AWS & Intel Research Webinar Series

Advancing the large scale weather and climate modeling data in the cloud

AWS & Intel Research Webinar Series

Introduction to AI/ML for research
Join this session for an introduction to AI/ML on AWS and see how AWS supports research AI/ML workloads. We will look at case studies and typical workflows for training and inference. Attendees will learn about Amazon SageMaker and see a case study from a research customer, in addition to a demo.
Speaker: Spencer Marley, Head of Business Development for AI/ML, Asia Pacific and Japan, Public Sector, AWS

Introduction to Quantum Computing for research
In this session, get an introduction to quantum computing and how industries and academic institutions are exploring using it to solve problems that are beyond the reach of classical computers. Check out a brief overview of Amazon Braket, a fully managed service from AWS that helps researchers to get started with quantum computing, including a walk-through of the AWS Management Console, the Amazon Braket SDK, and the multiple back-ends provided by AWS.
Speaker: Dr. Sanjay Padhi, Head of AWS Research, US Education, AWS

Enabling healthcare institution research in the cloud
Join this introductory session to learn how AWS is used by healthcare institutions and partners to support and enable care delivery and the intersection for researchers
Speakers: Dr. Julian Sham, Head of Health Business, Asia Pacific and Japan, AWS
Patrick C Brennan (Ph.D), Chair, Medical Imaging, Faculty of Medicine and Health, University of Sydney

Accelerating Genomics discoveries and enabling precision medicine in the cloud
Join this session for an overview of genomics on AWS, including solutions and customer examples for processing, analysis, and interpretation to accelerate scientific discoveries and enable precision medicine.
Speakers: Lisa McFerrin, Bioinformatician, AWS
Dr. Denis Bauer, Principal Research Scientist, CSIRO, Australia

AWS & Intel Research Webinar Series

Running molecular simulations and analysis workloads in the cloud
Join this session for an overview of how molecular dynamics research workflows with its associated applications (e.g. NAMD, GROMANCS) and tools (e.g. VMD) have adapted to operate on the AWS Cloud. The research scientists across the globe will be able to remotely run simulation workloads without any investment in local computing resources and reduce expertise requirement in setting up the HPC technologies.
Speakers: Austin Cherian, Technical Lead, High Performance Computing, Asia Pacific and Japan, Public Sector, AWS
Debamitra Chakravorty (Ph.D.), Project Lead – Computational Biology, Novel Techsciences (OPC) Private Limited, India

Towards next-generation computational fluid dynamics approaches on the cloud
This session will focus on how you can run CFD applications on AWS and how this can make a difference to its increase use in academia and industry. Go through the services that underpin this like Amazon EC2, Amazon S3 and FSx for Lustre. We will also have a live demo of running OpenFOAM and STAR-CCM+ on AWS.
Speaker: Dr. Neil Ashton, Principal Computational Fluid Dynamics Specialist Solution Architect, AWS

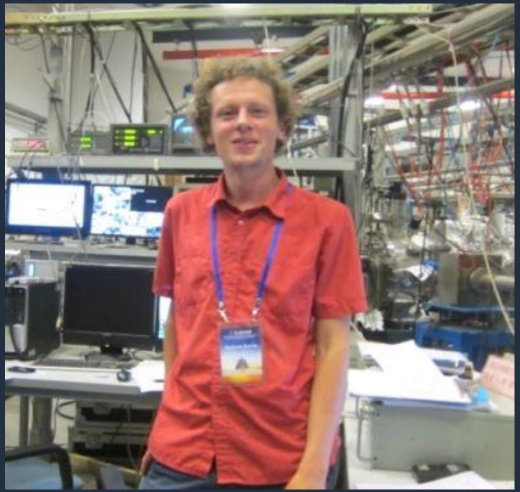
Advancing the large scale weather and climate modeling data in the cloud
In this session, learn about recent HPC results in large-scale numerical weather prediction and the AWS Earth on AWS public dataset repository with its NOAA and NASA datasets and sustainability initiatives. Also, discover community initiatives promising democratization of earth science resources, including Pangeo.
Speakers: Kevin Jorissen (Ph.D), Specialist Solutions Architect, High Performance Computing (HPC) and Numerical Weather Prediction, AWS
Justin Freeman, Senior Research Scientist, Bureau of Meteorology, Bureau of Meteorology, Australian Government

Please leave your questions in the chat box.
Our experts will answer thru chat
or during the live Q&A at the end.

AWS & Intel Research Webinar Series

Advancing the large scale weather and climate modeling data in the cloud

Speakers



Kevin Jorissen
Specialist Solutions Architect, High Performance Computing (HPC) and Numerical Weather Prediction, AWS

Justin Freeman
Senior Research Scientist, Bureau of Meteorology, Australian Government

Moderator

Austin Cherian
Technical Lead,
High Performance Computing,
Asia Pacific and Japan,
Worldwide Public Sector, AWS

AWS Research Webinar Series

Advancing large-scale weather and climate modeling in the cloud

Justin Freeman, Sr Research Scientist, Bureau of Meteorology (Australia)
Kevin Jorissen, Weather & Climate Lead, Amazon Web Services

PART 1:

State of the art of NWP and HPC on AWS

The what and why of it

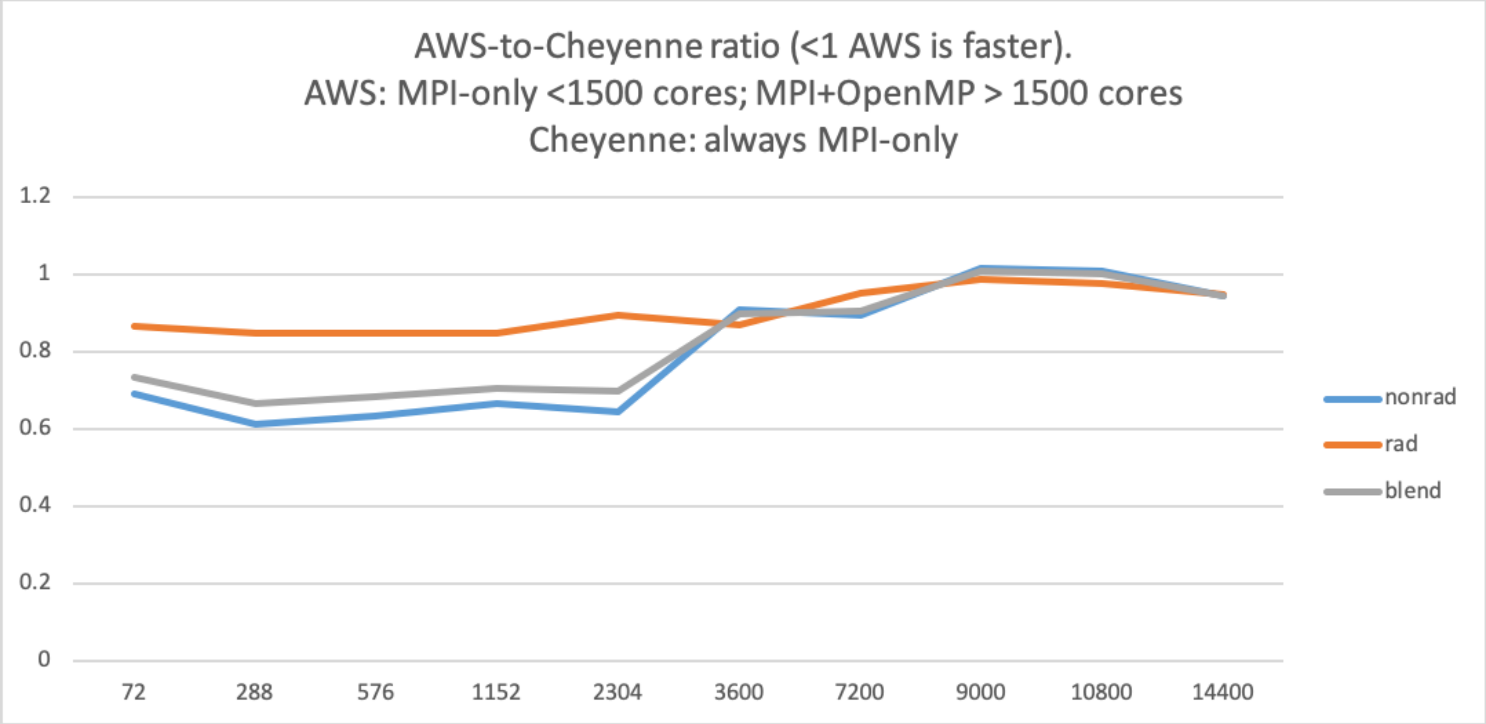
Similar HPC performance results

For other NWP models

WRF v.4

Work done with NCAR (Colorado, USA)

AWS matches turnaround time of new WRF benchmark on NCAR Cheyenne supercomputer up to 14,400 cores.



MPAS

Commercial weather customer deployed MPAS model on AWS at 3,000 – 4,000 cores. Performance matched or beat other HPC platforms.

GEOS-Chem

NAVGEM

Efficient high-performance computing on the Amazon Web Services cloud with over a thousand cores: application to global atmospheric chemistry simulations at 50-km resolution with GEOS-Chem 12.3.2

Jiawei Zhuang[1], Daniel J. Jacob[1], Haipeng Lin[1], Elizabeth W. Lundgren[1], Robert M. Yantosca[1], Judit Flo Gaya[1], Melissa P. Sulprizio[1], Sebastian D. Eastham[2], Kevin Jorissen[3]

[1]School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA
[2]Laboratory for Aviation and the Environment Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[3]Amazon Web Services, Seattle, WA 98109, USA

Correspondence to: Jiawei Zhuang (jiaweizhuang@g.harvard.edu)

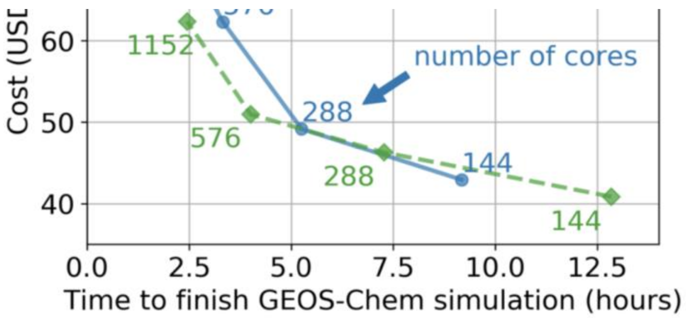
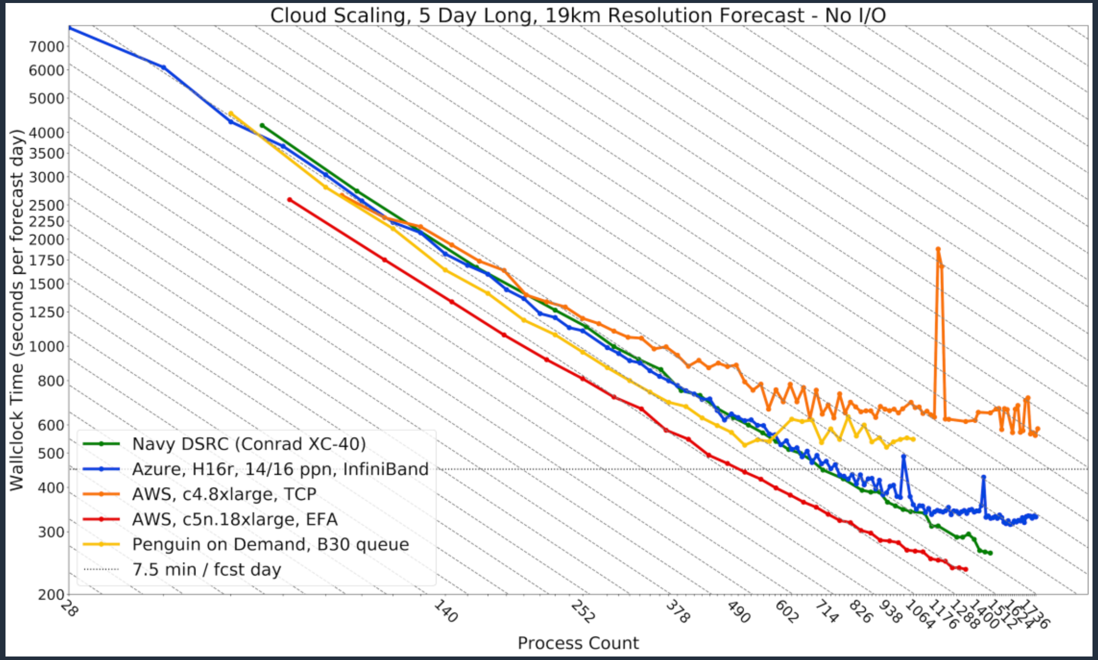


Figure 6. Cost of a 7-day global GEOS-Chem simulation of tropospheric-stratospheric chemistry at cubed-sphere C180 (≈50 km) resolution. The time required to finish the simulation (x-axis) varies with the number of cores (indicated as text next to the data point), which in turn affects the total cost because scalability is less than 100%,. Cost of the AWS EC2 cluster is compared to that of the NASA Pleiades cluster. The AWS EC2 time and cost are based on Intel-MPI and EFA performance and assume "us-east-1" spot pricing. The NASA Pleiades cost is based on the Standard Billing Unit (SBU) model.



US Navy ran NAVGEM benchmarks at ~1,500 core scale. AWS was most performant platform.
https://www.youtube.com/watch?v=GTHWf0OVGrw

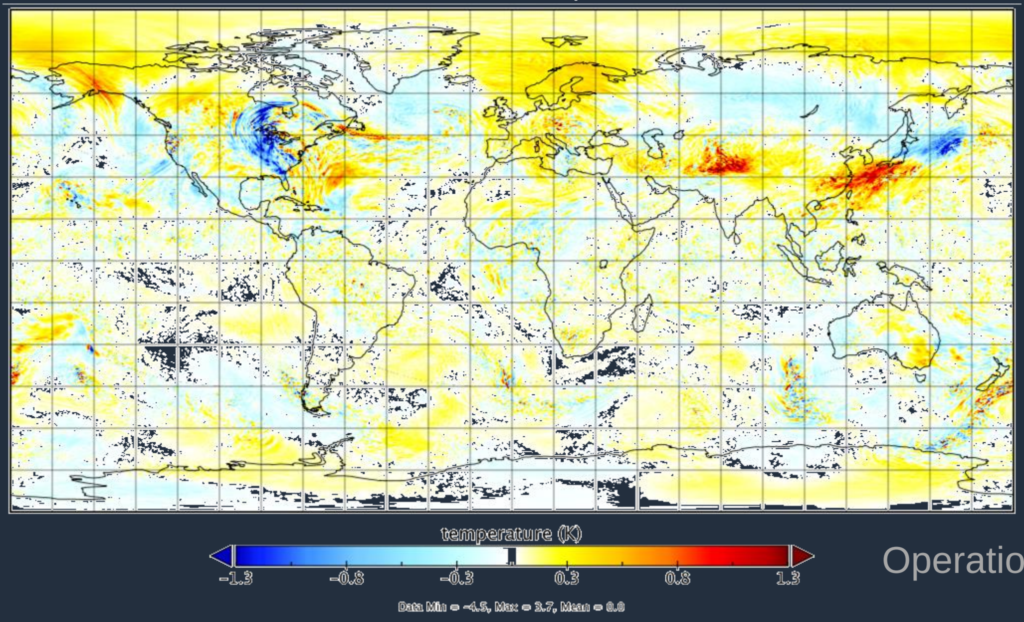
https://jiaweizhuang.github.io/blog/aws-hpc-guide/

FV3

Work by JCSDA:

- UFS (pre-)operational C768 (13km) global configuration
- 10-day forecast with NEMSfv3gfs on 48 AWS c5.x18large nodes
  - 36 cores, 144GB mem, 25GB/s network, Intel 19 with iMPI
  - 1728 cores in total
- model run takes 7 min 22s real time per simulated day

JEDI temperature increment propagated with
24hr UFS forecast valid at 00z on April 16, 2018, near 200 hPa

