



What is Solaris Nevada? Nevada at 37

Daniel Price

daniel DOT price AT sun DOT com

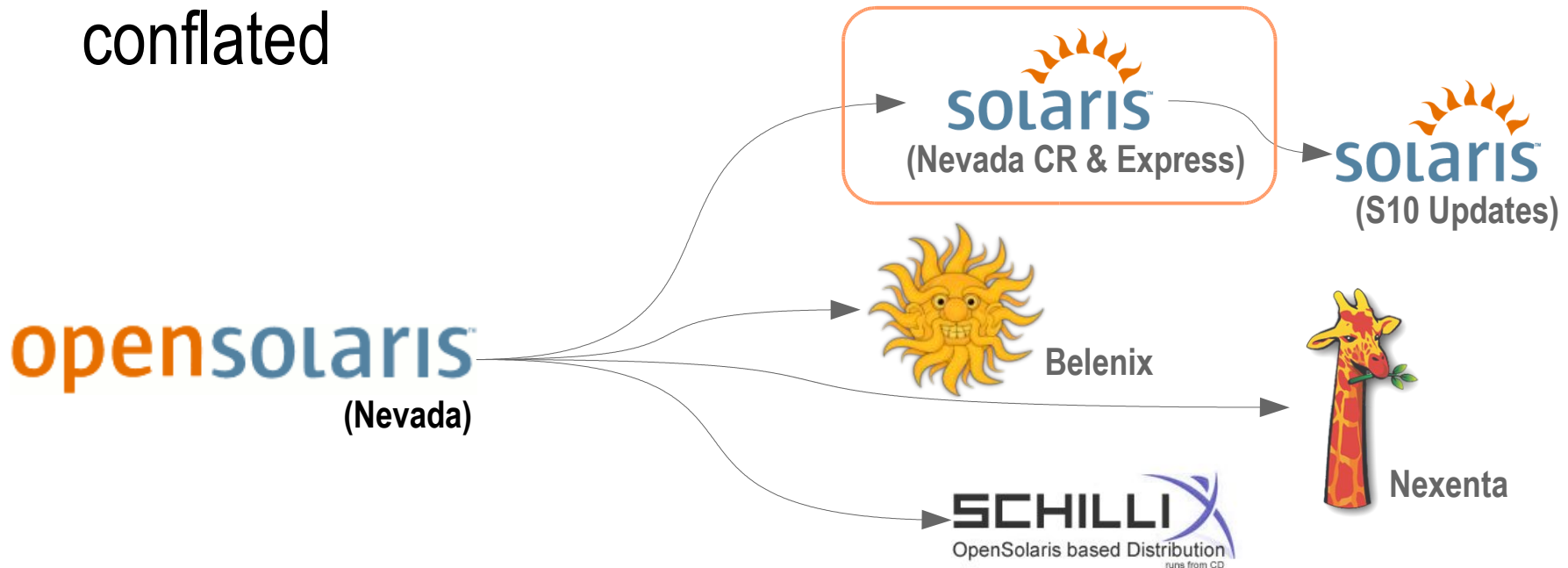
<http://blogs.sun.com/dp>

License

- This work is licensed under the Creative Commons Attribution-ShareAlike 2.5 License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-sa/2.5/> or send a letter to Creative Commons, 543 Howard Street, 5th Floor, San Francisco, California, 94105, USA.
- Copyright 2006 Sun Microsystems, Inc. All rights reserved.
- Use is subject to license terms.

What is Solaris Nevada?

- The next generation of Solaris
- A distribution based on OpenSolaris
- A community of engineers/developers/hackers
- “Nevada” codebase / “Nevada” distro codename conflated



37?

