Applications and their challenges on Blue Waters

Greg Bauer
Overview

- What is running on Blue Waters?
- What are the issues and what to do about them?
  - Scalability
  - Runtime consistency
  - Other job interference
  - IO
  - Congestion Protection
  - Interrupts
Changes to the system

• More XK nodes
  • From 3,072 to 4,224.
• Flattened XK region in torus
  • From 8x8x24 to 15x6x24.
• LNET nodes redistributed across XE and XK
  • Good – Improved aggregate bandwidth within the XK region of the torus (more X links, fewer Y links). LNETs in XK region provide possible (future) co-location of compute and IO.
  • Not so good – LNETs in region (IO was going through XK region anyway). X dimension now greater than ½ total X dimension. Requires topology aware scheduling.
• Testing with XK acceptance applications showed either little change or improved performance for ‘before/after’ comparison.
XE Usage in the last 3 months

- 50% of usage is 1,024 nodes or larger.
- Two teams using 5,000 and 8,192 nodes.
- During Friendly User period, several teams sustained runs at full system.
- Nothing prevents users from submitting very large jobs and priority goes to larger jobs.
- Average expansion factor for large jobs < 10.
Turbulence

PSDNS – 3D FFTs, off-node transposes using CAF replacement for the concurrent Alltoalls. Routinely running at 8,192 nodes (262,144 tasks) for $8,192^3$ problem in 48 hr. chunks.

- DISTUF – DNS using PETSc CG for direct Poisson solve. Looking at using MG. Scaling and code validation underway. Up to 512 nodes.
Cybershake

- Scalability issues with jobs on busy system. Cray (Fiedler) Topaware improved node selection and rank ordering.
- Looking at ways of using host CPU on XK nodes for part of workflow while GPU is doing computation.

<table>
<thead>
<tr>
<th>#nodes</th>
<th>Default</th>
<th>Topaware</th>
<th>Speedup</th>
<th>Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>64</td>
<td>4.006</td>
<td>3.991</td>
<td>0.37%</td>
<td>100% → 100%</td>
</tr>
<tr>
<td>512</td>
<td>0.572</td>
<td>0.554</td>
<td>3.15%</td>
<td>87.5% → 90%</td>
</tr>
<tr>
<td>4096</td>
<td>0.119</td>
<td>0.077</td>
<td>35.29%</td>
<td>52.6% → 81%</td>
</tr>
</tbody>
</table>

Paper in Extreme Scaling Workshop 2013
Coarse Grained MD

- Novel MD algorithm.
- Improved memory usage.
- Hilbert space filling curve (SFC) for load balancing.
- Dynamic communications mapping to handle irregular SFC boundaries.
- Scaling to 16,250 nodes (260,000 FP cores). Earlier data shown at right.
- Mostly MPI but replaced some functionality with DMAPP when faster.
- BW Symposium - https://bluewaters.ncsa.illinois.edu/web/portal/symposium-may-2013
Multi-Scale Fluid-Kinetic Simulation

- MHD-kinetic code to modeling the solar wind.
- Chombo framework for AMR and dynamic load balancing.
- P3DFFT
- Good strong (starting at 1,250 nodes) and weak scaling to 7,500 nodes.
- http://adsabs.harvard.edu/abs/2013ASPC..474..165P
XK jobs as of end of September

By looking at aprun instances and not job node count we can see when workloads are many single nodes bundled in a larger job.

A large number of >3,000 node apuns.
XK use scenarios

- Adoption of GPU
  - SIAL (ACESIII) – user annotation (SIAL directives) to assist CUDA code generator to get best speed-up. (T) – triples from CCSD(T).
  - https://bluewaters.ncsa.illinois.edu/web/portal/symposium-may-2013

- NEMO – PETSc + MAGMA to utilize GPU. Working on issues with sparse matrices and developing load balancing strategy across GPU and host CPU.
  - https://bluewaters.ncsa.illinois.edu/web/portal/symposium-may-2013
TorusView of 10 largest running jobs

- Relatively compact allocations.
- Some scattered clustering.
- Lots of concave shapes.
- Not showing all the small jobs filling in the rest of the torus.
TorusView of 10 largest running jobs

- Allocations shift planes as the end of the Z direction is hit.
- Voids where larger job allocations wrap around smaller ones.
Better nid allocation

- Would be better to have one of the following ...
- More about this tomorrow.
Impact of nid allocation

- Job – Job interaction
  - Analysis of key application communication intensity and sensitivity
  - 20% slowdown typical, 100% or more possible.

<table>
<thead>
<tr>
<th>Communication</th>
<th>MILC</th>
<th>NAMD</th>
<th>NWCHEM</th>
<th>PSDNS</th>
<th>WRF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intensive</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Sensitive</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

1 – low  3 – high as viewed by convex app.
Impact of poor nid allocation - Consistency

- Two jobs (8,192 nodes) with nearly same nid allocation (s10_8972n). Red job affected by other workload communicating through the region.

- Run time variation - poor wallclock accuracy (padding wallclock).
IO

- LNETs scattered across the torus (orange colored geminis).
- Specific OSTs served by specific LNETs (not a full fat tree for the IB between OSTs and LNETs).
- IO is “topology sensitive”.

JLPC 2013
Routing of IO write

- 15 compute geminis (●) (30 nodes) writing to files served by a LNET pair (●).
- Color scale is the number of convergent routes on the link.
“Topology” aware IO library

- Analysis of the Blue Waters File System Architecture for Application IO Performance - CUG 2013, May 6, 2013 Authors: Kalyana Chadalavada, Rob Sisneros
Congestion Protection

- To avoid data loss, traffic injection is throttled for a period of time, when reaching a point where forward progress is stalling. Throttling is applied and removed until congestion is cleared.
- System monitors percentage of time that traffic trying to enter the network from the nodes and percentage of time network tiles are stalled.
- Fortunately not a common occurrence. It does happen, typically in bursts.
- Can happen with node-node (MPI, PGAS) or node-LNET (IO) traffic.
- Many-to-one and long-path patterns.
- Libraries and user can control node injection as a precaution.
- In CP reports, flit rates represent data arriving at the node from the interconnection network.
Congestion Protection Burst
Congestion Protection Analysis

• Look at application to node relation.
• wrf listed as top application and the top 10 nodes are wrf nodes.
• nwchem running at same time (listed #4).
• The OVIS state of the network data should help here.
Interrupts

- We provide to the user a checkpoint interval calculator based on the work of J. Daly, using recent node and system interrupt data.
- September data
  - 22,640 XE nodes MTTI ~ 14 hrs.
  - 4,224 XK nodes MTTI ~ 32 hrs.
  - System interrupts MTTI ~ 100 hrs.
- Checkpoint intervals on the order of 4 – 6 hrs. at full system (depending on time to write checkpoint).