

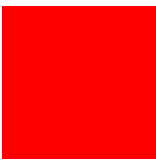
ORACLE®



**ORACLE®**

## **Benefit from Oracle Solaris Zones**

Duncan Hardie  
Oracle Solaris Product Management



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.

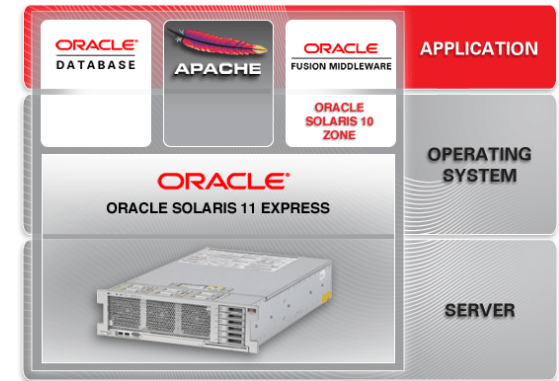
# Agenda

- Who are you, what do you want?
- Using Oracle Containers Today
  - A quick history lesson
  - Walking the walk
- What's new in Solaris 11 Express
  - New and improved!

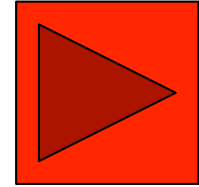


# Zones: Feature Evolution since S10

- Up to 33% adoption in production
- Improvements in all areas
  - Core:
    - *configure privileges, move, rename, attach, detach, update on attach, DTrace in a zone, boot args*
  - Packaging:
    - *live upgrade, parallel patching, turbo packaging*
  - ZFS:
    - *assign ZFS datasets to zones, clones and snapshot*
  - Networking:
    - *stack instances, default router*
  - Resource Management:
    - *simplify, CPU Caps, observability*
  - Brands and Ecosystem:
    - *S8C, S9C, TX, Cluster, Linux, OEM Ops Center integration*



## Hope you didn't miss




Show Terry's  
P2V Demo

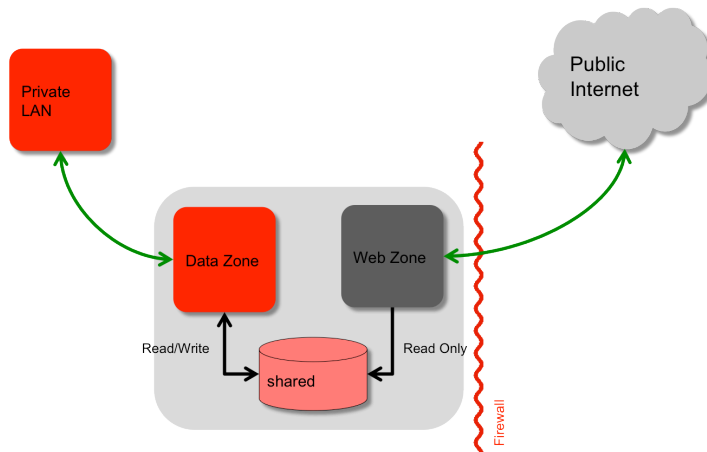
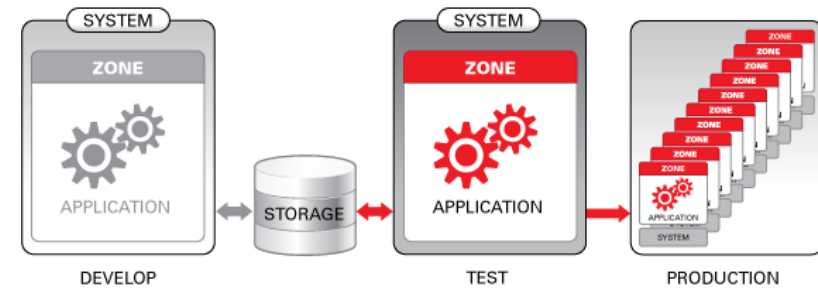
- Solaris 10 Physical to Virtual (P2V)
  - Provides consolidation capability
    - Also in Solaris 8 Containers, Solaris 9 Containers
  - Create a system image, transfer and install the zone
  - flar, cpio, pax xustar, ufsdump, directory
  - Image automatically updated during installation
  - Host ID emulation
  - <http://www.oracle.com/technetwork/articles/servers-storage-admin/p2vvirtualizationmigration-170693.pdf>
- Update on attach
  - Existing update-on-attach is maximally conservative
  - “Update All”-on-attach is more liberal
    - -U (upper-case-U)

