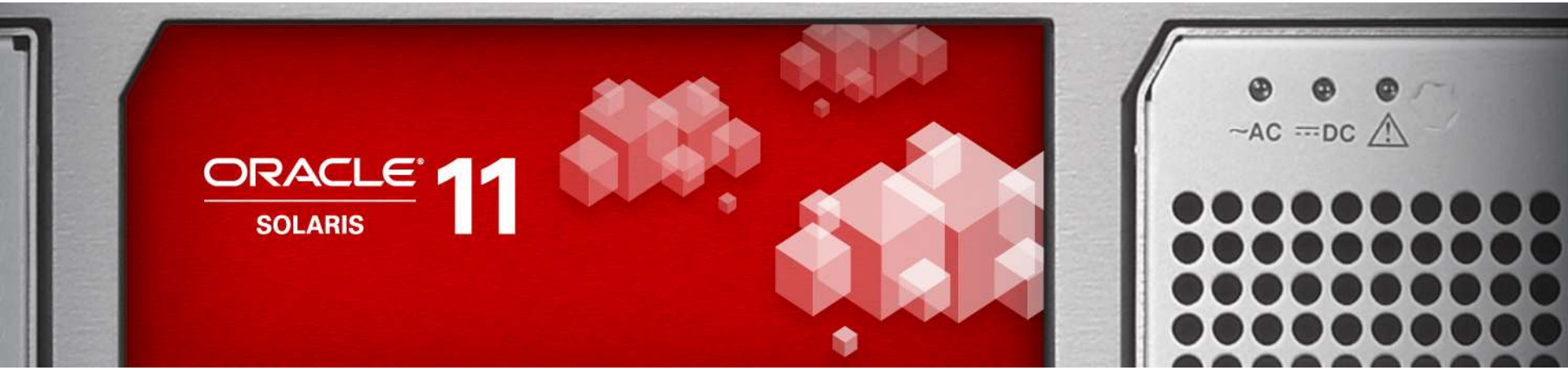

ORACLE®

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.



ORACLE

Oracle Solaris Zones Best Practices

Detlef Drewanz
Principal Sales Consultant, EMEA Server Presales

Agenda

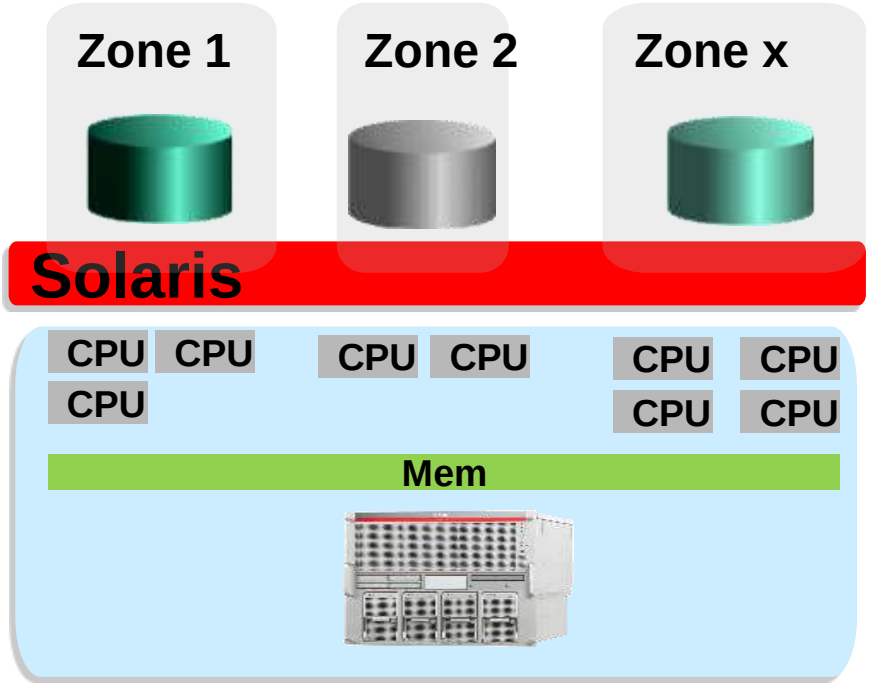
- Solaris Zones
- Zones Use Cases
- Resource Management
- Solaris 11 Zones News
- Networking and Solaris 11 Zones
- Security Improvements in Solaris 11 Zones
- Updating Zones

Agenda

- **Solaris Zones**
- Zones Use Cases
- Resource Management
- Solaris 11 Zones News
- Networking and Solaris 11 Zones
- Security Improvements in Solaris 11 Zones
- Updating Zones

What is a Zone (a.k.a. Container)

- Virtualization above kernel
- Shared Kernel
- Separate file systems
- Complete software isolation
- Sub-thread granularity
- Low overhead



Oracle Solaris Zones (a.k.a. Containers)

- One Solaris, 1000's of Zones, no additional costs
- Available on every system where Solaris runs
- OS-virtualization, isolation and resource limitation
- For efficient and secure consolidation
- Easy partitioning on application layer
- Fast and easy cloning and migration of zones
- Instant restart

Agenda

- Solaris Zones
- **Zones Use Cases**
- Resource Management
- Solaris 11 Zones News
- Networking and Solaris 11 Zones
- Security Improvements in Solaris 11 Zones
- Updating Zones

Zones use cases

- Consolidation
  - Optimize workload and licences
  - Consolidate small, test, development, learning systems
- Isolation
  - Multiple instances of the same service on a system
  - Development/Test/Quality Ensurance/Production on one system
- Create new architectures
  - Flexible service operation (instant restart)
  - Encapsulate applications
  - Operate legacy applications on new hardware

