Report on cHiPSet Training School 2016

NEW TRENDS IN MODELING AND SIMULATION IN HPC SYSTEMS

Bucharest, Romania, 21-23 September 2016

The training school within COST Action IC1406 featured presentations and hands-on practice and demonstration of novel methods, mechanisms, techniques and technologies in Modelling and Simulation (MS), with a special emphasis on evaluation of HPC Systems. Today MS is widely considered the essential tool in science and engineering to substantiate the prediction and analysis of complex systems and natural phenomena. MS offers suitable abstractions to manage the complexity of analyzing Big Data in various scientific and engineering domains. Unfortunately, Big Data problems are not always easily amenable to efficient MS over HPC. Also, MS communities may lack the detailed expertise required to exploit the full potential of HPC solutions, and HPC architects may not be fully aware of specific MS requirements. Thus, the goal of the training school was to offer to participants, PhD students and Early Stage Researchers, coming from these two worlds, the skills to understand and work with models and concepts coming from HPC, to design accurate modeling and simulation strategies for the evaluation of HPC solutions, to design, construct and use complex MS tools that capture many of the HPC modeling needs, from scalability to fault tolerance and beyond. At the end, participants learned how to efficiently turn massively large HPC data into valuable information and meaningful knowledge, with the help of covered new trends in MS.

Description of courses

The logical structure of the Training School programme, in terms of reached topics and subjects, allowed participants to understand and grasp concepts related to the terminology, properties and the models used to evaluate HPC systems using modeling and simulation. This was continued with a set of lectures and hands-on exercises, on tools to evaluate using modeling and simulation systems and applications, either related to distributed processing infrastructures, Cloud systems, or various other applications. In the end, the Training School ended with lectures on what does it mean to develop good evaluation methodologies.

The first day of the Training School was dedicated to classes on the theoretical implications of the evaluation using modeling and simulation as the instrument-of-choice, for HPC systems. Two lectures covered from the theoretical aspects behind *performance evaluation*, to the use of tools for the analysis of complex HPC (and HTC) applications.

The first course, on “High-Performance Computing: Gossip, Lies, & Secrets”, was taught by *Horacio González-Vélez*, from National College of Ireland, Ireland. The training course covered an introduction into the principles and methods for High-Performance Computing. It made trainees familiar with the tools to develop HPC applications, and form the set of skills for them to understand the pitfalls and

subtle details behind optimizing such applications when running them on large distributed infrastructures.

In fact, Horacio González-Vélez is a world-wide expert in this topic. Today he directs the NCI's cloud infrastructure, postgraduate programmes, and research with a clear emphasis on innovation, education inclusion, and industrial partnerships. In the past, he spent over a decade working in engineering and product marketing for innovation-driven companies such as Silicon Graphics and Sun Microsystems. Award-winning lecturer and researcher, Horacio has also carried out applied research in parallel and distributed computing, funded by a number of public and industrial organizations including the European Commission FP6 & FP7, UK NESTA, the Scottish Funding Council, NVidia, Dell, and Microsoft. He is also an accredited expert on cloud computing for the BCS, the UK Chartered Institute for IT. He has been recognized with the European Commission ICT award for his efforts on scientific dissemination and the UK NESTA Crucible Fellowship for his inter-disciplinary research on computational science. He holds a PhD in Informatics from the University of Edinburgh.

The second course, on "Leveraging Modeling and Simulations in HPC Infrastructures via Increased Usability", was taught by *Sandra Gesing*, from University of Notre Dame, USA. The lectures approached the development of applications over Science gateways. Modeling and simulations, which necessitate HPC infrastructures, are often based on complex scientific theories and involve interdisciplinary research teams. IT specialists support with the efficient access to HPC infrastructures. They design, implement and configure the simulations and models reflecting the sophisticated theoretical models and approaches developed and applied by domain researchers. Roles in such interdisciplinary teams may overlap dependent on the knowledge and experience with computational resources and/or the research domain. Domain researchers are mainly not IT specialists and the requirement to employ HPC infrastructures via command line often forms a huge hurdle for them. Thus, there is the need to increase the usability of simulations and models on HPC infrastructures for the uptake by the user community.

