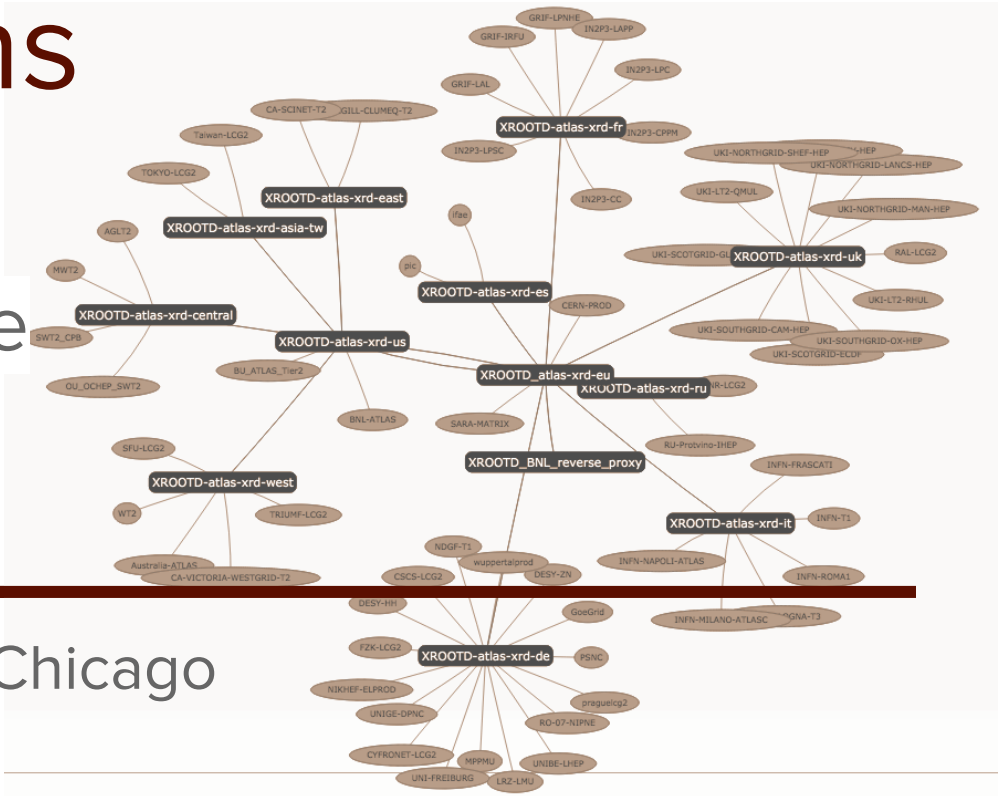
ATLAS Lessons Learned

& might do in the future



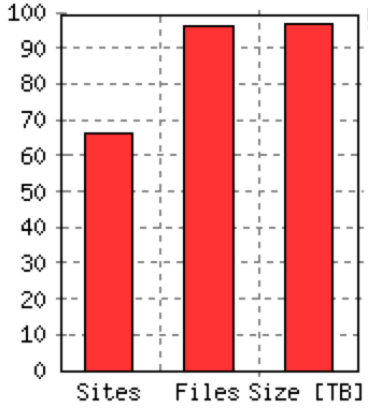
Rob Gardner • University of Chicago

XRootD Workshop
UCSD, January 27, 2015

Introduction

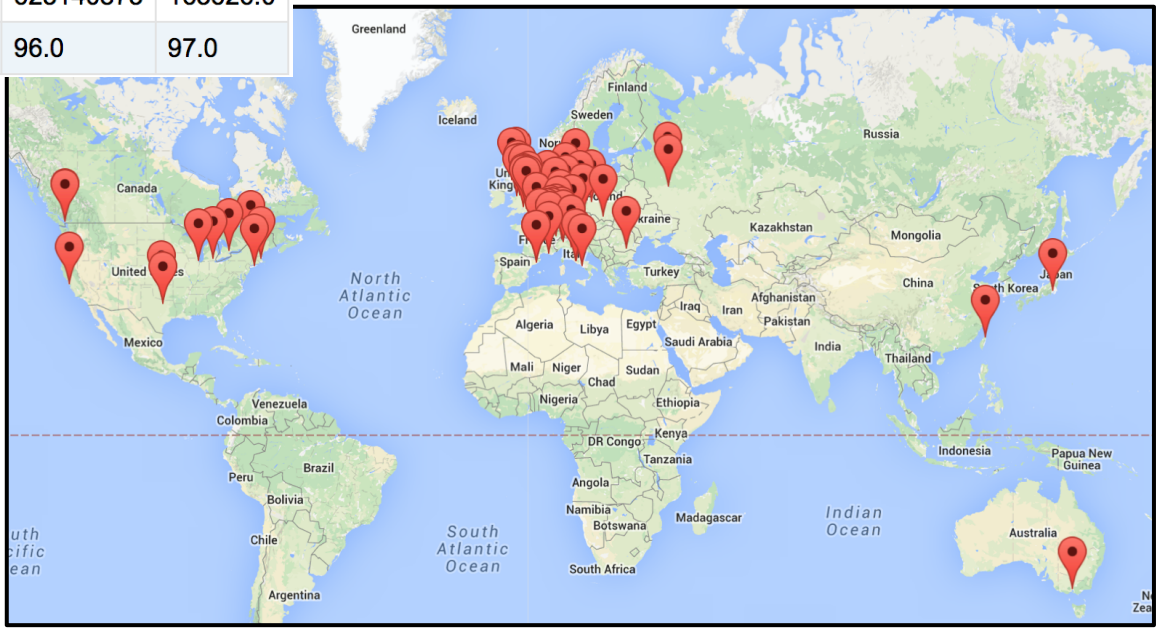
- Quick recap of FAX status
- ...and functional usage modes
- Headaches
- Site worries!
- Some metrics
- Odd ideas moving forward
- Summary and Conclusions

FAX deployment status as of Dec 2014



	Sites	Files	Size [TB]
FAX	59	600013168	158812.0
Total	90	623140873	163626.0
Coverage	66.0	96.0	97.0

