The Fermilab Data Storage Infrastructure

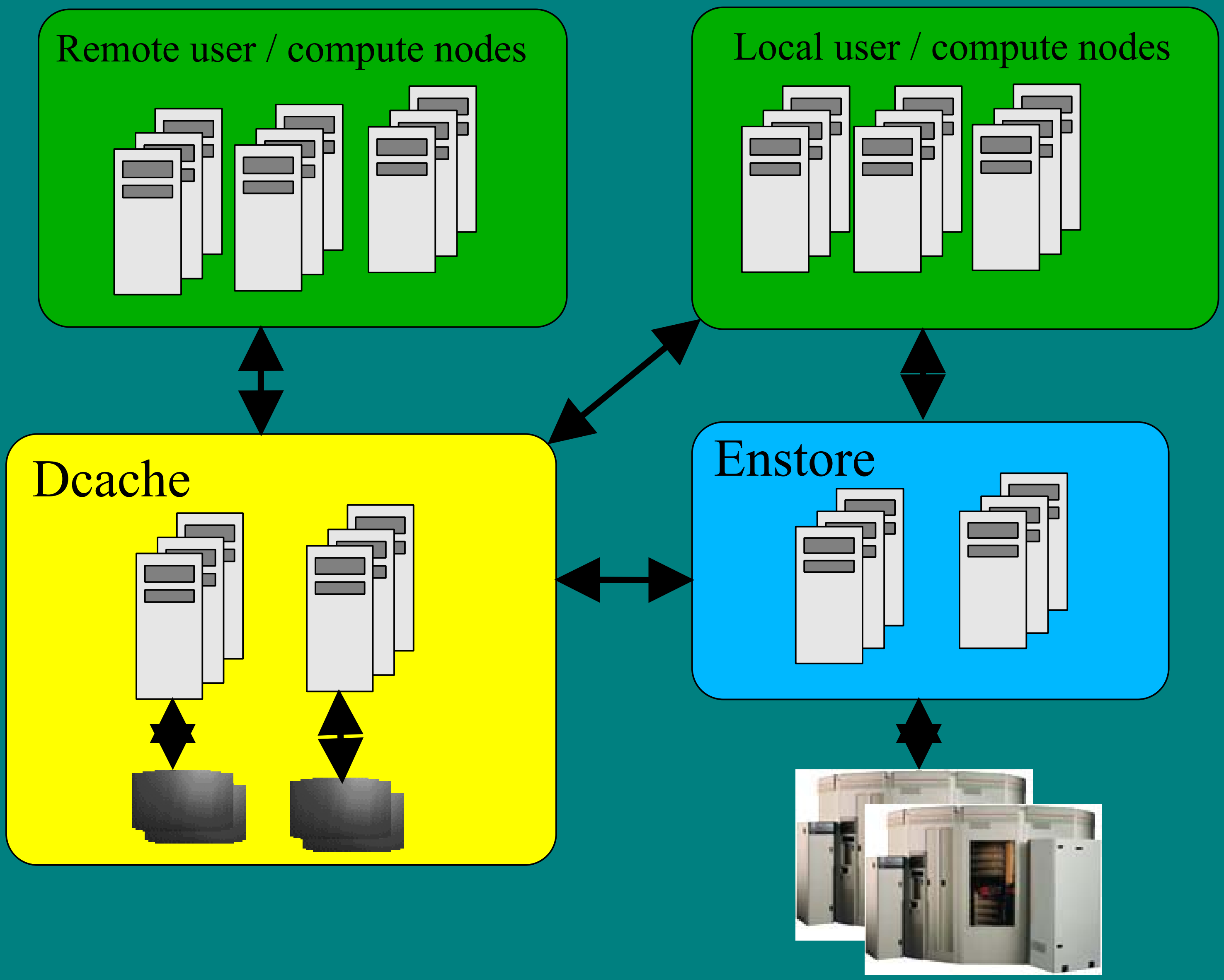
Alexander N. Moibenko
Fermi National Accelerator Laboratory

April 8, 2003

Fermilab Storage Requirements

- Capacity: several Petabytes
- Performance
- Data acquisition rate: 20 MB/s and more sustained for 30 days
  - Overall rate: 250 MB/s and more
- Data Access: need control of data placement on tape
- Import/export: easy data exchange with other labs.
- Operation:
  - Efficient resource management
  - Fault tolerance
  - Easy administration and monitoring
  - Lights-out operation
- Media: flexibility in media selection
- Flexibility: addition of new features and quick bug fixes
- Scalability: capacity, rates, concurrent access

Fermilab data storage infrastructure



Enstore Data Storage System

Primary Data Store for large data sets

- Distributed access to data on tapes
- High fault tolerance and availability
- Priority based request handling including interception of resources by DAQ requests
- Request handling with look ahead and sorting capabilities
- Configurable resource management: storage groups
- Grouping of similar sets of data: file families
- Unix like data storage presentation
- Encp as enstore user interface:

  encp [options] <source> <destination>

  - Self described data on tapes
  - Tape import /export in robot: exchange data with other labs.

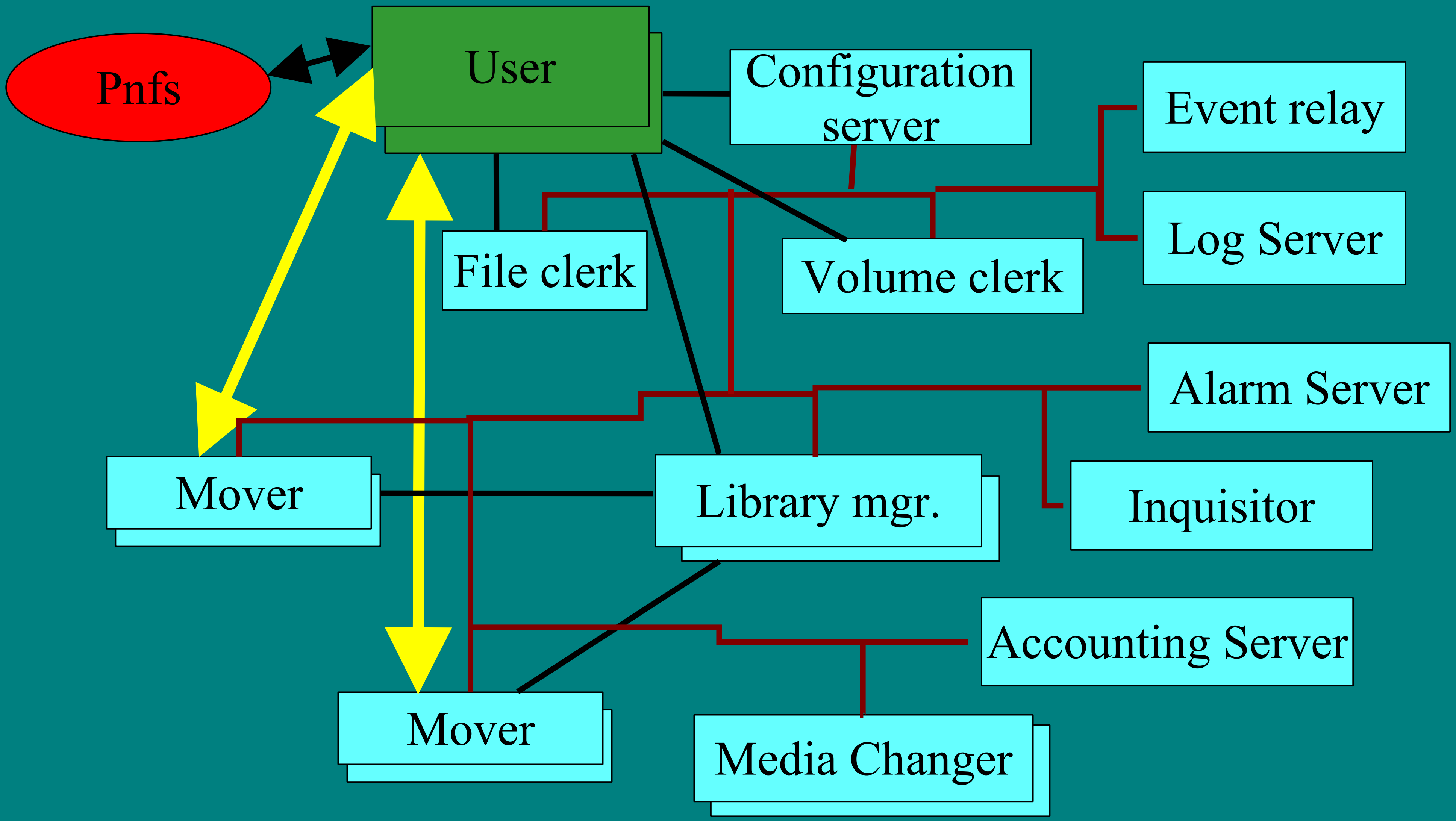
Hardware

- 1 ADIC AML2 and 4 STK robotic tape libraries
- 28 STK 9940A Tape Drives
- 12 STK 9940B Tape Drives
- 8 9840 Tape Drives
- 9 IBM LTO 1 Tape drives
- About 90 PC nodes

Software Design Approach

- Open Source and free ware software
  - Libdb and Postgres DB
  - GNU Plot Tools
  - Apache web server
- Portable code not depending on HW and OS
  - Python as major programming language
  - Time critical code in C with ability to compile in different OS
- Modularity
- Client / server architecture
- Reuse products
  - FTT - Fermi Tape Tool
  - PNFS - namespace

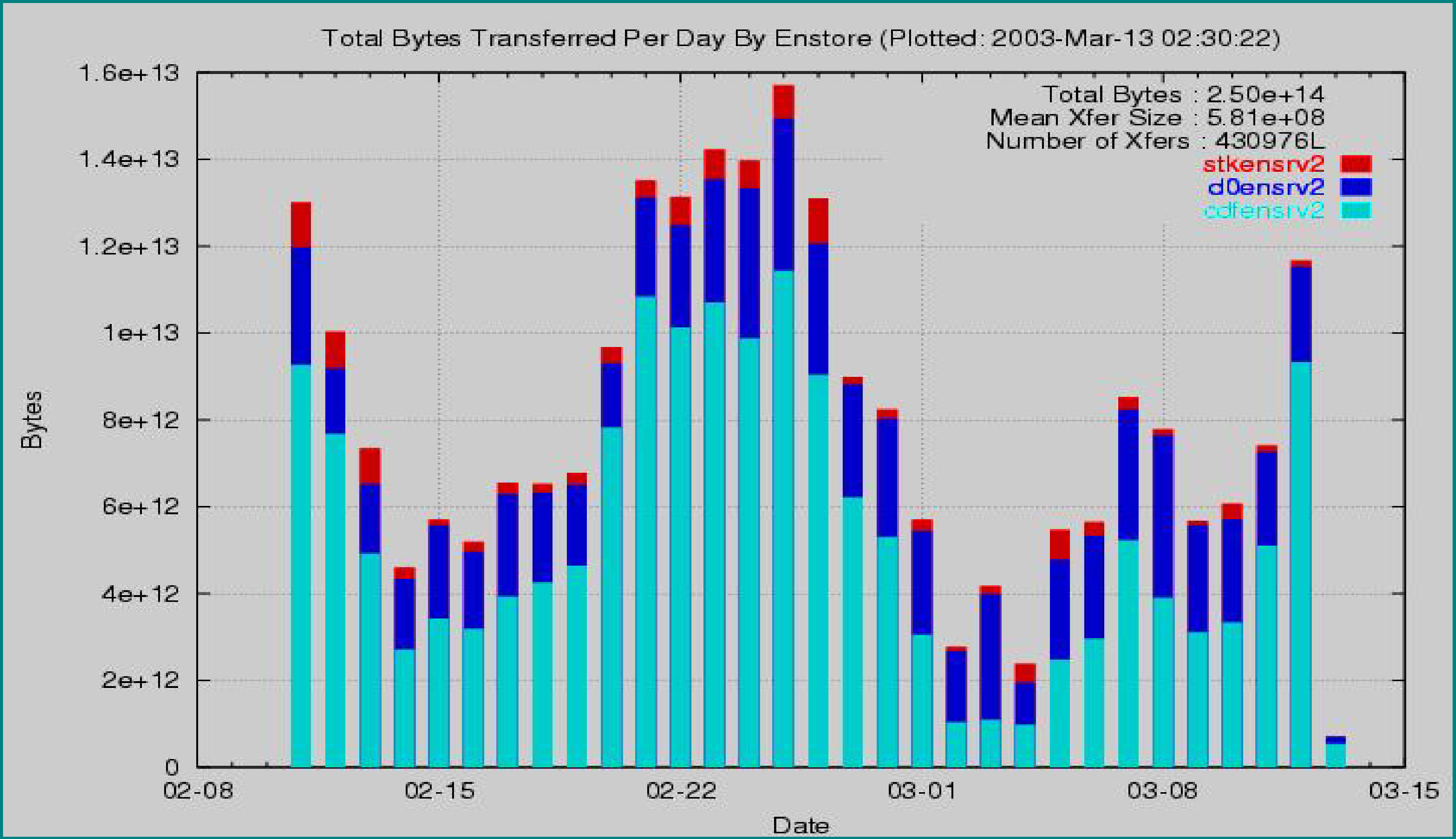
Enstore Architecture



Enstore monitoring

- State of enstore servers
- Amount of resources
- Request queues
- Data movement, throughput, tape mounts
- Volume information
- Alarms
- Completed transfers
- All information is published on the Enstore web site: http://www-hppc.fnal.gov/enstore/
- Entv - graphical data transfer presentation

