ORACLE®

What's New In Oracle Solaris 10 9/10

Isaac Rozenfeld
Oracle'iz'ing Solaris
http://blogs.sun.com/unixman

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions.

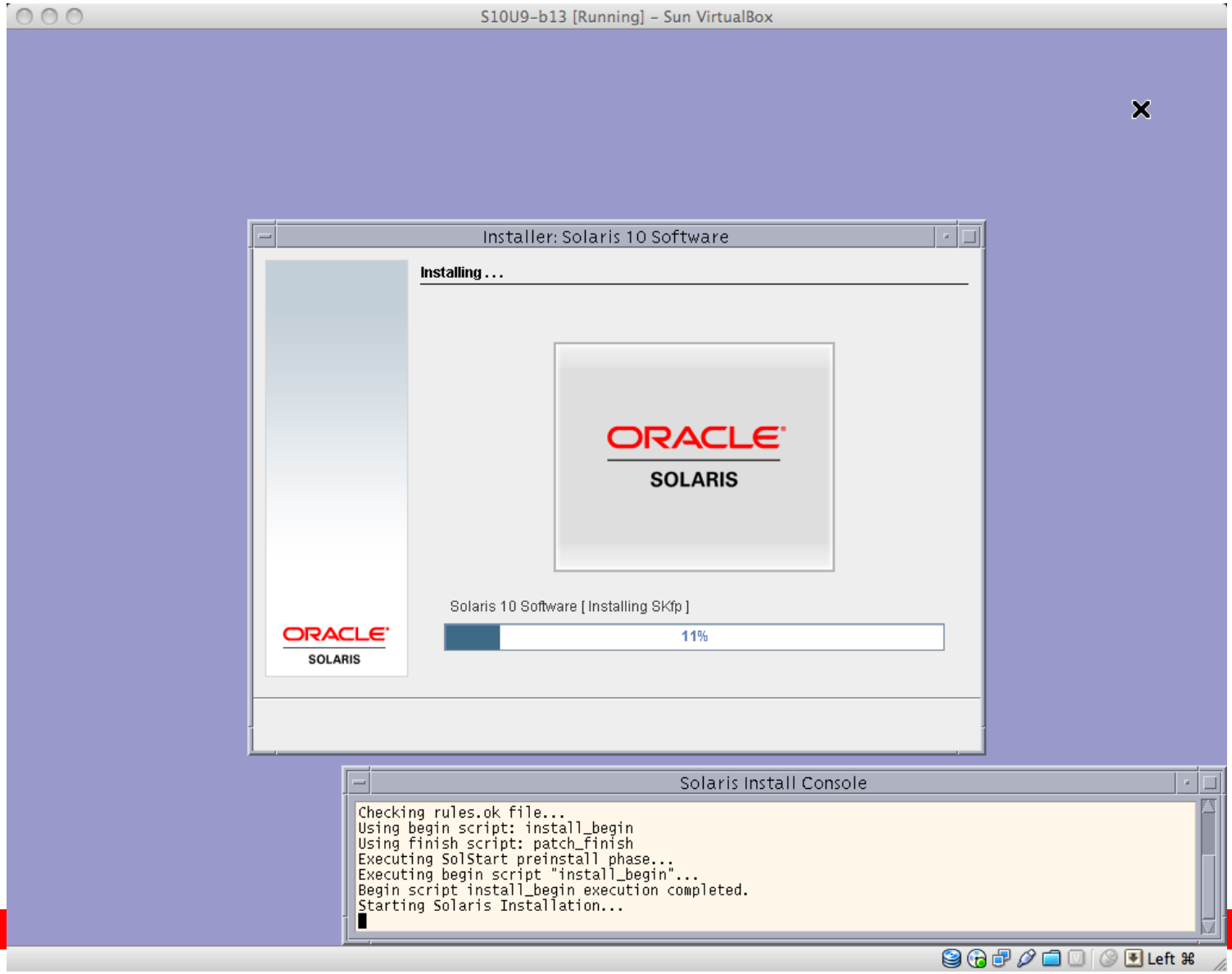
The development, release, and timing of any features or functionality described for Oracle's products remains at the sole discretion of Oracle.

