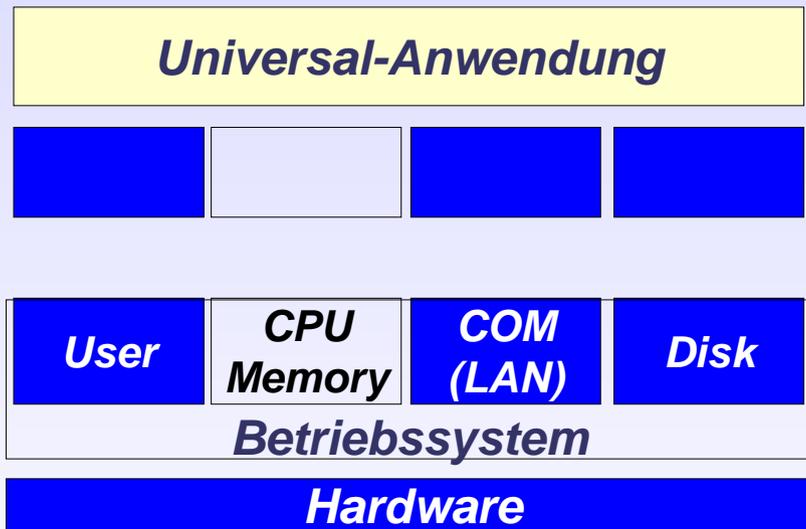
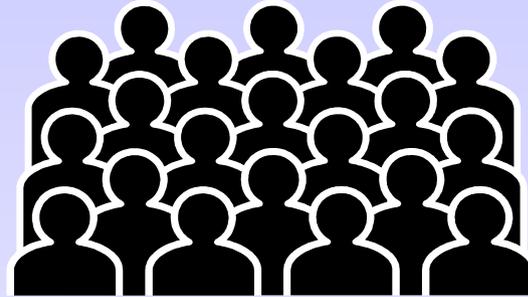


Herzlich Willkommen!

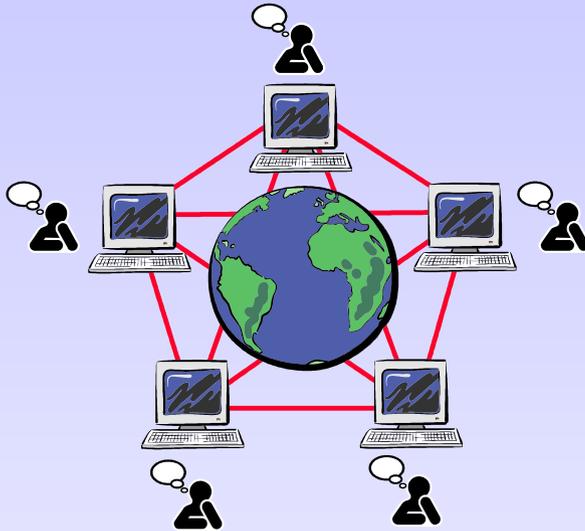
*Von der Speichervirtualisierung
zur RZ-Lösung*

OSL Storage Cluster 3.1



Ein zentrales System:

- mit universellen Einsatzmöglichkeiten, möglichst einer Anwendung für alles
- das nie ausfällt und sicher ist
- mit genügend Ressourcen für beliebige Problemstellungen
- mit einfachster Bedienung
- mit Zugang für alle
- bezahlbar

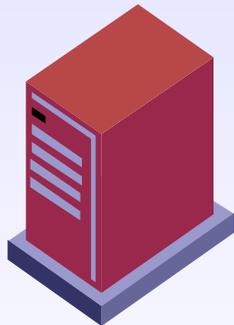


Dezentralität

- Benutzer wollen von verschiedenen Orten aus zugreifen
- verteilte System- und Anwendungsarchitekturen

Leistung = Preis²

- Preis für symmetrische Systeme steigt nahezu quadratisch mit der Größe / Zahl der CPU



Ausfallrisiko

- Systeme können ausfallen
- je größer und komplexer ein System, desto wahrscheinlicher ist ein Ausfall
- je zentraler und konsolidierter ein System, desto gravierender die Konsequenzen eines Ausfalls

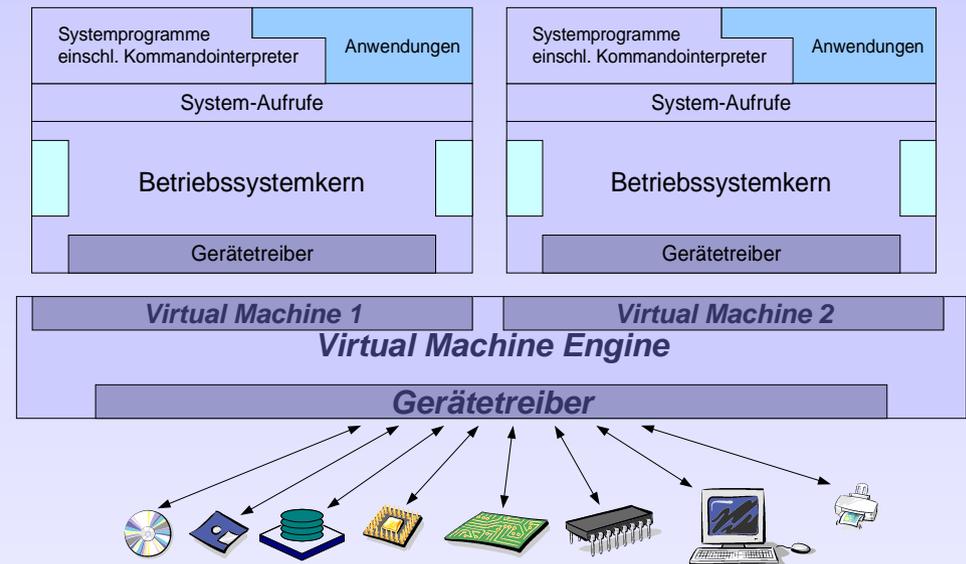
Zwei alternative Lösungsmöglichkeiten?

Cluster OS



- **faszinierende Möglichkeiten**
- **komplexe Herausforderungen in der Administration**
- **gute Eignung für HV, hohe Flexibilität**
- **geringere HW-Kosten**
- **hohe Effizienz und Integration nur mit Spezial-OS (Marktakzeptanz und Preis?)**
- **extreme Kommunikationsanforderungen**
- **Speicher- und Geräteanbindung schwierig**

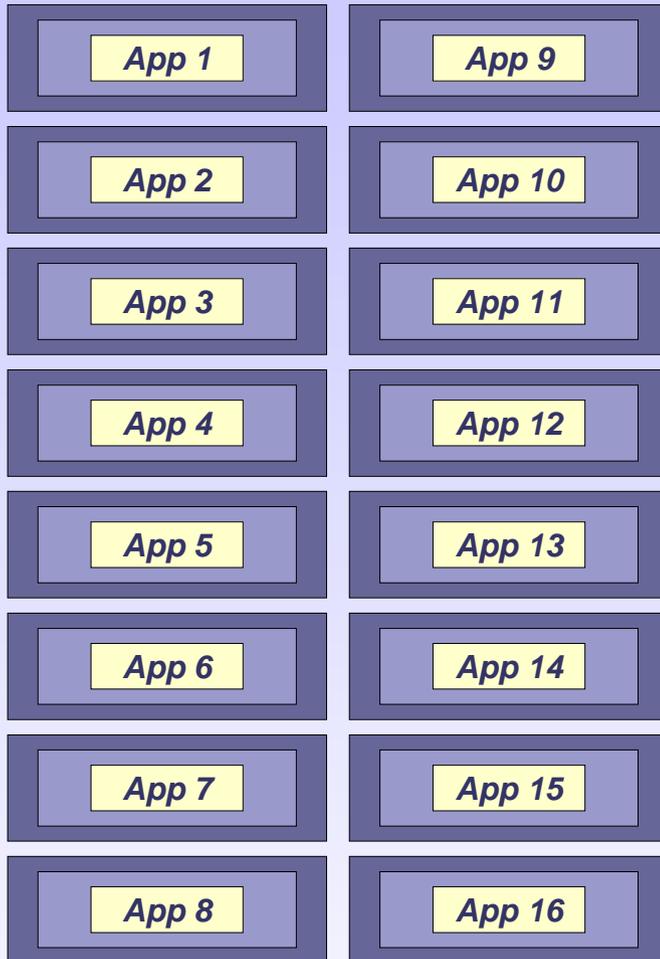
Virtual Machines



- **weniger komplexe Administration**
- **höhere HW-Kosten bei geringerer HW-Redundanz**
- **Sonderproblem Devicesharing**
- **je nach Implementierung deutlicher Overhead**
- **eliminiert einige Vorteile von SMP-Systemen**

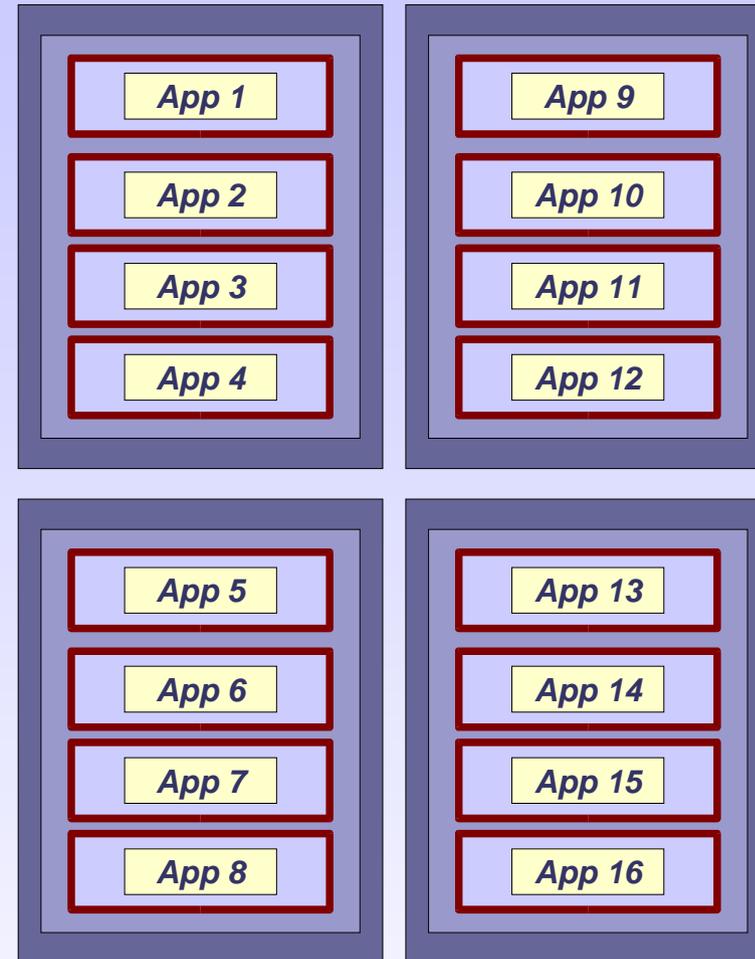
Der Sieger in der Praxis

ohne virtuelle Maschinen



**aus 16 realen Maschinen
aus 32 Softwareumgebungen
aus dem Nichts**

mit virtuellen Maschinen

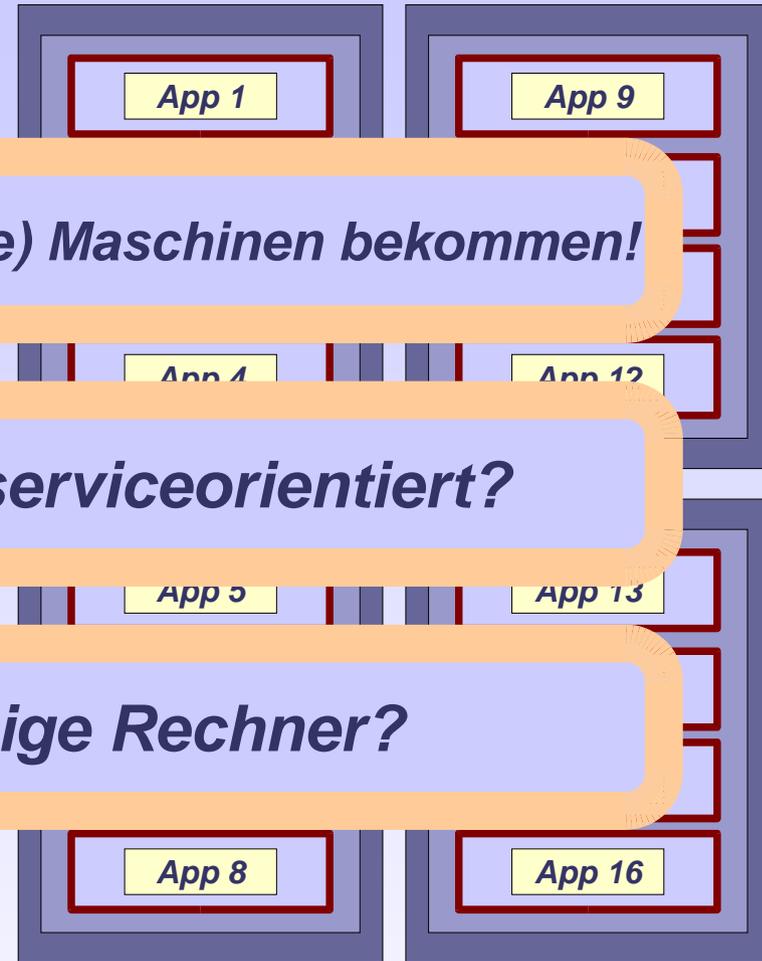
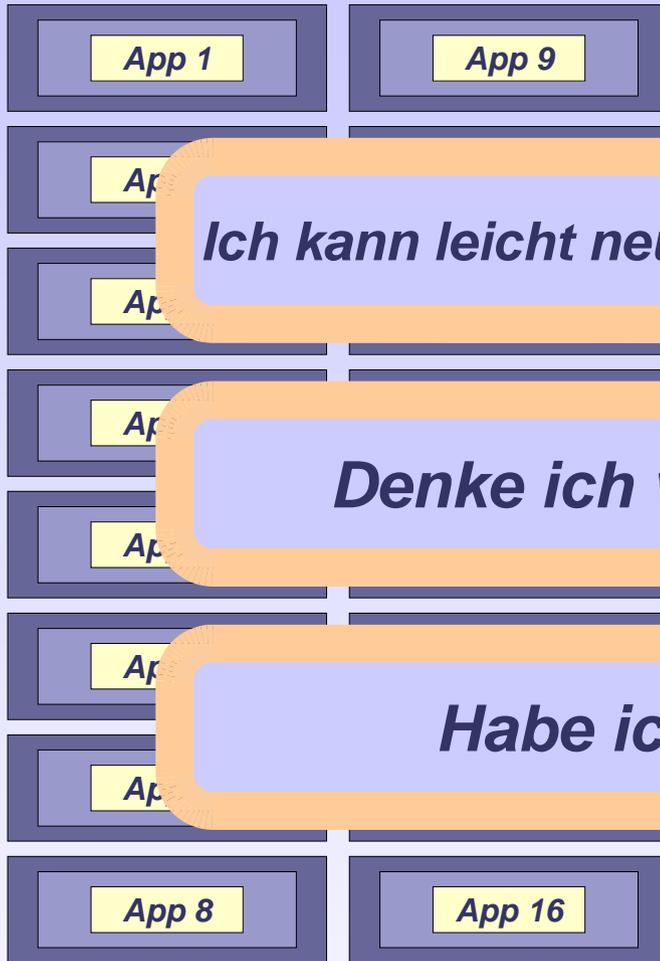


**wird
werden
entstehen** **die Illusion von 16 realen Maschinen
mindestens 36 Softwareumgebungen
mindestens 12 neue SW-Abhängigkeiten**

Der Sieger in der Praxis

ohne virtuelle Maschinen

mit virtuellen Maschinen



Ich kann leicht neue (virtuelle) Maschinen bekommen!

Denke ich wirklich serviceorientiert?

Habe ich zu wenige Rechner?

