02/12/2013

Research Computing on Amazon Web Services

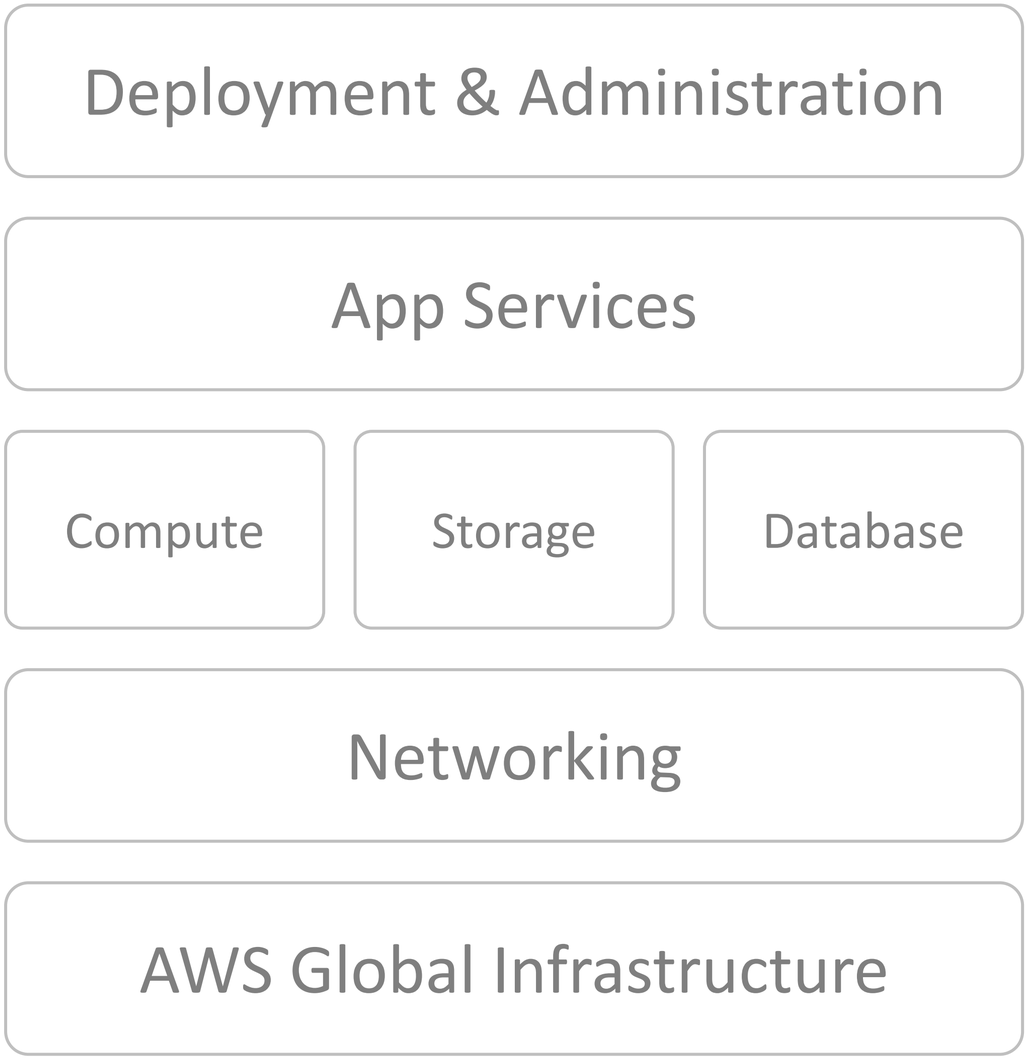
Chris Hayman | Principal Solutions Architect

