ATLAS XRootd Demonstrator

Doug Benjamin
Duke University
On behalf of ATLAS

Preamble

- This talk represents the work of many people within ATLAS. Would like to recognize and thank them.
  - Idea – Massimo Lamanna
  - Coordination/ Tier 3 testing - DB
  - Tier 1/Tier 2 interface (coding/testing) – Charles Waldman, Rob Gardner, Brian Bockelman(CMS)
  - XRootd expert guidance/help – Andy Hanushevsky, Wei Yang, Dirk Duellman, Andreas Joachim Peters, Massimo Lamanna
  - ATLAS DDM - Angelos Molfetas , Simone Capanna, Hiro Ito
  - US ATLAS Tier 1/2 – Patrick McGuigan, Shawn McKee, Saul Youseef, John Brunelle, Hiro Ito, Wei Yang
  - ATLAS Tier 3 site (participate in immediate future) – Justin Ross (US), Santiago Gonzalez de la Hoz (Spain), Wahid Bhimji (UK), Andreas Penzold (Germany)

Description

- XRootd makes an excellent data discovery and data transfer system when used with a global name space (ATLAS DDM folks call it the Global Physical Dataset Path)
- It is used to *augment* and not replace the existing ATLAS DDM system
- Tier 1 and Tier 2 sites would act as seeds for the data sources (Read only)
- Tier 3 both fetch the data and serve it (sink and source – reducing load between T1/T2 and T3)

Advantages

- Tier 1 and Tier 2 sites can limit the bandwidth using XRootd (overloading the system not an issue)
- Tier 3 sites would participate in delivering data to other Tier 3 sites
- Popular data sets would be replicated as required based on usage (similar to ATLAS PD2P)
- Tier 3 sites co-located with Tier 2 (or Tier 1) sites could use XRootd data servers attached T2 storage for read only access
- Complex Tier 2 sites (those with separate physical locations) can share data to jobs through XRootd protocol
- CERN based physicists can easily fetch their own ROOT files from their home institute clusters to their local desktop

Required Steps

- Global name space convention and implementation (including client tools)
- Couple XRootd to LFC and local storage space at ATLAS DDM sites (Tier 1/Tier 2 and some Tier 3 sites)
- Local site caching of files at Tier 3's
- Implement ATLAS data security policies

Global Physical Dataset Path (PDP)

(from Angelos Molfetas slides last ATLAS software week)

- Background:
  - Requested for Tier-3 sites
  - Store datasets in a commonly accessible area
- Support in the dq2 clients:
  - Download to PDP using dq2-get
  - List datasets and files in PDP using dq2-ls
  - Base path and prefix can be specified
- Implications:
  - Lesser reliance on Site Services for Tier-3
  - Lesser reliance on LFC for Tier-3
- PDP example:

DSN:

mc10_7TeV.105563.Pythia_Gee_003_800.merge.AOD.e574_s933_s946_r1652_r1700_tid192840_00

PDP:

/grid/atlas/dq2/mc10_7TeV/AOD/e574_s933_s946_r1652_r1700/mc10_7TeV.105563.Pythia_Gee_003_800.merge.AOD.e574_s933_s946_r1652_r1700_tid192840_00/

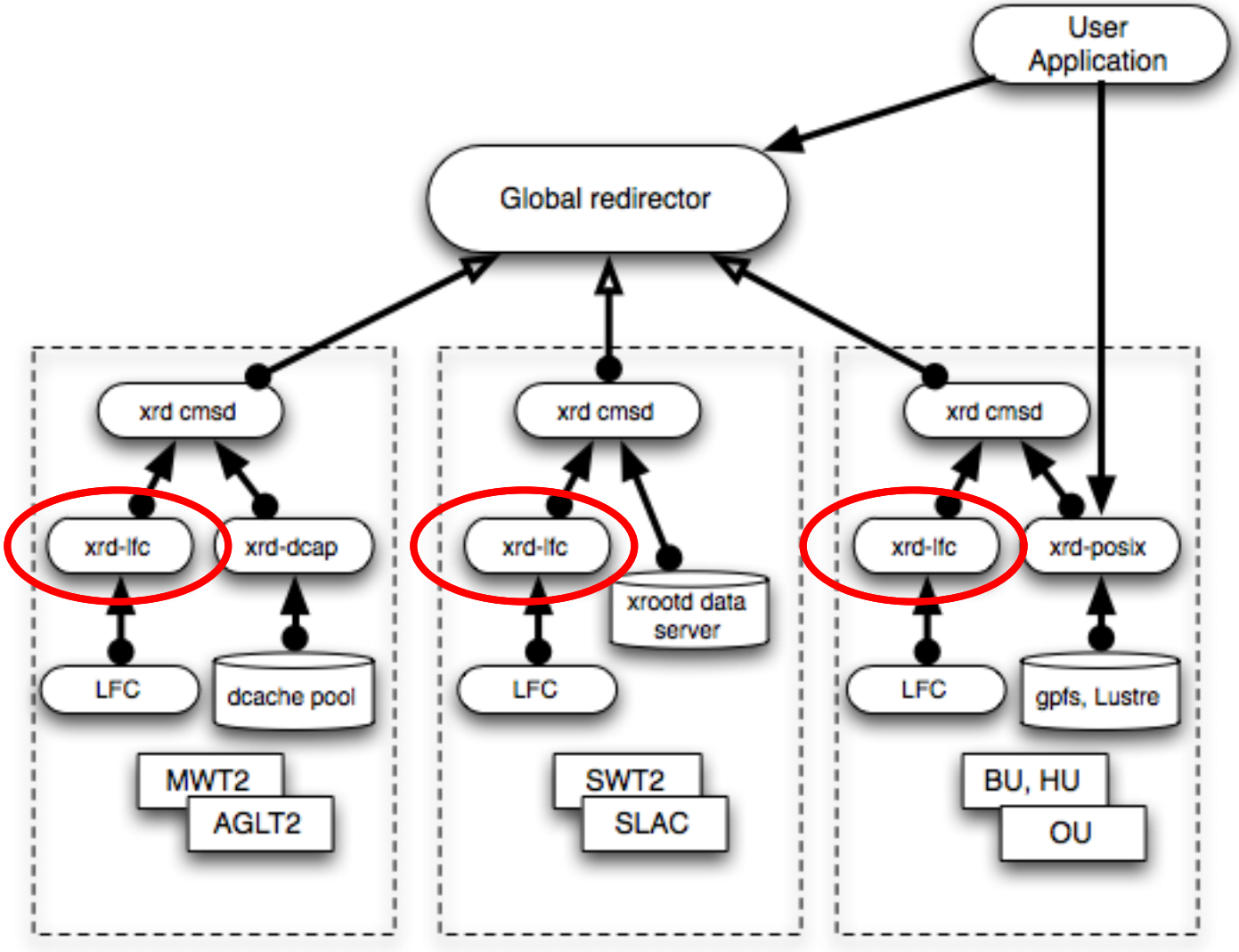
base path and prefix:

LFC to global name space (PDP) coupling

(from Charles Waldman's slides last ATLAS software week)

- Use Xrd plug-in architeture to read from non-xroot storage ("fslib"), and also to map global namespace to local storage paths ("Name2Name") [initial idea – Brian Bockelman]
- Supports dCache, xrootd, GPFS, Lustre, etc as storage backends
- Tier 3 sites can store files according to global namespace (no LFC)
- Code (dev)

  http://repo.mwt2.org/viewvc/xrd-lfc



(from Charles Waldman's slides last ATLAS software week)

Data caching in and out of Tier 3 sites

XRootd transfer service (FRM) plugin

- Transfer process calls simple shell script to fetch the data
- Currently testing with XRootd copy command xrdcp in simple bash script

  Could use another transfer command if needed

Some Tier 3's (and Tier 2 sites) have data servers on Private networks

- Transfer files to and from such a site using the XRootd Proxy service

Current Status

- ATLAS DDM Client tools (dq2-get,dq2-ls) modified to use global physical dataset path (PDP)
- Deployed services at four Tier 2's in US, 2 Tier 3's in US, 1 Test site in US
  - MWT2, AGLT2 (dCache), SLAC, SWT2 (XRootd)
  - Will deploy at 5th NET2 (GPFS) shortly.
- Functional testing successful:
  - Xrd-native, Xrd-dcap, Xrd-direct pool, XRootd extreme copy
  - Testing ganglia based monitoring for XRootd data flows out of Tier 2/3 sites.
- Tier 3 XRootd transfer service (FRM) testing
  - Transferred files automatically between Tier 3 sites
  - Native XRootd copy (xrdcp)
  - Triggered from user root analysis job
  - Can use sites with data clusters on private networks – need edge service (Proxy machine)

Plans

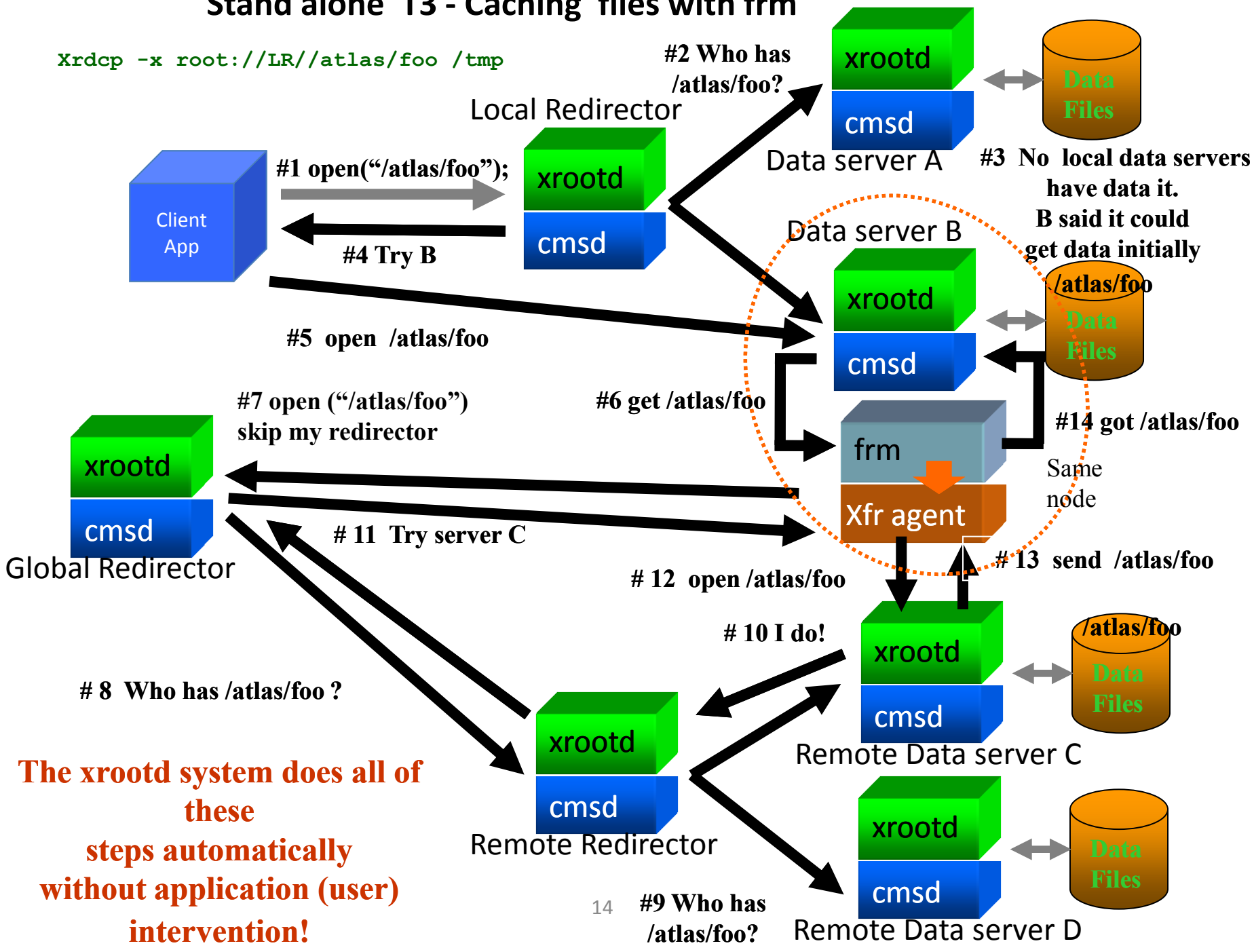
- Near term (< 6 weeks)
  - Add a few addition Tier 2 sites in US and/or Europe
  - Add a few addition Tier 3 sites in US and Europe
  - Test Tier 3 site caching (for Lustre, GPFS back ends)
  - Add automated testing (HAMMER Cloud)
  - Would like to add additional global redirector in Europe (redundancy – reduce RTT for European sites)
- Longer term
  - Increase the number of Tier 3 sites testing the system on both sides of the Atlantic (10-15)
  - Authentication
    - Use X509 to allow authorized users access ATLAS data
    - Write activity triggered by trusted servers

Conclusions

- XRootd is a good match for data deliver amongst Tier 3 sites while reducing load on rest of ATLAS DDM system
- ATLAS has made much progress in developing the components needed to use such a system
- Should help us keep on top of users data needs

Backup Slides

Stand alone T3 - Caching files with frm



Direct Pool Access