Éligible à une allocation de type : Titre du sujet :

UCA-EDSTIC-EUR-DS4H

Autonomic Adaptation in the context of Distributed Streams Processing Systems

Mention de thèse :

Informatique

HDR Directeur de thèse inscrit à l'ED STIC :

Françoise Baude

Co-encadrant de thèse éventuel :

Nom :

HUET

Prénom :

Fabrice

Email :

fabrice.huet@unice.fr

Téléphone :


Email de contact pour ce sujet :

fabrice.huet@unice.fr

Laboratoire d'accueil :

I3S

Description du sujet :

Today numerous Distributed Stream Processing Systems (DSPS) are available. They differ by the way: 1) the user expresses the treatment to the data streams, 2) the format to which it is translated, 3) how the underlying supporting platform handles the resulting format and may introduce some sort of automatic reconfiguration, for instance to handle elasticity in the computing resources.

In this proposed PhD research, we are focusing on the programming capabilities for both functional and non-functional properties of DSPS systems. In particular, we will focus on the autonomic reconfiguration of these systems at run-time. We will consider only DSPS models where the user defines the system as a graph of operators onto which data flow, in a similar vein as Infosphere/System S [1], Storm, PiCo resulting skeletons, Vienna VTDL/VISP [2] ... Indeed, in the team, we have developed a programming library targeted at expressing a DSPS program as a hierarchical composition of distributed software components. The library is entitled “GCM-streaming” [3] https://github.com/moliva/gcm-streaming, relying onto the GCM / ProActive OW2 Java-based platform [4], which thus offers reconfiguration capabilities.
So far, these reconfiguration capabilities have not been very well explored in the case of DSPS applications. This is the goal of this PhD research to push forward in this direction, by exploring 1) not only non-functional needs (like performance, that conducts to elasticity, like this is done in many systems proposed eg Twitter Heron, Microsoft Dhalion [5], ...) 2) but also more innovative, to explore functional needs: depending of what is “found” by the computation, it may be relevant to adapt the functional operators without completely stopping the system, in order for example to focus more on a specific situation detection and understanding. E.g, if tweets sentiment analysis shows a very bad sentiment about a specific topic, it may be relevant to extract which are the words that appear the most in these tweets, in order to understand why the negative sentiment. Our goal is to study how to express and enact management rules in GCM-streaming and similar middle-sized grain programmable DSPS to not only dictate reconfiguration of the graph of operators pertaining to performance improvement, fault-tolerance, etc., but also to functional needs. Indeed, more and more frequently, DSPS applications are long-running deployed applications whose adaptation of the graph of operators is needed, but being critical, dynamic adaptation is needed in order to minimize the stopping duration and prevent to have to stop/destroy/redeploy a complete newer version of the application.

Dimensions we aim to study and have some contributions:
- expressiveness of the GCM-streaming model based on the underlying ProActive multi-activeobject with future model: GCM component model can be extended with the notion of Future on Streams [6], to better control the way stream data flow onto the graph of operators, aiming better balancing the data load of each operator on the way
- capability to express and enact the logic of adaptation, which pertain to various aspects: programming and plugging a) of functional and non-functional oriented sensors, b) of reactive or proactive rules through external high-level adequate programming languages, e.g. GCM-script extended with Machine Learning based strategic decisions, c) dynamic enactment of adaptation of the user level graph, in ways that suits with the underlying representation of the stream application

We target to produce prototypes to support these ideas by extending both GCM-streaming and/or the Apache STORM open source framework the team is also currently working adding some optimizations in the Storm middleware. Use cases to illustrate the need of functional- level triggered dynamic adaptation will be defined by building realistic scenario requiring for example Natural Language Processing, within the context of Smart cities (eg disaster management [7]). Besides functional-level triggered reconfiguration, we will also target performance optimization (scaling up and down) and fault-handling scenario.


English version: