



Integriert denken

Storage – Virtualization - Clustering

OSL Technologietage
Berlin • 12./13. September 2012

Bert Miemietz

OSL Gesellschaft für
offene Systemlösungen mbH

12. 9.

- 10 Jahre OSL
- Virtualisierung mit Oracle SPARC T4-Systemen (Detlef Drewanz - Oracle)
- SUSE Linux Enterprise Server (Ralf Dannert - SUSE)
- OSL Storage Cluster 4.0 und RSIO (Bert Miemietz – OSL)

Schiff Ahoi – ca. 16:30 Uhr

Hausaufgaben

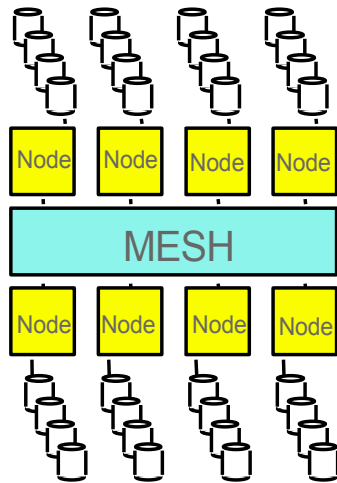
13. 9.

- Solaris 11 – Best Practices (Detlef Drewanz - Oracle)
- RSIO - Tutorial (Oliver Härtel - OSL)
- Linux: Storage-Attachement und HV in 5 Minuten (Christian Schmidt - OSL)
- High-Speed I/O mit SSD (Andre Hirschberg – Cristie Data Products)

10 Jahre OSL
Auf dem Weg zu
ganzheitlichen RZ-Konzeptionen

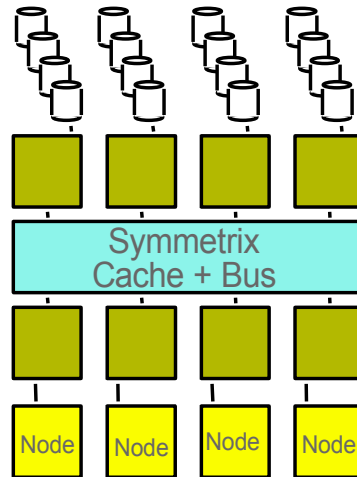
So fing alles an ...

1994



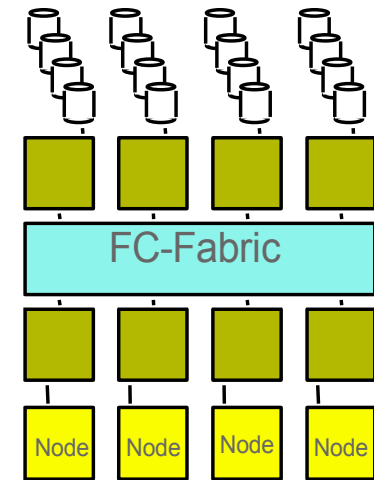
RM1000 Massive Parallel
Computer with integrated
shared storage

1998



RM600 Clusters with
shared Symmetrix storage
system and DiskView

2000



Starting RM and
PrimePower **sharing**
storage over SAN

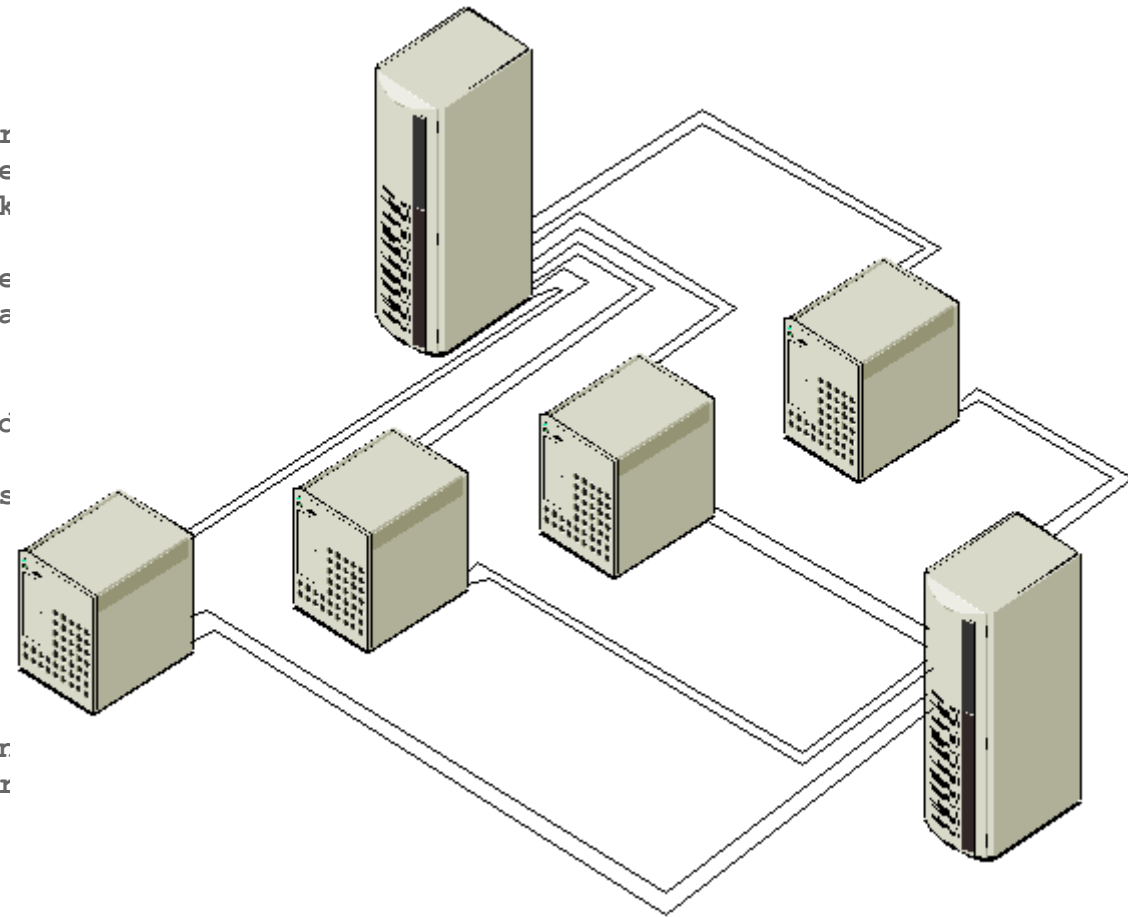
- *Start als Migrationslösung RM1000 -> RM600-Cluster*
- *von Anfang an Fokussierung auf Shared Storage*
- *erste Versionen als Toolkit ohne Lizenzgebühren*
- *zunächst reine Aufsatzlösung für DC/OSx (Pyramid) bzw. Reliant Unix (Siemens Nixdorf)*

So fing alles an ...

