# MONITORING A PETABYTE SCALE STORAGE SYSTEM


Jon Bakken – Fermilab
Eileen Berman – Fermilab
Chih-Hao Huang – Fermilab
Alexander Moibenko – Fermilab
Don Petravick – Fermilab
Michael Zalokar - Fermilab


Fermilab operates a petabyte scale storage system, Enstore, which is the primary data store for experiments' large data sets. The Enstore system regularly transfers greater than 15 Terabytes of data each day. The goal of the monitoring implemented to support this facility, is to provide sufficient high quality monitoring to -

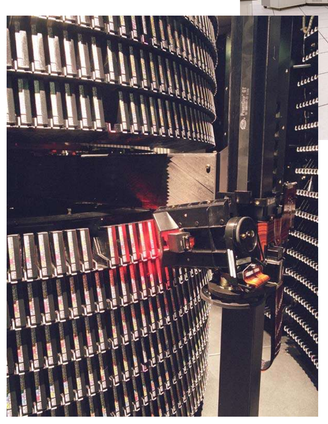
- Insure data integrity.
- Enable problems to be quickly and efficiently located and solved.
- Enable users to monitor progress of their requests.
- Provide information to be used to predict future needs (tapes, tape drives...)

To accomplish this, monitoring of a wide-variety of components is done -

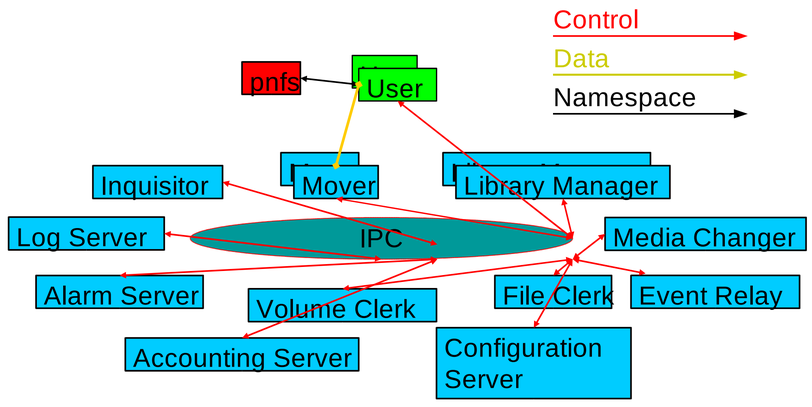
- Enstore software and transfer status
- Volumes and Automated Tape Libraries
- Hardware and Firmware
- Security
- Networks
- Metrics

for several different roles -

- Users
- Administrators
- Developers
- Site system operators
- Managers



**STK Automated Tape Library (5 silos total)**




★★ Used frequently by developers

## Networks

*(mostly automated to generate alarms)*

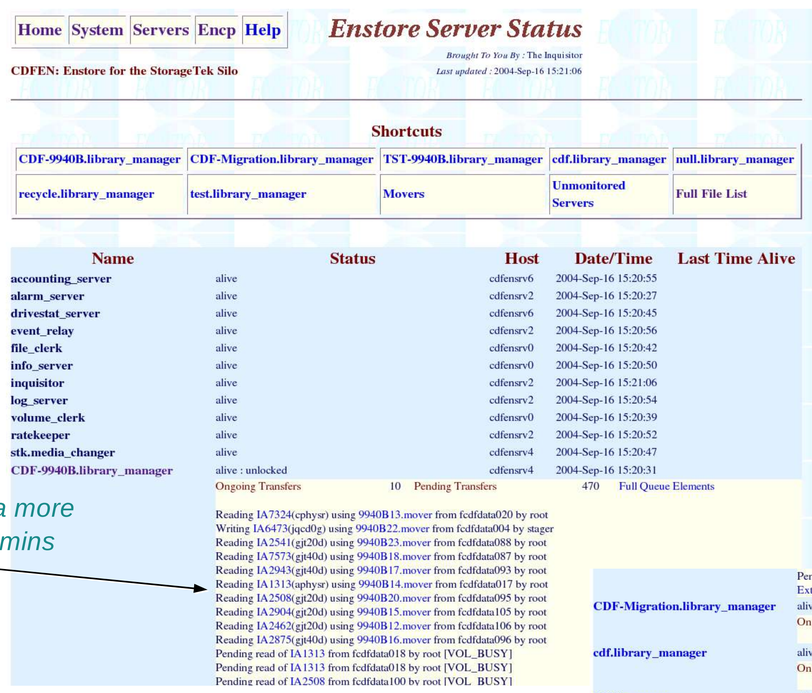
**Transfer rate – simulate a real transfer**
**UDP network monitoring**
**Network rate**

