Experiences with gStore, a scalable Mass Storage System with Tape Backend

Chep 2007,Victoria
Sep 3, 2007

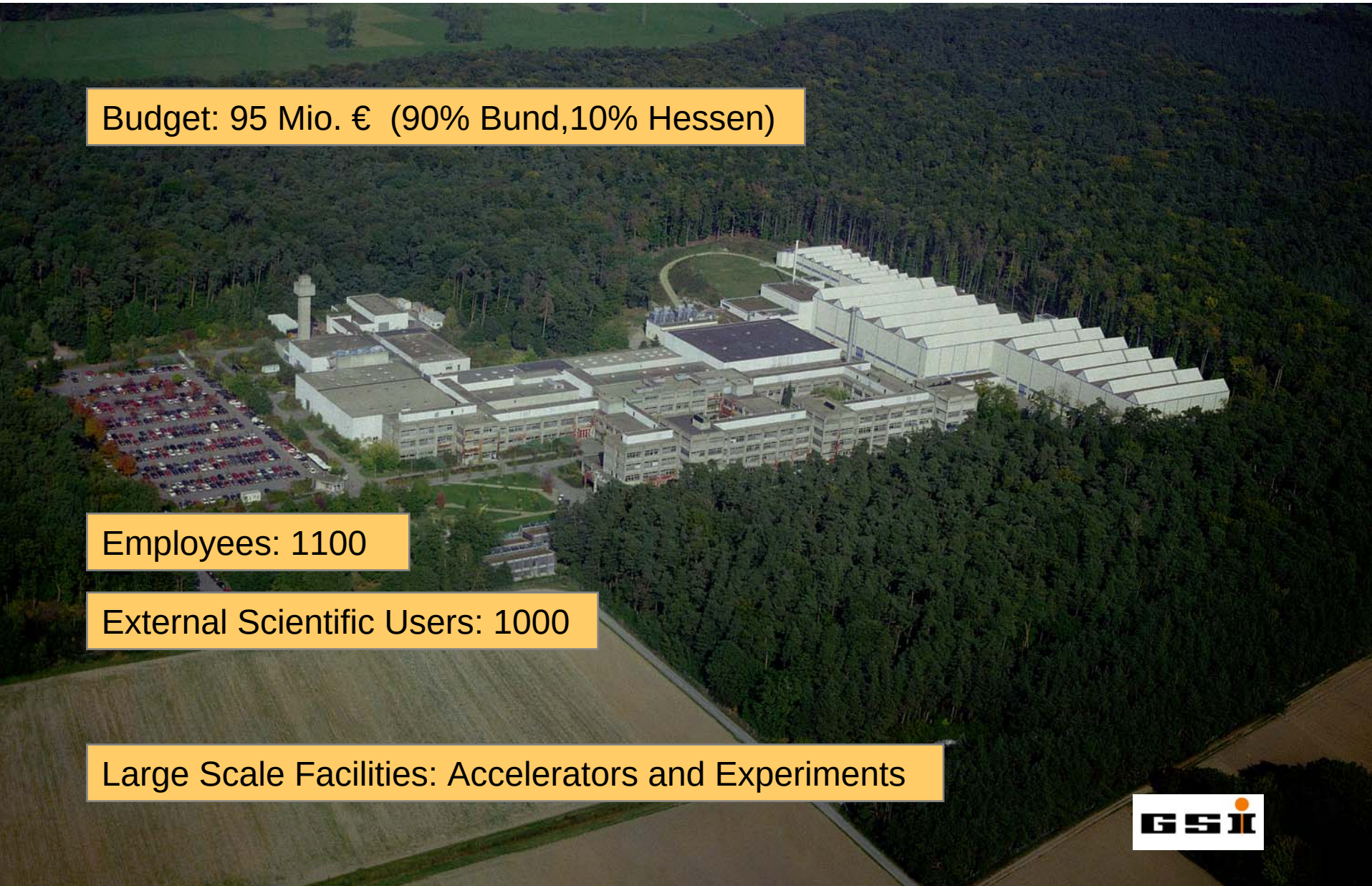
Horst Goeringer, Matthias Feyerabend, Sergei Sedykh

