

Accelerating Time to Science on AWS

IS&T Colloquium, NASA Goddard Space Flight Center, April 6th, 2016

Jamie Baker

Principal, Federal Scientific Accounts
bakjames@amazon.com





Amazon.com...



... AWS: the leader in utility (aka cloud) computing

2004

amazon.com

\$7B retail business
~10,000 employees
A whole lot of servers

2013



Every day, AWS adds enough
server capacity to power this
\$7B enterprise

Said another way...

AWS is deploying the equivalent of a top-20
supercomputer **every few days**

AWS is Leader and Visionary

Gartner Magic Quadrant for Cloud Infrastructure as a Service, Worldwide

Figure 1. Magic Quadrant for Cloud Infrastructure as a Service, Worldwide



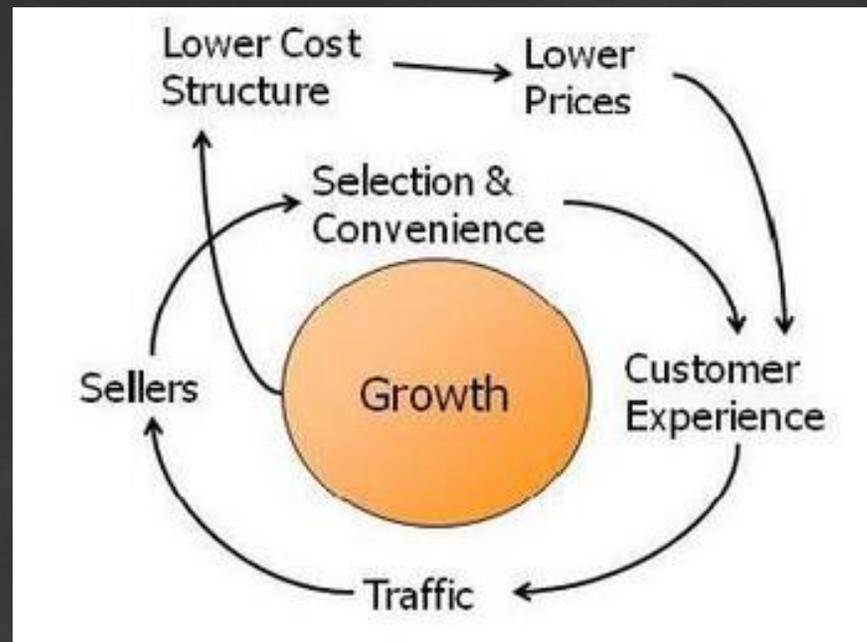
Source: Gartner (May 2015)

Source: Gartner (May 2015)

Gartner "Magic Quadrant for Cloud Infrastructure as a Service, Worldwide," Lydia Leong, Douglas Toombs, Bob Gill, May 18, 2015. This Magic Quadrant graphic was published by Gartner, Inc. as part of a larger research note and should be evaluated in the context of the entire report. The Gartner report is available at <http://aws.amazon.com/resources/analyst-reports/>. Gartner does not endorse any vendor, product or service depicted in its research publications, and does not advise technology users to select only those vendors with the highest ratings or other designation. Gartner research publications consist of the opinions of Gartner's research organization and should not be construed as statements of fact. Gartner disclaims all warranties, expressed or implied, with respect to this research, including any warranties of merchantability or fitness for a particular purpose.

