VANDERBILT UNIVERSITY

ACCRE
Advanced Computing Center for Research & Education

Storage @ Vanderbilt

Kevin L. Buterbaugh
klb@accre.vanderbilt.edu

OSG Storage Forum
July 1st, 2009

Who are we? (from a grid perspective)

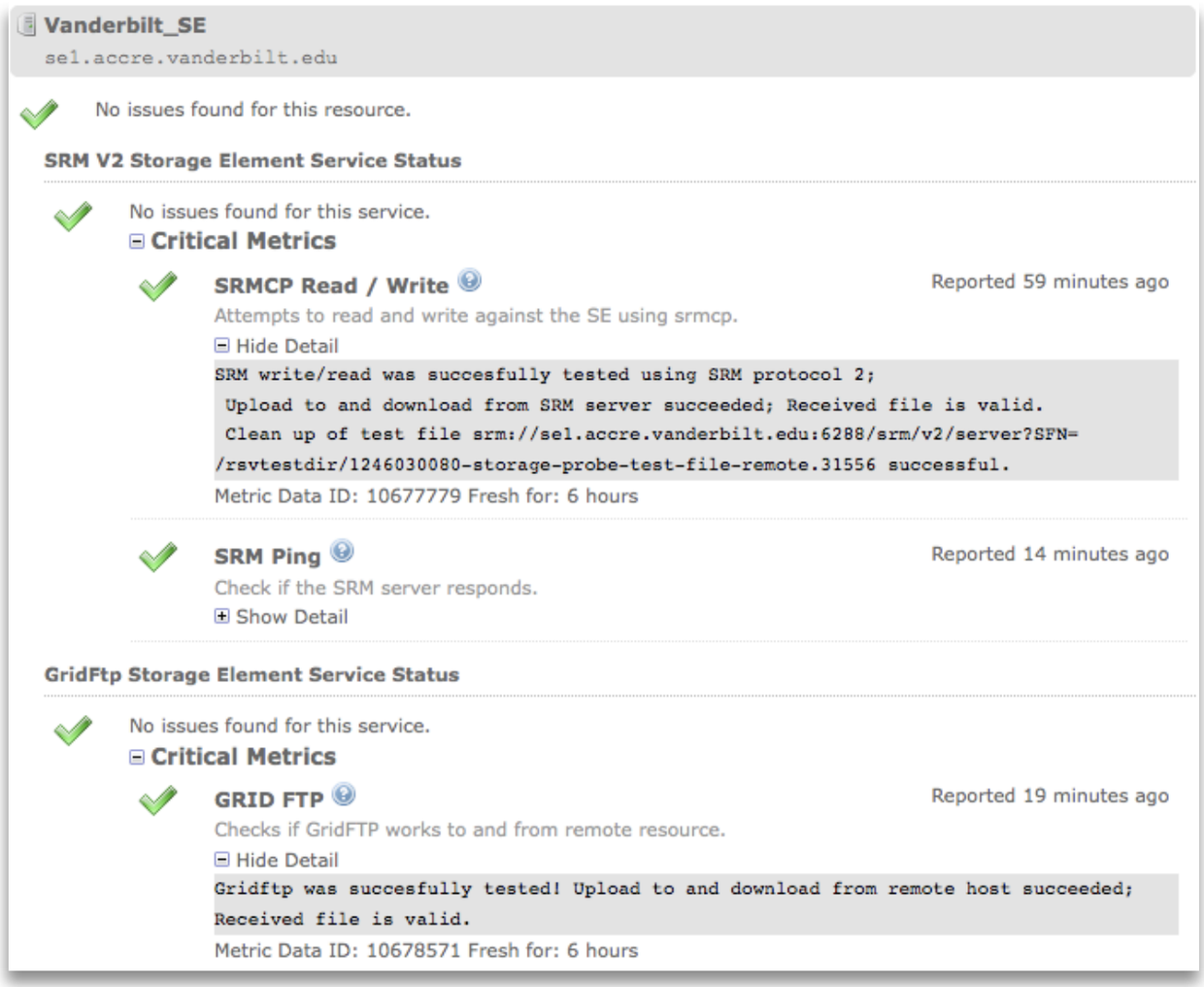
* Tier 3 associated with CMS
* One of the largest US Tier 3's
* Tier 1.x associated with CMS-HI
* More than 1700 batch slots (and that will very soon be more than 2000 batch slots)
* One CE (vampire.accre.vanderbilt.edu) that also is our GUMS server; 2 dual-core Opteron CPUs, 8 GB RAM
* Me personally?
* I'm a cluster / grid system administrator...
* I'm not a physicist!

What about storage?

* Two identical SE's (one production, one test / development)
* 2 quad-core Opteron CPU's, 16 GB RAM, 10 Gbit Ethernet
* Production SE is se1.accre.vanderbilt.edu
* BeStMan
* So ... what makes us different?
* Our back end storage!
* Instead of the standard SAN, NAS, and / or NFS setup, we use storage "depots" via REDDnet
* Why?
* NSF funded project to develop a tool to break the CMS "data tether"
* Locally supported SE that will be used by CMS-HI Tier 1.x

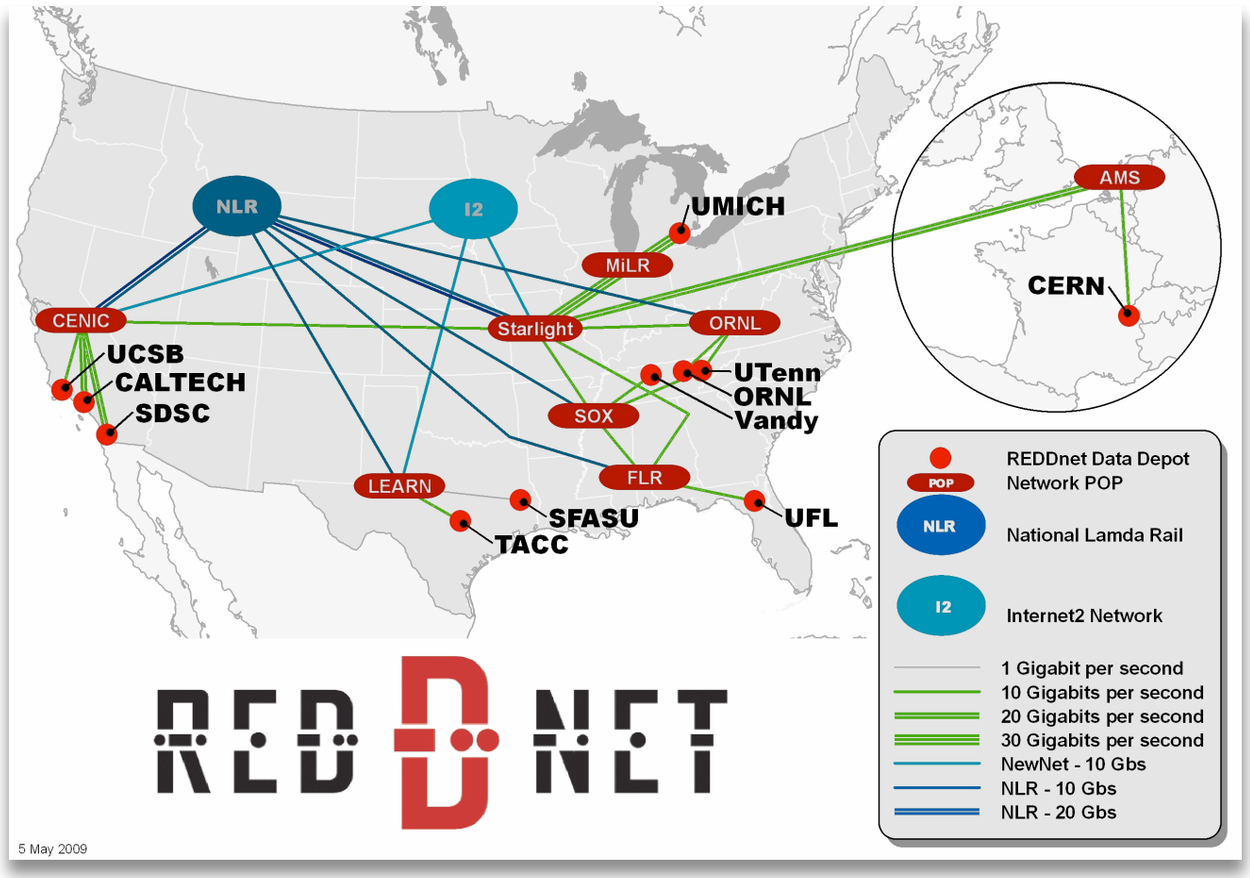
Where are we with this?

* We've only recently reintegrated REDDnet as our SE
* We also recently upgraded to the new more powerful SE listed previously ... and changed it's name - which is not a trivial thing to do
* Reading / writing to REDDnet via GridFTP and / or SRM commands works!



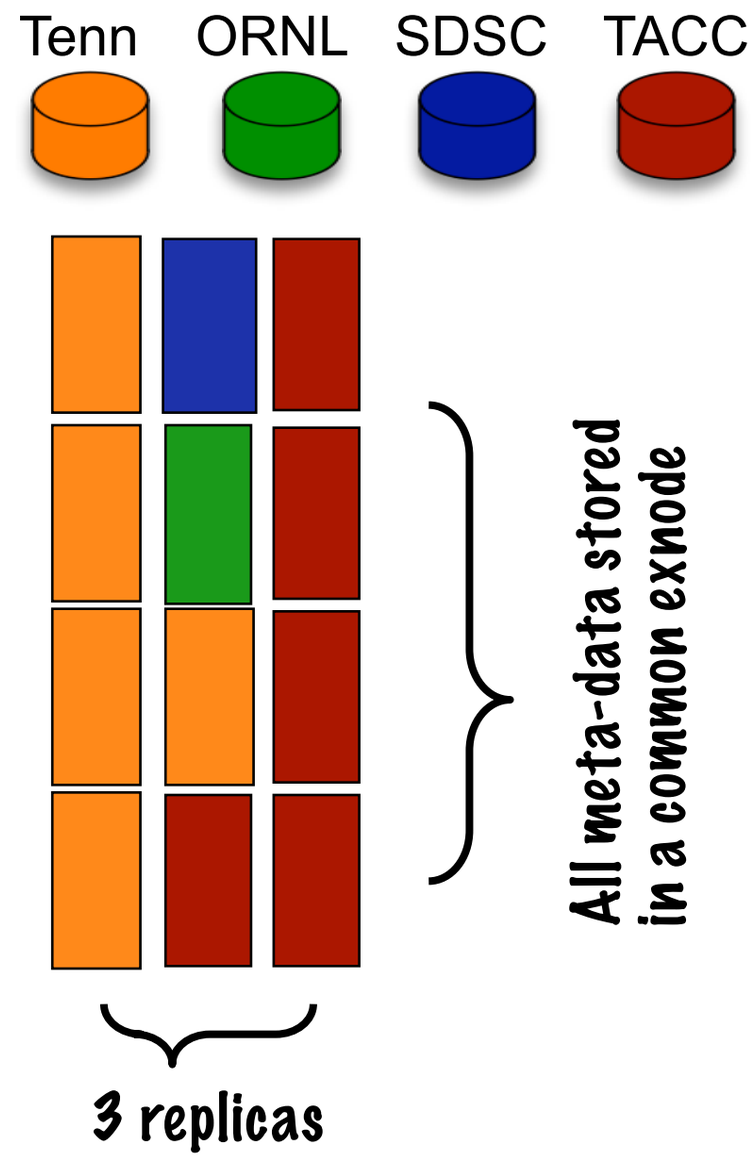
What is REDDnet?

* Research and Education Data Depot Network
* NSF Funded
* "Working storage" - manage the logistics of sharing, moving, and staging large datasets across wide areas and distributed collaborations
* "Data Pulse" - strongly interested in a data set for a brief period … then another … then another …



REDDnet - Replication

* File is broken into "slices" of fixed size
* Can replicate a file or migrate it
* Replication info is stored in the same exnode as the original file
* Example: 3 replicas across 4 depots
* If a depot (or its' network) goes down, you can still access the file (transparently)



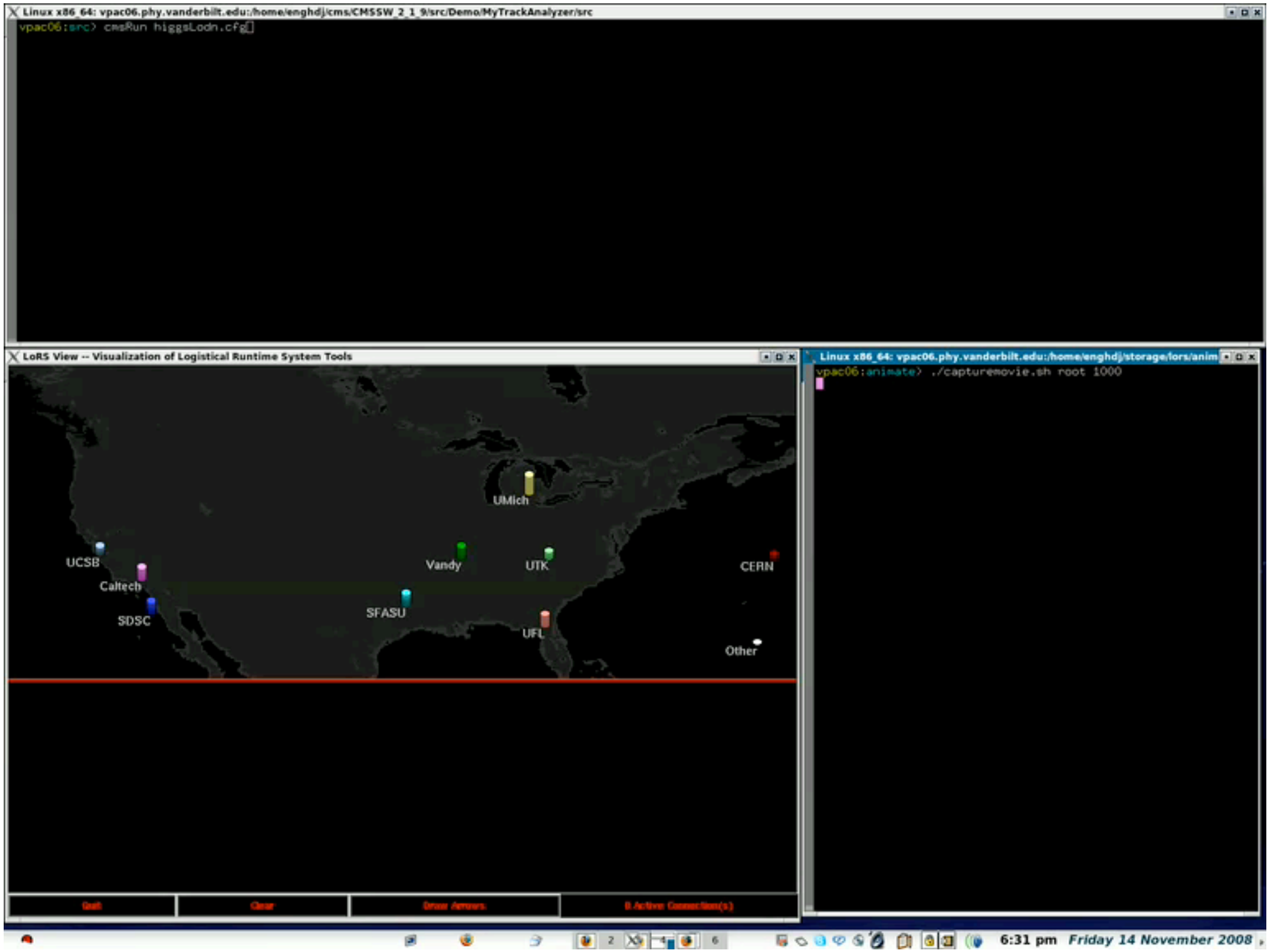
REDDnet - Policy Driven

* Can set replication / movement for a directory, sub-directory, or file
* A "generic" (non-CMS) REDDnet user may want to upload a file to a local depot at ORNL, then have it moved across the country to their university
* A CMS REDDnet user will want to replicate the file and put copies on depots near several collaborators (which addresses the current CMS latency issue)
* All of this occurs in the background (automatically) based on policy
* Security - all files can be read by anyone who knows the exnode, but to upload a file you have to have a grid certificate or a REDDnet account

CMS Tier 3 Analysis

* Inject data into REDDnet using standard CMS data movement tools (PhEDEx / gridFTP)
* Special gridFTP backend then uploads data into REDDnet
* The data is then replicated to ensure that copies are near the users' CPUs
* Plugin for CMS software - reads directly from REDDnet
* All of this is transparent to the user
* Demo...

REDDnet CMS Demo



GridFTP Reliability

* The graph below shows sustained 1-2 Gbit / second transfers to a REDDnet gridFTP server, 24 x 7 for a full month
* Transfers attempted to mimic real world usage
* The first 3 weeks worth of transfers were across campus
* The last week worth of transfers were from Caltech to Vanderbilt