Find information about the course and ask questions here:

https://siili.rahtiapp.fi/lumi-ai-workshop-oct25

LUMI

Welcome

Moving your AI training jobs to LUMI workshop

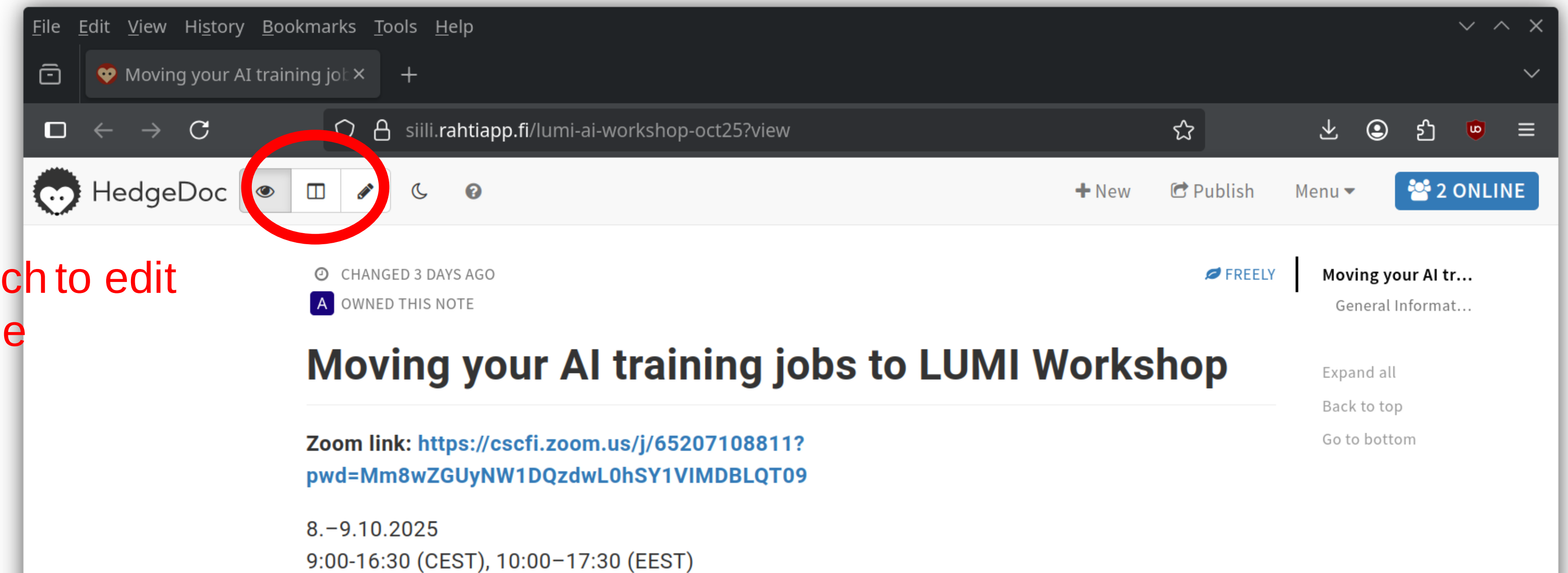
8.10.2025

EURO

FINLAND

Find information about the course and ask questions here:

https://siili.rahtiapp.fi/lumi-ai-workshop-oct25


File Edit View History Bookmarks Tools Help
Moving your AI training jot
siili.rahtiapp.fi/lumi-ai-workshop-oct25?view
HedgeDoc
+ New
Publish
Menu
2 ONLINE
FREELY
Moving your AI tr...
General Informat...
Expand all
Back to top
Go to bottom

CHANGED 3 DAYS AGO

A OWNED THIS NOTE

Moving your AI training jobs to LUMI Workshop

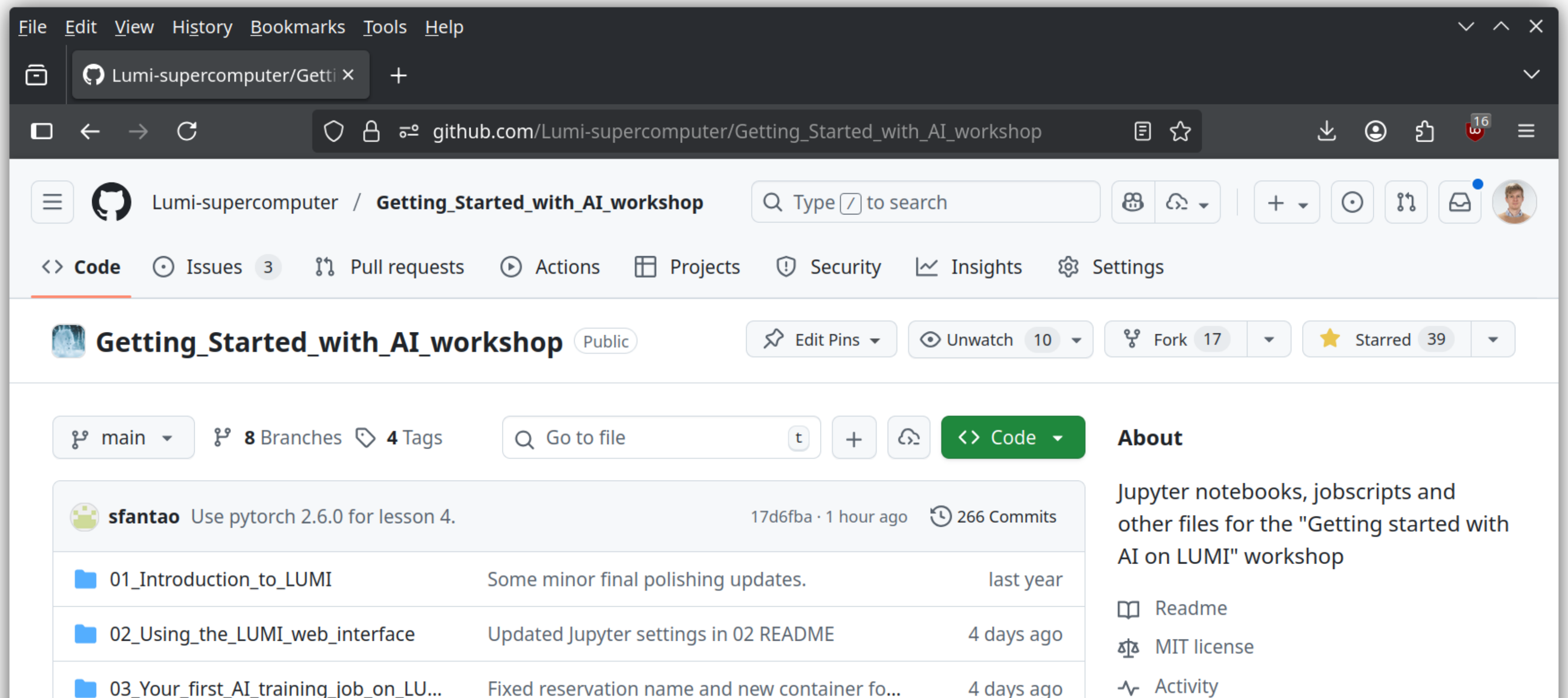
Zoom link: https://cscfi.zoom.us/j/65207108811?pwd=Mm8wZGUyNW1DQzdwL0hSY1VIMDBLQT09

