Ref. Ares(2019)7372492 - 29/11/2019

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

EXDCI-2

European eXtreme Data and Computing Initiative - 2

Grant Agreement Number: 800957

D2.1
HPC-HPDA technology roadmap

Final

Version: 1.11
Author(s): Maike Gilliot, ETP4HPC
Michael Malms, ETP4HPC
Jean-Philippe Nominé, ETP4HPC
Marcin Ostasz, ETP4HPC
Date: 29/11/2019

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: D2.1	
	Deliverable Nature: Report	
	Dissemination Level: PU *	**Contractual Date of Delivery:** 30/11/2019
		Actual Date of Delivery: 29/11/2019
	EC Project Officer: Evangelos Floros	

* Public as referred to in Commission Decision 2991/844/EC

Document Control Sheet

Document	**Title: HPC-HPDA technology roadmap**	
	ID: D2.1	
	Version: 1.1	**Status:** ***Final***
	Available at: http://www.exdci.eu	
	Software Tool: Microsoft Word 2016	
	File(s): D2.1	
Authorship	**Written by:**	Maike Gilliot, ETP4HPC Michael Malms, ETP4HPC Jean-Philippe Nominé, ETP4HPC Marcin Ostasz, ETP4HPC
	Contributors:	Paul Carpenter, BSC Marc Duranton, CEA Hans-Christian Hoppe, INTEL Guy Lonsdale, SCAPOS Sai Narasimhamurthy, SEAGATE Dirk Pleiter, FZJ/JSC Pascale Rossé-Laurent, ATOS Laurent Cargemel, ATOS Estela Suarez, JSC Gabriel Antoniu, INRIA Manolis Marazakis, FORTH Erwin Laure, KTH Andreas Wierse, SICOS Andre Brinkman, JGU Adrian Tate, CRAY Marc Asch, U-PICARDIE Jean-Francois Lavignon, TS-JFL Peter Bauer, ECMWF Cristiano Malossi, IBM Ovidiu Vermesan, SINTEF Benny Koren, MELLANOX Maria Perez, UPM Jens Krueger, FRAUNHOFER
	Reviewed by:	François Bodin, U. Rennes Mathis Bode , RWTH
	Approved by:	MB/TB

Document Status Sheet

Version	Date	Status	Comments
0.1	21/10/2019	Draft	First full text
0.2	22/10/2019	Draft	Some restructuration
0.3	24/10/2019	Draft	First full editorial pass
0.4	27/10/2019	Draft	(Almost) last contents addenda
0.9	28/10/2019	Draft	Full draft for internal review
1.0	15/11/2019	Final	First corrections after internal reviews
1.1	24/11/2019	Final	Final corrections for submission to TB§MB
1.11	29/11/2019	Final	Version following minor corrections from TB

Document Keywords

Keywords:	ETP4HPC, Strategic Research Agenda, Multi-Annual Roadmap, HPC Technologies, Research and Innovation, Digital Continuum

Table of Contents

List of Figures

List of Tables

References and Applicable Documents

[1] http://www.etp4hpc.eu/sra
[2] http://www.etp4hpc.eu
[3] http://www.exdci.eu
[4] https://eurohpc-ju.europa.eu/
[5] http://www.exascale.org/bdec/
[6] https://aioti.eu/
[7] http://www.bdva.eu/
[8] https://exdci.eu/collaboration/coe
[9] https://ecs-org.eu/
[10] https://ec.europa.eu/digital-single-market/en/news/digital-europe-programme-proposed-eu92-billion-funding-2021-2027
[11] https://ec.europa.eu/info/horizon-europe-next-research-and-innovation-framework-programme_en
[12] https://ec.europa.eu/info/sites/info/files/mazzucato_report_2018.pdf

List of Acronyms and Abbreviations

AI	Artificial Intelligence
AIOTI	Alliance for the Internet of Things Innovation
BDEC	Big Data and Extreme-scale Computing
BDVA	Big Data Value Association
CoE	Centre of Excellence (for Computing Applications)
CPS	Cyber- Physical System
DCP	Digital Continuum Platform
DoA	Description of Activity
DPU	Data Processing Unit
DX.Y	Deliverable Number X.Y
EC	European Commission
ECSO	European Cyber security Organisation
ETP4HPC	European Technology Platform for High Performance Computing
EU	European Union
EXDCI	European Extreme Data and Computing Initiative
FET	Future and Emerging Technologies
H2020	Horizon 2020 - The EC Research and Innovation Programme in Europe
HiPEAC	European Network of Excellence on High Performance and Embedded Architecture and Compilation
HPC	High Performance Computing
HPDA	High-Performance Data Analytics
HW	Hardware
IMC/PIM	In Memory Computing; Processor In Memory
IoT	Internet of Things
I/O	Input/Output
IT	Information Technology
MFF	Multiannual Financial Framework
NAS	Network Area Storage
NVM	Non-Volatile Memory
PRACE	Partnership for Advanced Computing in Europe
R&D	Research and Development
R&D&I	Research and Development and Innovation
R&I	Research and Innovation
RIAG	EuroHPC R&I Advisory Group, one of the two Advisory Groups of EuroHPC Industrial and Scientific Advisory Board[5]
SME	Small and Medium-Sized Enterprise
SRA	Strategic Research Agenda
SRA4	Fourth edition of ETP4HPC SRA

Executive Summary

This is a report on the fourth High Performance Computing technology roadmap developed and maintained by ETP4HPC, the European Technology Platform for High Performance Computing, with the support of the EXDCI-2 project. This roadmap, called Strategic Research Agenda (SRA; SRA4 is the abbreviation used for this version), adopts a structured approach to the identification of the key research objectives in the area of High Performance Computing and High-Performance Data Analytics in the 2021 – 2022 timeframe in collaboration with the stakeholders of the Internet of Things, Cyber Physical Systems and Artificial Intelligence.

