Operations in Italy and plans for INFN sites

Domenico Elia, Stefano Bagnasco,
Stefano Piano
INFN sez. Trieste

Outline of the talk

- Operations organization
- Networking status
- Available resources and evolution
- The Tier-1 at INFN-CNAF
- Status of Tier-2 sites
- Mandatory plots
- Other activities:
  - The site dashboard (see Giacchino's talk)
  - Feasibility study to run on HPC (see Matteo's talk)

Operations Organization

- ALICE-IT Computing Coordination:
  - Domenico Elia
- Deputy:
  - Stefano Bagnasco
- Tier-2 Operations Coordination:
  - Stefano Piano
  - Monthly phone conference for coordination and performance monitoring
  - Yearly face-to-face workshop (2012 @ Catania, 2013 @ Trieste, 2014 @ Frascati, 2015 @ CNAF, 2016 @ PD)
- Monthly Tier-1 Management Board at CNAF

Networking

- Tier-1 at 60 Gbps (LHCONE + LHCOPN)
  - 20 Gbps reserved for LHC ONE
  - Upgrade to 100 Gbps connection in 2017
- All Tier-2's connected to LHCONE with at least 10 Gbps
  - Through GARR-X
  - All Tier-2's easily upgradable to 40 Gbps
  - Bari, Catania, Padova-LNL already at 20 Gbps
  - Bari, Catania aim to upgrade at 40 Gbps in 2017
- IPV6
  - All INFN sites will act coordinately at the same time but no exact estimate yet

Resources available for Alice

- **Tier-1 at CNAF, Bologna**
  - Shared with other LHC experiments and a large number of others
- **4 official Tier-2 centers**
  - Bari, Catania, Padova-LNL and Torino
  - “Official” means “directly funded by INFN according to plans and pledges”
- **Additional (minor) centers**
  - Cagliari, Trieste
  - Local resources, different creative funding, mostly out of pledge
- Projects providing resources in the ALICE INFN sites over the last years
  - **ReCaS (BA and CT)**, sizeable contribution to 2014 and 2015 pledges
  - **CyberSar (CA) and TriGrid (CT)**, both ended, resources becoming obsolescent

Current Status

- Tier-1: Pledge 2016 fully deployed:

  CPU: 29045 HS06 / **DISK**: 3900 TB / **TAPE**: 5500 TB

  / (94% used) / (99% used)

- Tier-2: available at the end of 2016

	Bari	Catania	LNL-Padova	Torino	Total
HS06	10512	13147	11385	10289	45333
TB	1244	1204	1202	1226	4876

Pledge 2016: CPU: 43840 HS06 / **DISK**: 4830 TB

Computing resources at INFN T1/T2s
Expected in 2017

- Tier-1:

  Pledge 2017: CPU: 38300 HS06 / **DISK**: 4480 TB / **TAPE**: 10800 TB

  CPU already at plegde 2017, DISK/TAPE at pledge by ~July/Sept

- Tier-2:

  Pledge 2017: CPU: 50870 HS06 / **DISK**: 5790 TB

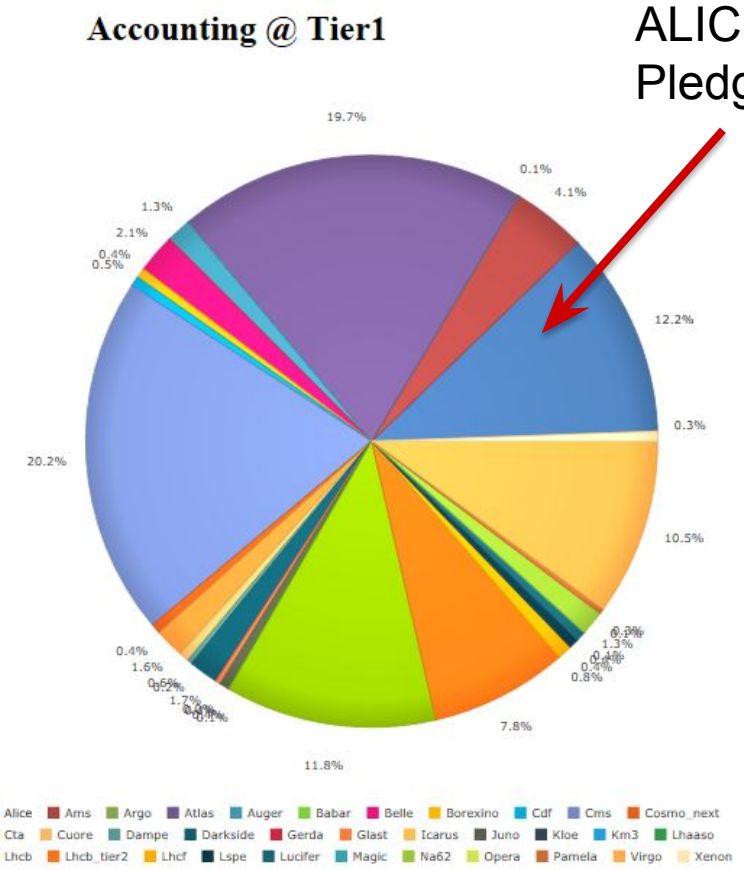
  Resource procurement:

  - global financial budget corresponding to pledge (as in the previous years)
  - INFN pushing for (as much as possible) centralized/grouped tenders
  - CPU: well advanced (handled by CNAF for all Tiers/all LHC exps)
  - DISK: ongoing, grouped “ALICE+CMS” procedure (for all Tiers)

Tier-1 at CNAF

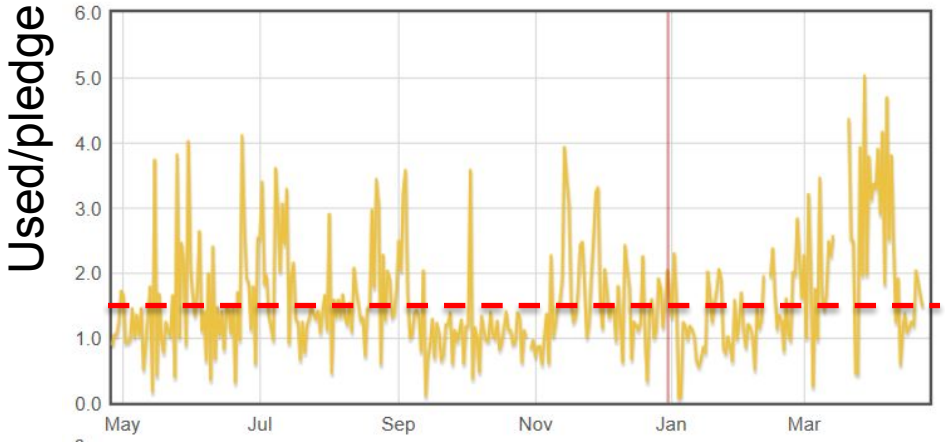
- CPU : ~1000 WNs, ~21500 computing slots, ~220 kHS06
  - ALICE share is 29045 HS06 (2016) => 38295 HS06 (2017)
  - LSF as current Batch System, HTCondor migration foreseen
  - ALICE dedicated resources: 2 VOBoxes
- CNAF::SE : 3.4 PB (94% used)
  - 4 xrootd (v4.1.3) as server+redirector (40Gb/s x 4)
  - 1 alias DNS which redirects to the 4 servers
- CNAF::TAPE : 5.5 PB (99% used)
  - gpfs filesystem (plugin xrootd to manage recalls)
  - 4 xrootd (v4.1.3) servers configured as server+redirector (40Gb/s x 4)
  - 1 alias DNS which redirects to the 4 servers
  - TAPE 1 extra PB available (anticipation for this year)

Used/Pledge at CNAF



ALICE
Pledge 12.2%

2016: 29045 HS06
(2017: 38295 HS06)



Opportunistic usage of resources (pledge+50%)

Tier-1 remote extension to Bari ReCaS

- 48 WNs (~26 kHS06) and ~330 TB of disk allocated to Tier-1 farm for WLCG experiments in Bari-ReCaS data center
- ~10% of CNAF total resources, ~13% of resources pledged to WLCG experiments
- Goal: direct and transparent access from CNAF
- Similar to CERN/Wigner extension
- Goal: transparent access from CNAF farm
  - Should be indistinguishable for users
- CNAF LSF master dispatches jobs also to Bari-ReCaS WNs
  - BARI WNs considered as local resources
- CEs (grid entry points for farm) at CNAF
- Auxiliary services installed in Bari-ReCaS
  - CVMFS Squid servers (for software distribution)
  - Frontier Squid servers (used by ATLAS and CMS for condition db)

