

‘data-access’ pre-GDB Summary

(hastily prepared - apologies for omissions...)

Also see Michel’s excellent notes at
[https://twiki.cern.ch/twiki/bin/view/LCG/
GDBMeetingNotes20140513](https://twiki.cern.ch/twiki/bin/view/LCG/GDBMeetingNotes20140513)

Wahid Bhimji

14 May 2014



pre-GDB (Data access)






Tuesday, 13 May 2014 from **11:00** to **17:25** (Europe/Zurich)
at **CERN (31-S-028)**

[Manage](#) ▾

Description Covering local and remote data access including data federations: interesting studies, technologies and expectations...

Video Services Vidyo public room : pre-GDB__Data_access_ [More Info](#) | [Join Now!](#) | [Connect 31-S-028](#)

Tuesday, 13 May 2014

- | | | |
|---------------|--|---|
| 11:00 - 11:10 | Intro 10' | ▾ |
| | Speaker: Wahid Bhimji (University of Edinburgh (GB)) | |
| 11:10 - 11:30 | Federation workshop summary 20' | ▾ |
| | Speaker: Andrew Bohdan Hanushevsky (SLAC National Accelerator Laboratory (US)) | |
| | Material: Slides   | |
| 11:30 - 11:50 | Monitoring 20' | ▾ |
| | Speakers: Alexandre Beche (CERN), Dr. Domenico Giordano (CERN) | |
| 12:00 - 12:30 | Data access analysis 30' | ▾ |
| | Speakers: Christian Nieke (Brunswick Technical University (DE)), Matevz Tadel (Univ. of California San Diego (US)), Valentina Mancinelli (Universita e INFN (IT)), Nicolo Magini (CERN) | |
| | <hr/> | |
| | Data access - from infrastructure point of view 15' | ▾ |
| | Speaker: Christian Nieke (Brunswick Technical University (DE)) | |
| | Material: Slides   | |
| | <hr/> | |
| | Data access - from experiment point of view 15' | ▾ |
| | Speaker: Nicolo Magini (CERN) | |
| | <hr/> | |
| 12:50 - 14:00 | Lunch | |
| 14:00 - 14:20 | CMS plans expectations on sites 20' | ▾ |
| | Speaker: Kenneth Bloom (University of Nebraska (US)) | |
| 14:20 - 14:40 | Alice plans expectations on sites 20' | ▾ |
| | Speaker: Costin Grigoras (CERN) | |
| 14:40 - 15:00 | LHCb plans expectations on sites 20' | ▾ |
| 15:00 - 15:20 | ATLAS plans & expectations on sites 20' | ▾ |
| | Speaker: Robert William Gardner Jr (University of Chicago (US)) | |
| | Material: Slides  | |
| 15:25 - 15:45 | Tea | |
| 15:45 - 16:05 | German sites perspectives and plans 20' | ▾ |
| | Speakers: Guenter Duckeck (Ludwig-Maximilians-Univ. Muenchen (DE)), Guenter Duckeck (Experimentalphysik-Fakultaet fuer Physik-Ludwig-Maximilians-Uni) | |
| 16:05 - 16:25 | Dynamic federations and http plugin for xrootd 20' | ▾ |
| | Speaker: Fabrizio Furano (CERN) | |
| 16:25 - 16:45 | ATLAS plans for Http/Dav 20' | ▾ |
| | Speaker: Cedric Serfon (CERN) | |
| 16:45 - 17:05 | Root I/O - status & plans 20' | ▾ |
| | Speaker: Philippe Canal (Fermi National Accelerator Lab. (US)) | |

- ❖ Scope of this meeting was on **data-access**: including local and WAN (federation or otherwise).



pre-GDB (Data access)

Tuesday, 13 May 2014 from **11:00** to **17:25** (Europe/Zurich)
at **CERN (31-S-028)**

Manage ▾

Description Covering local and remote data access including data federations: interesting studies, technologies and expectations...

Video Services Vidyo public room : pre-GDB__Data_access_ [More Info](#) | [Join Now!](#) | [Connect 31-S-028](#)

Tuesday, 13 May 2014

11:00 - 11:10	Intro 10' Speaker: Wahid Bhimji (University of Edinburgh (GB)) Material: Slides	▾
11:10 - 11:30	Federation workshop summary 20' Speaker: Andrew Bohdan Hanushevsky (SLAC National Accelerator Laboratory (US)) Material: Slides	▾
11:30 - 11:50	Monitoring 20' Speakers: Alexandre Beche (CERN), Dr. Domenico Giordano (CERN) Material: Slides	▾
12:00 - 12:30	Data access analysis 30' Speakers: Christian Nieke (Brunswick Technical University (DE)), Matevz Tadel (Univ. of California San Diego (US)), Valentina Mancinelli (Universita e INFN (IT)), Nicolo Magini (CERN)	▾
..... Data access - from infrastructure point of view 15' Speaker: Christian Nieke (Brunswick Technical University (DE)) Material: Slides		▾
..... Data access - from experiment point of view 15' Speaker: Nicolo Magini (CERN) Material: Slides		▾
.....		

