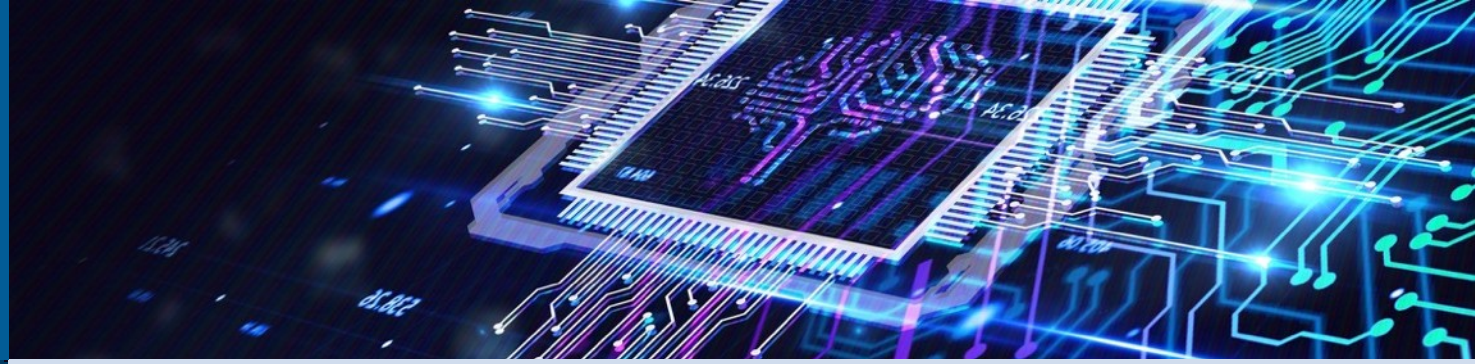
CSC

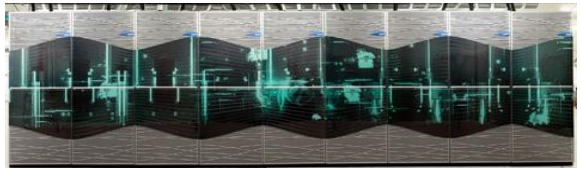
ICT Solutions for Brilliant Minds




# Tools for machine learning research at CSC

Machine learning with spatial data seminar – 27.2.2024
Mats Sjöberg, CSC

# CSC's computing infrastructure

**LUMI supercomputer**

11,912 GPUs
AMD MI250X

**Mahti supercomputer**

96 GPUs
NVIDIA A100

**Puhti supercomputer**

320 GPUs
NVIDIA V100

openstack

**cPouta** VMs

openstack

**ePouta** private VMs

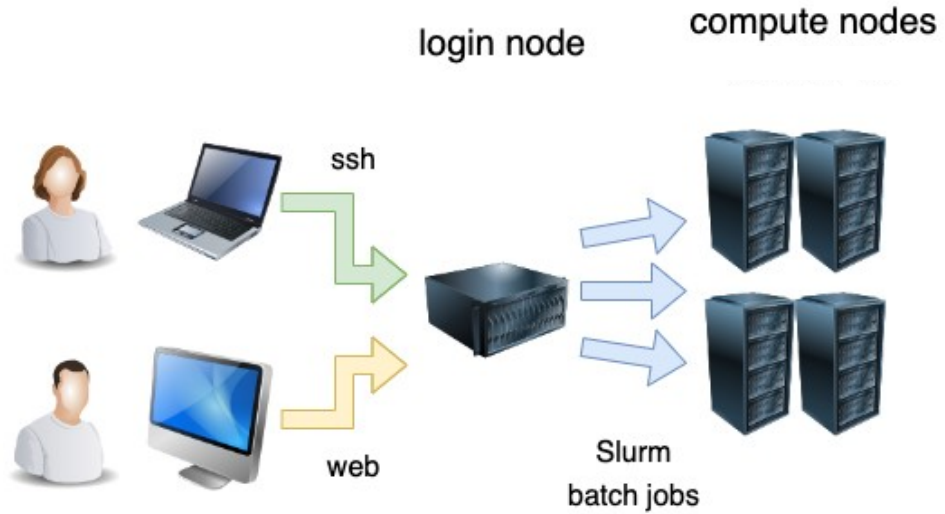
**Allas** object storage

**Rahti** containers

OPENSHIFT

# What is a supercomputer?

- HPC = high-performance computing
- A large number (cluster) of "normal" computers (nodes)
  - Actually pretty good computers: >100 GB RAM, 10s-100s CPU cores
  - Sometimes also GPUs
- Supercomputer = 100s or 1000s of such computers connected with
  - Shared file system
  - Fast network (100-200 GB/s)
- Dividing a heavy computation into smaller subtasks = *parallelism*



# GPU computing


- CPUs are optimized for latency whereas GPUs are optimized for throughput
  - That is a CPU calculates one thing very fast
  - GPU calculates many things simultaneously (a bit slower)

