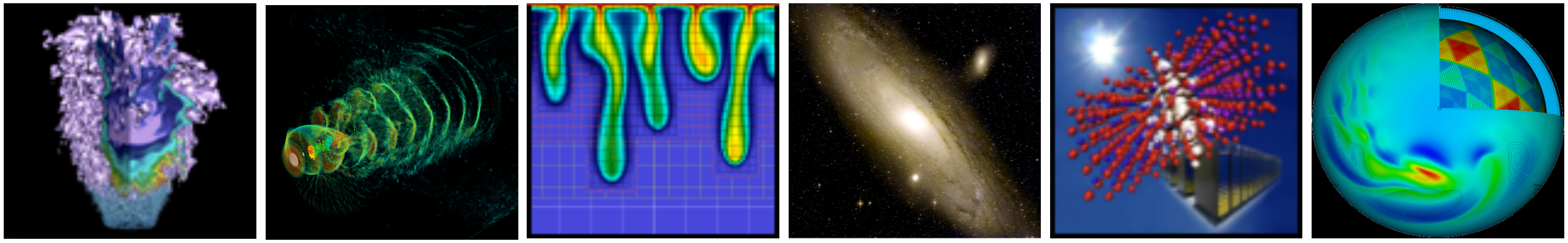


Quick Introduction to HPSS at NERSC

Nick Balthaser
NERSC Storage Systems Group
nabalthaser@lbl.gov

Joint Genome Institute, Walnut Creek, CA
Feb 10, 2011

U.S. DEPARTMENT OF ENERGY | Office of Science

NERSC National Energy Research Scientific Computing Center

BERKELEY LAB Lawrence Berkeley National Laboratory

Agenda

- **NERSC Archive Technologies Overview**
- **Use Cases for the Archive**
- **Hands-on:**
  - Authentication
  - Client Usage and Examples
- **Client Installation**

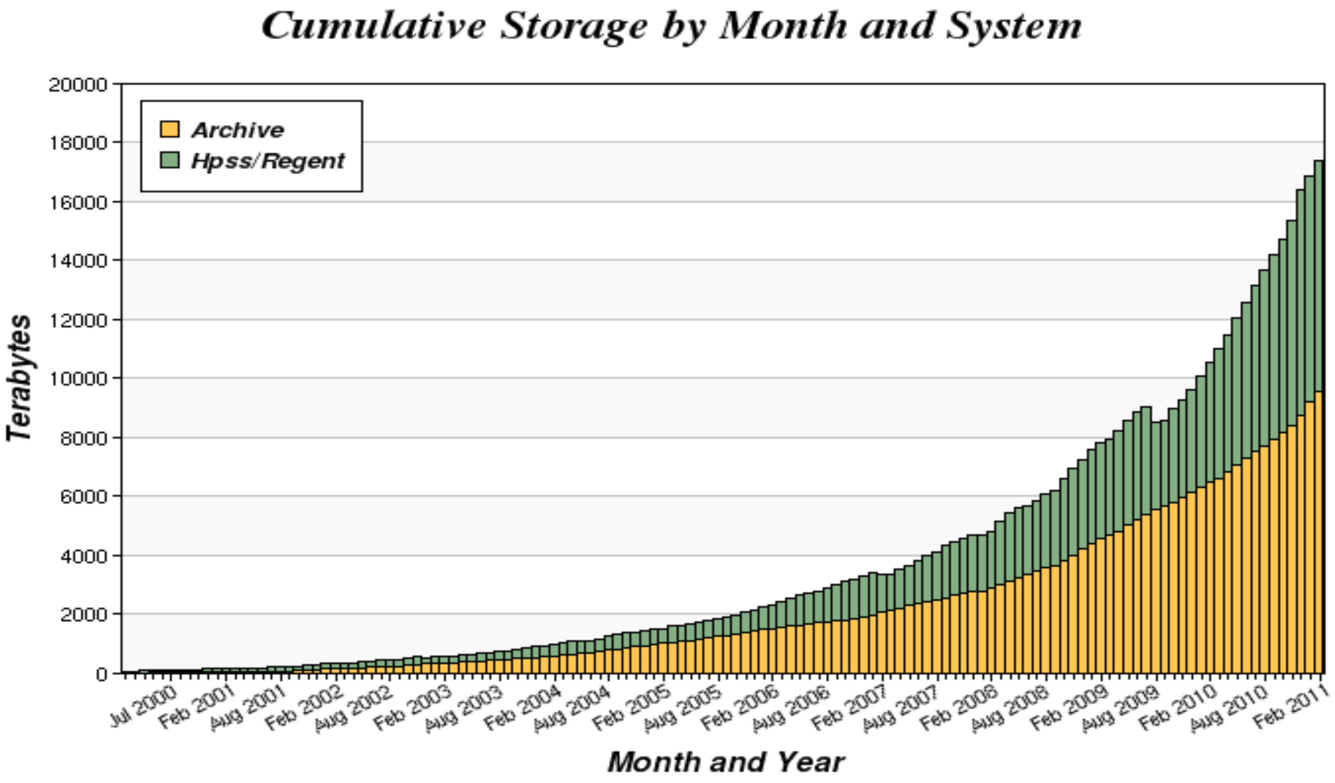
NERSC Archive Has 2 Levels, Fast Front-end Disk Cache and Enterprise Tape

- **Current data volume: 12PB in 100M files written to 26k tapes (user system)**
- **Permanent storage is magnetic tape, disk cache is transient**
  - All data written to HPSS goes through the disk cache
  - Disk to tape migration occurs every 30 minutes
  - Data retained on disk approximately one week, on average
- **Tapes and tape drives are contained in robotic libraries**
  - Cartridges are loaded/unloaded into tape drives by sophisticated library robotics
- **110 tape drives in user (archive) system**
  - 3 cartridge and drive technologies in use: Oracle T10KB/T10KC (1TB/5TB, high capacity) and 9840D (fast access, 80GB)

Front-ending the Tape Subsystem is 240TB Fast-access disk

- **Disk cache hardware: Data Direct Networks 9550 FC and 9900 SAS disk arrays**
- **User system has 13 server nodes, IBM p4/p5/p7 running AIX**
  - 12 IO nodes called data movers: read/write to network, disk and tape devices
  - 1 core server: coordinates system activity and serves metadata
- **HPSS storage application is under active development by IBM, LBNL, LLNL, LANL, SNL, and ORNL.**
  - NERSC has 2 full-time HPSS developers on staff
  - New features, stability improvements, and bug fixes are continually being developed.

Approximately 50% data growth per year



- **NERSC has 4 dedicated DTN nodes for high-speed transfers**
  - Transfer rates over 1GB/sec are possible

It is Important to Know How to Store and Retrieve Data Efficiently

- **HPSS clients can emulate file system qualities**
  - FTP-like interfaces can be deceiving: the archive is backed by tape, robotics, and a single DB2 database instance for metadata
  - Operations that would be slow on a file system, e.g. lots of random IO, can be impractical on the archive

- **HPSS does not stop users from making mistakes**
  - It is possible to store data in such a way as to make it difficult to retrieve
    - Tape storage systems do not work well with small files
  - The archive has no batch system. Inefficient use affects others.

Use Cases for the Archive

- Typical use case: long-term storage and retrieval of very large raw data sets
  - Good for incremental processing
- Long-term storage of result data
- Data migration between compute platforms
- Backups (/project and system/server backups)

Authentication is easy

- **NERSC storage uses a token-based authentication method**
  - User places encrypted authentication token in ~/.netrc file at the top level of the home directory on the compute platform
  - Token information is verified in the NERSC LDAP user database
  - All NERSC HPSS clients can use the same token
  - Tokens are username and IP specific—must generate a different token for use offsite

Authentication Hands-on

- **Authentication tokens can be generated in 2 ways:**
  - Automatic – NERSC auth service (recommended):
    - Log into any NERSC compute platform
    - Type “hsi”
    - Enter NERSC password
  - Manual – https://nim.nersc.gov/ website
    - Under “Actions” dropdown, select “Generate HPSS Token”
    - Copy/paste content into ~/.netrc
    - chmod 600 ~/.netrc
- **Use NIM website to generate token for alternate IP address**

