Data-Aware Support for Hybrid HPC and Big Data Applications

Silvina Caíno-Lores, Florin Isaila, Jesús Carretero
Department of Computer Science
University Carlos III of Madrid
Leganés, Spain
{scaino, fisaila, jcarrete}@inf.uc3m.es

Nowadays, high-performance computing (HPC) applications are increasingly demanding data analysis and visualization over major datasets, which is shifting these originally computationally intensive systems towards parallel data-intensive problems. On the other hand, Big Data (BD) applications are requiring the performance level of the supercomputing ecosystem. As a result, this general trend is leading to greater overlapping between the HPC and BD paradigms.

Nevertheless, HPC and BD systems have been traditionally built to solve different problems: HPC focuses on CPU-intensive tightly-coupled applications, and BD tackles large volumes of loosely-coupled tasks. These objectives have determined the underlying architectures of HPC and BD infrastructures. In particular, most HPC infrastructures are architected so that compute and storage are decoupled but connected through high-speed interconnections, as in grids or clusters. In contrast, BD environments co-locate computation and data and focus on elasticity, thus clouds become their preferred infrastructure.

In this scenario, upcoming applications will suffer the lack of an ideal environment able to cope with their computing and data requirements. Recent works have suggested the opportunity of combining the HPC and BD approaches to alleviate this issue, but there is still no current approach to exploit these opportunities for generalist hybrid applications. As a consequence, our ultimate research problem is building a platform able to manage applications built for computationally-intensive simulations, data-intensive analysis, or both, with a focus on data-awareness and fault-tolerance.

In this presentation we will present the results of the first stage of our research, in which we explored the effects these BD-inspired paradigms could have in current scientific applications by analyzing a meaningful use-case. In addition, we will introduce our research roadmap to build a system able to bridge the gap for data-intensive scientific applications, while taking advantage of upcoming advances in supercomputing infrastructures like accelerators and NVRAM.