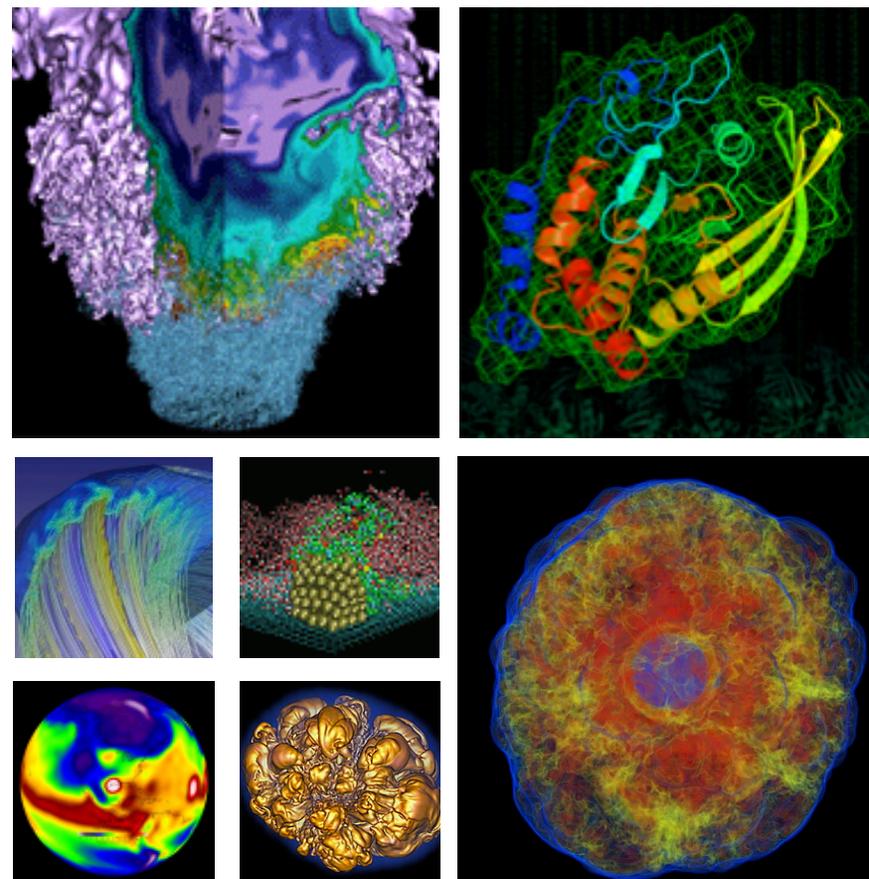


NERSC File Systems and How to Use Them

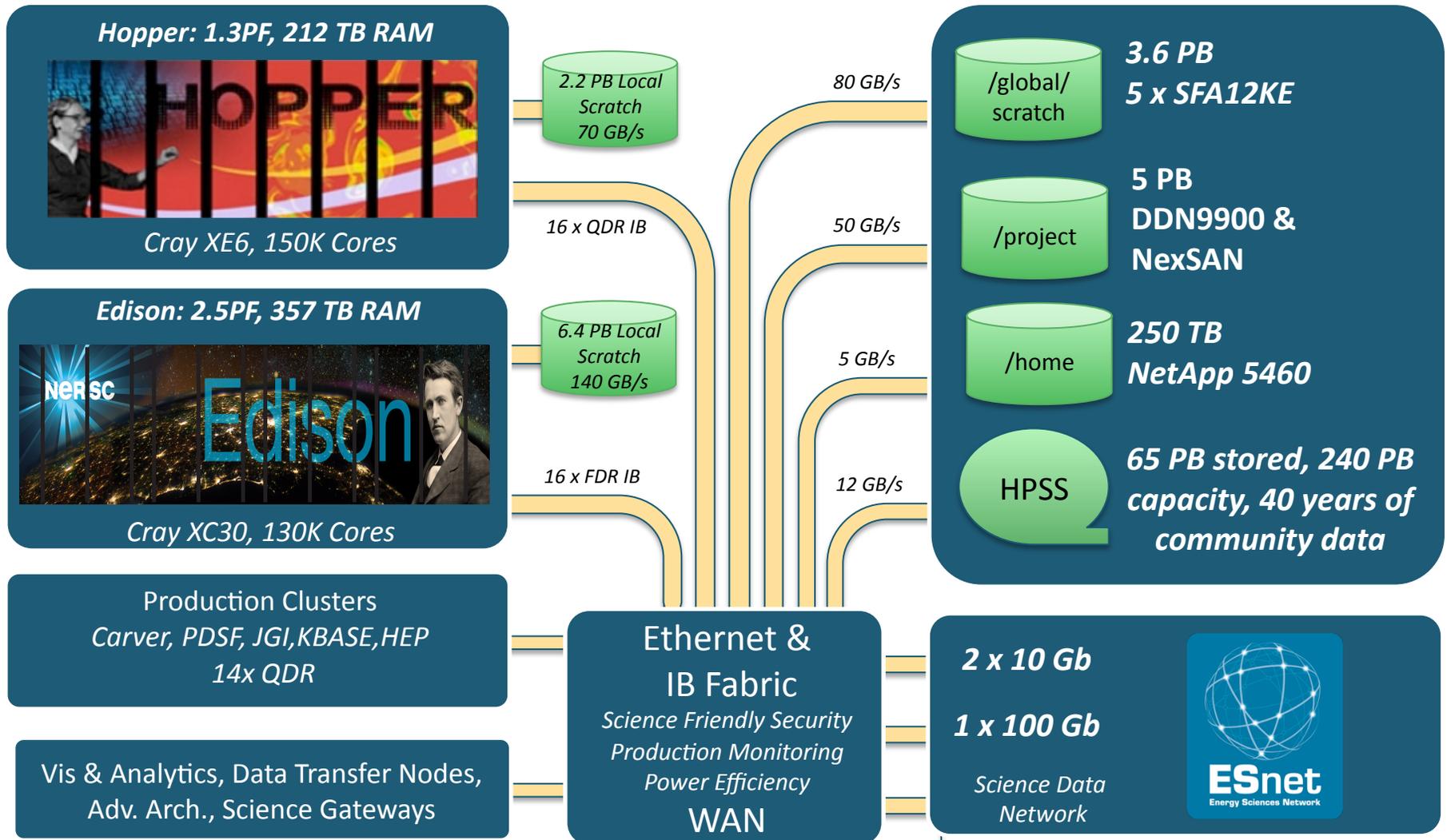


David Turner
NERSC User Services Group

Joint Facilities User Forum on Data-Intensive Computing
June 18, 2014



The compute and storage systems 2014



Overview



- **Focus on user-*writable* file systems**
- **Global file systems**
- **Local file systems**
- **Policies**
- **Performance**
- **Platform summary**
 - Edison, Hopper, Carver

Protect Your Data!



- Some file systems are backed up
- Some file systems are not backed up
- Restoration of individual files/directories may *not* be possible
- Hardware failures and human errors *will* happen

BACK UP YOUR FILES TO HPSS!

Global File Systems



- **NERSC Global Filesystem (NGF)**
- **Based on IBM's General Parallel File System (GPFS)**
- **Architected and managed by NERSC's Storage Systems Group**
- **Provides directories for home, global scratch, and project**
- **Also provides /usr/common**
 - NERSC-supported software

Global Homes File System Overview



- **Provided by two ~100 TB file systems**

`/global/u1`

`/global/u2`

– 5 GB/s aggregate bandwidth

- **Low-level name**

`/global/u1/d/dpturner`

`/global/u2/d/dpturner ->`

`/global/u1/d/dpturner`

- **Better name**

`/global/homes/d/dpturner`

- **Best name**

`$HOME`

Global Homes Use



- **Shared across all platforms**
 - `$HOME/edison`, `$HOME/hopper`, etc.
 - “dot files” (`.bashrc`, `.cshrc.ext`, etc.) might contain platform-specific clauses

```
if ($NERSC_HOST == "edison") then
...
endif
```
- **Tuned for small file access**
 - Compiling/linking
 - Job submission
 - **Do not run batch jobs in \$HOME!**

Global Homes Policies



- **Quotas enforced**
 - 40 GB
 - 1,000,000 inodes
 - Quota increases rarely (i.e., never) granted
 - Monitor with **myquota** command
- **“Permanent” storage**
 - No purging
 - Backed up
 - Hardware failures and human errors *will* happen

BACK UP YOUR FILES TO HPSS!

