CICompass

Funded by the National Science Foundation, Grant #2127548

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

USC Viterbi School of Engineering Information Sciences Institute

INDIANA UNIVERSITY

UNIVERSITY OF NOTRE DAME

THE UNIVERSITY OF UTAH

TEXAS TECH UNIVERSITY

renci

Enabling Earth System Science in the Cloud

Brian Dobbins

CI4MF 2022

March 1st, 2022

This material is based upon work supported by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement No. 1852977.

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:

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

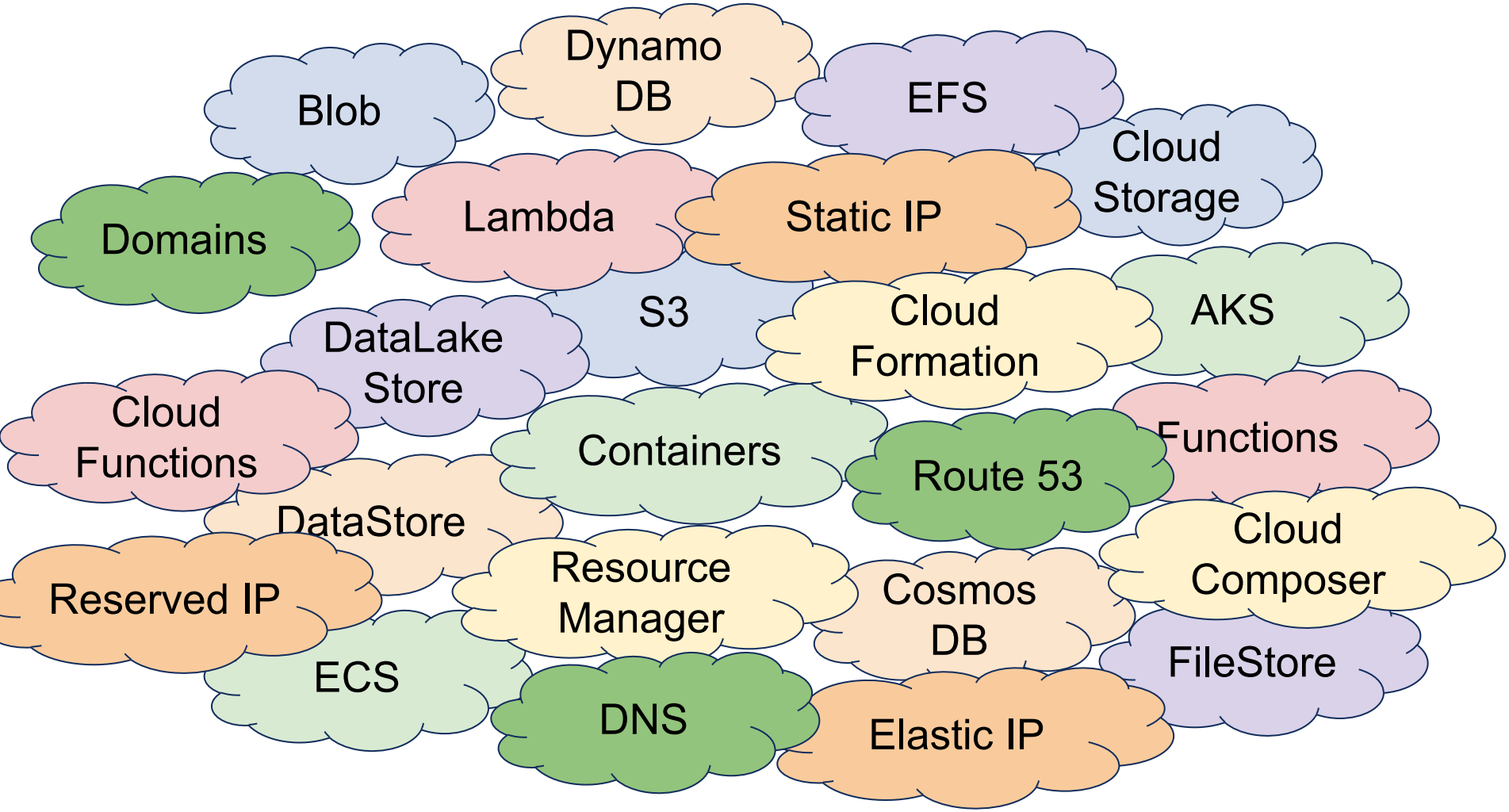
NCAR & NSF-funded researchers have access to on-prem systems:

- Cheyenne (5 PF, retiring!)
- Derecho (20 PF, *soon*!)

… But that's only a *subset* of all the people who use the model.

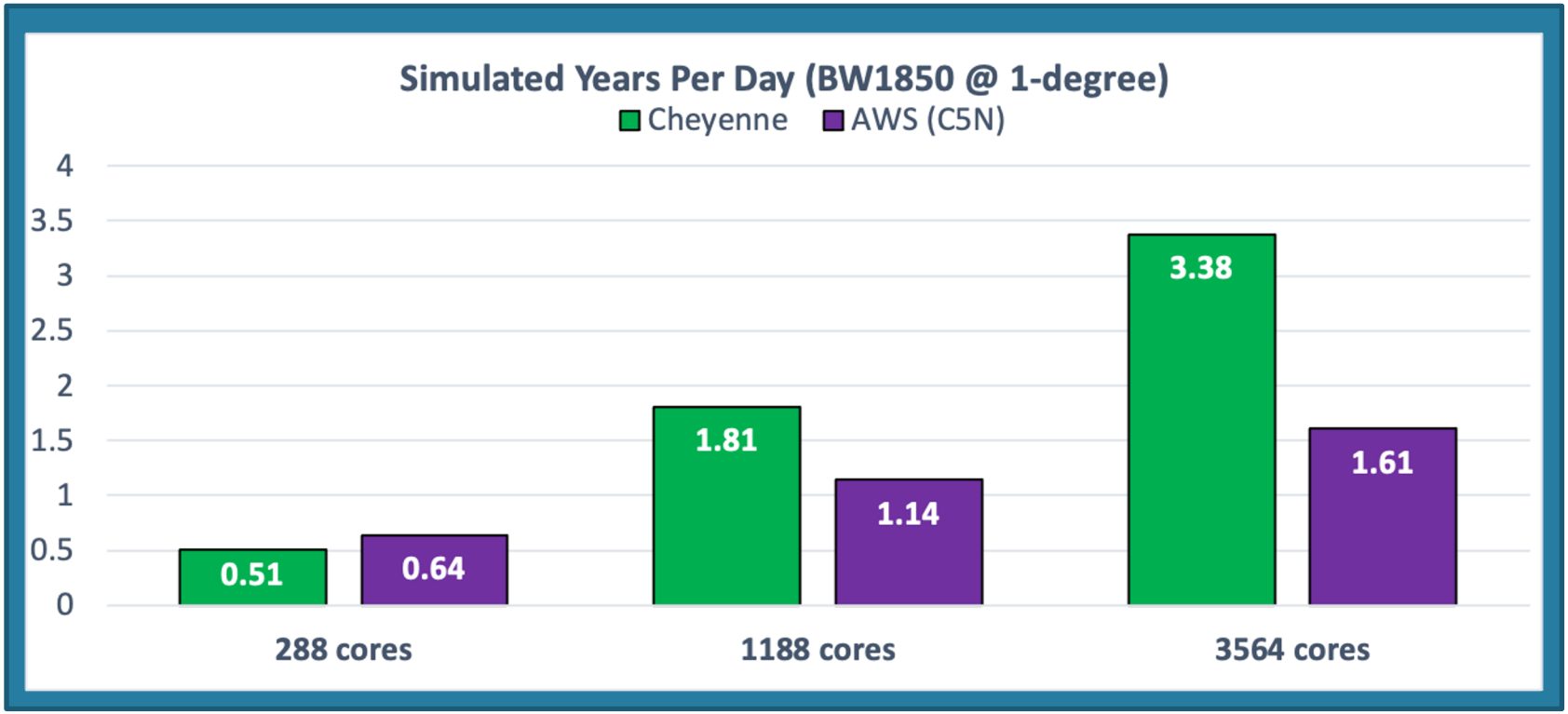
- Our cloud work is mostly focused on *external* users of CESM.

