Status report from Tokyo Tier2 at ICEPP

Tomoe Kishimoto

ICEPP, The University of Tokyo

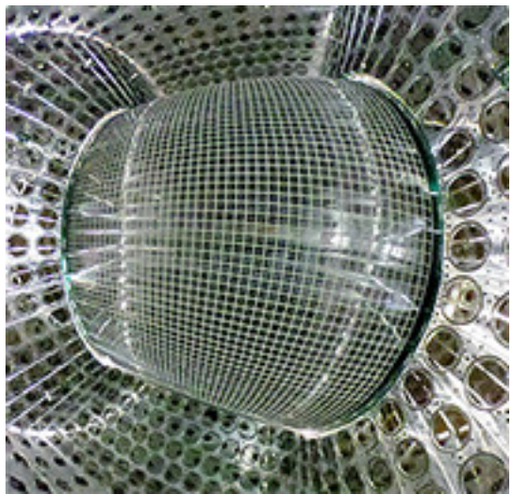
Jan. 29 2017

International Center for Elementary Particle Physics

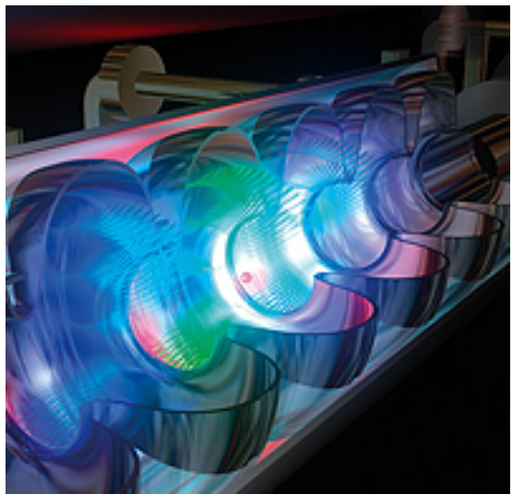
東京大学
素粒子物理国際研究センター
International Center for Elementary Particle Physics
The University of Tokyo

✓ Main projects in ICEPP

ATLAS experiment at LHC



MEG experiment at PSI (μ→eγ rare decay)



R & D for ILC

ATLAS Japan

✓ ATLAS-Japan group
- 17 institutes and ~150 members
- Tokyo Tier2 is the only WLCG site in ATLAS-Japan

ICEPP regional analysis center

✓ Resource overview

- Support only ATLAS VO in WLCG (**Tier2**) and provide ATLAS-Japan dedicated resources (local use)
- Hardwares are leased, and are replaced in every three years
- ~10000 CPU cores including service instances and ~10 PB disk storage (T2 + local use)
  - 18.11HS06/core (Intel Xenon E5-2680 v3)

4th system (2016-2019)

Single VO and uniform architecture

✓ Operation team

- H.Sakamoto (will retire in next Mar.), J.Tanaka, T.Mashimo, N.Tomoaki, T.Kishimoto, N.Matsui

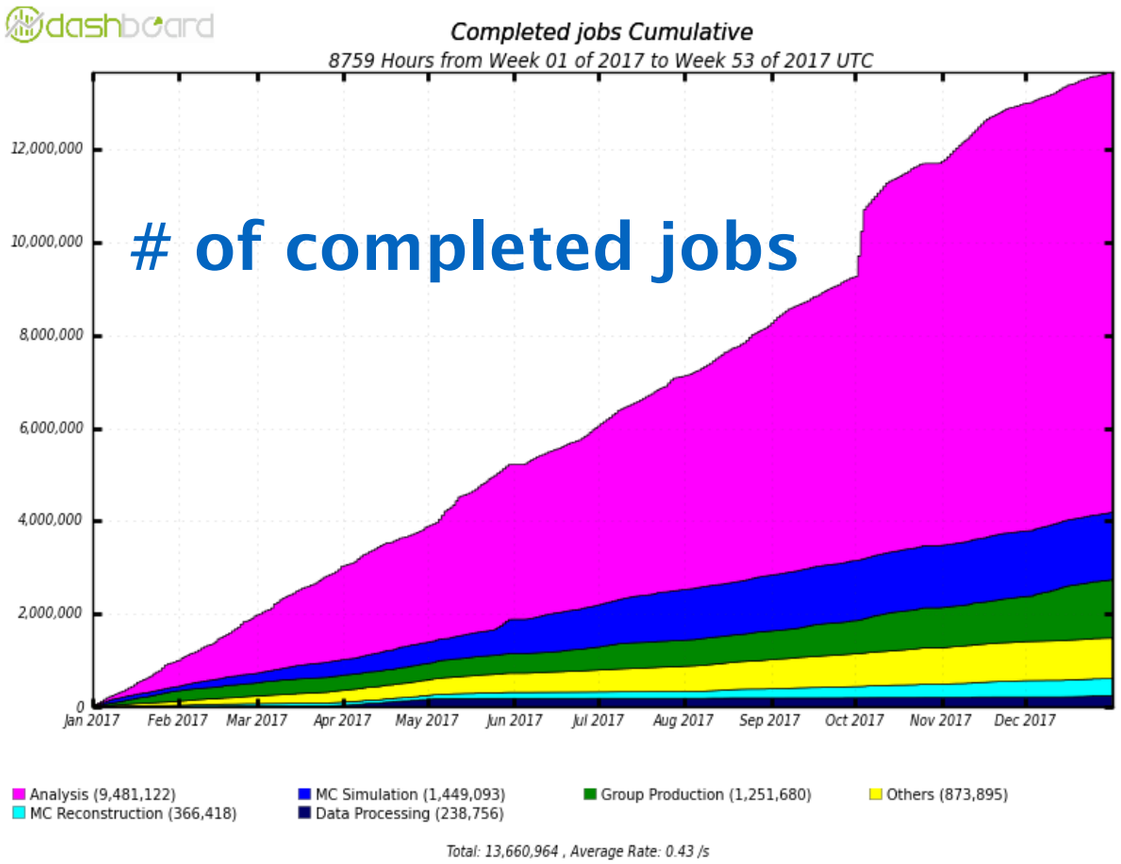
WLCG pledge

		CPU [HS06]	DISK [TB]	(*)LOCALGROUPDISK [TB]
2017	Pledge	34,000	4,000	-
	Deployed	111,268	4,000	1,000
2018	Pledge	40,000	4,800	-
	Deployed	111,268	4,800	1,000

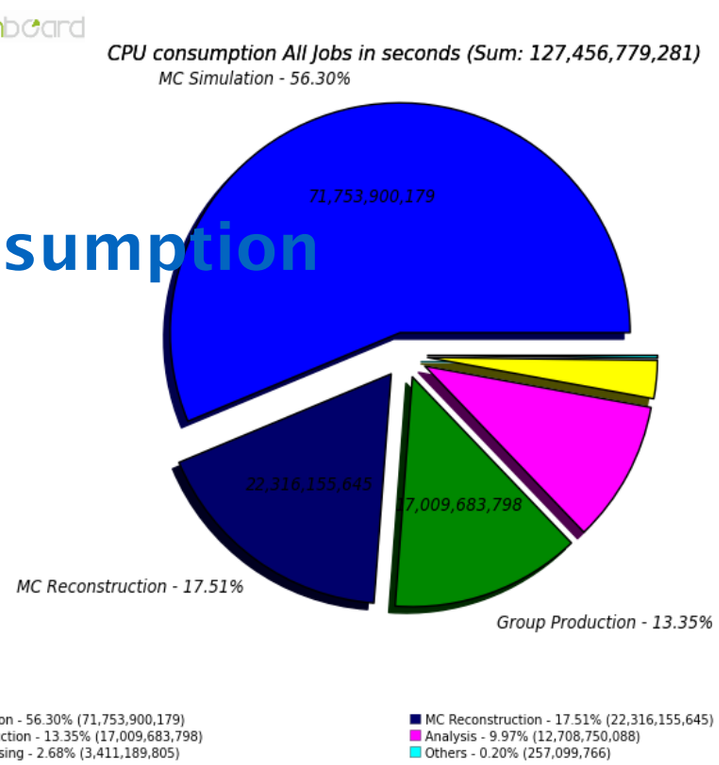
(*) Grid disks for ATLAS-Japan group

- ✓ Tier2 resources
  - The current system (4th system) satisfies 2018 WLCG pledge
  - New system will be provided for 2019-2021
    - (Need to migrate 5.8 PB data to the new system...)

Site status in ATLAS



# of completed jobs



CPU consumption

- ✓ Fraction of # of completed jobs for the last year:
  - Production: **4.0% (Tier2)** – 2.2% (All)
  - Analysis: **6.3% (Tier2)** – 4.1% (All)
    ← Good contributions

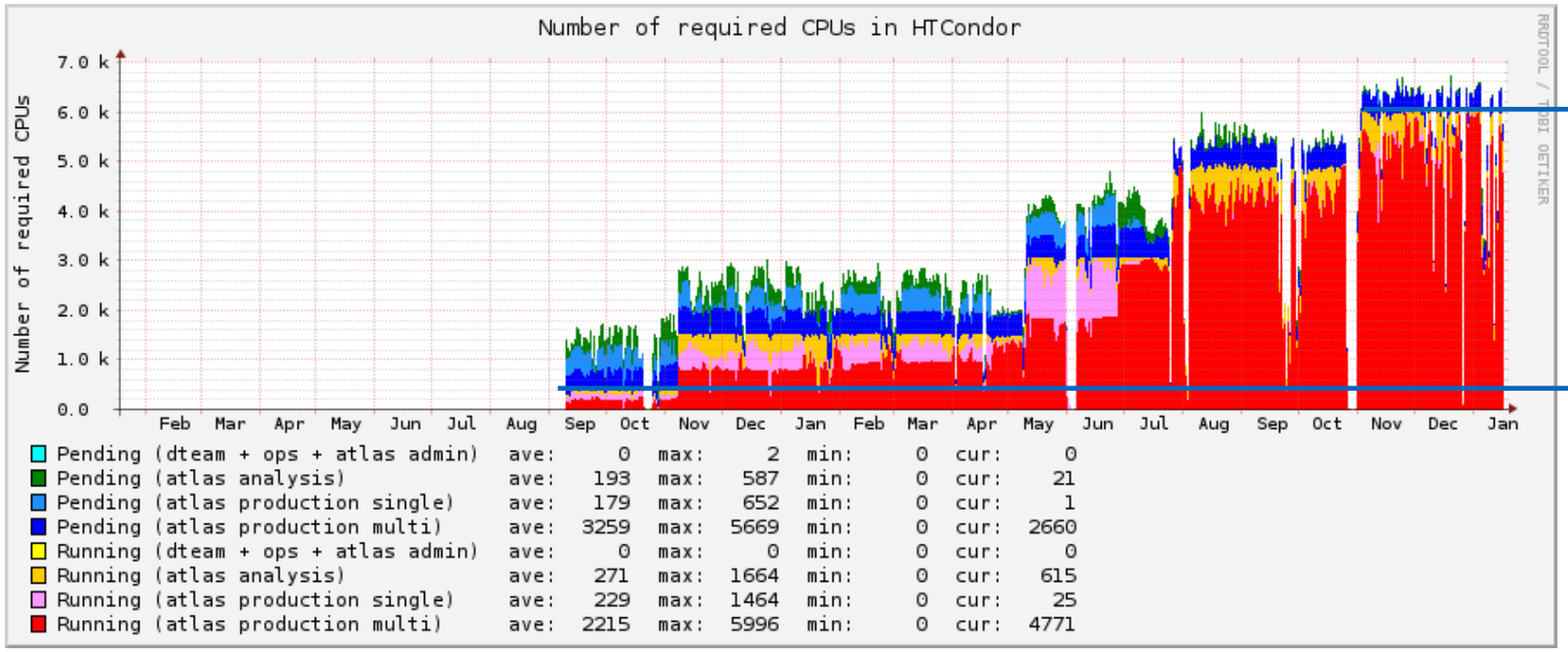
# of ATLAS-J authors ∼ 150
# of ATLAS authors ∼ 3000

- ✓ > 99% site availability has been achieved using the 4th system (for 2 years)

