

22nd International Conference on Computing in High Energy and Nuclear Physics, Hosted by SLAC and LBNL, Fall 2016

---

# Development of stable Grid service at the next generation system of KEKCC

---

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

Computing Research Center  
HIGH ENERGY ACCELERATOR RESEARCH ORGANIZATION, KEK



Computing Research Center

Tomoaki Nakamura, KEK-CRC

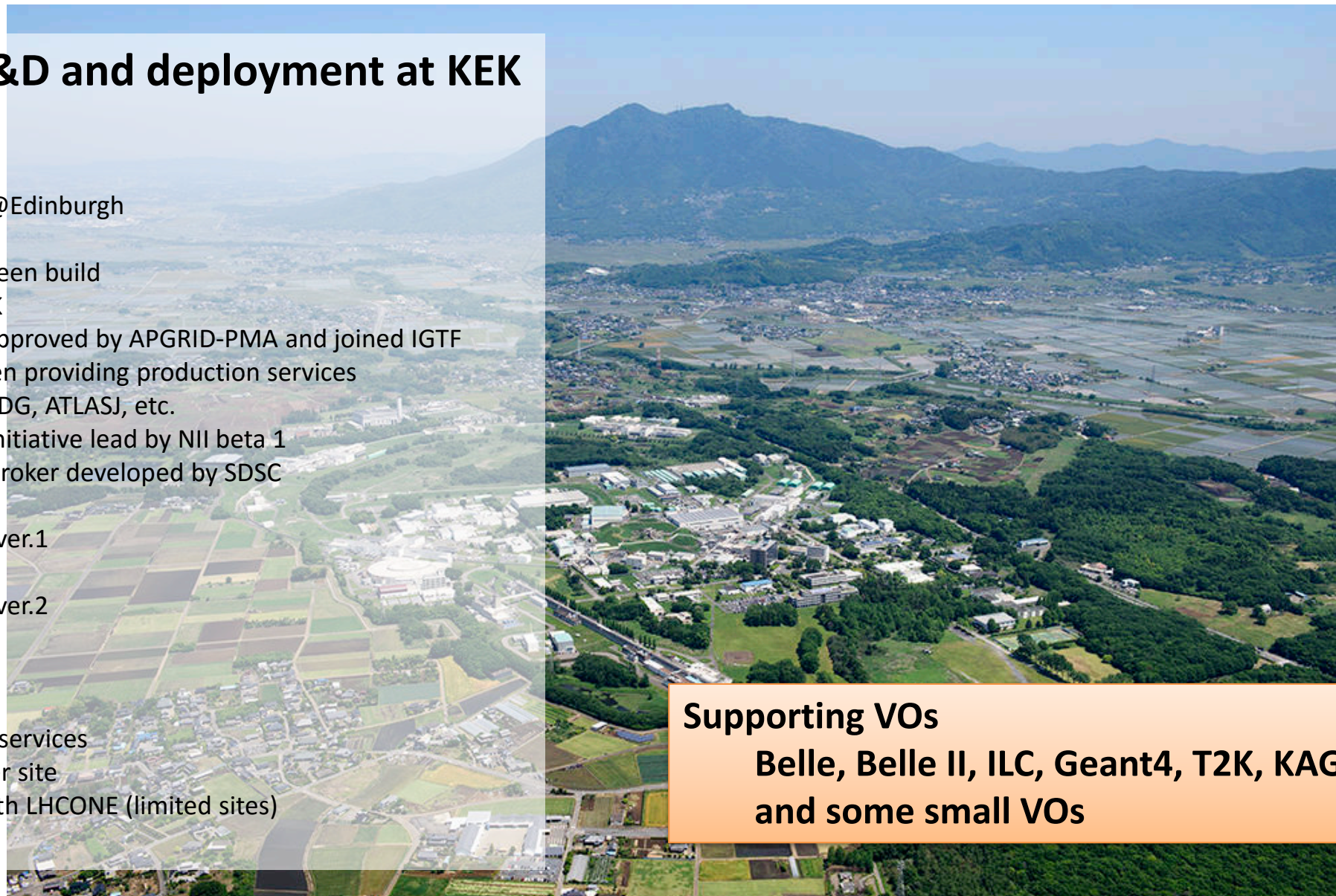


## History of the Grid R&D and deployment at KEK

- 2002 R&D on Grid with GT2  
Attend Globus tutorial @Edinburgh
- 2004 Private CA in production  
Test bed for LCG2 has been build  
HEP Data Grid WS @KEK
- 2006 KEK Grid CA has been approved by APGRID-PMA and joined IGTF  
LCG2 site of KEK has been providing production services  
VO supported: Belle, APDG, ATLASJ, etc.  
NAREGI: National Grid Initiative lead by NII beta 1  
SRB: Storage Resource Broker developed by SDSC  
Deployment of gLite 3.0
- 2007 Deployment of NAREGI ver.1
- 2009 Start iRODS service
- 2010 Deployment of NAREGI ver.2
- 2011 Deployment of EMI-1
- 2012 Deployment of EMI-2
- 2013 Deployment of UMD-3  
Termination of NAREGI services
- 2015 Joined WLCG as observer site  
Start test connection with LHCONE (limited sites)
- 2016 Full LHCONE connection

