

ANSYS FLUENT 13

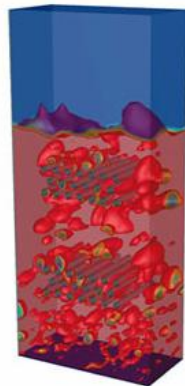
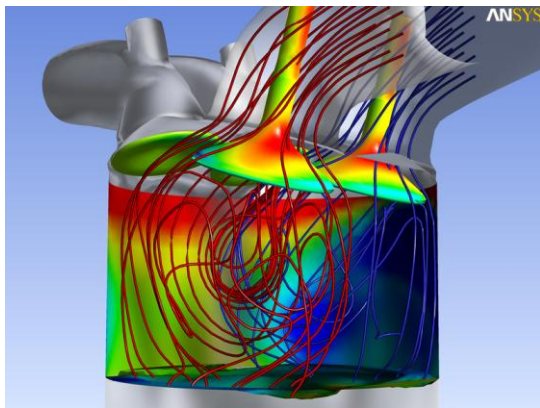
Performance Benchmark and Profiling

April 2011



- **The following research was performed under the HPC Advisory Council activities**
 - Participating vendors: AMD, Dell, Mellanox
 - Compute resource - HPC Advisory Council Cluster Center
- **For more info please refer to**
 - [http:// www.amd.com](http://www.amd.com)
 - [http:// www.dell.com/hpc](http://www.dell.com/hpc)
 - <http://www.mellanox.com>
 - <http://www.ansys.com>

- **Computational Fluid Dynamics (CFD) is a computational technology**
 - Enables the study of the dynamics of things that flow
 - By generating numerical solutions to a system of partial differential equations which describe fluid flow
 - Enable better understanding of qualitative and quantitative physical phenomena in the flow which is used to improve engineering design
- **CFD brings together a number of different disciplines**
 - Fluid dynamics, mathematical theory of partial differential systems, computational geometry, numerical analysis, Computer science
- **ANSYS FLUENT is a leading CFD application from ANSYS**
 - Widely used in almost every industry sector and manufactured product



- **The following was done to provide best practices**
 - ANSYS FLUENT performance benchmarking
 - Interconnect performance comparisons
 - CPU performance
 - Understanding FLUENT communication patterns
 - Ways to increase FLUENT productivity
 - MPI libraries comparisons

- **The presented results will demonstrate**
 - The scalability of the compute environment
 - The capability of FLUENT to achieve scalable productivity
 - Considerations for performance optimizations

- **Dell™ PowerEdge™ R815 11-node (528-core) cluster**
- **AMD™ Opteron™ 6174 (code name “Magny-Cours”) 12-cores @ 2.2 GHz CPUs**
- **4 CPU sockets per server node**
- **Mellanox ConnectX-2 VPI adapters for 40Gb/s QDR InfiniBand and 10Gb/s Ethernet**
- **Mellanox MTS3600Q 36-Port 40Gb/s QDR InfiniBand switch**
- **Fulcrum based 10Gb/s Ethernet switch**
- **Memory: 128GB memory per node DDR3 1333MHz**
- **OS: RHEL 5.5, MLNX-OFED 1.5.2 InfiniBand SW stack**
- **MPI: Platform MPI 7.1**
- **Application: ANSYS FLUENT version 13.0.0**
- **Benchmark workload:**
 - sedan_4m (External Aerodynamics Flow Over a Passenger Sedan)
 - truck_poly_14m (External Flow Over a Truck Body with a Polyhedral Mesh)

- **HPC Advisory Council Test-bed System**
- **New 11-node 528 core cluster - featuring Dell PowerEdge™ R815 servers**
 - Replacement system for Dell PowerEdge SC1435 (192 cores) cluster system following 2 years of rigorous benchmarking and product EOL
 - System to be redirected to explore HPC in the Cloud applications
- **Workload profiling and benchmarking**
 - Characterization for HPC and compute intense environments
 - Optimization for scale, sizing and configuration and workload performance
 - Test-bed Benchmarks
 - RFPs
 - Customers/Prospects, etc
 - ISV & Industry standard application characterization
 - Best practices & usage analysis



About Dell PowerEdge™ Platform Advantages

Best of breed technologies and partners

Combination of AMD™ Opteron™ 6100 series platform and Mellanox ConnectX InfiniBand on Dell HPC

Solutions provide the ultimate platform for speed and scale

- Dell PowerEdge R815 system delivers 4 socket performance in dense 2U form factor
- Up to 48 core/32DIMMs per server – 1008 core in 42U enclosure

Integrated stacks designed to deliver the best price/performance/watt

- 2x more memory and processing power in half of the space
- Energy optimized low flow fans, improved power supplies and dual SD modules

