Preparing, developing, running and analyzing large scale Melissa enabled applications. The engineer will support a larger variety of ensemble run based computations. Working in close direction with the tram scientific instruments for instance. This includes computing surrogate models (reduce models) using deep simulation. Our team is today investigating other ensemble runs based computations where Melissa takes over the running simulation that are processed in parallel on-line as soon as available, thus they dynamically connect to a parallel data processing parallel server. This server gets the data running in the background. Once these simulations are submitted to the machine batch scheduler independently. Once these simulations start the contract sooner if we find a good candidate or even later to accommodate some specific situations (titulation for instance).

Hiring date is flexible. We expect to hire the candidate in June 2021, but we have the possibility to start the contract sooner if we find a good candidate or even later to accommodate some specific situations (titulation for instance).

The city of Grenoble is surrounded by the Alps mountains, offering a high quality of life and where you can experience all kinds of mountain related outdoors activities and more.

Principales activités

Numerical simulations are today commonly used for modeling complex phenomena or systems in different fields such as physics, chemistry, biology or industrial engineering. Some of these numerical simulations require supercomputers to run high-resolution models. In general, a numerical simulation needs a set of input parameters in order to produce the simulation outputs. The input parameters and the often complex internal model produce outputs that can be very large. Analyzing these outputs, tuning the input parameters, or adjusting the internal simulation state based on observation data provided by scientific instruments is still very challenging. Ensemble runs consist in running many times the same simulation with different parameter sets to get a steady analysis of the simulation behavior in the parameter space. The simulation can be a high resolution large scale parallel code or a smaller scale lightweight version called meta-model or surrogate-model. This is a process commonly used for Uncertainty Quantifications, Sensibility Analysis, Data Assimilation, Deep Learning based Surrogate training. This may require to run from thousands to millions of the same simulation, making it an extremely compute-intensive process that will fully benefit of Exascale machines.

Existing approaches to handle large ensemble runs show a limited scalability. They either rely on intermediate files where each simulation run writes outputs to file that are next processed for result analysis. This makes for a flexible process, but writing these data to the file system and reading them back for the analysis step is a strong performance bottleneck at scale. The alternative approach consists in aggregating all simulation runs into the same large monolithic MPI job. Results are processed as soon as available avoiding intermediate files. However, by not taking benefit of the loose synchronization needs between the different runs, this over-constrains the execution regarding compute resource allocation, application flexibility or fault tolerance. To overcome these limitations we recently proposed a new framework, Melissa, relying on an elastic architecture. Simulations are submitted to the machine batch scheduler independently. Once these jobs start they dynamically connect to a parallel data processing parallel server. This server gets the data from the running simulation that are processed in parallel on-line as soon as available, thus avoiding intermediate files. The computed partial results can be retro-fed to the simulation (needed for data assimilation for instance). They can also be used to support an adaptive sampling process where the set of parameters for the next simulation runs are defined according to these partial results. Such framework enables to fully take benefit of the loose synchronization capabilities between simulation runs: simulations are submitted and allocated independently enabling to better use the machine resources and to support efficient fault tolerance mechanisms.

Melissa framework has been developed for sensitivity analysis where the server gets the data from the simulation to compute statistics without influencing directly the execution of the simulation. Our team is today investigating other ensemble runs based computations where Melissa can bring an original solution. This includes computing surrogate models (reduce models) using deep learning, data assimilation (steering the simulation execution with data obtained from other scientific instruments) for instance.

The engineer will work on the Melissa development and side activities, mainly:

1. Melissa framework consolidation and extension. One objective is to make it more generic to support a larger variety of ensemble run based computations. Working in close direction with the tram master and PhD Students as well as external collaborators.
2. Preparing, developing, running and analyzing large scale Melissa enabled applications. The engineer
will have access to a panel of world-class supercomputers like Jean-Zay at Idris or Juwels at Julich, as well as smaller ones for prototyping.

Through this work the engineer will consolidate her/his expertise in high performance computing, high performance data analysis and gain knowledge on deep learning, reinforcement learning, data assimilation. She/he will integrate a dynamics research team, be involved in research activities including preparation of scientific publications. She/he will have the opportunity to interact with other research teams as well as industrial partners involved with Melissa (EDF R&D for instance), and as such may have to make some short travels in France and abroad.

References:
- Melissa for sensibility analysis: https://hal.inria.fr/hal-01607479v1 and https://melissa-sa.github.io/
- Melissa for data assimilation: https://hal.archives-ouvertes.fr/LIG/hal-03017033v2

Compétences
We welcome candidates with a master in computer science, experience in high performance computing and scientific computing, and ideally knowledge related to sensitivity analysis, data assimilation or deep learning. No previous work experience required as long as you are motivated and ready to train yourself to complement your skills. Income will be adjusted to your experience.

Expected technical skills include Linux expertise, C/C++ and Python programming, parallel programming experience, a good mastering of development processes including git, continuous integration, containers, etc.

Work will be split in between software development, experiments on large scale supercomputers, explorative research.

Candidates should be motivated by research work, and with good abilities for team work, oral and written communication skills.

A good level of English is thus required. French is not mandatory and INRIA will provide French classes if needed.

To apply submit your CV, references, recent marks, and if available your last Internship/Master Thesis manuscript. With your application provide any element (github account, etc.) that could help us assess your skills beyond your academic record, as well as a few references to persons we can contact to get some feedback on your qualities.

Avantages
- Subsidized meals
- Partial reimbursement of public transport costs
- Leave: 45 days/year
- Social, cultural and sports events and activities (CAESUS)
- Access to vocational training
- Social security coverage

As UGA employee

Rémunération
Income will depend on the candidate's work experience and titles.

For instance a candidate with 3 years of experience and an Engineering title or PhD can reach 2800 euros/month gross income.