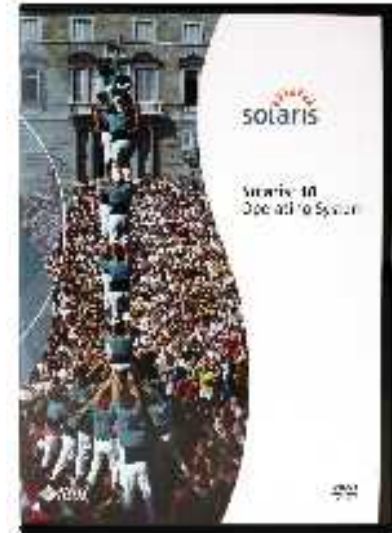




OpenStorage Technologien und Solaris

Detlef Drewanz

Senior Systems Engineer
Ambassador Operating Systems
Sun Microsystems GmbH



opensolaris™



SOLARIS OPERATING SYSTEM

BE BRILLIANT FASTER



The Solaris ecosystem consists of the Solaris OS release plus the open source OpenSolaris OS release. The free and open Solaris OS is a proven, industry leading operating system with features designed to save your business time and money; the OpenSolaris community is where the next generation of Solaris is being built, and the OpenSolaris OS is where to find the latest results of these efforts.

Already Using Solaris? Start Here:



**For Solaris administrators,
internal/third-party enterprise developers**

What you get:

- Access to thousands of 3rd party software applications
- A 10 year+ support life cycle to create rock-solid, long-term deployments
- The latest updates to supplement my existing Solaris deployments
- Unmatched enterprise support offerings

[Learn More](#)

New to Solaris? Start Here:



**For students, Web 2.0 developers,
open source OS developers**

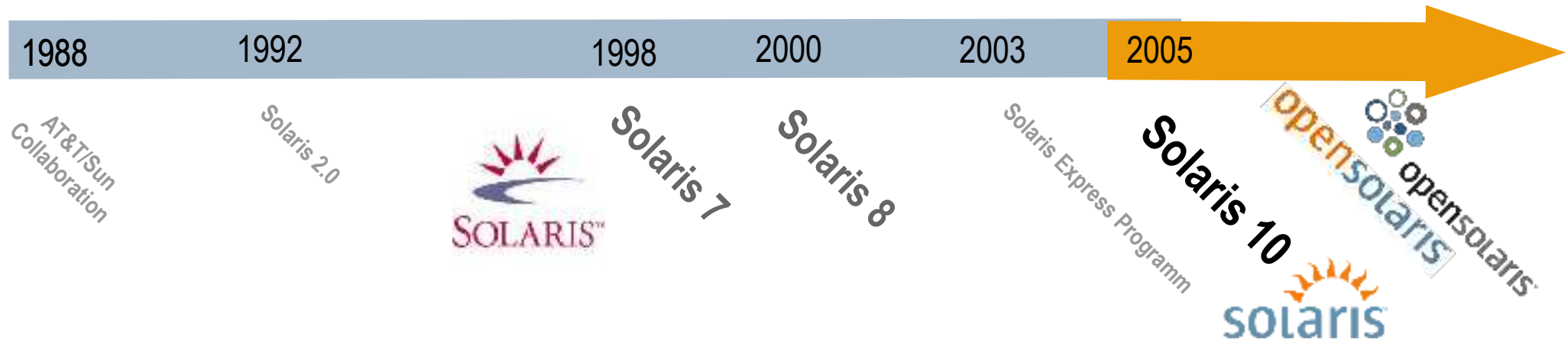
What you get:

- The latest Web 2.0 free/open source applications
- Frequent releases; the latest Solaris features under development
- LiveCD, the easy way to evaluate Solaris technologies risk free
- Comprehensive repository of packages

[Learn More](#)

20 Jahre Solaris

Kompatibel, Stabil, Performant - per Design



- Besondere Produktlebenszyklen
 - > Für Entwickler und Kunden
 - > Funktionalitäten erweitern anstatt zu ersetzen
 - > Garantierte Kompatibilität
- Vereinigt Innovation und Kompatibilität
- Flexibel einsetzbar durch umfangreiche Eigenschaften

