

Implementation of tracc as-a-service

Particle tracking at Large Hadron Collider (LHC) experiments is a crucial component of particle reconstruction, yet it remains one of the most computationally challenging tasks in this process. As we approach the High-Luminosity LHC era, the complexity of tracking is expected to increase significantly. Leveraging coprocessors such as GPUs presents a promising solution to the rising computational demands. The tracc project is a tracking demonstrator under the ACTS software designed to harness GPU resources for tracking. Despite promising initial results, the deployment of GPU algorithms such as tracc in production chains remains a significant challenge. In this talk, we present an as-a-service (aaS) approach to address these deployment challenges. A dedicated backend written to efficiently manage multiple concurrent requests from the server and load multiple model instances onto a dedicated GPU server is presented. Our results demonstrate increased resource utilization and a significant improvement in throughput compared to standalone tracc implementations.

Focus areas

HEP

Primary author: COCHRAN-BRANSON, Miles (University of Washington (US))

Co-authors: ZHAO, Haoran (University of Washington (US)); JU, Xiangyang (Lawrence Berkeley National Lab. (US)); CHOU, Yuan-Tang (University of Washington (US))

Presenter: COCHRAN-BRANSON, Miles (University of Washington (US))