

Research Computing, HPC & Deep Learning on AWS

Adrian White, Research & Technical Computing
Amazon Web Services

November 2017

Agenda

10:00	Research on AWS introduction	(30 mins)
10:30	Machine Learning & AI on AWS	(30 mins)
	<i>Demo: Deep Learning AMI</i>	
11:00	Lab (Part 1): Accessing AWS	(1 hours)
12:00	Lunch break	(1 hour)
13:00	Ronin – Unleash your Research	(1 hour)
14:00	Lab (Part 2): Jupyter & MXNET on AWS	(1 hours)
15:00	Wrap-up & Next Steps	(15 mins)

AWS for Research & Technical Computing



Time to Science

Access research infrastructure in minutes



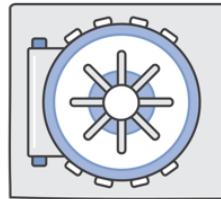
Globally Accessible

Easily collaborate with researchers around the world



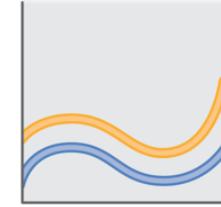
Low Cost

Pay-as-you-go pricing



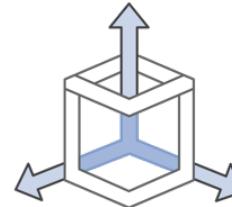
Secure

A collection of tools to protect data and privacy



Elastic

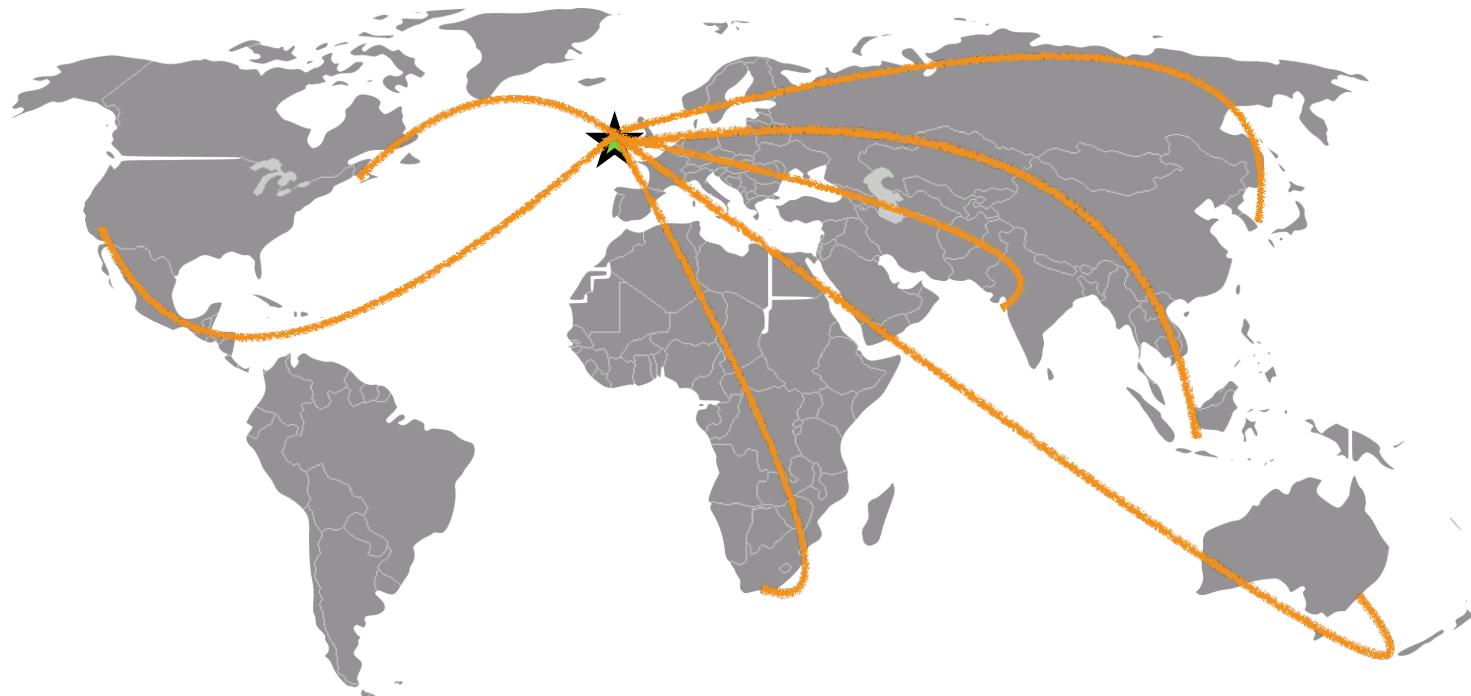
Easily add or remove capacity



Scalable

Access to effectively limitless capacity

Collaboration is easier in the cloud

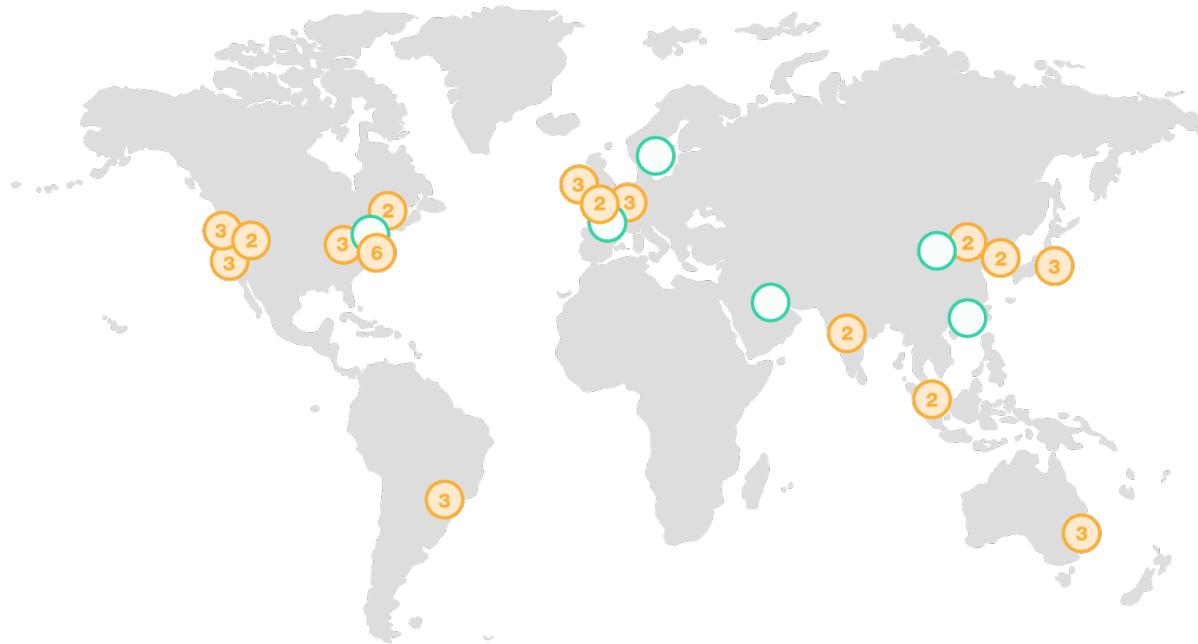


More time spent computing the data than moving the data.

Global AWS Regions

Current: 44 Availability Zones within 16 geographic Regions

Announced: 17 more Availability Zones and 6 new regions



AWS Region = A cluster of Availability Zones
Availability Zone = A cluster of data centers

All regions are sovereign, meaning your data never leaves that location unless you cause it to.

Americas

- AWS GovCloud (2)
- **AWS GovCloud East**
- US West
- Oregon (3)
- Northern California (3)
- Northern Virginia (5)
- Ohio (3)
- Montreal (2)
- São Paulo (3)

