



Introduction to Amazon Cloud & EC2 Overview

David Christian
Enterprise Solutions Architect
chradavi@amazon.com



Agenda

- Introduction to AWS Cloud
- Global Reach
- EC2 Overview
- EC2 Details

What is AWS?

AWS provides a highly reliable, scalable, low-cost infrastructure platform in the cloud that powers millions of businesses in over 190 countries around the world.

Benefits

- Low Cost
- Elasticity & Agility
- Open & Flexible
- Secure
- Global Reach



What sets AWS apart?



Experience

Building and managing cloud since 2006



Service Breadth & Depth

175+ services to support any cloud workload



Pace of Innovation

History of rapid, customer-driven releases



Global Footprint

24 regions, 76 availability zones, 216 edge locations



Pricing Philosophy

77 proactive price reductions to date



Ecosystem

Thousands of consulting/system integrator & technology partners

Experience with Operational Reliability

Our goal is to make our operational performance indistinguishable from perfect. We are driven to remove any all causes of failure.

- We have spent over a decade building the world's most reliable, secure, scalable, and cost-effective infrastructure.
- Service SLAs between 99.9% and 100% availability. Amazon S3 is designed for 99.99999999% durability.
- Availability Zones exist on isolated fault lines, flood plains, and electrical grids to substantially reduce the chance of simultaneous failure.
- The AWS Service Health Dashboard provides 24/7 visibility in the real-time operational status of all services around the globe.

Pricing Philosophy

High volume / low margin businesses are in our core DNA

Trade CapEX for
variable expense

Pay for what
you use

Our economies of
scale provide us
with lower costs

77 price
reductions
since 2006

Pricing model
choice to support
variable and stable
workloads

On-demand
Reserved Instances
Spot

Save more money as
you grow bigger

Tiered pricing
Volume discounts
Custom pricing

Customer obsessed



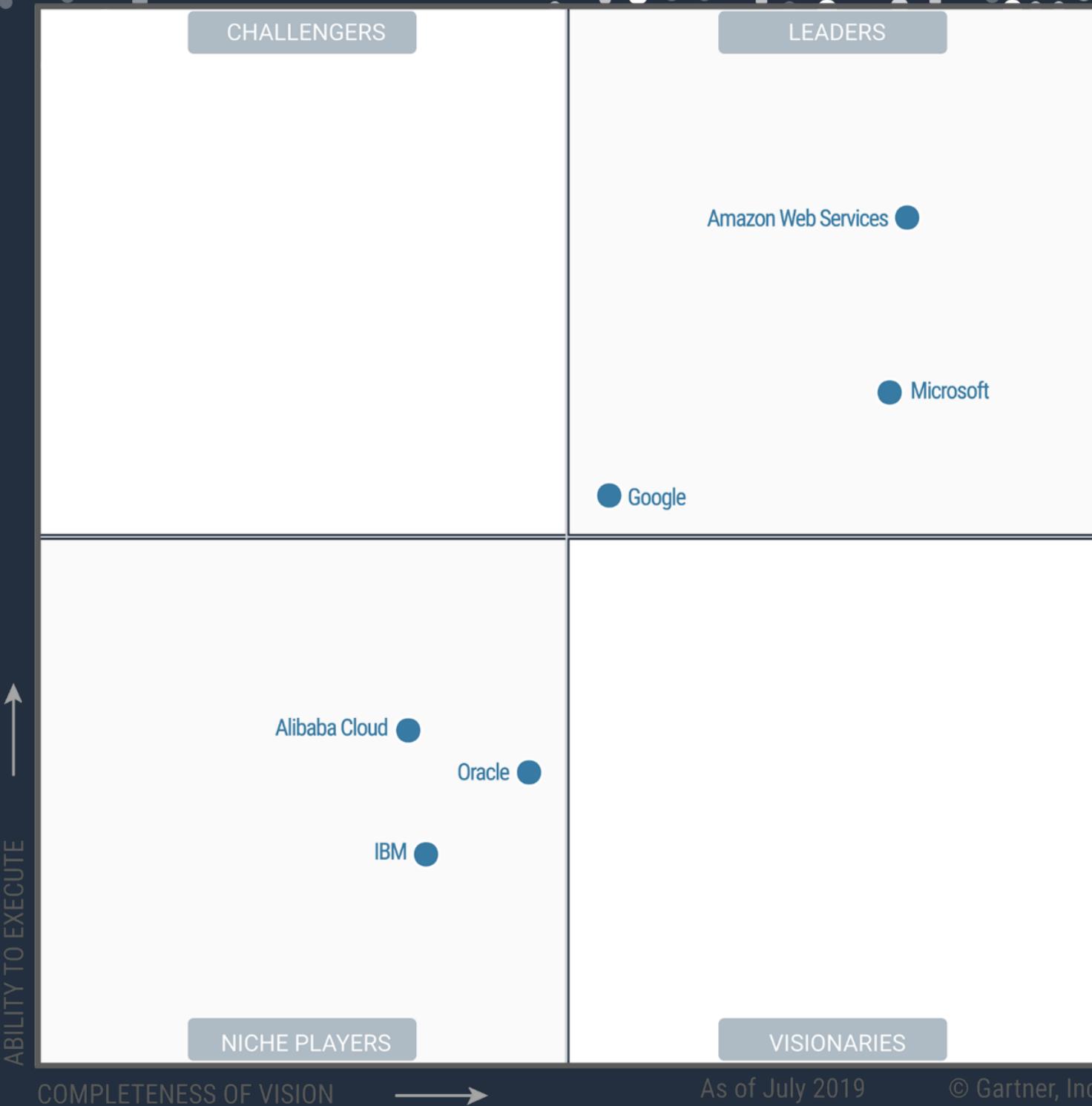
90%
of roadmap originates with customer requests
and are designed to meet specific needs



“Performance, reliability, and responsiveness are fundamental to our customer experience, and T3 instances help us to deliver on that customer promise while also controlling our costs.”

—Heroku

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



AWS Recognized as
a Cloud Leader for the
9th Consecutive Year

Gartner, Magic Quadrant for Cloud Infrastructure as a Service, Worldwide, Raj Bala, Bob Gill, Dennis Smith, David Wright, July 2019. ID G00365830. Gartner does not endorse any product or service depicted in its research publications, and does not advise technology users to select only those vendors with the highest ratings. 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. The Gartner logo is a trademark and service mark of Gartner, Inc., and/or its affiliates, and is used herein with permission. All rights reserved.

