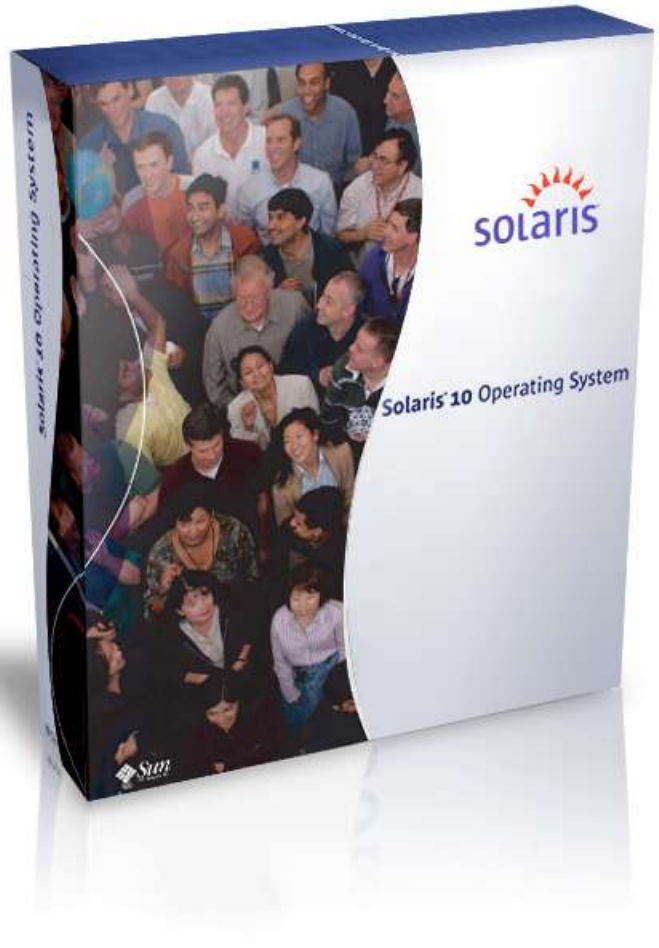
I'm zany for zones!

Linda Kateley
Solaris 10 Adoption Specialist
linda.kateley@sun.com

Agenda

- Zone Basics
- Zones/Containers Admin
  - Filesystem
  - Patching
  - migration
- Next generation SCLA
- Next generation Xen

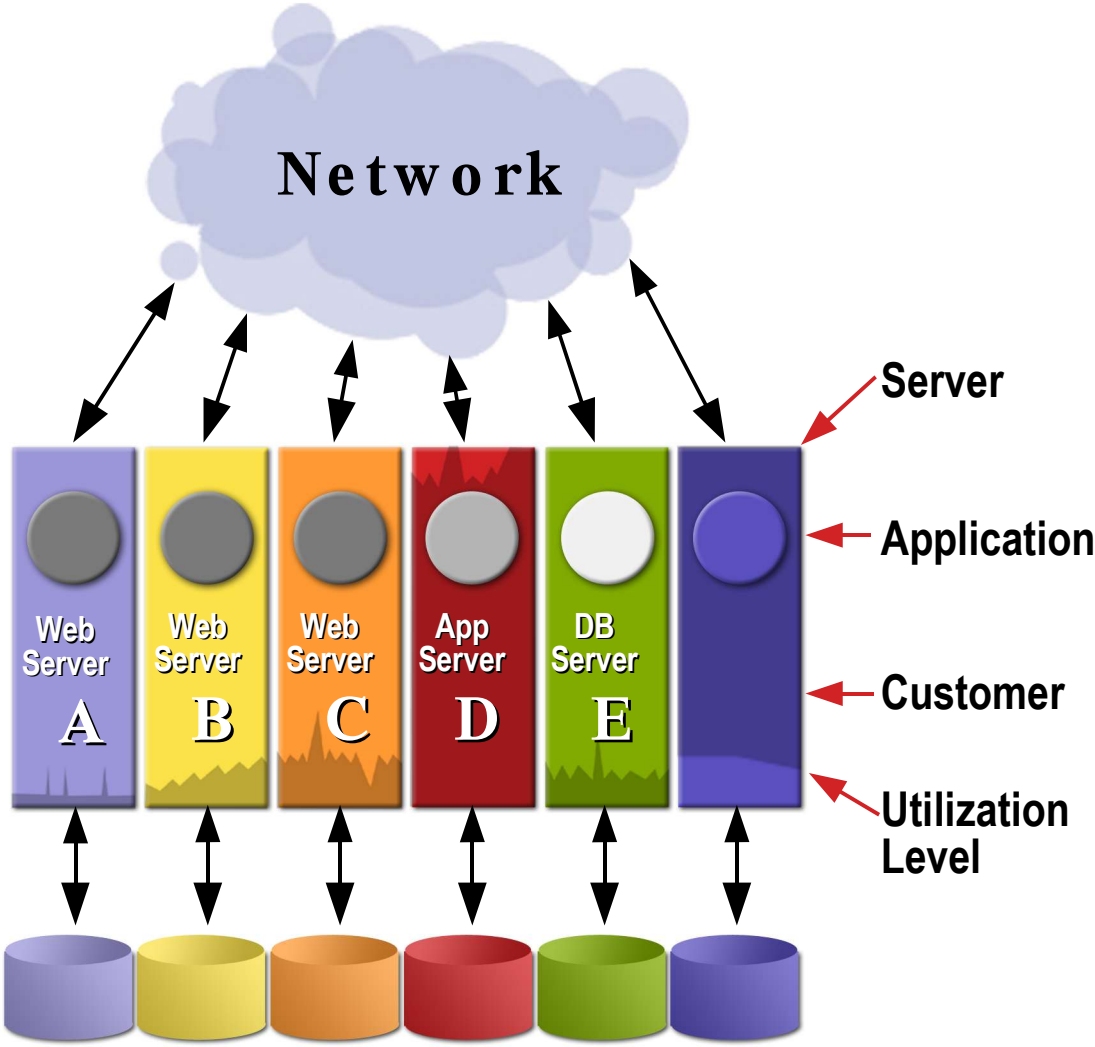


Solaris 10

- Dynamic Tracing (DTrace)
- Solaris Containers
- Predictive Self-Healing
- ZFS
- Secure Execution
- Integrated SAN Support
- Compatibility Guarantee

