

# HPC I/O in the Data Center

6th HPC-IODC Workshop

Julian Kunkel     Jay Lofstead     Jean-Thomas Acquaviva

November 23, 2019

## Workshop Organizer Information

### Dr. Julian Kunkel

Dr. Kunkel is a Lecturer at the Computer Science Department at the University of Reading. Previously, he worked as postdoc in the research department of the German Climate Computing Center (DKRZ) that partners with the Scientific Computing group at the Universität Hamburg. He manages several research projects revolving around High-Performance Computing and particularly high-performance storage. Julian became interested in the topic of HPC storage in 2003, during his studies of computer science. Besides his main goal to provide efficient and performance-portable I/O, his HPC-related interests are: data reduction techniques, performance analysis of parallel applications and parallel I/O, management of cluster systems, cost-efficiency considerations, and software engineering of scientific software.

### Dr. Jay Lofstead

Dr. Jay Lofstead is a Principal Member of Technical Staff at Sandia National Laboratories in Albuquerque, New Mexico. Since 2010, Jay has been working on HPC simulation workflows focusing on data management issues and as well as general I/O and storage issues for HPC. His prior work includes the R&D100 Award winning ADIOS I/O componentization framework in use in more than 30 production scientific simulations. He is a member of several conference and workshop program committees.

### Dr. Jean-Thomas Acquaviva

successively worked for Intel, the University of Versailles and the French Atomic Commission (CEA). He participated to the creation of their joint laboratory the Exascale Research Centre, where he led the Performance Evaluation Team. Today he's actively contributing the development of the DDN Storage Advanced Technology Centre in France. Jean-Thomas has hands-on experience in the HPC

ecosystem, with positions in start-up, large research public institutions, SME or global organizations. He has been a direct contributor to the ETP4PH Strategic Research Agenda and he's chairing two conferences on high-performance storage. Jean-Thomas was main organizer of the **Workshop On Performance and Scalability of Storage Systems (WOPSSS)** that run from 2016-2018 at ISC HPC.

## Abstract

Managing scientific data at large scale is challenging for scientists but also for the host data center. The storage and file systems deployed within a data center are expected to meet users' requirements for data integrity and high performance across heterogeneous and concurrently running applications.

With new storage technologies and layers in the memory hierarchy, the picture is becoming murkier. To effectively manage the data load within a data center, I/O experts must understand how users expect to use these new storage technologies and what services they should provide in order to enhance user productivity. We seek to ensure a systems-level perspective is included in these discussions.

The HPC-IODC workshop addresses the storage challenge from a unique perspective, moving the focus from the application-centric perspective to the perspective of data centers and operators. In the workshop, we bring together I/O experts from data centers and application workflows to share current practices for scientific workflows, issues and obstacles for both hardware and the software stack, and R&D to overcome these issues. To focus on relevant aspects and streamline the discussion, a list of relevant topics is provided as common structure of the talks. We welcome submission of scientific papers with state-of-the-practice and specifically focused on I/O in the datacenter. For further information regarding participation, the topics and the mailing list, see the web page.

## Keywords

HPC Centre Planning and Operations, File systems, Storage, Performance Analysis and Optimization

## Workshop length

Full day. Each prior time we have offered this workshop, we have easily filled the entire day.

## Workshop webpage

<https://hps.vi4io.org/events/2020/iodc>

## Targeted audience

- I/O experts from data centers and industry.
- Researchers/Engineers working on high-performance I/O for data centers.
- Interested domain scientists and computer scientists interested in discussing I/O issues.
- Vendors are also welcome, but their presentations must align with the same topics and not focus on commercial aspects.

## Estimated attendance

In the past, we had more than 45 attendees.

## Early Acceptance

We will contribute to the joint Springer LNCS post-conference proceedings.

## Workshop format

The workshop content is structured in three parts, firstly, in the morning we organize a keynote and have best papers and discussion of community activities. The afternoon session of HPC-IODC puts an emphasis on data center issues and covers research paper presentations regarding state of the practice in the data center, discussion slots, and talks from I/O experts.