1

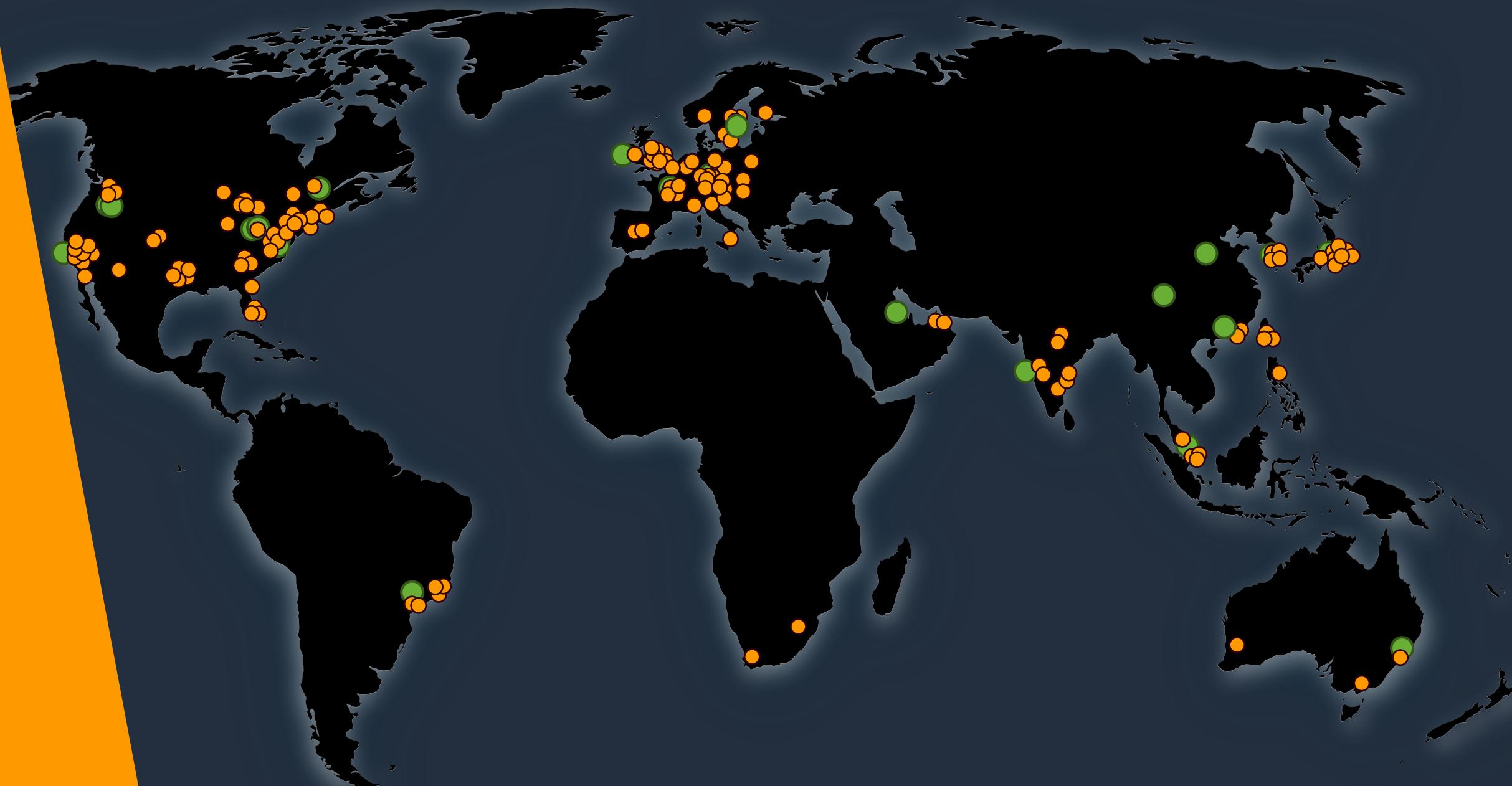
AWS Global Reach

24
Regions

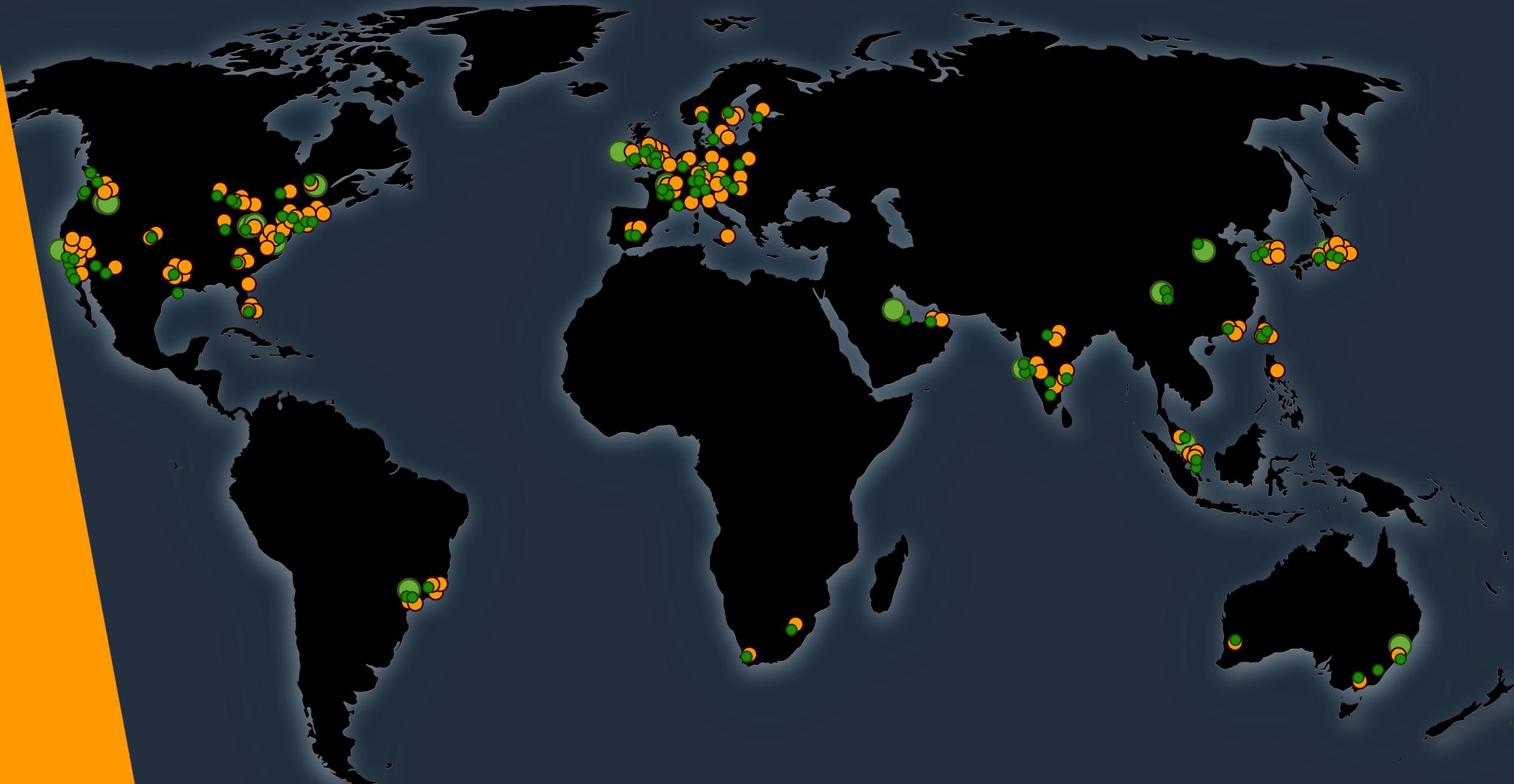


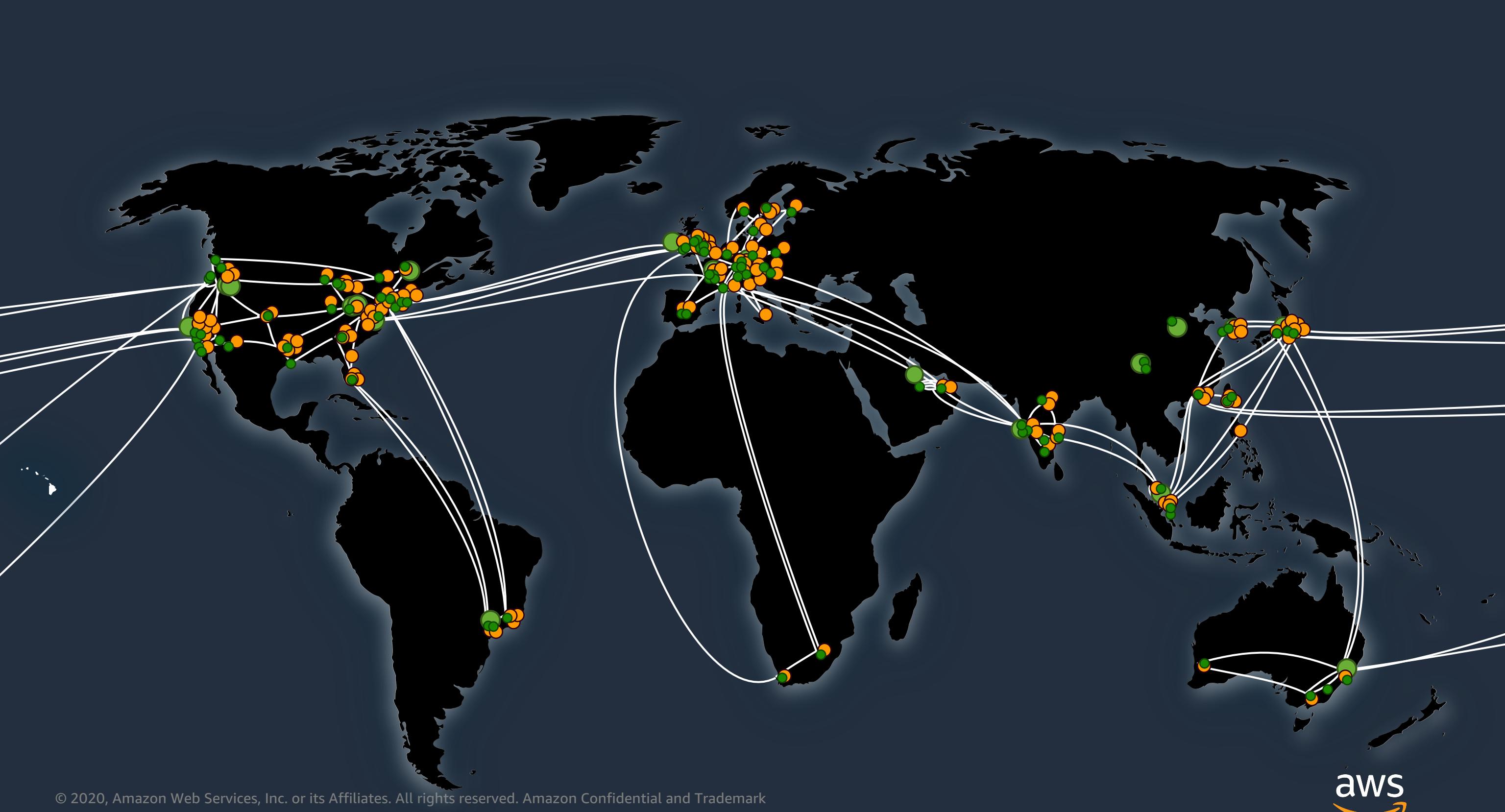
216

Amazon
CloudFront
Points of
Presence



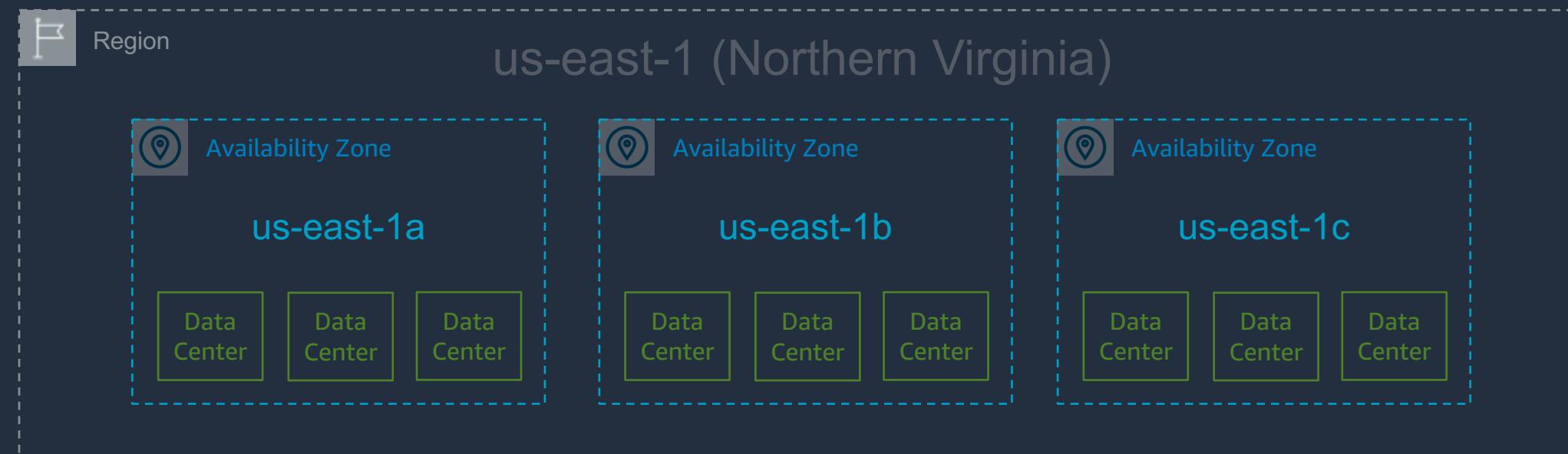
100
AWS Direct
Connect
locations





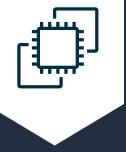
Availability Zones

- A region is comprised of multiple Availability Zones (typically 3)
- Fully independent partitions on isolated fault lines, flood plains, and power grids
- Each AZ: redundant power and redundant dedicated network
- Each AZ: typically multiple data centers
- Between AZs: high throughput, low latency (<10ms) network
