Data and application management strategies of SBGrid

Ian Stokes-Rees

OSG All Hands Meeting, March 2011

Design Objectives

A.

Web-accessible environment for collaborative, compute and data intensive science

B.

Extensible infrastructure to facilitate development and deployment of novel computational workflows

Data Tiers - Scoping

- **VO-wide**: all sites, admin managed, very stable
- **User project**: all sites, user managed, 1-10 weeks, 1-3 GB
- **User static**: all sites, user managed, indefinite, 10 MB
- **Job set**: all sites, infrastructure managed, 1-10 days, 0.1-1 GB
- **Job**: direct to worker node, infrastructure managed, 1 day, <10 MB
- **Job indirect**: to worker node via UCSD, infrastructure managed, 1 day, <10 GB

Data Tiers - Locality

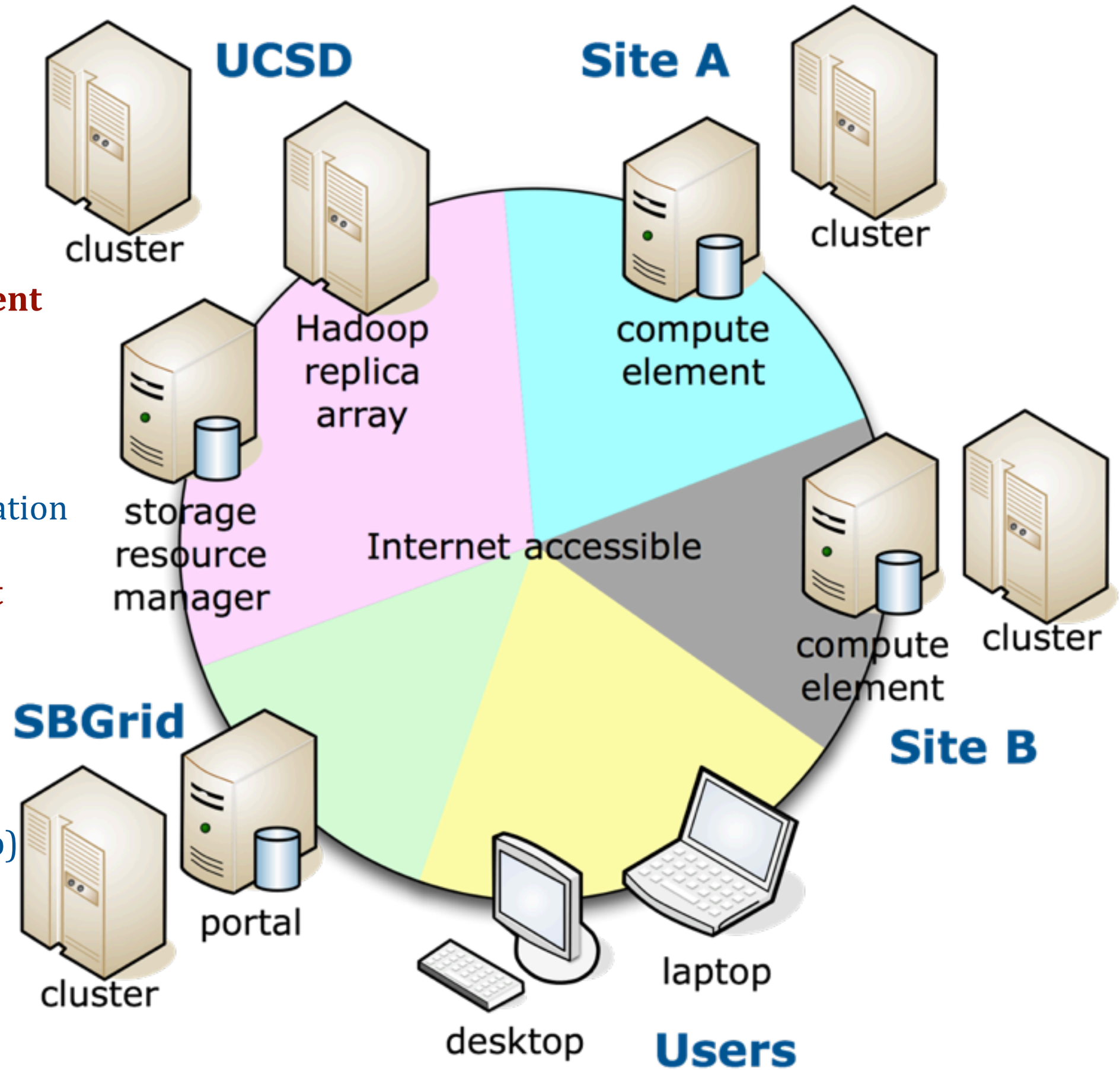
- **UCSD**
  - Bulk storage: > 20 GB available to us
  - lcg-cp access: push files from our gateway node, jobs pull to WN
  - HadoopFS: Replicated, load-balanced access
- **Per-site $OSG_{DATA,APP}**
  - Sync from master at our gateway node (rsync or UCSD indirect tarball)
  - User file space in **`$OSG_DATA/sbgrid/user/`*`UNAME`*`/workflow`**
- **Per-job pre-stage**
  - Condor file movement of small scripts, STDOUT, STDERR, logging, results
- **Per-job run-time fetch**
  - curl/wget to fetch by HTTP(S) (via Squid/http_proxy, if available)
  - lcg-cp from UCSD if possible (preferred)

