



Contribution ID: 316

Type: Poster

Data Sampling challenges in the ALICE O2 distributed processing system

The ALICE experiment at the CERN LHC focuses on studying the quark-gluon plasma produced by heavy-ion collisions. After the Long Shutdown 2 in 2019-2020, the ALICE Experiment will see its data input throughput increase a hundredfold, up to 3.4 TB/s. In order to cope with such a large amount of data, a new online-offline computing system, called O2, will be deployed. By reconstructing the data online, it will be possible to compress the data stream down to 90GB/s before storing it permanently.

One of the key software components of the system will be the data Quality Control (QC) that replaces the existing online Data Quality Monitoring and offline Quality Assurance. It is this framework and infrastructure, which will be responsible for all aspects related to the analysis software aimed at identifying possible issues with the data itself, and indirectly with the underlying processing done both synchronously and asynchronously. Since analyzing the full stream of data online would exceed the available computational resources, a reliable and efficient sampling will be needed. It should provide a few percent of data selected randomly in a statistically sound manner with a minimal impact on the main dataflow. Extra requirements include the option to choose messages corresponding to the same events over a group of computing nodes and the optional possibility to ensure getting a fixed amount of data at the cost of blocking the main dataflow.

In this paper we present the design of the O2 Data Sampling software. In particular, we highlight our requirements for pseudo-random number generators to be used for sampling decisions, as well as the results of the benchmark we performed to evaluate different possibilities. Finally we report on a large scale test of the O2 Data Sampling we carried out.

Authors: KONOPKA, Piotr (AGH University of Science and Technology (PL)); Mr VON HALLER, Barthelemy (CERN)

Presenter: Mr VON HALLER, Barthelemy (CERN)

Session Classification: Poster Session

Track Classification: Track 1: Computing Technology for Physics Research