AWS Cloud 101 for DoD

Austin Buckingham

Solutions Architect, Air Force

Why customers choose AWS

EXPERIENCE

17 years

helping millions of customers

GLOBAL REACH

32 Commercially available regions (and more for DoD/IC) spanning 102 Availability Zones and over 125 Direct Connect locations

SECURITY

300+

security features

INNOVATION

200+

service offerings

AWS INFRASTRUCTURE

3.6x+

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

TCO

111

price reductions since 2006

ECOSYSTEM

4500

software listings from 1,400 ISVs

The **most secure, extensive, and reliable** Global Cloud Infrastructure

Infrastructure allows you to run workload

You have the same access and capabilities no matter where you are

200+ fully featured services from data centers globally

Understanding the AWS Global Infrastructure

REGION

A physical location where we cluster data centers. Each Region consists of a minimum of three isolated, and physically separate AZ's.

AVAILABILITY ZONE (AZ)

One or more discrete data centers with redundant power, networking, and connectivity in an AWS Region.

LOCAL ZONES

AWS Local Zones place compute, storage, database, and other AWS services closer to end-users so you can run latency-sensitive applications and meet data residency requirements in more locations

POINTS OF PRESENCE

Reduce latency by delivering data through 450+ globally dispersed Points of Presence (PoPs) with automated network mapping and intelligent routing using Amazon CloudFront.

WAVELENGTH ZONES

AWS infrastructure deployments that embed AWS services within telecommunications providers' data centers at the edge of the 5G network.

DIRECT CONNECT LOCATIONS

The Direct Connect service establishes a private, physical network connection between AWS and your data center or office environment – bypassing the public internet.

OUTPOSTS

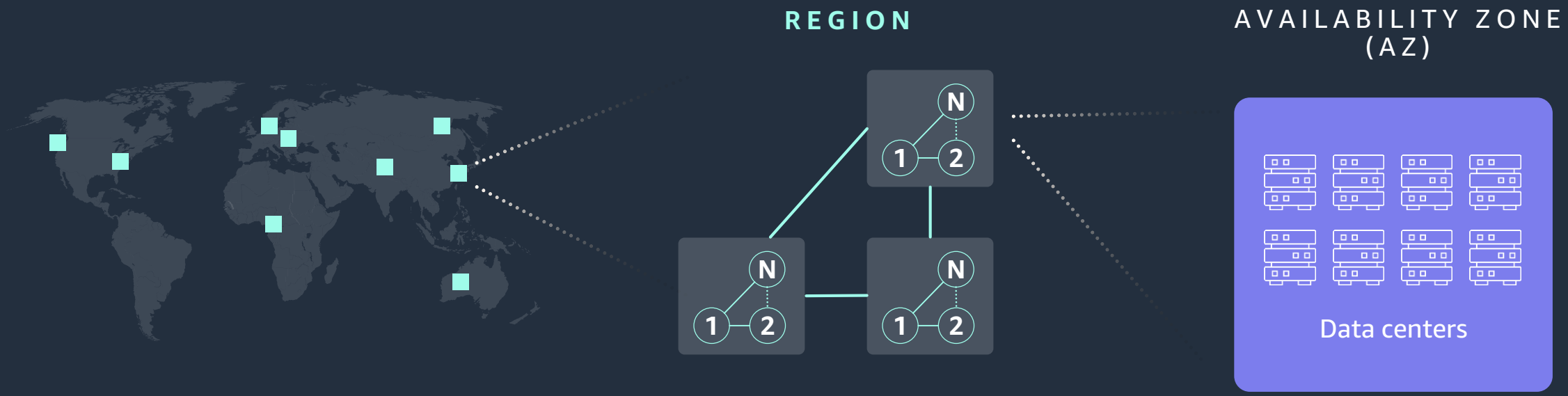
AWS Outposts are a family of fully managed solutions delivering AWS infrastructure and services to virtually any on-premises or edge location for a truly consistent hybrid experience.

Snow Family

The AWS Snow Family supports intense edge compute and storage workloads in the most austere environments, to include Denied, Degraded, Intermittent, and Latent (DDIL).

AWS Region design

AWS Regions are comprised of multiple AZs for high availability, high scalability, and high fault tolerance. Applications and data are replicated in real time and consistent in the different AZs.



Regional Expansion

32 total regions with five coming soon

2006-2011

8 New Regions

N. AMERICA
AWS GovCloud West
Oregon
N. California
N. Virginia

S. AMERICA
São Paulo

EUROPE
Ireland

ASIA PACIFIC
Singapore
Tokyo

2012-2017

8 New Regions

N. AMERICA
Ohio
Canada Central

EUROPE
Frankfurt
London
Paris

ASIA PACIFIC
Seoul
Mumbai

AUSTRALIA &
NEW ZEALAND
Sydney

2018-2023

14 New Regions

N. AMERICA
AWS GovCloud East

EUROPE
Milan
Stockholm
Spain
Zurich

MIDDLE EAST
Bahrain
UAE
Israel

ASIA PACIFIC
Hong Kong
Osaka
Jakarta
Hyderabad
Beijing*
Ningxia*

AFRICA
Cape Town

AUSTRALIA &
NEW ZEALAND
Melbourne

COMING SOON

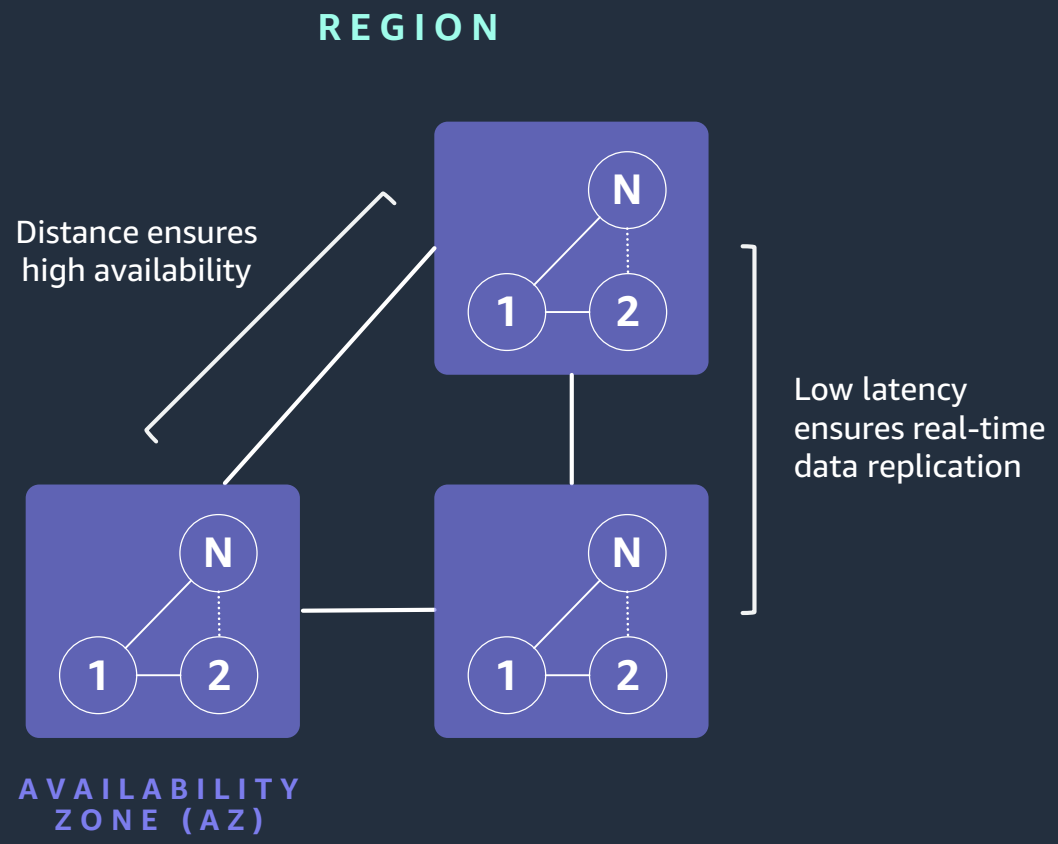
4 New Regions

N. AMERICA
Canada West

ASIA PACIFIC
Thailand
Malaysia

AUSTRALIA &
NEW ZEALAND
Auckland

Availability zones design for resiliency



100K+ servers at scale

ISOLATED PARTITION

Fully isolated with one or more datacenters

POWER

Highly available, fault tolerant, and scalable

DISTANCE

Physically separated by a meaningful distance – all within 60 miles (100km) of each other

