

From laptop to LUMI - CSC services for researchers

Jussi Enkovaara, HPC support



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

Haka

by CSC

eduroam

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, more on request)
- 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
 - Intel CPUs, 40 cores per node
 - GPU partition with NVIDIA V100 GPUs
- Mahti
 - geared towards medium and large scale parallel simulations
 - AMD CPUs, 128 cores per node
 - small GPU partition with NVIDIA A100 GPUs
 - subset can be sliced to smaller GPUs for interactive workloads, code development etc.

LUMI

- LUMI is pan-European supercomputer hosted by CSC
 - user support is distributed to LUMI consortium countries
- For Finnish users access is via CSC
- Over 10 000 AMD GPUs
- “small” (about the size of Mahti) CPU partition
- auxiliary partition for data-analytics with lots of memory (32 TB per node)

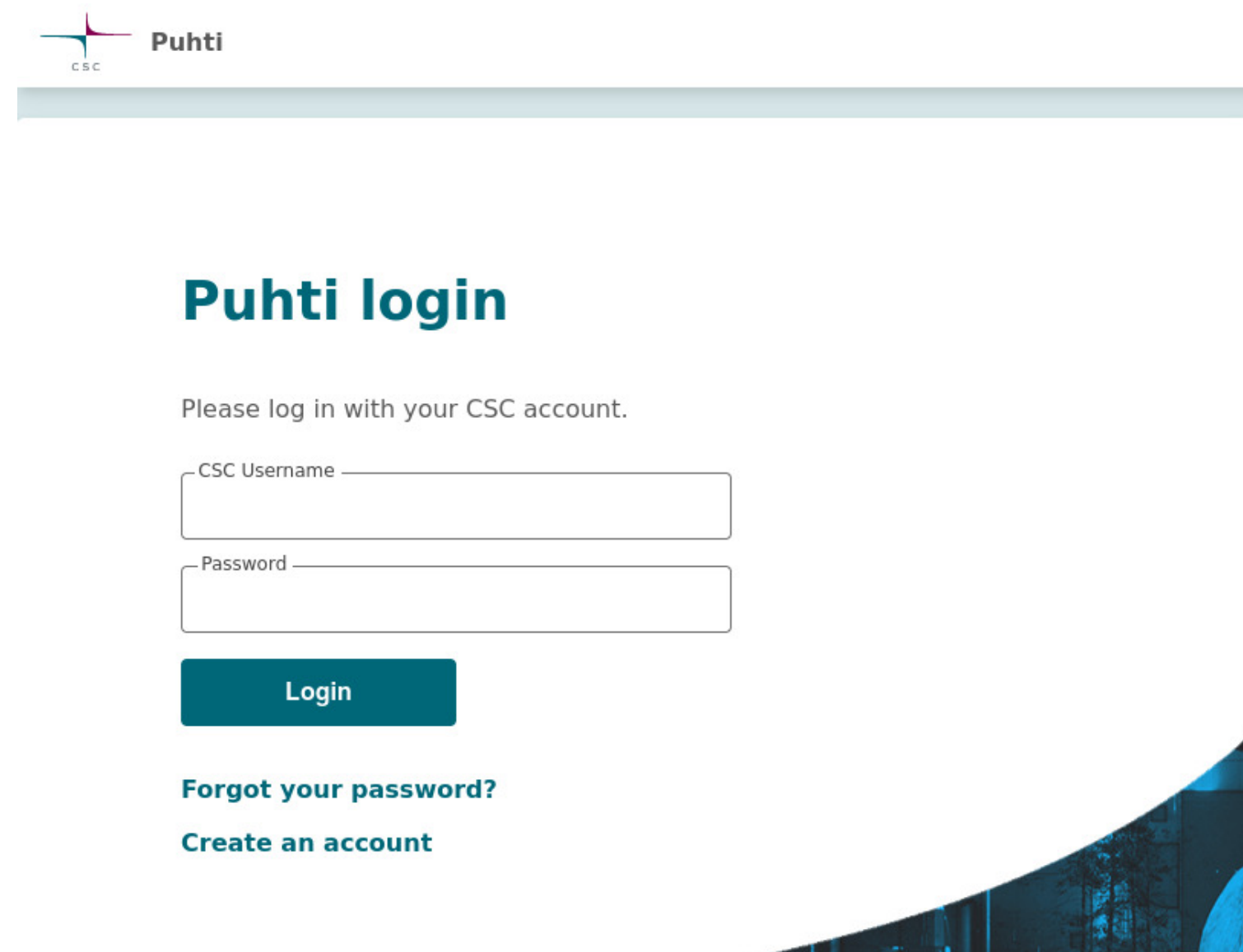



Should I use Puhti, Mahti or LUMI?