The SRA4 document is the work of over 80 experts representing the member organisations of ETP4HPC. It also includes the contributions of external technical leaders representing those areas of technology that, together with High Performance Computing, form what we have come to call 'The Digital Continuum'. This new concept well reflects the main trend of this SRA – it is not only about developing High Performance Computing technology in order to build competitive related European systems but also about making our solutions work together with Big Data, Artificial Intelligence, Internet of Things and other similar technologies.

The targeted audiences of this document are:

- EuroHPC Joint Undertaking (EuroHPC) and in particular its Research and Innovation Advisory Group which will use the research objectives identified in this SRA to build its own roadmap;
- entities interested in forming project consortia in response to the EuroHPC (and related) calls;
- anyone interested in the development of High-Performance Computing technology in Europe.

Apart from the well-developed eight technical focus areas of High-Performance Computing related technology and their research priorities, this documents also presents an argument in favour of placing it in the Digital Continuum spectrum. The current state of technology available for implementing the next generation of High Performance Computing systems in Europe is analysed, as well as that of the supporting technologies, roadmaps and projects. New Research and Innovation priorities are defined as a potential basis for future EuroHPC calls. The advancements in High Performance Computing and High-Performance Data Analytics in Europe are presented in the context of similar developments in the United States, China and Japan. Also, upstream technologies with a potential to impact commercially available technologies within the next 5-10 years, such as nanoelectronics and photonics, are discussed. A number of operational suggestions in relation to the future work programmes conclude the document.

SRA4 also addresses three important aspects concerning the implementation of the research priorities presented in the document. First, arguments in favour and against implementing open source hardware or software are discussed. Then, the document recommends a balanced approach towards technology provision when deciding on whether to source from European or global vendors and at the same time ensuring Europe's independence in technology provision.

Thirdly, operational recommendations are presented concerning the definition of work programmes and calls for the 2021 to 2024 period.

As the next step, ETP4HPC will reach out to the co-players and stakeholders across the digital continuum to jointly propose synchronised research actions. During the last six months, ETP4HPC has participated in a multitude of conferences and horizontal collaborative work sessions, aiming at uniting the critical forces needed to tackle the multi-disciplinary technical

challenges associated with the solutions to problems the European society will face in the next ten years.

1 Introduction

The SRA, whose contents and elaboration process are described in this deliverable, operates within the context of EuroHPC Joint Undertaking (EuroHPC JU [4]). Launched in November 2018, EuroHPC is a joint initiative of the EU and at date 30 European states aimed at developing a world-class supercomputing ecosystem in Europe, with 1 billion Euro budget for 2019-2020 initial period [4]. It is intended to pool EU and national resources in High-Performance Computing (HPC) with the aim of:

- Acquiring and providing a world-class multi-petascale and pre-exascale supercomputing and data infrastructure for Europe's scientific, industrial and public users, matching their demanding application requirements by 2020. This would be widely available to users from the public and private sector, to be used primarily for research purposes;
- Supporting an ambitious research and innovation agenda to develop and maintain in the EU a world-class HPC ecosystem, exascale and beyond, covering all scientific and industrial value chain segments, including low-power processor and middleware technologies, algorithms and code design, applications and systems, services and engineering, interconnections, know-how and skills for the next generation supercomputing era.

The overall goal of ETP4HPC and its SRAs is to strengthen the European HPC technology provision value chain and improve the competitiveness of the entire European ecosystem: HPC technologies and systems, infrastructures (i.e. supercomputing facilities), and application expertise. The development of HPC capabilities and maintaining them at a globally competitive level has been identified as one of the European priorities.

Chapter 2 of this document explains the global vision on which this SRA is based – i.e. HPC being at the heart and spearhead of the digital continuum.

Chapter 3 presents the objectives, structure and organisation of the SRA in more detail as well as the concepts that have been used in its preparation.

Chapter 4 shows the emerging important aspects developed in this SRA, as compared with previous releases.

Annex 6 presents the detailed outline of the SRA, which, as of writing this report, is being finalised and will only be published and printed in December 2019/January 2020.

Annex 7 illustrates the timeline of SRA4 generation and the related salient events and milestones.

Annex 8 contains a list of the working groups, including their leaders, whose work constitutes the core of this SRA.

2 Vision on the current and emerging HPC landscape : towards an integrated digital continuum

The rapid proliferation of digital data generators, the unprecedented growth in the volume and diversity of the data they generate, and the intense evolution of the methods for analysing and using that data are radically reshaping the landscape of scientific computing. The most critical problems involve the logistics of wide-area, multistage secured workflows that will move back and forth across the computing continuum, between the multitude of distributed sensors, instruments and other devices at the network's edge, and the centralised resources of commercial clouds and HPC centres [5]. The objective of this SRA is to put HPC into perspective of this new paradigm of 'The Digital Continuum'.