8.–9.10.2025
9:00-16:30 (CEST), 10:00–17:30 (EEST)

Switch to edit mode

Exercise material

https://github.com/Lumi-supercomputer/Getting_Started_with_AI_workshop



File Edit View History Bookmarks Tools Help
Lumi-supercomputer/Getti
github.com/Lumi-supercomputer/Getting_Started_with_AI_workshop
Lumi-supercomputer / Getting_Started_with_AI_workshop
Type / to search
Code
Issues 3
Pull requests
Actions
Projects
Security
Insights
Settings
Getting_Started_with_AI_workshop
Public
Edit Pins
Unwatch 10
Fork 17
Starred 39
main
8 Branches
4 Tags
Go to file
t
Code
About

Jupyter notebooks, jobscripts and other files for the "Getting started with AI on LUMI" workshop

- Readme
- MIT license
- Activity

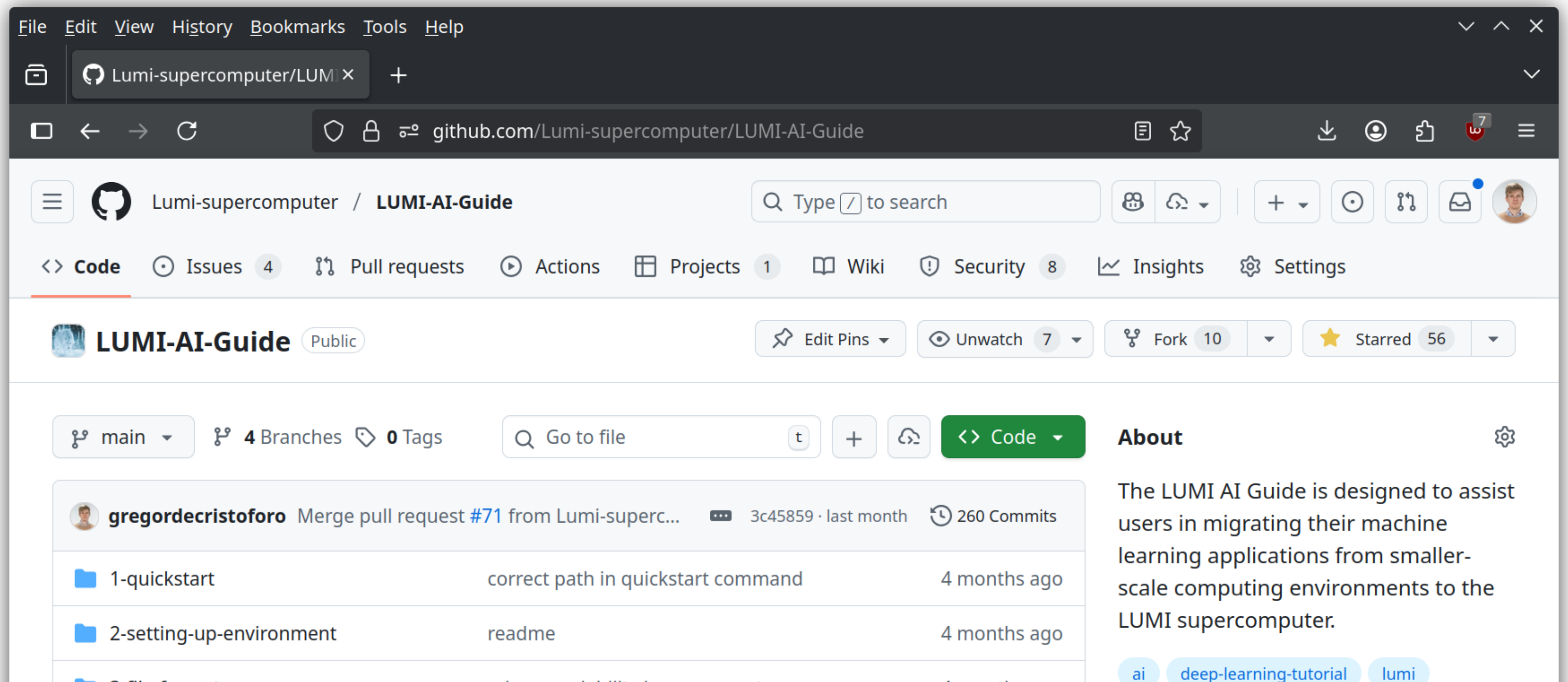
sfantao Use pytorch 2.6.0 for lesson 4. 17d6fba · 1 hour ago 266 Commits

01_Introduction_to_LUMI	Some minor final polishing updates.	last year
02_Using_the_LUMI_web_interface	Updated Jupyter settings in 02 README	4 days ago
03_Your_first_AI_training_job_on_LU...	Fixed reservation name and new container fo...	4 days ago

LUMI AI Guide

LUMI

https://github.com/Lumi-supercomputer/LUMI-AI-Guide



gregordecristoforo Merge pull request #71 from Lumi-superc... 3c45859 · last month 260 Commits

1-quickstart	correct path in quickstart command	4 months ago
2-setting-up-environment	readme	4 months ago

About

The LUMI AI Guide is designed to assist users in migrating their machine learning applications from smaller-scale computing environments to the LUMI supercomputer.