CE and batch system update

✓ Migration from “CREAM+Torque/Maui" to “ARC+HTCondor” has been completed

HTCondor pool occupancy



6144 CPU cores deployed

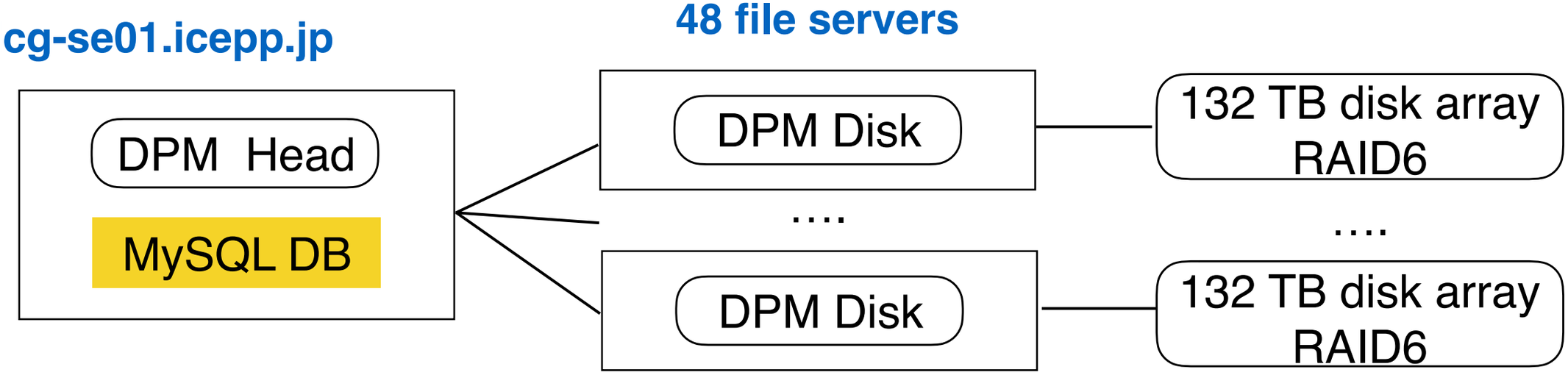
384 CPU cores deployed

✓ Introduced dynamic partitioning for single- and multi-core jobs

- Improvement of CPU utilization was observed
- (Reported at AFAD2017, see backup)

SE and database update

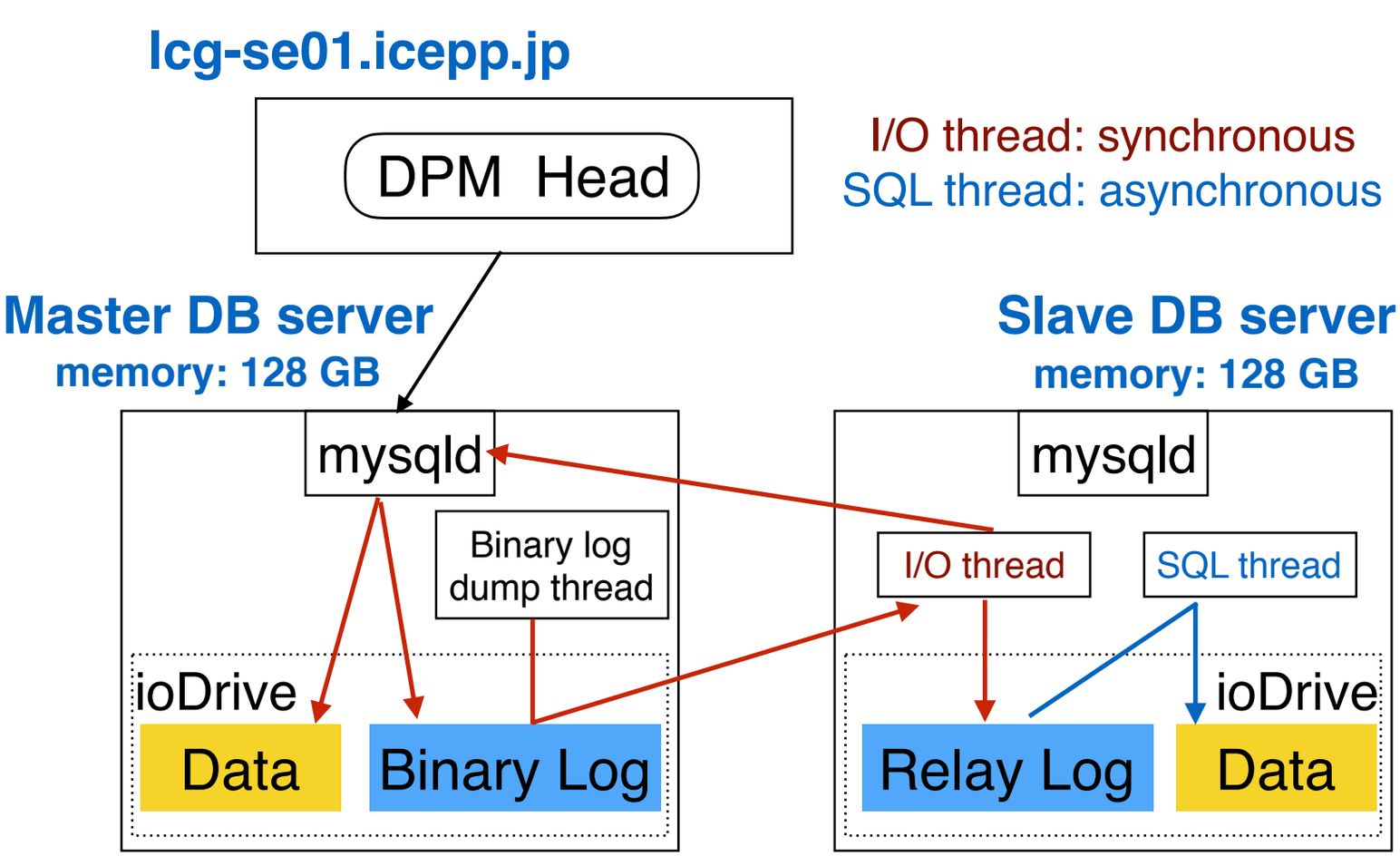
- ✓ Disk storage is managed by **DPM**, and its database is **MySQL**
- ✓ Previous configuration of SE:



- No redundancy in MySQL database…, risk of producing dark data

MySQL replication

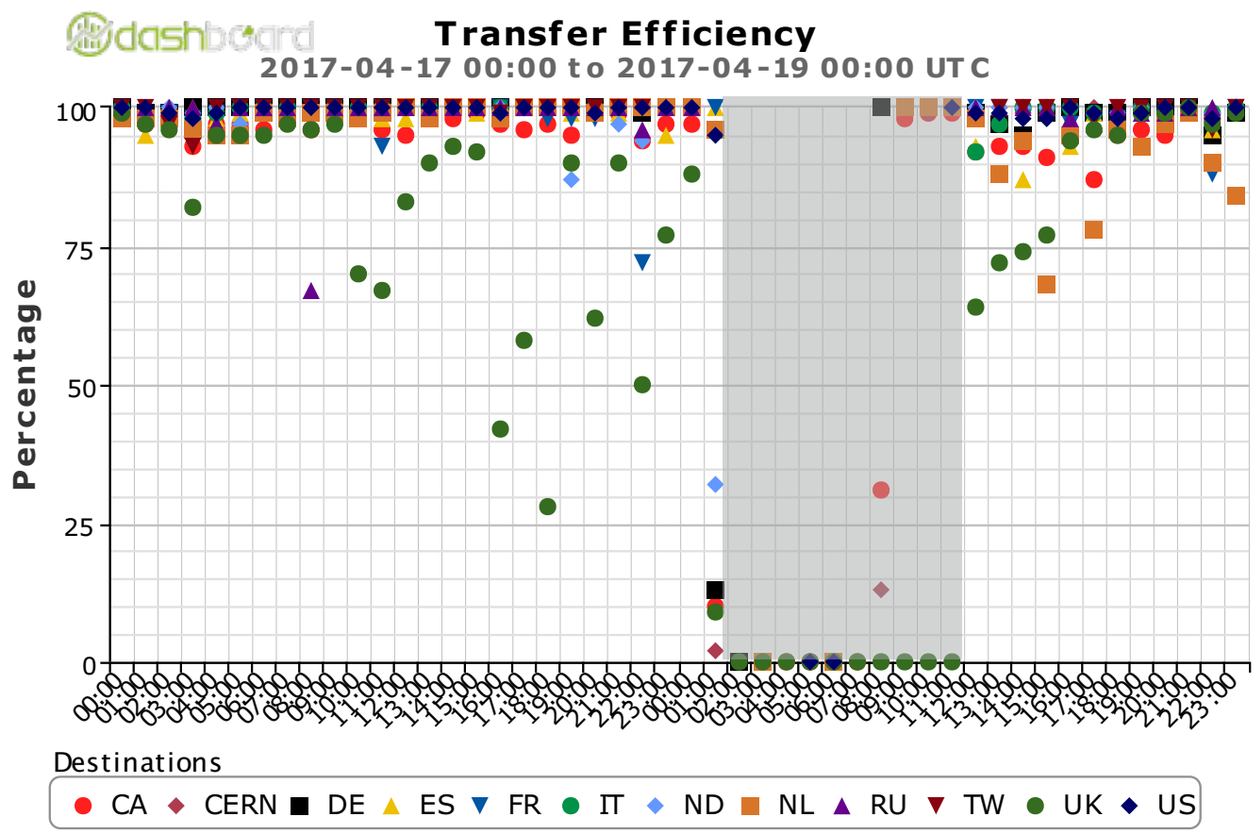
- ✓ Semi-synchronous replication in MySQL has been implemented
  - Master server is replicated to slave server automatically
    → Can use slave server as new master server when a trouble occurs in master server
  - Daily backup from slave server (takes ~10 mins)
    → No impact on master server performances



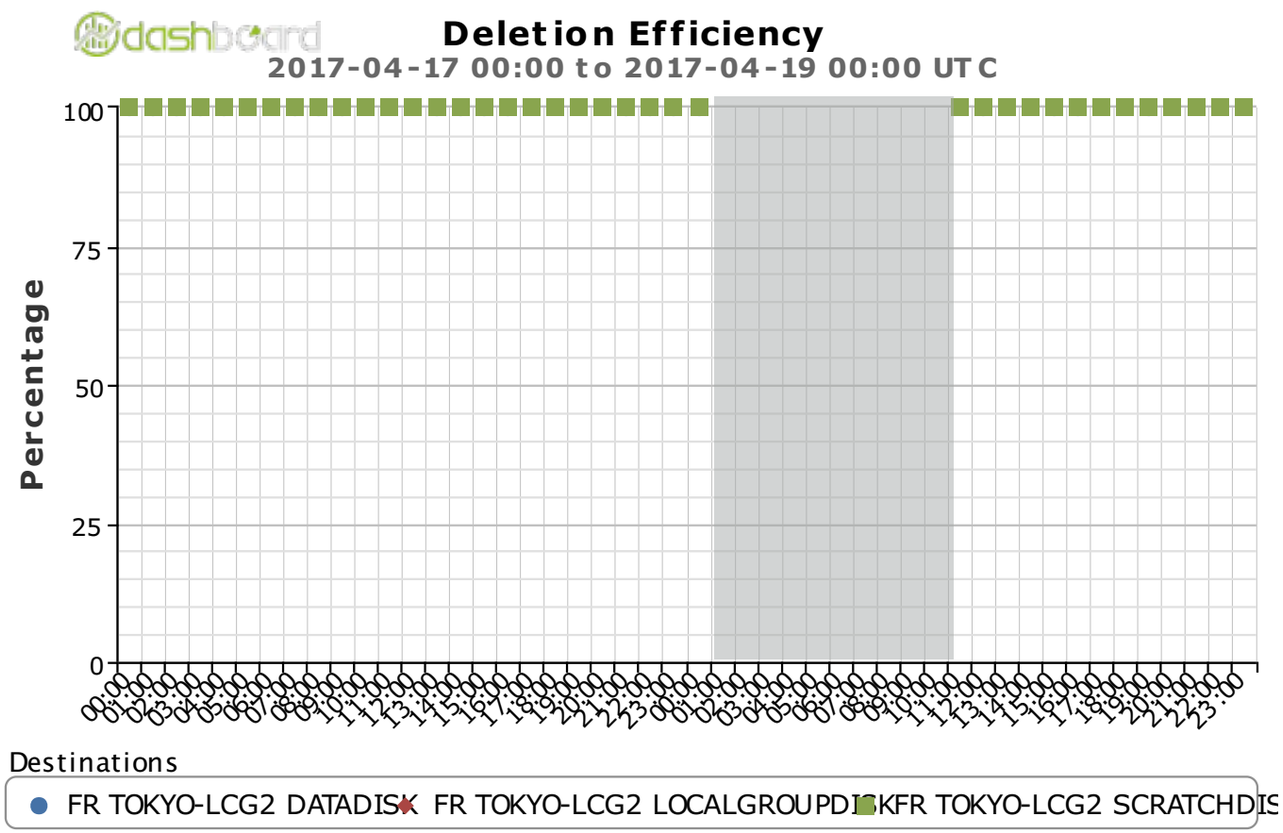
- Fusion-IO ioDrive has been attached for database spaces to reduces time for maintenances
- Binary log increases by 8GB per day