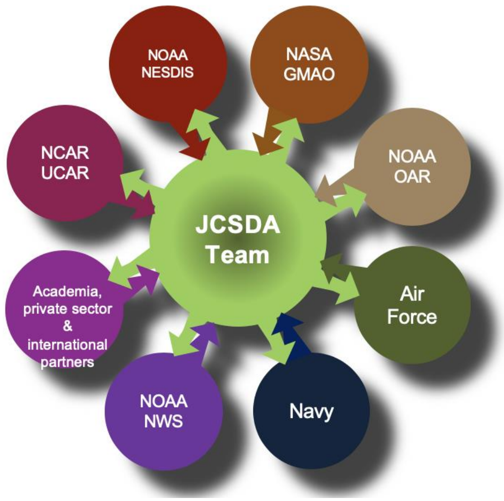


Operational GFS v15

Additional HPC work by JCSDA:
Benchmarks of JEDI using 864 MPI threads using singularity on Amazon EC2

Summary of JCSDA Cloud Activities

https://www.jcsda.org
https://www.jcsda.org/jcsda-project-jedi



- **Infrastructure as a Service**
  - Collaborative code management
  - Unified build system (EcBuild)
  - Unified workflow (Rapids)
  - Containers
  - Automated Testing

AWS ParallelCluster
Amazon EC2
docker
Charliecloud
Travis-CI
DockerHub
AWS CodeBuild
AWS ECR

- **Platform as a Service**
  - Training
  - Data Warehouse
  - Serverless plots on-demand
  - NRT Web applications
  - AI/ML

jupyter
amazon S3
DynamoDB
Amazon API Gateway
TensorFlow
Keras
Amazon SageMaker

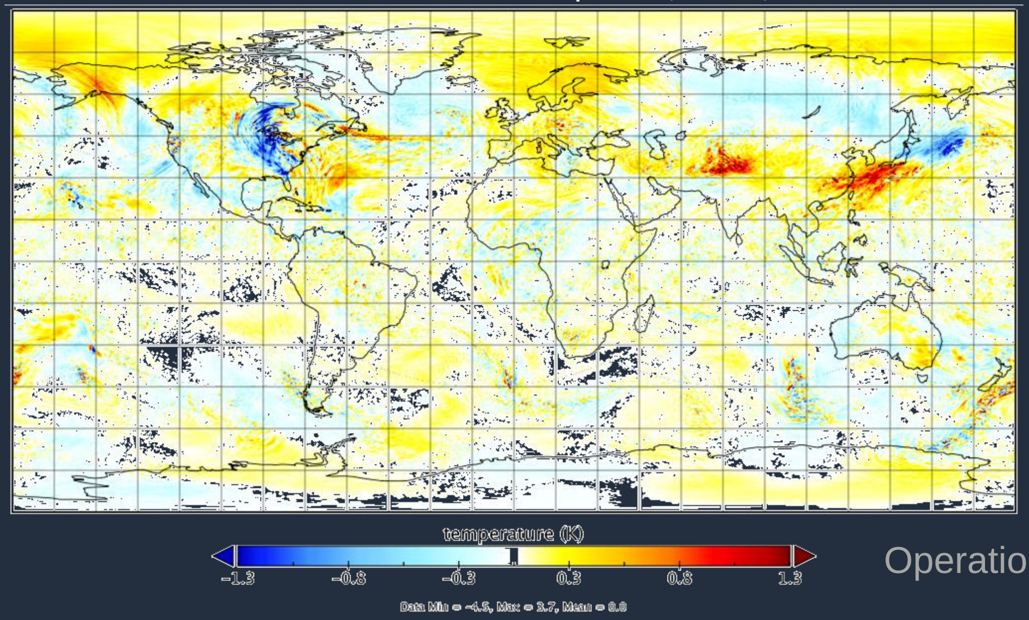
Courtesy of Tom Auligné, Director, JCSDA

FV3

Work by JCSDA:

- UFS (pre-)operational C768 (13km) global configuration
- 10-day forecast with NEMSfv3gfs on 48 AWS c5.x18large nodes
  - 36 cores, 144GB mem, 25GB/s network, Intel 19 with iMPI
  - 1728 cores in total
- model run takes 7 min 22s real time per simulated day

JEDI temperature increment propagated with
24hr UFS forecast valid at 00z on April 16, 2018, near 200 hPa



Operational GFS v15

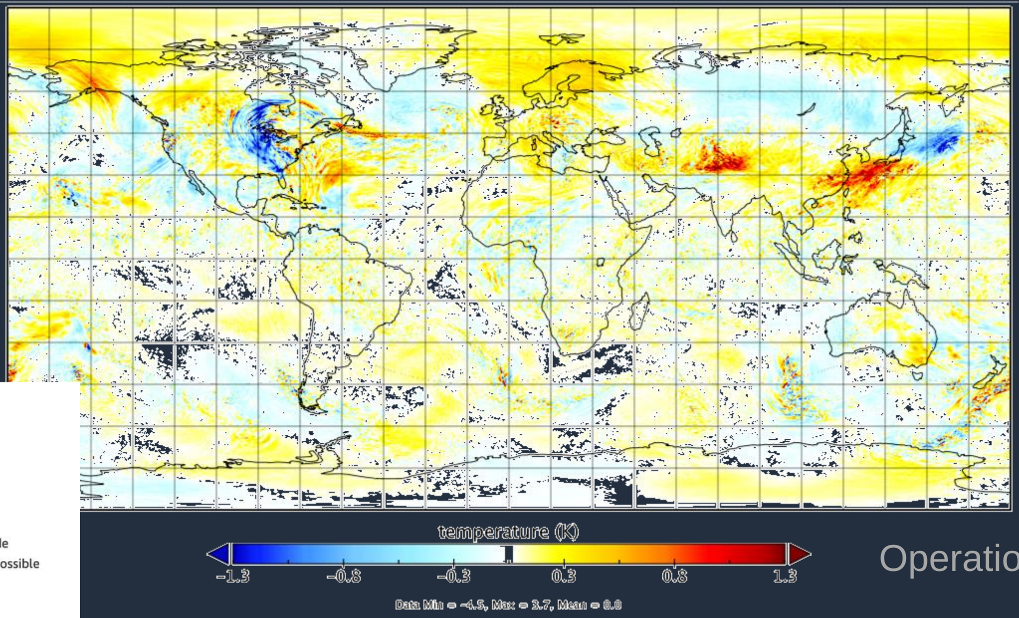
Additional HPC work by JCSDA:
Benchmarks of JEDI using 864 MPI threads using singularity on Amazon EC2

FV3

Work by JCSDA:

- UFS (pre-)operational C768 (13km) global configuration
- 10-day forecast with NEMSfv3gfs on 48 AWS c5.x18large nodes
  - 36 cores, 144GB mem, 25GB/s network, Intel 19 with iMPI

JEDI temperature increment propagated with
24hr UFS forecast valid at 00z on April 16, 2018, near 200 hPa



Operational GFS v15

Maxar Uses AWS to Deliver Forecasts 58% Faster Than Weather Supercomputer

2020

When weather threatens drilling rigs, refineries, and other energy facilities, oil and gas companies want to move fast to protect personnel and equipment. And for firms that trade commodity shares in oil, precious metals, crops, and livestock, the weather can significantly impact their buy-sell decisions. To limit damage, these companies need the earliest possible notice before a major storm strikes. That's the challenge Maxar Technologies set out to solve.

Historically, many industries have relied on reports generated by the on-premises supercomputer operated by the National Oceanic and Atmospheric Administration (NOAA). However, the weather predictions take an average of 100 minutes to process global data. Over time, many companies began to realize they would require much faster weather warnings to protect their interests. Similar to how NASA has expanded its partnerships with private firms to acquire commercial space hardware and services, the processing and delivery of critical weather data products could also be effectively commercialized.

"With the fast networking speed provided by AWS, we accomplished what many IT experts considered impossible."

Stefan Cecelski
Data Scientist, Maxar Technologies

Additional operational FV3 results reported by Maxar:
https://aws.amazon.com/solutions/case-studies/maxar-case-study/
Daily FV3 forecasts on >5,000 cores (156x c5n.18xlarge)
Forecast delivered > 2x faster (42 min)
See AWS and Maxar presentations at UFS workshop next week

Performance for other HPC codes…

Lots of HPC codes, including tightly coupled ones, routinely run at several-thousand-core-per-job scale with comparable performance to on-premises HPC, including Unified Model, WRF, NAVGEM, FV3, MPAS, GEOSChem, CBAM.

Beware of outdated studies.

We see a performance step-up with each new generation of servers.

There are similar results for computational fluid dynamics, molecular dynamics, genomics, machine learning…

https://d1.awsstatic.com/whitepapers/architecture/AWS-HPC-Lens.pdf: Well-Architected best principles for HPC on AWS

Big picture for AWS and research computing

And reasons to do your compute in the cloud beyond base HPC performance

The long-term trends in scientific computing/HPC

Increased diversity of

- HPC practitioners
- hardware platforms
- applications and frameworks

It's a data-driven new world

- Data gravity: massive volumes of data
- Cross-disciplinary research
- Research Data Management: compliance, security, reproducibility

The public cloud (AWS) has the right characteristics to address these challenges.

Key strengths of AWS for scientific discoveries

Improve time-to-results

- Compute resources are available anytime
- Build scientific pipelines faster with advanced services for ML, data analytics, workflows…

Collaboration and big data

- Store massive data sets
- And share them with your collaborators
- With compute/analytics/ML available
- In a secure and compliant way

Open data on AWS

AWS hosts a selection of data sets that anyone can access for no cost.

https://aws.amazon.com/public-datasets/

Life science

- TCGA and ICGC
- 1000 Genomes
- Genome in a Bottle
- Human Microbiome Project
- 3000 Rice Genome

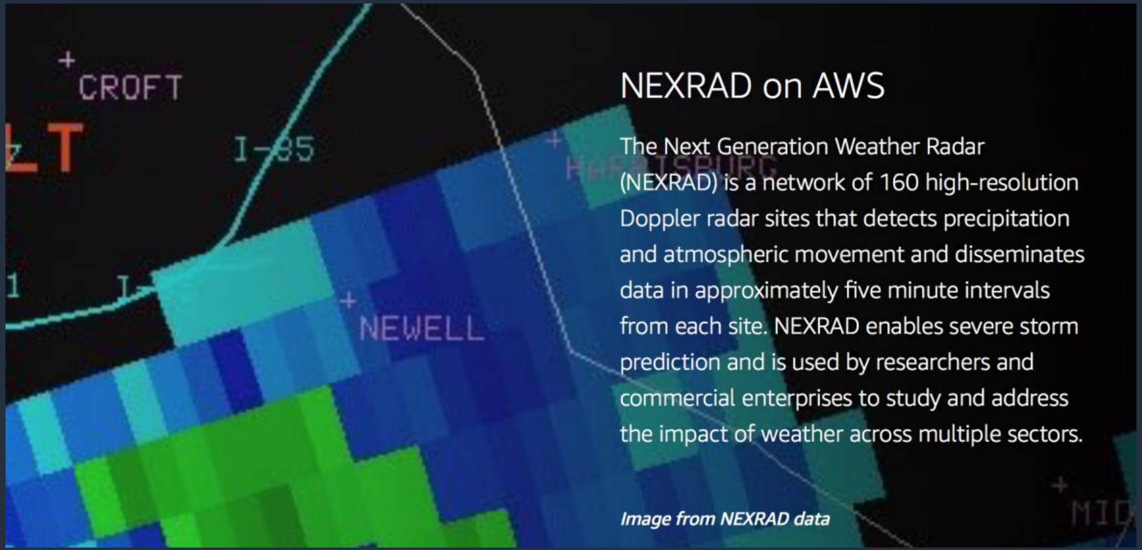
Earth science

- Landsat
- NEXRAD
- NASA NEX

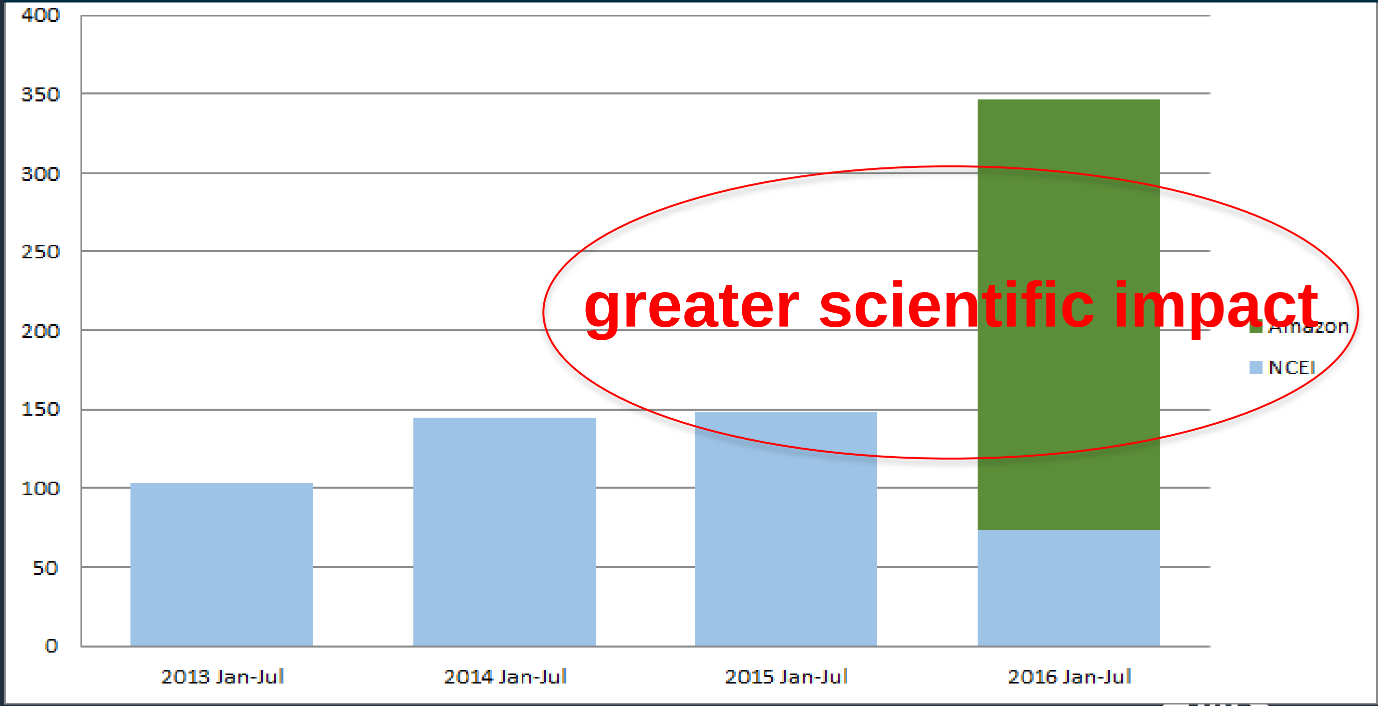
Internet science

- Common Crawl Corpus
- Google Books Ngrams
- Multimedia Commons

NOAA Big Data program



NEXRAD on Amazon S3, usage increased 2.3x

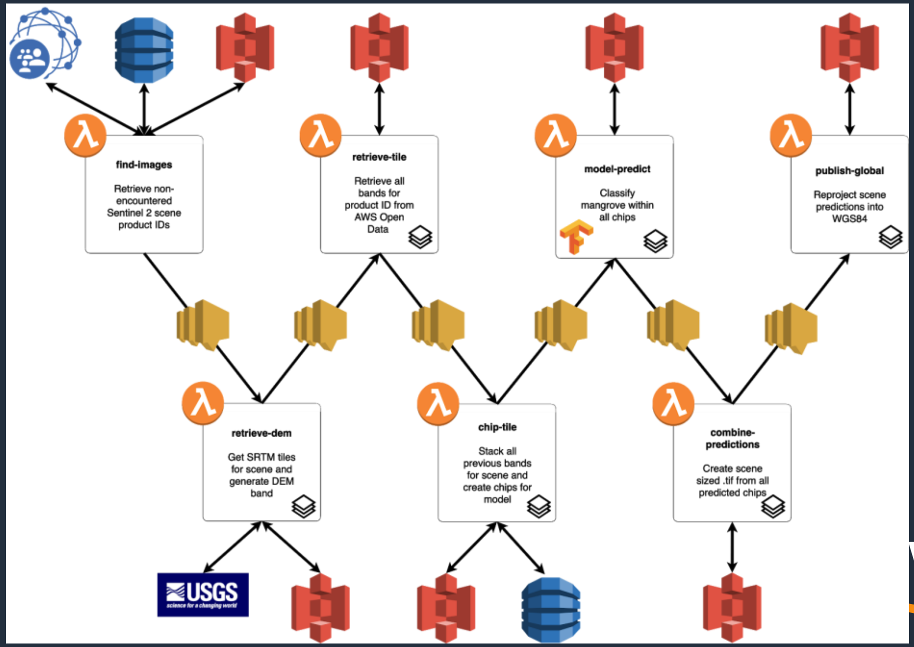


Collaborating on scientific data in the cloud

UK Hydrologic Office – Sentinel-2 data

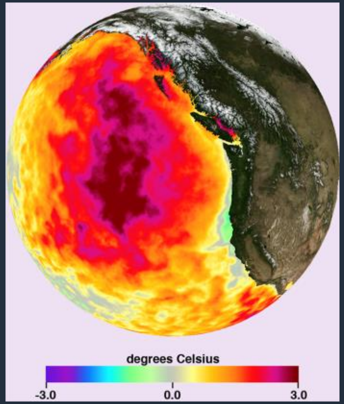
The UK Hydrologic Office used the Sentinel-2 satellite data shared via the AWS Public Dataset program to ***identify mangroves around the world***: https://medium.com/uk-hydrographic-office/creating-a-global-dataset-using-serverless-applications-and-deep-learning-c4e267fa810c

"Our compute provider of choice for this pipeline was AWS. The strongest deciding factor for this was the availability of the European Space Agency's Sentinel 2 (S2) imagery, which is available as an open data set, detailed here. The data set gives almost complete global coverage at 10m resolution. Having the availability of the data set beside our compute without storing it ourselves was a winner for us."



