EIRO forum

serving european science

ESI 2019 Project Day Technologies - CERN

ROOT
An object-oriented Data Analysis Framework

The ROOT system provides a set of frameworks with all the functionality needed to handle and analyse large amounts of data in a very efficient way. It defines the data as a set of objects, and then specialised storage methods are used to get direct access to the separate attributes of the selected objects, without having to touch the bulk of the data. Included in the framework are histogramming methods in an arbitrary number of dimensions, curve fitting, function evaluation, minimisation, graphics and visualisation classes to allow the easy setup of an analysis system that can query and process the data interactively or in batch mode, as well as a general parallel processing framework that can considerably speed up any data analysis process.

AREA OF EXPERTISE

- Information Technology

APPLICATIONS

- Data Analysis in various fields. Examples are: high-energy physics, astronomy, biology, finance and medicine.

IP STATUS

- Available under the Open Source License LPGL (http://root.cern.ch/drupal/ content/ license)

RELATED PUBLICATIONS

- R. Brun, F. Rademakers, ROOT — An object oriented data analysis framework, Nucl. Instr. and Methods in Physics A 389, Issues 1–2, 11 April 1997, (81-86), (http://www.sciencedirect.com/science/article/pii/S016890029700048X).
- I. Antcheva, et al, ROOT — A C++ framework for petabyte data storage, statistical analysis and visualization, Computer Phys. Com., 180, Issue 12, December 2009, (2499-2512), (http://www.sciencedirec com/science/article/pii/S0010465509002550).

CONTACT PERSON

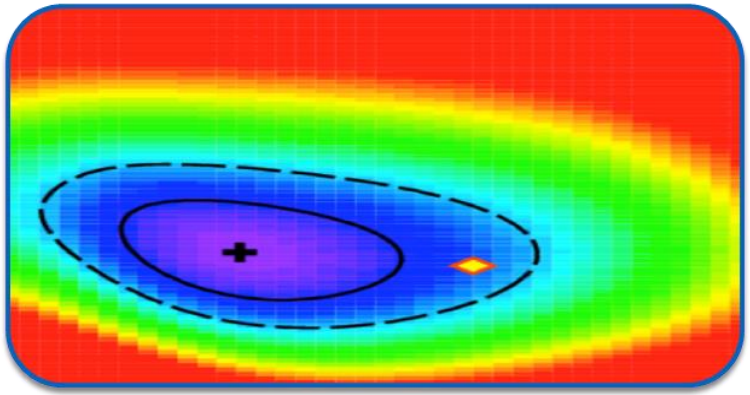
nick.ziogas@cern.ch

Find out more at:
kt.cern

ADVANTAGES

- The command language, the scripting, or macro, language and the programming language are all C++. This is due to a built-in C++ programming language interpreter CINT that also allows for fast prototyping of the macros to be processed since it removes the time-consuming compile/link cycle.
- It also provides a good environment to learn C++. If more performance is needed the interactively developed macros can be compiled using a C++ compiler via a machine independent transparent compiler.
- The system has been designed in such a way that it can query its databases in parallel on clusters of workstations or many-core machines.
- ROOT is an open system that can be dynamically extended by linking external libraries. This makes ROOT a premier platform on which to build data acquisition, simulation and data analysis systems.

ROOT-TMVA (Toolkit for Multivariate Data Analysis)



WHAT

ROOT / TMVA is a modular big data software framework, providing the functionalities needed to deal with big data processing, statistical analysis, visualisation and storage. It is mainly written in C++ but integrated with other languages such as Python and R. Integrated machine learning environment.

HOW

- Deep Learning
- Artificial neural networks
- Rectangular cut optimisation
- Projective likelihood estimation
- Multidimensional estimations
- Linear discriminant analysis
- Function discriminant analysis
- Boosted/bagged decision trees
- Predictive learning
- Support Vector Machine

WHY

- **Good for analysis of extreme large sets of homogeneous data**
- **Used in physics, biology, finance and insurance fraud analysis**
- **Possible application in processing and analysis of large medical datasets, for example**
  - Genomics data
  - EEG/ECG data
  - Biosensor data
- **Open source**

Thank you

Lorenzo Moneta	CERN EP/SFT
Nick Ziogas	CERN IPT/KT
Enrico Chesta	CERN IPT/KT & TWG-IMKTT