LSST Infrastructure and Activities at CC-IN2P3

Yvan Calas

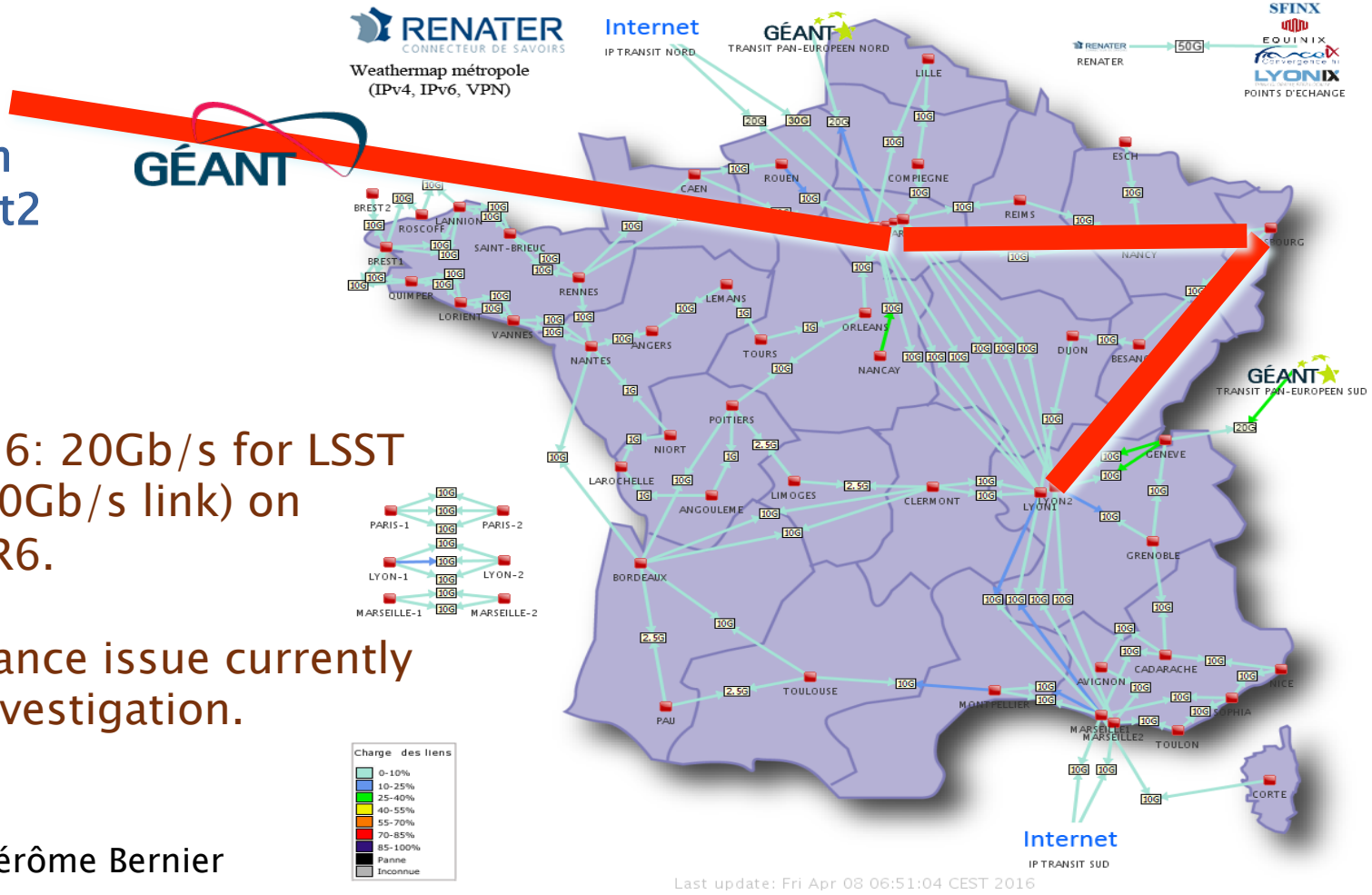
CCIN2P3

LSST France, 8 juin 2016

cnrs

Dedicated 20Gb/s link between CCIN2P3 and NCSA

