



Offer #2024-07285

PhD Position F/M Scheduling of task-based parallel applications on heterogeneous Cloud computing environments (IDP 2024)

Contract type : Fixed-term contract

Level of qualifications required : Graduate degree or equivalent

Fonction : PhD Position

Level of experience : Recently graduated

About the research centre or Inria department

The Inria center at the University of Bordeaux is one of the nine Inria centers in France and has about twenty research teams.. The Inria centre is a major and recognized player in the field of digital sciences. It is at the heart of a rich R&D and innovation ecosystem: highly innovative SMEs, large industrial groups, competitiveness clusters, research and higher education players, laboratories of excellence, technological research institute...

Context

Cloud computing systems give the general public easy access to computing, communication and storage resources. Its pay-per-use model leads to a reduction in the barriers to setting up a large-scale IT infrastructure [Buyya2019], thereby reducing risks and operational costs, which can be exceptionally important for small and medium-sized businesses and research institutes. In the context of High-Performance Computing (HPC), the use of Cloud resources enables running parallel applications and obtaining results at low cost and waiting time. However, migrating HPC applications to the Cloud is not a trivial task.

The challenges involved in migrating HPC applications to the Cloud come from key differences between Cloud and HPC infrastructures. The heterogeneity of computing resources, well known in HPC with the use of CPUs and GPUs in parallel, is increased in the Cloud due to the constant updating and integration of new resources in the infrastructure [Guidi2021]. Virtualization in the Cloud leads to resource sharing and increased interference, leading to greater performance variation than observed in HPC infrastructures. Finally, Clouds' usage-based payment model (which does not exist in HPC) offers optimization possibilities linked to the cost and performance of different resource options. In particular, spot instances represent resources made available by the Cloud provider at very low cost that can be reclaimed by the Cloud provider at any time [Munhoz2023], giving way to the possibility of optimizing cost and performance under risk of interruption.

Given the current differences between Cloud and HPC platforms, automatic and transparent mechanisms are needed to enable HPC applications to run on the Cloud in a correct, high-performance and low-cost way. What's more, these mechanisms need to be implemented outside the applications themselves, to facilitate the development and migration of new applications.

Assignment

The aim of this thesis is to propose and adapt scheduling and resource management techniques in runtime systems used for programming HPC applications. These techniques should enable an efficient, economical and resource-saving execution of parallel applications in the Cloud.

To this end, this research will use StarPU [Augonnet2011] (<https://starpu.gitlabpages.inria.fr/>), a runtime system developed at LaBRI and the Inria research center at the University of Bordeaux that enables the execution and scheduling of parallel application tasks in heterogeneous environments for applications following the sequential task flow (STF) model, and the HPC@Cloud toolkit [Munhoz2023] (<https://github.com/lapesd/hpcac-toolkit>), a unified framework for migrating, testing and running parallel applications in Cloud environments developed by the international partner laboratory LaPESD from the Federal University of Santa Catarina in Brazil.

On the application side, our main focus will be on the dense linear algebra methods supported by the Chameleon library (<https://solverstack.gitlabpages.inria.fr/chameleon/>) and on the finite element methods of FEniCS (<https://fenicsproject.org/> - in the context of the Maelstrom associate team with SIMULA). Regarding the Cloud environment, we will be focusing on standard commercial Clouds such as Amazon AWS, with the possibility of moving to energy-efficient infrastructures such as the ones provided by Qarnot Computing (<https://qarnot.com/en> - in the context of the Défi Inria PULSE).

Main activities

Main activities:

- Extensive bibliographical research
- Research software development
- Scientific experiments
- Publication of research results in scientific papers
- Presentation and dissemination of the results at conferences and workshops

Additional activities:

- Attendance at the mandatory Doctoral School courses
- Optional participation in Master-level internship student advising
- Optional teaching at the University of Bordeaux or at the ENSEIRB-MATMECA Engineering School

Skills

- Intermediary knowledge of high-performance computing and Cloud computing;
- Good level of software development under UNIX-like operating systems;
- Experience with programming for Cloud computing platforms;
- Good writing skills;
- Willingness to work in a diverse and international environment.

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

- 2100€ / month (before taxes) during the first 2 years,
- 2190€ / month (before taxes) during the third year.

General Information

- **Theme/Domain** : Distributed and High Performance Computing
Scientific computing (BAP E)
- **Town/city** : Talence
- **Inria Center** : [Centre Inria de l'université de Bordeaux](#)
- **Starting date** : 2024-10-01
- **Duration of contract** : 3 years
- **Deadline to apply** : 2024-05-03

Contacts

- **Inria Team** : [STORM](#)
- **PhD Supervisor** :
Lima Pilla Laercio / laercio.lima@inria.fr

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 3,500 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

- Curiosity and appetite for exploration and experimentation;
- Openness to work in a team;
- Motivation to implement ethical and rigorous scientific practices.

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.

Instruction to apply

Thank you to send:

- CV
- Cover letter
- Master marks and ranking
- Support letter(s)

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.