



ALICE
A JOURNEY OF DISCOVERY

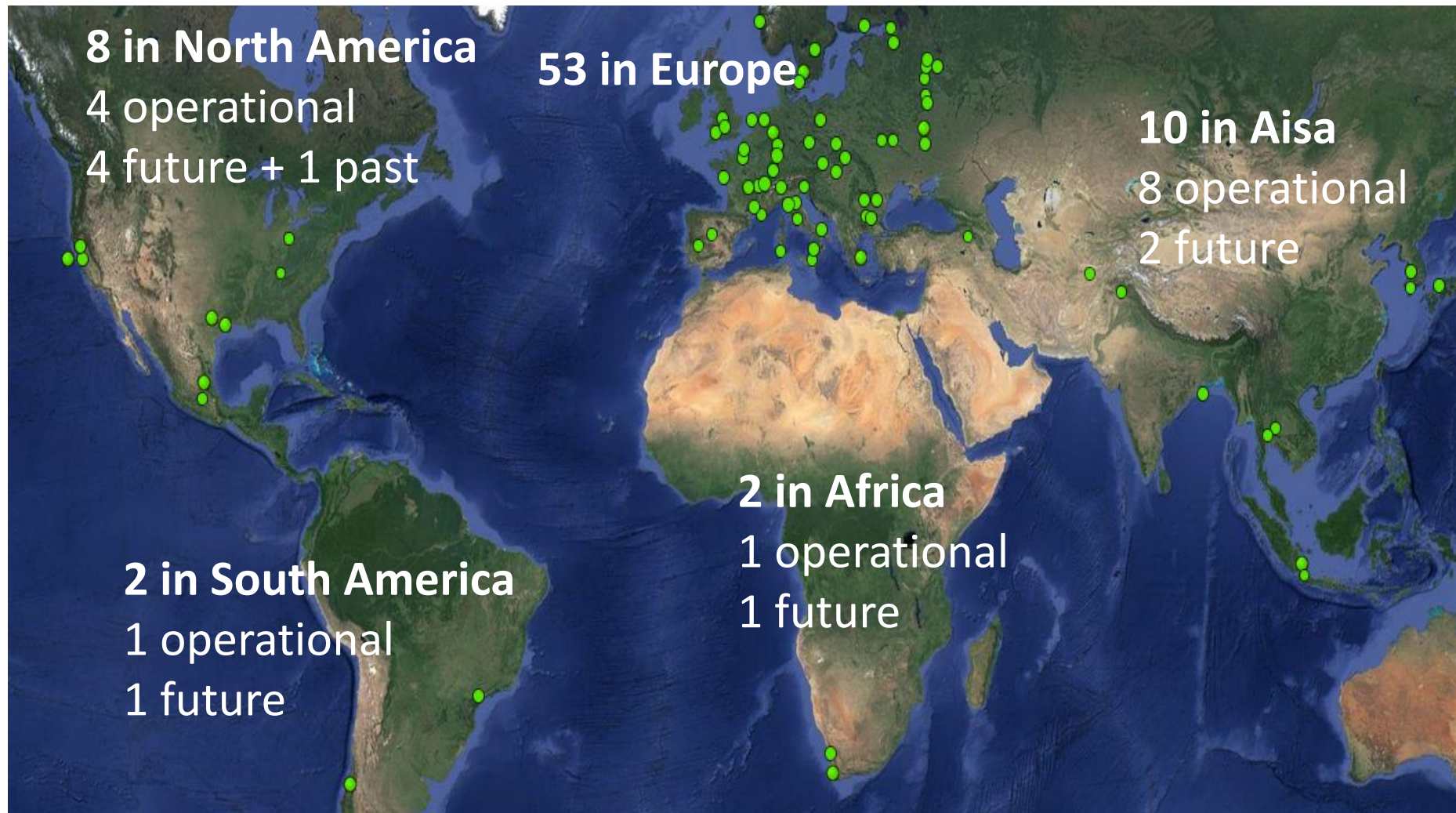
ALICE Grid operations +some specific for T2s

US-ALICE Grid operations review

7 March 2014

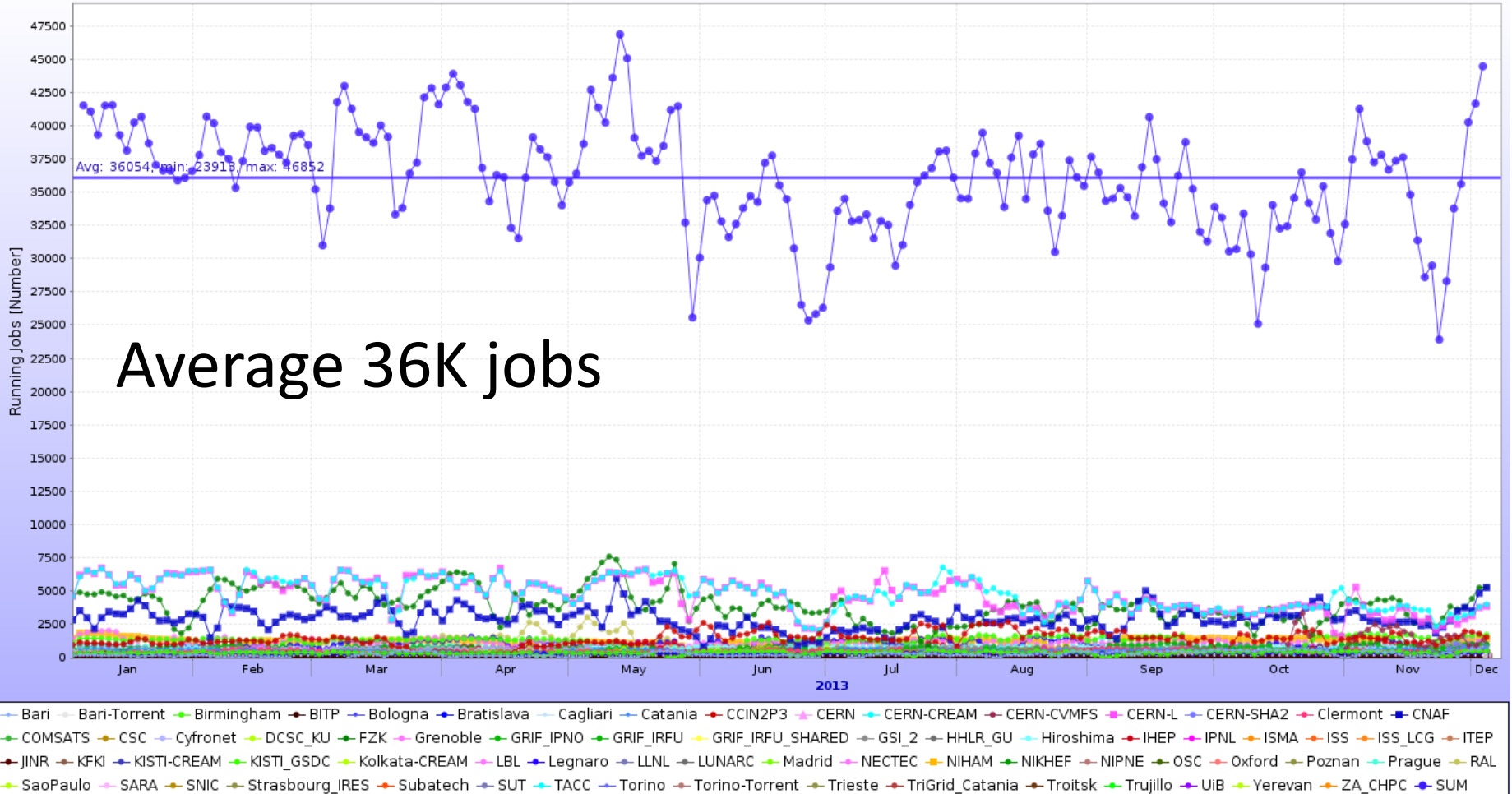
Latchezar Betev

The ALICE Grid



Grid job profile in 2013

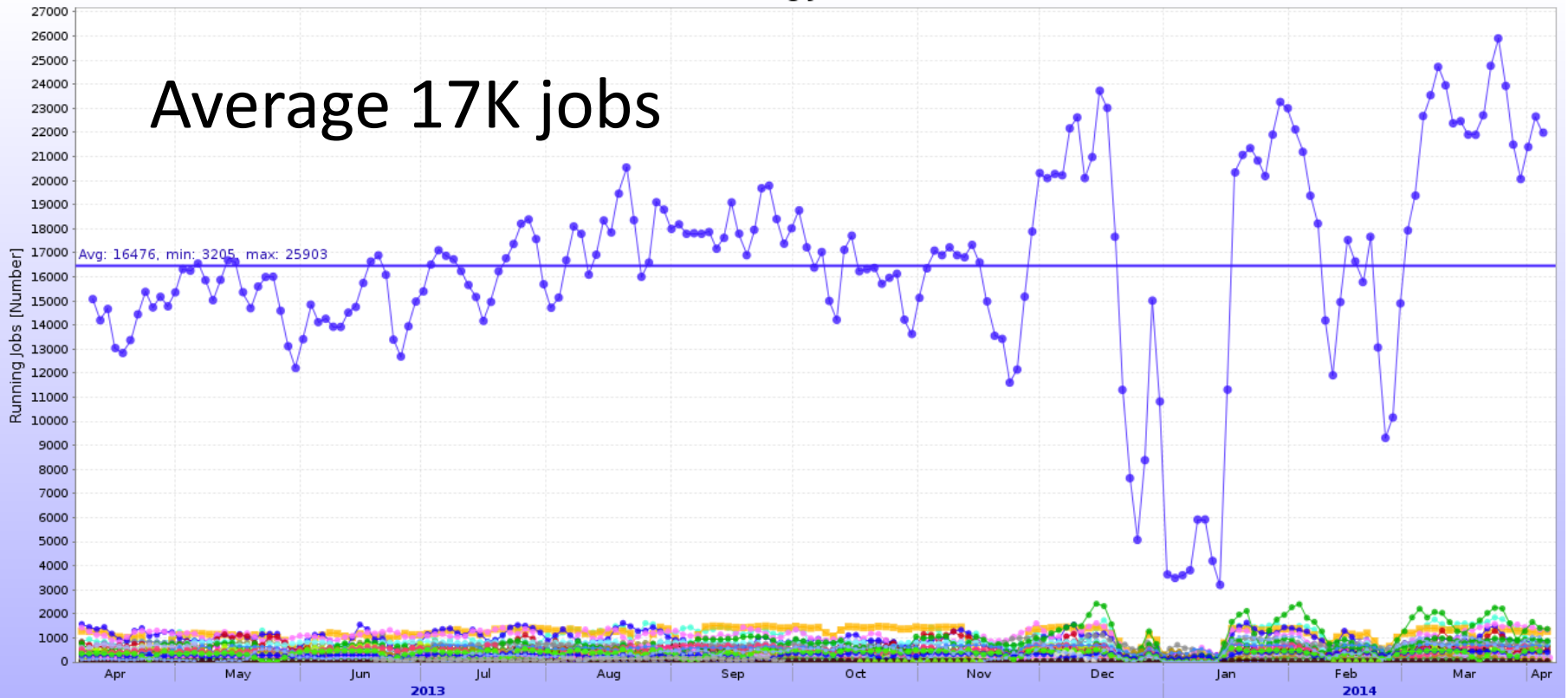
Running Jobs



Grid job profile in 2013 – T2s

Running Jobs

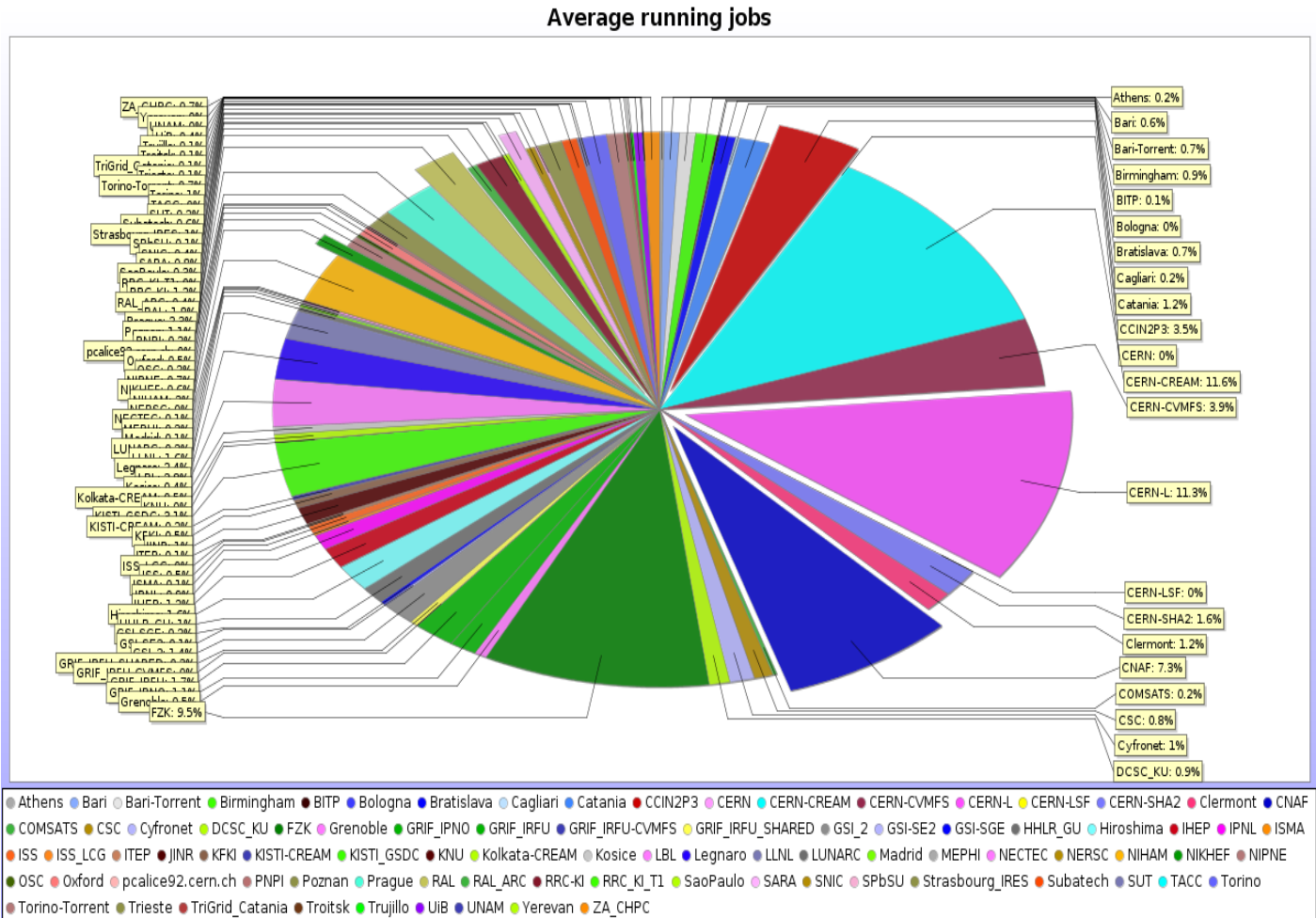
Average 17K jobs



- Athens
- Bandung
- Bari
- Bari-Torrent
- Birmingham
- BITP
- Bologna
- Bratislava
- Cagliari
- Catania
- Clermont
- COMSATS
- CSC
- Cyfronet
- DCSC_KU
- Grenoble
- GRIF_IPNO
- GRIF_IRFU
- GRIF_IRFU-CVMFS
- GRIF_IRFU_SHARED
- GSI
- GSI-SE2
- GSI_2
- HHLR_GU
- Hiroshima
- IHEP
- IPNL
- ISMA
- ISS
- ISS_LCG
- ITEP
- JINR
- KFKI
- KISTI-CREAM
- KNU
- Kolkata
- Kolkata-CREAM
- Kosice
- LBL
- Legnaro
- LLNL
- LUNARC
- Madrid
- MEPHI
- NECTEC
- NERSC
- NIHAM
- NIPNE
- ORNL
- OSC
- Oxford
- PAKGRID
- pcalice92.cern.ch
- PNPI
- Poznan
- Prague
- RRC-KI
- RRC_KI_T1
- SaoPaulo
- SNIC
- SPBSU
- Strasbourg_IRES
- Subatech
- SUT
- TACC
- Talca
- Torino
- Torino-Torrent
- Trieste
- TriGrid_Catania
- Troitsk
- Trujillo
- UIB
- UNAM
- Yerevan
- ZA_CHPC
- SUM

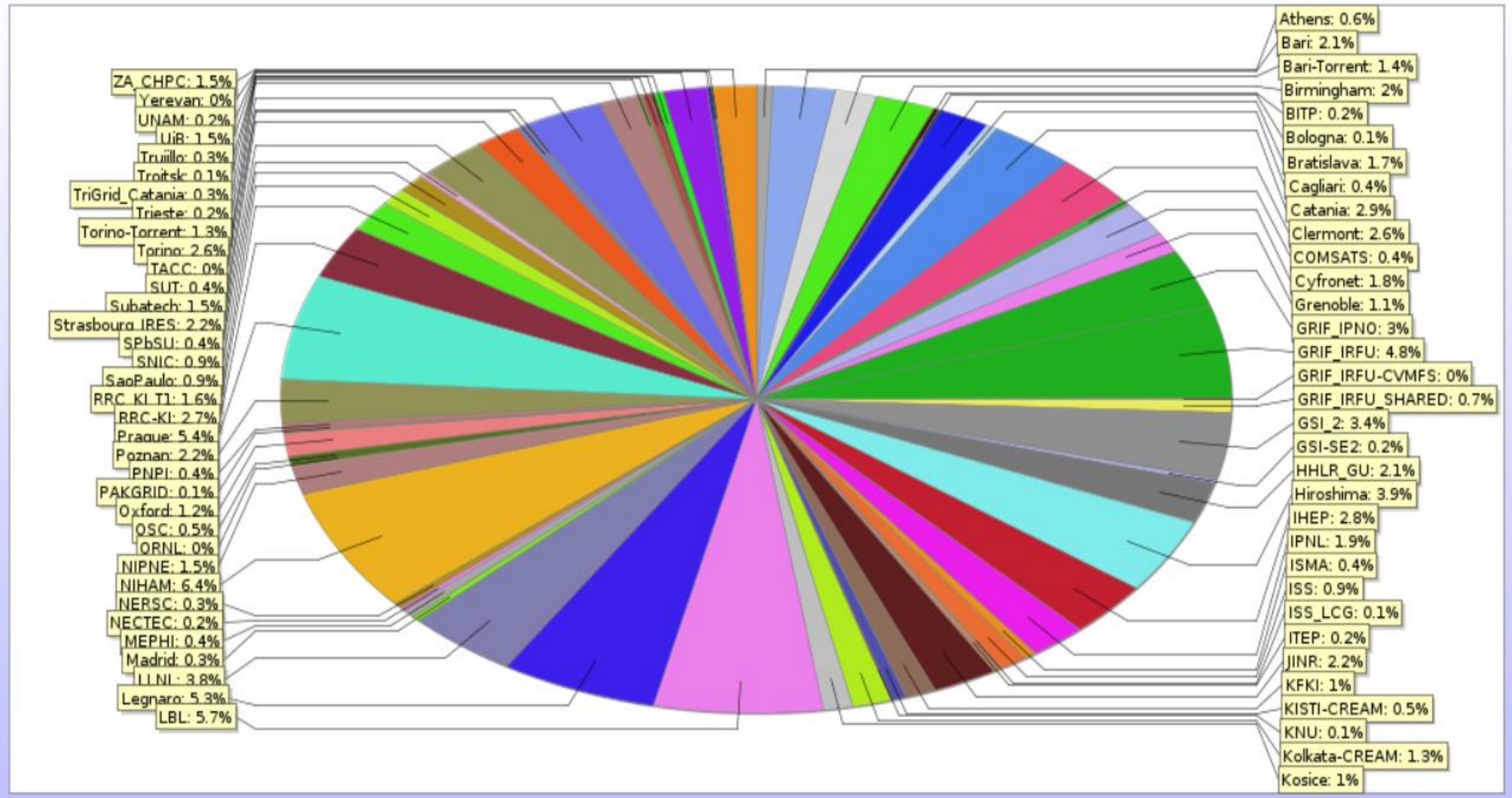
Resources delivery distribution

The remarkable 50/50 share T1/T2 is still alive and well



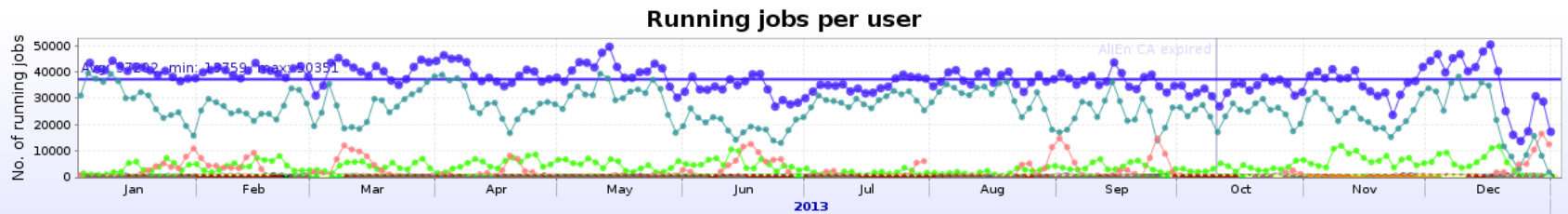
Resources delivery T2s

LBL (6%) + LLNL (4%) + OSC (0.5%) = 10.5% of total T2



Job mixture

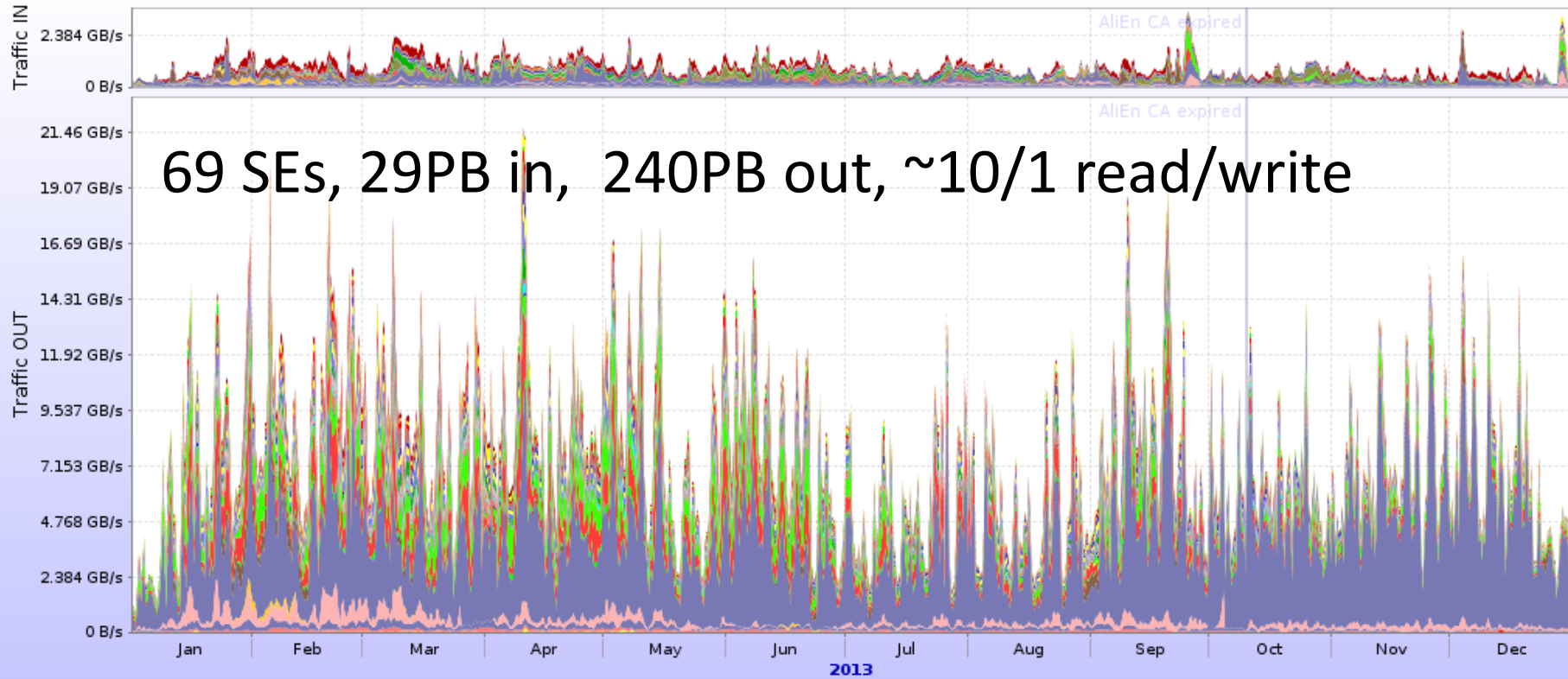
69% MC, 8% RAW, 11% LEGO, 12% individual, 447 individual users



- aabramya → aagrigror → aalici → aalkin → abergogn → abilandz → aborisso → adash → adobrin → adubla → afestant → aggarwal → agheata → agomezra → agostine → agrelli → agrigora
- aherghel → akalweit → akarasu → alardeux → alcaliva → alidaq → aliproduct → alitrain → alla → altsybee → amas → amastros → amatyja → amishra → amorreal → anolivei → ansharma
- antoniol → aortizve → apalaha → apandey → arauf → arnaldi → arossi → ataurro → atimmis → atsuji → attilio → audupa → auras → aveen → awhitehe → ayut → azaborow → azaroche
- baek → bastid → bdoenigu → bedanga → beole → betevl → bgruberg → bguerzon → bianchil → bkileng → bnorris → bogdan → bpaul → bpelesser → bsahlmul → bschang → candrei → canoa
- cbedda → cbianchi → cferreir → cholm → cjahnke → cjena → cluzzi → cmayer → cmohler → cnattras → coppedis → covisan → cperez → cristeia → csilvest → csoegaar → cterrevo → cuautle
- cyaldo → cynthia → czach → dainesea → das → dblau → dcaffarr → dcoella → ddegrutt → ddoobrigk → ddomenic → decaro → defalco → dgangadh → dgomezco → dialexan → djkim
- dkeijden → dleyvape → dlodato → dlohner → dmuhlhei → dpant → dpatalak → dpiyarat → dponomar → drathee → dsakata → dsarkar → dsekihat → dstocco → dthomas → dwatanab
- eabbas → ebruna → ebuthete → ecalvovi → ecasula → echeilad → ekryshen → elumens → emeninno → epereira → epezzele → epohjois → erogocha → eserradi → fbarile → fbellini → fbock
- fbossu → fcolamar → ffionda → filimon → fkrizek → freidt → frprino → fzhou → gbencedi → gconesab → germain → ginnocen → gkoyitha → gluparel → goerlich → gonzalez → grigras
- gsimatov → guernane → gulbrand → gvolpe → habeck → hamagaki → hansena → hbelloma → herdal → hleovar → hljunggr → hongyan → hosokawa → hozhu → hpoppenb → htjung
- hupereir → iarsene → ibhat → idas → ikoutche → ilakomov → imaldona → imartash → ivoroby → janielsk → jaroslav → javander → jbohm → jbook → jcastill → jcunning → jdo → jgamble
- jgcn → jgradosl → jgrosseo → jikumar → jinkim → jisong → jklay → jklein → jkral → jmartinb → jmazer → jmercado → jmjnyar → jrak → jsalwede → jseger → jstiller → jungyu → jviinika
- jwilkins → kamin → kgunji → kharlov → kimb → kiselev → kkobayas → kleinb → kmikhail → kobdaj → kong91 → konush → koshiba → kschwarz → ksenosi → kshtejer → kskjerda → kthomps
- kujjer → kumara → ladrón → lagana → laphacet → lbarnby → lbrenner → lcalerod → lcuunquei → lfeldkam → lgraczyk → lish → lleardin → lmalinin → lmanceau → lmassacr → lmilano → lmolnar
- loizides → lolah → lramona → lronflet → lvalenci → mafontan → majanik → mamukher → marene → maszyman → matarzil → mazimmer → mbombara → mbroz → mchojnc → mcolocci
- mconnors → mcosenti → mewang → mfasel → mfiguere → mgagliar → mguilbau → mgumbo → mhecker → minkim → miweber → mkim → mkohler → mkour → mkrzewic → mleoncin
- mmalayev → mmarchis → mmeres → mmmartin → morsch → mploskon → mrodrigu → mrwilde → msong → mspryop → msteinpr → mstolpov → mtangaro → mvala → mvarygas → mvassili
- mveldhoe → mverweij → mvl → mzesko → nagrawal → nbehera → nilsen → nmanukya → nmohamma → nnovitzk → noferini → nsharma → ntanaka → nystrand → nzhighare → odjuvsia
- okovalen → pachmay → paganop → pchrist → pcrochet → pdinezza → pdutoit → pganoti → pgonzale → pkalinak → pkhan → pkurash → ploenne → pluettig → podesta → poghos → polishch
- ppareek → ppillot → prabhat → prosnet → prsnko → psahoo → psaz → pscott → psrisawa → pversteer → raul → rbala → rbaral → rbelmont → rbertens → rcruzalb → rdang → rgrajcar
- rgrosso → rhaake → richterm → rirusso → rkhandel → rma → rmazumde → rodrigua → rpregen → rromita → rsarneck → rscott → rsingh → rsultano → rtanizak → sahil → sahn → saiola
- salapoin → saltinpi → sbansal → sbjelogr → sbufalin → sdash → sde → sefcik → sesumi → sevdokim → sgaur → shabetai → sharma → shayashi → sheckel → sjena → skar → slindal
- smanconi → smhlanga → soh → spahulah → spflitsc → spiano → spochybo → sprasad → srajput → srasanen → ssakai → sschrein → ssingha → subasu → subikash → svallero → syano
- syasnopo → takim → takobaya → tapiata → tbrownin → tchujo → tjurik → tmoon → tschuste → tsinha → tsokubo → ttsuji → turrisi → tyuasa → unknown → uwesterh → vajzerm → vbairath
- venaruzz → veral → vgrabski → victor → vkovalen → vkucera → vpapikya → vramilli → vrزازi → vriabov → wislavi → vzaccolo → wsato → xizhu → xlopez → xsanchez → xzhang → ycorrale
- yhori → ynam → yozhang → yozhou → yryabov → yzhan → zahammed → zampolli → zconesa → zhangh → zhuj → zhwu → zuzhang → zyin → zzhou → SUM

Access to data (disk SEs)