# Real world examples

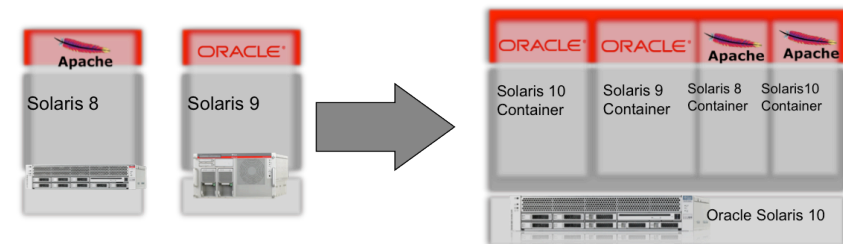
## Eco Consolidation

40 x 8 way systems	One 16 way system	Total Savings
<b>Before</b>  <ul style="list-style-type: none"> <li>• 320 CPUs</li> <li>• OS support: \$140,000</li> </ul>	<b>After</b>  <ul style="list-style-type: none"> <li>• 16 CPUs</li> <li>• OS support: \$42,000</li> </ul>	<b>Total</b>  <ul style="list-style-type: none"> <li>• 95% less CPUs</li> <li>• 70% lower annual support cost</li> </ul>

## Rapid Application Deployment



## Protecting Web Applications



## Legacy Investment Protection



# **Oracle Solaris 11 Express Enhancements**





## Zones for Oracle Solaris 11: Big Ideas

- Rationalized install and update
- Greater runtime “completeness”
- Focus on data center class storage
- Oracle Solaris 10 Zones
- Exclusive-IP network stack enhancements
- Integration with new network stack architecture (Crossbow)
- Accurate Utilization Monitoring



## Packaging and Installation

- IPS folds patching, packaging, ZFS, live upgrade, update on attach, and networked package repositories into a unified solution
- Packages have sane names, proper dependencies
- Newly installed zones use IPS for minimization; install is fast
  - User or enterprise tools can add additional software without involving global zone admin
  - Enterprises can use signed pkg support to limit contents
- Boot Environment management extends to zones



## Packaging and Installation (2)

- Whole and Sparse Zones implementations merging
  - Write anywhere you like in the zone
  - Removes ISV confusion
  - Work to preserve read-only-ness is being studied
- To update zones in Oracle Solaris 11 Express:
  - Update global zone (pkg update)
  - Reboot to new BE
  - For each zone:
    - Detach zone
    - Attach zone -u



# Zones Observability

- Improved Utilization Monitoring
  - CLI and Ops Center integration
  - Use extended accounting for accuracy
  - Report shared and dedicated resources
  - Utilization against configured limits

# Introducing zonestat(1m)

- `$ zonestat 5`

- ...

- SUMMARY Cpus/Online: 32/32 Physical: 32.0G Virtual: 47.9G
- -----CPU----- ----PHYSICAL----- -----VIRTUAL-----
- ZONE USED %PART %CAP %SHRU USED PCT %CAP USED PCT %CAP
- [total] 1.57 4.92% - - 5660M 17.2% - 9.9G 20.6% -
- [system] 0.09 0.28% - - 5086M 15.5% - 9275M 18.8% -
- kodiak-dp 1.00 100% - 100% 46.0M 0.14% 4.49% 36.2M 0.07% 1.17%
- global 0.48 1.56% - 1.56% 419M 1.27% - 673M 1.37% -
- kodiak-ab 0.00 0.00% - 0.01% 67.0M 0.20% - 115M 0.23% -
- kodiak-rie 0.00 0.00% - 0.02% 41.6M 0.12% - 62.4M 0.12% -

