The peril of the petascale: looming challenges in large-scale computational science

John Clyne
Computational and Information Systems Laboratory
NCAR

Emerging petascale computing platforms will offer unprecedented opportunities for large-scale numerical simulation, but at the same time will pose extraordinary challenges for data analysis and visualization. While CPU and GPU performance has advanced at a rate meeting or exceeding Moore's law, other computing technologies essential to the end-to-end computational science process, most notably IO, have lagged far behind. Frequently the bottleneck in interactive data exploration is not the speed at which we can compute or render, but simply the rate at which we can recover data from disk. Petascale computing will only exacerbate this problem. Without significant improvements in data transfer rates, or radical changes in the composition and operation of today's HPC centers, the utility and the scientific return on the investment in these petaflop capable machines will be limited.

Progressive data access strategies have the potential to decrease the analysis and visualization demands on storage, communication, processing, and graphics by orders of magnitude. But their effective use in a petascale environment will require a departure from present practices that will impact both scientists and HPC resource providers alike. This talk will present results of our recent work in the area of wavelet-based compression and progressive data access. We have incorporated progressive access into an open source visual data analysis tool, VAPOR, which is capable of interactively exploring terascale scientific data sets using only an ordinary laptop computer. We will also discuss our thoughts on extending our current methods to support future petascale applications.