



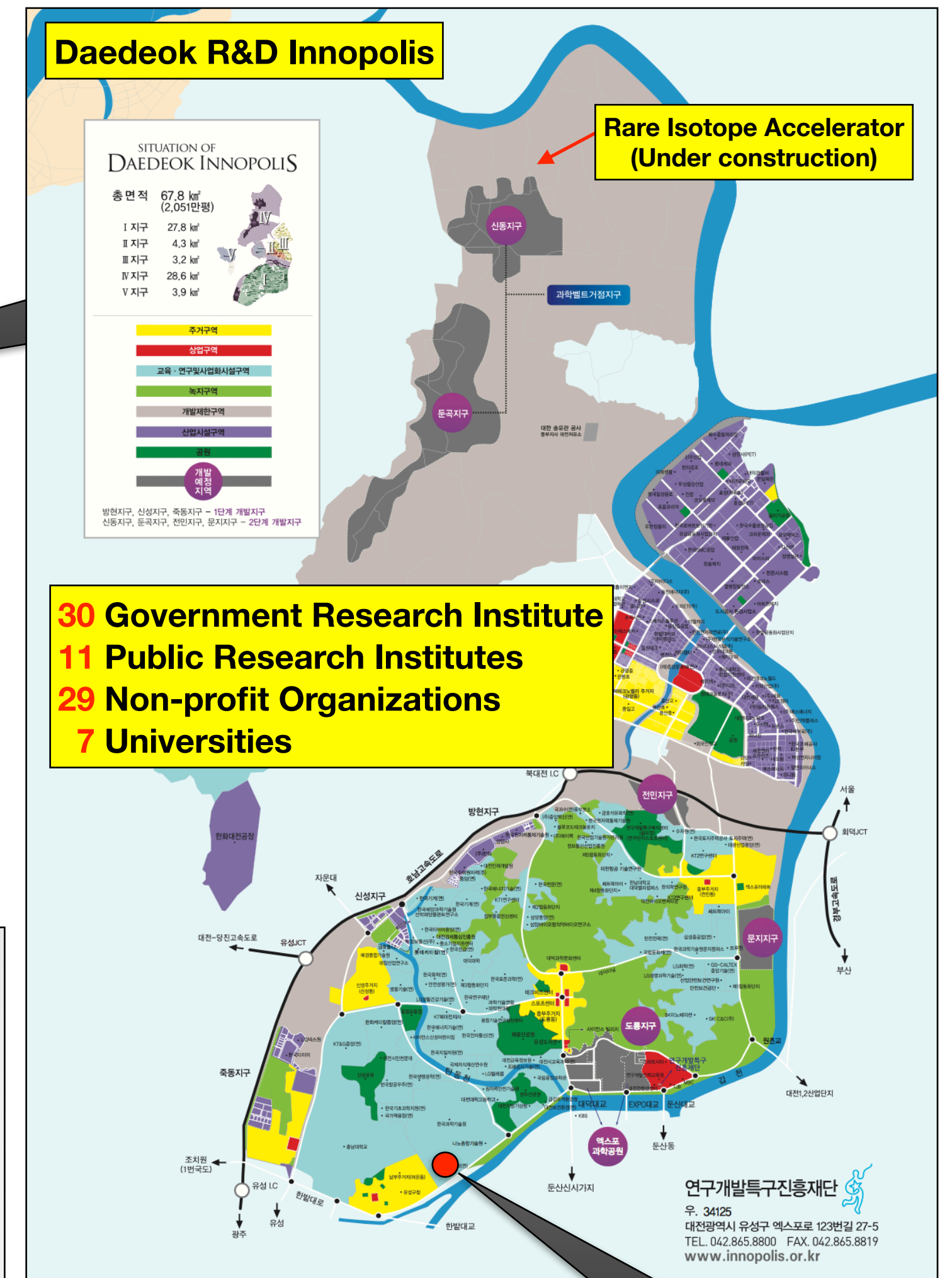
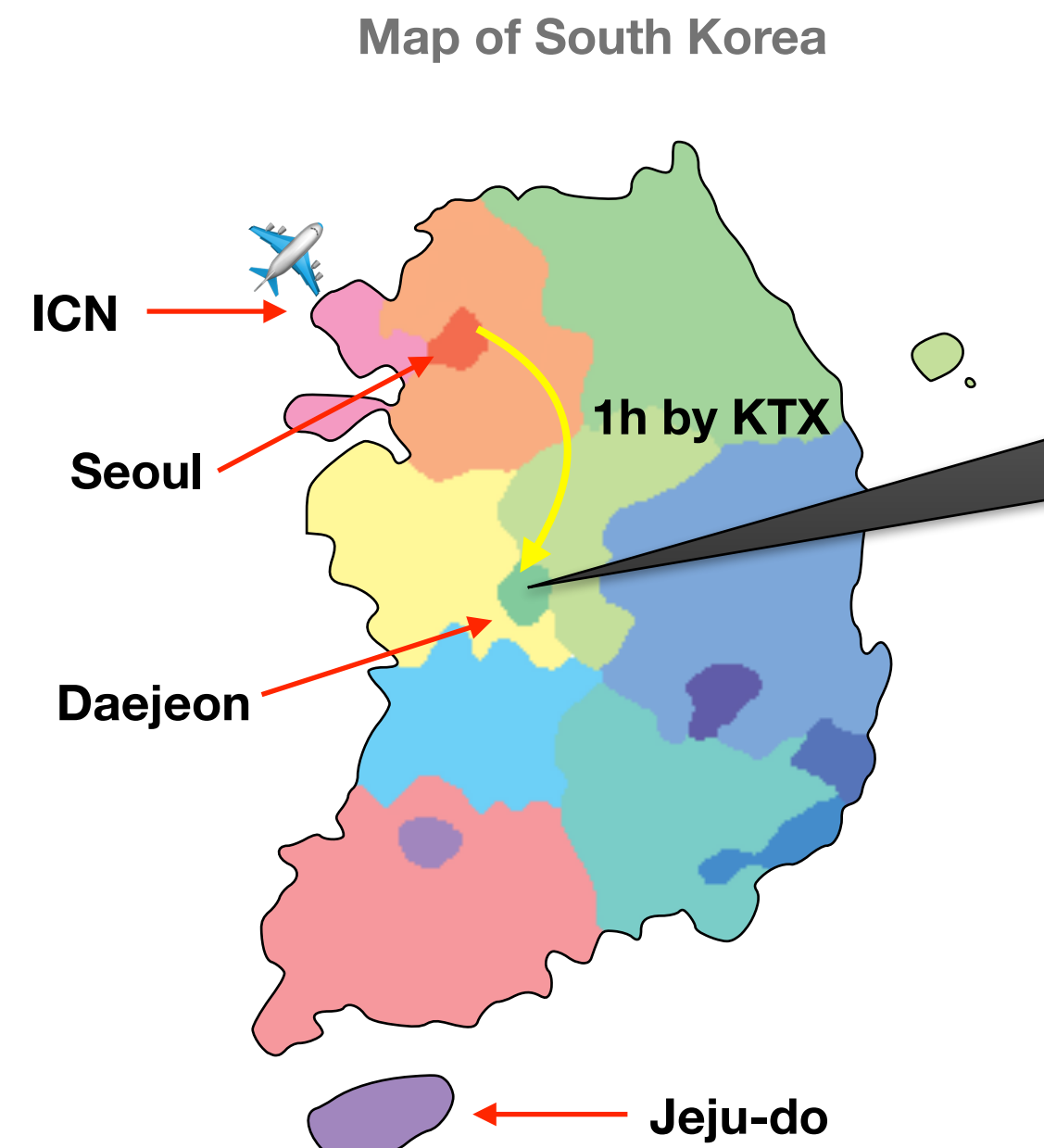
GSDC: Datacenter for Data-intensive Research

Ahn Sang-Un for KISTI-GSDC
“e-Science Activities in Korea” @ ISGC2021

KISTI

Korea Institute of Science and Technology Information

- Government-funded research institute founded in 1962 for National Information Services and Supercomputing
- National Supercomputing Center
 - **Nurion** - Cray CS500 system
 - 25.7 PFlops at peak, ranked 11th of Top500 (2018) => 21st (Nov 2020)
 - **Neuron** - GPU system, 1.24 PFlops
 - **KREONet/KREONet2** - National R&E network



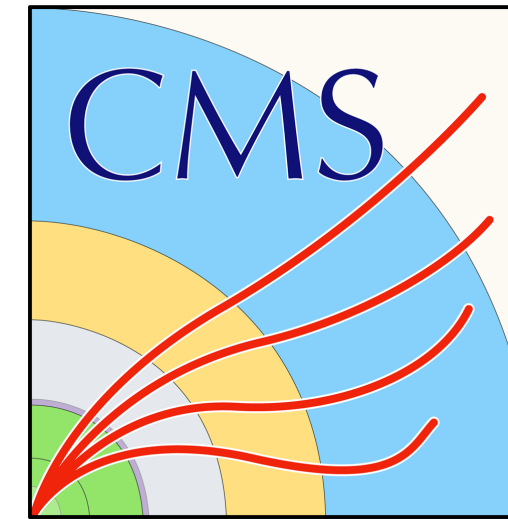
GSDC

Global Science experimental Data hub Center

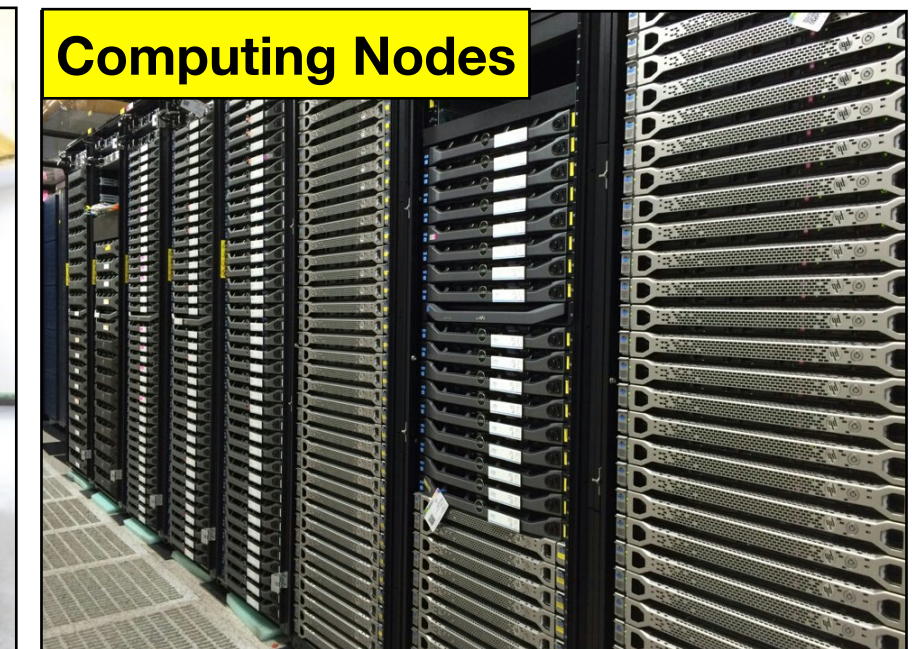
- Government-funded project, started in 2009 to promote Korean fundamental research through providing computing power and data storage
- Datacenter for data-intensive fundamental research
 - 17 staff: system administration, experiment support, external-relation, management and planning