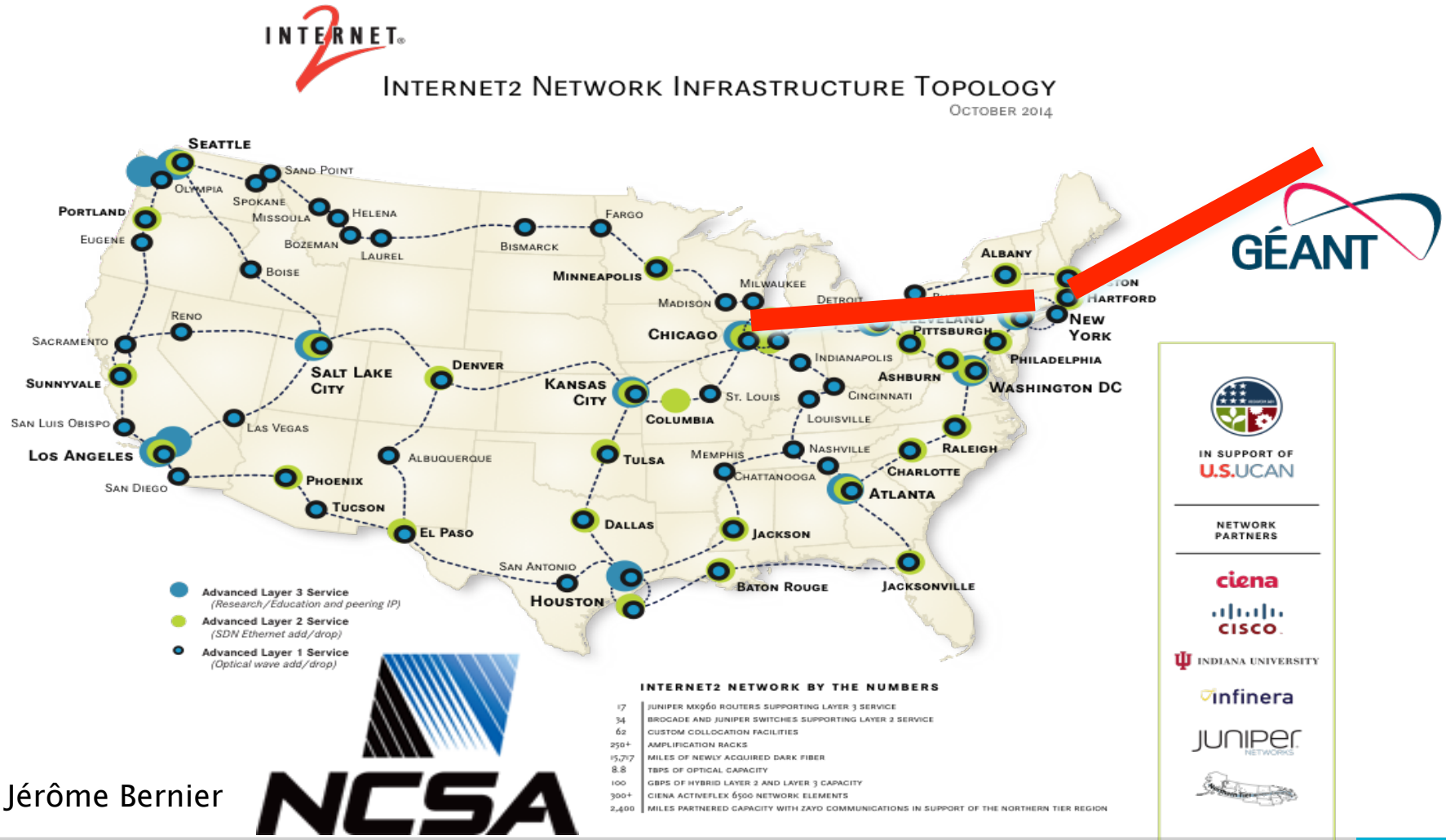
USA
ManLan
Internet2
NCSA



- May 2016: 20Gb/s for LSST (over 100Gb/s link) on RENATER6.
- Performance issue currently under investigation.