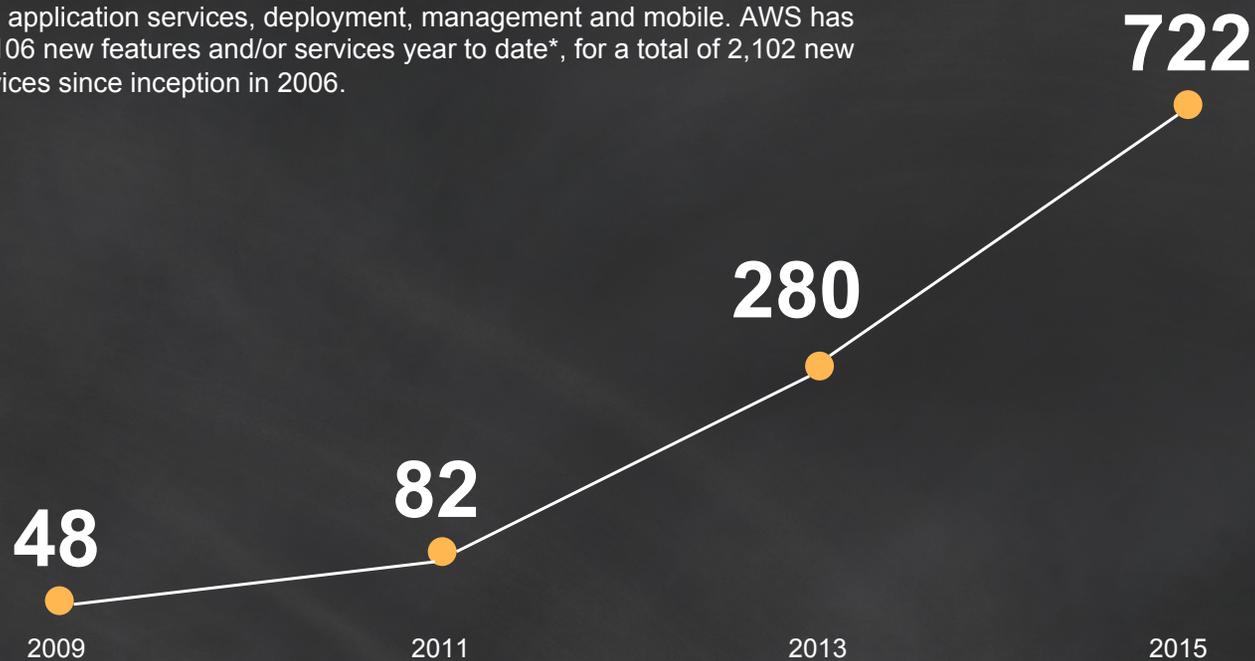


Amazon Virtuous Cycle

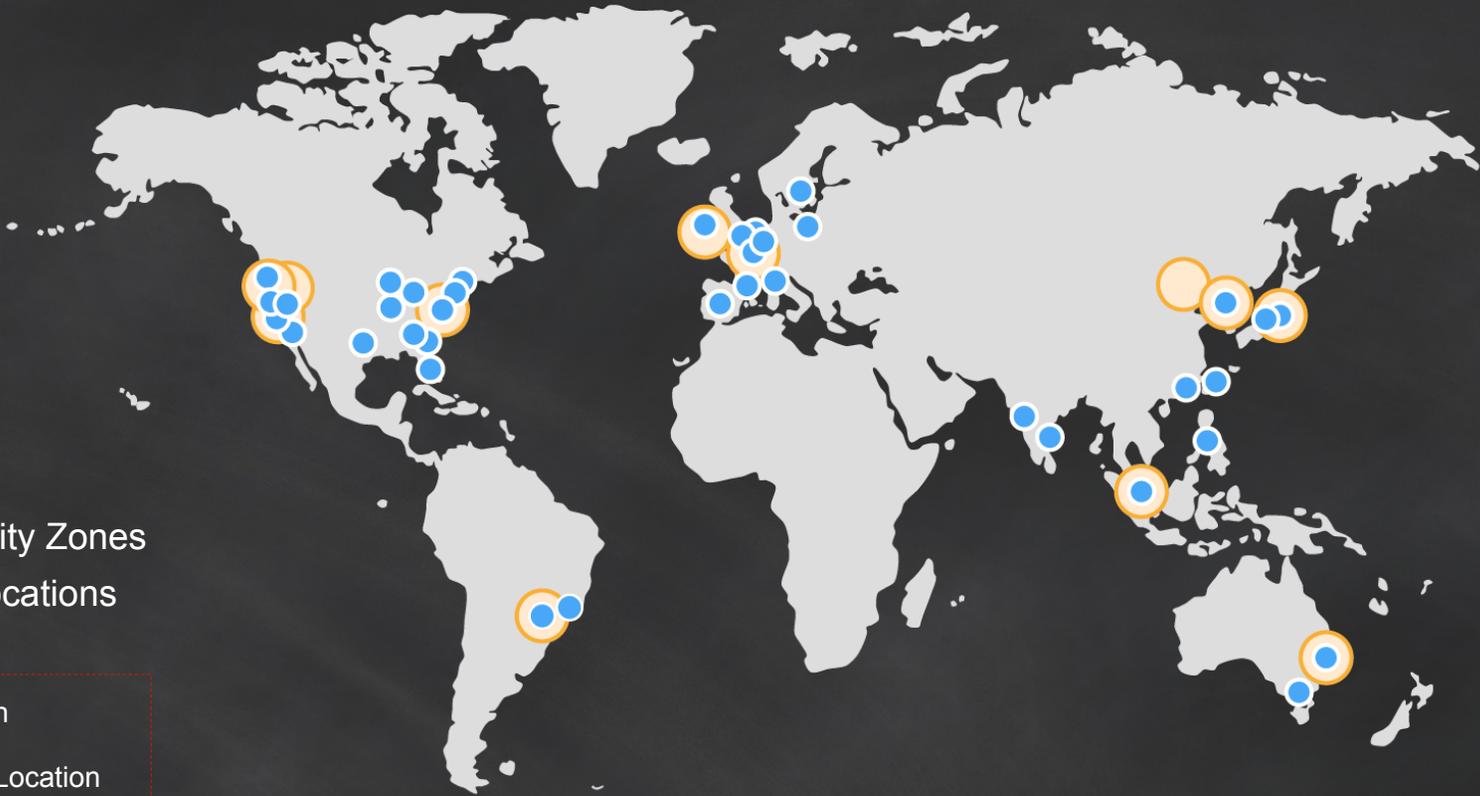


AWS Rapid Pace of Innovation

AWS has been continually expanding its' services to support virtually any cloud workload and now has more than 70 services that range from compute, storage, networking, database, analytics, application services, deployment, management and mobile. AWS has launched a total of 106 new features and/or services year to date*, for a total of 2,102 new features and/or services since inception in 2006.



Globally Accessible: AWS Global Infrastructure



12 Regions
33 Availability Zones
54 Edge Locations

○ Region
● Edge Location



You choose where your apps and data go!

North America

US East (N. Virginia) Region

EC2 Availability Zones: 5

Launched 2006

US West (N. California) Region

EC2 Availability Zones: 3

Launched 2009

Canada (Montreal) Region

Announced

US West (Oregon) Region

EC2 Availability Zones: 3

Launched 2011

AWS GovCloud (US) Region

EC2 Availability Zones: 2

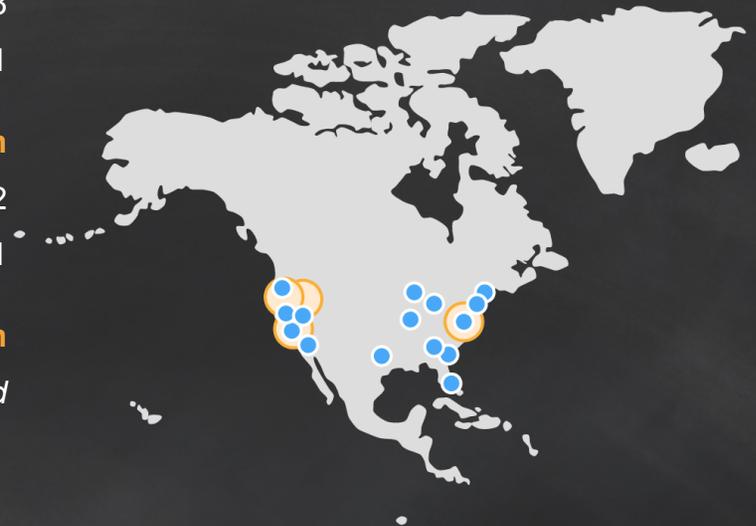
Launched 2011

US Central (Ohio) Region

Announced

AWS Edge Locations

Ashburn, VA (3); Atlanta, GA; Chicago, IL; Dallas/Fort Worth, TX (2); Hayward, CA; Jacksonville, FL; Los Angeles, CA (2); Miami, FL; New York, NY (3); Newark, NJ; Palo Alto, CA; San Jose, CA; Seattle, WA; South Bend, IN; St. Louis, MO



