

Internship offer for Master Student – Computer Science

Laboratory / University: IRIT / Université Paul Sabatier

Research teams: PYRAMID & SEPIA

Title : Smart Data Replication

Motivation:

Serverless environments allow the execution of functions on demand. Most of the literature focuses on optimizing its scalability, but little attention is paid to the data replication problem. Data is distributed in chunks located in different nodes and replicated for fault tolerance. The student will explore the utilization of AI to design a scheduler that optimizes the number and location of data (chunk) replicas as a trade-off between task completion time and the data center energy consumption. The next section addresses two possible directions for the application of ML/DL.

Project Outline:

The project builds on top of a prior publication [1]. To limit the scope, the student will focus on the following tasks:

- Propose a multi-objective scheduler that determines the best number of replicas per data chunk, and their location. The objective is a trade-off between execution time and energy consumption.
- This can be formulated as a traditional ILP problem and then explore the application of Reinforcement Learning algorithms (e.g. DQN, PPO, Bandits).
- The work can be done iteratively: focus first on the number of replicas, then on their location.
- Develop or adapt forecasting capabilities of the future resource consumption and/or incoming number of requests.

- This can be formulated as a regression or time series prediction problem. Techniques range from traditional (SVN, linear regression) to the use of DNN such as LSTMs.
- Assess the quality of the proposed multi-objective scheduler using both a perfect prediction (using an Oracle) and the forecasted resource consumption (with errors). The quality will be studied in terms of energy consumption and execution time. Based on the student background, simulation (using Simgrid), real deployment (on Grid5000) or both will be conducted.

Scientific Extension:

As a first interest extension, the prior formulation can be extended to consider an unreliable environment where multiple nodes (FaaS) fail at different times. The work can be extended in a second step to take economic issues into consideration for the choice of the number of replicas, and their locations (different locations may have different prices to host data and execute FaaS).

After the Project:

The Master Thesis could continue after the internship upwards a PhD, depending on the candidate and fundings opportunity.

Location: Toulouse, France

Supervision:

Jean-Marc Pierson, University of Toulouse, Institute for Research in Informatics at Toulouse (IRIT), Jean-Marc.Pierson@irit.fr

Tania Lorida Botran, Roblox Inc., tbotran@roblox.com

Riad Mokadem, University of Toulouse, Institute for Research in Informatics at Toulouse (IRIT), Riad.Mokadem@irit.fr