

From laptop to LUMI - CSC services for researchers

Jussi Enkovaara, Sebastian von Alfthan



CSC – Finnish expertise in ICT for research, education and public administration

Outline

- What is CSC
- CSC computing services
- CSC data services
- Other services
- How to get access?

CSC - IT center for science

- CSC is non-profit company producing IT services for research and higher education
- Owned by ministry of education and culture (OKM) and higher education institutions
- Most CSC services are free of charge for end users

FUNET

by CSC



by CSC



When I might need CSC?

- My calculation takes very long, but could be run in parallel
- My calculation needs a lot of memory
- I need lots of storage space
- Scientific application is available at CSC
 - expensive commercial application
 - open source applications without need to install myself
- I want to share data
- I want to publish data and make it available

How CSC supercomputers differ from university cluster?

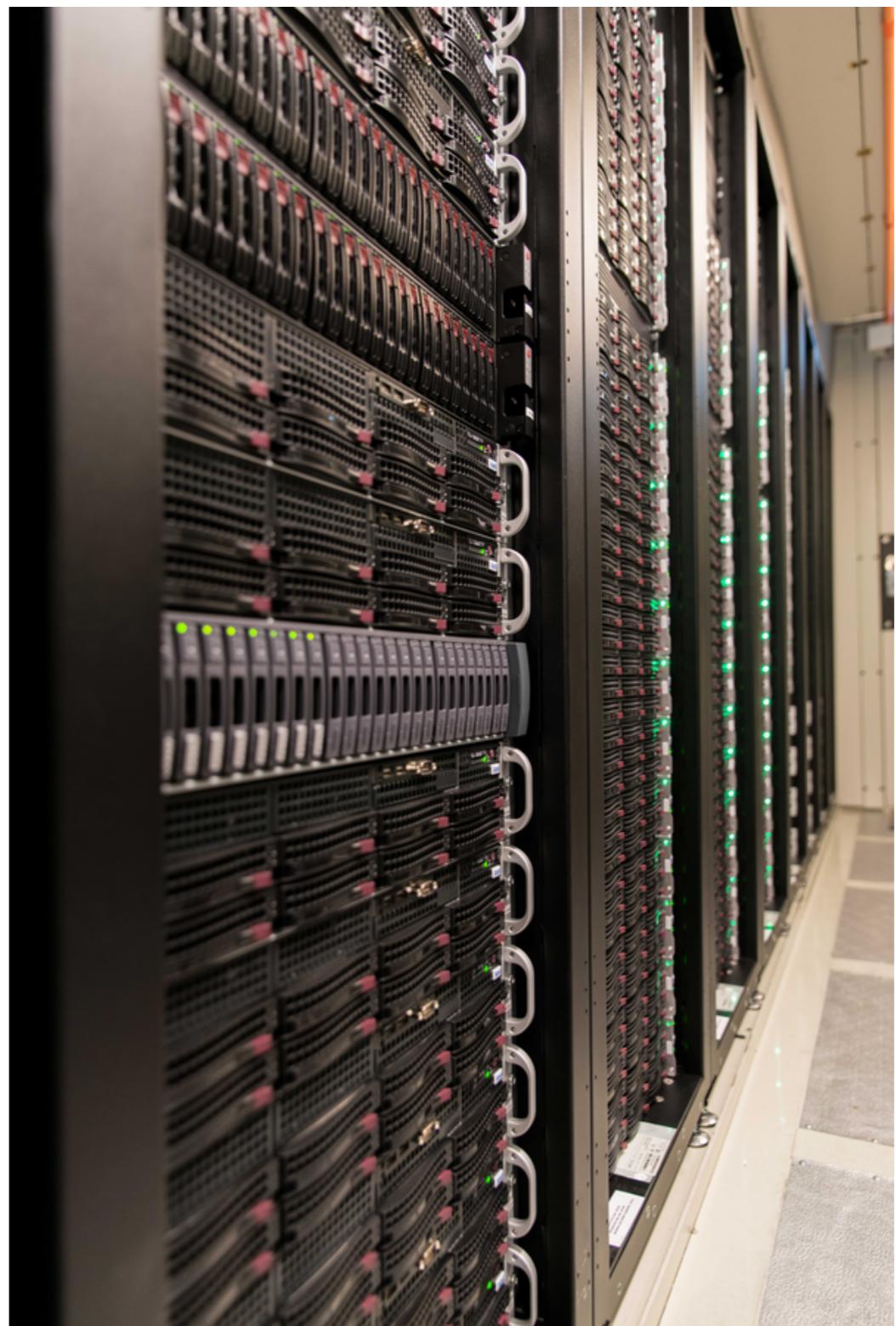
- Main difference is in scale
- Parallel calculations up to 25 000 cores (Mahti) or hundreds of GPUs (LUMI)
- Larger storage space available (1 TiB default in Puhti)
- Similar command line based terminal access
 - As a new service CSC provides also access via web browser
- Similar **module** system and **batch queue** system
- CSC resources consume “billing units”