Enterprise Applications



Virtual Desktops



Sharing and Collaboration

Databases

Relational

NoSQL

Caching

Analytics

Hadoop

Real-Time

Data Warehouses

Data Workflows

App Services

Queuing

Orchestration

App Streaming
Transcoding

Email

Search

Deployment and Management

Containers

DevOps Tools

Resource Templates

Usage Tracking

Monitoring and Logs

Mobile Services

Identity

Sync

Mobile Analytics

Notifications

Foundation Services



Compute
(VMs, Auto Scaling
and Load Balancing)



Storage
(Object, Block
and Archive)



**Security and
Access Control**



Networking

Infrastructure



Regions



Availability Zones

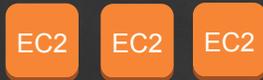


**Content Delivery Networks and Points of
Presence**

Compute Services

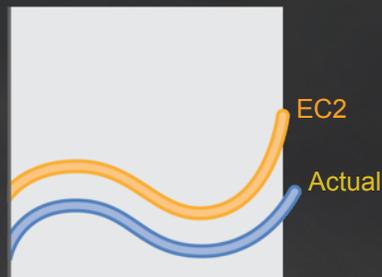
Amazon EC2

Elastic virtual servers
in the cloud



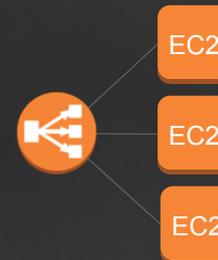
Auto Scaling

Automated scaling
of EC2 capacity



Elastic Load Balancing

Dynamic
traffic distribution



EC2

There are 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.

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 fps).
- 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.

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

C4

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

Features:

- High frequency Intel Xeon E5-2666 v3 (Haswell) processors optimized specifically for EC2
- EBS-optimized by default and at no additional cost
- Ability to control processor C-state and P-state configuration on the c4.8xlarge instance type
- Support for [Enhanced Networking](#) and Clustering

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

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.



Networking Services

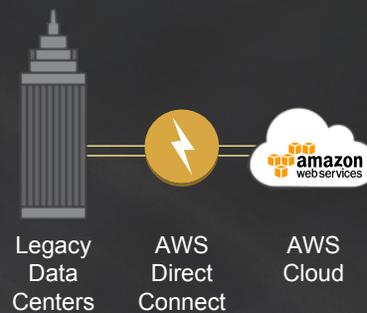
Amazon VPC

Private, isolated section of the AWS cloud



AWS Direct Connect

Private connectivity between AWS and your data center



Amazon Route 53

Domain Name System (DNS) web service



Storage Services

Amazon EBS

Block storage
for use with
Amazon EC2



Amazon S3

Internet
scale storage
via API



Images
Videos
Files
Binaries
Snapshots

Amazon Glacier

Storage for
archiving
and backup



Images
Videos
Files
Binaries
Snapshots

AWS Storage Gateway

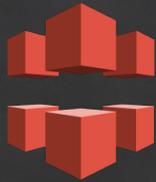
Integrates
on-premises IT
and AWS storage



Application Services

Amazon CloudFront

Distributing
content globally



Amazon CloudSearch

Managed search
service



Amazon Elastic Transcoder

Video transcoding



Database Services

Amazon RDS

Managed **relational** database service



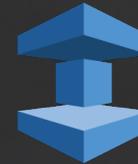
Amazon DynamoDB

Managed **NoSQL** database service



Amazon ElastiCache

In-memory **caching** service



Big Data Services

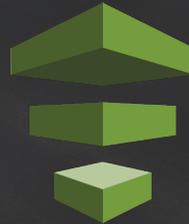
Amazon Elastic MapReduce (Amazon EMR)

Hosted **Hadoop**
framework



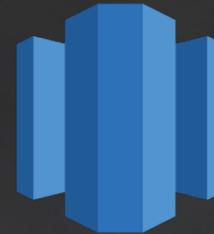
AWS Data Pipeline

Move data among
AWS services and
on-premises data sources



Amazon Redshift

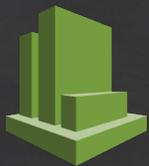
Petabyte-scale data
warehouse service



Deployment and Administration

Amazon CloudWatch

Monitor resources



AWS Identity and Access Management (IAM)