★ Used by admins , managers

## Security

*(automated to generate alarms)*

**State of write tabs on tapes – augments security of data**
**Time synchronization**
**Check for open LISTEN sockets– part of security checks**




★★★ Used by users, admins, operators

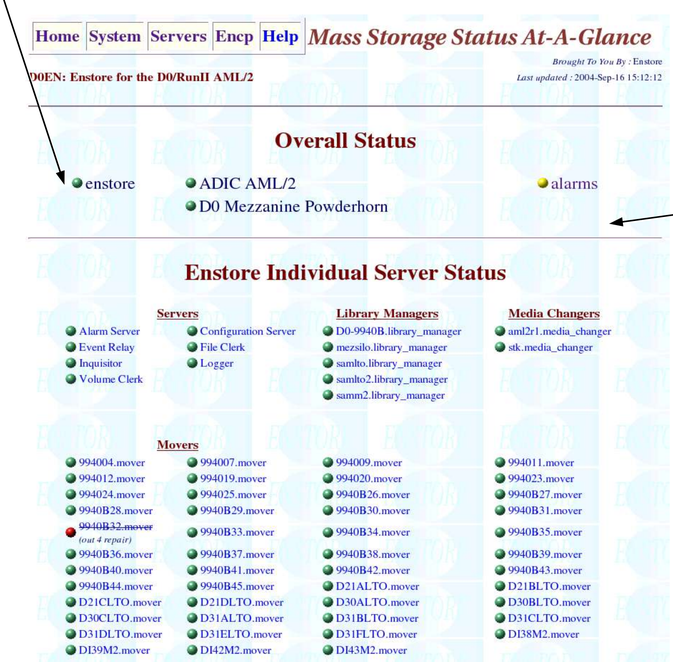
*Gives immediate view of system as a whole, problems result in automated pages of administrators*

★★ Used by users, admins, developers

*Request queues monitored for a more detailed view of problems by admins*

★ Used by admins

*Current transfer activity is useful to administrators for quick problem location*



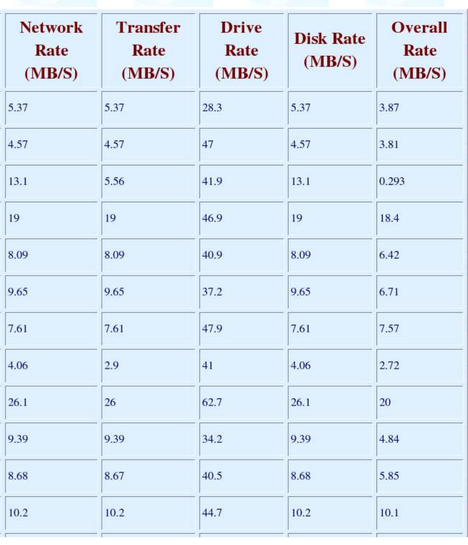

## Enstore

*(some automated to generate alarms, others admins watch)*

**Status-at-a-Glance**- gives single view of system state
**Request queues**
**Server/Mover status**
**Transfer history**– successful and error
**Current Enstore configuration**
**Active alarms**– work list for administrators
**Log files**– available on web for access and search
**User files transferring or queued**
**Size of PNFS databases**
**Enstore log file size**

★★ Used by users, admins, developers

*Transfer history monitored to identify transfer , network, and drive problems*




★★★ Used heavily by admins, and by developers and managers

*Active alarm page is the work list for the administrators*

Current configuration obtained thru CLI not web page.

User files list not referenced, users go straight to queues.

