represent the new HPC workload, selected from the LNCC’s workload:

The goal of this post-doctoral research is to study and model the performance of applications that use the high-performance computing (HPC) architecture. In 2017, the LNCC's workload included applications that are important to the LNCC because they are representative of the supercomputer's workload. The team, headed by Santos Dumont, was the largest in Latin America and used a diverse scientific community, thus it runs applications from many fields. Therefore, its workload allows for drawing conclusions that can be generalized for many similar applications and systems. The generated knowledge will guide the proposal of monitoring and profiling techniques for applications, and the design of new coordination mechanisms to arbitrate resources in HPC environments.

Trips between Bordeaux and Petrópolis are planned during the contract. Travel expenses are covered by the joint team within limits set by Inria.

Scientific context

HPC architectures, the supercomputers, were conceived to efficiently run traditional HPC applications, namely numerical simulations. However, in the context of the convergence between HPC and Big Data [1], the notion of scientific application is evolving into a scientific workflow composed of CPU-intensive and data-intensive tasks. This evolution characterizes the new HPC workload.

In this new scenario, efficient application execution becomes more challenging due to a mismatch between systems and applications. New applications include new methods, libraries, and runtime systems that may not have been properly optimized to the supercomputer, leading to problems such as load imbalance and poor communication performance. Meanwhile, supercomputers' resources are arbitrated between applications using little information as the number of CPUs and the estimated execution time, which potentially wastes resources that are unused at different moments during application execution [2]. Additionally, although running on independent nodes, concurrent applications still share the network and I/O infrastructures, which means they can interfere with each other. The contention in the access to shared I/O resources has shown to affect applications' performance non-uniformly, depending on their characteristics [3, 4]. Hence, these problems are expected to become worse as the new HPC workload includes more diverse codes and should be tackled by better scheduling at application and system levels, and consider applications' characteristics to avoid issues such as interference [5].

References


Assignment

The goal of this post-doctoral research is to study and model the performance of applications that represent the new HPC workload, selected from the LNCC's workload:

- A numerical simulation library, called MHM, developed by the LNCC [6]. This library implements a number of finite element methods and offers support to hybrid parallelism (OpenMP + MPI) for classic and multiscalar numerical simulations.
- Data analysis tasks and workflows from the BioInfoPortal science gateway [7] developed by the LNCC to allow for easy execution of bioinformatics applications on the Santos Dumont machine.

About the research centre or Inria department

Team STORM combines strengths on high level DSLs, heterogeneous runtimes and performance analysis tools to help programmers get the highest efficiency from modern computer architectures in a portable manner.

Context

This work will be developed within the framework of the HPEProSolv joint team. This team was established in 2021 as a collaboration between Inria Bordeaux (TADAaM and STORM teams) and the National Laboratory for Scientific Computing (LNCC) in Petrópolis, Brazil.

The team's main goal is to study and characterize the new High-Performance Computing workload, represented by a set of scientific applications that are important to the LNCC because they are representative of the supercomputer's workload. Their machine, named Santos Dumont, was the largest in Latin America and used by a diverse scientific community, thus it runs applications from many fields. Therefore, its workload allows for drawing conclusions that can be generalized for many similar applications and systems. The generated knowledge will guide the proposal of monitoring and profiling techniques for applications, and the design of new coordination mechanisms to arbitrate resources in HPC environments.

Trips between Bordeaux and Petrópolis are planned during the contract. Travel expenses are covered by the joint team within limits set by Inria.

About Inria

Inria is the French national research institute dedicated to digital science and technology. It employs 2,600 people. Its 200 agile project teams, generally run jointly with academic partners, include more than 1,500 scientists and engineers working to meet the challenges of digital technology, often at the interface with other disciplines. The Institute also employs numerous talents in over forty different professions. 900 research support staff contribute to the preparation and development of scientific and entrepreneurial projects that have a worldwide impact.

The keys to success

Important qualities to succeed in this work include the capacity for initiative and autonomy, integrity, a willingness to learn, and relational abilities to work in a diverse and geographically-distributed team. A thesis in computer science is required. A thesis in performance profiling or modeling is a real asset.

Instruction to apply

Thank-you to send:
- CV
- Cover letter
- Support letters (mandatory)
- List of publication

Defence Security:

This position is likely to be situated in a restricted area (ZRR), as defined in Decree No. 2011-1425 relating to the protection of national scientific and technical potential (PPST). Authorisation to enter an area is granted by the director of the unit, following a favourable Ministerial decision, as defined in the decree of 3 July 2012 relating to the PPST. An unfavourable Ministerial decision in respect of a position situated in a ZRR would result in the cancellation of the appointment.

Recruitment Policy:

As part of its diversity policy, all Inria positions are accessible to people with disabilities.

Warning: you must enter your e-mail address in order to save your application to Inria. Applications must be submitted online on the Inria website. Processing of applications sent from other channels is not guaranteed.
The recruited person will work in collaboration with researchers from the joint team to profile these applications at different scales, and in concurrence with other codes and stress benchmarks. The recruited person will also be responsible for modeling the applications’ performance, for finding ways to generalize these profiles to similar applications, and for identifying the information that should be obtained during application execution. This information should be useful for obtaining new profiles automatically, and to compute metrics that can help the runtime to predict deviations from the standard application behavior (for instance, if the input phase of an HPC simulation lasts longer than expected, it is possible the application is treating a larger problem and thus will run longer, with longer and more spaced output phases).

References


Main activities

Main activities

- Design and run experiments with applications on a supercomputer
- Model application performance and the effects of interference
- Identify useful information and performance metrics for modeling and predicting application behavior
- Write reports and papers on the subject
- Organize scripts and datasets for the reproduction of results and statistical analyses

Skills

- Knowledge of parallel computing, HPC, and performance profiling and modeling are required.
- Communication skills in English (reading, writing, presenting) are required.
- Knowledge of the French and Portuguese languages are a plus.
- Technical skills: command line usage of Linux-based HPC systems; script programming; ability to modify the source code of applications written in different programming languages; statistical analysis using R or Python.

This is a post-doctoral position for 12 to 24 months, offered in the context of the collaboration between Inria and the LNCC. The candidature must be submitted by email to postdoc-dri@inria.fr before July 10th 2021 with all the documents listed below. If you want to apply, please contact us beforehand at laercio.lima-pilla@labri.fr, francieli.zanon-boito@u-bordeaux.fr, and jean-francois.mehaut@univ-grenoble-alpes.fr.

- The summary sheet;
- A research project detailing the research program, the work plan, the visits expected during the post-doc and when it should begin (November the 1st by default, January the 1st 2022 at the latest);
- A detailed CV of the candidate including a description of the work conducted during the Ph.D., a complete list of the publications, and the 2 most important publications;
- A motivation letter of the candidate;
- Two recommendation letters (from people working in France or outside);
- A support letter from the Inria team;
- A copy of the passport.

Benefits package

- Subsidized meals
- Partial reimbursement of public transport costs
- Possibility of teleworking and flexible organization of working hours
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Social, cultural and sports events and activities
- Access to vocational training
- Social security coverage

Remuneration

2653€ / month (before taxs)