- Virtual: Really “swap reservation”



## Introducing zonestat(1m)

- `zonestatd` Daemon performs monitoring
  - Allows non-root users and non-global zones to see (some of) the information
- Zonestat can monitor:
  - virtual-memory, physical-memory, locked-memory, pool-psets, lwps, processes, shm-memory, shm-ids, sem-ids, msg-ids
- Limit output to specific zones
- Sort by various columns
- Machine parseable output mode
- End-of-run reporting for average, high, total usage.
- Drill down by resource type

# Introducing zonestat(1m)

- Example: Monitor lwps & processes:

```
$ zonestat -r processes,lwps 5
PROCESSES          SYSTEM LIMIT
system-limit      292K
      ZONE  USED   PCT   CAP  %CAP
    [total]  191 0.63%   -    -
   [system]    0 0.00%   -    -
     global  167 0.55%   -    -
        foo   24 0.08%  300 8.00%

LWPS          SYSTEM LIMIT
system-limit  2047M
      ZONE  USED   PCT   CAP  %CAP
    [total]  713 0.00%   -    -
   [system]    0 0.00%   -    -
     global  618 0.00%   -    -
        foo   95 0.00% 1000 9.50%
```

# Zones Security

- Delegated administration (via RBAC authorizations)
  - Authorizations can be configured directly in zonecfg(1m):

```
example# zonecfg -z myzone
zonecfg:myzone> add admin
zonecfg:myzone:admin> set user=jack
zonecfg:myzone:admin> set
auths=login,manage
zonecfg:myzone:admin> end
zonecfg:myzone> commit
```

- Authorizations are implemented via /etc/user\_attr and synced there by zonecfg.