~/.netrc example

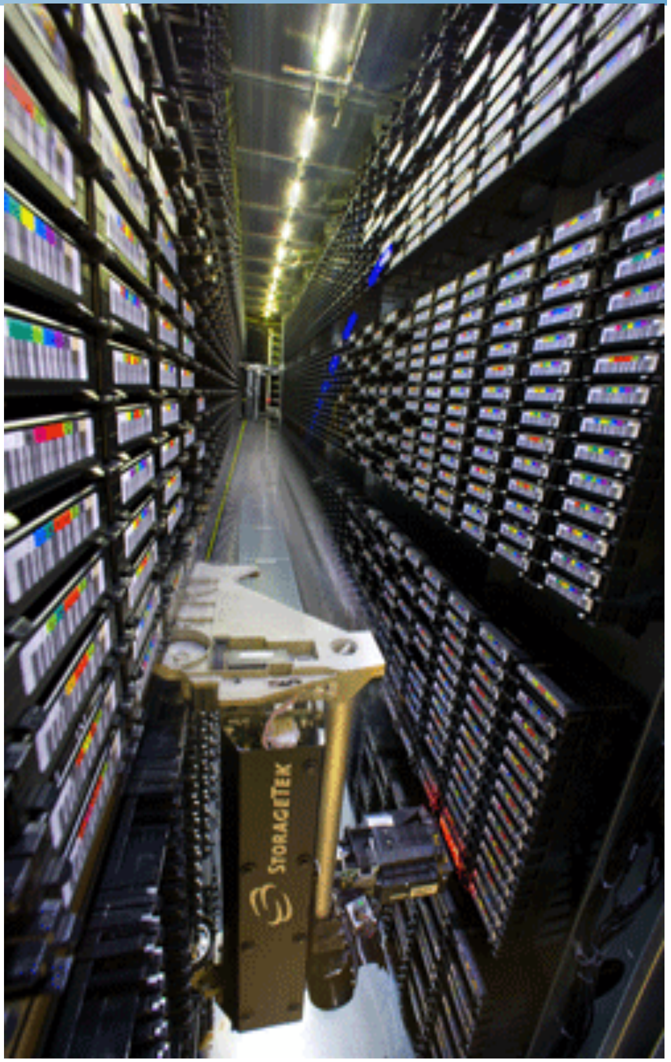
```
machine archive.nersc.gov
login joeuser
password 02UPMUezYJ/Urc7ypflk7M8KHLITsoGN6ZIcfOBdBZBxn+BViShg==


machine ftp.nersc.gov
login anonymous
password joeuser@nersc.gov
```

- **Check permissions on this file**
  - Should be rw for user only

HPSS Client Overview

- **Parallel, threaded, high performance:**
  - HSI
    - Unix shell-like interface
  - HTAR
    - Like Unix tar, for aggregation of small files
  - PFTP
    - Parallel FTP
- **Non-parallel:**
  - FTP
    - Ubiquitous, many free scripting utilities and APIs
- **GridFTP interface (garchive)**
  - Connect to other grid-enabled storage systems



Hands-on Examples: HSI

- **Most flexibility, many features and options**
- **Can cause problems if not used correctly (supports recursive transfers of small files/ directories)**
- **Features:**
  - Parallel, high speed transfers
  - Interactive and non-interactive modes
  - Common shell commands: chown, chmod, ls, rm, etc.
  - Recursion
  - Command-line editing and history
  - Wildcards
- **Connecting to the archive: type “hsi”**

```
bash-4.0$ hsi
[Authenticating]
A:/home/j/joeuser->
```

Interactive HSI

- **Transfer**

```
A:/home/j/joeuser-> put myfile
put 'myfile' : '/home/j/joeuser/myfile' ( 2097152 bytes, 31445.8 KBS (cos=4))
```

- **Retrieve**

```
A:/home/j/joeuser-> get myfile
get 'myfile' : '/home/j/joeuser/myfile' (2010/12/19 10:26:49 2097152 bytes,
 46436.2 KBS )
```

- **Full pathname or rename**

```
A:/home/j/joeuser-> put local_file : hpss_file
A:/home/j/joeuser-> get local_file : hpss_file
```

- **Wildcards**

```
A:/home/j/joeuser-> prompt
prompting turned off
A:/home/j/joeuser-> mput .bash*
```

Non-interactive HSI

- **One-line mode**

```
bash-4.0$ hsi "mkdir mydir; cd mydir; put myfile; ls –l"
```

