

AI Confidential Workshop

Tomáš Martinovič



SUPERPOČÍTAČE IT4INNOVATIONS

LUMI AI Factory

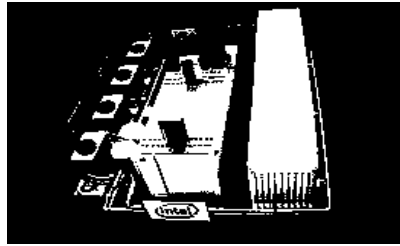


EuroHPC
Joint Undertaking

VŠB TECHNICKÁ
UNIVERZITA
OSTRAVA

IT4INNOVATIONS
NÁRODNÍ SUPERPOČÍTAČOVÉ
CENTRUM

Anselm
(0.1 PF)



July 2013

DGX-2
(2 PF^{AI})



March 2019

Karolina
(15.7 PF)



August 2021

LUMI-Q



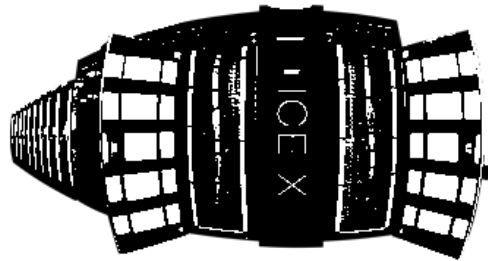
September 2025

2011

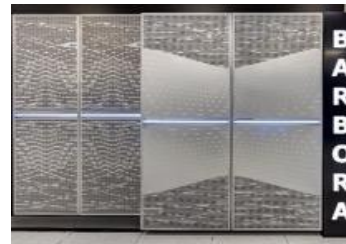
July 2015

October 2019

Autumn 2021



Salomon
(2 PF)



Barbora
(0.8 PF)



LUMI
(550+ PF, #9, in Kajaani, FI)



JAK NA SUPERPOČÍTANÍ ...



```
----- /apps/modules/bio -----
GROMACS/4.5.5-gompi-2020c-ORCA-5.0.1
GROMACS/2021.4-foss-cuda-2020b-PLUMED-2.7.3
GROMACS/2022-foss-cuda-2020b
GROMACS/2023-foss-2022a-CUDA-12.2.0-PLUMED-2.9.0-mpi
GROMACS/2023-foss-2022a-CUDA-12.2.0-PLUMED-2.9.0
GROMACS/2023.4-gompi-2020c-ORCA-5.0.1-double-precision
GROMACS/2024.1-foss-2023a
(O)

----- /apps/modules/chem -----
ABINIT/9.10.3-intel-2022a
ASE/3.22.1-pfaff-2023a
Amber/22.0-foss-2022a-Ambertools-23.3-CUDA-11.7.0
BEEF/0.1.1-intel-2020a
CP2K/9.1-foss-2022a
Critic2/1.1stable-intel-2020b
DFT-D4/3.6.0-intel-2022a-Python-3.10.8
DFT-D4/3.6.0-intel-2023b-Python-3.11.5
k
mpir
sockets
mpir

your public key has been saved in e
The key fingerprint is:
SHA256:fFp7L5cfGT0jmWRJ1TvVwHRJCzUR
The key's randomart image is:
+---[RSA 3072]----+
| ..o+=00l
| ..++*|
| ...o|
| .. .o |
| S.o. +=.|
| +o.o + *|
| .....B + |
```

```
...running on Rocky 8.X

Public Service Announcement: Apptainer on the Karolina cluster
Posted: (2024-05-10 10:23:47)

Apptainer is now a part of the operating system, you do not need to load the
module.

$ apptainer --version
apptainer version 1.3.1-1.el8

Last login: Mon May 27 10:05:28 2024 from 89.24.247.60
[mgolas@login2.karolina ~]$
```

- Vybrat si superpočítač
- Vyplnit žádost o zdroj
- Čekat na potvrzení
- Vytvořit účet
- Nastavit SSH klíč
- Znalost práce s Linux terminálem
- Přihlásit se na superpočítač
- Naučit se používat dávkový plánovač SLURM
- Naučit se používat moduly
- Naučit se pracovat s daty na superpočítači
- ... konečně něco spočítat

UMÍME TO LÍP

LUMI AI Factory

VŠB TECHNICKÁ
UNIVERZITA
OSTRAVA | IT4INNOVATIONS
NÁRODNÍ SUPERPOČÍTAČOVÉ
CENTRUM

- Let's have a nice web interface
- Alokace na jednom místě
- Napříč více výpočetními clustery
- Spuštění počítání jedním tlačítkem
- Mějte logy
- Správa dat
- Jednotný login
- ... a mnoho dalšího

The screenshot displays the 'AI TRAINING' workspace interface. It features a sidebar with navigation options: Home, Workspace, Data, Upload DataSets, Workflow DataSets, Workflows, and Executions. The main content area is titled 'Manage your projects, team members, and computing resources'. It includes a 'Projects' table with columns for SELECT, PROJECT NAME, SHORT NAME, DESCRIPTION, START DATE, END DATE, CONTACT EMAIL, and ACTIONS. Below this is a 'Team members' table with columns for NAME, ROLE, and ACTIONS. To the right, there is a 'Resources' section with a table listing resources like 'iRODS 5.1' and 'singlet' with their respective start and end dates and assigned status.

The screenshot shows a dashboard titled 'AI TRAINING Portal' with a 'Cluster Status' indicator. It features three main status cards: '5 Datasets' (with an 'upload + workflow' button), '2 Projects' (with an 'active' status), and '0 Access Requests' (with a 'none pending' status). Below these cards, there are sections for 'Recent Workflows' and 'Quick Actions'. The 'Recent Workflows' section lists 'TagWorkflow', 'Builder_Workflow', and 'Whisper_finetuning'. The 'Quick Actions' section includes 'New Workflow', 'View Executions', 'Upload Dataset', and 'Workspace'.

Založené na LEXIS Platformě:
docs.lexis.tech
opencode.it4i.eu/lexis-platform



The screenshot shows the 'AI TRAINING' workflow editor interface. It displays a workflow titled 'Whisper_finetuning' with a grid of nodes. The 'Input_data' node is connected to the 'Whisper_finetuning' node. A 'Properties' dialog box is open for the 'Input_data' node, showing fields for 'Node Name', 'Dataset id', 'File path', and 'Source Directory'. The 'Dataset id' is 'a8fba6c2-1870-11f1-a23a-d2a67a6ec3c1' and the 'File path' is 'Conversations 2025-10-10---its---tempZone...'. The 'Source Directory' is 'd8d1/a8fba6c2-1870-11f1-a23a-d2a67a6ec3c1'.

DATASET MANAGEMENT

AI TRAINING

Active Project: AI Confidential Workshop

Home > Data

Upload Datasets

9 datasets available · Updated just now

Search by title, topic, subtopic or keywords...

Dataset Name	Topic	Created
Efficientnet_data_inference	Computer Vision	02/04/2026
Efficientnet_model_inference	Computer Vision	02/04/2026
YOLO_model_inference	Computer Vision	02/04/2026
YOLO_data_finetuning	Computer Vision	01/04/2026

User: tomas.martinovic@vs...

AI TRAINING

Active Project: AI Confidential Workshop

Home > Workflow DataSets > YOLO_output_finetuning > Details > File List

YOLO_output_finetuning

Refresh

YOLO_output_finetuning > output > model_finetune > weights

File Name	Size
best.pt	5.96 MB
last.pt	5.96 MB

Drag 'n' drop some files here

User: tomas.martinovic@vs...

Copyright © 2026 IT4Innovations, All rights reserved. Powered By LEXIS

WORKFLOWS

The screenshot displays the LUMI AI Factory workflow editor. The interface includes a sidebar with navigation options: Home, Workspace, Data, Upload DataSets, Workflow DataSets, Workflows (selected), and Executions. The main workspace shows a workflow named 'Whisper_finetuning' on a grid. It consists of an 'Input_data' node (INPUT) connected to a 'Whisper_finetuning' node (COMPUTE). A properties panel is open for the 'Input_data' node, showing the following details:

- Node Name:** Input_data
- Dataset id:** a6fba6c2-1870-11f1-a23a-d2a67a6ec3c1
- File path:** /
- Select Dataset:** Conversations 2025-10-10---tts---tempZone...
- Source Directory:** ddi://a6fba6c2-1870-11f1-a23a-d2a67a6ec3c1/
- Source directory (automatically generated from dataset selection):** /input/user
- Target Directory *:** /input/user
- Target directory (on target machine or DDI):** /input/user



WORKFLOW MONITORING A LOGY

The screenshot displays a workflow monitoring interface for an HPC job named "Whisper_Finetuning".