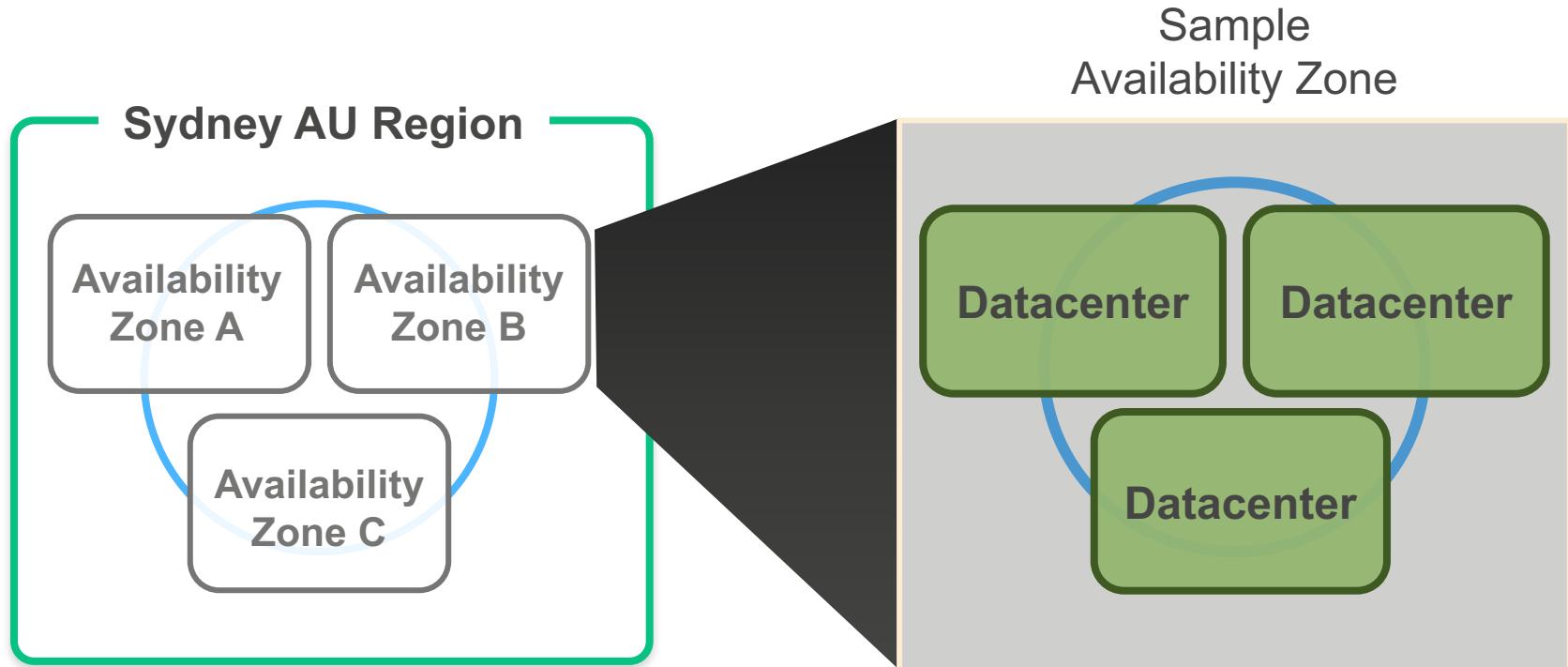
Europe / Middle East

- Ireland (3)
- Frankfurt (3)
- London (2)
- **Paris**
- **Stockholm**
- **Bahrain**

Asia Pacific

- Singapore (2)
- Sydney (3)
- Tokyo (3)
- Seoul (2)
- Mumbai (2)
- Beijing (2)
- **Ningxia**
- **Hong Kong(3)**

Anatomy of a region



A tool for almost every job

A tool for almost every job

because you'll want tools you don't know about yet

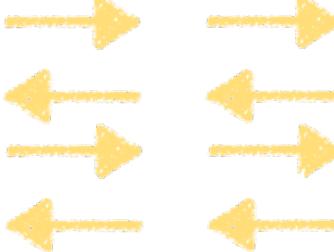
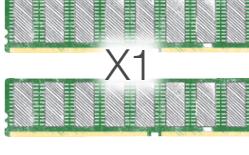
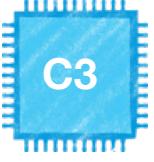
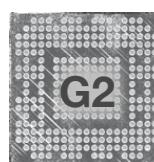
The AWS Platform

Account Support
Support
Managed Services
Professional Services
Partner Ecosystem
Training & Certification
Solution Architects
Account Management
Security & Pricing Reports
Technical Acct. Management

Marketplace	Mgmt. Tools	Analytics	Dev Tools	Artificial Intelligence	IoT	Mobile	Enterprise Applications	Game Development
Business Applications	Monitoring	Query Large Data Sets						
DevOps Tools	Auditing	Elasticsearch						
Business Intelligence	Service Catalog	Business Analytics						
Security	Server Management	Hadoop/Spark			Rules Engine	Build, Test, Monitor Apps	Document Sharing	
Networking	Configuration Tracking	Real-time Data Streaming	Private Git Repositories	Voice & Text Chatbots	Local Compute and Sync	Push Notifications	Email & Calendaring	
Database & Storage	Optimization	Orchestration Workflows	Continuous Delivery	Machine Learning	Device Shadows	Build, Deploy, Manage APIs	Hosted Desktops	
SaaS Subscriptions	Resource Templates	Managed Search	Build, Test, and Debug	Text-to-Speech	Device Gateway	Device Testing	Application Streaming	3D Game Engine
Operating Systems	Automation	Managed ETL	Deployment	Image Analysis	Registry	Identity	Backup	Multi-player Backends
Migration	Application Discovery	Application Migration	Data Migration	Database Migration	Server Migration			
Hybrid	Data Integration	Integrated Networking	Identity Federation	Resource Management	VMware on AWS	Devices & Edge Systems		
Application Services	Transcoding	Step Functions	Messaging					
Security	Identity & Access	Key Storage & Management	Active Directory	DDoS Protection	Application Analysis	Certificate Management	Web App. Firewall	
Database	Aurora	MySQL	PostgreSQL	Oracle	SQL Server	MariaDB	Data Warehousing	NoSQL
Storage	Object Storage	Archive	Exabyte-scale Data Transport	Block Storage	Managed File Storage			
Compute	Virtual Machines	Simple Servers	Web Applications	Auto Scaling	Batch	Containers	Event-driven Computing	
Networking	Isolated Resources	Dedicated Connections	Global CDN	Load Balancing	Scalable DNS			
Infrastructure	Regions	Availability Zones	Points of Presence					

AWS Instance Types

Broad Set of Compute Instance Types for HPC and Deep Learning

General purpose	Compute optimized	Storage and I/O optimized	Memory optimized	GPU or FPGA enabled
 T2	 M4			
 M3	 C4		 X1	 P2
	 C3		 R4	 G2
			 R3	

EC2

There's a couple dozen EC2 compute instance types alone, each of which is optimized for different things.

One size does not fit all.

Memory Optimized

R3

R3 instances are optimized for memory-intensive applications and have the lowest cost per GiB of RAM among Amazon EC2 instance types.

Features:

- High Frequency Intel Xeon E5-2670 v2 (Ivy Bridge) Processors
- Lowest price point per GiB of RAM
- SSD Storage
- Support for [Enhanced Networking](#)

Model	vCPU	Mem (GiB)	SSD Storage (GB)
r3.large	2	15.25	1 x 32
r3.xlarge	4	30.5	1 x 80
r3.2xlarge	8	61	1 x 160
r3.4xlarge	16	122	1 x 320
r3.8xlarge	32	244	2 x 320

Use Cases

We recommend memory-optimized instances for high performance databases, distributed memory caches, in-memory analytics, genome assembly and analysis, larger deployments of SAP, Microsoft SharePoint, and other enterprise applications.

C4

C4 instances are the latest generation of Compute-optimized instances, featuring the highest performing processors and the lowest price/compute performance in EC2.

Model	vCPU	Mem (GiB)	Storage	Dedicated EBS Throughput (Mbps)
c4.large	2	3.75	EBS-Only	500
c4.xlarge	4	7.5	EBS-Only	750
c4.2xlarge	8	15	EBS-Only	1,000
c4.4xlarge	16	30	EBS-Only	2,000
c4.8xlarge	36	60	EBS-Only	4,000

GPU

G2

This family includes G2 instances intended for graphics and general purpose GPU compute applications.

Features:

- High Frequency Intel Xeon E5-2670 (Sandy Bridge) Processors
- High-performance NVIDIA GPU with 1,536 CUDA cores and 4GB of video memory
- On-board hardware video encoder designed to support up to eight real-time HD video streams (720p at 30fps) or up to four real-time FHD video streams (1080p at 30 ps).
- Support for low-latency frame capture and encoding for either the full operating system or select render targets, enabling high-quality interactive streaming experiences.

Model	vCPU	Mem (GiB)	SSD Storage (GB)
g2.2xlarge	8	15	1 x 60

Use Cases

Game streaming, video encoding, 3D application streaming, and other server-side graphics workloads.

C3

Features:

- High Frequency Intel Xeon E5-2680 v2 (Ivy Bridge) Processors
- Support for [Enhanced Networking](#)
- Support for clustering
- SSD-backed instance storage

Model	vCPU	Mem (GiB)	SSD Storage (GB)
c3.large	2	3.75	2 x 16
c3.xlarge	4	7.5	2 x 40
c3.2xlarge	8	15	2 x 80
c3.4xlarge	16	30	2 x 160
c3.8xlarge	32	60	2 x 320

M3

This family includes the M3 instance types and provides a balance of compute, memory, and network resources, and it is a good choice for many applications.

Features:

- High Frequency Intel Xeon E5-2670 v2 (Ivy Bridge) Processors*
- SSD-based instance storage for fast I/O performance
- Balance of compute, memory, and network resources

Model	vCPU	Mem (GiB)	SSD Storage (GB)
m3.medium	1	3.75	1 x 4
m3.large	2	7.5	1 x 32
m3.xlarge	4	15	2 x 40
m3.2xlarge	8	30	2 x 80

Use Cases

Small and mid-size databases, data processing tasks that require additional memory, caching fleets, and for running backend servers for SAP, Microsoft SharePoint, and other enterprise applications.

<http://aws.amazon.com/ec2/instance-types/>

C4

Intel Xeon E5-2666 v3, custom built for AWS.

Intel Haswell, 16 FLOPS/tick

2.9 GHz, turbo to 3.5 GHz

Feature	Specification
Processor Number	E5-2666 v3
Intel® Smart Cache	25 MiB
Instruction Set	64-bit
Instruction Set Extensions	AVX 2.0
Lithography	22 nm
Processor Base Frequency	2.9 GHz
Max All Core Turbo Frequency	3.2 GHz
Max Turbo Frequency	3.5 GHz (available on c4.2xLarge)
Intel® Turbo Boost Technology	2.0
Intel® vPro Technology	Yes
Intel® Hyper-Threading Technology	Yes
Intel® Virtualization Technology (VT-x)	Yes
Intel® Virtualization Technology for Directed I/O (VT-d)	Yes
Intel® VT-x with Extended Page Tables (EPT)	Yes
Intel® 64	Yes

AWS Official Blog

New Compute-Optimized EC2 Instances

by Jeff Barr | on 13 NOV 2014 | in [Amazon EC2](#) | [Permalink](#)

Our customers continue to increase the sophistication and intensity of the compute-bound workloads that they run on the [Cloud](#). Applications such as top-end website hosting, online gaming, simulation, risk analysis, and rendering are voracious consumers of CPU cycles and can almost always benefit from the parallelism offered by today's multicore processors.

The New C4 Instance Type

Today we are pre-announcing the latest generation of compute-optimized [Amazon Elastic Compute Cloud \(EC2\)](#) instances. The new C4 instances are based on the Intel Xeon E5-2666 v3 (code name Haswell) processor. This custom processor, designed specifically for EC2, runs at a base speed of 2.9 GHz, and can achieve clock speeds as high as 3.5 GHz with Turbo boost. These instances are designed to deliver the highest level of processor performance on EC2. If you've got the workload, we've got the instance!

Here's the lineup (these specs are preliminary and could change a bit before launch time):

Instance Name	vCPU Count	RAM	Network Performance
c4.large	2	3.75 GiB	Moderate
c4.xlarge	4	7.5 GiB	Moderate
c4.2xlarge	8	15 GiB	High
c4.4xlarge	16	30 GiB	High
c4.8xlarge	36	60 GiB	10 Gbps

These instances are a great match for the [SSD-Backed Elastic Block Storage](#) that we introduced earlier this year. [EBS Optimization](#) is enabled by default for all C4 instance sizes, and is available to you at no extra charge. C4 instances also allow you to achieve significantly higher packet per second (PPS) performance, lower network jitter, and lower network latency using [Enhanced Networking](#).

<http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/c4-instances.html>

M4.16XL

Intel Xeon 2686v4, custom built for AWS.

Intel Broadwell, 16 FLOPS/tick

2.2 GHz, turbo to 3.60 GHz

Performance	
# of Cores	22
# of Threads	44
Processor Base Frequency	2.20 GHz
Max Turbo Frequency	3.60 GHz
Cache	55 MB SmartCache
Bus Speed	9.6 GT/s QPI
# of QPI Links	2
TDP	145 W
VID Voltage Range	0

Expanding the M4 Instance Type – New M4.16xlarge

by Jeff Barr | on 27 SEP 2016 | in [Amazon EC2](#), [Launch](#) | [Permalink](#) | [Comments](#)

EC2's M4 instances offer a balance of compute, memory, and networking resources and are a good choice for many different types of applications.

We launched the M4 instances last year (read [The New M4 Instance Type](#) to learn more) and gave you a choice of five sizes, from **large** up to **10xlarge**. Today we are expanding the range with the introduction of a new **m4.16xlarge** with 64 vCPUs and 256 GiB of RAM. Here's the complete set of specs:

Instance Name	vCPU Count	RAM	Instance Storage	Network Performance	EBS-Optimized
m4.large	2	8 GiB	EBS Only	Moderate	450 Mbps
m4.xlarge	4	16 GiB	EBS Only	High	750 Mbps
m4.2xlarge	8	32 GiB	EBS Only	High	1,000 Mbps
m4.4xlarge	16	64 GiB	EBS Only	High	2,000 Mbps
m4.10xlarge	40	160 GiB	EBS Only	10 Gbps	4,000 Mbps
m4.16xlarge	64	256 GiB	EBS Only	20 Gbps	10,000 Mbps

The new instances are based on Intel Xeon E5-2686 v4 ([Broadwell](#)) processors that are optimized specifically for EC2. When used with Elastic Network Adapter (ENA) inside of a placement group, the instances can deliver up to 20 Gbps of low-latency network bandwidth. To learn more about the ENA, read my post, [Elastic Network Adapter – High Performance Network Interface for Amazon EC2](#).

C5

Intel Xeon Skylake.

Supports AVX-512
NVMe support
ENA for 25 Gbps network

**Up to 72 vCPUs and 144 GiB memory
25 Gbps network interconnect**

Ideal for:

- Batch processing
- HPC
- ML / DL inference
- Video encoding

Introducing Amazon EC2 C5 Instances, the next generation of Compute Optimized instances

Posted On: Nov 6, 2017

Amazon EC2 C5 instances are the next generation of Amazon EC2's compute optimized instances and are powered by 3.0 GHz Intel® Xeon® Scalable processors (Skylake). C5 instances are built using a new light-weight hypervisor, which provides practically all of the compute and memory resources to customers' instances. As a result, C5 instances deliver the best price/compute performance in the EC2 product family, with a 25% improvement in price/performance compared to C4 instances and certain applications delivering greater than 50% improvement. C5 instances are ideal for running compute-heavy workloads like batch processing, distributed analytics, high-performance computing (HPC), machine/deep learning inference, ad serving, highly scalable multiplayer gaming, and video encoding.

C5 instances are available in 6 sizes and introduce a new larger size, c5.18xlarge, which provides 72 vCPUs and 144 GiB of memory that will allow customers to effectively consolidate workloads. Next generation Elastic Network Adapter (ENA) and NVM Express (NVMe) technology provide C5 instances with high throughput, low latency interfaces for networking and Amazon Elastic Block Storage (Amazon EBS). C5

P2 (K80)

- Up to 16 x Nvidia K80 GPUs in a single instance
- Including peer-to-peer PCIe GPU interconnect
- Supporting a wide variety of use cases including deep learning, HPC simulations, and batch rendering

Instance Size	GPUs	GPU Peer to Peer	vCPUs	Memory (GiB)	Network Bandwidth*
p2.xlarge	1	-	4	61	1.25 Gbps
p2.8xlarge	8	Y	32	488	10 Gbps
p2.16xlarge	16	Y	64	732	25 Gbps

P3 (Volta)

- Up to 8 x Nvidia Volta GPUs in a single instance
- One petaflop of mixed-precision, 125 teraflops of single-precision, and 62 teraflops of double-precision floating point performance
- 300GB/s NVIDIA NVLink
- Well-suited for distributed deep learning frameworks, such as MXNet

Instance Size	GPUs	GPU Peer to Peer	vCPUs	Memory (GiB)	Network Bandwidth*
p3.2large	1	-	8	61	Up to 10 Gbps
p3.8xlarge	4	NVLink	32	244	10 Gbps
p3.16xlarge	8	NVLink	64	488	25 Gbps

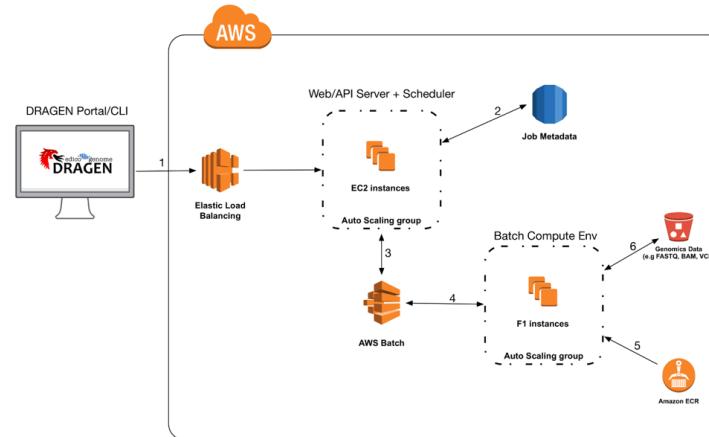
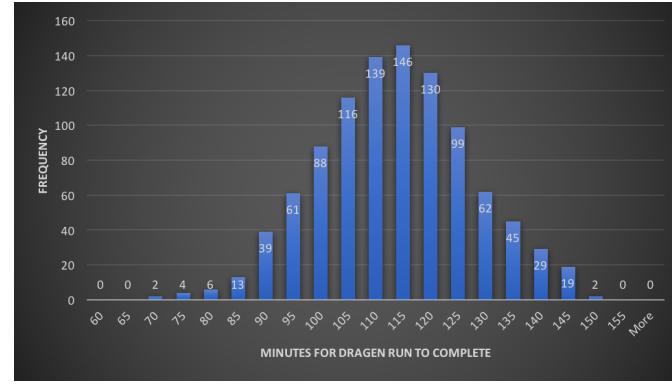
F1 (FPGAs)

- Up to 8 Xilinx Virtex UltraScale Plus VU9p FPGAs in a single instance with four high-speed DDR-4 per FPGA
- Largest size includes high performance FPGA interconnects via PCIe Gen3 (FPGA Direct), and bidirectional ring (FPGA Link)
- Designed for hardware-accelerated applications including financial computing, genomics, accelerated search, and image processing

Instance Size	FPGAs	FPGA Link	FPGA Direct	vCPUs	Memory (GiB)	NVMe Instance Storage	Network Bandwidth*
f1.2xlarge	1	-		8	122	1 x 480	5 Gbps
f1.16xlarge	8	Y	Y	64	976	4 x 960	25 Gbps

