



GSDC: Datacenter for Data-intensive Research

Sang-Un Ahn for KISTI-GSDC
PRAGMA36

Contents

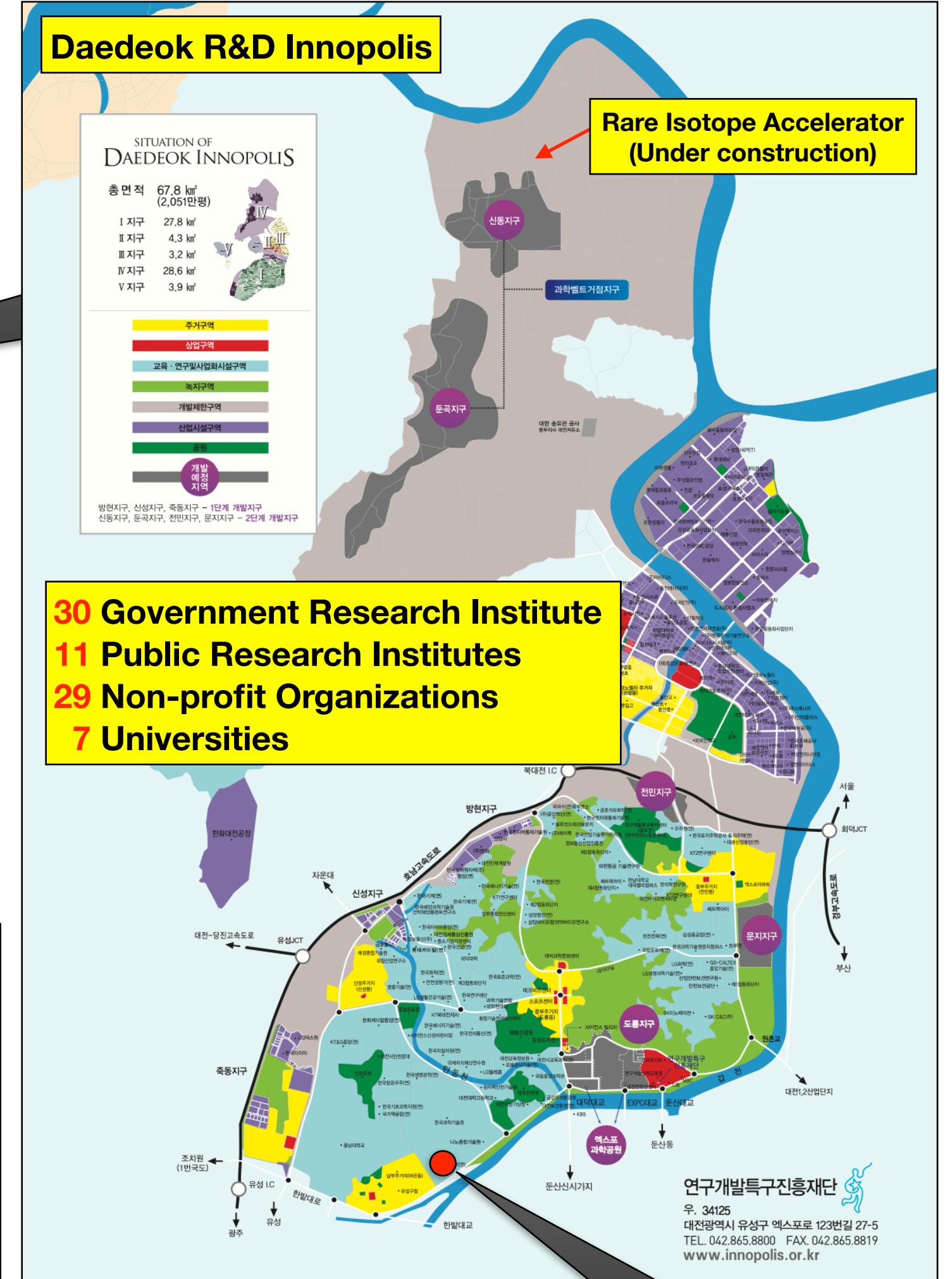
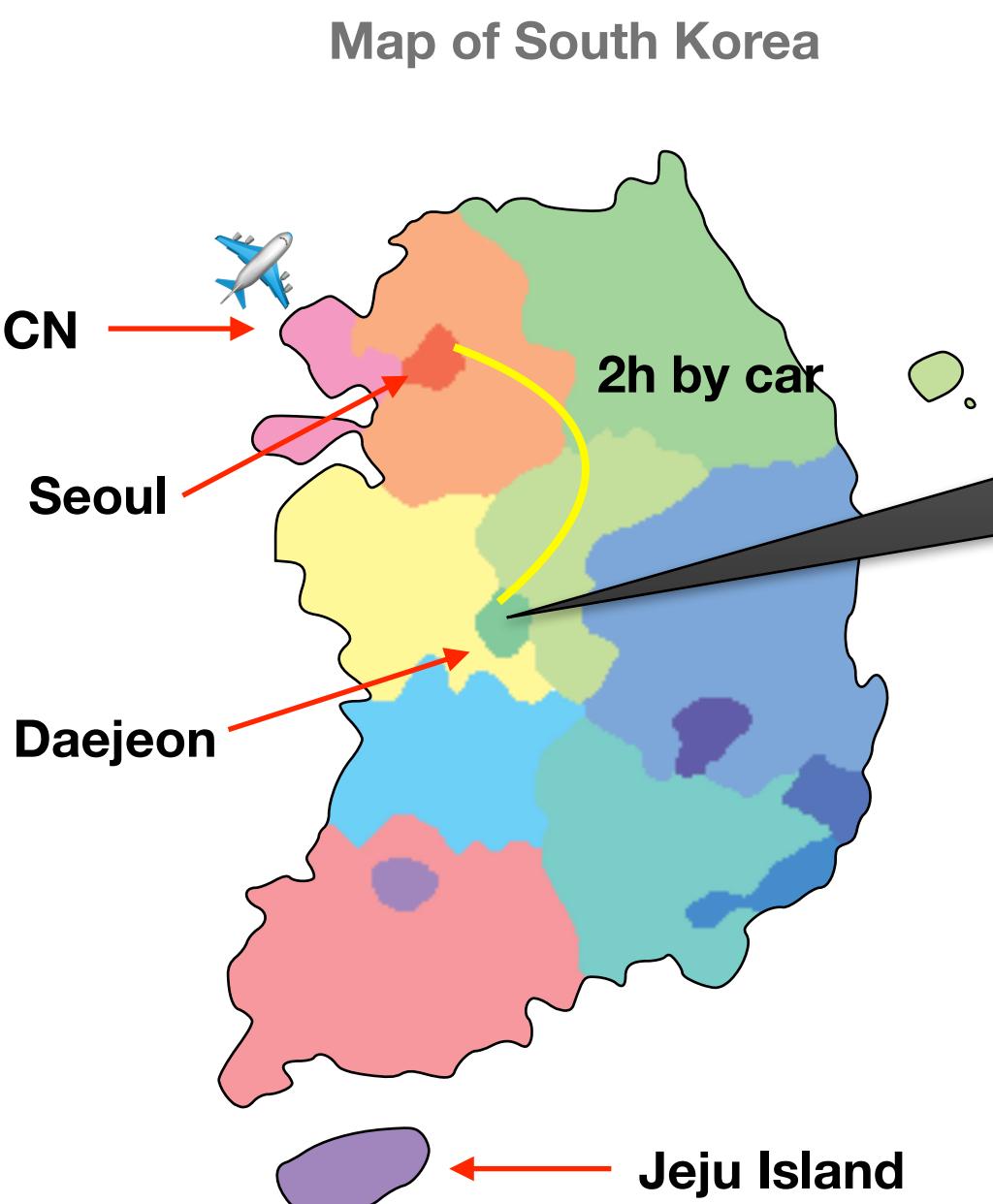
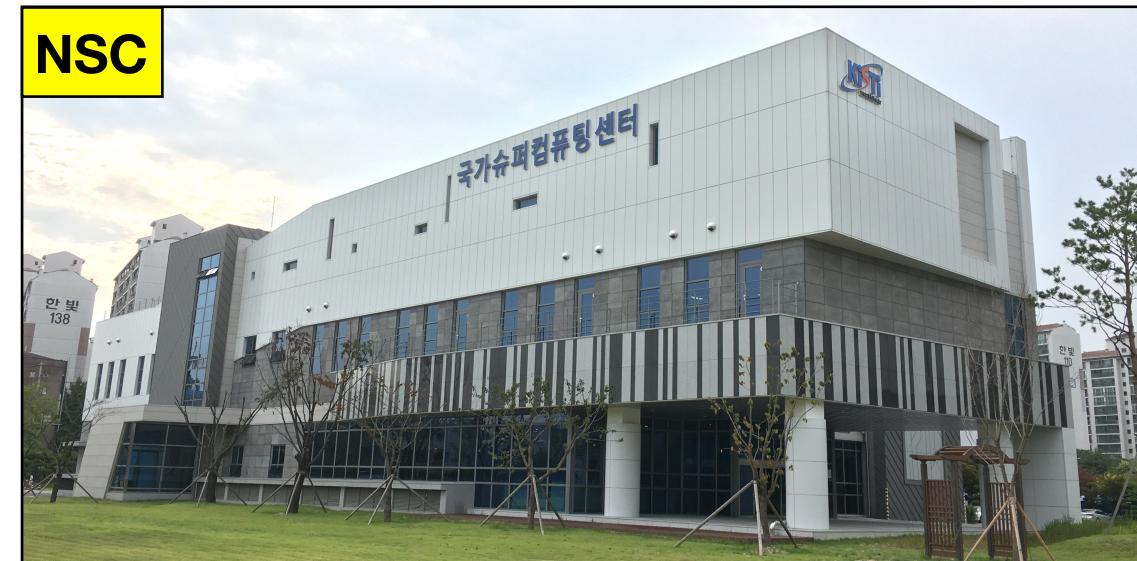
- Introduction to KISTI-GSDC
- Grid Services and Operations
- Role of Datacenter to support Domestic R&D activities
- Summary

Introduction to KISTI - GSDC

KISTI

Korea Institute of Science and Technology Information

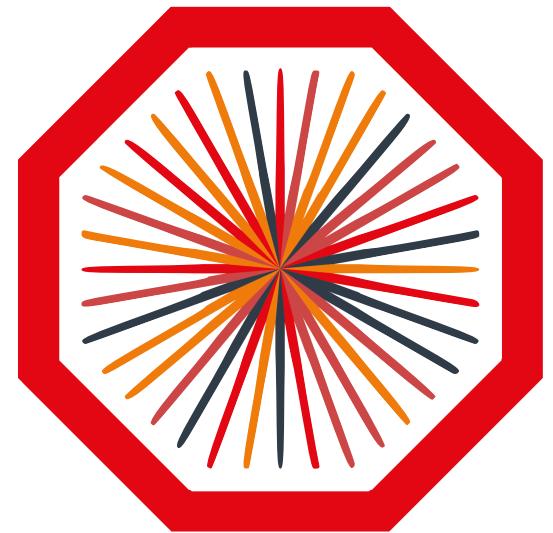
- Government-funded research institute founded in 1962 for **National Information Services and Supercomputing**
- National Supercomputing Center
 - Cray CS500 system (installed & commissioned)
 - 25.7 PFlops (Rpeak), 13.9 PFlops (Rmax) ranked 11th of Top500 (Jun 2018)
 - Tachyon II system (~307.4 TFlops, 2008)
 - KREONet/KREONet2 - National R&E network



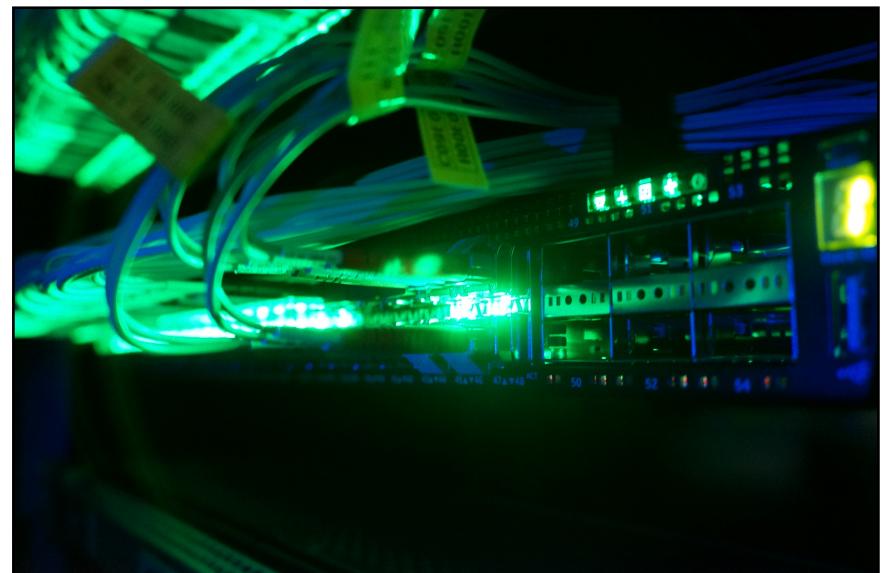
GSDC

Global Science experimental Data hub Center

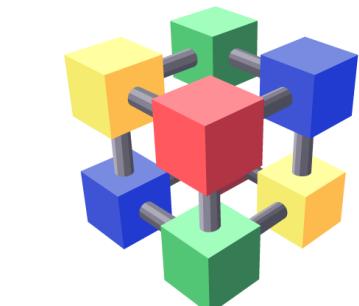
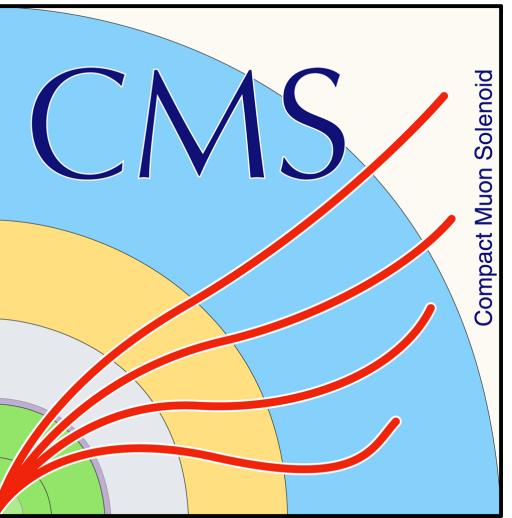
- Government-funded project, started in 2009 to promote Korean fundamental research through providing computing power and data storage
- Datacenter for data-intensive fundamental research
 - 17 staff: system administration, experiment support, external-relation, management and planning



