Audio trouble?

- Try logging out and back in
- If that doesn't fix it, try calling in

  +1 929-252-0881 PIN: 989 421 120#

- If you have bandwidth issues, Google Meet can call you - just click the 3 dots in the bottom right and choose "use phone for audio".

Please mute yourself if you are not speaking

Place any questions in the chat

Coastal Coupling Community of Practice Webinar Series

February 22, 2021

1:00 - 3:00 pm CT

www.weather.gov/watercommunity

Agenda

1. Introduction
2. Panelists
   - Patrick Tripp - cloud sandbox
   - Rich Signell - coastal coupling cloud pilot
   - Jena Kent - BDP
   - Dan Morris - Microsoft AI for Earth
3. Discussion
4. Future opportunities

CC CoP Data Need

Easily accessible, open-source, quality-controlled data at high resolution that is updated with a regular frequency for model initialization, verification, and validation

- Need community or community-contributed tools to read, process the high resolution open source data
- Data is needed particularly during extreme events (model development and validation)
- Focus should not be on access, but rather on regularly updating the data

Patrick Tripp

- Senior software engineer at RPS Group
- Leading IOOS cloud sandbox development
- B.S. in Computer Science from UC San Diego
- patrick.tripp@rpsgroup.com

IOOS Cloud Sandbox

Patrick Tripp, RPS Group

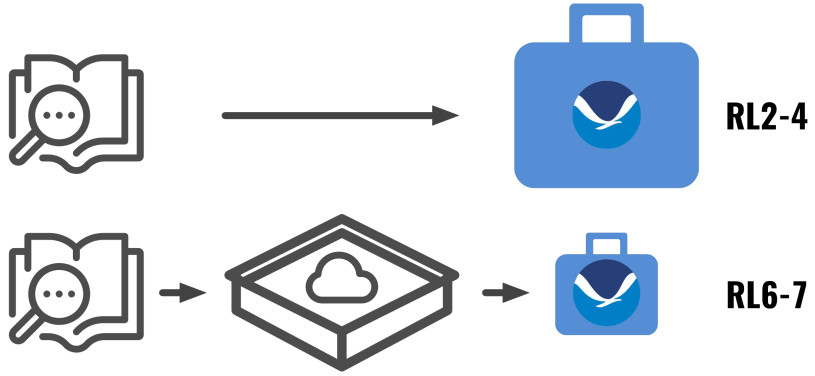
February 22, 2021

MAKING COMPLEX EASY

Sandbox Goals

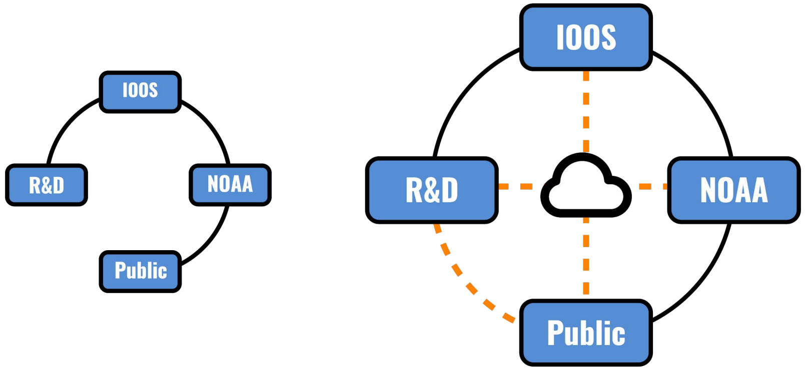
Improve Research To Operations

improve efficiency, speed and accessibility
sandbox supports integrated transition and effective use of significant new R&D products

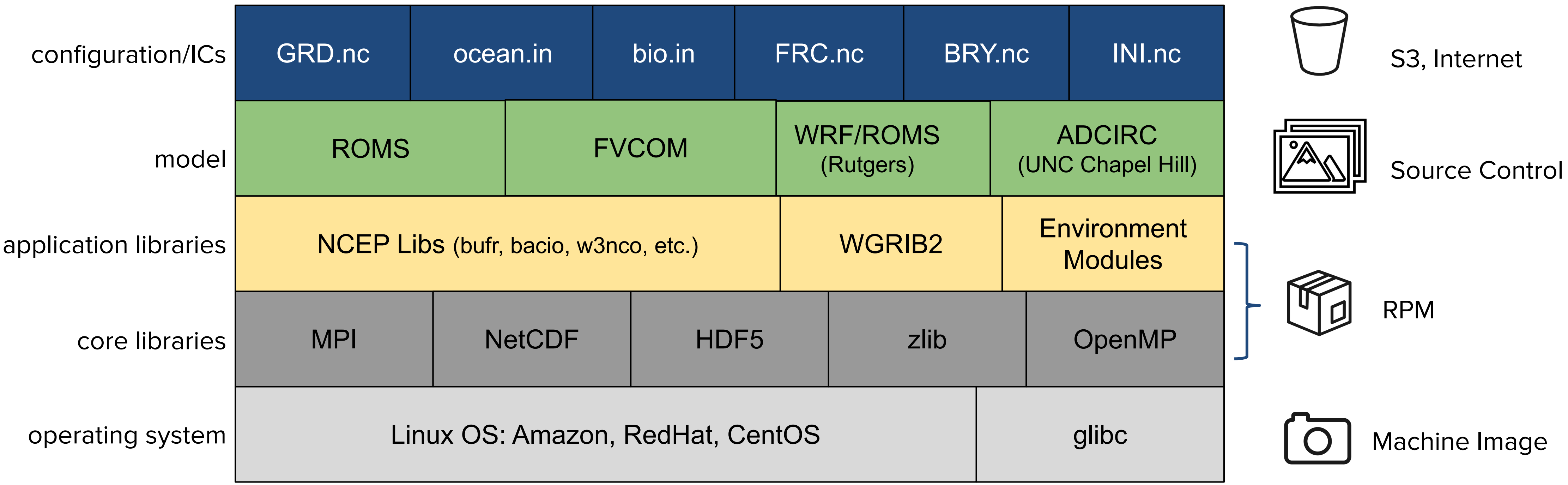


Complement Research To Operations

create environment for collaboration, innovation and iteration
dotted-line relationships complement solid-line.
decoupling opens new opportunities for improved R2O/O2R



