Informations générales

- **Thème/Domaine** : Représentation et traitement des données et des connaissances
- **Systèmes d'Information (BAP E)**
- **Ville** : Montbonnot
- **Centre Inria** : CRI Grenoble - Rhône-Alpes
- **Département** : Département de l'Inria Grenoble
- **Date de prise de fonction souhaitée** : 2019-10-01
- **Durée de contrat** : 3 ans
- **Date limite pour postuler** : 2019-04-28

Contacts

- **Equipe Inria** : TYREX
- **Directeur de thèse** : Bonifati Angela / angela.bonifati@inria.fr

**A propos d'Inria**

Inria, l'institut national de recherche dédié aux sciences du numérique, promeut l'excellence scientifique et le transfert pour avoir le plus grand impact. Il emploie 2400 personnes. Ses 200 équipes-projets agiles, en général communes avec des partenaires académiques, impliquent plus de 3000 scientifiques pour relever les défis des sciences informatiques et mathématiques, souvent à l'interface d'autres disciplines. Inria travaille avec de nombreuses entreprises et a accompagné la création de plus de 160 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**

There you can provide a "broad outline" of the collaborator you are looking for what you consider to be necessary and sufficient, and which may combine :

- tastes and appetencies,
- area of excellence,
- personality or character traits,
- cross-disciplinary knowledge and expertise.

This section enables the more formal list of skills to be completed and tightened (reduced) :

- "Essential qualities in order to fulfil this assignment are feeling at ease in an environment of scientific dynamics and wanting to learn and listen."
- "Passionate about innovation, with expertise in Ruby on Rails development and strong influencing skills. A thesis in the field of "**" is a real asset."

**Consignes pour postuler**

The campaign is not open to local students who have not done any significant mobility.

**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 candidats sont appelés à postuler.

**2019-01476 - PhD Position F/M Distributed Query Analytics on Property Graphs [PhD Campaign 2019 - Campagne Doctorants Grenoble Rhône-Alpes]**

**Type de contrat** : CDD de la fonction publique

**Niveau de diplôme exigé** : Bac + 5 ou équivalent

**Fonction** : Doctorant

**A propos du centre ou de la direction fonctionnelle**

Grenoble Rhône-Alpes Research Center groups together a few less than 800 people in 35 research teams and 9 research support departments.

Staff is localized on 5 campuses in Grenoble and Lyon, in close collaboration with labs, research and higher education institutions in Grenoble and Lyon, but also with the economic players in these areas.

Present in the fields of software, high-performance computing, Internet of things, image and data, but also simulation in oceanography and biology, it participates at the best level of international scientific achievements and collaborations in both Europe and the rest of the world.

**Contexte et atouts du poste**

The candidate will pursue a PhD thesis under the supervision of Angela Bonifati and Pierre Genevès.

**Scientific Context.** Graphs and property graphs [BFV18] are becoming ubiquitous in many settings such as social and professional networks, collaborative networks for governmental agencies, health and energy consumption monitoring, scientific networks and knowledge graphs, alongside recommendation and fraud detection systems.

Property graphs represent the newest graph data model that enhance the existing RDF and graph database models with list of properties attached to nodes and edges. In property graphs and related query languages for property graphs (out of which ongoing standardization activities bringing to proposals such as GQL, GQL and G-Core [AAB18]), paths become first-class citizens in querying/analytical tasks, while key-value pairs are queried together with recursive paths in the underlying graphs.

As query languages for such graphs are under development, the respective modification operations are also newly defined [CAP, cypher]. The combinations of queries and updates bring to new analytical operations for such graphs, whose execution requires scalable platforms.