ALICE



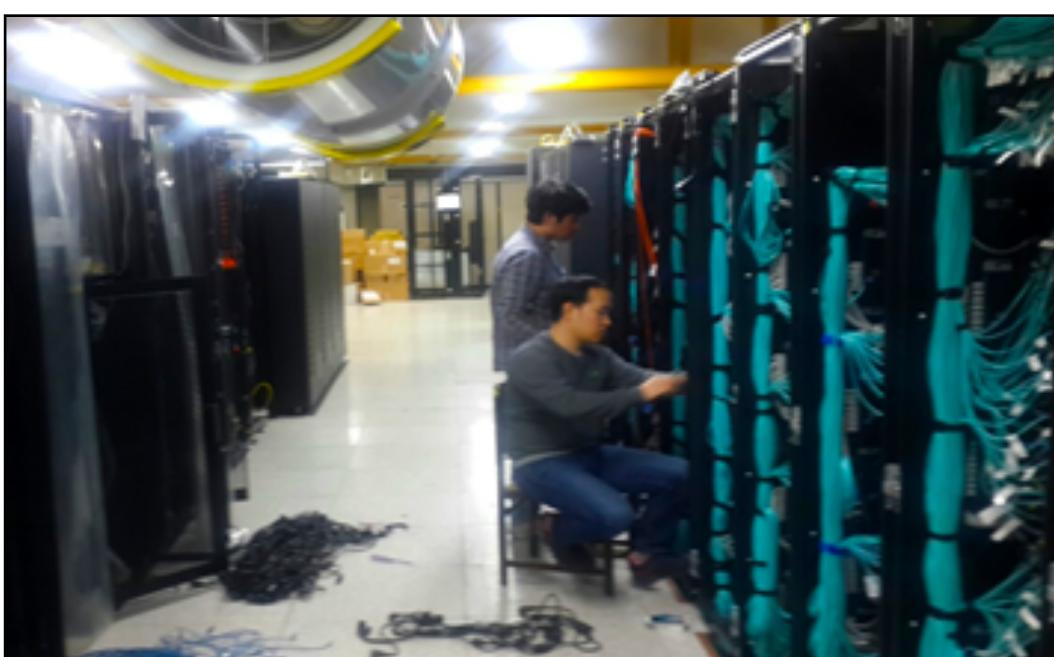
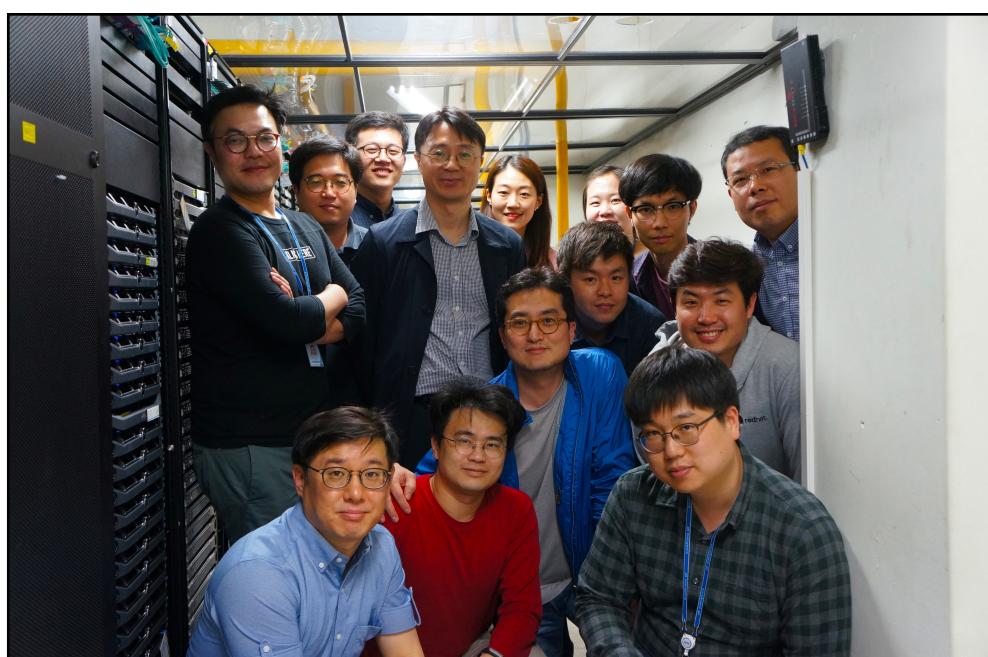
Disk Storage



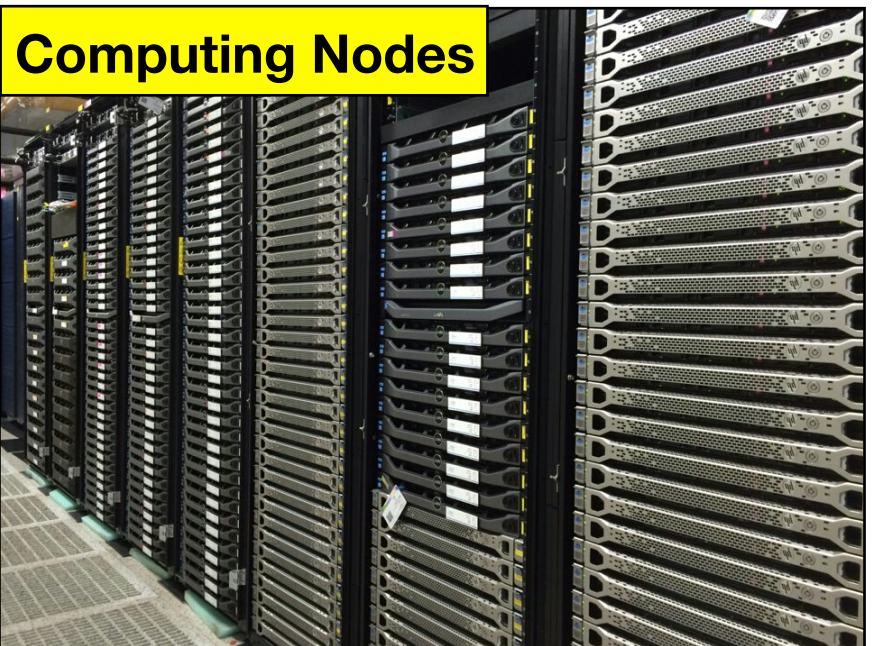
WLCG
Worldwide LHC Computing Grid



International Cancer Genome Consortium



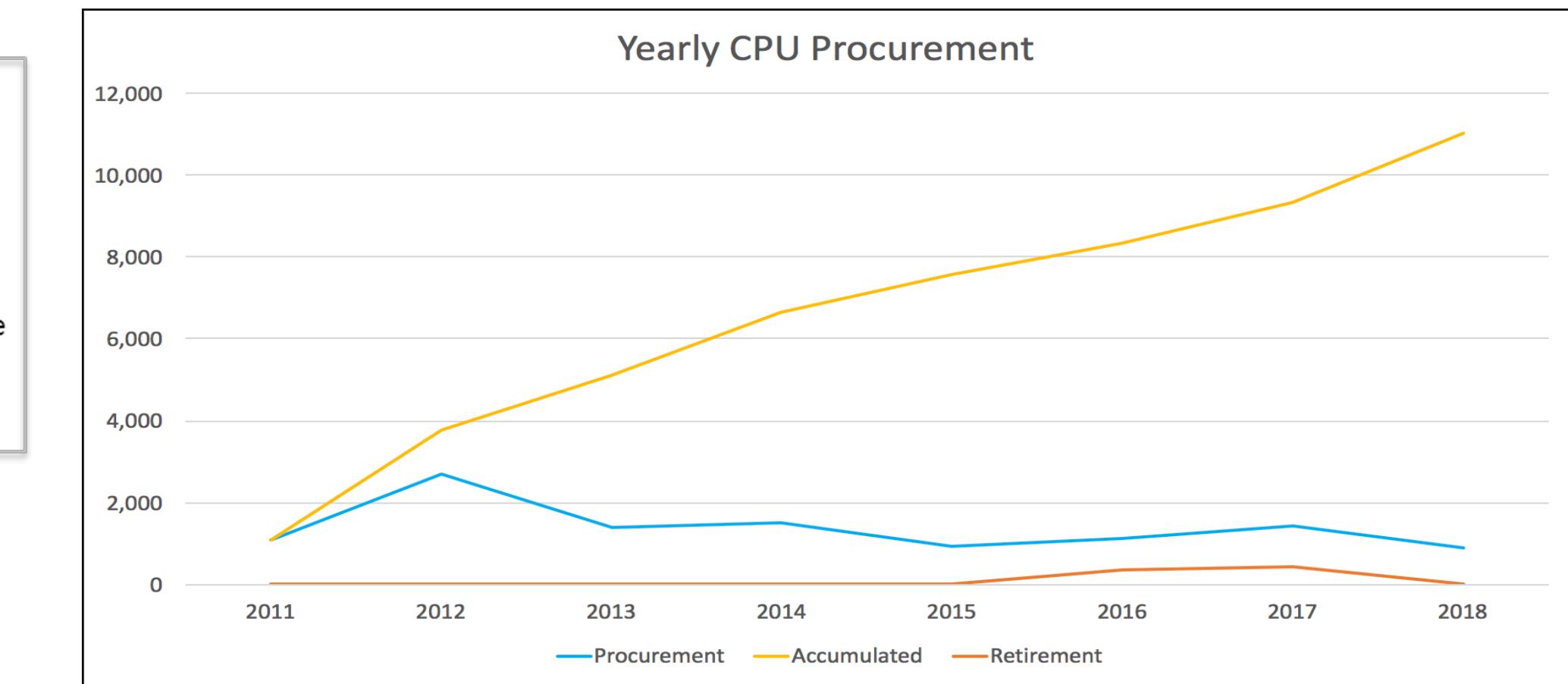
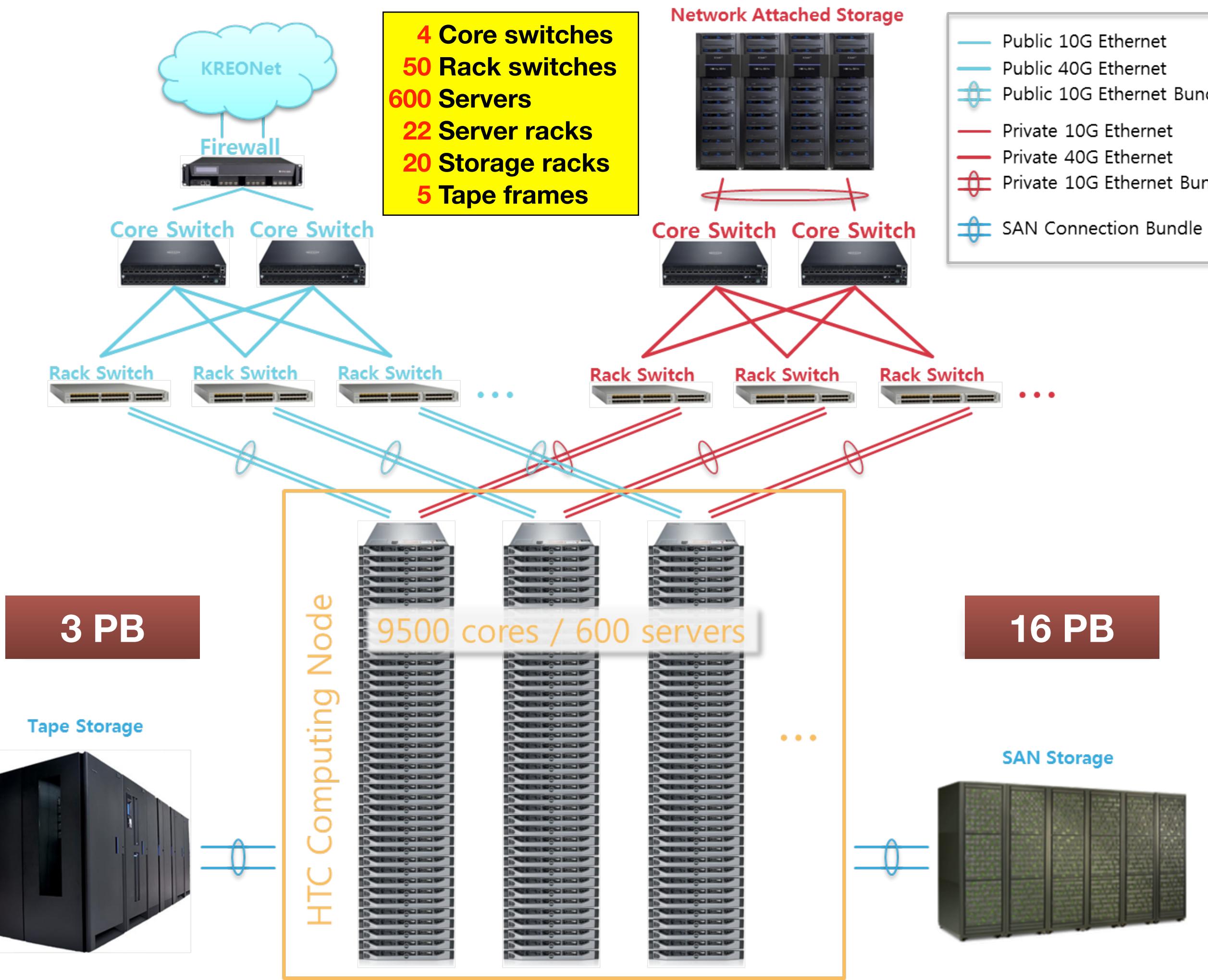
Tape System



