

XSEDE 2014 Tutorial Proposal

XSEDE Application Development with RADICAL Cybertools

XSEDE Application Development with RADICAL Cybertools (½ Day)

This tutorial provides an overview of RADICAL Cybertools (<http://radical-cybertools.github.io/>), a suite of tools that enable Scalable, Interoperable and Sustainable Science applications on XSEDE. It consists of a short theoretical part in which we cover general topics related to application development on XSEDE, and a longer practical part with hands-on training and examples of using RADICAL Cybertools. Using the RADICAL.Cybertools Python libraries, we will guide participants through the development of a skeletal prototype of quintessential distributed applications, as well the most common usage patterns of RADICAL Cybertools.

Instructors

- Shantenu Jha (Rutgers University, shantenu.jha@rutgers.edu)
 - Ole Weidner (Rutgers University, ole.weidner@rutgers.edu)
 - Andre Merzky (Rutgers University, andre.merzky@rutgers.edu)
 - Mark Santcroos* (Rutgers University, mark.santcroos@rutgers.edu)
 - Vivek Balasubramanian* (Rutgers University, vb224@scarletmail.rutgers.edu)
- * Students

Target Audience

Ever wondered how you can write a job-submission script once and execute on any XSEDE machine? Ever wondered how to run thousands of jobs all over XSEDE? Ever tried to move your data “automagically” where your jobs get submitted? If you’ve ever wondered about any of these and more and are interested in learning about tools that provide you with both explicit capabilities as well as the building blocks to develop customized capabilities, then this tutorial will interest you. This tutorial will interest anyone who is looking to develop scalable and interoperable approaches to utilizing the set of heterogeneous XSEDE resources.

This tutorial is targeted towards both the end-scientist and tools and service developers. It is most appropriate for intermediate and advanced users of XSEDE, that have a basic knowledge of the Python programming language as well as some UNIX skills (ssh, sftp, etc.) are required to participate in the tutorial’s hands-on session.

Requirements

- If participants want to use their own laptops:
 - Linux or OS X
 - Python >= 2.5
- All participants need:
 - XSEDE accounts, access to at least one, better two HPC clusters + password-less SSH-key setup
- Ideally we have an on-site VM (provided by XSEDE) on which participants can log-in and run the tutorial code so we can keep the load on the login nodes as low as possible. If that is not possible, we will try to provide an off-site VM, e.g., on EC2 for the tutorial.

XSEDE 2014 Tutorial Proposal

XSEDE Application Development with RADICAL Cybertools

Sample Tutorial Exercises and Slides

Slides from tutorials at XSEDE'12 and XSEDE'13 can be found at Ref [2] and [3] respectively. We will be significantly improving these slides.

RADICAL-Cybertools[1] is the next incarnation of the tools and programming libraries that provide the core for a consistently growing number of distributed, data-intensive scientific applications that are responsible for an annual consumption of more than 14M SUs on XSEDE resources [Figure 1].

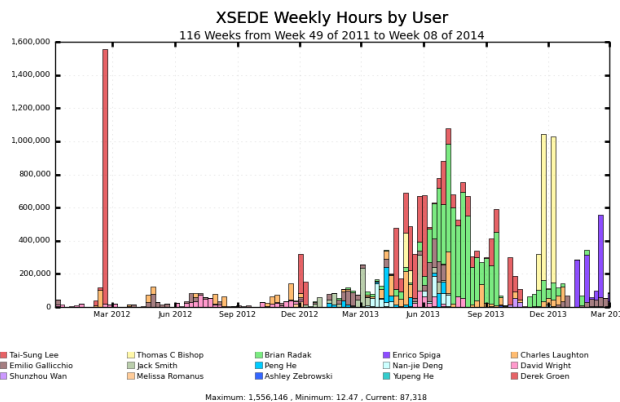


Figure 1: XSEDE SU consumption between 2012 and 2014 generated by applications based on RADICAL-Cybertools [4].

Much has happened since our previous XSEDE Tutorials in 2013 “*Introduction to BigJob - A SAGA-Based Interoperable, Extensible and Scalable Pilot-Job for XSEDE*” [2] and in 2013 “*An Introduction to BigJob and BigData Using Pilots to Manage Computation and Data on XSEDE*” [3]. BigJob has undergone major revisions and a rigorous redesign and has been integrated into the RADICAL-Cybertools stack as *RADICAL-Pilot*. Similarly, SAGA-Python has been hardened and extended to support additional HPC schedulers, new and upcoming HPC resource as well as IaaS-style resources. It is now part of the RADICAL-Cybertools stack, named RADICAL-SAGA.

