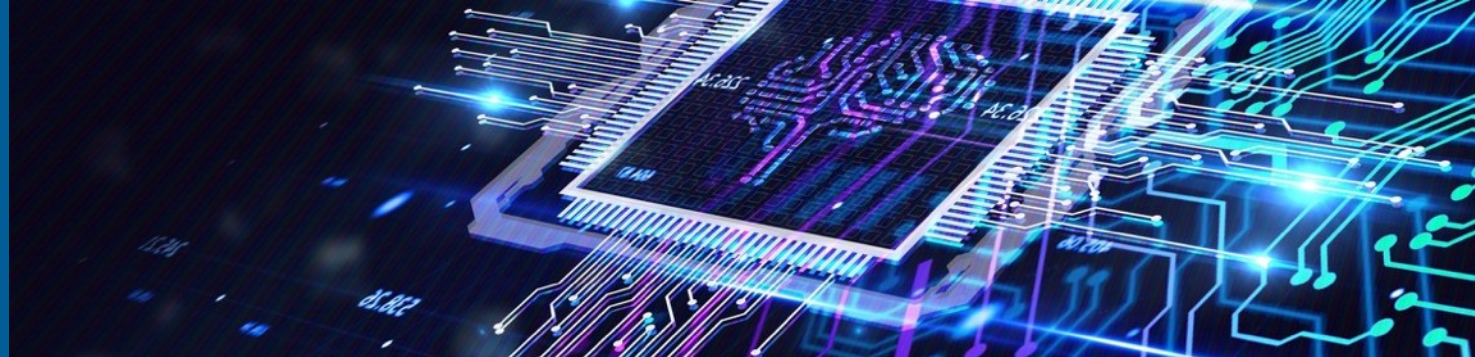




CSC

ICT Solutions for
Brilliant Minds

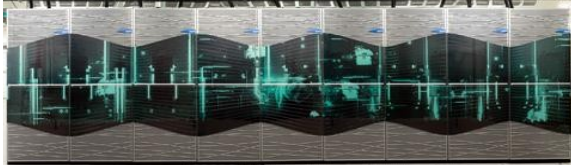


Tools for machine learning research at CSC

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



CSC's computing infrastructure



cPouta
VMs



ePouta
private VMs

LUMI
supercomputer

11,912 GPUs
AMD MI250X

Mahti
supercomputer

96 GPUs
NVIDIA A100

Puhti
supercomputer

320 GPUs
NVIDIA V100

Allas
object storage

Rahti
containers

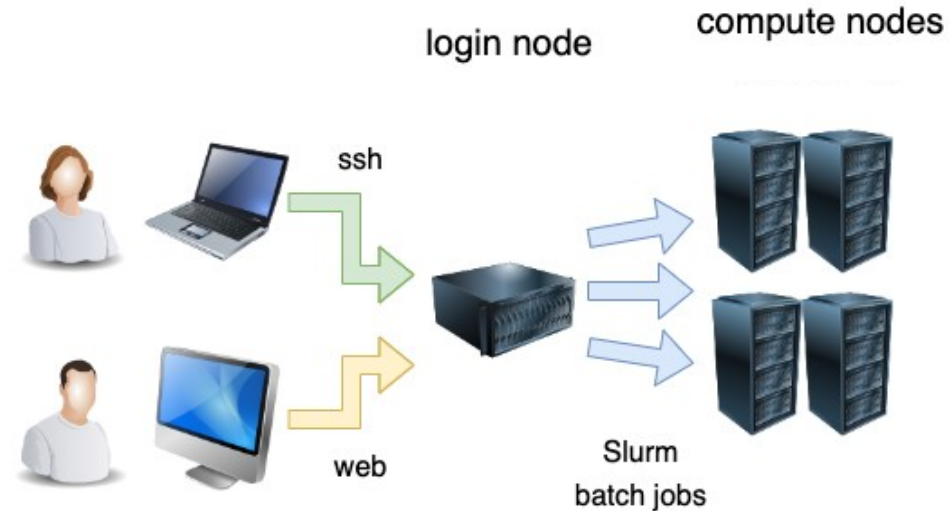