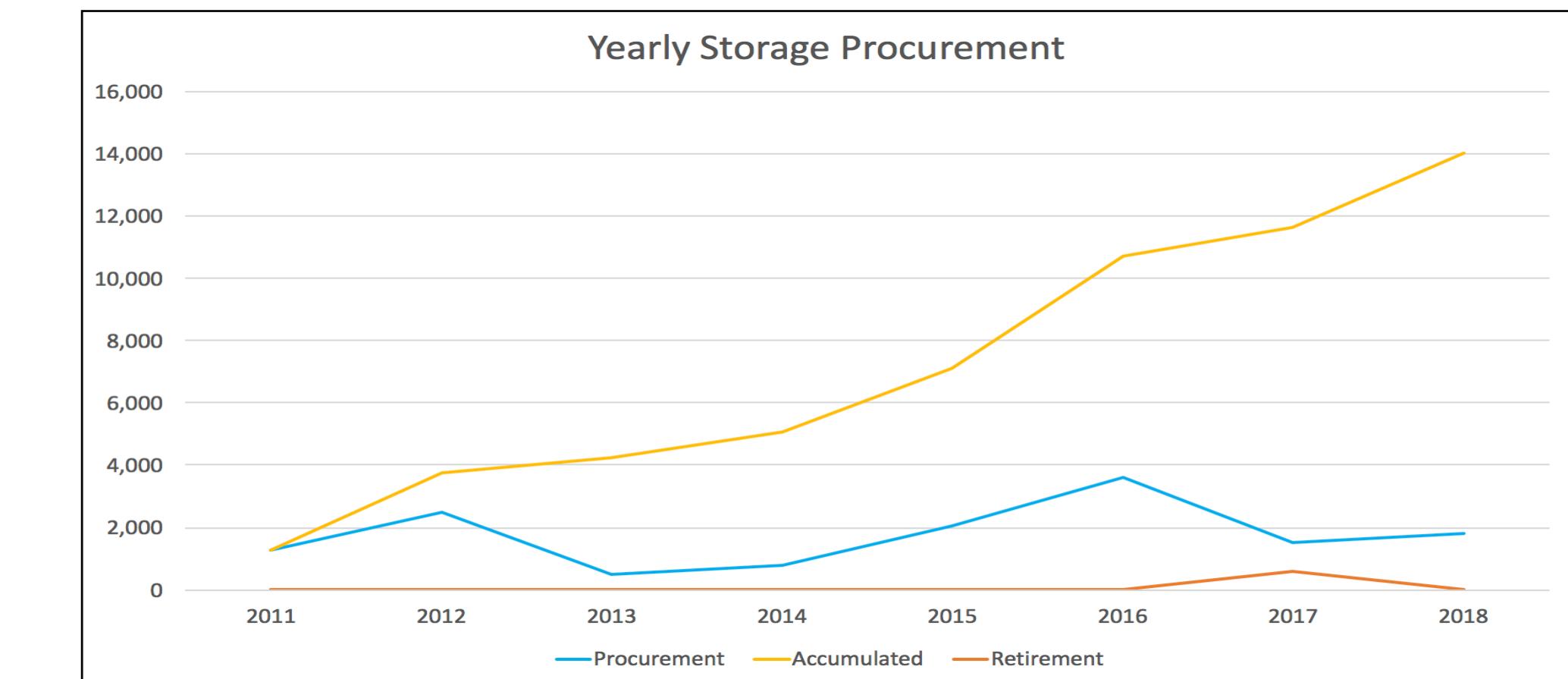
Experiments Support



Infrastructure

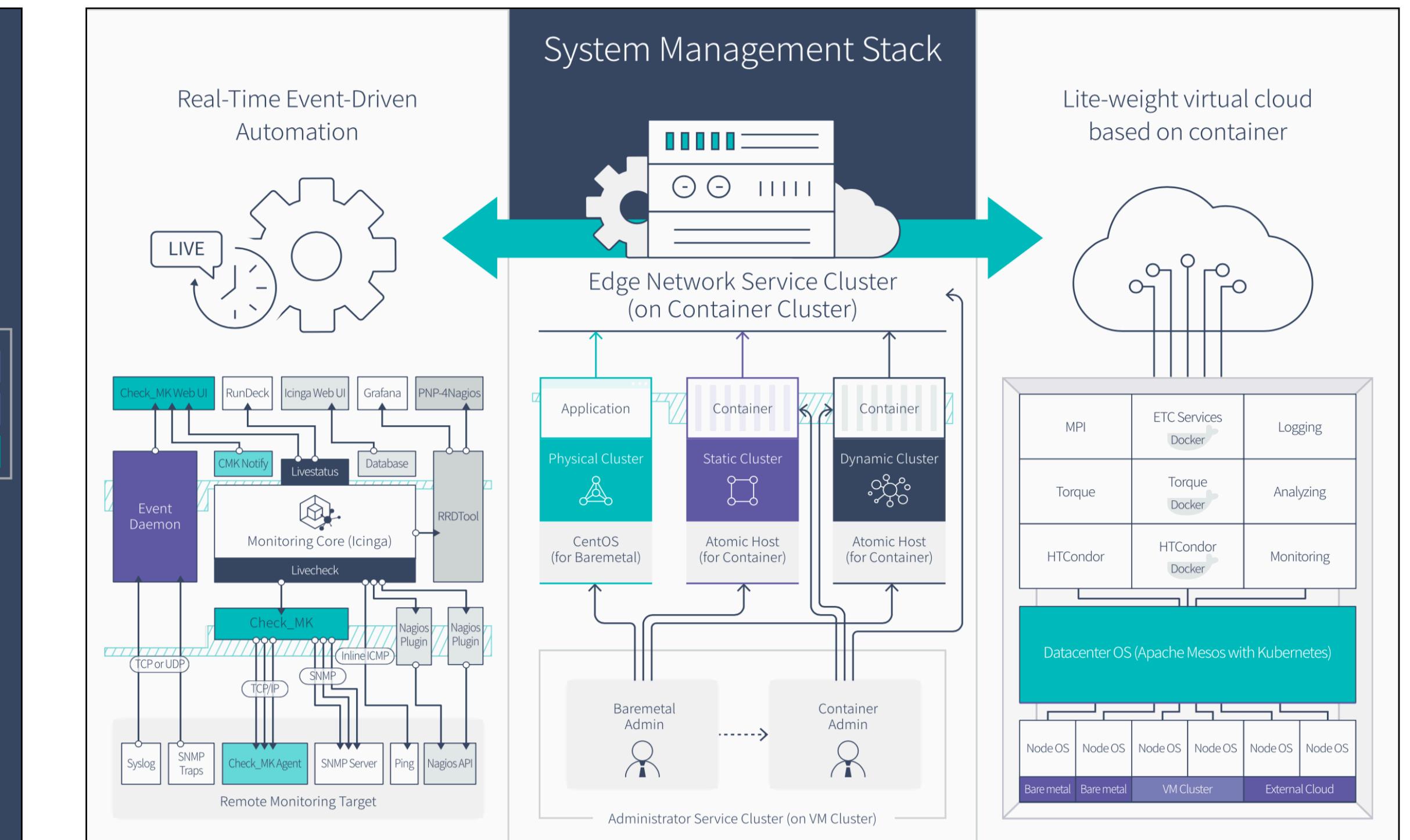
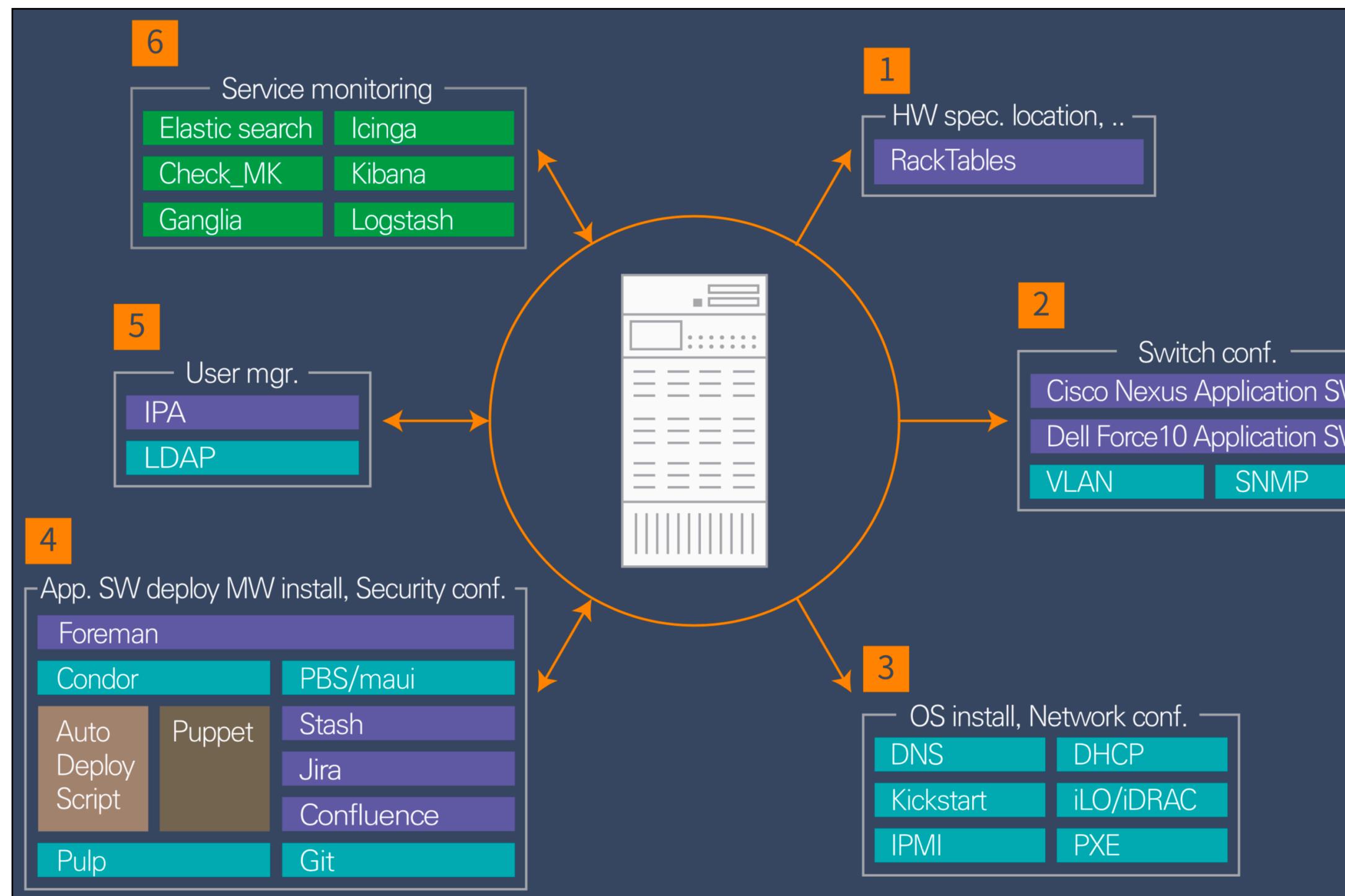