ATLAS data management monitor

Transfer efficiency: source is Tokyo



File deletion efficiency

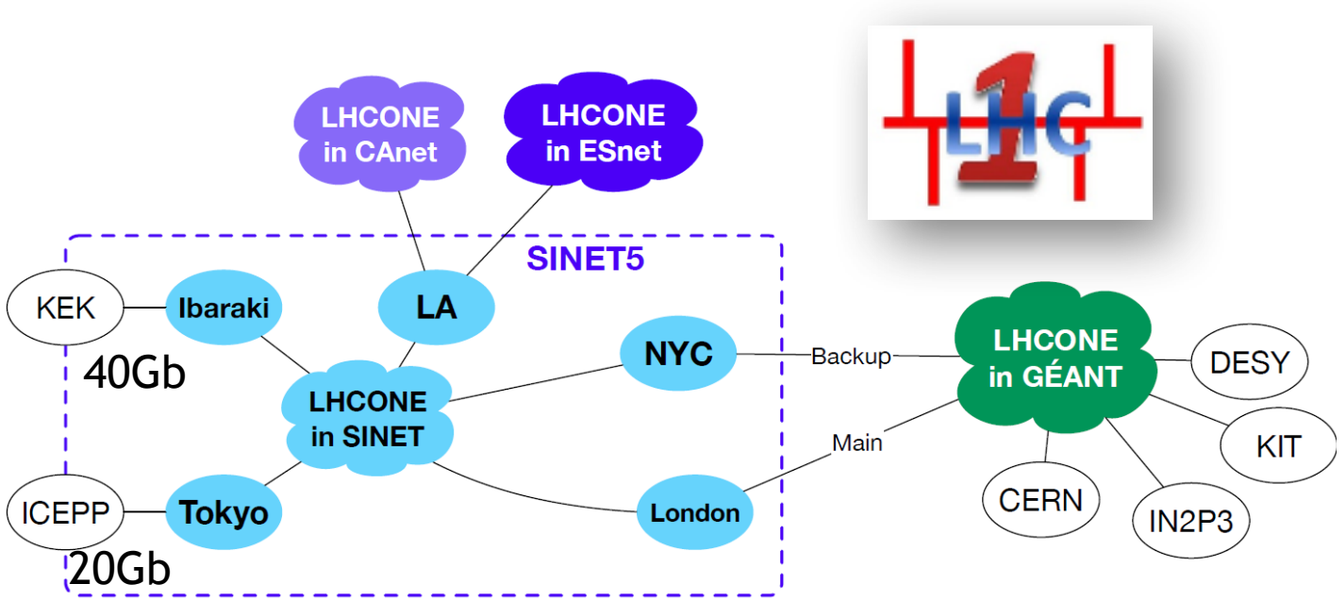


Downtime for the database upgrade

✓ No issues have been observed after the database upgrade

International network status

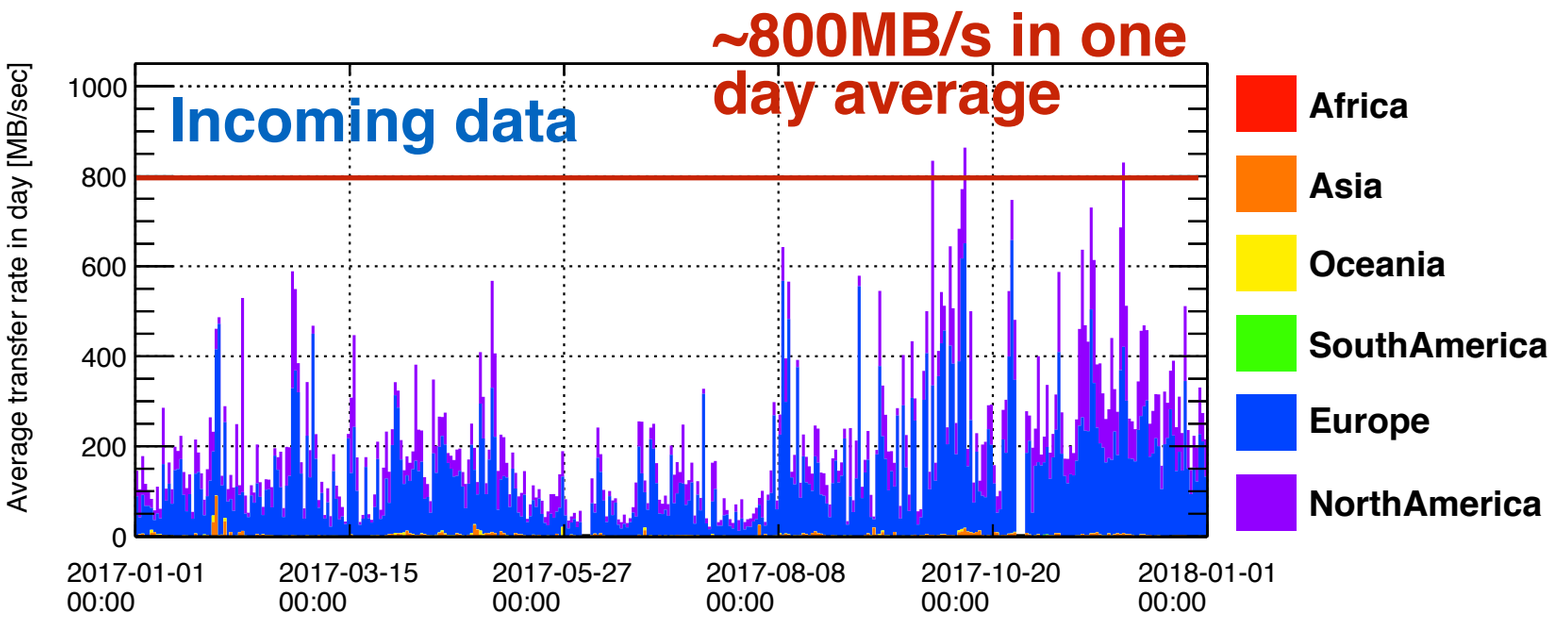
- ✓ SINET5 is a NREN in Japan
  - 2016 Mar. : **20 Gbps for London and 100 Gbps for LA** become available
  - 2016 Apr. : LHCONE peering for EU sites
    - ICEPP⇆CERN latency improved by 30%
  - 2016 Sep. : LHCONE peering for US sites



ICEPP and KEK use common LHCONE VRFs in SINET since 2016 Sep.