Job Details:

- Selected Cluster: ATR-25-5
- Started: 2026-04-06 17:05
- View HPC Job Logs

INPUT DATASETS:

- Whisper Input Data
- Whisper Model

OUTPUT DATASETS:

- Whisper_Finetuning_r

EXECUTION STEPS:

- Check Status
- Xcom Storage (0.9s)
- Select Operator (9.1s)
- HEAppE Session (2.8s)
- HEAppE Prepare Job (3.4s)
- Input Transfers (1.5s)
- Input Transfers (1.4s)
- Input Transfers (1.0s)
- Input Transfers (1.0s)

Workflow Diagram:

```
graph TD; start --> Whisper_Finetuning; Whisper_Finetuning --> finish;
```

Log Output:

```
0% | 0/3 [00:00<?, ?it/s] 33% | 1/3 [00:38-01:16, 38.14s/it]
0% | 0/1 [00:00<?, ?it/s] [A
[A 33% | 1/3 [00:41-01:16, 38.14s/it]
100% | 1/1 [00:00-00:00, 41.39it/s] [A
[A 67% | 2/3 [00:56-00:26, 26.22s/it]
0% | 0/1 [00:00<?, ?it/s] [A
[A 67% | 2/3 [01:00-00:26, 26.22s/it]
100% | 1/1 [00:00-00:00, 113.66it/s] [A
[A100% | 3/3 [01:16-00:00, 23.73s/it]
0% | 0/1 [00:00<?, ?it/s] [A
[A100% | 3/3 [01:21-00:00, 23.73s/it]
100% | 1/1 [00:00-00:00, 120.32it/s] [A
[A
100% | 3/3 [01:21-00:00,
23.73s/it]100% | 3/3 [01:21-00:00, 27.07s/it]
Writing model shards: 0% | 0/1 [00:00<?, ?it/s]Writing model shards: 100% | 1/1 [00:01-00:00,
1.23s/it]Writing model shards: 100% | 1/1 [00:01-00:00, 1.23s/it]
```



Dedikované komponenty

- Nové kontejnery
- Šablony



MLflow integrace

- Metriky modelu
- Monitorování využití systému



Integrace
managementu
inference služby

PŘÍSTUP K SUPERPOČÍTAČŮM

LUMI AI Factory

VŠB TECHNICKÁ
UNIVERZITA
OSTRAVA

IT4INNOVATIONS
NÁRODNÍ SUPERPOČÍTAČOVÉ
CENTRUM



Podpora
v rámci evropských
a národních
projektů
EDIH,
AI Factories

Komerční
pronájem

Národní, výzkumné
a jiné otevřené
výzvy

Obrazové modely

- Retinopatie
- Analýza částic
- Analýza buněk

Dálkový průzkum Země

- Detekce změn

Analýza časových řad

- Predikce poptávky
- Predikce zpoždění
- Spotřeba energie

Agentní systémy

- Podpora rozhodování - požáry
- AI bezpečnost pomocí agentních interakcí

Zajištění kvality

- Ansámby modelů

Jednoduchý přístup k superpočítačům

- Bezpečný transfer velkých dat
- Bezpečné spouštění výpočtů pomocí API
- Webové aplikace na míru



NENÍ TO JEN O TRÉNOVÁNÍ MODELŮ...

Trénink

- Předem dané množství dat – dávková úloha
- Potřeba efektivně přenést data k výpočetní infrastruktuře – GPUs
- Poskytnout výsledky uživateli

Inference

- Dávková úloha nebo požadavky zasílané uživateli
- Nutnost specifické orchestrace – provoz služeb čekajících na požadavky není efektivní v dávkových plánovačích

HPC Cluster



HyperQueue



CPU Partition

GPU Partition

Děkuji za pozornost

Tomáš Martinovič

tomas.martinovic@vsb.cz

IT4Innovations National Supercomputing Center

VSB – Technical University of Ostrava

Studentská 6231/1B

708 00 Ostrava-Poruba, Czech Republic

www.it4i.eu



EuroHPC
Joint Undertaking



Spolufinancováno
Evropskou unií

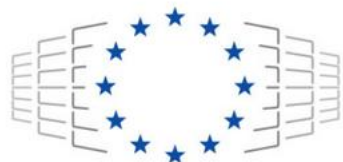
MŠMT
MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY

LUMI AI Factory

KONTAKTUJTE NÁS

EXPERTI LUMI AI FACTORY
ai-factory@lumi-supercomputer.eu

ČESKÝ KONTAKT LUMI AI FACTORY
ai-factory@it4i.cz



EuroHPC
Joint Undertaking



Spolufinancováno
Evropskou unií

MŠMT
MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY