There's no queue in the cloud

Science, not servers

AWS cloud services for weather data and prediction models

Kevin Jorissen Amazon Web Services jorissen@amazon.com

NSF Workshop: Modeling Research in the Cloud Thursday, June 1 2017

BIO

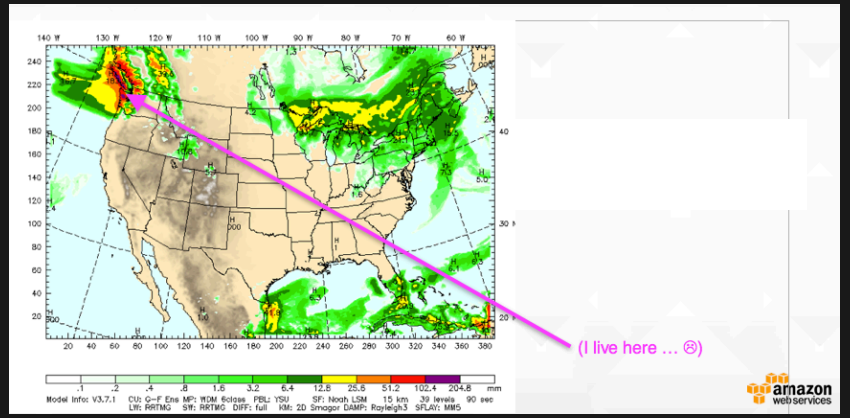
Kevin Jorissen

Seattle

Kevin Jorissen has 10 years of experience in computational science. He developed software solving the quantum physics equations for light absorption by materials, taught workshops to scientists worldwide, and wrote about high performance computing in the cloud before it was fashionable. He worked as a postdoctoral researcher in Antwerp, Lausanne, Seattle, and Zurich.

Kevin joined Amazon in 2015 to help accelerate the adoption of cloud computing in the scientific community globally. He holds a Ph.D. in Physics, but feels equally proud of growing parsnips, surviving a night bus from Darjeeling to Kolkata, and adapting to medium-spicy Sichuan food. He thinks a five-hour run in the Cascade mountains is a fine way to spend a free Saturday.

AWS and the world



amazon web services

“Public Cloud”, or, as we call it, “Cloud”

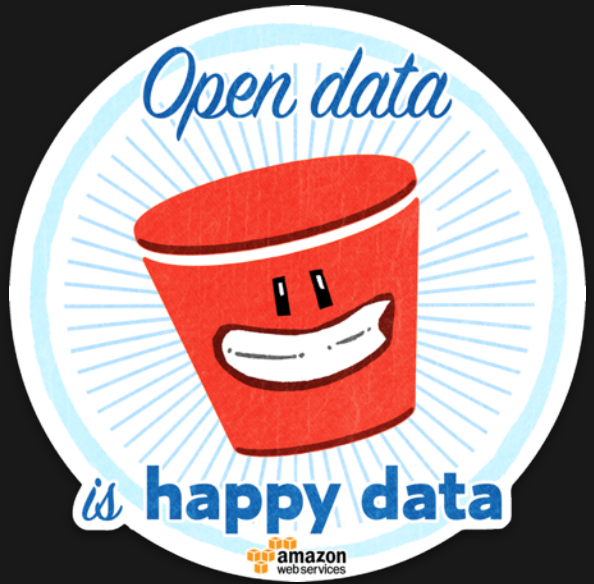
Only the public cloud offers:

- The benefits of hyperscale (does Intel make custom processors for you?)
- Real elasticity/scalability
- Global footprint that supports collaboration
- Real-world security
- 90+ services for data analytics/streaming/machine learning/…
- 1000+ new features for those services each year
- 20,000+ AWS Partners with 3rd party offerings
- Many customers state large cost savings on moving to AWS
- Agility promotes reduced time-to-science

The AWS flavor

- Long experience in building secure, reliable infrastructure at scale
- Wide array of services
- Large footprint
- Invested in research, education, and HPC

AWS and data



amazon
web services

The Big Data Challenge

It's typically consuming and expensive to acquire, store, and analyze large data sets. Accessing data at scale is often a prohibitive challenge.

Our Solution – Shared Open Data on AWS

AWS global footprint makes it a powerful platform for scientific collaboration.
Users and compute can be brought to the data.
AWS offers many advanced big data related services.

Sharing data on AWS makes it accessible to a large and growing community of researchers who use the AWS cloud.

AWS "Open Data" program showcases potential and best practices by hosting several key datasets.

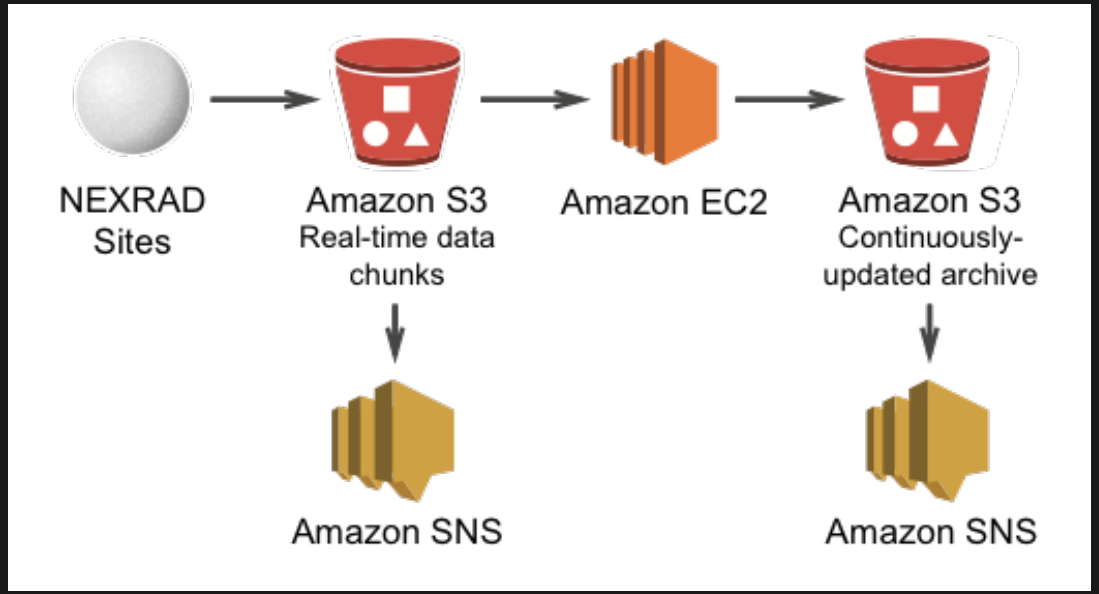
Opening data is the beginning, not the end. Users need to be **educated** and have access to **tools** to analyze and process the data.

When data is shared in the cloud, **anyone** can analyze **any volume of data** without needing to download or store it themselves.

BDP in Action: NEXRAD on AWS

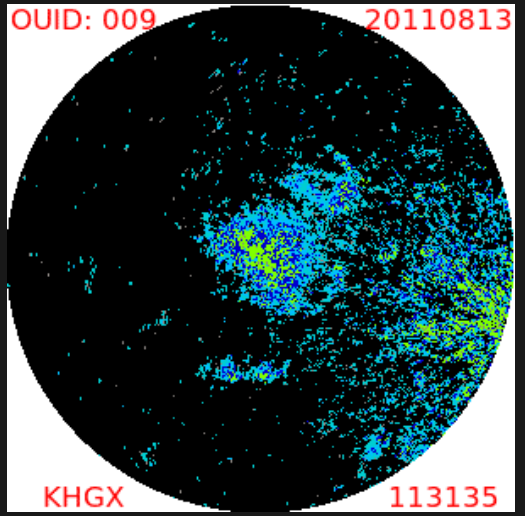
NEXRAD on AWS makes 270TB of individual volume scan files and real-time chunks as objects on Amazon S3.

- Data can be accessed programmatically via a RESTful interface and quickly deployed to any of our products for analysis and processing.
- Amazon Simple Notification Service (SNS) allows subscription to notifications of new data.



NEXRAD on AWS

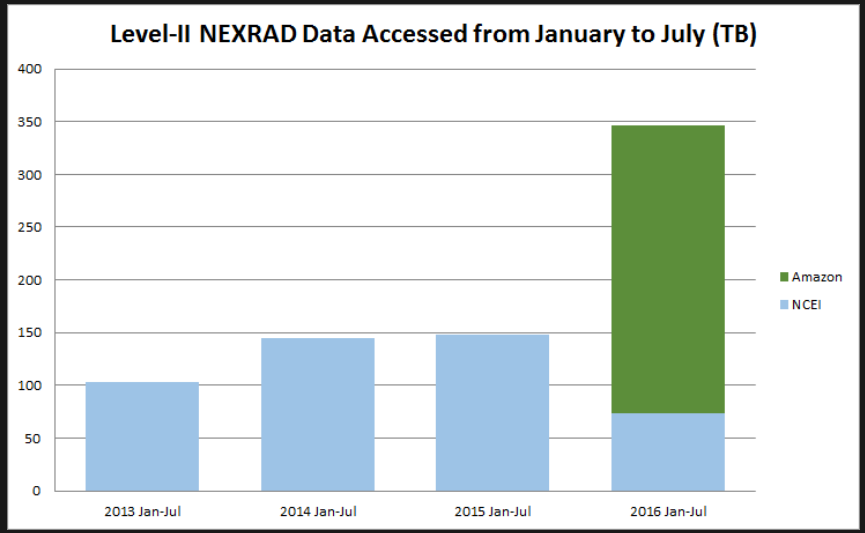
Dr. Eli Bridge uses NEXRAD data to study birds and other flying animals as part of the Oklahoma Biological Survey at the University of Oklahoma. “One of the biggest challenges in our work was simply obtaining large chunks of radar data to work with. Having NEXRAD on AWS is a major help to us. I can download a radar scan in about 2-4 seconds. So there’s really no need for us to store raw data anymore.”



NEXRAD on AWS

Immediate usage:

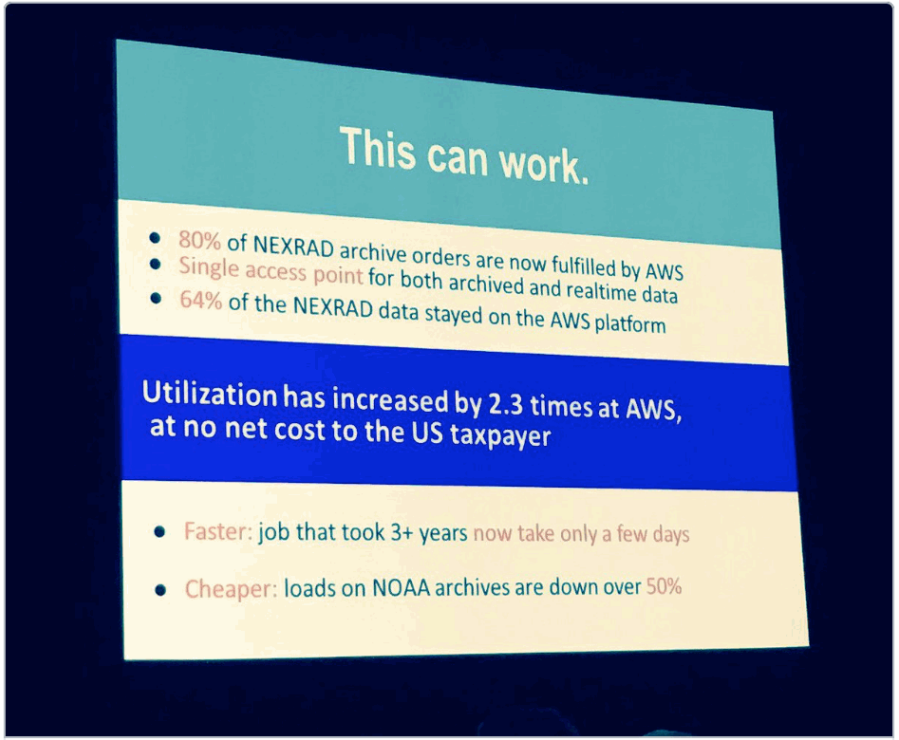
- Climate Corporation **cut two weeks** out of an analysis pipeline
- Increased NEXRAD usage **2.3✕**
- A weather data company stopped storing their own NEXRAD archive, **freeing up revenue** to build new products.



NEXRAD on AWS

Immediate usage:

- Climate Corporation **cut two weeks** out of an analysis pipeline
- Increased NEXRAD usage **2.3✕**
- A weather data company stopped storing their own NEXRAD archive, **freeing up revenue** to build new products.



