



Welcome to hybrid seminar

LUMI Supercomputing for geospatial companies

08.06.2023 | Katri Tegel, CSC – IT Center for Science Ltd.



Finnish Geospatial
Research and
Education Hub



The background of the slide is a light blue aerial topographic map. A semi-transparent blue overlay is applied to the map, featuring a network of thin blue lines and numerous small blue dots scattered across the terrain. The text is centered in white on this background.

We use Mentimeter for Questions and feedback



Non-profit state organization with special tasks



Turn over in 2022

64,4M€



LUMI

Owned together with EuroHCP JU and unique consortium of 10 countries – Belgium, Czech Republic, Denmark, Estonia, Finland, Iceland, Norway, Poland, Sweden and Switzerland.



Hewlett Packard Enterprise

Headquarters in Espoo, data center in Kajaani



Owned by the Finnish state (70%) and all Finnish higher education institutions (30%)



652 employees (06.2023)



EuroHPC
Joint Undertaking

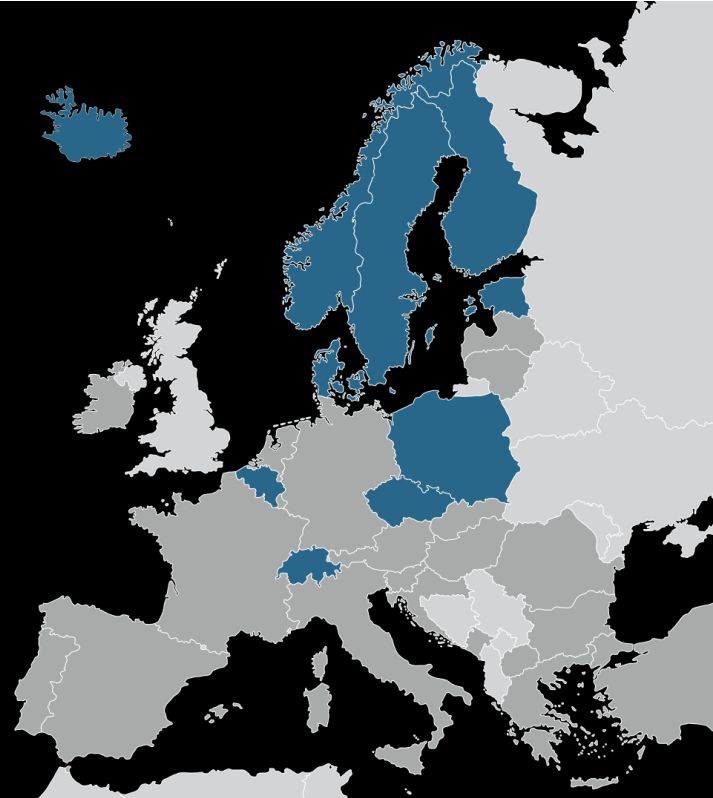


LUMI

EuroHPC Joint Undertaking: Europe's world-class scientific computing and data management research infrastructure (high-performance computing, HPC)

EuroHPC funding: EU and participating countries

- **The supercomputer hosted by the LUMI consortium (10 countries) in Kajaani**
- Developing competence in high-performance computing
 - **CSC operates EuroCC Finland, National competence center in Finland**



Geoportti and Location Innovation Hub EDIH



Geoportti – The Hub for for Finnish Geospatial Research and Education Resources

- A shared service for using geospatial data and geocomputing tools. Geoportti RI helps the researchers in Finland to use, to refine, to preserve and to share their geospatial resources.
- Coordinated by National Land Survey of Finland (NLS)

<https://www.geoportti.fi/>



Location Innovation Hub (LIH) European Digital Innovation Hub (EDIH)

- Large network that consists of companies, research organizations and universities, as well as the public sector.
- Goal: to expand the use of geospatial data in services, new business and technologies, and produce new benefits and innovations for society based on them.
- Coordinated by NLS

<https://locationinnovationhub.eu/en/home/>

Seminar agenda

- Welcome and practicalities
- Geocomputing, Q&A
- 9.45 –10.00 Break
- LUMI Supercomputing for Business, Q&A
- ICEYE use case, Q&A
- Wrap-up and where to go from here

The background of the slide is a topographic map of a region, likely in Finland, showing terrain contours and some urban areas. The map is overlaid with a grid of blue dots and thin blue lines, suggesting a geospatial data visualization or a network of points.

Geocomputing

Samantha Wittke and Markus Koskela
CSC – IT Center for Science Ltd.


Agenda

- What is a supercomputer?
- Why use a supercomputer?
- Geocomputing
- Deep Learning
- What is needed?
- Tools
- Training and examples

What is a supercomputer?

Clusters of powerful processors working together to solve a task.

