H2020-FETHPC-2014

Coordination of the HPC strategy

EXDCI

European eXtreme Data and Computing Initiative

Grant Agreement Number: FETHPC-671558

D4.1
First Report on cross-cutting issues and disruptive innovation strategies

Final

Version:	1.0
Author(s):	François Bodin, University of Rennes 1 Carlo Cavazzoni, CINECA Giovanni Erbacci, CINECA
Contributors:	Guillaume Colin de Verdière, CEA Jacques-Charles Lafoucrière, CEA Franz-Josef Pfreundt, Fraunhofer Mario Rosati, CINECA
Date:	25.08.2016

Project and Deliverable Information Sheet

EXDCI Project	**Project Ref. №: FETHPC-671558**	
	Project Title: European eXtreme Data and Computing Initiative	
	Project Web Site: http://www.exdci.eu	
	Deliverable ID: D4.1	
	Deliverable Nature: Report	
	Dissemination Level: PU	**Contractual Date of Delivery:** 31 / 8 / 2016
		Actual Date of Delivery: 25 / 8 / 2016
	EC Project Officer: Beatrice Marquez-Garrido	

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

Document Control Sheet

Document	**Title: First Report on cross-cutting issues and disruptive innovation strategies**	
	ID: D4.1	
	Version: 0.9	**Status:** draft
	Available at: http://www.exdci.eu	
	Software Tool: Microsoft Word 2013	
	File(s): EXDCI-Deliverable-Template-Word2013-1.1.docx	
Authorship	**Written by:**	François Bodin, University of Rennes 1 Carlo Cavazzoni, CINECA Giovanni Erbacci, CINECA
	Contributors:	Guillaume Colin de Verdière, CEA Jacques-Charles Lafoucriére, CEA Franz-Josef Pfreundt, Fraunhofer Mario Rosati, CINECA
	Reviewed by:	Natale Curatolo, PRACE aisbl Guy Lonsdale, SCAPOS
	Approved by:	MB/TB

Document Status Sheet

Version	Date	Status	Comments
alpha	25/7/2016	Draft	Initial contributes
0.9	06/08/2016	Draft	Draft version ready for internal review
1.0	25/08/2016	Final version	

Document Keywords

Keywords:	EXDCI, Exascale, disruptive technology, cross-cutting

Table of Contents

References and Applicable Documents

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

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

[3] http://www.etp4hpc.eu

[4] EESI, http://www.eesi-project.eu

[5] http://www.eesi-project.eu/wp-content/uploads/2015/05/EESI2_D2.5_Final-report-on-co-design-centres.pdf

[6] Workshop "*Machine learning at Scale*" during SC 2015, http://ornlcda.github.io/MLHPC2015/ ISC 2016 Kenote Talk by Andrew Ng (Baidu Research)

[7] http://www.exascale.org/bdec/

[8] http://openmp.org/wp/about-openmp/

[9] https://www.open-mpi.org/about/members/

[10] https://hadoop.apache.org/who.html

[11] Wilde et al. 2015 "*Taking Advantage of Node Power Variation in Homogenous HPC Systems to Save Energy*"

[12] Inadomi et al 2015 "*Analyzing and Mitigating the Impact of Manufacturing Variability in Power-Constrained Supercomputing*"

[13] F. Beneventi et al. 2014 "*An effective gray-box identification procedure for multicore thermal modeling*". IEEE Transactions on Computers, 63(5):1097–1110.

[14] P. Hammarlund et al. 2014 "*Haswell: The fourth-generation Intel core processor*" IEEE Micro, (2):6–20.

[15] Borghesi et al. 2016 "*Predictive Modeling for Job Power Consumption in HPC Systems*"

[16] Sirbu et al 2016 "*Predicting system-level power for a hybrid supercomputer*"

[17] Shoukourian et al 2015 "*Predicting Energy Consumption Relevant Indicators of Strong Scaling HPC Applications for Different Compute Resource Configurations*"

[18] https://repo.anl-external.org/repos/PowerMonitoring/trunk/bgq/

[19] http://www.montblanc-project.eu/

[20] http://www.deep-er.eu/project

[21] https://github.com/LLNL/msr-safe

[22] http://www-micrel.deis.unibo.it/monitoring/wordpress/

[23] http://geopm.github.io/geopm/

[24] https://tu-dresden.de/die_tu_dresden/zentrale_einrichtungen/zih/forschung/projekte/hdeem

[25] http://powerapi.sandia.gov/

[26] https://hal.inria.fr/hal-00799904 (XKaapi)

[27] http://starpu.gforge.inria.fr/ (StarPU)

[28] https://en.wikipedia.org/wiki/Non-volatile_memory

[29] Dong Li, Jeffrey S. Vetter, "*A Survey Of Architectural Approaches for Managing Embedded DRAM and Non-volatile On-chip Caches*", IEEE Transactions on Parallel and Distributed Systems, 23 May 2014. IEEE computer Society Digital Library. IEEE Computer Society, http://doi.ieeecomputersociety.org/10.1109/TPDS.2014.2324563

[30] http://events.linuxfoundation.org/sites/events/files/slides/Wheeler_LinuxForum_Korea_2013.pdf

[31] https://www.arm.com/

[32] http://www.nextplatform.com/2016/01/21/light-at-the-end-of-the-silicon-photonics-tunnel/

[33] https://newsroom.intel.com/news-releases/intel-completes-acquisition-of-altera/

[34] https://ec.europa.eu/digital-single-market/en/news/european-commission-will-launch-eu1-billionquantum-technologies-flagship

[35] https://ec.europa.eu/programmes/horizon2020/en/news/eight-new-centres-excellence-computingapplications

[36] Seth Lloyd, Silvano Garnerone, Paolo Zanardi. Quantum algorithms for topological and geometric analysis of data. Nature Commun. 7, 10138 (2016)