Sun Solaris Strategie

**Innovative
Technologie**



**Offen, Open Source
und kostenfrei**



**Multi-Plattform
Sun / non-Sun
x86 / SPARC**

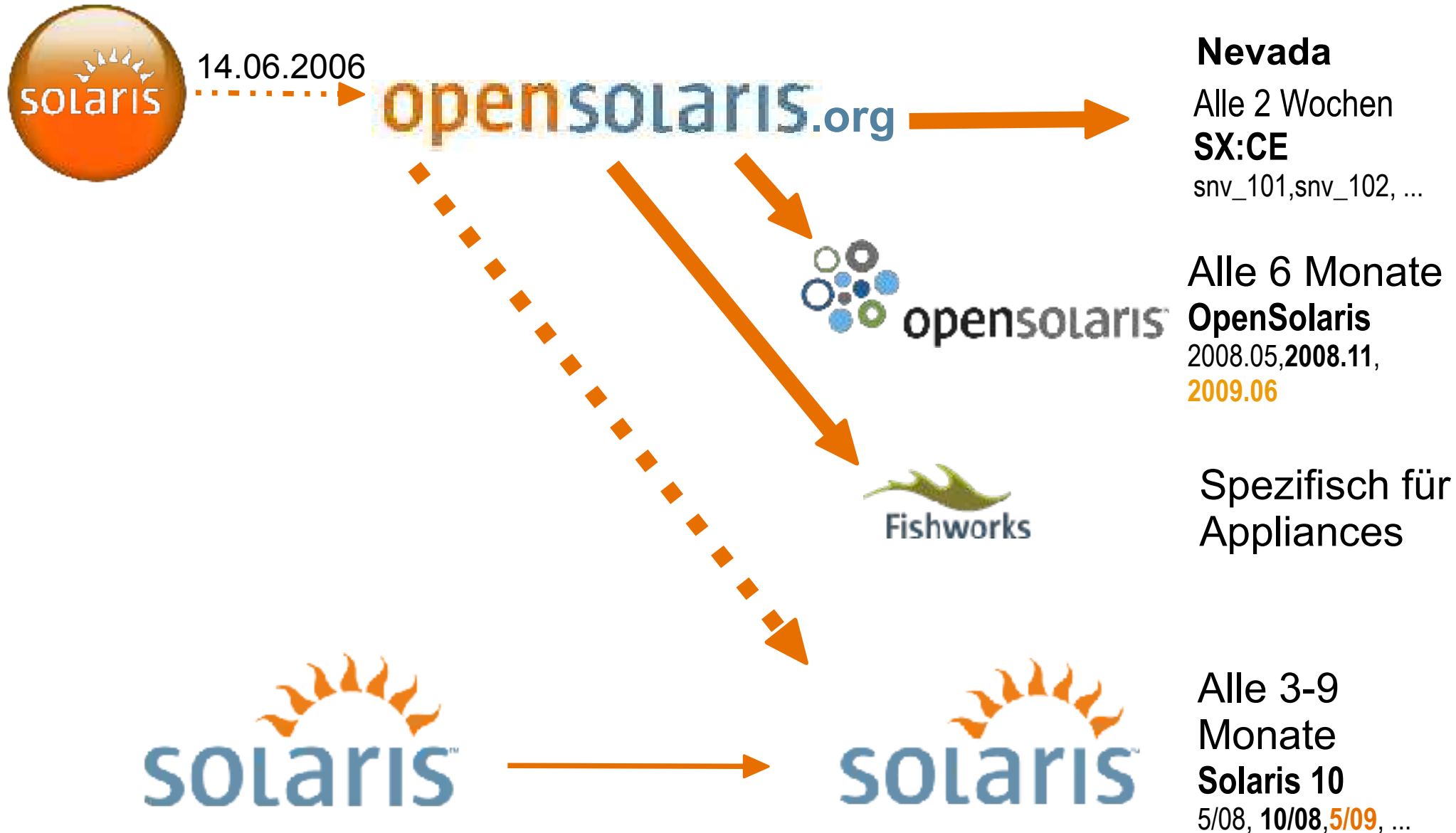


opensolaris™



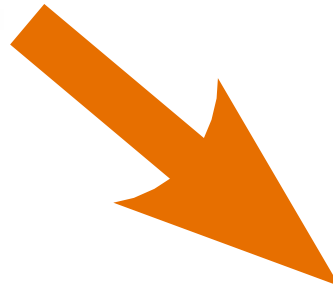
“ otevřený 열린 مفتوح ανοικτό মুক্ত libre
 मुक्त öppen open פתוח 开放的
 開 オープン ಪುಕ್ತ libero nyílt
 放的 வெளிப்படை açık ::::: livre 的
 的 открытый açık ::::: livre offen

Innovation in Solaris über OpenSolaris



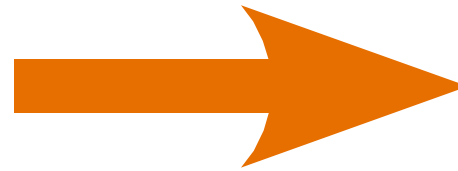


Aus vielen Teilen kann man
ein Auto bauen.



Ein Auto ?





Das Endprodukt ist entscheidend !

- Innovative Ideen
- Gereiftes Design
- Erprobte Setups
- Zuverlässiger und Langfristiger Support

Duales Release-Konzept

Eine Solaris Plattform, zwei Distributionen



solaris™

- Unternehmen, geschäftskritische Anwendungen
- Unübertroffene Unterstützung
>10 Jahre und mehr
- Lange Release-Zyklen
>3-5 Jahre mit viertel-jährlichen Updates

opensolaris™

- Entwickler, Web2.0, Early Adopter
- Neueste Innovationen
- Support-Optionen bis 7x24
- Kurze Release-Zyklen
>Updates alle 6 Monate

Solaris 10 Update Highlights

- Solaris 10 10/08 (Update 6)
 - > Newboot SPARC
 - > ZFS Root & Boot
 - > Diverse weitere ZFS Erweiterungen
 - > Integration ZFS und LiveUpgrade
- Solaris 10 5/09 (Update 7)
 - > Optimierungen für den Xeon 5500 “Nehalem”
 - > Überarbeitete iSCSI Target Implementierung
 - > Neue Treiber etc.
- Siehe auch: What's New
 - docs.sun.com > Solaris 10 > Solaris 10 What's New Collection