Capped Containers

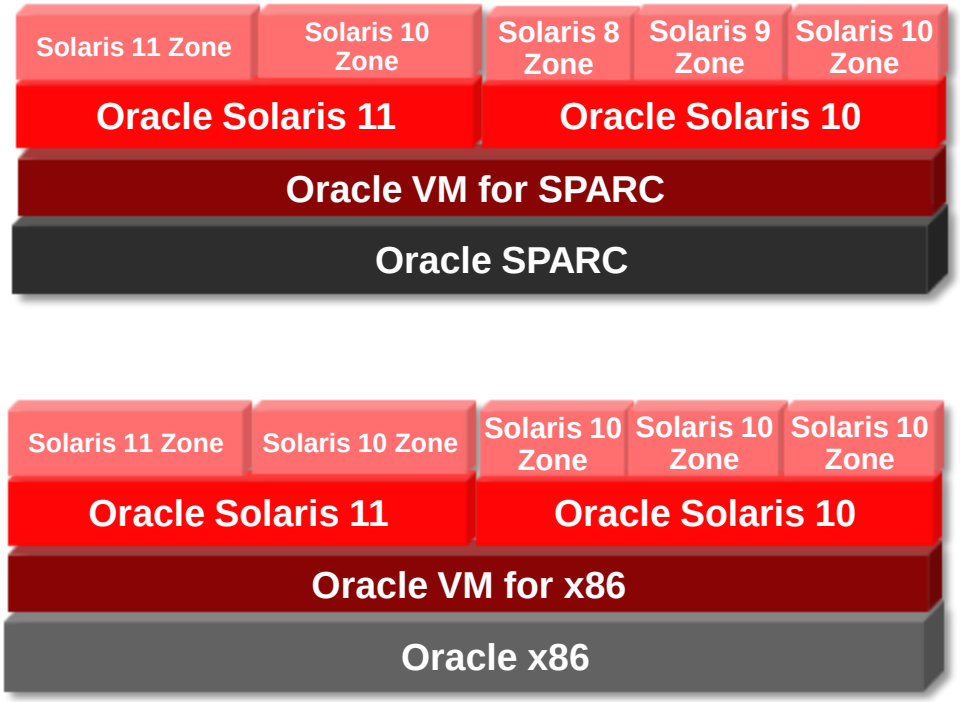
To optimize license usage

- Capped Container accepted as Hard Partitions
  - Create processor set
  - Assign pset to resource pool
  - Bind Zone to resource pool
- Licence assigned CPU in resource pool
- Not to confuse with capped-cpu feature

Terminology

- One **Global Zone** per System
  - Installed directly on bare metal or into VM
- Multiple **Non-global Zones** sharing one global Zone
  - Virtualized Environment
- A **Branded Zone** emulates a non-native OS Environment
  - **Solaris 10 Zone**
    - A branded Zone used to run a Solaris 10 user space
  - **Solaris Legacy Container**
    - A branded Zone used to run a Solaris 8 or Solaris 9 user space

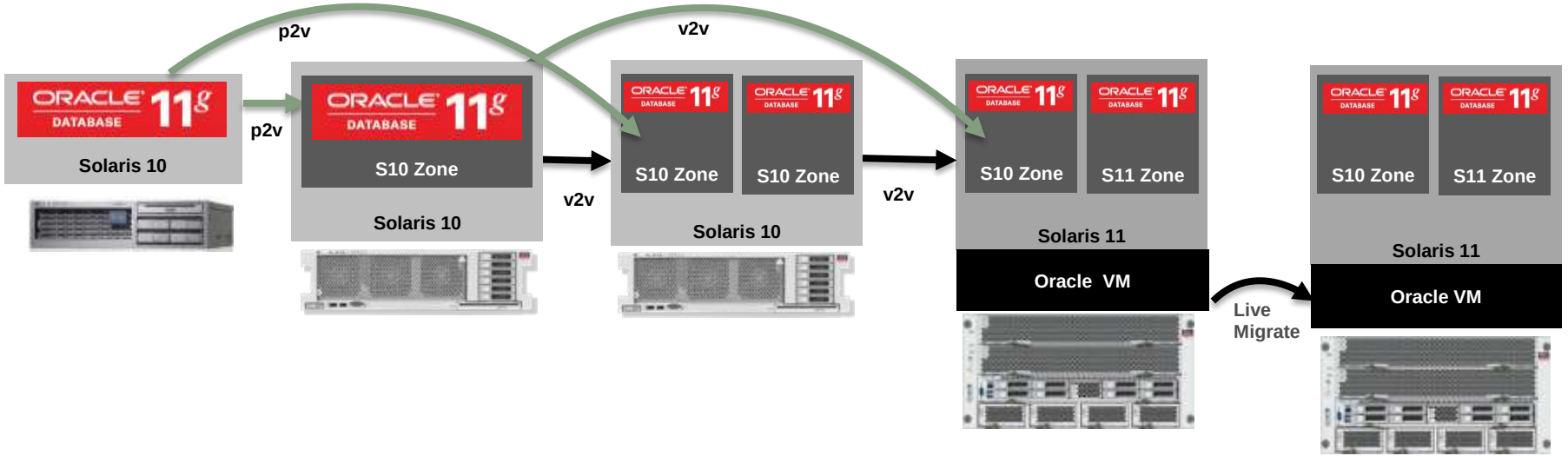
Application Investment Protection



Seamless Upgrades

Oracle Solaris 11 Zones, Oracle VM

- Seamless upgrades from previous version
  - Assisted with a built-in pre-flight checker
- Live migration with OVM SPARC and OVM x86



Agenda

- Solaris Zones
- Zones Use Cases
- **Resource Management**
- Solaris 11 Zones News
- Networking and Solaris 11 Zones
- Security Improvements in Solaris 11 Zones
- Updating Zones

Resource Management

- Provide and limit shared Resource to all Zones
- Consolidation of workloads with different requirements
  - Throughput oriented
  - Response time critical
  - Availability
- Enable through resource management
- Keep resource assignment changeable through runtime

Plan for Capacity

With Clear Observability

- To plan you need to clearly see your environment
  - zonestat
  - flowstat
  - DTrace
- Plan for consolidation
- Plan for expansion
- Plan for re-design
- Plan for re-design connect

Zones Resource Management

- Balance
  - Faire Share Resources through rules
  - Assigned based on shares in the event of 100% utilization (no limit below 100%)
- Capping
  - Cap Resources on a Limit
- Partitioning
  - Assign and use Resources Exclusively

Resource Management

Helps organizations meet service level agreements

CPU cap

CPU shares

Memory Cap

Swap Cap

Resource management with zones (keep SLAs!)

