Computation Institute

Site issues and deployment of federated XrootD infrastructure in ATLAS

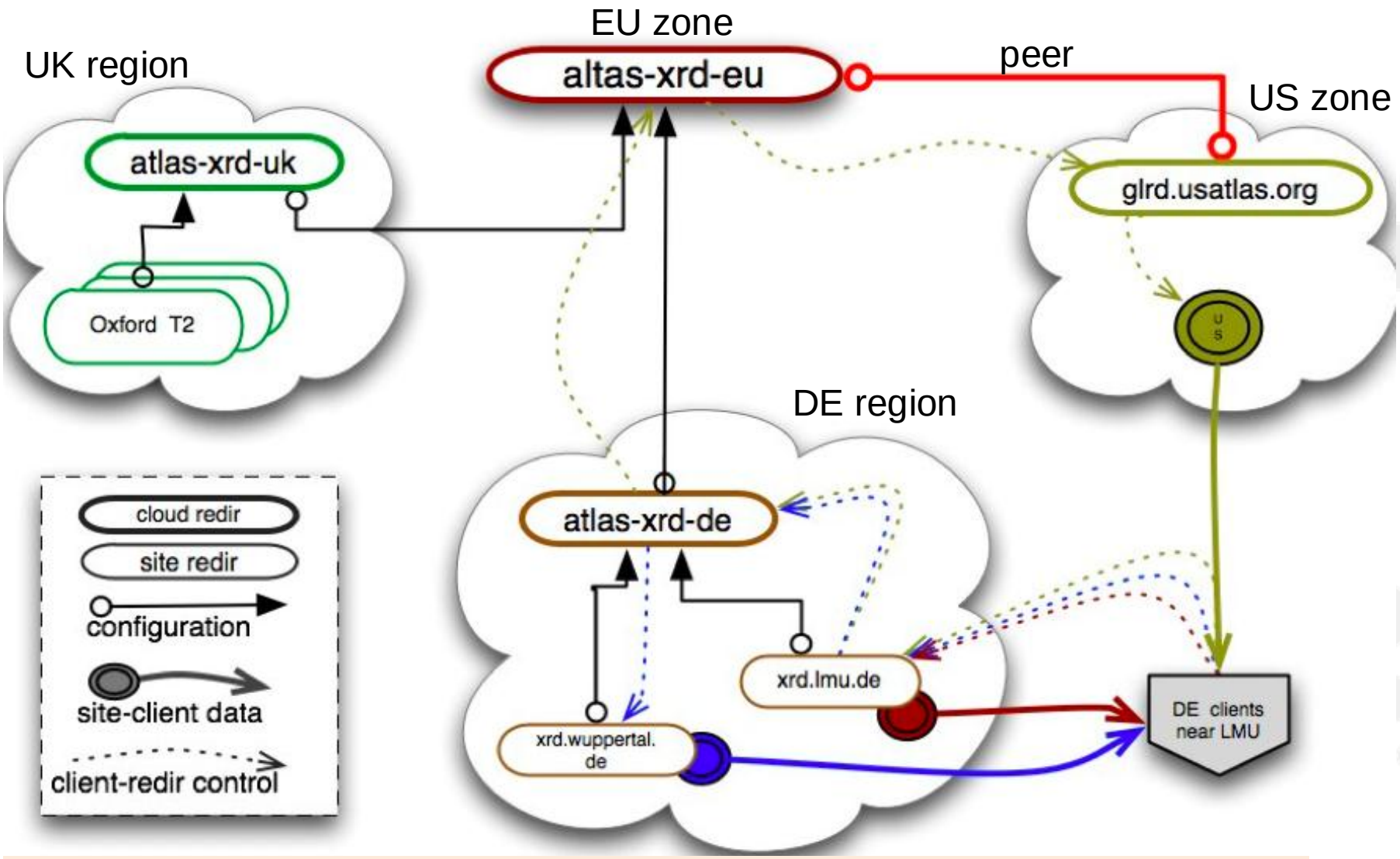
Rob Gardner
Computation and Enrico Fermi Institutes
University of Chicago