Tier-2 sites
General remarks

- New Data Centers (HPC, Grid and Cloud):
  - Bari, Catania and Torino are rather large
  - funding for infrastructural upgrades provided by special project (ReCaS in Bari and Catania) and for hardware by bank foundation (Compagnia S.Paolo in Turin)
  - Pd-LNL Tier-2 spread in 2 sites
- All sites working to expand support to more VOs beyond LHC ones, to allow for resource optimization (e.g. Padova-LNL with OpenStack based cloud infrastructure and Torino with OpenNebula multi-tenancy cloud)
- All sites can allow for more resources coming in without big infrastructural investments although manpower is tight

Tier-2 sites
Bari

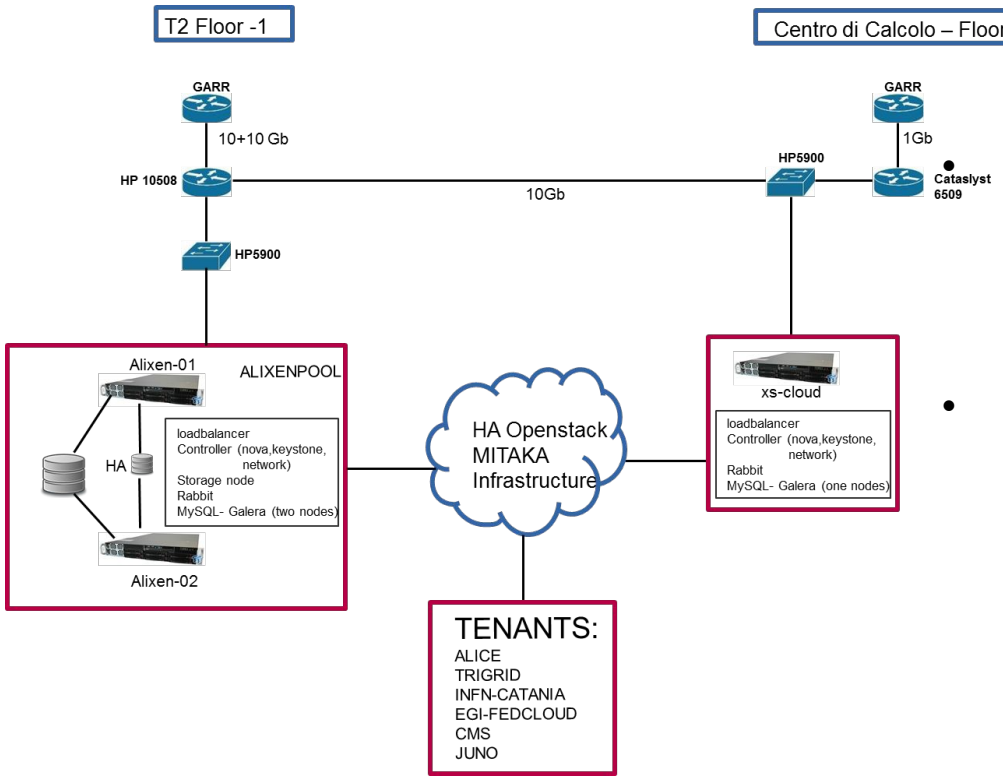
Since July 2015 new ReCaS Data Center:

- 420 m^2 with 4x20 racks
- 10 Gbps point-to-point
- 6 Computer Room Air Conditioners
- UPS 800kW x 7 min + GP1650
- 128 servers (CPU 64 bit AMD)
- 8192 cores (2304 INFN - 5888 UNIBA)
- 3552 TB DELL (1152 INFN - 2400 UNIBA)
- IBM System Storage TS3500 Tape Library (UNIBA)
- HPC cluster (800 core Intel, Infiniband and 20 NVIDIA K40 boards) (UNIBA)

Tier-2 sites
Bari

- 20 Gbps link GARR-X/LHC-ONE (upgrade to 40 Gbps)
- Storage:
  - 2016 storage pledged: 1200TB
  - Actually ~900TB are available and the remaining will be soon deployed
  - Used 626 TB (70%)
  - Native XRootD version 4.0.4
- CPU (ALICE and opportunistic use):
  - 2016 HS06 pledged: 10500 (~1100 Jobs)
  - Max Running Jobs: 5231
  - Avg Running Jobs: ~1900
  - Efficiency: 91,5%
- Testbed for «ALICE Dashboard» and «Monitoring system for large and federated datacenters» (see Gioacchino's slides)

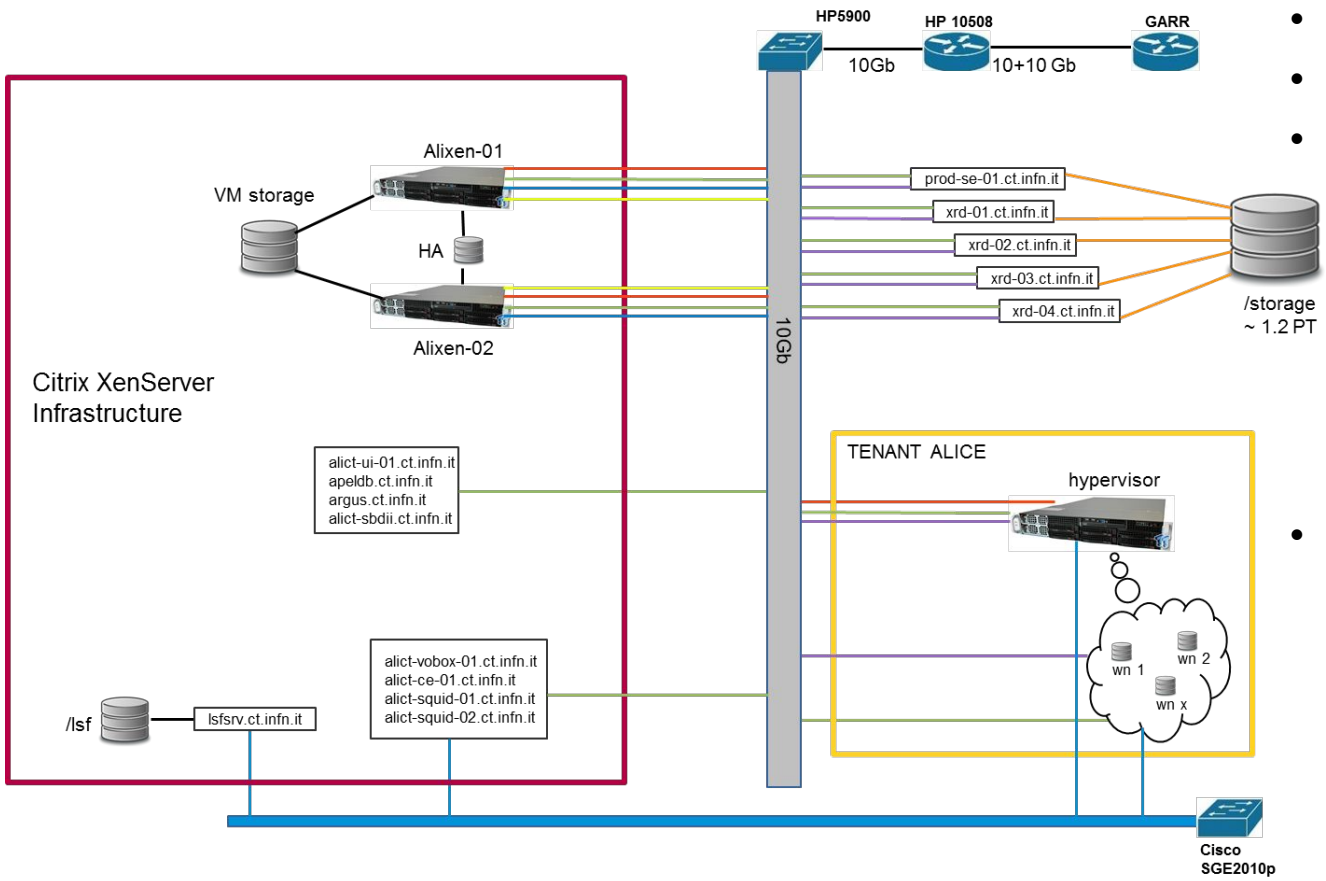
Tier-2 sites
Catania



ALICE INFN-CATANIA T2 Virtualization:

