LUMI

### Find information about the course and ask questions here: https://md.sigma2.no/lumi-ai-workshop-may25

## Welcome

LU



NETHERLANDS

Moving your AI training jobs to LUMI workshop 27.5.2025



# Find information about the course and ask questions here:

https://md.sigma2.no/lumi-ai-workshop-may25

	<u>F</u> ile <u>E</u> dit <u>V</u> iew Hi <u>s</u> tory <u>B</u> ookmarks <u>T</u> ools <u>H</u> elp Moving your Al training jobs × +					
	$\leftrightarrow \rightarrow \mathbf{C}$	s://md. <b>sigma2.no</b> /lumi-ai-workshop-may25?view 🗉 🕁 🔍 Search	ٹے لیے	<b>02</b> 🖲 🖆 🖷 🖕 ≫ 🚍		
Switch	HedgeDoc HedgeDoc	Changed a minute ago	+ New 😁	Publish Menu ▼ 😤 2 ONLINE		
		Moving your AI training jobs to LUMI	Workshop	General Information Schedule (all times i Events Slides. exercises & re		
		2728.05.2025 9:00-16:30 (CEST), 10:00-17:30 (EEST) Amsterdam & online		Q&A of day 1 Expand all Back to top		
		Please ask your questions at the bottom of this document < click here	9	Go to bottom		

# New LUMI AI Guide

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

<u>F</u> ile <u>E</u> dit <u>V</u> iew Hi <u>s</u> tory <u>B</u> ookmarks <u>T</u> ools <u>H</u> elp			_ + 😣		
O Lumi-supercomputer/LUMI- × +					
$\leftarrow$ $\rightarrow$ C O A https://github.com/Lumi-	supercomputer/LUMI-AI-Guide	🗐 110% ☆ 🥸 Search	😽 🔮 🦉 ජු 💷 ≫ ≡		
E C Lumi-supercomputer / LUMI-AI-Guid	e	Q Type () to search	🗑 •   + • O II 🕛		
<> Code 💿 Issues 2 🏌 Pull requests 💿 Actions 🖽 Projects 🖽 Wiki 🕕 Security 3 🗠 Insights 🕸 Settings					
UMI-AI-Guide Public		🖈 Edit Pins 🗸 💿 Unwatch 8 👻	양 Fork 5 ★ ☆ Star 8 ★		
រេះ main 👻 រេះ 10 Branches 🛇 0 Tags	Q Go to file	t + <> Code -	About ®		
gregordecristoforo Merge pull request #35 from maciejjan/patch-1 🚥 31962fe · 4 days ago 🕚 186 Commits UMI AI Guide is designed to assist users in migrating their machine					
MLflow-visualization	fix broken links	4 days ago s	scale computing environments to the LUMI supercomputer.		
TensorBoard-visualization	fix broken links	4 days ago			
assets/images	move assets in image directory	5 days ago	ai deep-learning-tutorial lumi		

## Introduction to LUMI

Moving your AI training jobs to LUMI workshop 27.5.2025

B

LUMI

# LUMI is not one single computer

It behaves quite a bit different than your local computer

## LUMI is a very fast computer in Europe **LUMI**

- 8<sup>th</sup> 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 equivalent from each country
  - Offer email support, courses, workshops, ...
  - Responsible of software stack



#### LUMI is a cluster of individual computers **LUMI**

- 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

#### LUMI

#### Command line interface



#### 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.

#### **Pinned Apps**



## LUMI consists of different parts

LUMI

- 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

<u>L U M I</u>



#### 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

LUMI



aroup

- 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

<u>L U M I</u>

# AMD is not Nvidia

But the differences are quite small

## Our GPUs are confusing

#### LUMI



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 **LUM** 

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

#### LUMI

- 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:~\$

<u>L U M I</u>

#### 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



UΜ

### LUMI has three storage systems

#### LUMI



- 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></username>	Configuration files
Project persistent	/project/ <project></project>	Installations + final results
Project scratch	/scratch/ <project></project>	Input + Intermediate results
Project flash	/flash/ <project></project>	Input if high bandwidth is needed

## Lustre doesn't like many small files

#### LUMI



#### 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

#### LUMI



#### 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

#### LUMI



#### 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

LUMI



Use them well and you will get great results

LUMI

## Questions?