NUMA-Aware Thread and Resource Scheduling for Terabit Data Movement

Taeuk Kim, Awais Khan, Youngjae Kim, Sungyong Park, Scott Atchley

PDSW-DISCS 17 WIP session
November 13, 2017, Denver, USA
Need for Data Coupling over ESnet

- Data Coupling across HPC facilities
  - Nuclear interaction datasets generated at NERSC needed at the OLCF for Peta-scale simulation
  - Climate simulations run at ALCF and OLCF validated with BER datasets at ORNL data centers
Terabits Network Environment

- Terabits network improvement only contributed the network transfer rate.

But, data sets are stored at slow storage systems!
LADS: Layout-Aware Data Scheduling [FAST’15]

• LADS offers an end-to-end data transfer optimization.

LADS solved the impedance mismatch problem between the faster network and slower storage system!
What Memory Bottleneck Occurs in LADS?
Architectural Overview for LADS

- NUMA-based DTN Architecture in Source and Sink
Memory Bottleneck with Single RMA Buffer

- NUMA-based DTN Architecture in Source and Sink
Memory Bottleneck with Single RMA Buffer

- NUMA-based DTN Architecture in Source and Sink

Remote Memory Accesses!!!

CPU Socket 1 accessing RMA Buffer hosted by CPU Socket 0
Multiple RMA Buffers

- Distributing the RMA buffer to all CPU sockets
  - To reduce the remote socket’s memory access
Multiple RMA Buffers

Possibility for accessing remote socket’s memory reduced!
Memory-aware Thread Scheduling (MTS)

- Binding all threads to in-socket RMA buffer
- Load balancing among in-socket NUMA nodes
Memory-aware Thread Scheduling (MTS)

- Local Memory Accesses & Load Balancing
Test-bed Configuration

• Data Transfer Nodes (DTNs)
  • 2 CPU sockets, 4 NUMA nodes, 24 cores
  • 128GB memory
  • InfiniBand EDR (100Gb/s)

• Workloads
  • 8x3GB files (Big file workload)
  • 24,000x1MB files (Small file workload)

• Storage Systems
  • We used the memory file system (tmpfs) to eliminate storage bottlenecks.
Evaluation

Throughput increased to an average of 24.3%!
Contact:
Taeuk Kim (taugi323@sogang.ac.kr)
Department of Computer Science and Engineering
Sogang University, Seoul, Republic of KOREA