



Contribution ID: 138

Type: Oral

High Performance Data Format for CLAS12

Monday, November 4, 2019 2:15 PM (15 minutes)

With increasing data volume from Nuclear Physics experiments requirements to data storage and access are changing. To keep up with large data sets new data formats are needed for efficient processing and analysis of the data. Frequently, in the experiments data goes through stages from data acquisition to reconstruction and data analysis and data is converted from one format to another causing wasted CPU cycles.

In this work we present High Performance Output (HIPO) data format developed for CLAS12 experiment at Jefferson National Laboratory. It was designed to fit the needs of data acquisition and high level data analysis, to avoid data format conversions at different stages of data processing. The new format was designed to store different event topologies from reconstructed data in tagged form for efficient access by different analysis groups. In centralized data skimming applications HIPO data format significantly outperforms standard data formats used in Nuclear and High Energy Physics (ROOT) and industry standard formats, such as Apache Avro and Apache Parquet.

Consider for promotion

No

Primary author: Dr GAVALIAN, Gagik (Jefferson Lab)

Presenter: Dr GAVALIAN, Gagik (Jefferson Lab)

Session Classification: Track 4 – Data Organisation, Management and Access

Track Classification: Track 4 – Data Organisation, Management and Access