**Main activities**

- Analyze and characterize existing storage solutions.

---

**2018-01037 - Post-Doctoral Research Visit / Converged Storage for Joint HPC and Big Data Processing**

**Contract type:** Public service fixed-term contract  
**Level of qualifications required:** PhD or equivalent  
**Fonction:** Post-Doctoral Research Visit

---

**About the research centre or Inria department**

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.

---

**Context**

The proposed position is located within the KerData (http://www.irisa.fr/kerdatal/doku.php?id=kerdata) research team at Inria Rennes. Led by Gabriel Antoniu, the KerData team focuses on scalable Big Data storage and processing on clouds and post-Petascale platforms, according to the current needs and requirements of data-intensive applications.

This work will be supervised by Alexandru Costan and Gabriel Antoniu. It will be done in the context of the frameworks of Package of the HPC-BigData IPL, where the KerData team (Rennes) and the Zenith team (Montpellier) are collaborating. It will use a concrete application scenario available thanks to the P4Go-Net application—one of the reference applications of the HPC-Big Data IPL. This work will be complementary to the work plan scheduled for a research master internship position (M2 level) that is proposed by the KerData team for the 2019 summer, focusing on the HPC-Big Data convergence at processing level (combining in situ/intransit processing with stream-based processing).

The postdoc hired on this position will co-advise the master intern.

---

**Assignment**

In the High Performance Computing (HPC) area, the need to get fast and relevant insights from massive amounts of data generated by extreme-scale computations led to the emergence of in situ and in transit processing approaches. They allow data to be visualized and processed in real-time, in an interactive way as they are produced, in opposition to the traditional offline analysis. In the Big Data area, however, the search for real-time analysis was materialized through a different approach: stream-based processing, which consists of treating an unbounded flow of small data items generated by many data sources arriving at high speed rates.

This illustrates how tools and cultures from HPC and Big Data Analytics (BDA) have evolved in divergent directions: essentially, they were motivated by different optimization criteria. However, recently a new converging processing paradigm is starting to be explored, by coupling state-of-the-art BDA tools with HPC simulations in order to predict and react to different situations in real-time.

This new hybrid HPC / BDA paradigm is illustrated in Figure 1. There are several use cases for this type of system. For instance, for traffic flow optimization, HPC simulations can be used to predict an optimized theoretical distribution of cars on a city's road structure and their results merged with predictions based on historical data in order to have more accurate predictions. Another example is connected car maintenance, where HPC simulations can be used to predict stress caused to car parts and, similarly, their results can be crossed with BDA predictions calculated using historical and real-time data.

To define such a system, it is necessary to address some issues related to the integration of those two very heterogeneous worlds. In this post-doc subject, we focus on storage issues. HPC and BDA applications have different requirements in terms of data storage. HPC applications are usually run on supercomputers and, hence, their file systems must allow for massive concurrent accesses of processes and tasks. On the other hand, BDA applications commonly follow a "write once, ready many" model, meaning that their file systems must be optimized for multiple parallel reading operations. In this context, HPC applications commonly use parallel file systems, such as Lustre, PVFS or OrangeFS while BDA applications use distributed key-value stores like DynamoDB or blob-based storage such as Ceph.

A first storage approach for hybrid HPC/BDA systems, would be to simply use separated native solutions for HPC and BDA frameworks, namely parallel file systems and key-value/blob-based storage systems, respectively. In the state of the art, however, there are also hybrid approaches, i.e. file systems or converged blob-based storage systems that can be used on HPC and BDA systems, such as HDFS and Thy, respectively. Such approaches, which had interesting preliminary performance results, open the possibility of using the same data storage strategy for HPC and BDA frameworks.

Our objective will be to analyze the available storage solutions, identify and group different extreme scale scenarios, and, for each of them, evaluate data storage systems' performances in terms of metrics such as processing latency (time taken to calculate a prediction, cf. Figure 1) and data throughput (observed data throughput from HPC and BDA frameworks to the storage systems, cf. Figure 1).

---

**General Information**

- **Theme/Domain:** Distributed Systems and middleware  
  Software Experimental platforms (BAP E)  
  **Town/City:** Rennes  
  **Inria Center:** CR - Rennes - Bretagne Atlantique  
  **Starting date:** 2019-01-01  
  **Duration of contract:** 12 months  
  **Deadline to apply:** 2018-11-14

---

**Contacts**

- **Inria Team:** KERDATA  
  **Recruiter:** Antoniu Gabriel / gabriel.antoniu@inria.fr

---

**About Inria**

Inria, the French National Institute for computer science and applied mathematics, promotes "scientific excellence for technology transfer and society". Graduates from the world's top universities, Inria's 2,700 employees rise to the challenges of digital sciences. With its open, agile model, Inria is able to explore original approaches with its partners in industry and academia and provide an efficient response to the multidisciplinary and application challenges of the digital transformation. Inria is the source of many innovations that add value and create jobs.

---

**The keys to success**

Essential qualities in order to fulfill this assignment are feeling at ease in an environment of scientific dynamics and willingness to learn and evolve in an international collaborative setting, in a strong team spirit.

---

**Conditions for application**

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

For more information, please contact gabriel.antoniu@inria.fr

---

**Defence Security:**  
This position is likely to be situated in a restricted area (ZRR), as defined in Decree No. 2011-1452 relating to the protection of national scientific and technical potential (PPST). Authorization to enter an area is granted by the director of the unit, following a favourable Ministerial decision, as defined in the decree of 3 July 2012 relating to the PPST. An unfavourable Ministerial decision in respect of a position situated in a ZRR would result in the cancellation of the appointment.

**Recruitment Policy:**  
As part of its diversity policy, all Inria positions are accessible to people with disabilities.

---

**Warning:** you must enter your e-mail address in order to save your application to Inria. Applications must be submitted online on the Inria website. Processing of applications sent from other channels is not guaranteed.
Performance evaluation of state-of-the-art storage solutions through large-scale experiments
Architecture design and implementation of a unified framework for HPC-Big Data processing
Large-scale evaluation of the unified framework with real applications available within the HPC-Big Data Inria Project Lab (IPL)
Write research papers and present them in reference international and national venues

Additional activities
- Co-supervision of master interns
- Participate to meetings with partners in the HPC-Big Data IPL and related collaborative projects

**Skills**
Applicants should have a doctoral degree in computer science and a strong background in operating systems and distributed computing. Excellent programming skills and experience with distributed experimental platforms are appreciated. Knowledge of C# is helpful.

**Benefits package**
- Subsidised catering service
- Partially-reimbursed public transport

**Remuneration**
monthly gross salary amounting to 2653 euros