

**ORACLE®**

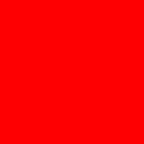


**ORACLE®**

## **Netzwerkvirtualisierung in Oracle Solaris**

Detlef Drewanz

EMEA Server Principal Sales Consultant



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.

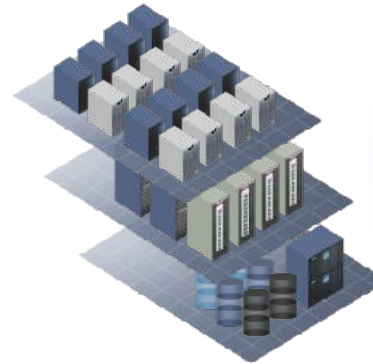
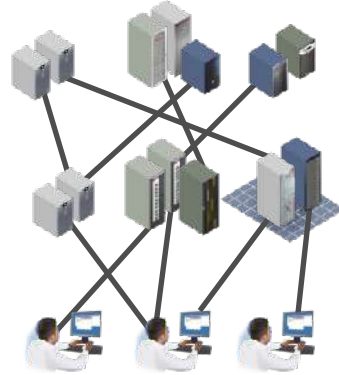
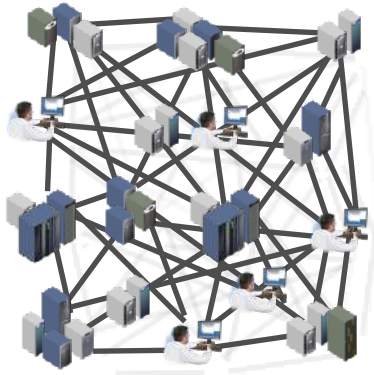
In addition, the following is intended to provide information for Oracle and Sun as we continue to combine the operations worldwide. Each country will complete its integration in accordance with local laws and requirements. In the EU and other non-EU countries with similar requirements, the combinations 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.

# Agenda

- Virtualisierung und Netzwerke
- Netzwerkvirtualisierung in Oracle Solaris
  - VNIC
  - Etherstub
  - Flows
  - Ressourcenmanagement
  - Accounting und Statistiken
- Anwendungsbeispiele



# Architekturen im RZ verändern sich



Silos

Konsolidiert

Optimiert

Cloud

# Virtualisierung auf allen Ebenen

- Anwendungen/Services
- OS
- Netzwerk
- Storage



## Wikipedia Zum Thema Virtualisierung

Virtualisierung bezeichnet Methoden, die es erlauben, Ressourcen eines Computers zusammenzufassen oder aufzuteilen.

Primäres Ziel ist, dem Benutzer eine Abstraktionsschicht zur Verfügung zu stellen, die ihn von der eigentlichen Hardware isoliert.

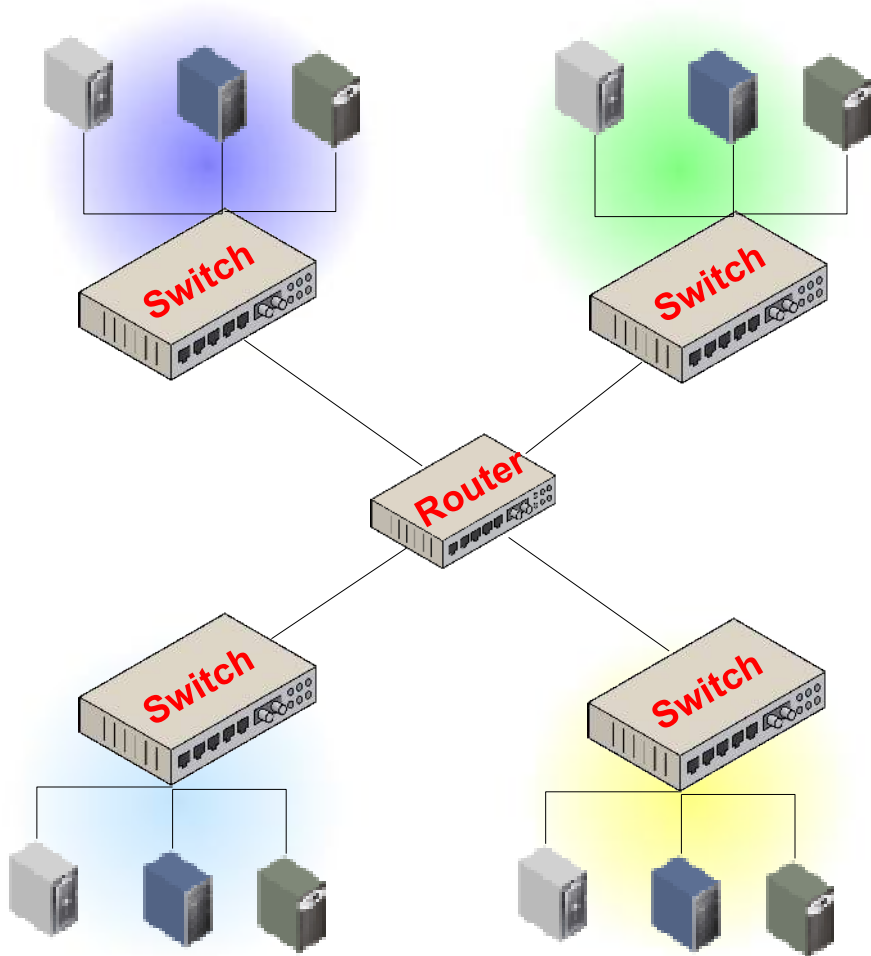
Eine logische Schicht wird zwischen Anwender und Ressource eingeführt, um die physischen Gegebenheiten der Hardware zu verstecken.

# Veränderungen durch Virtualisierung

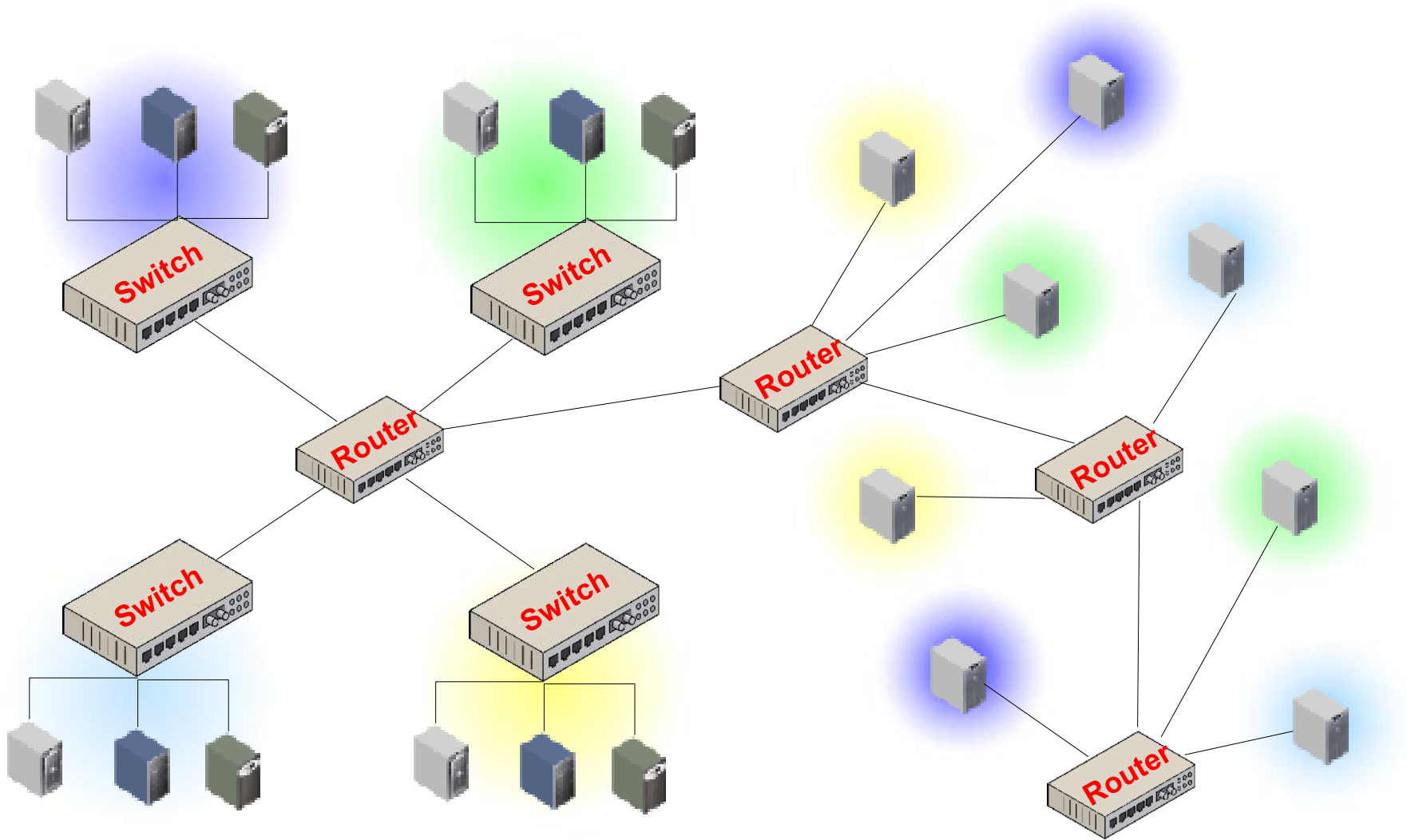
- Konsolidierung
  - Systeme
  - Services
  - Netzwerke
- Flexibilisierung von Diensten
  - Security Kapselung
  - Schneller Restart
  - Verschiebbar
  - Live Migration
- Modularisierung von Diensten
  - Jeder Dienst in einer eigenen Umgebung
  - Jeder Umgebung ihr eigenes Netzwerk