### Supporting VOs

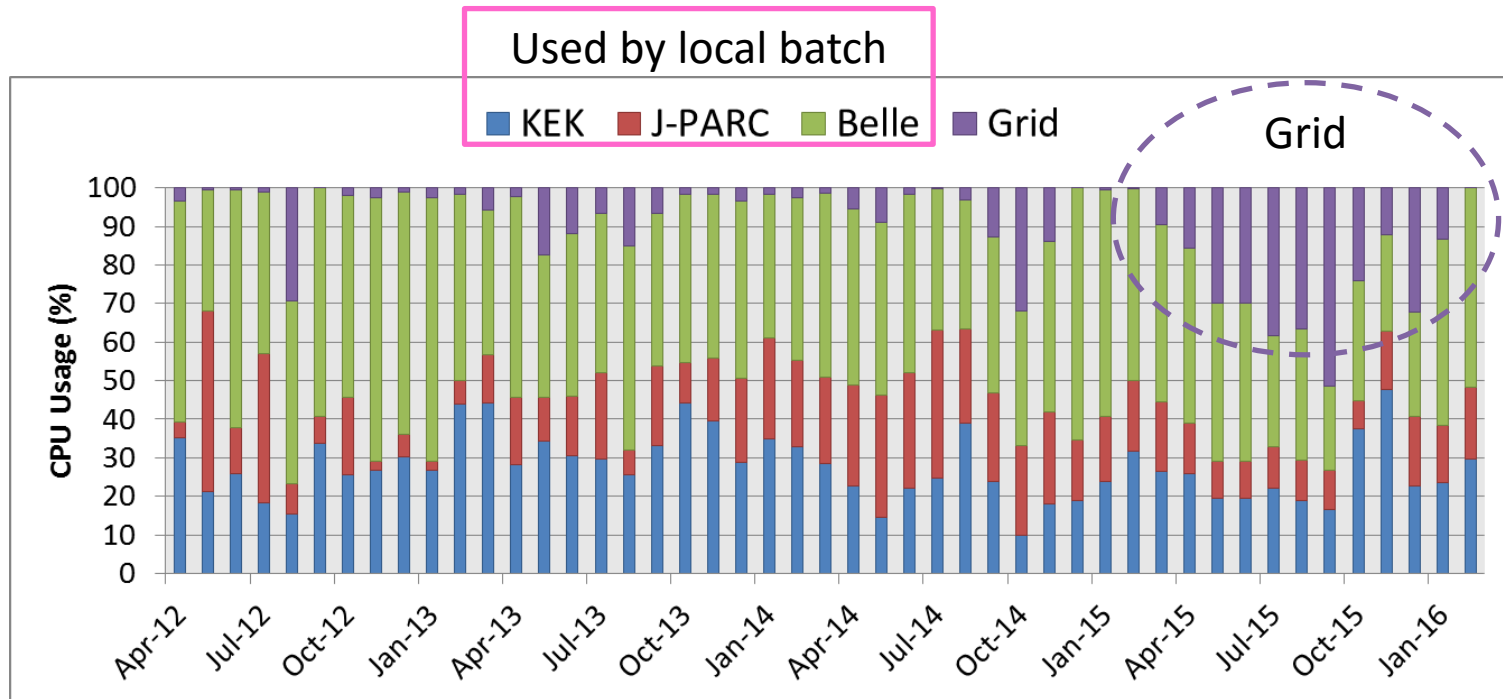
**Belle, Belle II, ILC, Geant4, T2K, KAGRA  
and some small VOs**



# Computing demand at KEK



CPU usage already reached at 94% of the total resource in KEK, used via local batch system (LSF) and also from Grid at the previous system (until Aug. 2016). The fraction of usage from Grid reached ~50% mostly coming from Belle II MC production. Apparently computing resource is not enough, Need to Upgrade.



Fraction of CPU usage, break down by groups (2012 - 2015)

## Requirement for the next 4 years

	CPU (cores)	Disk (PB)	Tape (PB)
Belle	1,000	1.2	3.5
Belle II	7,500	9	29
ILC	400	0.3	1.5
CMB	250	0.5	1
J-PARC	1,650	5.9	27
KOTO	1,000	5	15
T2K	300	0.2	1
MLF	50	0.5	8
Others (J)	300	0.2	3
<b>Total</b>	<b>10,800</b>	<b>17</b>	<b>65</b>
<b>Current Sys.</b>	<b>4,000</b>	<b>7</b>	<b>18</b>
<b>Next Sys.</b>	<b>10,000</b>	<b>13</b>	<b>70</b>

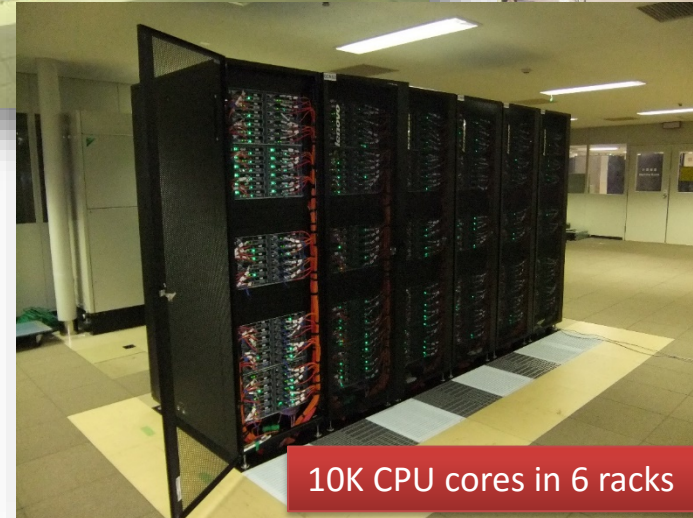
# New KEK Central Computer system (KEKCC)



## System specification

	Current	New	Upgrade Factor
CPU Server	IBM iDataPlex	Lenovo NextScale	
CPU	Xeon 5670 (2.93 GHz ,6core)	Xeon E5-2697v3 (2.6GHz, 14cores)	
CPU cores	4,000	10,000	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	250 MB/s	350 MB/s	
Tape max capacity	16 PB	70 PB	x4.3
Power Consumption	200 kW (actual monitored value)	< 400 kW (max estimation)	

## Network switch



Computing node



Disk

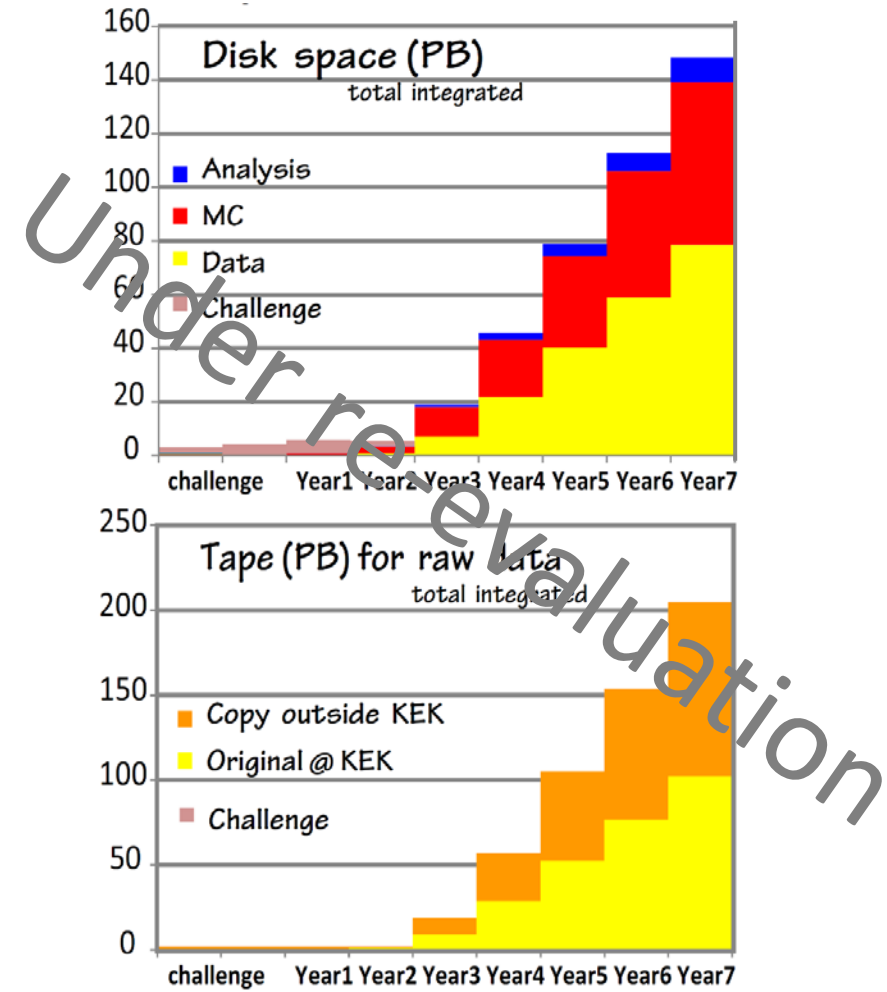
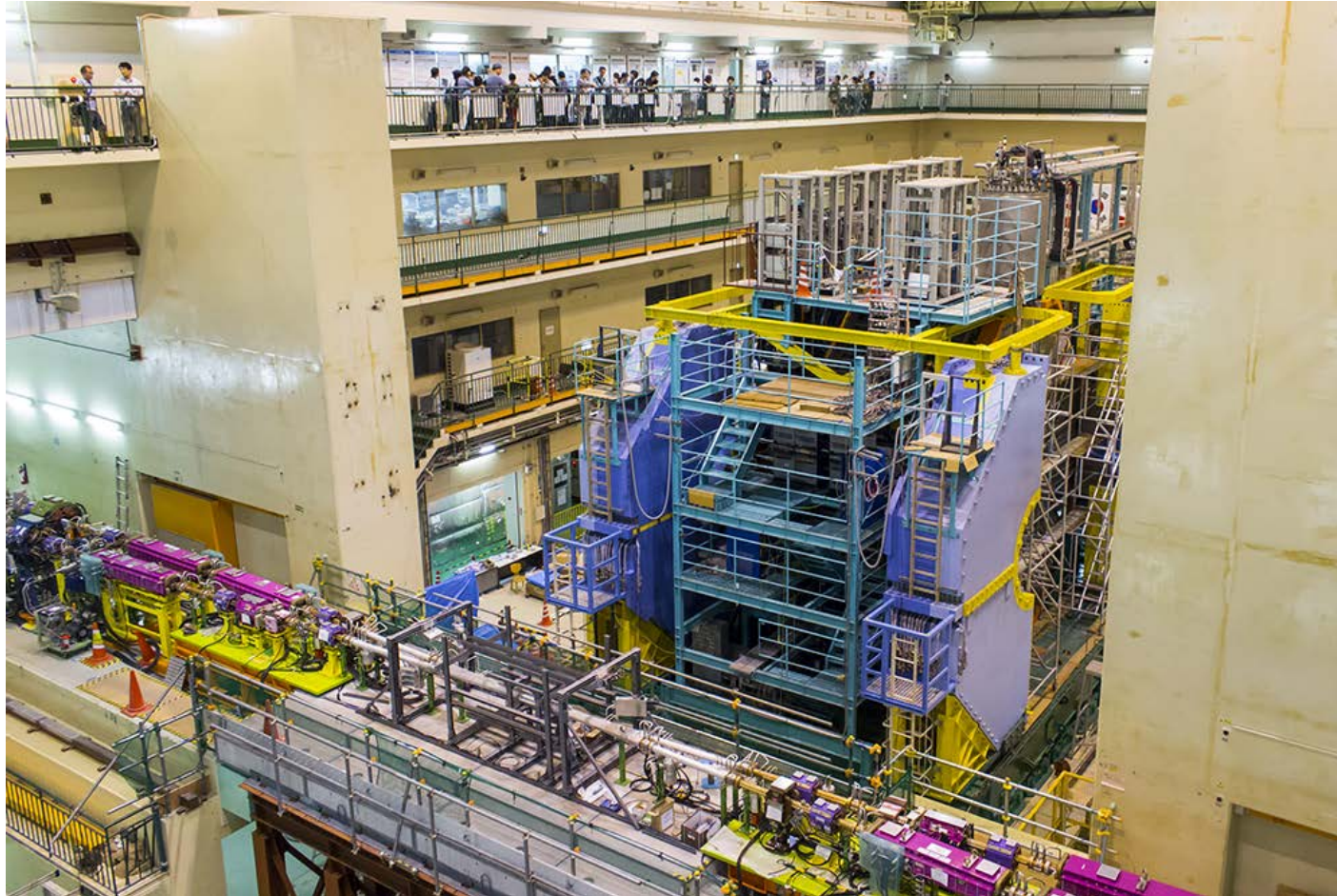
Tape