ZFS



ZFS Design Prinzipien

- **Pooled Storage**

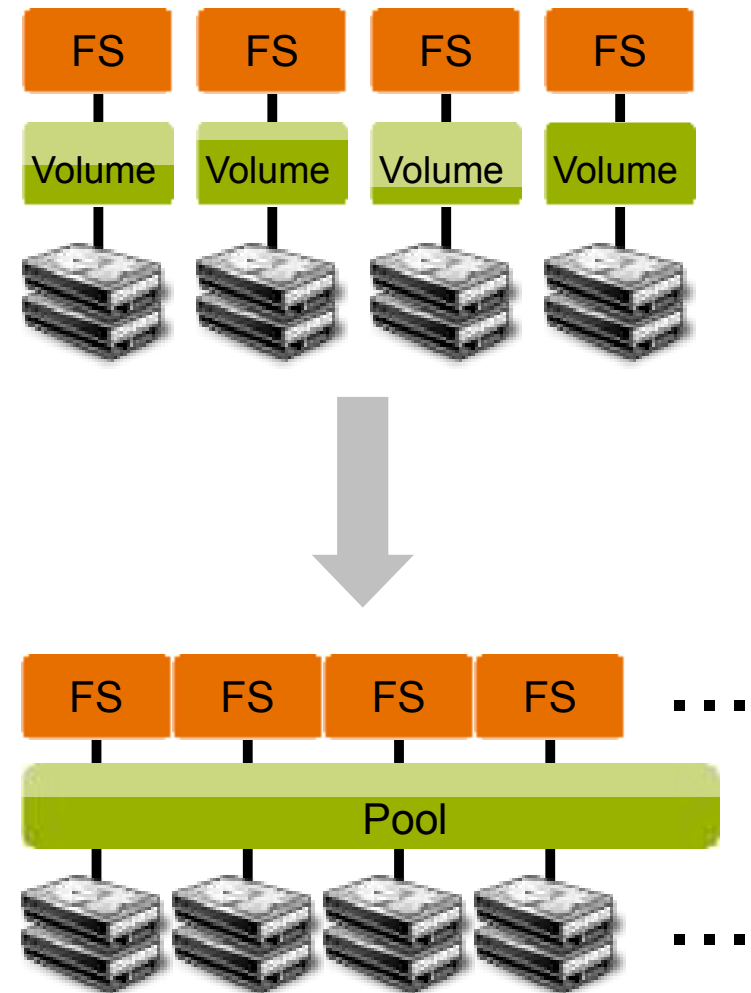
- > Filesysteme über Platten/LUNs verteilt
- > Separierte 'read/write caches'
- > Sehr einfache Administration

- **End-to-end Daten Integrität**

- > Prüfsummen für alle Blöcke
- > „Selbstheilung“ bei Spiegeln
- > RAID5 und RAID6 im Filesystem

- **Transaktionsorientiert**

- > Kein fsck mehr
- > Online Konsistenz Checks
- > Hohe Schreibperformance (*next free space*)
- > Copy-on-write