NASA Space Act Agreement

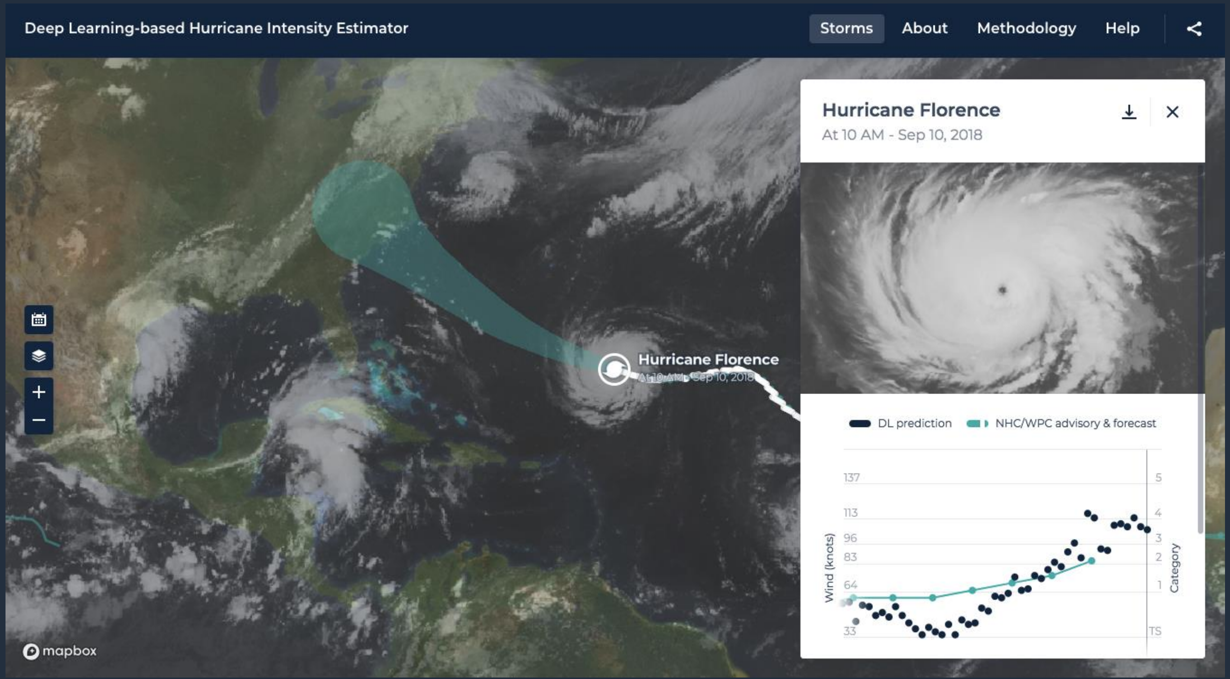
https://registry.opendata.aws/collab/nasa/ – "Multi-Scale Ultra High Resolution ***Sea Surface Temperature" dataset***:



Dr. Chelle Gentemann, sr. scientist at Farallon Institute: "Something that took me three months and 3,000 lines of code now takes me 10 minutes with 20 lines of code. Now you don't have to have a big supercomputer and a system administrator to figure out how to download, store, and access the data. This is a transformative technology that's paving the way for the democratization of science."

https://earthdata.nasa.gov/learn/articles/tools-and-technology-articles/mur-sst-in-the-cloud

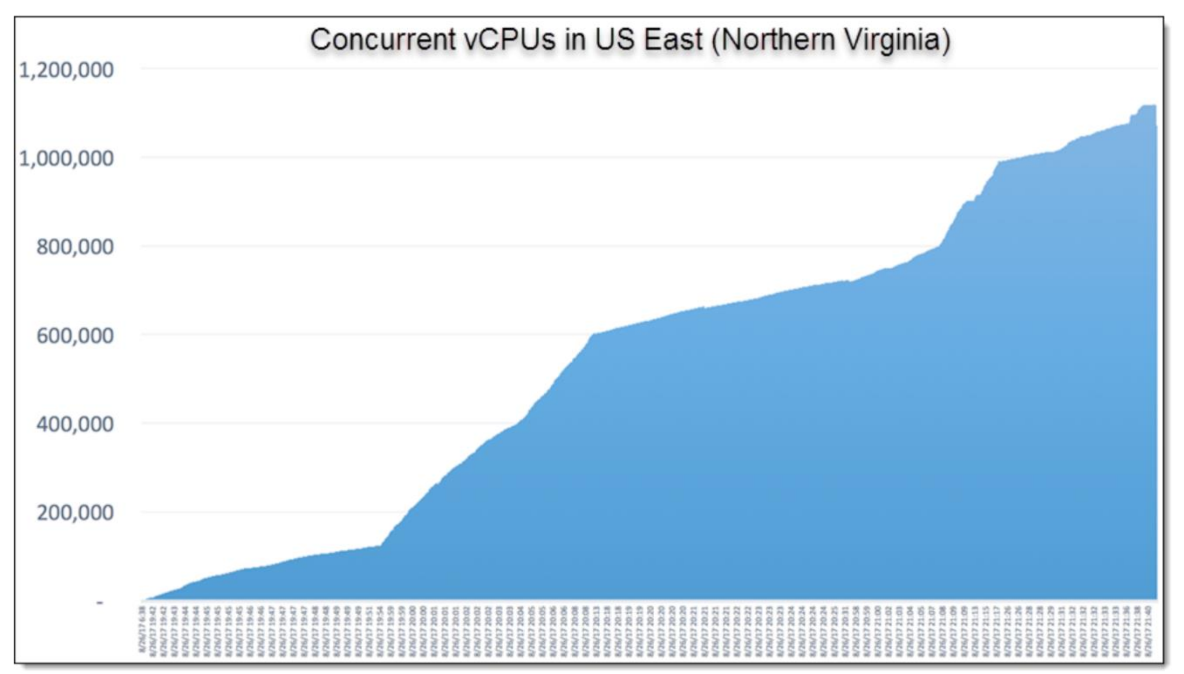
Estimating hurricane wind speeds with ML

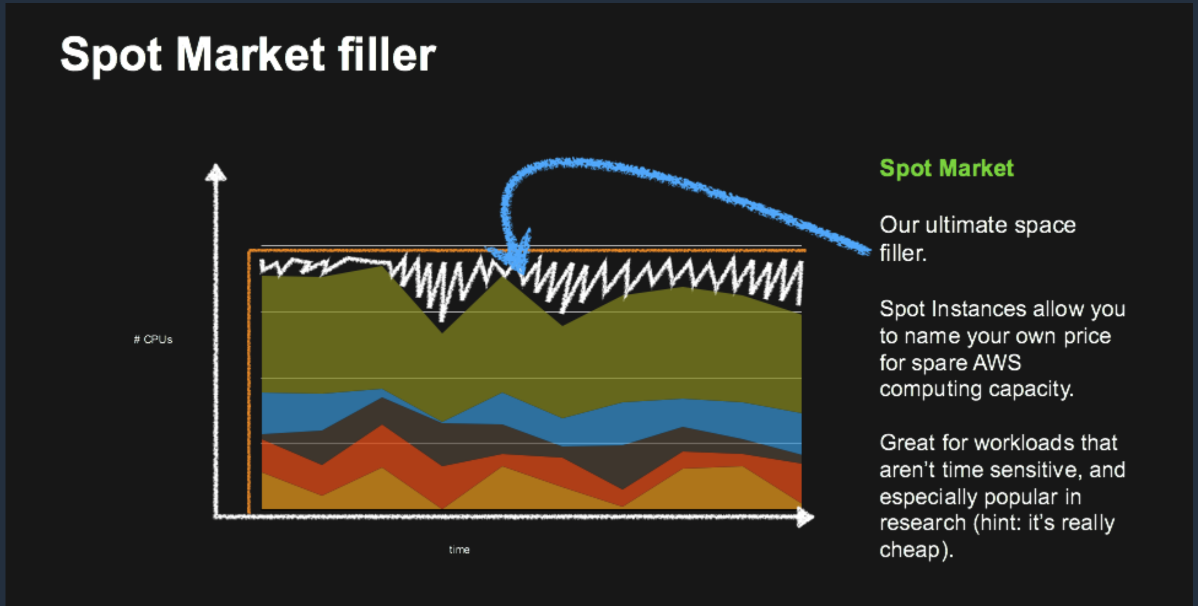


Uses the GOES16 satellite dataset on AWS, updated every 15 minutes. More accurate and much faster. (6h → 15 min)

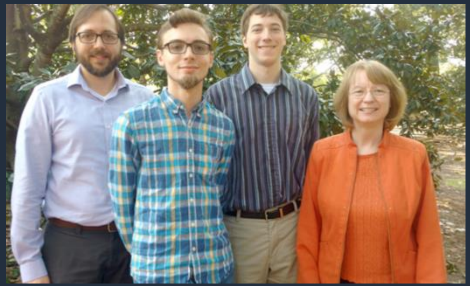
https://aws.amazon.com/blogs/publicsector/estimating-hurricane-wind-speeds-with-machine-learning/?did_trk=psr_card&trk=psr_card

Your very own national supercomputer: 550,000 cores for NLP at Clemson U





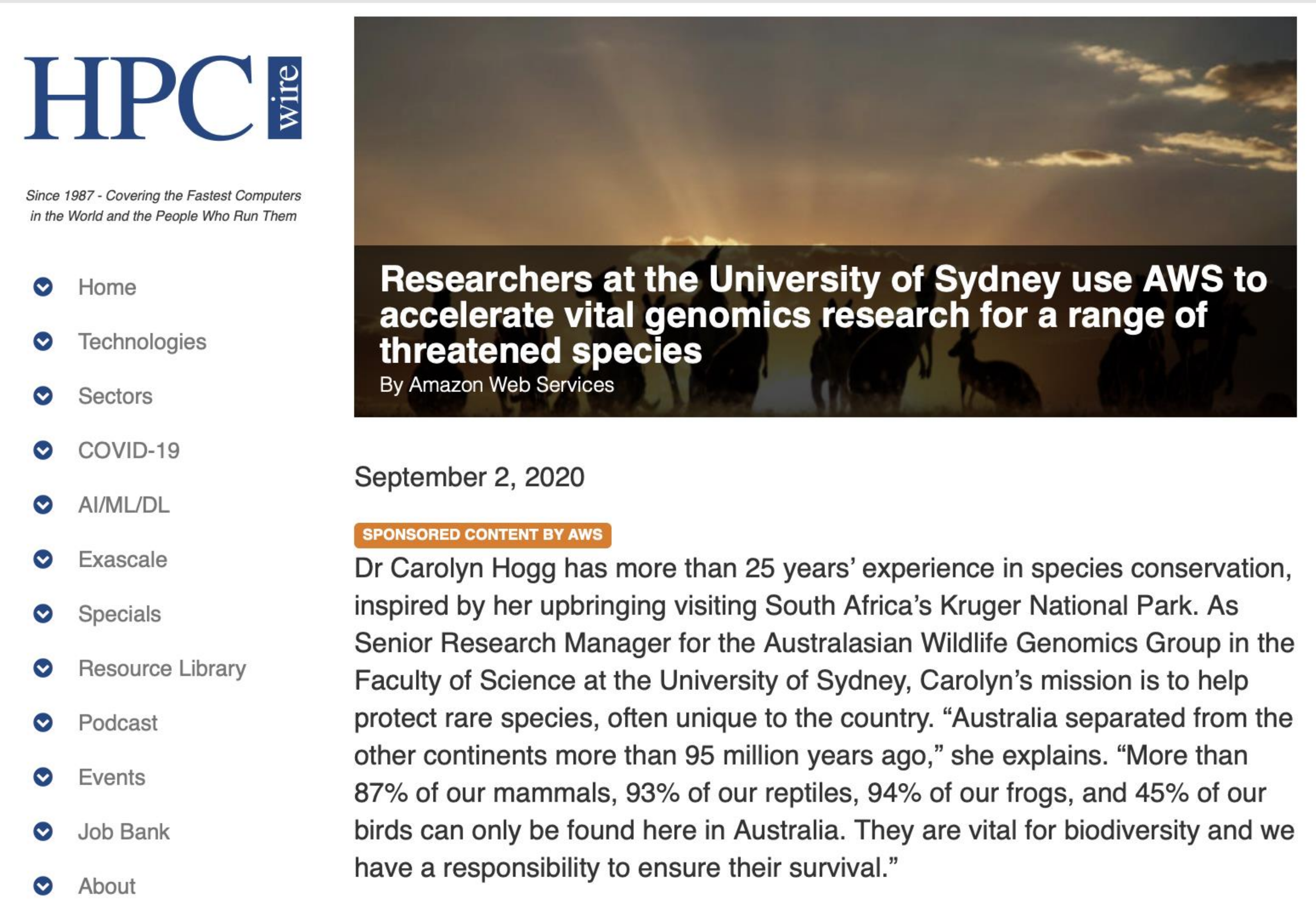
"spot market": cheap AWS computing –a good fit for research



"I am absolutely thrilled with the outcome of this experiment. The graduate students on the project […] used resources from AWS and Omnibond and developed a new software infrastructure to perform research at a scale and time-to-completion not possible with only campus resources." – Prof. Amy Apon, Co-Director of the Complex Systems, Analytics and Visualization Institute

https://aws.amazon.com/blogs/aws/natural-language-processing-at-clemson-university-1-1-million-vcpus-ec2-spot-instances/

Sydney Uni & cute endangered mammals



HPCwire

Since 1987 - Covering the Fastest Computers in the World and the People Who Run Them

- Home
- Technologies
- Sectors
- COVID-19
- AI/ML/DL
- Exascale
- Specials
- Resource Library
- Podcast
- Events
- Job Bank
- About

Researchers at the University of Sydney use AWS to accelerate vital genomics research for a range of threatened species