OPENSIFT



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 |



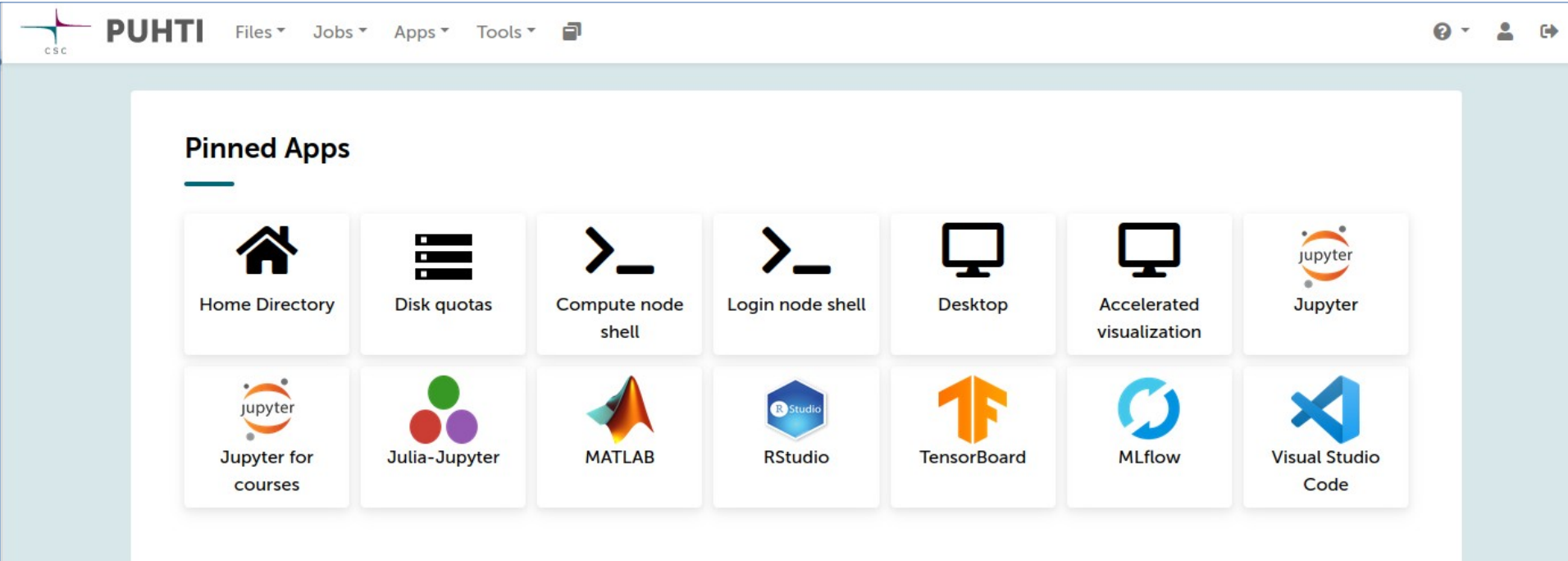
Machine learning frameworks

- 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!



The screenshot displays the Open OnDemand web user interface. At the top, there is a navigation bar with the PUHTI logo and the CSC logo, followed by menu items: Files, Jobs, Apps, Tools, and a folder icon. On the right side of the navigation bar, there are icons for help, user profile, and a share icon.