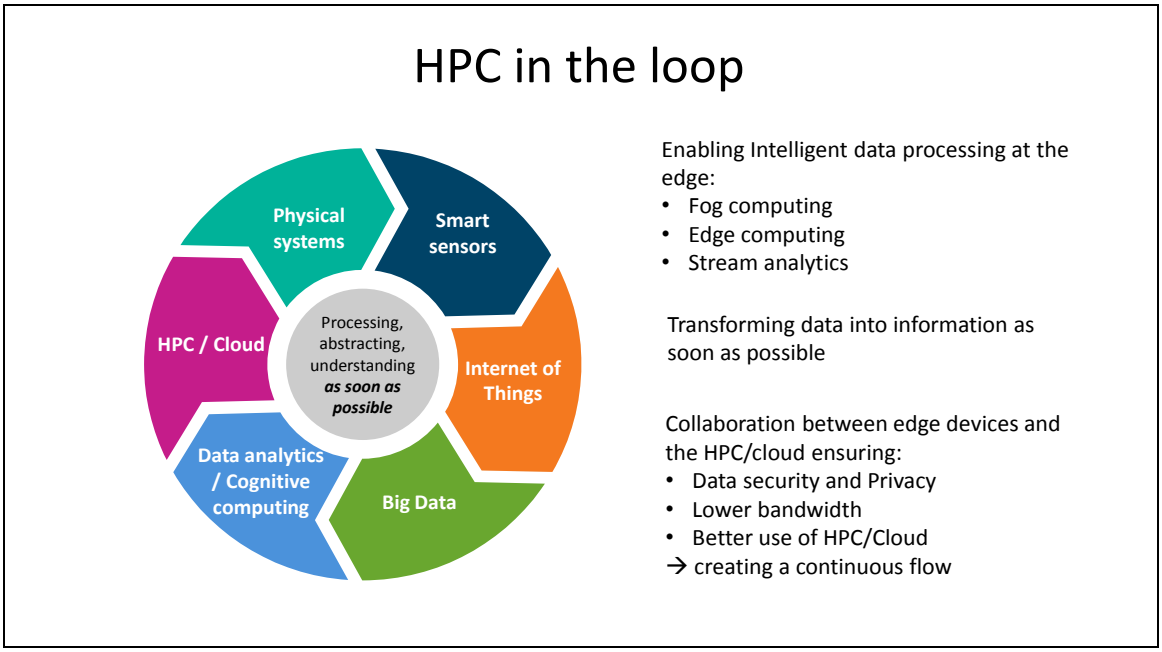

Figure 1: HPC in the loop

Figure 1 illustrates HPC as one element of a circular workflow ("HPC in the loop") starting with data generated at smart sensors in an IoT environment. Data is being locally pre-processed at the edge, relevant parts are forwarded to decentralised "fog nodes" close to the edge. A subset of data is then transferred for centralised Data Analytics in clouds or simulation and modelling in centralised HPC centres. In an increasing number of use scenarios based on the concept of the "Digital Twin" a "twin-copy" of a physical entity is held and continuously updated on these central compute infrastructures (this corresponds to the Centre to Edge Framework vision, a so-called "Research Cluster" in SRA4, as described in section 3.3, and in Figure 4 in this latter Section). The final outcome of the loop is a set of optimised actions in the "Cyber Physical Entanglement" representing physical systems (e.g. robots, vehicles, industrial processes) interconnected in complex intelligent networks.

3 Role and structure of the SRA

3.1 The role of this SRA4

The role of this Strategic Research Agenda (SRA) remains mostly the same as that of the previous three SRAs: its main part outlines research priorities in the area of technology – hardware and software throughout the entire stack of HPC IT infrastructure - for the next 3-4 years. As outlined in Figure 2, the document feeds this information into EuroHPC Research and Innovation Advisory Board (RIAG) as recommendations for the upcoming definition of research calls to be launched in 2021 and 2222.

The SRA is meant to describe major trends in the deployment of HPC and HPDA (High-Performance Data Analytics) methods and systems, driven by economic and societal needs in Europe, taking into account the changes expected in underlying technologies and the overall architecture of the expanding underlying IT infrastructure. The goal is a complete picture of the state-of-the- art and the challenges for the next 3-4 years instead of focusing on specific technologies, implementations or solutions. Any reference to specific products or solutions is intended as a reference to better explain the context, not as an implied promotion. In result, this SRA remains agnostic in the choice of vendors while aiming to maintain the diversity of implementation options.

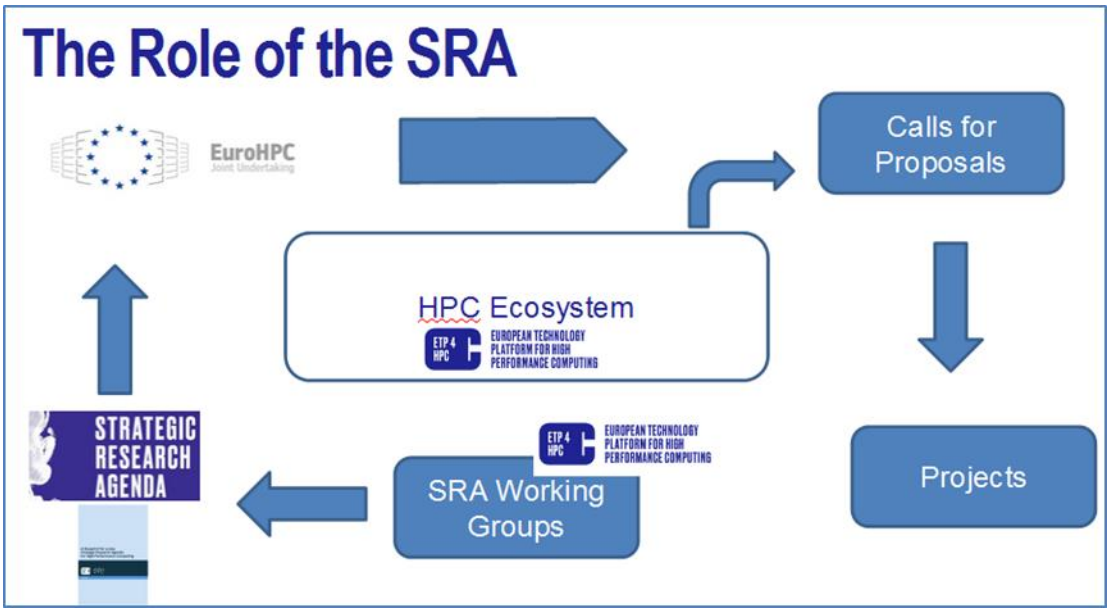

Figure 2: The role of SRA 4
It reflects the principles defined in the ETP4HPC Blueprint (2019) and its main role is to feed the European HPC technology research priorities into the EuroHPC RIAG.