- CPU
  - CPU Fair Shares
  - CPU Capping
  - CPU Partitioning
- Memory
  - Used phys memory (RSS)
  - Virtual memory (= Swap)
  - Non-Pageable Memory (Locked)
  - Shared Memory, Semaphores
- Processes
  - LWP‘s

```
zone.max-swap
        system          16.0EB    max   deny    -
zone.max-locked-memory
        system          16.0EB    max   deny    -
zone.max-shm-memory
        system          16.0EB    max   deny    -
zone.max-shm-ids
        system          16.8M     max   deny    -
zone.max-sem-ids
        system          16.8M     max   deny    -
zone.max-msg-ids
        system          16.8M     max   deny    -
zone.max-lwps
        system          2.15G     max   deny    -
zone.cpu-cap
        system          4.29G     inf   deny    -
zone.cpu-shares
        privileged          1     -     none    -
        system          65.5K     max   none    -
```

Agenda

- Solaris Zones
- Zones Use Cases
- Resource Management
- **Solaris 11 Zones News**
- Networking and Solaris 11 Zones
- Security Improvements in Solaris 11 Zones
- Updating Zones

Oracle Solaris 11 Zones Improvements

- ZFS datasets
- Boot environments in Zones
- Delegated administration
- Observability via zonestat
- Solaris 10 Zones
- NFS Server in a Zone
- Network virtualization and resource management

Oracle Solaris 11: Integrated Virtualization

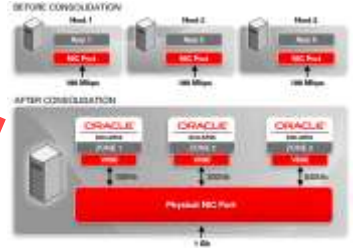
Provisioning

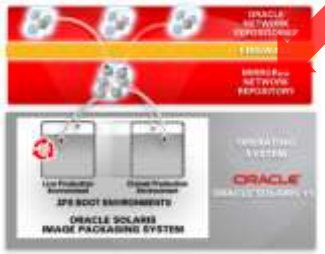
Security

Network Virtualization

Software Management

Data Management



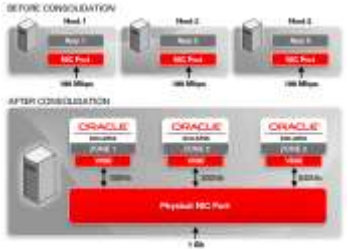


Oracle Solaris 11: Integrated Virtualization

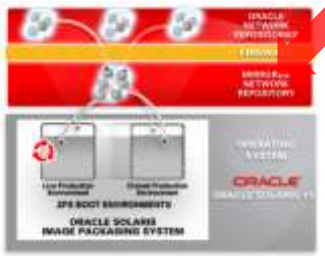
Provisioning
-Automated Installer
-Distro Constructor

Security
-ZFS Encryption
-Immutable Zones
-Delegated Admin

Network Virtualization
-Network in a box
-Bandwidth Control
-Resource Mgmt



Software Management
-IPS
-Repositories
-Boot Environments



Data Management
-ZFS
-COMSTAR

Resource Management with Zones - New in S11

- Processes
  - Processes (new with S11)
- Networking (new with S11)

Resource Management in Solaris 11

- New max-processes resource control

```
cantaloup# zonecfg -z keetonga
  zonecfg:keetonga> set max-processes=300
```

- prctl(1) now shows resource utilization:

```
cantaloup# prctl -i zone keetonga
  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
```

Resource Management

Helps organizations meet service level agreements

CPU cap

CPU shares

Memory Cap

Swap Cap

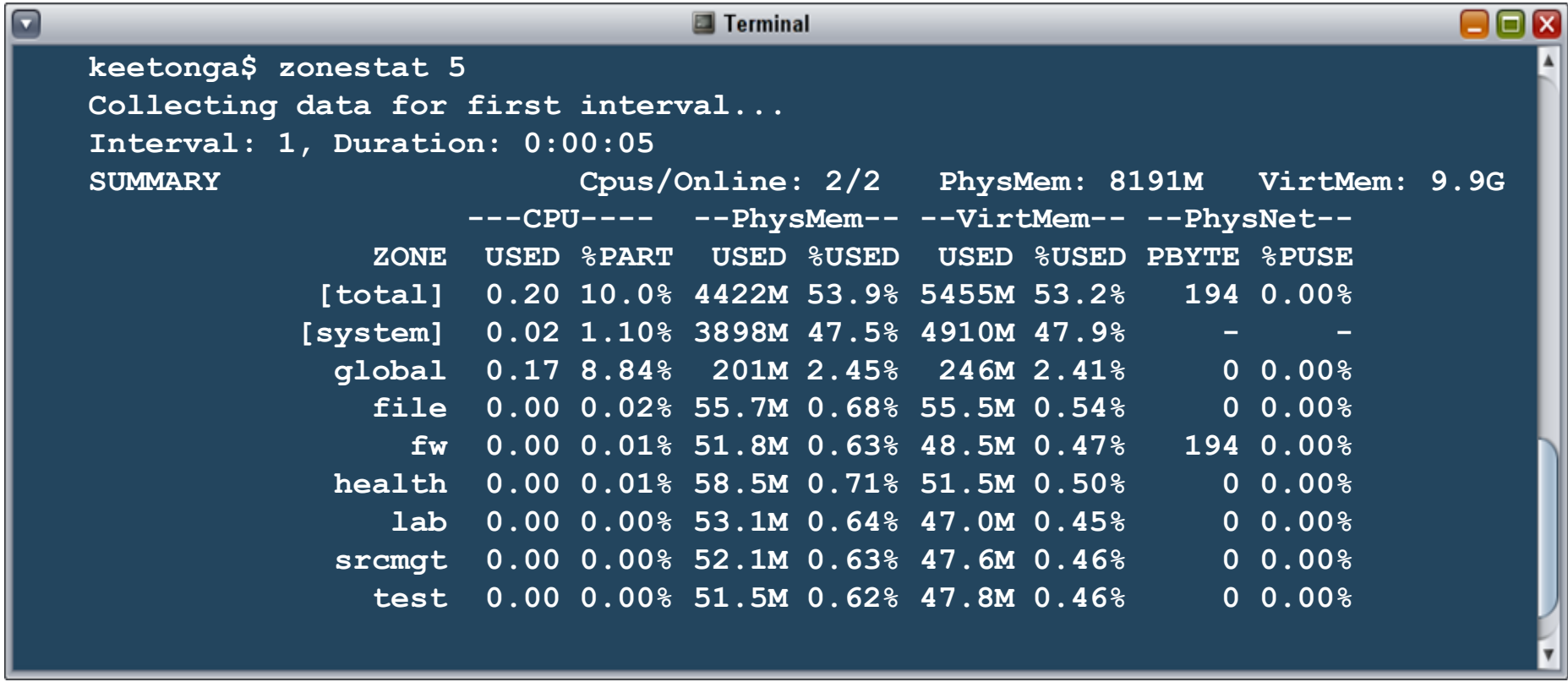
New for Oracle Solaris 11:

Bandwidth Cap

CPUs for Networking

Max Processes per Zone

Observability



Solaris 11 Zone Installation

- Zone root by Default on own ZFS Dataset (compressed)
- One Zones model (no more to distinguish sparse/whole)
- Zones Minimization
  - Install by default *pkg://solaris/group/solaris-small-server*
- Zone Installation
  - Automatic: with profiles and Automated Installer (AI)
  - Interactive: similar to AI based install
  - Automatic Zone upgrade: (pkg update in global zone)

Zone installation (2)

- IPS Necessary to Install Packages in a Zone
  - Set http_proxy or https_proxy for GZ if behind a firewall
- Zones inherit Publishers from Global Zone
  - No need to manage repositories in the zones
- IPS proxy to global zone
  - Allows zones to install pkg regardless of network config

Solaris 11 Zones Deployment

- AI is also used when installing zones interactive
- Default manifest

  /usr/share/auto_install/manifest/zone_default.xml
- Default profile enables interactive system configuration
- Provide alternate manifest and/or profile with

  zoneadm -z <zone> install -m <manifest> -c <profile>.xml
- Create profile with

  sysconfig create-profile –o <profile>.xml

Bootenvironments and Zones

- Nested boot environment for Zones (zbe)
- Can be different from global Zone (non-kernel pkgs)
- A zone admin can create boot environments
- A zbe in a zone belongs to the be of the global zone
- A zbe can be booted, mounted, cloned
- A zbe belongs to one be in a global Zone
- A be mount in the global zone mounts all belonging zbe

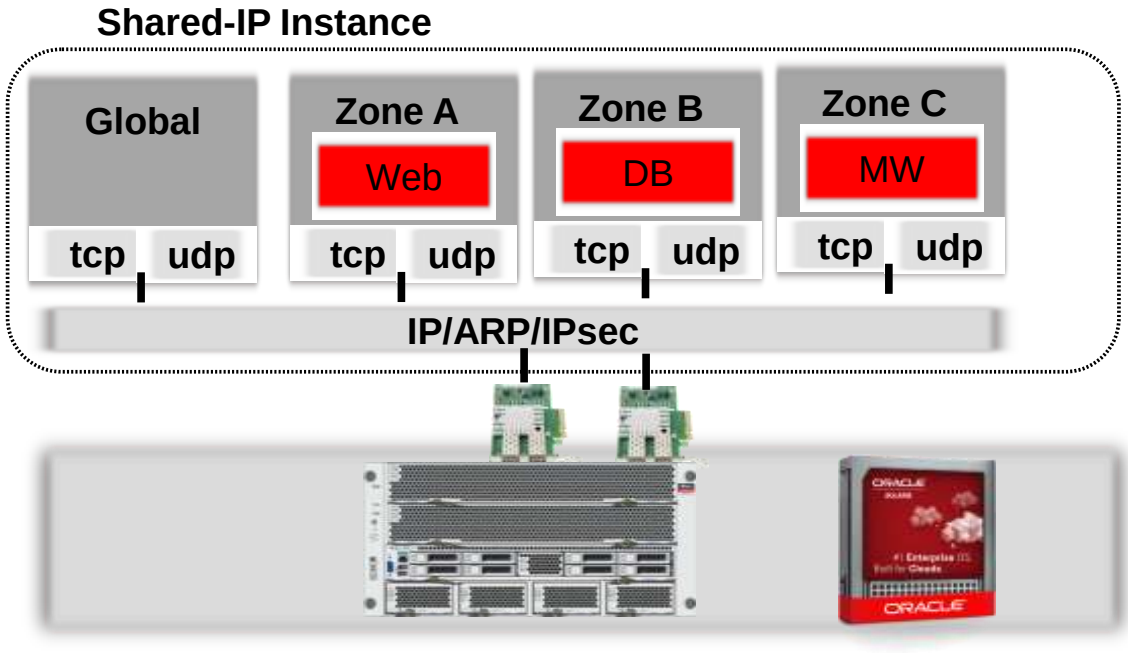
Solaris 10 Zones

- Runs a Solaris 10 userland in Solaris 11
  - Is able to use vnics
- v2v: move a Zone from Solaris 10 to Solaris 11
- p2v: move a Solaris 10 global zone to a Solaris 10 Zone
  - The Solaris 10 global Zone runs then as a non-global zone
  - Non-global Zones need to be moved first
- Must be Solaris 10 10/09 or newer
- Solaris 8 or Solaris 9 zones do not run in Solaris 11

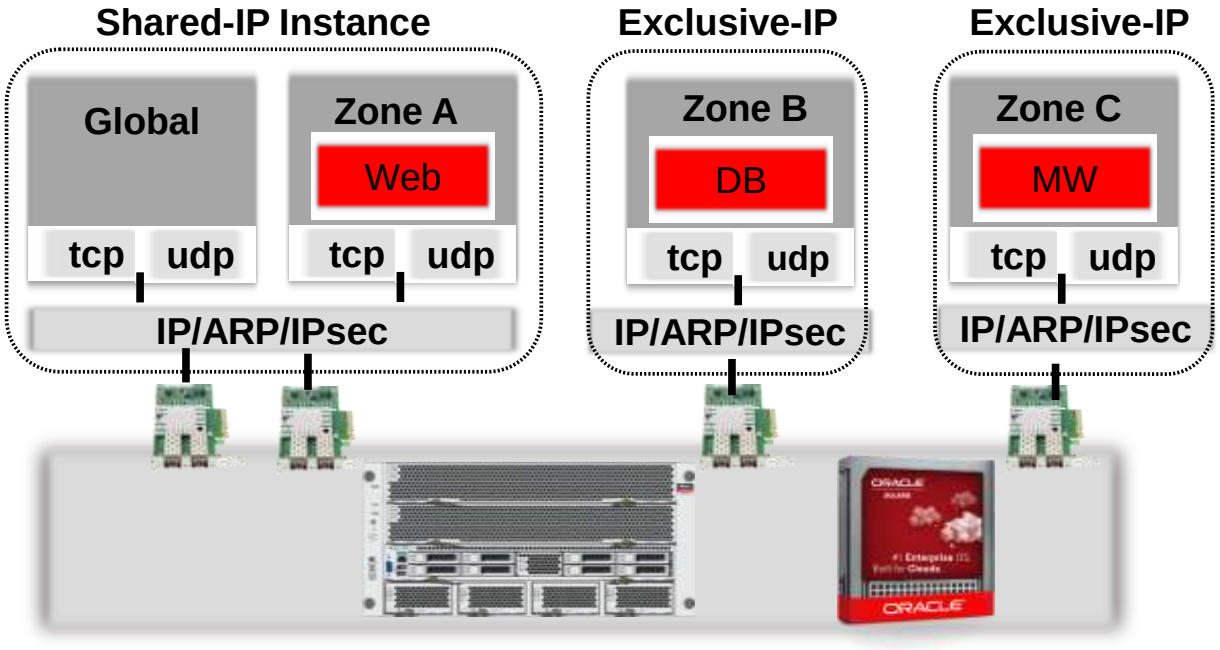
Agenda

- Solaris Zones
- Zones Use Cases
- Resource Management
- Solaris 11 Zones News
- **Networking and Solaris 11 Zones**
- Security Improvements in Solaris 11 Zones
- Updating Zones