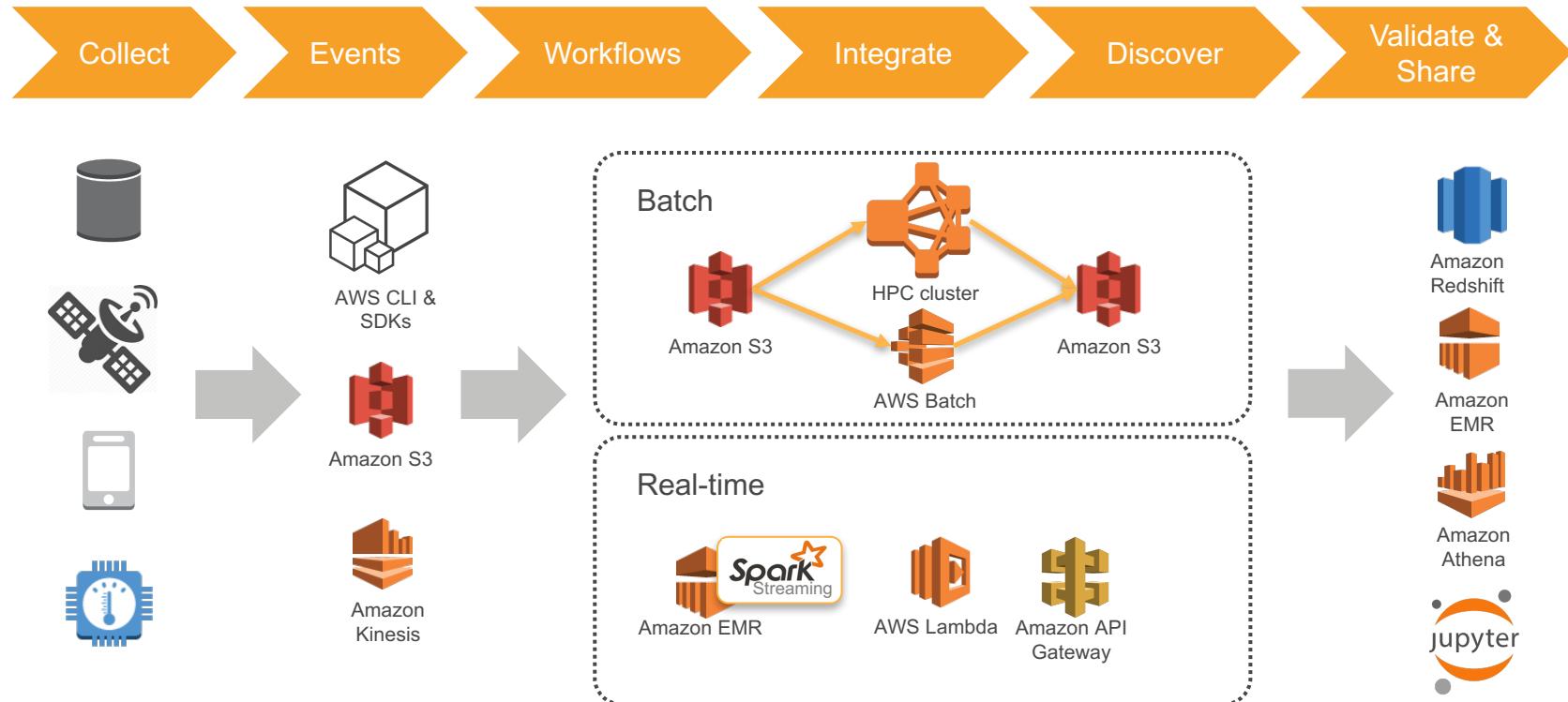
Accelerating Precision Medicine at Scale

- Children's Hospital of Philadelphia (CHOP) analysed 1000 genomes in 2.5 hours using **Edico DRAGEN**
- Sourced data from the Center for Applied Genomics Biobank
- Averaged approximately \$3 per genome (AWS cost)
- Used 1000 x f1.2xlarge instances in a single AWS region
- Orchestrated with **AWS Batch** with all DRAGEN binaries in Docker

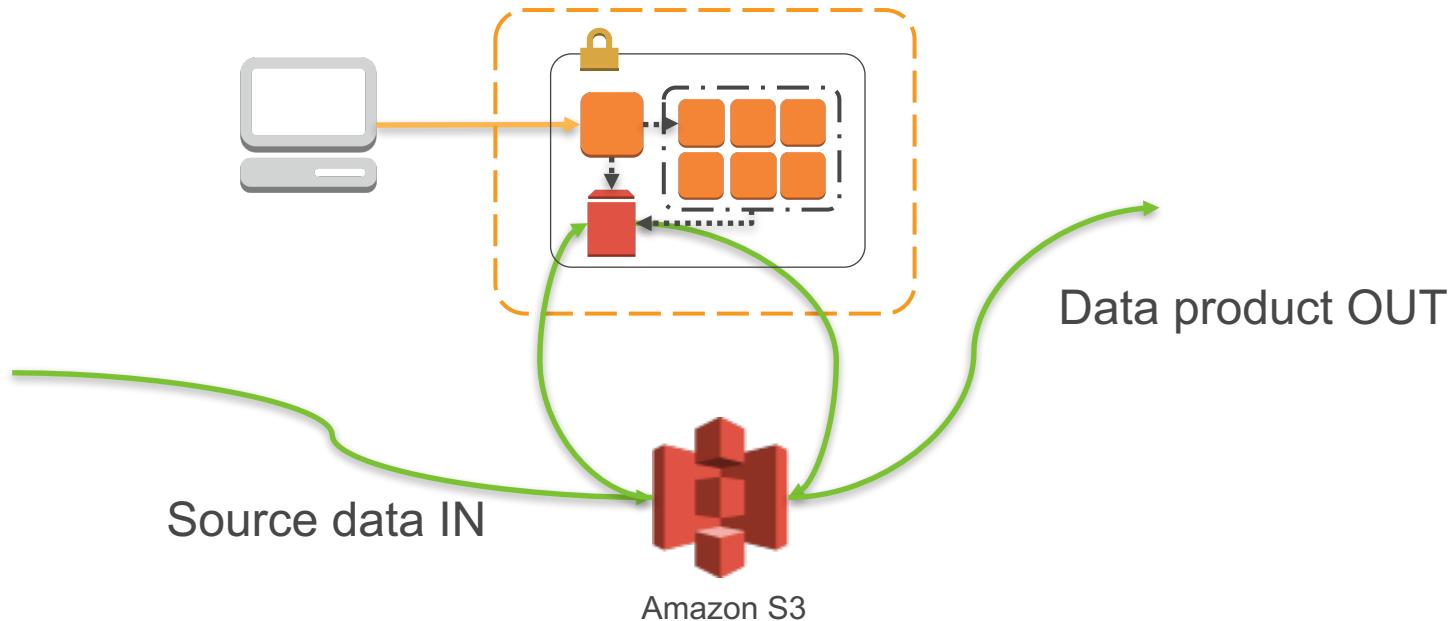


Research as workflow

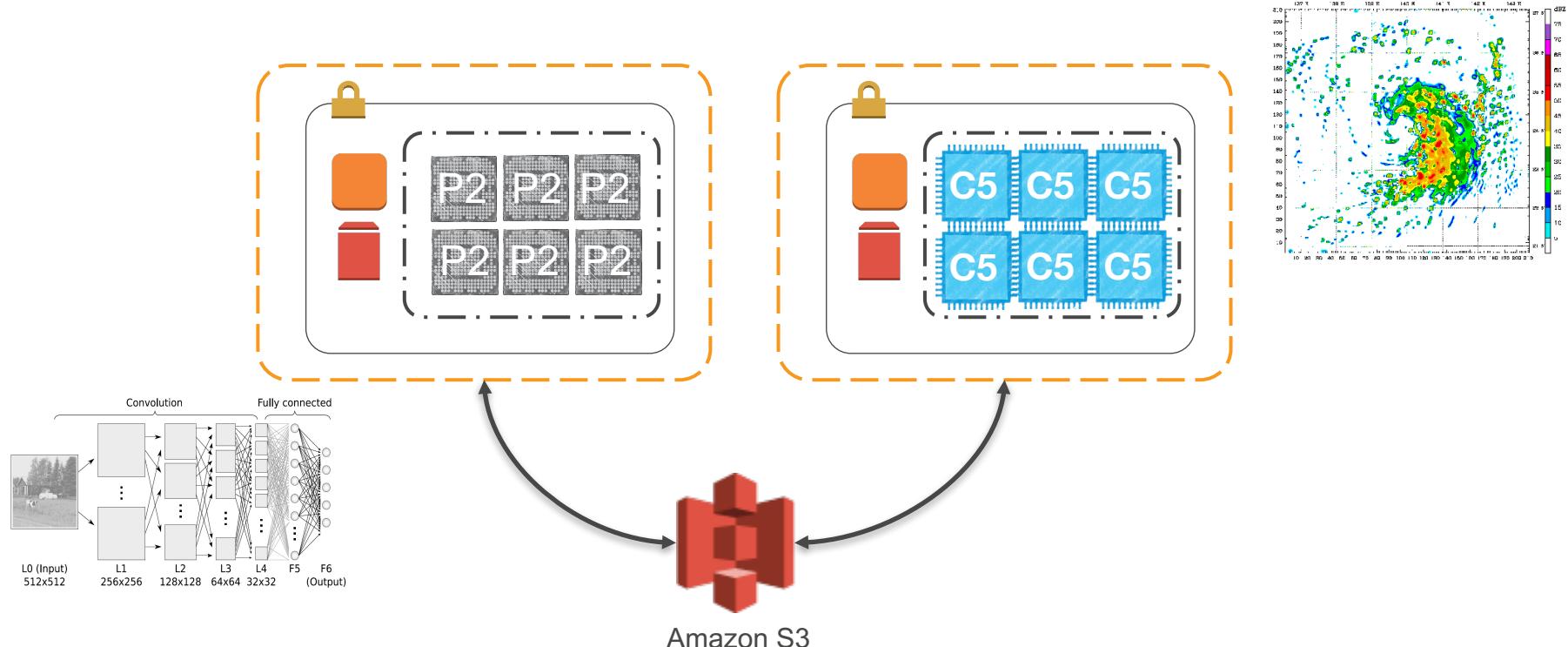
Model Capability based on Research Workflows



