HPC ADVISORY COUNCIL


# OpenFOAM Performance Benchmark and Profiling

October 2012

hp

Mellanox TECHNOLOGIES

OpenFOAM

# Note

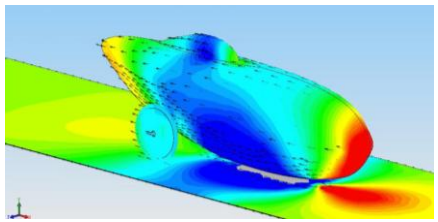
- **The following research was performed under the HPC Advisory Council activities**
  - Special thanks for: HP, Mellanox

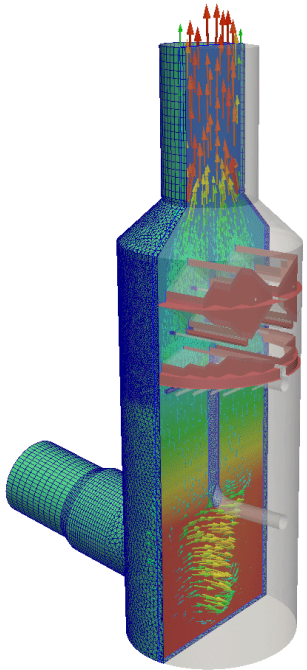

- **For more information on the supporting vendors solutions please refer to:**
  - www.mellanox.com, http://www.hp.com/go/hpc

- **For more information on the application:**
  - http://www.openfoam.com/

# OpenFOAM Applications

- **OpenFOAM® (Open Field Operation and Manipulation) CFD Toolbox in an open source CFD applications that can simulate**
  - Complex fluid flows involving
    - Chemical reactions
    - Turbulence
    - Heat transfer
  - Solid dynamics
  - Electromagnetics
  - The pricing of financial options





- **OpenFOAM support can be obtained from OpenCFD Ltd**

# Objectives

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

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

# Test Cluster Configuration

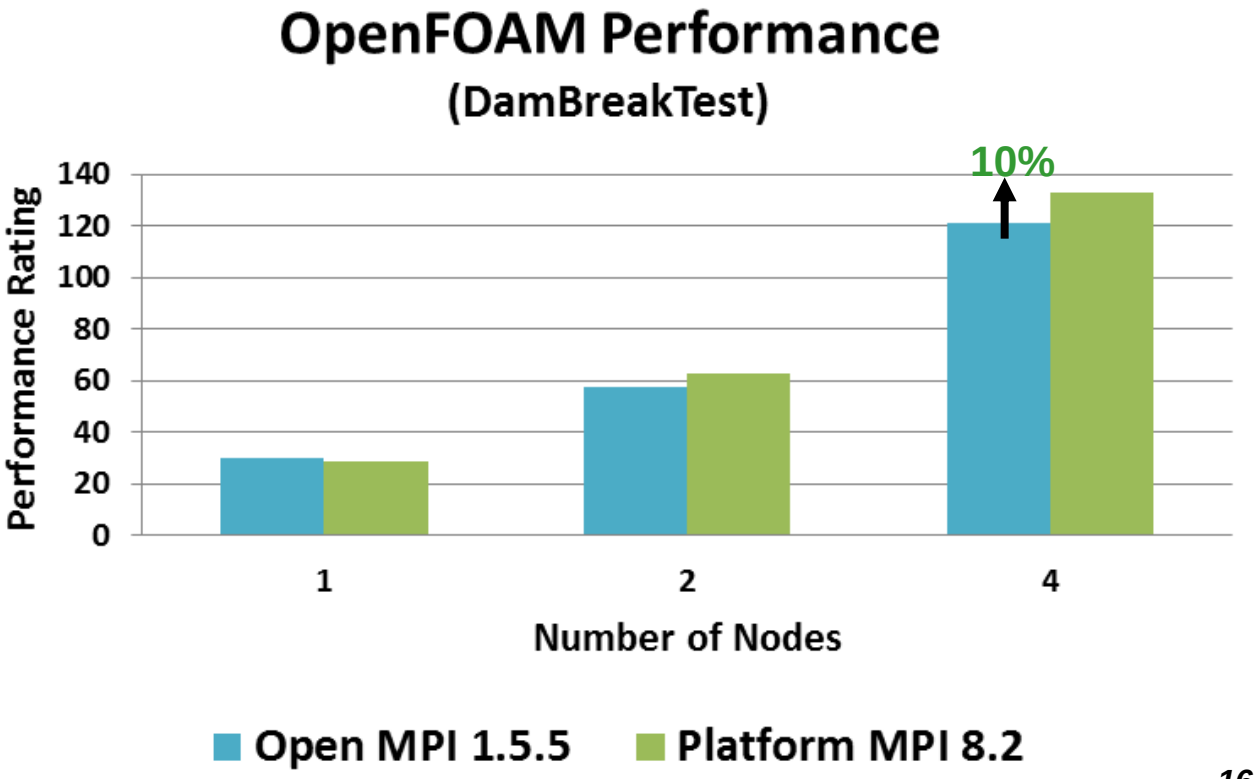
- **HP ProLiant SL230s Gen8 4-node "Athena" cluster**
  - Processors: Dual Eight-Core Intel Xeon E5-2680 @ 2.7 GHz
  - Memory: 32GB per node, 1600MHz DDR3 DIMMs
  - OS: RHEL 6 Update 2, OFED 1.5.3 InfiniBand SW stack
- **Mellanox ConnectX-3 VPI InfiniBand adapters**
- **Mellanox SwitchX SX6036 56Gb/s InfiniBand and 40G/s Ethernet Switch**
- **MPI: Open MPI 1.5.5, Platform MPI 8.2**
- **Application: OpenFOAM 2.1.0 (double precision)**
- **Benchmark Workload:**
  - damBreakTest (using interFoam solver)