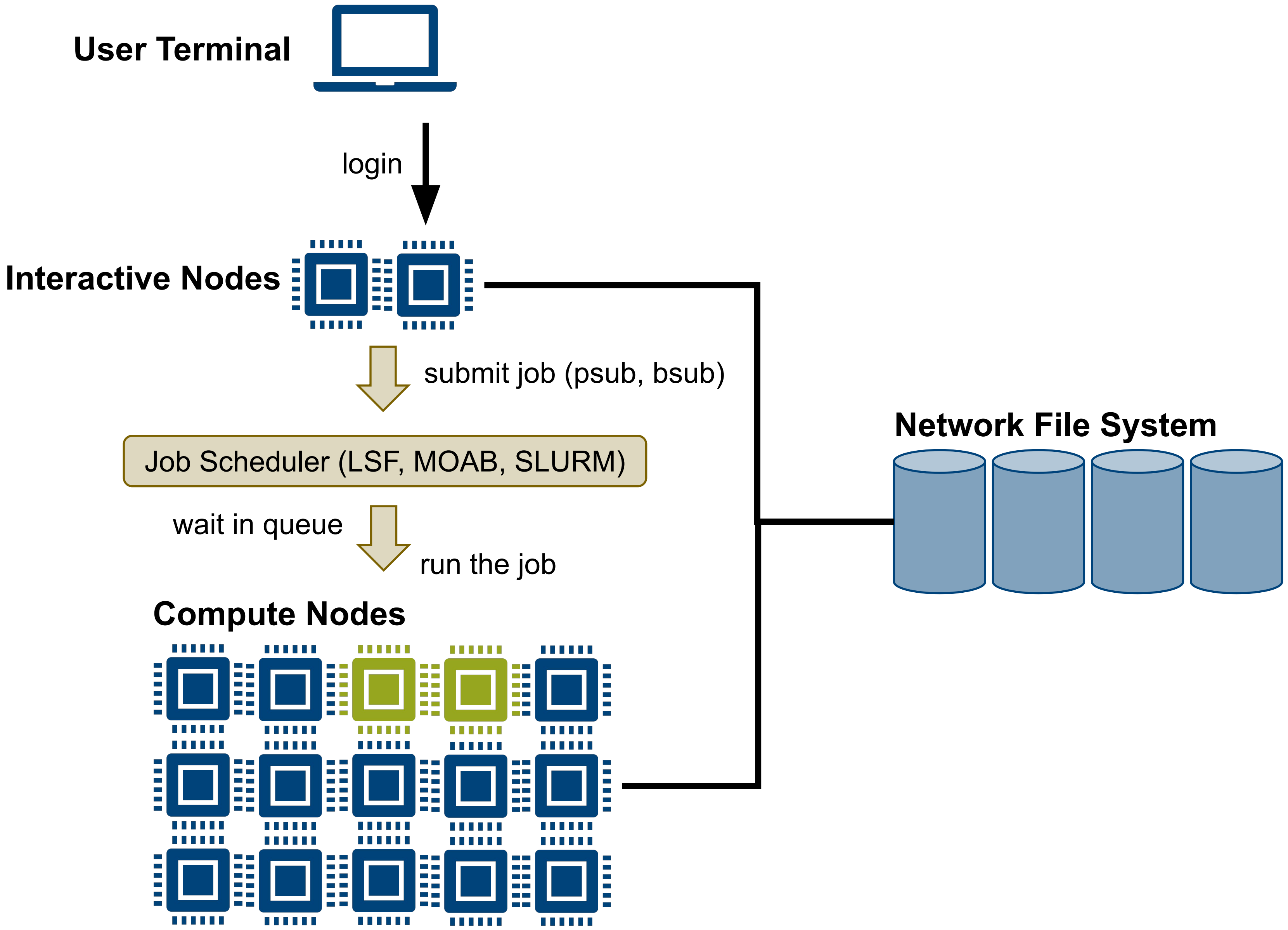
Sandbox Environment



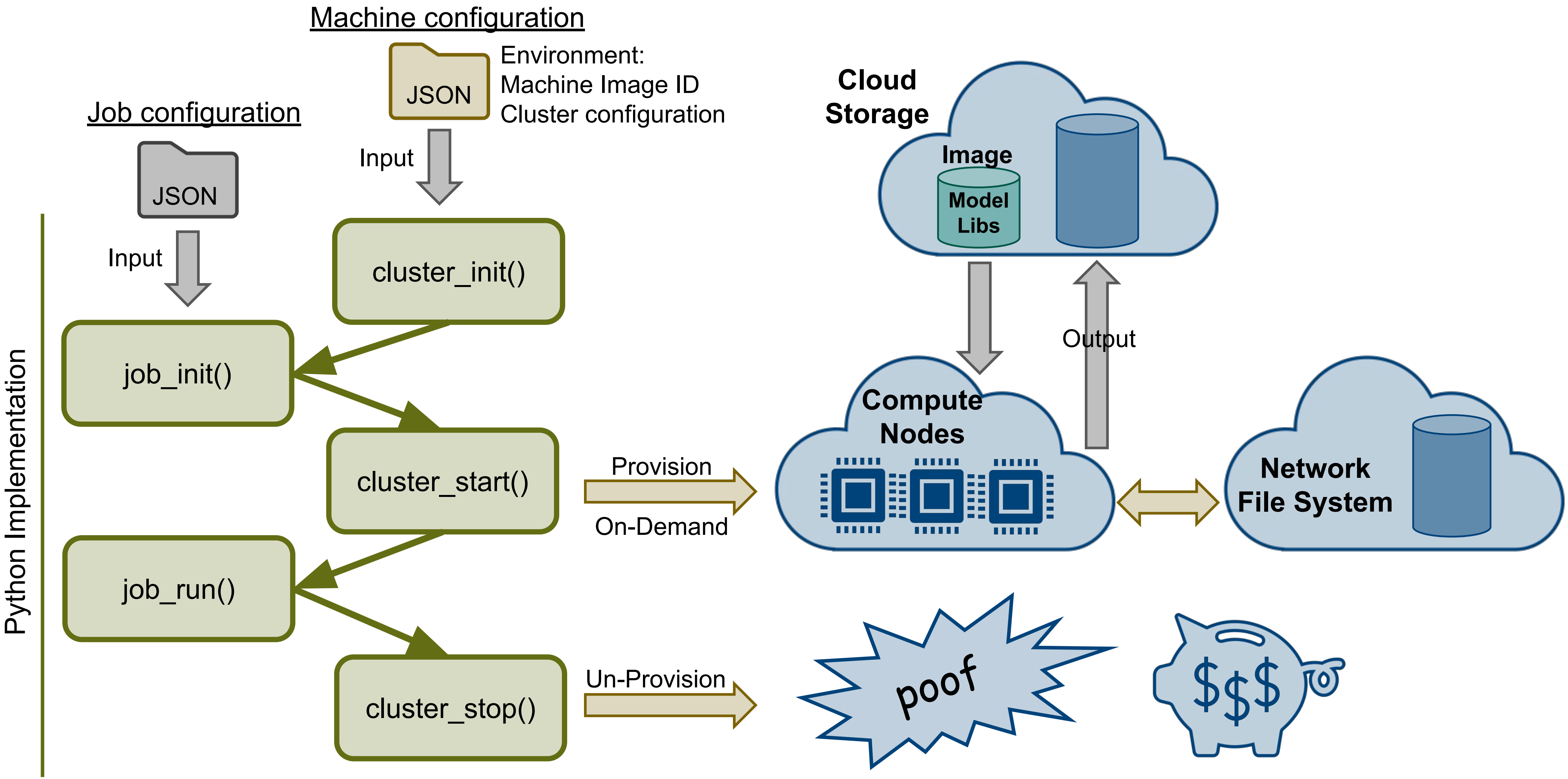
Currently Supported and Tested Models

- NOSOFS ROMS and FVCOM Models
  - CBOFS, CIOFS, DBOFS, GOMOFS, TBOFS, LEOFS, LMHOFS, NEGOFS, NGOFS, NWGOFS, and SFBOFS
  - Quasi-operational
- LiveOcean
  - Quasi-operational with operational fail-over triggering
- Coupled WRF/ROMS w/ ESMF v8
  - Hurricane Irene test case
- ADCIRC
  - Full test suite in ADCIRC repository
  - Hurricane Florence test case

Traditional Datacenter Cluster



On-demand Cloud Compute Cluster



Job: NPROCS, ROMS Tiling, depends on machine size.

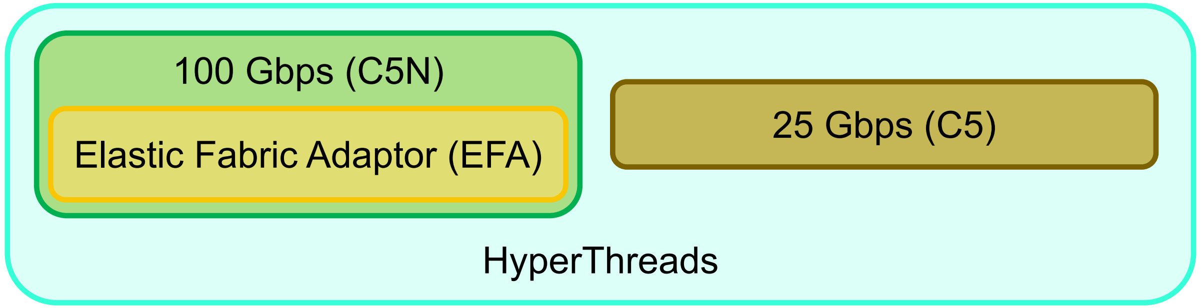
AWS HPC Performance Testing

AWS provides a maze of different options. What options and configurations are most practical for this application?

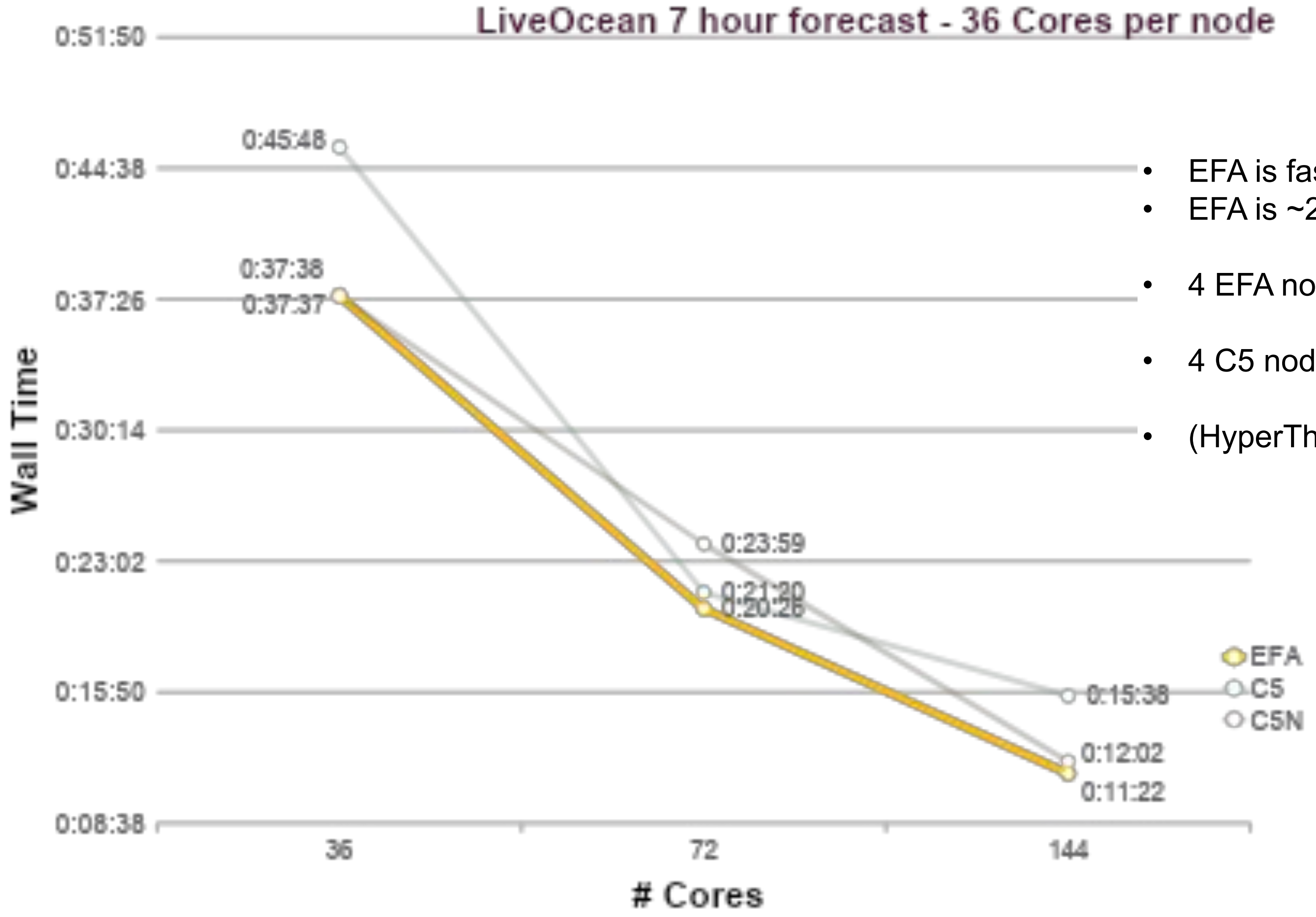
<u>Cluster Configurations</u>

- 1 Node
  - 36 Cores (6x6 tiling)
  - 72 VCPUs (9x8 tiling) w/ HyperThreads
- 2 Node
  - 72 Cores (9x8 tiling)
  - 144 VCPUs (12x12 tiling) w/ HyperThreads
- 4 Node
  - 144 Cores (12x12 tiling)
  - 288 VCPUs (18x16 tiling) w/ HyperThreads

