EUROPEAN HIGH PERFORMANCE COMPUTING
JOINT UNDERTAKING

Research & Innovation Advisory Group

Strategic Research and Innovation Agenda 2019

Date:	05/04/2019
Version:	2.4
Authors:	EuroHPC RIAG

Table of Contents

Document History

Version	Date	Comment
1.0	20/03/2019	Document created from the EuroHPC orientation paper
2.0	27/03/2019	Document updated with context drawn from these sources: https://www.etp4hpc.eu/pujades/files/SRA%203.pdf http://bdva.eu/sites/default/files/BDVA_SRIA_v4_Ed1.1.pdf https://www.etp4hpc.eu/pujades/files/bigdata_and_hpc_FINAL_20Nov18.pdf
2.1	27/03/2019	Changed document formatting
2.2	28/03/2019	Edited text and added section on Conclusions and Outlook
2.3	29/03/2019	Corrections and final edits
2.4	05/04/2019	Final version after editing during the 2nd RIAG meeting

1 INTRODUCTION

The EuroHPC Joint Undertaking supports an ambitious research and innovation agenda to develop and maintain in the Union a world-class high-performance computing (HPC) ecosystem towards exascale and beyond, to strengthen competitiveness and to help create new markets and societal applications. It will also promote the uptake and systematic use of research results generated in the Union by users from science, industry, including SMEs, and the public sector. These objectives of improving EU's capabilities in HPC will contribute to the digitisation of industry and the transition towards a European data economy overall and in line with the EU policy priorities on the Digital Single Market.

The Petascale to Exascale transition is very complex for several reasons. Firstly, architectural heterogeneity and specialisation, new memory technologies, storage and interconnects challenge application development. Secondly this transition is happening at a time when a data deluge is taking place. An important challenge is how to deal with the data generated by sensors as well as the numerical simulations themselves. The convergence of extreme data and computing with new considerations such as edge computing, in-transit computing, etc. come with many new capabilities for doing science and industrial applications (e.g. AI technologies), connection to data driven AI, and with an enlargement of the HPC ecosystem that will shape the future.

The Multiannual Strategic Research and Innovation Agenda will identify research and innovation priorities for the development and adoption of technologies and key competences for HPC and Big Data. It will cover all scientific and industrial value chain segments, including low-power processor and middleware technologies, algorithms and code design, services and engineering, interconnections, up to system architectures for the next generations of supercomputing. It will support the integration of technologies into supercomputing systems through a co-design and piloting approach to test the readiness of the European technology projects, vendors and users to produce a globally competitive HPC system. Finally, it will highlight the importance of achieving excellence in HPC applications through re-engineering of codes and application workflows with scientific, industrial and societal impact and supporting emerging HPC-enabled lead-market applications. The Strategic Research and Innovation Agenda 2019 is the first in the realisation of these ambitious goals.

The Strategic Research and Innovation Agenda addresses aspects of training and science education of the next generation, skills development and services to industry, illustrating the potential of HPC through industrial use cases creating business value to help Europe compete in this complex market. Customized training programs should be developed for the specific needs of the different stakeholders, in particular SMEs.

The Multiannual Strategic Research and Innovation Agenda shall be reviewed regularly in accordance with the evolution of the scientific and industrial demands.

2 TOWARDS EXTREME SCALE TECHNOLOGIES AND APPLICATIONS

The support for a sustainable extreme-scale HPC ecosystem in Europe requires mastering the R&D process with a co-design approach and a holistic view on the technology supply, hardware, software stack, and applications. The goal is to improve time and energy to solution, robustness, reliability, portability, maintainability and productivity on upcoming exascale and extreme performance computing capabilities for scientific, industrial or societal challenges.

Europe has a particularly strong role in scientific and industrial applications with a large fraction of the applications used in the world being originally developed in Europe. While different application areas have specific challenges, there are a number of cross-cutting issues.

Portability and maintainability are a key concern for applications, also when approaching the extreme scale. It is therefore mandatory to have well-defined, stable, and portable interfaces, where possible relying on international standards, particularly when it comes to new developments such as heterogeneous architectures, deeper memory hierarchies, and new networking concepts. Furthermore, there is a need for improved software engineering practices including testing for both correctness and performance, collaborative software development, benchmarking and validation, and code re-factoring and modernisation.

The increased complexity of future HPC architectures is making productive coding and performance tuning an even more difficult task. Novel approaches, e.g. Domain Specific Languages (DSL) relying on smart underlying system software layers, can help abstracting the complexity of future HPC architectures. Scaling towards the extreme scale will increase the data requirements drastically. Applications will need high bandwidth memory and networks, as well as new techniques to deal with data in a more energy- and performance-efficient way, such as in-situ and in-transit post processing coupled with the full range of AI techniques e.g. for automatic features detection/tracking.

Knowing the hardware and software constraints brought by extreme-scale architectures, the need to optimise both performance and energy consumption and the vast number of applications to consider, it is important to develop low overhead, accurate (high resolution) and possibly automatic sampling tools, able to profile performance, power consumption and data movement in a scalable way. Impact on application performance should be minimised. Information collected could feed a central database of thermal and performance signatures allowing the user to know exactly the behaviour of their application, using performance analysis and decision-making tools to adapt at run-time in an automated way.

The increased size of extreme-scale systems will likely increase their exposure to faults. Fault tolerance and resilience techniques, both on the algorithmic/solver level as well as on the runtime/programming environment level, will be required.

As extreme scale will allow not only to support capability-based applications able to scale out on a full machine, but also capacity-based applications based on coupled multi-physics and multi-scale approaches, it is mandatory at the European level to develop unified scalable environments allowing to support complex end-to-end workflows, smart coupling of applications, the assessment of uncertainties quantification or perform large optimisation studies.

However, the development or adaptation of applications and codes to extreme-scale architectures should not overlook the objective of the European HPC strategy of strengthening the European supply industry and technological autonomy of the Union. It should duly take into account the efforts of any EU initiative to develop European exascale components and systems, like for example the European Processor Initiative, and the associated software stack in a co-design approach. The close interaction of the two is essential to reach the objective of designing, developing and procuring in 2022/2023 an exascale supercomputer based to the extent possible on European technology.

Recommendation 1: Extreme scale computing and data driven technologies

World-class, extreme-scale, High Performance Computing and data-driven technologies should be developed that use software engineering techniques, programming tools and libraries that can be adapted and retargeted to rapidly evolving HPC architectures, in view of maximising application performance and efficiency, in next generation supercomputers. Actions should allow leveraging the efforts on the European low-power processing technologies as well as the Centres of Excellence and build a sustainable exascale HPC ecosystem in Europe, enabling collaborations among the relevant stakeholders. Actions should aim at mature solutions and exploitation in future European exascale HPC system.

The focus should be on the development of extreme-scale computing technologies and system architectures, programming models, mathematical methods and algorithms in an increasingly complex environment of heterogeneous computing with memory and storage hierarchies.

The design of the technologies should respond to critical demands of application performance, energy efficiency, scale, resilience, programmability, etc. across the levels of the compute stack, including compute elements, networking, data storage and data handling.

A co-design approach should be followed, covering from the application to the hardware, and involve the work done at the Centres of Excellence and the data provided e.g. by the Digital Innovation Hubs.

In some sectors, Big Data Computing applications are expected to move towards more compute-intensive algorithms to reap deeper insights across descriptive (explaining what is happening), diagnostics (explaining why it happened), prognostics (predicting what can happen) and prescriptive (proactive handling) analysis. HPC capabilities in HPC centres or at the edge are expected to be of assistance to faster decision making across more complex data sets. For example, modelling, simulation or AI applications are workloads that readily stand to benefit from HPC's experience in optimising and parallelising difficult optimisation algorithms problems. Major requirements include highly scalable performance, high memory bandwidth, low power consumption, and excellent reduced-precision arithmetic performance.

On the other hand, analytics is expected to become a fully-fledged software component of the HPC ecosystem to process the massive results of large numerical simulations (e.g. to apply advanced and highly complex in-situ analytics and processing including visualisation to this data to generate insights) or to feed numerical models with complex data produced by scientific and industrial instruments/systems (e.g. telescope, satellite, sequencers, particle accelerators, etc) or by large-scale systems of systems (e.g. edge devices, smart sensors, IoT devices, cyber-physical systems, etc).

In addition, HPC simulations could profit significantly from iterative refinements of their underlying models effected by advanced data analytics tools and machine learning techniques, e.g. by accelerated convergence. HPC can also benefit from Big Data management approaches, especially in the case of dynamic scenarios (Big Data Computing is much more flexible with the notions of data at rest, data on move, data in change).

Workloads combining HPC and Big Data involve multiple variables and diverse criteria, such as data types, scalability and type of processing, communication between tasks or processes, dynamic and adaptive cluster resource management, dimensioning and configuration according to economic cost, required quality of service, and availability among many others. Software engineering has a critical role to play in ensuring those workloads and platforms can make best use of the underlying hardware resources, lower barriers between development and operations, and offer access to solutions for relevant stakeholders.

While most extreme-scale efforts happen in the context of open-source software developments, many commercial codes also stand to embrace advances made by open source developments. Specific efforts should be taken to address the adoption of both open source software and commercial software to the proposed Exascale architecture and stimulate a rich eco-system of available applications.

Recommendation 2: HPC and data-driven application-oriented platforms

High Performance Computing (HPC) and data-driven application-oriented platforms should be developed, driven by complex application workflows (for instance, High Performance Data Analytics (HPDA), combining artificial intelligence and simulation modelling, exploiting underlying hardware architectures' heterogeneity/modularity, integrating cloud-based solutions etc.) and offer solutions to key application areas, including industrial use cases. The use of HPC solutions to generate innovation and value creation should be clearly demonstrated (for instance in manufacturing, farming, health, mobility, natural hazards, climate or cybersecurity), aiming at providing secure and simple access and service provisioning to relevant stakeholders based on such HPC solutions.

The focus should be on the development of energy-efficient HPC solutions supporting the adoption of applications with industrial and societal relevance on evolving HPC hardware and system software/programming environments. It should include co-design in close cooperation with the scientific disciplines to explore and demonstrate the technical feasibility and value of advanced workflows, e.g. mixed/integrated simulation, HPDA & AI, and ensuring wide adoption in production use.

Recommendation 3: Improve the usage of industrial applications on extreme-scale computing environments

It is necessary to enable industrial applications to fully exploit the evolving HPC hardware and software landscape and seek synergies with open-source components.

The focus should be to improve software and codes for industrial users to fully exploit the new capabilities of extreme performance HPC environments. This includes aspects such as novel algorithms, efficiency, refactoring, scalability, porting and optimisation to novel HPC hardware and software architectures of increased supercomputing performance. The software and codes should be target areas of significant demonstrable market impact, where Europe is leader or should achieve leadership and create value in Europe.

3 Innovating and widening HPC use and skills base

The take-up of HPC services by industry and SMEs is a crucial element for the full development of a sustainable HPC ecosystem in Europe. At the same time, the lack of availability of skilled staff is one of the most pressing problems of SMEs in Europe[1]. Addressing the skills gap and investing in skills development in HPC-related areas is equally important to ensure that Europe has the human capital to boost productivity, innovation and economic growth in Europe.

According to the Coordination and Support Action SESAME Net[2] for example, essential efforts are needed at national and local level to identify and engage SMEs and help them to increase HPC skills and overall use of HPC services. Furthermore, many EU Member States were lacking the competences to facilitate access and uptake of HPC services by industry and SMEs.

The creation of HPC Competence Centres across Europe complementing existing European initiatives would be an effective instrument to provide, at a national level, HPC services to industry, SMEs, academia and public administrations. Moreover, these centres could be the focal point for training, skills development and increase the local HPC knowledge and human capital required to ease and foster the transition towards wider uptake of HPC in Europe.

HPC Competence Centres should be open to develop and deliver tailored solutions for targeted regional/national industries. To support this, the necessary expertise and applications know-how should be made available close to the users. The focus should be on leading-edge, innovative solutions for specific industries/ applications (e.g. automotive, healthcare/medical and/or industrial markets).

The creation of a Pan-European network of HPC Competence Centres in Europe would facilitate the exchange of best practices, sharing of HPC knowledge and expertise. Activities should include twinning and mentoring schemes, sharing of existing libraries of HPC codes and, inter alia facilitate access to upgraded HPC application codes, in order to encourage mutual learning and strive towards reaching a uniform maturity level across HPC Competence Centres, and help close the gap between experienced and less experienced countries. The Pan-European network of HPC Competence Centres would ensure that the specific "vertical" expertise and solutions of each national Competence Centre provides support to other HPC Competence Centres.

Recommendation 4: Support existing or the creation of new HPC Competence Centres and their networking in Europe

To achieve the creation of an exascale ecosystem in the Union it is necessary to support the creation, networking and coordination across Europe of up to one national HPC Competence Centre (HPC-CC) in a maximum number of EuroHPC Participating States. These should provide HPC services to industry (including SMEs), academia and public administrations, increasing the local HPC knowledge and human capital required to ease and foster the transition towards wider uptake of HPC in Europe. HPC Competence Centres should be open to delivering tailored solutions for targeted regional/national vertical industries.

The HPC-CCs should act as national hubs, complementary to existing European initiatives, such as Centres of Excellence, facilitating access of their national stakeholders to European HPC competence and opportunities in different sectors and domains. In this way, they will contribute to validating and disseminating developments of interest to their national constituencies.

[1] https://ec.europa.eu/docsroom/documents/32601/attachments/5/translations/en/renditions/native

[2] Funded under the H2020 EINFRA-6-2014 call – https://sesamenet.eu/

SMEs are the backbone of the European economy. Digital transformation and industry 4.0 offer SMEs opportunities to innovate and grow. It is therefore important to offer SMEs mechanisms to be successful and to stay competitive and to enable smaller actors to enter the market and exploit new business opportunities. The uptake of HPC is an essential enabler to support the SMEs in the transition towards the digital economy.

The EU-funded FORTISSIMO project for example enabled European SMEs to be more competitive globally through the use of simulation services running on HPC systems. Further efforts are however necessary to offer more SMEs and from more industrial sectors to use HPC technology to the benefit of their businesses.

Recommendation 5: Support to training and education for next generation of HPC and HPDA experts and ecosystem
Training and education schemes should be developed to enlarge the pool of highly skilled professionals who will master the next generation of HPC and HPDA ecosystem. This should include engagement with the higher education system, encompassing a multidisciplinary palette from computer science and engineering to computational and data-driven sciences, as well as training courses aimed at improving the skills of existing professionals. Special efforts should be devoted to increasing diversity of both technology developers and lead users, to enlarge the pool of available competence and enable a more inclusive and representative HPC and HPDA ecosystem. The actions should be coordinated with those of the European Centres of Excellence and other existing European initiatives.

Recommendation 6: Stimulating the innovation potential of SMEs
Effective mechanisms for inclusion of innovative, agile SMEs should be provided in a landscape characterised by strong presence of public pan-European e-infrastructures, lowering the barriers for small actors to enter the market and exploit new business opportunities. The focus should be on SMEs whose innovation potential could be increased as users of advanced HPC services, and a mechanism involving financial support to third parties which will adequately stimulate such innovation potential of SMEs participating in the action. The main target should be European engineering and manufacturing SMEs, or any other fast-growing sector of the economy. Links should be made with future national HPC Competence Centres, other relevant domain-oriented centres, Digital Innovation Hubs, or industry associations.

4 FROM A EUROPEAN PROCESSOR AND OTHER EUROPEAN TECHNOLOGIES TO EUROPEAN EXASCALE SYSTEMS

In recent years, significant investments have been made at private, national and European levels in the development of HPC core technologies for the exascale era.

The core objective of the European Processor Initiative (EPI) is the development of a European processor which should be ready for integration in the exascale supercomputers to be acquired by the EuroHPC JU in 2022/23. This presupposes different generations of microprocessors, of growing maturity and market-readiness, accompanied by large-scale demonstrators where they are integrated into prototype systems to assess the progress towards the capability of delivering exascale supercomputers. The EPI is therefore structured in different phases, mapping the development of the different generations of microprocessors and the large-scale demonstrators.

The EPI Phase 1 is funded by the Horizon 2020 Framework Program. It is a crucial European effort to develop European microprocessor technology for HPC, but also for embedded applications that demand high-end computing capabilities (i.e. “embedded HPC”), notably automotive and future autonomous driving capabilities. EPI Phase 2 is expected to further advance the low-power microprocessor designs and bring them closer to an operational device. The architecture and the platforms would evolve and should be integrated in extreme-scale demonstrators to test the technology readiness of the underlying concept, scalability to exascale and compatibility with the applications.

The integration of EPI and other technologies in pilots are a key step towards establishing European exascale capabilities and solutions. The pilots are vehicles to optimise and synergise the effectiveness of the entire European HPC strategy through the integration of R&D outcomes into fully integrated HPC system pilots. The primary focus of pilots should be of establishing proof-points for the readiness, usability and scalability potential of the successful technologies developed in previous European actions, national actions of EuroHPC Participating States or private European actions, when deployed in conjunction with market technologies.

Pilots create “ready to use” systems commensurate with exascale objectives. They encourage a strong co-design approach between technology and applications providers, producing tangible results. Deployed by HPC centres, pilots are used by the involved application providers for their development of new and relevant applications. Hence, the pilots have an up-time compatible with the development and execution of applications, which also implies that applications by way of co-design need to be advanced to a readiness level that allows exploiting the technological advancements pilots are expected to provide. Applications in this context comprise existing and emerging use cases that require substantial performance boosts to address key science challenges with high relevance to European society and economy.

Sustainability and economic viability of the developed solutions are key aspects. However, the chosen applications must have high-volume market potential. Wherever appropriate to address specific technology needs and/or activities, the EPI initiative should remain open to adapt the partnership.

In complement to phase-2 of the EPI initiative, it is necessary to demonstrate in operational environments the successful integration of technology building blocks developed in EPI and other previous R&I actions into world-class exascale-class pilot systems. The challenges of power efficiency, resiliency and scalability of these systems require a strong co-design approach driven by ambitious applications involving technology suppliers, system integrators, supercomputing infrastructure providers and user communities.

5 CONCLUSIONS AND OUTLOOK

This document provides a strategic research and innovation agenda for 2019, supporting the ambition of the EuroHPC Joint Undertaking to develop and maintain in the Union a world-class High Performance Computing ecosystem towards exascale and beyond. The described actions are essential milestones to reach the exascale area but are not an end in themselves. The content of the Strategic Research agenda will be updated regularly to adapt to the changing environment and to take into account the progress of the state-of-the-art. It will also include a long-term vision and an update of Europe’s HPC technology strategy as there is a persistent need for basic research, prototyping and aligning priorities with other stakeholders, all of which should take place in the context of European competitiveness in science and industry.

HPC, data driven and AI ecosystems in particular should continue to identify research challenges particular to their respective domains, but not in isolation from each other. Fostering increased collaboration between HPC and Big Data ecosystems will lead to the identification and exchange of complementary capabilities (e.g. applications, setups, workflows, etc.) that will accelerate the pace of innovation for their respective stacks, i.e. by transplanting specific HPC-inspired-capabilities for addressing the compute intensity of certain Big Data Computing workloads, and similarly transplanting specific Big Data-inspired-capabilities for addressing the data-intensity of certain HPC simulation and modelling workloads.