Researchers and I/O experts can submit their proposals for a paper or talk according to a call for participation for papers and speakers. We provide a common list of topics to be addressed in each I/O expert's talk on our webpage, such that the individual presentations are aligned.

## Previous Instances

We held this workshop the first time ISC extended its program to include workshops (ISC 2015). Since then, the workshop is held annually attracting more than 40 attendees in subsequent instances. In 2019, we accepted 10 papers and a handful of expert talks attracting about 45 attendees.

## History

In ISC 2015, about 30 attendees were attracted, the workshop was organized around 10 invited presentations, and two discussion sessions.

In 2016, the workshop format was changed, in addition to the expert talks, we added a research section allowing submission of short scientific papers and a keynote talk. The attendance to the workshop was similar to 2015, but finally we accepted four research papers.

In 2017, more than 50 people attended the workshop; we accepted 5 papers. Despite the large attendance of previous instances, we were forced by the organizers to run only a half day. That year we teamed up with WOPSSS (a workshop focusing on performance aspects in I/O) and organized a full storage together promoting the two events on a joint webpage: <https://isc-hpc-io.org>.

In 2018, we organized a shared morning session with WOPSSS (45+ people) and then split after lunch into two fully autonomous workshops. We believe that collaboration with WOPSSS was fruitful and beneficial for the community still it didn't increase the number of attendees of either workshop.

In 2019, we jointly agreed with the WOPSSS organizers to apply for an independent workshop for the following reasons: the attendance of both workshops was high; the central theme differ<sup>1</sup>; we need enough discussion slots to preserve the unique workshop character of HPC-IODC. As WOPSSS didn't have sufficient submissions, WOPSSS was canceled last minute and we accepted three papers from the workshop. The WOPSSS organizers said they will move to a different conference leaving HPC-IODC as the sole storage and IO related workshop at ISC.

## Expected outcome

- 1) Sparking impulses for the development of data center I/O.
- 2) The networking among the participants will be improved.
- 3) We contribute a summarizing workshop paper to ISC's post-conference workshop proceedings in LNCS.

## Details on the Call for Papers

We accept short papers with up to 12 pages (excl. references) in LNCS format. We cover data center research and state-of-the-practice papers. Our targeted proceedings are ISC's post-conference workshop proceedings in Springer LNCS. We will use Easy-chair for managing the proceedings and PC interaction. Additionally, we explore the usage of an open review model, e.g., paper drafts will be publicly available and can be assessed/discussed by the public.

Expected number of papers: We expect 14 submissions<sup>2</sup>. With an acceptance rate of 50%, we aim to fill at least half of the presentation slots with papers.

---

<sup>1</sup>There was only little overlap in respect to research papers as WOPSSS focus was on general performance analysis, tools, results and characterization, while HPC-IODC addresses the operational side and data center aspects.

<sup>2</sup>We invited the PC to submit papers and have external agreements.

**Schedule:**

- Announcement/CfP: As soon as workshop notifications are sent by ISC
- Submission deadline: 2020-02-10
- Author notification: 2020-04-24
- Workshop: 2020-06-25

Note that our call for speakers follows a similar schedule, but they'll have to submit only a talk abstract and a brief bio.

Members of the program committee

- George S. Markomanolis (Oak Ridge National Laboratory)
- Suren Byna (Lawrence Berkeley National Laboratory)
- Adrian Jackson (The University of Edinburgh)
- Javier Garcia Blas (Carlos III University)
- Bing Xie Oak (Ridge National Lab)
- Sandro Fiore (CMCC)
- Glenn Lockwood (Lawrence Berkeley National Laboratory)
- Michael Kluge (TU Dresden)
- Jean-Thomas (Acquaviva DDN)
- Robert Ross (Argonne National Laboratory)
- Wolfgang Frings (Juelich Supercomputing Centre)
- Feiyi Wang (Oak Ridge National Laboratory)
- Thomas Boenisch (High performance Computing Center Stuttgart)

To increase the participation further, we firstly will actively invite representatives of different data centers and representatives of active workflow users/developers to give a talk and/or submit short papers regarding recent issues. Secondly, we will announce the workshop on typical mailing lists and send the call for participation.

For further information, see: <https://hps.vi4io.org/events/2020/iodc>.