EUROPEAN HIGH PERFORMANCE COMPUTING
JOINT UNDERTAKING

Infrastructure Advisory Group

Multiannual Strategic Agenda

For the acquisition of Supercomputers

Date:	11/01/2019
Version:	1.6
Authors:	The members of the EuroHPC JU Infrastructure Advisory Group
Approved by:	Sanzio Bassini, Zoe Cournia, Sergi Girona, Branislav Jansík, Lene Krøl Andersen, Stoyan Markov, Norbert Meyer, Peter Michielse, Claus Axel Muller, Minna Palmroth, Sinead Ryan, Daniel Verwaerde
Public:	EuroHPC JU Governing Board

Table of Contents

Document History

Version	Date	Comment
1.0	26/11/2018	Document created for The members of the EuroHPC JU Infrastructure Advisory Group
1.1	28/11/2018	Updates based on feedback from the 1st INFRAG meeting. Removed references to Top500 ranking goals. Introduced chapter on European aspects and moved the European Technologies section there. Removed detailed list of high-level services and re-phrased the relevant reccommendation. Changed performance target of pre-exascale from peak performance to theoretical (Rmax) performance. Added references to sustainable and green energy solutions. Added recommendaiton on hosting entitiy competences. Emphasis on applications has been underlined in various places in the text and in recommendations.
1.2	02/11/2018	Updates reflecting off line comments. Renumbering of recommendations. Various text improvements.
1.3	03/11/2018	Elaborated text on system co-design, system support aspects, benchmarking and experimental platforms.
1.4	04/11/2018	Clarifications regarding systems characteristics.
1.5	20/12/2018	Various changes based on comments and suggestions of INFRAG members
1.6	10/1/2019	Updates incorporating results from 2nd INRAG meeting discussions.

1 INTRODUCTION

The Multiannual Strategic Agenda has been defined by the EuroHPC Infrastructure Advisory Group (INFRAG) according to the mandate specified by the EuroHPC Regulation. The purpose of this document is to advise the Governing Board of the EuroHPC Joint Undertaking on the infrastructure development roadmap that EuroHPC should follow in the coming years and in particular on the number and types of supercomputing systems that EuroHPC should procure, deploy and operate in the Participating States, and in a longer term, on the pooling and federation of the HPC resources in Europe.

The INFRAG has taken into consideration the current views of HPC User Communities in Europe, as identified by the two informal EuroHPC working groups on User Requirements and on Hosting & Procurement, as well as by the PRACE Scientific Steering Committee. The current roadmap of scientific usage of HPC in Europe has also taken into account the *PRACE Scientific Case for Computing in Europe 2018-2026.*

This report also outlines the technological and architectural state of the art and presents options that EuroHPC should take into consideration when preparing the upcoming procurements of supercomputing systems. A recommended timeline for the execution of procurements to meet the goals of EuroHPC is also provided.

2 THE EUROPEAN DIMENSION

2.1 European Technologies

The European supply of HPC Technology currently lags behind the rest of the world. As it has been noted, although Europe consumes 30% of global HPC resources the European industry contributes to only 5% of them.

In the recent years the European Commission has funded a large number of projects with the aim to research and develop novel HPC technologies. These projects have already started delivering results, with some of them reaching a high maturity level and are currently commercialized. Moreover, the recently established European Processor Initiative (EPI) with the goal to foster the supply of a European processor following ARM and RISC-V instruction sets is an important opportunity for the future exascale machines and holds the promise to accelerate technological development in Europe, impacting various industry sectors (e.g. ICT, Automotive, IoT).

This technological development effort has to be accompanied with a strong co-design approach, ensuring that the technologies meet the users' demands (in hardware, software and development tools) and that users adapt their codes and applications to the new technologies developed. Keeping user communities involved in this process is crucial for the successful exploitation of future computing platforms. For this purpose, mechanisms should be in place to link the users' demands with the technology supply and their integration all the way to the supercomputers in operation. Porting of existing scientific codes into new architectures requires considerable effort and investment of many person years from the side of application development teams. It also requires, significant effort in preparing the programming environment and the management systems able to manage much bigger configurations (towards exascale).This fact has to be taken into consideration when making architectural choices. However, in the longer term, to exploit the exascale capabilities, a porting of application codes is unavoidable.

Recommendation (1): The supercomputers acquired and operated by the EuroHPC Joint Undertaking should aim to maximize the uptake of European technologies. For this purpose, the EuroHPC Joint Undertaking should make a best effort to link the development of competitive EU technology supply with the users demand and to explore ways to best integrate the results of H2020 projects on HPC technology and integration in the target Pre-exascale and Petascale systems.

2.2 The European Infrastructure Ecosystem

The EuroHPC infrastructures activities are not run in isolation. There are also a large number of national initiatives taking place in parallel, focusing primarily on local and regional user requirements and development strategies. Systems operating on a national level can play an important role in the provisioning of a full-scale HPC ecosystem in Europe. In some cases, they can provide small-scale versions of the pre-exascale systems, or incorporate novel technologies offering a platform of testing and application porting paving the way for bigger systems of similar architectures.

In parallel, Europe is investing considerably in Research & Innovation programs, giving priority to domains considered strategic both at National and EU level. They are improving computational applications, are supporting the development of new computing paradigms, the evolution of programming paradigms and the improvement of both development and system level tools, as well as the developing the underlying hardware components, systems and the integration into experimental machines. This work should not take place in isolation but in close collaboration with European infrastructure activities, including PRACE for HPC and EUDAT for data-related aspects. This calls for a close coordination between the Research & Innovation and the Infrastructure development pillars of EuroHPC, ensuring their complementarity.

Recommendation (2): National HPC infrastructure initiatives should align with JU efforts in deploying and offering a diversity of architectures able to support the full range of applications used in Europe.

3 USER AND TECHNOLOGY CONSIDERATIONS

3.1 Challenges and opportunities of user communities

EuroHPC will provide important HPC resources both for academia, industry and public administration. The main goal of these systems should be to satisfy the demand of the scientific applications, without neglecting industrial applications, and should help overcome the challenges and limitations currently faced by them. This means that the EuroHPC supercomputers should not only be designed based on theoretical FLOP/s rates but should aim for improving the actual application performance.

Future pre-exascale and exascale systems are expected to expand the frontiers of **Fundamental Sciences** and also address the need for advances in **Next Generation Computing** with new algorithms that scale better and completely new approaches to solve mathematical problems in more energy efficient manners.

In brief the areas to be supported by the EuroHPC infrastructure cover a broad spectrum or academic and industrial applications including:

- Climate, Weather and Earth Science
- Life Sciences and Medicine
- Energy, Oil & Gas exploration, Fusion & Plasma simulations

- High Energy Physics and Astrophysics
- Engineering, & Manufacturing
- Infrastructure & logistics
- Chemistry & Material Sciences,
- Emerging domains including Artificial Intelligence, Deep Learning, Big data analytics, Internet of Things (IOT) and cybersecurity.

Recommendation (3): EuroHPC supercomputers should not target a limited number of applications but rather should aim to cover in the best way possible the computational requirements of a broad range of European scientific, industrial and public administration applications.

Recommendation (4): The pre-exascale supercomputers should be capability systems, designed to support large applications requiring a maximum of resources. They should take into account both the peak performance and the applications performance, with priority given to the latter aspect, in order to provide leadership-class capabilities in Europe. For the measurement of the application performance appropriate benchmarks should be defined, for example inspired by the benchmarks and the corresponding workloads in use by PRACE.

Recommendation (5): Peta-scale supercomputers should complement as much as possible the pre-exascale supercomputers with less capability but covering a wider range of users, usages and applications.

3.2 Architectural aspects

This section analyses the main hardware and software needs for the EuroHPC supercomputers, driven by the HPC user needs.

3.2.1 Processing units

There are three trends in the recently deployed HPC systems: a) HPC standalone processors (multi-core), b) many-core processors and c) Hybrid architectures incorporating accelerators (such as GPUs) attached to generic processors, While multi-core processors provide a generic approach for most common HPC workloads, accelerators and many-core CPUs provide a higher number of threads with a lower performance. This implies certain limitations to extract good performance from the latter, the clearest being the more restricted scalability to several cores, which might not be possible for many applications. Typically, CPU-based systems impose a lower adaptation barrier for generic users as application porting doesn’t require specific modifications. The opposite is typically true for accelerators, which demand adoption of specific APIs and multithread parallelization approaches but the performance gains in the end usually compensate for this extra effort.