In this regard, this SRA differs from the roadmap documents issued by EuroHPC, which tend to emphasise specific implementation options, originating from its vision and objectives as agreed by the members of EuroHPC.

The majority of this document has been developed by eight SRA Working Groups, composed of technical experts recruited from ETP4HPC member organisations. The leaders and co-leaders of these working groups are well recognised HPC specialists within the European and international HPC community. In addition, partner organisations have provided significant contributions, e.g. the “Big Data-Extreme Compute” – BDEC -project [5], the “Alliance for

Internet of Things Innovation" [6], the "Big Data Value Association" [7], the Centres of Excellence [8] and the European Cyber security Organisation [9]). Further details on the milestones and the interactions among the parties involved are presented in Annex (7).

3.2 Approach for identifying the key topics

Figure **3** shows a layered approach to the identification of relevant research objectives in the area of HPC and HPDA in the 2021 – 2022 timeframe, in collaboration with Internet of Things (IoT), Cyber Physical Systems (CPS) and Artificial Intelligence (AI):

- The top centre layer represents the political framework driving an extended use of HPC and innovation in technology provisioning in Europe in the forthcoming years. Under the "single digital market strategy", the next Multi Annual Funding Framework 2020-2027 (MFF) of the European Commission includes the "Digital Europe programme" [10] to fund the digital transformation beyond 2020 and "Horizon Europe" with "Thematical Clusters" and "Missions" containing societal challenges stimulating R&I in HPC and HPDA. Five thematic clusters address the full spectrum of global challenges through top-down collaborative R&I activities. A small number of missions with specific goals will lead to a comprehensive portfolio of projects cutting across several clusters. The first few missions will be introduced in the first strategic planning phase for Horizon Europe[1] [11].

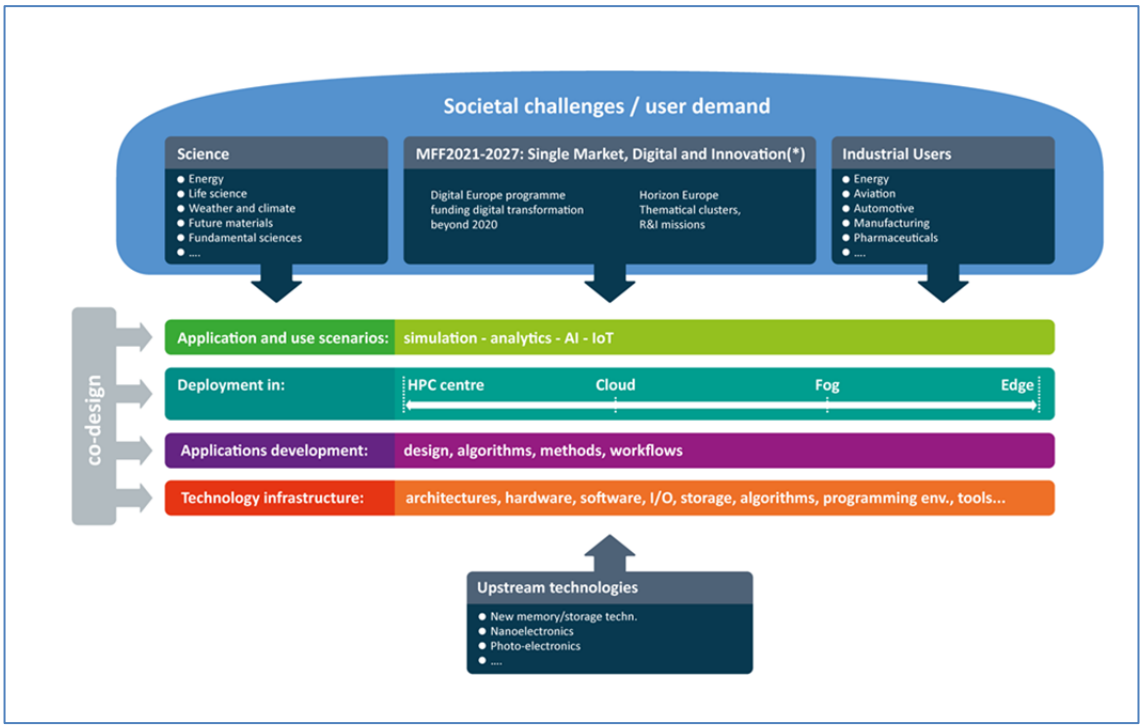

Figure 3: A structured approach to derive research priorities for HPC technology and its application

[1] The examples shown here are preliminary and taken from the Mazzacuto report available at https://ec.europa.eu/info/sites/info/files/mazzucato_report_2018.pdf [12]

- The second source of drivers for future technology improvements is represented on the upper right side by commercial and industrial users of HPC. Especially in this category, new use-patterns for HPC are emerging in the context of new products and services.
- Science has a well-established role in providing major users and in driving the architectural development of HPC systems. Although some fields covered here also are at the centre of the topical clusters and missions referred to in the Horizon Europe framework, it is important to acknowledge the influence of all scientific domains.
- The next layer down ("Application and use scenarios") translates the use-cases defined in the clusters/missions into application and technology use scenarios across the domains of "simulation", "AI", "Analytics" and "Internet of things (IoT)". As argued below, these domains can no longer be handled separately as they are all required to implement solutions to the current and future problems.
- HPC technology will not only be deployed in dedicated data centres in the future. A federation of systems and functions with a consistent communication and management mechanism across all participating systems will be required creating a "continuum" of computing. The "Deployment" layer describes the challenges associated with this change, where HPC functionality is now extended to clouds, fog computing and edge computing.
- The next layer down ("Applications development: design, algorithms, methods, workflows") addresses the software development aspects of the application portfolio.
- The next layer outlines the technologies used to implement the IT infrastructure discussed above. While most of the described components, functions and features will be deployed in data centres, local small-scale deployments (edge/fog) will integrate the technology stack as well. The range of technologies covers algorithms, programming languages and tools, system software, architectures, hardware components, I/O and storage as well as addressing critical features such as reliability and energy efficiency. This is the core of the SRA.
- The emergence of upstream technologies for future HPC system/component architectures complements the influence of the societal challenges and is expected to facilitate novel and superior solutions. The related chapter outlines those candidate technologies which are most likely to be applicable within the timeframe of Horizon Europe.

3.3 The structure of SRA4

At the core of the document the outline and prioritisation of HPC research domains for the next work programme period can be found in a structured way using the concept of "Research Clusters" and "Research domains". The model used to identify the elements discussed in these two views is the layered approach shown in Figure **3**. The layers "Applications development" and "Technology infrastructure" are covered in detail in this chapter.

- **Research domains** are needed to describe the essential layers and elements of an HPC functional stack. While "System Architecture" applies a holistic view an all elements of the stack, "System Hardware Components" covers the lowest hardware level followed by "System Software and Management". "Programming environment" adds a user's (programmers) view to using the HPC hardware and system software infrastructure. "I/O and storage" covers the space of feeding the compute nodes with data and storing data. "Math and algorithms" provides yet another view on any level of software used in HPC systems, not to confuse with algorithms used in the application layer. "Application co-design" offers an application writer's input to the needs for the implementation of next generation HPC infrastructure. Finally, given the new "HPC in a digital continuum-trend, "Centre to edge" covers any aspects of HPC functionality starting to be used in complex multi-tiered workflows.

- **Research clusters** represent crosscutting "themes" which encapsulate the research priorities for the next generation of HPC infrastructure. They are shown as orthogonal elements in Figure 4 as their impact crosses over all or almost all research domains. Some are traditional themes such as "Energy efficiency" or "Resilience", some are new and originate from outside the traditional HPC, e.g. "Data everywhere" or "AI everywhere". Under the heading of each "Cluster" several aspects with a high degree of similarity are grouped (or 'clustered') together. The clusters are introduced in the following chapter.

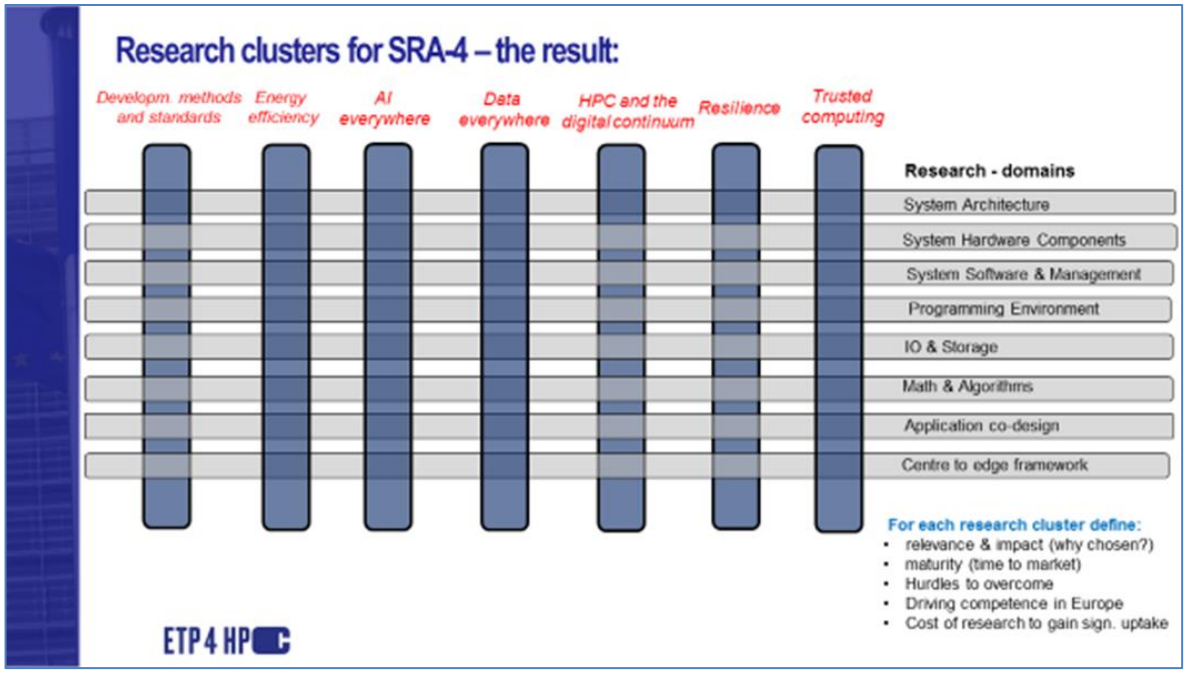

Figure 4: The sources for the SRA4 research priorities: the concept of clusters and domains

- As outlined in the chapter on Upstream Technologies, significant investments are made in R&D on next- or even further technology generations not directly linked to HPC, but to a broader use in computing and data processing. In some cases, a clear deployment within the next years can be foreseen, such as components for neuromorphic computing or "in memory computing". In other cases, their commercial exploitation is not yet obvious, but they deserve attention today, as a path for their utilisation (e.g. quantum accelerators) needs to be addressed now. In this chapter, these options will be presented starting from emerging ones, or the ones that are almost implemented, and then moving on to the current state of the art to the most disruptive ones.

