2021-03337 - PhD Position F/M Efficient geo-distributed data stream processing

Type de contrat : CDD
Niveau de diplôme exigé : Bac + 5 ou équivalent
Autre diplôme apprécié : A master degree in distributed systems and/or Cloud computing.
 Fonction : Doctorant

A propos du centre ou de la direction fonctionnelle

The Inria Rennes - Bretagne Atlantique Centre is one of Inria's eight centres and has more than thirty research teams. The Inria Center is a major and recognized player in the field of digital sciences. It is at the heart of a rich R&D and innovation ecosystem: highly innovative PMEs, large industrial groups, competitiveness clusters, research and higher education players, laboratories of excellence, technological research institute, etc.

Contexte et atouts du poste

This project will be conducted within the IRISA Myriads team which is working on the design of innovative infrastructures and middleware for future fog computing platforms. The team leader, Guillaume Pierre, is also the coordinator of the FogGuru European project.

Rennes is the capital city of Brittany, in the western part of France. It is easy to reach thanks to the high-speed train line to Paris. Rennes is a lively city and a major center for higher education and research. The job will take place within the INRIA/IRISA research center, which is internationally recognized for its research in the domain of information and communication sciences.

Mission confiée

The infrastructures which host most of the large internet applications are becoming increasingly distributed. To deliver excellent performance to their users while reducing the usage of wide networking resources, fog computing extends the traditional cloud computing model with additional resources located close to the end-user devices [1].

Fog platforms have very different geographical distribution compared to traditional clouds. Classical datacenter clouds are composed of many reliable and powerful machines located in a very small number of data centers and interconnected by very high-speed networks. In contrast, fog nodes are composed of a very large number of points-of-presence with a couple of weak and potentially unreliable servers, interconnected with each other by low-speed long-distance networks.

Data stream processing is an attractive paradigm for analysing IoT data in the fog before transmitting processed results to the cloud [2]. Stream processing engines allow programmers to express applications as a workflow of data transformations (operators) which execute over unbounded data streams. Workflows are organized as a directed acyclic graph where vertices represent operators and edges represent data streams. However, current data stream processing platforms such as Apache Flink [3] were not designed to operate in geo-distributed environments such as fog computing platforms, and parts of their implementation creates suboptimal performance in this context.

Principales activités

The objective of this thesis is to propose alternative implementations to deliver maximum processing efficiency in a fog computing environment. This will require the PhD student to identify the main sources of inefficiency and to propose alternative techniques. For example, stream processing engines introduce a variety of stateless or stateful operators to transform one or more input streams in one or more output streams. When the operator state cannot be easily partitioned between multiple replicas of the operator, it becomes necessary to replicate the state and to maintain its consistency every time this state is updated. This requires costly communications in case the replicas were geographically distributed. A possible approach to address this issue would be to exploit so-called "conflict-free replicated data types" [4] to better control the tradeoff between inter-replica synchronization and computation accuracy.


Compétences

- Excellent programming skills in Linux environments.
- Excellent communication and writing skills.
- Good command of English.
- Knowledge of the following technologies is not mandatory but will be considered as a plus:
  - Cloud resource scheduling
  - Distributed container systems: Kubernetes, Docker Swarm.
  - Single-board computers such as Raspberry PI
  - Python and shell scripting
  - Revision control systems: git, svn.

Informations générales

- Thème/Domaine : Systèmes distribués et intergiciels
- Système & réseaux (BAP E)
- Ville : Rennes
- Centre Inria : CRI Rennes - Bretagne Atlantique
- Date de prise de fonction souhaitée : 2021-10-01
- Durée de contrat : 3 ans
- Date limite pour postuler : 2021-04-15

Contacts

- Equipe Inria : MYRIADS
- Directeur de thèse : Pierre Guillaume / Guillaume.Pierre@irisa.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 scientifiques 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 et à l’innovation contribuent à faire émerger et grandir des projets scientifiques ou entrepreneuriaux qui impactent le monde. Inria travaille avec de nombreuses entreprises et a accompagné la création de plus de 180 start-up. L’institut s’efforce ainsi de répondre aux enjeux de la transformation numérique de la science, de la société et de l’économie.

L’essentiel pour réussir

Doing a PhD is a job unlike any other. Please read this document to decide whether a PhD is the right career move for you: https://medium.com/great-research/do-you-need-a-ph-d-7f629e6e28e0

Consignes pour postuler

Please submit online : your resume, cover letter and letters of recommendation eventually.

For more information, please contact guillaume.pierre@irisa.fr

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.

Attention : Les candidatures doivent être déposées en ligne sur le site Inria. Le traitement des candidatures adressées par d’autres canaux n’est pas garanti.
Linux distributions: Debian, Ubuntu.

Note that knowledge of French is not required for this position.

Avantages
- Subsidized meals
- Partial reimbursement of public transport costs
- Leave: 7 weeks of annual leave + 10 extra days off due to RTT
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Social, cultural and sports events and activities
- Access to vocational training
- Social security coverage

Rémunération
Monthly gross salary:
- amounting to 1982 euros for the first and second years and
- 2085 euros for the third year