Job offer –Post-doctorate in Computer Science
(*For non-French scientists only*)

Research Project Short Title as Submitted to CEFIPRA: "StreamSpan : Advancing Stream Data Systems with Spanning Events**"**
Principal Investigator contact (Name and email id): "Guillaume RASCHIA, LS2N; Polytech Nantes; Université de Nantes, guillaume.raschia@ls2n.fr"
Reference Number of the Job Offer: IFI_CEF_25_18

Project description

- **Keywords :** Database, Data Stream Processing, Streaming Systems, Time Window Aggregation

- **Context :** The post-doctorate candidate aims at reinforcing the *StreamSpan* project team such that s/he is expected to become a key task force to achieve scientific and technical goals. The *StreamSpan* CEFIPRA project is an Indian-French collaboration between IIIT Bangalore, India, and LS2N, France. The lab of digital sciences of Nantes (LS2N) is a large laboratory of 450 people. It is located in Nantes on 5 geographical sites. Its research activity is structured in 5 areas of expertise and 5 cross-cutting themes. Post-doc candidate will be member of the DUKe (Data User Knowledge) research team in the Data Science and Decision-Making area. Job location is Polytech Nantes, the Engineering School of Nantes University.

- **Abstract of the Research Project :** In today's data-driven world, the efficient processing of information relies heavily on understanding and leveraging the *time* dimension, especially in data streams. Events are either time points or time intervals, with the latter being prevalent across many application domains. For example, in a telco network, real-time monitoring of phone calls requires tracking numerous Key Performance Indicators (KPIs) to ensure high Quality of Service (QoS) for customers. Similarly, in Industrial IoT and digital twins, atomic machine tasks with start and end times could be collected and analyzed to prevent breakdowns or optimize processes, while music and video streaming platforms track listening/watch times. However, stream processing systems, including those designed for big data, currently lack native support for handling events with a lifespan. As a result, developers are burdened with the responsibility of designing and implementing complex data processing pipelines to address this limitation. The StreamSpan CEFIPRA project outlines a novel approach to address the challenges posed by so-called *spanning events* and introduces a data processing system architecture that efficiently handles spanning event streams.

  The proposed solution aims to develop an open-source framework capable of processing streams and big data in an *asynchronous* and *distributed* manner, while providing a *unified processing model* for both point and spanning events. Additionally, new benchmark schemes are proposed to assess and compare spanning events processing capabilities based on real-time industrial use cases, and a baseline implementation is developed on top of modern streaming system. Furthermore, the project aims at exploring *multi-query optimization* for stream processing systems, considering several factors such as window query types, time slicing schemes, processing time cost, input data stream profile as well as in- or out-of-order messages. By addressing these challenges, we aim both to contribute to the theory of data stream systems and to empower developers with off-the-shelf robust solutions for processing point events and spanning events seamlessly.

- **Scientific Objectives of the Project :** The post-doctorate candidate is expected to fully contribute to achieve the *StreamSpan* objectives, especially on three dimensions:

  - Theoretically, s/he participates to define how to integrate spanning event processing into big data stream systems. S/he specifically studies the slicing schemes to incorporate to streaming engines.
  - Practically, s/he develops prototypes and glues existing systems in a way that it facilitates re-use of the complex technology by end-users. S/he also validates the proposal against extensive benchmarks.

- Besides, s/he writes docs and project deliverables (related to her/his tasks), and s/he disseminates the *StreamSpan* outcomes with publication of research papers.

- **Methodology and Timeline of the Project :** In the first year (renewable for a second year), the post-doc candidate will spend 2 months to build and set up the technical environment as well as to update the state-of-the-art on streaming engines and spanning event stream processing. Next 4 months should be dedicated to co-elaborating the scientific proposal (how to handle spanning events in big data stream systems). Then, 3 more months are dedicated to implementing on the targeted platform. After 9 months, one enter a refinement step to fix and enhance the proposal as well as to debug and improve the implementation. All along the process, concepts and implementation should be well-documented enough to communicate and transmit to teammates. A first draft of a demo paper and a research paper as well are expected before the end of the year.

  All the tasks require to constantly discuss and share opinions and issues with the other *StreamSpan* team members, either on site (Polytech Nantes) or abroad (IIIT Bangalore). It is a 6-to-8 team.

Candidate profile

- Candidates can be all nationalities except French. In case of double nationality (French and another one), the candidate is not eligible. In the context of CEFIPRA, Indian candidates are preferred
- Applicants for PhD must have a master's degree (or be in the process of obtaining one) or have a University degree equivalent to a European Master's (5-year duration) to be eligible at the time of the deadline of the call
- Applicants for post-doctorate must have a PhD degree (or be in the process of obtaining one)
- No competences in French language is required
- Candidate competences : algorithms (and analysis of), functional programming, advanced data structures, databases. [bonus] data stream systems, parallel computing and distributed systems
- Candidate know-how : multi-paradigm programming, complexity analysis, research paper writing, scientific communication
- Expected starting date: September onwards
- Expected duration: 12 months

How to candidate ?

Documents to be provided :

1. A cover letter (reasons for the candidature, professional project …) max 2 pages
2. A copy of the master's degree or a proof of the program followed (and expected date of end) OR A copy of the PhD degree or a proof of the PhD program followed (and expected date of defense) max 1 page
3. A copy of results for previous scholarship (max 3 pages)
4. International curriculum vitae (max 2 pages)
5. Two letters of recommendation: one from any Indian institution and one from the French institution planned to host the candidate –mandatory- (max 2 pages)
6. All should be submitted within 1 pdf file of no more than 10 pages.

Applications should be submitted to the following email address: msi@ifindia.in mentioning the reference number of the Job offer clearly.

Research Project Title as Submitted to CEFIPRA: "StreamSpan: Advancing Stream Data Systems with Spanning Events**"**

Candidates are requested to contact the French scientific principal investigator of the project before submission. A recommendation letter from the scientific principal investigator is mandatory.

Benefits:

- Monthly allowance of 2400 euros for Post-Doc
- Travel allowance
- University fee
- Carte de séjour fee
- Campus France management fee
- Registration to the French social security scheme

Selection process:

Selection is made by a dedicated selection committee of at least 4 persons. Decisions will be transmitted by the Embassy of France to CEFIPRA. **No consideration will be given for candidates with no recommendation letter from the French institution.**

Criteria for applicants' selection:

Academic excellence

- Excellence of the Academic background, Academic records, Honors, Letters of support, Participation to international research projects, exchange programmes and conferences.

Motivation and qualities

- Academic maturity: appropriation of the thesis project (stakes and contexts) • Quality of the presentation (oral expression, skills for synthesis, English level) • Maturity of the professional project: capacity to project her/himself within five years in terms of career development.

About CEFIPRA:
Indo-French Center for the Promotion of Advanced Research (CEFIPRA/IFCPAR) is an Indian body which promotes scientific cooperation between France and India in advanced fields of Science and Technology. It is supported by the Department of Science and Technology, Government of India and the Ministry of Europe and Foreign Affairs of the French government