Grid Computing Infrastructure at KEK

Hiroyuki Matsunaga
(Computing Research Center, KEK)

LCWS2016, Morioka
Dec. 8, 2016

KEK Central Computer System (KEKCC)

- Only large system that supports many projects at KEK
  - Belle/Belle II
  - J-PARC (Neutrino, Kaon, Muon, Hadron, Material and Life Science)
  - ILC
- Japanese procurement/operation model is very different from that of US/EU
  - The whole system is replaced every 4-5 years
    - According to the Japanese Government procurement model
  - System is operated mostly by company's system engineers
    - Lab staff members are very limited (25 in Computing Research Center)
  - Bidding takes 1 year.
  - Major changes occur only at the replacement
  - Downtime is unavoidable
  - Data migration is another concern as data volume gets bigger
- Current system started the operation in September this year

Recent system upgrade

	Old	New	
CPU Server	IBM iDataPlex	Lenovo NextScale	
CPU	Xeon 5670 (2.93 GHz ,6core)	Xeon E5-2697v3 (2.6GHz, 14cores)	
CPU cores	4,000	10,024	**x2.5**
IB	QLogic 4xQDR	Mellanox 4xFDR	
Disk Storage	DDN SFA10K	IBM Elastic Storage System (ESS)	
HSM Disk Storage	DDN SFA10K	DDN SFA12K	
Disk Capacity	7 PB	13 PB	**x1.8**
Tape Drive	IBM TS1140 x 60	IBM TS1150 x54	
Tape Speed	4TB/vol, 250 MB/s	10TB/vol, 350 MB/s	
Tape max capacity	16 PB	70 PB	**x4.3**
Power Consumption	200 kW	250 - 300 kW	

10k CPU cores in 6 racks

Disk

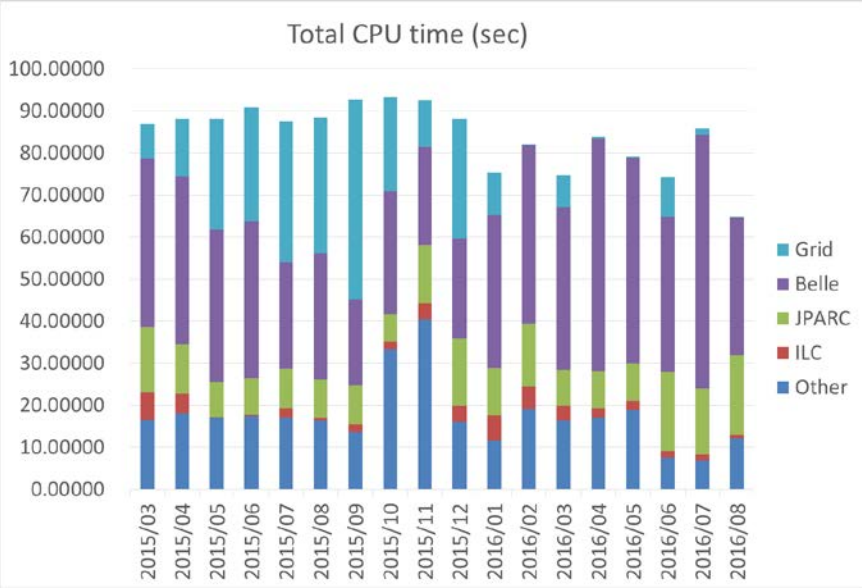
Tape

- Job scheduler: LSF v9
- HSM: GPFS + HPSS (GHI)