- Citrix XenServer pool
- running VMs for central services
  - CE (LSF Batch Server)
  - VOBOX, site BDII, UI, ARGUS, APELDB, CVMFS SQUID
- VMs per WNs
  - 5 VMs for each hypervisor
  - 14(8) core, 56(24) GB RAM 100GB disk
  - 160 VMs providing 2048 virtual cores
- OpenStack in HA + Zen LB
  - Controller Node, RabbitMQ, MariaDB (Galera Cluster), Keystone
  - NetworkNode configured using L2 linuxbridge
  - testing the infrastructure using latest OS release Mitaka

Tier-2 sites
Catania

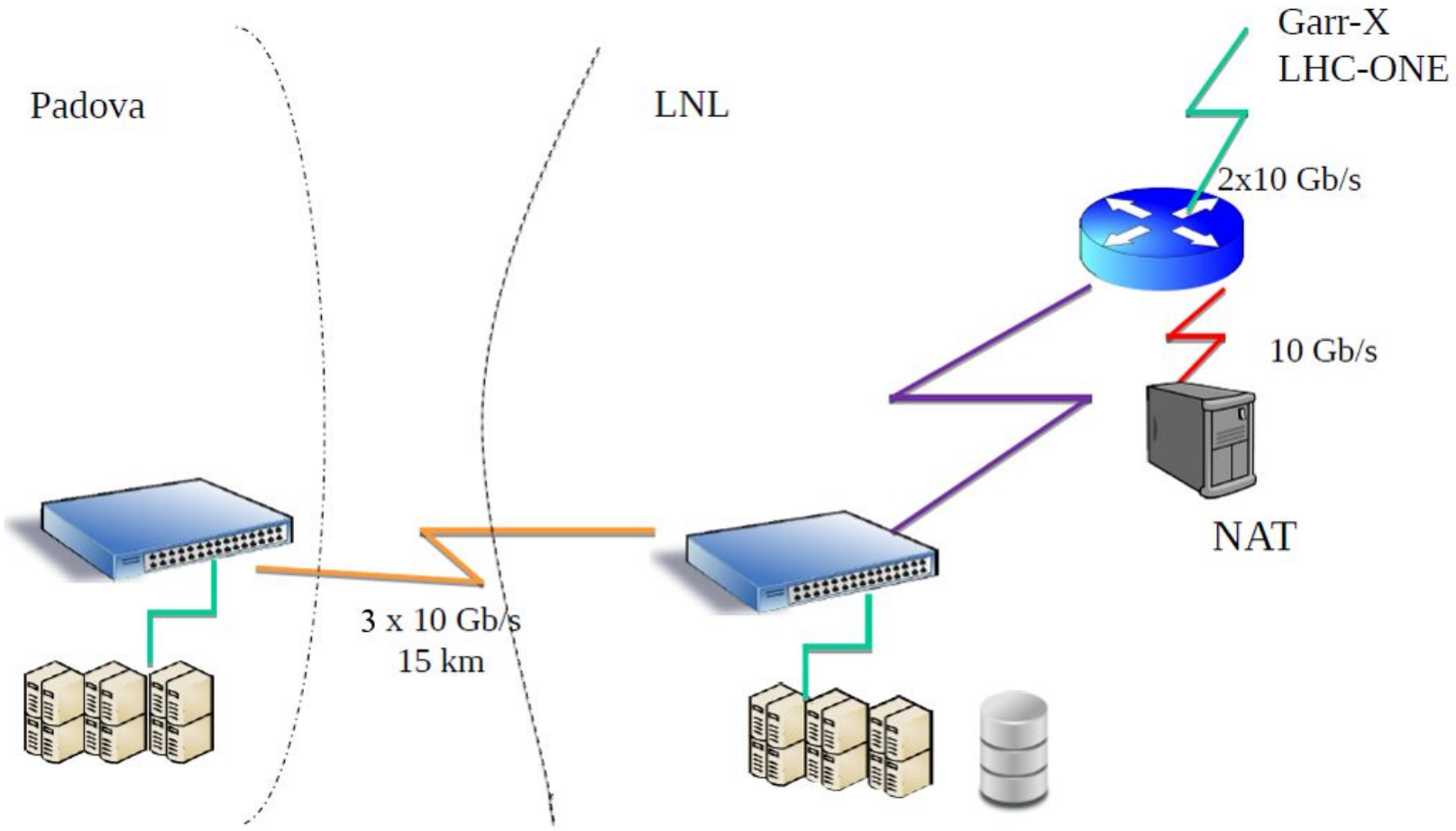


ALICE::CATANIA::SE:

- ~1.2PB (92% used)
- GPFS: v3.5.0.10
- 5 xrootd server (1 redirector)
  - xrootd v4.3.0
  - network bandwidth 50Gbps (10 x 2Gbps)
- 20 Gbps link GARR-X/LHC-ONE

Fibre Channel
PUBLIC – 90.147.17.0/24 – 10 Gb
LSF – VLAN 1921 - 192.168.80.0/23 – 1 Gb
Mysql - Galera – VLAN 3306 – 172.16.17.0/24 - 1Gb
XROOTD VLAN 1025 – 10.255.0.0/23 – 10 Gb
Openstack management – VLAN 1010 - 10.10.0.0/24 – 1 Gb

Tier-2 sites
Padova-LNL



Tier-2 sites
Padova-LNL

Tier-2 spread in 2 sites:

- INFN Legnaro Nat. Lab (LNL) and INFN Padova to share resources, infrastructures, manpower
- Tier-2 for both ALICE and CMS experiments
- Other VOs can use the resources if not used by ALICE and CMS

Computing services:

- 6 CREAM CE in LNL(used by all VOs)
- VOBOX in Legnaro
- 2 squid servers (for CVMFS): one in LNL, one in Padova
- Monitoring services (Ganglia, nagios, cacti) in Padova
- Site BDII in LNL
- Argus server (used only for glexec) in LNL

Computing resources and services:

- All resources managed by a single LSF cluster
- ALICE resources: 67 WNs, 1864 cores, 18584 HS06
- Possibility to use CMS resources, when not used by CMS: 156 WNs, 3356 cores, 36535 HS06
- Actually CMS migrated to multi-core jobs, and therefore the sharing of resources between ALICE and CMS is more difficult now

Storage:

- “Pure” xrootd (v. 4.2.3)
- 1 redirector and 6 servers (all in Legnaro)
- No intermediate layers
- ~ 1.128 PB (~ 98 % used)

Tier-2 sites
Torino

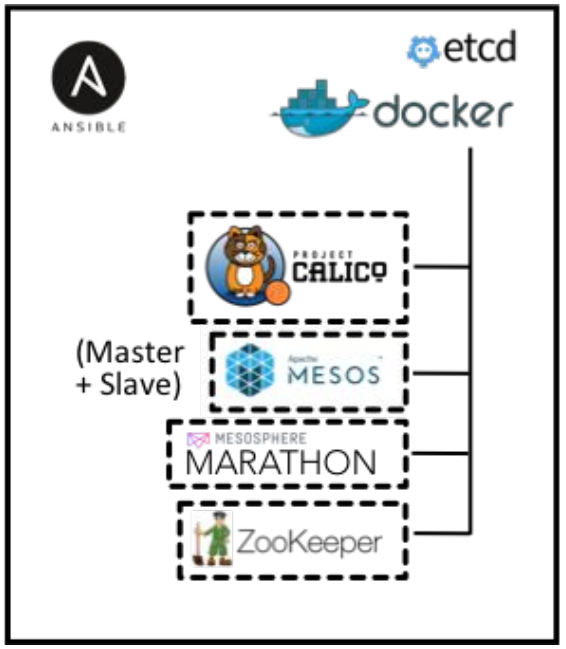
Tier-2 site:

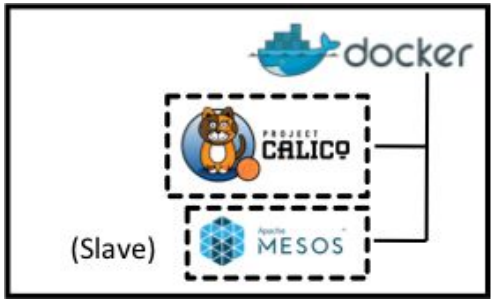
- Nothing really new here…
- Private Cloud Infrastracture hosts Tier2 site and several [Elastic] Virtual Clusters
- OpenNebula software stack
- 9(+1) Gbps link GARR-X/LHC-ONE
- Total Storage 2.2 PB (+360 TB coming soon) including 1.1 PB for xrootd

