GSDC: Datacenter for Data-intensive Research
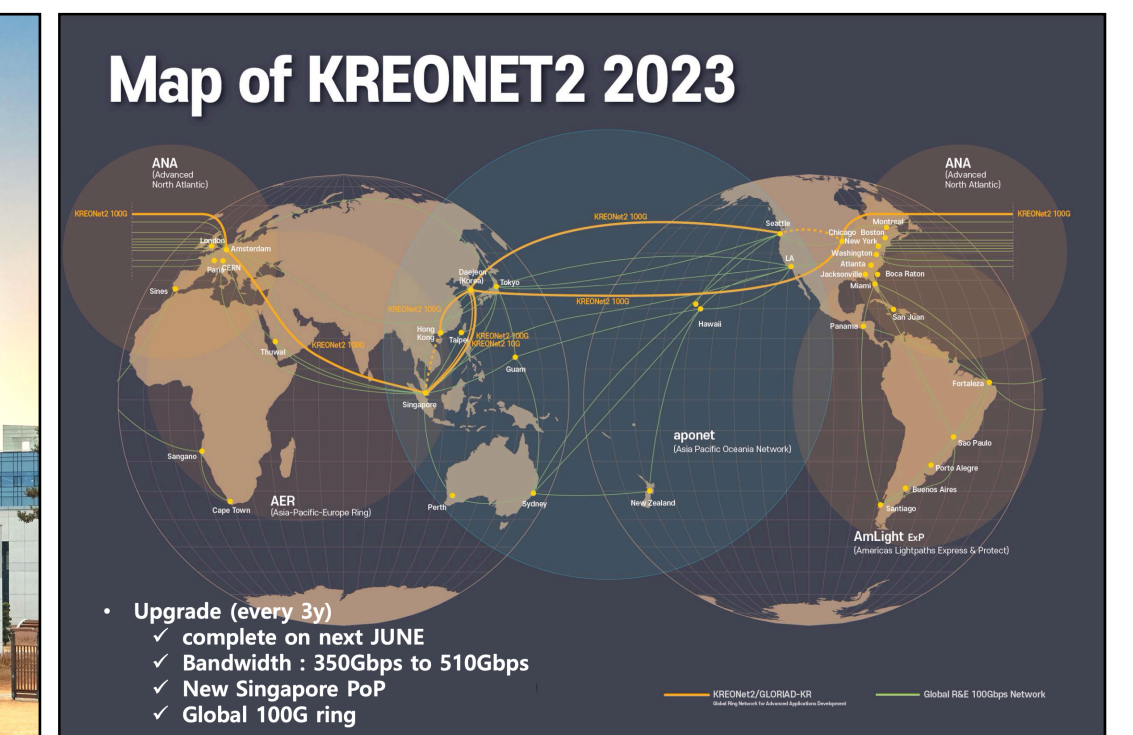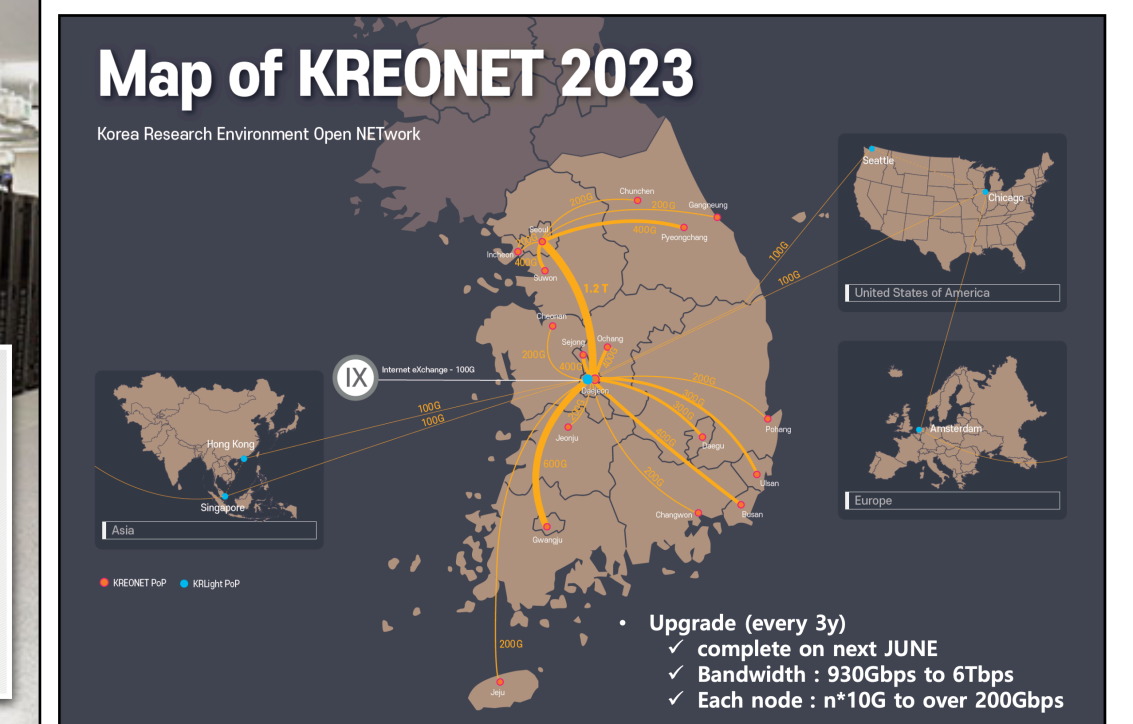
e-Science Activities in Korea @ ISGC 2023

Ahn Sang-Un

KISTI

Korea Institute of Science and Technology Information

- Government-funded research institute founded in 1961 for national information services and supercomputing

- Information Services: **ScienceON**, **NTIS**, etc.

- National Supercomputing Center

  - **Nurion** - Cray CS500 system

    - 25.7 PFlops at peak, ranked 11th of Top500 (2018) ⇨ 46th (Nov 2022)

  - **Neuron** - GPU system, 1.24 PFlops

  - **KREONet/KREONet2** - National/International R&E network

슈퍼컴퓨터 5호기

TOP 500 CERTIFICATE

Map of KREONET 2023
Korea Research Environment Open NETwork

- Upgrade (every 3y)
  - ✓ complete on next JUNE
  - ✓ Bandwidth : 930Gbps to 6Tbps
  - ✓ Each node : n*10G to over 200Gbps

Map of KREONET2 2023

- Upgrade (every 3y)
  - ✓ complete on next JUNE
  - ✓ Bandwidth : 350Gbps to 510Gbps
  - ✓ New Singapore PoP
  - ✓ Global 100G ring

GSDC

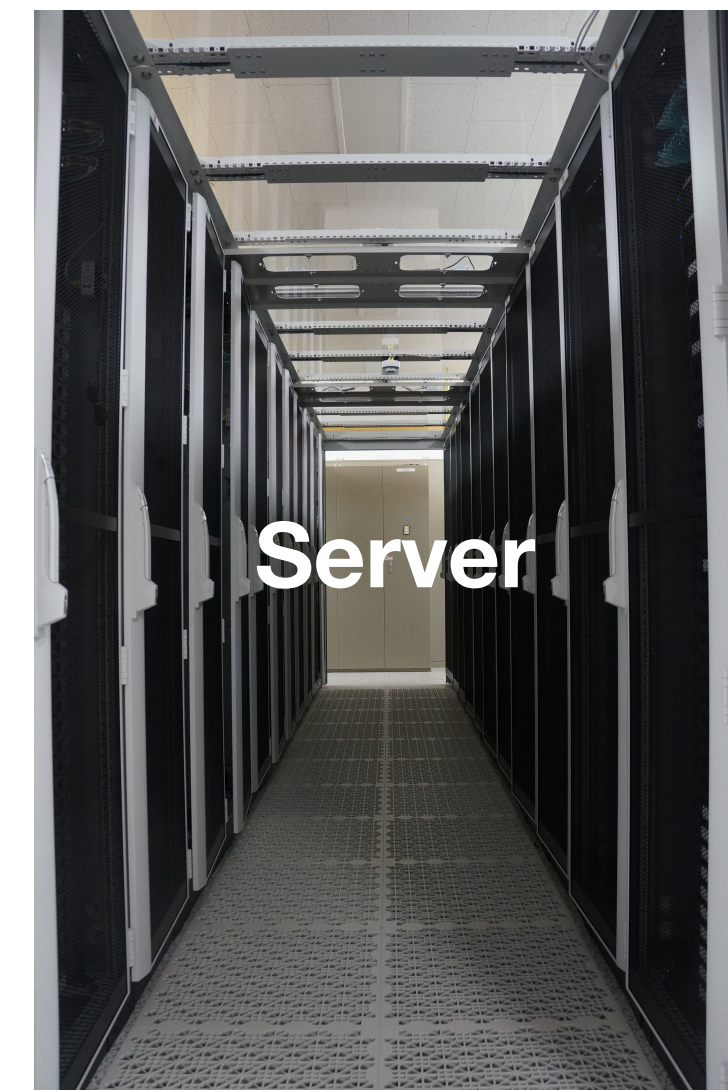## Global Science experimental Data hub Center

- Government-funded project, started in 2009 to promote Korean fundamental research through providing computing power and data storage

- **Datacenter for data-intensive fundamental research**

  - Preserving data from domestic or overseas large and complex scientific instruments as well as bio-medical and simulation-R&D activities

  - Providing services based on technology development: distributed computing structure, high availability storage system, infra integrated management, disk-based custodial storage
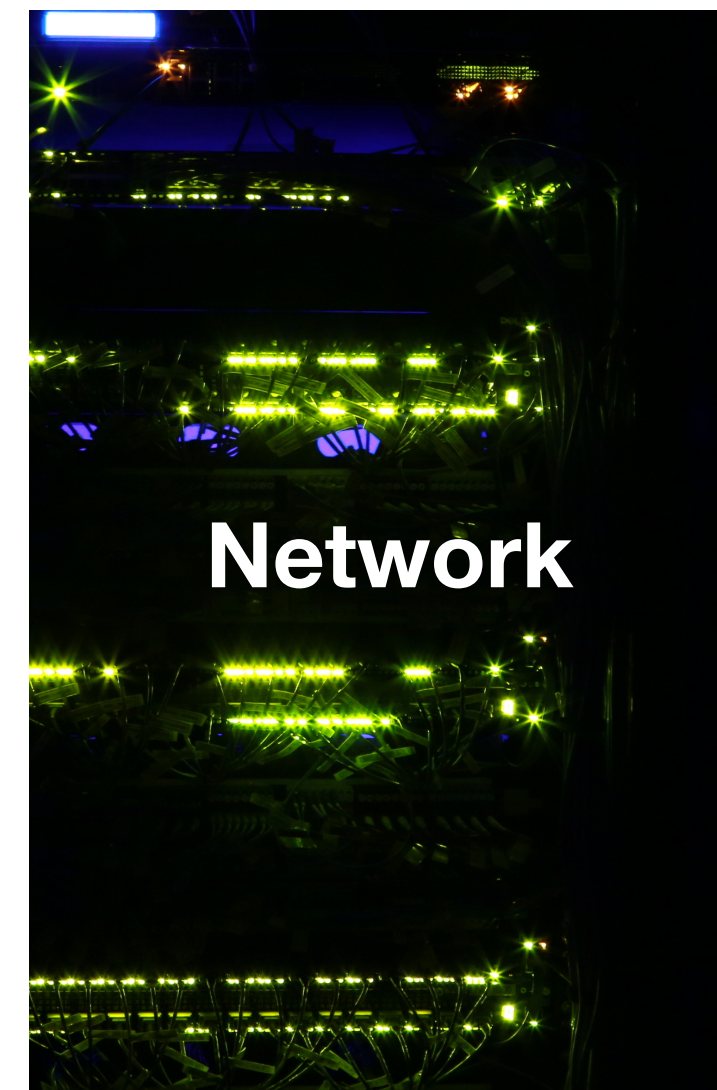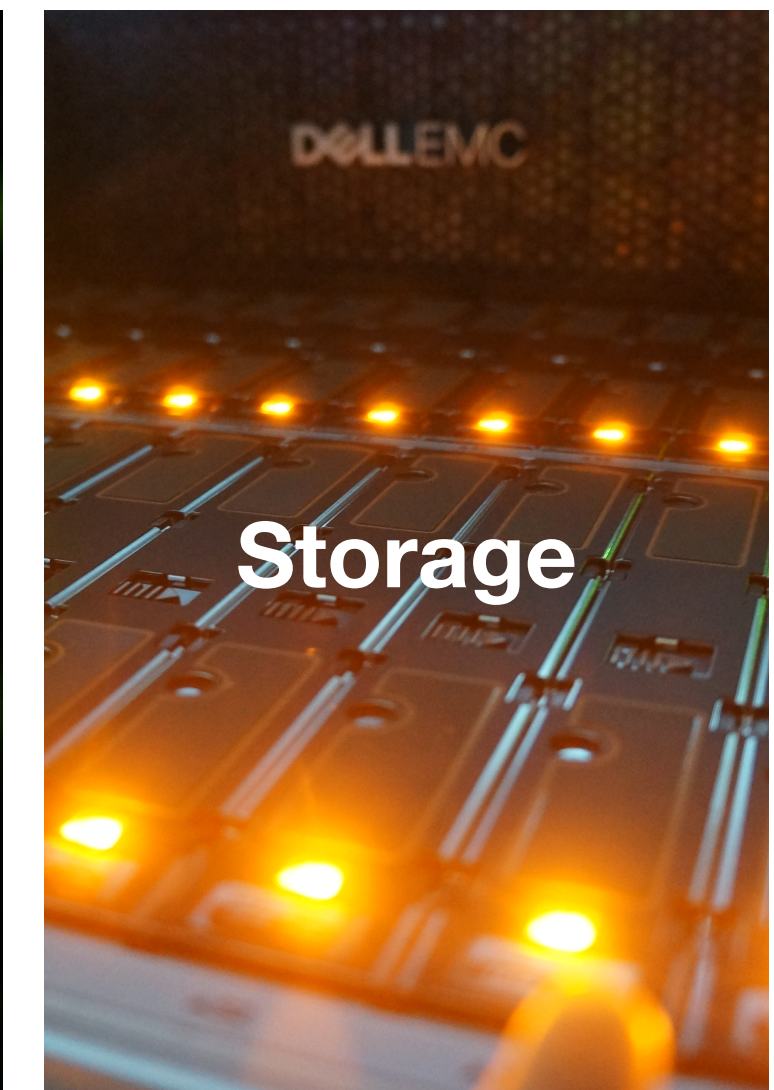
Server **Network** **Storage**

Supporting Experiments

Heavy-Ion Physics
WLCG Tier-1 (2014)

WLCG
Worldwide LHC Computing Grid

ALICE

Elementary Particle Physics
WLCG Tier-2 (2018)

CMS
Compact Muon Solenoid

Astrophysics
LDG Tier-2 (2019)

LIGO

GLOBAL SCIENCE EXPERIMENTAL DATA HUB CENTER
글로벌대용량실험데이터허브센터
GSDC

Belle II
Elementary Particle Physics
B2G Tier-2 (TBD)

RENO
ν θ_{13}
Neutrino Physics
RAW Storage

Medical Science
ICGC-ARGO Center

PAL

Structural Biology using Cryo-EM/XFEL
Data Storage & Analysis Pipeline
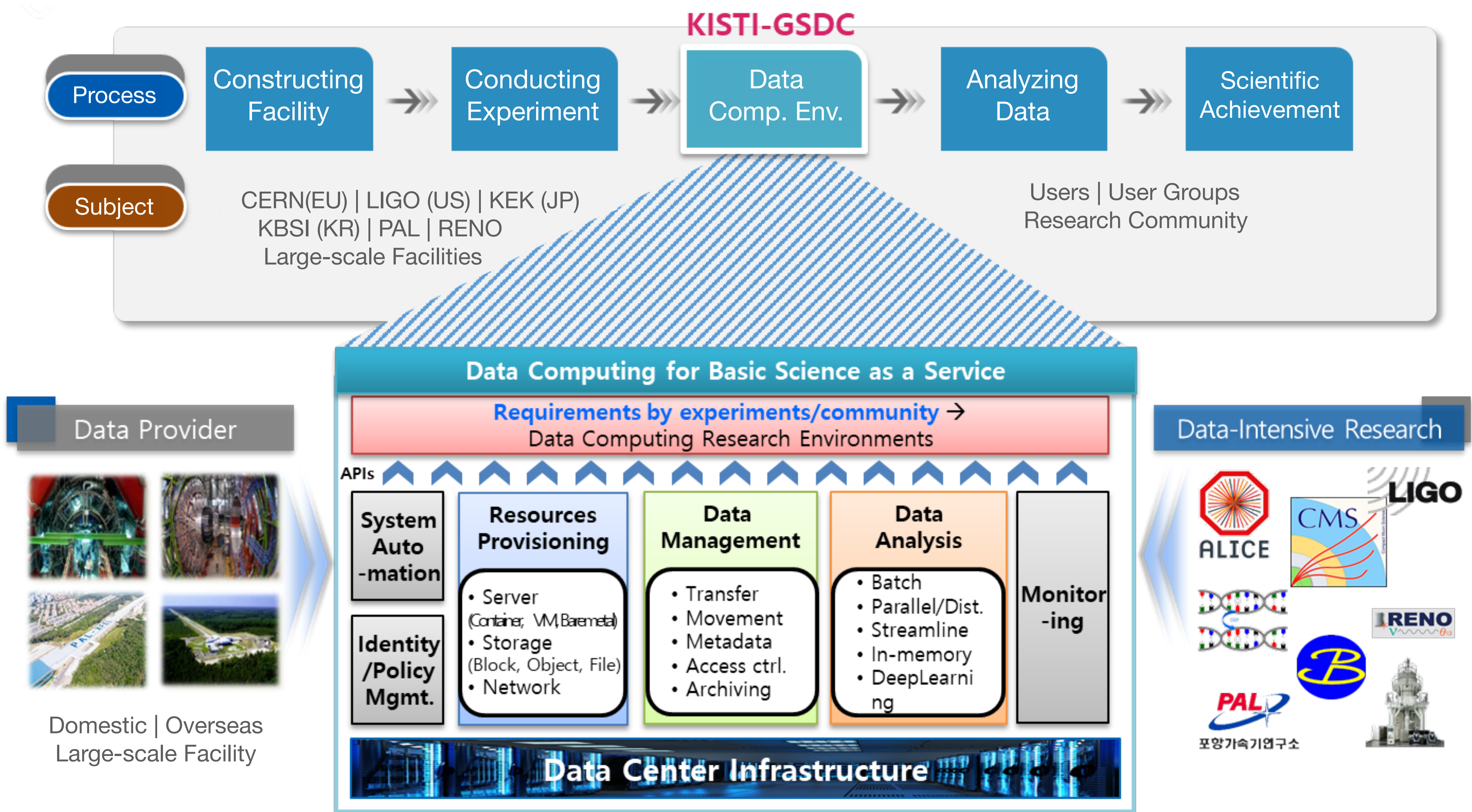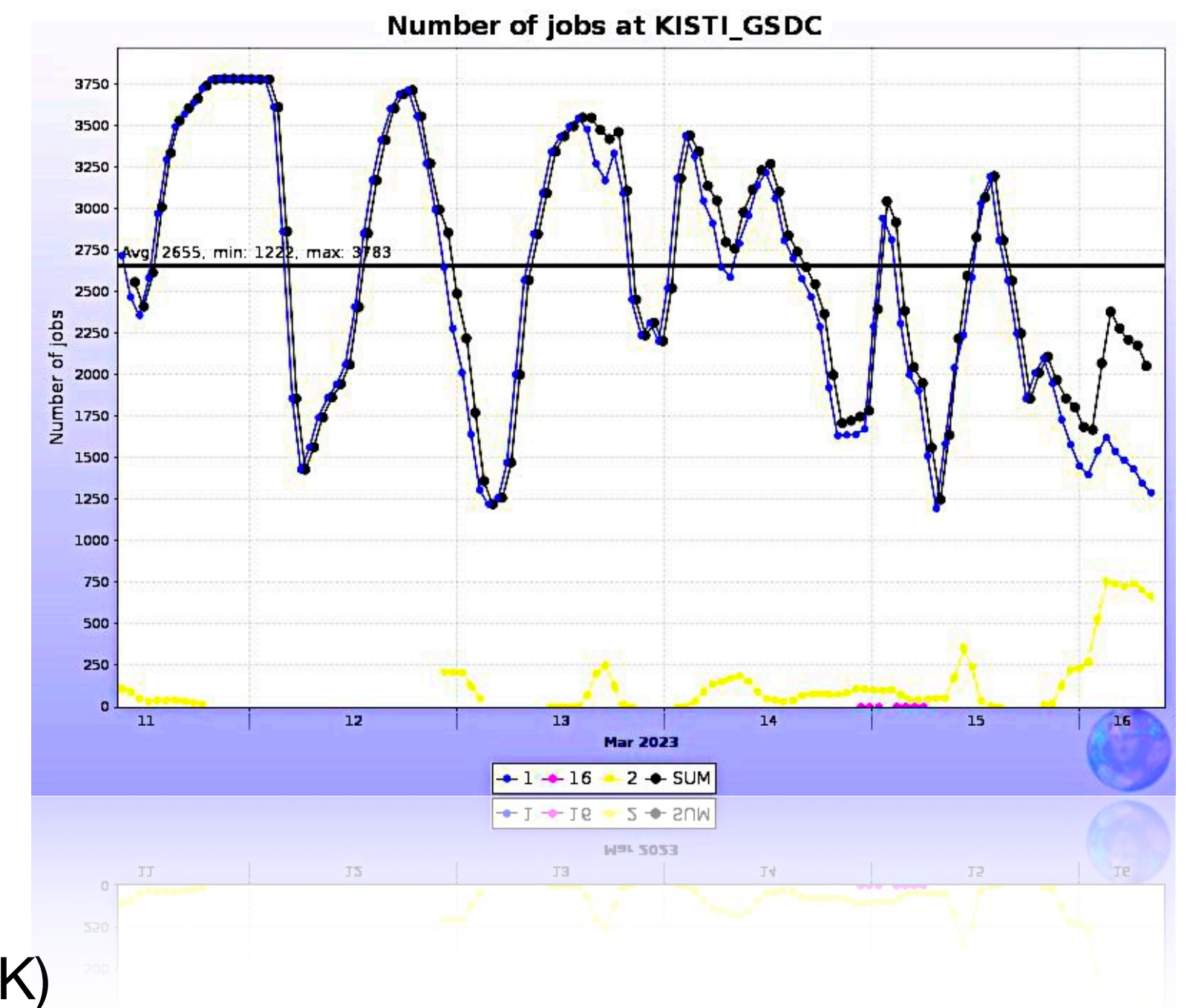
Role of GSDC for Data-intensive Research

WLCG Tier-1 @ KISTI-GSDC
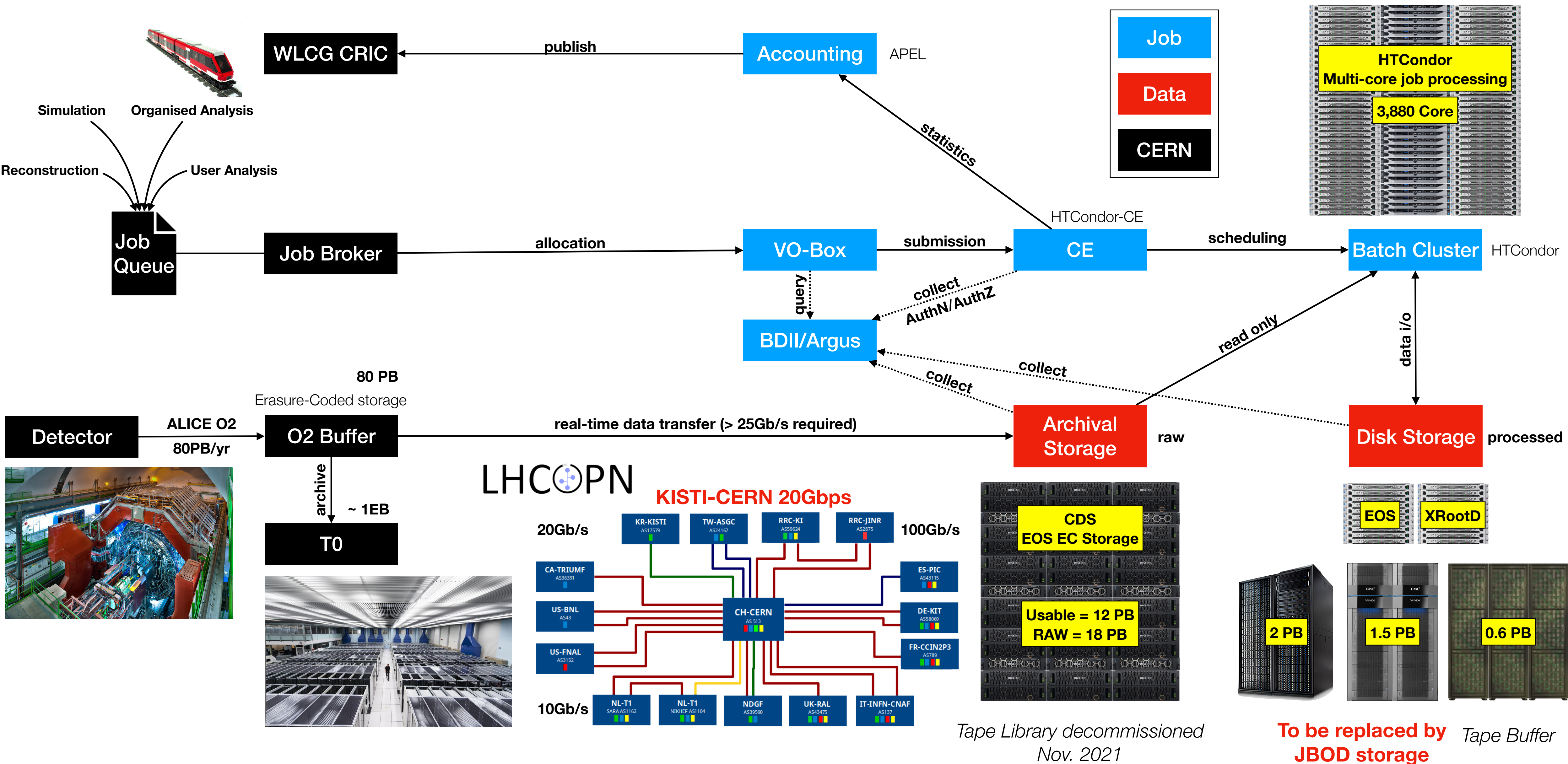## Flagship Service for Data-intensive Computing

