Supercomputing Services e-INFRA CZ

Current status and plans

Vít Vondrák
IT4Innovations, VSB-TUO

cesnet
MUNI CERIT-SC
VŠB TECHNICKÁ UNIVERZITA OSTRAVA | IT4INNOVATIONS NÁRODNÍ SUPERPOČÍTAČOVÉ CENTRUM

Basic supercomputing services

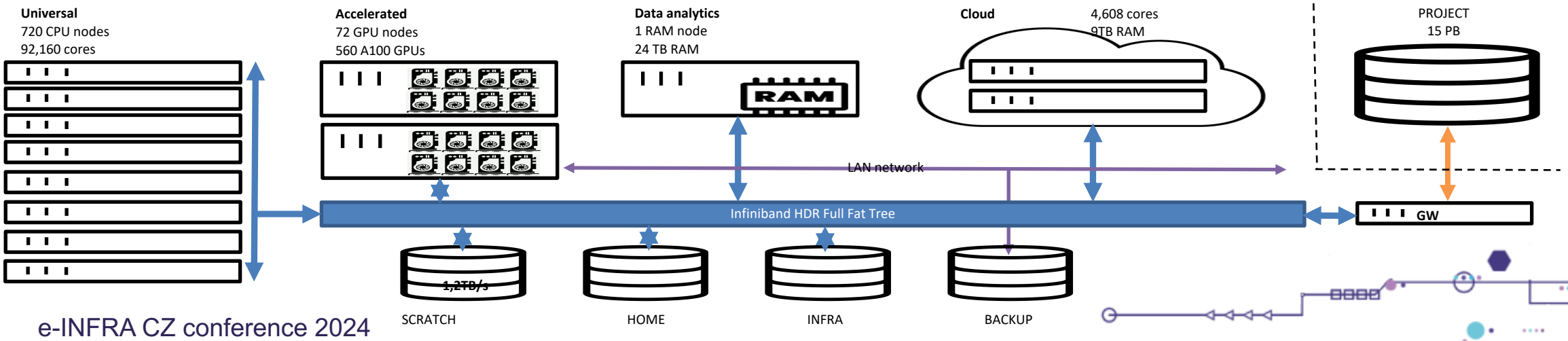
Services provided by **IT4Innovations National Supercomputing Center @ VSB-Technical University of Ostrava**

- Operation and provisioning supercomputing and data resources to the national scientific community and industry
  - **Open access** - Allocation based on open grant competition
  - Thematic access - Societal end economic needs incl. cooperation with industry, education, …
- National node of EU HPC & Data infrastructures
  - EuroHPC Joint Undertaking (EuroHPC JU petascale + pre-exascale + NCC)
  - Partnership for Advanced Computing in Europe (PRACE)
  - EUDAT Collaborative Data Infrastructure
  - Memberships in ETP4HPC, BDVA, iRODS, EOSC association
- Training and educational activities
- User support (incl. application level support)

VSB TUO | IT4I

Karolina supercomputer

- In operation from 2021
- EuroHPC supercomputer
  - 65% Czech funding
  - 35% EU funding (for EuroHPC users)
- Total investment approx. EUR 15 million
- Total theoretical performance 15.7 PFlop/s
- Used by more than 1,700 users to solve more than 700 projects.
- Expected end of operation for EuroHPC users 2026



Karolina upgrade

- Software upgrade scheduled from 8.4.2024 for about 4 weeks
  - Operating system from CentOS 7.9 to Rocky Linux 8.9
  - SCRATCH storage backend from ClusterStore 4.5 to ClusterStore 6.6
  - Management software HPCM 1.4 to HPCM 1.11
- Main impact to users:
  - Rocky Linux 8 **modern up-to-date operating system** on login, compute, and visualization nodes
  - Updated the operating **system kernel**, **system libraries**, system **utilities** and **application libraries**
  - Rebuilt and updated user applications **/apps**
  - Updated **SLURM 23.02.7** scheduler
  - Improved **SCRATCH** performance and stability
  - Availiability of **LibSci** BLAS libraries
  - Updated **GPU and MOFED** drivers
  - Enhanced system **security**

Given the extensive nature of the upgrade, it essentially entails a **complete reinstallation** of the Karolina supercomputer.

Karolina cloud partition

E-INFRA CZ cloud

Key features:

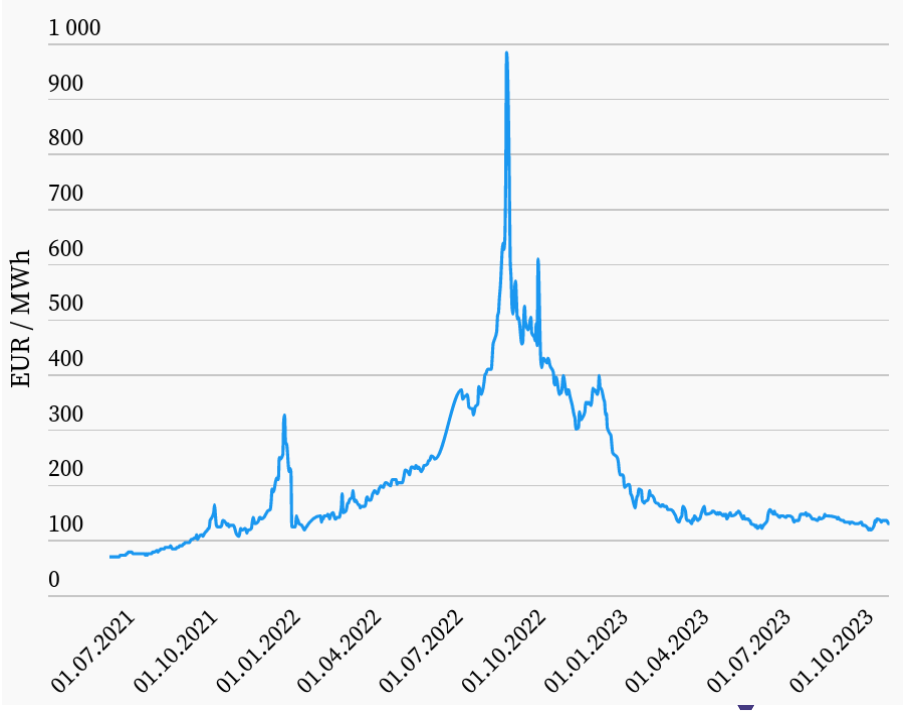
- 22 nodes from Karolina supercomputer, 2816 phys. CPU cores, 5.5TB RAM
- Based on OpenStack
- GUI access: https://ostrava.openstack.cloud.e-infra.cz/
- Documentation: https://docs.it4i.cz/cloud/einfracz-cloud/
- Available for e-INFRA users with active projects at IT4Innovations National Supercomputing Center
- Users have default, limited resources. Apply to increased resources within IT4I Open Access Competition

The main advantages:

- Conveniently accessed via a web interface
- Scalability – virtual machines (number and configuration) can be created as needed
- Resource flexibility – root privileges on virtual machine. You can install your own programs, set quotas, run tasks
- Orchestration – VM visible from HPC nodes, VMs can orchestrate the HPC jobs
- Scale-out - you can use OpenStack tools (Terraform) to automate configuration

Motivation to save energy

- Energy price raised significantly in 2022 and beyond with huge impact on operation of our clusters
- IT4I operates with reduced budget for support and operation of the large research infrastructures
- Implies necessity to reduce energy cost, which means we must improve energy efficiency of the infrastructure
- In general reducing carbon footprint
- From 1.2. 2023 Karolina cluster is using hardware configuration increasing energy efficiency

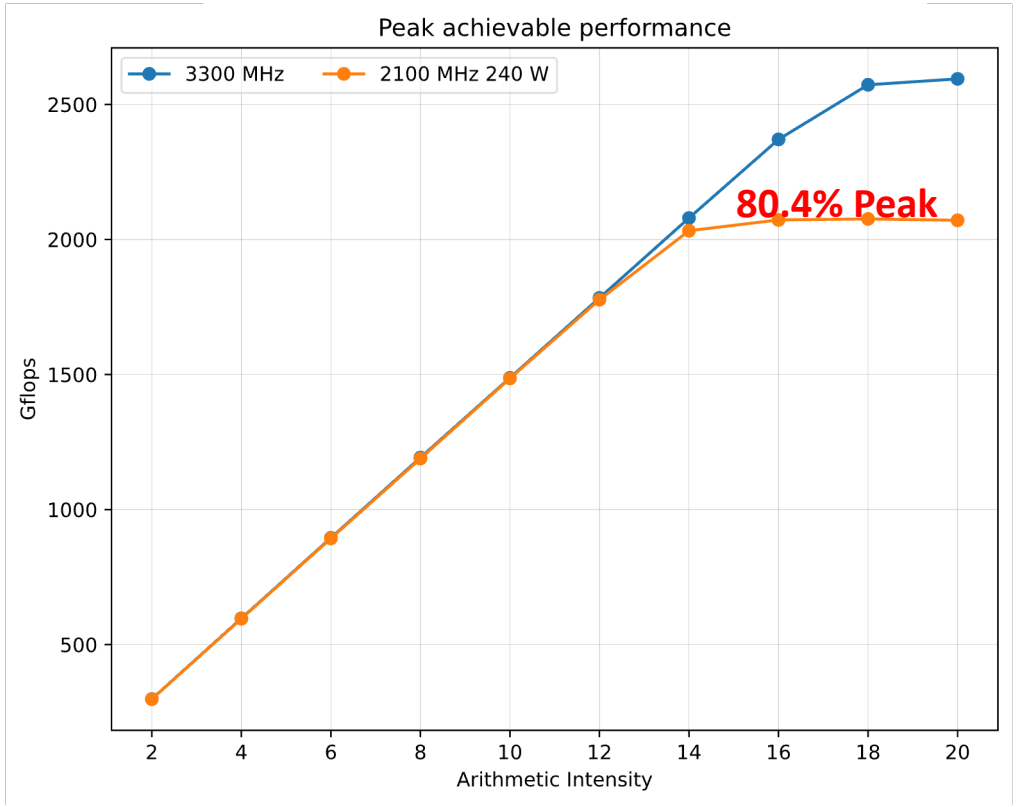


https://www.e15.cz/cena-elektriny-burza-graf-aktualne

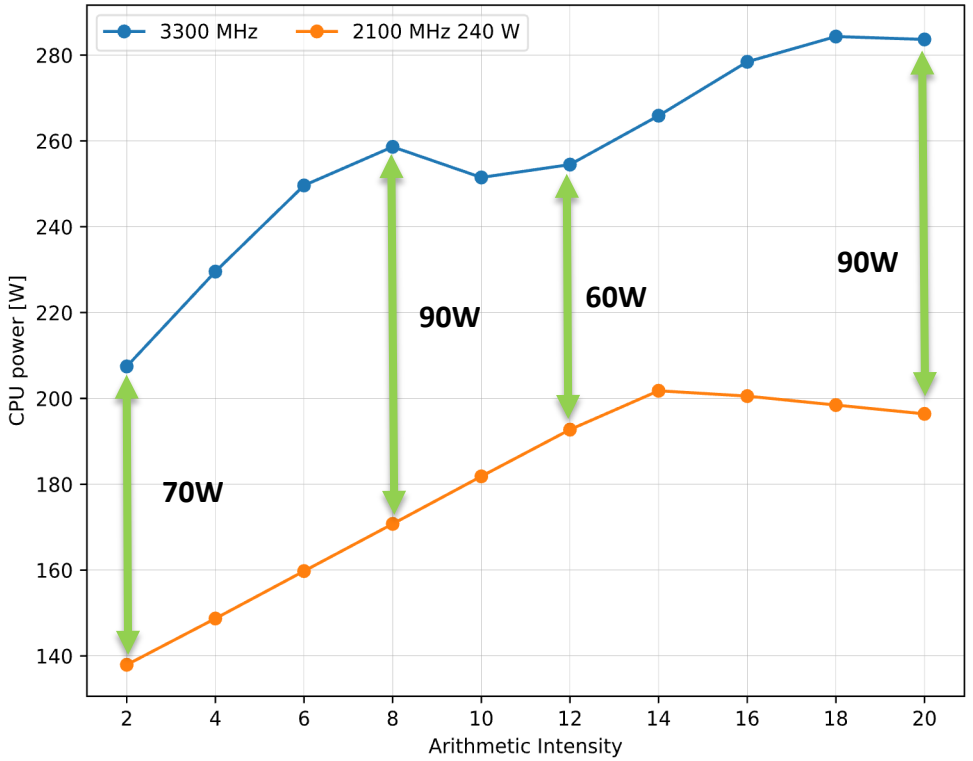
Karolina energy optimization

AMD 7H12 AVX2 BOOST LIMIT

Performance [GFlops]



Power [W]

