

WRF

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.wrf-model.org/>

- **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: Open MPI 1.4.3**
- **Compilers: Intel Compilers 10.4**
- **Application: WRF 3.0.1**
- **Benchmark workload**
 - 12km CONUS benchmark case

- **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

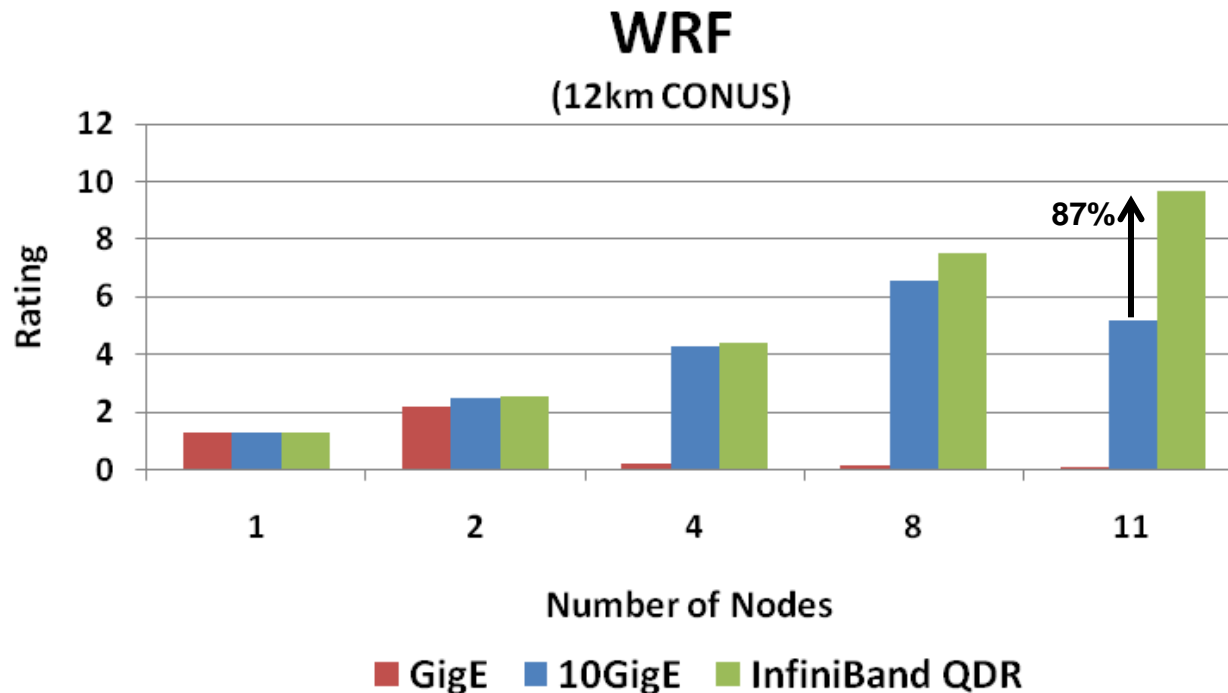
Optimized for long-term capital and operating investment protection

- System expansion
- Component upgrades and feature releases



WRF Performance - Interconnects

- **InfiniBand shows continuous gain as the cluster scales**
 - Up to 87% faster than 10GigE at 11 nodes
- **10GigE stops scaling after 8 nodes**
- **GigE scales only up to 2 nodes**