- The only WLCG Tier-1 in Asia for the ALICE experiment

  - Small (compact) but contributing about 10% of T1 resource requirements of ALICE

    - More than 2% of total (T0+T1+T2+AFs) resource requirements of ALICE

- CE

  - HTCondor-based, whole-node submission enabled (for N-core jobs)

- SE

  - XRootD/EOS based disk storage

  - Archival SE : CDS, the disk-based one powered by EOS

- Networking

  - LHCOPN : 20G dedicated link between Daejeon (KR) and Geneva (CH)

 - LHCONE : 100G provisioned by KREONet connecting to EU, US and Asia (SG/HK)

Number of jobs at KISTI_GSDC

KISTI ALICE T1 Structure Overview

To be doubled with new clusters

Job (blue)
Data (red)
CERN (black)

HTCondor Multi-core job processing
3,880 Core

WLCG CRIC ← **publish** ← Accounting — APEL

Simulation, Organised Analysis, Reconstruction, User Analysis → Job Queue

Job Queue → Job Broker → **allocation** → VO-Box → **submission** → CE → **scheduling** → Batch Cluster — HTCondor

Accounting — **statistics**

HTCondor-CE

VO-Box — **query** → BDII/Argus

BDII/Argus — **collect AuthN/AuthZ** → CE

80 PB
Erasure-Coded storage

Detector → **ALICE O2 / 80PB/yr** → O2 Buffer → **real-time data transfer (> 25Gb/s required)** → Archival Storage — **raw**

O2 Buffer → **archive** → **~ 1EB** → T0

BDII/Argus — **collect** → Archival Storage

BDII/Argus — **collect** → Disk Storage

Archival Storage — **read only** → Batch Cluster

Batch Cluster — **data i/o** → Disk Storage — processed

EOS, XRootD

LHCOPN

KISTI-CERN 20Gbps

20Gb/s ... 100Gb/s ... 10Gb/s

KR-KISTI AS17579, TW-ASGC AS24167, RRC-KI AS59624, RRC-JINR AS2875
CA-TRIUMF AS36391, ES-PIC AS43115
US-BNL AS43, CH-CERN AS 513, DE-KIT AS58069
US-FNAL AS3152, FR-CCIN2P3 AS789
NL-T1 SARA AS1162, NL-T1 NIKHEF AS1104, NDGF AS39590, UK-RAL AS43475, IT-INFN-CNAF AS137

CDS EOS EC Storage
Usable = 12 PB
RAW = 18 PB

Tape Library decommissioned Nov. 2021

2 PB ... **1.5 PB** ... **0.6 PB**

To be replaced by JBOD storage   *Tape Buffer*

7

CDS in one slide

Custodial Disk Storage
Tapeless Archiving

EC Layout using 4 parity nodes

- The first disk-based custodial storage replaced tape for ALICE experiment
- 12 PB usable space with 12+4 erasure coding for data protection (powered by CERN EOS)
- Fully automated deployment of EOS components using Linux containers
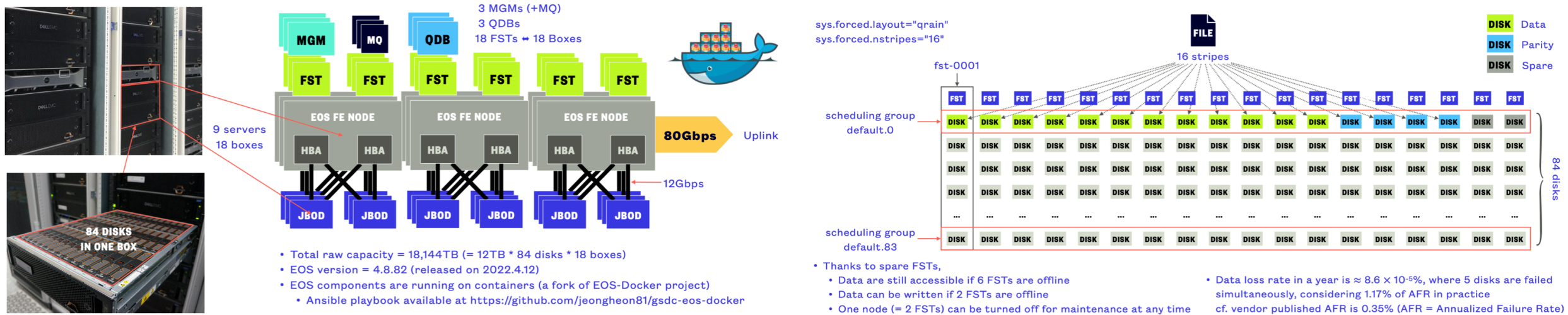
System Architecture

QRAIN(12+4) Layout
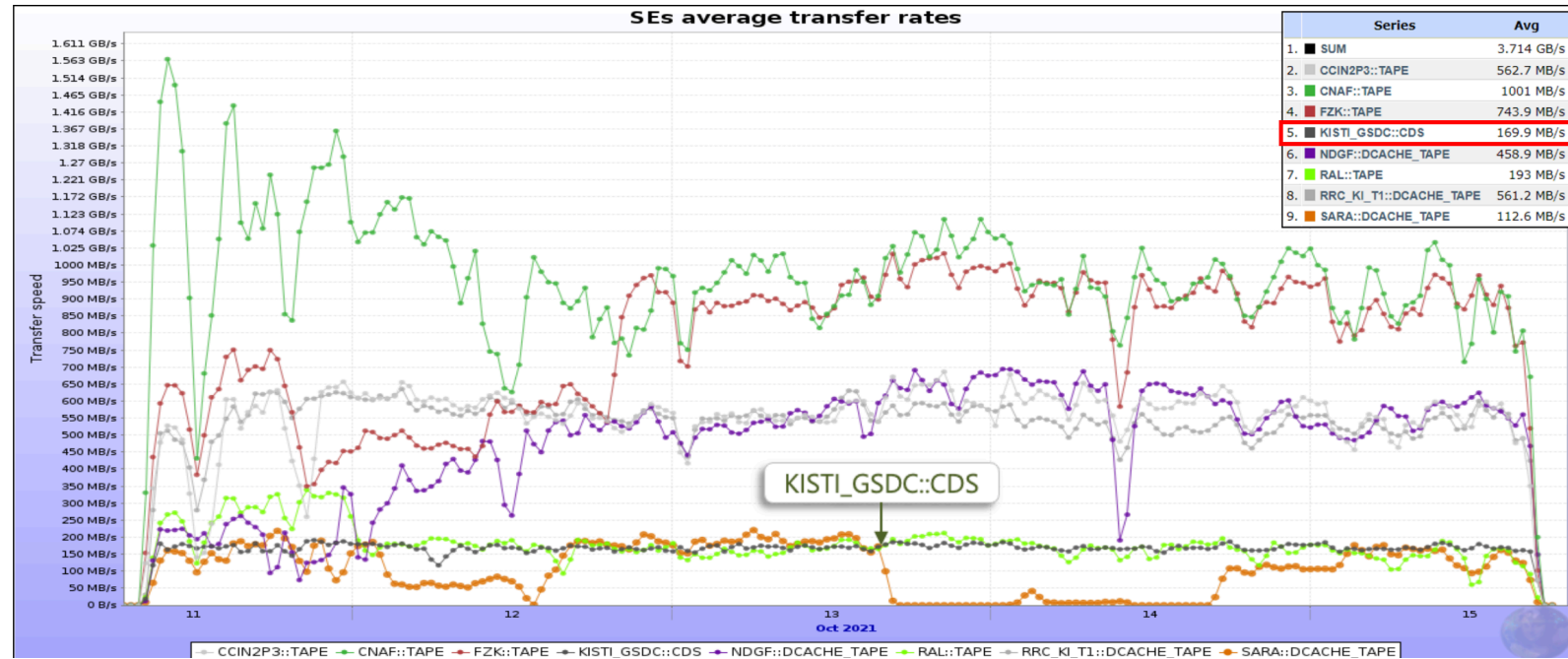
WLCG Data Challenges (Oct 2021)
Preparation for LHC RUN3 raw data transfer

- Participation as a Tape (custodial storage) for the ALICE experiment
- Joined efforts of the WLCG Collaboration preparing for LHC RUN3 data taking
- Successful to meet the target (stable) transfer performance (150MB/s)

170MB/s on average for 5-day of transfer
101.4TB of data (51k files) transferred

Individual files 1.953GB, total transferred 1.766PB

Centre	Files	size
CCIN2P3	143230	279.7TB
CNAF	239913	468.6TB
GridKA	187327	368.9TB
KISTI	51914	101.4TB
RAL	45023	87.9TB
NDGF	100635	196.5TB
RRC_KI	110479	216.8TB
SARA	23566	46TB

