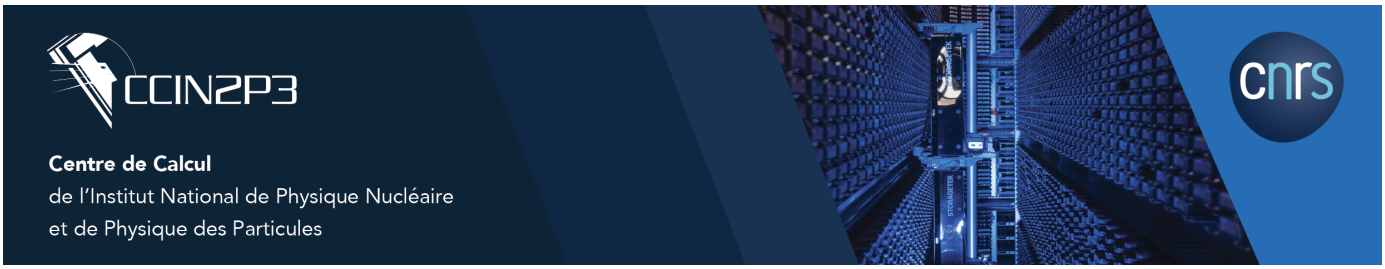

Workflow vs. Dataflow: Concepts, challenges, and simulation for High Performance Computing Internship/Ph.D. proposal

This internship requires solid skills in C/C++ programming, Linux, and concurrent programming. The main objective is to combine the respective strenghts of two renowned simulation toolkits: PREESM [1] et SimGrid [2]. The former is *dataflow-oriented* [3] and able to accurately simulate the performance of a single heterogeneous node that combines CPUs, GPUs, and FPGAs. The latter is *workflow-oriented* [4] and can accurately simulate the communications between heterogeneous nodes.

Workflows and dataflows are conceptually close. Both model applications that exhibit task and data parallelism, but they also show some fundamental differences. A first milestone will be to identify the commn concepts and main differences between these two application models. Then, the main part of the internship will be to adapt, extend, and/or modify the workflow-oriented programming interface of the SimGrid toolkit to enable the simulation of dataflow-oriented applications.

Depending on the capacity of the candidate to achieve these two milestones within the duration of the internship, the development of a full-fledge simulator and the performance evaluation of scheduling algorithms will be considered.

This internship will indeed be done in the context of the funded ANR DARK-ERA (2021-2024) project and can lead to a Ph.D program. This project is connected to the international *Square Kilometer Array* (SKA) radio-telescope that will generate data at an unprecedented rate (several Terabits per second). A chain of complex algorihtms will be executed on this data in real time, without any available storage capacity, and with an enercy budget of less than a MWatt for 250 Petaflops.

Such constraints require to design an innovative and highly heterogeneous architecture. To help at this design, the Dark-Era project aims at developping a rapid prototyping tools of dataflow applications to quickly assess memory usage, latency, throughput, or energy consumption and explore a large design space.

To apply to this internship, please send your resume and cover letter to martin.quinson@ens-rennes.fr and frederic.suter@cc.in2p3.fr.

References

[1] M. Pelcat, K. Desnos, J. Heulot, C. Guy, J.-F. Nezan, and S. Aridhi. PREESM: A Dataflow- based Rapid Prototyping Framework for Simplifying Multicore DSP Programming. In Proceedings of the 6th European Embedded Design in Education and Research Conference (EDERC), pages 36–40, 2014.

[2] H. Casanova, A. Giersch, A. Legrand, M. Quinson, and F. Suter. Versatile, scalable, and accurate simulation of distributed applications and platforms. Journal of Parallel and Distributed Computing, 74(10):2899–2917, 2014.

[3] E. A. Lee and D. G. Messerschmitt. Synchronous dataflow. In Proceedings of the IEEE 75(9):1235–1245, 1987)

[4] E. Deelman, T. Peterka, I. Altintas, et al. The Future of Scientific Workflows. The International Journal of High Performance Computing Applications. 32(1):159-175, 2018.

21 avenue Pierre de Coubertin
CS 70202 - F-69627 Villeurbanne CEDEX

Tél. : +33 (0)4 78 93 08 80
Fax : +33 (0)4 72 69 41 70

Web : https://cc.in2p3.fr
Twitter : @cc.in2p3