HEPiX Autumn 2020 online Workshop

KEK Site Report

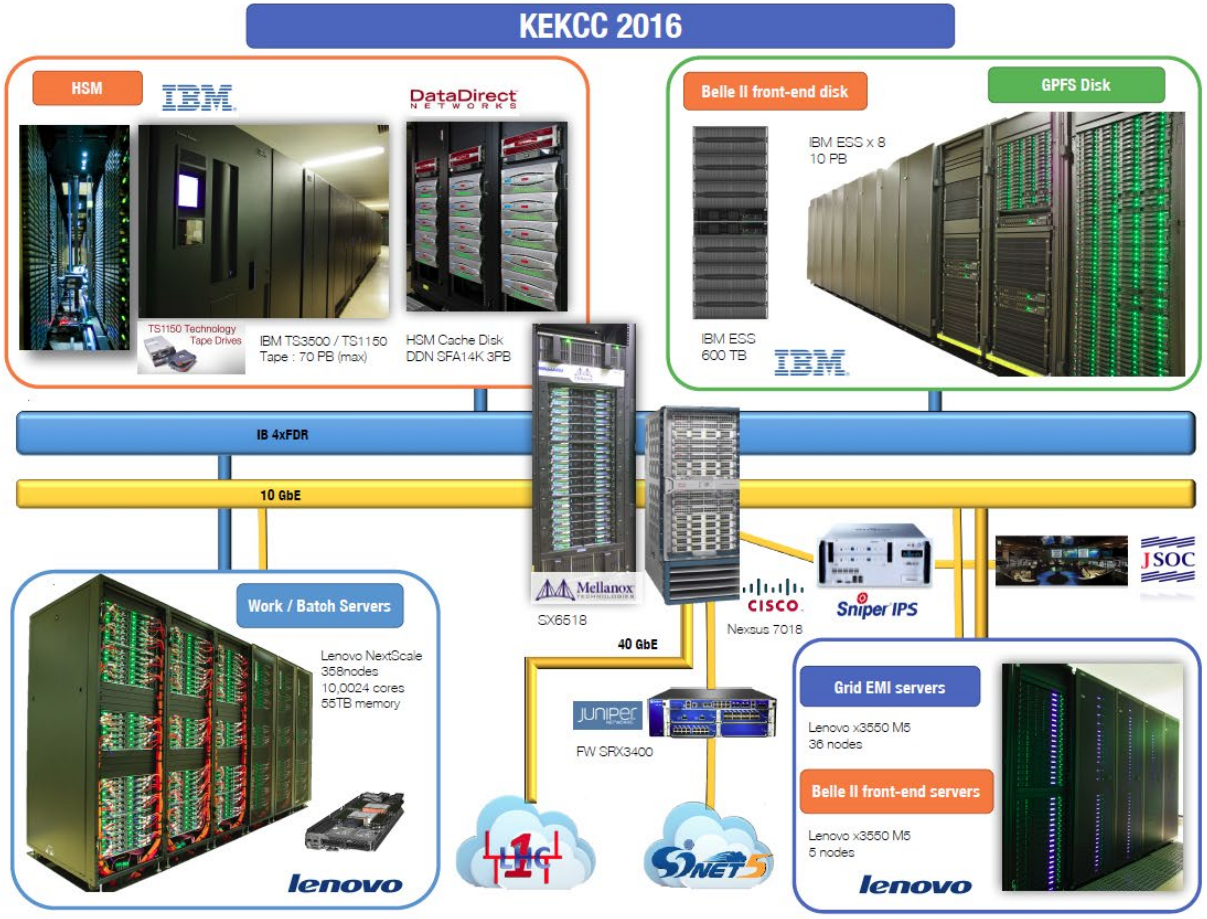
T. Nakamura, G. Iwai, K. Murakami,
T. Sasaki, S. Suzuki, W. Takase

Computing Research Center
Applied Research Laboratory
HIGH ENERGY ACCELERATOR RESEARCH ORGANIZATION, KEK

Computing Research Center

KEKCC: KEK Central Computer System

Launched in Sep. 2016.
No major upgrade in terms of the HWs during the 4 years operation period. Quite a stable phase.



SYSTEM RESOURCES

CPU : 10,024 cores
- Intel Xeon E5-2697v3 (2.6GHz, 14cores) x 2
  358 nodes
- 4GB/core (8,000 cores) /
  8GB/core (2,000 cores) (for app. use)
- 236 kHS06 / site

Disk : 10PB (GPFS) + 3PB (HSM cache)

Interconnect : IB 4xFDR

Tape : 70 PB (max cap.)
HSM data : 8.5 PB data, 170 M files,
5,000 tapes

Total throughput : 100 GB/s (Disk, GPFS),
50 GB/s (HSM, GHI)

JOB scheduler : Platfrom LSF v9

K. Murakami

Breakdown of CPU consumption

Compute node

CPU: Intel Xeon E5-2697v3 (2.6GHz, 14cores) x 2
358 nodes, 10,024 cores, 236kHS06/site

Memory: 4GB/core (8,000 cores)
8GB/core (2,000 cores)

Storage

Disk: 10PB (GPFS, IBM ESS x8 racks)
3PB (HSM cache)

Interconnect: InfiniBand 4xFDR (56 Gbps)

