Python Machine Learning Example Projects using the Modular Supercomputing Architecture (MSA)

PROF. DR. – ING. MORRIS RIEDEL, UNIVERSITY OF ICELAND / JUELICH SUPERCOMPUTING CENTRE (JSC)

24TH FEBRUARY, 12TH JOINT LABORATORY FOR FOR EXTREME-SCALE COMPUTING (JLESC) WORKSHOP, ONLINE

@ProfDrMorrisRiedel @Morris Riedel @MorrisRiedel @MorrisRiedel https://www.youtube.com/channel/UCWC4VKHmL4NZgFfKoHtANKg

Joint Laboratory for Extreme-Scale Computing

EuroHPC Joint Undertaking

EOSC NORDIC

RAISE Center of Excellence

ADMIRE

UNIVERSITY OF ICELAND
SCHOOL OF ENGINEERING AND NATURAL SCIENCES
FACULTY OF INDUSTRIAL ENGINEERING, MECHANICAL ENGINEERING AND COMPUTER SCIENCE

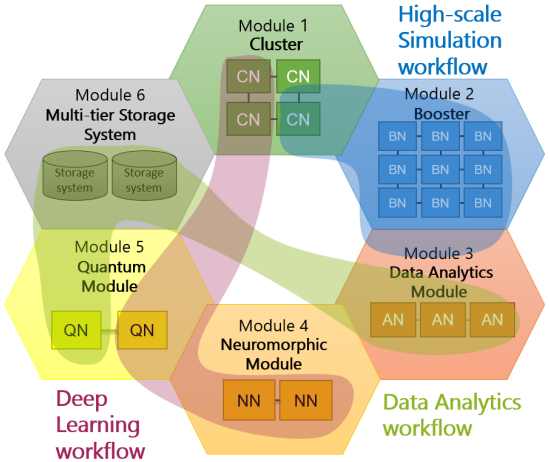
EURO

HELMHOLTZAI | ARTIFICIAL INTELLIGENCE COOPERATION UNIT

DEEP Projects

JÜLICH Forschungszentrum | JÜLICH SUPERCOMPUTING CENTRE

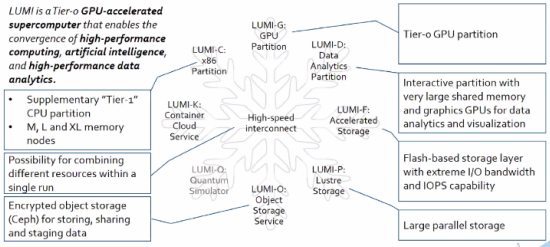
Modular Supercomputing Architecture (MSA) & Exascale Roadmap @ EU



EuroHPC
Joint Undertaking

[5] EuroHPC Joint Undertaking

LUMI system architecture



[6] LUMI Supercomputer

2023

1 EF

2020

85 PF

python

python

[1] DEEP Series of Projects Web Page

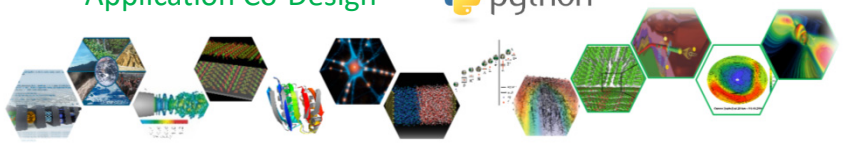
DEEP Projects

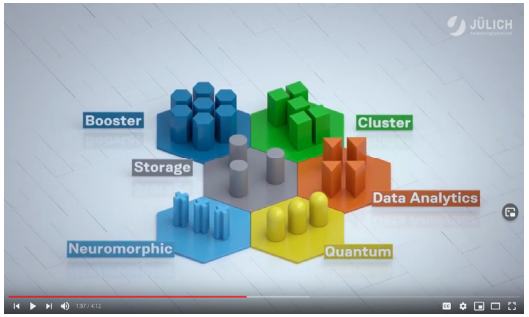
2018

12 PF

Application Co-Design

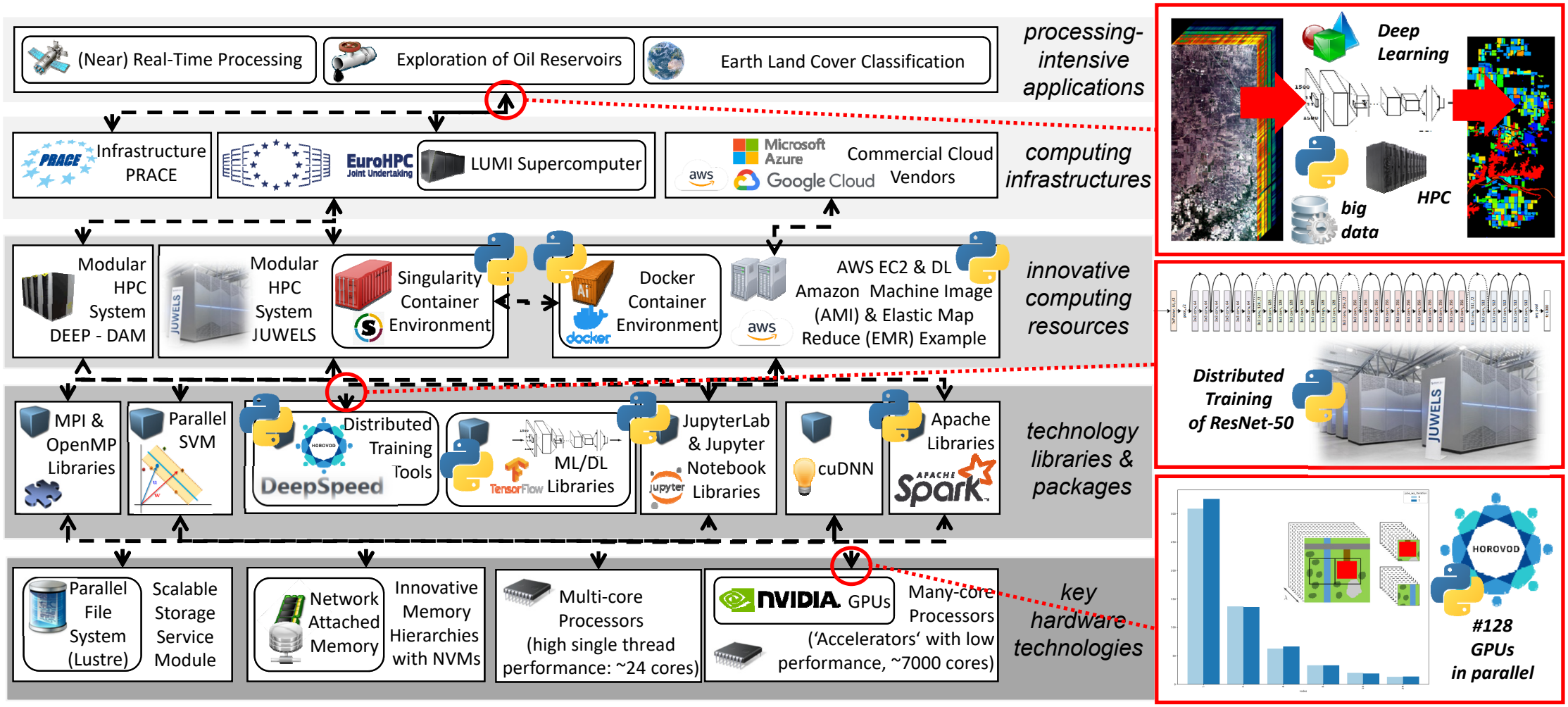
python



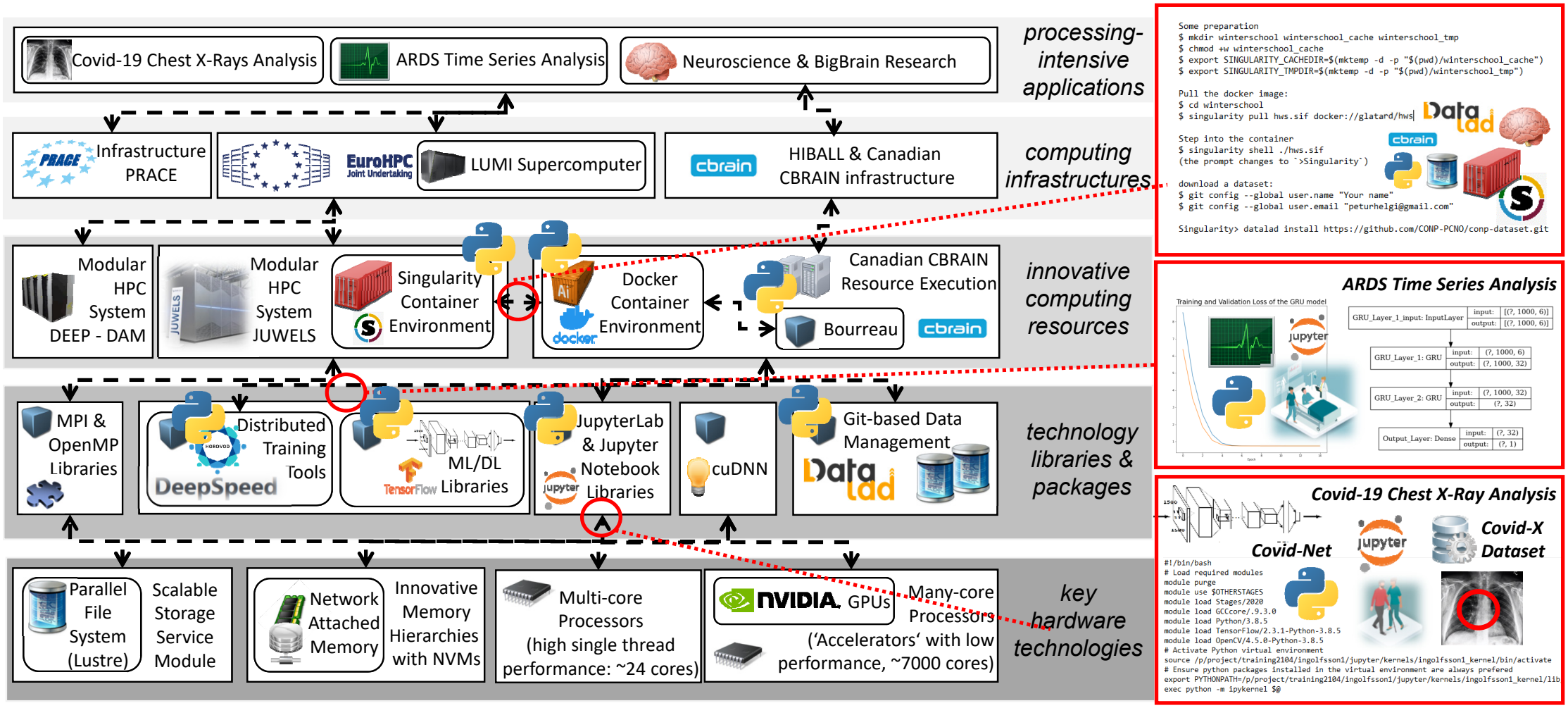


[2] YouTube, 'flexible and energy-efficient supercomputer: JUWELS is faster than 300 000 modern PCs

Python Machine Learning Example – Remote Sensing AI & HPC Applications



Python Machine Learning Example – Health & Medical AI & HPC Applications



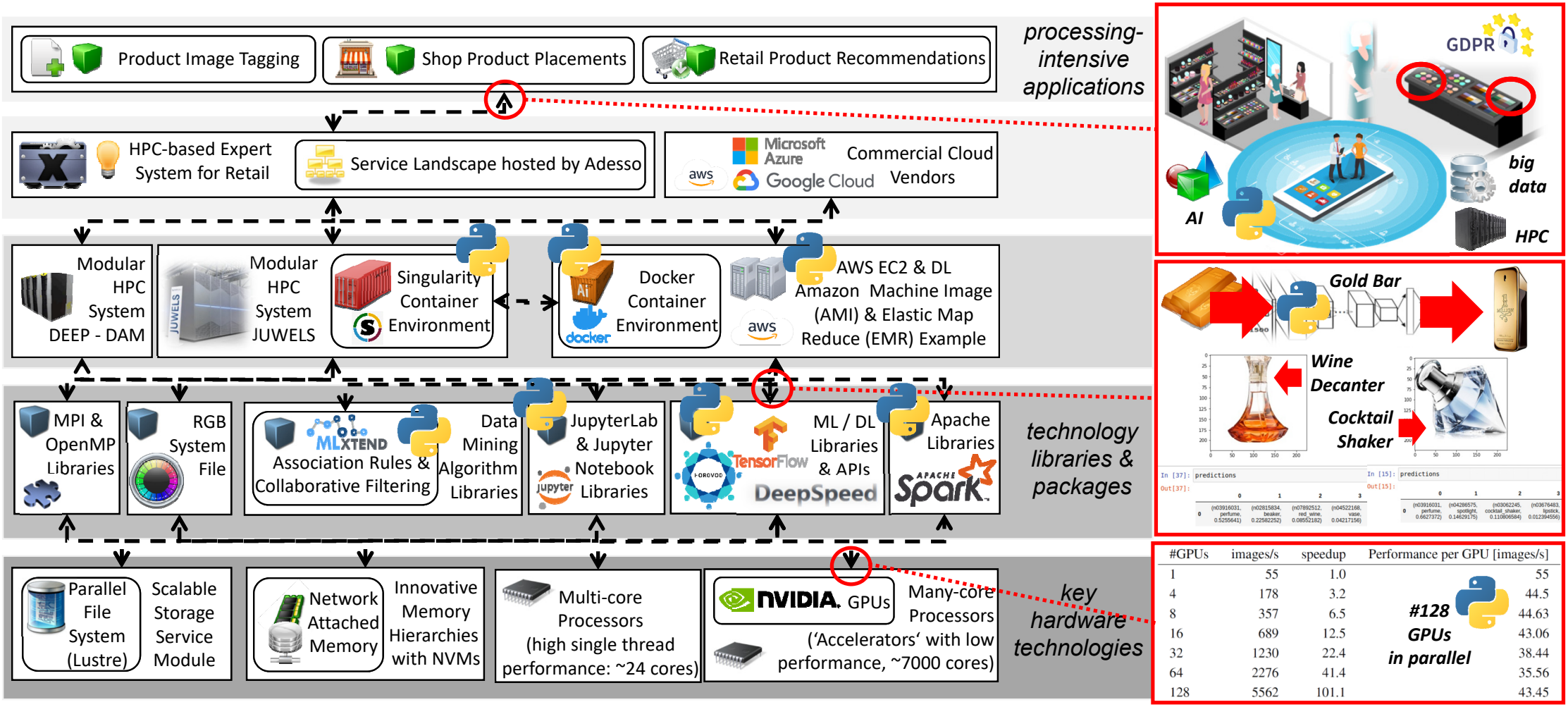

```
Some preparation
$ mkdir winterschool winterschool_cache winterschool_tmp
$ chmod +w winterschool_cache
$ export SINGULARITY_CACHEDIR=$(mktemp -d -p "$(pwd)/winterschool_cache")
$ export SINGULARITY_TMPDIR=$(mktemp -d -p "$(pwd)/winterschool_tmp")

Pull the docker image:
$ cd winterschool
$ singularity pull hws.sif docker://glatard/hws

Step into the container
$ singularity shell ./hws.sif
(the prompt changes to `>Singularity`)

download a dataset:
$ git config --global user.name "Your name"
$ git config --global user.email "peturhelgi@gmail.com"

Singularity> datalad install https://github.com/CONP-PCNO/conp-dataset.git
```