Intro 10'

Speaker:

Wahid Bhimji (University of Edinburgh (GB))

❖ Some suggested key questions :

- ❖ Do we understand our data access well enough? Are I/O performance wins out there?
- ❖ Data federations are in production and offer increased flexibility and resource usage:
 - ❖ But do we have everything needed to work at scale?
 - ❖ Do sites need to plan or provision more?
 - ❖ How do we use this software: monitoring; caches etc.?
- ❖ Are we employing solutions compatible with wider communities? Should we? (c.f. Big Data etc.)
- ❖ Is our protocol zoo growing (http / xrootd / (rfio) / gridftp etc..)? Are there paths to simplification?

- ❖ Also one slide on the progress on allowing WLCG Tier 2 disk-only sites to not have SRM in Run2 - look at if you care

Federation workshop summary 20'

Speaker:

Andrew Bohdan Hanushevsky (SLAC National Accelerator Laboratory (US))

- ❖ <https://indico.fnal.gov/conferenceDisplay.py?confId=7207>
- ❖ Experiment Reports; Australia and UK site perspectives
- ❖ Monitoring ; Scale testing;
- ❖ Developments: Panda, HTTP/Dav, Caching proxy
- ❖ Many items summarised and updated in this meeting
- ❖ From discussion: xrootd4 in RC - progressing to available soon..

Monitoring 20'

Speakers:

Alexandre Beche (CERN), Dr. Domenico Giordano (CERN)

- ❖ Comprehensive monitoring system built for xrootd (plans to extend / adapt for http) . Valuable info - one example below

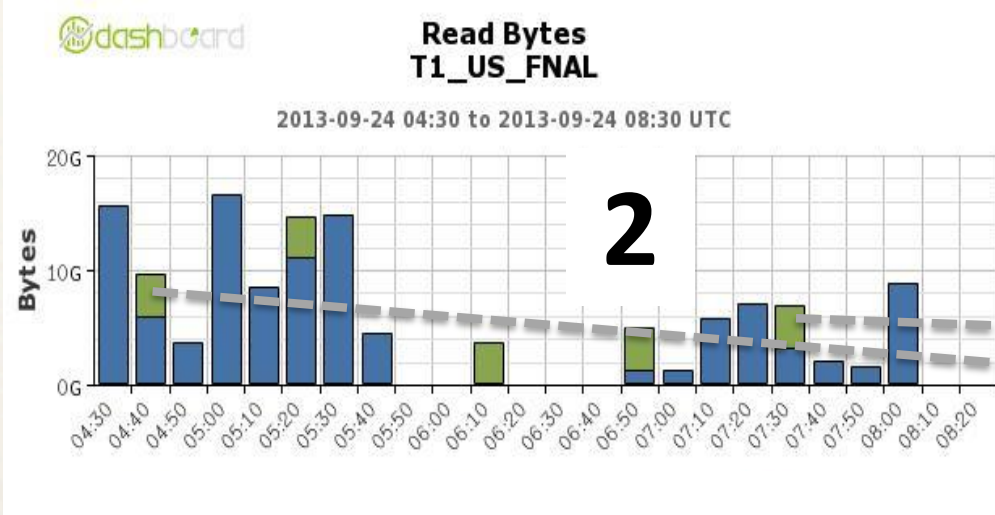
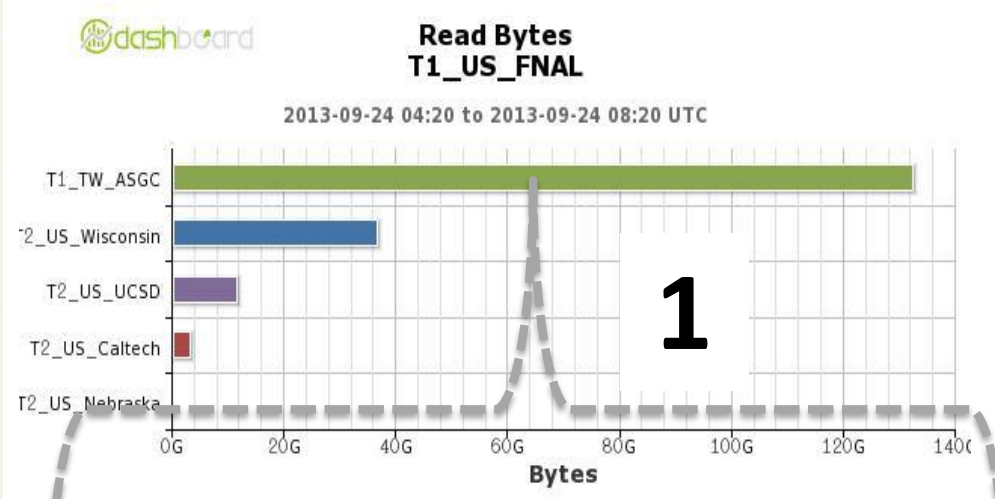
- ❖ Not (yet) used as much as it could.