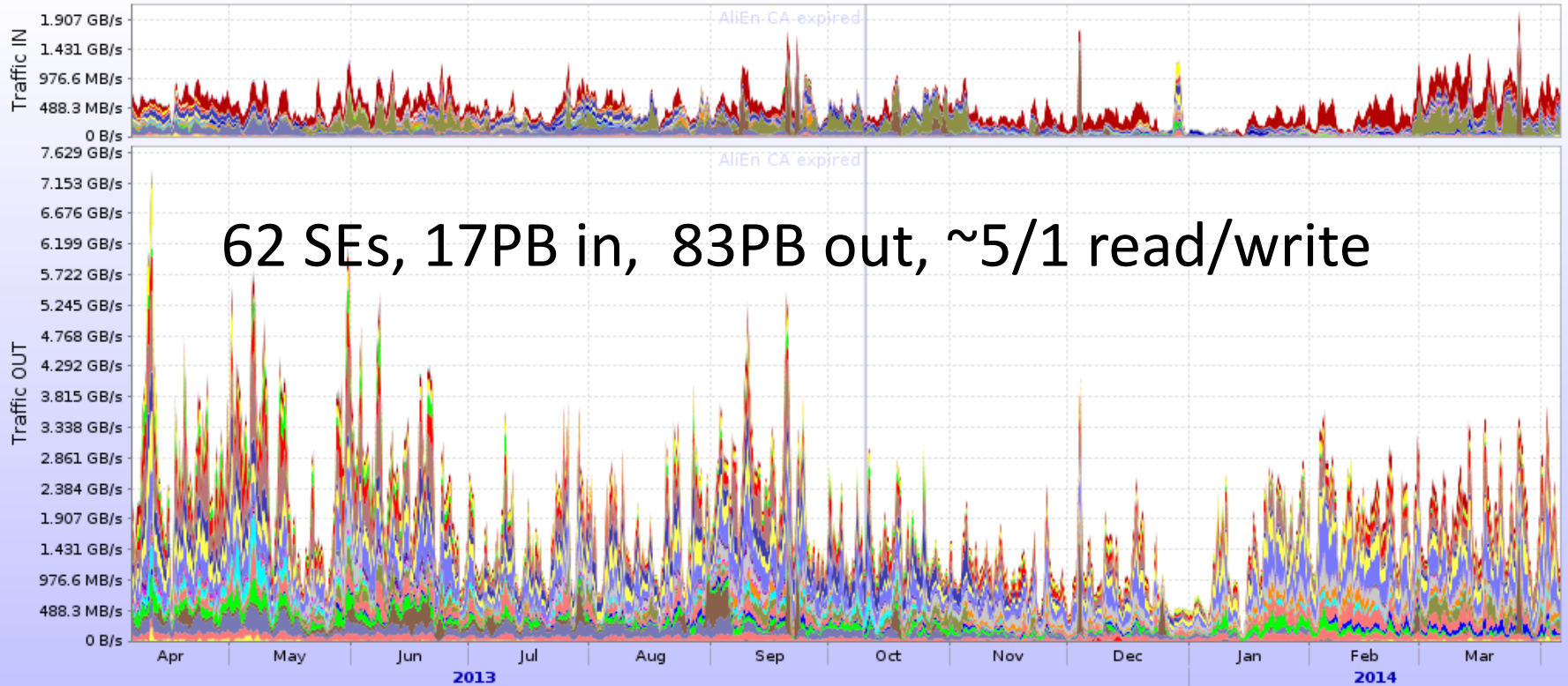
Aggregated network traffic per SE



- SE
- Bari::SE
- BARI::SE
- Birmingham::SE
- BITP::SE
- Bologna::SE
- Bratislava::SE
- Catania::SE
- CCIN2P3::SE
- CCIN2P3::TAPE
- CERN::ALICEDISK
- CERN::EOS
- CERN::TOALICE
- Clermont::SE
- CNAF::SE
- CNAF::TAPE
- CyberSar_Cagliari::SE
- Cyfronet::XRD
- FIXME::SE
- FZK::SE
- FZK::TAPE
- Grenoble::SE
- GRIF_IPNO::SE
- GRIF_IRFU::DPM
- GSI::SE2
- GSI::SE
- HHLR-GU::SE
- Hiroshima::SE
- IHEP::SE
- IPNL::SE
- ISMA::SE
- ISS::FILE
- ITEP::SE
- JINR::SE
- KFKI::SE
- KISTI::SE
- KISTI_GSDC::SE2
- KISTI_GSDC::TAPE
- KISTI_GSDC::TE
- Kolkata::SE
- Kosice::SE
- LBL::SE
- Legnaro::SE
- LLNL::SE
- Madrid::SE
- MEPHI::SE
- NECTEC::SE
- NIHAM::FILE
- PNPI::SE
- Poznan::SE
- Prague::SE
- RRC-KI::SE
- RRC_KI_T1::EOS
- SaoPaulo::SE
- SPbSU::SE
- Strasbourg_IRES::SE
- Subatech::SE
- SUT::SE
- Talca::SE
- Torino::SE
- Trieste::SE
- Trigridd::SE
- Troitsk::SE
- Trujillo::SE
- UNAM_T1::SE
- Wuhan::SE
- WUT::SE
- YERPHI::SE
- ZA_CHPC::SE

Access to data (disk SEs, T2s)

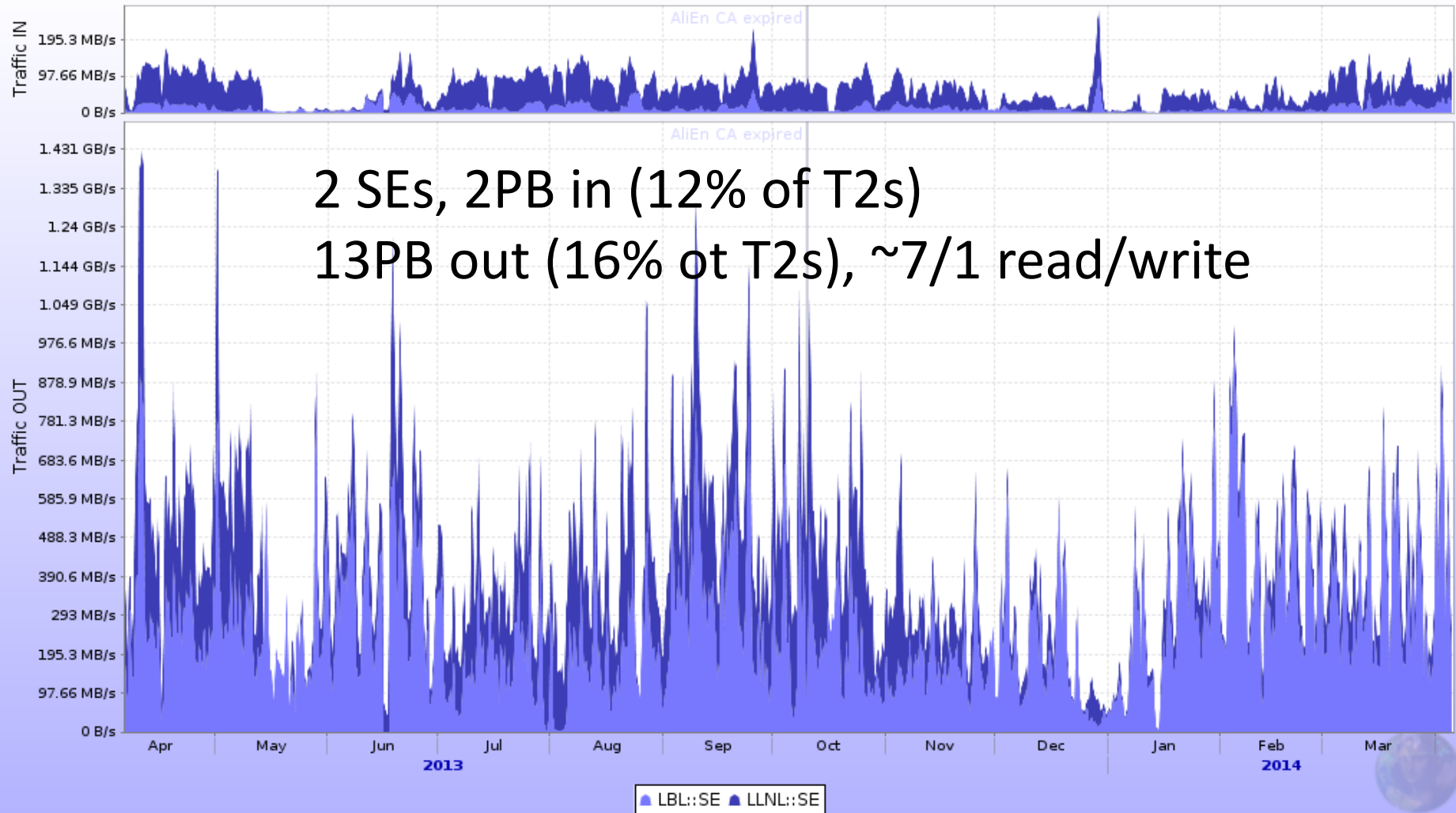
Aggregated network traffic per SE



- ::SE ● Bari::SE ● BARI::SE ● Birmingham::SE ● BITP::SE ● Bo::SE ● Bologna::SE ● Bratislava::SE ● Catania::SE ● CERN::T0ALICE ● Clermont::SE
- CyberSar_Cagliari::SE ● Cyfronet::XRD ● Grenoble::SE ● GRIF_IPNO::SE ● GSI::SE2 ● GSI::SE ● HHLR-GU::SE ● Hiroshima::SE ● IHEP::SE ● IPNL::SE ● ISMA::SE
- ISS::FILE ● ITEP::SE ● JINR::SE ● KFKI::SE ● KISTI::SE ● Kolkata::SE ● Kosice::SE ● LBL::SE ● Legnaro::SE ● LLNL::SE ● Madrid::SE ● MEPHI::SE ● NECTEC::SE
- NIHAM::FILE ● PNPI::SE ● Poznan::SE ● Prague::SE ● RRC-KI::SE ● RRC_KI_T1::EOS ● SaoPaulo::SE ● SPbSU::EOS ● SPbSU::SE ● Strasbourg_IRES::SE
- Subatech::EOS ● Subatech::SE ● SUT::SE ● Torino::SE ● Trieste::SE ● Troitsk::SE ● Trujillo::SE ● UNAM_T1::SE ● Wuhan::SE ● WUT::SE ● YERPHI::SE
- ZA_CHPC::SE

Access to data (disk SEs, US)