[1] Package History

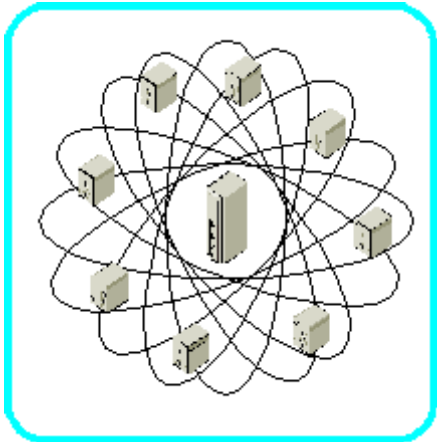
DISKVIEW SMART CLUSTER 1.0 Ein Shared-Storage Cluster

Version	Load	Functions/Changes
0.5	10	Base functions, operable.
0.6	10	Added fsadmin command. Improved r check for configuration mismatches vdisks. They now can handle vdisk than 6 chars without problems.
11		First version of dvscdocs package
14		Includes UCB-compatible online ma
14		Additional manpages
0.7	10	Added /etc/clustername workaround
		Additional manpages. New command
11		extensions to dkadmin and dkblocs
12		previous consistency check for
		application vdisk config files
14		implementation of cluster root
18		New licence procedure,
		some minor changes
19		Included 5.43 workaround for
20		disks > 4GB, enhancement to vdcor
20		moved variable files to /dvsc/var
		dktab.user is no longer used!
21		some minor enhancements
0.8	10	included beta of node state monitoring for facultative
		testing by customers. dvbcd has to be started manually.
		See dvbcd(1) and cfadm(1) for details.
12		some minor changes to ndinfo
13		fixed vdconfig -e problems with concat disks



und noch unter Reliant Unix ging es weiter ...

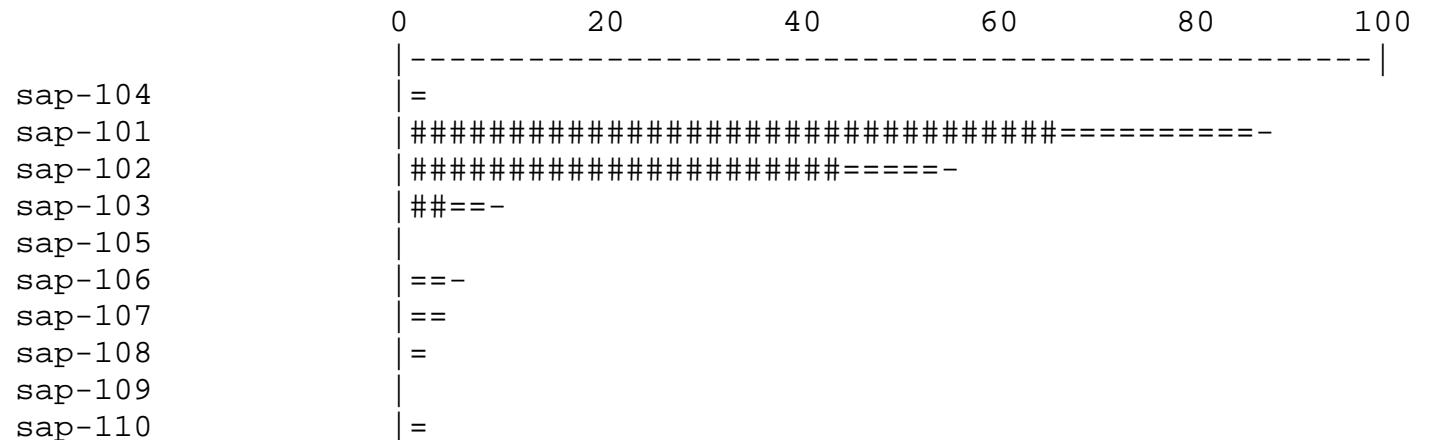
- 1998 1.1.
- Failover Option
 - aggregierte, clusterweite Performancedaten
 - Filesystem-Administration



Example CPU data graph

DiskView Smart Cluster performance graph: cpu and memory usage
=====

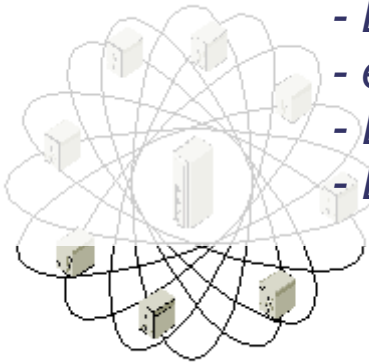
user mode, = system mode, - wait for io



und noch unter Reliant Unix ging es weiter ...

- 1998 1.1.
- einfacher Disk Access Manager (Userspace)
 - Failover Option
 - aggregierte, clusterweite Performancedaten
 - Filesystem-Administration

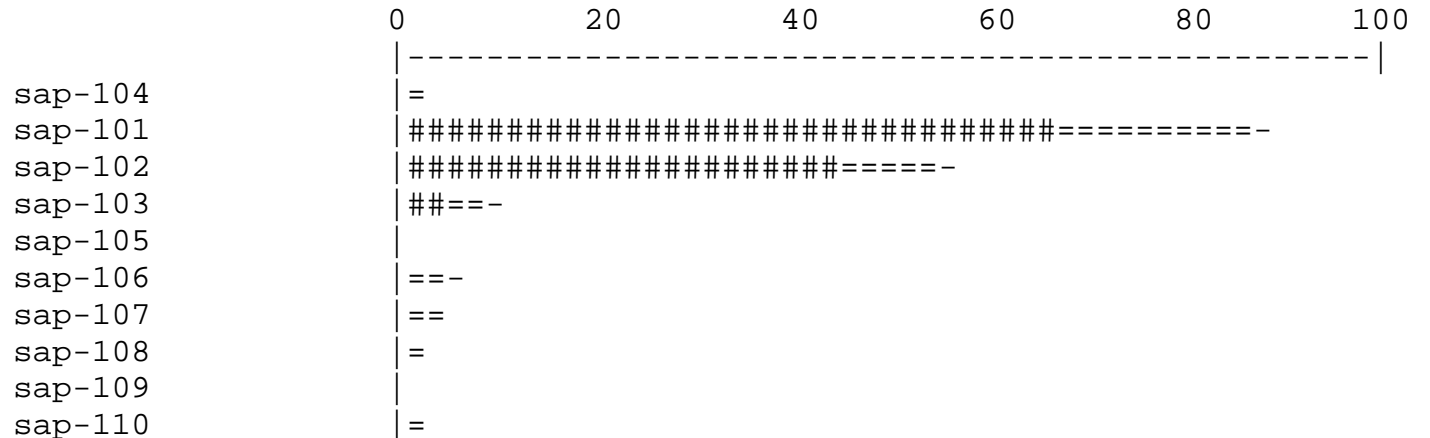
- 1.2
- Universe Extensions (heute: XDM)
 - Lösungen für Timefinder und SRDF
 - erstmals Cluster-Command
 - Einbindung Non-Symmetrix-Systeme
 - Device Classes (Storage Classes)



