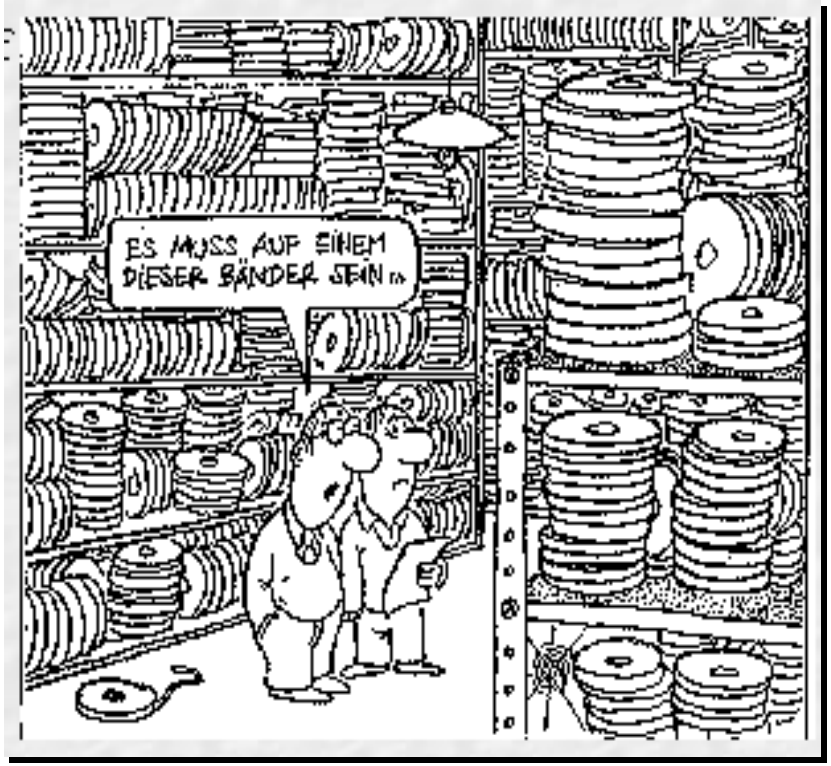
Langzeitarchivierung – Perspektive eines Ressourcenproviders



TMF Workshop

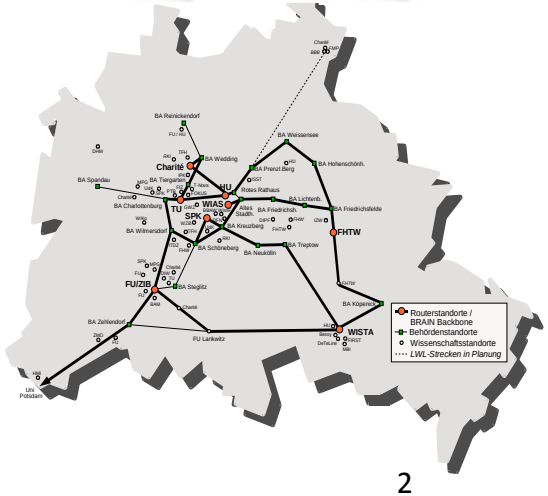
Langzeitarchivierung von medizinischen Forschungsdaten

13. April 2010, Berlin

Thomas Steinke
Zuse Institute Berlin (ZIB) <www.zib.de>
steinke@zib.de

ZIB Kurzvorstellung

- Supercomputing
  - SGI Altix: 15,000 CPU Cores, 150 TFlop/s, 46 TB Hauptspeicher, 810 TB mit parallelem Dateisystem
- Storage
  - SAN > 1,000 TB
- Datenarchivierung
  - 2 Sun STK SL8500, 14,000 Slots, 28 Laufwerke
  - DMF (Cray), HSM/Tivoli (IBM), SAM-FS (Sun)
- Netzwerk-Management
  - Verwaltung des Berliner Metropolitan Area Network (BRAIN) auf dedizierten LWL



steinke@zib.de

Datenmanagement am ZIB

- Service
  - Archivierung der Simulationsdaten vom HPC-System (seit 1989)
    - klassische Communities wie Klimaforschung, CFD, & QCD
    - Migration von On-Line ←→ Off-Line
  - Speicherung von Daten von D-Grid-Projekten
    - Datenmanagement, z.B. mit **SRB** & **iRODS**
  - Realisierung von LZA-Projekten
  - Betrieb von Archivroboter (StorageTek / Sun)