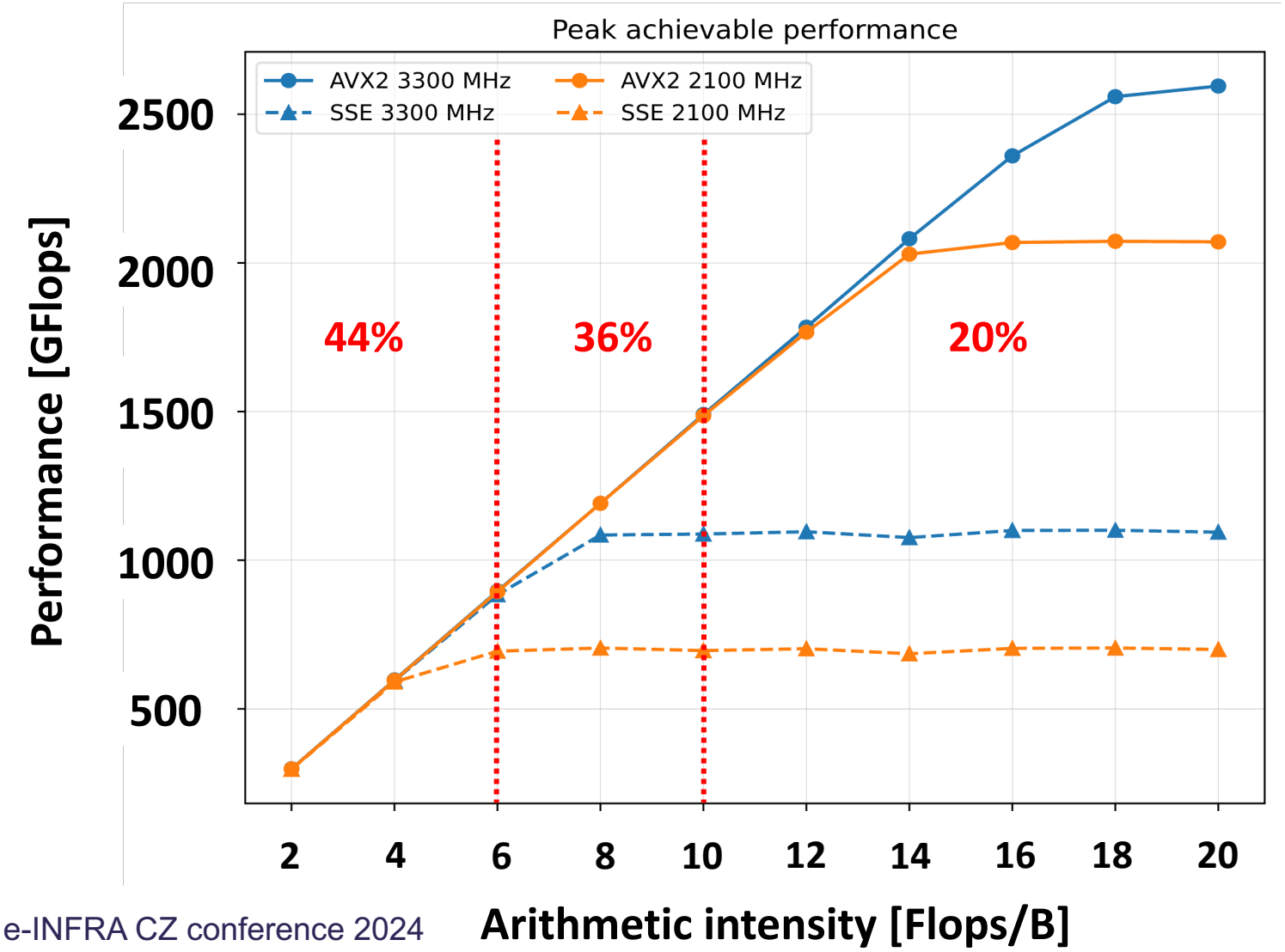


default

2100 MHz

Roofline model

AMD 7H12 – AVX2 + SSE

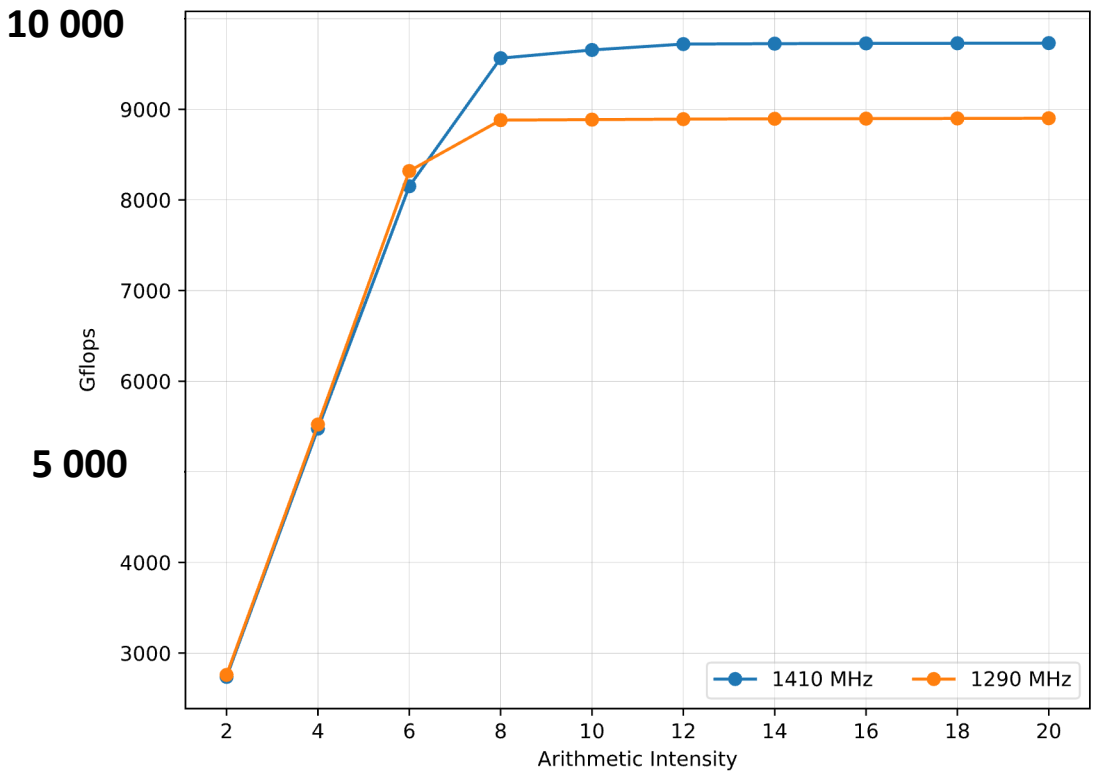


AVX2 default	100%
AVX2 2.1 GHz limit	80.4%
SSE default	43.1%
SSE 2.1 GHz limit	27.5%

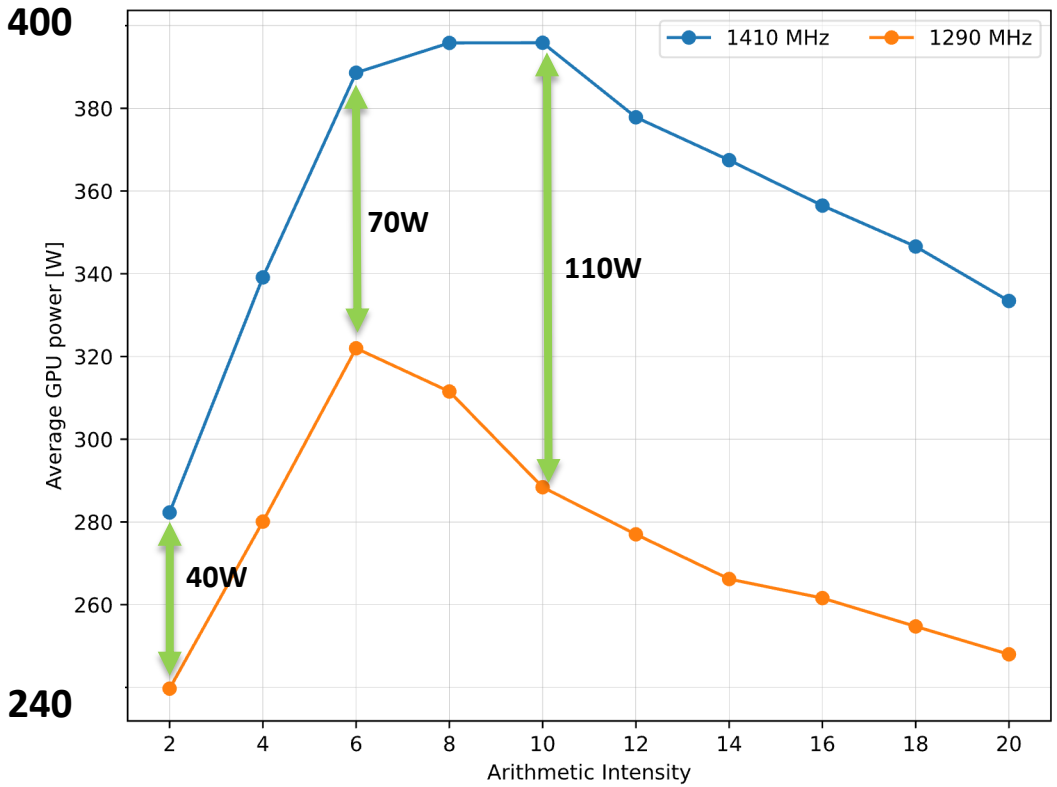
Karolina energy optimization

NVIDIA A100 (40GB HBM2)

Performance [GFlops]



Power [W]



default

2100 MHz

Energy optimization result

Karolina CPU partition

- Utilization **72.7% -> 88.4% (+15.7%)**
- Elimination of Free and Preempt queues
- System average power (without cooling) **327 kW -> 264 kW (63 kW, -19.3%)**
- Expected power consumption for 88.4% utilization 351 kW => 87 kW savings + additional power savings from reduced cooling
- Equivalent to turning off **220 fully utilized nodes (31% out of 720 nodes)**
- 87 kW equals to 743 MWh / year
- 1MWh ~ 6000 CZK => 4.5 M

Karolina GPU partition

- Utilization **56.9% -> 72.2% (+15.3%)**
- System average power (without cooling) **141 kW -> 139 kW (2 kW, -1.4%)**
- Expected power consumption for 72.2% utilization 155 kW => 16 kW savings (10.3%) + additional power savings from reduced cooling
- Equivalent to turning off **4 fully utilized nodes (6% out of 72 nodes)**
- 16 kW equals to 140 MWh / year
- 1 MWh ~ 6000 CZK => 840 k CZK