# Vom einfachen Netzwerk



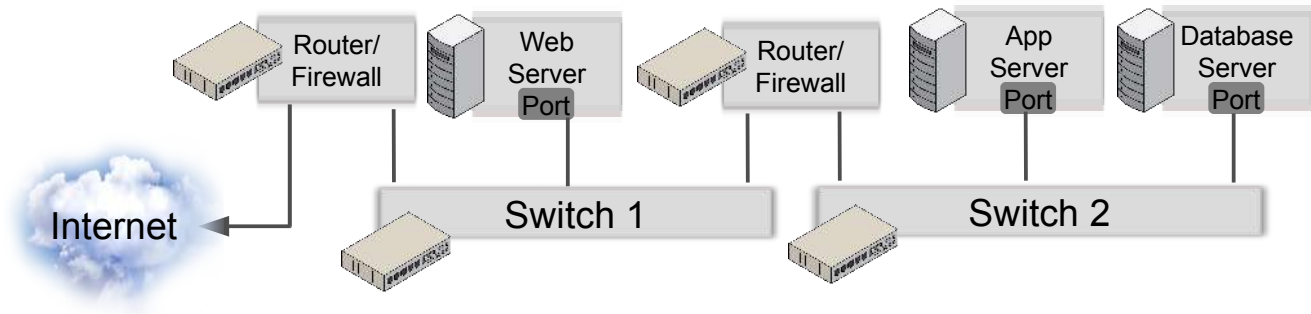
# ... zu immer mehr Komplexität



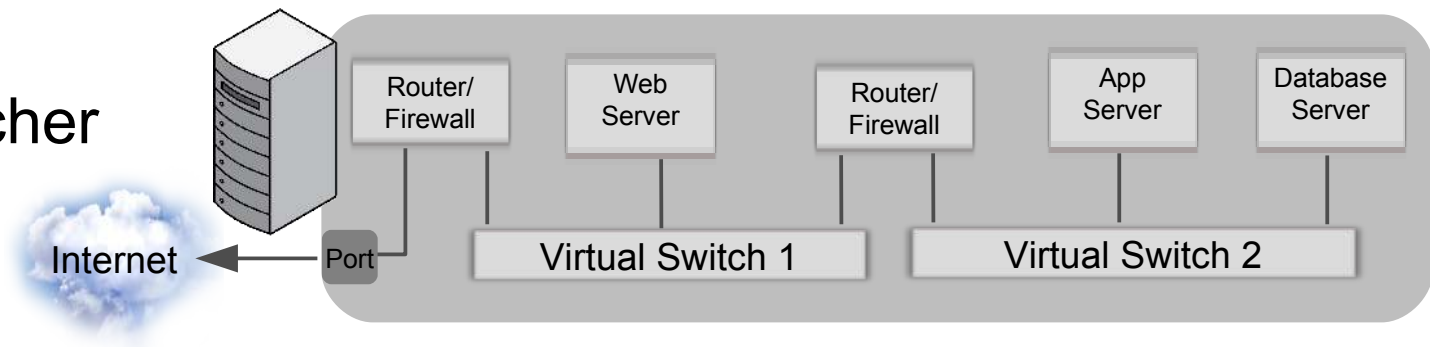
# Konsolidierung in Netzwerken

## Zielvorstellung

- Vorher



- Nacher



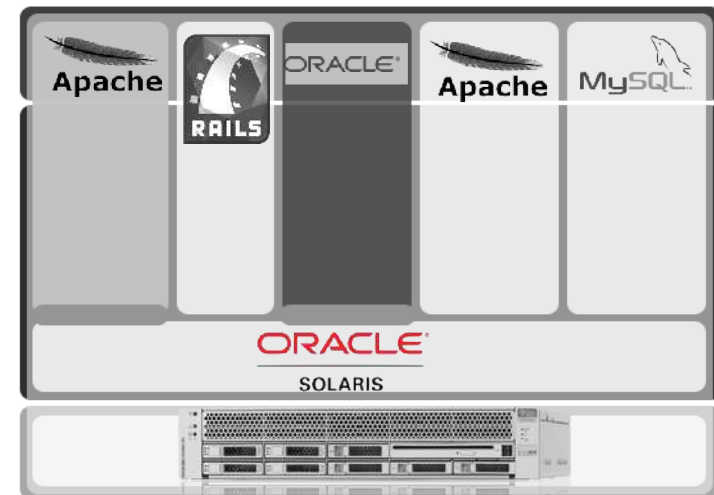
# Virtualisierung und Netzwerke

- Bildung virtueller Netzwerke durch tagged VLAN
  - Konfiguration in den Systemen und an den Netzwerkkomponenten
  - Welche Ausdehnung ist möglich ?
  - Welcher Aufwand bei Verschiebung/Migration ?
  - Abhängigkeit von der benutzten Netzwerk-Hardware ?
- Erzeugung virtueller Interfaces durch Hypervisor
  - Sind diese Interfaces transparent wie phys. Interfaces nutzbar ?
  - Gibt es separate Statistiken dafür ?
  - Wie ist die Ressourcenzuteilung zwischen den virtuellen Interfaces ?
  - Wo wird QoS realisiert ?
- ...

# Oracle Solaris Zones

Virtualisierung - enthalten in jedem Oracle Solaris

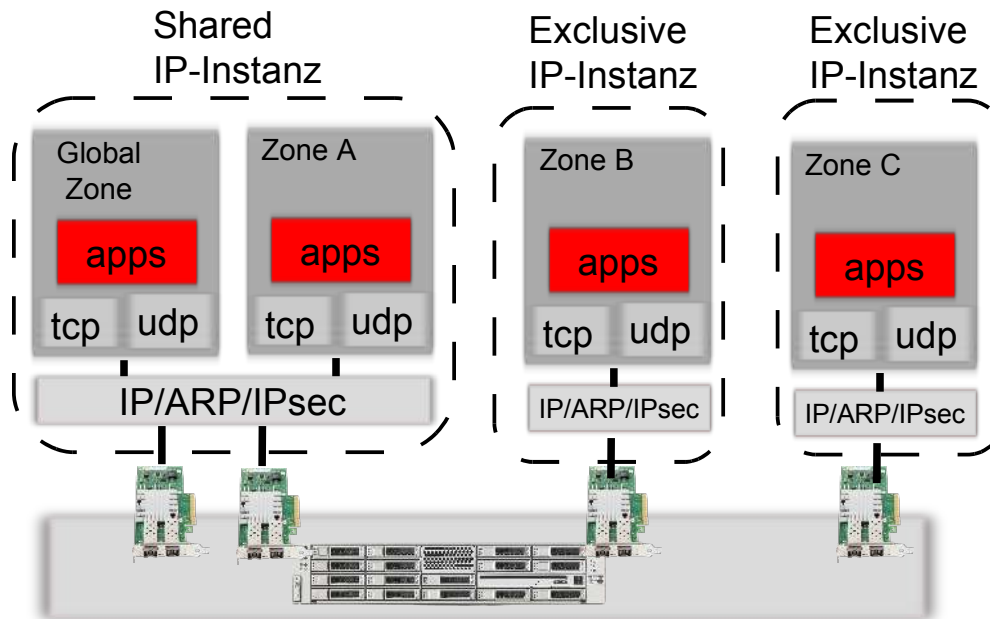
- Hocheffiziente, hochperformante OS-Virtualisierung
  - Tausende von Applikationen unabhängig auf einem System
- Virtualisierte Laufzeitumgebungen
  - Gemeinsamer Solaris Kernel
  - Gemeinsame Geräte
  - Separierte Filesysteme
  - Separierte Prozeßumgebungen
  - Dediziertes Management der Ressourcen
  - Gemeinsam genutzte oder separierte TCP/IP-Stacks mit eigenen Interfaces



# Netzwerkanbindung von Solaris Zonen

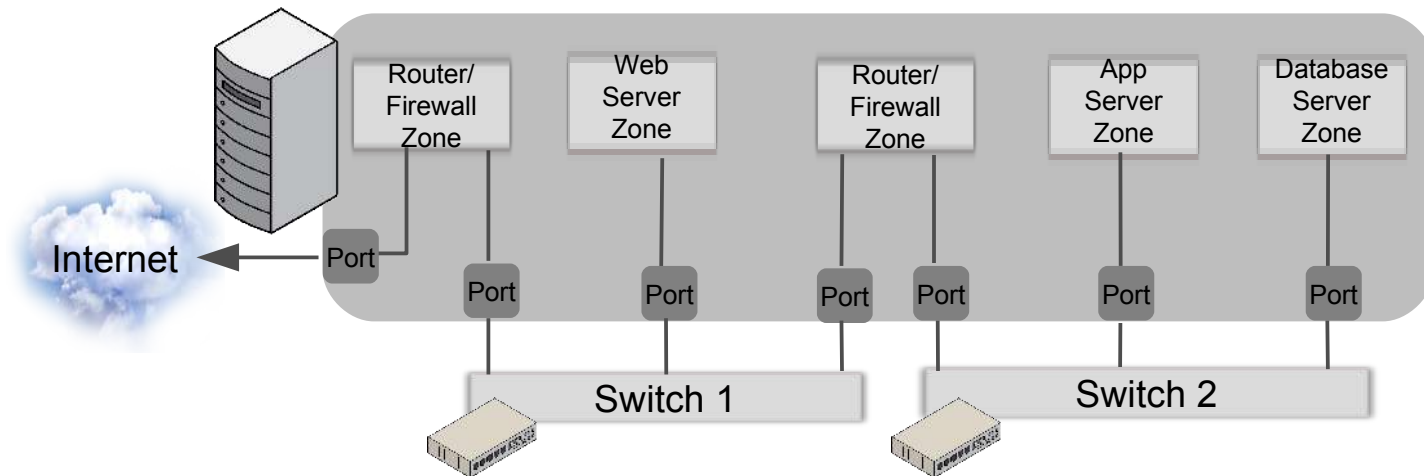
## Shared-IP und Exclusive-IP Instanzen

- Shared IP-Instanzen
  - Separierte Kommunikation
  - Gemeinsamer IP-Stack
  - Gemeinsame Interfaces
- Exclusive IP-Instanzen
  - Separierter IP-Stack
  - Separierte IP-Parameter
  - Exclusive Interfaces



