



aix*celerate*

Machine Learning (ML) on NVIDIA GPUs

Tuning Workshop 2024

December 9th-11th, 2024 | Aachen

Dr. Sandra Wienke, Dr. Jannis Klinkenberg

Who We Are? – Organizations

- **NHR4CES**: NHR for Computational Engineering Science
 - NHR = National High Performance Computing: alliance of 9 German universities to provide resources and competencies to scientists at all German universities
 - RWTH + TU Darmstadt = NHR4CES: focus on computational engineering sciences
- **IT Center** of RWTH Aachen University, HPC team
 - Provides consulting and training services for HPC users
 - Conducts own HPC research and lecturing (Chair for High Performance Computing)
- **HPC.NRW**: The North Rhine-Westphalian Competence Network for HPC
 - Connection of expertise of the large HPC centers (Tier 2) with the consulting services of smaller centers (Tier 3)
- **EDIH Rheinland**: European Digital Innovation Hub
 - Advice on Digitalization, Artificial Intelligence and High Performance Computing
 - Focus on small and medium-sized enterprises, Start-ups, Spin-offs and public facilities
- **WestAI**: AI services from NRW for Germany
 - Consulting services and access to hardware resources on Nvidia H100 for academia as well as small and medium-sized enterprises*



* constraint to research and exploration activities

Who We Are? – People

Speakers, Supporters, HPC & ML Experts

IT Center of RWTH Aachen University

- Tim Cramer
- Marc-André Hermanns
- Jannis Klinkenberg
- Jan Kraus
- Radita Liem
- Fabian Orland
- Felix Tomski
- Dominik Viehhauser
- Christian Wassermann
- Sandra Wienke

NVIDIA

- Fabian Berressem



Overview

Topics

- Analyzing Performance of ML Codes (Dec 9th)
- Scaling ML Codes across Multiple GPUs/ Nodes (Dec 10th)
- Handling Datasets of ML Codes (Dec 11th)

Note

- This is **not** an introductory course on ML
- Focus is on using GPUs of RWTH's HPC cluster "CLAIX"

Organization

- Presentations
 - Hybrid format
 - Open for everyone
- Code & Tuning Activities → BYO code
 - On-premise only
 - Selection process due to limited seats

Overview

Questions – any time!

- In-person or
- “Raise hand” in Zoom & unmute your mic, or “Q&A” feature in Zoom

Course material

- Will be continually added to website
- Website: <https://blog.rwth-aachen.de/itc-events/en/event/aixcelerate-2024/>