# MONITORING A PETABYTE SCALE STORAGE SYSTEM


Jon Bakken – Fermilab
Eileen Berman – Fermilab
Chih-Hao Huang – Fermilab
Alexander Moibenko – Fermilab
Don Petravick – Fermilab
Michael Zalokar - Fermilab


Hardly used at all, took much effort to implement and make stable

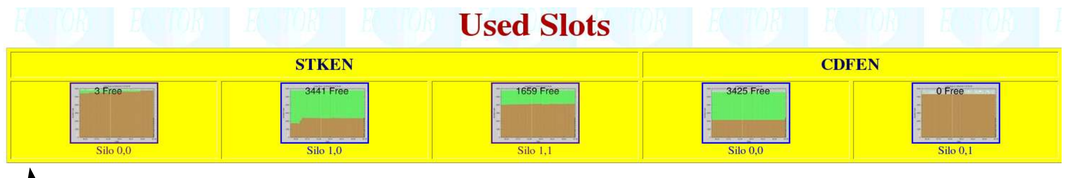
## Hardware/Firmware

*(mainly automated to generate alarms)*

- **Node configuration**
- **Status of raid disk**
- **Node CPU/Memory Information**
- **System event log**
- **Voltages – baseboard, processor**
- **Fan speeds – processor**
- **Temperatures – baseboard, processor**
- **Console logs**

**ADIC Automated Tape Library**

Used by users, admins



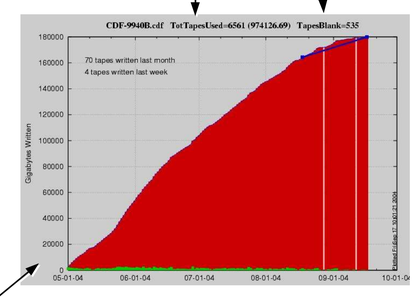

## Volume/Library

*(mainly monitored by administrators)*

- **Library Quota/experiment/type**
- **Cleaning tape status**
- **Rate of blank tape usage**– per experiment and per type
- **Empty slots in automated tape libraries**
- **Failed tape drive transfers**
- **Automated tape library logs**
- **Tape Drives**– serial #, type, firmware version
- **Volume information**
- **Volumes with too many mounts**
- **Volumes with errors**
- **Volume inventory**
- **Read random files**– part of data integrity checks