Complementary systems

Key features:

- Represent emerging, non-traditional, and highly specialized hardware architectures that are not yet common in Czech Republic or Europe supercomputing data centers.
- Enable new programming models, libraries, and application development tools
- Allow research teams to test and compare with traditional architectures
- Available for **all e-INFRA users with active projects at IT4Innovations**

Complementary systems I

- Fujitsu ARM A64FX,
- Intel FPGA (Altera),
- AMD FPGA (Xilinx),
- Edge server (NVIDIA Tesla T4 GPU)

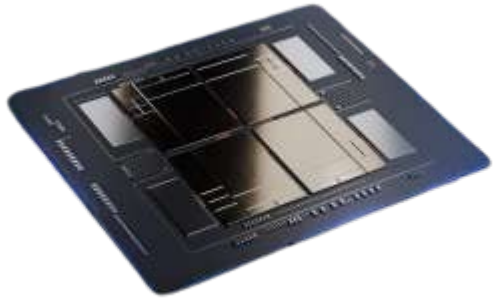
Complementary systems II

- ARM + Nvidia GPU + DPU,
- IBM Power10,
- AMD Milan with very large L3 cache,
- Virtual Desktop Infrastructure (VDI) virtual GPU accelerated workstations,
- **Intel Sapphire Rapids HBM**
- **NVIDIA Grace CPU Superchip**

Complementary systems – latest additions

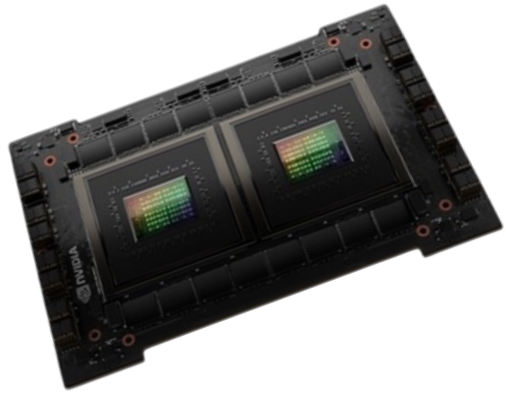
Intel Sapphire Rapids HBM node:

- 2x Xeon CPU Max 9468, 2.10GHz – **most powerful CPU** at IT4I
- 128GB HBM2 memory - **highest CPU to RAM bandwidth** at IT4I
- Unique (at IT4I) combination of HBM and DDR5 memory



NVIDIA Grace CPU Superchip node:

- ARM CPU, 144 cores, 3.1GHz and 1 TB/s RAM BW, **second most powerful CPU** at IT4I
- **ARM and NVIDIA Linux software ecosystem** incl. NVHPC libs and compilers
- ARM architecture: **ASIMD and SVE, 128bit**
  Neon Advanced SIMD ASIMD and Scalable Vector Extensions SVE, 128bit registers
- **4 FMA or FMLA vector instructions** per clock cycle
- Fast **RAM (LPDDR5X)** via NVIDIA NVLink-C2C technology

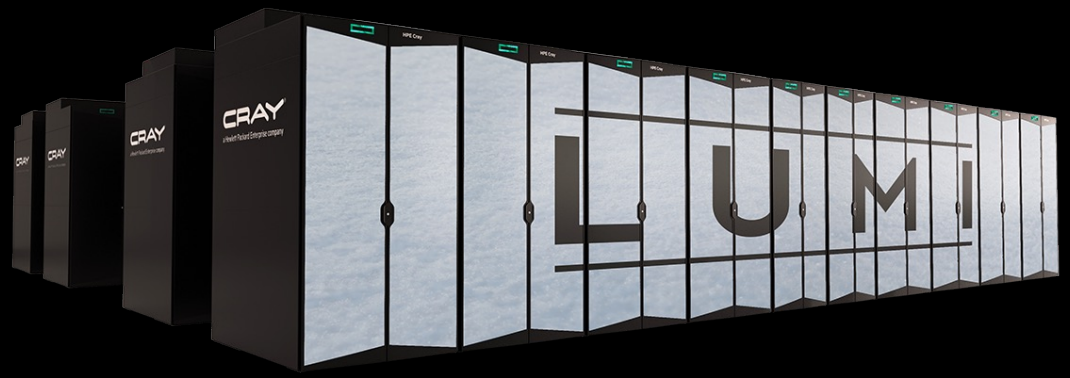


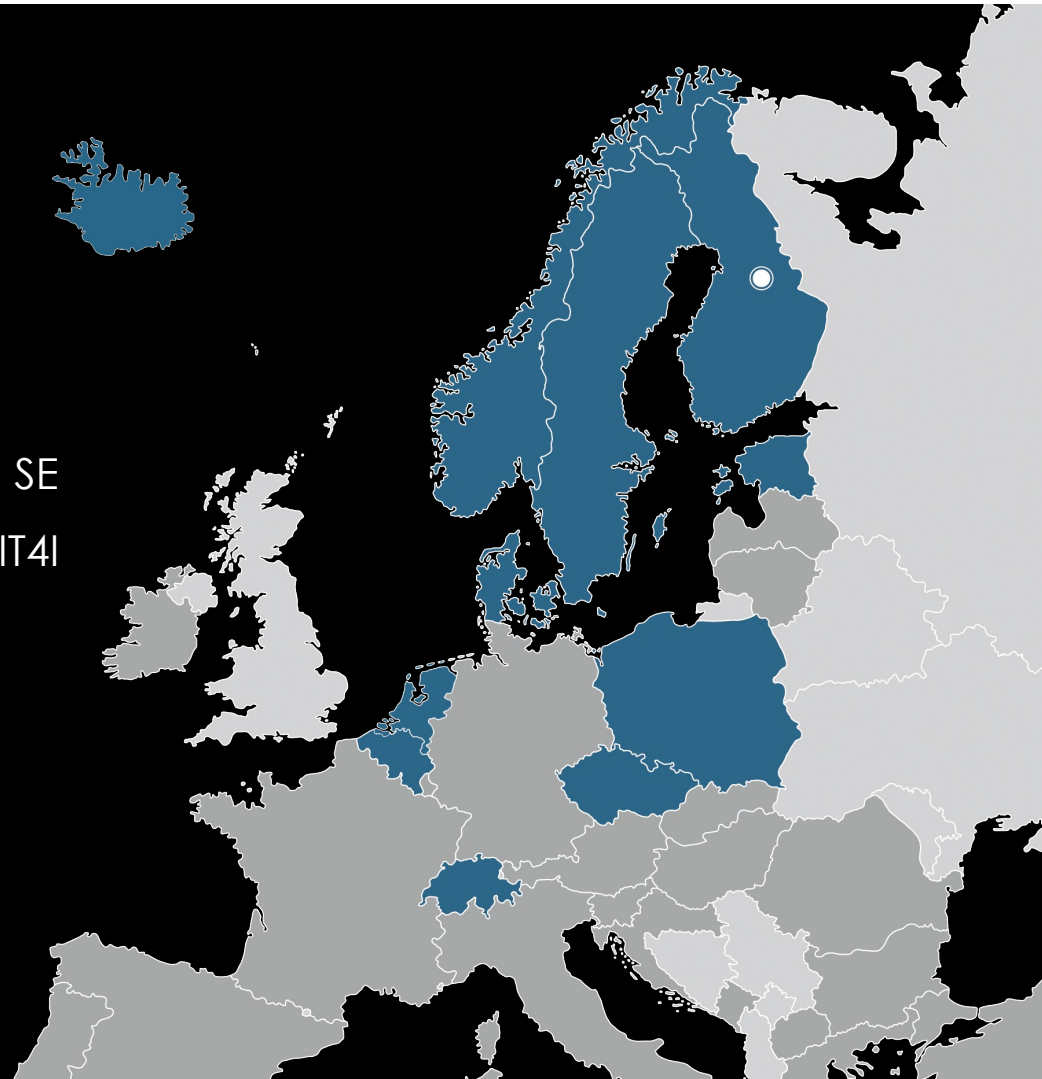
Unique opportunity to migrate applications, test performance, and perform optimization.

EuroHPC
Joint Undertaking

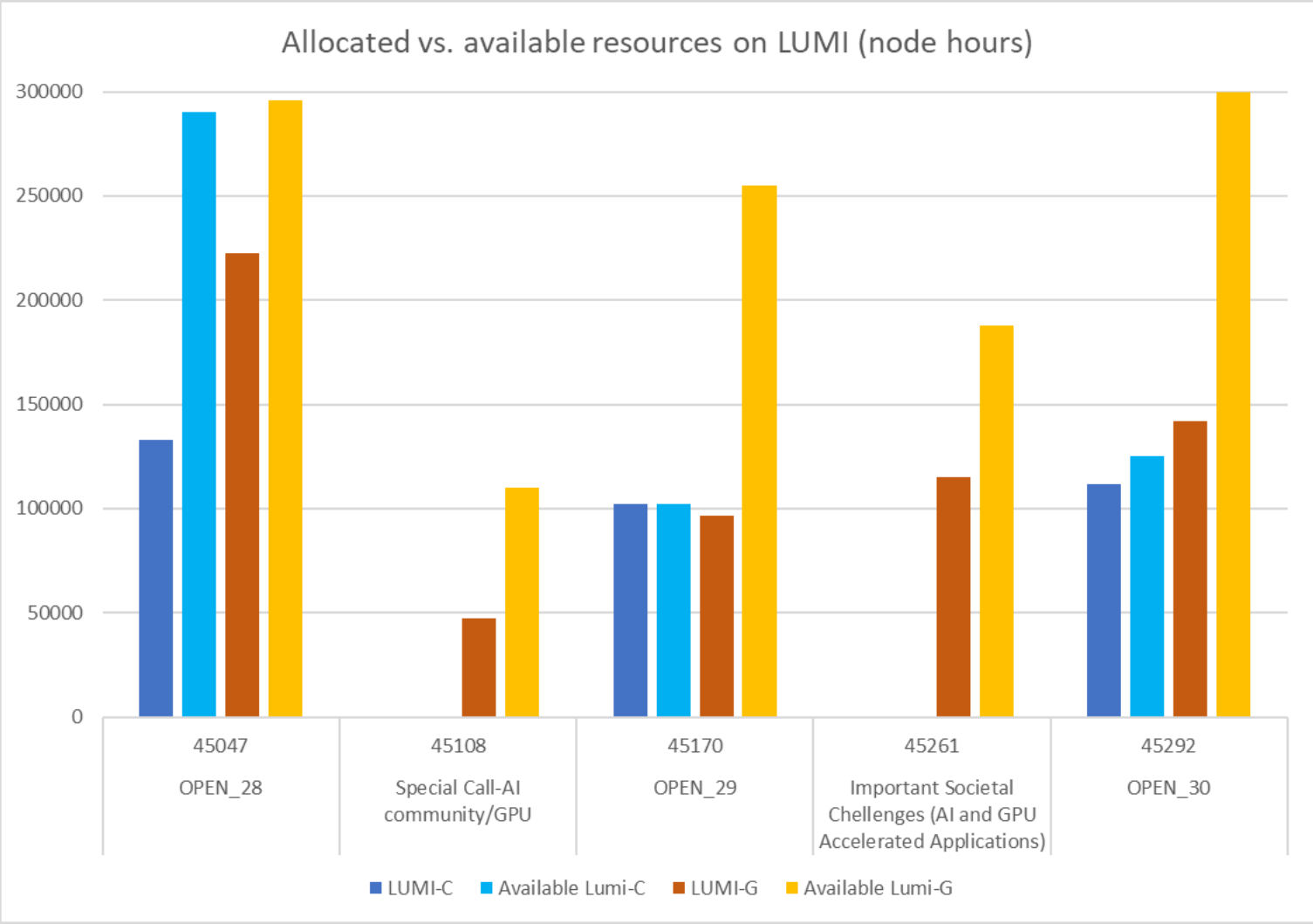
LUMI

- Manufactured by **Hewlett Packard Enterprise**
- Total budget: 207.1M€ **(50 % EuroHPC JU)**
- **LINPACK Performance: 379,7 Pflop/s**
- #5 in TOP500, #7 in GREEN500, and **#1 in Europe**
- Period of operation: 2021–2027
- Consortium: **FI (CSC in Kajani)**, BE, CH, **CZ,** DK, EE, IS, NL, NO, PL, SE
- Ca **2,5% of the resources available to the Czech users** through IT4I Open Access calls





LUMI national allocations



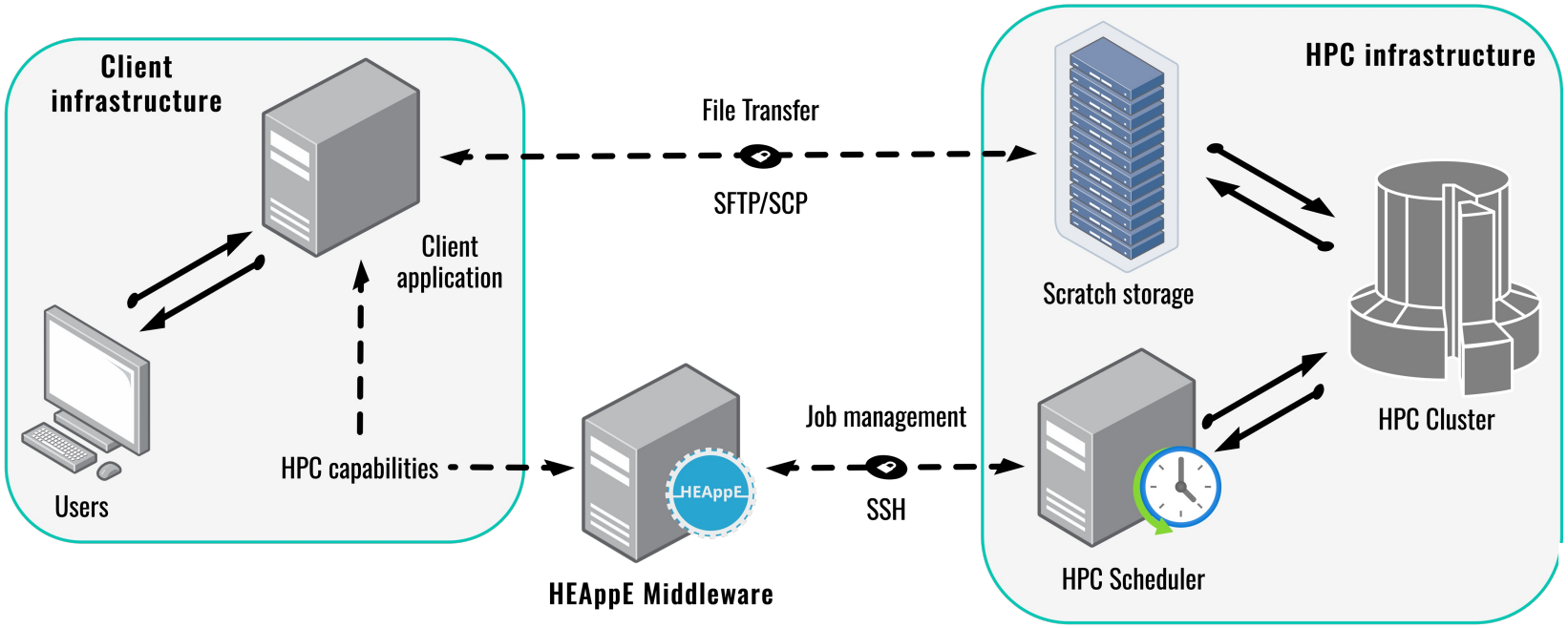
Great for PyTorch and AI

- PyTorch is an optimized tensor library for deep learning using GPUs and CPUs.
- PyTorch/LUMI-G proven in many successful projects, **incl. 70 Billion parameter** OLMo LLM.
- Comprehensive Guide on PyTorch at LUMI is available at¶ https://docs.it4i.cz/lumi/pytorch/
- **Do not miss Lukas Prediger talk on PyTorch on LUMI**

