HPC/HPDA research engineer

Overview

Maison de la Simulation (https://mdls.fr/) offers 2-3 engineer positions to join the newly created PDI team.

- Contact: Yushan Wang (yushan.wang@cea.fr) & Benoît Martin (bmartin@cea.fr)
- Duration: 3 years
- Start date: ASAP

Application

Candidates must have at least a Master's degree or equivalent in Computer Science, Applied Mathematics, or other relevant fields. A PhD degree and working experience in a relevant domain are appreciated. Good programming skills are required.

Applications should be sent to Yushan Wang and Benoît Martin. They should include:

- a curriculum vitae
- a motivation letter
- at least two referees with their email addresses
- links to Masters/PhD thesis and publications
- links to software contributions

Context

With the increasing complexity of numerical simulation codes, new approaches are required to analyze the ever-growing amount of data. This requires coupling up-to-date data analysis libraries with the existing highly optimized numerical simulation codes. The PDI Data Interface code coupling library is designed to fulfill this goal.

The open-source PDI Data Interface library (https://pdi.dev/) is designed and developed for process-local loose coupling in high-performance simulation codes. PDI supports the modularization of codes by inter-mediating data exchange between the main simulation code and independent modules (plugins) based on various libraries. It is developed in modern C++ and offers C, Fortran, and Python application programming interfaces.

PDI offers a reference system similar to Python or C++'s shared_ptr with locking to ensure coherent access by coupled modules. It provides a global namespace (the data store) to share references and implements the Observer pattern to enable modules to react to data availability and modifications. It implements a metadata system that can specify a dynamic type for references based on the value of other data (e.g., array size based on the value of a shared integer). Codes using PDI's declarative API expose the buffers in which they store data and trigger notifications when significant steps in the simulation are reached. Third-party libraries such as HDF5, SIONlib, or FTI are wrapped in a PDI plugin. A YAML configuration file is used to interleave plugins and additional code without modifying the original application.

Another aspect we explore with PDI is in-situ data analysis, which performs numerical analytics during the simulation. This is necessary due to the ever-growing gap between file system bandwidth and compute capacities. To this end, we are developing the Deisa (https://github.com/pdidev/deisa) plugin. This plugin is based on the open-source Dask framework and allows us to transfer data to dedicated processes to perform in-situ analysis.

One of our goals is to establish a feedback mechanism between the in-situ data analysis and the numerical simulation. This allows better resource allocations and on-the-fly simulation monitoring. Another aspect that in-situ analysis enables is using AI methods for HPC and HPDA. For instance, we can have unsupervised detection of rare events during the simulation, which can greatly reduce the amount of produced data, thus reducing stress on the file system.

Work environment

At Maison de la Simulation laboratory, you will join a group of engineers and scientists focusing on all aspects of high-performance computing (HPC). You will have the opportunity to collaborate with PDI users and to introduce new features in the PDI plugin family. As a member of the PDI team, you will also have the opportunity to exchange with the developers of other HPC codes to enrich your skills in HPC code development. To validate your developments, you will be provided with access to the top European supercomputers (Adastra, Jean-Zay, etc.).

Work content

As a member of the newly created PDI team, your primary focus will be developing and maintaining the PDI library.

- Develop core functionalities and new plugins for PDI
- Develop the Deisa library
- User-support
- Organize training sessions
- Library packaging and deployment

Skills

The successful candidate will master the following skills and knowledge:

- Proficiency in modern C++ (C++14 and above)
- Software engineering and library design
- Modern development environment (Linux, git, CMake, etc.)
- Communication (writing, presenting, and training)
- Team-work and integration in an international environment

In addition, the following will be considered a plus:

- Data analysis libraries such as Dask
- Knowledge and experience with Python, Fortran and/or GPU computing
- HPC and parallel libraries such as OpenMP and MPI
- HPC parallel IO libraries such as HDF5 or NetCDF
- Experience with supercomputers tools (slurm, sbatch, etc.), packaging and deployment

Salary and benefits

The CEA offers salaries based on your degrees and experience. This position provides several advantages:

- The possibility of joining collaborations with other European laboratories, the United States, and Japan
- Numerous opportunities to travel internationally (exchanges, conferences, workshops and more)
- 5 weeks of paid vacation and 4 weeks of RTT per year, and up to 2 days of remote work per week.
- Reimbursement of up to 75% of public transport cards and a free transport network throughout the Ile-de-France region
- Complimentary health insurance and several company savings plans