**aus 16 realen Maschinen
aus 32 Softwareumgebungen
aus dem Nichts**

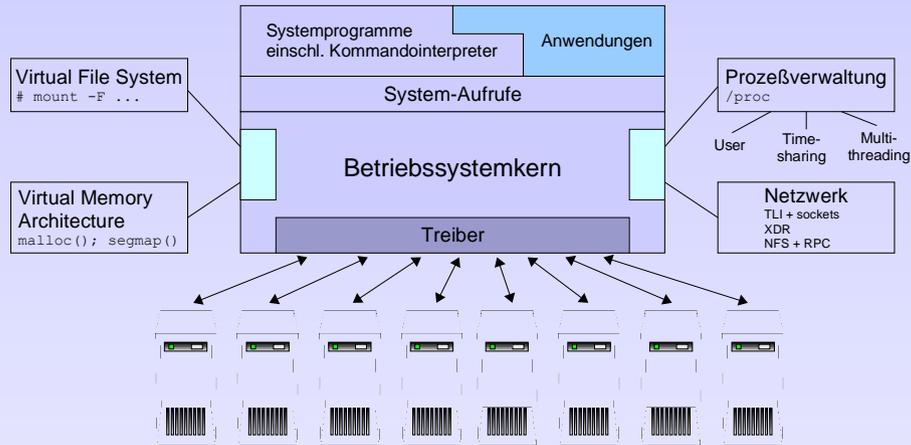
**wird
werden
entstehen**

**die Illusion von 16 realen Maschinen
mindestens 36 Softwareumgebungen
mindestens 12 neue SW-Abhängigkeiten**

Services / Applikationen bereitstellen:

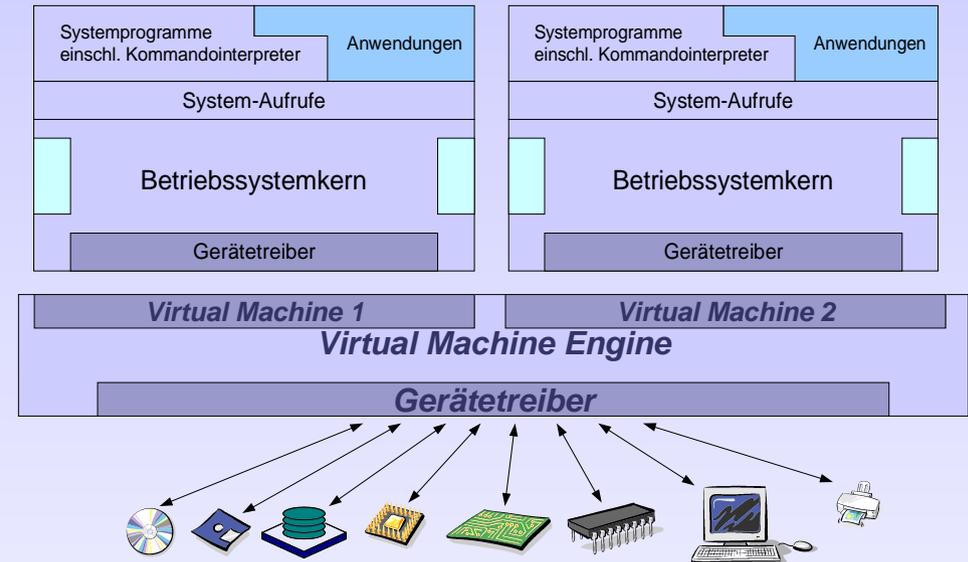
- ***mit hoher Verfügbarkeit***
- ***mit optimaler Performance***
- ***mit minimalem Aufwand***

Cluster OS



- **faszinierende Möglichkeiten**
- **komplexe Herausforderungen in der Administration**
- **gute Eignung für HV, hohe Flexibilität**
- **geringere HW-Kosten**
- **hohe Effizienz und Integration nur mit Spezial-OS (Marktakzeptanz und Preis?)**
- **extreme Kommunikationsanforderungen**
- **Speicher- und Geräteanbindung schwierig**

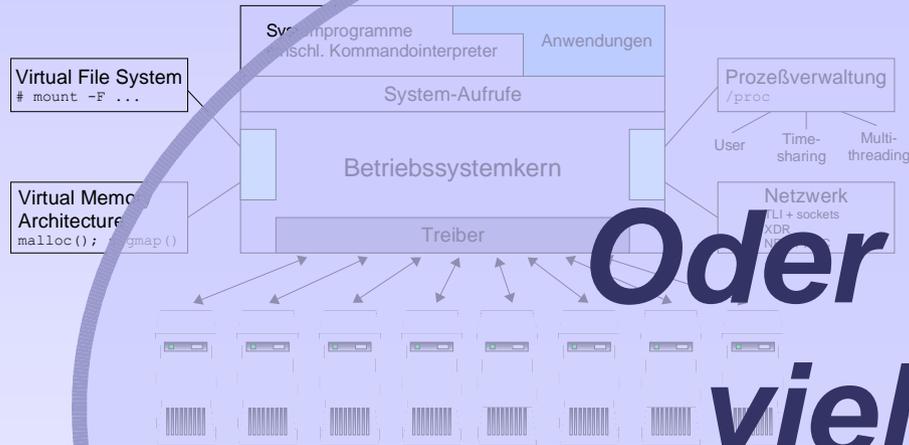
Virtual Machines



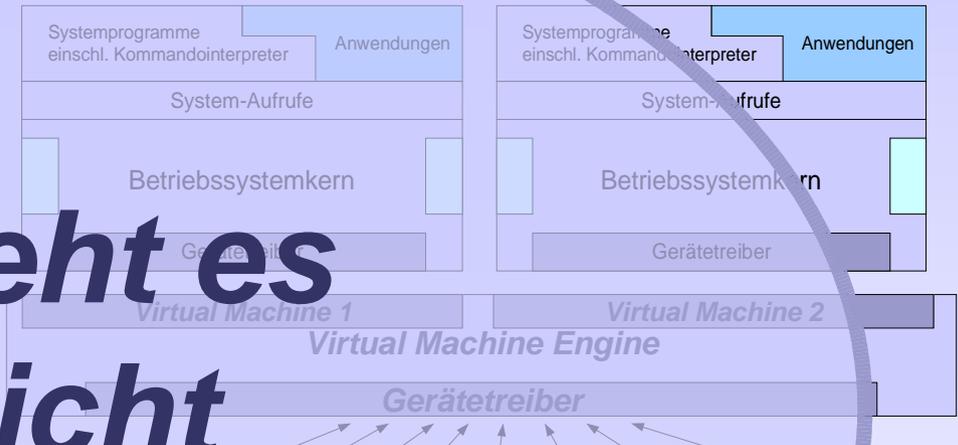
- **weniger komplexe Administration**
- **höhere HW-Kosten bei geringerer HW-Redundanz**
- **Sonderproblem Devicesharing**
- **je nach Implementierung deutlicher Overhead**
- **eliminiert einige Vorteile von SMP-Systemen**

Zwei alternative Lösungsmöglichkeiten?

Cluster OS



Virtual Machines



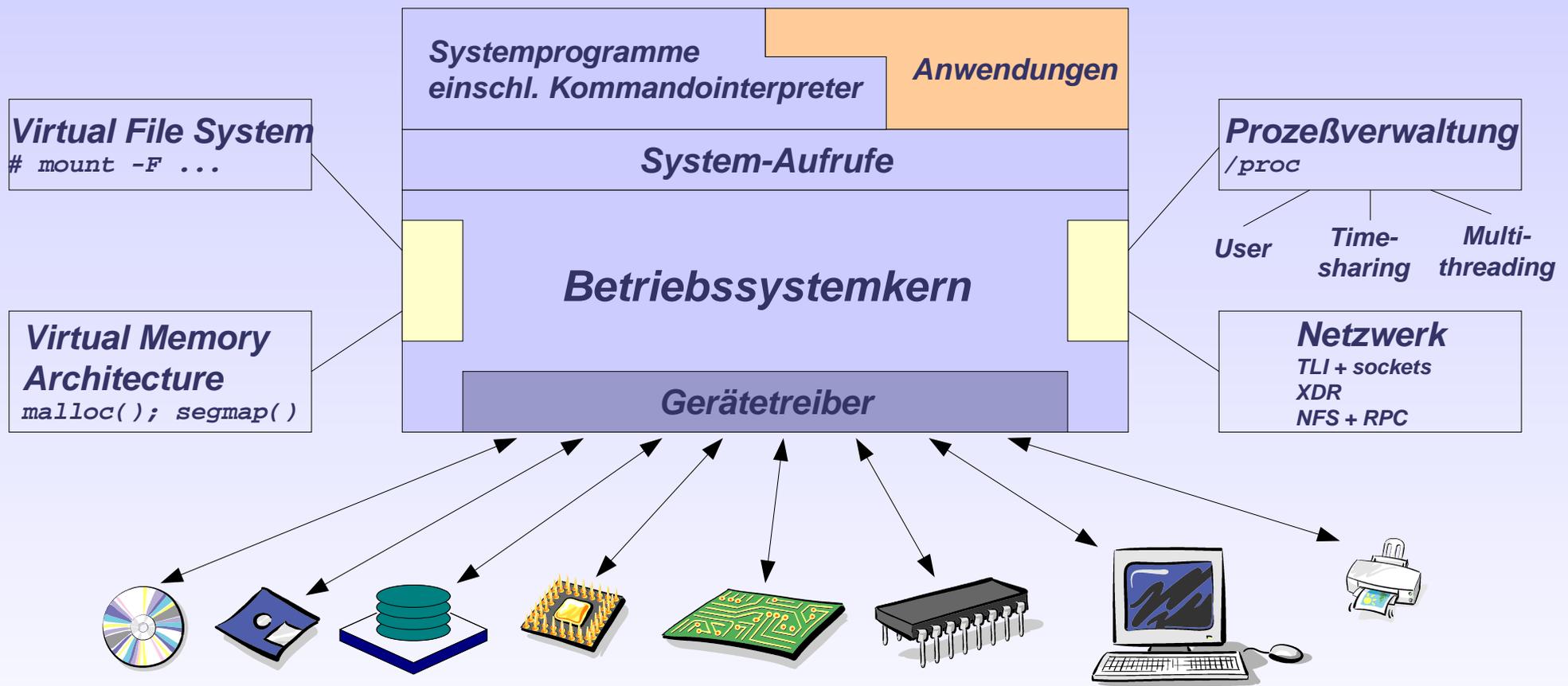
**Oder geht es
vielleicht
auch ganz anders?**

- **faszinierende Möglichkeiten**
- **komplexe Herausforderungen in der Administration**
- **gute Eignung für HV, hohe Flexibilität**
- **geringere HW-Kosten**
- **hohe Effizienz und Integration nur mit Spezial-OS (Marktakzeptanz und Preis?)**
- **extreme Kommunikationsanforderungen**
- **Speicher- und Geräteanbindung schwierig**

- **weniger komplexe Administration**
- **höhere HW-Kosten bei geringerer HW-Redundanz**
- **Sonderproblem Devicesharing**
- **je nach Implementierung deutlicher Overhead**
- **eliminiert einige Vorteile von SMP-Systemen**

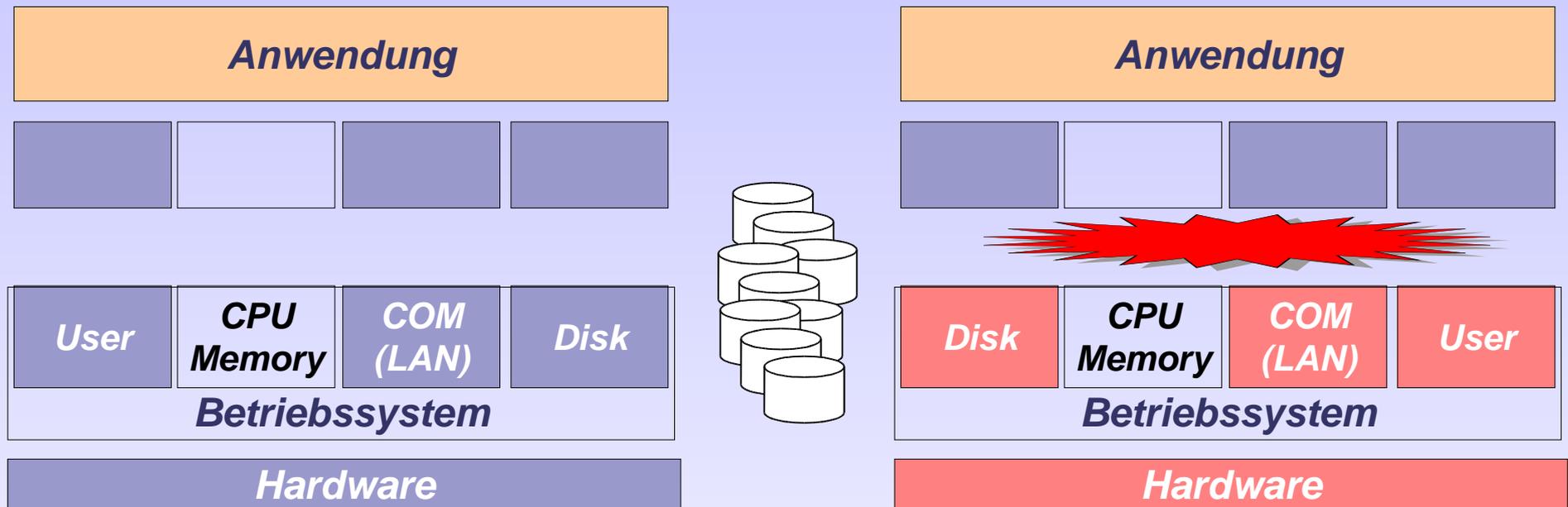
Was bietet mir das Standard-Unix?

Eine weitgehend virtuelle und selbstoptimierende Ablaufumgebung



Mindestens ein Problem bleibt

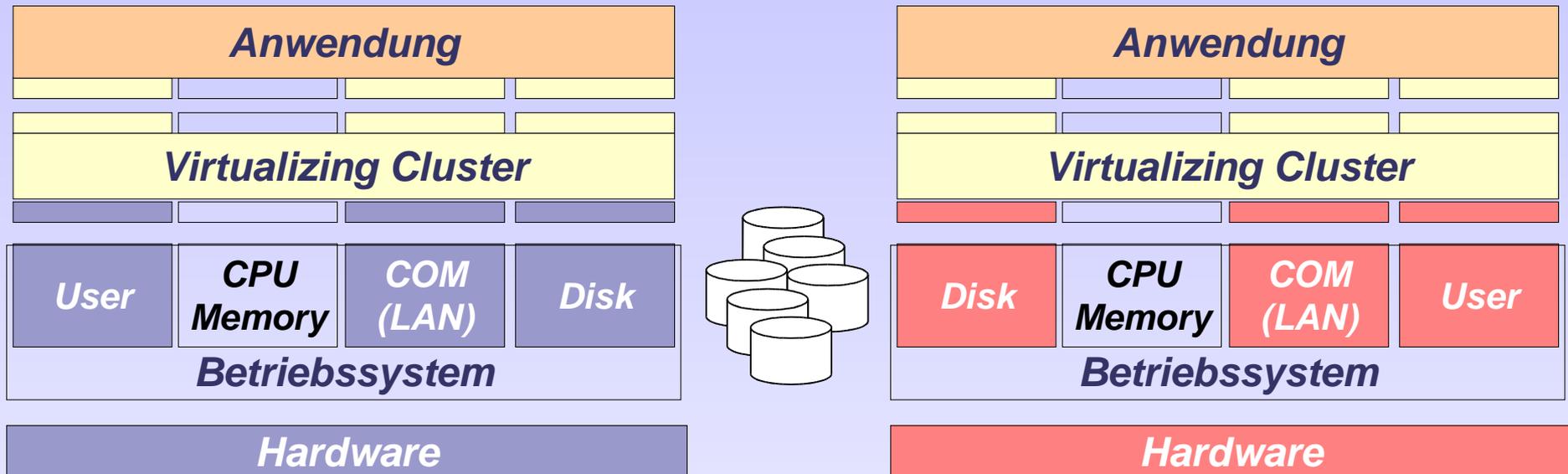
Hardwarebezug und unflexible Installation als Bremse für HV und Optimierung



- **Installation der Anwendung heißt:**
 - Kopieren von Programmen und Daten auf bestimmte Ressourcen des Systems
 - Parametrisierung entsprechend den Ressourcen von Hardware und Betriebssystem
- **Mit der Installation entstehen Abhängigkeiten von einem bestimmten System**
- **Ablauf der Applikation auf anderem System nicht möglich**

Unser Lieblingsansatz

zusätzliche rechnerübergreifende Virtualisierung

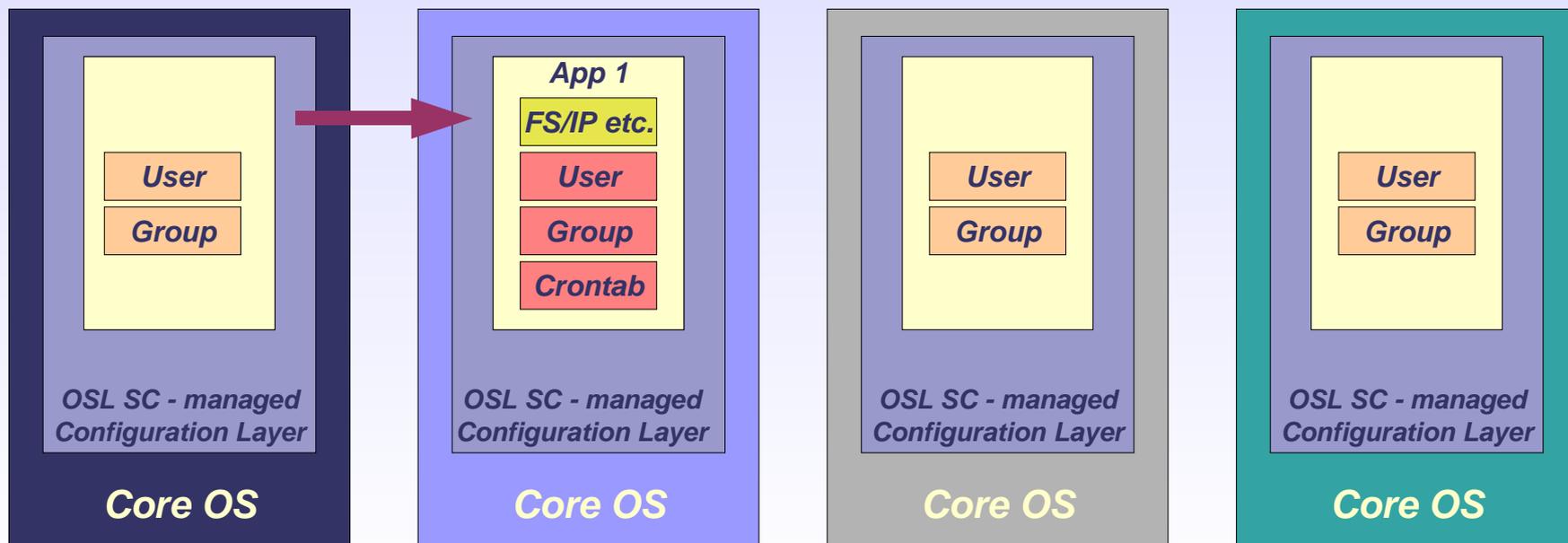


- **Nutzung eines Standard-Betriebssystems (Solaris)**
- **daher keine eigenen Parallelisierungsmöglichkeiten**
- **Vereinfachung der Administration durch Virtualisierung und Clusterdienste**
- **geeignet für heutige Anwendungen und für Hochverfügbarkeit**
- **hohe Flexibilität, optimale Anpassung an heutige IT-Infrastrukturen**
- **geringere HW-Kosten**
- **niedrige Kommunikationsanforderungen**
- **einheitliche Speicheranbindung**
- **Integration mit Storage-Virtualisierung bietet zusätzlichen Nutzen**

Und was brauchen wir noch?

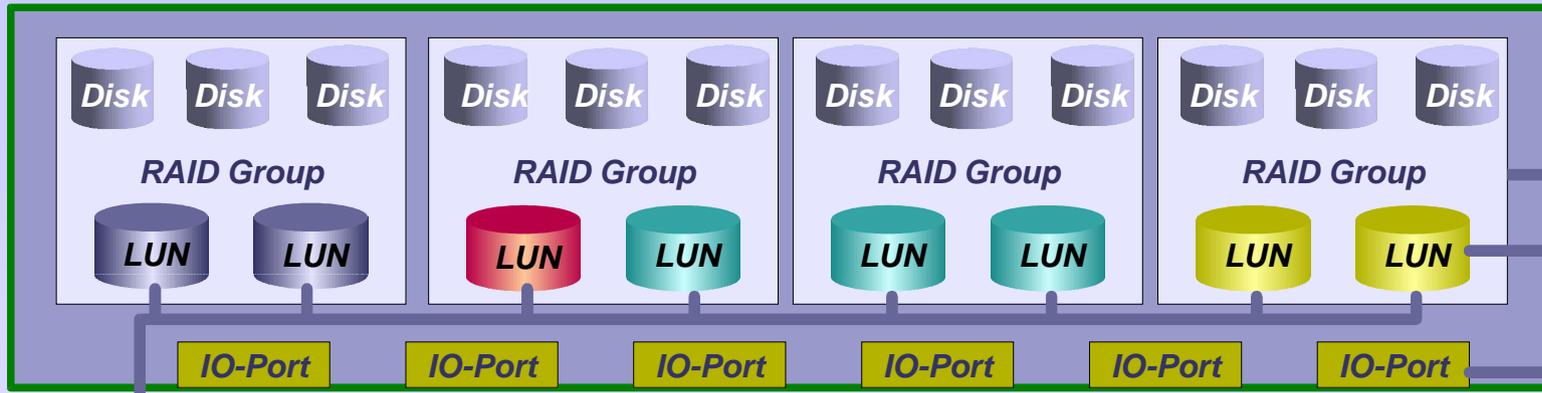
Globale Nutzerverwaltung komplettiert virtualisierte Ablaufumgebungen

- geeignet für Server / Application Service User
- Unabhängig von externen Services wie NIS/LDAP/ADS
- Vermeidung von Konflikten, Synchronisation, automatische Reparatur
- User kann einer Applikationen zugeordnet werden
- Crontab und Login-Möglichkeit wandern mit der Applikation
- auch nach Neuinstallation sofort wieder verfügbar



Blick in das durchschnittliche Rechenzentrum

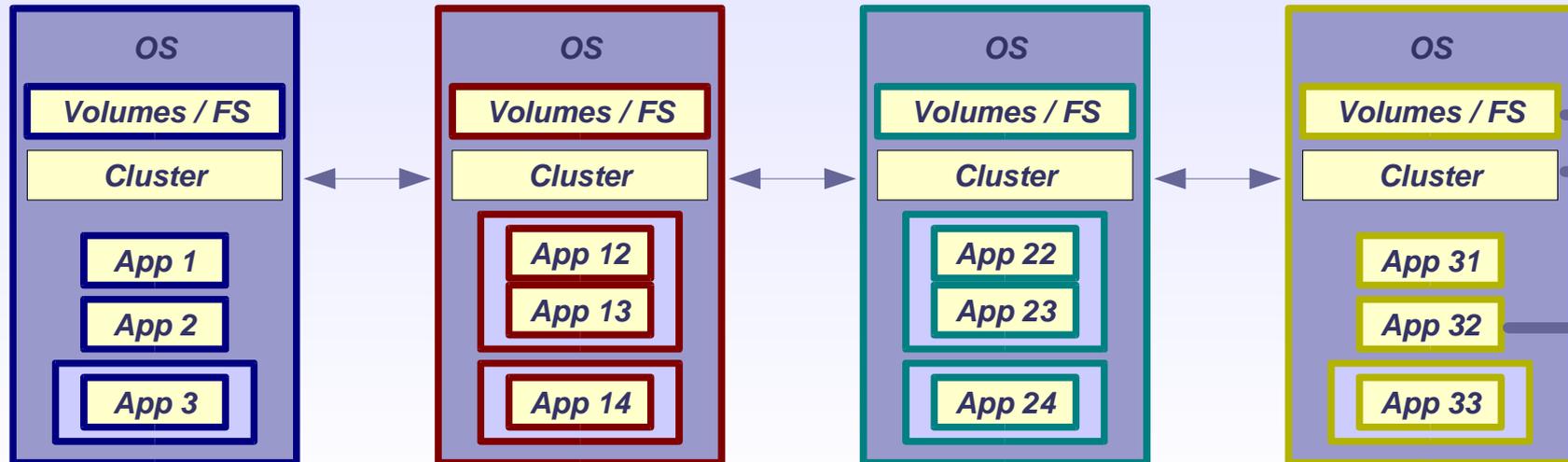
Moderne Infrastruktur, vielschichtige Administration



unflexibler, hostbezogener Storage mit Verschnitt

S A N

Isolierte Administration

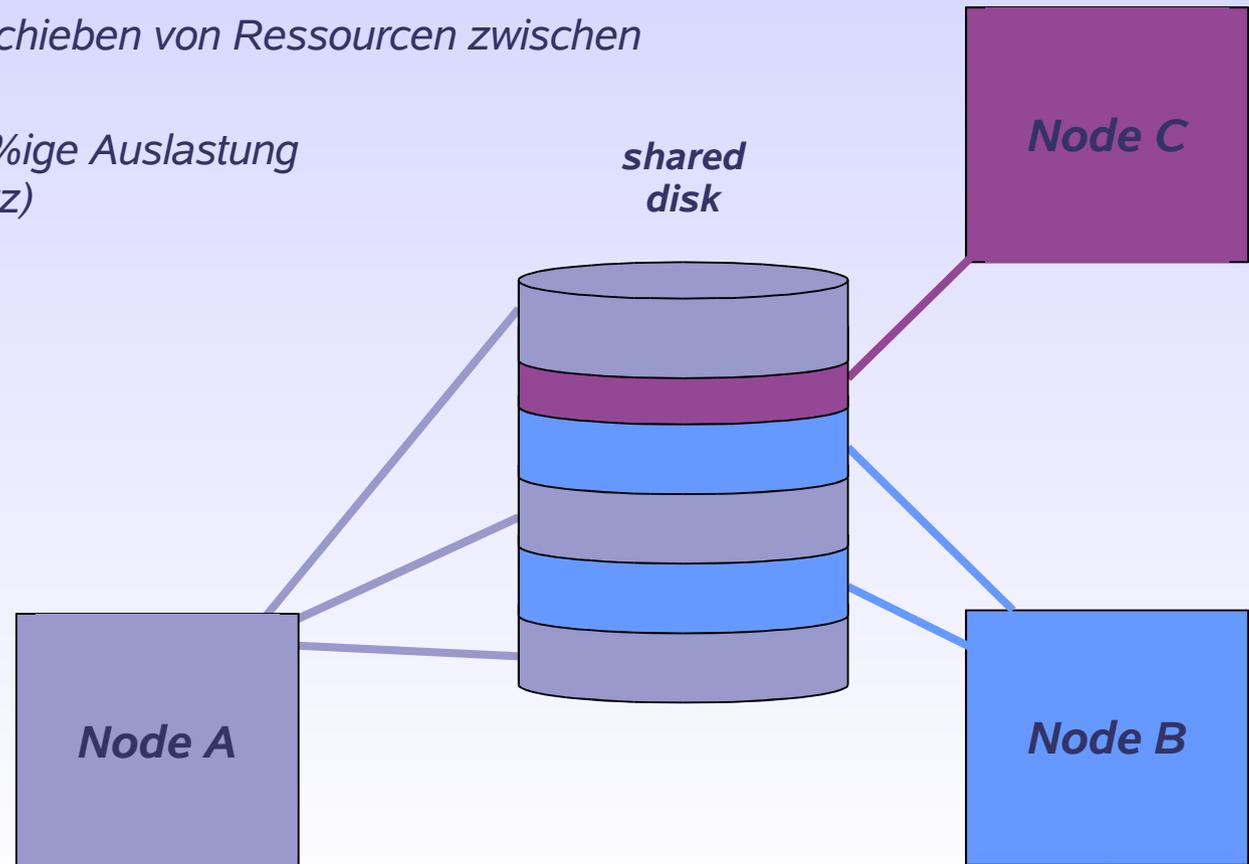


Ungenutztes Potential: Disk-Sharing

Vorteile mit einem alternativen Konzept

Vorteile des Disk-Sharings

- RAID systeme können große LUNs bereitstellen / Volume Sharing erhöht Flexibilität und verbessert Nutzung des Speichers
- Load-Balancing zwischen allen Servern (IO)
- Flaschenhalse durch IO-Peaks und Disk Overload vermeiden
- ermöglicht problemsloses Verschieben von Ressourcen zwischen Servern
- globale Pools ermöglichen 100%ige Auslastung (kein Verschnitt von Plattenplatz)
- einfachere Administration

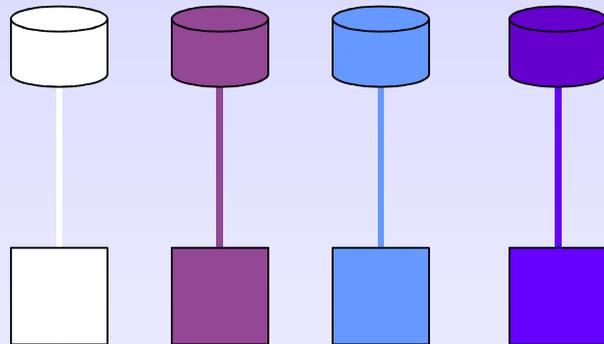


Verbesserung IO-Verhalten durch Disk-Sharing

Vorteile überwiegen, Nachteil kann durch Software aufgefangen werden

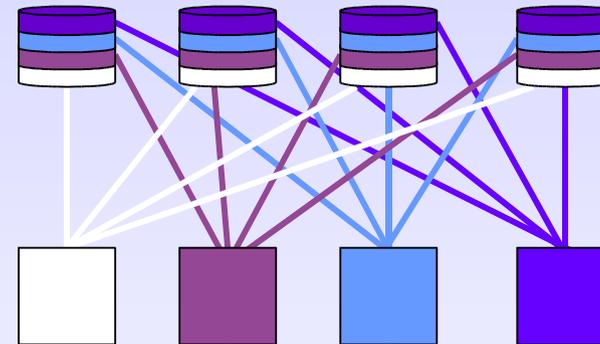
Erhöhte Bandbreite mit Disk-Sharing und multithreaded IO

no disk sharing



- Bandbreite genau einer Platte, auch für Peak-Loads
- keinerlei Flexibilität

shared disk



- 4x Bandbreite einer Platte für Peak-Loads
- Seiteneffekt: Mehr Flexibilität
- komplexere Administration durch OSL Storage-Cluster verborgen

Was haben wir erreicht?

Vereinfachung in der Administration und mehr Flexibilität

RAID-View

Spezifische Darstellung interner Ressourcen

Host-View

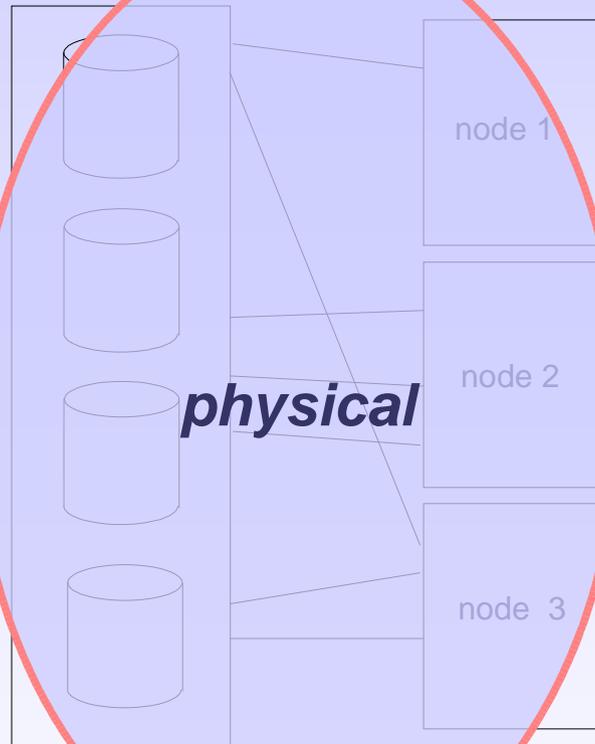
Hardwareabhängige, rechner-spezifische Darstellung externer Ressourcen

Virtueller "Disk-View"

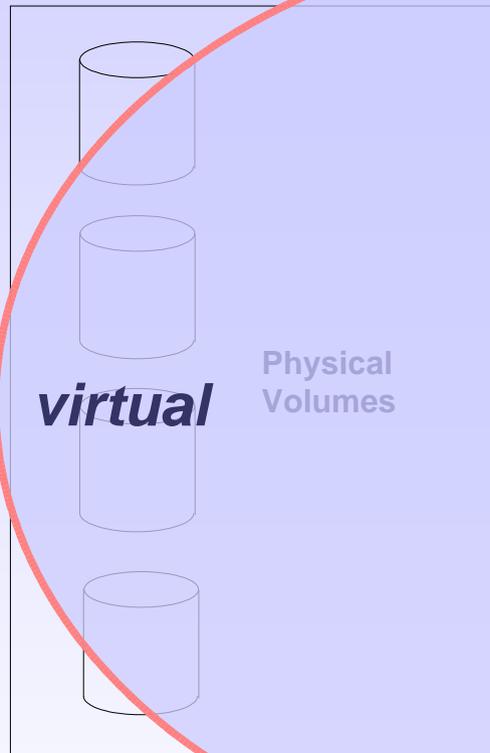
Clusterweit einheitliche, hardwareunabhängige Darstellung aller RAID-Ressourcen

Frei definierte Virtuelle Volumes

Clusterweit einheitliche, hardwareabstrakte Bereitstellung frei definierbarer, bedarfsgerechter virtueller Volumes

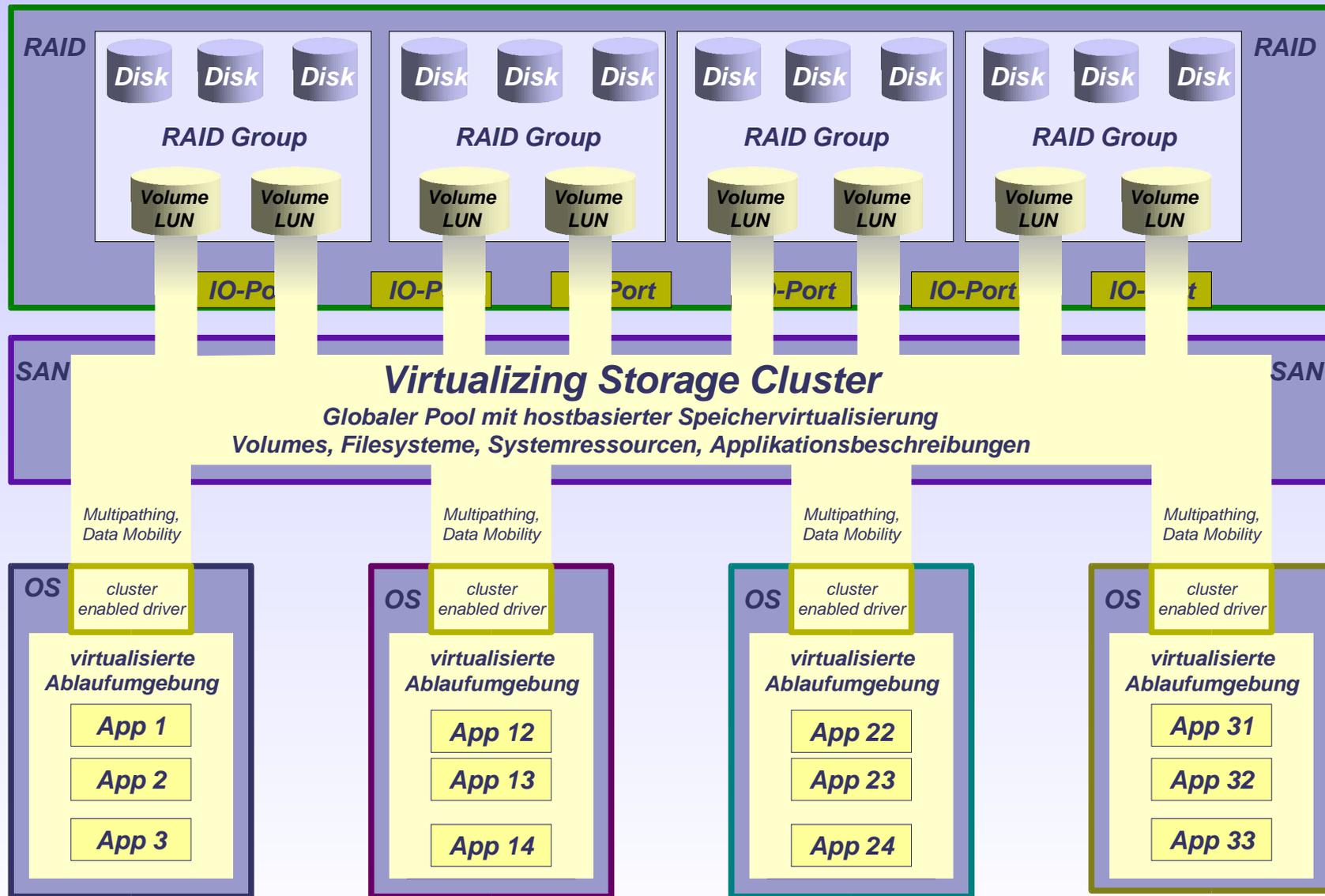


Stufe 1



Zentralisierung + Netztopologie = Cluster

OSL Storage Cluster: Vereinfachung durch Integration



Was heißt Shared Storage Clustering

Nutzung moderner RZ-Infrastrukturen für neuartige Management-Konzepte

OSL Storage Cluster:

- Lösung zur Integration von Unix-Servern mit modernen, RAID-basierten Speicherinfrastrukturen
- erweitert OS um aufeinander abgestimmte Virtualisierungs-, Management- und Cluster-Funktionalitäten
- Speicher- bzw. Volume-Management, Virtualisierung, System- und Applikationsmanagement sowie Clustering werden als Einheit begriffen
- das administrative Konzept und die Software selbst zielen auf flexible, virtualisierte Administrations- und Ablaufumgebungen
- deutliche Vereinfachung der Abläufe und administrativen Aufgaben im RZ

"Die Verbindung aus Softwaretechnologie und durchdachter, langfristig angelegter RZ-Organisation beim Anwender hilft, Ressourcen effektiv auszunutzen, Kosten zu senken und zusätzliche Freiheit bei der Auswahl der Systemplattformen zu gewinnen."