[37] Lloyd, S., Mohseni, M. & Rebentrost, P. Quantum algorithms for supervised and unsupervised machine learning. Preprint at http://arxiv.org/abs/1307.0411 (2013)

[38] Rebentrost, P., Mohseni, M. & Lloyd, S. Quantum support vector machine for big feature and big data classification. Phys. Rev. Lett. 113, 130503 (2014)

[39] Lloyd, S., Mohseni, M. & Rebentrost, P. Quantum principal component analysis. Nat. Phys. 10, 631–633 (2014)

[40] Alejandro Perdomo-Ortiz, Neil Dickson, Marshall Drew-Brook, Geordie Rose, Alán Aspuru-Guzik. Finding low-energy conformations of lattice protein models by quantum annealing. Sci. Rep. 2, 571 (2012)

[41] R. Li, R. Di Felice, R. Rohs, D. Lidar, work in progress

[42] [9] Keith A. Britt and Travis Humble. High-Performance Computing with Quantum Processing Units. arXiv:1511.04386v1 (2015)

[43] Thomas Häner, Damian S. Steiger, Mikhail Smelyanskiy, and Matthias Troyer. High Performance Emulation of Quantum Circuits. arXiv:1604.06460v1 (2016)

[44] B. P. Lanyon, J. D. Whitfield, G. G. Gillett, M. E. Goggin, M. P. Almeida, I. Kassal, J. D. Biamonte, M. Mohseni, B. J. Powell, M. Barbieri, A. Aspuru-Guzik and A. G. White. Towards quantum chemistry on a quantum computer. Nat. Chem. 2, 106 (2010).

[45] Nicolas P. D. Sawaya, Mikhail Smelyanskiy, Jarrod R. McClean, and Alán Aspuru-Guzik. Error sensitivity to environmental noise in quantum circuits for chemical state preparation. arXiv:1602.01857v2 (2016)

[46] Mikhail Smelyanskiy, Nicolas P. D. Sawaya, and Alán Aspuru-Guzik. qHiPSTER: The Quantum High Performance Software Testing Environment. arXiv:1601.07195v2 (2016)

[47] Ryan Babbush, Peter J. Love and Alán Aspuru-Guzik. Adiabatic Quantum Simulation of Quantum Chemistry. Sci. Rep. 4, 6603 (2014)

[48] P. J. J. O'Malley, R. Babbush, I. D. Kivlichan, J. Romero, J. R. McClean, R. Barends, J. Kelly, P. Roushan, A. Tranter, N. Ding, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, E. Jeffrey, A. Megrant, J. Y. Mutus, C. Neill, C. Quintana, D. Sank, A. Vainsencher, J. Wenner, T. C. White, P. V. Coveney, P. J. Love, H. Neven, A. Aspuru-Guzik, J. M. Martinis. Scalable Quantum Simulation of Molecular Energies. arXiv:1512.06860v1 (2016)

[49] Nikolaj Moll, Andreas Fuhrer, Peter Staar, Ivano Tavernelli. Optimizing qubit resources for quantum chemistry simulations in second quantization on a quantum computer. arXiv:1510.04048v3 (2016)

[50] Borzu Toloui and Peter Love. Quantum Algorithms for Quantum Chemistry based on the sparsity of the CI-matrix. arXiv:1312.2579v2 (2013)

[51] IBM Research Report (RC25482)

[52] https://www.nersc.gov/news-publications/nersc-news/nersc-center-news/2015/shifter-makes-container-based-hpc-a-breeze/

List of Acronyms and Abbreviations

BDEC	Big Data and Extreme scale Computing
BSP	Bulk Synchronous Programming
CAPEX	Capital expenditure
CoE	Centre of Excellence
CPU	Central Processing Unit
DDR	Double Dynamic Rate (volatile) memory used in a computer
DNN	Deep Neural Network
ETSI	European Telecommunications Standards Institute
FET-HPC	Future and Emerging Technologies - High Performance Computing
FPGA	Field Programmable Gate Array
GPU	Graphics Processing Unit
HPC	High Performance Computing
HPDA	High Performance Data Analytics, usually implying a mix of HPC and DA (Data Analytics)
IaaS	Infrastructure as a Service
ISA	Instruction Set Architecture
KNL	Knights Landing: 2nd Generation Intel® Xeon Phi™ Processor
ML	Machine Learning
MRAM	Magnetic Random Access Memory
NVM	Non Volatile Memory
OPEX	Operational expenditure
PaaS	Platform as a Service
PCM	Phase-Change Memory
PRAM	Phase change Random Access Memory
QC	Quantum Computing
SaaS	Software as a Service
SATA	Serial Advanced Technology Attachment
SRA	Strategic Research Agenda
SSD	Solid State Device
SWOT	Initials of strengths, weaknesses, opportunities, and threats
TRL	Technology Readiness Level ranges from 1 (principles observed) to 9 (proven system)
VM	Virtual Machine

Executive Summary

Deliverable D 4.1 "*First report on cross-cutting issues and disruptive innovation strategies*", presents the first year activity of EXDCI Task 4.1 "*Cross cutting issues & Disruptive innovations*".

Cross cutting issues are defined as issues common to many layers of a system and which are going to strongly influence the design of future HPC and exascale systems. As analyzed during the EESI 2 EU Project [4] these issues clearly justify co-design approaches for exascale computing.

However, the translation into practice remains extremely difficult due to the scattering of the ecosystem over multiple actors with disjoint constraints and interests. On the other hand, disruptive innovation, whether in usage, in algorithms, in software or in hardware may lead to a more competitive system design.

Thus, on the path toward exascale it is of strategic importance to analyze the impact of cross cutting issues while taking into account potential disruptions, identifying implementation strategies that minimize both the technical as well as the economic risk.

