ODOS: OpenMP DPU Offloading Support

Data processing units (DPUs) as network co-processors are an emerging trend in our community, with plenty of opportunities yet to be explored. These have been generally used as domain-specific accelerators transparent to application developers; In the HPC field, DPUs have been used as MPI accelerators, but also to offload some tasks from the general-purpose processor. However, the latter required application developers to deploy MPI ranks in the DPUs, as if they were remote (weak) compute nodes, hence considerably hindering programmability. The wide adoption of OpenMP as the threading model in the HPC arena, along with that of GPU accelerators, is making OpenMP offloading to GPUs a wide trend for HPC applications. ODOS is the OpenMP offloading support for network co-processor DPUs. ODOS implements in LLVM support for OpenMP standard offloading semantics.

Barcelona Supercomputing Center - Centro Nacional de Supercomputación