CSC computing services

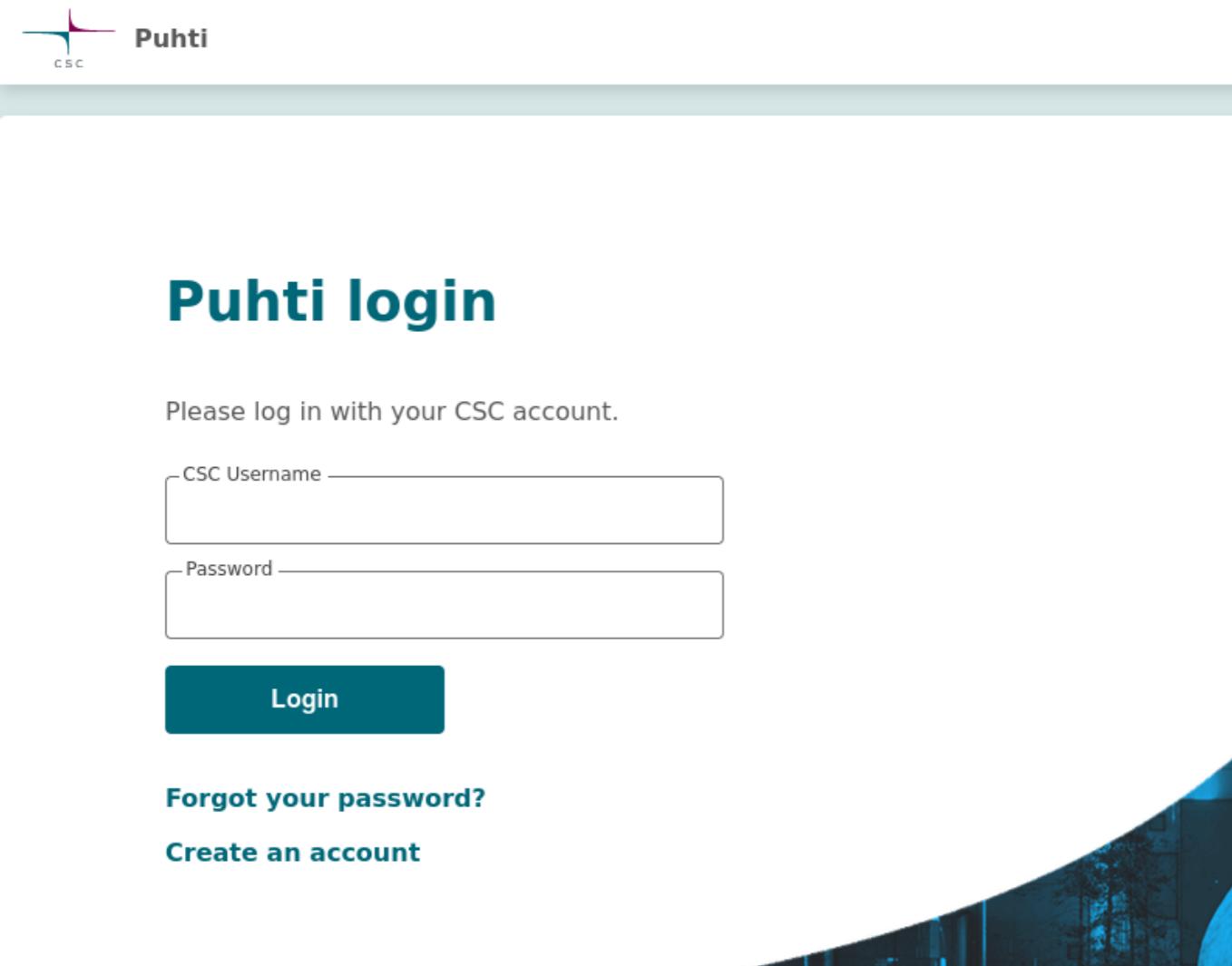
- Puhti
 - use cases from interactive single core data processing to medium scale parallel simulations
 - some large memory nodes
 - new web interface for accessing Puhti
 - Intel CPUs, 40 cores per node, 682 nodes
 - Nvidia V100 GPUs, 4 GPUs per node, 80 nodes
- Mahti
 - geared towards medium and large scale parallel simulations
 - AMD CPUs, 128 cores per node, 1404 nodes
 - Nvidia A100 GPUs, 4 GPUs per node, 24 nodes

