Contribution ID: **70** Type: **Poster**

Using Functional Languages and Declarative Programming to analyze ROOT data: LINQtoROOT

Tuesday 2 September 2014 08:00 (1 hour)

Modern high energy physics analysis is complex. It typically requires multiple passes over different datasets, and is often held together with a series of scripts and programs. For example, one has to first reweight the jet energy spectrum in Monte Carlo to match data before plots of any other jet related variable can be made. This requires a pass over the Monte Carlo and the Data to derive the reweighting, and then another pass over the Monte Carlo to plot the variables the analyzer is really interested in. With most modern ROOT based tools this requires separate analysis loops for each pass, and script files to glue to the results of the two analysis loops together. A framework has been developed that uses the functional and declarative features of the C# language and its Language Integrated Query (LINQ) extensions to declare the analysis. The framework uses language tools to convert the analysis into C++ and runs ROOT or PROOF as a backend to get the results. This gives the analyzer the full power of an object-oriented programming language to put together the analysis and at the same time the speed of C++ for the analysis loop. The tool allows one to incorporate C++ algorithms written for ROOT by others. A byproduct of the design is the ability to cache results between runs, dramatically reducing the cost of adding one-more-plot and also to keep a complete record associated with each plot for data preservation reasons. The code is mature enough to have been used in ATLAS analyses. The package is open source and available on the open source site CodePlex.

Summary

A tool that converts functional queries over ROOT data in the C# language into C++ that runs in the ROOT or PROOF environment is described. This open source translation software uses declarative programming to describe the tasks an analyzer wants accomplished. A few lines can generate a large number of plots from several different datasets, and the results can be combined, all in one program. A byproduct of the technique means that previously generated plots can be cached to avoid having to re-run later. This could also have data preservation implications.

Primary author: WATTS, Gordon (University of Washington (US))

Presenter: WATTS, Gordon (University of Washington (US))

Session Classification: Poster session

Track Classification: Computing Technology for Physics Research