Ref. Ares(2020)7989126 - 29/12/2020

H2020-FETHPC-3-2017 - Exascale HPC ecosystem development

EXDCI-2

European eXtreme Data and Computing Initiative - 2

Grant Agreement Number: 800957

D3.5
Second report on the organisation of WP3 workshops during EuroHPC Summit Week 2020

Final

Version: 1.0
Author(s): Mathis Bode (RWTH Aachen University), Giovanni Aloisio (CMCC)
Date: 20/12/2020

Project and Deliverable Information Sheet

EXDCI Project	**Project Ref. №: FETHPC-800957**	
	Project Title: European eXtreme Data and Computing Initiative - 2	
	Project Web Site: http://www.exdci.eu	
	Deliverable ID: **D3.5**	
	Deliverable Nature: Report	
	Dissemination Level: PU	**Contractual Date of Delivery:** 31 / December / 2020
		Actual Date of Delivery: 29 / December / 2020
	EC Project Officer: Athanasia-Charalampia Evangelinou	

* - The dissemination levels are indicated as follows: **PU** – Public, **CO** – Confidential, only for members of the consortium (including the Commission Services) **CL** – Classified, as referred to in Commission Decision 2991/844/EC.

Document Control Sheet

Document	**Title:** Second report on the organisation of WP3 workshops during EuroHPC Summit Week 2020	
	ID: D3.5	
	Version: 1.0	**Status:** ***Final***
	Available at: http://www.exdci.eu	
	Software Tool: Microsoft Word 2013	
	File(s): EXDCI-2-D3.5-v1.0.docx	
Authorship	**Written by:**	Mathis Bode, RWTH Aachen Univ. Giovanni Aloisio, CMCC
	Contributors:	Stéphane Requena, GENCI
	Reviewed by:	Maria Toth, ETP4HPC Janez Povh, PRACE & Univ. of Ljubljana
	Approved by:	MB/TB

Document Status Sheet

Version	Date	Status	Comments
0.1	30/11/2020	Draft	Before internal review
0.2	10/12/2020	Draft	After internal review
0.3	20/12/2020	Draft	After TB/MB
1.0	20/12/2020	Final version	Approved by TB/MB

Document Keywords

Keywords:	PRACE, Research Infrastructure, HPDA, Big data, AI, HPC, Workflows.

Table of Contents

List of Figures

References and Applicable Documents

[1] http://www.exdci.eu

[2] http://www.prace-project.eu

List of Acronyms and Abbreviations

AI	Artificial Intelligence
CNN	Convolutional Neural Network
CoE	Centres of Excellence for Computing Applications
CPU	Central Processing Unit
D	Deliverable
EC	European Commission
EIP	European Innovation Partnership
EU	European Union
FETHPC	Future and Emerging Technologies in High Performance Computing
FPGA	Field Programmable Gate Array
GPU	Graphics Processing Unit
H2020	Horizon 2020 – The EC Research and Innovation Programme in Europe
HEP	High Energy Physics
HTC	High Throughput Computing
HPC	High Performance Computing
HPDA	High Performance Data Analytics
IoT	Internet of Things
LES	Large-Eddy Simulation
M	Month
MS	Milestone
WP	Work Package

Executive Summary

This deliverable summarizes the most relevant information about the EXDCI-2-WP3-organized "*2nd European Communities Workshop on Exascale Computing*" with a focus this year on "*AI and HPC convergence*" [1] originally planned as physical meeting during the EuroHPC Summit Week 2020 but finally held on 26th November, 2020 as a virtual meeting due to the restrictions by the Covid-19 pandemic.

The workshop was organized in two sessions. Speakers covered a broad range from the Centre of Excellences (CoEs) and the Future and Emerging Technologies in High Performance Computing (FETHPC) projects to science and industry/vendors. In both sessions, the representatives of the different projects reported about the major challenges and issues faced, the outcomes achieved so far as well as plans for the next years, with respect to Artificial Intelligence (AI) and High Performance Computing (HPC) convergence.

The First Session was chaired by Mathis Bode (RWTH Aachen University, EXDCI-2-WP3 leader) whereas the second one was led by Giovanni Aloisio (CMCC, EXDCI-2-Task 3.2 leader).

The 12 presentations covered various aspects of AI and HPC convergence from big data processing and parallel data analysis, machine learning and AI, scientific simulations over HPC, etc.

A wrap-up discussion was also held at the end of the workshop to better clarify some aspects not addressed in previous questions and to summarize some conclusions. In this respect, the discussions and questions showed that many European communities are facing similar challenges and would be thankful for more joint efforts and combined solutions as well as networking activities like this workshop to ease experiences, results, solutions and best practices sharing. For WP3, it was important to get this feedback for improving joint work with various European communities.

Overall, the workshop was successful, with about 86 people attending the two sessions.

1 Introduction

Work Package 3 (WP3) "Excellence in HPC applications and usages" focuses on applications and best practice usage in the context of potential requirements towards Exascale platforms. This concerns classical High Performance Computing (HPC) applications but also High Performance Data Analytics (HPDA), High Throughput Computing (HTC), and Artificial Intelligence (AI). As ascertained during the EXDCI project, Europe is developing a significant fraction of the applications used in the world and the biggest producer of data. Therefore, this effort must be continued as science opportunities are evolving very quickly with the expected availability of Exascale supercomputers (and corresponding HPDA and AI tools) since: a) large research infrastructure designs are evolving with new capabilities of HPC, Big Data, and data-driven workflows; b) new tools and approaches are increasingly needed to take into account new technical realities; c) as users get a better understanding of the potential of HPC, Big Data, and data-driven applications, possibly combined, science goals evolve rapidly and new application domains appear.

For scientific applications, WP3 relies on the PRACE [2] Scientific Case and interacts with the PRACE user communities, and the HPC Center of Excellence for Computing Applications (CoEs), to focus on the Exascale aspects, as well as on the influence of specific technological or algorithmic innovations. WP3 roadmaps HPC applications and usages and coordinates with HPC user communities and CoEs.

Similar considerations apply to industrial applications, but in this case, it is often required to fit the application in a complex proprietary workflow, for instance a "digital twin" of a product or manufacturing setup. This requires engaging and working directly with the users or their collective organisations including European Innovation Partnerships (EIPs) and national initiatives. The roadmap results are of importance to avoid gaps in the value chain, but also to evaluate scenarios where "disruptive innovation" entails changes in the value chain or permit different entry strategies.

To fulfil this role within the EXDCI-2 project, WP3 is divided into three tasks:

- Task 3.1 – Roadmap of HPC applications and usages
- Task 3.2 – Engagement with HPC users' communities and CoEs
- Task 3.3 – Preparation of industrial codes to exascale

Mathis Bode (RWTH Aachen University) is WP3 leader and Stéphane Requena (GENCI/PRACE) is WP3 co-leader.

This deliverable entitled "D3.5 – Second report on the organisation of WP3 workshops during HPC Summit Week 2020", originally planned to be held in Porto at the EuroHPC Summit Week 2020, was postponed due to Covid-19 restrictions and held virtually on November 26, 2020 Relevant milestones during this period for this deliverable are:

- MS7 – 2nd brainstorming session
- MS9 – Workshop during HPC Summit Week 2020

As stated in the D3.1 "Roadmap in HPC applications and usages" issued by EXDCI and followed by the third edition of PRACE' Scientific Case, today capability and capacity (ensemble, uncertainties) multi-scale and multi-physics simulations as well as large scale scientific instruments (satellites, telescopes, sequencers, accelerators, …) and Internet of Things (IoT) devices are generating massive amount of data. In order to value this data in competitive times, scientific and industrial communities are using HPC and HPDA now supported by the use of innovative feature detection/analysis based on AI.

After a first workshop focused on how HPC and HDPA are now addressing the analysis and the management of big data, this second workshop was focused on the convergence between HPC and AI, how supercomputers users can benefit from the use of AI.

A detailed description of MS6 can be found in the deliverable "D3.2 – First Report on joint brainstorming sessions among scientific and industrial users' communities" due in M16 (postponed from M15). This deliverable provides a summary of the second workshop organized by WP3. The aimed target of the workshop is explained and details about different contributions to the workshop are summarized. Conclusions with respect to the focus topic of the workshop are given.

