LBNL/NERSC Site Presentation for HPSS User Forum 2003

Shreyas Cholia (scholia@lbl.gov)

Lawrence Berkeley National Laboratory

National Energy Research Scientific Computing Center

Mass Storage Group:

Nancy Meyer (Group Lead), Matthew Andrews,

Harvard Holmes, Damian Hazen, Wayne Hurlbert,

Nancy Johnston, Shreyas Cholia

Outline

- Site configuration
- Statistics
- Special features unique to the NERSC storage system
- Problems encountered
- Future plans
- Wish list

Site Configuration

NERSC production storage system consists of:

- **2 production HPSS systems (Regent and Archive)**
- **NERSC DCE Cell**
- **AFS**
- **HPSS Test Systems (formerly PROBE)**

Site Configuration

Hardware:

- **Nodes – 6M1, Nighthawk II, Winterhawk I, Winterhawk II**
- **STK Silo – Mix of 9840 (SCSI) and 9940B (Fibre) drives**
- **Disk Hardware – LSI E4400, DDN SA8000, IBM SSA drives**
- **DCE Cell – Sun E250**

Site Configuration

Software:

- **AIX 4.3.3 on core server nodes, AIX 5.1 on GridFTP/Mover node**
- **HPSS 4.3 patch level 3**
- **DCE 3.1**
- **Encina 4.3.0.5**
- **Solaris 2.8 on DCE servers**
- **Globus 2.2.4**

Site Configuration

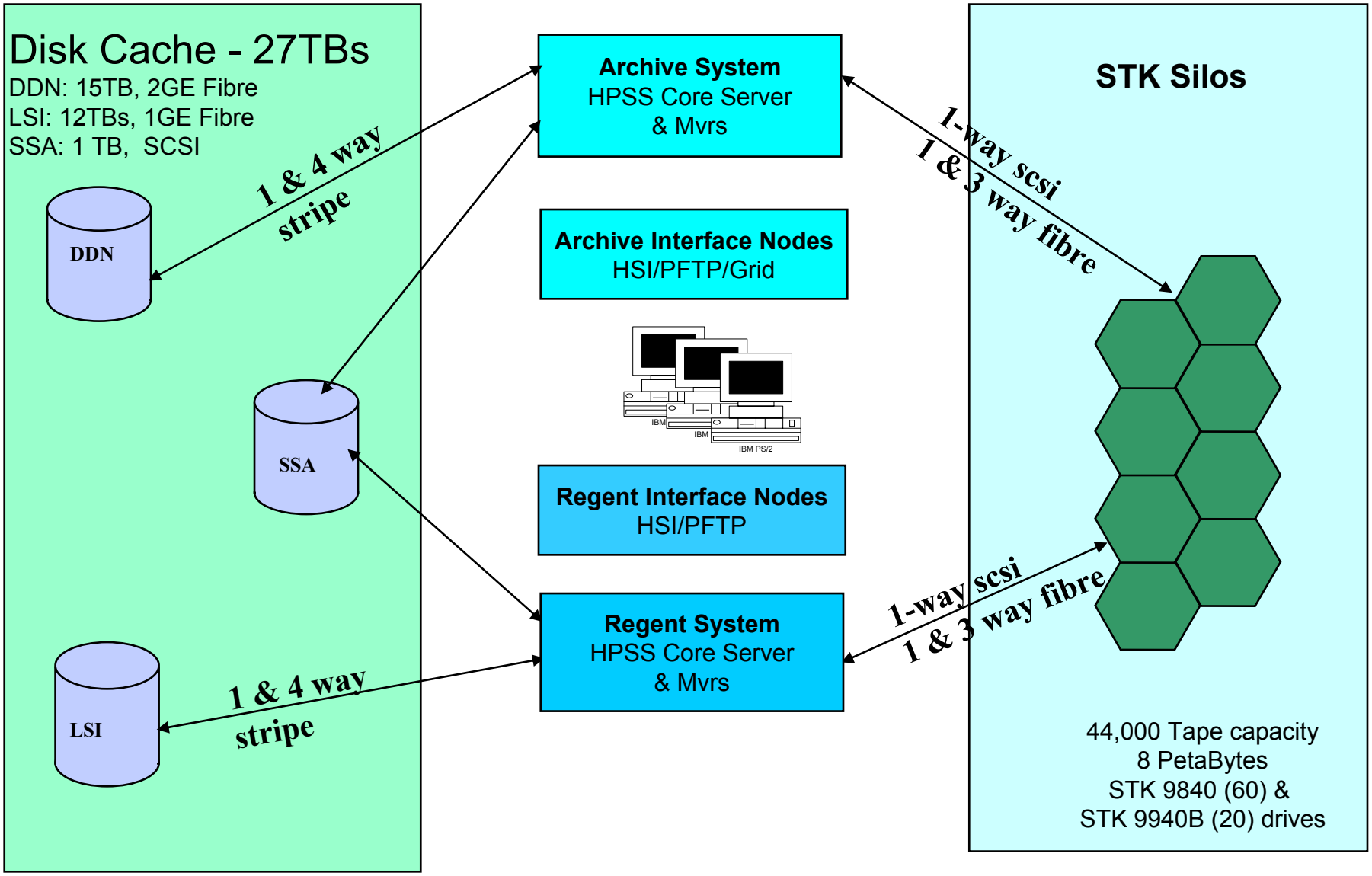
Interfaces to Storage:

- **PFTP**
- **HSI 2.6.2**
- **Grid enabled PFTP**

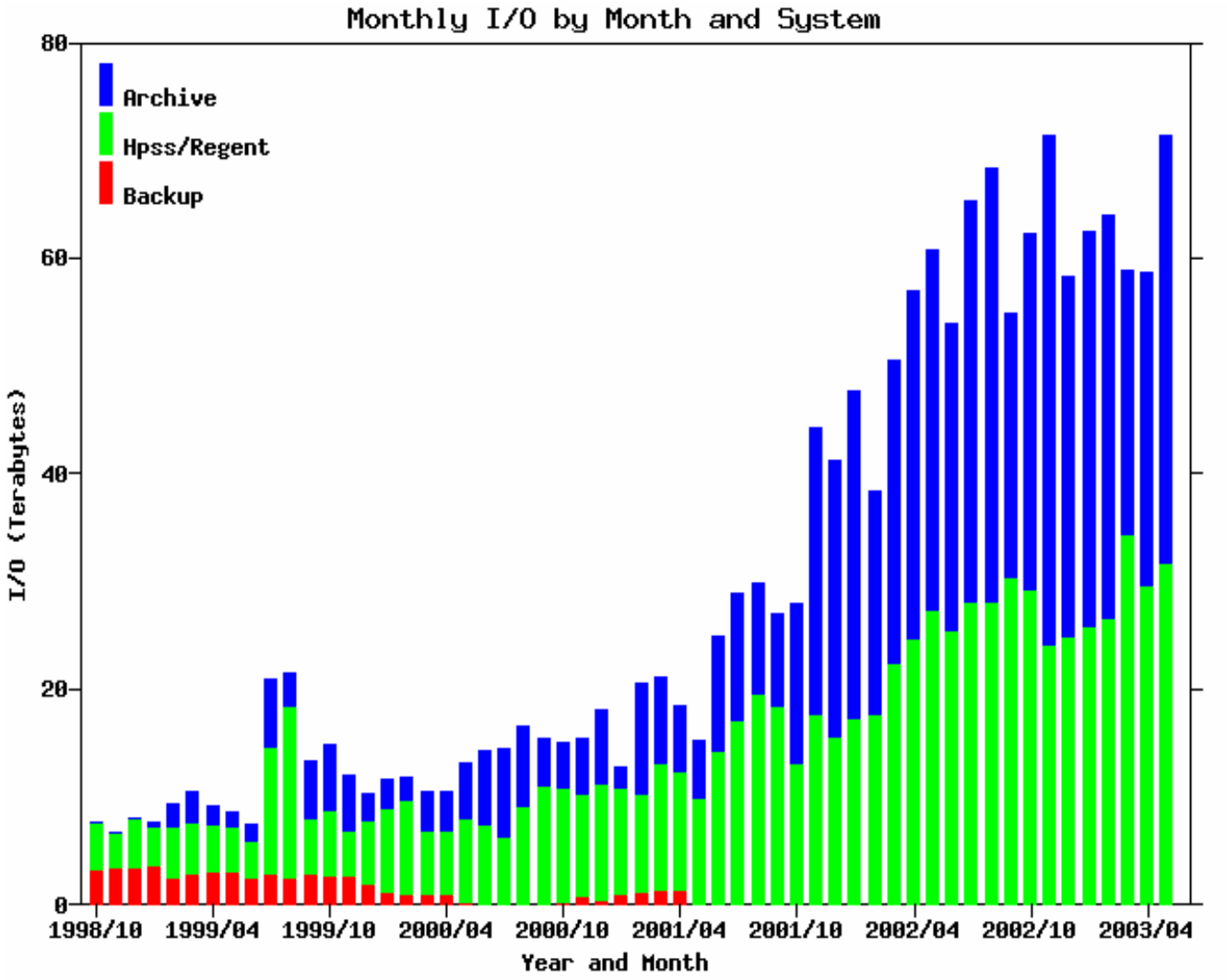
Authentication schemes supported:

- **Kerberos**
- **Encrypted username/password access**
- **GSI (Grid)**

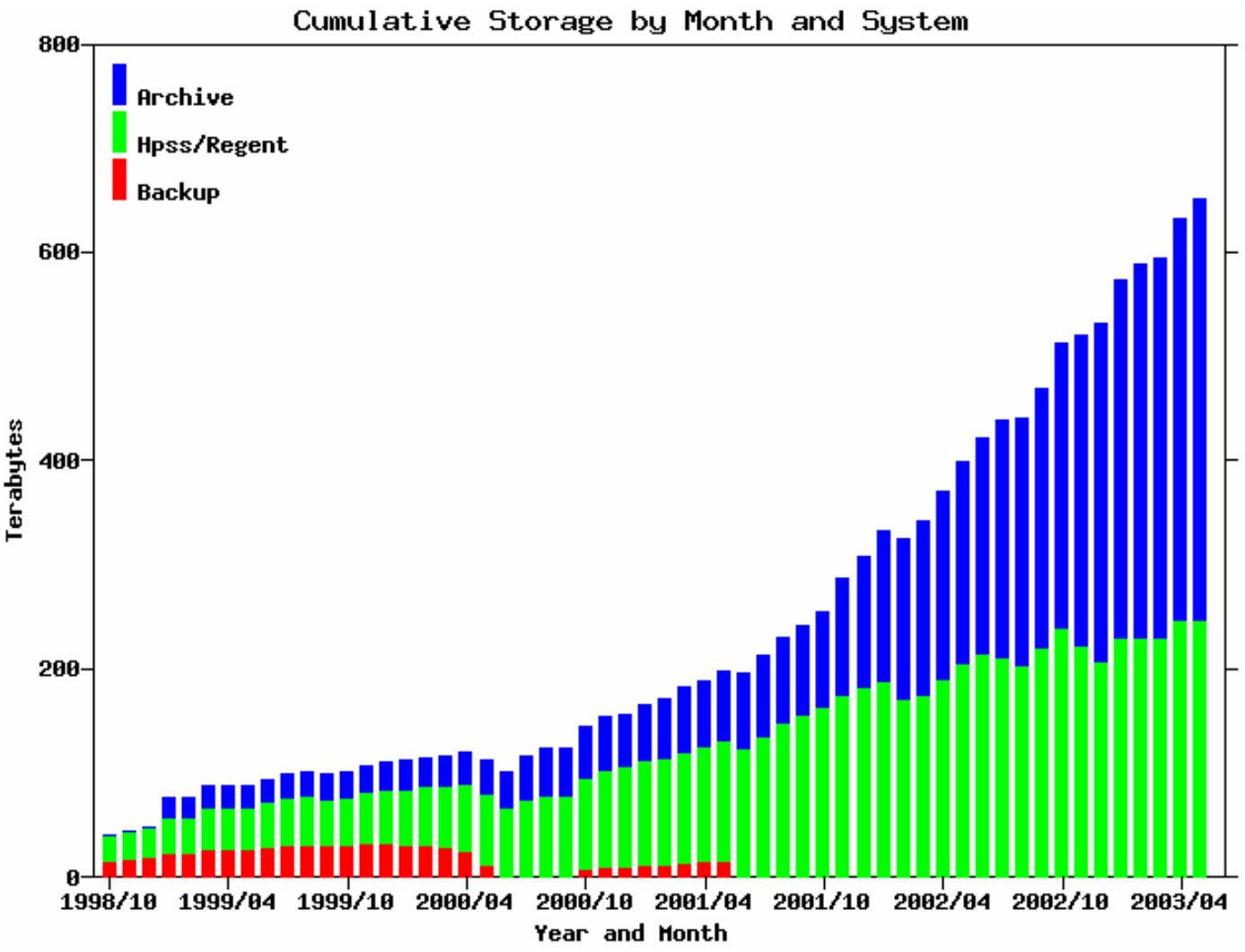
Site Configuration



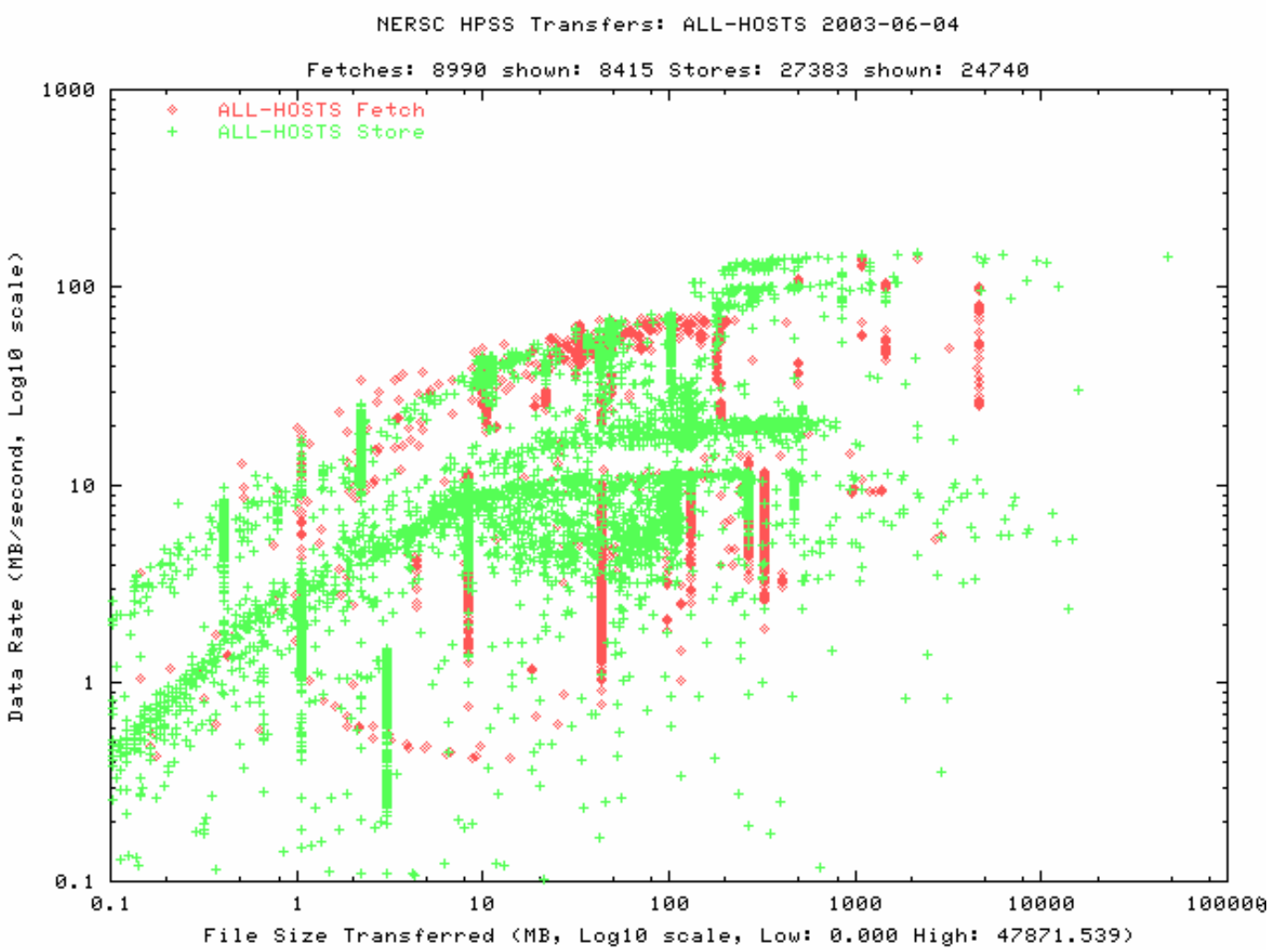
Statistics


Monthly I/O by Month and System
Archive
Hpss/Regent
Backup
I/O (Terabytes)
0
20
40
60
80
1998/10
1999/04
1999/10
2000/04
2000/10
2001/04
2001/10
2002/04
2002/10
2003/04
Year and Month

Statistics


Cumulative Storage by Month and System
Archive
Hpss/Regent
Backup
Terabytes
800
600
400
200
0
1998/10
1999/04
1999/10
2000/04
2000/10
2001/04
2001/10
2002/04
2002/10
2003/04
Year and Month