# Zonen und Netzwerk-Konsolidierung

- Anwendung von Solaris Zonen für die Virtualisierung
- Shared IP-Instanzen verbinden alle Zonen über einen Stack mit separiertem Netzwerkverkehr
- Physische Separierung mit exklusiven IP-Instanzen
  - Benötigen exclusive Interfaces



# Netzwerkvirtualisierung in Oracle Solaris

- Als Projekt Crossbow in OpenSolaris
- Solaris 11 Express basiert auf Technologien in Solaris 10 und OpenSolaris
- Einzigartige Plattform für die Netzwerkvirtualisierung mit Ressourcenkontrolle und einfachem Management
  - Effektive gemeinsame Nutzung von Netzwerk Ressourcen
  - Aufteilung einzelner physischer Interface in viele virtuelle
  - Network-in-a-box
- Eingebettet in Oracle Solaris
  - Gesamtheitlicher Ansatz
    - Interface Hardware, Treiber, Interruptbehandlung, Kernel, verschiedene Protokollschichten, CPU-Nutzung



# Netzwerk-Virtualisierung

## Überblick

- Dynamisches Polling von Netzwerkkarten
- Konsequente Nutzung der H/W-Unterstützung moderner Netzwerkkarten
- NIC-Virtualisierung (vNIC) in beliebiger Anzahl
- Software-basierte Netzwerkschitches (Etherstub) und Bridges
- Klassifikation von Netzwerkverkehr in Flows
- Ressourcenzuteilung zu vNIC und Flows
- Ein einfaches Managementinterface

# Netzwerk-Virtualisierung in Oracle Solaris

## Anwendungsbeispiele

- Konsolidierung von Server aus umfangreichen Netzwerkkumgebungen
- Realisierung typischer Netzwerk-Funktionsgeräte auf beliebigen Solaris Server-Systemen
- Aufbau von Clouds innerhalb von Systemen
  - “Network-in-a-Box”

# Virtual NIC (VNIC)

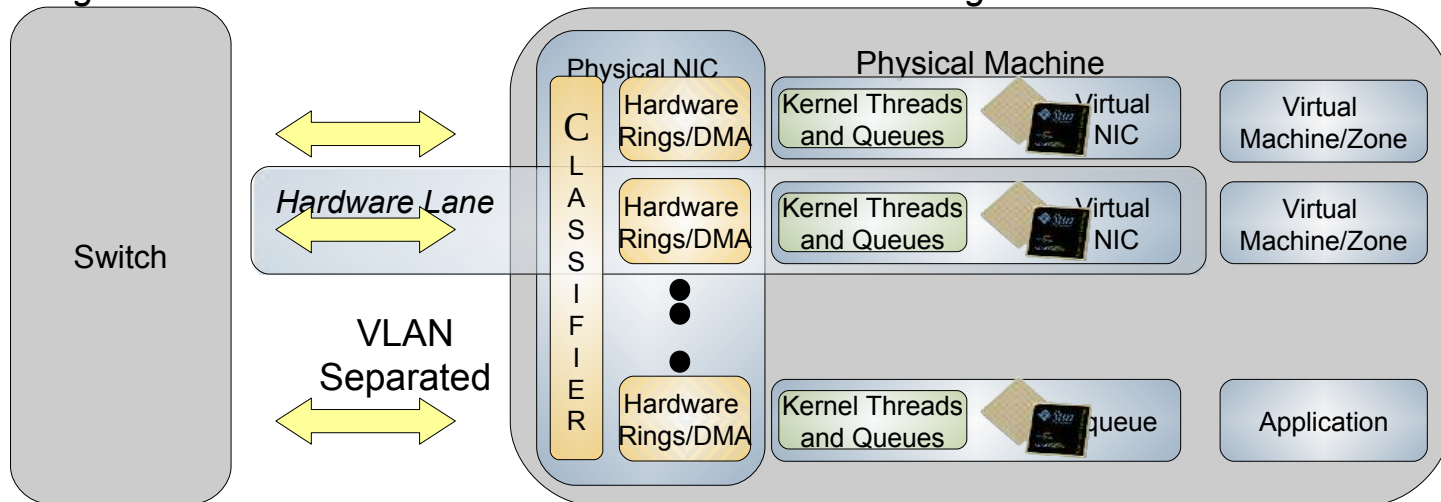
- Virtual NICs auf phys. Interfaces
  - Wie physikalische NICs nutzbar
    - Eigene MAC-Adresse
    - Eigene Kernel Statistik
  - Einstellbare Parameter
    - LinkSpeed, CPU, Priorität, VLAN Tags, Priority Flow Control (PFC)
- Eine Broadcast-Domain für alle VNIC auf einem Interface
- Nutzung von HW-Features der Netzwerkkarte wenn möglich
  - Weitere MAC-Adressen der Netzwerkkarte
  - HW-Rings

# Hardware Lanes

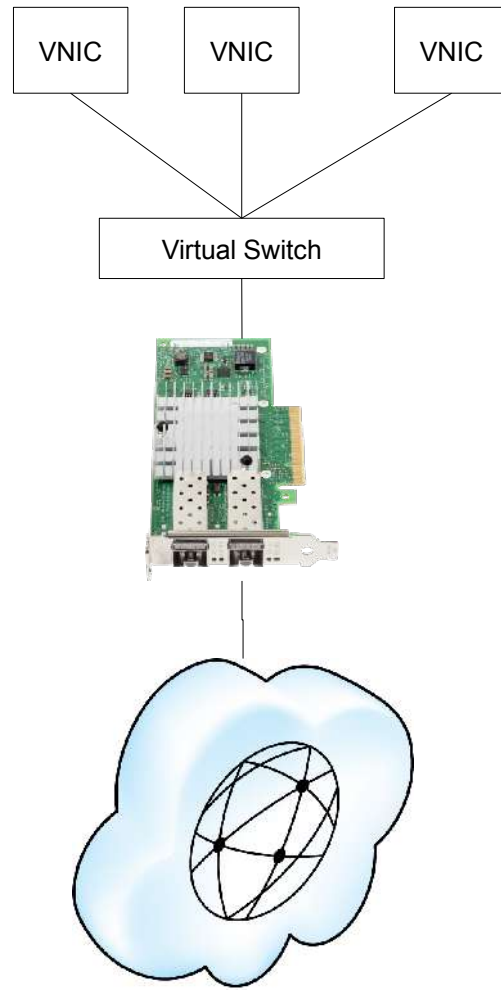
- Nutzung von HW auf modernen Netzkarten
  - Erhöht die Skalierbarkeit
  - Reduziert Aufwand der Systeme für Paketverarbeitung
- Anzeige der verfügbaren HW-Ringe der Netzkarten