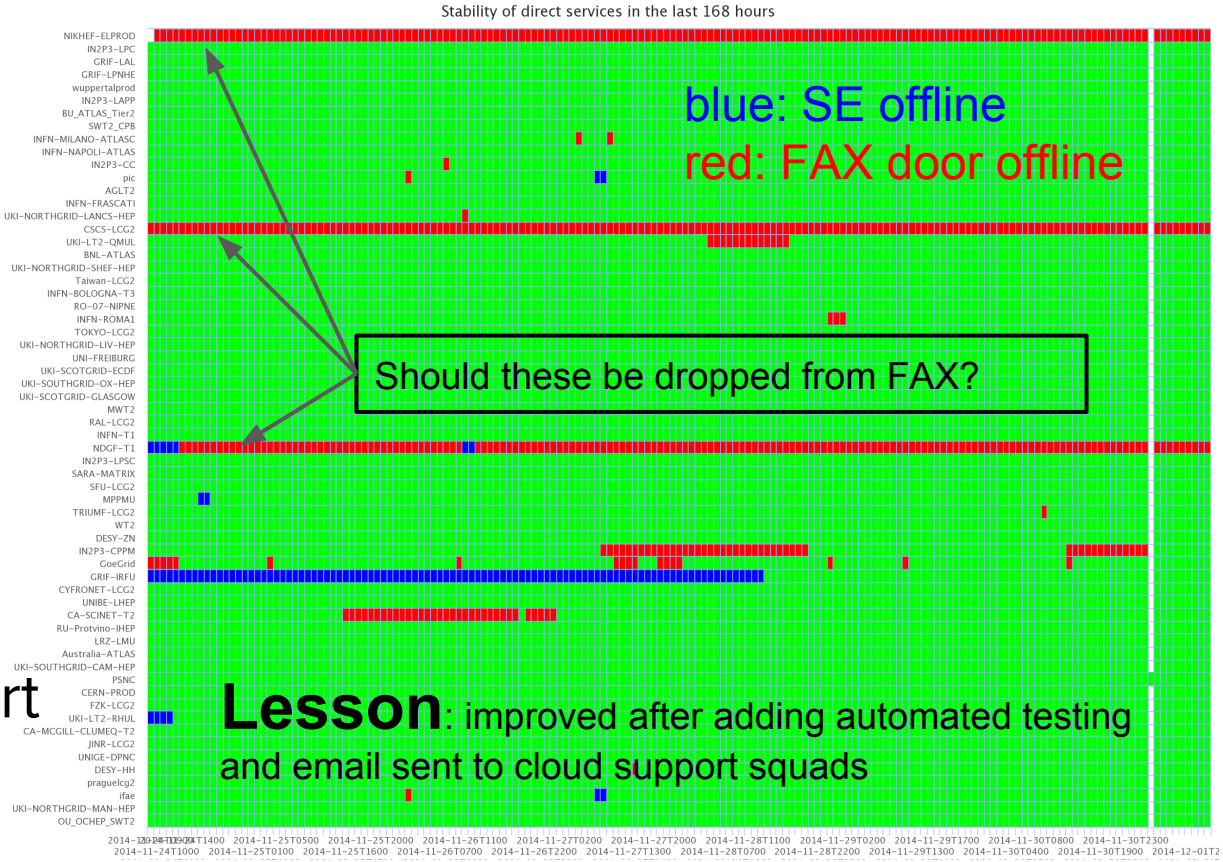
Goal reached ! >96% files covered



Regional redirectors deployed in a hierarchy (US and EU zones, and regional EU clouds)

Direct access always very stable

- Stable
- No load issues
- Monitoring change for EU “privacy” caused 3 sites to unplug (1 has since rejoined - SARA)
- Most EU sites report to US

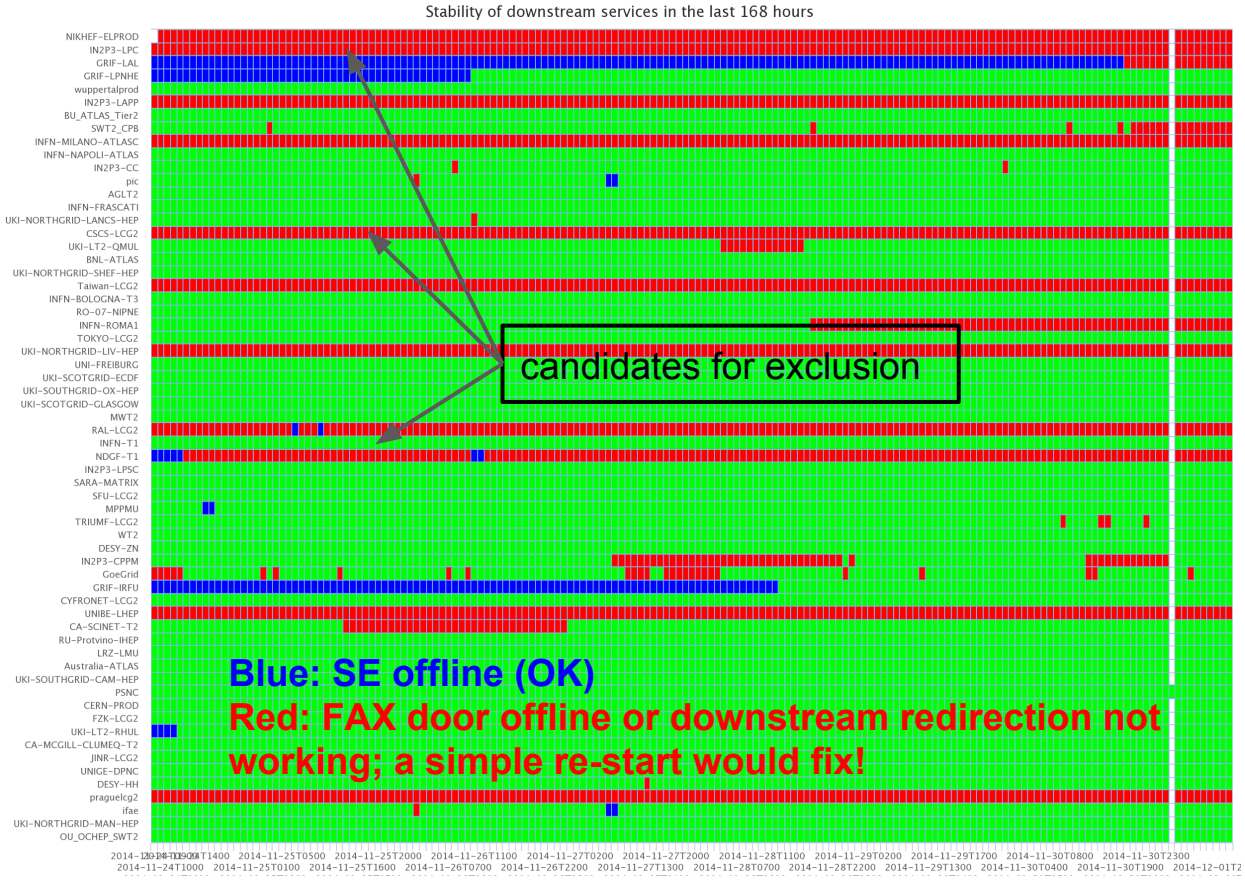


E.g. disruptions in redirection

FAX redirectors at CERN moved to AI infrastructure:

- this breaks downstream redirection
- **the change requires sites to restart their endpoint**
- this is important since overflow jobs get the files through downstream redirection

Lesson: XRootD door and FAX redirector not currently part of a blacklist, so we don't always have full attention of site admins (yet).



Monitoring challenges

headaches

Lesson: a long and fitful path to standardize formats from various SEs; still not 100% successful

- EU sites have started to send monitoring data to the CERN collector
  - Thanks to Igor Pelevanyuk current state may be seen here: http://dashb-xrootd-comp.cern.ch/cosmic/ATLASmigrationMonitoring/ (most EU sites still send to US)
- Still **a lot of effort needed by FAX and dashboard teams** to make summary and detailed monitoring match (its a long chain from site to dashboard):
  - http://dashb-ai-621.cern.ch/cosmic/DB_ML_Comparator/
- Started deeper analysis of Panda job info data transported into Hadoop at CERN.
- Further improvements in Fax Status Board (http://waniotest.appspot.com/)

WAN access modes implemented in ProdSys2

- Failover

