



Contribution ID: 145

Type: **Poster presentation**

Towards Provenance and Traceability in CRISTAL for HEP

Monday, October 14, 2013 3:00 PM (45 minutes)

Efficient, distributed and complex software is central in the analysis of high energy physics (HEP) data. One area that has been somewhat overlooked in recent years has been the tracking of the development of the HEP software and of its use in data analyses and its evolution over time. This area of tracking analyses to provide records of actions performed, outcomes achieved and (re-)design decisions taken is an active part of computer science research known as provenance data capture and management. In recent years there has been a wealth of research conducted in the computer science community on this topic, however very little work has been done to address the requirements that have emerged from the HEP domain in the LHC era.

This paper discusses a system known as CRISTAL which has been in development and active use at CERN for the past decade. CRISTAL is a mature and very stable system which was originally developed to track the construction of the ECAL element of CMS. The current usage is discussed in this paper in the context of its application at CMS. CRISTAL has also been commercialised by two external companies. The first company, M1i (Annecy France), has developed a purely BPM (Business Process Management) solution and has sold the product (Agilium) in the retail, logistics and manufacturing sectors; in the second company, Technoledge (Geneva, Switzerland) it is being applied to fuel cell production lines with a focus on provenance data capture and management, therefore demonstrating its maturity as a provenance system.

CRISTAL is currently being moved towards an open source license, and is being used in several EC projects, one example of which is N4U (or neuGRID for Users) where a so-called Analysis Service is being developed to enable neuroimaging researchers to record and track their complex workflows and analyses. This Analysis Service allows for the reuse of clinical research analysis workflows by scientists proving its generic application as a tool for the management of scientific data. We are currently aiming to apply CRISTAL for the indexing of previous HEP data with our team based at CERN. We also feel that CRISTAL can be applied to aid scientists at CERN in creating their experiments through use of the N4U Analysis Service built on top of CRISTAL, allowing them to share, reuse or amend past HEP analyses.

In addition to analysis reuse and sharing, CRISTAL's unique approach to provenance capture provides a means for scientists to log errors and to audit which analyses can be used in conjunction with various datasets. Consequently, CRISTAL provides a unique viewpoint for investigators to see where and more importantly why their experiments may have failed and to store their results. Some initial ideas for the use of CRISTAL in HEP are outlined in detail in this paper. Currently we are investigating the feasibility of using the N4U Analysis Service or a derivative along with CRISTAL to address the requirements of data and analysis logging and provenance capture within the HEP environment.

Primary author: BRANSON, Andrew (University of the West of England (GB))

Co-authors: SHAMDASANI, Jetendr (University of the West of England (GB)); Prof. MCCLATCHEY, Richard (University of the West of England (GB))

Presenter: SHAMDASANI, Jetendr (University of the West of England (GB))

Session Classification: Poster presentations

Track Classification: Distributed Processing and Data Handling A: Infrastructure, Sites, and Virtualization