Example CPU data graph

Example graph: cpu and memory usage
=====

user mode, = system mode, - wait for io

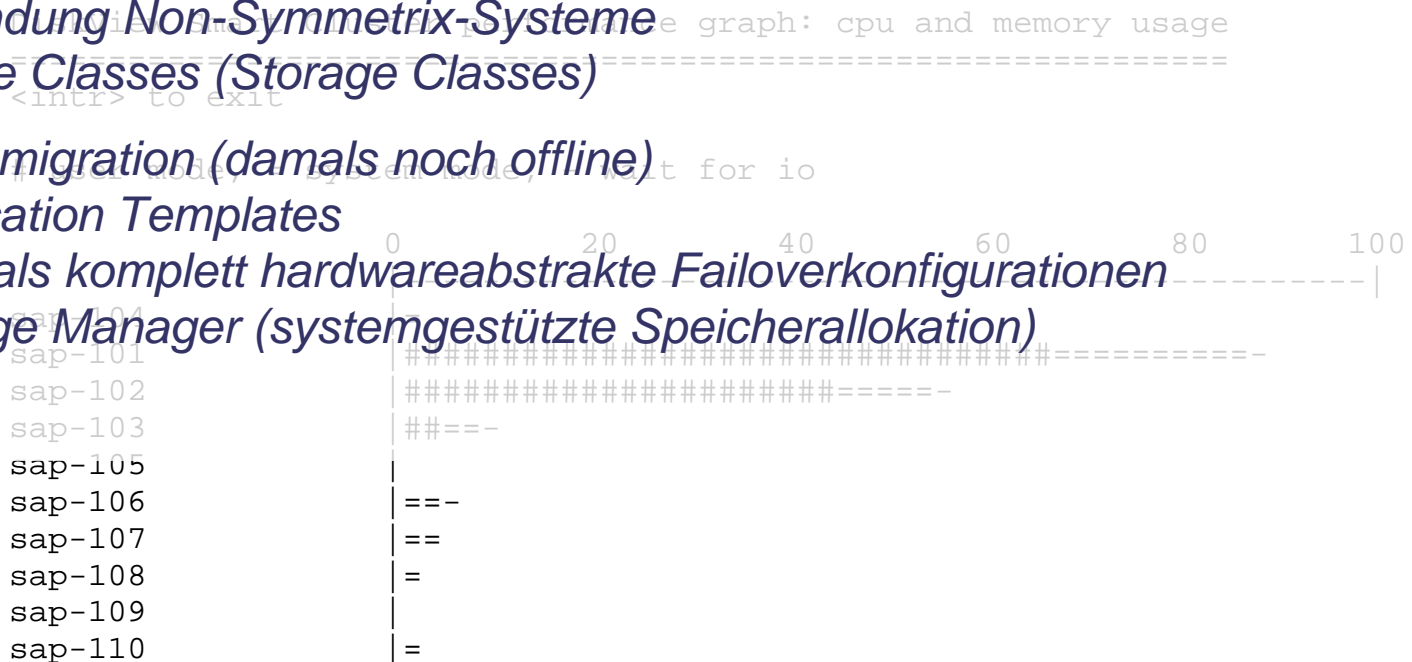


und noch unter Reliant Unix ging es weiter ...

- 1998 1.1.
- einfacher Disk Access Manager (Userspace)
 - Failover Option
 - aggregierte, clusterweite Performancedaten
 - Filesystem-Administration

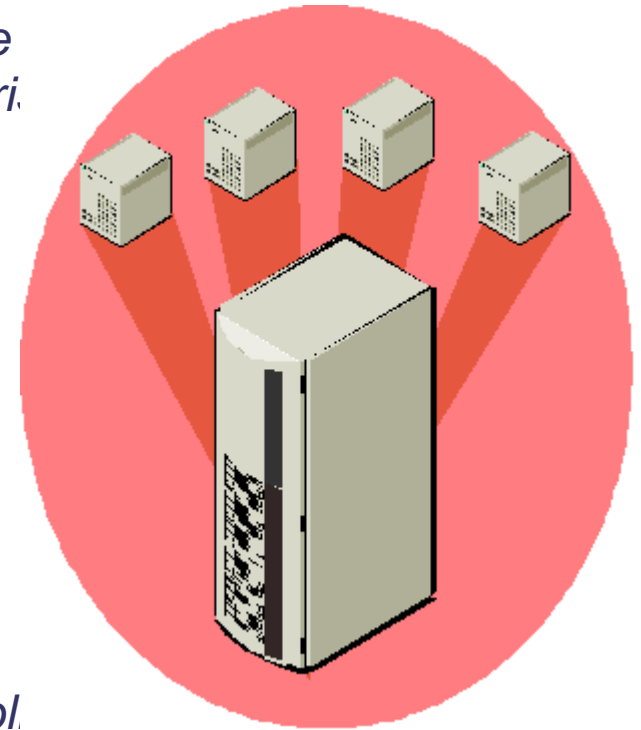
- 1.2
- Universe Extensions (heute: XDM)
 - Lösungen für Timefinder und SRDF
 - erstmals Cluster-Command
 - Einbindung Non-Symmetrix-Systeme
 - Device Classes (Storage Classes)

- 2000 1.3
- Datenmigration (damals noch offline)
 - Application Templates
 - erstmals komplett hardwareabstrakte Failoverkonfigurationen
 - Storage Manager (systemgestützte Speicherallokation)



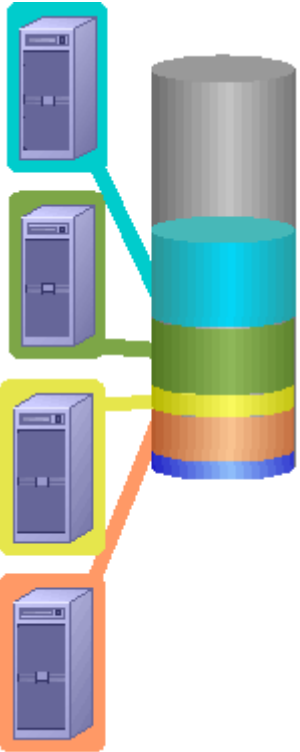
2001 - Start in die Solaris-Welt mit eigenem Treiber

- 2001
- Virtual Volumes (eigener Treiber, Handhabung ähnlich RUX Vdisk)
 - simple, concat, stripe
- 2002 2.0
- DiskView Storage Cluster 2.0 für Solaris
 - von Anfang an mit eigenem Multipfad (eine der ersten offenen Lösungen!)
 - Clonen ganzer Datenbestände per BCV mit m2slock
- 2002
- ## Gründung der OSL Gesellschaft für offene Systemlösungen mbH
- 2.1
- Erste Storage-Cluster-Version unter OSL-Regie
 - Wiedereinführung der HA (nunmehr unter Solari.
 - Dynamic Node Lists (Selbstkonfiguration)
 - Integration von Sun-Hardware (Vxxx ...)
- 2003 2.2
- Raw Device Sharing
 - Devices online vergrößern
 - 100 kleine Verbesserungen
- 2005 3.0
- Host-based Mirroring zwischen Universen
 - Clonen von Volumes
 - Live Data Migration (Device Mobility)
 - Priorisierung Anwendungs-I/O
 - Extended Volume Controls (XVC) & "Atomic Spl.