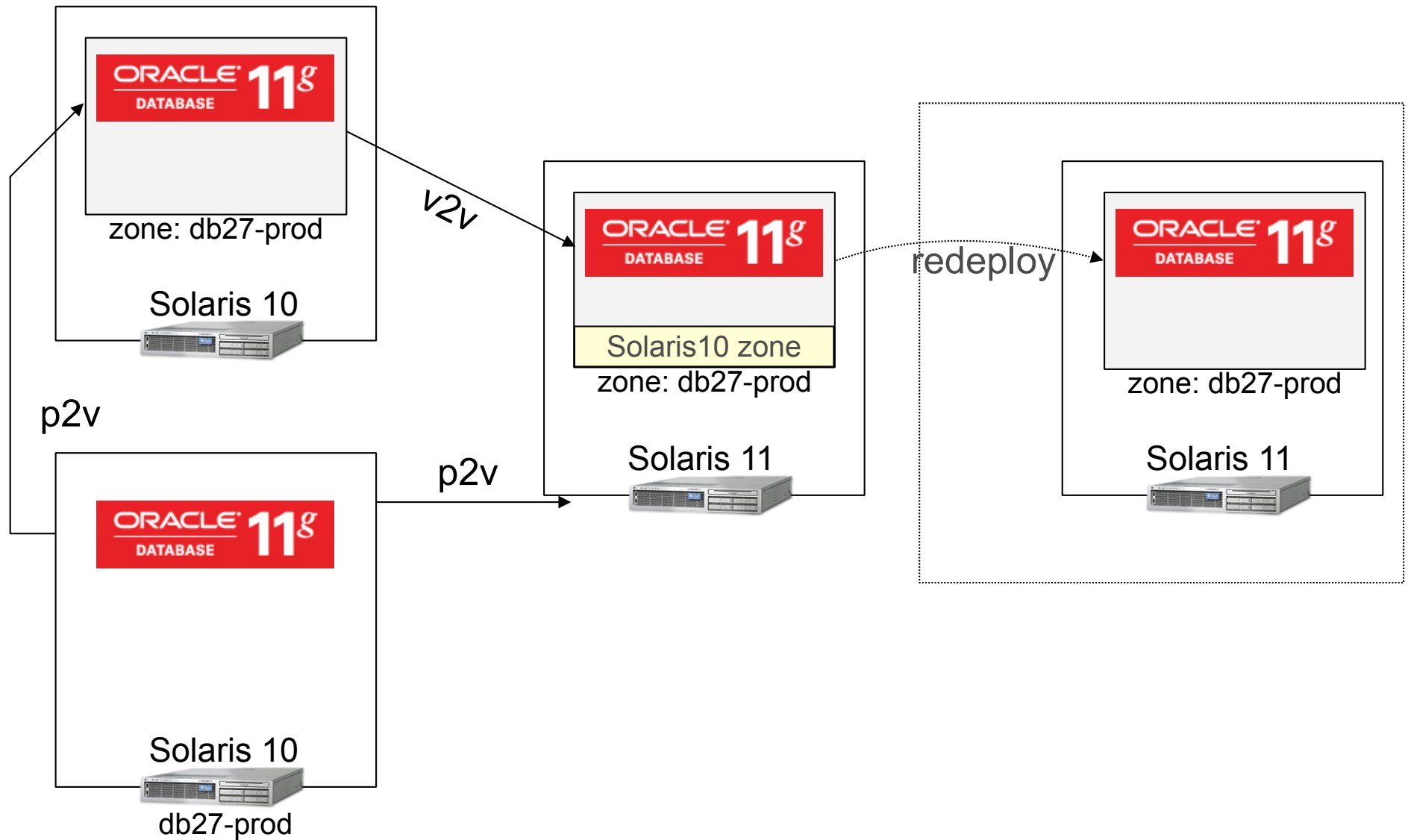




# Solaris 10 Zones on Solaris Next

- *solaris10* branded zone
  - Similar to existing solaris8 and solaris9 brands on S10
- Adoption and compatibility aid for Solaris Next
  - Protect investment in S10 (infrastructure, training, support)
  - Leverage new technology in an S10 context
    - e.g. Crossbow for Solaris 10
  - Avoid required application recertification
- p2v installation process
  - Also v2v for moving Solaris 10 native zones
- Supports Solaris 10 10/09 or later within the zone

# Investment protection



# Networking: Exclusive-IP Zones

- Extend and improve exclusive IP stack:
  - New `allowed-address` property constrains which IP addresses zone can use (via in-kernel L2/L3 protection)
  - `defrouter` property now supported for exclusive-IP zones

```
example# zonecfg -z myzone
zonecfg:myzone> set ip-type=exclusive
zonecfg:myzone> add net
zonecfg:myzone:net> set allowed-
address=11.1.1.32/24
zonecfg:myzone:net> set physical=vnic0
zonecfg:myzone:net> set defrouter=11.1.1.1
zonecfg:myzone:net> end
zonecfg:myzone> commit
```

