Tutorial: Open Source Scientific Computing

Amit Saha

Red Hat, Inc, 193 North Quay, Brisbane, Australia. *asaha@redhat.com*

In this tutorial, the audience will learn how they can adopt Open Source tools and methodologies in their research.

1 Why Open Source?

The Open Source ecosystem has excellent tools for the scientific community. Over the years, thanks to contributions by the various communities, these tools have matured and have been adopted for scientific research in various fields. They are now more than capable alternatives to expensive proprietary software. Besides their economic advantages, they also bring with them the freedom to be used, tinkered with and redistributed with liberal licenses.

Adopting Open Source methodologies in research enables sharing of tools and data with an appropriate license. This serves the dual purpose of open-ness of the research and also protects the intellectual rights from being abused.

The next section outlines the various software tools that will be addressed in this tutorial.

2 Open Source Tools

Operating System: The *Fedora Scientific Spin* is a freely available Linux based operating system, exclusively targeting computer users conducting scientific research. It has a plethora of tools already installed which makes it easy to get started with, out of the box. Other choices include *Scientific Linux*.

Numerical Computing tools: GNU Octave, *Scilab* and *Sage* aim to provide an all-in-one scientific software package with support for various numerical and scientific computing requirements.

C/C++ Libraries: The *GNU Scientific Library* is a library for C/C++ implementing an extensive list of mathematical routines.

The *GNU multi precision* library adds support for multi-precision arithmetic for C/C++ programs.

Python for Scientific Computing: The *Python* programming language combines rapid prototyping features with an easy syntax. Third party libraries such as SciPy with a gentle learning curve makes Python a first class choice for scientific computing for beginners and veterans alike.

R for Data analysis: *R* is a programming language and environment for statistical computing. It provides a wide variety of modeling and data analysis techniques and has become a de-facto standard for analyzing data and making statistical inferences based on them.

Parallel and Distributed Computing: Certain areas of scientific research demand the availability of high performance computing power. Traditional open source solutions such as *Open MPI* and modern cloud computing based solutions make it easy to harness the collective power of a lab of low cost computers. These tools are supported by libraries for popular programming languages to make them easy to use.

Drawing and Plotting: Plotting tools such as *Gnu plot* and drawing tools such as *Xfig* and *Dia* allow creation of high quality graphs and figures and exporting them to *encapsulated postscript* (eps) to be used with the LaTex typesetting system.

3 Open Source Methodologies

Open Source Licensing: Computer programs – either developed from scratch or built upon someone else's work form the cornerstone of computational scientific research. Any researcher or interested reader can easily verify and better understand a research article's results simply by running those programs. Hence, it is imperative that such programs are made publicly available. Sharing programs using a suitable Open Source license protects the interests and rights of the creator of the program, while also giving anyone else the right to read, enhance and improve those programs. This goes a long way in increasing transparency of the whole scientific process.

Version Control: Version control enables inter and intra group research collaboration. The benefits of version control are not only limited to computer programs, but any other data – research papers, data sets, figures and others.

4 Conclusion

Scientific computing stands to benefit highly if it adopts Open Source computing tools and methodologies. Ultimately, it will result in a more open scientific research community and hence enable and foster innovation on a larger scale than it is.