Master Thesis Proposal (30 ECTS credits):

A live programming environment for Big Data applications

Lab: Software Languages Lab (SOFT), VUB.

Advisors: Elisa Gonzalez Boix, Matteo Marra

Period: Academic year 2019/2020.

Domain interests of the candidate: debuggers, programming languages, parallel programming, Map/Reduce

Context:

Hardware advances in storage capacity and CPU processing have given rise to the concept of Big Data, characterized by the so-called 3 Vs (Volume, Velocity and Variety). Novel software platforms have emerged to analyze and store large data sets in a scalable way. The two most prominent programming models are Hadoop Map/Reduce and Apache Spark which typically embrace a batch-oriented data processing to achieve a high parallelisation of data analysis.

Current trends indicate that the volume, velocity and variety of data are increasing quickly due to an explosion on diversity and number of sources of information (as a result of the digitalization of data, e.g. smart objects and sensors, interconnectivity of data and popularity of social media data. This poses challenges for Big Data frameworks to be able to meet the new requirements of the emerging real-time streaming data processing applications. The 2017 Hadoop perspective annual report by Syncsort, a leading company in (Big) data integration, estimates the need of new tools to simplify the interaction of the programmers with different evolving frameworks and datasets. In particular, very few debugging support is currently provided in those frameworks. When an error happens, developers are lost in trying to understand what happened from the information provided in log files. Alternatively, few solutions allow to replay the execution, but they are slow and time-consuming.

In previous work at the Software Language Lab, Port is a Big Data framework that allows Pharo programmers to write highly parallel applications and to execute them on state-of-the-art clusters using Hadoop Yarn. Through a middleware, called Pharo On Yarn, different Pharo instances can be deployed on Yarn, and Port can then model the execution using a Map/Reduce like model. Port already runs different applications on Isabelle, a ten nodes cluster of the Software Languages Lab with a total of 80 cores and 320 GB of RAM in the worker nodes.

Proposal:

The goal of this thesis proposal is to explore IDE features which enables a live coding and debugging of Big Data applications written in Port deployed on clusters like Isabelle. We envision a number of features that a student could explore in the master thesis depending of his/her background and interests including a monitoring system to track the status of the cluster and of a running application in Port, an automized distributed application testing which allows to translate benchmarking suites like HiBench in Port, online distributed debugging, and managing of streaming applications.

Studying at Brussels and SOFT:

Brussels is the official capital of Belgium and unofficial capital of Europe. From the city in which Charles V was declared king of the unified Spain, to hosting the headquarters of the European Union today, Brussels has always played an important political role and is defined by a vibrant mix of old and new. Brussels is also a cosmopolitan city like no other. 31% of its population is of foreign origin, and this makes for a unique atmosphere in which cultures interact easily with one another. As such, VUB receives every year international students in the context of different programs such as Erasmus, Socrates or International Masters.

SOFT is active in the design, formalisation, implementation and application of new programming languages and new software engineering technologies. This includes inventing, improving and applying state-of-the-art programming technologies, software components, and development tools to support the software construction process of the future. SOFT aims to cover all aspects of this domain varying from (formal) foundational research up until industrial applications. SOFT currently has a headcount of 5 full-time professors (V. Jonckers, W. De Meuter, E. Gonzalez Boix, C. De Roover, D. Devriese), 3 part-time professors, 6 post-docs and 25 pre-docs. The lingua franca of the lab is English as many of its current members come from different countries such as Kenya, Cuba, Italy, and USA. The SOFT lab organizes regular group meetings next to the weekly progress meetings with the supervisor(s) to follow up on the work conducted by master students.

More information:

- Read more about SOFT http://soft.vub.ac.be/
- Read more about VUB http://www.vub.ac.be/en/

- Questions on the proposal? Mail Elisa Gonzalez Boix egonzale@vub.be (in Catalan/Spanish/English)