Distributed Data Management at DKRZ

Wolfgang Sell
Deutsches Klimarechenzentrum GmbH
sell@dkrz.de

Table of Contents

DKRZ - a German HPCC

- Mission of DKRZ
- DKRZ and its Organization
- DKRZ Services
- DKRZ Restructuring

Mission of DKRZ

In 1987 DKRZ was founded with the Mission to

- ***Provide state-of-the-art supercomputing and data service to the German scientific community to conduct top of the line Earth System and Climate Modelling.***

- ***Provide associated services including high level visualization.***

DKRZ and its Organization (1)

Deutsches KlimaRechenZentrum = *DKRZ*
German Climate Computer Center

- organised under private law (GmbH) with 4 shareholders
- investments funded by federal government, operations funded by shareholders

DKRZ and its Organization (2)

DKRZ internal Structure

- 3 departments for
  - systems and networks
  - visualisation and consulting
  - administration
- 20 staff in total
- until restructuring end of 1999 a fourth department supported climate model applications and climate data management

DKRZ Services

- operations center: ***DKRZ***
  - technical organization of computational ressources (compute-, data- and network-services, infrastructure)
  - advanced visualisation
  - assistance for parallel architectures (consulting and training)

Model & Data Services

Application center: ***Model & Data***

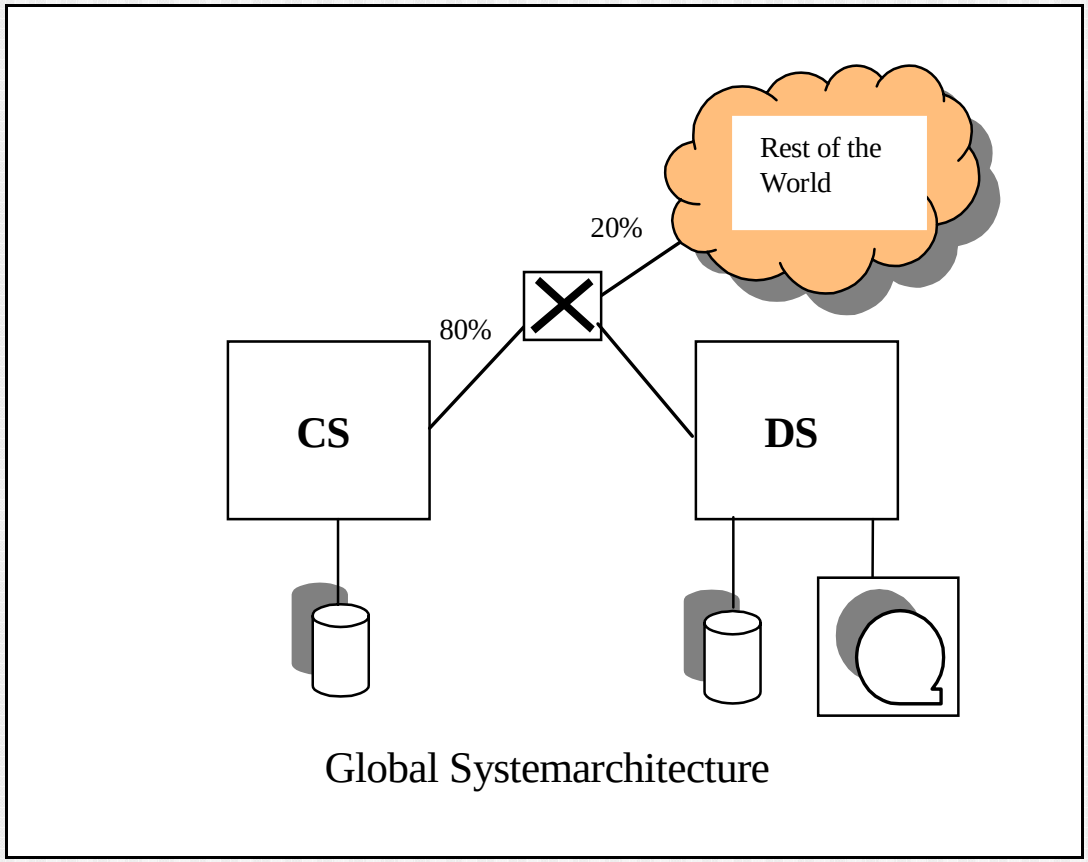
- professional handling of community models
- specific scenario runs, e.g. IPCC
- scientific data handling

Model & Data Group external to DKRZ, administered by MPI for Meteorology, funded by BMBF

HPC Systemarchitecture suited for Earth System Modeling

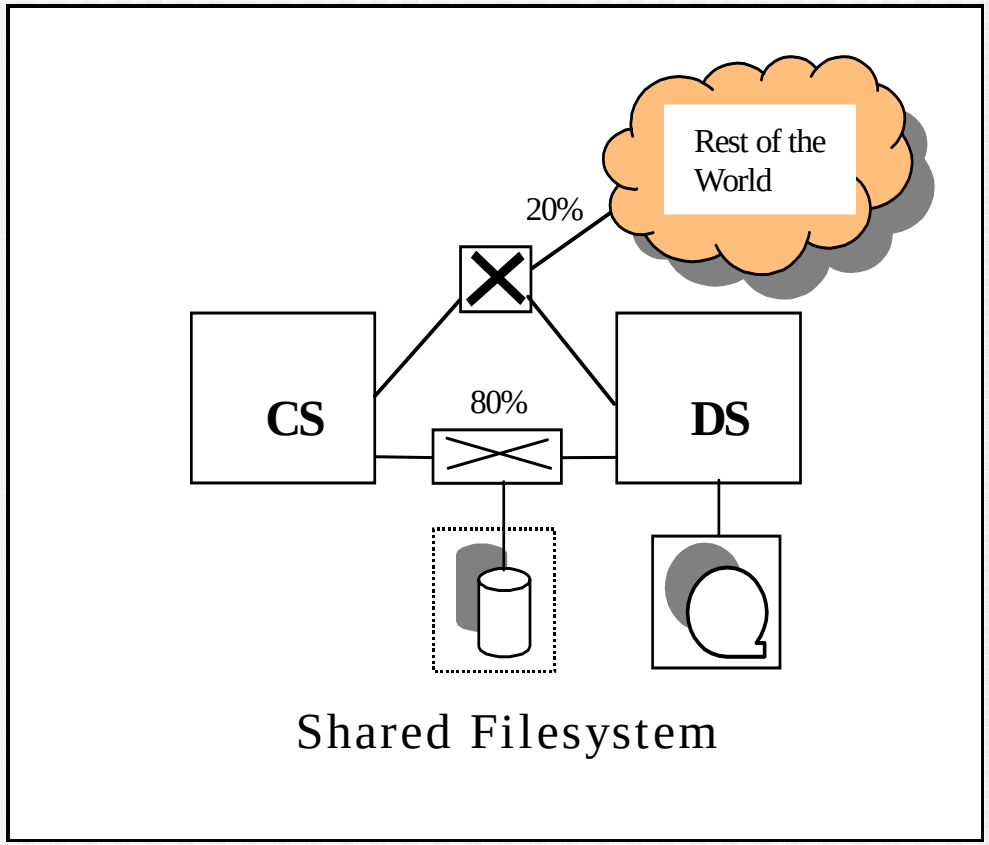
- **Principal HPC System Configuration**
- **Configuration Variants**
- **Links between Different Services**
- **The Data Problem**
- **Pros and Cons of Shared Filesystems**

Generic HPC System Configuration

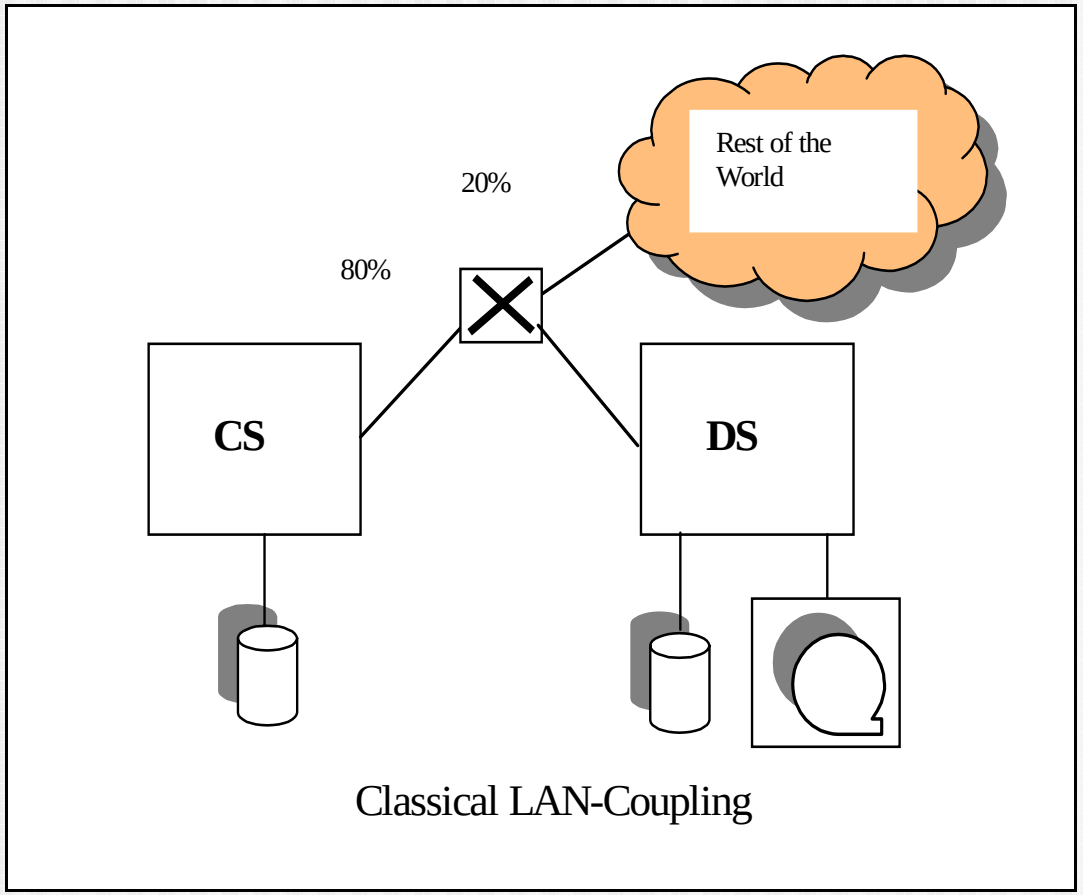


Global Systemarchitecture

Variants of System Configuration (1)