H.Goeringer@gsi.de

Overview

1. Introduction GSI
2. design principles of gStore
3. the current system
4. current projects
5. final remarks

GSI - Gesellschaft für Schwerionenforschung
Centre for Heavy Ion Research



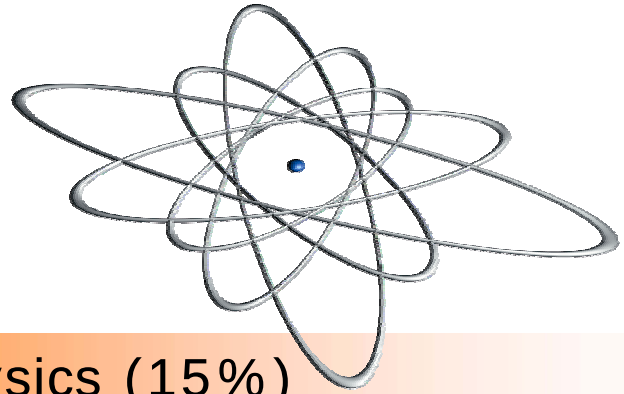
Budget: 95 Mio. € (90% Bund,10% Hessen)

Employees: 1100

External Scientific Users: 1000

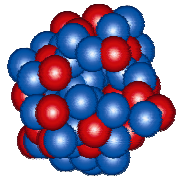
Large Scale Facilities: Accelerators and Experiments

Research Areas at GSI



Nuclear Physics (50%)

- **Nuclear reactions up to highest energies**
- **Superheavy elements**
- **Hot dense nuclear matter**

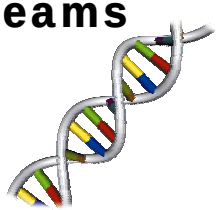


Atomic Physics (15%)

- **Atomic Reactions**
- **Precision spectroscopy of highly charged ions**

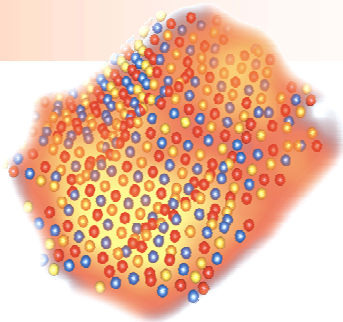
Biophysics and radiation medicine(15%)

- **Radiobiological effect of ions**
- **Cancer therapy with ion beams**



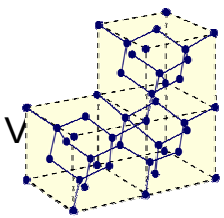
Plasma Physics (5%)

- **Hot dense plasma**
- **Ion-plasma-interaction**



Materials Research (5%)

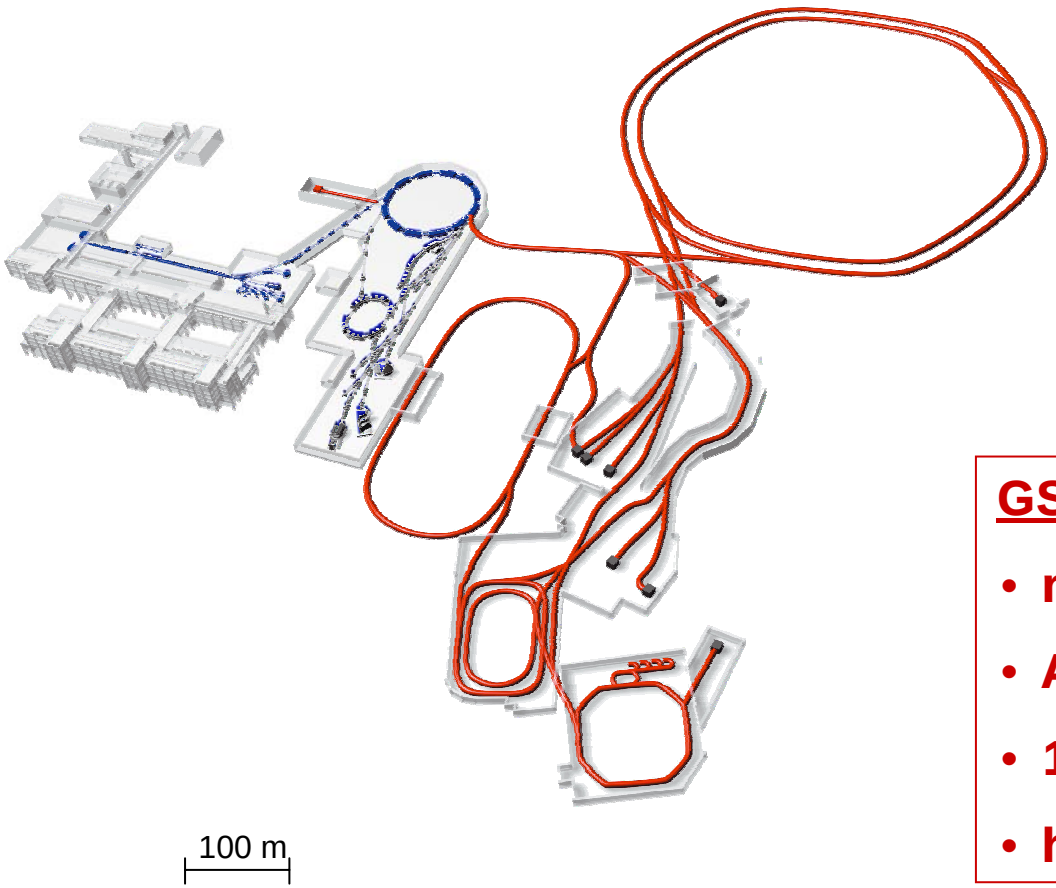
- **Ion-Solid-Interactions**
- **Structuring of materials with ion beams**