Aggregated network traffic per SE



Data access 2

- 99% of the data read are input (ESDs/AODs) to analysis jobs, the remaining 1% are configurations and macros
- From LEGO train statistics, ~93% of the data is read locally
 - The job is sent to the data
- The 7% is file cannot be accessed locally (either server not returning it or file missing)
 - In all such cases, the file is read remotely
 - Or the job has waited for too long and is allowed to run anywhere to complete the train (last train jobs)
- Eliminating some of the remote access (not all possible) will increase the global efficiency by few percent
 - This is not a showstopper at all, especially with better network

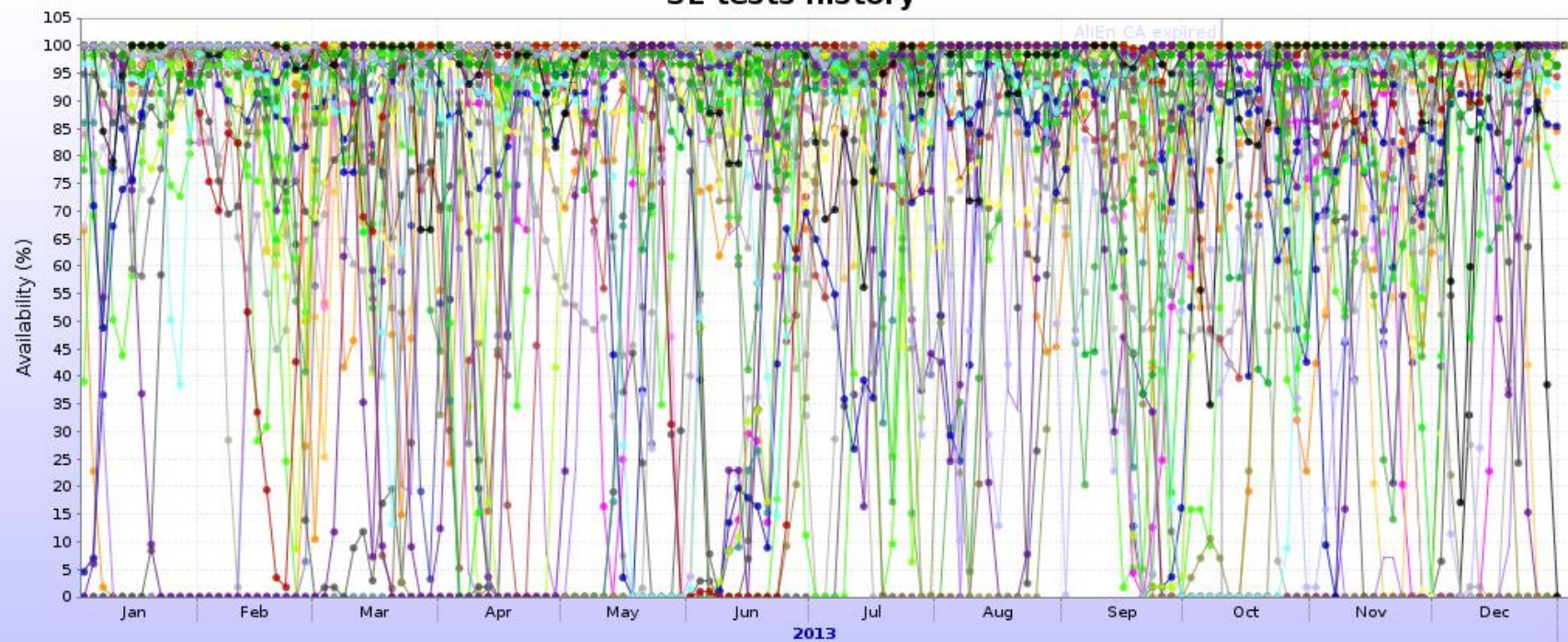
Storage availability

- More important question – availability of storage
- ALICE computing model – 2 replicas => if SE is down, we lose efficiency and may overload the remaining SE
 - The CPU resources must access data remotely, otherwise there will be not enough to satisfy the demand
- In the future, we may be forced to go to one replica
 - Cannot be done for popular data

Storage availability (2)

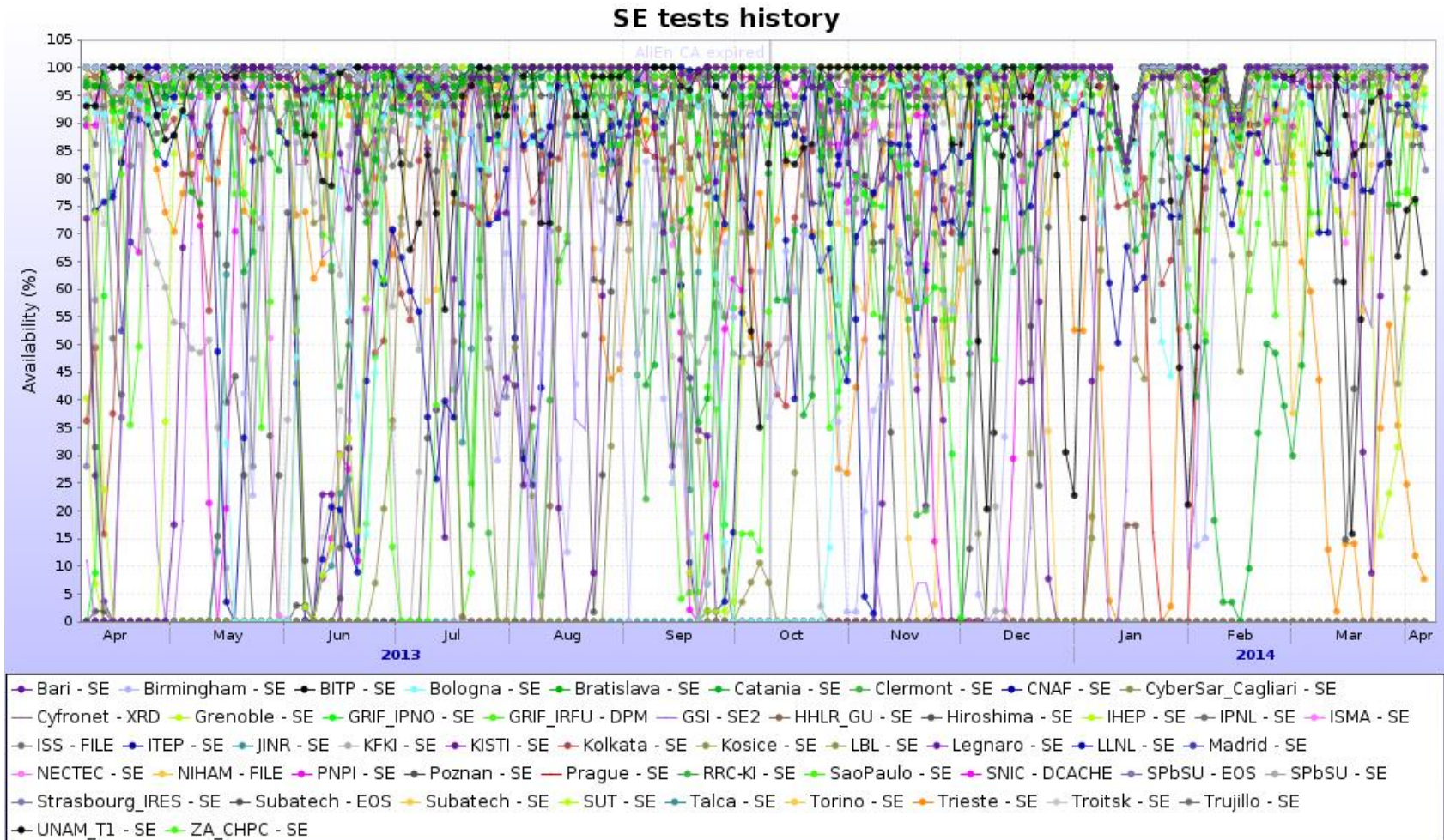
- Average SE availability in the last year: 86%