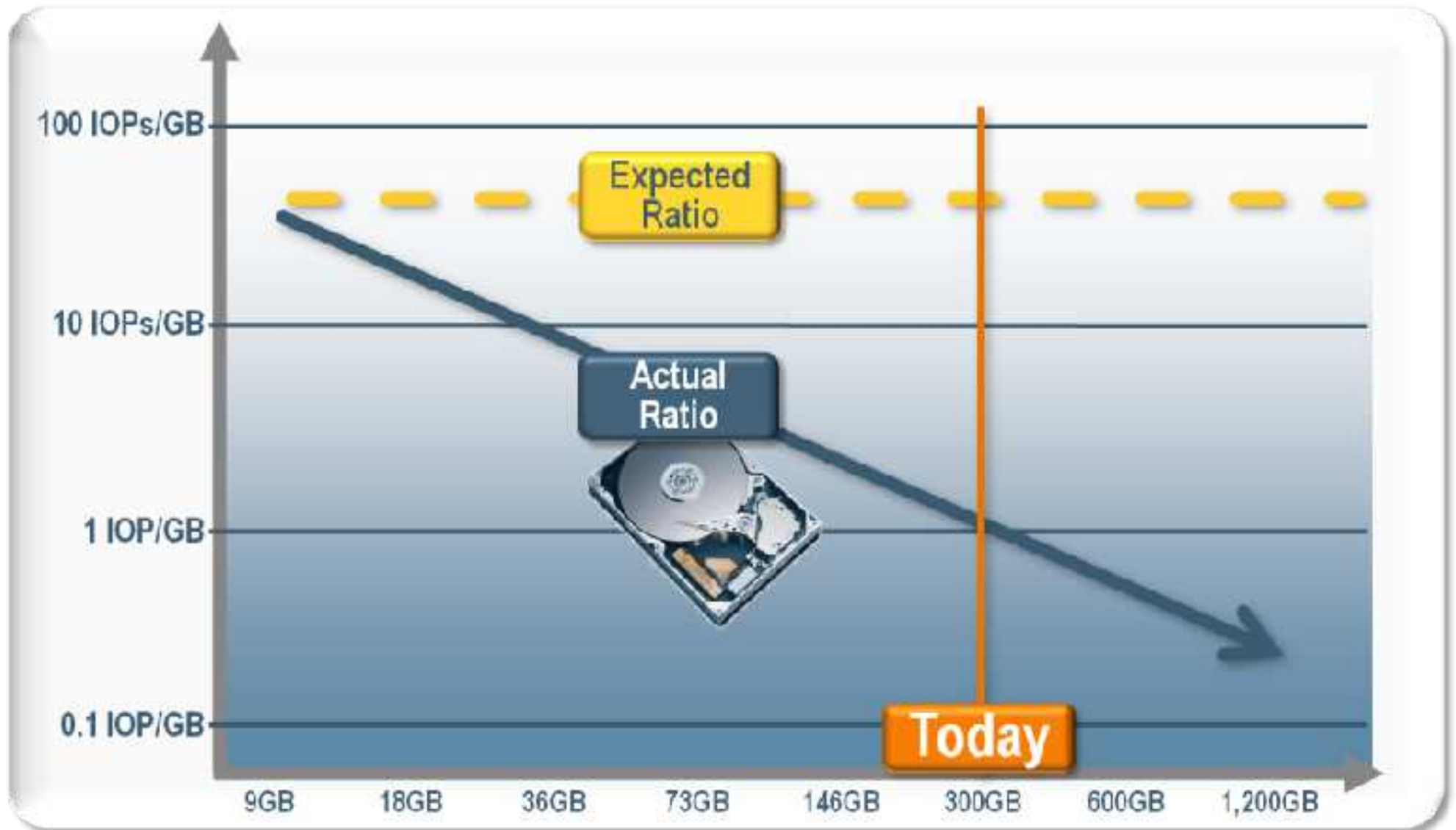


Trends bei plattenbasierten Speichern

- Schnellere Laufwerke
 - > 7200 rpm → 10k rpm → 15k rpm
- Performance-Metriken
 - > Durchsatz, IOPS, Latenz
- (NV-)RAM für bessere Schreib-Performance

