Offer #2023-05814

Engineer F/M: Strategies and tools for placing HPC tasks and data on heterogeneous memory architectures

Contract type : Fixed-term contract

Renewable contract : Yes

Level of qualifications required : Graduate degree or equivalent

Fonction : Temporary scientific engineer

About the research centre or Inria department

The Inria Bordeaux Sud-Ouest centre is one of Inria's nine centres and has around 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 memory hierarchy of HPC platforms keeps becoming even more complex. Now that all processors have multiple levels of caches, memory itself is changing with the advent of different technologies with different performance characteristics. Heterogeneous memories first appeared in HPC several years ago with the Intel Xeon Phi processor. Custom strategies were implemented to place bandwidth-sensitive data buffers on the appropriate memory. The emergence on non-volatile memory DIMMs (that may be used either as fast storage or as slow but high-capacity memory) brought similar questions. Moreover several ARM and x86 platforms are going to combine HBM and DRAM on the road to exascale, and the CXL technology is bringing new ways to connect additional memory targets to modern processors. There is a critical need for software stacks to learn portable ways to identify different kinds of memories, to abstract their characteristics in terms of performance, and to decide where to allocate which data buffer.

Heuristics for Heterogeneous Memory (H2M) is an ANR-DFG joint-project between the Inria TADaaM team in Bordeaux (France) and RWTH Aachen (Germany) to address these questions. The goal is to develop a hierarchy of programming abstractions to expose heterogeneous memory at different levels of detail and control, complemented by a set of required, vendor-neutral capabilities to be provided by standards and intelligent runtime systems.

This engineer position is funded by the H2M ANR-DFG project (https://h2m.gitlabpages.inria.fr/).

Assignment

The role of the engineer will be:
* Survey of new and upcoming heterogeneous memory hardware (HBM, NVM, CXL, etc)
* Deployment of software techniques to simulate heterogeneous memory (virtual machines, hardware features to limit bandwidth, etc)
* Analysis of the performance characteristics of heterogeneous platforms
* Development of the H2M software stack, e.g. memory allocation strategies
* Development of the hwloc library, e.g. exposing more information about heterogeneous memory
* Study of methods and tools to identify the sensitivity of data buffers and compute kernels to memory bandwidth or latency
* Evaluate the contributions with different applications on different hardware architectures

Main activities

The engineer will develop these ideas in the software projects developed by the H2M project and the Inria TADaaM team. This notably includes the hwloc library for low-level hardware management and operating system interaction. Higher level ideas will be development as part of the H2M project in collaboration with the RWTH Aachen team, with the goal of proposing ideas to the OpenMP standard committee. Modifications of the LLVM compile and runtimes might be used as a way to identify sensitivity and apply placement heuristics.

The candidate will be in charge of implementing the ideas in collaboration with other members, as well as evaluating them with different applications on either real or simulated platforms.

Skills

Technical skills and level required:

- Basic knowledge of parallel architectures (multicore, NUMA, etc), operating systems and compilation techniques.
- C/C++ programming as well as some development tools (CMake, autotools, etc)
- Unix environment (shell, ssh, etc)

Languages:

- English mandatory, French appreciated.

Remuneration

The gross monthly salary will be between € 2652 and € 3039, depending on your qualifications and professional experience (before social security contributions and monthly withholding 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 :** 2023-05-01
- **Duration of contract :** 1 year, 6 months
- **Deadline to apply :** 2023-04-30

Contacts

- **Inria Team :** TADAAM
- **Recruiter :**
  Goglin Brice / Brice.Goglin@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.

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 by this job, Please could you apply on website jobs.inria with the following documents :

CV
cover letter
recommendation letters

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.