



REDDNET & OSG

Tier-3 Analysis with
Distributed Data

Daniel Engh

OSG/CMS T3 Meeting – Nashville, TN

Aug 11, 2010



REDDNET

Logistical Networking

- ▣ Designed for Wide Area data access
- ▣ Confluence of Data and Networking
 - Think of data as communication not static storage
 - Use layered communications model (like OSI)
 - ▣ IBP protocol (like IP)
 - Simple, limited, -- scalable --
 - ▣ Higher Layers (like TCP, sessions, ...)
 - LoDN, Phoebus, PerfSonar, Posix libs
- ▣ REDDnet
 - ▣ a deployment of LN tools
 - ▣ 700+ TB hardware, fast networking

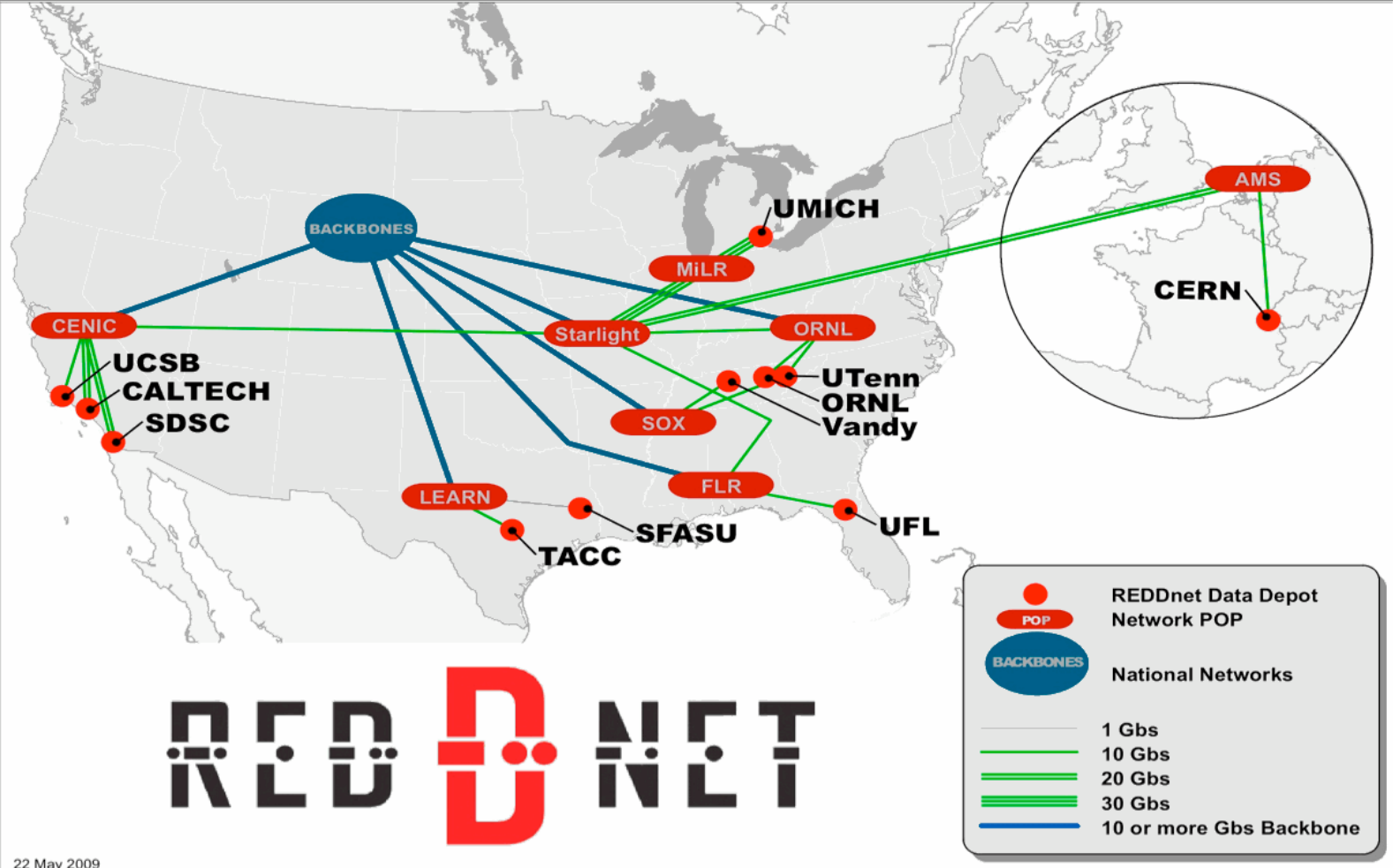
wins 2010 Internet 2 IDEA award

The Core: IBP Depots and Exnodes

- ▣ IBP Depots
 - ▣ Simple, basic, limited, distributed
 - ▣ Store data blocks (not files)
 - ▣ IBP keys – security for each block
 - ▣ Best effort (no advance reservation, etc)
 - ▣ No info on files, permissions, owners, etc.

- ▣ Exnode
 - ▣ Assemble your file (like UNIX inode)
 - ▣ For each data block:
 - URLs
 - IBP keys (read, write, manage)
 - length, offset

REDDnet Research & Education Data Depot Network



REDDNET

Mid Layer: File Services

- ▣ LoDN and L-Store
 - store exnodes
 - Add file system services
 - Directories, Owner, permissions, xtrr
 - Add data placement policies
 - How many replicas, where?
 - Dispatch the data
 - Block-level replication
 - Maintain data integrity
 - Check/repair holes/re-dispatch
 - Maintain data placement policies

