Contribution ID: **266**   Type: **not specified**

Portable Acceleration of CMS Mini-AOD Production with Coprocessors as a Service

Thursday, 9 November 2023 16:50 (15 minutes)

Computing demands for large scientific experiments, such as the CMS experiment at CERN, will increase dramatically in the next decades.To complement the future performance increases of software running on CPUs, both in online TDAQ systems and offline data processing, explorations of coprocessor usage hold great potential and interest. We explore the novel approach of Services for Optimized Network Inference on Co-processors (SONIC) and study the deployment of this as-a-Service approach in large-scale data processing. In this setup, the main CMS Mini-AOD creation workflow is executed on CPUs, while several machine learning (ML) inference tasks are offloaded onto (remote) coprocessors, such as GPUs. With experiments performed at Google Cloud and the Purdue Tier-2 computing center, we demonstrate the ML algorithm acceleration individually and the throughput improvement for the entire workflow. We also show the generalizability of the approach, demonstrating deployment on CPUs without performance decrease. SONIC enables high coprocessor usage with portability to different hardware types enabled. This is the first demonstration of a realistic CMS workflow with co-processors as-a-service computing paradigm. Future plans and challenges will also be discussed.

Early Career

Presenter: LIU, Mia (Purdue University)

Session Classification: RDC5