

# bamboo: easy and efficient analysis with python and RDataFrame

*Tuesday 24 November 2020 16:39 (10 minutes)*

The bamboo analysis framework [1] allows to write simple declarative analysis code (it effectively implements a domain-specific language embedded in python), and runs it efficiently using RDataFrame (RDF) - or viewed differently: it introduces a set of tools to efficiently generate large RDF computation graphs from a minimal amount of user code (in python), e.g. a simple way to specify selections and outputs, automatically filling a set of histograms with different systematic variations of some input variables.

It is currently being used for several analyses on the full CMS Run2 dataset, and thus provides an example of a very analysis description language-like approach that is compatible with the practical needs of modern HEP data analysis (different types of corrections, machine learning inference, user-provided extensions, combining many input samples and scaling out to a batch cluster etc.).

[1] <https://cp3.irmp.ucl.ac.be/~pdavid/bamboo/>

**Primary author:** DAVID, Pieter (Universite Catholique de Louvain (UCL) (BE))

**Presenter:** DAVID, Pieter (Universite Catholique de Louvain (UCL) (BE))

**Session Classification:** Software