

## **cHiPSet Training School 2016**

---

### **New Trends in Modeling and Simulation in HPC Systems**

Bucharest, Romania, 21-23 September 2016

#### **Goal**

---

The training school within COST Action IC1406 features 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 is to offer to participants 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 will be able to efficiently turn massively large HPC data into valuable information and meaningful knowledge, with the help of covered new trends in MS.

#### **Scientific Programme**

---

**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)

## Description of courses

---

<b>Course title</b>
1. High-Performance Computing: Gossip, Lies, & Secrets
<b>Trainer</b>
<b>Horacio González-Vélez</b> National College of Ireland, Ireland Horacio.Gonzalez-Velez@ncirl.ie
<b>Description of the training course</b>
The training course will cover an introduction into the principles and methods for High-Performance Computing. It aims to make 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.
<b>Trainer Biography</b>
Horacio González-Vélez directs the NCI's cloud infrastructure, postgraduate programmes, and research with a clear emphasis on innovation, education inclusion, and industrial partnerships. 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 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.

<b>Course title</b>
2. Evaluation of Cloud Systems
<b>Trainer</b>
<b>Florin Pop</b> University Politehnica of Bucharest, Romania florin.pop@cs.pub.ro
<b>Description of the training course</b>
The course will cover 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.
<b>Trainer Biography</b>
Florin Pop received his PhD in Computer Science at the University POLITEHNICA of Bucharest in 2008. He received his MSc in Computer Science in 2004 and the Engineering degree in Computer Science in 2003, at the same University. He is Associate Professor within the Computer Science Department 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.

**Course title**

3. Leveraging Modeling and Simulations in HPC Infrastructures via Increased Usability

**Trainer**

**Sandra Gesing**

University of Notre Dame, USA  
sandra.gesing@nd.edu

**Description of the training course**

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, a demo will take place on the MoSGrid science gateway [7]. 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 will be provided for a hands-on tutorial to explore the system and receive molecular structures via integrated visualization. An open discussion will suggest measurements for developers deciding which technology is suitable for a community under consideration of the community's preferred tools, methods and available data. The intention of the open discussion is to create a catalogue of requirements on developing science gateways starting with a set of suggestions assembled beforehand.

### **Trainer Biography**

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.

Her research resulted in over 70 scientific publications. Additionally, she has been editing over 10 special issues or proceedings, respectively. The successful European workshop series IWSG (International Workshop on Science Gateways) has been founded by her in 2009 and she has guided since. IWSG is the partner workshop series of the US workshop series on science gateways that exists for over 10 years and the Australian IWSG-A, which has been founded in 2015. The significance of science gateways has been also recognized by IEEE and she chairs and coordinates the IEEE technical area on science gateways. She has been chairing international workshops and has been serving on dozens of international conference and workshop program committees. Supporting open-access initiatives, she is an academic editor of the open-access journal PeerJ Computer Science.

Prior to the position at Notre Dame, she was a research associate in the Data-Intensive Research Group at the University of Edinburgh, UK, in the area of data-intensive workflows and in the Applied Bioinformatics Group at the University of Tübingen, Germany, in the area of science gateways and grid computing. Additionally, she has perennial experience as a project manager and system developer in industry. As head of a system programmer group, she has led long-term software projects (e.g. infrastructure on web-based applications). She received her German diploma in computer science from extramural studies at the FernUniversität Hagen and her PhD in computer science from the University of Tübingen, Germany.

Online CVs:

<http://sandra-gesing.com/cv/>

<http://www.adjacentgovernment.co.uk/stakeholders/dr-sandra-gesing/25440/>

### **Course title**

4. Challenges and solutions in Simulating Clouds

### **Trainer**

**Marc Frincu**

West University of Timisoara, Romania

[mfrincu@info.uvt.ro](mailto:mfrincu@info.uvt.ro)

### **Description of the training course**

The course focuses on the theory and hands-on behind some of the most widely used tools for simulating a Cloud environment. The activities are done over

SimGrid, once widely used simulator for HPC and distributed systems.

### **Trainer Biography**

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.  
More details available at: <http://web.info.uvt.ro/~mfrincu/>.

### **Course title**

5. 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  
[karatza@csd.auth.gr](mailto:karatza@csd.auth.gr)

### **Description of the training course**

Large scale distributed systems such as grids and clouds offer computational services to scientists, consumers and enterprises. However, 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 we will present 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.

### **Trainer Biography**

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.

Professor Karatza 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.  
For additional details: <http://agent.csd.auth.gr/~karatza/>