CHEP 2016 Conference, San Francisco, October 8-14, 2016

Contribution ID: **337** Type: **Oral**

LHCb trigger streams optimization

Tuesday 11 October 2016 14:15 (15 minutes)

The LHCb experiment stores around 10ˆ11 collision events per year. A typical physics analysis deals with a final sample of up to 10ˆ7 events. Event preselection algorithms (lines) are used for data reduction. They are run centrally and check whether an event is useful for a particular physical analysis. The lines are grouped into streams. An event is copied to all the streams its lines belong, possibly duplicating it. Due to the storage format allowing only sequential access, analysis jobs read every event and discard the ones they don't need.

This scheme efficiency heavily depends on the streams composition. By putting similar lines together and balancing the streams sizes it's possible to reduce the overhead. There are additional constraints that some lines are meant to be used together so they must go to one stream. The total number of streams is also limited by the file management infrastructure.

We developed a method for finding an optimal streams composition. It can be used for different cost functions, has the number of streams as an input parameter and accommodates the grouping constraint. It has been implemented using Theano [1] and the results are being incorporated into the streaming [2] of the LHCb Turbo [3] output with the projected analysis jobs IO time decrease of 20-50%.

[1] Theano: A Python framework for fast computation of mathematical expressions, The Theano Development Team
[2] Separate file streams https://gitlab.cern.ch/hschrein/Hlt2StreamStudy, Henry Schreiner et. al
[3] The LHCb Turbo Stream, Sean Benson et al., CHEP-2015

Tertiary Keyword (Optional)

Data processing workflows and frameworks/pipelines

Secondary Keyword (Optional)

Distributed workload management

Primary Keyword (Mandatory)

Distributed data handling

Authors: PANIN, Alexander (Yandex School of Data Analysis (RU)); USTYUZHANIN, Andrey (Yandex School of Data Analysis (RU)); Mr REDKIN, Artem (Yandex Data Factory); DERKACH, Denis (Yandex School of Data Analysis (RU)); Mr TROFIMOV, Ilya (Yandex Data Factory); VESTERINEN, Mika Anton (Ruprecht-Karls-Universitaet Heidelberg (DE)); KAZEEV, Nikita (Yandex School of Data Analysis (RU)); NEYCHEV, Radoslav (Yandex School of Data Analysis (RU))

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

Session Classification: Track 4: Data Handling

Track Classification: Track 4: Data Handling