- Between AZs: physical separation < 100km (60mi)

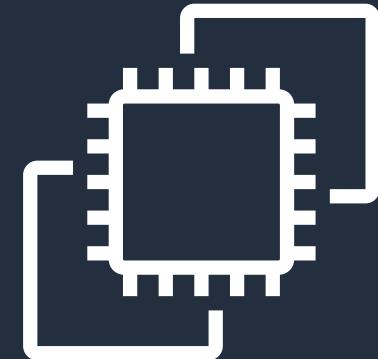


2

EC2 Overview



Choices for Compute



Amazon EC2

Virtual server instances
in the cloud



Amazon ECS, EKS, and Fargate

Container management service
for running
Docker on a managed
cluster of EC2

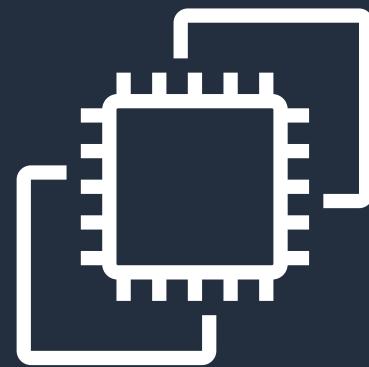


AWS Lambda

Serverless compute
for stateless code execution in
response to triggers



Amazon EC2



Amazon EC2

Linux | Windows

Arm and x86 architectures

General purpose and workload optimized

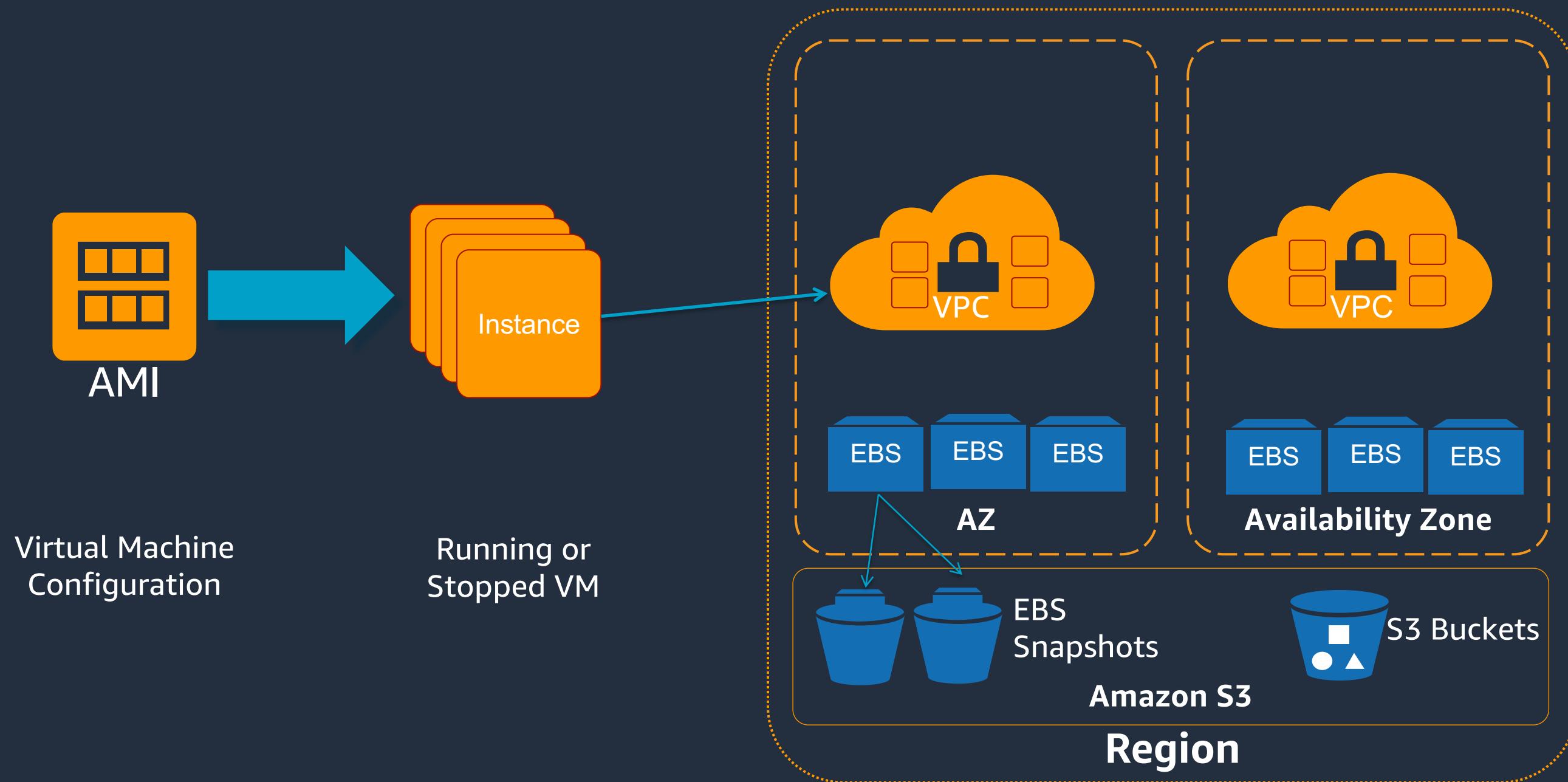
Bare metal, disk, networking capabilities

