Final Progress Report

Petascale Data Storage Institute

DE-FC02-06ER25766

Peter Honeyman, Principal Investigator
University of Michigan
2260 Hayward St., Ann Arbor, MI 48109-2122
(734) 763-4413, honey@citi.umich.edu

1. Technical Areas of Progress

pNFS research and development

pNFS is an extension to NFSv4 that helps clients overcome NFS scalability and performance barriers. Like NFS, pNFS is a client/server protocol implemented with secure and reliable remote procedure calls. pNFS departs from conventional NFS by allowing clients to access storage directly and in parallel. This helps overcome server bottlenecks inherent to NAS access methods.

Making pNFS available to petascale researchers required coordinated progress in several dimensions, a confluence of multiple processes that took nearly ten years. The protocol had to be specified and published by the IETF. Nascent implementations had to track changes in the draft specification, changes in the Linux kernel, and interoperate with one another. The process by which modifications are accepted into the Linux kernel by maintainers and developers (and ultimately by Linus Torvalds himself) itself required consensus and compromise. Finally, pNFS support needed to be provided by the major Linux distributors (Red Hat, SUSE, etc.) and storage vendors (Netapp, EMC, IBM, Microsoft, etc.) on both the engineering and product sides.

The specification requirement was met in December 2008, when the IETF approved the NFSv4.1 specification, which incorporates pNFS and pNFS file layouts, as a Proposed Standard as well as the (separate) specifications of pNFS object and block layouts.

Throughout the PDSI project, CITI was the key contributor to the Linux-based, open source implementation of NFSv4.1 and pNFS. Considerable effort was devoted to the process of refining the IETF specification, which underwent dozens of preliminary drafts, rebasing implementations to the latest Linux kernel, which itself changed quarterly, and to testing interoperability with other developers. At last, the Linux pNFS implementation was incorporated into the Linux mainline kernel, although this was achieved after the PDSI project ended.

Scalability test bed

CITI research interns developed a small-scale test bed to explore pNFS scaling properties. Most of what we learned is that a scalable iSCSI-based storage backend is beyond the present capacity of the mainline Linux kernel. We had more success with Windows 2008 Server. Small scale testing demonstrated linear scaling up to the limits of the storage backend.

2. Publications

None

3.Personnel

Faculty (2), graduate students (2), undergraduate students (4).

Peter Honeyman, Research Professor of Computer Science and Engineering (University of Michigan Principal Investigator for PDSI).

J. Bruce Fields, Assistant Research Scientist (PDSI-related research and development at the University of Michigan).

Josef Sipek, University of Michigan doctoral pre-candidate in Computer Science and Engineering.

Eaman Jahani, University of Michigan doctoral pre-candidate in Computer Science and Engineering.

Eric Anderle, Research Intern, University of Michigan undergraduate studying Computer Science and Engineering.

Michael Groshans, Research Intern, University of Michigan undergraduate studying Computer Science and Engineering.

Shatarupa Nandi, Research Intern, University of Michigan undergraduate studying Computer Science and Engineering.

Bryan Smith, Research Intern, University of Michigan undergraduate studying Computer Science and Engineering.

4. Outreach and Tech Transfer

Peter Honeyman organized a BoF session on HPC Storage at the FAST Conference, February 2010.

CITI participated in the spring 2010 pNFS/NFSv4.1 interoperability workshop (Connect-a-thon), held in Santa Clara, CA.

CITI organized and hosted the summer 2010 pNFS/NFSv4.1 interoperability workshop (Bake-a-thon), attended by dozens of technologists from around the world.

CITI participated in the fall 2010 pNFS/NFSv4.1 interoperability workshop (Bake-a-thon), held in Hopkinton, MA.

CITI hosted a weekly conference call devoted to engineering concurrency and parallelism in NFSv4. Attendees included engineers from CITI, SGI, CGS, IBM, Google, and Red Hat.

5. External Presentations

None.