DAY 1: INTRODUCTION

Scalable Deep Learning

2021-02-01 | Stefan Kesselheim | Helmholtz AI @ JSC

COURSE COMMUNICATION

- Big Blue Button
  - Link: `https://webconf.fz-juelich.de/b/kes-ogt-3cv-yew`
  - Please mute yourself, when not speaking.
  - Please do unmute yourself and contribute!
- Slack
  - `#general` for general stuff
  - `#day-n-[lectures|tutorials]` for questions and comments
- Coffee break: `https://www.wonder.me/r?id=c18a9b22-3e52-41f6-a860-63a52b077173`
- Course Material
  - Course Material:
    `https://gitlab.version.fz-juelich.de/MLDL_FZJ/juhaicu/jsc_public/sharedspace/teaching/intro_scalable_dl_2021/course-material`
  - Course environment (usually no need to clone):
    `https://gitlab.version.fz-juelich.de/MLDL_FZJ/juhaicu/jsc_public/sharedspace/teaching/intro_scalable_dl_2021/course2021_working_environment`

COURSE GOALS

Machine Learning

+

Supercomputing

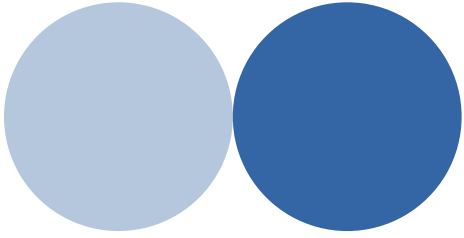
COURSE GOALS

Machine Learning

+

Supercomputing

YOU?



Strong Machine Learning Background

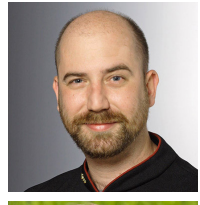
Strong Supercomputing Background

COURSE TEAM

Jenia Jitsev

Roshni Kamath

Mehdi Cherti



Alexandre Strube

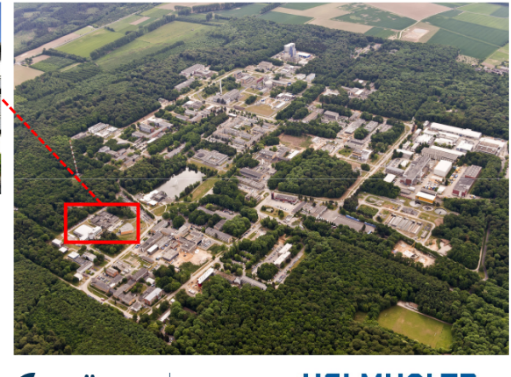
Jan Ebert

Stefan Kesselheim

WELCOME TO JSC

Welcome @ Juelich Supercomputing Centre (JSC) of Forschungszentrum Juelich

- Selected Facts
  - One of EU largest inter-disciplinary research centres (~5000 employees)
- Special expertise
  - Physics
  - Material sciences
  - Nanotechnology
  - Neuroscience & Medicine
  - Information technology (HPC, Big Data, Quantum Computing, Clouds, etc.)
  - Artificial Intelligence (AI)



JÜLICH Forschungszentrum | JÜLICH SUPERCOMPUTING CENTRE

[25] Juelich Supercomputing Centre

HELMHOLTZ
RESEARCH FOR GRAND CHALLENGES

[24] Helmholtz Association Web Page

WELCOME TO FZJ

1956

FOUNDATION

on 12 December

Shareholders

90 % Federal Republic of Germany

10 % North Rhine-Westphalia

11

INSTITUTES

2 project management organizations

609.3

million euros **REVENUE total**

(40 % external funding)

