How to understand and improve the performance of your parallel applications using the POP Methodology

Fouzhan Hosseini [1]

1 The Numerical Algorithms Group (NAG)

HPC applications are often very complex and their behavior depends on a wide range of factors from algorithms, to programming models, library and language implementations and hardware. The task of understanding performance bottlenecks of a parallel code and making improvements often ends up being a daunting trial and error process. To make the problem even more complicated, many HPC applications inherit different layers of legacy code, written and optimized for a different computing era. To optimize the performance of a parallel application, the first step is to understand the behavior of the application. However, there is often a lack of quantitative understanding of the actual behavior of HPC applications. The Performance Optimisation and Productivity (POP) Centre of Excellence, funded by the EU under the Horizon 2020 Research and Innovation Programme, attempts to establish a quantitative methodology for the assessment of parallel codes. This methodology uses a set of hierarchical metrics, where each metric represents relative impact of one cause of inefficiency in parallel codes. These metrics provide a standard, objective way to characterise different aspects of the performance of parallel codes. In this talk, I will review the development of these metrics and give examples of how use of these metrics allowed us to quickly identify the performance bottlenecks of applications from different domains of science and engineering. In addition, the POP methodology facilitates training HPC experts and performance analysts by defining a common and systematic approach to assessing and improving parallel codes.