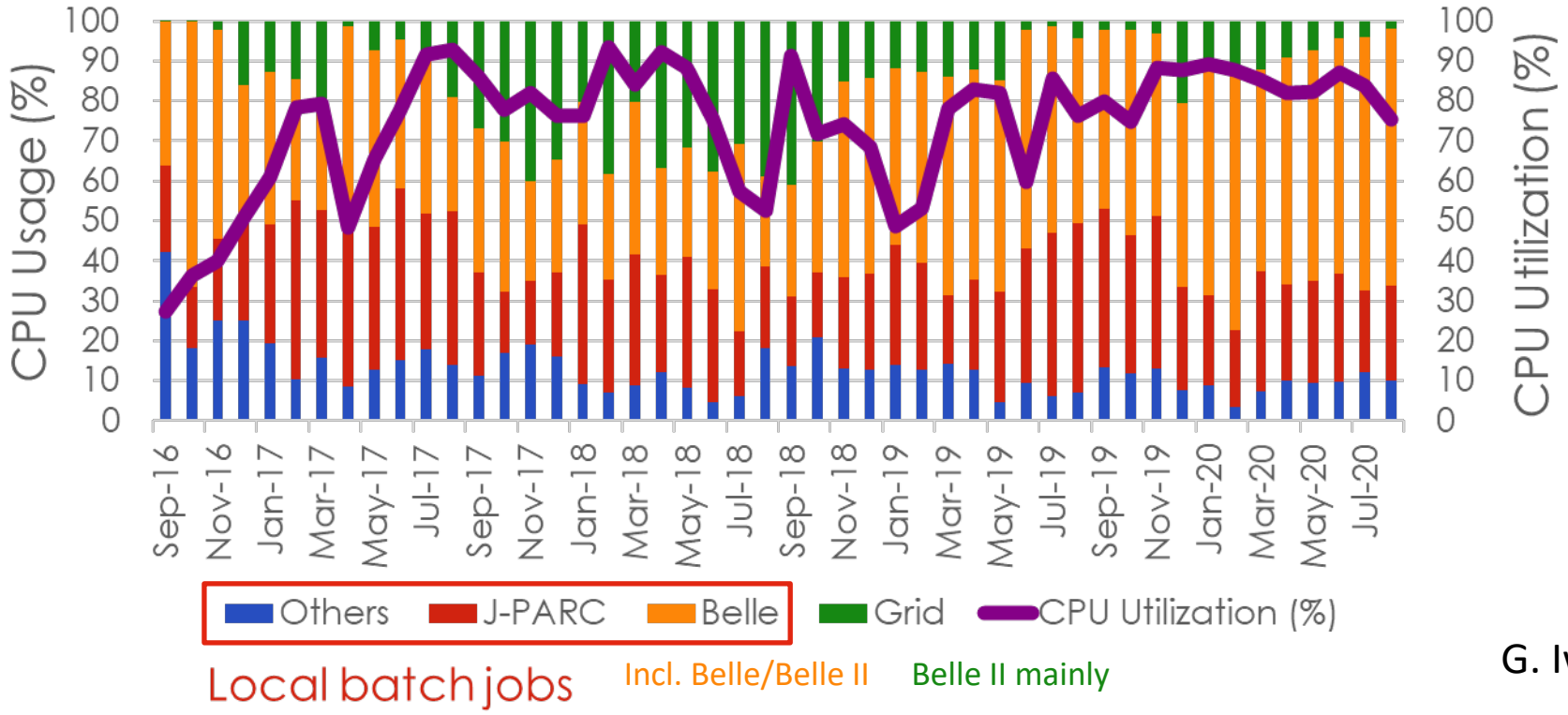
Tape: 70 PB (max cap.)

CPU usage: breakdown by groups,
normalized by the total CPU usage per month .

CPU usage has been reached more than 90 % of total resource.

Throughput

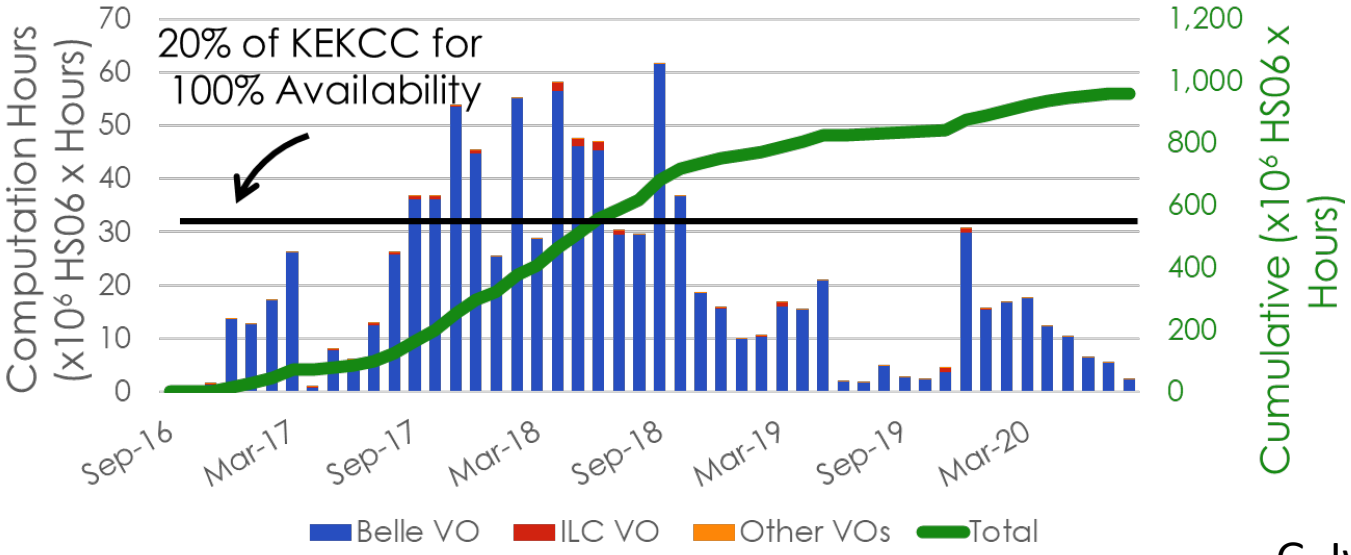
100 GB/s (Disk, GPFS), 50 GB/s (HSM, GHI)



