EXDCI Case Studies: HPC talent generation
Henri Doreau: CEA Research Engineer

Henri Doreau is as a Research Engineer at CEA, the French Alternative Energies and Atomic Energy Commission. As a software developer, he contributes to numerous projects in high-performance storage and big data domains. Away from the keyboard, Henri nurtures a passion for literature and likes oenology, cooking, running and cycling.

EXDCI is a European-funded project led by PRACE and ETP4HPC, the two most significant HPC bodies in Europe. It aims to co-ordinate the development and implementation of a common strategy for the European HPC ecosystem by supporting road-mapping, strategy-making and performance-monitoring activities.

This is one of a series of case studies designed to demonstrate the range of interesting careers in High Performance Computing (HPC). More case studies are available at http://www.exdci.eu/CaseStudies/

Henri, tell us a bit about yourself and how your career path has led you to where you are now.

I grew up in Angers, in the Loire valley (France), where I also studied electronics and computer sciences at ESEO engineering school.

I have a strong interest in software development, and especially free software. In the past, I used to focus mostly on IT security, and discovered HPC through an internship at CEA[1]. I applied out of curiosity, enjoyed it, and I am still in the same team today! Back then, my goal was to implement scalable command execution techniques within the ClusterShell library[2]. It matched perfectly my inclination to push IT to the limits and design algorithms that "do more".

Free software communities (and the nmap security scanner project in particular) had taught me to adopt a rigorous approach to try to produce elegant and robust code. I realised that this was even more relevant within CEA due to the nature of the concerns behind HPC.

I enjoyed the culture of excellence, learning things every day, working on challenging problems, and the dual research and engineering aspect of the work. Free software is really important to me, and the fact that CEA values it too made it a perfect place for me to work.

---

[1] http://www.cea.fr/english-portal
[2] https://github.com/cea-hpc/clustershell

So what does your job involve?

Today, my work is divided into two parts.

Most of my time is spent on system development of high performance storage solutions. This includes contributions to the POSIX filesystem world, but also more innovative techniques combining various approaches such as object storage or machine learning. I started working on the Lustre filesystem to fix a couple of small things we observed in production, and then I increased the scope of my contributions to bigger and more complex patches.

I am also a system administrator on the storage infrastructure of our Tera (defence) and TGCC (academic and industrial research) computing centres. My role is to ensure that the storage system runs smoothly and with the expected performance.

These two activities are mutually self-exclusive, so I cannot work on them both at the same time, but production experience is incredibly valuable when developing. Having a deep understanding of the critical tools (namely the Lustre filesystem and RobinHood policy engine) makes me a much more efficient administrator. This is usually referred to as "DevOps" in the computer literature.

What are the exciting aspects of working in an HPC-related career?

My activities allow me to meet with many people from very different technical areas. First, there are the users of the supercomputers, who include physicists, mechanical engineers and biologists, among others. HPC has become a key tool for many scientists working on major questions of our time. Activities range from climate modelling to personalised medicine or high energy physics. It is important to understand what the users do and how their code operates in order to provide them with sensible advice for using the existing systems, and to design even more efficient ones.

I also travel to meet with various stakeholders in the storage community, including vendors, other HPC sites, and researchers. This really is a job where you learn things every day. It shapes your way of thinking. Being exposed to such a variety of approaches, smart people, innovative ideas and technologies forces you into considering many more options to solve what you are working on. You must simultaneously stay receptive to new techniques and yet know precisely what you want to achieve, in order to filter out what may currently be trendy, but is not appropriate in your case.

What is your outlook on the use of HPC in your field?

As a team and an organisation we put significant effort into making our system more efficient, smarter and more robust. We are working hard on improving interfaces between layers of our storage stack, the mid-term goal being to prepare it for new models. We are currently integrating machine-learning techniques into our centralised logging infrastructure to re-structure information and assist troubleshooting. I can see that we are making progress despite being completely new to this particular field, and this is very exciting!

Exascale is a major goal, and there are still many open questions before we can achieve it. We are investing significant amounts of time in order to design secure and reliable systems that will scale to this extent. I am fully committed to these collective and ambitious targets.

Coming up this year is LAD'16, an international workshop on Lustre, which we are organising in Paris. I expect that this event will be another opportunity to strengthen the Lustre community, share ideas and start projects between teams.

Where do you see your career leading you next?

I have a very strong interest in solving technical problems. I would like to continue accumulating experience and develop my expertise in order to be able to tackle larger and more complex challenges.

Organisational questions behind HPC are also especially interesting. This field has only recently been identified as a strategic one here in Europe. From my position, I have the feeling that parallel storage is not researched as much as its critical role requires. I have also noticed that many groups are not aware of others' work and so miss opportunities for highly fruitful collaborations. HPC plays a pivotal role between research and industry, which I believe can be strengthened. I hope that I can contribute to that at some point.

CEA's Tera 100 supercomputer