This report analyses three cross-cutting issues in conjunction with nine disruptive innovation strategies and for each of them proposes recommendations.
The cross-cutting issues analysed are:

1. Extreme Computing and Data Analytics Convergence
2. Standardization Efforts
3. Monitoring, analysis and data collection framework/tools in real time and at high frequency.

The analysed disruptive innovation strategies are in the areas:

1. Task based programming practices
2. Non Volatile Memory
3. Byte Addressable IO
4. ARM based Architectures
5. Photonics in Silicon
6. Reconfigurable Supercomputing
7. Integration and synergies between Quantum Computing and HPC
8. Virtualization and Containers Technologies
9. HPC Cloud.

These lists are not exhaustive and will be completed in further studies during the second year of the EXDCI Project.

The analysis presented in this Deliverable is based on three main criteria: TRL (Technology Readiness Level), potential impact on roadmaps and exascale systems, SWOT analysis.

1 Introduction

Contrary to the development of Petascale machines that was a direct extension of the Terascale machines, the race to Exascale is taking place in an extremely volatile context. On the one hand, the energy and resilience requirements (cross-cutting issues) are pushing technology to the limits. On the other hand, managing the data deluge requires the integration of HPC exogenous technology and practices mostly originated in the Big Data community (disruptive innovations) [1], [2].

As a consequence, the design of future exascale systems must integrate the cross-cutting issues while at the same time consider new capabilities. For many aspects of this design it is very unlikely that it will be possible to produce efficient systems without considering emerging hardware and software technologies such as new programming environments or non-volatile memories [3].

This document proposes an extension of the EESI2 efforts on disruptive and cross-cutting issues, exploring new directions and recommendations which were not considered within EESI, [4].
The work here is intended to identify core items that may strongly impact the design of future systems. It is not conceived to be a state of the art analysis or to be exhaustive. For each item we propose a SWOT analysis and a set of recommendations. This is complementary to the ET4HPC SRA-v2[1] and hopes to provide insights for the next SRA.

The work documented in this deliverable represents the WP4 activity done in the first year of the EXDCI Project, a deeper and more complete analysis will be achieved during the next year activity.

1.1 Organization of the Document

This report is organized as follows: Section 2 considers a subset of cross cutting issues and proposes a set of approaches to address them. Section 3 deals with disruptions most likely to have an impact in the 2020 time frame. The last section summarizes this report in a few recommendations.

[1] ETP4HPC, The Strategic Research Agenda, http://www.etp4hpc.eu/en/sra.html

2 Cross-Cutting Issues

Cross cutting issues are defined as topics that are common to many layers of an exascale system (e.g. energy constraints). Cross-cutting issues are the main reason justifying co-design approaches for exascale computing. In the context of EXDCI we extend this definition to issues such as standardization since this strongly influences the development of new systems.

By their nature, cross-cutting issues are very related to the co-design approaches that have been proposed for crafting future systems [5]. However, while they are to some extent easy to identify they nevertheless pose very serious interdisciplinary organizational issues. Furthermore, the issues are interrelated. For instance, a very energy efficient system may prove to be impossible to program. The importance of separate issues must be balanced because optimizing one specific cross-cutting issue may impact negatively on another one (e.g. energy versus programmability). In other words, cross-cutting issues must be considered in the light of a CAPEX (capital expenditure) / OPEX (operational expenditure) analysis with evaluation of the "global efficiency".

The following cross-cutting issues have been selected for analysis in the context of EXDCI. We have left out issues such as energy or resilience that have already received quite some attention in EESI1/2 [4]:

1. Extreme Computing and Data Analytics Convergence;
2. Standardization Efforts;
3. Monitoring, analysis and data collection framework/tools for real time and at high frequency.

They are presented in the remainder of this section.

2.1 Extreme Computing and Data Analytics Convergence

Currently, the Big Data community and the HPC community are mainly active in separate spheres of operation, have different interests and "speak" different languages [6]. Hadoop and Spark, which represent to some extent the tool-chain used in the analytics community, try to make the use of the software/hardware systems as easy as possible and topics like performance and energy consumption do not play a major role. Most Hadoop installations have less than ten nodes.

But things are slowly changing and as compute performance and I/O become bottlenecks the convergence happens [7]. Problems like graph and uncertainty analysis together with data growth are true challenges and define the regime of High Performance Data Analytics. Deep learning has become a serious performance challenge in the industry as Deep neural networks (DNNs) are becoming larger and the technology is very broadly used. The main obstacle here is the difficult balance of achieved flop rates and communication which leads to strong limitations in scalability. At the same time Deep Learning gets better with increasing amounts of data causing challenges in the I/O subsystem. As a consequence, typical hardware systems are quite small. The best indication that this is a place where the stakeholders should talk more with each other is, for example, when they must use an MPI DL application that is started within a Spark/Hadoop tool-chain using GPUs and Infiniband as communication network.

SWOT analysis

Strengths	**Weaknesses**
Large investments by industry and research High impact on industry	Scalability problems in Deep Learning A few very strong players (Google, …)
Opportunities	**Threats**
Algorithmic/scalability challenges that can be attacked by the HPC community to stimulate interaction Deep Learning goes HPC: new applications on large HPC machines	New methods that solve the same problems with less computer power Impact on HPDA technology

Recommendations

- Stimulate the interaction between the HPC Community as a whole and the analytics research community
- Stimulate algorithmic research
- Make large data sets available for the research community
- Synchronization with Centres of Excellence looking at HPDA i.e. CoeGSS (http://coegss.eu)

2.2 Standardization Efforts

International standards de-facto-standards[2] are strongly influencing research experiments, further developments, and industrial products. In the HPC domain they provide long-term visibility for developers, and are warrantors of a widely shared and agreed on basis for further developments. As such, they are the most used basis for large software and hardware products.