G. Iwai

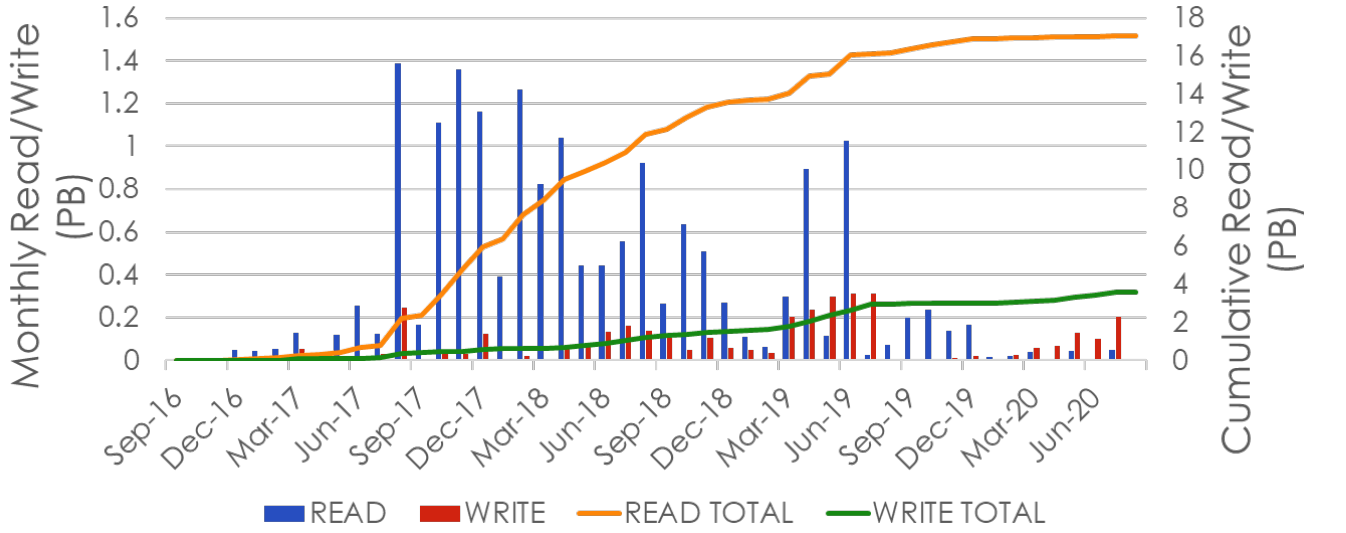
Grid Jobs and Data

Grid Jobs
168M HS06 hour/month
(23.5 HS06/core)