1K CPUs & 2PB Storage Procured in 2018
New Hardware Purchased Every Year

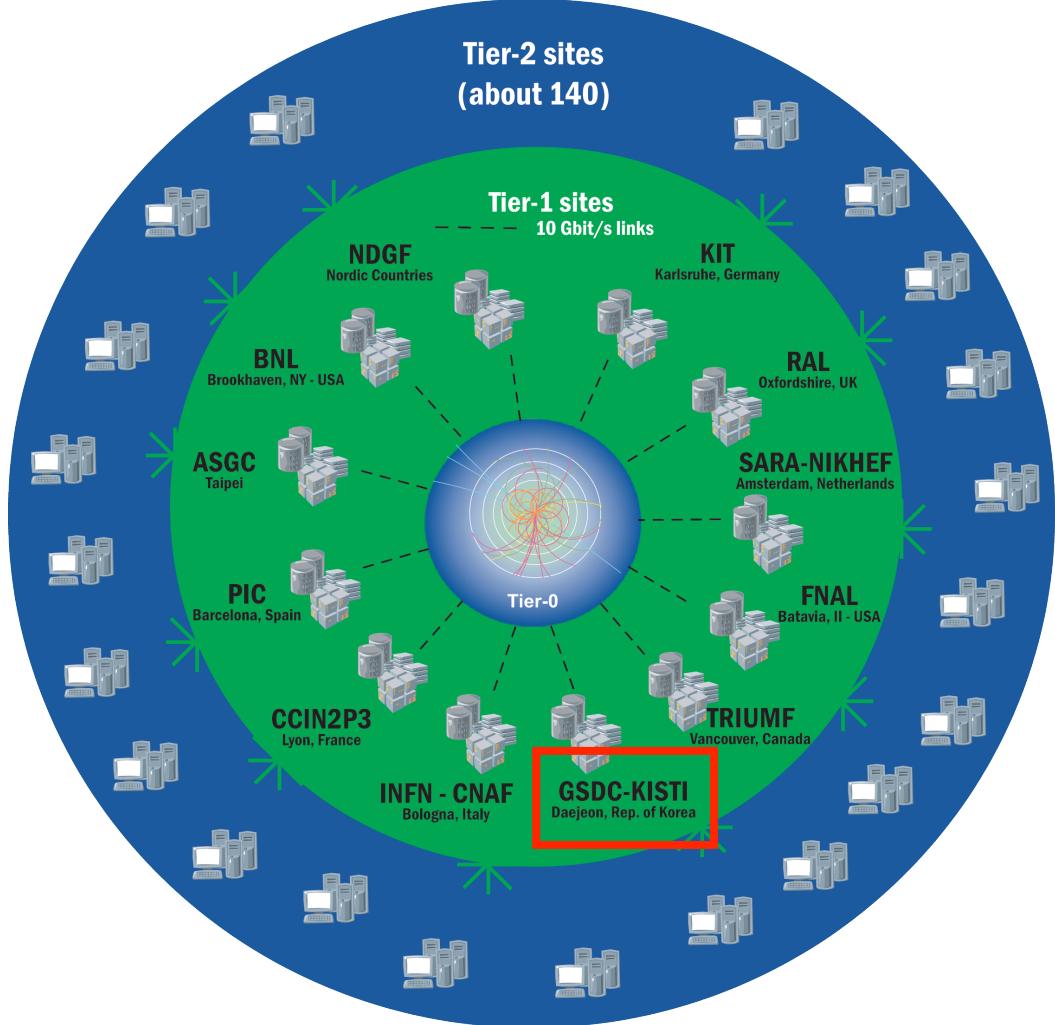


System Management

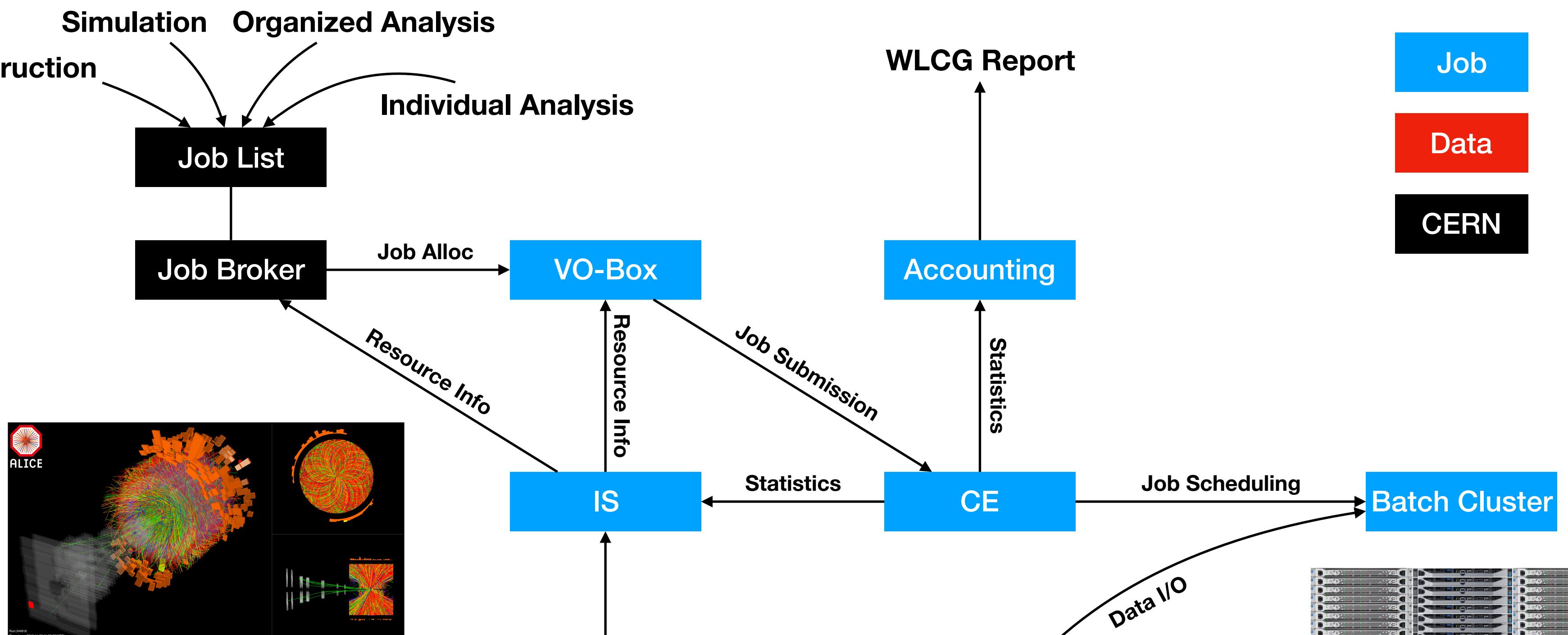
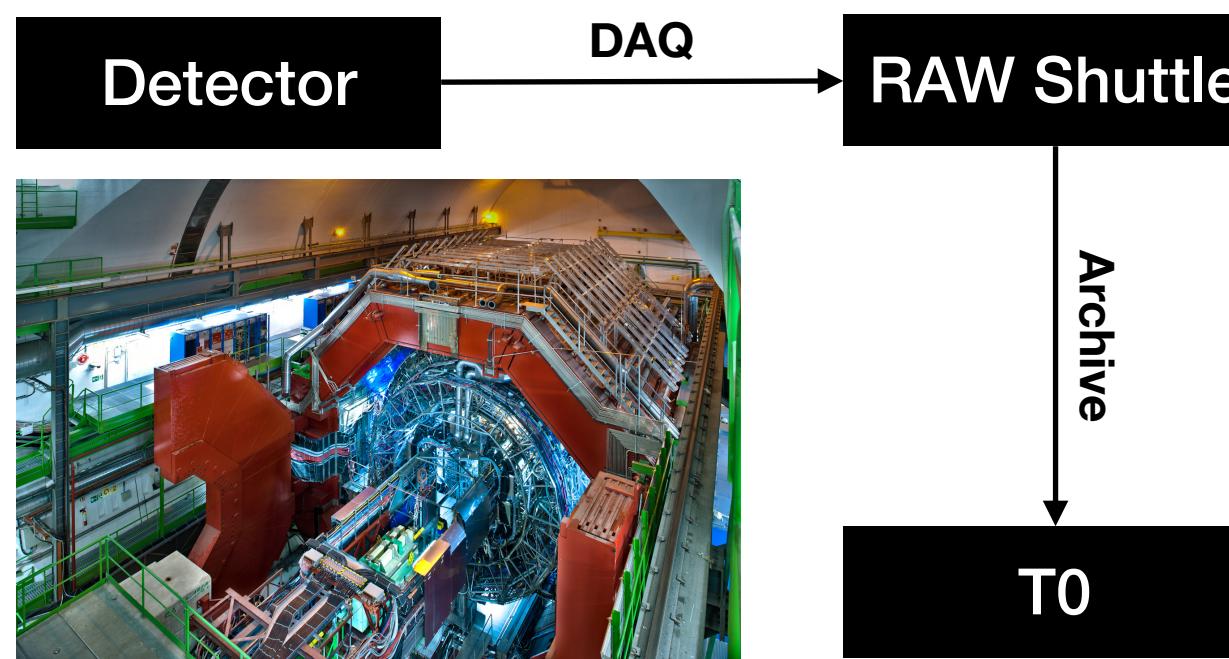
- +600 Servers and +25 Storage
- Every year of procurement... heterogenous systems with different vendors and specifications : Challenges !!
- Automated provisioning and configuration management are key for efficient and stable operations
- Moving towards virtual infrastructure based on container



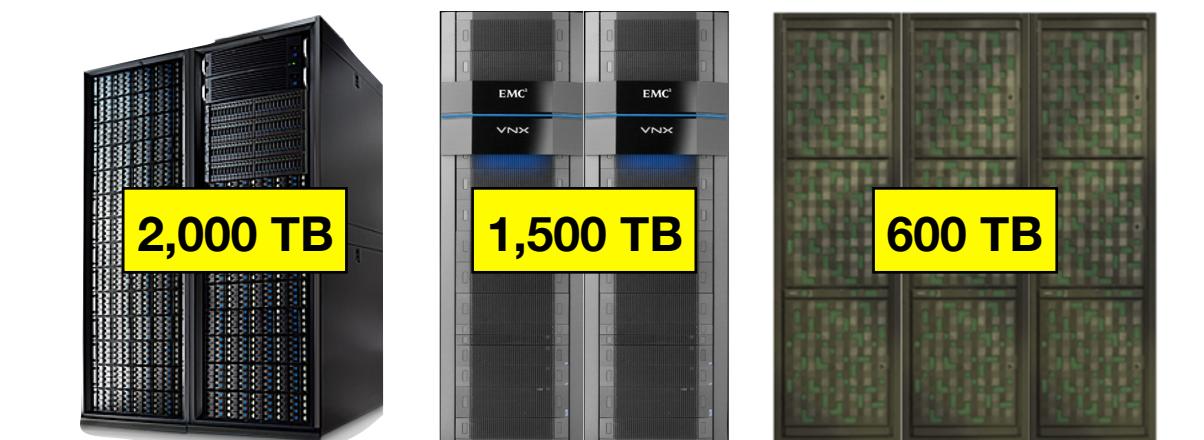
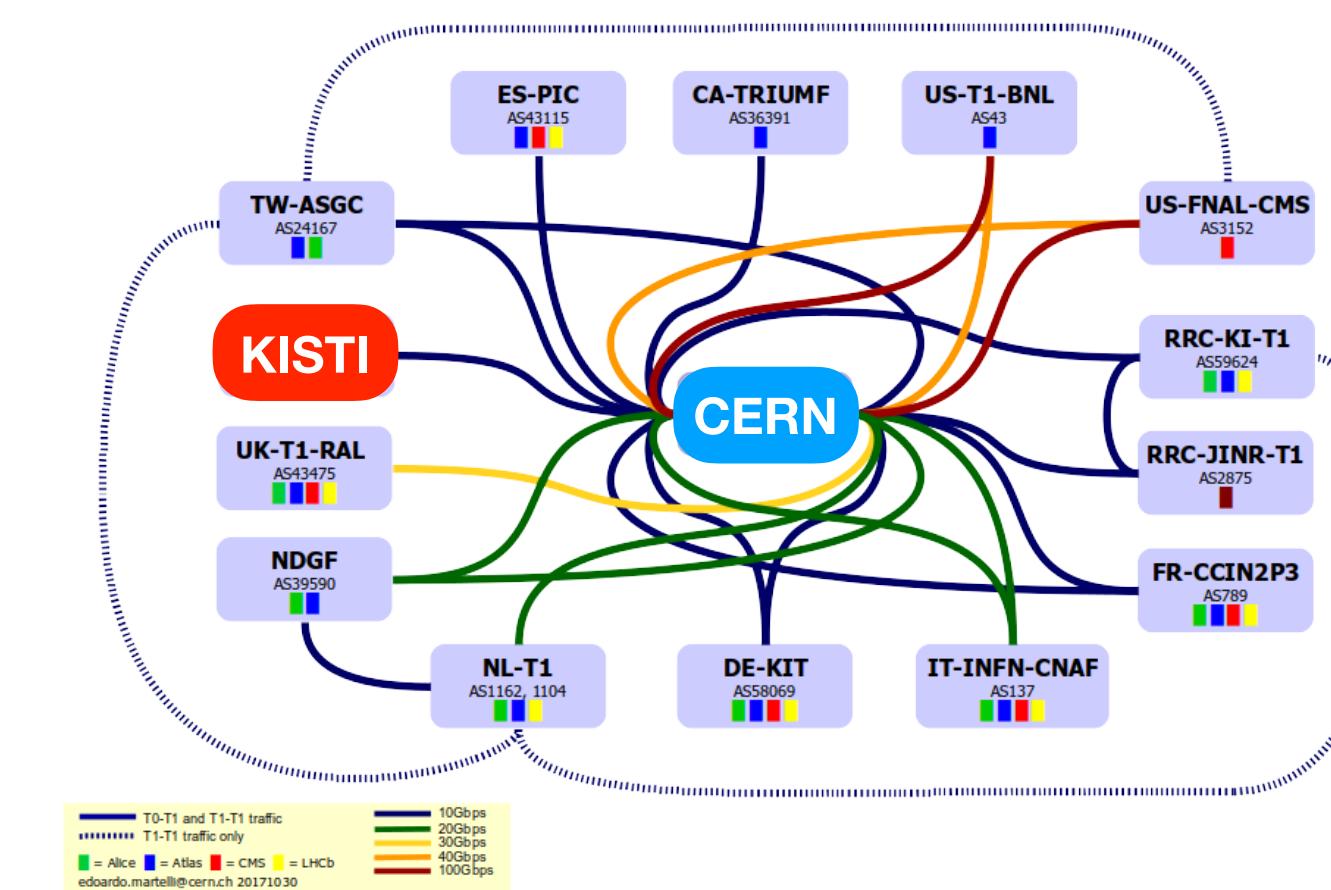
Grid Services and Operations



Worldwide LHC Computing Grid



LHCOPN



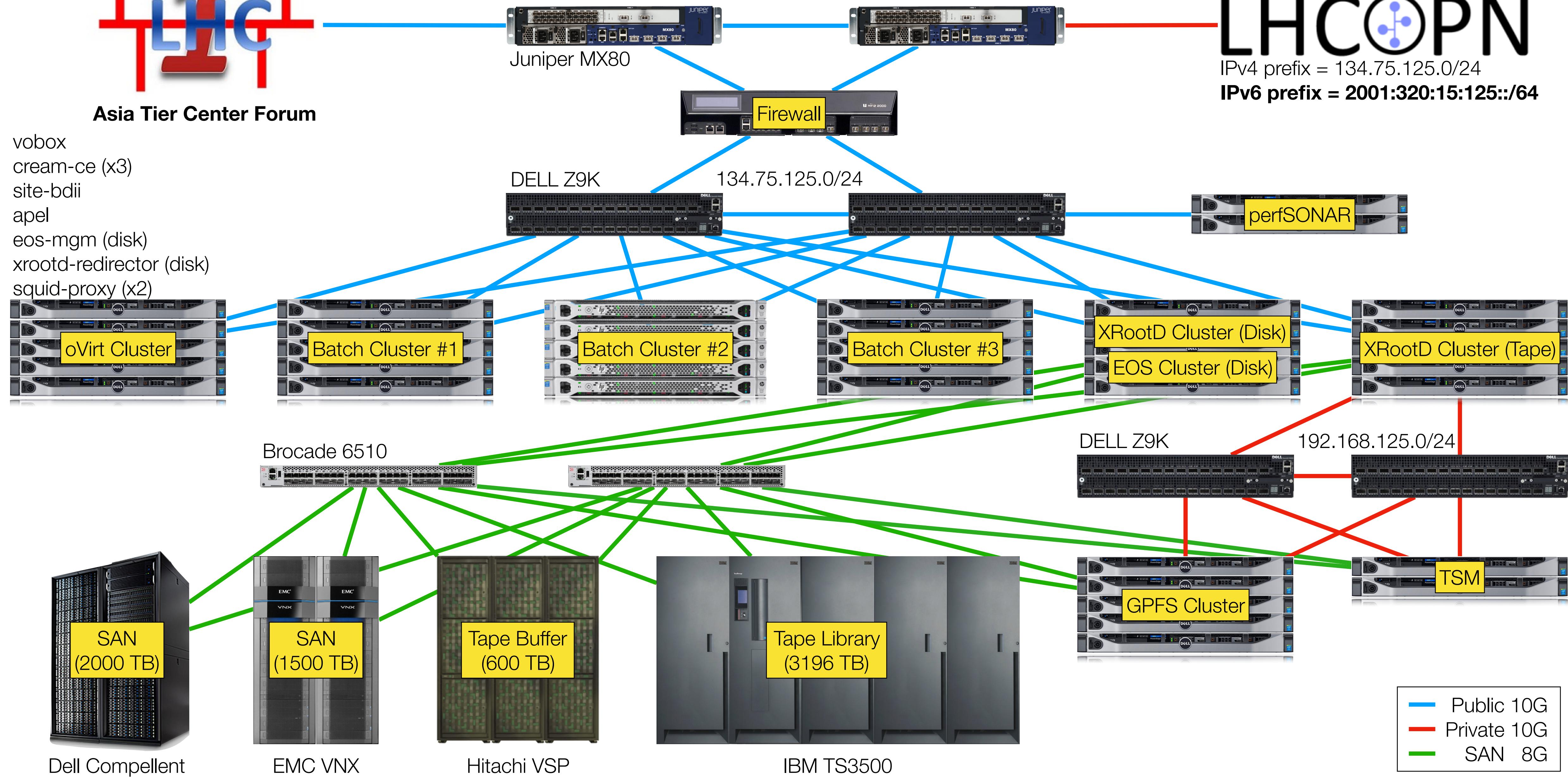
Job
Data
CERN

WLCG Tier-1 System Topology



Asia Tier Center Forum

vobox
cream-ce (x3)
site-bdii
apel
eos-mgm (disk)
xrootd-redirector (disk)
squid-proxy (x2)