Variants of System Configuration (2)

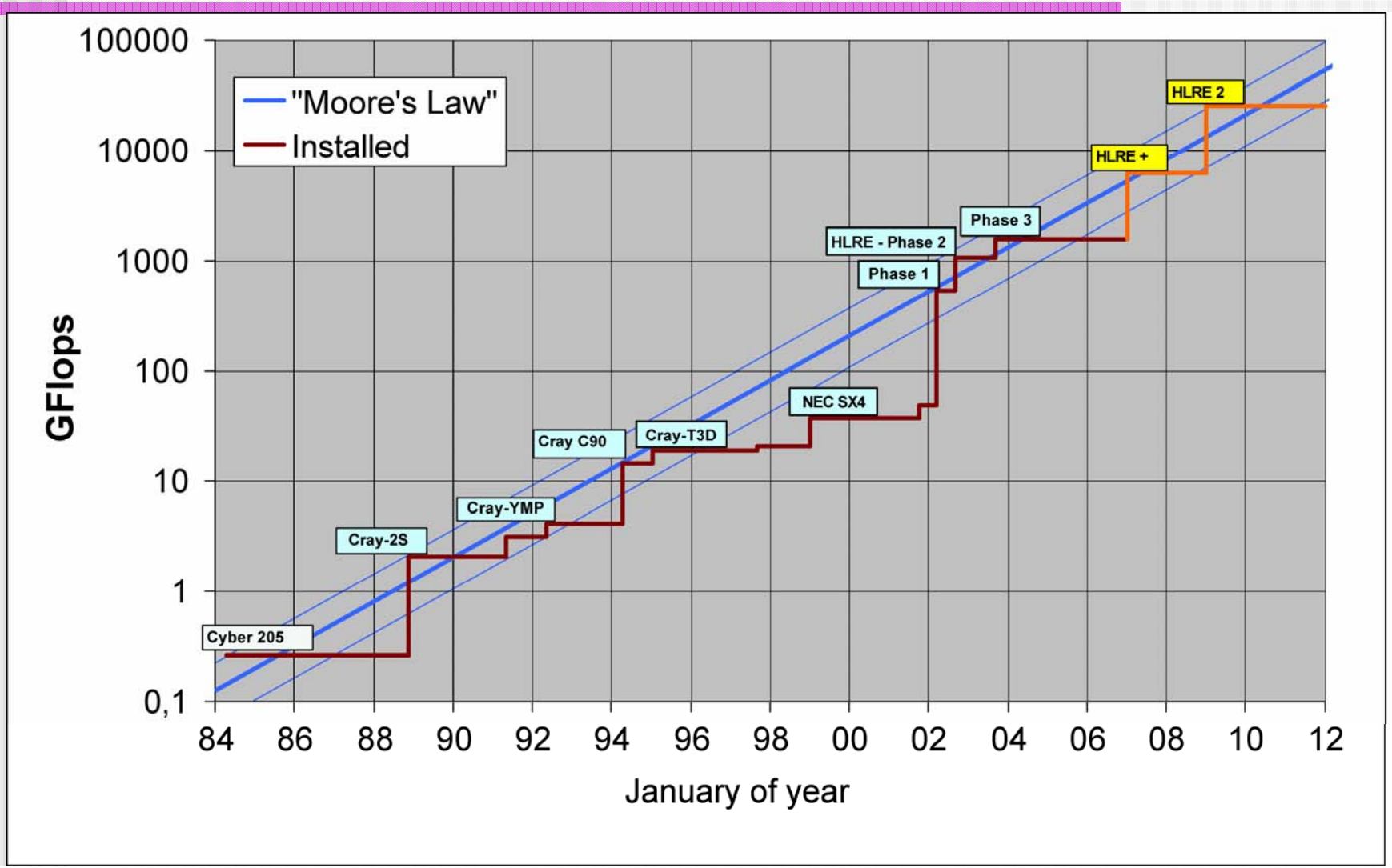


Classical LAN-Coupling

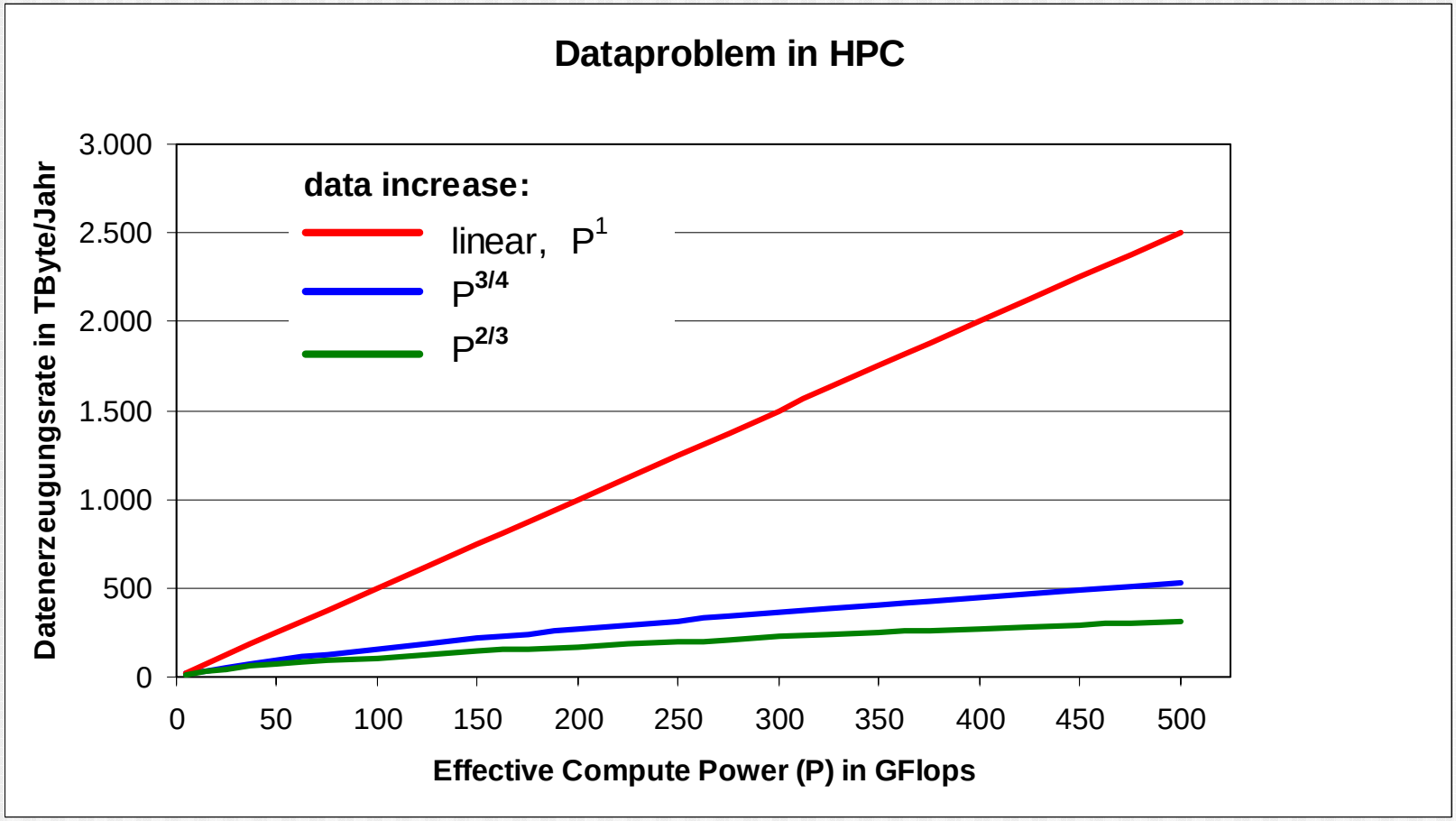
Link between Compute Power and Non-Computing Services

- Functionality and Performance Requirements for Data Service
  - Transparent Access to Migrated Data
  - High Bandwidth for Data Transfer
  - Shared Filesystem
- Possibility for Adaptation in Upgrade Steps due to Changes in Usage Profile
- Balance between Computational and Data Management Capabilities

Evolution of Computing Power at DKRZ



Adaptation Problem for Data Server



Pros of Shared Filesystem Coupling

- High Bandwidth between the Coupled Servers
- Scalability supported by Operating System
- No Needs for Multiple Copies
- Record Level Access to Data with High Performance
- Minimized Data Transfers