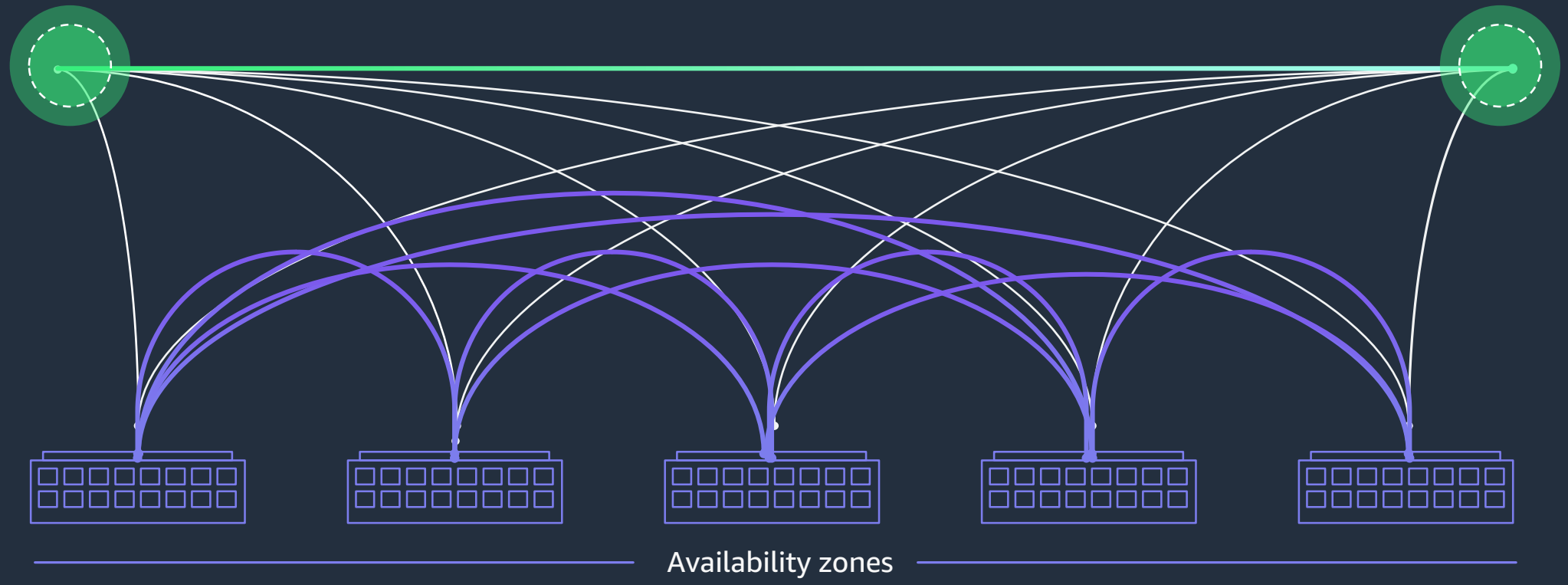
INTERCONNECTION

Datacenters connected via fully redundant and isolated metro fiber

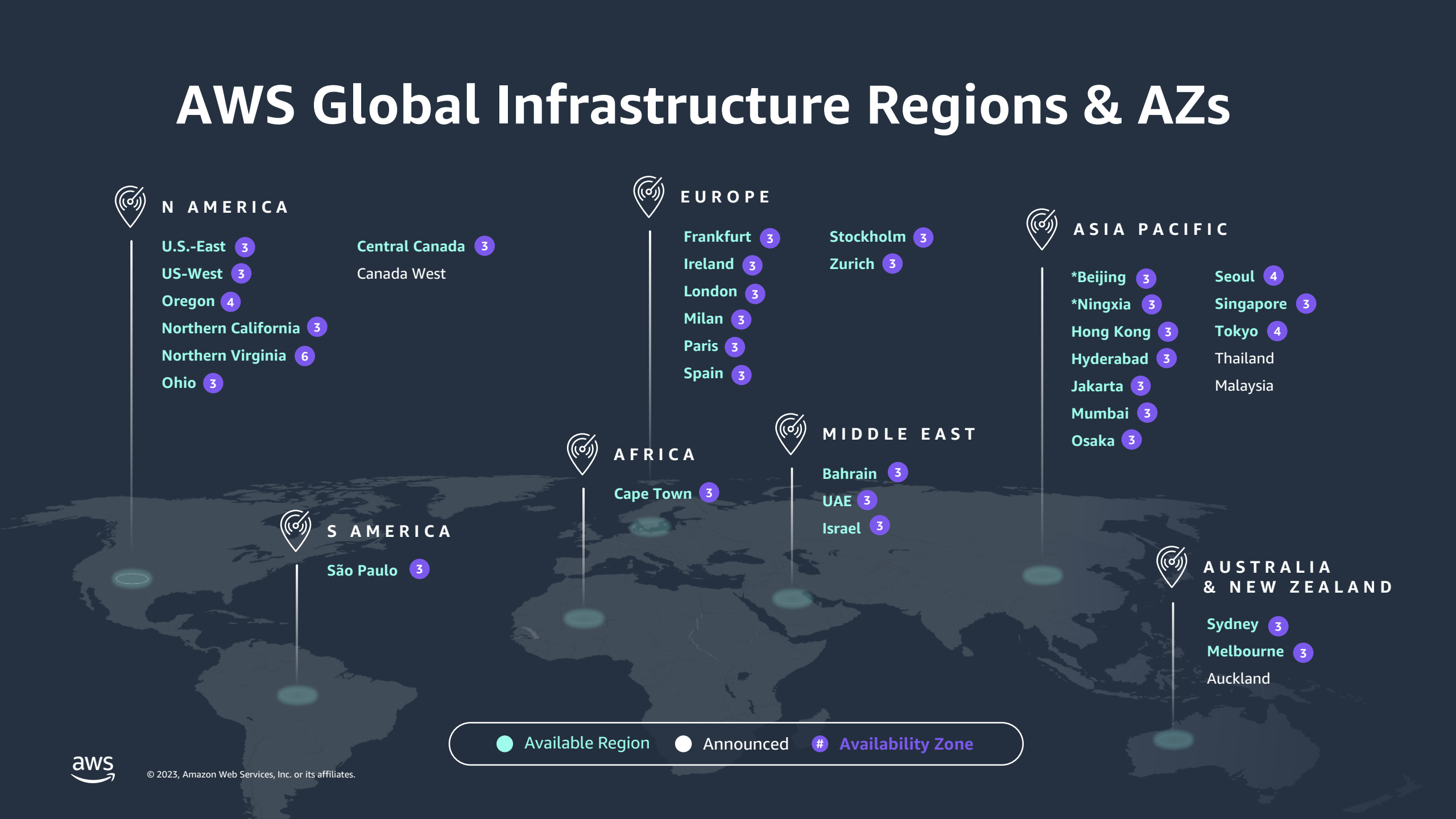
AWS network design

At least 2 redundant transit centers

Highly peered & connected

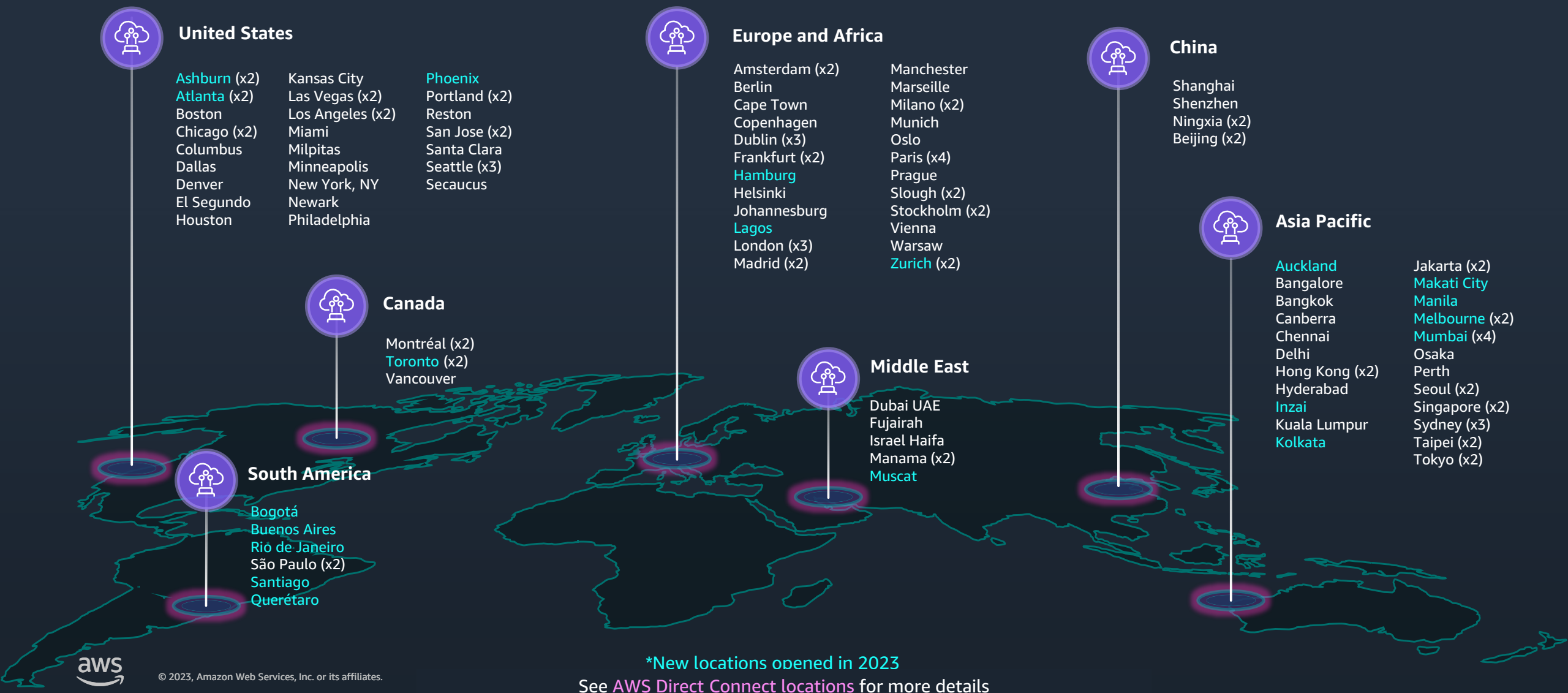


AWS Global Infrastructure Regions & AZs



AWS Direct Connect

Reach every AWS Region, GovCloud, and Local Zones from over 125 locations worldwide (excluding China)



United States

Ashburn (x2)
Atlanta (x2)
Boston
Chicago (x2)
Columbus
Dallas
Denver
El Segundo
Houston
Kansas City
Las Vegas (x2)
Los Angeles (x2)
Miami
Milpitas
Minneapolis
New York, NY
Newark
Philadelphia
Phoenix
Portland (x2)
Reston
San Jose (x2)
Santa Clara
Seattle (x3)
Secaucus

Canada

Montréal (x2)
Toronto (x2)
Vancouver

South America

Bogotá
Buenos Aires
Rio de Janeiro
São Paulo (x2)
Santiago
Querétaro

Europe and Africa

Amsterdam (x2)
Berlin
Cape Town
Copenhagen
Dublin (x3)
Frankfurt (x2)
Hamburg
Helsinki
Johannesburg
Lagos
London (x3)
Madrid (x2)
Manchester
Marseille
Milano (x2)
Munich
Oslo
Paris (x4)
Prague
Slough (x2)
Stockholm (x2)
Vienna
Warsaw
Zurich (x2)

Middle East

Dubai UAE
Fujairah
Israel Haifa
Manama (x2)
Muscat

China

Shanghai
Shenzhen
Ningxia (x2)
Beijing (x2)