- ❖ Thanks to Alexandre (who is moving on)

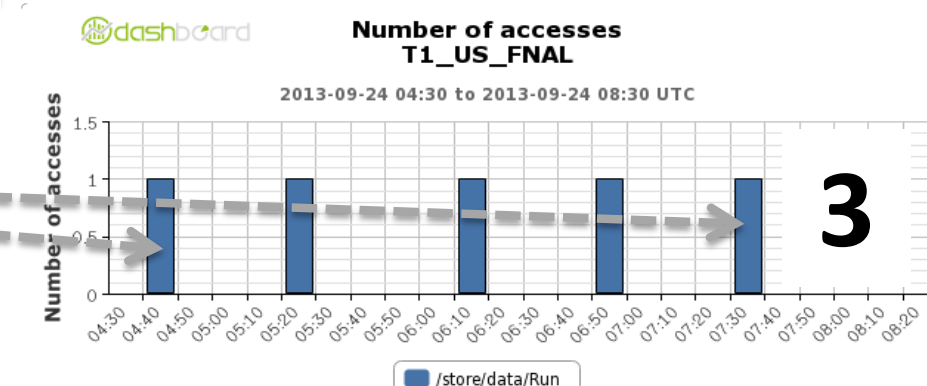
- ❖ From discussion: need for VO filtering at xrootd source

Use case example

Understand site access patterns



1. Which sites are reading from FNAL
2. Zoom to a specific site to understand which users are reading
3. Understand which files are read by a user



Data access - from infrastructure point of view 15'

Speaker:

Christian Nieke (Brunswick Technical University (DE))

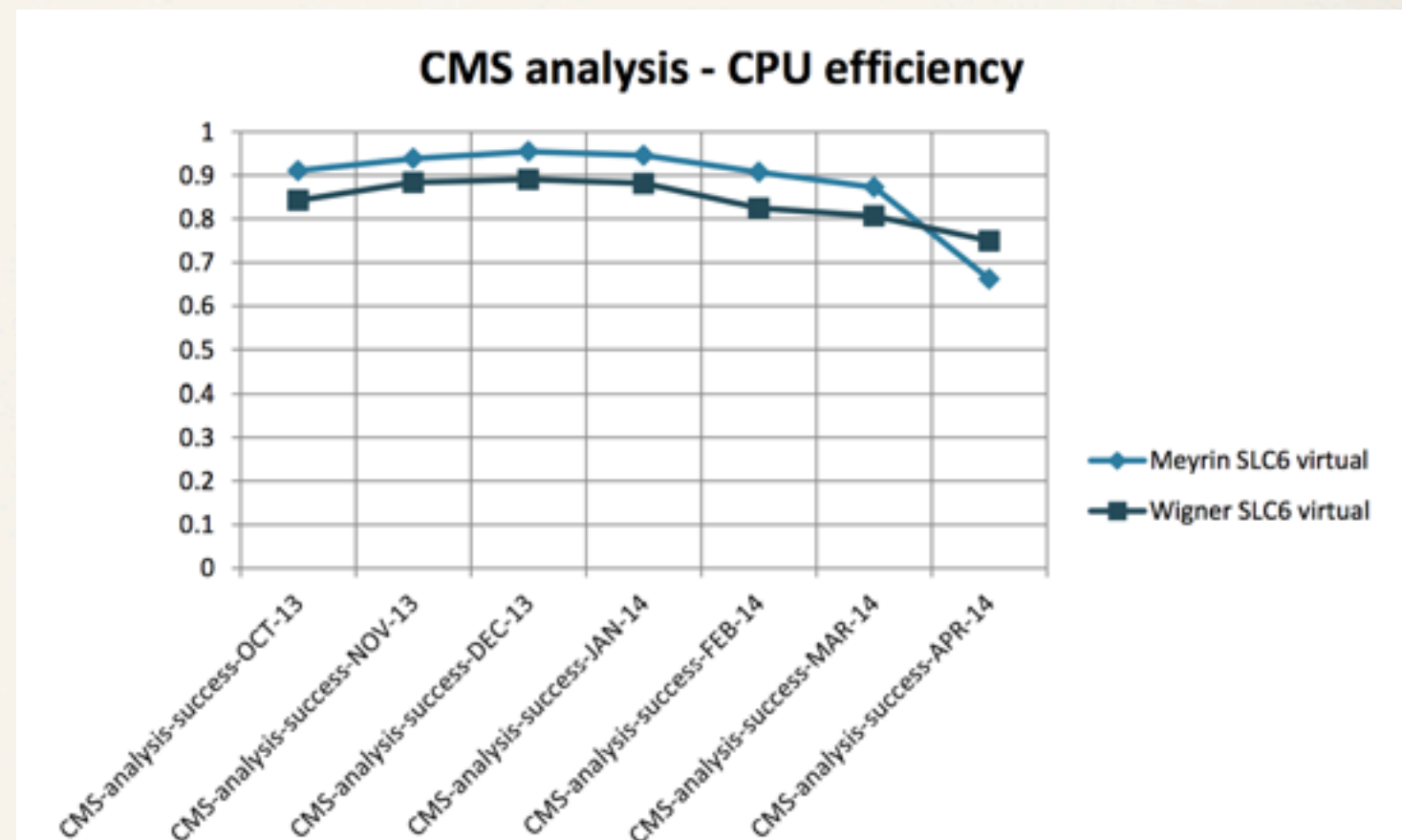
- ❖ To “Understand; Improve; Predict..”
- ❖ Sources from EOS logs, LSF, dashboard
- ❖ “Semi-automatic” detection of performance anomalies
- ❖ Metric definition needs to be appropriate (e.g. not always CPU “eff.”)
- ❖ Proposal (need) to fill “app info” field in xrootd records (to allow matching of workload)

















