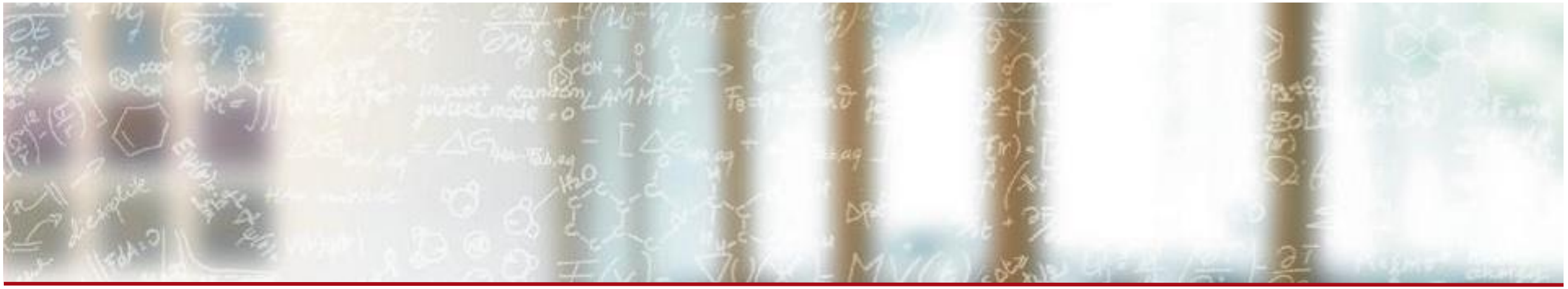
CSCS
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

ETH zürich



CSCS: From the Mountains to the Alps

Schichtwechsel, 5th September 2025

Nina Mujkanovic et al.

CSCS (Centro Svizzero di Calcolo Scientifico)

- A unit of the Swiss Federal Institute of Technology Zurich (ETHZ)
- Founded in 1991 in Manno, Ticino
- Moved to Lugano in 2012

CSCS: The Beginnings

1991

- "SWITCH" network established, linking the Federal Institutes of Technology with CSCS.
- CSCS begins operation on 30 September 1991.
- Installation of the first supercomputer, a NEC SX-3/22 "Adula".

1990

- The center's location is decided: a new building in Manno.

1988

- Start of discussions on locating the Swiss National Supercomputing Centre in Ticino.

1986

- Validity of the Federal decree limited to five years starting 1 October 1986; funding thus had to be used by 30 September 1991.

1985

- Federal decision to invest 40 million Swiss francs in a high performance computing center (> CSCS) and 15 million Swiss francs in developing a national research network for making the supercomputer accessible to Switzerland's leading universities (> SWITCH).

1980

- The Federal Government identifies a lack of well-trained computer scientists in Switzerland.

Key Facts

Staff

- 127 staff members
- 25 nationalities
- English as official language

Yearly Budget

- CHF 30 Mio. operational budget
- CHF 20 Mio. IT Investments

Building

- 2’600 m² offices
- 2’000 m² machine room

Electrical Power

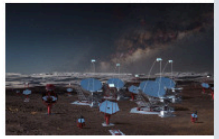
- Today 11 MW
- Possible extension to 25 MW
- 100% green energy (hydroelectrical)

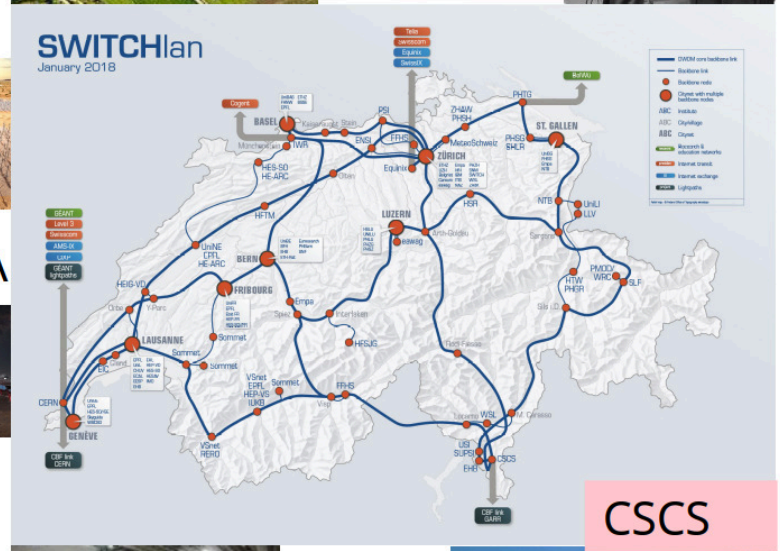
National and International Collaboration

PSI

MCH

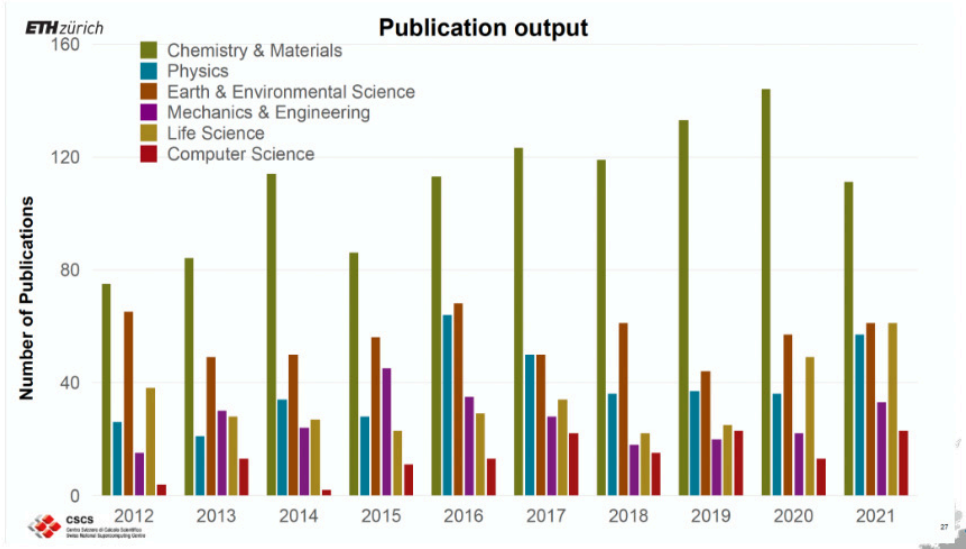
SKA / CTA

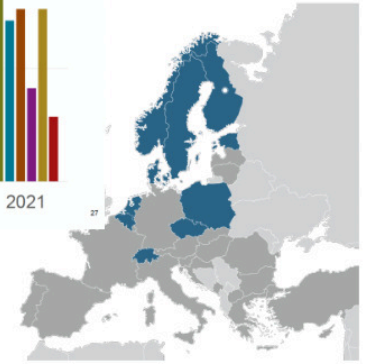


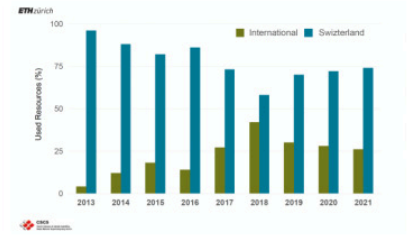


CSCS

CERN







CSCS: the ALPS research infrastructure (RI)

CSCS develops and operates a high-performance computing and data research infrastructure that supports world-class science in Switzerland.

Why Supercomputing?

Predicting the Weather

Weather is chaos

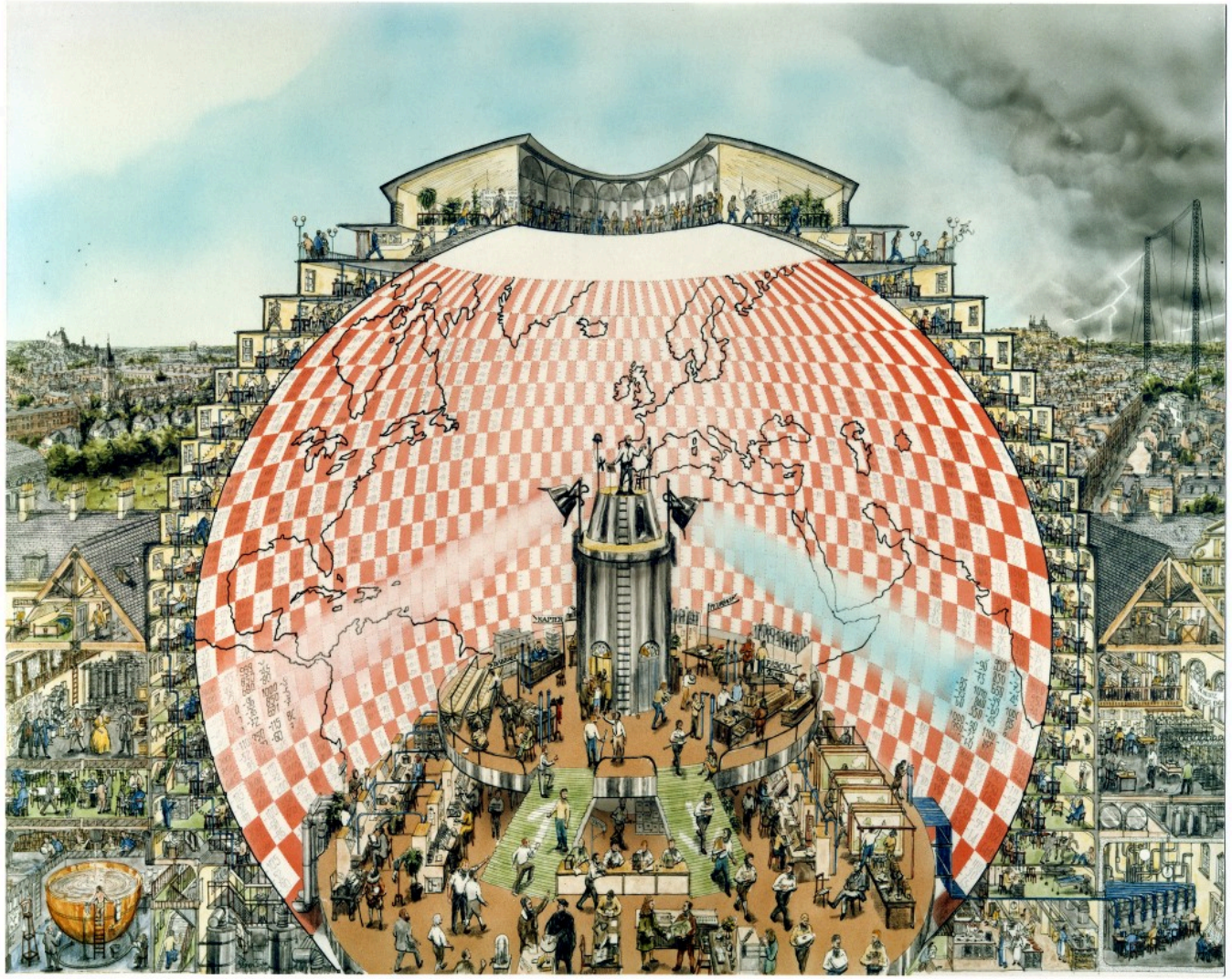
Predicting the Weather

- Met Office established in 1854 by Admiral Robert FitzRoy for maritime meteorology
  - established data collection system
  - production of weather charts including earliest wind roses
  - founded science of weather forecasting
  - concept of synoptic meteorology
- Aided by new electric telegraph
- Storm warnings issued from 1860
- General forecast 2 days ahead from 1861
- Publications of forecasts ceased in 1866

Richardson's Fantastic Forecast Factory

"A myriad [64000] computers are at work upon the weather of the part of the map where each sits ... . The work of each region is coordinated by an official of higher rank."

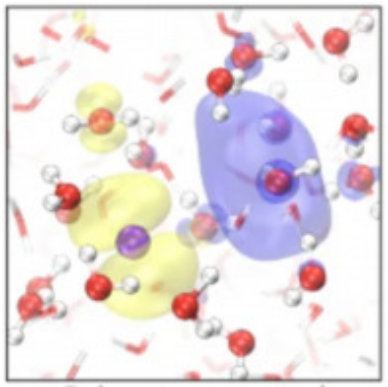
- Weather Forecasting by Numerical Process - Lewis Fry Richardson, 1922
  - Mathematical equations that describe atmospheric flow solution via dividing globe into cells

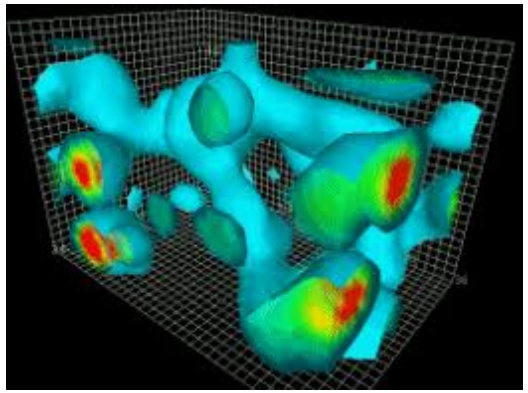


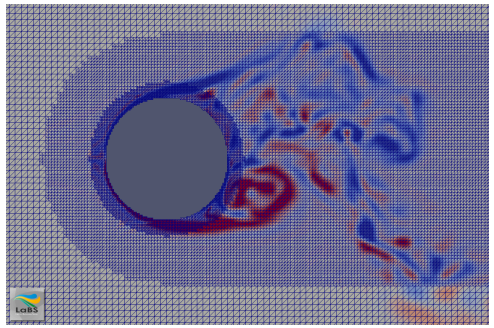
[1] https://www.emetsoc.org/resources/rff/

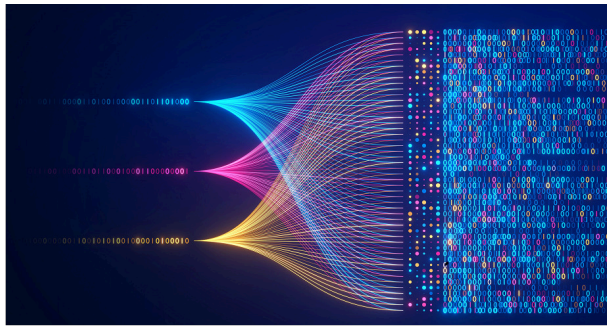
MeteoSwiss at CSCS

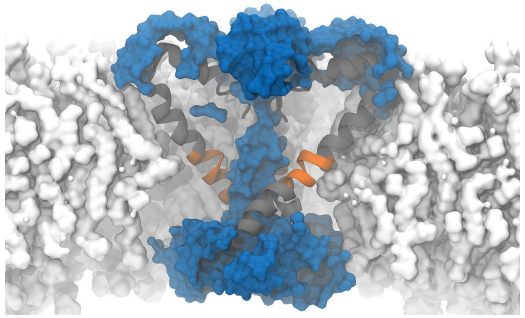
HPC and Examples

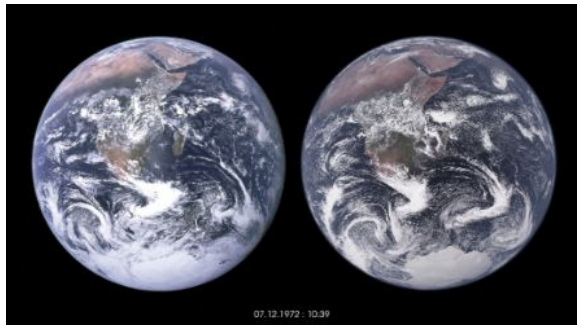










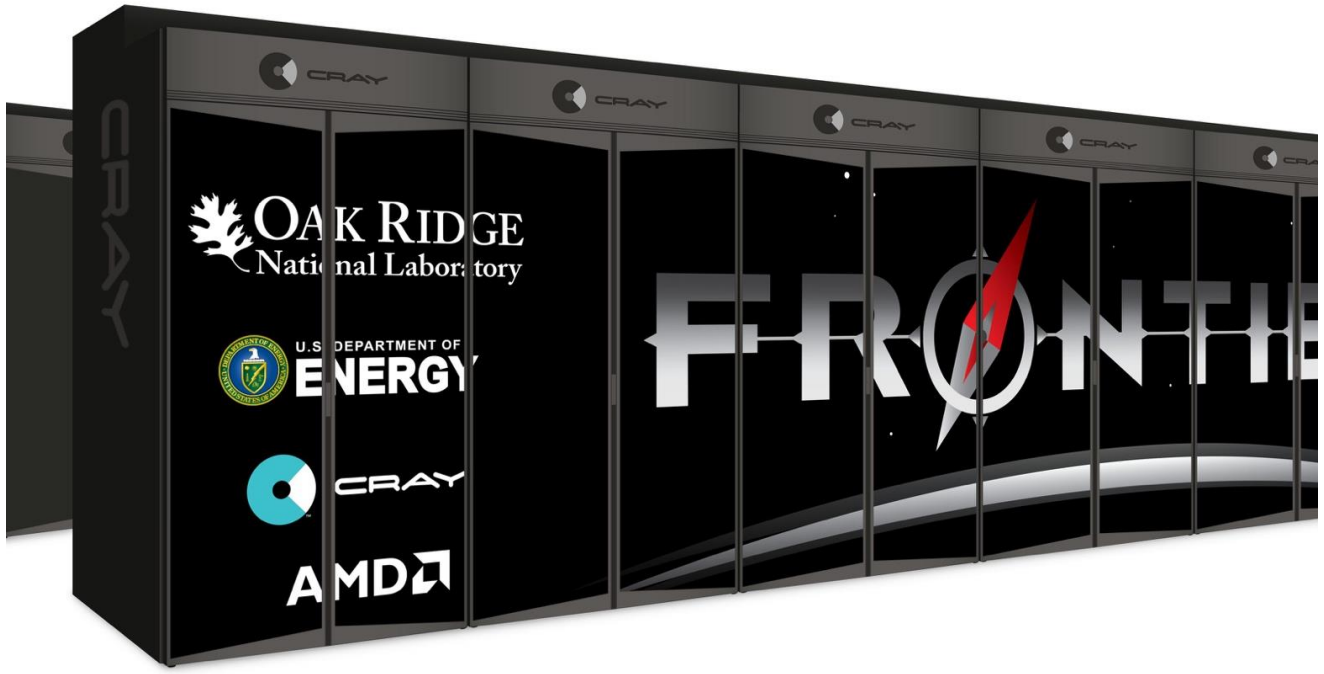


- Chemistry & Materials
- Physics
- Life Sciences
- Earth and Environmental Sciences
- Mechanical Engineering
- Computer Science

[1] https://doi.org/10.1038/s41467-024-46772-0
[2] http://www.physics.adelaide.edu.au/cssm/lattice/
[3] https://www.cscs.ch/science/life-science/2022/hpc-simulations-predict-mysterious-biological-processes-of-the-cell
[4] https://www.meteoswiss.admin.ch/about-us/media/press-releases/2024/02/milestone-in-climate-and-weather-research.html
[5] http://perso.ens-lyon.fr/emmanuel.leveque/labs.php

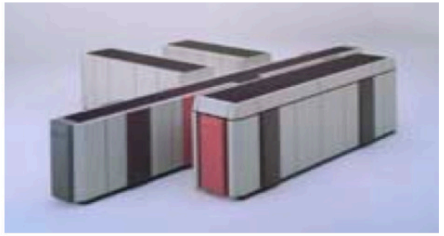
Then and Now

Cray-1, 1975



Frontier, 2022

Compute Infrastructure over time



1991 NEC SX3
5.5 GF Adula

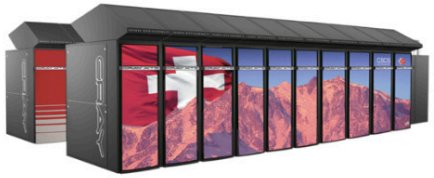
1996 NEC SX4
10 GF Gottardo

1999 NEC SX5
64 GF Prometeo

2002 IBM SP4
1.3 TF Venus

2005 Cray XT3
5.8 TF Palu

2006 IBM P5
4.5 TF Blanc



2009-12 Cray XE6 **402 TF**
Monte Rosa

2012-13-16 Cray XC40 / XC50
25 + 2 PF Piz Daint

2020 HPE Cray EX
0.3 – 0.4 EF Alps

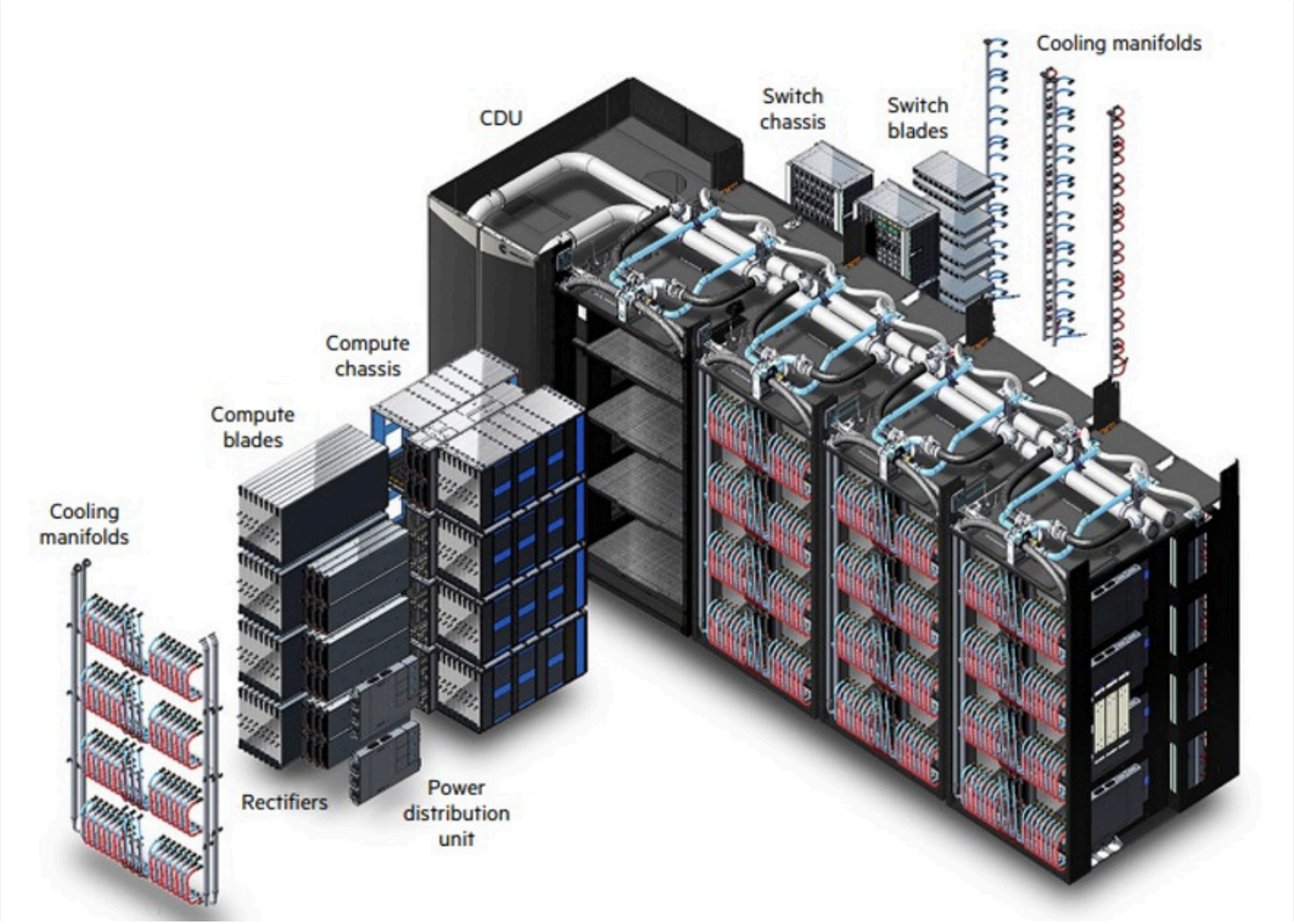
Alps Research Infrastructure

- Heterogeneous supercomputer
- HPE/Cray technology
  - HPE/Cray Shasta Slingshot network
  - Cabinets, power distribution, cooling
- Power envelope: 11 MW
- vClusters: dedicated partitions with tailored software environments
  - Daint, Eiger – for the User Lab
  - Clariden – for AI
  - Santis, Tasna & Balfrin – MeteoSwiss' numerical weather forecasts
  - Beverin - AMD MI300A cluster
- Geo-distributed hardware:
  - Lugano (CSCS)
  - Lausanne (EPFL)
  - Villingen (PSI) for data Archives.
  - Bologna for data access to ECMWF.

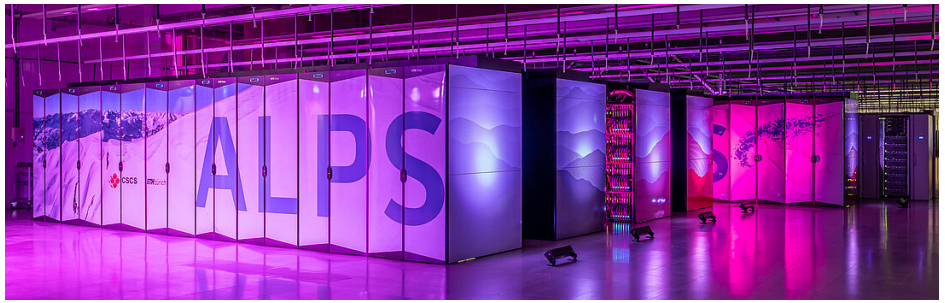
Multiple Architectures

- 1’024 multicore nodes
  - with 2 AMD Rome processors
- 2’688 hybrid ARM64+Nvidia nodes
  - with 4 NVIDIA Grace-Hopper superchip (288/4)
  - 10’752 GPUs
  - 6.9 PB of RAM
  - additional special purpose nodes
- 128 hybrid MI300A nodes
  - 8 APU's per node (292/8)
  - 24 AMD ‘Zen 4’ x86 CPU cores
  - 228 AMD CDNA 3 compute units / 912 Matrix Cores
  - 1024 GPUs
- hardware will be upgraded over time
- Memory:
  - 100 + 10 PB scratch disk
  - 5 + 1 PB Solid State Disk (SSD)
  - 2 tape libraries of ca. 130 PB each

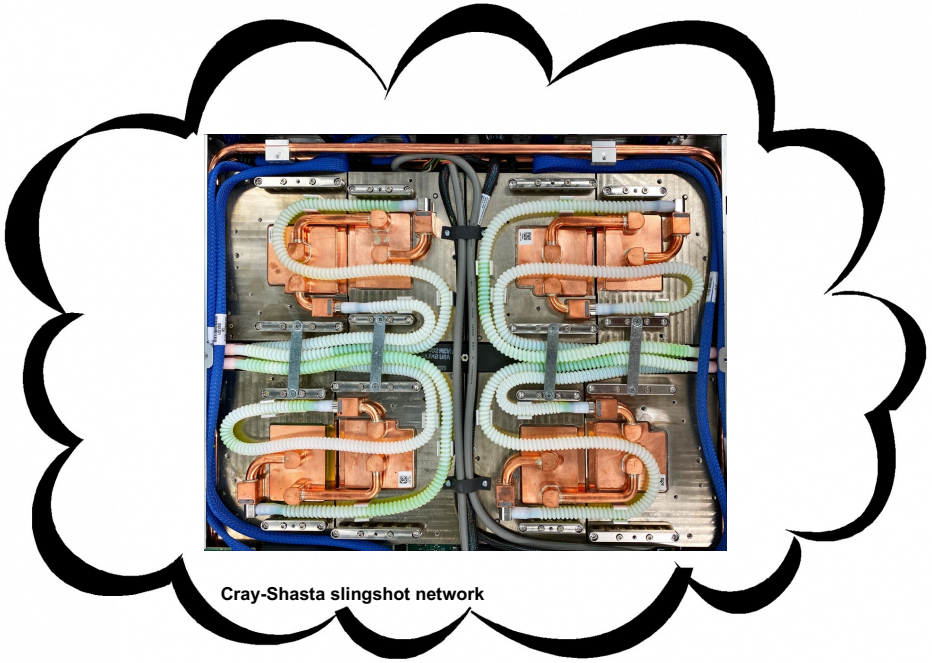
Alps Cray-EX - the Makings of a Supercomputer



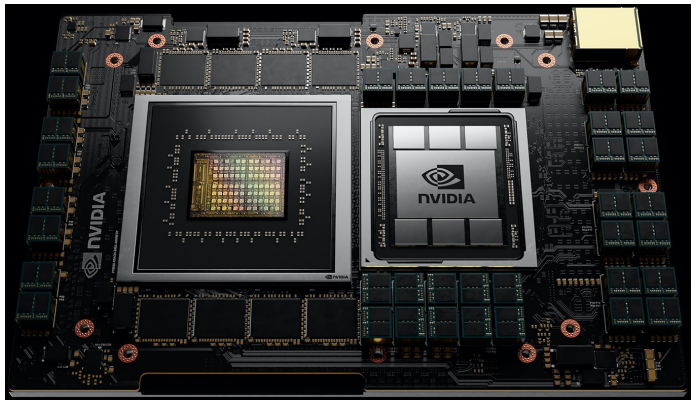
ALPS: 2,688 nodes = 10,752 Superchips



ALPS



GH200 = Grace + Hopper
= Superchip

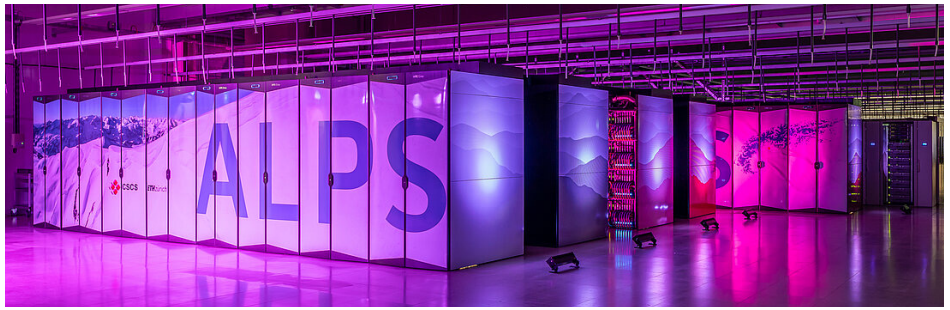


Power envelope: 10 MW

CSCS: Pioneering accelerated computing & cloud native HPC since 2011

Piz Daint, 2017

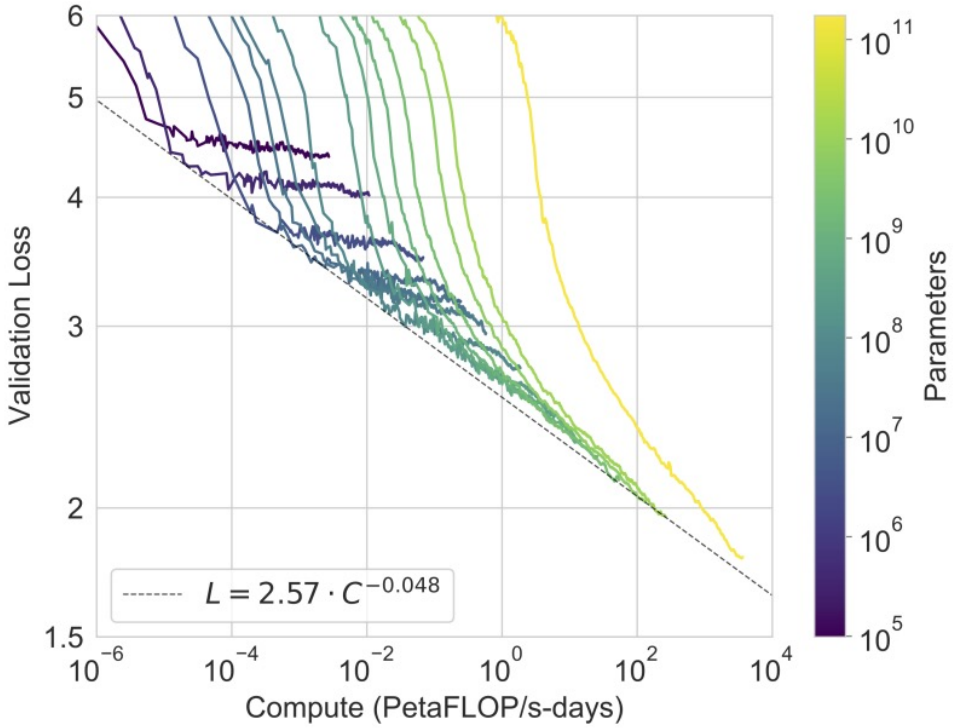
- Deployed GPU supercomputers in 2011
- Operationalized GPU based weather prediction in 2015
- Deployed Europe's largest and the world's greenest supercomputer in 2017
- NVIDIA's launch customer for P100

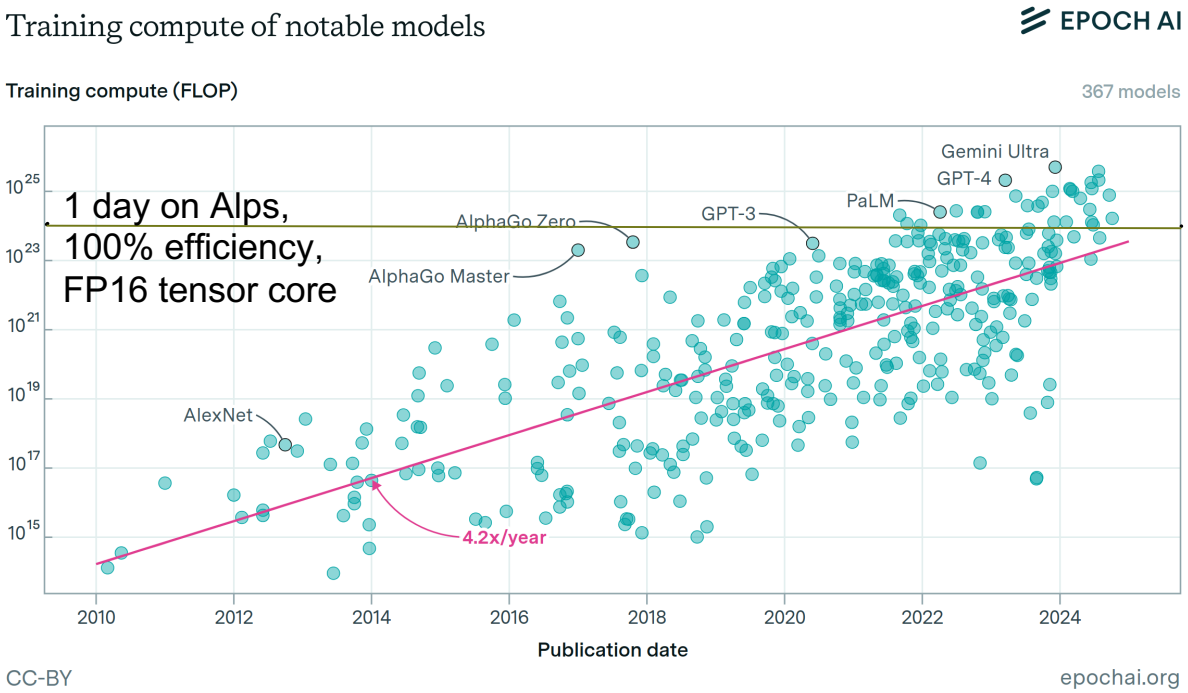


Alps, 2024 (inauguration)

- Europe's largest and greenest supercomputer in 2024 (?)
- Europe's most capable, open, AI supercomputer
- NVIDIA's launch customer for GH200

Infrastructure is key for modern AI developments

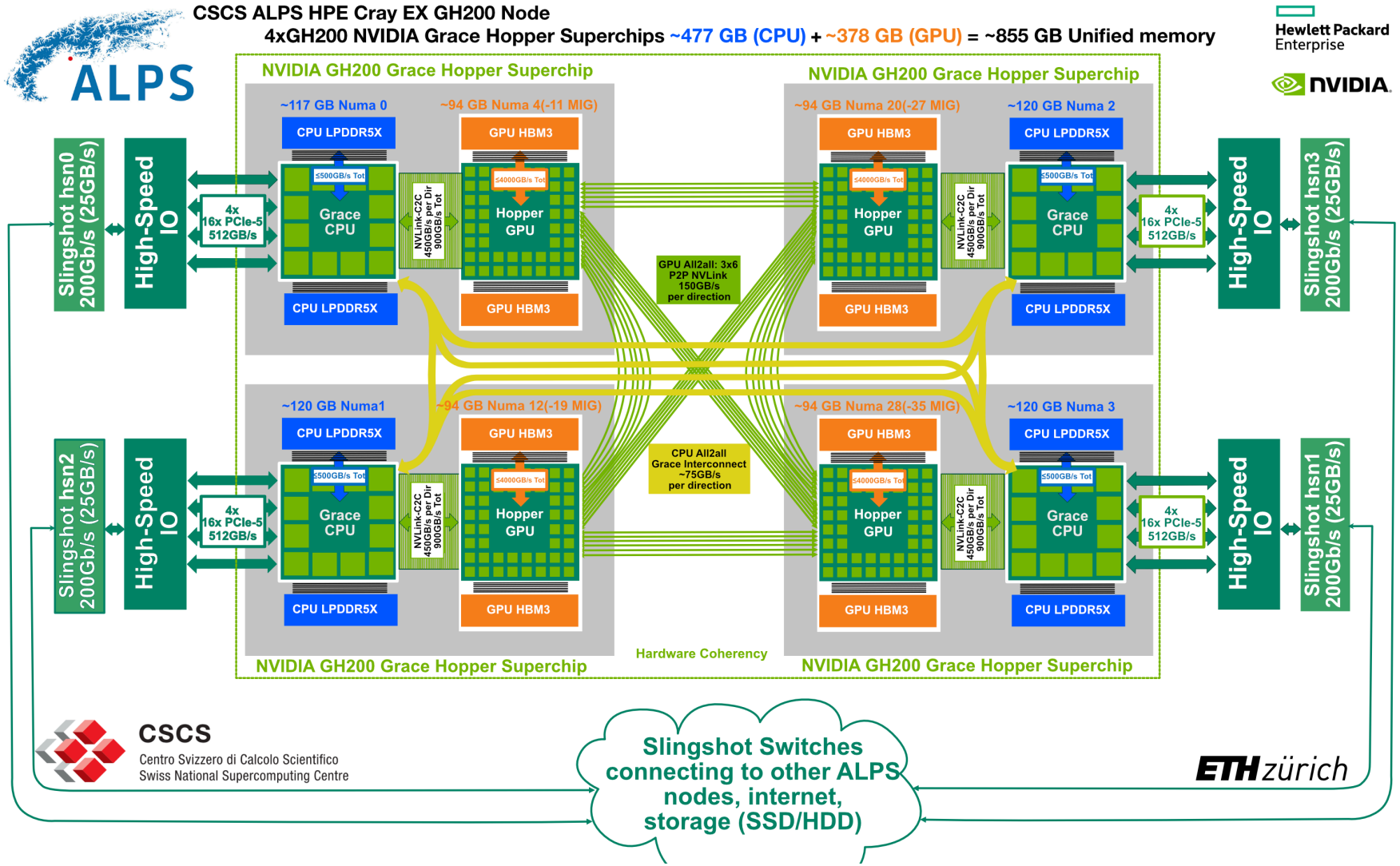




- No infrastructure, no modern AI. Accuracy and computational power are directly related.
- The rate of increase in FLOPs needed is staggering, growing much faster than Moore's law once did.

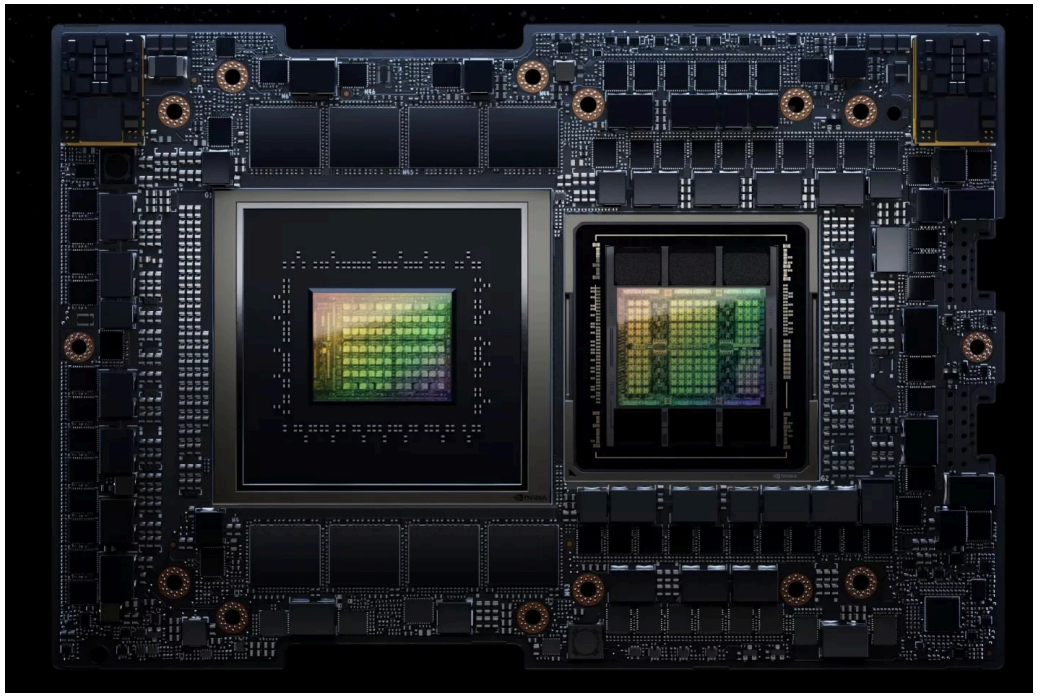
Alps Blade

ALPS Grace-Hopper node



ALPS Grace-Hopper module

- 72 ARM Grace CPU cores and one H100 GPU
- 128GB host memory and 96GB GPU memory
- Fast chip-to-top (C2C) link between CPU and GPU with 900GB/s
- Current Power Limit: 624.15 W
  - GPU and CPU share power budget
  - CPU has precedence over GPU

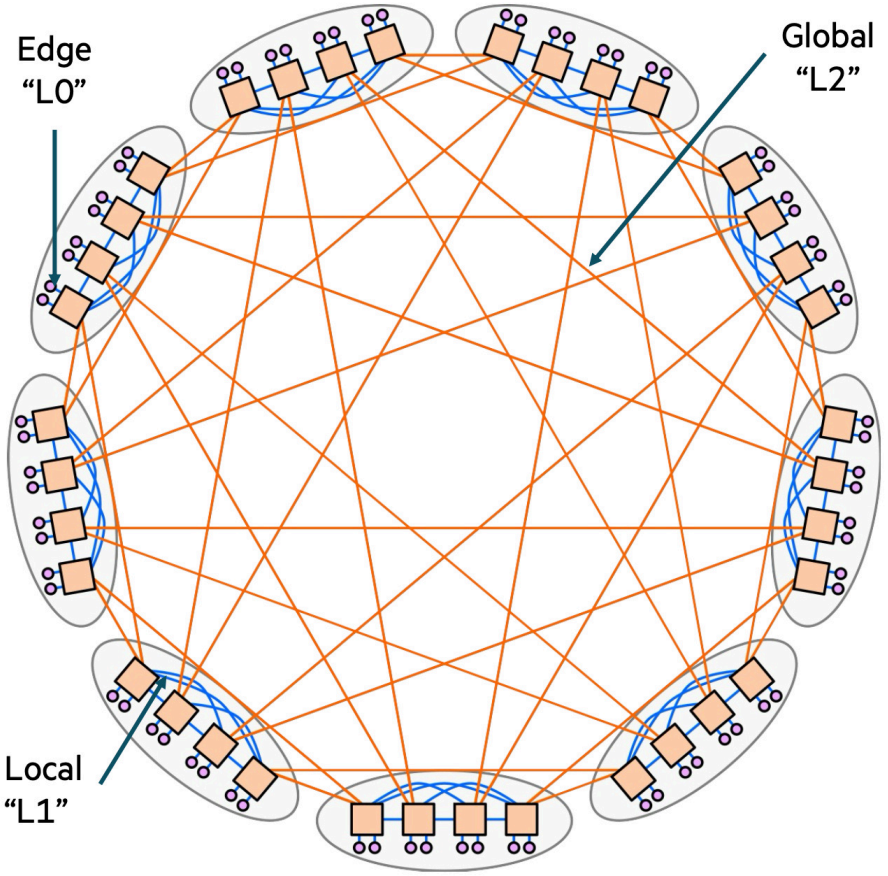


HPE Slingshot Network

- Built from Cassini NIC (PCIe Gen4) and Rosetta Switch (64-port).
  - one NIC (Cassini) per GH200 module (4 per node)
- Uses 200 Gbps Ethernet links for dragonfly networks
  - 25 GB/s unidirectional peak throughput per NIC
- Supports 64 to 250,000+ endpoints (largest system: ~85,000 endpoints).
- Optimized Ethernet protocol
- Adaptive routing & congestion management optimize performance.
- support remote memory operations

Dragonfly Network Topology

- Low-diameter “direct” network with no external fat-tree top switches.
- Fewer optical links needed → reduces cost for large systems.
- Groups of routers act as high-radix virtual routers to improve scalability.
- Three levels of connectivity:
  - Edge links (L0): Nodes to local routers.
  - Intra-group links (L1): Routers within a group.
  - Inter-group links (L2): Optical connections between groups.
- Scales linearly in cost while providing high bandwidth.
- Max three hops between components in the network.