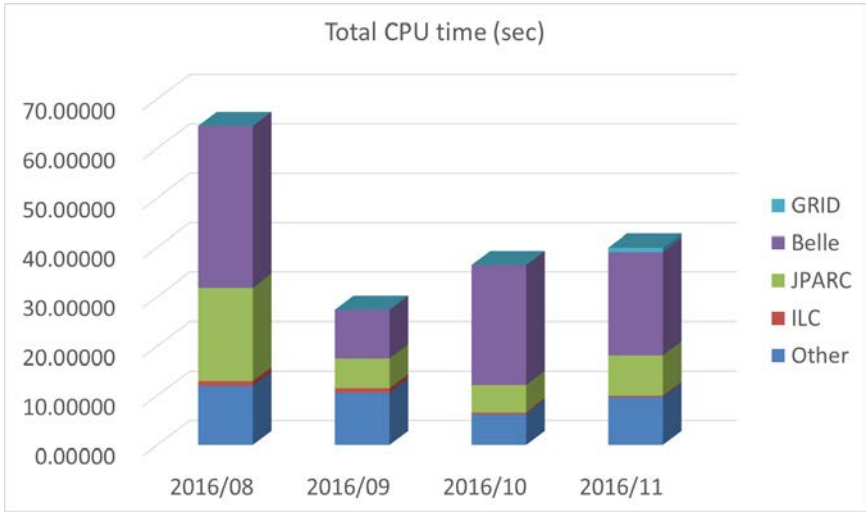
HPSS
High Performance Storage System

CPU usage in KEKCC

Previous system



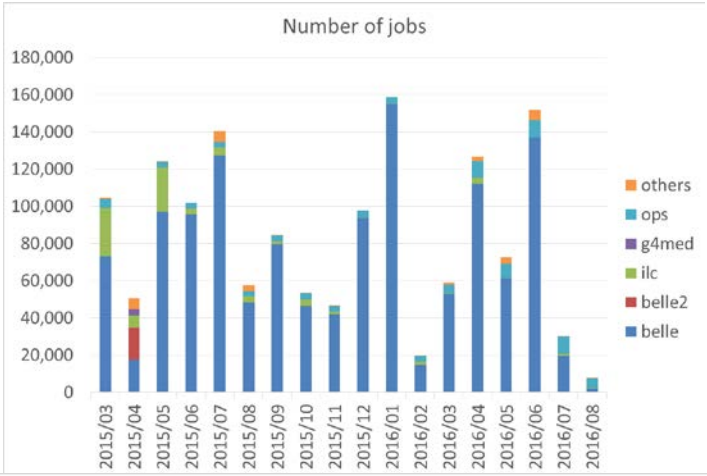
New system



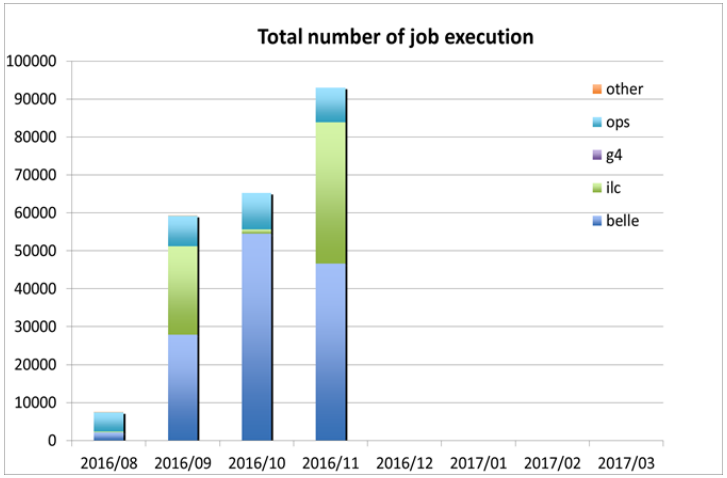
- CPU resources are shared by local and Grid jobs
- New system are still in the process of tuning and getting stable
- Grid jobs consume half of CPU time at peak