Cons of Shared Filesystem Coupling

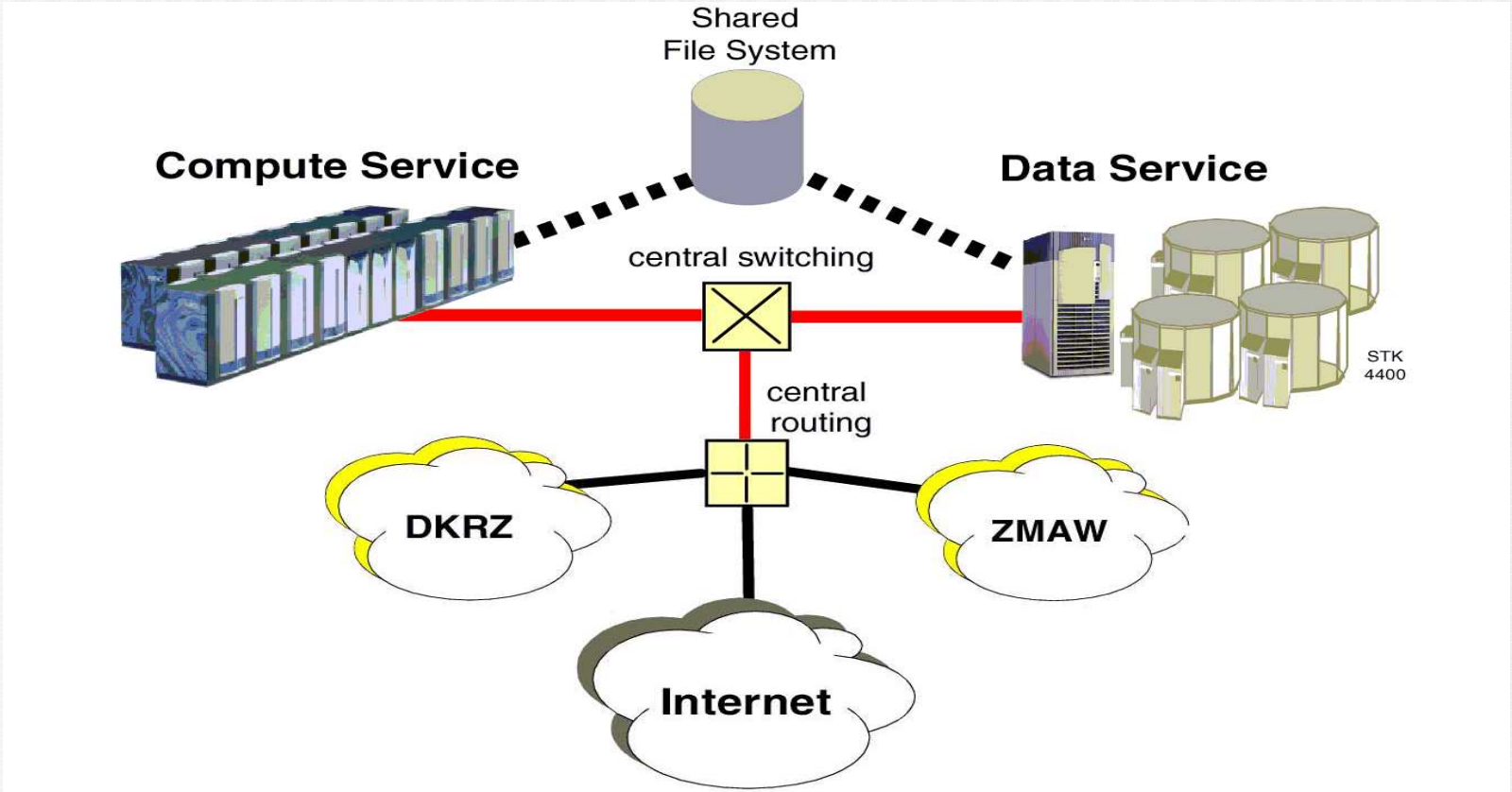
- Proprietary Software needed
- Standardisation still missing
- Limited Number of Systems that can be connected

HLRE Implementation at DKRZ

HöchstLeistungsRechnersystem für die Erdsystemforschung
= *HLRE*
High Performance Computer System for Earth System Research

- **Principal HLRE System Configuration**
- **Requirements and Constraints**
- **Links between Different Services**
- **Option for Systemoperation**

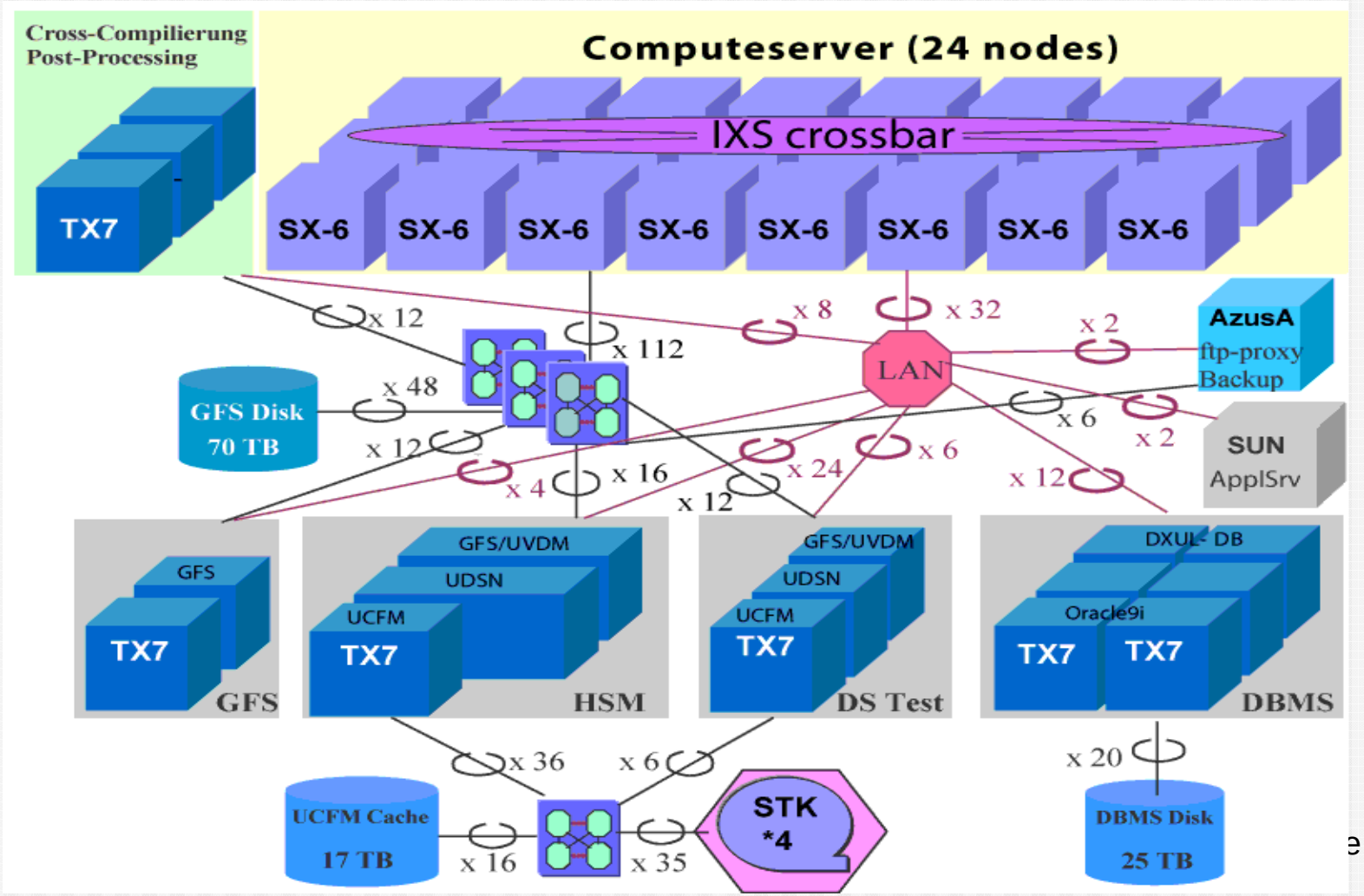
Principal HLRE System Configuration



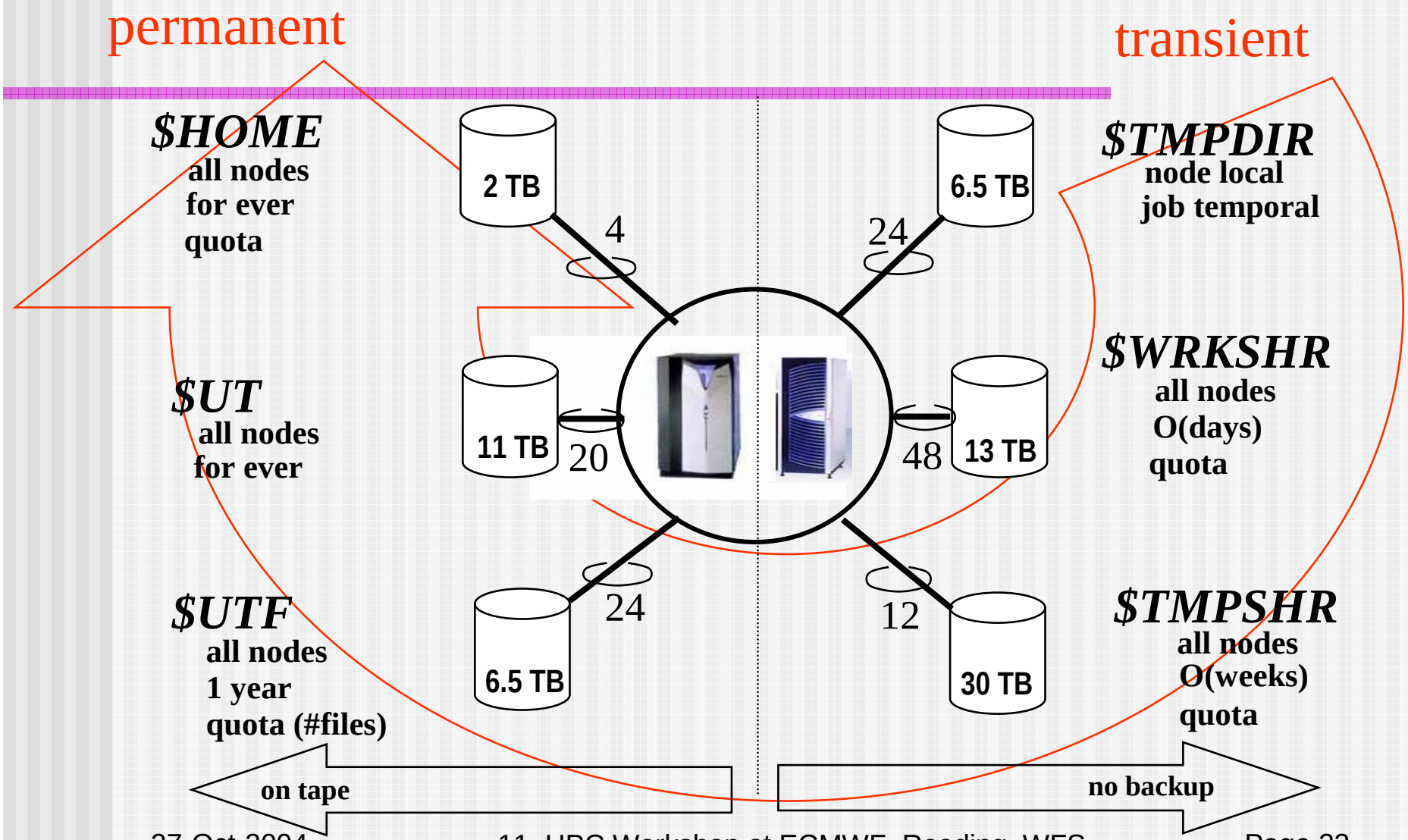
Hardware at DKRZ (October 2004)

