



Solaris 10 Zones (AKA “N1 Grid Containers”, NKA “Solaris Containers”)

Harry J. Foxwell, Ph.D.
Senior System Engineer
Sun Microsystems



Related Technologies

- Sun Enterprise Server Domains (HW)
- IBM mainframe LPAR
- IBM AIX WorkLoad Manager
- HP vPar (virtual partition)
- HP PRM (Process Resource Manager)
- VMWare
- Linux
 - <http://user-mode-linux.sourceforge.net/>
 - <http://sourceforge.net/projects/xen>
 - <http://www.linux-vserver.org/>

Resources

- `www.sun.com/solaris/10`
- `http://www.sun.com/bigadmin/content/zones/`
- `http://www.blastwave.org/docs/Solaris-10-b51/DMC-0002/dmc-0002.html`

Zones can be used for Server Consolidation

- Run multiple applications securely and in isolation on the same system
- Utilize the hardware resources more effectively
- Allow delegated administration of the application environment
- Streamline the effort in maintaining the system

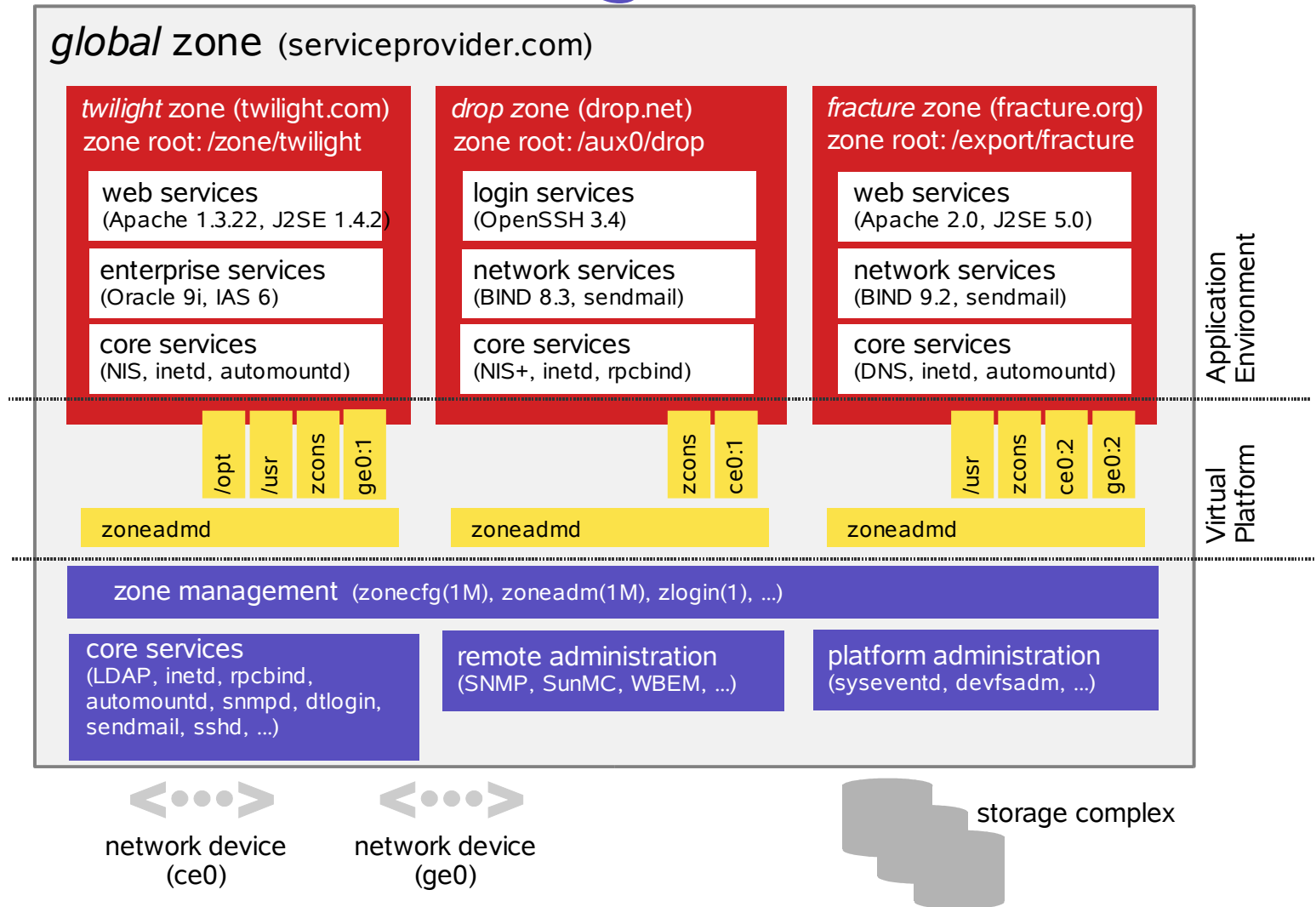
Zones Summary

- Isolated application environments within a **single Solaris instance**
- Resource, name space, security and failure **isolation**
- Efficient and granular using a lightweight OS layer
- Delegated, simplified administration
- No porting as ABI/APIs are the same

Typical Uses for Zones

- Consolidating data center workloads such as multiple databases
- Hosting **untrusted** or **hostile** applications or those that require global resources like IP port space
- Hosting “complete” environments
- Deploying Internet facing services
- Software development