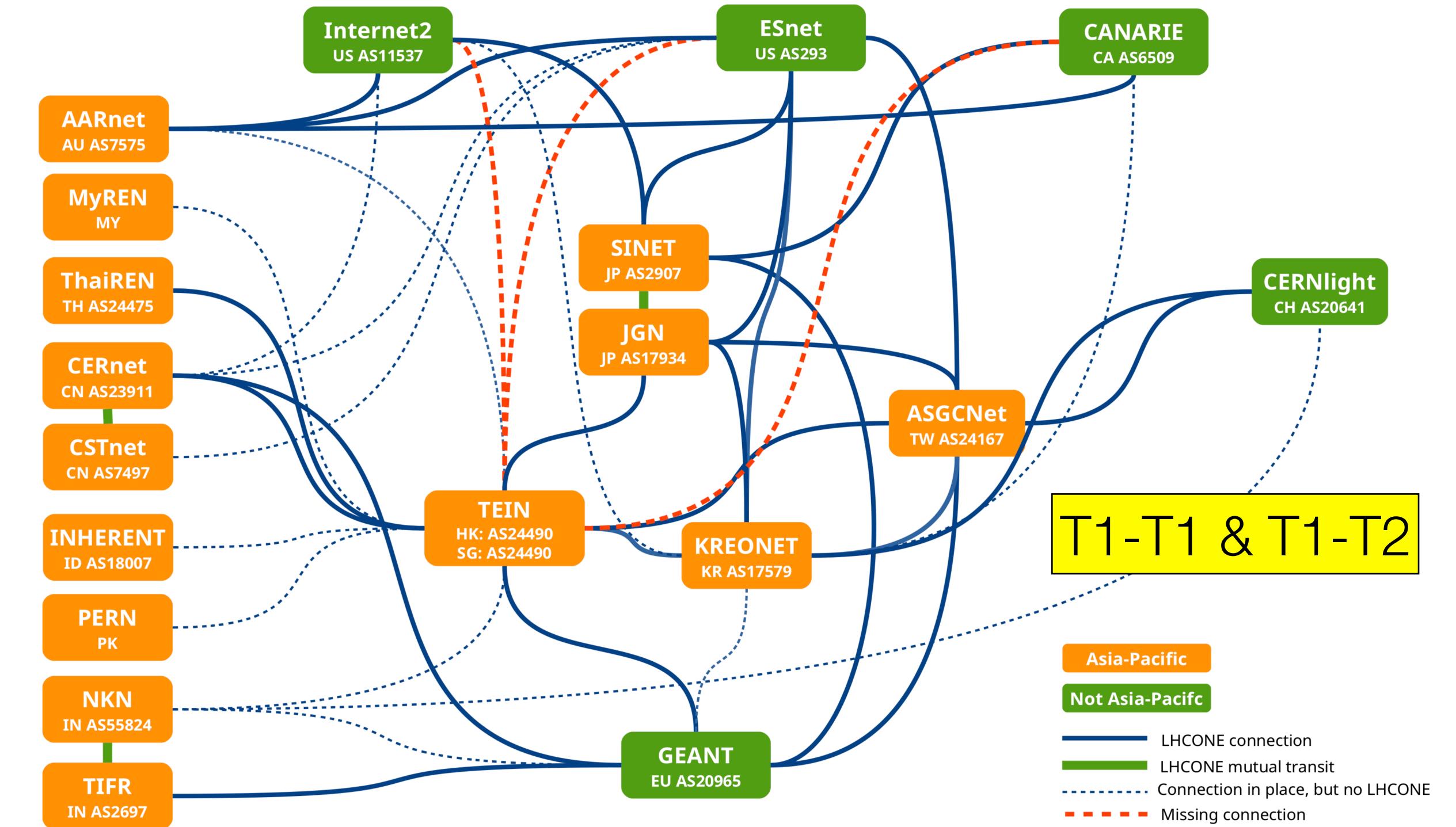
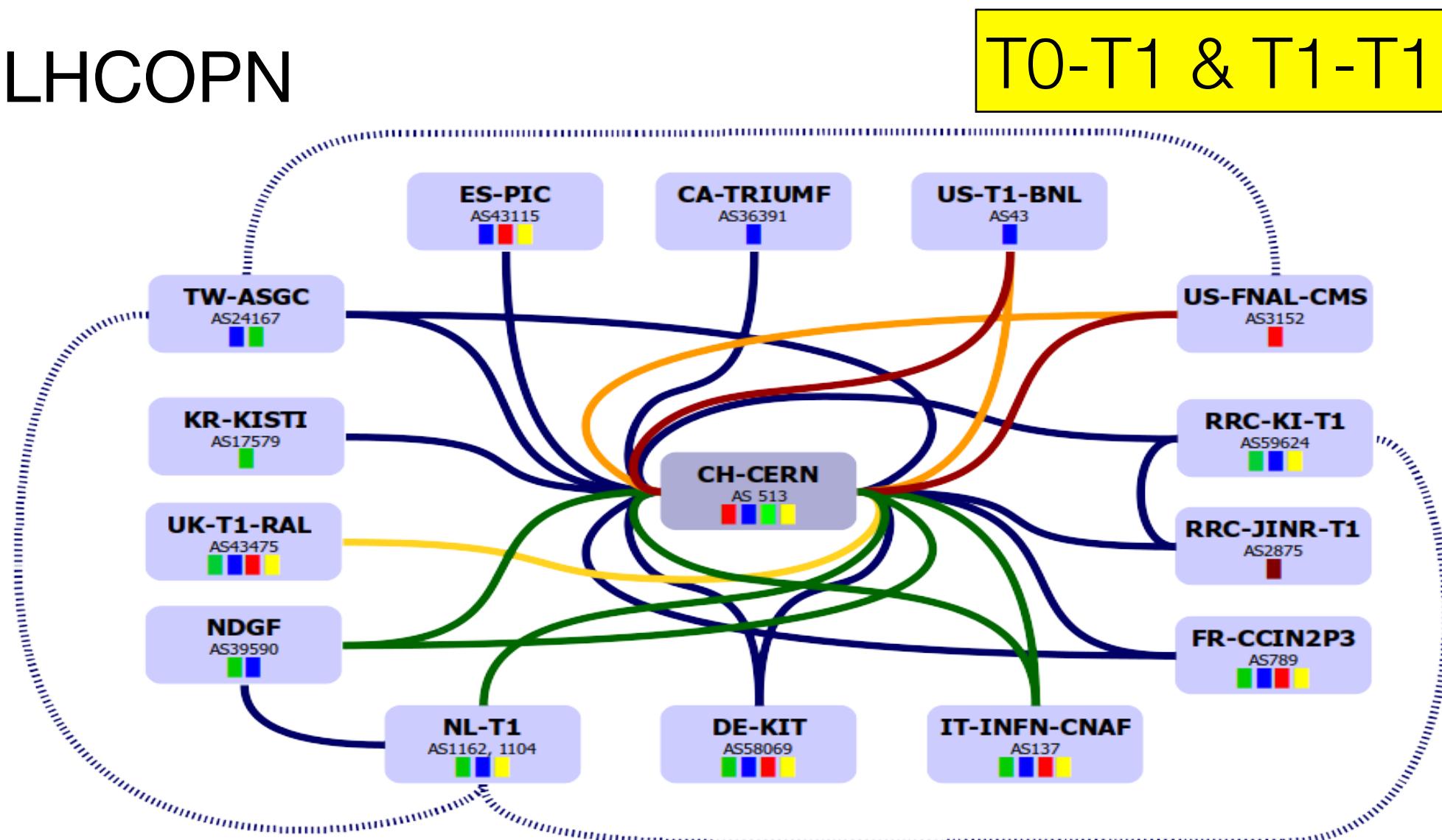


LHCONE in Asia

www.atcforum.org

Networking

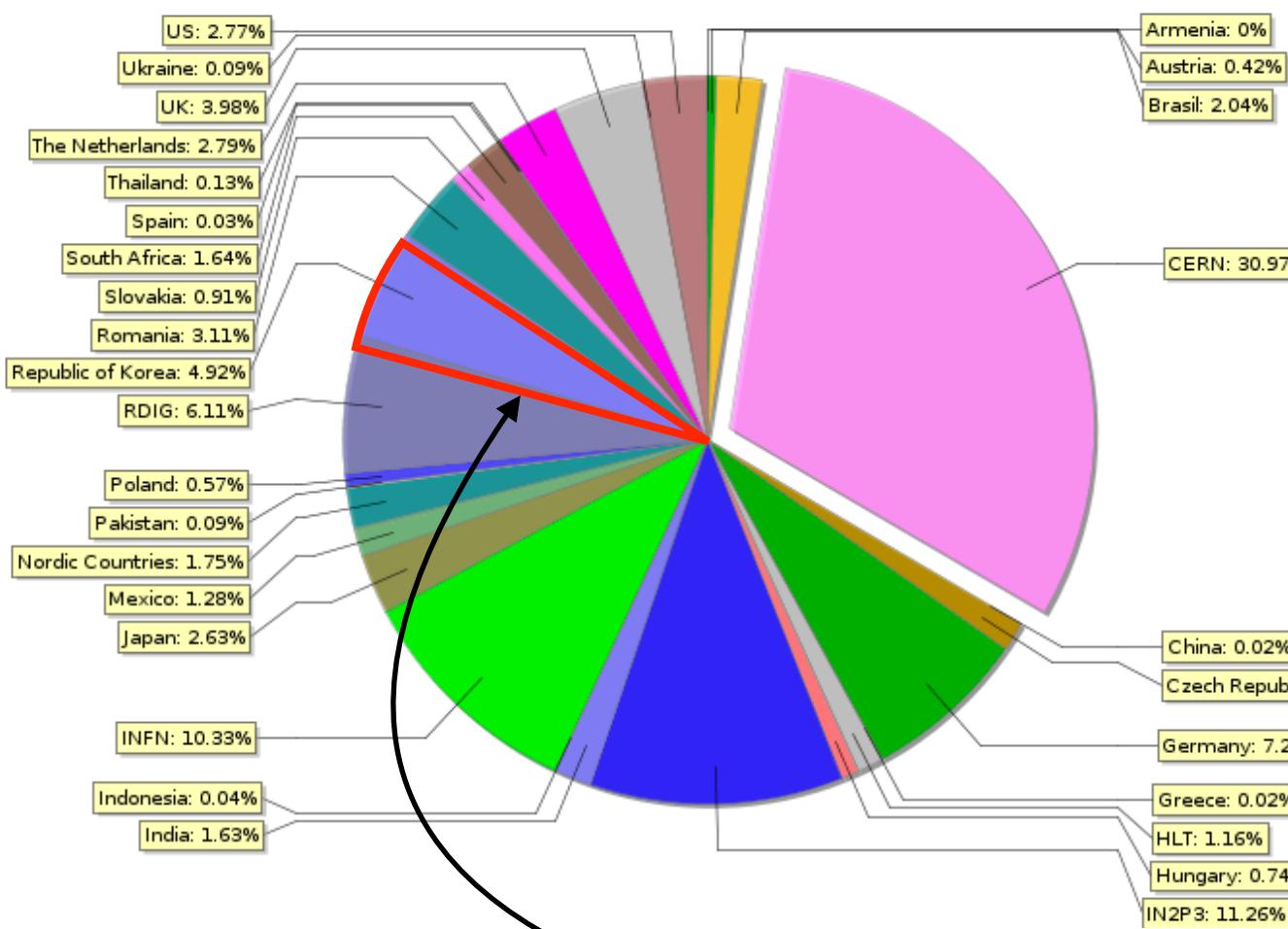
LHCOPN



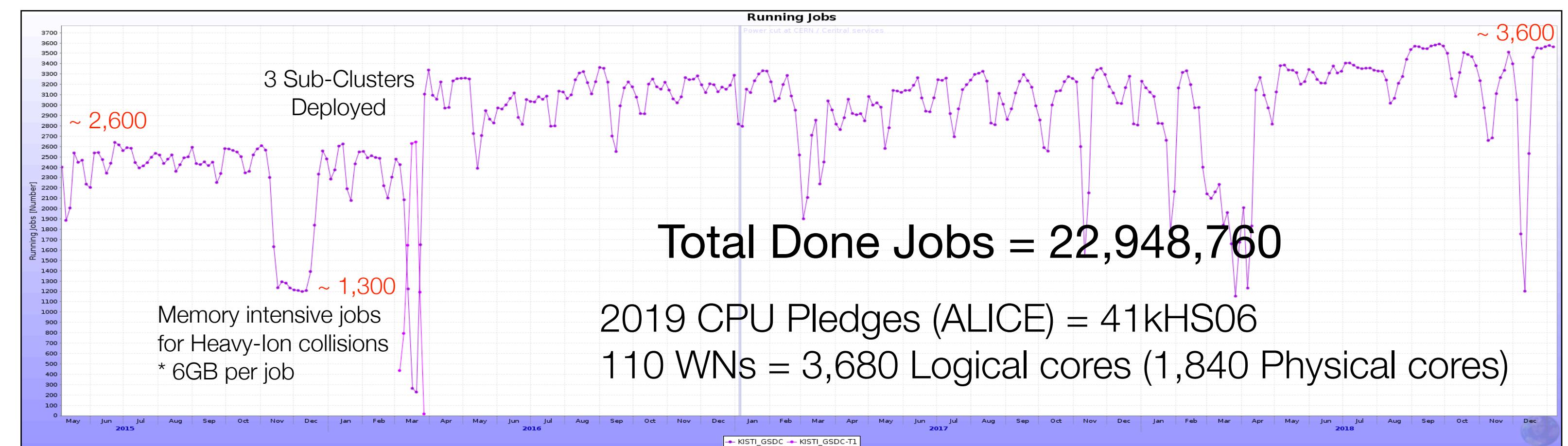
WLCG Tier-1 for ALICE (LHC RUN2 Operations)