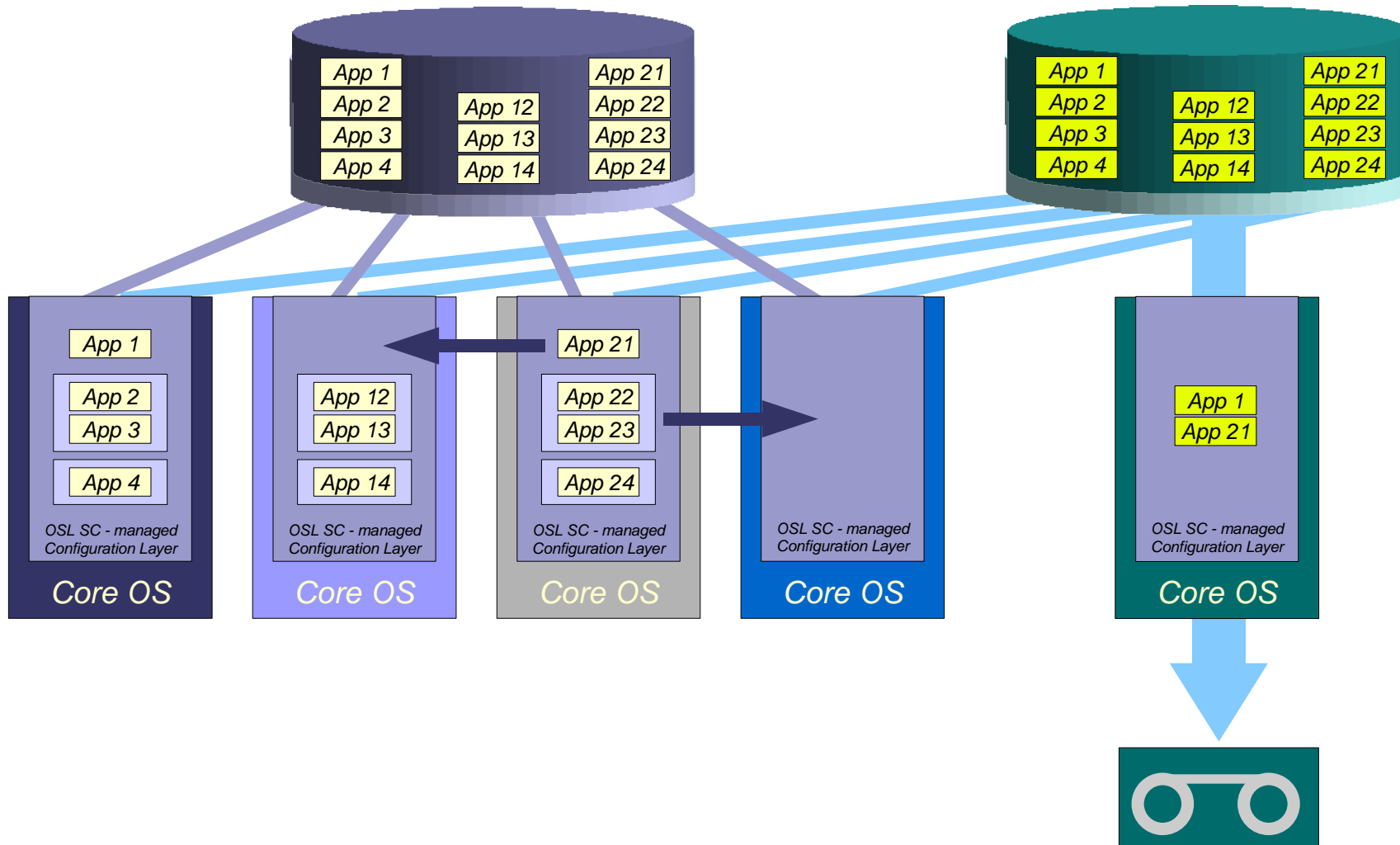
2006 – Orientierung auf ein integriertes Paket

- 2006 - Erste Überlegungen zu RSIO (Virtual Storage über Ethernet)
- 2008 3.1 - Verknüpfung von Speichervirtualisierung, HA-Environments und OS-Administration zu einem integrierten Paket
- > Speichervirtualisierung mit Applikationsbezug
 - clusterweites Benutzer- und Gruppenmanagement
 - neues XVC-Konzept mit weiteren nützlichen Triggern und Bandbreitensteuerung
 - applikationsbezogene Bandbreitensteuerung
 - lineare und integrierte Volumetypen
 - Solaris 10: Integration ZFS und Zonen
 - komplette Lösungen für Backup und Disaster Recovery
 - > einzigartige Kombination aus synchroner und asynchroner Spiegelung und damit Schutz auch gegen Softwarefehler!
 - restorefreies Instant Recovery (Riesendatenbanken in wenigen Minuten!)
 - Lösungen für Systemkopien z. B. für Oracle-DB + SAP
 - integrierte Backup-/Restore-Lösungen für Informix, Oracle ...
 - Musterlösungen für Postgres-SQL
 - clusterweites LAN-less Backup und Redolog-Archivierung
 - Integrationslösungen mit Networker und Hypertape
 - DB-abstraktes Recovery-Interface mit Point-in-Time-Fähigkeiten (dvamrestore, appdbroll)



Der Stand mit OSL SC 3.1

Dynamische, flexible und hochverfügbare RZ-Infrastrukturen

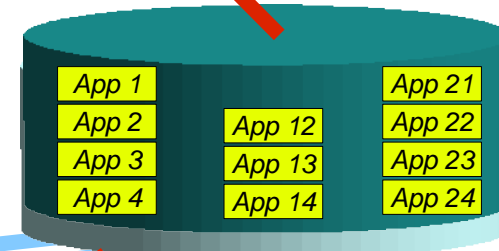
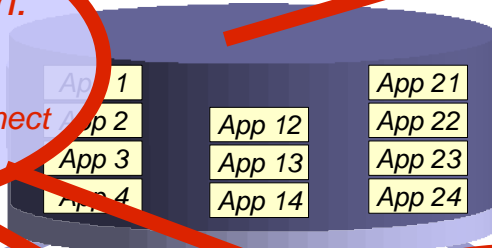


Der Stand mit OSL SC 3.1

Dynamische, flexible und hochverfügbare RZ-Infrastrukturen

Trennung I/O OLTP <-> Backup:
spezialisierte Systeme ersetzen Universalsysteme
homogenes Lastprofil

HW-Einsparungen:
weniger Controller
weniger Kabel
kein dedizierter Interconnect
kein Split Brain



Highspeed Tape Backup
zentralisiert, über Spezialmaschine
getrennt von
Produktionssystemen

Backup to Disk:
sehr hoher Durchsatz
zentrale Steuerung
inkrementell
Recovery ohne Restore

zentrale Administration
jede Applikation überall starten
ressourcenbasierte Optimierung
zentrales Monitoring
Hochverfügbarkeit
HW- und OS-Abstraktion
Systemkopien

geeignet für Disaster Recovery:
mit oder ohne autonome RAID-Spiegel
integriert existierende RAID-Spiegel
Restart von beliebigem RAID-System
Restart auf beliebigem Host

volle Transaktionsabsicherung:
Kombination mit permanenten Spiegeln
Restart vom Disk-Backup (Spiegel)
Roll forward (full / until time / until cancel)

Auf dem Weg zur ganzheitlichen RZ-Lösung

- 2009 - Erste Vorstellung von RSIO (Solaris)
- 2010 - Vorstellung von RSIO für Linux auf der SNWE
- 2011
 - Erste Demonstrationen Linux-Cluster und XDM via RSIO
 - Erarbeitung Vorzugskonfigurationen mit Cosifan und etomer
 - Rückübernahme des Service in die eigene Verantwortung (7x24h)
 - Umstellung Fernwartung auf OSL Shared Terminal (Eigenentwicklung)
- 2012 4.0
 - Netzwerkkonvergenz SAN-Ethernet
 - flexible Bausteine für eine dynamische RZ-Umgebung
 - Integration von Solaris und Linux einschl. HA
 - Virtual Nodes
 - Integration verschiedener OS-Virtualisierungstechnologien
 - Erhöhung Knoten- und Volumenzahl

Parallel: Entwicklung des Unternehmens

- stabile wirtschaftliche Entwicklung, Bewahrung Entscheidungsfreiräume für konsequente Kunden- und Lösungsorientierung
- 10. 8. 2012 – offizielle Eröffnung neuer Standort
Weichenstellung für:
 - deutlich verbesserte, professionelle Arbeitsbedingungen
 - Workshops, Schulungen, Expertentreffen mit praxisnaher technischer Ausstattung
 - behutsames personelles Wachstum



Hier stehen wir heute

Ein modulares, aber integriertes Gesamtsystem

<p><i>Application Awareness</i></p> <p><i>Bandbreitensteuerung</i></p> <p><i>User-Management</i></p>	<p>Application Control Option</p>	<p><i>Applikationen spiegeln</i></p> <p><i>Applikationen klonen</i></p> <p><i>B2D / DASI / DR -Werkzeuge</i></p>
	<p><i>clusterweite Steuerung von Applikationen</i></p>	
	<p><i>virtualisierte (hardwareabstrakte) Ablaufumgebungen</i></p>	
	<p><i>Hochverfügbarkeit</i></p>	
	<p><i>ressourcenbasiertes Selbstmanagement</i></p>	
	<p>Application Resource Description</p>	

