OSDC/PIRE research proposal

LARC (Laboratory of Computer Networks and Architecture)
University of São Paulo (USP)

LARC (Laboratory of Computer Networks and Architecture) is a laboratory inside the Computing and Digital Systems Department of Escola Politecnica (School of Engineering) from University of São 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 M.Sc. 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/

Analyzing Big Data Applications over ESNet's Science DMZ Model

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. Usually the data is collected/stored away from the processing facility, posing significant challenges for its transport on top of data networks, especially when this data has to cross firewalls. One possible solution is the use of ESNet's Science DMZ model, where scientific data is segregated in a specialized, low-latency, high-bandwidth network segment with different security policies.

The goal of this research is to investigate the viability and benefits of the Science DMZ model combined with data streaming solutions such as Apache Spark for a Big Data scenario where data and processing are geographically distant. In this particular research, the idea to have an application running at an OSDC cluster being fed by data sources at our University, both in and out of the Science DMZ segment. The student will get familiar with cloud-computing environment and architect data-stream collection and processing while taking into account network requirements and challenges for this kind of application.

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
- Strong networking concepts.