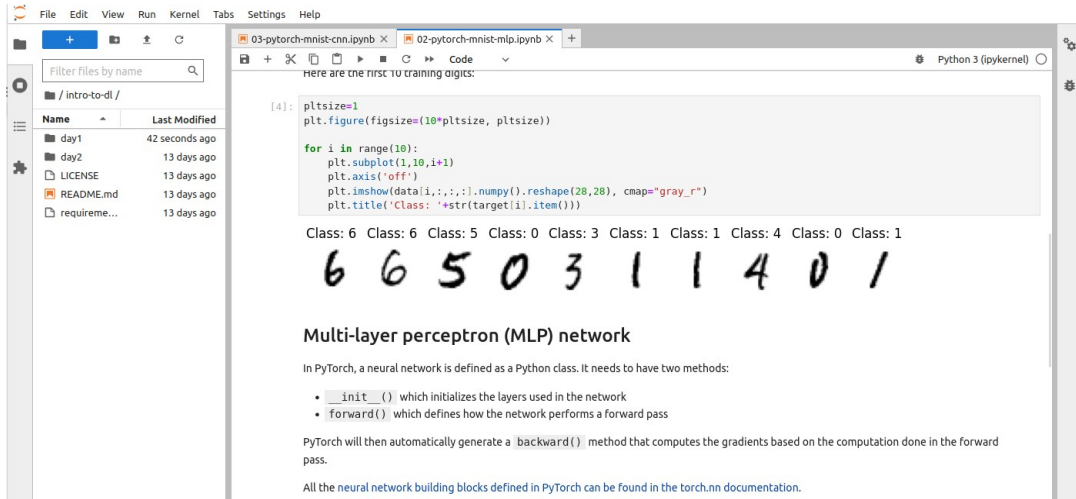
The main content area is titled "Pinned Apps" and features a grid of 14 application tiles, arranged in two rows of seven. Each tile contains an icon and a label:

- Home Directory (house icon)
- Disk quotas (three horizontal bars icon)
- Compute node shell (terminal icon)
- Login node shell (terminal icon)
- Desktop (monitor icon)
- Accelerated visualization (monitor icon)
- Jupyter (Jupyter logo)
- Jupyter for courses (Jupyter logo)
- Julia-Jupyter (Julia logo)
- MATLAB (MATLAB logo)
- RStudio (RStudio logo)
- TensorBoard (TensorBoard logo)
- MLflow (MLflow logo)
- Visual Studio Code (Visual Studio Code logo)

Graphical tools

Easy to launch in web UI:

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



```
File Edit View Run Kernel Tabs Settings Help
+ - + - + -
Filter files by name
/ intro-to-dl /
Name Last Modified
day1 42 seconds ago
day2 13 days ago
LICENSE 13 days ago
README.md 13 days ago
requireme... 13 days ago

03-pytorch-mnist-cnn.ipynb x 02-pytorch-mnist-mlp.ipynb x
Python 3 (pykernel)
Here are the first 10 training digits:

[4]: pltsize=1
plt.figure(figsize=(10*pltsize, pltsize))

for i in range(10):
    plt.subplot(1,10,i+1)
    plt.axis('off')
    plt.imshow(data[i,:,:].numpy().reshape(28,28), cmap='gray_r')
    plt.title('Class: '+str(target[i].item()))

Class: 6 Class: 6 Class: 5 Class: 0 Class: 3 Class: 1 Class: 1 Class: 4 Class: 0 Class: 1
6 6 5 0 3 1 1 4 0 1

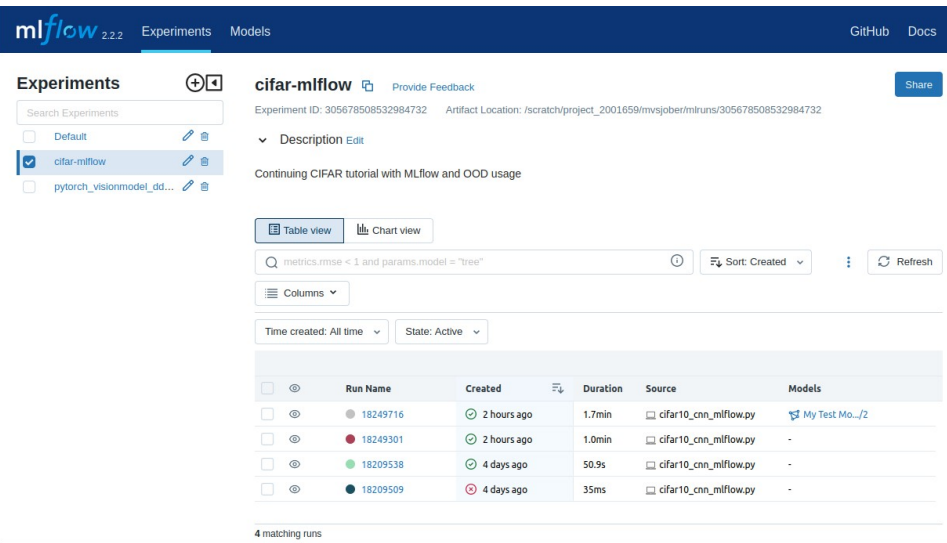
Multi-layer perceptron (MLP) network

In PyTorch, a neural network is defined as a Python class. It needs to have two methods:

- __init__() which initializes the layers used in the network
- forward() which defines how the network performs a forward pass

PyTorch will then automatically generate a backward() method that computes the gradients based on the computation done in the forward pass.

All the neural network building blocks defined in PyTorch can be found in the torch.nn documentation.
```



