KEK Grid System Status and Future

Go Iwai

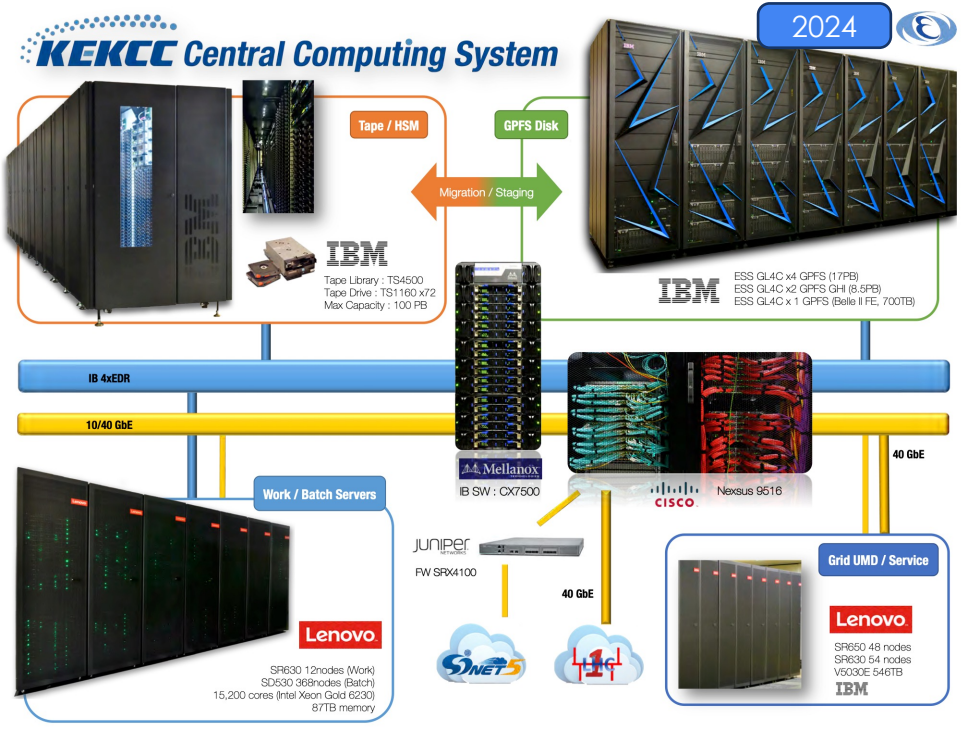
High Energy Accelerator Research Organization (KEK)

Computing Research Centre (CRC)

KEKCC – A Largest Scale Computer System

1.5 Years of Production Experience

KEKCC: Entering the 2nd Year of the 4-year Contract



Grid instances are running in KEKCC

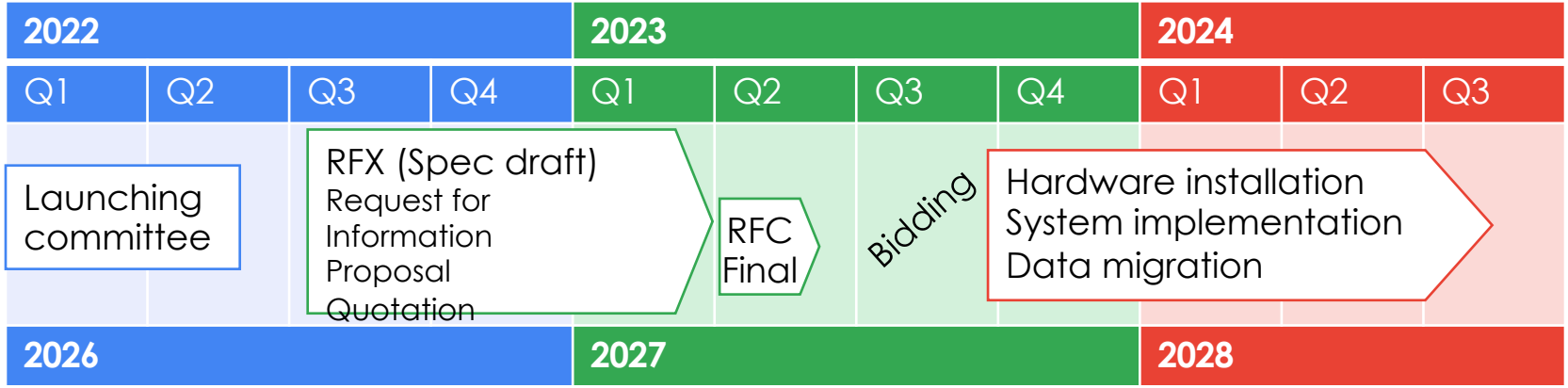
Stable production with ~90% utilisation

- Linux Cluster + Storage System (GPFS/HSM)
  - Accommodate multiple experiments in a single system
- In production since September 2024

- CPU: 12,096 cores
  - AMD EPYC 9654, 2.4 GHz (Base clock, up to 3.7 GHz)
  - 192 cores/node (2 CPU/node, 96 cores/node)
  - 63 calculation nodes
  - 382K HS23 (31.56 HS23/core)
    - 6K HS23/node @3.5 GHz Measured w/o SMT
    - LINPACK: R_{max}: 0.52 / R_{peak}: 0.45 (PFlop/s)
- Memory: 56 TB
  - 4 GB/job
- Disk: 30 PB
  - 20 PB: GPFS for experimental groups
    - 5.4 PB for Belle II, and 2.4 PB for Belle
  - 10 PB: GPFS-HPSS-Interface (GHI) as an HSM cache
- Tape: 120 PB as maximum capacity