DJ Patil
@DJ44

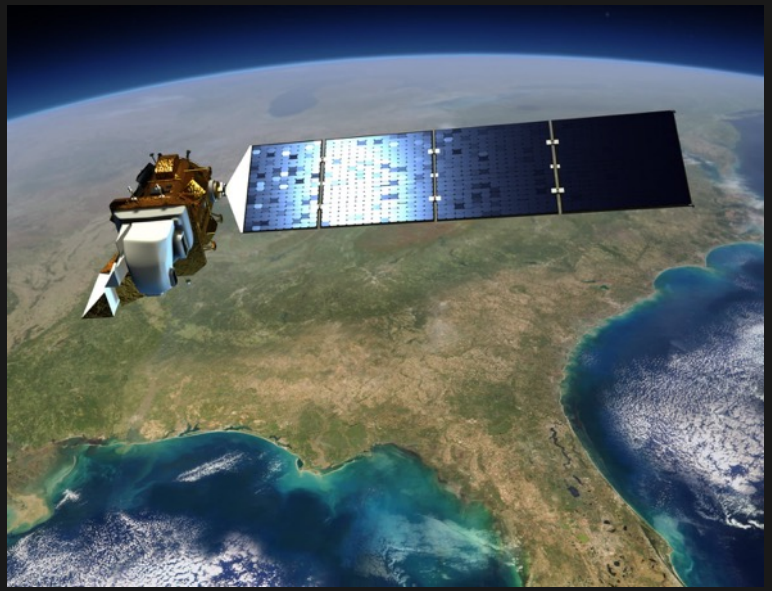
Follow

Wow. What happens when @NOAA puts their data on the cloud. #WHOpenData 👊👍🙌🇺🇸

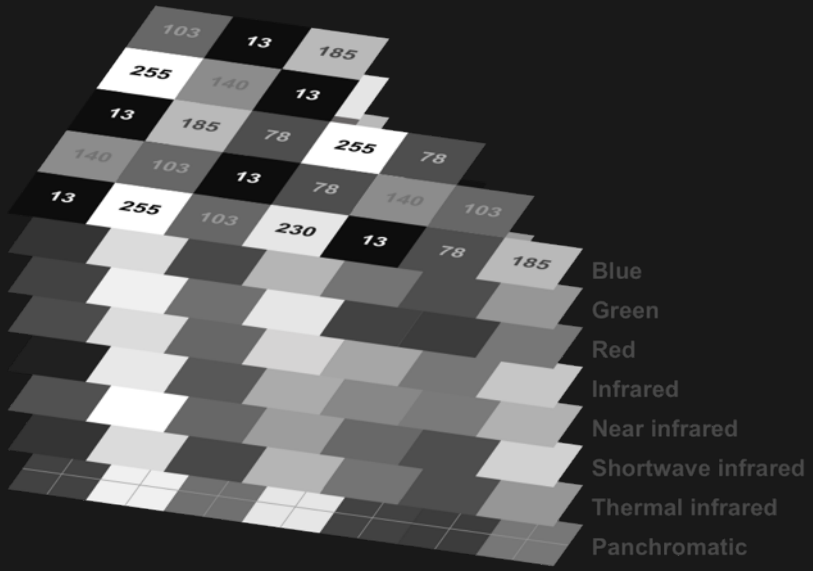
4:04 PM - 28 Sep 2016

92 124

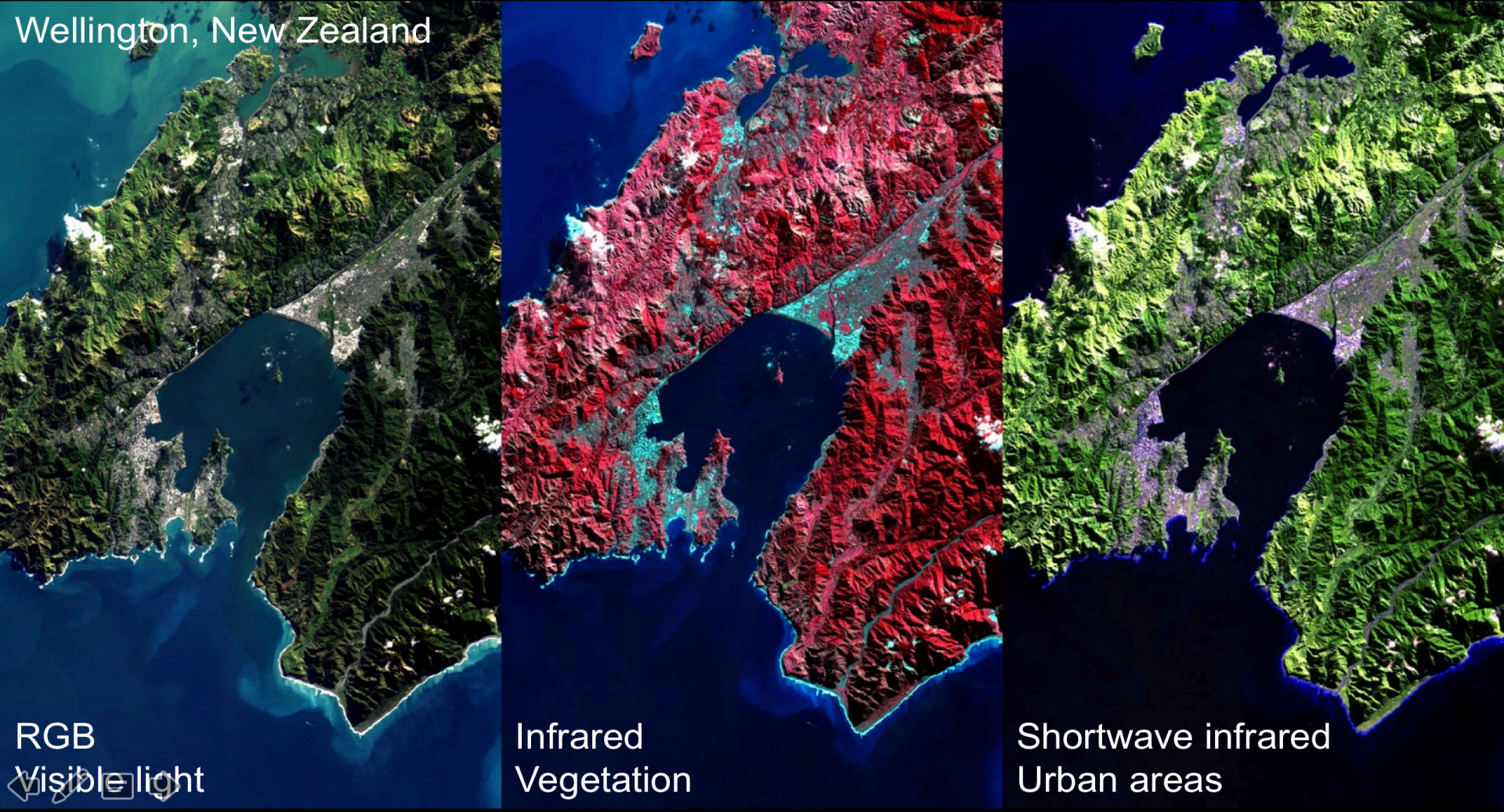
Landsat on AWS



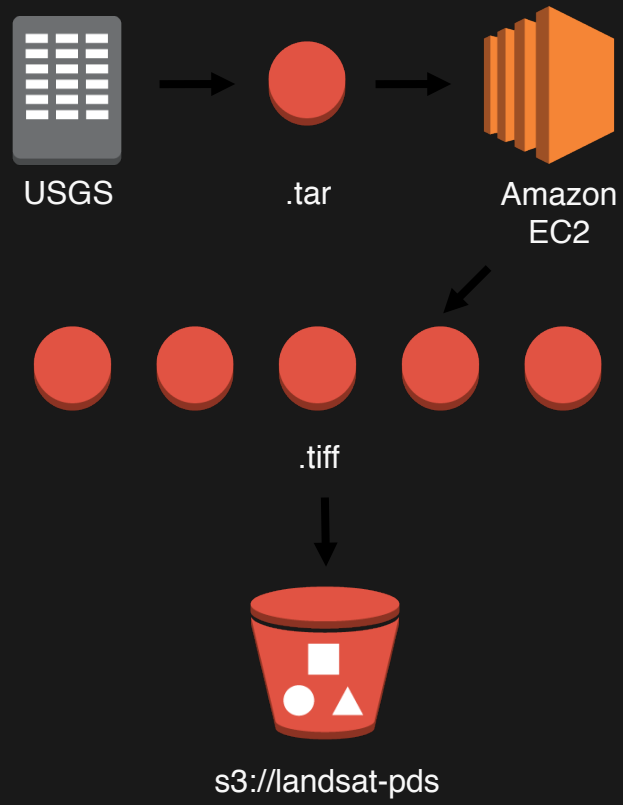
Landsat 8 satellite



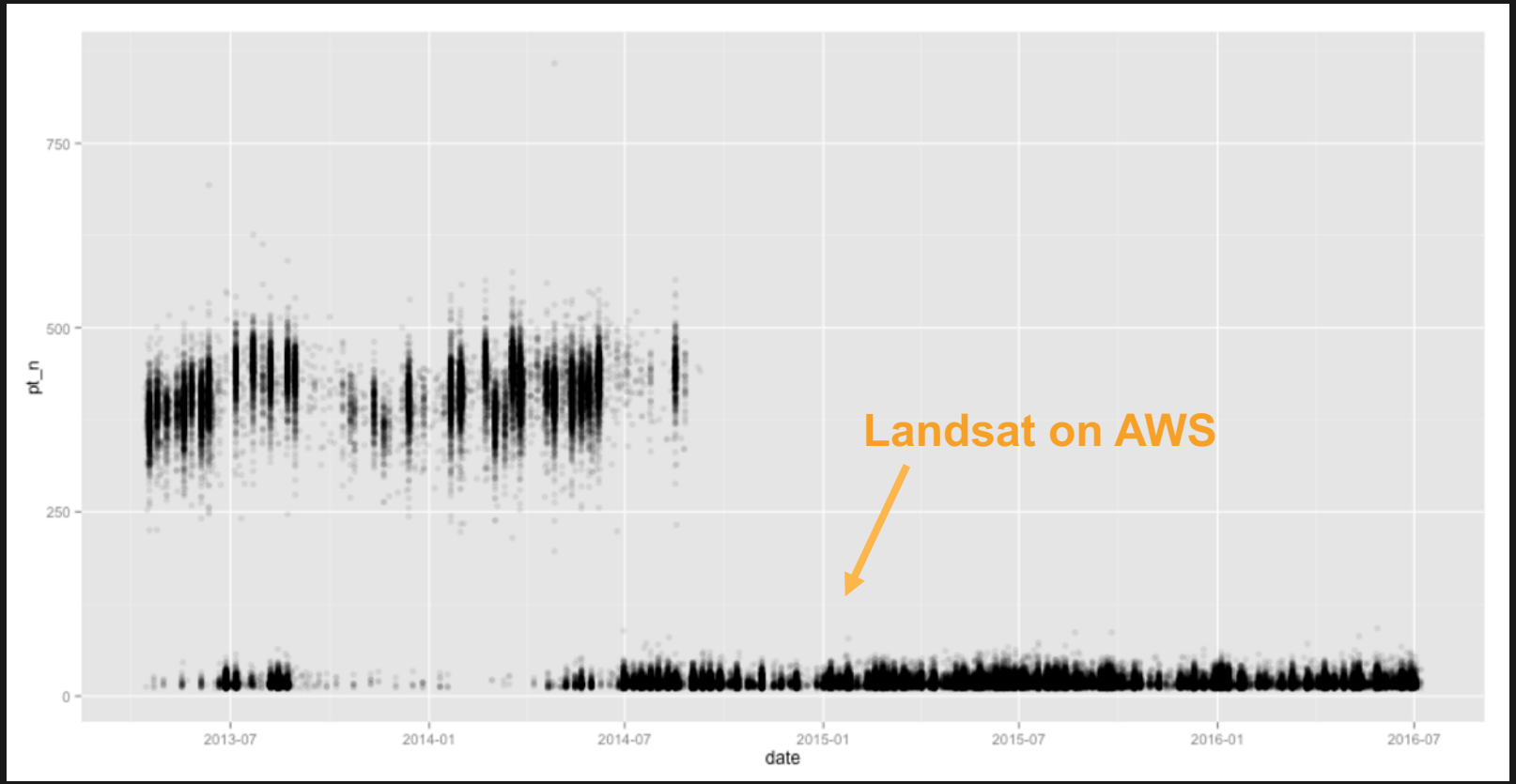
Raster data


Wellington, New Zealand
RGB
Visible light
Infrared
Vegetation
Shortwave infrared
Urban areas

Landsat on AWS



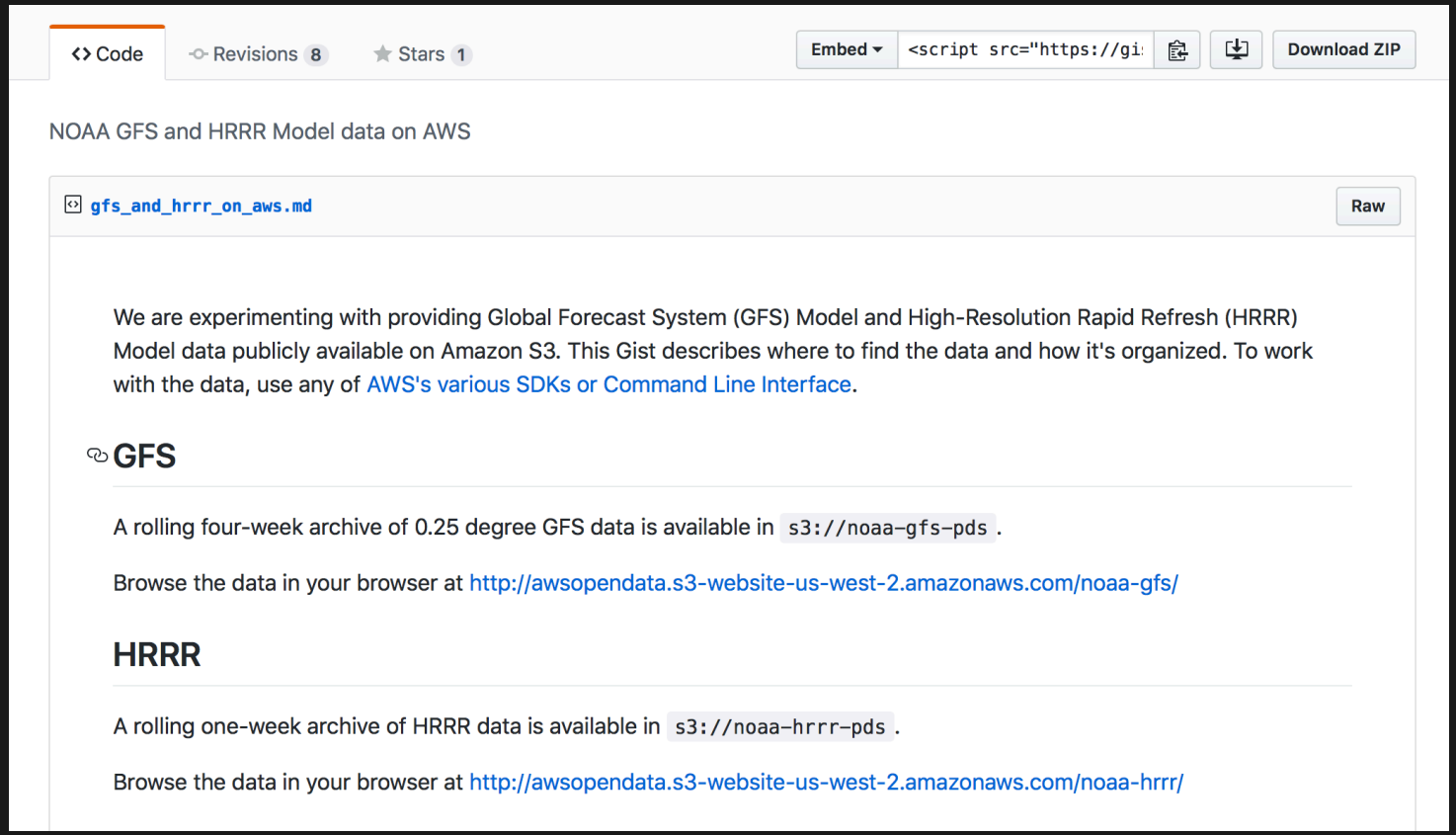
Within the first year of Landsat on AWS, the data has been requested over 1 billion times, globally. Over 400,000 scenes are now available



Graph by Drew Bollinger (@drewbo19) at Development Seed

amazon web services

GFS, HRRR on AWS



Code | Revisions 8 | Stars 1 | Embed <script src="https://gi | Download ZIP

NOAA GFS and HRRR Model data on AWS

gfs_and_hrrr_on_aws.md | Raw

We are experimenting with providing Global Forecast System (GFS) Model and High-Resolution Rapid Refresh (HRRR) Model data publicly available on Amazon S3. This Gist describes where to find the data and how it's organized. To work with the data, use any of AWS's various SDKs or Command Line Interface.

GFS

A rolling four-week archive of 0.25 degree GFS data is available in `s3://noaa-gfs-pds`.

Browse the data in your browser at http://awsopendata.s3-website-us-west-2.amazonaws.com/noaa-gfs/

HRRR

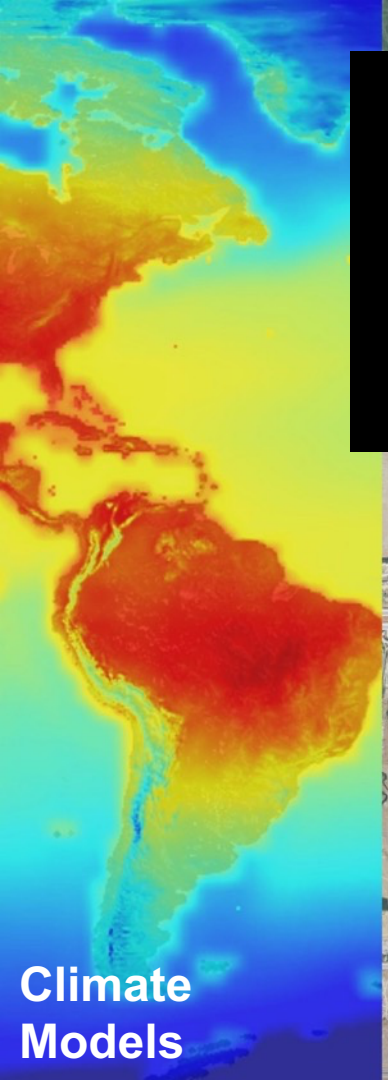
A rolling one-week archive of HRRR data is available in `s3://noaa-hrrr-pds`.