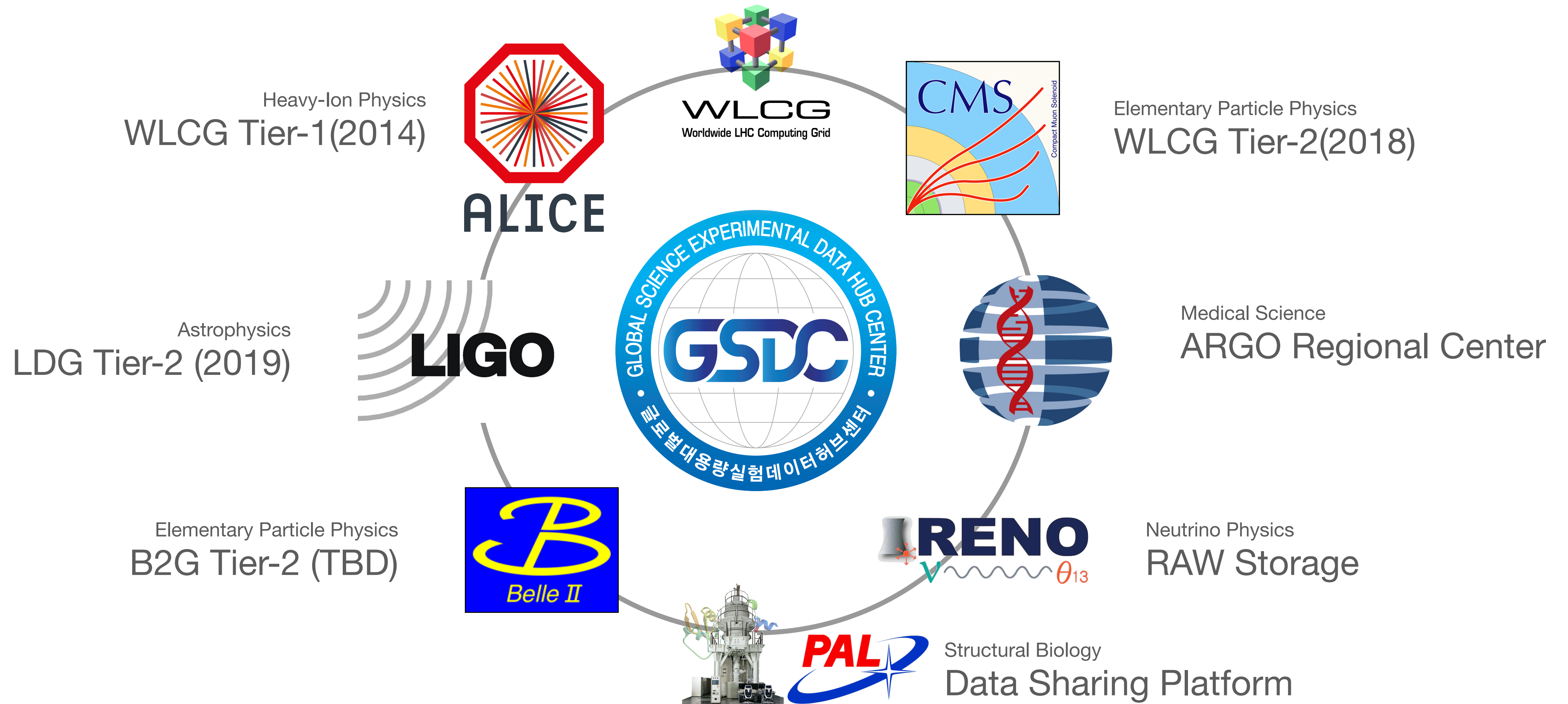
ALICE



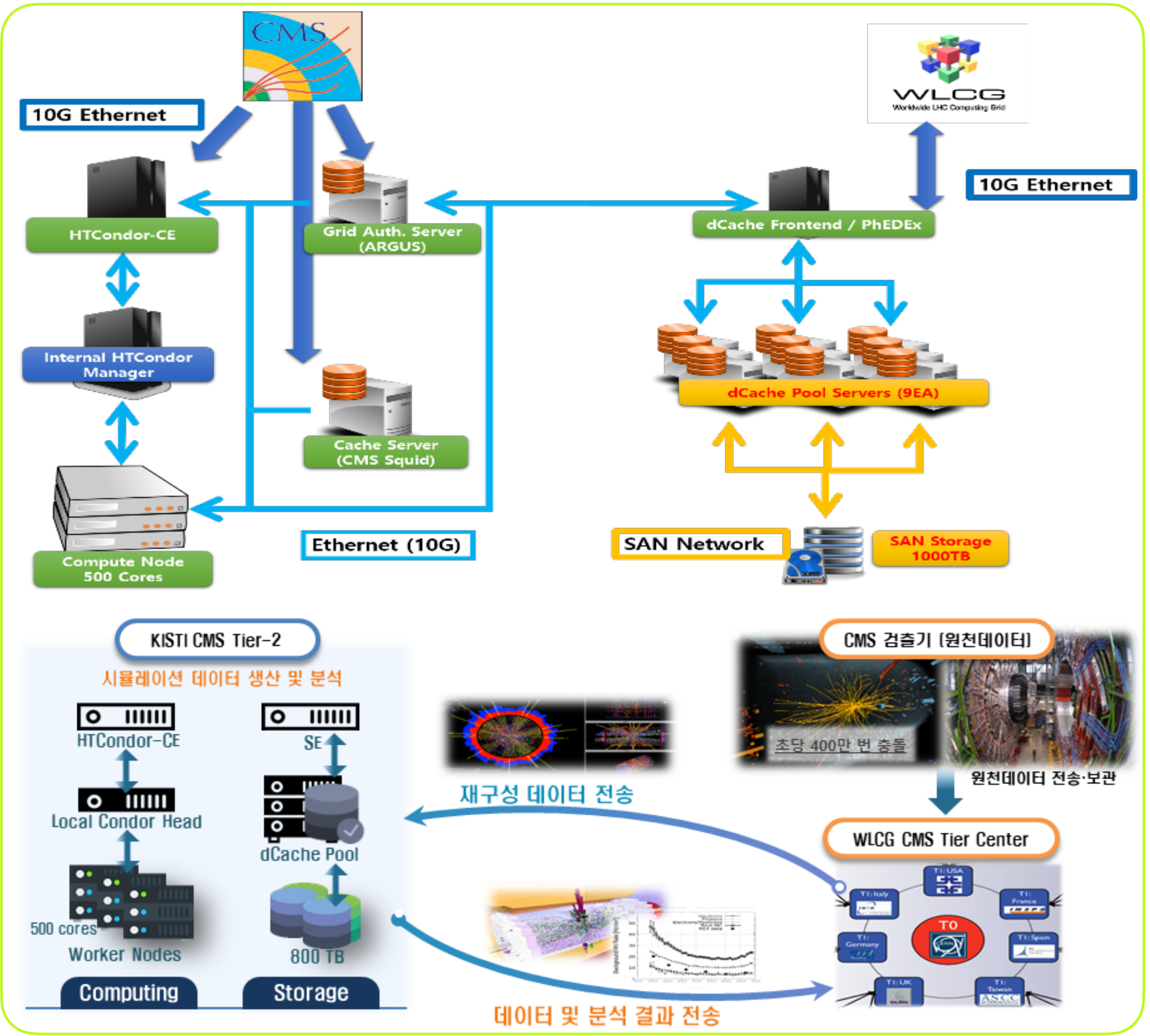
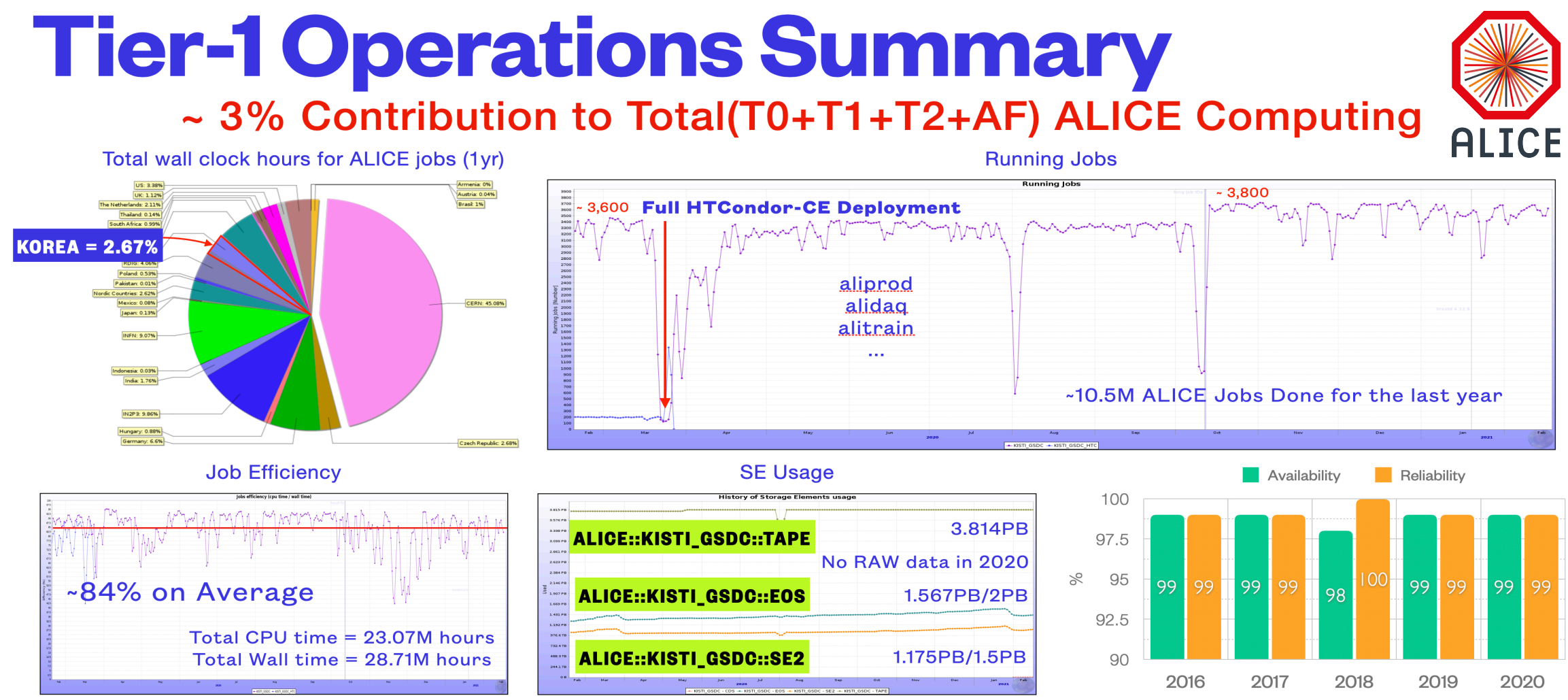
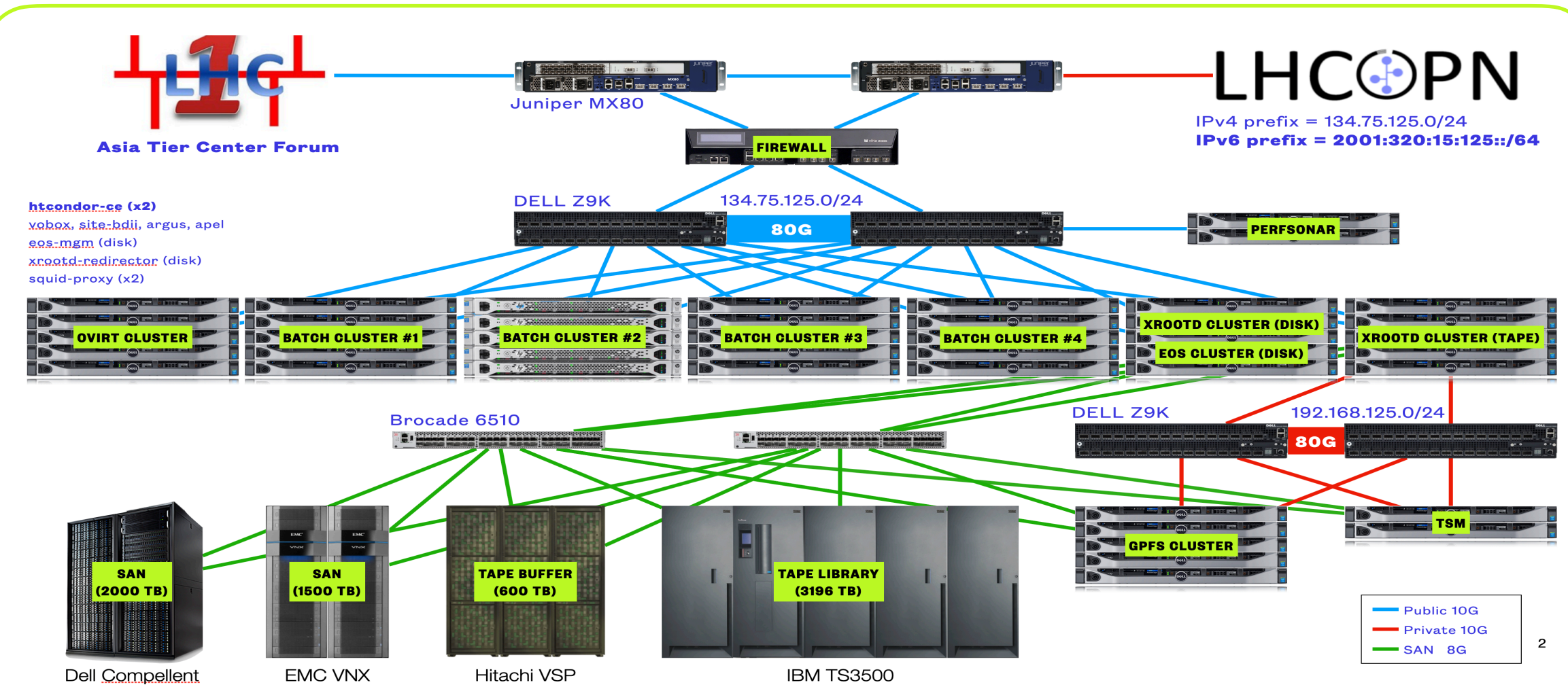
International
Cancer Genome
Consortium