Agenda

Amazon Web Services Reference Model

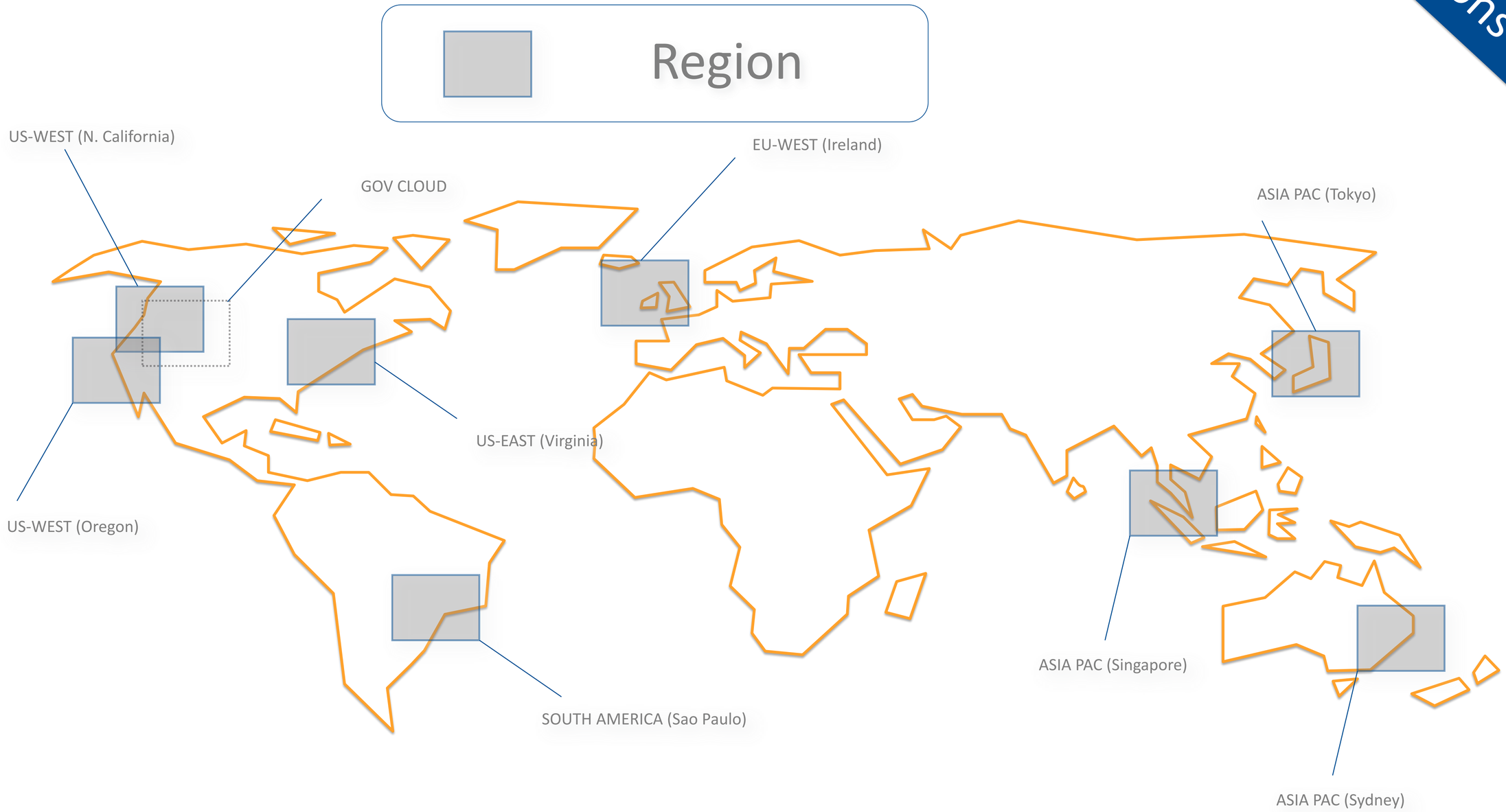
1) Compute: Amazon EC2
   Storage: Amazon EBS, S3 & Glacier
2) Public Datasets
3) Managed Analytics: Amazon Kinesis & EMR
4) Bring Your Own (HPC Tools)