ARDS Time Series Analysis

Covid-19 Chest X-Ray Analysis


```
#!/bin/bash
# Load required modules
module purge
module use $OTHERSTAGES
module load Stages/2020
module load GCCcore/.9.3.0
module load Python/3.8.5
module load TensorFlow/2.3.1-Python-3.8.5
module load OpenCV/4.5.0-Python-3.8.5
# Activate Python virtual environment
source /p/project/training2104/ingolfsson1/jupyter/kernels/ingolfsson1_kernel/bin/activate
# Ensure python packages installed in the virtual environment are always prefered
export PYTHONPATH=/p/project/training2104/ingolfsson1/jupyter/kernels/ingolfsson1_kernel/lib
exec python -m ipykernel $@
```

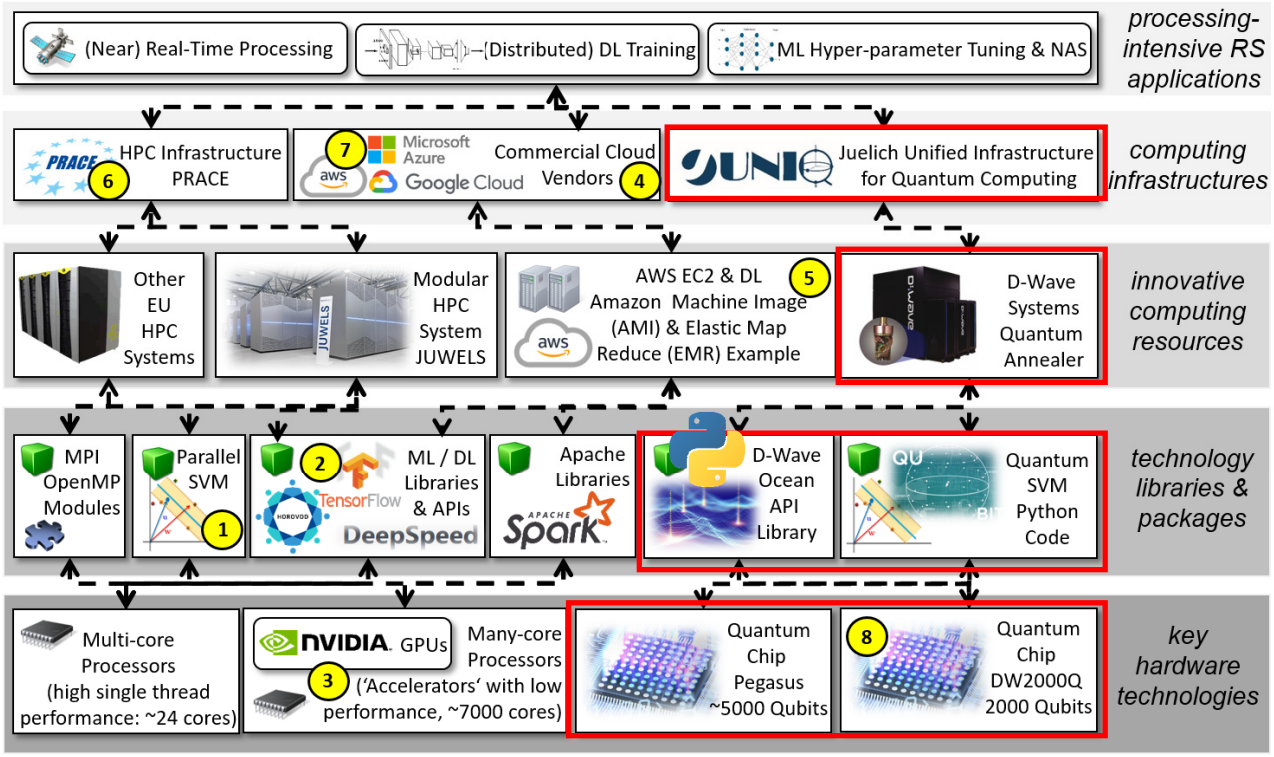
Python Machine Learning Example – Retail AI & HPC Applications