The following is intended to provide information for Oracle and Sun as we continue to combine the operations worldwide. Each country will complete its integration in accordance with local laws and requirements. In the EU and other non-EU countries with similar requirements, the combinations of local Oracle and Sun entities as well as other relevant changes during the transition phase will be conducted in accordance with and subject to the information and consultation requirements of applicable local laws, EU Directives and their implementation in the individual members states. Sun customers and partners should continue to engage with their Sun contacts for assistance for Sun products and their Oracle contacts for Oracle products.

Solaris 10 U9 Key Facts

- ~14 Builds
- Solaris 10 9/10
- Oracle branding changes
- Oracle University updating Solaris Curriculum training
- Immense focus on quality and integration work
- Hardware Platform support enhancements; MU9 for boot
- Virtualization Portfolio enhancements
- Solaris 10 is part of DB and Middleware nightly validation tests

Installation Re-branding Example



Release Highlights

- More SPARC, Intel, AMD platforms supported
- Lots of Enhancements to Virtualization
  - Oracle Solaris Zones
  - Oracle VM Server for SPARC (Logical Domains)
- iSCSI boot support for T-series and X-series
- I/O management and performance
- Lots of ZFS Enhancements
  - ~150 bugfixes/RFE's (snv_132)
- Service Tags 1.1.5 (snv_127)
- Patching improvement in determining space

Solaris Auto Registration

- Oracle Solaris Auto Registration is new (optional)
  - During the initial reboot, after installed/upgraded, configuration data bout your system is automatically communicated through the existing service tag to Oracle Product Registration System.
  - Goal of using service tag data about your system to help Oracle enhance customer support and services.
  - More details about Service Tags *http://wikis.sun.com/display/ServiceTag/Sun+Service+Tag+FAQ*

Solaris Auto Registration (2)

- Oracle Solaris Auto Registration is new (optional)
  - You can elect to have your configuration data sent anonymously
    - Config data sent to Oracle would have no link to customer name
    - You can disable Auto Registration altogether
  - You can use the same config data to create and manage entire inventory of systems
  - Registering with your support credentials using one of the registration options, you have a way to inventory your systems, by recording and tracking the service tags for the systems and for the s/w products installed
  - For instructions on tracking your registered products, see: *http://wikis.sun.com/display/SunInventory/Sun+Inventory*

Solaris Zones Enhancements

- Non-global zone unique hostid emulation (snv_108)
  - Allows GZ administrator to set a unique hostid for a NGZ
  - Needed by software that's licensed to a specific hostid
  - Particularly useful when virtualizing older environments on newer/modern hardware whilst preserving full functionality
- Physical-to-Virtual (P2V) for native zones (snv_109)
  - Continues to leverage and demonstrate value of Zones
  - Ability to migrate existing Solaris 10 physical nodes and treat them as Zones going forward
  - For the "native" brand and it will support installing from flash archives, along with all of the other archive formats that p2v of s8 & s9 have supported
  - Side-note: solaris10 Brand in development for Solaris Next

Solaris Zones Enhancements (2)

- Improved *zoneadmd*
  - Better integration with Oracle Cluster brand
  - Ability to upgrade Solaris on a Zone-cluster node (which is a cluster brand zone) via CD or Jumpstart; previously only doable by Live Upgrade
- Use of *zoneadm attach -U* for patching
  - Updates all zone's packages so that the rev's of packages would be the same had this zone actually been installed from scratch (as opposed to being attached)
  - To update packages on a system, detach a zone, update the Global zone and then re-attach all zones with *-U*

Solaris Zones Enhancements (3)

- No reboot upon TZ update
  - RFE by customers with 100+'s of domains on high-end systems
- Oracle Solaris Cluster 3.3 is available
  - Oracle WebLogic, Siebel CRM, SAP ERP are supported in clustered Zones

Oracle VM Server for SPARC

- Cooperative Guest Domain Migration (snv_129)
  - Next phase of migrating Logical Domains
  - Critical LDoms 2.0 Virtualization feature
  - Lifts certain previously enforced requirements for Warm Migration
- Enhanced sun4v diagnostics (snv_106)
  - Improves CPU and Memory diagnostics for T5440's
- Support for Rainbow Falls (snv_131)
- CMT observability APIs (snv_131)
- Console logging enhancement for sun4v (snv_120)