May 2015 - Dec 2018

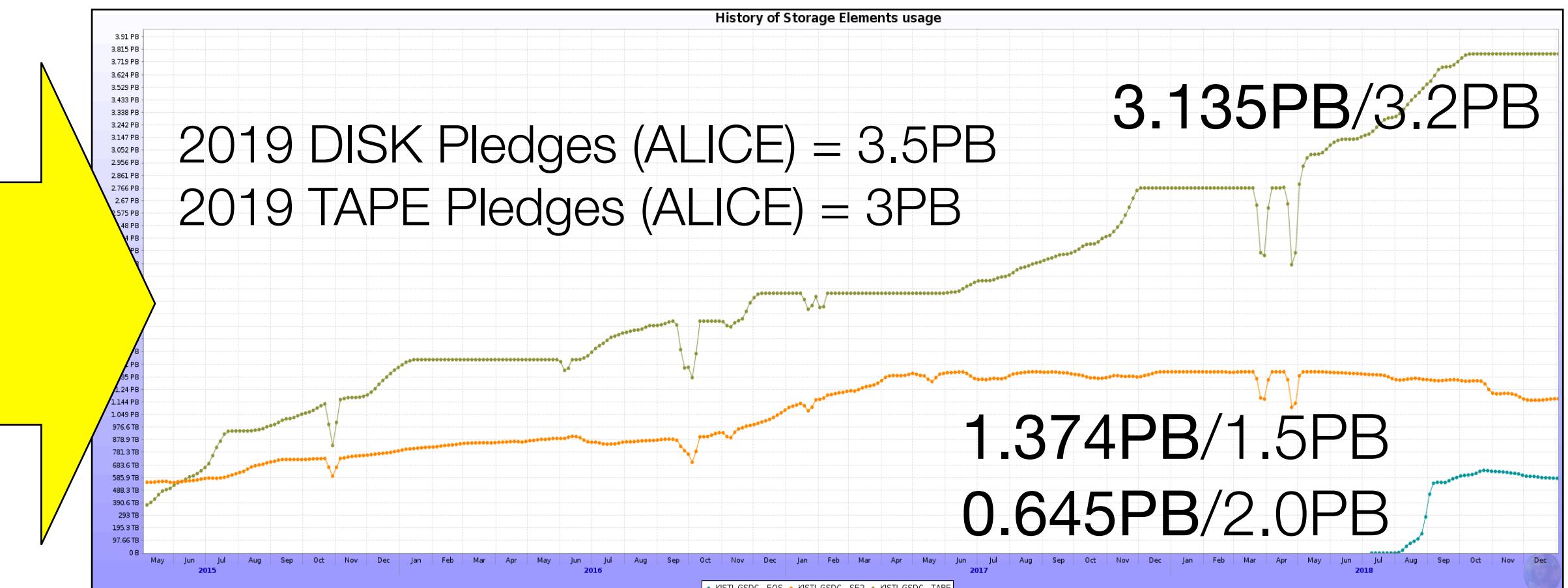
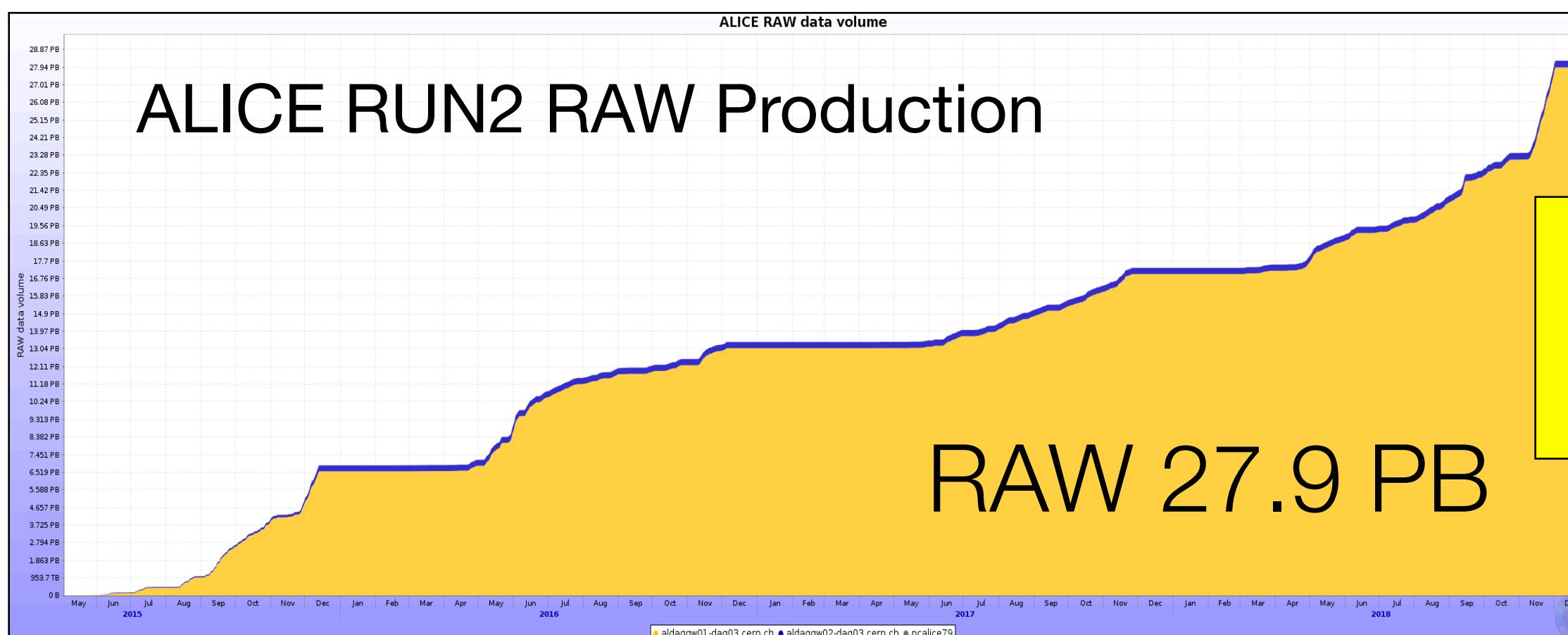
Total CPU power for ALICE jobs (RUN2)



Active Jobs (RUN2)

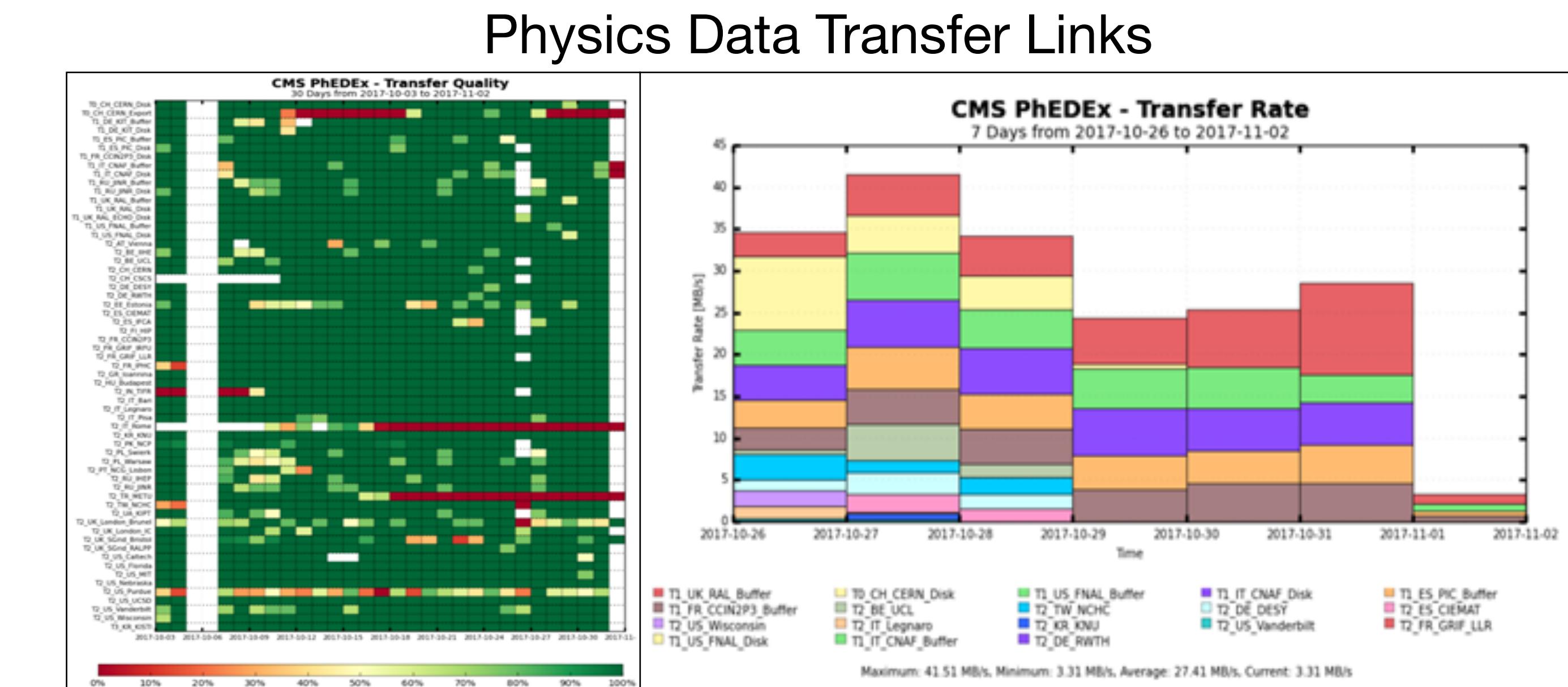
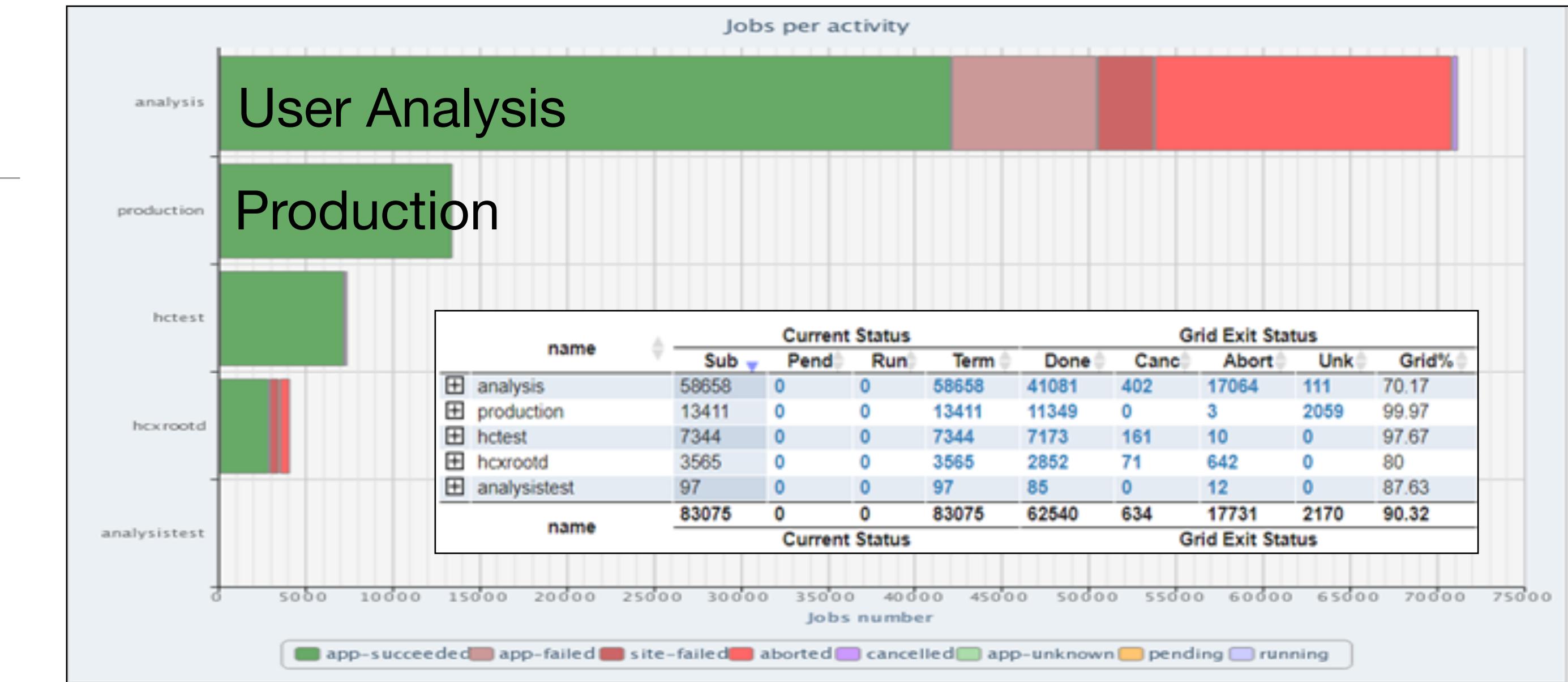
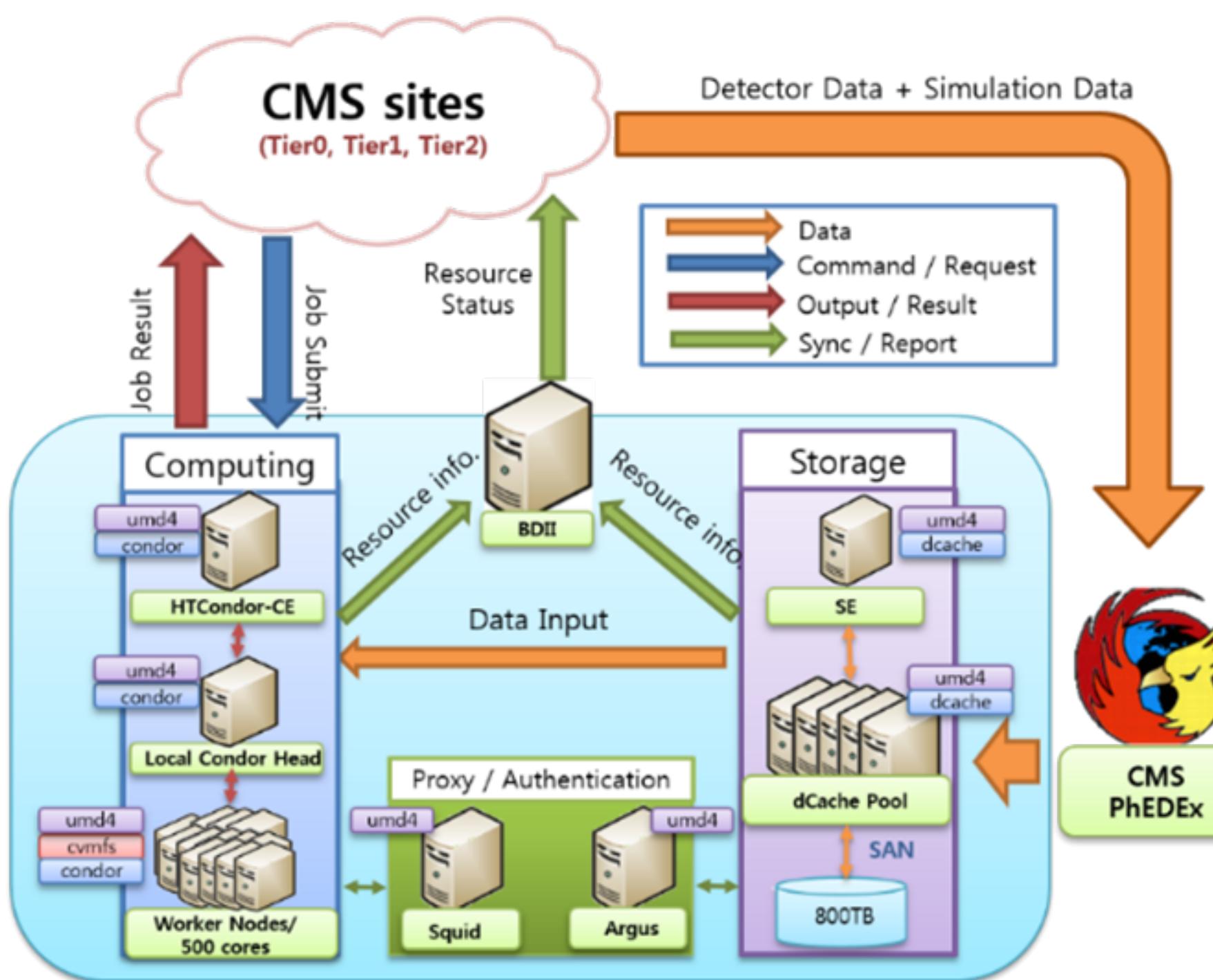


