Packaging the HEP simulation stack on conda-forge

Lev Pambuk
Odesa National University of Technology

June 2024

Project proposal

One common toolchain used in high energy physics for simulation is: MadGraph5_aMC@NLO [1], PYTHIA8 [2], Delphes [3]. Installing these tools can be challenging at times, especially for new users. The conda-forge [4] packaging infrastructure system and package registry allows for distribution of complex binaries across multiple platforms though the Conda package management ecosystem. As ROOT [5] and the PYTHIA8 library with Python bindings [6] have been successfully packaged and distributed on conda-forge it should be possible to package all the components of the HEP simulation stack and distribute them on conda-forge. Packaging these components will allow them to have the ability to not only be installed as individual tools, but also be installed together in a coherent package environment.

However, the interconnected nature of some of these tools requires that multiple dependencies are first packaged and distributed on conda-forge before the full stack can be. This project will attempt to package as many of the dependencies of the HEP simulation stack on conda-forge as possible starting with LHAPDF [7] and adding Python 3 bindings for the HepMC2 [8] and HepMC3 [9] conda-forge feedstocks. These selected projects provide a good testbed for the challenge of the full stack as they require supporting multiple build toolchain and compiler requirements, while also being reasonably easy to build without a large amount of domain expertise. The authors of these projects have agreed to provide assistance with technical questions on build systems and testing of the packaged artifacts, making them good candidates for initial work. If time allows, the FastJet [10] project will also be packaged. Providing even a partial selection of these tools through conda-forge would already significantly lower the barrier to use for many researchers.

The project will be done under the mentorship of Dr. Matthew Feickert.

Project Deliverables

- Distributions on conda-forge as conda-forge feedstocks:
    - LHAPDF with Python 3 bindings
    - HepMC2 with Python 3 bindings
    - HepMC3 with Python 3 bindings
    - FastJet tool (stretch goal)
- Contributions to the documentation of the upstream projects that indicate that the libraries and command line tools can be installed through conda-forge.
- A project GitHub repository with usage examples of how to install and use the packaged tools.

These deliverables are aimed to speed up creating a coherent package environment of HEP simulation stack that will make the process of installing needed tools much easier.

Timeline

- **Week 1-2**: Refresh knowledge, familiarize myself with software development tooling (like CI/CD, Git, etc) and packaging build systems (especially their dependency management) that will be used for a project. Learn about tools that will be packaged and how conda-forge works (creating package recipes, overall build infrastructure, etc). Create a developer environment with all the required tooling on a local machine.
- **Week 3-4**: Add a conda-forge feedstock for Python 3 bindings for the HepMC2 in collaboration with Chris Burr, the maintainer for the HepMC2 conda-forge feedstock.
- **Week 5-6**: Add a conda-forge feedstock for Python 3 bindings for the HepMC3 in collaboration with Chris Burr, the maintainer for the HepMC3 conda-forge feedstock..
- **Week 7-8**: Package LHAPDF on conda-forge.
- **Week 9-10**: Package FastJet on conda-forge (stretch goal).
- **Week 11**: Working with the maintainers of the packaged tools to help them update their documentation and examples to show the conda-forge distributed tools.
- **Week 12**: Summary report and slides of achieved results.

References

[1]MadGraph5_aMC@NLO overview: https://launchpad.net/mg5amcnlo

[2]PYTHIA8 main page: https://pythia.org/

[3]Delphes repository: https://github.com/delphes/delphes

[4]Conda-forge main page: https://conda-forge.org/

[5]ROOT conda-forge package: https://anaconda.org/conda-forge/root/

[6]PYTHIA8 Python bindings: https://anaconda.org/conda-forge/pythia8/

[7]LHAPDF main page: https://lhapdf.hepforge.org/

[8]HepMC2 repository: https://github.com/conda-forge/hepmc2-feedstock

[9]HepMC3 repository: https://github.com/conda-forge/hepmc3-feedstock

[10]FastJet main page: http://fastjet.fr/