CDS Operation for ALICE
Fully commissioned since Nov 2021

Significant but endurable EC induced traffic observed

Current snapshot of the CDS in the ALICE monitoring system

http://alimonitor.cern.ch/stats?page=SE/table

	Total	Used
Bin	15.79	6.89
Dec	17.77	7.76

ALICE RAW data replication to the CDS
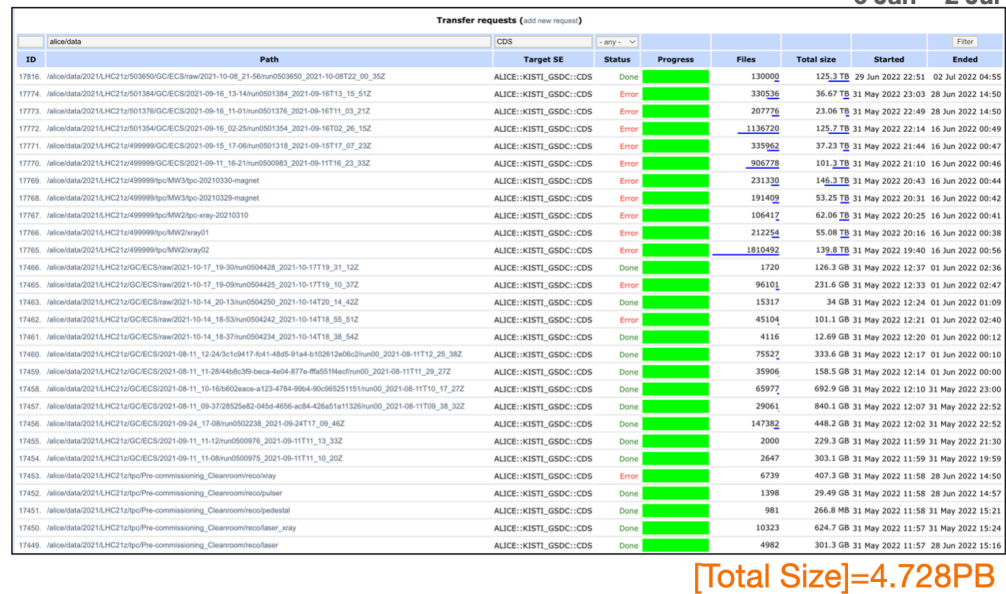
6 Jan ~ 2 Jul

Peak traffic IN + OUT = 4.172GB/s + 3.218GB/s
= 7.39GB/s ≈ 60Gb/s

IN: transfer + re-distribution = 14.3PB

Δ=5PB

OUT: re-distribution = 9.294PB
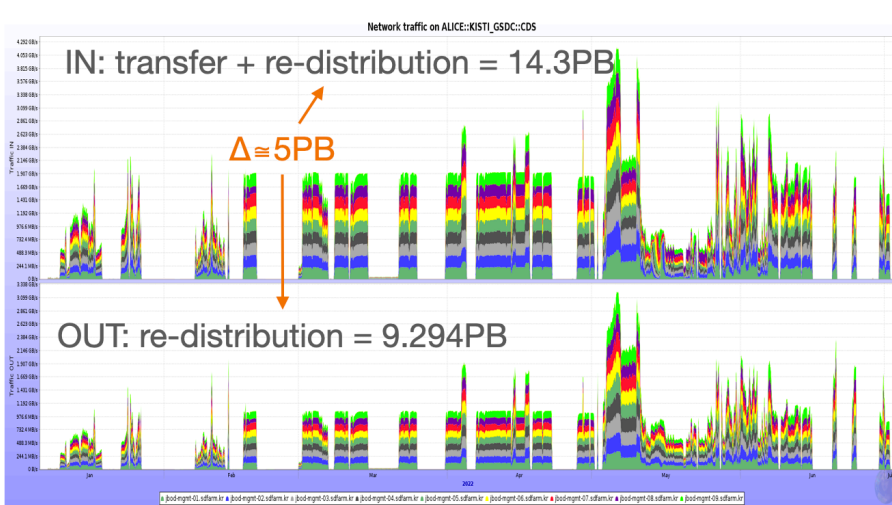
Re-distribution Traffic induced by EC

Average transfer rate = 328MB/s

6 Jan ~ 2 Jul

7.5Gb/s

LHCOPN - KREONet2

6 Jan ~ 2 Jul

[Total Size]=4.728PB

Power Consumption Measurement
Implication of cost that is not uncomfortably higher than tape

Instantaneous power consumption (kilowatt) per minute (Jan 2021 - Feb 2022)

16.49kW

Gradually increasing of the mean power consumption over time

17.02kW

8/1/21   6/3/22

10 - 19 Feb
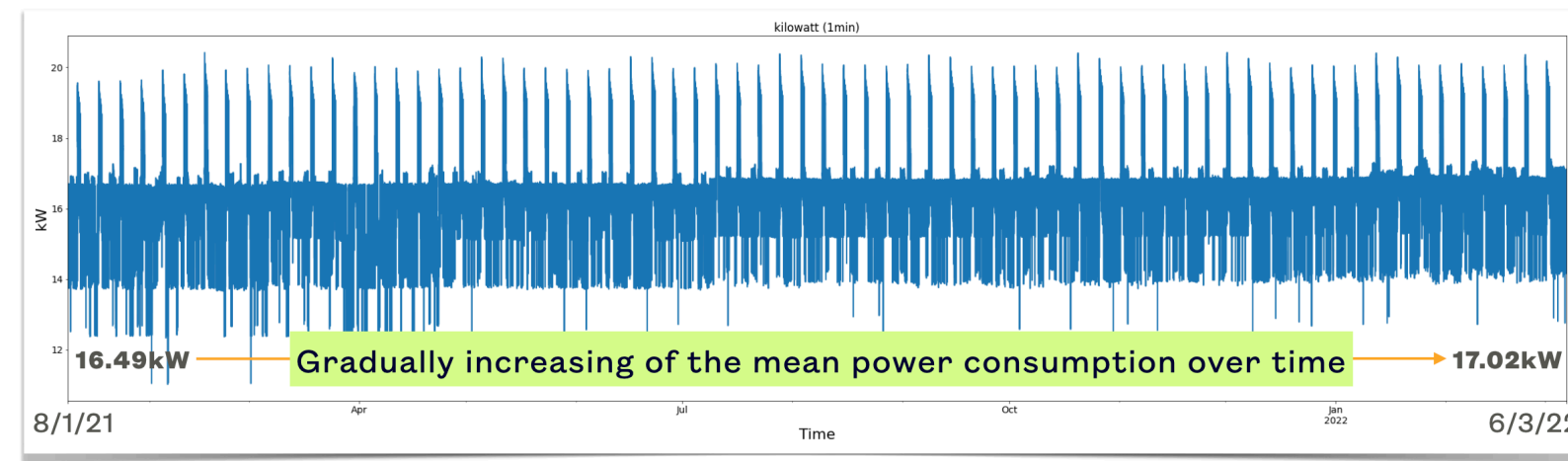
Comparison with other storage at KISTI
1.125W/TB for full load (cf. 0.5W/TB for Tape)

	Capacity (TB)	Max		Min		Mean	
		kW	W/TB	kW	W/TB	kW	W/TB
CDS	18,144	20.426	**1.125**	11.015	**0.607**	16.85	**0.923**
TS3500	3,200	1.6	**0.5**	-	-	-	-
SC7020	2,500	12.120	4.8	-	-	-	-
Isilon	2,950	13.730	4.6	-	-	-	-
Isilon	2,360	12.88	9	-	-	-	-
VNX	2,000	5.1	2.2	-	-	-	-
VSP	1,430	18.3	9.15	-	-	-	-
CX4-960	1,500	14.9	9.9	-	-	-	-

Remarkable result for idle state (0.6W/TB)

Periodic full load activities that last 24hours for every 6 days
(Interval = 518400s) ≠ (EOS scan-interval = 604800s (7 days))
Uncorrelated with data transfers
So far, no clue found for these activities
(Stable electrical characteristics (currents, voltages, etc.)
No special features embedded in ME484)