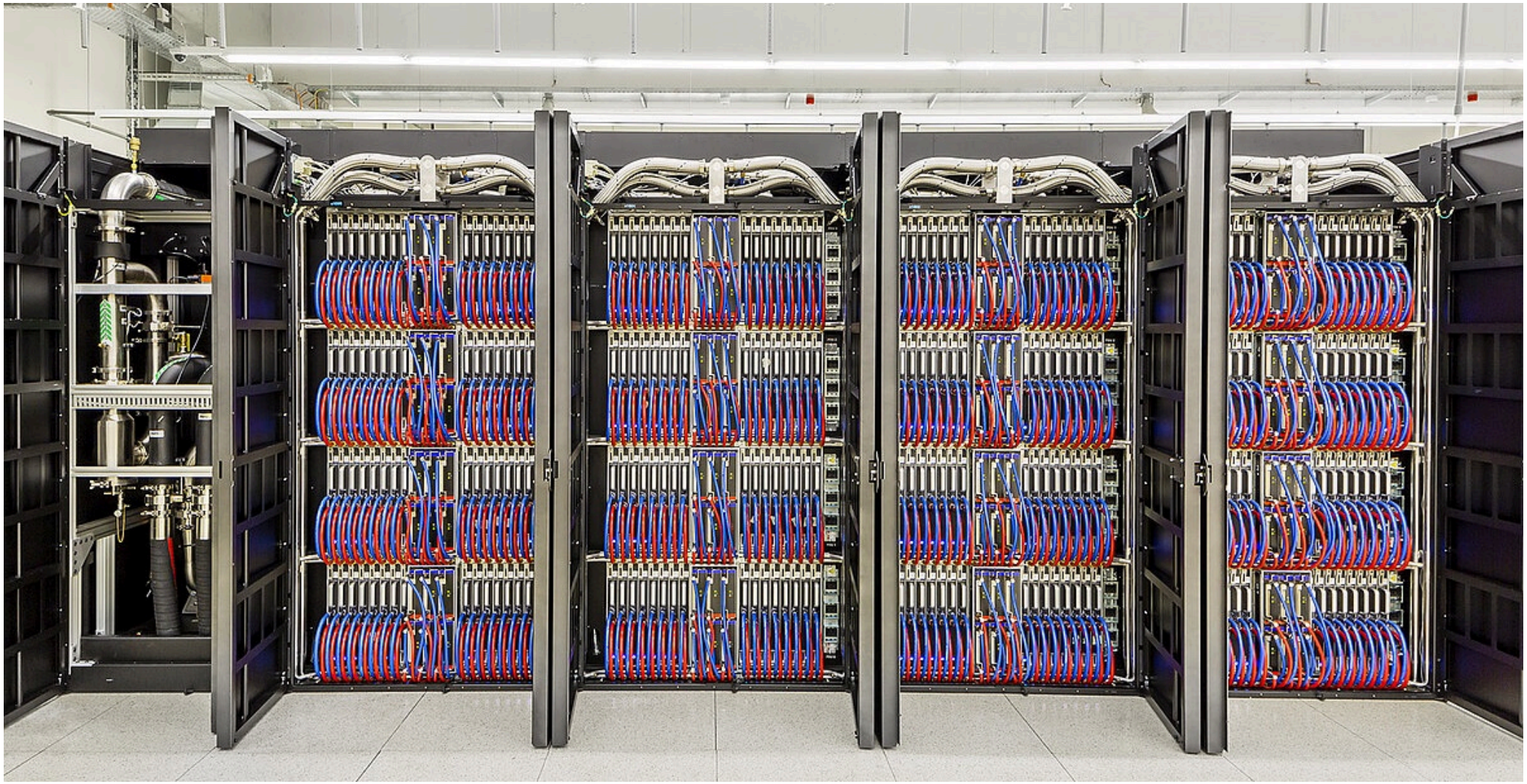


Cooling considerations

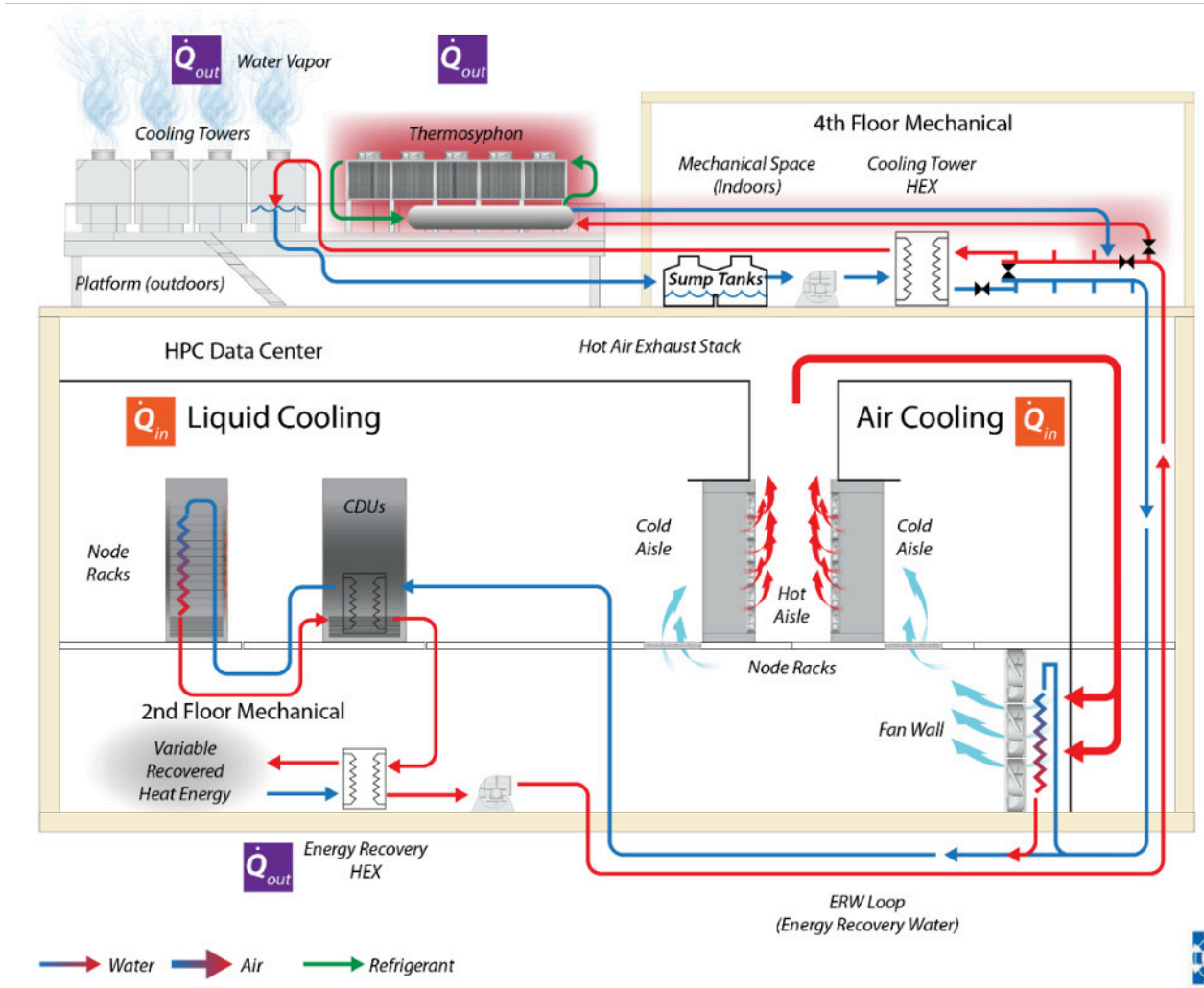
- Cooling requirements
- Thermal footprint
- Cooling installation (rack density, raised floors, aisle containment, humidity, filtration,...)
- Power consumption (air cooling fans account for 20% of power consumption)
- Water flow rates, fluid distribution, leak detection
- Chemical mix of fluid (biocides)
- Heat circulation
- Environmental cooling

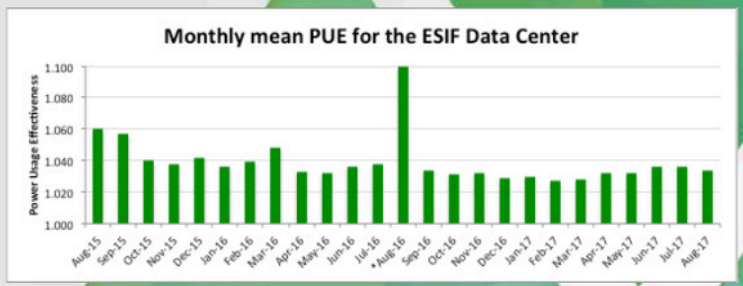
Cooling at CSCS

Cooling at CSCS



Cooling Integration





NREL NATIONAL RENEWABLE ENERGY LABORATORY

Johnson Controls

Sandia National Laboratories

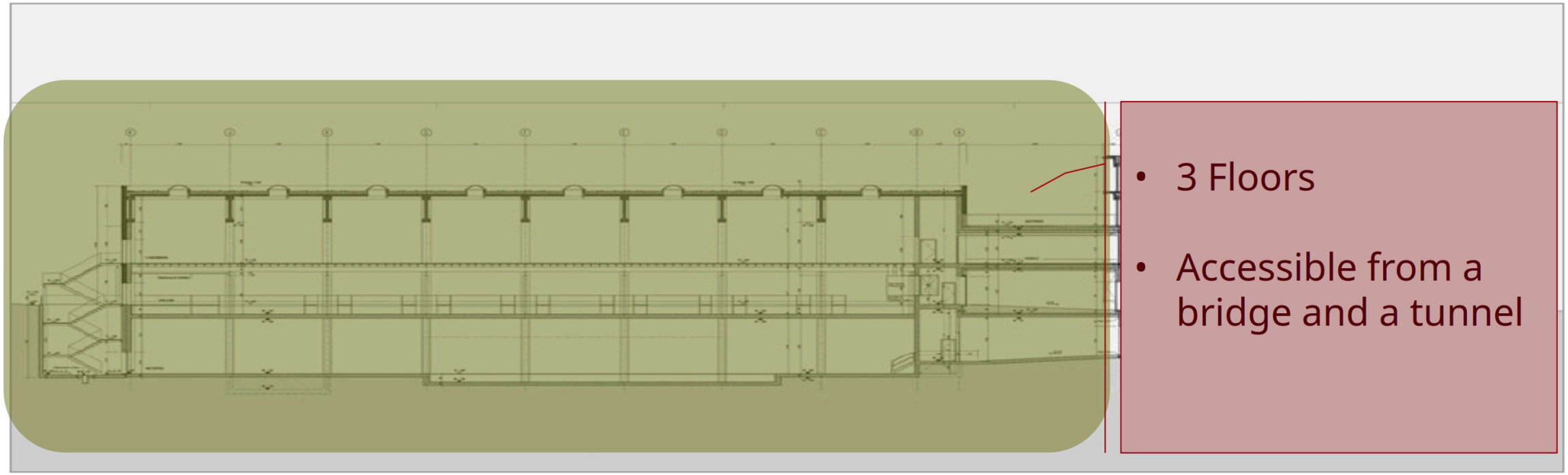
CSCS: From Manno...

… To Lugano

Office Building

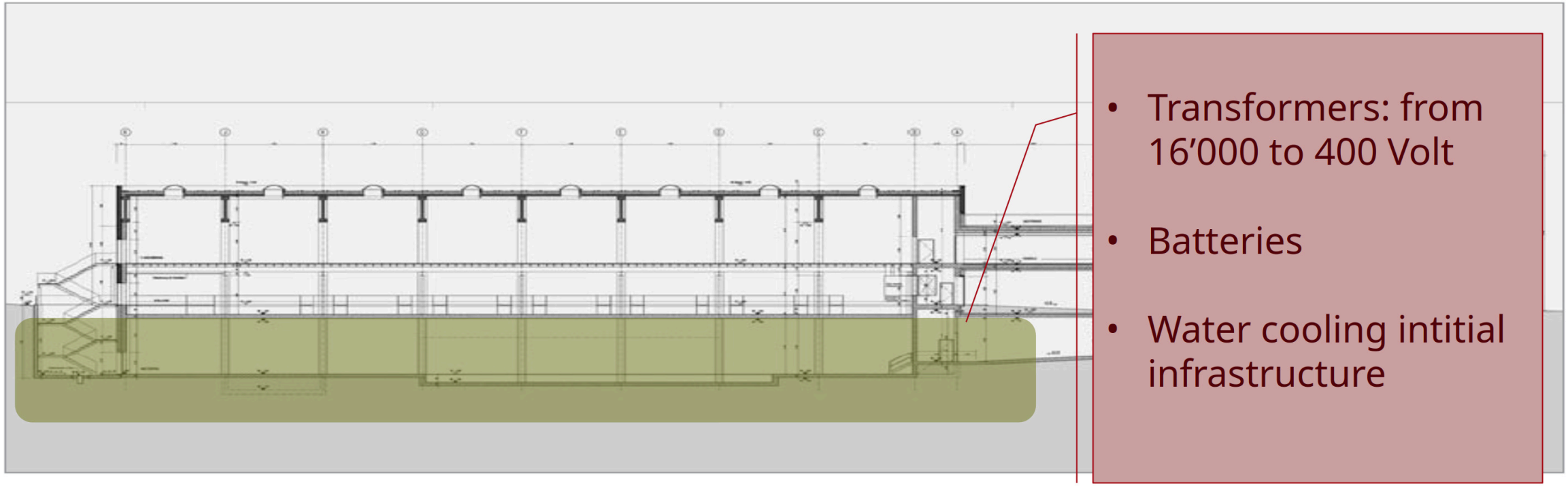
Data Center

Data Center: Overview



- 3 Floors
- 2000 m²
- Next to fire fighters

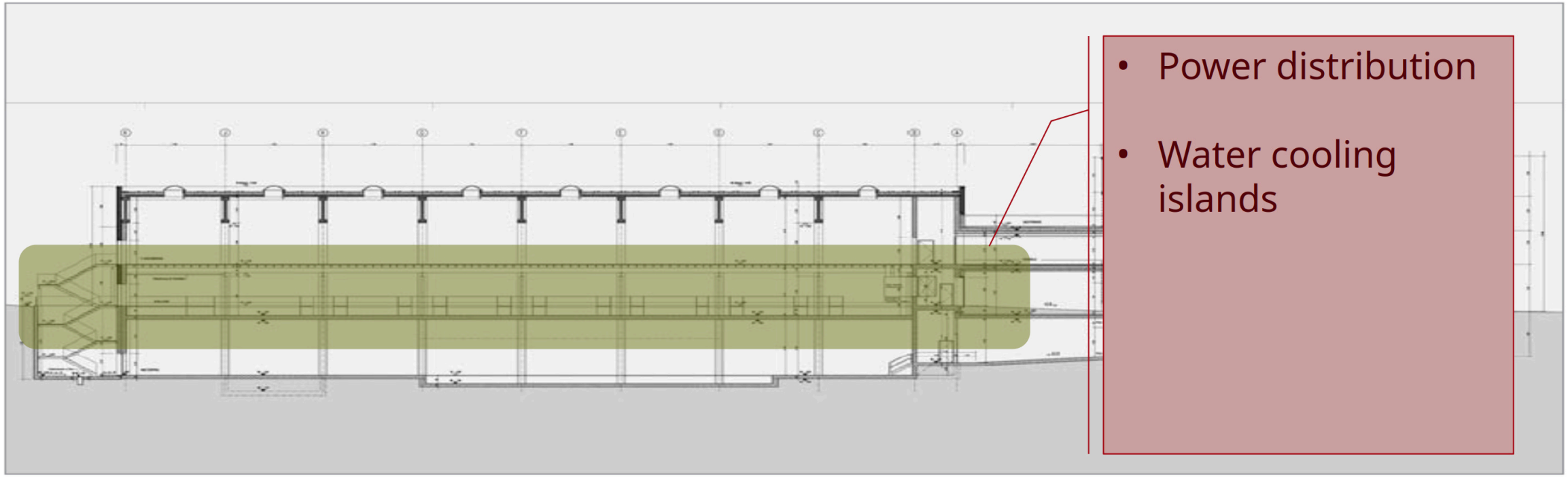
Data Center: Resource Deck



- Transformers: from 16’000 to 400 Volt
- Batteries
- Water cooling intitial infrastructure

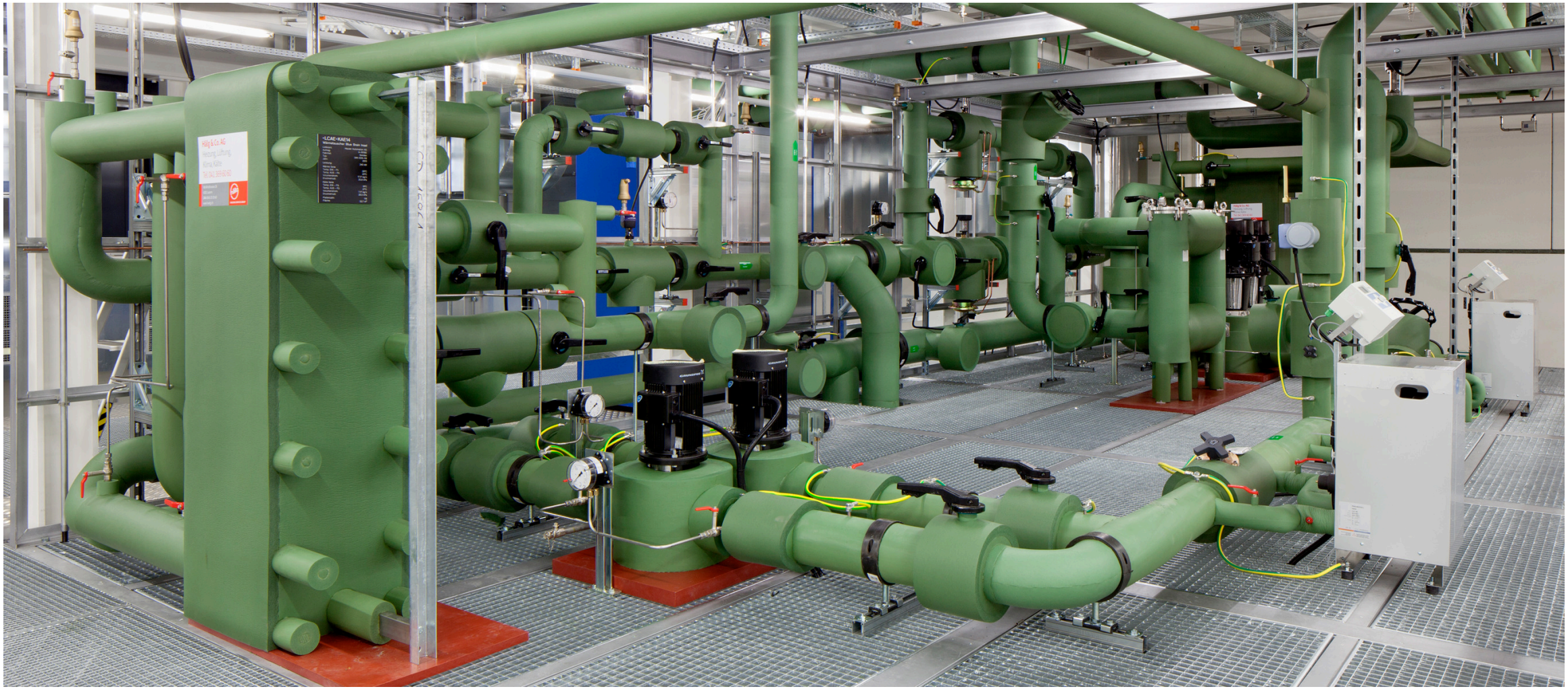
Data Center: Resource Deck

Data Center: Distribution Deck

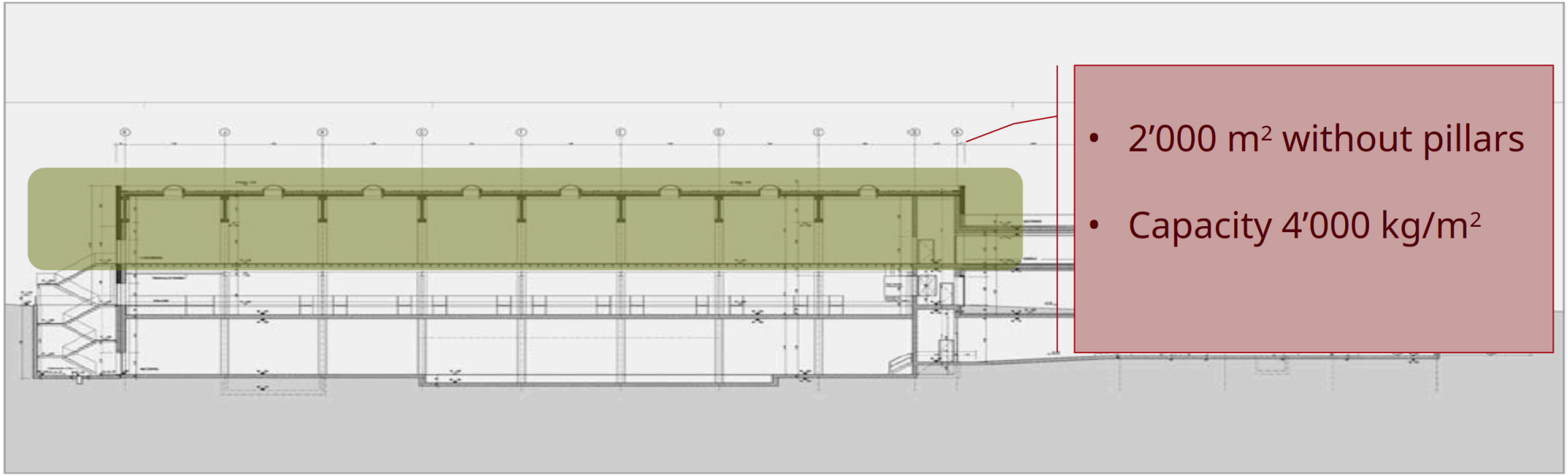


- Power distribution
- Water cooling islands

Data Center: Distribution Deck



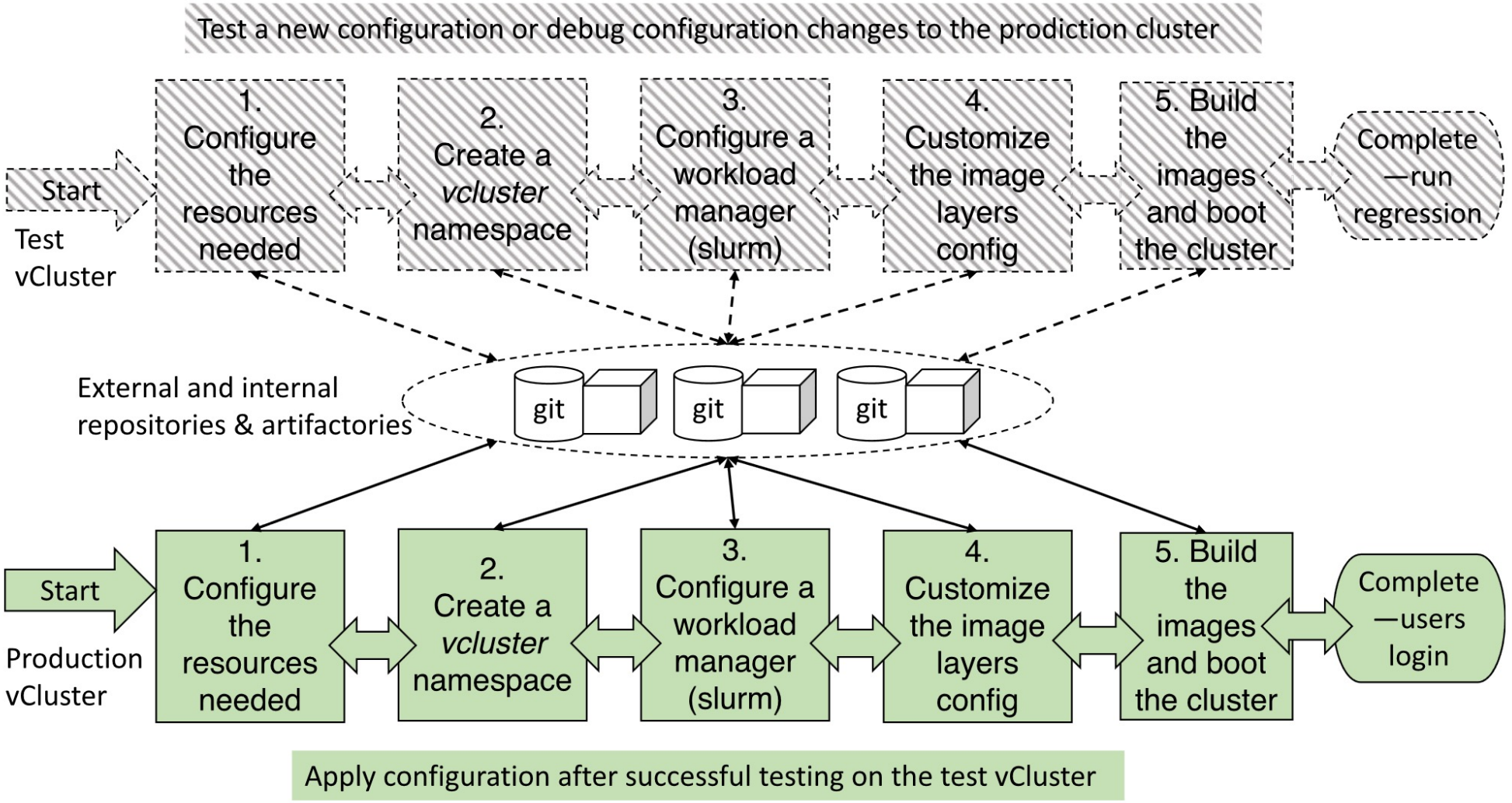
Data Center: Machine Room



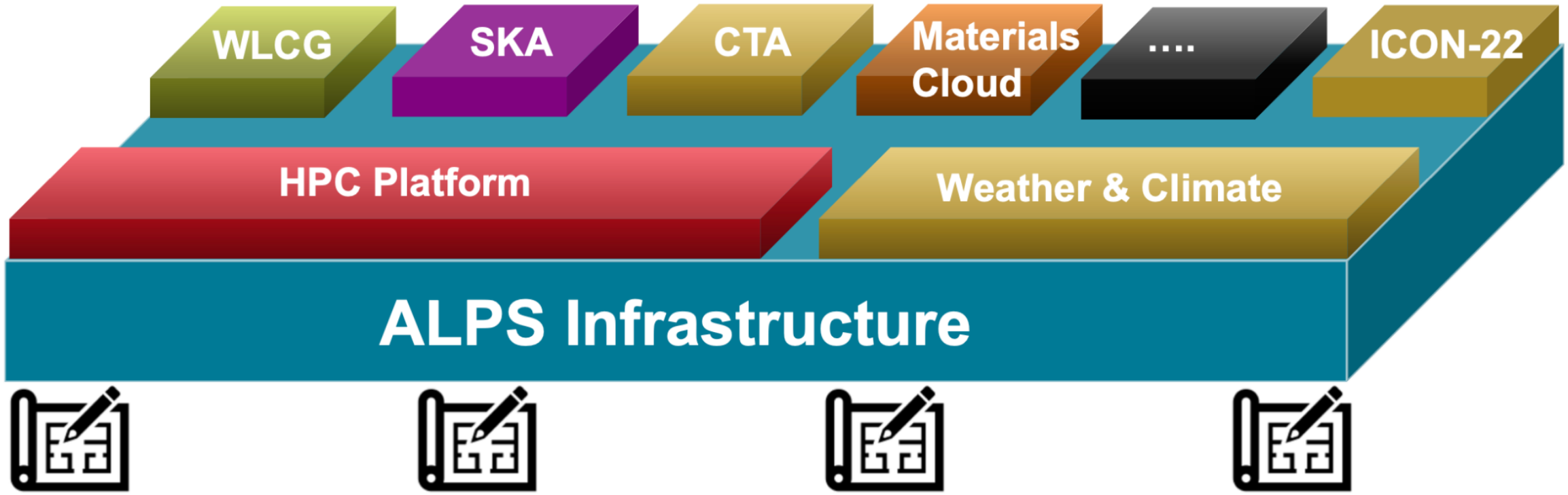
Alps Research Infrastructure

- Heterogeneous supercomputer
- HPE/Cray technology
  - HPE/Cray Shasta Slingshot network
  - Cabinets, power distribution, cooling
- Power envelope: 11 MW
- vClusters: dedicated partitions with tailored software environments
  - Daint, Eiger – for the User Lab
  - Clariden – for AI
  - Santis, Tasna & Balfrin – MeteoSwiss' numerical weather forecasts
  - Beverin - AMD MI300A cluster
- Geo-distributed hardware:
  - Lugano (CSCS)
  - Lausanne (EPFL)
  - Villingen (PSI) for data Archives.
  - Bologna for data access to ECMWF.