Packaged | Custom | Community AMIs

Multiple purchase options: On-demand, RI, Spot



EC2 Terminology





What's a virtual CPU? (vCPU)

- A vCPU is typically a hyper-threaded physical core*
- Divide vCPU count by 2 to get core count
- On Linux, "A" threads enumerated before "B" threads
- On Windows, threads are interleaved
- Cores by Amazon EC2 & RDS DB Instance type:
<https://aws.amazon.com/ec2/virtualcores/>

* *CPU Optimizing options allow disabling hyperthreading and reduce number of cores*



Memory and Storage

What's a GiB?

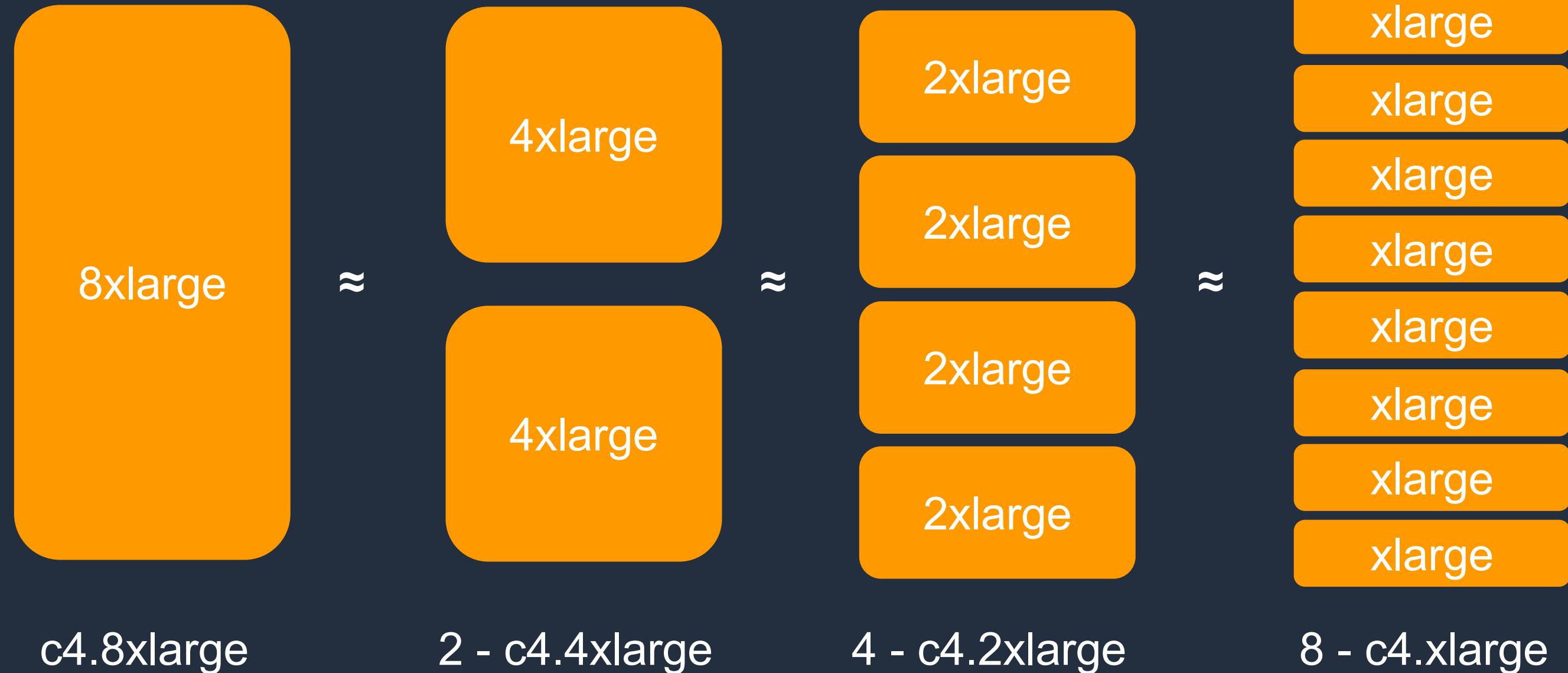
- Memory is presented as GibiBytes (GiB) and not Gigabytes (GB)
- $256 \text{ GiB} = 275 \text{ GB}$

What about storage?

- Storage is independent of compute
- You allocate drives known as EBS volumes
- Max 16 TiB per volume
- Some instance types provide physically attached (ephemeral) storage



Instance sizing





Resource allocation

- All resources assigned to you are dedicated to your instance with no over commitment*
 - All vCPUs are dedicated to you
 - Memory allocated is assigned only to your instance
 - Network resources are partitioned to avoid “noisy neighbors”
- Curious about the number of instances per host?
 - See “Dedicated Hosts Configuration Table” for a guide.

*Again, the “T” family is special



Choose your processor and architecture



Intel® Xeon® Scalable
(Skylake) processor



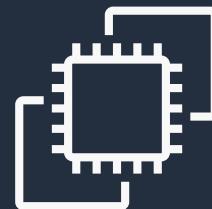
NVIDIA V100
Tensor Core GPUs



AMD EPYC processor



Amazon ARM based
Cloud Processor



FPGAs for custom
hardware acceleration

Right compute for the right application and workload



EC2 Naming Explained

Instance generation

The diagram illustrates the structure of an EC2 instance name, **c5n.xlarge**, which is composed of three main parts:

- Instance family**: The prefix **c5n**.
- Attribute**: The separator **.** between the instance family and instance size.
- Instance size**: The suffix **xlarge**.