Collected power-related metrics
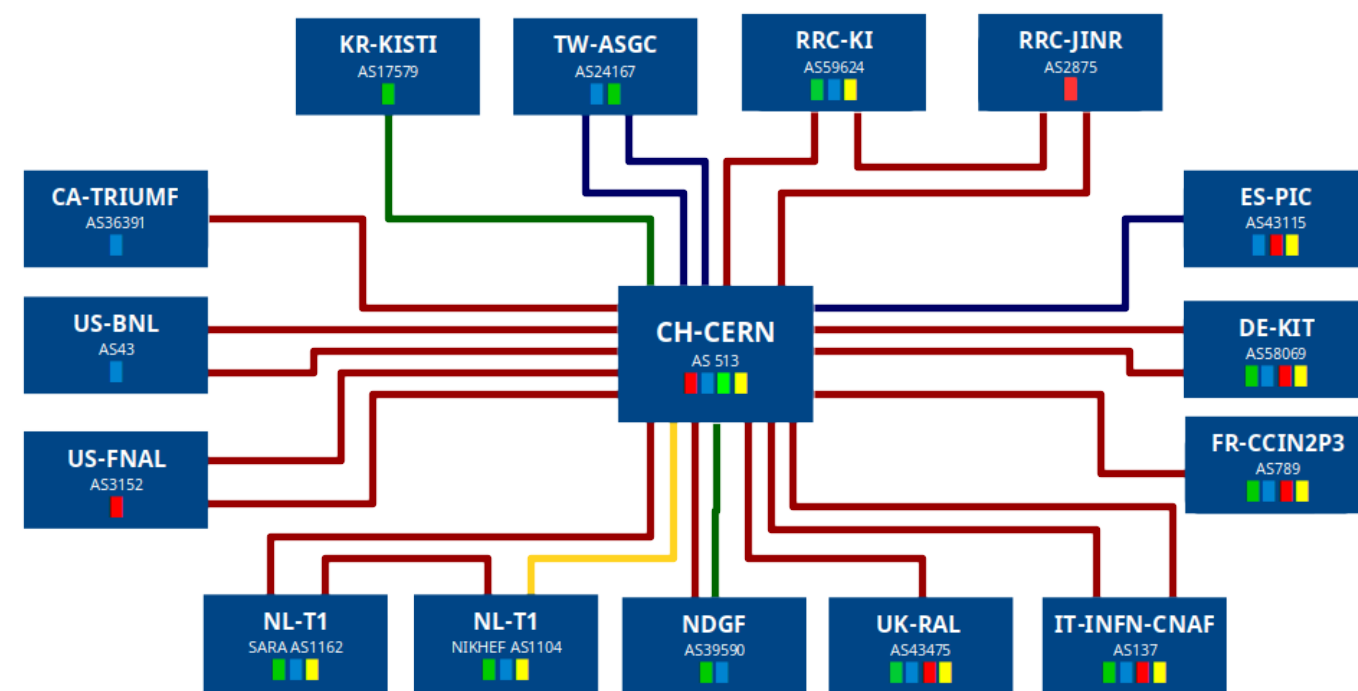for every minute via SNMP
from 12 PDUs in 3 racks

Current operations and planning of CDS to be presented @ EOS Workshop 2023

8

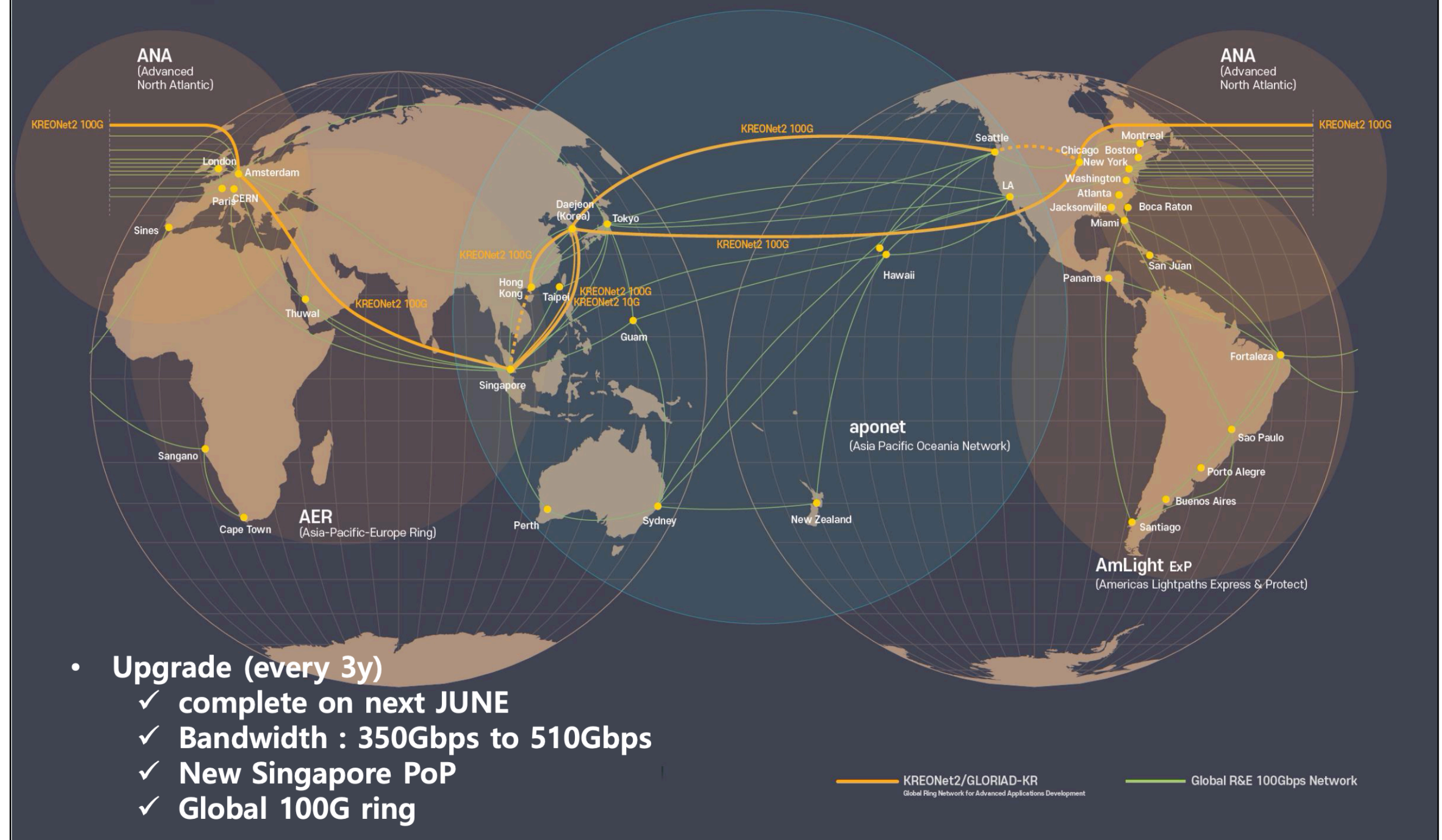**Cost analysis Tape vs. CDS to be presented @ CHEP2023**
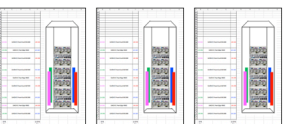
Networking - LHCOPN
Dedication to LHC Raw Data Transfer between T0 and T1s

