GSDC: Datacenter for Data-intensive Research

Sang-Un Ahn for KISTI-GSDC
“e-Science Activities in Korea” @ ISGC2018

Contents

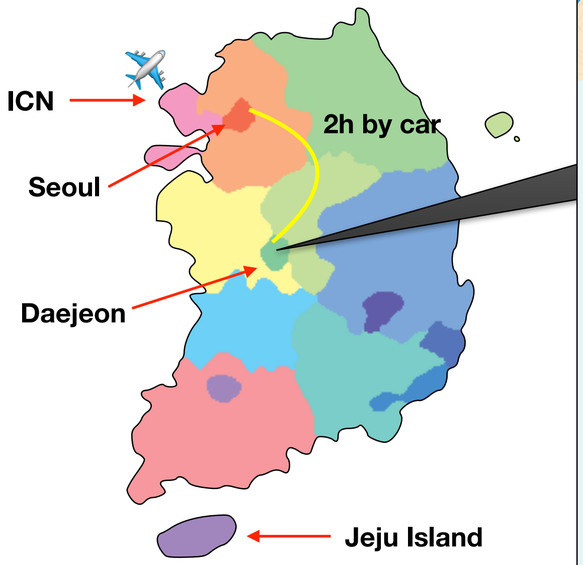
Introduction to KISTI - GSDC

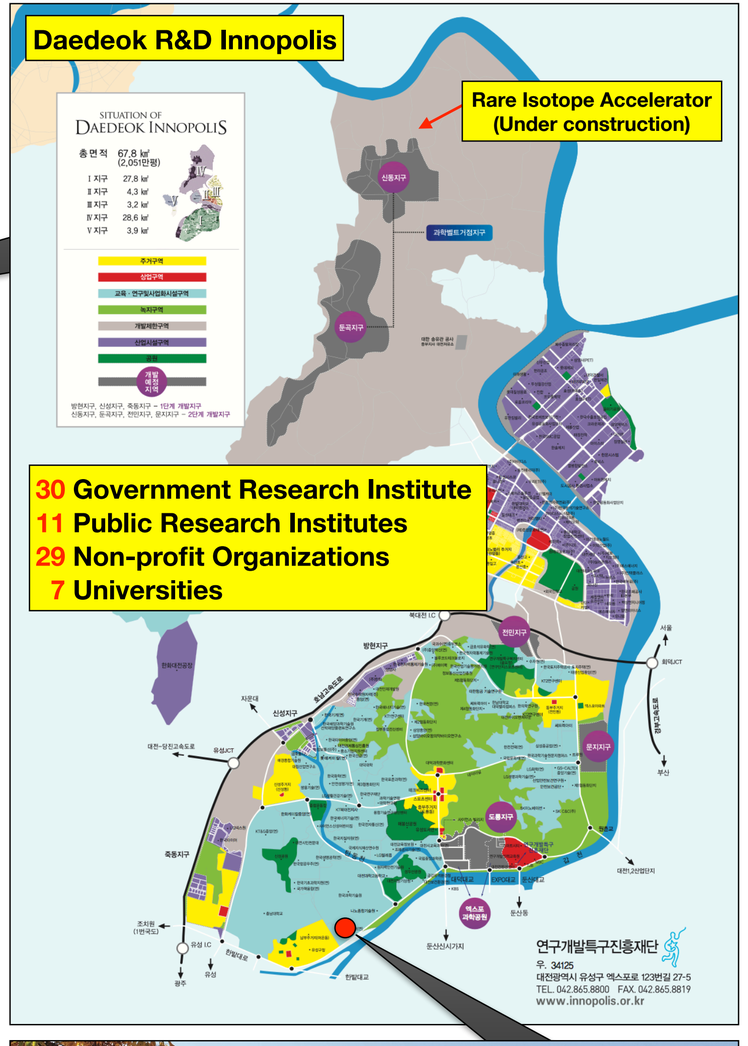
KISTI

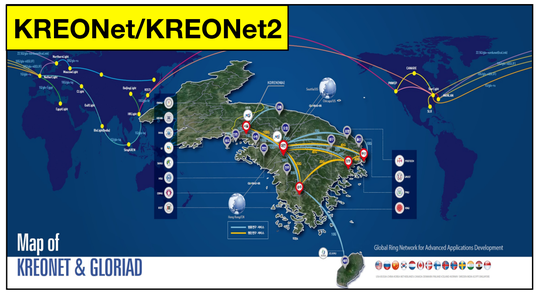
Korea Institute of Science and Technology Information

- Government-funded research institute founded in 1962 for National Information Services and Supercomputing
- National Supercomputing Center
  - Cray CS500 system (being installed)
    - 25.7 PFlops at peak, expected to be ranked 10th of Top500 (2018)
  - Tachyon II system (~307.4 TFlops, 2008)
  - KREONet/KREONet2 - National R&E network

Map of South Korea



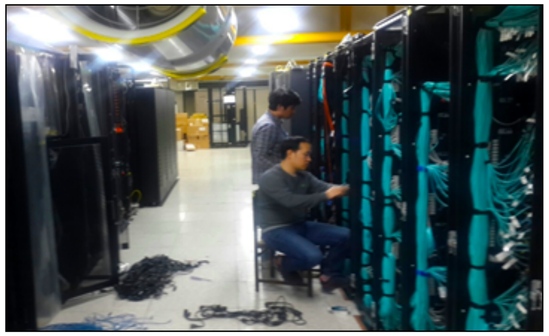


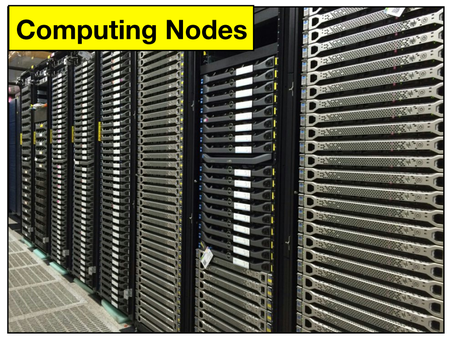


GSDC

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
  - 16 staff: system administration, experiment support, external-relation, management and planning





Experiments Support

Heavy-Ion Physics
WLCG Tier-1
(CPU) 3,500 cores / (DISK) 3,000 TB / (TAPE) 3,000 TB
Tier-3 : (CPU) 400 cores / (DISK) 600 TB

Elementary Particle Physics
Tier-3 : (CPU) 1,000 cores / (DISK) 600 TB
WLCG Tier-2
(CPU) 1,000 cores / (DISK) 800 TB

Astrophysics
LDG Tier-3
(CPU) 800 cores / (DISK) 600 TB
LDG Tier-2 (TBD)