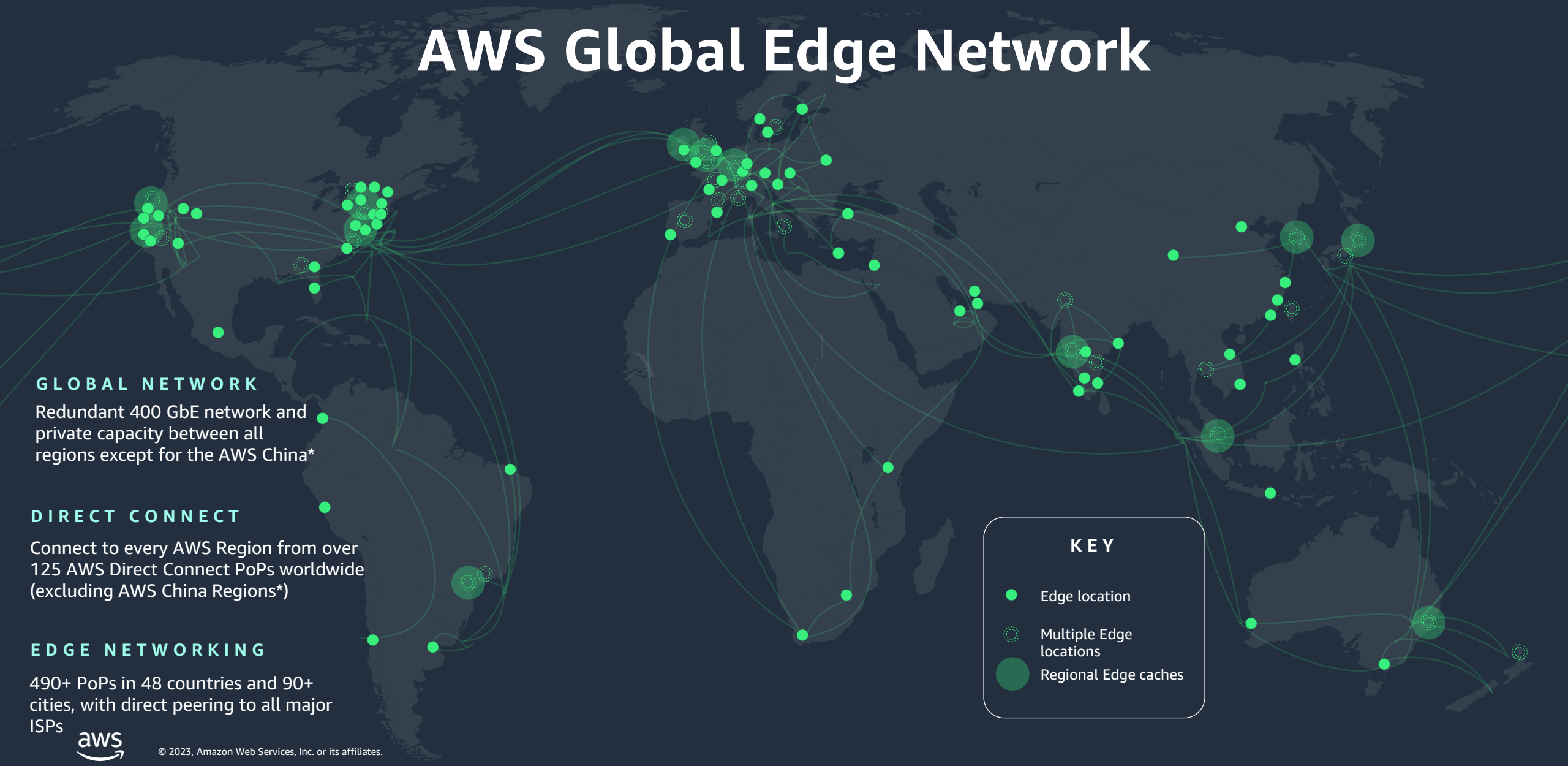
Asia Pacific

Auckland
Bangalore
Bangkok
Canberra
Chennai
Delhi
Hong Kong (x2)
Hyderabad
Inzai
Kuala Lumpur
Kolkata
Jakarta (x2)
Makati City
Manila
Melbourne (x2)
Mumbai (x4)
Osaka
Perth
Seoul (x2)
Singapore (x2)
Sydney (x3)
Taipei (x2)
Tokyo (x2)

*New locations opened in 2023

See AWS Direct Connect locations for more details

AWS Global Edge Network



GLOBAL NETWORK

Redundant 400 GbE network and private capacity between all regions except for the AWS China*

DIRECT CONNECT

Connect to every AWS Region from over 125 AWS Direct Connect PoPs worldwide (excluding AWS China Regions*)

EDGE NETWORKING

490+ PoPs in 48 countries and 90+ cities, with direct peering to all major ISPs

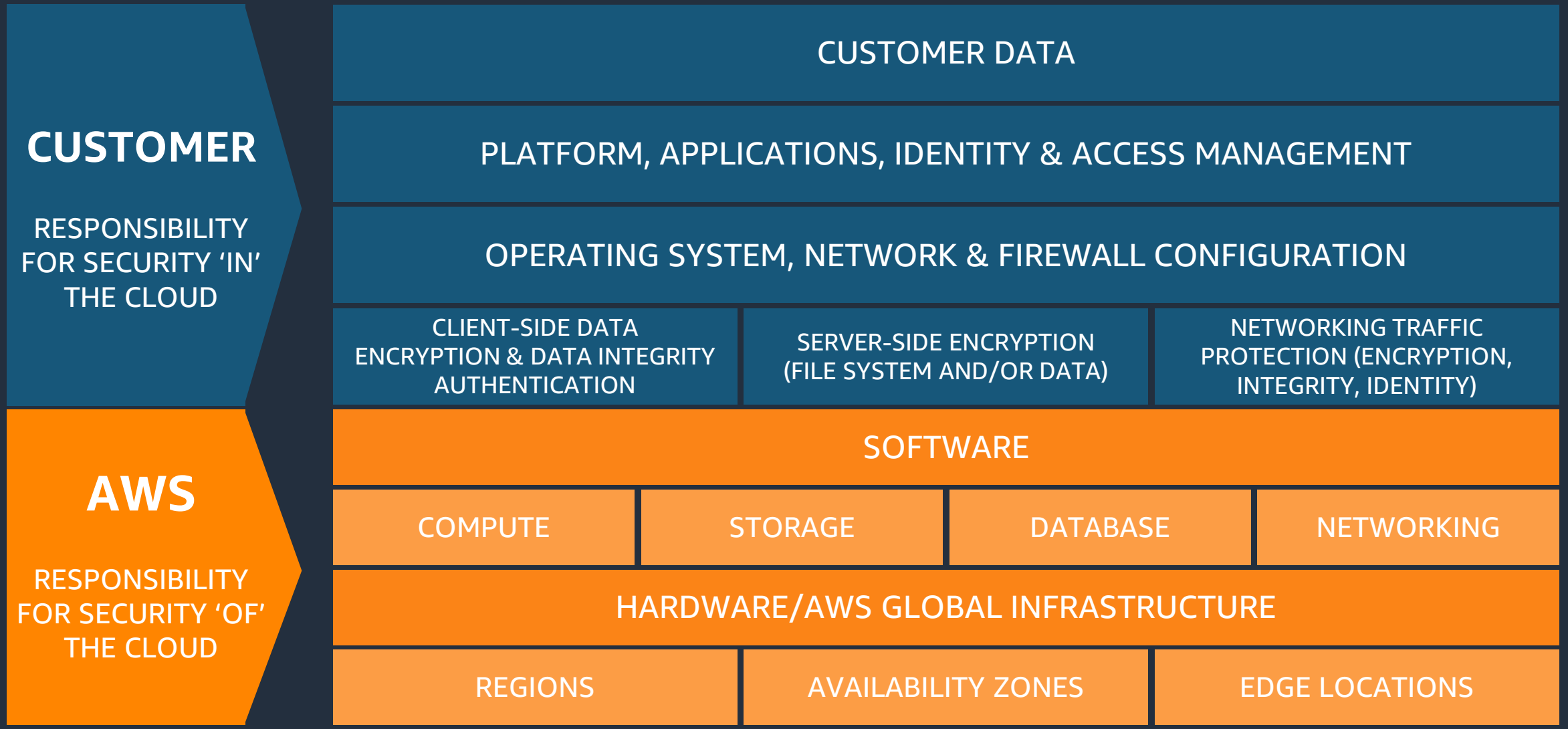
KEY

- Edge location
- Multiple Edge locations
- Regional Edge caches

aws

1
2

AWS Shared Responsibility Model



The line varies…

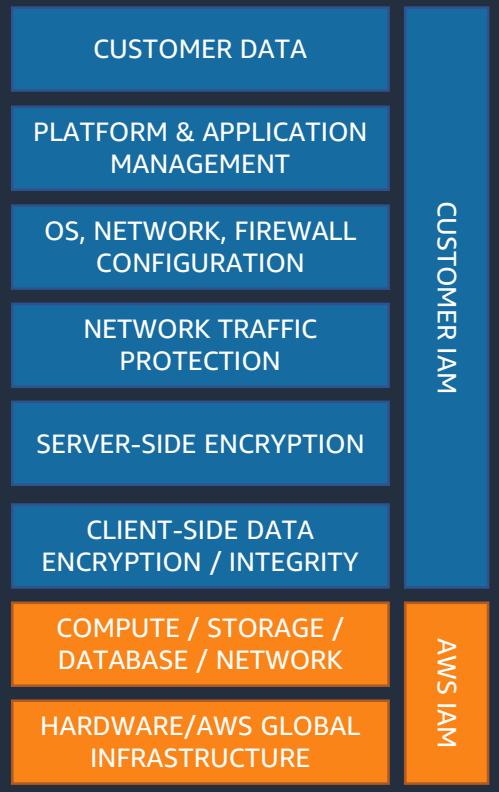
Amazon EC2

Amazon RDS

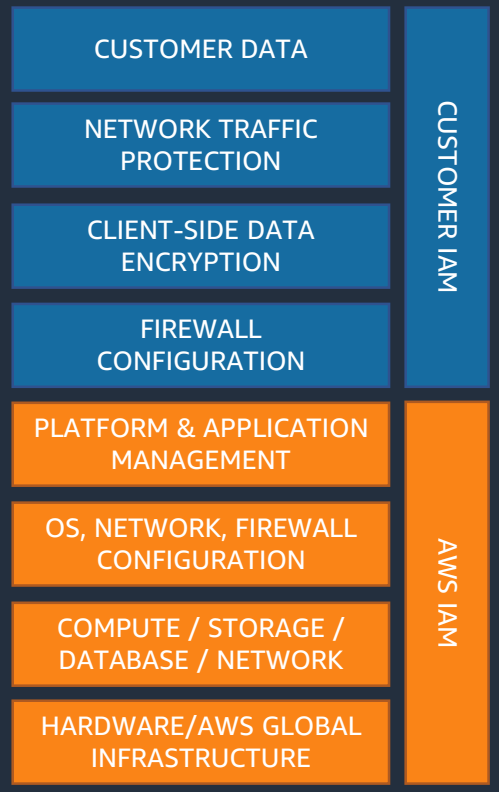
Amazon S3

AWS KMS

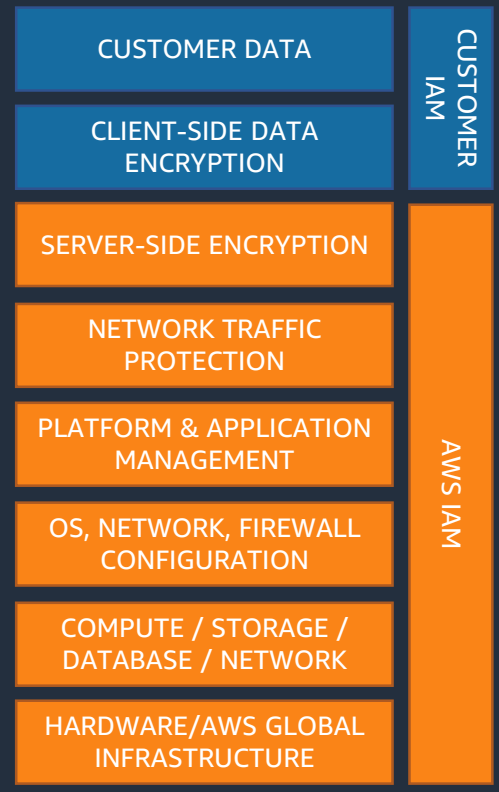
Amazon DynamoDB



INFRASTRUCTURE SERVICES



CONTAINER SERVICES



ABSTRACTED SERVICES

More Customizable
+
More Customer Responsibility

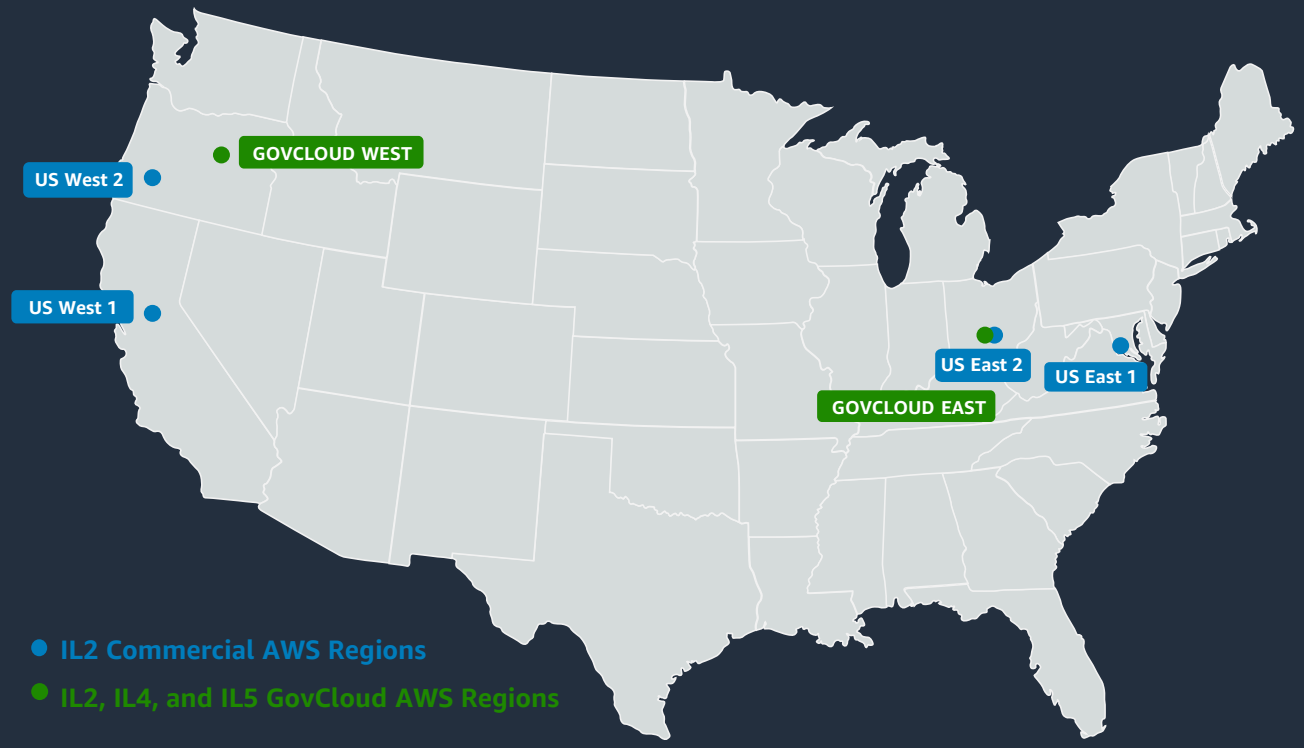
Less Customizable
+
Less Customer Responsibility
+
More Best Practices built-in

DoD Information Impact Levels

SRG v1r4 Impact Level	Maximum Data Type	Information Characterization
2	Non-Controlled Unclassified Information	Unclassified information approved for public release
		Unclassified, not designated as controlled unclassified information (CUI) or critical mission data, but requires some minimal level of access control
4	Controlled Unclassified Information	Requires protection from unauthorized disclosure as established by Executive Order 13556 (Nov 2010); Education, Training, SSN, Recruiting (if medical is not included), Credit card information for individuals (i.e., PX or MWR events)
		PII, PHI, SSN, Credit card information for individuals, Export Control, FOUO, Law Enforcement Sensitive, Email
5	Controlled Unclassified Information + NSS	National Security Systems and other information requiring a higher level of protection as deemed necessary by the information owner, public law, or other government regulations
6	Classified up to SECRET	Pursuant to EO 12958 as amended by EO 13292; classified national security information or pursuant to the Atomic Energy Act of 1954, as amended to be Restricted Data (RD)

DoD Cloud Computing Security Requirements Guide (SRG):
https://public.cyber.mil/dccs/dccs-documents/