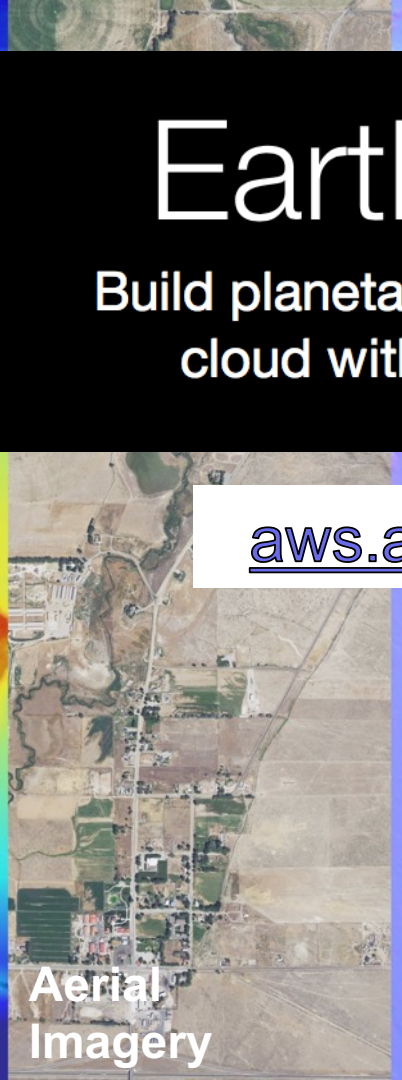
Browse the data in your browser at http://awsopendata.s3-website-us-west-2.amazonaws.com/noaa-hrrr/

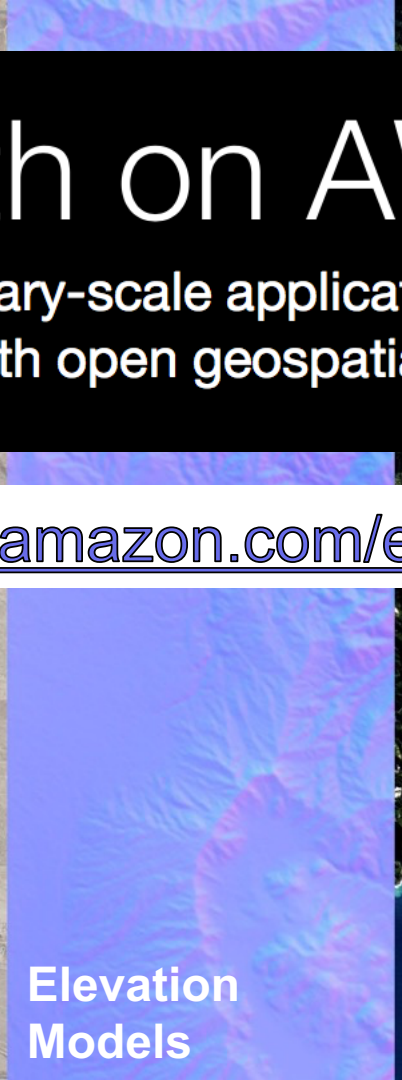
Earth on AWS

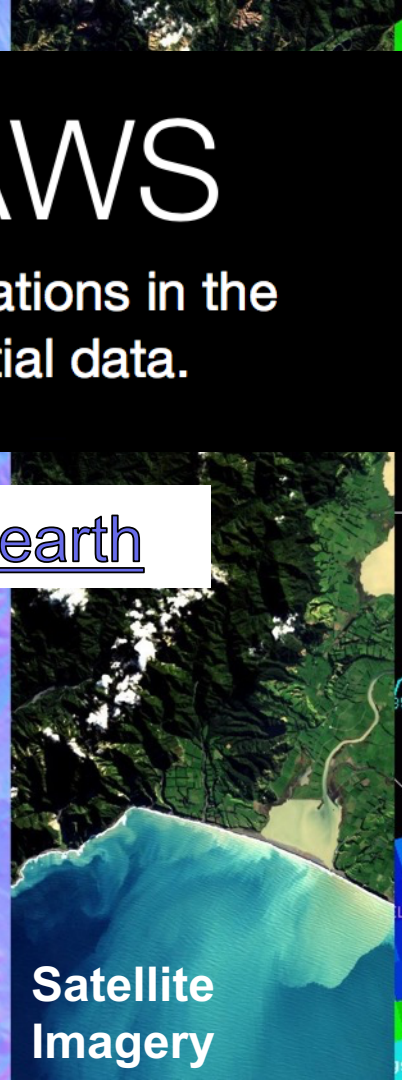
Build planetary-scale applications in the cloud with open geospatial data.

aws.amazon.com/earth









Bring important tools to the cloud

Thredds (see Docker talks John/Josh/Carlos)
Netcdf
HDF5
…

amazon web services

Beyond hot storage (S3) and cold storage (Glacier)

AWS Machine Learning

AWS Kinesis: ingest and process data streams

AWS IoT: capture data from lots of small devices, e.g. sensor networks, smart cities, …

AWS Athena: run SQL queries straight on S3 data (no cluster required)

AWS Lambda: define automated “triggered” compute actions on S3 objects (no servers required)

AWS Rekognition: Image recognition

… 90+ services to make the most of your data

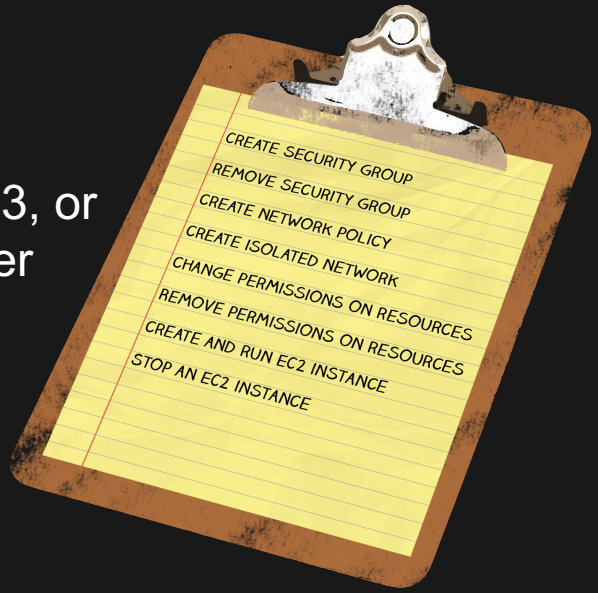
Architected and Audited for Security

Certifications and accreditations for workloads that matter

AWS CloudTrail - AWS API call logging for governance & compliance

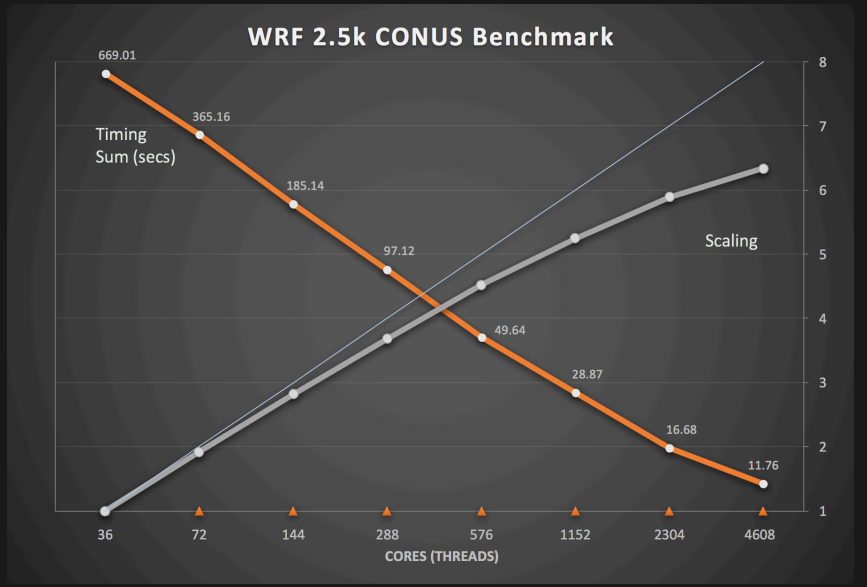
Log and review user activity

Stores data in S3, or archive to Glacier



amazon web services

AWS and HPC



amazon web services

AWS for High Performance Computing…

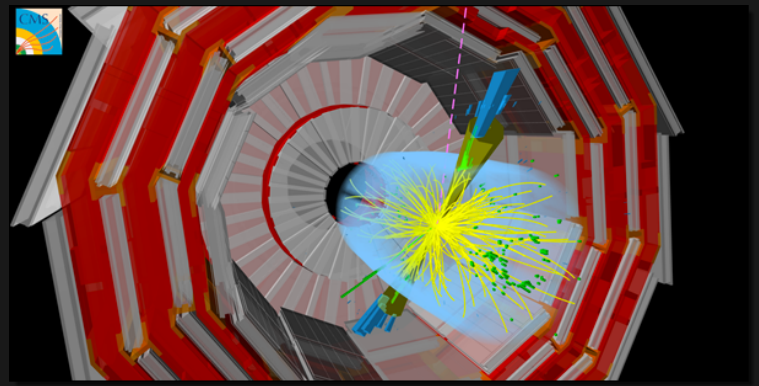
- Many types of research or commercial HPC on AWS:
  - High-energy physics simulations
  - Weather and climate modeling and prediction
  - Analysis of fluids, structures, and materials
  - Thermal and electromagnetic simulations
  - Genomics, proteomics, and molecular dynamics
  - 3D rendering and visualizations
  - Deep learning training and inference
  - Seismic and reservoir simulations

- Many HPC applications benchmarked on 100s-1000s of compute cores.
- AWS recently contributed performance optimizations to the OpenMPI code base.
- AWS Partners make a living offering HPC solutions built on AWS.
- Ongoing investments in CPU/GPU/FPGA EC2 instances *(Skylake instances coming!)*, other hardware, graphics, automation, ISV partnerships, etc.

amazon web services

High Energy Physics with CERN and Fermilab

- Fermilab is one of the Tier 1 data centers for the CMS experiment (at CERN)
- Participated in finding the Higgs Boson to understand mass
- Launched the High Energy Physics Cloud Project in June, 2015
- Recently **added 58,000 cores** (or 4x increase in Fermilab capacity) to simulate 500 million events over 10 days
- AWS allowed FermiLab to burst capacity for large-scale data analysis, which on-prem systems were unable to do



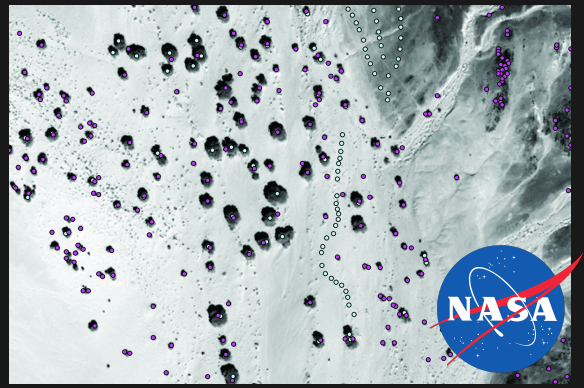
Source: https://aws.amazon.com/blogs/aws/experiment-that-discovered-the-higgs-boson-uses-aws-to-probe-nature/

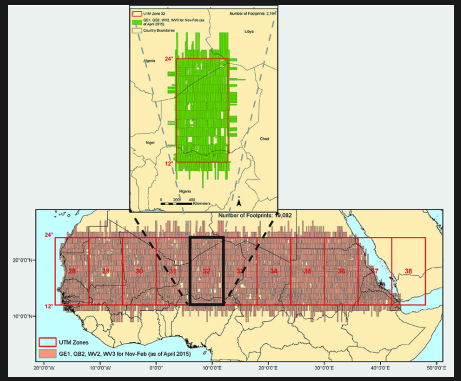
amazon web services

NASA & Cycle Computing – Climate Research

- Mosaicking 2,500+ QuickBird satellite images into 100-kilometer (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 potentially reduce total processing time to less than 1 month from 10 months





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

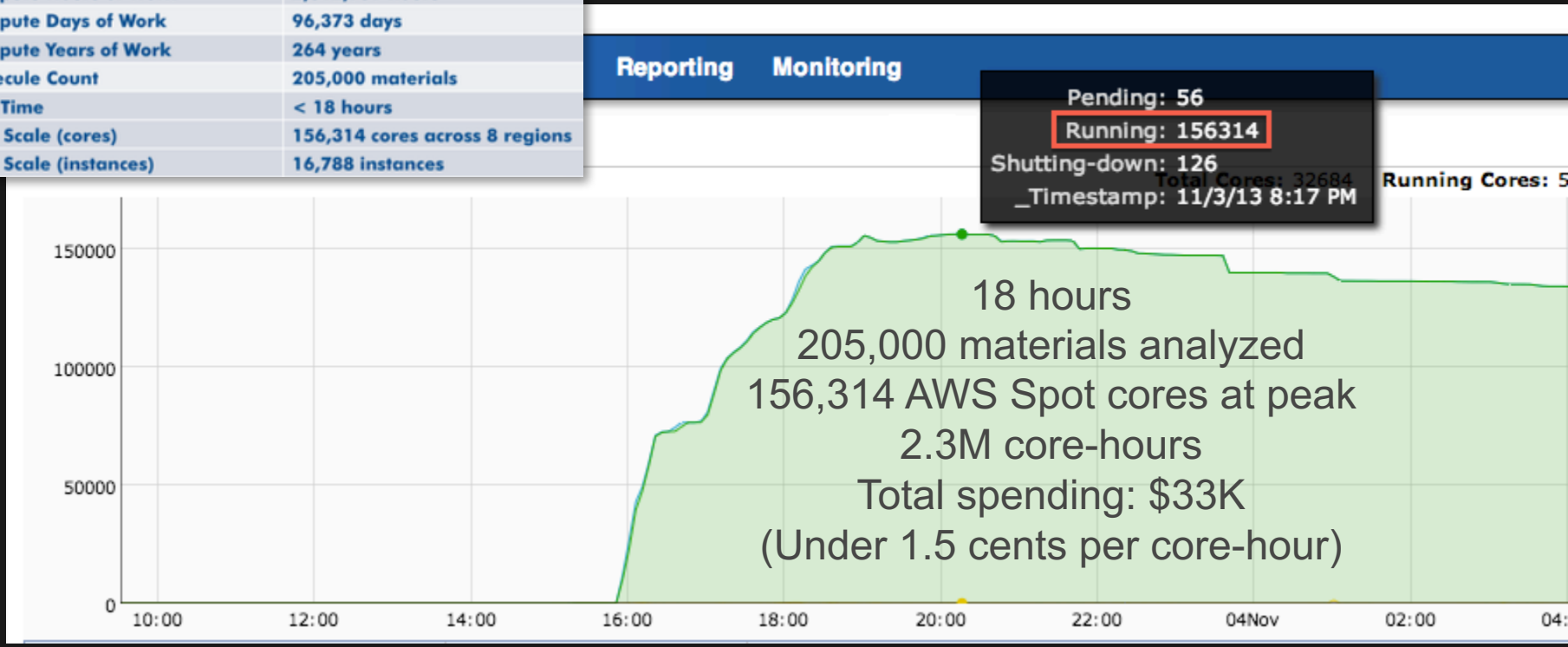
CYCLECOMPUTING

amazon web services

Scale in the Cloud

Metric	Count
Compute Hours of Work	2,312,959 hours
Compute Days of Work	96,373 days
Compute Years of Work	264 years
Molecule Count	205,000 materials
Run Time	< 18 hours
Max Scale (cores)	156,314 cores across 8 regions
Max Scale (instances)	16,788 instances

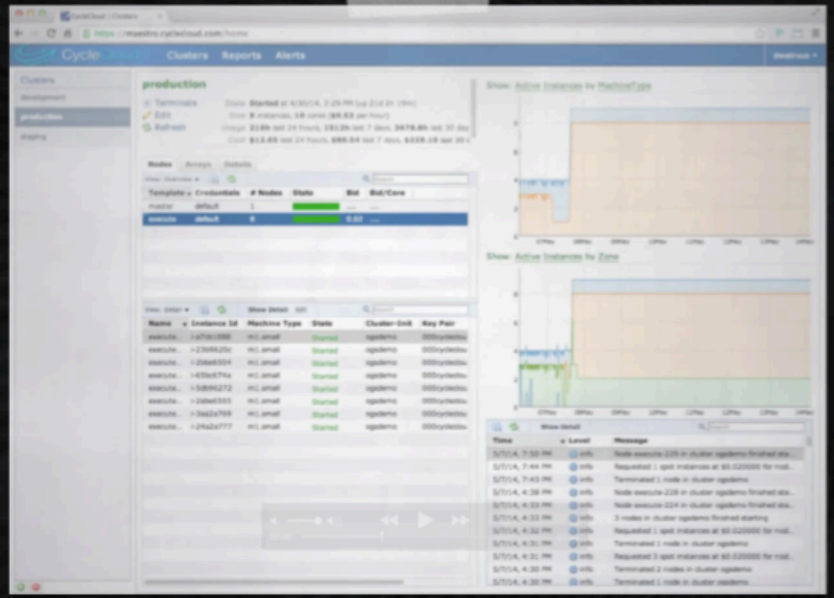
CYCLECOMPUTING
LEADER IN CONDOR GRID COMPUTING SOLUTIONS



Novartis

39 years of computational chemistry in 9 hours

Novartis ran a project that involved virtually screening 10 million compounds against a common cancer target in less than a week. They calculated that it would take 50,000 cores and close to a $40 million investment if they wanted to run the experiment internally.