Oracle VM Server for SPARC (2)

- n2cp to support CKM_AES_GCM (snv_107)
  - Multi-threaded, loadable device driver supporting hw-assisted acceleration of crypto ops on a T2 processor
- Virtual Disk Enhancements (snv_128)
  - vdisk handling path failures (a la mpxio)
  - Virtual disk multipathing allows you to configure a virtual disk for a domain, and have it's back-end storage be accessible by more then one path
  - The paths would lead through different service domains
  - Improves redundancy if one service domain were to fail
- Static Direct IO (snv_133)
  - Combined with LDoms 2.0, guest domains can directly access PCIe endpoint devices.
  - Used to obtain higher granularity and throughput/latency of device access for guest domains

Oracle VM Server for SPARC (3)

- LDoms Memory DR (snv_120)
  - Dynamically add/delete memory to/from a LDom
- Logical Domains Agents (snv_124)
- /usr/sbin/virtinfo and libv12n.so libraries (snv_135)
  - Offers easy identification of LDoms domain capabilities and features, such as: *what type of domain this is, a guest domain or control domain, what the domain name (the name as used by the LDoms manager) of this domain is, what the UUID of this domain is, what the control domain network nodename is and what the chassis serial number of the platform its running on*

More SPARC Enhancements

- Support for ITU construction tools
  - Modified *itu* utility to allow invocation of install-time update process
  - 3rd party vendors can now deliver updates on CD/DVD/USB
  - Gain the benefit of modifying installer media with new packages/patches that might be introduced
  - Aids in delivering software updates for hardware platforms and creation of customized (minimized!) install media
    - itu(1M)
    - mkbootmedia(1M)
    - pkg2du(1M)
    - updatemedia(1M)

More SPARC Enhancements

- Automatic Boot Archive Recovery
  - See the man page for boot(1M)
- Netlogic NLP2020 PHY in *nxge* driver
  - Provides 10G QSFP support on RF, Seville and Solana platforms
  - Achieves 40G from 4 x 10G NIU ports justing one QSFP connector/cable instead of 4 x 10G SFP+/XFP connectors/cables

x64/x86

- FMA support for AMD Magny Cours processor (snv_128)
  - AMD's Multi-chip module
  - FMA CPU and Memory topology enhancements
- FMA support enhancements for:
  - Nehalem EX (snv_127)
  - Jasper Forest (Hamilton) (snv_125)
  - Westmere (Lynx+, Virgo+) (snv_125)
- FBDIMM idle power enhancement (snv_116)
- Intel Westmere CPC updates (snv_134)
- AMD Shanghai/Istanbul performance counters (snv_114)

x64/x86 (2)

- Intel Xeon 5600 series
  - supports IA32_ENERGY_PERF_BIAS Model Support Register
  - You can set the MSR to the desired energy and performance bias
  - Done at boot time in /etc/system, via:
    - `set cpupm_iepb_policy=`value``
    - value is a number from 0 to 15
- 64-bit `libc` string functions improvement with SSE
  - String functions have been enhanced with streaming SIMD extensions
  - Offer performance improvements in `strcmp()`, `strcpy()`, `strlen()` for 64-bit applications
  - If copying or comparing strings of 2 MB or more, use `smemcpy()`, `memmove()` instead

x64/x86 (3)

- Automatic Recovery of boot archives
  - A new *auto-reboot-safe* property added to boot configuration service, *svc:/system/boot-config:default*
  - Default value is *false* to prevent booting into an unknown boot device
  - If the system is configured to point to the BIOS boot device and default GRUB menu entry where Solaris is installed, set to *true*
    - Enables auto-reboot of the system to recover an out-of-date boot archive
  - Use *svccfg* and *svcadm* to change the property's value
- Broadcom NetXtreme II 10 GE NIC driver (bnxe)
- Broadcom HT1000 SATA Controllers (bcm_sata)
- SATA/AHCI Port Multiplier (ahci)

x64/x86 (4)

- Intel AES-NI Optimization
  - Advanced Encryption Standard has been accelerated by Intel through the introduction of New Instructions beginning with Xeon 5600-series
  - There are 6 instructions and they offer significant performance on AES operations
    - Reduces CPU overhead during IPsec operations
    - Tests show 50% CPU reduction compared to Xeon 5500 CPUs
  - Automatically detected and used by Solaris Cryptographic Framework