AWS Regions for DoD



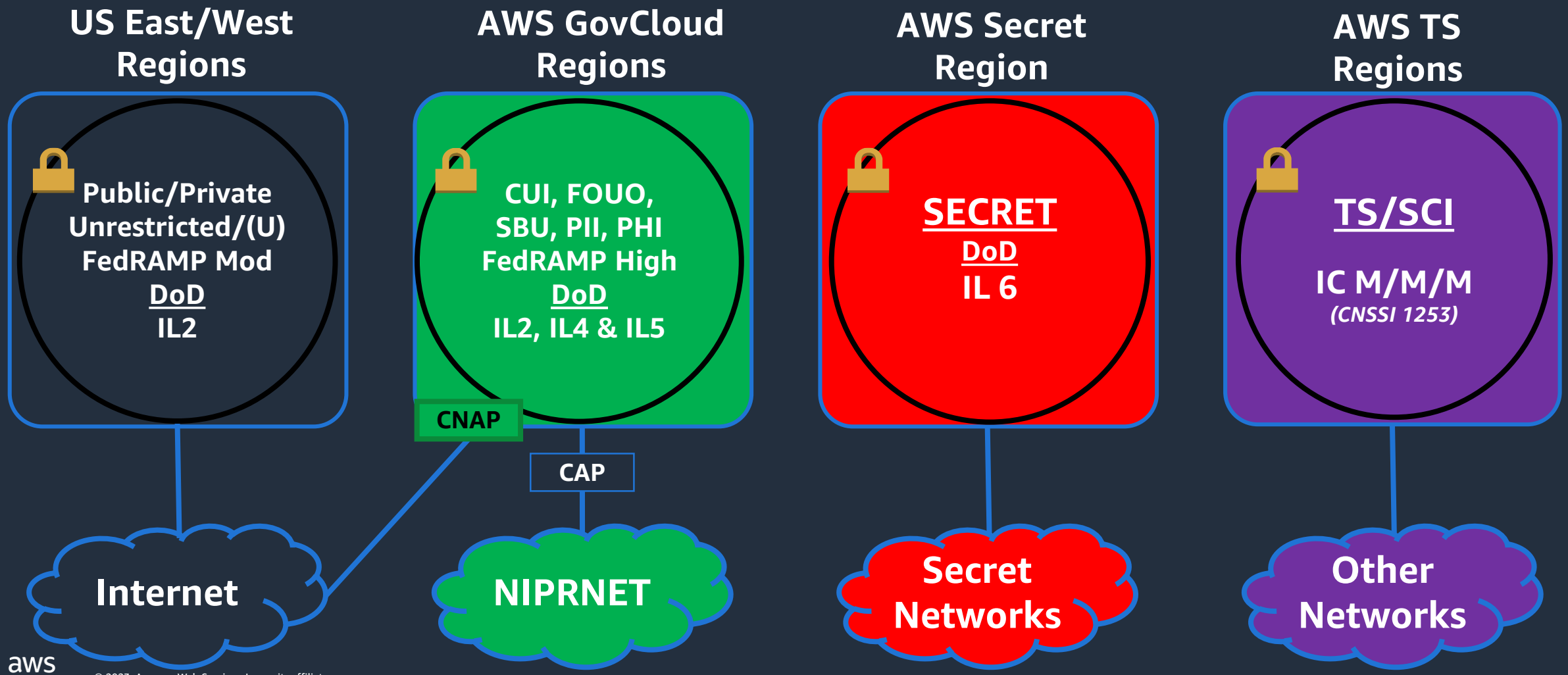
AWS GovCloud (US)

- 2 regions: US-East & US-West
- Accounts issued to US persons/companies only (to administer according to their security compliance regime)
- Data, network, and machine isolation
- Unique authentication with separate & isolated credential database and dedicated management console
- Managed by U.S. citizens on U.S. soil
- FIPS 140-2 compliance
- Service(s) are deployed into Region based on customer demand

Additional Regions Available at S and TS

GovCloud Regions offers the same high level of security as the other AWS Regions… restricted to approved US Persons

Cloud Infrastructure to meet DoD needs



DoD CC SRG IL Compliant Services

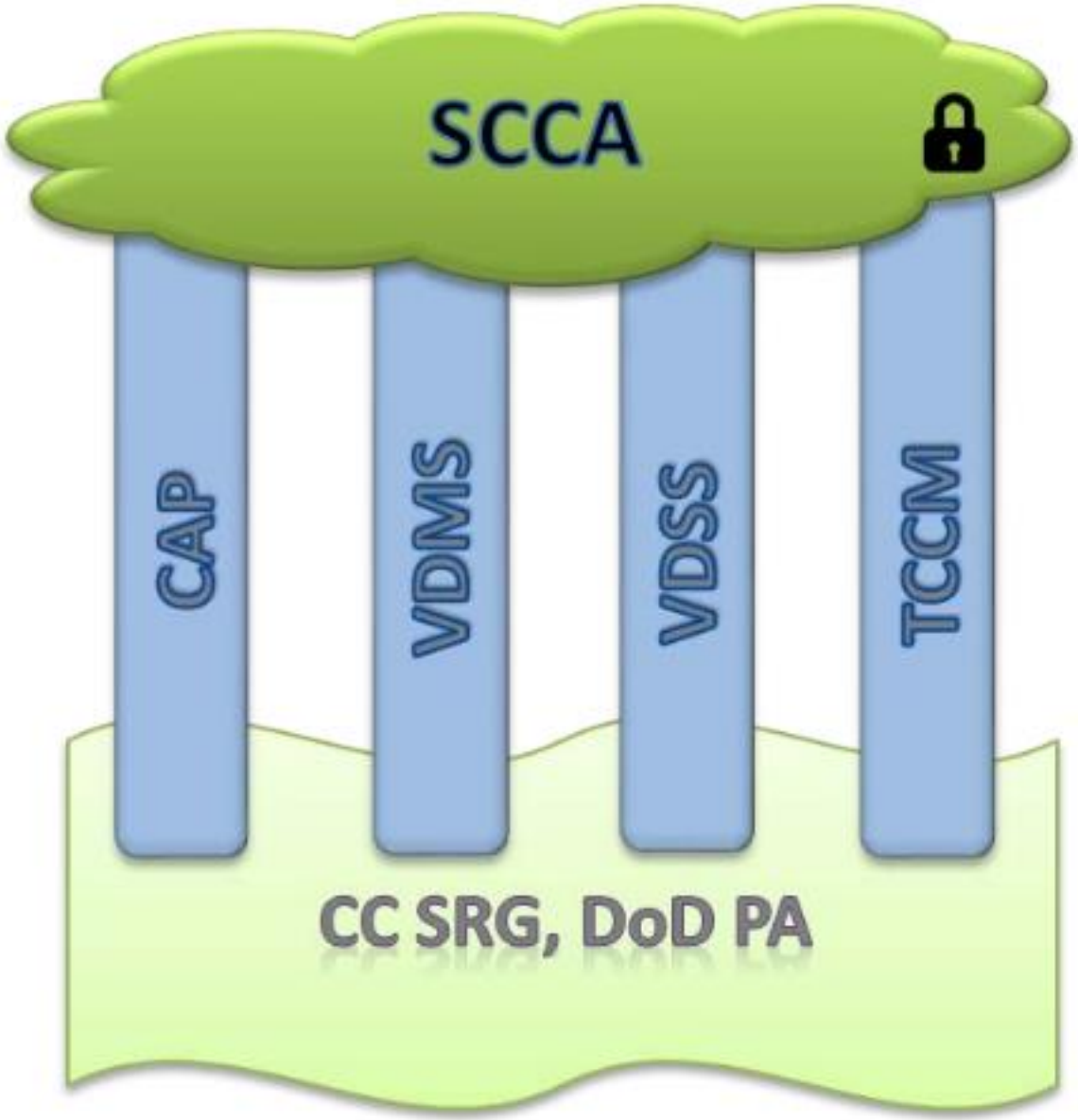
- Provides a list of AWS Services in Scope of AWS assurance programs

SERVICES / PROGRAMS	SDKs	DoD CC SRG IL2 (East/West)	DoD CC SRG IL2 (GovCloud)	DoD CC SRG IL4 (GovCloud)	DoD CC SRG IL5 (GovCloud)	DoD CC SRG IL6 (AWS Secret Region)
Amazon API Gateway	apigateway	✓	✓	✓	✓	✓
Amazon AppStream 2.0	appstream	✓	✓	✓	✓	
Amazon Athena	athena	✓	✓	✓	✓	
Amazon Aurora (MySQL)		✓	✓	✓	✓	
Amazon Aurora (Postgres)		✓	✓	✓	✓	
Amazon Chime	chime	✓				
Amazon Chime SDK	meetings-chime		✓	DISA Review	DISA Review	
Amazon Cloud Directory	clouddirectory	✓	✓	✓	✓	
Amazon CloudFront	cloudfront	✓				
Amazon CloudWatch	cloudwatch	✓	✓	✓	✓	✓
Amazon CloudWatch Logs	logs	✓	✓	✓	✓	✓
Amazon Cognito	cognito-idp, cognito-identity, cognito-sync	✓	✓	✓	✓	