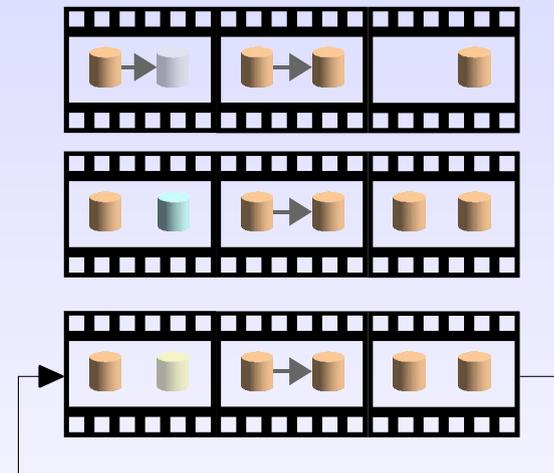
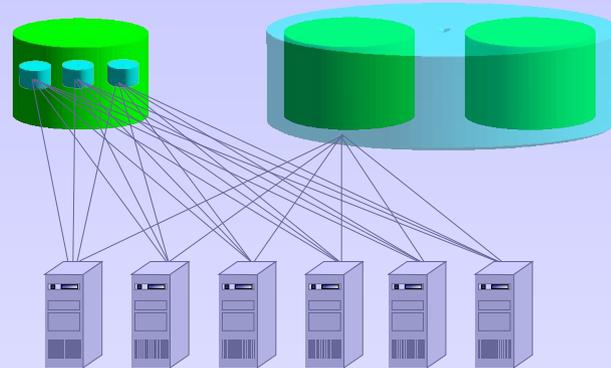
Kern des OSL Storage Clusters:

globale, hostbasierte Speichervirtualisierung

Was bietet OSL Storage Cluster im Speichermanagement?

Blockbasierte Virtualisierung – bedarfsgerecht und zuverlässig

Basis-Virtualisierung
clusterweit
globale Pools
Daten verschieben
Daten clonen
Daten spiegeln
Sonderfunktionen



Physical Volumes + Application Volumes
linear oder integriert (simple, concat, stripe)
HW-Abstraktion und IO-Multipathing
systemgestützte Allokation
Online-Konfig./-Dekonfig./-Vergrößerung

global devices / global namespace
integrated access management

rechnerübergreifend
global inventory
verschnittfreie Ausnutzung

online Daten verschieben / reorganisieren
automatische Priorisierung Anwendungs-IO

einmalig online auf beliebige Ziele kopieren
atomare Operationen für mehrere Volumes

Dauerhafte Beziehung Master -> Image
bis zu 3 Images
inkrementelle Resynchronisation
atomare Operationen für mehrere Volumes
Überbrückung Fehler auf Master

XVC (Extended Volume Controls)
z. B. Pause, Stop, Trigger, Aktionen
Bandbreitensteuerung
detaillierte Statistik

Leistungsumfang im Detail

Speicher-Virtualisierung, Anwendungssteuerung, Hochverfügbarkeit, Backup und DR

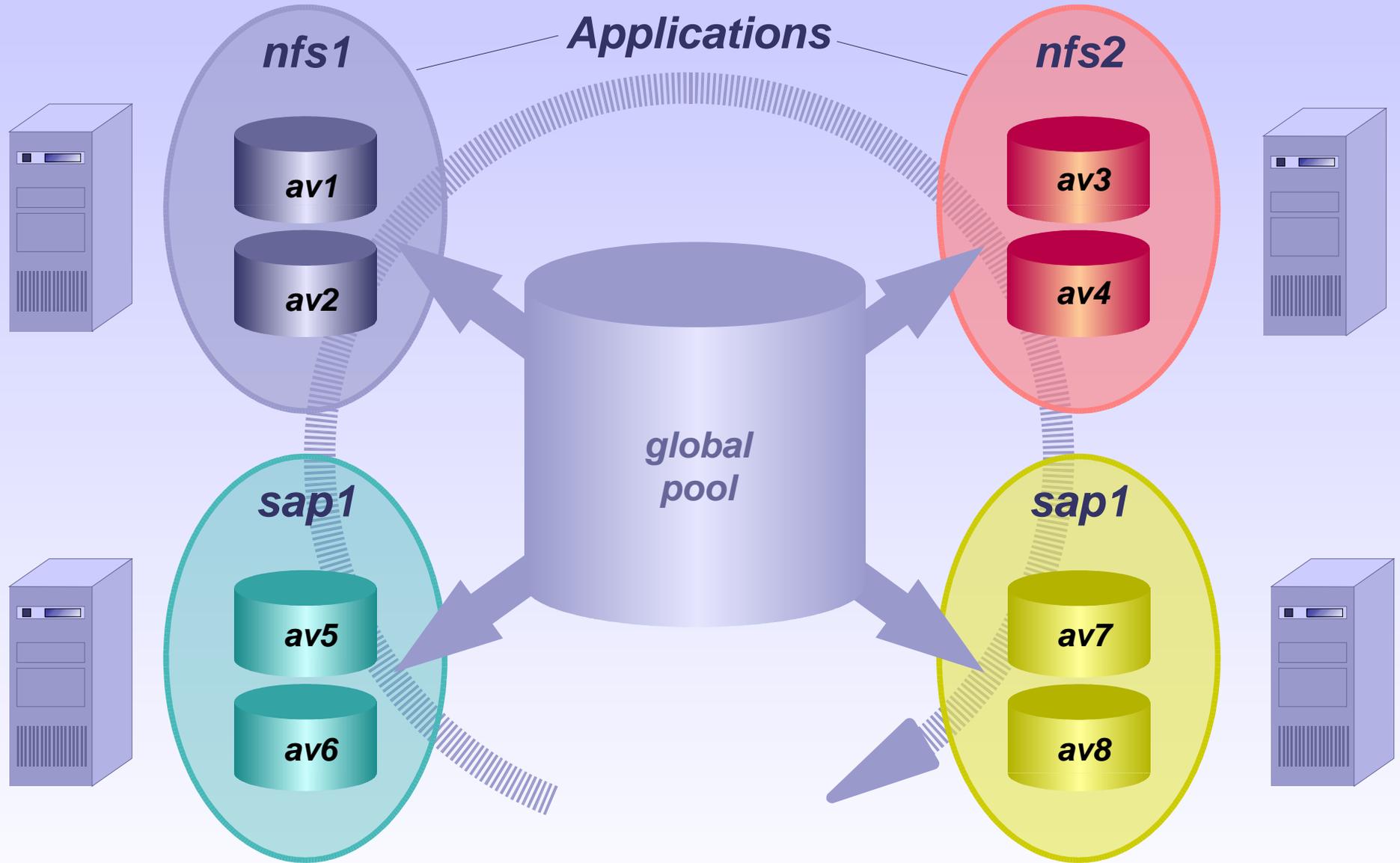


Application Awareness	Application Control Option	Application Mirrors
	<i>clusterweite Steuerung von Anwendungen</i>	
	<i>virtualisierte (hardwareabstrakte) Ablaufumgebungen</i>	
	<i>Hochverfügbarkeit</i>	
	<i>ressourcenbasiertes Selbstmanagement</i>	
Bandbreitensteuerung	Application Resource Description	Application Clones
User-Management		B2D / DASI / DR-Tools

Clusterfähige Speichervirtualisierung	Extended Data Management
<i>globale (hostübergreifende) Storage Pools</i>	<i>Integration RAID-basierter Datenkopien / Snapshots</i>
<i>Global Disk Inventory</i>	<i>Hostbasierte Datenspiegelung</i>
<i>Global Devices / Global Namespace</i>	<i>Live Data Migration</i>
Cluster-Volumenmanager mit automatischer Allokation	<i>Data Clones</i>
<i>Disk Access Management</i>	
<i>IO-Multipathing</i>	

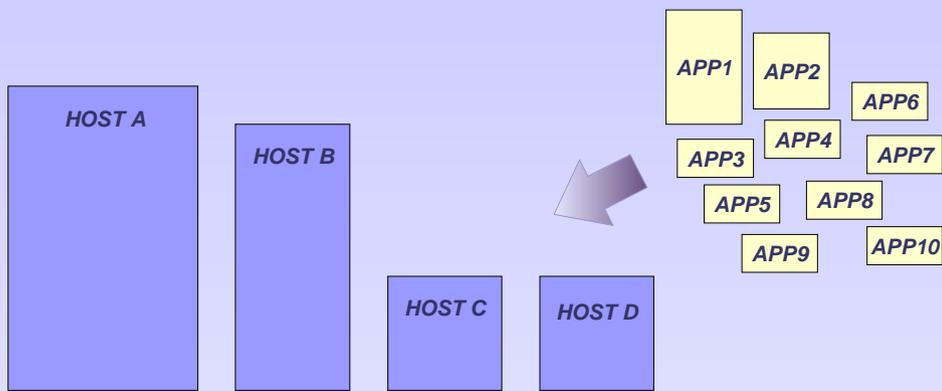
ist die Integration mit den Anwendungen

Speichervirtualisierung, Clustering und Applikationsmanagement gehören zusammen

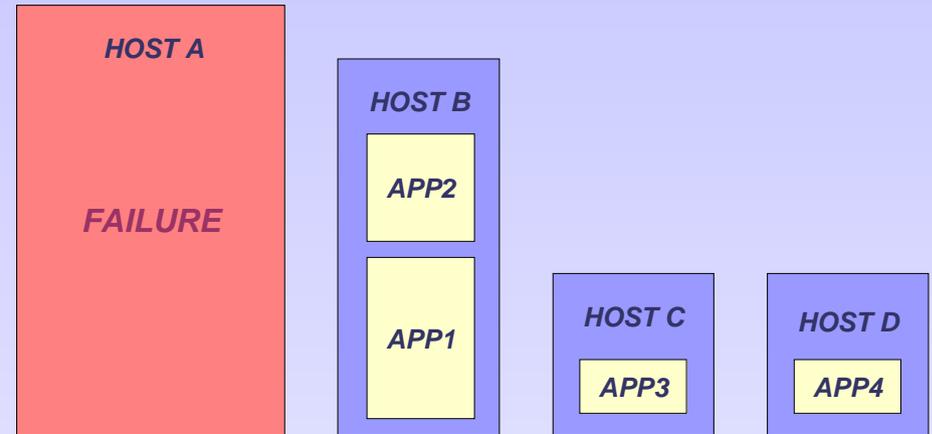


Und das geht natürlich auch

High Availability und Adaptive Computing

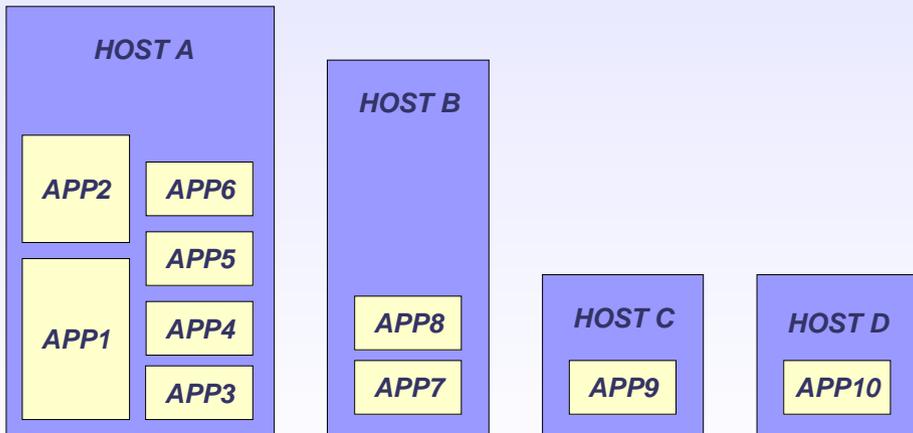


1

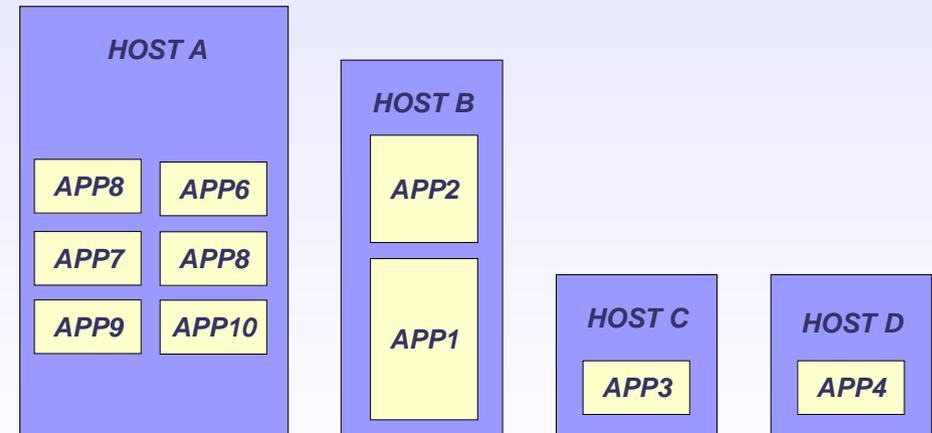


3

2



4



● Warum?

- Sättigung IO-Kanäle
- Sättigung Speichersystem(e)
- Konkurrenz Applikationen

● Was?

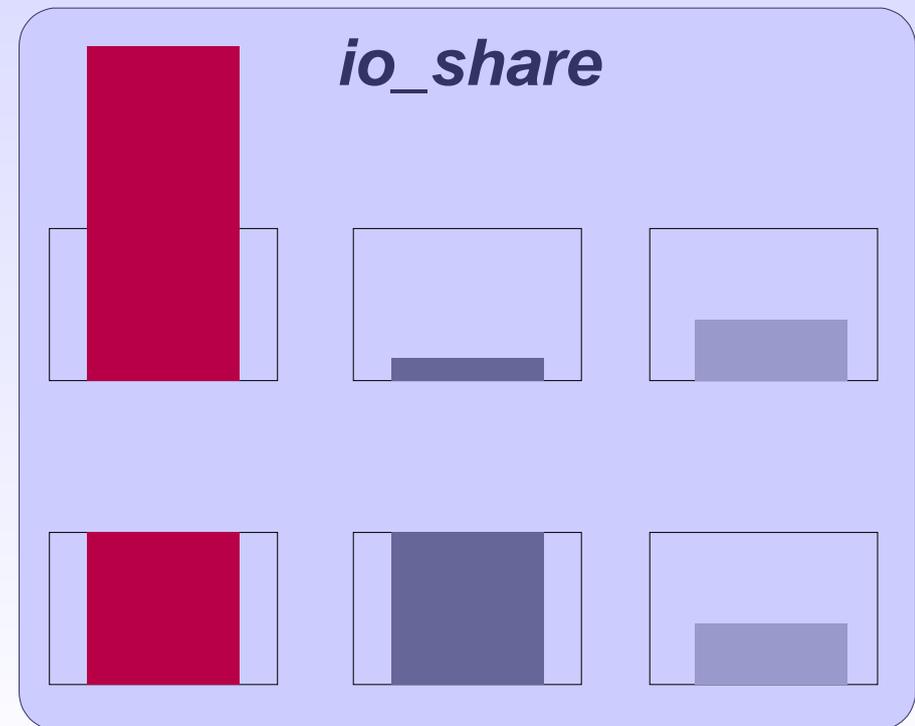
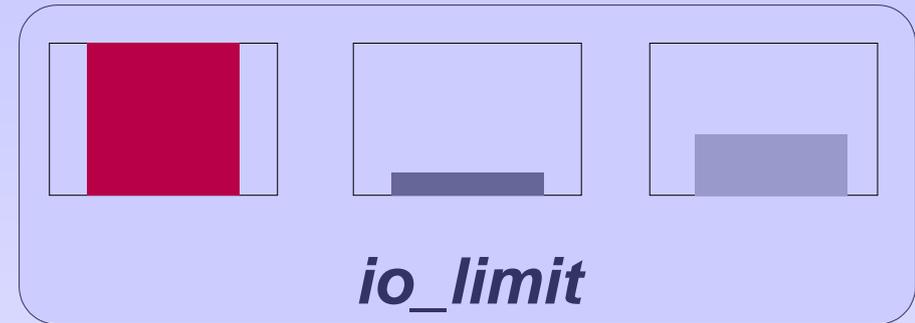
- einzelne Volumes
- Gruppen von Volumes
- Applikationen

● Wie?

