Computation Institute

FAX

status
developments
performance
future

Rob Gardner
Yang Wei
Andrew Hanushevsky
Ilija Vukotic

THE UNIVERSITY OF CHICAGO
efi.uchicago.edu
ci.uchicago.edu

Current Focus: RUCIO, RUCIO, RUCIO

FAX Site Certification Phase 1 Status

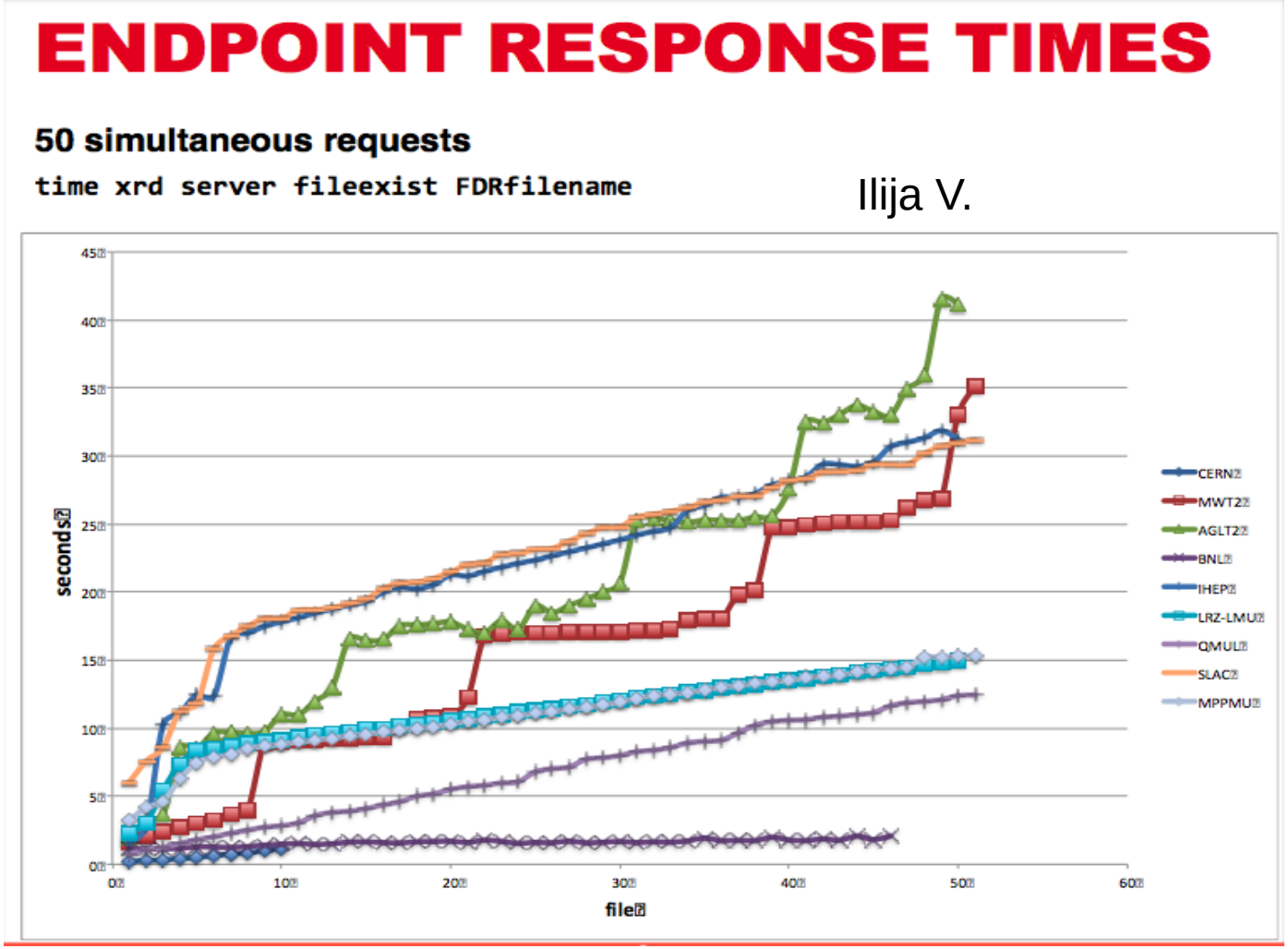
Notation: Completed | In progress | unknown | Not applicable

(click the X509 column head to visit the X509 migation twiki page)