Traditional Resource Management

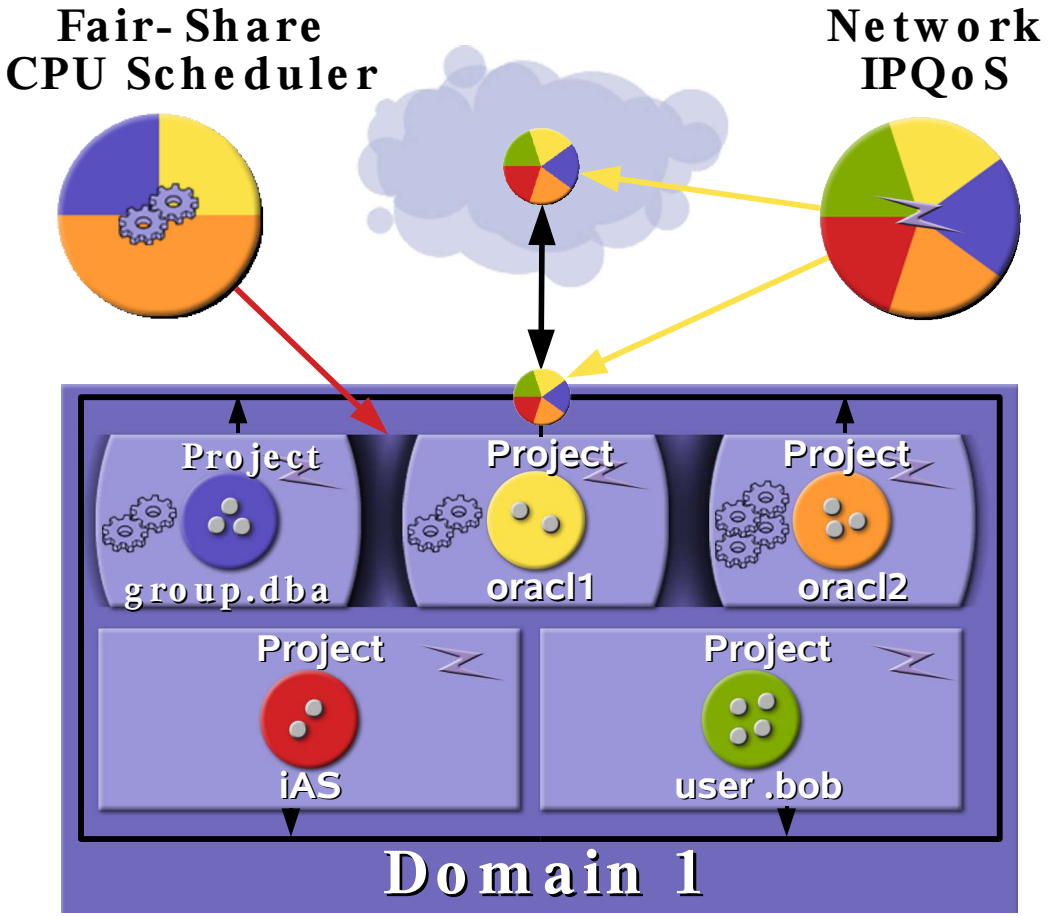
- One application per server
- Size every server for the peak
- Avg. utilization rate is 20%–30%



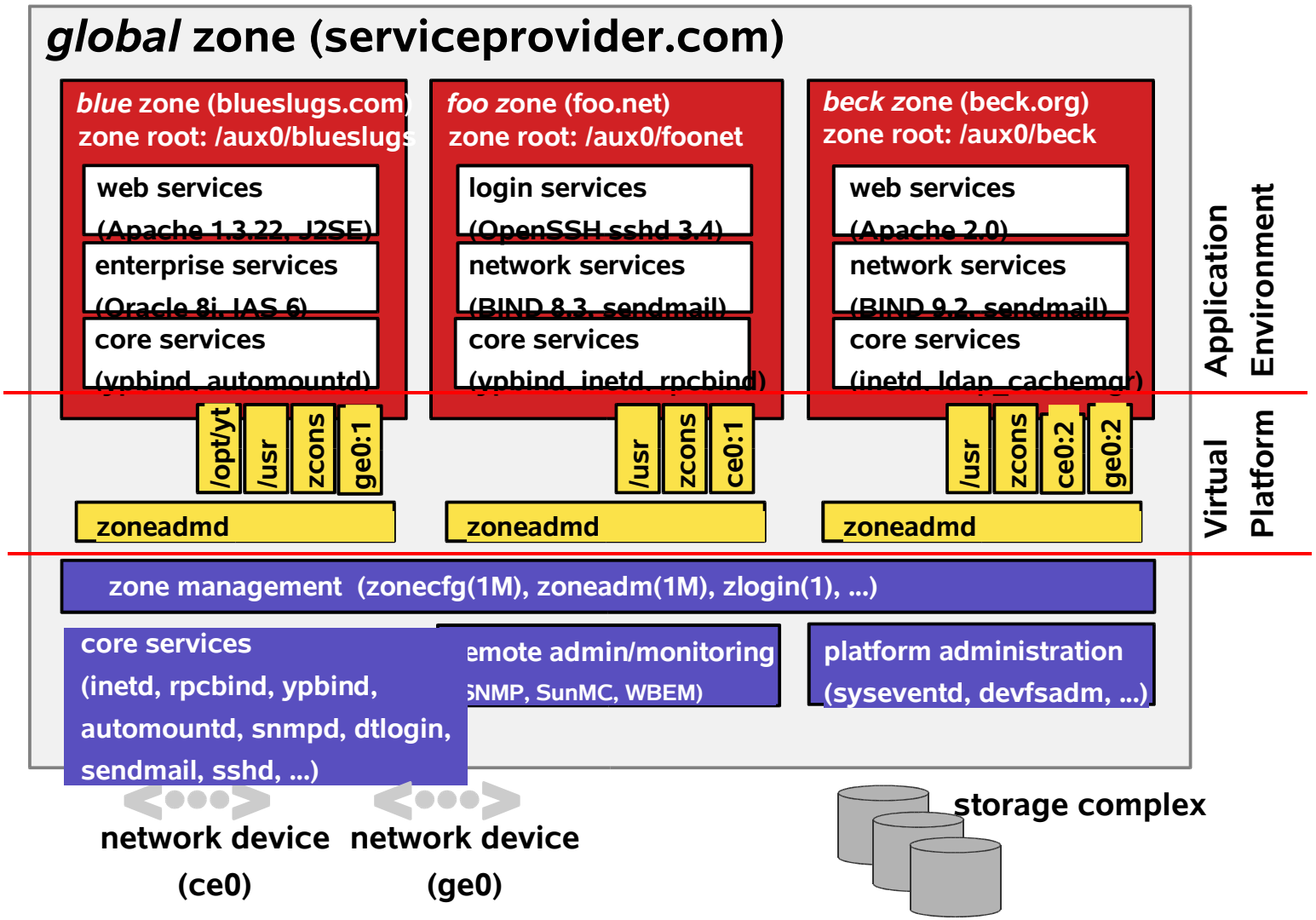
Solaris Container

Resource Management

- Workload Metering
- Sub-CPU Partitioning
- Control CPU, Memory, and Network



Zones Block Diagram



Creating a zone

```
global# zonecfg -z zone1
zone1: No such zone configured
Use 'create' to begin configuring a
  new zone.
zonecfg:zone1> create
```