Tier-2 sites
Torino

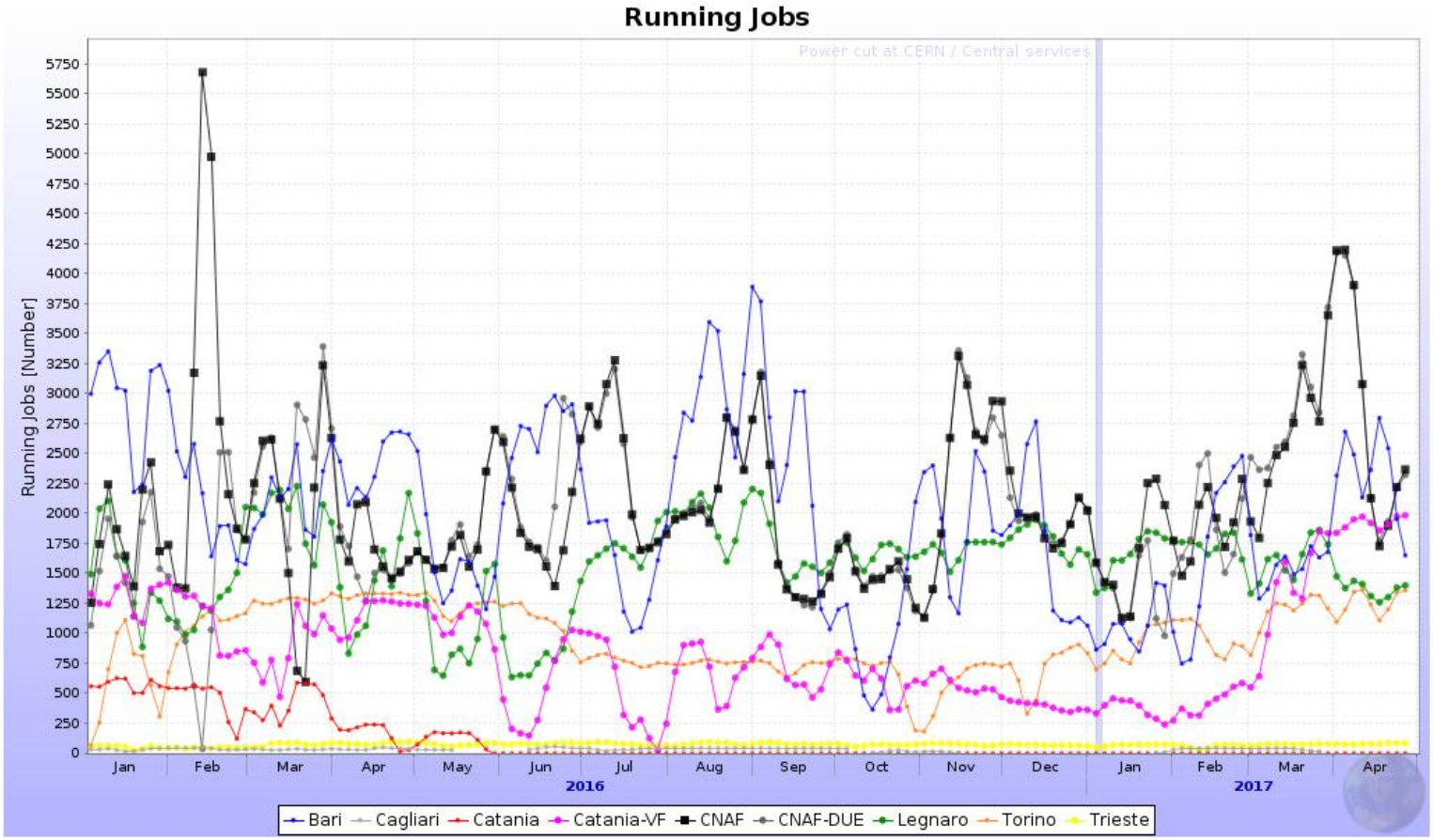
C3S Project:

- A joint Competence Centre for Scientific Computing with UniTO managing OCCAM, a multipurpose HPC facility:
  - –About 1000 CPU cores
  - 4x dual-GPU nodes
  - 4x 4-socket high-memory nodes
  - 256 TB high performance Lustre storage
  - 768 TB expandable archival storage
  - 56Gb/s InfiniBand + 10Gb/s ethernet
  - *À la page* resource provisioning model through containers & Docker
- Resources are NOT for LHC
  - But some activities will be possible (e.g. GPU code development and testing)
  - And opportunistic usage through Plancton (see Matteo's talk this morning)