Shared-IP Instance



Shared-IP vs. Exclusive IP-Instances



Cloud-Scale Networking With Solaris 11

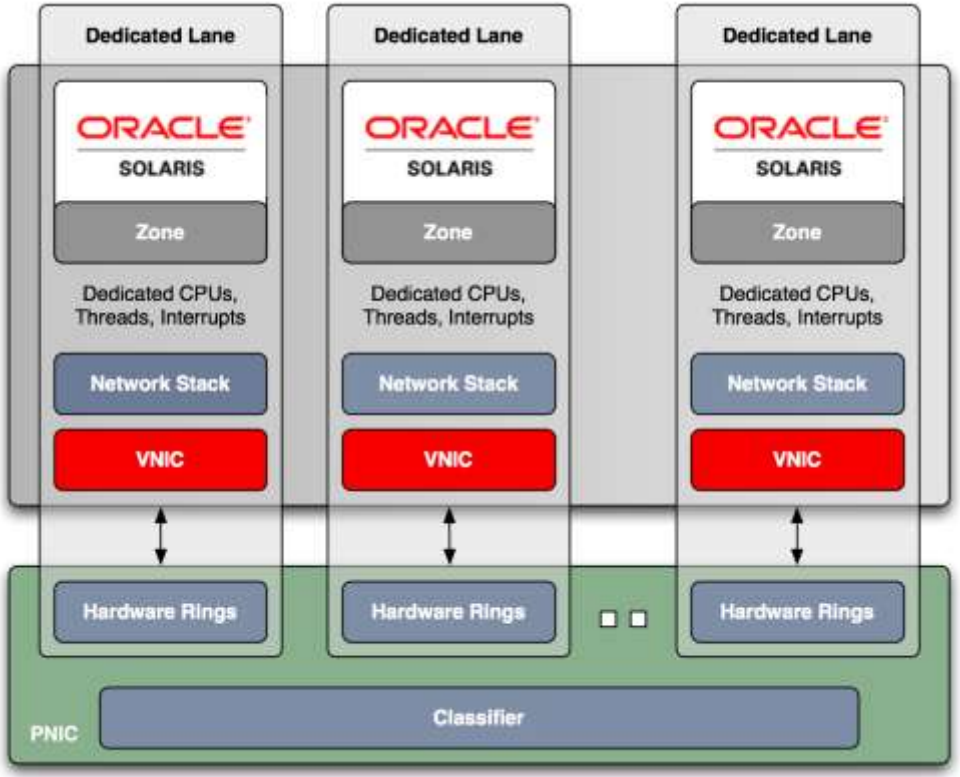
Network Virtualization	Virtual NICs (VNICs), Virtual switching, Hardware-assisted virtualization, Automatic VNICs for zones, SR-IOV Integration, VLAN isolation, Anti-spoofing protection
Resource Control	Integrated QOS, Bandwidth limits, Mapping to CPUs or CPU pools for isolation
Performance	Parallel stack, NUMA I/O Framework, SR-IOV Integration, Dynamic Polling, Buffer Management, Pre-mapped buffers, Kernel Socket API, 4x Lower latency vs KVM, Converged Ethernet
Built-in Network Funtionality	Routing, Firewall, Load Balancing, VRRP, Bridging
Management	IPMP re-architecture, Vanity naming, Automatic IP configuration, Centralized IP administration, Centralized data link administration, Consolidated data link properties, GLDv3 unification for legacy drivers
Observability	Real-time data link, hardware, and flow statistics. History integrated with extended accounting. Capture local traffic through through virtual switch and IP loopback path.
APIs	Committed GLDv3 APIs, pluggable TCP congestion algorithms, IP Filter Hooks, Kernel socket API

Solaris 11 Zones and Networking

- Uniqe network names: net0, net1, ...
- Default: Exclusive-IP
  - Automatic creation of vnic for Zones (add anet)
  - Restrict mac-address and ip-address
  - Maxbw, priority, vlan-id configurable
- Configuration moves with Zone
- Shared IP stack is still supported and configurable
- Snoop for loopback

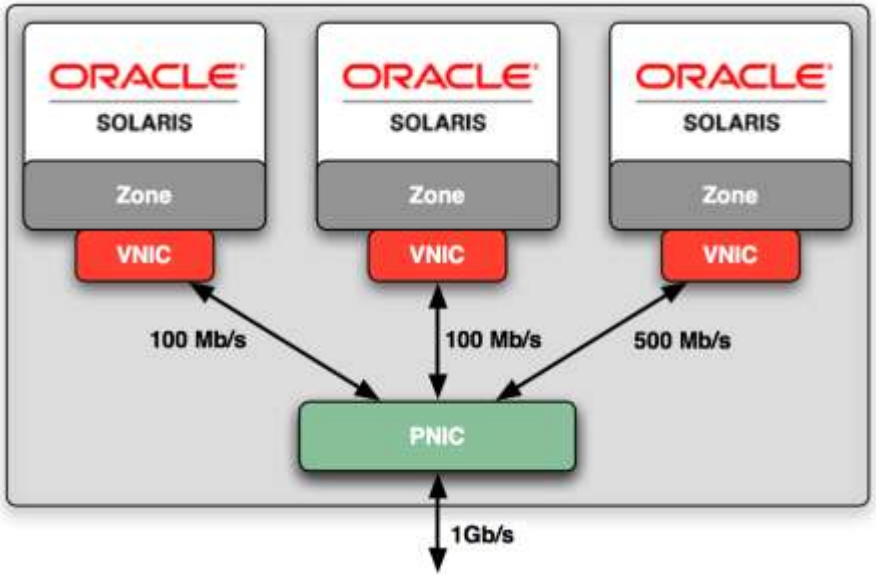
Parallel Network Virtualization Architecture

