HTC on HPC

Parsl's Role in Mapping Earth's Topography from Space

Claire Porter
Polar Geospatial Center
Sept 2022
PGC DEM INITIATIVES
REMA, ARCTICDEM, EARTHDEM

1. Stereoscopic high-resolution imagery coverage
2. Scalable terrain extraction algorithm
3. Ridiculous compute resources

source: Polar Geospatial Center
## HTC ON HPC

### PROBLEM SCALE

<table>
<thead>
<tr>
<th></th>
<th>ArcticDEM</th>
<th>REMA</th>
<th>EarthDEM</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strip DEMs</td>
<td>365,766</td>
<td>232,898</td>
<td>950,070</td>
<td>1,548,734</td>
</tr>
</tbody>
</table>

- 15,177,593 DEM extraction tasks
- 1,548,734 DEM assembly tasks
Direct Job Submission:
1 to 1 task-job bundling

Swift/Parsl:
100+ to 1 task-job bundling

• A set of tasks can number over 10,000, each using ½ a single node

• If project queue limits are high, jobs cripple the scheduler by raising iteration time

• If project queue limits are low, jobs can’t effectively use the available compute nodes

• Parsl allows tasks to be bundled into groups, reducing the impact of the scheduler while achieving high compute node usage

• Parsl also can dynamically allocate tasks to open nodes within a running job – they are not limited to a static list of tasks to run (inefficient if run time varies greatly)
PARSL’S PLACE IN THE ECOSYSTEM

PGC

Assemble Source and Globus to Frontera

Assemble Results and Globus Home

DEM Extraction using SETSM

Data Merging

PARSL

Node Count

Reserved

Used

Time
PARSL’S PLACE IN THE ECOSYSTEM

Data Gathering

Assemble Source and Globus to Frontera

Assemble Results and Globus Home

Dem Extraction using SETSM

Data Merging

Node Count

Time

Reserved

Used
PARSL’S PLACE IN THE ECOSYSTEM

- Assemble Source and Globus to Frontera
- Assemble Results and Globus Home
- Data Merging
- DEM Extraction using SETSM

PGC

PARSL