*Sample view; this list is updated daily here: https://aws.amazon.com/compliance/services-in-scope/

DoD SCCA FRD V2.9

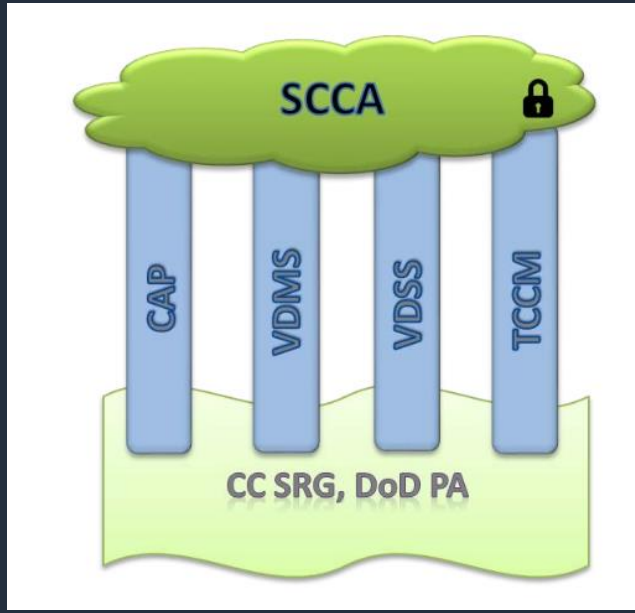
- Secure Computing Architecture (SCCA) Functional Requirements Document (FRD)
- Released March 9th 2017
- Replaces the Draft CAP FRD
- Provides implementation **flexibility**
- **Freedom** to architect and manage as a shared services enclave



DoD SCCA Component Functional Requirements

Virtual Datacenter Security Stack (VDSS)
Provides network and application security capabilities such as an application-aware firewall and/or intrusion prevention system.

Virtual Datacenter Management Stack (VDMS)
Provides system support services for mission owner environments (AD/LDAP, DNS, Patch Repos). Potentially CSSP offerings as well.

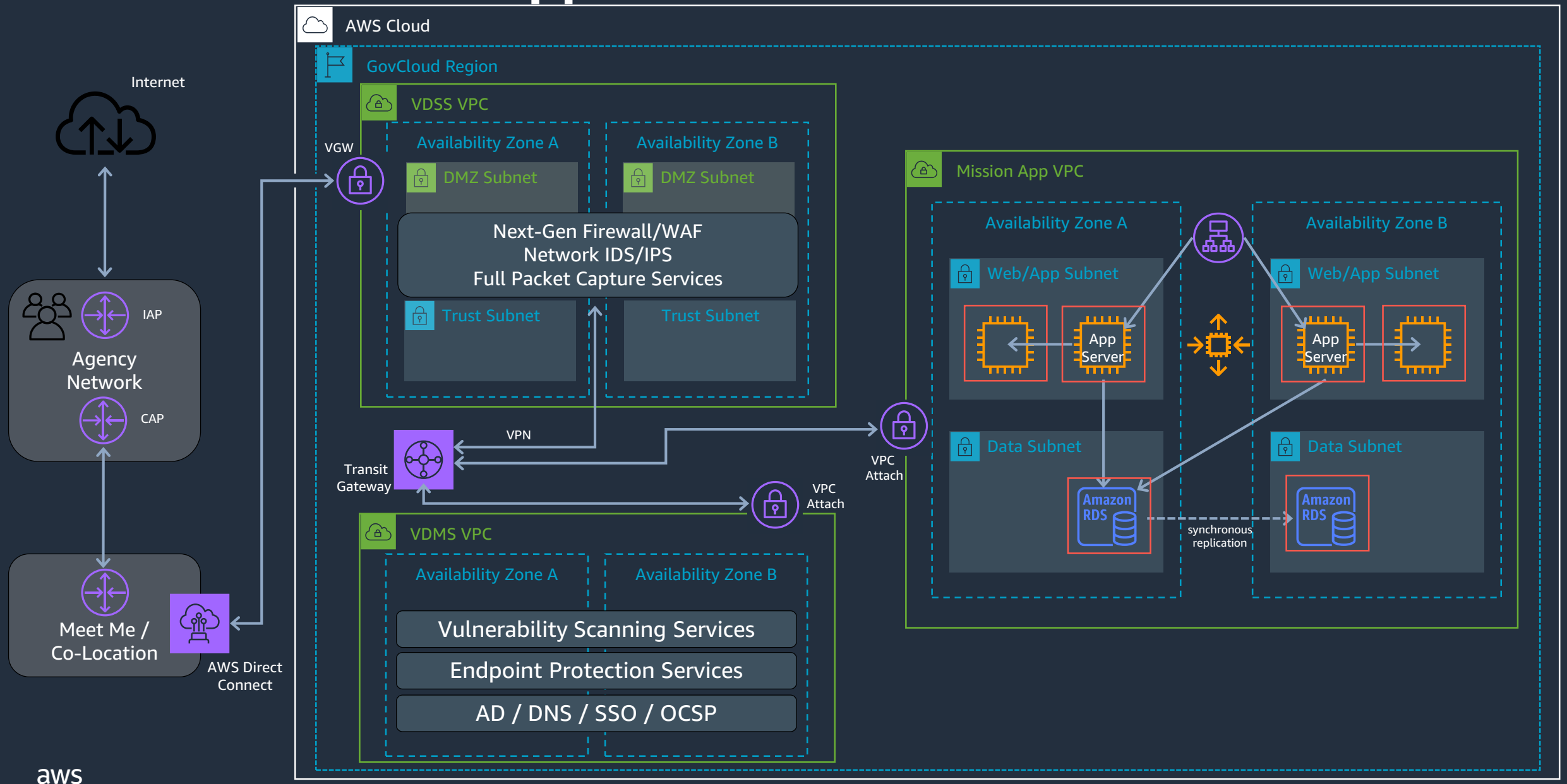


Trusted Cloud Credential Manager (TCCM)
An individual or entity appointed by the Authorizing Official to establish **policies** for controlling privileged user access to connect Virtual Private Clouds to DISN and for administrating cloud services

Cloud Access Point (CAP)
Provides network access to the cloud and boundary protection of DISN from the cloud.