G. Iwai

Grid storage read/write
external data transfer



17 PB delivered to the other sites, 4 PB transferred to KEKCC during the 4 years.

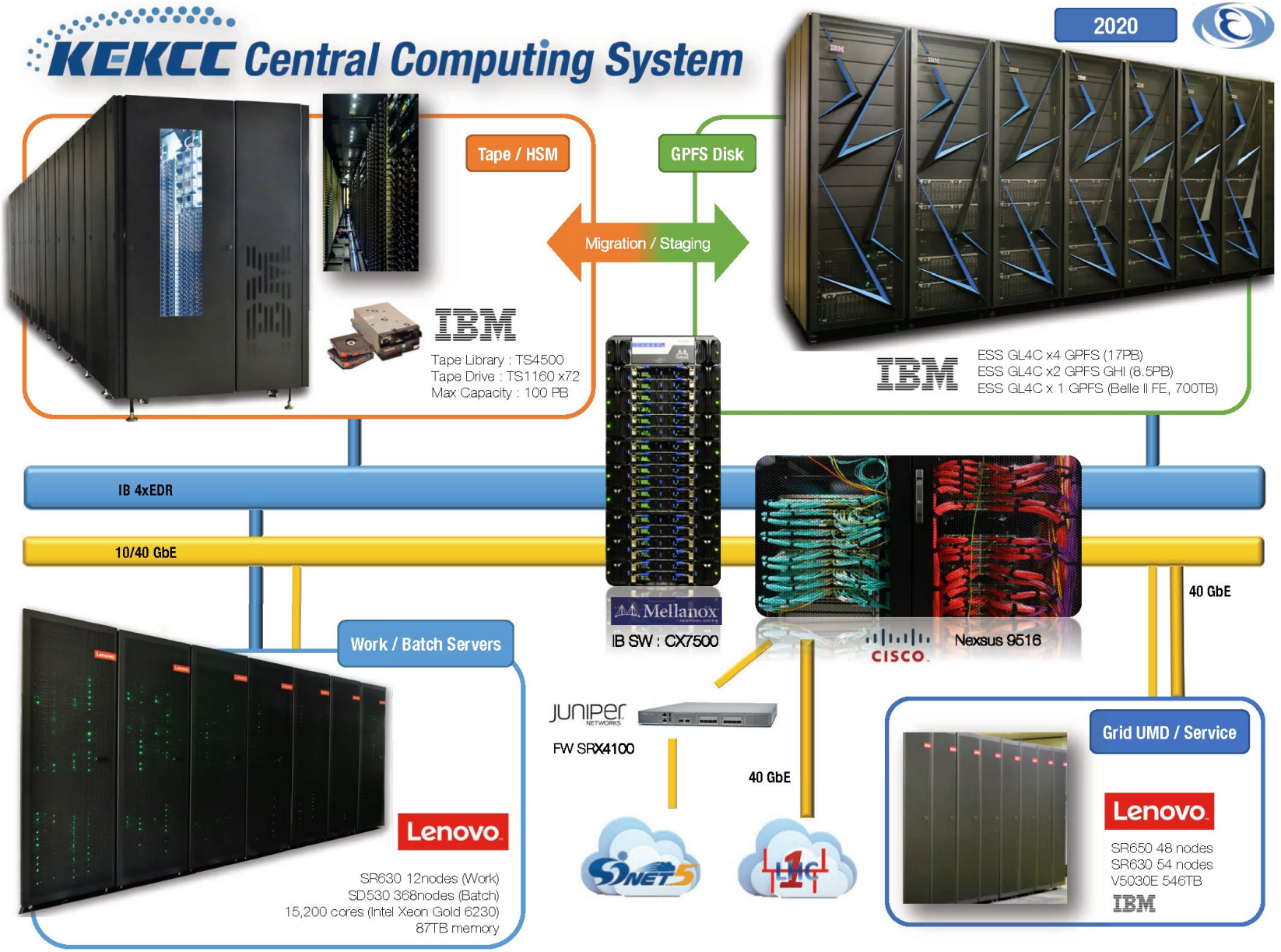
System migration

Computer South bldg.
(Previous system)

Computer North bldg.
(New system)

Dec. 2019:	End of procurement, Performance evaluation
Jan. - Mar. 2020:	Hardware delivery
Apr. - Jul. 2020:	System construction and setup
Aug. 2020:	Data migration and System stress test
Sep. 2020:	Start operation

New KEKCC 2020



Resource comparison

Launched on Sep. 1st

K. Murakami

	2016	2020	Upgrade Factor
CPU	Xeon E5-2697v3 Haswell (2.6GHz, 14cores)	Xeon Gold 6230 Cascade Lake (2.1 GHz, 20 cores)	
CPU cores	10,024	15,200	**x1.5**
HS06	236k	480k (est.)	**x2**
OS	SL 6.10	CentOS 7.X (7?)	
Disk Capacity	10 + 3 PB (HSM)	17 + 8.5 PB (HSM)	**x2**
Tape Drive	IBM TS1150 x54	IBM TS1160 x72	
Tape Media	7 TB/vol (JC) 10 TB/vol (JD), 360 MB/s	7 TB /vol (JC) 15 TB/vol (JD-Gen6) 20 TB/vol (JE), 400 MB/s	
Tape max capacity	70 PB	100 PB	**x1.4**

Worker node configuration

- CPU: 40 cores / node
- Memory: 4.8GB/core (304 nodes), 9.6GB/core (72 nodes)
- Storage: 960GB SATA SSD / node