  In the case stage-in fails due to a temporary SE related problem, the [illegible] attempt the stage-in a second time after a few minutes. If that fails as well [illegible] to attempt stage-in from a remote SE using FAX.

- Overflow

  When deciding where [illegible] estimate that it is better to send it to a site that does not [illegible] let it read from FAX, rather than let it sit in the queue of the site that h[illegible]ited to analysis queues.

- Explicit overflow

  If a user explicitly requires CE that does not have the input data, the task will be brokered to that CE and FAX used to get the data.

Lesson: having better monitoring tools to assess functional behavior, performance, and cost-benefit to controls

Lesson: Soft bandwidth controls at the site to relieve site admin anxiety

Failover (*)

Setup per Panda queue using two AGIS fields:

- **allowfax**=True will enable FAX retries.
- **faxredirector** sets the FAX access point to be used. For optimal performance it should be set to the site's closest redirector.

(*) Enabled by default for all Panda queues March 2014

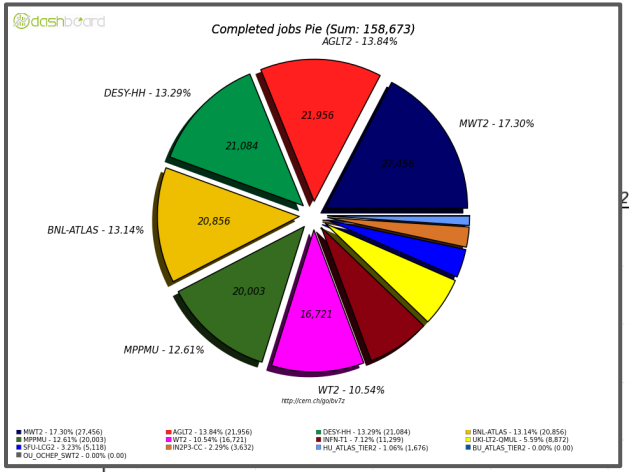
Overflow

Other queue settings:

- **wansinklimit**** - limits the bandwidth that jobs overflown to the site can use.
- **wansourcelimit**** - limits the bandwidth that site's FAX endpoint can deliver to jobs overflown elsewhere.

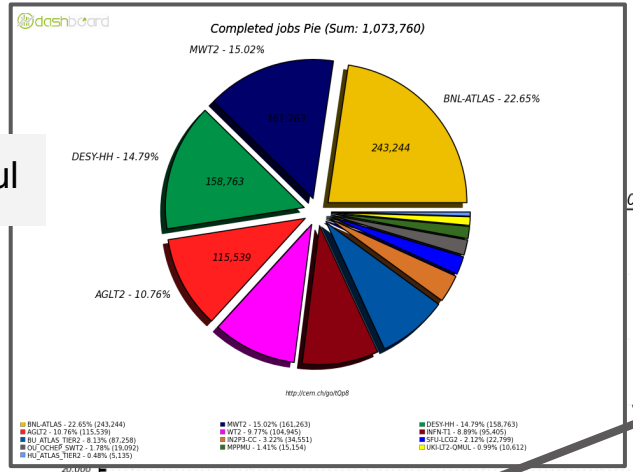
** zero value turns off overflow in that direction.

Overflow activated during peak periods



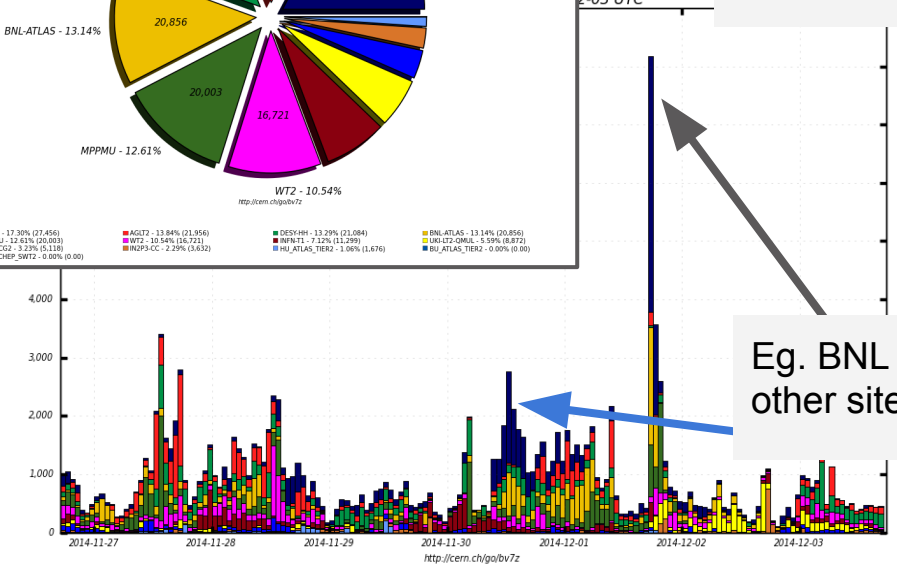
Overflow

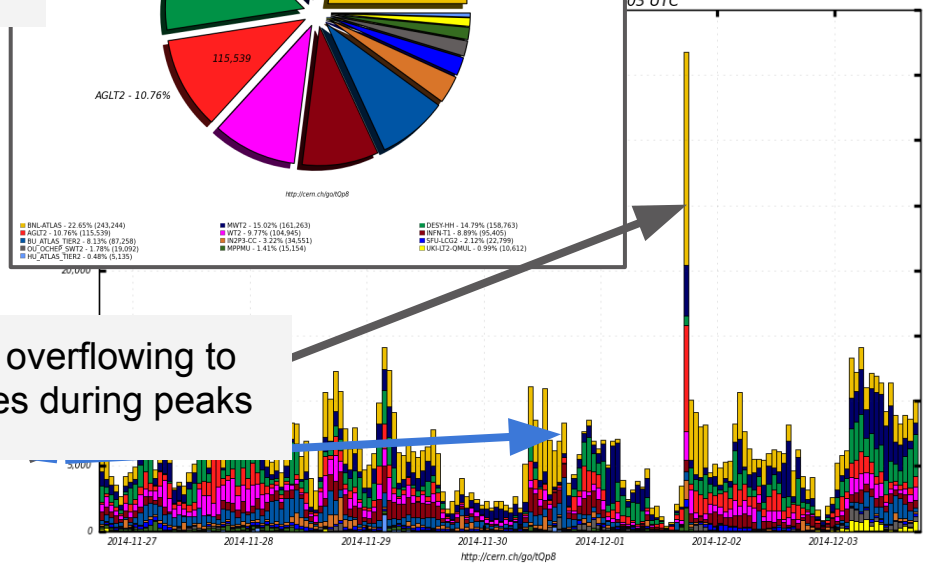
Successful



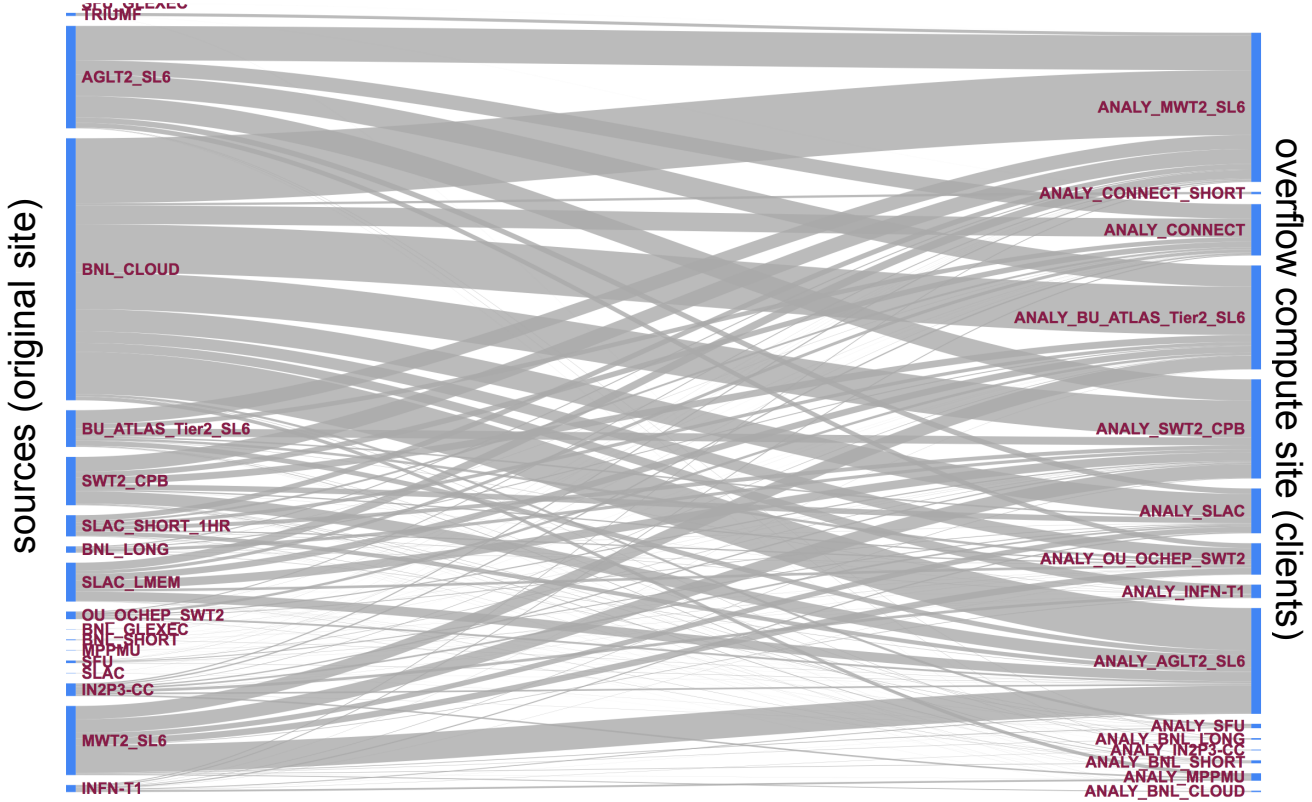
Regular

Eg. BNL overflowing to other sites during peaks





Overflow data flow patterns

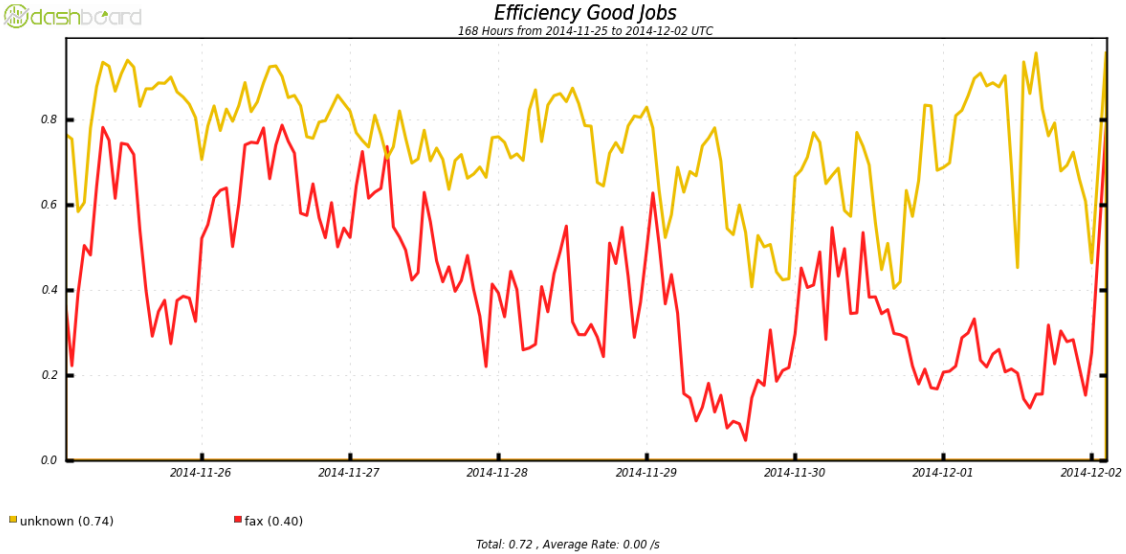


Linewidth is proportional to log(number of jobs) between a source and destination.

Plot is integral over last 4 months, so recently added sites show small values.

As expected largest source was BNL, but Tier2-Tier2 sharing possible

CPU efficiency



Strong dependence on job mix (user code).

Jobs with TTreeCache have roughly the same efficiency as local jobs.