Versatile Software-Defined Cluster



Versatile Software-Defined Cluster



Data Center: Machine Room

Learnings from the last 1-2 years

- HW, good choice, very well suited to ML
- Collaboration with our partners: HPE/NVIDIA
- Requires whole CSCS Team
- GH200
  - Delays
  - ARM
  - Unified memory
- Slingshot (nice but when it fails affects everything)
  - Network isolation possible
- Cooling & Power
  - lucky to have a very good team
  - had to be careful when old+new system were both running
- Storage (lustre, VAST)
- SW: platform
- vServices
  - Infrastructure as code
  - for internal use first, but interested in sharing
- User and resource management compatible with cloud approaches
- SLURM
  - Locality/topology aware
  - Fair scheduler also when overbooked
  - Much experience in-house
- Kubernetes
  - Tricky multitenancy
  - Good for services (wand-db, inference service,...)
  - First applications on GH200

Top 500 (Linpack)

- ALPS is the only GH200 system in the Top10

Rank	System	Cores	Rmax (PFlop/s)	Rpeak (PFlop/s)	Power (kW)
1	**El Capitan** - HPE Cray EX255a, AMD 4th Gen EPYC 24C 1.8GHz, AMD Instinct MI300A, Slingshot-11, TOSS, HPE DOE/NNSA/LLNL United States	11,039,616	1,742.00	2,746.38	29,581
2	**Frontier** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE Cray OS, HPE DOE/SC/Oak Ridge National Laboratory United States	9,066,176	1,353.00	2,055.72	24,607
3	**Aurora** - HPE Cray EX - Intel Exascale Compute Blade, Xeon CPU Max 9470 52C 2.4GHz, Intel Data Center GPU Max, Slingshot-11, Intel DOE/SC/Argonne National Laboratory United States	9,264,128	1,012.00	1,980.01	38,698
4	**Eagle** - Microsoft NDv5, Xeon Platinum 8480C 48C 2GHz, NVIDIA H100, NVIDIA Infiniband NDR, Microsoft Azure Microsoft Azure United States	2,073,600	561.20	846.84	
5	**HPC6** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, RHEL 8.9, HPE Eni S.p.A. Italy	3,143,520	477.90	606.97	8,461
6	**Supercomputer Fugaku** - Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D, Fujitsu RIKEN Center for Computational Science Japan	7,630,848	442.01	537.21	29,899
7	**Alps** - HPE Cray EX254n, NVIDIA Grace 72C 3.1GHz, NVIDIA GH200 Superchip, Slingshot-11, HPE Cray OS, HPE Swiss National Supercomputing Centre (CSCS) Switzerland	2,121,600	434.90	574.84	7,124

HPCG 500 (High-Performance Conjugate Gradient)

- Different Application, different ranking
- Not all systems are in both lists

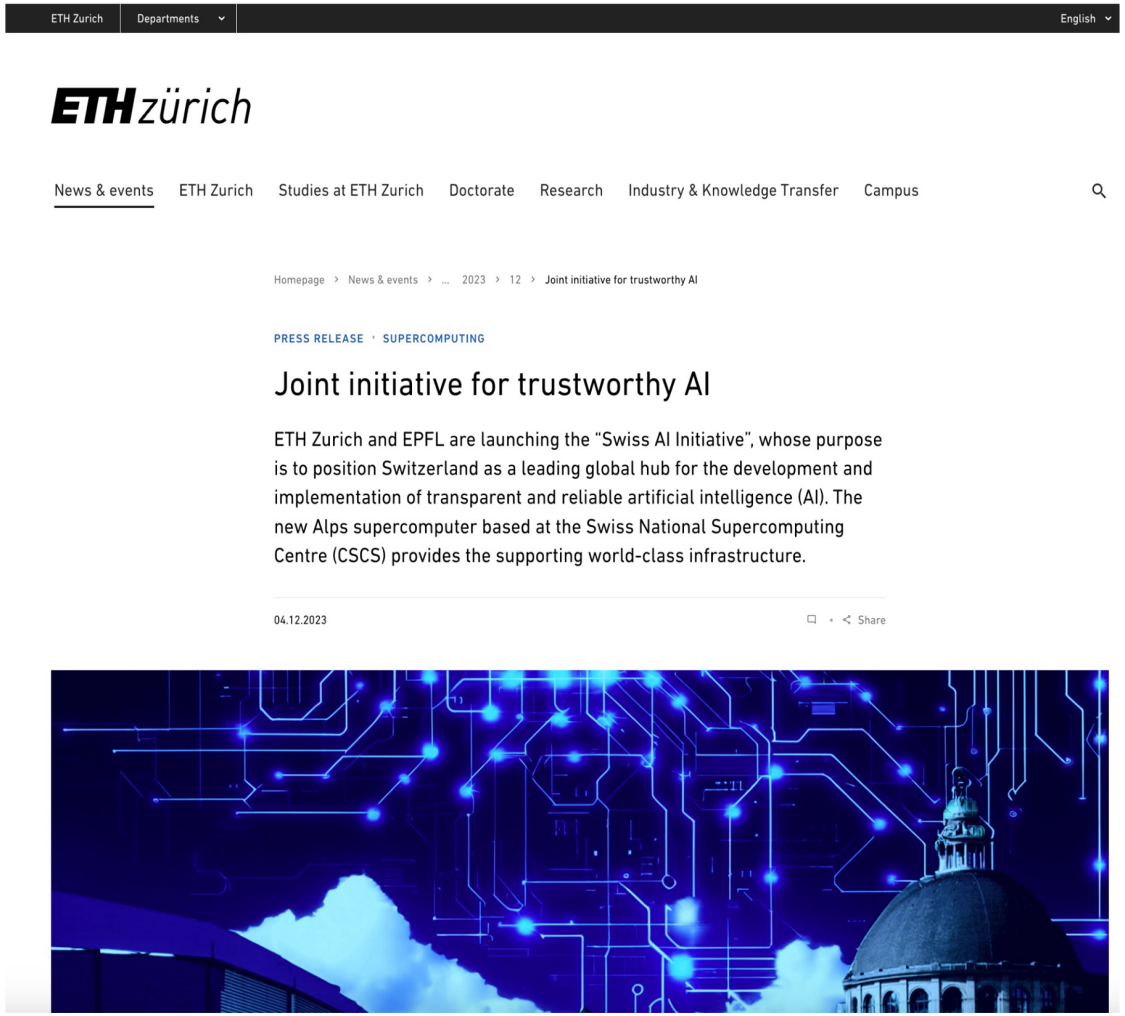
Rank	TOP500 Rank	System	Cores	Rmax (PFlop/s)	HPCG (TFlop/s)
1	6	**Supercomputer Fugaku** - Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D, Fujitsu RIKEN Center for Computational Science Japan	7,630,848	442.01	16004.50
2	2	**Frontier** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE Cray OS, HPE DOE/SC/Oak Ridge National Laboratory United States	9,066,176	1,353.00	14054.00
3	3	**Aurora** - HPE Cray EX - Intel Exascale Compute Blade, Xeon CPU Max 9470 52C 2.4GHz, Intel Data Center GPU Max, Slingshot-11, Intel DOE/SC/Argonne National Laboratory United States	9,264,128	1,012.00	5612.60
4	8	**LUMI** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE EuroHPC/CSC Finland	2,752,704	379.70	4586.95
5	7	**Alps** - HPE Cray EX254n, NVIDIA Grace 72C 3.1GHz, NVIDIA GH200 Superchip, Slingshot-11, HPE Cray OS, HPE Swiss National Supercomputing Centre (CSCS) Switzerland	2,121,600	434.90	3671.32

Green 500

- GH200 have top spots in the green top 10
- Leaders are smaller systems
- ALPS is more efficient than the 6 faster systems

Rank	TOP500 Rank	System	Cores	Rmax (PFlop/s)	Power (kW)	Energy Efficiency (GFlops/watts)
1	222	**JEDI** - BullSequana XH3000, Grace Hopper Superchip 72C 3GHz, NVIDIA GH200 Superchip, Quad-Rail NVIDIA InfiniBand NDR200, ParTec/EVIDEN EuroHPC/FZJ Germany	19,584	4.50	67	72.733
2	122	**ROMEO-2025** - BullSequana XH3000, Grace Hopper Superchip 72C 3GHz, NVIDIA GH200 Superchip, Quad-Rail NVIDIA InfiniBand NDR200, Red Hat Enterprise Linux, EVIDEN ROMEO HPC Center - Champagne-Ardenne France	47,328	9.86	160	70.912
3	440	**Adastra 2** - HPE Cray EX255a, AMD 4th Gen EPYC 24C 1.8GHz, AMD Instinct MI300A, Slingshot-11, RHEL, HPE	16,128	2.53	37	69.098
14	7	**Alps** - HPE Cray EX254n, NVIDIA Grace 72C 3.1GHz, NVIDIA GH200 Superchip, Slingshot-11, HPE Cray OS, HPE Swiss National Supercomputing Centre (CSCS) Switzerland	2,121,600	434.90	7,124	61.047

The Swiss AI initiative

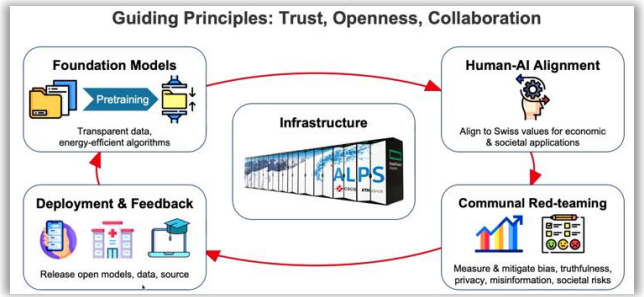


Develop capabilities, know-how & talent to build trustworthy Generative AI aligned with Swiss values

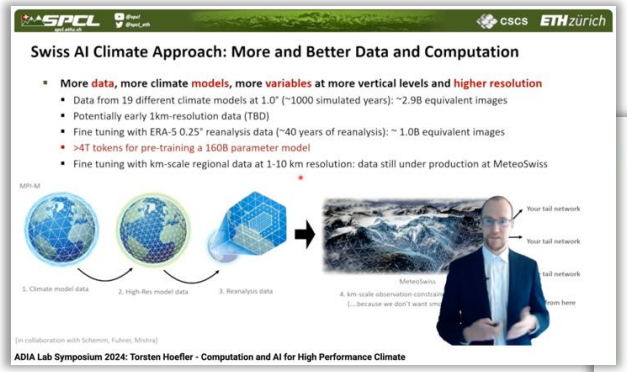
Make these resources available for the benefit of Swiss society and global actors

Machine Learning on Alps and the Swiss AI Initiative

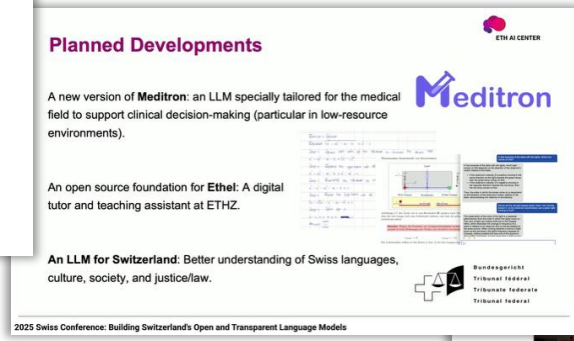
www.swiss-ai.org



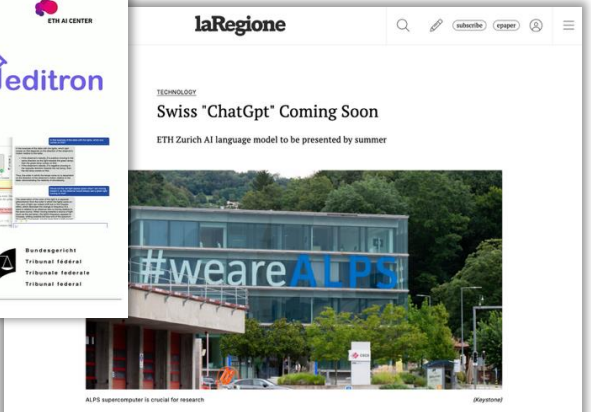
- Our efforts are guided by our tight collaborations with world-class ML experts working across scientific domains.
- The Swiss AI Initiative is a current and prominent example of that.
  - 80+ research groups, representing a wide range of domains.
  - Wide spectrum of model architectures, usage modes, datasets, scales and scaling libraries, and project phases.
- Through the SwissLLM execution, about half of Alps (through multiple vClusters) has been dedicated to the ML community.



"Computation and AI for High Performance Climate" by Torsten Hoefler - https://www.youtube.com/watch?v=QLmqJmj39qs



"Building Switzerland's Open and Transparent Language Models" by Imanol Schlag https://www.youtube.com/watch?v=gjvakQQOCdw

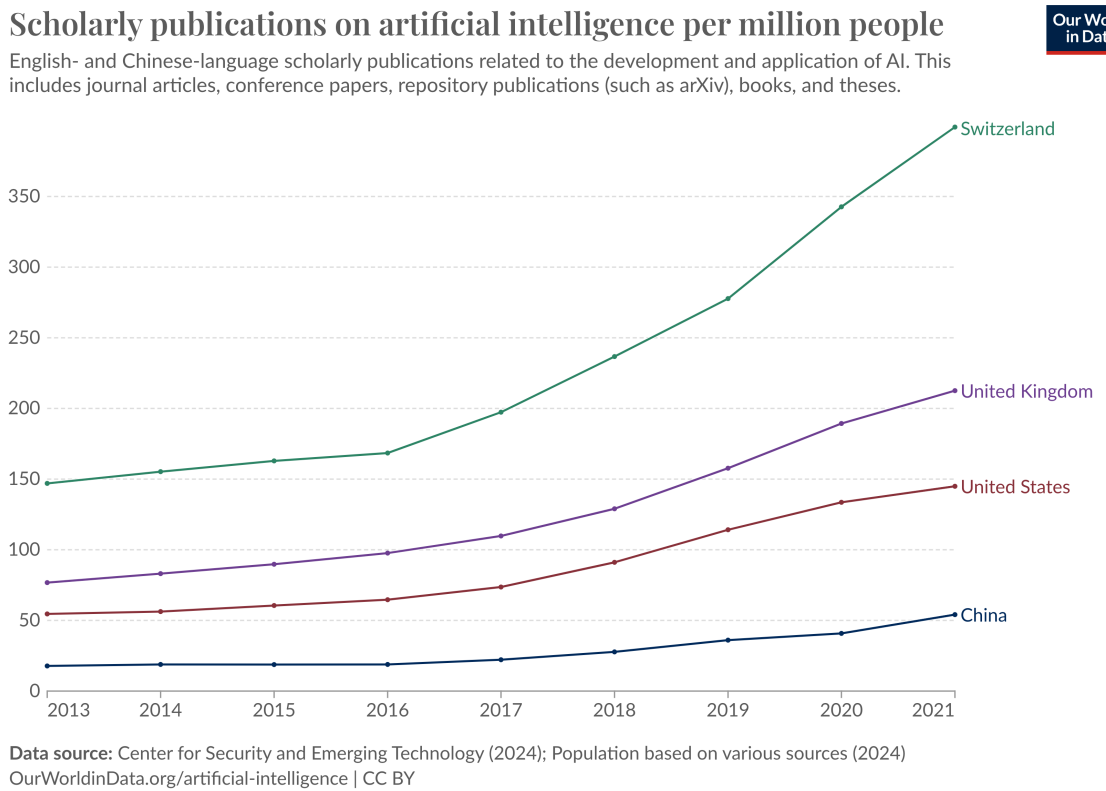


From "*Community access for ML on ALPS: The Swiss AI Initiative*" *by Joost VandeVondele* - https://www.youtube.com/watch?v=xLuuE0HuXAU

slated from https://www.laregione.ch/rubriche/tecnologia/18305519/zurigo-politecnico-svizzero-modello-ia