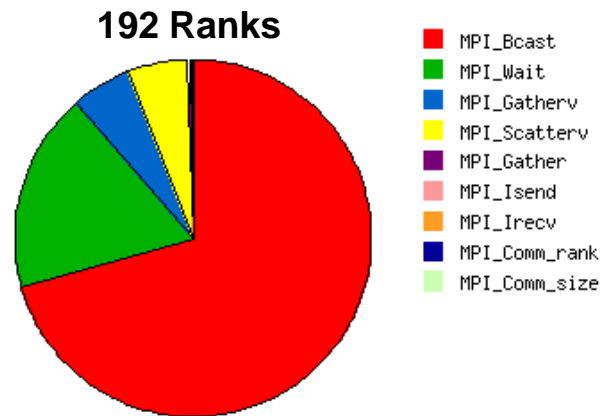


Higher is better

48 Cores/Node

WRF MPI Profiling – MPI Functions

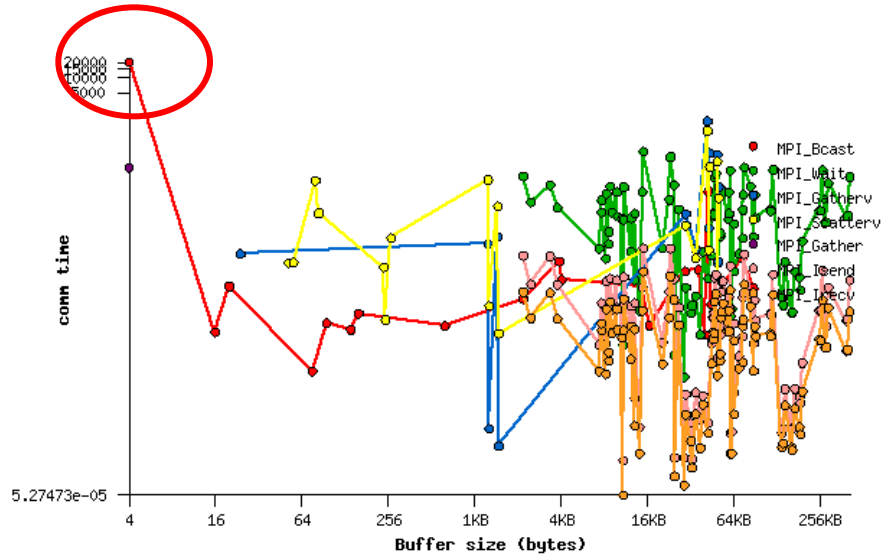
- **MPI_Bcast dominates MPI communication time**
 - 75% with 528 MPI ranks