Reference Model



On a global footprint

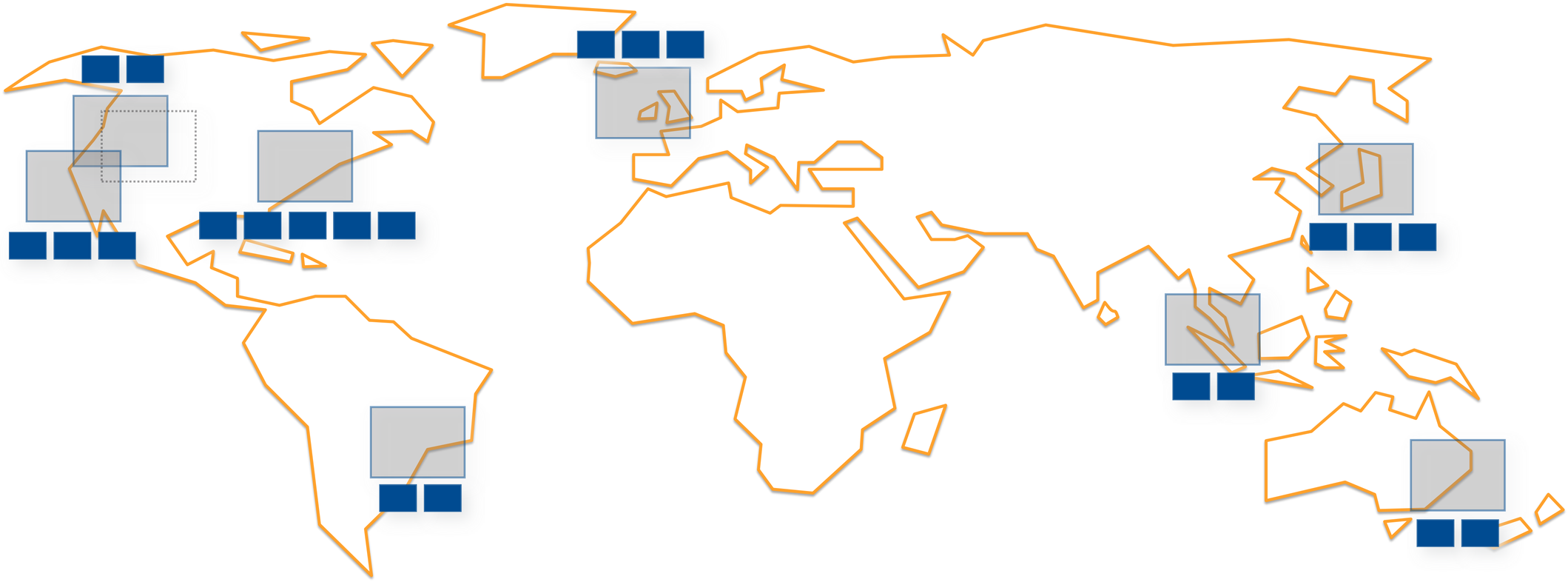
Regions



On a global footprint

Zones

Availability Zone



On a global footprint

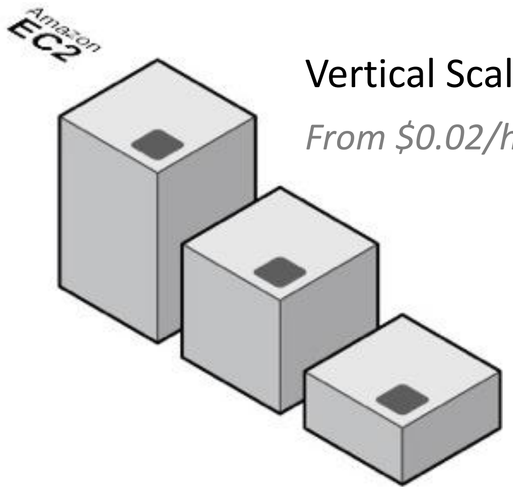
Edges

Compute

• • • •

Storage

Compute



Vertical Scaling

From $0.02/hr

Deployment & Administration

App Services

Compute | Storage | Database

Networking

AWS Global Infrastructure

Elastic Compute Cloud (EC2)

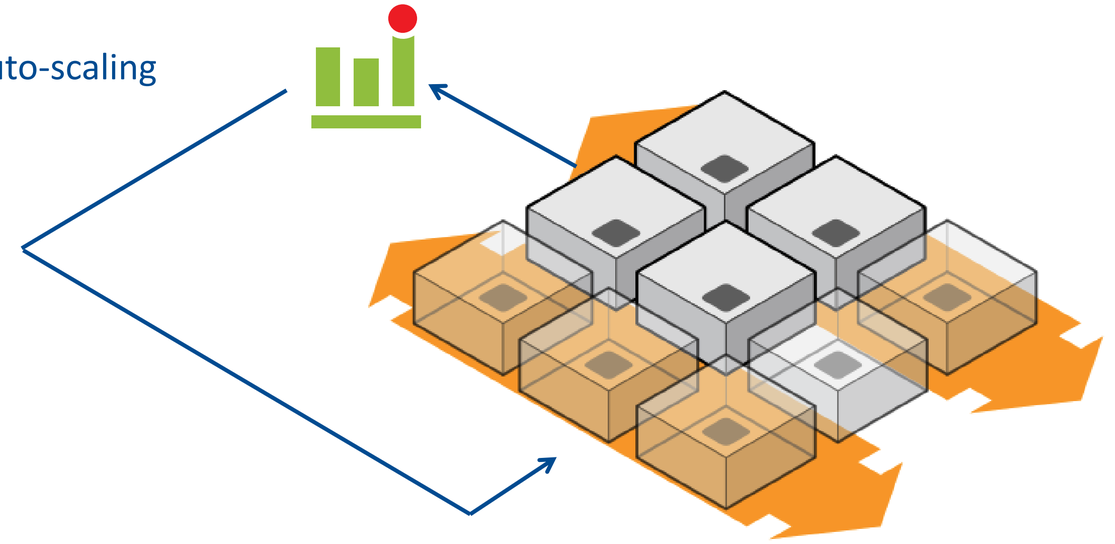
Basic unit of compute capacity

Range of CPU, memory & local disk options

23 Instance types available, from micro to cluster compute

Feature	Details
Flexible	Run windows or Linux distributions
Scalable	Wide range of instance types from micro to cluster compute
Machine Images	Configurations can be saved as machine images (AMIs) from which new instances can be created
Full control	Full root or administrator rights
Secure	Full firewall control via Security Groups
Monitoring	Publishes metrics to Cloud Watch
Inexpensive	On-demand, Reserved and Spot instance types
VM Import/Export	Import and export VM images to transfer configurations in and out of EC2

Compute




```
as-create-auto-scaling-group MyGroup
    --launch-configuration MyConfig
    --availability-zones eu-west-1a
    --min-size 4
    --max-size 200
```

- Deployment & Administration
- App Services
- **Compute**
- Storage
- Database
- Networking
- AWS Global Infrastructure