Medical Science
(CPU) 600 cores / (DISK) 700 TB

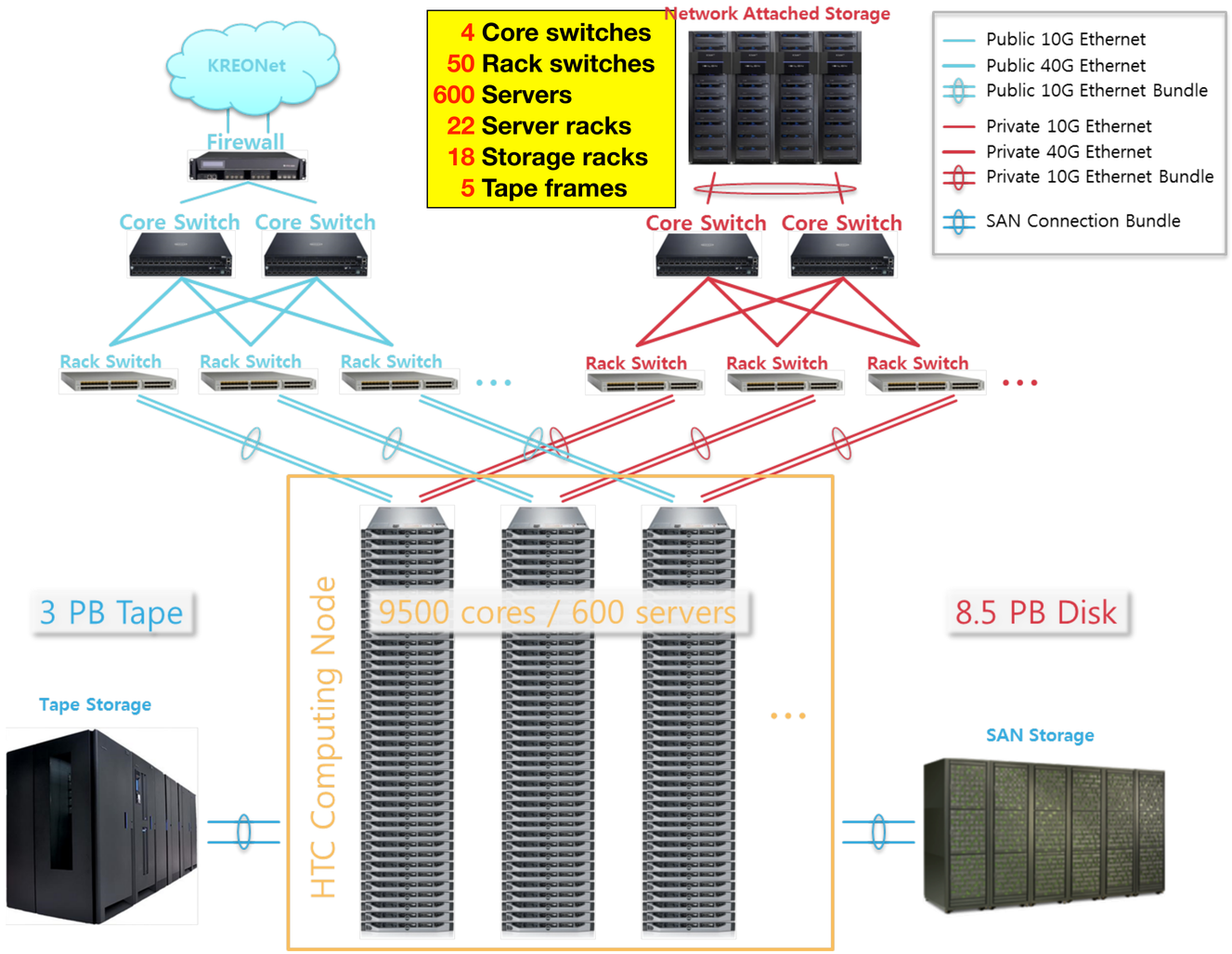
Elementary Particle Physics
B2G Tier-2 (TBD)
(CPU) 300 cores / (DISK) 100 TB

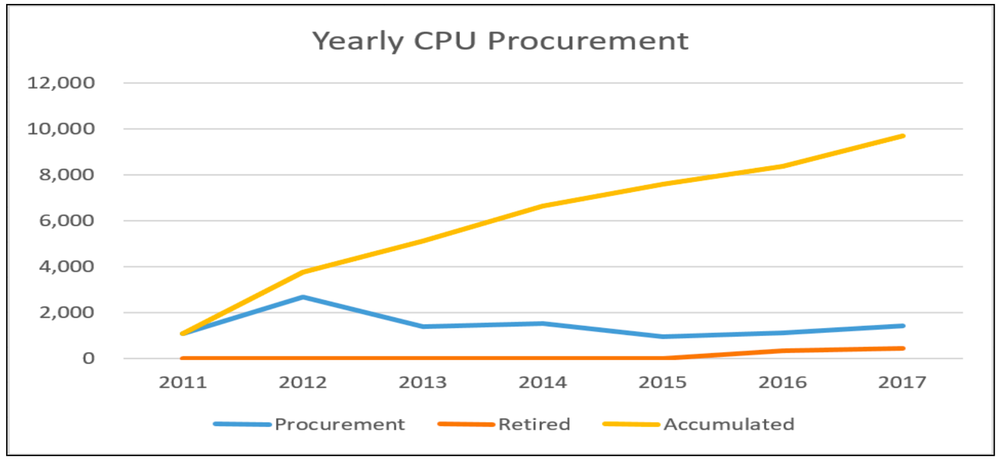
Neutrino Physics
RAW storage
(CPU) 360 cores / (DISK) 700 TB

Structural Biology
(CPU) 700 cores / (GPU) 6 nodes / (DISK) 700 TB

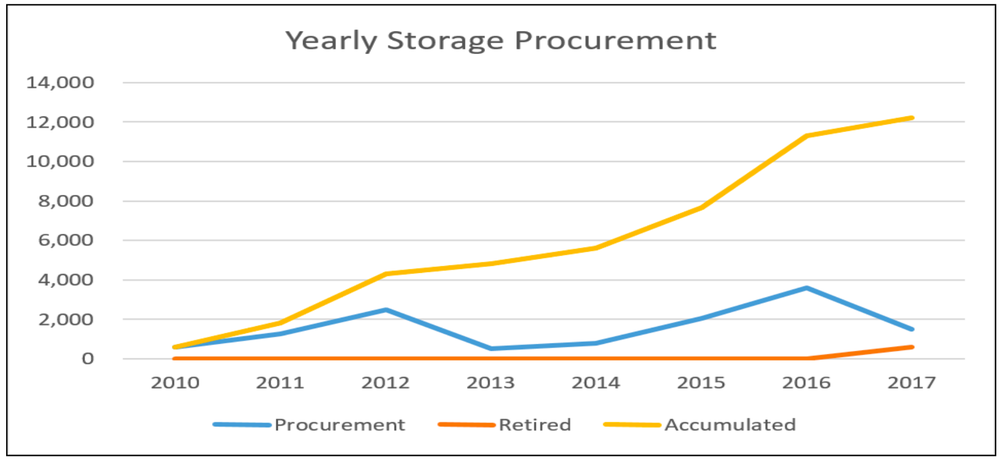
Datacenter Infrastructure

Infrastructure



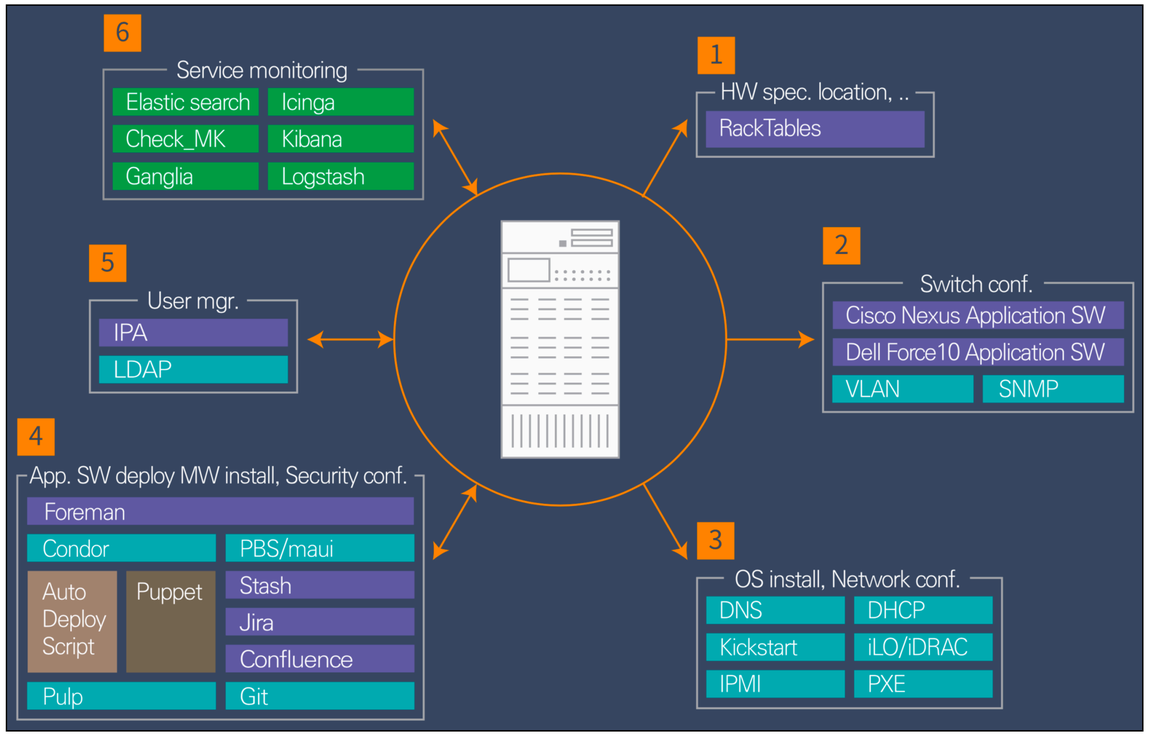


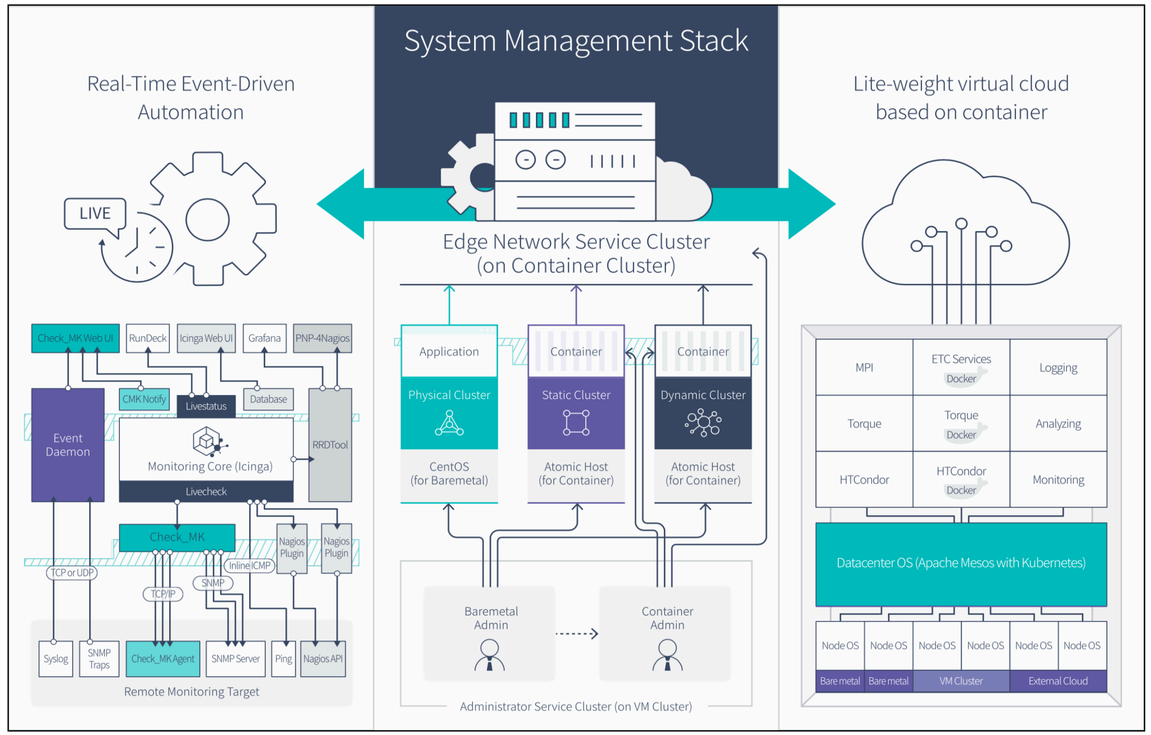
1.5K CPUs & 1.5PB Storage
Procured Every Year on Average



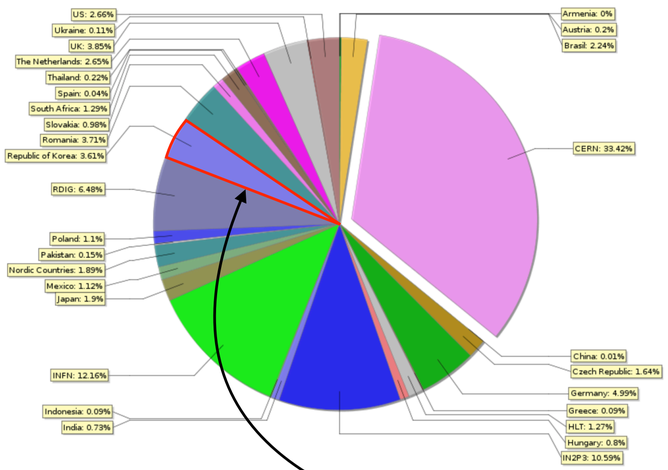
System Management

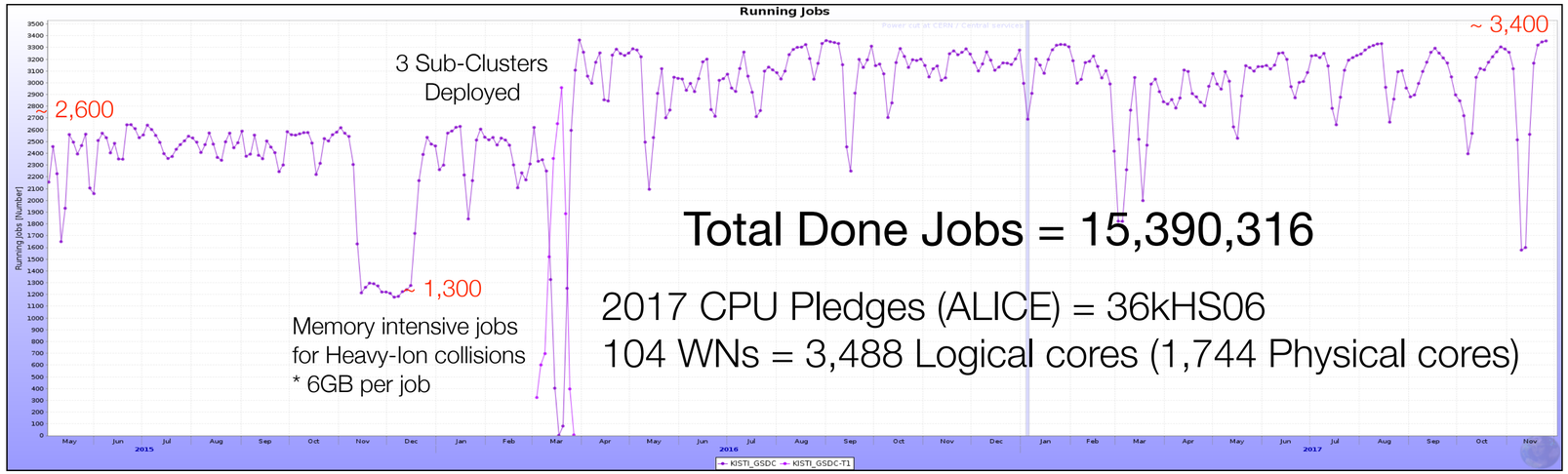
- +600 Servers and +25 Storage
- Every year of procurement… heterogenous systems with different vendors and specifications : Challenges !!
- Automated provisioning and configuration management are key for efficient and stable operations
- Moving towards virtual infrastructure based on container



