



Contribution ID: 231

Type: **Poster presentation**

Hepdoop

Monday, 14 October 2013 15:00 (45 minutes)

“Big Data” is no longer merely a buzzword, but is business-as-usual in the private sector. High Energy Particle Physics is often cited as the archetypal Big Data use case, however it currently shares very little of the toolkit used in the private sector or other scientific communities.

We present the initial phase of a programme of work designed to bridge this technology divide by both performing real HEP analysis workflows using predominately industry “Big Data” tools, formats and techniques, and the reverse: to perform real industry tasks with HEP tools. In doing so it will improve interoperation of those tools, reveal strengths and weakness and enable efficiencies within both communities.

The first phase of this work performs key elements of an LHC Higgs Analysis using very common Big Data tools. These elements include data serialization, filtering and data mining. They are performed with a range of tools chosen not just for performance but also for ease-of-use, maturity and size of user community. This includes technologies such as Protocol Buffers, Hadoop and Python Scikit and for each element we make comparisons with the same analysis performed using current HEP tools such as ROOT, Proof and TMVA.

Summary

Primary author: BHIMJI, Wahid (University of Edinburgh (GB))

Co-authors: WASHBROOK, Andrew John (University of Edinburgh (GB)); BRISTOW, Timothy Michael (University of Edinburgh (GB))

Presenter: BHIMJI, Wahid (University of Edinburgh (GB))

Session Classification: Poster presentations

Track Classification: Event Processing, Simulation and Analysis