Contribution ID: **41**                                                                 Type: **not specified**

# Use of the Hadoop structured storage tools for the ATLAS EventIndex event catalogue.

*Friday, 2 October 2015 16:45 (15 minutes)*

The ATLAS experiment collects billions of events per year of data-taking, and processes them to make them available for physics analysis in several different formats. An even larger amount of events is in addition simulated according to physics and detector models and then reconstructed and analysed to be compared to real events. The EventIndex is a catalogue of all events in each production stage; it includes for each event a few identification parameters, some basic non-mutable information coming from the online system, and the references to the files that contain the event in each format (plus the internal pointers to the event within each file for quick retrieval). Each EventIndex record is logically simple but the system has to hold many tens of billions of records, all equally important. The Hadoop technology was selected at the start of the EventIndex project development in 2012 and proved to be robust and flexible to accommodate this kind of information; both the insertion times and query response times are acceptable for the continuous and automatic operation that started in spring 2015.

This talk will describe the EventIndex data input and organisation in Hadoop and explain the operational challenges that were overcome in order to achieve the expected good performance.

**Primary authors:**   Dr FAVARETO, Andrea (University and INFN Genova (Italy));   Prof. BARBERIS, Dario (University and INFN Genova (Italy))

**Presenter:**   Dr FAVARETO, Andrea (University and INFN Genova (Italy))

**Session Classification:**   Workload Management Systems in Applied Research and BigData