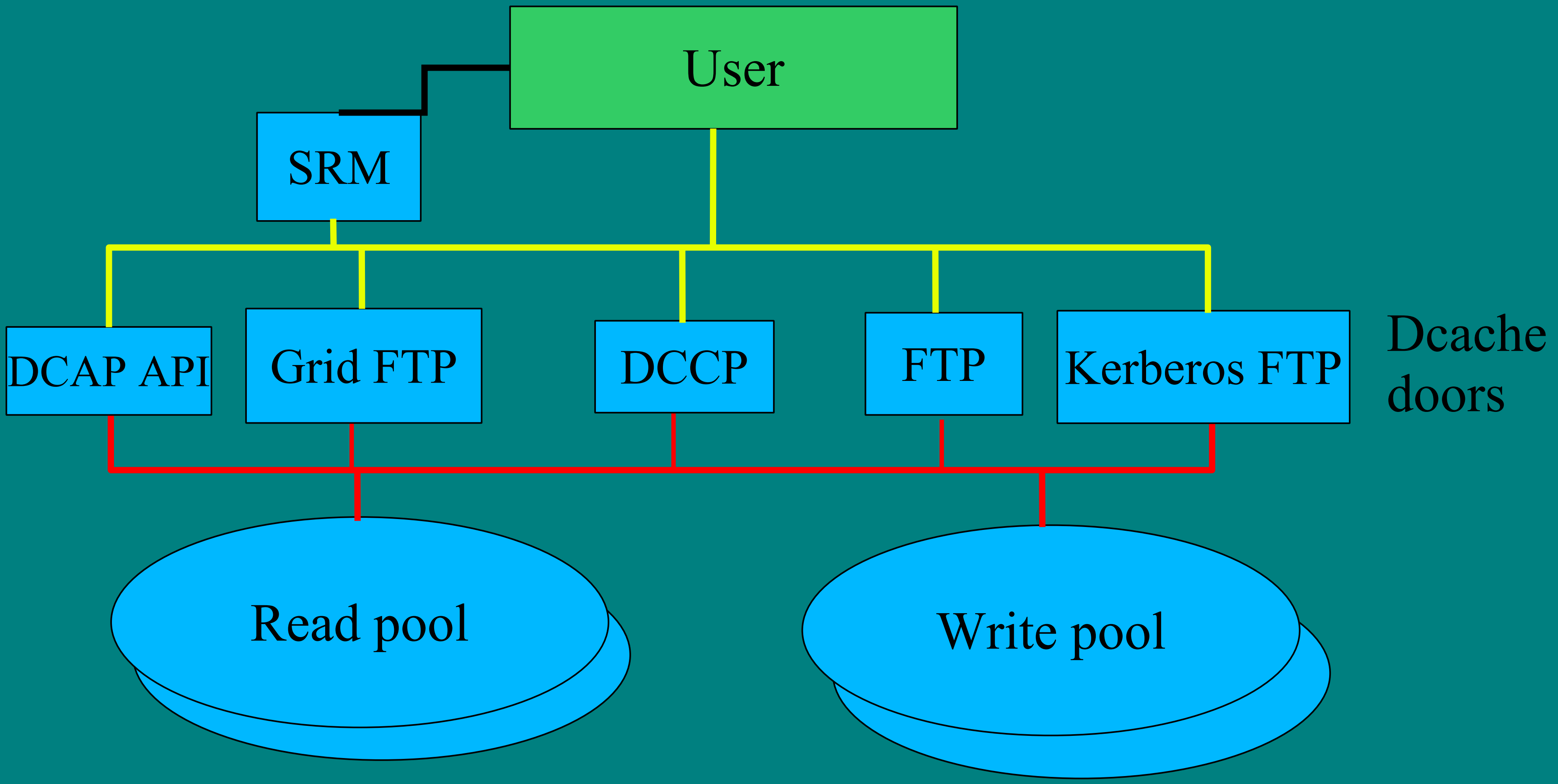
Data Transferred Per Day by Enstore



Dcache

- Developed primarily by DESY in collaboration with FNAL
- Buffers data on disks
- Migrates data to Enstore
- Rate adaptation
- Deferred writes
- Data staging
- Read ahead
- Partial file reads (required by root)

Simplified structure of Dcache



Dcache configuration

- Based on use of inexpensive computers
  - Usually PCs running Linux
  - Few Suns
- 4 separate Dcache systems
- Overall capacity 100 TB
- 150 nodes
- Rapidly growing
- More doors added to integrate into GRID computing

Conclusion

- The Fermilab Data storage infrastructure has been successfully used for several years
- Meets the requirements of experiments
- Uses inexpensive hardware
- Robust and fault tolerant
- Easily scales to meet increasing capacity and throughput requirements
- Reacts fast to DAQ requests
- Accessible from everywhere