# dladm show-phys -H

LINK	GROUP	GROUPTYPE	RINGS	CLIENTS
e1000g0	0	RX	1	e1000g0

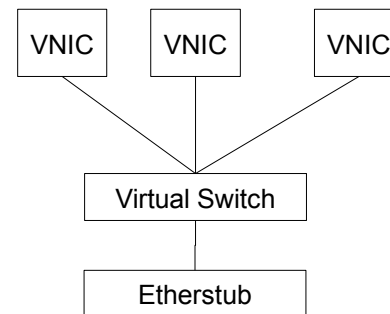
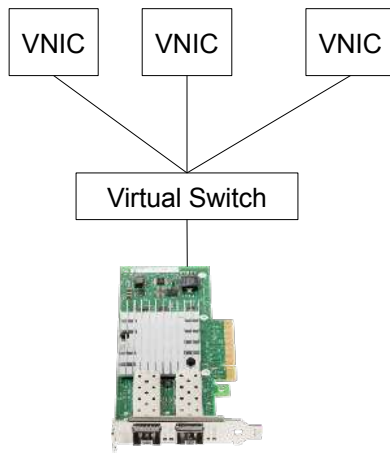


# VNIC und Virtual Switches



# Etherstub

- Virtual NICs auf pseudo Interfaces (Etherstub)
  - Ohne Existenz von phys. Interfaces nutzbar
  - VNIC auf phys. Interface ohne Unterschied zu VNIC mit Etherstub



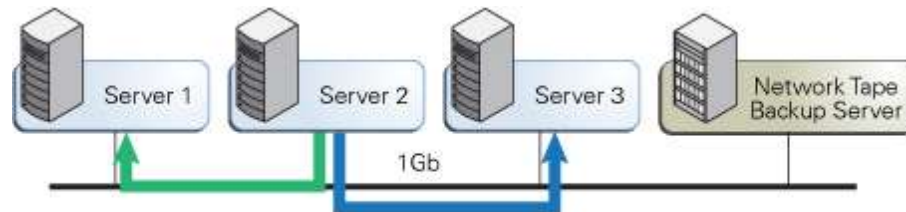
# Flows

- Teilmenge des Netzwerkverkehrs
  - Über Netzwerkport, Socket und Kernelthread
- Festlegung basierend auf
  - Services (protocol + remote/local ports)
  - Transport (TCP, UDP, SCTP, iSCSI, etc)
  - Remote und local IP addresses
  - Remote IP Subnets
- Attribute per flow setzbar
  - Unabhängig von verwendeter Hardware
  - Bandbreiten-Limits
  - Prioritäten
  - CPUs

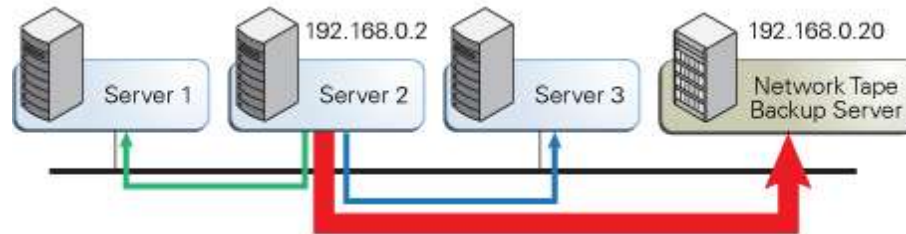
# Bandbreitensteuerung durch Flows

- Flows können Überlastsituationen in Netzwerken regeln oder DoS Attacken eindämmen

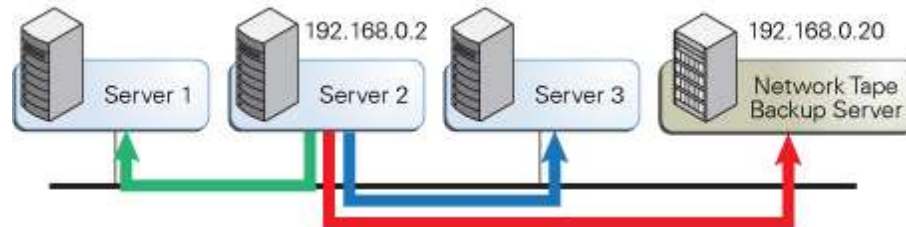
– Ausgangslage



– Überlast und Verschiebung der Last



– Regulierung durch Flow





# Managementinterface

- `dladm(1M)`
  - Erzeugen, Konfigurieren, Löschen von VNIC und Etherstub
- `flowadm(1M)`
  - Erzeugen, Konfigurieren, Löschen von Flows
- `dlstat(1M)`
  - Laufzeitstatistiken von VNIC
- `flowstat(1M)`
  - Laufzeitstatistiken von Flows
- `acctadm(1M)`
  - Extended Accounting Facility
  - U.a. zur Aufzeichnung von Netzwerkstatistiken zur späteren Auswertung nutzbar

# Management: VNIC

```
# dladm create-vnic -l e1000g0 vnic1
```

```
# dladm create-vnic -l e1000g0 -v 17 -m 0:1:2:3:4:5 \  
-p maxbw=2M -p cpus=4,5,6 vnic1_in_vlan17
```

```
# dladm show-vnic
```

<i>LINK</i>	<i>OVER</i>	<i>SPEED</i>	<i>MACADDRESS</i>	<i>MACADDRTYPE</i>	<i>VID</i>
<i>vnic1</i>	<i>e1000g0</i>	<i>1000</i>	<i>2:8:20:a9:7c:d2</i>	<i>random</i>	<i>0</i>
<i>vnic1_in_vlan17</i>	<i>e1000g0</i>	<i>2</i>	<i>0:1:2:3:4:5</i>	<i>fixed</i>	<i>17</i>