Data access - from experiment point of view 15'

Speaker:

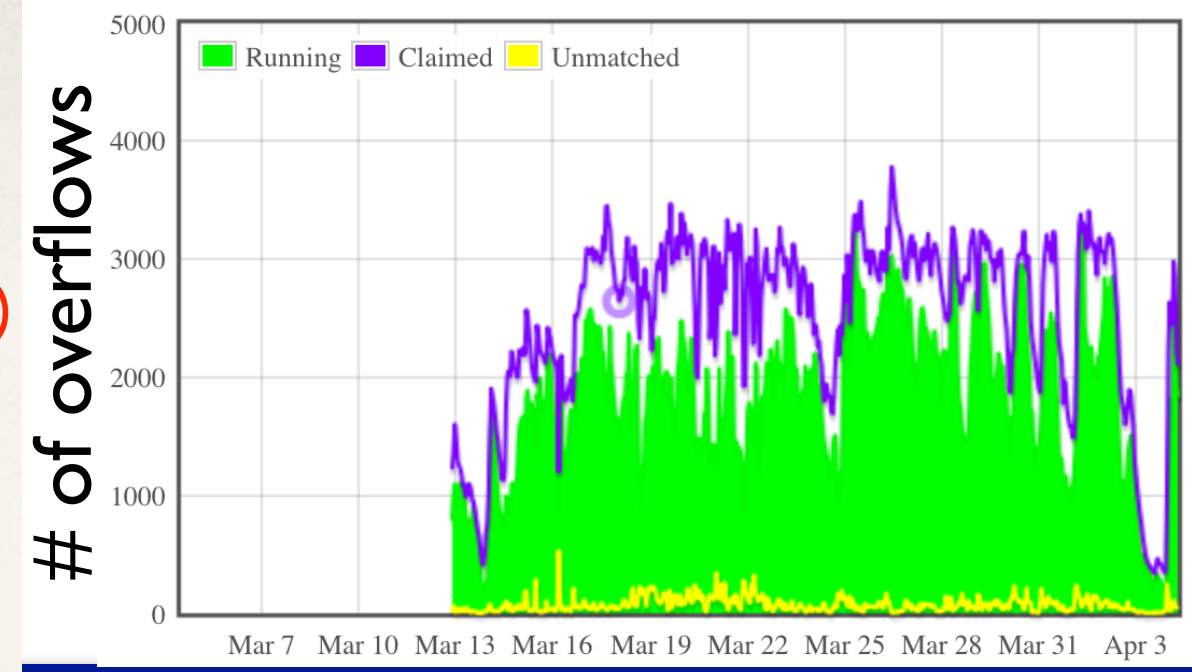
Nicolo Magini (CERN)

- ❖ Meyrin <-> Wigner
- ❖ Tests by experiments
- ❖ Some differences but can be down to cpu / VM / OS mix as well as latency so further investigation



14:00 - 14:20	CMS plans expectations on sites 20' Speaker: Kenneth Bloom (University of Nebraska (US)) Material: Slides 	
14:20 - 14:40	Alice plans expectations on sites 20' Speaker: Costin Grigoras (CERN) Material: Slides  	
14:40 - 15:00	LHCb plans expectations on sites 20'	
15:00 - 15:20	ATLAS plans & expectations on sites 20' Speaker: Robert William Gardner Jr (University of Chicago (US)) Material: Slides 	
15:25 - 15:45	Tea	
15:45 - 16:05	German sites perspectives and plans 20' Speakers: Guenter Duckeck (Ludwig-Maximilians-Univ. Muenchen (DE)), Guenter Duckeck (Experimentalphysik-Fakultaet fuer Physik-Ludwig-Maximilians-Uni) Material: Slides 	
16:05 - 16:25	Dynamic federations and http plugin for xrootd 20' Speaker: Fabrizio Furano (CERN) Material: Slides 	
16:25 - 16:45	ATLAS plans for Http/Dav 20' Speaker: Cedric Serfon (CERN) Material: Slides 	
16:45 - 17:05	Root I/O - status & plans 20' Speaker: Philippe Canal (Fermi National Accelerator Lab. (US)) Material: Slides 	