HEAppE middleware

High-End Application Execution Middleware

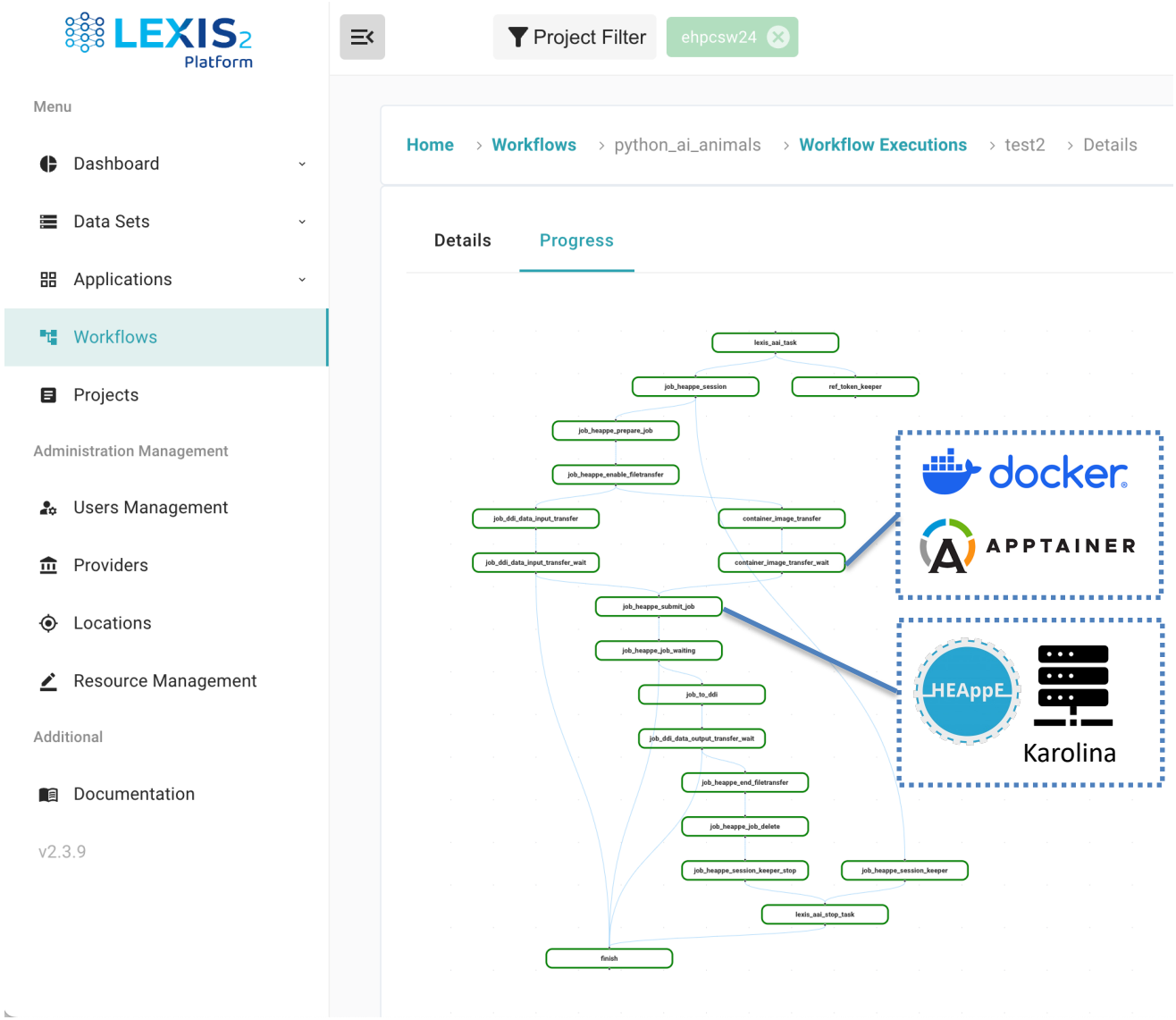
- Providing HPC capabilities as a service to client applications and their users
- Secured and restricted access to HPC infrastructure
- REST API for easy access and integration
- Authentication and authorization to provided functions
- Monitoring and reporting of executed jobs and their progress



LEXIS2 Platform allows containers on HPC

- Run your AI workloads on Karolina supercomputer from your browser
- Upload your AI container and input dataset in the LEXIS Platform
- Select input dataset and set run-time parameters of your container through the LEXIS workflow execution
- Monitor the execution & view logs from the container running on HPC
- Download and examine the results https://portal.lexis.tech

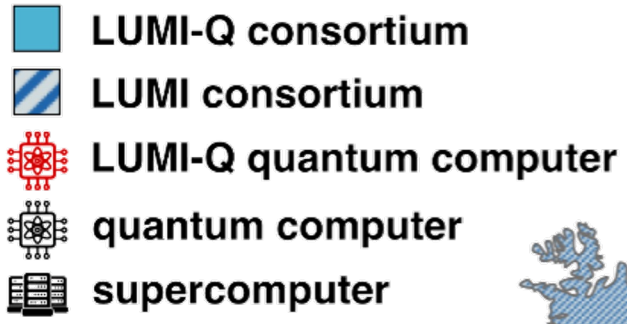
- **More tomorrow in Martin Golasowski presentation**