- Example: CSC's Puhti GPU nodes with V100's:

| | #cores | max clock speed | memory |
|---|---|---|---|
| 2 x Xeon CPUs | 2 x 20 | 3.9 GHz | 384 GB |
| 4 x V100 GPUs | 4 x 5120 | 1.455 GHz | 4 x 32 GB |

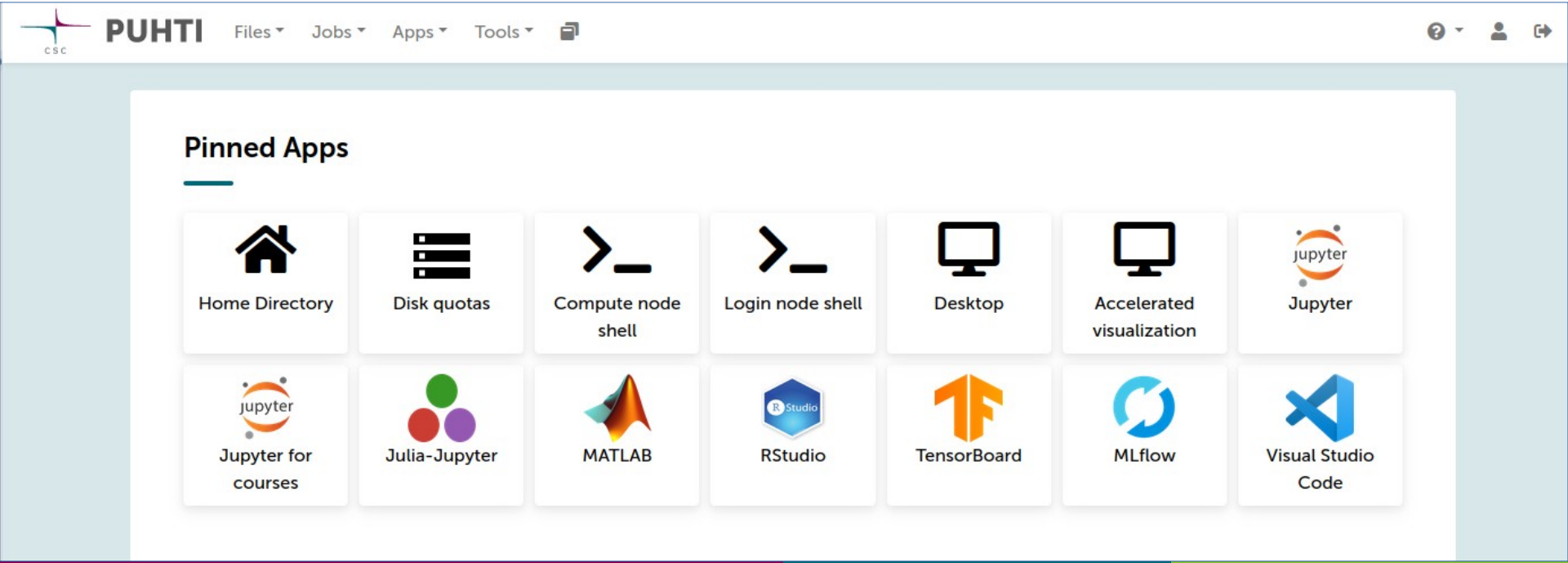
# Machine learning frameworks

- Major machine learning frameworks supported for NVIDIA and AMD GPUs: **PyTorch, TensorFlow** and **JAX**
  - For full application support see: https://docs.csc.fi/apps/
- Multi-GPU and multi-node jobs support, e.g., with PyTorch DistributedDataParallel and DeepSpeed
  - Machine learning guide: https://docs.csc.fi/support/tutorials/ml-guide/



# Open OnDemand web user interface

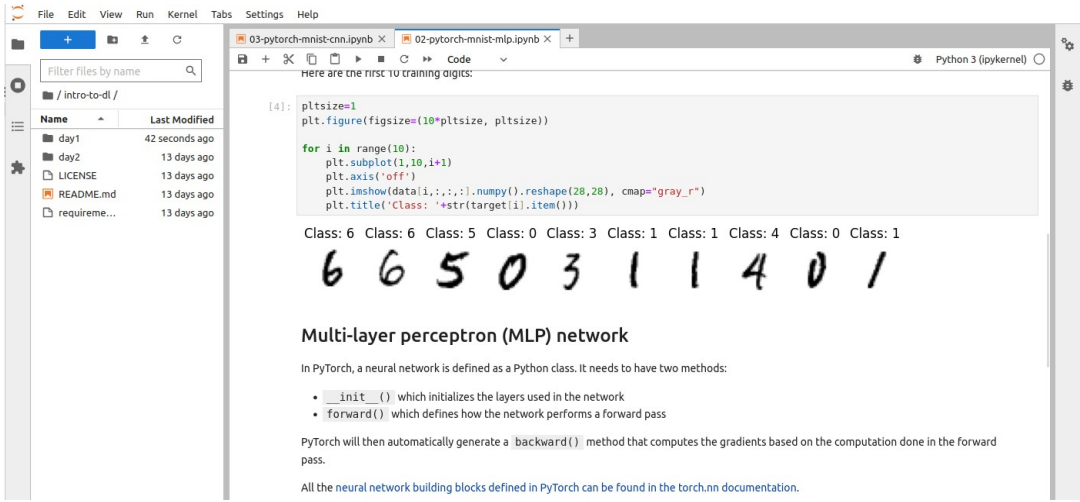
- Now also on Mahti and LUMI!