Should I use Puhti or Mahti?

- Puhti
 - more software available
 - up to few hundreds of CPU cores
 - CPU nodes with fast local disk
 - Jupyter notebooks, RStudio, ...
- Mahti
 - minimum of 128 CPU cores
 - Often shorter queues (in CPU partitions)
 - Very large scale simulations
 - Scalability tests needed for over 20 nodes (2560 cores)



Web interface: www.puhti.csc.fi



The Puhti login page features the CSC logo and the word "Puhti". It includes fields for "CSC Username" and "Password", a "Login" button, and links for "Forgot your password?" and "Create an account". A large blue graphic of a mountain or wave is visible at the bottom.

Puhti login

Please log in with your CSC account.

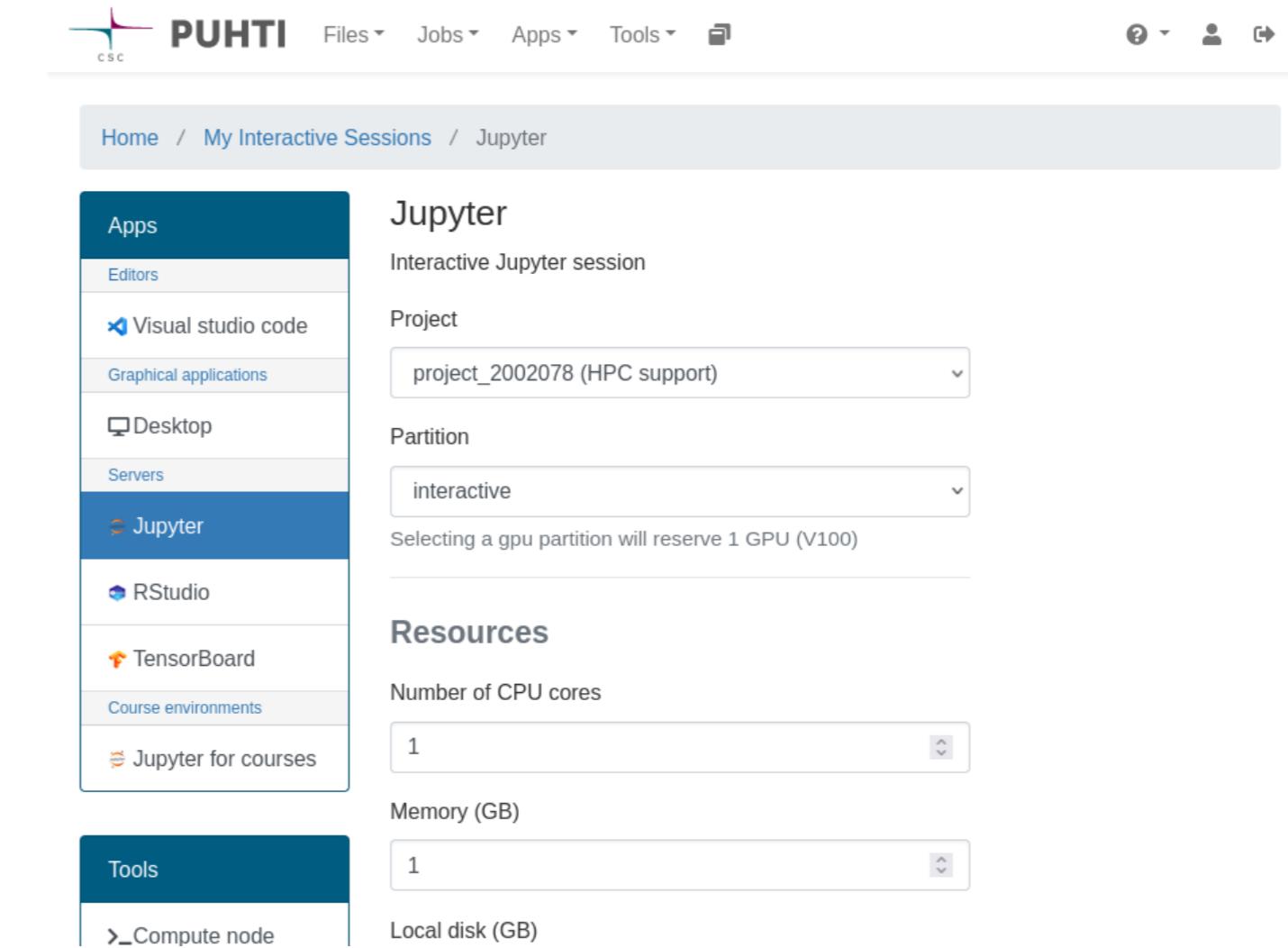
CSC Username _____

Password _____

Login

[Forgot your password?](#)

[Create an account](#)



The PUHTI web interface shows the "Jupyter" section. The "Apps" sidebar lists "Editors", "Visual studio code", "Graphical applications", "Desktop", "Servers", "Jupyter" (which is selected and highlighted in blue), "RStudio", "TensorBoard", "Course environments", and "Jupyter for courses". The main area shows "Interactive Jupyter session" under "Project" (set to "project_2002078 (HPC support)"), "interactive" under "Partition" (with a note about GPU reservation), and resource settings for "Number of CPU cores" (1), "Memory (GB)" (1), and "Local disk (GB)".