Contrary to some application domains such as represented by ETSI or the OpenFOAM initiative (now driven by the ESI group[3]), the European HPC community is not much involved in technological-oriented standards specifications efforts. For instance, in OpenMP [8] only ARM is a Permanent Members of the Architecture Review Board (ARB) (13 members), the most influential committee of the association. Europe's representation is stronger at the Auxiliary Members of the ARB (14 members): BSC, Bristol University, EPCC, and RWTH Aachen University.

In the case of OpenMPI [9], European technology providers are more present, with ARM and Bull being contributor to the initiatives.

[2] We use standards in the reminder for both kinds.
[3] http://www.esi-group.com/

Regarding the Big Data community, whose technology is becoming crucial to future HPC applications, the EU stakeholder influence is close to nil. For instance EU stakeholders are not present in the Hadoop Project Management Committee [10]. Overall there is a lack of involvement in technology-oriented initiatives contrary to application-oriented initiatives

This absence has multiple consequences for Europe's position in HPC: First, standards specifications are by construction best suited for the products and technologies represented by the most active members. Not being involved thus means the necessity to adapt to choices and decisions taken by parties following their interests and objectives. Second, not being involved in the specification effort introduces a delay in the acquisition of the corresponding knowledge (it is usually very clear to committee members which direction is being taken year(s) in advance). The effort to catch-up is then important and there is an inherent delay to reach the market.

The low degree of involvements is not due to a lack of competencies in the EU ecosystem. We believe that it is mainly due to a lack of incentive for actions that have long-term effect and require a strong commitment. Indeed, being influential in these bodies, implies the need to allocate time and resources for highly skilled experts to attend meetings, as well as for development engineers to perform the necessary experiments to build the technical contributions.
If the current status is to be changed, i.e., to increase participation in standardization bodies, it is mandatory to define efficient incentives.

SWOT analysis

Strengths	**Weaknesses**
EU has a large body of suitably qualified experts, either in academia or in industry.	Involvement of EU stakeholders in HPC / DB related international standards and initiative is not strong enough to influence evolution.
Opportunities	**Threats**
Many new initiatives responding to current changes are opportunities for newcomers.	Fast acting ecosystems (e.g. US or China) may be in place before the EU community can react or propose alternatives.

Recommendations

Incentives for EU stakeholders to participate in international standardization in the extreme scale computing and big data must be increased. New forms of support must be invented to ensure the presence of high-profile scientists and EU industry stakeholders in existing and emerging initiatives.

2.3 Monitoring, analysis and data collection framework/tools in real time and at high frequency

In an exascale system, the number of active components will increase significantly with respect to current petascale systems. This is mainly due to the fact that the performance of the

elemental components delivering the computational power are not going to improve their performance significantly (e.g. the single FPU performance will be close to a few GFlops, or the core's frequency will remain close to 1-2GHz), so you then need to pack more components to get more aggregate performance. The situation is even worse due to the fact that in modern chips the components can behave quite differently with respect to one another, so heterogeneity is everywhere, with transient time of the order of milliseconds/microseconds (e.g. frequency stepping) [11][12][13][14].

Thus, monitoring and the capacity to analyze data coming from all the components at high frequency becomes a fundamental tool to assess the performance and the efficiency of the machine and to find out opportunities for improving the system throughput and energy to solution for the user application.

Because of the high frequency and the number of components/available counters, the monitoring and analysis of exascale system parameters immediately becomes a big data problem, where "classical" approaches fail to scale.

Moreover, with appropriate APIs applications can also benefit from the information collected by the monitoring framework, both from post mortem trace analysis and runtime adaptive optimization [15][16][17].

There are promising ongoing studies and activities on these topics, both at the EU level (under the FET-HPC program) and also in the individual R&D of many institutions, under nationl/regional funding programs. What is probably missing is an effort to converge to some basic standard to allow planning and deployment of such tools at scale and independently from a specific vendor [18]-[25]. A focus group, including research organizations, computing centres and vendors would be required, leading to a proposal for a coordinated action to work on the standardization aspect. Indeed, today each organization builds its own monitoring framework and spends significant effort to optimize it to the underlying HW and low-level APIs as well as to its own management facility. Oblige all the European institutions receiving Commission support to define common interfaces between each of the software levels and components involved in the monitoring, management and storage of the activity traces will increase significantly the efficiency of each supercomputer centre as well as creating a common ground and data-sharing framework for creating innovation potentials which span all European supercomputer centres.

SWOT analysis

Strengths	**Weaknesses**
HW components built-in sensors accessible from the software stack allow fine grain monitoring of architectural and physical parameters and furthermore gathering of data out of band.	Current monitoring infrastructures lack: a common standard, low-overhead interfaces for fine grain monitoring as well as non-root privilege access.
Opportunities	**Threats**
Fine Grain Monitoring Systems are an enabling technology for dynamic management of performance, energy, power and resiliency of the	HW manufacturer fragmentation as well as system integrator's proprietary monitoring technologies reduces the permeability of those technologies at site level.

supercomputing machine	Lack of common interfaces and standards for connecting the infrastructure level monitored data with machine level ones. Problems of time synchronization in between the different data sources.

Recommendations

- A focus group should be established, including research organizations, computing centres and vendors, leading to a proposal for a coordinated action to work on the standardization aspect.

3 Disruptive Innovations

The HPC landscape is subject to many disruptions. Some are due to technical trends that no longer apply (e.g. processor frequency increase) while others are due to new technologies that are reaching the market (e.g. non volatile memory). Hopefully, these new technologies may help to propose new hardware and software systems capable of handling existing obsolete operational practice, while also helping to design efficient exascale systems. For instance, the introduction of the data analytics capabilities in HPC machine requires new memory systems taking advantage of non-volatile memories and new interconnect capabilities offered by photonics in silicon (to be friendly to data analytics algorithms) as well as a large extension of the HPC software stack to integrate big-data originated libraries (that can be facilitated by the use of containers technology).

Some disruptive innovations / emerging technologies may have an impact on the short-term (e.g. new programming tools) while others are not expected to be available soon (e.g. quantum computing). In this report we mainly consider short-term topics and a few long-term topics that are likely to have a huge impact.

Of course, when considering disruptions, forecasting is extremely delicate and often proven wrong. However, ignoring upcoming innovations is also very likely to produce poor design. In EXDCI we study a set of topics that we see at the core of exascale systems at some point in the future. The purpose of the analysis is to provide some insight on the potential issues and benefits of current (coming) innovations. The list of topics is the following:

1. Task based programming practices
2. Non Volatile Memory
3. Byte Addressable IO
4. ARM based Architectures
5. Photonics in Silicon
6. Reconfigurable Supercomputing
7. Integration and synergies between Quantum Computing and HPC
8. Virtualization and Containers
9. HPC Cloud

It should be noted that this list is not exhaustive and will be complemented in further EXDCI study. In the remainder of the section we propose an analysis based on the following criteria:

- TRL (Technology Readiness Level) analysis
- Potential impact on roadmaps and exascale systems
- SWOT analysis

For each topic we elaborate a recommendation for addressing this topics in the H2020 timeframe.

3.1 Task based programming practices

Task-based parallelism is gaining momentum due to new architectures (KNL Intel Xeon Phi second Generation, GPGPU) and the maturity of runtimes (examples include XKaapi [26] and StarPU [27]). Using task-based programming methods breaks the classical BSP algorithmic

approach[4]. An evaluation of the TRL of such technology is 5-6 since it requires a better integration with standards (OpenMP for example) and compiler technologies clearly needed to make it widely acceptable. Task-based programming methods may lead to a better use of exascale architectures, especially if those architectures are heterogeneous (i.e. having GPU and/or FPGA).

SWOT analysis

Strengths	**Weaknesses**
Allows enhanced parallelism inside an application Allows potential optimal usage of multiple heterogeneous resources inside a supercomputer	Lack of tools that hinders development Lack of standardization
Opportunities	**Threats**
Mechanisms that can be used in HPC and Data Analytics	Numerical schemes are not updated to cope with tasks (various inertia)

Recommendations

- Through peer review processes, make sure that new numerical schemes or algorithms are task-aware from day one, where appropriate. It means that a transverse link between "mathematicians" and language/runtime developers must be maintained.
- Synchronization with the FETHPC projects AllScale (http://www.allscale.eu) and INTERTWinE (http://www.intertwine-project.eu)

3.2 Non Volatile Memory

NVM stands for Non Volatile Memory [28], [29]. NVMs come in many flavors (examples include: PCM, MRAM) that exhibit different properties. They may change the way we think about internal and external storages in a computer. Since some NVMs are already commercially available, the TRL level is 9. Introducing NVM in the architecture of a computer impacts:

1. The potential size of the core memory since NVM has a better GB/€ ratio than common DDR;
2. The ratio between internal and external storage. With large NVM memories, disks (and/or SSDs) might disappear;
3. The design of operating systems will have to cope with information being still available after a power off/on cycle and security issues.

SWOT analysis

Strengths	**Weaknesses**

[4] This could also be listed in the cross-cutting issues sections since it impacts directly numerical methods and algorithms.

HPC builds on a mass market (cell phones, tablets, etc.)	Durability of the NVRAM implementation has to improve beyond current value (3-5000 erase cycles for consumer parts)
Opportunities	**Threats**
Boost the HPDA market with a price reduction of its storage subsystem	Security of the solution

Recommendations

- Include NVM technologies in any EU-based design for future exascale systems
- In EU-based prototypes of exascale systems, request large enough NVM memory sizes to be able to change storage paradigms
- Synchronization with the FETHPC project nextgenio (http://www.nextgenio.eu).

3.3 Byte addressable I/O

A byte addressable I/O subsystem is able to access data one byte (8 bits) at a time [30]. Classical I/O subsystems are accessing data by "pages". The size of the page depends on the underlying file system implementation and tuning. Classically a page is 4KB wide but it can be much larger (several MB). It means that updating only a few bytes in a file requires moving unnecessarily large quantities around. It means also that a lot of software layers could be simplified. New NVMe based SSD already support byte I/O. Therefore TRL is 9. The main impact foreseen is in the database field where storing key-value pairs is hindered by the page access.

SWOT analysis

Strengths	**Weaknesses**
NVMe storage is displacing SATA in mainstream IT systems	Lack of complete software stack adaptation
Opportunities	**Threats**
Impacts Data Analytics and large data bases technologies	Lack of adoption due to developments costs (time, wide scope…)

Recommendations

- Start working on updating the software stack in anticipation of the new hardware capabilities.
- Synchronization with the FETHPC project nextgenio (http://www.nextgenio.eu), ExaNest (http://www.exanest.eu) and Sage (http://www.sagestorage.eu).

3.4 ARM Based Architectures

ARM ltd [31] is a technology provider that offers several IP items:

- An ISA (Instruction Set Architecture), currently dubbed as armv8;
- CPU and GPU designs;
- Libraries of physical designs to build components.

Depending on their needs, companies can choose one of those IP. Companies which have the know-how to build a CPU or a GPU will select the ISA license whereas newcomers to the CPU/GPU market can choose to buy designs and libraries IP for a starter. ARM is the reference in the embedded world. Its use would go beyond the X86 dominance in the HPDA market. TRL is 9 since armv8 implementations are already available by multiple vendors even if not for large scale HPC usage. ARM implementation introduces flexibility on the market since it can easily be tuned to different market segments. It also helps to build more energy aware supercomputers.