Setting's for the zone

```
zonecfg:zone1> set zonepath=/zoneroots/zone1
zonecfg:zone1> set autoboot=true
zonecfg:zone1> add net
zonecfg:zone1:net> set address=192.9.200.67
zonecfg:zone1:net> set physical=hme0
zonecfg:zone1:net> end
zonecfg:zone1> ^D

#zoneadm list -c
```

Installing the zone

```
global# zoneadm -z zone1 install
Constructing zone at /zoneroot/zone1/root
Creating dev directories
Creating dev links
Copying packages and creating contents file
Copying files and directories
Setting up /etc/motd
Setting up /etc/inittab
Setting up /etc/vfstab
Setting up /var/yp/aliases
Configuring files
```

boot the zone

global# zoneadm -z zone1 boot

- Took about .6 seconds on ferrari

- global# zlogin -C zone1
- [Connected to zone 'mydesktop' console]
- <Run through sysid tools as usual to do initial customization>

Solaris 10 Containers

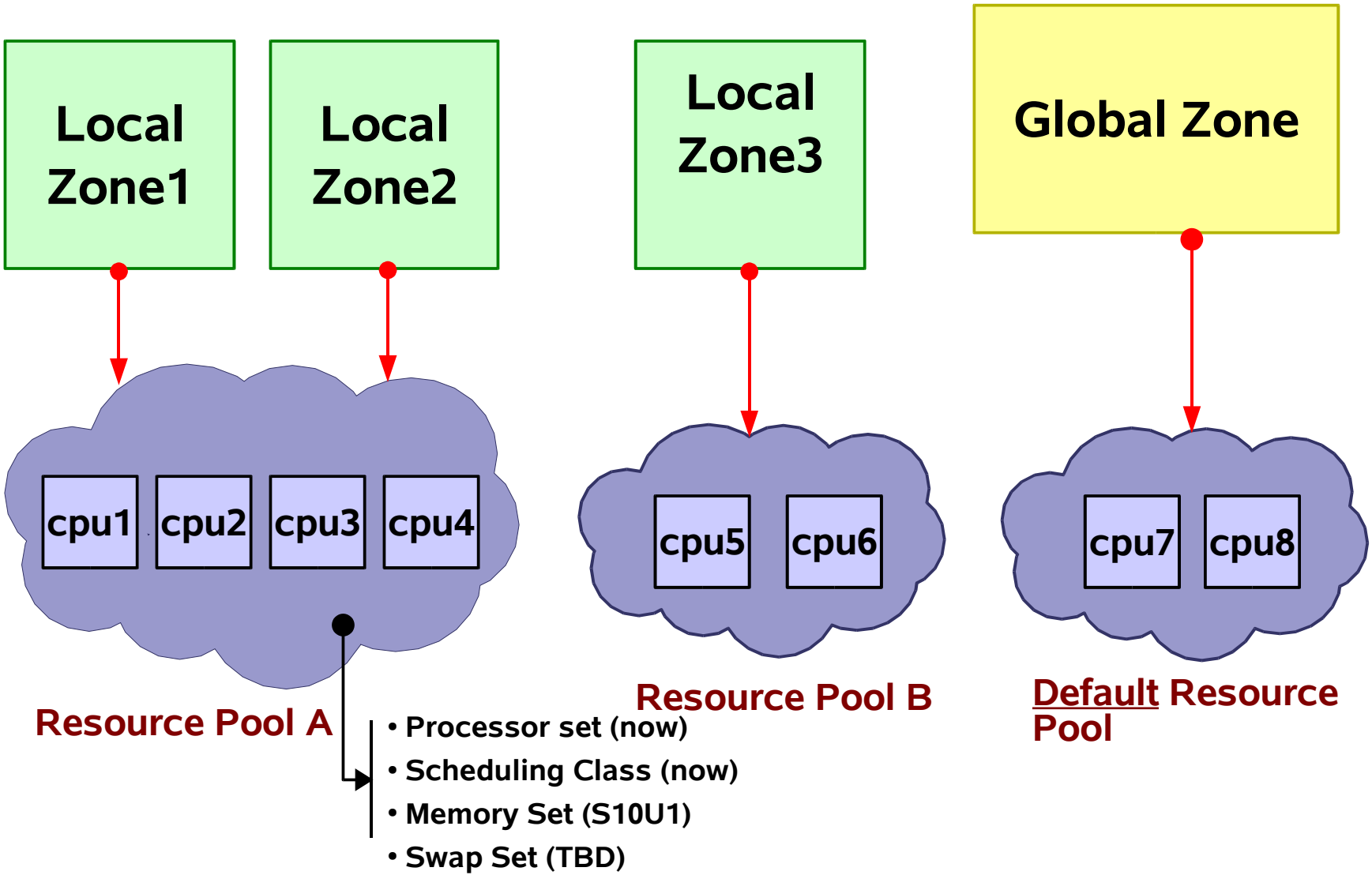
demo

Solaris 10 Containers

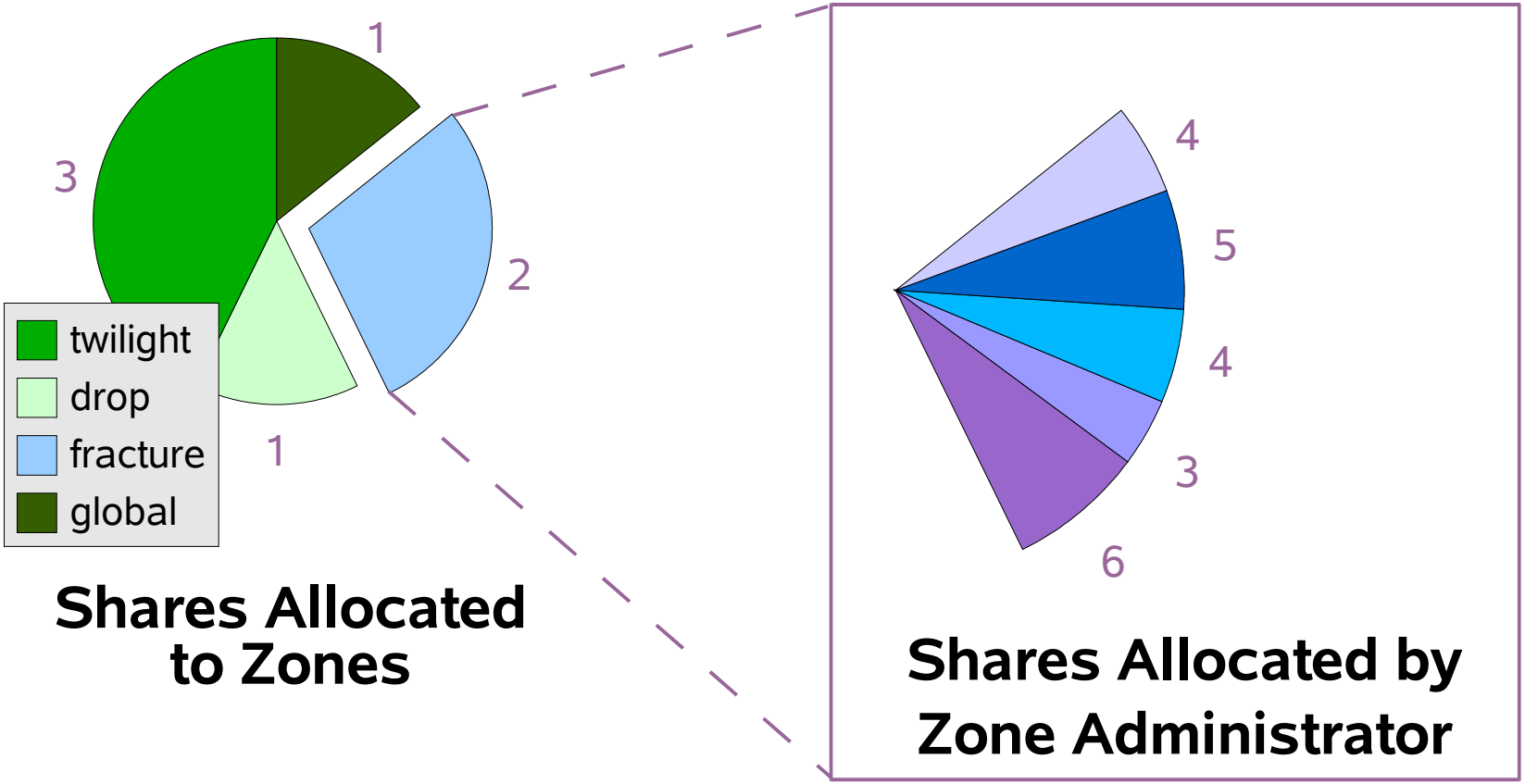
Administrating zones

- resource management
- pools
- patching
- files
- backup

Zones and Resource Pools



Two Level FSS



FSS-TS-IA

Controlling CPU Consumption

The Fair Share Scheduler can be used to control CPU consumption of the instances.

The Fair Share Scheduler is not the default scheduler and must be enabled using the dispadmin(1M)command:

```
# dispadmin -d FSS
```

Projects

Command	Description
projadd(1M)	adds a new project to the local project database
projmod(1M)	modifies a project entry in the local project database
projdel(1M)	deletes a project entry from the local project database
projects(1)	displays project membership for a user
newtask(1)	switches to a project

Projects

/etc/project

projname:projid:comment:user-list:group-list:attributes

/etc/project contains five standard projects:

system, user.root, noproject, group.staff, default

The system project is used for all system processes and daemons.

All of roots processes run in the user.root project.

The noproject project is a special for IPQoS.

The group.staff project will be used for all users in the group staff

The default project serves as a catch-all and will be used for users not matching any of the other projects.

Projects

/etc/project

projname:projid:comment:user-list:group-list:attributes

#projadd

-U user,user

-G group,group

-c comment or description

-K value=attributes

-p unique project number (if not given will give next available

name

#projects -l

Projects

Admin commands

#projects -l will show all defined projects

#id -p – will show users project

#newtask -p project exec – allows us to execute in a project

#prstat -J – show per project consumption

#prstat -T – show per task consumption

Projects
cpu control- priv

#projmod -K “project.cpu-shares=(priv,value,action)” project

Privilege level determines who can modify

There are three privilege levels:

basic -the owner of the calling process

privileged -only privileged (superuser)users can change

system -the threshold is fixed for the lifetime of the operating system instance

Projects
cpu control- value

#projmod -K “project.cpu-shares=(priv,value,action)” project

CPU Shares Configuration

