Improving Data Retrieval Rates Using Remote Data Servers

Ted D'Ottavio
Bartosz Frak
Seth Nemesure
John Morris

Brookaven National Laboratory

MOMAU002
Project Goal

- Order of magnitude faster data delivery to client applications
- Unnoticeable ( < 1 sec) delays when working with live data
- Removing addressable memory constraints on 32bit client applications
- Future proofing
  - Extensible, pluggable architecture
  - Web application friendly data formats (XML, JSON)
Solution Outline

• Thin out the clients
  • Migrate proven, tested codebase to middleware services
  • Parallelize processing (culling) with a compute grid
  • Deliver culled data to clients based on the request context

• Move closer to the data stores
  • Connect the middleware platform using high speed network to NAS and archive stores

• Cache
  • Cache resources in a distributed non-replicating memory caches across the cluster
  • Keep evicted resources in a local SSD cache
## Results

<table>
<thead>
<tr>
<th>Throughput (MB/sec)</th>
<th>Speedup</th>
</tr>
</thead>
<tbody>
<tr>
<td>Client to Remote Disk-store</td>
<td>5.4</td>
</tr>
<tr>
<td>Client through Data Server to Remote Disk-store</td>
<td>146</td>
</tr>
<tr>
<td>Client through Data Server to SSD Cache</td>
<td>245</td>
</tr>
<tr>
<td>Client through Data Server to RAM</td>
<td>968</td>
</tr>
</tbody>
</table>