Auto-scaling

Automatic re-sizing of compute clusters based upon demand

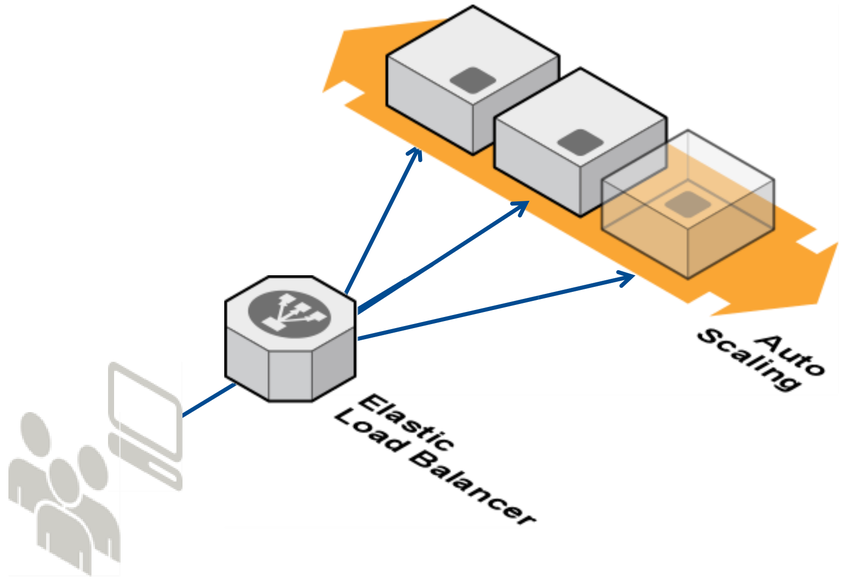
Feature	Details
Control	Define minimum and maximum instance pool sizes and when scaling and cool down occurs
Integrated to CloudWatch	Use metrics gathered by CloudWatch to drive scaling
Instance types	Run auto scaling for on-demand instances and spot. Compatible with VPC

Compute

Elastic Load Balancing

Create highly scalable applications

Distribute load across EC2 instances in multiple availability zones



Deployment & Administration

App Services

Compute | Storage | Database

Networking

AWS Global Infrastructure

Feature	Details
Auto-scaling	Automatically scales to handle request volume
Available	Load balance across instances in multiple availability zones
Health checks	Automatically checks health of instances and takes them in or out of service
Session stickiness	Route requests to the same instance
Secure sockets layer	Supports SSL offload from web and application servers with flexible cipher support
Monitoring	Publishes metrics to Cloud Watch

amazon web services

New C3 Instances

Instance Name	vCPU Count	Total ECU	RAM	Local Storage	Hourly On-Demand
c3.large	2	7	3.75 GiB	2 x 16 GB SSD	$0.15
c3.xlarge	4	14	7 GiB	2 x 40 GB SSD	$0.30
c3.2xlarge	8	28	15 GiB	2 x 80 GB SSD	$0.60
c3.4xlarge	16	55	30 GiB	2 x 160 GB SSD	$1.20
c3.8xlarge	32	108	60 GiB	2 x 320 GB SSD	$2.40

2.8 GHz Intel Xeon E5-2680 (Ivy Bridge) processor.

amazon web services

New I2 Instances (Coming Soon)

Instance Name	vCPU Count	RAM	Instance Storage (SSD)
i2.large	2	15 GiB	1 x 360 GB
i2.xlarge	4	30.5 GiB	1 x 720 GB
i2.2xlarge	8	61 GiB	2 x 720 GB
i2.4xlarge	16	122 GiB	4 x 720 GB
i2.8xlarge	32	244 GiB	8 x 720 GB

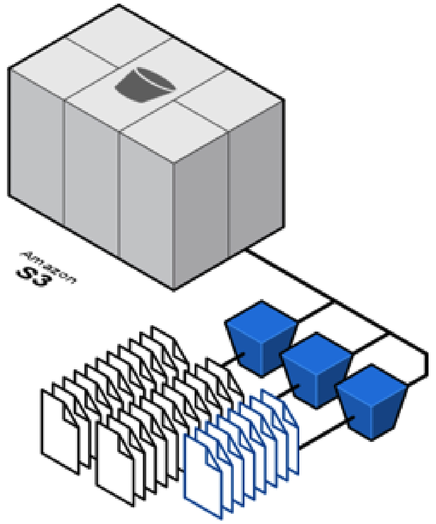
2.5 GHz intel Xeon E5-2670v2

i2.8xlarge instances will be able to deliver 350,000 random read IOPS and 320,000 random write IOPS.

New G2 Instances

NVIDIA GRID™ (GK104 "Kepler") GPU (Graphics Processing Unit),
1,536 CUDA cores and 4 GB of video (frame buffer) RAM.
Intel Sandy Bridge processor running at 2.6 GHz with Turbo
Boost enabled, 8 vCPUs (Virtual CPUs).
15 GiB of RAM.
60 GB of SSD storage.

Storage



Deployment & Administration

App Services

Compute | **Storage** | Database

Networking

AWS Global Infrastructure

S3 - Durable storage, any object

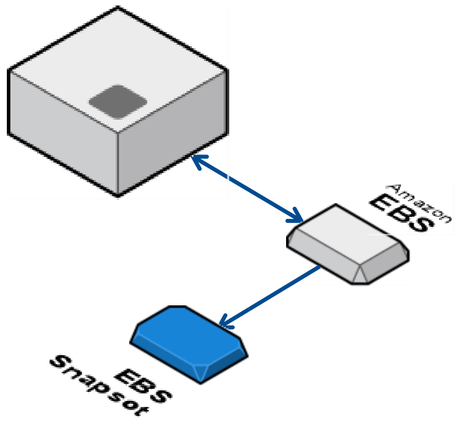
99.999999999% durability of objects

Unlimited storage of objects of any type

Up to 5TB size per object

Feature	Details
Flexible object store	Buckets act like drives, folder structures within
Access control	Granular control over object permissions
Server-side encryption	256bit AES encryption of objects
Multi-part uploads	Improved throughput & control
Object versioning	Archive old objects and version new ones
Object expiry	Automatically remove old objects
Access logging	Full audit log of bucket/object actions
Web content hosting	Serve content as web site with built in page handling
Notifications	Receive notifications on key events
Import/Export	Physical device import/export service

Storage



Elastic Block Store

High performance block storage device

1GB to 1TB in size

Mount as drives to instances

Deployment & Administration

App Services

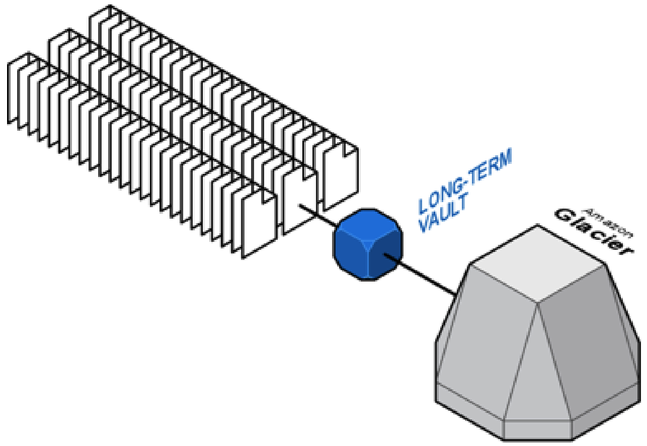
Compute | **Storage** | Database

Networking

AWS Global Infrastructure

Feature	Details
High performance file system	Mount EBS as drives and format as required
Flexible size	Volumes from 1GB to 1TB in size
Secure	Private to your instances
Performance	Use provisioned IOPS to get desired level of IO performance
Available	Replicated within an Availability Zone
Backups	Volumes can be snapshotted for point in time restore
Monitoring	Detailed metrics captured via Cloud Watch

Storage



Deployment & Administration

App Services

Compute | **Storage** | Database

Networking

AWS Global Infrastructure

Glacier

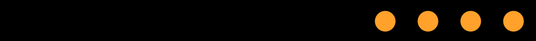
Low-cost storage service

Secure and durable storage for backup and archive

For data that is infrequently accessed

Feature	Details
Low cost	$0.01/GB/month with no up-front capital commitments
Durable	Same 99.999999999% durability as S3
Flexible	Store any amount of data on-demand. Eliminate the need for capacity planning
Secure	Leverage AWS' robust security platform. Control access to your data.
API	REST-based API to send and receive data
Import/Export	Optionally use portable storage devices to import/export mass data
Vault inventory	Index for real-time view of the contents of the vault

Public Datasets



AWS Public Data Sets

- A centralized repository of public datasets
- Seamless integration with cloud based applications
- No charge to the community
- Some of the datasets available today:
  - 1000 Genomes Project
  - Human Microbiome Project
  - Ensembl
  - GenBank
  - Illumina – Jay Flateley Human Genome Dataset
  - YRI Trio Dataset
  - UniGene
  - Influenza Virus
  - PubChem
- **Tell us what else you'd like for us to host …**

Managed Analytics

Amazon Kinesis 02/12/2013

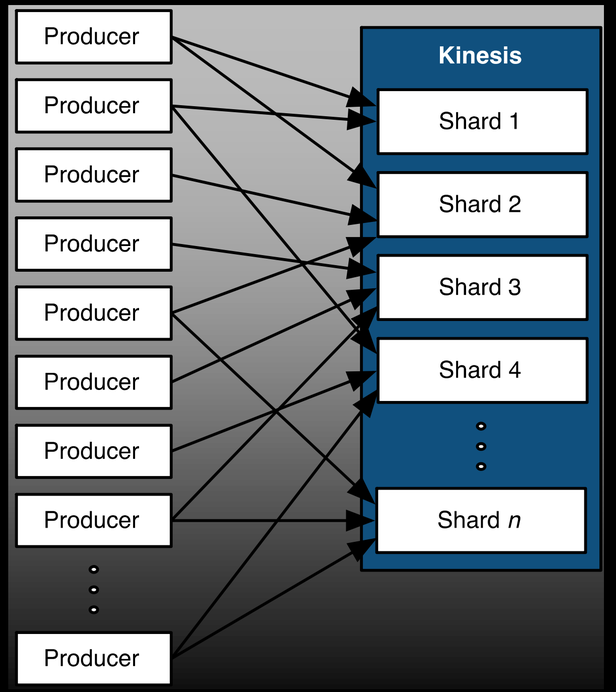
Currently in Limited Preview

Managed Service for Real-Time Processing of Big Data

Putting data into Kinesis
Managed Service for Ingesting Fast Moving Data

- Streams are made of Shards
  - A Kinesis Stream is composed of multiple **Shards**
  - Each Shard ingests up to 1MB/sec of data and up to 1000 TPS
  - All data is stored for 24 hours
  - You scale Kinesis streams by adding or removing Shards