Introduction to LUMI

Moving your AI training jobs to LUMI workshop

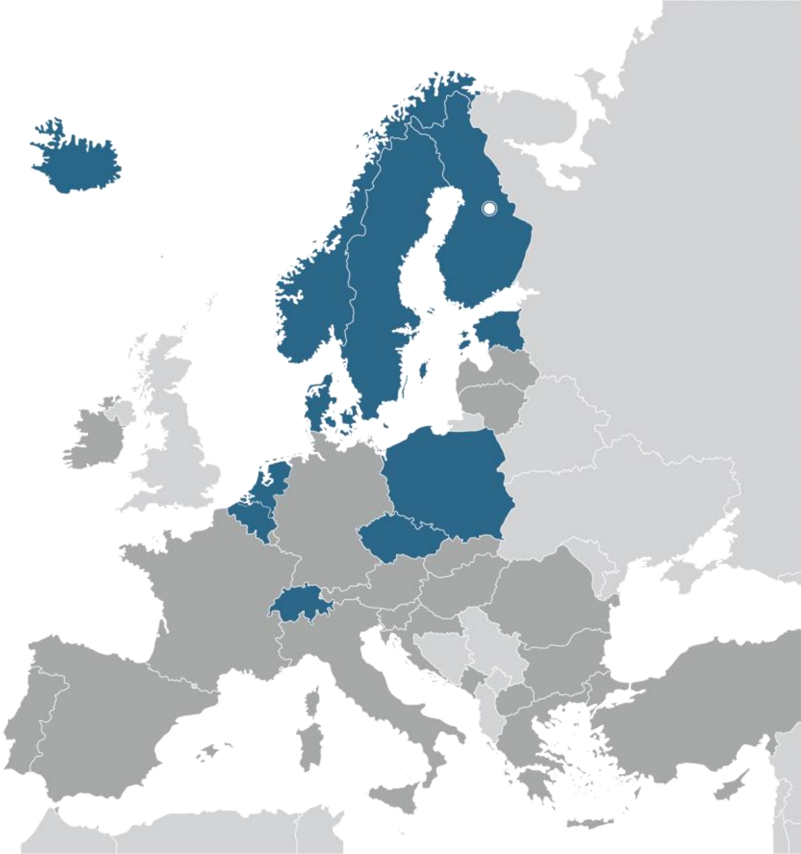
8.10.2025

LUMI is not one single computer

It behaves quite a bit different than your local computer

LUMI is a very fast computer in Europe

- 9th fastest computer in world (TOP500)
- Operated by LUMI consortium
  - 11 countries collaborating
  - 50 % financed by EuroHPC JU
- Located in Kajaani, Finland
- Distributed LUMI user support team (LUST)
  - One full time employee equiv alent from each country
  - Offer email support, courses, workshops, ...
  - Responsible of software stack

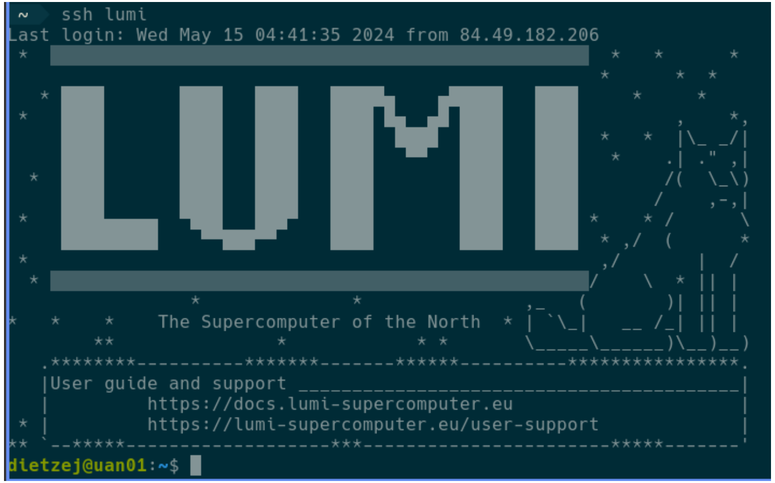


LUMI is a cluster of individual computers

- LUMI is not one superfast computer
- Instead it consists of a few thousand individual computers (“nodes”)
- All of them are connected by a fast interconnect
- Speed comes from parallelization

Two ways of connecting

Command line interface

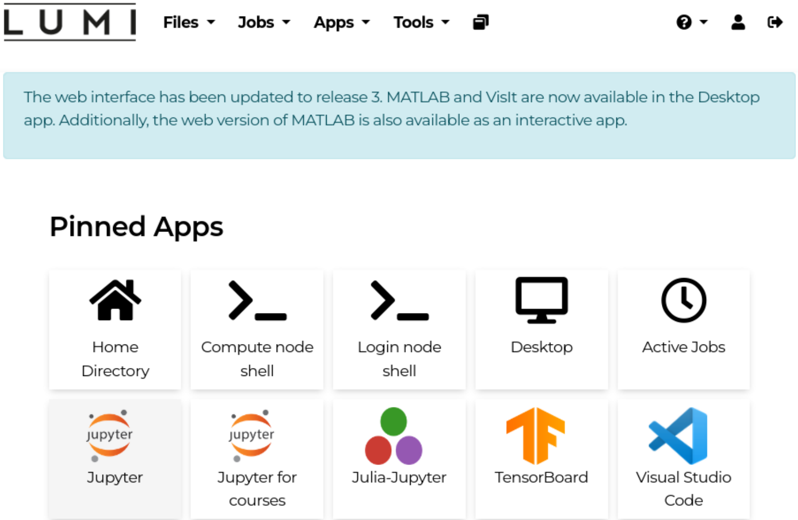


```
~  ssh lumi
Last login: Wed May 15 04:41:35 2024 from 84.49.182.206

                The Supercomputer of the North

  .********_________*******______******__________****************.
  |User guide and support ________________________________________|
  |          https://docs.lumi-supercomputer.eu                   |
  |          https://lumi-supercomputer.eu/user-support           |
  `--*****_____________________***_______________________*****______'
dietzej@uan01:~$ 
```

