2020-02397 - PhD Position F/M [AR] Hypersteaming Program Evaluation for Monitoring Dynamic Data

Type de contrat : CDD  
Niveau de diplôme exigé : Bac + 5 ou équivalent  
Fonction : Doctorant

A propos du centre ou de la direction fonctionnelle

The Inria Lille – Nord Europe Research Centre was founded in 2008 and employs a staff of 360, including 300 scientists working in sixteen research teams. Recognised for its outstanding contribution to the socio-economic development of the Hauts-de-France region, the Inria Lille – Nord Europe Research Centre undertakes research in the field of computer science in collaboration with a range of academic, institutional and industrial partners.

The strategy of the Centre is to develop an internationally renowned centre of excellence with a significant impact on the City of Lille and its surrounding area. It works to achieve this by pursuing a range of ambitious research projects in such fields of computer science as the intelligence of data and adaptive software systems. Building on the synergies between research and industry, Inria is a major contributor to skills and technology transfer in the field of computer science.

Contexte et atouts du poste

Complex event processing [1,2,3] is fundamental for dynamic data management [4]. The basic problem is to monitor data streams in order to react on complex events with low latency. Complex events are usually defined by logical queries on the logical structure that is communicated over the stream. Monitoring processes can then be decomposed into two tasks: the first is the detection of complex events, and the second the creation of a reaction to each of the complex events. The first task is known as query answering on streams and has been largely studied in the last 15 years [5-14]. The study of the second task is still at the beginning. In the simplest case, it amounts to send the complex event to the output stream or to some database. In more complicated cases, it requires to change the state of the monitoring processes itself. In such cases, both steps are mutually depend, so that they cannot be decomposed sequentially any more. The recursive dependency of such adaptive monitoring processes can be described by query-based programs, rather than by single queries.

Mission confiée

1) The research topic and its scientific and economic context

We propose to develop general algorithms and tools that can evaluate query-based programs in streaming mode. Only few previous work on streaming evaluation such programs exists so far [16,18]. In order to remedy this situation, we propose in this PhD project to lift recent streaming tools and algorithms for the evaluation queries to the evaluation of query-based programs. We propose to start from path queries on streams that contain data trees and afterwards continue with the monitoring of streams that beside of data trees may contain data graphs as used in knowledge representation in the semantic Web, (i.e., in the RDF format (Resource Description Framework)).

2) Subject's condition in the receiving lab

This will be the fourth PhD thesis on stream processing in sequence, under the direction of Joachim Niehren at Inria Lille: O. Gauwin (2008-10,[14]), T. Sebastian (2010-15, [13]), Momar Sakho (2016-19, [18]). The results obtained there will be most relevant for the thesis proposed here. On tool site, we plan to reuse the modules produced by M. Sakho, and those implemented by T. Sebastian in the context of the industrial transfer project QuickPoc (2010-15) on XML stream precisions with the company Innovimax. An important point here is to break software development limitations that are due to shared-ownership with Innovimax, so that novel transfer activities on the domain of streaming support for graph databases become envisagable.

3) Objectives, expected results:

The first question is which kind of programming language to use to enable monitoring processes based on query answering. We propose to use higherorder purely functional programs -- without imperative updates -- that permit to apply queries defined in some external query language. The starting points will be the language X-Fun [16], that was proposed as a common core language for the XML transformation languages XSLT and XQuery. XFun was developed within the LINKS team in cooperation with the University of Bratislava. This language is well suited for monitoring data trees. We believe that it can be adapted so that it can be used to monitor data graphs. This should take less than the first 6 month of the PhD project.

The second challenge is to reduce the high risk of blocking parallel threads, which typically arises when composing multiple monitoring processes, i.e., when monitoring a stream that was produced by monitoring some other stream. In order to avoid this frequent limitation, we propose to use the hyperstreaming approach, in which streams are generalized to streams with holes called hyperstreams. The holes can be filled later on by some concurrent thread, while the main stream can be processed without blocking, by jumping over the hole based on automata techniques. Query answering algorithms and tools for path queries on data trees were developed in the PhD project Momar Sakho (2016-19) within the Links project [18] supervised by Niehren & Boneva. These are to be lifted to path queries on data graphs.

Principales activités

The work programme with deliverables and provisional timetable:

Year 1: Develop the monitoring language G-Fun. Adapt X-Fun from data trees to data graphs. Merge the resulting version of XFun with XQP so that path queries on graphs can be defined within the language rather than externally as in X-Fun. We call the resulting language G-Fun

Year 2: Development and Implementation of G-Fun evaluator. Develop and implement an hyperstreaming evaluator for G-Fun. Extend the hyperstreaming evaluator for X-Path developed by Momar Sakho from data trees to data graphs. Lift the streaming implementation of XQP done by T. Sebastian (2011-15) to a hyperstreaming evaluator for G-Fun.

Informations générales

- Thème/Domaine : Représentation et traitement des données et des connaissances
- Statistiques (Big data) (BAP E)
- Ville : Villeneuve d'Ascq
- Centre Inria : CRI Lille - Nord Europe
- Date de prise de fonction souhaitée : 2020-10-01
- Durée de contrat : 3 ans
- Date limite pour postuler : 2020-04-23

Contacts

- Equipe Inria : LINKS
- Directeur de thèse : Niehren Joachim / Joachim.Niehren@inria.fr

A propos d’Inria

Inria, est l’institut national de recherche dédié aux sciences et technologies du numérique. Il emploie 2600 personnes. Ses 200 équipes-projets agiles, en général communes avec des partenaires académiques, impliquent plus de 3500 chercheurs et ingénieurs pour relever les défis du numérique, souvent à l’interface d’autres disciplines. L’institut fait appel à de nombreux talents dans plus d’une quarantaine de métiers différents. 900 personnels d’appui à la recherche contribuent à faire émerger et grandir des projets scientifiques ou entrepreneuriaux qui impactent le monde.

Consignes pour postuler

Sécurité défense : Ce poste est susceptible d’être affecté dans une zone à régime restrictif (ZRR), telle que définie dans le décret n°2011-1425 relatif à la protection du potentiel scientifique et technique de la nation (PPST). L’autorisation d’accès à une zone est délivrée par le chef d’établissement, après avis ministériel favorable, tel que défini dans l’arrêté du 03 juillet 2012, relatif à la PPST. Un avis ministériel défavorable pour un poste affecté dans une ZRR aurait pour conséquence l’annulation du recrutement.

Politique de recrutement : Dans le cadre de sa politique diversité, tous les postes Inria sont accessibles aux personnes en situation de handicap.
Year 3: Application and Experimentation with G-Fun

Develop prototypes for demonstration for RDF graphs based on G-Fun. Develop and perform an experimental evaluation of the new tool for hypstreaming G-Fun. Write up of the PhD thesis.

Planned collaborations:

This topic subscribes to the ANR Headwork (2016-21). The partners are the Université de Rennes, Inria Paris, Inria Lille, and the Museum national d'Histoire Naturel in Paris, the crowd sourcing company Wirk, and the coordinator is D. Gross-Amblard. The topic of Headwork is to develop general frameworks for data-centric workflows with human interaction, as needed for crowd sourcing systems. The language G-Fun will permit to develop monitoring tools for such datacentric workflows, or even as implementation platform for such workflows enhanced with streaming abilities. In particular, cooperations with the crowd sourcing company Wirk (previously FoolFactory) are envisaged.

Avantages

- Partial reimbursement of public transport costs
- Subsidized meals
- Leave: 7 weeks of annual leave + 10 extra days off due to RTT (statutory reduction in working hours)
- Possibility of teleworking (after 6 months of employment) and flexible organization of working hours
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Access to vocational training
- Possibility of French courses
- Social, cultural and sports events and activities
- Administrative support: Social security coverage / Help for Housing / Scientific Resident card and help for visa

Rémunération

1st and 2nd year: 1,982€ Gross monthly salary (before taxes)
3rd year: 2,085€ gross monthly salary (before taxes)