Future Directions for CMS Tier-3s

Ian Fisk

Tier-3s

- Tier-3s continue to be a growth area in CMS Computing
  - They now outnumber the Tier-2s (53 in SiteDB as compared to 52 Tier-2s)
  - More than half of these are in the US.
    - 12 countries have Tier-3s.

Future Directions in Tier-3s

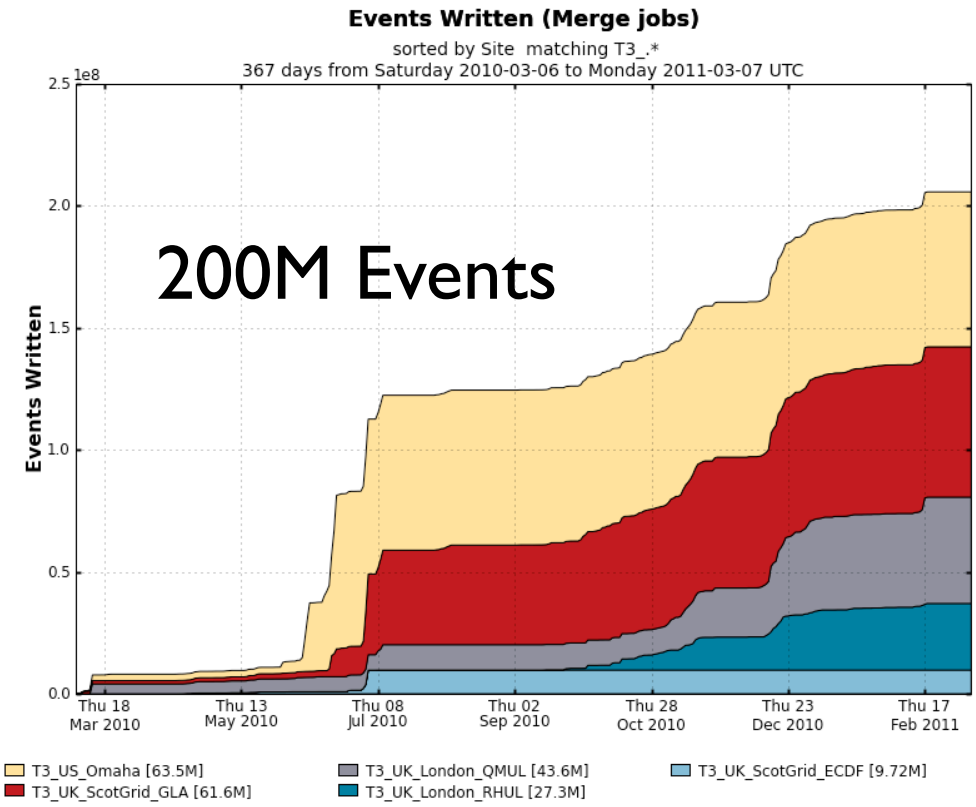
- In CMS Tier-3s cover a diverse set of resources
  - Big variations in the amount of processing and storage
  - Dedicated facilities to fractions of share installations
- A common element is the lack of a lots of effort dedicated to CMS activities
  - Typically Tier-3s are operated by fractions of physicist time or effort from a central IT organization
- A number of the items we're looking at are intended to reduce the effort needed to operate a Tier-3

How Tier-3s Benefit CMS

- In many ways CMS has been treating Tier-3s like lower capacity Tier-2s
  - All the functionality of a Tier-2 was possible, but a full set of services was need (Grid interfaces, PhEDEx, CMS Environment, /store/user)
- When looking at simplifying we need to assess what sites gain and what CMS might potentially lose
  - CMS gains by having analysis resources operated at Tier-3s.
    - It's good for the science, but it also adds resources to CMS and it lowers the load on the Tier-2s
  - CMS gains by having opportunistic access to Tier-3s for simulated event production
    - Adds resources to the experiment

Contributions of Tier3s to Simulation

- In the last year, Tier-3s have produced about 200M simulated events
  - This is a small fraction of the total, but it's additional resources
  - We could potentially use more



Quandary

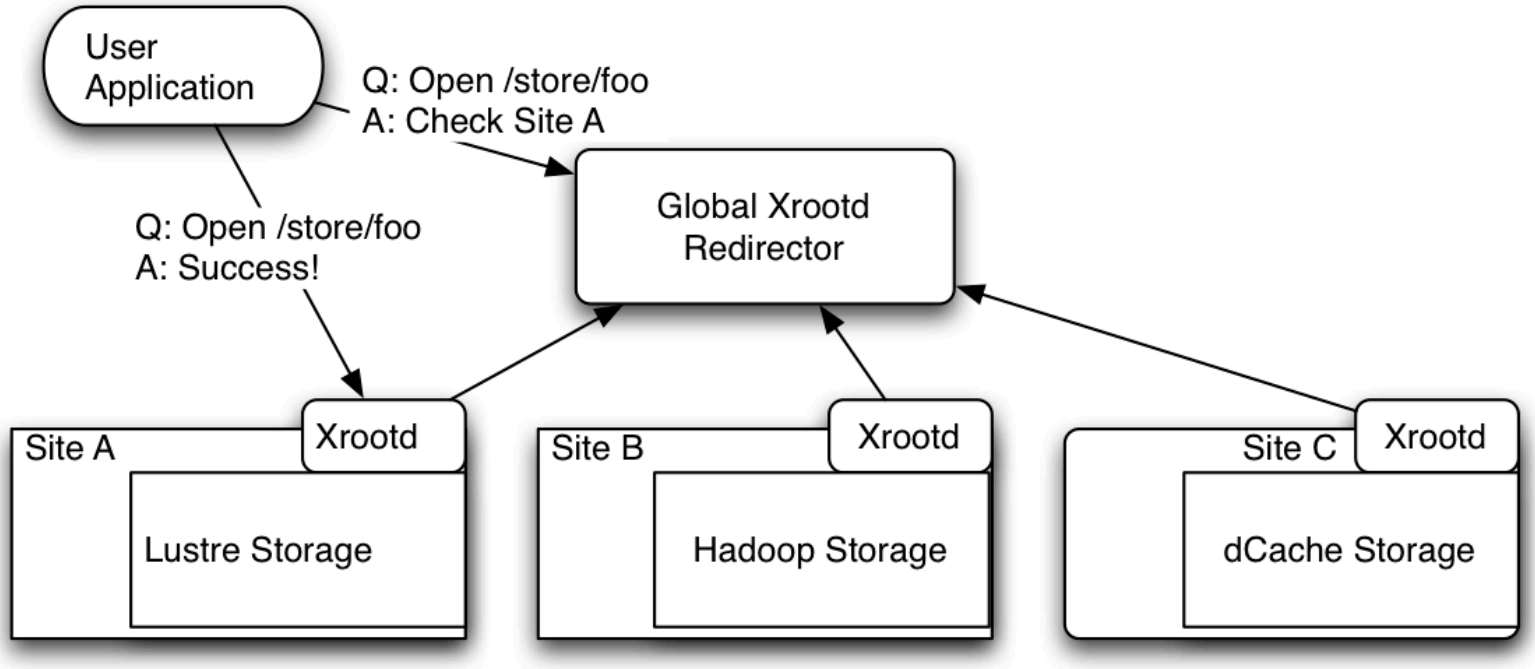
- How do we simplify the operations so that more sites are useful for analysis
  - While not overloading the rest of the infrastructure
  - And maintaining the other beneficial contributions from Tier-3s to CMS

Some items in integration

- Biggest complaints we get from the Tier-3 community is the effort needed to install and operate the grid services, and the effort needed to install and operate PhEDEx
  - We also get complaints about the effort needed to install and operate a cluster.
    - If you are complaining about this, please do not operate a Tier-3. It's better not to have a resource than it is to have a insecurely run center. There is enough analysis computing in centrally provided systems. Some Tier-2s with power and cooling infrastructure have been willing to host equipment in exchange for access to it.
- In response to the first 2 complaints we are investigating solutions
  - The Xrootd redirector work, should allow Tier-3s to operate diskless* (* Some limitations apply)
  - Investigating the use of CVMFS for software distribution

Diskless Tier-3s

- Instead of transferring data in advance with PhEDEx we can stream the data in with xrootd



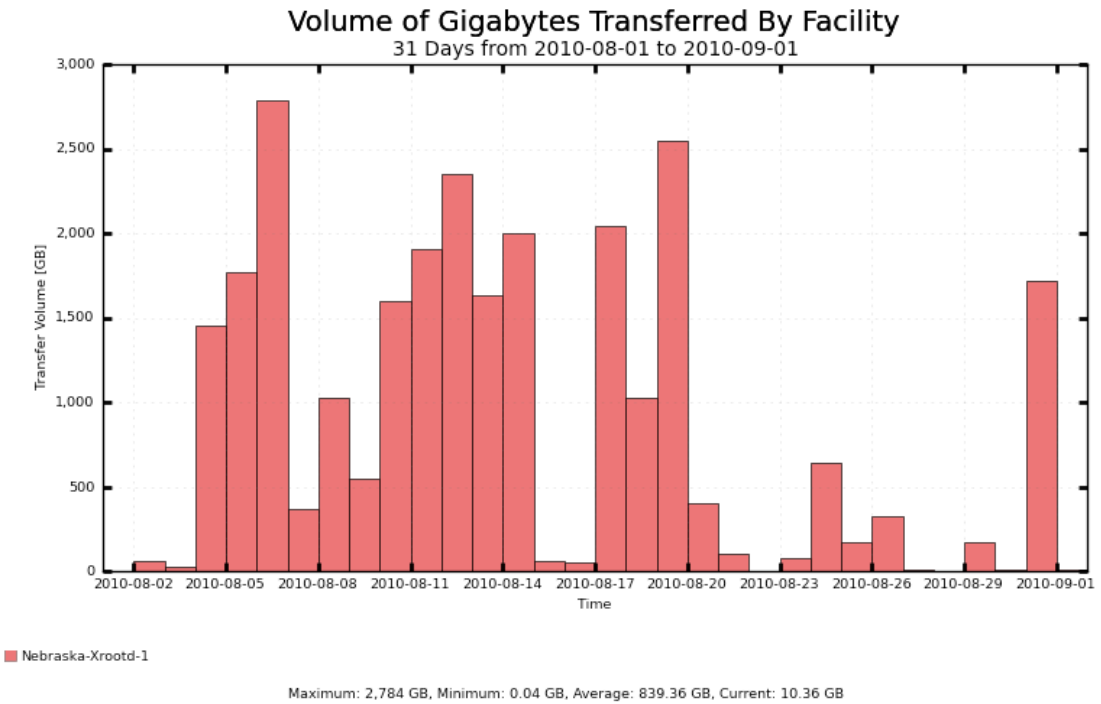
Xrootd layers on top of *existing storage element*.
Think of it as a proxy (or a door) to the site's data

Benefits

- The streaming works because people have spent considerable effort trying to optimize the IO
  - Useful in particular for samples that will only be accessed once
  - Need to look more seriously at xrood local site caching
- You don't need a local PhEDEx installation or even a grid SE
- The US is a good place to test this
  - Reasonable networking
  - 2 copies of the heavily accessed AOD: a Tier-2 copy and a FNAL copy

Working

- Omaha Tier-3 is not typical
  - 10Gb/s networking
  - 5000 cores
- CPU efficiency looks competitive



In use at Omaha - diskless T3!

Drawbacks

- We need to introduce some throttling on the source sites to make sure the reading centers can't knock something over
- We have estimated the impact caused by modest sized Tier-3s
  - Need to watch the relative fraction of analysis computing at Tier-3s and the level of activity
- A diskless Tier-3 is potentially operating with a storage element
  - This is fine, but need to think about what to do with user created data

Software Distribution

- CMS is currently looking at CVMFS to distribute software packages
  - CERN Virtual Machine File System is a read-only software distribution built around squid and fuse. It's used for distributing software and environments to CERN virtual machines, but it can also be used in regular machines
  - Appears to be a scalable way to have a distributed read-only file system
  - Still in the testing phase, but FNAL currently has replaced one of the read-only NFS nodes with CVMFS

Benefits

- We currently install software for sites that want it though the CE
  - This would allow sites to have the software independently of a functional CE
    - (Installing the CMS Software is not a big deal even manually)

Drawbacks

- The CVMFS system is still in integration, but is expected in production soon
- The lack of a CE would spare the site some operations but has 2 significant drawbacks
  - Local analysis users will need to run specialized versions of CRAB (supported by the sites)
  - CMS will not be able to access the site for opportunistic access

Future Directions for Tier-3s

- The goal for CMS is to make the Tier-3s efficient analysis resources
  - We are trying to respond to operations issues as they have come up
- CMS continues to benefit from sites being part of OSG
  - I think the sites also benefit because the services developed for the Tier-2s are more easily used
- There are some interesting tasks in development that may increase the utility of the Tier-3s
  - Hoping to also maintain the opportunistic access.