Global Scratch File System Overview



- **Provides 3.6 PB high-performance disk**
 - 80 GB/s aggregate bandwidth
- **Primary scratch file system for Carver**
 - Also mounted on Edison, Hopper, Datatran, etc.
- **Low-level name**
 - `/global/scratch2/sd/dpturner`
- **Better name**
 - `$GSCRATCH`
 - AKA `$SCRATCH` on Carver and Datatran

Global Scratch Use



- **Shared across many platforms**
 - `$GSCRATCH/carver`, `$GSCRATCH/edison`, etc.
- **Tuned for large streaming file access**
 - Running I/O intensive batch jobs
 - Data analysis/visualization

Global Scratch Policies



- **Quotas enforced**
 - 20 TB
 - 4,000,000 inodes
 - Quota increases may be requested
 - Monitor with `myquota` command
- **Temporary storage**
 - Bi-weekly purges of *all* files that have not been accessed in over 12 weeks
 - List of purged files in `$GSCRATCH/purged.<timestamp>`
 - Hardware failures and human errors *will* happen

BACK UP YOUR FILES TO HPSS!

Project File System Overview



- **Provides 5.1 PB high-performance disk**
 - 50 GB/s aggregate bandwidth
- **Widely available**
- **Intended for sharing data between platforms, between users, or with the outside world**
- **Prior to AY14**
 - Must be requested
 - `/project/projectdirs/bigsci`
- **Beginning AY14**
 - Every MPP repo gets project directory
 - `/project/projectdirs/m9999`

Project Use



- **Tuned for large streaming file access**
 - Running I/O intensive batch jobs
 - Data analysis/visualization
- **Access controlled by Unix file groups**
 - Group name usually same as directory
 - Requires administrator (usually the PI or PI Proxy)
 - Can also use access control list (ACL)

Project Policies



- **Quotas enforced**
 - 1 TB
 - 1,000,000 inodes
 - Quota increases may be requested
 - Monitor with `prjquota` command
 - `% prjquota bigsci`
- ***Permanent storage***
 - No purging
 - Backed up if quota \leq 5 TB
 - Hardware failures and human errors *will* happen

BACK UP YOUR FILES TO HPSS!

Science Gateways on Project



- **Make data available to outside world**

```
mkdir /project/projectdirs/bigsci/www
```

```
chmod o+x /project/projectdirs/bigsci
```

```
chmod o+rx /project/projectdirs/bigsci/www
```

- **Access with web browser**

```
http://portal.nersc.gov/project/bigsci
```

Local File Systems on Edison



- **Edison *scratch* file systems**
 - `/scratch1`
 - `/scratch2`
 - Each has 2.1 PB
 - Each has 48 GB/s aggregate bandwidth
 - `/scratch3`
 - 3.2 PB
 - 72 GB/s aggregate bandwidth
- **Provided by Cray, based on Lustre**

Edison Scratch Use



- **Each user gets a scratch directory in `/scratch1` or `/scratch2`**
 - `/scratch2/scratchdirs/dpturner`
 - `$SCRATCH`
- **Access to `/scratch3` must be requested**
 - Large datasets
 - High bandwidth
- **Tuned for large streaming file access**
 - Running I/O intensive batch jobs
 - Data analysis/visualization

Edison Scratch Policies



- **Quotas enforced in \$SCRATCH by submit filter**
 - 10 TB
 - 10,000,000 inodes
 - Quota increases may be requested
 - Monitor with `myquota` command
 - No quota enforcement in /scratch3
- **Temporary storage**
 - Daily purges of *all* files that have not been accessed in over 12 weeks
 - List of purged files in \$SCRATCH/purged.<timestamp>
 - Hardware failures and human errors *will* happen

BACK UP YOUR FILES TO HPSS!

