



Contribution ID: 283 Contribution code: TUE 02

Type: Poster

Surrogate Modeling for Scalable Evaluation of Distributed Computing Systems for HEP applications

Tuesday 22 October 2024 15:36 (15 minutes)

The sheer volume of data generated by LHC experiments presents a computational challenge, necessitating robust infrastructure for storage, processing, and analysis. The Worldwide LHC Computing Grid (WLCG) addresses this challenge by integrating global computing resources into a cohesive entity. To cope with changes in the infrastructure and increased demands, the compute model needs to be adapted. Simulations of different compute models present a feasible approach for evaluating different design candidates. However, running these simulations incurs a trade-off between accuracy and scalability. For example, while the simulator DCSim can provide accurate results, it falls short on scalability when increasing the size of the simulated platform. Generative Machine Learning as a surrogate is successfully used to overcome these limitations in other domains that exhibit similar trade-offs between scalability and accuracy, such as the simulation of detectors.

In our work, we evaluate the usage of three different machine learning models as surrogate models for the simulation of distributed computing systems and assess their ability to generalize to unseen jobs and platforms. We show that those models can predict the simulated platforms' main observables derived from the execution traces of compute jobs with approximate accuracy. Potential for further improving the predictions lies in using other machine learning models and different encodings of the platform-specific information to achieve better generalizability for unseen platforms.

Primary authors: SCHMID, Larissa (KIT - Karlsruhe Institute of Technology (DE)); HORZELA, Maximilian Maria (KIT - Karlsruhe Institute of Technology (DE)); ZHYLA, Valerii (KIT - Karlsruhe Institute of Technology (DE)); GIFFELS, Manuel (KIT - Karlsruhe Institute of Technology (DE)); QUAST, Gunter (KIT - Karlsruhe Institute of Technology (DE)); KOZIOLEK, Anne (KIT - Karlsruhe Institute of Technology (DE))

Presenter: SCHMID, Larissa (KIT - Karlsruhe Institute of Technology (DE))

Session Classification: Poster session

Track Classification: Track 4 - Distributed Computing