From the energy efficiency point of view, the accelerator-based and many-core approaches provide a very good Gigaflops/watt ratio, while the HPC standalone processors usually give a lower Gigaflops/watt.

Considering the performance targets of the pre-exascale machines, the options to provide a pure HPC standalone-processor machine is limited, as with the current technologies it will require an excessive power consumption to provide a machine in the order of 300 petaflops peak performance. Also, the current offerings for many-core processors is very limited in the time-frame for pre-exascale machines, but this option could be evaluated for the future machines beyond 2023.

While the vast majority of codes today have been optimized for x86 systems, user community feedback indicates that most users will not have a problem switching to a different architecture. However, several user groups do express the importance of high clock, high floating-point rate, and lots of cache per core and node. This means it will likely not be a good alternative to consider systems that compensate low clock with many more nodes. The peak performance of the system is the same in the end but it will not be optimal for actual applications.

For what concerns accelerators, there is undoubtedly a widespread adoption of GPU accelerators across the full range of stages in a simulation, from ab initio simulation and data generation to post-processing and data analysis. Accelerators are also used by a significant fraction of the HPC users' community from biosciences to physical and engineering applications. The downside of current GPU offerings is that there is a strong bias towards a single vendor with more than 95% of the codes ported on GPUs using CUDA. This significantly limits the architectural options and leads to potential technology and vendor lock-in.

Recommendation (6): Upcoming pre-exascale and exascale systems are expected to rely on a combination of multi/manycore CPU with accelerators. The implementation of the EuroHPC supercomputers' should provide as much diversity of architectures to satisfy the application demands, and should align, to the extent possible, with the hardware efforts in Europe and in particular with the architectural choices of EPI. The effort required for users for porting their applications to new architectures should be taken into consideration, in particular for the short term. Nevertheless, in the long term the codes should be developed for the new architectures.

Recommendation (7): EuroHPC has an important role to promote competition and open standards in the entire market. Considering the level of investment about to make on HPC infrastructures, it is reasonable to expect that EuroHPC has the necessary weight to influence the accelerator industry to move towards open programming standards (e.g. OpenMP and OpenCL). This must be combined with a common risk-sharing approach between EuroHPC investments and technology providers in moving key software to such APIs.

3.2.2 Memory

Although CPUs and HPC accelerators have improved drastically, the peak performance of the compute nodes and the delivered performance at the application level have been lagging. The growing gap between theoretical and delivered performance is directly connected to slow improvement in memory speed; this is often referred to as the "memory wall".

Upcoming Non-Volatile Memory (NVRAM) technologies are opening new opportunities for HPC systems. New NVRAM will feature larger byte-addressable capacities as DRAM allowing new approaches where only compute nodes are populated with High Bandwidth Memory (HBM). Their performance (read or write bandwidth) will be higher than current flash-based NVRAM, approaching DRAM levels. Furthermore, their endurance should be comparable with DRAM at least in combination with some hidden wear-levelling technology. They could be used in HPC systems, both as main memory and ultra-fast IO.

It is clear that these new technologies will provide new options of memory speed and size within each node but the way to access to this memory hierarchy may increase the complexity of the codes and limit its portability.

For HPC applications, the memory used largely depends on the problem to be solved, moving from few MB per core to several GB/core.

Recommendation (8): Considering the diversity of the codes that the supercomputers will have to serve, the fact that not all the codes pose similar memory requirements and the fact that memory is an expensive system component, nodes with variable memory profiles (in terms of total size) should be considered to fulfil the different users' requirements while keeping the overall system budget within acceptable limits. The actual choice of memory size per node shouldn't sacrifice the effective bandwidth, which remains crucial for application performance, but should take into consideration the requirements of the applications.

3.2.3 Systems Interconnect

The application performance relies on parallelisation and depends directly on the efficiency of the interconnect unifying the compute nodes into a single system. The HPC system interconnection network must scale together with the compute nodes and the storage performance. The HPC networks bandwidth is planned to grow from a 100 Gb/s today, up to 200-400 Gb/s in the coming years thanks to the development of new generations of *SerDes* (Serializer/Deserialize) circuits.