- Builds are released every 2 weeks.
- Build 37 is currently open.
 - > As of March 23, 2006, Nevada is 74 weeks old

By the Numbers

- ~7,500 bug fixes / RFEs (since S10)
- 45 community contributions
- Oldest bug fixed (so far):
 - > 1156383 Internet hosts requirements not met with respect to loopback addresses (filed 2/2/1994)

Participation

opensolaris™

- <http://www.opensolaris.org>
- <http://bugs.opensolaris.org>
- User groups worldwide
- Students: Solaris University Challenge
 - > Teams up to 4
 - > \$5000 per team member + Ultra20
 - > \$100,000 in gear for your school
 - > Ends June 2006

Solaris Enterprise System

- The whole stack is **free**
 - > Over time, open source as much as possible
 - > No support, no indemnification, but free!
- Free Components:
 - > Solaris 10
 - > Java Enterprise System (Access Mgr, Directory Server EE, ID Mgr, App Server, Msg Q, Web Server, Proxy Server, Calendar, Mail, Messaging, Portal Suite, Sun Cluster Suite)
 - > N1 (SPS 5.1, SMC 3.6, GridEngine, SMS)
 - > Studio 11, Java Studio Creator, Java Studio Enterprise
- Downloads immediately available
 - > <http://www.sun.com/software/solaris/get.jsp>

Nevada Themes

- System & Networking Performance
- Core Networking features and File Sharing
- Observability
- New x64 and SPARC Platform Support
- Device Support
- Desktop Goodies
- Virtualization: Zones, Xen, BrandZ
- Reduce Irritations: Approachability
- ZFS

Annotations

- Features already integrated
- *Features planned for Nevada in Italics*
 - > No guarantees: feature may not complete, or may not make the release
 - > This talk does not discuss what is in, or will, or won't appear in Solaris Update releases

System Performance

- Large text pages for executables & libs
- Large Page OOB (or Out-Of-the-Box)
 - > large pages automatically supplied to applications based on segment sizes, alignment and TLBs.
- Large Pages for Kernel Memory (LP Kmem) on sun4u
- Hierarchical L-group support
 - > Useful on 4-way+ Opteron Systems
 - > Unbundled: plgrp(1) lgrpinfo(1) tools
 - > <http://opensolaris.org/os/community/performance/numa>

System Performance

- Cool stuff
 - > 64-bit division on AMD-64 platforms: 50% faster
 - > SPARC RSA (for SSL, for example) in the kernel is now about twice as fast as before
 - > rand(3c), rand_r(3c), malloc(3c), free(3c) all faster
 - > SVM default interlace & resync buffer sizes expanded
- x86/x64
 - > Non-temporal load/store access optimizations for x64 performance
 - > Much faster memmove on x86/x64, strcpy, etc. faster on x86
- *Significant malloc rewhack in the works*

Core Network Performance

- GLDv3 (codename Nemo)
 - > Dynamic switching between interrupt and polling
 - > 10Gbps NIC support
 - > Vlan and Trunking support for off the shelf NICs
 - > Near linear trunk scalability: 4x1Gbps NICS → 3.6Gbps
 - > 7Gbps receive with 1500 byte frames!
- UDP Performance (Yosemite)
 - > “FireEngine for UDP”
 - > Improve TIBCO perf 70-80% (recv) 90-130% (xmit)
- *Forwarding Performance (Surya)*
 - > Close to 1 million pkts/sec forwarding on Opteron

Core Networking: Clearview

- Improve integration between key networking technologies
 - > IPMP, IP Tunnels, DHCP, Zones, Dynamic Reconfiguration, Network Observability Tools
- Vanity names for network interfaces
- More snoop-ability:
 - > Tunnels, inter-zone traffic, loopback, inside zone
- IPMP interfaces represent IPMP groups
- Unified interface provides `dladm(1m)` for all interfaces
 - > All interfaces: VLANs, aggregations, vanity names

Core Networking Crossbow

- Virtualize NICs based on protocol, service or container
- Enable interesting resource management configurations
- Advanced stack virtualization project
- Take advantage of NIC memory partitioning features