Experiments Support



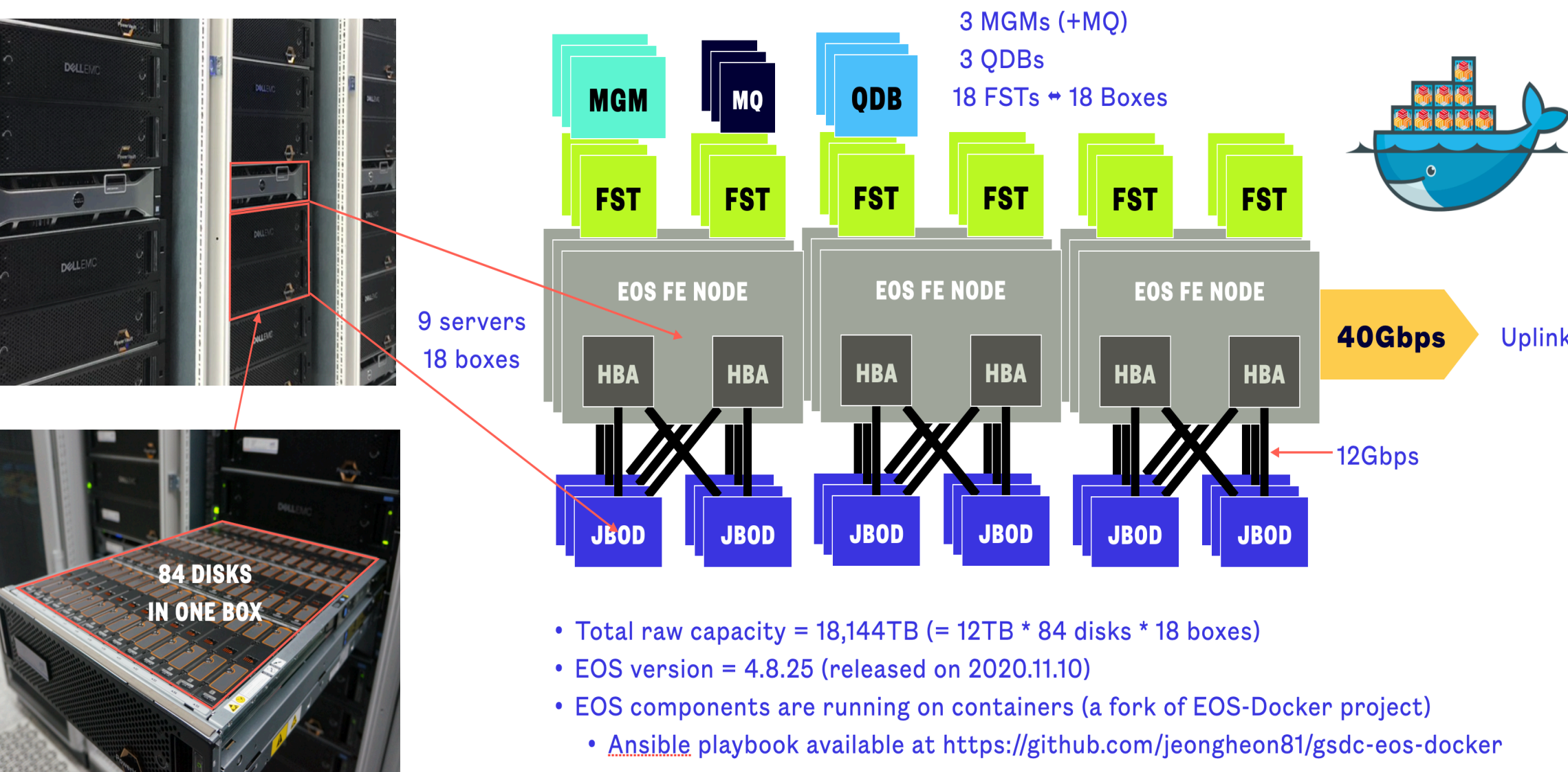
WLCG Tier-1/Tier-2



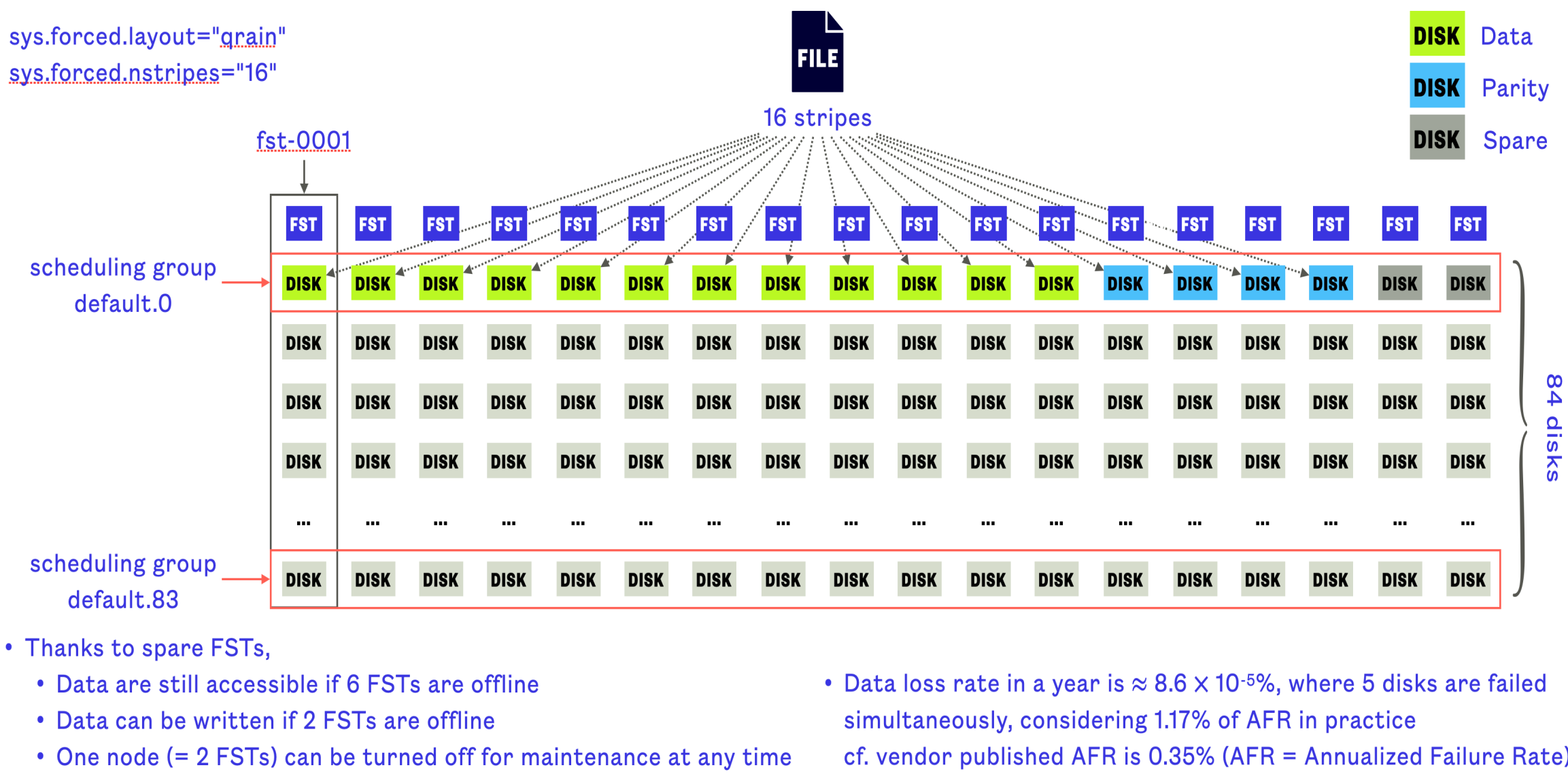
Custodial Disk Storage (Tapeless)