SE tests history



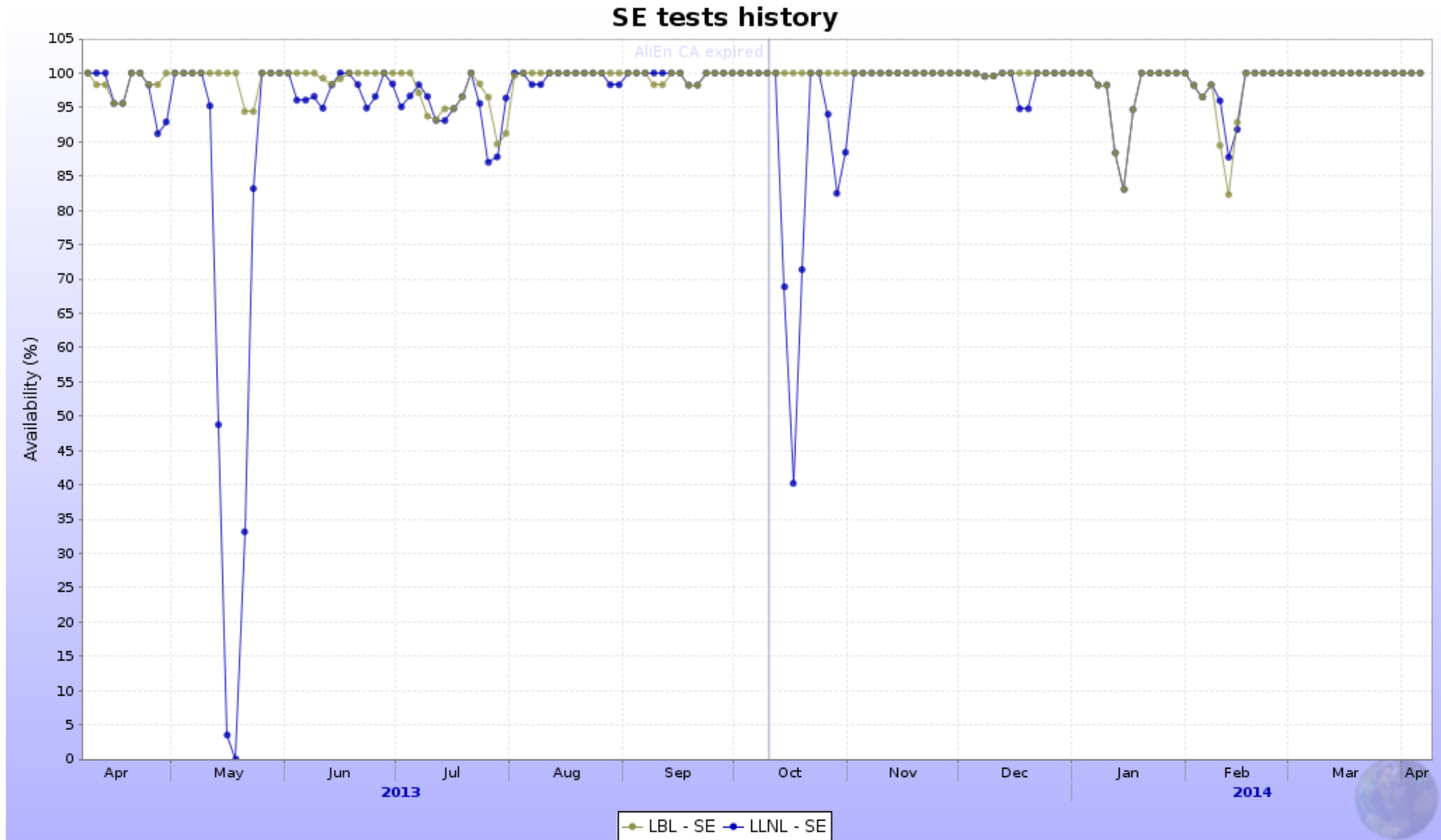
Storage availability – T2s

- Average SE availability in the last year: 80%



Storage availability – US

- Average SE availability in the last year: 97% (!)

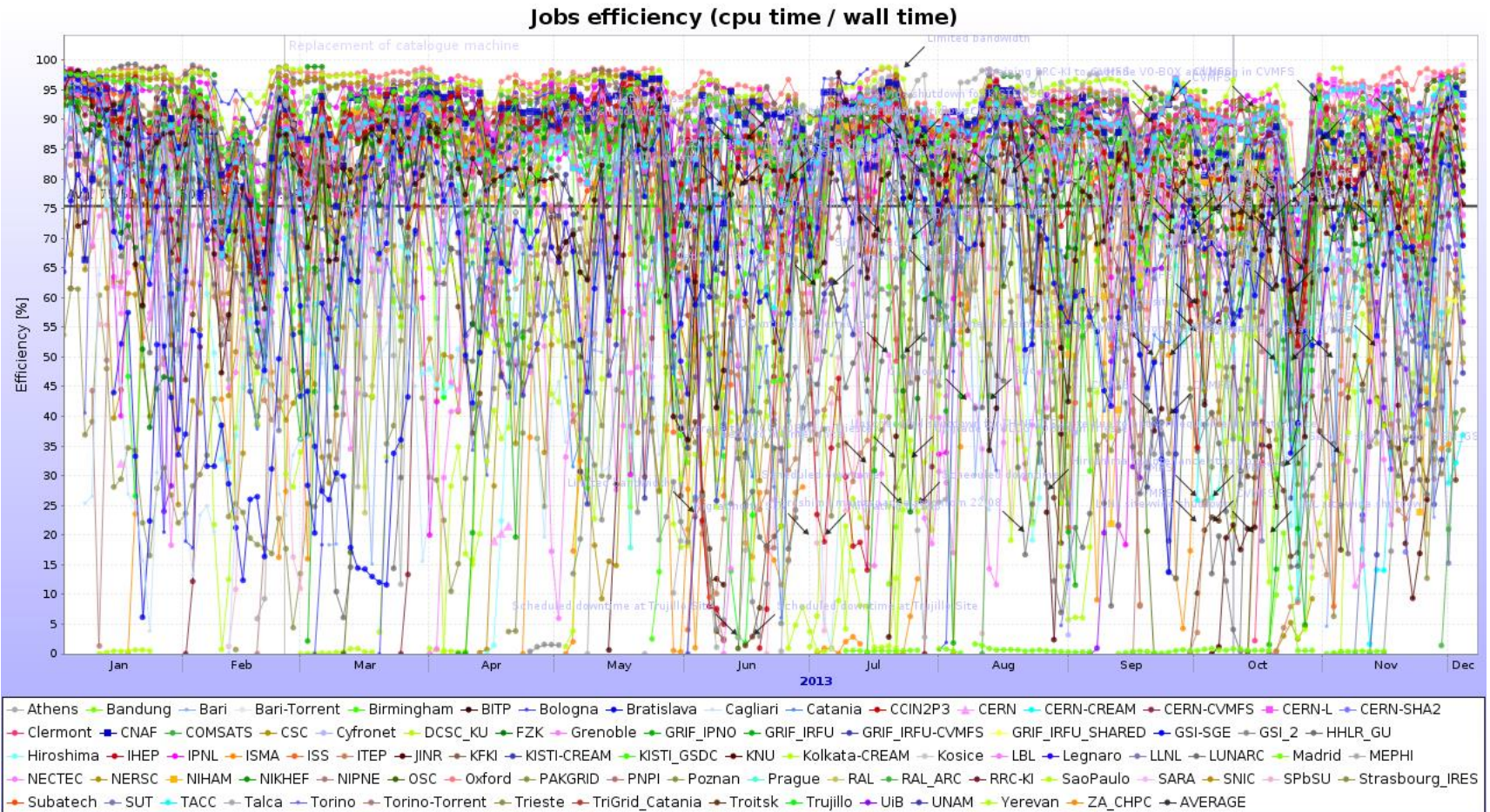


Other services

- Nothing special to report
 - Services are mature and stable
 - Operators are well aware of what is to be done and where
 - Ample monitoring is available for every service (more on this will be reported throughout the workshop)
 - Personal reminders needed from time to time
 - Several services updates were done in 2013...

