2 Solaris Nevada build 31 soon to move to Nevada 38
  - > Will show up on OpenSolaris when ready
- Solaris Nevada build 31 on Xen 3.0
  - > 32-bit, domU, uniprocessor
  - > Available now for download from OpenSolaris

Solaris Zones and Xen

- Zones
  - Scalable, fast, virtual platform, sparc & x64
  - Emphasis on sharing, simpler admin
  - Improved fault isolation over "single system."
  - Alternate brands
- Xen (hardware virtualization)
  - Familiar admin model
  - Emphasis on separation
  - Better performance isolation
  - Better fault isolation, SPOFs remain

Solaris Zones and Xen (continued)

- Complementary, not alternatives
  - Use one, the other, or both
- Key differentiators for HW virtualization
  - Different OSes on same machine
  - Different patch levels on same machine
  - Suspend/Resume, and Live Migration
- Technology Lifecycle
  - Zones – nearly two years old via Solaris Express, mission-critical deployment >today<
  - Xen – still being created, will take some time to mature

Summary

- Xen with Solaris...
  - provides new capabilities from a new virtualization layer
  - is a work-in-progress
  - will further commoditize HW virtualization

- Questions?

OpenSolaris on Xen Community

- Just getting going:
  - Email, blogs, FAQs
- Find it at:

  http://www.opensolaris.org/os/community/xen

- Plan
  - Develop Solaris-on-Xen in public view
  - Release ON codebase for port asap
    - Track OpenSolaris SCM tool(s)
- Details
  - NV b31 (and later) domU and dom0 plus patches for Xen 3.x