Computing and Informatics, Vol. 39, 2020, 617–621, doi: 10.31577/cai_2020_4_617

# PREFACE TO THE SPECIAL ISSUE ON PROVIDING COMPUTING SOLUTIONS FOR EXASCALE CHALLENGES

Dieter Kranzlmüller

*Munich Network Management Team (MNM-Team)*
*Ludwig-Maximilians-Universität*
*Munich, Germany*
*&*
*Leibniz Supercomputing Centre (LRZ)*
*Bavarian Academy of Sciences and Humanities*
*Garching, Germany*
*e-mail:* kranzlm@mnm-team.org

Maximilian Höb

*Munich Network Management Team (MNM-Team)*
*Ludwig-Maximilians-Universität*
*Munich, Germany*
*e-mail:* hoeb@mnm-team.org

Today, as we expect the appearance of the first supercomputer with exascale capabilities, some tough questions remain to be solved. A main challenge of the upcoming exascale era will appear around the ability to process extremely large datasets of several peta- and exabytes effectively and efficiently. This needs to be addressed with new management and architectural approaches, focusing on modular, integrated and light-weight solutions, which enable end-users across all scientific domains to utilize these future supercomputing systems. This special issue gives an overview on approaches focusing of these challenges and exacting applications, requiring exascale-ready compute and data services.

In their paper Bobák et al. [2] present an exascale reference architecture as a layered framework to overcome exascale challenges, developed within the PROCESS project[1]. Their modular and service-oriented approach enables low-overhead

[1] www.process-project.eu

usage of the underlying HPC, cloud or accelerated-based infrastructure, completely abstracted to end-users and simplifying the configuration, deployment and data-handling process.

Data management will be a major part of this work – the building block of any successful ecosystems. All services about processing multiple sources of distributed data must be smart, since more and more data sets are geographically spread across also various administrative domains. Cushing et al. [5] propose a scalable, programmable and easy to deploy data infrastructure supporting data-intensive applications.

Together with data management, future systems need to implement efficient deployment and computing services. Meizner et al. [10] detail their scalable computing platform, enabling complex workflow deployments on heterogeneous systems. Their evaluation includes data intensive applications, showing the ability of the approach. One of those applications is based on observation data from the LOFAR radio telescope. Madougou et al. [9] describe their simple point-and-click-reduction of the complex workflow to compute calibrated sky maps out of the LOFAR observations. This workflow relies on efficient data movement and makes use of the data infrastructure presented above.

Based on extremely large digital images of tumor tissue, histopathology evolved to a computational and storage demanding application. Graziani et al. [7] present a modular, containerized pipeline for the detection of tumor regions in digital specimens of breast lymph nodes with deep learning models. The three independent layers of this pipeline were evaluated on different computing resources accessible through the services described in [5] and [10].

Having a much deeper dive into HPC systems, supercomputers face not only efficiency challenges across different systems, but also inside single clusters. Parallel applications demand efficient distributed memory access to achieve large-scale performance. To decrease the complexity of implementing scientific applications, Gschwandtner et al. [8] propose a programming interface with the ease of shared memory programming models.

As Data analysis gains more importance as more data is available while today's systems are capable of processing more data. This can be seen not only on HPC clusters, but also on cloud resources. To assist end-users building scientific, cloud-based data analytic pipelines, Baranowski et al. [3] present the Cookery framework.

Bystrov et al. [4] investigate the performance of haemodynamic flow computations on a cloud infrastructure, focusing on the parallel performance analysis, energy consumption and virtualization overhead of the software service based on the ANSYS Fluent platform. Such evaluations are crucial to guarantee an efficient overall system usage and enable other domains to execute their applications also on such large-scale systems.

One such application is the calculation of the Levenshtein distance between two strings like the DNA, which was known as a sequential-only application. To make use of the ecosystem presented before, it requires a scalable parallel implementation.

Sadiq and Yousaf [6] introduce a distributed, parallel algorithm to calculate the Levenshtein distance.

Combining all these software and middleware oriented approaches, it is mandatory to consider the actual hardware resources available. Berberich et al. [1] provide an overview on the European HPC landscape and introduce a pan-European HPC portal collecting all information and facilitate access to the portfolio of services offered across Europe.

The collection of papers presented in this special issue provides some valuable solutions for using the future exascale supercomputers effectively and efficiently. However, at the same time, many aspects are useful on smaller systems as well and demonstrate the utilization of today's technology, such as for example containers, for applications in science and research. It is clear that the exascale threshold is a merely artificial construct. The true challenge stems from the ever increasing computational performance, and the necessity of scaling applications efficiently. Under this perspective, the tools and applications presented here demonstrate what is possible with today's technology, and even more, and the mindset needed to utilize the supercomputing systems tomorrow.

## Acknowledgment

We would like to thank Dr. Ladislav Hluchy, the Editor-in-Chief of Computing and Informatics (CAI) for his timely advice on this special issue. A big thanks also goes to Viera Jablonska, the CAI journal editorial assistant for her great support in publication of the special issue. This work has only be possible through the support of the European Commission within the EU project PROCESS, under the grant agreement No. 777533.

## REFERENCES

[1] Berberich, F.—Liebmann, J.—Teodor, V.—Nominé, J.-P.—Pineda, O.—Segers, P.: European HPC Landscape. Computing and Informatics, Vol. 39, 2020, No. 4, pp. 622–643, doi: 10.31577/cai_2020_4_622.

[2] Bobák, M.—Hluchý, L.—Habala, O.—Tran, V.—Cushing, R.—Valkering, O.—Belloum, A.—Graziani, M.—Müller, H.—Madougou, S.—Maassen, J.: Reference Exascale Architecture (Extended Version). Computing and Informatics, Vol. 39, 2020, No. 4, pp. 644–677, doi: 10.31577/cai_2020_4_644.

[3] Baranowski, M.—Belloum, A.—Cushing, R.—Valkering, O.: Cookery: A Framework for Creating Data Processing Pipeline Using Online Services. Computing and Informatics, Vol. 39, 2020, No. 4, pp. 678–694, doi: 10.31577/cai_2020_4_678.

[4] Bystrov, O.—Kačeniauskas, A.—Pacevič, R.—Starikovičius, V.—Maknickas, A.—Stupak, E.—Igumenov, A.: Performance Evaluation of Parallel

Haemodynamic Computations on Heterogeneous Clouds. Computing and Informatics, Vol. 39, 2020, No. 4, pp. 695–723, doi: 10.31577/cai_2020_4_695.

[5] Cushing, R.—Valkering, O.—Belloum, A.—Souley, M.—Bobak, M.—Habala, O.—Tran, V.—Graziani, M.—Müller, H.: Process Data Infrastructure and Data Services. Computing and Informatics, Vol. 39, 2020, No. 4, pp. 724–756, doi: 10.31577/cai_2020_4_724.

[6] Sadiq, M. U.—Yousaf, M. M.: Distributed Algorithm for Parallel Edit Distance Computation. Computing and Informatics, Vol. 39, 2020, No. 4, pp. 757–779, doi: 10.31577/cai_2020_4_757.

[7] Graziani, M.—Eggel, I.—Deligand, F.—Bobák, M.—Andrearczyk, V.—Müller, H.: Breast Histopathology with High-Performance Computing and Deep Learning. Computing and Informatics, Vol. 39, 2020, No. 4, pp. 780–807, doi: 10.31577/cai_2020_4_780.

[8] Gschwandtner, P.—Jordan, H.—Thoman, P.—Fahringer, T.: AllScale API. Computing and Informatics, Vol. 39, 2020, No. 4, pp. 808–837, doi: 10.31577/cai_2020_4_808.

[9] Madougou, S.—Spreeuw, H.—Maassen, J.: Processing Radio Astronomical Data Using the PROCESS Software Ecosystem. Computing and Informatics, Vol. 39, 2020, No. 4, pp. 838–859, doi: 10.31577/cai_2020_4_838.

[10] Meizner, J.—Nowakowski, P.—Kapala, J.—Wojtowicz, P.—Bubak, M.—Tran, V.—Bobák, M.—Höb, M.: Towards Exascale Computing Architecture and Its Prototype: Services and Infrastructure. Computing and Informatics, Vol. 39, 2020, No. 4, pp. 860–880, doi: 10.31577/cai_2020_4_860.

**Dieter Kranzlmüller** is Chairman of the Board of Directors at Leibniz Supercomputing Centre of the Bavarian Academy of Sciences and Humanities. In 2008 he joined the Board of Directors at LRZ and became Full Professor of computer science at the Chair for Communication Systems and System Programming at Ludwig-Maximilians-Universität Munich (LMU). His scientific focus lies in e-infrastructures, including network and IT management, grid and cloud computing, as well as in high performance computing, virtual reality and visualisation. He graduated from the Johannes Kepler University Linz, Austria. Having spent a number of years working in the IT industry, he returned to academia to work at the universities of Reading and TU Dresden, and to act as deputy director of the EGEE project at CERN in Geneva. He is truly internationally oriented and is a member of many European and international organizations in the field of IT.

**Maximilian Höb** is Associate Scientist in the Munich Network Management Team at Ludwig-Maximilians-University Munich and Co-Coordinator of the PROCESS project, in which he also contributes to two Use Cases in the area of data management and agricultural simulation based on the Copernicus data sets. His research focuses on large scale system architectures and performance-aware containerization of HPC applications.