<u>Node Configurations</u>



- C5: 25 Gbps Network
- C5N: 100 Gbps Network
- C5N with EFA: Low-latency, high-throughput for MPI
- HyperThreads: 2 hardware threads per core



- EFA is fastest with more than 1 node
- EFA is ~27% faster than C5
- 4 EFA nodes is ~3.36 times faster than 1 EFA node
- 4 C5 nodes is ~3 times faster than 1 C5 node
- (HyperThreads not viable - crashes on 4 nodes)

EFA : Elastic Fabric Adaptor
C5 : 25 Gbps network
C5N : 100 Gbps network

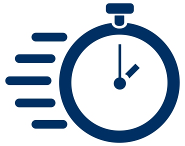
AWS performance is faster than University of Washington's HPC cluster.

Practical Use Case: If UW cluster suffers an outage AWS can fill the gap in a timely manner.

LiveOcean : AWS EC2 Instance Cost Comparison

Instance Type	AWS Price/Hour
c5.18xlarge :	$ 3.06
c5n.18xlarge :	$ 3.888

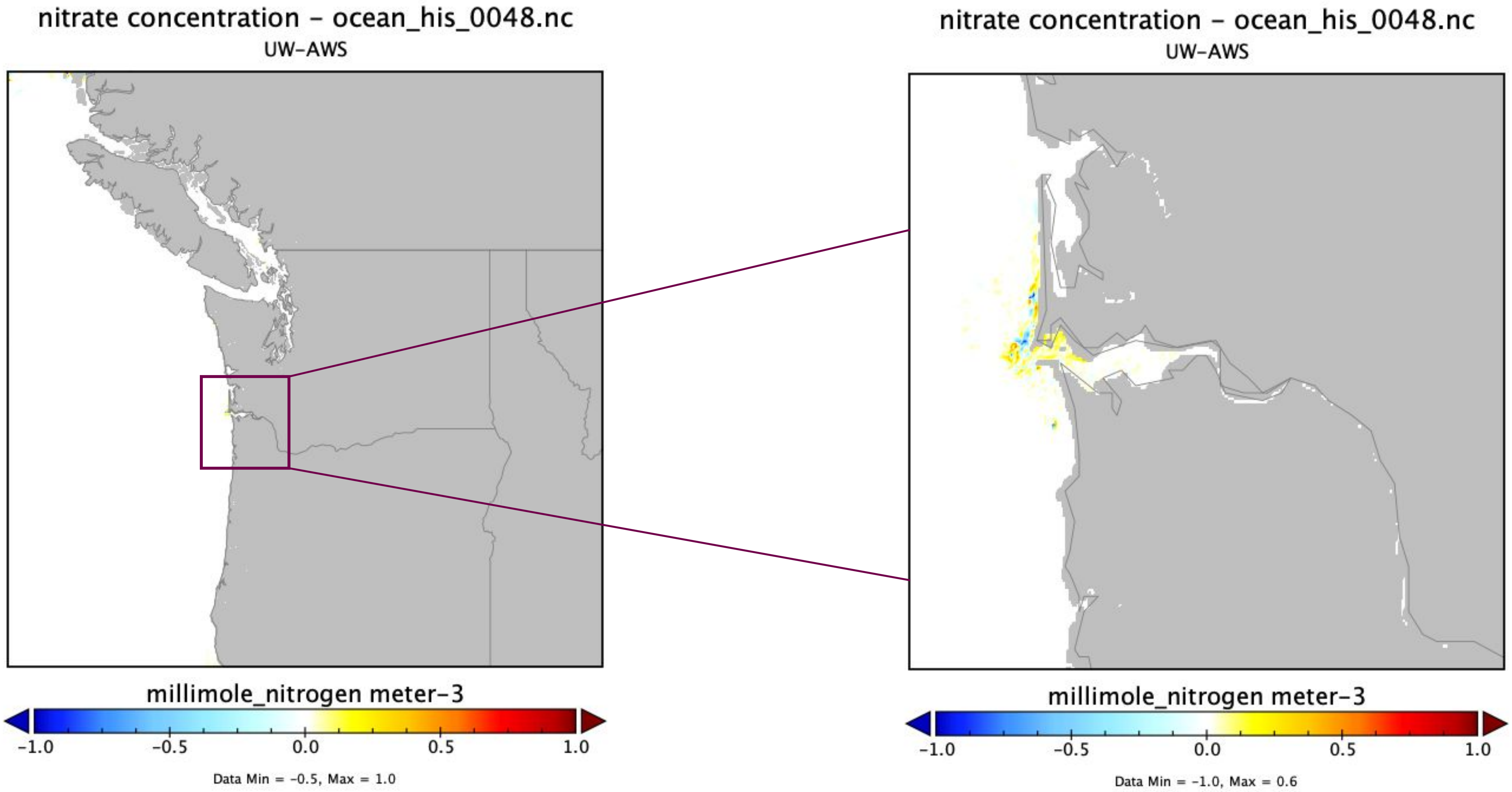
EFA C5n 18x	LiveOcean	1 Node	2 Nodes	4 Nodes
	Test timing / fhr (minutes)	0:05:23	0:02:55	0:01:37
	73 hour forecast time (minutes)	393	213	118
	73 hour forecast cost	$ 25.47	$ 27.59	$ 30.59
	Monthly 1 fcst/day	$ 763.96	$ 827.82	$ 917.70



C5 18x	LiveOcean	1 Node	2 Nodes	4 Nodes
	Test timing / fhr (minutes)	0:06:33	0:03:03	0:02:14
	73 hour forecast time (minutes)	478	222	163
	73 hour forecast cost	$ 24.36	$ 22.69	$ 33.26
	Monthly 1 fcst/day	$ 730.77	$ 680.78	$ 997.76

What is the optimal configuration that meets the requirements?

