



Contribution ID: 16

Type: **Oral presentation to parallel session**

BESIII physical analysis on hadoop platform

Thursday 17 October 2013 13:30 (22 minutes)

This paper brings the idea of MapReduce parallel processing to BESIII physical analysis, gives a new data analysis system structure based on HADOOP framework; Optimizes the process of data processing, by establish an event level metadata(TAG) database and do event pre-selection based on TAGs, significantly reduce the number of events that need to do further analysis by 2-3 classes, which reduces the I/O volume and improves the efficiency of data analysis jobs; The event storage structure in DST files are re-organized to optimize the selective reading patterns with event pre-selection. Designs the MapReduce models for TAG generation, TAG based event pre-selection and event analysis, and develop proper MapReduce libs that fit for the ROOT framework to do things such as data splitting, event fetching and result merging. An 8-nodes cluster is used for system test, the testing result shows that the new system shortens the data analyzing time by 80%, and the cluster system shows great scalability when adding more worker nodes.

Author: Dr SUN, Gongxing (INSTITUTE OF HIGH ENERGY PHYSICS)

Co-authors: Mr ZANG, Dongsong (Chinese Academy of Sciences (CN)); Mr HUO, Jing (IHEP); Ms LEI, Xiaofeng (IHEP)

Presenter: Dr SUN, Gongxing (INSTITUTE OF HIGH ENERGY PHYSICS)

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

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