- Forschung & Entwicklung
  - geographisch verteiltes föderiertes Dateisystem – **XtreemFS**
    - entwickelt im Rahmen EU FP7 XtreemOS
    - Einsatz in laufenden Projekten
  - verteilte skalierende fehlertoleranter Key-Value-Store -- **Scalaris**

Langzeitarchivierung

Wir wissen warum…

Amalia Bibliothek Weimar

Stadtarchiv Köln

steinke@zib.de

Langzeitarchivierung am ZIB – laufende Projekte

Max-Planck-Institut für Gravitationsphysik (Albert-Einstein-Institut)

GEO 600

GEO600

Deutsche Archäologische Institut

DEUTSCHE KINEMATHEK MUSEUM FÜR FILM UND FERNSEHEN

Deutsche Kinemathek – Museum für Film und Fernsehen

AIP

Astrophysikalisches Institut Potsdam

[weitere:] **HLRN II**, Lattice QCD

ZIB

steinke@zib.de

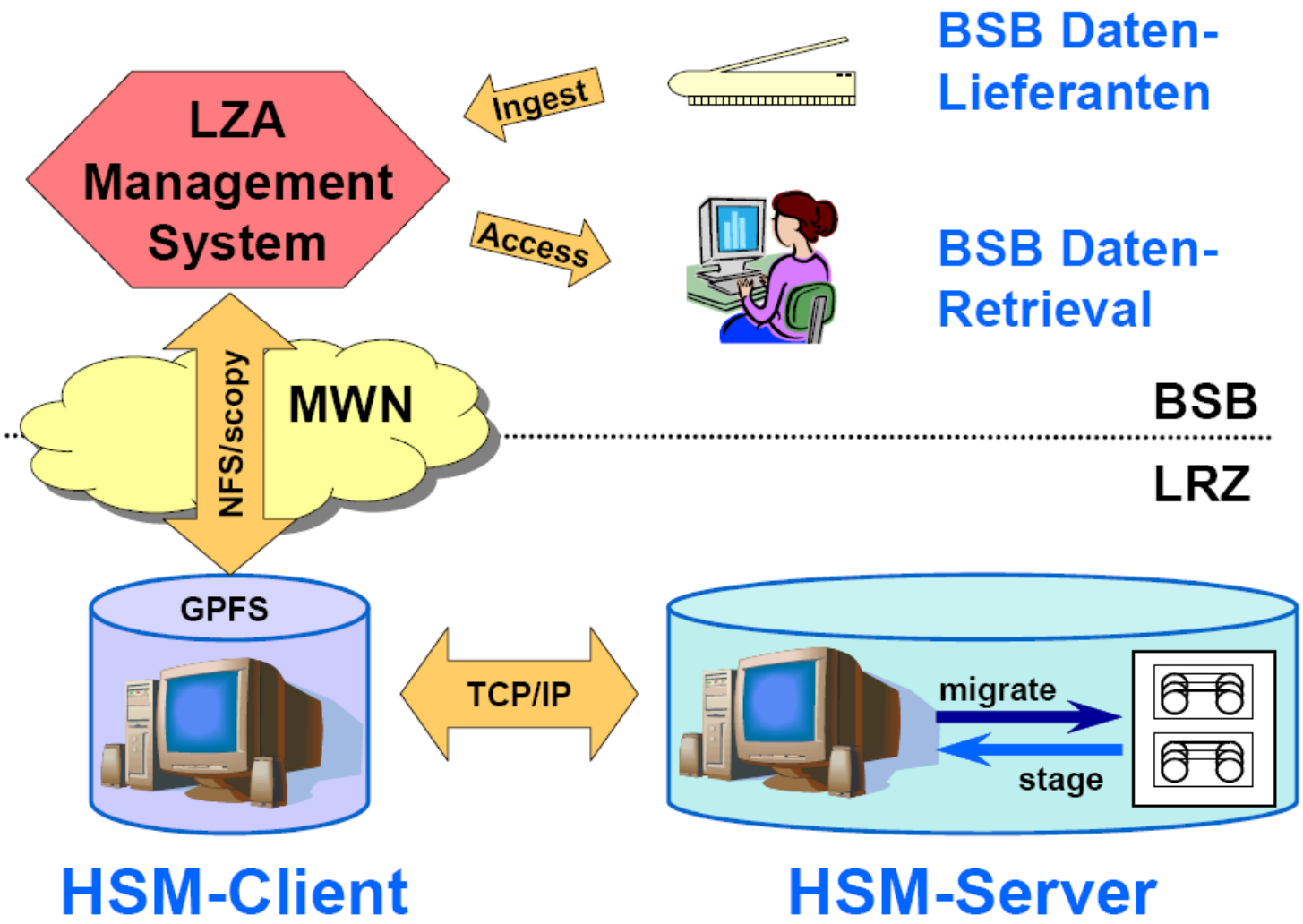
GEO600: Anforderungen in Zahlen

- Projekt läuft seit 2002
  - 25,3 MByte/min
  - pro Tag: 1.440 Dateien, 36,4 GByte
  - pro Monat: 43.200 Dateien, 1,1 TByte
  - pro Jahr: 518.400 Dateien, 13,1 TByte

… seit 2002: > 90 TByte Daten

Beispiel einer LZA-Architektur (BABS, München)

Bibliothekarisches Archivierungs und Bereitstellungssystem



Quelle: Bernd Reiner, LRZ München

Herausforderungen für LZA

Langzeitarchivierung…

- Zeitachse beachten – wie sieht der Datenzugriff in 2100 aus?
- Einschränkungen durch physiko-chemische Prozesse
  - Lebensdauer der Lesbarkeit auf den Medien
- technischer Innovationszyklus
  - ca. alle 2-3 Jahre neue Laufwerks- / Band-Generation
  - Herstellerabhängigkeiten
  - Migrationsstrategien

Rolle des ZIB bei der Langzeitarchivierung

- Service auf **Bit-Preservation-Level**:
  - Sicherstellung der Konsistenz der Daten
- Archivierung der Daten
  - Betrieb von Speicher- und Server-Ressourcen
  - **Monitoring** & Wartung