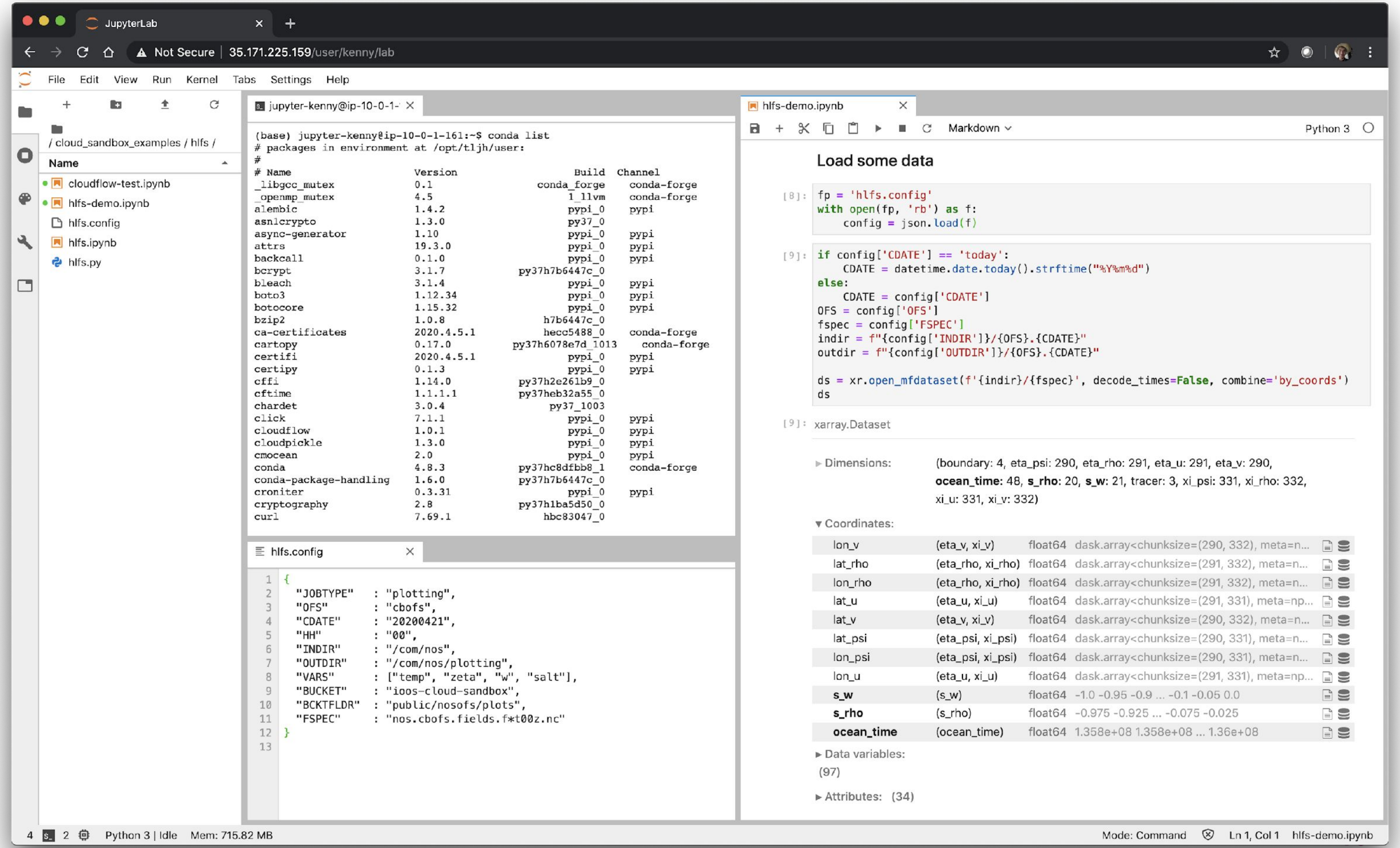
Difference



Small differences due to rounding of floating point values, sequence of operations, different tiling, etc.