Science gateways also called portals, virtual research environments or virtual labs form a solution, which offer a graphical user interface tailored to a specific research domain with a single point of entry for job and data management hiding the underlying infrastructure. In the last 10 years quite a few web development frameworks, containerizations, science gateway frameworks and APIs with different foci and strengths have evolved to support the developers of science gateways in implementing an intuitive solution for a target research domain. The selection of a suitable technology for a specific use case is essential and helps reducing the effort in implementing the science gateway by re-using existing software or frameworks. Thus, a solution for a user community can be provided more efficiently. Additionally, novel developments in web-based technologies and agile web frameworks allow for supporting developers in efficiently creating web-based science gateways.

The topic science gateways and related technologies have gained also importance in the last 10 years for the HPC community. The first time in the history of such solutions, providers of HPC, grid and cloud infrastructures have reported in 2014 that more of their resources have been accessed via science gateways than via command line. The US National Science Foundation (NSF) has recommended a Science Gateway Community Institute for funding, which will provide services starting in July 2016. Additionally, IEEE launched a technical area on science gateways as part of the Technical Committee on Scalable Computing.

During this class, Dr. Sandra Gesing showed a demo on the MoSGrid science gateway. The MoSGrid science gateway is a portal for the computational chemistry community enabling researchers to create, edit, invoke and monitor molecular simulations prepared via workflows targeting quantum chemistry, molecular dynamics and docking tools. The intuitive user interface does not require any prior knowledge of the research domain and serves as example for a workflow-enabled portal with enhanced distributed data management features exploiting HPC infrastructures. The complex infrastructures are

hidden from the users via intuitive user interfaces. All necessary files and access to the system were provided by the Trainer, for a hands-on tutorial to explore the system and receive molecular structures via integrated visualization.

Dr. Sandra Gesing is a research assistant professor at the Department of Computer Science and Engineering and a computational scientist at the Center for Research Computing at the University of Notre Dame, USA. Her research interests include science gateways especially for bioinformatics applications and distributed and parallel computing. In this context, she also works on analysis frameworks for modeling and simulations. She is heavily involved in the proposed US Science Gateway Community Institute, where her role focuses on outreach and community engagement.

The second day of the Training School was dedicated to hands-on sessions, with classes on the use of tools for MS of HPC Systems. Two widely-used tools were selected for "dissection": CloudSim was the topic of the first lecture, and SimGrid was the one of the second lecture of the day.

The third course, on the "Evaluation of Cloud Systems", was taught by *Florin Pop*, from University Politehnica of Bucharest, Romania. The course covered the fundamental skills for a practitioner working in the field of Cloud Systems to have, for the development of a correct methodology for the evaluation using simulation of Cloud services and components. The hands-on activities focused on the use of a well-known MS tool, CloudSim, and trainees were given the chance to try for their own and complete a series of assignments designed to make them familiar with an evaluation process designed over modeling and simulation.

Florin Pop is Professor within the Computer Science Department at UPB, and also an active member of Distributed System Laboratory. His research interests are in scheduling and resource management (decentralized techniques, re-scheduling), multi-criteria optimization methods, Grid middleware tools and applications development (satellite image processing an environmental data analysis), prediction methods, self-organizing systems, contextualized services in distributed systems.

The fourth lecture, on "Challenges and solutions in Simuling Clouds", was taught by *Marc Frincu*, from West University of Timisoara, Romania. The course focused on the theory and hands-on behind some of the most widely used tools for simulating a Cloud environment. Hands-on activities were done with the Trainers over SimGrid, once widely used simulator for HPC and distributed systems.

Dr. Marc Frincu is an assistant professor with the Department of Computer Science at UVT working on cutting edge topics related to clouds, smart grids, and Big Data. He received his PhD from UVT in 2011 working on Adaptive Scheduling for Distributed Systems. His full thesis (in English) can be downloaded from here. Prior to rejoining UVT in 2015, he was a postdoctoral research associate at the University of Southern California working with prof. Viktor Prasanna on smart grids and clouds. At UVT he leads the CER research group focusing on applying cloud computing to areas such as smart grids and Big Data.

Finally, the third day of the Training School was dedicated to the presentation of ups and downs in choosing the right steps for performing scientific evaluations of HPC systems.