PUHTI

Files ▾ Jobs ▾ Apps ▾ Tools ▾

Home / My Interactive Sessions / Jupyter

Jupyter

Interactive Jupyter session

Project

project_2002078 (HPC support)

Partition

interactive

Selecting a gpu partition will reserve 1 GPU (V100)

Resources

Number of CPU cores

1

Memory (GB)

1

Local disk (GB)

Web interface: www.puhti.csc.fi

- Graphical access to files
- Visual studio code for code development
- Easy to launch jupyter notebooks, Rstudio, Tensorboard
- Graphical desktop available, now also accelerated one for heavy visualization tasks
- Enable users to create custom course environments
- And more...

LUMI

- LUMI is pan-European supercomputer hosted by CSC
- For Finnish users access is via CSC
- Over 10 000 AMD GPUs (Lumi-G)
 - available in autumn 2022
- “small” (about the size of Mahti) CPU partition
- auxiliary partition for data-analytics with lots of memory (32 TB per node), and 8x8 Nvidia A40 GPUs

Programming for CSC supercomputers

- C++/C, Fortran, Python, R, Julia, ...
- Parallel programming with MPI and OpenMP
- Vectorization important for single CPU performance
- High performance libraries
 - BLAS, LAPACK, ScaLAPACK, FFTW
- GPU programming with OpenMP, OpenACC, CUDA, HIP
- Parallel performance analysis and debugging tools
- Machine learning frameworks
 - PyTorch, Tensorflow, ...

Using Supercomputers efficiently

- A supercomputer is not a laptop, and scaling workloads up can hit many bottlenecks: slurm, IO, compute, memory, scaling
- Using CSC env. efficiently:
<https://ssl.eventilla.com/csccompenvsselflearn>
- Workflow recommendations (NEW):
<https://docs.csc.fi/computing/running/throughput/>
- ML getting started:
<https://docs.csc.fi/support/tutorials/ml-starting/>

CSC cloud computing services

- Sometimes one needs more flexibility than available in supercomputers
 - different operating system and system libraries
 - usage without batch job system
- CSC provides variety of cloud computing services
 - cPouta: general computing cloud
 - ePouta: computing cloud for sensitive data
 - Rahti: container cloud e.g. for running web services

Puhti: some technical details

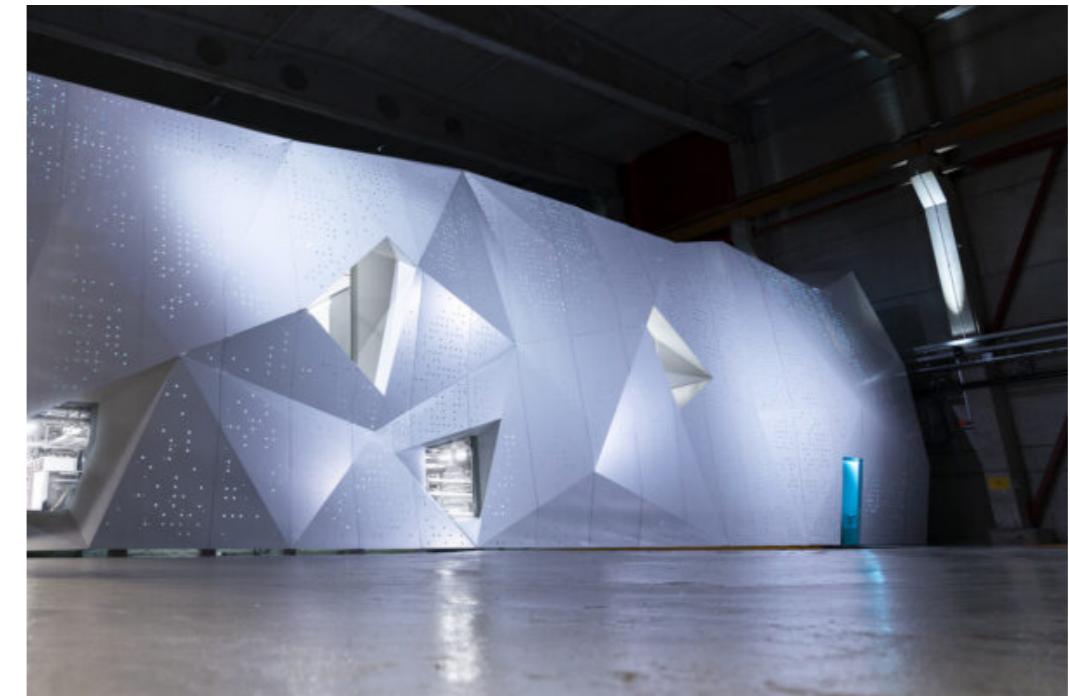
- 700 CPU nodes with range of memory and local storage options
 - Intel Xeon 2 x 20 cores @ 2.1 GHz (AVX512)
 - 192 GB - 1.5 TB memory per node
 - Some nodes with fast local NVME disks
 - Infiniband HDR 100 Gb/s
 - Pure MPI performs typically best
- 80 GPU nodes
 - 4 NVIDIA V100, 32 GB of memory
 - Infiniband HDR 200 Gb/s

Mahti: some technical details

- 1404 CPU nodes
 - AMD EPYC 2 x 64 cores @ 2.6 GHz (AVX2)
 - 256 GB of memory per node
 - Infiniband HDR 200 Gb/s
 - Hybrid MPI/OpenMP can often improve performance
 - Memory bound applications may perform better when not using all the cores
 - Non-uniform memory access (NUMA) with 8 NUMA domains per node
 - Correct binding of processes / threads to cores important for performance
- 24 GPU nodes
 - 4 NVIDIA A100, 32 GB of memory
 - Infiniband HDR 200 Gb/s

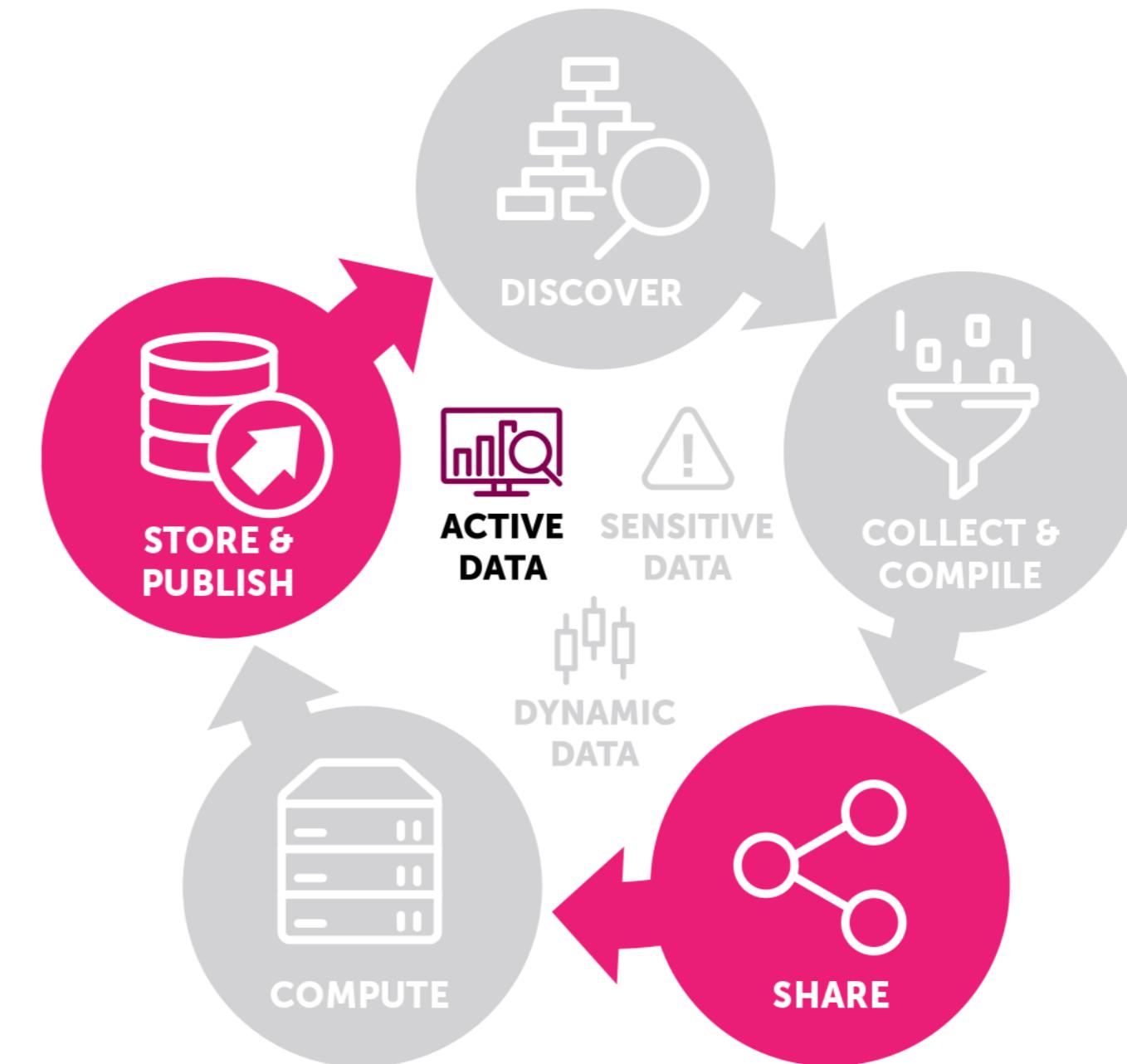
LUMI: some technical details

- 2560 GPU nodes
 - 4 AMD MI250X GPUs with 128 GB of memory per GPU
 - HPE Slingshot interconnect
 - four 200 Gb/s network cards, 800 Gb/s bandwidth in total
- Theoretical peak performance of 550 petaflops / s
- Storage
 - 8 PiB flash lustre
 - 80 PiB traditional lustre
 - 30 PiB object storage (Lumi-O)



