



Politecnico
di Torino

SmartData@PoliTO



April 26th, 2022, 2:00 PM CEST

Sala Ciminiera

5th floor DAUIN

Corso Castelfidardo 34



Marco Canini Associate Professor - KAUST



Accelerated Deep Learning via Efficient, Compressed and Managed Communication

ABSTRACT

Scaling deep learning to a large cluster of workers is challenging due to high communication overheads that data-parallelism entails. This talk describes our efforts to rein in distributed deep learning's communication bottlenecks. We describe SwitchML, the state-of-the-art in-network aggregation system for collective communication using programmable network switches. We introduce OmniReduce, an efficient streaming aggregation system that exploits sparsity to maximize effective bandwidth use. We touch on our work to develop compressed gradient communication algorithms that perform efficiently and adapt to network conditions. Lastly, we take a broad look at the challenges to accelerated decentralized training in the federated learning setting where heterogeneity is an intrinsic property of the environment.