SWOT analysis

Strengths	**Weaknesses**
Allows for multiple original implementations Energy efficient technology	Not yet available for HPC nor for Big Data machines
Opportunities	**Threats**
Introduces healthy competition Address the energy efficiency challenge	Maturity of HPC/HPDA implementations Competition might kill this initiative Softbank strategy for ARM

Recommendation

- Implement one or two ARM based systems for HPDA using different implementations to foster a new market as well as show commitment to broad deployment.
- Synchronization with the FETHPC project Mont-Blanc 3 (http://montblanc-project.eu)

Special note:

After the initial writing of this section, the Japanese company Softbank announced the takeover of ARM ltd late July 2016 (http://www.softbank.jp/corp/d/sbg_press_en/). For the foreseeable future, this acquisition doesn't change the recommendations above since Softbank's CEO doesn't plan to change ARM activities. Yet, the evolution of ARM should be monitored by all European HPDA actors to anticipate any (unwanted) inflexion in ARM's evolution.

3.5 Photonics in Silicon

Photonics in silicon [32] is a hardware solution that replaces metal connections with light transmission. Intra-chip and inter-chip connections are potential targets for this technology. Supercomputer architectures as well as their compute elements will be impacted by this technology. The TRL of this technology is difficult to assert but should be around 5. The potential impact of such technology is to decouple memory storage from CPU while reducing energy consumption.

SWOT analysis

Strengths	**Weaknesses**
Long range signaling Energy efficient connectivity	Standards for interoperability
Opportunities	**Threats**
Might be the only viable solution to build 1Eflop in 15MW	Technological challenge of coupling new generations of transistors and optical devices that might delay mass availability

Recommendations

- Propose innovative exascale designs relying extensively on silicon photonics
- Foster a viable silicon photonics ecosystem

3.6 Reconfigurable Supercomputing

At the time of writing of this document, most, if not all, supercomputers are either homogeneous (CPUs only) or heterogeneous using GPUs as accelerators. In both cases, the functional units are fixed i.e. built-in the silicon. A reconfigurable supercomputer, on the other hand, is an heterogeneous architecture using FPGAs (Field Programmable Gate Array) that can be reprogrammed according to the code needs [33].

Using reprogrammable hardware allows for efficient algorithms in specific cases such as data stream filtering (data mining). FPGAs usually outperform CPU implementations in such cases especially considering their GFlops/W specification.

TRL is already 9 for the hardware side, yet on the software side the level is at most 6.

Machine Learning and Data Analytics are two obvious targets for reconfigurable machines. This is one of the reasons that Intel acquired Altera in 2015.

SWOT analysis

Strengths	**Weaknesses**
Best possible energy solution for a dedicated	Hardware that is not general purpose

solution	Long development cycle (at least as of 2016)
Opportunities	**Threats**
Major advances in machine learning (ML) and Data Analytics (DA)	Lack of ease of use that favors other architectures such as GPUs

Recommendations

- Start working on programming models and environments that go beyond OpenCL, which is seen as too low level by most programmers.
- Synchronization with the FETHPC projects EXTRA (https://www.extrahpc.eu) and EcoScale (http://www.ecoscale.eu/)

3.7 Integration and synergies between Quantum Computing and HPC

Quantum Computing is the usage of a controlled quantum system to compute solutions of computational problems. The intrinsic property of the superposition of states of a quantum system is the key feature that allows a quantum computer to compute solutions of complex problems at a speed which is beyond comparison with a typical digital computer, e.g. a quantum computer can in principle solve a problem with factorial complexity with a single instruction. Given that, a full featured quantum computer with a complete instruction set can in theory outperform any digital computer, but the reality is quite far from that. Quantum computers today are able to perform only few instructions (in many cases just one), on a limited number of q-bits, moreover the "CPU" is huge in size and weighs tons, this is because to keep the quantum state under control the chip should be cooled down near to the absolute zero, and thermal fluctuations kill the superposition of states and stop the CPU.
HPC is among those field where quantum computing could be used in synergy with digital computers to speed-up complex numerical kernels. Traditional HPC systems could take care of all the instructions which cannot be executed on the quantum computer (typically I/O, and control code) offloading to the quantum engine by mean of a dedicated library of computational kernels that can be reformulated using the quantum logic.

Then, the quantum computers will not replace the digital computers but most probably will integrate them. In principle one quantum computer could be shared among different applications and users (probably being accessible through a cloud), but for HPC the most probable configuration is the shipping of a quantum engine together with the HPC system. Today, commercially available quantum computers are quite expensive (of the order of 10M€) but the price is less than the price of high-end HPC systems.

The quantum computers could disrupt both the HW configuration of HPC systems, and applications. Clearly a quantum-enabled HPC world will look completely different from what we know today. Programming paradigms will need to incorporate quantum logic, facility management and system administration functions will need to acquire new competences to manage the quantum computers.

Despite there are already quantum computers commercially available (i.e. D-Wave), these are sort of proof-of-concept systems with very limited capability and complex operational procedures. These first example of quantum computers are closer to a cryogenic laboratory than a computational system. Thus, with respect to general HPC applications and workloads the TRL is not greater than 3.

The European Commission recently proposed to make €1 billion available for a Quantum Flagship, a large-scale European research program for quantum technology [34]. At the "Quantum Europe Conference" held in Amsterdam on May 17-18, 2016, academia and industry convened to define a roadmap for quantum technology; the industrial investment in quantum technologies was massively embodied by the giants Intel, Microsoft, Google and IBM, among others.

Some of the proposed activities call for a dialogue between the quantum computing (QC) community and the high-performance computing (HPC) community, well represented also by the eight new centres of excellence for computing applications, recently funded by the European Commission, including biomolecular systems, materials design and energy problems [35].

Quantum technology objectives that resonate with the HPC community are:

- Simulator of motion of electrons in materials;
- Development and design of new complex materials;
- Versatile simulator of quantum magnetism and electricity;
- Simulators of quantum dynamics and chemical reaction mechanisms to support drug design;
- Solving chemistry and materials science problems with special purpose quantum computer > 100 physical qubits;
- General purpose quantum computers exceed computational power of classical computers.