Local File Systems on Hopper



- **Hopper *scratch* file systems**
 - `/scratch`
 - `/scratch2`
 - Each has 1.0 PB
 - Each has 35 GB/s aggregate bandwidth
- **Provided by Cray, based on Lustre**

Hopper Scratch Use



- Each user gets a scratch directory in ***/scratch1 and /scratch2***
 - /scratch/scratchdirs/dpturner***
 - **`$SCRATCH`**
 - /scratch2/scratchdirs/dpturner***
 - **`$SCRATCH2`**
- **Tuned for large streaming file access**
 - Running I/O intensive batch jobs
 - Data analysis/visualization

Hopper Scratch Policies



- **Quotas enforced by submit filter**
 - Combined (scratch/scratch2) quotas
 - 5 TB
 - 5,000,000 inodes
 - Quota increases may be requested
 - Monitor with `myquota` command
- **Temporary storage**
 - Daily purges of *all* files that have not been accessed in over 12 weeks
 - List of purged files in `$SCRATCH/purged.<timestamp>`
 - Hardware failures and human errors *will* happen

BACK UP YOUR FILES TO HPSS!

Long-Term File Systems



- **Global home directories**
 - Source/object/executable files, batch scripts, input files, configuration files, batch job summaries (*not* for running jobs)
 - Backed up
 - 40 GB permanent quota
 - \$HOME
- **Global project directories**
 - Sharing data between people and/or systems
 - All MPP repos have one
 - Backed up if quota less than or equal to 5 TB
 - 1 TB default quota

Short-Term File Systems



- **Local scratch directories**
 - Cray (Edison, Hopper) only
 - Large, high-performance parallel Lustre file system
 - Not backed up; files purged after 12 weeks
 - Hopper: 5 TB default quota; Edison: 10 TB default quota
 - \$SCRATCH, \$SCRATCH2
- **Global scratch directories**
 - All systems
 - Large, high-performance parallel GPFS file system
 - Not backed up; files purged after 12 weeks
 - 20 TB default quota
 - \$GSCRATCH

File System Suggestions



- **DO NOT RUN BATCH JOBS IN \$HOME**
 - Use \$SCRATCH for running Edison/Hopper batch
 - Use \$GSCRATCH for running Carver batch
- **Performance can be limited by metadata**
 - Do not store 1000s of files in single directory
- **Use “tar” to conserve inodes**
- **Use HPSS to archive important data**
 - Protection against hardware failure
 - Quota management
- **DO NOT USE /tmp!**

File Systems Summary



File System	Path	Type	Default Quota	Backups	Purge Policy
Global Homes	\$HOME	GPFS	40 GB / 1M inodes	Yes	Not purged
Global Scratch	\$GSCRATCH	GPFS	20 TB / 4M inodes	No	12 weeks from last access
Global Project	/project/ projectdirs/ projectname	GPFS	1 TB / 1M inodes	Yes, if quota less than or equal to 5TB	Not purged
Hopper Scratch	\$SCRATCH and \$SCRATCH2	Lustre	5 TB / 5M inodes (combined)	No	12 weeks from last access
Edison Scratch	\$SCRATCH	Lustre	10 TB / 5M inodes (none in /scratch3)	No	12 weeks from last access

Resources



<http://www.nersc.gov/users/data-and-file-systems/>

[http://www.nersc.gov/users/data-and-file-systems/
file-systems/](http://www.nersc.gov/users/data-and-file-systems/file-systems/)

[http://www.nersc.gov/users/computational-systems/
edison/file-storage-and-i-o/](http://www.nersc.gov/users/computational-systems/edison/file-storage-and-i-o/)

[http://www.nersc.gov/users/computational-systems/
hopper/file-storage-and-i-o/](http://www.nersc.gov/users/computational-systems/hopper/file-storage-and-i-o/)



Thank you.