- absolute Bandbreite (*io_limit*)
- adaptives Konzept (*io_share*)
- Limit für Synchronisationsvorgänge (*sync_limit*)

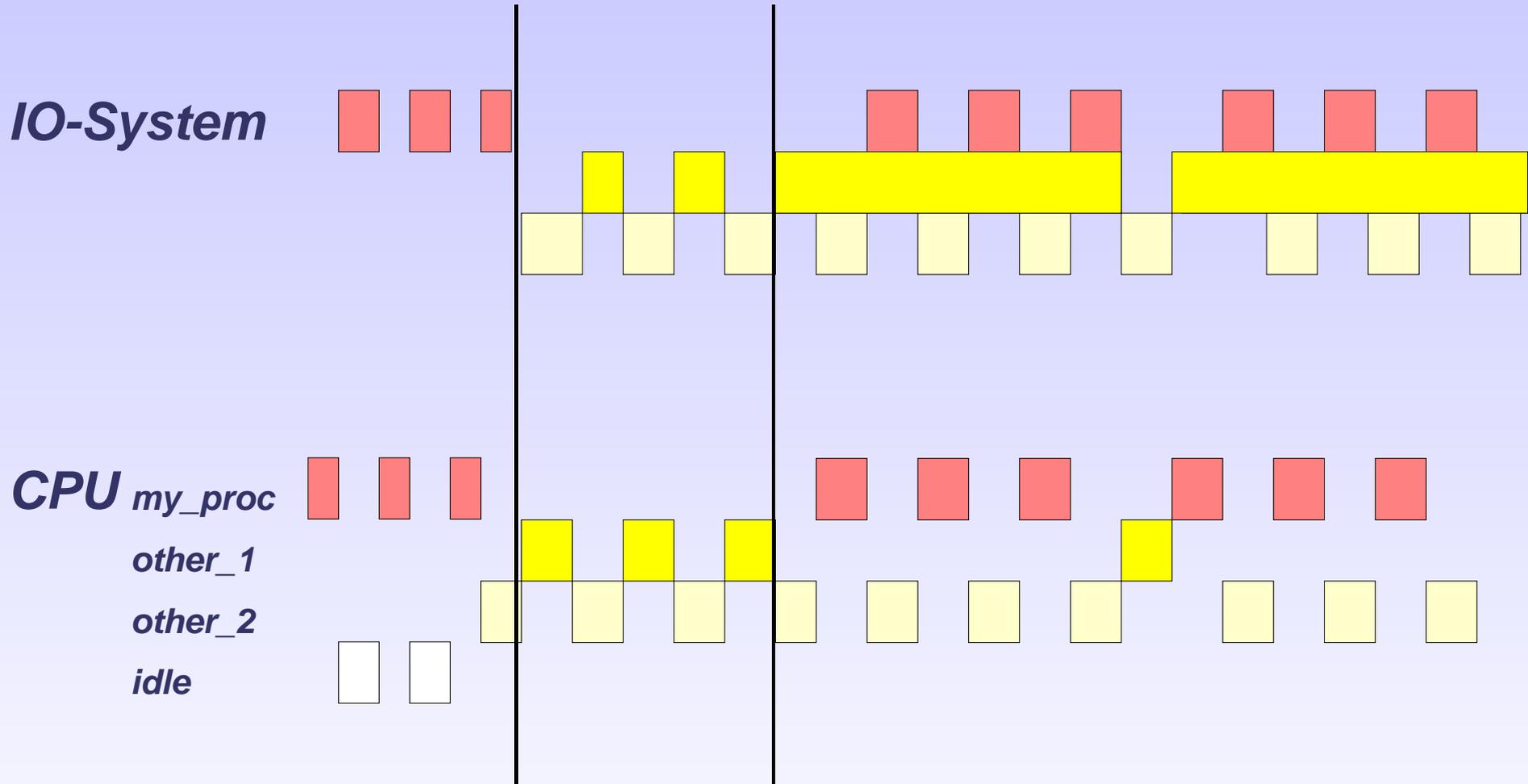
● Mit welchem Resultat?

- verbessertes Antwortzeitverhalten
- faire Verteilung von IO und CPU-Bandbreite
- reduzierte CPU-Belastung
- gesteigerter Gesamtdurchsatz



IO- und CPU-Verteilung gemeinsam steuern

Über die Wechselwirkung von IO-Performance und CPU-Belastung

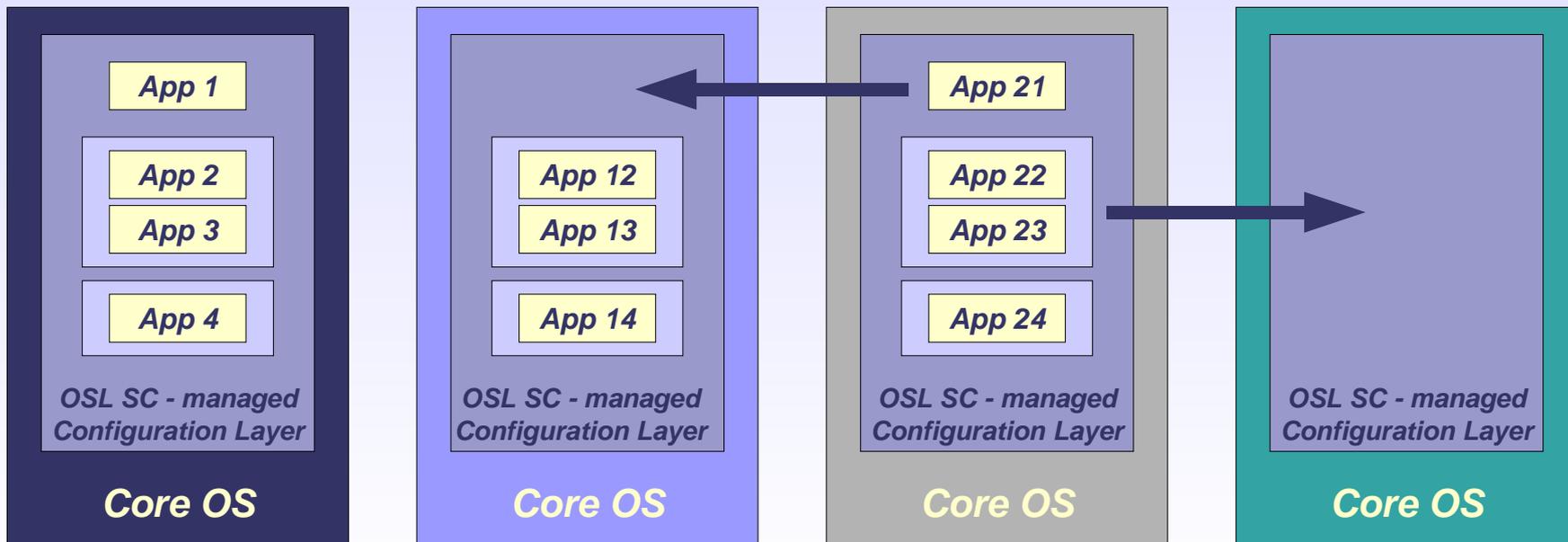


Was haben wir gewonnen?

Organisation in Applikationen ermöglicht Virtualisierte Ablaufumgebungen

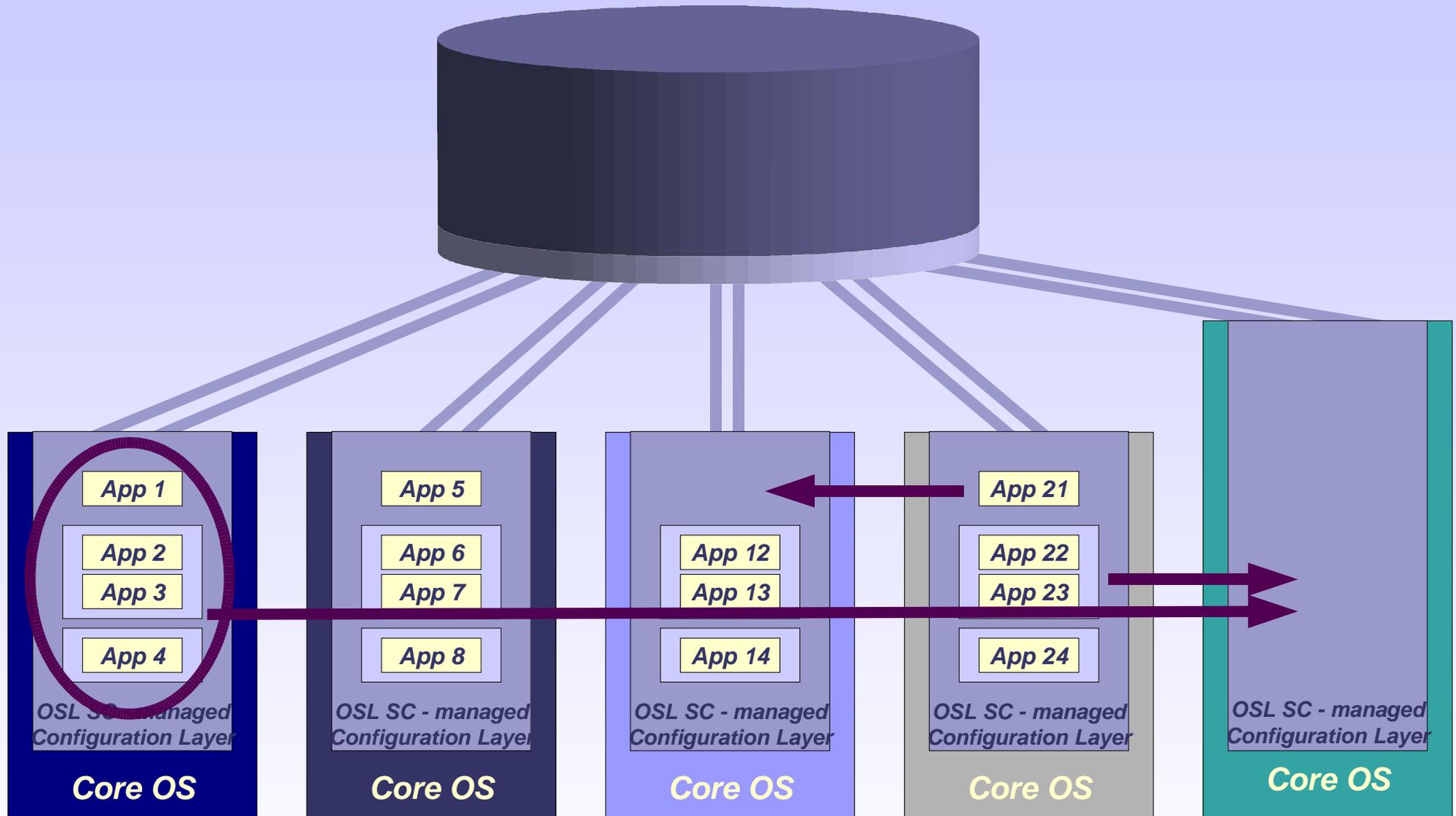
- *Global Devices und Application Resource Description*
 - Raw- und Blockdevices + Dateisysteme
 - ZFS
 - IP-Adressen und NFS
- *Globales Nutzer- und Gruppenmanagement*
- *Automatische Adaption ASCII-Konfigurationsdateien*
- *Globales Management und Migrationsdienste für Zonen*

Umgebung lebt außerhalb der Maschine weiter



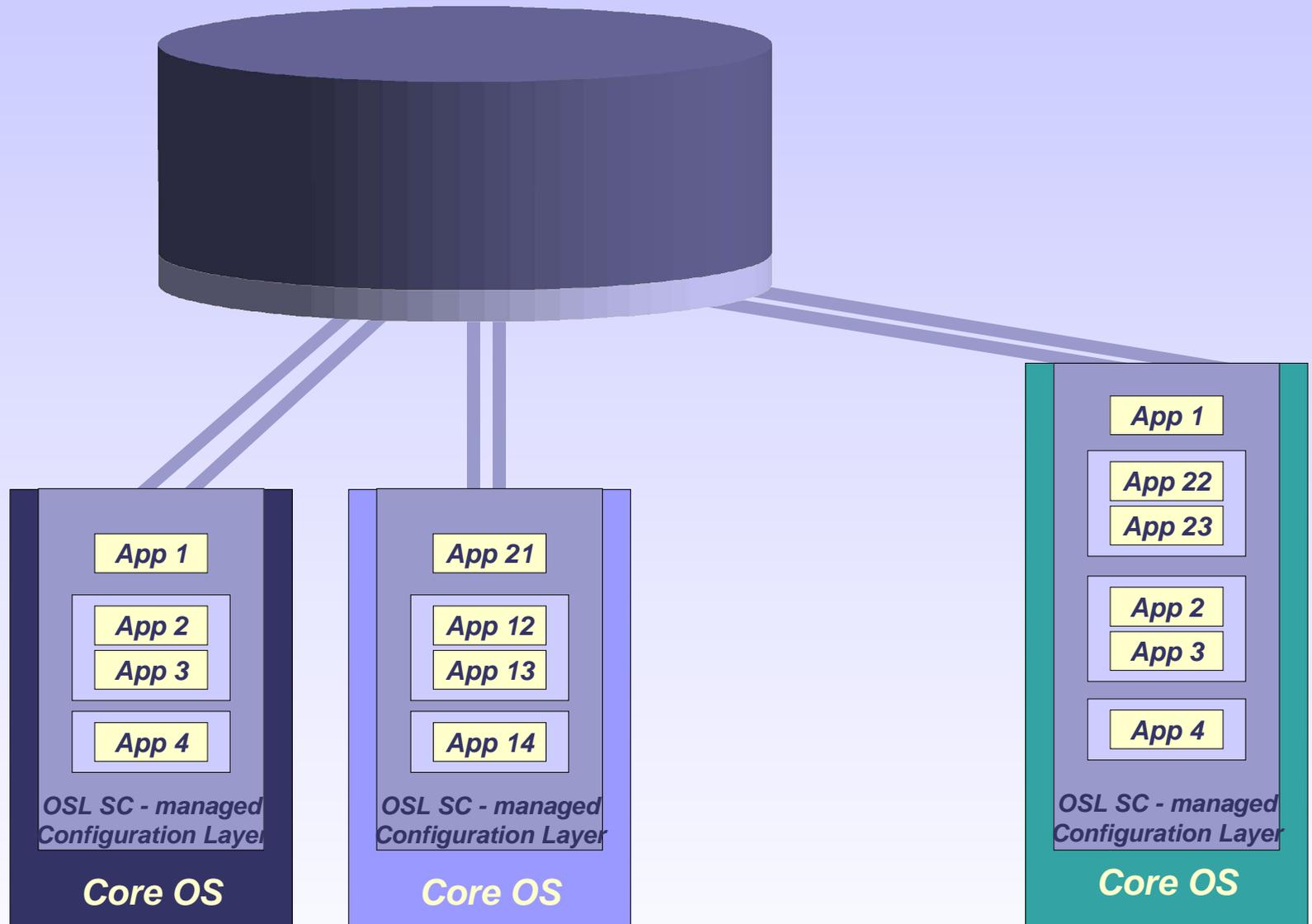
Was sind dynamische Knotengruppen?

Hardware tauschen und Cluster verändern ohne Konfigurationen anzupassen



Was sind dynamische Knotengruppen?

Hardware tauschen und Cluster verändern ohne Konfigurationen anzupassen



Core OS

Core OS

Core OS

Core OS

Applikation Awareness braucht die passende Organisation

Technologie allein kann das Problem nicht lösen

- **organisatorische Maßnahme: Encapsulated Application Setup (EAS)**

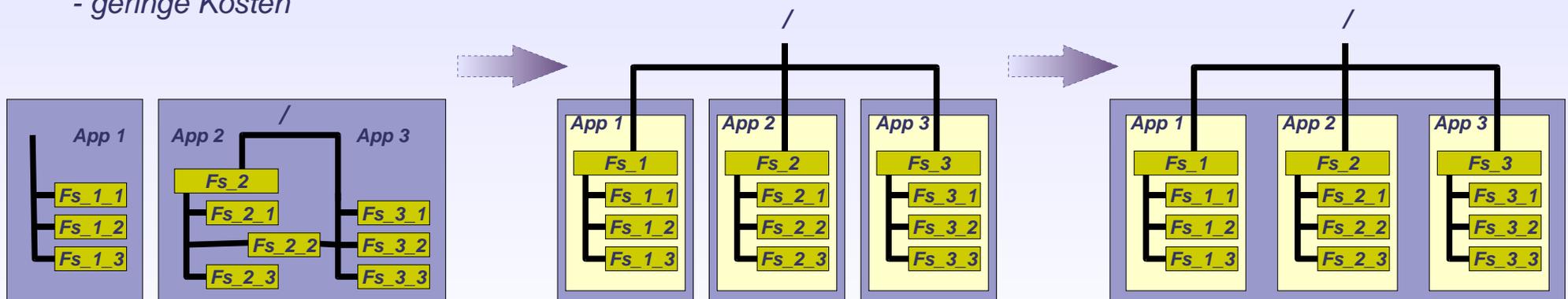
- klare Strukturierung der Installation
- kostet als organisatorische Maßnahme vergleichsweise wenig
- Voraussetzung für flexible Zuordnung von Anwendungen zu Hosts
- Datenaustausch und auch die Datensicherungswelt profitieren
- Problem: katastrophales Installationsdesign vieler Anwendungen

- **der Cluster**

- liefert das Framework für den EAS
- stellt eine global einheitliche Ablaufumgebung bereit
- sorgt für Datenkonsistenz (Konfiguration)
- kann die Anwendungen und Hochverfügbarkeit steuern und überwachen

