esiwace

CENTRE OF EXCELLENCE IN SIMULATION OF WEATHER AND CLIMATE IN EUROPE

Training Programme

Table of contents

ESiWACE2 offers and supports various training programmes on pre-exascale HPC software engineering, methods and tools for engineers and scientists in the domain of weather and climate. These trainings constitute a transfer of general knowledge from ESiWACE2 experts to the community.

This document provides an overview of the trainings that will be offered during the runtime of the ESiWACE2 project from 2019 to 2022.

Calendar of the Trainings

Check the contents and calendar of the training: https://www.esiwace.eu/services/trainings

Impact

The ESiWACE2 trainings are offered in different HPC areas of ESiWACE2 pre-exascale expertise i.e. I/O, computation (DSLs, C++ and coupling software), data analytics and containerisation. ESiWACE2 also plans to organise two summer schools to train scientists in these matters. Our trainings result in a larger number of scientists and engineers with higher qualifications in the use and optimisation of climate and weather applications on tier-0 machines, increasing HPC awareness in that community.

For all trainings, we embrace the creation of online digital media as Open Educational Resources (OER) material (https://www.oercommons.org/) that is available under a permissive CC-by licence and can be reused by teachers and researchers outside the consortium.

Container Hackathon for Modellers

Atmosphere and ocean models are characterised by complex dependencies, external configurations, and performance requirements. The objective to containerise such software stacks helps to provide a consistent environment to ensure security, portability and performance. Since the container is built only once, but then can be deployed on multiple platforms, productivity is increased. ETH Zurich provides subsequent Docker/Shifter support for the teams to complete the containerisation of their models.

Aim of the hackathon is to create containerised versions of Earth system models (e.g. COSMO, ICON, Nemo, OpenIFS, EC-Earth) using a Docker-compatible technology such as Shifter and to identify representative test cases.

In order to help evaluating performance when containers are used and which adaptations are necessary, the ported Earth system models from the hackathon is going to be deployed on CSCS supercomputer (Piz Daint) using an HPC-oriented container runtime. The Docker-compatible format of runtime will ease container portability across different systems and help collecting the performance figures from the representative test case.

Audience: Open to all modelling teams involved in ESiWACE2 but we are open also for groups outside the ESiWACE2 community.

Schedule: Three days in autumn/winter 2019.

Contact Person: Lucas Benedicic, CSCS (ETH Zurich)

Training on Data Input/Output and HPC

Topics:

- Software stack as a standardised and efficient platform for Data Input/Output (I/O)
- Storage application programming interfaces (APIs) and file systems
- Efficient creation of I/O workflows
- Performance and efficiency of file systems, non-volatile memory (NVM), tape or object storage

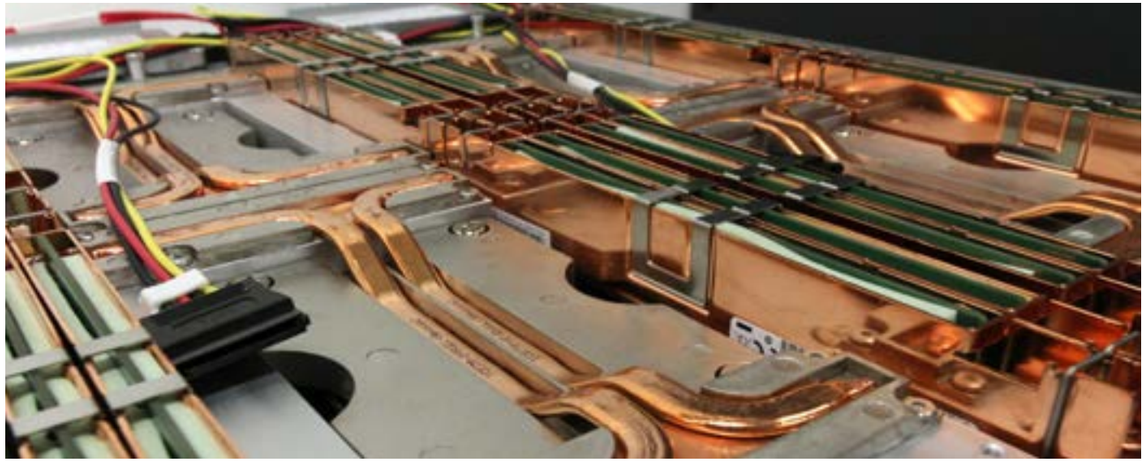
ESiWACE2 plans to create descriptive course material covering the co-designed software stack of WP4 as a standardised and efficient platform for I/O and on how the analysis stack (Cylc, XIOS, ESDM, NetCDF, HDF5) integrates with the underlying storage APIs and file systems (e.g., Mero/Clovis, Lustre), efficient creation of I/O dominated workflows, performance and efficiency considerations when dealing with file systems, NVM, tape or object storage, and cost-considerations in storage architectures.

