

# Index files for Belle II - very small skim containers

*Thursday, 13 October 2016 14:45 (15 minutes)*

The Belle II experiment will generate very large data samples. In order to reduce the time for data analyses, loose selection criteria will be used to create files rich in samples of particular interest for a specific data analysis (data skims). Even so, many of the resultant skims will be very large, particularly for highly inclusive analyses. The Belle II collaboration is investigating the use of “index-files” where instead of the skim recording the actual physics data, we record pointers to events of interest in the complete data sample. This reduces the skim files by 2 orders of magnitude. While this approach was successfully employed by the Belle experiment where the majority of the data analysis was performed on a single center cluster, index files are significantly more challenging in a distributed computing system such as the Belle II computing grid.

The approach we use is for each skim file to record metadata identifying the original parent file, as well as the location of the event within the parent file. Since we employ the ROOT file container system it is straight forward to read just the identified events of interest. For remote file access, we employ the XROOTD protocol to both select identified events and transfer them to the worker nodes used for data analysis.

This presentation will describe the details of the implementation of index-files within the Belle II grid. We will also describe the results of tests where the analyses were performed locally and situations where data was transferred internationally, in particular, from Europe to Australia. The configuration of the XROOTD services for these tests has a very large impact on the performance of the system and will be reported in the presentation

## Tertiary Keyword (Optional)

Distributed workload management

## Secondary Keyword (Optional)

Computing middleware

## Primary Keyword (Mandatory)

Data processing workflows and frameworks/pipelines

**Primary author:** Prof. SEVIOR, Martin (University of Melbourne)

**Co-authors:** Ms HSU, Chia-Ling (University of Melbourne); Dr MIYAKE, Hideki (High Energy Accelerator Research Organization, Tsukuba, Japan); Prof. UEDA, Ikuo (High Energy Accelerator Research Organization, Tsukuba, Japan); Prof. HARA, Takanori (High Energy Accelerator Research Organization, Tsukuba, Japan); Prof. KUHR, Thomas (LMU, Munich, Germany); Mr BLOOMFIELD, Tristan (University of Melbourne)

**Presenter:** Prof. SEVIOR, Martin (University of Melbourne)

**Session Classification:** Track 7: Middleware, Monitoring and Accounting

**Track Classification:** Track 7: Middleware, Monitoring and Accounting