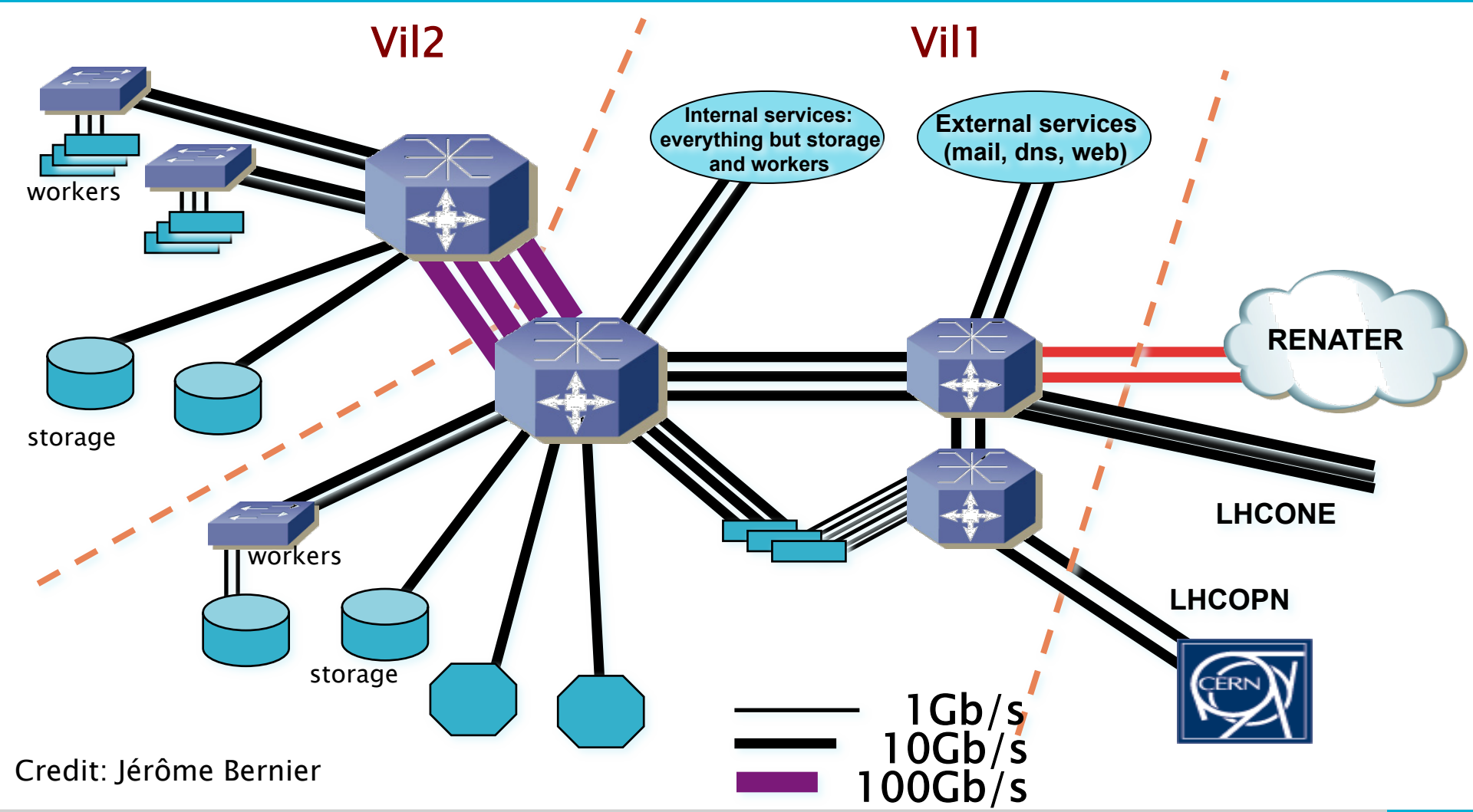
Credit: Jérôme Bernier

Internet2 - NCSA



Credit: Jérôme Bernier

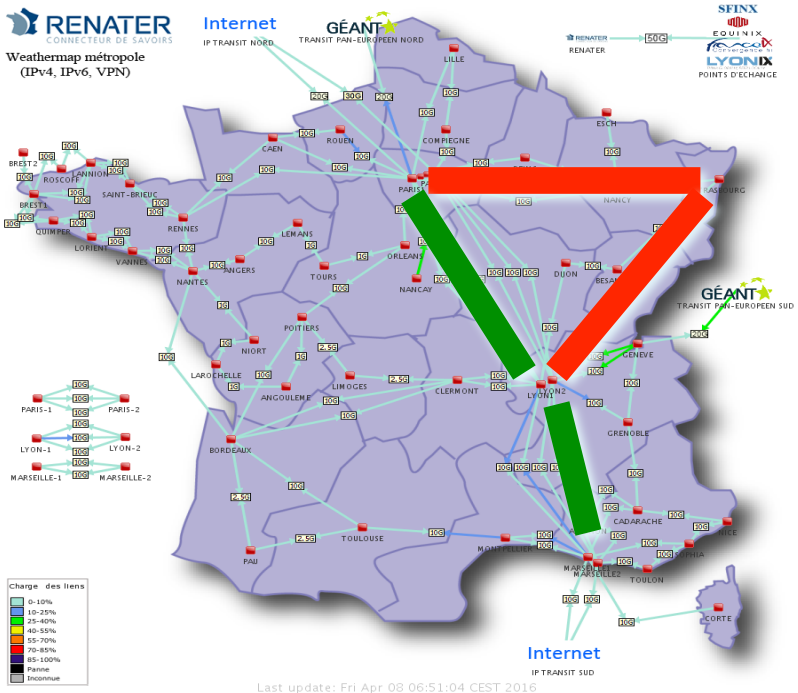
CC-IN2P3 Local Area Network (LAN)



