How Open Source Supports the Largest Computers on the Planet

Best Practices for HPC Software Developers

July 18, 2018

Ian Lee
Lawrence Livermore National Laboratory
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CDC 3600</td>
<td>CDC 7600</td>
<td>CRAY 1</td>
<td>ASCI Blue-Pacific</td>
<td>BlueGene</td>
<td>Petascale and exascale</td>
</tr>
<tr>
<td>Pioneering</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>computing</td>
</tr>
<tr>
<td>simulations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>of particle</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tracking</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ozone mixing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>models</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dynamics in</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>three dimensions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Global climate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>modeling</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unprecedented</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dislocation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dynamics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>simulations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Helping the</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>medical</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>community</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>plan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>radiation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breakthrough</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>visualizations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>of mixing fluids</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Detailed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>predictions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>of ecosystems</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Top500.org

- 3 out of 16 #1 systems over last 20 years

ASCI White
Nov 2000 – Nov 2001

BlueGene/L
Nov 2004 – Nov 2007

Sequoia
June 2012

https://www.top500.org/resources/top-systems/
Sierra
ZFS is an open source filesystem and volume manager designed to address the limitations of existing storage solutions

- 2011: Available for Linux
- Ten LLNL filesystems, totaling ~ 100PB
- Ships in Ubuntu 16.04
Contributing to ZFS on Linux

ZFS

GitHub repository: [zfs CONTRIBUTING.md](https://github.com/zfscontributing.md)

Branch: master

Commits:
- [commitcheck](https://github.com/zfscontributing.md/commit/commitcheck): Multiple OpenZFS ports in commit

Contributors:
- dinatale2
- [commitcheck](https://github.com/zfscontributing.md/commit/commitcheck)

File information:
- 291 lines (230 sloc) 11.6 KB
First of all, thank you for taking the time to contribute!

By using the following guidelines, you can help us make ZFS on Linux even better.

Table Of Contents

What should I know before I get started?

- Get ZFS
- Debug ZFS
- Where can I ask for help?

How Can I Contribute?

- Reporting Bugs
- Suggesting Enhancements
- Pull Requests
- Testing

Style Guides

- Coding Conventions
- Commit Message Formats
  - New Changes
  - OpenZFS Patch Ports
  - Coverity Defect Fixes
  - Signed Off By
The latest stable and development versions of this port can be downloaded from the official ZFS on Linux site located at:

http://zfsonlinux.org/

This ZFS on Linux port was produced at the Lawrence Livermore National Laboratory (LLNL) under Contract No. DE-AC52-07NA27344 (Contract 44) between the U.S. Department of Energy (DOE) and Lawrence Livermore National Security, LLC (LLNS) for the operation of LLNL. It has been approved for release under LLNL-CODE-403049.

Unless otherwise noted, all files in this distribution are released under the Common Development and Distribution License (CDDL).

Exceptions are noted within the associated source files. A few notable exceptions and their respective licenses include:

Skein Checksum Implementation: module/icmp/algsskein/THIRDPARTYLICENSE
AES Implementation: module/icmp/asm-x86_64/aes/THIRDPARTYLICENSE.gladman
AES Implementation: module/icmp/asm-x86_64/aes/THIRDPARTYLICENSE.openssl
Parallel solvers for sparse linear systems featuring multigrid methods. This repository houses releases and test releases. Pull requests can still be addressed. LLNL users should use the main repository on MyBitbucket. [https://www.llnl.gov/casc/hypre/](https://www.llnl.gov/casc/hypre/)
## 432 Software Repositories