WLCG Storage Federations meeting – 'Toward Sites, Part 2'
November 8, 2012

Quick review of goals

- Common ATLAS namespace across all storage sites, accessible from anywhere
- Easy to use, homogeneous access to data
- Identified initial use cases
  - Failover from stage-in problems with local SE
    - Now implemented, in production on several sites
  - Gain access to more CPUs using WAN direct read access
    - Allow brokering to Tier 2s with partial datasets
    - Opportunistic resources without local ATLAS storage
  - Use as caching mechanism at sites to reduce local data management tasks
    - Eliminate cataloging, consistency checking, deletion services
- WAN data access group formed in ATLAS to determine use cases & requirements on infrastructure

Our concept of federation

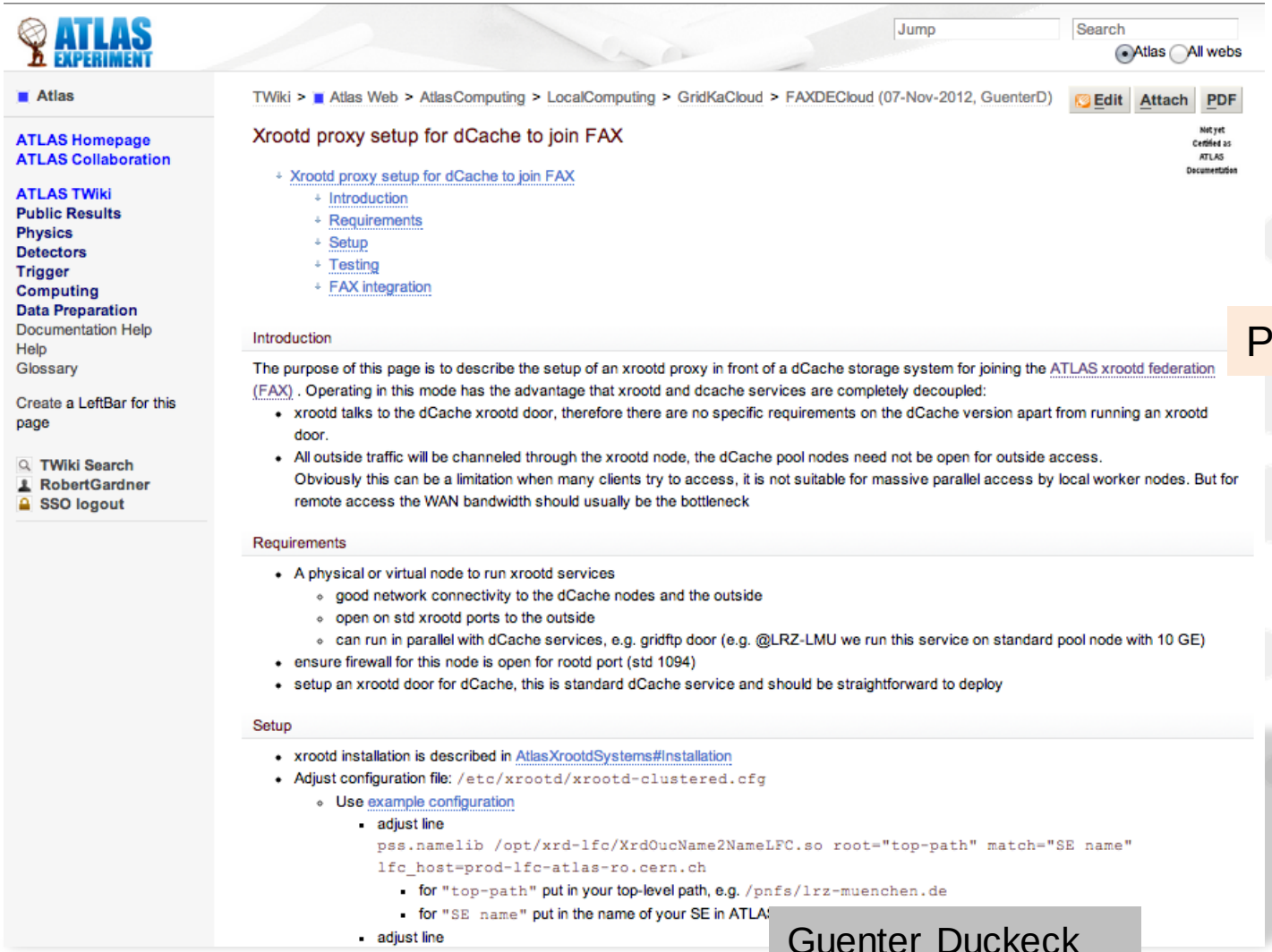


Start search locally, redirect as needed (local, cloud region, zone, global)
Uniform access to loosely coupled storage resources