#GPUs	images/s	speedup	Performance per GPU [images/s]
1	55	1.0	55
4	178	3.2	44.5
8	357	6.5	44.63
16	689	12.5	43.06
32	1230	22.4	38.44
64	2276	41.4	35.56
128	5562	101.1	43.45

#128 GPUs in parallel

Python Machine Learning Example – Quantum Module with D-Wave Systems



1. Parallel ML implementations still rare (MPI/OpenMP)
2. Open source tools good, but all need to fit in versions
3. Using very many GPUs beyond NVlink could be tricky
4. Look & feel of CC vendor ML services differ significantly
5. Costs of GPUs of CC vendors (e.g., EC2) tough, 24$/hour
6. GPU hours are free, but requires time grant proposal
7. Free GPUs in Google Colab vary in the available types
8. Works not yet with multi-class problems & large data

Legend: (N) Highlighted Challenges & Experiences

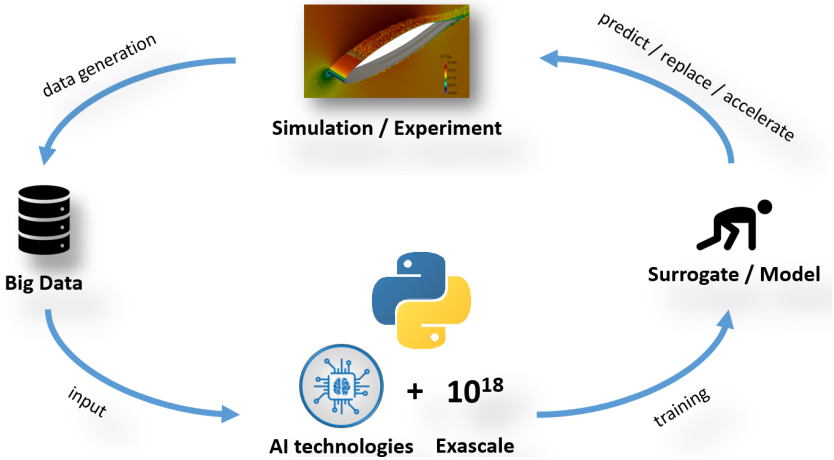
```
from quantum_SVM import *
import numpy as np
from utils import *
from sklearn.model_selection import KFold
from sklearn import preprocessing

# Write the data
experiments=1
slice=40 # Number of samples to use for the training
fold=int(len(X_train)/40)

print(fold)

for i in range(0,experiments):
    cv = KFold(n_splits=fold, random_state=i, shuffle=True)
    count=0
    for test_index, train_index in cv.split(X_train):
        #print("Train Index: ", len(train_index), "\n")

        X_train_slice, y_train_slice = X_train[train_index], Y_train[train_index]
        X_train_slice = preprocessing.scale(X_train_slice)

        X_test_slice, y_test_slice = X_train[test_index], Y_train[test_index]
        X_test_slice = preprocessing.scale(X_test_slice)
```


[8] Approaching Remote Sensing Image Classification with Ensembles of SVMs on the D-Wave Quantum Annealer, G. Cavallaro & M. Riedel et al.

(ensembles due to small datasets compared to full datasets on CPUs/GPUs)