- Garantie der Korrektheit und Verfügbarkeit
  - Erhaltungsmaßnahmen

Technische Herausforderungen

- *Bit Preservation: A Solved Problem?*
  - David S. H. Rosenthal, Stanford University Libraries

    "Indeed, current digital storage technologies are not merely astoundingly cheap and capacious, they are astonishingly reliable. Unfortunately, these attributes drive a kind of "Parkinson's Law" of storage, in which demands continually push beyond the capabilities of systems implementable at an affordable price."

  - Herstellerangaben (MTTDL): keine verwertbare Information
    - um Größenordnungen zu optimistisch
    - neue messbare Metriken (z.B. Bit-Lebenszeit) notwendig

  - gegenwärtige Speichertechnologien (Hardware + Software) zu komplex/anfällig, um Anforderung nach Bit-Preservation zu genügen

If proponents really believed that bit preservation was solved, they wouldn't bother with backups. – D. S. Rosenthal

Strategien für technische Realisierung LZA

Tradition und Neuentwicklungen

- traditionelle Wege:
  - Erstellung und Verwaltung von Kopien
    - z. B. Kopien auf >2 Bänder, (mehrere Standorte)
    - Limitierung vor Hintergrund exponentiell wachsender Datenvolumina
  - bauliche Maßnahmen (Lambertz-Raum)

- Entwicklung neuer Methodiken, z. B. …
  - Low-Level HW-Überwachung von Bändern
    - Realisierung am ZIB (auch am SDSC u. ETHZ)
    - frühzeitige Erkennung von “Soft Errors” auf Bändern

Migration bei Technologiewechsel (2008)

2008

2 x STK Datenroboter (Powderhorn)
20 Jahre in Betrieb (1989-2009)

2x STK SL8500

Migration der Nutzerdaten **1.3 PByte**

ca. 4 Wochen, eigene CRC-Software, spez. konfigurierter Migrationsserver für max. Durchsatz (6x Lese-/ 4x Schreiblaufwerke über Fibre Channel)