Manage users, groups, and permissions



Amazon OpsWorks

DevOps framework for application lifecycle management



AWS CloudFormation

Templates to deploy and manage



Web App



Enterprise App



Database

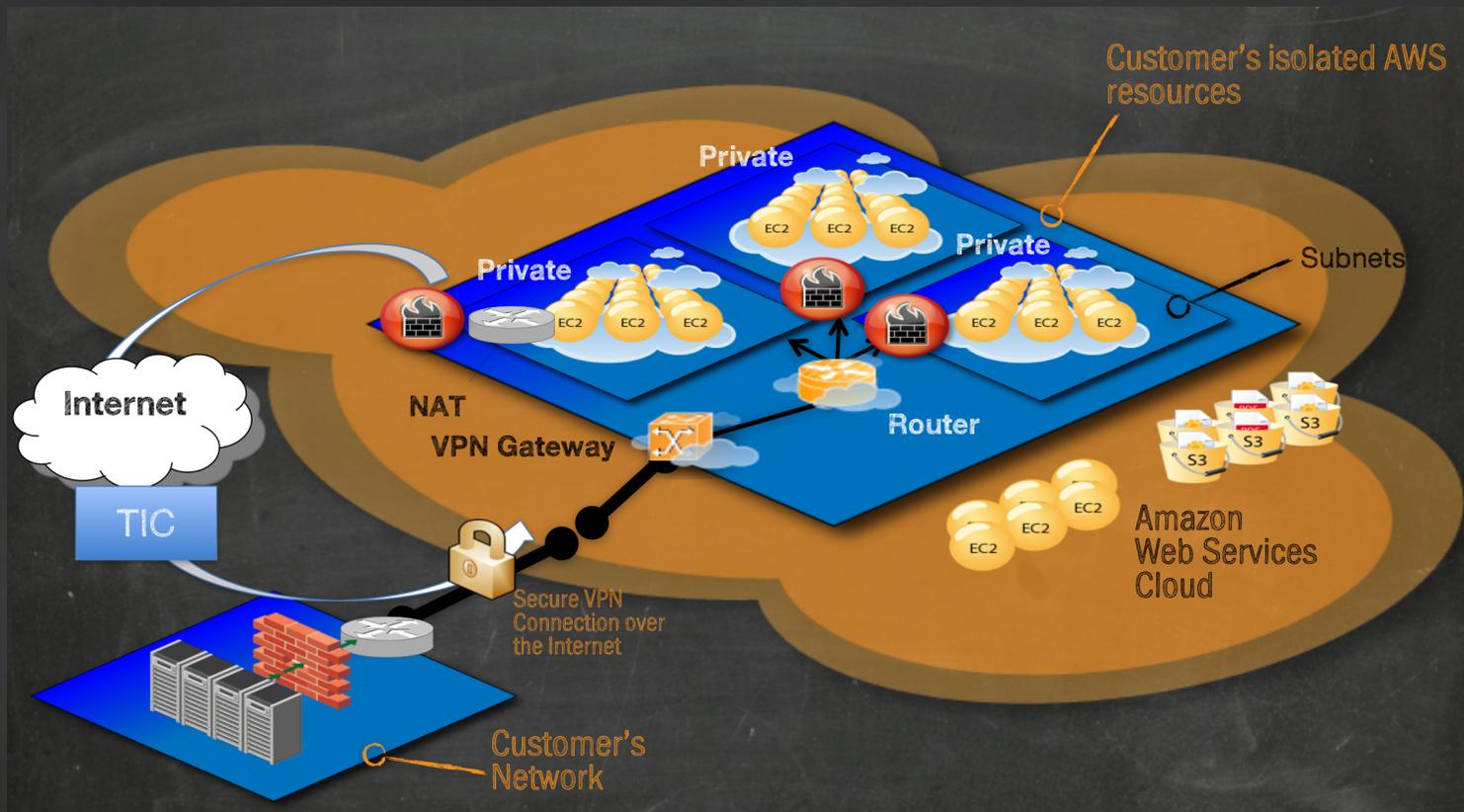
AWS Elastic Beanstalk

Automate resource management

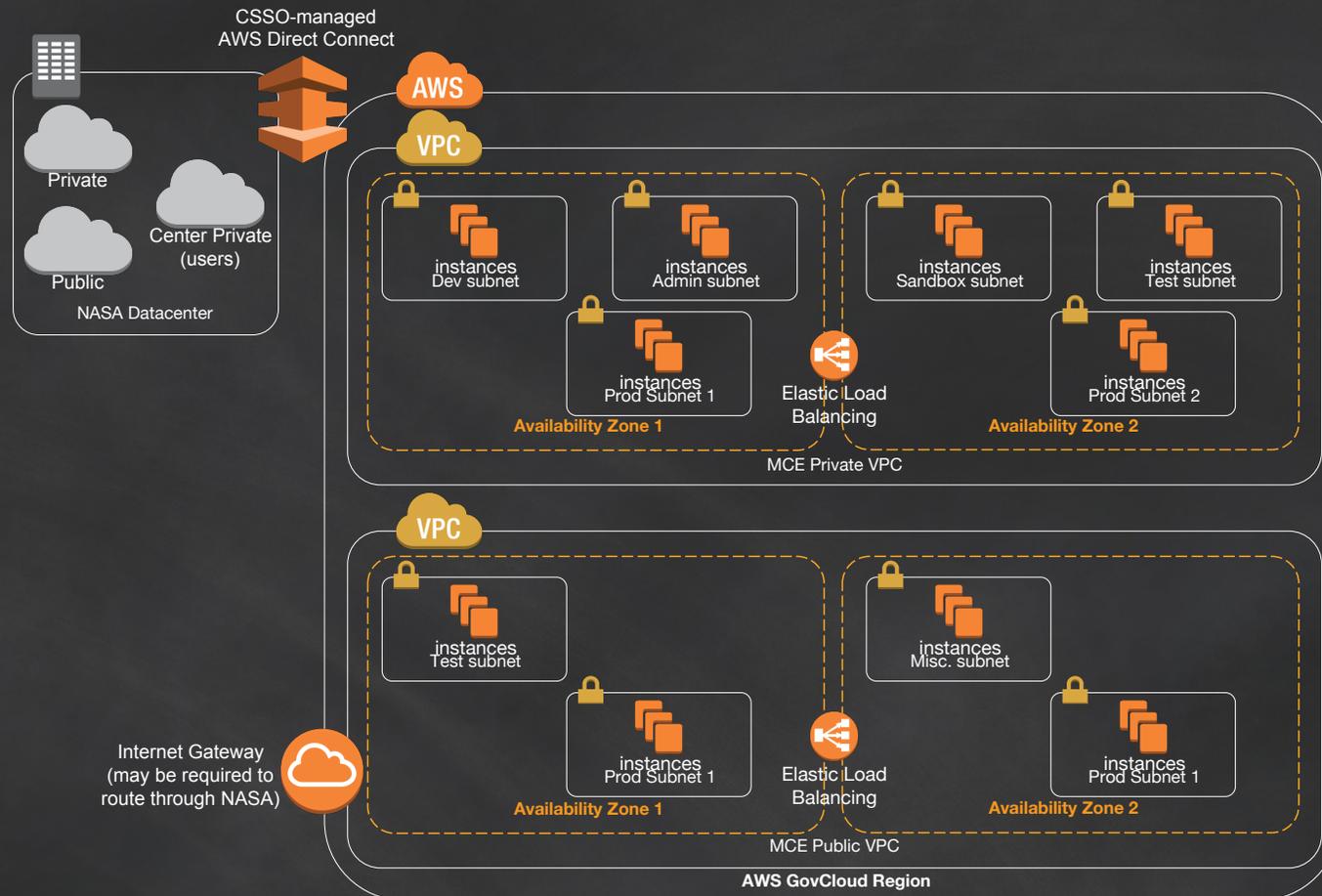


Secure: Virtual Private Cloud

Extremely secure by default.