CMS plans expectations on sites 20'
Speaker:
Kenneth Bloom (University of Nebraska (US))



- ❖ CMS data federation (AAA) working and widely used
 - ❖ For fallback, planned remote work, opportunistic etc.
- ❖ Scale tests of infrastructure: some observed limits on some sites or systems - need to be understood.
- ❖ Want remaining few sites deployed - and tuned...
- ❖ From discussion: Throttling plugin for xrootd - will be in main release by July - sites should use that if they need

Alice plans expectations on sites 20'

Speaker:

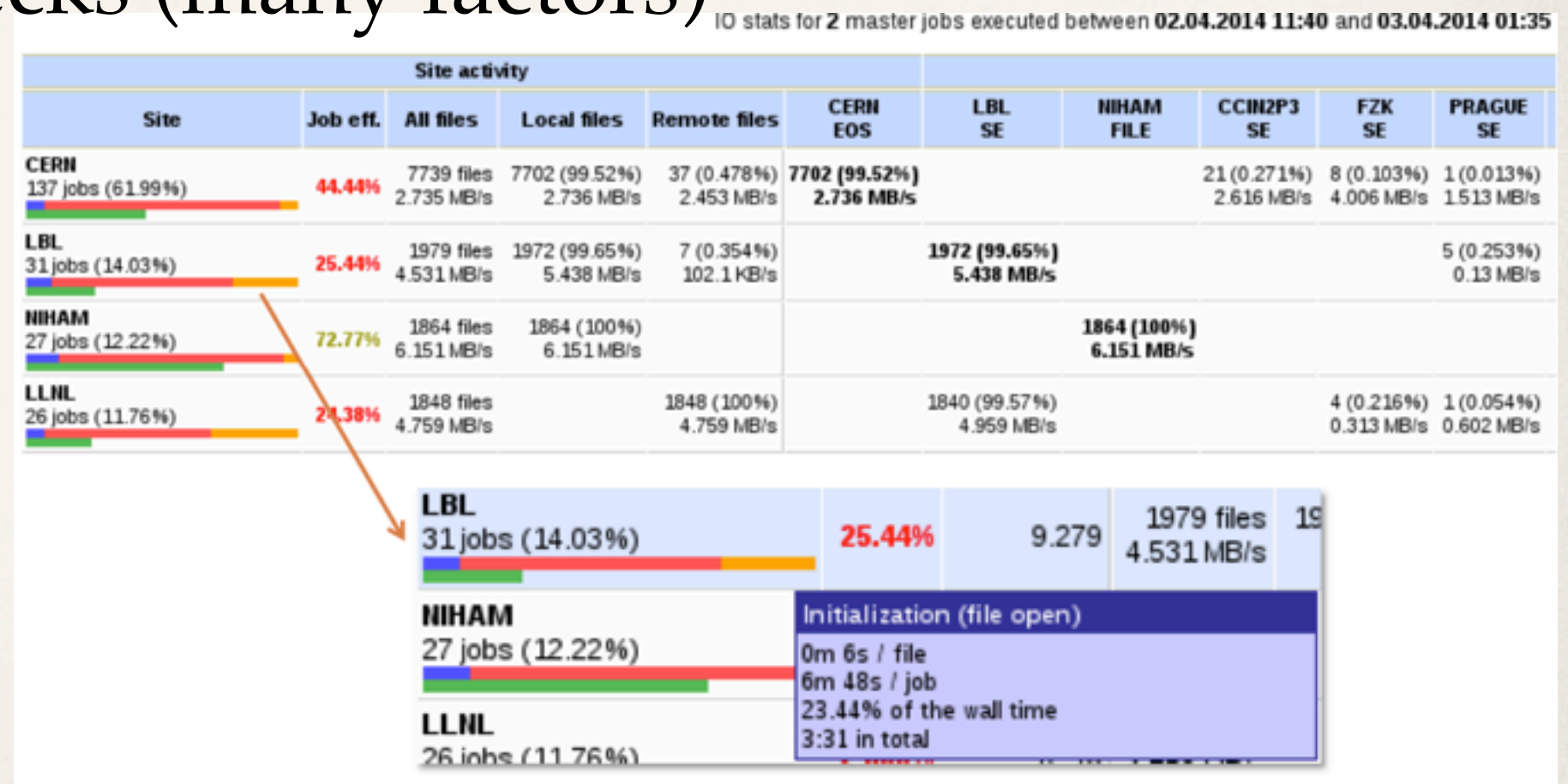
Costin Grigoras (CERN)

- ❖ Remote reading for “urgent” tasks - to get job done.