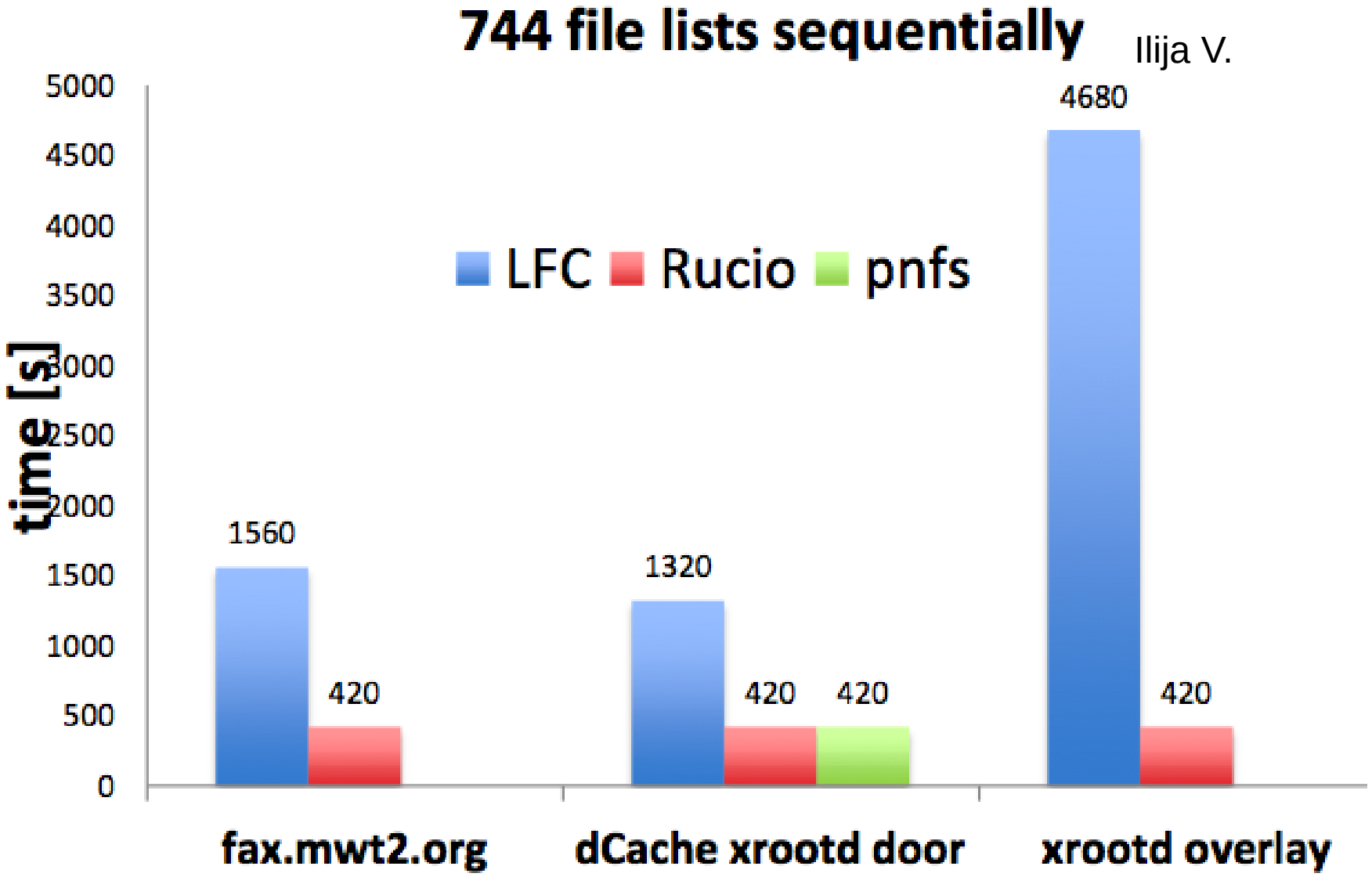
FAX site name	X509	Xrootd 3.3.3 Upgrade	Config Change	RUCIO N2N	Comments
FZK-LCG2					
IFAE					
PIC					
AGLT2					
BU_ATLAS_TIER2					
CERN-PROD					
LRZ-LMU					
MPPMU					
MWT2					
OU_OCHEP_SWT2					
SWT2_CPB					
UKI-LT2-QMUL					
UKI-NORTHGRID-LANCS-HEP					
UKI-NORTHGRID-MAN-HEP					
UKI-NORTHGRID-SHEF-HEP					
UKI-SCOTGRID-ECDF					
UKI-SCOTGRID-GLASGOW					
UKI-SOUTHGRID-CAM-HEP					
UKI-SOUTHGRID-OX-HEP					

Why RUCIO is important to FAX

- LFC based gLFN /atlas/dq2/… not scale well
- RUCIO based gLFN /atlas/rucio/scope:file



RUCIO gLFN improves performance



Status

Site Name	Direct	Upstream redirection	Downstream redirection	X509	Rucio N2N
UKI-NORTHGRID-LANCS-HEP	OK	OK	OK	On	OK
pic	OK	OK	OK	On	NoRucioNamesSupport
UKI-SCOTGRID-ECDF	OK	OK	OK	On	OK
UKI-NORTHGRID-LIV-HEP	OK	OK	OK	On	NoRucioNamesSupport
wuppertalprod	OK	OK	OK	Off	NoRucioNamesSupport
GRIF-LAL	OK	OK	OK	On	NoRucioNamesSupport
RU-Protvino-IHEP	OK	OK	OK	On	NoRucioNamesSupport
IN2P3-CPPM	OK	OK	OK	On	NoRucioNamesSupport
CERN-PROD	OK	OK	OK	On	OK
GRIF-IRFU	OK	OK	OK	On	NoRucioNamesSupport
BNL-ATLAS	OK	OK	OK	On	NoRucioNamesSupport
OU_OCHEP_SWT2	OK	OK	OK	On	OK
GRIF-LPNHE	OK	OK	OK	On	NoRucioNamesSupport
UNI-FREIBURG	OK	OK	OK	On	OK
MWT2	OK	OK	OK	On	OK
GoeGrid	OK	OK	OK	Off	NoRucioNamesSupport
JINR-LCG2	OK	OK	OK	On	OK
MPPMU	OK	OK	OK	On	OK
RAL-LCG2	OK	NoUpstreamRedirection	OK	On	NoRucioNamesSupport
INFN-FRASCATI	OK	OK	OK	On	NoRucioNamesSupport
DESY-ZN	OK	OK	OK	On	NoRucioNamesSupport
AGLT2	OK	OK	OK	On	OK
UNIGE-DPNC	OK	OK	OK	On	OK
praguelcg2	OK	OK	OK	On	NoRucioNamesSupport
INFN-NAPOLI-ATLAS	OK	OK	OK	On	NoRucioNamesSupport
UKI-SCOTGRID-GLASGOW	OK	OK	OK	On	OK
UKI-LT2-QMUL	OK	OK	OK	On	OK
ifae	OK	OK	OK	On	NoRucioNamesSupport
UKI-NORTHGRID-MAN-HEP	OK	OK	OK	On	NoRucioNamesSupport
UKI-SOUTHGRID-CAM-HEP	OK	OK	OK	On	OK
UKI-SOUTHGRID-OX-HEP	OK	OK	OK	On	OK
IN2P3-LAPP	OK	OK	OK	On	NoRucioNamesSupport
BU_ATLAS_Tier2	OK	OK	OK	On	OK
FZK-LCG2	OK	OK	OK	On	NoRucioNamesSupport
INFN-T1	OK	OK	OK	On	NoRucioNamesSupport
WT2	OK	OK	OK		OK
UKI-NORTHGRID-SHEF-HEP	OK	OK	OK		NoRucioNamesSupport
INFN-ROMA1	OK	OK	OK		NoRucioNamesSupport
LRZ-LMU	noDirect	NoUpstreamRedirection	NoFirstLevelRedirec		NoRucioNamesSupport
PSNC	OK	OK	NoFirstLevelRedirec		OK
Taiwan-LCG2	noDirect	NoUpstreamRedirection	NoFirstLevelRedirection	Off	NoRucioNamesSupport
IN2P3-LPSC	OK	OK	NoFirstLevelRedirection	On	NoRucioNamesSupport
CYFRONET-LCG2	OK	OK	NoFirstLevelRedirection	On	NoRucioNamesSupport
DESY-HH	OK	OK	NoFirstLevelRedirection	On	NoRucioNamesSupport
UKI-LT2-RHUL	noDirect	NoUpstreamRedirection	NoFirstLevelRedirection	Off	NoRucioNamesSupport
SWT2_CPB	noDirect	NoUpstreamRedirection	NoFirstLevelRedirection	Off	NoRucioNamesSupport