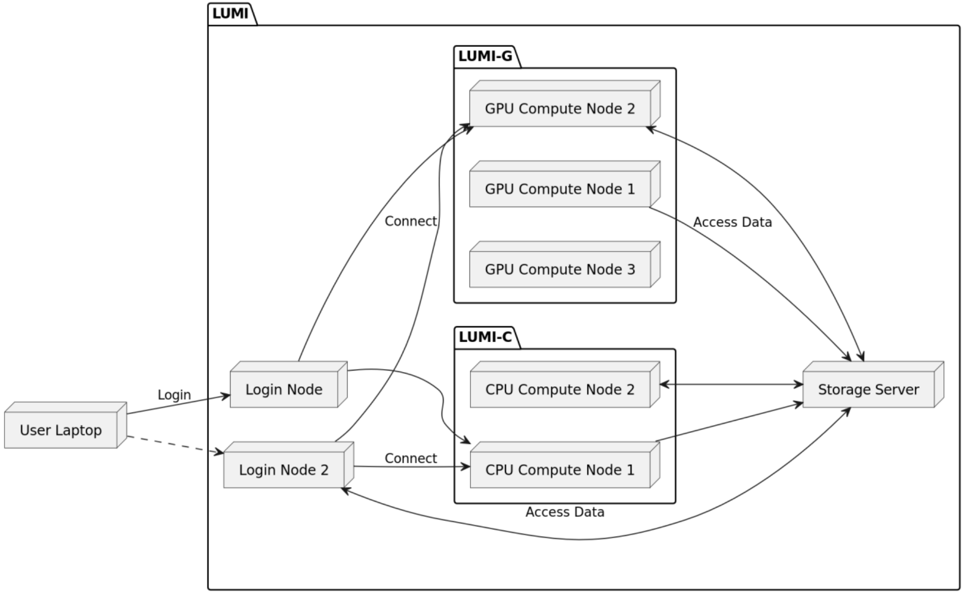
Browser based interface (OpenOnDemand)



The web interface has been updated to release 3. MATLAB and VisIt are now available in the Desktop app. Additionally, the web version of MATLAB is also available as an interactive app.

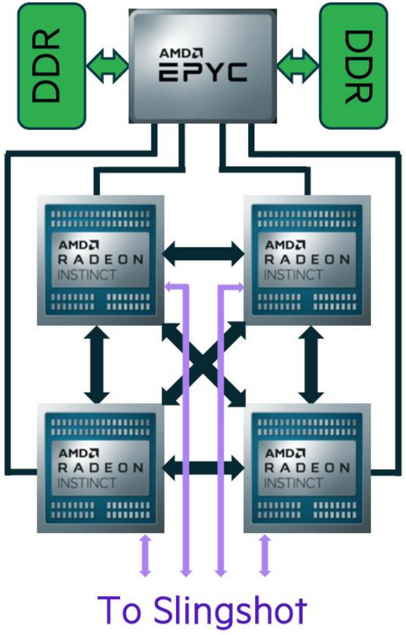
LUMI consists of different parts

- Computers
  - Login nodes – UAN (user access nodes)
  - CPU compute nodes – LUMI-C
  - GPU compute nodes – LUMI-G
  - Visualisation nodes – LUMI-D
- Storage
  - 80 PB main parallel storage – LUMI-P
  - 8.5 PB accelerated storage – LUMI-F
  - 30 PB object-based storage – LUMI-O
- Interconnect
  - HPE Slingshot 13
  - Connects everything

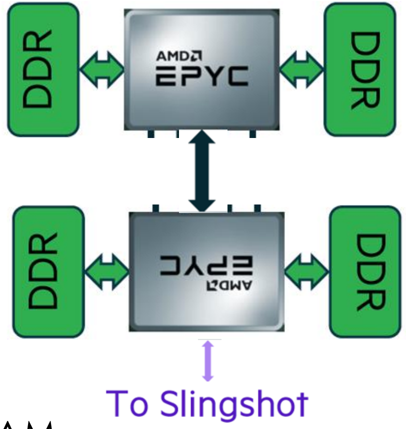


LUMI-C and -G are quite different

LUMI-G



2978 nodes with
4x MI250X (2 x 64GB)
1x AMD Trento CPU
512 GB RAM
4x 200 Gbit/s NIC