Sorted list of replicas

- ❖ Deep understanding/ monitoring of job efficiency and site bottlenecks (many factors)

- ❖ Average analysis requires 2MB/s/core



- ❖ “For newly deployed storage we plan to use EOS”

ATLAS plans & expectations on sites 20'

Speaker:
Robert William Gardner Jr (University of Chicago (US))

Site	Jobs	Files(FAX)	Files(Local)	GB(FAX)	GB(Local)	Files/hr	Gbps
FR: ANALY_GRIF-LPNHE	10395	12105	37535	33107.6	46637.66	6.0	0.036
US: ANALY_BNL_LONG	8872	11343	9744	10307.4	9997.28	5.6	0.011
IT: ANALY_INFN-MILANO-ATLASC	7924	13433	45648	13853.1	43028.52	6.7	0.015
IT: ANALY_INFN-T1	7259	9718	27777	2575.64	8950.93	4.8	0.003
FR: ANALY_ROMANIA07	6063	12137	85078	15452.9	83675.55	6.0	0.017
DE: MPPMU	5938	23820	2119	13995.3	2639.46	11.8	0.015
DE: DESY-ZN	5825	16054	553	8052.63	732.89	8.0	0.009
IT: INFN-T1	5587	8873	12004	3619.25	8238.22	4.4	0.004
TW: ANALY_TAIWAN_PNFS_SL6	4999	5127	37594	506.51	27414.86	2.5	0.001

- ❖ ATLAS Federation (FAX) , in production and stable.
(Improvement with “Rucio” (no LFC lookup))
- ❖ Failover used widely (not causing a big WAN load).
- ❖ “Overflow” (rebrokering if needed to remote queue)

Recommendations to WLCG sites (1)

- In the Feb 2014 ATLAS S&C Week ADC Operations session it was agreed as policy that T1s and T2Ds are to offer XRootD & HTTP/WebDAV access to storage, where the storage technology allows
 - ADC furthermore asks and encourages sites not yet in the FAX federation to take the modest additional step beyond supporting XRootD of joining FAX
- We intend to demonstrate WAN data access at scale (<~10% of data access) in DC14
- Consequently, timescale for installation is in time for pre-DC14 testing

- now working in testing

Recommendations to WLCG sites (2)

- Priorities (in order)
 - Enable XRootD data access
 - Enable FAX
 - Enable HTTP/WebDAV data access
 - More details in Cedric’s talk later
- If there are problems with either XRootD or HTTP/WebDAV we encourage sites to contact

LHCb Federation plans / feedback (email from Philippe Charpentier)

1. For production jobs, we always download the input dataset to the WN as these jobs are CPU-bound.
2. For working group or user analysis:
 - we access files using xroot (at all of our sites)
 - jobs are brokered to a site when the full dataset is supposed to be present (according to our FC)
 - we create an XML file catalog in the job that contains all replicas of all files, starting with the local replica
 - Gaudi is dereferencing the LFN using the XML: catalog, and tries to open the replicas in turn until successful.

Therefore in summary we access files on the WAN only in case a file is not reachable locally due to any reason (file actually missing, disk server down, overloaded....)
3. Interactive usage: users may access files from anywhere using xroot. Currently they need to specify from which SE, and we are going to implement a client that will find out the most appropriate location according to the FC (again with failover if the file cannot be accessed)

German sites perspectives and plans 20'

Speakers:

Guenter Duckeck (Ludwig-Maximilians-Univ. Muenchen (DE))