Jupyter Lab Features

- File explorer
- Terminal
- Text editor
- Interactive notebooks

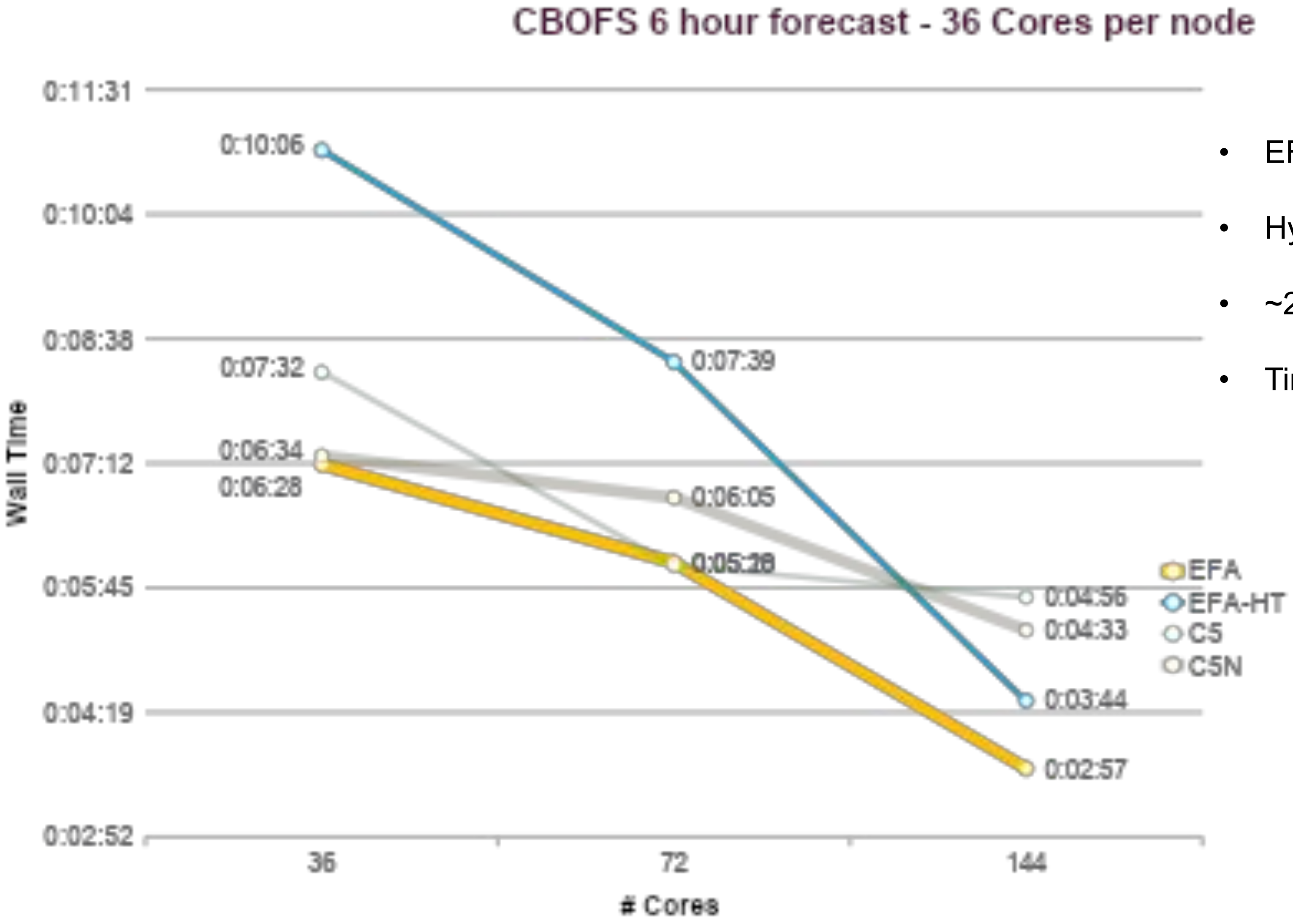


Links

GitHub Repositories

https://github.com/ioos/nosofs-NCO - copy of NOAA operational version with local changes
https://github.com/ioos/LiveOcean - scripts to retrieve, build, and run LiveOcean
https://github.com/ioos/Cloud-Sandbox - Python and BASH solutions
https://ioos.github.io/Cloud-Sandbox - API documentation for above

Questions and Comments



- EFA is usually fastest
- HyperThreads show no benefit
- ~2.16 times faster on 4 EFA nodes vs 1
- Times were fairly consistent

EFA : Elastic Fabric Adaptor
HT : HyperThreads
C5 : 25 Gbps network
C5N : 100 Gbps network

Does this performance meet the requirements?

CBOFS : AWS EC2 Instance Cost Comparison

Instance Type	AWS Price/Hour
c5.18xlarge :	$ 3.06
c5n.18xlarge :	$ 3.888

EFA C5n 18x

CBOFS	1 Node	2 Nodes	4 Nodes
Test timing / fhr (minutes)	0:01:05	0:00:53	0:00:29
48 hour forecast time (minutes)	52	43	24
48 hour forecast cost	$ 3.35	$ 5.53	$ 6.12
Monthly 1 fcst/day	$ 100.57	$ 165.89	$ 183.51

C5 18x

CBOFS	1 Node	2 Nodes	4 Nodes
Test timing / fhr (minutes)	0:01:15	0:00:53	0:00:49
48 hour forecast time (minutes)	60	42	39
48 hour forecast cost	$ 3.07	$ 4.32	$ 8.05
Monthly 1 fcst/day	$ 92.21	$ 129.74	$ 241.54

~2.5x speedup for ~2x cost.

Cluster Configuration JSON File

```
Example:
{
"platform" :"AWS",
"region"   :"us-east-1",
"nodeType" :"c5n.18xlarge",
"nodeCount":2,
"image_id" :"ami-0abc123abcdef012345”,
"tags"     : [{ "Key":"",”Value":””}],
"subnet_id":"subnet-0f1234abcdef8901",
"placement_group":"cloud-sandbox”,
"key_name” :”your_private_key”,
"sg_ids"   : ["sg-0012345678abc6b012”]
}
```

Job Configuration JSON File

```
Example:
{
"JOBTYPE"   :"forecast",
"OFS"       :"cbofs",
"CDATE"     :"today",
"HH"        :"00",
"COMROT"    :"/com/nos",
"TIME_REF"  :"20160101.0d0",
"BUCKET"    :"cloud-sandbox",
"BCKTFLDR"  :"/nos/cbofs",
"NTIMES"    :"34560",
"OUTDIR"    :"auto",
"OCEANIN"   :"auto",
"OCNINTMPL":"cbofs.ocean.in"
}
```

To run it: `./workflows/workflow_main.py job/jobs/cbofs.00z.fcst`

Rich Signell

- Research Oceanographer at USGS in Coastal and Marine Hazards and Resources Program
- Expertise in geoinformatics, physical oceanography, numerical modeling, Python, and web services
- rsignell@usgs.gov

https://wiki.esipfed.org/User:Rsignell

Jena Kent

- Oceanographer at NOAA's Center for Operational Oceanographic Products and Services (CO-OPS)
- Big Data Program Communications and User Engagement Lead
- jena.kent@noaa.gov

https://www.noaa.gov/organization/information-technology/big-data-program

NOAA Big Data Program

CC CoP Webinar | Monday, February 22, 2021

Jena Kent, Comms & Engagement Lead
NOAA Big Data Program | NOAA Office of Chief Information Officer

NOAA Team
Jonathan O'Neil | Adrienne Simonson | Patrick Keown

NOAA CISESS/NCICS Team
Otis Brown | Jonathan Brannock | Jenny Dissen

2 0002 HIMAWARI-8 2 21 JUN 19172 190000 01865 12455 01.00 McIDAS

INTRODUCTION

The NOAA Big Data Program (NOAA BDP)

Enables Innovation in Environmental Services using NOAA Data Accessed Through the Cloud Service Providers

TODAY….

- ➔ **NOAA BDP Overview and Status**
- ➔ **NOAA Ocean Data in the Cloud**
- ➔ **How Can You Get Involved**

NOAA Moving to the Cloud

- NOAA in the midst of a science and technology transformation
  - Newly approved NOAA Cloud Strategy
  - New NOAA Data Strategy
- Criticality of utilizing and leveraging emerging technologies for data access and analytics → meeting our users needs
- NOAA's growing observation sources and networks, scientific computing needs, and limitations of resources
- Enabling value of environmental data in socio-economic contexts for our society

NOAA BDP Leverages Partnerships

Provides access to 15PB of NOAA's environmental data as part of a 10 year contract established with CSPs for NOAA data in the cloud

Democratizes access to NOAA Data via Partnerships

- Reduces and removes obstacles to the public use of NOAA data, and help avoid costs and risks associated with federal data access services

Enables risk sharing, resources, rewards, leverage greater efficiencies in program delivery, skill sharing --- and accelerate innovation

Status of the Big Data Program

- BDP is under the NOAA OCIO, Chief Data Officer
- Program is operational and in year 2 of 10 year contract
- There are over 8 PB of NOAA data publicly accessible through 3 Cloud Service Providers ...over 145+ datasets in the cloud
- BDP has seen and enabled significant increase in data usage… supporting users and decision makers across various sectors of the economy

NOAA BIG DATA PROGRAM
Accelerating Access to Earth Data

Federal Security Boundary

Data transfer is one way

Data Broker

CISESS
Cooperative Institute for Satellite Earth System Studies
NC STATE UNIVERSITY

Only trusted user inside security boundary

Data Collaborators

- Microsoft Azure
- Google Cloud
- amazon web services

Users

- Research/ Academia
- Business
- Public
- Value-Add Providers

NOAA BDP Website

NOAA S-111 Surface Water Currents Data

oceans sustainability water

Description

S-111 is a data and metadata encoding specification that is part of the S-100 Universal Hydrographic Data Model, an international standard for hydrographic data. This collection of data contains surface water currents forecast guidance from NOAA/NOS Operational Forecast Systems, a set of operational hydrodynamic nowcast and forecast modeling systems, for various U.S. coastal waters and the great lakes. The collection also contains surface current forecast guidance output from the NCEP Global Real-Time Ocean Forecast System (GRTOFS) for some offshore areas. These datasets are encoded as HDF-5 files conforming to the S-111 specification, and are geospatially subset into individual tiles conforming to the NOAA/OCS Nautical Product Tiling Scheme, with filenames indicating the corresponding NOAA Electronic Navigational Chart (ENC) Cell Identifier. A full set of S-111 tiles is created for each new model run cycle, which occurs four times per day for all models except for RTOFS, which updates only once per day. Files are organized using a path naming convention that includes the OFS identifier (e.g. 'cbofs' corresponding with output from the Chesapeake Bay Operational Forecast System) as well as the year, month, day, and hour corresponding with each model run initialization time. Each individual S-111 (HDF-5) file contains all forecast projections from a single model run for that geographic area. In other words, a single S-111 file will contain multiple gridded arrays each containing a forecast valid at a distinct time in the future, out to the forecast horizon of the underlying modeling system. All surface currents forecasts in this collection are computed at a depth of 4.5 meters below water surface, or half the water column depth, whichever is shallower.

Update Frequency

Four times per day for each OFS, except Global RTOFS which updates once per day

License

Open Data. There are no restrictions on the use of this data.

Documentation

https://noaa-s111-pds.s3.amazonaws.com/README.html

Managed By

See all datasets managed by NOAA.

Contact

For any data delivery issues, please contact the NOAA Big Data Project Team at: noaa.bdp@noaa.gov. For general questions or feedback about the data, please submit inquiries through the NOAA Office of Coast Survey (OCS) ASSIST Tool at https://www.nauticalcharts.noaa.gov/customer-service/assist/.

Resources on AWS

Description
NOAA S-111 Surface Water Currents Datasets

Resource type
S3 Bucket

Amazon Resource Name (ARN)
`arn:aws:s3:::noaa-s111-pds`

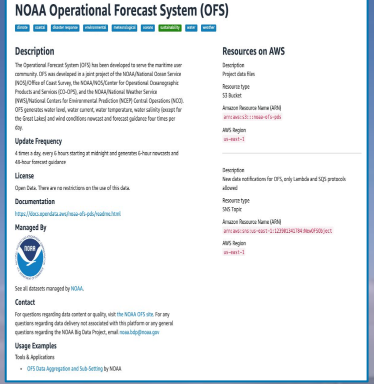
AWS Region
`us-east-1`

Description
NOAA S-111 Surface Water Currents New Dataset Notification

Resource type
SNS Topic

Amazon Resource Name (ARN)
`arn:aws:sns:us-east-1:123901341784:NewS111Object`

AWS Region
`us-east-1`



NOAA Operational Forecast System (OFS)

climate coastal disaster response environmental meteorological oceans sustainability water weather

Description

The Operational Forecast System (OFS) has been developed to serve the maritime user community. OFS was developed in a joint project of the NOAA/National Ocean Service (NOS)/Office of Coast Survey, the NOAA/NOS/Center for Operational Oceanographic Products and Services (CO-OPS), and the NOAA/National Weather Service (NWS)/National Centers for Environmental Prediction (NCEP) Central Operations (NCO). OFS generates water level, water current, water temperature, water salinity (except for the Great Lakes) and wind conditions nowcast and forecast guidance four times per day.

Update Frequency

4 times a day, every 6 hours starting at midnight and generates 6-hour nowcasts and 48-hour forecast guidance

License

Open Data. There are no restrictions on the use of this data.

Documentation

https://docs.opendata.aws/noaa-ofs-pds/readme.html

Managed By

See all datasets managed by NOAA.

Contact

For questions regarding data content or quality, visit the NOAA OFS site. For any questions regarding data delivery not associated with this platform or any general questions regarding the NOAA Big Data Project, email noaa.bdp@noaa.gov

Usage Examples

Tools & Applications

- OFS Data Aggregation and Sub-Setting by NOAA

Resources on AWS

Description
Project data files

Resource type
S3 Bucket

Amazon Resource Name (ARN)
`arn:aws:s3:::noaa-ofs-pds`

AWS Region
`us-east-1`

Description
New data notifications for OFS, only Lambda and SQS protocols allowed

Resource type
SNS Topic

Amazon Resource Name (ARN)
`arn:aws:sns:us-east-1:123901341784:NewOFSObject`

AWS Region
`us-east-1`

amazon web services

Improved data access via cloud services to support end users like mariners, maritime pilots, port authorities, and shipping companies optimally access and integrate disparate data sources to determine the best route when navigating congested waterways and ports… and assist with ship clearance and dredging

Improvements in Precision Navigation has historically saved shippers an estimated $10 million per year (NOAA 2017) --- Improving access and analytics to data and forecasts, will further these savings and efficiencies

