

Wissenschaftliches Rechnen - Scientific Computing - 2016

The group **Wissenschaftliches Rechnen** (Scientific Computing) is composed of members of Universität Hamburg and the German Climate Computing Center (DKRZ).

We conduct research on high performance storage systems and energy efficiency of cluster infrastructure. We have expertise in parallel programming and environmental modeling. In 2016, we participated in eight funded research projects.



Top (LTR): K. Chasapis, M. Hensel, T. Ludwig, M. Zimmer, E. Betke
Middle (LTR): N. Hübbe, H. Lenhart, M. Kuhn, C. Hovy, J. Kunkel, Y. Alforov, N. Jumah
Bottom (LTR): M. Heidari, A. Novikova, F. Große, J. Lüttgau
Missing: A. Fuchs, J. Squar

Ongoing Projects

BigStorage (Yevhen Alforov, Michael Kuhn)

BigStorage is a European Training Network (ETN) whose main goal is to train future data scientists in order to enable them to apply holistic and interdisciplinary approaches for taking advantage of a data-overwhelmed world. This requires HPC and Cloud infrastructures with a redefinition of storage architectures underpinning them while focusing on meeting highly ambitious performance and energy usage objectives. During the last year, several applications from the domain of earth system science were analyzed with regard to their I/O requirements within a subproject. After communication and cooperation with climate scientists I/O and data management requirements were collected and outlined in an official document. In addition, advantages and limitations of widely used and available parallel and distributed file systems were identified and presented in one of the deliverables. According to the main objectives of BigStorage power-saving and energy-efficient solutions (data reduction techniques) and approaches for measuring and modeling power consumption were examined. Results of this work were published in one of the deliverables.

Keywords: data reduction, climate science, high level I/O, energy-efficiency.

ESiWACE (Julian Kunkel, Jakob Lüttgau, Philipp Neumann)

Within the Centre of Excellence in Simulation of Weather and Climate in Europe (ESiWACE), a H2020 funded project, we started the design of an I/O middleware to improve performance for climate and weather related use cases. The middleware will integrate into NetCDF and HDF5 but utilize multiple storage tiers as back-ends concurrently. We created a report on the cost and future of data center storage infrastructure.

Keywords: climate and weather science, I/O, middleware, exascale computing, big data

LSDMA (Konstantinos Chasapis, Michael Kuhn)

The LSDMA project is funded by the Helmholtz Association of research centers in Germany and targets to extend and improve their data services. The project structure is divided into Data Life Cycle Laboratories (DLCLs) and the Data Services Integration Team (DSIT). DLCLs collaborate with the scientists during the entire data life cycle. From this collaboration they define the characteristics of community-specific tools and mechanisms that will improve their workflow efficiency. On the other side, DSIT's responsibility is to implement generic cross-community technologies based on the requirements set of the DLCs. Our group is involved in the Performance and Power Optimization working package that is part of DSIT. To this extend we have contributed to the project in the following manners: i) providing best practices for Scientific I/O with HDF and NetCDF, ii) evaluating metadata operation in Lustre file system, iii) modeling power consumption of HPC servers and iv) exploring the compressibility of scientific data. The results of our research have been published on several conferences the past years.

Keywords: data compression, data management, energy-efficiency, efficient I/O

Publication: **A Best Practice Analysis of HDF5 and NetCDF-4 Using Lustre** (Christopher Bartz, Konstantinos Chasapis, Michael Kuhn, Petra Nerge, Thomas Ludwig), In High Performance Computing, Lecture Notes in Computer Science (9137), pp. 274–281, (Editors: Julian Martin Kunkel, Thomas Ludwig), Springer International Publishing Switzerland), ISC 2015, Frankfurt, Germany, ISBN: 978-3-319-20118-4, ISSN: 0302-9743, 2015-06, http://dx.doi.org/10.1007/978-3-319-20119-1_20

Ufo 16 (Fabian Große, Hermann Lenhart)

Ufo 16, a project from the German environmental protection agency (Umweltbundesamt UBA) was set up as an environmental impact assessment in relation to the reduction of river nutrient loads in the frame of new EU legislation. One special focus was the development of ecological dangerous oxygen deficiency zones. The process understanding on the North Sea oxygen dynamics resulted in a high-impact publication (Große et al., 2016), while the extension of the trans-boundary nutrient transports method on the oxygen dynamics formed the basis for the PhD thesis of Fabian Große, which was finalised in May 2017.

Keywords: ecosystem modelling, north sea, eutrophication, TBNT modelling, oxygen deficiency

Publication: **Looking beyond stratification: a model-based analysis of the biological drivers of oxygen deficiency in the North Sea** (Fabian Große, Naomi Greenwood, Markus Kreuz, Hermann Lenhart, Detlev Machoczek, Johannes Pätsch, Lesley A. Salt, Helmut Thomas), In Biogeosciences, Series: 13, pp. 2511–2535, (Editors: Veronique Garçon), Copernicus Publications (Bahnhofsallee 1e, 37081 Göttingen, Germany), 2016, <http://dx.doi.org/10.5194/bg-13-2511-2016>

New Projects in 2016

AIMES (Nabeeh Jumah, Anastasiia Novikova, Julian Kunkel)

With the Advanced Computation and I/O Methods for Earth-System Simulations (AIMES) project we address the key issues of programmability, computational efficiency and I/O limitations that are common in next-generation icosahedral earth system models. To increase programmability, in a co-design approach of scientists, we defined a domain-specific language (DSL) that abstracts from technical details. A prototypical tool to translate the DSL into a general purpose language has been created. To improve I/O, we developed the Scientific Compression Interface Library (SCIL) that decouples the usage of data compression algorithms from the requirements of data quality and, thus, will increase code portability. SCIL will choose the best available algorithm to meet various user-defined criteria regarding data quality and required performance.

