

# SciDAC PDSI LANL 2<sup>nd</sup> half FY07 report info

## Accomplishments

### *LANL Planned Data Release*

Los Alamos National Lab has an ongoing effort to release data helpful to and requested by the research community. In the second half of the reporting period, LANL continued to release trace data and file system statistics. At the 2007 PDSI Meeting, Andy Konwinski, a LANL summer student, reported on his efforts to test out and evaluate the tracing various testing methods looking for the method with the least amount of overhead yet supplies the system and desired library calls. In addition, LANL has released their tracing mechanism. LANL has also released the first file system statistics data using a tool from Carnegie-Mellon University and Panasas; fsstats. In the first half of FY08, LANL plans to release fsstats data from thousands of LANL workstation over a five to ten year period along with current data from our production “scratch” file systems.

### *HECURA File Systems and I/O Projects, review, coordination and Workshop*

As part of the PDSI, LANL coordinated the second half FY07 HECIWG activities. As a part of this activity, 9 HECURA projects were reviewed (mid year review) to ensure the HEC organizations are being well served by the research being conducted. The 2007 HECIWG Workshop was held and attended by over 100 university and industry researcher influential in the FSIO area. Feedback from the workshop and HECURA reviews led to the development of a HECIWG FSIO website (<http://institute.lanl.gov/hec-fsio/>) including documentation from past workshops, abstracts and information on all the HECURA FSIO and CPA projects, and links to relevant project web sites.

## Publications

We had no publications from LANL directly but our failure data release, synthetic application code release, and consultation assisted with the following:

Eric Lalonde. “A Characterization of LANL HPC Systems”, Masters Thesis, University of California, Santa Cruz. 2007

Olen Davis, Kari Macklin, Baily Kelly “Parallel Search using multiple Google Desktops in Parallel”, Colorado School of Mines, Technical report, 2007.

Clay Baenziger, Bruce Bugbee, Ryan Ford, Charlie, Grammon. “LANL Supercomputing Data Analysis”, Colorado School of Mines Technical Report, 2007.

## Supported people

Gary Grider – LANL Staff – PDSI Task PI

James Nunez – LANL Staff

John Bent – LANL Staff

Milo Polte – CMU student summer intern at LANL

Andy Kowinski – Wisconsin student summer intern at LANL

## **Outreach, Collaborations and Tech Transfer**

### Outreach/Collaborations

- Educational Outreach:
  - Two summer interns at LANL worked on PDSI related topics; Andy Konwinski worked on application tracing and Milo Polte worked on multi-dimensional file systems.
  - Worked with two groups of undergraduates students on summer projects at the Colorado School of Mines' Field Session; applying a statistical data analysis to the LANL event/usage and parallel search.  
([http://www.mines.edu/Academic/courses/math\\_cs/mac370/](http://www.mines.edu/Academic/courses/math_cs/mac370/))
- HECIWG FSIO 2007 Activities
  - Mid year reviews for 9 HECURA FSIO projects completed where guidance was provided on the research being done.
  - HECIWG FSIO 2007 Workshop held
  - HECIWG FSIO 2007 Workshop document is being worked on
  - Web site created and populated collecting HECIWG FSIO information and links to each of the projects and the relevant institutions

### Tech Transfer

- Release of tracing mechanism, this application captures system and library calls using strace, collects timing information to detect skew and node drift, and summarizes
- Initial clearing of first file system statistics (fsstats), clearing the way for future large set of fsstats releases
- Released initial FSSTATS data from scratch file systems on open supercomputer at LANL

## **External Presentations**

- Data release update, including planned fsstats data and trace availability, talk given at the HECIWG Workshop in Arlington, VA
- Educational Outreach:
  - One hour lecture on HPC Storage, File Systems, and I/O past, present, and future was given to Colorado School of Mines parallel programming class
  - One hour lecture on HPC Storage, File Systems, and I/O past, present, and future was given to Colorado School of Mines faculty

## LANL SOW Tasks Status

Task Description	% of scope	% complete 1st half FY07	% complete 2nd half FY07	Milestone
1-2 LANL, SNL Participate in HPC I/O and file storage systems curriculum development at participating institute universities including courses and parenthetical degrees.	10	70	100	Conduct guest lectures/seminar s
1-5 ALL Develop & host I/O and file storage workshop for science application developers and users; and for I/O and file storage researchers	8	80	100	SC06 Apps/IO wor
1-5 LANL Sponsor SciDAC and HEC/URA/NSF I/O and file storage R&D showcase as an extension to the HEC/IWG I/O FSIO workshop to showcase the R&D	24	40	100	Plan and hold HEC FSIO2007 with HE status
1-5 LANL, PNNL, SNL Assist in the validation of emerging HPC storage related standards and API's such as pNFS, iSCSI Enhanced RDMA (iSER), and active storage, and enhanced POSIX I/O.	10	60	100	Get POSIX HECE and 1 out, Get pNF into POSIX
1-2 LANL, SNL Provided parallel I/O traces of unclassified parallel applications.	14	50	75	Important app IO tr public
1-2 LANL, SNL Provide parallel I/O traces of synthetic parallel benchmarks as well as source for the benchmarks to enable base lining for parallel I/O trace analysis and replay research.	14	40	100	Important synthetic sets public
1-2 LANL, PNNL, NERSC Collect up to a decade of supercomputer, high hp networking, and I/O and storage system reliability data, machine/environment config info, mtti/mtr) and failure cause data.	10	40	100	Collect/release reliability/usage/ev for systems
1-2 LANL, PNNL Collect up to a decade of supercomputer, hp networking, and I/O and file storage system usage data, including job length, size, processor usage and other usage profile data.	10	40	75	Collect/release reliability/usage/ev for disks (one syste