- Virtualization and QoS designed-in
- Independent Hardware Lanes with dedicated resources (CPUs, I/O threads, interrupts): from the NIC to applications
- VNIC behaves just like a regular NIC (link speed, stats, MAC address)
- Hardware and software fanouts for best scalability
- Adaptive polling mode depending on load



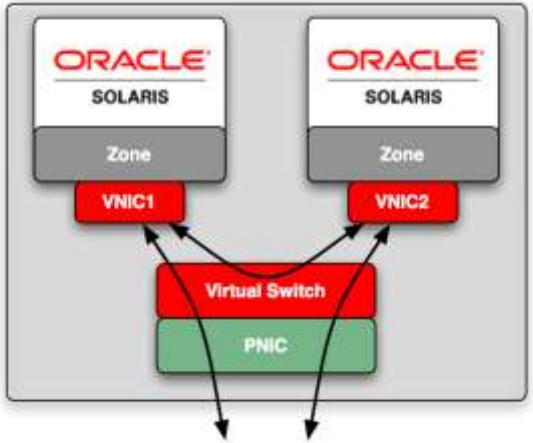
Network Resource Control

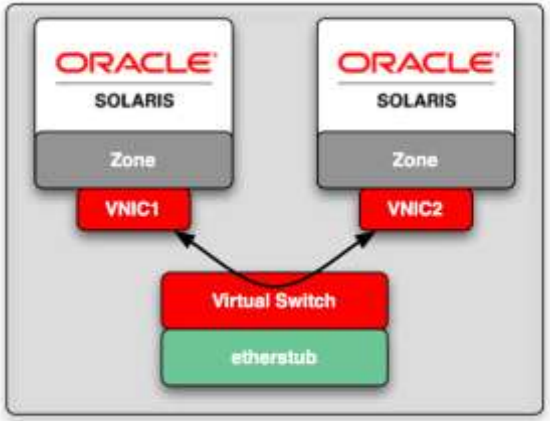
- Set bandwidth limit on a VNIC (virtual link speed)
- QoS integrated in the core stack, no separate component to configure
- Constrain the CPUs used by VNICs or data links by CPU ids or pool names
- Integrated with Solaris resource management and zones




```
# dladm create-vnic -l net0 \
-p maxbw=100M vnic0
```

Virtual Switching



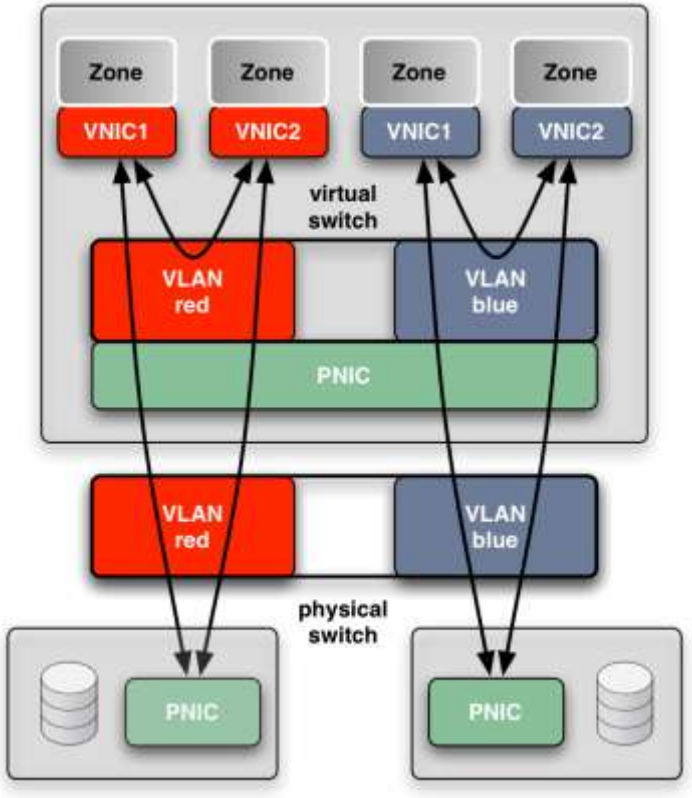


- A virtual switch is created automatically when VNICs are configured
- Virtual switches allow VNICs to communicate with each other and with hosts on the network

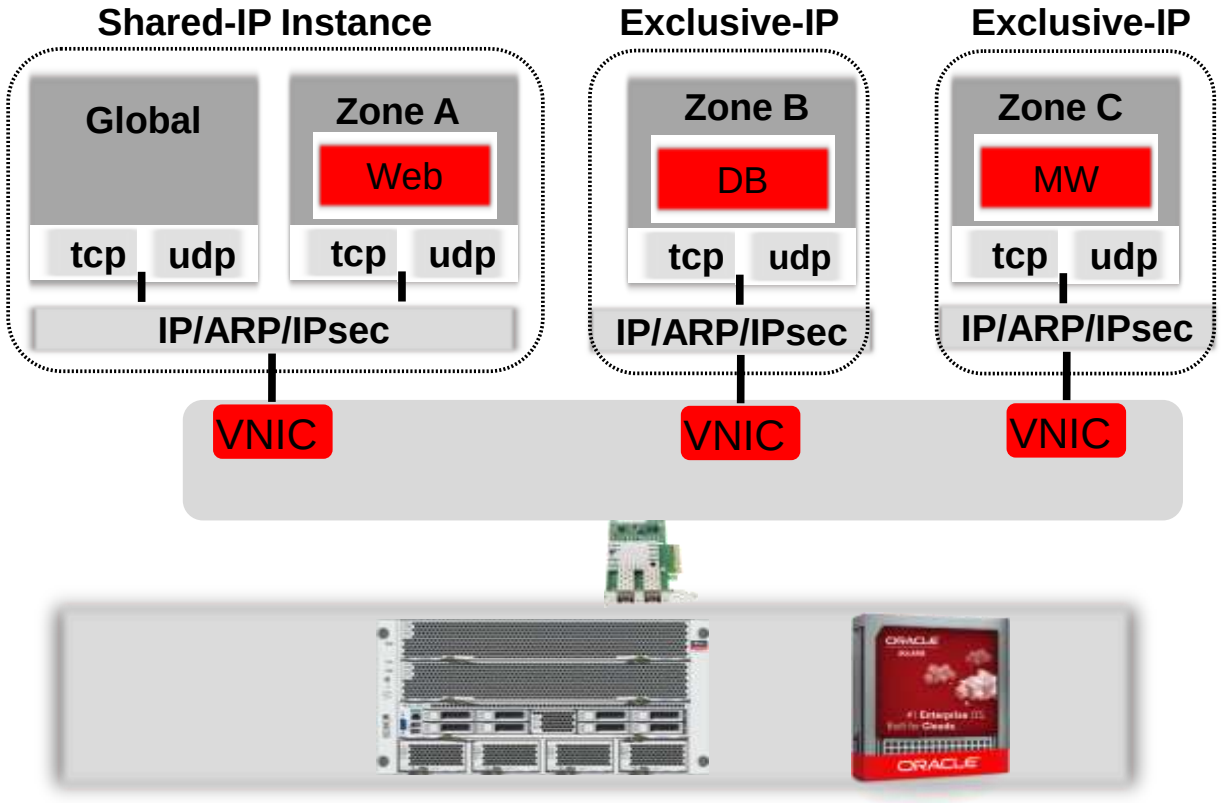
- Use etherstubs instead of physical NICs
- Build virtual switches that are independent from any hardware
- As many as you want on a single host