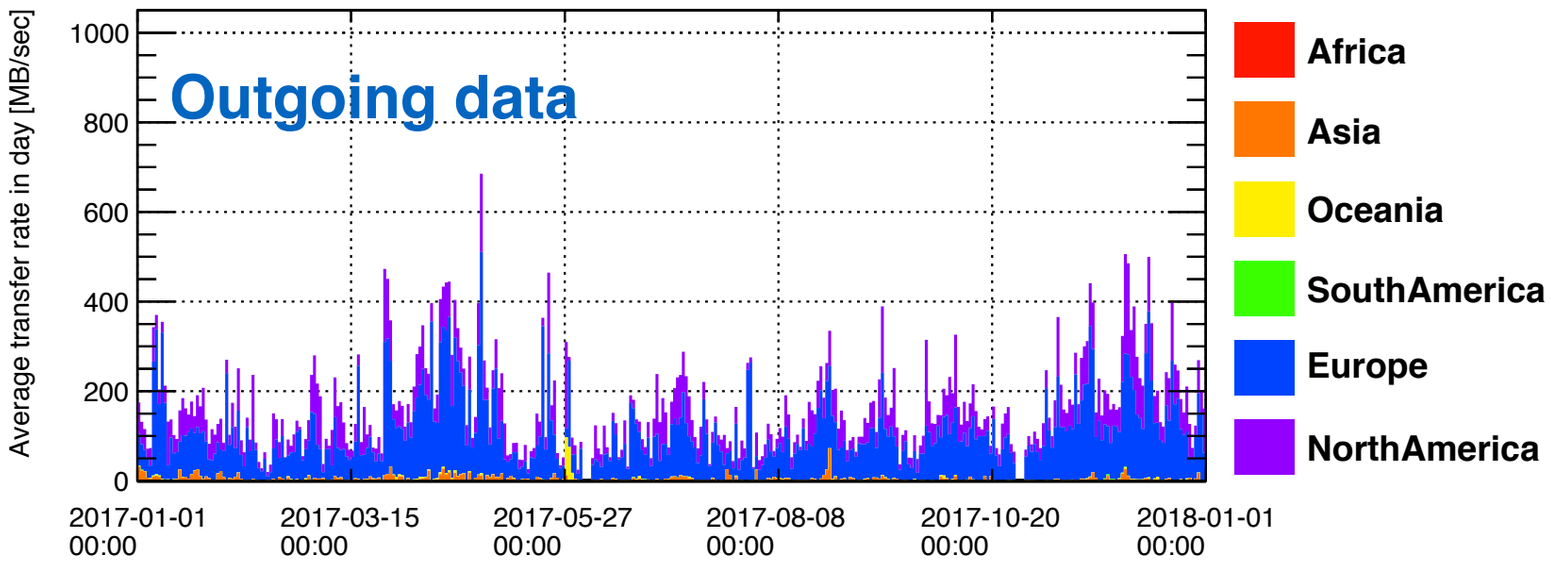
- 2017 Sep. : **LHCONE peering for ASGC, KREONET2 and TEIN via JGN-X VRF in HongKong (100 Gbps for Tokyo ⇄ HongKong)**

Data transfer with other site



Total transfer volumes last year

Europe: 4.2 PB (67 %)
North America: 2.0 PB (32 %)
Asia: 94 TB (2%)



Europe: 2.6 PB (64 %)
North America: 1.3 PB (31 %)
Asia: 206 TB (5%)

Status of IPv6 migration

- ✓ Long pause due to problems of main switch firmware…
  - The firmware was fixed last year, and our procedure/ experience for IPv6 filtering have been matured

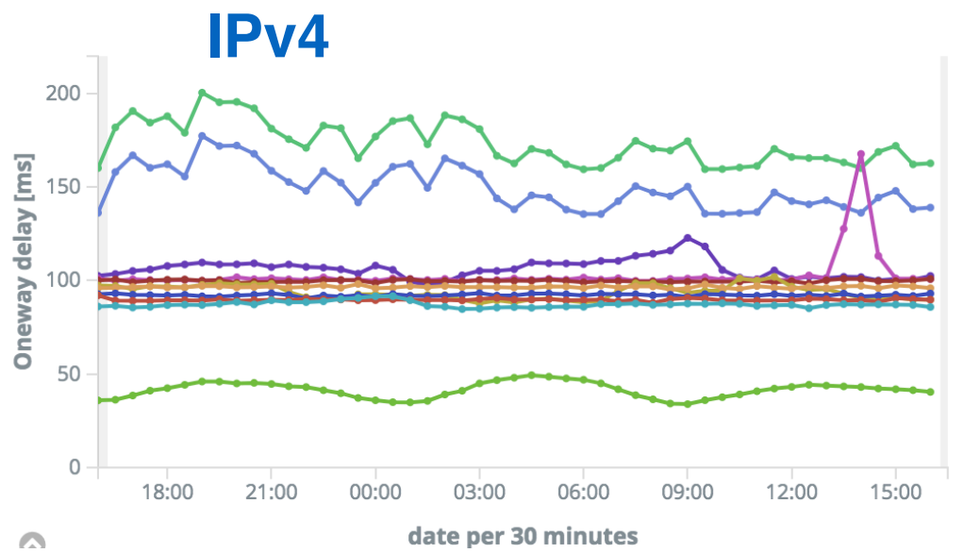
perfSONAR

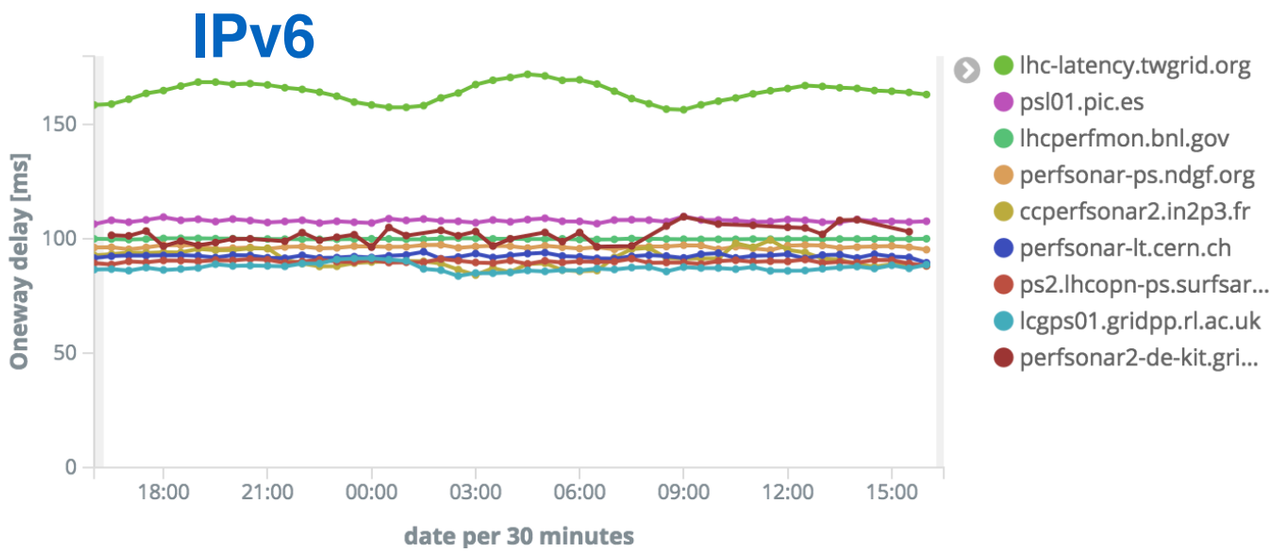
- ✓ IPv6 migration plan:
  1. Enable the dual stack mode of perfSONARs (**done**)
  2. Enable LHCONE peering via IPv6, need to discuss with SINET and University network team (by end of Aug. 2018)
  3. Enable the dual stack mode of storage system (by end of Dec. 2018)

perfSONAR is a key tool to measure IPv6 performances

PerfSONAR measurements

- ✓ Data measured by PerfSONARs are also stored to ELK stack for good visualization
- ✓ Latency tests with ATLAS Tier1s:

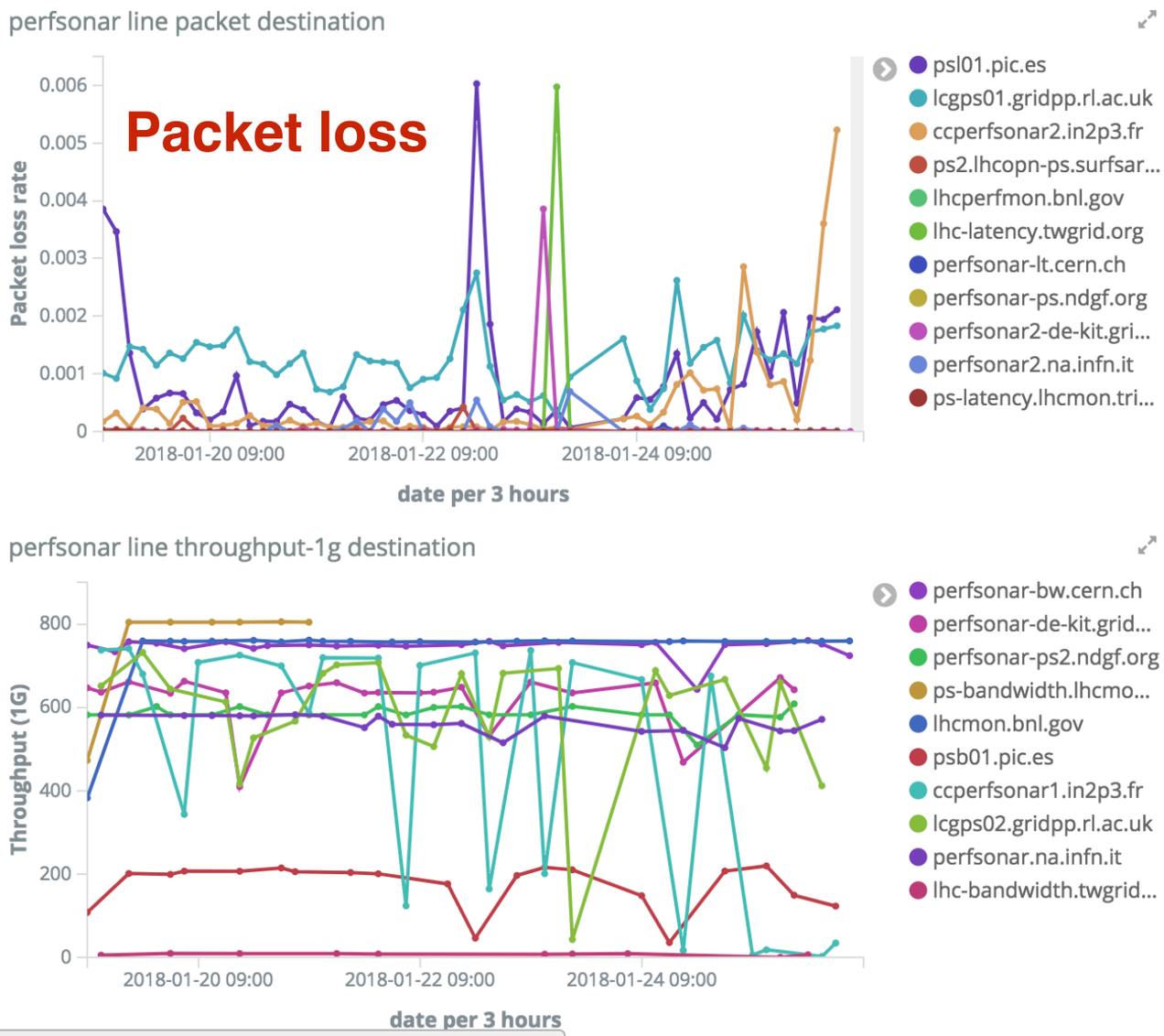




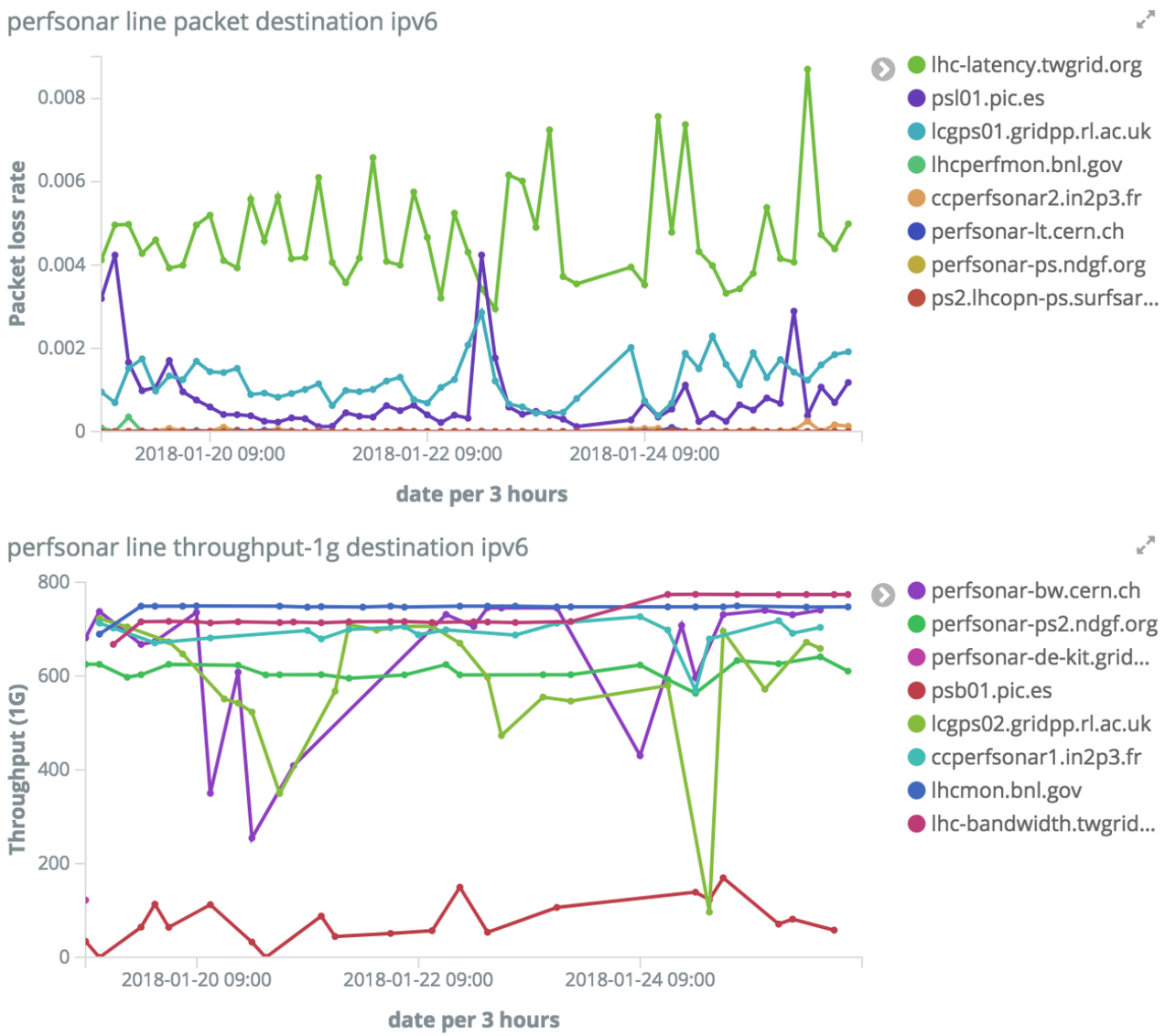
- IPv6 tests are stable so far, but differences of performance are expected since LHCONE peering for IPV6 is not ready yet