Service in production since September 1st 2016

K. Murakami et al. (CHEP2016)

<https://indico.cern.ch/event/505613/contributions/2227443/>

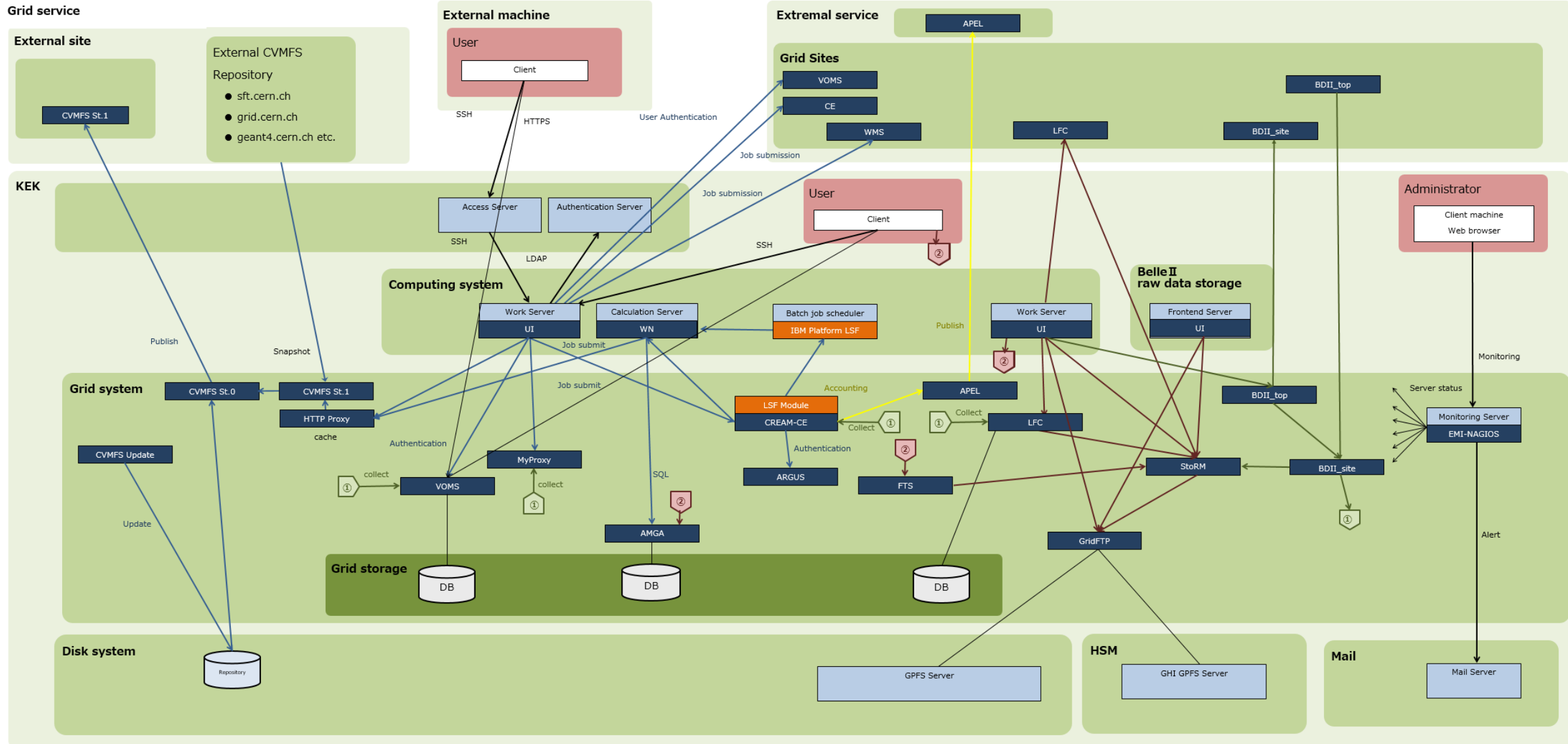
# Coming massive Belle II data



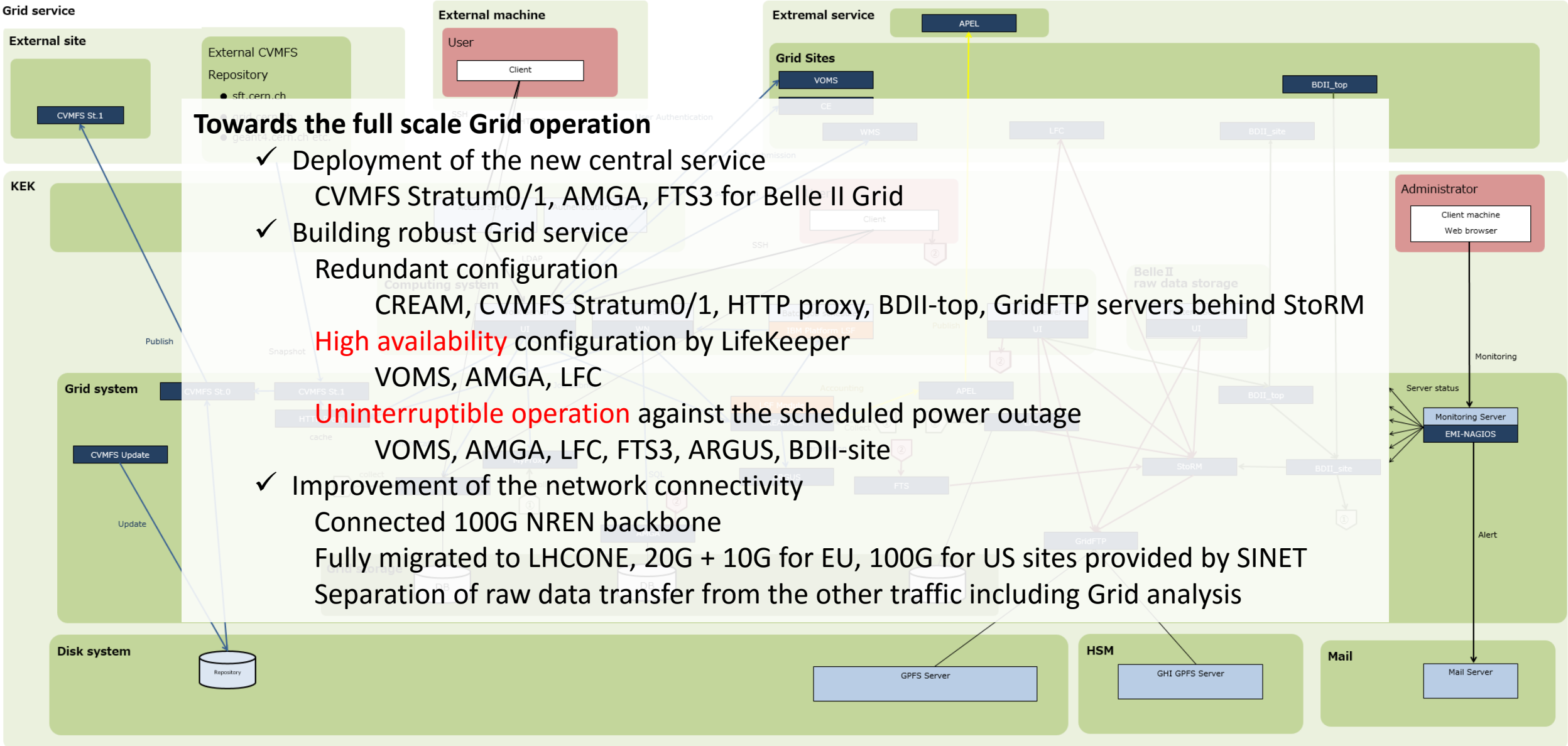
T. Hara et al. (CHEP2016)

<https://indico.cern.ch/event/505613/contributions/2228504/>

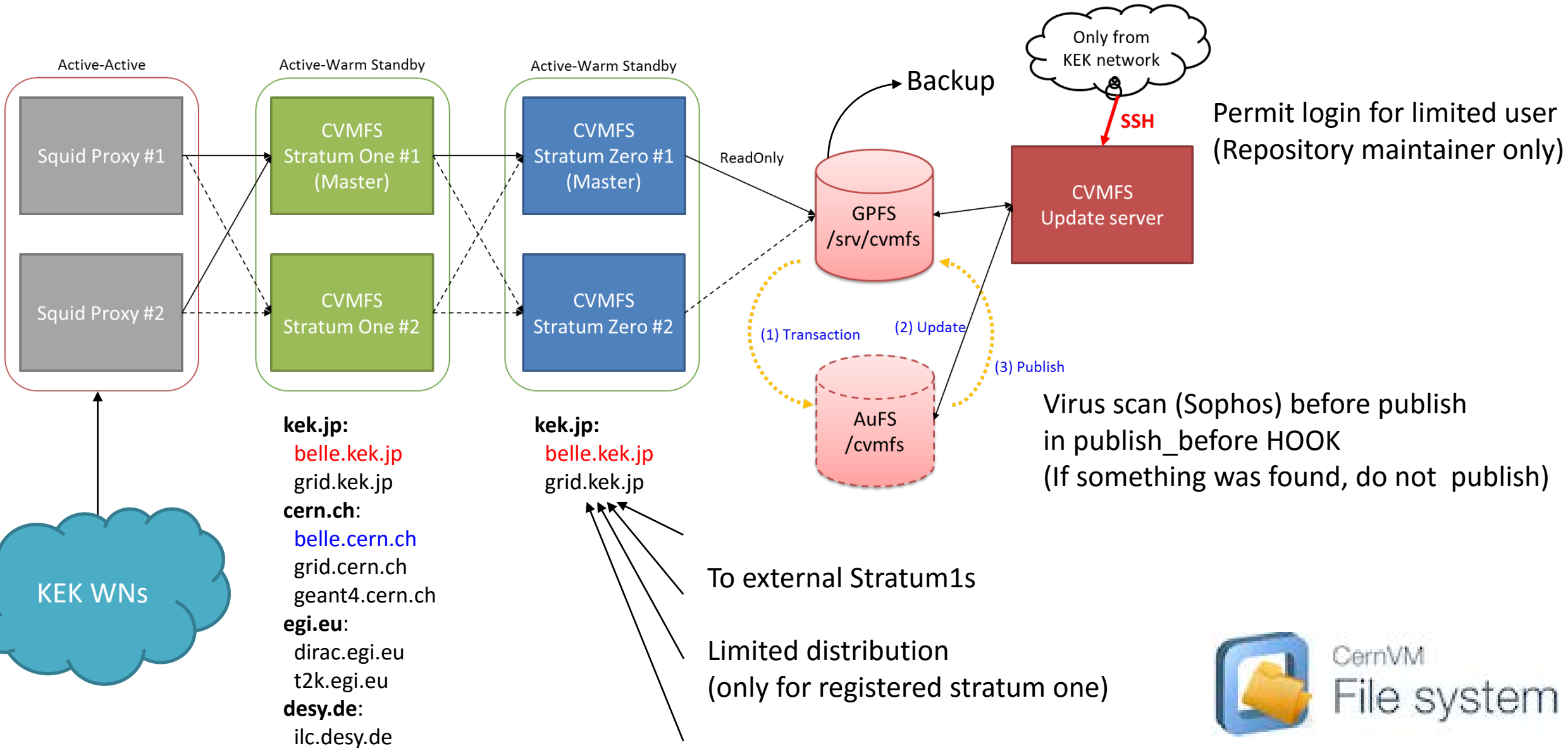
# Overview of the new Grid system



# Overview of the new Grid system



# Deployment of new central service (CVMFS)



# Building robust Grid service



LifeKeeper

LifeKeeper

VOMS #1

VOMS #2

DB

AMGA #1

AMGA #2

DB

Belle II, KAGRA etc...

Belle II

Critical service for the other sites in Belle II Grid.  
In case of failure, switch without service stop.

Update LFC

LifeKeeper

LFC #1

LFC #2

DB

Dedicated to Belle II

Read only without  
GSI authentication

Active-Active

RO LFC #1

DB (SSD)

RO LFC #2

DB (SSD)

replication

**No interference between  
Belle II and the other VOs**

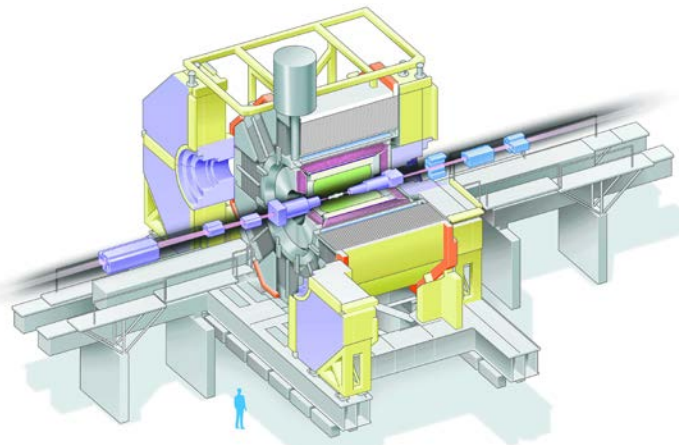
Update LFC

LFC

DB (SSD)

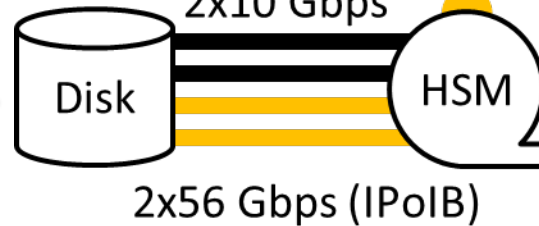
For the other VOs, e.g. ILC etc.