- The SRA will also provide non-technical information useful to understand the context of the core technical chapters:
    - A separate chapter is dedicated to the international HPC and HPDA arena presenting a snapshot on the most prominent ecosystems, namely those of the US, Japan and China, and their strategies and achievements in order to understand how Europe can either benefit from their work or compete with them effectively, see "International arena: HPC and HPDA in Europe/China/US/Japan".

- Operational Recommendations for the implementation of Work Programme 2021/2022 will be presented in chapter “Operational recommendations for the implementation of Work Programme 2021-2022”. These recommendations include proposals for new project types and support instruments.

4 Main novelties in SRA4

Compared to the previous SRAs which addressed mainly HPC stack-centric priorities, this SRA reaches out to include the priorities gaining importance due to the increasing use of HPC functionality in complex, multi-disciplinary workflows together with AI, IoT, and HPDA workflow phases. The interaction with these domains makes the technology provisioning more complex, as new requirements need to be met, e.g. security and privacy constraints. In many new deployment scenarios, the challenge is to intelligently adapt to and adopt technology, methods and best practices invented and in use elsewhere and on the other side provide interfaces, functions and services for other workflow stages to connect and exploit the HPC capabilities.

4.1 Application and use case scenarios

HPC has always been one of the main tools advancing science by delivering results, which can be attained by the use of cutting-edge computer technologies only. Throughout the last decade, numerical computing has been growing rapidly in many directions: higher fidelity, multi-physics models; deluge of observational data from sensors and of simulated data; semi-automatic data analysis and post-processing; uncertainty qualification and AI-based models. Combining all these aspects will result in a highly complex application (software) architecture, currently a focus area in related research.

In reference to Figure **4**, this layer is driven by the thematic clusters and missions as well by industrial and scientific needs. The extraction of IT/HPC requirements out of representative and strategically important use case scenarios is necessary in order to drive HPC R&I in the right direction. They are key to assess new architectures or infrastructure as well as provide testbeds to research & industrial teams.

In the context of promoting innovations for the HPC, HPDA and IoT ecosystem, the use cases identified must be such that alignment with technology “silos” is avoided, which would strongly restrict the shaping capabilities for the R&I work program. Furthermore, fully addressing the societal challenge can only be achieved when considering end-to-end approaches where data production is integrated with data analytics, machine learning, numerical simulation, data archiving as well as the final use of the results (availability and reprocessing of open data through portals). Underlying the use cases are applications relying on complex workflows within which individual tasks are executed on a wide variety of systems and whereby the complete data management cycle is addressed.

However, many representative use case scenarios are difficult to analyse since they combine many heterogeneous components (e.g. relying on different software stacks) as well as different resources or user governance strategies. For instance, this is about applications across a federation of systems - that includes HPC centres, cloud facilities, fog and edge components, and networks - while at the same time preserving security and privacy from end-to-end. Furthermore, the economics aspects of the deployment of these applications must be considered.

As a consequence, this means facing extreme scale heterogeneity where, in the worst case, the common denominator may be a common governance and resource allocation policy. At a high level, the main technical challenges are how to achieve interoperability between the application workflow components, their orchestration as well as reproducibility of execution in order to allow debugging and ease of deployment. In addition, infrastructure management and resource allocation policies are also strong roadblocks to overcome. For instance, supercomputers today

are typically deployed in a way that they become silos, with limited external connectivity, proprietary access processes, relatively rigid operational models that expect users to submit batch jobs, and limited flexibility in terms of software stack provisioning. It is difficult to make them part of an application workflow that would include components deployed in the Cloud, handle streaming of data, for example.

4.2 Work flow and capabilities

Understanding the workflow and data flows is of crucial importance for an analysis of real use cases. Each case (e.g. autonomous driving, personalised medicine, wind park operation, etc.) has its unique composition of basic "functional capabilities":

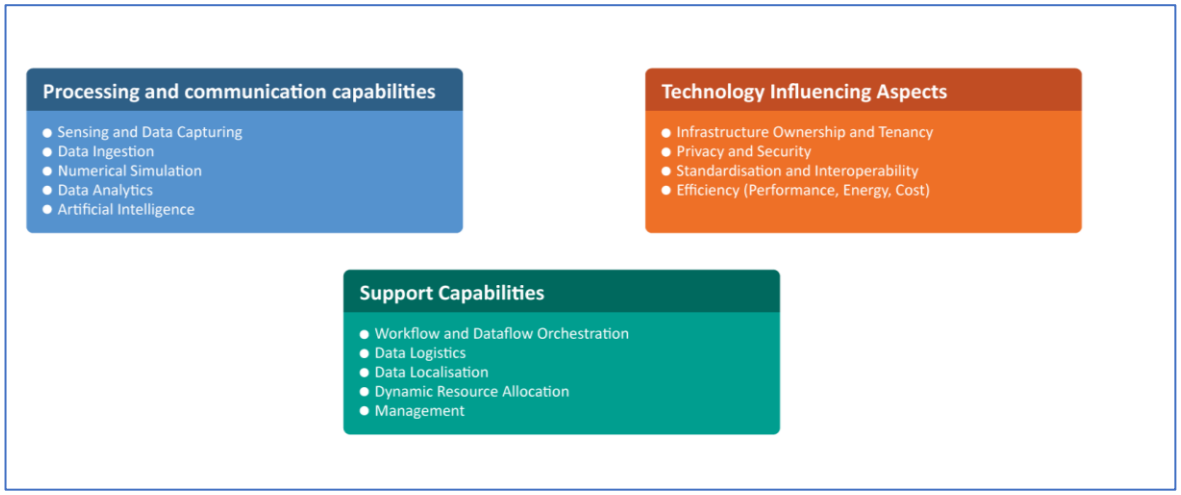

Figure 5: Categories of capabilities in mixed Simulation, Analytics, AI and IoT use scenarios