Statistics

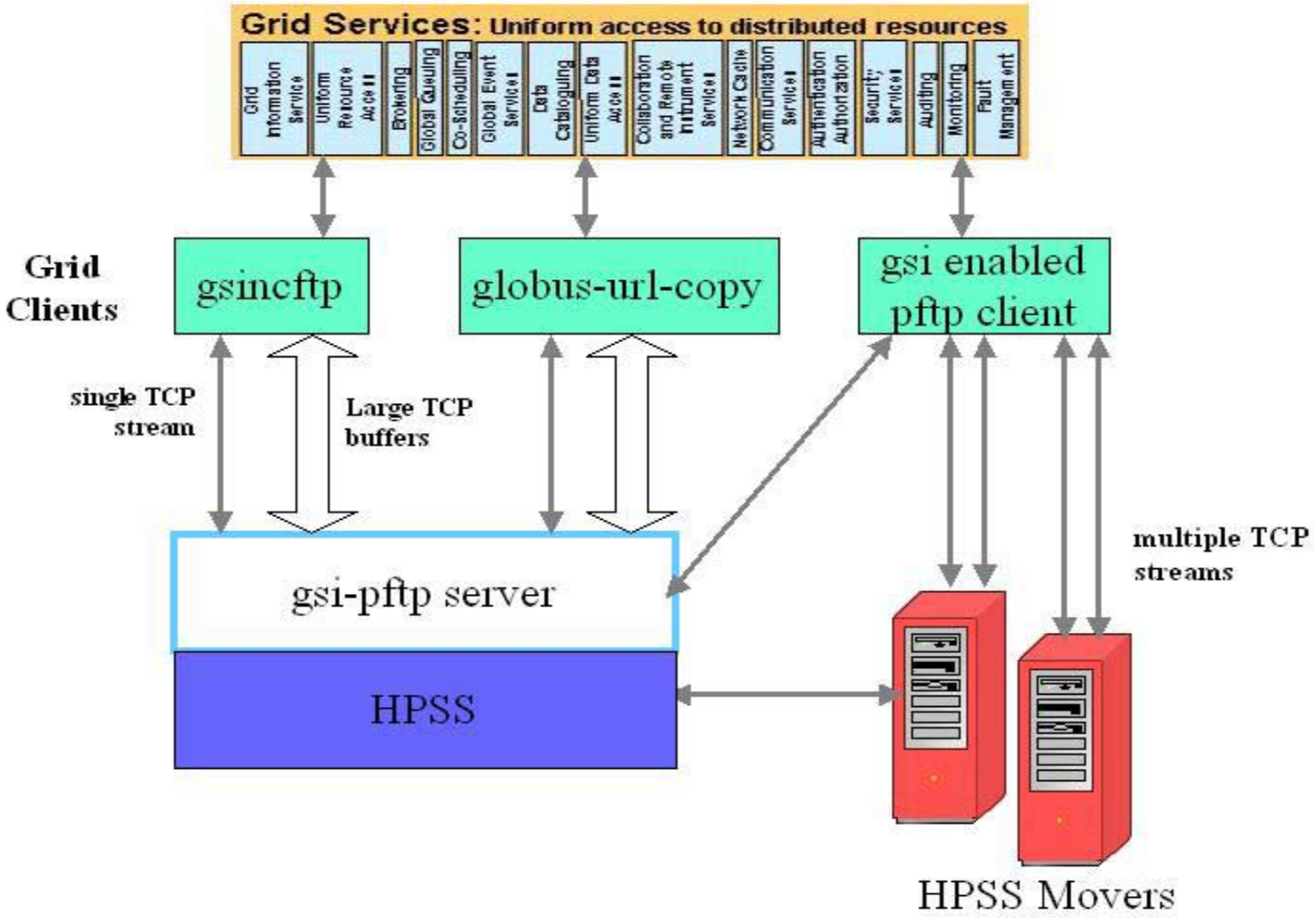

NERSC HPSS Transfers: ALL-HOSTS 2003-06-04
Fetches: 8990 shown: 8415 Stores: 27383 shown: 24740
ALL-HOSTS Fetch
ALL-HOSTS Store
Data Rate (MB/second, Log10 scale)
1000
100
10
1
0.1
0.1
1
10
100
1000
10000
100000
File Size Transferred (MB, Log10 scale, Low: 0.000 High: 47871.539)

Unique NERSC Features

GridFTP:

- **Grid Enabled HPSS PFTP Server.**
- **Allows users to authenticate to HPSS with Grid (GSI) credentials.**
- **Talks to all regular GridFTP clients (globus-url-copy, gsincftp, Java COG kit).**
- **Can perform 3rd Party Copy between 2 GridFTP servers.**
- **Grid Enabled PFTP client can perform Grid authentication to Grid enabled PFTP server.**
- **This client can be used for parallel transfers to/from HPSS Movers.**
- **Now in Production on Archive.**

Unique NERSC Features



Grid access to NERSC

Unique NERSC Features

Gatekeeper:

- **Used to enforce quotas.**
- **HPSS Accounting and NIM (NERSC Information Management) System are used generate a restricted user list, based on the user's allocation.**
- **Before a file open or create, the gatekeeper checks if the user is on restricted user list. If so, then the user is not allowed write access to the file.**
- **Read access is permitted for all users.**

Unique NERSC Features

lstape:

- **tool to generate list of all files on HPSS tapes.**
- **Stores information about files, tapes and tape position in a flat file database.**
- **Useful for disaster recovery scenarios.**
- **Also helps us locate files easily.**

Unique NERSC Features

FTP authentication:

- User performs an ssh into the authentication server.
- Generates an encrypted username/password combo by performing using his DCE username and password.
- Provides this encrypted login/password combo to the HPSS PFTP daemon to authenticate to HPSS.
- The PFTPD has local modifications to decrypt this combo and performs the authentication with the decrypted combo.
- Prevents having to send cleartext passwords over a potentially insecure network.

Problems Encountered

- Disk Storage MAP problems when initially upgrading to HPSS 4.3 – Dave Fisher helped us resolve these promptly so we had minimal fallout.
- Mixing disk with different Storage Segment size in COS results in wasted disk space.
- When converting to striped SC we did not fully export volumes which resulted in disk failures (Media Block Size inconsistency).

On the whole HPSS 4.3 has been running very smoothly and we have not encountered any major problems.

Future Plans

- Upgrade to HPSS 4.5.
- Providing direction and resources to define the role of HPSS on the Grid.
- Grid web service interface to storage – prototype (Globus File Yanker) was demonstrated at SC2002.
- XFS access to HPSS.
- Linux Movers.

Wish List

- Web Portal to Storage.
- Fully functional HPSS GridFTP, including the ability to perform parallel striped transfers with other GridFTP servers.
- SAN aware HPSS movers.
- Ability to backend newer distributed filesystems (GPFS, lustre, Storage Tank etc.) to HPSS.