Procurement

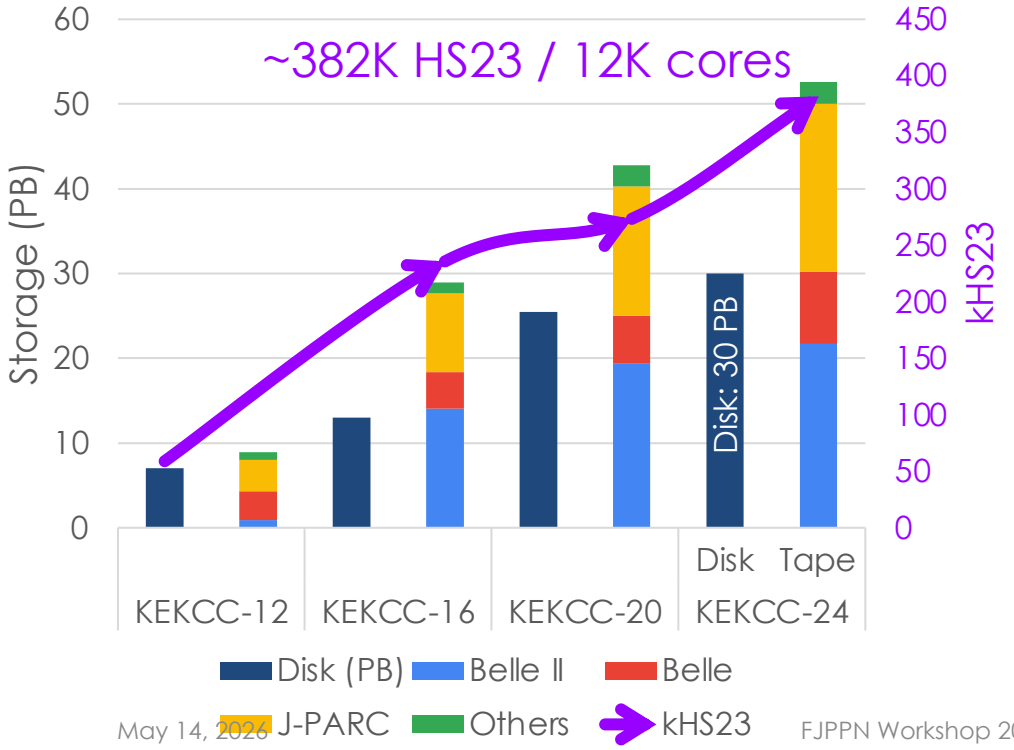
- A multiple-year rental system contract: KEKCC is entirely replaced every 4-5 years.
  - KEKCC has started in September 2024 and will end in August 2028
  - Need to migrate data, service, configuration, everything else., from the old system to the new one
  - Completely different purchase/operation model from EU/US sites: NOT in-house scale-out model, BUT rental system not only for hardware but also including installation, operation, uninstallation, and labor costs for almost everything
- Bidding process: 1.5 years in usual, but 2+ years last time
  - Unusually long duration due to the uncertain delivery time, bad currency rate to US$, and rising electricity costs
  - Launch the committee in early 2026 (delayed already) toward the next system introduced in 2028
    - Situation has not significantly changed