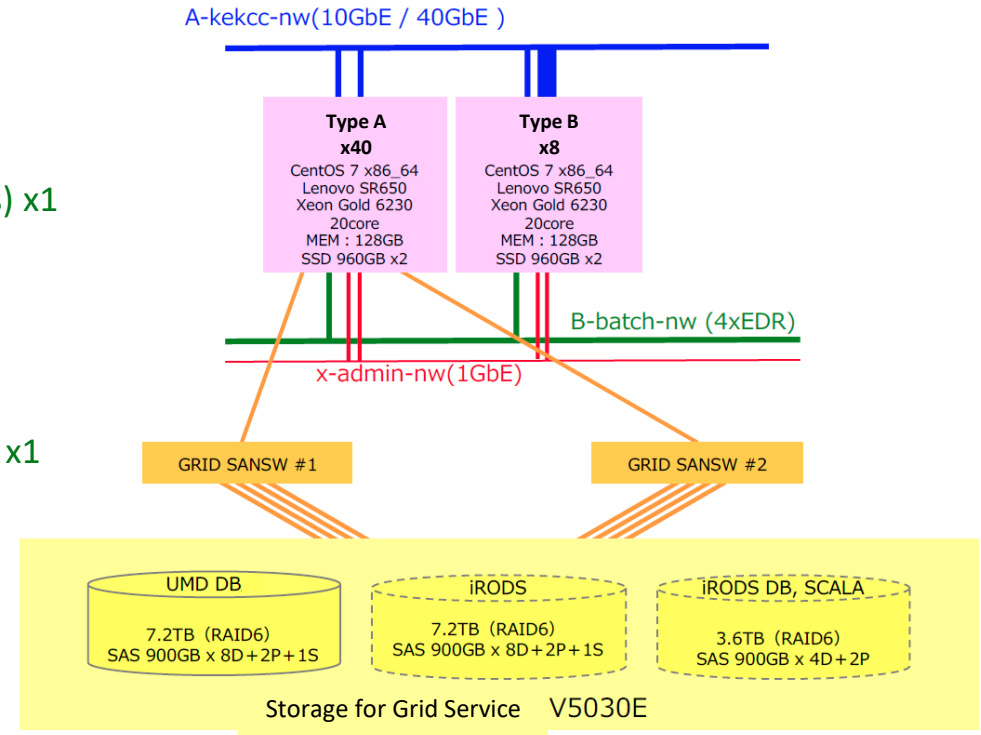
Specification of Grid instances

Type A: Head node of Grid instance x40

CPU: Xeon Gold 6230 2.10GHz 20Cores x1
Memory: 128GB
Disk: 960GB SATA-SSD x2
Network: 1GE x2, 10GE x2, 16GFC x2, IB (4xEDR, 100Gbps) x1

Type B: Data transfer node x8

CPU: Xeon Gold 6230 2.10GHz 20Cores x1
Memory: 128GB
Disk: 960GB SATA-SSD x2
Network: 1GE x2, 10GE x2, **40GE x1**, IB (4xEDR, 100Gbps) x1



Type A

BDII-top (AA), BDII-site (AA, UP)
VOMS (HA, UP)
LFC-Belle-RW (HA, UP), LFC-Belle-RO (AA, UP), LFC-Other (UP)
SE-StoRM-FE-Belle-Raw (AA)
SE-StoRM-BE-Belle-Raw
SE-StoRM-Belle-Ana (AA)
SE-StoRM-Other
CE-ARC (AA)
APEL
CVMFS-Stratum0 (AS)
CVMFS-Update
CVMFS-Stratum1 (AS)
FTS (AA, UP)
AMGA (AA, UP)
HTTP-Proxy (AA)
ARGUS (AA, UP)
Nagios