- **beides zusammen:**

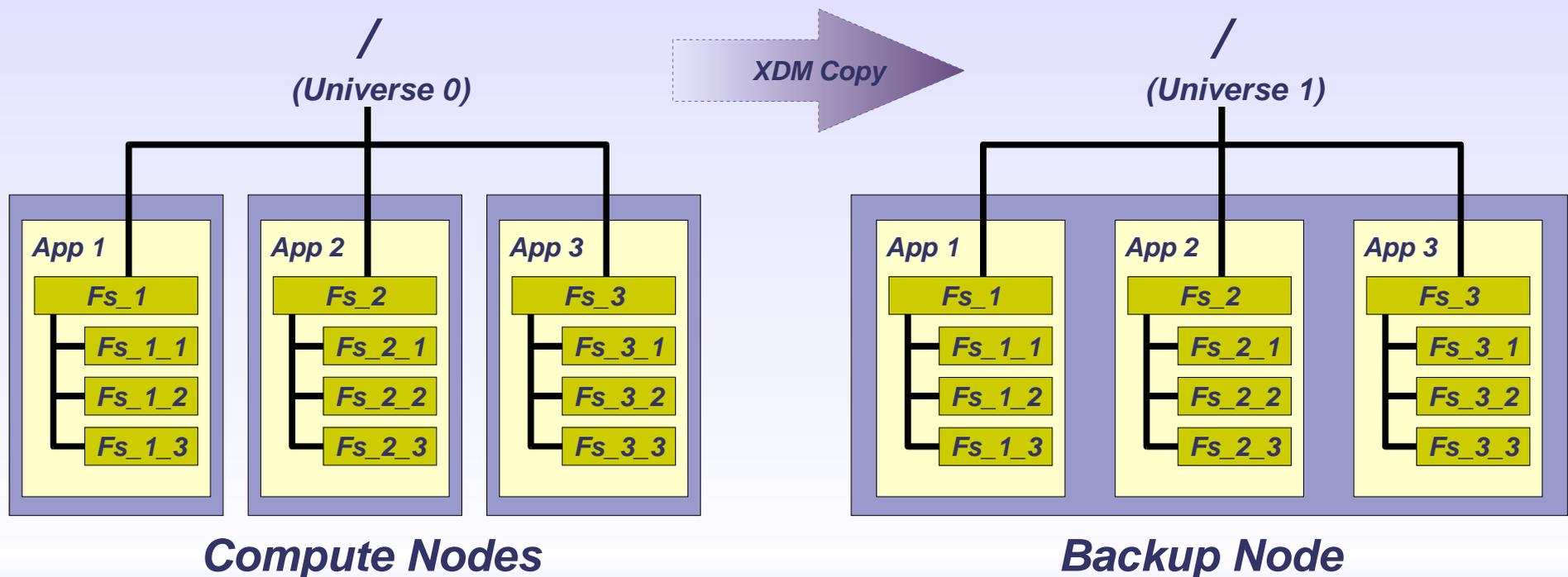
- Vereinfachung
- Flexibilität und Verfügbarkeit
- geringe Kosten



EAS – die „technologiefreie“ Virtualisierung

Encapsulated Application Setup

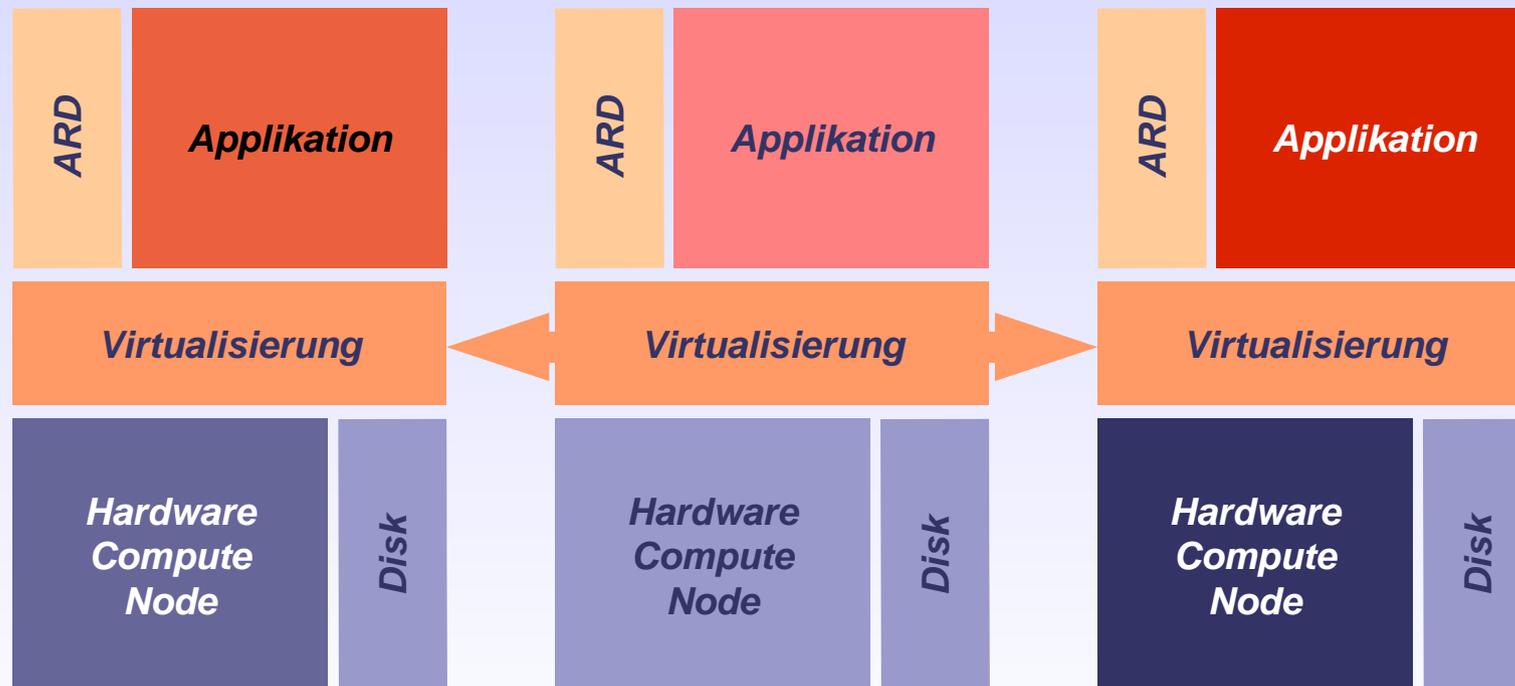
- Trennung von Knoten, Betriebssystem und Applikationen
- Global eindeutige Pfade für Applikationen (Global Root)
- Freie Beweglichkeit der Applikationen zwischen Rechnern
- Leichter Austausch von Knoten
- Vereinfachungen Backup
- Instant Restart in Verbindung mit XDM



EAS – die „technologiefreie“ Virtualisierung

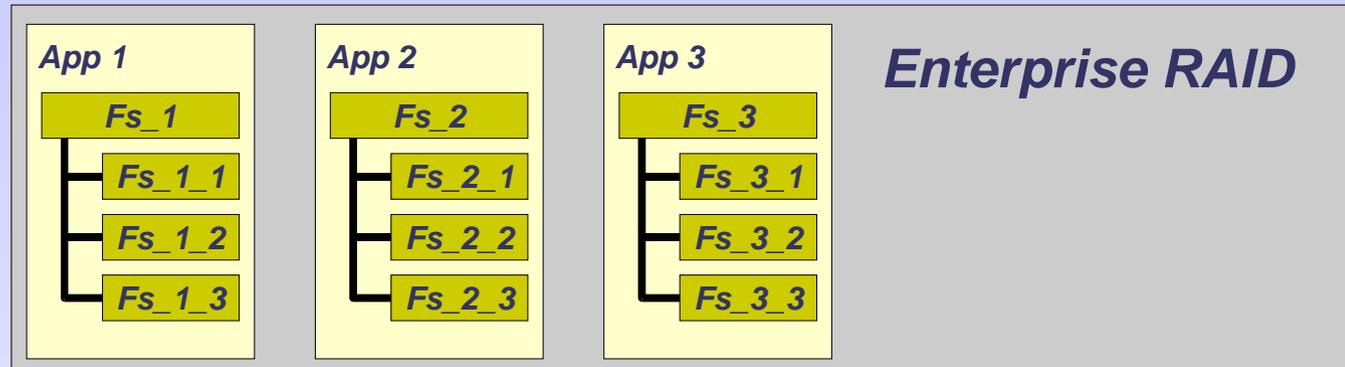
Systemkopien mit XDM

- Gemeinsam mit Kunden erarbeitete EAS-Guides und Beispiele
- Tools für interaktive oder Batch-Bearbeitung
- Weitgehend automatisierte Abläufe
- Intelligente Nachsynchronisation

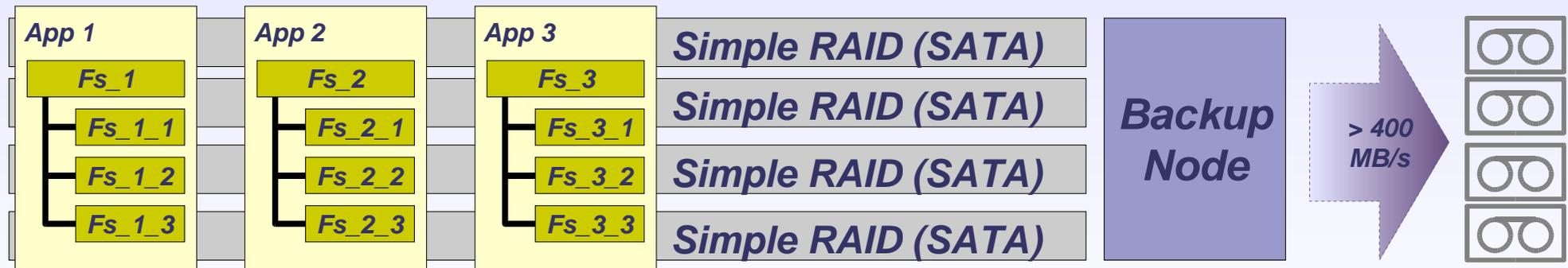


High Speed Backup mit XDM

"Abfallprodukt" existierender Anwendungsbeschreibungen



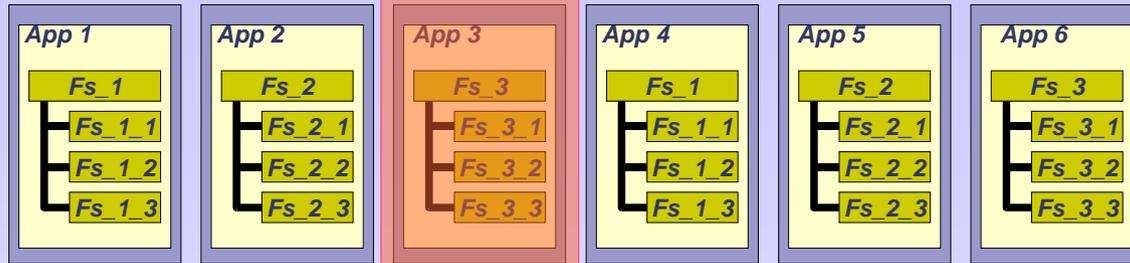
- Extrem kurzes Backup für Compute-Nodes
- High-Speed Streaming to Tape
- Keine Belastung der Compute-Nodes während Backup to Tape
- Restart-fähige Images der Applikation im Backup-Universum