5,914

EMPLOYEES

2,165scientists

536 doctoral researchers

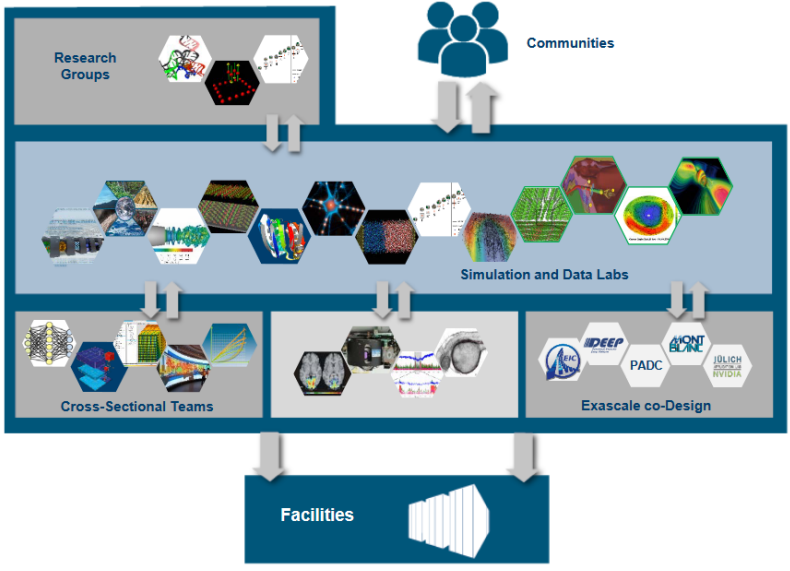
323 trainees and students on placement

867

VISITING SCIENTISTS

from **65** countries

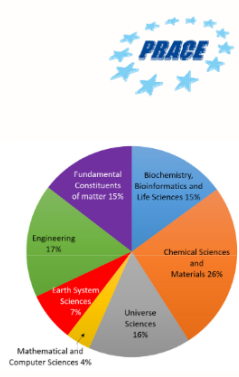
SUPPORT AND RESEARCH LANDSCAPE AT JSC

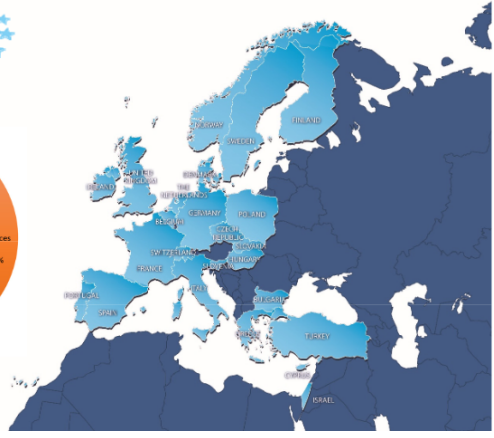


PRACE

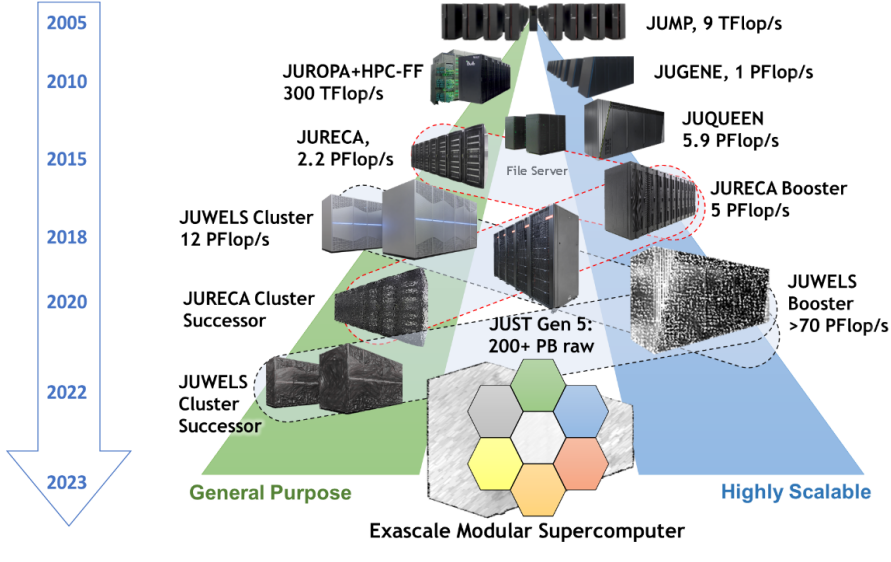
Partnership For Advanced Computing in Europe

- Basic Facts
  - HPC-driven infrastructure
  - An international not-for-profit association under Belgien law (with its seat in Brussels)
  - Has 25 members and 2 observers
  - Governed by the PRACE Council in which each member has a seat
  - Daily management of the association is delegated to the Board of Directors





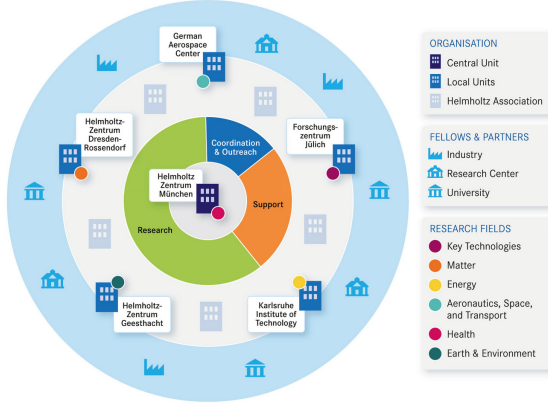
(DUAL) HARDWARE STRATEGYAT JSC



HELMHOLTZ AI

- Helmholtz AI Central Unit
  - Helmholtz Zentrum Muenchen (HMGU)
- Other Helmholtz AI Local Units
  - Energy: Karlsruhe Institute of Technology (KIT)
  - Earth & Environment: Geesthacht Centre for Material & Costal Research (HZG)
  - Aernoautics, Space & Transport: German Aerospace Centre (DLR)
  - Matter: Helmholtz Centre Dresden-Rossendorf (HZDR)
- Information: Helmholtz AI Local @ Forschungszentrum Juelich
  - Young Investigator Group @ INM-1
  - High Level Support Team (HLST) @ JSC