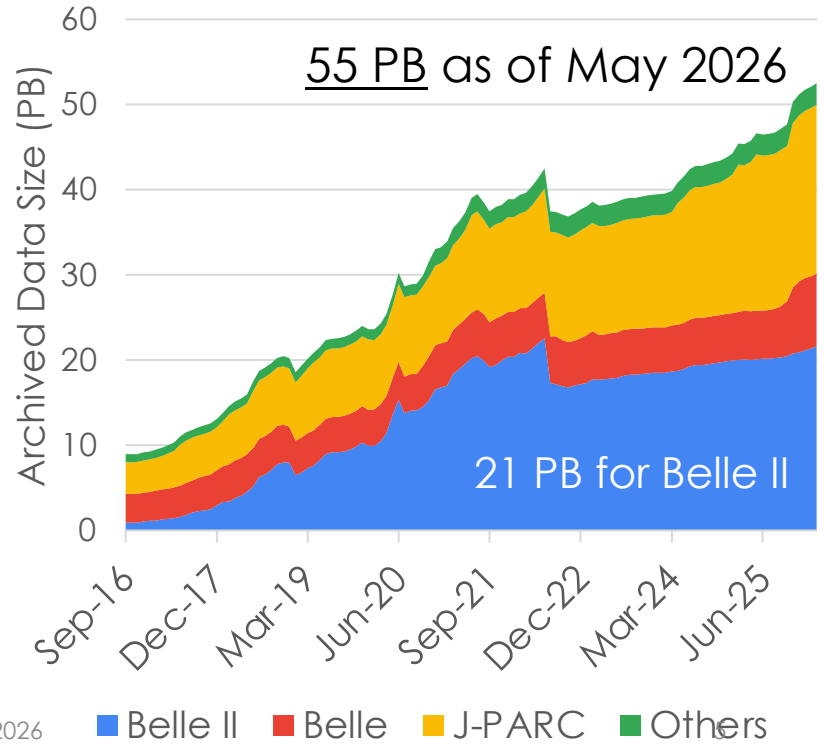


Expected timeline for the next system 2028

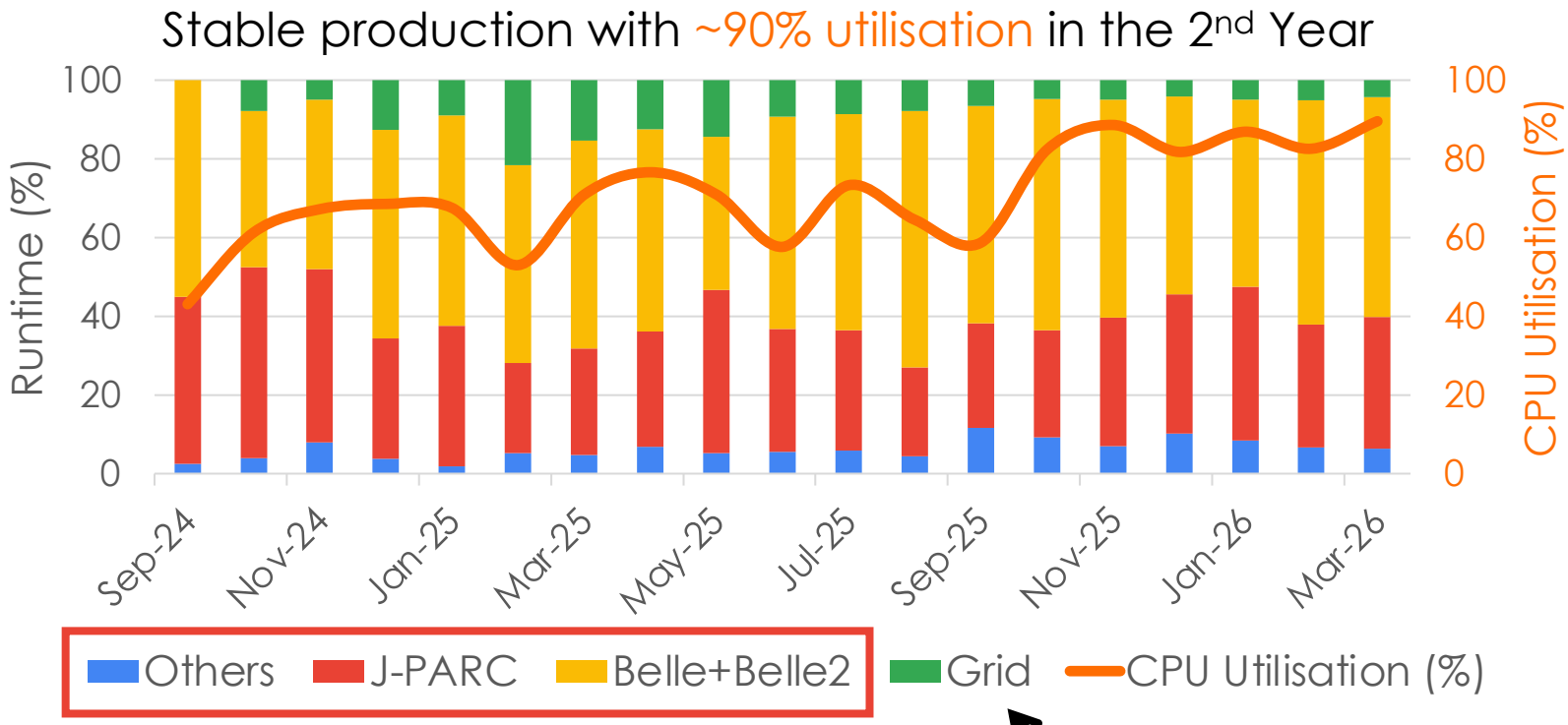
Site Scale Evolution

Resource History (Last 4-Gen)