VLAN Separation

- VNICs can be assigned a VLAN id
- Virtual switch provides VLAN separation
  - Local traffic between VNICs
  - Traffic to and from external hosts
- Extend VLAN separation from physical network into virtual switch

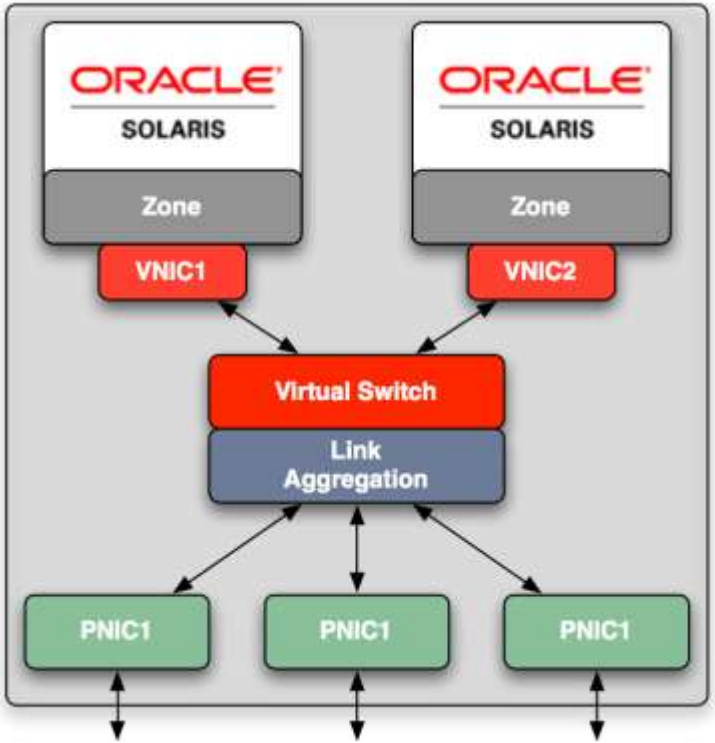


Shared-IP vs. Exclusive IP-Instances



Highly Available VNICs

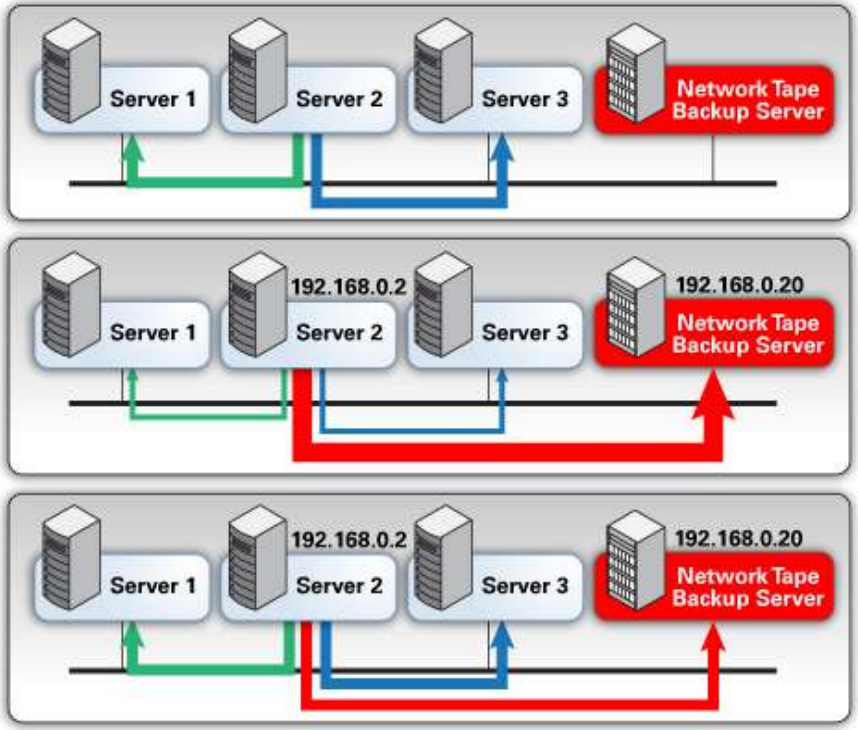
- Link Aggregation provides **transparent** failover and increased throughput to VNICs and zones
- Compliant with IEEE 802.3ad
- IP Multipathing (IPMP) can also be used, but needs to be configured from within zones



