

High performance computing in big data analytics

Virginia NICULESCU

Babeş-Bolyai University, Cluj-Napoca, M. Kogălniceanu Str., no. 1, 400084 Cluj-Napoca, Romania
E-mail: vniculescu@cs.ubbcluj.ro

Abstract

For long time High-Performance Computing (HPC) has been critical for running large-scale modeling and simulation using numerical models. The big data analytics domain (BDA) has been rapidly developed over the last years to process torrents of data now being generated in various domains. But, in general, the data analytics software was not developed inside the scientific computing community, and new approaches were adopted by BDA specialists. Data-intensive applications are needed in varied field ranges from advanced research— as genomics, proteomics, epidemiology and systems biology—to commercial initiatives to develop new drugs and medical treatments, agricultural pesticides and other bio-products. Big data processing is still needed in the more HPC traditional domains as physics, climate, and astronomy, but even there adopting data-driven paradigms could bring important advantages. On the other side BDA needs the infrastructure and the fundamentals of HPC in order to face with the needed computational challenges. There are important differences in the approaches of these two domains: those that are working in BDA focus on the 4Vs of big data which are: volume, velocity, variety, and veracity, while HPC scientists tend to focus on performance, scaling, and the power efficiency of a computation. As we are heading towards extreme-scale HPC coupled with data intensive analytics, the integration of BDA and HPC is a necessity and a current hot topic of research.

Keywords:

High Performance Computing; Big Data Analytics; Computational Intensive; Data Intensive