Every project can be assigned a project.cpu-shares resource control.Projects that do not have this resource control are assigned 1 share by the system.

Shares are numeric values

Shares are not percent

projecta 50, projectb 50 is the same as projecta 200, projectb 200

Projects
cpu control-action

#projmod -K “project.cpu-shares=(priv,value,action)” project

The action defines the action to be taken when the threshold is exceeded.

There are three possible actions:

deny -this denies resource requests for an amount that is greater than the threshold

signal -this sends the specified signal to the process exceeding the threshold value.

none -this causes no action when the threshold is exceeded

Projects

Available Resource Controls

Resource Control	Description
process.max-port-events	maximum allowable number of events per event port
process.crypto-buffer -limit	maximum number of bytes allocated for copying
process.max-crypto-sessions	maximum number of entries in the session table
process.add-crypto-sessions	number of entries added when enlarging the session table
process.min-crypto-sessions	minimum number of entries in the session table
process.max-msg-messages	maximum number of messages on a message queue
process.max-msg-qbytes	maximum number of bytes of messages on a message queue
process.max-sem-ops	maximum number of semaphore operations per semop call
process.max-sem-nsems	maximum number of semaphores per semaphore set
process.max-address-space	maximum size of the address space in bytes
process.max-file-descriptor	maximum index in filedescriptor table
process.max-core-size	maximum core file size in bytes
process.max-stack-size	maximum size of the stack segment in bytes
process.max-data-size	maximum size of the data segment in bytes
process.max-file-size	maximum file size in bytes

Projects

Available Resource Controls

Resource Control	Description -cont
process.max-cpu-time	maximum CPU time in seconds
task.max-cpu-time	maximum CPU time in seconds
task.max-lwps	maximum number of simultaneously available LWPs
project.max-port-ids	maximum allowable number of event ports
project.max-shm-memory	maximum size of System V shared memory in bytes
project.max-shm-ids	maximum number of System V shared memory segments
project.max-msg-ids	maximum number of System V message queues
project.max-sem-ids	maximum number of System V semaphores
project.cpu-shares	the number of CPU shares
zones.cpu-shares	number of CPU shares per zone

Projects
cpu control

You can also control cpu shares dynamically with

prctl(1M) get or set resource controls on a running process,task or project

rctladm(1M)display or modify global state of system resource controls

prctl -n project.cpu-shares -r -v # -i project projname

-n name of value

-r replace

-v new value

-i project, task, process

Configuring per zone shares

```
#dispadmin -d FSS
#reboot

#zonecfg -z name
zonecfg:zone1> add rctl
zonecfg:zone1:rctl> set name=zone.cpu-shares
zonecfg:zone1:rctl> add value
    (priv=privileged,limit=10,action=none)
zonecfg:zone1:rctl> end
zonecfg:zone1> verify
zonecfg:zone1> commit
zonecfg:zone1> ^D

#prctl -n zone.cpu-shares -r -v 25 -i zone zonename
```

Solaris 10 Containers

Rm demo

Pools

Since solaris 2.6 we have had psrset.