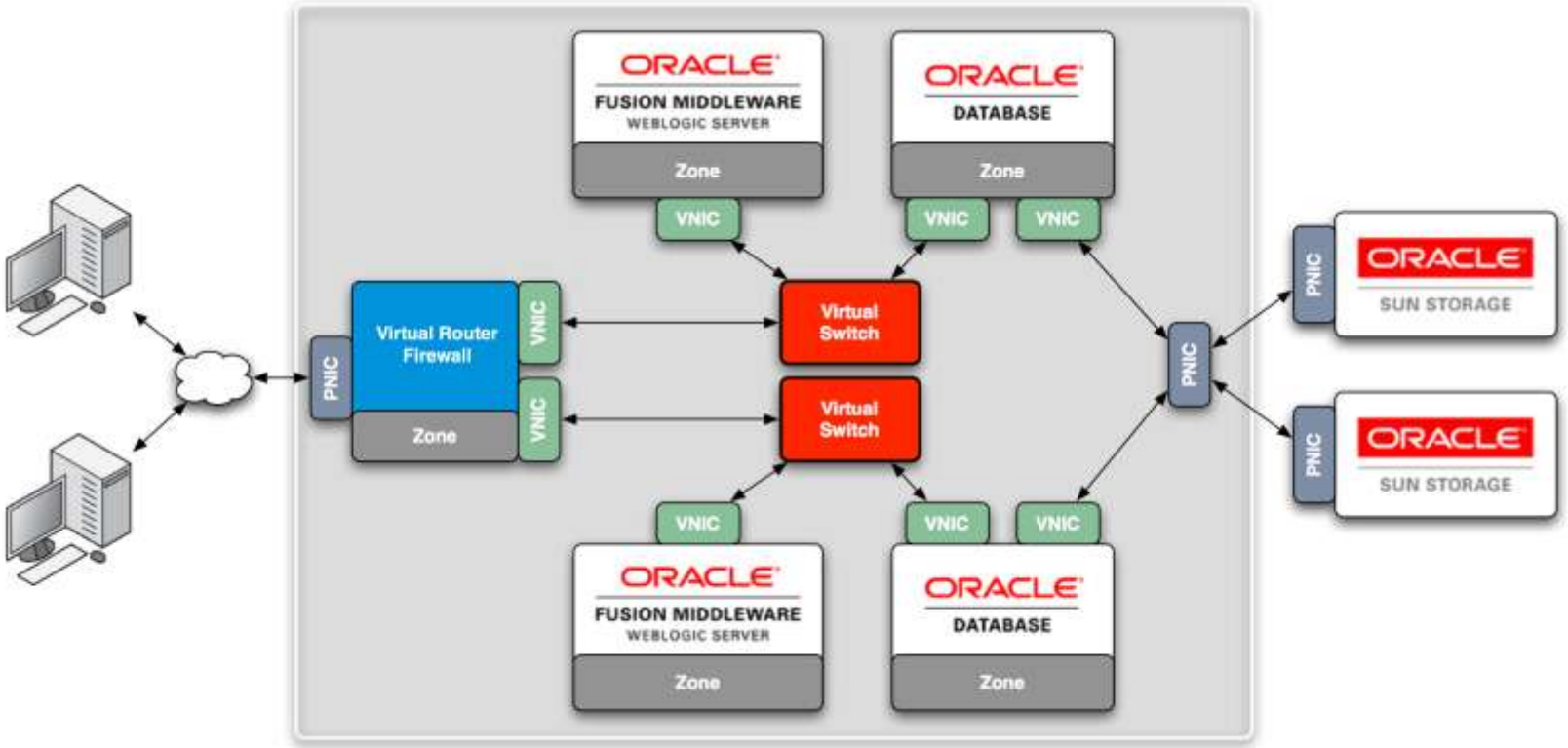
Controlling and Observing Flows

Control the Un-Controllable

- Built-in QoS can be applied to traffic flows specified by the administrator
- Managed by flowadm(1M) and specified by source and destination IP addresses, protocol, port number, etc.
- Flows can be observed in real time with flowstat(1M), or a history can be obtained using extended accounting



Virtual Multi-Tiered Architecture



Agenda

- Solaris Zones
- Zones Use Cases
- Resource Management
- Solaris 11 Zones News
- Networking and Solaris 11 Zones
- **Security Improvements in Solaris 11 Zones**
- Updating Zones

Zones Security Enhancements

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

```
cantaloup# zonecfg -z keetonga
   zonecfg:keetonga> add admin
   zonecfg:keetonga:admin> set user=detlefd
   zonecfg:keetonga:admin> set auths=login,manage
   zonecfg:keetonga:admin> end
   zonecfg:keetonga> commit
```

- Authorizations are implemented via /etc/user_attr and synced there by zonecfg.

Immutable zone

- Zones root filesystem set to read-only
  - Uses a mandatory write access control (MWAC) kernel policy
- Only OS part of zone (not the added filesystems)

  zonecfg set file-mac-profile = <config>
  - none:
  - strict: all files of root are read-only, only remote logging
  - fixed-configuration: some of /var writable, except system config directories)
  - flexible-configuration: /var and /etc files can be modified, ~sparse zones in S10
- Writable for administrative tasks: zoneadm boot -w

Zones for Solaris 11: Summary

- Rationalized installation, system configuration, update
- NFS server in a Zone, lofi improved
- Networking
  - Exclusive Stack now the default
  - Automatic networking
  - Network resource management in the zones config
- Immutable Zones and Delegated Administration
- Solaris 10 Zones

Agenda

- Solaris Zones
- Zones Use Cases
- Resource Management
- Solaris 11 Zones News
- Networking and Solaris 11 Zones
- Security Improvements in Solaris 11 Zones
- **Updating Zones**

Patching of Solaris Zones

- Patchlevel NGZ = Patch-Level GZ
- Delivered with Patch-Tools by default
  - Patching of GZ leads to patching of NGZ
- How to update „imported“ Zones on systems ?
  - Update-on-attach *zoneadm -z <name> -U*
  - Compares Patch and Package level between GZ and NGZ
  - Updates based on comparizon
  - Downgrade not possible

Patching of Systems with many Zones installed

- Patching of Zones requires careful of strategy
- Systems with many installed Zones demand one common downtime for all services in all Zones
- Challenge to coordinate downtime

Challenges

- Many installed Zones, but minimum downtime
- Update Zone by Zone not possible
- Behavior of Application through/after update
- Automate the update process
- How to Fallback ?

Update-Methods for non-global Zones

- Update a copy of a non-global Zone
  - Use Live Upgrade (Solaris 10) or Image Update (Solaris 11)
  - Clone a Zone and use Update-on-attach
- Create a new identical non-global Zone
  - Copy and update a reference Zone
  - Create an identical Zone by a script
- Important
  - Place application data onto separate storage (zpool)
  - Keep the application binaries in the Zones

Isolated Zones Update with Minimized Downtime

- Snapshot and clone running Zone to a separate zpool
- Transfer clone to system with newer software level
- Update clone
- Test the updated zone in a „sandbox“
- Activate the cloned
  - Move the application data in a separate zpool to the clone
- Downtime only during changing the roles

Summary

- Lightweight Virtualization technology at application level
- Isolation and Resource management
- Flexible technology with low overhead
- Careful Deployment planning recommended
- Now fully integrated into Solaris 11

Q&A

Detlef.Drewanz@oracle.com

Hardware and Software

ORACLE

Engineered to Work Together


ORACLE®