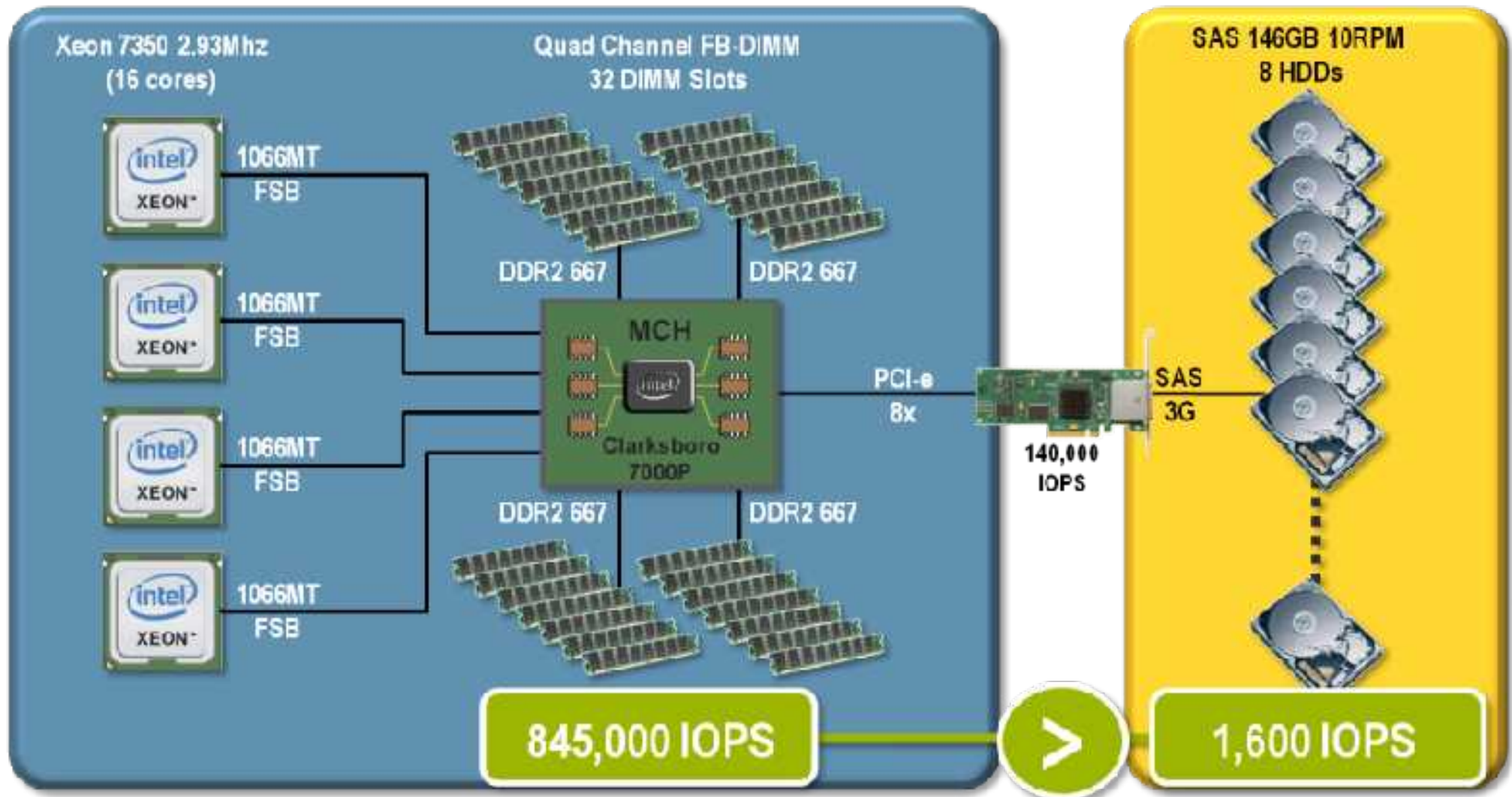


Der Fluch der steigenden Kapazität oder "Früher war alles besser"



Die CPU/Storage Lücke

SunFire x4450 ist 500 mal schneller als 8 HDDs



Was tun ?

Noch mehr Festplatten kaufen ?

Noch größere Storage Arrays kaufen ?

Noch mehr Strom verbrauchen ?

Flash Memory

- Ein Retter in der Not?
 - > Drei Größenordnungen schneller
 - > Niedrige Leistungsaufnahme
 - > Rapide fallende Preise
- Warum jetzt?
 - > 1980 erfunden von Dr. Fujuo Masuoka bei Toshiba
 - > 1988/1989 erste Produkte von Intel (NOR) und Toshiba (NAND)
 - > Heute 92% der Produktion im Consumer-Markt
 - MP3-Player, Speicherkarten für Kameras, Memory-Sticks
 - > Flash SSDs tauchen 2007 im Enterprise-Markt auf.
 - SLC versus MLC



Anatomie eines Solid State

- Eine SSD besteht aus:
 - > A) Controller
 - > B) DRAM
 - > C) Flash
- Flash ist aus Speicherblöcken aufgebaut
- Schreiben funktioniert nur blockweise !
- Controller ist zuständig für
 - > Wear Leveling (Schreibzyklen zählen)
 - > Bit Error Correction
 - > Bad Block Mapping
 - > Disk Interface (SATA, SAS, FC)