# Management: Etherstub

```
# dladm create-etherstub luftnummer1
```

```
# dladm create-vnic -l luftnummer1 luftnummer_vnic1
```

```
# dladm show-vnic
```

LINK	OVER	SPEED	MACADDRESS	MACADDRTYPE	VID
vnic1	e1000g0	1000	2:8:20:a9:7c:d2	random	0
vnic1_in_vlan17	e1000g0	2	0:1:2:3:4:5	fixed	17
luftnummer_vnic1	luftnummer1	0	2:8:20:13:c7:17	random	0

# Management: Flows

```
# flowadm add-flow -l e1000g0 -a transport=icmp \  
-p maxbw=1200K,priority=low icmp-flow
```

```
# flowadm show-flow
```

FLOW	LINK	IPADDR	PROTO	LPORT	RPORT	DSFLD
icmp-flow	e1000g0	--	icmp	--	--	--

# Anzeige von Laufzeitstatistiken

# flowstat -i 5

FLOW	IPKTS	RBYTES	IERRS	OPKTS	OBYTES	OERRS
icmp-flow	11	1,08K	0	11	1,08K	0
icmp-flow	5	490	0	5	490	0

# dlstat e1000g0 -i 5

LINK	IPKTS	RBYTES	OPKTS	OBYTES
e1000g0	41,51K	2,93M	2,27K	195,97K
e1000g0	1	60	1	42

# Aufzeichnung und Anzeige von Statistiken

- Aktivierung der Aufzeichnung

```
# acctadm -e extended -f /var/tmp/netlog.acct net
```

- Auswertung der Aufzeichnung

```
# dlstat show-link -h -f netlog.acct e1000g0
```

LINK	START	END	RBYTES	OBYTES	BANDWIDTH
e1000g0	08:11:24	08:11:44	0	42	0 Mbps
e1000g0	08:11:44	08:12:04	326	42	0 Mbps

```
# flowstat show-link -h -f /tmp/netlog icmp-flow
```

FLOW	START	END	RBYTES	OBYTES	BANDWIDTH
icmp-flow	08:11:24	08:11:44	1960	1960	0.001 Mbps
icmp-flow	08:11:44	08:12:04	1960	1960	0.001 Mbps

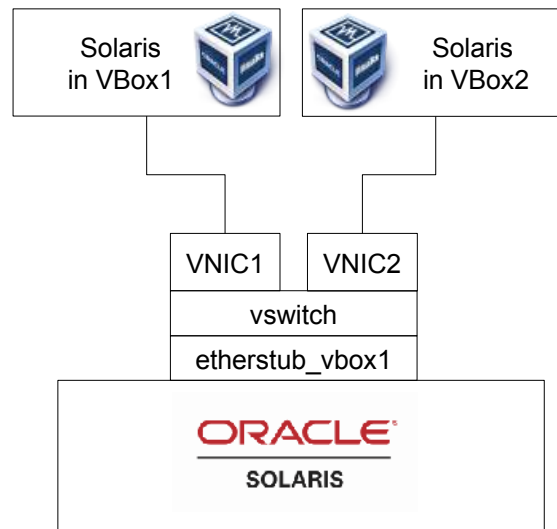
- Deaktivierung der Aufzeichnung

```
# acctadm -x net
```

# Anwendungsbeispiel

## VNIC und VirtualBox

- Zuweisung von VNIC zu VirtualBox Instanzen
- Separierung der Kommunikation durch Verwendung von Etherstubs



# Anwendungsbeispiel

VNIC und VirtualBox - Setup VNIC und Etherstub

```
# dladm create-etherstub etherstub_vbox1
# dladm create-vnic -l etherstub_vnic1 vnic1
# dladm create-vnic -l etherstub_vnic1 vnic2
# dladm show-vnic
```

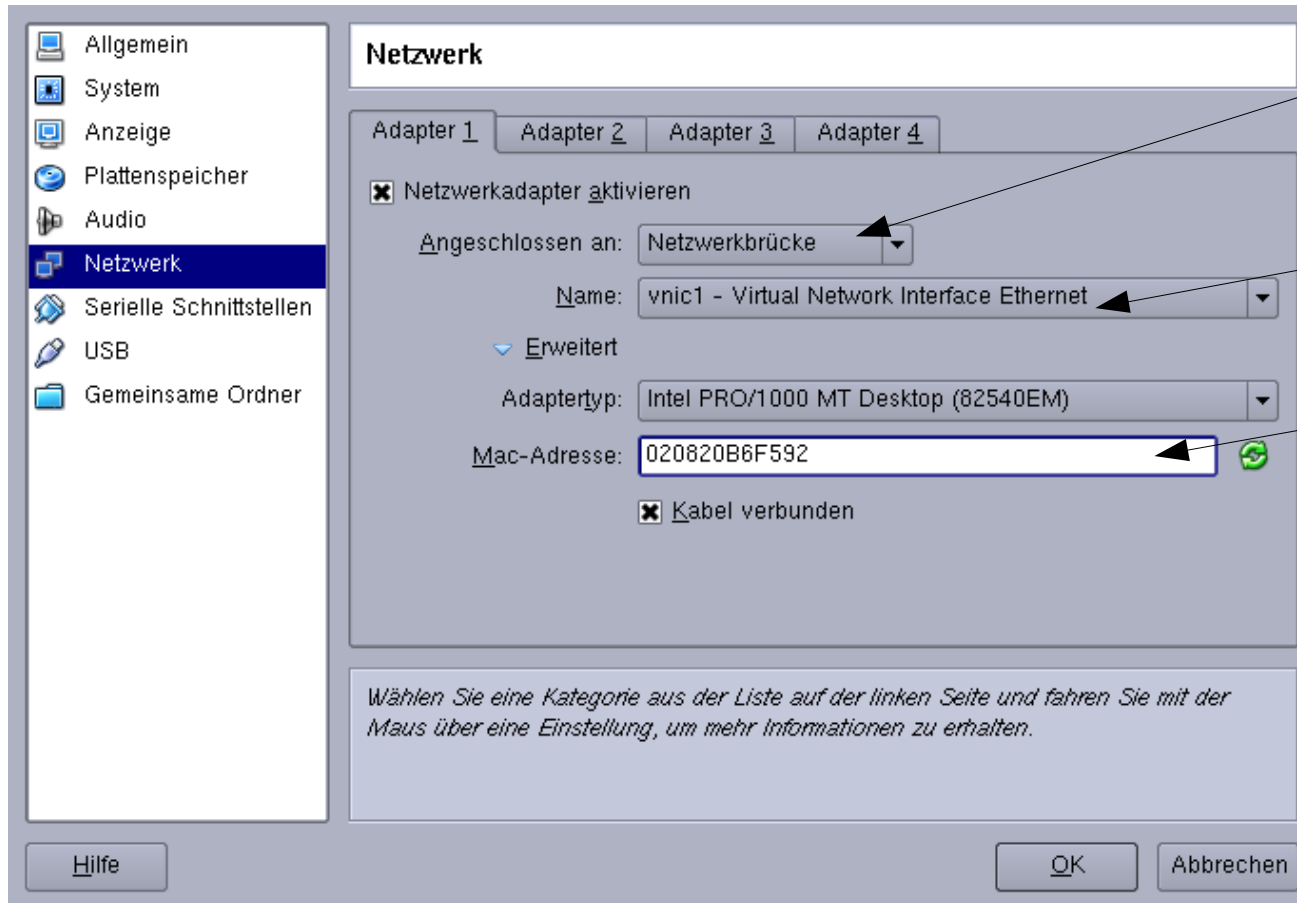
LINK	OVER	SPEED	MACADDRESS	MACADDRTYPE	VID
vnic1	etherstub_vnox1	0	2:8:20:b6:f5:92	random	0
vnic2	etherstub_vbox1	0	2:8:20:8e:29:5f	random	0