For upcoming pre-exascale systems, the minimum required bandwidth should be around 2Gb/s per MPI (Message Passing Interface) task. Higher bandwidths will permit sharing of multiple workloads such as MPI communications and the parallel filesystem I/O over the same network.

For what concerns latency, current technologies are providing around 0.5-0.9µs, adding to this number the latency of each of the hops required in order for a package to arrive to the final destination. Most of the network technologies used in HPC should provide at least 100-150 million packages per second. Finally such interconnects should be able to provide the hardware to offload the MPI traffic (including asynchronous collective operations), provide adaptive routing mechanisms, ensure end-to-end reliability of communication and allow the support of multiple topologies.

Recommendation (9): System interconnect is a critical factor to achieve the best possible application performance efficiency. EuroHPC systems should aim for state-of-the-art high-bandwidth low-latency interconnects, adequately leveraging the blocking factor, to benefit the application performance.

3.2.4 Storage

Applications are expected to use at least 100x more computational resources within the next few years, but also require similar increases of storage and I/O bandwidth. Parallel file systems of future systems need to support POSIX [1]accesses, MPI-IO, and object storage. One of the major concerns of applications in the exascale is how to manage system reliability and high-chance of a failure of one of the components involved in large execution. In this case and for performance issues, the fault-tolerant techniques of MPI and at the application level will be very important to restart big executions. In order to handle these restarts a fast and persistent local file system would be beneficial in which cases technologies like NVRAM can be a good approach.

The main interest for HPC users is the maximum capacity and performance of the parallel filesystems, as such, the number of users using local disks is reduced. However, having a fast local storage is a good approach for specific I/O that cannot be afforded by the parallel filesystem.

[1] Portable Operating System Interface (http://pubs.opengroup.org/onlinepubs/9699919799/)

Recommendation (10): Following the latest technology developments, a tiered storage high-bandwidth/high-throughput parallel file system should be implemented.

3.2.5 New HPC architectures for exascale machines

Besides HPC accelerators (local GPUs, pool of remote GPUs, many-cores, etc.), other types of processing elements such as Field Programmable Gate Arrays (FPGA), Digital Signal Processors (DSP), etc. have also been proposed for various dedicated applications. However, the specificities of their programming model have limited their adoption so far, emphasizing again the need of providing portable and standard programming models. The situation must be reassessed in light of the new developments underway, which integrate these devices more tightly with the rest of the system's resources.

New architectural approaches for enabling near-data processing, such as processing-in-memory or in-network-processing, could help mitigate data transport bottlenecks and reduce data transport. The latter could help to improve energy efficiency, as an increasingly larger fraction of the consumed power is spent on data transport.

Another interesting research area that could lead to disruptive changes in system architectures is photonics especially if new interconnects standards like Gen-Z[2] will allow to export easily pools of accelerators, remote memory or I/O. Full optical switches supporting new optical networks could completely change the way the system's resources (processors, memory and storage) are organized.

EuroHPC should already be paving the way to exascale computing, taking into consideration upcoming technologies and especially the results of European efforts in various layers of system architecture, from new power-efficient processors to novel compute node architectures. These developments should find their way already in the pre-exascale and petascale systems, foreseeing for example fraction of these systems to be developed based on novel experimental platforms.

Recommendation (11): EuroHPC systems should evaluate the adoption were possible, of alternative experimental architectural approaches, at least in some percentage of the systems, if these are rendered beneficial for a large fraction of European application domains. The systems might not be implement a homogenous (monolithic) configuration, but allow the combination of different components or technologies.

3.3 Energy efficiency

Moving to exascale implies more powerful compute nodes in a larger count. This transition has to take into account the power consumption. Consequently, energy efficiency of compute processors, memory, storage, interconnects, power systems, cooling and power delivery. Likewise, the coding of the applications is a major issue for the design of exascale HPC systems.

Recommendation (12): Due to the power consumption limitation and the size of an exascale or pre-exascale supercomputer, improving energy efficiency is a prerequisite on the path to Exascale. The advances which are necessary in this context need to address:

(1) Improvement of applications in optimising them in parallel for energy efficiency.
(2) Reduction of the "innate" electrical power consumption of the next generation HPC system's hardware infrastructure, including the multiple power conversion steps,

[2] Gen-Z Open Systems Interconnect (https://genzconsortium.org/)

(3) Improvement of applications making them more resilient to system failures

Recommendation (13): The supercomputers should aim at minimizing carbon footprint and environmental impact.