- The "Processing and Communication Capabilities" listed in Figure **5** cover all areas which require compute capabilities, be in a datacentre, edge or fog node or an IoT device – each of them with a different application scope. For a given workflow (use-case), the individual processing capabilities are expected to be spread meaningfully across locations and systems. We distinguish between data capture from devices, data ingestion into a compute environment, the typical HPC capability of numerical simulation, and the Big Data capabilities of data analysis and artificial intelligence. To address such new compute requirements, HPC capabilities must provide the processing capabilities for the Big Data environment, which includes interactive analytics as well as batch and real-time processing of data streams.
- The "Technology Influencing Aspects" are properties that significantly impact the design, implementation and integration of the processing capabilities but do not directly provide any data processing capabilities. They must be provided by the processing infrastructure in ways that satisfy the end-user requirements to result in an effective and efficient solution. The governance of compute infrastructure and data imposes policies on the data processing. Security and privacy must be considered in such an environment for most use cases to comply with regulatory and end-user needs. Interoperability and standards increase the trust in developed workflows and accelerate the adoption by users. The efficiency of a solution is relevant insofar that the costs of a solution limits the adoption in use-cases that yield limited revenue. A performant, energy and cost-efficient system maximizes industrial and commercial competence by enabling novel scenarios.
- "Support Capabilities" describe the crucial implementation aspects of a mixed scenario. As shown in Figure 5, the workflow reflects the interconnection of actions and data between the IoT devices, processing entities and data repositories. The identified capabilities are

currently underdeveloped for the environment discussed here and require further R&D efforts.

The orchestration of workflows and automatic / efficient deployment across the complex hardware-landscape provided, is required to exploit such systems. For instance, data must be placed and migrated intelligently to match the storage and processing capabilities of (IoT or edge) systems. Finally, workflows must adapt their processing capabilities dynamically depending on the input, or other external parameters such as the number of users or availability of processing capacity. This requires software layers that enable such dynamic, ad-hoc changes. We recognise that management procedures must be developed that deal with the distributed nature of computation, ownership, and conformance to standards while considering the efficiency aspects.

4.3 Data life cycle and dataflow: an example

Understanding the necessity for a dataflow orchestration in mixed Simulation & Big Data use scenarios is important. The capacity of storage infrastructure, the increased sophistication and deployment of sensors, the ubiquitous availability of computer clusters, allow the development of new analysis techniques and real time capabilities to ingest “fresh” data during simulation.

There are multiple scenarios:

- Input data coming from experimentations is injected into simulation to enhance it and potentially provide feedback to the instruments in order to steer them. In this case, the quality of the simulation will depend on the availability of this new set of data on the HPC system.
- Data is produced by sensors in a streaming mode and HPC resources are used to train the model. The model- training frequency will depend on data source obsolescence.
- Output or step-by-step data can be extracted from simulation for new in-situ processing, visualisation, post processing/inference and simulation context modification.

For these new scenarios, we observe a need for different levels of curation (sensors producing non-curated data versus use of databases with curated data):

- Unstructured data for instance issued from major scientific instruments or experimental facilities, which may be residing outside of supercomputer centralised facilities, will require non-trivial transformations before an ingestion could be realised by a simulation or Machine Learning or other HPDA steps. Depending on the real-time availability and quality of this data, the transformation and availability for simulation need a strong coordination effort (near real time data preparation).
- Qualified structured data resources shared by the communities through archives, databases or any specifics formats accessible through the internet have a well-known preparation process to enable their use in a simulation. The data transfer, compression process, encryption process could slow down the simulation workflow and could require some provisioning or concurrent data and simulation processing.

Then the challenge is to add and to coordinate the integration of these new data resource types in end-to-end application workflows without drastically increasing storage space dedicated to data availability.

For data driven application workflows, a well-balanced architecture will mostly depend on the efficiency of the dataflow and on the capability to reduce, filter, pre-process data close to the

source (on edge computing devices or fog nodes). The objective is to limit network and global storage congestion. The pre-emption of HPC resources depends on data workflow optimisation. This distributed data transformation must be integrated in a Big Data life cycle model that includes activities to more closely combine data curation with the research life cycle. These activities address planning, acquiring, preparing, analysing, preserving, and discovering data, describing the data and assuring its quality.

The relationship between scientific community data repositories and new distributed data workflows as well as the reproducibility in computational science are to be studied. Documenting data sources, experimental conditions, instruments and sensors, simulation scripts, processing of datasets, analysis parameters, thresholds, and analysis methods ensures not only a much-needed transparency of the research, but also data discovery and future data use in science.

Orthogonally, new HPC systems will have to consider, at the same time, where data is stored and how/where the same is accessed for computation. In a federated scenario, data could be stored across distributed Edge, Fog and possibly multiple centralized “data-centre-like” systems, e.g., reflecting the data production sites or specific access policies. Solutions to allow simulation, analytics or AI applications access data across federated and heterogeneous sites must be designed and built to strike the proper balance between data access performance, cost and consistency, while at the same time satisfying access control and privacy constraints. For example, AI-assisted intelligent caching systems could be designed and deployed to take advantage of the read-only nature of many workloads (e.g. Deep Learning training or Big Data input), either using high performance node-local storage such as NVM disks or leveraging existing NAS infrastructure.

The design of a global infrastructure allowing one to combine external edge- or peripheral environments with a central, shared infrastructure will require the analysis of the heterogeneity of the entire software environment, the identification of new data sources and the quality of the data. The diversity of application requirements, in terms of workflow patterns and data distribution will define the rules for new combined data use solutions.

5 Conclusion and outlook

The development of "High-Performance Computing" is undergoing a significant change. 'HPC' does not apply to only supercomputers in large datacentres but also to other scales like embedded or edge deployment; and HPC now finds itself at the heart of a converged compute infrastructure supporting simulation, modelling and data analysis in a global digital computing continuum. Core HPC technologies and methodologies are being used to enable concurrent processing to permeate all levels of that digital computing continuum.