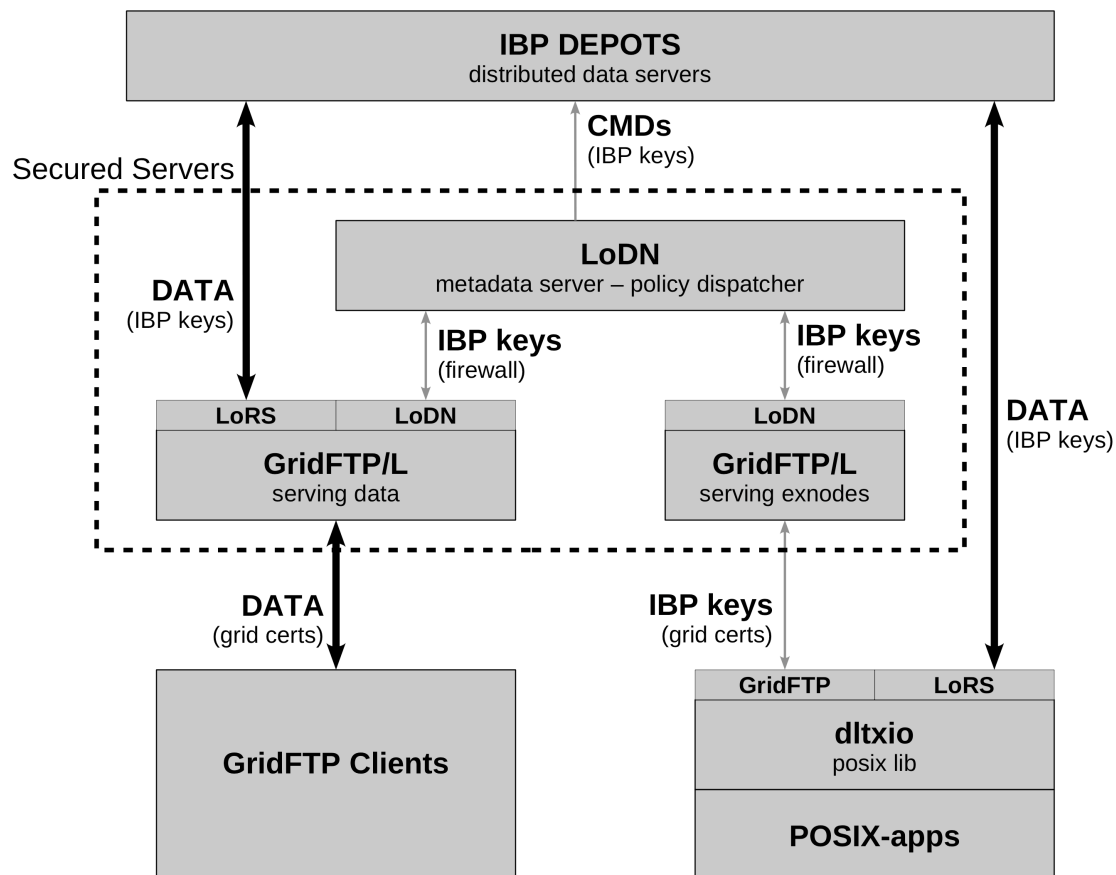
High Layer: User Interfaces

- ▣ GridFTP/L
 - ▣ Standard front-end, standard client tools, GUMS, etc
 - ▣ Backend talks to REDDnet service
 - Optimally access fastest (nearest?) data copy
 - ▣ Compatibility with:
 - SRM, Bestman, Phedex
- ▣ POSIX I/O
 - ▣ ROOT/L plugin developed
 - CMSSW compatible
 - ▣ Grid-secure, user certs GUMS, etc.
- ▣ Site setup: Use just like dCache, for example

File Services Grid Security

- ▣ Each data block secured with IBP keys
 - E.G. Need the read key to read the data
- ▣ Access to the exnode = access to data
 - GridFTP/L
 - ▣ Default Mode
 - Serves data as usual
 - IBP keys stay in GridFTP backend
 - ▣ Exnode mode
 - Serves IBP Keys
 - Small, fast transfer
 - Authenticated, encrypted transfer
 - Used by POSIX lib

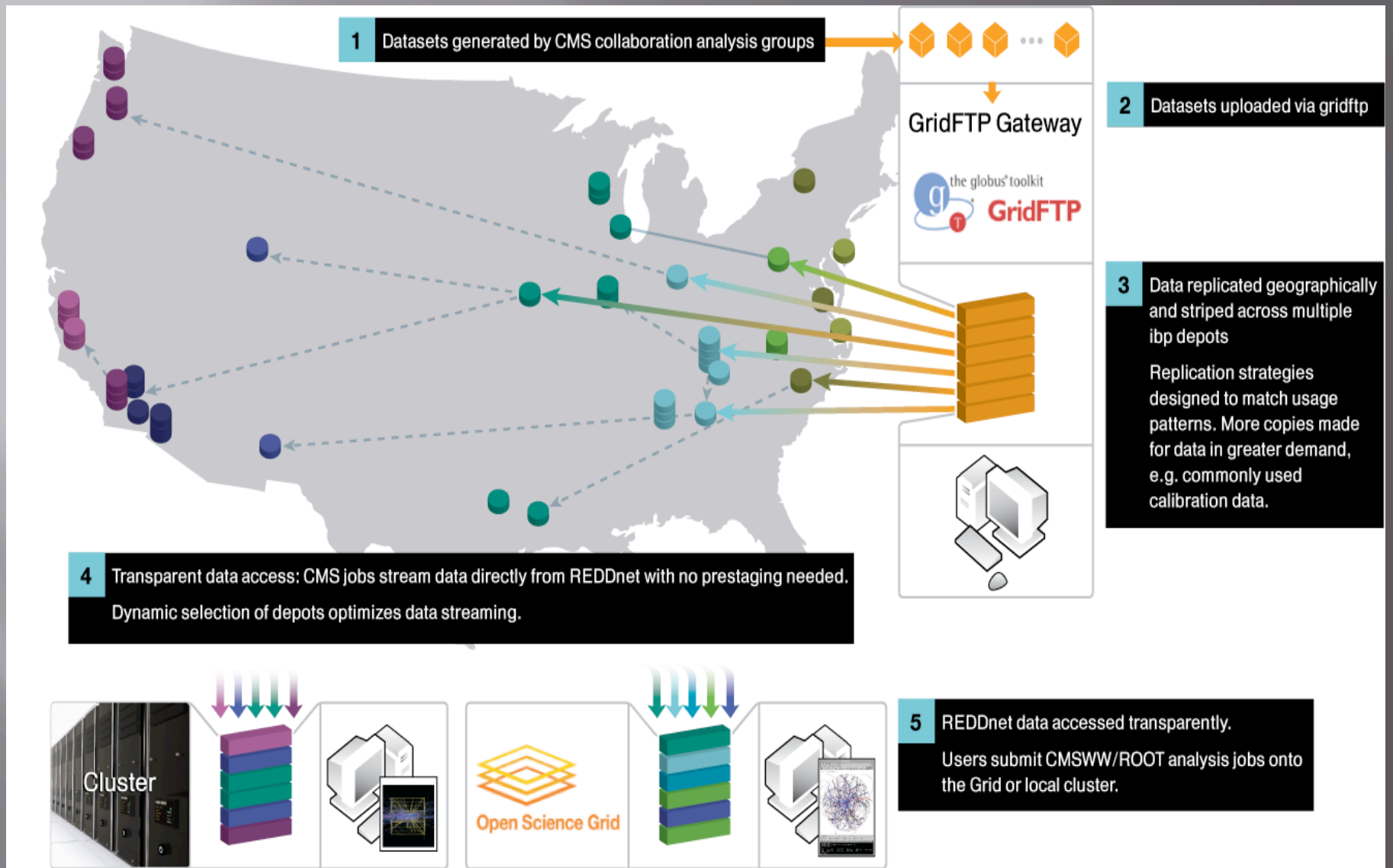
REDDnet Grid Security



Payload
(security mechanism)

Block-level augmentation

- ▣ Big advantages over file replication
 - Smaller units
 - ▣ Security embedded within block-level layer (IBP keys)
 - ▣ Retries less costly
 - ▣ Placement flexibility (space left on device)
 - File semantics reside at higher layer
 - ▣ Start, end, size, name
 - File-System semantics & services avoided
 - ▣ Files remain in same directory, policy, etc.
 - ▣ Owner, group, same
 - ▣ The blocks are replicated, the file is “augmented”
 - ▣ F.S. security (eg) not used for augmentation



How to set up

Entering phase for T3 test community

- Join CMS REDDnet mailing list
- Vanderbilt initially deploys/maintains
 - IBP Depots
 - LoDN/Lstore File Services
 - distributed GridFTP&SRM servers

Toolkit for easy installation coming...
- User sets up ROOT-based analysis
 - ROOT, FwLite, CMSSW
 - Manual install 2 REDDnet libs
 - Manually Adapt procedures, scripts.
 - Library will added to CMSSW IO protocol suite

POSIX lib available for recompiling apps

How to use it

- ▣ Request LoDN policy
 - Directory
 - Sites for replication
- ▣ Upload/Download data
 - globus-url-copy, uberftp, srm-copy
 - Your standard globus tools will work
 - OSG VO's already set up.
- ▣ Stream Data
 - ROOT plugin avail for download.
 - Run data off of local or nearby depots.

Usage examples

- ▣ globus-url-copy <file:///tmp/myfile> gsiftp://se3.accre.vanderbilt.edu/mydir/myfile
- ▣ uberftp se3.accre.vanderbilt.edu "ls /mydir"
- ▣ ROOT-based analysis
 - Specify physical file name
 - TFile::Open("lors://se3.accre.vanderbilt.edu/mydir/myfile");

Where to use it

- ▣ Vanderbilt Maintains depots and Gateways
- ▣ IBP depots Currently at:
 - CERN, Vandy, UFL, Umich, Caltech, SDSC, UCSB, SFASU, TACC, ORNL, UTK
 - 10-15 more sites will be added
 - ▣ Who is interested?
 - ▣ Email me

Daniel.Engh@vanderbilt.edu

- ▣ We'll bring more info to CMS/OSG T3 regular mtgs

CMSSW data streaming

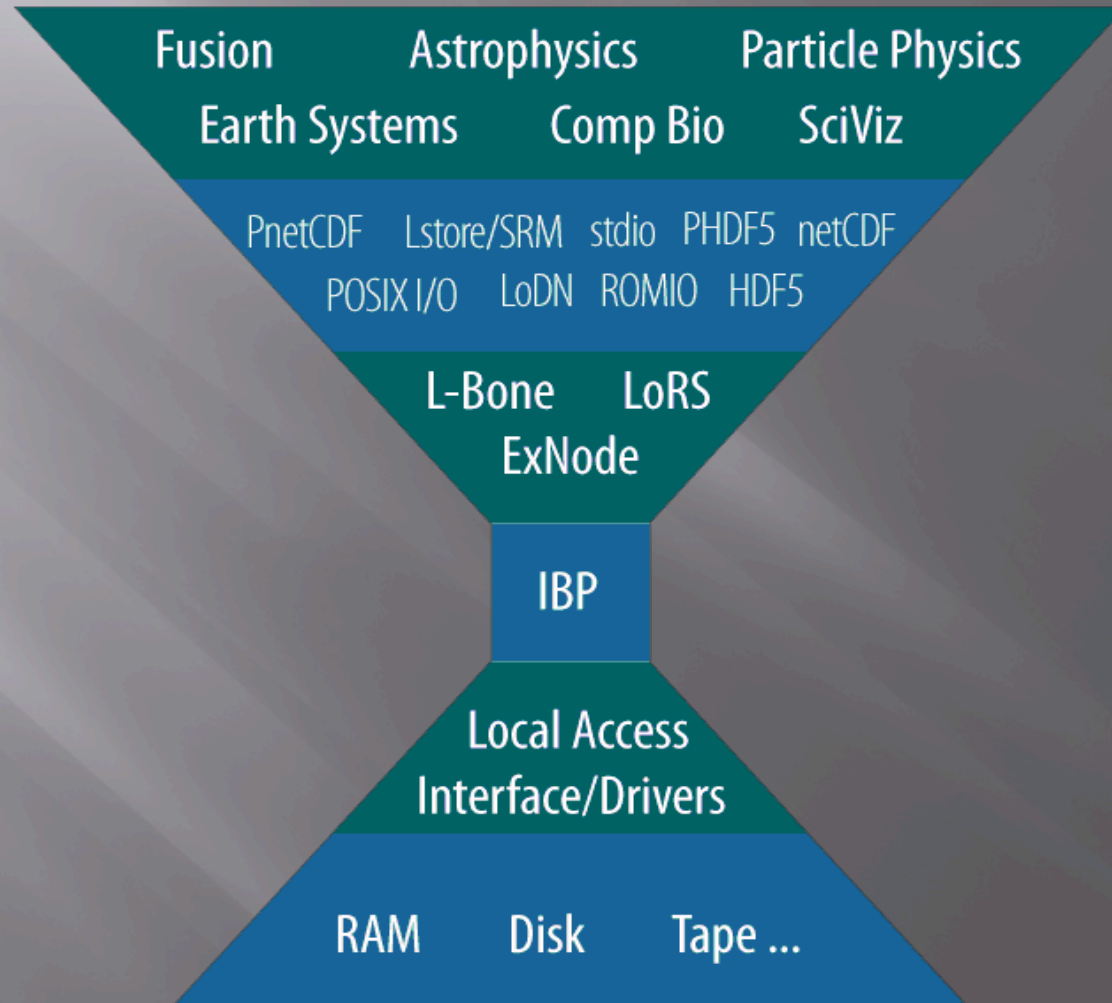
The screenshot displays a Linux desktop environment with three windows open:

- Top Window:** A terminal window titled "Linux x86_64: vpa006.phy.vanderbilt.edu:/home/anghuj/cmss/CMSSW_2_3/Store/Demo/MyTrackAnalyzer". The terminal shows the command `vpa006@larc1:~$ ssh -t joggel@lon.cern.ch`.
- Bottom-Left Window:** A window titled "LARS View - Visualization of Logical Runtime System Tools". It displays a map of North America with several colored markers representing different sites: UCSE, Caltech, SDSC, SFASU, Vandy, UTK, LBNL, and CERN. A red horizontal line is visible below the map.
- Bottom-Right Window:** A terminal window titled "Linux x86_64: vpa006.phy.vanderbilt.edu:/home/anghuj/storage/fermion". The terminal shows the command `vpa006@larc1:~$ ssh -t joggel@lon.cern.ch`.

