



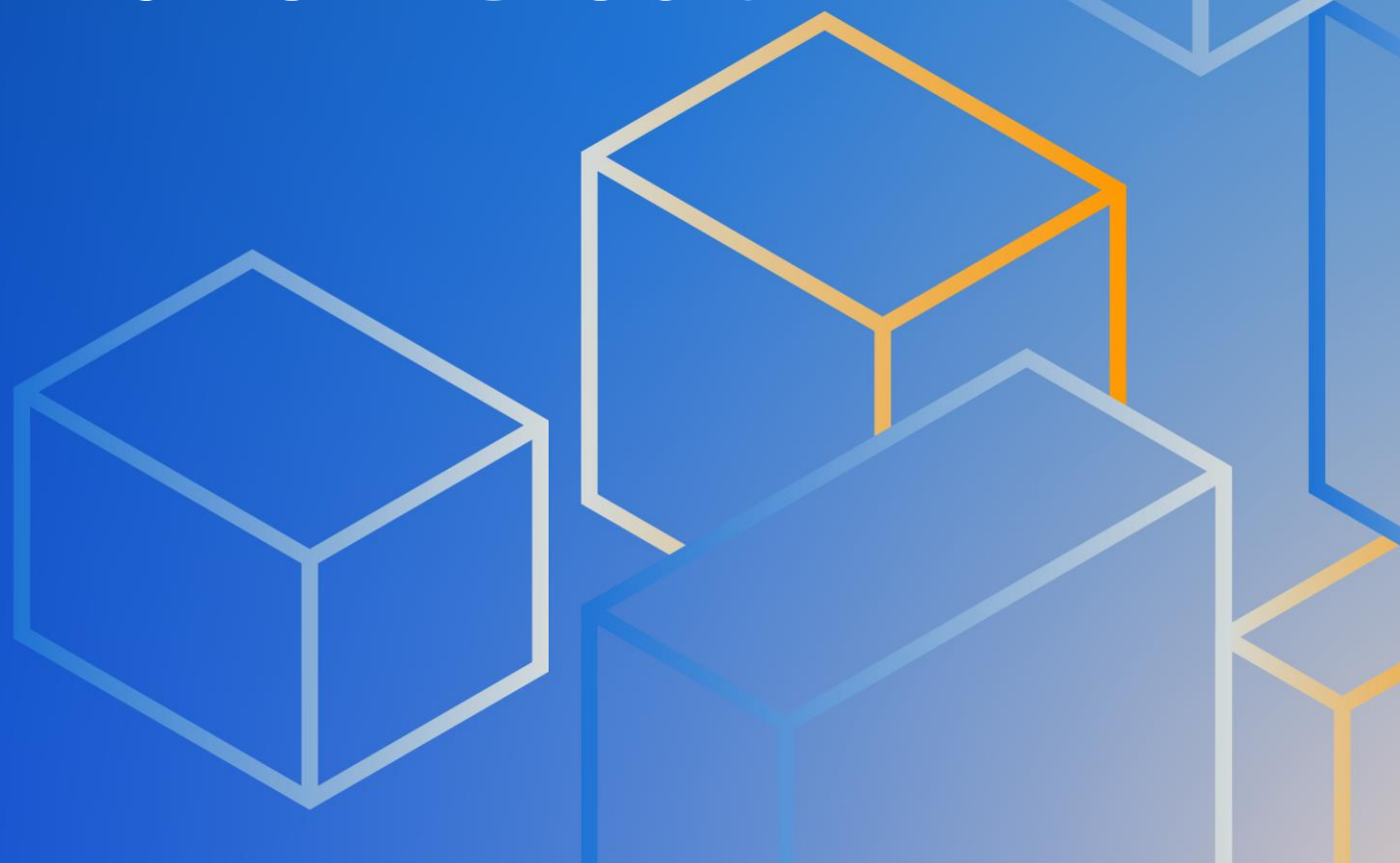
# Introduction to Amazon Cloud

## Amazon EC2 Overview

Noel Mathew

Solutions Architect

3/5/2025



# Agenda

- Introduction to Amazon Cloud
- AWS Global Reach
- Amazon EC2 Overview
- Amazon EC2 Design

# What is AWS?








AWS provides a highly reliable, scalable, low-cost infrastructure platform in the cloud that powers millions of businesses in over 245 countries and territories around the world.

## Benefits

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



# What sets AWS apart?

	Security	Fine-grained identity and access control; Build with the Highest Standards for Privacy and Data Security
	Service Breadth & Depth; Pace of Innovation	200+ fully featured services to support any cloud workload; AWS released 3,332 significant features and services in 2022
	Experience: 1M+ customers	Building and managing cloud since 2006
	Global Footprint	114 Availability Zones within 36 geographic Regions, 35 Local Zones, 550+ Points of Presence and 13 regional edge caches in 100+ cities across 50 countries
	Machine Learning	More machine learning happens on AWS than anywhere else. Machine learning in the hands of every developer and data scientist.
	Ecosystem	140,000+ APN partners from over 200 countries. The AWS Marketplace 12,000+ Products across 65 Categories
	Enterprise leader	AWS positioned as a Leader in the Gartner Magic Quadrant for Cloud Infrastructure and Platform Services



# Pricing Philosophy

High volume / low margin businesses are in our core DNA

Trade capital  
expense for  
variable expense

Pay for what  
you use

Our economies of  
scale provide us  
with lower costs

151 price  
reductions  
since 2006

Pricing model  
choice to support  
variable and  
stable workloads

On-demand  
Savings Plans  
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

“

We were able to easily support the jump from 30 billion records to 70 billion records a day because of the flexibility and scalability of Amazon S3 and Amazon Redshift. ”

– Robert Hunt, Vice President of Software Engineering, Nasdaq



It's **greener** in the cloud.

AWS's infrastructure is

**3.6x more energy efficient**

than the median of the surveyed U.S. enterprise data centers

AWS performs the same task with an

**80% lower carbon footprint**

Source: 451 Research, 2019, all rights reserved



© 2025, Amazon Web Services, Inc. or its affiliates.

# Responsible **water** use

AWS has multiple initiatives to improve our **water use efficiency** for cooling data centers:

- Evaporative cooling
- Reduce potable water usage
- Recycled cooling water
- Invest in reclaimed water infrastructure

AWS has committed to being **water positive** by 2030. Our four pillars:

- Water efficiency
- Sustainable Sources
- Water reuse in communities
- Water replenishment



# AWS Global Reach

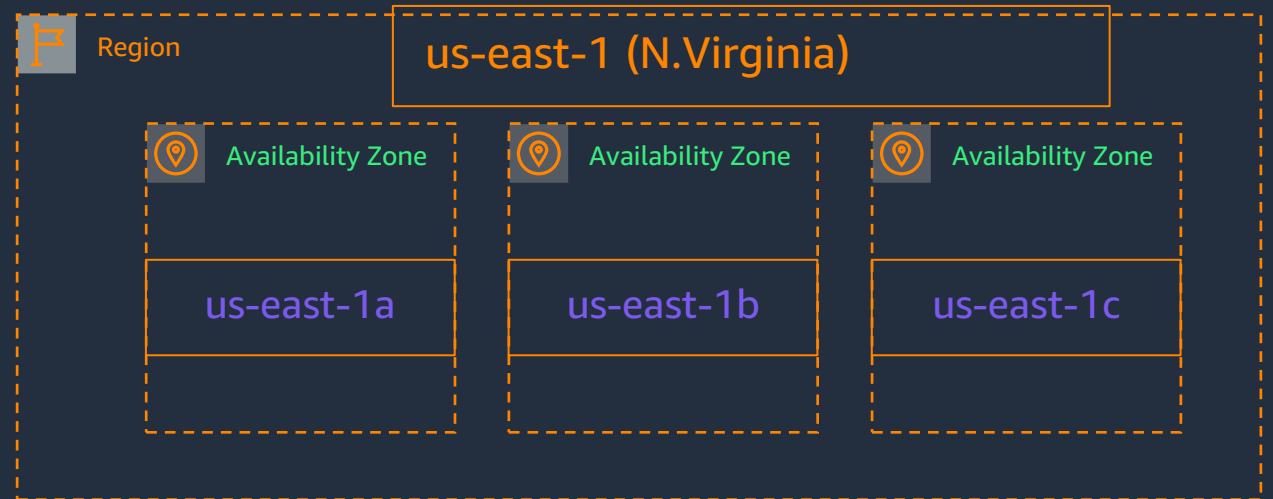


36  
Regions



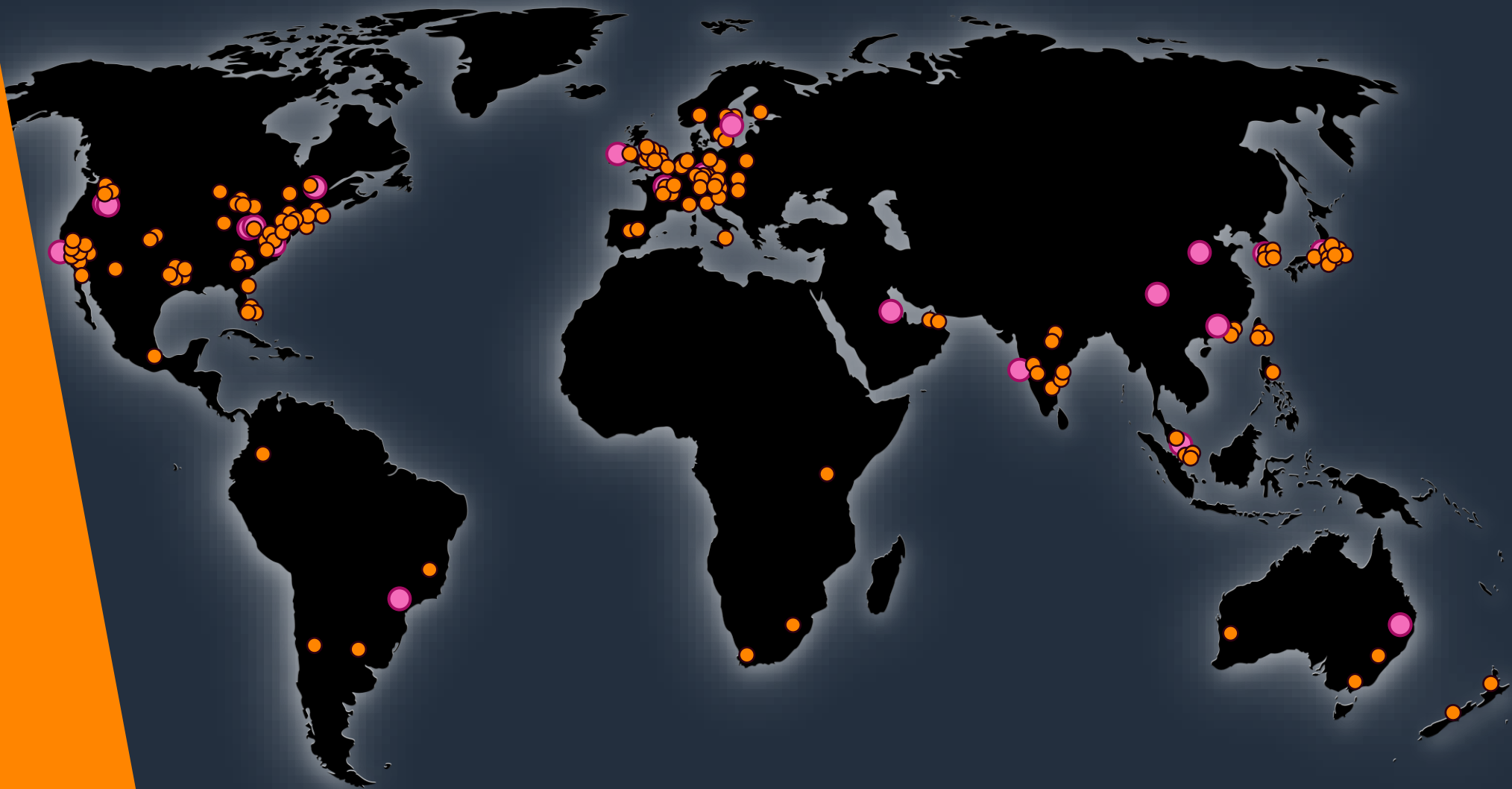
# Availability Zones

- Each AWS Region consists of multiple, isolated, and physically separate AZs within a geographic area
- An Availability Zone (AZ) is one or more discrete data centers with redundant power, networking, and connectivity in an AWS Region
- High throughput, low latency (< 10 ms) network between Availability Zones
- All traffic between AZs is encrypted
- Physical separation with-in 100 km (60 miles)