3.4 Operations & Environments

The scientific computing world is fully standardized on Linux today, and it is imperative that any future architecture provides a full Linux environment. This includes a POSIX-compliant file systems, the common development tools with free C/C++/Fortran compilers from GNU and LLVM, as well as potential vendor-provided compilers, parallel debuggers, and all common libraries used in HPC – all of which need to support the latest language standards. Nodes should also support execution of scripting languages such as Perl/Python, and accelerator hardware should provide support for some sort of standardized open API to enable long-term portability of codes.

As scientific computing becomes more diverse, high-end resources will also be expected to cater to much broader categories of users. Queue systems will need to be adapted to handle several orders of magnitude more jobs than previously possible to enable ensembles of simulations as well as new users using many independent tasks e.g. in deep learning or bioinformatics, and we expect rapidly-increasing demands for high-performance computing environments to support containerized applications using e.g. Docker/Singularity with MPI support.

Recommendation (14): Future architectures should exploit well-established open source software solutions. They should provide a mostly Linux based OS and support tools software stack anticipating the evolution of disruptive software technologies by evaluating novel tools, extensions and developments. This includes resource managers with advanced workflow execution capabilities and the support for high-performance oriented software containers.

3.5 Application and User support

Supercomputing centres, with their expertise in computing architectures and programming models, have to recast their service activities in order to support, guide and enable scientific program developers and researchers in refactoring codes and re-engineering algorithms - influencing the development process at its root and making sure that highly-skilled staff develop close relations with key user communities. These services should be provided for longer periods than presently, and the resulting codes adapted to the whole spectrum of supercomputers, with users of the community codes determining what resources are most suitable to apply for. For Europe to have a leading infrastructure, it is also critical that compute centres have staff participating in the development of e.g. programming environments, scientific software, etc...

These trends are not limited to the high end of supercomputing, but apply to all tiers of the HPC ecosystem. The compute node architectures will follow current technology trends and integrate thousands of cores, including accelerators. The effort involved in software refactoring is substantial and requires a deep knowledge of the algorithms, of the codes and, in many cases, of the science at stake. Therefore, a successful practice is to place these activities in community projects where substantial code development activities are common and long-term software development activities can be sustained.

Equally important for the success of EuroHPC infrastructure development, apart from user support, will be the level of competence in systems support. EuroHPC is expected to leverage the existing experience of supercomputing centres in Europe in order to provide stable, high-quality computing services. Investment and development of system level skills (administrators, operators and managers of supercomputing systems) should also be part of Europe's competence agenda for the next years. One important aspect on this direction is the provision of clear career paths and stable working conditions for supercomputing centre technical and support staff.

Finally, for scientists to perceive the European-level organization as their primary community, it is imperative that existing HPC training programs (e.g. those organised from PRACE) are expanded, and that there is a clear progression where users start with introductory courses, continue with more advanced activities, and eventually engage and start to influence the community through masterclass-level education.

Recommendation (15): Applications are essential for the success of EuroHPC infrastructure initiative. It is imperative that HPC users and application developers, are given the necessary training and support in order to exploit the full potential of the envisioned supercomputing ecosystem. They should also be given the opportunity to try new technologies when these become available. Funding of application-centric activities should continue and intensify in the upcoming years.

Recommendation (16): The importance of applications in the acquisition of EuroHPC supercomputers should be reflected by the selection of a proper set of application benchmarks and datasets, to be used for their evaluation. The application benchmarks should correspond to the scientific priorities that Europe will define, promoting the selection of system architectures that are optimal for these scientific domains.

3.6 Services

Today, the main usage model for large-scale HPC systems is still based on the scheduling of batch jobs. This access model on the other hand differs when taking into consideration personal workstations or shared memory servers, where timesharing interactive executions are the norm. The adoption of computational approaches in new scientific domains, the availability of huge amount of scientific data, and the development of advanced and automated workflows, are pushing toward the adoption and the integration on new usage modes and new way to provision high-end computing resources.

