Contribution ID: **305** Type: **Oral**

LHCb data and software dependencies in the long-term future preservation

Thursday, 13 October 2016 14:30 (15 minutes)

The Large Hadron Collider beauty (LHCb) experiment at CERN specializes in investigating the slight differences between matter and antimatter by studying the decays of beauty or bottom (B) and charm (D) hadrons. The detector has been recording data from proton-proton collisions since 2010. Data preservation (DP) project at the LHCb insures preservation of the experimental and simulated (Monte Carlo) data, the scientific software and the documentation. The project will assist in analysing the old data in the future, which is important for replicating the old analysis, looking for the signals predicted by a new theory and improving the current measurements.

The LHCb data are processed with the software and hardware that have been changing over time. Information about the data and the software have been logged in the internal databases and web portals. A current goal of the DP team is to collect and structure these information into a singular, solid database that can be used immediately for scientific purposes and for the long-term future preservation. The database is being implemented as a graph in Neo4j 2.3 and the supporting components are done in Py2Neo. It contains complete details of the official production of LHCb real, experimental data from 2010 to 2015. The data is represented as nodes with collision type, run dates and ID, that is related to the data-taking year, energy of the beam, reconstruction, applications etc. The model applies to the both simulated and the real data.

Data taken at different points in time is compatible with different software versions. The LHCb software stack is built on Gaudi framework, which can run various packages and applications depending on the activity. The applications require other components and these dependencies are captured in the database. The database will recommend the user which software to run in order to analyse specific data. Interface to the database is provided as a web page with a search engine, which allows the users to query and explore the database. The information in the database are worthwhile in the current research; and from DP point of view, they are crucial for replicating an old analyses in the future.

Once the database is fully completed and tested, we will use the graph to identify legacy software and critical components, for example those used in official collaboration productions. These components should be specially treated for the long-term preservation.

Tertiary Keyword (Optional)

Secondary Keyword (Optional)

Primary Keyword (Mandatory)

Preservation of analysis and data

Primary author: TRISOVIC, Ana (University of Cambridge (GB))

Co-authors: COUTURIER, Ben (CERN); JONES, Christopher Rob (University of Cambridge (GB))

Presenter: TRISOVIC, Ana (University of Cambridge (GB))

Session Classification: Track 8: Security, Policy and Outreach

Track Classification: Track 8: Security, Policy and Outreach