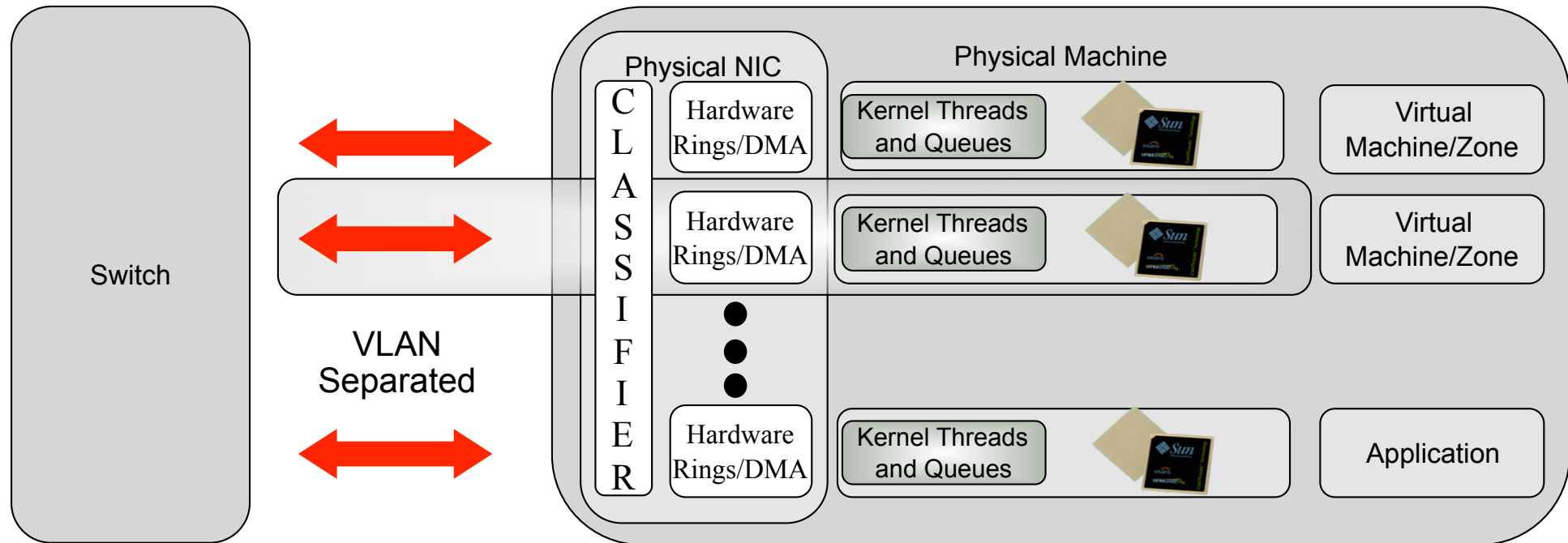
- Result: Shared-Stack style control over IP addresses with