- 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)
- LUMI
 - applications benefitting from AMD GPUs
 - CPU applications with similar usage profile as in Mahti
 - software availability more limited

Web interfaces

- All CSC supercomputer can be used also via web interface
 - <https://puhti.csc.fi>, <https://mahti.csc.fi>, <https://lumi.csc.fi>



 **Puhti**

Puhti login

Please log in with your CSC account.

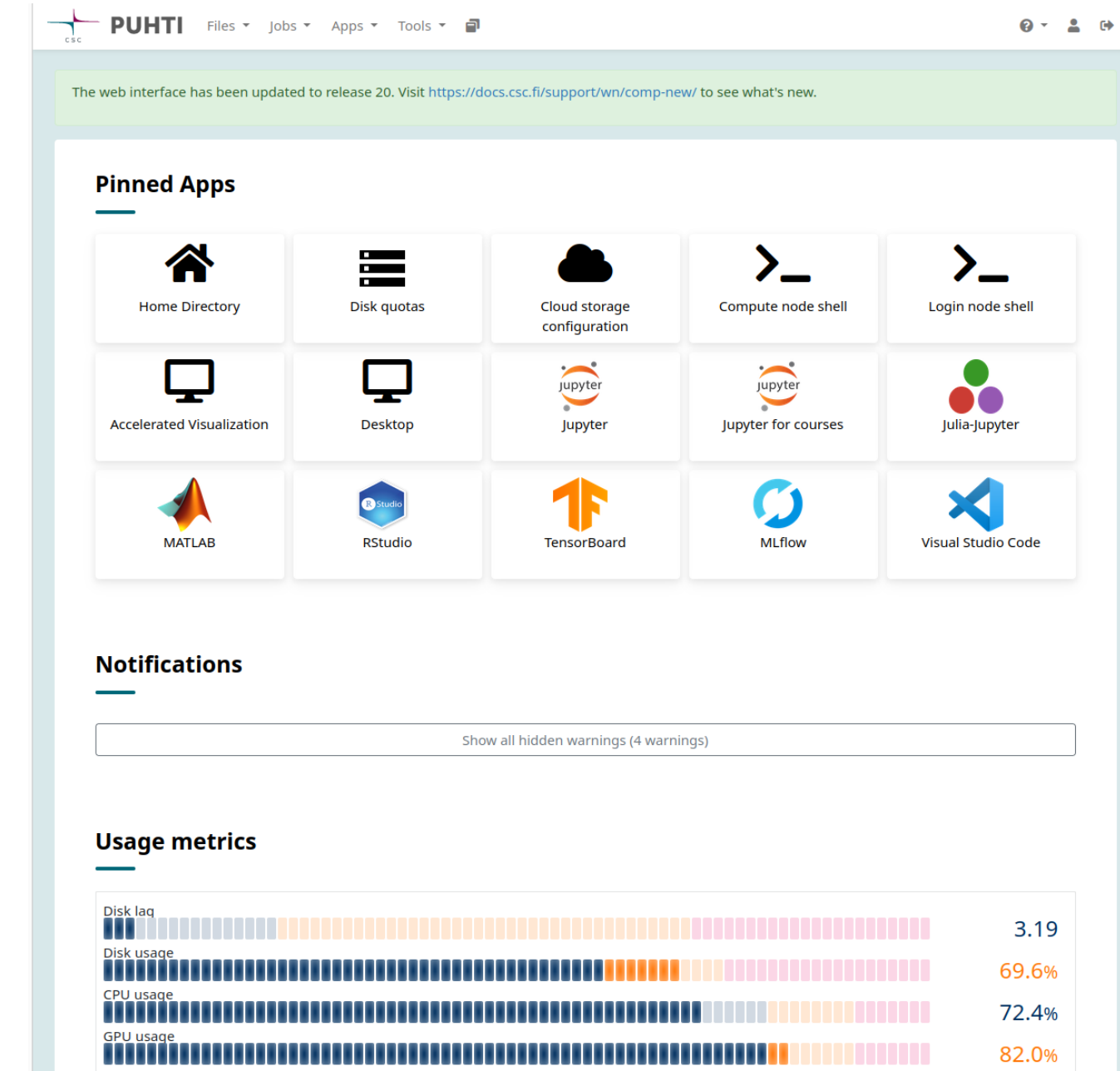
CSC Username


Password

Login

[Forgot your password?](#)
















[Create an account](#)



 **PUHTI** Files Jobs Apps Tools

The web interface has been updated to release 20. Visit <https://docs.csc.fi/support/wrn/comp-new/> to see what's new.