Credit: Jérôme Bernier

Evolution of RENATER6

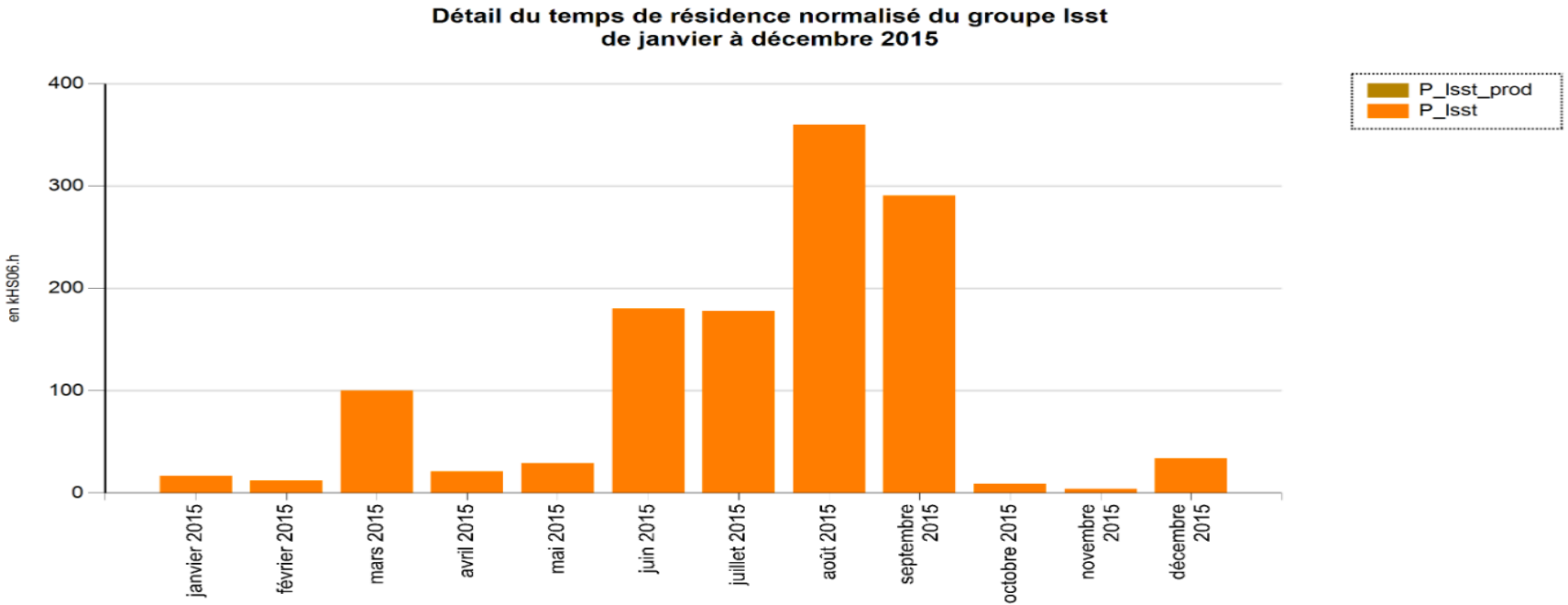
- RENATER6 100Gb/s deployment between Paris, Lyon and Marseille.
- CCIN2P3 100Gb/s connectivity.



