HP-C/DA:

Workshop on the In Situ Co-Execution of High-Performance Computing & Data Analysis

Program overview

2:00pm	–	**2:20pm**	Workshop introduction
2:20pm	–	**4:00pm**	HPC/HPDA co-execution support
4:00pm	–	**4:15pm**	— *break* —
4:20pm	–	**5:40pm**	HPC & HPDA work-loads
5:40pm	–	**6:00pm**	Closing remarks

Detailed program

2:00pm – 2:20pm: Workshop introduction

2:00pm **HP-C/DA, an overview**

Julien Bigot, CEA/MdlS and Bruno Raffin, Inria

Workshop introduction.

2:20pm – 4:00pm: HPC/HPDA co-execution support session

2:20pm **HPC-HPDA Coupling at Scale: Experience and Perspectives**

Amal Gueroudji, CEA

The goal of HPC/HPDA coupling at scale is to contribute to building digital twins capable of combining data from sensors, simulations and advanced data analysis processes. Achieving this goal is still a challenge today. This talk will rely on state of the art work and experience with the development of frameworks like FlowVR, Melissa, Tins and various associated use cases for in situ processing, computational steering, ensemble runs, data assimilation, to analyze the state of progress and identify some key challenges.

2:40pm **Maestro: Towards a Data- and Memory-aware Middleware**

Dirk Pleiter, FZJ & Univ. Regensburg

Performance bottlenecks of modern HPC and HPDA systems are usually in the memory and storage systems. Still few abstractions exist that capture data semantics of applications and capabilities of the underlying hardware are often largely unknown to the software stacks. Therefore, reasoning about data movement and memory in software is impossible. With other words: Data- and memory-awareness is lacking. The Maestro project addressing this shortcoming by building a middleware library that allows characterising application data, to reason about how to load and store that data, to assess the cost of moving it and to automatise data movement across diverse memory systems. In this presentation, we will provide an overview of Maestro's concepts and discuss the details of its emerging implementation.

3:00pm **Advancing Object storage architectures for Extreme scale**

Sai Narasimhamurthy, Seagate

The landscape and architecture assumptions for extreme scale computing is changing with the increase in heterogeneity (GPUs, FPGA, etc) in the computing subsystem along side multi/many core processing. There is also the arrival of a new extremely fast and high capacity Non-volatile storage class memories within the storage device hierarchy along side traditional disk drives, archives and flash technology. The workload assumptions for extreme scale computing is also changing with AI/Deep learning components as part of the highly data centric processing pipeline that will become common place. All this means that the traditional assumptions of the storage system stack is changing. We will no longer be able to rely on parallel file systems, POSIX interfaces etc as they had been designed for a different generation with different architecture assumptions in mind. In this talk, we will discuss an object storage based storage system architecture that is designed to address the new requirements.

3:20pm **On the Convergence of HPC, Cloud and Data Analytics for Exascale Weather Forecasting - ECMWF Present and Future**

Tiago Quintino, ECMWF

Starting 2014, ECMWF has embarked on a 10 year research programme on HPC Scalability, aiming to achieve Exascale numerical weather prediction systems by 2025. The programme is now entering its 2nd phase, where focus is now on improvements to the operational forecasting system that incur deeper restructuring changes. ECMWF operational forecast generates massive amounts of I/O in short bursts, accumulating to tens of TiB in hourly windows. From this output, millions of user-defined daily products are generated and disseminated to member states and commercial clients all over the world. These products are processed from the raw output of the IFS model, within the time critical path and under strict delivery schedule. Upcoming rise in resolution and growing popularity will increase both the size and number of these products.

The adoption of a new object store (FDB version 5) for the time-critical operations has opened the door for more comprehensive improvements to the post-processing chain and enabled new access paths to very high-resolution time critical datasets. These improvements will bring product generation and data analytics closer to the NWP model and the model output data, to build true data-centric processing and analytics workflows.

These are part of ECMWF plans to achieve Exascale NWP by 2025 and to empower our users and member states with novel and increased usage of our weather forecast data. As Exascale NWP datasets are expected to feature between 250 TiB to 1 PiB per forecast cycle, the data-centric approach is critical to enable their efficient usage, by minimising data transport and bringing post-processing and insight discovery closer to the data source.

We present the latest ECMWF developments in model I/O, product generation and storage, and how we are reworking our operational workflows to adapt to forthcoming new architectures and memory-storage hierarchies, as we build bridges from HPC data producer to Cloud based data analytics workflows.

3:40pm **HPC storage and processing convergence using Mochi data services**

Matthieu Dorier, ANL

The limitations of parallel file systems in terms of performance and flexibility have led many scientists to turn towards in situ analysis techniques, in which as much of the data processing as possible is done while the application is running, to reduce I/O pressure and produce early results. An alternative to this radical paradigm consists of designing data services that are highly tailored to their target applications. These data services can at the same time store and process data, present an interface that is suitable to their end users, and adapt to the platform they run on. However, these data services bring an important challenge. How can one easily design and implement such a service on a per-application basis? In this talk, we will present our work around the Mochi project, a set of composable libraries that can be used to efficiently design HPC data services. We will show multiple examples of such services: HEPnOS, which targets high energy physics workflows, and FlameStore, a service design for ensemble training of deep neural networks.

4:00pm – 4:20pm: Afternoon break

4:20pm – 5:40pm: HPC & HPDA work-loads session

4:20pm **Multi-scale brain simulation in the Human Brain Project: The EBRAINS in-transit simulation and analysis infrastructure**

Wouter Klijn, FZ-Julich

The Human Brain Project (HBP), the ICT-based Flagship project of the EU, is building research infrastructure for understanding the brain and finding new treatments for brain disease. Important capabilities are brain simulations of large- and multiscale experimental and clinical data sets with integrated analysis toolkits. The HBP offers simulation engines at different scales of abstraction. These simulators are now been integrated to enable multi-scale model simulation. These models will be an important tool for answering system level questions of (whole) brain ailments How to create a science end-user friendly production system from multiple individual applications, each simulating specific scales, is an open question. One challenge is that the probability of a fault condition grows as a multiplicative function of the number of applications in a workflow and their failure probabilities. This talk will introduce the infrastructure being build by HBP for the EBRAIN research infrastructure to address this and other challenges.

4:40pm **Resilience and scalability of deep learning in supercomputers**

Leonardo Bautista-Gomez, BSC

In recent years, Deep Neural Networks (DNNs) have achieved outstanding results in a wide range of applications, including language processing, speech, and visual recognition. In the quest to increase solution accuracy, there are trends of increasing the size of training datasets as well as introducing larger and deeper DNN models. In addition, applying Deep Learning in new domains, such as health care and scientific simulations, introduce bigger data samples and more complex DNN models. Those trends make the DNN training computationally expensive for a single node. Therefore, large-scale parallel training on High-Performance Computing (HPC) systems or clusters of GPUs is becoming increasingly common to achieve faster training time for larger models and datasets. In this talk, we will present recent developments in both, DNN's scalability as well as resilience while training at a large scale in HPC systems.

5:00pm **Catalyst, SENSEI and ALPINE: a cost comparison**

Charles Gueunet, Kitware

In this era closer to exascale than ever, supercomputers are able to generate data faster than they can write to disk. In this context, the results of High Performance Computing (HPC) approaches used for simulation need to be analysed without saving the whole data to disk.

In situ analysis is the process to visualize, explore, and analyse data living on the memory without transferring it to the disk. This brings data analysis algorithms to run directly on the supercomputers running the simulations, allowing them to take benefits of their large computational power through High Performance Data Analysis (HPDA).

Our presentation will start with a general introduction to in situ analysis along with a brief presentation of three existing solutions related to the VTK/Paraview suite: Catalyst, SENSEI and ALPINE. We will then present a comparison of these solutions in terms of development and computational costs.

- Development costs: we give an insight of the amount of code that needs to be added to use each of these libraries along with the impact on compilation time and binary size.
- Computational costs: we evaluate the amount of memory and cycles added by the library on the simulation execution.

At the end, the goal of this talk is to help researchers choose the solution that best suits their needs.

5:20pm **VESTEC – Interactive Supercomputing for Urgent Decision Making**

Achim Basermann, DLR

The use of extreme computing in real-time applications with high velocity data and live analytics is within reach. The availability of fast growing social and sensor networks raise new possibilities in monitoring, assessing and predicting environmental, social and economic incidents as they happen. Add in grand challenges in data fusion, analysis and visualization, and extreme computing hardware has an increasingly essential role in enabling efficient processing workflows for huge heterogeneous data streams. VESTEC is creating the software solutions needed to realize this vision for urgent decision making in various fields with high impact for the European community. VESTEC is building a flexible toolchain to combine multiple data sources, efficiently extract essential features, enable flexible scheduling and interactive supercomputing, and realize 3D visualization environments for interactive explorations by stakeholders and decision makers.

5:40pm – 6:00pm: Un-Panel & Closing remarks

5:40pm **Un-Panel discussions & Closing remarks**

Moderated by Julien Bigot, CEA/MdlS and Bruno Raffin, Inria

Building on the different talks of the day, we will engage attendees into a discussion on the challenges of In Situ Co-Execution of HPC and DA, and how to address them as a community. The moderators will drive the discussion based on oral and written (on-line open pad) questions and remarks.

List of speakers

Achim Basermann, DLR

Dr Achim Basermann is head of the department "High-Performance Computing" at German Aerospace Center's (DLR) Simulation and Software Technology institute and German Research Foundation (DFG) review board member in computer science, topic "Massively Parallel and Data Intensive Systems". In 2019, he became chairman of the strategy commission for national high-performance computing (NHR) in Germany. He coordinated the application workpackage in the European Grid computing project NextGRID (2004-2007), the pre- and postprocessing activities in the European Exascale computing project CRESTA (2011-2014) and the algorithmic research in the Exascale computing projects ESSEX I and II (2013-2018) of DFG. In 1995, he obtained his Ph.D. in Electrical Engineering from RWTH Aachen followed by a postdoctoral position in Computer Science at Research Centre Jülich GmbH, Central Institute for Applied Mathematics. From 1997 to 2009 he led a team of HPC application experts at the C&C Research Laboratories, NEC Europe Ltd., in Sankt Augustin, Germany and contributed to the Japanese Earth Simulator project. Current research is focussed on massively parallel linear algebra algorithms, partitioning methods, optimization tools in the area of computational fluid dynamics for many-core architectures and GPGPU clusters, high-performance data analytics and quantum computing.

Leonardo Bautista-Gomez, BSC

Dr. Leonardo Bautista-Gomez is a Senior Research Scientist at the Barcelona Supercomputing Center where he work on resilience and scalability for high-performance computing and machine learning. He was awarded the 2016 IEEE TCSC Award for Excellence in Scalable Computing (Early Career Researcher). Before moving to BSC he was a Postdoctoral researcher for 3 years at the Argonne National Laboratory, where he investigated data corruption detection techniques and error propagation. Prior to that, he did his PhD. in resilience for supercomputers at the Tokyo Institute of Technology. He developed a scalable multilevel checkpointing library called Fault Tolerance Interface (FTI) to guarantee application resilience at extreme scale. For this work, he was awarded the 2011 ACM/IEEE George Michael Memorial High-Performance Computing Ph.D. Fellow at Supercomputing Conference 2011 (SC11), Honorable Mention. Before moving to Tokyo Tech, he graduated in Master for Distributed Systems from the Paris 6 University.

Julien Bigot, CEA/MdlS

Dr. Julien Bigot is a CEA Researcher at Maison de la Simulation. His main research interest is related to programming models and software engineering issues for HPC applications. During his Ph.D. Thesis (2007-2010) at IRISA Rennes and ENS Lyon, he proposed a HPC-dedicated software component model based on assembly compilation for hardware adaptation. He applied this approach to real life applications such as a MapReduce skeleton, a parallel FFT code and the extreme-scale plasma simulation code Gysela5D. Since 2014, he holds a permanent CEA researcher position at Maison de la Simulation. There, he conducts research on the separation of concern between optimization and algorithm in HPC codes. He also leads the development of the PDI library used for code modularization to support separation of concern between IO optimization and the main code as well as integration of codes in complex workflows.

Matthieu Dorier, ANL

Matthieu Dorier graduated with a PhD from Ecole Normale Supérieure de Rennes, France, In 2014, and followed with a two-year postdoc at Argonne National Laboratory. Since 2017, Matthieu is a software development specialist at Argonne. His work revolves around data management for HPC, including parallel and distributed storage, data services, workflows, in situ analysis and visualization.

Virginie Grandgirard, CEA/IRFM

Dr. Virginie Grandgirard received the PhD degree in mathematics and applications from Besancon University, France, in 1999. She then obtained the Habilitation à Diriger des Recherches in 2016. She is presently researcher with CEA, France. She is one of the main developer of the 5D non-linear gyrokinetic semi-Lagrangian code GYSELA used for plasma turbulence simulations. This code is highly parallelized up to hundreds of thousands cores. Her research interests focus on numerical methods for Vlasov equations, high performance computing and tokamak plasma turbulence. She has co-authored 50 publications in peer-reviewed journals.

Sai Narasimhamurthy, Seagate

Sai Narasimhamurthy PhD is currently Managing Principal Engineer, Seagate (formerly Lead Researcher, Emerging Tech, Xyratex) working on Research and Development for next generation storage systems and responsible for EU R&D for the Seagate Systems business. Sai currently also holds the position of vice-chair of industry for the ETP4HPC organisation and leads the storage and I/O working group for developing ETP4HPC's Strategic Research Agenda (SRA). He has also actively led and contributed to many European led HPC and Cloud research initiatives currently coordinating and providing technical leadership for SAGE and Sage2 consortia. Previously, Sai was CTO and Co-founder at 4Blox, inc, a venture capital backed storage infrastructure software company in California addressing IP SAN(Storage Area Network) performance issues.

Dirk Pleiter, FZJ & Univ. Regensburg

Prof. Dr. Dirk Pleiter is a research group leader at FZJ, professor of theoretical physics at the University of Regensburg and adjunct professor at Cyprus Institute. At FZJ he is leading the work on application-oriented technology development. Currently, he is the principal investigator of the Power Acceleration and Design Center as well as the NVIDIA Application Lab at Jülich. He has played a leading role in several projects for developing massively-parallel special purpose computers, including QPACE. He is involved in several H2020 FETHPC projects, including EXA2PRO, Maestro, and Sage2, as well as in the Human Brain Project, where he (among others) acts as technical coordinator of the ICEI project.

Charles Gueunet, Kitware

Charles Gueunet joined Kitware in February 2016. For the first three years, he worked on its PhD on the topic of "High Performance Level-set based Topological Data Analysis", and became one of the main contributors of the Topology ToolKit (TTK). After defending in February 2019, Charles joined the Scientific Visualization team at Kitware. He now works on various projects involving parallel programming, discrete geometry and data analysis algorithms.

Wouter Klijn, FZ-Julich

Wouter Klijn completed a MSc in Artificial Intelligence from the University of Groningen in the Netherlands. His Master thesis was on the information content of cell species in a 3 layer model of a cortical micro-column. He currently is a software architect in the Simlab Neuroscience at the Forschungzentrum Jülich with a focus on in Artificial Intelligence, information theory of neural networks, big data real-time streaming systems and development of complex HPC processing pipelines. He is responsible for science and use case management in the Human Brain Project, an EU Flagship Project and ICEI, the Interactive Computing E-Infrastructure for the Human Brain Project. He is currently creating the science and software infrastructure architecture for the HBP. He also works with advanced HPC oriented AI solutions and multiple neural simulators. His modelling work is focused on self-organizing dynamics of extremely large neural networks with a 2d spatial structure.

Tiago Quintino, ECMWF

Dr Tiago Quintino is a Senior Analyst and Team Leader for Development of Production Services at ECMWF.

He and his team develop high-throughput specialist software that supports ECMWF's operational meteorological forecast model, systems for acquisition of incoming observations, management of direct model output, perpetual archival of weather observations and forecast data, and post-processing, generation and dissemination of meteorological products. His team also develops cloud meteorological and climate data provisioning services (Data-as-a-Service) in support of ECMWF's cloud activities.

Dr Quintino's career spans 20 years researching numerical algorithms and developing high performance scientific software in the areas of Aerospace and Numerical Weather Prediction. Lately, his research focuses on scalable data handling algorithms for generation of meteorological forecast products, optimising their workloads and I/O of massive data-sets.

Bruno Raffin, Inria

Bruno Raffin is Research Director at Inria and leader of the DataMove team. He led the development of the FlowVR/Melissa middleware for large-scale data-flow oriented parallel applications, used for scientific visualization, computational steering, in situ analytics for large-scale parallel applications. He also worked on parallel algorithms and cache-efficient parallel data structures (cache oblivious mesh layouts, parallel adaptive sorting), strategies for task-based programming of multi-CPU and multi-GPU machines. Bruno Raffin accounts for more than 60 international publications, advice 16 PhD students. Bruno Raffin has been involved in more than 30 program committees of international conferences. He was responsible for INRIA of more than 15 national and European grants and was the co-founder of the Icatis start-up company (2004-2008). He leads the INRIA Integrated Project Lab focused on the convergence between HPC, AI and Big Data (2018-2021).

Alejandro Ribes, EDF R&D

Dr. Alejandro Ribés graduated in computer science (bachelor's and master's) from the Universitat Jaume I, Castelló (ES). He later graduated, from Université de Nice Sophia-Antipolis (FR), in a master in image processing and computer vision. Alejandro Ribés also holds a Ph.D. in multispectral imaging applied to fine art paintings, from the Ecole Nationale Supérieure des Télécommunications (FR). He later was a postdoctoral fellow at the CEA laboratory in Orsay (FR), working on parallel MRI reconstruction. During this postdoc he was appointed as a lecturer at the Computer Science Department of Ecole Polytechnique, Palaiseau, France, where he taught for two years. Alejandro also worked in MRI technology, during more than two years, as a visiting scholar at the National Yang-Ming University, Taipei, Taiwan. In 2009, Alejandro Ribés became a Research Scientist at the R&D department of EDF. In December 2016, he became Principal Research Scientist. He recently introduced AI based methods on the context of advanced numerical simulation, especially deep neural networks trained using GPU clusters. From 2013, Alejandro Ribés also collaborates with Sorbonne Université (FR).