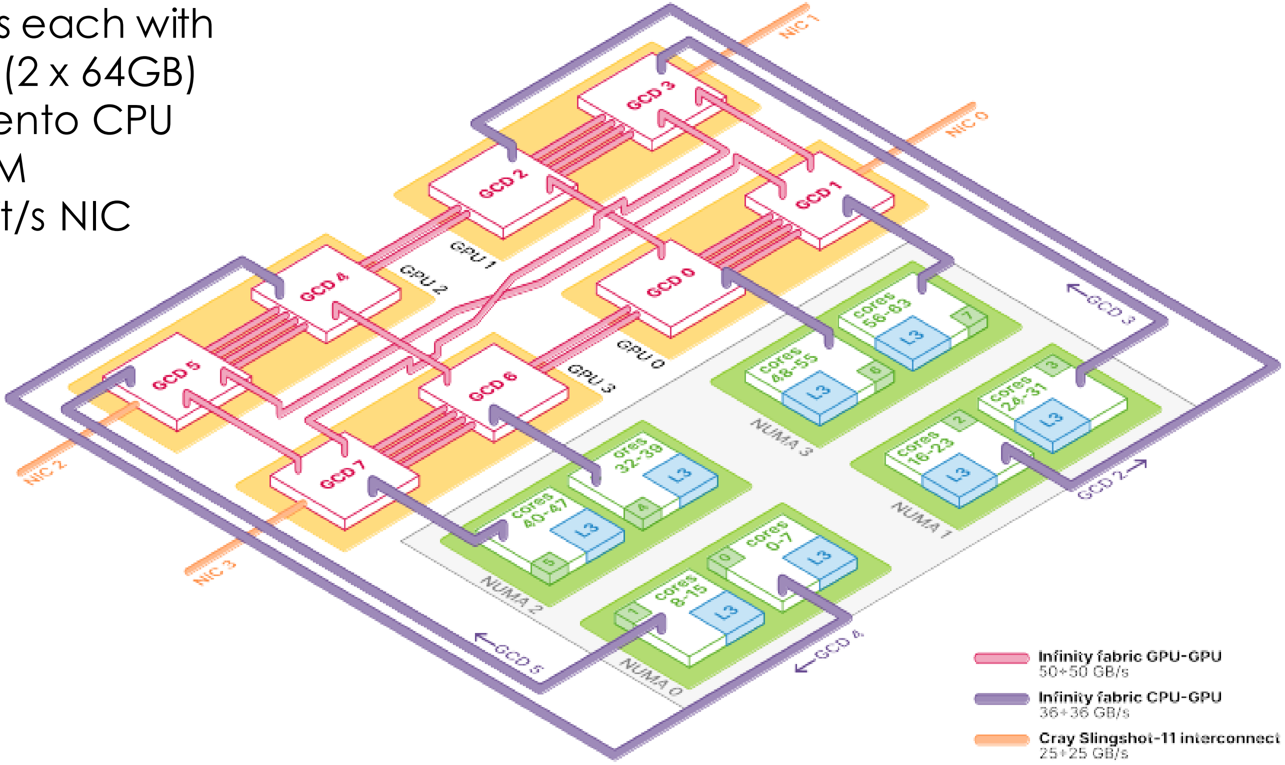


1888 nodes with 256 GB,
128 with 512 GB and 32 with 1 TB RAM
2x 64-core AMD Milan CPUs
1 x 200 Gbit/s NIC

LUMI-C

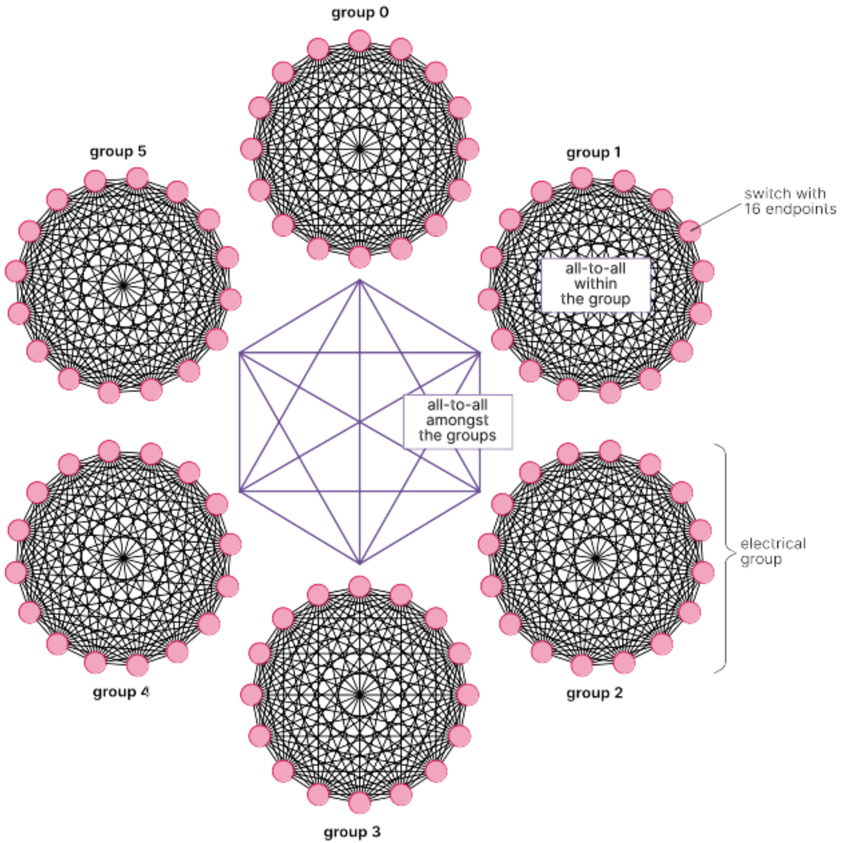
GPU nodes are the center of LUMI

2978 nodes each with
4 x MI250X (2 x 64GB)
1 x AMD Trento CPU
512 GB RAM
4 x 200 Gbit/s NIC



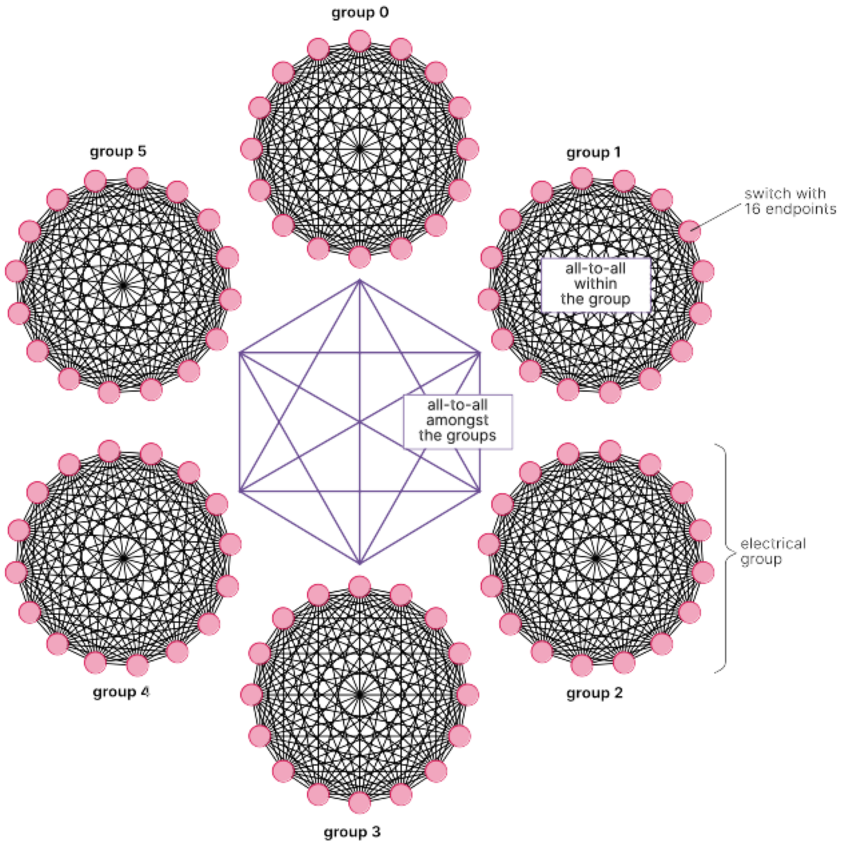
Interconnect is the fast backbone of LUMI