Cloud Complexity - Configuration



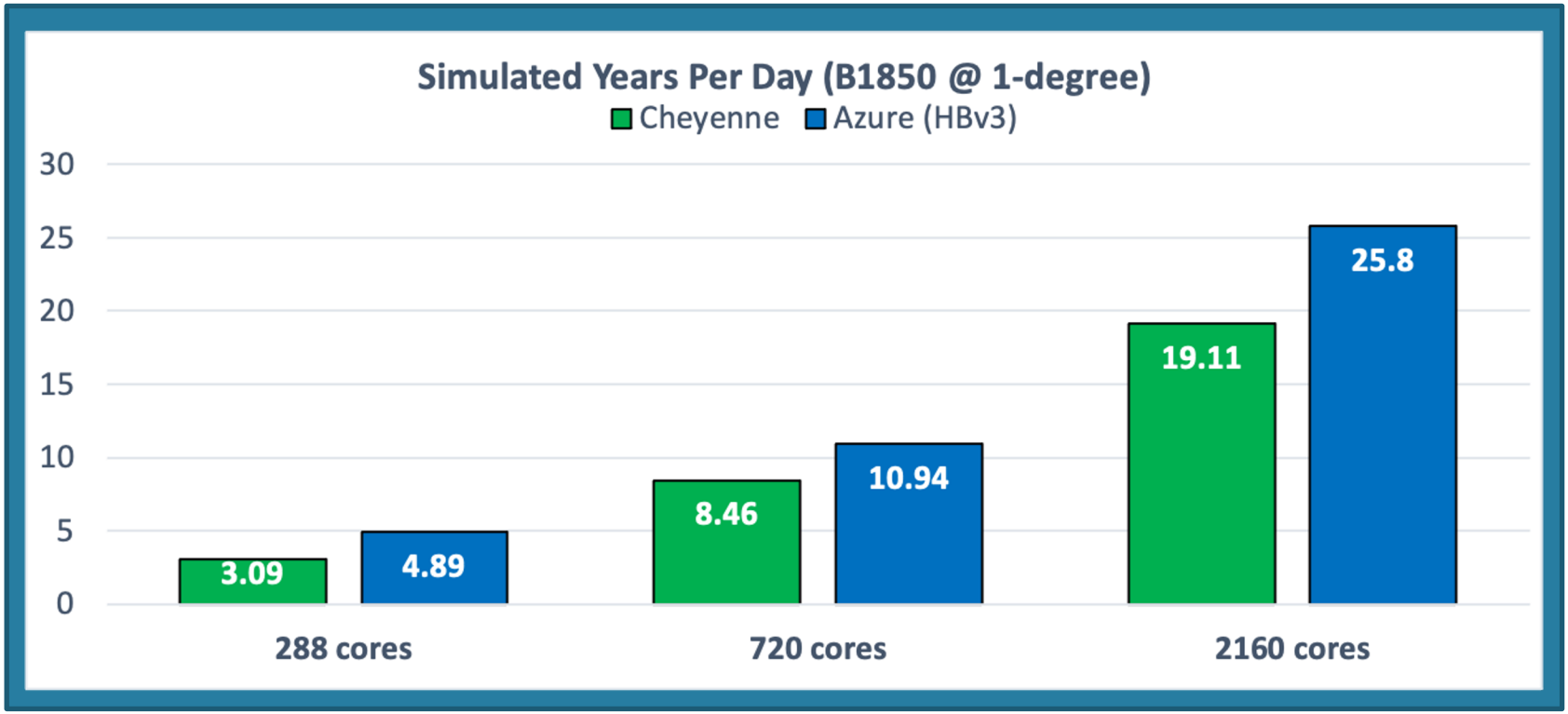
Cloud Complexity - Price & Performance

Even within a single cloud vendor, performance is complex…

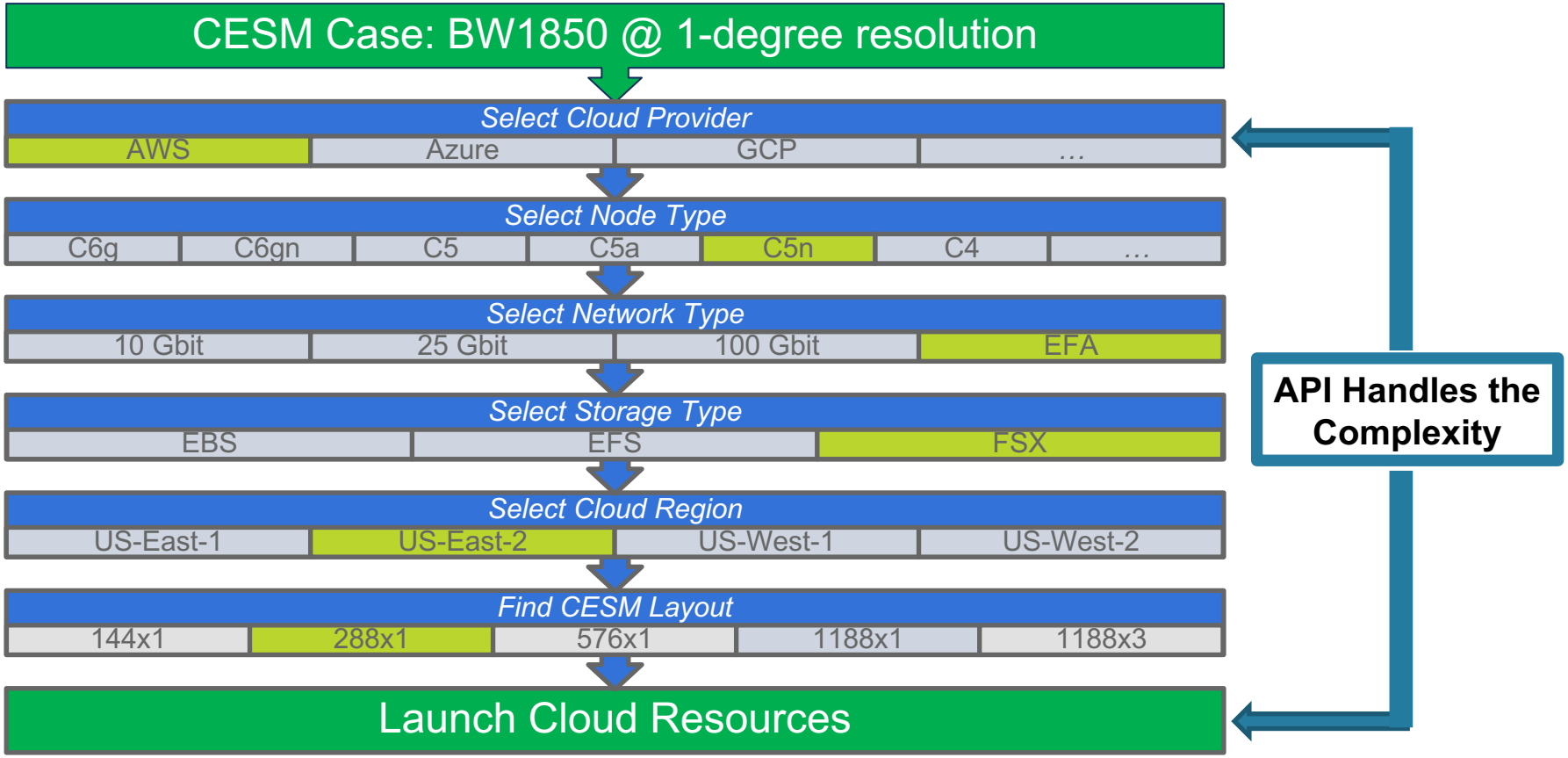


Cloud Complexity - Price & Performance

And going multi-cloud, it gets even *more* complicated!