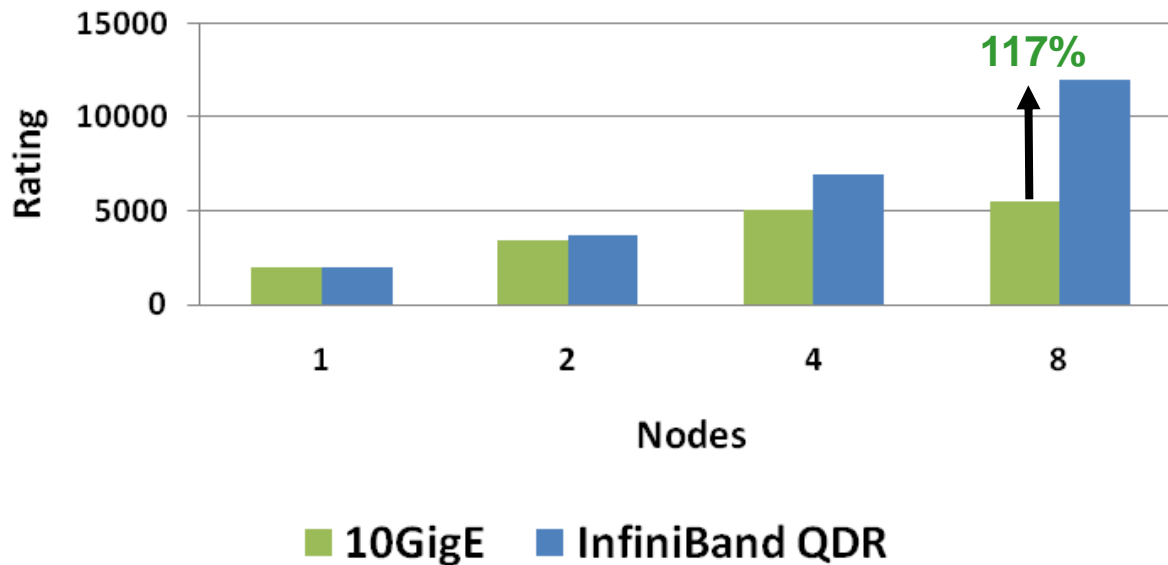
Optimized for long-term capital and operating investment protection

- System expansion
- Component upgrades and feature releases



- **Dataset: sedan_4m**
 - External Flow Over a Passenger Sedan
 - 3.6 million cells of mixed type, k-epsilon model, pressure-based coupled solver
- **InfiniBand shows continuous gain as the cluster scales**
 - Up to 117% higher performance than 10GigE

FLUENT Benchmark (sedan_4m)

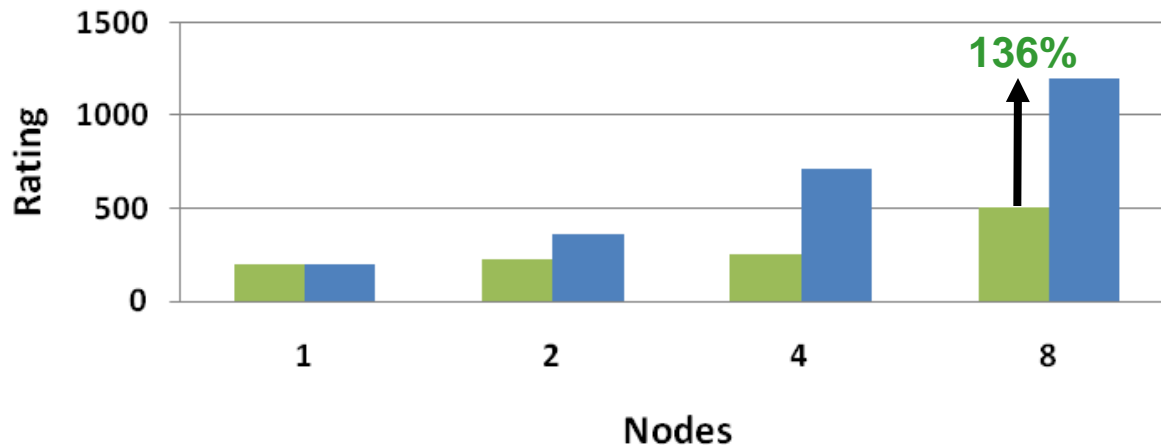


Higher is better

48 Cores/Node

- **Dataset: truck_poly_14m**
 - External Flow Over a Truck Body with a Polyhedral Mesh
 - 14 million polyhedral cells, DES model with the segregated implicit solver
- **InfiniBand shows continuous gain as the cluster scales**
 - Up to 136% higher performance than 10GigE

FLUENT Benchmark (truck_poly_14m)



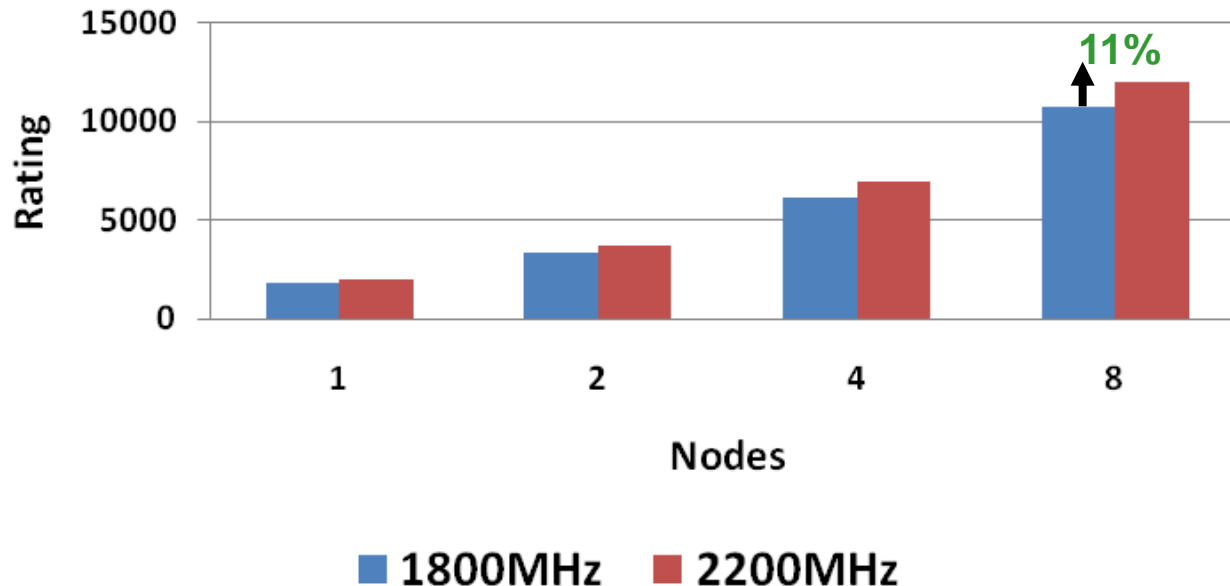
Higher is better

■ 10GigE ■ InfiniBand QDR

48 Cores/Node

- **Increasing CPU core frequency enables higher job efficiency**
 - Up to 11% better job performance between 2200MHz vs 1800MHz on 8-node
 - Delivers a gain of 10-13% on average in better job performance

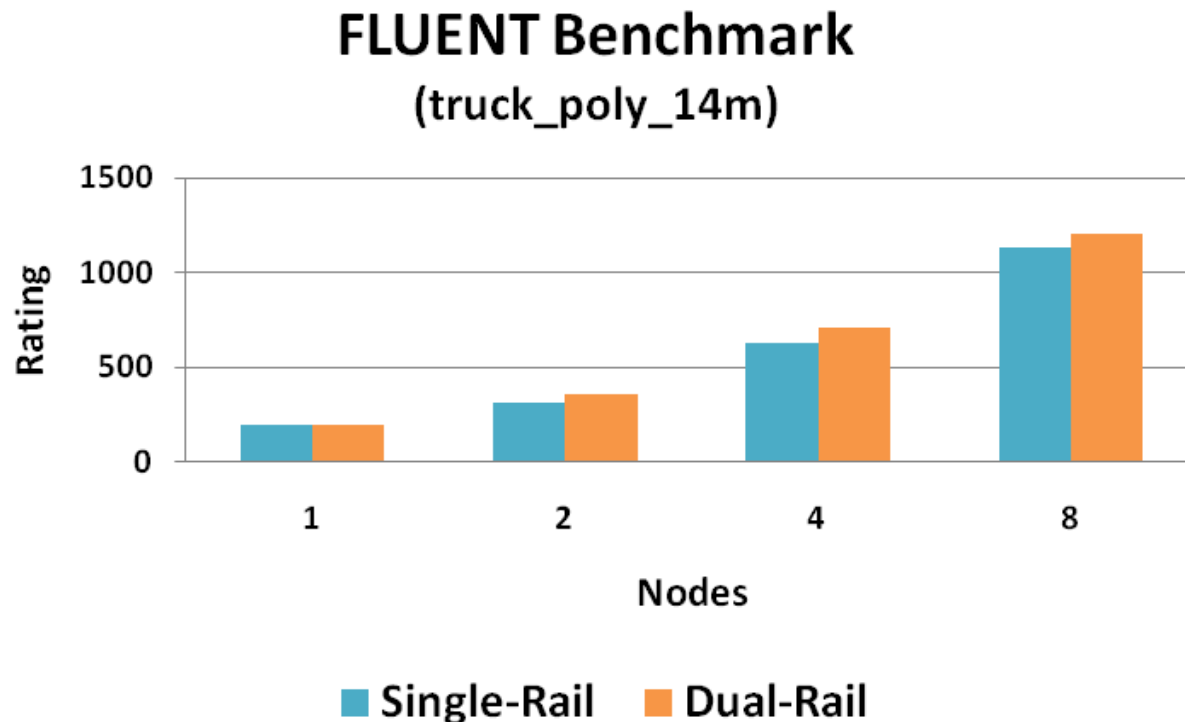
FLUENT Benchmark (sedan_4m)



Higher is better

48 Cores/Node

- **Dual-rail (Dual InfiniBand cards) enables better performance than single-rail**
 - Up to 15% better job performance when equipped with 2 InfiniBand cards per node
 - Delivers network bandwidth that requires for data communications

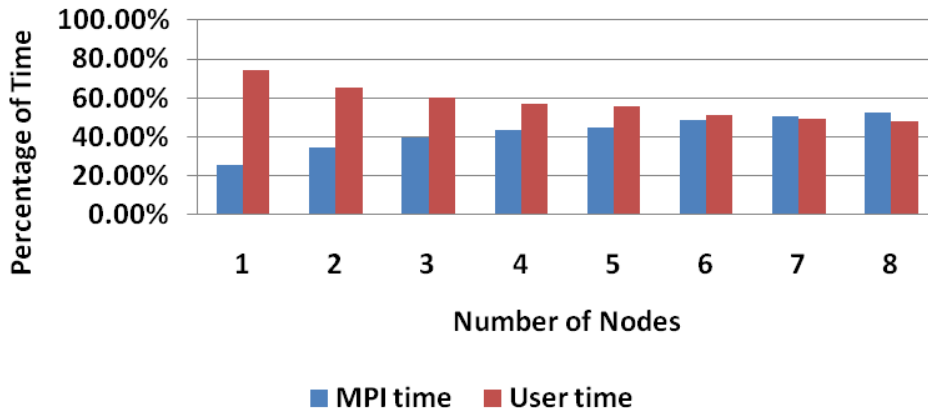


Higher is better

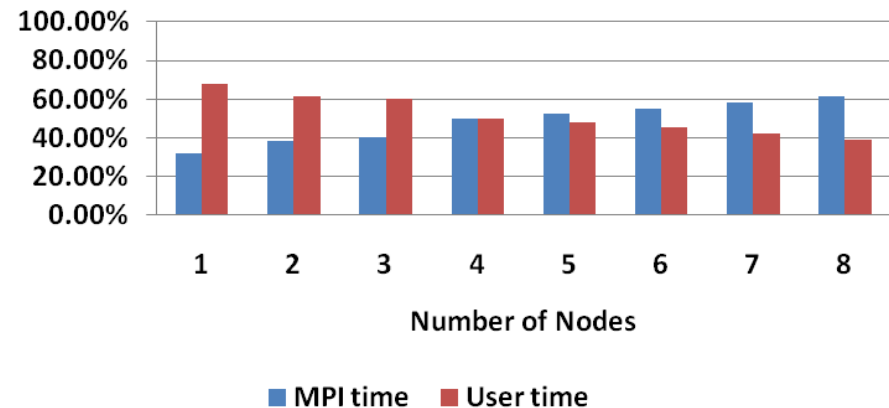
48 Cores/Node

- **Gradual increase in communications time as the cluster scales**
 - More time is spent on communications than computation after 6-node in sedan_4m
 - More time is spent on communications than computation after 4-node in truck_poly_14m