Technische Details -- 5 Folien für Freaks

Magnetbandspeicher

Roboter Sun STK SL8500

Ein Roboter als Modular erweiterbares System

Grundmodul + 5 Erweiterungen **Grundmodul + 1 Erweiterung** **beide mit 4 Durchreichen verbunden**

Insgesamt

13.500 Stellplätze für Bänder
16 Handbots
38 Laufwerke

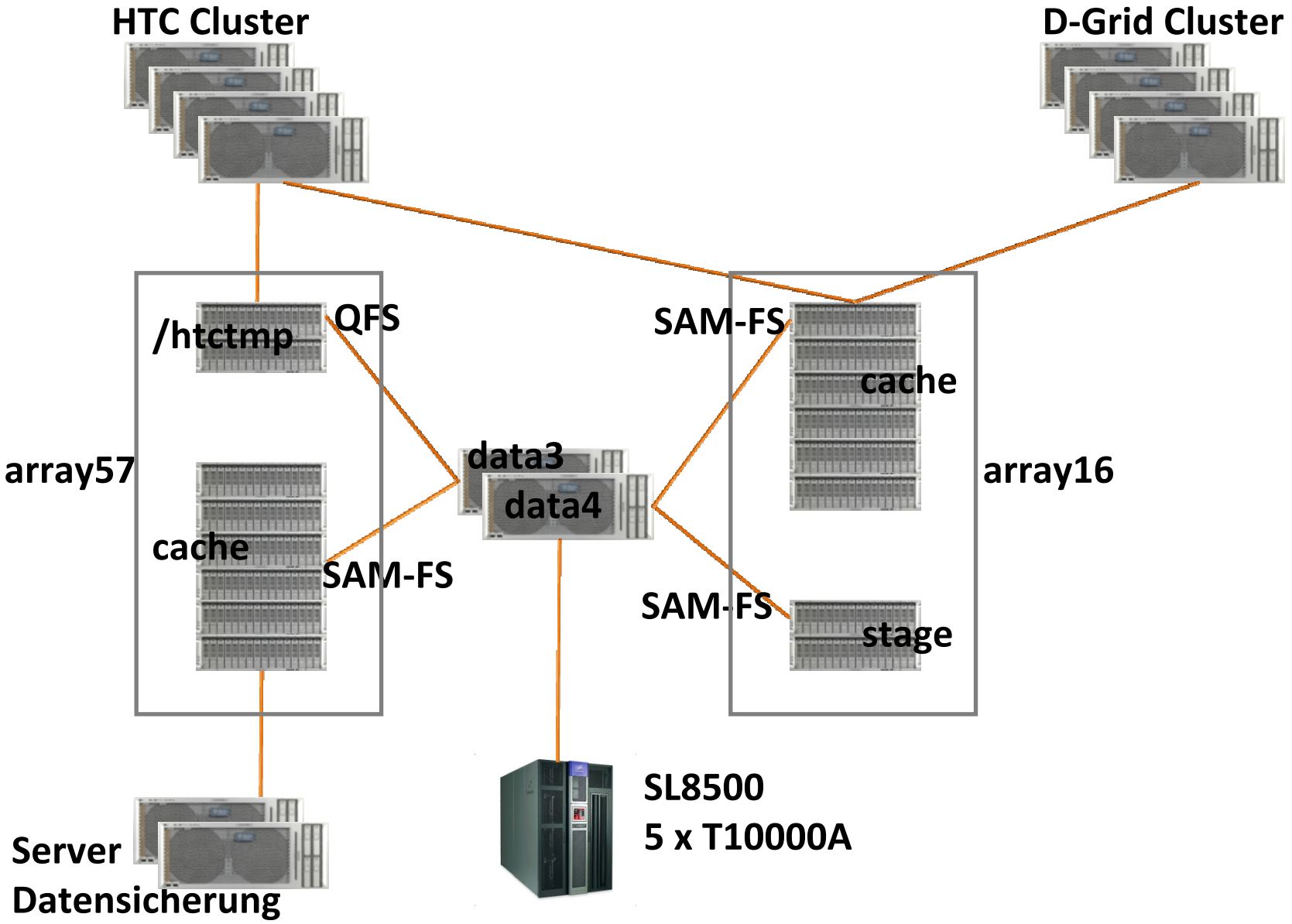
Speicherkapazität **(bei Komprimierung 1:1,5)**