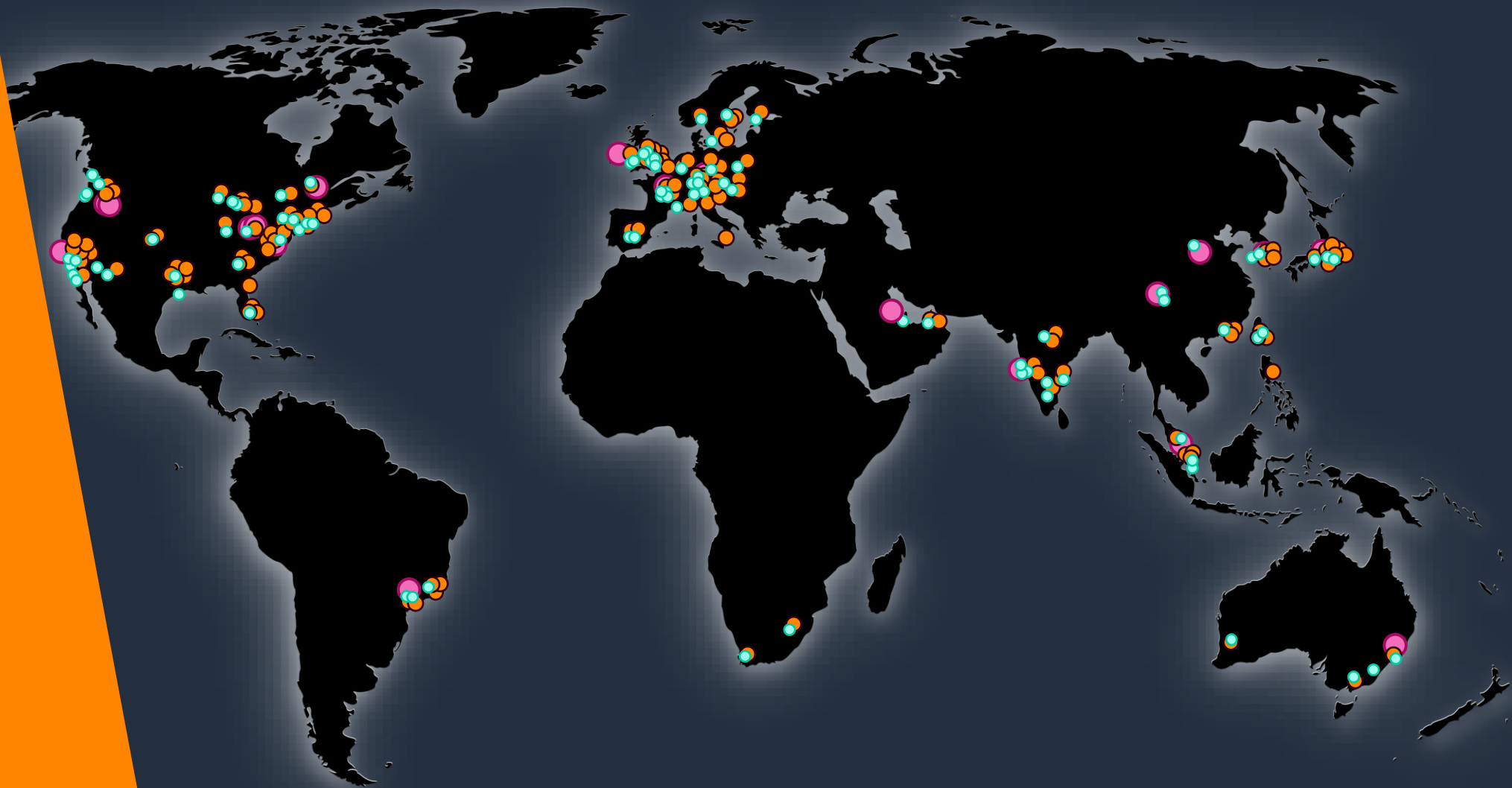
600+

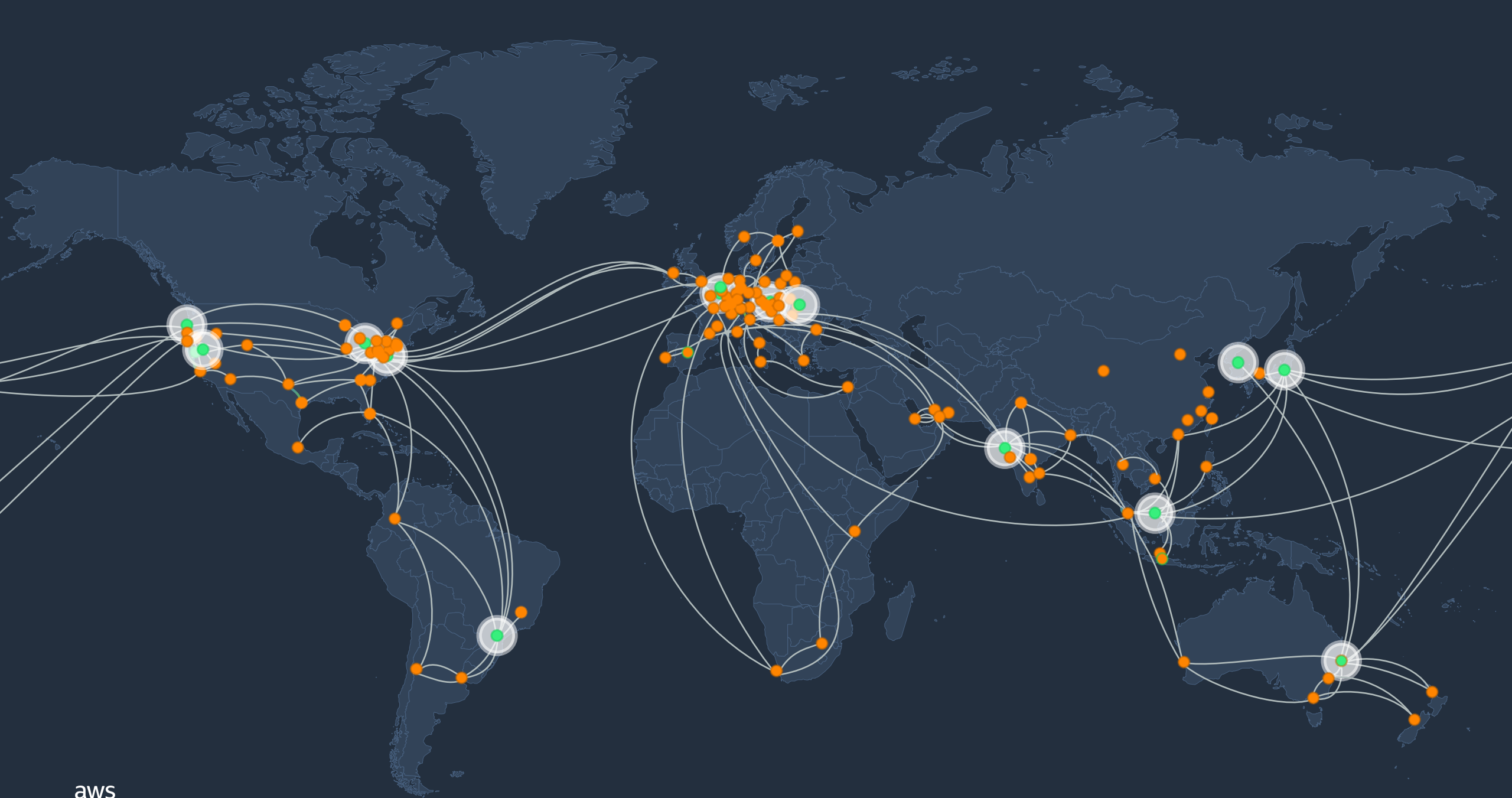
Amazon  
CloudFront  
Points of  
Presence



# 135

AWS Direct  
Connect  
locations





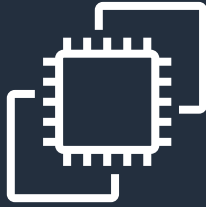
# Amazon EC2 Overview





# Choices for Compute

World-class performance, security, and innovation



## 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

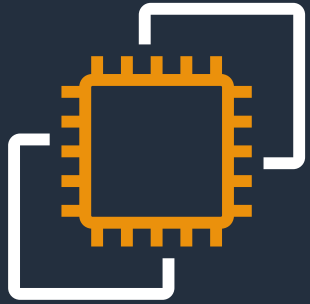


## AWS LAMBDA

Serverless compute  
for stateless code execution  
in response to triggers



# Amazon Elastic Compute Cloud (Amazon EC2)



**AMAZON EC2**

Linux | Windows | Mac

Arm and x86 architectures

General purpose and workload optimized

Bare metal, disk, networking capabilities

Packaged | Custom | Community AMIs