Research on both HPC applications as well as on HPC technology will expand from the current fields deploying HPC solutions to adjacent fields to address AI, Data Analytics and IoT-related challenges. This will influence the selection and definition of research priorities in the next SRA and this can only be effective and meaningful as the result of a true interdisciplinary effort.

Several workshops and collaborative sessions have been held during the preparation time for the upcoming SRA4. The analysis of a diverse set of "digital continuum use scenarios plays a significant role in determining the research focal points for the next SRA.

Ultimately, the recommendations defined in the next agenda will serve as an input to the research and innovation advisory process of EuroHPC, assisting in the definition of its R&D&I work programme for the period 2021-2022 (and beyond).

6 Annex: Strategic Research Agenda – Full table of contents

As mentioned above the document currently is in its integration stage with a deadline of end of year 2019. Given the three month delayed start of the actual work within EXDCI-2 for preparing and writing the document, only a temporary (although almost final) table of contents can be provided together with a description of the intended contents of the major chapters - see "The structure of SRA4" in section 3.3.

---

7 Annex: Timeline, milestones and meetings related to the SRA writing process

The following tables show the sequence of events for the creation of the SRA document (Table **1**), then the list of salient working sessions and workshops held in the process of SRA elaboration (Table 2). At the time of issuing this report, the integration of the chapters developed by the 8 working groups took place.

- March 19th : SRA4 process communicated at General Assembly
- March 21st : Invitation to apply for SRA-4 working groups to be sent out to ETP4HPC members
- March 31st : FP9 vision document electronic version available, registration for working groups incl. suggestions for research clusters
- April 12th : Deadline for working group registration and collection of proposals for research clusters
- April 15th - April 19th : we analyse your input and set up working groups
- May 17th : SRA-4 working group leaders internal workshop during European HPC Summit week (May 13th to 17th)
- May 20th – June 14th: Kick-off calls with working groups (8 calls, set up by office)
- June 19th : SRA-4 working group leaders internal workshop during ISC 19, start 18:30, Citadines Hotel, Frankfurt
- July/August/early September: writing complete text, individual working group calls (organized by working group leaders)
- Sept. 10th deadline for completing research domains and working group domain descriptions
- Sept. 10th to Sept. 24th : first peer review cycle (2 groups of four working domains)
- Sept. 24th to Oct. 2nd : working group leaders working in review results
- Oct. 4th – October 25th : integration of all contributions, first rendition of document
- Nov. 13th: f2f work session in Muenchen
- November: text refinements, reviews, corrections
- optional Dec. 9th week: closing SRA-4 workshop, IBM ZRL Rueschlikon
- December: language checks, document design, release

Table 1 : Sequence of events for the creation of the SRA document

April 2018	Conducted an "extreme compute – extreme data" use case analysis workshop with technical experts from BDVA, Exdci-2 and external HPC experts. Out of 15 use cases, 5 were analysed further to understand the diversity of the two technology stacks. A joint BDVA - ETP4HPC document was generated discussing the options for convergence of the two stacks and their basic characteristics and usage patterns. See: https://www.etp4hpc.eu/bigdata.html
June 24th, 2018	Work session with international contribution. Focus was a vision for the next 5 years for HPC technology and use evolution. The 5 presentations led to a second workshop on June 27th with the SRA technical experts and workgroup leaders to identify the main drivers for the next generation of research priorities and the implications with adjacent domains like AI, Data Analytics and IOT.
December 2018	At ICT 2018, BDVA and ETP4HPC held a session on "HPC & Bid Data" on the digital continuum between HPC centres, cloud, flog and edge computing.
December 2018	At the EBDV-Forum : participation in a panel discussion on the collaboration necessity between European associations promoting HPC, Big Data, Cyber Security, Robotics, 5G and IOT. Triggered by this event a closer collaboration between ETP4HPC, ECSO and AIOTI started and the joint work with BDVA was intensified.
December 2018	A MoU with BDVA and one with AIOTI was signed focussing on joint road mapping work for future SRAs.
May 17th, 2019	Conducted an EXDCI-2 SRA work session with technical experts, as part of the EuroHPC HPC Summit week. New workgroup leaders came on board and a consensus was reached on how to qualify research priorities for the SRA. Also, several industrial use cases were presented and examined following the approach outline in the Blueprint paper
May 24th, 2019	started the set of 8 working group kick-off calls (in total with over 100 workgroup members from industries, academia, research centres and organisations like Hipeac, BDVA and AIOTI.

Table 2 : Working sessions and workshops held in the process of SRA elaboration

8 Annex: SRA4 working groups and leads

System Architecture	• Laurent Cargemel, ATOS • Estela Suarez, JSC • Herbert Cornelius, MEGWARE
System Hardware Components	• Marc Duranton, CEA (HiPEAC) • Benny Koren, MELLANOX
System Software and Management	• Pascale Rosse-Laurent, ATOS • Maria Perez, UPM (BDVA) • Manolis Marazakis (FORTH)
Programming Environment	• Guy Lonsdale, SCAPOS • Paul Carpenter, BSC • Gabriel Antoniu, INRIA (BDVA)
I/O & Storage	• Sai Narasimhamurthy, SEAGATE • Andre Brinkmann, JGU
Mathematics & Algorithms	• Dirk Pleiter, JSC • Adrian Tate, CRAY (NAG)
Application co-design	• Erwin Laure, KTH • Andreas Wierse, SICOS
Centre-to-edge-framework	• Jens Krueger, FRAUNHOFER • Hans-Christian Hoppe, INTEL

Table 3 : SRA4 working groups and leads