Vergleich Disk versus SSD

Ein IOPS kostet 70 mal weniger in der SSD



High Speed Disk

180 Write IOPS

320 Read IOPS

300 GB

18 Watt

2.43 \$ pro IOPS



Enterprise SSD

7.000 Write IOPS

35.000 Read IOPS

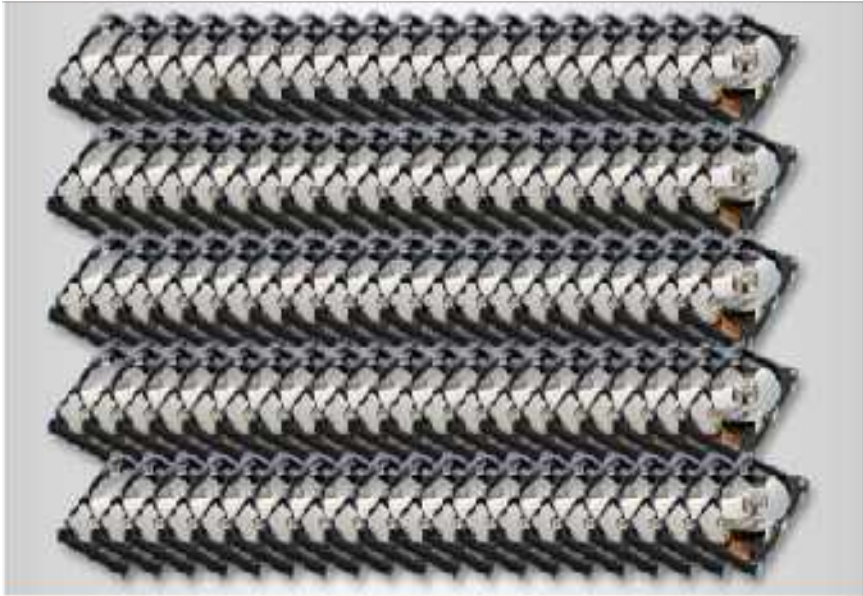
32 GB

3 Watt

0.04 \$ pro IOPS

Ein Beispiel

100 Enterprise HDDs



Capacity: 30TB
 Performance: 30K IOPS
 Cap/Op-X: \$55,000 – 1.75kWh

Hybrid Storage Pool



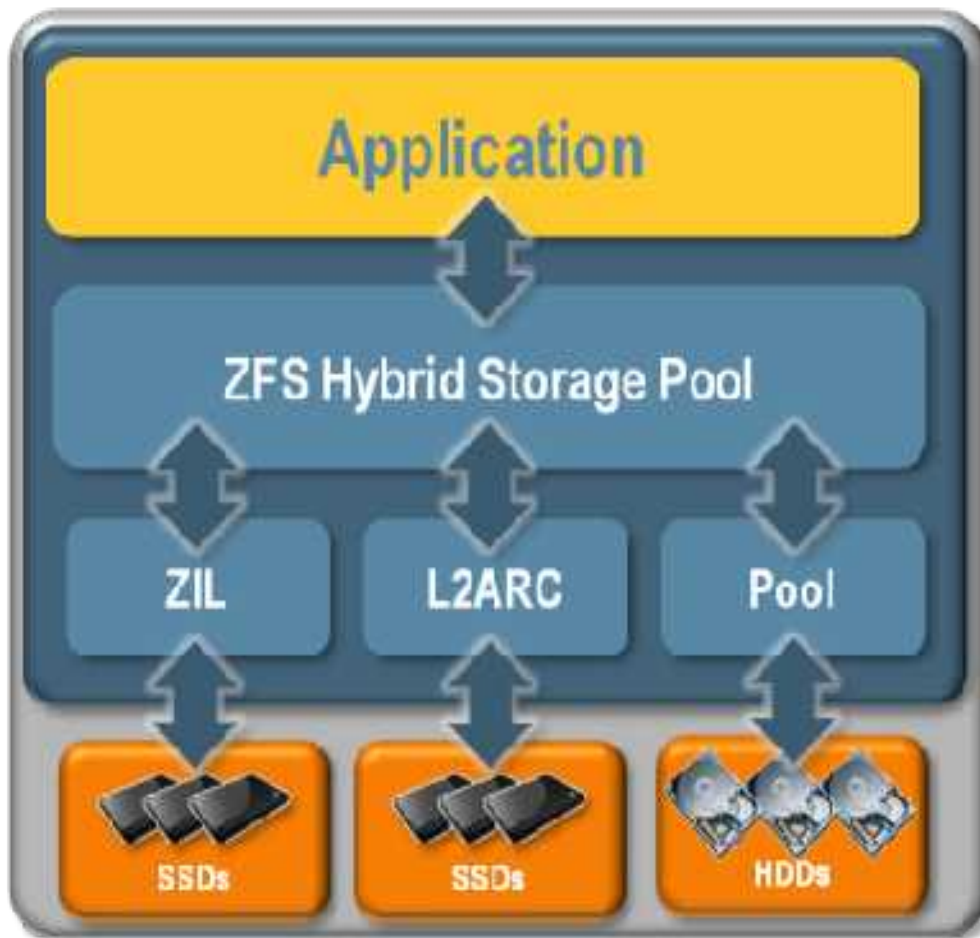
Capacity: 30TB
 Performance: 30K IOPS
 Cap/Op-X: \$6,040 - 0.392kWh

Die neue Speicherhierarchie

- SSDs heute kein 1:1 Ersatz für HDDs
 - > bezogen auf die Kapazität wesentlich teurer
 - > geringere Speicherdichte
 - > “eigene” Performance-Charakteristika
- Nutzung als neue Hierarchieebene



Der Hybride Storage Pool mit ZFS



- Speicher wird transparent verwaltet als ein Pool mit einer optimierten Hierarchie
- ZFS verbindet die Attribute der unterschiedlichen Device-Typen zu einer optimierten Einheit