Credit: Jérôme Bernier

LSST resources: CPU consumption in 2015

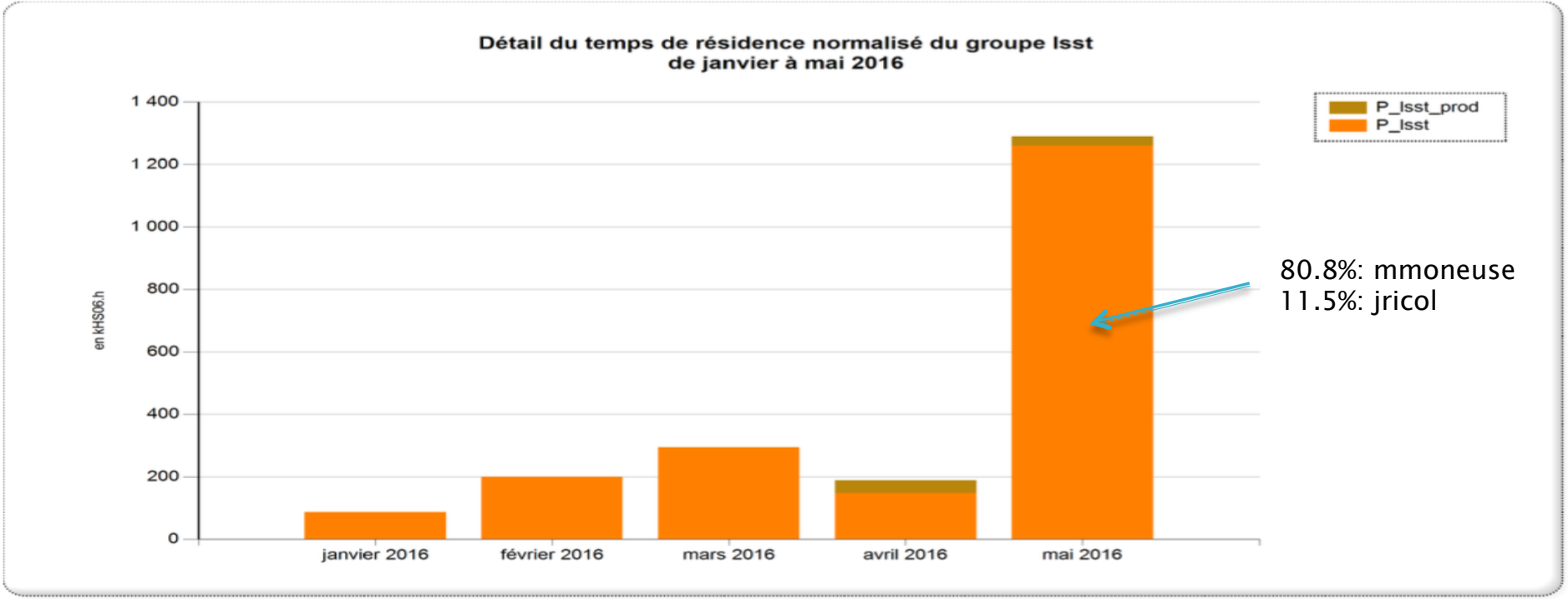
- 2015 CPU allocation: 10 MHS06.hours.
- “Real” consumption: 1,2 MHS06.hours (12%).
- Average efficiency: 91% (good).