Network Driver Work

- 10GBs Drivers
 - > Neterion driver (xge) migrated to GLDv3
 - > Chelsio driver
 - > Intel 10GB (ixgb)
- e1000g (Intel gigabit) overhaul
 - > Migrated to GLDv3: supports vLAN, Link Agg., etc.
 - > Now also supported on SPARC
 - > H/W Checksum Offload
- Nvidia ck8-04 now supported
- RealTek 8169S (gigabit) now supported

Network Interoperability

- *Active Directory*
 - > Allow Solaris clients to be well behaved in a Microsoft administrative domain
 - > PAM, GSSAPI, Single Sign-On for MSPAC
 - > Unix GID, UID mapping to MS SID

Security Technologies

- Greyhound
 - > Kernel SSL proxy, offload userland SSL processing
 - > Boost SSL performance 20%-80%
 - > Cut copy cost and context switching
 - > Application talks cleartext to the proxy
 - > Results at <http://www.spec.org/web2005/results/web2005.html>

Security Technologies (2)

- *Trusted Extensions*
 - > Brings Trusted Solaris codebase into the existing Solaris Codebase
 - > Trusted Solaris will become an add-on
 - > pkgadd to regular Solaris to become Trusted
 - > Based on Solaris Zones
 - > Trusted Desktop being folded into GNOME as well

NFS

- 200 megabytes/sec (1.6Gbs) on x64/10Gbs gear
 - > *Future: async RPC, request scheduling. wirespeed!*
- Full End-to-End NFSv4 ACL support
 - > ZFS support for NFSv4 ACLs
- *Automounter improvements*
- *shareadm(1m) command*
- *High integrity checksums*
- *V4.1-- working on IETF specs*
 - > *Directory Delegations*
 - > *pNFS*

Observability

- DTrace
 - > First stages (POCs) of integration with higher level languages: java, php, perl, ruby, python
 - > ISVs pursuing stable providers
 - > *DTrace/Zones mashup (dtrace_proc and dtrace_user)*
 - > 3rd party toolkits evolving (google for “dtrace toolkit”)
 - > JNI bindings for Dtrace
- *fsstat: filesystem statistics tool*
 - > Additional features in planning
- *nfs4trace prototype*
 - > <http://opensolaris.org/os/project/nfs4trace/>

Fault Management

- To date: 25-40% reduction in annual downtime
- 2005 *InfoWorld* Innovation Award for “Predictive Self-Healing”
- Opteron/Athlon64/Turion FMA
 - > ChipKill on by default
 - > CPU & Memory error detection
 - > Offlining CPUs & Cores, memory retire
 - > *PCI-E to follow later*
 - > Works fine on non-Sun Opteron-based systems
- SPARC FMA
 - > US-IV+, US-T1, PCI-E I/O

Fault Management: Feature Work

- SNMP MIB for FMA, Trap generation
- *Predictive Self-Healing and Virtualization*
 - > *Distributed fault management with Xen*
- *FMA Sensors/Health Monitors*
 - > Common architecture for continuous sensor monitoring and analysis
 - > Pluggable backends to provide raw sensor readings
 - > IPMI, perf. Counters, SMBus, SMART, etc.
- *FMA Simulator and Tools*
 - > End-to-end scenario design, injection, verification

x64/x86 Support

- Isimega driver
 - > PERC 4e/Si, PERC 4e/Di, PERC 4e/DC and MegaRAID 320-2e, 320-2x adapters
- Adaptec SATA raid improvements
- ACPI improvements
- SMBIOS support
 - > prtdiag on x64/x86, smbios(1m) command

x64/x86 Support: New Boot

- Goal: Simplify x86 boot, add features, improve first impressions, reduce boot time.
- No more real mode drivers or configurator!
- GRUB is used as the booter
 - > Other OS's detected at installation
 - > Boots RAMdisk “boot archive”
 - > Recovery archive supplied
 - > pxegrub used for PXE boot
- Facilitates USB boot, Live CDs, *ZFS root*