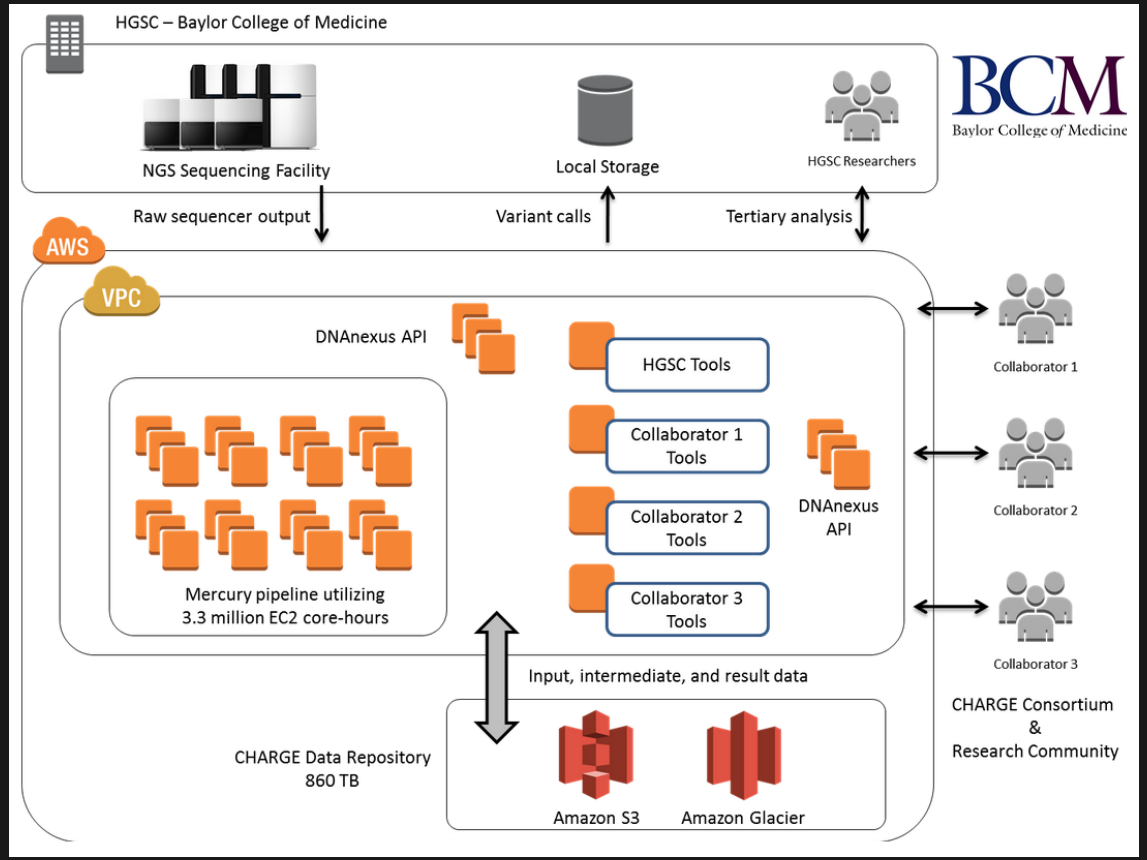


Partnering with Cycle Computing and Amazon Web Services (AWS), Novartis built a platform thst ran across 10,600 Spot Instances (**~87,000 cores**) and allowed Novartis to conduct 39 years of computational chemistry in 9 hours for a cost of **$4,232**. Out of the 10 million compounds screened, three were successfully identified.

amazon web services

Life Sciences

DNAnexus



Baylor CHARGE project:

- Genomics analysis on 14,000 participants
- 24 terabases of sequencer content each month
- 1PB of raw data storage
- 21,000 AWS compute cores at peak
- Initial analysis completed in 10 days

amazon web services

WRF is already used on AWS

Scientists at research institutions

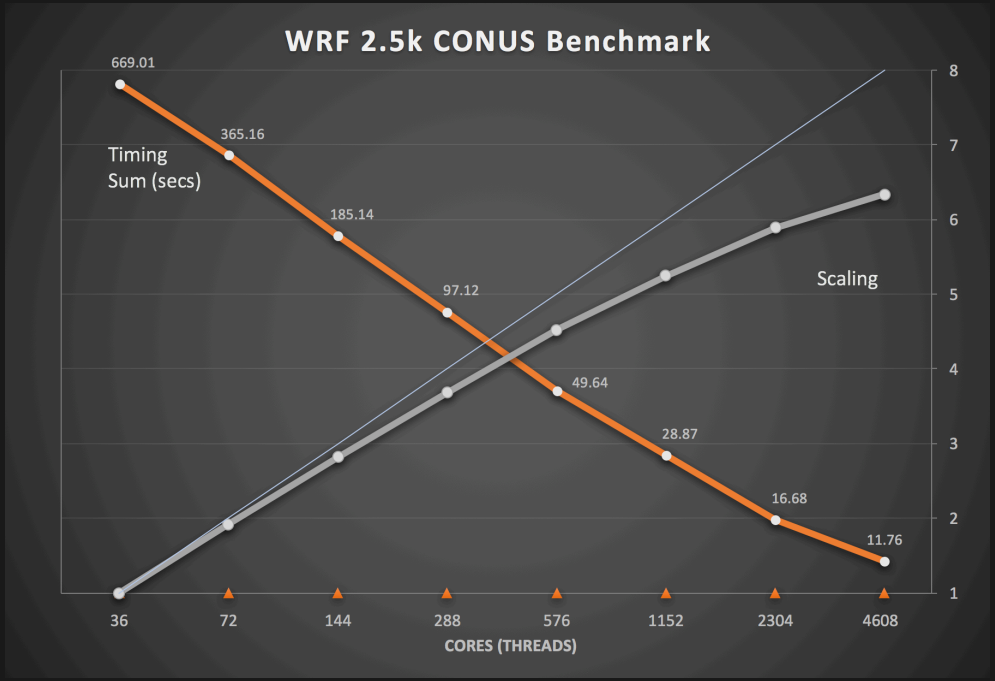
Financial sector

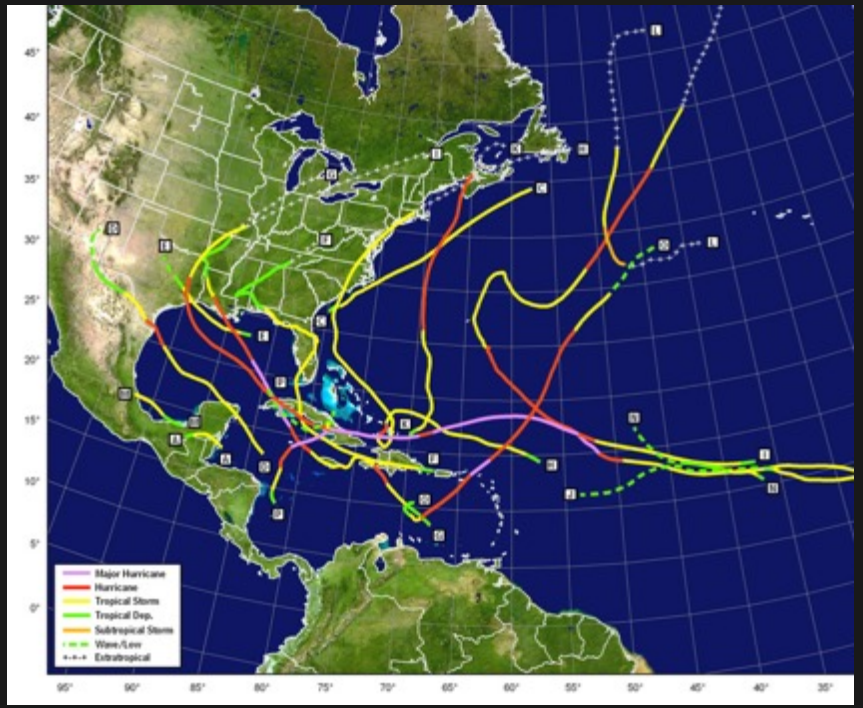
Commercial entities *(e.g. The Weather Company, WeatherRisk)*

EC2 compute clusters

Docker (container) solutions

Weather Prediction





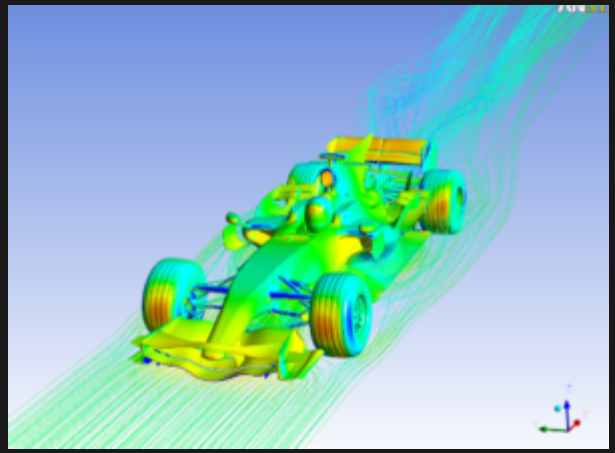
WRF Scaling and Performance on AWS

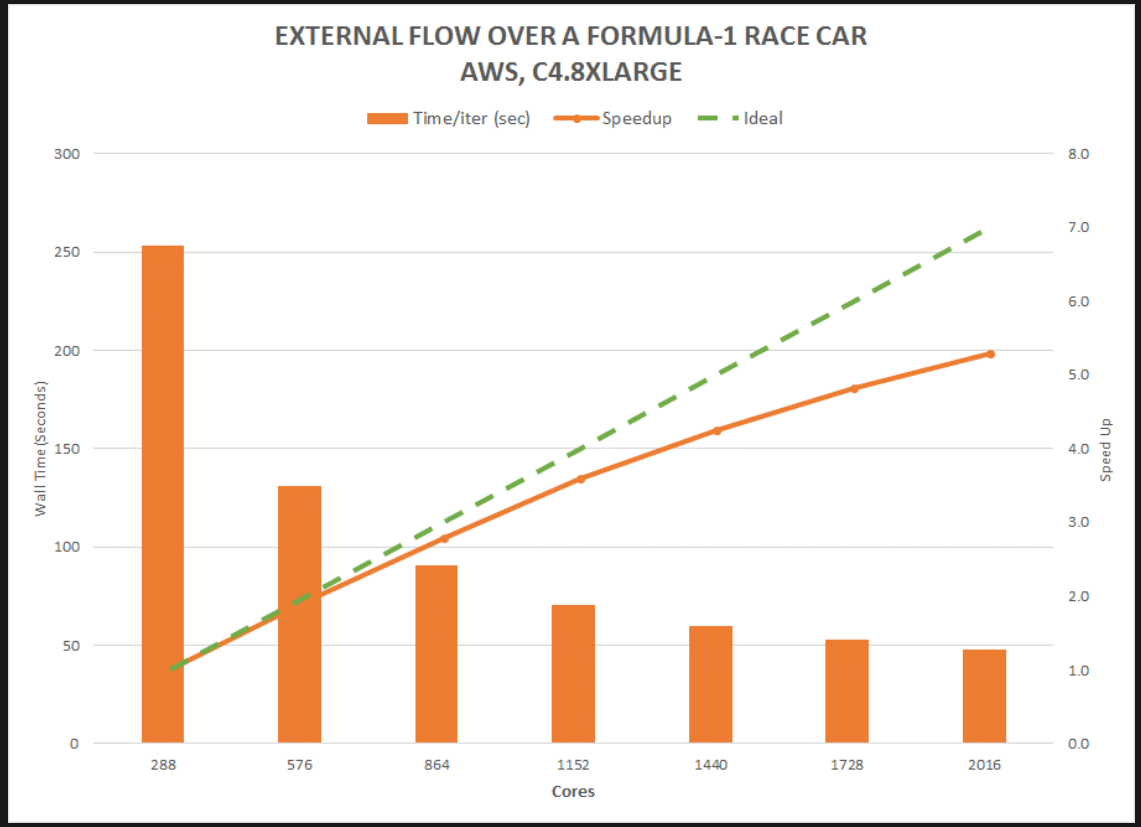
amazon web services

Performance for Fluid Dynamics on AWS

ANSYS Fluent

- AWS c4.8xlarge
- 140M cells
- F1 car CFD benchmark





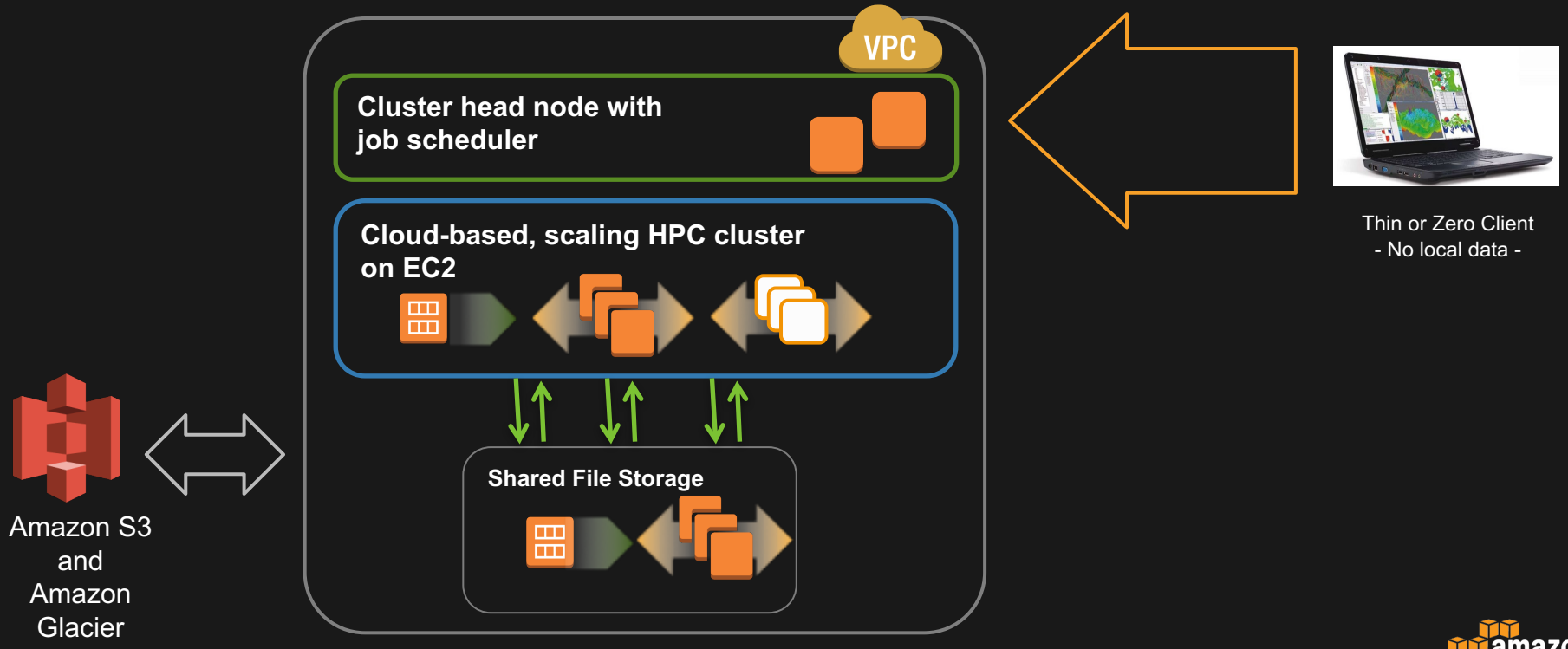
http://www.ansys-blog.com/simulation-on-the-cloud/

! Don’t forget:

Cutting-edge activities steal the show

But most of the users run small and medium size jobs.