Pinned Apps

 Home Directory	 Disk quotas	 Cloud storage configuration	 Compute node shell	 Login node shell
 Accelerated Visualization	 Desktop	 Jupyter	 Jupyter for courses	 Julia-Jupyter
 MATLAB	 RStudio	 TensorBoard	 MLflow	 Visual Studio Code

Notifications

Show all hidden warnings (4 warnings)

Usage metrics

Disk I/O	<div><div></div></div>	3.19
Disk usage	<div><div></div></div>	69.6%
CPU usage	<div><div></div></div>	72.4%
GPU usage	<div><div></div></div>	82.0%

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

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, ...

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 benefit from using less than 128 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, 40 GB of memory
 - Infiniband HDR 2 x 200 Gb/s
 - Subset can be sliced into smaller GPUs with MIG

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 530 Pflops / s
 - Linpack performance 380 Pflops / s
 - HPCG performance 4.6 Pflops / s



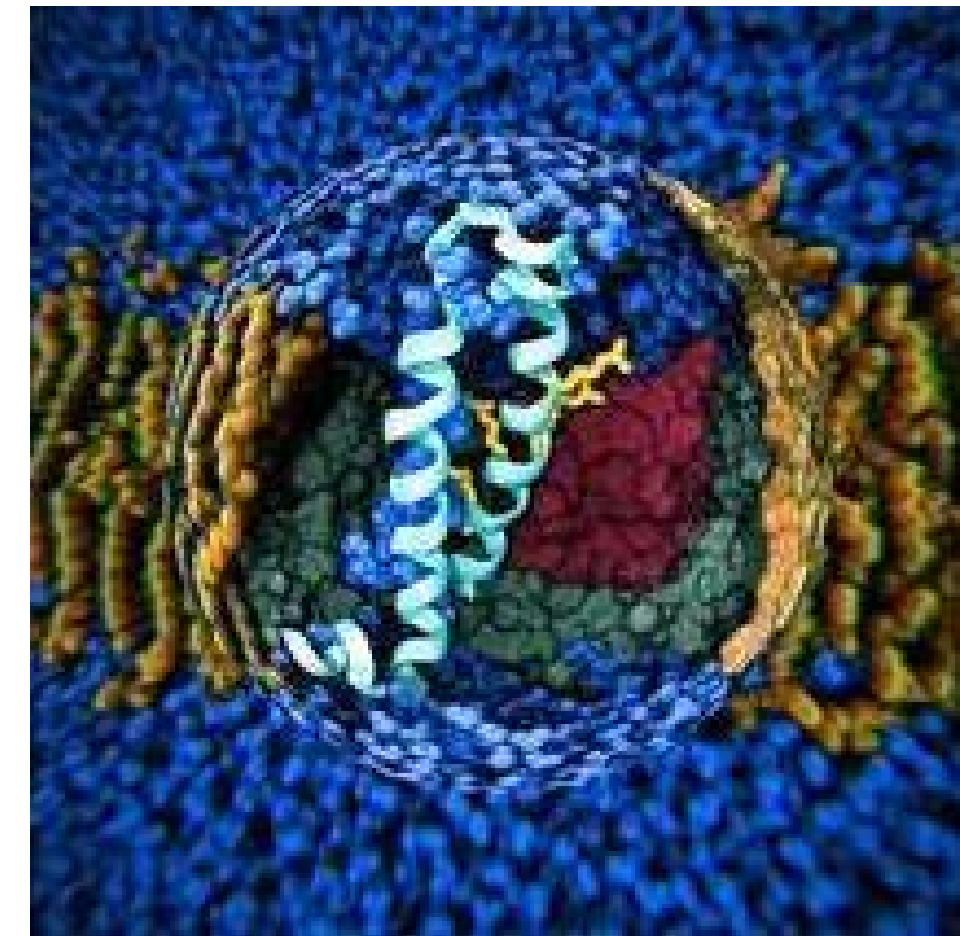
CSC data management and storage services

- 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
- Visualization
- Expert services
 - Help in using scientific software
 - Help in optimizing application performance



Topical trainings

- Online self-learning course: Elements of supercomputing
<https://edukamu.fi/elements-of-supercomputing>
- Online self-learning course: CSC computing environment
<https://csc.fi/en/training-calendar/csc-computing-environment-self-learning/>
- High Performance R, 26.-27.9
<https://csc.fi/en/training-calendar/high-performance-r/>

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
- 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
- Teachers may create a project for a course
 - small fixed amount of resources
 - participants need their own CSC accounts

Questions ?

- Getting support: servicedesk@csc.fi
- 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>