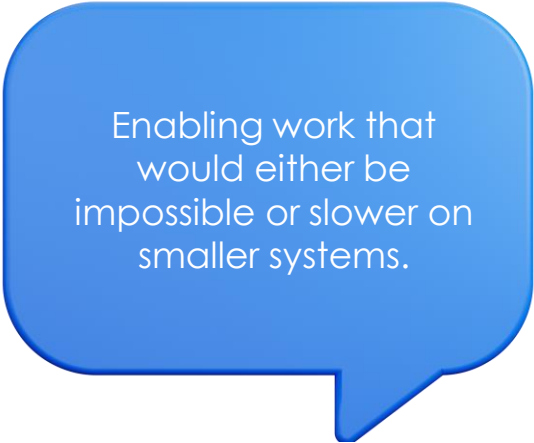
- Central and Graphical Processing Units (CPU/GPU)
- Large memory
- Fast connections
- Linux
- Shared by multiple users
- Standard interaction via command line interface



Also called
High-Performance
Computing (HPC)

Why use a supercomputer?

- "Outsource" heavy/specialized computations
- Resource needs
 - time, memory, storage, GPU
- Prebuilt environments
- Run many experiments at same time
- Parallelization
- Collaboration possibility
- **CSC specialist support**



Enabling work that would either be impossible or slower on smaller systems.

Geocomputing

- Mapsheets / Tiles
 - ▶ Naturally parallel
- Large point clouds, images, vector data
 - ▶ Resource needs
- Large scale calculations and simulations

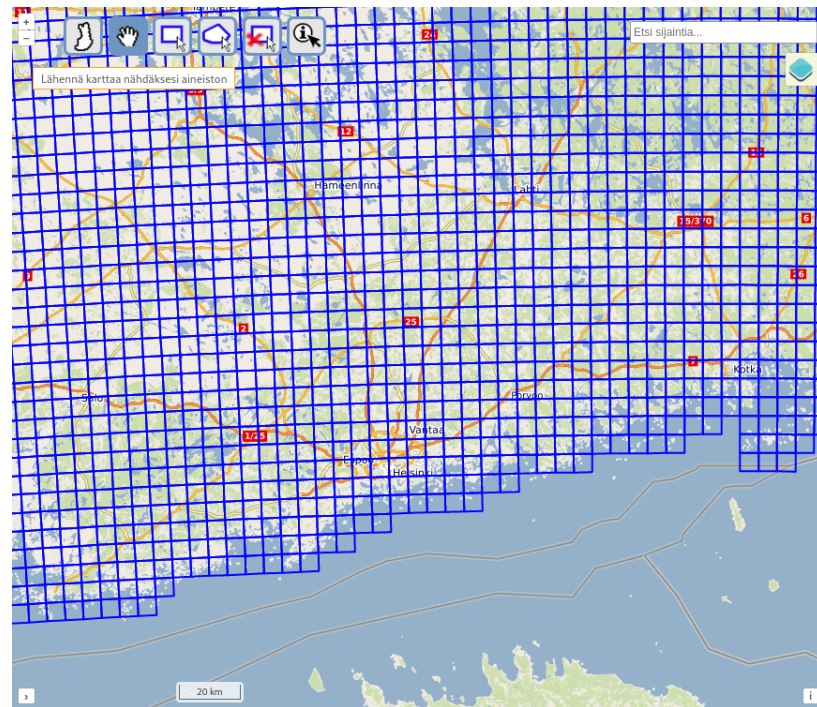
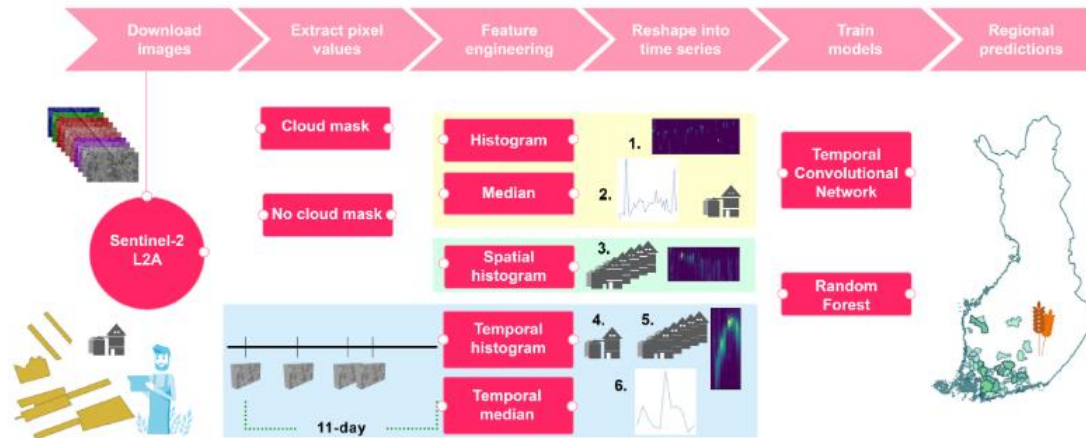