LUMI



- Connnects all nodes
- Has a similar role to ethernet
- Much higher speed bandwidth

Interconnect is the fast backbone of LUMI

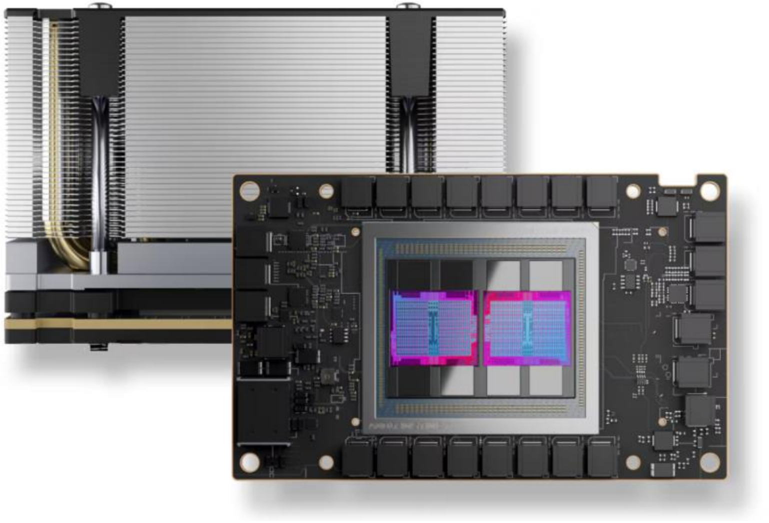


- Slingshot in Dragonfly topology
  - Each G node is connected to 4 switches
  - All-to-all amongst switches in a group
  - All-to-all between groups
  - Max of 3 switch hops
- Make sure to use it

AMD is not Nvidia

But the differences are quite small

Our GPUs are confusing



Each AMD Instinct MI250X

- 2 Graphics Compute Die (GCD)
- 110 compute units per GCD with each 64 stream processors
- 64 GB HBM GPU memory per GCD
- Each process can only use 64GB max – not 128GB

Different names but usually same concept

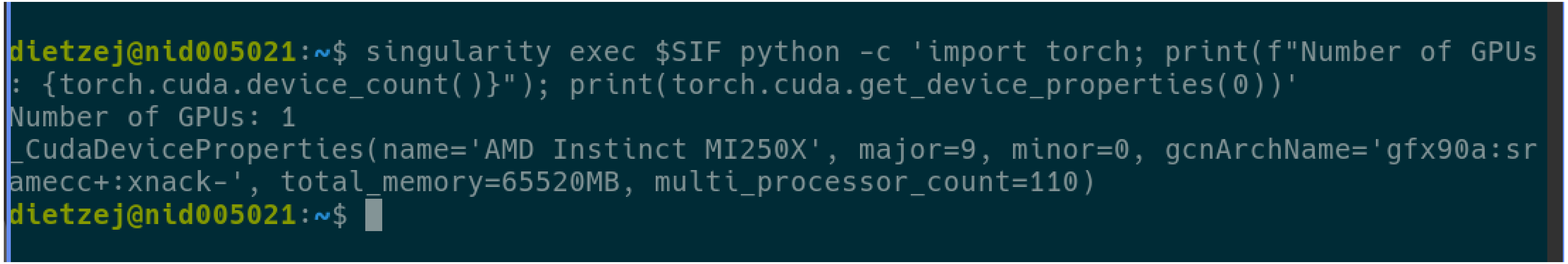
NVIDIA		AMD
PyTorch	ML Training	PyTorch
Infiniband / RoCE	Networking Between Nodes	HPE Slingshot
NCCL	Cross-GPU Communication	RCCL
CUDA / CuDNN	Software Stack	ROCm
A100, H100	GPU	MI250X, MI300X

ROCm is not CUDA

- ROCm is the equivalent software stack to Nividia’s CUDA
- Basically drop-in replacement
- Very similar concept
- Some small differences
- Consists of
  - GPU drivers
  - Compilers and profilers
  - Math and communication libraries

PyTorch makes it simple

- Both CUDA and ROCm are loaded with `cuda` submodule
- Check whether you can see any GPUs with `torch.cuda.device_count()`

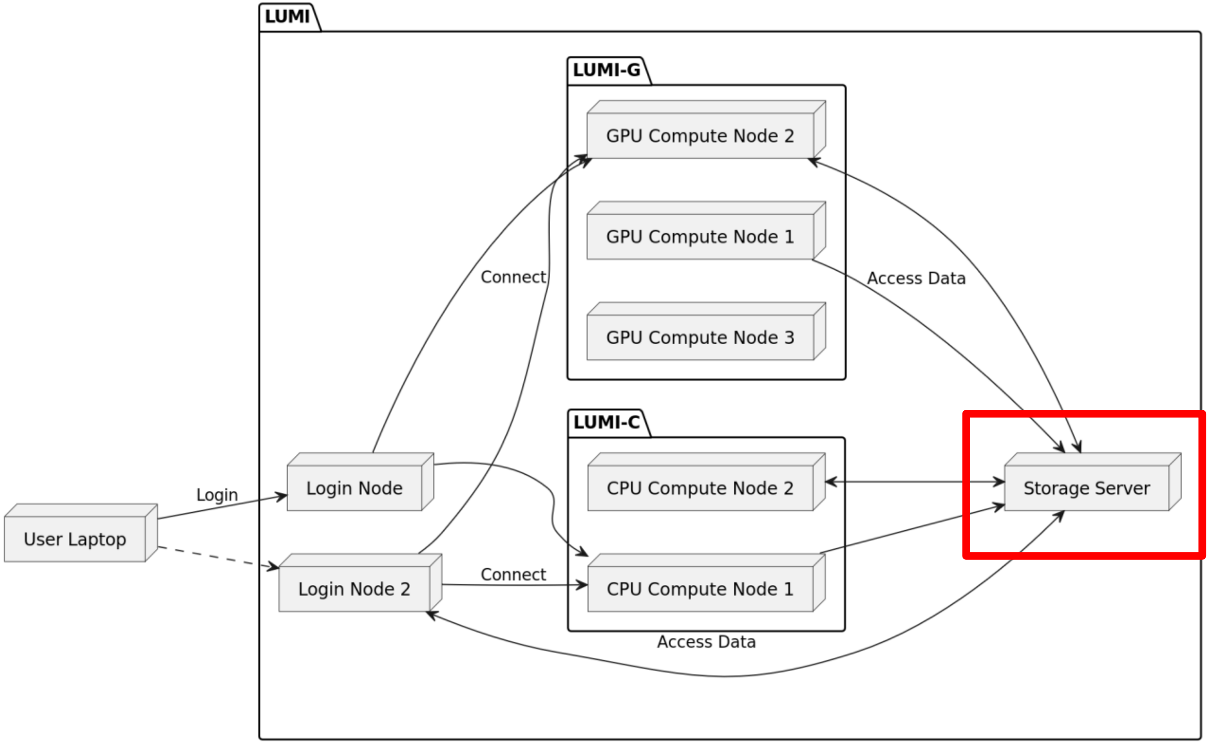


```
dietzej@nid005021:~$ singularity exec $SIF python -c 'import torch; print(f"Number of GPUs
: {torch.cuda.device_count()}"); print(torch.cuda.get_device_properties(0))'
Number of GPUs: 1
_CudaDeviceProperties(name='AMD Instinct MI250X', major=9, minor=0, gcnArchName='gfx90a:sr
amecc+:xnack-', total_memory=65520MB, multi_processor_count=110)
dietzej@nid005021:~$
```

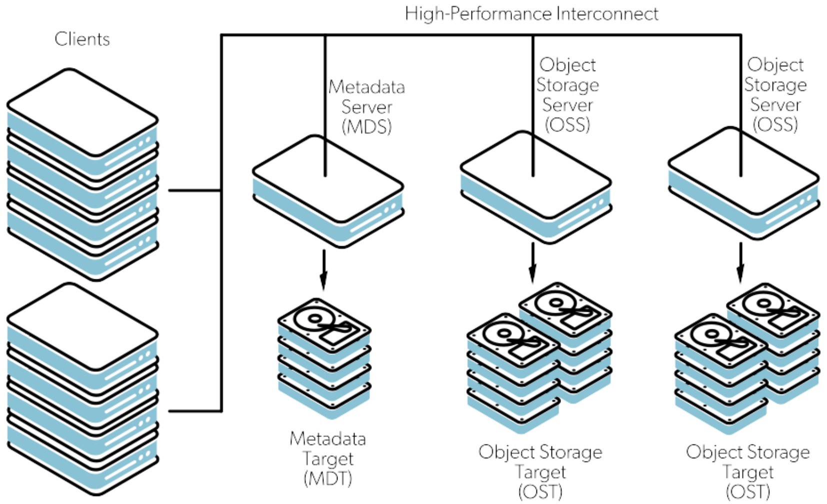
Storage is not as easy as on your laptop

But if you follow some rules you will be fine

There is more than one storage server



LUMI has three storage systems


High-Performance Interconnect
Clients
Metadata Server (MDS)
Object Storage Server (OSS)
Object Storage Server (OSS)
Metadata Target (MDT)
Object Storage Target (OST)
Object Storage Target (OST)

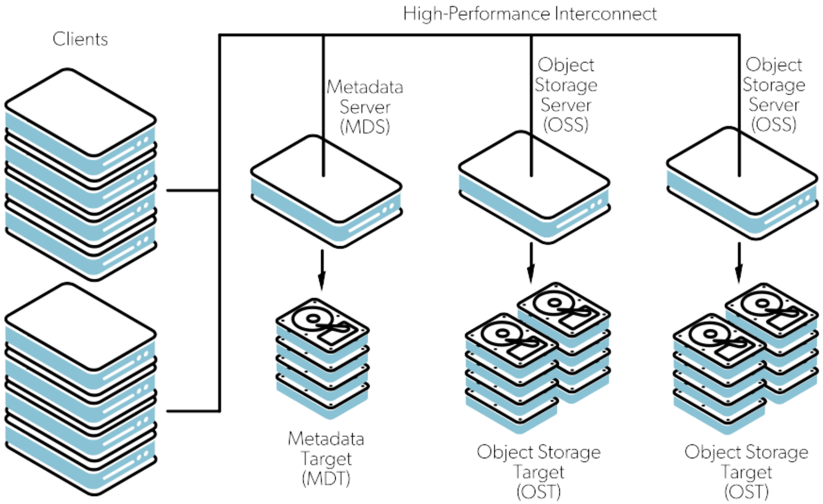
- LUMI-P
  - Lustre file system
  - Disk based
  - 4 independent systems with each 20 PB
- LUMI-F
  - Lustre file system
  - Solid-state (flash) based
  - 8.5 PB
- LUMI-O
  - Object storage based
  - Disk based
  - 30 PB

There are no local disks

- Compute nodes have no local disks
- Instead network storage (LUMI-P & -F) has to be used
- 4 storage areas

Area	Path	Usage
User home	/users/<username>	Configuration files
Project persistent	/project/<project>	Installations + final results
Project scratch	/scratch/<project>	Input + Intermediate results
Project flash	/flash/<project>	Input if high bandwidth is needed

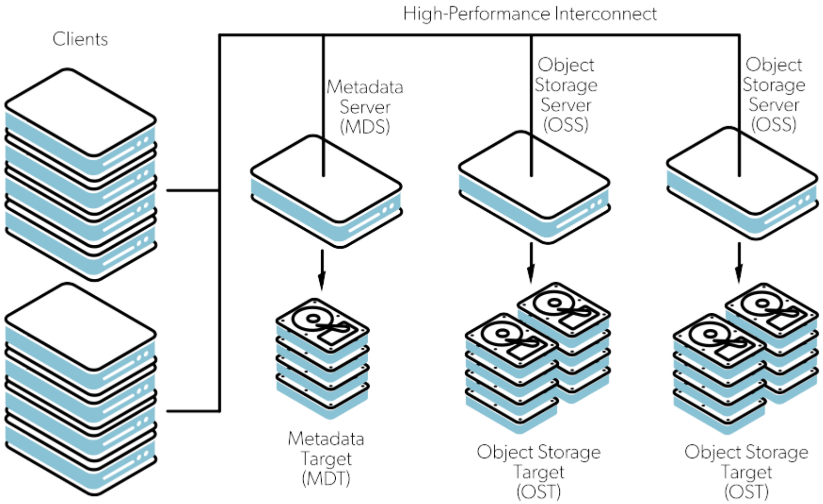
Lustre doesn’t like many small files


High-Performance Interconnect
Clients
Metadata Server (MDS)
Object Storage Server (OSS)
Object Storage Server (OSS)
Metadata Target (MDT)
Object Storage Target (OST)
Object Storage Target (OST)