FLUENT Profiling
(sedan_4m)
MPI/User Time Ratio



FLUENT Profiling
(truck_poly_14m)
MPI/User Time Ratio

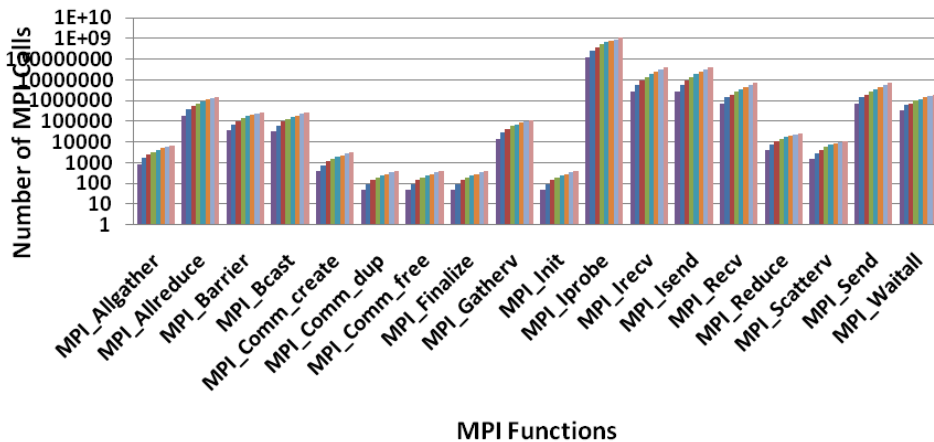


48 Cores/Node

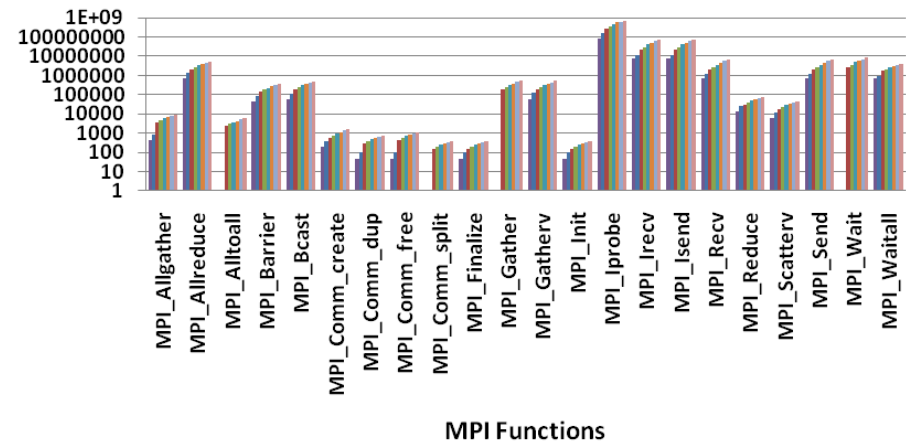
FLUENT Profiling – Number of MPI Calls

- **The most used MPI function is MPI_Iprobe**
 - MPI_Iprobe does non-blocking test for a message
 - Represents 92% of MPI calls used for 8-node in sedan_4m, 82% in truck_poly_14m
- **FLUENT uses a full range of MPI calls**
 - For blocking, non-blocking and point-to-point and collective communications
- **Data communications increases for larger dataset**
 - MPI_Irecv and MPI_Isend at a higher rate for truck_poly_14m

FLUENT Profiling
(sedan_4m)
Number of MPI Calls



Fluent Profiling
(truck_poly_14m)
Number of MPI Calls



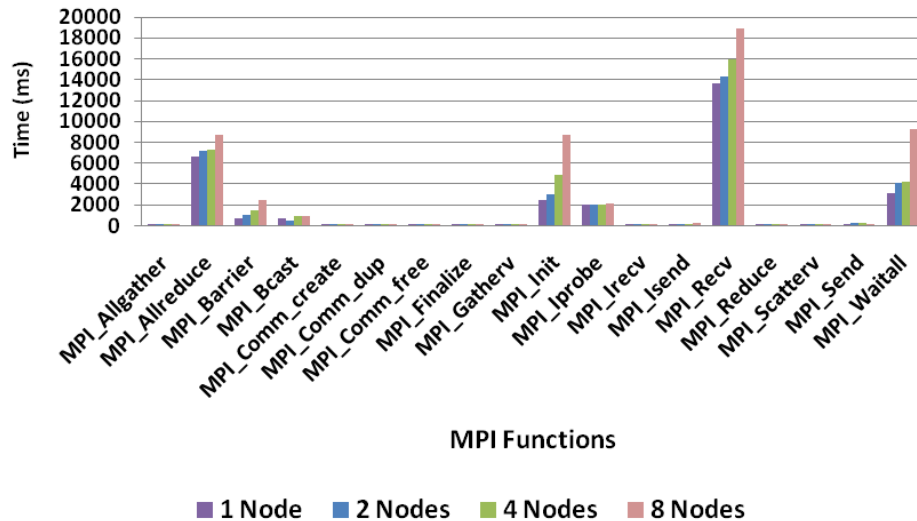
■ 1 Node ■ 2 Nodes ■ 3 Nodes ■ 4 Nodes ■ 5 Nodes ■ 6 Nodes ■ 7 Nodes ■ 8 Nodes