Image from <https://paituli.csc.fi/> Spatial data download service: 2m DEM tiles

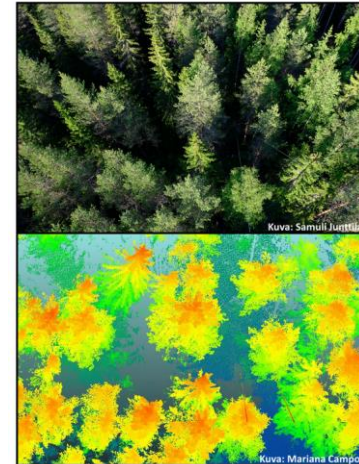
Examples from research

Scalable Crop Yield Prediction with Sentinel-2 Time Series and Temporal Convolutional Network

(Maria Yli-Heikkilä, LUKE)

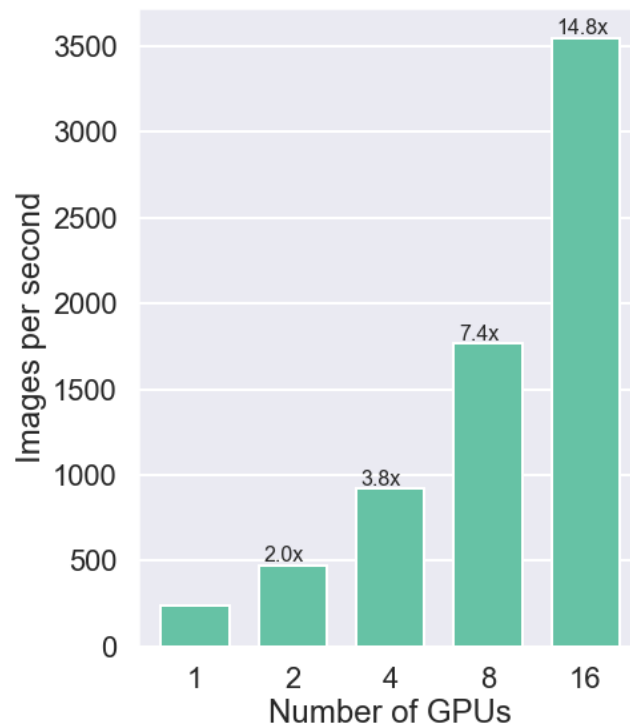


Multi platform point cloud processing and analysis (FGI, UEF, LUKE)



Supercomputers for Deep Learning

- All our supercomputers have **GPU acceleration**
- Resources for parallel training and inference
- Multi-GPU and multi-node jobs support



Some Deep Learning Research Projects

- **Open NLP models for Finnish:** F3AI – Foundation For Finnish Artificial Intelligence ([GPT-3 for Finnish](#))
- **Open LLM:** AI2 OLMo (Open Language Model)
- **Cancer diagnosis and grading:** ComPatAI – Artificial-intelligence driven computational pathology
- **Prostate cancer detection:** HistoEncoder – A foundation model for all digital histological samples
- **Vision-language models:** MMATT - Large-scale multi-modal data analysis with attention models
- **Climate modelling:** Active machine learning methods for atmospheric science application

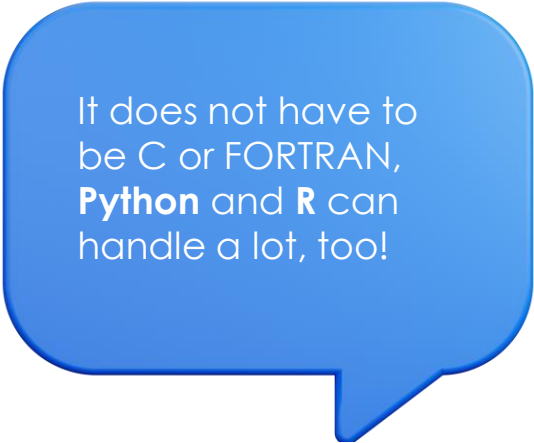
Suitable tools (Linux only)

- CloudCompare
- FORCE
- GDAL/OGR*
- GRASS GIS*
- LasTools
- MatLab
- OpenDroneMap
- Orfeo Toolbox
- PCL
- PDAL*
- Python geospatial packages
- QGIS*
- R geospatial packages
- SagaGIS*
- SNAP, Sen2cor, sen2mosaic
- WhiteboxTools
- Zonation
- Deep learning*: pytorch, tensorflow, Jax

*Already available in LUMI

What is needed?

- **Scripting**
- Basic Linux commands
- Using remote service
- Parallelisation (speed-up)
- Resource management



It does not have to be C or FORTRAN, **Python** and **R** can handle a lot, too!

What is needed?


- Scripting
- **Basic Linux commands**
- Using remote service
- Parallelisation (speed-up)
- Resource management



cd, pwd, cp, ls, mv, ...

What is needed?


- Scripting
- Basic Linux commands
- **Using remote service**
- Parallelisation (speed-up)
- Resource management



Ssh connections
via terminal or tools
like Putty.