# Reinforcement of data transfer

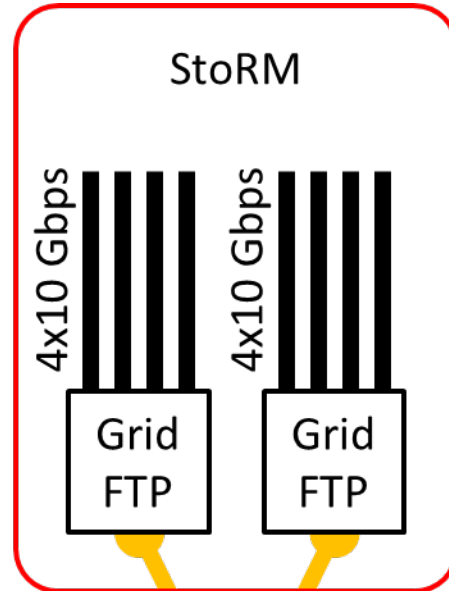


Online storage

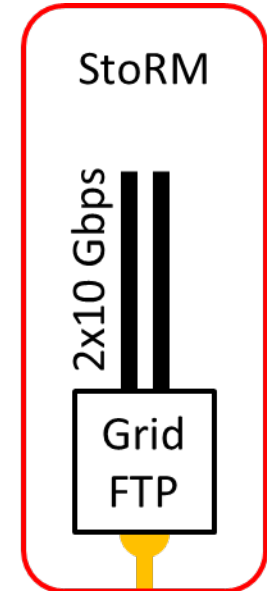
~3GB/s



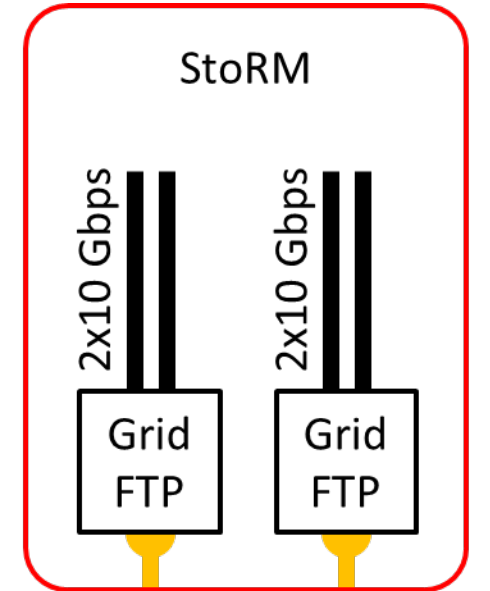
Belle II raw data