exclusive-stack features.



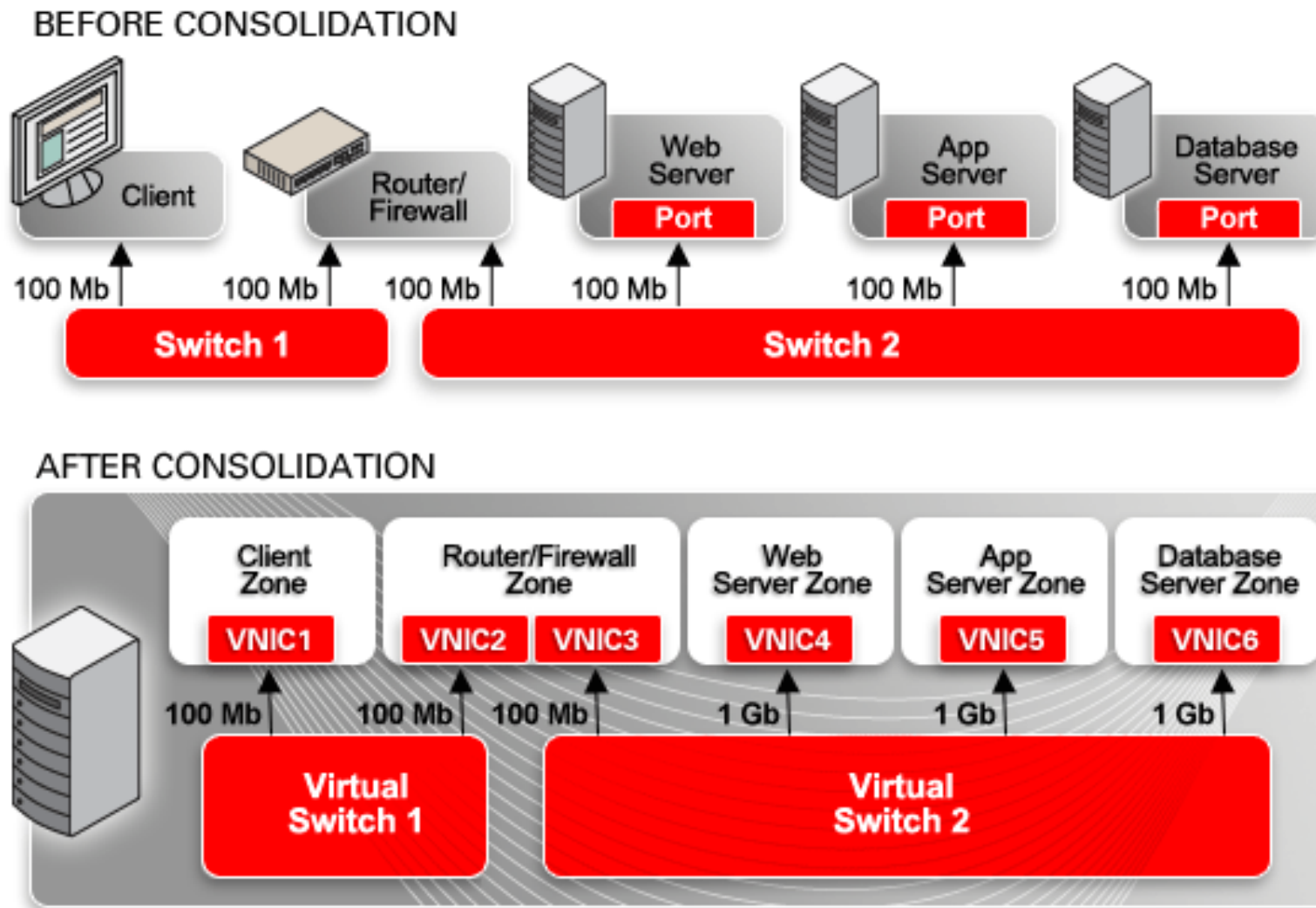
## Networking: Exclusive-IP Zones (2)