- **24 SX-6 Nodes**
  (192 Vector CPUs, 1,5 TByte CM and 1,5 Tflops peak)
- **IXS Crossbar switch**
  (24 x 24, 2*8*24 GByte/s cross section bandwidth)
- **10 NEC AsAmA Nodes**
  (132 Itanium-2, 1,0 and 1,5 GHz, Linux)
- **1 NEC AzusA**
  (8 Itanium-1; 800 MHz; Linux)
- **4 STK Silos**
  (total capacity ca. 3.5 PetaByte)
- **4 SUN Fire 4800 (Oracle Appl. Service)**

DKRZ Hardware
Current Configuration



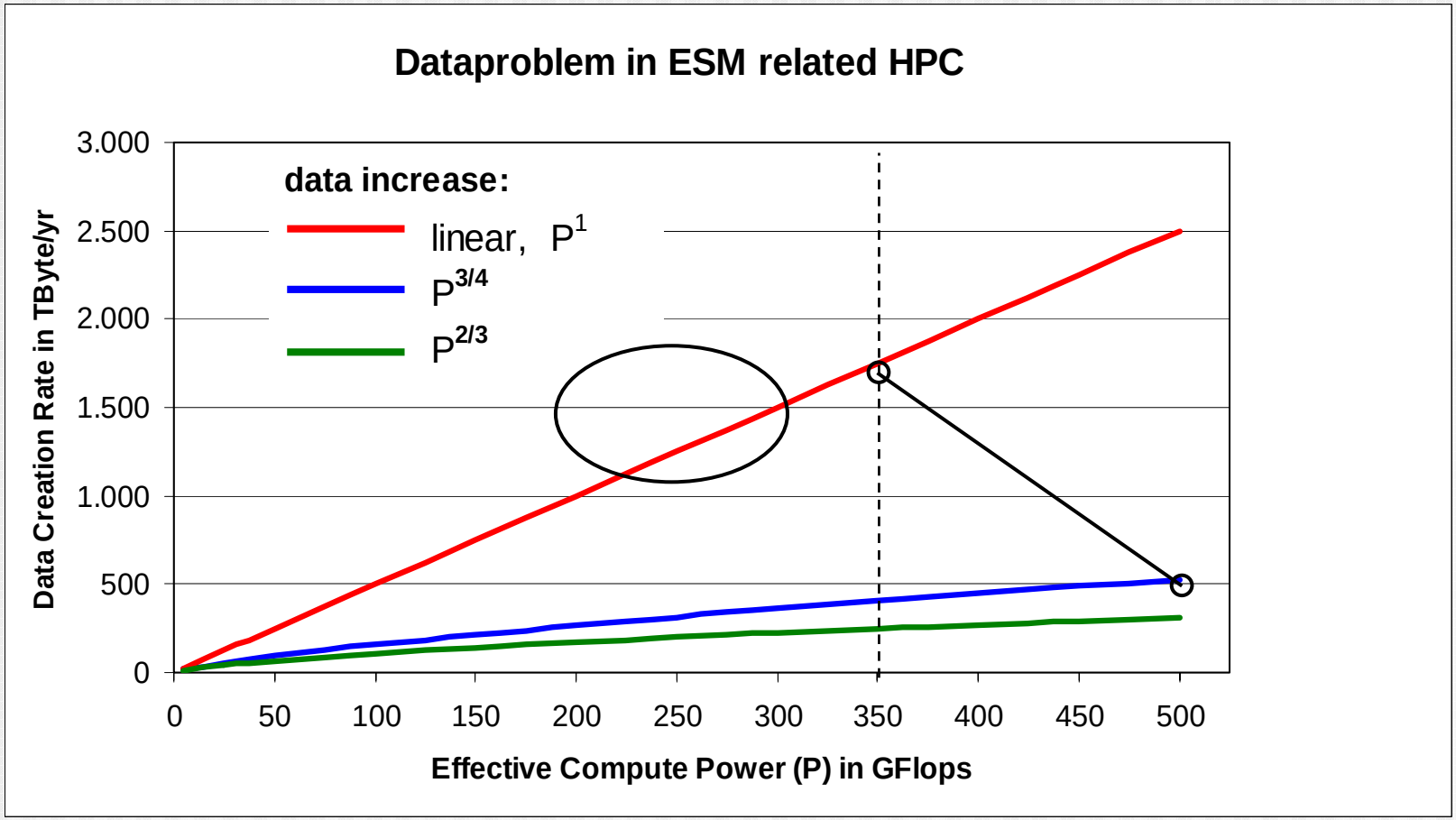
Filesystem Systematics
CS View



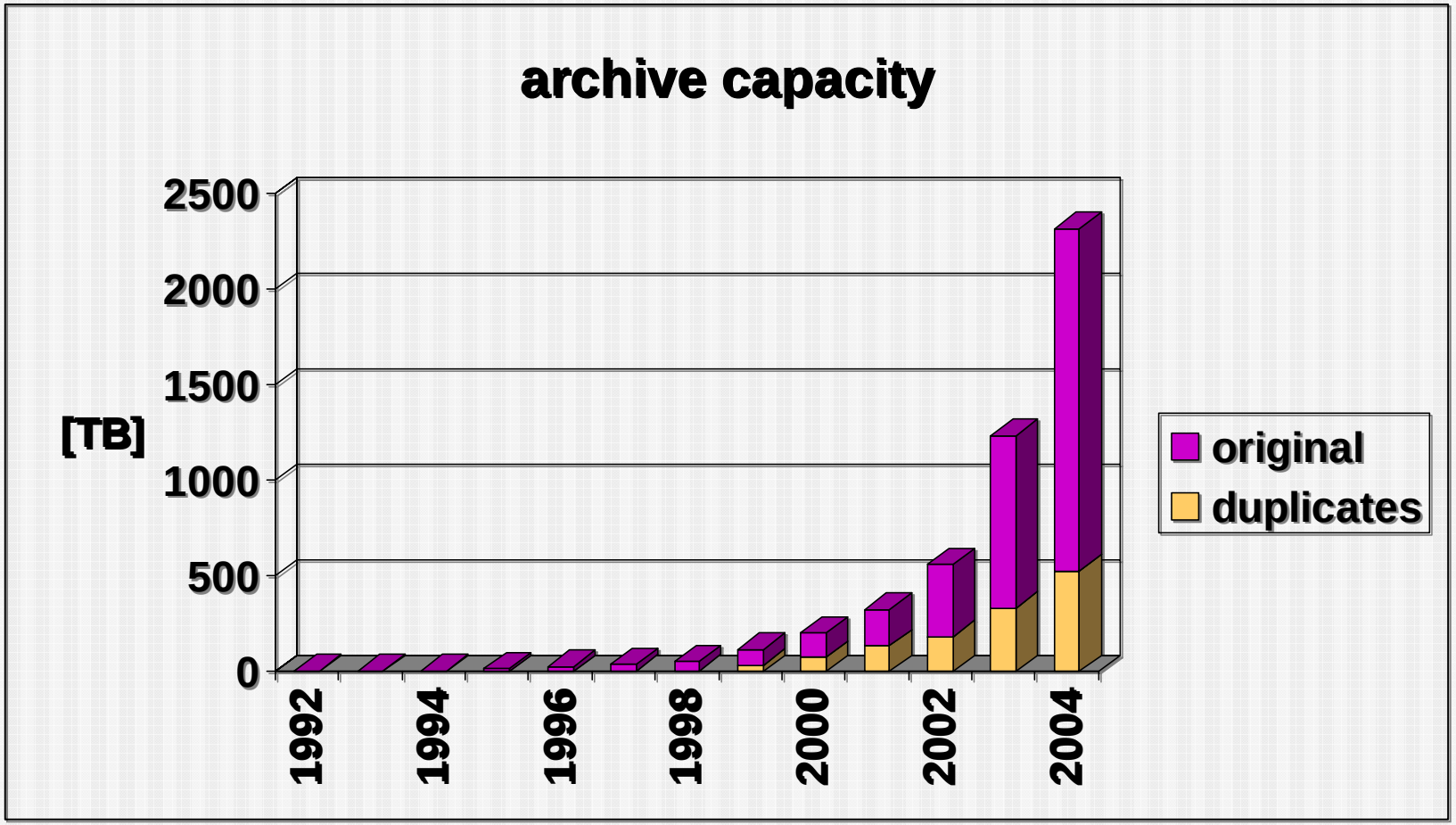
Some Results

- Point of Operation in CS-DS-Space
- Growth of the Data Archive
- Growth of Transferrate
- Variability of Transferrates

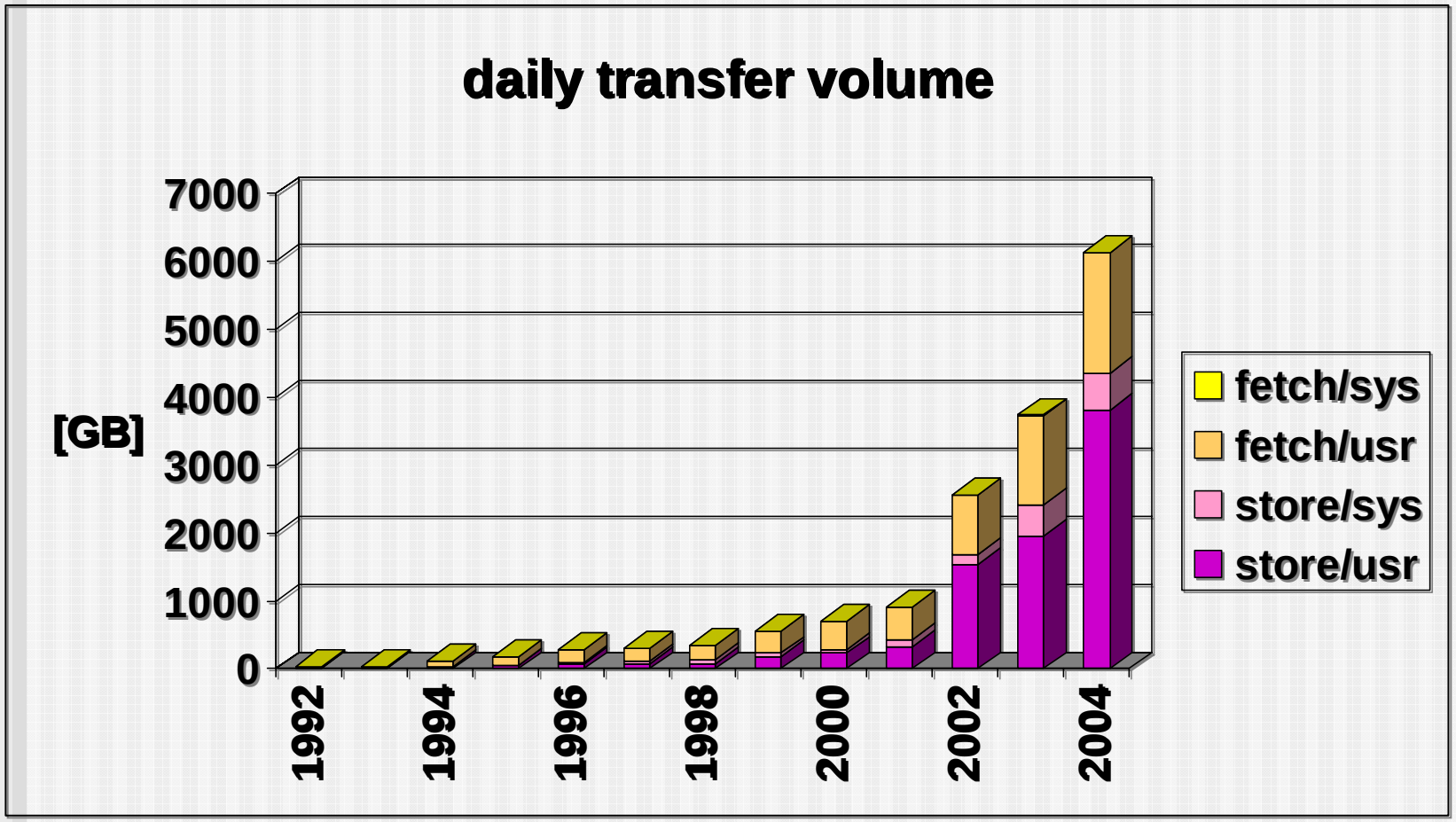
Point of Operation in CS-DS-Space



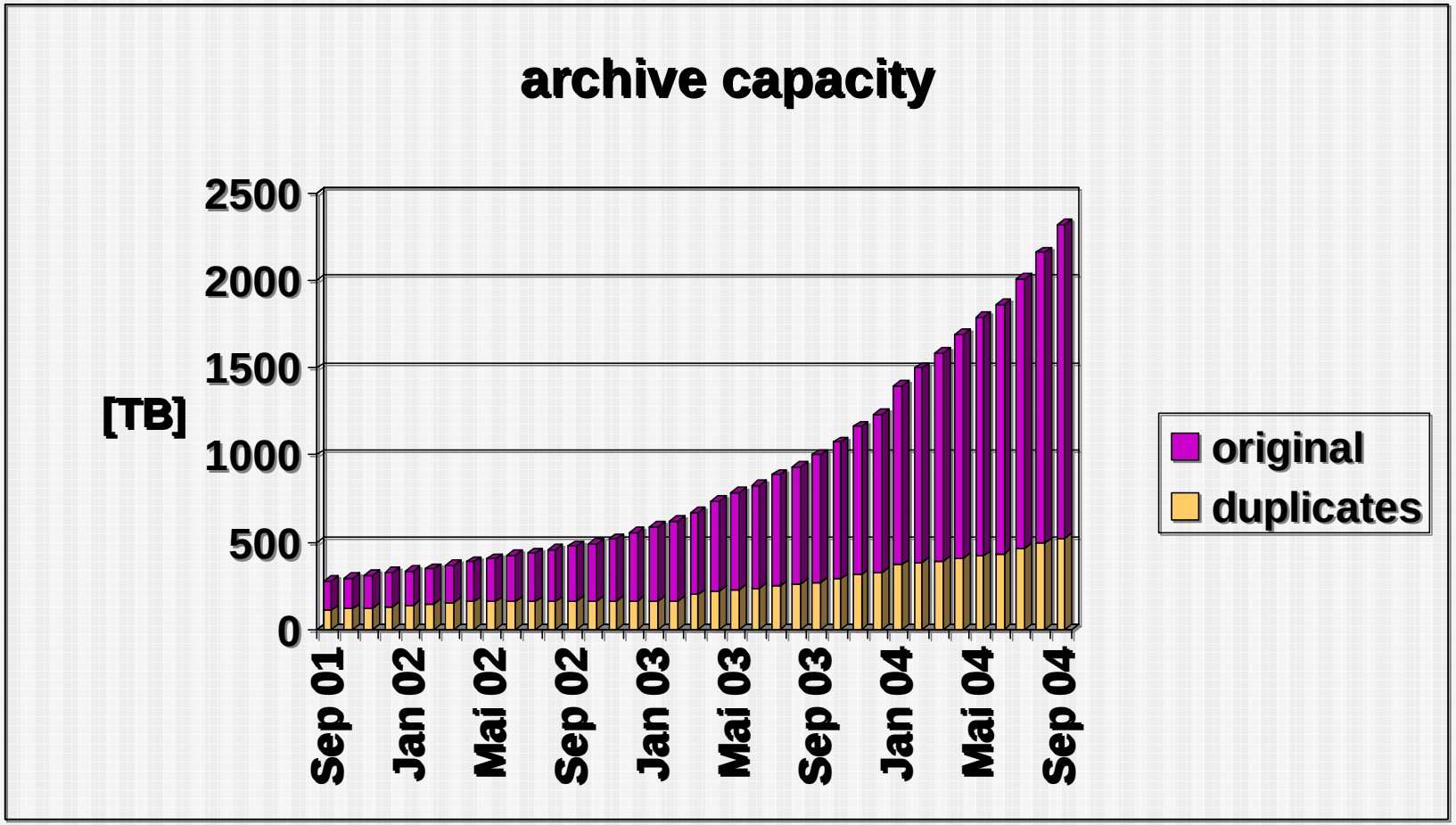
DS archive capacity (1)



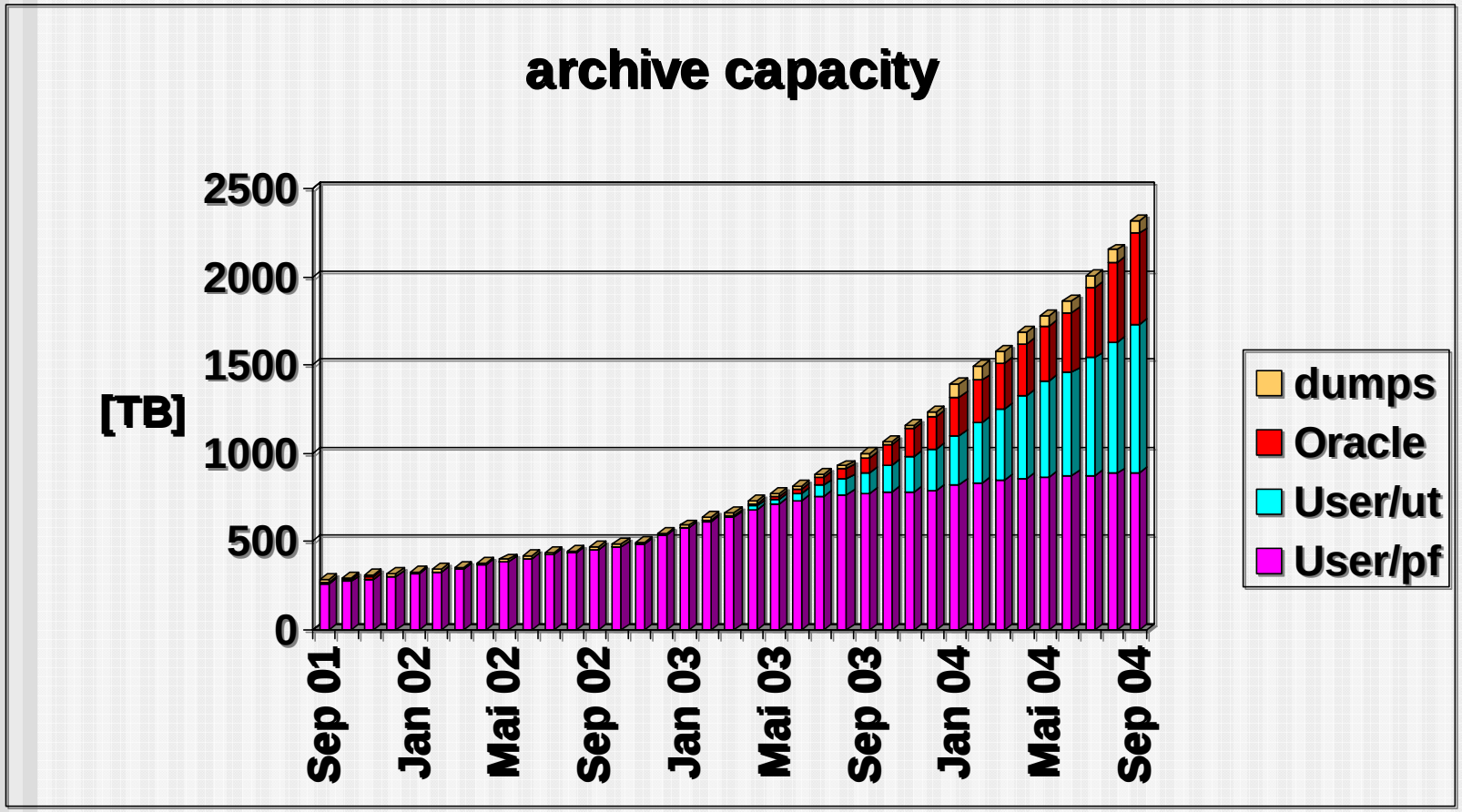
DS transfer rates (1)



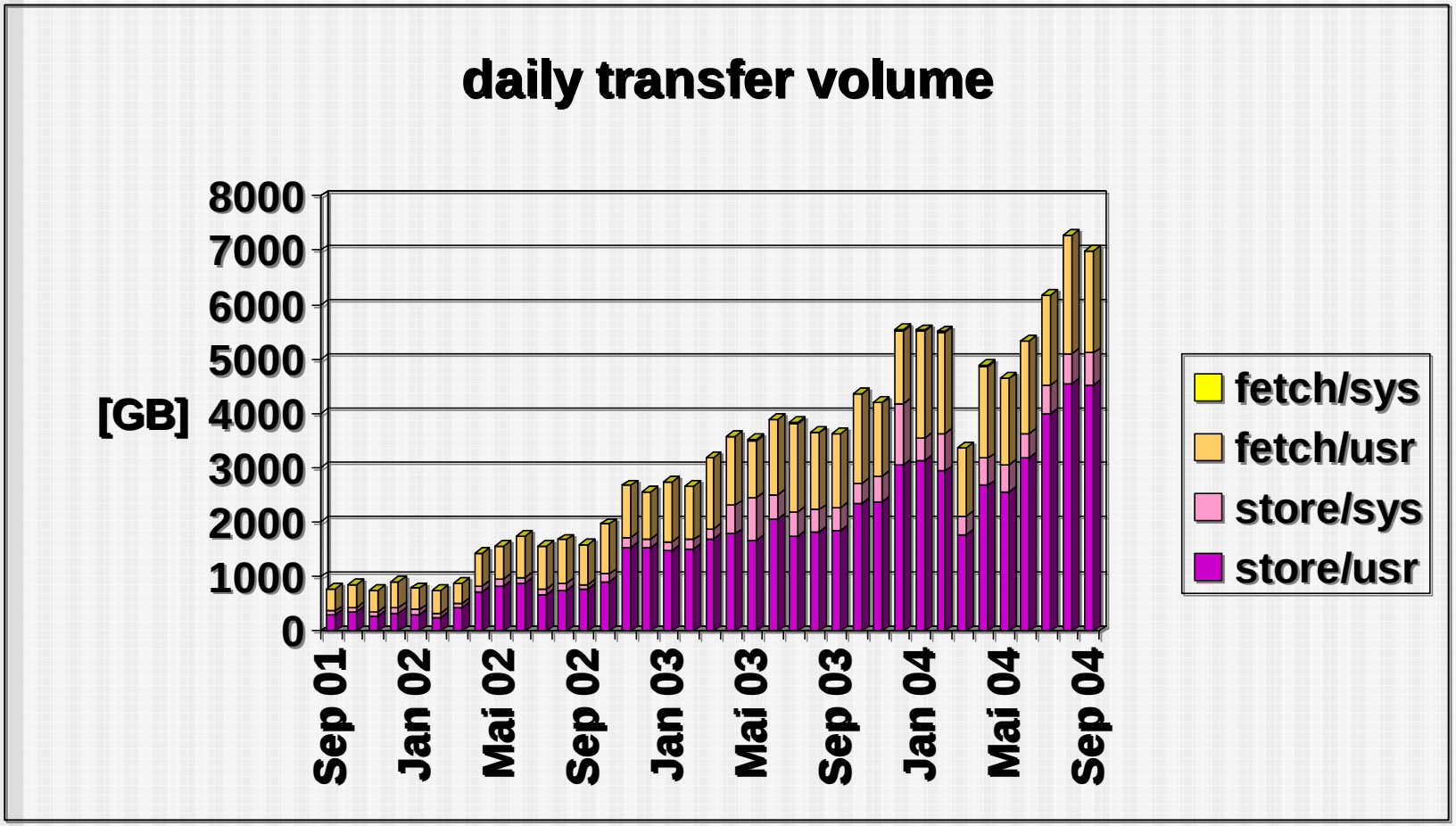
DS archive capacity (2001-2004)



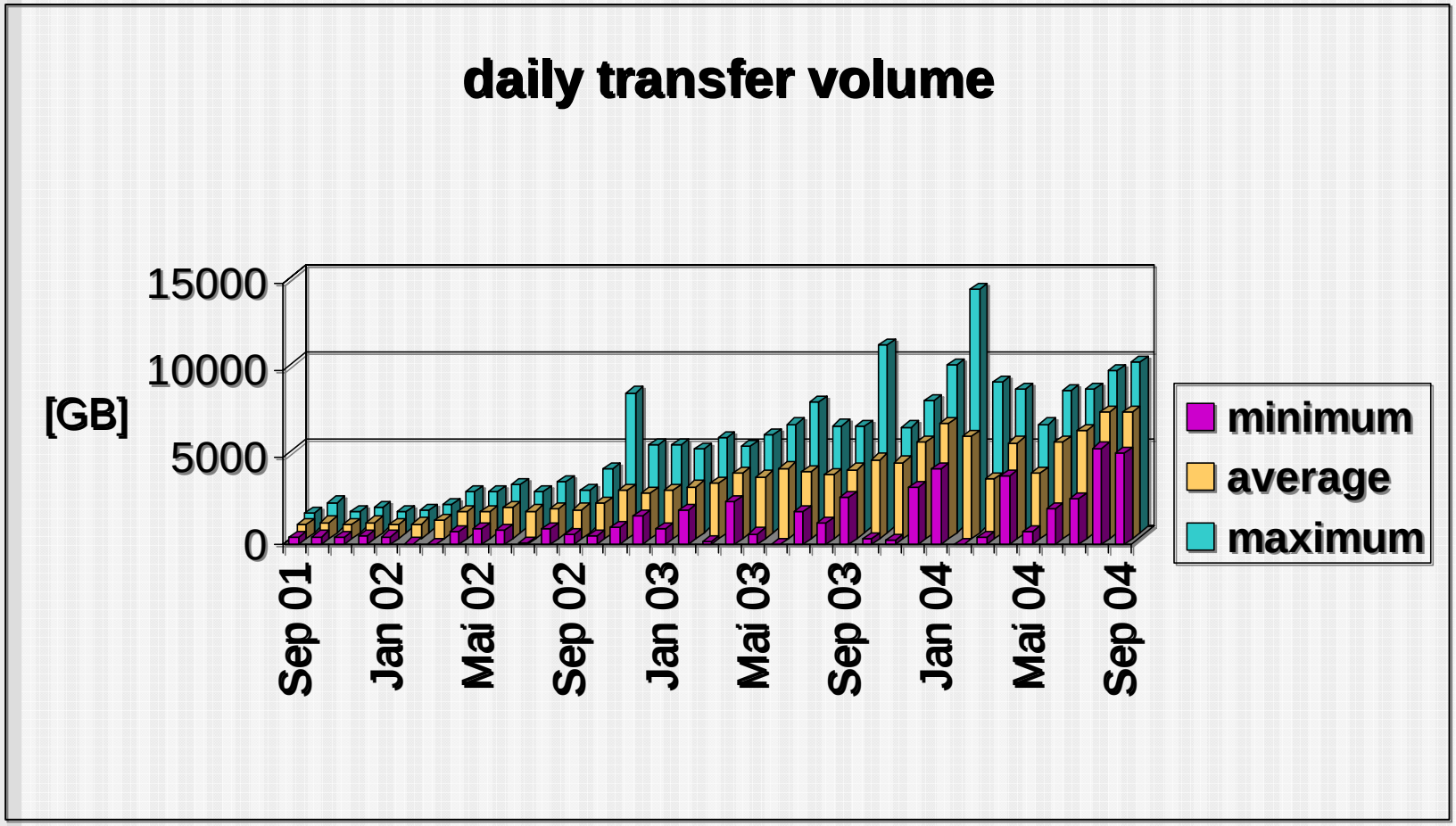
DS archive capacity (2001-2004)



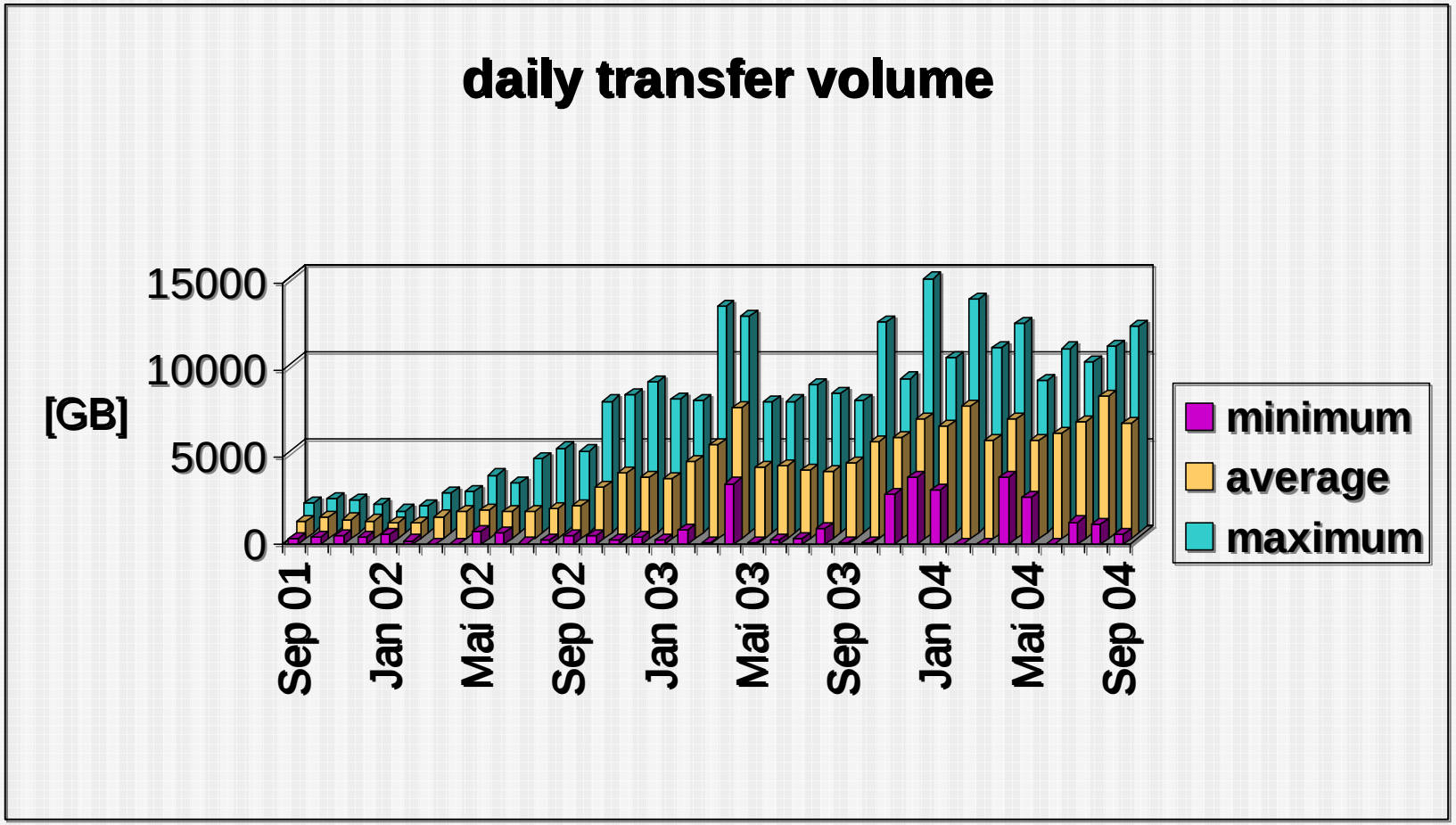
DS transfer rates (2001-2004)



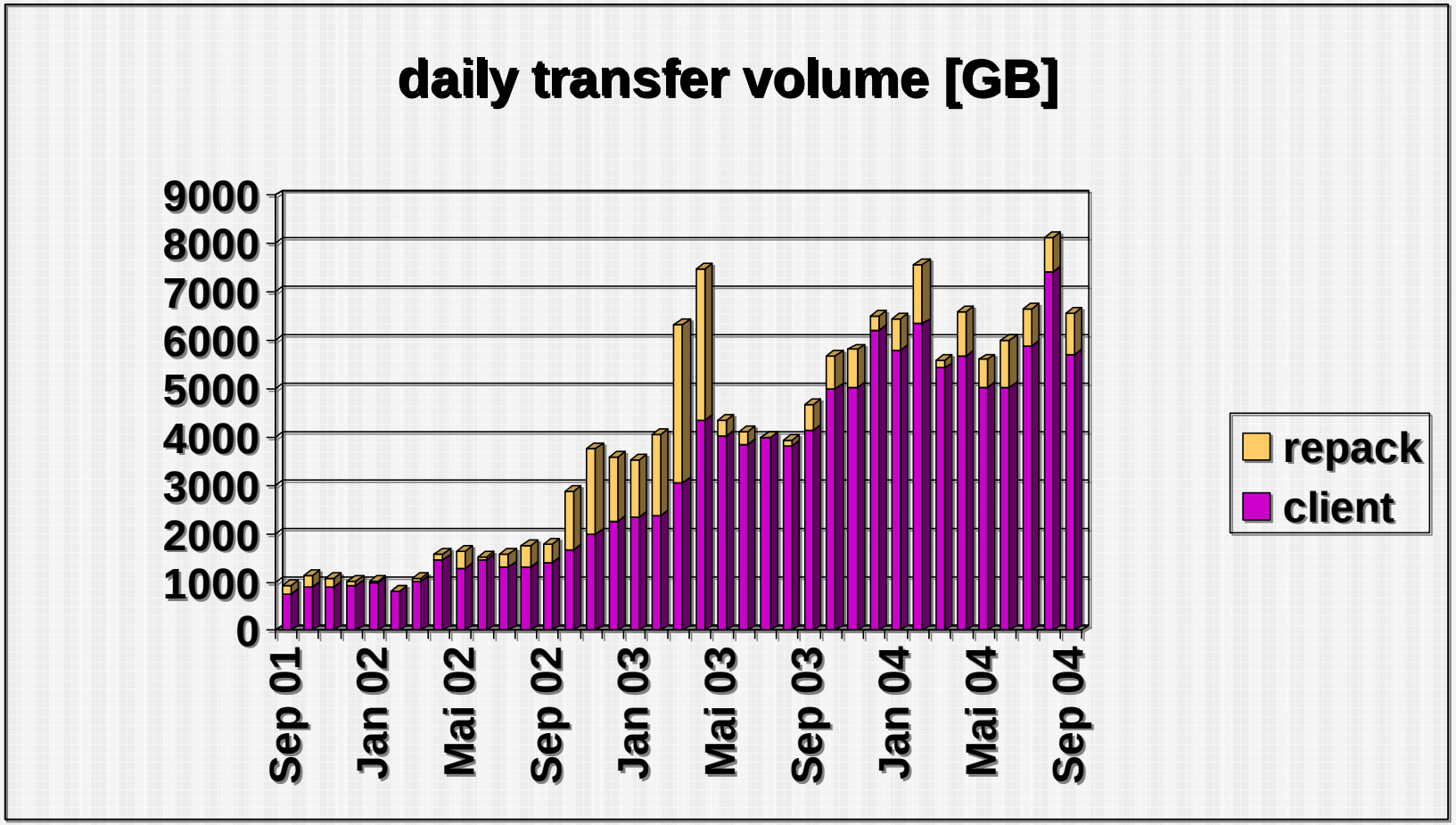
DS transfer rates (2001-2004)



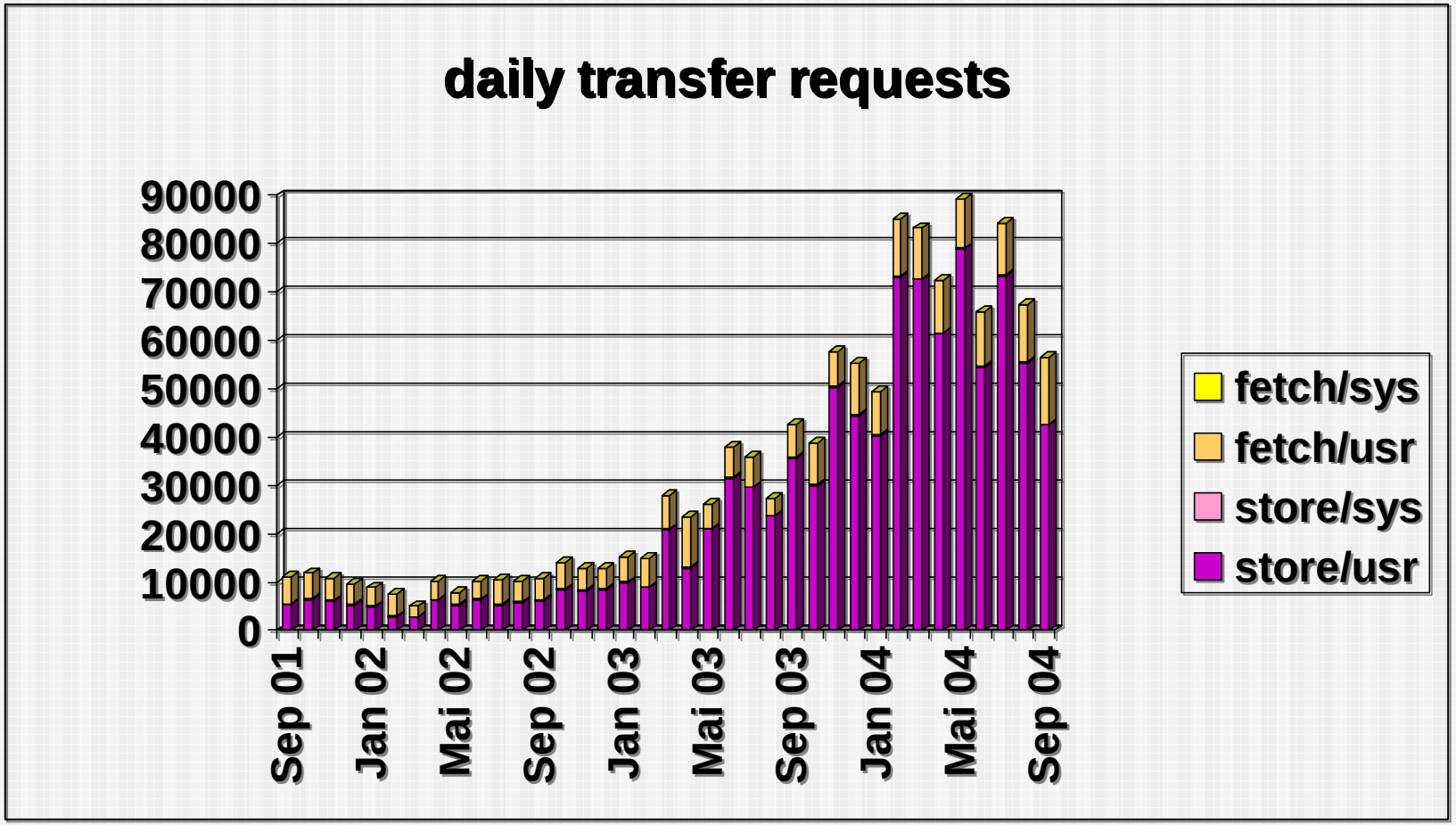
Tape transfer rates (2001-2004)