What is needed?

- Scripting
- Basic Linux commands
- Using remote service
- **Parallelisation (speed-up)**
- Resource management




Make efficient use of
the available resources
for speed-up.

What is needed?

- Scripting
- Basic Linux commands
- Using remote service
- Parallelisation (speed-up)
- **Resource management**

Our courses and example materials are designed to help you with these :-)



Resource reservation
and analysis.

Training and Examples

- 13. June 2023 STAC workshop - How to find and use spatiotemporal data easily?
 - https://ssl.eventilla.com/stac_2023
- 12.–13. October 2023 Geocomputing course
 - https://ssl.eventilla.com/geocomputing_2023
- Geocomputing examples: <https://github.com/csc-training/geocomputing>
- Previous training materials: <https://research.csc.fi/gis-learning-materials>

Summary

- Supercomputers are large shared computation systems that we can connect to remotely
- They enable data processing and analysis that would otherwise be impossible or very time consuming
- Supercomputers excel with tile-wise geospatial processing and Deep Learning tasks

It takes a bit of practice,
but we are here to help:

firstname.lastname@csc.fi



Katri Tegel



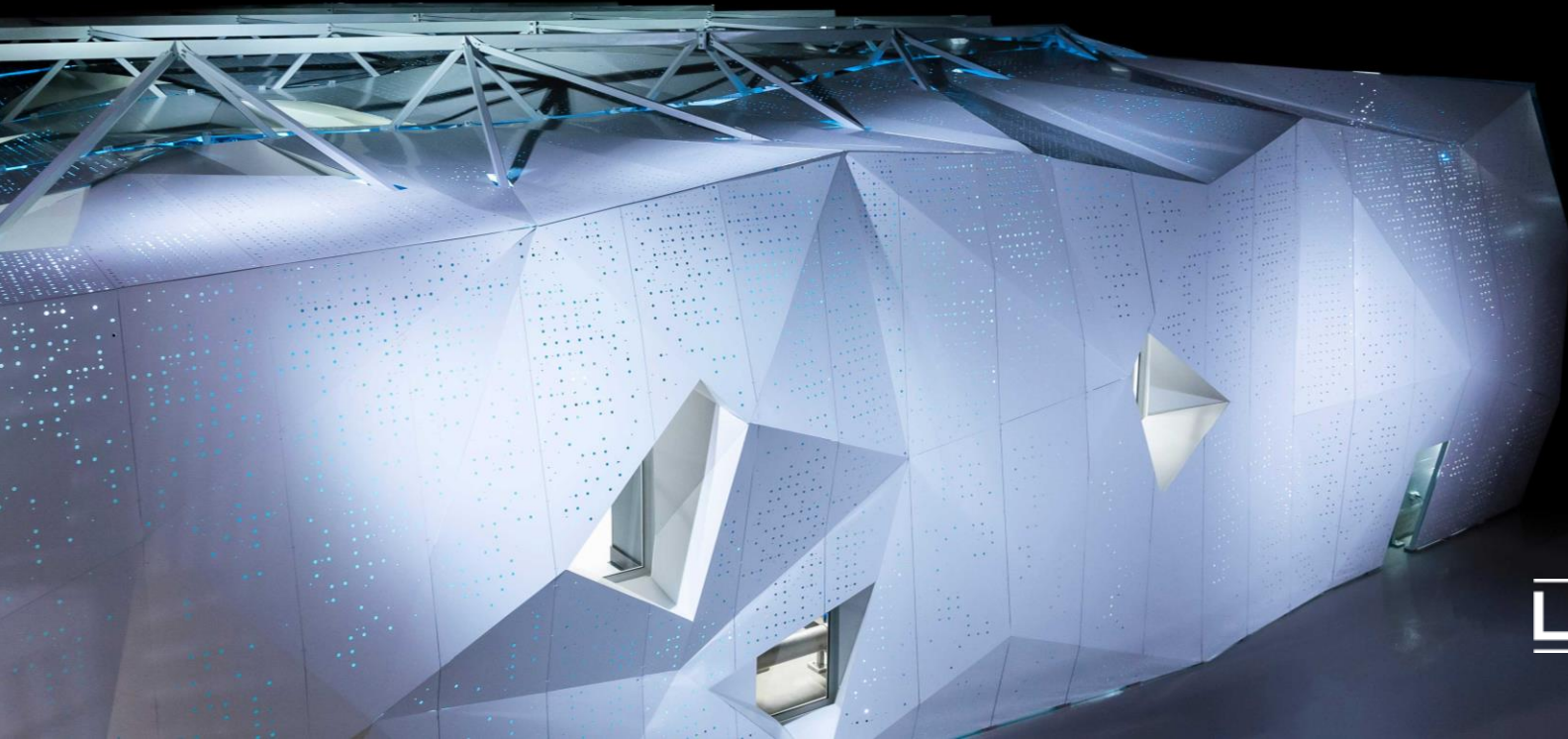
Samantha Wittke



Markus Koskela

Supercomputing for Business

Juhani Huttunen, CSC – IT Center for Science Ltd.



