June 20th, 2019

Use cases of lossy compression for floating-point data

Franck Cappello
University of Illinois, USA

Abstract
Architectural and technological trends of systems used for scientific computing call for a significant reduction of scientific datasets that are composed mainly of floating-point data. In this talk, we present experimental results of currently identified use cases of generic lossy compression to address the different limitations of scientific computing systems. We show from a collection of experiments run on parallel systems of a leadership facility that lossy data compression not only can reduce the footprint of scientific datasets on storage but also can reduce I/O and checkpoint/restart times, accelerate computation, and even allow significantly larger problems to be run than without lossy compression. These results suggest that lossy compression will become an important technology in many aspects of high-performance scientific computing.

Short bio

Franck Cappello received his Ph.D. from the University of Paris XI in 1994. He is a permanent senior researcher at INRIA, and a visiting research professor at the University of Illinois. He led the roadmap and strategy efforts for projects related to resilience at the extreme scale since 2008, as a member of the executive committee of the International Exascale Software Project. He is an IEEE Fellow and recipient of the 2018 IEEE TCPP Outstanding Service award. He initiated and directed the Grid'5000 project from 2003 to 2008, helping hundreds of researchers for their experiments in parallel and distributed computing. In 2009, he created with Marc Snir the Joint-Laboratory on Petascale (now Extreme Scale) Computing, that gathers since 2014 seven of the most prominent research and production centers in supercomputing. In 2016, Cappello became the director of two Exascale Computing Project software projects related to resilience and lossy compression. Through his 25 years research career, Cappello has directed the development of several high impact software, including XtremWeb, a pioneer Desktop Grid software, the MPICH-V fault tolerance MPI library, the VeloC multilevel checkpointing environment, the SZ lossy compressor and the Z-Checker tool to assess the errors produced by lossy compressors. https://www.anl.gov/profile/franck-cappello