

## **OSDC/PIRE research proposal**

**LARC** (Laboratory of Computer Networks and Architecture) is a laboratory inside the Computing and Digital Systems Department of Escola Politecnica (School of Engineering) from Universidade de Sao Paulo (PCS-EPUSP), Brazil. LARC technical staff is responsible for academic activities (undergraduate and graduate courses), researches and projects in the fields of Computer Networks and Architecture. Currently, over 50 professionals work at LARC, among them Ph.D., and Master candidates, engineers, and interns. These professionals have large academic, professional and scientific experience and conduct relevant, state-of-art projects in their fields.

LARC develops different types of projects, such as Research Projects (in order to advance knowledge) and Corporate/Community Projects (using the knowledge to contribute with the country's development).

For more detailed information please visit: <http://www.larc.usp.br/en>  
<http://www5.usp.br/en/>

### **Real-time Big Data Analysis**

Big data is data that exceeds the processing capacity of conventional database systems. Processing this data can uncover great values to better define organization's strategies and product development. The big data can be processed either in batch or real time. There are different frameworks that are applied for each one of these scenarios. One of the research groups at LARC has been developing code and testing Hadoop using climate related data sets and considering batch distributed processing.

The goal of this research is to develop and test code for Big data frameworks that have been developed to process big data in real or near-real time. The processing will be performed using the Open Science Data Cloud infrastructure.

The student will get familiar with the cloud environment, design predictive analytics solutions using popular Big data frameworks, such as Storm and SQLStream, and architect data stream collection and processing.

The student should meet the following skill requirements:

- Software development (Python, Java)
- Web application frameworks
- Linux OS
- Strong foundation knowledge on distributed systems
- Ability to architect high-volume solutions