Award 1440677 - Annual Project Report

Supplementary Material

Metrics:

- *5 new users or applications using RADICAL-Cybertools (on XSEDE, Campus resources or OSG).*

Thanks to working with XSEDE ECSS Workflow Community Applications Team [1], we were able to get RADICAL-Cybertools adopted as one of the four 'workflow tools" supported on XSEDE [2]. In conjunction, thanks to an active "CampusChampions" mailing, list, where RADICAL-Cybertools has been mentioned several times over the past year, there are many new users that have "silently" started using RADICAL-Cybertools. It is difficult to track all new users and their use cases; if they have basic requirements, they just get on with the job without needing further assistance. In advanced use cases or when there is a problem, there is a need to interact and thus we are able to monitor use-cases. Thus, in contrast to original metrics, we now primarily rely on depsy (an NSF ACI funded tool) to track impact. Links to radical-cybertools entry on depsy are provided at Ref [3, 4], **according to which one of our components (saga-python) is in the 97th percentile of impact.** Downloads for both components are ~4,000.

Nonetheless, here are five new users in the past year that we have supported, and links to their "use-cases".

- Oliver Beckstein & David Dotson (ASU) https://github.com/radical-cybertools/radical.pilot/issues/773
- Eric Irrgang (Michigan) https://github.com/radical-cybertools/radical.pilot/commit/99deda1a7d861228fa9ea6410cdb5c1f75a77d7d
- Leela Dodd (Yale) https://github.com/radical-cybertools/radical.pilot/issues/869
- Joohyun Kim (LSU): Pipelines of bioinformatics applications on LSU's SuperMIC and other XSEDE resources.
- Tom Cheatham (Utah): Ensemble runs on Blue Waters (work in progress).

[1] https://www.xsede.org/ecss-workflows
[2] https://sites.google.com/site/xsedeworkflows/
[3] http://depsy.org/package/python/saga-python
[4] http://depsy.org/package/python/radical.pilot

- *total of 5M SUs consumed on XSEDE via RADICAL Cybertools*

Informally, tracking the usage of the above five users and other previous (e.g., Darrin York (Rutgers): Replica-Exchange simulations on XSEDE resources) and our own usage on XSEDE, our estimate is that we have supported more than 5M. We are not providing "statistics" but just an estimate, Our customized scripts to analyze the usage of RADICAL-Cybertools was erroneous (as in it would provide a lower bound, not an upper bound!) and we are disbanding it favour of a more "reliable" measure. In order for data on usage of a tool to be reliable must be provided by an independent party and audited. Thus we are working XSEDEs XDMoD to integrate the usage of RADICAL-Cybertools into XDMoD. This is work in progress.

Community Outreach:

- *RADICAL Cybertools tutorial at XSEDE and OSG,*
    - https://github.com/radical-cybertools/supercomputing2015-tutorial/wiki
    - http://www.hpcwire.com/2015/11/13/hadoop-and-spark-get-radical-at-sc15/
    - Link to the tutorial material: http://nbviewer.jupyter.org/github/radical-cybertools/supercomputing2015-tutorial/tree/master/

- *Tutorials at NSF's PIRE (Open Science Data Cloud)*
  This tutorial was organized and prepared after the award had been made but just before the start date of the project:

    - Tutorial at NSF PIRE Workshop 2014, Amsterdam
      http://pire.opensciencedatacloud.org/radical-cybertools-saga-and-pilot/

- *EarthCube workshops and seminars*

    - We presented a RADICAL-Cybertools approach to data-intensive computing at the following EarthCube workshop at San Diego:
        - http://earthcube.org/workspace/architecture-wg/earthcube-architecture-workshop
        - https://drive.google.com/folderview?id=0BxsfXNGQGIv8fjJwWC02TmhxZXZleG0weE9OYzl5SGQxUGpLc0ltNk9EajM5aF9Mend0dDA&usp=sharing&tid=0BxsfXNGQGIv8fmtZdXYtbU1CWnMtbEJRYkJpOXpUQmtxV1RhZndVYTVTMTBCVFV0aThfcWc

    - We also presented an overview at the EarthCube Technology & Architecture Committee on Dec 7th, 2015:
        - https://docs.google.com/document/d/1tMLYVwAX9-bx29zh1S_PHwgXbtpPDuDG6zvWuukV5-8/edit

- *Presentations/tutorials at domain conferences, such as American Chemical Society and its publications,*
    - We did not present at the ACS, but rather at the European equivalent, namely at a CECAM workshop. We organized a 2 day workshop on "High Performance Distributed Computing tools for Biomolecular Simulations" (and contributed to another two day workshop on Advanced Sampling and Long Timescale Molecular Dynamics):
        - https://github.com/radical-cybertools/tutorials/wiki/CECAM-2015
        - http://www.cecam.org/workshop-0-1214.html

    - Furthermore, we also presented at another "domain" conference, the European Geophysical Union meeting in Vienna, Austria. This tutorial was organized and prepared after the award had been made but just before the start date of the project:
        - *Tutorial at EU funded DRIHM Workshop 2014, Madrid: http://www.drihm.eu/index.php/documents/drihm-summer-school-2014*

- *Work with other SI2 awardees (across all three levels: elements, integration and pre-institute conceptualization levels):*
    - We worked with (i) Swift (Wilde), (ii) Tom Cheatham (RAPID) and (iii) Cecilia Clementi. Different projects were at different levels spanning research, prototype and development/production.

- *Work with Futuregrid to enable uptake of RADICAL Cybertools as part of suite of research software available to users of Futuregrid.*
    - Futuregrid has come to an end; a successor project (of sorts) is Chameleon Cloud https://www.chameleoncloud.org/. Here is a link to the set of tools that are publicly available for use of radical-cybertools as a "research harness" on ChameleonCloud.
        - *https://github.com/radical-cybertools/chameleon-utils*
    - Here is a link to our project on Chameleon Cloud.

RADICAL-Cybertools: Enabling Scalable Compute and Data Intensive Science | CH-816985

Abstract: To support science and engineering applications that are the basis of many societal and intellectual challenges in the 21st century, there is a need for comprehensive, balanced, and flexible distributed cyberinfrastructure (DCI). The process of designing and deploying such large-scale DCI, however, presents a critical and challenging research agenda. One specific challenge is to produce tools that provide a step change in the sophistication of problems that can be investigated using DCI, while being extensible, easy to deploy and use, as well as being compatible with a variety of other established tools. RADICAL-Cybertools will meet these requirements by providing an abstractions-based suite of well-defined capabilities that are architected to support scalable, interoperable and sustainable science on a range of high-performance and distributed computing infrastructure. RADICAL Cybertools builds upon important theoretical advances, production-software-development best practices, and carefully-analyzed usage and programming models. RADICAL Cybertools is posed to play a role in grand-challenge problems, ranging from personalized medicine and health to understanding long-term global and regional climate. All software developed through the project will be open source and will be licensed under the MIT License (MIT). Version control on the SVN repository will be accessible via http://radical.rutgers.edu.