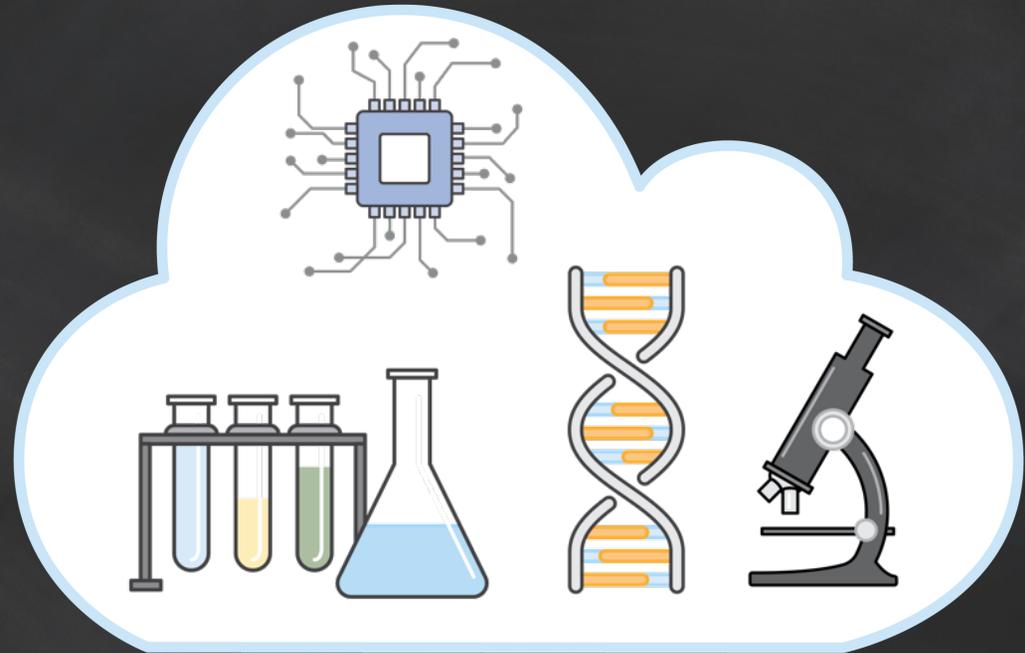


High-level Architecture



What do we mean by Scientific Computing?

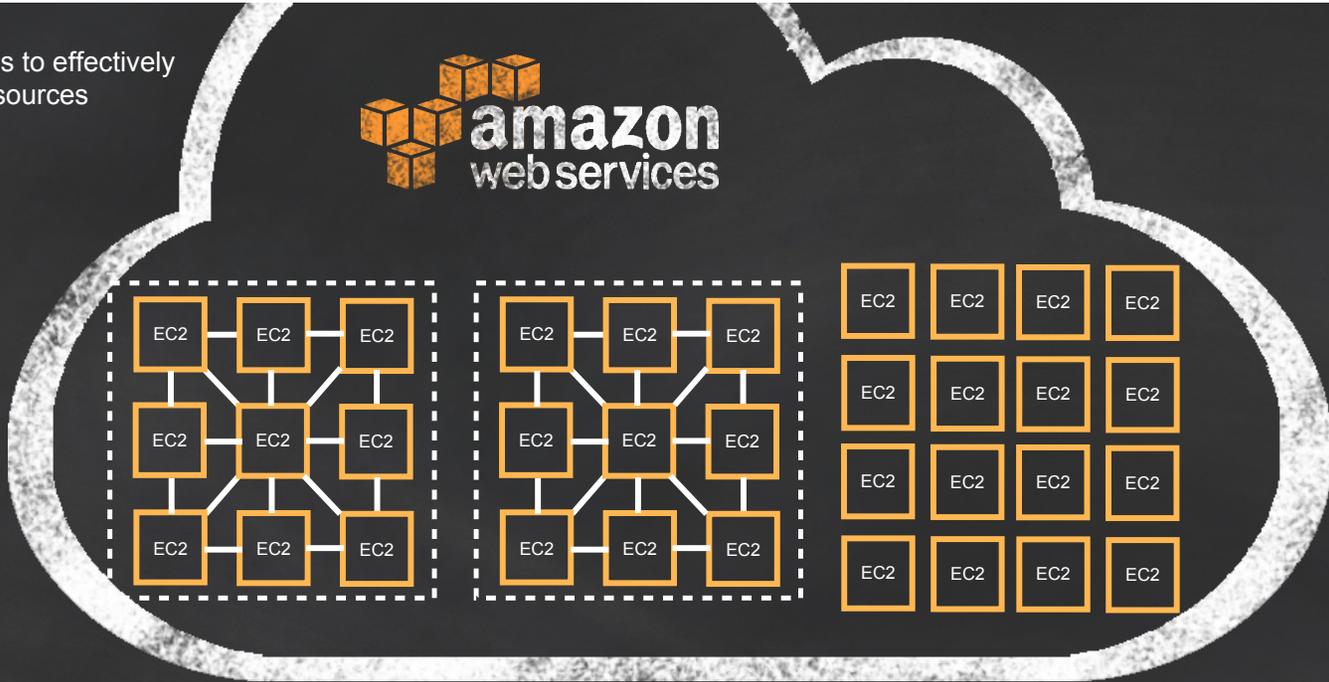
Scientific Computing refers to the application of **simulation**, mathematical **modeling** and **quantitative analysis** to analyze and solve **scientific problems**.



How is AWS used for Scientific Computing?