LUMI

Agenda

- LUMI (/HPC) overview
- LUMI Supercomputing for business – benefits and use cases
- How to start using LUMI in a company?
 - User paths, Test before use (“Try&Buy”), Service pricing
- Working with LUMI
 - Principles, Limitations, Steps & Tools & Technologies
- How to get help / Contact info

LUMI is an HPE Cray EX Supercomputer

LUMI

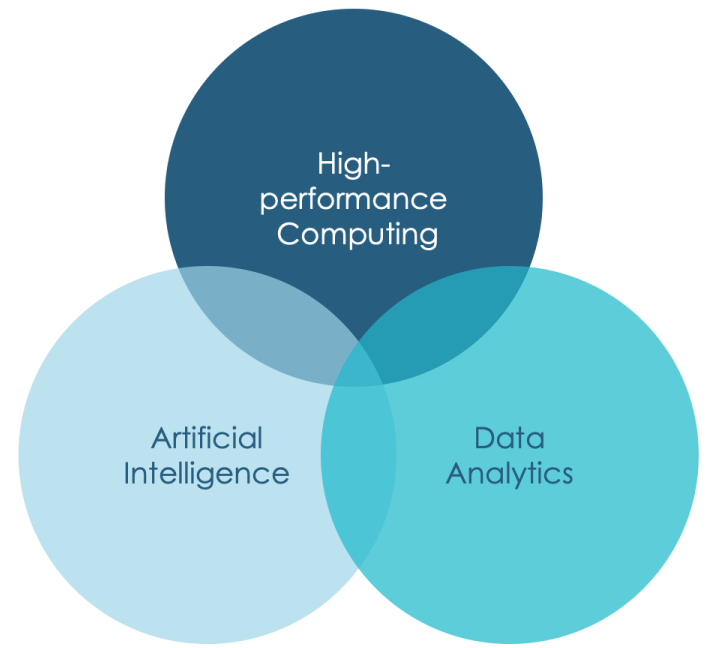



**Hewlett Packard
Enterprise**

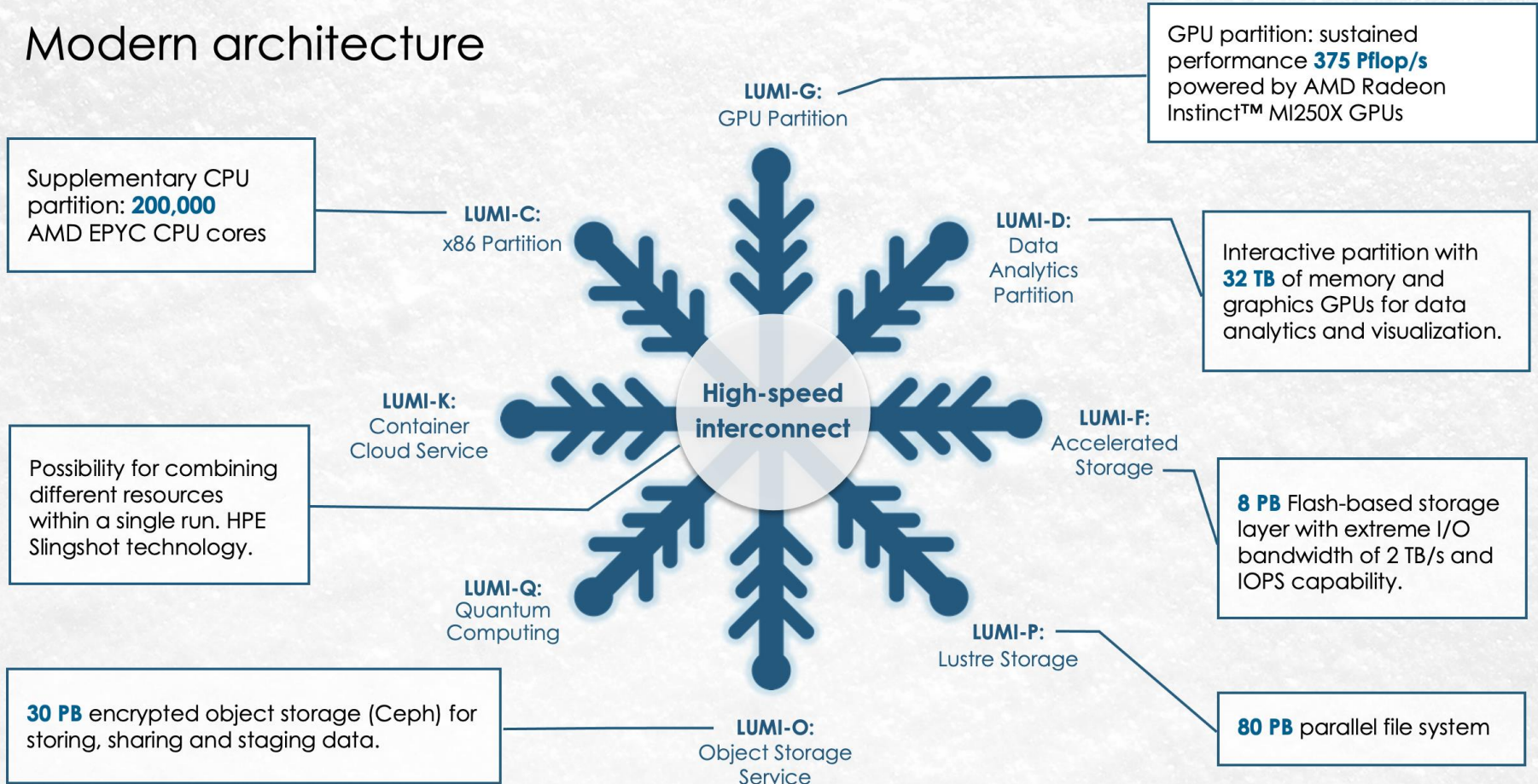
Its powerful processors can perform millions of calculations simultaneously, so it can operate millions of times faster than conventional computers and help us solve highly complex problems.