<p>Cluster- Speichervirtualisierung</p>	<p>RSIO</p>	<p>Extended Data Management</p>
<p><i>globale Storage-Pools (rechnerübergreifend)</i></p>	<p><i>Datacenter Block-I/O z. B. via Ethernet</i></p>	<p><i>Integration RAID-basierter Datenkopien & Snapshots</i></p>
<p><i>globales Disk-Inventory</i></p>	<p><i>Integration mit Speichervirtualisierung</i></p>	<p><i>hostbasierte Spiegelung</i></p>
<p><i>globale Geräte / globaler Namensraum</i></p>	<p><i>Multipathing + Trunking</i></p>	<p><i>online Datenmigration</i></p>
<p><i>Clustervol-Manager/automatische Allokation</i></p>	<p><i>Cross-Platform</i></p>	<p><i>Daten klonen</i></p>
<p><i>Disk Access Management</i></p>		
<p><i>I/O-Multipathing</i></p>		

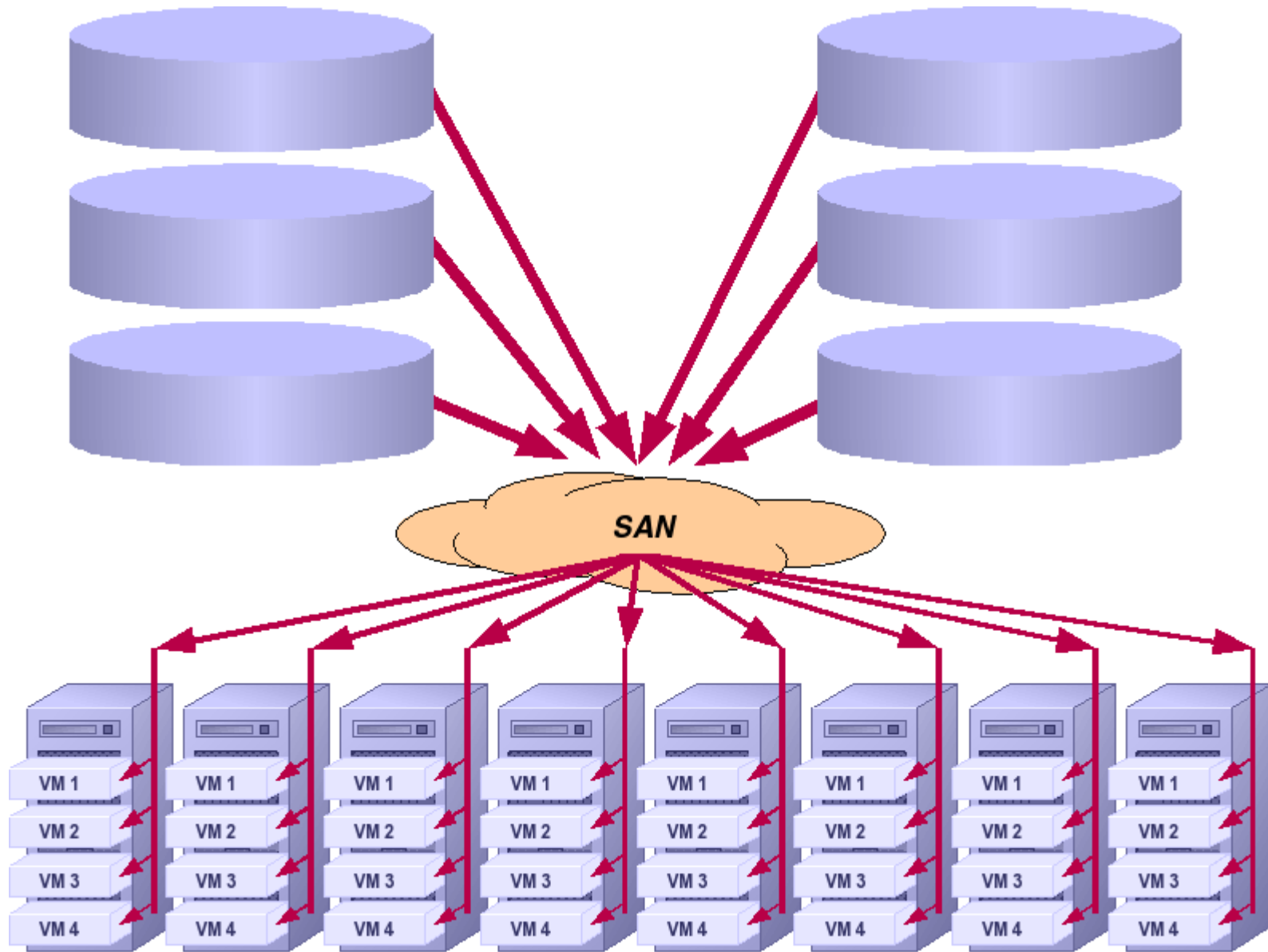
Hier stehen wir heute

Ein modulares, aber integriertes Gesamtsystem



Herausforderungen: 1. Virtuelle Maschinen

Anders strukturierte Datenströme / andere Prioritäten



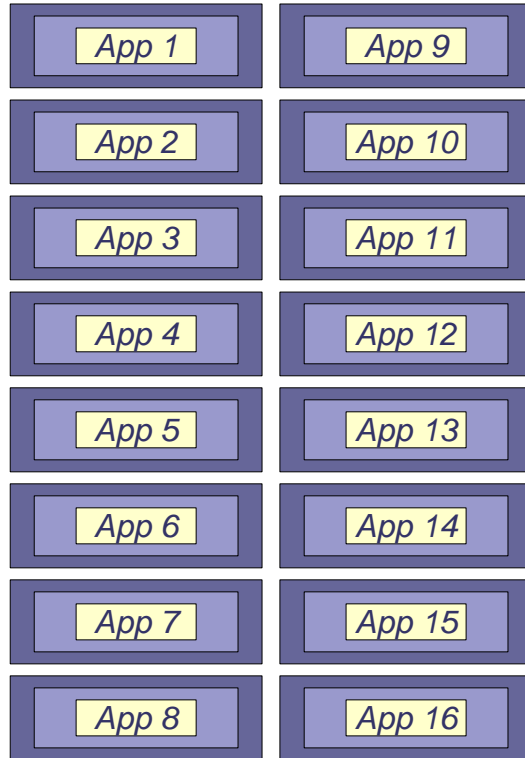
OSL Gesellschaft für offene Systemlösungen mbH