The syntax looked like

```
#psrset -a name cpu0 cpu1
```

We could then bind a process to the set using

```
#pbind pid name
```

When the cpu was idle nothing else could use it

Pools

Enter pools

We can set a min and max number of cpu's in a pool which one or more processes, projects or task can be assigned to.

The controlling daemon is the poold which will start at boot with the existence of a

/etc/pooladm.conf file

Pools- config

Enabling pools

#pooladm -e

Disabling pool

#pooladm -d

remember that pools will be enabled at boot with the existence of the file.

Pools- config

Creating the file

```
#pooladm -s
```

This will create an xml /etc/pooladm.conf file which is best viewed with

```
#poolcfg -c info
```

Which says give me info about the currect config.

Pools- config

Modifing the config- first create the set

```
# poolcfg -c 'create pset linda (uint pset.min =
  2; uint pset.max = 10)'
```

Then create a pool

```
# poolcfg -c 'create pool kateley'
```

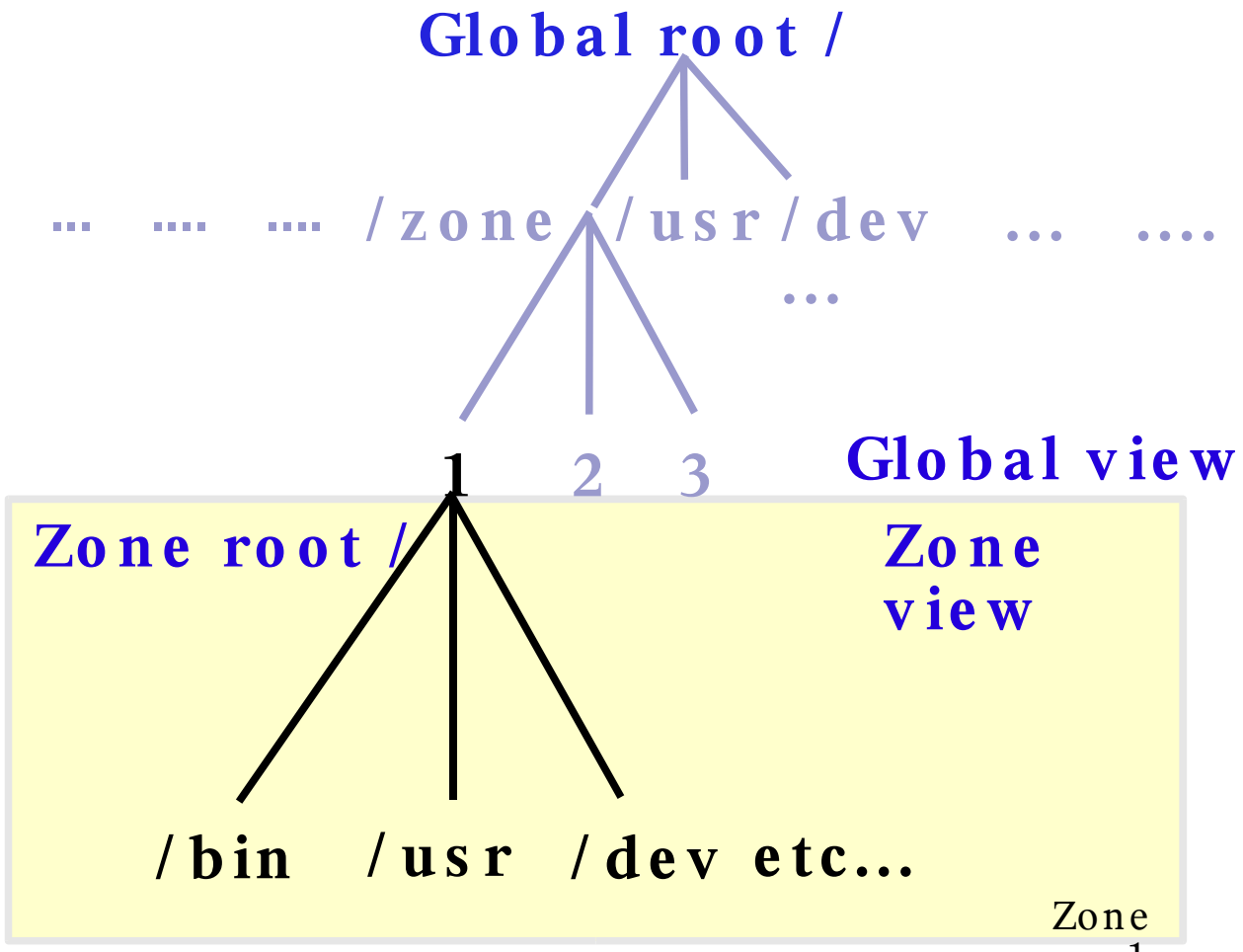
Connect the set to the pool

```
# poolcfg -c 'associate pool kateley (pset
  linda)'
```

Zone Pools

- Pools
- Zones may be bound to pools
  - Automatically via zone configuration
  - #poolbind(1M) -p poolname -i zoneid zonename
- All processes in zone bound to same pool

Zone File Systems



Solaris 10 Containers

File Systems

- Sparse-root vs. whole-root
- Read-write vs. read-only
- File access vs. device access
- Backups

Solaris 10 Containers

File System Creation – Direct Mount

- RW or RO access in LZ and GZ
- Easily accessible from GZ (by root)
- Can be unmounted and remounted by GZ (if not used)
- Simplest method
- Method:

```
global# mount /dev/dsk/c1t0d0s6 /export/zones/zone1/opt/local
global#mount -F lofs /dir /export/zones/zone1/dir
```

Solaris 10 Containers

File System Creation - lofs

- Can mount in multiple zones
- dir= is mount point in zone, special=name of dir to mount
- Method:

```
global# zonecfg -z zone1
      add fs
            set dir=/opt/local
            set special=/export/opt/local
            set type=lofs
            end
      exit
global# zoneadm -z zone1 boot
```

Solaris 10 Containers

File System Creation – UFS Mount

- After LZ boots, GZ can unmount and re-mount
- Method:

```
global# newfs /dev/dsk/c1t0d0s6
global# zonecfg -z zone1
      add fs
            set dir=/opt/local
            set special=/dev/dsk/c1t0d0s6
            set raw=/dev/rdsk/c1t0d0s6
            set type=ufs
            add options [ro,nodevices]
            end
      exit
global# zoneadm -z zone1 boot
```

Solaris 10 Containers

File System Creation – device in zone

- Method:

```
global# zonecfg -z zone1
        add device
                set match=/dev/dsk/c1t0d0s
        exit
global# zoneadm -z zone1 boot
```