### Filter Repositories

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Platform(s)</th>
<th>Repository</th>
<th>Stars</th>
<th>Pulls</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>zfs</td>
<td>ZFS on Linux - the official OpenZFS implementation for Linux.</td>
<td>zfs</td>
<td>3316</td>
<td>699</td>
<td></td>
<td></td>
</tr>
<tr>
<td>spack</td>
<td>A flexible package manager that supports multiple versions, configurations, platforms, and compilers.</td>
<td>spack</td>
<td>571</td>
<td>417</td>
<td></td>
<td></td>
</tr>
<tr>
<td>zfs-auto-snapshot</td>
<td>ZFS Automatic Snapshot Service for Linux.</td>
<td>zfs</td>
<td>373</td>
<td>104</td>
<td></td>
<td></td>
</tr>
<tr>
<td>spl</td>
<td>A shim layer which adds the core interfaces required for OpenZFS.</td>
<td>spl</td>
<td>284</td>
<td>168</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pkg-zfs</td>
<td>Native ZFS packaging for Debian and Ubuntu.</td>
<td>pkg-zfs</td>
<td>282</td>
<td>34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mfem</td>
<td>Lightweight, general, scalable C++ library for finite element methods.</td>
<td>mfem</td>
<td>193</td>
<td>79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rose</td>
<td>A release version of ROSE that passes all regression tests.</td>
<td>rose</td>
<td>167</td>
<td>59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pdsh</td>
<td>A high performance, parallel remote shell utility.</td>
<td>pdsh</td>
<td>162</td>
<td>27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dcp</td>
<td>dcp is a distributed file copy program that automatically distributes and dynamically balances work equally across nodes in a large distributed system without centralized state.</td>
<td>dcp</td>
<td>154</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>diod</td>
<td>Distributed I/O Daemon - a SIF file server.</td>
<td>diod</td>
<td>145</td>
<td>29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>zfp</td>
<td>Library for compressed numerical arrays that support high throughput read and write random.</td>
<td>zfp</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>magpie</td>
<td>Magpie contains a number of scripts for running Big Data software in HPC environments.</td>
<td>magpie</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
LLNL Open Source Presence

Repo Creation History

- Git Released [2005 - 04 - 07]
- GitHub Founded [2008 - 02 - 06]
- LLNL Soft Portal [2015 - 11 - 24]

- First Commit
- Added to GitHub

https://software.llnl.gov/explore
LLNL Open Source Engagement

179
LLNL Org. Members
- 53% Contributing Externally
- 47% No External Repos

432
LLNL Repositories
- 75% External Contributors
- 25% Only LLNL Contributors

https://software.llnl.gov/explore
LLNL Open Source Activities

https://software.llnl.gov/explore
“Our large collection of software is a precious Laboratory asset, one that benefits both Lawrence Livermore, and in many cases, the public at large.”

- Bruce Hendrickson
  Associate Director, Computation
AN UP-CLOSE VIEW OF THE SOFTWARE THAT UNDERPINS THE EXASCALE COMPUTING PROJECT
08/30/17

When exascale systems become a reality, the Exascale Computing Project (ECP) will bring to those systems both existing high-performance computing (HPC) software and promising emerging research. Accordingly, one of the objectives of the ECP is to create a production-quality base—a software stack—to support the scientific applications that will run on these systems.

Scientists developing applications for exascale systems depend on an intricate set of software that makes the computing system usable and the job of the application developer easier. The broad services this software provides are often collectively referred to as the software stack.

Virtually all of the ECP software stack developed by the US Department of Energy (DOE) is composed of open-source code, which makes the software broadly available and appealing to other programmers to contribute when the base capabilities are established. Software provided by the platform vendors, however, often consists of a combination of open-source and proprietary components.
Federal Source Code Policy


- “Agencies shall make custom-developed code available for Government-wide reuse and make their code inventories discoverable at https://www.code.gov (“Code.gov”) [...]”

- “[...] establishes a pilot program that requires agencies, when commissioning new custom software, to release at least 20 percent of new custom-developed code as Open Source Software (OSS) [...]”

GitHub and Government

Who's using GitHub  Peer Group  Accessibility  Contact

How agencies build software

Collaborate on code, data, policy, or procurement within your agency or with the public.