www.osl.eu

Apropos virtuelle Maschinen

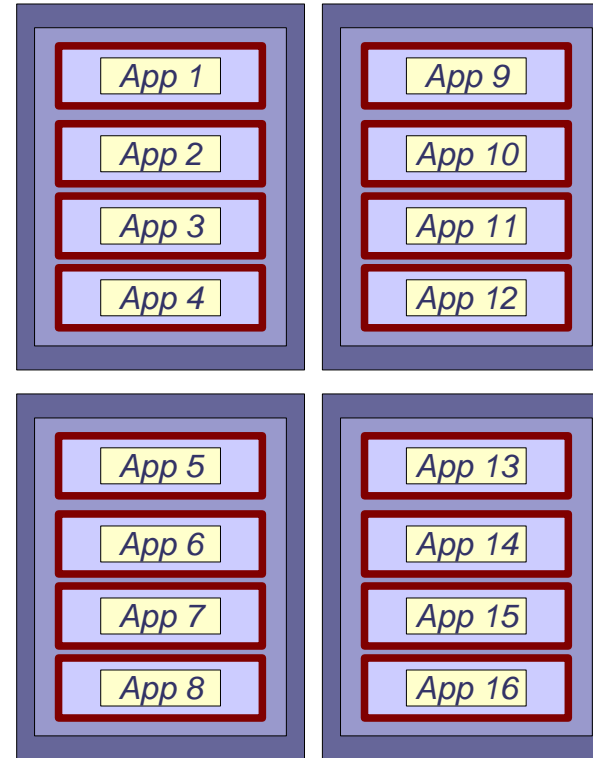
Mit OSL sind Sie hier schon weiter

ohne Virtualisierung



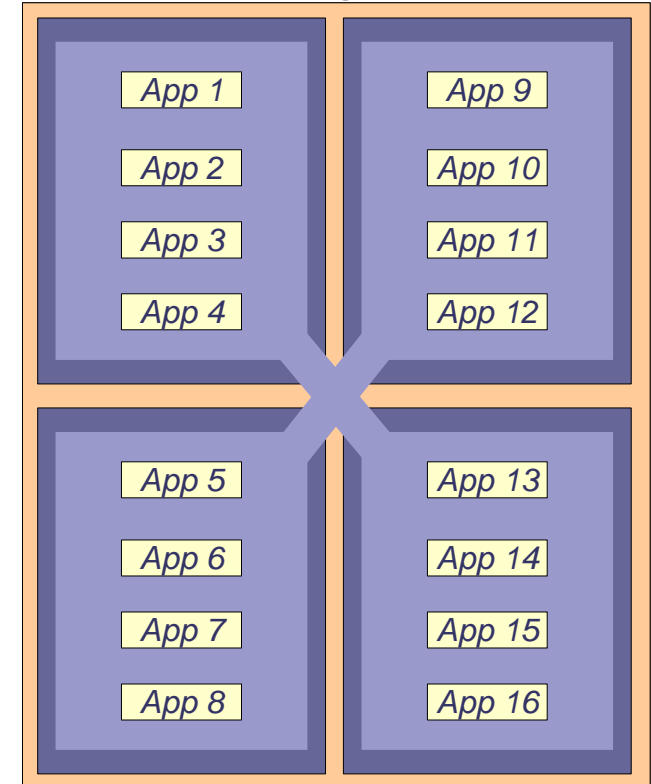
32 Softwareobjekte

mit virtuellen Maschinen



36 Softwareobjekte

OSL Storage Cluster



21 Softwareobjekte

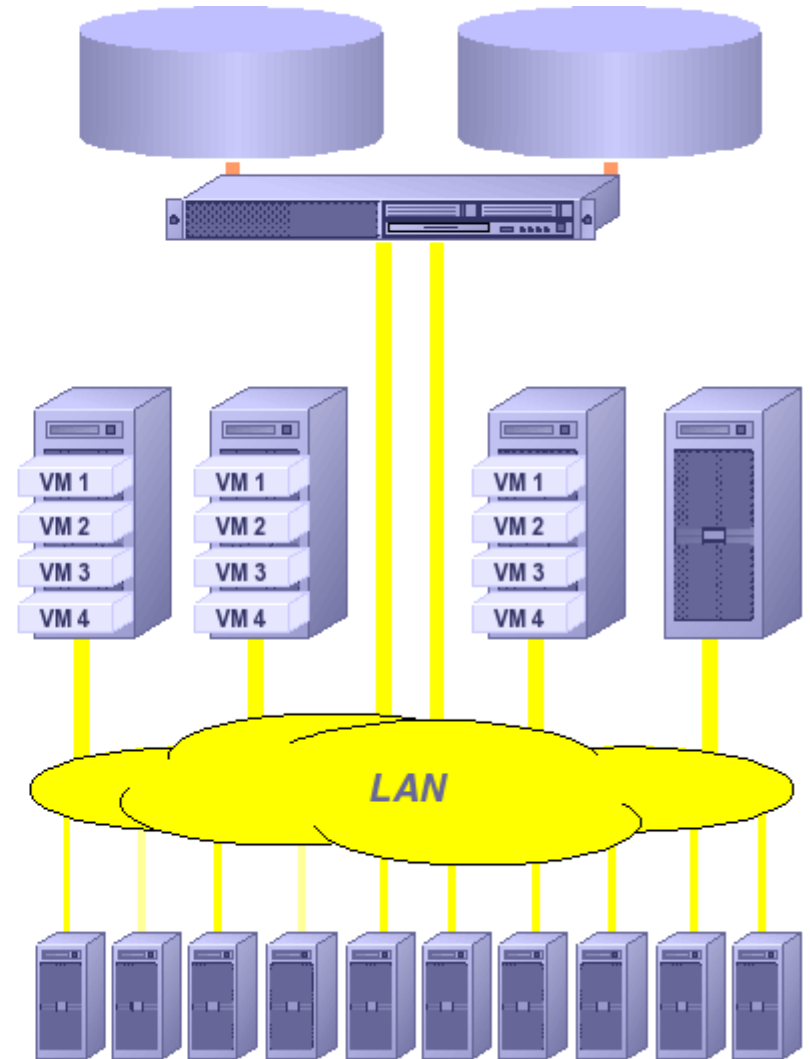
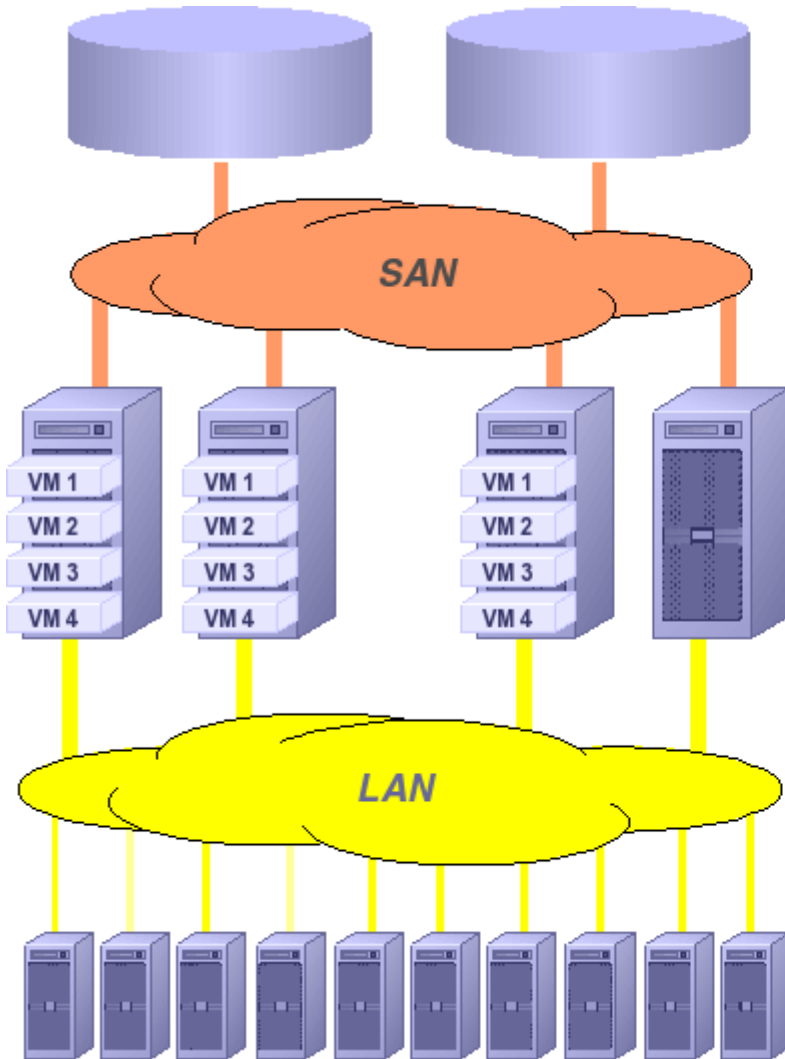
- Genau das Gegenteil des VM-Ansatzes: Strikte Trennung von OS und Applikation
- Applikation und ihr Runtime-Environment leben außerhalb des OS weiter
- erhebliche Reduzierung des administrativen Aufwandes / leichteres Lifecycle-Management

