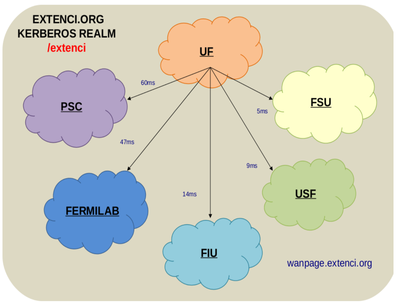
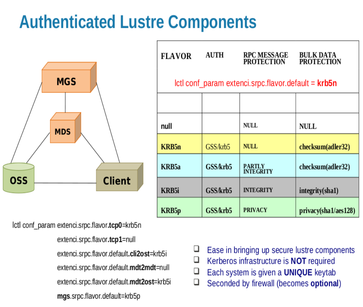
Secure Wide Area Network Access to CMS Analysis Data Using the Lustre Filesystem

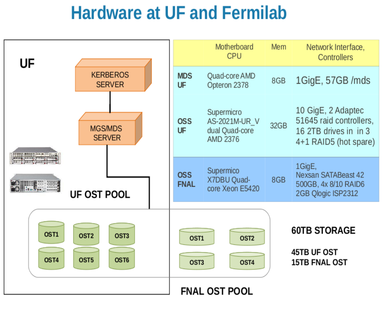
CHEP 2012, New York, USA, 2012

D. Bourilkov, P. Avery, M. Cheng, Y. Fu, B. Kim University of Florida
J. Palencia, R. Budden, K. Benninger, Pittsburgh Supercomputing Center
J. L. Rodriquez, J. Dilascio Florida International University
D. Dykstra, N. Seenu Fermi National Accelerator Laboratory

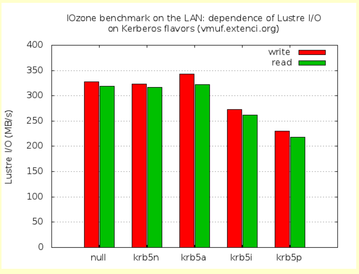
The ExTENCI project has designed and implemented a secure, distributed over the wide area network filesystem, based on the Lustre filesystem. The system comprises pools located at the University of Florida and at Fermilab, providing 60 TB of storage. It is used for remote access to analysis data from the CMS experiment at the Large Hadron Collider, and from the Lattice Quantum ChromoDynamics (LQCD) project. Security is provided by Kerberos authentication and authorization with additional fine grained control based on Lustre ACLs (Access Control List) and quotas. We investigate the impact of using various Kerberos security flavors on the I/O rates of CMS applications on client nodes reading and writing data to the Lustre filesystem, and on LQCD benchmarks. The clients can be real or virtual nodes. We are investigating additional options for user authentication based on user certificates.



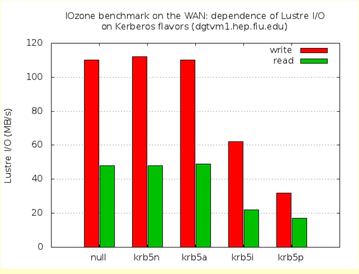


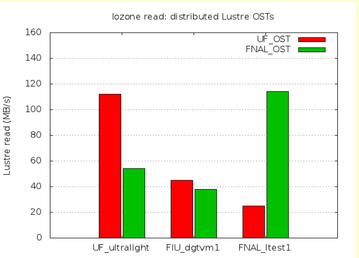


The Lustre filesystem is used in some of the world's largest and most complex computing environments. It provides high performance, scaling to tens of thousands of nodes and petabytes of storage with groundbreaking I/O and metadata throughput. The Lustre 2.* release offers a number of significant features and enhancements, including statehead, asynchronous glimpse lock (AGL), imperative recovery, large xattrs, MDS-survey, multi-threaded ptlrpcd, OSD API, parallel directory operations (pDirOps), quota protocol, code improvement, client parallel checksums, changelogs, commit on share (COS), lustre rsync, and size-on-MDS (preview) while supporting OEL 5, RHEL 5, SLES 10/11, and Fedora 11 (clients only).

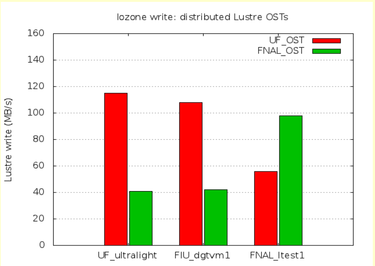


Impact of Kerberos flavors on I/O rates: on the LAN (left) and on the WAN (right). The effects are much more pronounced when accessing files over the WAN. We use krb5n, which has very minor overhead, for all subsequent tests.

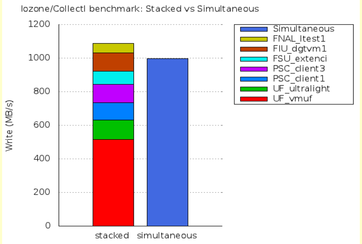


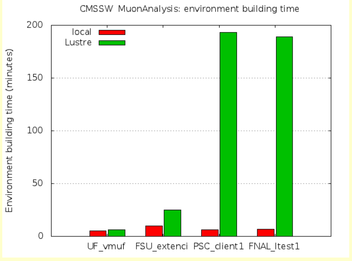


Distributed OST pools at the University of Florida and FNAL: I/O rates (as determined by IOzone) depend on the "distance" from the client to the server: best for LAN, worst for WAN over longer distances. The RTT: UF-FNAL 47 ms, UF-FIU 14 ms, FIU-FNAL 51 ms.

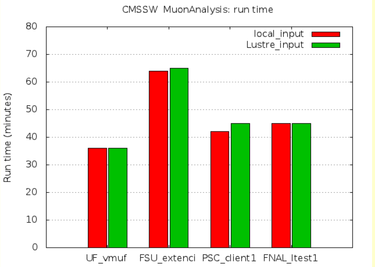


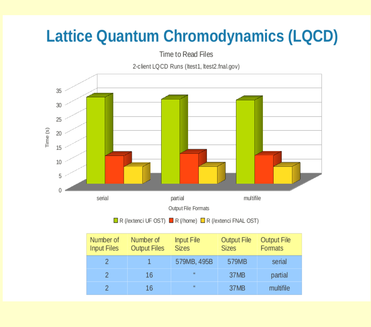
Scalability of our system: For a realistic mix of clients accessing data over the LAN or the WAN, we ran sets of tests first sequentially (stacked), then all clients in parallel. The simultaneous run produces I/O rates close to the stacked rates.





Tests with CMSSW. Compilation and link (involving many small files) time (left) is good over the LAN or for "close" clients: FSU (RTT 4 ms), much slower for "far" clients: FNAL and PSC (RTT 46-47 ms). A CPU dominated run test performing reconstruction over a 2 GB data file gives comparable run times (right).





Lattice QCD benchmarks: read times (left) and write times (right); shorter is better. A distributed Lustre filesystem with OSS at UF and OST pool and FNAL gives results competitive with using local files.