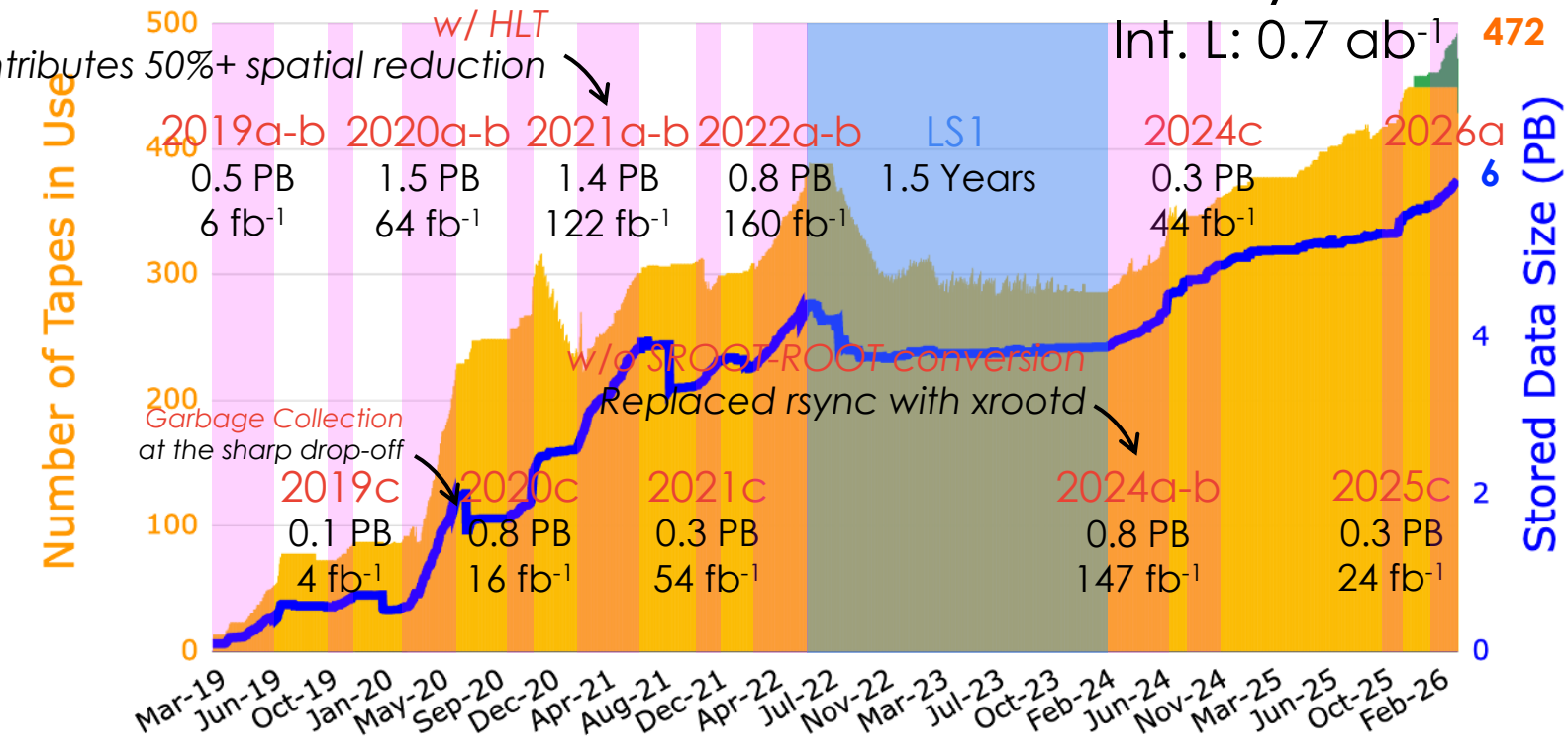


Runtime in the Entire System



* Belle2 RAW data centres: BNL, UVic, CNAF, DESY, KIT, and CCIN2P3

6 PB of Belle2 RAW data for 0.7 ab^{-1} of total int. luminosity



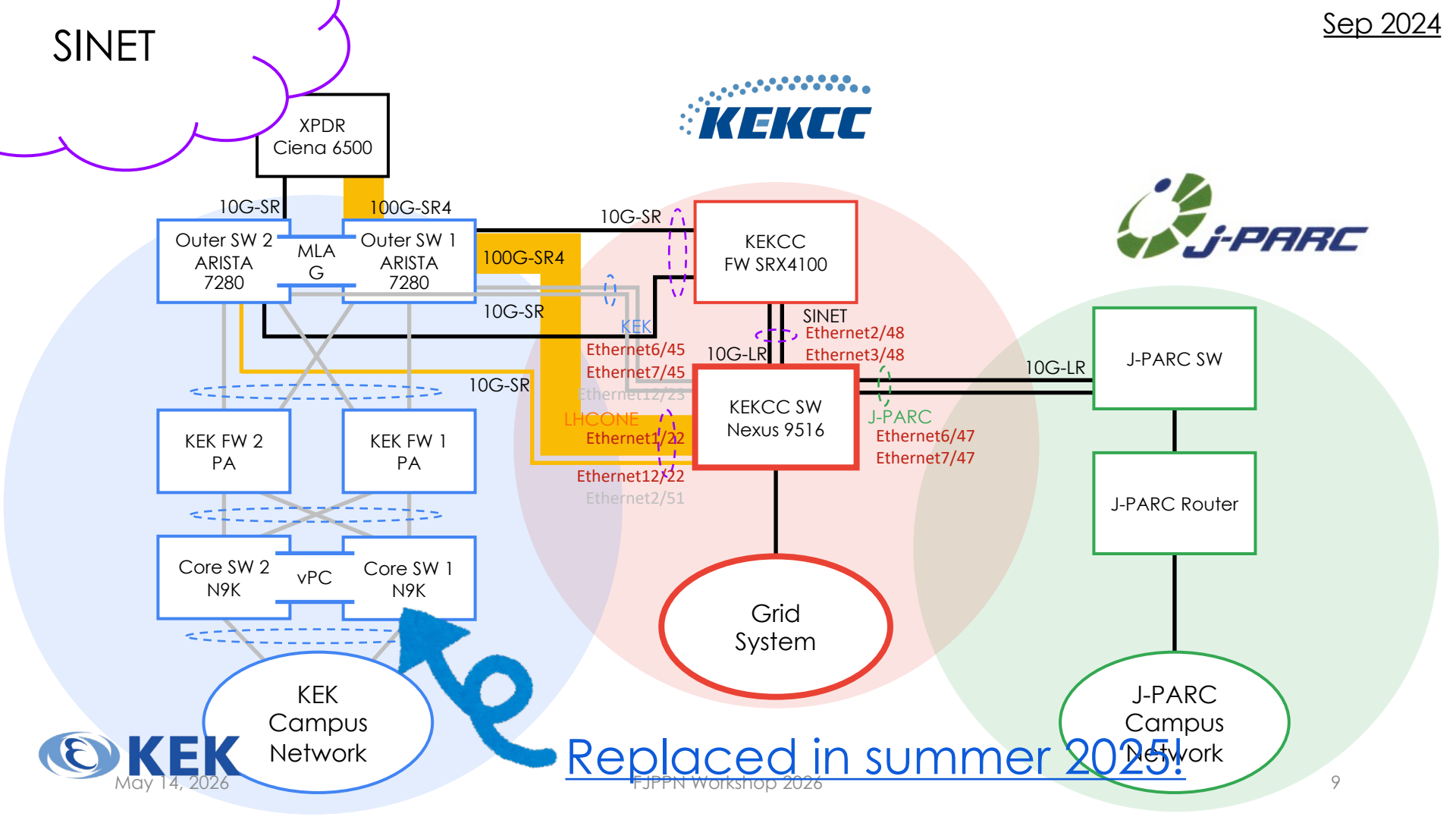
Networking

Network hardware replacement in summer 2025

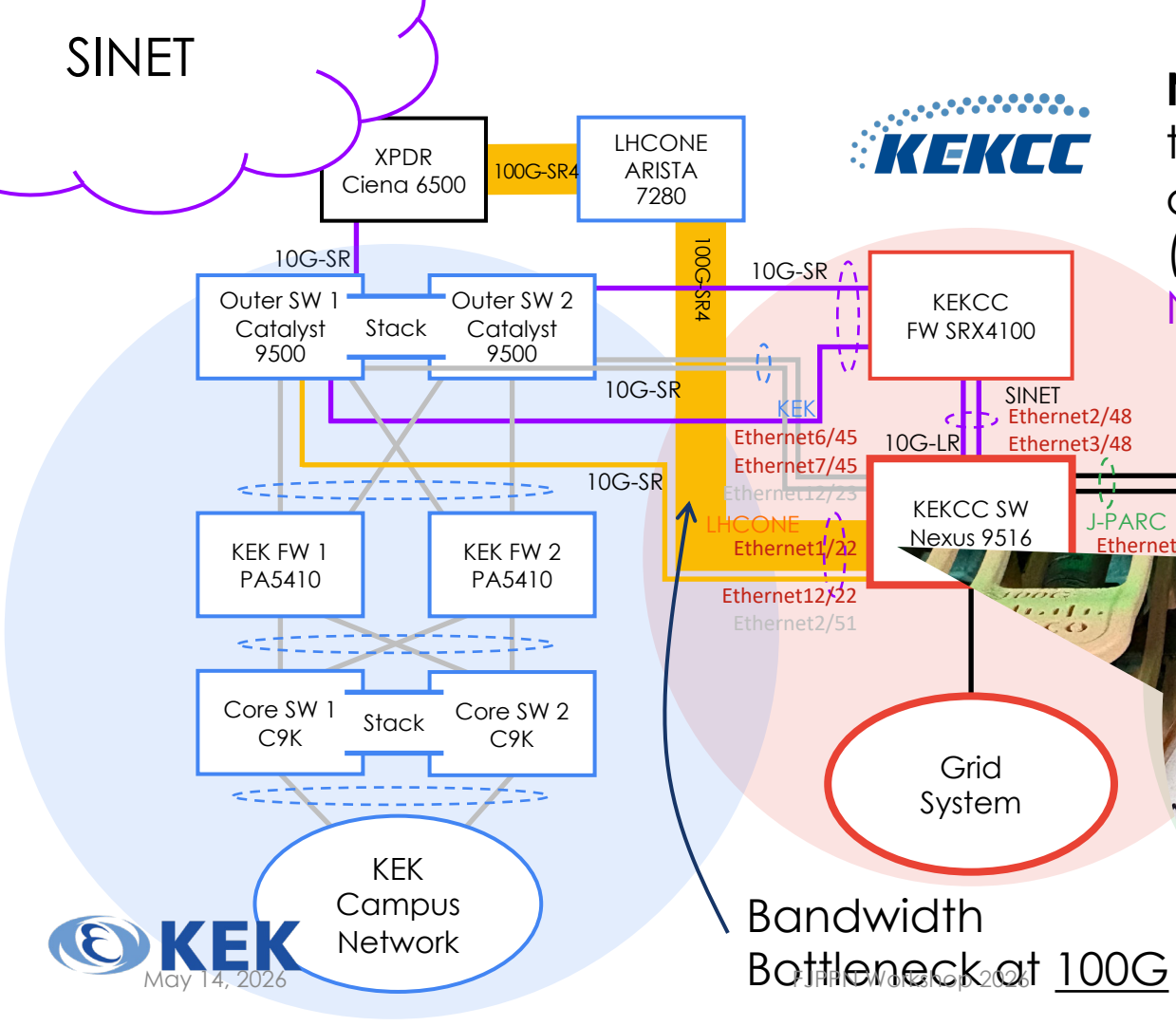
International network backbone

SINET

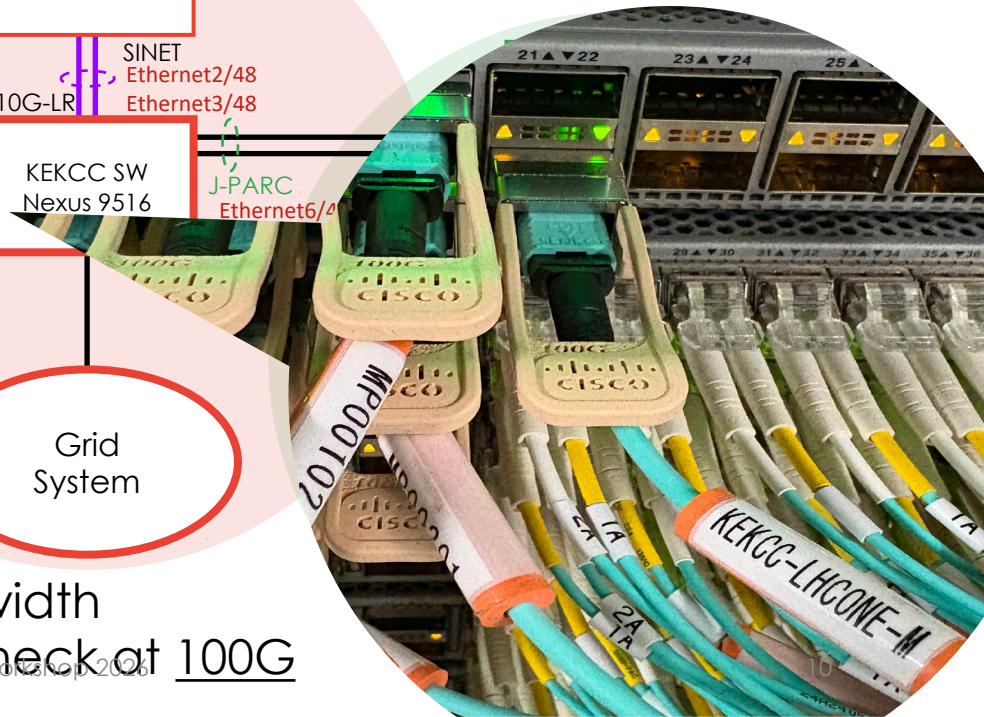
Sep 2024



SINET

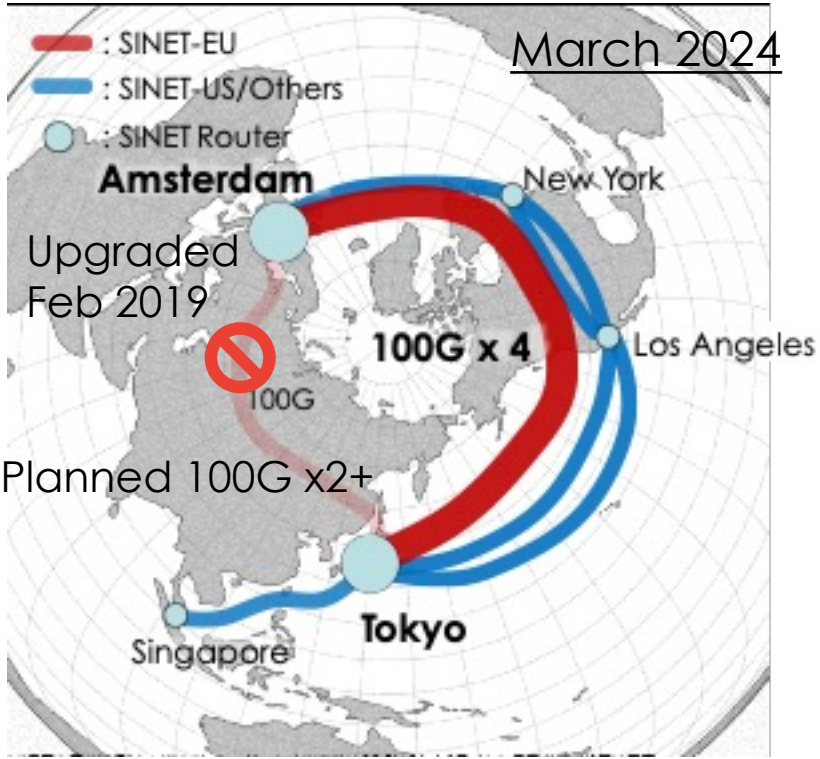


No significant difference in total bandwidth before and after hardware replacement (100G for LHCONE, 10G for Non-LHCONE traffic).

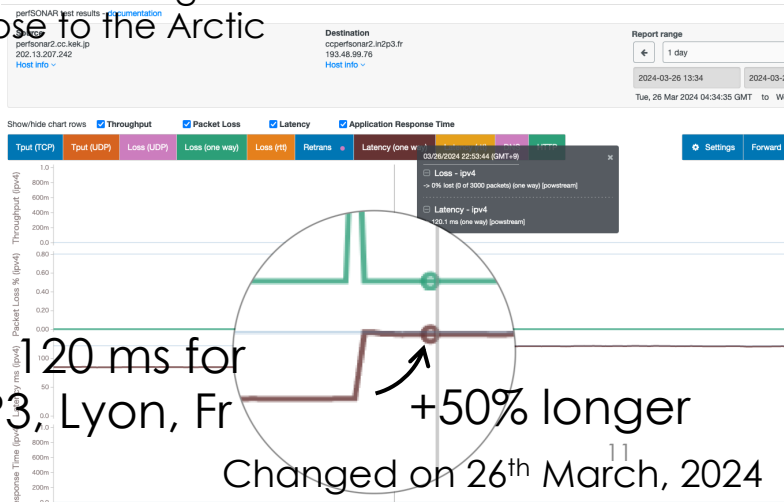


Bandwidth Bottleneck at 100G

International Network Route Transatlantic Back Again



- Siberian 100G route for Euro has been upgraded to the transatlantic route with 100G x4 lines
- Dedicated line for the traffic between Japan and Euro, not shared with traffic for the US
- To minimise the latency:
  - traffic passes through fewer routers on the shortest route close to the Arctic



84 ms to 120 ms for CC-IN2P3, Lyon, Fr

+50% longer

Changed on 26th March, 2024

Grid Service Deployment Status

OS migration RHEL7 to RHEL9

Data transfer test

Service Deployment as of Sep 2024

RHEL7 with ELS running a lot

as Belle2 dedicated

Service	OS	VM/Bare metal	Ethernet	IPv6	HA	UPS
StoRM	RHEL7 + ELS	Bare metal	10GE	✅	✅	
VOMS	RHEL7 + ELS	VM	10GE	✅	✅ SIOS LifeKeeper™	✅
IAM	RHEL9	Bare metal	10GE	✅	✅	✅
AMGA	RHEL7 + ELS	Bare metal	10GE	✅	✅ SIOS LifeKeeper™	✅
Top BDII	RHEL9	VM	10GE	✅	✅	✅
Site BDII	RHEL9	VM	10GE	✅	✅	✅
FTS3	RHEL9	VM	10GE	✅	✅	✅
ARC-CE	RHEL7 + ELS	Bare metal	10GE	✅	✅	
CVMFS Stratum Zero	RHEL9	Bare metal	10GE	✅	✅	
CVMFS Stratum One	RHEL7 + ELS	Bare metal	10GE	✅	✅	
CVMFS publisher	RHEL9	VM	10GE	✅		
Frontier Squid HTTP Proxy	RHEL9	VM	10GE	✅	✅	✅

