Journal of Computational Science 14 (2016) 74–77

Contents lists available at ScienceDirect

Journal of Computational Science

journal homepage: www.elsevier.com/locate/jocs

BSC Best Practices in Professional Training and Teaching for the HPC Ecosystem

Maria-Ribera Sancho

Barcelona Supercomputing Centre and Universitat Politècnica de Catalunya, Barcelona, Spain

ARTICLE INFO

Article history:
Received 23 August 2015
Received in revised form 7 December 2015
Accepted 10 December 2015
Available online 13 December 2015

Keywords:
HPC Ecosystem
HPC professional training strategies
HPC centre–University relationship

ABSTRACT

This paper outlines the key components of the European HPC ecosystems, analyses the major challenges as well as corresponding specific technical focus areas which needs to be addressed in European strategic research agenda for HPC leadership. Further the need for education and training is clearly identified and the BSC approach, BSC model and best practices are presented and analyzed. Further a generalization and analysis of the approach are given.

1. The Context: HPC Ecosystem

The High Performance Computing (HPC) eco-system has three major components: computing infrastructure, technological developments and applications. Pre-exascale (post-petascale) systems and future exascale computers are expected to have an ultra large scale and highly hierarchical architectures with nodes of many-core processors and accelerators ([1], [2]). That implies that existing systems, languages, programming paradigms and parallel algorithms would have, at best, to be adapted, and often become obsolete. To manage these ultra large scale parallel systems, new adaptive runtime systems, allowing managing huge distributed heterogeneous data, huge distributed load balanced tasks, minimizing the energy consumption, and with fault resilient properties will become necessary. New scientific methods and scalable algorithms for these systems will have to be introduced enabling to run applications with increasingly large-size. Many issues related to the emergence of post-petascale computing are all linked together and define the framework of High Performance Computing and Computational Science. To meet the challenges of this interdisciplinary, complex and highly strategic field, it is imperative and urgent to form a new generation of scientists through education and training activities of which the international dimension is an important feature. This is further emphasized by the latest published findings in USA in terms of needs for education and training specifically in HPC and Compute Intensive Science. [3]

The European Technology Platform for High Performance Computing (ETP4HPC) http://www.etp4hpc.eu/, is a Public Private Partnership that is a focal point in defining the Strategic Research Agenda (SRA) in HPC in Europe [1]. It proposes a roadmap to develop the HPC technology ecosystem and European leadership for the benefit of all HPC users, both academic and industrial ones. It has been prepared by a broad team of experts taking into account various inputs and the European strategic position. It sets objectives for the different technologies required to produce leading-edge HPC supercomputers up to the exascale level [1].

The major challenges to be addressed as identified by ETP4HPC are:

- System Architecture level challenges: future HPC platform architectures, energy cost and power consumption, concurrency and data locality, I/O latency and bandwidth, extreme scaling: from sub-component to total system, resilience, reliability, availability and serviceability.
- System Software level challenges: scalability, modularity, robustness, virtualization capability, increased system heterogeneity, awareness of data-movement cost
- Programming environment challenges: hierarchical models, data distribution and locality, performance analytics, novel parallel algorithms, data-movement cost, code migration and re-writing.

HPC usage models challenges: dealing with Big Data, increasing heterogeneity of data, HPC workloads in cloud computing, etc.

The technical focus areas that address the above challenges are specified below ([1] and [4]):

E-mail address: maria.ribera@bsc.es

http://dx.doi.org/10.1016/j.jocs.2015.12.004
1877-7503/

- HPC System Architecture and Components
- Energy and Resiliency
- Programming Environment
- System Software and Management
- Balance Compute, I/O and Storage Performance
- Big Data and HPC usage models
- Mathematics and Algorithms for Extreme scale HPC systems
- Extreme Scale Demonstrators

In order to address the above exascale computing challenges a sustained effort in HPC and related areas ranging from the development of novel multi- and many-core architectures, to new programming models, scalable algorithms, and new modelling techniques and paradigms is required. For example, HPC and Computational Science have an important role to play at graduate level education as well as at professional training level. The critical need to educate the next generation of computational scientists has been widely recognized in both the European [5] and American [6] contexts. As noted there, personnel shortages in computational science and HPC are a product of historical trends. Funding for these areas declined sharply in the 1990s following the end of the cold war and did not recover until about 2002 when readily available HPC clusters lead to wide-scale scientific advancements. During the decline, the number of university programs in computational science plummeted and young people who might have gone into this field opted instead for employment in areas such as large scale Internet and web computing. This has led to a greying of the talent pool, with a large number of computational scientists being within 10 years of retirement, just at a time when our needs are the greatest.

IDC surveys prepared for the EC [7] and for US Department of Energy [6] has identified clearly the inflection points and skills gap in Computational Science and HPC as follows: Parallelism, and how to use it, Petascale/Exascale computing, HPC system heterogeneity, HPC system architectural balance, HPC system reliability, HPC system and data center power and cooling.

The IDC update to their report for the European Commission prepared this year shows no significant differences in terms of talent gap while pointing to significant increase in investment in HPC technology [8]. To reach the above objectives and address successfully the skills gap there is an important ingredient lacking, namely an adequate supply of well-rounded and highly trained HPC and computational scientists with the ability to understand complex scientific problems and the skills in mathematical modelling, simulation, Big Data techniques, and HPC to address them. Indeed this is well understood also by the EC funding bodies and Education and Training is in the heart of the research agenda in the next H2020 program period. [2]

2. The Model: BSC Education and Training strategy for the EU HPC Ecosystem

2.1. Training at Barcelona Supercomputing Centre (BSC)

BSC mission to educate is reflected in an extensive training and education program providing training and professional skills for researchers in academia and industry, from computer science and the supercomputing (SC) user scientific communities. Moreover the centre aims to engage and capture talent at undergraduate level through various internship and scholarship programs and is educational and hub for young researchers with HPC career path.

As part of the Partnership for Advanced Computing in Europe (PRACE) consortium and one of the designated PRACE Advanced Training Centres (PATC) BSC, http://www.bsc.es/PATC, is working to develop appropriate European HPC professional training curricula, available beyond the PRACE partnership. Education and Training team at BSC is building curricula based on cutting-edge scientific research and development of models, software tools and simulation environments for high-performance computing and application areas, targeting research communities and industry with supercomputing needs. PATC HPC curriculum as envisaged by the PATC OMB (Operational Management Board) that caters for users with varying degrees of HPC competence/knowledge. One may enter the curriculum at any level according to their skills, and by having achieved the learning outcomes of a course at that level, if required to progress to courses at higher levels (which often demands skills from courses at lower levels as pre-requisites). The PATC OMB are aware that training is a demand driven learning and the attendees have current need requiring a course with practical outcomes to increase the level of performance of the taught subject. The curriculum was designed to showcase the skills basis, creating understanding of the HPC ecosystem at entry point through to acquiring the rights level of competence to use the SC technology advanced level.

The basic level (L1) is designed for users with little knowledge of fundamental HPC concepts and parallel programming that have some programming experience. Those coming from non-computer science background may be encouraged to attend entry-level preparatory courses (e.g. scientific programming and computing), if necessary. The intermediate level (L2) for users who have at least some experience of developing parallel code and are familiar with fundamental HPC concepts (e.g. having achieved the learning outcome of L1 courses) and the advanced level "L2" is aimed at users who are scaling up or moving to Tier-0 systems[1]; again the learning outcome of courses are typically pre-requisites. [9]

At BSC specialized training courses in HPC related subjects and seminars organized by the Education and Training team are carried out annually:

- PRACE Advanced Training Center Courses
- RES (Spanish Supercomputing Network) training sessions
- Project training initiatives
- CUDA Center of Excellence courses
- Workshops
- Summer Schools

The largest program with around 14 training courses is that running by the PRACE Advanced Training Centre at BSC which mission is to carry out and coordinate training and education activities that enable the European research community to utilize the computational infrastructure available through PRACE. The long-term vision is that such centres will become the hubs and key drivers of European high-performance computing education.

The students trained through the program can benefit from the hands-on sessions and communication with internationally renowned instructors. Participants are issued a Certificate of Attendance and in the case of BSC PATC, the Master in Innovation and Research in Informatics program in the Barcelona School of Informatics is crediting the courses taken by its students as seminars counting towards their optional subjects. PATC courses are attended by specialists from across Europe and all lectures and materials are in English. All BSC PhD students are encouraged to take the courses on the program as part of their professional development. The PATC courses cover different topics and offer several training levels. BSC is one of the PATC offering not only core HPC subjects but as well courses with relevance to the scientific community such as Introduction to

[1] Tier-0 systems, e.g. the 3 to 6 most advanced computing systems in Europe available for deployment and benchmarking of prototypes at petascale level as well as porting, optimizing and peta-scaling of applications. (http://www.prace-ri.eu/)