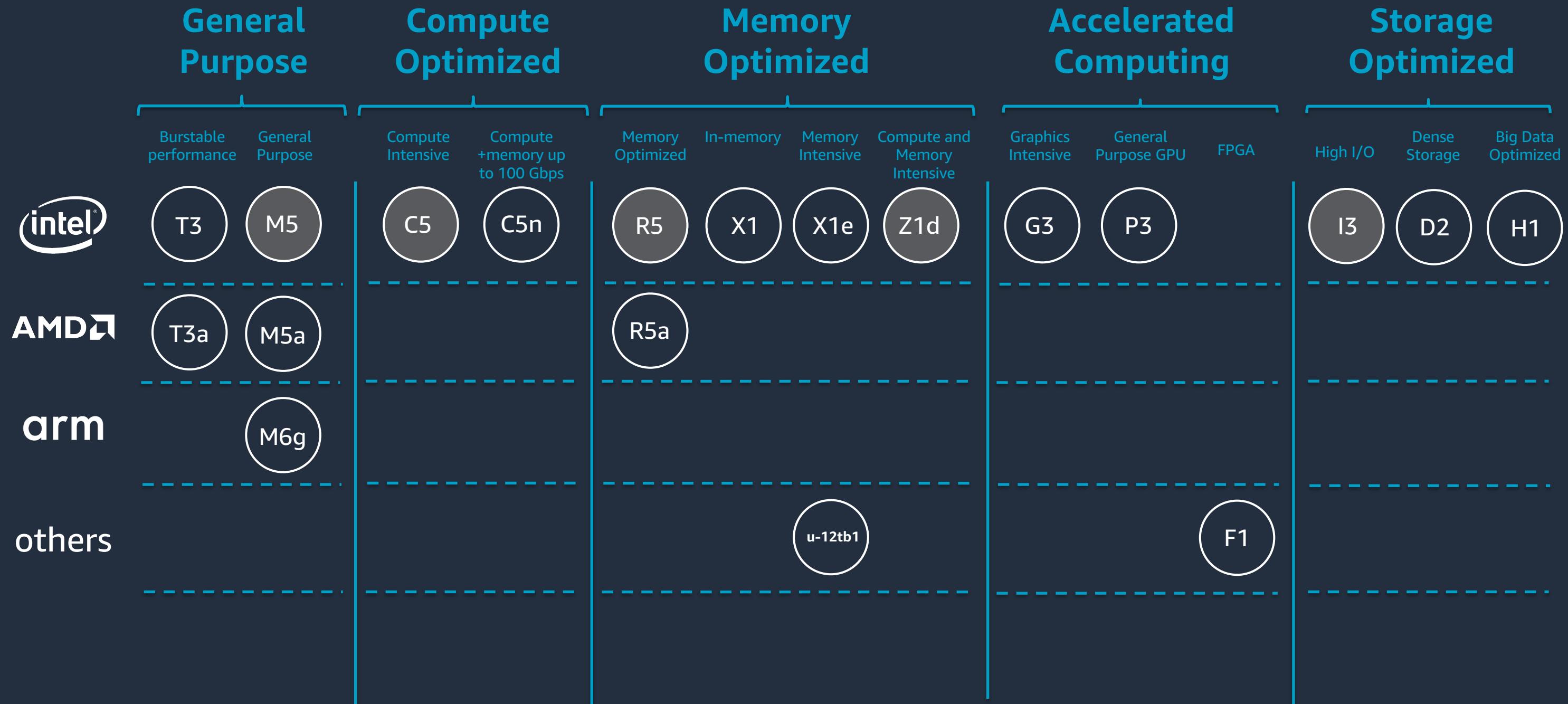
Yellow curly braces above the text group the parts: the first brace groups "c5n" and the separator ".", the second brace groups the separator "." and "xlarge", and the third brace groups "c5n" and "xlarge".

c5n.xlarge

Instance family Attribute Instance size



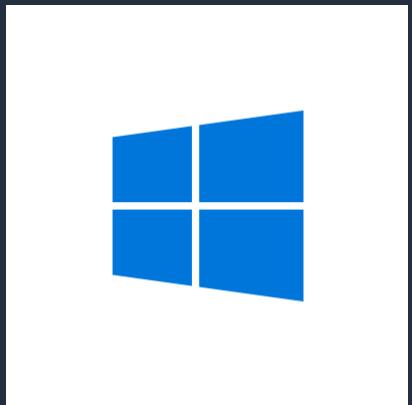
Instance Types





EC2 Operating Systems Supported

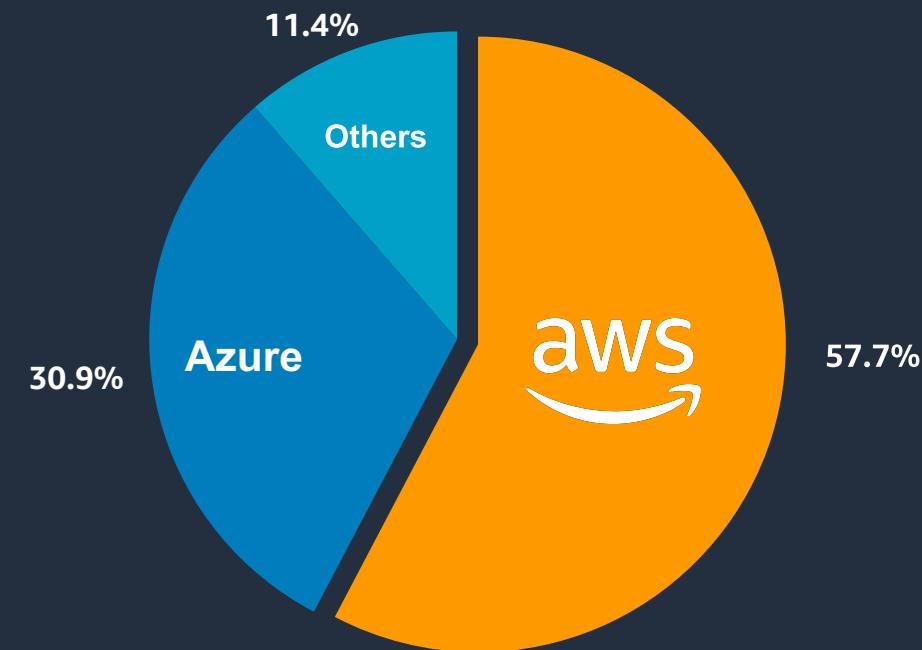
- Windows 2003R2/2008/2008R2/2012/2012R2/2016/2019
- Amazon Linux
- Debian
- Suse
- CentOS
- Red Hat Enterprise Linux
- Ubuntu



for more OSes see: <https://aws.amazon.com/marketplace/b/2649367011>



Windows Licenses by Cloud Provider



Note: Includes Windows instances deployed in the public cloud IaaS market during 2017 Source: IDC estimates, 2018

IDC, Windows Server Operating Environment Market Update, Doc # US44217118, Aug 2018

https://d1.awsstatic.com/analyst-reports/IDC_Slide_WindowsonAWS_JM181015.pdf

What is an Amazon Machine Image (AMI)?



Provides the information required to launch an instance

Launch multiple instances from a single AMI

An AMI includes the following

- A template for the root volume (for example, operating system, applications)
- Launch permissions that control which AWS accounts can use the AMI
- Block device mapping that specifies volumes to attach to the instance



Choosing an AMI

AWS Console

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

Step 1: Choose an Amazon Machine Image (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; or you can select one of your own AMIs.

Quick Start

My AMIs AWS Marketplace Community AMIs

Free tier only

Image	Name	Type	Select	Architecture
Amazon Linux	Amazon Linux 2 AMI (HVM), SSD Volume Type - ami-04681a1dbd79675a5	Free tier eligible	Select	64-bit
Amazon Linux	Amazon Linux AMI 2018.03.0 (HVM), SSD Volume Type - ami-0ff8a9107f77f867	Free tier eligible	Select	64-bit
Red Hat	Red Hat Enterprise Linux 7.5 (HVM), SSD Volume Type - ami-6871a115	Free tier eligible	Select	64-bit

AWS Marketplace

aws marketplace

View Categories ▾ Migration Mapping Assistant Your Saved List Sell in AWS Marketplace Amazon Web Services Home Help

Operating Systems (336 results) showing 1 - 10

1 2 3 4 5 ... 34 ▶

Image	Name	Rating	Version	Sold by
CentOS	CentOS 7 (x86_64) - with Updates HVM	★★★★★ (58)	1805_01	Sold by CentOS.org
CentOS	CentOS 6 (x86_64) - with Updates HVM	★★★★★ (33)	1805_01	Sold by CentOS.org
Debian	Debian GNU/Linux 8 (Jessie)	★★★★★ (86)	Version 8.7	Sold by Debian
CentOS	CentOS 6.5 (x86_64) - Release Media	★★★★★ (55)	Version 6.5 - 2013-12-01	Sold by CentOS.org

Use the AMI ID to launch through the API or AWS Command Line Interface (AWS CLI)

```
aws ec2 run-instances --image-id ami-04681a1dbd79675a5 --instance-type c4.8xlarge --count 10 --key-name MyKey
```