Solaris 10 Containers

Whole root zone

- Can only be done before zone install
- Method:

```
global# zonecfg -z zone1
      remove inherit-pkg-dir dir=/usr
      remove inherit-pkg-dir dir=/lib
      remove inherit-pkg-dir dir=/platform
      remove inherit-pkg-dir dir=/sbin
exit
  global# zoneadm -z zone1 boot
```

Solaris 10 Containers

Creating

- Can be created using a script or template
- Method:

```
global# zonecfg -z zone1
        create -t zone
```

```
global#zonecfg -z zone1
        export -f filename
global#zonecfg -z zone2
        create -f filename
```

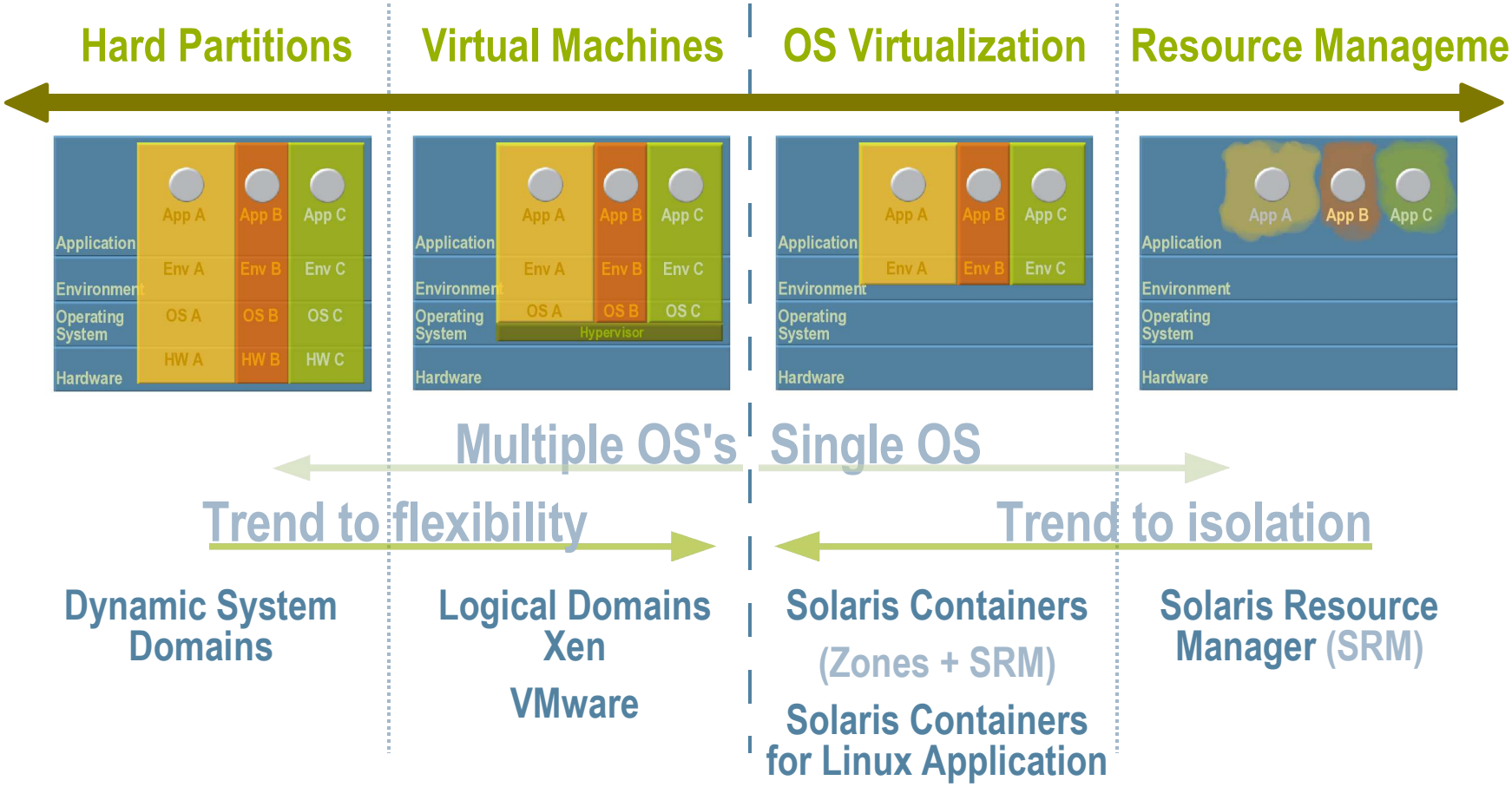
Solaris 10 Containers

Info

- http://www.opensolaris.org/os/community/zones/faq/

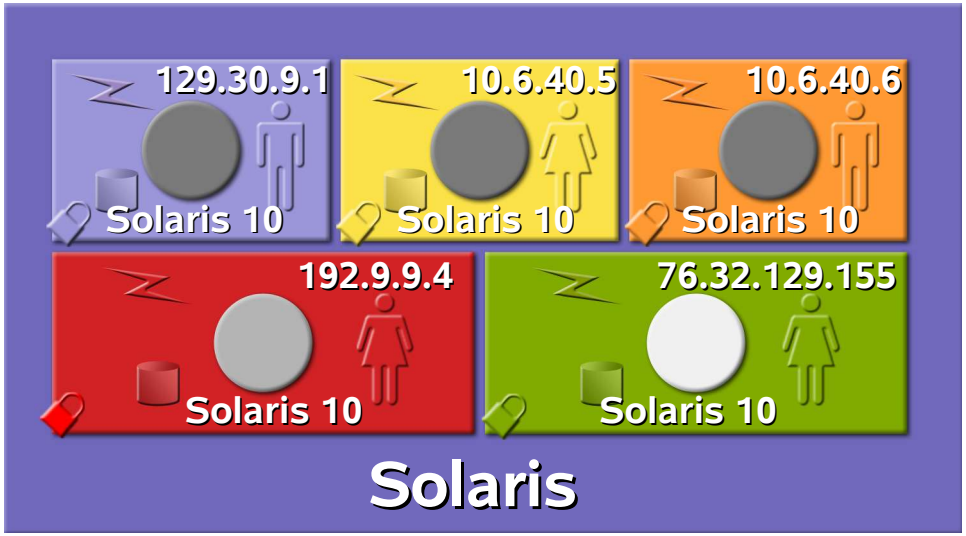
Solutions from Sun

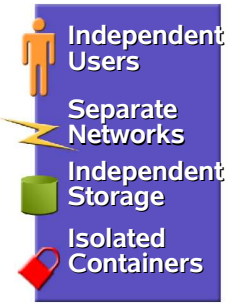
- It's all about **Customer Choice**



Extending Solaris Containers

Today ...