Recommendation (17): The EuroHPC supercomputers should aim to integrate range of novel service and tools in order to make them accessible to scientific cases and challenges integrating HPC, Big Data and cloud-like delivery of resources. These among others include services like virtualisation, on-site result visualization, support for containers and support for interactive application execution.

4 Systems Procurement Considerations

4.1 Market consultation

Prior market consultations aim at informing the tender-preparation process by gathering information directly from suppliers, provided that such advice does not have the effect of distorting competition and does not result in a violation of the principles of non-discrimination and transparency.

There are some identified options:

- Interviewing market stakeholders or contacting knowledgeable people in the relevant field, for example independent experts, specialised bodies, business organisations or economic operators.

- An open dialogue event/workshop.

They must also ensure that any information shared with a company as a result of its prior involvement is made available to the other participating companies. It should be designed in way that the fundamental principles of non-discrimination, equal treatment and transparency are safeguarded.

The following measures could help contracting authorities to ensure fair competition and avoid excluding a more advantaged tenderer:

- Openly announcing the preliminary market consultation (e.g. by publishing a prior information notice in national procurement portals and TED);
- Sharing with other candidates and tenderers all relevant information that results from involving one candidate or tenderer in the preparation of the procurement procedure;
- Fixing adequate time limits for the receipt of tenders in order to give all candidates sufficient time to analyse the information.

Recommendation (18): An open market consultation should be run for both Pre-exascale and Petascale systems. The market consultation should cover both technical aspects as well as procedural questions of the procurements.

4.2 Involvement of the Hosting Entities

For each supercomputer, the selected hosting entity needs to be involved in the preparation of the tender. The hosting agreement to be signed between the Joint Undertaking and the Hosting Entity must define the roles and responsibilities between both, first during the procurement procedure and then during the performance of the contract signed between the Joint Undertaking and the vendor.

Recommendation (19): Candidate hosting entities should be invited to take part in the market consultation. This will allow them to prepare for the upcoming procurements and also keep them informed from the early stages of the process.

4.3 Use of Competitive Dialogue vs. Open procedure

Most of the supercomputing centres in the past have not used an open procedure but some form of negotiated procedure. Due to the complexity and uniqueness of the procured solution, it is typically very difficult to precisely specify in advance the final configuration. Instead, the optimal solution to address the needs known upfront is identified during negotiations with a small set of qualified potential suppliers. This approach worked well due to the high level of technical expertise available to the contracting authority.

Both “competitive procedure with negotiation” and “competitive dialogue” have been applied. The latter gives the largest freedom, because it allows both adjusting the technical specifications based on the needs of procurer and negotiating during the meetings with the candidates. Given the tight timeline of EuroHPC, enabling such flexibilities is recommended since this will make the convergence toward the best solution faster.

In Table 1 the advantages and disadvantages of an open procedure and a competitive dialogue are compared. Independently of the procurement procedure, companies based in countries who are part of the WTO[3] are eligible to apply even if the competitive dialogue is divided in two phases, namely a Request for Information (RFI) and a Request for Proposals (RFP) phase.

[3] World Trade Organisation - https://www.wto.org/

Table 1: Comparison of advantages and disadvantages of the competitive dialogue vs. open procedure.

	Open procedure	**Competitive dialogue**
Advantages	• Simple procedure • Execution of the procedure is fast	• Most often used by the Supercomputers centres • Optimal solution to address needs known upfront is identified during negotiations with a small set of qualified candidates • Largest freedom to adjust the technical specifications based on the needs of procurer during the negotiations
Disadvantages	• The need/solution has to be defined very precisely and therefore tender preparation could take more time • Less commonly used by the Supercomputers centres in Europe • Empty tender if none of the candidates can meet the requirements	• Takes typically more time than an open procedure • Documentation of the dialogue sessions

Recommendation (20): Considering the innovative nature of EuroHPC procurements and based on best practices in Europe, a competitive dialogue is the recommended approach for Pre-exascale systems. For petascale supercomputers the common practice of the supercomputing centres is to follow a competitive dialogue, although this approach might prove to be too complex for the joint procurement.