- Simple PUT interface to store data in Kinesis
  - Producers use a **PUT** call to store data in a Stream
  - A **Partition Key** is used to distribute the PUTs across Shards
  - A unique **Sequence #** is returned to the Producer upon a successful PUT call



Application Services

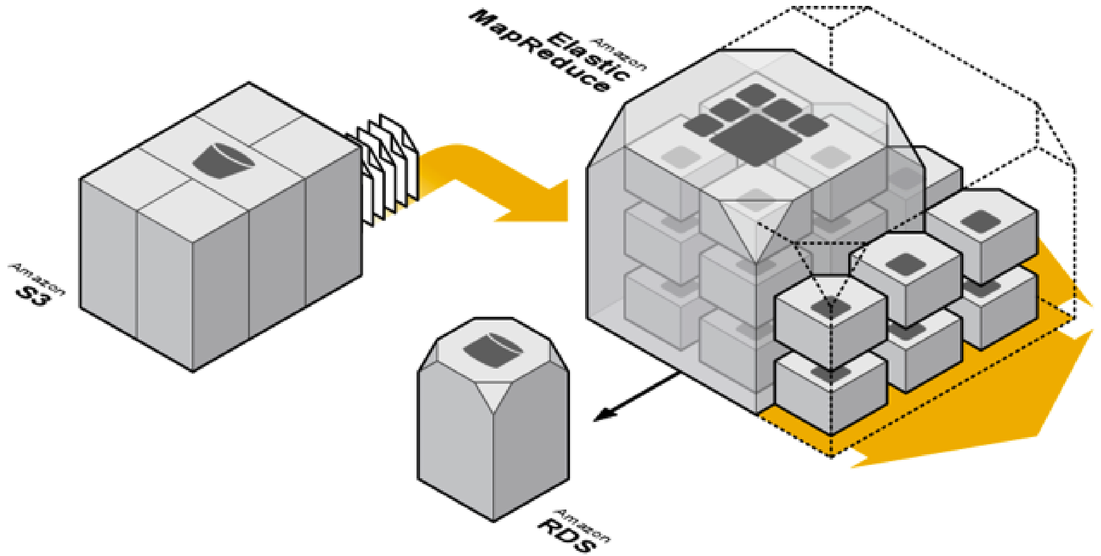
Elastic MapReduce

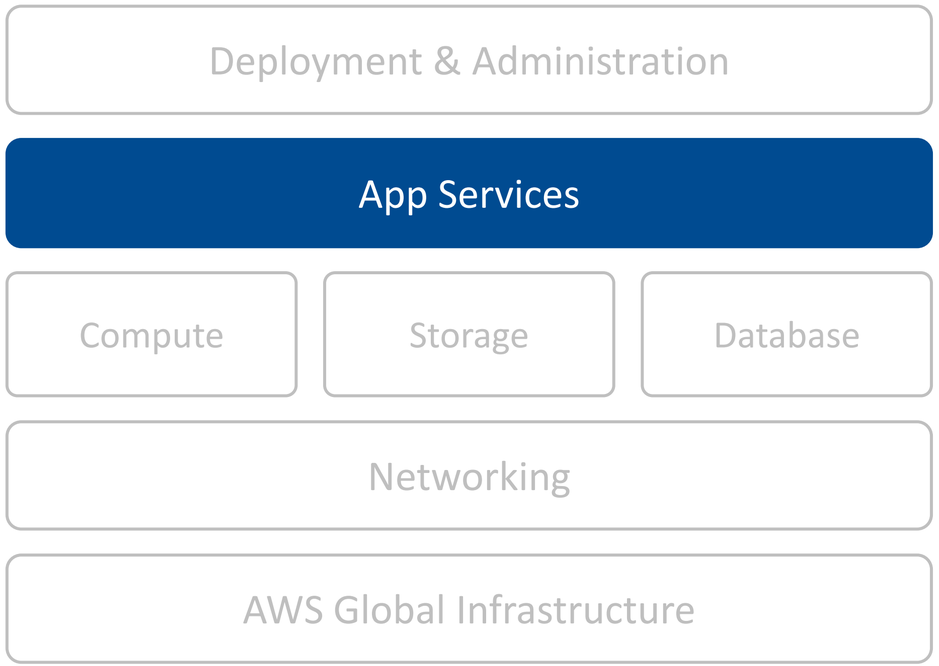
Managed, elastic Hadoop cluster

Integrates with S3 & DynamoDB

Leverage Hive & Pig analytics scripts

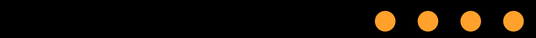
Integrates with instance types such as spot