4.9% Contribution to Total ALICE Grid Computing Power



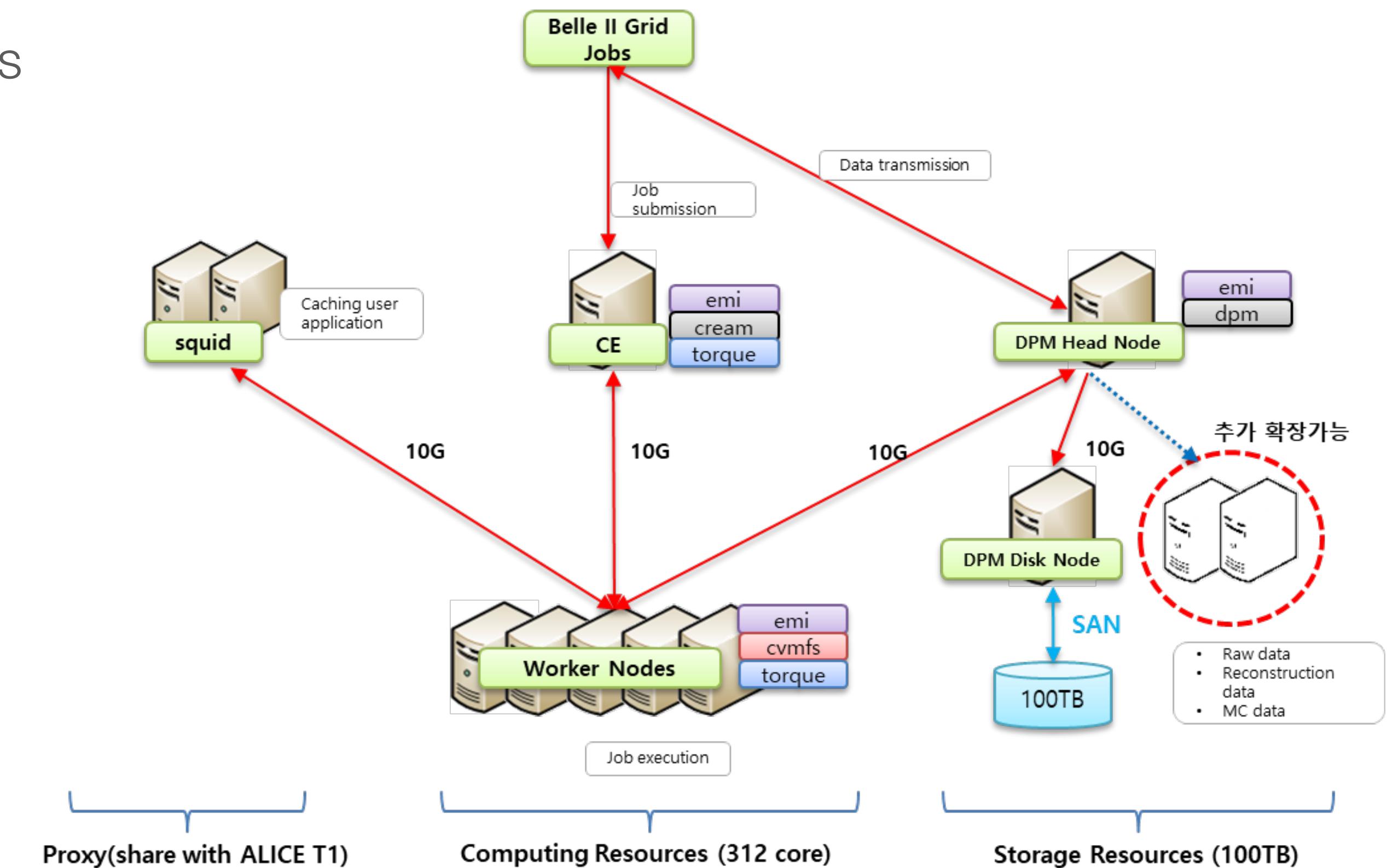
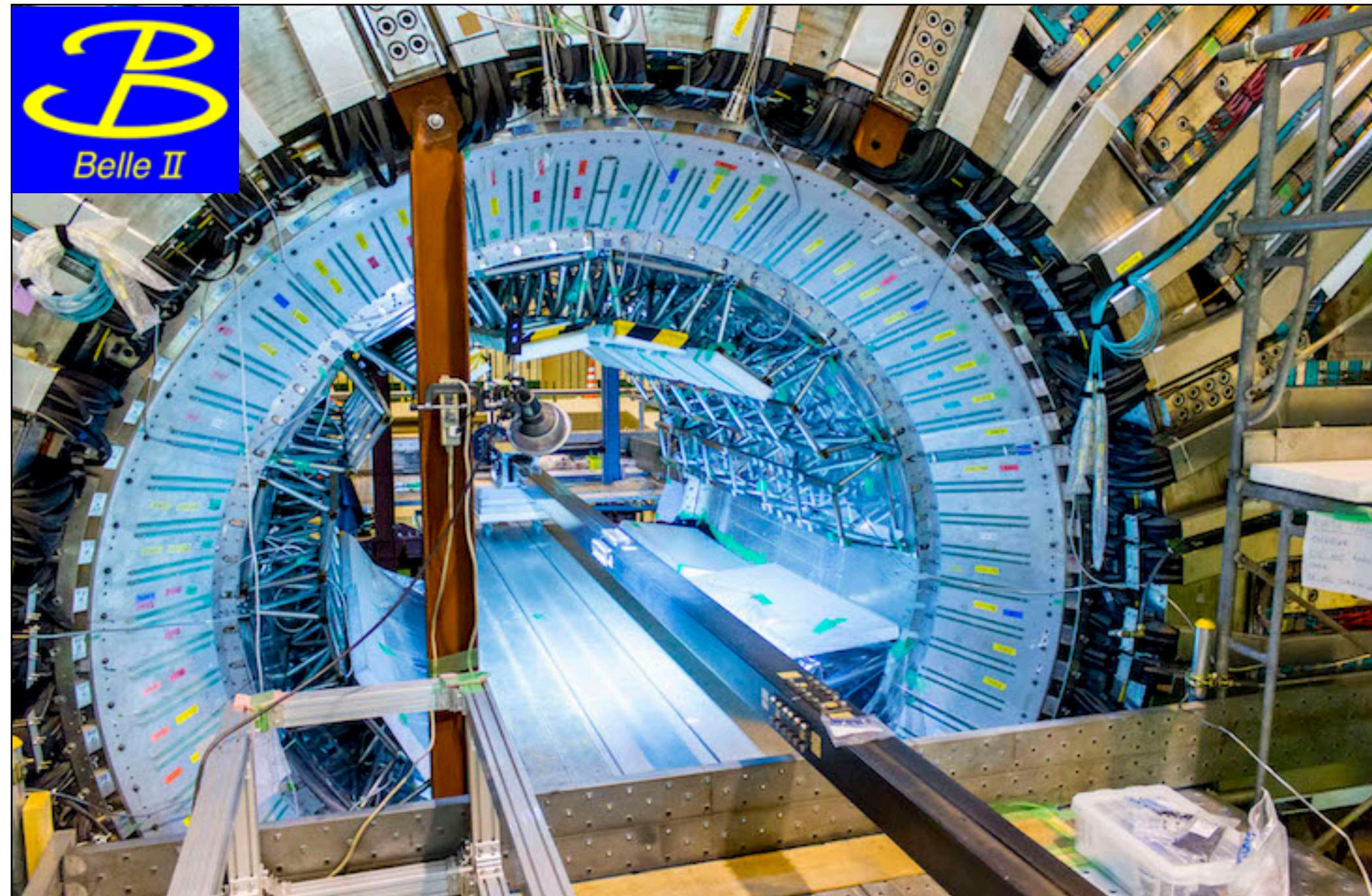
WLCG Tier-2 for CMS

- Service in production since October 2017
 - WLCG MoU signed in April 2018
 - Pledges: 10kHS06 CPU, 800TB Disk



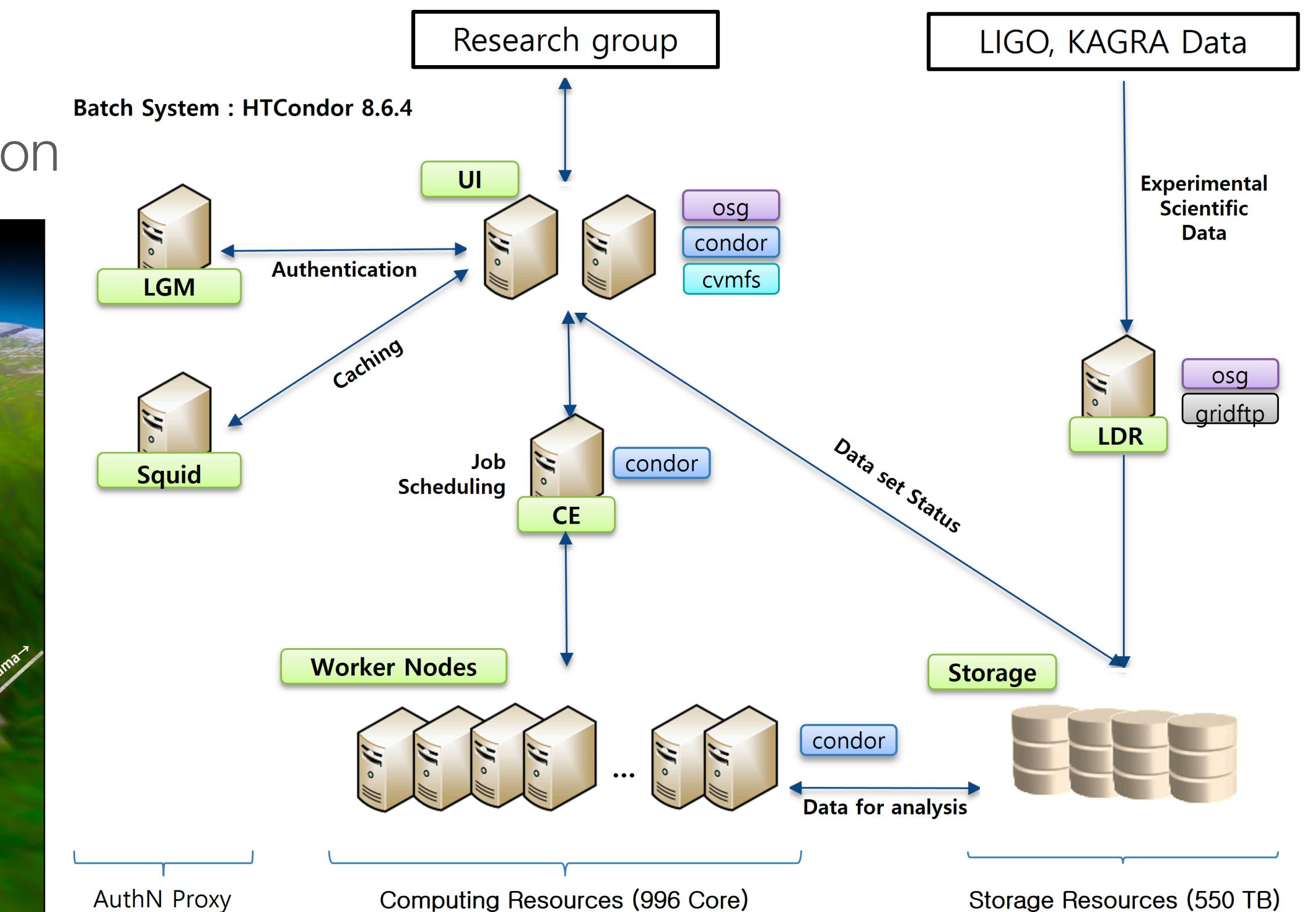
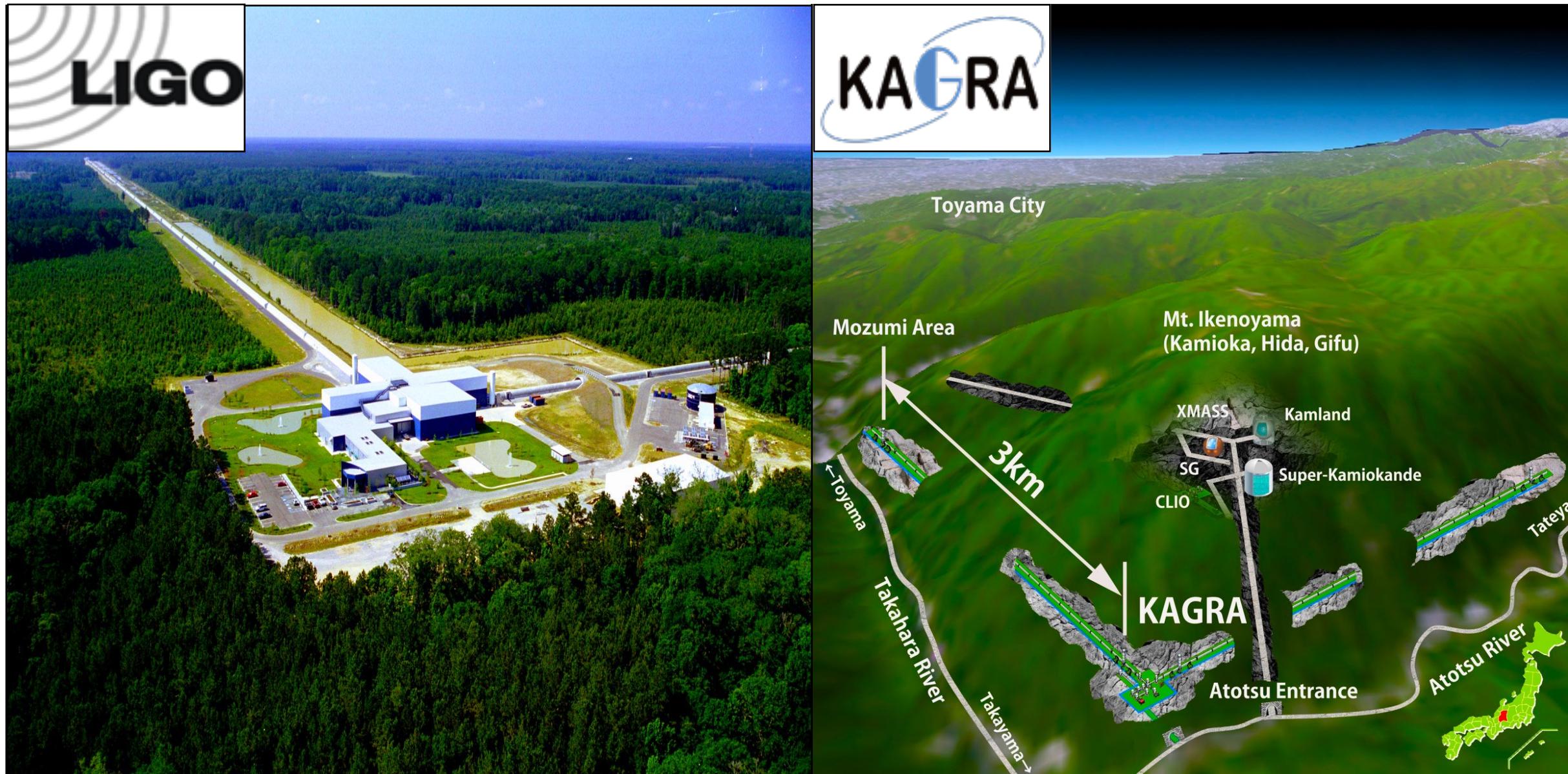
Belle II Computing Support

- General agreement on collaborating in computing for Belle II Experiment was established in Nov. 2016
- Additional agreement is required for computing resources contribution
- Belle II Grid services were deployed on the legacy system for Belle experiment
- Resources: 300 CPUs (=2.7kHS06), 100TB Disks