OSL Gesellschaft für offene Systemlösungen mbH

www.osl.eu

Herausforderungen: 2. Netzwerkkonvergenz

hat Vor- und Nachteile, ist aber unter vielen Aspekten interessant



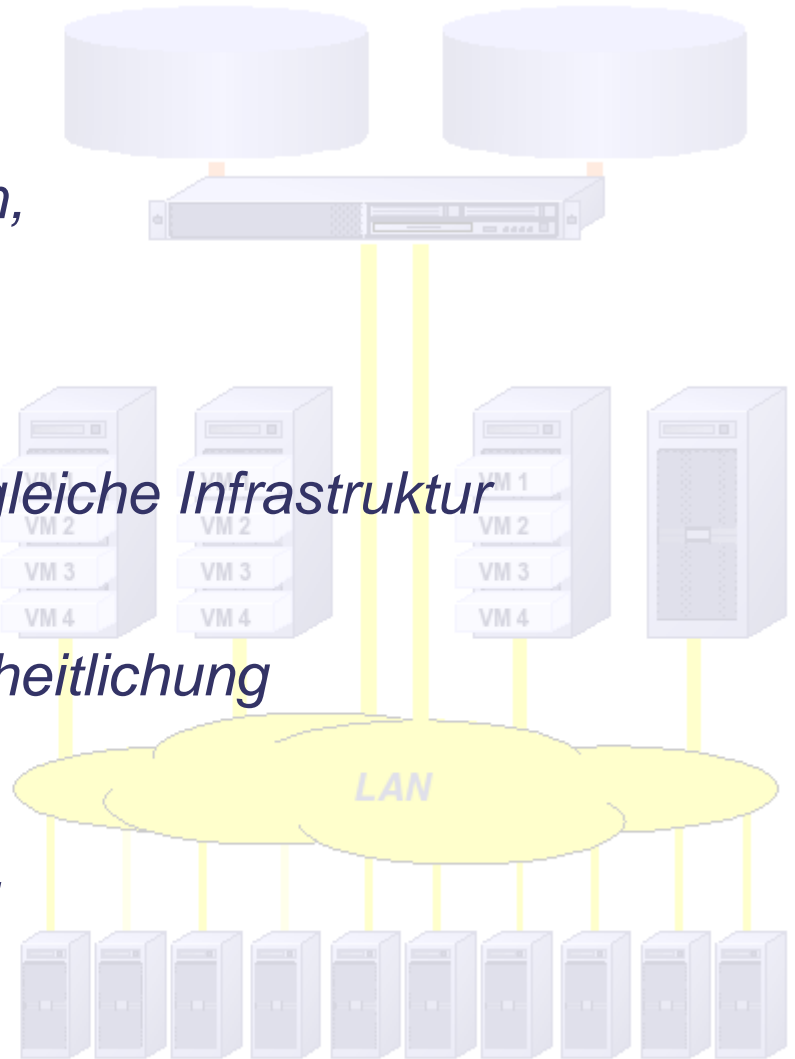
Richtig angepackt überwiegen ganz klar die Vorteile!

Herausforderungen: 2. Netzwerkkonvergenz

hat Vor- und Nachteile, ist aber unter vielen Aspekten interessant

- Ethernet überall verfügbar
- ich kann ein RZ ohne FC-SAN betreiben, kaum aber ohne (Ethernet-) Netzwerk
- günstige Kosten je Port
- Speicher- und VM-Anbindung über die gleiche Infrastruktur -> VM-Mobilität und Flexibilität
- organisatorische / administrative Vereinheitlichung
- standardisierte Infrastruktur (Hardware)

Und: Es muß nicht immer Ethernet sein!



Richtig angepackt überwiegen ganz klar die Vorteile!

Herausforderungen: 2. Netzwerkkonvergenz

Network Operations mit Unified I/O per FCoE

Server Operations

- Drivers, Bindings
- Miscellaneous SCSI-settings
- LUN-Setup
- Multipathing
- Disk Partitioning
- Network Management

SAN Operations

- FC Interface Mode and Speed
- Virtual Fabrics
- Port Channels & Trunks
- NPV
- Monitoring & SAN Mgmt. Systems
- Device Management

Converged Network Operations

- Switch Management
- Images, Users, Virtual Interfaces
- Management Interfaces
- System Level Support
- Span Ports

LAN Operations

- Drivers, Port Configurations
- VLANs
- Port Channels
- Trunks
- Monitoring + LAN Management
- Device Management

Vgl. "Design and Implementations of FCoE for the DataCenter" - SNIA Educational Paper, 2011

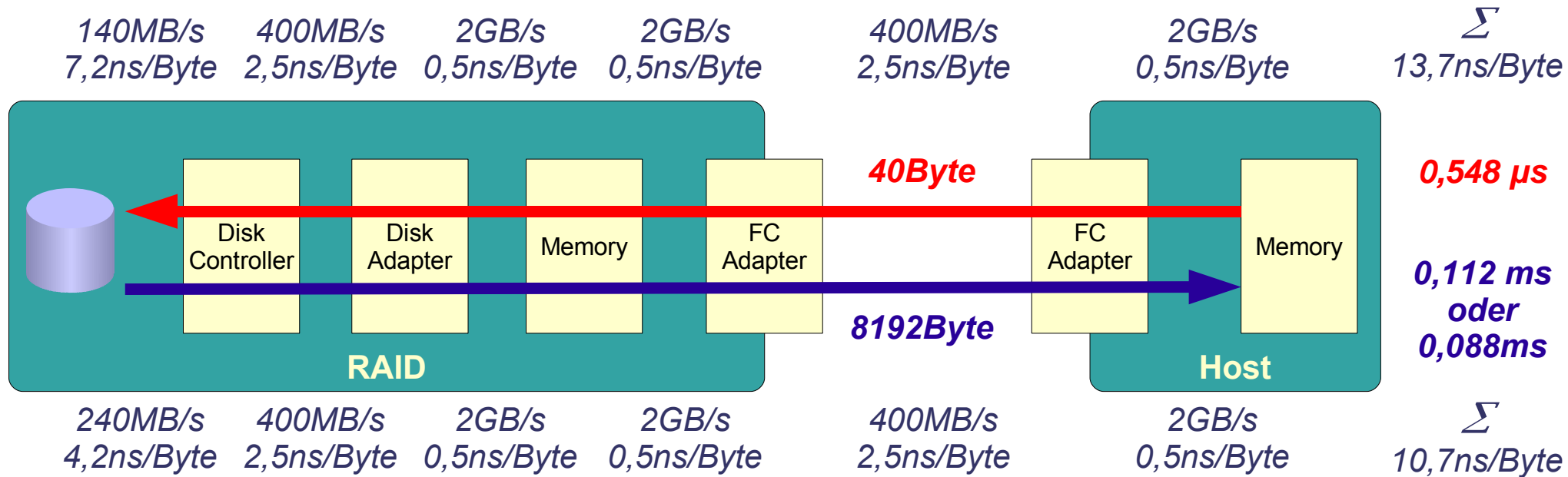


**Netzwerkkonvergenz bei Hardware ist nur die halbe Miete.
Genauso wichtig ist eine Vereinfachung der Administration.**

Herausforderungen: 3. Neue Speicherlösungen

Der Schritt von der HD zur SSD ist erst der Anfang

	HD	SSD	FeRAM	DDR SDRAM	MRAM
Kapazität	> 2 TB	> 512 GB	16 MB (Modul)	> 4 GB (Modul)	Universal RAM ?
Zugriff	4 ms	0,2 ms	0,05 μ s	10 ns	
Transfer	140 MB/s	280 MB/s	1,6 GB/s (DDR2)	40 GB/s	



bei 140 MB/s ca. 8900 Transfers/s ca. 70MB/s
 bei 280 MB/s ca. 11350 Transfers/s ca. 88MB/s