Summary

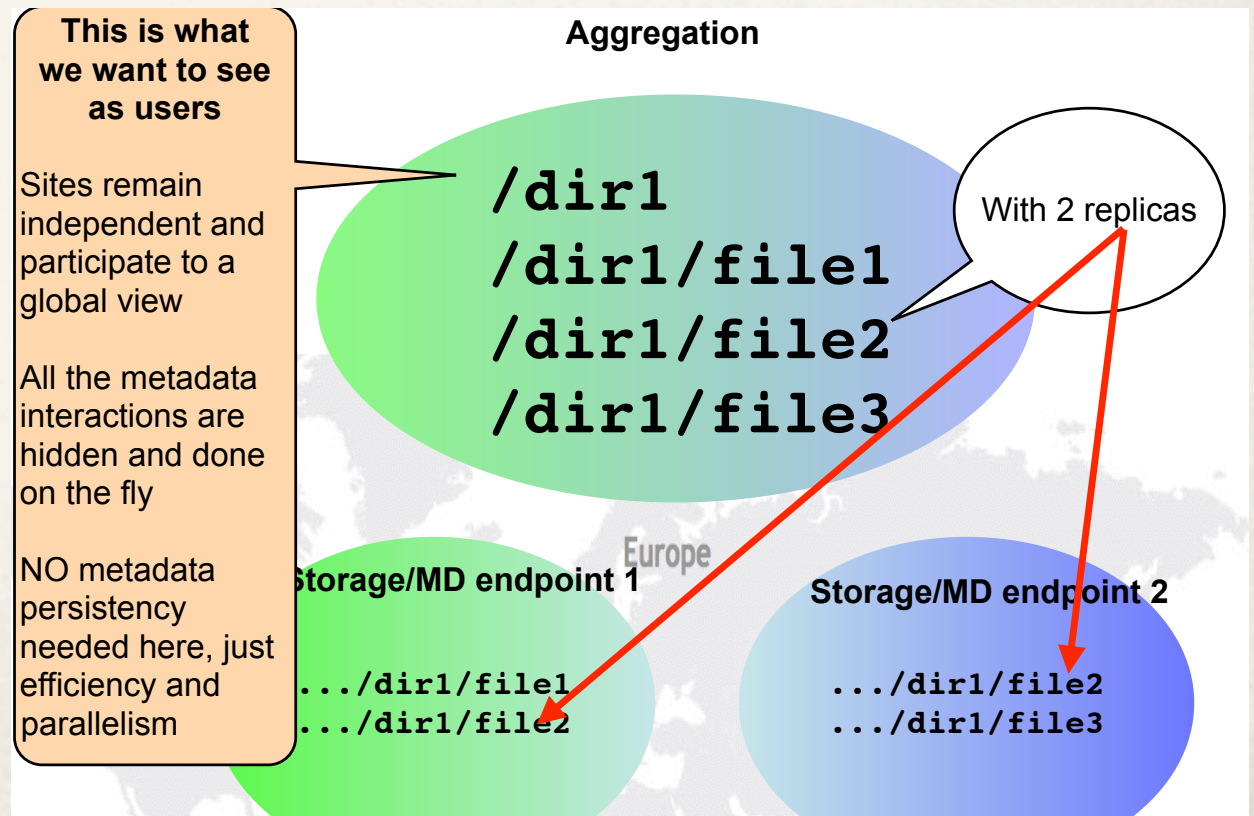
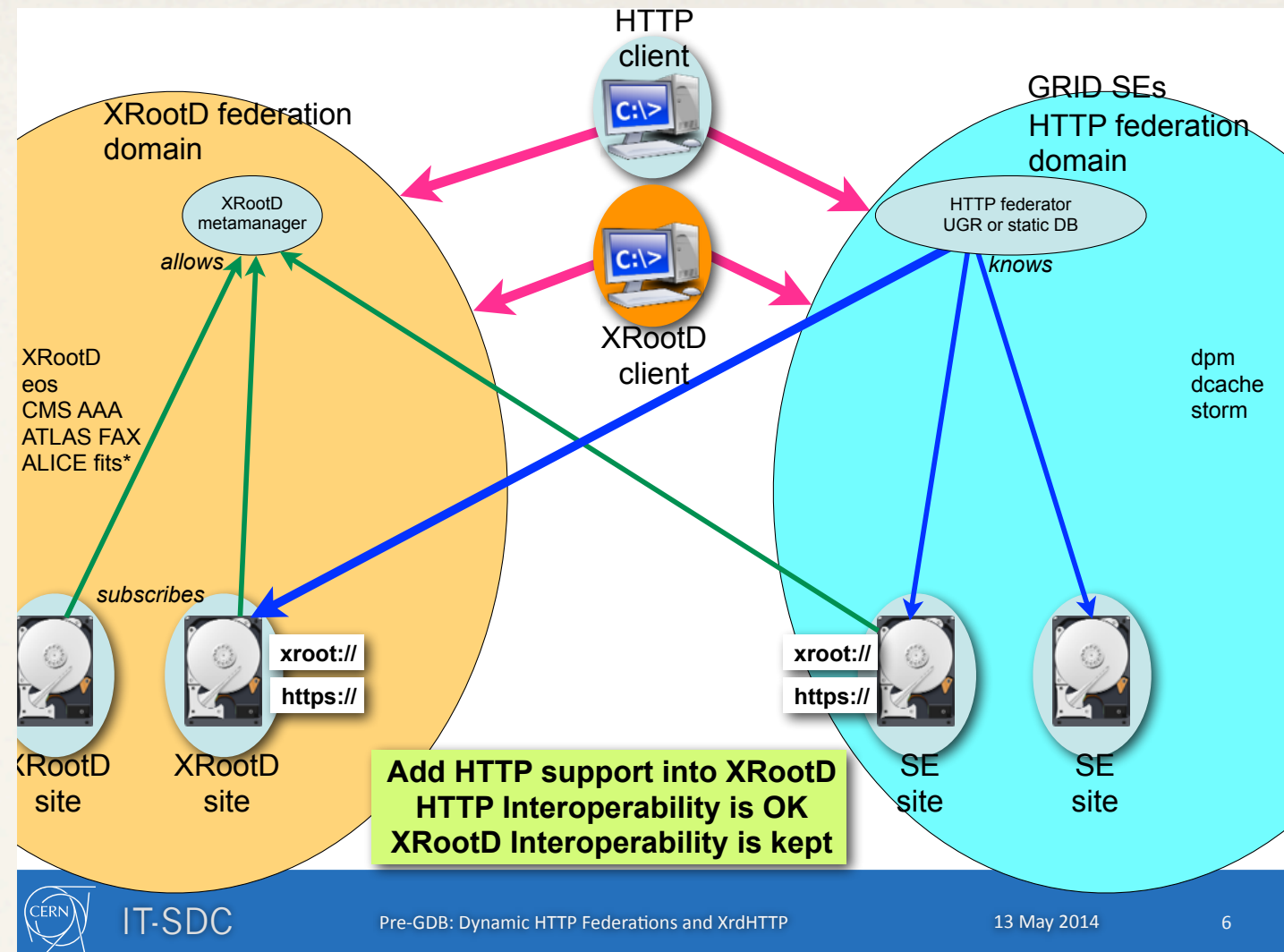
- Good experience with LAN direct IO at ATLAS-De sites since many years
- reducing protocol zoo and/or use of common std desirable
 - not easy to achieve in practice
- WAN/remote IO
 - FAX/xrootd largely deployed at DE sites
 - performance and stability looks promising
 - http/Webdav/Davix
 - in use for simulation input download (aria2c) at few sites
 - still testing for analysis direct IO
- CMS:
 - AAA in routine use at CMS DE sites