Decommissioning GridFTP, then SRM
As of March 2025

as Belle2 dedicated

Turned off GridFTP in January, SRM in March 2025

Service	OS	VM/Bare metal	Ethernet	IPv6	HA	UPS
StoRM	RHEL7 + ELS (WIP)	Bare metal	10GE	✅	✅	
VOMS	RHEL7 + ELS	VM	10GE	✅	✅ SIOS LifeKeeper™	✅
IAM	RHEL9	Bare metal	10GE	✅	✅	✅
AMGA	RHEL7 + ELS	Bare metal	10GE	✅	✅ SIOS LifeKeeper™	✅
Top BDII	RHEL9	VM	10GE	✅	✅	✅
Site BDII	RHEL9	VM	10GE	✅	✅	✅
FTS3	RHEL9	VM	10GE	✅	✅	✅
ARC-CE	RHEL7 + ELS	Bare metal	10GE	✅	✅	
CVMFS Stratum Zero	RHEL9	Bare metal	10GE	✅	✅	
CVMFS Stratum One	RHEL9 (Mar '25)	Bare metal	10GE	✅	✅	
CVMFS publisher	RHEL9	VM	10GE	✅		
Frontier Squid HTTP Proxy	RHEL9 (Minor fix)	VM	10GE	✅	✅	✅

Ongoing migration campaign to RHEL9
As of September 2025

as Belle2 dedicated

Migrated to SciTags-enabled DTNs on RHEL9

Service	OS	VM/Bare metal	Ethernet	IPv6	HA	UPS
StoRM	RHEL9	Bare metal	10GE	✅	✅	
VOMS	RHEL7 + ELS	VM	10GE	✅	✅ SIOS LifeKeeper™	✅
IAM	RHEL9	Bare metal	10GE	✅	✅	✅
AMGA	RHEL7 + ELS	Bare metal	10GE	✅	✅ SIOS LifeKeeper™	✅
Top BDII	RHEL9	VM	10GE	✅	✅	✅
Site BDII	RHEL9	VM	10GE	✅	✅	✅
FTS3	RHEL9	VM	10GE	✅	✅	✅
ARC-CE	RHEL7 + ELS	Bare metal	10GE	✅	✅	
CVMFS Stratum Zero	RHEL9	Bare metal	10GE	✅	✅	
CVMFS Stratum One	RHEL9	Bare metal	10GE	✅	✅	
CVMFS publisher	RHEL9	VM	10GE	✅		
Frontier Squid HTTP Proxy	RHEL9	VM	10GE	✅	✅	✅

Ongoing migration campaign to RHEL9
As of May 2026

as Belle2 dedicated

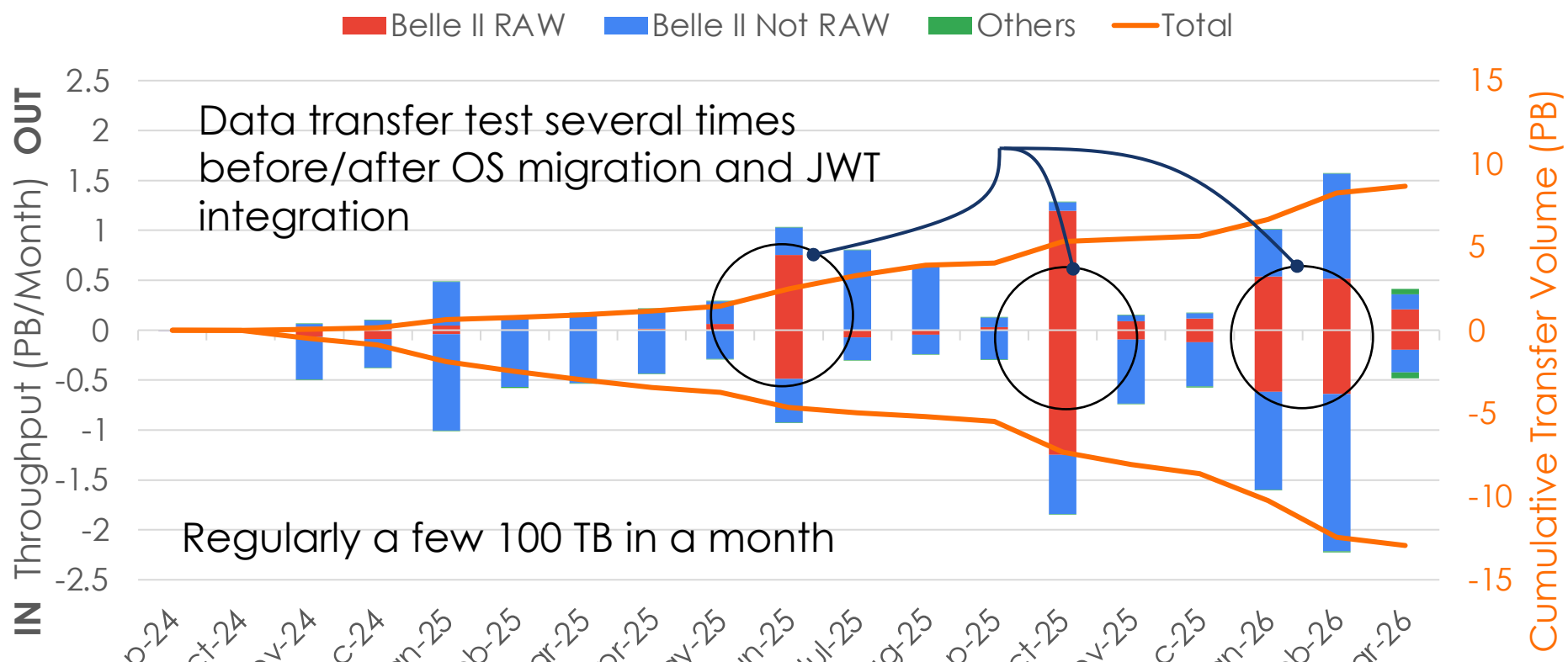
Service	OS	VM/Bare metal	Ethernet	IPv6	HA	UPS
StoRM	RHEL9	Bare metal	10GE	✅	✅	
VOMS	RHEL7 + ELS	VM	10GE	✅	✅ SIOS LifeKeeper	✅
IAM	RHEL9	Bare metal	10GE	✅	✅	✅
AMGA	RHEL7 + ELS	Bare metal	10GE	✅	✅ SIOS LifeKeeper	✅
Top BDII	RHEL9	VM	10GE	✅	✅	✅
Site BDII	RHEL9	VM	10GE	✅	✅	✅
FTS3	RHEL9 (v3.14)	VM	10GE	✅	✅	✅
ARC-CE	RHEL7 + ELS	Bare metal	10GE	✅	✅	
CVMFS Stratum Zero	RHEL9	Bare metal	10GE	✅	✅	
CVMFS Stratum One	RHEL9	Bare metal	10GE	✅	✅	
CVMFS publisher	RHEL9	VM	10GE	✅		
Frontier Squid HTTP Proxy	RHEL9	VM	10GE	✅	✅	✅

Completed service decommission (Mar 2026)

Upgrade to support unmanaged token (Jan 2026)

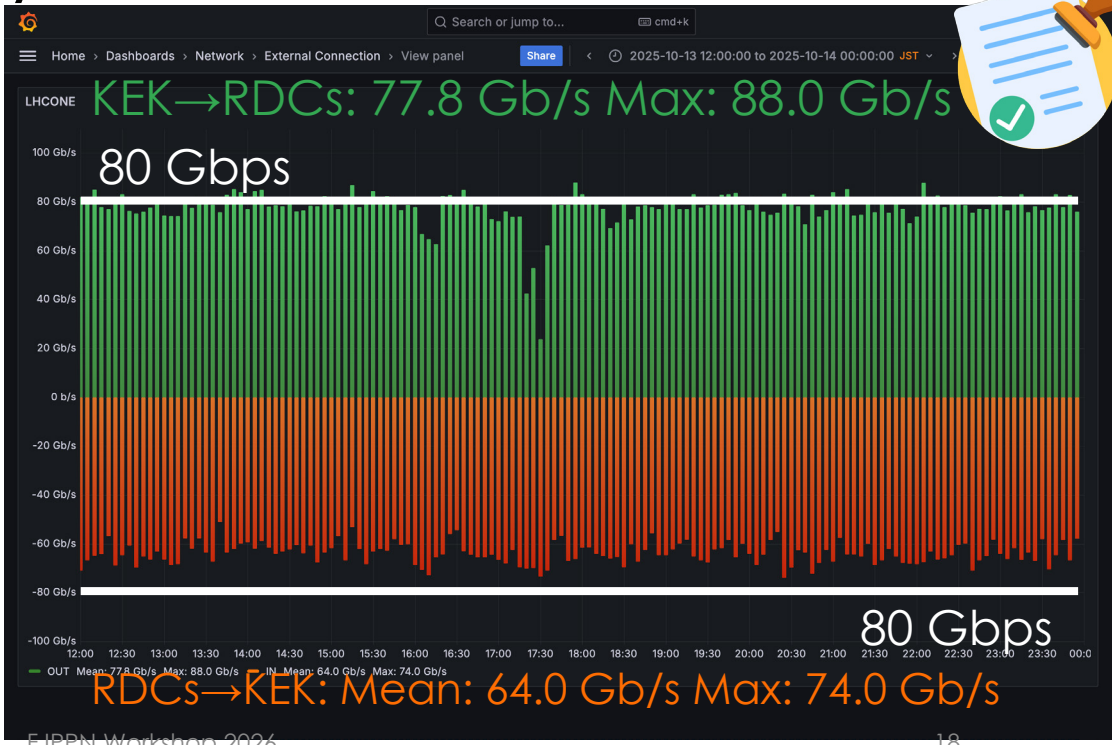
Working on migration to token-supported ARC7
Validation phase by experimental groups

Transfer Volume from/to StoRM (Not Including Internal Data Transfer)



Half a day data transfer test between KEK and Belle2 RDCs (BNL, UVic, CNAF, DESY, KIT, and CCIN2P3)

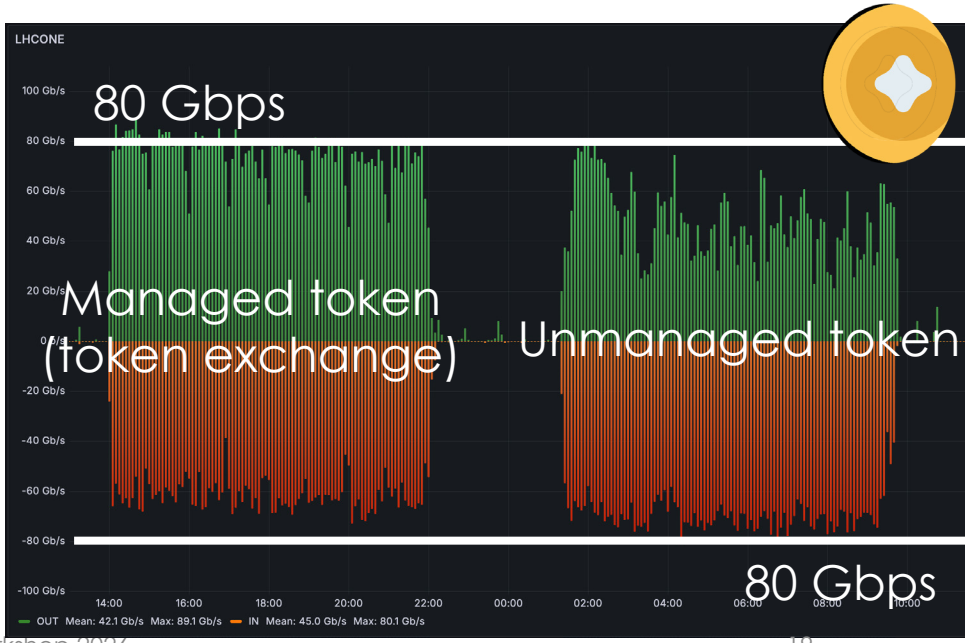
- Average throughput: 80 Gbps in both directions
- The bottleneck at 100G
- Most likely saturated at FTS optimisation limit (could be improved)
- Verified: No performance degradation observed with JWT authentication compared to X.509



Another Result: JWT version of data transfer test

- FTS dashboard:
  - KEK→B2RDCs: https://monit-grafana.cern.ch/goto/Q59XkjDDR?orgId=25
  - B2RDCs→KEK: https://monit-grafana.cern.ch/goto/Fyl3kCvvR?orgId=25
- Throughput: https://grafana.cc.kek.jp/goto/OhcXkjDDg?orgId=1