The move to new versions of ROOT and xAODs format should improve CPU efficiency considerably.

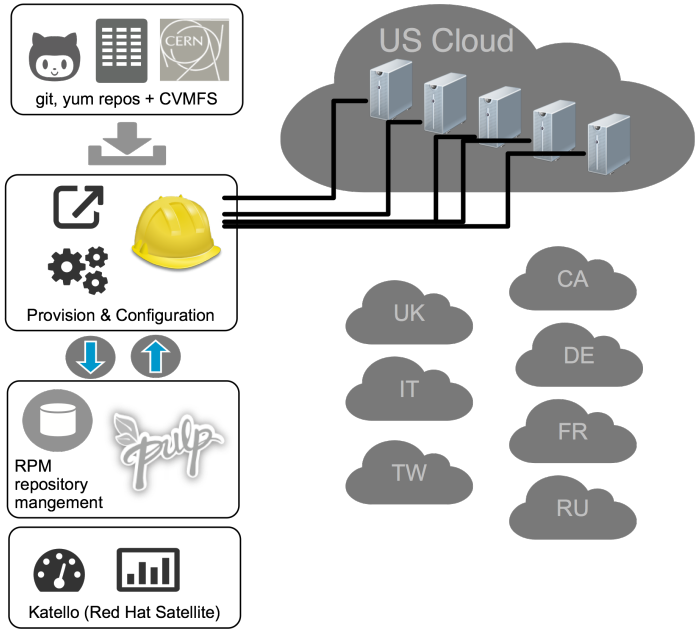
Not as much a worry with improvements in ROOT & as users migrate (should improve with time)

Challenges & time synchs

- Catalog dependence on N2N translation pre-Rucio
- Diversity of storage backends (dCache, DPM, EOS, XRootD, POSIX) ➔ distributed integration development
  - Synchronizing plugins across repositories in use
- Complex distributed monitoring infrastructure and difficulties in validation
- Deployment owned by the experiment, rather than WLCG deployment (grid)

Something different? (in theory)

- A trusted central configuration service so that local interfaces run “hands free”
- Leverage advances in data center management for remote provisioning
- Central expert team can monitor and make changes



Conclusions

FAX infrastructure

- Deployment is essentially finished (> 96% files are accessible, 600M files, 160 PB)
- Stable running for 100% failover deployment and 14 overflow-enabled analysis queues.

Enabling Overflow

- At the scale of 500-2000 finished jobs per hour not a serious load on the infrastructure.
- Brokering decisions could be analyzed to determine optimal tuning (analytics).
- CPU efficiency will improve with the adoption of the new ROOT release and switch to xAOD

Future

- Lightweight analytics stream based on fluming logs + Hadoop
- Can we do better with centralized configuration management?