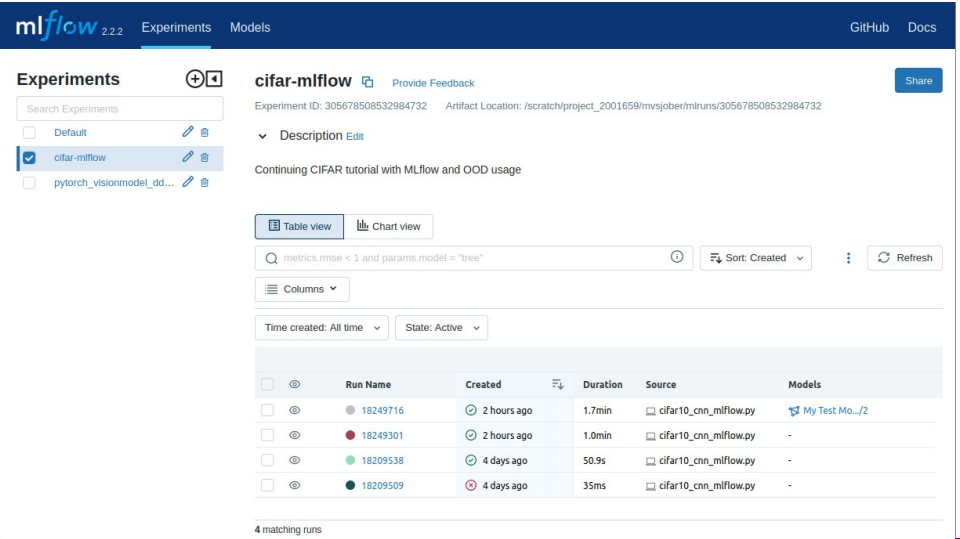


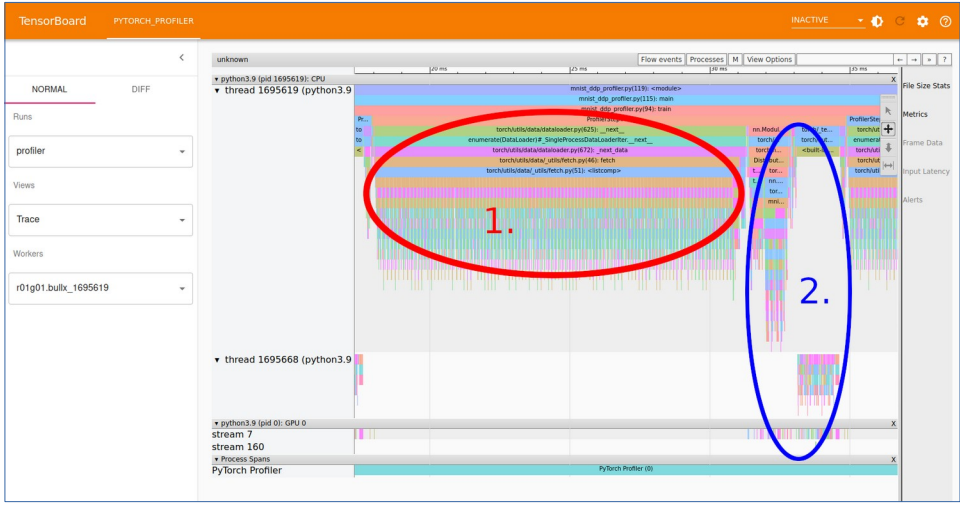
# Graphical tools

Easy to launch in web UI:

- JupyterLab Notebooks
- MLflow for tracking experiments
- TensorBoard (example: PyTorch profiler)



# CSC's cloud

- **Pouta** offers your own virtual server with full control of the software environment, but restricted computing performance compared to supercomputers
  - GPUs available, but somewhat limited
  - ePouta private cloud for sensitive data cases
- **Rahti** offers a more automatized container-based cloud environment, useful in particular for deploying web services
  - Limited GPU support (ask servicedesk!)


# Free science support!

- *Most resources and support **free of charge** for academic research and education by Finnish higher education institutions, and by state research institutes*

- In addition to normal technical support, CSC's specialists provide **science support**, e.g., in biosciences, chemistry, engineering, data analytics, machine learning, geosciences and digital humanities

- CSC provides training in many of these areas: https://www.csc.fi/training

**Mats Sjöberg**

Senior Machine Learning Specialist

CSC – IT Center for Science Ltd, Finland

mats.sjoberg@csc.fi

facebook.com/CSCfi

twitter.com/CSCfi

linkedin.com/company/csc---it-center-for-science

github.com/CSCfi