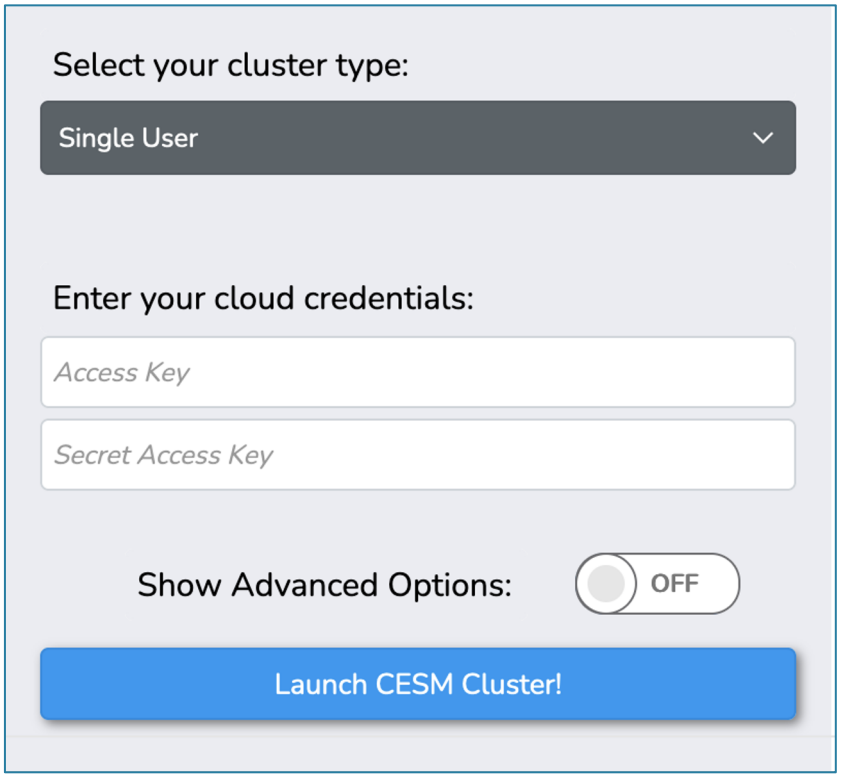


CESM Cloud API



CESM Cloud API

Web



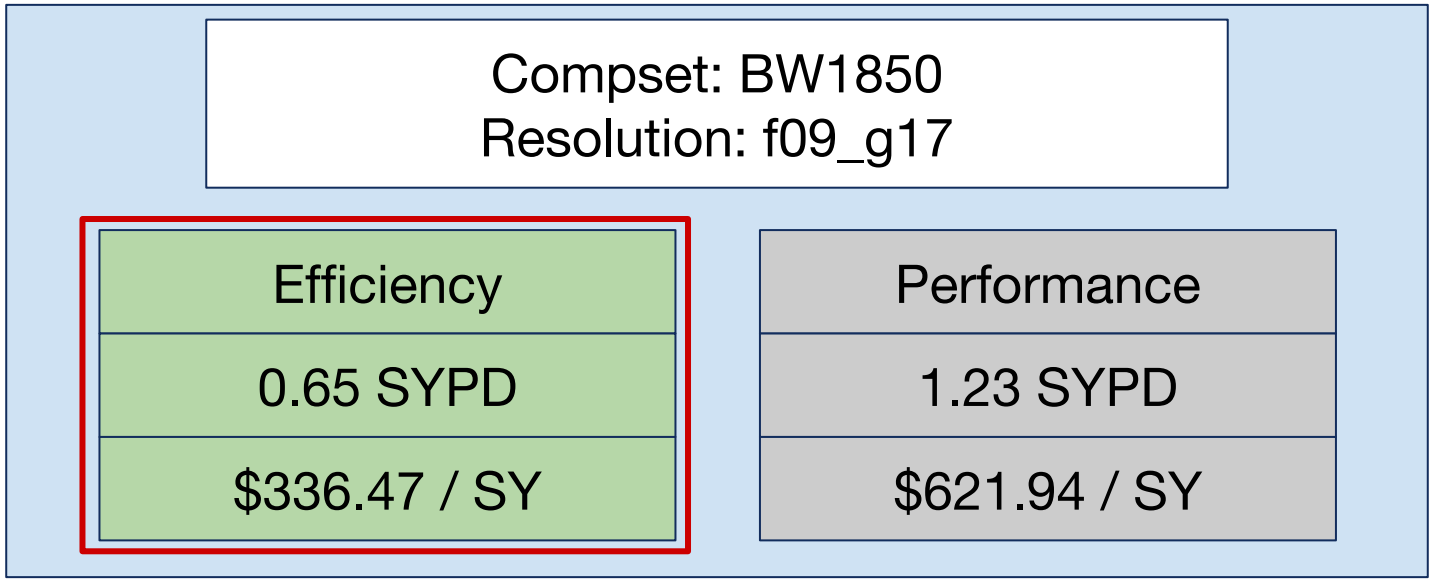
Linux CLI

```
[bdobbins@cheyenne1] $ cesm create cluster –v
Info: Found AWS credentials: **xT3g..
Info: Mode: Cluster[coupled]
Info: Nodes: C5N.18xlarge
Info: Querying regions for pricing
Info: Selected region US-East-1
Info: Requiring EFA
Info: EBS for /home (40G)
Info: Lustre for /scratch (2.4T)
Status: Creating cluster configuration
Status: Launching cluster
Status: Installing compilers on head node
…
  Success!
   Connect via: ssh bdobbins@52.15.115.58