[8] Quantum SVM, D. Willsch et al. *[7] M. Riedel, UTMessan 2020 YouTube Video*

Python Machine Learning Example – RAISE: Intertwined HPC Simulations & AI



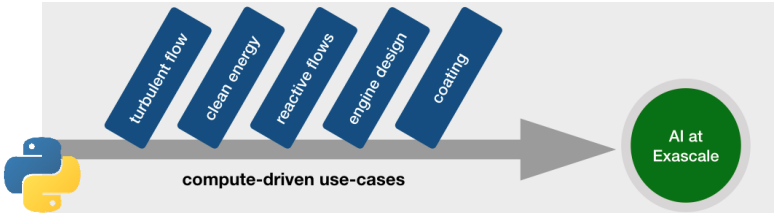
[4] CoE RAISE Web Page

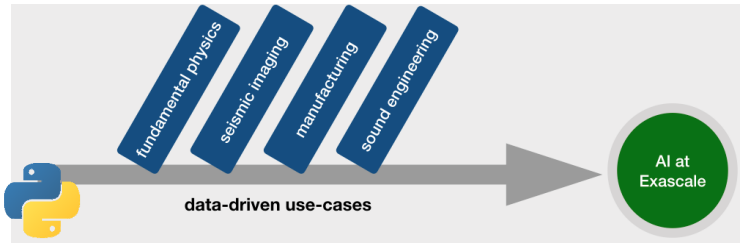
[3] Simulation Figure

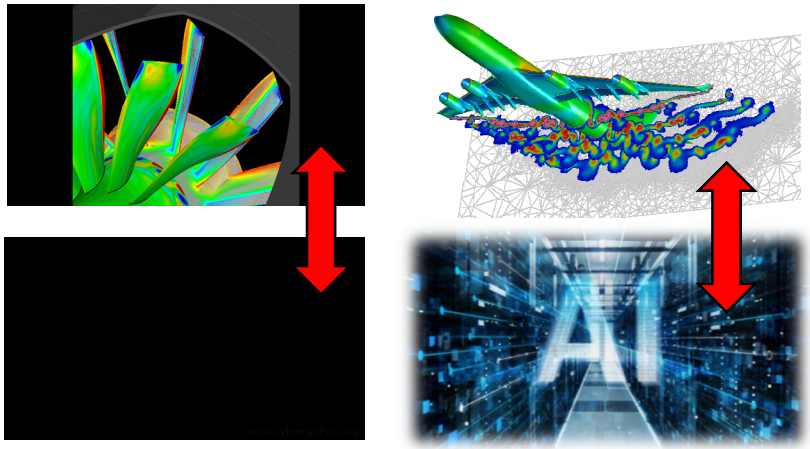
NEW

RAISE Center of Excellence

SAFRAN · JÜLICH Forschungszentrum · CERFACS · Atos · Delphi Consortium · 1862 RĪGAS TEHNISKĀ UNIVERSITĀTE · CERN · UNIVERSITY OF ICELAND · FLANDERS MAKE · PARTEC CLUSTER COMPETENCE CENTER The Modular Supercomputing Company · BSC Barcelona Supercomputing Center Centro Nacional de Supercomputación · RWTH AACHEN UNIVERSITY · The Cyprus Institute Research • Technology • Innovation







[8] Neural Network 3D Simulation

Open PhD Position Available in EU Project RAISE @ Iceland

Information

The PhD position is funded by the EU project Center of Excellence "Research on AI- and Simulation-Based Engineering at Exascale" (CoE RAISE). This project will be the excellent enabler for the advancement of European multi-physics and/or multi-scale applications on industrial and academic level and a driver for novel intertwined AI and HPC technologies.

Supervisor: Prof. Morris Riedel (University of Iceland)
Co-Supervisors: Dr. Gabriele Cavallaro (Jülich Supercomputing Centre) and Prof. Magnús Örn Úlfarsson (University of Iceland)

Starting date: January 2021
(Due to the current corona pandemic, the first work period can be conducted remotely)

Location: Reykjavík (Iceland). You will be employed at the University of Iceland. A research stay at the Jülich Supercomputing Centre (Forschungszentrum Jülich, Germany) is envisaged for a minimum period of time of 6 months. To obtain your PhD degree at the University of Iceland you will have to acquire 30 ECTS from courses and seminars. Your working hours will be not monitored and working from home will be largely permitted.

Goal: pioneer the research of advanced deep transfer learning methods in the context of complex learning scenarios in applications from remote sensing. The priority will be put on the investigation of the transferability capacity of Deep Learning (DL) models with meta-learning and Neural Architecture Search methods.

Research Group: be part of our joint research group "High Productivity Data Processing" at University of Iceland and Jülich Supercomputing Centre. The group is highly active in developing parallel and scalable machine (deep) learning algorithms for remote sensing data processing and many other types of applications (i.e., medical research and retail sectors).

Working Environment: Direct access to high performance multi-GPU systems equipped with the state-of-the-art of DL frameworks (TensorFlow, pyTorch, Chainer, Horovod, DeepSpeed). There is also the possibility to access innovative quantum computing systems.