- 20Gbps dedicated links from Daejeon to Geneva provided by KREONet2 with its 100Gbps lambdas
- Primary optical fibers: Daejeon-Chicago-Amsterdam-Geneva (Backup links through Daejeon-Seattle & GLORIAD-consortium)
- KREONet2 directly reaches Geneva from its Amsterdam PoP
- Provisioning of 100Gbps for OPN by the end of LHC RUN3 or before the start of HL-LHC (RUN4)
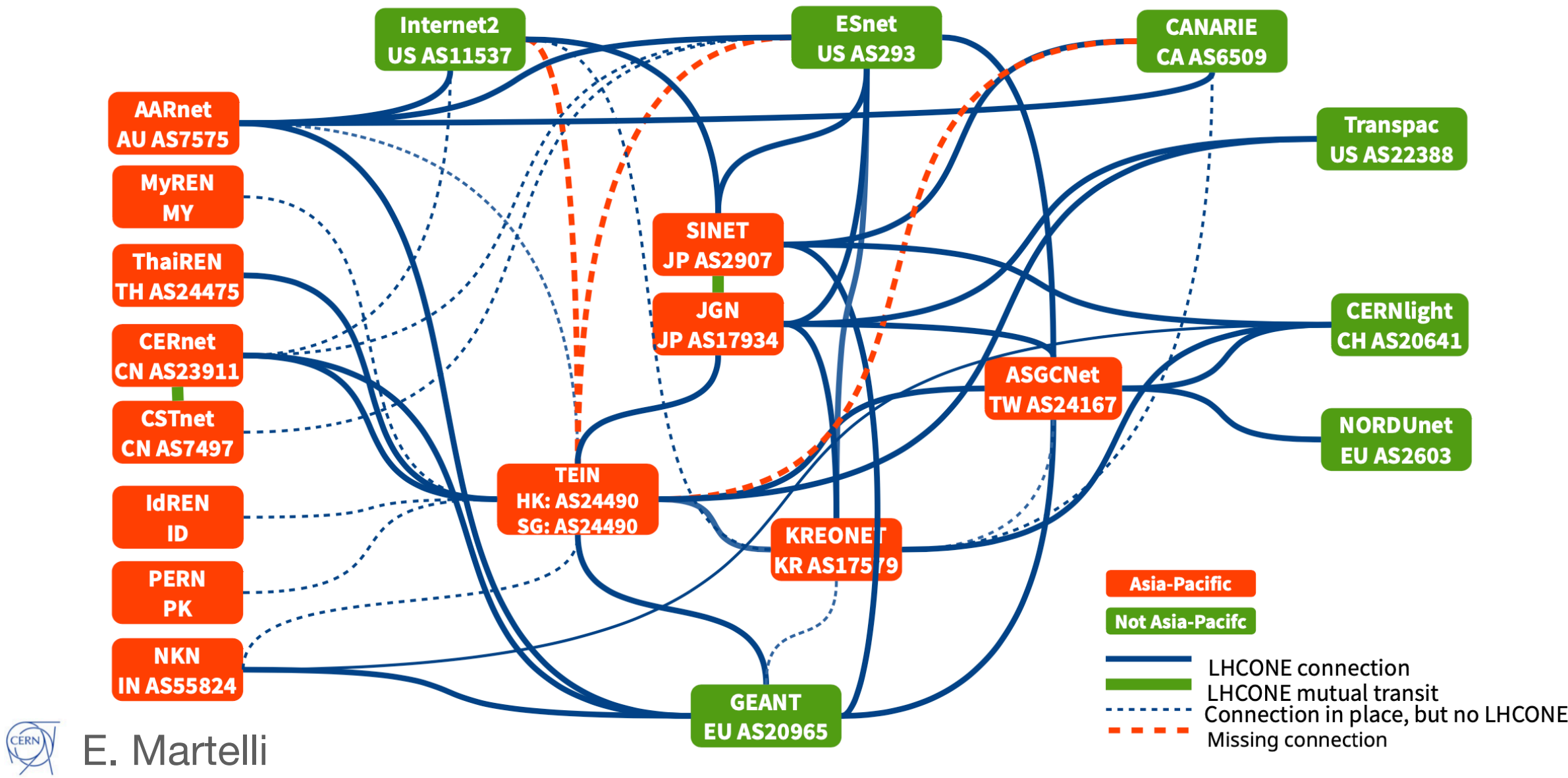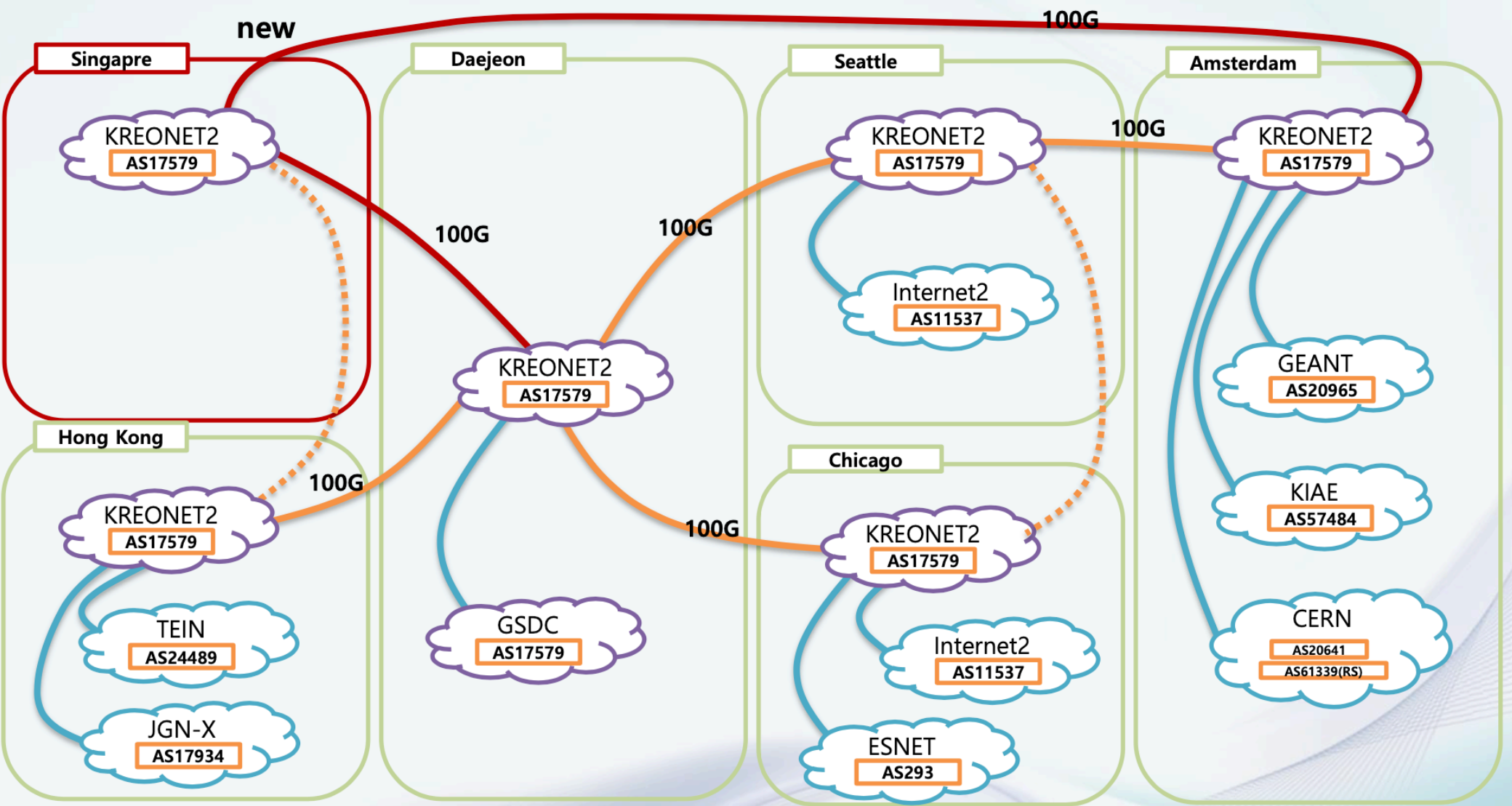

Map of KREONET2 2023

- Upgrade (every 3y)
  - ✓ complete on next JUNE
  - ✓ Bandwidth : 350Gbps to 510Gbps
  - ✓ New Singapore PoP
  - ✓ Global 100G ring

Networking - LHCONE

Towards full mesh reachability among Tier sites

Asia-Pacific VRFs – Current Status

E. Martelli

LHCONE on KREOENT2(2023)

- **Our Policy : allow transit**

C.-J. Park

- Policy that allows transit via KREONet2 resolves missing connections in Asia-Pacific region

10

Asia Tier Center Forum

- Its 6th series and the first F2F meeting among Asian Grid sites after the global pandemic

- Status and updates on Asian sites, experiments, and networking as well as fruitful discussions on Asian support model for computing cooperation

  - WG was formed and is working on developing proposed ideas and suggestions for the support model - a firm foundation supporting activities in similar time zones such as sharing expertises, exchange programs, training / developing technologies

- Centara Ao Nang, Krabi, Thailand
  - Co-hosted by SUT and KISTI-GSDC

- 24 registered participants (+4 remote)
  - 15 institutes from 8 countries
  - SUT(TH), Chulalongkorn U.(TH), Rajamangala University of Technology Isan (TH), BRIN(ID, former LIPI), VECC(IN), TIFR(IN), ASGC(TW), *Fudan U.(CN), *CCNU(CN), Hiroshima U.(JP), ICEPP(JP), *TEIN-CC(KR), KEK(JP), CERN(CH), KISTI(KR)
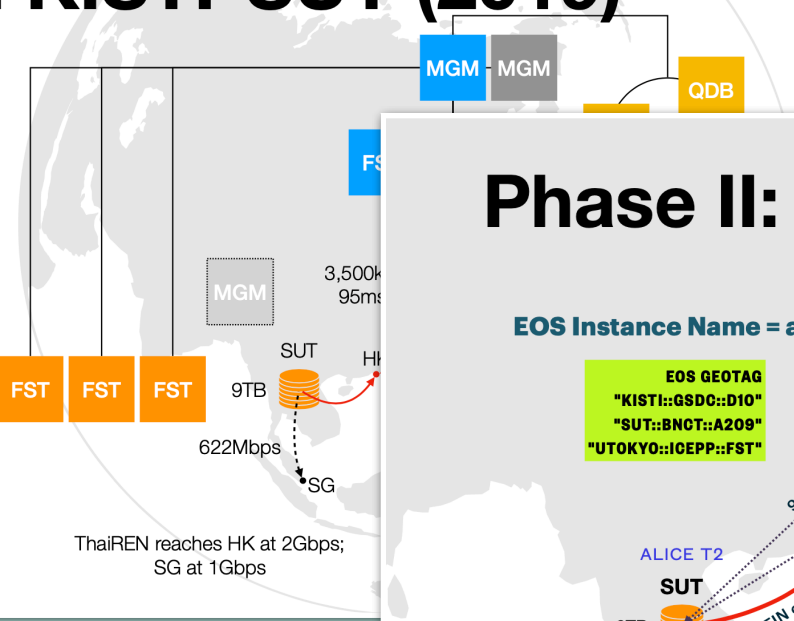
(*) Remote participation

Asian Distributed Storage Project

Phase I: KISTI-SUT (2019)