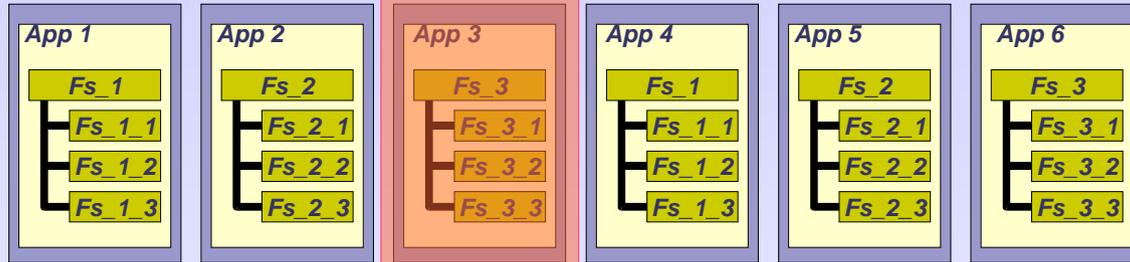
Application Aware Storage Management

Anwendungsbeschreibungen und Volume Management integriert

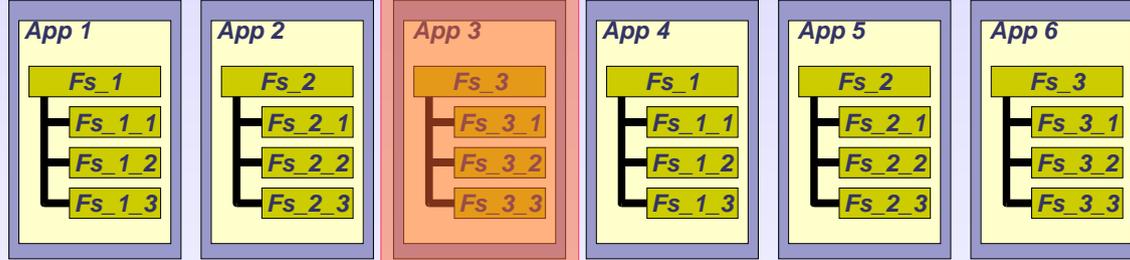
Universe 0
Produktion



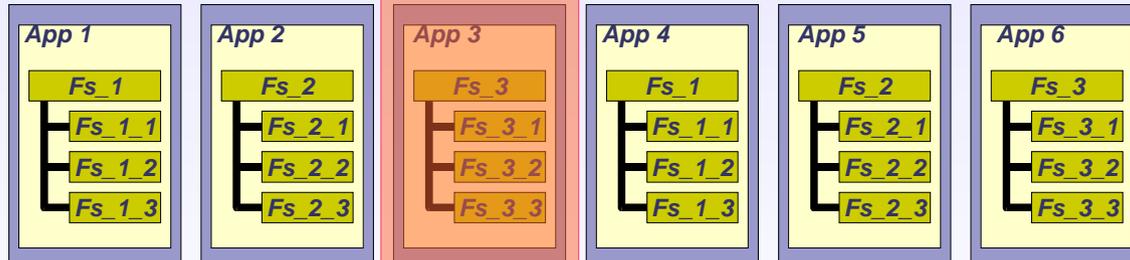
Universe 1
Backup 1



Universe 2
Backup 2



Universe 3
DR-Spiegel

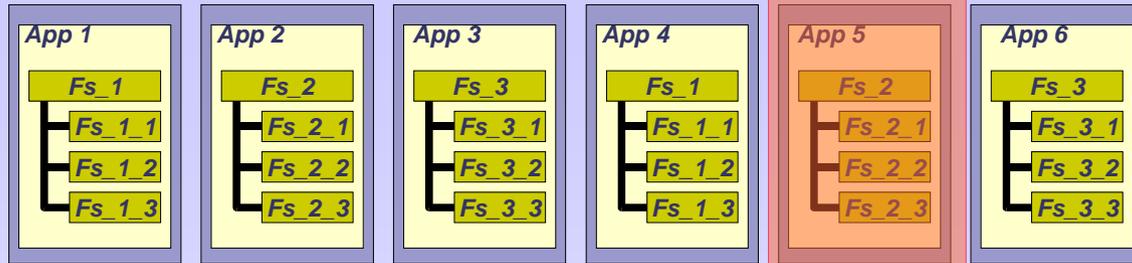


Application Specific View

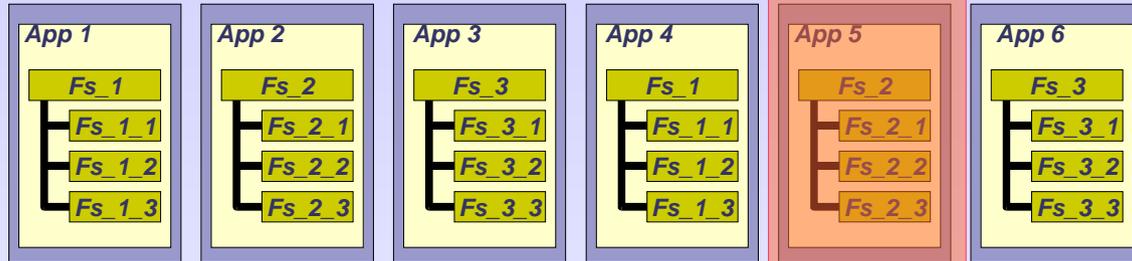
Application Aware Storage Management

Anwendungsbeschreibungen und Volume Management integriert

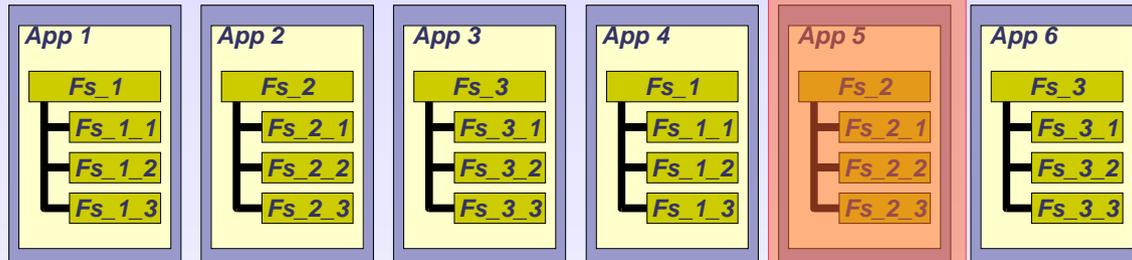
Universe 0
Produktion



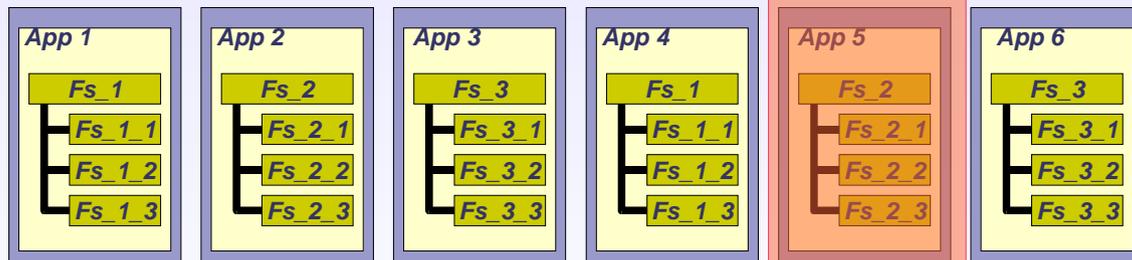
Universe 1
Backup 1



Universe 2
Backup 2



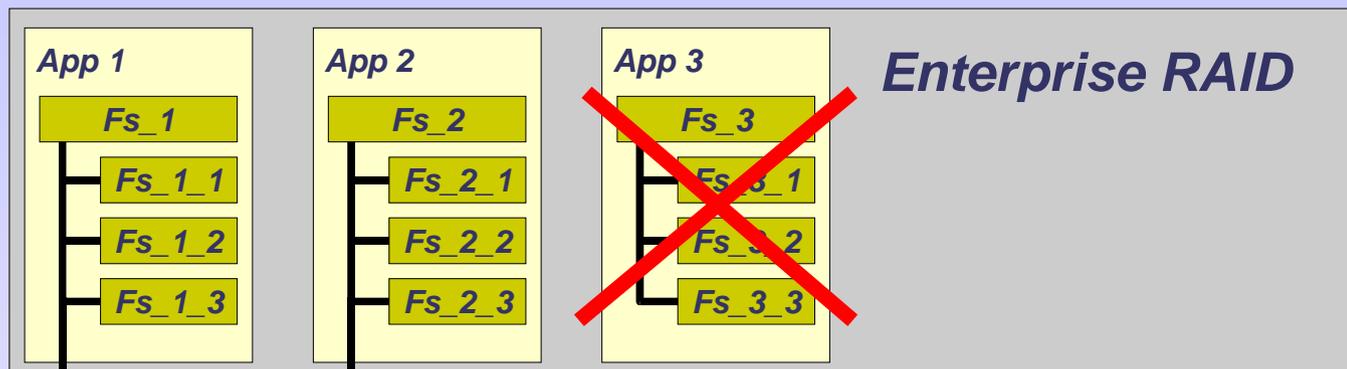
Universe 3
DR-Spiegel



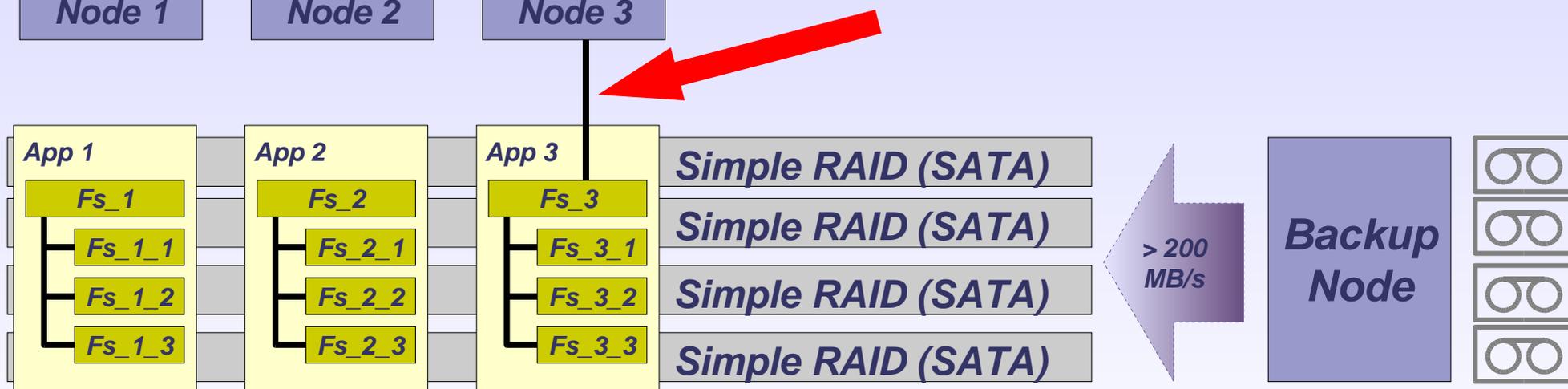
Application Specific View

Restorefreies Recovery mit XDM

Tape-Backup wird nur im Ausnahmefall benötigt

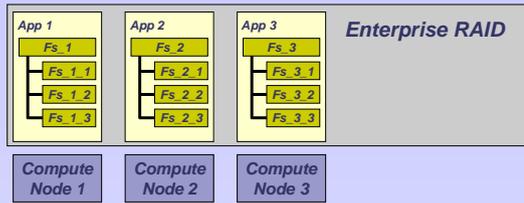


- Sofortiger Wiederanlauf
- Kein Restore vom Tape
- Preview-Möglichkeit
- Bei Bedarf High-Speed Streaming from Tape
- Resync auf Enterprise-Storage bei bereits laufender Anwendung



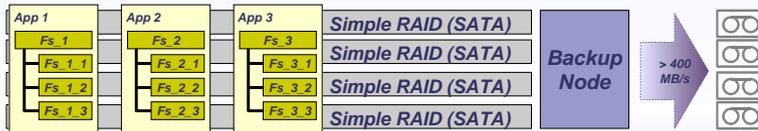
Differenzierung zu klassischen Backuplösungen

Möglichkeiten von Festplatten konsequent nutzen



- *Extrem kurzes Backup für Application Nodes*
- *Minimale CPU-Belastung auf den Application Nodes (keine Verarbeitung der Daten)*
- *Nutzung SAN statt LAN*
- *kein Backup-Client auf Application Nodes (kein dezentrales Pflegen von Konfigurationen)*
- *Atomarer Backup – Dauer: NULL damit konsistenter Zustand*
- *Restartfähige Images der Applikation damit extrem schneller Wiederanlauf*
- *SW für Tape-Backup nur auf DASI-Server*
- *Zentrale Administration*
- *extreme Durchsätze bei Tape-Backup/-Restore möglich*
- *niedrige Anforderungen an Backup-RAID*
 - ermöglicht SATA mit hoher Dichte
 - niedriger Platzbedarf
 - kürzere Backup-Zeiten
 - reduzierter Stromverbrauch / Wärmeabgabe

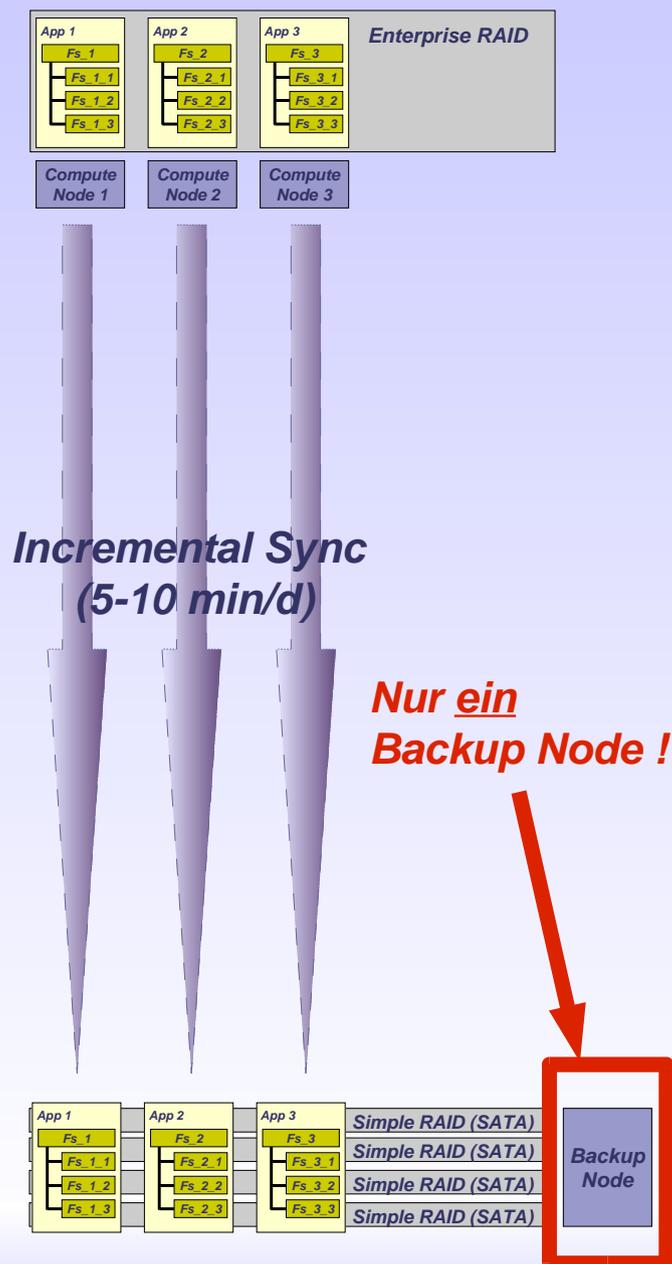
**Incremental Sync
(5-10 min/d)**



- *adaptive Fähigkeiten bzw. "selbstlernend"*
- *Integration mit HV*
- *leicht zu DR-Umgebung ausbaubar*

Integration mit Backup to Tape

Beispielimplementierung für Legato Networker

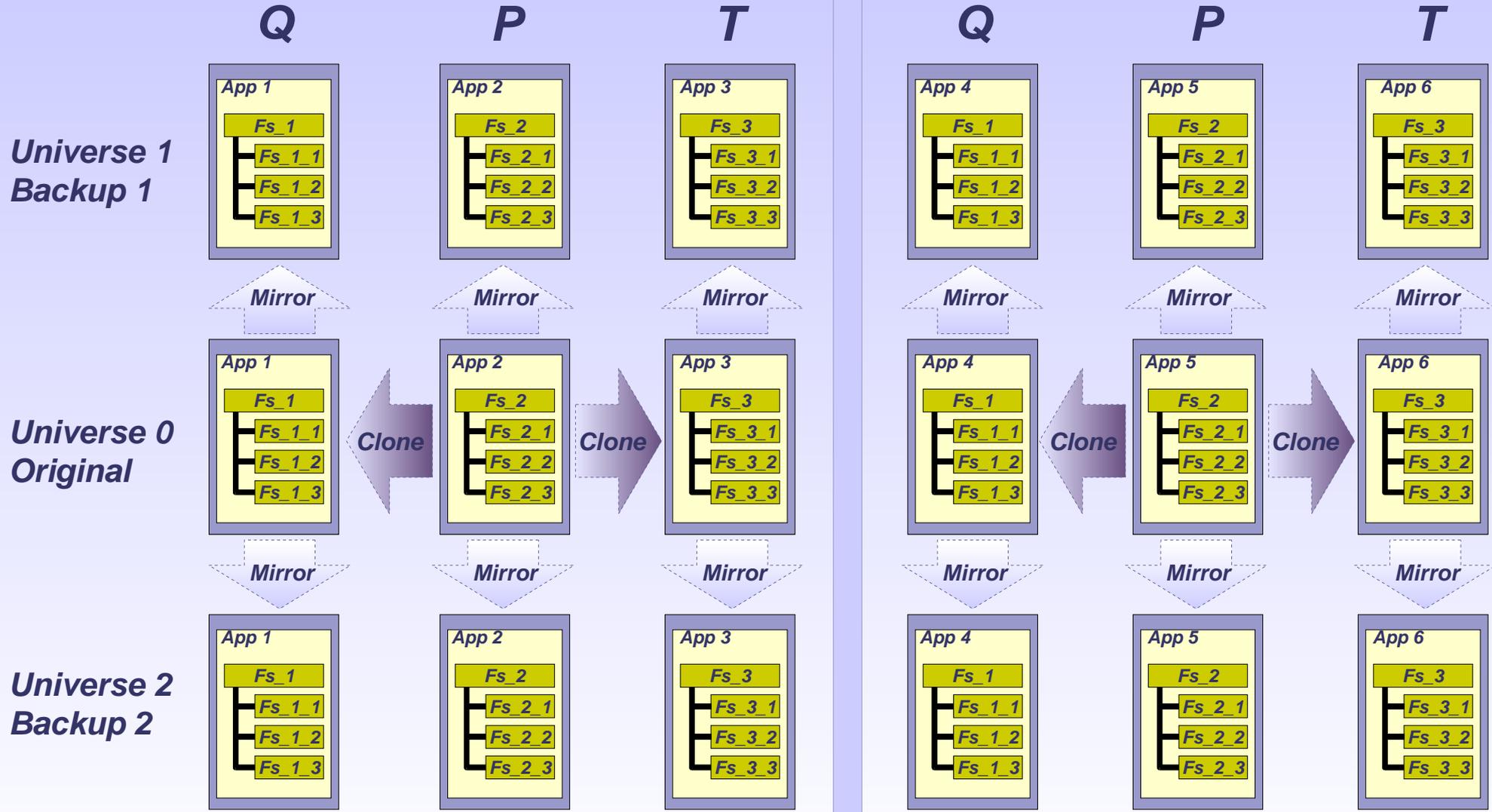


Was bietet die Integrationslösung von OSL?

- kombiniert B2D mit Bandsicherung
- applikationsorientiertes Verfahren
- sofortiger Neustart von Backup-Disk (kein Tape-Restore)
- Steuerung der Sicherungen über Networker-GUI oder CLI
- integriertes Pre- und Postprocessing
- mehrere Sicherungen pro Tag möglich
- differenzierte Erfolgskontrolle über die Networker-Indizes
- Aufzeichnung von Dateisystem-Informationen für Restore
- eindeutige Identifikation kompletter Sicherungen
- einfacher Restore kompletter Sicherungen
- weitere Funktionen für Oracle
 - Tool für Logrestore und Roll Forward
 - Archivierungslösung
- LAN-free Backup
- „cluster aware“, d.h. kein Eingriff bei Umschaltungen nötig
- automatische Gleichverteilung der Plattenlast
- hohe Durchsätze – VTL unter diesem Aspekt entbehrlich

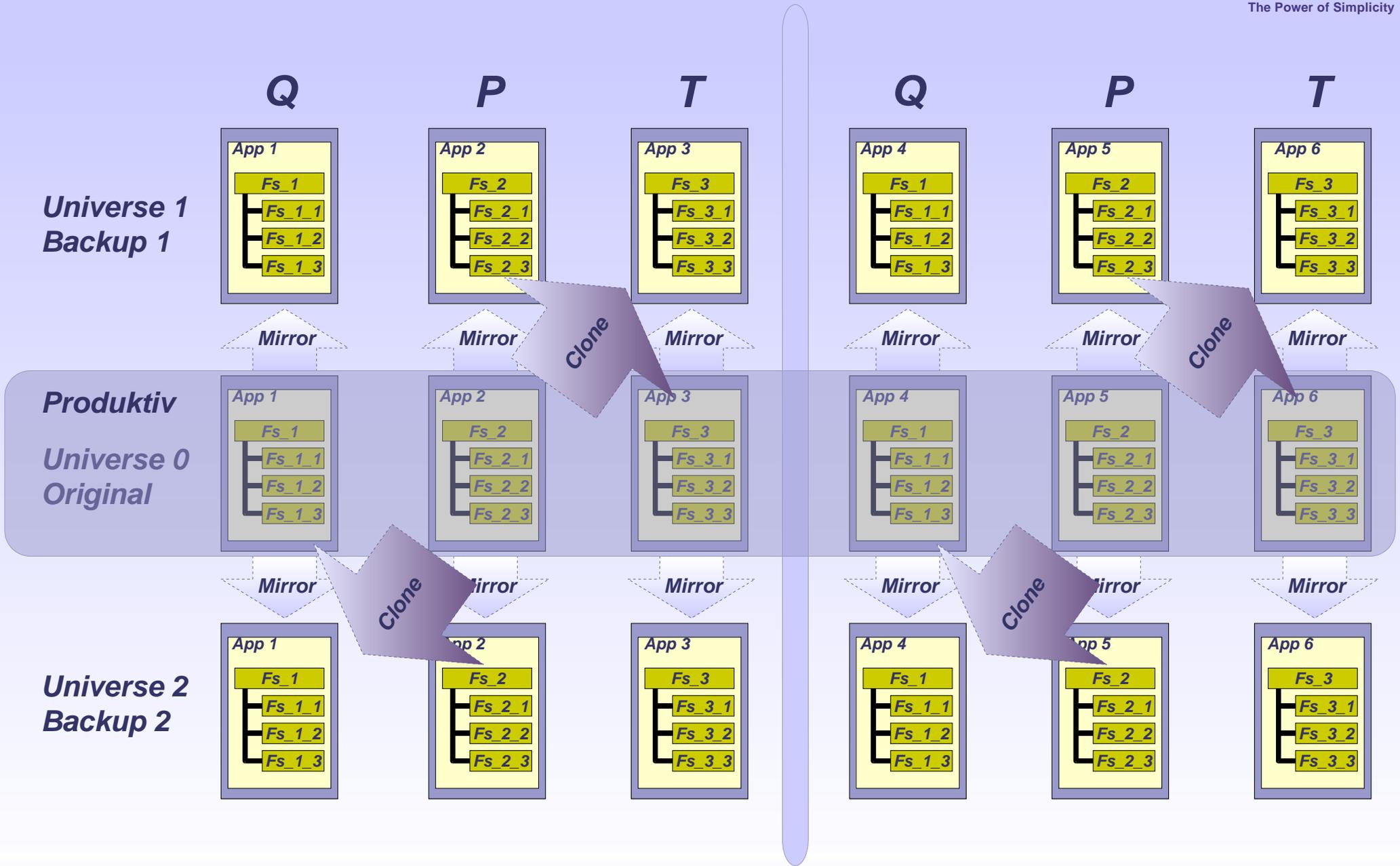
Spiegeln und Clonen

Für jeden Anwendungsfall den passenden Replikationstyp



Spiegeln und Clonen

Es geht auch anders ...