cmsd stability issue we are working on

- There will always be some “red” sites.
- <5% would be satisfactory.
- FAX is designed to be resilient to site failure

Need to update xrootd and n2n plugin
17 from 25 done

Status

- Stability is slowly improving
- New issues arose due to move to SL6
- Some issues giving false red addressed
- Moving away from LFC will make it even more stable
- Additional tools helping understand issues on the way

Other Activities

- Document Improvement
  - Simplify site storage integration instruction
  - Updated end user instruction for RUCIO
- Tools to enhance end-user experience
- Cost matrix
- Monitor and access info collection
- Panda job failover to FAX
- HC test
  - Jumpstart HC test for sites with RUCIO N2N

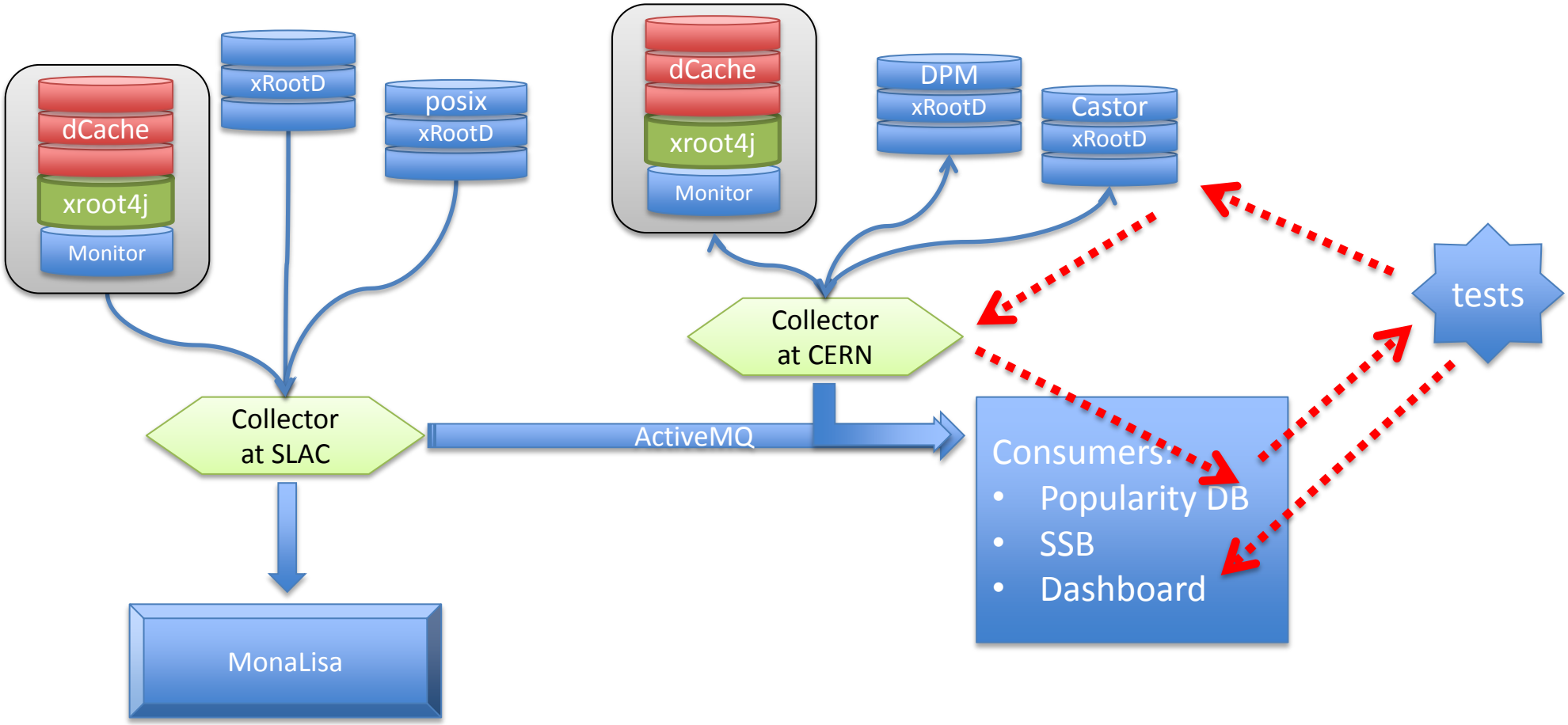
Simplify and Enhance End-User Experience