E.g. a cluster in the cloud is an ephemeral tool

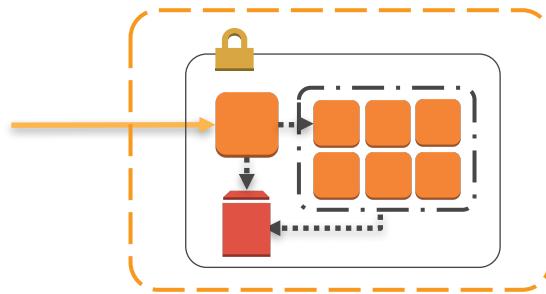


Clusters in the cloud are fit for purpose

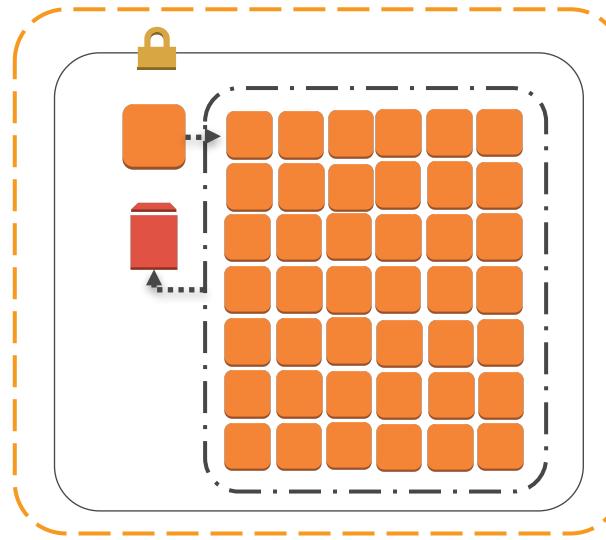


Clusters can scale and are elastic

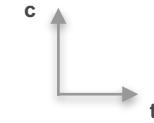
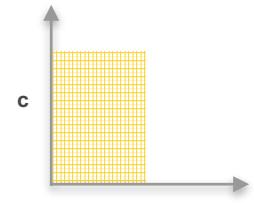
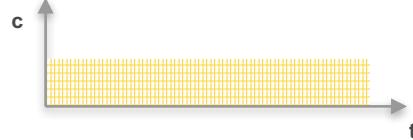
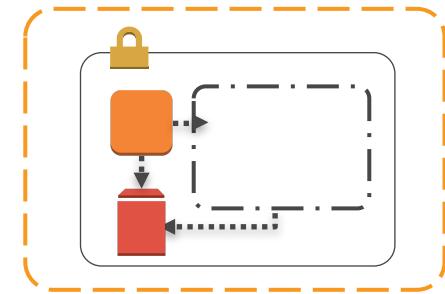
$W = 1, C = 1$



$W = n, C = n$



$W = 0, C \sim 0$



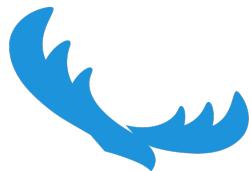
Ready-made HPC+HTC Tools

HPC and HTC tools on AWS



CfnCluster

CfnCluster is provided by AWS to quickly provision configurable HPC and HTC cluster environments



alcesflight

Alces Flight is available in the AWS Marketplace and bundles 1000+ commonly used scientific applications
<https://aws.amazon.com/marketplace/>



AWS Batch

AWS Batch provides compute resources via Docker containers with user-definable queues and an optimised job scheduler



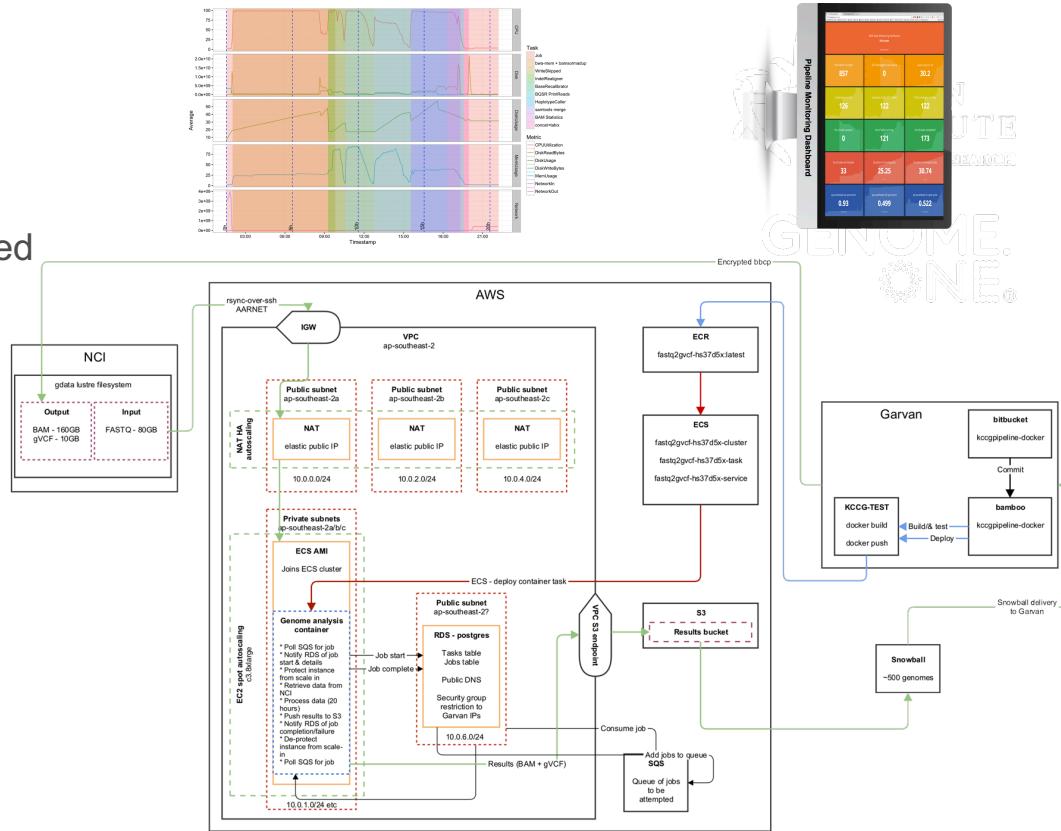
Amazon EMR

Amazon EMR provides a managed Hadoop framework supporting Apache Spark, HBase, Presto, and Flink, Spark MLLib on Amazon EC2 and EC2 Spot

**A familiar HPC architecture:
moved to AWS**

Garvan Institute – Containerized Bioinformatics

- The Garvan sequences **18,000** human genomes per year
- Have processed 4,500 genomes using 150 c3.8xlarge **spot instances**
- Used 3 million CPU-hours and processed 1.1 PB of data
- Use Docker with AWS ECS and ECR
- Realized a **20x cost saving** on AWS
- Have used AWS to containerize their genome analysis and run at scale - leveraging ECS, SQS and RDS

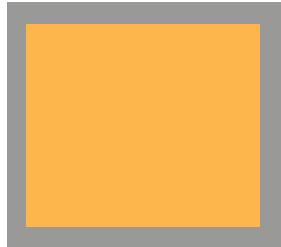


**But it's not (just) about
servers...**

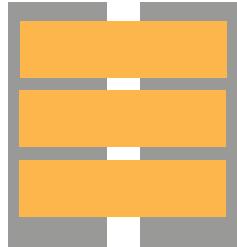
Evolving Compute Abstractions



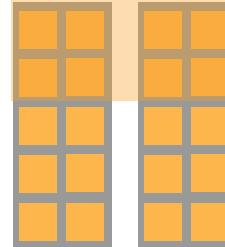
Physical



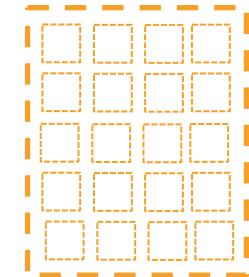
Virtualisation



Containerization



Serverless



AWS Lambda – How it Works



Bring your own code

Node.JS, Java, Python

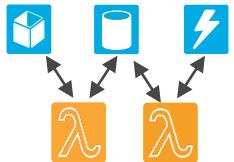
Java = Any JVM based language such as Scala, Clojure, etc.

Bring your own libraries



Simple resource model

- Select memory from 128MB to 1.5GB in 64MB steps
- CPU & Network allocated proportionately to RAM
- Reports actual usage



Flexible invocation paths

Event or RequestResponse invoke options

Existing integrations with various AWS services



Fine grained permissions

- Uses IAM role for Lambda execution permissions
- Uses Resource policy for AWS event sources

CSIRO – CRISPR search with AWS Lambda

GT-Scan2.0 is implemented as a microservices architecture using AWS Lambda

Serverless:

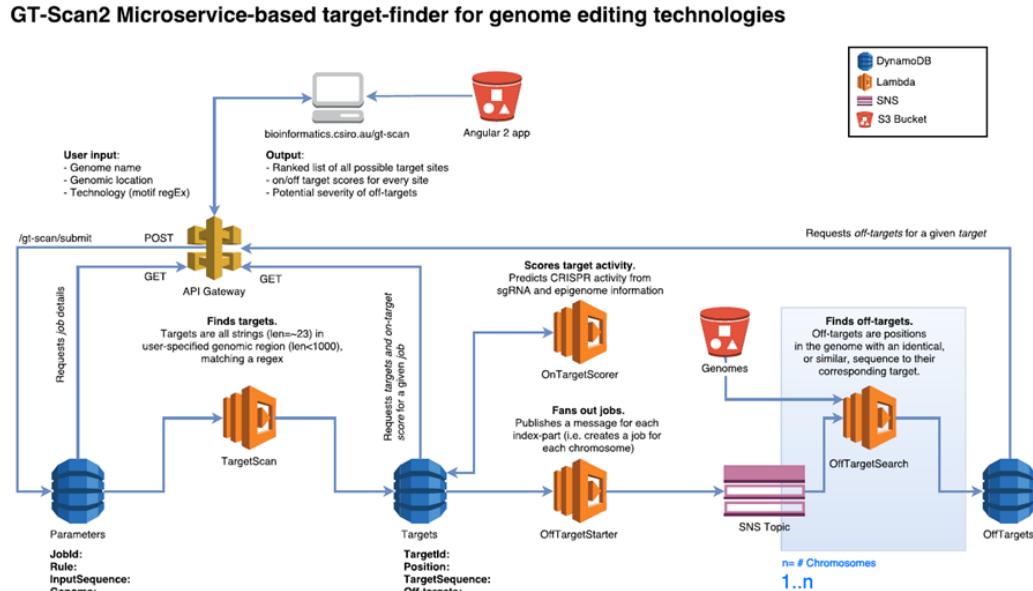
- Does not require users to have high-compute power

Scalable:

- Can be easily scaled to whole genome analysis

Also implement as a “stand-alone”

- Can be run on local servers
- Can incorporate your own ChIP-seq data rather than public data



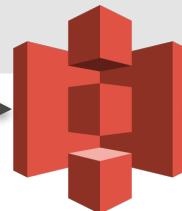
Data Analytics

S3 as a Data Lake

Collect



Share



Evolution of Data Analytics

Batch



Real time



Prediction



Amazon
EMR



Amazon
Redshift



AWS Batch



Amazon
Kinesis



Amazon Kinesis
Analytics



Amazon
SNS



AWS IoT

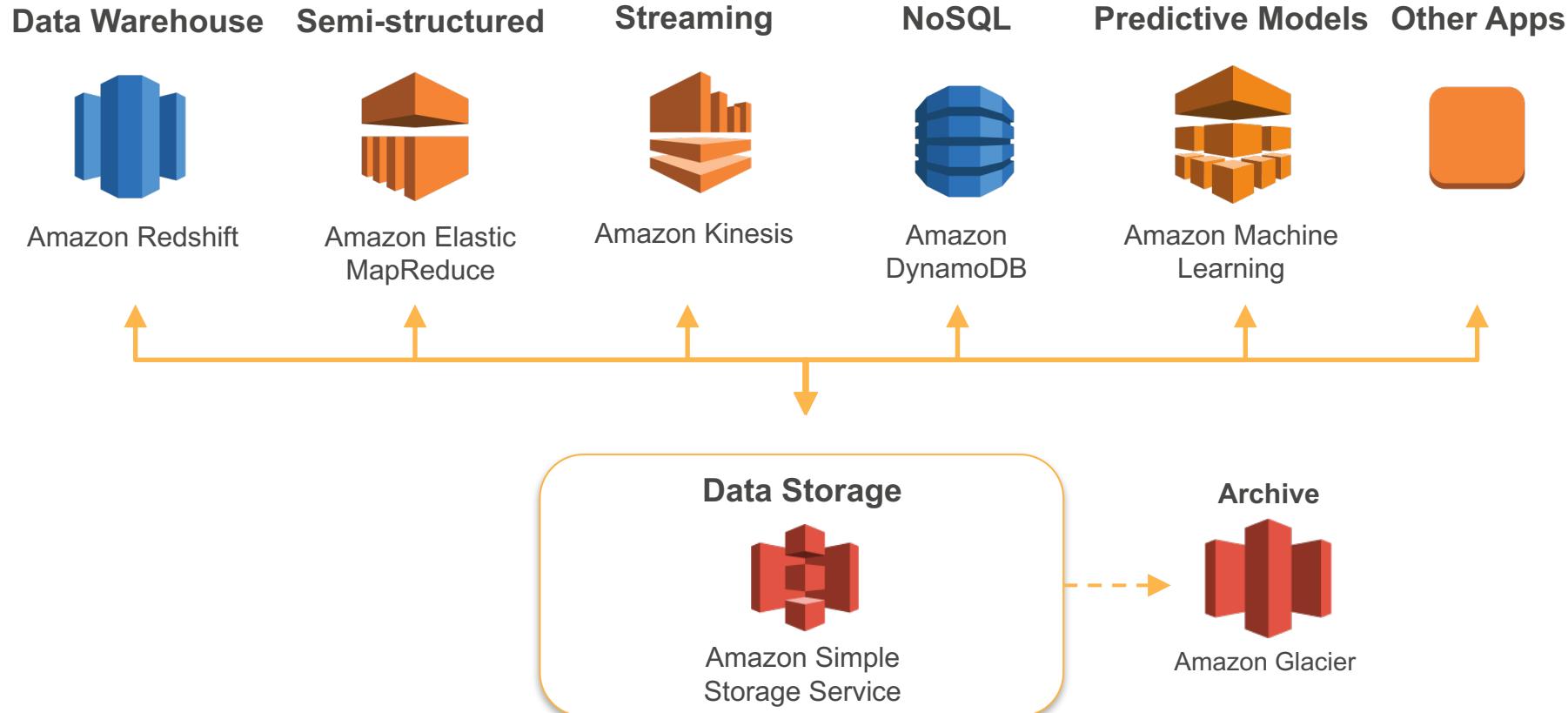


Amazon
Machine
Learning



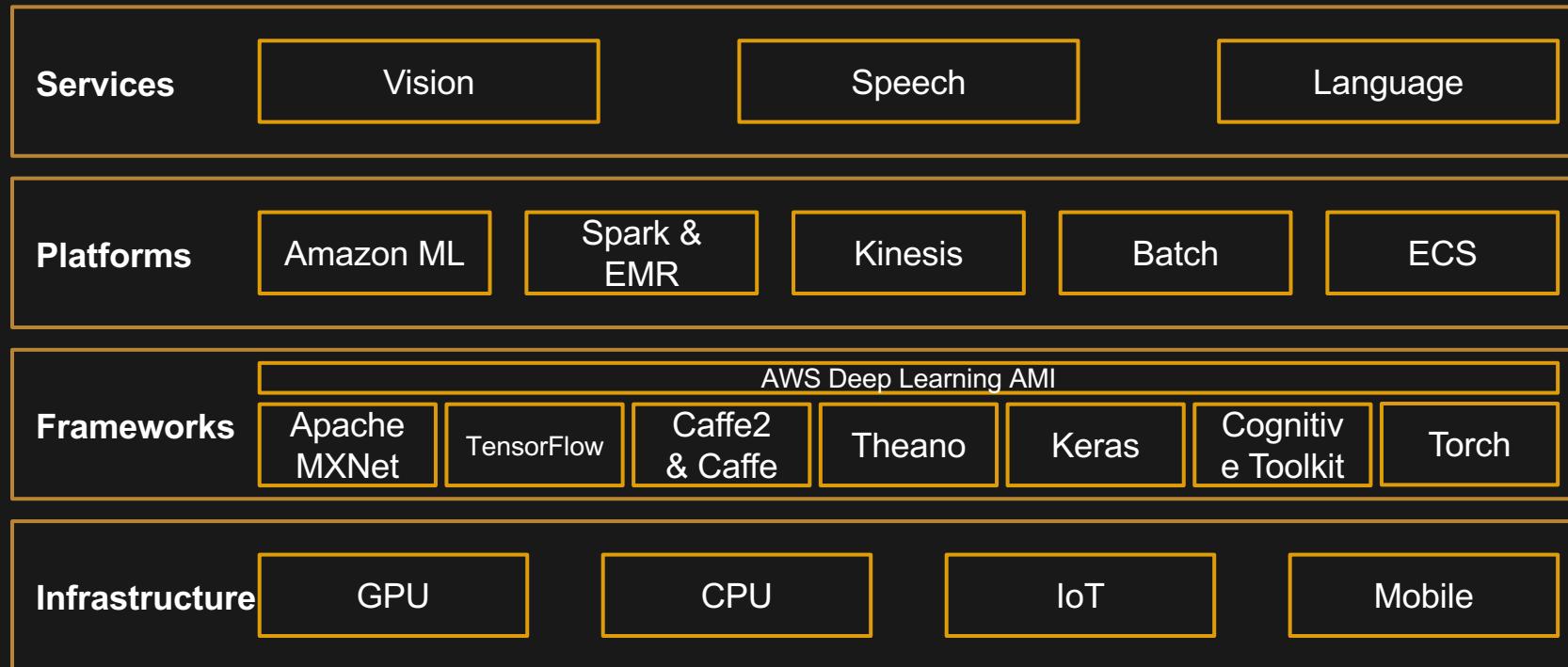
Amazon
Rekognition

Use an optimal combination of highly interoperable services

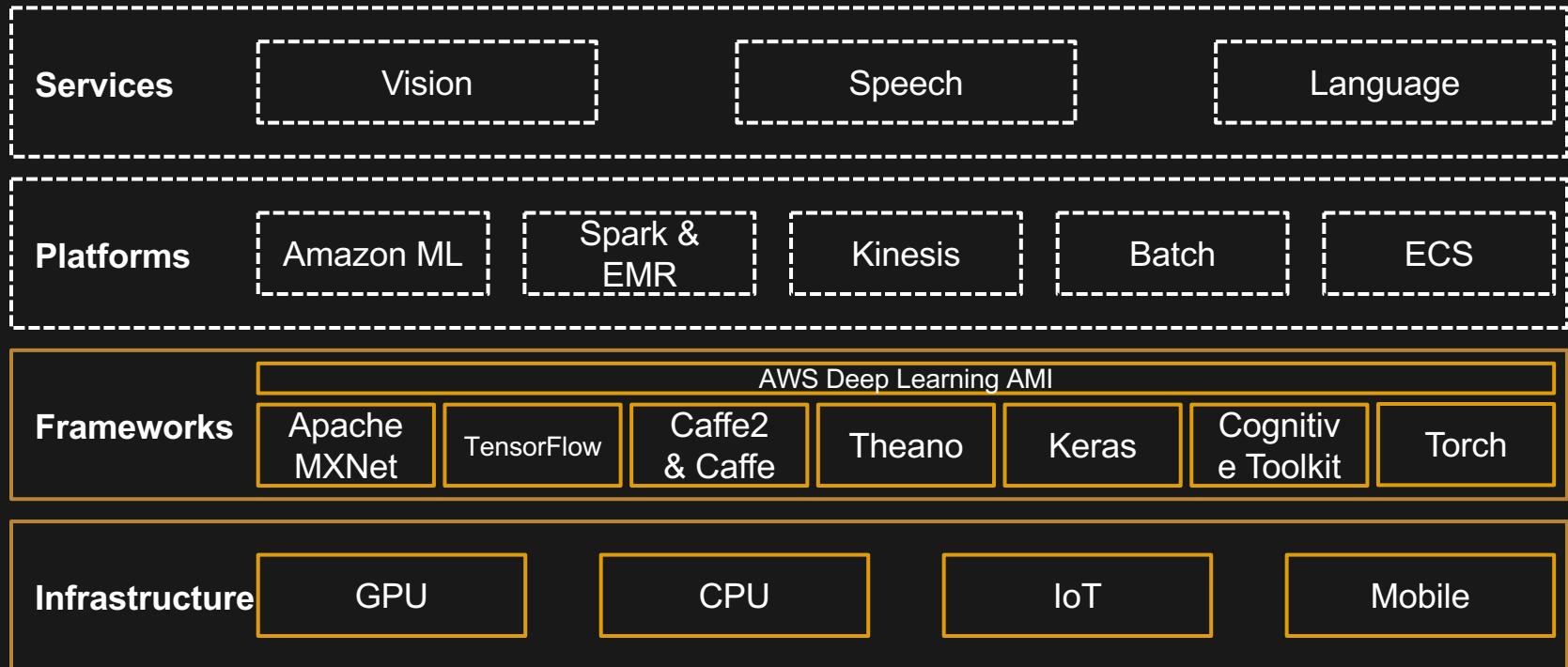


Machine Learning

The Amazon ML Stack



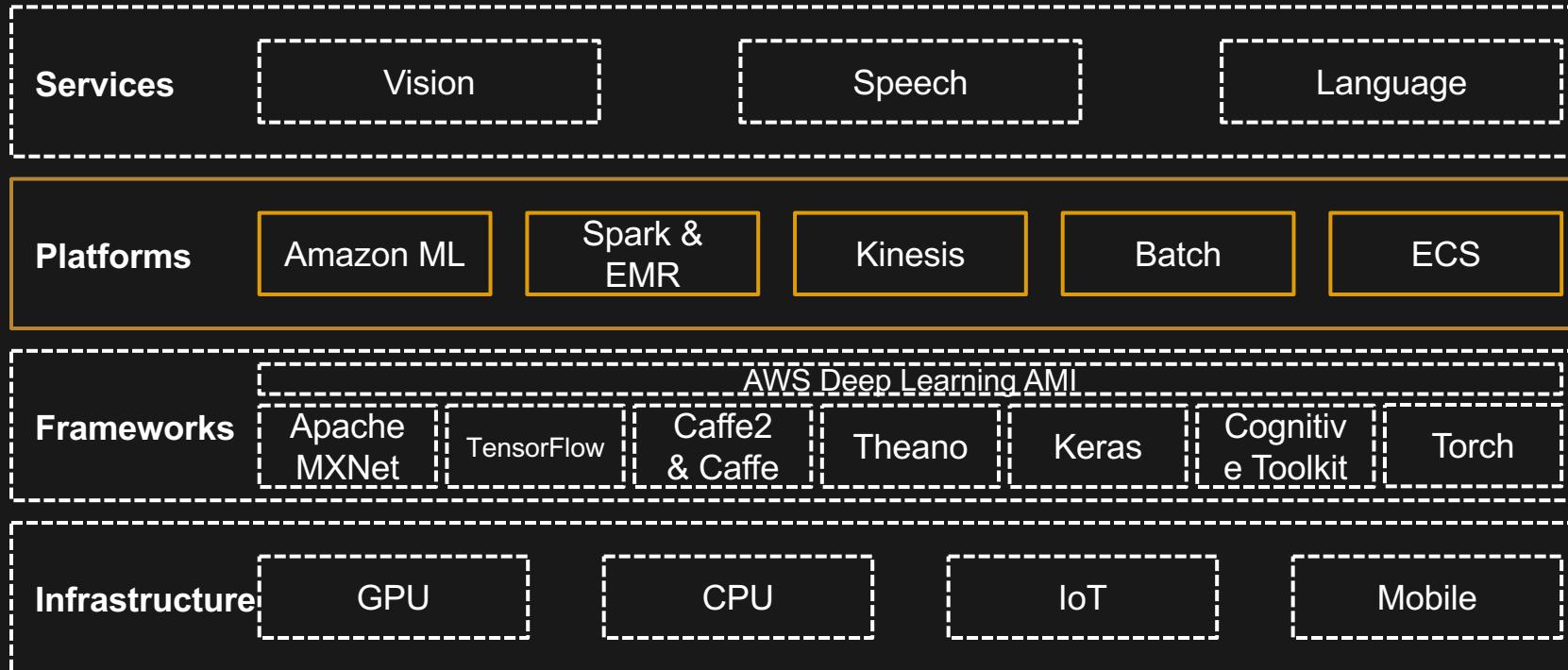
Frameworks & Infrastructure



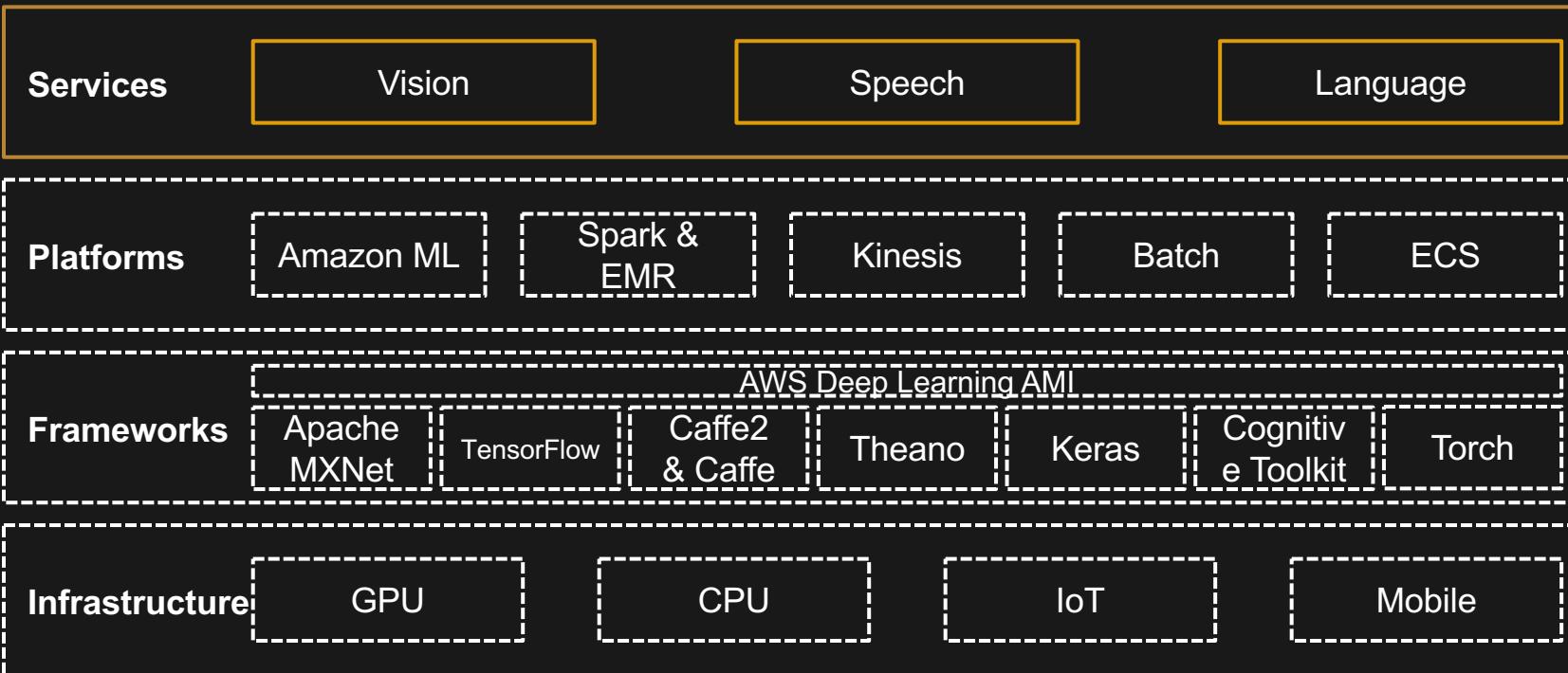
Support for all major frameworks



Machine Learning Platforms



API-driven Services



Amazon Rekognition



Potential Use Cases

Searchable Image Library

Face-based User Verification

Facial Recognition