The quantum information community and the HPC community have proceeded so far mostly on independent tracks, but recently bridging works have appeared. Such works address the development of quantum machine learning algorithms [36]-[39] that may revolutionize materials design and genomics, the application of a quantum annealer for a proof-of-concept molecular dynamics simulation [40] and for transcription factor-DNA binding specificity [41], the integration of quantum processing units in current and future HPC systems [42] and the solution of quantum algorithms on classical HPC platforms [43], the implementation of quantum chemistry on quantum computers [44]-[50].

SWOT analysis

Strengths	**Weaknesses**
Perform complex computations at almost infinite speed.	Very limited instruction set. Huge CPU that requires cryogenic cooling and set-up closer to a Physics laboratory than a typical IT installation.
Opportunities	**Threats**
Allow the computation of complex problems whose solution is practically	Application and kernel need to be rewritten using quantum logic instead of digital logic. The

impossible on any digital computers of exa- zetta- and expected scales beyond	facility to host a QC and the competences to manage it are quite different from those of typical IT equipment.

Recommendations

- In the Quantum Technology flagship promote an HPC research line with topics aligned with the R&D (HW and SW) covered by the Strategic Research Agenda of the ETP4HPC under the cPPP framework agreement: CoE, FET projects. This will allow QC to better integrate with HPC.
- Encourage the HPC centres to start an early evaluation of the quantum computer technologies (remote DEMO access is already available).
- Organise joint actions (workshops, symposia, training activities, etc.) to favor the interaction between the Quantum Computing community and HPC community.

3.8 Virtualization and Containers Technologies

During the last decade, hypervisor-based virtualization has been often the topic for experimentation in HPC, because of its potential advantages, in terms of flexibility of management and usage of resources, with respect to conventional HPC environments. Apart from a few simple workloads, the results were very unsatisfactory: the need to also manage hypervisors, the difficulty to provide native access to high performance file systems and networks and to virtualize GPUs and coprocessors, the overhead of the VMs and the unpredictable performance levels have been the main obstacles taking away momentum to the mainstream adoption of virtualized resources in HPC.
In the last two years, in the general IT sector, there has been a widespread adoption of a new virtualization technology: container-based virtualization. Containers are a lightweight virtualization method for running multiple isolated Linux systems under a common host operating system. Container-based computing is revolutionizing the way applications are developed and deployed. However, this revolution has yet to completely reach the HPC community.
Multiple studies have compared containers (i.e. Docker) to VMs; see [5[51] for a comparison based on a Linpack benchmark. Containers typically approach bare-metal performance running 10-1000 times faster than common VMs, and consume less memory since only few processes are needed to manage the workload. In principle, containers can directly use all computational resources available in a system. To fully explore containers' technology potentials, Cray and NERSC are piloting a concrete use case in the scientific domain of cosmology. The pilot will develop a customized software package, called Shifter, that will serve to easily deploy applications container within an HPC environment [52].

The adoption of a container-based virtualization technology in HPC could significantly expand the number of supported applications, and at the same time, reduce the time needed to get applications running on a specific supercomputer. The benefits include a reduced cost of application porting, a more predictable and consistent code behavior across platforms and operating systems, and an improved time-to-results.

Within the general ICT domain, container-based technologies have reached a high level of maturity and adoption, thus raising the TRL to 9. However, since the same level of adoption cannot be claimed for HPC applications, the TRL is between 5 and 7 as valid experiments do exist but have not moved to full production yet.

A widespread adoption of Containers in HPC may:

1. Enable the HPC Platforms to run more applications.
2. Increase flexibility for users.
3. Facilitate sharable and reproducible results.
4. Provide rich, portable environments managed by users in a simplified base system.

SWOT analysis

Strengths	**Weaknesses**
Containerized applications offer comparable performance with respect to bare-metal. All major ICT players are working hard on containers' technology: the rate of innovation and engineering is very high. A large number of containerized applications are already available and widely used.	Currently available containers technology do not provide direct access to underneath high performance storage and network resources inside the containers.
Opportunities	**Threats**
The containers technology adoption could significantly expand the number of applications available in a generic HPC system and dramatically facilitate the sharing of applications and reproducibility of results.	Some potential security issues (escalated privileges).

Recommendations

- EU should support HPC centres and vendors to better integrate the containers' technology into HPC systems, in particular for:
  a) extending the major HPC resource managers to natively support scheduling and management for containers;
  b) allowing, inside the container, a transparent and direct access to high performance subsystems and components (storage / network / accelerators).
- Foster the sharing of competences and expertise through the promotion of dedicated training events and the creation of specific user communities.
- Support the packaging of the most relevant HPC applications within standardized containers.

3.9 HPC Cloud

Today's Cloud Computing environments are built on a mix of technologies, including virtualization, automation and orchestration components. In some circles, virtualization is wrongly considered synonymous of cloud rather than presented as one of its possible components. A Cloud platform is composed by a pool of resources which are able to elastically scale and become accessible on-demand for users, regardless of the presence of any

virtualization layer. The virtualization layer which is common to many cloud environments is subject to performance degradation. While the hypervisor enables the visibility, flexibility and management capabilities required to run multiple virtual machines on a single box, it also creates additional processing overhead that can significantly affect performance.

The bare-metal cloud solutions are a way to offer dedicated servers without virtualization overhead, but with all the flexibility, scalability and efficiency of a traditional cloud infrastructure. In this way it is possible to combine the scalability and automation of virtualized cloud with performance, also easily guaranteeing direct access to High Performance Storage, to specialized networks and to accelerators, all components of an HPC system that are very difficult to virtualize.
Bare-metal instances can be provisioned via a web portal or using an API, providing access to HPC infrastructure on demand.