Lustre consists of 3 parts

- Client
  Compute or login node that wants to access a file
- Metadata Server
  - Doesn’t store file content
  - Just metadata like location, size, ...
  - Tells client where to find file
- Object Storage Server
  - Stores actual file content
  - Either complete file or parts
  - Sends and receives data to/from client

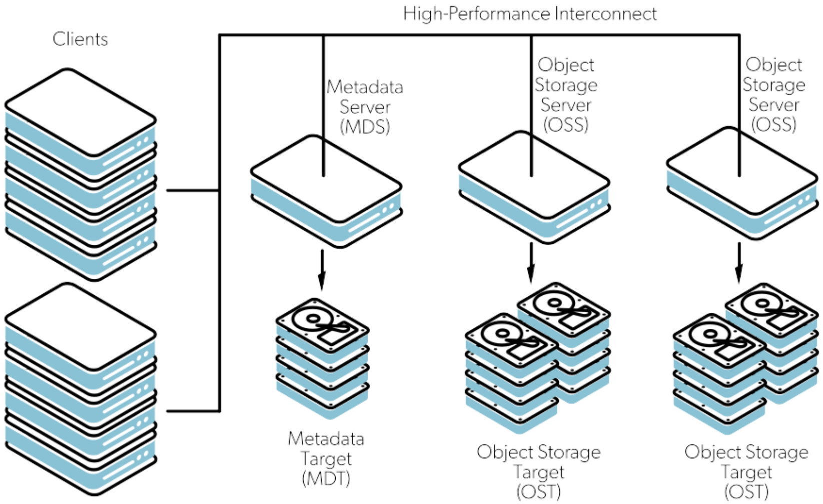
Lustre doesn’t like many small files


High-Performance Interconnect
Clients
Metadata Server (MDS)
Object Storage Server (OSS)
Object Storage Server (OSS)
Metadata Target (MDT)
Object Storage Target (OST)
Object Storage Target (OST)

Problem with many small files

- For each file the client queries the Metadata server (MDS)
- Many object storage servers but only one MDS
- MDS can get overloaded by queries if many clients ask for lots of small files each

Lustre likes few large files



High-Performance Interconnect
Clients
Metadata Server (MDS)
Object Storage Server (OSS)
Object Storage Server (OSS)
Metadata Target (MDT)
Object Storage Target (OST)
Object Storage Target (OST)

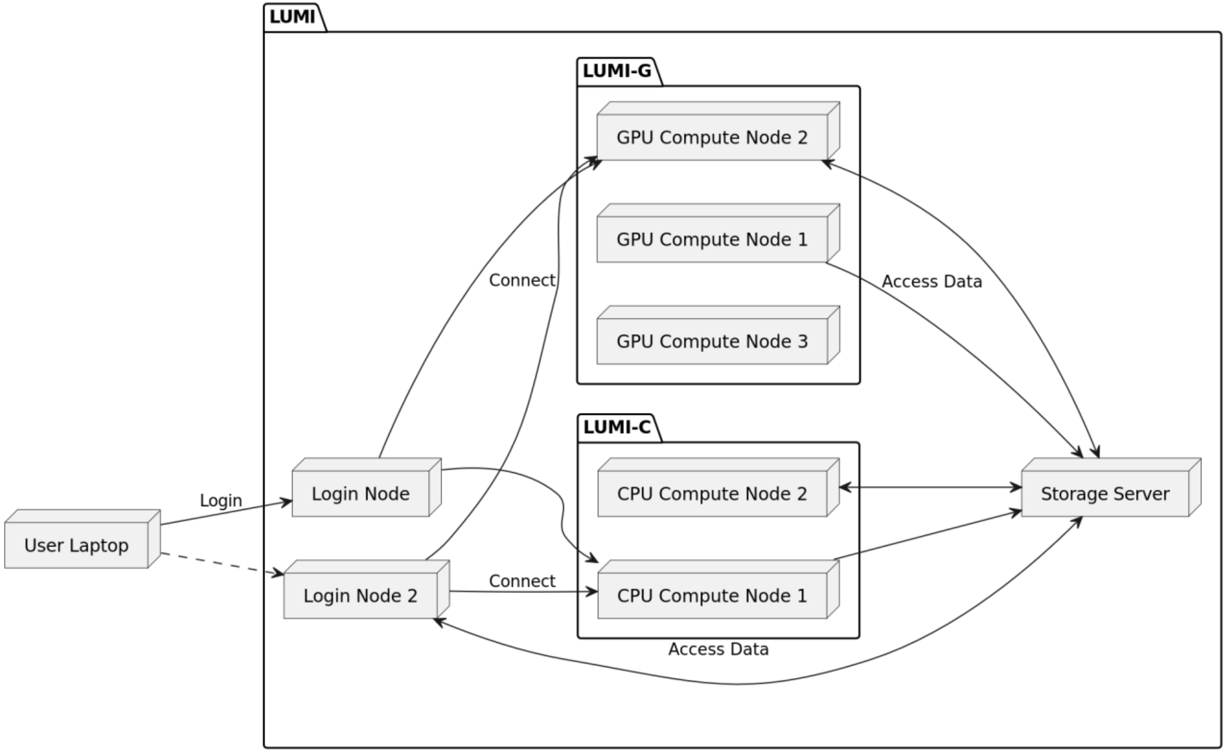
To avoid overloading MDS

- Avoid many (thousands) small files
- Avoid opening/closing many files in short time
- Bundle files together
- Python environments can be a problem → discuss later

What about /tmp?

- Compute nodes don’t have local disks/flash
- /tmp resides in memory
- Consumes space of your memory allocation
- Remember to allocate enough memory if you want to use /tmp

LUMI consists of different parts



Use them well and you will get great

Questions?