



Offre n°2024-07285

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

Type de contrat : Fixed-term contract

Niveau de diplôme exigé : Graduate degree or equivalent

Fonction : PhD Position

Niveau d'expérience souhaité : Recently graduated

A propos du centre ou de la direction fonctionnelle

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...

Contexte et atouts du poste

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.

Mission confiée

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).

Principales activités

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

Compétences

- 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.

Avantages

- 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

Rémunération

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

Informations générales

- **Thème/Domaine** : Distributed and High Performance Computing
Scientific computing (BAP E)
- **Ville** : Talence
- **Centre Inria** : [Centre Inria de l'université de Bordeaux](#)
- **Date de prise de fonction souhaitée** : 2024-10-01
- **Durée de contrat** : 3 years
- **Date limite pour postuler** : 2024-05-03

Contacts

- **Équipe Inria** : [STORM](#)
- **Directeur de thèse** :
Lima Pilla Laercio / laercio.lima@inria.fr

A propos d'Inria

Inria est l'institut national de recherche dédié aux sciences et technologies du numérique. Il emploie 2600 personnes. Ses 215 équipes-projets agiles, en général communes avec des partenaires académiques, impliquent plus de 3900 scientifiques pour relever les défis du numérique, souvent à l'interface d'autres disciplines. L'institut fait appel à de nombreux talents dans plus d'une quarantaine de métiers différents. 900 personnels d'appui à la recherche et à l'innovation contribuent à faire émerger et grandir des projets scientifiques ou entrepreneuriaux qui impactent le monde. Inria travaille avec de nombreuses entreprises et a accompagné la création de plus de 200 start-up. L'institut s'efforce ainsi de répondre aux enjeux de la transformation numérique de la science, de la société et de l'économie.

L'essentiel pour réussir

- Curiosity and appetite for exploration and experimentation;

- Openness to work in a team;
- Motivation to implement ethical and rigorous scientific practices.

Attention: Les candidatures doivent être déposées en ligne sur le site Inria. Le traitement des candidatures adressées par d'autres canaux n'est pas garanti.

Consignes pour postuler

Thank you to send:

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

Sécurité défense :

Ce poste est susceptible d'être affecté dans une zone à régime restrictif (ZRR), telle que définie dans le décret n°2011-1425 relatif à la protection du potentiel scientifique et technique de la nation (PPST). L'autorisation d'accès à une zone est délivrée par le chef d'établissement, après avis ministériel favorable, tel que défini dans l'arrêté du 03 juillet 2012, relatif à la PPST. Un avis ministériel défavorable pour un poste affecté dans une ZRR aurait pour conséquence l'annulation du recrutement.

Politique de recrutement :

Dans le cadre de sa politique diversité, tous les postes Inria sont accessibles aux personnes en situation de handicap.