Structure



DEEP-EST


High Energy Physics

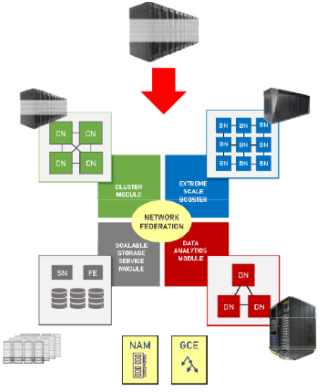

Earth Science

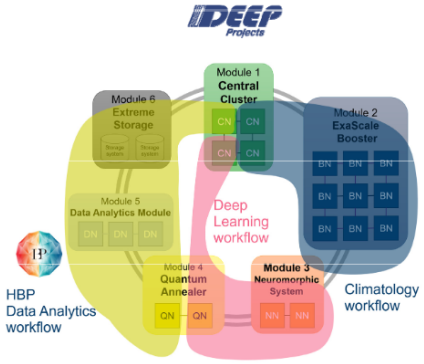

Space Weather


Molecular Dynamics


Neuroscience


Radio Astronomy





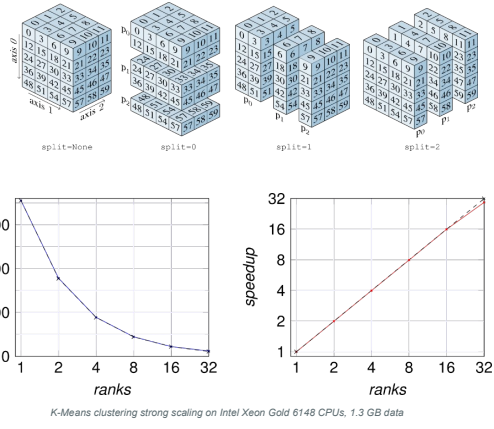
- **The modular supercomputing architecture (MSA) enables a flexible HPC system design co-designed by the need of different application workloads**

[12] DEEP Projects Web Page

HEAT

Helmholtz Analytics Toolkit

- **Python** framework for **data-intensive** computing
  - **n-dimensional** numerical **tensors**
  - Low-level operations and **high-level algorithms**
- **Transparent** data **scaling** via MPI
  - Start prototyping on **laptop**
  - Port without changes to **HPC clusters**
- Different **computational backends**
  - Traditional **CPU** computation
  - Advanced **GPU** acceleration
- Up to **three orders of magnitudes faster** than comparable frameworks (e.g. dask)



K-Means clustering strong scaling on Intel Xeon Gold 6148 CPUs, 1.3 GB data

Try it out: https://github.com/helmholtz-analytics/heat/

TIMETABLE

Day	Time	Title
1	9:00-9:30	Introduction
	9:30-10:00	Lecture 1.1: Intro to supercomputers
	10:00-11:00	Tutorial 1.1: First Steps on the Supercomputer
	11:00-11:30	Coffe Break
	11:30-12:00	Lecture 1.2: Supercomputer architecture and MPI primer
	12:00-13:00	Tutorial 1.2: Hello MPI World
2	9:30-10:00	Lecture 2.1: Motivation for Large Scale Deep Learning and Deep Learning Basics Recap
	10:00-11:00	Tutorial 2.1: Deep Learning Basics Recap
	11:00-11:30	Coffe Break
	11:30-12:00	Lecture 2.2: Distributed Training Schemes and Data Parallelism with Horovod
	12:00-13:00	Tutorial 2.2: A first parallelization with Horovod
3	9:30-10:00	Lecture 3.1: Large Datasets and Scaling Distributed Training
	10:00-11:00	Tutorial 3.1: Distributed Training with ImageNet and Scaling Basics
	11:00-11:30	Coffe Break
	11:30-12:00	Lecture 3.2: Is my code fast? Performance analysis
	12:00-13:00	Tutorial 3.2: Tools for training performance analysis
4	9:30-10:00	Lecture 4.1: Combating Accuracy Loss
	10:00-11:00	Tutorial 4.1 Part I : Combating Accuracy Loss, Basics
	11:00-11:30	Coffe Break
	11:30-12:30	Tutorial 4.1 Part II : Combating Accuracy Loss, Advanced
	12:30-13:00	Lecture 4.2 : Outlook beyond the basics
5	9:30-10:00	Lecture 5.1: Generative models, GAN basics
	10:00-11:00	Tutorial 5.1: Parallelizing a basic GAN: DCGAN architecture
	11:00-11:30	Coffe Break
	11:30-12:00	Lecture 5.2: Advanced GANs
	12:00-13:00	Tutorial 5.2: Parallelizing an advanced GAN: StyleGAN2 architecture
	13:00-13:15	Concluding remarks