Federated access, federated site deployment

- Model has been to work with ATLAS contacts from clouds
- US: the Tier1, all Tier2 centers
  - separately, a number of off-grid Tier 3 sites
- UK: four Tier 2 sites, working on N2N for Castor at RAL
- DE: three Tier 2 sites plus a CZ site, gearing up for more
- RU: two Tier 2 sites federated
- IT: DPM sites and atlas-xrd-it.cern.ch getting setup
- EOS – but with concerns about IO load from WAN accesses
- Network of redirectors and peering established, gaining practical operational experience

Site documentation (DE cloud region)



ATLAS EXPERIMENT

Jump | Search | Atlas | All webs

Atlas

ATLAS Homepage
ATLAS Collaboration

ATLAS TWiki
Public Results
Physics
Detectors
Trigger
Computing
Data Preparation
Documentation Help
Help
Glossary

Create a LeftBar for this page

TWiki Search
RobertGardner
SSO logout

TWiki > Atlas Web > AtlasComputing > LocalComputing > GridKaCloud > FAXDECloud (07-Nov-2012, GuenterD) | Edit | Attach | PDF

Xrootd proxy setup for dCache to join FAX

Not yet Certified as ATLAS Documentation

- Xrootd proxy setup for dCache to join FAX
  - Introduction
  - Requirements
  - Setup
  - Testing
  - FAX integration

Introduction

The purpose of this page is to describe the setup of an xrootd proxy in front of a dCache storage system for joining the ATLAS xrootd federation (FAX) . Operating in this mode has the advantage that xrootd and dcache services are completely decoupled:

- xrootd talks to the dCache xrootd door, therefore there are no specific requirements on the dCache version apart from running an xrootd door.
- All outside traffic will be channeled through the xrootd node, the dCache pool nodes need not be open for outside access. Obviously this can be a limitation when many clients try to access, it is not suitable for massive parallel access by local worker nodes. But for remote access the WAN bandwidth should usually be the bottleneck

Requirements

- A physical or virtual node to run xrootd services
  - good network connectivity to the dCache nodes and the outside
  - open on std xrootd ports to the outside
  - can run in parallel with dCache services, e.g. gridftp door (e.g. @LRZ-LMU we run this service on standard pool node with 10 GE)
- ensure firewall for this node is open for rootd port (std 1094)
- setup an xrootd door for dCache, this is standard dCache service and should be straightforward to deploy

Setup

- xrootd installation is described in AtlasXrootdSystems#Installation
- Adjust configuration file: `/etc/xrootd/xrootd-clustered.cfg`
  - Use example configuration
    - adjust line
      `pss.namelib /opt/xrd-lfc/XrdOucName2NameLFC.so root="top-path" match="SE name" lfc_host=prod-lfc-atlas-ro.cern.ch`
      - for `"top-path"` put in your top-level path, e.g. `/pnfs/lrz-muenchen.de`
      - for `"SE name"` put in the name of your SE in ATLAS
    - adjust line

Proxy for dCache

Guenter Duckeck

https://svnweb.cern.ch/trac/lcgdm/wiki/Dpm/Xroot/Setup#ATLASfederation

ATLAS federation

To join the ATLAS xroot federation you need to:

```
(i) choose and alert the admins of the regional redirect
and ask for its xrootd and cmsd port numbers,
e.g. atlas-xrd-eu.cern.ch uses xrootd port 1094 and cmsd port 1098
(ii) install "/usr/lib64/XrdOucName2NameLFC.so" (provided by atlas, see https://twiki.cern.ch/twiki/bin/viewauth
It requires lfc-devel as a dependency. e.g. "yum install lfc-devel"
(iii) set yaim variables e.g.

DPM_XROOTD_FEDREDIRS="atlas-xrd-eu.cern.ch:1094:1098,atlas,/atlas"
DPM_XROOTD_FED_ATLAS_NAMELIBPFX="/dpm/<sitename>/home/atlas"
DPM_XROOTD_FED_ATLAS_NAMELIB="XrdOucName2NameLFC.so root=/dpm/<sitename>/home/atlas match=dpmhost.example.com"
DPM_XROOTD_FED_ATLAS_SETENV="LFC_HOST=prod-lfc-atlas-ro.cern.ch LFC_CONRETRY=0 GLOBUS_THREAD_MODEL=pthread CSEC_
```

(change <sitename> and dpmhost.example.com as needed). The DPM_XROOTD_FEDREDIRS variable is a space separated list, add the above value as an item if you are joining more than one federation. If you would like to setup sending monitoring data to the central ATLAS collecting facility also set the monitoring directives described in the section below. After all the yaim configuration files are setup run yaim as usual on the head node. Also run yaim on the disk only nodes, unless dpm-xrootd has already been setup there. There is no federation specific configuration on the disk only nodes.