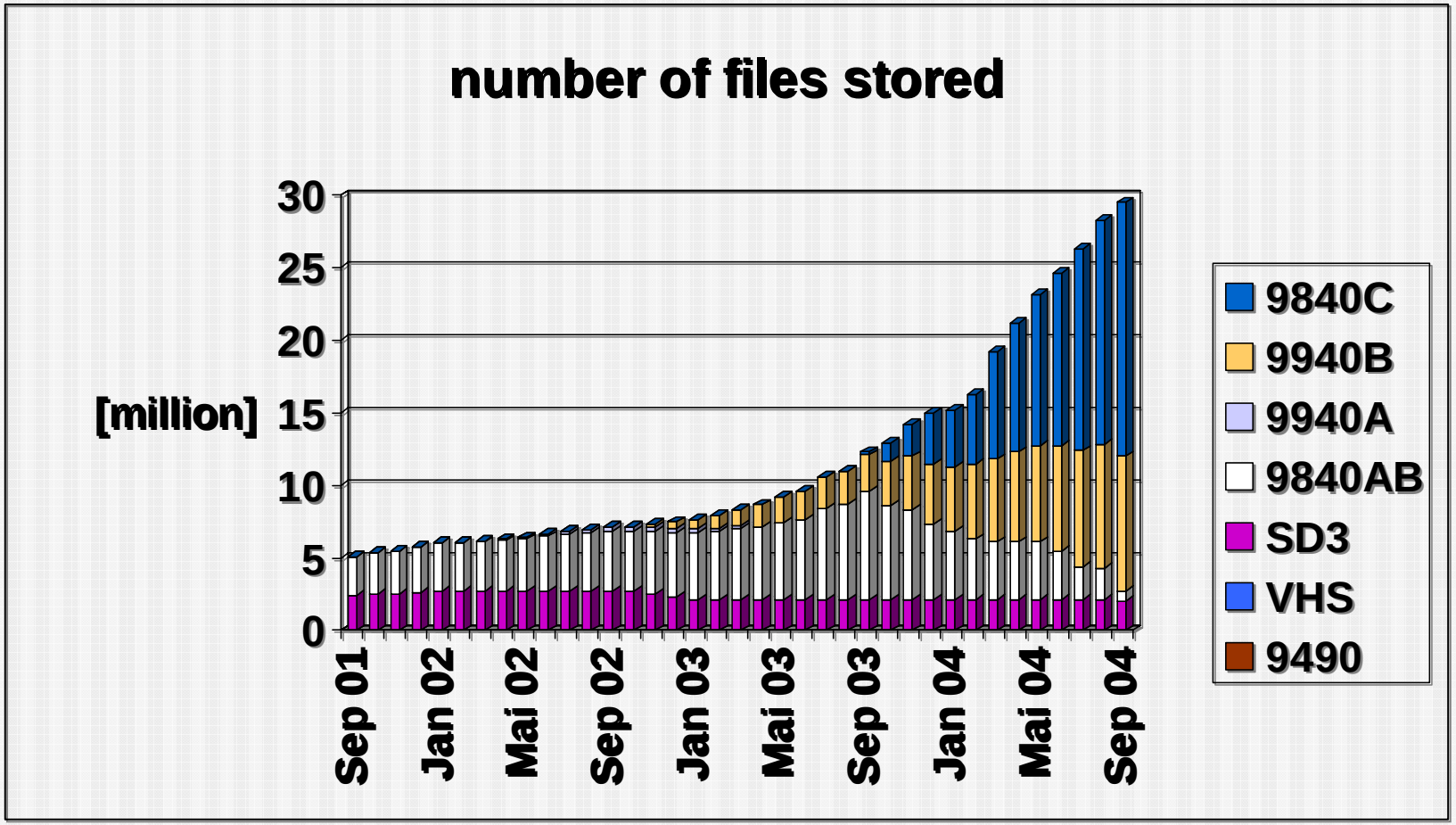
Tape transfer rates (2001-2004)



DS transfer requests (2001-2004)



DS archive capacity (2001-2004)



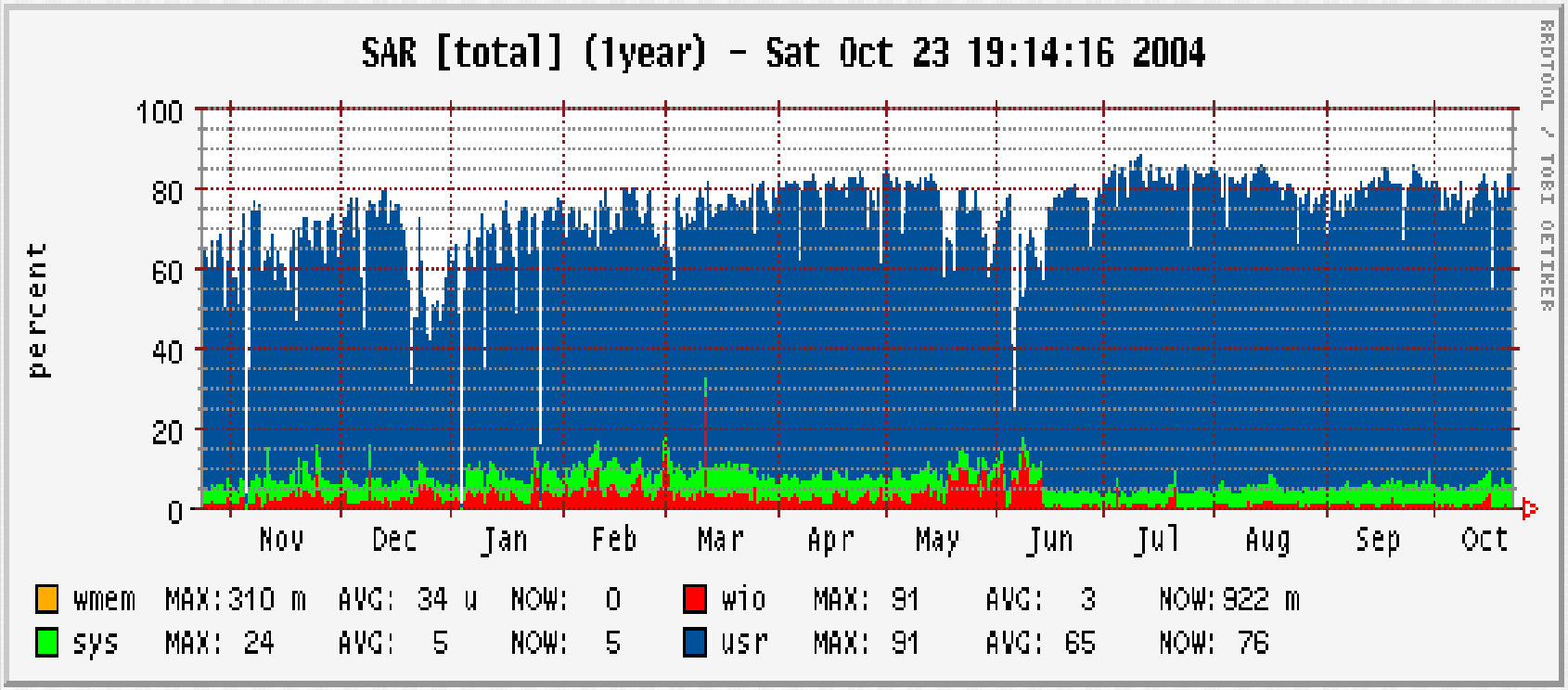
Some Lessons Learnt

- **Current Implementation of Non-Computing Services needs Significant Amount of Local Disk Space, e.g.**
  HSM and DBMS need their Own Cache
- **Lack of Standardisation for Shared Filesystems Dependence on Co-operativeness, e.g.**
  Graphics Server Integration
  Pre/Post-Processing Servers from Different Vendors
- **Fail-over Solutions needed in Complex Distributed Systems**

Some Lessons Learnt, cont.

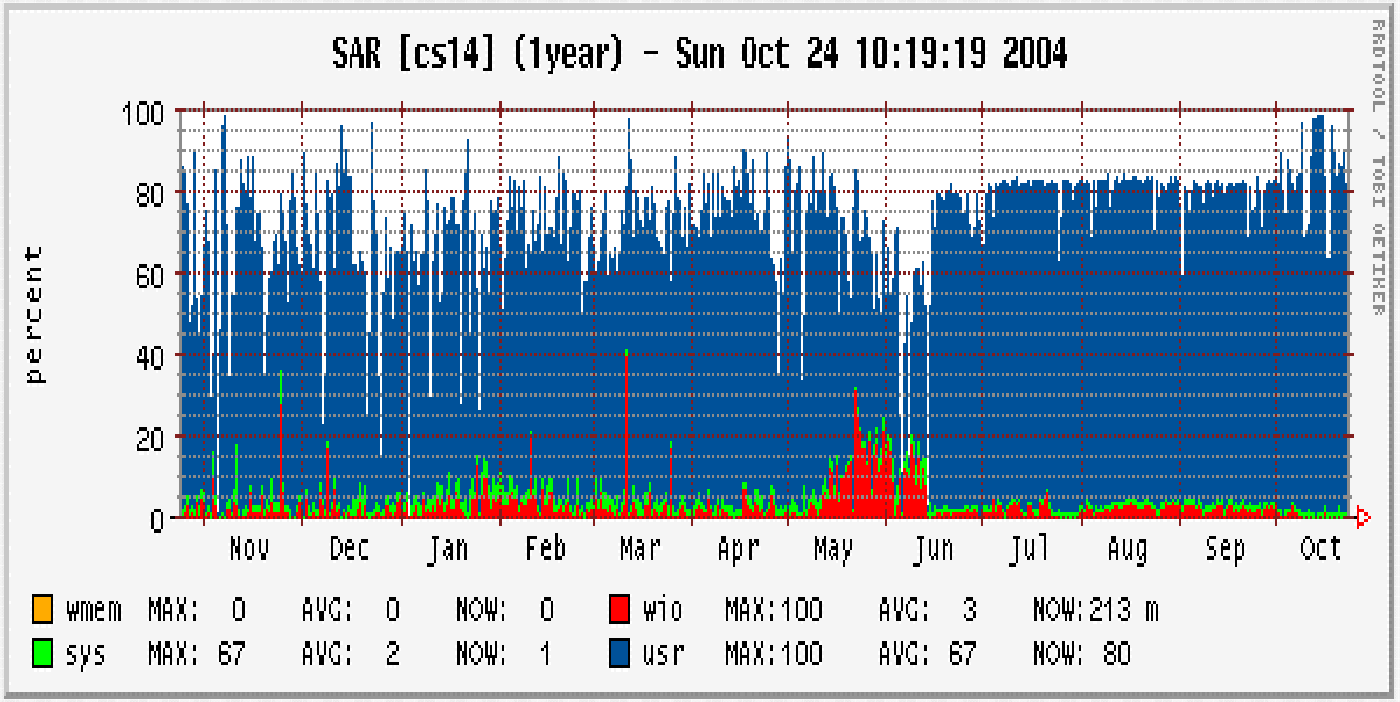
- **Server Scalability needed, but no Problem**
  Client Scalability may be a Problem, e.g
  128 LUN Limitation for Linux 2.4
- **Distributed Servers may Generate Intriguing Dependencies, i.e. clearly Structured High Level Services do not Guarantee Ease of Performant Operation**

Effect of Client/Server Interaction



Invocation Period and Lifetime of Dirty Pages for kupdated

Effect of Client/Server Interaction



Invocation Period and Lifetime of Dirty Pages for kupdated

Summary

- DKRZ provides Computing Resources for Climate Research in Germany on an competitive international level
- The HLRE System Architecture is suited to cope with a compute- and data-intensive Usage Profile
- Shared Filesystems today are operational in Heterogenous System Environments
- Standardisation-Efforts for Shared Filesystems needed


Thank you for your attention !