

## **Efficient Data Processing in High Performance Big Data Platforms**

**Nicola Tonellotto (ISTI CNR)**

Abstract: One of the largest and most used big data platforms used nowadays is represented by the Web search engines. With the ever-growing amount of data produced daily, all Web search companies must rely on distributed data storage and processing mechanisms hosted on computer clusters composed by thousands of processors and provided by large data centers. This distributed infrastructure allows to execute a vast amount of complex data processing that provide effective insights for the users, at near real-time with sub-second response times. Moreover, the most recent scientific advances in big data analytics exploit machine learning and artificial intelligence solutions. These solutions are particularly computationally expensive, and their energy consumption has a great impact on the overall energy consumption at a global scale. In this seminar we will discuss some recent investigations about (i) novel efficient algorithmic solutions to improve the usage of hardware resources (e.g., reducing response times and increasing the throughput) when complex machine learned models for processing large data collections and (ii) online management of computational load to reduce the energy consumption by automatically switching among their available CPU frequencies to adapt to external operational conditions.

Short bio: Dr. Nicola Tonellotto (<http://hpc.isti.cnr.it/~khast/>) is a researcher within the High Performance Computing Lab at Information Science and Technologies Institute of the National Research Council of Italy. His main research interests include high performance big data platforms and information retrieval, focusing on efficiency aspects of query processing and resource management. Nicola has co-authored more than 60 papers on these topics in peer reviewed international journal and conferences. He lectures on Computer Architectures for BSc students and Distributed Enabling Platforms for MSc students at the University of Pisa. He was co-recipient of the ACM's SIGIR 2015 Best Paper Award for the paper entitled "QuickScorer: a Fast Algorithm to Rank Documents with Additive Ensembles of Regression Trees".