High-performance, bare-metal cloud functionality is ideal to perform short-term, CPU-intensive, memory-intensive or data-intensive workloads without suffering from latency or overhead delays and while maintaining the same level of flexibility of a standard cloud environment. The adoption of a cloud model without any compromise on performance could expand the potential users of the new HPC systems

Existing technologies are mature enough to justify a TRL between 8 and 9. More effort should be invested in promoting their adoption within the HPC domain.

The delivery of HPC services through a bare metal cloud infrastructure could be an easy way to manage the convergence between traditional HPC workloads and HPDA activities. For example, an Apache Spark cluster would become a PaaS built on top a bare-metal based IaaS.
The submission model should evolve from job queuing and reservations towards an ad-hoc system deployment.
Last but not least, a bare-metal based cloud infrastructure could be the right platform for the extensive adoption of the SaaS paradigm even in the HPC world.

SWOT analysis

Strengths	**Weaknesses**
Bare-metal based Cloud platforms could guarantee a flexible and elastic scalable access to HPC resources while ensuring high performance. A bare-metal based HPC cloud is a simple and elegant way to also support HPDA workloads.	In the bare-metal cloud the minimum resource units to be allocated to the user is the single physical server and the efficient usage of physical resources is fully the responsibility of the user
Opportunities	**Threats**
Bare-metal HPC cloud may broaden the base of users of HPC systems by providing on-demand and scalable compute resources to enable scientific communities and private organizations to run large-scale HPC workloads in the cloud. A bare metal HPC cloud is the right platform for providing	With a bare-metal HPC cloud model, system management would be made by the end user, who might not have enough system administration skills to properly manage and configure the system.

scientific software as a service.	

Recommendations

- Extend HPC resource managers to support the provisioning of bare-metal Cloud machines.
- Explore the possibility to consolidate Cloud and Container technologies together to realize the full potential of available HPC resources by enabling a larger spectrum of use cases and applications.
- Improve existing virtualization technology to keep ensuring resource abstraction while reducing performance penalty.

4 Conclusion and Recommendations

This document analyzes a set of topics likely to have a strong impact on the design of future exascale system as well as on their uses. We have elaborated a set of recommendations that we hope, if implemented, will help the HPC ecosystem to grasp the challenges we are facing with disruption and cross-cutting issues.

For each topic we indicate the FETHPC project and CoE that should be involved in further analysis. A joint effort will be at the core of the EXDCI WP4 work plan for next year.

Here is a summary of the recommendations:

Topic	**Recommendation**
Big Data and Extreme Scale Computing (BDEC)	Stimulate the interaction between the large HPC Community as a whole and the analytics research community. Stimulate algorithmic research. Make large data set available for the research community.
Standardization	Incentives for EU stakeholders to participate in international standardization in the extreme scale computing and big data areas must be increased. New forms of support must be invented to ensure the presence of high-profile scientists and EU industry stakeholders in existing and emerging initiatives.
HPC System Real-Time Monitoring	A focus group should be established, including research organizations, computing centres and vendors, leading to a proposal for a coordinated action to work on the standardization aspects.
Task based programming practices	Through peer review processes, make sure that new numerical schemes or algorithms are task-aware from day one, where appropriate. It means that a transverse link between "mathematicians" and language/runtime developers must be maintained.
Non-Volatile Memory	Include NVM technologies in any EU-based design for future exascale systems. In EU-based prototypes of exascale systems, request large enough NVM memory sizes to be able to change storage paradigms.
Byte Addressable IO	Start working on updating the software stack in anticipation of the new hardware capabilities.
ARM based Architectures	Implement one or two ARM based systems for HPDA using different implementations to foster a new market as well as show commitment to broad deployment.
Photonics in Silicon	Propose innovative exascale designs relying extensively on silicon photonics.

	Foster a viable silicon photonics ecosystem
Reconfigurable Computing	Start working on programming models and environments that go beyond OpenCL, which is seen as too low level by most programmers.
Quantum Computing	In the Quantum Technology flagship promote an HPC research line with topics aligned with the R&D (HW and SW) covered by the SRA of the ETP4HPC under the cPPP framework agreement: CoE, FET projects. This will allow QC to better integrate with HPC. Encourage the HPC centres to start an early evaluation of the quantum computer technologies (remote DEMO access is already available). Organise joint actions (workshops, symposia, training activities, etc.) to favour the interaction between the Quantum Computing community and HPC community.
Virtualization and Containers	EU should support HPC centres and vendors to better integrate the containers' technology into HPC systems, in particular for: a) extending the major HPC resource managers to natively support containers scheduling and management; b) allowing, inside the container, a transparent and direct access to high performance subsystems and components (storage / network / accelerators). Foster the sharing of competences and expertise through the promotion of dedicated training events and the creation of specific users communities. Support the packing of mostly relevant HPC applications within standardized containers.
HPC Cloud	Extend HPC resource managers to support the provisioning of bare-metal Cloud machines. Explore the possibility to consolidate Cloud and Container technologies together to realize the full potential of available HPC resources by enabling a larger spectrum of use cases and applications. Improve existing virtualization technology to keep ensuring resource abstraction while reducing performance penalty.

5 Annex: TRL

Technology Readiness levels (**TRL)** describe the maturity of a system from fundamental research (TRL 1) to a product on the market (TRL 9). The table below lists all levels of the TRL. A system can be in our context, a compiler, a new programming language, a new CPU or any other technologies related to HPDA.

Level	Definition
1	Base principle observed
2	Technological concept formulated
3	Experimental proof of the concept
4	Key functions of the concept validated in a laboratory
5	Technology components validated in a representative environment
6	Demonstration of a prototype in a representative environment
7	Demonstration of a prototype in an operational environment
8	Full system qualified by tests and demonstrations
9	Full system validated in a real environment

See http://www.eesi-project.eu/wp-content/uploads/2015/05/EESI2_D6.1R_Final-report-on-operational-software-matutity-level-technology.pdf for more details.