**Mission confiée**

**Scientific Objectives.** We envision the study of scalable graph query and update batch processing in a distributed setting. The current state of the art is Cypher for Apache Spark [CAP], in which for instance named queries and updates are already supported along with an initial graph schema specification for property graphs. We believe that in this direction there are several milestones: out of which (i) plugging in a static analysis approach in order to capture the interference of batches of queries and updates prior to compilation [G10, G13]; (ii) proving the equivalence and bidirectionality of the operations in declarative and procedural batches of graph operations in the presence of a graph schema [BBF19, BFG19]; (iii) extending graph query workloads to the recently expressive query languages and considering the case of mixed query/update workloads [BBC17, CEG13, gmark]. Especially with massive graph data, static analysis verification is desirable in order to avoid inconsistent results. The ongoing definition of schema languages for graphs has also a huge impact on the above objectives.

**Scientific, Societal and Economic Impact.** We believe that this topic has many scientific, societal and economic outcomes in France. Many businesses are in fact collecting their data under the form of knowledge graphs but they do not know how to analyze them or they do not know how to do it efficiently. Both cases are covered by the development of this PhD topic.

**External Collaborations.** We would like to pursue our ongoing collaborations with Eindhoven University of Technology, Netherlands (Prof. George Fletcher) and his team, as well as collaborations with the most successful European graph database company, Neo4j (Neo4j) (Dr. Hannes Voigt and Dr. Petra Selner). During the course of the thesis, the student might be able to have exchanges with the above colleagues and also scientific stays at those universities/companies are possible.

**Publications of the Inria Tyrex team (related to the PhD topic).**

- [BBF05] Michael Benedikt, Angela Bonifati, Sergio Flesca, Avinash Yac: Verification of Tree Updates for Optimization. CAV 2005: 379-393

**A propos de la mission confiée**

Scientific Objectives. We envision the study of scalable graph query and update batch processing in a distributed setting. The current state of the art is Cypher for Apache Spark [CAP], in which for instance named queries and updates are already supported along with an initial graph schema specification for property graphs. We believe that in this direction there are several milestones: out of which (i) plugging in a static analysis approach in order to capture the interference of batches of queries and updates prior to compilation [G10, G13]; (ii) proving the equivalence and bidirectionality of the operations in declarative and procedural batches of graph operations in the presence of a graph schema [BBF19, BFG19]; (iii) extending graph query workloads to the recently expressive query languages and considering the case of mixed query/update workloads [BBC17, CEG13, gmark]. Especially with massive graph data, static analysis verification is desirable in order to avoid inconsistent results. The ongoing definition of schema languages for graphs has also a huge impact on the above objectives.

**Scientific, Societal and Economic Impact.** We believe that this topic has many scientific, societal and economic outcomes in France. Many businesses are in fact collecting their data under the form of knowledge graphs but they do not know how to analyze them or they do not know how to do it efficiently. Both cases are covered by the development of this PhD topic.

**External Collaborations.** We would like to pursue our ongoing collaborations with Eindhoven University of Technology, Netherlands (Prof. George Fletcher) and his team, as well as collaborations with the most successful European graph database company, Neo4j (Neo4j) (Dr. Hannes Voigt and Dr. Petra Selner). During the course of the thesis, the student might be able to have exchanges with the above colleagues and also scientific stays at those universities/companies are possible.

**Publications of the Inria Tyrex team (related to the PhD topic).**

- [BBF05] Michael Benedikt, Angela Bonifati, Sergio Flesca, Avinash Yac: Verification of Tree Updates for Optimization. CAV 2005: 379-393


[gmark] gMark https://github.com/graphMark/gmark (accessed on: 2019-03-17)

Other references.


Principales activités

Avantages

- Subsidized meals
- Partial reimbursement of public transport costs
- Leave: 7 weeks of annual leave + 10 extra days off due to RTT (statutory reduction in working hours) or possibility of exceptional leave (sick children, moving home, etc.)
- Possibility of teleworking (after 6 months of employment) and flexible organization of working hours
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Social, cultural and sports events and activities
- Access to vocational training
- Social security coverage

Rémunération


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.

posts Inria sont accessibles aux personnes en situation de handicap.