Accelerator Technology (10%)

- **Linear accelerator**
- **Synchrotrons and storage rings**

FAIR – Facility for Antiproton and Ion Research



GSI-today

- all kinds of ions
- max. 90 % speed of light

GSI-tomorrow / FAIR

- new isotopes
- Anti-Protons
- 10.000 times more sensitive
- higher speed

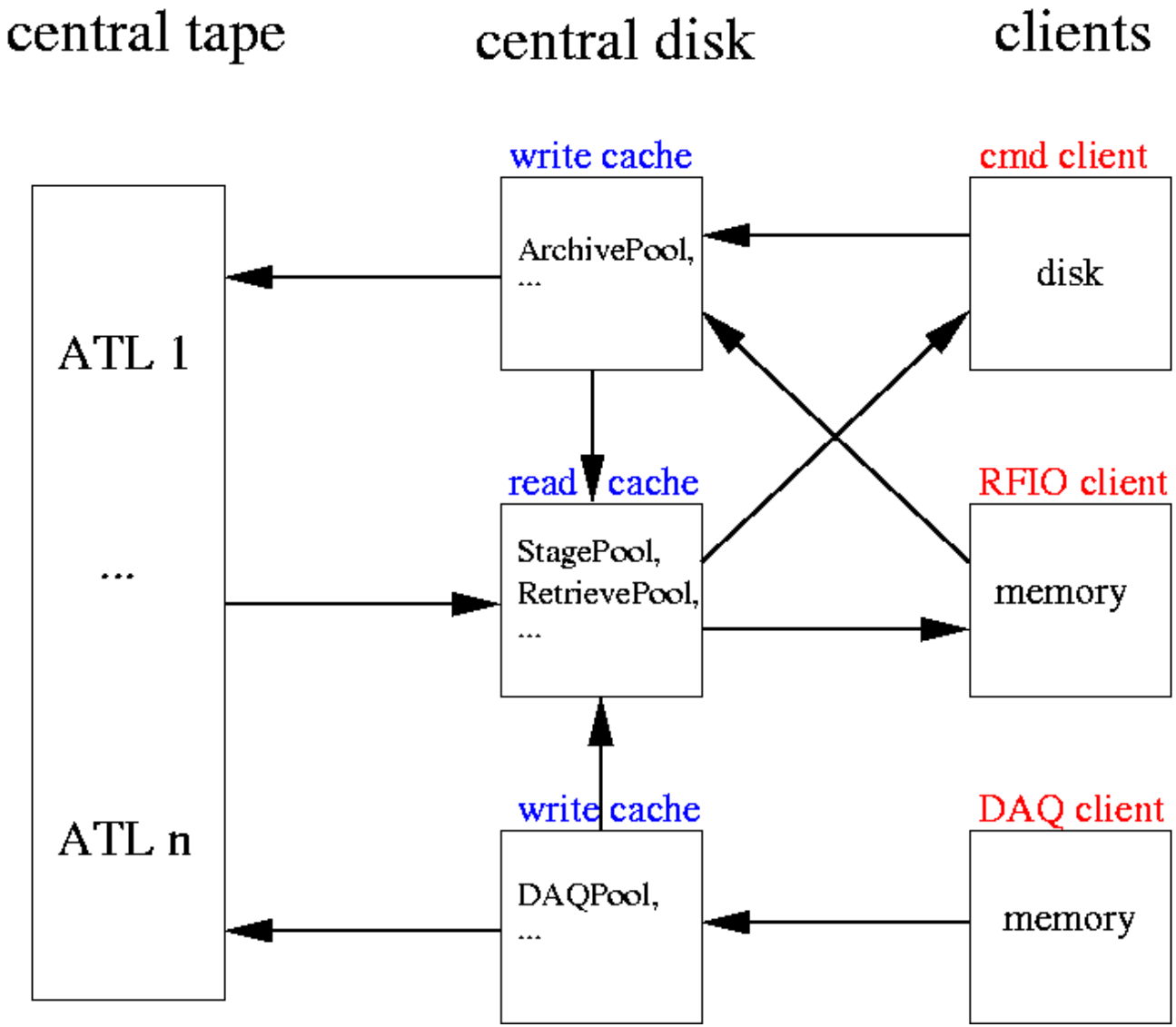
FAIR – Facility for Antiproton and Ion Research