4.4 Diversity of procured solutions

Aiming for a diversity of solutions for realising the pre-exascale systems (and later the exascale systems) will have several advantages. It will significantly reduce the risks for the EuroHPC pre-exascale program as technical and commercial roadmaps of suppliers may change at short notice.[4] Ensuring multiple supply chains for pre- and exascale technologies will promote competition. Furthermore, diversity and co-designing of architecture solutions will typically help to broaden the portfolio of applications that can make efficient use of the pre-exascale systems.

The pre-exascale systems could be purchased either totally separately or with a single procurement with two lots. It is necessary to take into account that:

- the supercomputers can be different (architectures and technologies);
- the supercomputers will be located different places following to the selection of the Hosting Entities, in compliance with the EuroHPC Regulation;
- the procurement could lead a vendor to be in a dominant position

With independent procurements, it would not be possible to ensure the procurement of solutions being as diverse as possible. It will therefore be recommendable to organise the procurement of pre-exascale systems as a single procurement with separate lots. This will allow for a coordinated assessment of the bids for each lot, while finally different contracts for different lots are being awarded.

Recommendation (21): For the procurement of the pre-exascale systems, a single procurement with different lots should be foreseen to facilitate diversity of the procured solutions. The rationale for a division of the procurement in separate lots shall be described in the tendering package and this division will be based at least on the place of the Hosting Entity in which the system will be installed.

Recommendation (22): As the main aim for the acquisition of the petascale supercomputers is to raise the capacity of national supercomputing centres a separate procurement should be implemented for each supercomputer.

4.5 Hosting Entity requirements

EuroHPC system procurements, both for pre-exascale and petascale machines, will be carried in close collaboration with supercomputing centres that will actually host and operate these systems. The selection of the appropriate hosting entities will be a crucial process, decisive for the overall success of the Joint Undertaking strategic goals. The selected hosting sites should demonstrate, based on a historical track record, a certain level of competence and assure the ability both in terms of infrastructure and manpower, to operate supercomputing systems, as well as provide a comprehensive set of services to end users.

The transition to exascale implies more powerful compute nodes in a larger count, but it has to take into account the electricity costs of hosting sites that must remain affordable. Consequently, energy efficiency (from the compute node to the full datacenter power/cooling infrastructure is a major issue for the design of exascale HPC systems. However, this topic is an essential motivation in all system components development (compute processors, memory, storage, interconnect, power system, cooling and power delivery, etc.).

Users should become engaged in the supercomputing sites efforts to reduce their ecological impact should taking into account the energy consumption of their runs. This requires the availability of appropriate monitoring and profiling tools as well as users training on how to use them and how to apply energy-oriented optimisations to their codes.

Recommendation (23): Based on the system performance targets, the size and the expected evolution of the HPC technologies state of the art, the selected hosting sites for the EuroHPC systems should ensure at least the following requirements:

- UPS power available to cover the critical systems including storage and access to data of the JU system.
- Power capacity and power quality for hosting a system in the range of 10 to 15 MW total consumption for the pre-exascale supercomputers.
- Enough capacity of air or liquid cooling
- At least 700 m² for the pre-exascale systems and 150 m² for the petascale systems of contiguous floor space available for hosting the supercomputer and auxiliary systems
- Flooring construction able to bear at least 2200 kg/m² distributed load
- At least 100 Gbit/s connectivity towards the rest of the GEANT Network (link capacity) for the Pre-exascale systems and 50 Gbit/s for the Petascale ones.
- Hosting physical security.
- Hosting fire mitigation equipment/procedures.

- Hosting up-to-date IT access security.
- Ability to perform at least a Level 1 measurement quality for a Top500 submission
- On call service support teams for IT issues during office hours
- A dedicated on-call service team for facilities issues during office hours
- The supercomputers should be operational and accessible to users 24/7.
- Regularly measure the satisfaction of the users with the service via a user survey

Recommendation (24): Candidate hosting entities should demonstrate a successful tracking record on user support activities and offer a high-level of competence in porting and running scientific applications.

Recommendation (25): Hosting sites should take into consideration where ever possible the implementation of sustainable solutions like heat re-usage and system cooling using "green" energy sources etc. aiming to minimise the operational Power Usage Effectiveness (PUE).