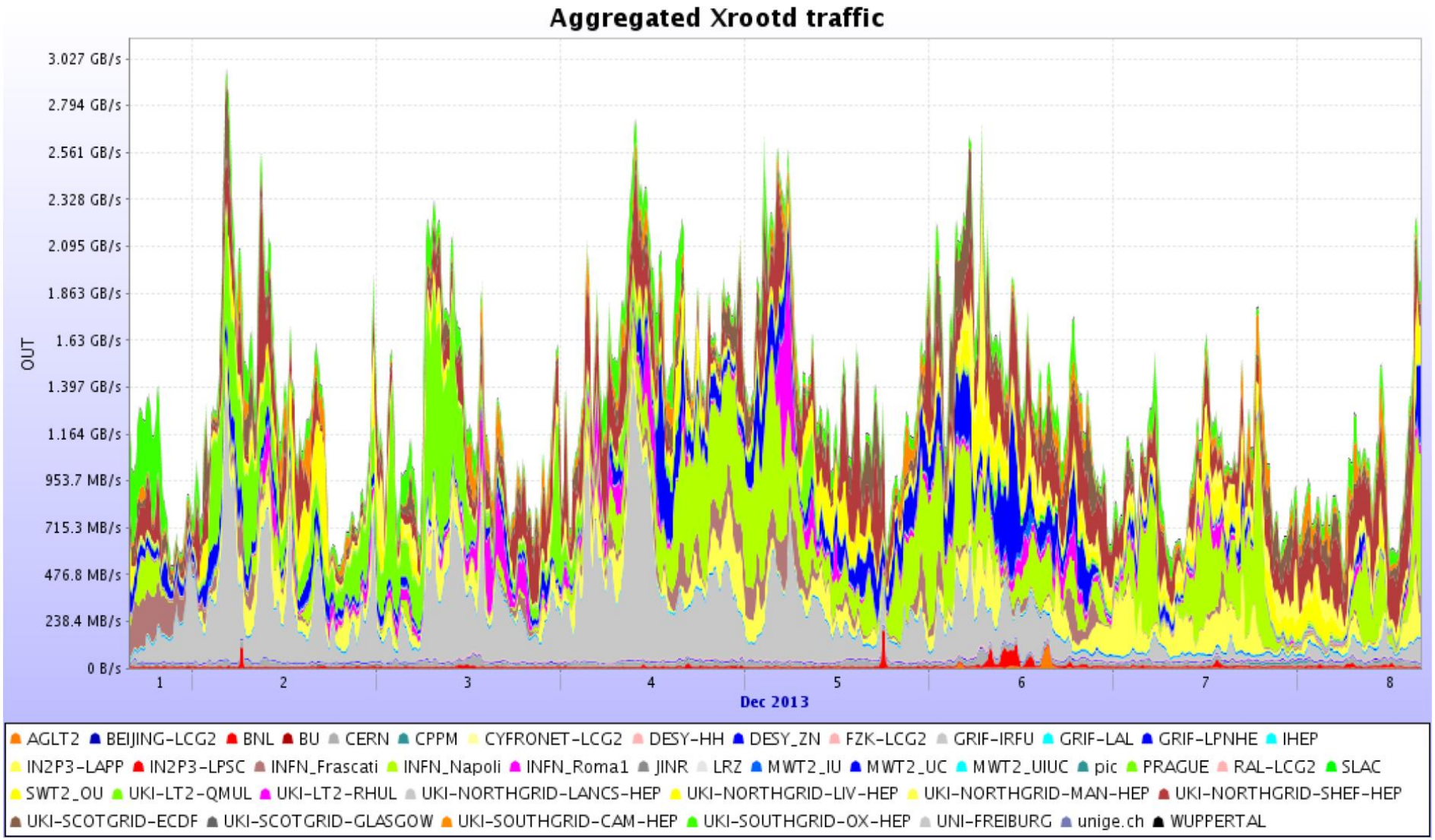
- localSetupFAX
  - Very important in improving an end-user experience. Big thanks to Asoka De Silva

- Sets up dq2-clients and grid middleware (emi-UI for 64-bit OS, gLite for 32-bit OS)
- Sets up xrootd (you can specify a different version from the default; see --help)
- Optionally setup ROOT (see --help)
- setup FAXTools; the tools are in the $PATH environment
  - isDSinFAX.py, fax_ls.py, setRedirector.sh, more to come
- determines and sets STORAGEPREFIX
  - To localredirector, if it can be resolved at the local domain name
  - If not to the geographically closest working endpoint
  - If not to glrd.usatlas.org

Monitoring and Access Info Collection

- Most of the monitoring system in place and functioning well
- New collector deployed at CERN. Now collecting only from INFN-T1. Will collect from all of EU endpoints.