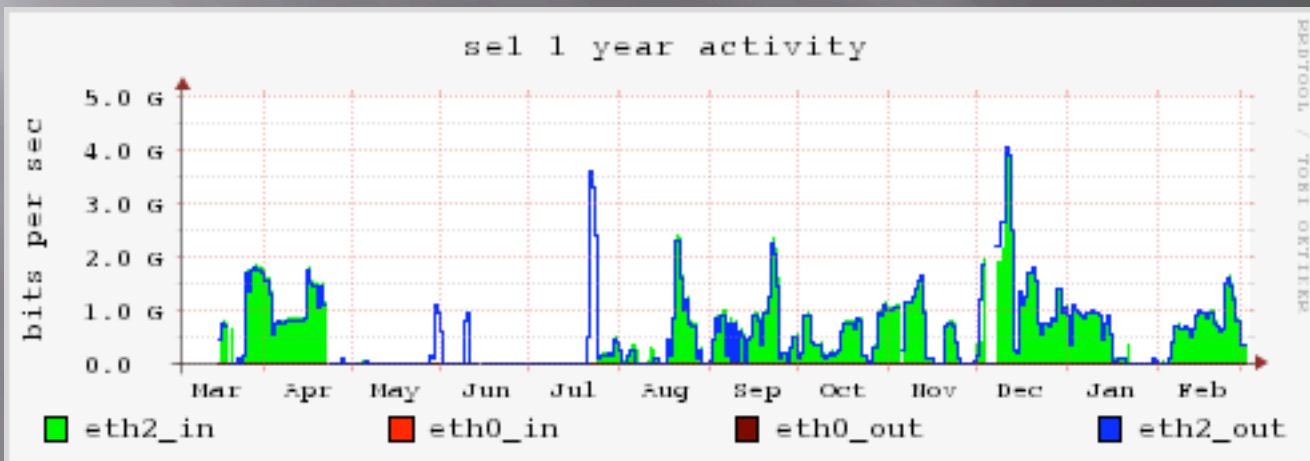
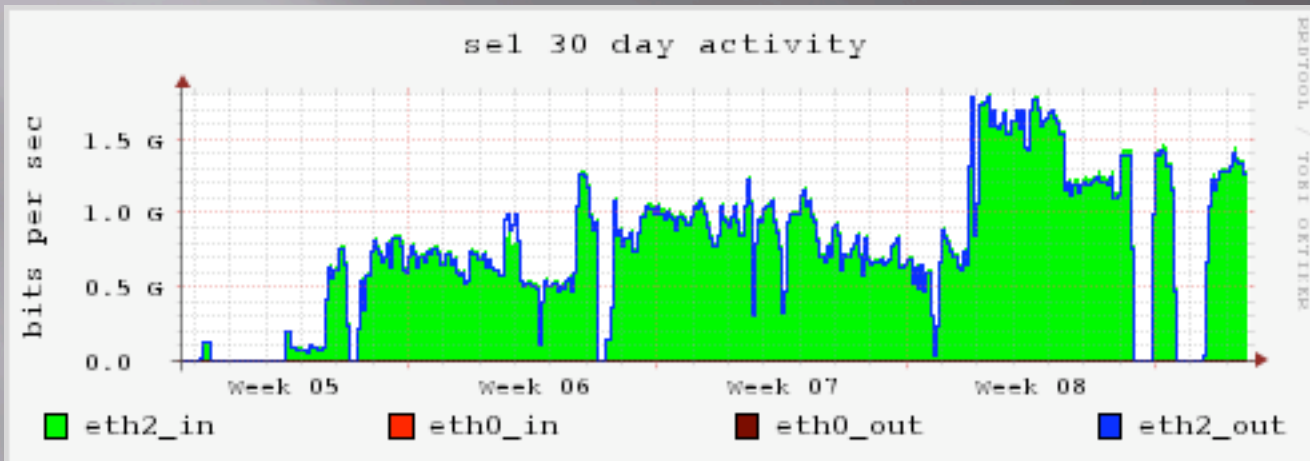
The system tray at the bottom of the screen shows the time as 6:31 pm on Friday 14 November 2008.