2 Summary of Deliverable

This delivery summarizes the organization of the WP3-organized workshop originally planned during the EuroHPC Summit Week 2020, but postponed to November 2020 due to Covid-19 restrictions. The description is separated into three parts:

- Summary of the Workshop (cf. Chapter 3)
- Summary of first session (mainly CoE presentations) (cf. Chapter 4)
- Summary of second session (mainly FETHPC presentations) (cf. Chapter 5)

The report finishes with some conclusions and a summary of next steps.

3 Summary of the Workshop

WP3 organized a workshop with the target of discussing important Exascale topics among various scientific communities. The workshop was designed as the continuation of a series of workshops and therefore denoted as "2nd European Communities Workshop on Exascale Computing ". In order to support fruitful discussions, AI and HPC convergence was chosen as the target topic, and **12 speakers** were invited to present their ideas/experiences/solutions/questions with respect to this topic.

Due to the importance of getting a detailed idea about the entire European Exascale landscape, the EXDCI-2-WP3 workshop was organized in close collaboration with Guy Lonsdale from FocusCoE. Finally, speakers covering a broad range from CoEs and FETHPC projects to science and industry/vendors were selected. A first session mainly featured speakers from CoEs and a second session focused on speakers mainly representing different FETHPC projects. The agenda of the workshop is shown in Figure 1 and Figure 2.

The virtual workshop took place on Thursday, 26th November, 2020 from 2:30pm to 6:00pm. About **86 people attended the workshop** in total, which was a success. The presentation covered various aspects of AI and HPC convergence, from Big Data processing and parallel data analysis, to provenance and reproducibility of simulation results. The discussions and questions showed that many European communities face similar challenges and would be thankful for more joint efforts and combined solutions. For WP3, it was important to get this feedback, to improve the collaboration with various actors of the European ecosystem.

The program is shown in Fig. 1 and 2. A snapshot of some participants during the meeting is presented in Fig. 3.

Virtual Workshop on AI and HPC convergence

organized by

EXDCI-2 (European eXtreme Data and Computing Initiative - Phase 2)

November 26th, 2020

Agenda November 26th, 2020

14:30	**Welcome and brief introduction**	
14:30-16:00	**Session 1 – AI and HPC Convergence**	**Chair: Mathis Bode**
14:30-14:45	HPDA for Biomolecular Research: The BioExcel Case	Rosa Badia (BSC)
14:45-15:00	Towards the next level of use of HPC in Engineering	Amgad Dessoky (HLRS)
15:00-15:15	Efficient CNN for space data classification on EuroEXA multi-FPGA platform	Iakovos Mavroidis (ICS-FORTH)
15:15-15:30	An HPC-enabled Data Science and Learning Environment for Climate Change Experiments at Scale	Donatello Elia (CMCC/ESiWACE)
15:30-15:45	Elastic Ensemble Run Data Processing with Melissa	Bruno Raffin (INRIA)
15:45-16:00	Towards the use of HPC for HEP workflows	Maria Girone (CERN OpenLab)

Figure 1 Agenda of the 2nd European Communities Workshop on Exascale Computing (1/2)

16:00-16:30	*Coffee break*	
16:30-18:00	**Session 2 – AI and HPC Convergence**	**Chair: Giovanni Aloisio**
16:30-16:45	Maestro: Orchestrating Data for HPC (and AI) Applications	Dirk Pleiter (Jülich Superc. Centre)
16:45-17:00	The Combination of Real-Time Data, HPC, and Interactive Visualization in the VESTEC project	Max Kontak (DLR)
17:00-17:15	ASPIDE: A data-oriented programming model for AI/HPC Convergence	Jesus Carretero (UC3M)
17:15-17:30	SAGE - Object Storage Platform for HPC & AI	Sai Narasimhamurthy (Seagate Technology, LLC)
17:30-17:45	Subgrid Modeling of Turbulent Reactive Multiphase Flows Using Physics-Informed Enhance Super-Resolution Generative Adversarial Networks	Mathis Bode (RWTH Aachen Univ.)
17:45-18:00	The New HPC: Accelerating AI, HPC and Visualisation	Tim Lanfear (NVIDIA)
18:00	**Wrap up and closing session**	**Stephane Requena**

Program Committee

Giovanni Aloisio (CMCC & University of Salento)	*giovanni.aloisio@unisalento.it*
Stephane Requena (GENCI)	*stephane.requena@genci.fr*
Mathis Bode (RWTH Aachen University)	*m.bode@itv.rwth-aachen.de*
Sandro Fiore (University of Trento)	*sandro.fiore@unitn.it*
Stefan Krieg (FZ-JUELICH)	*s.krieg@fz-juelich.de*
Jean-Claude André (Jca Consultance and Analysis)	*jc_andre@sfr.fr*
Guy Lonsdale (SCAPOS)	*guy.lonsdale@scapos.com*

This event is funded by EXDCI-2: the EXDCI-2 project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 800957 http://www.exdci.eu

Figure 2 Agenda of the 2nd European Communities Workshop on Exascale Computing (2/2)

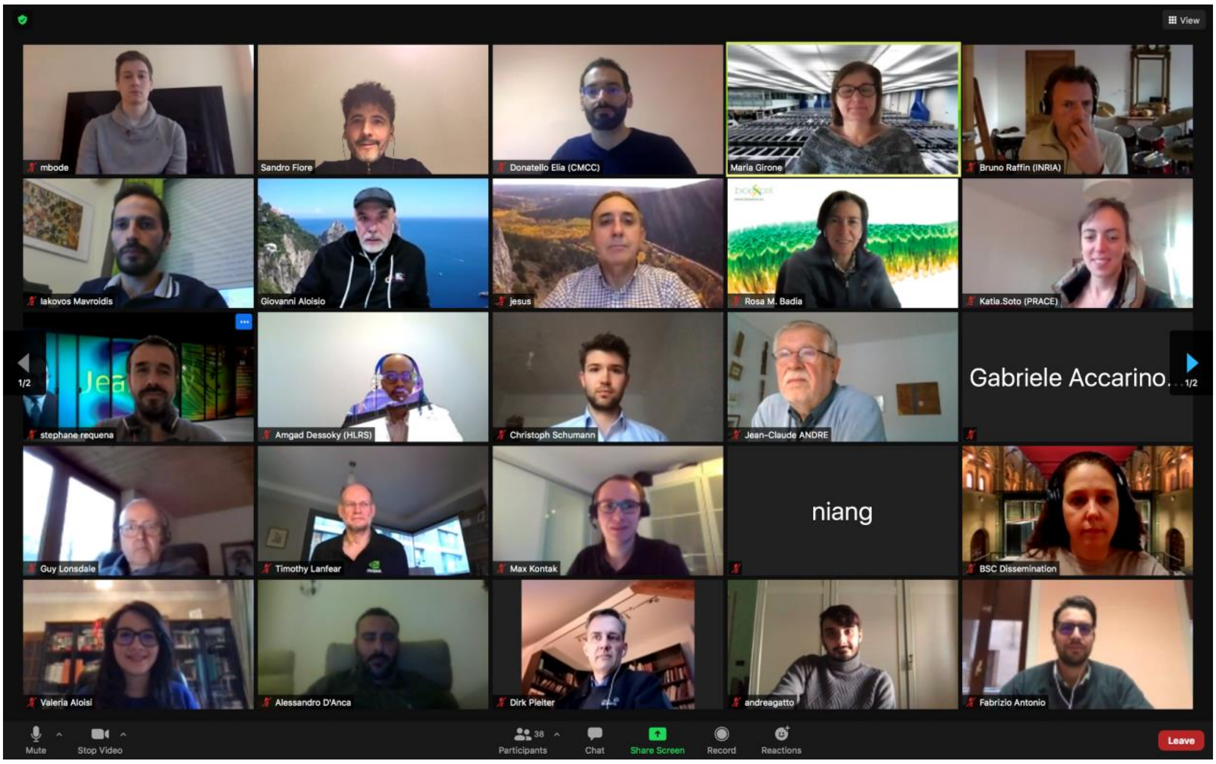

Figure 3 Overview of remote participants using the Zoom session

4 Session I

The next subsections present the Session I contributions regarding HPC and AI convergence.

4.1 HPDA for Biomolecular Research: The BioExcel Case

Speaker: Rosa M. Badia (BSC)

Rosa is the manager of the Workflows and Distributed Computing group at the BSC. Her current research interests include programming models for distributed computing platforms and its integration with novel storage technologies. Her group is developing the PyCOMPSs/COMPSs programming model, with the ambition of providing an environment able to build large workflows that combine HPC and HPDA.

Abstract

BioExcel is the European Centre of Excellence for provisioning support to academic and industrial researchers in the use of HPC and HTC in biomolecular research. The BioExcel was established to provide the necessary solutions for long-term support of the biomolecular research communities: fast and scalable software, user-friendly automation workflows and a support base of expert core developers. The main services offered by the center include hands-on training, tailored customization of code and personalized consultancy support. The talk will first introduce a quick overview of the BioExcel activities, and then will present the progress on the design and development of reproducible computational biomolecular workflows tackling real scientific problems. These workflows are based on the biobb library, powered by PyCOMPSs, and will combine HPC and Data Analytics.

Talk highlights

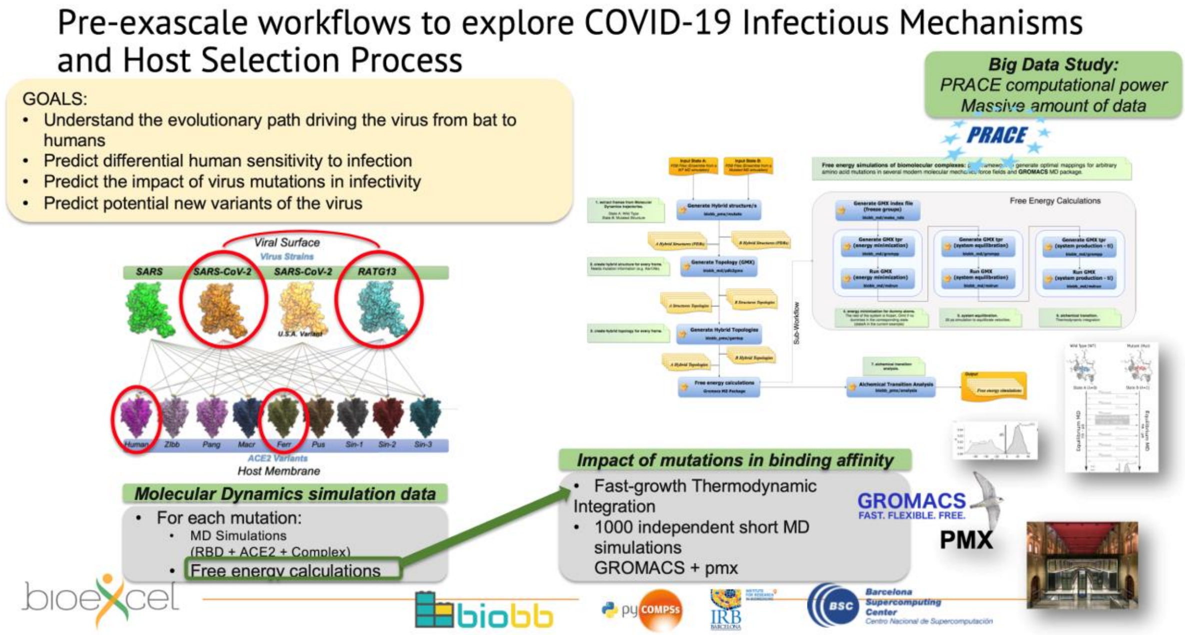

Figure 4 A pre-exascale workflow to explore COVID-19 infectious mechanisms

4.2 Towards the next level of use of HPC in Engineering

Speaker: Amgad Dessoky (HLRS)

Amgad Desssoky is the technical manager of EXCELLERAT Centre of Excellence for Engineering through the High Performance Computing Centre Stuttgart, before that he was a PhD researcher in institute of Aerodynamics and gas dynamics at University of Stuttgart over four years, thesis in Aerodynamics and Aeroacoustics of wind turbines was submitted and he was also ANSYS product manager for more than three years, and he is coming from a mechanical power engineering background.

Abstract

Engineering applications will be among the first to exploit Exascale, not only in academia but also in industry. For this reason, the EXCELLERAT activity brings together European experts to establish a CoE in Engineering Applications on HPC with a broad service portfolio, paving the way for the evolution towards Exascale. To fulfill its mission, EXCELLERAT will focus its developments on six carefully chosen references, which were analyzed on their potential to achieve Exascale performance in HPC for engineering. In the current talk, we will give an overview of EXCELLERAT objectives, approach, services, and AI activities within the project.

Talk highlights

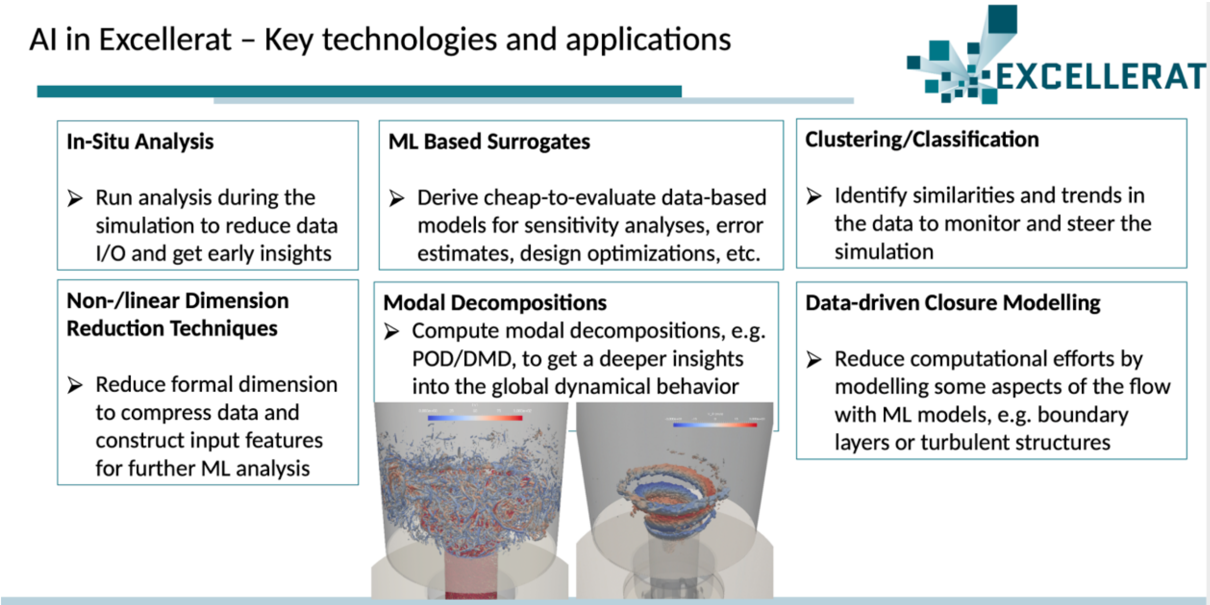

Figure 5 Key technologies and applications in EXCELLERAT to enable AI

4.3 Efficient CNN for space data classification on EuroEXA multi-FPGA platform

Speaker: Iakovos Mavroidis (ICS-FORTH)

Dr. Iakovos Mavroidis is a member of Foundation for Research and Technology – Hellas (FORTH). He received his M.Sc. degree from U. C. Berkeley in 2001 and his Ph.D. from the Technical University of Crete in Greece in 2011. He worked at Sun Microsystems in 2000 and he was with MIPS Technologies in 2001. From 2011 to 2016 he was a Teaching Associate at the University of Crete, Greece. He was also the project coordinator of the RAPID and ECOSCALE H2020 projects and he has participated in several other EU projects.

Abstract

Energy-efficient high-performance implementation of Convolutional Neural Networks (CNNs) directly in hardware can be a key enabling factor for future demanding machine learning applications. In this presentation, we exploit the implementation of a CNN on the EuroEXA Field Programmable Gate Array (FPGA)-based platform and explore different ways to minimize the impact of various hardware restrictions to improve performance. Our driving use case is a satellite-based remote sensing platform on which signal processing and classification tasks, with strict bandwidth and energy limitations, take place. We compare our results in terms of throughput, latency and energy against competing technologies such as Graphics Processing Units (GPU) and we demonstrate similar latency and throughput results while achieving an order of magnitude energy savings.

Talk highlights

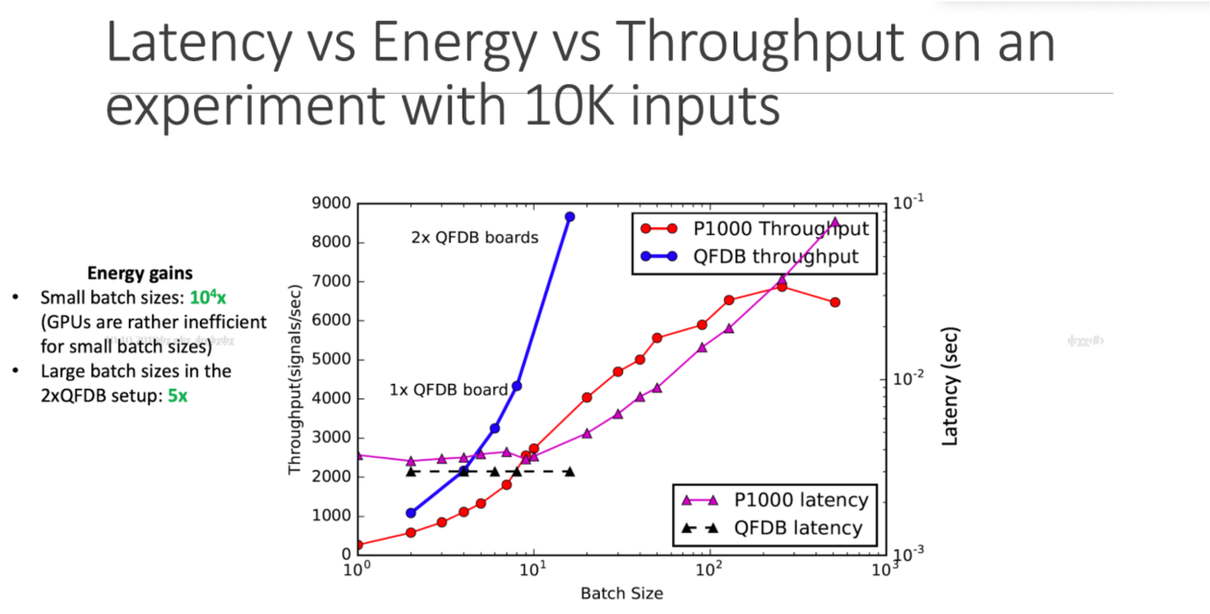

Figure 6 Latency vs Energy vs Throughput on an experiment with 10K inputs in the EuroEXA multi-FPGA platform

4.4 An HPC-enabled Data Science and Learning Environment for Climate Change Experiments at Scale

Speaker: Donatello Elia (CMCC)

Donatello Elia is a Junior Scientist for the Advanced Scientific Computing (ASC) division at the Euro-Mediterranean Center on Climate Change (CMCC) Foundation, which he joined in 2013. He holds a M.Sc. degree in Computer Engineering (2013) from the University of Salento in Italy, where he is currently pursuing a PhD degree in Engineering of Complex Systems. His main research interests include high performance data management, scientific data analysis, in-memory analytics and big data. Since 2018, he has been the Co-P.I. of the Ophidia project, a research effort on High Performance Data Analytics for eScience. He has been involved in various EU projects, such as EUBrazilCC, EUBra-BIGSEA, INDIGO-DataCloud, EOSC-Hub, IS-ENES3 and ESiWACE2. He has authored and co-authored various papers in refereed journals and international conference proceedings and is also a member of the IEEE Computer Society.

Abstract

Data science was envisioned by Jim Gray as a fourth paradigm of science. Also known as data-driven science, data science employs techniques from many fields within the context of mathematics, statistics, information science, and computer science. In many domains (e.g., life sciences, environmental science, and astrophysics) the different dimensions of data (volume, variety, velocity, veracity, and complexity) pose challenges that must be properly handled to efficiently address data-intensive scientific discovery at large-scale.
Climate is now a data problem and data science can be of great help in facing the scientific data challenges of this new era. In such a context, there is a strong interest in combining climate science with approaches from statistics, data analytics, big data, machine learning, HPC and data mining to comprehensively address knowledge extraction and actionable insights.

This contribution presents the main goal, scientific challenges and the first implementation of the Data Science environment developed at the Euro-Mediterranean Center on Climate Change as well as some insights related to the implementation of some representative climate change case studies, joining machine learning and HPC. A key component of the full stack is the Ophidia framework, which enables HPDA workflows at scale.

Talk highlights

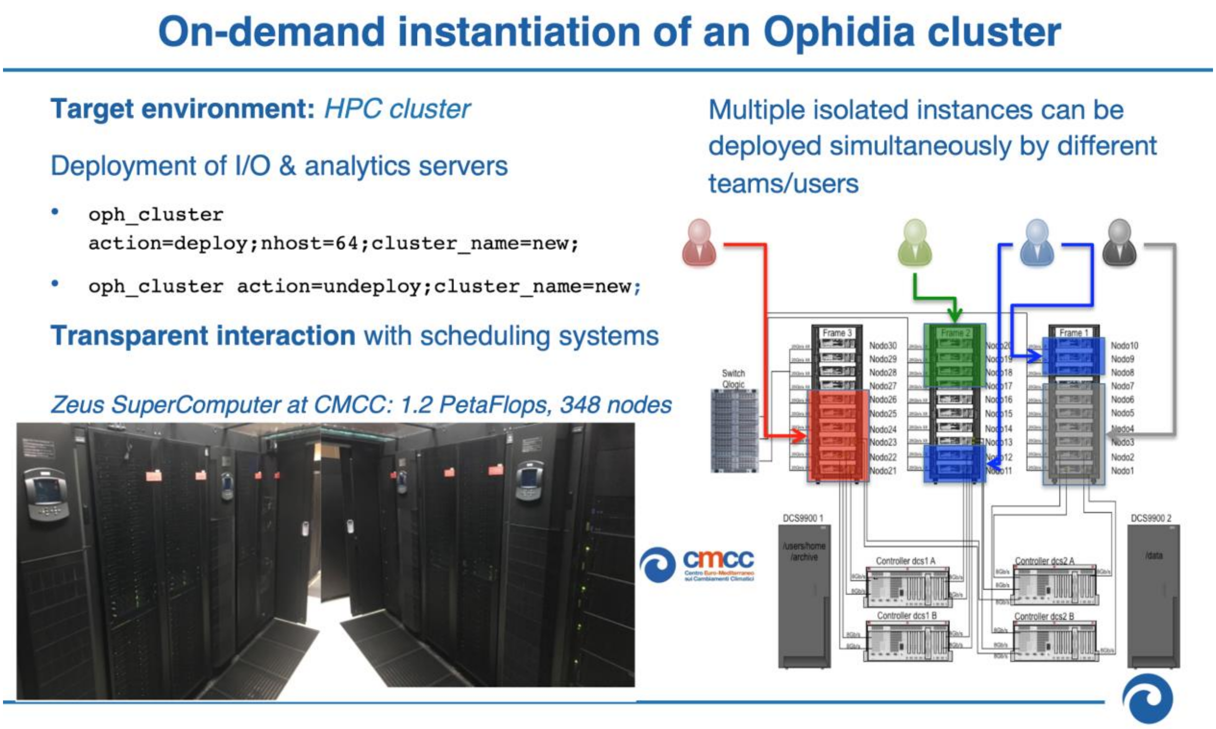

Figure 7 On-demand instantiation of an Ophidia cluster for HPDA in the HPC-enabled data Science environment at the Euro-Mediterranean Center on Climate Change

4.5 Elastic Ensemble Run Data Processing with Melissa

Speaker: 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 (Central Processing Unit) and multi-GPU machines. Bruno Raffin accounts for more than 60 international publications, advising 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. He leads the INRIA Integrated Project Lab focused on the convergence between HPC, AI and Big Data (2018-2021).

Abstract

Ensemble runs consist in running a given numerical simulation several times with different input parameters to sample the simulation behavior in the parametric space. Ensemble runs are performed for uncertainty quantification or data assimilation for instance, and its importance is growing with the availability of super large scale computers and the need to couple numerical simulations with data provided with other scientific instruments. But the amount of data produced when running many large scale simulations is prohibitive to store. In this talk we will present Melissa, a file avoiding, adaptive, fault tolerant and elastic framework designed to handle very large ensemble runs coupled with on-line data processing. We will present how, in the context of the AI and HPC convergence, Melissa can be leveraged for global sensitivity analysis, data assimilation and training of large scale surrogate models

Talk highlights

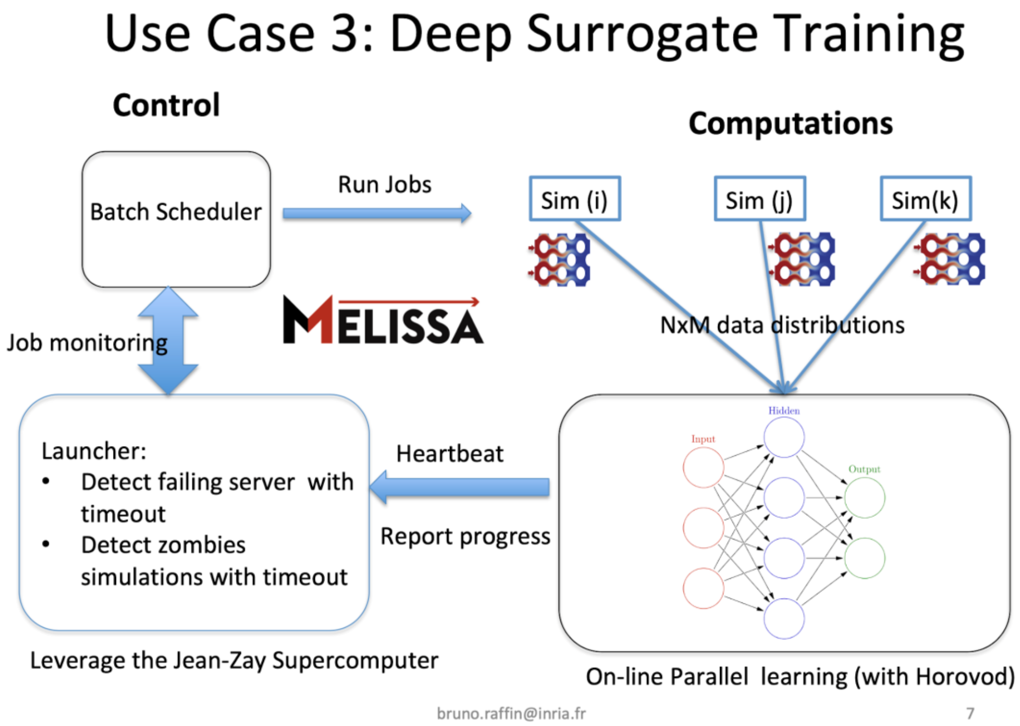

Figure 8 A large-scale ensemble use case for Deep Surrogate Training based on Melissa

4.6 Towards the use of HPC for HEP workflows

Speaker: Maria Girone (CERN OpenLab)

Maria Girone graduated from the University of Bari, Italy. She was awarded her Ph.D. in high-energy physics in 1994. Maria became a research fellow with the ALEPH experiment performing analysis and serving as accelerator liaison. Later, Maria developed detector hardware for the LHCb experiment as a research associate at Imperial College London. In 2002, Maria joined the IT Department as an applied scientist and CERN staff member. In 2004, Maria was appointed as coordinator of the Oracle database services for the LHC experiments. In 2009, she was appointed deputy group leader of the CERN IT Experiment Support group. In 2012, Maria became the founding chair of the WLCG Operations Coordination team, responsible for the core operations and commissioning of new services in the WLCG. In 2014, she was appointed the Computing Coordinator for the CMS Experiment at CERN for two years. As coordinator, Maria was responsible for 70 computing centers on five continents and more than 100 FTE of effort yearly to archive, simulate, process and serve petabytes of data. In 2016, Maria joined the management team of CERN openlab, taking over the position of CTO as of January 2016. In her role as CTO, Maria is managing the overall technical strategy of CERN openlab plans towards R&D in computing architectures, HPC and AI, in collaboration with the LHC experiments for the upgrade programs for software and computing, promoting opportunities for collaboration with industry.

Abstract

In this talk I will give an overview of the challenges in computing expected at CERN over the next decade. By 2027, the Large Hadron Collider will be upgraded to the High-luminosity program, bringing the computing and storage needs towards the exascale. I will discuss the approaches we are considering in order to handle this enormous amount of data, the use of heterogeneous architectures and the crucial role of HPC and AI.

Talk highlights

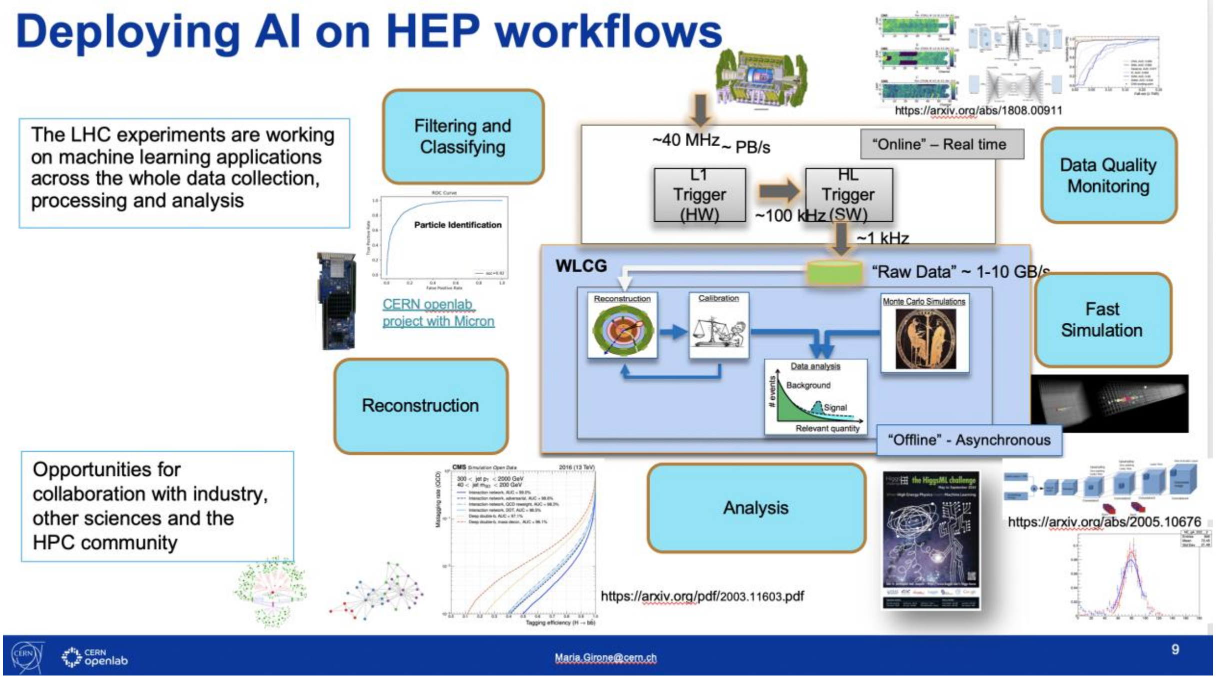

Figure 9 A workflow use case integrating AI from the High Energy Physics (HEP) community

5 Session II

The next subsection presents the Session II contributions regarding HPC and AI convergence.

5.1 Maestro: Orchestrating Data for HPC (and AI) Applications

Speaker: Dirk Pleiter (Jülich Supercomputing Centre)

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. He has played a leading role in several projects for developing massively-parallel special-purpose computers, including QPACE. He is involved in a leading role in several H2020 FETHPC projects, including EXA2PRO, Maestro, and Sage2, and he acts as technical coordinator of the ICEI/Fenix project.

Abstract

The Maestro project started with the ambition of designing and implementing a middleware that allows for data- and memory-aware implementation of HPC and AI applications. In this talk, we will report on Maestro's co-design process, the resulting design and its implementation status.

Talk highlights

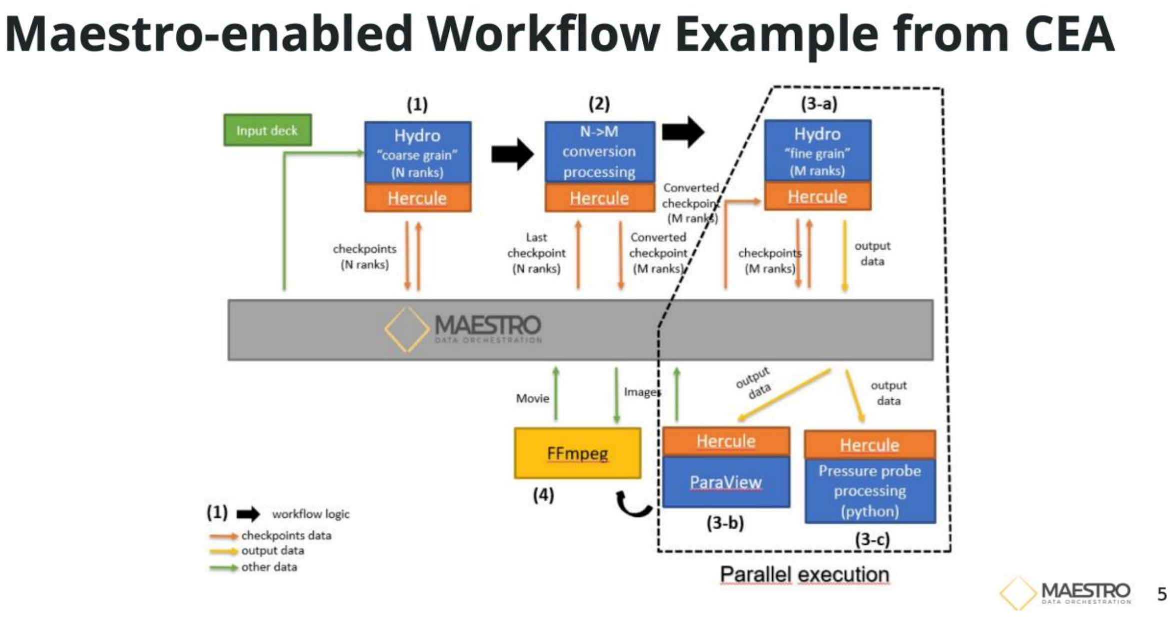

Figure 10 Example of a MAESTRO workflow in cosmology coupling simulation and in-transit post processing and visualization of the data

5.2 The Combination of Real-Time Data, HPC, and Interactive Visualization in the VESTEC project

Speaker: Max Kontak (DLR)

Dr. Max Kontak is a research scientist at the German Aerospace Center (DLR) in Cologne, working in the department for HPC. His current research activities include simulation methods for rotorcraft as well as both the data analytics and performance engineering aspects of the EU FETHPC project VESTEC.

Abstract

Traditionally, HPC has been used to simulate disastrous events such as wildfires, spread of diseases, or solar outbursts after the event, typically for post-disaster analysis. However, with the increasing availability of high-velocity sensor data and computational resources, as well as the development of elaborate in-situ data analytics and visualization techniques it is now possible to support urgent decision makers in real-time utilizing HPC infrastructure.
In this talk, we will present the approaches that are employed in the H2020 FETHPC project VESTEC (Visual Exploration and Sampling Toolkit for Extreme Computing) to tackle the challenges that arise when combining real-time data with HPC infrastructure and interactive visualization, both regarding technology and policies. Whilst the challenges are significant, so are the potential benefits of overcoming them, not only to the HPC community but also to society as a whole.

Talk highlights

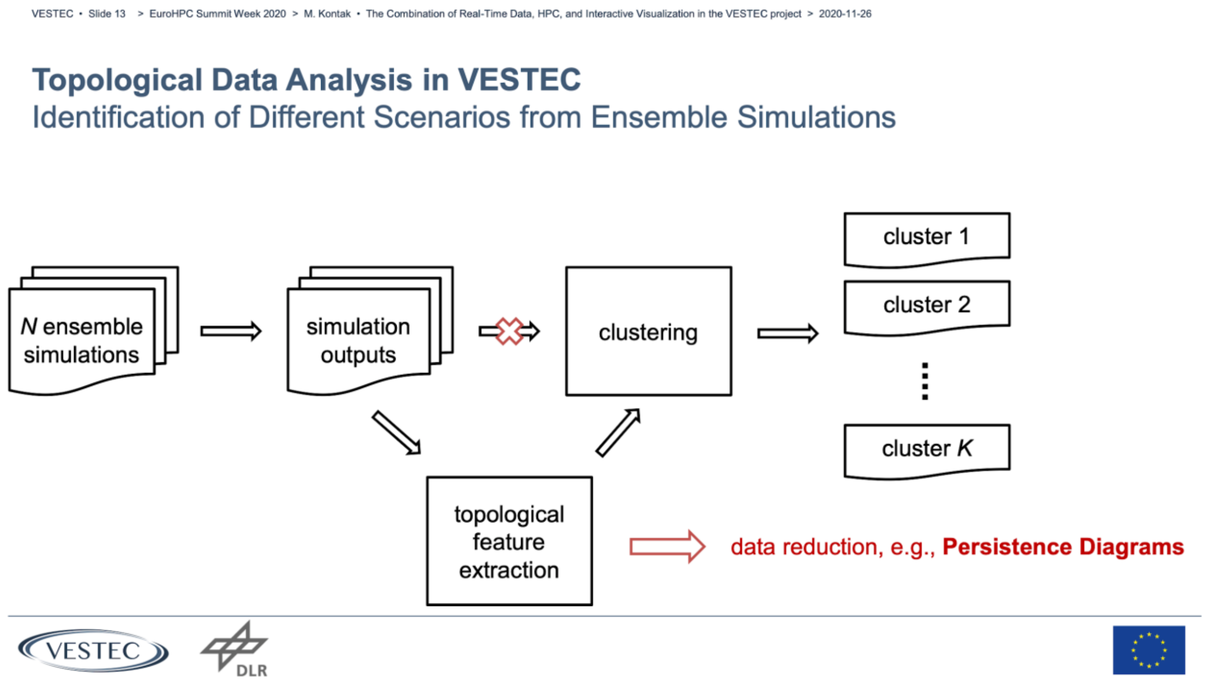

Figure 11 Typical workflows proposed by VESTEC for extracting and identifying using topological feature extraction and clustering techniques pertinent data structures

5.3 ASPIDE: A data-oriented programming model for AI/HPC Convergence

Speaker: Jesus Carretero (UC3M)

Jesus Carretero is a Full Professor of Computer Architecture and Technology at Universidad Carlos III de Madrid (Spain). His research activity is centered on high-performance computing systems, large-scale distributed systems, data-intensive computing, IoT and real-time systems. He was Action Chair of the IC1305 COST Action “Network for Sustainable Ultrascale Computing Systems (NESUS)”, and he is also currently involved in the H2020 FET-HPC program ASPIDE. He has participated and led several national and international research projects in these areas, founded by the Madrid Regional Government, Spanish Education Ministry and the European union. Prof. Carretero has published more than 250 papers in journals and international conferences, editor of several books of proceedings, and guest editor for special issues of journals as International Journal of Parallel Processing, Cluster Computing, and Future Generation Computer Systems, and he is coauthor of several text books related to Operation Systems and Computer Architecture. He has participated in many conference organization committees, and he has been General chair of HPCC 2011, ICPP 2016 and CCGRID 2017, and Program Chair of ISPA 2012, EuroMPI 2013, C4Bio 2014, and ESAA 2014. Prof. Carretero is a senior member of the IEEE Computer Society and member of the ACM.

Abstract

Extreme Data is an incarnation of Big Data concept distinguished by the massive amounts of data that must be queried, communicated and analyzed in (near) real-time by using a very large number of memory/storage elements and Exascale computing systems. Immediate examples are the scientific data produced at a rate of hundreds of gigabits-per-second that must be stored, filtered and analyzed, the millions of images per day that must be mined (analyzed) in parallel, the one billion of social data posts queried in real-time on an in-memory components database. Following the need of improvement of current concepts and technologies, ASPIDE project is focusing on data-intensive applications running on very large-scale computing elements to develop a data-oriented programming model for AI/HPC convergence. ASPIDE is providing a definition of new programming paradigms, APIs, runtime tools and methodologies for expressing data-intensive tasks on Exascale systems, which can pave the way for the exploitation of massive parallelism in data analysis tasks.

Talk highlights

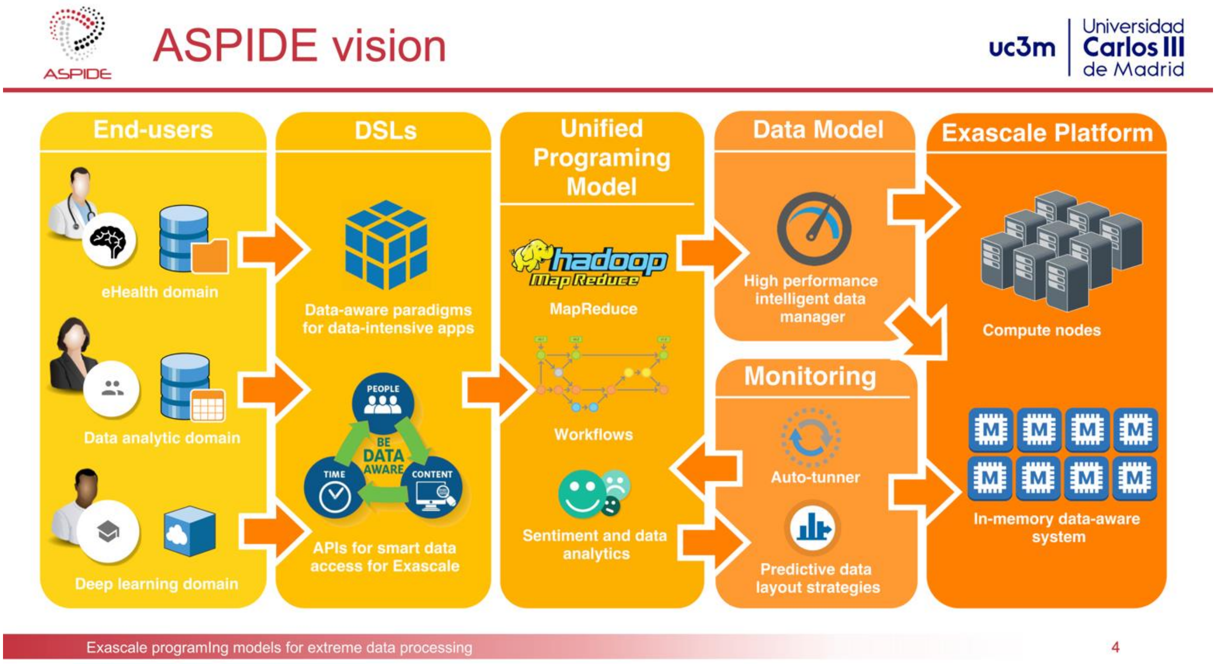

Figure 12 Vision of the ASPIDE project for extreme data processing

5.4 SAGE - Object Storage Platform for HPC & AI

Speaker: Sai Narasimhamurthy (Seagate Technology, LLC)

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.

Abstract

The landscape for storage for HPC is changing and existing parallel file paradigms which are designed for the petascale era are no longer suitable. There is a need to fundamentally change the software abstractions driving the storage infrastructure hardware especially with the arrival of newer non-volatile memories, which is expected to be part of the storage hierarchy for HPC along with Flash, and disk drives. Further we need to understand new use cases, and the storage and I/O needs of AI which is not very clear at this moment as it exploits the HPC infrastructure at extreme scales.
This talk will discuss and provide updates of the Sage2 (Percipient Storage for Exascale Data Centric Computing2) FETHPC project which addresses the above issues.

Talk highlights

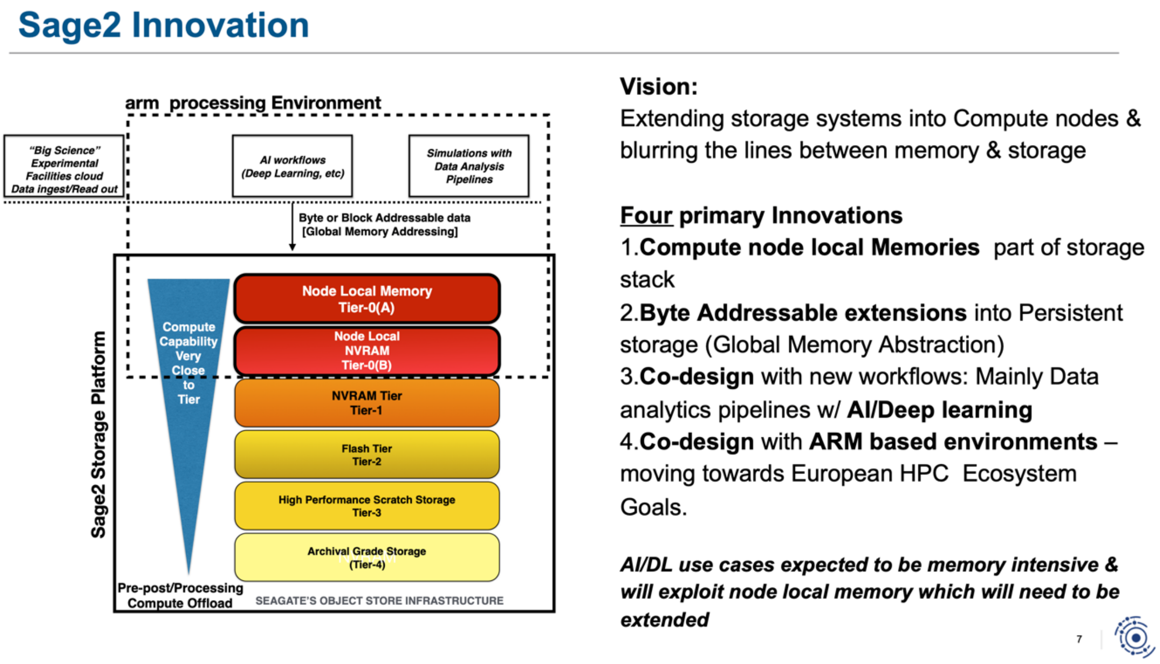

Figure 13 Innovation of Sage2 project towards memory intensive applications

5.5 Subgrid Modeling of Turbulent Reactive Multiphase Flows Using Physics-Informed Enhance Super-Resolution Generative Adversarial Networks

Speaker: Mathis Bode (RWTH Aachen University)

Mathis Bode works at RWTH Aachen University on problems related to HPC and AI with their application to multiphase flows and combustion. His research has been awarded multiple times and he contributed to various European projects.

Abstract

This talk shows examples how DL is used in the context of subgrid modeling for Large-Eddy Simulation (LES) of reactive/interfacial flows. Main focus is on the recently developed physics-informed enhanced super-resolution generative adversarial network. The performance in a priori and a posteriori tests is discussed, and the application to a multi-physics, multi-scale problem shown. Finally, benefits and shortcomings of AI compared to classical models is emphasized.

Talk highlights

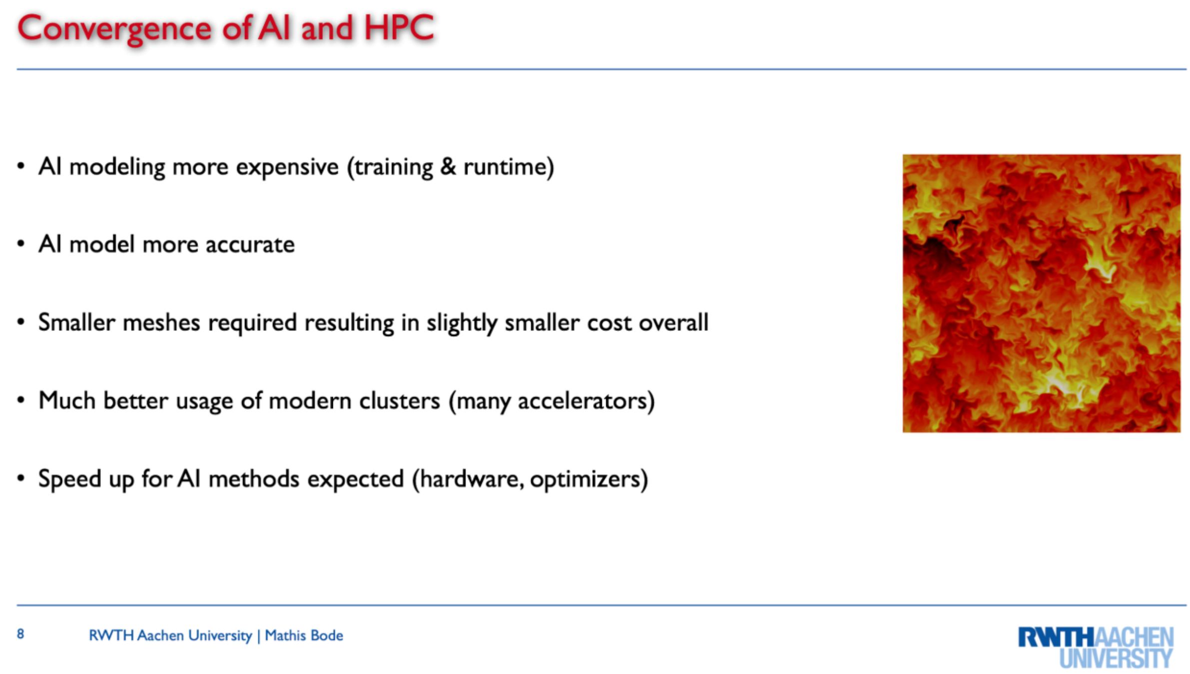

Figure 14 Conclusion on the use of AI based methods coupled to traditional LES simulations applied to reactive multiphase flows (in combustion for example)

5.6 The New HPC: Accelerating AI, HPC and Visualization

Speaker: Tim Lanfear (nVIDIA)

Timothy Lanfear manages the European solution architecture and engineering team in NVIDIA's Enterprise Solutions Group. He has twenty-five years' experience in HPC, starting as a computational scientist in British Aerospace's corporate research centre, and then moving to technical pre-sales roles with Hitachi, ClearSpeed, and most recently NVIDIA. He has a degree in Electrical Engineering and a PhD for research in the field of graph theory, both from Imperial College London. He is a Chartered Engineer and Member of the Institution of Engineering and Technology.

Abstract

The programmable GPU architecture designed for computer graphics has proved itself remarkably well-suited to the requirements of modern high-performance computing by unifying the requirements of AI, scientific simulation and scientific visualization in a single accelerated computing platform. Combining the two topics of HPC and AI in accelerated applications running on GPUs has been a fruitful endeavor leading to several finalists in the Gordon Bell prize. The ecosystem continues to grow with the recent full support for ARM processors hosting GPUs and accelerated I/O directly from GPU to storage systems.

Talk highlights

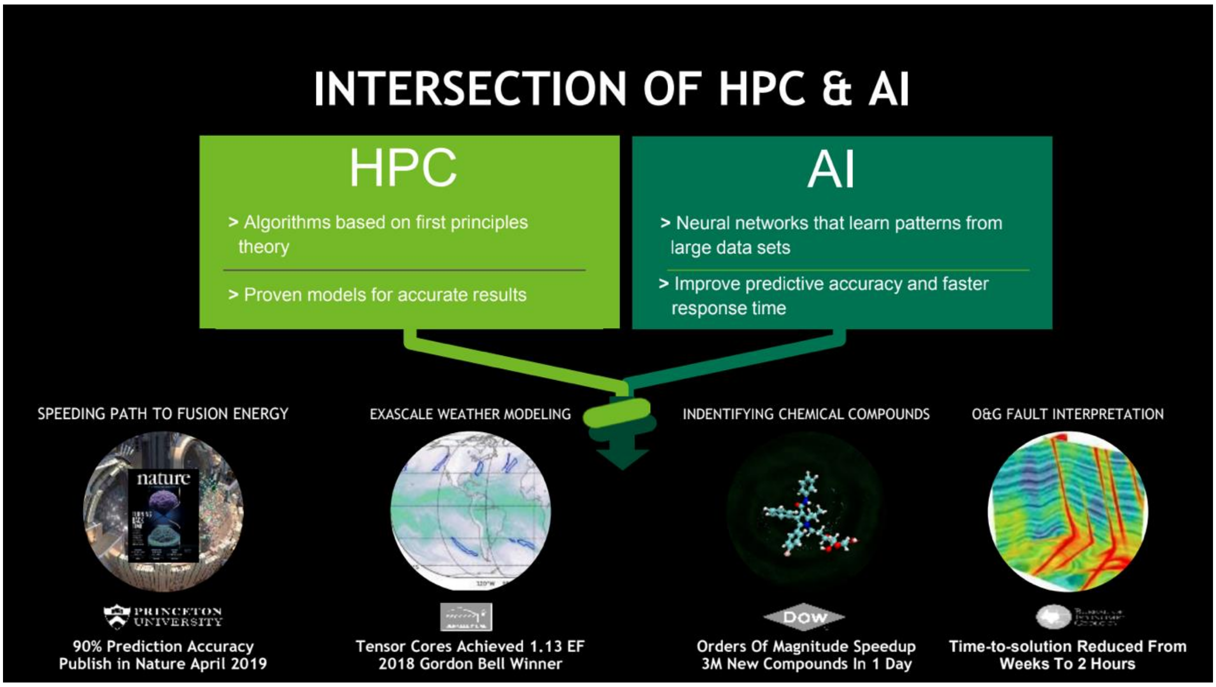

Figure 15 Illustration provided by nVIDIA on the convergence between HPC and AI with several key examples provided by customers

6 Conclusions and Next Steps

This deliverable summarizes the most relevant information about the WP3-organized "*2nd European Communities Workshop on Exascale Computing*" with a focus this year on "*HPC and AI Convergence*" held virtually on 26th November, 2020.

A total of 86 people attended virtually the 2 sessions, coming from end users communities (CERN and RWTH Aachen Univ) and CoE (EXCELERAT, BioExcel, ESIWACE and EoCoE), FET-HPC projects (EuroEXA, Maestro, VESTEC, Sage) to HPC vendors (nVIDIA).

The presentations given during the workshop highlighted that many modern workflows not only feature HPC and AI but also HPC, HTC, HPDA and AI at the same time. Thus, all of these methods must be considered when improving the European HPC landscape.

The discussions and questions showed that this integration of various methods is happening in Europe with concrete examples in end users applications (from climate, high energy physics, combustion, smart cities, medicine and health, astrophysics, …), new usages (processing of large scale scientific instruments, scaling of AI models, urgent and interactive capability and capacity computing and visualisation) served by converged middleware for smart schedulers, support of end-to-end workflows, data management, in-situ/in-transit post processing, security, new programming models to object data storage.

For WP3, it was important to get this feedback, for roadmapping strategies about HPC and AI in the future.