gStore: software view

gStore: GSI Mass Storage System

1. **TSM: Tivoli Storage manager**
   - **commercial**
   - **handles ATLs and tapes**
2. **GSI Software:**
   - **Interface to users**
   - **API to TSM**
   - **management read/write cache**

gStore: storage view

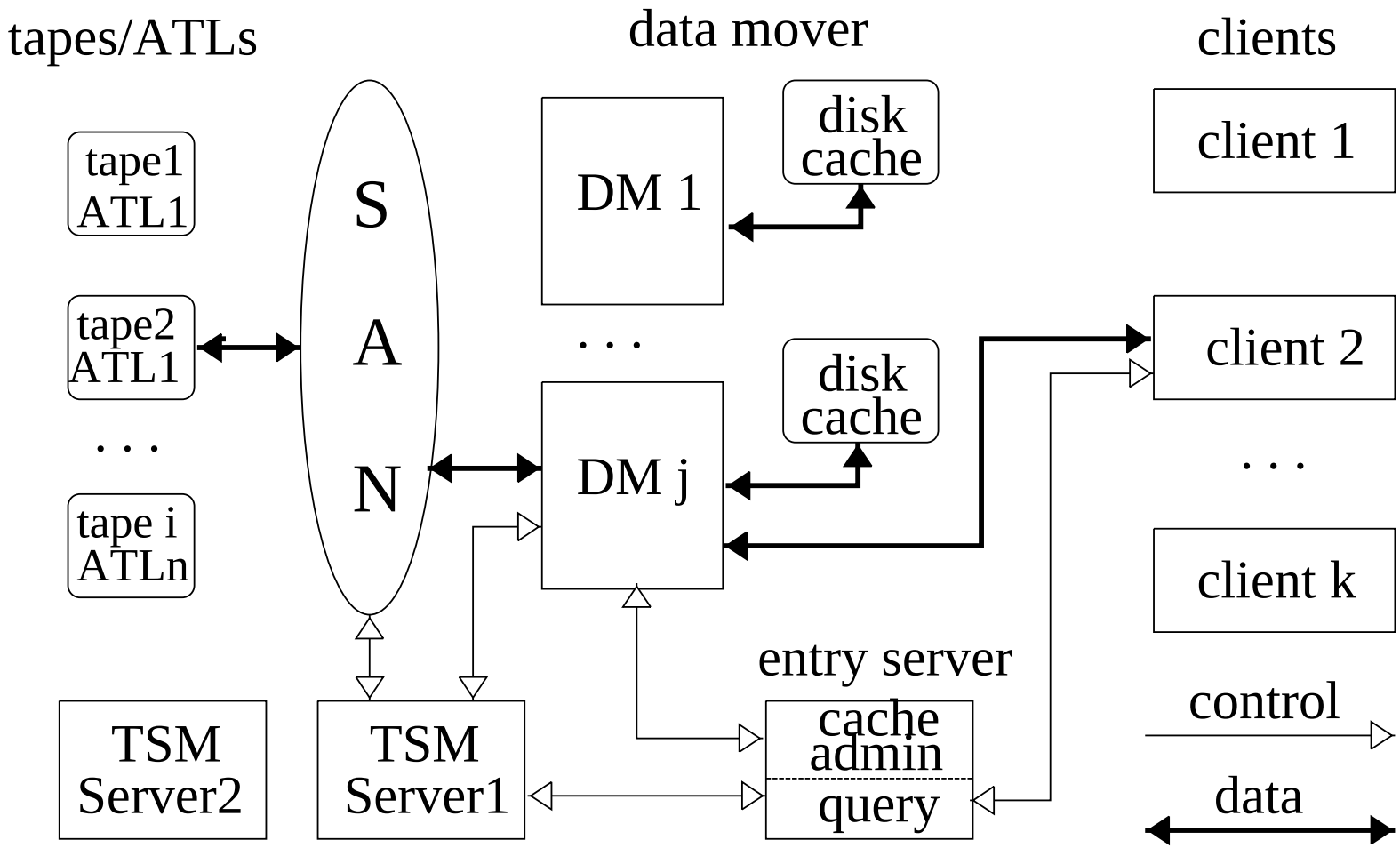


gStore: concept

Concepts:

- **separation of control and data flow:**
  - **control flow: TSM Server, Entry Server**
  - **data flow: Data Mover**
- **many DMs => many parallel data streams**
- **SAN: Storage Area Network**
- **Cache Manager: read and write cache**
- **direct connection DAQ (event builder) to exclusive gStore write cache**

gStore: overview



gStore: history

System scalable:

- **Feb 1997: Start**
  - **1 ATL, 1 TSM Server, 1 data mover**
  - **no SAN, no disk cache**
- **Jun 1998: 1st disk cache**
  - **80 GB, read cache only**
- **Jan 2003: SAN, 9 data movers**
- **Feb 2005: write cache, DAQ client**
- **May 2007: 2 TSM Servers**

developed with < 2 FTE!