Multiple purchase options: On-Demand, Spot instances, Reserved Instances, Savings Plans, Dedicated Hosts

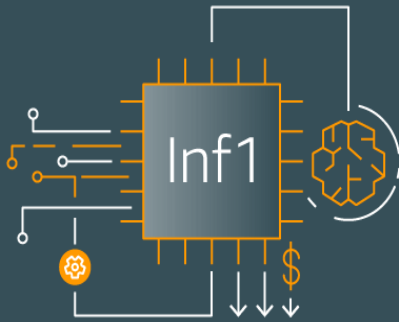
# Instance Types

	General Purpose		Compute Optimized		Memory Optimized				Accelerated Computing			Storage Optimized		
	Burstable performance	General Purpose	Compute Intensive	Compute + network up to 100 Gbps*	Memory Optimized	In-memory	Memory Intensive	Compute and Memory Intensive	Graphics Intensive	General Purpose GPU	FPGA	High I/O	Dense Storage	Big Data Optimized
intel  Local storage (NVMe SSD)	T3	M5	C5	C5n	R5	X1	X2iedn		G3	P2	F1	I3en	D3	H1
		M5d	C5d		R5d			Z1d				I3		
AMD	T3a	M5a			R6a				G5					
metal		M5	C5		R5	u-24tb1		Z1d				I3		
AWS Graviton	T4g	M7g	C7g	C7gn	R7g	X2gd			G5g			Im4gn		