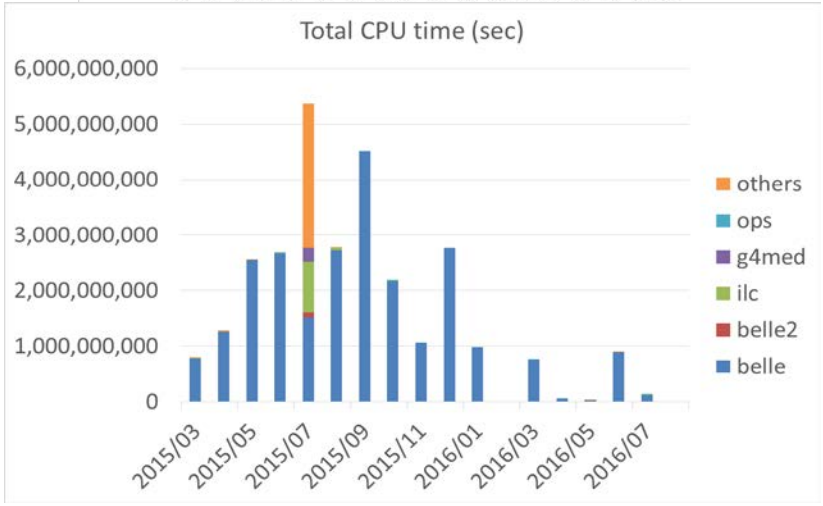
CPU usage of Grid jobs

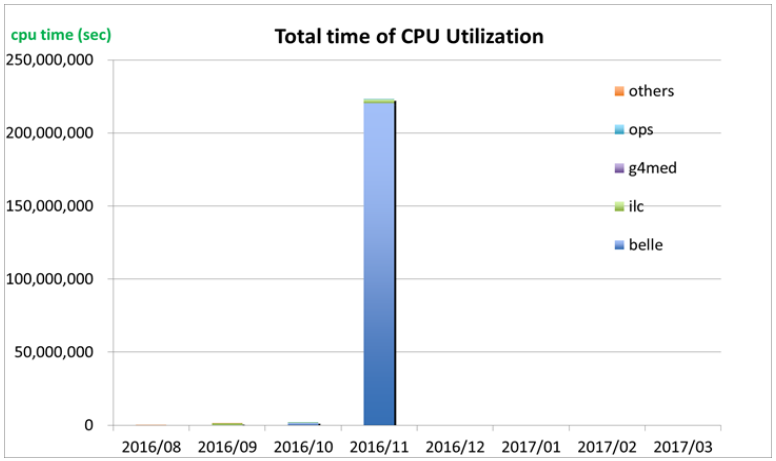
Previous system



New system







- Most of ILC jobs are short or not using CPU?