- Innerhalb der VirtualBox werden die VNIC wie phys. Interfaces erkannt (z.B. als e1000g0)



# Anwendungsbeispiel

## VNIC und VirtualBox - Setup von VirtualBox



Netzwerkbrücke auswählen

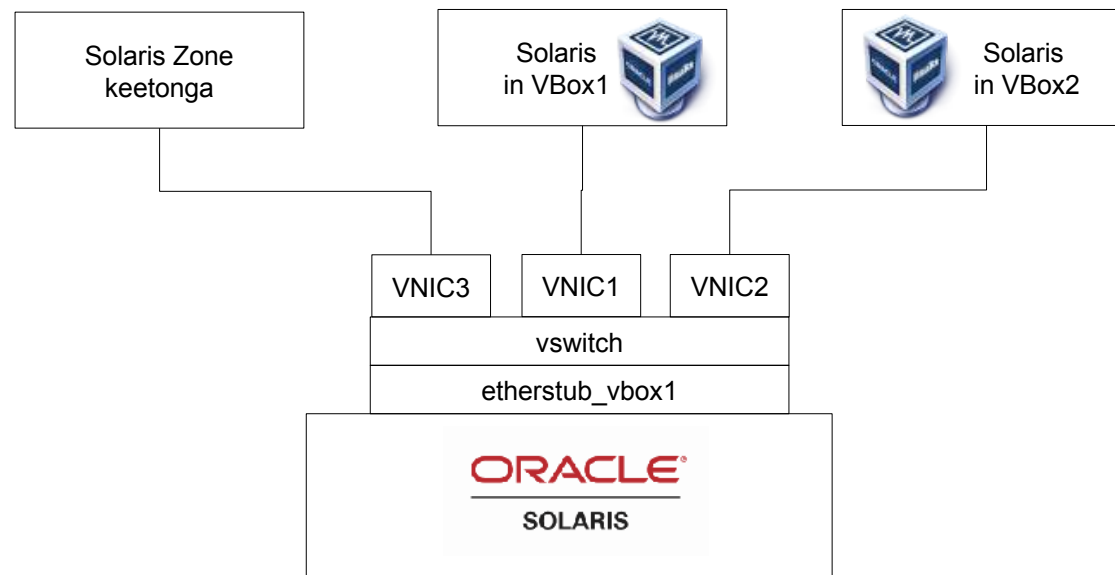
VNIC auswählen

MAC-Adresse der VNIC eintragen

# Anwendungsbeispiel

## VNIC und Oracle Solaris Zonen mit Exklusiver IP-Instanz

- Nutzung von VNIC als exklusives Interface
- Zone mit exklusiver IP-Instanz
- Separierte Kommunikation der Zone mit Instanzen von VirtualBox



# Anwendungsbeispiel

## VNIC und Oracle Solaris Zonen - Setup

```
# dladm create-vnic -l etherstub_vbox1 vnic3
```

```
# dladm show-vnic vnic3
```

LINK	OVER	SPEED	MACADDRESS	MACADDRTYPE	VID
vnic3	etherstub_vbox1	0	2:8:20:2b:f5:8e	random	0

```
# zonecfg -z keetonga info
```

```
ip-type: exclusive
```

```
net: physical: vnic3
```

```
keetonga # ifconfig vnic3 plumb
```

```
keetonga# ifconfig vnic3 192.168.100.3 up
```

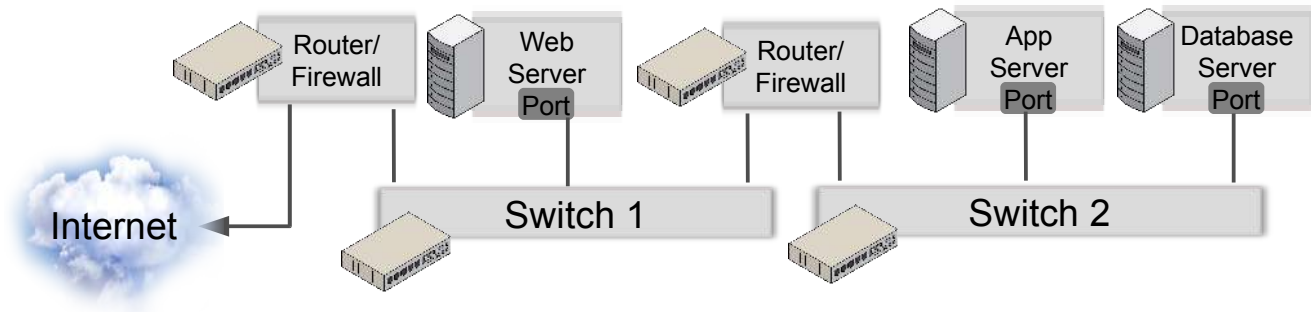
```
keetonga # ifconfig vnic3
```

```
vnic3: flags=1000843<UP,BROADCAST,RUNNING,MULTICAST,IPv4> mtu  
9000 index 2 inet 192.168.100.3 netmask ffffffff broadcast  
192.168.100.255 ether 2:8:20:2b:f5:8e
```