The convergence of High-performance Computing, Artificial Intelligence and Data Analytics will be key for solving the great scientific and societal challenges.



Modern architecture





Up to 20%

of LUMI's capacity
is reserved for
European industry
and SMEs

LUMI

For companies LUMI offer

- A world-class **supercomputing capacity with cost-efficient pricing**
- **Superfast product development** and **new business opportunities** in areas such as Data analytics and AI
- **Top technical expert and training support**
- A way to initiate or strengthen **cooperation with universities and research institutes**
- **Data security** based on ISO/IEC 27001 standard

Typical industrial applications



HEAVY SIMULATION OF COMPLEX SYSTEMS

Material sciences

- Development of new materials
- Development of new medicine

Climate & weather forecasting, numerical weather prediction

Understanding different physical phenomena, e.g.

- Fluid dynamics
- Structural mechanics
- Electromagnetics
- Heat transfer
- Acoustics
- Fire dynamics simulation

AI MODELS WITH MASSIVE DATA SETS

- Speech to text
- Speech recognition
- Image Recognition, e.g. for defect detection



Supercomputing use cases



Ramboll is a global engineering, architecture and consultancy company with a leading edge in creating sustainable cities and societies. The company employs 17,500 experts worldwide, and 2,500 people across Finland.

CHALLENGE

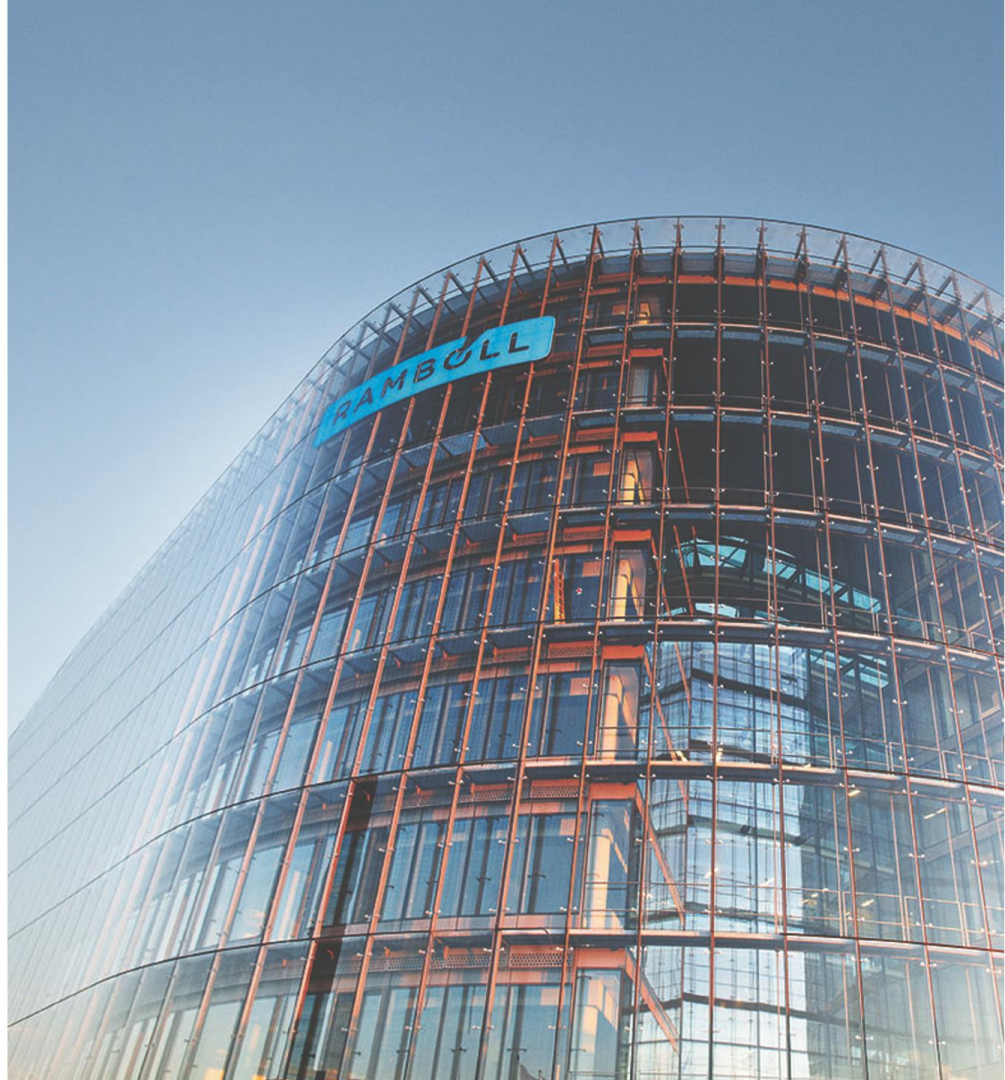
Ramboll Finland provides fire technical simulations of buildings, e.g., to authorities. Results are needed faster: in hours/a day rather than the current several days/a week.