Credit: Rachid Lemrani

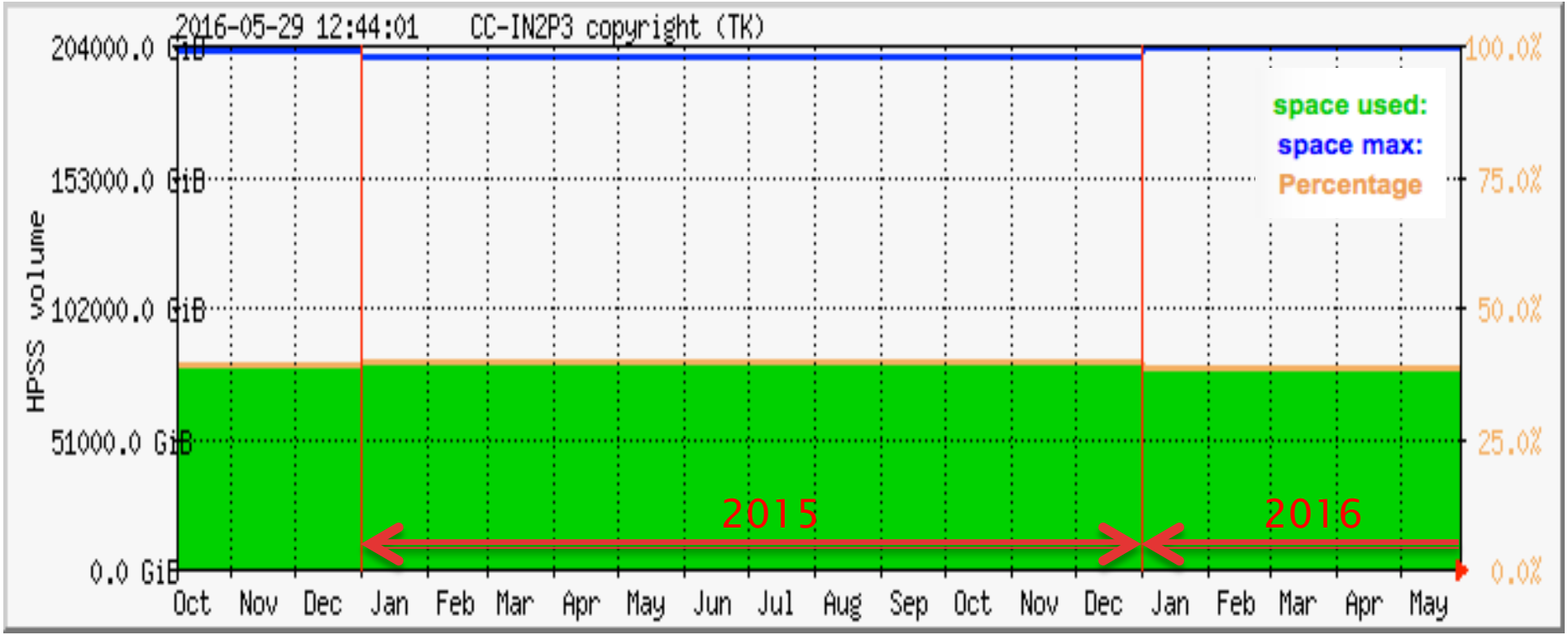
LSST resources: CPU consumption in 2016

- 2016 CPU allocation: 10 MHS06.hours.
- “Real” consumption (until May 30th): 1 807 304 kHS06.hours (18.07%).
- Average efficiency: 91% (good).