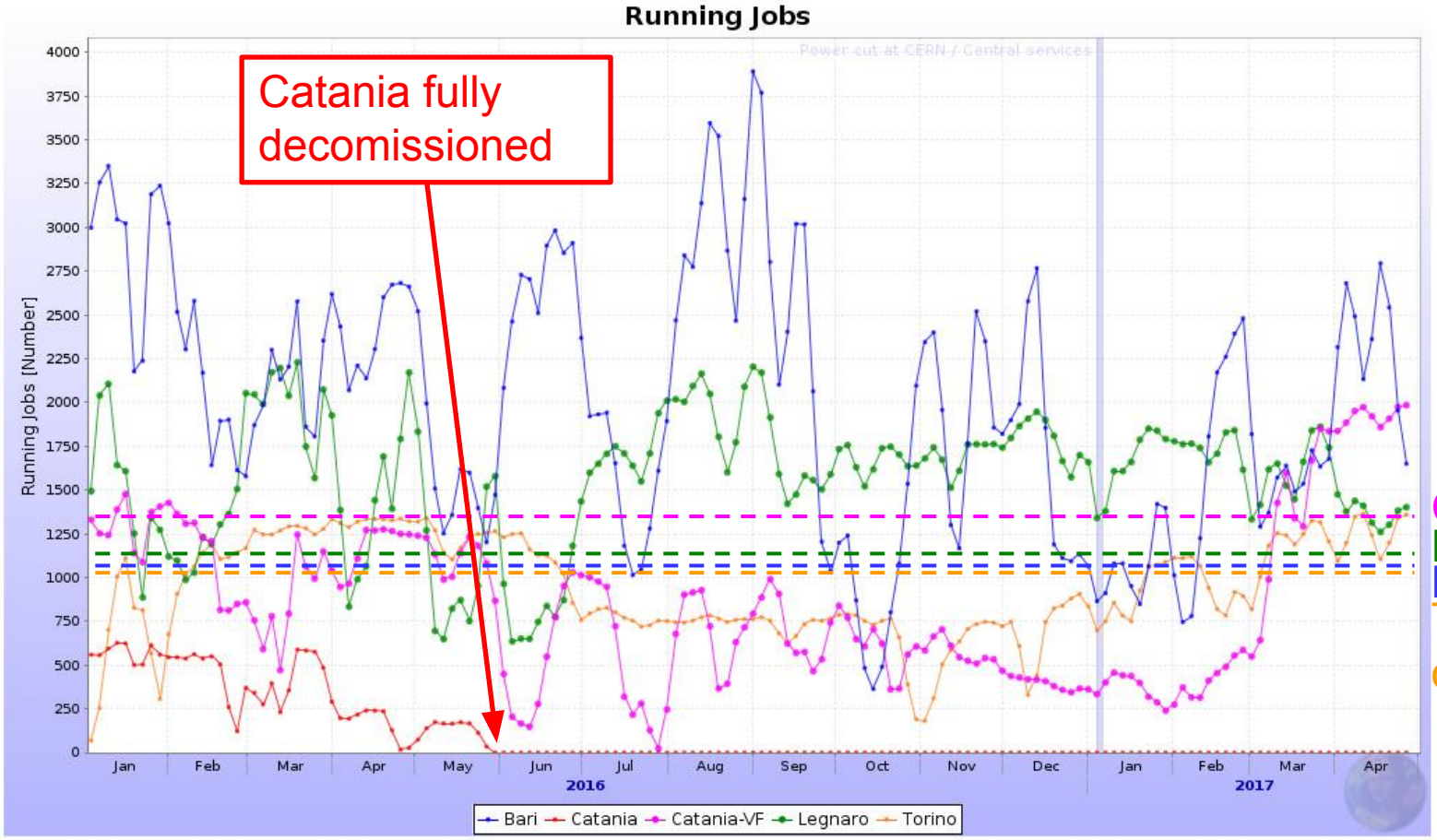




Tier-1 and Tier-2 sites
Running jobs profile



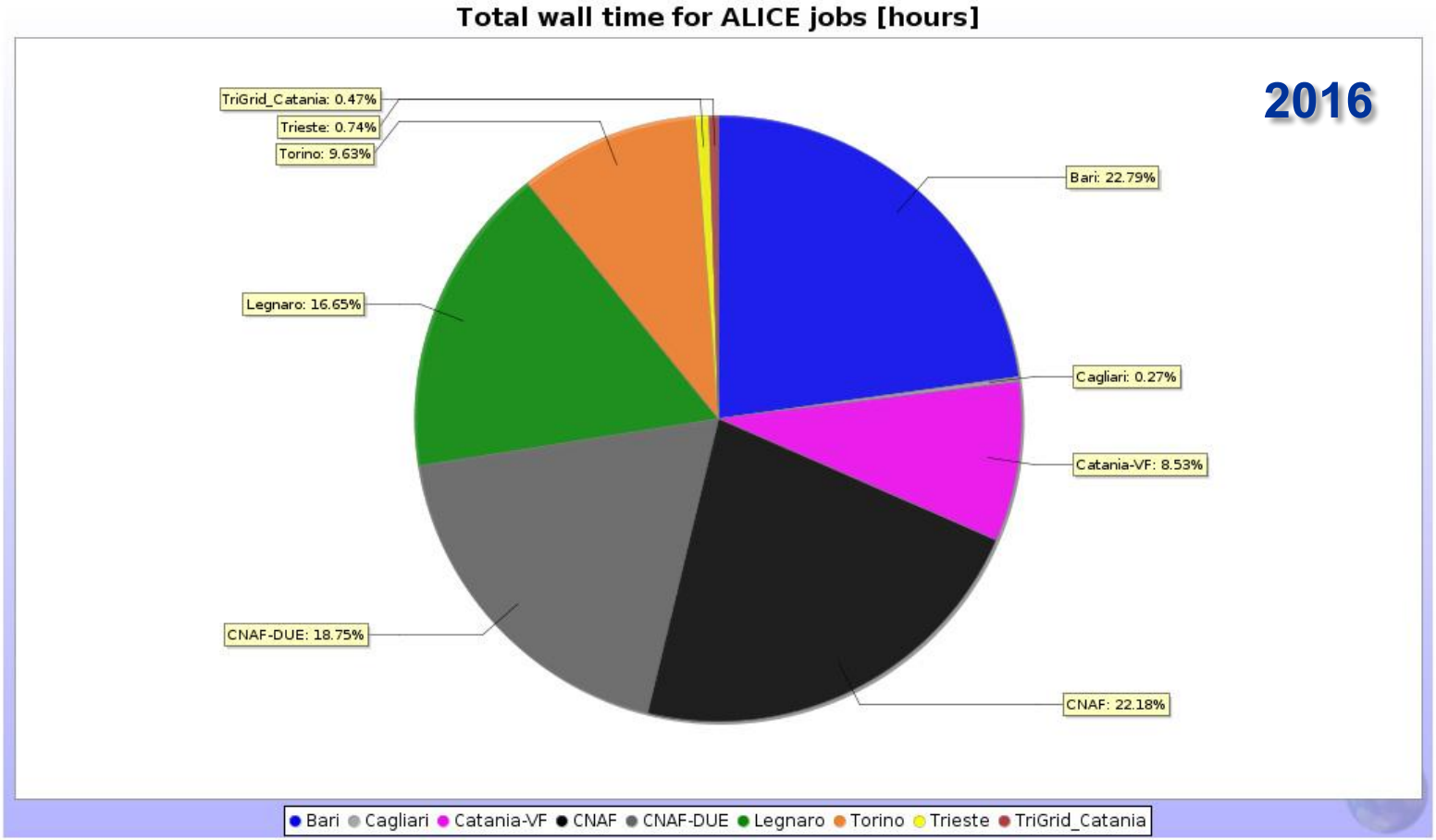
Tier-2 sites
Running jobs profile



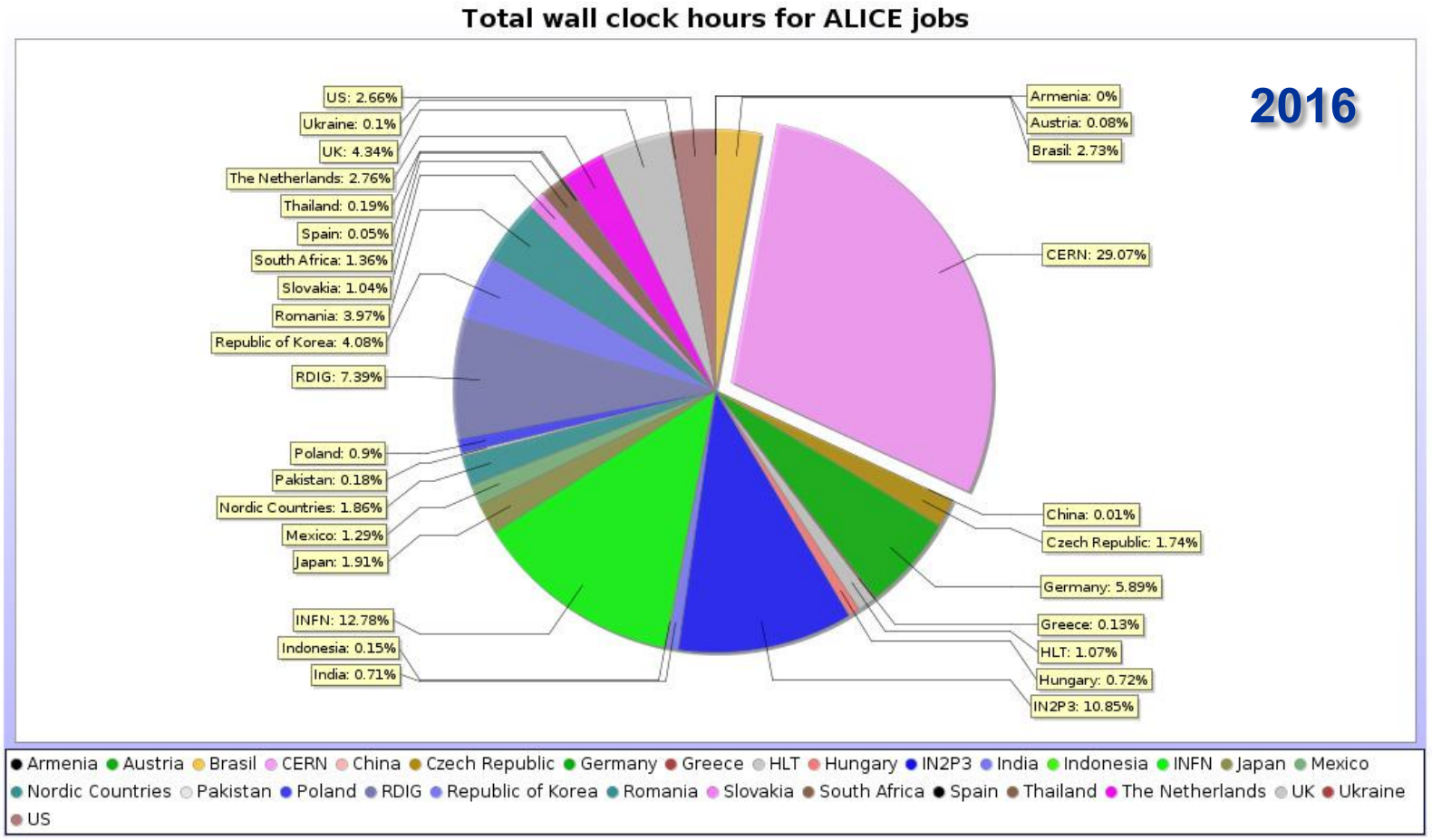
Pledge:

Catania-VF
PD-LNL
Bari
Torin
o

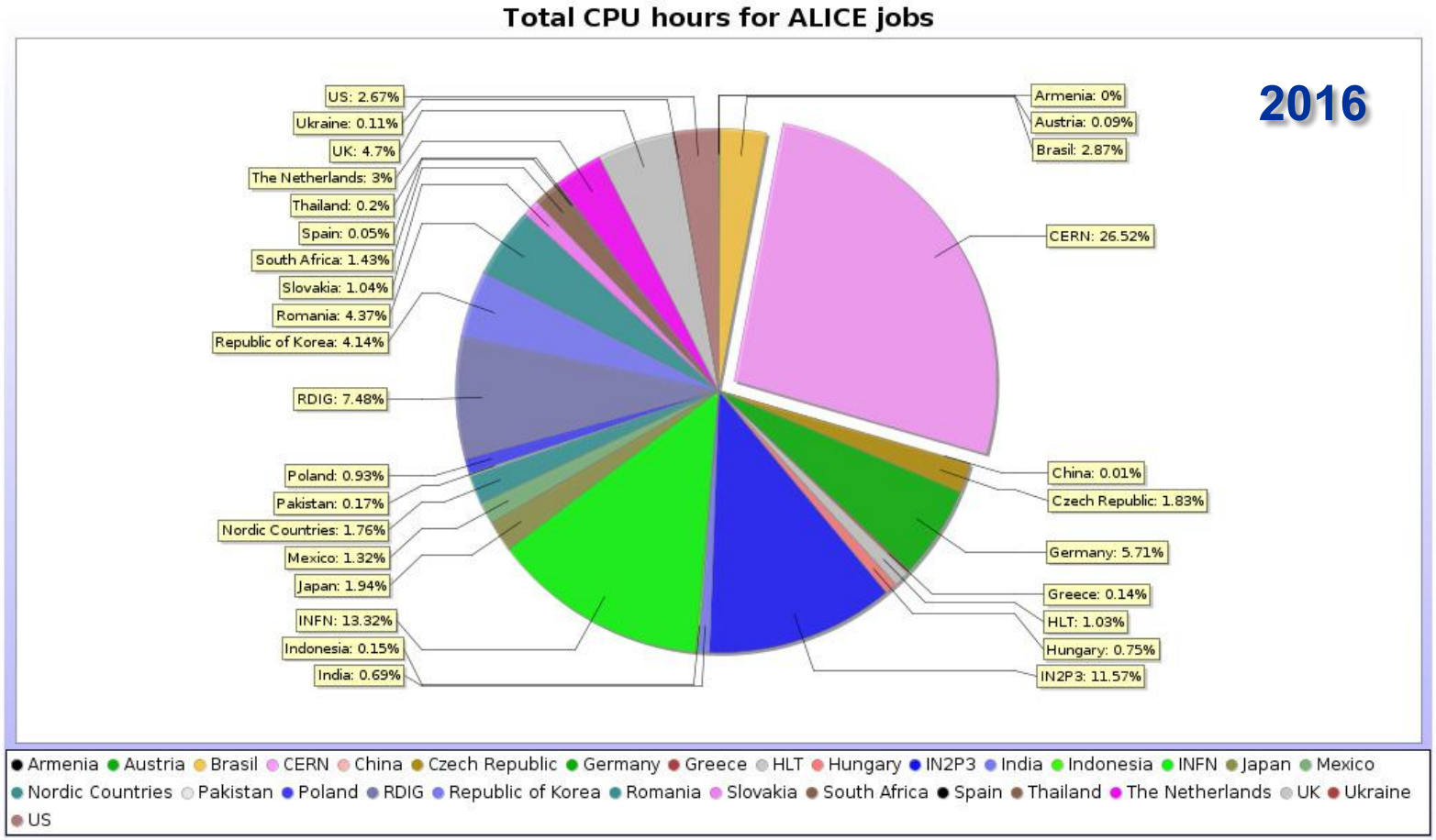
Tier-1 and Tier-2 sites
WallTime PieChart



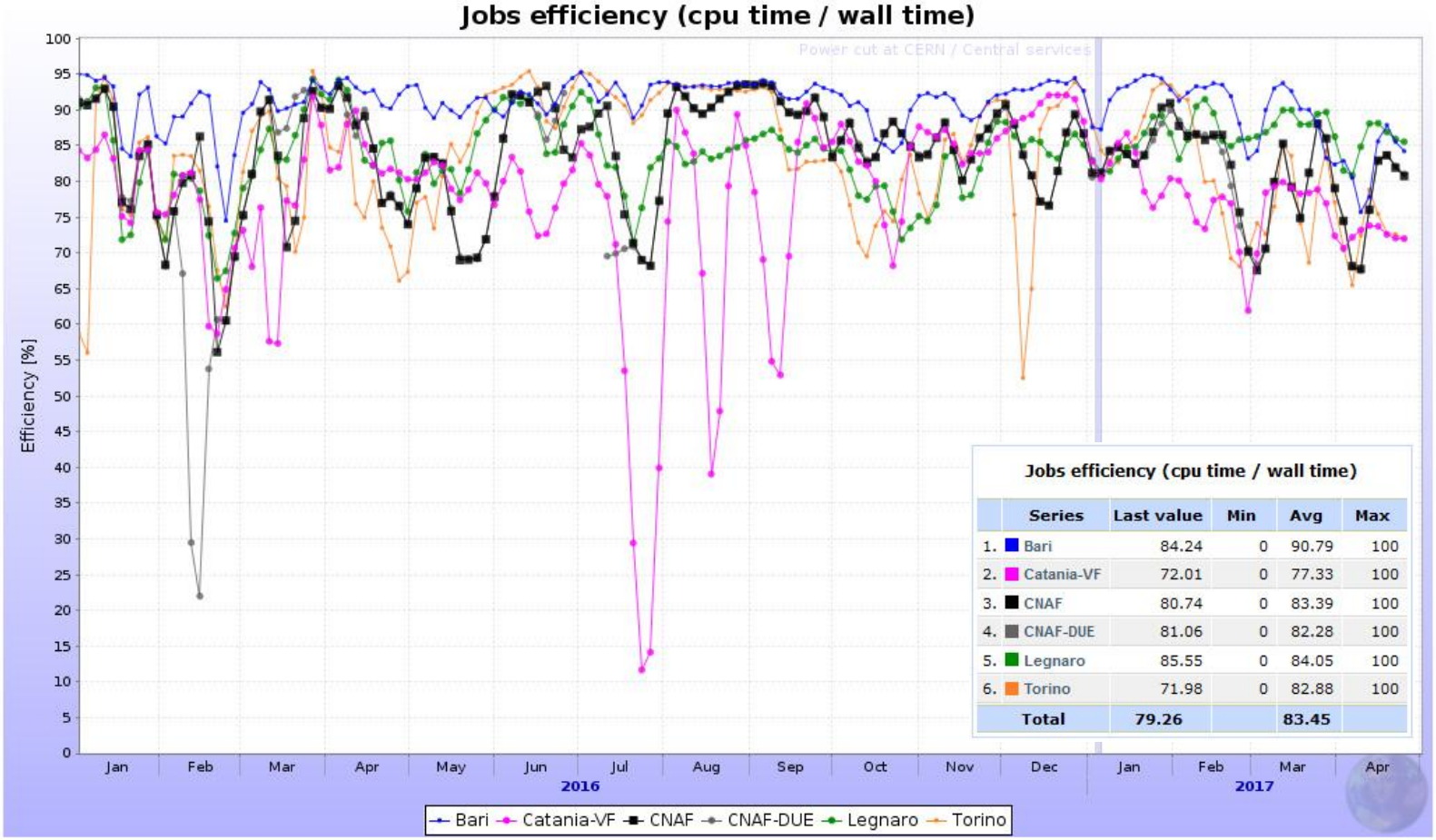
Tier-1 and Tier-2 sites
WallTime PieChart



Tier-1 and Tier-2 sites
CPUTime PieChart



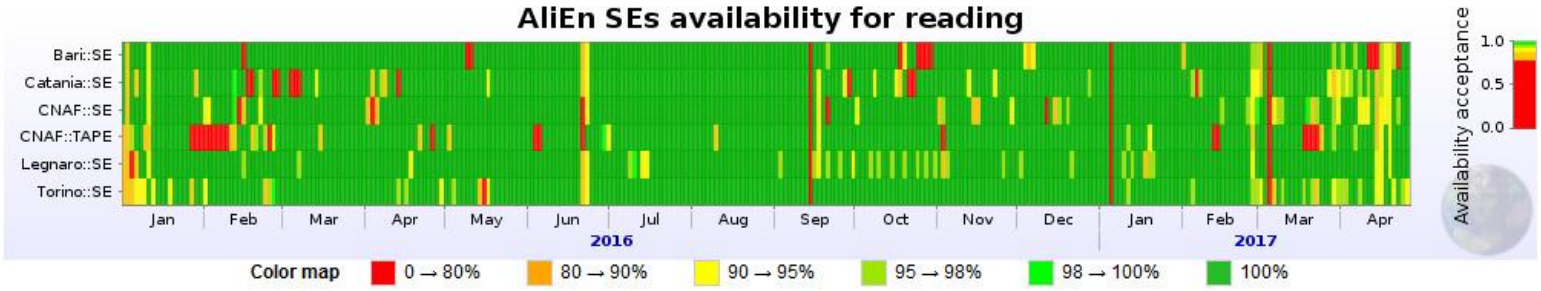
Tier-1 and Tier-2 sites
CPU Efficiency



Jobs efficiency (cpu time / wall time)