[bdobbins@cheyenne1] $
```

The API is accessible via web or command-line and *simplifies* using the cloud for scientists.

Real-Time Pricing



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.

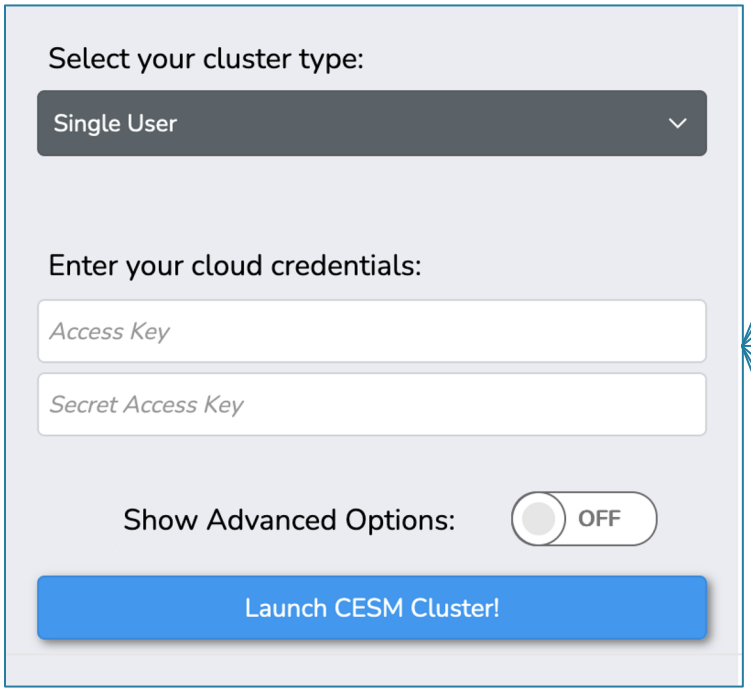
Benefits of an API

Standardization
Tested, common platform for workflows, training, etc

Support
One place to fix configurations ; ‘phone-home’ support

Metrics
Can (anonymously) track usage

Challenges



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

Thank You!

Brian Dobbins
bdobbins@ucar.edu