SOLUTION

CSC has provided EuroHPC/LUMI CPU resources, HPC software expert support to install, adapt and test the simulation software and compiler tools in the LUMI environment.

IMPACT

The typical fire technical simulations were finalised in hours compared to days earlier, shortening the time for fire safety approvals from the authorities.



GROKE®

Groke Technologies Oy is a Finnish company that focuses on developing intelligent methods for autonomous navigation to improve maritime safety.

CHALLENGE

How do you create an awareness system that monitors the environment and detects and recognizes objects around a vessel, for example from the side of another vessel to a far-away sea buoy? How can machine learning models be programmed for object detection with an optimal trade-off between accuracy and performance.

SOLUTION AND IMPACT

By building scalable data processing pipelines for training, evaluating and optimizing machine learning algorithms, and running large-scale training on a GPU-equipped HPC system, better training speed and better target detection were achieved leading to lower unit costs.



How to start using LUMI in a COMPANY?

Private–Public engagement	Pay per use model	Business Finland funding
<ul style="list-style-type: none"> • Project in cooperation with Finnish university or research organization (academic partner) • Project lead (PI) assigned from academic partner • Free of charge if results are published 	<ul style="list-style-type: none"> • National LUMI capacity Company pays market price to CSC • EuroHPC JU capacity Company pays market price to JU. Funding possible through PRACE SHAPE and EuroHPC JU programs for SMEs 	<ul style="list-style-type: none"> • Start-up and SMEs can request HPC grant at a value of 20,000-80,000 €. Can be added to an already running project. • Large and mid cap companies can include computing capacity into their R&D project budget, 40 % of costs covered • Capacity is valued at market price

Test Before Use (“Try&Buy”)



Do you want to develop your company's business faster and more efficiently? Why don't you try high-performance computing, data analytics or artificial intelligence to support research and development work? The LUMI supercomputer can now be used by your company.

FREE-OF-CHARGE "TEST BEFORE USE"-PROJECT

In the project companies will receive

- CPU-, GPU, and data storage resources
- our expertise support

If interested to take LUMI in use, please contact our experts!



**Funded by
the European Union**

Green
Data Centre
of the
Year 2023



Price list – LUMI Computing Services



Service	€, VAT 0%
LUMI computing project base package	1000 €
LUMI-C – computing nodes with CPU-processors (AMD Milan) • 1 CPU-node hour equals to 128 CPU-core hours	0,57 € / CPU-node-hour ~0,445 cents / CPU-core-hour
LUMI-G – GPU graphics processing units (AMD MI250)	0,535 € / GPU-hour
LUMI-P – Lustre parallel file system	0,005 € / TiB/h
LUMI-F – Flash memory storage	0,05 € / TiB/h
LUMI-O – LUMI-O – CEPH object storage	0,0025 € / TiB/h

LUMI Projects and User Accounts

- The use of LUMI is based on LUMI projects and personal user accounts.
- Each project has a Principal Investigator (PI) who is the lead contact and the responsible person for the project.
- A commercial LUMI project and user accounts for commercial company persons are requested via CSC ServiceDesk servicedesk@csc.fi

Running your jobs with LUMI Resources

- LUMI projects have CPU / GPU / data storage resources allocated according to your contract made with CSC.
- Running your application, i.e. a job, is based on batch jobs.
 - Your job is put to a queue to wait for requested resources, and run when they are available.
 - Note: **You pay for the used resources only**, not for service in general, like in monthly based cloud services.

LUMI service promise is not "24/7"

- LUMI is a highly advanced, multi-purpose "scientific instrument" that is mainly purposed to support research & development.
 - LUMI may not be able to fulfill the most demanding service quality requirements for the continuous production runs.

