Enabling the optimization of open-source biological computational tools with scripting languages

Stefan Popa [1]

[1] Intel Inc. / University POLITEHNICA Bucharest Romania. Email: stefan.popa@acm.org / stefan.a.popa@intel.com

Project Website: http://languagesperformance.intel.com
Source Code: https://github.com/php/php-src, https://hg.python.org/cpython/
License: PHP License Agreement (GPL-compatible--see https://github.com/php/php-src/blob/master/LICENSE), Python License Agreement (GPL-compatible--see https://hg.python.org/cpython/file/tip/LICENSE)

Main Text of Abstract

In this presentation it is described software optimizations for open-source biological computational tools like BioPHP or BioPython through Intel Architecture software optimizations of the open-source scripting languages used, like PHP or Python.

Today's compilers are providing various optimization techniques, such as loop unrolling, code layout optimizations, inlining and many other. All these techniques have a common point, they make their decisions based on the analysis made to the application source code. Even though this approach provides good results for general case, compilers can optimize furthermore an application if we know the workload, the input data or any other information about runtime by applying sampling or instrumentation Profile Guided Optimizations (PGO).

In Bio-related Cloud Computing applications, knowing in advance the workload or how the application must behave at runtime, can lead to a great degree of improvement and better end-user experience. In this paper we talk about how can we apply PGO on ones of the most used open-source biological computational tools, BioPHP and BioPython. The presentation will detail the way we used PGO to optimize PHP and Python interpreters (https://github.com/php/php-src/commit/7dac4d449f72d7eb029aa1a8ee87aaf38e17e1c5, https://hg.python.org/cpython/rev/7fcff838d09e).

```
Results for project PHP master, build date 2016-02-26 02:09:41+02:00
commit:         b378c7c
previous commit:        a1c9bd8
revision date:  2016-02-25 16:07:38+08:00
environment:    Haswell-EP
        cpu:            Intel(R) Xeon(R) CPU E5-2699 v3 @ 2.30GHz 2x18 cores, stepping 2, LLC 45 MB
        mem:            128 GB
        os:             CentOS 7.1
        kernel: Linux 3.10.0-229.4.2.el7.x86_64

Baseline results were generated using release php-7.0.0beta3, with hash 1674bd9
from 2015-08-05 04:56:40+00:00

------------------------------------------------------------------------------------------
                        benchmark   relative   change since   change since  current rev run
                                    std_dev*       last run       baseline        with PGO
------------------------------------------------------------------------------------------
:-|   biophp_minitools cgi -T1000       0.06%         -0.34%          1.32%            4.86%
------------------------------------------------------------------------------------------
* Relative Standard Deviation (Standard Deviation/Average)
```

```
Results for project Python default, build date 2016-02-26 04:42:07+02:00
commit:         ed30eac
previous commit:        2c4448b
revision date:  2016-02-26 00:42:33+00:00
environment:    Haswell-EP
        cpu:            Intel(R) Xeon(R) CPU E5-2699 v3 @ 2.30GHz 2x18 cores, stepping 2, LLC 45 MB
        mem:            128 GB
        os:             CentOS 7.1
        kernel: Linux 3.10.0-229.4.2.el7.x86_64

Baseline results were generated using release v3.4.3, with hash b4cbecb from
2015-02-25 12:15:33+00:00

------------------------------------------------------------------------------------------
                        benchmark   relative   change since   change since  current rev run
                                    std_dev*       last run       baseline        with PGO
------------------------------------------------------------------------------------------
:-|           biopython       0.35%         -0.66%          0.35%            4.04%
------------------------------------------------------------------------------------------
* Relative Standard Deviation (Standard Deviation/Average)
```

Profile Guided Optimizations brings more than 4% performance increase for open-source biological computational tools like BioPHP and BioPython. The initiative to have optimized open-source scripting languages is on-going. The large and very active communities behind each of these open-source scripting languages like PHP and Python generate daily commits, bug fixes, new language features, and performance optimizations. The overall performance status of PHP and Python language can be tracked daily using https://01.org/lp and http://languagesperformance.intel.com.