Job offer: Research Engineer

The Big Data Analytics Team at Ecole Polytechnique is seeking a research engineer in the area of big data analytics and cloud computing. The appointment would be for two years, with possible renewal for additional years.

I. Overall Project

Today's big data analytics systems are best effort only: despite the wide adoption, they still lack the ability to take user budgetary constraints and performance goals, and automatically configure an analytic job to achieve those goals. In this project, we aim to transform data analytics in the cloud by developing a next-generation unified data analytics optimizer: the optimizer takes as input a user analytic task as well as a set of user budgetary constraints and performance goals, and produces as output a cloud instance and runtime parameters of the job that best meet the user objectives. Our data analytics optimizer employs state of the art machine learning and optimization techniques to find the best cloud instance and runtime parameters for each analytical task. This project is funded by a prestigious 5-year research award from the European Research Council (ERC).

II. Research Environment

The research engineer will find an active and collaborative environment at Ecole Polytechnique, with world-renowned researchers in data analytics systems, data mining, machine learning, and statistics. Ecole Polytechnique is a French public institution of higher education and research, located in Palaiseau 45 minutes southwest of Paris. It is considered the most prestigious engineering school in France, with well-known educational programs in science and engineering. Among its alumni are three Nobel prize winners, one Fields Medalist, three Presidents of France, and many CEOs of French and international companies.

III. Job Description

The engineering support of the ERC project focuses on building and maintaining software systems such as HDFS, Yarn, Spark on a 1000-core data processing cluster, performance profiling through trace collection, performance modeling using machine learning techniques, and running and profiling analytical jobs in the cloud such as Amazon EC2.

We seek applicants with a strong interest in Big Data technology and a fast learning ability to master the technical skills related to Hadoop, Spark, container technology, cloud images, profiling, resource management and optimization over the course of the employment. The engineer will also have a rare, valuable opportunity to work with world-renowned researchers and participate in cutting-edge research that has the potential to transform data analytics in the cloud in the near future.

Responsibilities include:
* Installing and ongoing administration of HDFS, Yarn, Spark, and other relevant software tools on a 1000-core cluster.
* User account management on the cluster, including setting up Linux users and accounts for specific software systems.
* Running large-scale analytical workloads using Spark, and collecting OS and application traces for performance modeling.
* Running the above analytical workloads in the cloud such as Amazon EC2.
* Profiling different cloud instances.
* Developing an online service that runs the optimizer for user analytical jobs in the cloud.

**IV. Skills Required from the Applicant**

* Most essential requirements: The applicant must have good knowledge of computer systems, in particular, Unix or Linux, network protocols, and web services. The applicant is also expected to have general operational expertise such as good troubleshooting skills, understanding of system's capacity, bottlenecks, basics of memory, CPU, OS, storage, and networks.

* It is strongly preferred that the applicant has prior experience in deploying software systems such as Hadoop and Spark in dedicated clusters or in the Cloud, as well as developing analytical programs using these systems.

* Knowledge of SQL, relational databases, and the design principles of Spark is a plus.

**V. Application**

The position is available immediately. Application is open until the position is filled. The interested applicant is asked to submit a resume and 2 names of references to Prof. Yanlei Diao at jessica.gameiro@polytechnique.edu.