CSCS

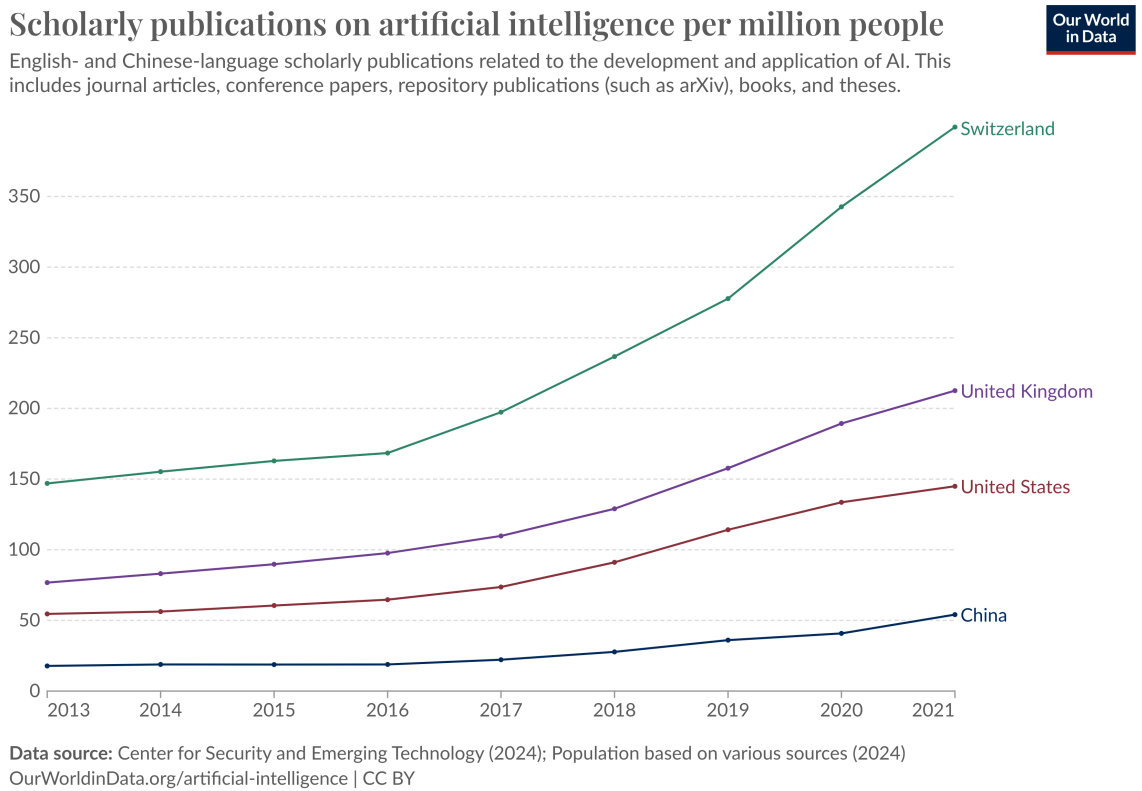
Swiss talent in AI

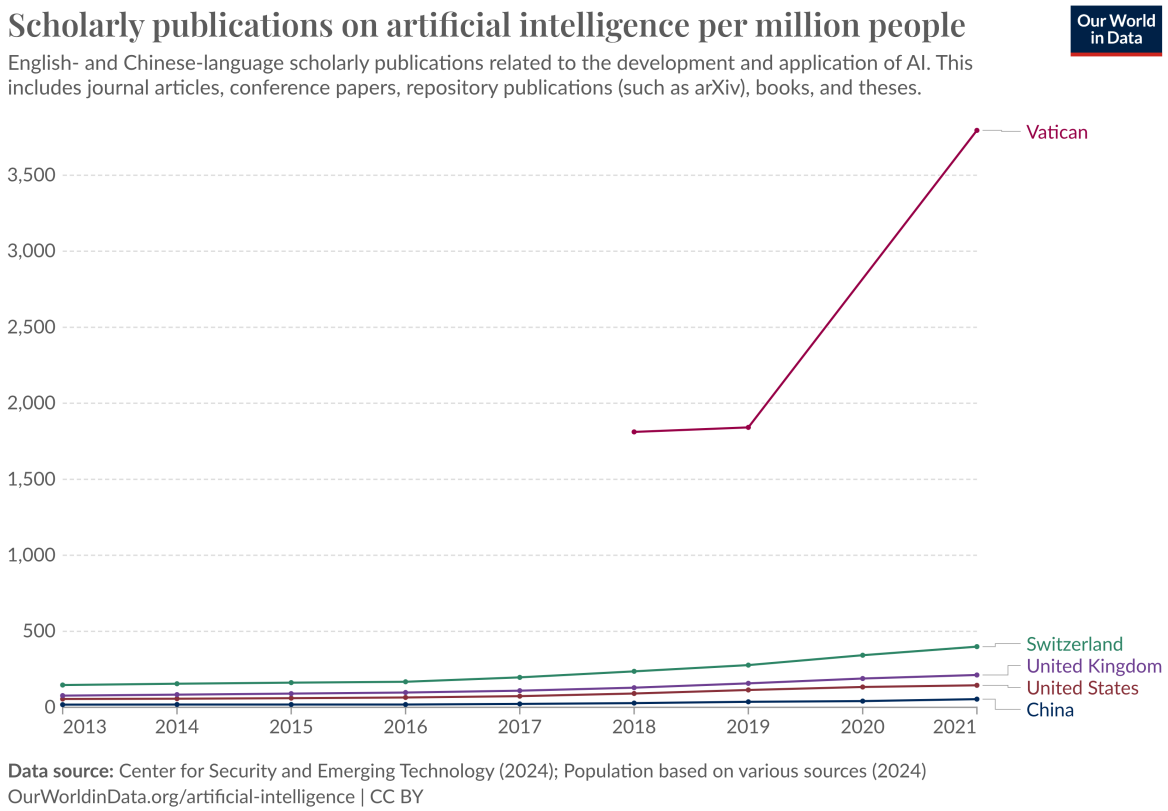


The impact of a strong educational system:

- Federal institutes of technology
- Cantonal Universities
- Universities of applied sciences.

Swiss talent in AI





But beware of small denominators when analyzing data…

Swiss AI structure

Verticals

Foundation model for sciences
Prof. Brbic, Prof. Schwaller, Prof. Marinkovic

Foundation model for education
Prof. Käser, Prof. Sachan

Foundation model for ego-centric vision & robotics
Prof. Alahi, Prof. Pollefeys, Prof. Katzschmann

Foundation model for health
Prof. Rätsch, Prof. Salathé, Prof. Fellay

Foundation model for sustainability / climate
Prof. Mishra, Prof. Schemm, Prof. Hoefler, Prof. Schindler, Prof. Tuia

Horizontals

Fundamentals of foundation models
Prof. Yang, Prof. He, Prof. Zdeborova, Prof. Flammarion

LLM security, red teaming & privacy
Prof. Troncoso, Prof. Tramèr

Tools & infrastructure for scaling
Prof. Klimovic, Prof. Falsafi

Human-AI alignment
Prof. Ash, Prof. Gulcehre

Large-scale multi-modal models
Prof. Cotterell, Prof. Zamir

Advanced LLMs
Prof. Bosselut, Prof. Jaggi, Dr. Schlag

- **Uniting researchers in Switzerland to tackle challenges out-of-reach of single research groups**
- **70 professors, > 500 PhD and Postdocs**
- **Lead by the ETHZ and EPFL AI centers**
- **Involving many of the Swiss universities and universities of applied sciences.**

ETH AI CENTER

EPFL AI Center

Universität Zürich[UZH]

IDSIA SUPSI

Hes·so VALAIS WALLIS

idiap RESEARCH INSTITUTE

UNI FR UNIVERSITÉ DE FRIBOURG UNIVERSITÄT FREIBURG

u^b UNIVERSITÄT BERN

UNIVERSITÉ DE GENÈVE

zhaw School of Engineering CAI Centre for Artificial Intelligence

CSCS

https://www.swiss-ai.org/

CSCS

ETH zürich

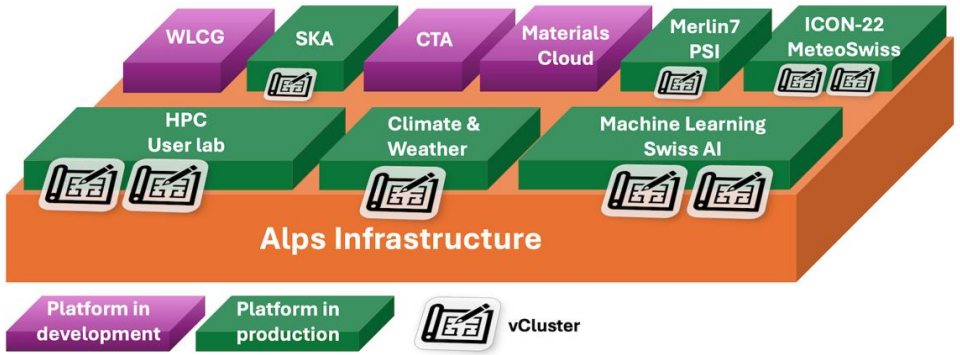
(Part of) Our Jurney So Far

- In 2022 we needed to prepare Alps to welcome this new community
  - CSCS had previous experiences with large ML workloads (incl. https://arxiv.org/abs/2110.11466), but this time it was different: it was a national effort involving a much larger community.

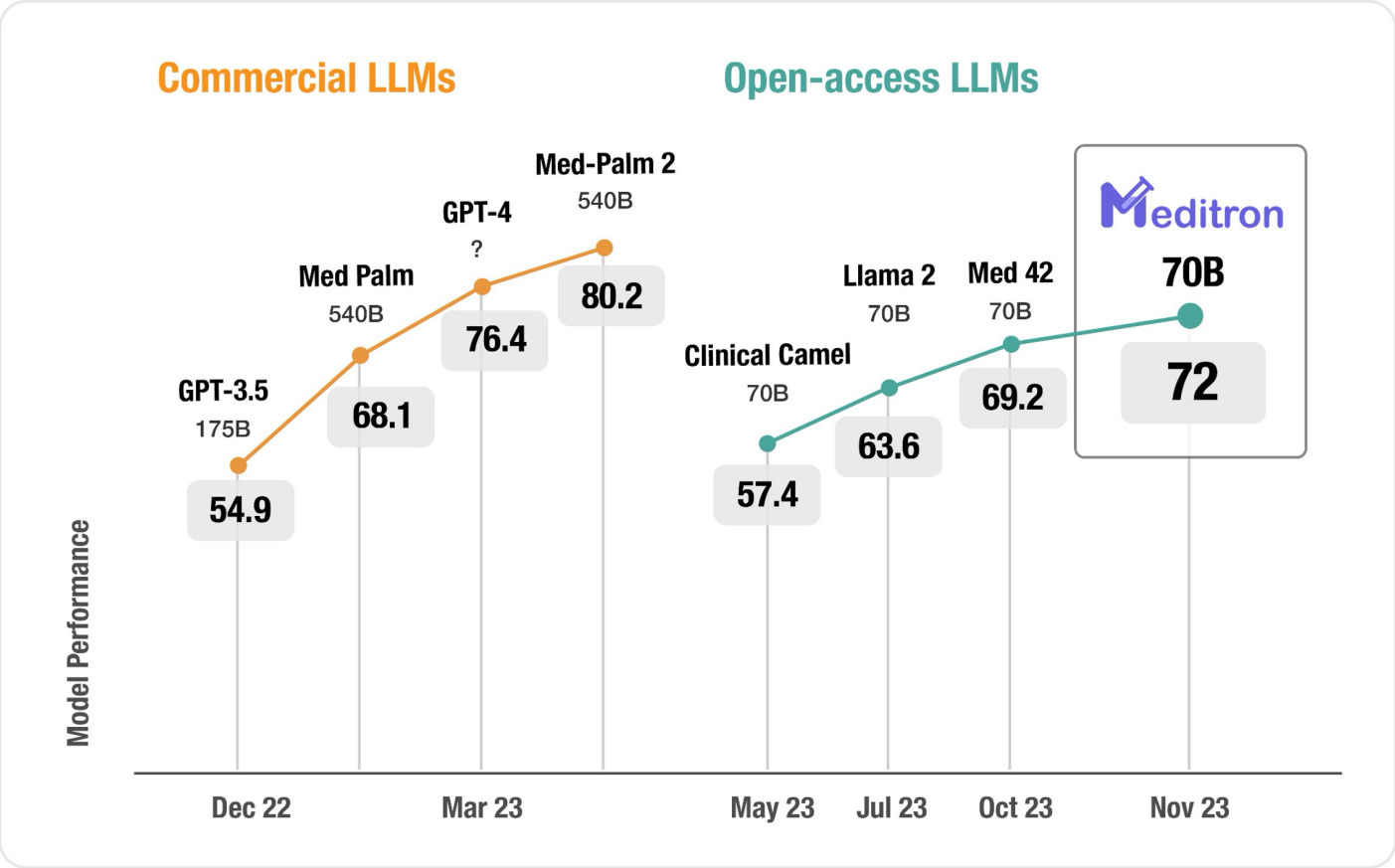
- By asking users for requirements, following the community trends, and even joining users' projects, we realized that interests evolve too fast to keep up effectively (e.g., 3Dp, GNNs, Attention mechanisms, MoE, Agents, …).

- By mid'24, we formed guiding principles for our ML platform design efforts:
  - We need to do something sufficiently general, i.e. relevant for all projects, model architectures, and for different project phases.
  - It needs to be compatible with our engineering capacity.
  - It should be resilient to the change of time, e.g., community interests shifts, future infrastructure updates.

- By now, quite a lot of experience accumulated through the SwissLLM effort.



Early Successes: Meditron

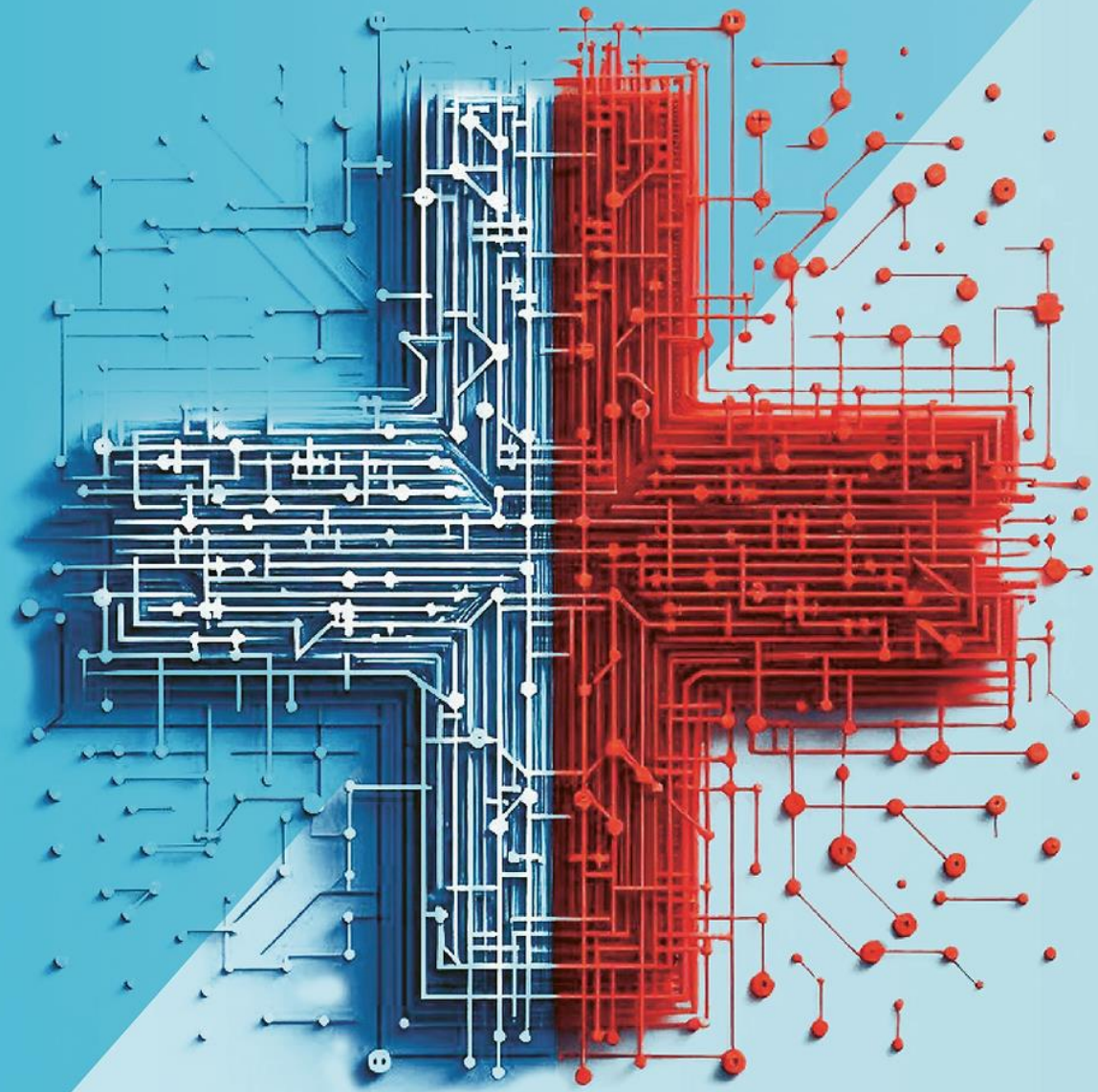


Transparent:
Open Data
Open Source
Open Weights

Performant:
Best open-source MedLM
In range of most powerful commercial LLMs

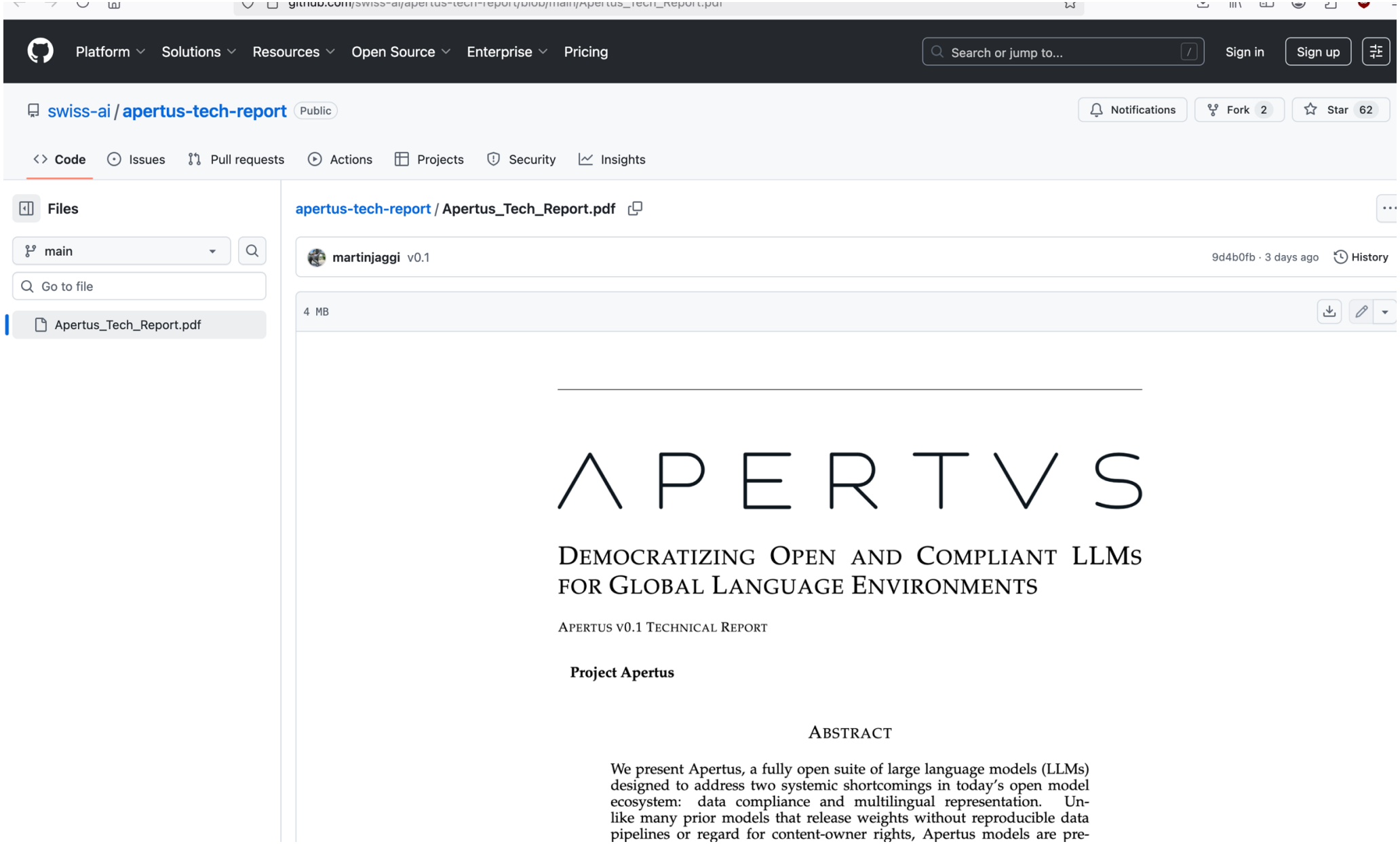
One use case: made available by Mary-Anne "Annie" Hartley and co-workers to doctors in Africa. Assessing and avoiding bias key for establishing trust and having impact.

https://arxiv.org/abs/2311.16079



APERTVS

EPFL ETH zürich CSCS


Platform
Solutions
Resources
Open Source
Enterprise
Pricing
Search or jump to...
Sign in
Sign up
swiss-ai / apertus-tech-report
Public
Notifications
Fork 2
Star 62
Code
Issues
Pull requests
Actions
Projects
Security
Insights
Files
main
Go to file
Apertus_Tech_Report.pdf
apertus-tech-report / Apertus_Tech_Report.pdf
martinjaggi v0.1
9d4b0fb · 3 days ago
History
4 MB
APERTVS
DEMOCRATIZING OPEN AND COMPLIANT LLMS FOR GLOBAL LANGUAGE ENVIRONMENTS
APERTUS V0.1 TECHNICAL REPORT
Project Apertus
ABSTRACT
We present Apertus, a fully open suite of large language models (LLMs) designed to address two systemic shortcomings in today's open model ecosystem: data compliance and multilingual representation. Unlike many prior models that release weights without reproducible data pipelines or regard for content-owner rights, Apertus models are pre-

Full Training Run Performance (see Apertus report)

A detailed timeline showing token throughput performance over the pretraining runs of the 70B and 8B Apertus models is displayed in Figure 10. We estimate that training of the 70B model for 15T tokens took 6.74×10^{24} FLOPs (details in Appendix D). In terms of usage, the main run has consumed around 6 million GPU hours, though this is underestimated as it does not count loading weights or overhead due to initial performance short-comings, failures or downtime. Once a production environment has been set up, we estimate that the model can be realistically trained in approximately 90 days on 4096 GPUs, accounting for overheads. If we assume 560 W power usage per Grace-Hopper module in this period, below the set power limit of 660 W, we can estimate 5 GWh power usage for the compute of the pretraining run. CSCS is almost carbon neutral, relying entirely on hydropower, and uses a sustainable cooling system that uses water from Lake Lugano in a closed cycle, with all the water returned to the lake and none consumed.[48]

vClusters via the (shared) filesystem

Swiss AI Initiative and the SwissLLM Pre-Training

Strong Scaling		
Nodes	Global Batch Size	TPSG
8	4096	892.53
16	4096	884.49
32	4096	876.74
64	4096	869.95
128	4096	856.31
256	4096	835.62
512	4096	790.70
1024	4096	722.82

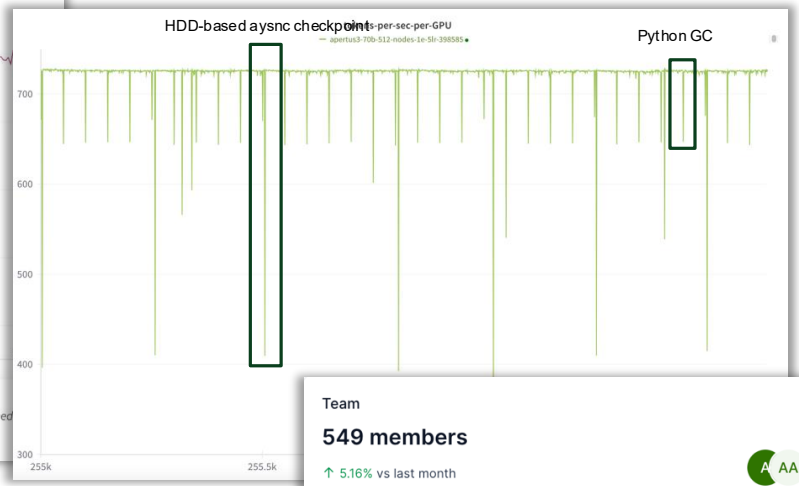
Weak Scaling		
Nodes	Global Batch Size	TPSG
8	32	819.20
16	64	800.57
32	128	792.07
64	256	783.75
128	512	773.12
256	1024	745.17
512	2048	761.65
1024	4096	722.82

Apertus: strong and weak scaling to 1024n / 4096 GPUs.
Data: Antoni-Joan Solergibert i Llaquet

Meta: Llama 70B, 1856 H100, Meta, ckpt/25 iters

https://pytorch.org/blog/6x-faster-async-checkpointing

CSCS: Apertus 70B, 2048 GH, ckpt/250 iters



A minor version upgrade of the mgmt plane

Cleaning of heat exchanger + major upgrade

Lightning strike near pumping station?

CSCS

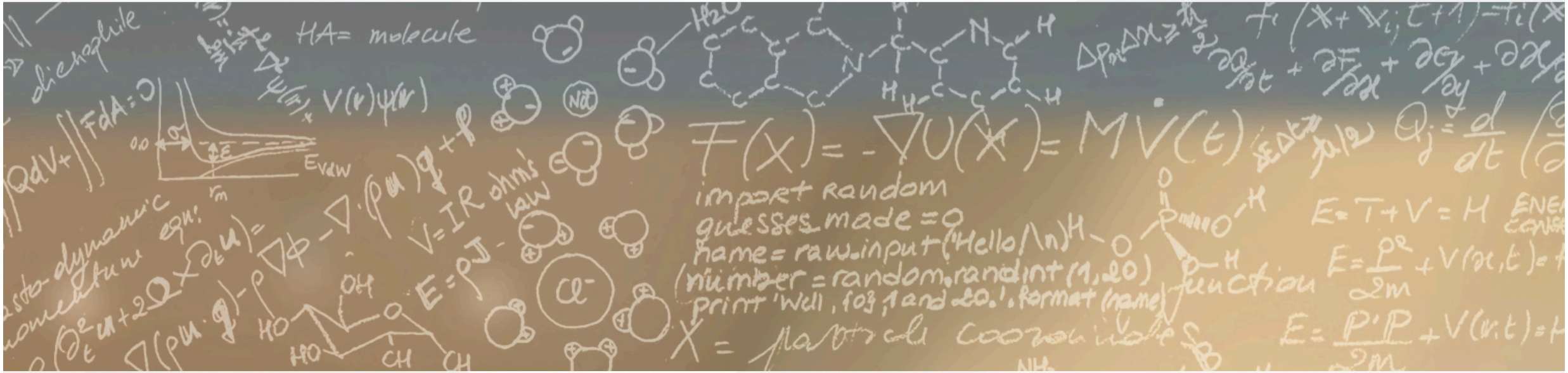
ETH zürich

Conclusions

- With Alps, CSCS has deployed a world-class AI capable supercomputer
- Infrastructure is key to modern AI and foundational models
- The Swiss AI initiative unites and will generate top talent in AI
- The initiative has a model for public-private partnerships

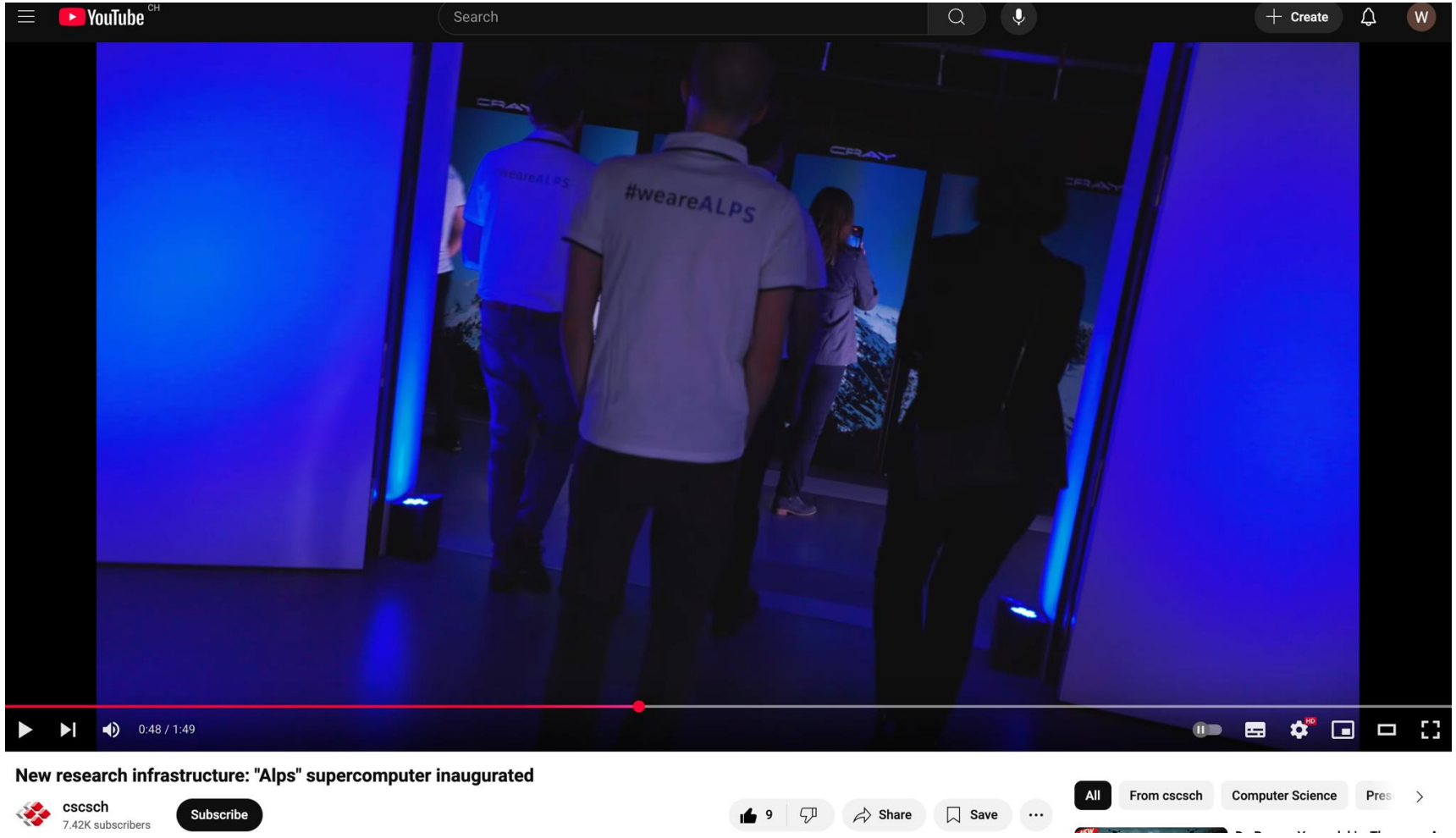
CSCS
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

ETH zürich



Thank you for your attention

Watch Me! https://youtu.be/yz9aCTG1mck?si=CK2urRgcGbASVeZN&t=48



Watch Me! https://www.youtube.com/watch?v=jGVRAK8KIBg