Aggregated Xrootd traffic
3.027 GB/s
2.794 GB/s
2.561 GB/s
2.328 GB/s
2.095 GB/s
1.863 GB/s
1.63 GB/s
1.397 GB/s
1.164 GB/s
953.7 MB/s
715.3 MB/s
476.8 MB/s
238.4 MB/s
0 B/s
OUT
1
2
3
4
5
6
7
8
Dec 2013
AGLT2 BEIJING-LCG2 BNL BU CERN CPPM CYFRONET-LCG2 DESY-HH DESY_ZN FZK-LCG2 GRIF-IRFU GRIF-LAL GRIF-LPNHE IHEP
IN2P3-LAPP IN2P3-LPSC INFN_Frascati INFN_Napoli INFN_Roma1 JINR LRZ MWT2_IU MWT2_UC MWT2_UIUC pic PRAGUE RAL-LCG2 SLAC
SWT2_OU UKI-LT2-QMUL UKI-LT2-RHUL UKI-NORTHGRID-LANCS-HEP UKI-NORTHGRID-LIV-HEP UKI-NORTHGRID-MAN-HEP UKI-NORTHGRID-SHEF-HEP
UKI-SCOTGRID-ECDF UKI-SCOTGRID-GLASGOW UKI-SOUTHGRID-CAM-HEP UKI-SOUTHGRID-OX-HEP UNI-FREIBURG unige.ch WUPPERTAL

Failover to FAX

- PanDA sends a job only to a site having all the input data. In case that an input file can not be obtained after 2 tries, the file will be obtained through FAX in case it exists anywhere else
- Job failure rate are small, but cause significant long turn-around time

Record count: 3763

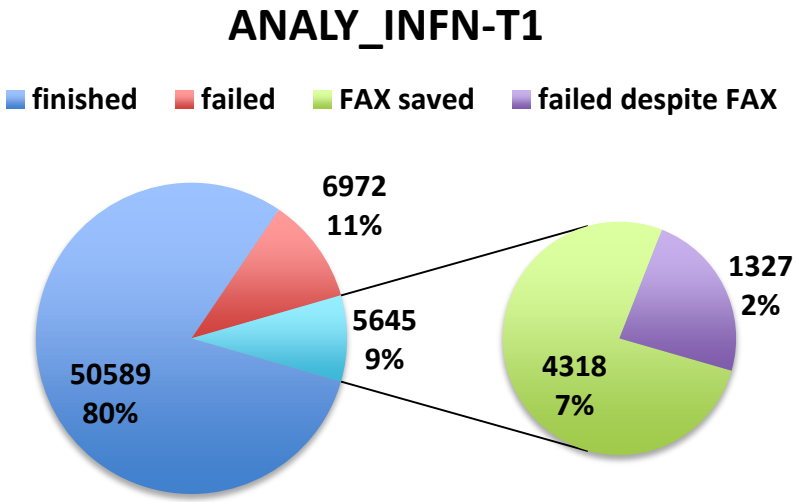
Show 50 entries Search:

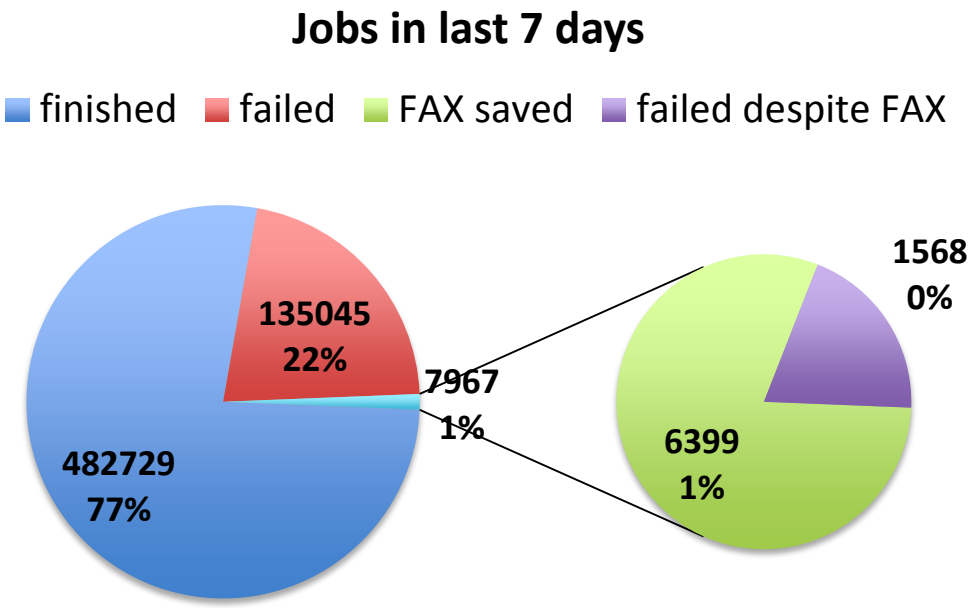
Site	Jobs	WithFAX [files]	WithoutFAX [files]	WithFAX [GB]	WithoutFAX [GB]
DE: GoeGrid	618	2419	3885	959.61	1456.27
DE: LRZ-LMU_HI	1	1	10	0.66	3.83
IT: ANALY_INFN-T1	2006	2185	7443	669.66	10744.11
TW: Taiwan-LCG2	21	42	193	18.08	77.42
UK: ANALY_LANCS_SL6	3	3	41	4.23	52.51
UK: UKI-NORTHGRID-LANCS-HEP_SL6	1	1	10	0.13	3.78
UK: UKI-SCOTGRID-GLASGOW_SL6	39	39	130	11.06	44.16
UK: UKI-SOUTHGRID-CAM-HEP_SL6	1	1	10	0.57	3.51
US: AGLT2_SL6	10	15	49	4.09	20.82
US: ANALY_BNL_LONG	335	403	12599	485.44	2883.31
US: ANALY_BNL_SHORT	364	388	7670	164.58	3650.47
US: ANALY_BU_ATLAS_Tier2_SL6	2	2	9	4.54	15.52
US: ANALY_HU_ATLAS_Tier2	77	81	381	25.51	686.26
US: ANALY_MWT2_SL6	11	11	148	3.01	134.23
US: ANALY_OU_OCHEP_SWT2	1	4	0	9.37	0.00
US: BU_ATLAS_Tier2_SL6	1	1	10	0.66	4.35
US: HU_ATLAS_Tier2	64	70	554	30.29	217.43
US: MWT2_SL6	6	25	33	1.71	9.70
US: OU_OCHEP_SWT2	198	235	1560	96.05	458.42
US: SWT2_CPB	3	23	0	8.52	0.00
US: UTA_SWT2	1	1	0	0.14	0.00

Failover to FAX

For some queues FAX “saves” a significant number of jobs that would otherwise fail.