InfiniBand QDR

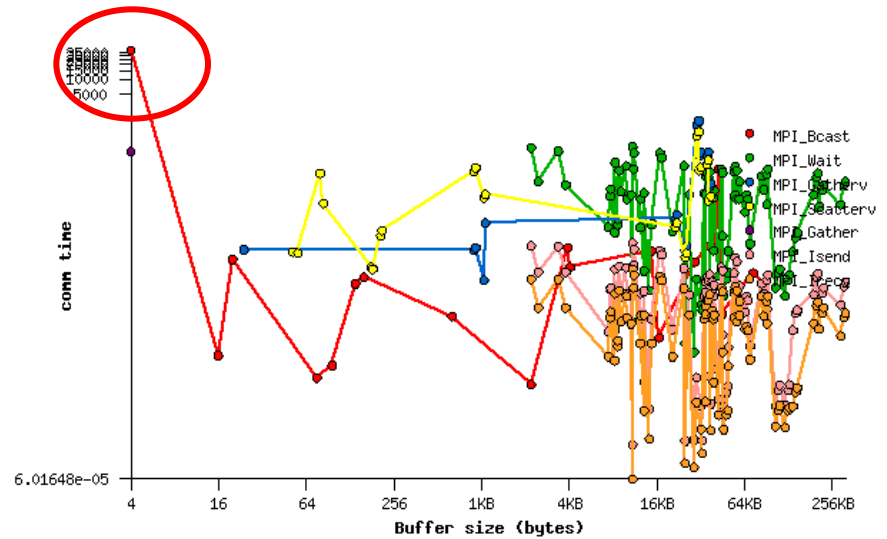
WRF MPI Profiling – Message Size

- **MPI_Bcast uses small size messages (4 Byte)**



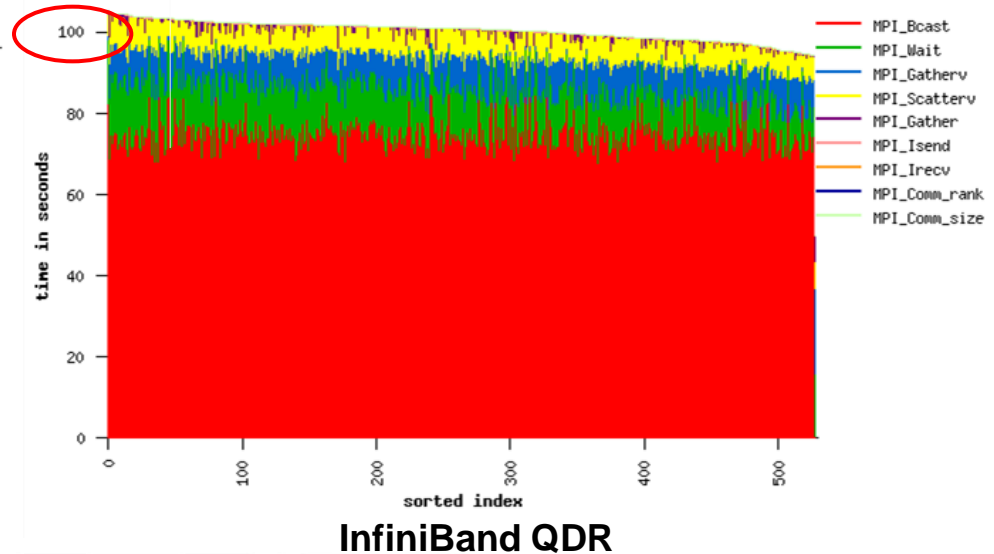
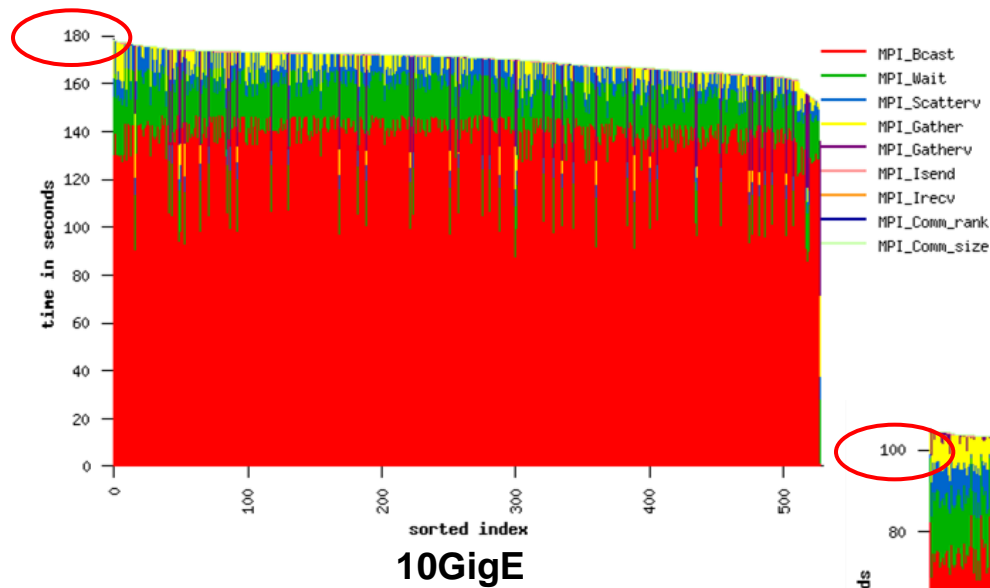
384 Ranks

528 Ranks



WRF MPI Profiling – 10GigE vs InfiniBand

- **10GigE consumes much higher communication time versus IB QDR**
 - Nearly double at 11 nodes



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