I/O

- Solaris Hotplug Framework (snv_128)
  - A common hotplug framework that can support the *hotplug* functionality for all industry standard PCI, PCI Express, IOV, Express Card, Cardbus and other virtual IO devices. The main goal of this project is to provide a generic common Solaris hotplug framework - a foundation which can support hotplug functionality for any hotpluggable bus
  - Use *hotplug* to enable/disable/offline/online supported PCI devices
- Fast Crash Dump (snv_127)
  - Saving a kernel crash dump takes a long time on machines with large memory; this project greatly reduces such time by 2-10 times
  - Equally reduces needed disk space in the `savecore` directory
  - New crash dump file, *vmdump.n*, is a compressed version of *vmcore.n* and *unix.n* files
  - To analyse with *mdb* use `savecore` to uncompress first

I/O (2)

- Multiple Disk Sector Size Support (snv_118)
  - Large block size for disks and SSDs
  - From current hard disks where sector size is 512 bytes to larger disks where sector sizes are 1K, 2K or 4K
  - Large sector size can provide higher reliability, correct labeling; greater efficiency for data transfer (raw & block), which means less overhead per sector written and read; less format time; and faster drive maintenance
  - Support for ZFS nonroot disk
  - Xen and Oracle VM Server for SPARC identify large sector sizes
- SSD Data Alignment (snv_122)
- HP SMART Array RAID device driver, cpqary3 (snv_131)
  - Integrated into Solaris 10; previously was a separate download from HP

I/O (3)

- iSCSI Initiator Parameters
  - Enable tuning iSCSI initiator parameters for accessing a given iSCSI target
  - Greatly improves iSCSI initiator connection response time for various network scenario
    - Particularly when the network between the iSCSI initiator and the target is slow or unstable
  - Use `iscsiadm` command or the library `libima` interface
- iSCSI Boot
  - Allows booting the OS over the network from a remote location
  - Loaded onto a diskless client with root residing on an iSCSI target that is on the network
  - Leverages standard Ethernet-based infrastructure

I/O (4)

- iSER Initiator
  - iSCSI Extensions for RDMA accelerate iSCSI protocol by mapping the data transfer to RDMA operations
  - This results in iSER initiator reading/writing data from iSER target at high data rates with low CPU utilization compared to iSCSI over TCP/IP
  - See iser(7D) man page
- IPoIB Connected Mode
  - An improved capability in the IPoIB driver
  - Instead of using UD datagrams for communication, now establish connected mode channels to peer nodes (with such capabilities)
  - Improves latency and bandwidth while lowering CPU utilisation
  - Automatically reverts to UD datagram mode if peer unable to support

I/O (5)

- Open Fabrics User Verbs Primary Kernel Components
  - Open Fabrics Enterprise Edition (OFED) RDMA KPIs
  - Kernel modules and drivers use OFED KPIs
  - Delivers kernel components required to interfaces OFED libs into InfiniBand Transport Framework (IBTF)
- InfiniBand Infrastructure Enhancements
  - Improves user experience and resiliency to fabric errors
  - Initial installation now allows IPoIB instances to joint appropriate partition without intervention
  - Faster detection and response to failures of subnet manager
  - Memory registration and speed have been improved
  - Combining this enhancement with ability to double MTU size on IB fabric, significantly improves usability

ZFS

- ZFS device replacement enhancements
  - We now provide a *sysevent* when an underlying device is expanded
  - *autoexpand* property determines pool behaviour when a LUN is expanded (disabled by default)
    - No export / import of pool required; no reboots, too!
    - *zpool online -e* expands the pool to full size of LUN
- Enhancement to *zpool list*
  - Provides better space allocation information
  - Replacing USED and AVAIL fields with ALLOC and FREE
  - ALLOC = physical space allocated to all datasets and metadata
  - FREE = unallocated disk space in the pool

```
# zpool list tank
NAME    SIZE  ALLOC   FREE    CAP  HEALTH  ALTROOT
tank    136G  55.2G  80.8G    40%  ONLINE  -
```