- High Performance Computing (HPC) for Engineering and Simulation
- High Throughput Computing (HTC) for Data-Intensive Analytics
- Hybrid Supercomputing centers
- Collaborative Research Environments
- Citizen Science
- Science-as-a-Service
- On-Demand Academic Training/Lab Environments

On-demand access to effectively limitless resources



GSFC,ARC, LARC



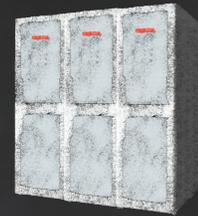
Specialized supercomputing resources



NASA Researcher



Small-scale shared compute



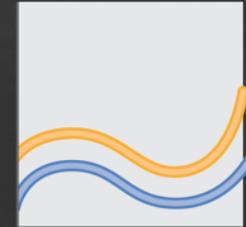
Why do researchers love using AWS?



Time to Science
Access research
infrastructure in minutes



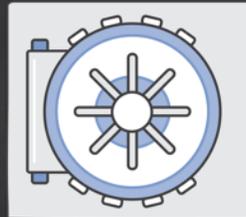
Low Cost
Pay-as-you-go pricing



Elastic
Easily add or remove capacity



Globally Accessible
Easily Collaborate with
researchers around the world



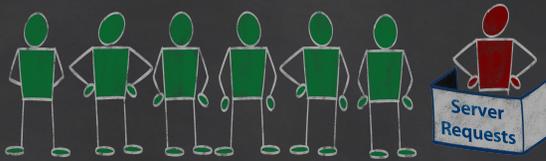
Secure
A collection of tools to
protect data and privacy



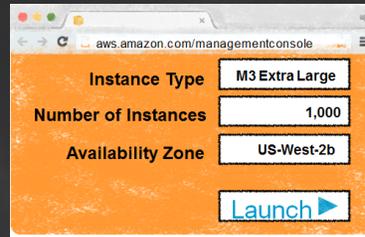
Scalable
Access to effectively
limitless capacity

Elastic: Research can't afford to be slow

Old World:
Infrastructure in Weeks



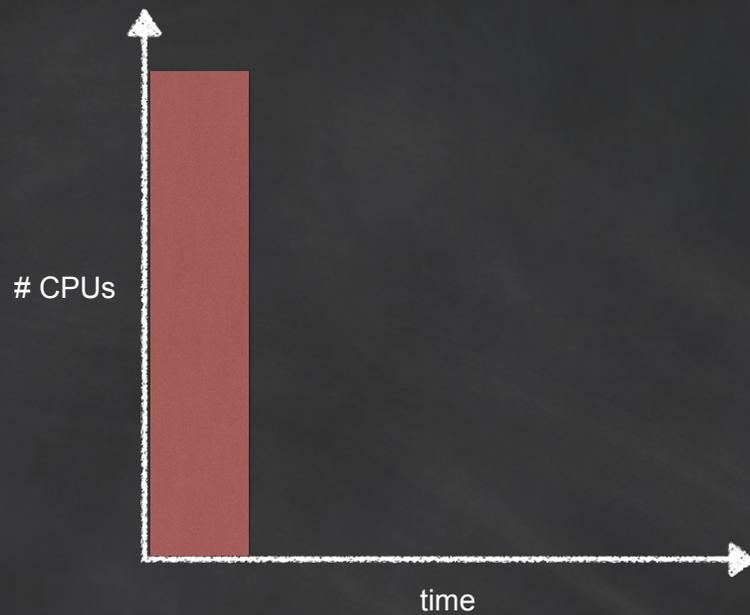
AWS:
Infrastructure in Minutes



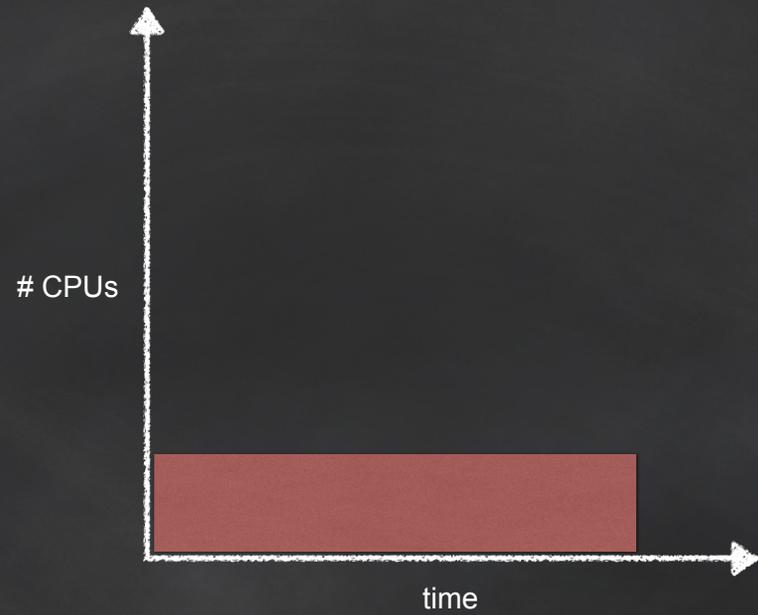
Add New Test Environment
Remove Test Environment
Add New DR Environment on East Coast
Add 100 Servers for Enrollment
Remove 100 Servers after Enrollment
Deploy 10,000 Core HPC Cluster
Shut down 10,000 Core HPC Cluster

Everything changes with this kind of agility

Scalable: Scientific results faster with scale



Wall clock time: **~1 hour**

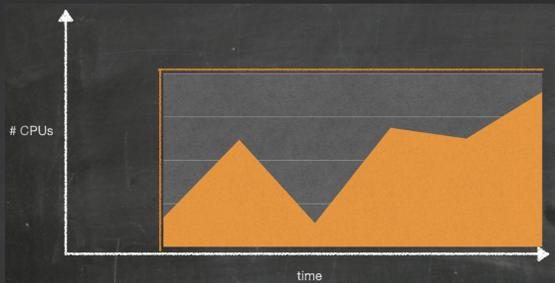


Wall clock time: **~1 week**

Cost: **equal**



Low Cost: pay-as-you-go pricing



When you only pay for what you use ...

- If you're only able to use your compute, say, 30% of the time, you only pay for that time.