Grid activity at KEK

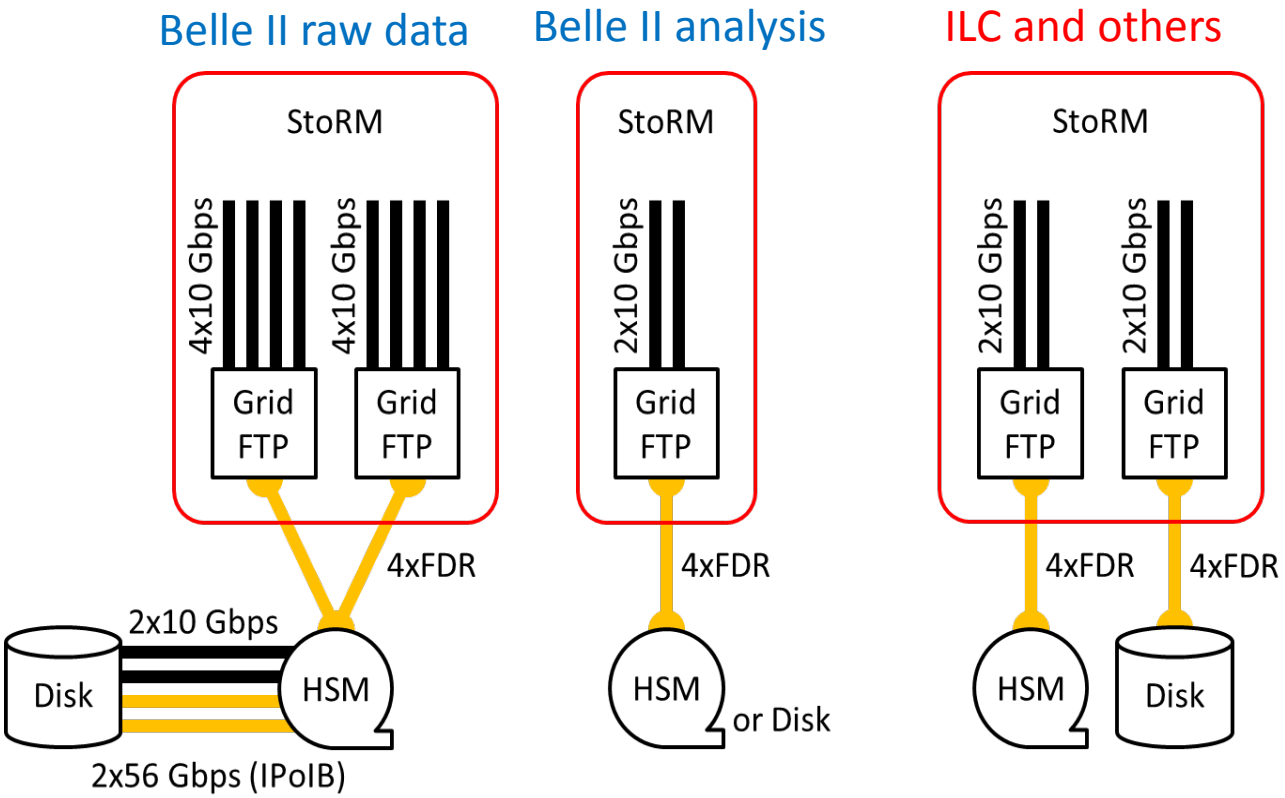
- Production services (Grid site, CA) since 2006
- Supports many VOs
  - Belle/Belle II (Tier-0), ILC, T2K, KAGRA, Geant4, etc.
  - KEK is involved in ATLAS, but not in computing
- Became an observer in WLCG in 2015
- Connect with LHCONE in 2016
  - LHCONE: Closed network for LHC and others (Belle II)
- Belle II will take physics data in the near future and be a major player

Grid Services on new KEKCC

- New services deployment
  - CVMFS
    - Stratum 0: belle.kek.jp, grid.kek.jp
    - Stratum 1: ilc.desy.de, dirac.egi.eu, t2k.egi.eu, …
      - Mainly for Japanese institutes and Asian sites

  - AMGA (Metadata Catalog), FTS3 (File Transfer Service)
    - For Belle II
- High availability configuration for important central services
  - VOMS, AMGA, LFC
- Reinforce data transfer performance
  - Increase number of servers and bandwidth
- DIRAC services are operated by Belle II people

