Desmond Performance Benchmark and Profiling

August 2009
Note

• The following research was performed under the HPC Advisory Council activities
  – Participating vendors: Intel, Dell, Mellanox
  – Compute resource - HPC Advisory Council Cluster Center
• For more info please refer to
Desmond is a molecular dynamics (MD) code
– Compute energies and forces for many standard fixed-charged force fields used in biomolecular simulation
– Compatible with polarizable force fields based on the Drude formalism

Desmond software includes
– Tools for minimization and energy analysis
– Methods for restraining atomic positions as well as molecular configurations
– Support for a variety of periodic cell configurations
– Facilities for accurate checkpointing and restart

Desmond performs high-speed molecular dynamics simulations of biological systems on conventional commodity clusters.
– Developed at D. E. Shaw Research
Objectives

• The presented research was done to provide best practices
  – Desmond performance benchmarking
  – Interconnect performance comparisons
  – Power-efficient simulations
Test Cluster Configuration

- Dell™ PowerEdge™ M610 16-node cluster
- Quad-Core Intel X5570 @ 2.93 GHz CPUs
- Intel Cluster Ready certified cluster
- Mellanox ConnectX MCQH29-XCC 4X QDR InfiniBand mezzanine card
- Mellanox M3601Q 32-Port Quad Data Rate (QDR-40Gb) InfiniBand Switch
- Memory: 24GB memory per node
- OS: RHEL5U3, OFED 1.4 InfiniBand SW stack
- MPI: Open MPI 1.3.2
- Application: Desmond
- Benchmarks: DHFR
  - Dihydrofolate reductase system with 23,558 atoms
Mellanox InfiniBand Solutions

- **Industry Standard**
  - Hardware, software, cabling, management
  - Design for clustering and storage interconnect
- **Performance**
  - 40Gb/s node-to-node
  - 120Gb/s switch-to-switch
  - 1us application latency
  - Most aggressive roadmap in the industry
- **Reliable with congestion management**
- **Efficient**
  - RDMA and Transport Offload
  - Kernel bypass
  - CPU focuses on application processing
- **Scalable for Petascale computing & beyond**
- **End-to-end quality of service**
- **Virtualization acceleration**
- **I/O consolidation Including storage**

The InfiniBand Performance Gap is Increasing

InfiniBand Delivers the Lowest Latency
Delivering Intelligent Performance
Next Generation Intel® Microarchitecture

Bandwidth Intensive
• Intel® QuickPath Technology
• Integrated Memory Controller

Threaded Applications
• 45nm quad-core Intel® Xeon® Processors
• Intel® Hyper-threading Technology

Performance on Demand
• Intel® Turbo Boost Technology
• Intel® Intelligent Power Technology

Performance That Adapts to The Software Environment
System Structure and Sizing Guidelines
- 16-node cluster build with Dell PowerEdge™ M610 blades server
- Servers optimized for High Performance Computing environments
- Building Block Foundations for best price/performance and performance/watt

Dell HPC Solutions
- Scalable Architectures for High Performance and Productivity
- Dell's comprehensive HPC services help manage the lifecycle requirements.
- Integrated, Tested and Validated Architectures

Workload Modeling
- Optimized System Size, Configuration and Workloads
- Test-bed Benchmarks
- ISV Applications Characterization
- Best Practices & Usage Analysis
Desmond Benchmark Results

- Input Dataset - DHFR
- InfiniBand QDR enables best performance and scalability
  - Up to 540% higher productivity versus Gigabit Ethernet
  - Up to 11% higher productivity versus InfiniBand DDR

**Desmond Benchmark Result**

(DHFR)

*Higher is better*
Power Cost Savings with Different Interconnect

- To finish the same number of Desmond jobs with IB QDR or GigE
  - InfiniBand QDR saves up to $9560 power
  - Yearly based for 16-node cluster
- As cluster size increases, more power can be saved

Desmond Power Consumption
(16 nodes)

$\text{$/KWh} = \text{KWh} \times 0.20$

Test Results

- **InfiniBand enables highest Desmond performance and scalability**
  - InfiniBand QDR delivers 540% higher productivity versus GigE

- **InfiniBand QDR enables highest return in investment**
  - Up to 11% higher productivity versus InfiniBand DDR
    - Measured with 16-node system
  - Performance gain increases with system size

- **InfiniBand power savings compared to GigE**
  - InfiniBand QDR enables up to $9560/year savings
## Estimated System Cost

<table>
<thead>
<tr>
<th>InfiniBand 40Gb/s Connected</th>
<th>Ethernet Connected</th>
</tr>
</thead>
<tbody>
<tr>
<td>$95K (blades servers)</td>
<td>Cost (estimation) $80K (blade servers)</td>
</tr>
<tr>
<td>149 jobs/day</td>
<td>Productivity</td>
</tr>
<tr>
<td>Cost per job: $637</td>
<td>23 jobs/day</td>
</tr>
<tr>
<td></td>
<td>Cost per job: $3478</td>
</tr>
<tr>
<td>16 servers provide performance equal to 16 servers capability</td>
<td>Utilization</td>
</tr>
<tr>
<td></td>
<td>16 servers provide performance equal to 2.5 servers capability</td>
</tr>
<tr>
<td>$95K - in actual compute capability</td>
<td>Return on Investment</td>
</tr>
<tr>
<td>$0 - wasted</td>
<td>$12.5K - in actual compute capability</td>
</tr>
<tr>
<td></td>
<td>$67.5K - wasted</td>
</tr>
</tbody>
</table>
Productive Systems = Balanced System

• Balanced system enables highest productivity
  – Interconnect performance to match CPU capabilities
  – CPU capabilities to drive the interconnect capability
  – Memory bandwidth to match CPU performance

• Applications scalability relies on balanced configuration
  – “Bottleneck free”
  – Each system components can reach it’s highest capability

• Dell M610 system integrates balanced components
  – Intel “Nehalem” CPUs and Mellanox InfiniBand QDR
    • Latency to memory and Interconnect latency at the same magnitude of order
  – Provide the leading productivity and power/performance system for Desmond simulations
Thank You
HPC Advisory Council