



Contribution ID: 219

Type: Oral

Multi-resource planning: Simulations and study of a new scheduling approach for distributed data production in High Energy and Nuclear Physics

Thursday 21 January 2016 14:25 (25 minutes)

Distributed data processing has found its application in many fields of science (High Energy and Nuclear Physics (HENP), astronomy, biology to name only those). We have focused our research on distributed data production which is an essential part of computations in HENP. Using our previous experience, we have recently proposed a new scheduling approach for distributed data production which is based on the network flow maximization model. It has a polynomial complexity which provides required scalability with respect to the size of computations. Our approach improves the overall data production throughput due to three factors: transferring input files in advance before their processing which allows to decrease I/O latency; balancing of the network traffic, which includes splitting the load between several alternative transfer paths; and transferring files sequentially in a coordinated manner, which allows to reduce the influence of possible network bottlenecks. In this contribution, we intend to present the results of our new simulations based on the Grid-Sim framework which is one of the commonly used tools in the field of distributed computations. In these simulations we study the behavior of commonly used scheduling approaches compared to our recently proposed approach in a realistic environment created by using the data from the STAR and ALICE experiments. We will also discuss how the data production can be optimized with respect to possible bottlenecks (network, storage, CPUs) and study the influence of the background traffic on the simulated schedulers. The final goal of the research is to integrate the proposed scheduling approach into the real data production framework. In order to achieve this we are constantly moving our simulations towards real use cases, study scalability of the model and the influence of the scheduling parameters on the quality of the solution.

Authors: MAKATUN, Dzmitry (Faculty of Nuclear Physics and Physical Engineering, Czech Technical University in Prague); Prof. RUDOVA, Hana (Faculty of Informatics, Masaryk University, Brno, Czech Republic); Dr LAURET, Jerome (Brookhaven National Laboratory); SUMBERA, Michal (Acad. of Sciences of the Czech Rep. (CZ))

Presenter: MAKATUN, Dzmitry (Faculty of Nuclear Physics and Physical Engineering, Czech Technical University in Prague)

Session Classification: Track 1

Track Classification: Computing Technology for Physics Research