4 PB bei T9940B
10 PB bei T10000A
20 PB bei T10000B

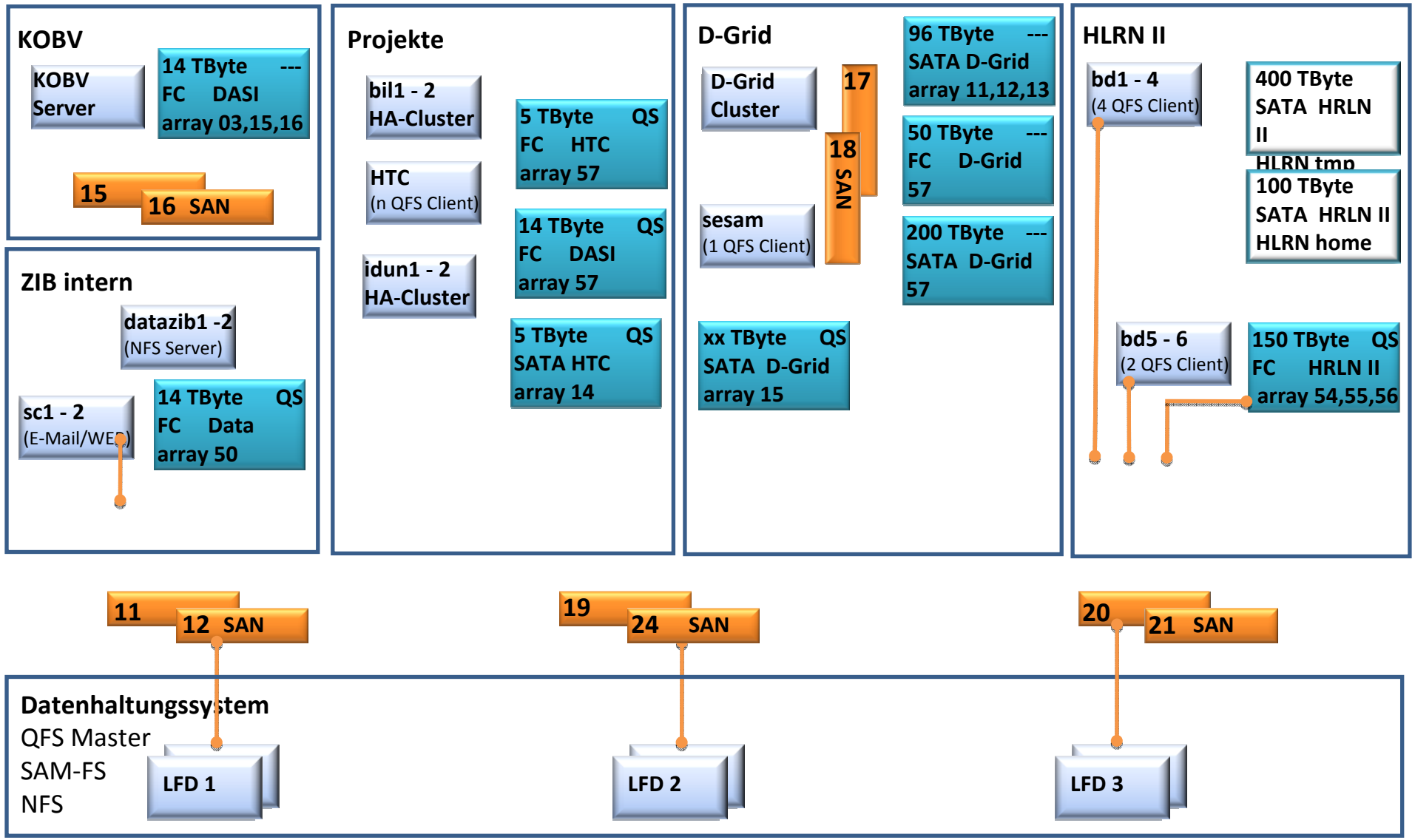
Laufwerke bis zu 64 pro Robot	**Typ**	**Kapazität eines Bandes (unkomprimiert)**	**Bandbreite**
	**17 x T9940B**	**200 GByte**	**30 MByte/s**
	11 x T10000A	**500 GByte**	**120 MByte/s**
	10 x T10000B	**1000 GByte**	**120 MByte/s**
	LTO und SDLT	**möglich, aber zur Zeit nicht geplant**	