And specific instructions for DPM sites
UK .. and now IT cloud; later FR

CMS federation

David Smith

To join the CMS xroot federation you need to:

```
(i) choose the appropriate regional redirector, e.g. xrootd.ba.infn.it (EU).
(ii) install the CMS Trivial File Catalogue name2name library, e.g. the latest xrootd-cmstfc package from:
http://repo.grid.iu.edu/osg-contrib/x86_64/
The xrootd-cmstfc package may require installation of xerces-c to satisfy the rpm dependencies, e.g. "yum instal
(iii) install the storage.xml file for your site in /etc/xrootd/storage.xml (available from $VO_CMS_SW_DIR/SITEC
(iv) Find the appropriate "protocol" to set. It is "direct" in the example below. Most sites will use be the sam
(v) set yaim variables e.g.

DPM_XROOTD_FEDREDIRS="xrootd.ba.infn.it:1094:1213,cms,/store"
DPM_XROOTD_FED_CMS_NAMELIBPFX="/dpm/<sitename>/home/cms"
DPM_XROOTD_FED_CMS_NAMELIB="libXrdCmsTfc.so file:/etc/xrootd/storage.xml?protocol=direct"
```

The DPM_XROOTD_FEDREDIRS variable is a space separated list, add the above value as an item if you are joining more than one federation. After all the yaim configuration files are setup run yaim as usual on the head node. Also run yaim on the disk only nodes, unless dpm-xrootd has already been setup there. There is no federation specific configuration on the disk only nodes.

VO central monitoring

xrootd has the capability to periodically send reports of many aspects of the service and file accesses via udp packets. Some VOs would like to collect this information at a central point for analysis.

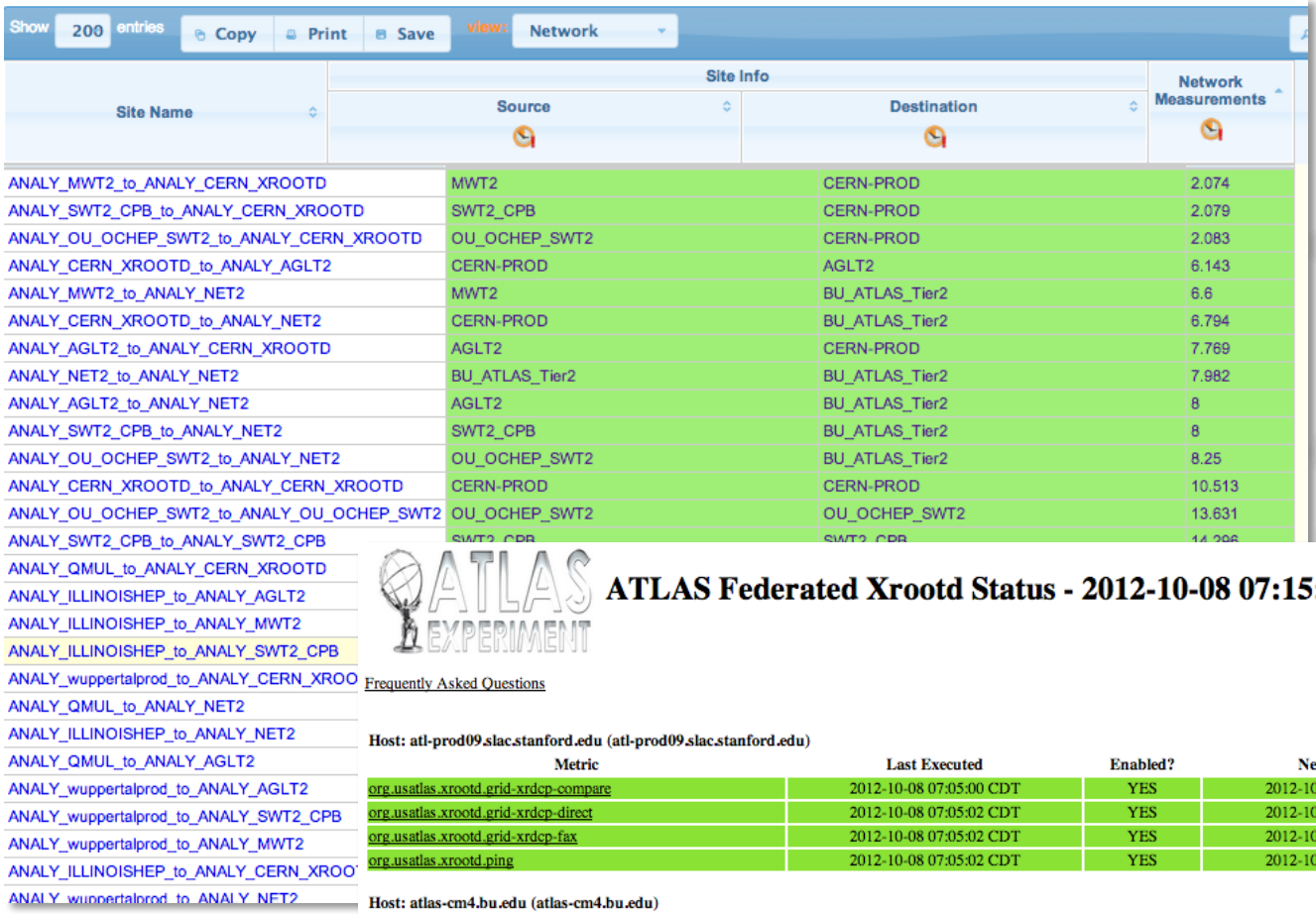
Sites are used in developing the federation

- Sites deploy a tandem of xrootd services
  - As easy as an Apache web server, in principle
- But the software, while a proven storage technology, requires additional development to become a federating technology:
  - Experiment-site specific file lookup service (i.e. N2N)
  - Customizations for backend storage types
  - Various 3rd party wide-area monitoring services (UCSD collector, ActiveMQ, Dashboards)
  - Security for read-only access: missing initially; still need gsi proxy validation
  - Standardizing monitoring metrics → further development
  - Status monitoring and alert systems for operations (RSV/Nagios)
  - New WLCG service definition (in GOCDB, OIM), similar to perfSONAR or Squid
  - Integration into ATLAS information system (AGIS)
  - Development in production & analysis systems: pilot & site movers
  - Accounting & caching will require more development, integration, testing, ..
- Good news many of these obstacles have been addressed in the R&D phase, and by CMS and ALICE before us.
- We benefit from vigorous developments by many groups working on various aspects of federation (AAA, XrootD & dCache teams, Dashboard…)

On-going work & issues

- SSB and WLCG transfer dashboard with cost matrix decision algorithm
- Xrootd instabilities seen in the UK cloud – perhaps related to N2N blocking at LFC
- FAX extensions to ATLAS information system AGIS
- Need new monitoring f-stream at all sites
- Stand-alone cmsd for dcache sites
- xrootd.org repository & EPEL policy (site guidance, esp. DPM)
- Several dCache specific issues, and many releases under test (1.9.12-22+, 2.2.4,…); f-stream, proper stat response and checksum support from dcache-xrootd doors
- Moving US sites to ro LFC
- Starting federating sites in Italy
- SLC6 issues – X509 and voms attribute checking
- Will update UDP collector service with f-stream format when available
- Functional testing probes & publishing into ActiveMQ and dashboards
- Monitoring will have to be validated at all stages
- FAX-enabled pilot site mover in production at several Tier 2s
- Documentation for users & site admins

Functional status & cost performance



Show 200 entries | Copy | Print | Save | View: Network

Site Name	Site Info: Source	Site Info: Destination	Network Measurements
ANALY_MWT2_to_ANALY_CERN_XROOTD	MWT2	CERN-PROD	2.074
ANALY_SWT2_CPB_to_ANALY_CERN_XROOTD	SWT2_CPB	CERN-PROD	2.079
ANALY_OU_OCHEP_SWT2_to_ANALY_CERN_XROOTD	OU_OCHEP_SWT2	CERN-PROD	2.083
ANALY_CERN_XROOTD_to_ANALY_AGLT2	CERN-PROD	AGLT2	6.143
ANALY_MWT2_to_ANALY_NET2	MWT2	BU_ATLAS_Tier2	6.6
ANALY_CERN_XROOTD_to_ANALY_NET2	CERN-PROD	BU_ATLAS_Tier2	6.794
ANALY_AGLT2_to_ANALY_CERN_XROOTD	AGLT2	CERN-PROD	7.769
ANALY_NET2_to_ANALY_NET2	BU_ATLAS_Tier2	BU_ATLAS_Tier2	7.982
ANALY_AGLT2_to_ANALY_NET2	AGLT2	BU_ATLAS_Tier2	8
ANALY_SWT2_CPB_to_ANALY_NET2	SWT2_CPB	BU_ATLAS_Tier2	8
ANALY_OU_OCHEP_SWT2_to_ANALY_NET2	OU_OCHEP_SWT2	BU_ATLAS_Tier2	8.25
ANALY_CERN_XROOTD_to_ANALY_CERN_XROOTD	CERN-PROD	CERN-PROD	10.513
ANALY_OU_OCHEP_SWT2_to_ANALY_OU_OCHEP_SWT2	OU_OCHEP_SWT2	OU_OCHEP_SWT2	13.631
ANALY_SWT2_CPB_to_ANALY_SWT2_CPB	SWT2_CPB		14.296
ANALY_QMUL_to_ANALY_CERN_XROOTD			
ANALY_ILLINOISHEP_to_ANALY_AGLT2			
ANALY_ILLINOISHEP_to_ANALY_MWT2			
ANALY_ILLINOISHEP_to_ANALY_SWT2_CPB			
ANALY_wuppertalprod_to_ANALY_CERN_XROO			
ANALY_QMUL_to_ANALY_NET2			
ANALY_ILLINOISHEP_to_ANALY_NET2			
ANALY_QMUL_to_ANALY_AGLT2			
ANALY_wuppertalprod_to_ANALY_AGLT2			
ANALY_wuppertalprod_to_ANALY_SWT2_CPB			
ANALY_wuppertalprod_to_ANALY_MWT2			
ANALY_ILLINOISHEP_to_ANALY_CERN_XROO			
ANALY_wuppertalprod_to_ANALY_NET2			

ATLAS EXPERIMENT

ATLAS Federated Xrootd Status - 2012-10-08 07:15:51

Frequently Asked Questions

Host: atl-prod09.slac.stanford.edu (atl-prod09.slac.stanford.edu)

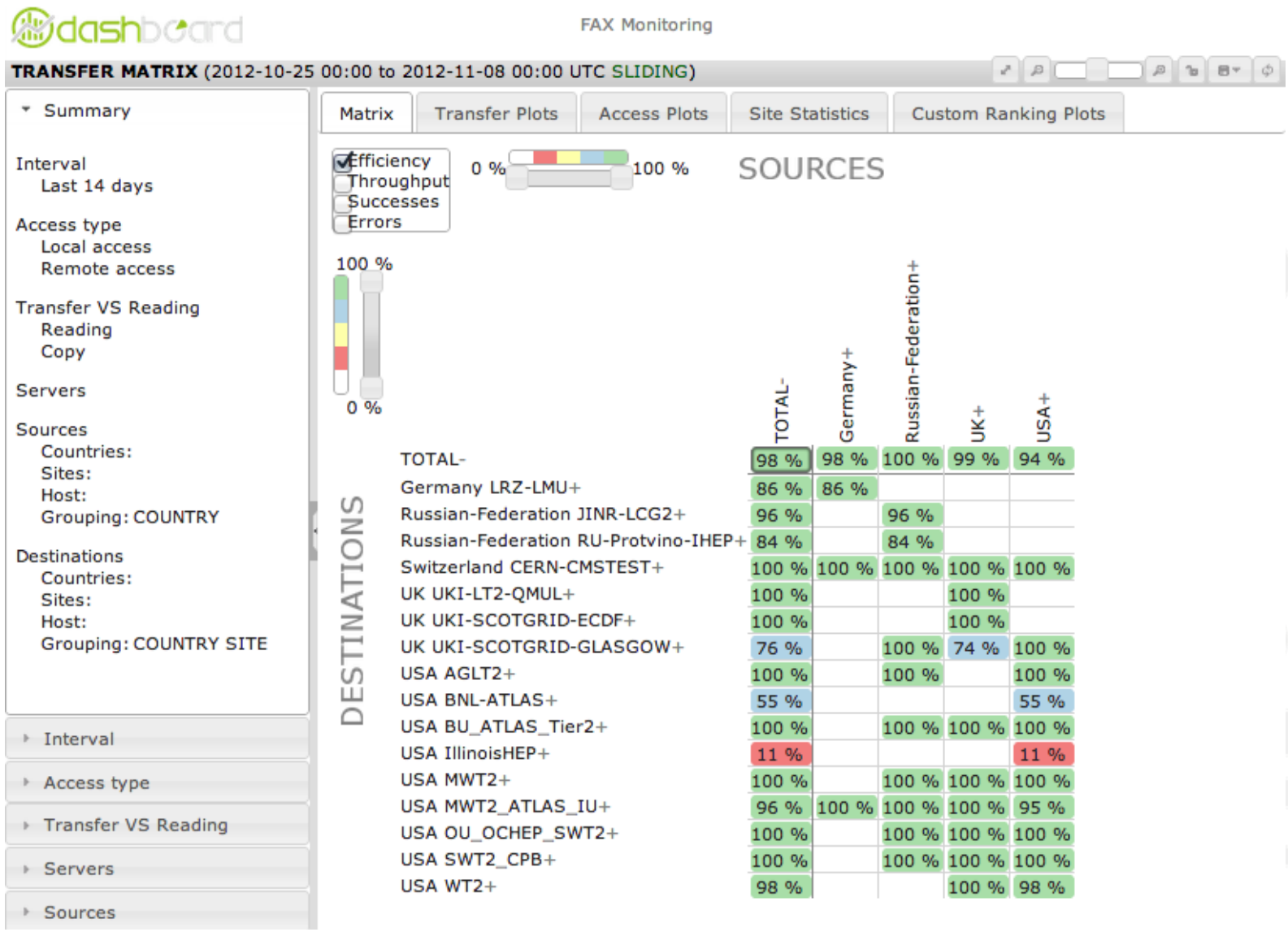
Metric	Last Executed	Enabled?	Next Run Time	Status
org.usatlas.xrootd.grid-xrdcp-compare	2012-10-08 07:05:00 CDT	YES	2012-10-08 07:20:00 CDT	OK
org.usatlas.xrootd.grid-xrdcp-direct	2012-10-08 07:05:02 CDT	YES	2012-10-08 07:20:00 CDT	OK
org.usatlas.xrootd.grid-xrdcp-fax	2012-10-08 07:05:02 CDT	YES	2012-10-08 07:20:00 CDT	OK
org.usatlas.xrootd.ping	2012-10-08 07:05:02 CDT	YES	2012-10-08 07:20:00 CDT	OK

Host: atlas-cm4.bu.edu (atlas-cm4.bu.edu)

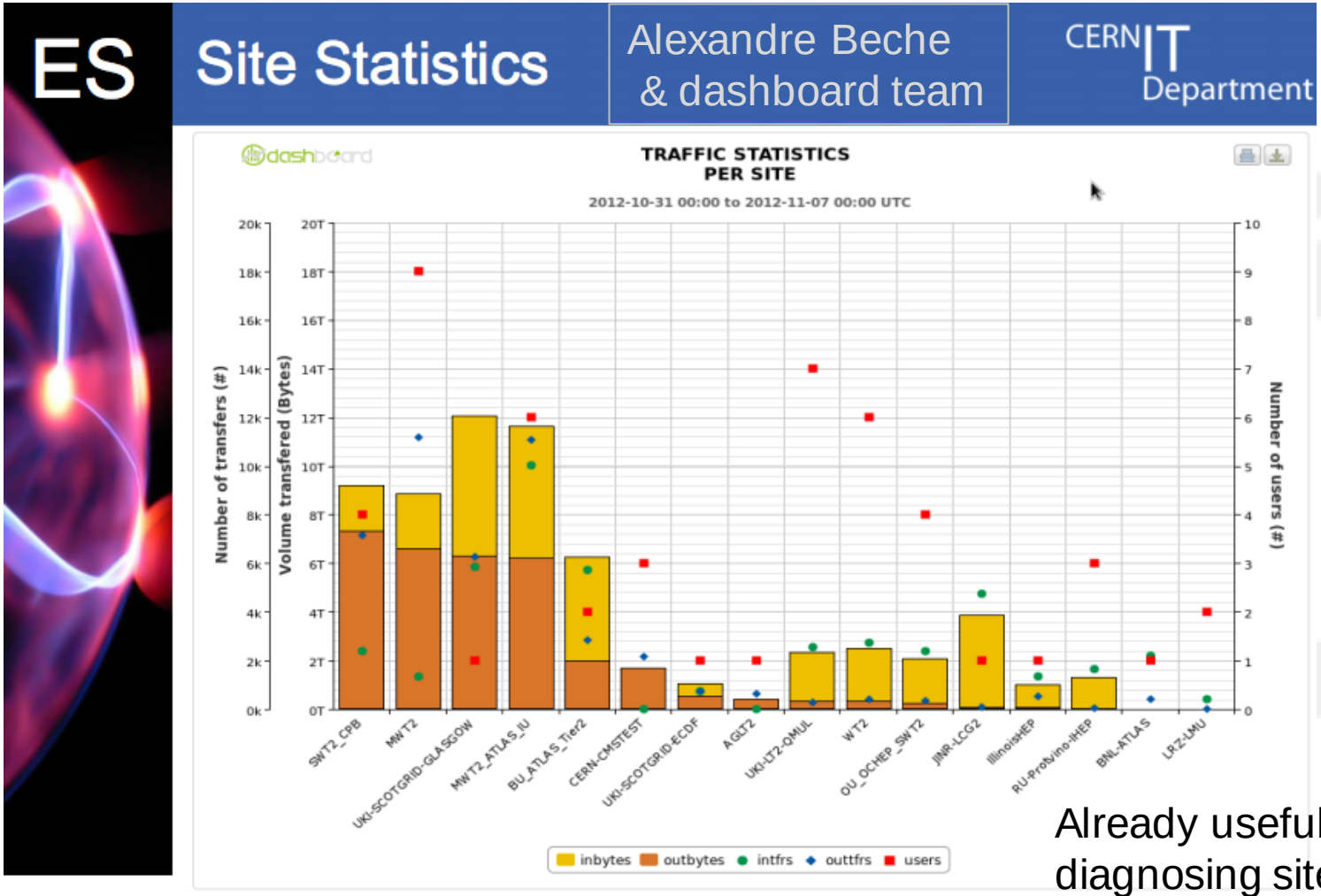
Metric	Last Executed	Enabled?	Next Run Time	Status
org.usatlas.xrootd.grid-xrdcp-compare	2012-10-08 07:05:01 CDT	YES	2012-10-08 07:20:00 CDT	OK
org.usatlas.xrootd.grid-xrdcp-direct	2012-10-08 07:05:01 CDT	YES	2012-10-08 07:20:00 CDT	OK
org.usatlas.xrootd.grid-xrdcp-fax	2012-10-08 07:05:01 CDT	YES	2012-10-08 07:20:00 CDT	OK
org.usatlas.xrootd.ping	2012-10-08 07:05:01 CDT	YES	2012-10-08 07:20:00 CDT	OK

There are many more components as discussed at the Lyon storage federations workshop in September

FAX dashboard – sites transfer matrix



FAX dashboard – prototyping extensions to give a better picture of site access



CERN IT Department
CH-1211 Geneva 23
Switzerland
www.cern.ch/it

07-11-2012 ***Alexandre Beche***

Already useful for diagnosing site config issues

9

Role of WLCG groups

- Serve as point of coordination for the common elements for the basic services that have emerged from the R&D programs
- Help drive requirements and priorities and liaison with software providing groups
- Help with packaging, deployment configurations, documentation, site support
- Coordinate extensions (protocols, caching) and drive consistency in the architecture
- Benefit, informed by related groups (e.g. networking)
- Liaise to sites through well-established mechanisms

Summary

- From the point of view of sites – we are engaging sites through cloud-region contacts and support teams to help identify and develop the infrastructure
- This activity drives a number of development and integration tasks for needed components (N2N interface to storage, monitoring providers, etc)
- As these are general issues and have natural overlap with CMS federating services it makes sense to continue fruitful collaboration
- WLCG working groups should provide a point of coordination and repository of expertise and operational support, and a context for evolving systems forward