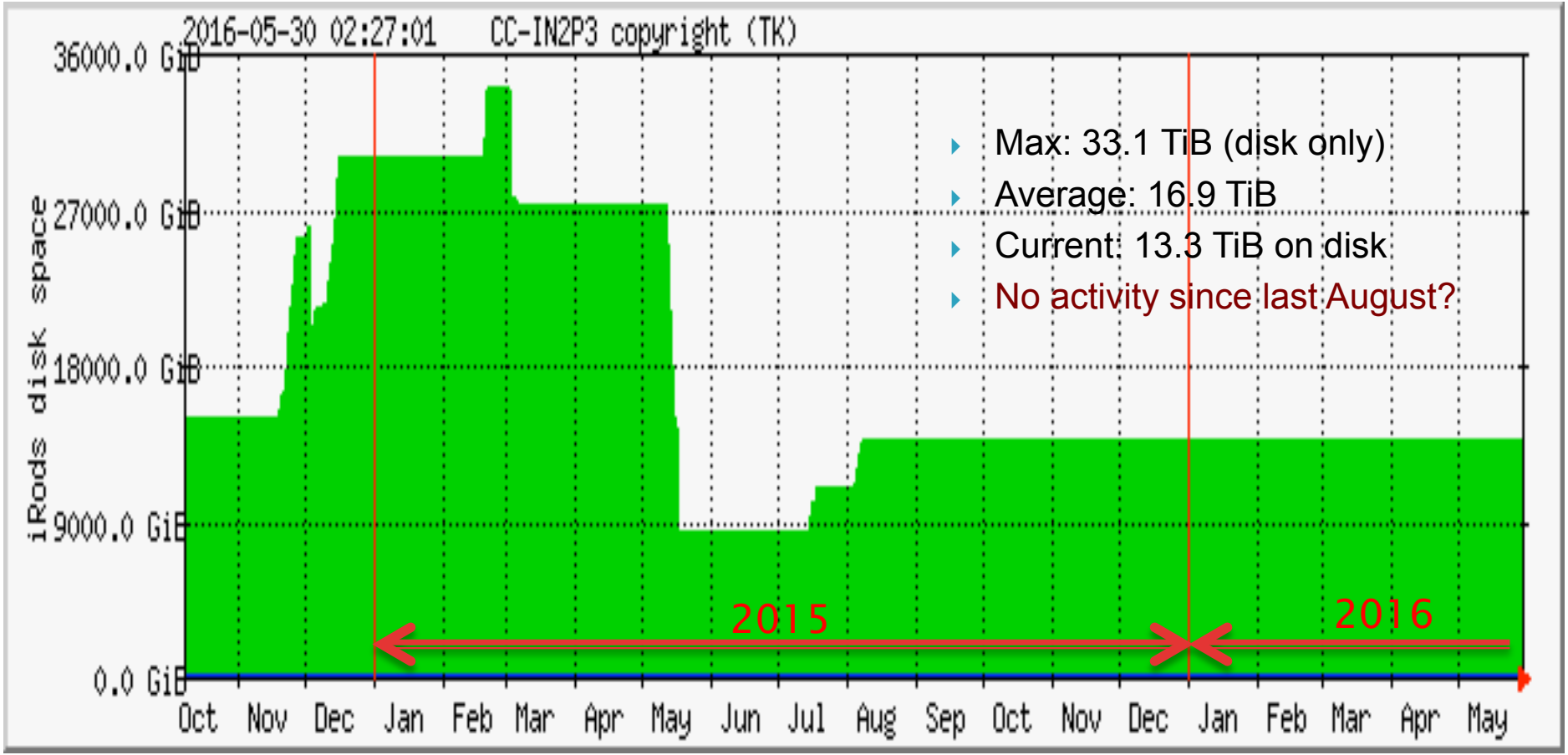
LSST resources: HPSS

- Total quota: 200 TB.
- Used: 76 TB (38%).



Credit: Rachid Lemrani

LSST resources: iRODS (disk buffer only)

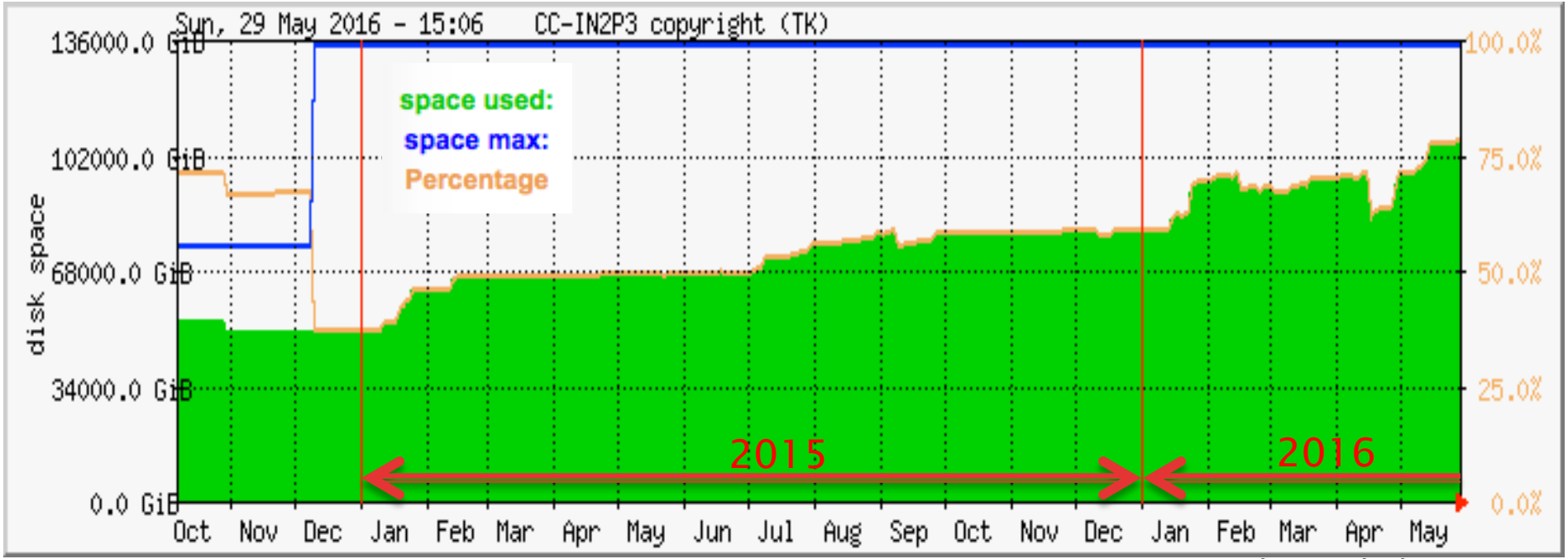


- Data stored on tape: 78GiB

Credit: Rachid Lemrani

LSST resources: SPS

- Total quota: 130 TB.
- Used: 100 TB (77%) for 54 users.
- 45 % of the used space has not been accessed for more than 2 years.
- 17 millions of files with more than 40% smaller than 10 kB.
- ~10 millions of files for lsstprod.