LUMI does not support processing of sensitive personal data

- At the moment LUMI have not the necessary security solutions in place to support processing of sensitive personal data (according to GDPR).

LUMI is batch-job computing environment

- LUMI is a secure computing environment with projects and data therein isolated from each other.
- However, it is possible to see metadata information of other projects in the batch job queues.
 - It is up to the company to decide based on LUMI security specifications whether the environment fulfills, e.g., your IP confidentiality requirements.

Main workflow steps with LUMI:

1. Prepare your data and application

- Move data to LUMI
- Install your application to LUMI, or
- Develop and test your application on LUMI, or
- Use available SW in LUMI

2. Run jobs i.e. your application under your project with LUMI job scheduler

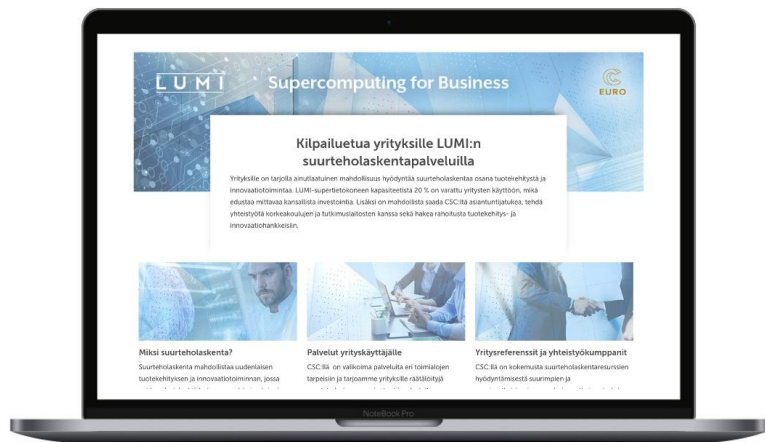
- Slurm Workload Manager

3. After job is done, you

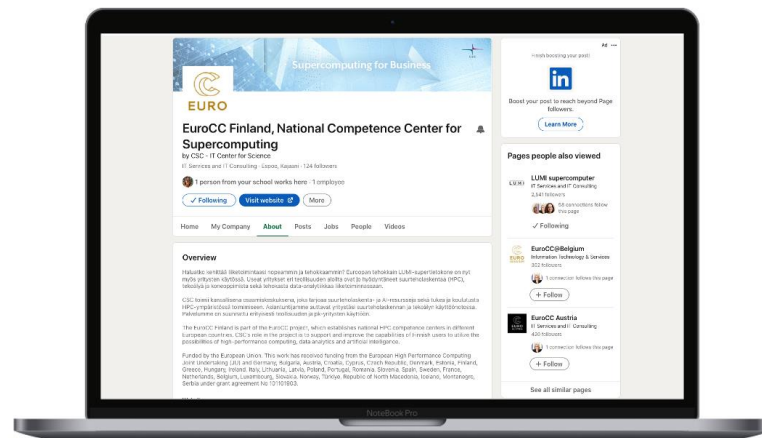
- Study the results on LUMI
 - LUMI-D nodes support interactive data analytics and visualization
 - Visualisation tools coming later along web remote access (Open OnDemand)
- In your own systems, after moving the results from LUMI

For more detailed documentation and guidelines visit: <https://docs.lumi-supercomputer.eu/>

More about Supercomputing for Business



www.csc.fi/en/solutions-for-business



www.linkedin.com/company/eurocc-finland

#supercomputingforbusiness #hpc #ai #data

How to get help / Further info



- CSC ServiceDesk servicedesk@csc.fi, (09) 457 2821, Mon-Fri 8:30 – 16:00
- CSC LUMI User Support <https://lumi-supercomputer.eu/user-support/need-help/>
- CSC LUMI documentation <https://docs.lumi-supercomputer.eu/>
- CSC training: <https://www.csc.fi/en/training#training-calendar>
- LUMI training: <https://www.lumi-supercomputer.eu/events/>
- CSC User guides and tutorials <https://docs.csc.fi/>
- Geocomputing with supercomputer: <https://research.csc.fi/geocomputing>

Extra:

- How does accessing a supercomputer look like?
 - <https://www.youtube.com/watch?v=HB9JUH0mPqI> (by EuroCC Sweden)

CSC's Business Solution Sales Team at Your Service



Pekka Uusitalo

Director

tel. +358 50 042 7720

pekka.uusitalo@csc.fi



Dan Still

Partnerships manager

tel. +358 50 381 9037

dan.still@csc.fi



Juhani Huttunen

Customer solution manager

tel. +358 40 581 1138

juhani.huttunen@csc.fi



Mikko Kerttula

Account manager

puh. +358 50 381 2766

mikko.kerttula@csc.fi



Morthen Mathisen

Senior Coordinator

puh. +358 50 3812935

morthen.mathisen@csc.fi