*Blank Tapes in Robot*

*Total Tapes in Robot*



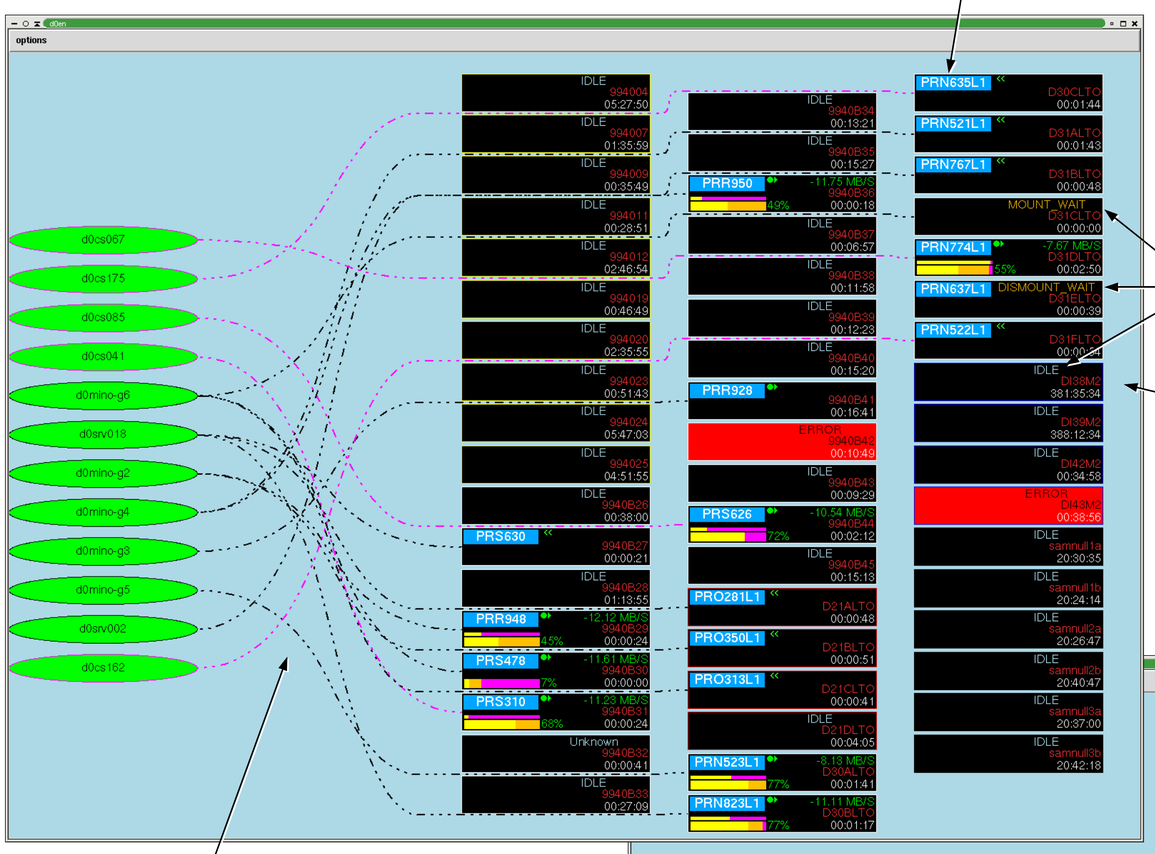

Used by users, admins

*Used to plan new purchases*

Used by everyone

**ENTV is a realtime monitoring program graphically representing the movement of data into and out of Enstore. It is the single most used monitoring tool provided.**