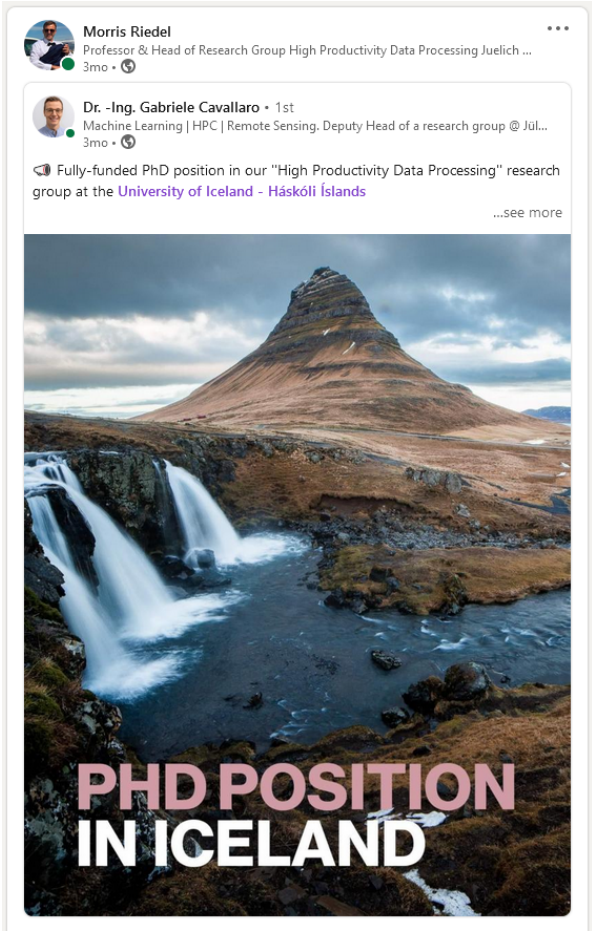
Other information: You will have the possibility to participate in international top conferences in the field of machine learning, HPC and remote sensing. You will be put in contact with several international partners for initiating research collaborations that match the topic of the PhD.

Background education: MSc degree in computer science or computer engineering. Level of English >= B2.

Required knowledge and experience: deep learning (Convolutional Neural Networks and/or Transformers) and Python programming (TensorFlow and/or pyTorch). Experience with parallel programming (OpenMP and MPI), High Performance Computing (HPC) and remote sensing data processing are a substantial plus.

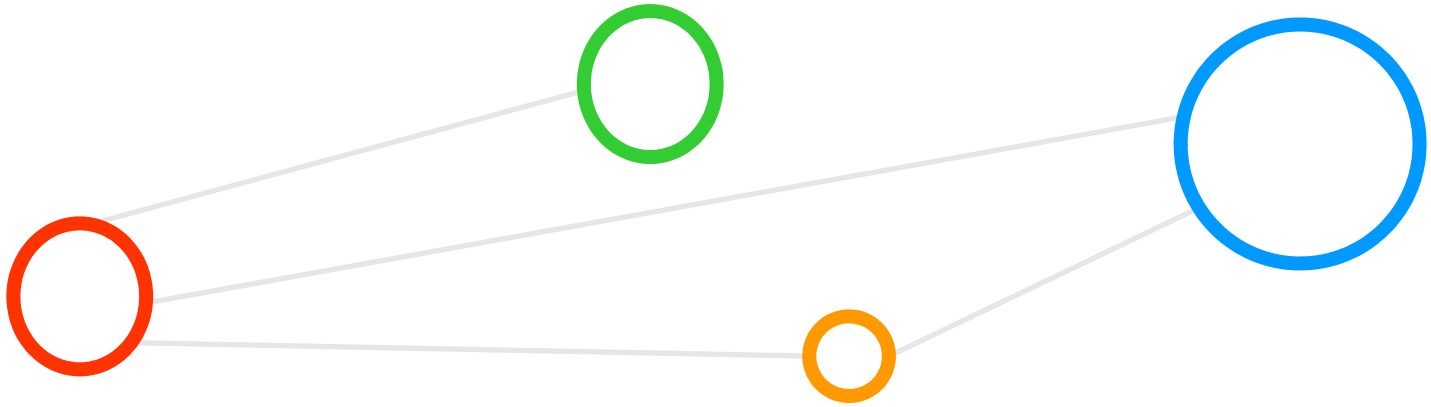
Apply: Send your CV, a cover letter and the transcripts of records of your bachelor and master to Gabriele Cavallaro: g.cavallaro@fz-juelich.de.

