UCX Latest and Greatest 2023

UCF 2023 – Dec 7

Yossi Itigin

Protocols v2 update

- **Enabled by default since v1.16**
  - Improved locality detection for GPU memory
  - Support GPU memory for RMA and Atomic operations
  - Reduced SW overheads for send request processing
  - Extensive performance tuning
- **Coming next**
  - Device memory pipeline – select RNDV bounce buffer according to memory locality
    - Introduce ”protocol variants”
  - Improved protocol performance estimation
    - Consider local and remote CPU utilization
  - Tuning for Grace-Hopper
    - Basic operation time cost per platform

Additional work

- **RDMO**
  - Offload shmem put+atomic to DPU
- **NCCL/UCX plugin**
  - Fixed multiple stability and deadlock issues
  - Performance improvements: worker per thread, support multi-receive
- **On demand paging**
  - Introduce UCX_REG_NONBLOCK_MEM_TYPES
  - WIP – Whole-VA memory key using ODP
- **XPMEM prefetch optimization**
  - During a page fault, prefetch and pin several pages forward and backward
  - Improves UCX intra-node bandwidth when communication buffer is not reused
- **shmem_lock to use MCS scheme**
  - Enabled by default

Plans for 2024

- **Active message extensions**
  - ucp_am_fetch_nbx() – Send active message and read data from the target
    - Use RDMA_WRITE from target to initiator
    - No need for progress on initiator side to receive the data
  - Multi-fragment receive callback
    - Instead of malloc+copy to a single buffer
- **Priority per operation**
  - Optional parameter per send request
  - IB: Open QPs on different SLs and schedule accordingly
  - No ordering guarantee between different priorities
- **NCCL/UCX plugin**
  - Performance tuning for all cases
- **Improved receive queue (ConnectX-8)**
  - Support multi-plane network and out-of-order receive
  - Small received messages do not need to consume the entire receive buffer
  - No need to repost WQE when data is scattered to CQE