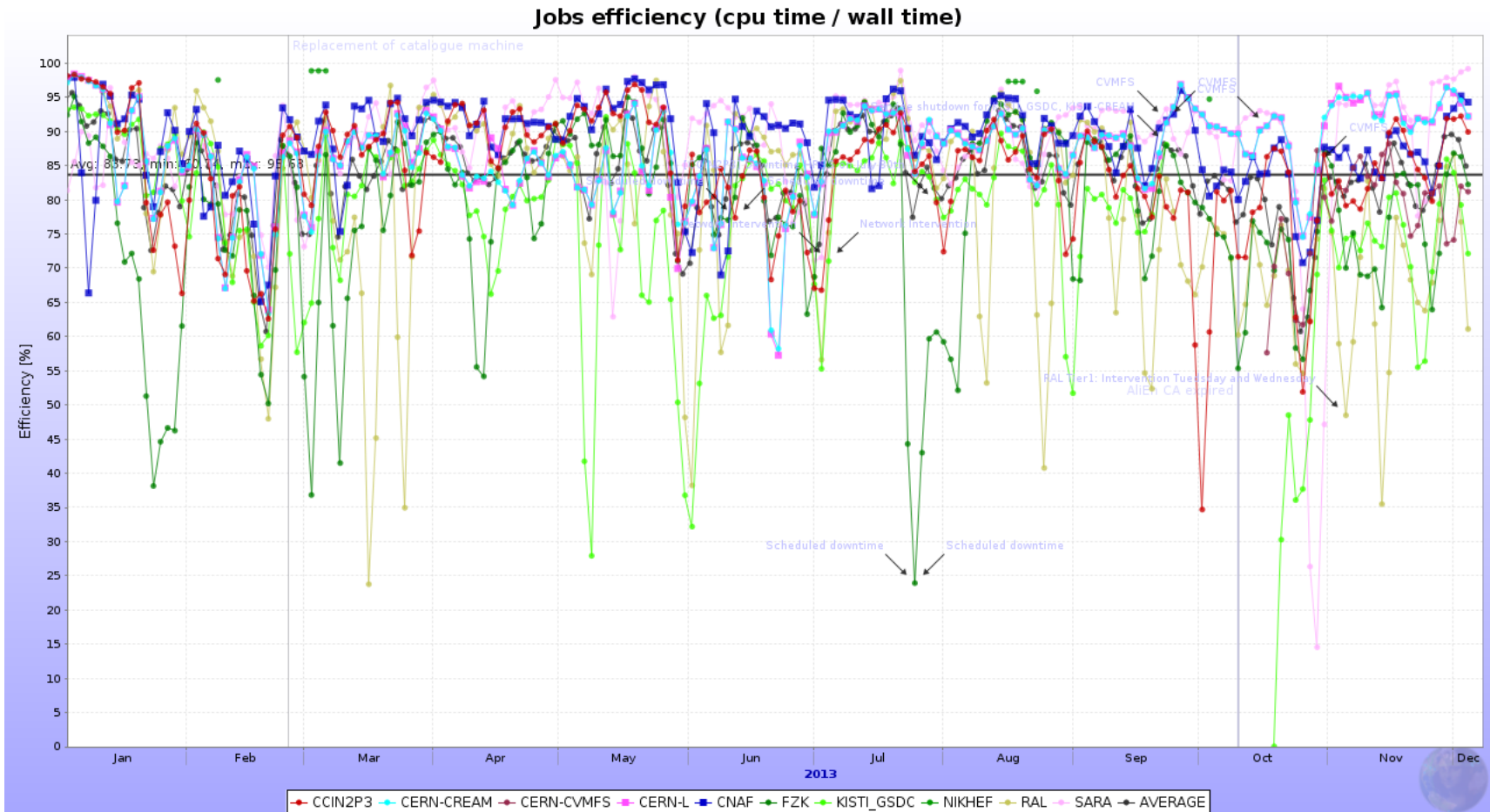
The Efficiency

Average of all sites: 75% (unweighted)



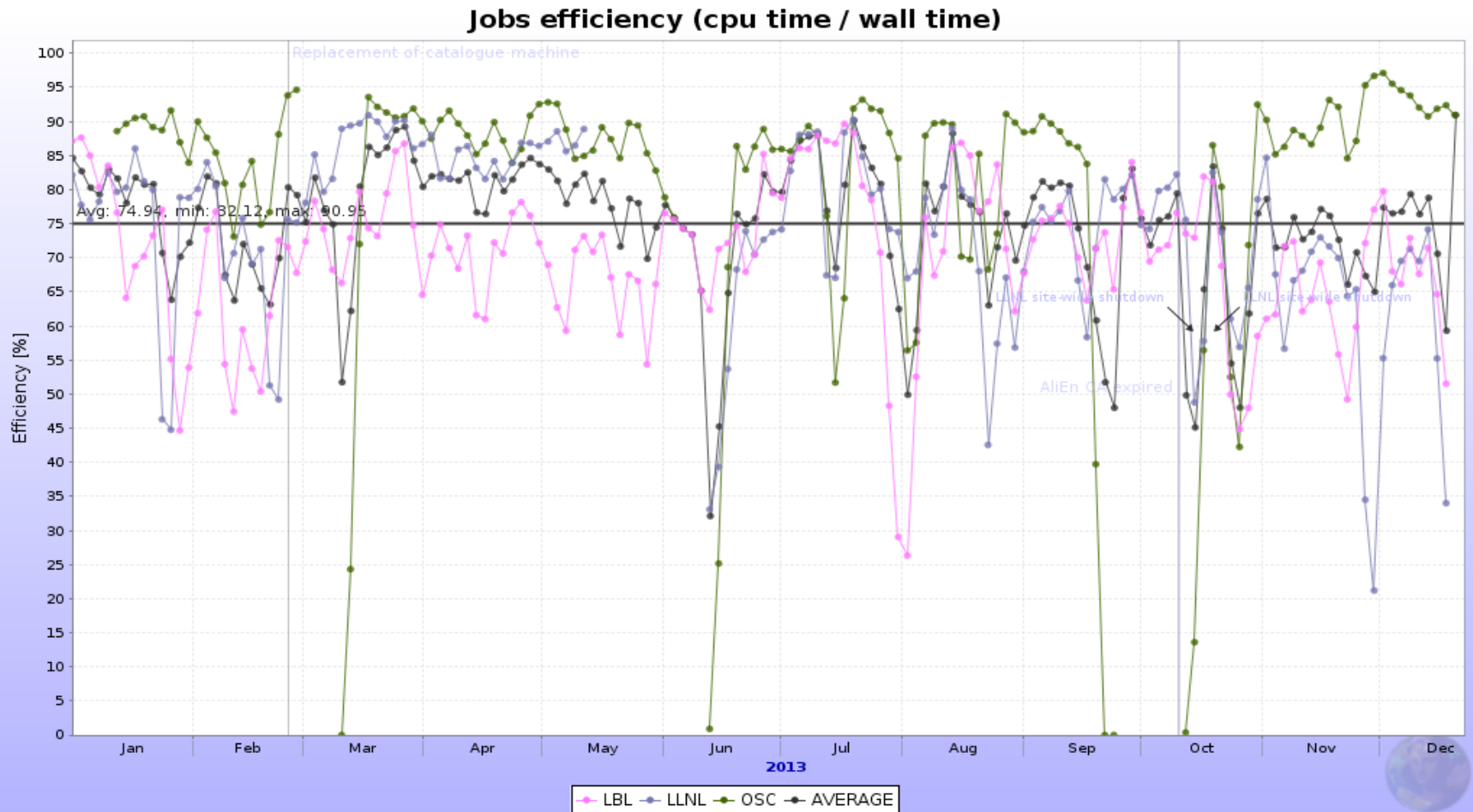
Closer look – T0/T1s

Average – 85% (unweighted)



Closer look – US

Average –75% (unweighted)



Summary on efficiency

- Stable throughout the year
- T2s efficiencies are not much below T0/T1s
 - It is possible to equalize all, it is in the storage and networking
- Biggest gains through
 - Inter-sites network improvement (LHCONE);
 - Storage – keep it simple – xrootd works best directly on a Linux FS and on generic storage boxes

What's in store for 2014

- Production and analysis will not stop – know how to handle these, nothing to worry about
 - Some of the RAW data production is left over from 2013
- Another 'flat' resources year – no increase in requirements
- Year 2015
 - Start of LHC RUN2 - higher luminosity, higher energy
 - Upgraded ALICE detector/DAQ – higher data taking rate; basically 2x the RUN1 rate

What's in store for 2014 - sites

- We should finish with the largest upgrades before March 2015
 - Storage – new xrootd/EOS
 - Services updates
 - Network – IPv6, LHCONE
 - New sites installation – Indonesia, US, Mexico, South Africa
 - Build and validate new T1s – UNAM, RRC-KI (already on the way)

Summary

- Stable and productive Grid operations in 2013
- Resources fully used
- Software updates successfully completed
- MC productions completed according to requests and planning
 - Next year – continue with RAW data reprocessing and associated MC
- Analysis – OK
- 2014 - focus on SE consolidation, resources ramp-up for 2015 (where applicable), networking, new sites installation and validation