Social Science Gateway to Teragrid @ Cornell University

(also known as VirtualRDC)
Social Science Gateway provides:

- Usual social science data management and analysis tools not available on TeraGrid (TG) itself
  - SAS, SPSS, Stata
- Also provides standard TG tools
  - R, compilers, etc.
- Graphical desktop access to SSG compute nodes
Social Science Gateway provides (cont)

- High-speed (10Gbps) access to TeraGrid and thus to thousands of compute nodes and/or data
- Local data storage (base allocation includes some storage, additional storage from Cornell storage cloud at cost)
Typical user

- Researcher has existing theory/model/data
- Project requires substantially more resources than typically locally available
- Researcher
  - is comfortable with the traditional data tools (SAS, SPSS)
  - Less comfortable with TeraGrid tools
Typical workflow

1. Project inception
2. Researcher requests TeraGrid allocation for full computational resources, obtains TG-wide credentials
3. Researcher requests Social Science Gateway (SSG) access with existing TG-wide credentials
Project Workflow
Researcher requests access to TeraGrid
Researcher requests access to SSG
Typical workflow (cont)

- (4) Data preparation and analysis on SSG
- (5) Full computation on TG
- Iterate (4) and (5) until satisfied
- End of project
Researcher uses SSG

Analysis/results appropriate
Researcher accesses SSG
Researcher submits job on TG

---

**TeraGrid User Portal**

![TeraGrid Systems Monitor](image)

**HPC Systems** | Storage Systems | Advanced Vis Systems | Special Purpose Systems

<table>
<thead>
<tr>
<th>Name</th>
<th>Institution</th>
<th>System</th>
<th>Peak TFlops</th>
<th>Memory TBytes</th>
<th>Status</th>
<th>Load</th>
<th>Running Jobs</th>
<th>Queued Jobs</th>
<th>Other Jobs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kären</td>
<td>NICS</td>
<td>Cray XT5</td>
<td>606.00</td>
<td>129.00</td>
<td>Up</td>
<td></td>
<td>121</td>
<td>103</td>
<td>77</td>
</tr>
<tr>
<td>Ranger</td>
<td>TACC</td>
<td>Sun Constellation</td>
<td>579.40</td>
<td>123.00</td>
<td>Up</td>
<td></td>
<td>382</td>
<td>145</td>
<td>184</td>
</tr>
<tr>
<td>Abe</td>
<td>NCSA</td>
<td>Dell Intel 54 Linux Cluster</td>
<td>93.47</td>
<td>9.36</td>
<td>Up</td>
<td></td>
<td>69</td>
<td>495</td>
<td>374</td>
</tr>
<tr>
<td>Lonestar</td>
<td>TACC</td>
<td>Dell PowerEdge Linux Cluster</td>
<td>62.16</td>
<td>11.50</td>
<td>Up</td>
<td></td>
<td>136</td>
<td>281</td>
<td>0</td>
</tr>
<tr>
<td>Steele</td>
<td>Purdue</td>
<td>Dell Intel 64 Linux Cluster</td>
<td>60.00</td>
<td>12.40</td>
<td>Up*</td>
<td></td>
<td>386</td>
<td>309</td>
<td>89</td>
</tr>
<tr>
<td>Queen Bee</td>
<td>LONI</td>
<td>Dell Intel 64 Linux Cluster</td>
<td>60.70</td>
<td>5.31</td>
<td>Up</td>
<td></td>
<td>113</td>
<td>205</td>
<td>35</td>
</tr>
<tr>
<td>Lincoln</td>
<td>NCSA</td>
<td>Dell Intel PowerEdge 1950</td>
<td>47.50</td>
<td>3.00</td>
<td>Up</td>
<td></td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Big Red</td>
<td>IU</td>
<td>IBM e135Q</td>
<td>30.60</td>
<td>6.00</td>
<td>Up</td>
<td></td>
<td>885</td>
<td>1183</td>
<td>19</td>
</tr>
<tr>
<td>Big Ben</td>
<td>PSC</td>
<td>Cray XT3</td>
<td>21.50</td>
<td>4.04</td>
<td>Up</td>
<td></td>
<td>16</td>
<td>139</td>
<td>50</td>
</tr>
<tr>
<td>Mercury</td>
<td>NCSA</td>
<td>IBM Itanium2 Cluster</td>
<td>10.23</td>
<td>4.47</td>
<td>Up</td>
<td></td>
<td>123</td>
<td>598</td>
<td>0</td>
</tr>
<tr>
<td>Cobalt</td>
<td>NCSA</td>
<td>SGI Altix</td>
<td>6.55</td>
<td>3.00</td>
<td>Up</td>
<td></td>
<td>33</td>
<td>253</td>
<td>112</td>
</tr>
<tr>
<td>Frost</td>
<td>NCFAR</td>
<td>IBM BlueGene.L</td>
<td>5.73</td>
<td>0.51</td>
<td>Up</td>
<td></td>
<td>1</td>
<td>25</td>
<td>0</td>
</tr>
<tr>
<td>Pope</td>
<td>PSC</td>
<td>SGI Altix 4700</td>
<td>5.00</td>
<td>1.04</td>
<td>Up</td>
<td></td>
<td>19</td>
<td>70</td>
<td>31</td>
</tr>
<tr>
<td>TeraGrid Cluster</td>
<td>SDSC</td>
<td>IBM Itanium2 Cluster</td>
<td>5.10</td>
<td>1.02</td>
<td>Up</td>
<td></td>
<td>7</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>TeraGrid Cluster</td>
<td>UC/ANL</td>
<td>IBM Itanium2 Cluster</td>
<td>0.61</td>
<td>0.24</td>
<td>Up*</td>
<td></td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>NSTG</td>
<td>ORNL</td>
<td>IBM IA-32 Cluster</td>
<td>0.34</td>
<td>0.07</td>
<td>Up</td>
<td></td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Total: 1580.89 | 314.58 | 2316 | 4618 | 973
Researcher...

- ...transfers data to TG (gridftp)
- ...submits jobs on TG
- ...iterates back and forth until project is completed
Other Cornell resources

- Cornell Matlab cluster (same 10Gbps network)
- Additional storage if needed by project (rental on a yearly basis)
VirtualRDC
@ Cornell University

(also known as Social Science Gateway to Teragrid)
VirtualRDC provides:

- Environment replicating the RDC compute center (Linux GUI, directory structure, etc.)
- Zero-obs datasets allowing some basic programming preparation
- Stepping-stone to projects in the Census RDC
Researcher PC

Social Science Gateway to TeraGrid

Confidential data at U.S. Census Bureau through RDC

TeraGrid
Social Science Gateway at Cornell

Funding through NSF (pending)

Available 4th quarter 2009 will be announced at http://www.vrdoc.cornell.edu

Contact:
Lars Vilhuber (lars.vilhuber@cornell.edu)
Or virtualrdc@cornell.edu
TODO

- Startup...
- Update website
- Tutorial on workflow, tools
- Integrate GUI tools for GridFTP transfers
  - Grifi
  - Globus Java Gui