- EOS @ KISTI
  - MGM (Master/Slave)
  - QuarkDB cluster (3 nodes)
  - 3 FSTs (30TB HDD NAS)
- EOS @ SUT
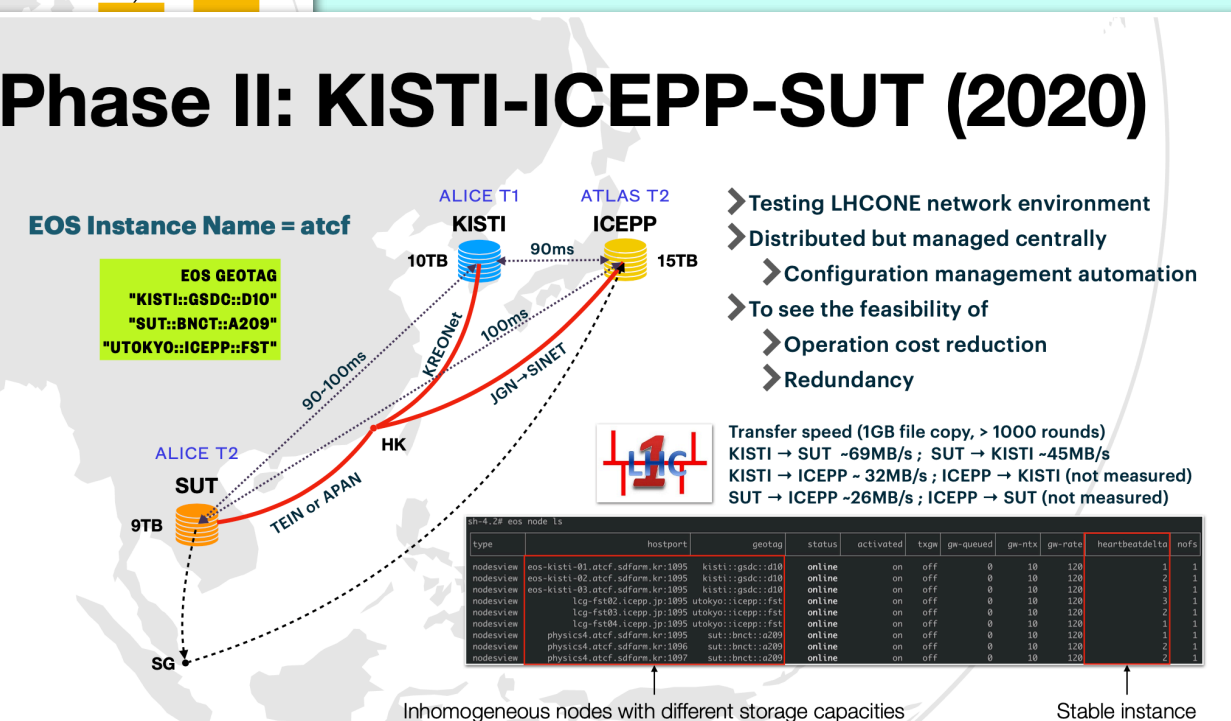 - 3 FSTs (9TB SSD NAS)
- EOS Instance Name = testatcf

ThaiREN reaches HK at 2Gbps; SG at 1Gbps

Phase II: KISTI-ICEPP-SUT (2020)

EOS Instance Name = atcf

- Testing LHCONE network environment
- Distributed but managed centrally
  - Configuration management automation
- To see the feasibility of
  - Operation cost reduction
  - Redundancy

Transfer speed (1GB file copy, > 1000 rounds)
KISTI → SUT ~69MB/s ; SUT → KISTI ~45MB/s
KISTI → ICEPP ~32MB/s ; ICEPP → KISTI (not measured)
SUT → ICEPP ~26MB/s ; ICEPP → SUT (not measured)

Inhomogeneous nodes with different storage capacities   Stable instance

Asian Support Model (ABC)

Baseline

- Individual efforts are limited and unsustainable
  - Envisaging an Asian version of computing cooperation to support S&T communities such as European projects (*EGI, EOSC, ESCAPE, ...) and US projects
  - Proposing "Asian Big science Clouds (ABC)"
    - A virtual collaboration of computing sites (centers) in Asia for science (small)
    - Discussing and seeking practical methods to support the related
      - Including actions in the previous slide
      - Starting with a joint-session (or workshop) in computing-related e.g. ISGC, APAN, WLCG Workshop, CHEP, etc.

(*) ASGC runs ROC_Asia/Pacific

Sustainable Model?

- With the outcome from ABC, we could try to establishing a practical S&T cooperation program (for funding) through ASEAN, ASEAN+3, EAS, etc.
  - ASEAN (Association of Southeast Asian Nations)
    - Member States: Brunei Darussalam, Cambodia, Indonesia, Lao PDR, Malaysia, Myanmar, Philippines, Singapore, Thailand, Viet Nam
    - ASEAN Dialogue Partner: Australia, China, India, Japan, Republic of Korea, New Zealand, Pakistan (Sectoral), EU, UK, US, etc.
  - ASEAN+3 (ASEAN Plus Three) : ASEAN, China, Japan, Republic of Korea
  - ASEAN PLUS THREE COOPERATION WORK PLAN (2023-2027)
  - EAS (East Asia Summit) : ASEAN+3, Australia, New Zealand, India

ASEAN PLUS THREE COOPERATION WORK PLAN 2023-2027

2.8 Strengthen cooperation in science, technology, and innovation (STI) through:

(a) Exploring cooperation in human resource development as well as development of technical and vocational skills and networking on STI, and promotion of public-private partnership;

(b) Fostering an open, equitable and environment for the development of STI, including by effectively protecting Intellectual Property Rights (IPR);

(c) Exploring joint capacity building activities, exchange of information, and sharing of best practices in areas of mutual interest such as STI policies, technology transfer, commercialisation, products and scientific standards, technical regulations, productivity, investment and intellectual property rights (IPR) management;

(d) Promotion of research and technology development in areas with potential for commercial applications such as robotics and automation, next-generation automobile, technology, and novel food, biotechnology, food technology, new materials, microelectronics and information technology, marine resources, new and renewable energy, climate change, life science, medical devices and technology, and space technology;

(e) Cooperation in meteorology addressing climate information and prediction services, weather observations and climate change; and

(f) Promotion and/or participation in relevant STI events, for example, the APT Centre for the Gifted in Science (ACGS), the APT Young Scientists Collaborative Innovation Forum and innovation youth camps and awards.

Supporting Domestic Research

Providing data storage, analysis pipeline and access

- Adapting the knowledge learned from operating Grid facilities to domestic region

- Dedicated optical links provided by KREONet for efficient data transfer and sharing
    - ‣ No need to move data by using external drives and overseas delivery

- Data analysis pipeline running on compute clusters
    - ‣ No need to own and maintain private cluster at individual labs

- User access to data and analysis pipeline without geographical constraints

➡ **Significant reduction of time in research activities**
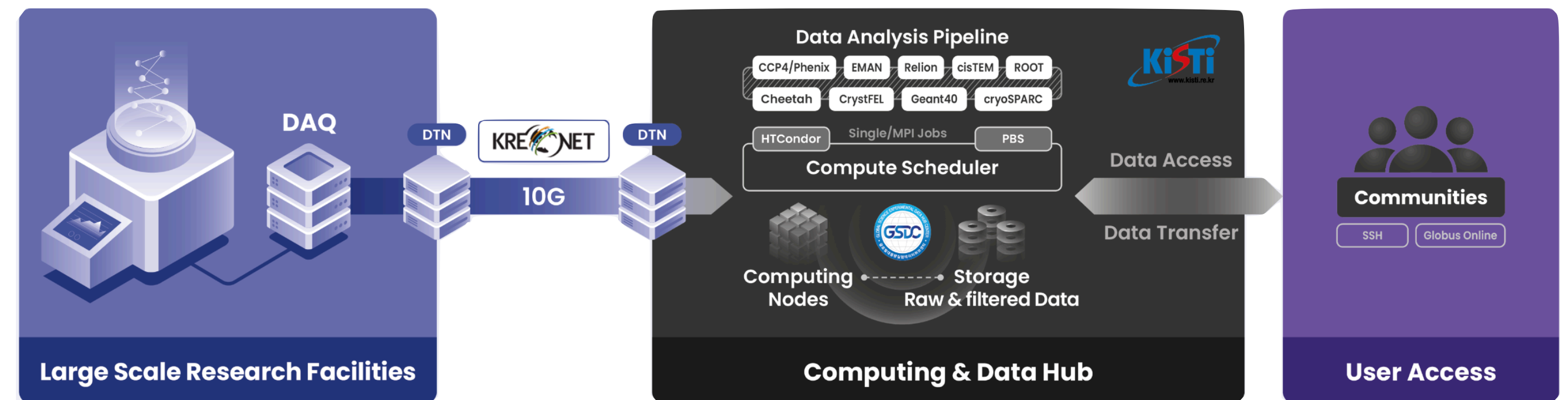
Summary

- KISTI-GSDC is a datacenter for data-intensive fundamental research

  - Its role is to provide data computing environment where needed

  - ALICE T1 is a representative example and there are other Grid facilities such as CMS T2, LDG T2 and Belle II T2

  - The knowledge has been transferred to support domestic research area and has helped to improve research activities

Thank you