Apply now



[10] Open PhD Position, RAISE EC Project @ Iceland

Lecture Bibliography



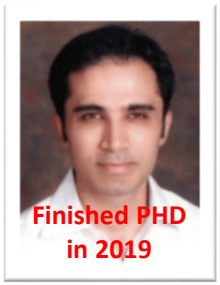
Lecture Bibliography

- [1] DEEP Series of Projects Web page, Online: http://www.deep-projects.eu/
- [2] YouTube Video, 'flexible and energy-efficient supercomputer: JUWELS is faster than 300 000 modern PCs' Online: https://www.youtube.com/watch?v=t5kNxPT5rSY&list=PLCer2BlxxQ2zToC6SRVlfwj0MO1-xIi6I
- [3] Copyright Institute of Aerodynamics and Chair of Fluid Mechanics, RWTH Aachen University, Online: https://www.aia.rwth-aachen.de
- [4] CoE RAISE Web page, Online: http://www.coe-raise.eu
- [5] EuroHPC Joint Undertaking Web page, Online: https://eurohpc-ju.europa.eu/
- [6] LUMI EuroHPC Supercomputer hosted at CSC Finland, Online: https://www.lumi-supercomputer.eu/
- [7] YouTube, Morris Riedel, UTmessan 2020 - Demystifying Quantum Computing, Online: https://www.youtube.com/watch?v=EQGshhspn9A
- [8] D. Willsch, M. Willsch, H. De Raedt, K. Michielsen, 'Support Vector Machines on the D-Wave Quantum Annealer', Online: https://www.sciencedirect.com/science/article/pii/S001046551930342X951733
- [9] Cavallaro, G., Willsch, D., Willsch, M., Michielsen, K., Riedel, M.: APPROACHING REMOTE SENSING IMAGE CLASSIFICATION WITH ENSEMBLES OF SUPPORT VECTOR MACHINES ON THE D-WAVE QUANTUM ANNEALER, in conference proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2020), September 26 – October 2nd, 2020, Virtual Conference, Hawai, USA, to appear, Online: https://igarss2020.org/Papers/ViewPapers.asp?PaperNum=1416
- [10] Open PhD Position for the RAISE EU project @ Iceland, Online: https://www.gabriele-cavallaro.com/news/fully-funded-phd-position

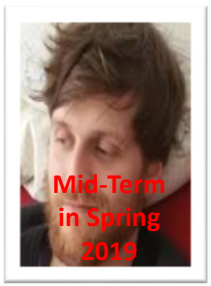
Acknowledgements – High Productivity Data Processing Research Group

PD Dr.
G. Cavallaro

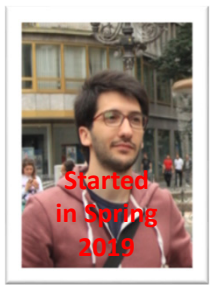
Senior PhD
Student
A.S. Memon



PD Dr.
M.S. Memon



PhD Student
E. Erlingsson

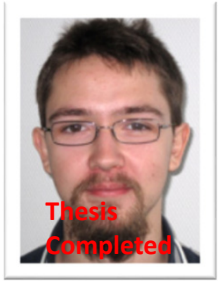


PhD Student
S. Bakarat

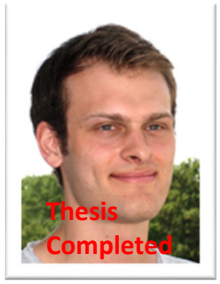
PhD Student
R. Sedona

PhD Student
P. H. Einarsson

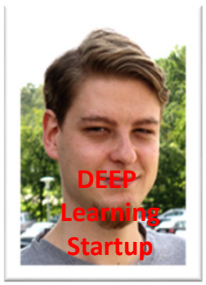
Dr. M. Goetz
(now KIT)



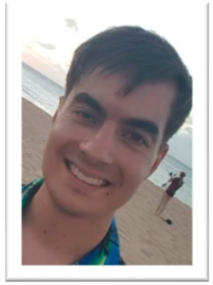
MSc M.
Richerzhagen
(now other division)



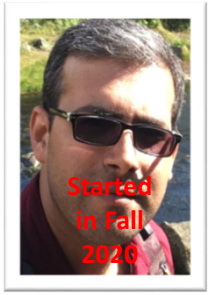
MSc
P. Glock
(now INM-1)



MSc
C. Bodenstein
(now
Soccerwatch.tv)



MSc G.S.
Guðmundsson
(Landsverkjun)



PhD Student
Reza

This research group has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 763558 (DEEP-EST EU Project) and grant agreement No 951740 (EuroCC EU Project) & 951733 (RAISE EU Project)


services
measurement
within
fusion
new
funding
Policy-based
concepts
device
Analysis
forms
cross-disciplinary
resources
Centre
climate
Computer
expertise
Approaches
construction
Fusion
year
like
Computational
Test
E.g
Computing
disciplines
results
Enable
different
Cross-Disciplinary
methods
High
big
nodes
PB
community
Environment
Modelling
EU
Science
centers
Key
scientific
important
files
model
international
etc
DKRZ
Use
storage
basis
Infrastructure
SMAQ
engineering
References
formats
simulations
ENES
technologies
brain
increasing
via
EUDAT
physical
IPCC
Big
Project
Data
processing
Common
TB
Methods
MapReduce
create
computing
Using
European
databases
GB
Supercomputing
Work
images
codes
analysis
Large
Summary
data-intensive
Service
Scientific
often
stored
Resources
analyze
levels
DLCL
Level
external
use
Research
Matters
Concrete
performance
JSC
computational
research
e.g
science
HPC
Juelich
datasets
acquisition
large-scale
system
Landscape
compute
management
System
Climate
modelling
AR5
Hardware
access
hundreds
Nutshell
Services
cores
distributed
current
network
DLCLs
Understanding
structures
Technologies
Earth
Structure
Germany
Simulation
directory
project
General
Provide
NASA
Energy
systems
day
Processing
Health