UltraSPARC-T1 Platforms

- Sun-Fire T1000/T2000 Released
 - > CMT: 8 cores * 4 threads/core = 32 threads
 - > Shared 3MB 12-way set associative L2 cache
 - > 4 on-chip DDR2 channels: 25.6GB/s total bandwidth
 - > ChipKill support (survive failure of a DRAM on a DIMM)
 - > Largest: 8-cores, 4 10K rpm SAS 2.5" disks, 32GB RAM
 - > Chip < 65W, System: 325W (nominal).

UltraSPARC-T1 Platforms (2)

- In Solaris: new sun4v architecture (still SPARCv9)
 - > FMA support; offlining of cores and strands
 - > CPC support (CPU perf. counters)
- Scheduler support:
 - > Load balancing across cores
 - > 32 cpus visible via psrinfo(1m)
 - > *Shared per-core run queues*
- Optimized crypto provider for public key crypto

Eco-friendlier x86: Tesla Project

- *E-Star compliance for Sun x86 Workstations*
- *Suspend to RAM (ACPI S3)*
- *Longer Term:*
 - > Powernow & Speedstep (and on MP systems)
 - > Not just laptops and desktops. Servers too!
 - > Reduce TCO by improving performance per KW/H
 - > Automatic CPU speed reduction
 - > Scheduler improvements
 - > Peripheral power management

Mobility Support

- Wireless command line tool
 - > wificonfig(1m)
- Driver suite
 - > In Nevada
 - > Atheros Chipset driver (ath)
 - > For download
 - > Upgraded cardbus driver
 - > Intel Pro/2100B (ipw), Intel Pro/2200BG 2915ABG (iwi), Prism II (pcwl), Cisco Aironet 340/350 (pcan)
 - > Ndis wrapper driver (bcmndis) for broadcom chips
 - > <http://opensolaris.org/os/community/laptop/wireless>

Device Support: Enterprise Storage

- Open Source as of 1/30/2006
- iSCSI initiator support
 - > iscsiadm(1m)
 - > Multiple sessions/target supported
 - > IMA (management API), iSNS support (naming services)
- Expanded HBA (Qlogic, Emulex), target support
- MPAPI support
 - > SNIA standard for multipath monitoring and control
- fcinfo(1m) command
- MPxIO support for many more devices
- CompactFlash ↔ ATA support

Device Support: SATA Framework

- Goals
 - > Avoid ATA framework
 - > Reuse SCSI as much as possible-- including device naming
 - > Support SATA hotplug
 - > Simplified HBA programming model
- Supported devices thus far
 - > Silicon Image 3124/3132
 - > Marvell 88SX: 5040, 5041, 5080, 5081, 6041, 6081
 - > No list just yet of which cards contain these chips

Device Support: Desktops

- Dcam1394 driver on x86
- USB auto-magic (ex. just plug in your digital cam and run gtkam)
- cdrecord now bundled
- dvd+rw tools bundled (for DVD writing)
- *Tamarack (vold.next & FreeDesktop HAL)*
 - > Replace vold with more modern implementation
 - > Integrate DBUS and do HAL port, integration
 - > Improve removable media security model
 - > Integrate smoothly with GNOME

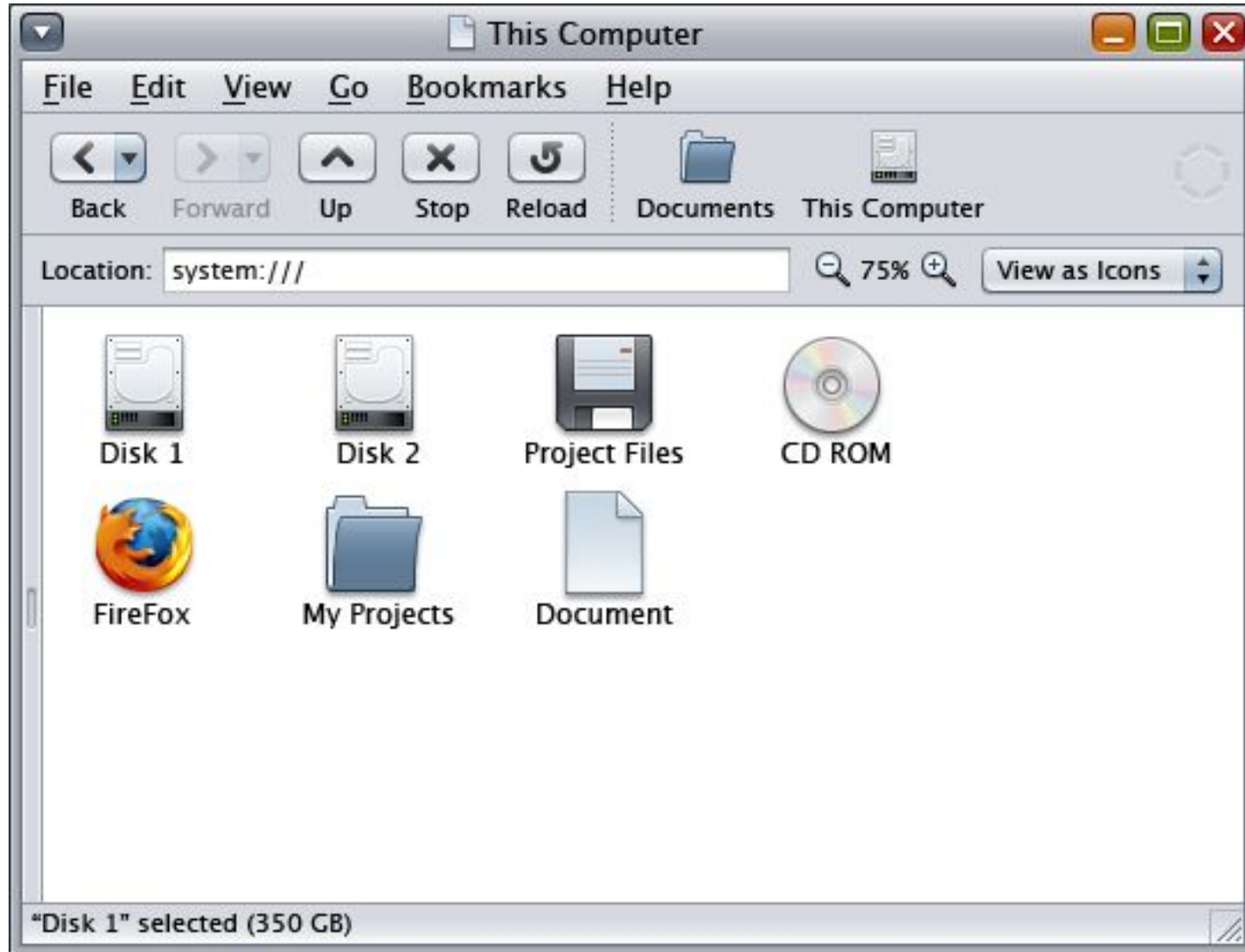
X Windows

- Xorg latest (currently in nv_36: Xorg 6.9 final)
- *nVidia driver, OpenGL for x86/x64*
 - > Download available. google: “solaris nvidia driver”
- MESA support
- *DRI support for H/W acceleration on Intel & ATI gfx*
- *Xorg work:*
 - > *Xorg on SPARC for XVR-2500 and future graphics cards*
 - > *Xorg support for Looking Glass*
 - > *64-bit Xorg server*
- *Migrate client apps and libs to X11R7*

Desktop Technologies

- *JDS4 (likely to be based on GNOME 2.14)*
 - > Current (2.12) snapshot released on opensolaris.org
 - > Improved L&F
 - > Improved applications: Evince, Ekiga (GnomeMeeting)

Desktop Technologies



Desktop Technologies (2)

- *Unified printer administration*
- Gnome-pilot with USB syncing (and Evolution support)
- RealPlayer 10 (/usr/bin/realplay)
- Acrobat 7 (only for SPARC) (/usr/bin/acroread)
- dvd+rw tools
- *Move from Mozilla to Firefox, Thunderbird*

Virtualization: Zones

- Management features: rename, move, clone
- “Attach”, “Detach”
 - > Foundation for moving zones from one system to another
 - > Should be helpful for backup, restore
- *ZFS filesystem autocreation at install*
- *Use of ZFS snapshots & clones for instant provisioning*
- mount/unmount functionality for upgrade

Virtualization: Zones (2)

- Configurable privileges
 - > Allows zones with more or fewer than the “default” set of privileges
- *Dtrace subset inside zones via `dtrace_proc` and `dtrace_user` privileges*
 - > Safe and virtualized to be scoped appropriately
- *Deeper SMF integration*
 - > Service instance per zone is likely
 - > Move configuration backend into SMF?
- *APIs for developers*

Virtualization: Zones Networking

- *Snoop support*
 - > *Global zone observability of cross-zone traffic*
 - > *Snoop support inside a zone*
- *Using DHCP to obtain a zone IP address*
 - > *(trickier than we thought)*
- *Allowing the global zone to NAT for other zones*

Virtualization: BrandZ

- “Branded Zones” present a non-native system model
 - > Brand is a zone attribute
 - > Brands provide custom installation routines
 - > Leverage existing zonecfg/zoneadm toolset
- Initial brand will be lx, supporting RHEL3
- Available today in preview form

Virtualization: Resource Management

- Resource Pools
 - > Convert to SMF services
 - > Disentangled from Java, making minimization easier
 - > *Pools ease of use*
- *CPU caps*
 - > *Based on FSS; Control maximum CPU usage*
- *Physical Memory Control*
 - > *Plugs into Resource Pools framework*
 - > *Workloads page against themselves when limit is reached*

Virtualization: Xen

- Xen is an open source hypervisor from Cambridge
- Xen is "paravirtualized" (OS has to be modified -- much better performance than VMWare)
 - > Ability to run other OS' (Linux, BSD, Windows), CPR, live migration
 - > Solaris port in progress
 - > dom0 and domU (Dom0 leverages Solaris' quality/scalability strengths)
- Initial demo & source for Solaris domU port to Xen 3.0 available on opensolaris.org
- Utilize upcoming CPU virtualization technology (AMD SVM, Intel VT) to run fully virtualized OS's

Freeware in Nevada: SFW

- “SFW” consolidation delivers **/usr/sfw**
 - > Components carry some level of support
- Pursuing enhanced free software strategy; hoping to leverage community interest
- *SFW Consolidation source tree / build env will be publicly released soon*
 - > Apache, Ant, bash, bind 9, cdrtools, GNU tools, gcc, gdb, glib, gtk+, Postgres, Tcl/Tk, Tomcat, zsh, Webmin, SNMP, Samba, Zebra, etc...
- *Sponsorship based contribution model*

Freeware in Nevada: Companion “CD”

- Companion software consolidation delivers `/opt/sfw`
 - > Components are unsupported, volatile
- *Hosted in SVN repository*
- *Direct contribution model*
 - > *Integrate directly after code review and c-team review*
- *Continuous delivery model to the web*
 - > *Potential to keep /opt/sfw continuously up to date*
- *Actively seeking maintainers/owners for packages*

Approachability: Reducing Irritations

- Remove barriers to adoption
- It just works, it's obvious, it's beautiful; plays well with others.
 - > Data should be easier to manage
 - > The system should be elegant and obvious to maintain
 - > Patch, install, packaging all need love

ZFS

- Blow away 20 years of assumptions about storage
- An always-checksummed copy-on-write filesystem
- Unlimited instant snapshots, clones
- Efficient remote replication
- Pooled storage eliminates storage “stranding”
- Radically simple system administration
- Integrated with FMA, Zones
- *ZFS Boot (aka ZFS Root) in the works*
 - > Imagine a world without UFS...