The material covers roughly 8 hour course work in the form of presentations and a summarising paper covering the individual aspects, with webinars presenting the created OER material and blog entries describing issues. Finally, we plan to harness existing community approaches like the Virtual Institute of I/O (VI4IO, https://www.vi4io.org/) for promoting this training.

Audience: End-users of the weather and climate data community, but the course is generally useful for any data-intensive science.

Schedule: This training is planned as an 8-hour course via webinars and/or blog entries.

Contact Person: Julian Kunkel, University of Reading (UREAD)



Training on Domain Specific Languages (DSL)

The rapid changes in the multiple supercomputing architectures used to run weather and climate codes and the different programming models used seriously affect the development productivity and the ability to retain a single source code running efficiently everywhere. Domain-specific languages provide a solution to portability of these codes. In this training, we provide insights into DSLs considered in ESiWACE2 (PSyclone, CLAW and GridTools ecosystem) and demonstrate how to apply them to weather and climate models.

Topics:

- Insights into PSyclone, CLAW and GridTools ecosystem
- Demonstration into how to apply them to weather & climate models

Audience: Participants interested in theoretically and practically learning how to use the DSL languages to implement partial differential equation operators. During a hands-on session, participants are encouraged to implement some of the benchmark models defined by WP2 using DSLs and to build their own toy models, followed by an in-depth evaluation of generated optimised implementation and performance benefits. This training is planned for 25 people.

Schedule: This training is planned as a 5-day face-to-face course in 2020.

Contact Person: Carlos Osuna, MeteoSwiss

Training on Advanced C++ for HPC

C++ is a very powerful programming language, used worldwide to develop complex and performance-critical applications. It is therefore an important candidate for developing HPC applications. Mastering the power of the language requires substantial effort but pays off as projects scale up in size and complexity. As the hardware architectures become more and more diverse and complex, C++ allows the implementation of the proper abstractions to make applications sustainable for the future. Specifically, C++ allows the development of type-safe, flexible and portable functionalities, with no runtime overhead.

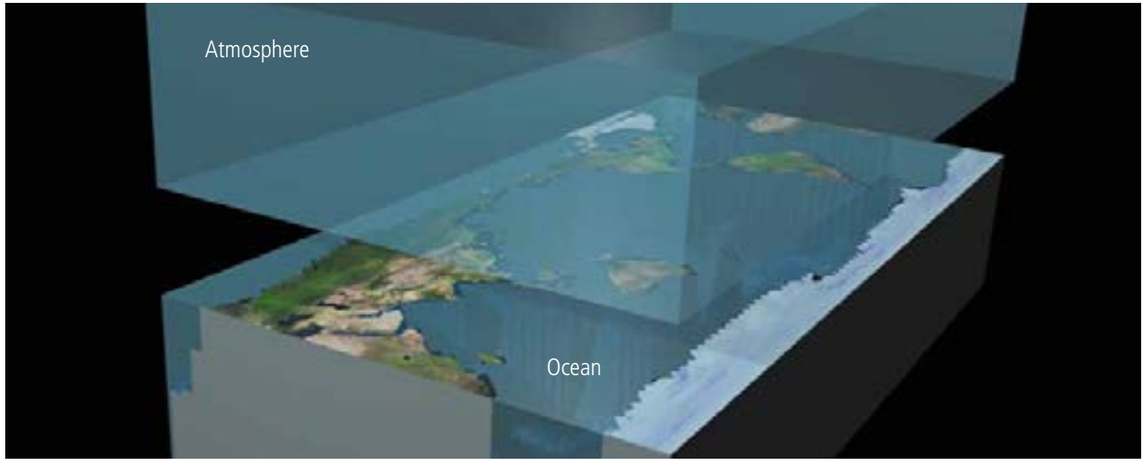
Topics:

- Generic programming techniques
- Application Programming Interface (API) development
- Specific C++-11/14 constructions

Audience: Up to 12 domain scientists can further improve their software-engineering skills by participating in the autumn 2021.

Schedule: This training is planned as a 3-day face-to-face course in 2021 at CSCS, Lugano (CH).