- The first disk-based custodial storage replacing tape for ALICE experiment
- 12 PB of usable space with 12+4 erasure coding for data protection (provided by EOS)

System Architecture

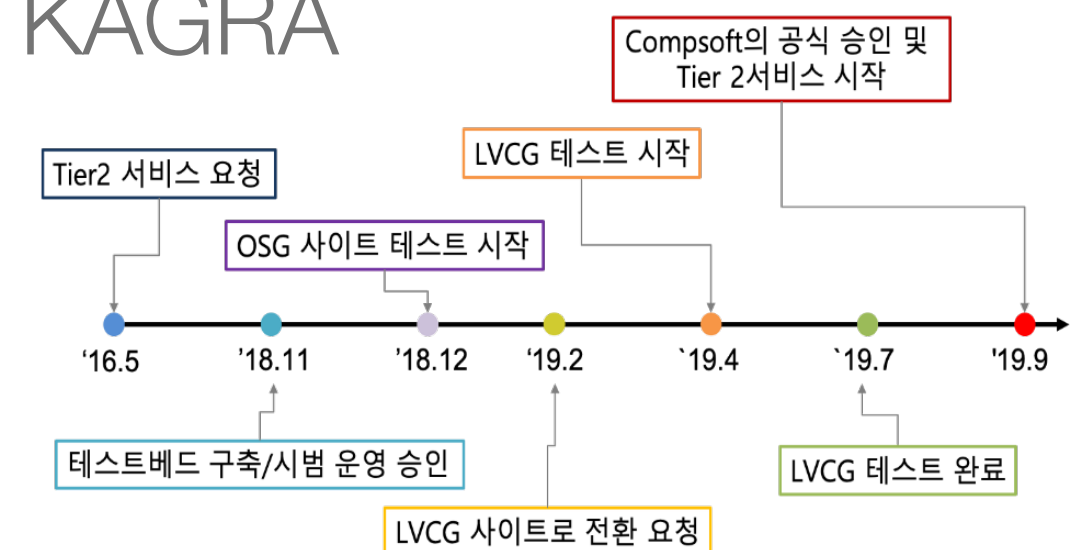


QRAIN(12+4) Layout



LIGO Tier-2

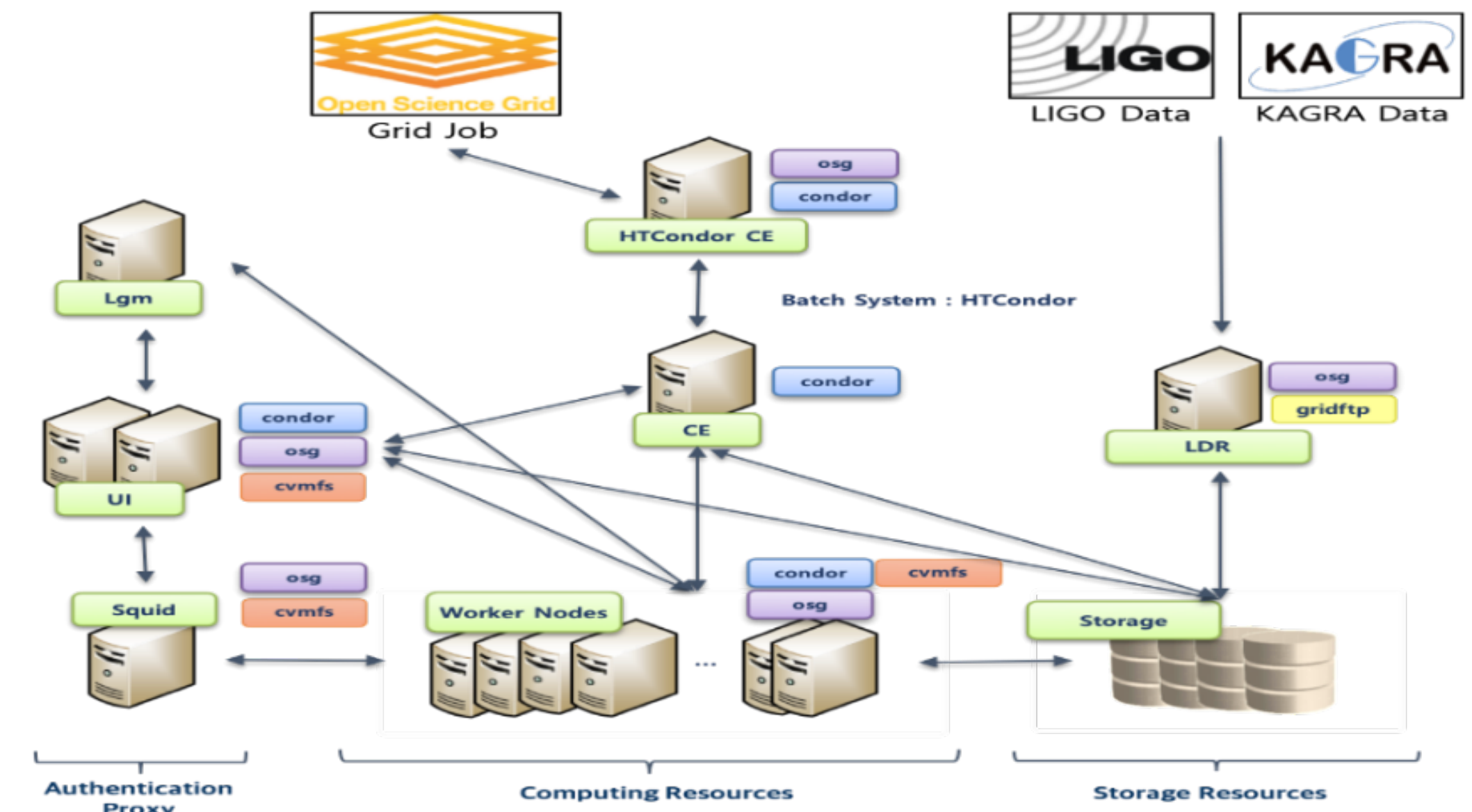
- Expanding LIGO Tier-3 service to the Global one, the Tier-2
 - Requested by Korean Gravitational Working Group that participating in LIGO and KAGRA
 - Integrated in Open Science Grid (OSG) for LDG services
- Test-bed started in 2018 and the Tier-2 in production since September 2019