■ 1 Node ■ 2 Nodes ■ 3 Nodes ■ 4 Nodes ■ 5 Nodes ■ 6 Nodes ■ 7 Nodes ■ 8 Nodes

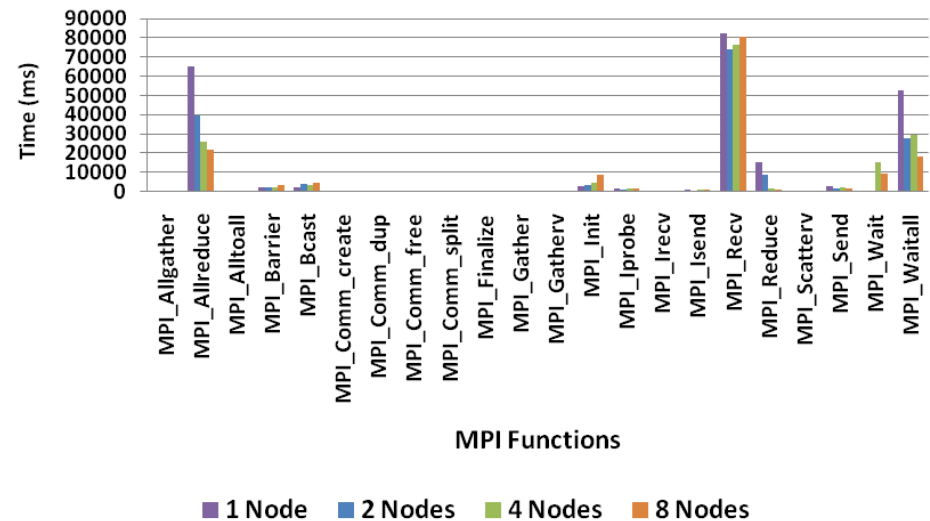
FLUENT Profiling – Time Spent of MPI Calls

- **The largest time consumer is MPI_Recv for data communications**
 - Occupies 37% of all MPI time for 8 node in sedan_4m
 - Occupies 53% of all MPI time for 8 node in truck_poly_14m
- **The next largest time consumer are MPI_Waitall and MPI_Allreduce**
 - MPI_Allreduce(17%) and MPI_Waitall(18%) for 8 node in sedan_4m
 - MPI_Waitall(14%) and MPI_Allreduce(12%) for 8 node in truck_poly_14m
- **More time spent on data MPI communication than MPI synchronization**

FLUENT Profiling
(sedan_4m)
Time Spent of MPI Calls

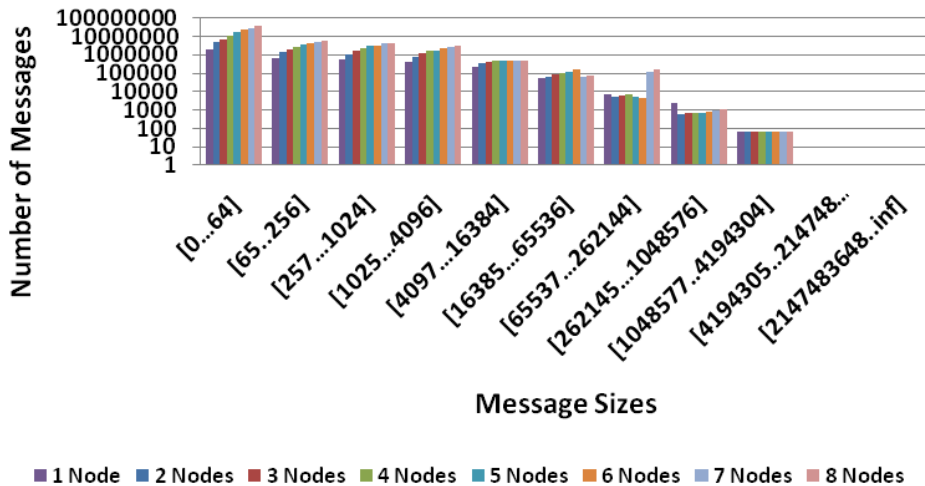


FLUENT Profiling
(truck_poly_14m)
Time Spent of MPI Calls

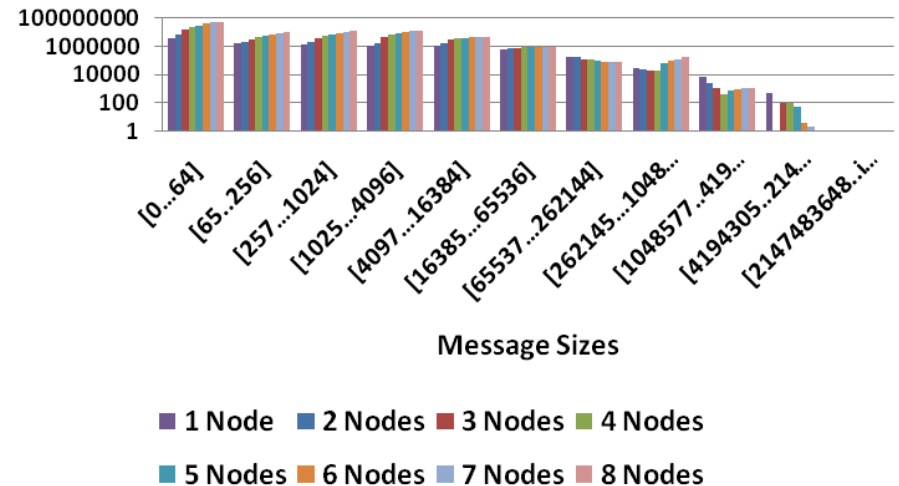


- **MPI message sizes are concentrated in range of small message sizes**
 - Majority are in the range of 0B and 64B
 - Small messages are typical used for synchronization, implies FLUENT is latency sensitive
- **Larger message sizes also appeared but at a smaller percentage**
 - Larger messages (65B to 4MB) responsible for data transfers between the MPI ranks
 - Implies that FLUENT also requires high network throughput