Ocean Data in the NOAA Big Data Program

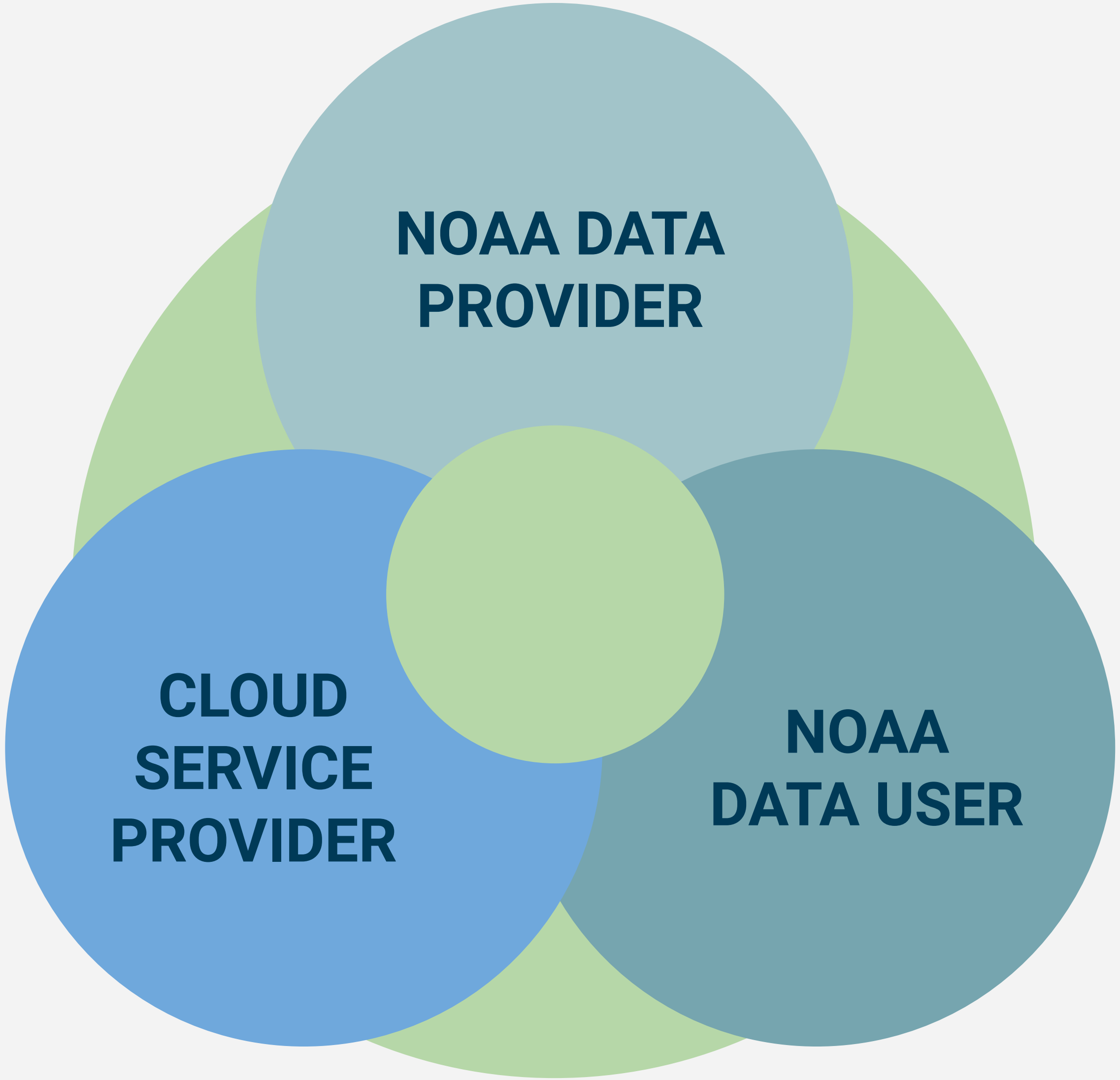
Ocean Datasets	Google	AWS	Microsoft
National Water Model	X	X	
Operational Forecast Systems		X	
NOAA S-111 Current Surface Data		X	
World Ocean Database	*	X	*
Tsunami	X		
Emergency Response Imagery		X	

**Planned*

Why Work with NOAA and BDP?

- Accelerate innovation in uses and applications in environmental sustainability areas and beyond
- Innovative data analytics and computational capabilities for earth system science
- Optimize cost and resource efficiencies
- Gain improved access to NOAA data for public use
- Support and revolutionize NOAA data delivery model
- Identify and stitch data issues and holes and support climate modeling

How Can You Get Involved?



- ➔ **Connect with the NOAA BDP Team:**
  - NOAA BDP Website
  - NOAA.BDP@NOAA.GOV
- ➔ **Explore the cloud service providers open data access pages**
  - Azure Open Data Catalog
  - AWS Registry of Open Data Sets
  - Google Cloud Public Datasets

Microsoft Azure

Google Cloud

amazon web services

NOAA Big Data Program

BDP Webpage | Data Intake Form

Thank you!

NOAA.BDP@NOAA.GOV

Jonathan O'Neil, NOAA OCIO, BDP Program Director
Adrienne Simonson, NOAA OCIO, Big Data Program Business Director
Patrick Keown, NOAA OCIO, Program Manager
Jena Kent, NOAA OCIO Communications and Engagement Lead
Otis Brown, Director, CISESS / NCICS / NC State University
Jonathan Brannock, IT Services, CISESS / NCICS / NC State University
Jenny Dissen, Engagement & Partnerships, CISESS / NCICS / NC State University

Appendix

NOAA Process for Working with BDP

- Initiate discussion with the EDMC Rep -

1

Identify your dataset

2

Submit a BDP request form (Intake Form)

3

Meet with BDP team to discuss details

4

Work with BDP and Data Broker (CISESS) to disseminate data on the cloud platform

5

Publicize the dataset's availability on the CSP when confirmation is received

Dan Morris

- Principal Scientist at Microsoft "AI for Earth" (and aspiring rock icon)
- Background in signal processing and machine learning for a variety of applications
- dan@microsoft.com

aka.ms/dan

Future Engagements

Date	Engagement	Other information
Mar 2	COMT proposal call	https://ioos.noaa.gov/about/funding-opportunities/
Mar 5	NSF proposal	https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=505594
31 Mar-1 Apr	Coastal Flood Modeling, Prediction and Observations for the U.S. West Coast	https://www.oceanvisions.org/2021-west-coastal-solutions
13-16 April	2nd NOAA General Modeling Meeting and Fair	https://www.star.nesdis.noaa.gov/ngmmf2021/
27-28 April	Coastal Flood Modeling, Prediction and Observations for the U.S. Gulf Coast	https://www.oceanvisions.org/2021-gulf-coast-solutions-workshop
10-14 May	ASFPM	www.asfpmconference.org
18-20 May	Ocean Visions 2021 Summit	https://www.oceanvisions.org/summit-2021
Early May	Annual meeting	

Thank you!

cayla.dean@noaa.gov

cell (865) 254-4098

www.weather.gov/watercommunity