HPC ADVISORY COUNCIL

DL-POLY Performance Benchmark and Profiling

August 2013

Note

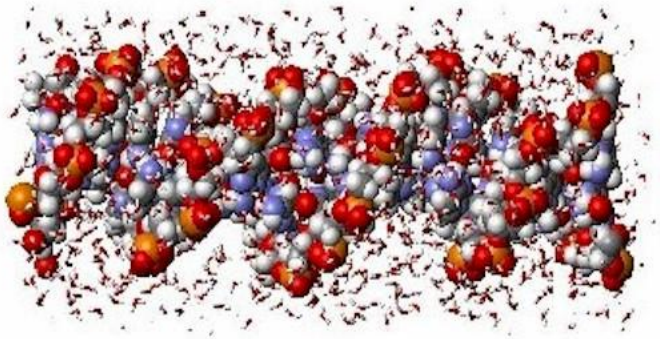
- **The following research was performed under the HPC Advisory Council activities**
  - Special thanks for: HP, Mellanox

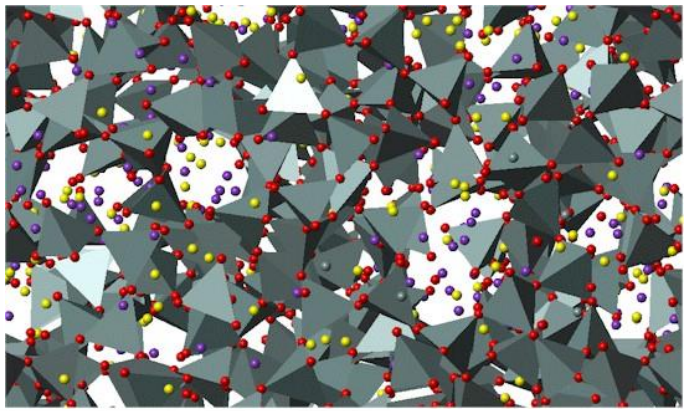
- **For more information on the supporting vendors solutions please refer to:**
  - www.mellanox.com, http://www.hp.com/go/hpc

- **For more information on the application:**
  - http://www.stfc.ac.uk/CSE/randd/ccg/software/DL_POLY/25526.aspx

DL-POLY

- **DL-POLY**
  - Is a general purpose classical molecular dynamics simulation software
  - Developed at Daresbury Laboratory by I.T. Todorov and W. Smith.
- **DL_POLY_4**
  - General design provides scalable performance from a single processor workstation to a high performance parallel computer.
  - Can be compiled a parallel application code, provided an MPI2 instrumentation is available on the parallel machine
  - DL_POLY_4 offers fully parallel I/O as well as a netCDF alternative (HDF5 library dependence) to the default ASCII trajectory file
  - It is supplied in source form under license





Objectives

- **The presented research was done to provide best practices**
  - DL-POLY performance benchmarking
  - Interconnect performance comparisons
  - MPI performance comparison
  - Understanding DL-POLY communication patterns

- **The presented results will demonstrate**
  - The scalability of the compute environment to provide nearly linear application scalability

Test Cluster Configuration

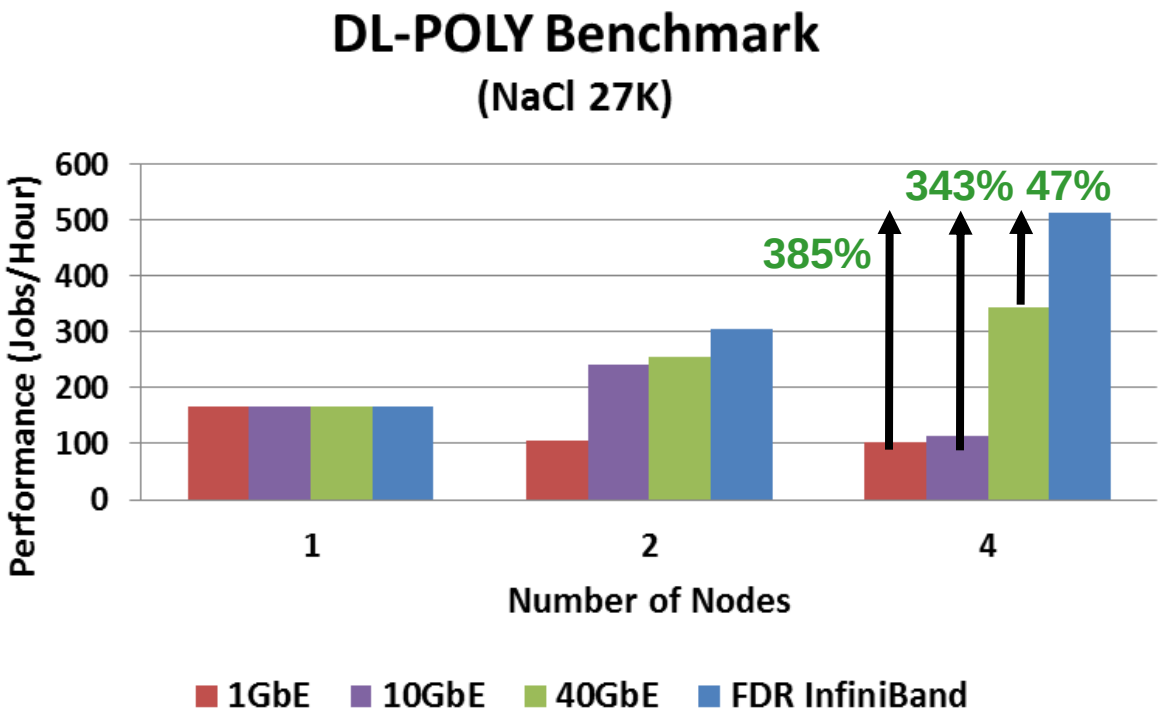
- **HP ProLiant SL230s Gen8 4-node "Athena" cluster**
  - Processors: Dual Eight-Core Intel Xeon E5-2680 @ 2.7 GHz
  - Memory: 32GB per node, 1600MHz DDR3 DIMMs
  - OS: RHEL 6 Update 2, OFED 2.0 InfiniBand SW stack
- **Mellanox Connect-IB FDR InfiniBand Adapters and ConnectX-3 VPI Adapters**
- **Mellanox SwitchX SX6036 56Gb/s FDR InfiniBand and 40G/s Ethernet VPI Switch**
- **MPI: Platform MPI 8.3**
- **Compiler: Intel Compilers Version 13 (Intel Composer XE 2013)**
- **Application: DL-POLY 4.04**
- **Benchmark Workload:**
- **Input dataset:**
  - Sodium Chloride with Ewald summation. System size is 27K ions

About HP ProLiant SL230s Gen8

Item	SL230 Gen8
Processor	Two Intel® Xeon® E5-2600 Series, 4/6/8 Cores,
Chipset	Intel® Sandy Bridge EP Socket-R
Memory	(512 GB), 16 sockets, DDR3 up to 1600MHz, ECC
Max Memory	512 GB
Internal Storage	Two LFF non-hot plug SAS, SATA bays or Four SFF non-hot plug SAS, SATA, SSD bays Two Hot Plug SFF Drives (Option)
Max Internal Storage	8TB
Networking	Dual port 1GbE NIC/ Single 10G Nic
I/O Slots	One PCIe Gen3 x16 LP slot 1Gb and 10Gb Ethernet, IB, and FlexF abric options
Ports	Front: (1) Management, (2) 1GbE, (1) Serial, (1) S.U.V port, (2) PCIe, and Internal Micro SD card & Active Health
Power Supplies	750, 1200W (92% or 94%), high power chassis
Integrated Management	iLO4 hardware-based power capping via SL Advanced Power Manager
Additional Features	Shared Power & Cooling and up to 8 nodes per 4U chassis, single GPU support, Fusion I/O support
Form Factor	16P/8GPUs/4U chassis

DL-POLY Performance - Interconnect

- **InfiniBand FDR is the most efficient inter-node communication for DL-POLY**
  - Outperforms 1GbE by 385% at 4 nodes
  - Outperforms 10GbE by 343% at 4 nodes
  - Outperforms 40GbE by 47% at 4 nodes
- **1GbE do not show performance gain beyond 1 node**

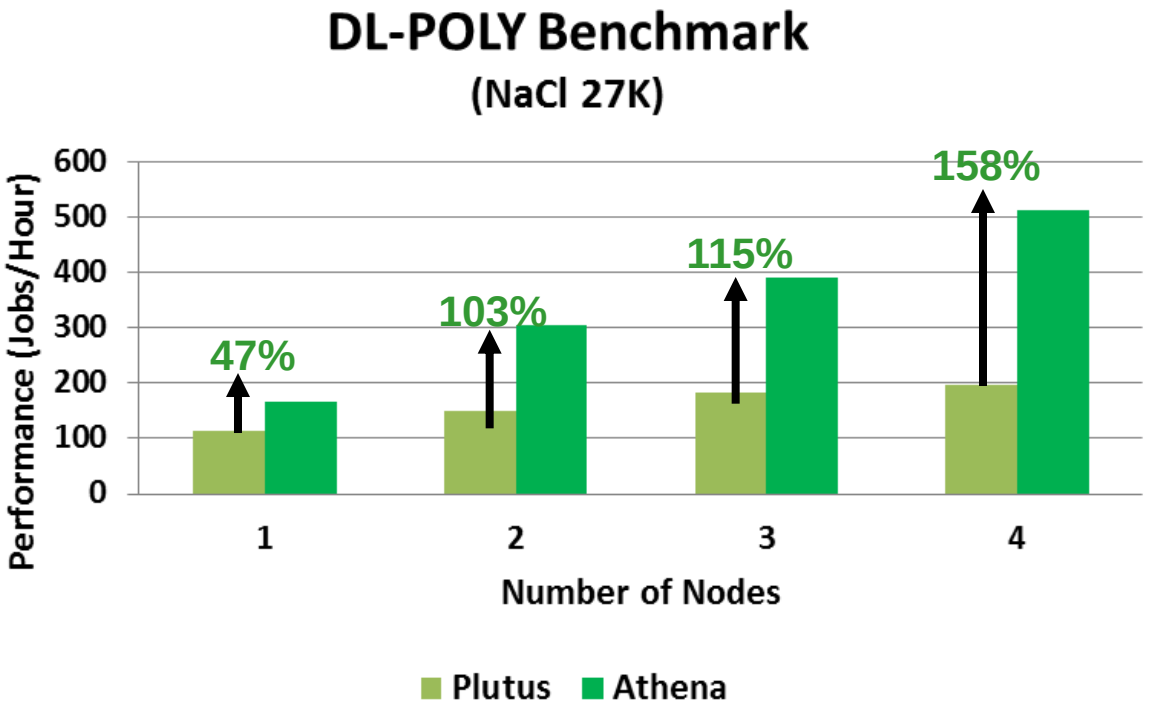


Higher is better

16 Processes/Node

DL-POLY Performance - Processors

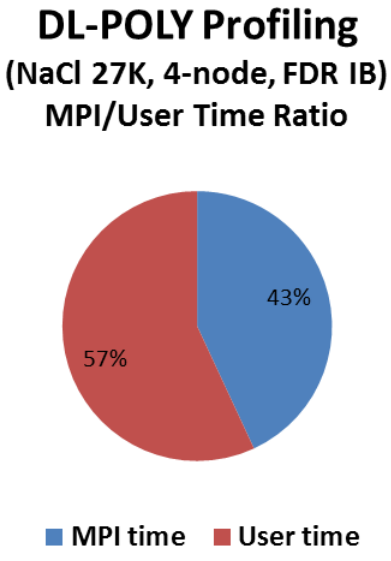
- **Intel E5-2680 processors (Sandy Bridge) cluster outperforms prior CPU generation**
  - Performs 158% higher than X5670 cluster at 4 nodes
- **System components used:**
  - Athena: 2-socket Intel E5-2680 @ 2.7GHz, 1600MHz DIMMs, FDR InfiniBand, 1HDD
  - Plutus: 2-socket Intel X5670 @ 2.93GHz, 1333MHz DIMMs, QDR InfiniBand, 1HDD
  - Athena has PCIe Gen3 bus which can enhance the communication at scale

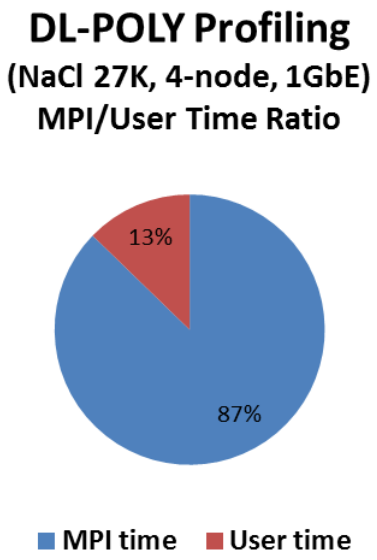


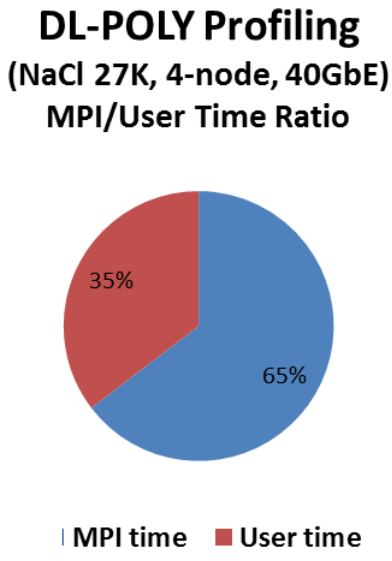
Higher is better

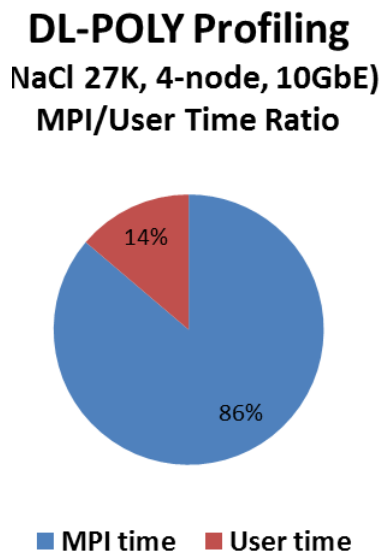
DL-POLY Profiling – MPI Time Ratio

- **InfiniBand FDR reduces the communication time at scale**
  - InfiniBand FDR consumes about 43% of total runtime
  - 40GbE consumes 65% of total time, while 10GbE and 1GbE consumes about 87%



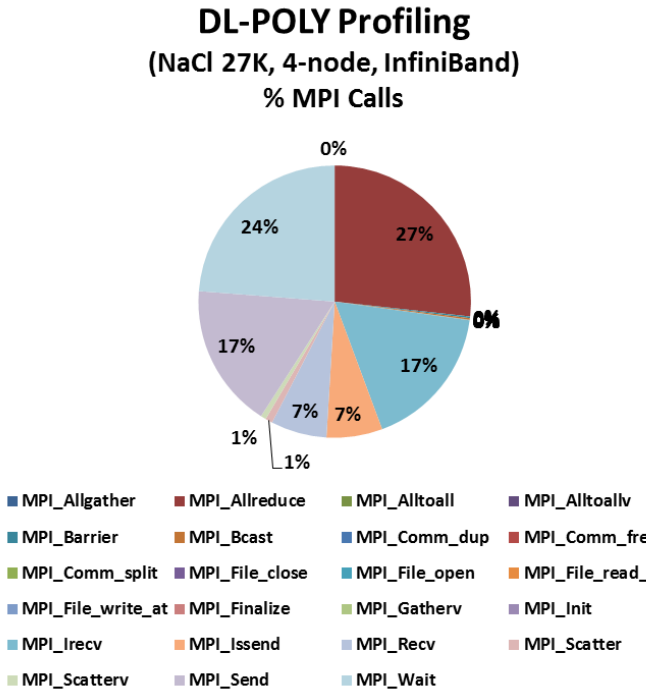


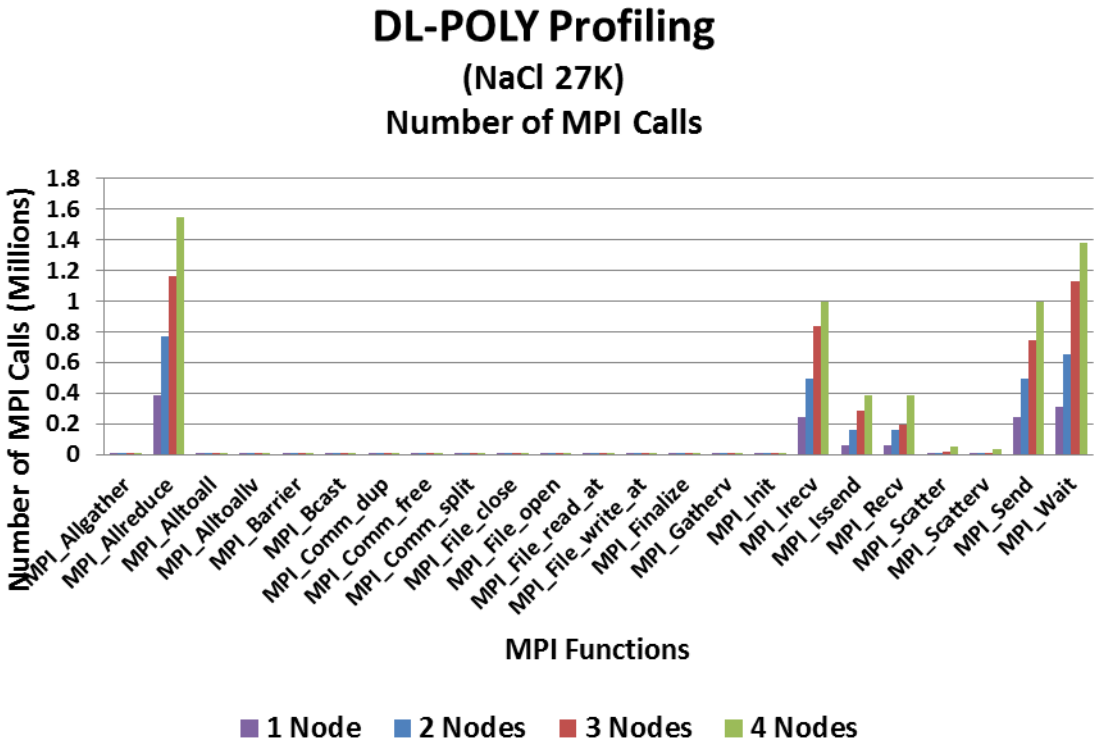




DL-POLY Profiling – MPI Functions

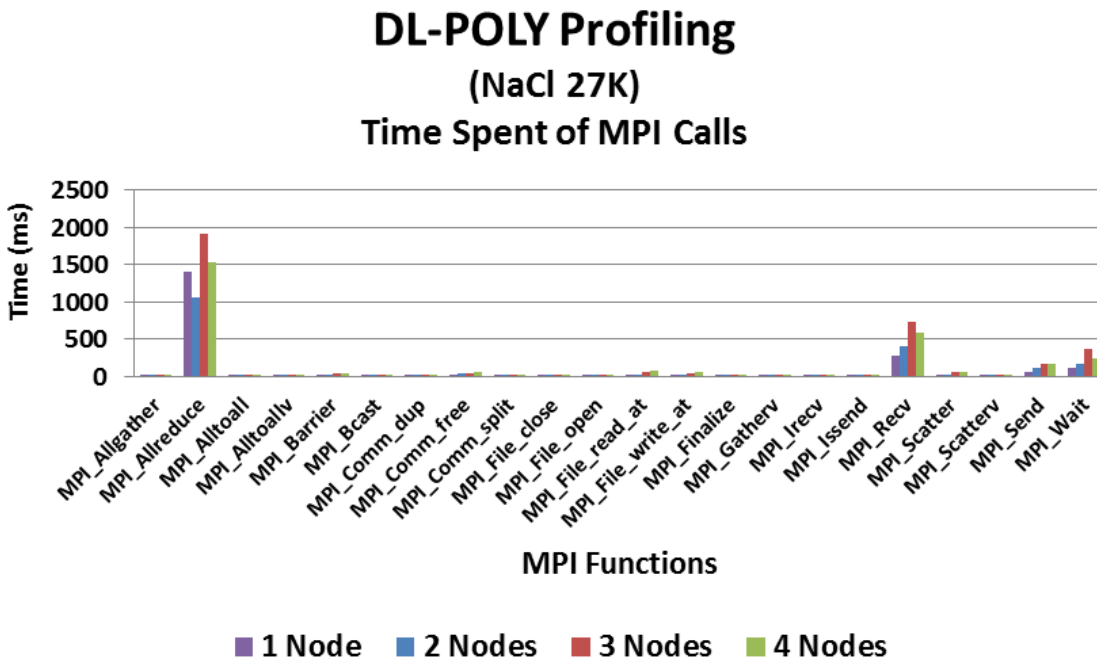
- **Mostl used MPI functions**
  - MPI_Allreduce (27%) and MPI_Wait(24%), MPI_Irecv (17%), MPI_Send (17%)