Zones Block Diagram



Zone Administration

- zoneadm (1M) is used by the global zone administrator to
 - **install** a new root file system for a configured zone
 - **list** zones and optionally their state
 - **verify** whether the configuration of an installed zone is semantically complete and ready to be booted
 - **boot** or **ready** an installed zone
 - **halt** or **reboot** a running zone
 - **uninstall** the root file system of an installed zone

Primary Zone States

- *Configured*: Configuration completely specified and committed to stable storage
- *Installed*: Packages have been installed under the zone's root file system
- *Ready*: Virtual platform has been established
- *Running*: User processes are executing in the zone application environment

Zone Console

- Zone pseudo-console available for each zone
 - Mimics a hardware console
 - Accessible via `zlogin -C`
 - Available prior to zone boot

```
global# zlogin -C zone1
[Connected to zone 'zone1' console]
twilight#
~.
[Connection to zone 'zone1' console closed]
```
- Publishes zone state change messages
[Notice: zone halted]

Security

- Each zone has a security boundary around it
- Runs with subset of privileges (5)
- A compromised zone is unable to escalate its privileges
- Important name spaces are isolated
- Processes running in a zone are unable to affect activity in other zones

Security in a Zone (2)

- Global zone root user is traditional root
- Activity is restricted inside a non-global zone at the system call boundary
 - Safe: `chmod(2)`, `chroot(2)`, `chown(2)` and `setuid(2)`
 - Unsafe: `memcntl(2)`, `mknod(2)`, `stime(2)`
 - Some calls, such as `kill(2)` are limited in scope
- Other restricted operations
 - Loading and unloading of kernel modules
 - Plumbing and modifying network interfaces

Process Model in a Zone

- Process namespace is partitioned
 - Processes may not see or interact with processes in other zones. Processes in other zones appear not to exist.
 - Processes running in the global zone can see all processes.
 - Processes in the same zone interact as usual.
 - `proc(4)` only provides information about processes in the zone.
 - Process tree is rooted by `zsched` rather than `init`

File Systems in a Zone

- Virtualized view of the file system namespace
- The zonepath is part of the configuration
- The root of the zone is located at `$zonepath/root`
- Restricted access to `$zonepath`
- Per-zone mount table:
 - Mounts from global zone into zone
 - Mounts from within zone limited by what is accessible

/dev Inside Zones

- No /devices in a zone
- /dev is constructed at zone boot at `$zonepath/dev`
- Loopback-mounted into the zone at `$zonepath/root/dev`
- /dev heavily restricted
 - `chmod(2)`, `chown(2)` and `chgrp(1)` are permitted
 - `link(2)`, `unlink(2)`, `symlink(2)`, `mknod(2)`, `creat(2)` and `rename(2)` are not allowed

Zone Commands

- Zone Configuration – `zonecfg`
 - Define what a zone looks like
- Console Access – `zlogin -C`
- Zone Administration – `zoneadm`
 - Install, Boot, Restart, Stop, List, Verify, Uninstall

Configuration/Administration

- `zonecfg` (1M) is used to specify resources (such as IP interfaces) and properties (such as a resource pool)
- `zoneadm` (1M) is used to perform administrative steps for a zone such as list, install, (re)boot, halt, et cetera
- Installation creates a root file system with factory-default editable files

zonecfg (1M) **Resources**

- `fs`: file system
- `inherit-pkg-dir`: directory which should have its associated packages “inherited” from the global zone
- `net`: network interface
- `device`: device
- `rctl`: resource control
- `attr`: generic attribute

Additional Features

- Support for read-only `lofs` (7FS)
- Configuration stored in a private XML file
- Zone ids are dynamically assigned at zone boot
- `ptree(1)` can displays a zone's process tree
- `traceroute(1M)` supported inside a zone
- `zonecfg(1M)`
 - `autoboot` property specifies action at global boot

- - NFSv4 client support
 - `nfsstat (1M)` virtualized per-zone
 - `ps (1)` can display processes from a list of zones or add a ZONE column to other reports
 - Support for `-p` option to `prtconf (1M)`

- CPU visibility

- Only take effect when resource pools are enabled
- Traditional commands and APIs that deal with processors will provide a “virtualized” view based on the pool (processor set) the zone is bound to
 - Including `iostat(1M)`, `mpstat(1M)`, `prstat(1M)`, `psrinfo(1M)`, `sar(1)` and `vmstat(1M)`
 - Including `sysconf(3C)` (when detecting number of processors configured/online) and `getloadavg(3C)`
 - Including numerous `kstat(3KSTAT)` values from the `cpu`, `cpu_info` and `cpu_stat` publishers

- `zones.max-lwps`
 - zone resource control
 - This resource control can be further subdivided within the zone itself using `project.max-lwps`
- Zone-aware auditing
 - Global zone administrator can specify whether auditing should be global or per-zone
 - If per-zone, each zone administrator can configure and process their audit trails independently

- - Support for `-l` and `-s` options to `swap` (1M)
 - Zones can be booted in single-user mode
 - Support for `sysdef` (1M) from within a zone
 - Zones where no `inherit-pkg-dir` resources have been defined are supported

Discussion

- How/Why would you use server virtualization technologies?
- Advantages?
- Disadvantages?



Zones



Harry.Foxwell@Sun.COM



A Multi-Platform OS Strategy



- Interoperate 'out of the box'
- Java support on Windows PCs
- Windows certification for Sun hardware



- Red Hat, SuSE
- 32- and 64-bit
- Sun and 3rd party hardware
- Complements Solaris
- Open Source



- SPARC and x86
- 32- and 64-bit
- Sun and 3rd party hardware
- Run Linux apps unchanged
- OpenSolaris