Feature	Details
Scalable	Use as many or as few compute instances running Hadoop as you want. Modify the number of instances while your job flow is running
Integrated with other services	Works seamlessly with S3 as origin and output. Integrates with DynamoDB
Comprehensive	Supports languages such as Hive and Pig for defining analytics, and allows complex definitions in Cascading, Java, Ruby, Perl, Python, PHP, R, or C++
Cost effective	Works with Spot instance types
Monitoring	Monitor job flows from with the management console

HPC Tools



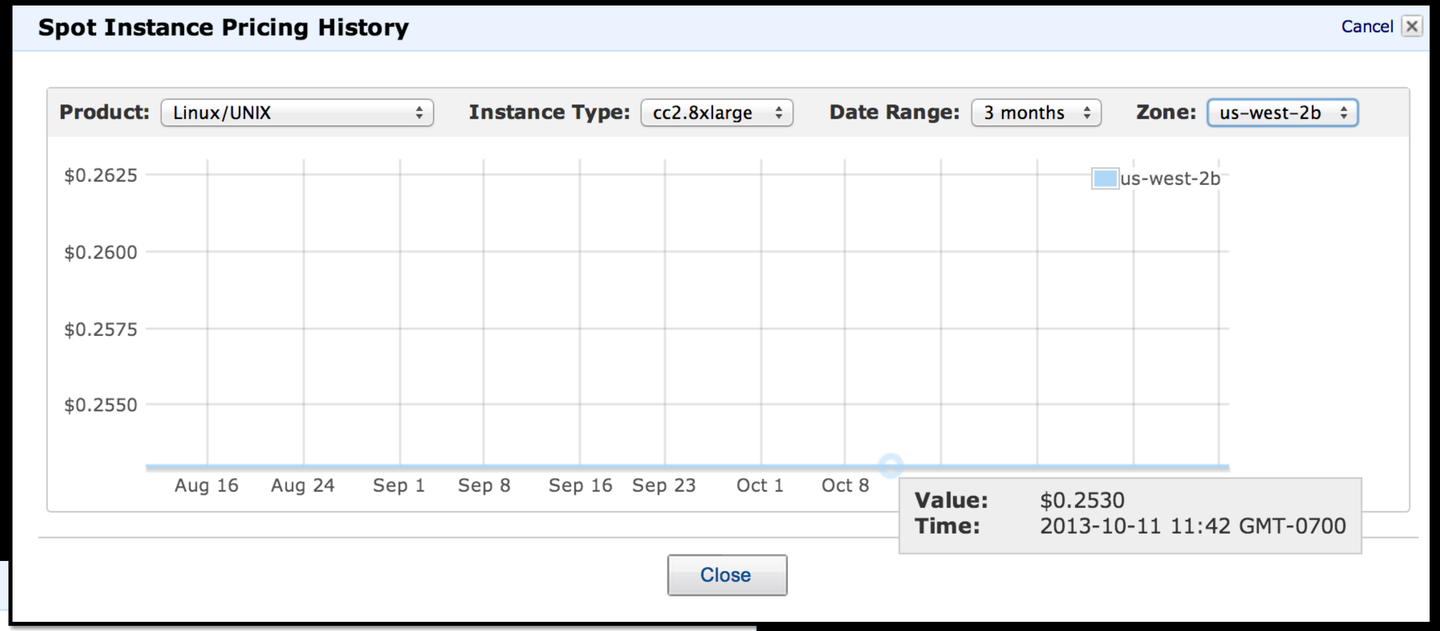
“Supercomputing simulation employs 156,000 Amazon processor cores

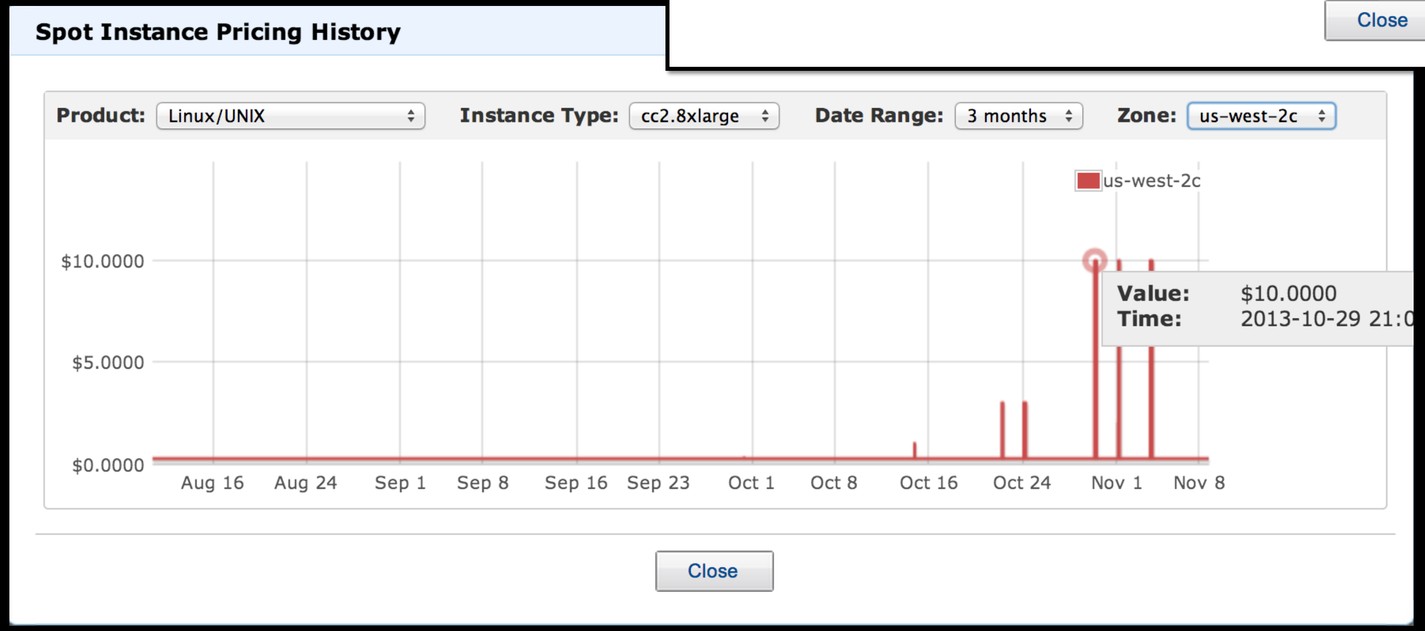
To simulate 205,000 molecules as quickly as possible for a USC simulation, Cycle Computing fired up a mammoth amount of Amazon servers around the globe.” [1]