As such, the fifth lecture, on "Performance evaluation and analysis of large scale distributed systems - Issues, Trends, Problems and Solutions", was taught by *Eleni Karatza*, from Department of Informatics, Aristotle University of Thessaloniki, Greece. As large scale distributed systems such as grids and clouds offer computational services to scientists, consumers and enterprises, there are important issues that must be addressed for large scale distributed systems, such as: performance, resource allocation, efficient scheduling, energy conservation, reliability, cost, availability, quality. Furthermore, due to the

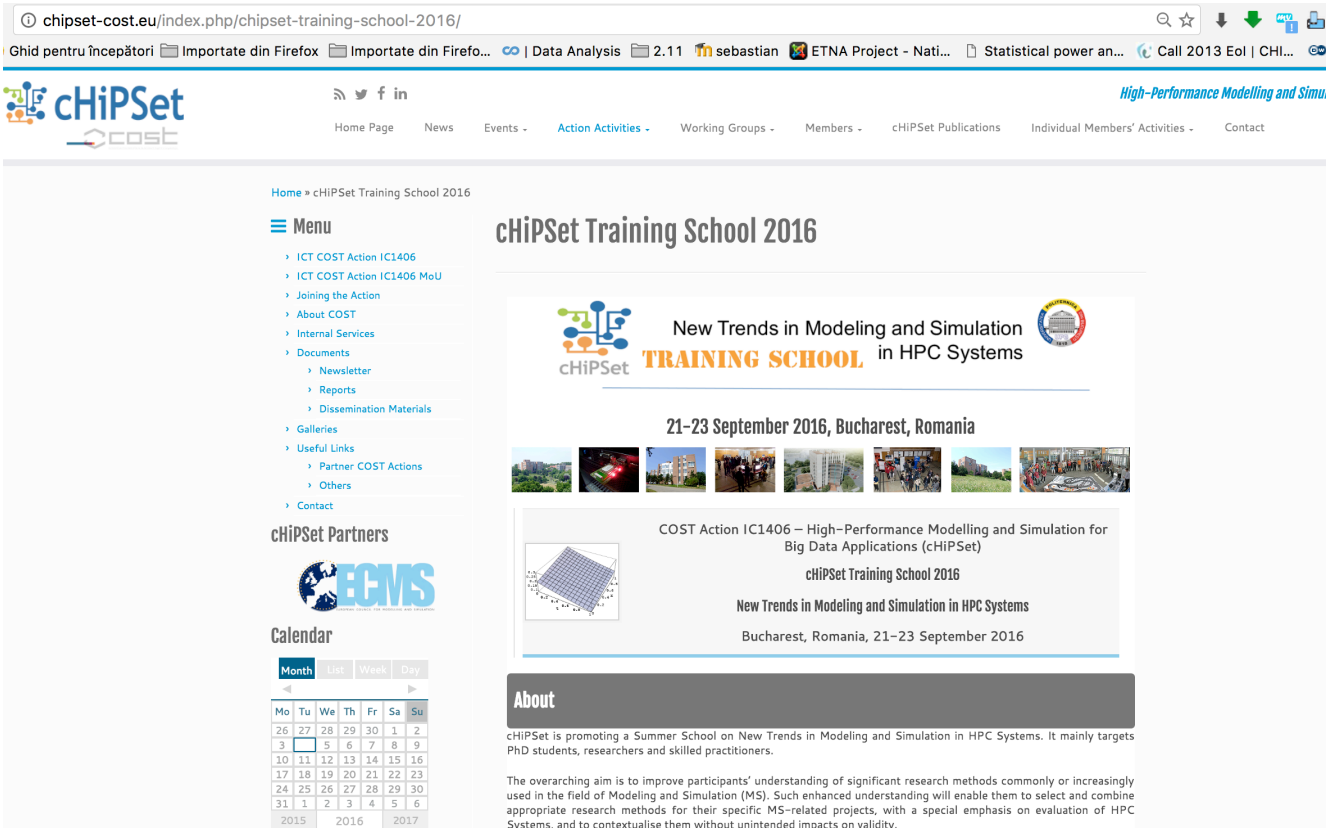
cost of electrical power consumption and the environmental impact, energy efficiency in large scale systems is a global IT concern.

Effective management of distributed resources is crucial to use effectively the power of these systems and achieve high system performance. Resource allocation and scheduling is a difficult task in large scale distributed systems where there are many alternative heterogeneous computers. The scheduling algorithms must seek a way to maintain a good response time along with energy-efficient solutions that are required to minimize the impact of grid and cloud computing on the environment. Furthermore, the simultaneous usage of computational services of different distributed systems such as clusters, grids and clouds can have additional benefits such as lower cost and high availability.

In this talk prof. Eleni Karatza presented state-of-the-art research covering a variety of concepts on HPC systems such as grids and clouds, based on existing or simulated grid and cloud systems, that provide insight into problems solving and we will provide future directions in the grid and cloud computing area. Advanced modelling and simulation techniques are a basic aspect of performance evaluation that is needed before the costly prototyping actions required for complex distributed systems. Queuing network models of large scale distributed systems will be described and analysed and performance metrics will be presented. Complex workloads will be examined including real time jobs and scientific workflows.

Eleni Karatza is a Professor in the Department of Informatics at the Aristotle University of Thessaloniki, Greece. Dr. Karatza's research interests include Computer Systems Modeling and Simulation, Performance Evaluation, Grid and Cloud Computing, Energy Efficiency in Large Scale Distributed Systems, Resource Allocation and Scheduling and Real-time Distributed Systems. She has authored or co-authored over 200 publications including four papers that earned best paper awards at international conferences. She is senior member of SCS, IEEE and ACM, and she served as an elected member of the Board of Directors at Large of the Society for Modeling and Simulation International (2009-2011). She has served as General Chair, Program Chair and Keynote Speaker in International Conferences. She has also been Editor and Guest Editor in International Journals.

The webpage of the event:



Scientific Programme of the Event:

Day 1 - Wednesday 21 September 2016

09:00 – Participants registration

09:15 – Official Welcome

09:30 – High-Performance Computing: Gossip, Lies, & Secrets (trainer: Horacio González-Vélez, National College of Ireland, Ireland) – part I

11:30 – Coffee break

11:45 – High-Performance Computing: Gossip, Lies, & Secrets (trainer: Horacio González-Vélez, National College of Ireland, Ireland) – part II

13:15 – Lunch

14:00 – Evaluation of Cloud Systems (trainer: Florin Pop, University Politehnica of Bucharest, Romania) – part I

16:00 – Coffee break

16:15 – Evaluation of Cloud Systems (trainer: Florin Pop, University Politehnica of Bucharest, Romania) – part II

17:45 – End of Training Sessions for the Day

Day 2 - Thursday 22 September 2016

09:00 – Participants registration

09:30 – Leveraging Modeling and Simulations in HPC Infrastructures via Increased Usability (trainer: Sandra Gesing, University of Notre Dame, USA) – part I

11:30 – Coffee break

11:45 – Leveraging Modeling and Simulations in HPC Infrastructures via Increased Usability (trainer: Sandra Gesing, University of Notre Dame, USA) – part III

13:15 – Lunch

14:00 – Challenges and solutions in Simulating Clouds (trainer: Marc Frincu, West University of Timisoara, Romania) – part I

16:00 – Coffee break

16:15 – Challenges and solutions in Simulating Clouds (trainer: Marc Frincu, West University of Timisoara, Romania) – part II

17:45 – End of Training Sessions for the Day

Day 3 - Friday 23 September 2016

9:30 – Performance evaluation and analysis of large scale distributed systems - Issues, Trends, Problems and Solutions (trainer: Eleni Karatza, Department of Informatics, Aristotle University of Thessaloniki, Greece) – part I

11:30 – Coffee break

11:45 – Performance evaluation and analysis of large scale distributed systems - Issues, Trends, Problems and Solutions (trainer: Eleni Karatza, Department of Informatics, Aristotle University of Thessaloniki, Greece) – part II

13:15 – Lunch

14:00 – Round table, discussions on topics related to Evaluation of HPC using Modeling and Simulation

14:30 – Group work, final examination

Afternoon – free time for sightseeing (more details on potential tours will be available online)

List of participants

Trainees:

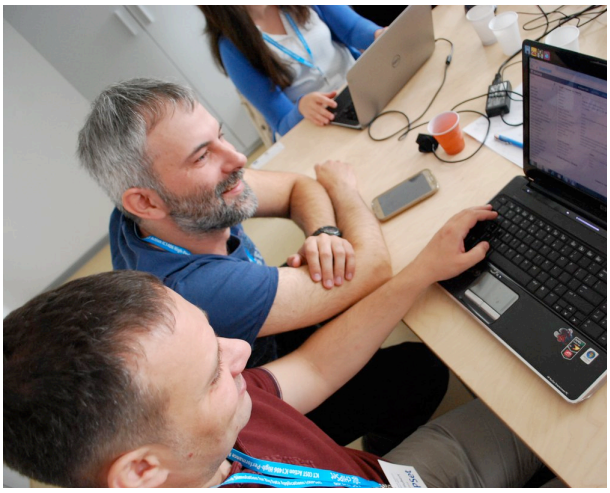
1. Qian Wang, Software Research Institute, Athlone IT, Ireland
2. Hubert Zarzycki, Wroclaw School of Information Technology, Poland
3. Mateusz Krzysztoń, Warsaw University of Technology, Poland
4. Olivera Novović, BioSense Institute, University of Novi Sad, Serbia
5. Fátima Manuela da Silva Leal, University of Vigo, Spain
6. Bruno Veloso, University of Vigo, Spain
7. Ilias Mavridis, Aristotle University of Thessaloniki, Greece
8. Jacek Czerniak, Casimir the Great University, Poland
9. Agnieszka Jakóbik, Cracow University of Technology, Poland
10. Pablo Padron Santana, University of La Laguna, Spain
11. Georgios Skourletopoulos, University of Nicosia, Cyprus
12. Adriana E. Chis, National College of Ireland, Ireland
13. Daniel Grzonka, Cracow University of Technology, Poland
14. Mauro Iacono, Seconda Università degli Studi di Napoli, Italy
15. Fabio Tordini, UNITO, Italy
16. Paweł Szynkiewicz, Polish Academy of Science, Poland
17. Milena Georgieva, Burgas Free University, Bulgaria
18. Chitra Balakrishna, Edge Hill University, UK
19. Marcel Antal, Technical University of Cluj-Napoca, Romania
20. Claudia Pop, Technical University of Cluj-Napoca, Romania
21. Elena Apostol, University Politehnica of Bucharest, Romania
22. Gabriel Apostol, University Politehnica of Bucharest, Romania
23. Luminica Borcea, University Politehnica of Bucharest, Romania
24. Radu-Ioan Ciobanu, University Politehnica of Bucharest, Romania
25. Cristian Chilipirea, University Politehnica of Bucharest, Romania
26. Vlad Dragoi, University Politehnica of Bucharest, Romania
27. George Iordache, University Politehnica of Bucharest, Romania
28. Catalin Gosman, University Politehnica of Bucharest, Romania

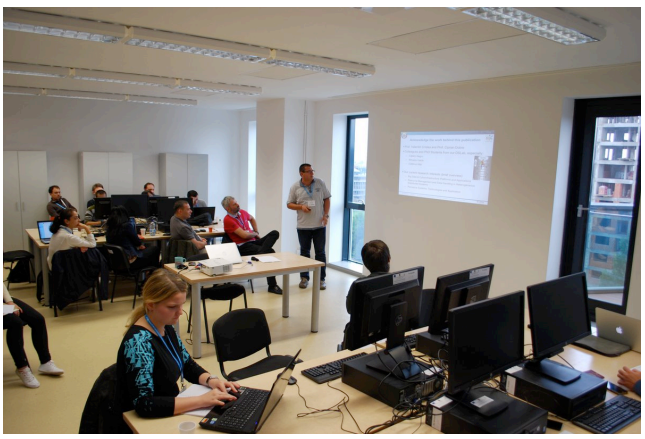
29. Catalin Leordeanu, University Politehnica of Bucharest, Romania
30. Radu-Corneliu Marin, University Politehnica of Bucharest, Romania
31. Bogdan Mocanu, University Politehnica of Bucharest, Romania
32. Alexandra Mocanu, University Politehnica of Bucharest, Romania
33. Andreea Chilipirea (Petre), University Politehnica of Bucharest, Romania
34. Valeriu Stanciu, University Politehnica of Bucharest, Romania
35. George Suciu, BEIA, Romania
36. Ciprian Truica, University Politehnica of Bucharest, Romania
37. Mihaela Vasile, University Politehnica of Bucharest, Romania
38. Liviu Hirtan, University Politehnica of Bucharest, Romania
39. Alexandru Gherghina, University Politehnica of Bucharest, Romania
40. Vlad Ciobanu, University Politehnica of Bucharest, Romania
41. Loredana Groza, University Politehnica of Bucharest, Romania
42. Catalin Negru, University Politehnica of Bucharest, Romania

Trainers:

1. Sandra Gesing, University of Notre Dame, USA
2. Eleni Karatza, Aristotle University of Thessaloniki, Greece
3. Horacio González-Vélez, National College of Ireland, Ireland
4. Marc Frincu, West University of Timisoara, Romania
5. Florin Pop, University Politehnica of Bucharest, Romania
6. Ciprian Dobre, University Politehnica of Bucharest, Romania

A selection of photos taken during the event:





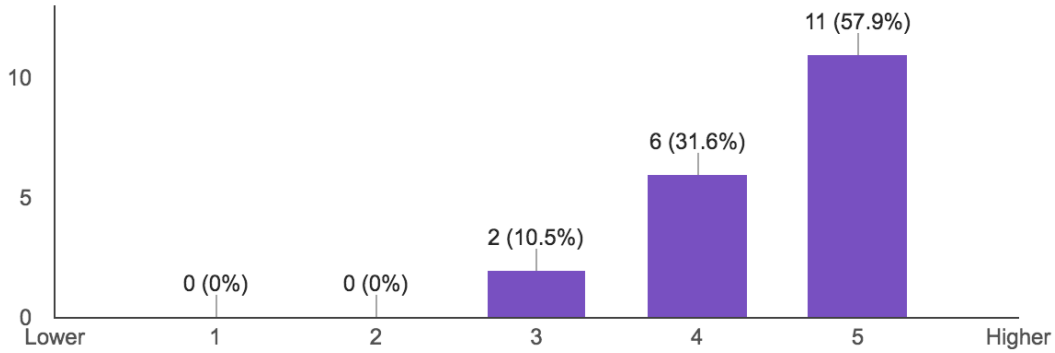
After the event, the participants were asked to fill in an online Evaluation Form. The participants were asked to grade the quality of the programme (scientific quality) and/or the (local) organization of the Training School. They were asked to give suggestion to improve the next editions, where they gave answers such as:

- The addition of a project within the summer school central topic would improve the results in terms of knowledge acquirement by students.

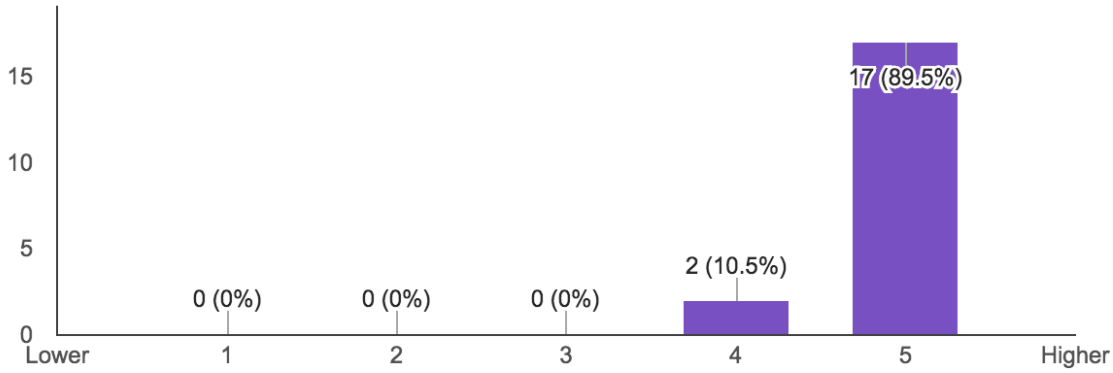
- Industry representatives can be invited to support the training. For example they can provide HPC, cloud platform (Microsoft Azure, IBM Bluemix) and/or demonstrate a technology. In-depth understanding of exemplary subject can be achieved by introducing a hackathon for a one or two days.
- The training school was well organized and for the future my only suggestion would be to have some more practical hands-on applications.

A summary of the grades receive is:

Please grade the quality of the programme in the Training School. (19 responses)



Please grade the organization of the Training School. (19 responses)



Would you consider participating again in the next edition(s) of this Training School?
(19 responses)