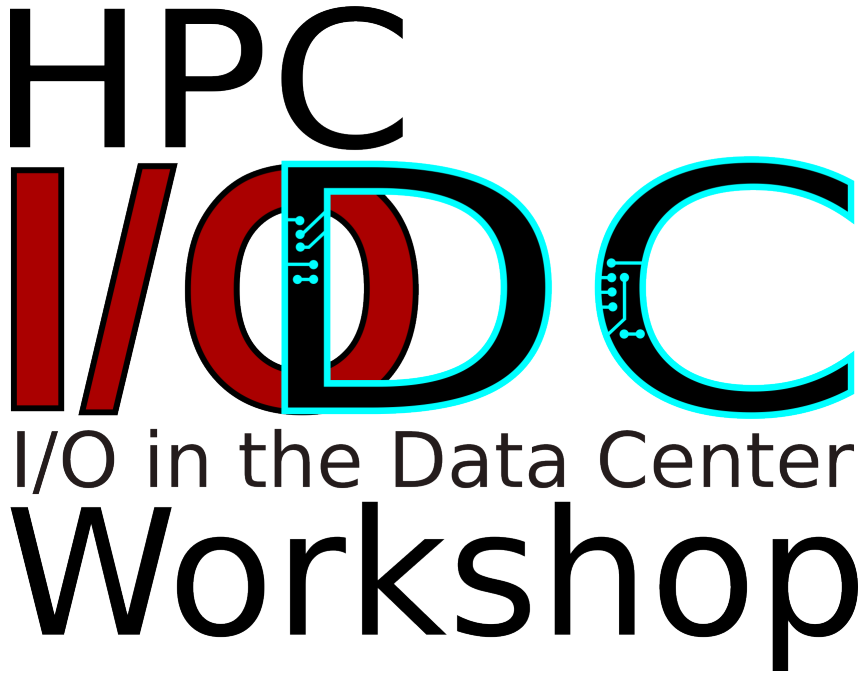
HPC
I/O
DC
I/O in the Data Center
Workshop

HPS
https://hps.vi4io.org

Julian M. Kunkel (University of Göttingen/GWDG)
Jay Lofstead (Sandia National Laboratories)
Jean-Thomas Acquaviva (DDN)

# 7th HPC I/O in the Data Center Workshop

# Sponsoring

The workshop is powered by

esiwace
CENTRE OF EXCELLENCE IN SIMULATION OF WEATHER AND CLIMATE IN EUROPE

The Virtual Institute for IO

The Journal of
High-Performance Storage

# EU funded Project: ESiWACE

## The Centre of Excellence in Simulation of Weather and Climate in Europe

- Representing the European community for
  - climate modelling and numerical weather simulation
- Goals in respect to HPC environments:
  - Improve efficiency and productivity
  - Supporting the end-to-end workflow of global Earth system modelling
  - Establish demonstrator simulations that run at highest affordable resolution
- Funding via the European Union's Horizon 2020 program (grant #823988)

http://esiwace.eu

esiwace
CENTRE OF EXCELLENCE IN SIMULATION OF WEATHER AND CLIMATE IN EUROPE

# The Virtual Institute for I/O

## Goals

- Provide a platform for I/O enthusiasts for exchanging information
- Foster training and collaboration in the field of high-performance I/O
- Track and encourage the deployment of large storage systems by hosting information about high-performance storage systems

https://www.vi4io.org

# Outlook for VI4IO

## Selection of ongoing activities

- Collaboration with the HPC Certification Forum
  - https://hpc-certification.org
  - Ensure that IO-related competences are properly represented in the skill-tree
- Capturing file-system specific characteristics properly
  - For the high-performance storage list and the IO500

## The Goals of the HPC Certification Forum (HPCCF)

- Fine-grained standardizing HPC knowledge representation
- Establishing international certificates attesting knowledge
- Supporting an ecosystem around the HPC competences

# The Journal of High-Performance Storage (ISSN 2748-7814)

## Features

- Open reviews, i.e., anyone can provide feedback
- Living papers, i.e., can improve over time
- Digital replicability (of analysis/experiments)
- Free open access

https://jhps.vi4io.org

## Status

- First issue published on 2021-01-29
- Two pending papers in the incubator will be published soon

# Motivation for the Workshop

- I/O perspective of centers is often ignored
- Data centers aim to provide optimal service and performance

## Providing a good storage strategy is challenging

- Zoo of emerging storage solutions/technology
- Variety of HPC file systems have pro/cons
- Management of large volume/file numbers (PByte/Billions of files)
- Middleware to fix file system issues present in all file systems
- Data center needs to consider IO interaction of applications/workflows

# Understanding Systems and Users

## Knowing the behavior would allow to provide a better system

- A perfect understanding of usage and efficiency would allow for
  - selection of the right storage technology
  - gearing optimization effort towards mostly used I/O libraries
  - understanding the requirements for the procurement
  - optimizing the data center's efficiency as a whole
- But users often don't know their I/O patterns (and workflows)
- The I/O stack is challenging even for experts

## Maybe I/O experts from data centers can make a difference

- From **individual** activity towards **community** effort and ultimately useful **conventions**

# About the HPC-IODC Workshop

## Goal: Bring together I/O experts from data centers

- Regardless of file system
- Foster information exchange
- Opportunity for networking

## Topics of interest

- Scientific workload
- Usage characteristics (file, folders, scientific libraries)
- System perspective
- Architecture
- Performance aspects and monitoring
- Issues during production and potential solutions

# Workshop

## Dissimination of results

- Presentations will be made available on our webpage
- Videos will be published on YouTube (and linked)
- Research Papers are published in Springer LNCS and potentially in JHPS
- We will write a preface and summarize the workshop results

## Interactivity

- Critical discussion is welcome and expected from attendees
- Put questions/comments in the VC chat (or verbally at end of a talk)
- Discussion time slots: open topics, everyone can raise/discuss issues
  - Ultimately controlled by a moderator
  - Discussion is documented on Google Doc (link on the webpage)

# Agenda

## Morning session (available in BigBlueButton)

8:55 *Welcome*

9:00 **Research session**

11:00 **Expert talks**

13:00 *Virtual Lunch Break*

## Afternoon (using ISC HPC Zoom)

14:00 **Welcome**

14:05 *Panel: The impact of HPC and Cloud convergence on storage*

15:30 **Student mentoring session / Research talks**

17:00 *Expert talk*

17:30 *Discussion of hot topics*

18:15 *Farewell*

The ESiWACE project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No **823988**

*Disclaimer: This material reflects only the author's view and the EU-Commission is not responsible for any use that may be made of the information it contains*