# 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<br>Four SFF non-hot plug SAS, SATA, SSD bays<br>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<br>1Gb and 10Gb Ethernet, IB, and FlexFabric 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<br>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 |

# OpenFOAM Performance - MPI

- **Platform MPI outperforms Open MPI when running at scale**
  - Up to 10% faster runtime achieved at 4 nodes
  - No optimization flags were used

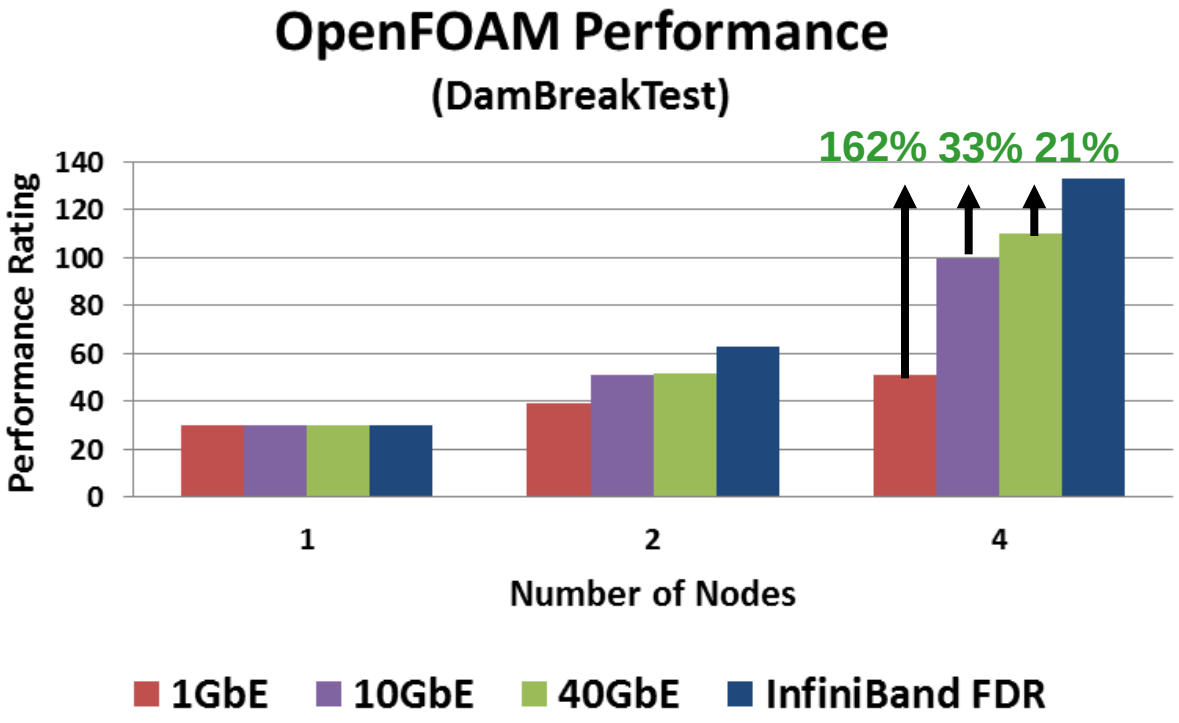



*Higher is better*

*16 Processes/Node*

# OpenFOAM Performance - Interconnect

- **InfiniBand FDR provides the best inter-node communication for OpenFOAM**
  - Outperforms 1GbE by 162% at 4 nodes
  - Outperforms 10GbE by 33% at 4 nodes
  - Outperforms 40GbE by 21% at 4 nodes
- **1GbE shows little performance gain beyond 2 nodes**