Dynamic federations and http plugin for xrootd 20'

Speaker:

Fabrizio Furano (CERN)

- ✧ XrdHTTP is done (in Xrootd4)
- ✧ Easy http(s)/ dav for xrootd sites
- ✧ Performance of xrd
- ✧ Dynamic federations allows listing and replica finding - testbed: <http://federation.desy.de>



ATLAS plans for Http/Dav 20'

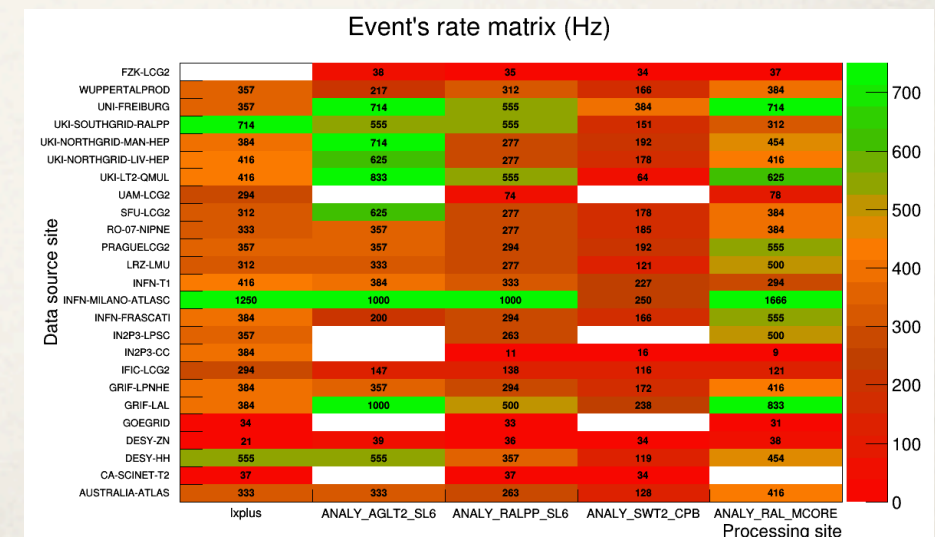
Speaker:

Cedric Serfon (CERN)

- ❖ Rucio uses DAV (where available):
 - ❖ Used already for renaming - now in FTs to inject/ delete.
 - ❖ Available at sites - but needs QoS to match srm etc.
 - ❖ User download via “rucio redirector” (random or geoip or selected) or Metalink server.
 - ❖ Direct access with Davix testing underway with some preliminary results.

Event rate matrix

- Disclaimer : Very preliminary with limited statistic and limited number of sites :



- TODO : Crosscheck with FAX results.

Root I/O - status & plans 20'

Speaker:

Philippe Canal (Fermi National Accelerator Lab. (US))

- ❖ TTreeCache configurable in environment
- ❖ “ROOT I/O is now thread friendly”
- ❖ Path to update ROOT IO for tomorrows need:
- ❖ ROOT I/O In Person Workshop coming up : June 25 at CERN: <http://indico.cern.ch/e/ROOT-IO-7>

Were my questions answered?

- ❖ We have lots of data to access on data-access performance
 - ❖ We still need to understand it and use it to improve
 - ❖ All the way from ROOT IO to site storage / network tuning
- ❖ Data federations are in production and various new use cases appearing
 - ❖ Expectation on remaining sites to enable
 - ❖ We have monitoring (need to look at it); we expect a few things (throttling plugin, VO filtering for monitoring etc.)
- ❖ Many interesting developments on http , brokering , caching etc
 - ❖ that can take this to new levels but also simplify and be used be used with other communities.