Running CESM2 in the AWS Cloud

Brian Dobbins

CESM Workshop 2018
June 19th – NCAR Center Green
Motivations

Just a few of many:
• Immediate, on-demand compute power
• Unified environment for training
• Greater access for scientists
What *is* ‘the cloud’?
What *is* ‘the cloud’?
What *is* the Cloud?

*(as it applies to running CESM!)*
What *is* the Cloud?

*(as it applies to running CESM!)*

Let’s start by comparing it, where possible, with Cheyenne:
What is the Cloud?

(as it applies to running CESM!)

Let’s start by comparing it, where possible, with Cheyenne:

<table>
<thead>
<tr>
<th></th>
<th>“Cheyenne”</th>
<th>‘The Cloud’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td>NWSC in Cheyenne, WY</td>
<td>It (mostly) doesn’t matter.</td>
</tr>
</tbody>
</table>
What *is* the Cloud?

*(as it applies to running CESM!)*

Let’s start by comparing it, where possible, with Cheyenne:

<table>
<thead>
<tr>
<th></th>
<th>“Cheyenne”</th>
<th>‘The Cloud’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td>NWSC in Cheyenne, WY</td>
<td>It (mostly) doesn’t matter.</td>
</tr>
<tr>
<td>Nodes</td>
<td>4,032 total (shared)</td>
<td>Whatever you request (*).</td>
</tr>
</tbody>
</table>
What *is* the Cloud?

*(as it applies to running CESM!)*

Let’s start by comparing it, where possible, with Cheyenne:

<table>
<thead>
<tr>
<th></th>
<th>“Cheyenne”</th>
<th>‘The Cloud’</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Location</strong></td>
<td>NWSC in Cheyenne, WY</td>
<td>It (mostly) doesn’t matter.</td>
</tr>
<tr>
<td><strong>Nodes</strong></td>
<td>4,032 total (shared)</td>
<td>Whatever you request (*).</td>
</tr>
<tr>
<td><strong>Processor Type</strong></td>
<td>2.3 Ghz Xeon E5-2697V4</td>
<td>Multiple types available.</td>
</tr>
</tbody>
</table>
What *is* the Cloud?

*(as it applies to running CESM!)*

Let’s start by comparing it, where possible, with Cheyenne:

<table>
<thead>
<tr>
<th></th>
<th>“Cheyenne”</th>
<th>‘The Cloud’</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Location</strong></td>
<td>NWSC in Cheyenne, WY</td>
<td>It (mostly) doesn’t matter.</td>
</tr>
<tr>
<td><strong>Nodes</strong></td>
<td>4,032 total (shared)</td>
<td>Whatever you request (*).</td>
</tr>
<tr>
<td><strong>Processor Type</strong></td>
<td>2.3 Ghz Xeon E5-2697V4</td>
<td>Multiple types available.</td>
</tr>
<tr>
<td><strong>Network</strong></td>
<td>Specialized (low latency)</td>
<td>Standard.</td>
</tr>
</tbody>
</table>

This is the first *key* difference – the low-latency network in supercomputers matters!
What *is* the Cloud?

*(as it applies to running CESM!)*

Let’s start by comparing it, where possible, with Cheyenne:

<table>
<thead>
<tr>
<th></th>
<th>“Cheyenne”</th>
<th>‘The Cloud’</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Location</strong></td>
<td>NWSC in Cheyenne, WY</td>
<td>It (mostly) doesn’t matter.</td>
</tr>
<tr>
<td><strong>Nodes</strong></td>
<td>4,032 total (shared)</td>
<td>Whatever you request (*).</td>
</tr>
<tr>
<td><strong>Processor Type</strong></td>
<td>2.3 Ghz Xeon E5-2697V4</td>
<td>Multiple types available.</td>
</tr>
<tr>
<td><strong>Network</strong></td>
<td>Specialized (low latency)</td>
<td><em>Standard.</em></td>
</tr>
<tr>
<td><strong>System Management</strong></td>
<td>NCAR (SSG, USS, CSG, etc.)</td>
<td><em>It depends!</em></td>
</tr>
</tbody>
</table>

Someone has to configure the environment! The end user? NCAR? But not AWS.
What *is* the Cloud?