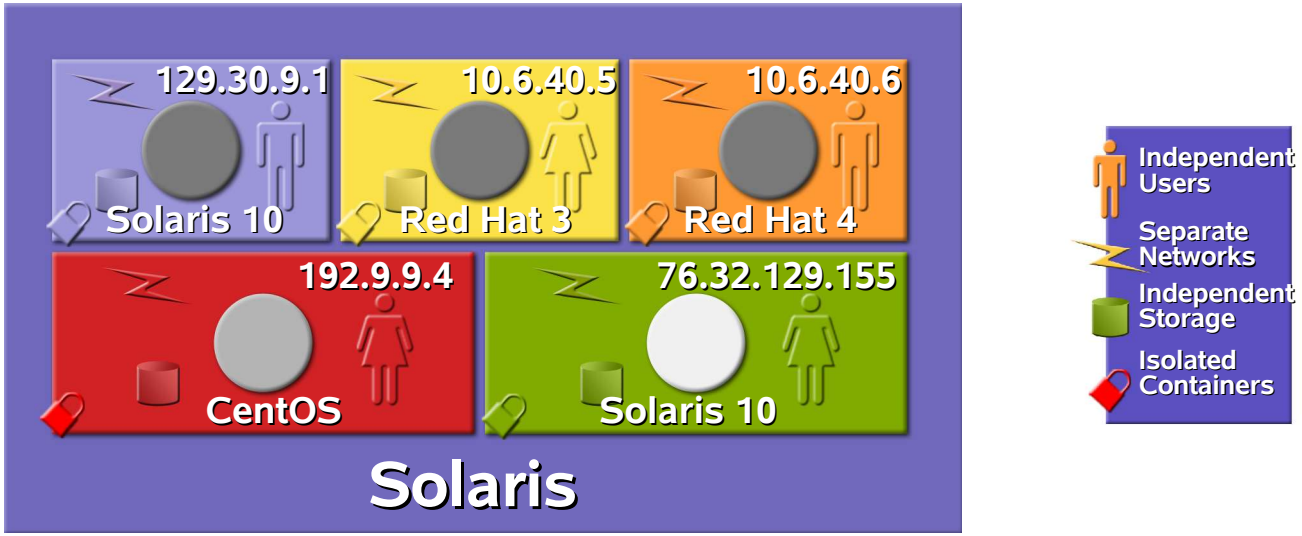


Single Kernel

Single Operating System

Extending Solaris Containers

... Tomorrow ...



Single Kernel

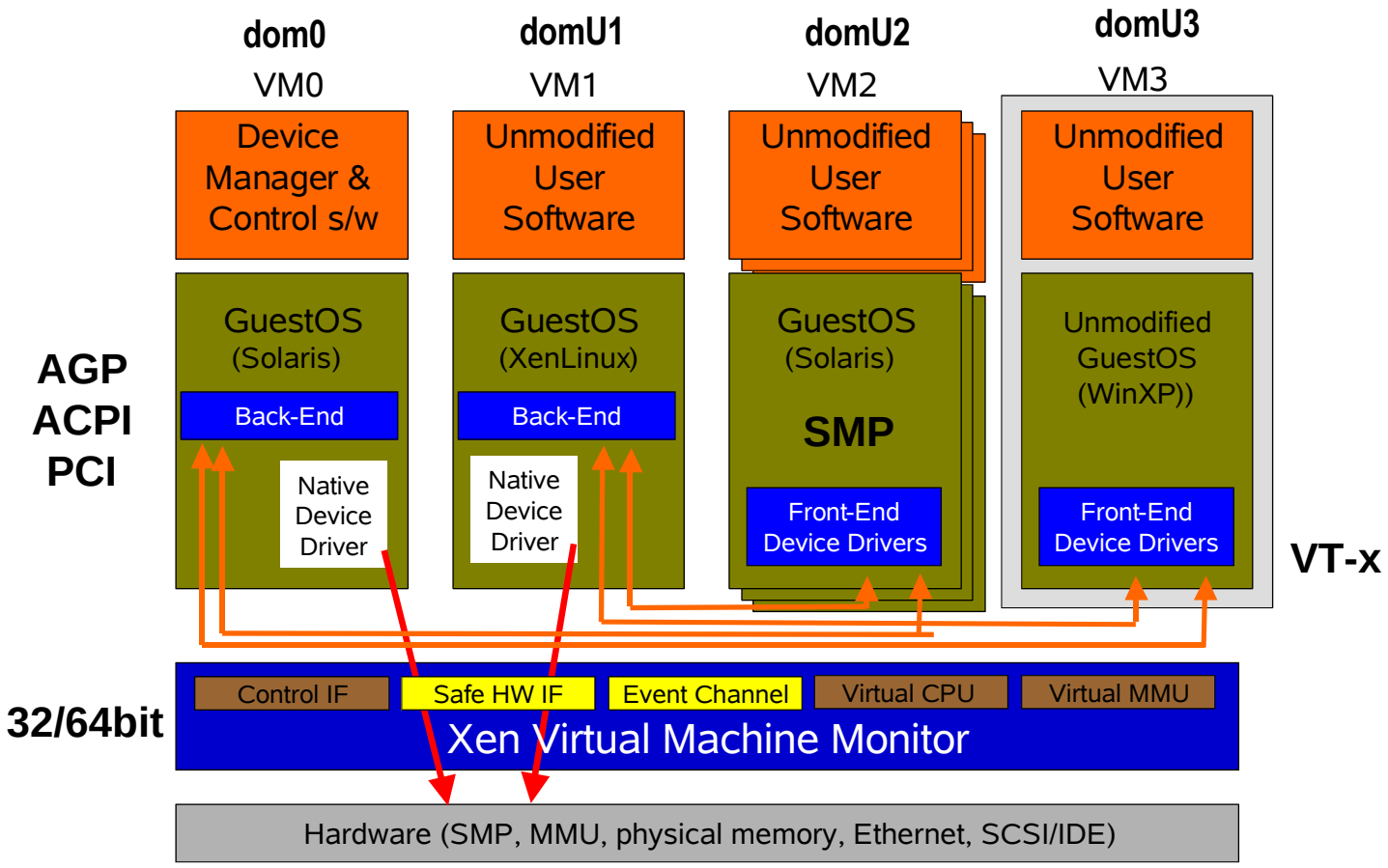
Multiple Operating Environments

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"