Besides the many improvements and new features in the software packages and our push towards a homogeneous sustainable and more user-friendly software ecosystem “RADICAL-Cybertools”, we have worked with many application groups and end-users / scientists to build next generation distributed applications on XSEDE. Applications have ranged from simple, loosely coupled tasks (e.g., [4]), to more complex Biochemistry applications (e.g., [5]) to web-based science gateways [6]. In the process, we have not only learned to understand (and improve!) our own tools better, but also how distributed applications should be developed in a loosely federated HPC cluster environment like XSEDE. One of the main goals of our tutorial is to share the practical (and often painful) lessons learned with the XSEDE community.

The RADICAL-Cybertools half-day tutorial will bring together RADICAL team members, XSEDE staff and XSEDE computational scientists and users to discuss how develop Python-based distributed science applications, tools and services that are capable of scaling computational and data-intensive workloads within and beyond the boundaries of XSEDE. In the first part of the tutorial we will share our experience building large-scale applications on XSEDE with RADICAL-Cybertools, discuss important considerations, patterns and common pitfalls. In the second half of the tutorial, we will interactively develop skeletal

XSEDE 2014 Tutorial Proposal

XSEDE Application Development with RADICAL Cybertools

prototype of distributed applications that incorporates many of the topics introduced in the first half. This tutorial is targeted towards both, end-users who are looking for a more 'programmatic' approach to run their computational workload efficiently across XSEDE, and application and tool developers who are looking to implement more complex application, tools and services to support the computational demands of their respective communities. The goal is to send the participants home with a better understanding how to get started with the development of sustainable and scalable distributed science applications on XSEDE. The prototype we develop during the tutorial can serve the participants as foundation and starting point for future developments.

Draft Tutorial Structure

1. Presentation: Developing Scalable and Interoperable Distributed Applications for XSEDE (25% of total time)
2. Presentation: A (Very) Brief Introduction to RADICAL-Cybertools (10% of total time)
3. Hands-On Session: Developing an Application Prototype with RADICAL-Pilot (65% of total time)

References

[1] RADICAL Cybertools <http://radical-cybertools.github.io/>

[2] XSEDE12: Tutorial: Slides and Program
<https://github.com/saga-project/BigJob/wiki/XSEDE-2012-Tutorial>

[3] XSEDE13 Tutorial: Slides and Program
<https://github.com/saga-project/tutorials/wiki/XSEDE13>
<https://www.xsede.org/web/xsede12/tutorials#09>

[4] Jack A. Smith, Melissa Romanus, Pradeep Kumar Mantha, Yaakoub El Khamra, Thomas C. Bishop, and Shantenu Jha. 2013. Scalable online comparative genomics of mononucleosomes: a BigJob. In *Proceedings of the Conference on Extreme Science and Engineering Discovery Environment: Gateway to Discovery* (XSEDE '13). ACM, New York, NY, USA, Article 23, 8 pages. DOI=10.1145/2484762.2484819 <http://doi.acm.org/10.1145/2484762.2484819>

[5] David Wright, Benjamin Hall, Owain Kenway, Shantenu Jha and Peter Coveney Computing clinically relevant binding free energies of HIV-1 protease inhibitors. *Journal of Chemical Theory and Computation*, (2014). <http://dx.doi.org/10.1021/ct4007037>.

[6] DARE Gateways: <http://dare.cct.lsu.edu/>.

[7]
http://glidein.unl.edu/osgtg/dn_hours_bar/dn_hours_bar?endtime=2014-02-28+23%3A59%3A59&starttime=2011-12-08+00%3A00%3A00&user=Ashley%20Zebrowski%7CMelissa%20Romanus%7CThomas%20C%20Bishop%7CCharles%20Laughton%7CEmilio%20Gallicchio%7CJack%20Smith%7CDavid%20Wright%7CBrian%20Radak%7CTai-Sung%20Lee%7CEnrico%20Spiga%7CBruce%20Boghosian%7CShunzhou%20Wan%7CDerek%20Groen%7CPeng%20He%7CNan-Jie%20Deng%7CWei%20Dai&title=XSEDE%20Weekly%20Hours%20by%20User&span=604800&facility=*&probe=*&resource-type=5EBatch%24%7CBatchPilot&vo=*&role=*&exclude-facility=NONE%7CGeneric%7CObsolete&exclude-user=NONE&includeSuccess=true&exclude-vo=unknown%7Cother&includeFailed=true&exclude-role=NONE