EP Software Seminar

SPEAKER: Enrico Guiraud, Elisabetta Manca

TITLE: **Using RDataFrame, ROOT's declarative analysis tool, in a CMS physics study**

DATE: 16 Oct 2019, 10:00

PLACE: 40/S2-A01 - Salle Anderson

ABSTRACT

With the expected large increase in the amount of available data in LHC Run 3, now more than ever HEP scientists must be able to efficiently write robust, performant analysis software that can take full advantage of the underlying hardware. This is the context in which the ROOT team (and here especially Enrico) developed RDataFrame, a swiss-army knife for data manipulation that provides a high-level interface, in C++ and Python, as well as transparent optimizations such as multi-thread data parallelism. This new tool supports typical HEP workflows and data formats and it has been designed to flexibly scale up from data exploration on a laptop to analysis of millions of events exploiting hundreds of CPU cores. As a result, ROOT users can now write simpler code that runs faster. The first part of the seminar will introduce RDF, showcase its most prominent features, outline current developments and several real-world use-cases. Precision measurements are often affected by large systematic uncertainties related to the models used in simulation, and progress can be made by the extraction of features directly from data. However, the analysis of unprecedented numbers of events in a sustainable scale of time is not possible with standard techniques. The possibilities of using the ROOT RDataFrame to overcome these limitations is demonstrated within the setup of a CMS physics study in the second part of this seminar.

Organised by: Andrea Bocci (EP-CMG), Dirk Düllmann (IT-ST-AD), Peter Hristov (EP-AIP), Axel Naumann (EP-SFT), Niko Neufeld (EP-LBC), and Andreas Salzburger (EP-ADP)
Coffee will be served at 11h00