FLUENT Profiling
(sedan_4m)
MPI Message Sizes



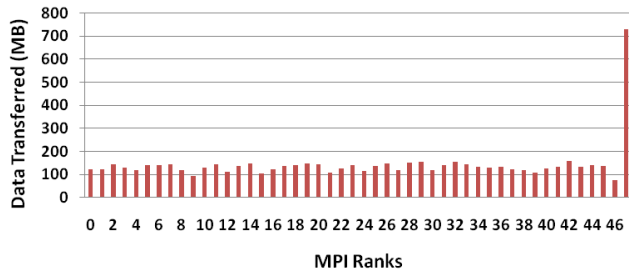
Fluent Profiling
(truck_poly_14m)
MPI Message Sizes



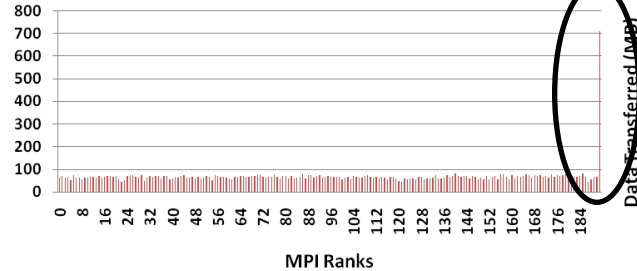
FLUENT Profiling – Data Transfer Per Process

- **Data transferred to each MPI rank is generally the same except for the last**
 - Around 450MB per MPI rank for truck_poly_14m, and 100MB for sedan_4m,
 - The last MPI rank has a significantly higher data rate than the rest
- **As the cluster scales, data transfers remains generally to the same level**

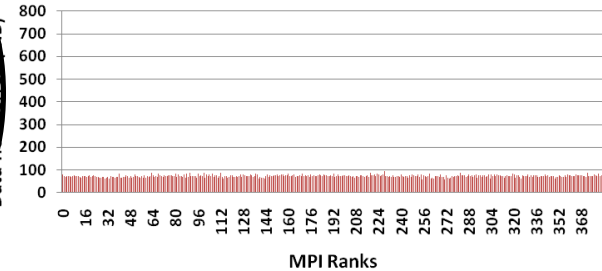
Fluent Profiling
(sedan_4m, 1-node)
Data Transferred by Ranks



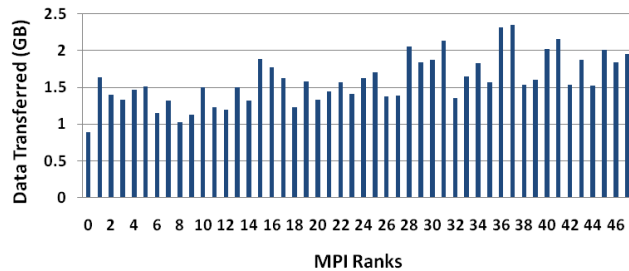
Fluent Profiling
(sedan_4m, 4-node)
Data Transferred by Ranks



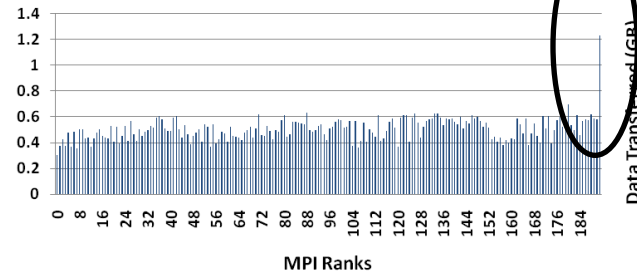
Fluent Profiling
(sedan_4m, 8-node)
Data Transferred by Ranks



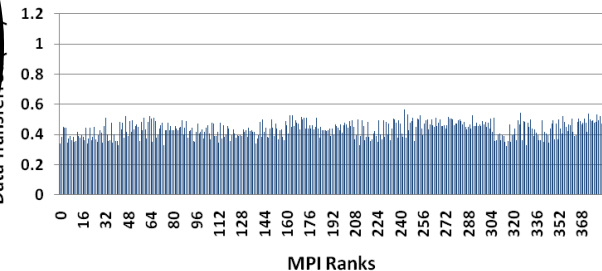
Fluent Profiling
(truck_poly_14m, 1-node)
Data Transferred by Ranks