Credit: Rachid Lemrani

LSST resources at CC-IN2P3

- LSST ressource usage: http://cctools.in2p3.fr/mrtguser/info_manips_detail.php?group=lsst
- SPS monitoring: https://ccspsmon.in2p3.fr/
- User Support for astro: Rachid Lemrani
- Please do not hesitate to open a ticket concerning incidents, requests, etc.

Current status of the Qserv cluster

- 50 compute nodes (*ccqserv1xx*):
  - Used to store the celestial objects catalog.
  - Dell R620 et R730, 400 CPU CPU cores, 800 GB RAM, 500 TB disk.
  - Private subnet.
- 1 virtual machine (*ccqservbuild*) for compilation, packaging and deployment.
- 1 *docker mirror registry* (*ccqservreg*) used to store *Docker* images on site (disk cache):
  - some issues with the deployment.
- 2 virtual machines (*ccqservmon*) for the cluster monitoring:
  - *Ganglia* et *Graphana* (OS monitoring part mainly).
  - *ElasticSearch* / *Kibana* (performance evaluation from Qserv log files).
  - « *nagios-like* » HTML page used to check the availability of some services.

Ongoing works

- *Docker* experimentation:
  - *processing CFHT data with LSST stack*:
    - Purpose: package single steps of a pipeline in the form of application containers and orchestrate their execution to compose the desired workflow.
    - Creation of dedicated *Docker* images.
    - Use of several storage systems (local disk, *CernVM-FS*, *SPS*).
    - Several input parameters (eg. visit, CCD) when a container is started.
    - Serialization of container execution (eg. *Mesos*).
    - Use of *Docker* with batch systems under investigation:
      - *new beta version of Univa Grid Engine* (UGE) & HTCondor.
  - for *Qserv*:
    - Quick deployment of Qserv (SLAC) – See Fabrice's talk.
    - Deployment and use of a *Docker mirror registry* (local disk cache deployed at CC-IN2P3).

Ongoing work (cont.)

- Binary distribution of the LSST software stack through CernVM-FS (Fabio Hernandez and Vanessa Hamar).
  - provide an easy-to-use mechanism for using official stable versions of the LSST software framework, available worldwide, for individual users, computing centers and grids.
- Exploration of Apache Spark for LSST workflows (Osman Aïdel).
  - To understand if and how platforms supporting the *Map-Reduce* model can be used for LSST workflows.
- Monitoring of the usable bandwidth via *Perfsonar* and start testing intersite data transfer (Fabio Hernandez).

Partnership with DELL: topics under discussion

- Utilization of NVM (*Non Volatile Memory*) disks with 2 use cases:
  - Qserv cluster:
    - Celestial objects catalog
    - The Qserv nodes need to make use of a two-column relation of ~40 billion rows while planning and dispatching certain queries.
    - Lookups within this table need to execute very quickly (order of milliseconds per lookup).
  - Direct image processing:
    - Large amount of I/O induced by the LSST image processing applications
    - Try to explore data locality to limit utilization of networked file systems (e.g. SPS) and even to consider building a tiered storage system within each compute node.
  - Hardware configuration of the 2 servers:
    - 14 cores, 512GB RAM, 2 disks of 1TB each, 2 SSD disks of 400GB each, 1 NVM Express (NVMe).
- 10 GPU servers with:
  - 4 GPU per serveur.
  - Intel Xeon E5-2640 v3 (2,6GHz), 128GB RAM, 400GB SSD.
  - Infiniband + 10Gbps.
  - Strong interest of the LSST community.

Forthcoming events

- CC-IN2P3 scheduled downtime on 14/06/2016.
- 2 people from LSST Data Management team are visiting CC-IN2P3 on 05/07/2016*:*
  - *Frossie Economu*: Science Quality and Reliability Engineering.
  - *Tim Jenness*: Deputy Data Management System Architect.
  - Topics:
    - Current status and perspectives of the LSST Data Management subsystem.
    - Infrastructure evolution at CC-IN2P3 and impact for LSST.
    - …
- *LSST 2016 Project and Community Workshop*, Tucson, USA, August 15-19.

Questions & Comments