

An interactive and comprehensive working environment for high-energy physics software with Python and jupyter notebooks

Thursday, October 13, 2016 11:15 AM (15 minutes)

Today's analyses for high energy physics experiments involve processing a large amount of data with highly specialized algorithms. The contemporary workflow from recorded data to final results is based on the execution of small scripts - often written in Python or ROOT macros which call complex compiled algorithms in the background - to perform fitting procedures and generate plots. During recent years interactive programming environments, such as jupyter, became popular. Jupyter allows to develop Python-based applications, so-called notebooks, which bundle code, documentation and results, e.g. plots. The advantages over classical script-based approaches is the feature to recompute only parts of the analysis code, which allows for fast and iterative development, and a web-based user frontend, which can be hosted centrally and only requires a browser on the user side.

In our novel approach, Python and jupyter are tightly integrated into the Belle II Analysis Software Framework 2 (basf2), currently being developed for the Belle II experiment in Japan. This allows to develop code in jupyter notebooks for every aspect of the event simulation, reconstruction and analysis chain. Basf2 is based on software modules implemented in C++11 which have Python bindings created with Boost Python and PyROOT. These interactive notebooks can be hosted as a centralized web service via jupyterhub with docker and used by all scientists of the Belle II Collaboration. Because of its generality and encapsulation, the setup can easily be scaled to large installations.

This contribution will describe the technical implementations of the jupyter integration into basf2. The required code is generic enough to be easily applicable to other high energy physics frameworks and even to software from other research domains. The talk presents a full example of a jupyter-based analysis and some notebooks already successfully used in outreach and educational purposes.

Tertiary Keyword (Optional)

Secondary Keyword (Optional)

Analysis tools and techniques

Primary Keyword (Mandatory)

Data processing workflows and frameworks/pipelines

Authors: Mr BRAUN, Nils (KIT); Dr HAUTH, Thomas (KIT)

Co-authors: Dr PULVERMACHER, Christian (KIT); Dr RITTER, Martin (LMU / Cluster Universe)

Presenter: Dr HAUTH, Thomas (KIT)

Session Classification: Track 5: Software Development

Track Classification: Track 5: Software Development