Purchasing Options

On-Demand

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

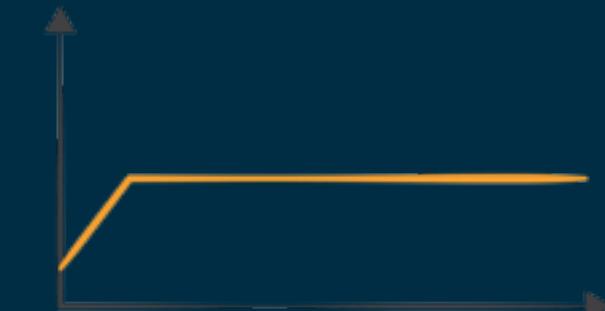
For Spiky workloads or to define needs



Reserved Instances

Make a 1 or 3-year commitment and receive a **significant discount** off On-Demand prices

For committed utilization



Spot Instances

Spare EC2 capacity at **savings of up to 90%** off On-Demand prices

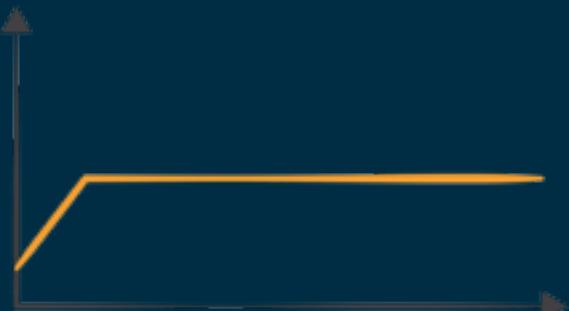
For time-insensitive or transient workloads
Need to be Fault-tolerant, stateless



Savings Plans

Commit to a \$/h spend and **share discount** across compute options and regions

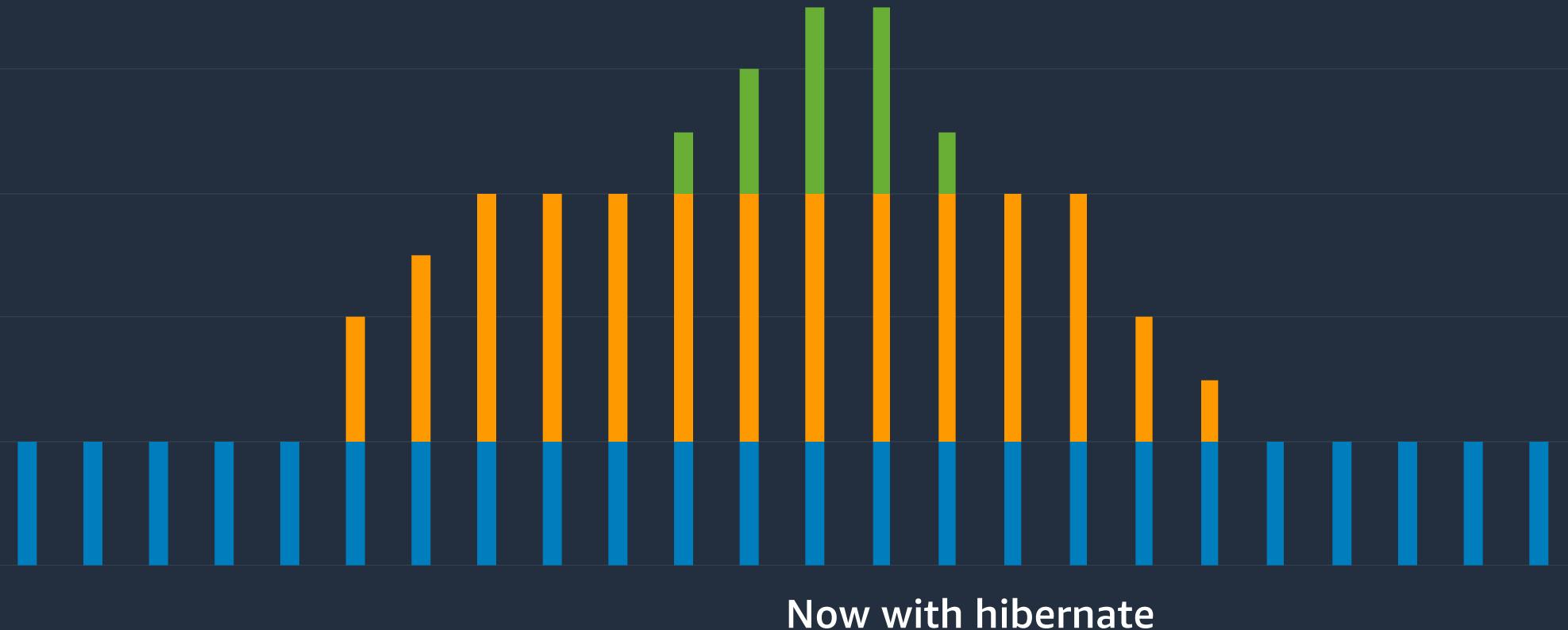
For committed utilization



To optimize EC2, combine all three purchase options!



Simplify capacity and cost optimization



Scale using
Spot,
On-Demand,
or both

Use **Reserved Instances**
for known/steady-state
workloads

AWS services make this easy and efficient



Amazon EC2
Auto Scaling



EC2 Fleet



Amazon Elastic
Container Service



Amazon Elastic
Container Service
for Kubernetes



AWS
Thinkbox



Amazon
EMR



AWS
CloudFormation

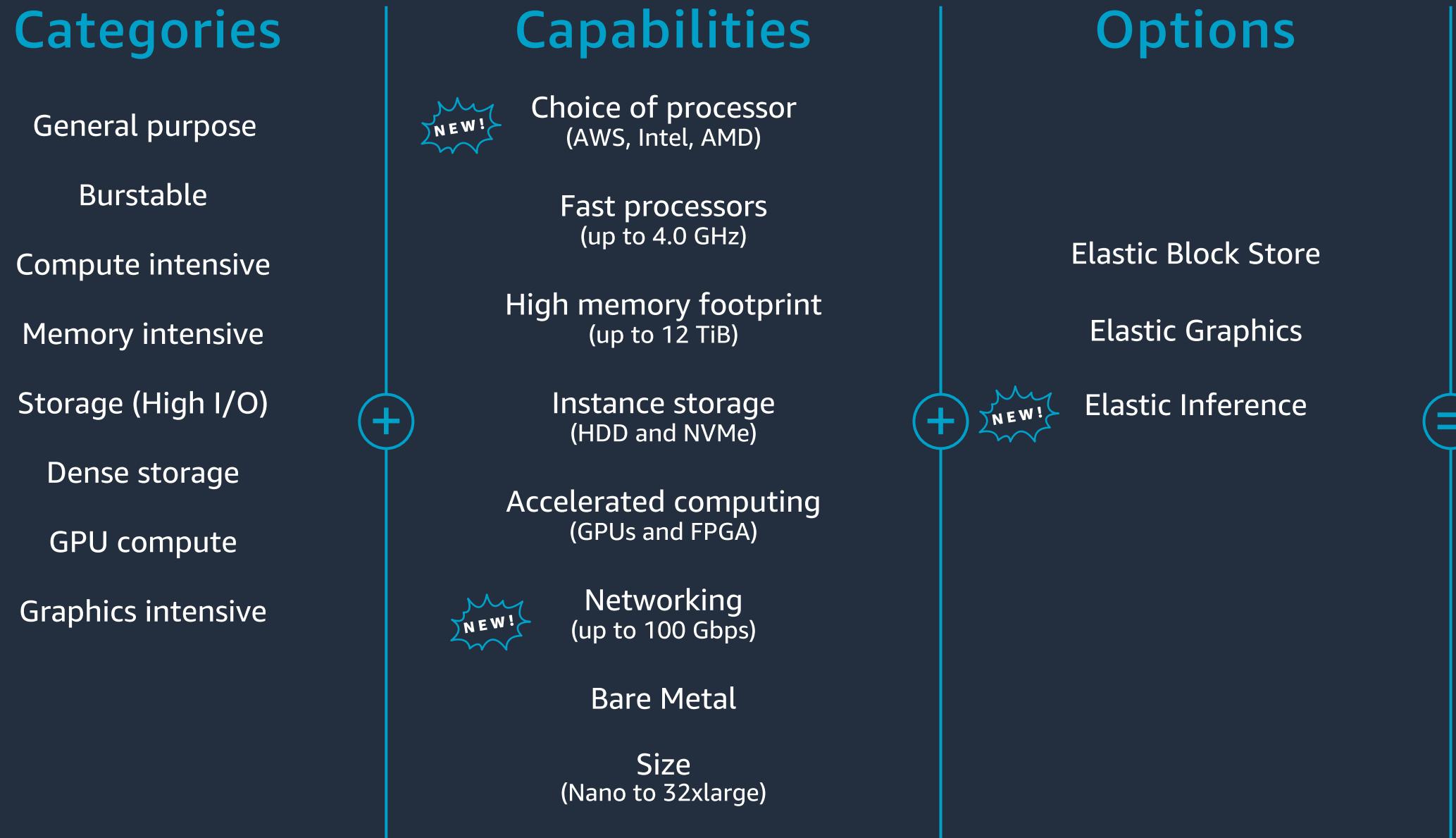


AWS Batch





EC2 Options



200+
instance types
for virtually
every workload
and business need



EC2 Design



Which hypervisor do we use?