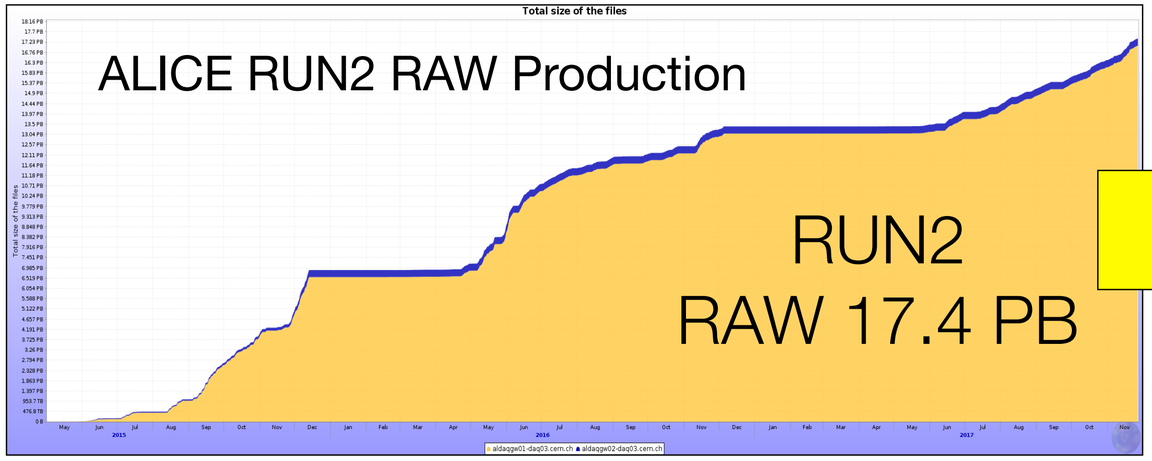


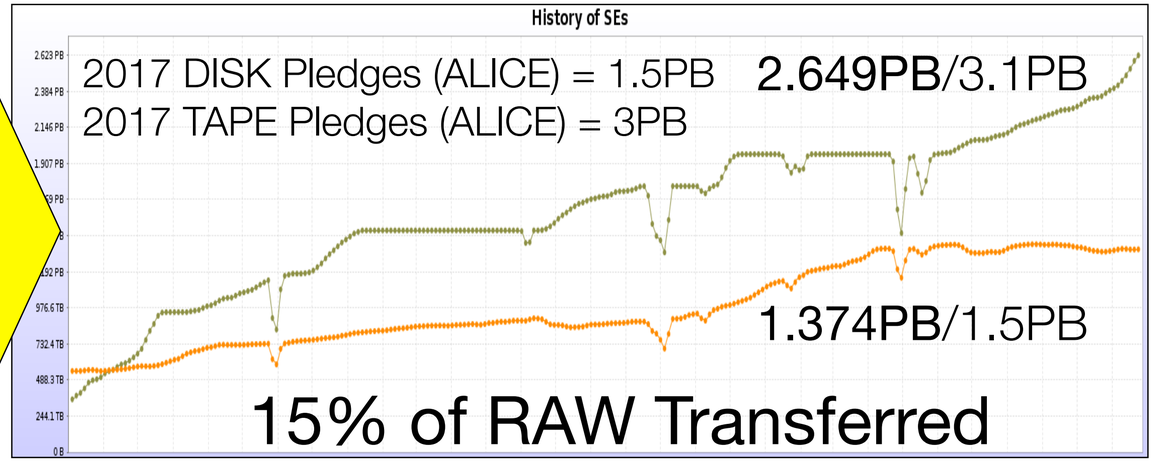
WLCG Tier-1 for ALICE (LHC RUN2 Operations)





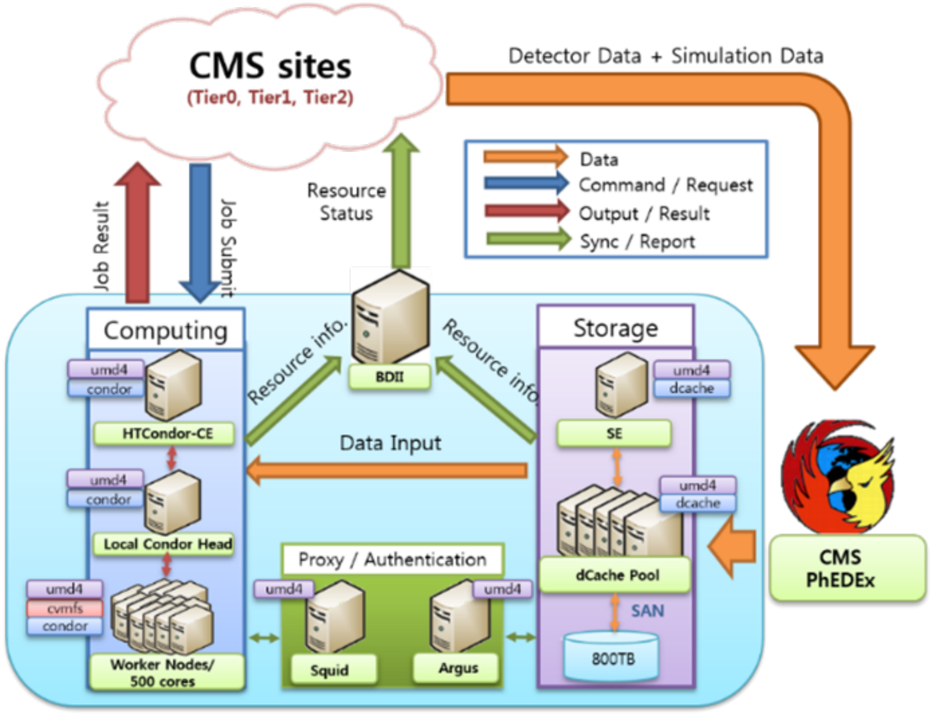
3.61% of Total ALICE Computing Resources

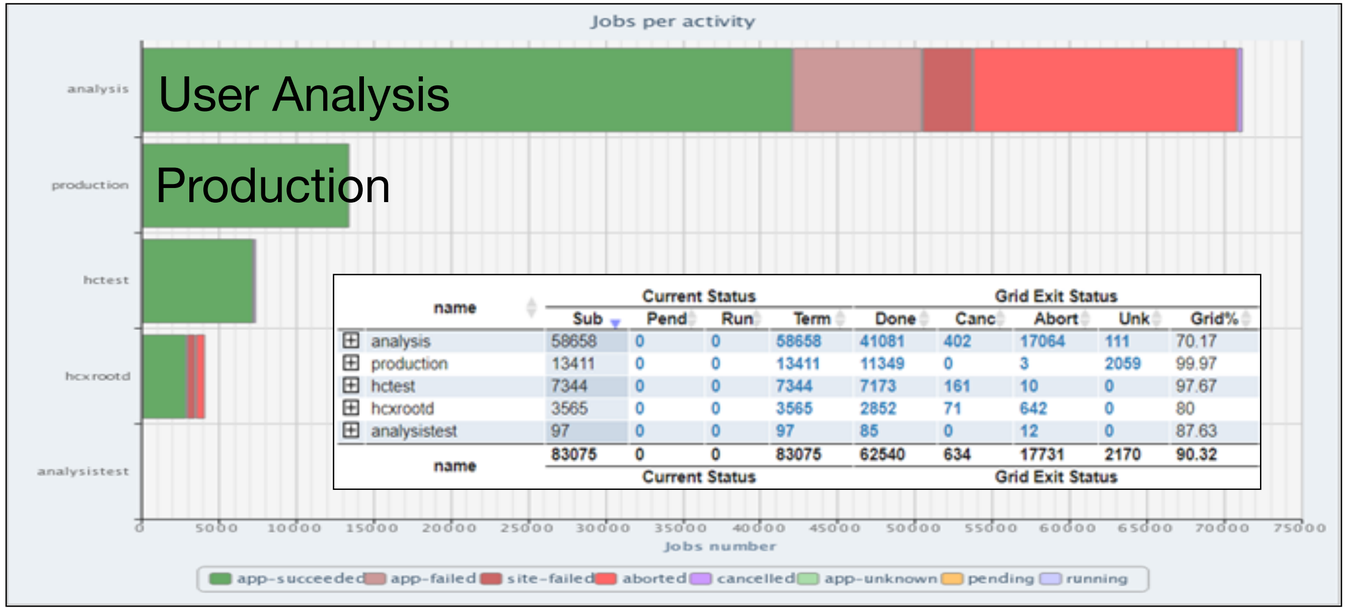




WLCG Tier-2 for CMS

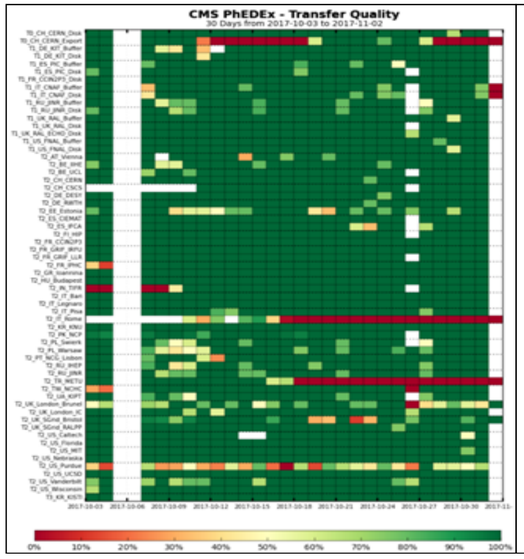
- Service in production since the last October
- WLCG MoU will be signed in April 2018
- Pledges: 10kHS06 CPU, 800TB Disk

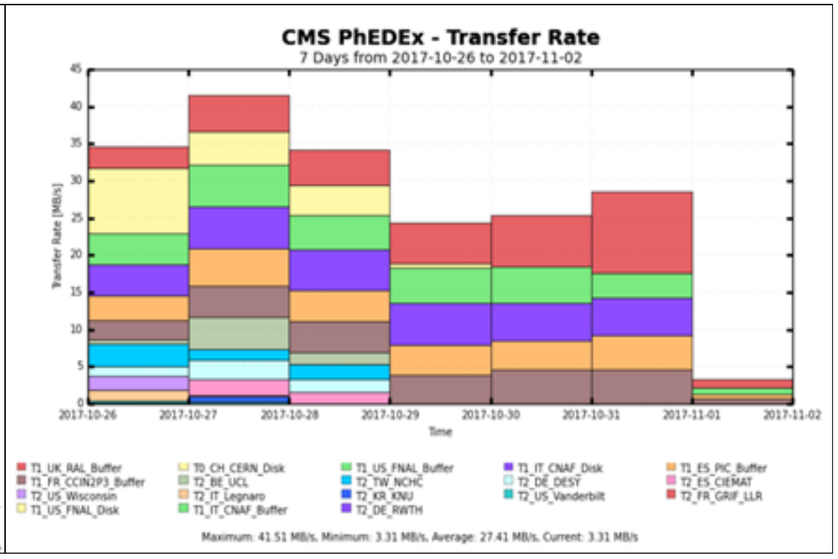




name	Current Status				Grid Exit Status				
	Sub	Pend	Run	Term	Done	Canc	Abort	Unk	Grid%
analysis	58658	0	0	58658	41081	402	17064	111	70.17
production	13411	0	0	13411	11349	0	3	2059	99.97
hctest	7344	0	0	7344	7173	161	10	0	97.67
hcxrootd	3565	0	0	3565	2852	71	642	0	80
analysistest	97	0	0	97	85	0	12	0	87.63
name	83075	0	0	83075	62540	634	17731	2170	90.32
	Current Status				Grid Exit Status				

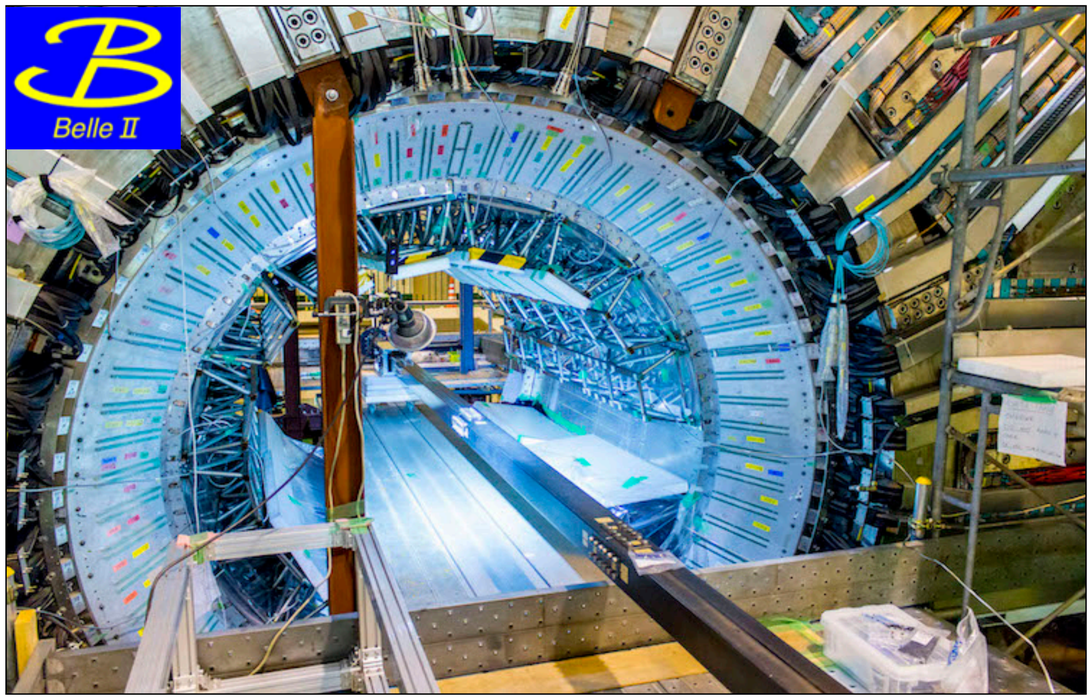
Physics Data Transfer Links

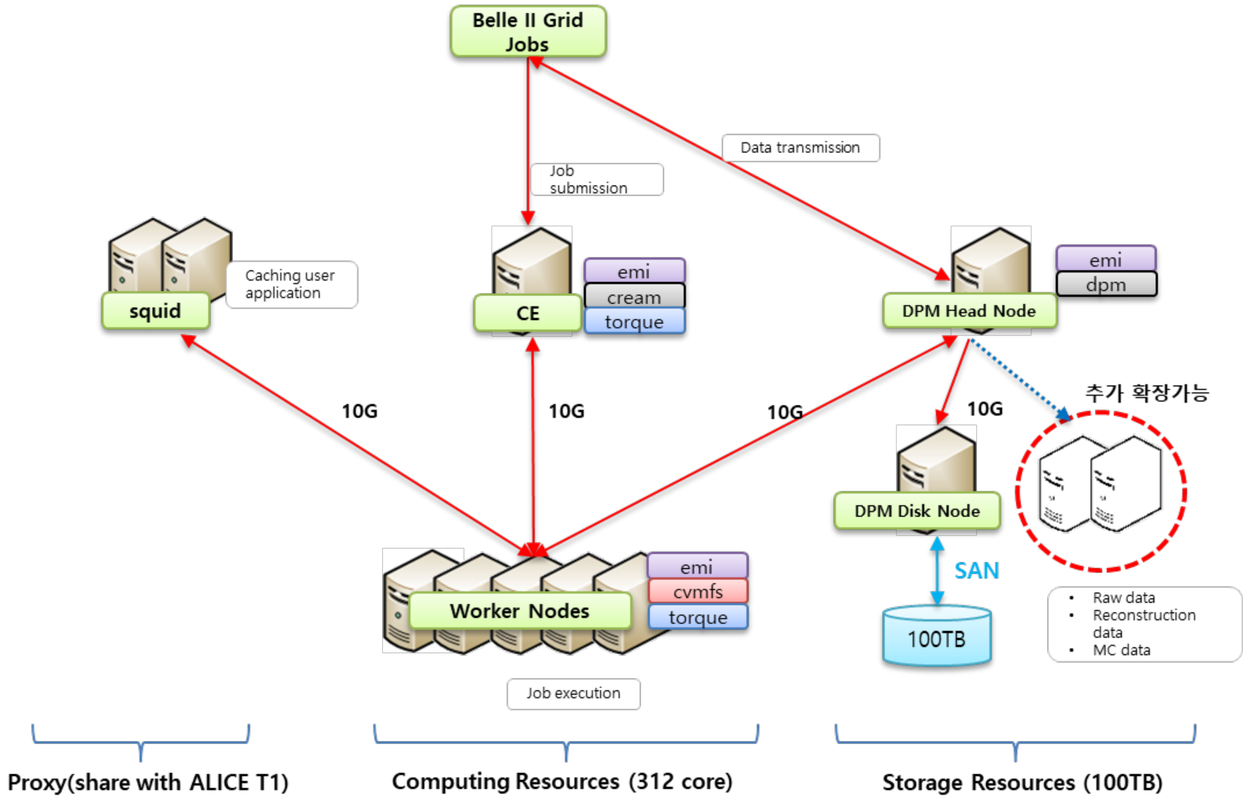




Belle II Computing Support

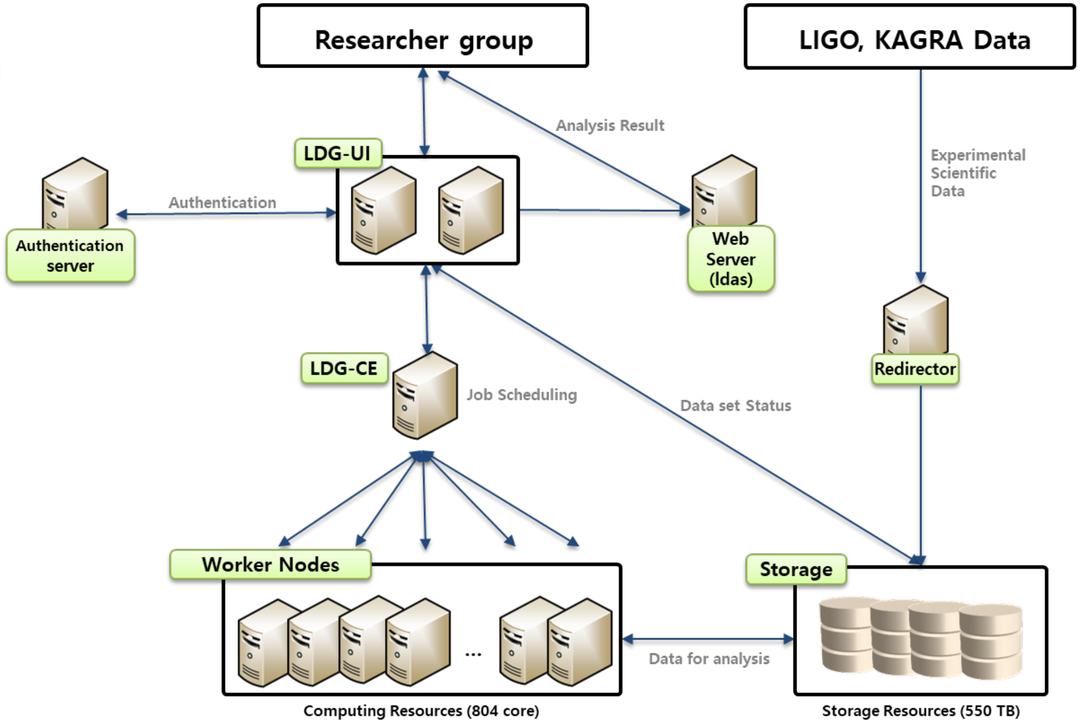
- **General agreement on collaborating in computing for Belle II Experiment** was established in Nov. 2016
- Additional agreement is required for computing resources contribution
- Belle II Grid services were deployed on the legacy system for Belle experiment
- Resources: 300 CPUs (=2.7kHS06), 100TB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





LIGO & KAGRA Supports

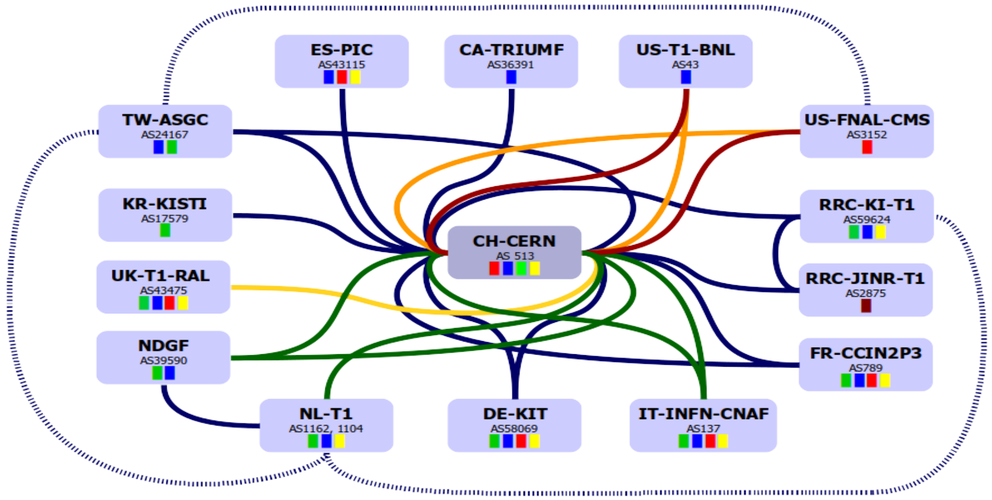
- Pioneering experiment to detect cosmic gravitational waves and to develop gravitational-wave observations
- **Whole observations data produced from LIGO experiments transferred to GSDC**
- Currently, supports at Tier-3 level for LIGO Data Grid
- Resources: 1,000 CPUs, 550TB Disk
- Global services for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