[1] c|net news

http://news.cnet.com/8301-1001_3-57611919-92/supercomputing-simulation-employs-156000-amazon-processor-cores/

amazon web services





Schedulers

- MIT Star Cluster
- Grid Engine
- HT Condor

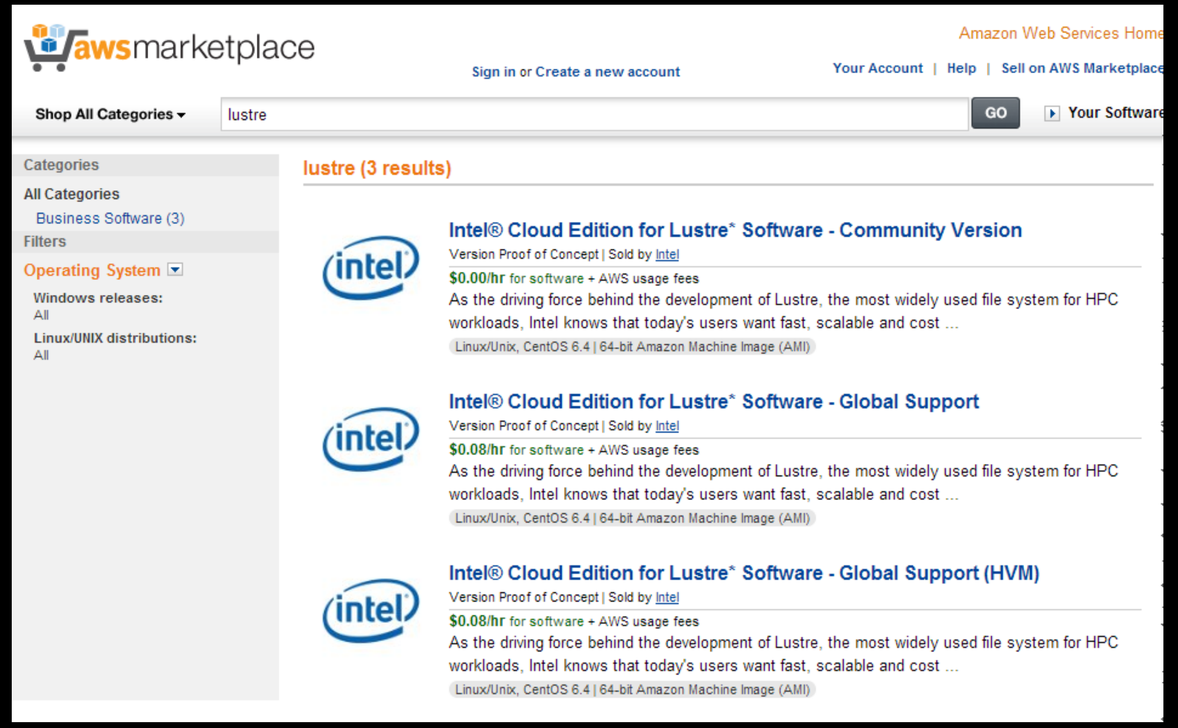
Filesystems on EBS / Local Storage

- Lustre
- OrangeFS
- Etc.

Network

- 10 GbE full bisection bandwidth
- C3 “Enhanced Networking Capability”

```
$ sudo ethtool -i eth0
driver: ixgbevf version: 2.11.3 firmware-version:
N/A
bus-info: 0000:00:03.0
supports-statistics: yes
supports-test: yes
supports-eeprom-access: no
supports-register-dump: yes
supports-priv-flags: no
```



aws.amazon.com/grants

hayman@amazon.com

Thank You

Chris Hayman | Principal Solutions Architect