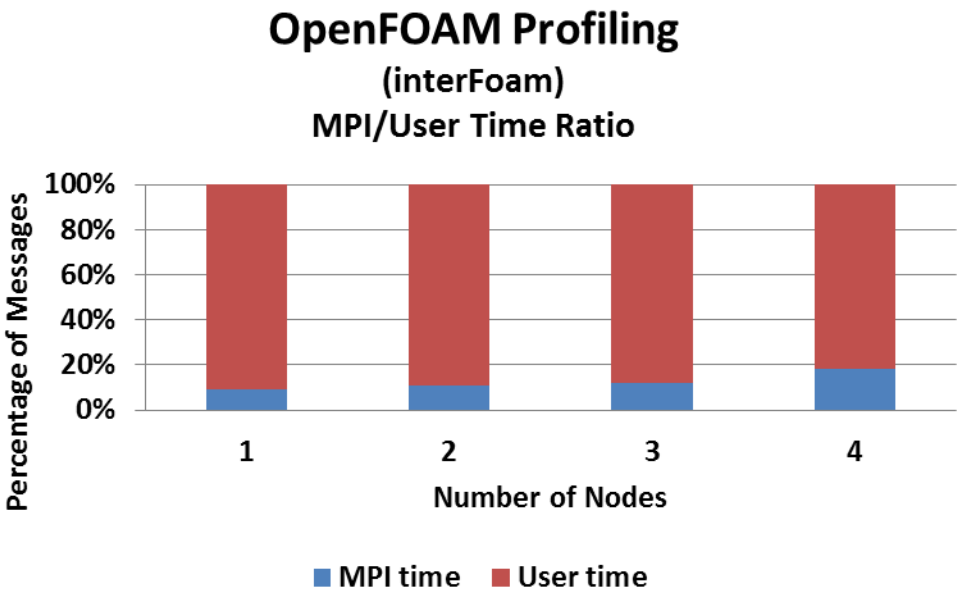



*Higher is better*

*16 Processes/Node*

# OpenFOAM Profiling – MPI Time Ratio

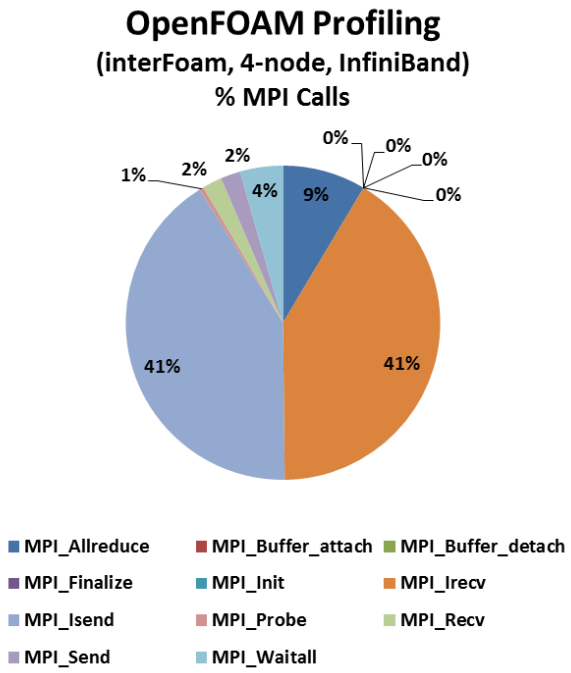
- **InfiniBand FDR reduces the communication time at scale**
  - InfiniBand FDR consumes about 20% of total runtime at 4 nodes