LIGO LVC Computing Resources

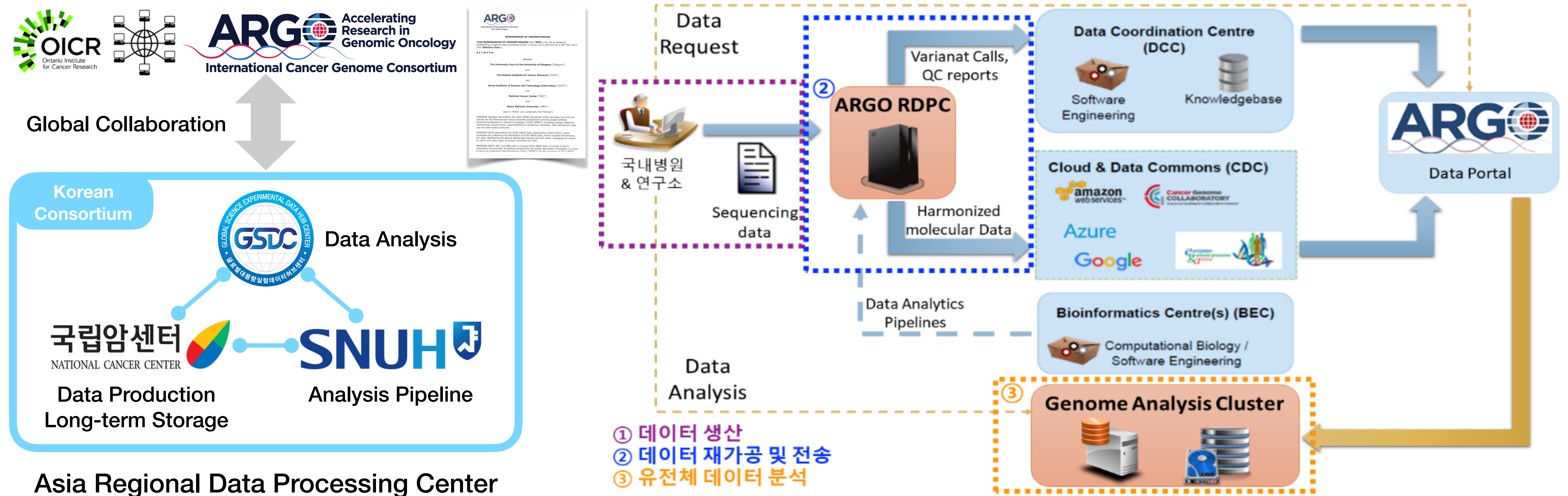
1. **Providing the entire LVC unified access** to dedicated and shared computing resources from internal and external computing centers

- OSG -> **LIGO-Virgo Computing Grid (LVCG)**
- Existing Resources:
 - LIGO Data Grid (CIT, LLO, LHO, Cardiff/ARCAA, GTech, UWM/Nemo)
 - Virgo computing centers (CNAF, IN2P3, Nikhef, Polgaw)
- New Resources
 - New KISTI Tier-2 Site!**
 - New Virgo computing centers (e.g., PIC)
 - XSEDE HPC/HTC (e.g., COMET, Stampede)
 - Louisiana State University (LSU) CPU and GPU clusters
 - Pacific Research Platform (PRP) GPU cluster
 - Other LVC campus/institutional clusters



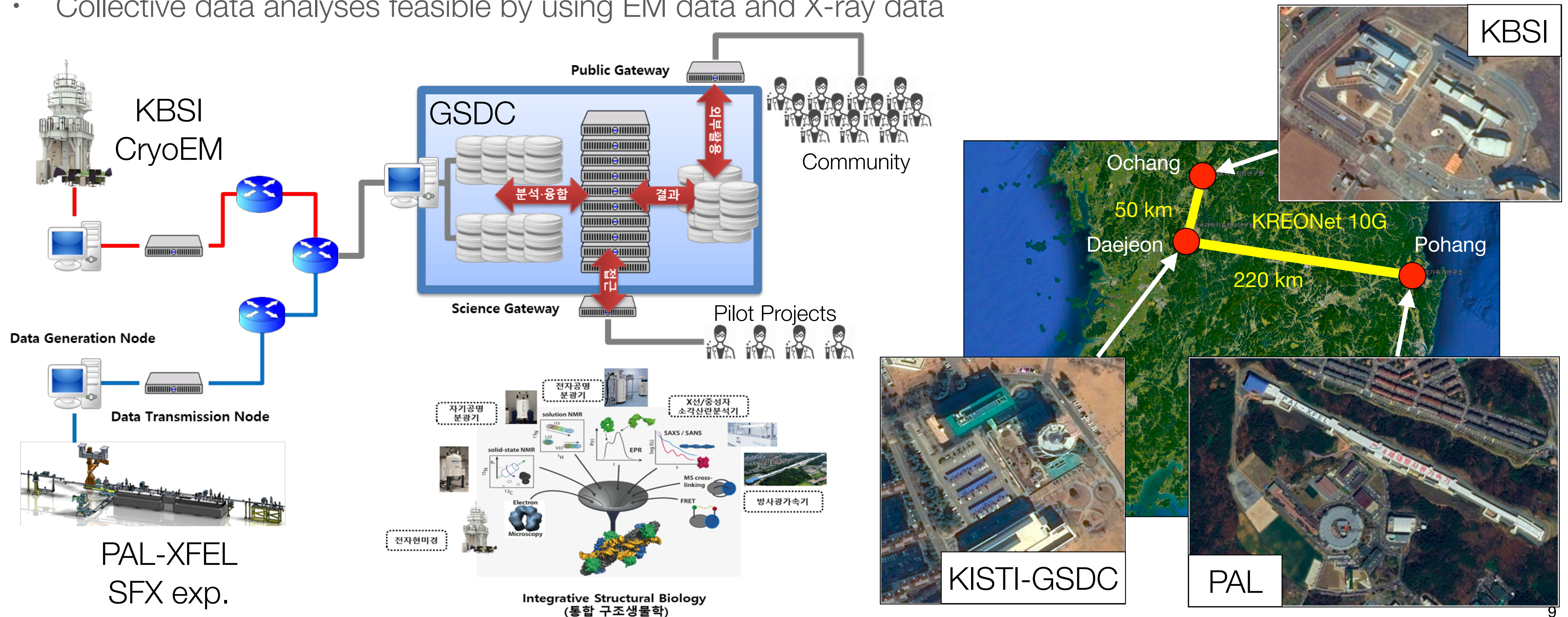
ICGC ARGO RDPC

- Hosting global cancer genome clinical data ~ 100k samples
- Collaboration with National Cancer Center and Seoul National University Hospital

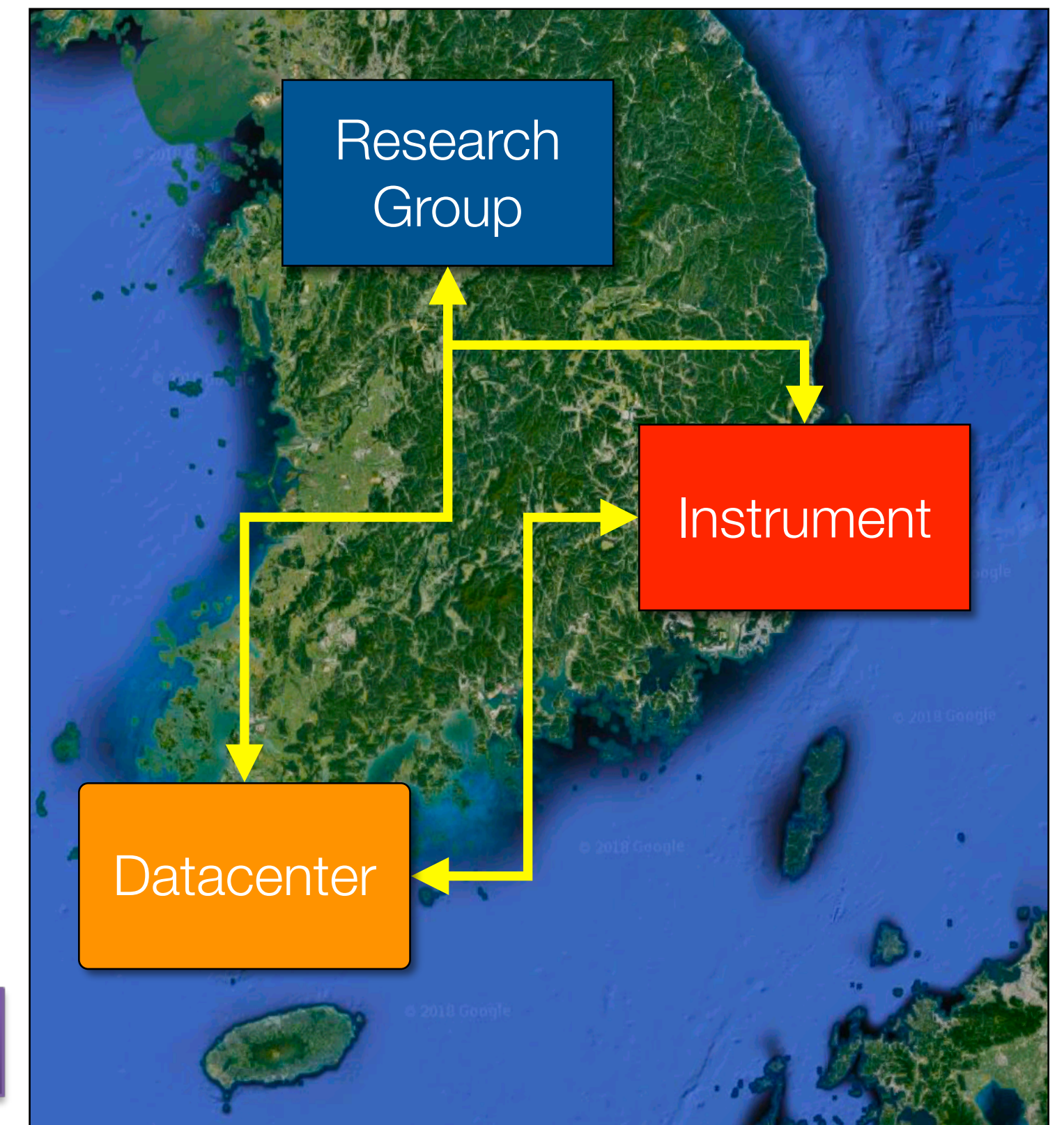
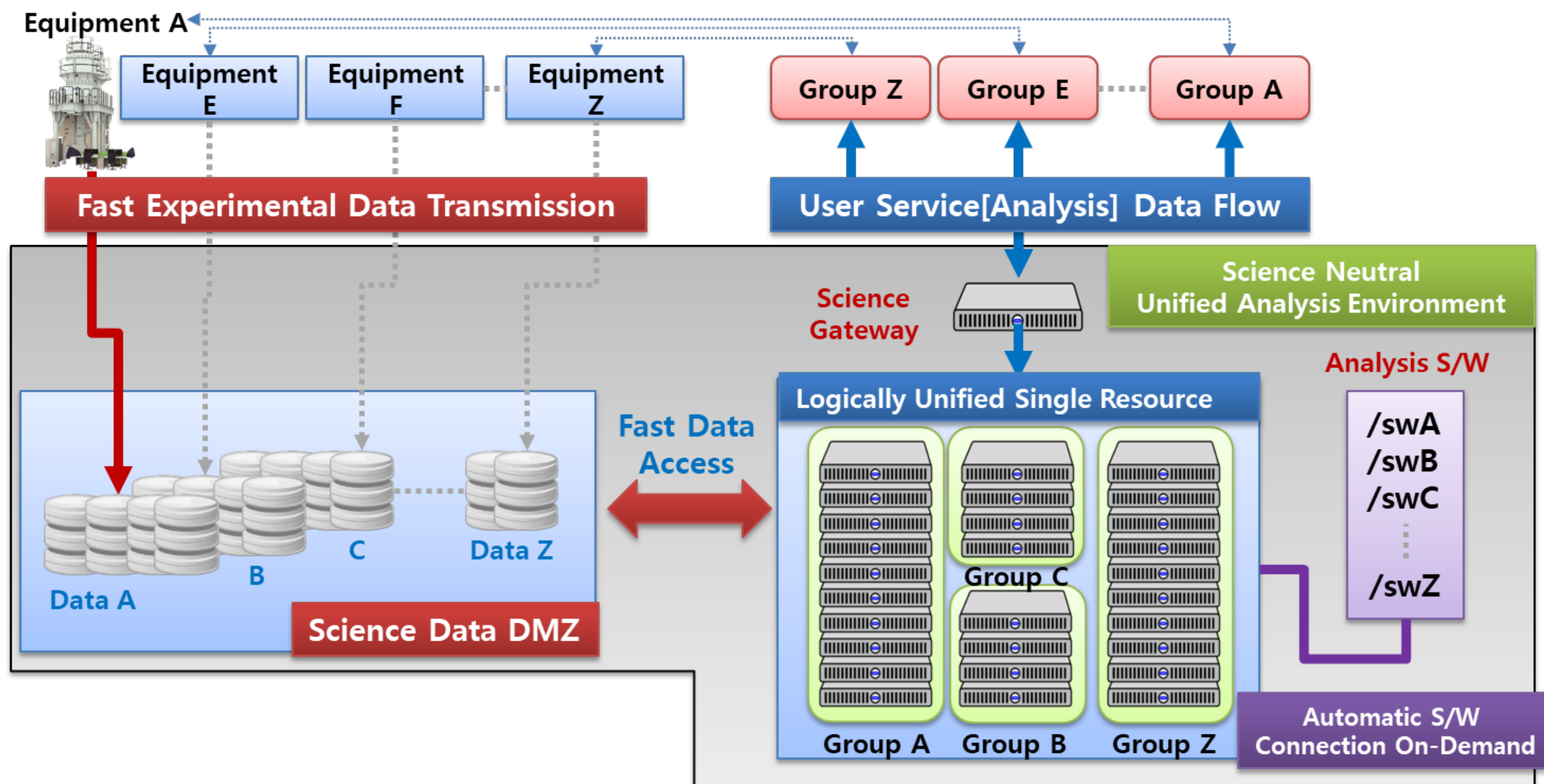


