Introduction

This is the Deployment Guide for Mellanox Scalable Hierarchical Aggregation and Reduction Protocol (SHARP)™.

This technology improves upon the performance of MPI and Machine Learning collective operation, by offloading collective operations from the CPU to the network and eliminating the need to send data multiple times between endpoints.

This innovative approach decreases the amount of data traversing the network as aggregation nodes are reached, and dramatically reduces collective operations time. Implementing collective offloads communication algorithms supporting streaming for Machine Learning in the network also has additional benefits, such as freeing up valuable CPU resources for computation rather than using them to process communication.