Original host architecture: **Xen-based**

- Hypervisor consumed resources from the underlying host
- Limited optimization

AWS Nitro Hypervisor: **Custom KVM based hypervisor**

- AWS Nitro System (launched on Nov 2017)
- Less server resources used, more resources for the customer
- AWS optimized

Bare metal: **Direct access to processor and memory resources**

- Built on the AWS Nitro system
- Enables custom hypervisors and micro-VM runtimes



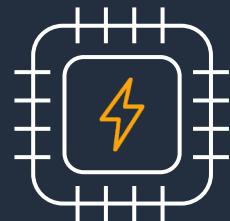
AWS Nitro System

Nitro Card



Local NVMe storage
Elastic Block Storage
Networking, monitoring,
and security

Nitro Security Chip



Integrated into motherboard
Protects hardware resources

Nitro Hypervisor



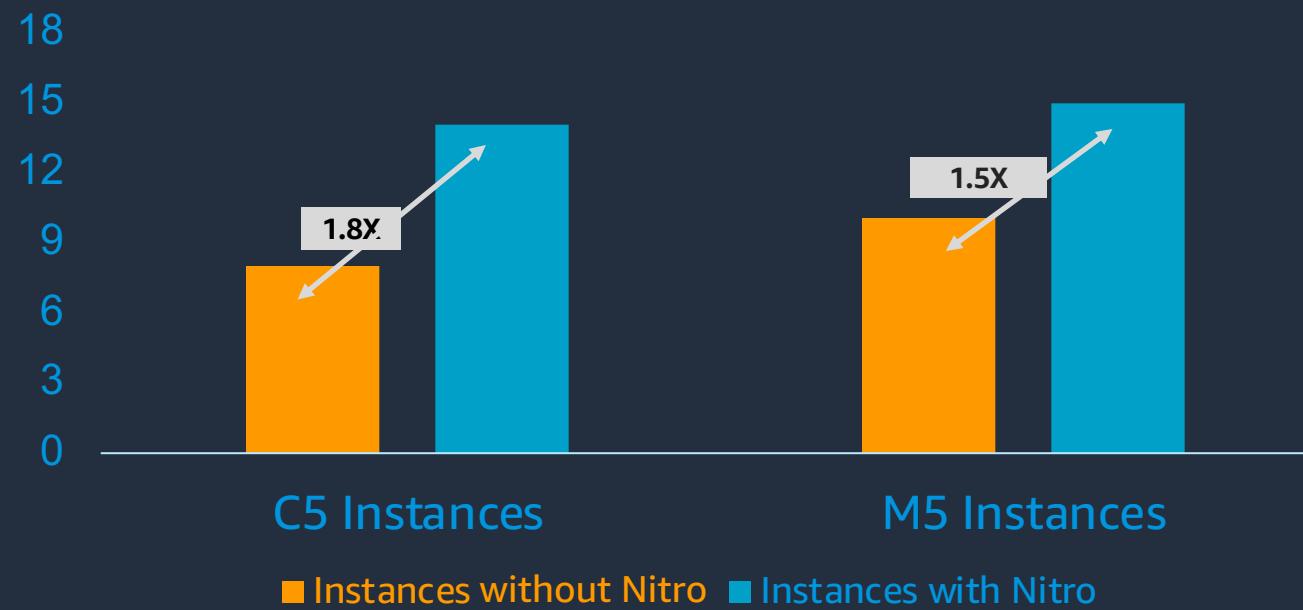
Lightweight hypervisor
Memory and CPU allocation
Bare metal-like performance

Modular building blocks for rapid design and delivery of EC2 instances

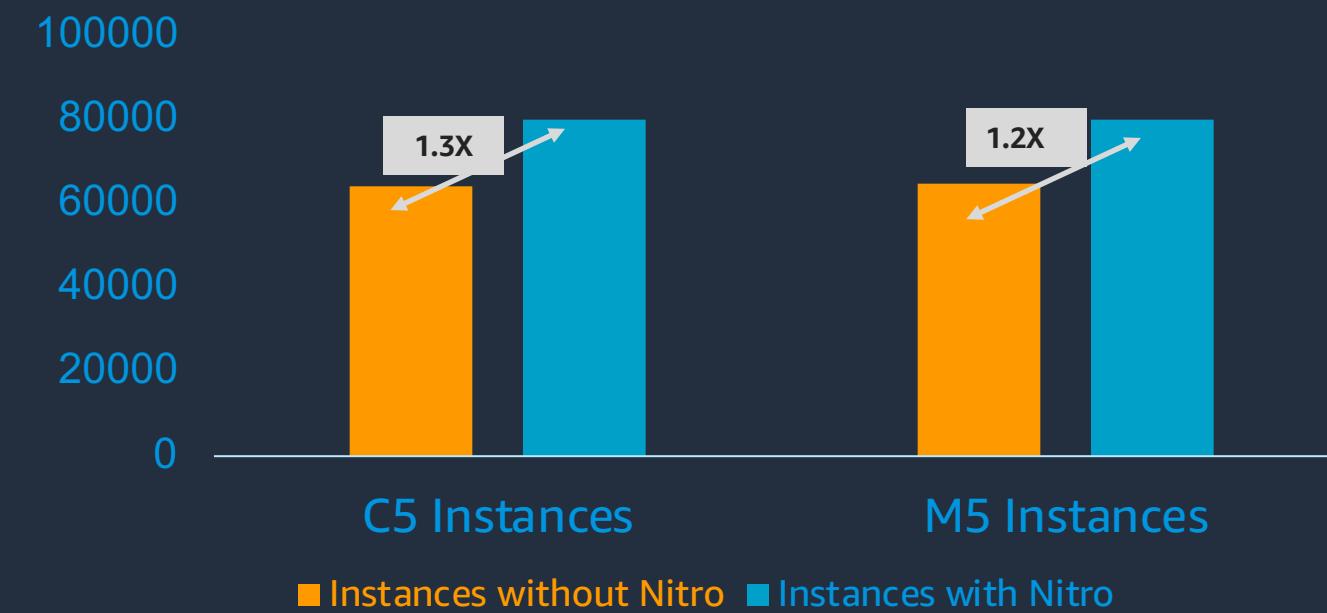


AWS Nitro System

EBS-Optimized Instance Bandwidth



EBS-Optimized Instance IOPS

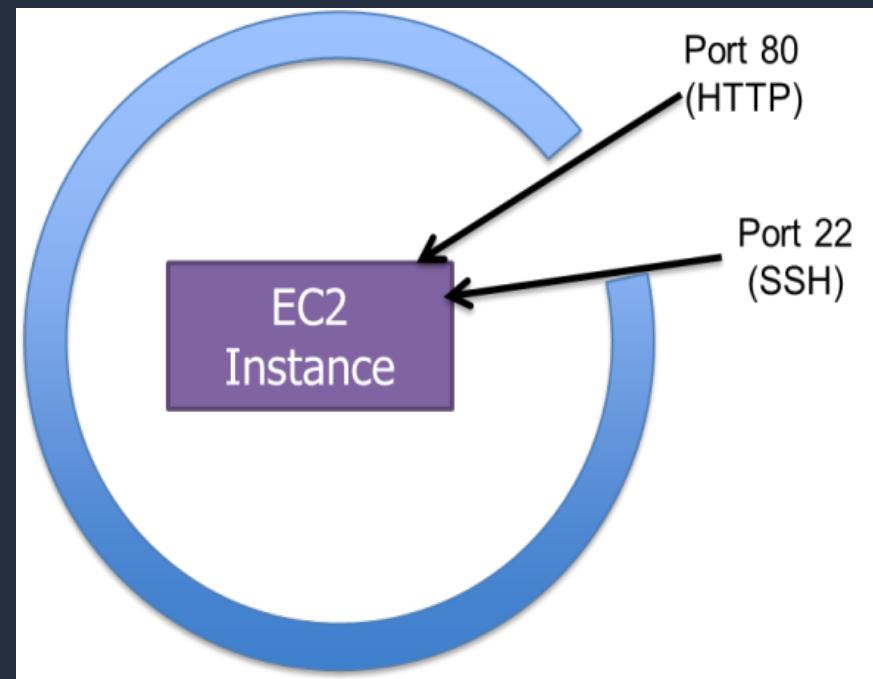


Nitro instances provide **bandwidth, performance, and price improvements** over previous instance generations