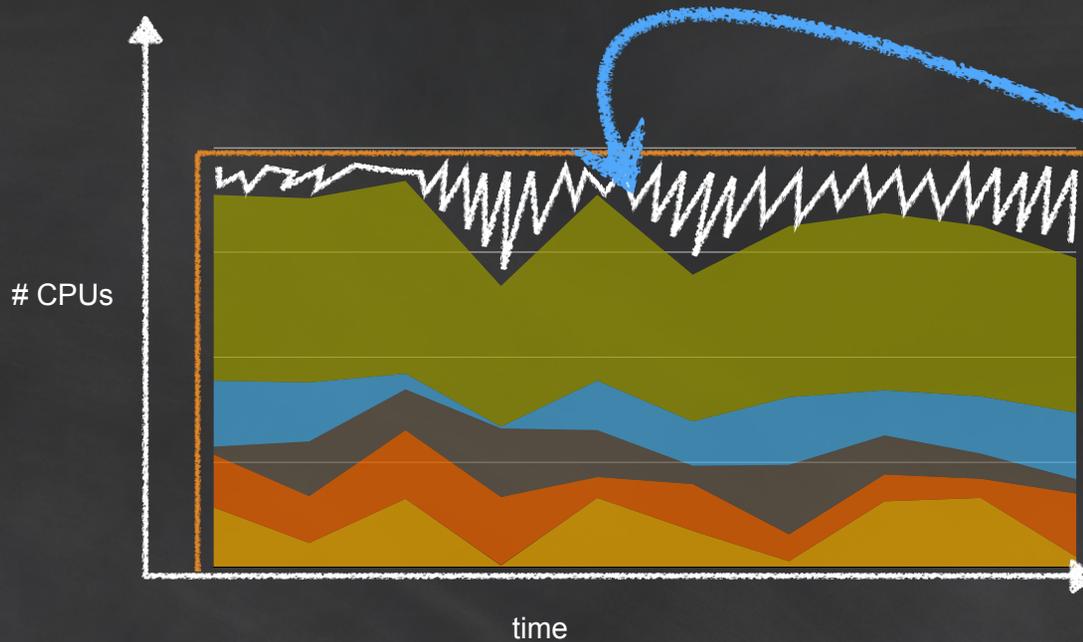
... you have options.

- 1 Pocket the savings**
 - Buy chocolate
 - Buy a spectrometer
 - Hire a scientist.

- 2 Go faster**
 - Use 3x the cores to run your jobs at 3x the speed.

- 3 Go Large**
 - Do 3x the science, or consume 3x the data.

Low Cost: Spot Market



Spot Market

Our ultimate space filler.

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

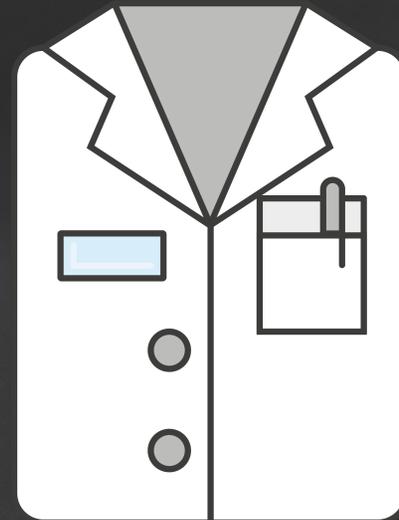
Great for workloads that aren't time sensitive, and especially popular in research (hint: it's really cheap).

Low Cost: Research Grants

AWS provides free usage credits to help researchers:

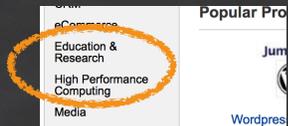
- Teach advanced courses
- Explore new projects
- Create resources for the scientific community

aws.amazon.com/grants

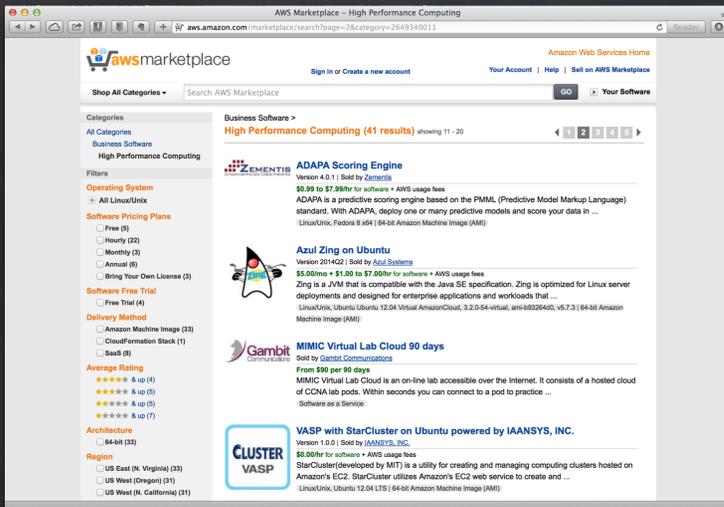


Time to Science: The AWS Marketplace

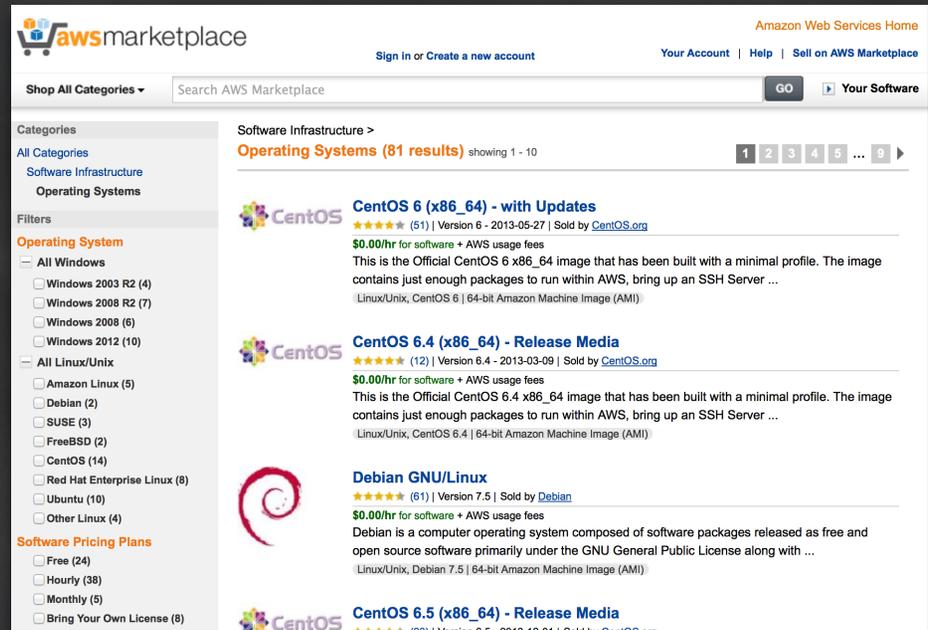
aws.amazon.com/marketplace



Dozens of scientific applications. Need more!