Keywords: climate science, data reduction, domain-specific language, performance-portability

Bull I/O cooperation (Eugen Betke, Julian Kunkel)

The Bull I/O cooperation is about detection, analysis and elimination of I/O problems. Within the scope of the project we developed a Lustre monitoring daemon and an online monitoring framework, whose purpose is to detect inefficient applications and provide hints where the problem may be caused. A major achievement of the project is the evaluation of the NVRAM-based burst buffer. The results will be considered in the architecture of the next HPC system at DKRZ. Another achievement is the development of an open source NetCDF benchmark that is able to mimic many popular access patterns.

Keywords: online monitoring, I/O analysis, burst buffer

Publication: **MPI-IO In-Memory Storage with the Kove XPD** (Dr. Julian Kunkel), PDSW-DISC Workshop, Salt Lake City, USA, 2016-11-14

IPCC-L (Anna Fuchs, Michael Kuhn, Julian Kunkel)

Universität Hamburg has become one of five Intel® Parallel Computing Centers for Lustre worldwide. The project “Enhanced Adaptive Compression in Lustre” was granted at the beginning of 2016 and has an initial runtime of two years. The aim of the project is to enable compression within the Lustre filesystem. At first, the infrastructure will be prepared to pass through the compressed data and make the backend (ZFS) handle them correctly. This already involves client- as well as server-side changes. Each stripe will be chunked, compressed and sent over the network. Preliminary user space analysis has shown that read-ahead can become a big problem when the chunks are read with logical gaps. The next technical challenge is to integrate the changes into ZFS. Once the infrastructure is done, the actual topic of adaptivity and dynamic decision making will be investigated.

Keywords: filesystem, Lustre, compression

Publication: **Data Compression for Climate Data** (Michael Kuhn, Julian Kunkel, Thomas Ludwig), In Supercomputing Frontiers and Innovations, Series: Volume 3, Number 1, pp. 75–94, (Editors: Jack Dongarra, Vladimir Voevodin), 2016-06, <http://dx.doi.org/10.14529/jsfi1601>

i_SSS (Jannek Squar, Michael Kuhn, Hermann Lenhart)

The project „integrated Support System for Sustainability“ started at the beginning of 2016 and has a project time of 5 years. The aim of the project is to enable farmers to determine site characteristics of their field in order to apply measures for a resource and environment-friendly agriculture. The information for the farmers will be provided by the geographic information system SAGA (System for Automated Geoscientific Analyses). Therefore, the SAGA tool is the central platform for development which also includes the incorporation of model information, e.g. on hydrological conditions as well as remote sensing data. To fulfil the targets of i_SSS, access to weather data is necessary to evaluate and run models for terrain analysis. For this purpose we gained access on historical forecast data of the last two years. We developed a tool to handle weather data from the Deutscher Wetterdienst (but also from other sources like Global Forecast System and RADOLAN) and to preprocess weather data before they are loaded into SAGA. Different approaches and tools for loading and pre-processing input data were evaluated to select the one with the most promising prospect.

Keywords: digital farming, geographical information system, OpenMP

Additional activities

Besides research activities, the group was active with respect to HPC and teaching as well.

We prepared workshops and BoFs for the Supercomputing Conference in Salt Lake City and the ISC High Performance Conference in Frankfurt. For the third time in a row, the research group has participated in the Student Cluster Competition with a team of students from Universität Hamburg. The team used a cluster consisting of two nodes with four NVIDIA Tesla K80s each and has achieved the second best HPL performance with 12.29 TFLOPS. More information can be found in the students' report: https://wr.informatik.uni-hamburg.de/media/teaching/sommersemester_2016/scc-16-bericht.pdf

In addition, we participate in the European Open File System (EOFS) organization and in the Virtual Institute for I/O. One goal is the community-driven development of the IO-500 benchmark and list, which, similar to the TOP500 list, allows the tracking of storage performance over the years.

In terms of teaching, Scientific Computing offers many modules and events. For instance, the research group has offered a project about parallel computing at the Informatics Department's Girls' Day. Furthermore, the expert module, concerning of the lecture of high performance computing by Prof. Thomas Ludwig and the introduction of parallel programming for geoscience student has become a well accepted choice for geoscience students as an extension of their lectures. For additional information about our teaching activity have a look at <https://wr.informatik.uni-hamburg.de/teaching/>

Further information about projects, publications and other activities are available online:

<https://wr.informatik.uni-hamburg.de/>