*(as it applies to running CESM!)*

Let’s start by comparing it, where possible, with Cheyenne:

<table>
<thead>
<tr>
<th></th>
<th>“Cheyenne”</th>
<th>‘The Cloud’</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Location</strong></td>
<td>NWSC in Cheyenne, WY</td>
<td>It (mostly) doesn’t matter.</td>
</tr>
<tr>
<td><strong>Nodes</strong></td>
<td>4,032 total (shared)</td>
<td>Whatever you request (*).</td>
</tr>
<tr>
<td><strong>Processor Type</strong></td>
<td>2.3 Ghz Xeon E5-2697V4</td>
<td>Multiple types available.</td>
</tr>
<tr>
<td><strong>Network</strong></td>
<td>Specialized (low latency)</td>
<td>Standard.</td>
</tr>
<tr>
<td><strong>System Management</strong></td>
<td>NCAR (SSG, USS, CSG, etc.)</td>
<td>It depends!</td>
</tr>
<tr>
<td><strong>Cost</strong></td>
<td>Upfront, single cost ($)</td>
<td>Metered by use ($$$).</td>
</tr>
</tbody>
</table>

Cloud pricing has a lot of factors – pre-buying, on-demand use, etc. Every option we looked at was more expensive than our own system.
Setting up a ‘Virtual Cluster’

The easiest (but not necessarily the best) way to run CESM on AWS was to spin up a virtual cluster:

- Linux + Compilers + Libraries
- Queuing system
- Multiple nodes

At this point, it’s like any other port – just modify the machine-specific settings (eg, config_machines.xml).
Using the Virtual Cluster

• With this environment, we can do all the usual stuff:
  – `create_newcase --case ... --compset ... --mach aws_c5`
  – `./case.build`
  – `./case.submit`
  – `qstat`

• *We* had to configure that functionality (once)

• From the science side of things, it *operated* the same as Cheyenne (or any other cluster).
Test Configuration

- CESM Configuration:
  - 1-degree CAM6(*) Aquaplanet case
  - 10 model days
  - From 36 to 2304 processors
  - Always used 3 threads per MPI rank

- Hardware:
  - Cheyenne (Xeon Broadwell, 36 cores per node, Infiniband net)
  - ‘C4’ nodes (Xeon Haswell, 18 cores per node, 10Gbit net)
  - ‘C5’ nodes (Xeon Skylake, 36 cores per node, 25Gbit net)
Performance Results

CAM6 Performance

Simulated Years Per Day (SYPD)

- Cheyenne
- AWS C4
- AWS C5

Cores

NCAR | Running CESM2 in the Cloud
Performance Results

CAM6 Performance

Simulated Years Per Day (SYPD)

Cores

Cheyenne
AWS C4
AWS C5

36 72 144 288 576 1152 2304
Performance Results

CAM6 Performance

- Cheyenne
- AWS C4
- AWS C5

Simulated Years Per Day (SYPD)

Cores

Running CESM2 in the Cloud
Performance Results

CAM6 Performance

Simulated Years Per Day (SYPD)

- Cheyenne
- AWS C4
- AWS C5

Cores

Running CESM2 in the Cloud
What’s next?

• A few new runs:
  – Can we trade threads for ranks?
  – Re-run the C5 cases

• Run on Azure’s Infiniband-connected nodes.
  – Performance *should* be closer to Cheyenne.

• Replace the ‘virtual cluster’ with a *cloud-centric* run script:
  – The `/case.submit` script can request nodes itself – no cluster!

NCAR | Running CESM2 in the Cloud
Looking deeper into the future...

• Do we integrate cloud functionality with CIME?
  – Like PBS, Slurm, etc.

• Do we remove compile-time processor counts from CESM?
  – Cloud hardware changes fast
  – Easier for end users
  – Load-balancing
Conclusions

• Running in the cloud is pretty easy
  – Hours -> Minutes -> ‘Instant’

• The Cloud can provide a common training environment
  – Hopefully we’ll do this at a tutorial at AGU!
  – Anyone, from any institution, can use our environment

• Performance is mixed
  – Good on small node counts (hardware improves fast!)
  – Bad on large node counts
A quick note of thanks to Amazon, and in particular John Ewart and Kevin Jorissen, for their support – we got to do all of this for free courtesy of their ‘Research Credits’ program. They’ve also said they’ll help with any future hands-on CESM2 training using the cloud.
Small vs. Large Scale Performance

(Relative to Cheyenne)

~ 2.5 SYPD for Cheyenne & AWS C5

~ 30 SYPD for Cheyenne, but only 11.33 for AWS C5
What *is* ‘the cloud’?

According to NIST, the cloud:
- is an on-demand self-service
- can be accessed via the internet
- pools resources across customers
- can scale to fit peaks in demand
- has metered charging like a utility

In short, it’s a lot of hardware that we can access when we want and use how we want... provided we pay for it.
Setting up a ‘Virtual Cluster’

Two ‘lessons learned’ while setting it up:

1. AWS ‘Placement Groups’ can ensure compute nodes are located as close as possible (lower network latency).

2. Making sure that PBS/Torque *placed ranks* efficiently. (Very important!)