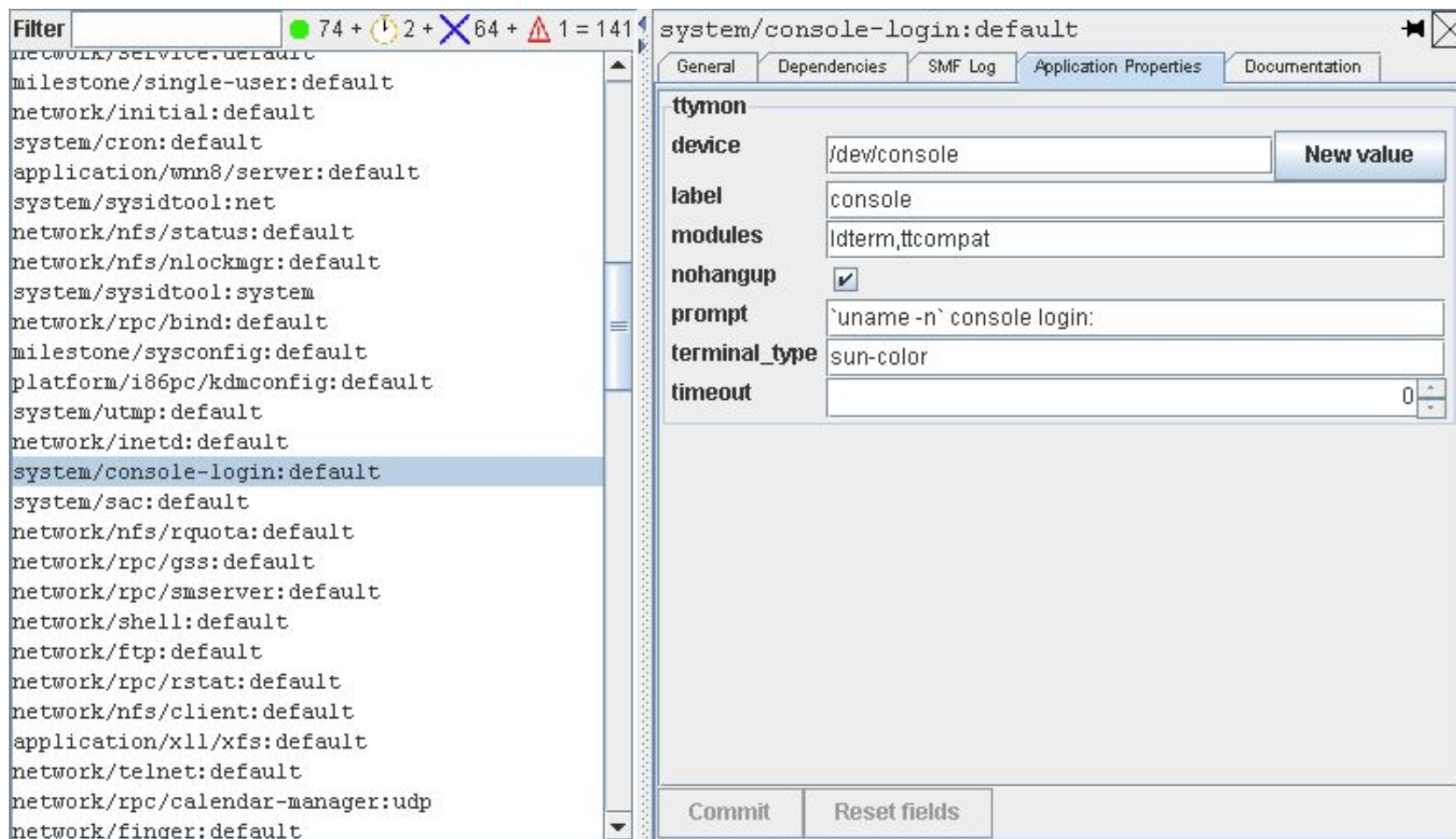
Service Management (SMF)