Networking

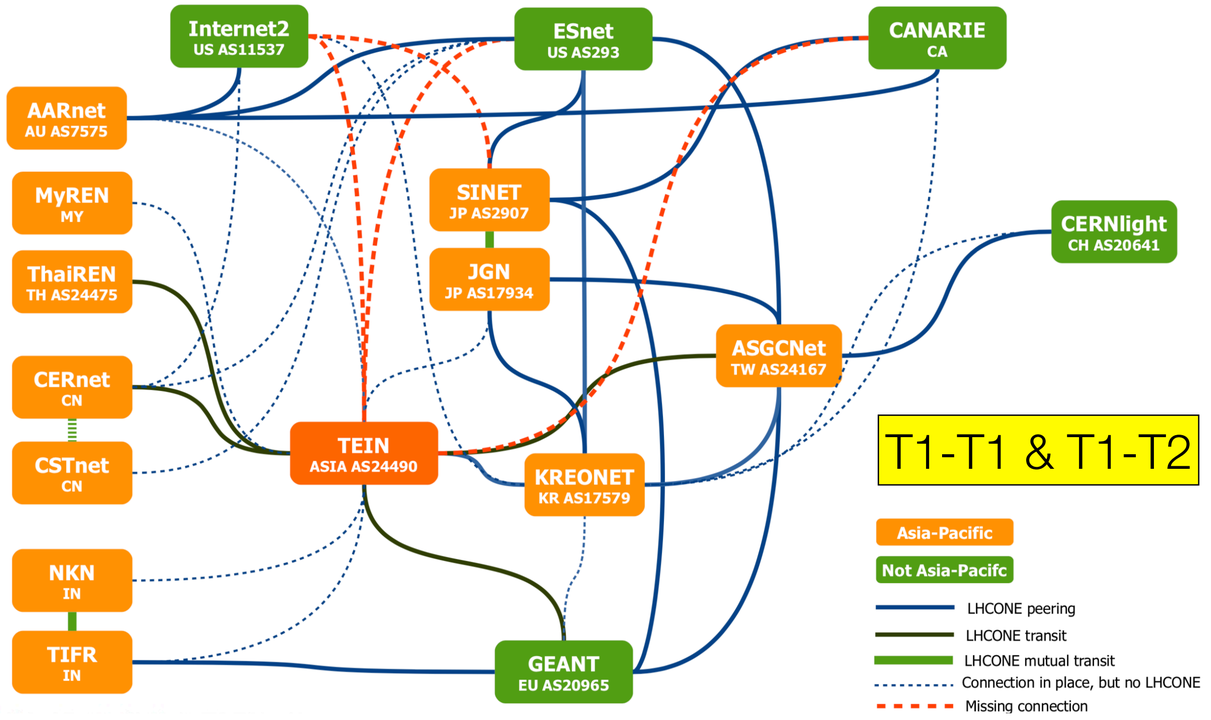
LHCOPN

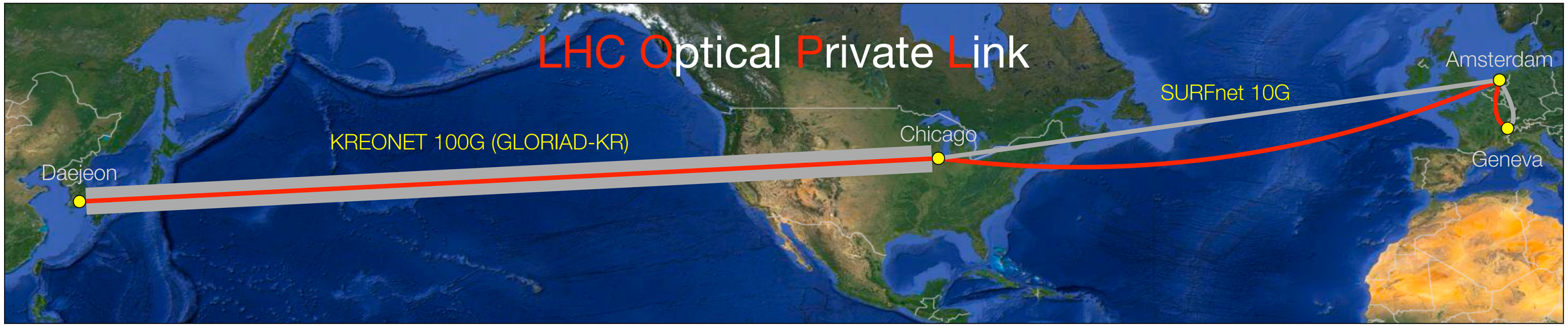
T0-T1 & T1-T1



LHCONE in Asia

www.atcforum.org

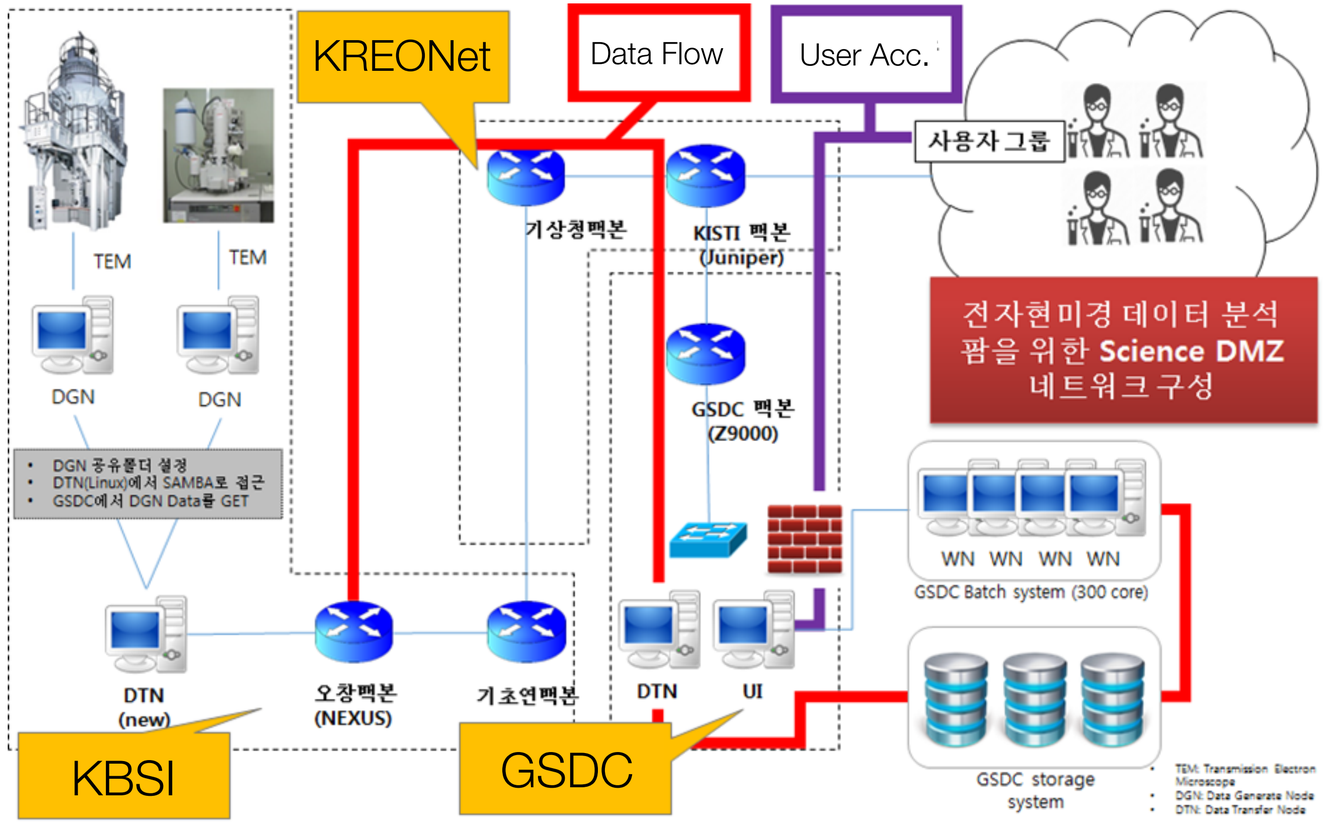




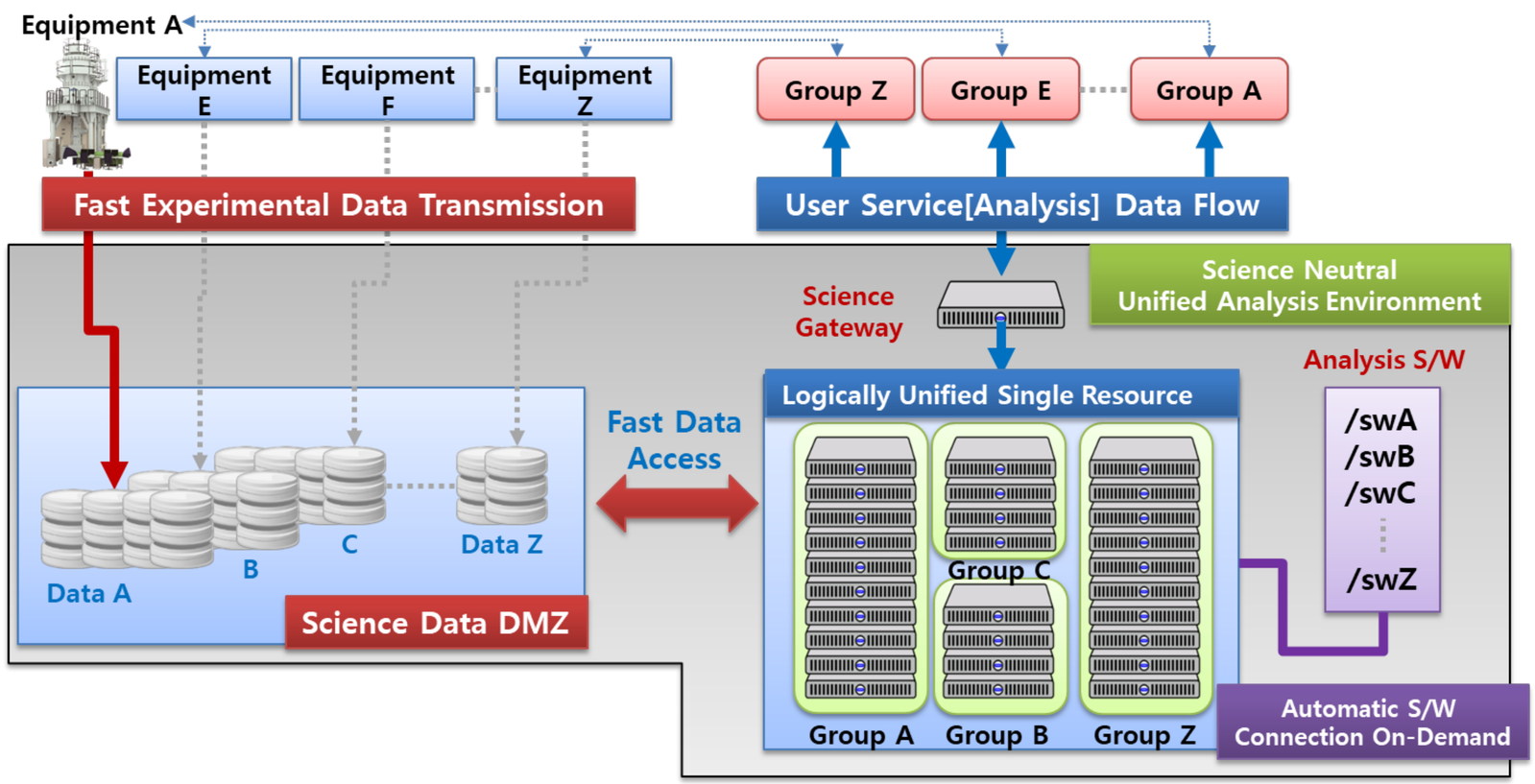
Role Expansion to National Datacenter

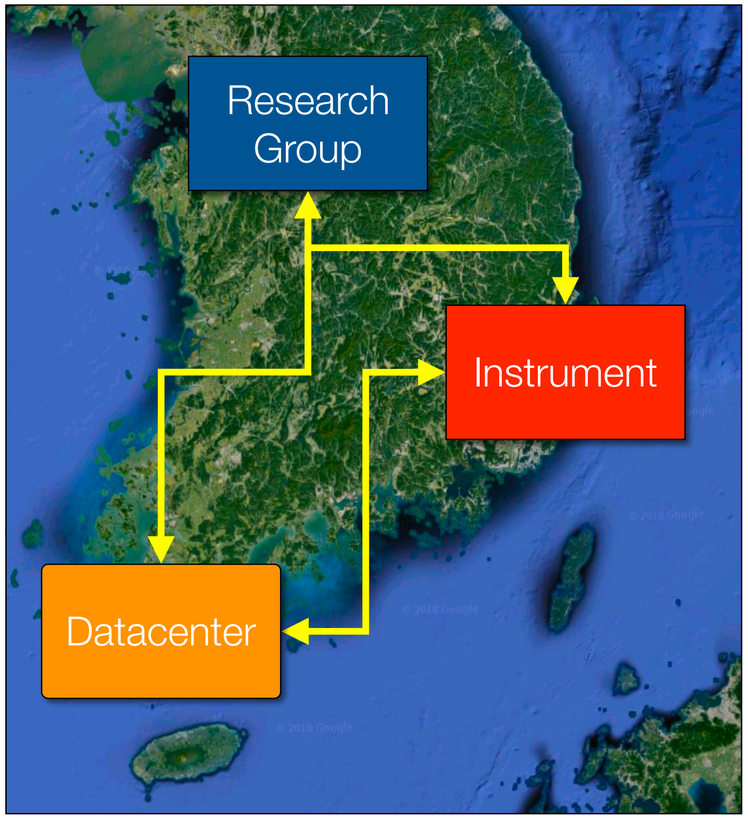
Transmission Electron Microscopy for Structural Biology