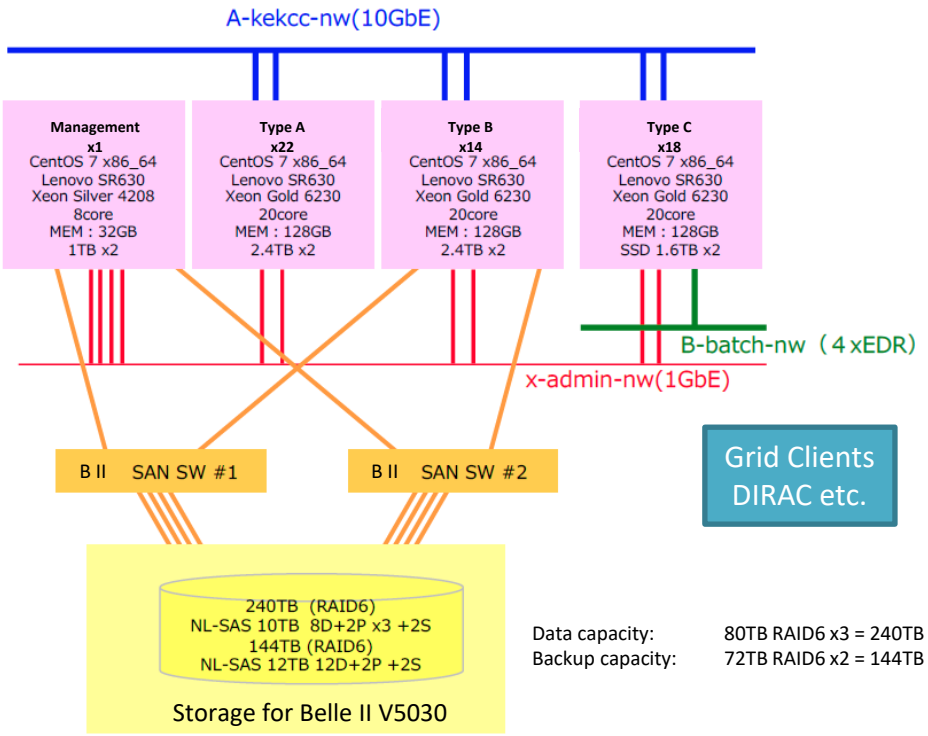
Type B

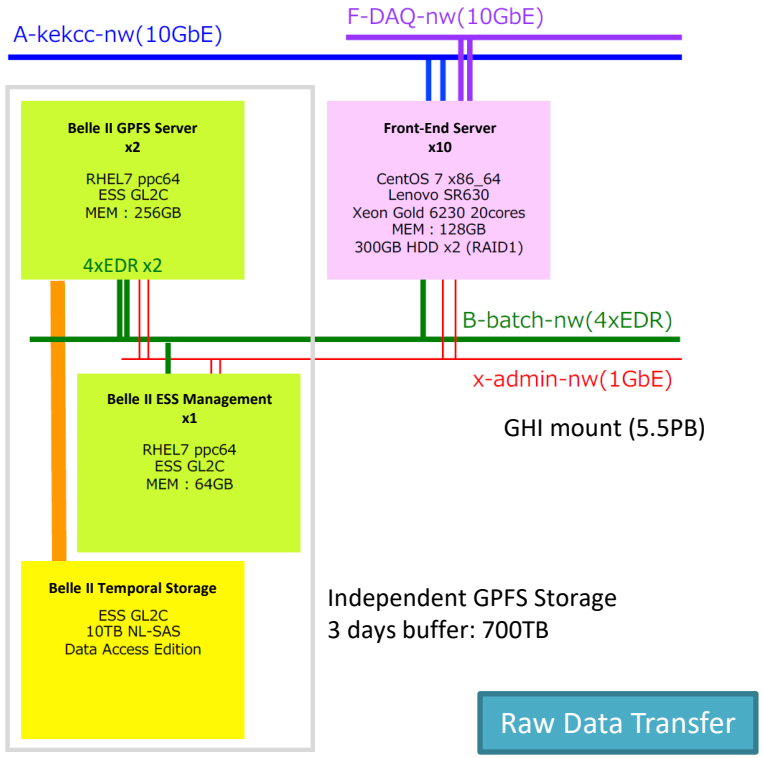
DTN-Belle-Raw (40GE x4)
DTN-Belle-Ana (40GE x2)
DTN-Other (40G x2)

- HA: High Availability
- AA: Active-Active
- AS: Active-Standby
- UP: External power supply

- Almost service instances are deployed by CentOS7.
- Some light-weight instances are running on KVM.
- StoRM, VOMS, LFC are deployed on RHEL6.
  - CentOS7 packaging is not in time for the construction.
  - OS Support will be terminated by the end of November.
  - Use End of Lifecycle Support add-on (not-free).

Experiments dedicated servers





Type A: Belle II x16

CPU:	Xeon Gold 6230 2.10GHz 20Cores x1
Memory:	128GB
Disk:	2.4TB SAS-HDD x2
Network:	1GE x2, 10GE x2

Type B: Belle II x14

CPU:	Xeon Gold 6230 2.10GHz 20Cores x1
Memory:	128GB
Disk:	2.4TB SAS-HDD x2
Network:	1GE x2, 10GE x2, 16GFC x2
Power:	External power supply (10 servers)

Type C: for Belle II x18

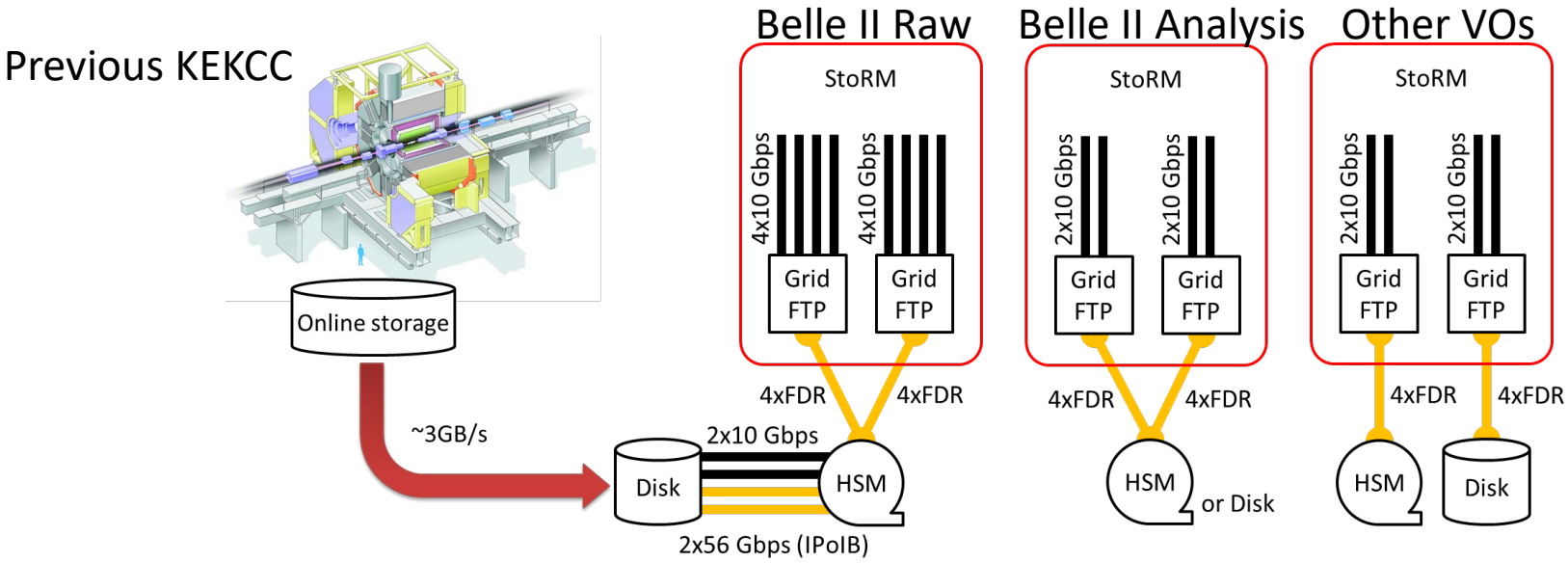
CPU:	Xeon Gold 6230 2.10GHz 20Cores x1
Memory:	128GB
Disk:	1.6TB SAS-SSD x2
Network:	1GE x2, 10GE x2, IB (4xEDR, 100Gbps) x1
Power:	External power supply (14 servers)