Collaborate on code

Code collaboration and review are built into the development process with GitHub. Share work, discuss changes, and get feedback in one place to write quality code.

On your servers or in the cloud

Run GitHub Enterprise on your servers as a virtual appliance, on AWS GovCloud or Azure, or let us host the code for you on GitHub.com, whatever your security requires.

https://government.github.com
US Government Organizations on GitHub

https://government.github.com/community/
Thank You!

ian@llnl.gov

@IanLee1521 // @LLNL_OpenSource

https://speakerdeck.com/IanLee1521
TOSS – Tri-Lab Operating System Software

- Built on Red Hat Enterprise Linux
  - Not an HPC distribution

- Adds LLNL developed additions and patches to support HPC
  - Low Latency Interconnect: Infiniband
  - Parallel File System: Lustre
  - Resource Manager: SLURM

- Work closely with open communities

TOSS is a software stack for HPC – large, interconnected clusters!
- Began as simple resource manager
  - Now scalable to 1.6M+ cores (sequoia)
- Launch and manage parallel jobs
  - Large, parallel jobs, often MPI
- Queuing and scheduling of jobs
  - Much more work than resources

http://slurm.schedmd.com
- Family of projects used to build site-customized resource management systems

- flux-core
  - Implements the communication layer and lowest level services and interfaces

- flux-sched
  - Consists of an engine that handles all the functionality common to scheduling

- capacitor
  - A bulk execution manager using flux-core, handles running and monitoring 1000’s of jobs

http://flux-framework.github.io
- Handles combinatorial explosion of ABI-incompatible packages
- All versions coexist, binaries work regardless of user’s environment
- Familiar syntax, reminiscent of brew, yum, etc

$ spack install mpileaks unconstrained
$ spack install mpileaks@3.3 @ custom version
$ spack install mpileaks@3.3 %gcc@4.7.3 % custom compiler
$ spack install mpileaks@3.3 %gcc@4.7.3 +threads +/- build option
$ spack install mpileaks@3.3 os=SuSE11 os=<frontend OS>
$ spack install mpileaks@3.3 os=CNL10 os=<backend OS>
$ spack install mpileaks@3.3 os=CNL10 target=haswell target=<cpu target>

https://spack.io
- Manages the first-ever decentralized database for handling climate science data
- Multiple petabytes of data at dozens of federated sites worldwide
- International collaboration for the software that powers most global climate change research
VisIt

- Originally developed to visualize and analyze the results of terascale simulations
- Interactive, scalable, visualization, animation and analysis tool
- Powerful, easy to use GUI
- Distributed and parallel architecture allows handling extremely large data sets interactively

https://visit.llnl.gov
Department of Energy

877 Projects

3D Torus Routing Engine Module for OFA OpenSM v 1.0
This OpenFabrics Alliance (OFA) OpenSM routing engine module provides credit-loop-free routing while supporting two quality of service (QoS) levels for an InfiniBand fabric with a 3D torus topology. L...

Open Source

A Distributed Parallel Visualization and Analysis Tool
Visit is an interactive parallel visualization and graphical analysis tool for viewing scientific data on UNIX and PC platforms. Users can quickly generate visualizations from their data, animate them...

Open Source

A Grand Canonical Monte Carlo simulation program for computing ion distributions around biomolecules in hard sphere solvents
The GIBB software program is a Grand Canonical Monte Carlo (GCMC) simulation program (written in C++) that can be used for computing the excess chemical potential of ions and the mean activity coeff...

Open Source

A Latency Tolerant Runtime System for Mass Market Computer Systems
Grappa is a latency tolerant runtime system for mass market computer systems comprised of multiple nodes having x86 multicore processors and infiniband interconnect. Grappa is designed to enable in-me...

Open Source
Public US Government GitHub Data Scrape

- 252 US Government Orgs
  - U.S. Federal (137)
  - U.S. Military and Intelligence (12)
  - U.S. Research Labs (103)

- 8716 Open Source Repositories