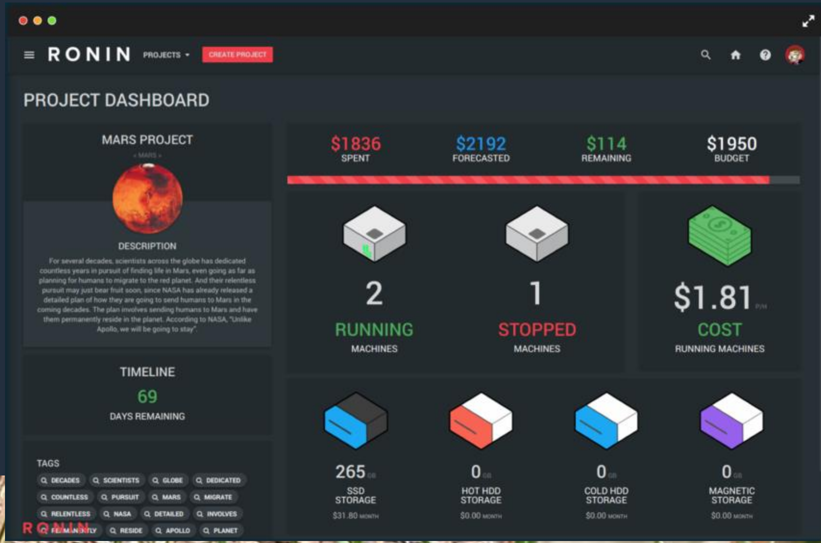
By Amazon Web Services

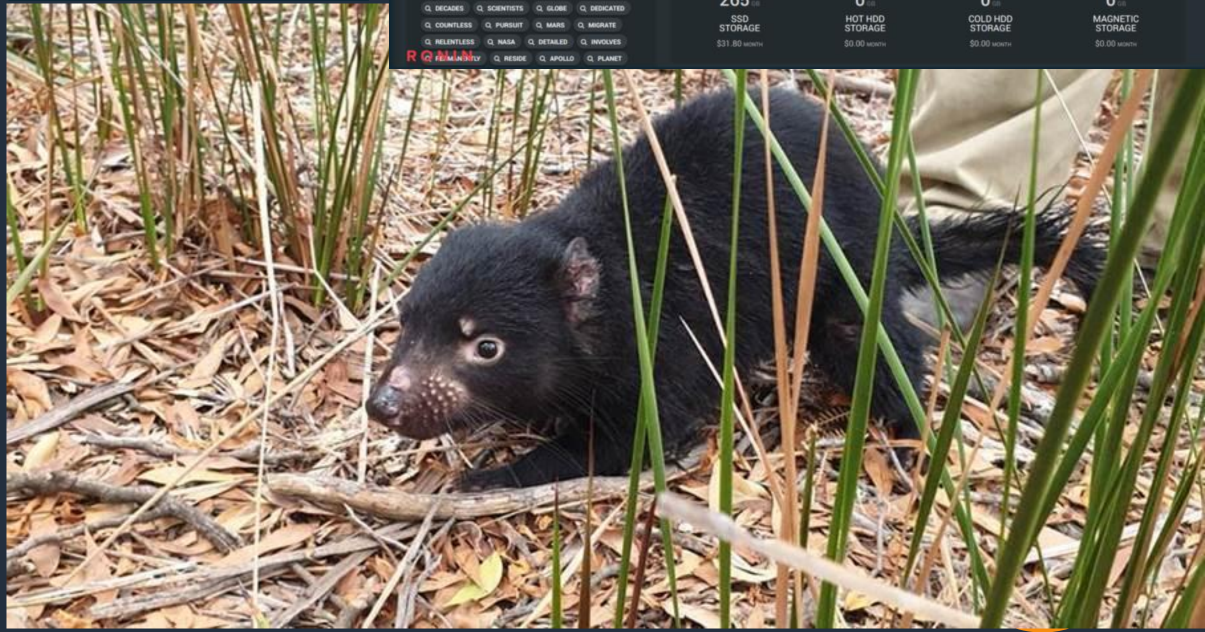
September 2, 2020

SPONSORED CONTENT BY AWS

Dr Carolyn Hogg has more than 25 years' experience in species conservation, inspired by her upbringing visiting South Africa's Kruger National Park. As Senior Research Manager for the Australasian Wildlife Genomics Group in the Faculty of Science at the University of Sydney, Carolyn's mission is to help protect rare species, often unique to the country. "Australia separated from the other continents more than 95 million years ago," she explains. "More than 87% of our mammals, 93% of our reptiles, 94% of our frogs, and 45% of our birds can only be found here in Australia. They are vital for biodiversity and we have a responsibility to ensure their survival."

RONIN IS A CLOUD ORCHESTRATION PLATFORM, LOWERING THE ENTRY LEVEL TO USING THE CLOUD





https://www.youtube.com/watch?v=PVC-63I29nY&feature=youtu.be

Getting to publication and conservation faster

nature genetics

ARTICLES

https://doi.org/10.1038/s41588-018-0153-5

OPEN

Adaptation and conservation insights from the koala genome

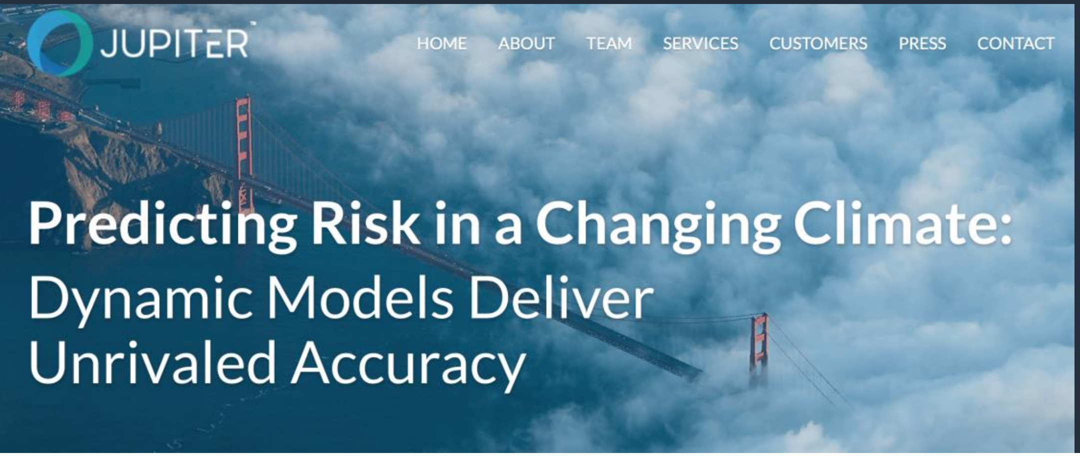
Rebecca N. Johnson[1,2,30,31*], Denis O'Meally[2,3,30], Zhiliang Chen[4,30], Graham J. Etherington[5], Simon Y. W. Ho[2], Will J. Nash[5], Catherine E. Grueber[2,6], Yuanyuan Cheng[2,7], Camilla M. Whittington[8], Siobhan Dennison[1], Emma Peel[2], Wilfried Haerty[5], Rachel J. O'Neill[9], Don Colgan[1], Tonia L. Russell[10], David E. Alquezar-Planas[1], Val Attenbrow[1], Jason G. Bragg[11,12], Parice A. Brandies[2], Amanda Yoon-Yee Chong[5,13], Janine E. Deakin[14], Federica Di Palma[5,15], Zachary Duda[9], Mark D. B. Eldridge[1], Kyle M. Ewart[1], Carolyn J. Hogg[2], Greta J. Frankham[1], Arthur Georges[14], Amber K. Gillett[16], Merran Govendir[8], Alex D. Greenwood[17,18], Takashi Hayakawa[19,20], Kristofer M. Helgen[1,21], Matthew Hobbs[1], Clare E. Holleley[22], Thomas N. Heider[9], Elizabeth A. Jones[8], Andrew King[1], Danielle Madden[3], Jennifer A. Marshall Graves[11,14,23], Katrina M. Morris[24], Linda E. Neaves[1,25], Hardip R. Patel[26], Adam Polkinghorne[3], Marilyn B. Renfree[27], Charles Robin[27], Ryan Salinas[4], Kyriakos Tsangaras[28], Paul D. Waters[4], Shafagh A. Waters[4], Belinda Wright[1,2], Marc R. Wilkins[4,10,30], Peter Timms[29,30] and Katherine Belov[2,30,31]

The koala, the only extant species of the marsupial family Phascolarctidae, is classified as 'vulnerable' due to habitat loss and widespread disease. We sequenced the koala genome, producing a complete and contiguous marsupial reference genome, including centromeres. We reveal that the koala's ability to detoxify eucalypt foliage may be due to expansions within a cytochrome P450 gene family, and its ability to smell, taste and moderate ingestion of plant secondary metabolites may be due to expansions in the vomeronasal and taste receptors. We characterized novel lactation proteins that protect young in the pouch and annotated immune genes important for response to chlamydial disease. Historical demography showed a substantial population crash coincident with the decline of Australian megafauna, while contemporary populations had biogeographic boundaries and increased inbreeding in populations affected by historic translocations. We identified genetically diverse populations that require habitat corridors and instituting of translocation programs to aid the koala's survival in the wild.

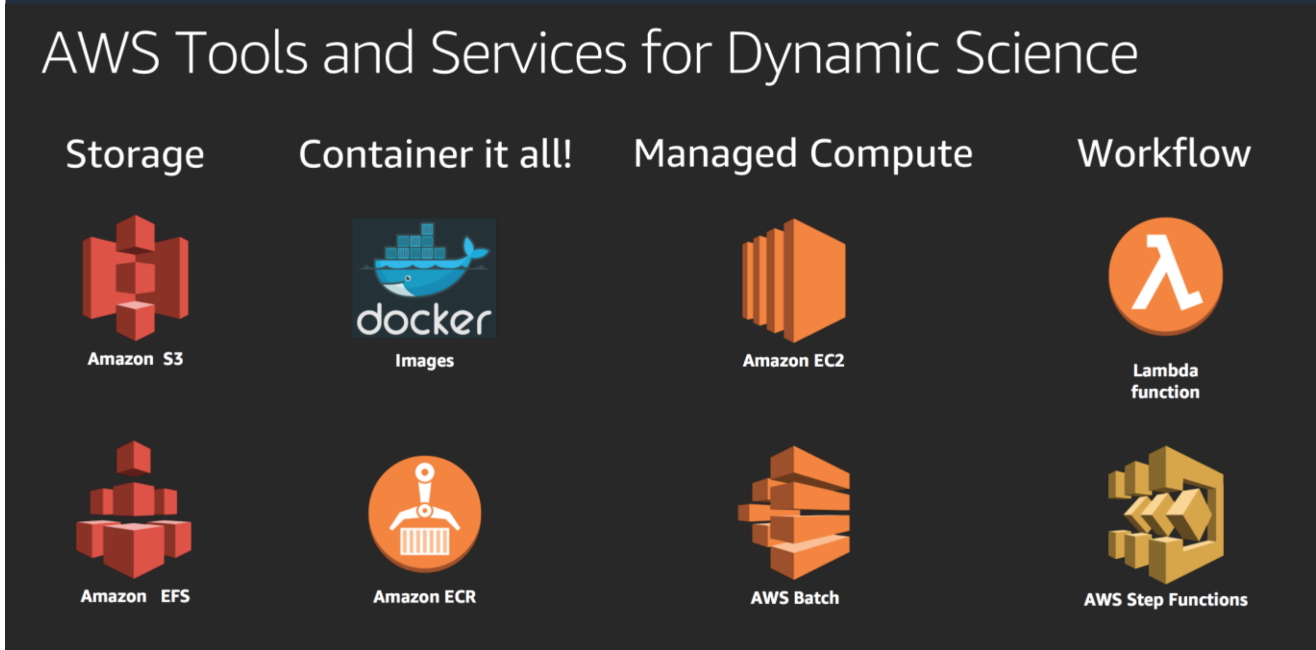
length of the reads at the 60% percentile was calculated as 10,889 bp. The FALCON assembly was run on Amazon Web Service Tokyo region using r3.8xlarge spot instances as compute node, with the number of instances varying from 12 to 20 depending on availability.

https://aws.amazon.com/blogs/aws/saving-koalas-using-genomics-research-and-cloud-computing/