Confirmed performance parity between X.509 and JWT AuthNZ



Authentication and Authorisation

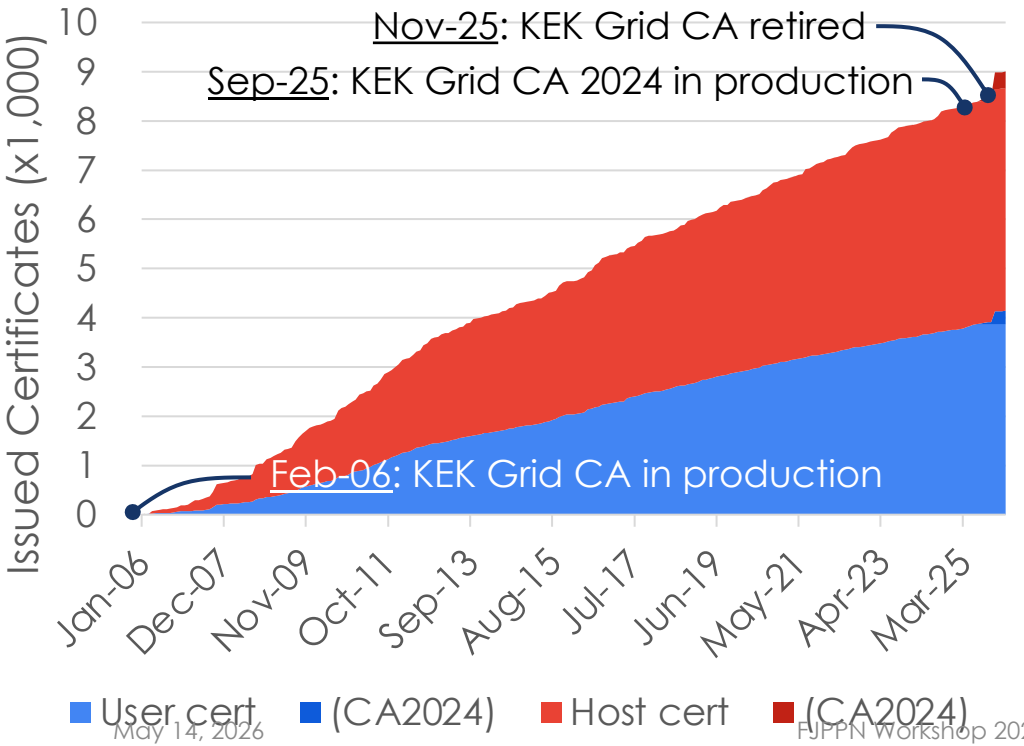
New CA is in production

Progress in Token-based AuthNZ Integration

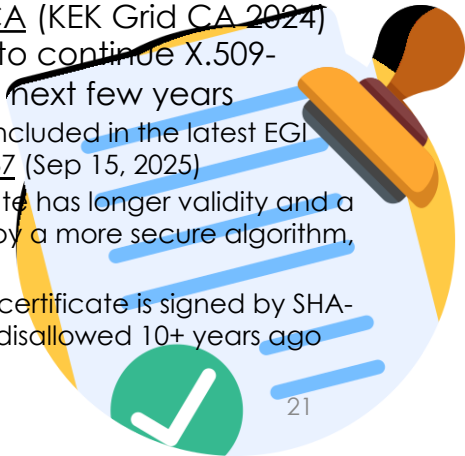
20+ years operation

Retirement of KEK Grid CA

4.1K user certs
4.8K host certs



- Initially only for domestic users, then extend the service for specific experiment members outside Japan, e.g.: Belle2
- CA's root certificate expired in November 2025
- Expected to migrate from X.509 to JWT authentication by the date – Unfortunately not!
- Launched a new CA (KEK Grid CA 2024) in **September** 2025 to continue X.509-based auth. for the next few years
  - Accredited and included in the latest EGI Trust Anchor v1.137 (Sep 15, 2025)
  - New root certificate has longer validity and a signature signed by a more secure algorithm, i.e.: SHA-2
  - The previous root certificate is signed by SHA-1, which NIST has disallowed 10+ years ago



A Long Journey: Migrating X.509/VOMS to JWT/IAM

2025	Q1	
	Q2	

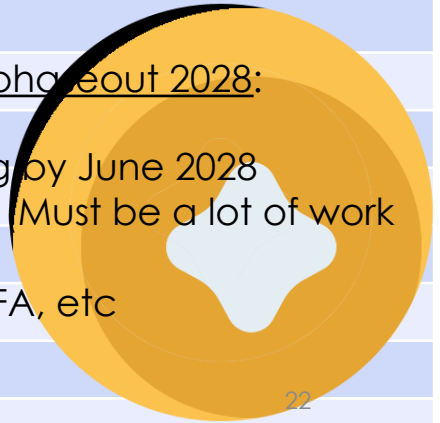
	Q3	Dual operation with VOMS and IAM VOMS Attribute Authority (AA)
		`/cvmfs/grid.cern.ch/etc/grid-security/vomses/belle-belle-auth.cc.kek.jp`
		`/cvmfs/grid.cern.ch/etc/grid-security/vomses/belle-voms.cc.kek.jp`

	Q4	JWT-supported StoRM
2026	Q1	FTS v3.14 for unmanaged token support
	Q2	JWT-supported ARC-CE
	Q3	
	Q4	
2027	Q1	WLCG Data Challenge 2027
	Q2	
	Q3	
	Q4	
2028	Q1	Completion of the X509 / VOMS phaseout (user certificate)
	Q2	

Steadily moving toward a pure token-based environment

Action Items toward X.509/VOMS phaseout 2028:

- mini-DC 2026
- VOMS service decommissioning by June 2028
- User registration directly to IAM (Must be a lot of work behind)
  - Non-F2F ID verification, MFA, etc



Summary

KEKCC Status & Operations

- ☺ Successfully entering the 2nd year of the 4-year contract with stable production
- ☺ High utilisation maintained (~90% CPU) with 55 PB of stored data (as of May 2026)

Network & Data Transfer Test

- ☺ Achieved 80+ Gbps throughput between KEK and B2RDCs
  - Belle II HL scenario: 40 TB/day (3.7 Gbps)
- ☺ Confirmed performance parity between X.509 and JWT-based AuthNZ

OS migration campaign to RHEL9 is still ongoing

- ☺ Done: CVMFS, StoRM
- ☺ Decommissioned: AMGA
- 😐 Working on: ARC-CE (in final phase)

Future Roadmap for AuthNZ

- 😐 Continued operation of the CA (KEK Grid CA 2024) and VOMS for the time being, until direct user registration to IAM is fully in production

At Social dinner, 2nd FJPPL collaboration meeting (COMP_X) in 2007

https://kds.kek.jp/event/6/

Cheers to 20 years of working together!

CCIN2P3

KEK
高エネルギー加速器研究機構