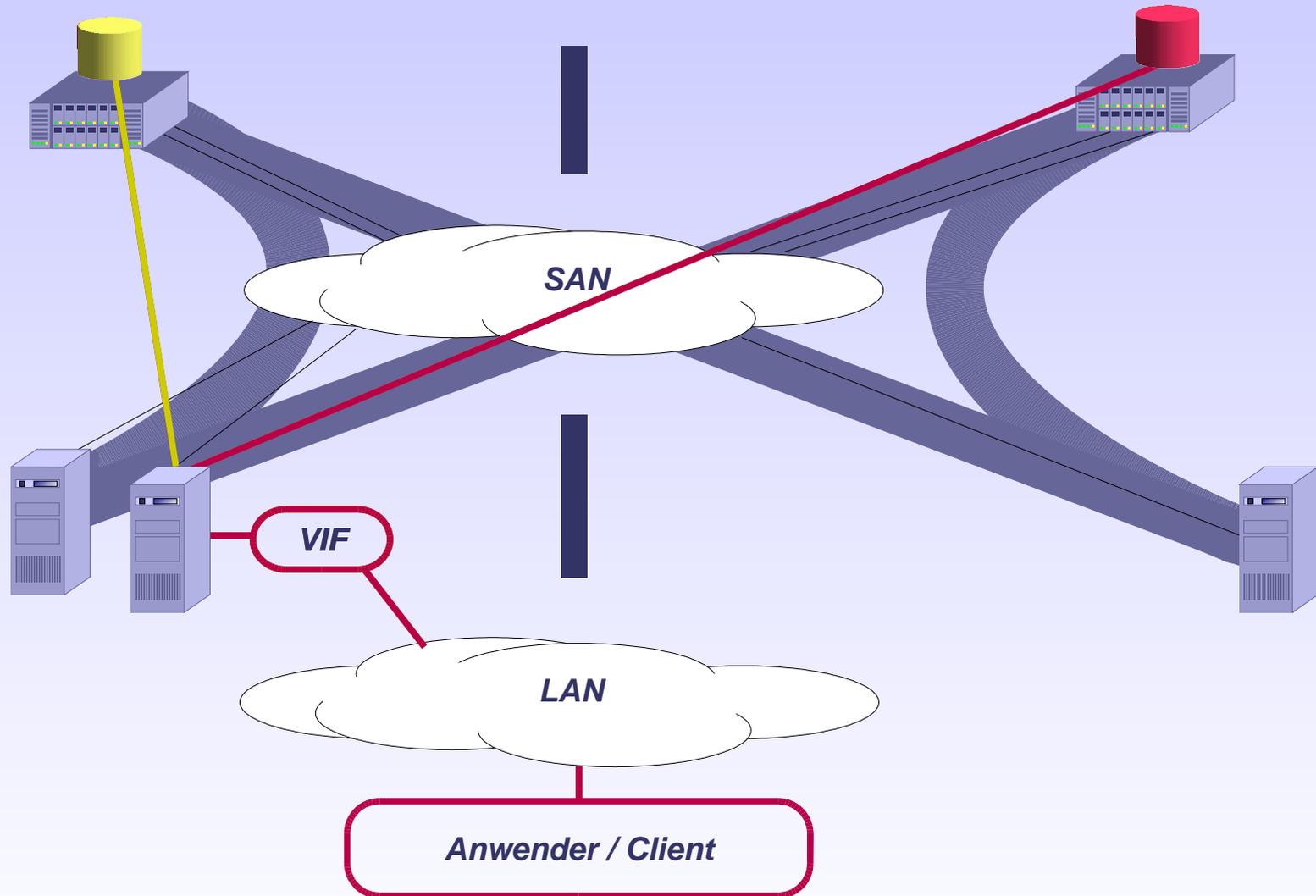
Spiegeln und Clonen

Es geht auch anders ...



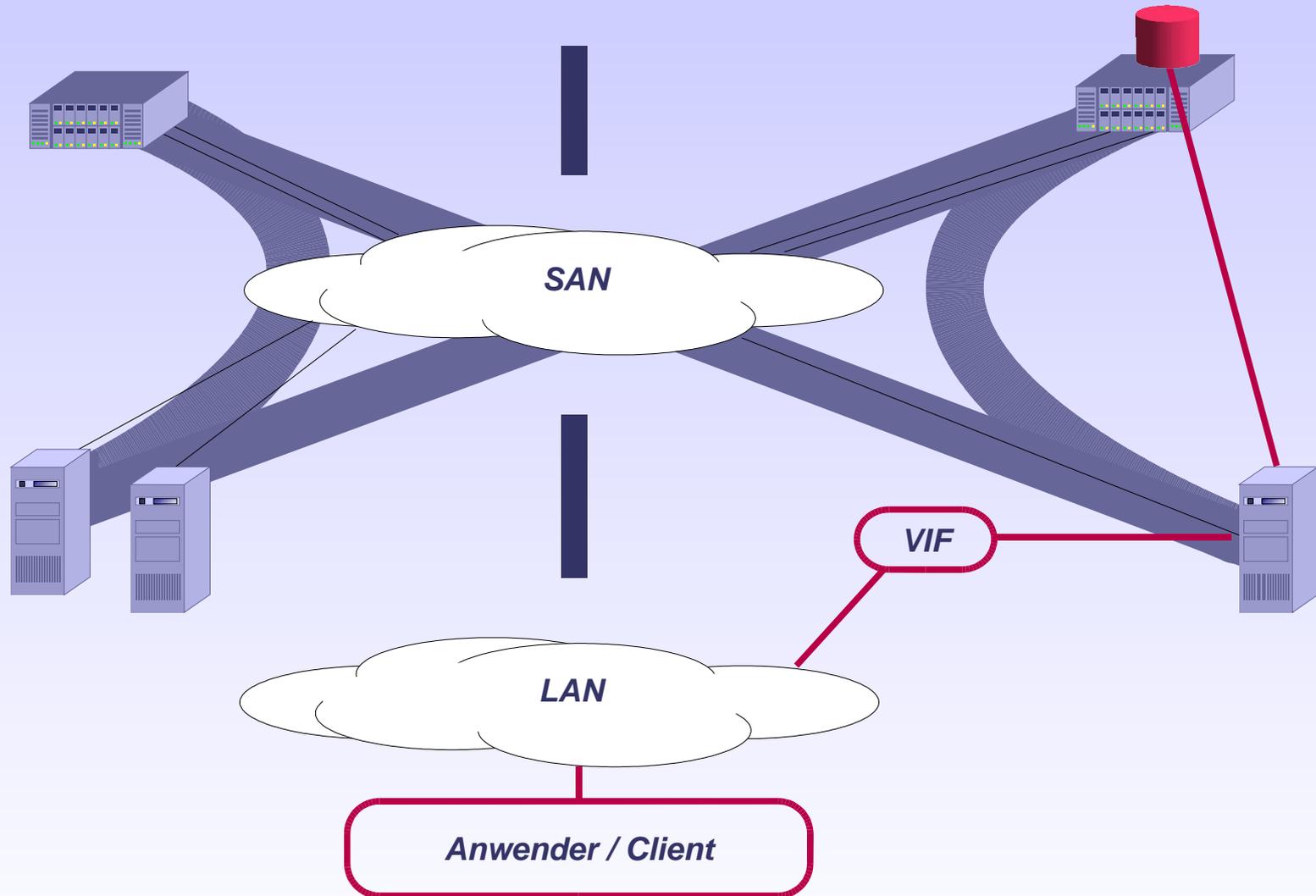
Alles zusammen

Clusterfähige Storage Virtualisierung, Backup, HV, Disaster Recovery



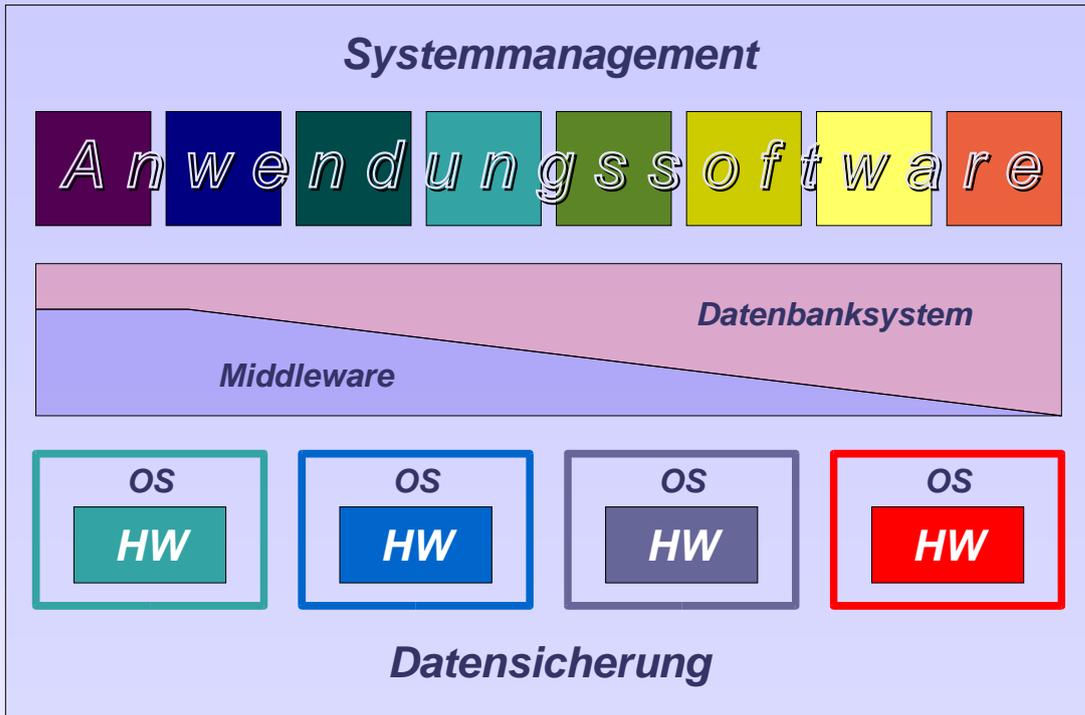
Alles zusammen

Clusterfähige Storage Virtualisierung, Backup, HV, Disaster Recovery

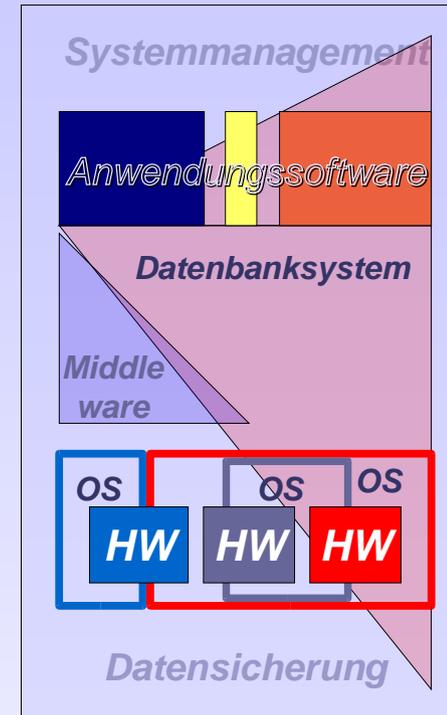


Und sonst – was kommt sonst im RZ?

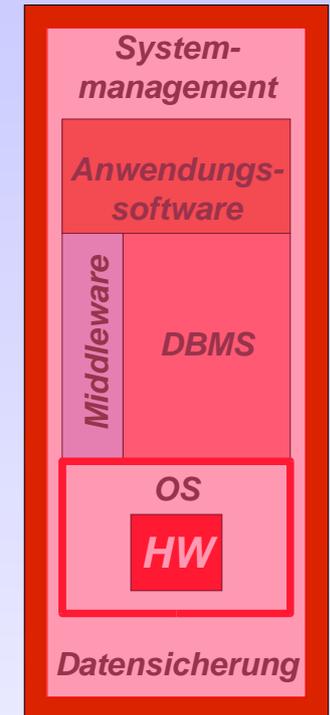
Klassische Systematik der Komponenten im kommerziellen RZ wird unschärfer



- Spezialisierung in allen Schichten
- vielfältige Betriebssysteme, standardisierte Schnittstellen
- plattformübergreifende Softwarelösungen
z. B. für Datensicherung und Systemmanagement
- Anwendungen oft portabel



- Branchen-SW wird seltener
- weniger Betriebssysteme
- proprietäre Lösungen aus Anwendungssoftware/
Datenbanken erobern Systemmgmt.
/ DASI / OS
- integrierende Komponente anders
als bei Heimanwendungen nicht
mehr das OS, sondern die
Applikation



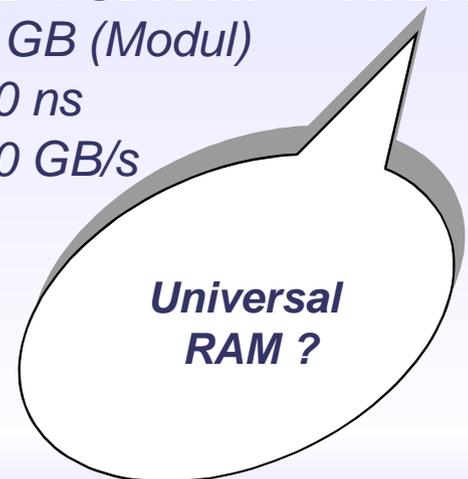
Vision:
all out of a
(red) box

Hardware und Systemsoftware

Es deuten sich Umbrüche an

- Peripherie steigert Rechenleistung dramatisch (Grafikkarten, Netzwerkkarten), gewisse Sättigung absehbar
 - Steigerung CPU-Leistung vorwiegend über Multicore- und CMT-Technologie
-> Druck auf Redesign der Software (Zunahme Komplexität)
 - mit aktuellen Systemen oft Überhang an CPU-Leistung
-
- SSD-Disks in der Einführung
 - F(e)RAM heute bereits bei 16MB pro Modul (Toshiba, Fujitsu)

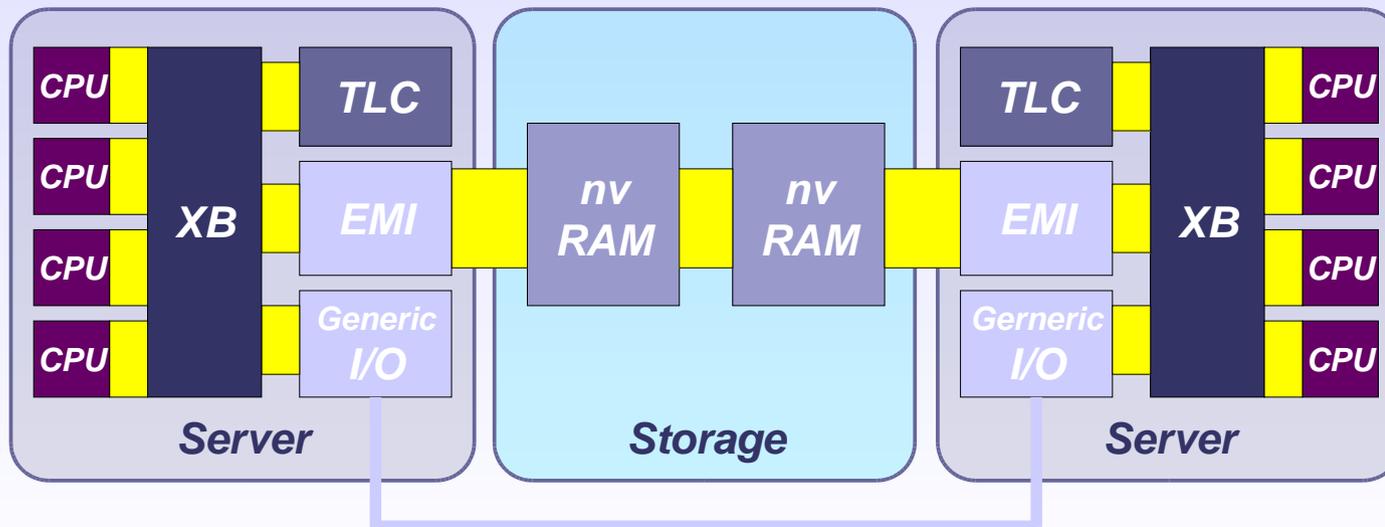
	HD	SSD	FeRAM	DDR SDRAM	MRAM
Kapazität	2 TB	>512 GB	16 MB (Modul)	4 GB (Modul)	
Zugriff	4 ms	0,2 ms	0,05 μ s	10 ns	
Transfer	140 MB/s	240 MB/s	1,6 GB/s (DDR2)	40 GB/s	



Konsequenzen der Speicher-Revolution

Vieles ist zu überdenken

- nicht mehr Platte sondern Anschlußtechnologie wird zum Flaschenhals
- Renaissance paralleler Zugriffsmethoden?
- klassische Serialisierung von Zugriffen verliert an Bedeutung, das hat Auswirkungen auf:
 - Filesysteme
 - Datensicherungsverfahren und Snapshots
 - Design von Datenbanken (Logs, Archive, Parallele Datenbanken)
 - Design von Applikationen
 - ganze RZ-Konzepte (Cluster, HA, Clouds)
- was wird aus SANs?
- wie sieht der Datacenter-Rechner der Zukunft aus?



Und in Zukunft

Alles was besteht ...

- *informiert bleiben und Freiheitsgrade bewahren*
- *in konkreter Hardware möglichst kurzfristig festlegen*
- *auf offene und flexible Systeme (HW und Software) setzen*
- *zukunftsstaugliche Konzeptionen erarbeiten
(HW tauschen kann einfach sein, Prozesse ändern eher kaum)*

***Nicht allein Wissen, sondern vor allem Lernfähigkeit
sichert langfristig den Erfolg!***