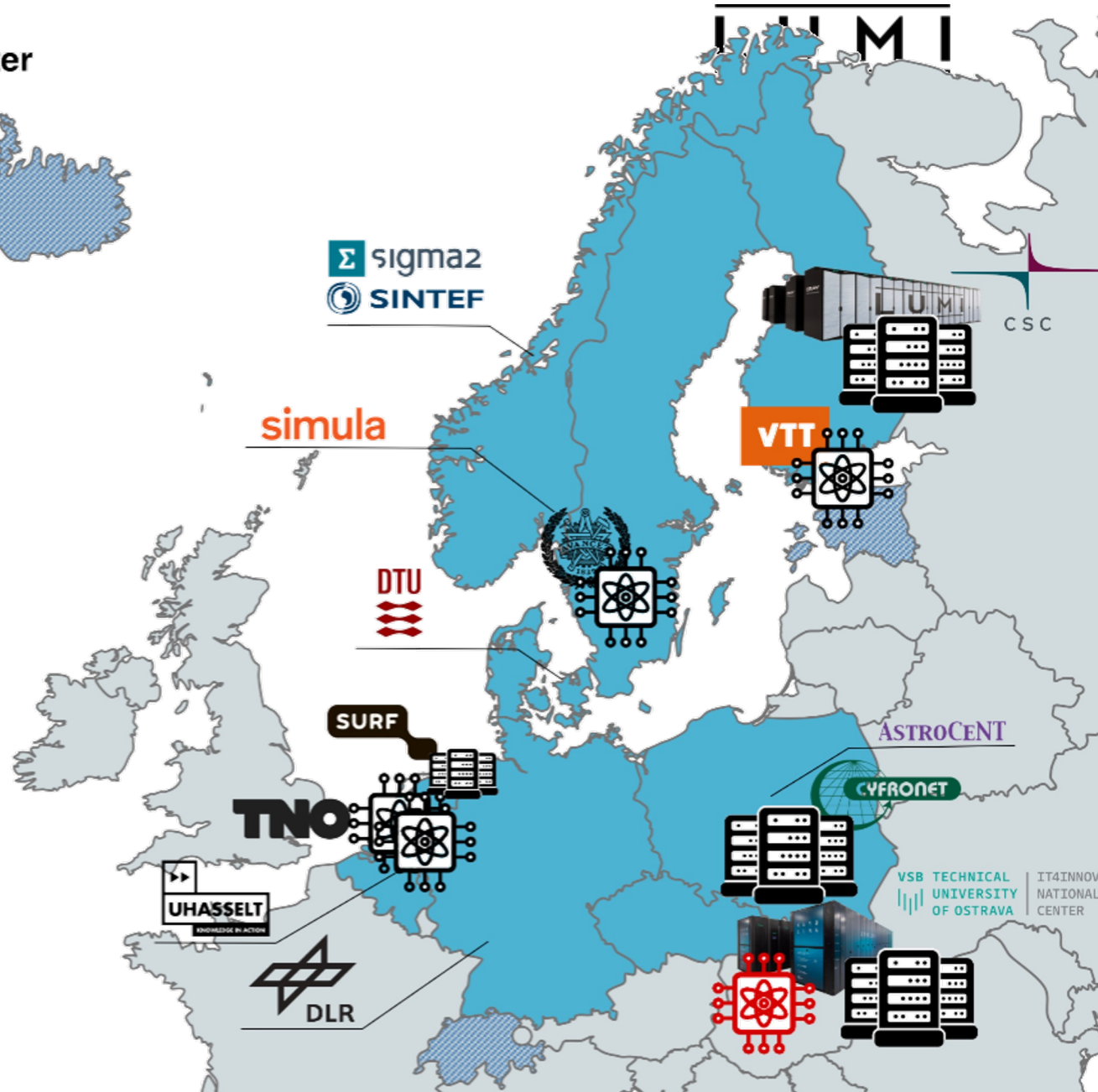


What we plan to install and operate

At IT4Innovations@VSB-TUO

LUMI-Q





Inclusive

- Builds on the 9-country pan-European LUMI consortium + The Netherlands and Germany
- The procured quantum computer is a quantum computer for Europe, not a single country

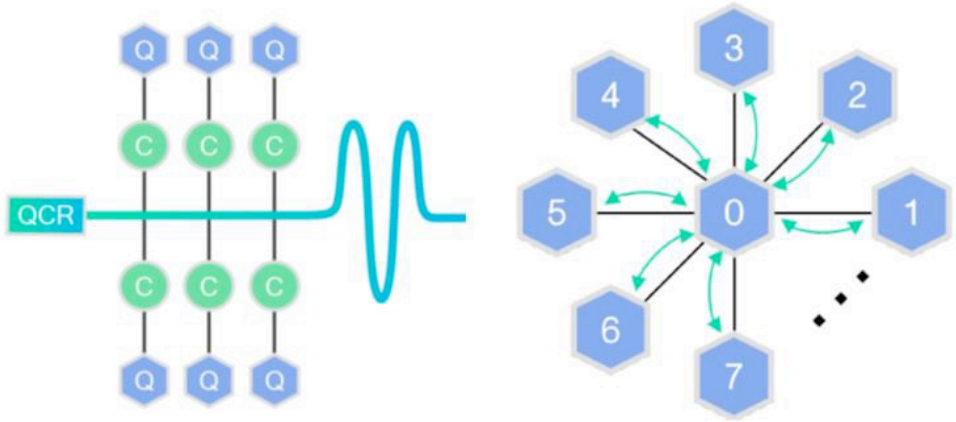
Diverse

- Getting several QCs to the fingertips of researchers and developers is crucial for catalysing software development.
- Different problems will fit different architectures and software stack infrastructure better

Accessible

- By being available through several platforms distributed throughout Europe, LUMI-Q provides a familiar interface to a uniquely large user base

LUMI-Q quantum computer



- Star-shape qubit topology, one-to-all qubit connectivity,
- Strong reduction of SWAP operations count
- High complexity quantum algorithms possible
- Massive advancement compared to anything presently available
- Increased performance
- **Bids evaluation ongoing**
- **Installation: Q4 2024**

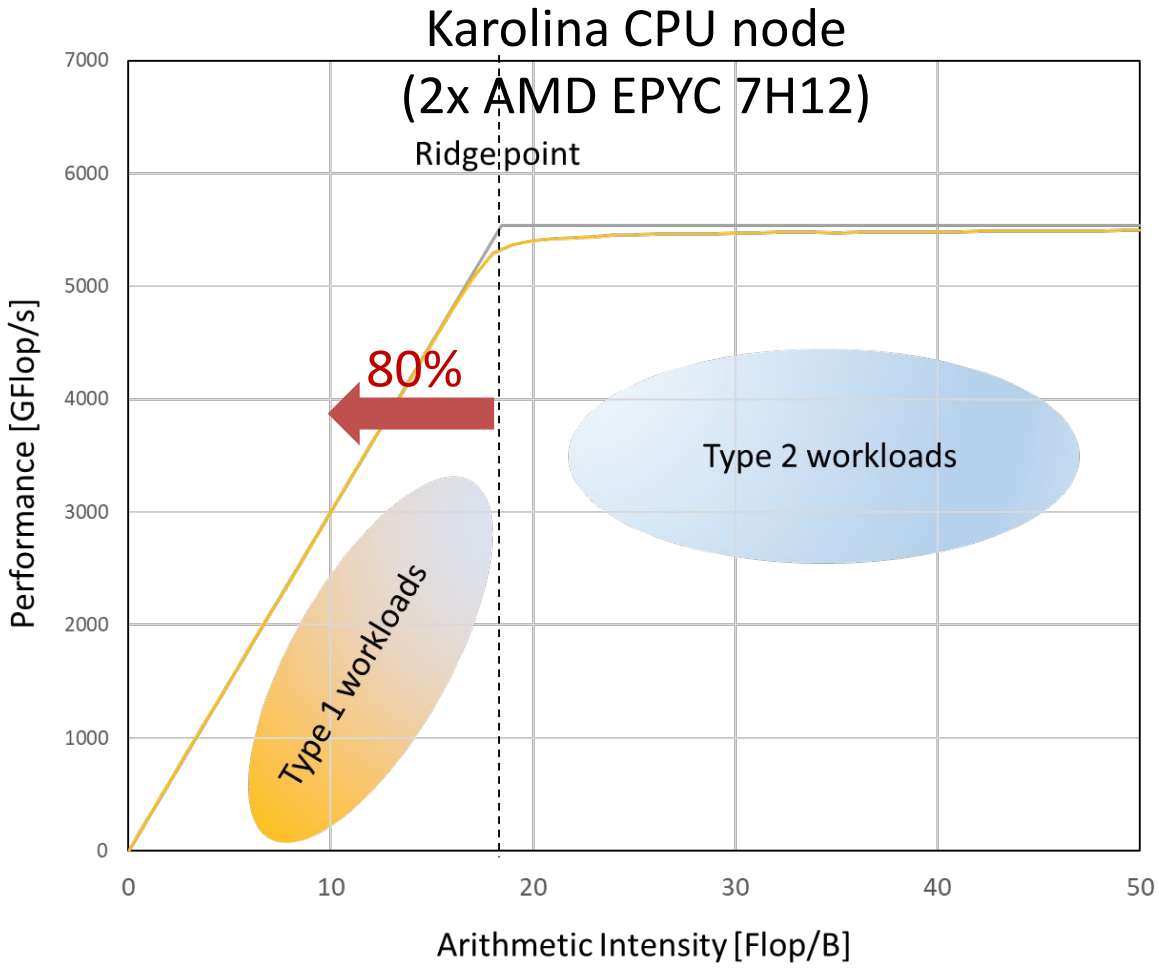
Metric	Value
Qubits	>= 20
Qubit connectivity	one-to-all, star-shape
T1 relaxation time	typically ~40 µs minimum for all qubits:15 µs
T2 dephasing time	typically ~20 µs minimum for all qubits: 15 µs
Readout fidelity	> 0.95