For most of sites percentage of jobs failing due to IO problems small.

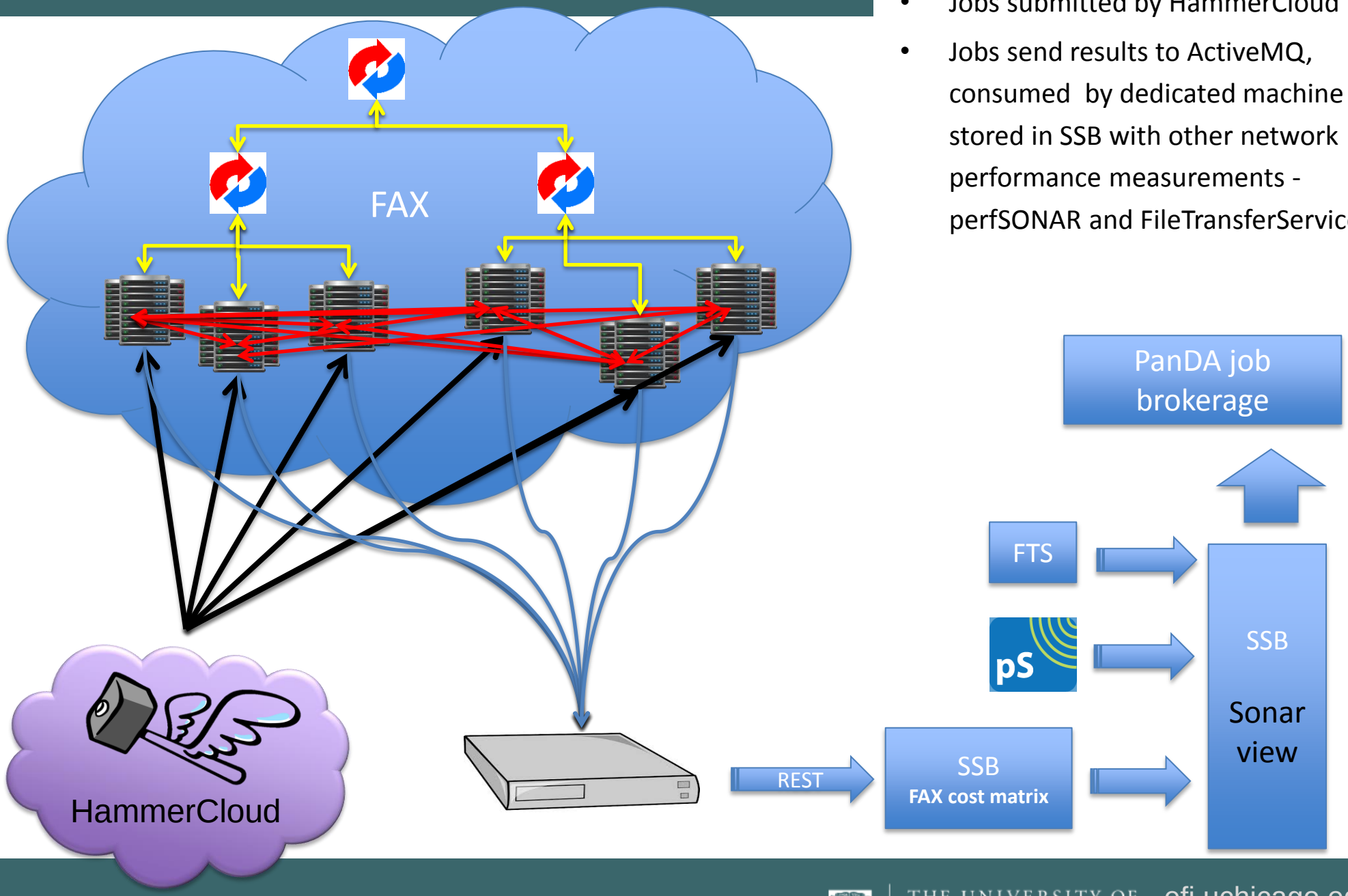




Failed-over jobs transferred **12324 (5 TB)** files from FAX and **55788 (29 TB)** from local storage.