Front-End Server: x 10

CPU:	Xeon Gold 6230 2.10GHz 20Cores x1
Memory:	128GB
Disk:	300GB SAS-HDD x2
Network:	1GE x2, 10GE x2, IB (4xEDR, 100Gbps) x1, 10GE-F-DAQ-NW x2

Performance upgrade of Grid system

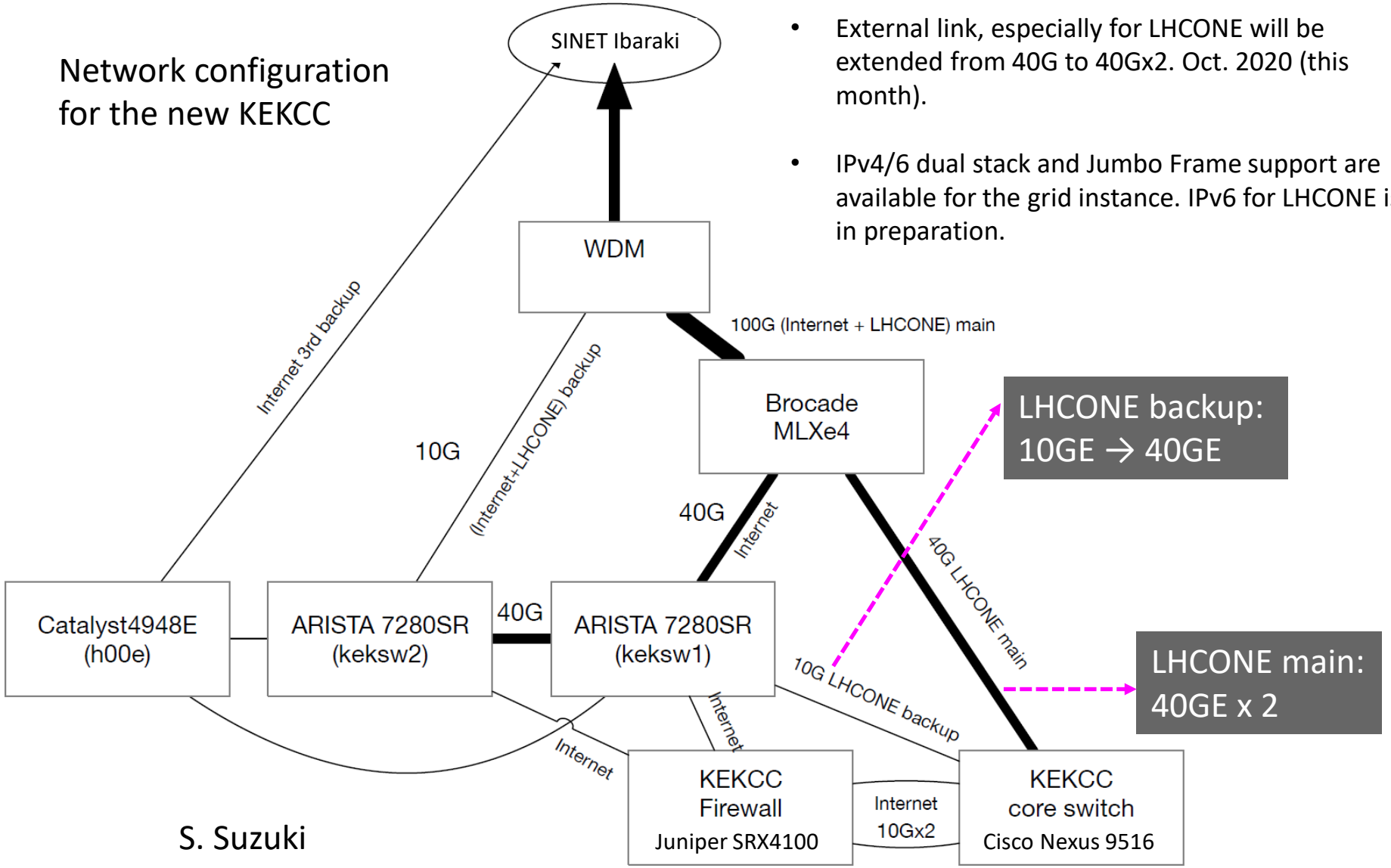
- Basic configuration does not change from the current system in terms of redundancy and robustness.
  - Redundant configuration
    - CE, CVMFS Stratum0/1, HTTP proxy, BDII-top, GridFTP servers behind StoRM
  - High availability configuration by LifeKeeper
    - VOMS, AMGA, LFC
  - Uninterruptible operation against the scheduled power outage
    - VOMS, AMGA, LFC, FTS3, ARGUS, BDII-site
- All of the systems are built based on Centos 7 (partially RHEL8).
- Some lightweight services are prepared by the virtual machine (KVM).
- CREAM computing element are replaced to ARC-CE.
- The capability of the data transfer nodes are strengthened.
  - Belle II Raw: 10Gx4 cables x 2 nodes (80G) → 40G x 4 nodes (160G)
  - Belle II Analysis: 10Gx2 cables x 2 nodes (40G) → 40G x 2 nodes (80G)
  - Other VO: 10Gx2 cables x 2 nodes (40G) → 40G x 2 nodes (80G)