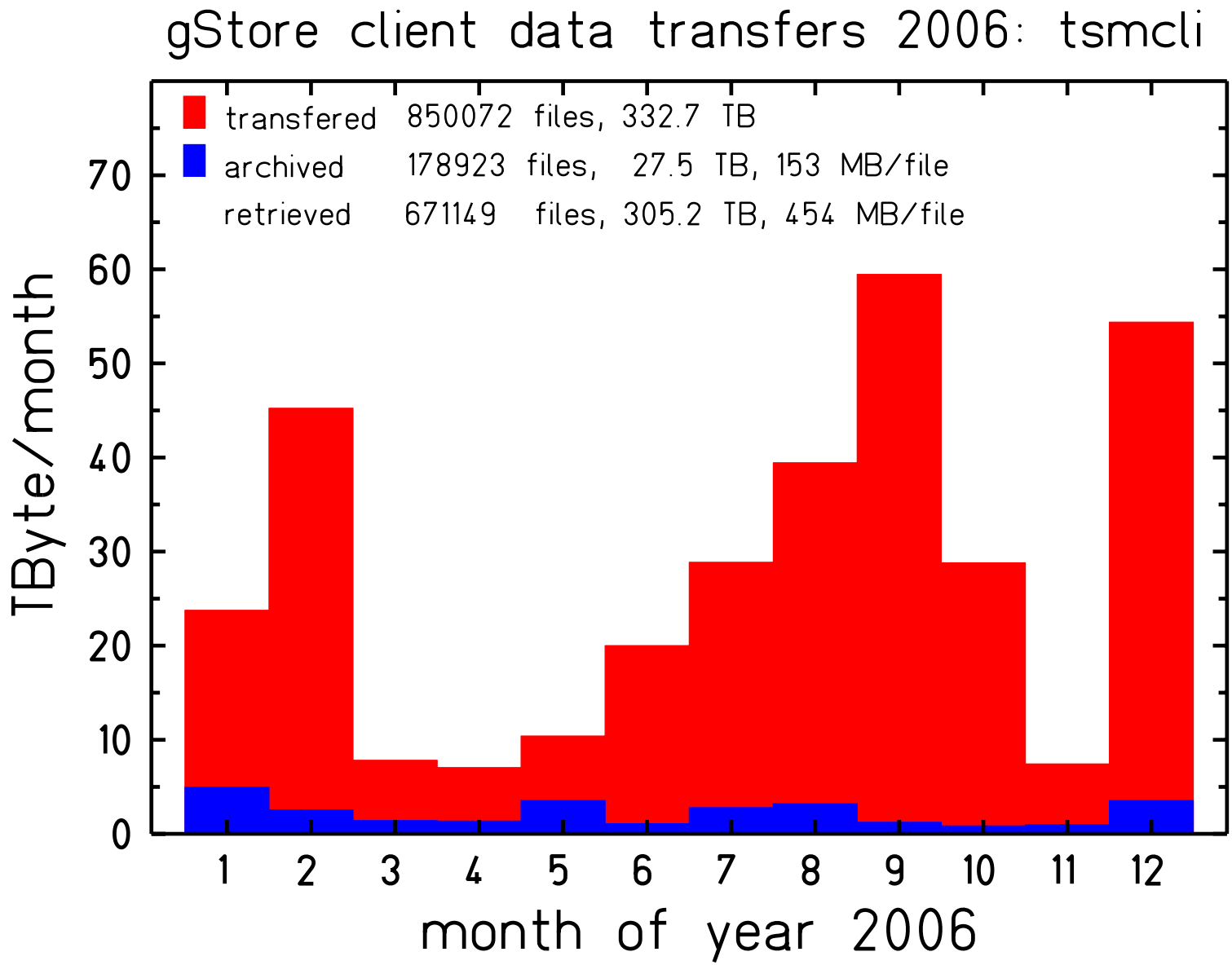
gStore hardware I

1. **TSM server:**
   - **POWER5 p5x0 (AIX 5, TSM 5.4)**
   - **ATL: IBM 3494**
     - **4 (6) tape drives IBM 3592:**
       - **100 MByte/s,**
       - **700 GByte/vol**
     - **max 1.6 PB data capacity**
   - **data mover:**
     - **8 x Linux, 30 TB disk cache**

gStore hardware II

2. TSM server:
   - x86 cluster (Windows 2000, TSM 5.4)
   - ATL: Sun StorageTek L700
     - 9 tape drives LTO2:
       - 35 MByte/s,
       - 200 GByte/vol
     - max 140 TByte data capacity
   - data mover:
     - 9 x Windows, 4 TB disk cache
     - 1 x Linux, 0.4 TB disk cache)

gStore usage 2006



gStore load 2006

overall data transfer 2006:

- **333 TB via LAN**
- **120 TB via SAN**
- **overall ~0.45 PB**

top data transfer in 2006: Dec 31

- **overall: 9.6 TB in 24 h**
  - **111 MB/s on average**
- **1 data mover: 2.9 TB in 24 h**
  - **33.6 MB/s on average**

current projects

- 2nd level DM: no SAN connection
- gStore as backend to xrootd
  - in test environment (Alice tier 2): gStore access for xrootd clients available
- Grid SRM (Storage Resource Manager)
  - under investigation
- lots of user requests ...

Final Remarks I

- currently ca. 250 TB of exp. data on tape
  - including 50 TB copies
- 1.6 – 2 PB max tape capacity
- 35 TB disk cache (1st level)
- I/O bandwidth
  - > 1 GB/s cache <-> clients (LAN)
  - < 0.9 GB/s cache <-> tape (SAN)
    - Hades DAQ end 2008: 200 MB/s

Final Remarks II

- **gStore fully scalable in data capacity and I/O bandwidth**
- **supports several TSM servers**
- **gStore fully flexible in hardware (TSM)**
- **gStore adaptable for cooperation with external software packages**

Final Remarks III

- in the past 10 years (<2 FTEs): managed
  - growth of 1-2 orders of magnitude
  - various hardwares and platforms
- gStore prepared for further growth
- FAIR requirements to IT in 2015

  (experiments CBM, Panda):

  similar to actual LHC experiments