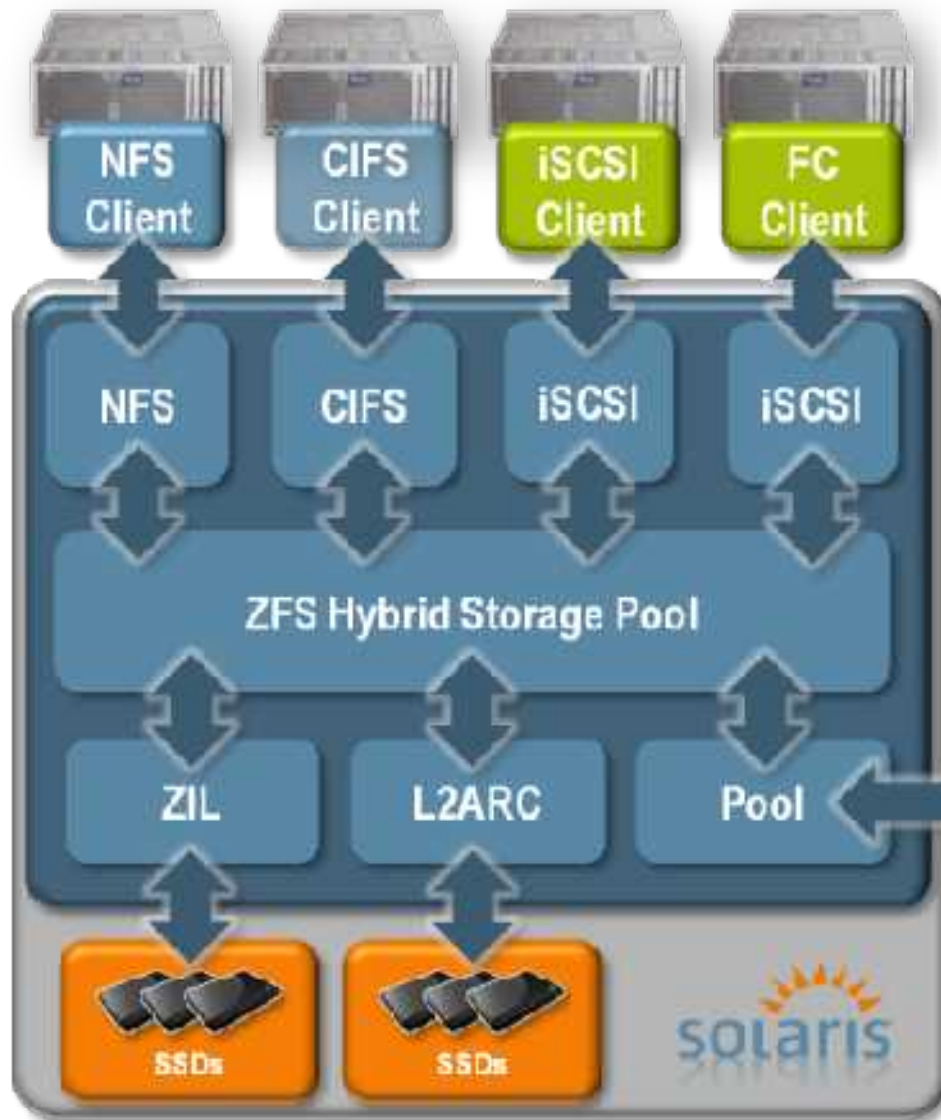
„Logzilla“: schreib-optimierte SSDs

- “Logzilla”
 - > Bezeichnung bei Sun für schreib-optimierte SSDs
- ZIL: **ZFS** Intent **L**og
- Problem: ODSYNC Workload (order synchronous writes) wartet bei jedem Schreib-IO auf das Ausschreiben auf Platten(n)
 - > DBMS, NFS
- Lösung: Umleiten der ODSYNC Schreiboperationen auf eine schreib-optimierte SSD reduziert Latenz
- Separierter ZIL in Solaris 10 10/08 und OpenSolaris

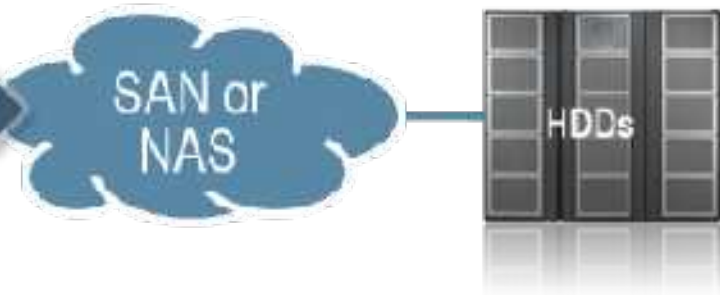
„Readzilla“: lese-optimierte SSDs

- “Readzilla”
 - > Bezeichnung bei Sun für lese-optimierte SSDs
- L2ARC: ZFS Level 2 Addaptive Read Cache
 - > Daten im DRAM werden vor Ihrer Verdrängung automatisch auf die L2ARC Devices (Readzillas) geschrieben
 - Mittels einem Thread: User I/O ohne erhöhte Latenz
 - > L2ARC Writes werden in Gruppen zusammengefasst zur Optimierung der Schreib-Prozesse. Der Schreibvorgang vom RAM zum L2ARC erfolgt “round robin” bzgl. der Flash Module
 - > “Dirty data” und “dirty pages” werden unverzüglich aus dem L2ARC gelöscht.
 - > L2ARC in OpenSolaris – noch nicht in Solaris 10
 - blogs.sun.com/brendan/entry/test

Storage Caching Server für SAN/NAS

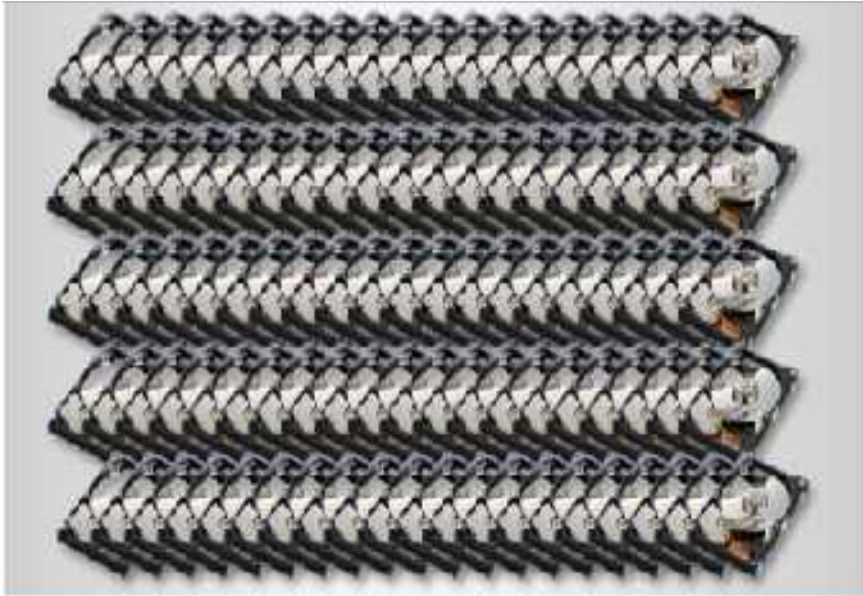


- Cache Gateway für vorhandene SAN/NAS Storage Systeme
- Häufig benutzte Daten verbleiben im schnellen SSD Storage Pool