	Series	Last value	Min	Avg	Max
1.	Bari	84.24	0	90.79	100
2.	Catania-VF	72.01	0	77.33	100
3.	CNAF	80.74	0	83.39	100
4.	CNAF-DUE	81.06	0	82.28	100
5.	Legnaro	85.55	0	84.05	100
6.	Torino	71.98	0	82.88	100
	Total	79.26		83.45	

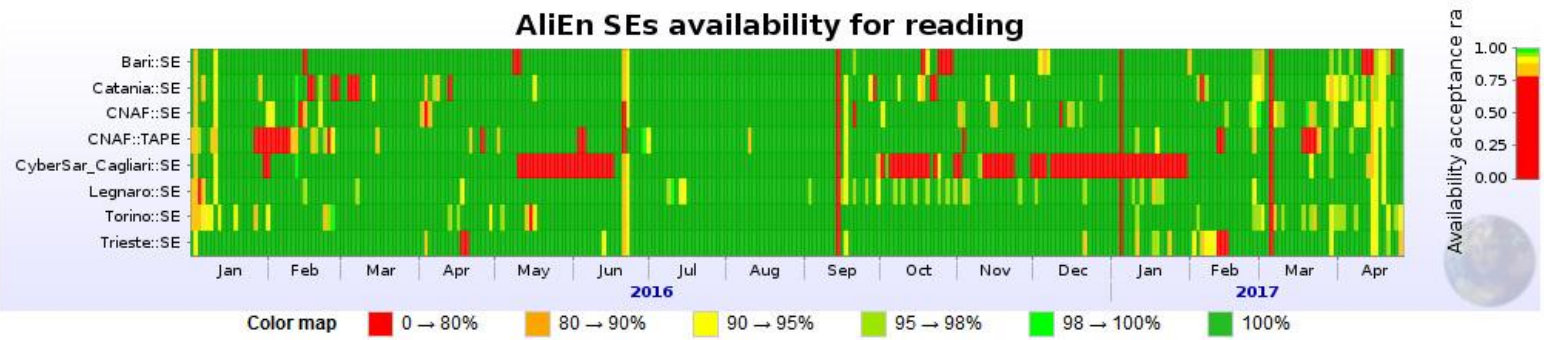
Tier-1 and Tier-2 sites
SE Availability



Statistics

Link name	Data		Individual results of reading tests			Overall
	Starts	Ends	Successful	Failed	Success ratio	Availability
Bari::SE	01 Jan 2016 00:00	27 Apr 2017 00:03	5612	170	97.06%	97.07%
Catania::SE	01 Jan 2016 00:03	27 Apr 2017 00:03	5577	205	96.45%	96.49%
CNAF::SE	01 Jan 2016 00:07	27 Apr 2017 00:06	5671	110	98.10%	98.16%
CNAF::TAPE	01 Jan 2016 00:07	27 Apr 2017 00:07	5555	226	96.09%	96.15%
Legnaro::SE	01 Jan 2016 00:18	27 Apr 2017 00:19	5699	81	98.60%	98.61%
Torino::SE	01 Jan 2016 00:30	27 Apr 2017 00:33	5692	89	98.46%	98.53%

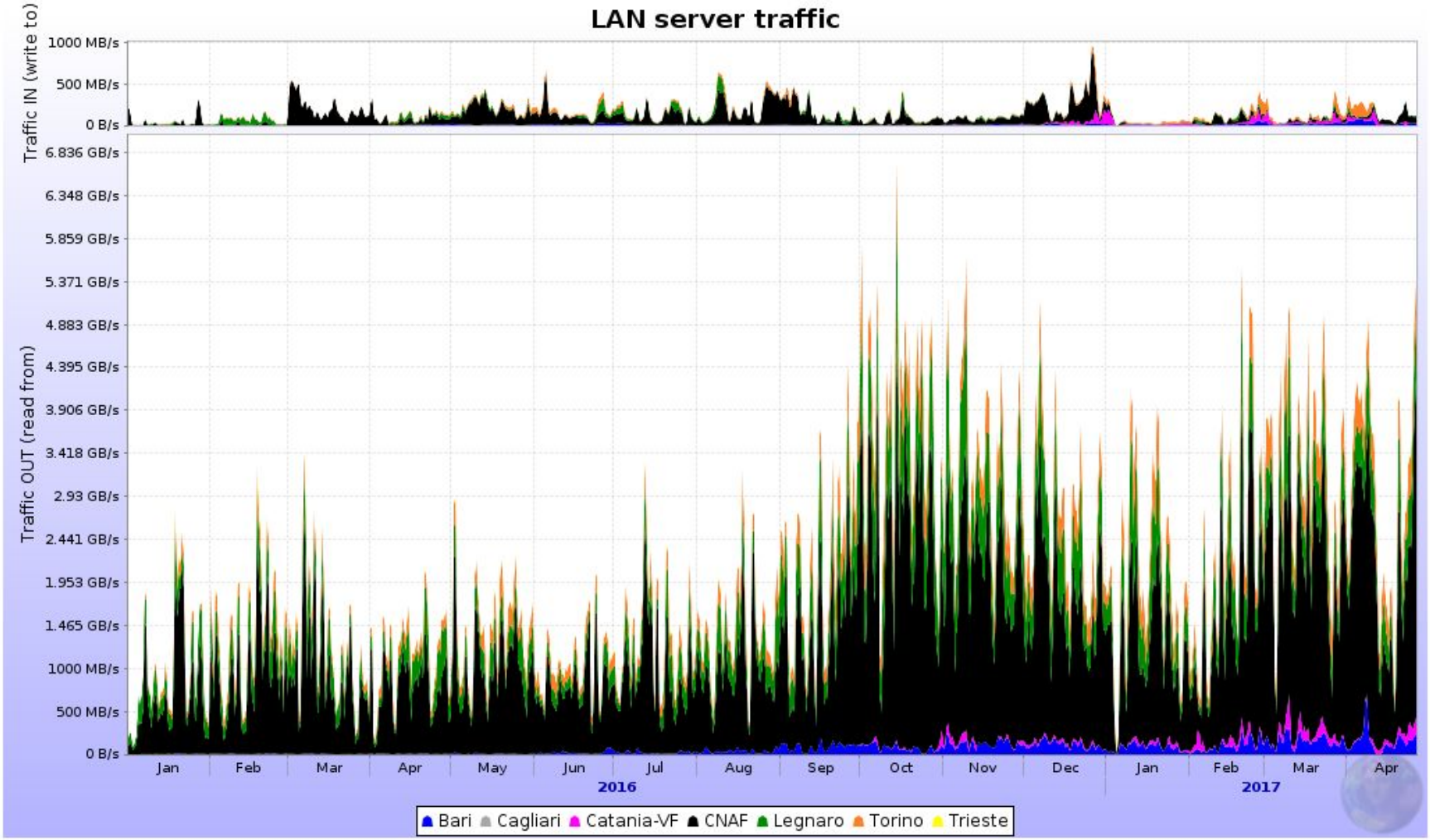
All Italian sites
SE Availability



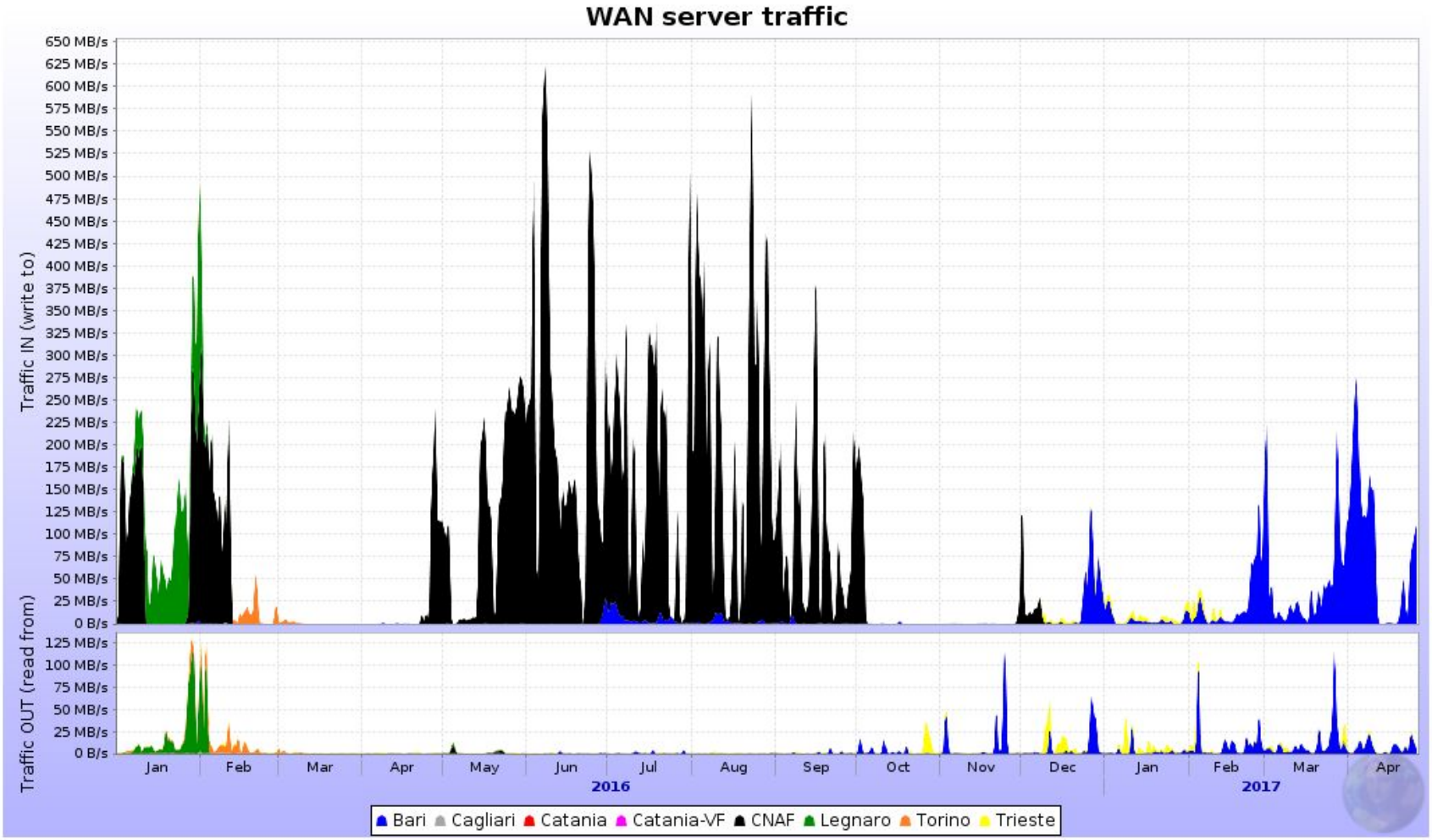
Statistics

Link name	Data		Individual results of reading tests			Overall
	Starts	Ends	Successful	Failed	Success ratio	Availability
Bari::SE	01 Jan 2016 00:00	27 Apr 2017 00:03	5612	170	97.06%	97.07%
Catania::SE	01 Jan 2016 00:03	27 Apr 2017 00:03	5577	205	96.45%	96.49%
CNAF::SE	01 Jan 2016 00:07	27 Apr 2017 00:06	5671	110	98.10%	98.16%
CNAF::TAPE	01 Jan 2016 00:07	27 Apr 2017 00:07	5555	226	96.09%	96.15%
CyberSar_Cagliari::SE	01 Jan 2016 00:08	27 Apr 2017 00:07	4216	1565	72.93%	72.88%
Legnaro::SE	01 Jan 2016 00:18	27 Apr 2017 00:19	5699	81	98.60%	98.61%
Torino::SE	01 Jan 2016 00:30	27 Apr 2017 00:33	5692	89	98.46%	98.53%
Trieste::SE	01 Jan 2016 00:31	27 Apr 2017 00:31	5688	93	98.39%	98.45%

All Italian sites
Network: LAN



All Italian sites
Network: WAN



All Italian sites
Network: LAN vs WAN

LAN server traffic

Traffic IN

	Series	Last value	Min	Avg	Max	Total
1.	Bari	24.53 MB/s	0 B/s	7.738 MB/s	1013 MB/s	306.9 TB
2.	Cagliari	0 B/s	0 B/s	0.989 KB/s	5.636 MB/s	39.23 GB
3.	Catania-VF	0.947 MB/s	0 B/s	19.12 MB/s	3.016 GB/s	331.2 TB
4.	CNAF	72.67 MB/s	0 B/s	100.2 MB/s	9.422 GB/s	3.88 PB
5.	Legnaro	20.02 MB/s	0 B/s	15.47 MB/s	3.645 GB/s	613.8 TB
6.	Torino	35.11 KB/s	0 B/s	16.85 MB/s	1.629 GB/s	667.8 TB
7.	Trieste	0 B/s	0 B/s	0.247 MB/s	112.5 MB/s	9.795 TB
	Total	**118.2 MB/s**		**159.6 MB/s**		**5.765 PB**

Traffic OUT

	Series	Last value	Min	Avg	Max	Total
1.	Bari	236.7 MB/s	0 B/s	58.73 MB/s	2.879 GB/s	2.275 PB
2.	Cagliari	0 B/s	0 B/s	0.447 MB/s	56.46 MB/s	17.73 TB
3.	Catania-VF	190.7 MB/s	0 B/s	44.47 MB/s	2.648 GB/s	770.3 TB
4.	CNAF	3.865 GB/s	0 B/s	1.313 GB/s	15.58 GB/s	52.09 PB
5.	Legnaro	718.6 MB/s	0 B/s	408.5 MB/s	10.21 GB/s	15.83 PB
6.	Torino	396.4 MB/s	0 B/s	217.2 MB/s	3.323 GB/s	8.406 PB
7.	Trieste	4.458 MB/s	0 B/s	5.286 MB/s	820.7 MB/s	209.7 TB
	Total	**5.376 GB/s**		**2.031 GB/s**		**79.57 PB**

WAN server traffic

Traffic IN

	Series	Last value	Min	Avg	Max	Total
1.	Bari	113.5 MB/s	0 B/s	10.83 MB/s	1.382 GB/s	429.5 TB
2.	Cagliari	10.89 KB/s	0 B/s	24.39 KB/s	63.98 MB/s	967.4 GB
3.	Catania	0 B/s	0 B/s	0 B/s	0 B/s	0 B
4.	Catania-VF	0 B/s	0 B/s	0 B/s	0 B/s	0 B
5.	CNAF	0 B/s	0 B/s	65.99 MB/s	2.587 GB/s	2.557 PB
6.	Legnaro	0 B/s	0 B/s	4.38 MB/s	1.037 GB/s	173.7 TB
7.	Torino	0 B/s	0 B/s	0.44 MB/s	372.7 MB/s	17.43 TB
8.	Trieste	2.783 KB/s	0 B/s	0.568 MB/s	683.3 MB/s	22.54 TB
	Total	**113.5 MB/s**		**82.23 MB/s**		**3.186 PB**

Traffic OUT

	Series	Last value	Min	Avg	Max	Total
1.	Bari	5.324 MB/s	0 B/s	3.159 MB/s	2.336 GB/s	125.3 TB
2.	Cagliari	0 B/s	0 B/s	0.184 MB/s	65.81 MB/s	7.29 TB
3.	Catania	0 B/s	0 B/s	0 B/s	0 B/s	0 B
4.	Catania-VF	0 B/s	0 B/s	1.557 KB/s	6.716 MB/s	26.97 GB
5.	CNAF	0 B/s	0 B/s	67.47 KB/s	291.7 MB/s	2.614 TB
6.	Legnaro	0 B/s	0 B/s	1.647 MB/s	951.8 MB/s	65.35 TB
7.	Torino	0 B/s	0 B/s	0.61 MB/s	195.1 MB/s	24.17 TB
8.	Trieste	1.206 MB/s	0 B/s	1.35 MB/s	954 MB/s	53.55 TB
	Total	**6.53 MB/s**		**7.017 MB/s**		**278.3 TB**

Conclusions

Thanks
for
your attention !