mlflow 2.2.2 Experiments Models GitHub Docs

Experiments Default cifar-mlflow pytorch_visionmodel_dd...

cifar-mlflow Provide Feedback Share

Experiment ID: 305678508532984732 Artifact Location: /scratch/project_2001659/mvsjobser/mirruns/305678508532984732

Description Edit

Continuing CIFAR tutorial with MLflow and OOD usage

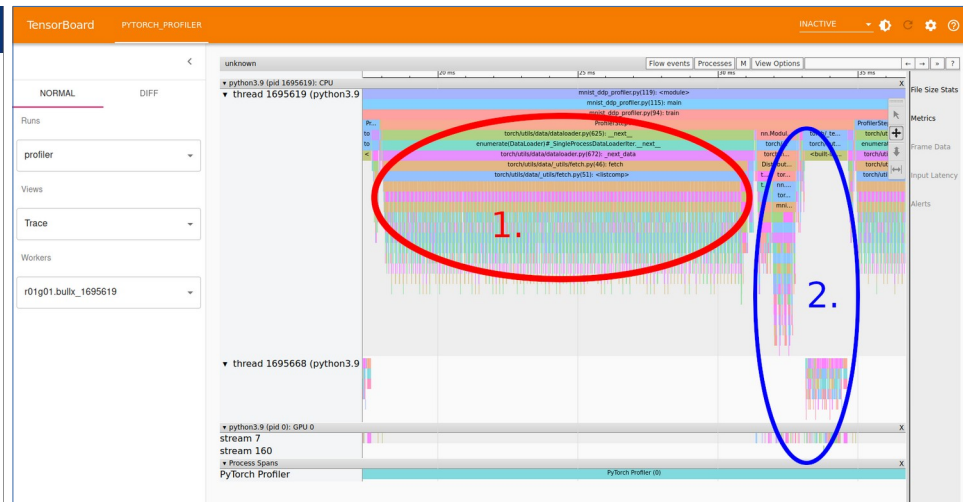
Table view Chart view

metrics.rmse < 1 and params.model = "tree" Sort: Created Refresh

Columns Time created: All time State: Active

| <input type="checkbox"/> | <input type="checkbox"/> | Run Name | Created | Duration | Source | Models |
|--------------------------|--------------------------|----------|-------------|----------|-----------------------|-----------------|
| <input type="checkbox"/> | <input type="checkbox"/> | 18249716 | 2 hours ago | 1.7min | cifar10_cnn_mlflow.py | My Test Mo.../2 |
| <input type="checkbox"/> | <input type="checkbox"/> | 18249301 | 2 hours ago | 1.0min | cifar10_cnn_mlflow.py | - |
| <input type="checkbox"/> | <input type="checkbox"/> | 18209538 | 4 days ago | 50.9s | cifar10_cnn_mlflow.py | - |
| <input type="checkbox"/> | <input type="checkbox"/> | 18209509 | 4 days ago | 35ms | cifar10_cnn_mlflow.py | - |

4 matching runs



TensorBoard PYTORCH_PROFILER INACTIVE

unknown 100 ms 100 ms 100 ms

python3.9 (pid 1695619): CPU

thread 1695619 (python3.9)

metrics_dig_profiler.py(115): main

torch/nn/modules/conv.py(141): train

torch/nn/modules/conv.py(23): next

torch/nn/modules/conv.py(72): next_data

torch/nn/modules/conv.py(46): torch

torch/nn/modules/conv.py(51): <listcomp>

1.

2.

python3.9 (pid 0): GPU 0

stream 7

stream 160

Process Spans

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