PerfSONAR measurements

IPv4



IPv6



Summary

- ✓ Tokyo Tier2 with the 4th system is running
  - Providing enough computing resources for ATLAS
  - > 99% site availability is achieved
- ✓ Migration from Torque/Maui to HTCondor has been completed
- ✓ Redundancy in MySQL database has been implemented
  - Reduced the risk of producing dark data
- ✓ International network connectivity has been improved thanks to Japanese NRENs (SINET and JGN)
- ✓ IPv6 migration is ongoing
  - PerfSONARs are IPv6 ready, tests are working well

Backup

CPU utilization

There was a pbs_server crush

	November 2016				December 2016			
	week1	week2	week3	week4	week1	week2	week3	week4
Static partitioning (Torque/Maui)	-	98.8%	91.5%	95.6%	97.7%	79.7%	99.6%	90.5%
Dynamic partitioning (HTCondor)	-	99.4%	97.0%	98.3%	99.1%	90.4%	99.5%	97.9%

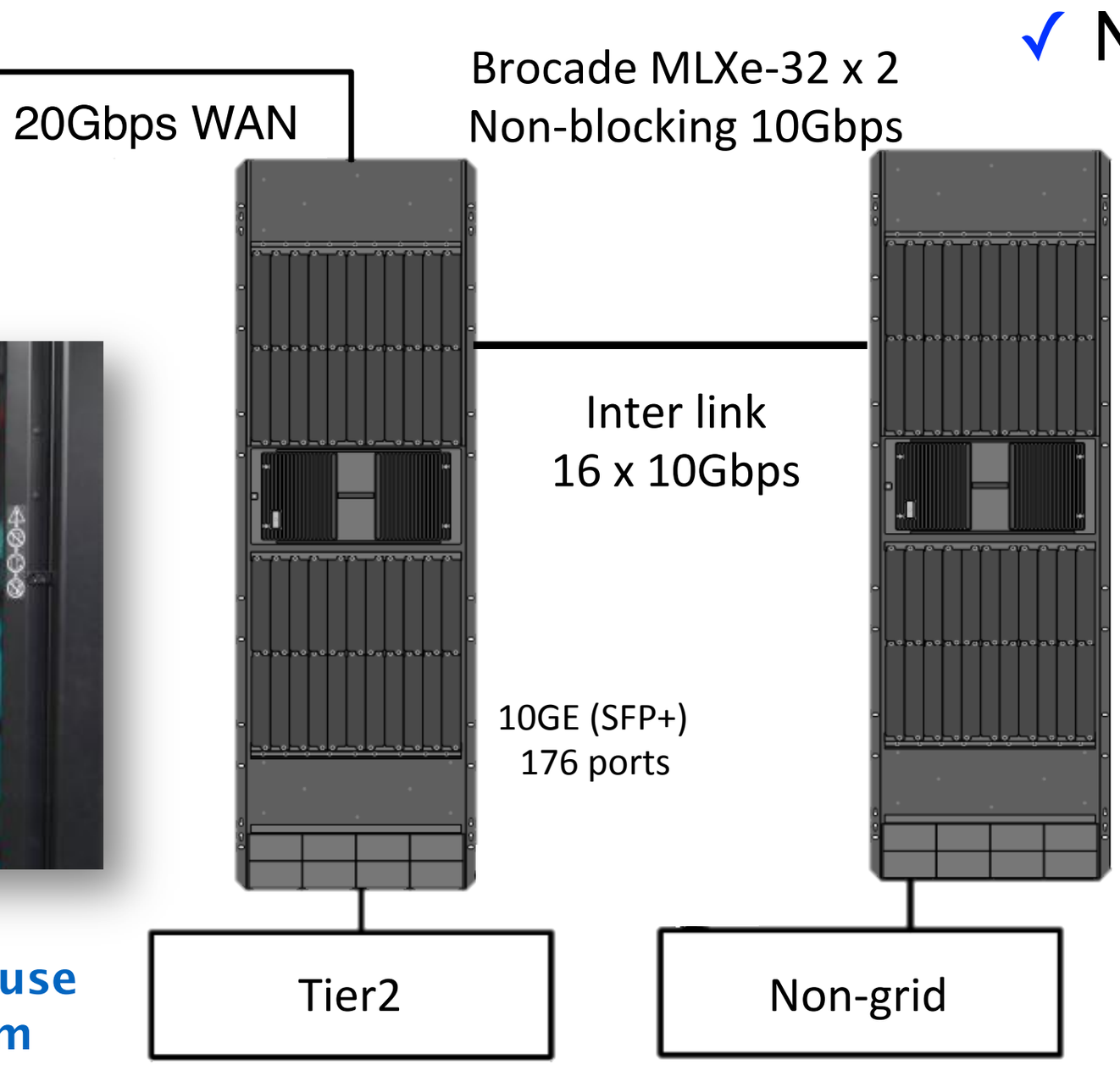
Test jobs (e.g. ops job) are overcommitted in HTCondor system

- Improvement of CPU utilization has been observed thanks to the dynamic partitioning.
- HTCondor is stable so far.

Tier2 configuration

✓ Network

Main switches: continued use from previous (3rd) system



DPM file servers
LCG service nodes
LCG worker nodes

GPFS/NFS file servers
Tape servers
Non-grid service nodes
Non-grid computing nodes