- 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

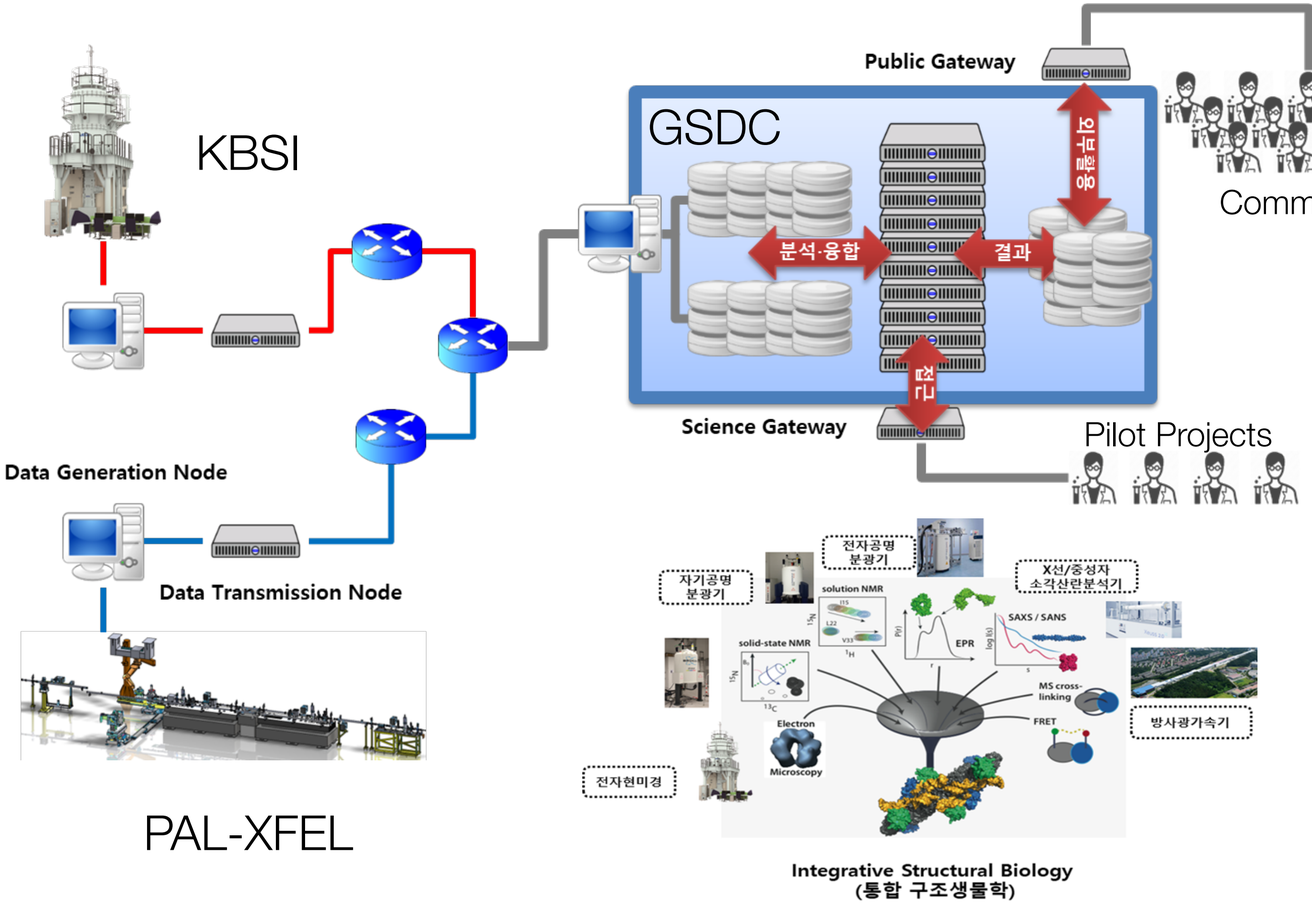


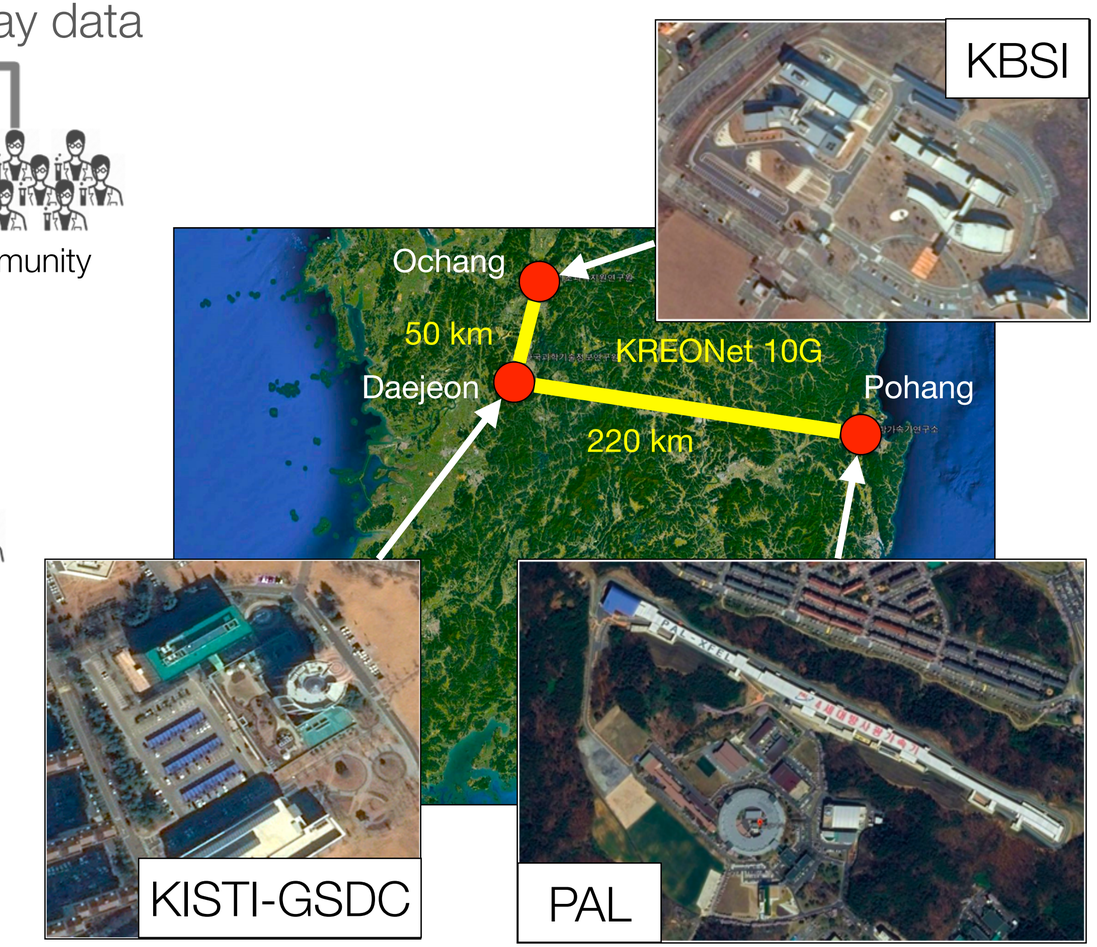


- Simplify R&D process from DAQ to Analysis
- Applicable to Korean R&D environment: research group, facility or instruments, datacenter are not in one place
- A transparent layer of computing platform is necessary to make R&D activities fluent and efficient

Pilot Projects 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





Government Movement towards Data Sharing and Application

- “모아서 새롭게(Collect and Renew)” TF
  - Establishing a governance for research data sharing and application in terms of “Open Science”
    - Preparing for legal and institutional basis to collect/manage/share R&D data and its application
    - Capacity building to strengthen the capability for collaborative work
    - Establishing a platform for R&D data share and application and supporting cloud-like infrastructure
  - Pilot projects in Bioinformatics, Materials, Large-scale Facilities & Instruments, and Artificial Intelligence

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 have implemented Science DMZ Model to support various kinds of data-intensive research in domestic region by considering typical Korean R&D environments and activities
- Recently we have launched pilot projects to support Integrative Structural Biology using Cryo-EM as well as XFEL, which is the accordance with the Government-driven strategy to foster data sharing and application in terms of “Open Science”
- KISTI-GSDC will expand its role towards a national datacenter for data-intensive research

GSDc Promoting Science

Thank you