2018-00467 - Monads and optimized compilation into efficient big data programs - Post-Doctorant Inria Research center

Contract type: Public service fixed-term contract
Level of qualifications required: PhD or equivalent
Fonction: Post-Doctoral Research Visit

About the research centre or Inria department

Grenoble Rhône-Alpes Research Center groups together a few less than 800 people in 35 research teams and 9 research support departments.

Staff is localized on 5 campuses in Grenoble and Lyon, in close collaboration with labs, research and higher education institutions in Grenoble and Lyon, but also with the economic players in these areas.

Present in the fields of software, high-performance computing, Internet of things, image and data, but also simulation in oceanography and biology, it participates at the best level of international scientific achievements and collaborations in both Europe and the rest of the world.

Context

The work will be carried out in the Tyrex research team common to Inria and CNRS LIG (http://tyrex.inria.fr). Tyrex is dedicated to study data-centric programming, with a long experience of conducting research both at the theoretical and implementation levels. This work is part of an ambitious program towards the synthesis of optimized big data programs.

Assignment

Building AI applications with amounts of data that exceed single computer capabilities remains a very time-consuming and expensive task.

As pointed out in a recent Stanford report [BORZ17], "This expense comes not from a need for new and improved statistical models but instead from a lack of systems and tools for supporting end-to-end machine learning application development, from data preparation and labeling to productionization and monitoring". In the CLEAR project [cle20] of the Tyrex team at LIG/Inria Grenoble, we seek to provide high-level programming abstractions and techniques to facilitate the construction of AI/big data applications.

Current big data toolboxes (such as Apache Spark) provide low-level primitives for distributed data-centric computations. These toolboxes are progressively extended with higher-level programming abstractions such as dataframes, ML pipelines [MBY+15], and queries [AXL+15, GCL+17]. The next challenge is to make such programming abstractions and techniques more efficient and more accessible in each stage of the construction of a big data application. This is what we explore in the CLEAR project [cle20], where we investigate the synthesis of code optimized for big data toolboxes. In particular, we explore how code can be automatically generated and optimized from higher-level descriptions. The overall goal is to provide high-level programming abstractions and techniques to facilitate the construction of more robust and more efficient AI/big data applications, such as the ones found in linked data [GJGL16], finance, retail, healthcare [GCL+17], etc.

Additional information/References:


[GCL+17] Pierre Genevès, Thomas Calmant, Nabil Layaida, Marion Lepelley, Svetlana Artemova, and
Main activities

Skills
We seek brilliant and motivated candidates with an equal passion for theory and practice. Knowledge and interest in programming, compilers, theory (e.g. monads) and distributed frameworks (e.g. Apache Spark) will be appreciated.

Benefits package
- Subsidised catering service
- Partially-reimbursed public transport
- Social security
- Paid leave
- Flexible working hours
- Sports facilities

Remuneration
Gross salary: 2650 Euros per month