# Konsolidierung in Netzwerken

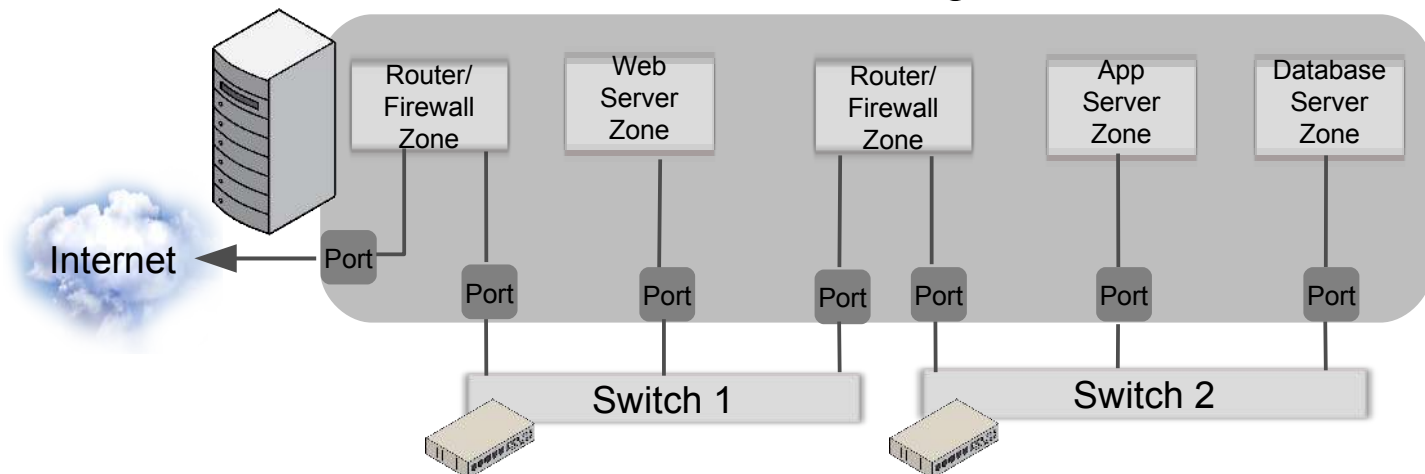
## Weiteres Beispiel

- Vorher



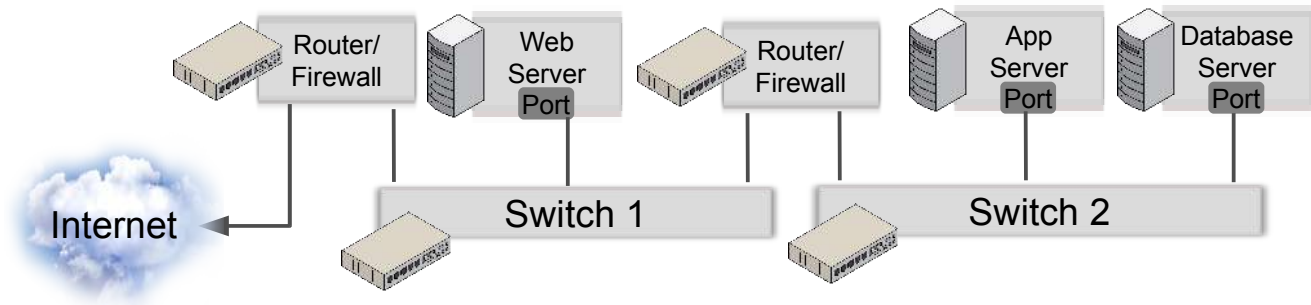
- Umsetzung mit Solaris Zones

- U.U. viele exclusive Interfaces notwendig

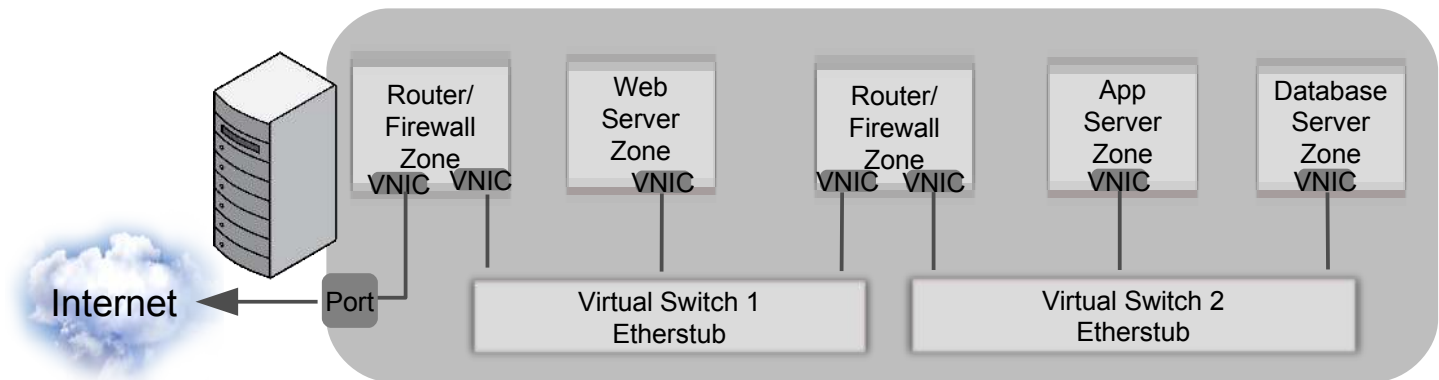


# Netzwerk-Konsolidierung mit Solaris

- Ausgangslage



- Nutzung von VNIC, Etherstub und Oracle Solaris Zonen



# Zusammenfassung

- Hohe Performance, Lineare Skalierung
  - Hardware-Lanes
  - Dynamic Polling
- Virtual NICs und Virtual Switching
- QoS überarbeitet
  - L3/L4 Flows, einfach zu konfigurieren
- Resource Partitionierung & DoS Abwehr
- Anwendungen
  - Network-in-a-Box
  - Konsolidierung von Netzwerken und Netzwerkgeräten
  - Aufbau von Netzwerken “in Software”

# Weitere Informationen

- Projekt Crossbow:  
<http://hub.opensolaris.org/bin/view/Project+crossbow/>
- Usenix Paper  
[http://www.usenix.org/events/lisa09/tech/full\\_papers/tripathi.pdf](http://www.usenix.org/events/lisa09/tech/full_papers/tripathi.pdf)
- Solaris 11 Express Dokumentation

# Q & A







# Hardware and Software

ORACLE

# Engineered to Work Together

**ORACLE®**