Sponsors for the catering: NEC & NVIDIA

 **Orchestrating a brighter world**

NEC



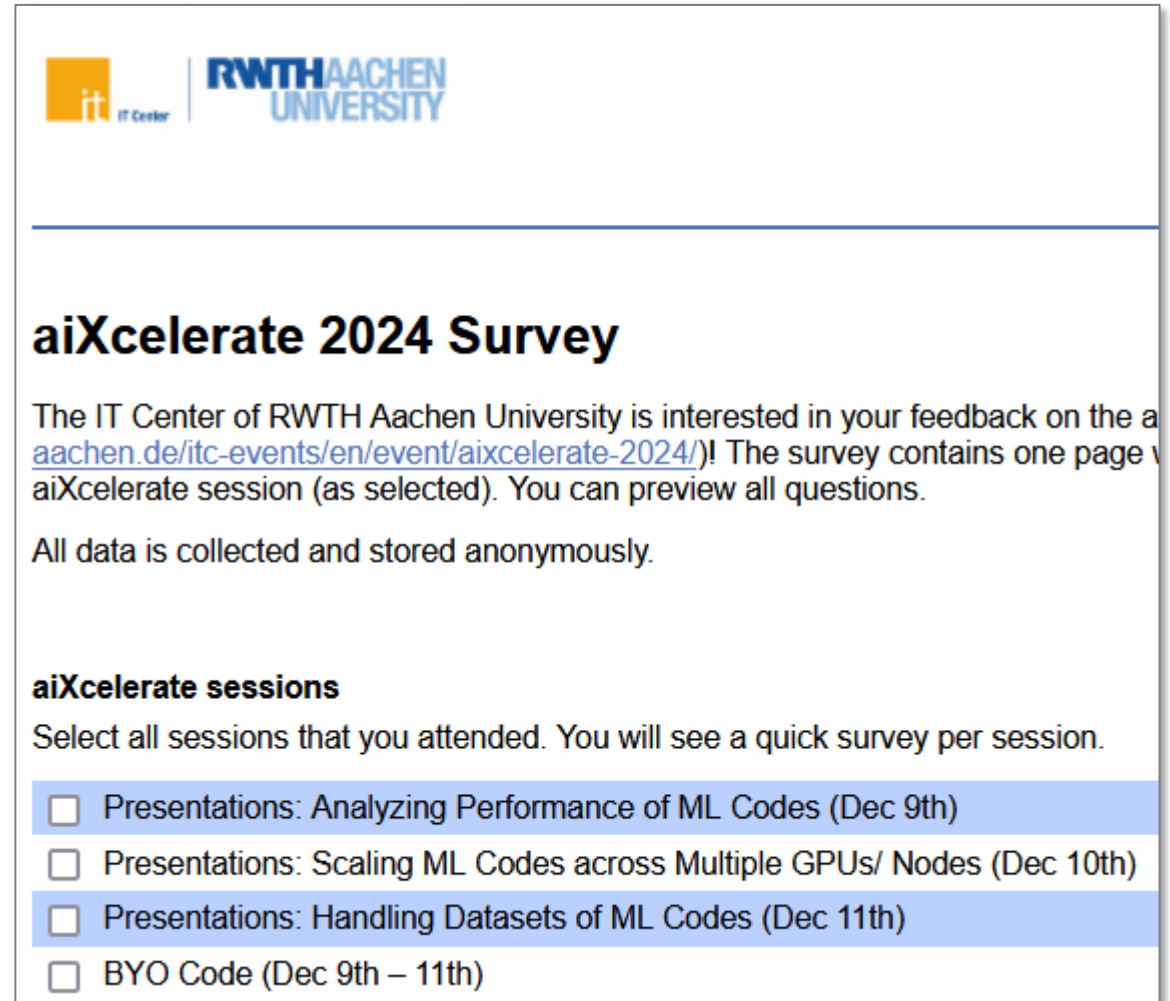
Evaluation – Feedback Wanted!



aiXcelerate Evaluation Form

(valid until December 23rd, 2024)

https://s2survey.net/aixcelerate_2024



aiXcelerate 2024 Survey

The IT Center of RWTH Aachen University is interested in your feedback on the [aiXcelerate 2024](https://www.rwth-aachen.de/itc-events/en/event/aixcelerate-2024/)! The survey contains one page per aiXcelerate session (as selected). You can preview all questions.

All data is collected and stored anonymously.

aiXcelerate sessions

Select all sessions that you attended. You will see a quick survey per session.

- Presentations: Analyzing Performance of ML Codes (Dec 9th)
- Presentations: Scaling ML Codes across Multiple GPUs/ Nodes (Dec 10th)
- Presentations: Handling Datasets of ML Codes (Dec 11th)
- BYO Code (Dec 9th – 11th)

Agenda for Today (Dec 10th, 2024)

Time	Topic	Speaker
Scaling ML Codes across Multiple GPUs/ Nodes		
9:00 – 9:05	Welcome	Sandra Wienke (RWTH)
9:05 – 9:15	Multi-GPU Setup on CLAIX	Jannis Klinkenberg (RWTH)
9:15 – 10:00	Using 'PyTorch Distributed'	Fabian Berressem (NVIDIA)
10:00 – 10:30	Using Tensorflow and Horovod	Jannis Klinkenberg (RWTH)
10:30 – 11:00	Coffee Break	
Code Activies (BYO Code)		
11:00 – 12:30	Code Activities	
12:30 – 13:30	Lunch Break (on your own)	
13:30 – 15:00	Code Activites	
15:00 – 15:30	Coffee Break	
15:30 – 17:30	Code Activities	