Enabling Earth System Science in the Cloud
by Brian Dobbins, NCAR

Lightning Talk
CI Compass Cyberinfrastructure for NSF Major Facilities 2022 Workshop
March 1, 2022
Enabling Earth System Science in the Cloud

Brian Dobbins
CI4MF 2022
March 1st, 2022
Overview

• The Community Earth System Model (CESM)

• Cloud Complexity
  – Configuration
  – Price & Performance

• CESM Cloud API
  – Real-time pricing
  – Benefits
  – Challenges
CESM

CESM is a sophisticated and complex model:

- Roughly 1.8 million lines of Fortran code
- Dependencies: compilers, MPI, NetCDF, PNetCDF, XML, Python
- Thousands of different built-in configurations

NCAR & NSF-funded researchers have access to on-prem systems:
- Cheyenne (5 PF, retiring!)
- Derecho (20 PF, soon!)

... But that’s only a subset of all the people who use the model.
- Our cloud work is mostly focused on external users of CESM.
Cloud Complexity - Price & Performance

Even within a single cloud vendor, performance is complex…

![Graph showing Simulated Years Per Day for Cheyenne and AWS (C5N) with different core counts. The graph indicates that Cheyenne performs better than AWS at all core counts, with the best performance at 3564 cores.]
And going multi-cloud, it gets even more complicated!

Cloud Complexity - Price & Performance

Simulated Years Per Day (B1850 @ 1-degree)
- Cheyenne
- Azure (HBv3)

<table>
<thead>
<tr>
<th>Cores</th>
<th>Cheyenne</th>
<th>Azure (HBv3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>288 cores</td>
<td>3.09</td>
<td>4.89</td>
</tr>
<tr>
<td>720 cores</td>
<td>8.46</td>
<td>10.94</td>
</tr>
<tr>
<td>2160 cores</td>
<td>19.11</td>
<td>25.8</td>
</tr>
</tbody>
</table>
CESM Cloud API

CESM Case: BW1850 @ 1-degree resolution

Select Cloud Provider
- AWS
- Azure
- GCP
...

Select Node Type
- C6g
- C6gn
- C5
- C5a
- C5n
- C4
...

Select Network Type
- 10 Gbit
- 25 Gbit
- 100 Gbit
- EFA

Select Storage Type
- EBS
- EFS
- FSX

Select Cloud Region
- US-East-1
- US-East-2
- US-West-1
- US-West-2

Find CESM Layout
- 144x1
- 288x1
- 576x1
- 1188x1
- 1188x3

Launch Cloud Resources

API Handles the Complexity
The API is accessible via web or command-line and simplifies using the cloud for scientists.
We’re planning on real-time querying of CPU costs, giving users the ability to estimate needs and choose between cost-efficient and high-performance options.

<table>
<thead>
<tr>
<th>Efficiency</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.65 SYPD</td>
<td>1.23 SYPD</td>
</tr>
<tr>
<td>$336.47 / SY</td>
<td>$621.94 / SY</td>
</tr>
</tbody>
</table>
**Benefits of an API**

<table>
<thead>
<tr>
<th><strong>Standardization</strong></th>
<th>Tested, common platform for workflows, training, etc</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Support</strong></td>
<td>One place to fix configurations; ‘phone-home’ support</td>
</tr>
<tr>
<td><strong>Metrics</strong></td>
<td>Can (anonymously) track usage</td>
</tr>
</tbody>
</table>
Challenges

- Web Design / Development
- API Design
- Containers
- Database Integration
- InfoSec
- Cloud-specific APIs
- … And the application itself
Thank You!

Brian Dobbins
bdobbins@ucar.edu