Extra Slides

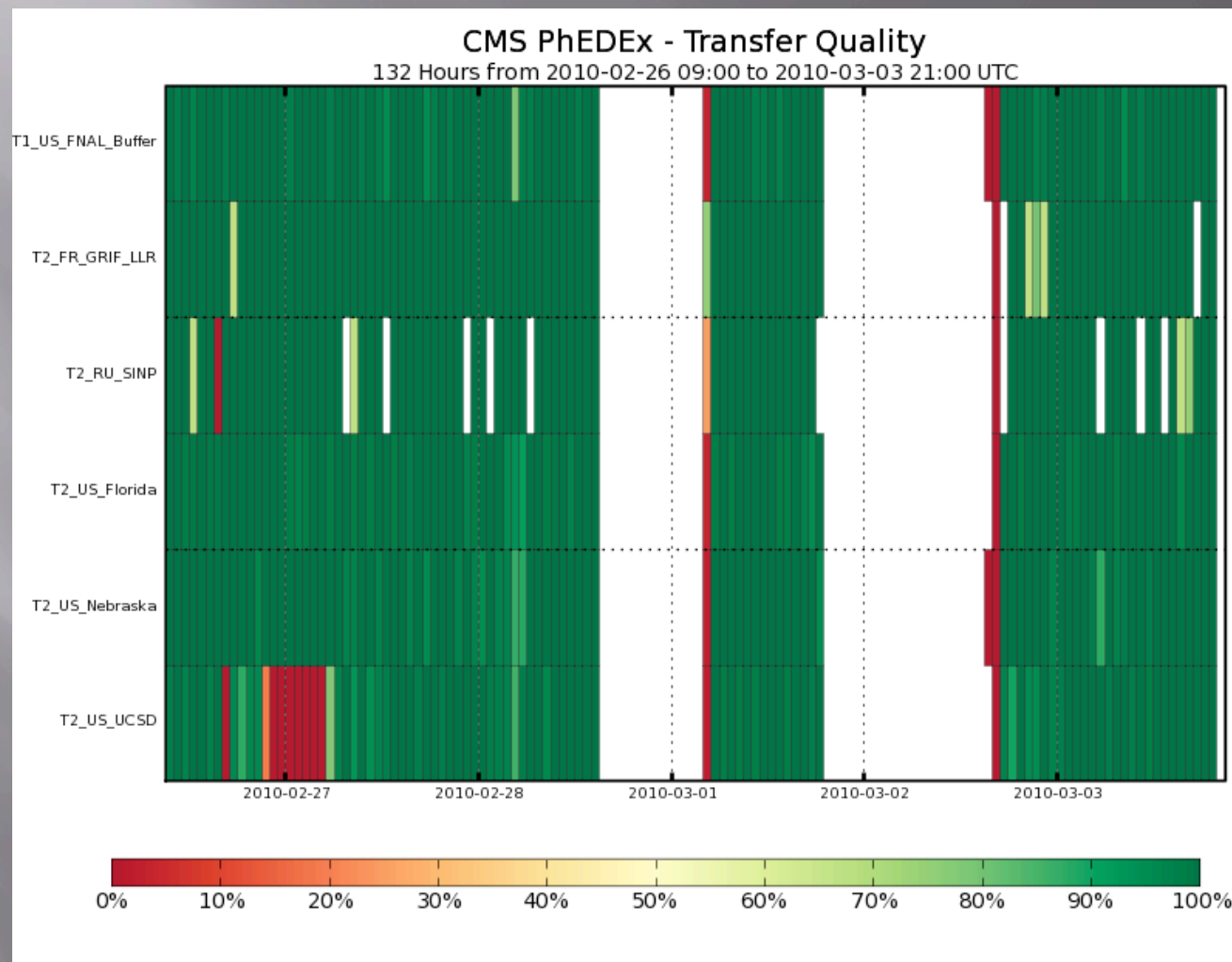
LN Layers



Vanderbilt GridFTP/L Gateway



CMS PhEDEx monitors REDDnet



CMS PhEDEx monitors REDDnet