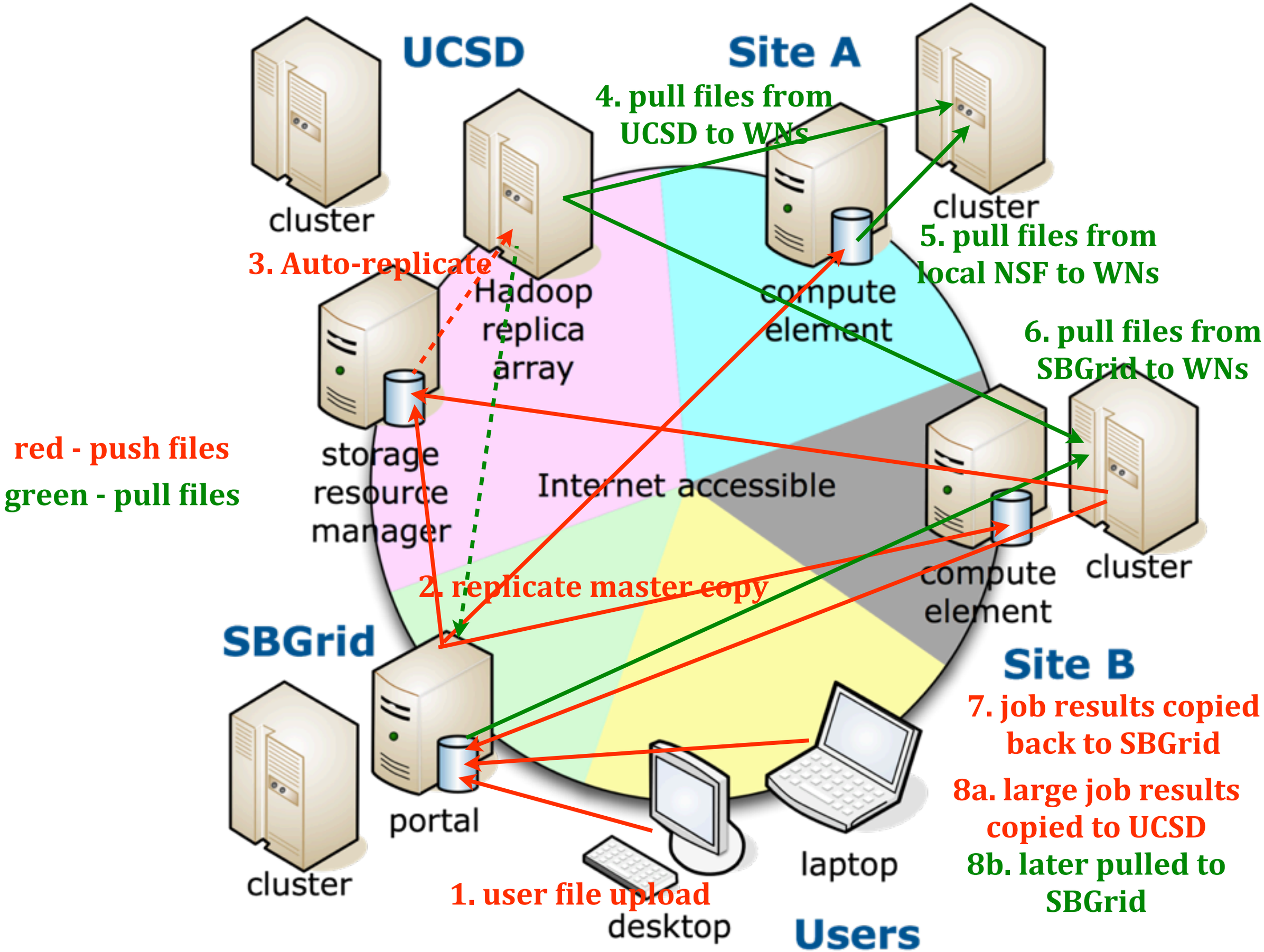
Typical SBGrid Workflow

- A single invocation may involve 10-100 GB of data input, generated data, and retained data across millions of files
- This creates a significant burden on file systems and network
- As usual, obviously essential to minimize data movement

Data Management
quota
du scan
tmpwatch
conventions
workflow integration

Data Movement
scp (users)
rsync (VO-wide)
grid-ftp (UCSD)
curl (WNs)
cp (NFS)
htcp (secure web)




UCSD
Site A
cluster
4. pull files from UCSD to WNs
cluster
3. Auto-replicate
Hadoop replica array
compute element
5. pull files from local NSF to WNs
6. pull files from SBGrid to WNs
red - push files
green - pull files
storage resource manager
Internet accessible
compute element
cluster
2. replicate master copy
SBGrid
Site B
7. job results copied back to SBGrid
portal
8a. large job results copied to UCSD
8b. later pulled to SBGrid
cluster
laptop
1. user file upload
desktop
Users

Lessons

- rsync is an effective way to replicate and maintain a master image at OSG sites
- lcg-cp is generally fast for big files
  - We often see 10 MB/s which equals 1 GB in < 2 minutes
- HTTP served content needs to be managed carefully
  - risk DDOS on web server
  - http_proxy (Squid) potentially helps a lot
- Heavy I/O is an easy way to bring down your OSG infrastructure

Issues and Future

- Users have big data sets on “non-grid” systems
- Privacy (and then ACLs) for grid data is limited
- Maintenance is manually intensive
- GASS cache should at least check timestamps
- Upcoming: remote compilation and configuration of applications
- Watching Globus Online closely (TPFT, web i/f)
- Some experimentation with DropBox user file syncing (with judicious use of “on” and “off”)

Acknowledgements

- Peter Doherty
  - SBGrid OSG Sysadmin
- Terrence Martin
  - UCSD, HadoopFS support and encouragement
- Brian Bockleman
  - University of Nebraska, suggestions
- Steve Timm
  - FNAL, patience for repeatedly overwhelming NFS or web traffic

Q&A

- What are the main pain points for you?
- What is upcoming in OSG?
- Are there obvious tools or strategies that should be more widely adopted?
- What do other VOs do to handle this?
- How do “regular users” handle their grid software and application deployment and management?