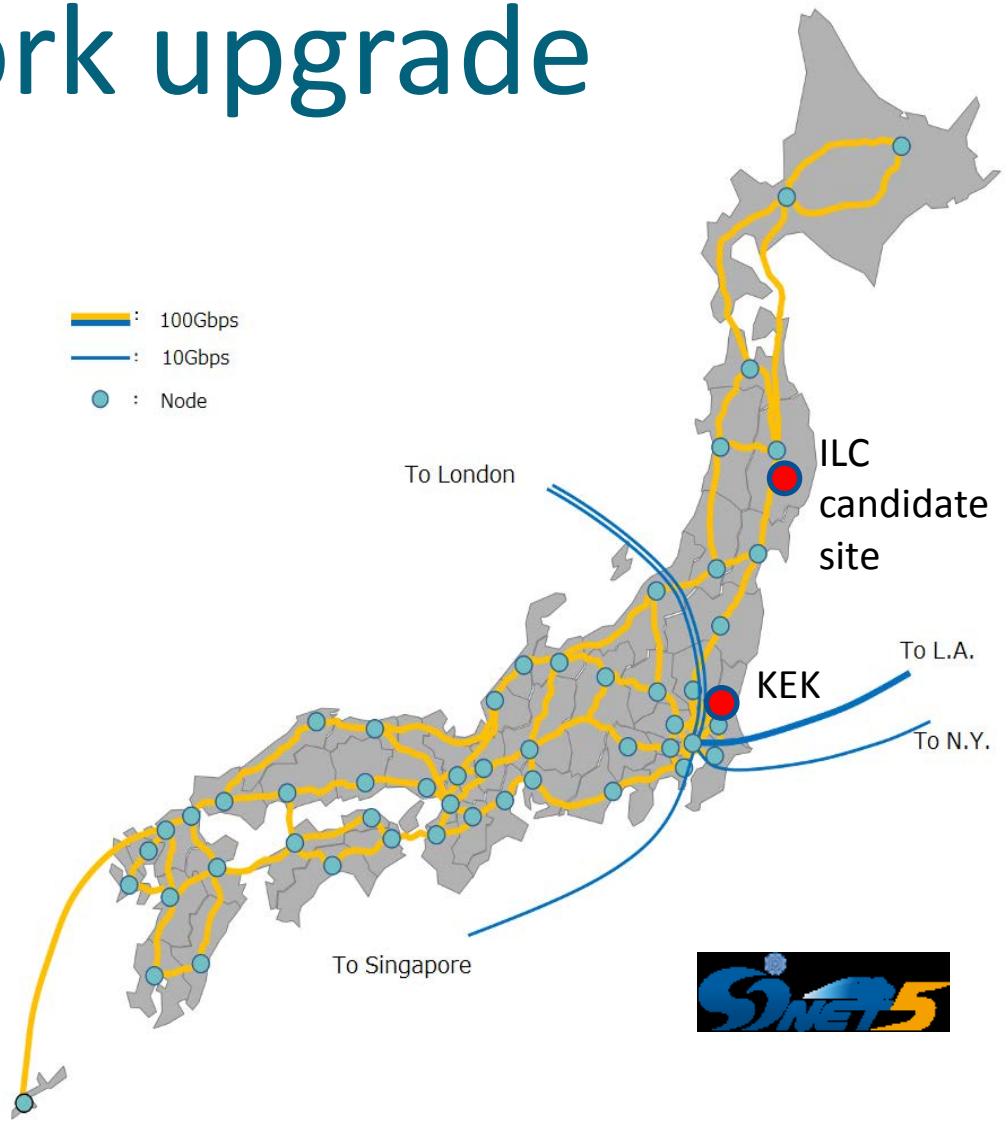
Data transfer



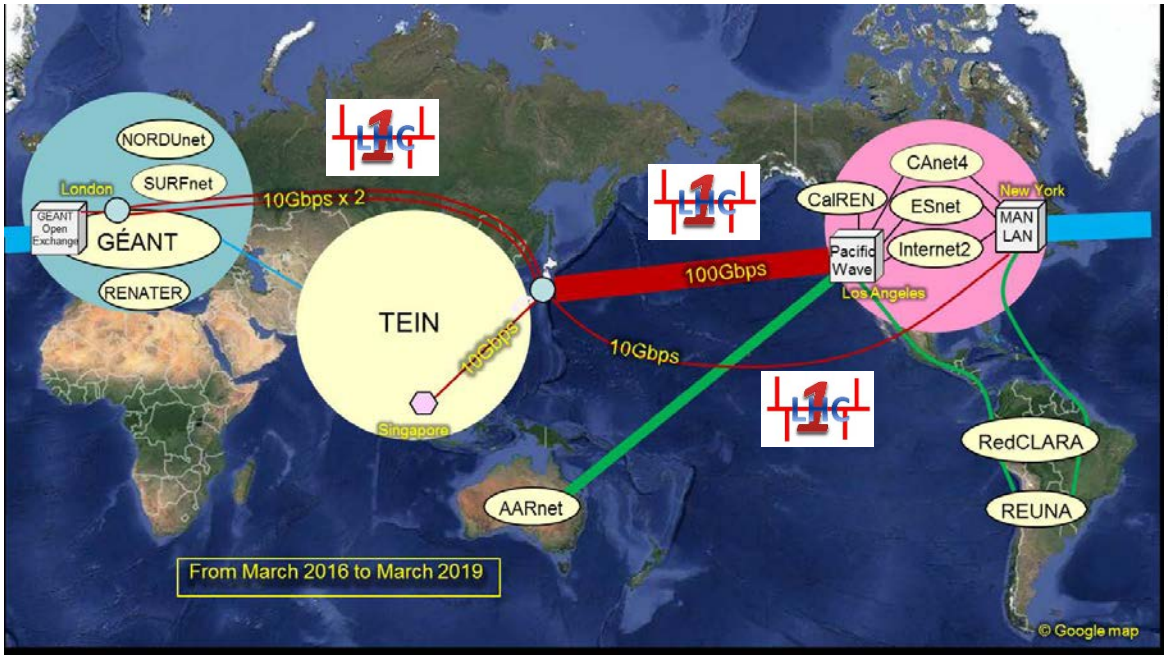
Total throughput:
HSM: 50GB/s (IBM GPFS+HPSS on DDNSFA12K)
Disk: 100GB/s (IBM GPFS on IBM ESS)

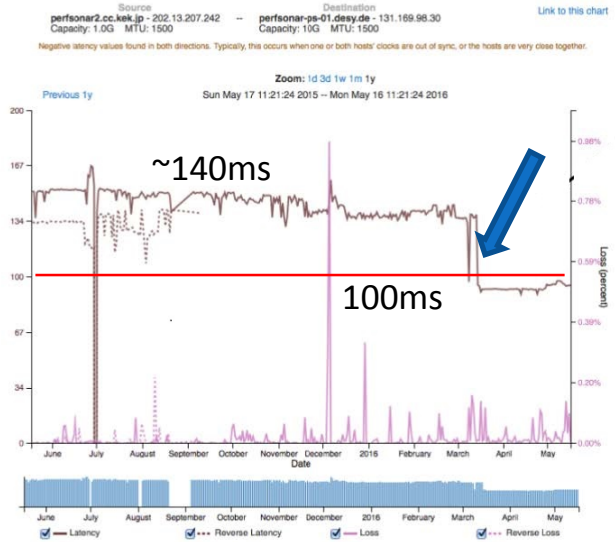
Network upgrade

- SINET (Japanese NREN) was upgraded in March 2016
  - SINET4 to SINET5
- Backbone upgrade from multi-10G or 40G to multi-100G
- JP-US: from 3x10G to 100G (L.A.) +10G (N.Y.)
- JP-EU(New): 2x10G
- At KEK: from 2x10G to 100G+10G



International Network





Latency between KEK and DESY (S. Suzuki)

- SINET5 provides 100G+10G to US and 2x10G for EU since Mar. 2016.
- LHCONE peering with GEANT, ESnet and CANARIE has started in Sep. 2016
  - After the launch of the new KEKCC

Performance improvement

KEK Outgoing

Destination	SINET4 [Gbps]	SINET5 old KEKCC [Gbps]	SINET5 new KEKCC LHCONE [Gbps]	Increase over old KEKCC & LHCONE
PNNL	3.6	3.9	8.4	115%
DESY	3	3	-	-
KIT	3.5	3.2	-	-
CNAF	-	3.8	9.0	136%
NAPOLI	3	3	8.8	190%

KEK Incoming

Source	SINET4 [Gbps]	SINET5 old KEKCC [Gbps]	SINET5 new KEKCC LHCONE [Gbps]	Increase over old KEKCC & LHCONE
PNNL	4.6	6.3	-	-
DESY	4	8	-	-
KIT	5	7	-	-
CNAF	7	7	13.5	93%
NAPOLI	5.5	6.6	13	97%

Vikas Bansal (PNNL) et al. (Belle II)

Summary

- New KEKCC has started operation in September 2016
  - Lease contract until August 2020
  - CPU : 10K cores (x2.5), 240 kHS06 (x4.0)
  - Disk : 13PB (x1.8), Tape : 70PB (x4.3)
- New services have been deployed
  - CVMFS stratum 0 and 1 servers
- Data transfer performance improved
  - SINET5 provides better network environments in Japan and to US and EU
  - Belle II services are well separated from other VO activities