Offer #2025-09040

PhD Position F/M Resilient network communications, tolerant to failures and volatility, and suited for artificial intelligence applications.

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

The position is part of a collaboration between the Hivenet company and the TOPAL team. The successful candidate will be part of the TOPAL team based at Inria Bordeaux.

About TOPAL: In TOPAL, we're tackling the evolving challenges at the intersection of high-performance computing (HPC), numerical simulation, and machine learning. As computing platforms grow in scale and complexity—featuring millions of cores and diverse hardware like GPUs—we're leveraging our long-standing expertise in dynamic runtime systems to make the most of these resources. Our work helps adapt to unpredictable workloads and optimize task scheduling without relying on rigid, pre-planned execution. We're now extending this know-how to emerging applications like deep neural network training, which places unique demands on both computation and memory. At the same time, we're addressing the urgent need to reduce energy consumption and carbon footprint in HPC. That means rethinking algorithms, data movement, and hardware use to build more sustainable systems. A key focus for us is efficient data management, since moving and storing data is increasingly more costly than computing itself. By combining our background in linear algebra, resource

scheduling, and algorithmic optimization with these new demands, we're shaping the future of scientific computing on next-generation platforms.

About the Hivenet company: Hivenet is shaping the future of cloud computing by leveraging unused computing capacity to provide a decentralized, environmentally friendly, and user-empowered alternative to traditional cloud services.

Assignment

Context

Hivenet is a company that offers individuals and businesses the ability to make their unused computing resources available. Hivenet offers a data storage service called **HiveDisk**, which utilizes the storage space contributed by HiveDisk users. This enables HiveDisk users to benefit from geo-distributed and replicated storage. Similarly, Hivenet aims to share, via **HiveCompute**, unused computing resources (primarily GPUs) to perform tasks related mainly to the training and inference of artificial intelligence applications. Through a web interface, users can request the allocation of a certain number of GPUs distributed across different machines and then access them to run their computations.
Initially, the allocated GPUs will be located on machines within the same local network (for example, within a company's site network or a PoliCloud container), but the long-term goal is to leverage GPUs located across different enterprise or community networks in various locations (e.g., all the sites of a company across a country) [1].

This project presents numerous challenges, mainly because the targeted environment differs from traditional HPC environments. From a hardware perspective, the machines are less powerful, heterogeneous, and interconnected via standard networks that are less efficient and reliable than HPC networks. It is also important to consider that computing resources are not available at all times (for example, machines may be less available during the day when employees are using them) and are more likely to become unavailable unexpectedly. Furthermore, using machines spread across different geographical locations results in a network with heterogeneous performance: latency between two sites is much higher than within a single site.

Objectives

The objective of this PhD thesis is to explore issues related to network communications in such a context. This will require an analysis of existing communication libraries (such as PCCL [3], MPI [2], or libp2p) in order to determine which one(s) can be adapted to the targeted environment. Once a suitable communication model is established, the work will focus—given a set of machines and their topology—on adapting communication patterns in learning applications to minimize communication overhead. This may involve using routing algorithms and better distribution of computation and data, tailored to the characteristics of the network interconnecting the machines.
The system will also need to detect the addition or loss of machines and respond accordingly—for example, by ignoring the contributions of lost machines in a

data-parallel setup, or by redistributing the data and computation.

In a second phase, the project will consider how to manage network usage when HiveDisk and HiveCompute are active simultaneously on the same networks and machines. The goal will be to maintain acceptable performance levels for both services by dynamically adjusting quality-of-service parameters based on network conditions and user requirements.

References

[1] N. T. Karonis, B. de Supinski, I. Foster, W. Gropp and E. Lusk, "A Multilevel Approach to Topology-Aware Collective Operations in Computational Grids." arXiv preprint cs/0206038, 2002

[2] L. Shalev, H. Ayoub, N. Bshara, and E. Sabbag, "A Cloud-Optimized Transport Protocol for Elastic and Scalable HPC", *IEEE micro*, *40*(6), 67-73.

[3] M. Keiblinger, M. Sieg, J. Min Ong, S. Jaghouar and J. Hagemann, "Prime Collective Communications Library -- Technical Report", *arXiv preprint arXiv:2505.14065*, 2025

Main activities

The PhD student will conduct original researches on the topic described above, and will collaborate with colleagues in the TOPAL team and Hivenet partners.

Activities includes, but are not limited to: bibliographical synthesis, research, software implementation, presentation of results at conferences, attending research schools, etc

Skills

Technical skills and level required :

- Solid understanding of network communication (socket, TCP/IP);
- Proficiency in systems programming (C/C++) and high level language (Python).
- Experience with AI frameworks (PyTorch, Tensorflow, ...) and software performance measurement is a plus
- Interest in distributed systems and HPC are also a plus

Languages :

Good communication skills in English (French is a plus)

Relational skills :

Ability to work collaboratively in an academic–industry setting

Benefits package

- Subsidized meals
- Partial reimbursement of public transport costs
- Leave: 7 weeks of annual leave + 10 extra days off due to RTT (statutory reduction in working hours) + possibility of exceptional leave (sick children, moving home, etc.)
- 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

The monthly salary will be 2200€, and 2300€ in 2026 (before social security contributions and witholding tax)

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 :** 2025-09-01
- **Duration of contract :** 3 years
- **Deadline to apply :** 2025-07-31

Contacts

- **Inria Team :** TOPAL
- **PhD Supervisor :**
  Herault Thomas / thomas.herault@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

Feeling comfortable in a dynamic scientific environment, enjoying learning and tackling real-world problems, and having an experimental mindset are essential qualities for succeeding in this role.

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

If you are interested, thanks to candidate on jobs.inria with the following documents :

- cv
- cover letter

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.