Integrating Solid State Disks in Storage Resource Management of Scientific Data

E. J. Otoo
School of Computer Science
University of the Witwatersrand
Johannesburg, South Africa

Abstract

The problem of how to organize the massively large volumes of scientific data, originating primarily from observational instruments and high performance computer simulations of phenomena, for subsequent insightful analysis pose enormous challenges to scientists and data-management experts. High resolution instruments and high performance computing, generate and stream data to disks at data centers and at rates much higher than the input output (I/O) bandwidth of parallel storage systems allow. Recent advances in solid state drives (SSDs), may now provide some solutions on how to manage the I/O of data-intensive scientific computing. The SSDs can be integrated as another tier-level between main memory and spinning Hard Disk Drives (HDDs) in the storage hierarchy. SSDs are now of sufficient capacity (320 ~ 512 GB) and long term reliability that they can be used, not only to boost the response time of data accesses to/from persistent storage but used for energy conservation at the data centers. Idle disks that are not accessed for long periods of time are spun down instead of retaining thousands of disks to be continuously spinning to maintain the massively large datasets. We present results of simulations of two different configuration schemes, for integrating SSDs in the multi-tier storage hierarchy, that improve both the effective I/O bandwidth and conserve up to 40% energy.