*Volume Label*

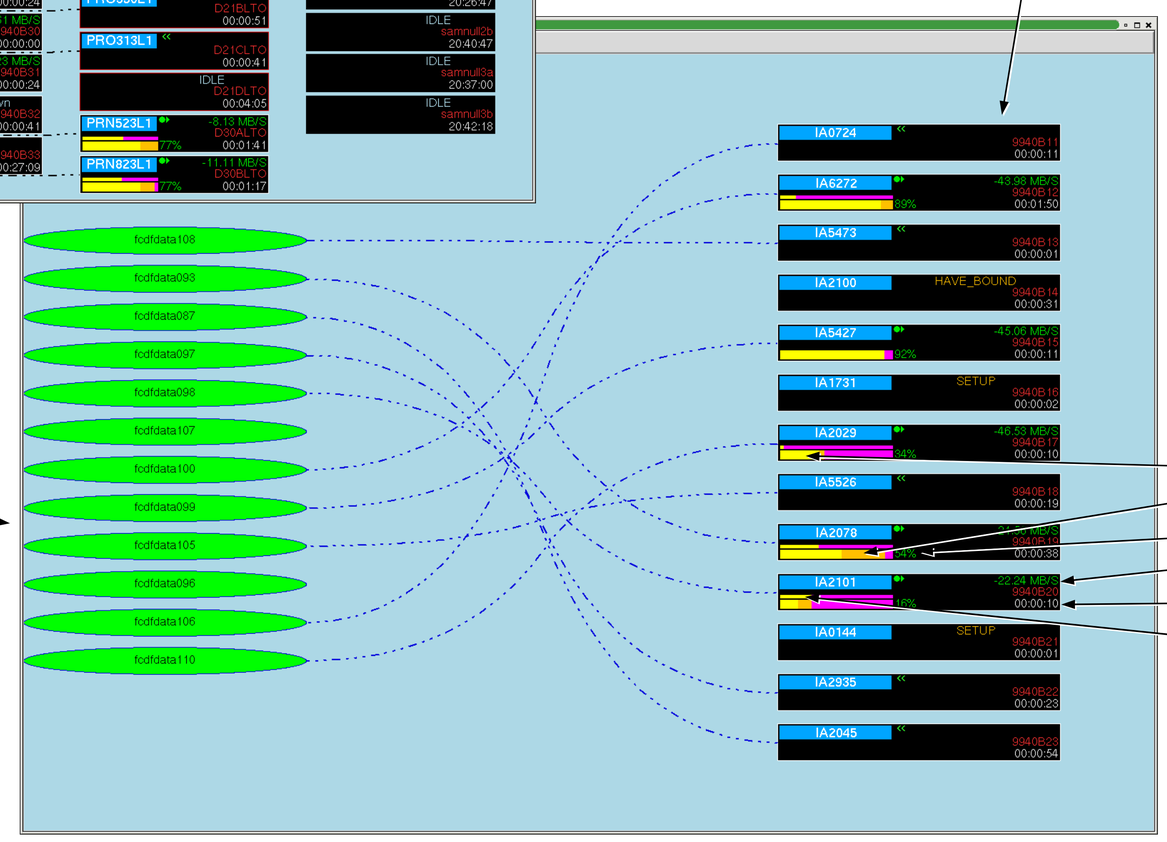



*State of transfer*

*Individual tape drives*

*Animated directional flow of data*

*User nodes accessing files on tape*




*Read*

- *Transferred to user*
- *Read from tape*
- *% transferred to user*
- *Network rate ('-' for read)*
- *Time in state*
- *In memory to be sent to user*

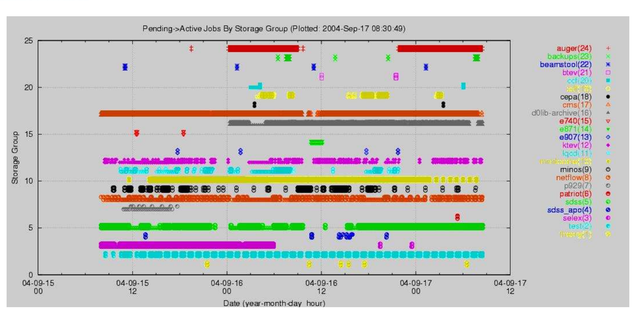
Used by users, admins, developers, managers

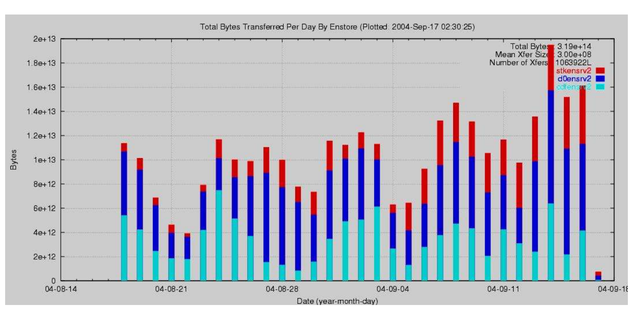
## Metrics

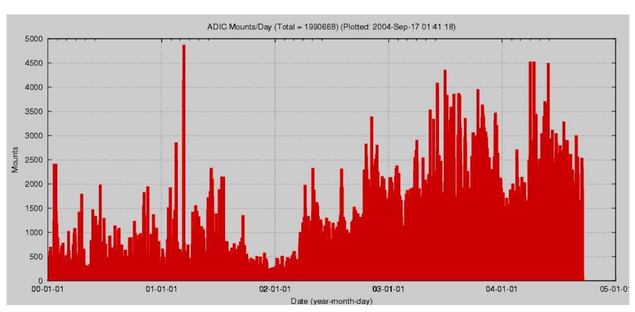
*(monitored by administrators)*

- **Total user data on tape**
- **Bytes written/day**
- **Bytes read/day**
- **Per Experiment use of the system**
- **Transfer activity**
- **Drive utilization snapshots**
- **Mounts/day**
- **Instantaneous rate of transfers**
- **Backup times**
- **Status of cronjob execution**

| Total User Data on Tape (Cdfen, D0en, Stken) : | 1884.795 TB |
|---|---|




## Conclusion

**The monitoring of the Mass Storage Systems at Fermilab has been a great success. Users, administrators, operators, developers, and managers make use of this monitoring to do problem resolution, maintain data integrity, follow file requests, and plan for future needs, both in the long term and the short term. Future improvements to the monitoring should include the ability to easily track the progress of a single job by a user.**