Contribution ID: **204** Type: **Oral**

LHCb data processing optimization using Event Index

Tuesday 19 January 2016 14:50 (25 minutes)

Experiments in high energy physics routinely require processing and storing massive amounts of data. LHCb Event Index is an indexing system for high-level event parameters. It's primary function is to quickly select subsets of events. This paper discusses applications of Event Index to optimization of the data processing pipeline.

The processing and storage capacity is limited and divided among different physics studies by expert assigned physics value. The selection pipeline consists of analyst-written algorithms (triggers and stripping lines). An event passes the selection if any of the algorithms finds it useful. Considering that some events mass more than one algorithm, the rate adjustment requires guesswork and has to be done in several iterations. In other words finding the optimal balance between the different algorithms is an unnecessary time-consuming burden an operator has to deal with.

Having access to the set of per-event decisions Event Index can be used to optimize the selection procedure, relieving the algorithms authors from manually adjusting the parameters and achieving better overall efficiency.

From the implementation point of view Event Index is based on Apache Lucene indices distributed over multiple shards on multiple nodes. The data is stored in a problem-neutral format, thus the system can easily be adapted for new tasks.

Authors: USTYUZHANIN, Andrey (Yandex School of Data Analysis (RU)); Mr REDKIN, Artem (Yandex Data Factory); Mr TROFIMOV, Ilya (Yandex Data Factory); KAZEEV, Nikita (Yandex School of Data Analysis (RU))

Presenter: KAZEEV, Nikita (Yandex School of Data Analysis (RU))

Session Classification: Track 1

Track Classification: Computing Technology for Physics Research