EC2 Elastic Compute Cluster Stack



Performance Considerations for HPC on AWS

Test using larger, real-world examples
- Use large cases for testing: do not benchmark scalability using only small examples

Domain decomposition
- Choose number of cells per core for either per-core efficiency or for faster results

Instance types
- C4 or M4 are best choices today

Network
- Use a placement group
- Enable enhanced networking

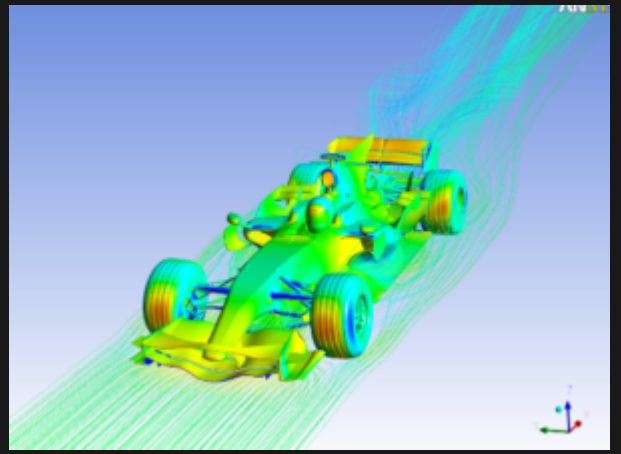
OS version
- Use Amazon Linux or a version 3.10 or later Linux kernel

Processor states and affinity
- Use P-states to reduce processor variability
- Use CPU affinity to pin threads to CPU cores

MPI libraries
- Intel MPI recommended

Hyper-threading
- Test
- Usu

"BEST PRACTICES" ARE AVAILABLE



Universities have an AWS Account Manager and Solutions Architect
Ask them for help!

amazon web services

Alces Flight for HPC compute clusters

Self-scaling HPC clusters instantly ready to compute, billed by the hour and use the AWS Spot market by default, so they're very low cost

850+ popular scientific applications

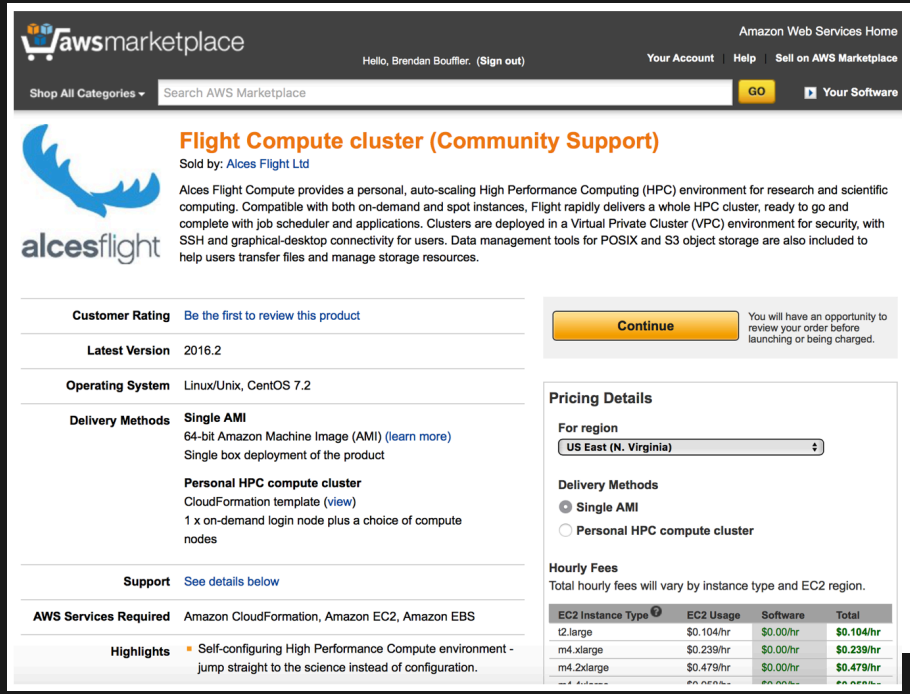
- **Pre-installed**
- **Multiple versions, complete with libraries and compiler optimizations, ready to run**

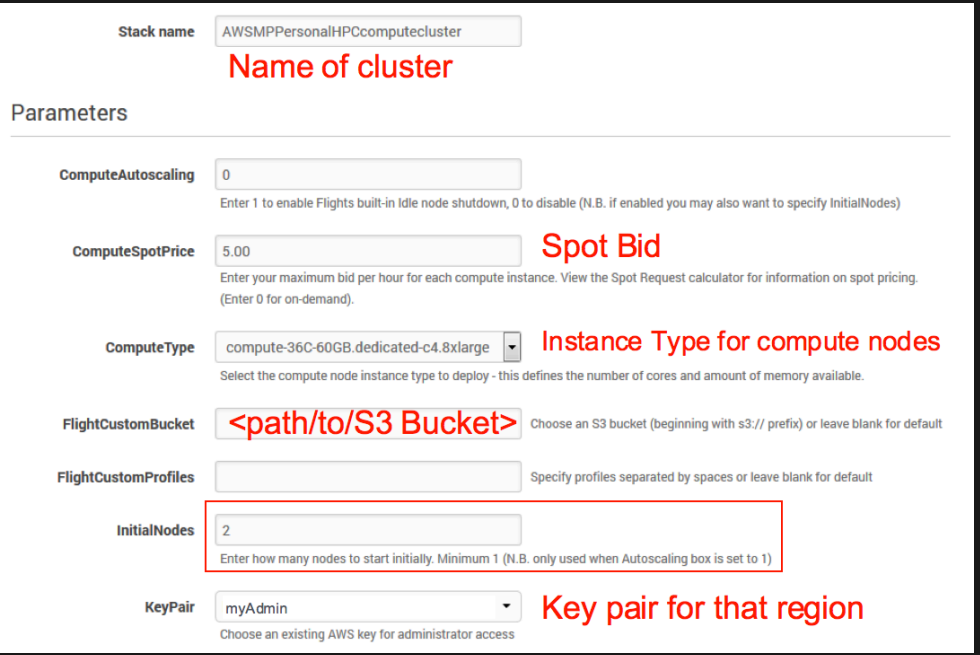
Available via the AWS Marketplace (the cloud's "App Store")

awsmarketplace

amazon
web services

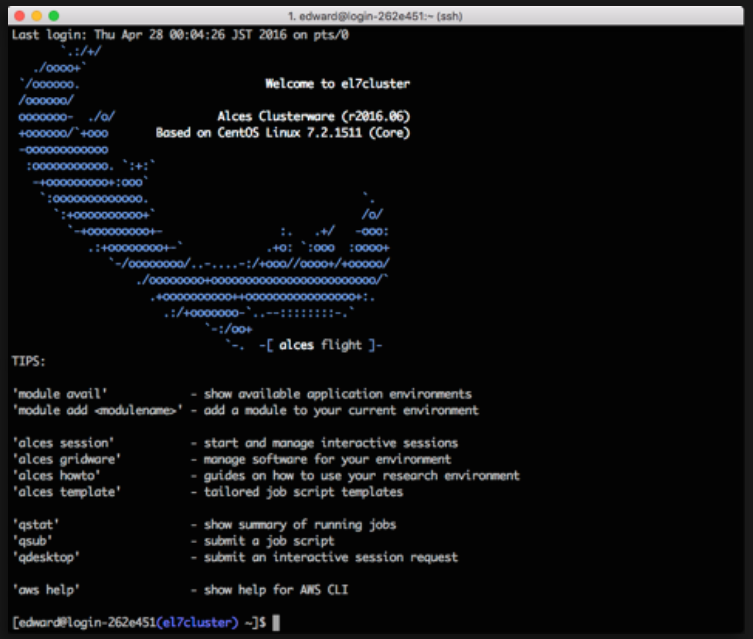
Create an Alces cluster from the AWS Marketplace



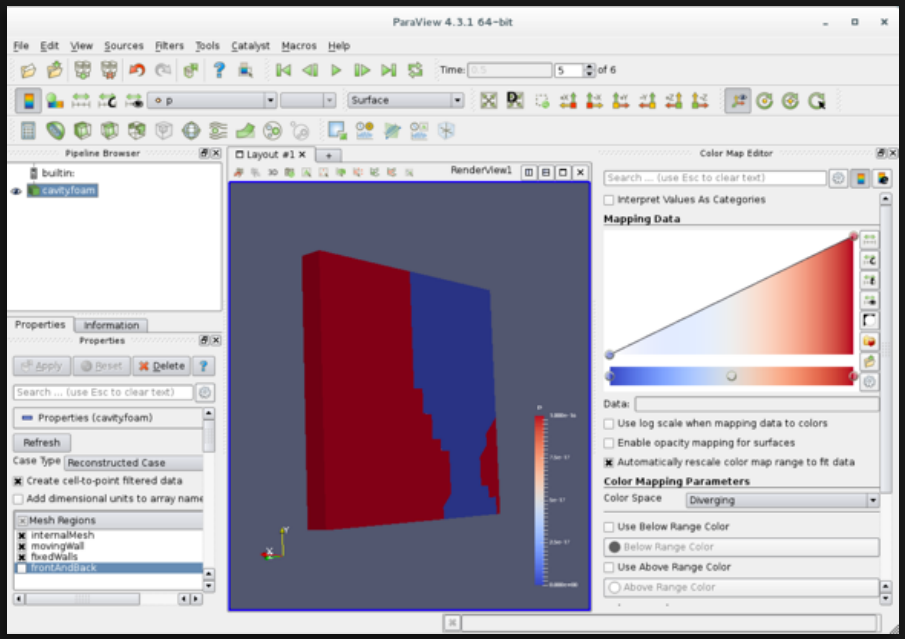


amazon web services

Then access from command line or console



Command Line (ssh)



Graphical Console

amazon web services

Familiar Features…

The headnode offers a job scheduler, compilers, libraries and MPI.

Compute nodes run your jobs.

The hundreds of scientific & HPC codes in the enormous catalogue are all the same apps that you go to big HPC centers to use.

Try Alces Flight: aws.amazon.com/marketplace

Help & docs: www.alces-flight.com

Tutorials: http://tinyurl.com/alcesFlightYoutube

+Additional Benefits

Your cluster is personal. You can change the way it works if you need to.

You only pay for what you use.

It can scale as large as required.

Deploys in minutes.

No job queues in the AWS cloud.

Install or create optimized software configurations.

No management overhead.

Administrator control for the freedom to customize the Alces environment .

Access to the graphical console of the control node to run visual apps at scale

amazon web services

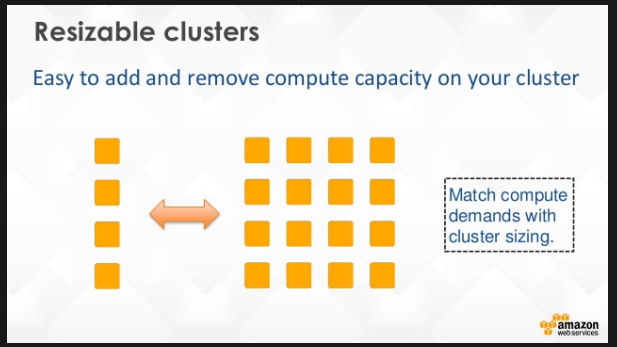
Schedulers

Your choice:

- SGE
- Slurm
- Torque
- OpenLava
- PBS Pro
- You have full rights so you can always install your favorite, custom scheduler
- Or skip it

Smart scaling:

- Scheduler knows how much work is waiting in the job queue
- Triggers expansion of compute fleet if needed (up to limit)
- Terminates idle compute nodes so you don't pay for idle nodes.



amazon web services

[alces@login1(myAWSomeHPCDemo) ~]$ alces gridware list

```
[alces@login1(myAWSomeHPCDemo) ~]$ alces gridware list
main/apps/Q/13012015                main/apps/R/3.2.3                   main/apps/R/3.2.5                   main/apps/R/3.3.0
main/apps/anaconda3/2.5.0           main/apps/bamtools/2.3.0            main/apps/bcftools/1.3              main/apps/bedtools/2.25.0
main/apps/bowtie/1.1.0              main/apps/bowtie2/2.2.6             main/apps/breakdancer/1.3.5.1       main/apps/bsoft/1.9.1
main/apps/bwa/0.7.8                 main/apps/cmake/3.4.3               main/apps/cmake/3.5.2               main/apps/codesaturne/4.0.5
main/apps/cortexcon/0.05            main/apps/cpanminus/1.5017          main/apps/cufflinks/2.2.2.20150701  main/apps/cython/0.23.4
main/apps/diffreps/20150410         main/apps/ead/1.0.3                 main/apps/emboss/6.3.1              main/apps/espressomd/3.1.2
main/apps/fastqscreen/0.4.1         main/apps/fasttree/2.1.3            main/apps/fastx/0.0.14              main/apps/ffmpeg/1.2.1
main/apps/freebayes/1.0.2           main/apps/ghc/7.8.2                 main/apps/gnuplot/5.0.2             main/apps/gpuburn/0.6
main/apps/grace/5.1.25              main/apps/hdf5/1.8.13               main/apps/hmmer/3.1b1               main/apps/hpl/2.1
main/apps/htseq/0.6.1p1             main/apps/htslib/1.3                main/apps/idr/2.0.2                 main/apps/igv/2.3.51
main/apps/imb/4.0                   main/apps/iozone/3.420              main/apps/jags/4.2.0                main/apps/khmer/2.0
main/apps/landsatutil/0.13.0        main/apps/macs/2.1.0.20150731       main/apps/mats/3.2.2b               main/apps/memesuite/4.11.1
main/apps/memtester/4.3.0           main/apps/mono/3.0.6                main/apps/nucleoatac/0.3.1          main/apps/openfoam/3.0.1
main/apps/paraview/4.3.1            main/apps/patchelf/0.9              main/apps/perl/5.20.2               main/apps/phantompeakqualtools/1.1
main/apps/picard/2.1.0              main/apps/pip/8.1.2                 main/apps/pycogent/1.5.3            main/apps/pyprophet/0.18.3
main/apps/python/2.7.8              main/apps/python3/3.3.3             main/apps/python3/3.4.3             main/apps/rdpclassifier/2.2
main/apps/rmblast/2.2.27            main/apps/samtools/0.1.18           main/apps/samtools/0.1.19           main/apps/samtools/1.3
main/apps/screed/0.9                main/apps/seqtk/1.0                 main/apps/setuptools/2.1            main/apps/setuptools/15.1
main/apps/sip/4.16.3                main/apps/soapdenovo/2r240          main/apps/stampy/1.0.22.1848        main/apps/star/2.4.2a
main/apps/star/2.5.2a               main/apps/tau/2.25.1                main/apps/tax2tree/1.0              main/apps/tophat/2.1.0
main/apps/trimmomatic/0.35          main/apps/trinity/2.2.0             main/apps/vcftools/0.1.12b          main/apps/velvet/1.2.10
main/apps/viennarna/2.1.1           main/apps/virtualenv/13.1.2         main/compilers/gcc/5.1.0            main/libs/arpack-ng/3.3.0
main/libs/atlas/3.10.2              main/libs/biomformat/1.1.2          main/libs/bioperl/1.6.923           main/libs/biopython/1.63
main/libs/blas/3.6.0                main/libs/boost/1.55.0              main/libs/boost/1.60.0              main/libs/eigen/3.2.4
main/libs/fftw3/3.3.4               main/libs/freetype/2.6.3            main/libs/gdal/2.1.0                main/libs/gmp/6.1.0
main/libs/gsl/1.16                  main/libs/gtextutils/0.6            main/libs/lapack/3.5.0              main/libs/lasagne/20160403
main/libs/libgit2/0.23.4            main/libs/llvm/3.7.1                main/libs/matplotlib/1.4.3          main/libs/matplotlib/1.5.1
main/libs/mgridgen/1.0              main/libs/mpfr/3.1.4                main/libs/numexpr/2.4.4             main/libs/numpy/1.9.2
main/libs/numpy/1.10.4              main/libs/nvidia-cuda/7.5.18        main/libs/openblas/0.2.15           main/libs/opencv/2.4.9
main/libs/opencv/2.4.12             main/libs/openlibm/0.4.1            main/libs/pandas/0.17.0             main/libs/pcre2/10.21
main/libs/protobuf/2.5.0            main/libs/protobuf/2.6.1            main/libs/pybigwig/0.2.7            main/libs/pysam/0.8.3
main/libs/scikit-image/0.12.3       main/libs/scipy/0.17.0              main/libs/scotch/6.0.3              main/libs/seqan/1.4.1
main/libs/suitesparse/4.5.1         main/libs/tbb/4.4.20160128          main/libs/theano/0.8.0              main/libs/tkimg/1.4.2
main/mpi/openmpi/1.8.5
```

Alces has >850+ applications : NCAR, want to add WRF?

https://gridware.alces-flight.com/software

amazon web services

Coming soon: easy HPC by Alces-Launch

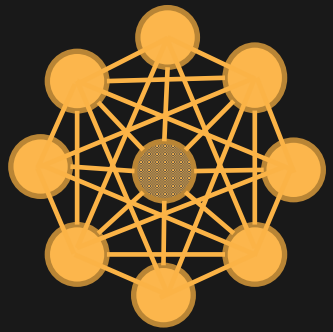
HPC Administrator:

* Predefines cluster templates for her users in terms of installed applications, number and type of instances, and lifetime of the cluster (capped core-hours and capped cost)

Scientist:

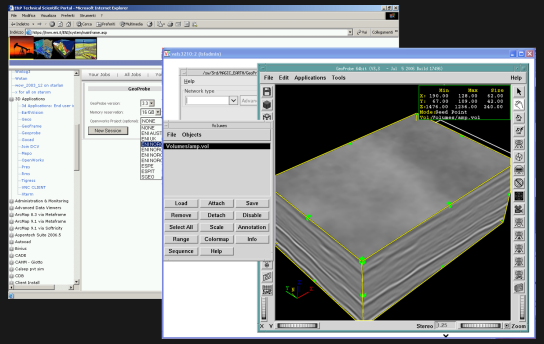
- Chooses right cluster from a few friendly tiles (“WRF, medium cluster, 400 core hours”)
- Up and running with a few clicks and without technical questions or unfamiliar cloud lingo

Other HPC Cluster strategies



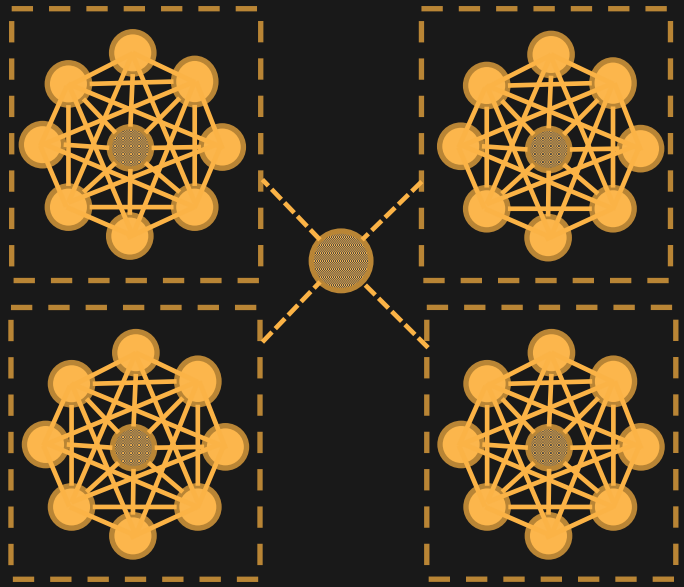
Other Cluster Tools

CfnCluster
CloudyCluster
~~StarCluster~~
…



Hide the Cloud

EnginFrame
HTCondor
…



Ensemble?

Run all members at once!

Other HPC Cluster strategies
AWS Marketplace – Technology Partners

INQDO

acellera®

SENTINEL Hub

intel®

AceCloud

Overleaf

DNAnexus

STERLING GEO
ENGINEERING GEOGRAPHY

PIRONET
A CANCOM COMPANY

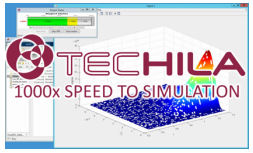
alcesflight

BeeGFS®
developed by Fraunhofer

figshare

SevenBridges
genomics

CFD Direct



the SERVER LABS
the IT architects

MathWorks®

arcus global

illumina®

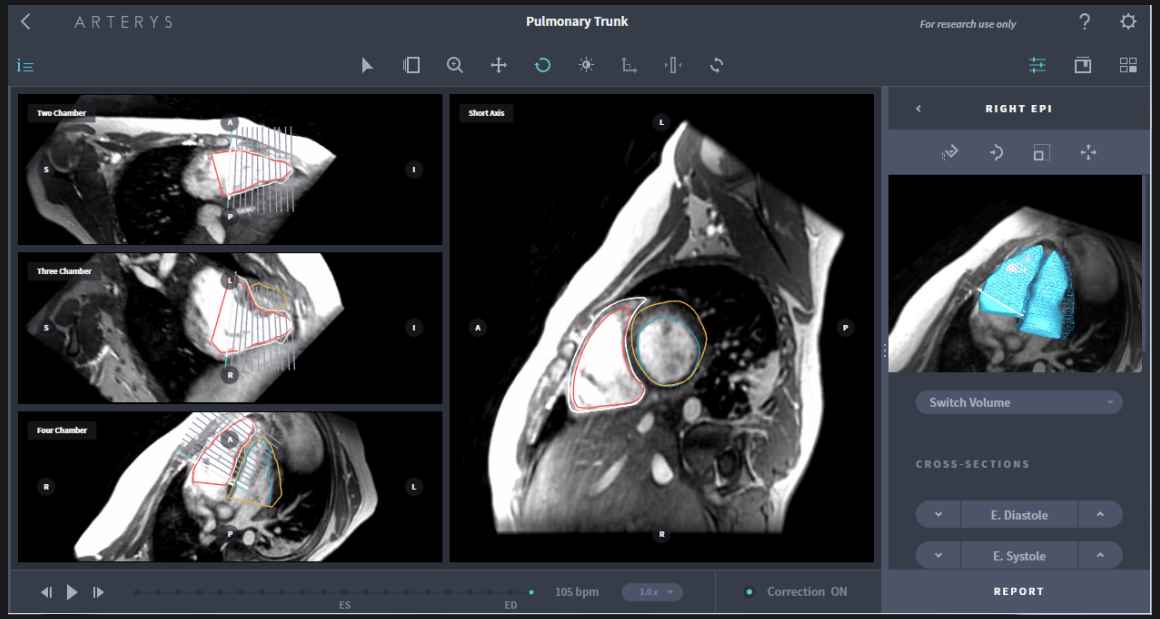
Secure Graphics and Collaboration

Cloud can be used for pre-and post processing as well as HPC

- Use GPUs in the cloud for remote rendering and remote desktops

Cloud is more secure for collaboration

- Encrypt the data in flight and at rest
- Manage your own keys and credentials
- Deliver pixels to your collaborators, not the actual data



amazon web services

Beyond servers (EC2)

AWS ECS: Run Docker containers

AWS Batch: just submit Docker jobs to a queue (AWS manages the compute infrastructure for you)

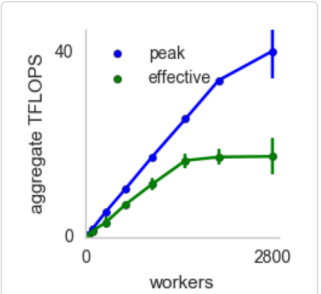
AWS Lambda: short-lived compute actions, e.g. Python code (serverless, billed by the second)

pywren

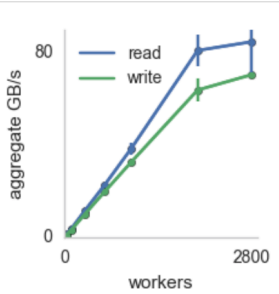
Pywren lets you run your existing python code at massive scale via AWS Lambda

```
def my_function(b):
  x = np.random.normal(0, b, 1024)
  A = np.random.normal(0, b, (1024, 1024))
  return np.dot(A, x)

pwex = pywren.default_executor()
res = pwex.map(my_function, np.linspace(0.1, 100, 1000))
```



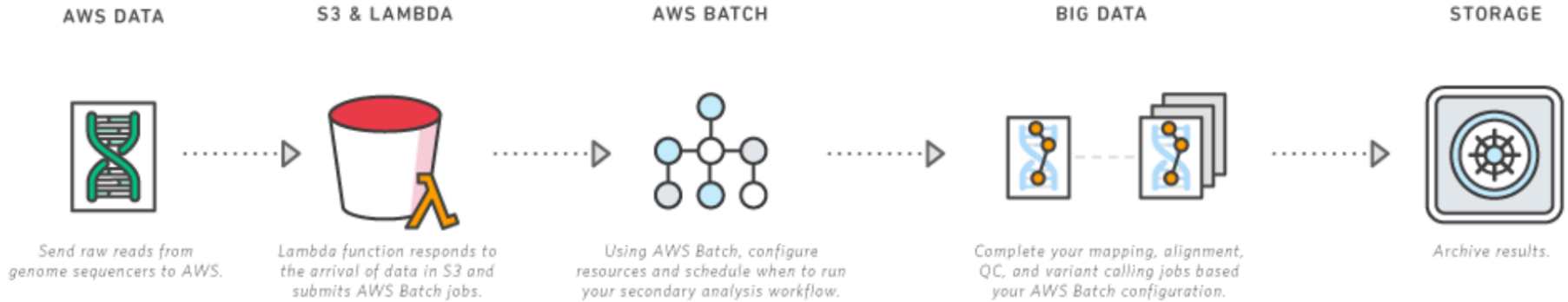
40 TFLOPS on Lambda [more]



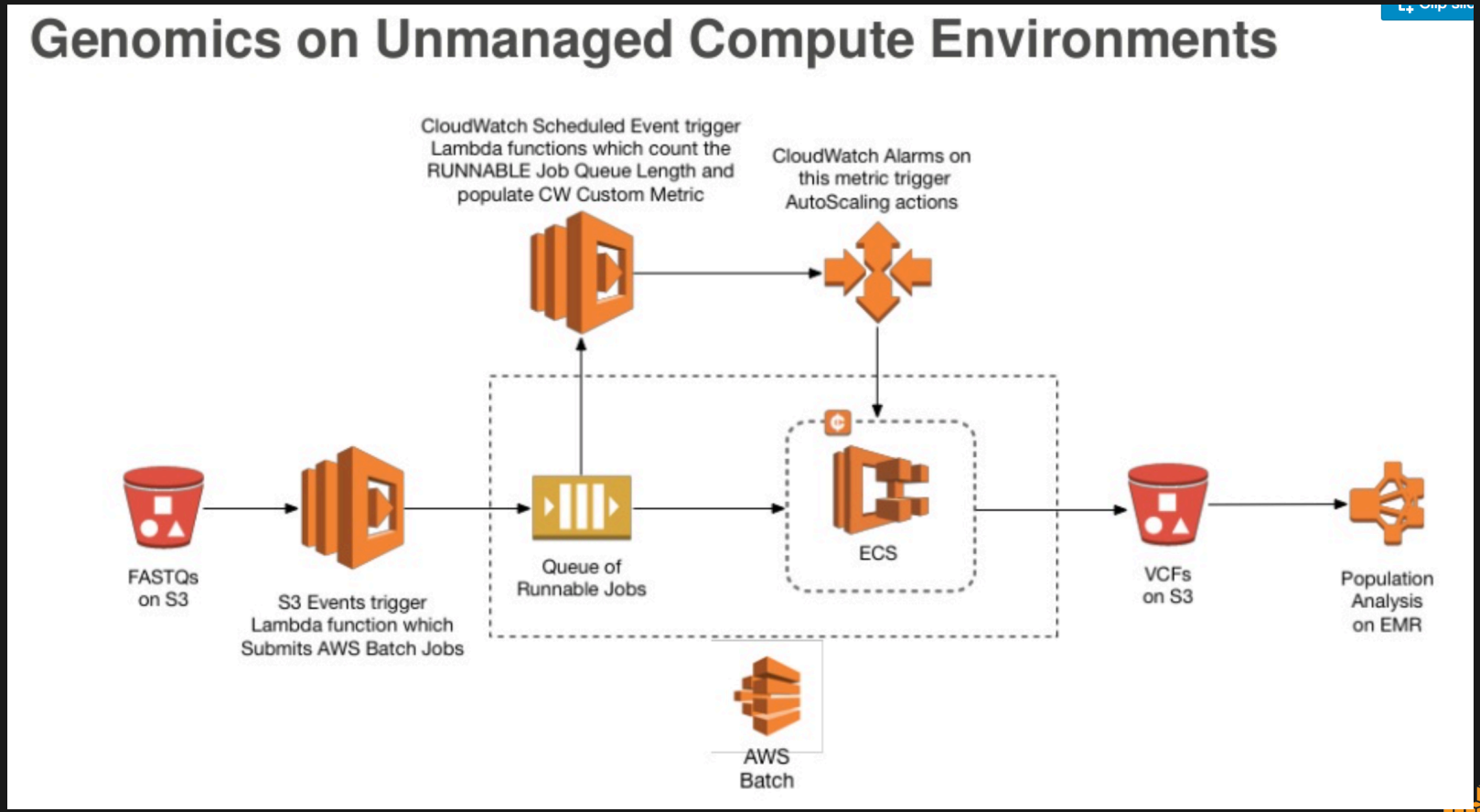
80 GB/sec read and 60 GB/s write to S3 [more]

pywren

Example Genomics workflow on AWS



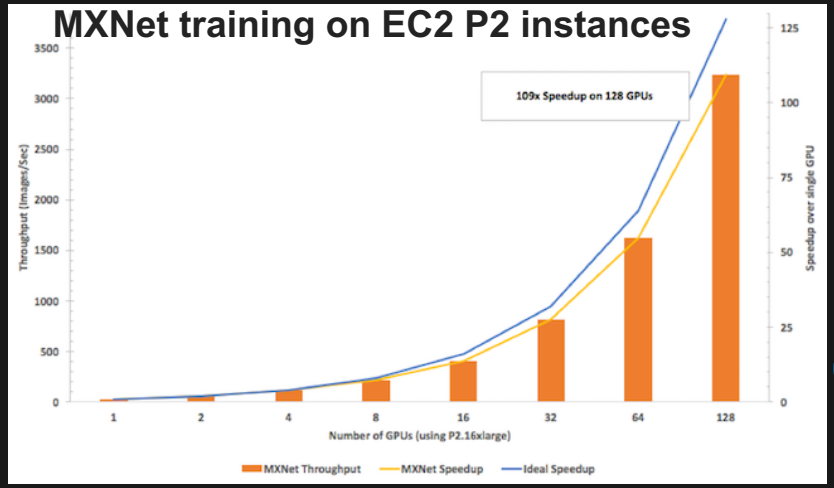
Example Genomics workflow on AWS



Biggest Trend Today in HPC:

Machine Learning, A.I., and Deep Learning

- Recommendation engines
- Voice recognition / chat bots
- Fraud and intrusion detection
- Picture recognition and tagging
- Document tagging and classification
- Autonomous driving and robotics



TensorFlow

Caffe

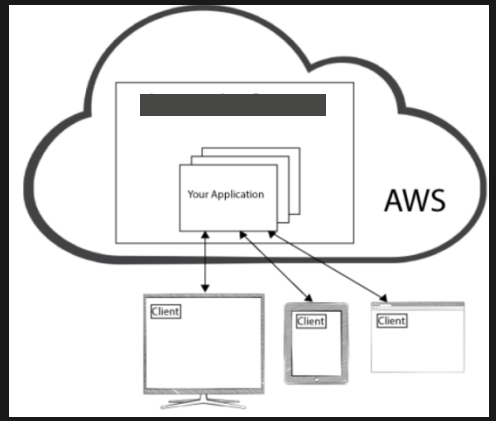
dmlc mxnet

theano

torch

AWS and collaboration

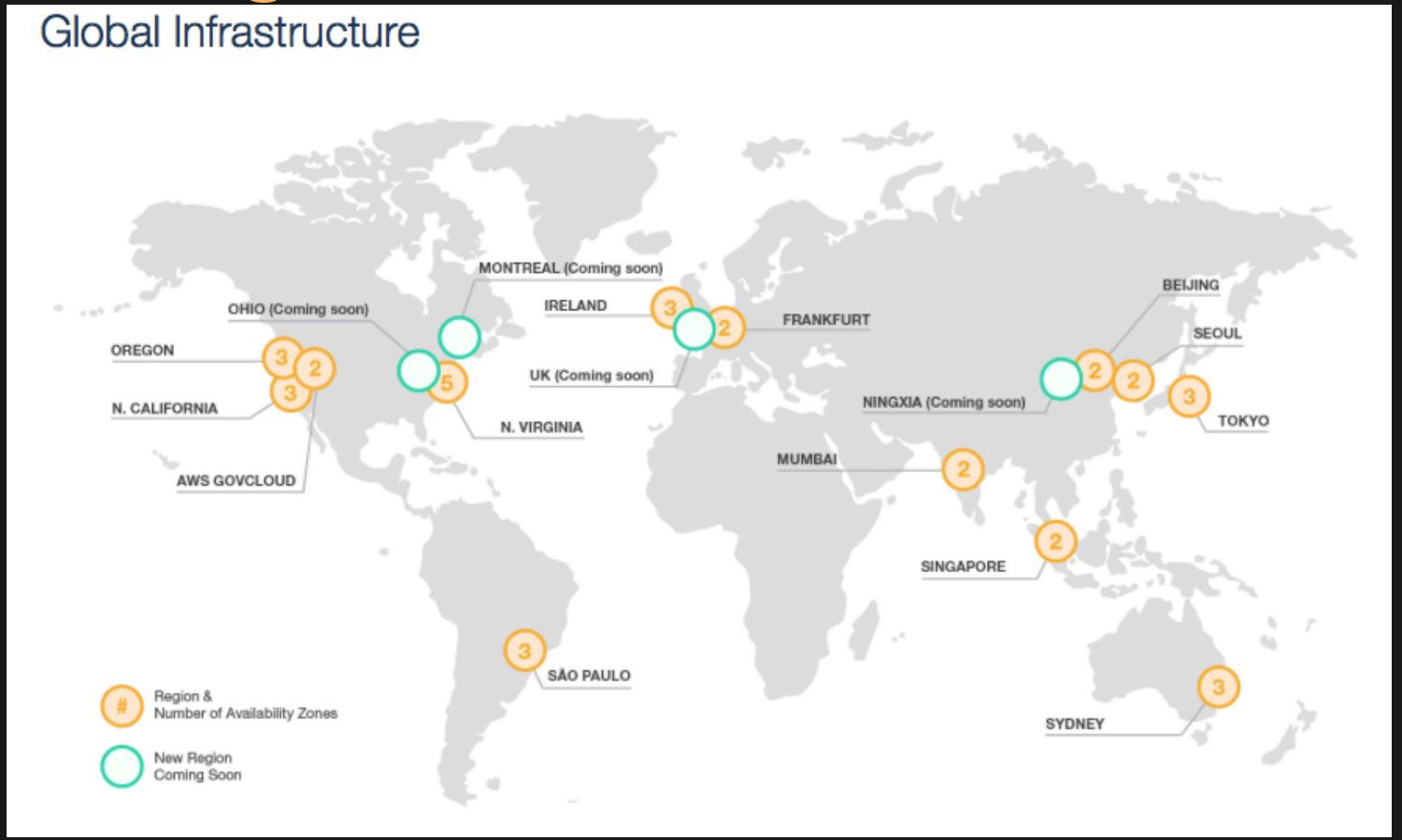
Data Sharing and Global Collaboration



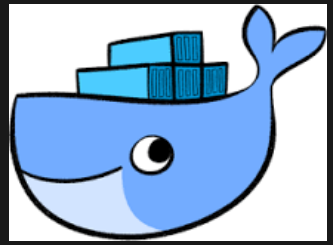
Bring the users to the data, don't send the data to the users

amazon web services

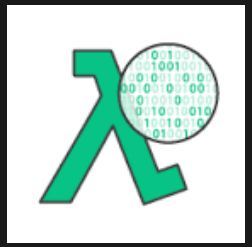
Enabling Global Collaboration



Sharing applications on AWS



- Save server as “virtual machine image” (AMI)
- Save container as Docker image
- Save disk volume as EBS snapshot



- Share your AWS Batch Job Definitions or Lambda functions
- Place your application in the AWS Marketplace (you don’t have to charge for it)

- Share as objects in S3

- Submit it as an Alces Gridware app

amazon web services

Sharing workflows and architectures

- Save entire stack as CloudFormation “template”
  - Infrastructure as code: all the elements of an AWS architecture are defined in a text file, called a “CloudFormation template”.
  - AWS executes the template and stands up the prescribed infrastructure.
  - Of course, you can customize it endlessly.
- Share “recipe” for compute cluster customization (Alces Flight customizer, CfnCluster config file, Chef recipe, …)
  - Define a ‘WRF cluster’ in CfnCluster config file:

```
compute_instance_type = c4.8xlarge
ebs_snapshot_id = snap-570ffb0e
Max_queue_size = 20
```

- Build & share platform, e.g. EnginFrame, Galaxy, …

Automate everything

You can use AWS console in web browser (point and click), BUT:

- AWS has APIs for EVERYTHING – you can script ANY AWS operation for automation, and execute API calls from a CLI
- AWS offers extensive SDKs for major languages, e.g. Java, Ruby, Python, .NET, iOS, and even Internet-of-Things (IoT) devices
- Many well-known businesses run on AWS, but hide the cloud completely from the user experience

AWS and the research community

“Where is the ‘Cloud’ button?”

amazon web services

AWS Research Cloud Program

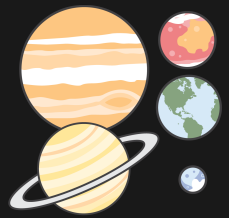
Science first, not servers. Researchers are not professional IT people (nor do they wish to be).

Simple and easily explained procedures to get set up with cloud access.

Budget management tools to ensure that over-spends do not happen.

Best practices to ensure both data and research budgets are safe and privacy is protected.

Fast track to invoice-backed billing & Egress Waiver. No credit card required.

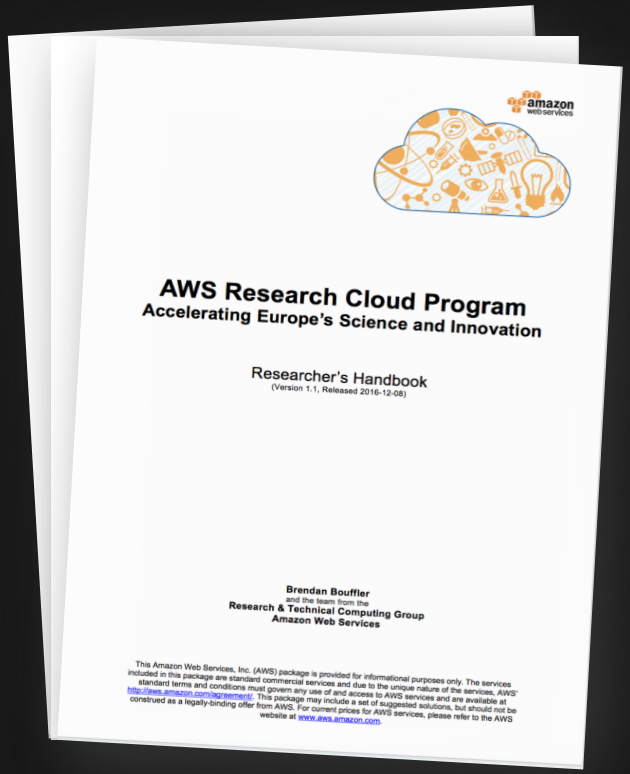


Large catalog of scientific Solutions from partners, including instant clusters from AWS Marketplace.

amazon web services

AWS Researcher's Handbook

The 150-page "missing manual" for science in the cloud.



Written by Amazon's Research Computing community for scientists.

- Explains foundational concepts about how AWS can accelerate time-to-science in the cloud.
- Step-by-step best practices for securing your environment to ensure your research data is safe and your privacy is protected.
- Tools for budget management that will help you control your spending and limit costs (and preventing any over-runs).
- Catalogue of scientific solutions from partners chosen for their outstanding work with scientists.

aws.amazon.com/rcp

amazon web services

Global Data Egress Waiver reduces data transfer cost

All qualifying research customers should use this!

Why?

Researchers strongly need Predictable Budgets

Who?

Available to Degree-granting / Research Institutions

What?

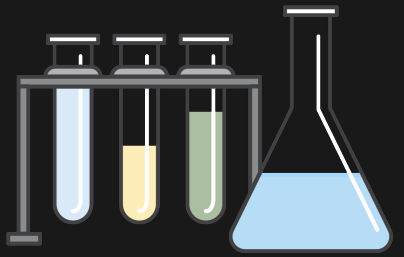
Waives data egress charges from Qualified Accounts (capped at 15% of Total Spend)

How?

Join Research Cloud Program,

Or talk to your Account Team.

AWS peers with National Research and Education Networks for reliable, high-performance connection to/from AWS
e.g. 100Gbps+ to ESnet on W. Coast; also Internet2, Geant, Jisc, SInet, AARnet, …



amazon web services

AWS Cloud Credits for Research provide promotional AWS cloud credits for **anyone** to conduct research on AWS.

aws.amazon.com/research-credits

Focus on building community tools, proof of concept, etc.

amazon web services

Partnership with research funding agencies

- AWS initiated collaborative program with the National Science Foundation (NSF)
- The program provides NSF funds up to $26.5 million in addition to $3 million in AWS Cloud Credits to researchers to perform **cutting edge Big Data research on cloud for a period of 3-4 years**
- Enables cloud based research to foster and accelerate innovation.
- Precedent for similar collaborative programs with other agencies, and international research entities.
- RFP awardees selected by NSF per usual review process, i.e. cloud part of grant funding process

(Q1 2017. Further initiatives under consideration.)

"In today's era of data-driven science and engineering, we are pleased to work with the AWS Research Initiative via the NSF BIGDATA program to provide cloud resources for our Nation's researchers to foster and accelerate discovery and innovation." -- **Dr. Jim Kurose,** Assistant Director of the National Science Foundation (NSF) for Computer and Information Science and Engineering Directorate (CISE)

AWS Educate & Academy

- Self-service membership
- AWS usage credits
- Access to AWS Training content
- Curated content from AWS and educators
- Self-study learning paths and digital badges for students
- Job board for students

- Authorized ~60-hour curriculum developed & maintained by AWS
- Aligned to industry-recognized AWS Certifications
- Educator training and “instructor accreditation”
- Educator & Student discount for AWS Certification exam
- Free Practice Exams

Institutions, educators, and students benefit from both.

amazon
web services

Billing: Budgets and Organizations

AWS Budgets:

- Track which project each expense belongs to
- Get notified when bill reaches a threshold
- Automatically shut down resources when limit exceeded
- See up-to-date spend anytime in web browser

AWS Organizations:

- Central management of multiple AWS accounts
- Control access policies and compliance
- Track costs & Control bill payment across accounts (& get volume discounts)
- Free

Overhead – U of Washington empowers researchers to choose the best solution

Indirect Cost (F&A) Waiver for UW-IT Research Storage, Compute and Cloud Services

Effective April 1, 2015, UW-IT research storage, compute and contracted cloud services are no longer subject to Indirect Costs, also known as Facilities & Administrative (F&A) charges for sponsored research expenditures.

This waiver applies to the following UW-IT services:

Contracted Cloud Services

- Microsoft Azure
- Amazon Web Services

Benefits

Extending the F&A waiver to these UW-IT services allows Principal Investigators (PIs) in a sponsored research program to choose the most appropriate solution to meet their computing needs, whether it be research storage, compute or contracted cloud services through UW-IT, or

U. Washington also created a very active “eScience Institute” that supports campus researchers/educators with Cloud adoption and other needs.

amazon web services

AMPLab & RISELab (Algorithms, Machines, People)

- Collaborative 5-year effort between UC Berkeley, NSF, and industry partners (2012-2016) – AWS is founding partner
- Students and researchers AMPLab used AWS to rapidly prototype and develop new systems at a scale and with a speed not possible before
- Resulted in Apache Spark, developed on AWS, and integrated with AWS core services

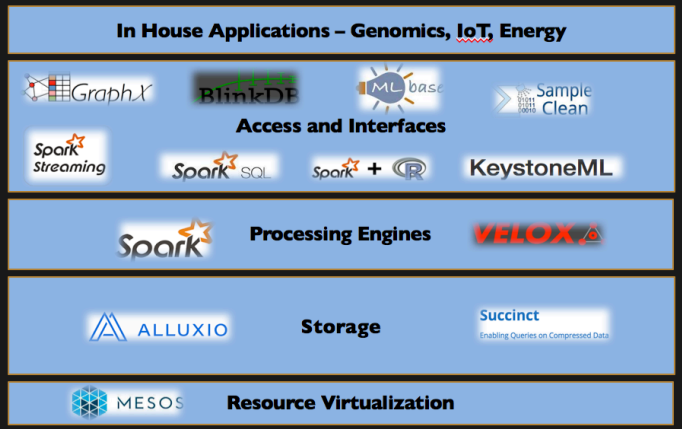
From batch data to advanced analytics

Algorithms
- Machine Learning, Statistical Methods
- Prediction, Business Intelligence