Herausforderungen: 3. Neue Speicherlösungen

Der Schritt von der HD zur SSD ist erst der Anfang



- *Erlauben die heutigen High-End-Speichersysteme Zugriff auf die modernsten Technologien im Storage-Bereich?*
- *Wenn die Grenzen zwischen Storage und Servern verschwimmen: Wie passen heutige Speichernetzparadigmen dazu?*
- *Die schnellsten Komponenten stehen quasi für PC-Technologie (Standardsysteme) zur Verfügung*
- *SCSI (auch über FC oder IP) ist keine adäquate Antwort (mehr):*
 - *keine Kenntnis von Vernetzung / verteilten Funktionen*
 - *kaum parallelisierbar*
 - *in heutigen Systemen unnütze I/O-Transformationen*
 - *damit Performance- und Verfügbarkeitsnachteile*

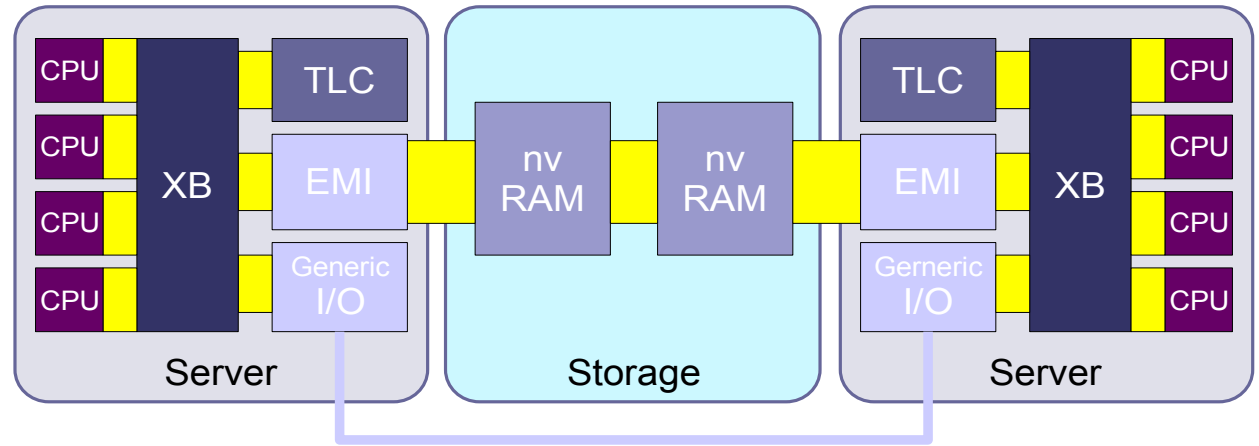
Schlußfolgerung

Es wird eine Aufwertung von Standard-Servern, Mehrzweck-Netzwerken und hostbasierter Software geben!

Herausforderungen: 3. Neue Speicherlösungen

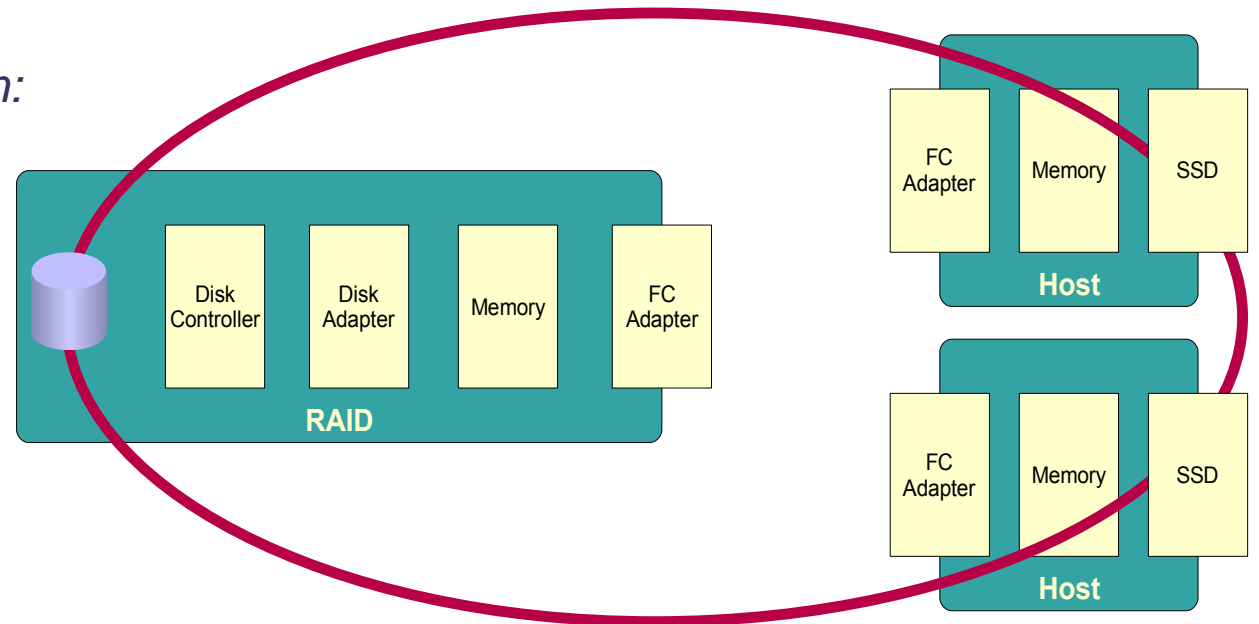
Der Schritt von der HD zur SSD ist erst der Anfang

Im RZ noch Phantasie:
Komplett neuartiges
Systemdesign



Ein erster Ansatz nimmt Gestalt an:

lokaler Massenspeicher
verspricht deutliche
Performancegewinne



Herausforderungen:

- Integration mit externem Storage
- Beherrschung im Cluster
- Umsetzung der Performancegewinne im Gesamtkonstrukt

Wohin wird es also in Zukunft gehen ?

- *Unterstützung mehrerer Plattformen*
- *Unterstützung mehrerer Speichernetzwerktechnologien*
- *RSIO als **das** neue I/O-Protokoll*
 - *auch zur Implementierung ultraschneller Block-I/O-Storage-Server unter Verwendung von Industrie-Standardkomponenten*
- *Leistungsfähige Bausteine für flexible Lösungen*
 - *Integration der führenden Storage-Lösungen*
 - *I/O-Beschleuniger*
 - *Kombination Netzwerk-Storage + lokaler Storage*
- *Applikationsbezug im gesamten Stack*
- *Einheitliches Virtualisierungsframework*

kooperative Ausrichtung unserer Konzepte

Portabilität und Standardisierung

Offenheit und Flexibilität

Danke !

***OSL wünscht eine interessante
Veranstaltung!***