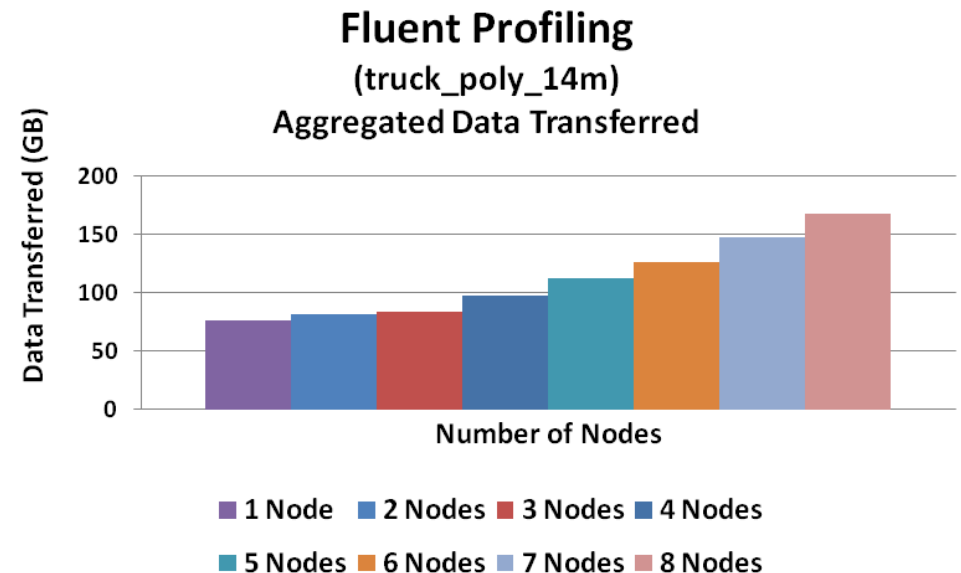
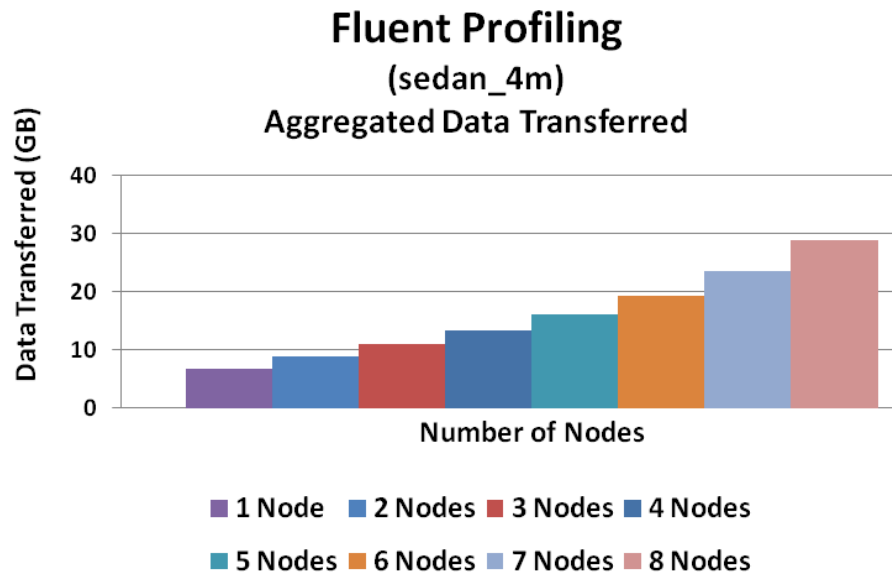
Fluent Profiling
(truck_poly_14m, 4-node)
Data Transferred by Ranks



Fluent Profiling
(truck_poly_14m, 8-node)
Data Transferred by Ranks



- **Aggregated data transfer refers to:**
 - Total amount of data being transferred in the network between all MPI ranks collectively
- **The total data transfer steadily increases as the cluster scales**
 - As a compute node being added, more data communications will happen
- **Significantly more communications happen for larger dataset**



InfiniBand QDR

- **FLUENT is a leading CFD application from ANSYS**
- **Networking**
 - InfiniBand QDR allows FLUENT to scale as it provides low latency and high throughput
 - Dual rail (two adapters) can increase the performance by 15% on a 4 socket server
- **CPU**
 - Shows gains in job productivity by using higher CPU frequency
- **Data transfer on the network**
 - Significantly more data being transferred for the larger dataset
 - Tends to increase steadily as cluster scales
- **MPI**
 - Shows FLUENT uses a range of MPI API for communications and synchronizations

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