Detect Inappropriate Content in Images

Sentiment Analysis

Celebrity Identification

C-SPAN

influicity

realtimes

SmugMug 

 **Kaliber**
ENABLING INFORMATION ASSURANCE

Openinfluence

 **Mobilink**

 **GO GIRL
apps**

THE TAKE


**MOTOROLA
SOLUTIONS**

Artfinder


bynder

SOCIAL SOUP

witlee



 **zmags**

 **Sturdy**

Amazon Polly (text-to-speech)



Potential Use Cases

Content Creation
Mobile & Desktop Applications
Internet of Things (IoT)

Education & E-learning
Customer Contact Center
Accessibility



GoAnimate

duolingo

amazon
RAPIDS



The Washington Post

Beeliked™
SOCIAL POLLINATION



iTranslate

inhealthcare

folio



Supporting people
with sight loss

aculabcloud
A true cloud telecoms platform

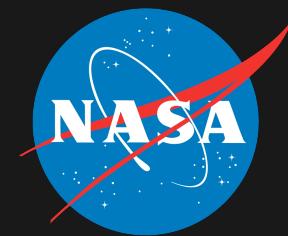
Amazon Lex (speech-to-text)



Potential Use Cases

Appointment Booking
Informational Services
Internet of Things (IoT)

Customer Support
Access Enterprise Data



Jupyter Notebooks on AWS

Research customers are increasingly doing exploratory data science and analytics work using notebooks.

