Introduction

Since its beginning, the lifecycle of LHC [1] has been composed of the data-taking (Run) and upgrade (Long Shutdown, LS) periods. During the Run periods, the LHC experiments, such as ALICE [2], require a significant amount of storage and computing resources to store, process and analyze the experimental and simulated data. In addition to this, the requirements for these resources are increasing over the years, at each LHC running period. In order to predict the resource requirements of the ALICE Experiment for a particular LHC Run period, we developed a flexible and highly configurable simulation tool, which does this prediction by discrete-event simulation (DES) of ALICE data flow processes. The tool provides a Web GUI, which allows entering of all the necessary parameters and to graphically visualize the results of simulations. The steps of the current development can be summarized as follows:

- Description of the ALICE data flow processes for certain LHC Runs
- Definition of input parameters necessary for simulations
- Automation of ALICE data flow in Run 3 and estimation of required storage resources
- Same simulation for 2017 year of Run 2 and comparison of the results with real numbers for the validation of tool results.

ALICE data flow during LHC Run3/Run4 [3]

The simulation tool architecture

The simulation tool provides a Web GUI, which allows to manipulate with the default input parameters and run the simulation with different scenarios. The interface is also used for the graphical representation of simulation results and allows to export these results in the form of PDF reports.

Resource types and their capacities

The amount of CPU and storage capacities of resource types involved in the data flow process was estimated. These parameters are required for the estimation of computing resource usage.

Data management policies

The data replication and removal policies are significantly influence the resource usage estimation results. The amount of data for each data type is kept on each storage resources, and then moved to space or removed.

Initial results (Estimated amount of created data during 1 year of Run 3)

The results of some simulations for the 2017 year of Run 3 showed that the total amount of RAW data is ~9 PB, meanwhile the real number is ~8 PB [1]. The works on the development of the tool are in progress to make the simulations as realistic as possible.

Conclusion and future work

This simulation software is a flexible and highly configurable tool, which presently allows to estimate and graphically visualize the volume of storage resources, necessary for each LHC Run and upgrade period. In order to validate the initial results of simulations, another simulation has been done for the storage resource usage during the 2017 year of LHC Run 2 and compared with real numbers. The works on the development of the tool are continued to make the simulations more precise, as well as with the vision to make the software applicable to the other experiments of LHC.

References