Belle II analysis



Other VOs

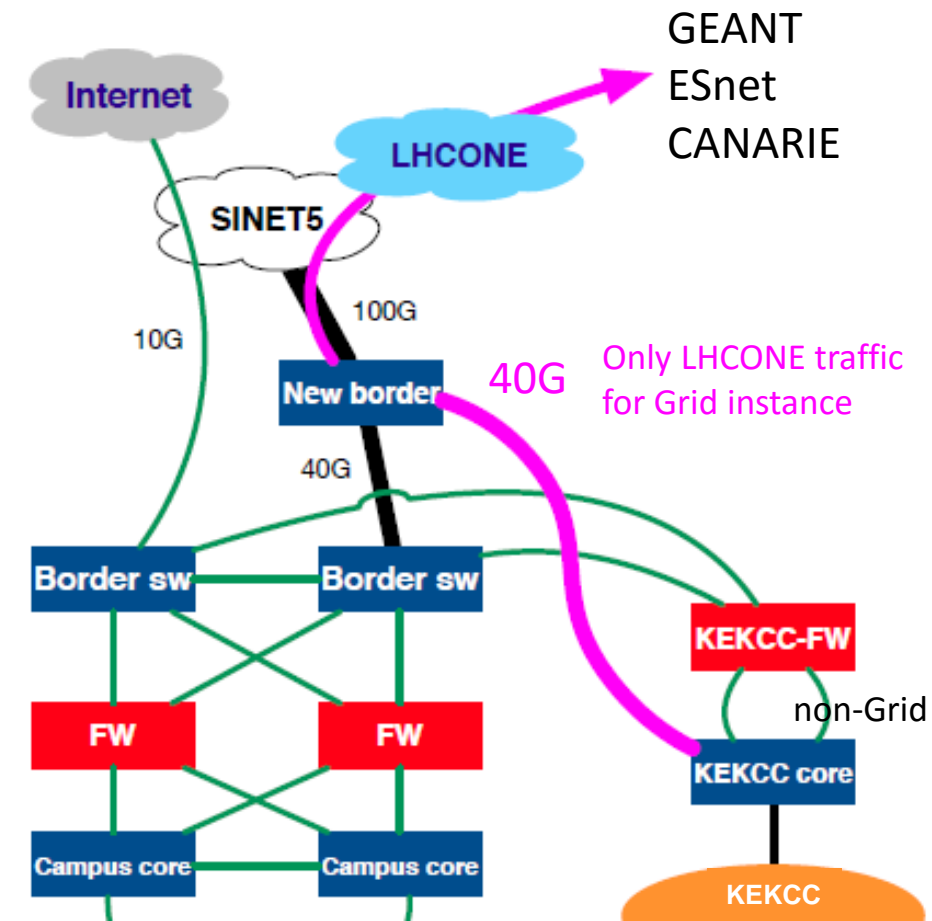
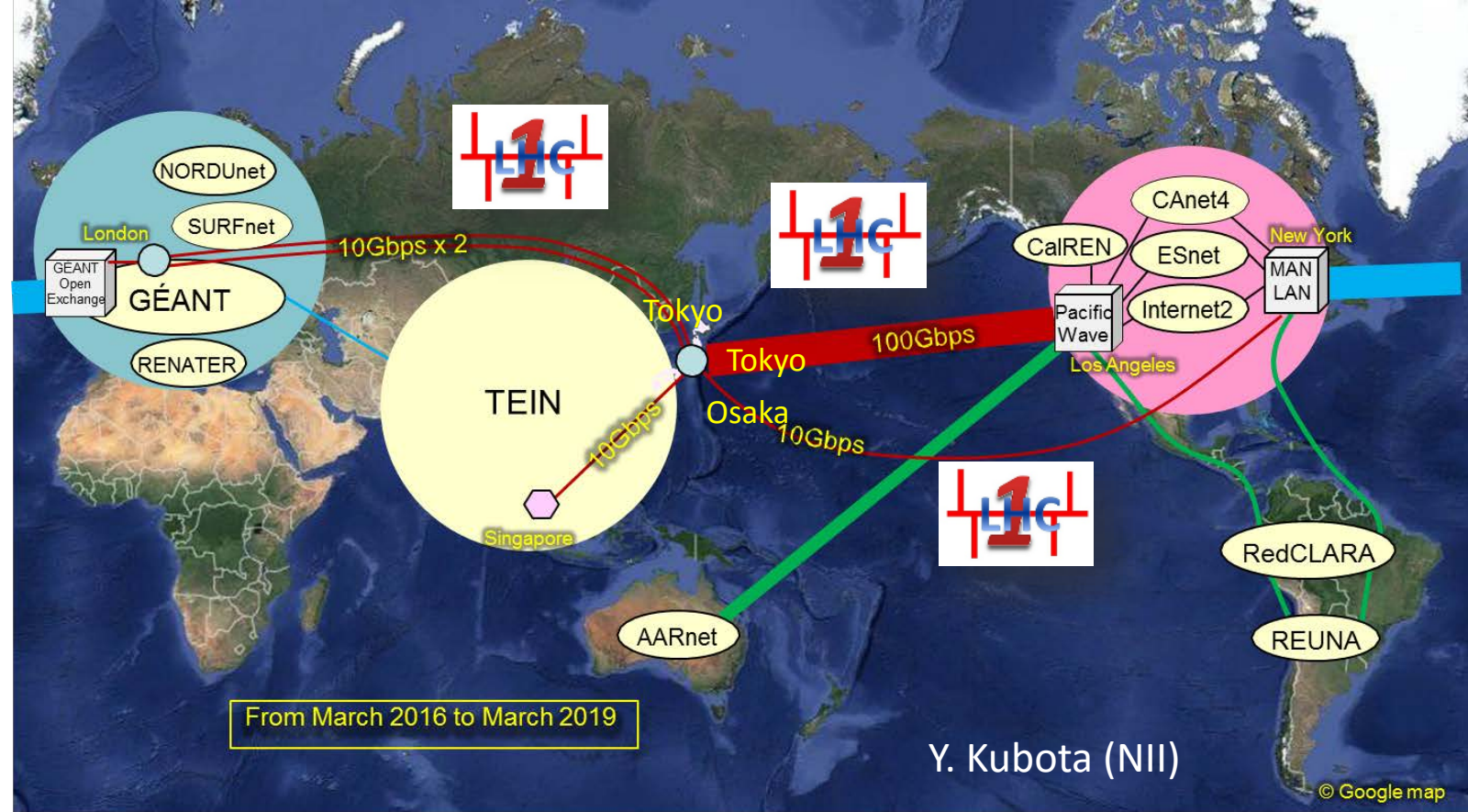


Complete separation of Belle II raw data transferring path from analysis and the other VOs activity.