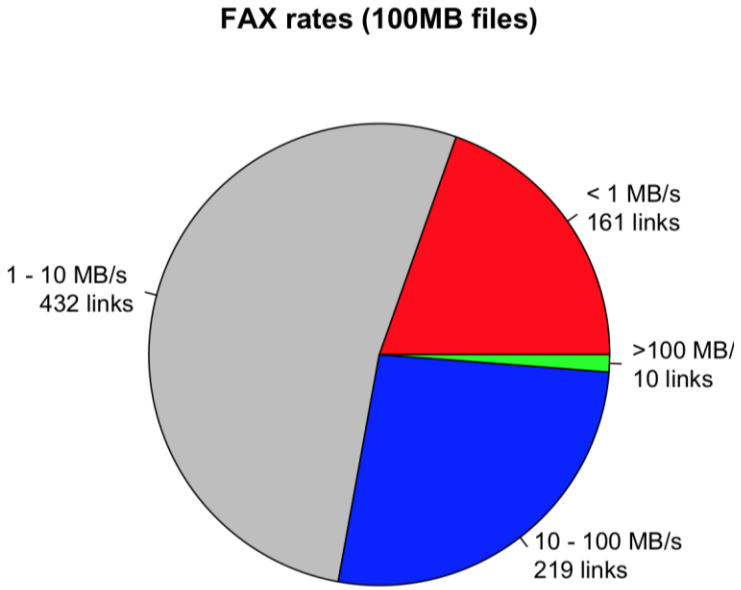
Cost matrix

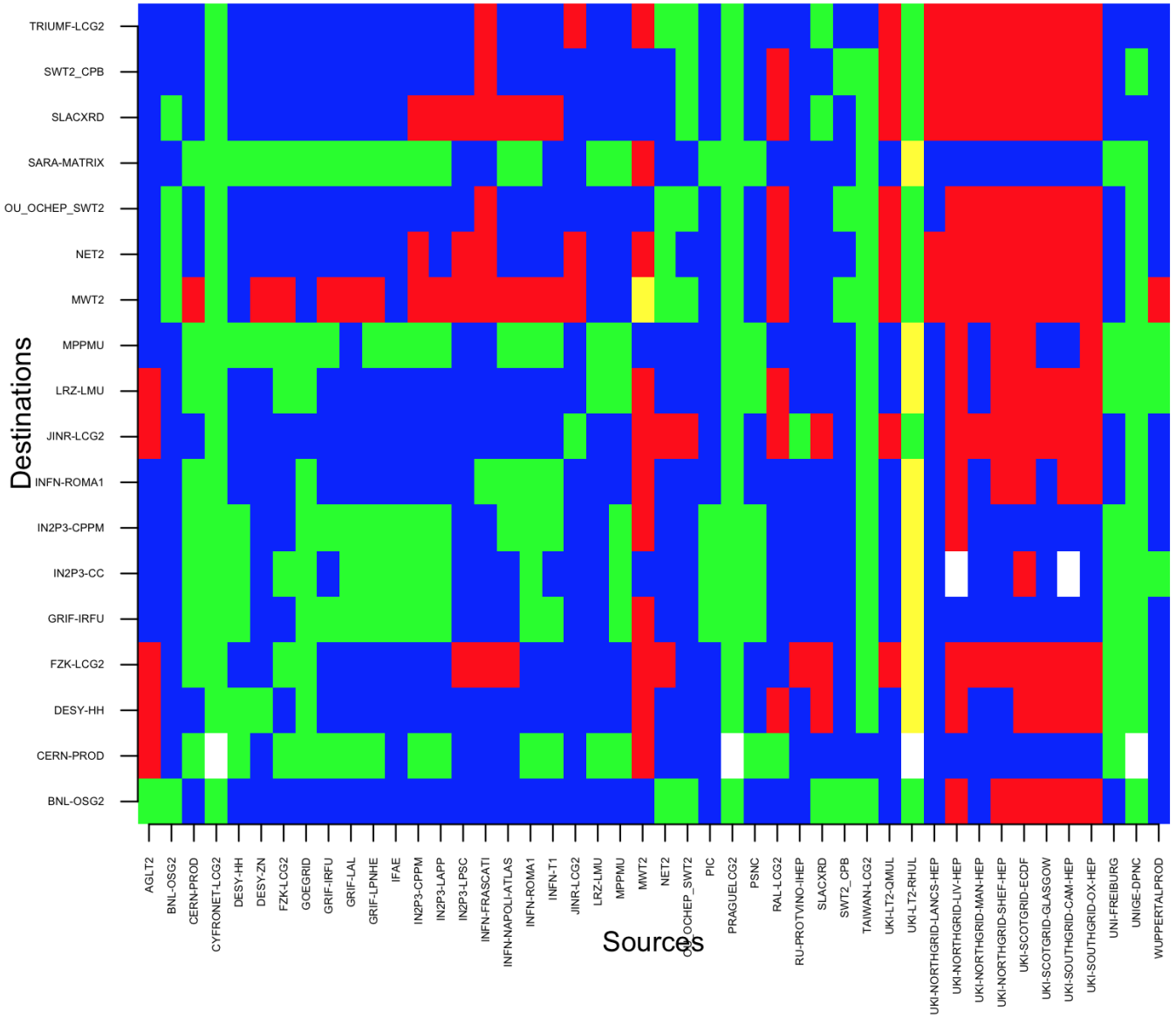
- Measures transfer rates (memory-to-memory) between 42 ANALY queues and each FAX endpoint
- Jobs submitted by HammerCloud
- Jobs send results to ActiveMQ, consumed by dedicated machine stored in SSB with other network performance measurements - perfSONAR and FileTransferService



Cost matrix

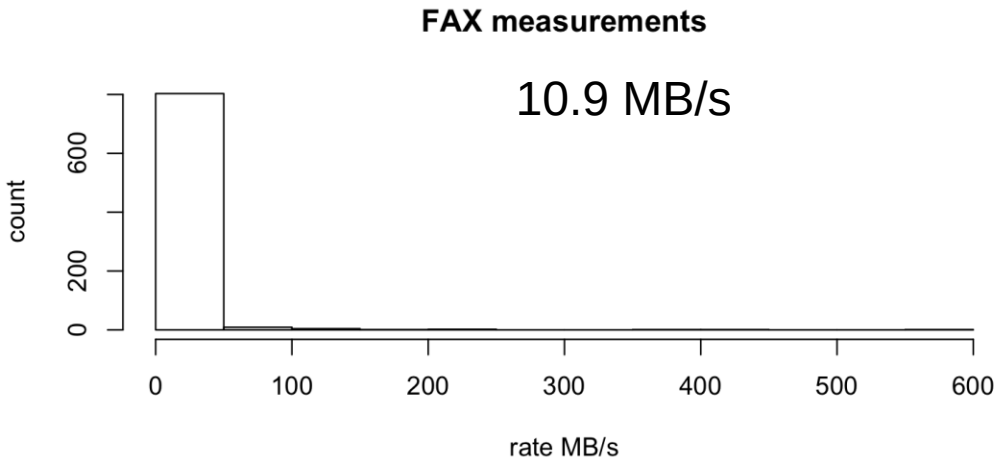
- All the chain is ready. Waiting for Tadashi to add cost into decision making.