LIGO & KAGRA Supports

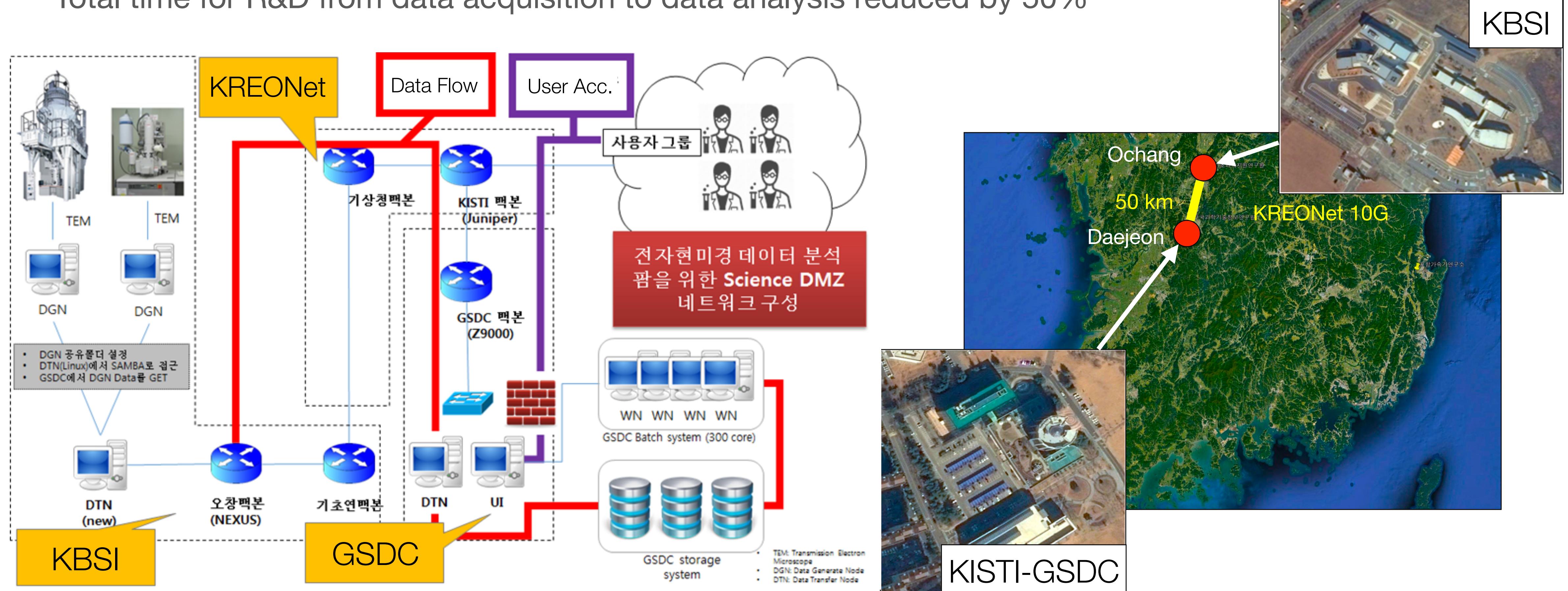
- Pioneering experiment to detect cosmic gravitational waves and to develop gravitational-wave observations
- Whole observations data produced from LIGO experiments transferred to GSDC
- Currently, supports at Tier-3 level for LIGO Data Grid
- Resources: 1,000 CPUs, 550TB Disk
- Global services for LDG are demanding and in preparation



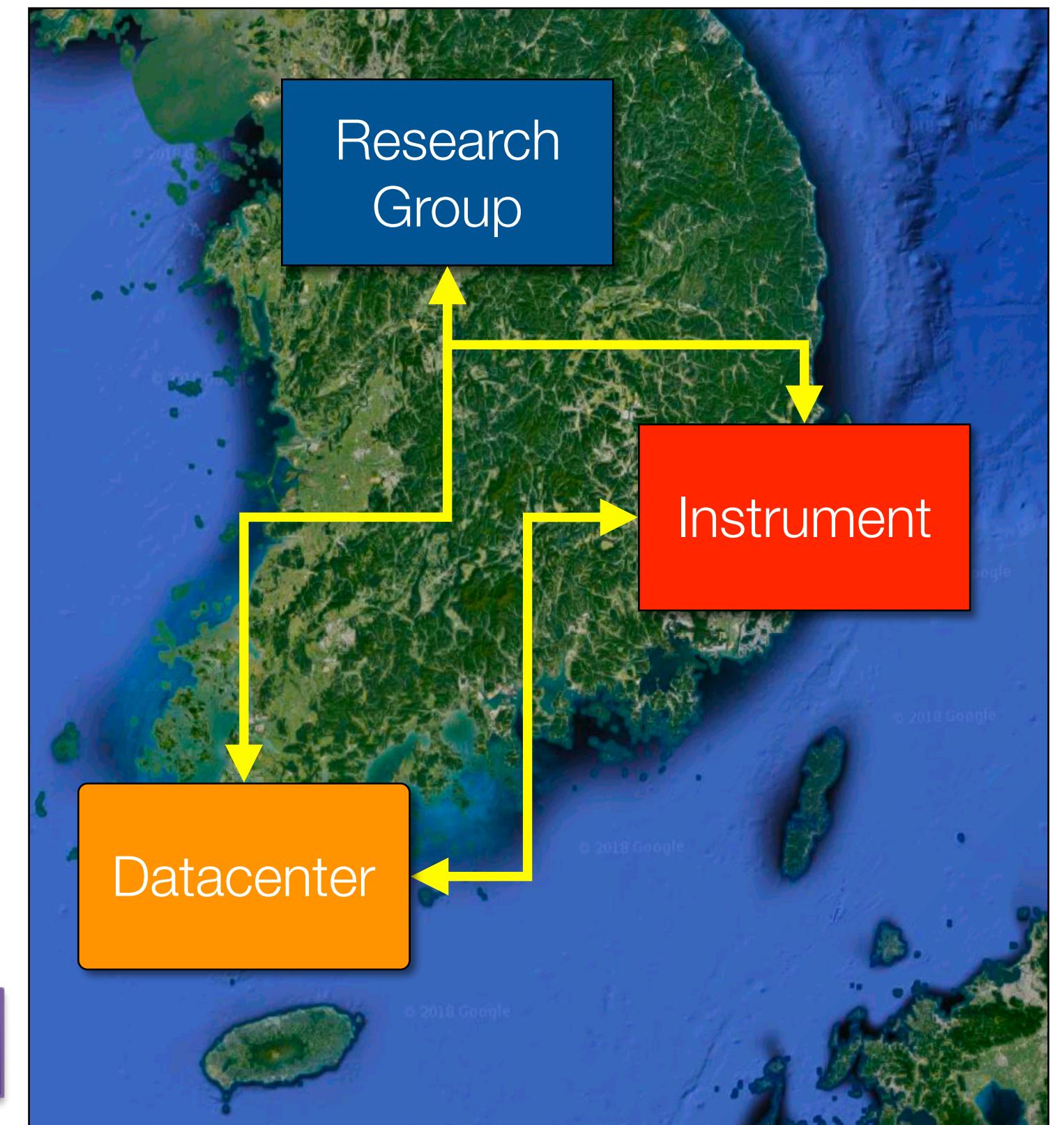
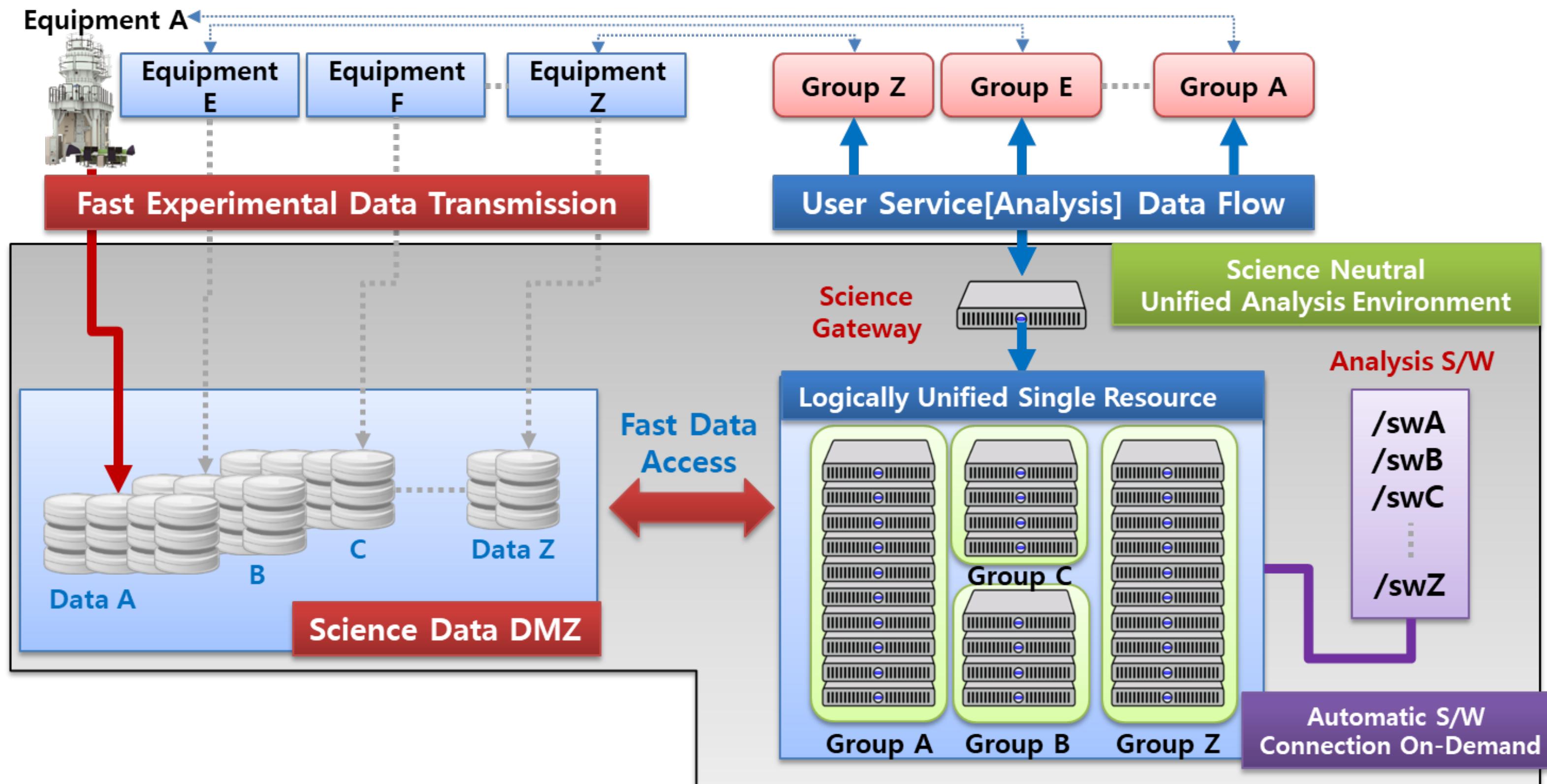
Role of Datacenter to support Domestic R&D activities

Transmission Electron Microscopy for Structural Biology

- First attempt to implement Science DMZ model for research community support
- Total time for R&D from data acquisition to data analysis reduced by 50%



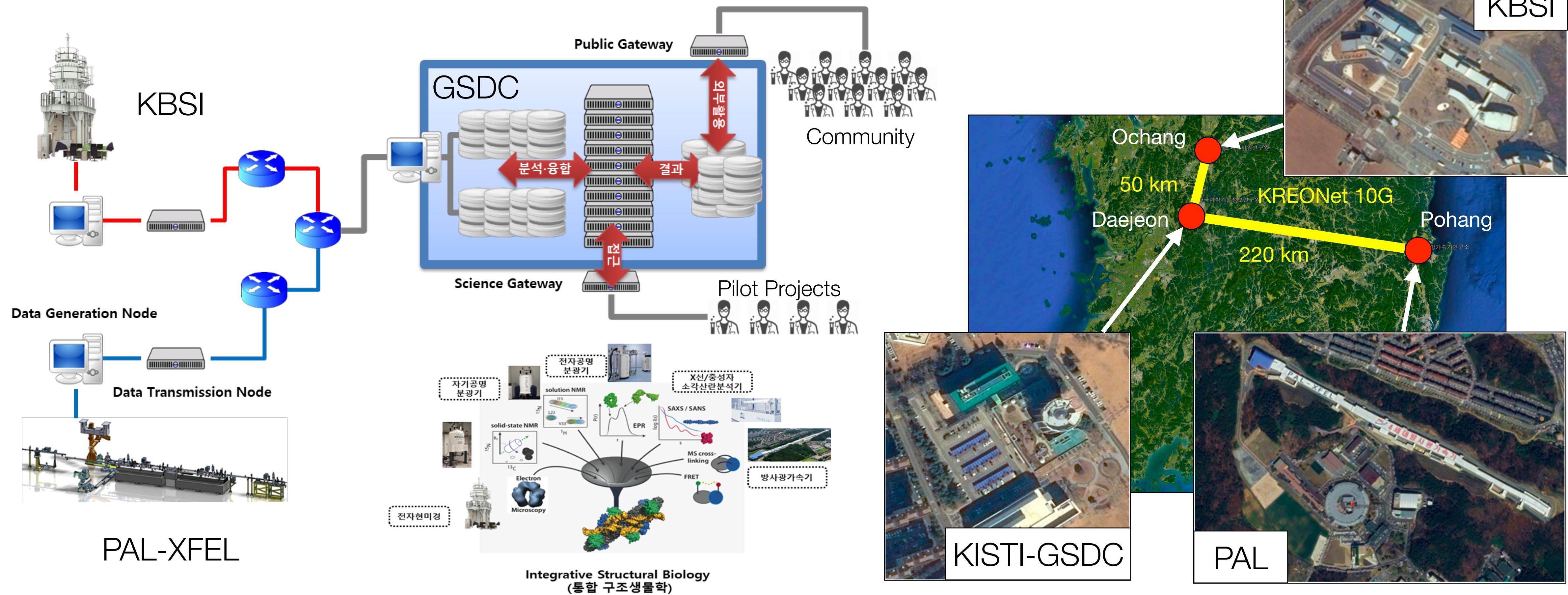
Unified Data Analysis Platform for Large-scale Facilities & Instruments



- Simplify R&D process from DAQ to Analysis
- Applicable to Korean R&D environment: research group, facility or instruments, datacenter are not in one place
- A transparent layer of computing platform is necessary to make R&D activities fluent and efficient

A Pilot Project for Sharing Data from Heterogenous Instruments

- Establishing physical links for data transfer and systems for sharing and analyzing data
- Collective data analyses feasible by using EM data and X-ray data



Summary

- KISTI-GSDC is a datacenter to support data-intensive research fields in Korea and its infrastructure has been growing
- We supports not only WLCG but also other VOs, e.g. Belle II and LIGO/KAGRA and so on
- We implemented Science DMZ Model to support various kinds of data-intensive research in domestic region by considering typical Korean R&D environments and activities
- Also we launched pilot projects to promote data sharing and application in the field where large scale research facilities are used under the governance of national research data platform

GSDc Promoting Science

Thank you