EC2 Security Groups

Security Group Rules

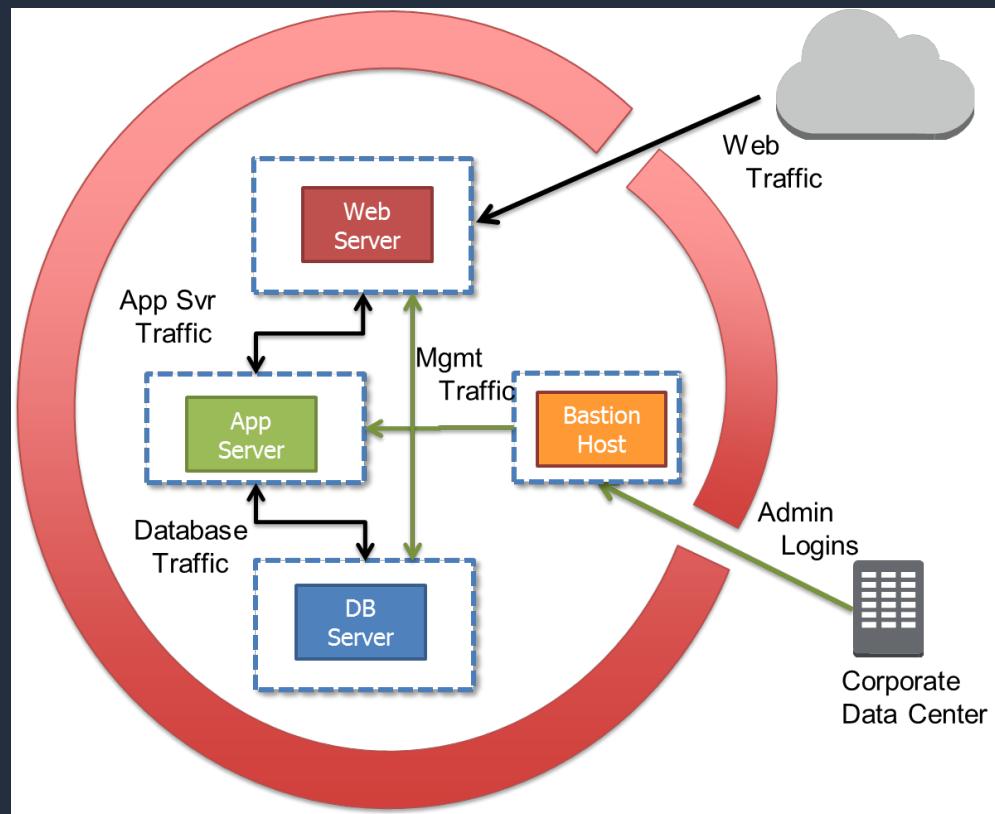
- Name
- Description
- Protocol
- Port range
- IP address, IP range, Security Group name



Tiered EC2 Security Groups

Hierarchical Security Group Rules

- Dynamically created rules
- Based on Security Group membership
- Create tiered network architectures



"Web" Security Group:

TCP 80 0.0.0.0/0

TCP 22 "Mgmt"

"App" Security Group:

TCP 8080 "Web"

TCP 22 "Mgmt"

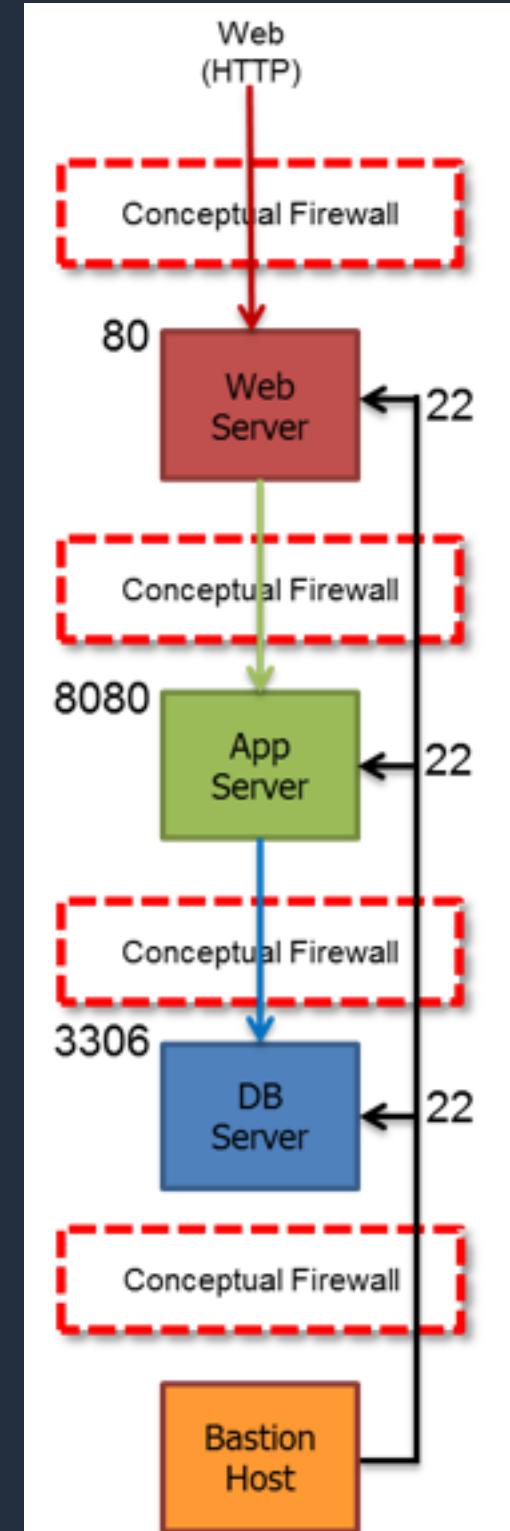
"DB" Security Group:

TCP 3306 "App"

TCP 22 "Mgmt"

"Mgmt" Security Group:

TCP 22 163.128.25.32/32



EC2 IP Addressing

Default VPC	Virtual Private Cloud
Dynamic Private IP	Dynamic or Static Private IP Address
Dynamic Public IP	None by default (can be created with publicIP=true)
Optional Static Public IP (EIP)	Optional Static Public IP (EIP), BYOIP
AWS-provided DNS names <ul style="list-style-type: none">• Private DNS name• Public DNS name	AWS-provided public DNS lookup AWS-provided private DNS names Customer-controlled DNS options

EC2-Specific Credentials

EC2 key pairs

- Linux – SSH key pair for first-time host login
- Windows – Retrieve Administrator password

Standard SSH RSA key pair

- Public/Private Keys
- Private keys are not stored by AWS

AWS approach for providing initial access to a generic OS

- Secure
- Personalized
- Non-generic (NIST, PCI DSS)

“Public Half” inserted by Amazon into each EC2 instance that you launch



“Private Half” downloaded to your desktop

Instance Metadata

<http://169.254.169.254/latest/meta-data/> contains a wealth of info

- ami-id
- ami-launch-index
- ami-manifest-path
- block-device-mapping/
- hostname
- instance-action
- ★ **instance-id**
- instance-type
- kernel-id
- local-hostname
- local-ipv4
- mac
- network/
- ★ **placement/availability-zone**
- profile
- public-hostname
- public-ipv4
- public-keys/



Any Questions?

