Towards Characterization of Data Movement in Large-Scale Scientific Applications

Jun Yi, Rajkumar Kettimuthu and Venkatram Vishwanath
Mathematics and Computer Science Division; Computational Institute
Argonne National Laboratory; University of Chicago
<kettimut,venkatv>@mcs.anl.gov

Application data flows have diverse needs, varying in characteristics including burstiness, latency, reliability, jitter, message size and priority. Parallel m-to-n flows are becoming increasingly common. Bulk data movement is moving beyond file transfers to memory-to-memory, memory-to-disk and disk-to-memory transfers. We aim to develop a framework to capture the requirements and characteristics of various science flows.

Application Characteristics

- Design concerted flows API
- Capture Application Requirements
  - Latency, jitter and bandwidth
  - Reliability and error tolerance
  - Deadline, start time, burstiness
  - Contiguous vs non-contiguous
  - Compression technique
  - M-to-N flows
- Create data transfer kernels for representative applications

Flow characteristics

<table>
<thead>
<tr>
<th>App</th>
<th>Type of Flow</th>
<th># of Streams</th>
<th>BW</th>
<th>Latency</th>
<th>Burstiness</th>
<th>Size</th>
<th>Protocol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global Online</td>
<td>Data</td>
<td>4 per app</td>
<td>High N</td>
<td>Y</td>
<td>Large TCP</td>
<td>TCP</td>
<td></td>
</tr>
<tr>
<td>AIS</td>
<td>Data</td>
<td>3 per app</td>
<td>High N</td>
<td>Y</td>
<td>Small TCP</td>
<td>TCP</td>
<td></td>
</tr>
<tr>
<td>ENZO Simulation</td>
<td>Data</td>
<td>4 per app</td>
<td>High N</td>
<td>Y</td>
<td>Variable TCP</td>
<td>TCP</td>
<td></td>
</tr>
<tr>
<td>Remote Vis</td>
<td>Data</td>
<td>3 per app</td>
<td>High N</td>
<td>Y</td>
<td>Large TCP</td>
<td>TCP</td>
<td></td>
</tr>
</tbody>
</table>

Acknowledgements: Office of Advanced Scientific Computing Research, Office of Science, U.S. Department of Energy, under Contract DE-AC02-06CH11357,