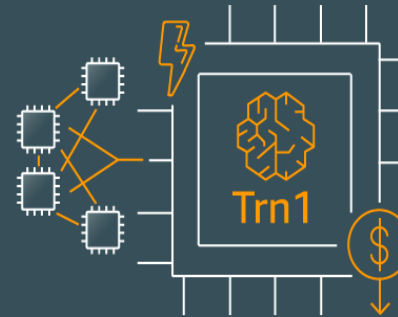
# AWS chips optimized for deep learning

## AWS Inferentia



Lowest cost inference in the cloud for running deep learning models—up to 70% lower cost than GPU instances

## AWS Trainium



The most cost-efficient high performance DL training instance

# Instance Naming

Instance generation

c7gn.xlarge

Instance  
family

Attribute(s)

Instance size

# Instance Sizing



# Choose your processor and architecture



Intel® Xeon® Scalable  
(Skylake) processor



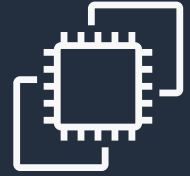
NVIDIA B100  
Tensor Core GPUs



AMD EPYC processor



AWS Graviton  
Processor (arm)



FPGAs for custom  
hardware acceleration

**Right compute for the right application and workload**



# AWS Graviton Processor

Enabling the best price/performance for your cloud workloads

## Graviton2 Processor



7x performance, 4x compute cores, and 5x faster memory



Built with 64-bit Arm Neoverse cores with AWS-designed silicon using 7 nm manufacturing technology

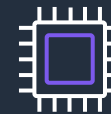


Up to 64 vCPUs, 25 Gbps enhanced networking, 19 Gbps EBS bandwidth

## Graviton3/3E Processor



25% higher performance, 2x higher floating-point performance, 2x faster cryptographic performance

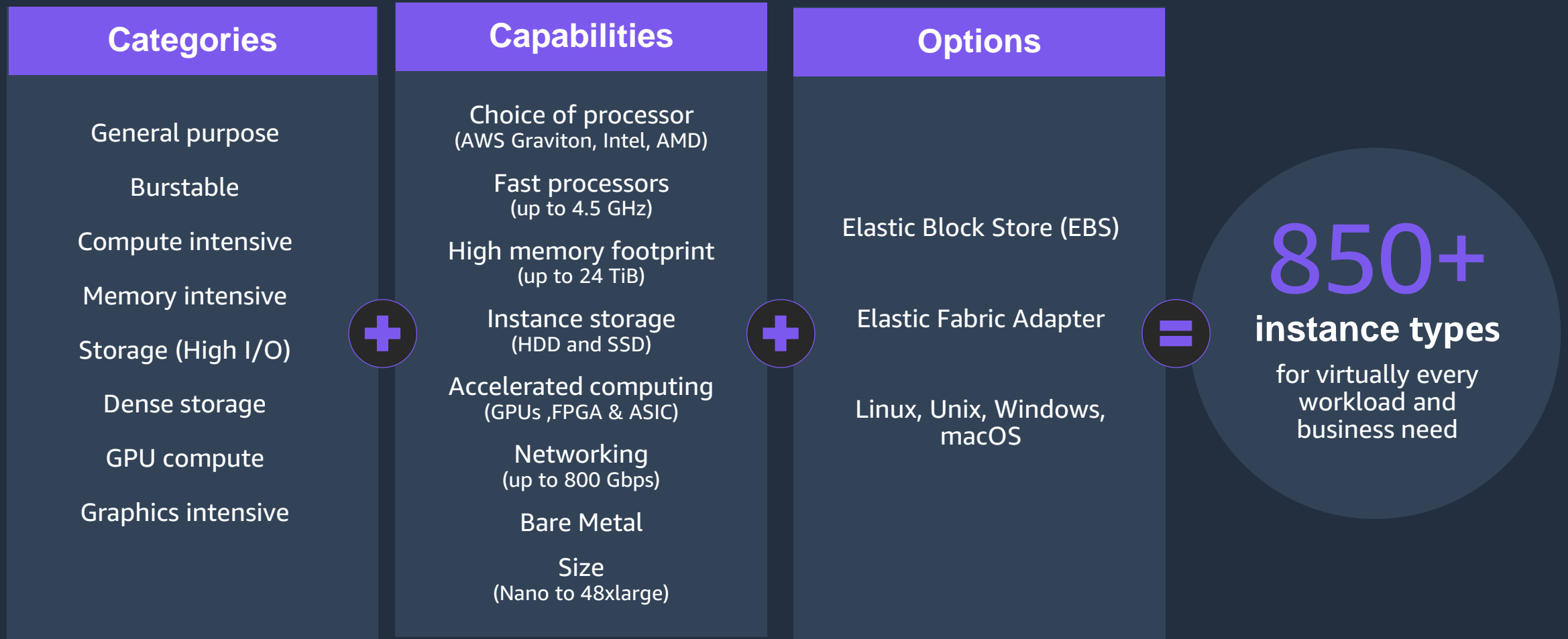


DDR5 memory provides 50% more memory bandwidth compared to DDR4



Support for bfloat16 and delivers up to 3x better performance for ML workloads

# Broadest and deepest platform choice





# Memory and Storage

## What's a GiB?

- Memory is presented as GibiBytes (GiB) and not Gigabytes (GB)
- 256 GiB = 275 GB

## What about storage?

- Storage is independent of compute
- You allocate drives known as Amazon Elastic Block Store (EBS) volumes
- Amazon EBS volumes support up to 64 TiB per volume
- Some instance types provide physically attached (ephemeral) storage

# EC2 Operating Systems

- Windows Server 2012/2012 R2/2016/2019/2022
- Amazon Linux (NEW: Amazon Linux 2023)
- Debian
- SUSE
- CentOS
- Red Hat Enterprise Linux (RHEL)
- Ubuntu
- Mac, including M1 Mac instances

Visit the [AWS Marketplace](#) for more Operating Systems



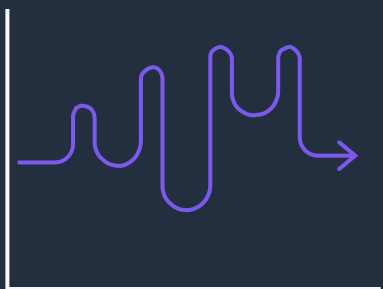
# What is an Amazon Machine Image (AMI)?

- Provides the information required to launch an instance
- Launch multiple instances from a single AMI with the same configuration
- An AMI includes the following:
  - One or more Amazon Elastic Block Store (Amazon EBS) snapshots, or a template for the root volume (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

# Amazon EC2 purchase options

## On-Demand

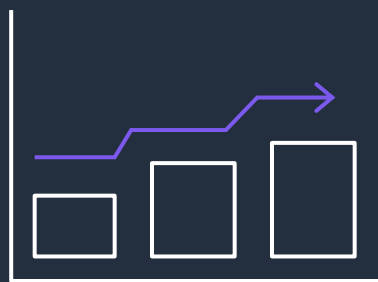
Pay for compute capacity by **the second** with no long-term commitments