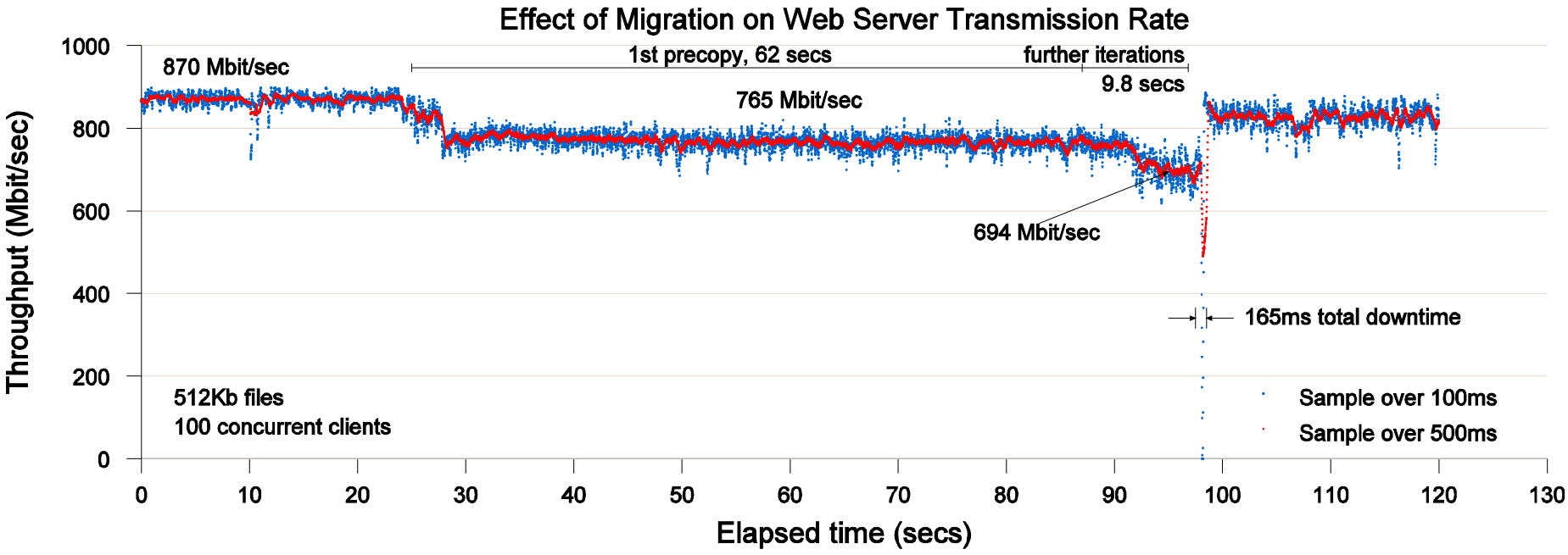
Xen 3.x Architecture



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

SPECweb99 Migration Experiment

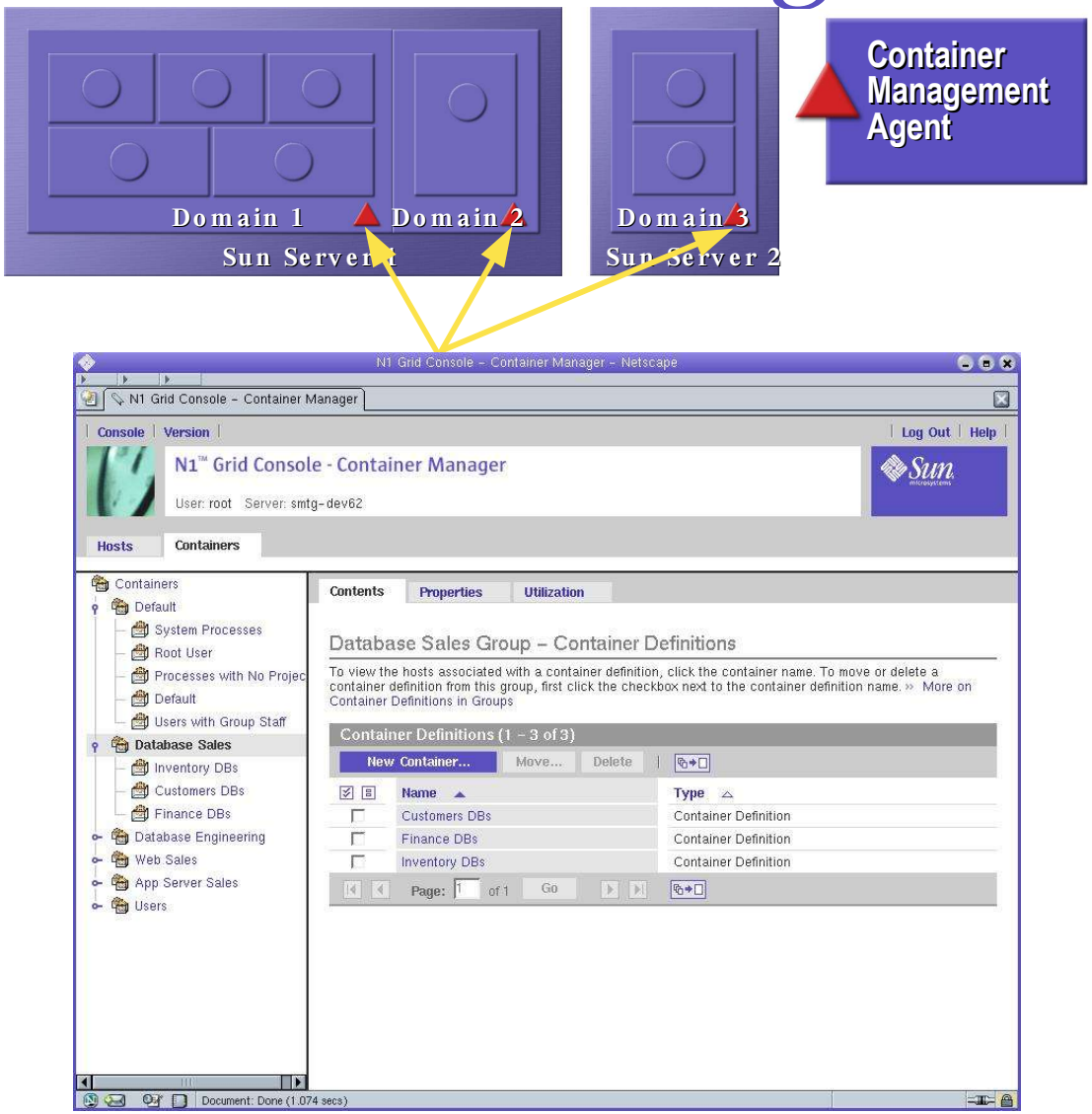


From LinuxWorld 2005 Virtualization BoF

Solaris Container Console

- Browser based GUI to manage Containers
- Controls resource management on Solaris 8 OS and Solaris 9 OS
- Controls Zones on Solaris 10
- Uses the Sun MC 3.5 Update 1 infrastructure

Container Management



Solaris Container Console

Features & Benefits

- Container Management
  - Create/Delete/Modify Containers
- Centralized Management of Multiple Systems
  - Manage all the Containers across the network
- Container Replication
  - Recreation a Container on a separate system
- Container and Process Monitoring
  - Zoom into a Container to verify its contents

Webmin

About Webmin

- A web-based interface for UNIX system administration
- It comes with Solaris 10, or get it at http://www.webmin.com
-

linda.kateley@sun.com

February 2006