Ein Beispiel

100 Enterprise HDDs



Capacity: 30TB
 Performance: 30K IOPS
 Cap/Op-X: \$55,000 – 1.75kWh

Hybrid Storage Pool



Capacity: 30TB
 Performance: 30K IOPS
 Cap/Op-X: \$6,040 - 0.392kWh

Sun Open Storage Appliances (ST7xxx)

Preis

ST7410C Clustered HA
288 x 3.5" SATAII Disks
Up to 287TB* total storage
Hybrid Storage Pool with Read and Write Biased SSD



ST7410
288 x 3.5" SATAII Disks
Up to 287TB* total storage
Hybrid Storage Pool with Read and Write Biased SSD



ST7210
48x 3.5" SATAII Disks
Up to 46TB total storage
Hybrid Storage Pool with Write Biased SSD



ST7110
16x2.5"SAS Disks, 2.3TB
Standard Storage Pool SSD is not used



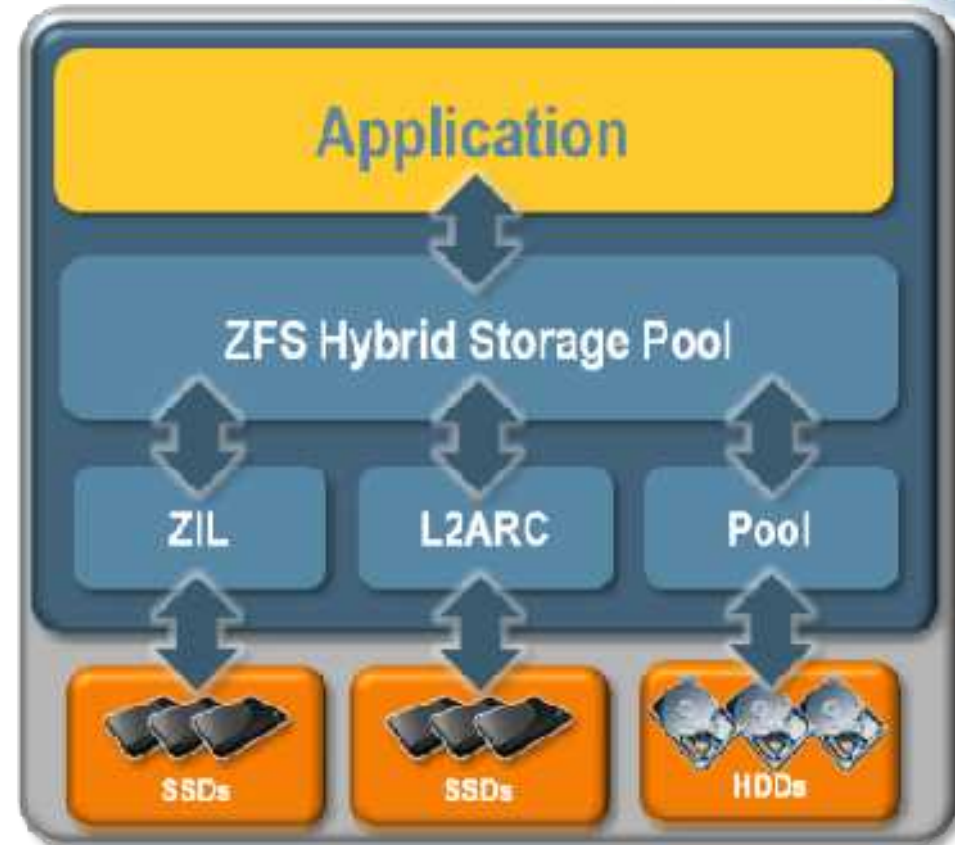
Kapazität / Performance

ZFS Hybrid Storage Pool

Der Turbolader für Applikationen



- Flash Memory basierte SSDs
 - > preiswerte IOPS
 - > teure Kapazität
- Hybrider Ansatz
- Schnelle Writes in den ZIL
“Logzilla”
- Schnelle Reads aus dem L2ARC
“Readzilla”
- Open Storage zum selber Bauen
mit OpenSolaris oder
fertig konfiguriert mit
Sun Unified Storage Appliances der S7000-Reihe



Das ist **OpenStorage**

- **Offene Architektur**
Standardisierte Hardware- und Softwarekomponenten mit Speicherfunktionen und höherer Skalierung bei besserer TCO als proprietäre Alternativen
- **Offene Software**
OpenSource Code und APIs für eine freie Entwicklergemeinschaft
- **Offene Interoperabilität**
Einfache Integration in heterogene Umgebungen durch Nutzung von OpenStandards

Dies ist der nächste Schritt der Entwicklung hin zu einem durchgängig offenen System.

Open Servers

Leveraging innovative product design and packaging, common components, open source software, and wide interoperability to deliver breakthrough economics for Sun customers

OpenStorage

A storage approach that leverages open software, an open architecture with common components, and open interoperability to create innovative storage products that deliver breakthrough economics for Sun customers

Open Network

Unified data center network that utilizes common components, open source software, and seamless integration with existing environments to deliver breakthrough economics for Sun customers

Sun's Solaris öffnet den Weg zu OpenStorage Architekturen

**Innovative
Technologie**



**Offen, Open Source
und kostenfrei**



**Multi-Plattform
Sun / non-Sun
x86 / SPARC**





Vielen Dank !

Detlef.Drewanz@sun.com