Jupyter on AWS allows researchers to take advantage of any AWS compute node type:

- Large memory, CPU optimized, IO optimized
- GPU nodes (e.g. multiple K80 GPUs)

Researchers can also access Batch, HPC and Spark/Mllib clusters with Jupyter

How to:

[Run Jupyter Notebook and JupyterHub on Amazon EMR](#)

[Creating and Using a Jupyter Instance on AWS](#)



[**Creating and Using a Jupyter Instance on AWS**](#)

Authors:

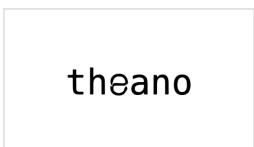
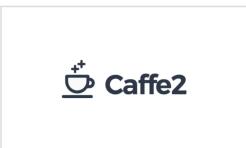
Jeff Layton, AWS Research and Technical Computing Team
Adrian White, AWS Research and Technical Computing Team

AWS Deep Learning AMI

The AWS Deep Learning AMI is for:

- data scientists
- machine learning practitioners
- research scientists

Supports multiple libraries and frameworks:



See also, <https://aws.amazon.com/blogs/ai/get-started-with-deep-learning-using-the-aws-deep-learning-ami/>

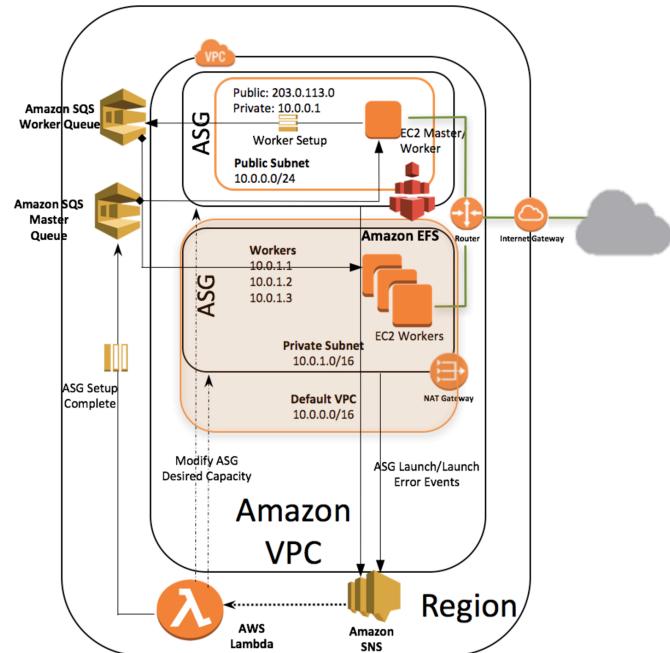
Distributed Deep Learning on AWS

- Distributed training across GPUs or CPUs using MXNet
- Spin up a cluster in minutes
- Automatically add or remove cluster nodes
- Supports Amazon EFS share filesystem
- Available on GitHub

<https://github.com/awslabs/deeplearning-cfn>

```
INFO:root:Epoch[18] Time cost=6.020
INFO:root:Epoch[18] Validation-accuracy=0.992237
INFO:root:Epoch[19] Batch [100] Speed: 9960.58 samples/sec Train-accuracy=1.000000
INFO:root:Epoch[19] Batch [200] Speed: 9830.53 samples/sec Train-accuracy=1.000000
INFO:root:Epoch[19] Batch [300] Speed: 10233.37 samples/sec Train-accuracy=1.000000
INFO:root:Epoch[19] Batch [400] Speed: 9747.18 samples/sec Train-accuracy=1.000000
INFO:root:Epoch[19] Batch [500] Speed: 9913.81 samples/sec Train-accuracy=1.000000
INFO:root:Epoch[19] Batch [600] Speed: 10109.95 samples/sec Train-accuracy=1.000000
INFO:root:Epoch[19] Batch [700] Speed: 9935.70 samples/sec Train-accuracy=1.000000
INFO:root:Epoch[19] Batch [800] Speed: 10210.48 samples/sec Train-accuracy=1.000000
INFO:root:Epoch[19] Batch [900] Speed: 9824.65 samples/sec Train-accuracy=1.000000
INFO:root:Epoch[19] Train-accuracy=1.000000
INFO:root:Epoch[19] Time cost=6.021
INFO:root:Epoch[19] Validation-accuracy=0.992237
[ec2-user@ip-10-0-0-224 image-classification]$
```

AWS Deep Learning Cluster



See also, <https://aws.amazon.com/blogs/compute/distributed-deep-learning-made-easy/>

Introducing Gluon

Gluon is a new machine learning library from AWS and Microsoft.

Gluon gives you:

- A friendly API (familiar for developers)
- Dynamic networks (construct networks at run time)
- The algorithm can adjust the network during definition and learning
- High performance operators for training



Gluon is available today in Apache MXNet, with support coming for the Microsoft Cognitive Toolkit in a future release.

Open source: <https://github.com/gluon-api/gluon-api/>

Open Data at AWS

Open Data at AWS

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

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.

Earth on AWS

Build planetary-scale applications
in the cloud with open geospatial

SpaceNet Machine Learning Imagery

SpaceNet is a corpus of high-resolution satellite imagery and labeled training data for researchers to utilize to develop and train machine learning algorithms. The data set is made up of roughly 1,990 square kilometers of imagery at 50 cm resolution and 220,594 corresponding building footprints.

Image from SpaceNet program

UK Met Office Weather Forecasts

Archive data from the UK Met Office Global and Regional Ensemble Prediction System (MOGREPS) is now available on Amazon S3. Both a high resolution weather forecast covering the United Kingdom and a global weather forecast are available.

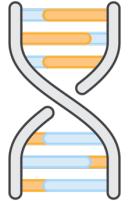
NASA Earth Exchange (NEX)

The NASA Earth Exchange (NEX) makes it easier and more efficient for researchers to access and process earth science data. NEX datasets available on Amazon S3 include downscaled climate projections, global MODIS vegetation indices, and Landsat Global Land Survey data.

Image from NASA

Research Programs at AWS

AWS Research Cloud Program



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



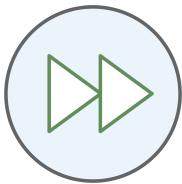
Simple and easily explained procedures to get set up with cloud access.



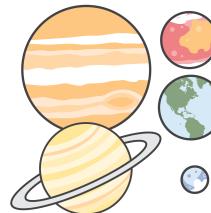
Budget management tools to ensure that over-spends do not happen.



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



Fast track to invoice-backed billing & Egress Waiver.



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

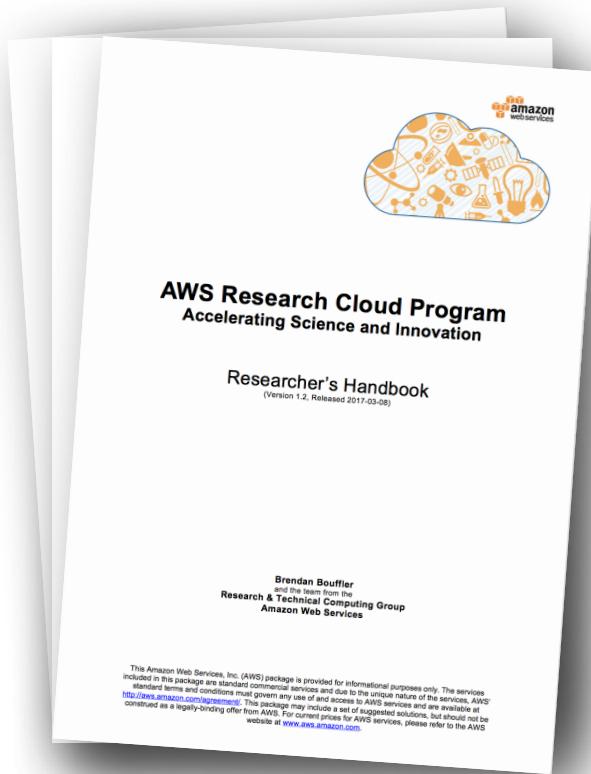
IT'S ABOUT SCIENCE, NOT SUSERS.

We recognize that whilst research is often a compute-intensive activity, **most researchers are not IT experts.**

We want to simplify research in the cloud with easy-to-use tools for researchers and their students, and share the catalogue of “**researcher-obsessed**” **products and services** created by many of our partners.

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.

Global Data Egress Waiver

Why?

Researchers
need predictable
budgets

Who?

Available to
Degree-granting
/ Research
Institutions in
APAC (and
elsewhere)

What?

Waives data
egress charges
from Qualified
Accounts
(capped at 15%
of total spend)

How?

Contract
Addendum
Required.
Talk to your
AWS account
team.

All qualifying research customers should use this!



Cloud Credits for Research

The Cloud Credits for Research program aims to support:

1. Building tools to facilitate future research
2. Performing proof of concept for research or open data workloads
3. Training the research community on the usage of the cloud

Also available to incubate "centers of excellence" in research on the AWS cloud.

More information, <https://aws.amazon.com/research-credits/>



Case Study: Detecting Meteors

Lab: Deep Learning on AWS with Jupyter and MXNet