Almost all major research-focused operating systems.



Time to Science: Community AMIs

Step 1: Choose an AMI

An AMI is a template that contains the software configuration (operating system, application server, and applications) required to launch your instance. You can select an AMI provided by AWS, our user community, or the AWS Marketplace;

Quick Start

My AMIs

AWS Marketplace

Community AMIs

Operating system

- Amazon Linux
- Cent OS
- Debian
- Fedora
- Gentoo
- OpenSUSE
- Other Linux
- Red Hat
- SUSE Linux
- Ubuntu
- Windows

EC2 Management Console

Services Edit

Brendan Bouffler Sydney Help

1. Choose AMI 2. Choose Instance Type 3. Configure Instance 4. Add Storage 5. Tag Instance 6. Configure Security Group 7. Review

Step 1: Choose an Amazon Machine Image (AMI)

Cancel and Exit

An AMI is a template that contains the software configuration (operating system, application server, and applications) required to launch your instance. You can select an AMI provided by AWS, our user community, or the AWS Marketplace; or you can select one of your own AMIs.

Quick Start

My AMIs

AWS Marketplace

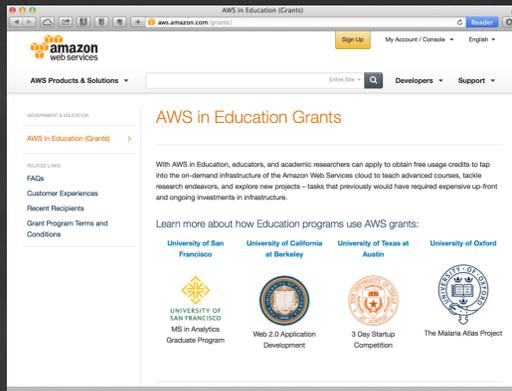
Search: hvm

35 results for "hvm" on AWS Marketplace
Partner software pre-configured to run on AWS

Operating system	AMI ID	Architecture
<input type="checkbox"/> Amazon Linux		
<input checked="" type="checkbox"/> Cent OS		
<input type="checkbox"/> Debian		
<input type="checkbox"/> Fedora		
<input type="checkbox"/> Gentoo		
<input type="checkbox"/> OpenSUSE		
<input type="checkbox"/> Other Linux		
<input type="checkbox"/> Red Hat		
<input type="checkbox"/> SUSE Linux		
<input type="checkbox"/> Ubuntu		
<input type="checkbox"/> Windows		
	brightheadnode-7.0-10-centos6-hvm - ami-233c5b19	64-bit
	Bright Cluster Manager v7.0	
	Root device type: ebs Virtualization type: hvm	
	CentOS 6.5 x86_64 (HVM) - by VisualOps - ami-3703640d	64-bit
	Minimal CentOS 6.5 + cloud-init, with VisualOps OpsAgent	
	Root device type: ebs Virtualization type: hvm	
	CentOS 6.4 x86_64 (HVM) - by VisualOps - ami-3d036407	64-bit
	Minimal CentOS 6.4 + cloud-init, with VisualOps OpsAgent	
	Root device type: ebs Virtualization type: hvm	
	RightImage_CentOS_6.5_x64_v13.5.2_HVM_EBS - ami-45950b7f	64-bit

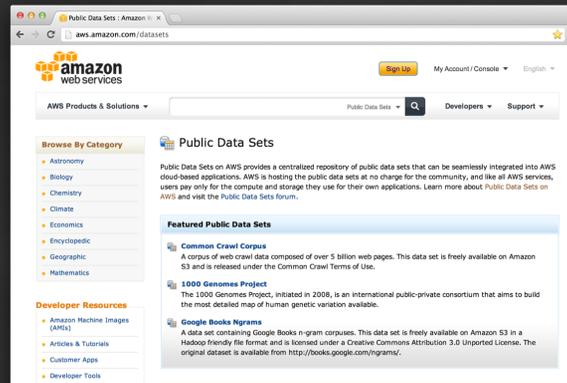
AWS Global Impact Initiatives for Science

AWS Research Grants



- Grants to initiate & support development of cloud-enabled technologies.
- Typically one-off grants of AWS resources like EC2 (compute) or S3 & EBS (storage) or more exotic like Kinesis & twitter feeds.
- Frequently results in reusable resources, like AMIs or open data, which we strongly encourage.
- Lowers the risk to try the cloud.

AWS Hosted Public Datasets



- Large and globally significant datasets hosted and paid for by AWS for community use.
- Data can be quickly and easily processed with elastic computing resources in the surrounding cloud.
- AWS hopes to enable more innovation, more quickly.
- Provided in partnership with content owners, who curate the data.

AWS SciCo Team

- **Dedicated team** focusing on Scientific Computing & Research workloads
- Globally focussed and engaged in Big Science projects like the **SKA**.
- Leveraging AWS resources all over the world.
- Ensuring the cloud is able to make a disruptive **impact on science**.



AWS and the NASA Earth eXchange (NEX)

- National Climate Assessment datasets hosted on AWS
- Machine images, tutorials and hosted workshops provided by NASA
- Data and Software now available to those without @nasa.gov email addresses
- Enables crowd-sourced citizen science applications like those found on the Zooniverse



Thank you



Visit the NASA EMCC online: <https://intranet.share.nasa.gov/agency/cloudservices>