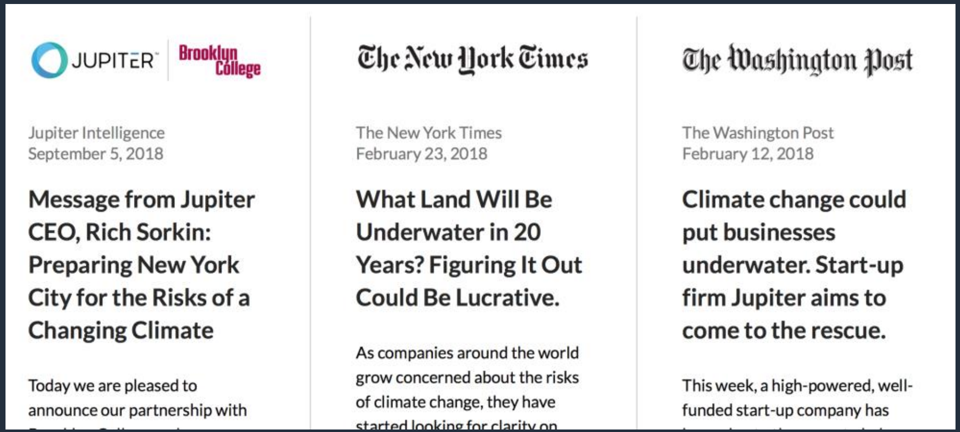
Startups in HPC using managed services

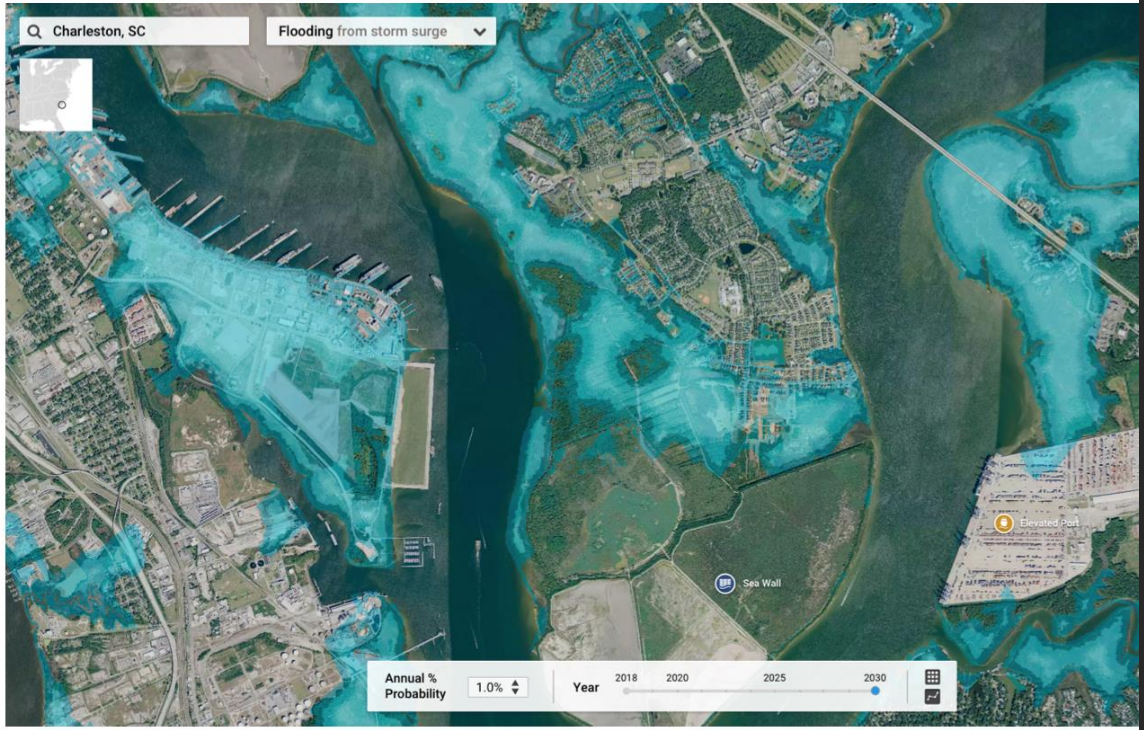


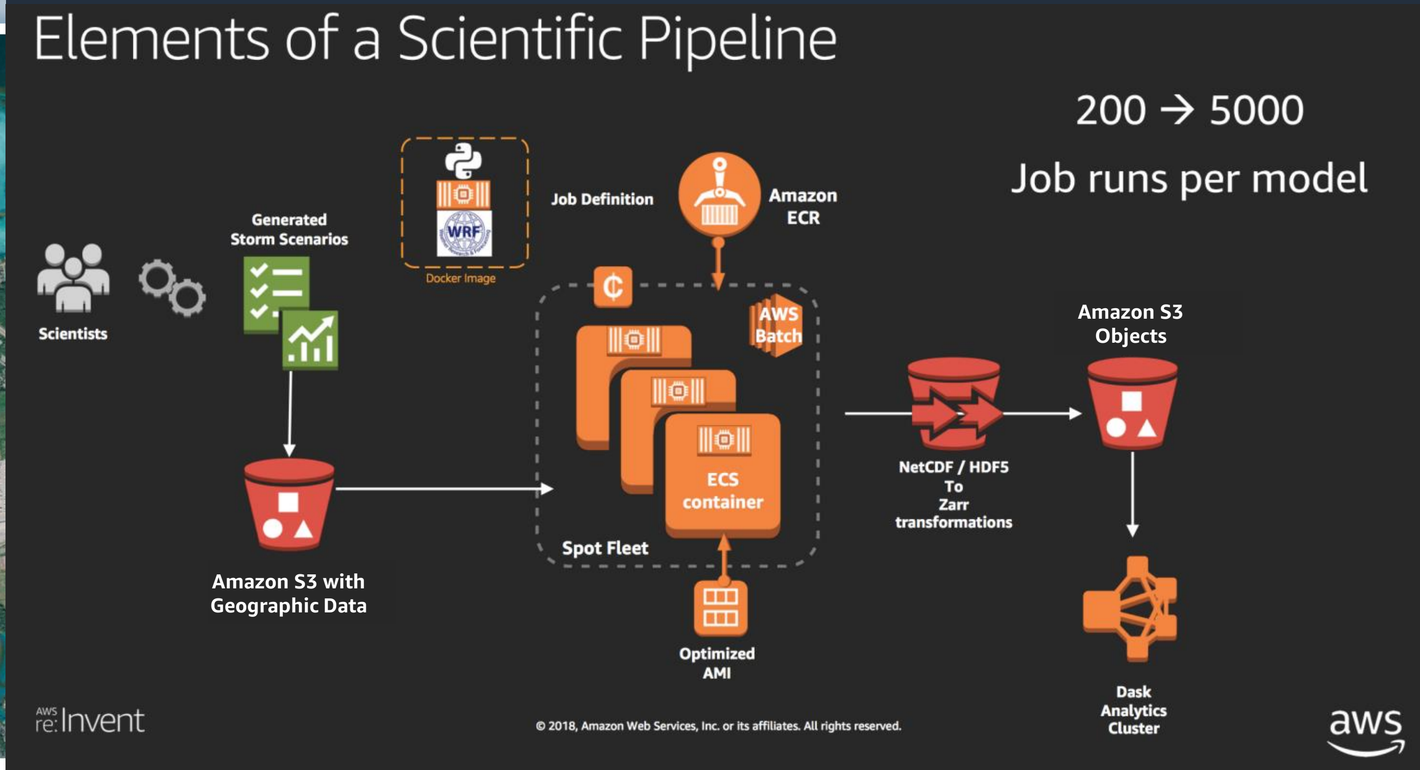
Startups need rapid development and robust, scalable deployment. Managed services—beyond basic compute and storage—are key.



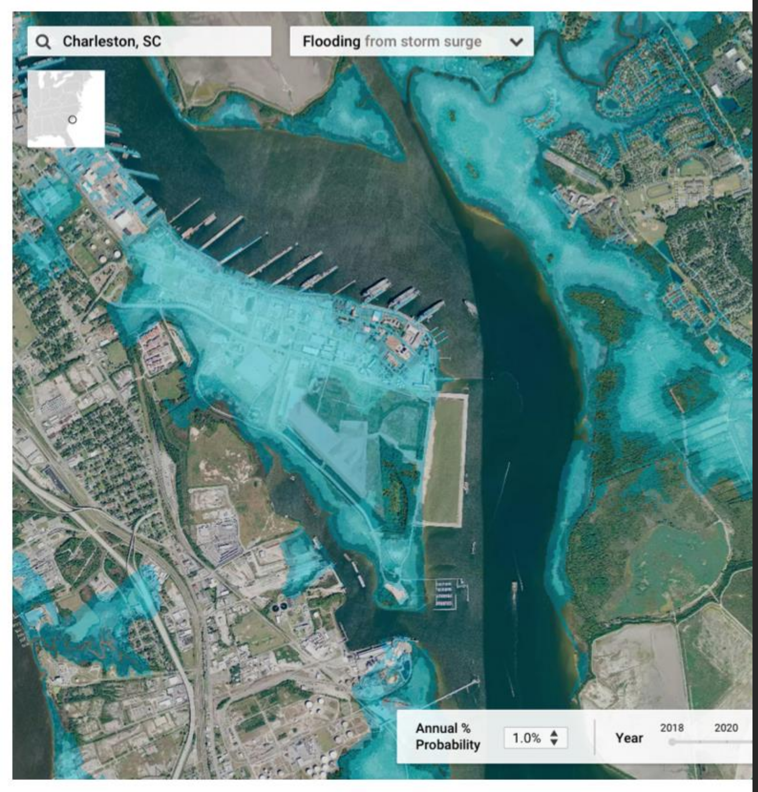
Weather and climate startups



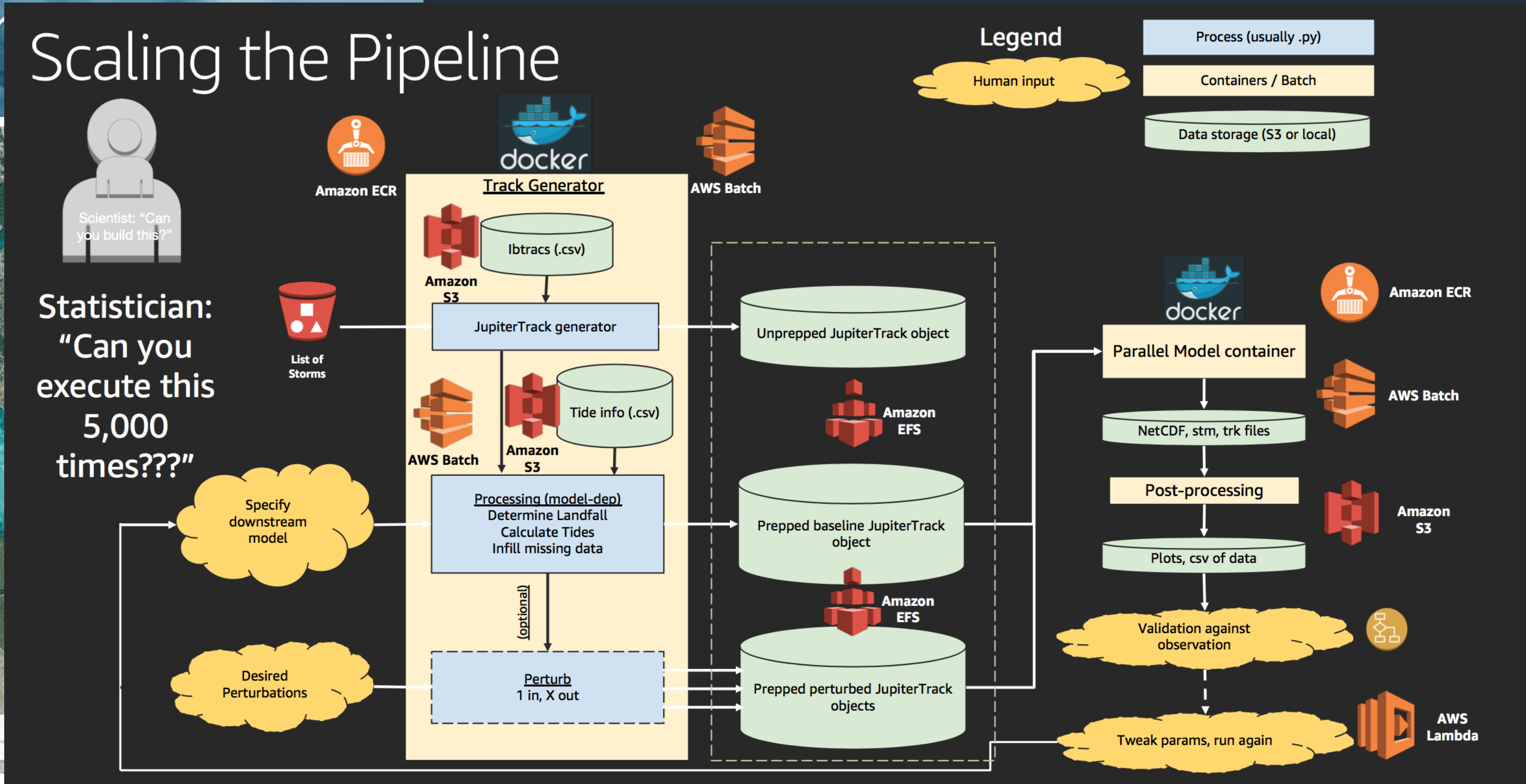




Weather and climate startups

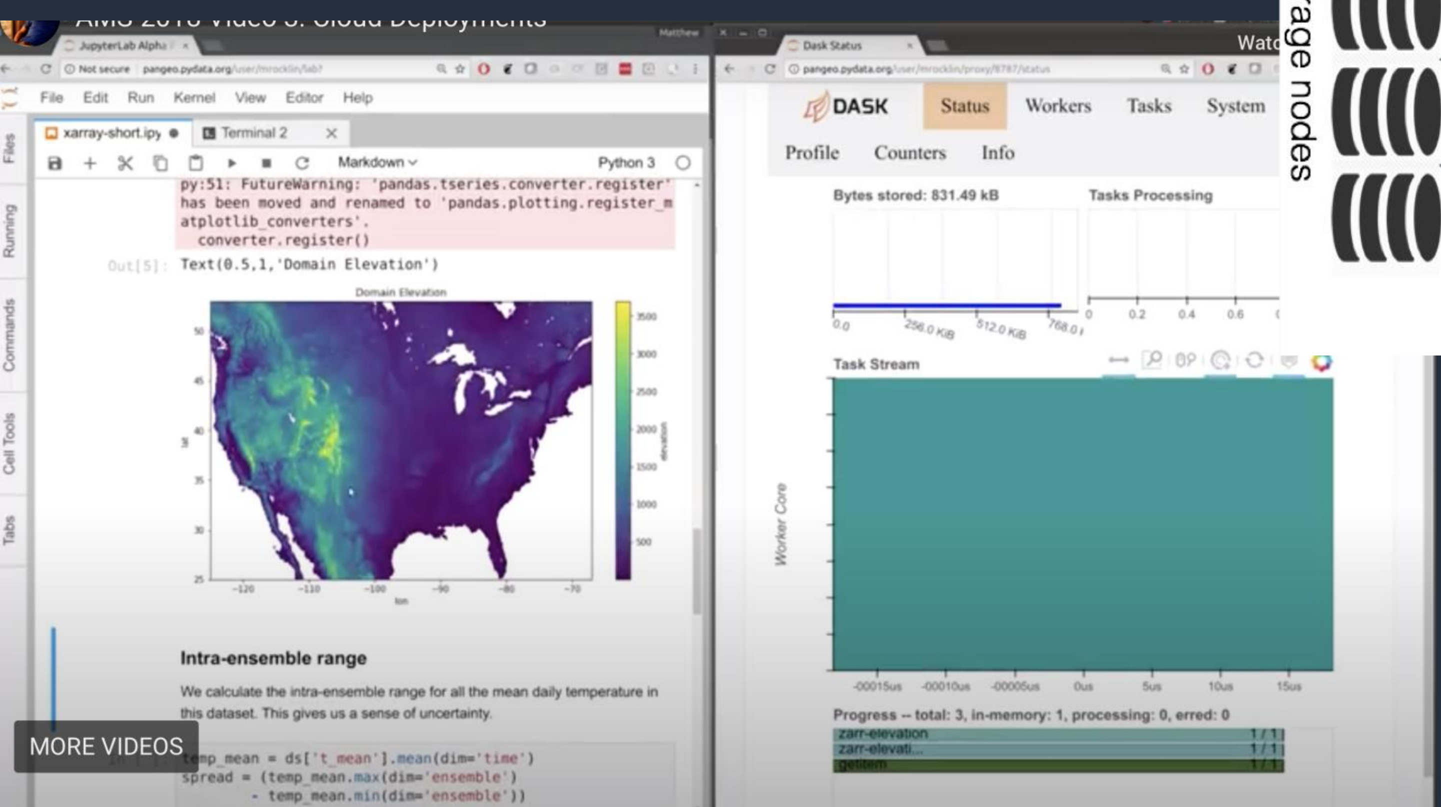


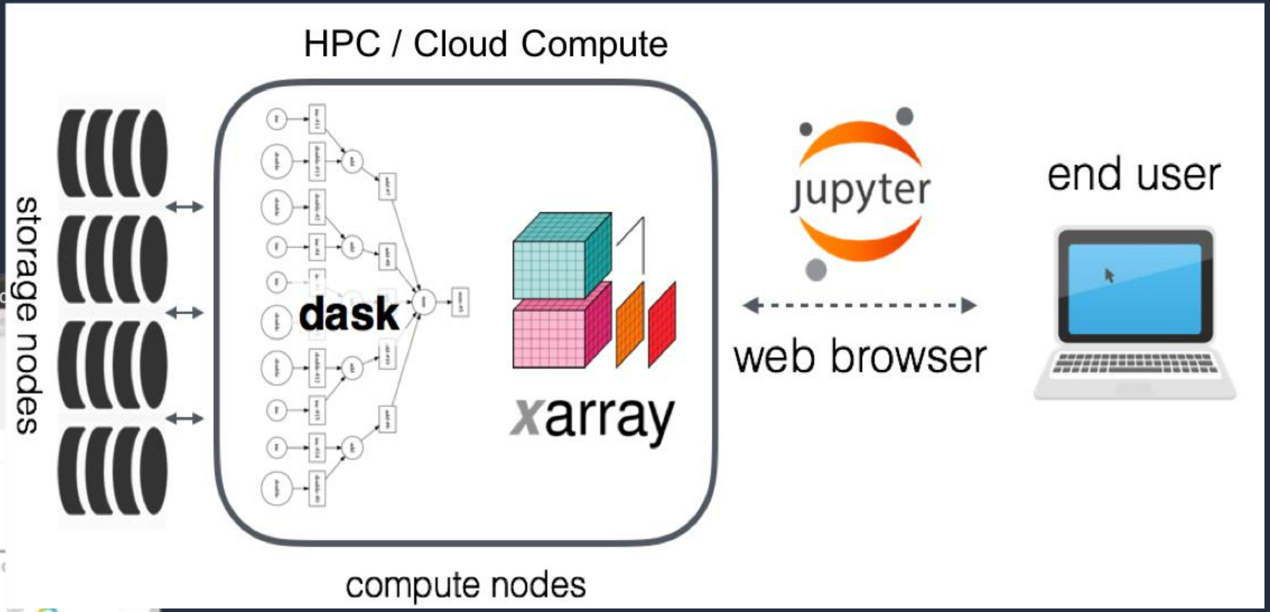
Complex end-to-end pipelines:



Pangeo: Notebook-based access to weather data

http://pangeo.io/architecture.html





aws

Pangeo: Notebook-based access to weather data

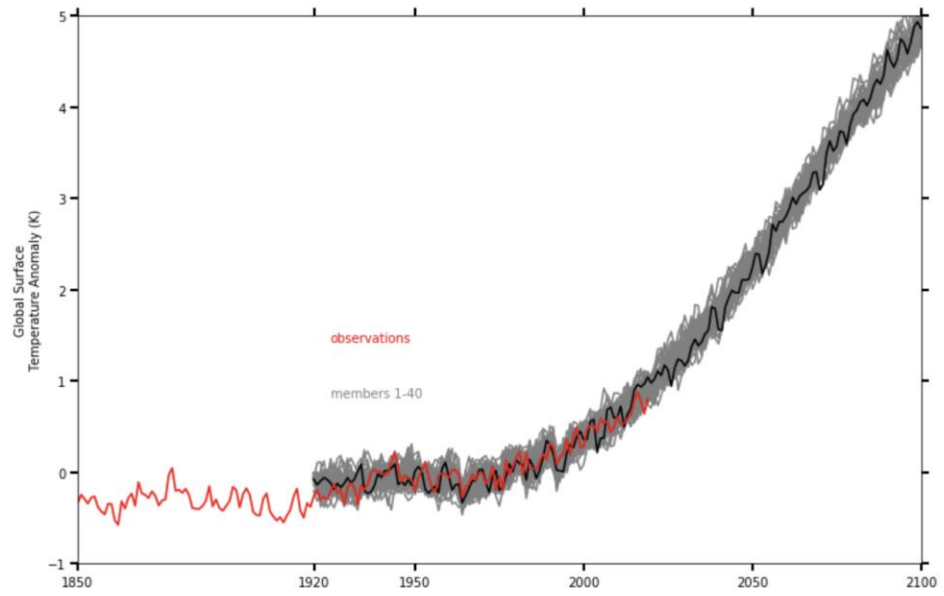
http://pangeo.io/architecture.html

Analyzing large climate model ensembles in the cloud

Joe Hamman Follow
Oct 9, 2019 · 4 min read

Written by Joe Hamman, posted on behalf of the NCAR Science at Scale Team.