G. Iwai

Reinforcement of network bandwidth

Network configuration for the new KEKCC



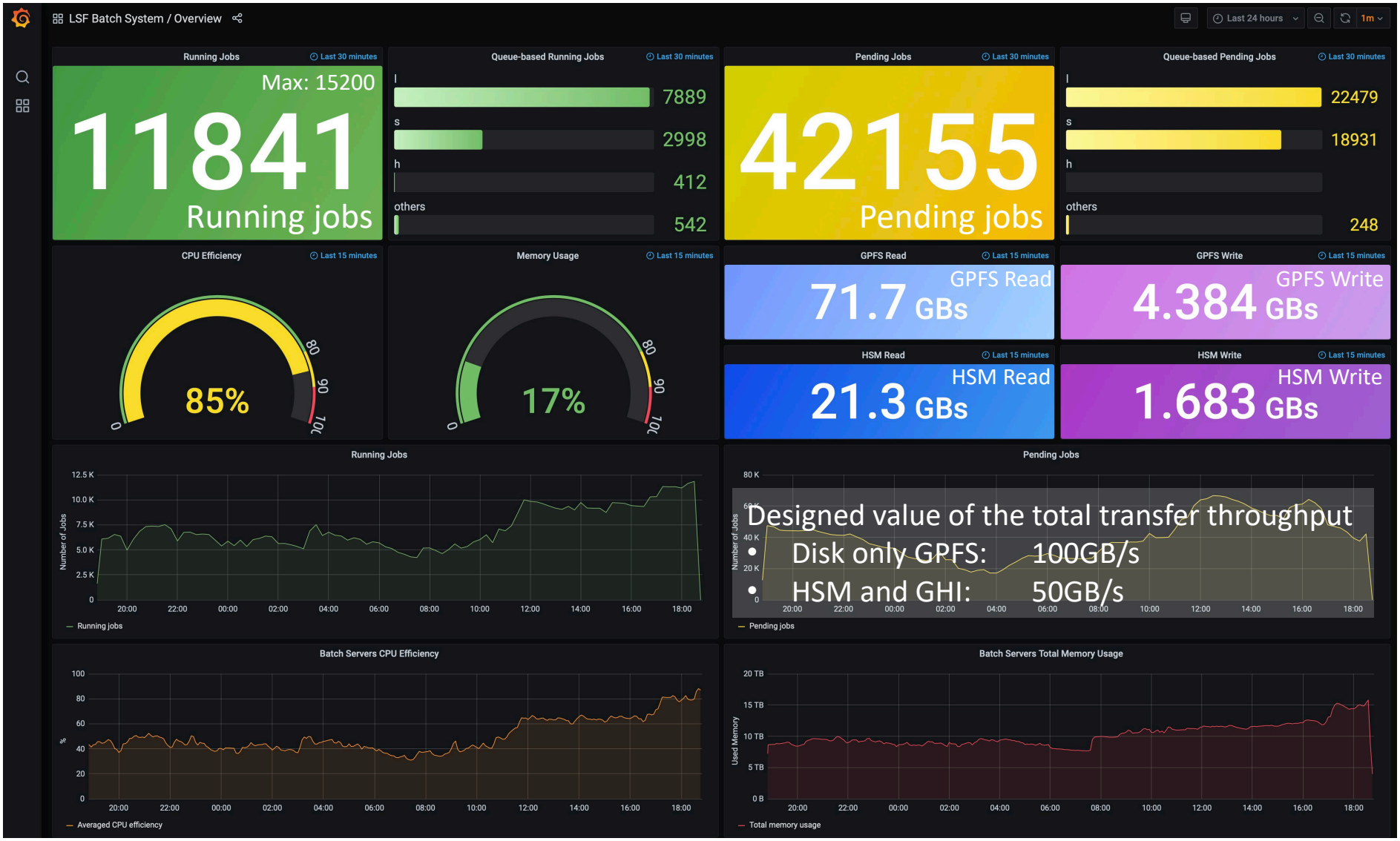
- External link, especially for LHCONE will be extended from 40G to 40Gx2. Oct. 2020 (this month).
- IPv4/6 dual stack and Jumbo Frame support are available for the grid instance. IPv6 for LHCONE is in preparation.

S. Suzuki

Status of the new KEKCC

Applied Research Laboratory

Snapshot taken on Oct. 6th



Summary

The new KEK central computer system (KEKCC) has been launched on Sep. 1st, 2020.

- Basic system functionality is all available not only for the data analysis system but also for the IT infrastructure including Grid-CA, email, Mailing list, Web, Indico, Wiki, Online storage, etc.
- A lot of minor issues still remain.
  - Performance and parameter tuning, and so on.
- Need to upgrade RHEL6 based Grid instances, e.g. StoRM, VOMS.
  - DB entry in LFC will be migrated to the Rucio at BNL.
  - Need to explore the quick DB upgrade of PostgreSQL for Belle II AMGA (Currently Ver. 8.4.20, ~250GB DB takes 1 week to take full dump)

Computing requirements from the next generation experiments hosted by and related to KEK are becoming high.

- Actually, we support Belle II, ILC, KAGRA and many pilot projects.
- Several projects have interests in going to utilize the Grid infrastructure.
  - J-PARC (muon g-2)
  - T2K / Hyper Kamiokande
  - Other small experiments