Pilot Project for Sharing Data from Heterogenous Instruments

- 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



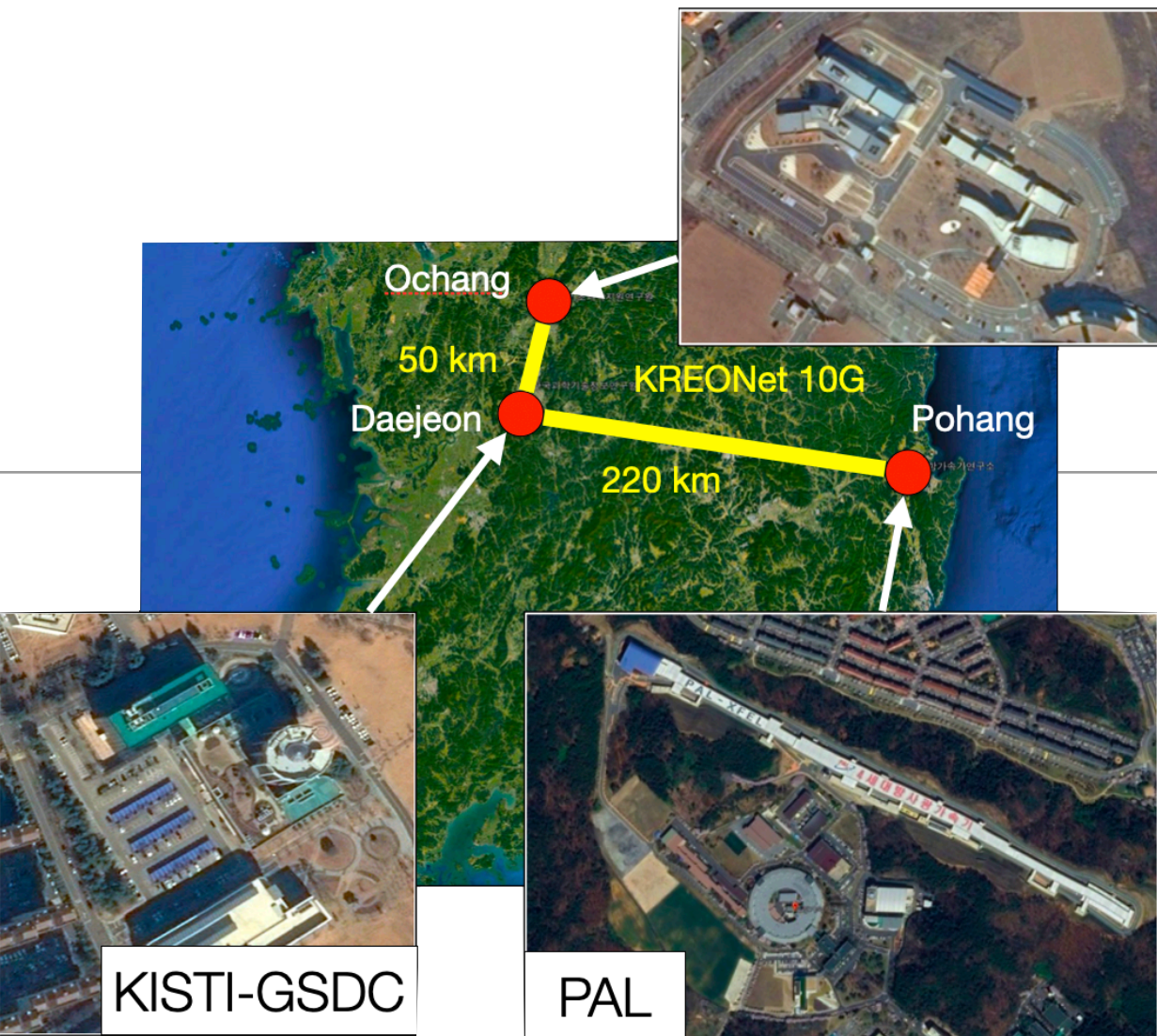
Unified Data Analysis Platform for Large-scale Facilities & Instruments



- 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

Big Data Express

- Collaboration with Fermilab (US)
- KISTI-GSDC developed BDE AmoebaNet
- Demonstrated intercontinental fast data transfer in SC
- Plan to implement BDE to support domestic experiments such as PAL-XFEL, KBSI-CryoEM



BigData Express – High Performance Data Transfer (I)

	mdtmFTP	FDT	GridFTP	BBCP
Large file data transfer (1 X 100G)	74.18	79.89	91.18	Poor performance
Folder data transfer (30 x 10G)	192.19	217	320.17	Poor performance
Folder data transfer (Linux 3.12.21)	10.51	-	1006.02	Poor performance

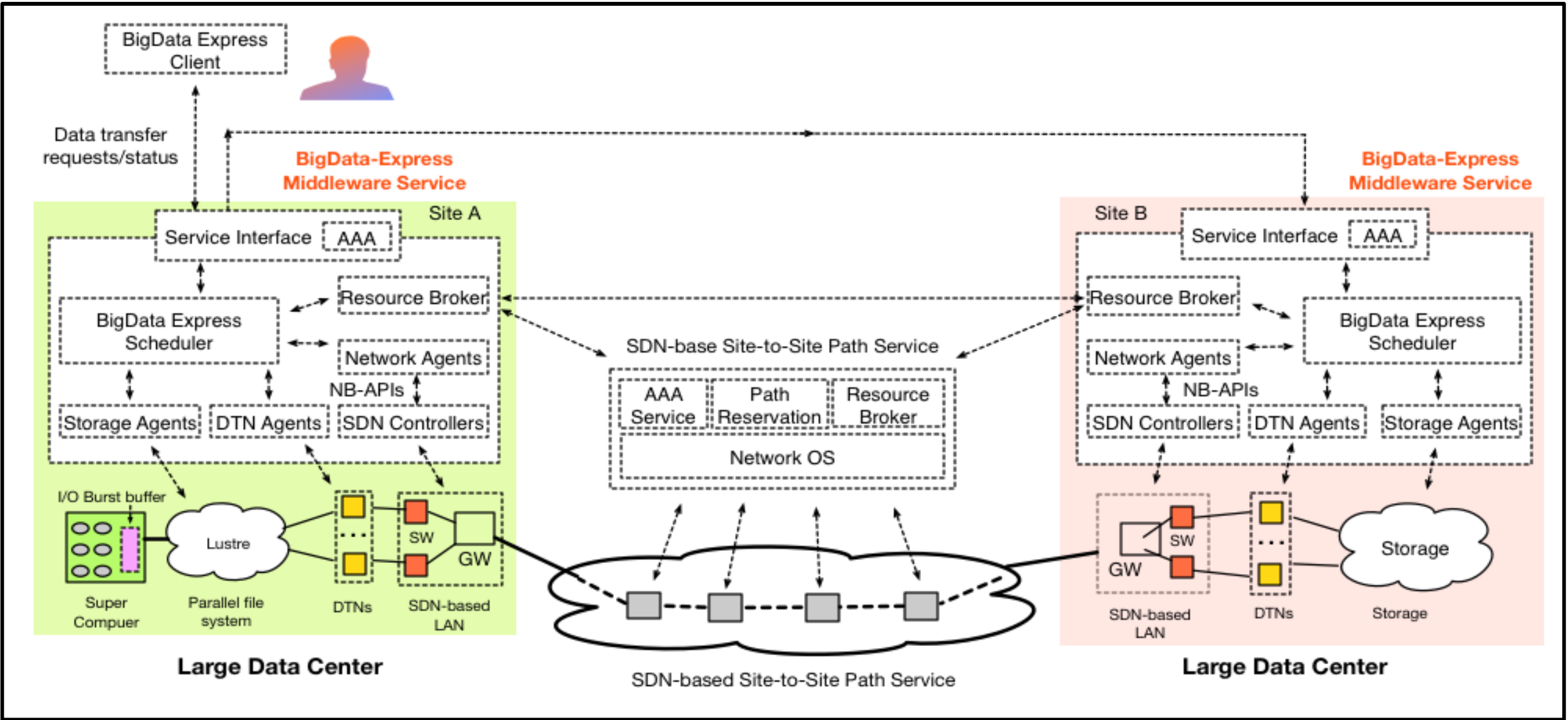
Time-to-completion (Seconds) – Client/Server mode Lower is better

	mdtmFTP	FDT	GridFTP	BBCP
Large file data transfer (1 X 100G)	34.976	N/A	106.84	N/A
Folder data transfer (30 x 10G)	95.61	N/A	-	N/A
Folder data transfer (Linux 3.12.21)	9.68	N/A	-	N/A

Time-to-completion (Seconds) – 3rd party mode Lower is better

Note 1: “-” indicates inability to get transfer to work
Note 2: BBCP performance is very poor, we do not list its results here
Note 3: BBCP and FDT support 3rd party data transfer. But BBCP and FDT couldn't run 3rd party data transfer on ESNET testbed due to testbed limitation

mdtmFTP is faster than existing data transfer tools, ranging from 8% to 9500%!
@ESnet 100GE SDN Testbed,



Summary

- GSDC is a dedicated datacenter at KISTI to promote fundamental research in South Korea by providing necessary computing power, storage and services
 - Tightly coupled with KISTI Supercomputer infrastructure and KISTI KREONet's reachability
 - Expanding its contribution to global research communities
 - More focus on improving R&D capability with advancing data sharing platform in line with the Government promotion