- Public manifest class-action scripts
- *Allow property customization during jumpstart and in profiles*
- *Notifications on service state transitions*
 - > *Send an SNMP trap, Send an email, run a script*
- *svcs -x diagnosis for inetd services*
- *Expanded template support*
- *More service conversions*

System Management

- “Visual Panels” proof of concept
 - > <http://opensolaris.org/os/project/vpanels>
- *Merged preferences management*
 - > Auto-generate GUIs as much as possible
 - > Build JMX agents once, reuse across many products
 - > <http://java.sun.com/products/JavaManagement>

System Management



The screenshot displays a system management interface. On the left, a list of services is shown, with 'system/console-login:default' selected. On the right, the configuration details for this service are displayed in a tabbed window titled 'system/console-login:default'. The tabs include 'General', 'Dependencies', 'SMF Log', 'Application Properties', and 'Documentation'. The 'General' tab is active, showing the following configuration:

Property	Value
ttymon	
device	/dev/console
label	console
modules	ldterm,ttcompat
nohangup	<input checked="" type="checkbox"/>
prompt	`uname -n` console login:
terminal_type	sun-color
timeout	0

At the bottom of the configuration window, there are two buttons: 'Commit' and 'Reset fields'.

Network “Auto Magic”

- Simplify and automate network configuration
- Integrate “Bonjour” service discovery
- Improve duplicate address detection
- Establish a mechanism for storing network information in a profile and auto-activating that as needed

Install/Packaging (short term)

- SVR4 packaging code released
 - > http://opensolaris.org/os/project/svr4_packaging
- *Install performance improvement*
 - > Trouble spots known, now we need to fix them...

Install/Packaging (longer term)

- Strategy draft open for comment until 14 Apr 2006
 - > <http://opensolaris.org/jive/thread.jspa?threadID=7070>
- Simplified system configuration, integrated with the post-installation experience
- Updated and simplified graphical and text interfaces
 - > Live CD/DVD experience
 - > Integrated hardware compatibility testing
 - > Integrated partition and filesystem resizing for multi-OS installs on x86
- Deep integration with latest features: ZFS and SMF
- Simple, fast, and reliable installation of additional software after installation

Good Night, and Good Luck

- These things are EOF (deleted), or will be soon:
- *DMI (Desktop Management Interface)*
 - > *Maybe also SEA (Sun Enterprise Agents)*
- ASET (Automated Security Enhancement Tool)
- AT&T FACE (Framed Access Command Env.)
- chs(7d) (IBM ServeRAID), dbri(7d) (ISDN) drivers
- pcscsi(7d) driver (some old scsi adapter)
- SunButtons/SunDials (bd(7M))
- *SBUS & UPA framebuffers (maybe)*
- *STSF (Standard Type Services Framework)*

Cool Things for the Community to Do

- File bugs: <http://bugs.opensolaris.org>
- SMF service conversions for free software
 - > Get them into the base source distributions
 - > Apache, Squid, MySQL, Postgres, ...
- MythTV/Freevo
- More & Better: WineX, pearpc, bochs, Qemu, UML
- Build something out of Java/DTrace
- Help the Earth: Speedstep, PowerNow!
- Drivers: bluetooth, usb-networking, irda
- Give this presentation to your local user group

Conclusions

- It's the people!
- Nevada is looking healthy

What is Solaris Nevada? Nevada at 37

Daniel Price

daniel DOT price AT sun DOT com

<http://blogs.sun.com/dp>