Using the Common Workflow Language (CWL) to run portable workflows with Arvados and Toil

Peter Amstutz[1], Brian O'Connor[2], Benedict Paten[2], Alexander Wait Zaranek[1]

[1] Curoverse, Boston. peter.amstutz@curoverse.com, awz@curoverse.com
[2] University of California, Santa Cruz. broconno@ucsc.edu, benedict@soe.ucsc.edu

With special thanks to all contributors to Common Workflow Language, Toil, and Arvados.

Project Websites:
> http://commonwl.org, http://arvados.org, https://toil.readthedocs.org/en/latest/

Source Code:
> https://github.com/common-workflow-language/common-workflow-language,
> https://github.com/curoverse/arvados, https://github.com/BD2KGenomics/toil

License:
https://github.com/common-workflow-language/common-workflow-language/blob/master/LICENSE.txt
https://github.com/curoverse/arvados/blob/master/COPYING
https://github.com/BD2KGenomics/toil/blob/master/LICENSE.txt

The Common Workflow Language (CWL) is a community effort that started at the BOSC Codefest 2014 to create a common specification for describing analysis tools and workflows that is portable and scalable across a variety of hardware and software platforms. Arvados is a cluster and cloud compute platform developed by Curoverse consisting of a content-addressed storage system "Keep", a compute management system "Crunch". Arvados can run in a variety of cloud and cluster computing configurations. Toil is a workflow management engine developed by the University of California Santa Cruz Genomics Institute. Toil also supports a number of cluster and cloud computing environments.

Some of the benefits of a common, community-developed language for computational workflows include the ability of scientists to collaborate and leverage the work of others, enhanced development of tooling around a common format, code that can "travel to the data" avoiding the need for large downloads and addressing legal issues around data migration, and ease of benchmarking methods across tools, techniques and platforms.

This talk will briefly introduce CWL and provide a project and community update since BOSC 2015. The talk will discuss the implementation of CWL on the Arvados platform and in the Toil workflow engine and how compatibility with CWL was achieved differences in the design of each system. The talk will discuss how to run workflows written in CWL on each system. The talk will then present results from running a representative genomics workflow written using CWL on both Arvados and Toil, using different cloud providers, with no workflow customization or porting required.