30 Years of High Performance Storage System Collaboration

June 1, 2022

This year marks the 30th anniversary of the High Performance Storage System (HPSS) Collaboration. The HPSS Collaboration is comprised of five US DOE HPC laboratories: Lawrence Berkeley, Lawrence Livermore, Los Alamos, Oak Ridge, and Sandia, and, as partner from industry, IBM. As Development Members, the six parties created HPSS which is a software-defined scalable long-term datastore. Today, HPSS is currently used by sites around the world serving more than 4.5 Exabytes of production data that is growing at rate of over 500 Petabytes a year and is still developed and supported by the same parties.

In the late 1980s, HPC leaders recognized the need for long-term archiving that was high-speed and massively scalable which would leverage distributed hierarchical storage capacity management to meet performance requirements of their supercomputers. Moreover, HPC leaders were keenly aware that applying best practices in software development, management, and quality from industry would be key to success. In response, the National Storage Laboratory (NSL) was organized to investigate, demonstrate, and commercialize high-performance hardware and software storage technologies focused on removing network computing bottlenecks. The HPSS Collaboration, established in 1992, grew out of the NSL's results and experience.

A software product that thrives for three decades is exceedingly rare, particularly one that is actively developed in a collaborative manner by multiple geographically distributed organizations. When the HPSS Collaboration began, the terascale era—10^{12} CPU floating point operations per second (flops)—was still 5 years away. The HPSS architecture, implementation, and focus on collaboration have allowed it to evolve to meet the demands of that era as well as the subsequent petascale era (10^{15} flops). DOE and industry investments in HPSS have resulted in a team spanning generations with skills and experience in software development as well as storage systems engineering, complex systems integration, and remote deployment and support. The team has provided over 10 major releases of HPSS yielding remarkable operational efficiencies, performance, and storage capabilities all while achieving the greater speeds demanded of scientific, research, and commercial endeavors. HPSS is poised to continue its evolution throughout the emerging exascale era (10^{18} flops).

HPSS Collaboration's Development Members realized early on that no single organization has the necessary experience and resources to meet all the challenges represented by the growing imbalance between computing power and data storage capabilities. Designed not only to provide scalable high performance and capacity, HPSS takes advantage of hierarchies of storage technologies including solid state disk, magnetic disk, tape, and cloud allowing it to provide balanced total cost of ownership for archival storage. Additionally, needs for data accessibility and related protocols change over time. As a result, many industry products have come and gone in three decades. Meanwhile, a hardware-agnostic approach from the beginning has allowed production instances of HPSS smoothly transition across the constant network, server, and storage media changes. HPSS continues to lead and adapt to needs of the time while honoring its primary mission of long-term data stewardship of the crown jewels of data for government, academic, and commercial organizations around the world.

For more information:
http://www.hpss-collaboration.org

High Performance Storage System