CLIP SYMPOSIUM · 13 NOVEMBER 2026

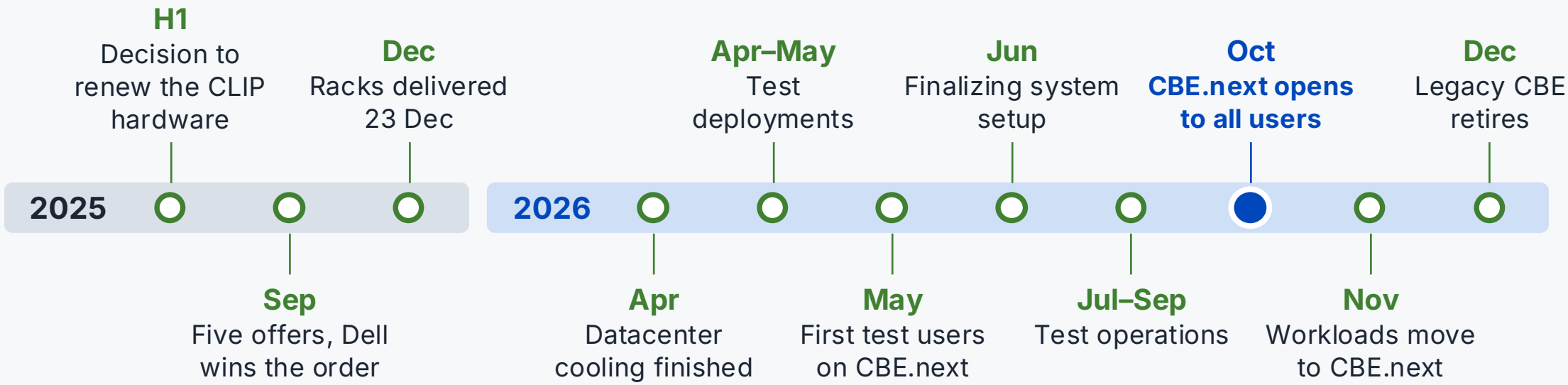
New CLIP on the Block

The story of the year

Erich Birngruber (GMI) · Ronny Zimmermann (IMBA)

3,456	**128**	**800 GB/s**	**400 Gbit/s**
Zen 5 CPU cores	GPUs: B200, A100, RTX 6000	storage throughput	RoCEv2 network

CBE.next timeline: from delivery to login



3.5 million Euro investment

64	3,456	8 TB	1.2 PB	200 kW
B200 GPUs	**Zen 5 CPU cores**	**High-memory nodes**	**Parallel flash storage**	**Datacenter cooling modifications**

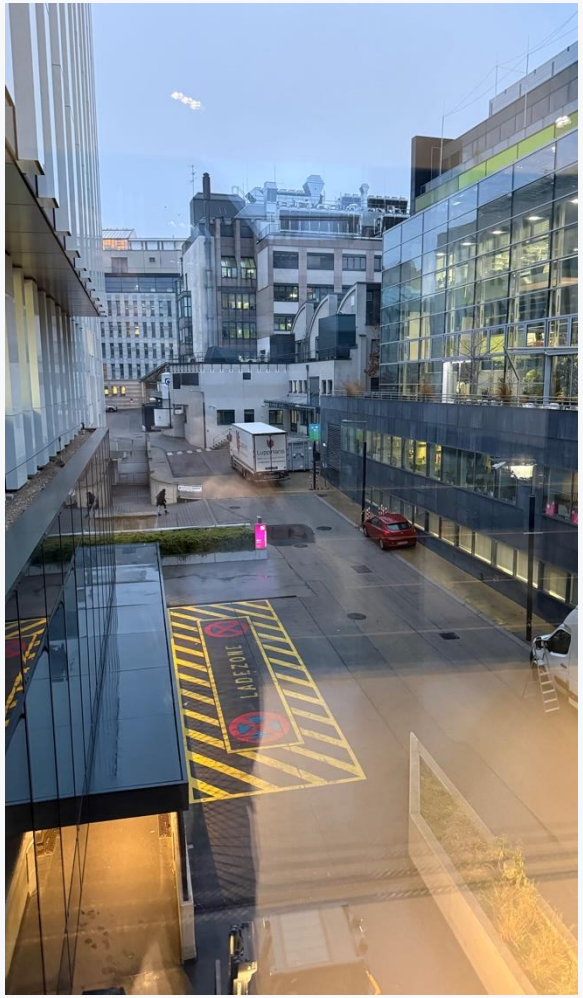
Four racks replace 200 previous-gen nodes.

6 → 4

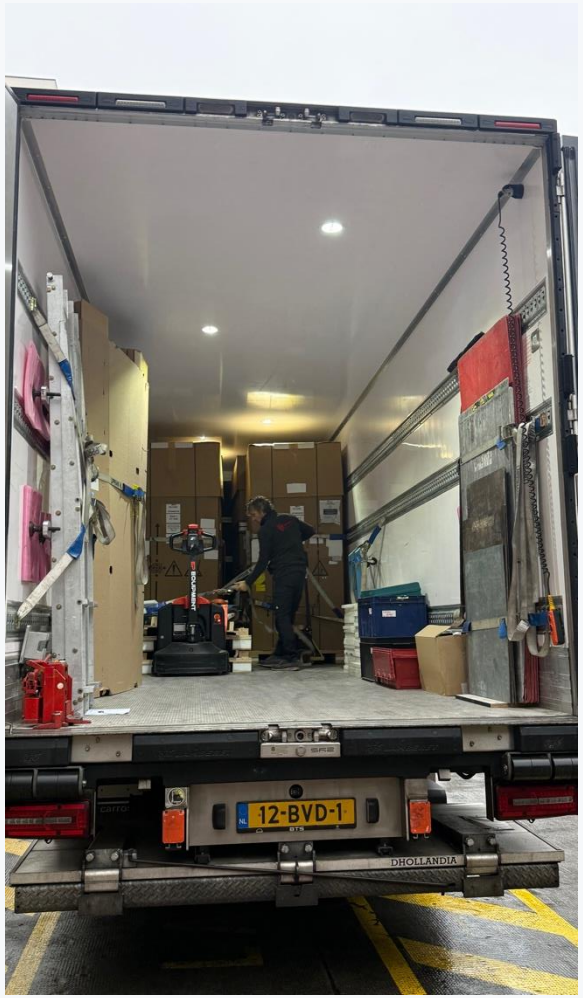
racks

100 → 200 kW

power consumption



The way in
A truck waiting for us



23 December 2025
Four racks unloaded



Unpacked outside
Crates opened in the courtyard



The last metres
Each rack squeezed into the lift

Power and cooling upgraded in a live datacenter

Power

12 × 63 A three-phase lines, up from **10 × 32 A**

Water

Floor pipes abandoned. New **200 mm** pipes welded in at the ceiling

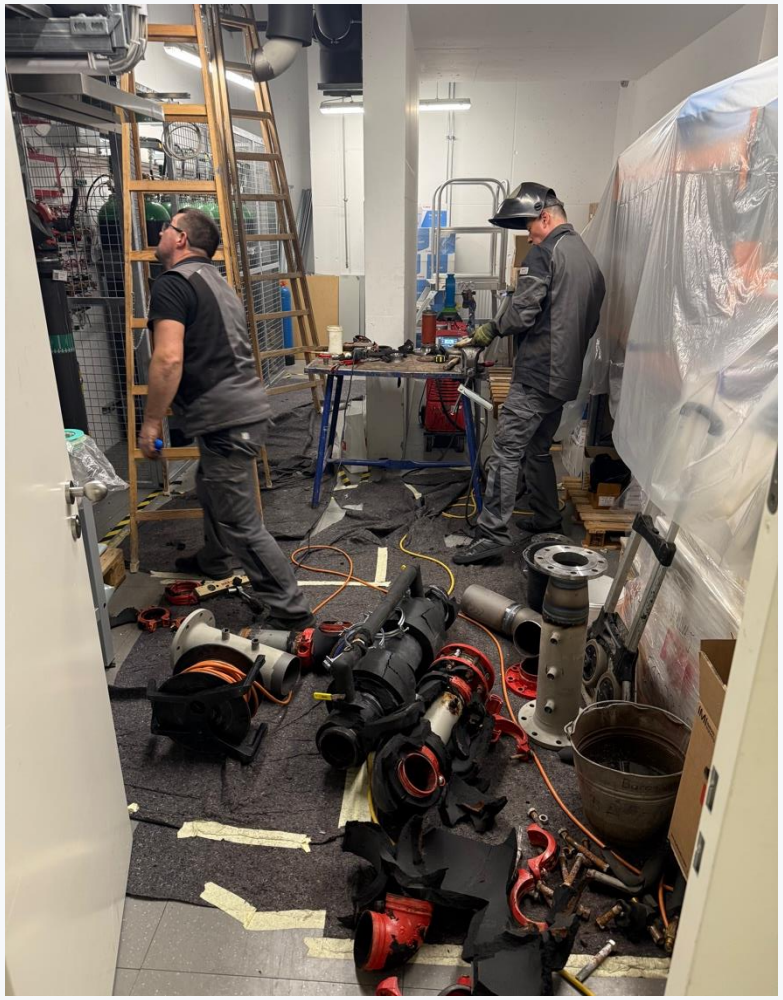
Cutover

Both new supplies switched on during one planned shutdown, **14 to 17 April**

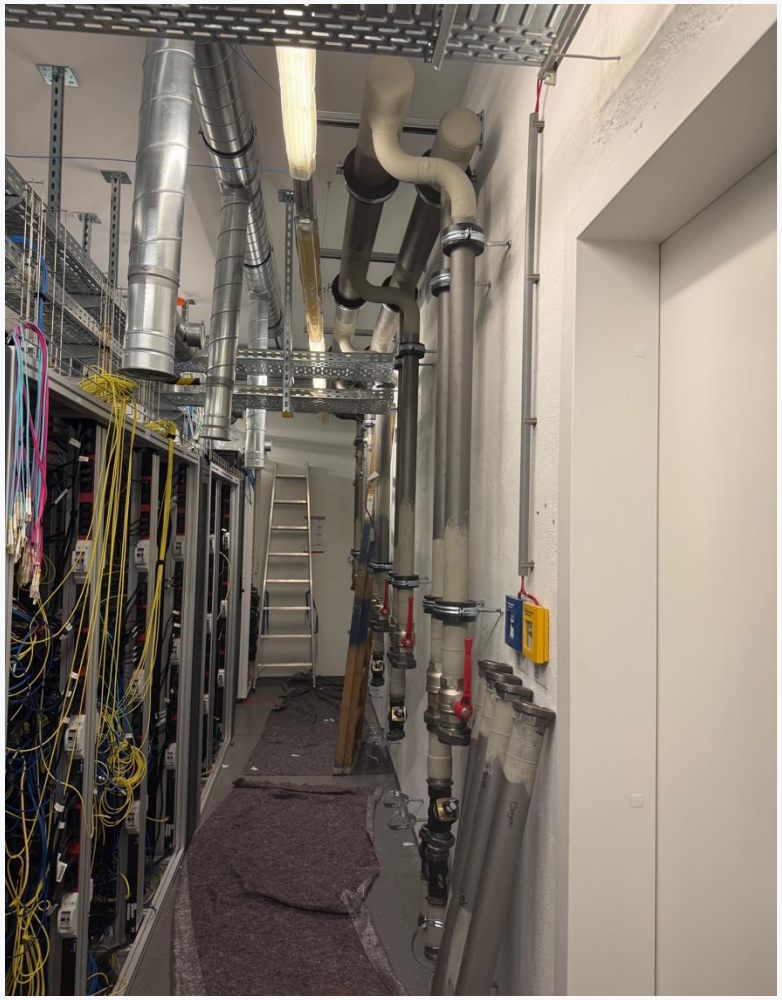
Critical network components in the same room stayed on through the cutover.

> **Thank you, Facility Management team**
>
> Thomas Tluchor · Martin Roijdl · Franz Hickl





Tapping into the building
New lines joined to the house system



Up to the ceiling
New 200 mm pipes above the racks



Under the floor
Old floor pipes taken out of service

New cooling, baptism by a hot summer

In-house

Water-cooled racks, new to everyone

New to the vendor, Facility Management and us. Steep learning curve: water cooling is nothing like air cooling. Every challenge was solved in-house.

40+ °C

Summer heatwave

Building cooling maxed out. Legacy CBE and CBE.next in parallel added to the load.

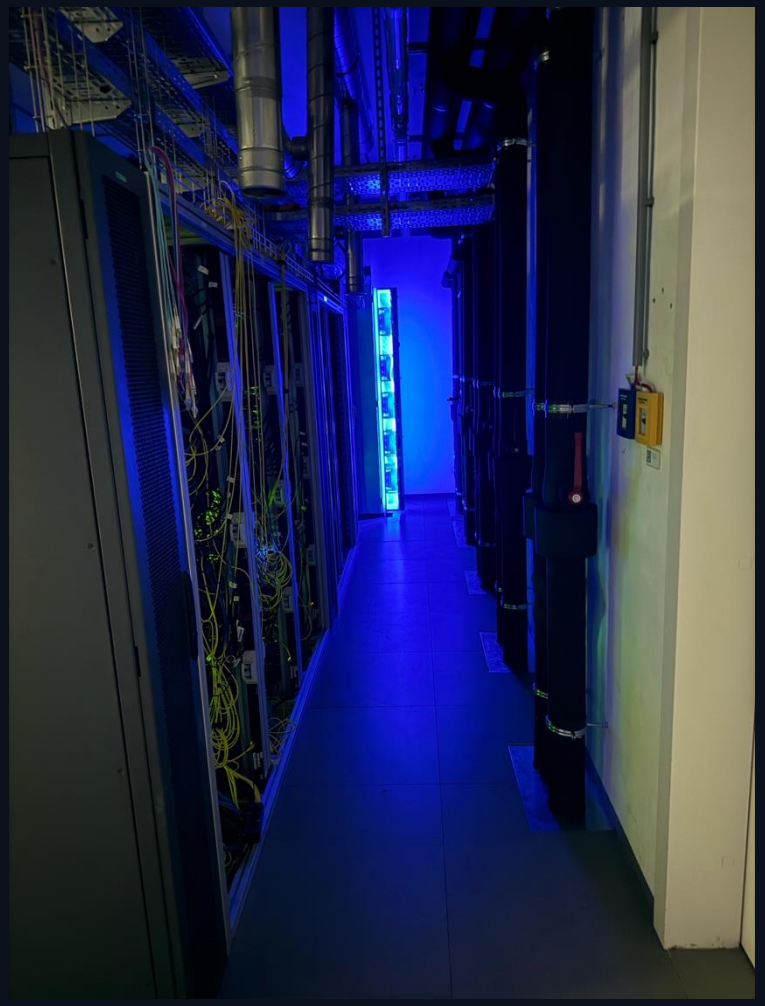
Before next summer

More capacity
In-house work is extending the cooling system.

Less load
Legacy CBE will already be powered off.

Close watch
We monitor cooling temperatures through the summer.





All in order
Racks closed and running



Running cool
Cooling working as planned

Nodes that set themselves up

Zero-touch

Planned in one system. From there, the node installs itself and starts running jobs.

Image-based

Software baked into the node image once. At first boot the node configures itself.

Low latency

400 Gbit/s RDMA network keeps the B200 GPUs fed with data from storage and other nodes.

Presented at FOSDEM 2026, Brussels: “Zero-Touch HPC Nodes”

archive.fosdem.org/2026/schedule/event/QTFECS-zero-touch-hpc-nodes

Thank you, systems engineering team

Ümit Seren · Leon Schwarzäugl

CBE.next: easier to reach, faster to compute

Open to all users since **31 October 2026**

Log in from anywhere

All access without VPN.
Secured by short lived credentials

Work in your browser

OpenOnDemand, with apps made for our users.

Ready for AI agents

Open-source agent skills (CLIP-HPC/skills) and agent-readable documentation.

340 GPUs

Bigger GPUs

MIG slices included in the count. 180 GB memory per B200.

1.2 PB

Data that keeps up

Flash storage at 800 GB/s.
GPUDirect streams data straight into GPU memory.



A soft landing on CBE.next

Get answers

Questions about logging in, access or the new system? We answer them and help you find your way around CBE.next.

Move your work

A step-by-step migration guide. We port your job scripts and workflows to CBE.next together with you.

Adapt and optimize

We help you tune your jobs for the new GPUs and storage, and build new workflows and apps on OpenOnDemand.

Thank you, user support team

Alexander Bindeus · Felix Schmidt

Building the future

NOW

Move in

Bring your work over, node by node

18 DECEMBER

Old CBE retires

Your deadline for data and workflows

2027

Build on it

More apps, LLM hosting, the CLIP 2030 strategy

Together with YOU



Keep exploring

clip.science

About CLIP, access, solutions and support

docs.vbc.ac.at

Research computing documentation: CBE.next, migration guide

HPC service desk: contact@clip.science · CLIP after the presentations

Happy Computing!