## Total throughput

HSM: 50GB/s (IBM GPFS+HPSS on DDN SFA12K)  
Disk: 100GB/s (IBM GPFS on IBM ESS)

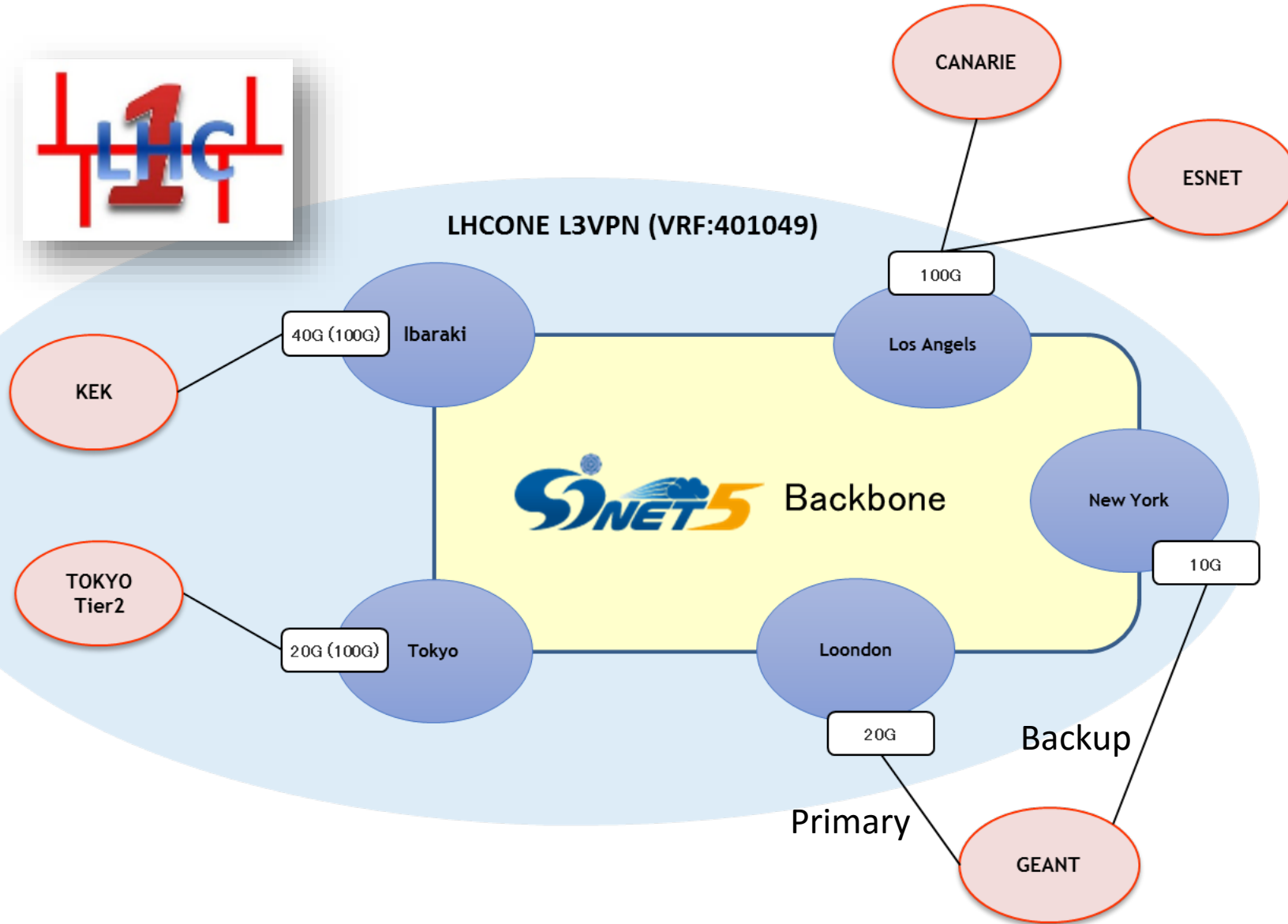
# Upgrade of network connectivity



SINET5 (NII) provides 100G+10G to US and 2x10G for EU since Mar. 2016.  
 LHCONE peering with GEANT, ESnet and CANARIE have been started Sep. 2016.

Policy routing at KEKCC core switch  
 Bypassing FW for LHCONE traffic

# Further extension of LHCONE connection



Now full migration was completed and then,

- ✓ LHCONE connection with Asian sites (Taiwan, Korea, Hon Kong etc.)
- ✓ LHCONE backup of trans-pacific connection (TransPAC-Pacific Wave 100G, Seattle)
- ✓ Upgrade bandwidth for London line
- ✓ IPv6 on LHCONE



**The new Grid service at KEK is ready for massive production with the launch of new KEK Central Computer System (KEKCC) at September 1st, 2016.**

## **Service level improvement:**

Many kinds of the central services are newly introduced by **High Availability Configuration** to achieve **Uninterruptible Operation** also in terms of the electric power cut for the facility maintenance, e.g. CVMFS Stratum0/1, VOMS, LFC, AMGA and FTS3 dedicated to Belle II Grid.

## **Performance improvement:**

Data transfer performance is upgraded significantly by the high bandwidth internal network and powerful GridFTP servers. Belle II raw data transfer to the other sites is not affected by any other activities at KEK. We expect the smooth data transfer to the other sites with the LHCONE routing.