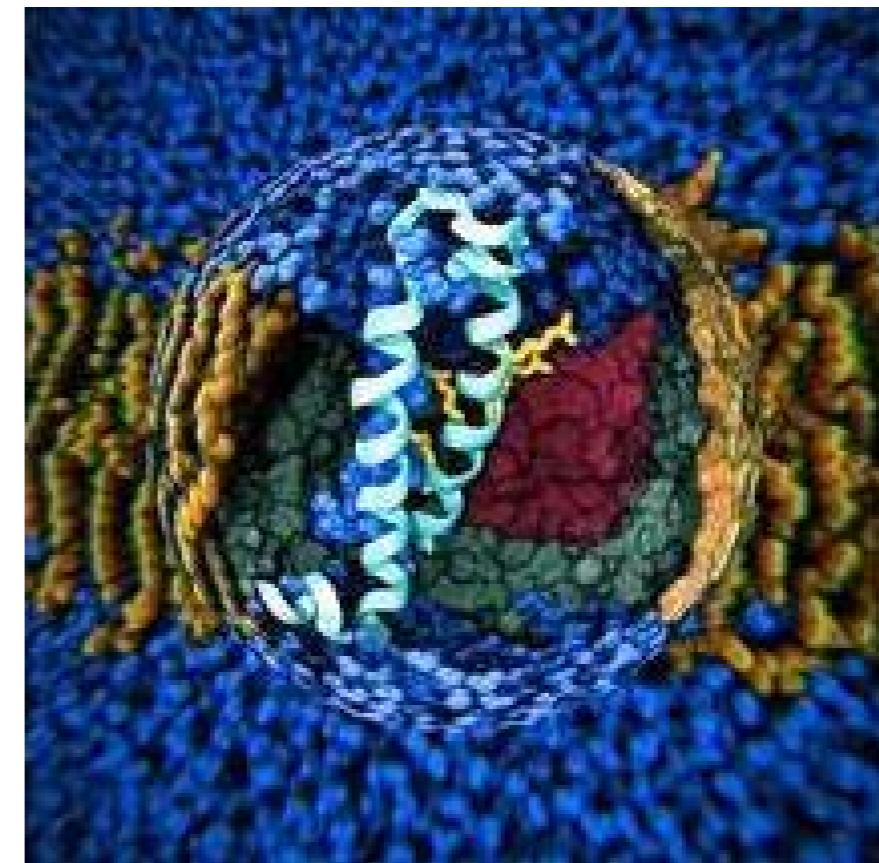
CSC data management and storage services

- Allas object storage
 - Data can be accessed anywhere from internet
 - Data can be shared with different levels of access control
- FAIR data services
(Findable, Accessible, Interoperable, Reusable)
 - Publishing datasets together with metadata
 - Searching public datasets



Other services

- Training
 - Large number of training courses in various aspects of scientific computing
- Expert services
 - Help in using scientific software
 - Help in optimizing application performance



Topical trainings

- Elements of supercomputing
<https://edukamu.fi/elements-of-supercomputing>
- CSC training portal:
<https://www.csc.fi/en/training>
- Self learning material:
<https://docs.csc.fi/support/training-material/>
- Lumi training portal:
<https://www.lumi-supercomputer.eu/events/>
- EuroHPC CC portal:
<https://www.eurocc-access.eu/services/training/>



The CSC Summer School in High-Performance Computing

⌚ 26.6.2022 - 5.7.2022 ⌂ Haltia Lake Lodge, Solvalla Nuksio

The CSC Summer School in High Performance Computing is back in Nuksio!



⌚ Event time
Starts: 26.6.2022
Ends: 5.7.2022

📍 Event location
Haltia Lake Lodge,
Solvalla Nuksio

Getting access to CSC supercomputers and services

- User needs to create a CSC account at my.csc.fi
 - With HAKA authentication only couple of mouse clicks
- Access to both Lumi and national resources available from same portal
- For computational and storage resources CSC project is needed
 - project manager needs to be an experienced researcher (post doc or higher)
 - project manager applies for billing units
 - project manager can add users to project

Questions ?

- Getting support: servicedesk@csc.fi
- Lumi support: <https://www.lumi-supercomputer.eu/user-support/>
- CSC user guide: <https://docs.csc.fi>
- FAIR data services: <https://fairdata.fi>
- Services for researchers: <https://research.csc.fi>
- CSC Training: <https://csc.fi/training>