5 OVERALL RECOMMENDATIONS

5.1 Performance targets and number of systems

Based on the initial goals set by the EuroHPC initiative to close the gap with the leading HPC regions in the world and in particular the target for acquiring supercomputing systems that would position Europe at the forefront of the global supercomputing race, the pre-exascale machines should be capability systems aiming an aggregated performance level of at least 150 Petaflops sustained performance. Petascale machines should aim for 10-50 Petaflops sustained performance. This would provide a wide coverage for both the top capability computing in the leading class category and the needs of certain classes of applications requiring smaller peak performance.

5.2 Proposed Petascale and pre-Exascale machine architectures

The architectures of the EuroHPC JU supercomputers should aim to diversify by implementing different approaches with the ultimate goal to satisfy a wide spectrum of application domains. While several architectural options still remain open for the Exascale area in terms of processor, accelerators, memory and I/O subsystems, the multi-petascale and the pre-Exascale production systems planned to be deployed by the EuroHPC Joint Undertaking should in principle take into consideration 2 different kind of hardware architectures.

The systems should consider a mixture of thin and fat node configurations in order to accommodate different memory requirements.

Both implemented architectures should cater for optimal interconnect topologies ensuring high-bandwidth, low latency capabilities; covering the demanding performance requirements of the majority of HPC applications including the ability to scale to full system executions.

Through their diversity, these systems should be able to cover the needs of both capability (execution of large scale applications spanning potentially to all the system) and capacity (execution of coupled multiscale/multiphysics applications, ensemble optimization and uncertainty quantification studies). Especially for what concerns the accelerated based system, it should be able to accommodate a mixture of HPC and HPDA/AI complex workflows orchestrated by a flexible resource manager able to support coexistence of batch and interactive stream access mode.

5.3 Timeline for acquisition and operation of the systems

5.3.1 Pre-exascale supercomputers

The steps and indicative times for the procedure from publication to expected start of the mandate for the selected Hosting Entities for pre-exascale are in the table below:

Selection of HE Pre-exascale milestones	Date and time or indicative period
Call for Expression of Interest Publication	Jan 2019
Submission of applications	Mar 2019
Evaluation and notification of results	Apr 2019
Signature of Hosting Agreements	May-June 2019
Hosting Site Preparation	June 2019 – June 2020
Site adaptation to procured system	July 2020 – Sep 2020

For what concerns the procurement of the pre-exascale supercomputers the following table provides a guideline of the dates to follow in order to ensure that the EuroHPC systems will be operational by the end of 2020.

Pre-exascale procurement milestones	Date and time or indicative period
Market Consultation with Vendors	May 2019
National and EuroHPC JU Commitment budget commitments for acquisition of pre-exascale supercomputers	May-June 2019
Preparation of Tender for Supercomputers with JU - HE	June-July 2019
Launch of the call for Procurement of Pre-exascale Supercomputers	July 2019
Evaluation of Final bids and tender award	May 2020
Construction & Installation of Supercomputer	June to Oct 2020
Expected Installation of the Supercomputer in Hosting Entity	October 2020
Expected Supercomputer Acceptance	November 2020
Start of Operations	End of 2020

5.3.2 Petascale supercomputers

The steps and indicative times for the procedure from publication to expected start of the mandate for the selected Hosting Entities for petascale are in the table below:

Selection of HE petascale milestones	Date and time or indicative period
Call for Expression of Interest Publication	Jan 2019
Submission of applications	Mar 2019
Evaluation and notification of results	Apr 2019
Signature of Hosting Agreements	May-June 2019
Hosting Site Preparation	June 2019 – March 2020
Site adaptation to procured system	April 2020 – May 2020

For what concerns the procurement of the petascale supercomputers the following table provides a guideline of the dates to follow in order to ensure that the EuroHPC systems will be operational by July 2020.

Petascale procurement milestones	**Date and time or indicative period**
Market Consultation with Vendors	May 2019
National and EuroHPC JU Commitment budget commitments	May-June 2019
Preparation of Tender for Supercomputers with JU - HE	June-July 2019
Launch of the call for Procurement	July 2019
Evaluation of Final bids and tender award	March 2020
Construction & Installation of Supercomputer	March to June 2020
Expected Installation of the Supercomputer in Hosting Entity	June 2020
Expected Supercomputer Acceptance	July 2020
Start of Operations	July 2020