Small cluster III

Installation Q1 2025

Compute	180 nodes
Home	25 TB
Scratch	300 TB
Login	2 nodes
Mgmt	2 nodes
Infiniband 200GBps	194 ports
LAN	196 ports

- Memory throughput 900-1200GB/s
- Intel Granite Rapids CPU
- AMD Turin CPU
- Nvidia Grace Superchip CPU



Big cluster III

Installation Q4 2026

Non-accelerated	300 nodes
Accelerated	74 nodes
Home	25 TB
Scratch	2000 TB
Login	4 nodes
Mgmt	4 nodes
Infiniband	530 ports
LAN	460 ports

Architectural choices

- x86_64 architecture
- ARM architecture
- NVIDIA Architecture (GRACE+HOPPER)

Considered configuration

- 4x Grace Hopper NVIDIA per node
- includes ARM Grace CPU
- GPU memory perf 4 TB/s
- CPU memory perf 0.9 TB/s
- Estimated 30 FP64 PFlop/s peak.

IT4I data room infrastructure

- **Existing cooling capacity and performance metrics**
  - 3 cold-water circuits with operating temperature 6 – 15°C
    - Theoretical total cooling performance – 1.4 MW
    - Cold-water circuits help to cool down warm water circuits at higher outdoor temperatures
  - 2 warm-water circuits with operating temperature 30 - 35°C
    - Theoretical total cooling performance – 1.2 MW
- **Existing parameters of the power supply**
  - Reserved power – 2MW
  - All devices can be backed up from two independent power branches

CHALLENGES WE FACE

- **The need to place new systems in IT4I data room** – increased heat load
- **Spatial constraints**
  - Larger dimensions and weight of more powerful cooling devices (chillers, dry coolers)
  - Roof platform load capacity
  - Acoustic parameters of new chillers
- **Temperature changes**
  - Increasing number of days with high outdoor temperatures (> 30°C)

Modernization of data room - plan

- **Cold-water circuits**
  - New theoretical cooling capacity 1.8 MW
  - Replacement of outdoor chillers on the roof equipped with the free-cooling
  - One dry cooler in each circuit
- **Warm-water circuits**
  - New theoretical cooling capacity 2.1 MW
  - Installation of a new additional pair of coolant pipes with two new dry coolers
  - Replacement of existing dry coolers on the roof
  - Additional adiabatic cooling of dry coolers
- **Power supply**
  - Increase reserved power to 2.8 MW (includes replacement of measuring transformers)
  - Change in backup concept - some devices backed up from only one power branch
- **Expected outcome - Enhanced system reliability and efficiency**
  - Reduced risk of system failures
  - Reduced operating costs
  - Sufficient power supply and cooling capacity for future technologies

Training in 2023

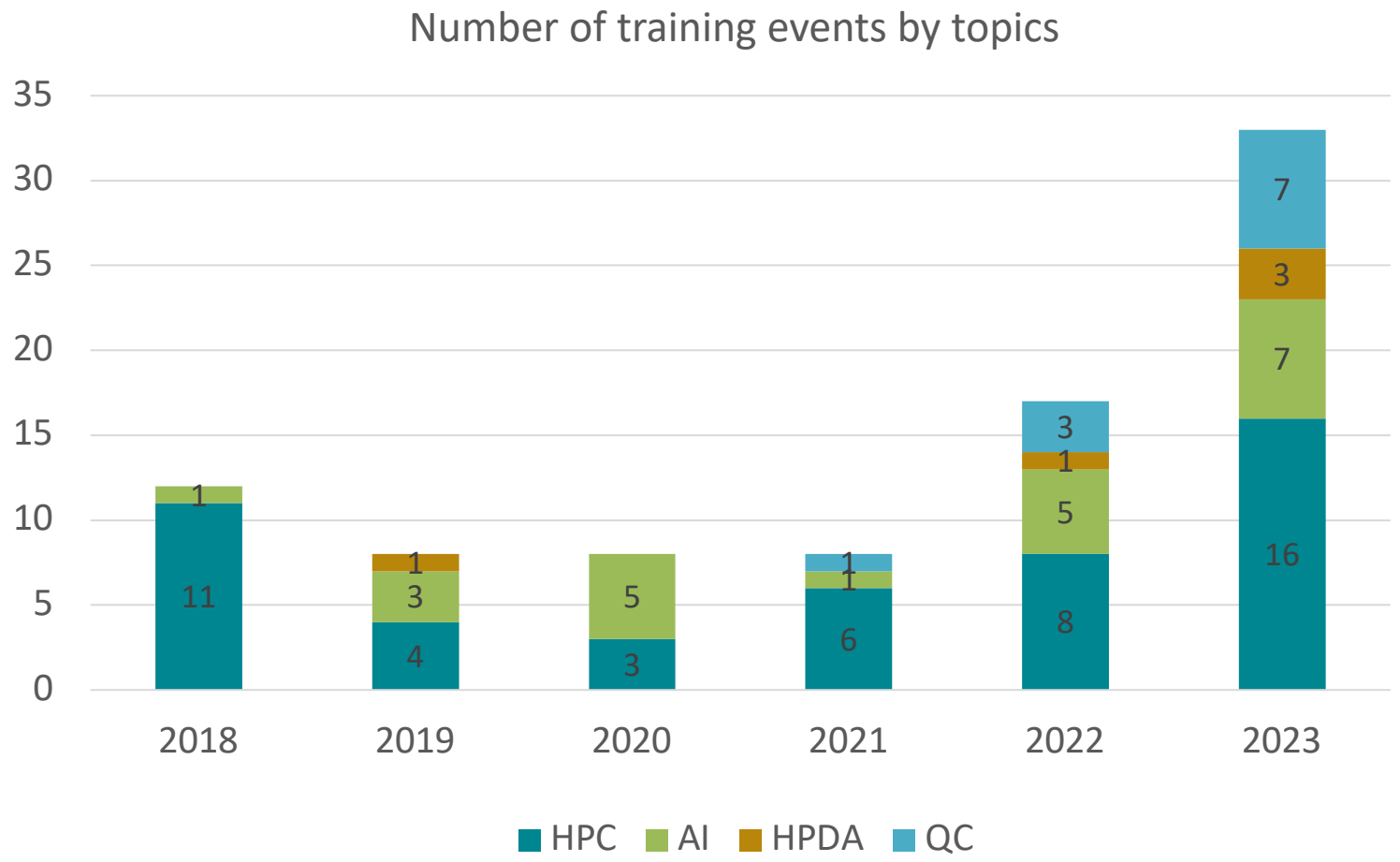
Total number HPC, HPDA, AI and QC events: 33

Total number of participants: 873

External participants 778 (89%)

Participants from industry and public administration 135 (15%)

International participants 443 (50%)



Training plan 2024

May	28.5.	Variational Quantum Algorithms
	29.-30.5.	Parallel R
June	3.-5.6.	Introduction to HPC
	13.6.	Quantum Computing Seminar: Solving Optimisation Problems Using the NISQ Era Quantum Computers
	19.-21.6.	Quantum Espresso
August	17.-25.8.	EUMaster4HPC Summer School: HPC in Data Science
September	12.9.	Quantum Computing Seminar
	4.-6.9.	POP3 profiling and optimisation tools
October	TBA	Python in HPC
	TBA	Programming GPUs with CUDA and C/C++
November	TBA	LUMI: AI workshop
	14.11.	Quantum Computing Seminar

https://events.it4i.cz

e-INFRA CZ

info@e-infra.cz

info@it4i.cz