Machines
- Clusters and Clouds
- Warehouse Scale Computing

People
- Crowdsourcing, Human Computation
- Data Scientists, Analysts

Berkeley Data Analytics Stack



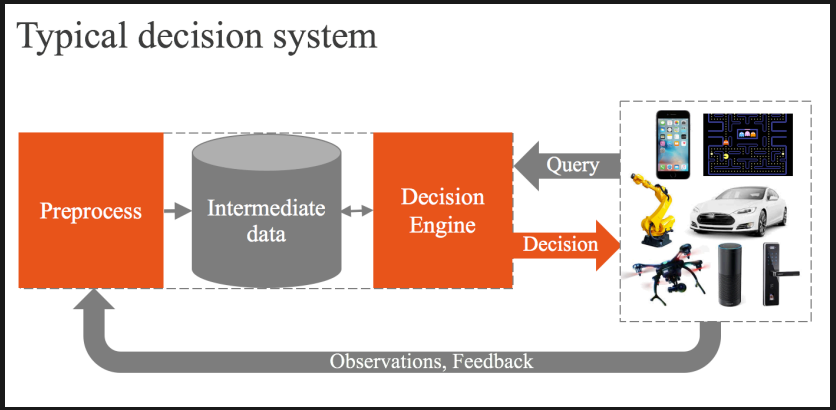
https://amplab.cs.berkeley.edu

AMPLab & RISELab (Real-time Intelligent Secure Execution)

- Collaborative 5-year effort between UC Berkeley, National Science Foundation, and industry partners. (2017-2021) – AWS is founding partner

Data only as valuable as the decisions it enables

Develop open source platforms, tools, and algorithms for intelligent real-time decisions on live-data



From live data to real-time decisions

Thank You

jorissen@amazon.com

Monday morning, 9AM – let's:

DATA

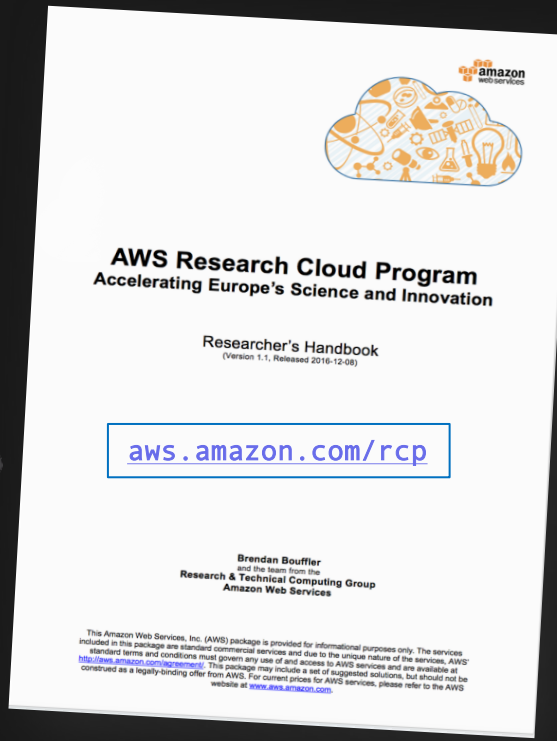
- Identify top 2-3 impactful data sets in need of better accessibility
- Discuss hosting on AWS cf. NEXRAD

MODELING

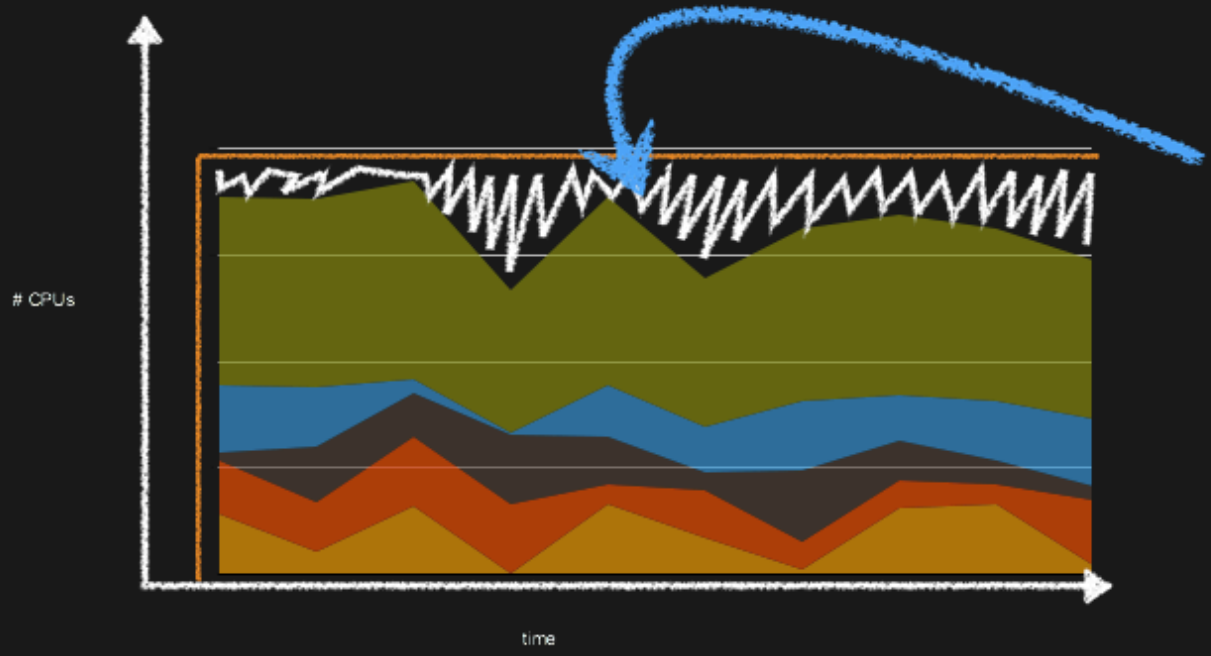
- Look at handful of main models (WRF, etc.)
- Can NCAR & AWS create approved WRF-on-AWS etc. for community?

WORKFLOWS

- Integrate models with HPC tools & platforms?
- Which workflows to prioritize?



Alces uses the “Spot Market” to save $$ (unless you tell it not to)



Spot Market

Our ultimate space filler.

Spot Instances allow you to name your own price for spare AWS computing capacity.

Great for workloads that aren’t time sensitive, and especially popular in research (hint: it’s really cheap).

AWS Compute Consumption Models

On-Demand

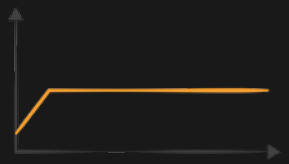
Pay for compute capacity by the hour with no long-term commitments

For spiky workloads, or to define needs

Reserved

Make a low, one-time payment and receive a significant discount on the hourly charge

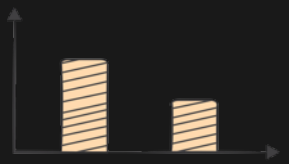
For committed utilization



Spot

Bid for unused capacity, charged at a Spot Price which fluctuates based on supply and demand

For time-insensitive or transient workloads

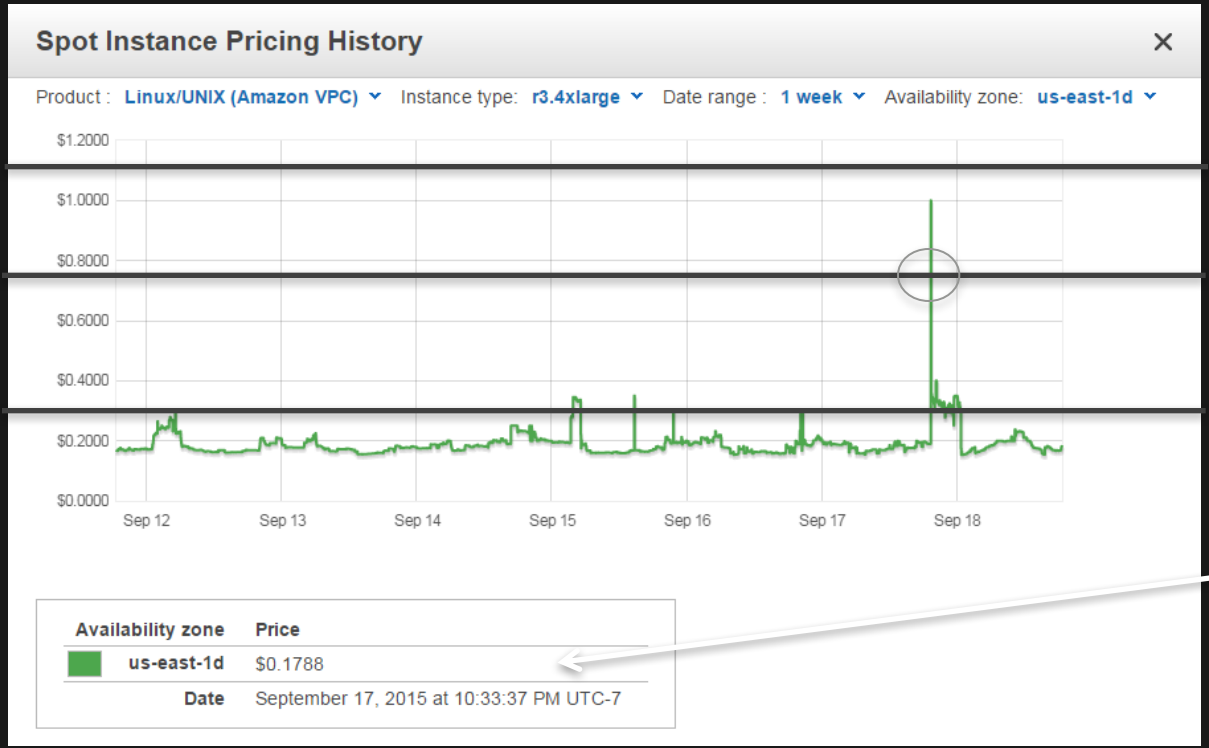


Spot Rules are Simple

Spot is a market in which the price of compute changes based on supply and demand

You'll never pay more than your bid. When the market exceeds your bid you get 2 minutes to wrap up your work. Time to checkpoint!

Bid Price Vs Market Price



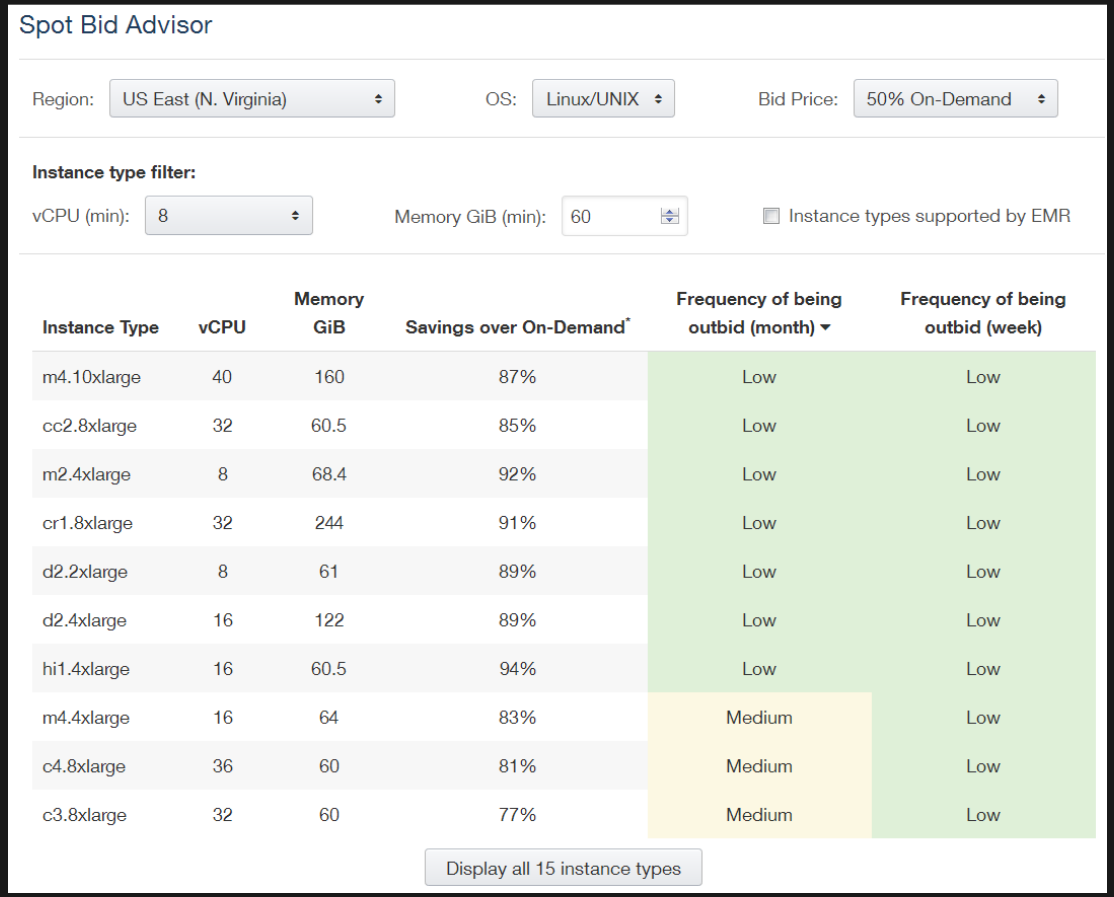
75% Bid

50% Bid

25% Bid

You pay the market price

Spot Bid Advisor



Instance Type	vCPU	Memory GiB	Savings over On-Demand*	Frequency of being outbid (month)	Frequency of being outbid (week)
m4.10xlarge	40	160	87%	Low	Low
cc2.8xlarge	32	60.5	85%	Low	Low
m2.4xlarge	8	68.4	92%	Low	Low
cr1.8xlarge	32	244	91%	Low	Low
d2.2xlarge	8	61	89%	Low	Low
d2.4xlarge	16	122	89%	Low	Low
hi1.4xlarge	16	60.5	94%	Low	Low
m4.4xlarge	16	64	83%	Medium	Low
c4.8xlarge	36	60	81%	Medium	Low
c3.8xlarge	32	60	77%	Medium	Low

1) We make this easier using Spot Bid Advisor

2) With careful and automated pool selection and bidding, you can keep your Spot cluster running as long as you need to

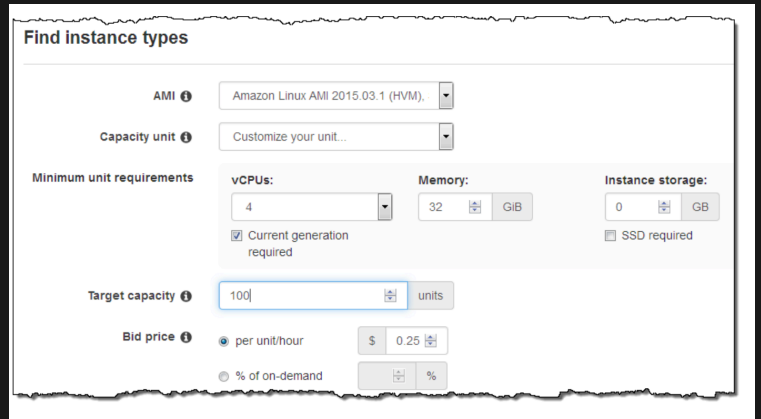
Advanced Spot usage

Spot Fleet

- “Give me 400 cores- choose the best Availability Zone and instance types for me”
- You can select and weigh instance types
- AWS chooses best and cheapest compute fleet for you

Spot Block

* Your spot instances are guaranteed for up to 6 hours
* Slightly lower discount



amazon web services

Price Example: Spot vs. On-Demand (YMMV)

4 compute nodes, 2 hours

- On-Demand, us-east
  - $19.13
- Spot (us-west-1)
  - $7.22
- Almost 1/3rd the cost!

16 compute nodes, 32 hours

- On-Demand, us-east
  - $1,018.77
- Spot (us-west-1)
  - $223.11
- Almost 1/5th the cost!

High-Performance Interconnects in the AWS cloud

- AWS provides 10Gbps and 20Gbps bi-directional, non-oversubscribed Ethernet between EC2 compute instances.
- Proprietary hardware and optimized kernels make the most of the network fabric.
- HPC performance is strong for a large share of real-world HPC use cases. (Synthetic benchmarks and micro-benchmarks can be misleading and irrelevant.)
- “Best practices” help achieve best possible performance.