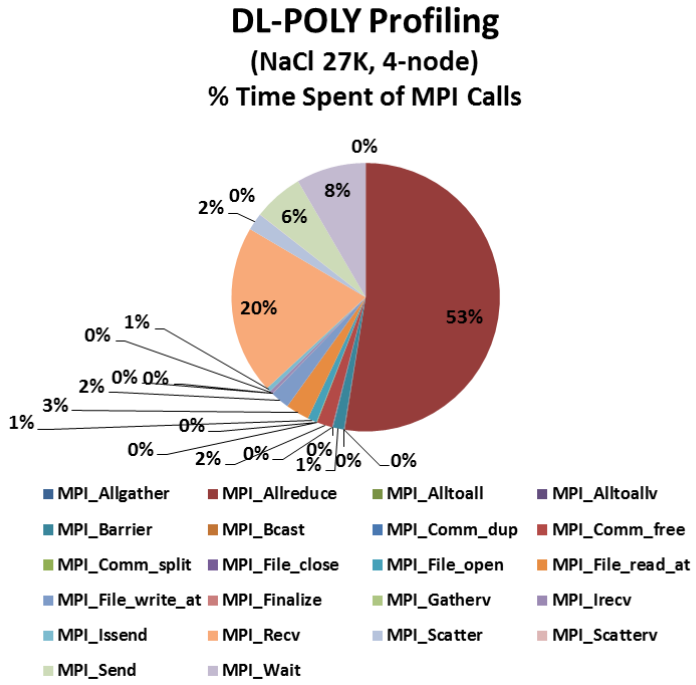




DL-POLY Profiling – MPI Functions

- **The most time consuming MPI functions:**
  - MPI_Allreduce (53%), MPI_Recv(17%), MPI_Wait (8%)





DL-POLY Summary

- **HP ProLiant Gen8 servers delivers better DL-POLY Performance than its predecessor**
  - ProLiant Gen8 equipped with Intel E5 series processes and InfiniBand FDR
  - Provides 158% higher performance than the ProLiant G7 servers when compare at 4 nodes
- **InfiniBand FDR is the most efficient inter-node communication for DL-POLY**
  - Outperforms 1GbE by 385% (or by over 3x) at 4 nodes
  - Outperforms 10GbE by 343% at 4 nodes
  - Outperforms 40GbE by 47% at 4 nodes
- **DL-POLY MPI Profiling**
  - Heavy MPI communications are seen between MPI processes
  - InfiniBand FDR reduces communication time; leave more time for computation
    - InfiniBand FDR consumes 43% of total time, versus 65% 40GbE, versus 87% 1GbE and 10GbE
  - Non-blocking communications are seen:
    - Time spent: MPI_Allreduce (53%), MPI_Recv(17%), MPI_Wait (8%)
    - Most used: MPI_Allreduce (27%) and MPI_Wait(24%), MPI_Irecv (17%), MPI_Send (17%)

Thank You

HPC Advisory Council

All trademarks are property of their respective owners. All information is provided "As-Is" without any kind of warranty. The HPC Advisory Council makes no representation to the accuracy and completeness of the information contained herein. HPC Advisory Council Mellanox undertakes no duty and assumes no obligation to update or correct any information presented herein