Contact Person: William Sawyer, CSCS (ETH Zurich)



Training on OASIS3-MCT

This training offers basic concepts in code coupling using the OASIS3-MCT coupler, focussing on the ocean-atmosphere context.

Topics: The material covers theoretical concepts of code coupling, instructions on how to download, install and compile OASIS3-MCT and finally the implementation of the coupling between two toy models in a hands-on tutorial session.

Audience: Engineers, physicists, and computer scientists interested in using this code coupling software in their own coupled model.

Schedule: This training is planned once a year, over four consecutive weeks, in 2020, 2021 and 2022, with weekly online learning activities requiring about 2-hour of work per week from the participants.

Contact Person: Sophie Valcke, CERFACS

Training on High Performance Data Analytics (HPDA)

The objective of this training course is to increase scientists' expertise on scientific data analysis at scale applied to climate and weather domains, using high-performance data analytics tools available from the open source market (i.e., Ophidia). The training covers from simple analytics tasks to workflows and applications (e.g., Python-based) and provides best practices and guidelines on dealing with massive scientific datasets on HPC architectures. The training foresees hands-on exercises carried out through Jupyter Notebooks.

Topics:

- Big data
- Introduction to scientific data management
- Scientific data analytics at scale
- Analytics workflows for eScience
- Big Data in HPC: High-performance data management
- Open-source High Performance Data Analytics (HPDA) tools

Audience: The training addresses several vertical training levels (e.g., intermediate, expert) from different horizontal perspectives (e.g., end-user, developer, administrator).

Schedule: This training is planned as three 8-hour online training courses in 2020, 2021 and 2022, over 4 consecutive weeks, with 2 hours of work per week required for the participants.

Contact Person: Sandro Fiore, CMCC, and Niklas Röber, DKRZ

Training on Containerisation

The traditional HPC landscape was led by computational physics, chemistry and other disciplines requiring huge processing performance and running on large number of nodes. Recently, more scientific disciplines rely on computing resources and introduce new requirements for the HPC environments, some needing only a fraction of one compute node for a short time while others need more powerful nodes than those regularly available. Such diversity is pushing software development and distribution towards portable virtualization technologies that improve research reproducibility, and scientists rely more and more on containers to run their experiments. Specifically, containers address many of the requirements needed by researchers of the different scientific disciplines. In this context, Docker is the de-facto standard as a platform for the creation and distribution of containers.

This course is an introduction to an end-to-end scientific computing workflow utilizing Docker containers. Attendees learn about the fundamentals of containerisation and the advantages it brings to scientific software. Participants then familiarize with Docker technologies and tools, discovering how to manage and run containers on personal computers, and how to build applications of increasing complexity into portable container images. Particular emphasis is given to software resources which enable highly-efficient scientific applications, like MPI libraries and the CUDA Toolkit. The training also explains the Docker distribution model based on cloud registries. The last part of the tutorial focuses on deploying Docker images on high-end computing systems, using a container platform capable of leveraging the performance and scalability of such machines, while maintaining a consistent user experience with Docker. After the tutorial, a hands-on session gives attendees the chance to freely experiment with building and running their own containers. Materials are provided to replicate the examples from the tutorial or tackle more advanced exercises.

Topics:

- Containers general concepts
- Introduction to Docker
- Running containers on personal workstations
- Managing images
- Packaging scientific applications with container images
- Distributing images: remote registries and Docker Hub
- Running containers on high-performance systems

Audience: Scientists, engineers and students interested in introducing container technologies in their activities. No previous container knowledge required. Familiarity with Linux and using a command-line terminal is highly beneficial.

Schedule: This training is planned as a 4-hour introductory tutorial and 2-hour hands-on course. The date is not fixed yet.

Contact Person: Lucas Benedicic, CSCS (ETH Zurich)

Contact

Visit our website: https://www.esiwace.eu/

Email: esiwace@dkrz.de

Follow us on Twitter: @esiwace
https://twitter.com/esiwace/

Access our Results

We support Open Access!

All our results – publications and reports – are available on Zenodo:
https://zenodo.org/communities/esiwace/

Funding

ESIWACE2 has received funding from the European Union's Horizon 2020 research and innovation programme (H2020-INFRAEDI-2018–1 call) under grant agreement No 823988.

Print: dieUmweltDruckerei GmbH Hannover · print using a carbon-neutral process on FSC certified paper
Editor: DKRZ (German Climate Computing Center) · Bundesstrasse 45a · 20146 Hamburg · 2019 · info@dkrz.de