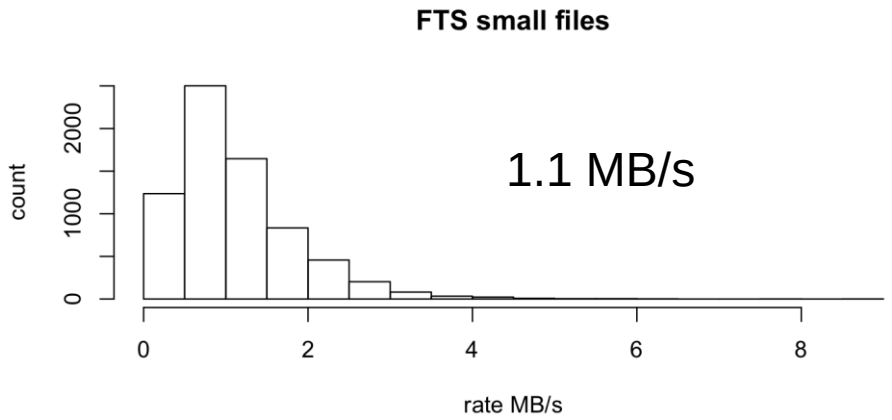


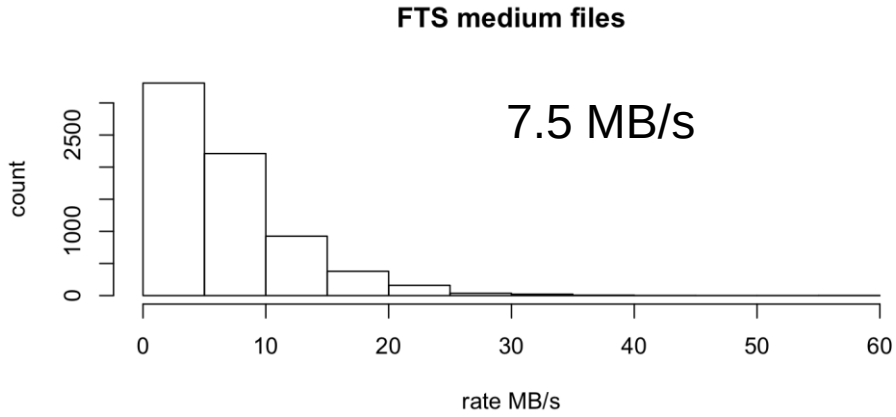


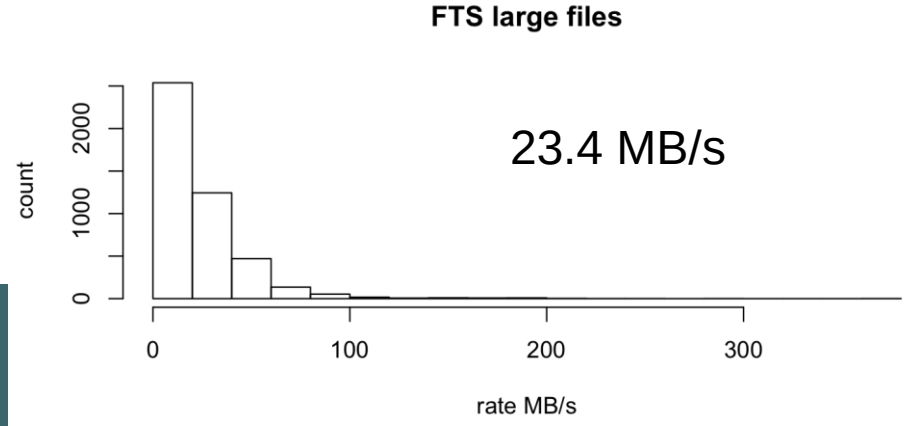
Compared to FTS

- An effort to analyze and understand measurements of FAX, FTS and PerfSONAR
- Jorge, Artem, Ilija









FAX test files 100 MB

FTS small <100 MB, medium 100MB – 1GB, large >1GB

Next Focus: Stress Test (RUCIO only)

Already doing small scale stress test

Sequential list of 744 files

	old gLFNs (LFC)		new gLFNs (RUCIO)	
Site	**time [mm:ss]**	**failures**	**time [mm:ss]**	**failures**
AGLT2	35:18	3	13:14	0
WT2	40:00	3	11:06	3
CERN-PROD	56:19	0	not yet in RUCIO	
MWT2	30:52	0	10:56	0

Feedback from power users indicate improved stability

Next:

HC test on sites stability and performance, or FDR again (especially for cross site test)

Time line: Jan-Feb 2014?
prefer to do this before DC14?

Pre-requirement for Stress test:

- RUCIO renaming completed
- RUCIO N2N deployed and functioning
- Redirector network fully functioning
  - DE cloud redirector isn't working well

Beyond the Next Focus

Future Site Integration:

- Develop redundant setup architecture at site level
  - Increase optimization of site architecture (e.g. proxy)
- Drop LFC support in N2N

Continue Improve on "Other Activities"