ZFS (2)

- ZFS snapshots retention policies
  - Useful if you allow `zfs receive` to inadvertently destroy older snapshots because they no longer exist on the sending side
  - You can now hold a snapshot to prevent its destruction
  - A snapshot with clones can now be deleted pending removal of the last clone, using `zfs destroy -d`
  - Use `zfs hold` to hold a snapshot (or a set of snapshots)
- Triple parity RAID-Z (raidz3)
  - Starting now, the redundant RAID-Z config may have either single-, double-, or triple-parity, allowing for one, two or three devices to fail *without any data loss*
  - When creating a pool, use `raidz3` as an argument

ZFS (3)

- ZFS log device enhancements
  - Use `logbias` property to tell ZFS how to handle synchronous requests for a specific dataset
    - If `logbias` is set to `latency` (default) ZFS uses the pool's separate log devices, if any, to handle the requests at low latency
    - If `logbias` is set to `throughput` ZFS does not use pool's separate log devices; instead optimizes synchronous operations for global pool throughput and for the efficient use of resources; `logbias=throughput` is known to improve DB write performance

ZFS (4)

- ZFS log device enhancements
  - Log device removal
    - `zpool remove` allows you to remove a log device from pool
    - A mirrored log devices is removed by specifying the top-level mirror for the log device
    - Removing a separate log device causes ZIL transaction records to be written to the main pool
    - Redundant top-level virtual devices now have a numeric identifier
      - In a mirrored storage pool of 2 disks, the top level virtual device is `mirror-0`

ZFS (5)

- ZFS storage pool recovery
  - If underlying devices become unavailable due to power failure or more then the maximum allowable number of devices fail (in a redundant ZFS configuration), the pool becomes damaged
  - `zpool clear` and `zpool import` allow you to attempt to recover a damaged pool via the `-F` flag
  - `zpool status`, `zpool clear` and `zpool import` automatically report a damaged pool and describe how to recover it
- New ZFS system process per zpool
  - Each pool now has an associated process: `zpool -poolname`
  - Threads in this process are the pool's I/O processing threads used to handle I/O tasks (ex: compression, checksum validation)
  - Gives observability of pool's CPU utilization, with traditional `ps` and `prstat` interfaces; see SDC(7) man page for details
  - Only in the global zone

ZFS (6)

- Splitting a mirrored ZFS storage pool
  - `zpool split` allows you to split a mirrored storage pool
  - Detaches a disk (or disks) in the original mirrored pool to create another identical pool

New Unicode Locales added

af_ZA.UTF-8 (South Africa)

en_SG.UTF-8 (English, Singapore) zh_SG.UTF-8 (Chinese, Singapore)

ms_MY.UTF-8 (Malaysia)

id_ID.UTF-8 (Indonesia)

bn_IN.UTF-8 (Bengali, India) en_IN.UTF-8 (English, India)

gu_IN.UTF-8 (Gujarati, India) kn_IN.UTF-8 (Kannada, India)

mr_IN.UTF-8 (Marathi, India) te_IN.UTF-8 (Telugu, India)

ta_IN.UTF-8 (Tamil, India)

Additionally ...

- Sparse File Support for `cpio`
  - As used by Live Upgrade tools, that previously filled holes
- `sendmail` has new properties
  - Allow for automatic rebuilding of `sendmail.cf` and `submit.mc` files
  - Now split into 2 instances for easier management of the traditional daemon and the client queue runner
- New privilege in the Basic privilege set
  - `net_access` allows processes to create a network endpoint
  - By removing this privilege, an administrator restricts network access and IPC
  - See the privileges(5) man page

Additionally (2)

- Generic FMA quality improvements (snv_121)
- Historic Analysis Enhancements (snv_125)
  - Improves managing cost of service
- FireFox 3.5 bundled (snv_119)
- Java 6 Update 20 (rebranding + security)
- Ghostscript upgraded to 8.64 (snv_124)
- Adobe Flash 10.0.xx security updates (snv_121)
- Gtar 1.22 (snv_116)
- BIND 9.6.1
- SunVTS 7.0ps9 (snv_137)
- Cacao 2.2.4.1
- Less v.436 (v.381 is pretty old)


Sun
ORACLE®


ORACLE®