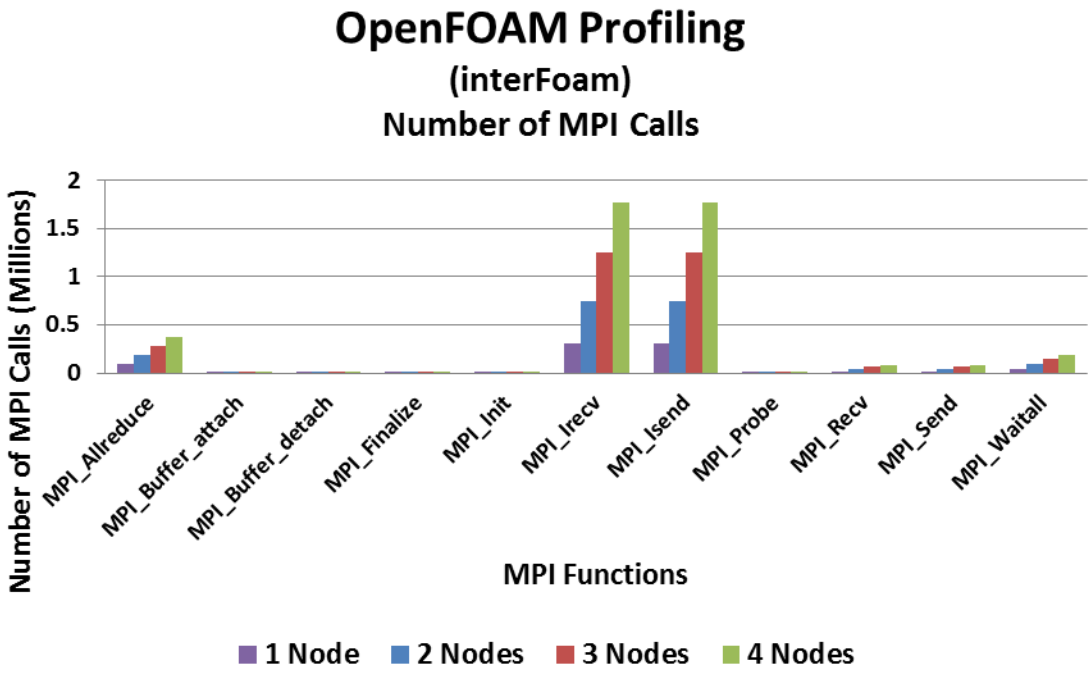



***16 Processes/Node***

# OpenFOAM Profiling – MPI Functions

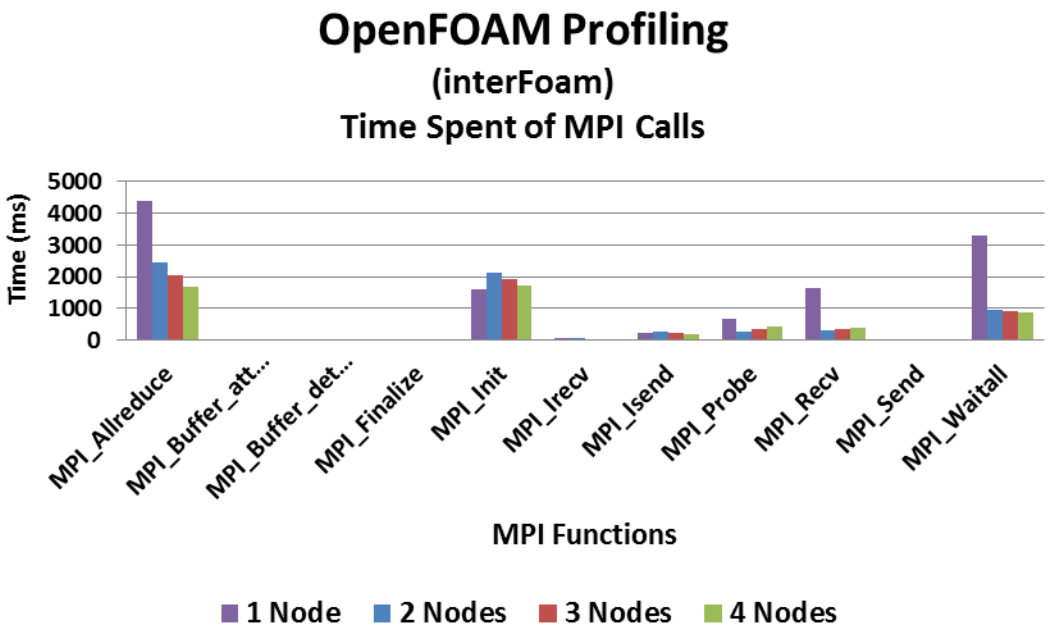
- **Mostly used MPI functions**
  - MPI_Irecv (41%) and MPI_Isend (41%), MPI_Waitall (9%), MPI_Allreduce (4%)

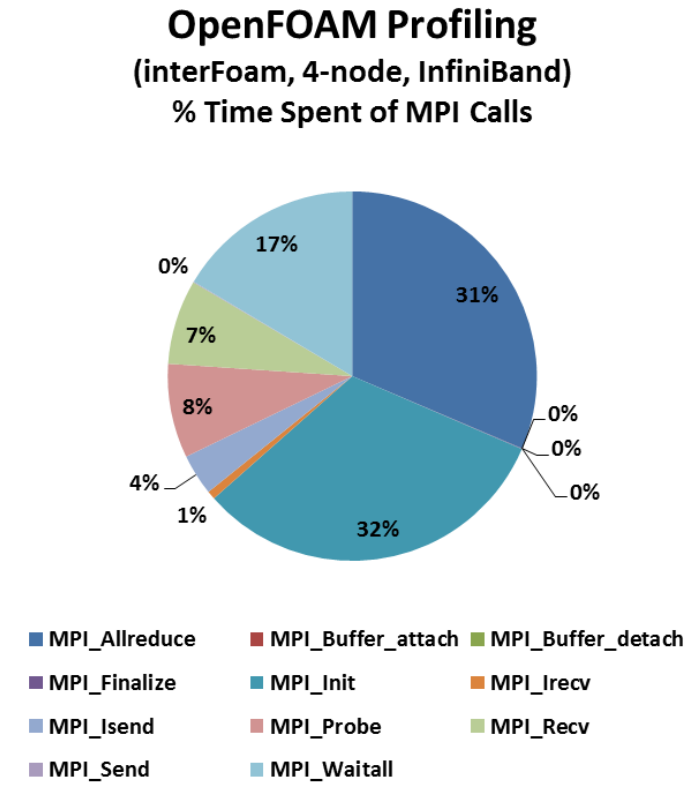


# OpenFOAM Profiling – MPI Functions

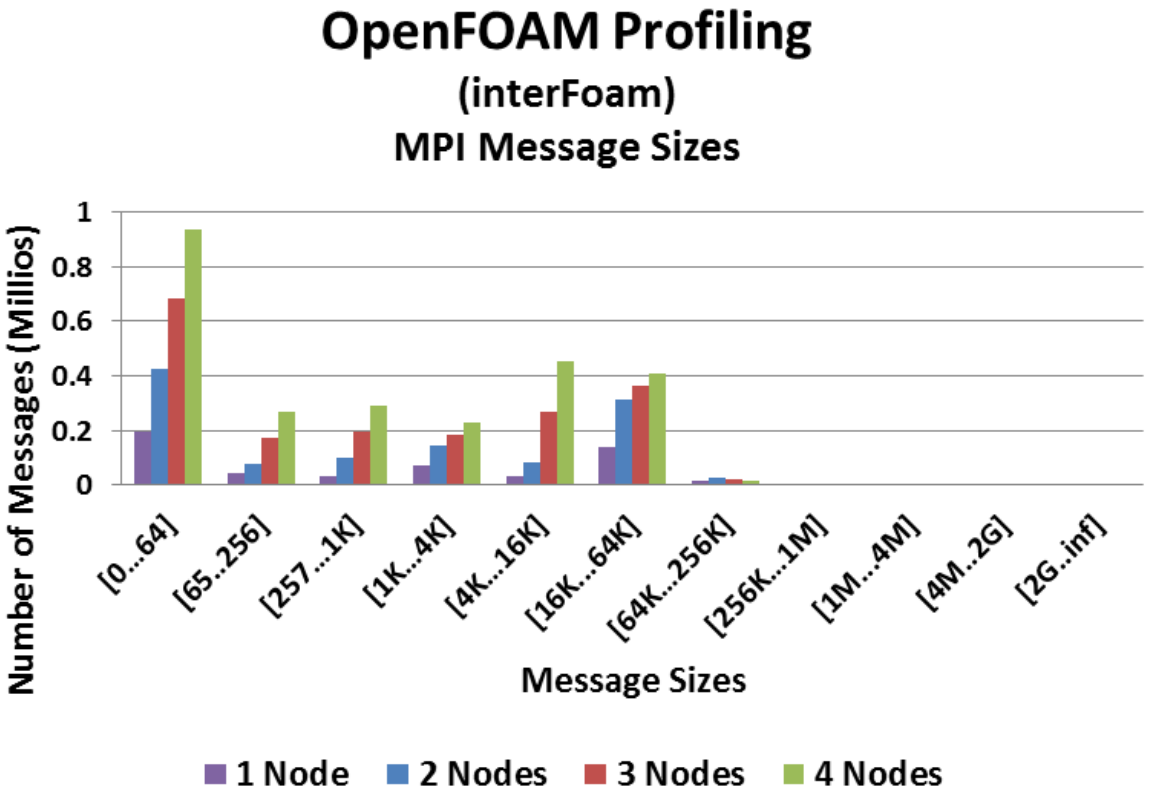
- **The most time consuming MPI functions:**
  - MPI_Init (32%) MPI_Allreduce (31%), MPI_Waitall (17%)



# OpenFOAM Profiling – Message Size

- **Distribution of message sizes for the MPI calls**
  - Peak between 0B to 64KB
  - Large concentration in the small and the medium message sizes



# OpenFOAM Summary

- **OpenFOAM performance benchmark demonstrates**
  - InifiniBand FDR delivers higher application performance and linear scalability
    - Outperforms 1GbE by 162%, 10GbE by 33%, 40GbE by 21% at 4 nodes
  - Platform MPI tuning can boost application performance by 10% over Open MPI at 4 nodes
- **MPI profiling on the OpenFOAM interFoam solver**
  - Message send/recv creates big communication overhead
  - Most are small message used by OpenFOAM
  - Collectives (MPI_Alltoall) overhead increases as cluster size scales up
  - Heavy MPI communications are seen between MPI processes
  - InfiniBand FDR reduces communication time; leave more time for computation
    - InfiniBand FDR consumes 20% of total time
  - Non-blocking communications are seen:
    - MPI_Irecv (41%) and MPI_Isend (41%), MPI_Waitall (9%), MPI_Allreduce (4%)

# Thank You

## HPC Advisory Council