The Science at Scale Team at the National Center for Atmospheric Research (NCAR) is excited to announce the release of the Community Earth System Model (CESM) Large Ensemble Numerical Simulation (LENS) dataset published in the Amazon Public Dataset Program (link to dataset). In this blog post, we give a brief overview of 1) the LENS dataset, 2) how you can access the data, and 3) a Binder-ready Jupyter Notebook that reproduces a few key analyses of the LENS dataset — originally presented in the Kay et al. 2015 paper.



Reproduced figure 2 from Kay et al. 2015. Original Caption: "Global surface temperature anomaly (1961–90 base period) for the 1850 control, individual ensemble members, and observations (HadCRUT4; Morice et al. 2012)." The 1850 control run is not shown in the reproduction.

https://medium.com/pangeo/cesm-lens-on-aws-4e2a996397a1

aws

MetPy: Python library access to weather data

https://unidata.github.io/python-gallery/examples/Nexrad_S3_Demo.html#sphx-glr-examples-nexrad-s3-demo-py

Plotting AWS-hosted NEXRAD Level 2 Data

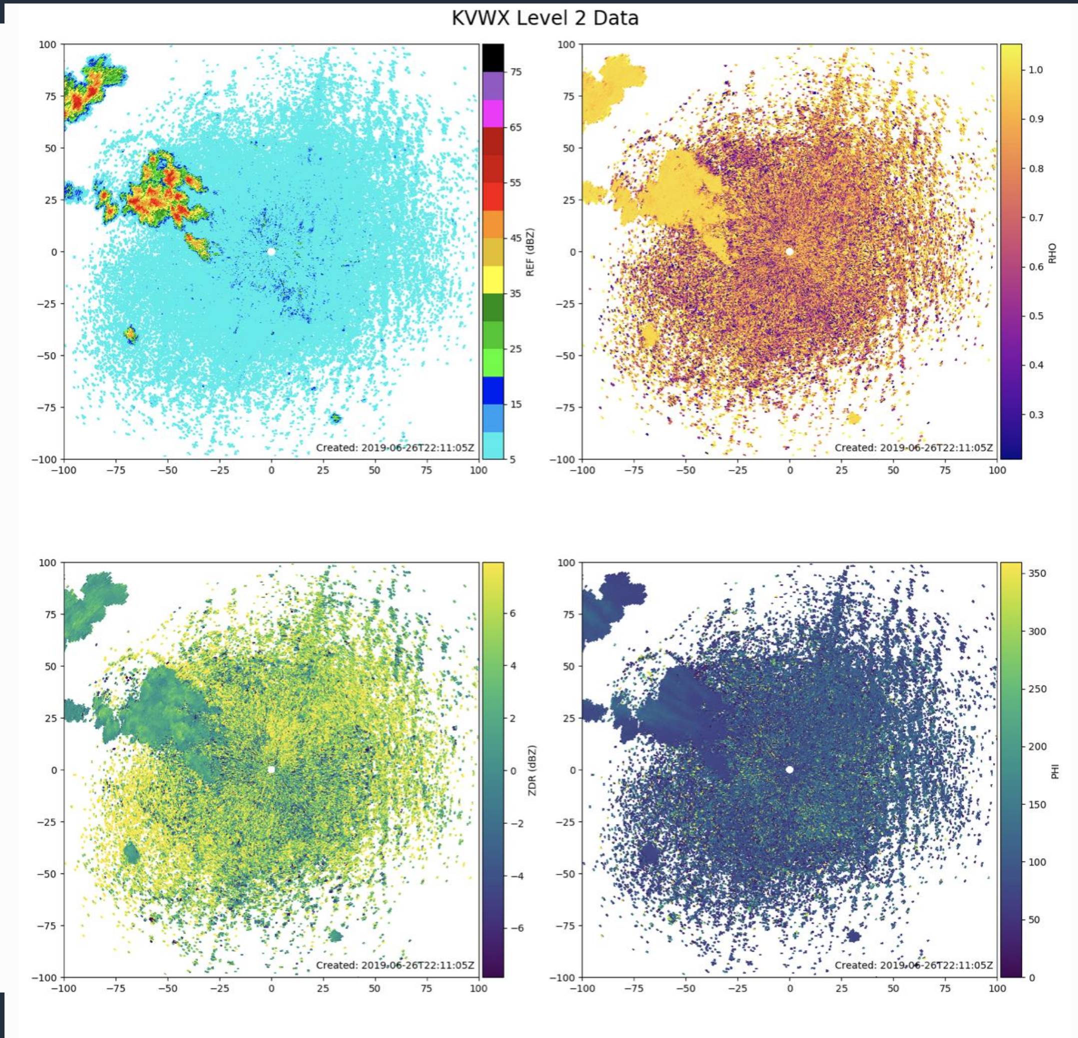
Access NEXRAD radar data via Amazon Web Services and plot with MetPy

Accessing data remotely is a powerful tool for big data, such as NEXRAD radar data. By accessing it in the cloud, you can save time and space from downloading the data locally.

```
import boto3
import botocore
from botocore.client import Config
import matplotlib.pyplot as plt
from metpy.io import Level2File
from metpy.plots import add_timestamp, ctables
from mpl_toolkits.axes_grid1 import make_axes_locatable
import numpy as np
```

Access the data in the AWS cloud. In this example, we're plotting data from the Evansville, IN radar, which had convection within its domain on 06/26/2019.

```
s3 = boto3.resource('s3', config=Config(signature_version=botocore.UNSIGNED,
                                        user_agent_extra='Resource'))
```



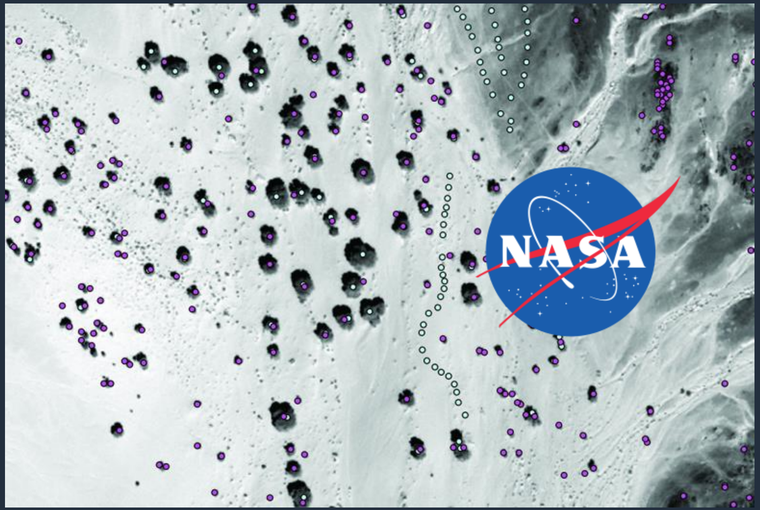
Satellite imagery: Research, agriculture, finance…

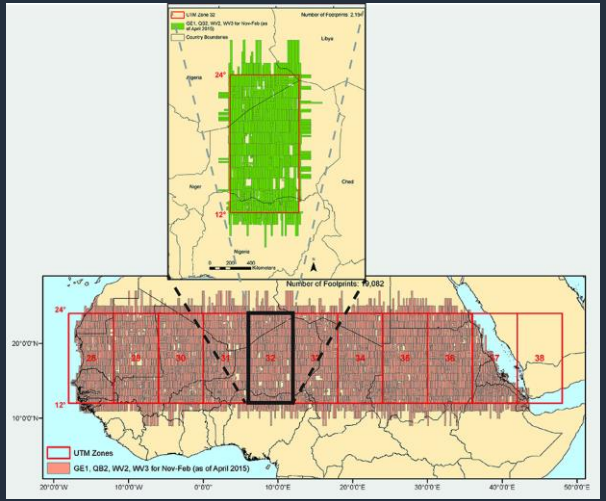
NASA – climate research

- Mosaicking 2,500+ QuickBird satellite images into 100-km x 100-km tiles, which are then broken into 25-km x 25-km sub-tiles for processing

- Orthorectifying and mosaicking all satellite data in ADAPT

- Identifying trees and shrubs using adaptive vegetation classifier algorithms; estimating biomass; incorporating algorithms to calculate tree and shrub height for biomass estimates





The combined resources of ADAPT and AWS reduce total processing time from 10 months to less than 1 month

https://www.nas.nasa.gov/SC15/demos/demo31.html

Teaching

User workshops for WRF, CESM, GEOSChem, and FV3 have been run in AWS

Repeatable environment for instructors

Students report faster learning outcomes

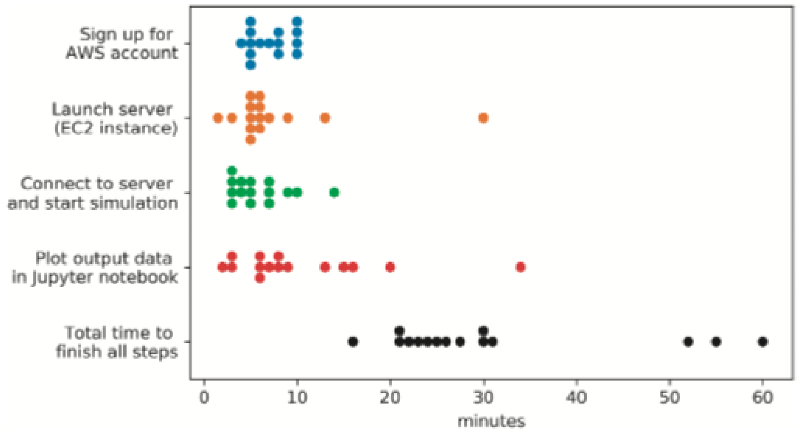
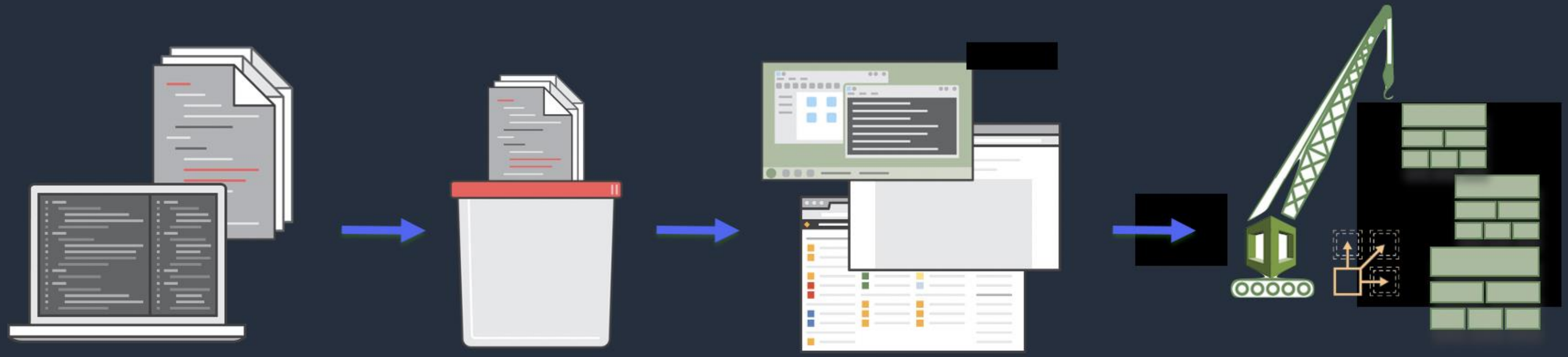


FIG. 3. Wall-clock time required for 15 GEOS-Chem users with no prior cloud experience to complete a demo GEOS-Chem simulation on the AWS cloud and visualize the output. The actual simulation runtime is not included and is a few minutes for this demo case.

8 | BAMS OCTOBER 2019

How AWS facilitates reproducibility

- Shared data
- Containers
- Infrastructure as code with AWS CloudFormation



In conclusion (Part 1)

HPC results and beyond

In summary…

- AWS is a **viable HPC platform for operational-scale Numerical Weather Prediction**, as shown by running Unified Model weather forecasts at ~2,500 cores within comparable time as on premises; and capable of speeding up 2x beyond that by increasing cluster size.

- These results are not a fluke: AWS posts strong performance results for **many weather and climate models, and for many other types of HPC codes (CFD, MD…)**

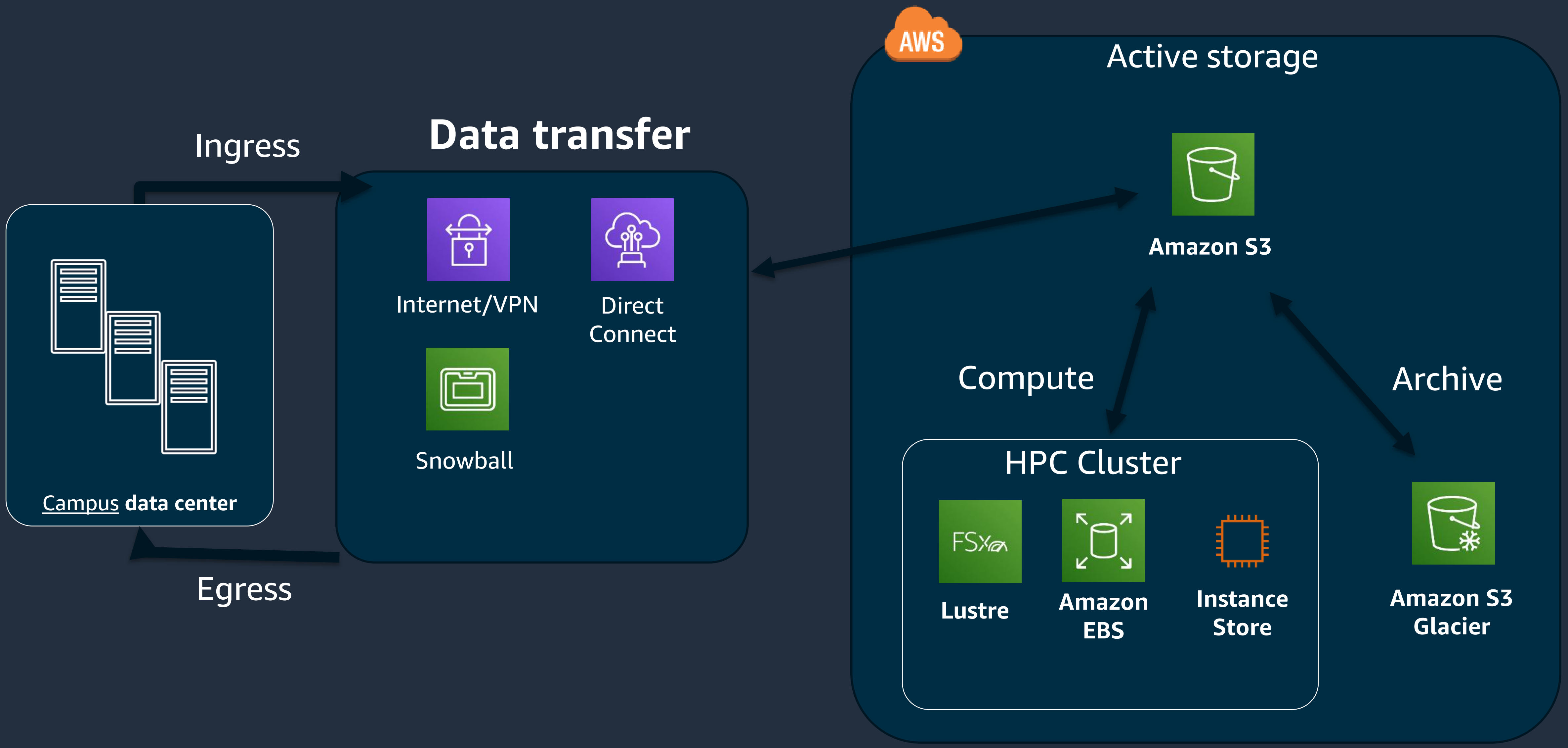
- AWS is also a strong platform for hosting and distributing massive **datasets**, building complex operational **pipelines**, doing **research**, and for **ML/AI** or analytics – e.g. NOAA and NASA datasets in the AWS Open Data platform, Pangeo, etc.

PART 2:

How to do HPC in the cloud

And how it is different from on-premise HPC

Research data flow



Cloud HPC: the basics

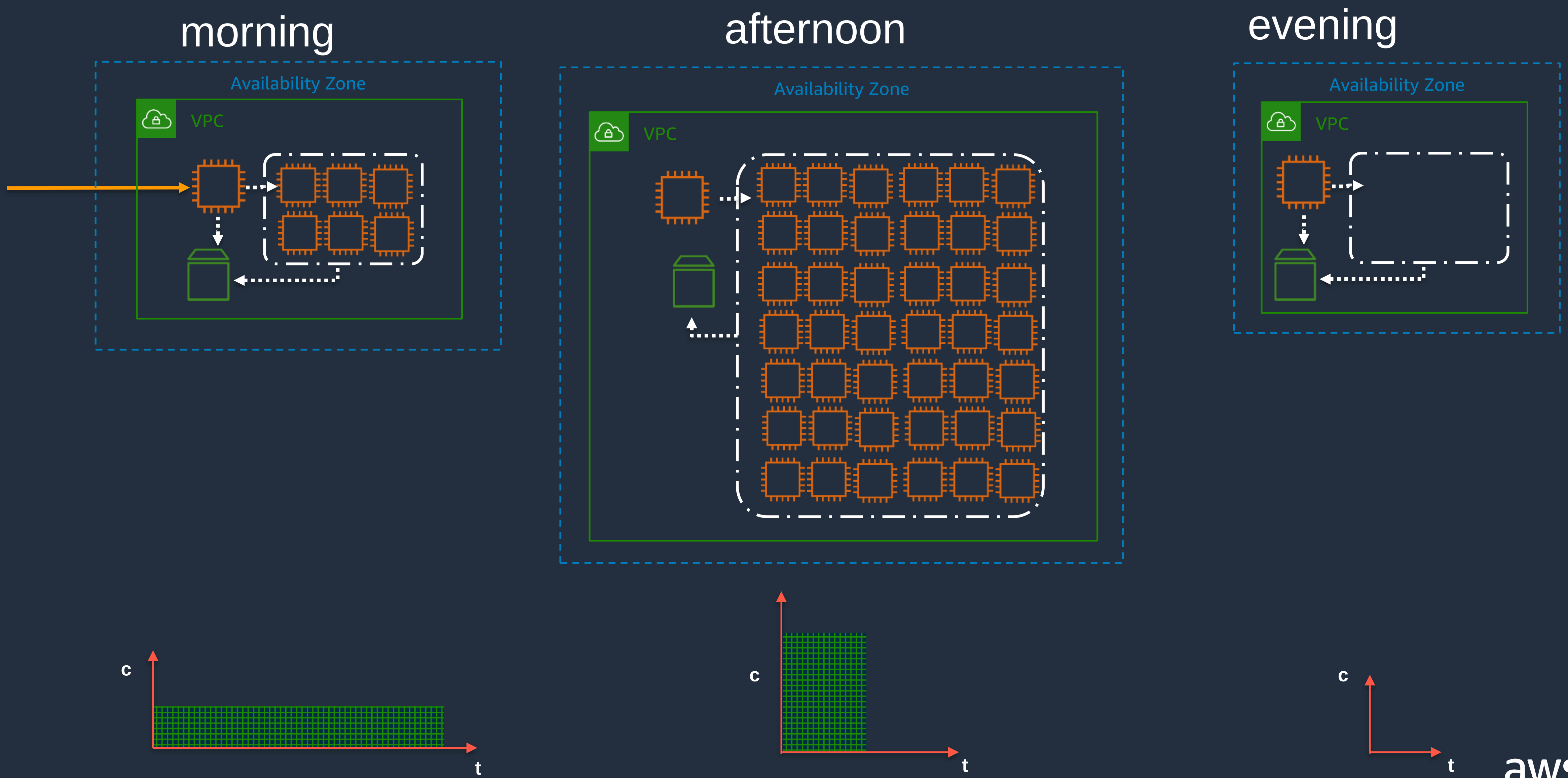
1. Several ways of doing compute/HPC

- **HPC cluster**
  - **Traditional look-and-feel (CLI, scripts, scheduler…)**
  - **Good for large MPI jobs**
- AWS Batch managed computing
  - Based on Docker containers
  - modern, automated
- Serverless (Lambda)
- Notebook based (e.g. Pangeo)
- Amazon EKS
- ML/AI services

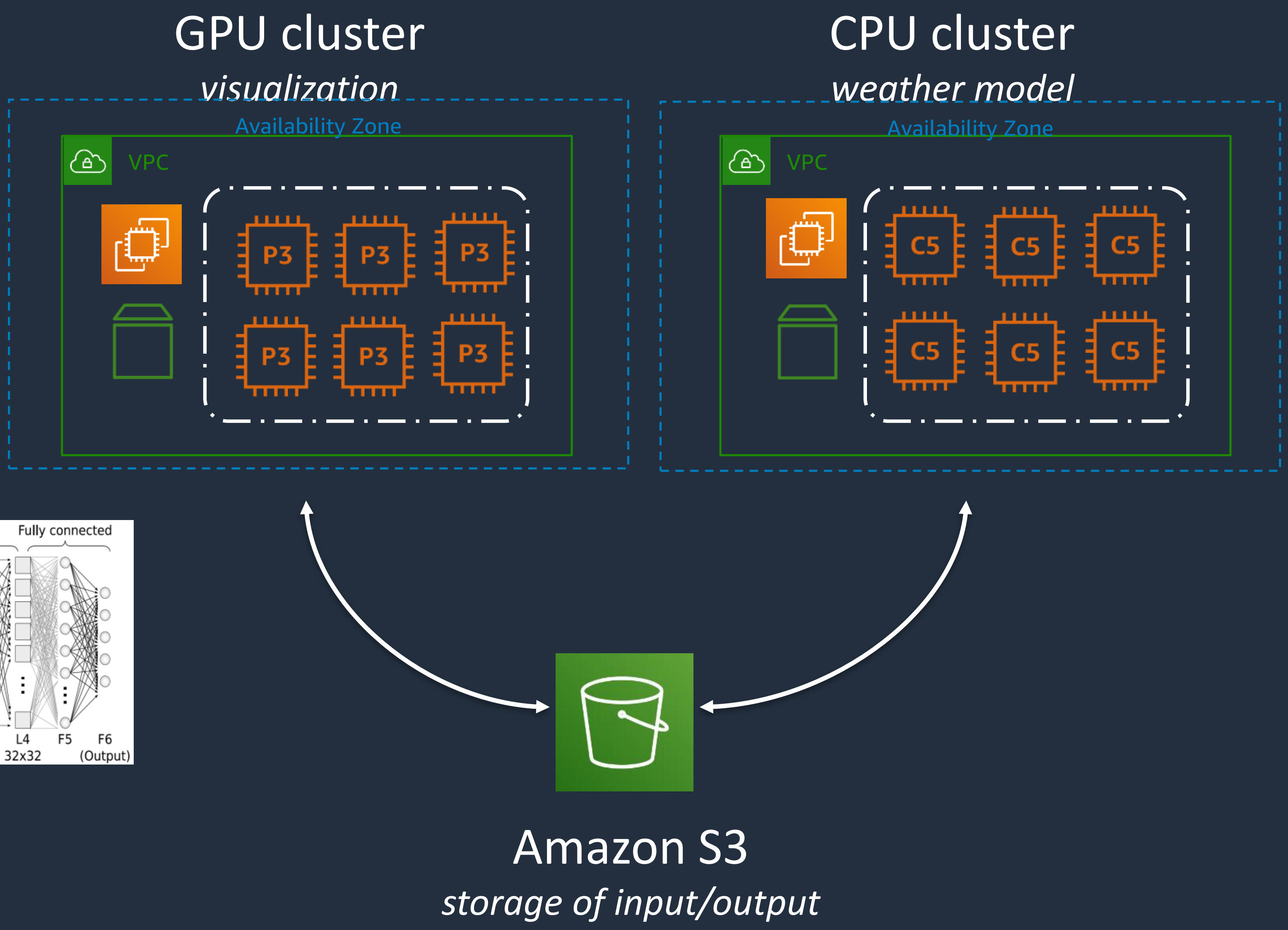
2. Smash the monolith!

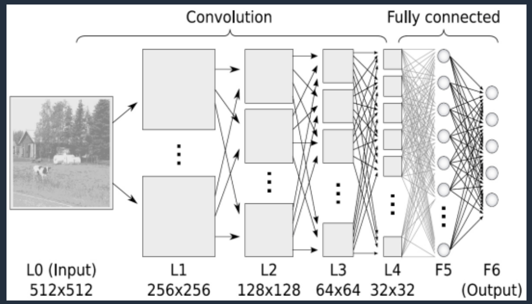
- Cluster is ephemeral (data lives forever)
- Embrace flexibility
- Smaller units tailored to specific jobs
- More frequent upgrades and changes
- No competition for resources

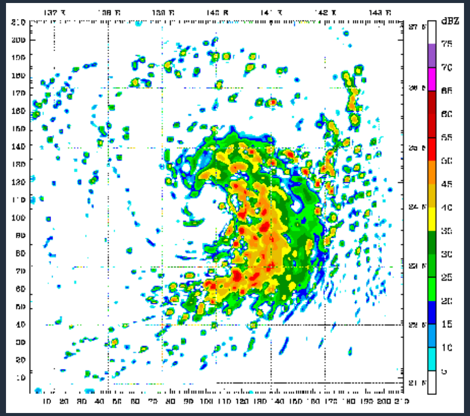
Clusters are flexible



Clusters are fit for purpose

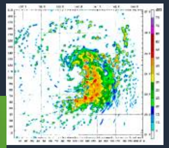






aws

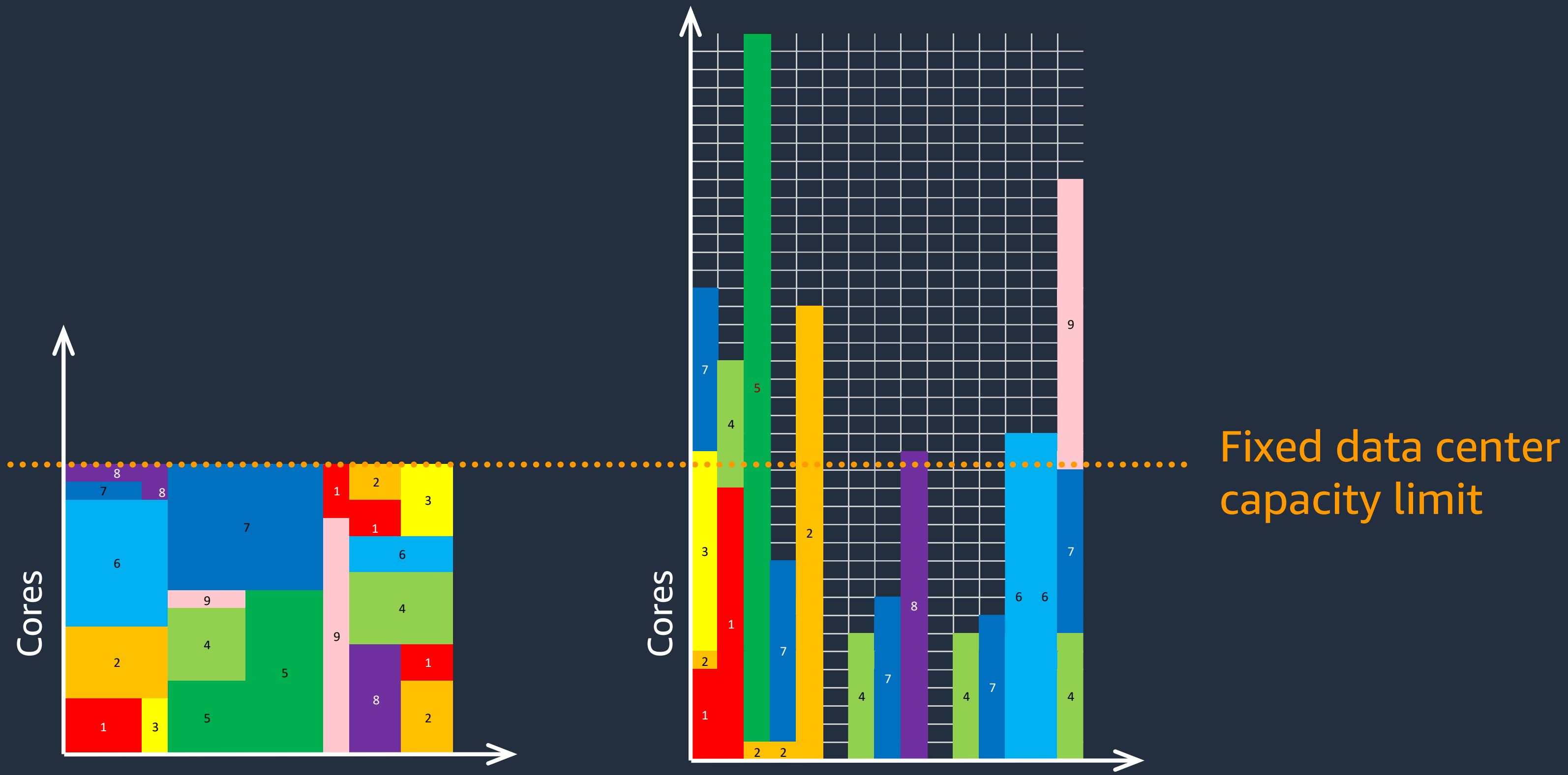
Clusters are ephemeral



Amazon S3

storage of input/output

Scalability improves time-to-results



Finite capacity, usually with long queues to wait in

Massive capacity when needed to speed up time to results, and agile environment when additional hardware and software experimentation is needed

Traditional HPC center **AWS**

Wide choice of server type

Categories

General purpose

Burstable

Compute intensive

Memory intensive

Storage (High I/O)

Dense storage

GPU compute

Graphics intensive

Capabilities

Choice of processor
(AWS, Intel, AMD)

Fast processors
(up to 4.0 GHz)

High memory footprint
(up to 24 TiB)

Instance storage
(HDD and NVMe)

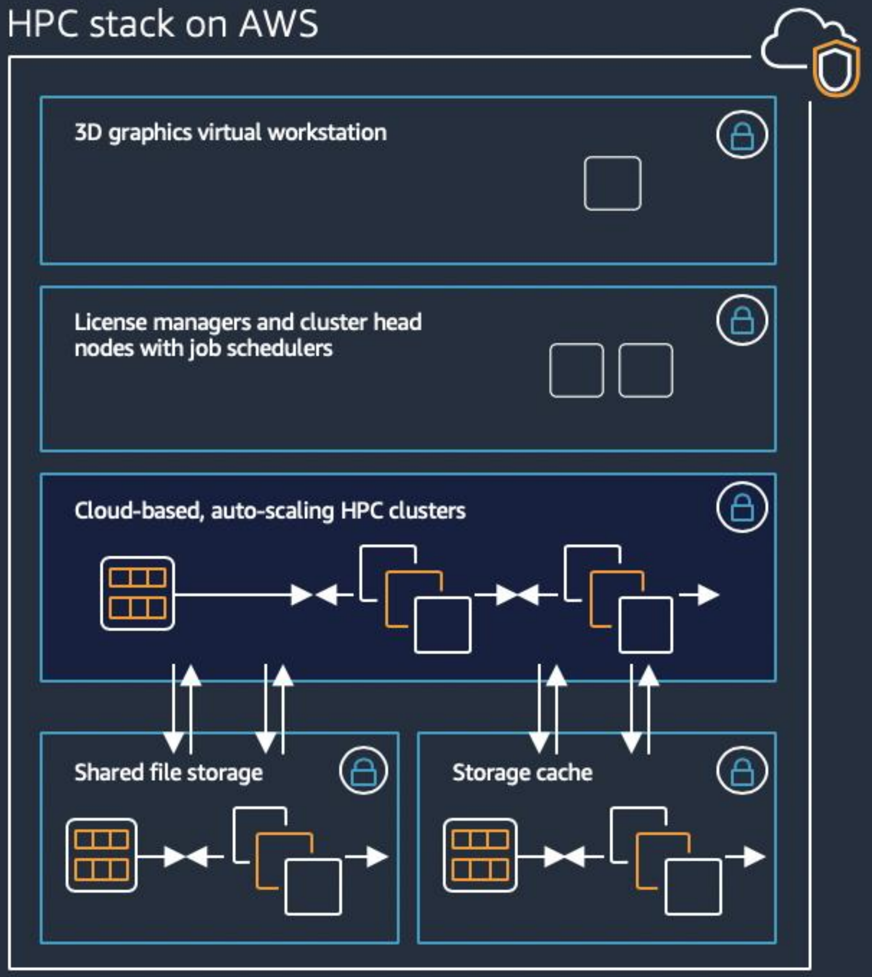
Accelerated computing
(GPUs and FPGA)

Networking
(up to 100 Gbps)

Bare Metal

Size
(Nano to 32xlarge)

HPC stack on AWS



Intel® Xeon Scalable processors

AMD EPYC processors

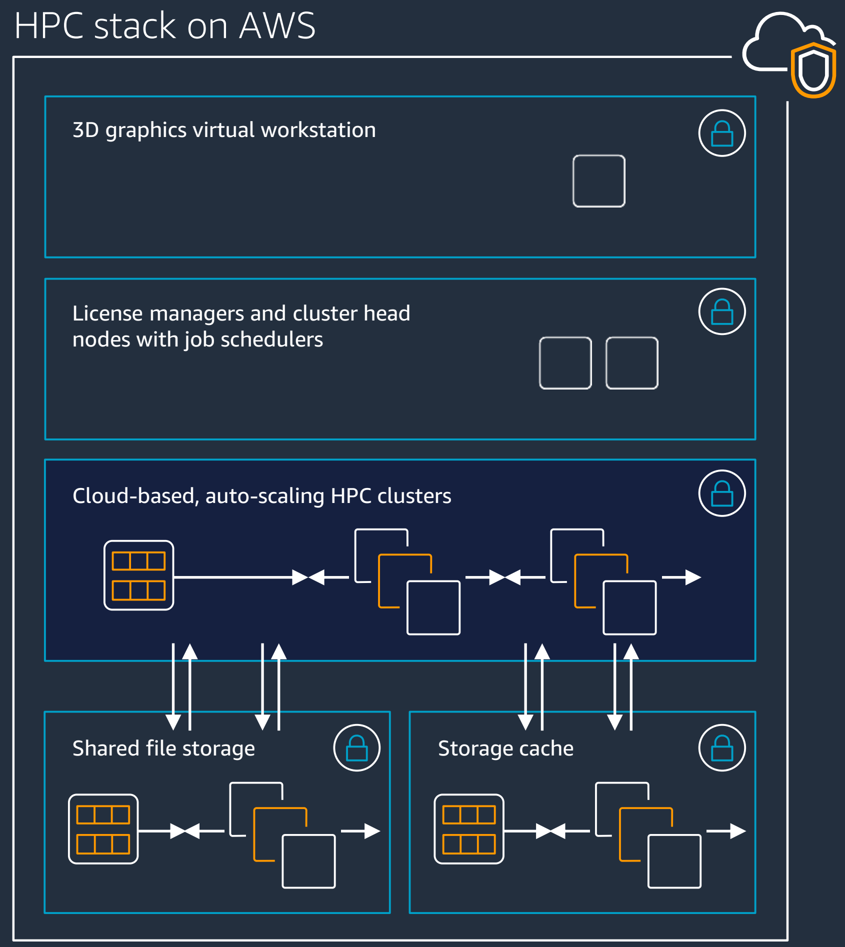
Graviton processors

High bandwidth compute instances: C5n

Massively scalable performance

- C5n Instances will offer up to 100 Gbps of network bandwidth
- Significant improvements in maximum bandwidth, packet per seconds, and packets processing
- Custom designed Nitro network cards
- Purpose-built to run network bound workloads including distributed cluster and database workloads, HPC, real-time communications and video streaming

Featuring Intel Xeon Scalable (Skylake) processor



Benefits of latest generation intel-based instances

	General Purpose Computing	High Performance Computing
M4 → M5	Up to **24%** EC2 Instance Cost Savings	Up to **3.5x** Performance Improvement
C4 → C5	Up to **50%** EC2 Instance Cost Savings	Up to **4.6x** Performance Improvement

[1] Source: https://www.dbbest.com/blog/validating-aws-ec2-sql-server-deployments-using-benchmark-tools/
[2] Source: Intel benchmarks
[3] Source: : TSO Logic Research https://tsologic.com/wp-content/uploads/2018/11/TSOLogic_Intel_Research_Final_2018-1.pdf

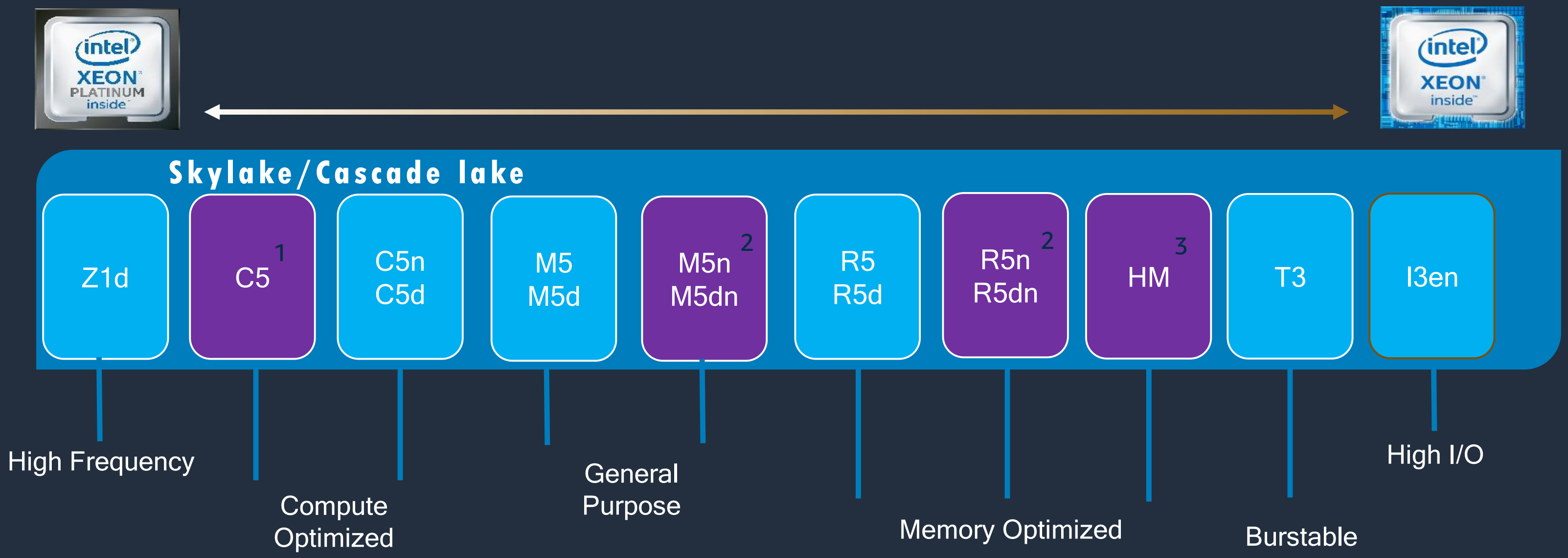
Source: : TSO Logic Research https://tsologic.com/wp-content/uploads/2018/11/TSOLogic_Intel_Research_Final_2018-1.pdf

AWS EC2 Instances Powered By latest gen Intel processors

EC2 Instance Details: https://aws.amazon.com/ec2/

Skylake/Cascade Lake - instances available

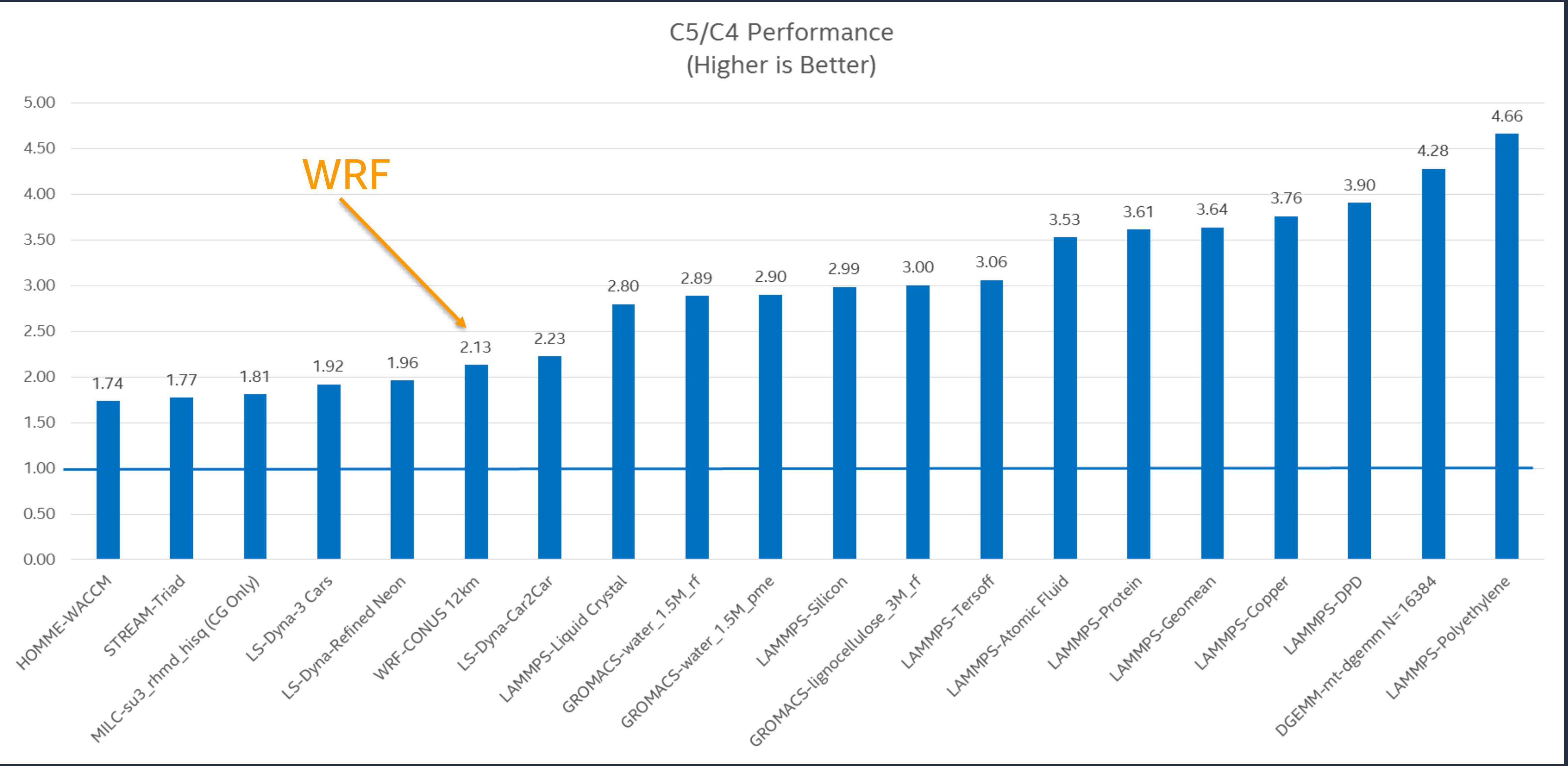
Skylake-based instance



CASCADE LAKE INSTANCES with 2nd Generation Xeon Scalable Processor and Deep Learning Boost (VNNI)

1 - Cascade Lake instances available: C5.12xlarge, C5,24xlarge, C5.metal
2 - Cascade Lake instances available: M5n, M5dn, R5n, R5dn (32 instance sizes)
3 - Cascade Lake instances available: 18 TiB, 24 TiB

Application Workloads Performance

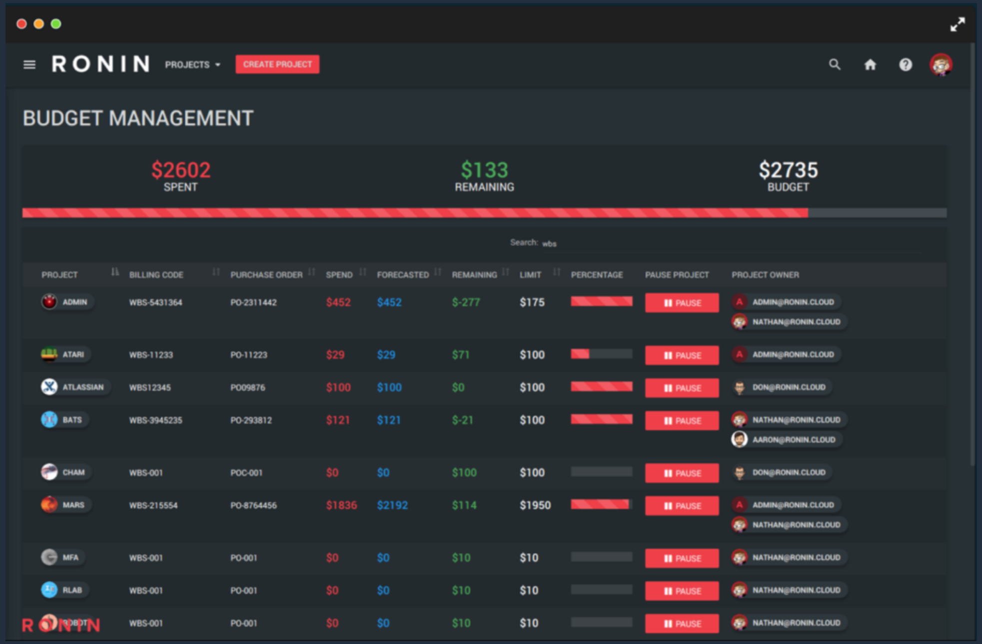


Delivering campus-wide cloud HPC capabilities

RONIN IS A CLOUD ORCHESTRATION PLATFORM, LOWERING THE ENTRY LEVEL TO USING THE CLOUD

ENTERPRISE BUDGET CONTROL.

ONE ACCOUNT, PROJECT CENTRIC BILLING.



https://ronin.cloud

Research institutions and AWS

How to make the cloud part of your research portfolio

What are the trends of today?

- HPC capabilities now capable of tightly coupled jobs on 1000s of cores
  - Realistic scaleout or migration option
  - New server types *(AMD/ARM/GPUs)*
- Research data lakes keep growing
- Increased collaborations with leading agencies and institutions. *(NASA/NOAA/NIH/NSF/DoE labs; Universities…)*
- Toolset maturing
  - AWS services: servers/containers/serverless; storage; AI/ML; …
  - Deployment/management *(Ronin; Emory)*
  - Ecosystem
- Metrics that Matter
  - Time to discovery, learning outcomes, …

Moving beyond HPC scaling curves: Cloud as contributor to institutional goals

Can cloud make university's research data more reusable?

Can cloud make students more employable after graduation?

Can cloud shorten average time-to-discovery and boost impact?

Can cloud raise the university's profile for research (inter)nationally?

Can cloud help make competitive faculty hires? *(Extra resources allow the competitive new hire to stay on top of the field)*

Can cloud help new faculty build impact faster? *(Put cloud $ in every startup package and see citations build up faster.)*

Can cloud democratize compute/analytics/ML/AI across all departments?

Can cloud help grad students finish up faster?

Can cloud boost the approval rate of grant applications?

Thank you!