Cluster LFD2

QFS/SAM-FS Server data3 und data4

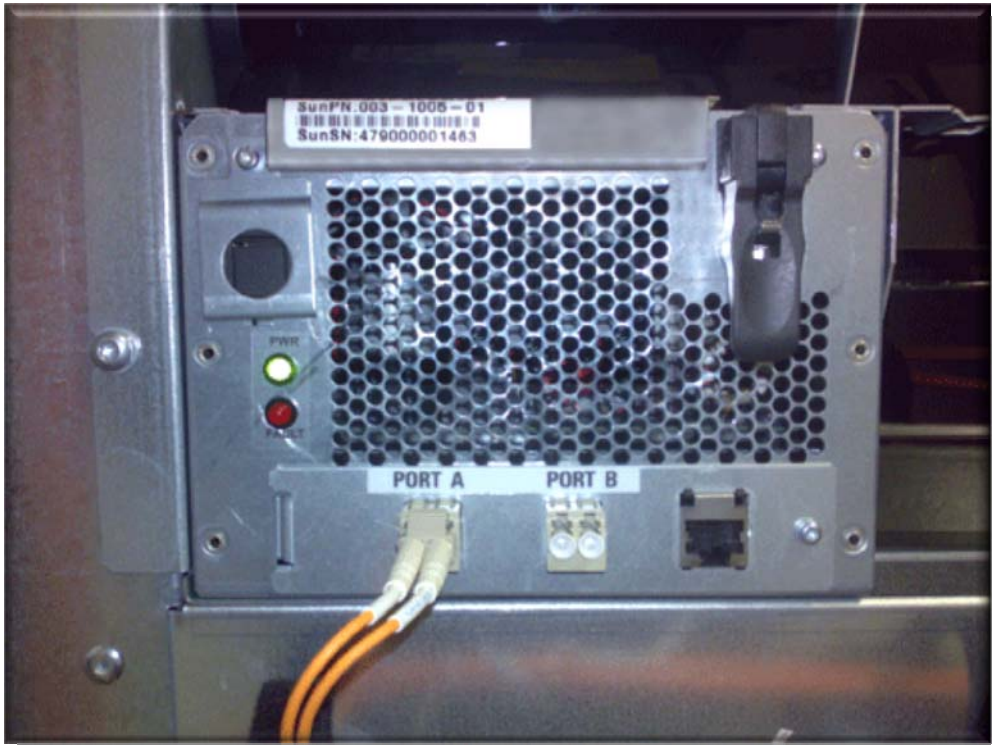


Anbindung des Plattenspeichers und der Bandlaufwerke



Magnetbandspeicher

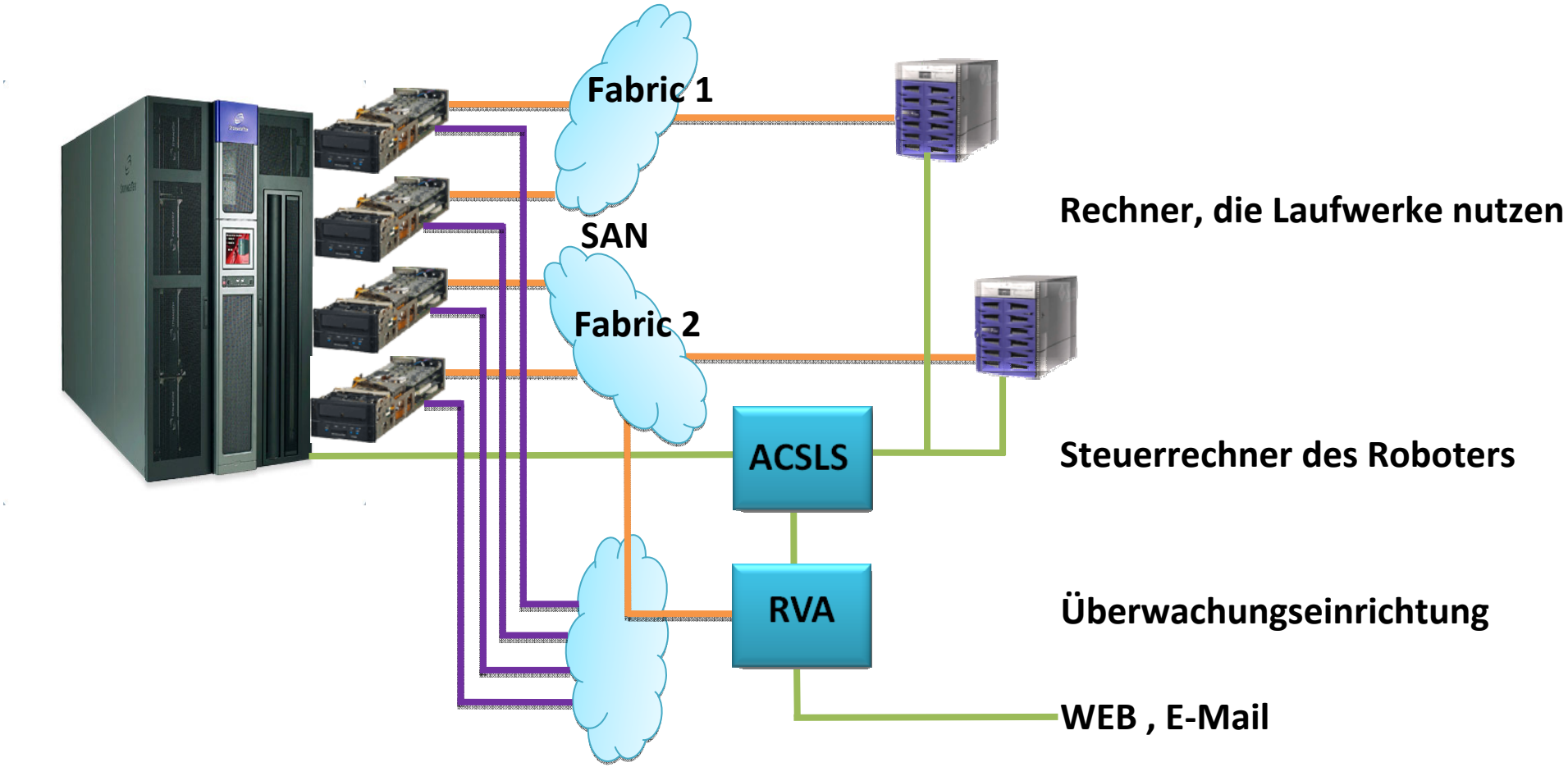
Bandlaufwerke



Port B, zumeist ungenutzt, wird jetzt für ein RVA private out off band SAN genutzt

Magnetbandspeicher

Anschluss der RVA über ein physisch getrenntes FC Netz



Die B-Ports der Laufwerke wurden auf zwei eigene FC-Switche geführt, womit ein physisch getrenntes FC Netz nur für die Funktionen des Management über RVA existiert.

IP - Steuerung
FC- Daten
FC - Monitoring

Rollen eines LZA-Providers (I)

Archivierung

- Speicherung der Rohdaten + Metadaten
  - Schwerpunkt: Bit-Preservation
  - Zugriffskontrolle / -management
  - Auditing, Monitoring
  - Ressourcenmanagement auf unterer Ebene: Zugriffspfade
  - Replika an geographisch verteilten Standorten wünschenswert
- keine Interpretation der Daten
  - Datenformate sind “egal” – ein Bitstrom wird gespeichert
  - Anwendungen/obere SW-Schicht sollte Daten interpretieren
    - Erfolg der Speicherung von Daten in (Unix) Dateisystemen im akademischen Umfeld – keine SQL-Datenbanken
- (üblicherweise) keine Erzeugung von Metadaten
- rechtliche Aspekte müssen vorab geklärt sein

Rollen eines LZA-Providers (II)

Zugriffsdiensten

- Konfiguration + Betrieb von Zugriffsdiensten
  - Nutzung der Erfahrungen aus Grid-Projekten möglich

Metadaten-Management

- Definition ist getrieben durch Community
  - Inhalte, Strukturen, Standards
  - Provider kann technisch Prozess begleiten
- daher bislang nicht typische Aufgabe des LZA-Provider, wenn mehrere Communities unterstützt werden
- aber: Betrieb des Ingest-Service möglich/nützlich