- Most of the benefits of network administrative rework accrue to zones: dladm, ipadm, IP tunnels, IPMP, etc.
- Crossbow and Zones work well together:
  - Create arbitrary numbers of vnics and assign them to zones, lifting S10 limitations for exclusive stack
  - Zones work seamlessly with crossbow virtual networks
  - Resource management and flow controls
- Layer 2 & 3 Networking Protections
  - Prevents exclusive stack zones from emitting various forms of mischevious traffic (MAC spoofing, IP spoofing, et cetera)
  - (Shared stack zones cannot by definition do these things)

# Redesigned network stack

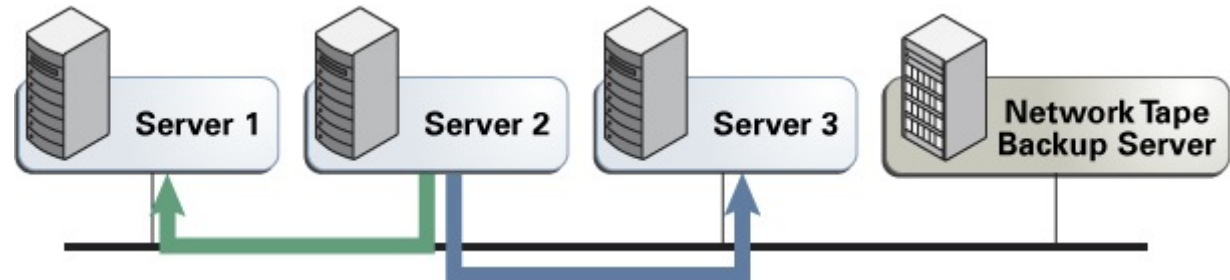


# Network in a box

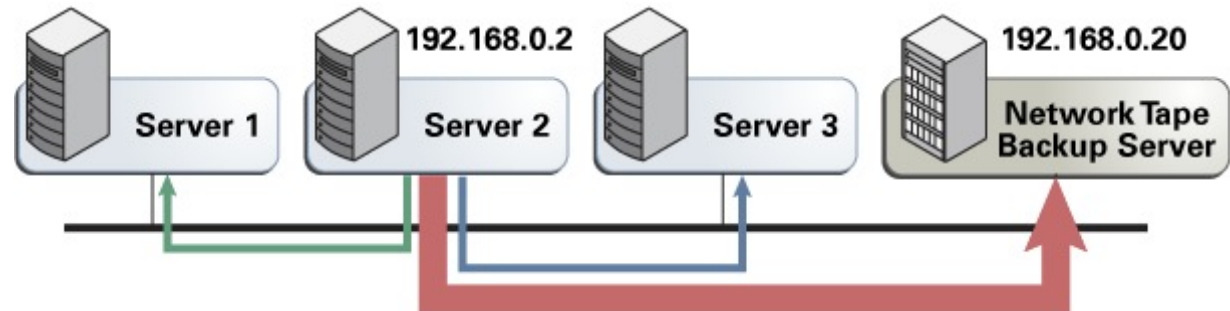


# Traffic flows, resource management and monitoring

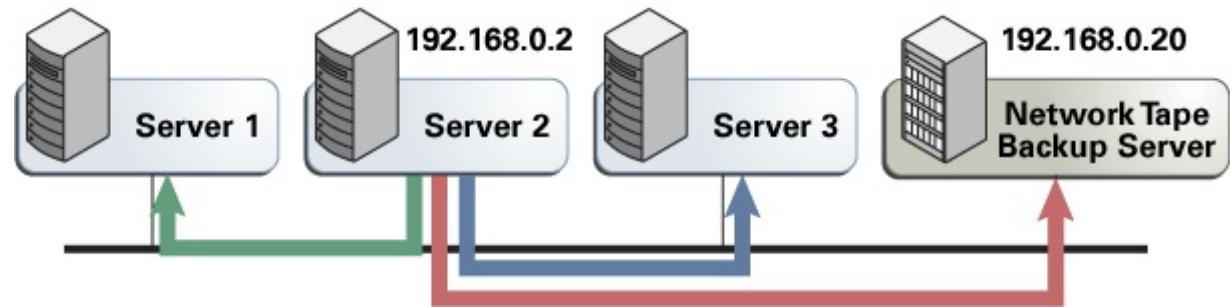
Prior to starting network backup



Network backup soaking up so much net bandwidth it impacts other communication



After limiting bandwidth used by backup, other traffic no longer suffers





## Shared-IP Zones: IPMP

- Address moves during failover/failback plagued the old model:
  - Zone would have ce0:2 one moment, ce1:1 the next
  - Zone administrator had no fixed point of control for the zone's IP addresses, nor any clues that the names might change
- With New IPMP:
  - IPMP no longer affects the zone – e.g., ipmp0:2 remains ipmp0:2 over the lifetime of the zone
  - Should the zone administrator be interested, the IFF\_IPMP flags will make it clear that the address is highly available





## Shared-IP Zones: Misc

- `snoop(1m)` support in the global zone for loopback (i.e. `lo0`) devices
  - Snoop traffic flowing between zones!

# Resource Management

- New max-processes resource control

```
example# zonecfg -z myzone
zonecfg:myzone> set max-processes=300
```

- prctl(1) now shows resource utilization:

```
example# prctl -i zone foo
zone: 4: foo
NAME      PRIVILEGE      VALUE      FLAG      ACTION
zone.max-lofi
          usage      0
          system    18.4E      max      deny
zone.max-swap
          usage    28.3MB
          privileged 3.00GB      -      deny
          system    16.0EB      max      deny
```



# Storage

- lofiadm(1m) and lofi(7d) now supported inside of zones
  - Resource control to limit max lofi devices
- Zones storage device support
  - Add block/raw storage devices to a zone without reducing system security

## If it must run, run it on Solaris



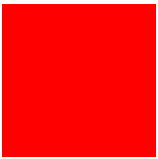
- Solaris 10 : great new features
- Solaris 11 Express: also great new features
- More importantly: S11E Integrated new features



## Next Steps

- Product overview, download and support information
  - [oracle.com/solaris](http://oracle.com/solaris)
- System administrators
  - [oracle.com/technetwork/systems](http://oracle.com/technetwork/systems)
- Oracle Technology Network
  - [oracle.com/technetwork/server-storage/solaris](http://oracle.com/technetwork/server-storage/solaris)





*Q&A*

# **Hardware and Software**

## **Engineered to Work Together**