Spiky workloads,  
to define needs

## Reserved Instances

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



Committed and  
steady-state usage

## Savings Plans

Same great discounts as Amazon EC2 RIs with **more flexibility**



Committed flexible  
access to compute

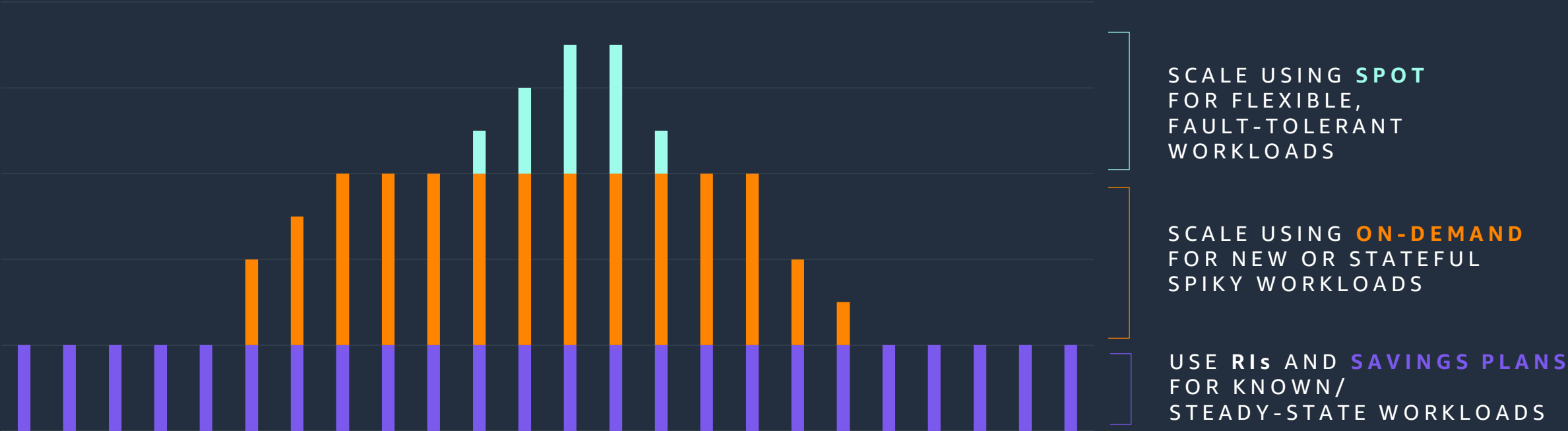
## Spot Instances

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



Fault-tolerant, flexible,  
stateless workloads

# Simplifying capacity and cost optimization



AWS services make this easy and efficient



Amazon EC2  
Auto Scaling



EC2 Fleet



Amazon Elastic  
Container Service  
(Amazon ECS)



Amazon Elastic  
Kubernetes Service  
(Amazon EKS)



AWS  
Thinkbox



Amazon  
EMR



AWS  
CloudFormation



AWS Batch



# EC2 Security Groups

- Virtual firewall
- Security Group Rules
  - Security Group name
  - Description
  - Protocol
  - Port range
  - IP address, IP range

**Basic details**

Security group name [Info](#)  
MyWebServerGroup  
Name cannot be edited after creation.

Description [Info](#)  
Security for Production Web Server

VPC [Info](#)  
Q

---

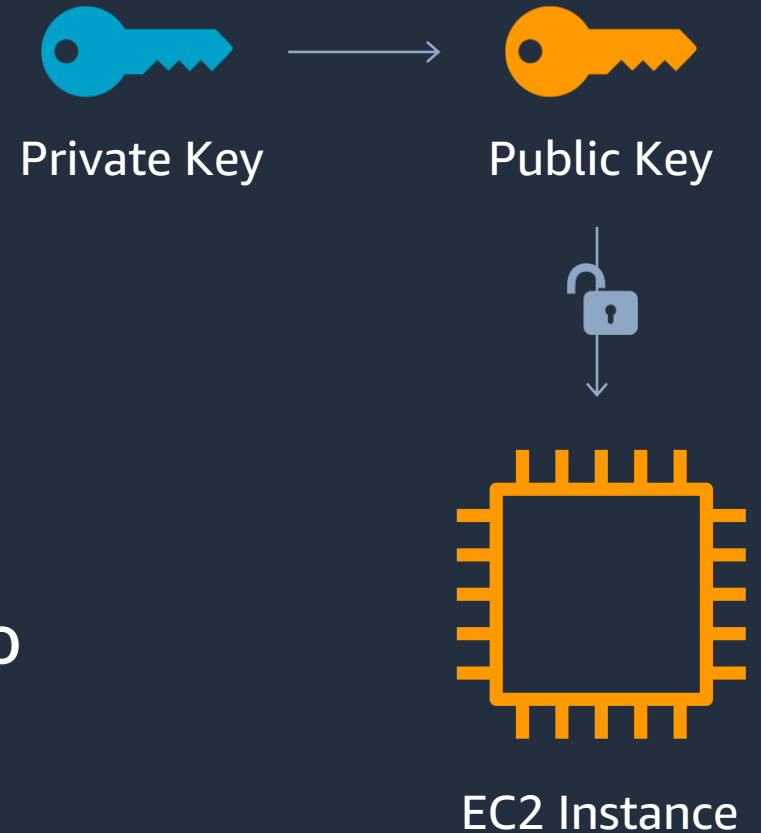
**Inbound rules** [Info](#)

Type <a href="#">Info</a>	Protocol <a href="#">Info</a>	Port range <a href="#">Info</a>	Source <a href="#">Info</a>	Description - optional <a href="#">Info</a>	
SSH ▼	TCP	22	Custom ▼ Q 1.2.3.4/32 ✕	Admin access	Delete
HTTP ▼	TCP	80	Anywhere-I... ▼ Q 0.0.0.0/0 ✕	Web traffic	Delete
HTTPS ▼	TCP	443	Anywhere-I... ▼ Q 0.0.0.0/0 ✕	Secure web traffic	Delete

Add rule

# 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)



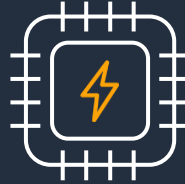
# AWS Nitro System

## 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 **Amazon EC2** instances





# Thank you!

Noel Mathew

[nomath@amazon.com](mailto:nomath@amazon.com)