- **Command File**

```
bash-4.0$ cat mycommands.txt
put myfile
ls -l
quit
bash-4.0$ hsi "in mycommands.txt"
```

- **Here Document**

```
bash-4.0$ hsi <<EOF
put myfile
ls -l
quit
EOF
```

- **Standard Input**

```
bash-4.0$ echo 'mkdir mydir; cd mydir; put myfile; ls -l; quit' | hsi
```

Hands-on Examples: HTAR

- **Similar to Unix tar**
- **Parallel, high speed transfers, like HSI**
- **Recommended utility for archiving small files**
  - Faster/safer than running Unix tar via pipeline
  - Creates index for fast file retrieval
- **HTAR traverses subdirectories to create tar-compatible aggregate file in HPSS**
- **No staging space required**
- **Limitations:**
  - Aggregate file can be any size, recommend 500GB max
  - Aggregates limited to 5M member files
  - Individual HTAR member files max size 64GB
  - 155/100 character prefix/filename limitation

HTAR, Continued

- **Create archive**

```
bash-4.0$ htar –cvf /home/n/nickb/mytarfile.tar ./mydir
HTAR: a   ./mydir/
HTAR: a   ./mydir/foofile
HTAR: a   /scratch/scratchdirs/nickb/HTAR_CF_CHK_50212_1297706778
HTAR Create complete for /home/n/nickb/mytarfile.tar. 2,621,442,560 bytes
    written for 1 member files, max threads: 3 Transfer time: 11.885 seconds
    (220.566 MB/s)
```

- **List archive**

```
bash-4.0$ htar –tvf /home/n/nickb/mytarfile.tar
```

- **Extract member file(s)**

```
bash-4.0$ htar –xvf /home/n/nickb/mytarfile.tar ./mydir/foofile
```

PFTP and FTP

- **PFTP**
  - Standard FTP-like interface distributed with HPSS
  - Implements parallel transfers for performance
  - FTP-compatible syntax
  - Scriptable with some effort (Here doc or command file)
  - NERSC compute platforms only

```
bash-4.0$ pftp –i < cmds.txt
```

- **FTP**
  - Available everywhere, but non-parallel, low performance
  - Free utilities such as ncftp, curl, and Perl Net::FTP add flexibility for scripting
- **Both interfaces implement *ALLO64 <filesize>* for writing files to the correct COS**

GridFTP

- **GridFTP uses a certificate based authentication method—not ~/.netrc**
  - Users can use grid credentials to transfer data between other grid-enabled sites
- **GridFTP is the server**
  - Clients include uberftp and globus-url-copy
- **Clients often support user-tunable parameters for WAN transfer**

HPSS Client Download and Installation

- **HPSS clients are provided on NERSC systems (hopper, franklin, etc.) No download/installation necessary**
  - HSI and HTAR are now installed on JGI system phoebe
- **HSI and HTAR are licensed for binary download for NERSC users (workstations, servers, offsite platforms)**
  - Go to the NERSC software download page
    - https://www.nersc.gov/users/data-and-networking/hpss/storing-and-retrieving-data/software-downloads/
  - Select appropriate version for your hardware/OS (NERSC username/ password required)
    - Minor OS version differences ***may*** be Ok
- **FTP client is usually available on most operating systems**
  - Lower performance on high-speed networks
  - Problems with authentication on Windows7

Reporting Problems

- **NERSC Staff: Contact Storage Systems**
  - Email *ssg@nersc.gov*
  - 24x7 NERSC Operations: 510-486-6821

- **NERSC Users: Contact NERSC Consulting**
  - Toll-free 800-666-3772
  - 510-486-8611, #3
  - Email *consult@nersc.gov*.

Further Reading

- NERSC Website
  - http://www.nersc.gov/users/data-and-networking/hpss/
- NERSC Grid documentation
  - http://www.nersc.gov/users/software/grid/data-transfer/
- HSI, HTAR, PFTP man pages should be installed on NERSC compute platforms
- Gleicher Enterprises Online Documentation (HSI, HTAR)
  - http://www.mgleicher.us/GEL/
- "HSI Best Practices for NERSC Users" – LBNL Report #LBNL-4745E

NERSC

National Energy Research
Scientific Computing Center