SCCA Architecture Approach in AWS

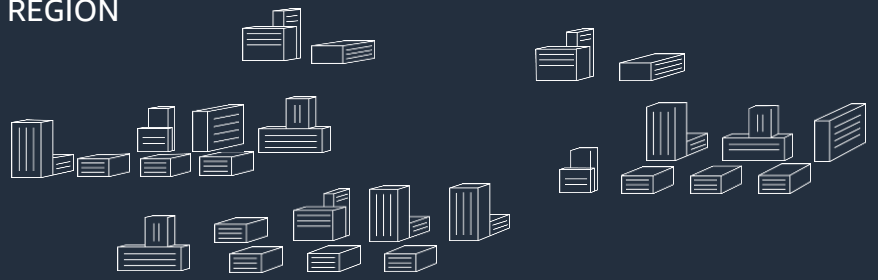
From DoD SCCA section 2.2



Networking

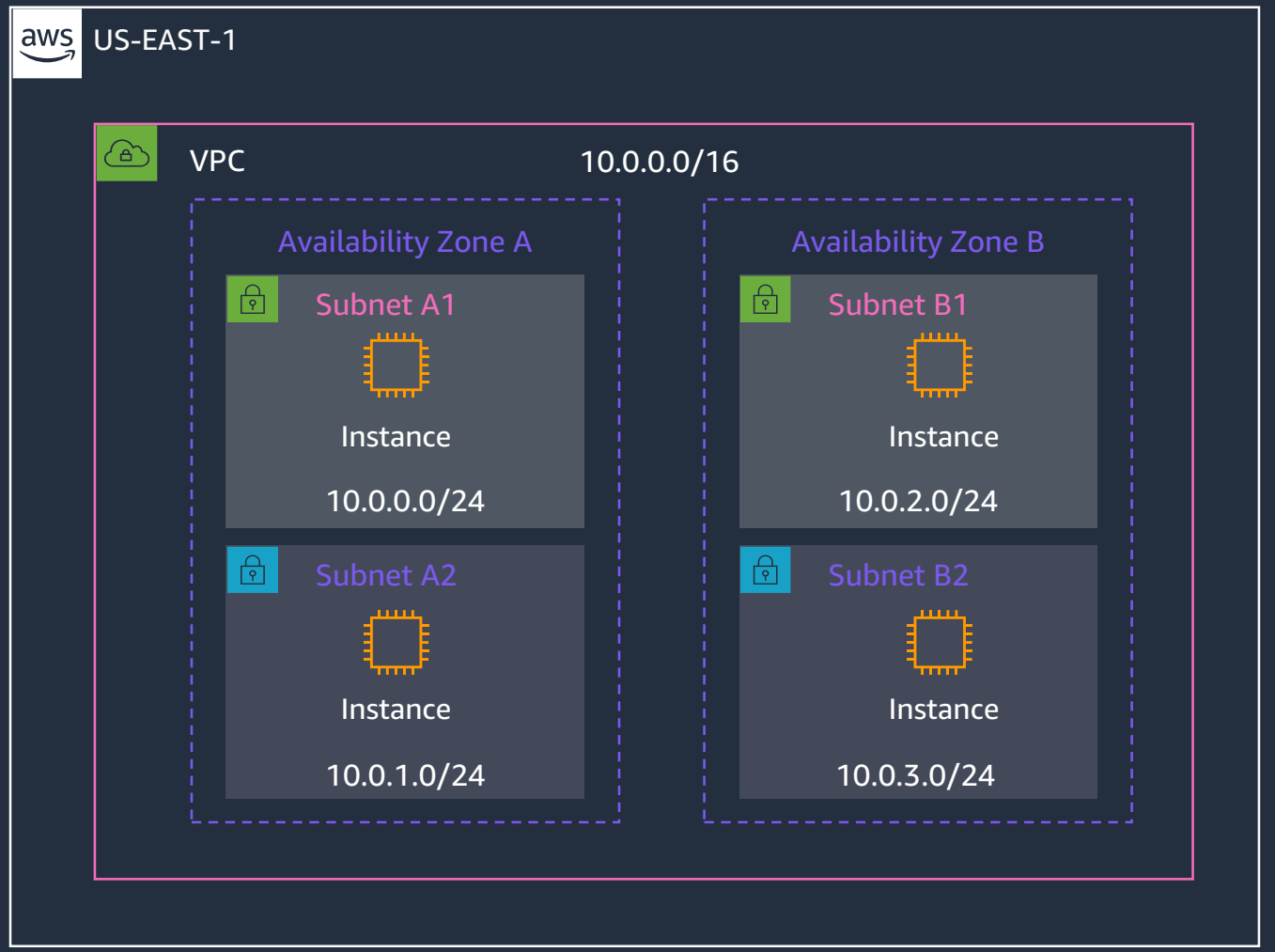
Amazon Virtual Private Cloud (VPC) overview

REGION



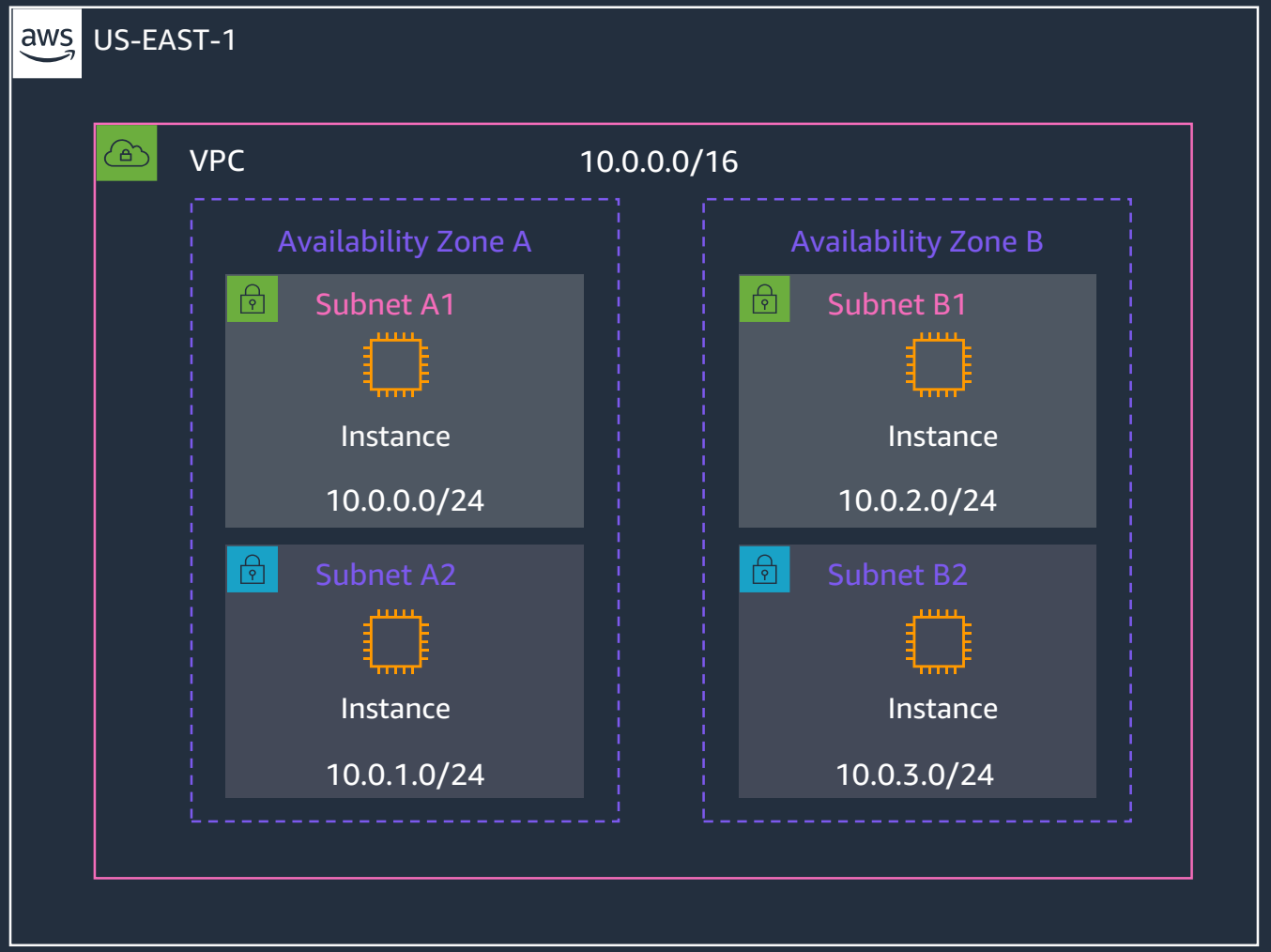
AVAILABILITY ZONE

DATA CENTER, RACK, HOST



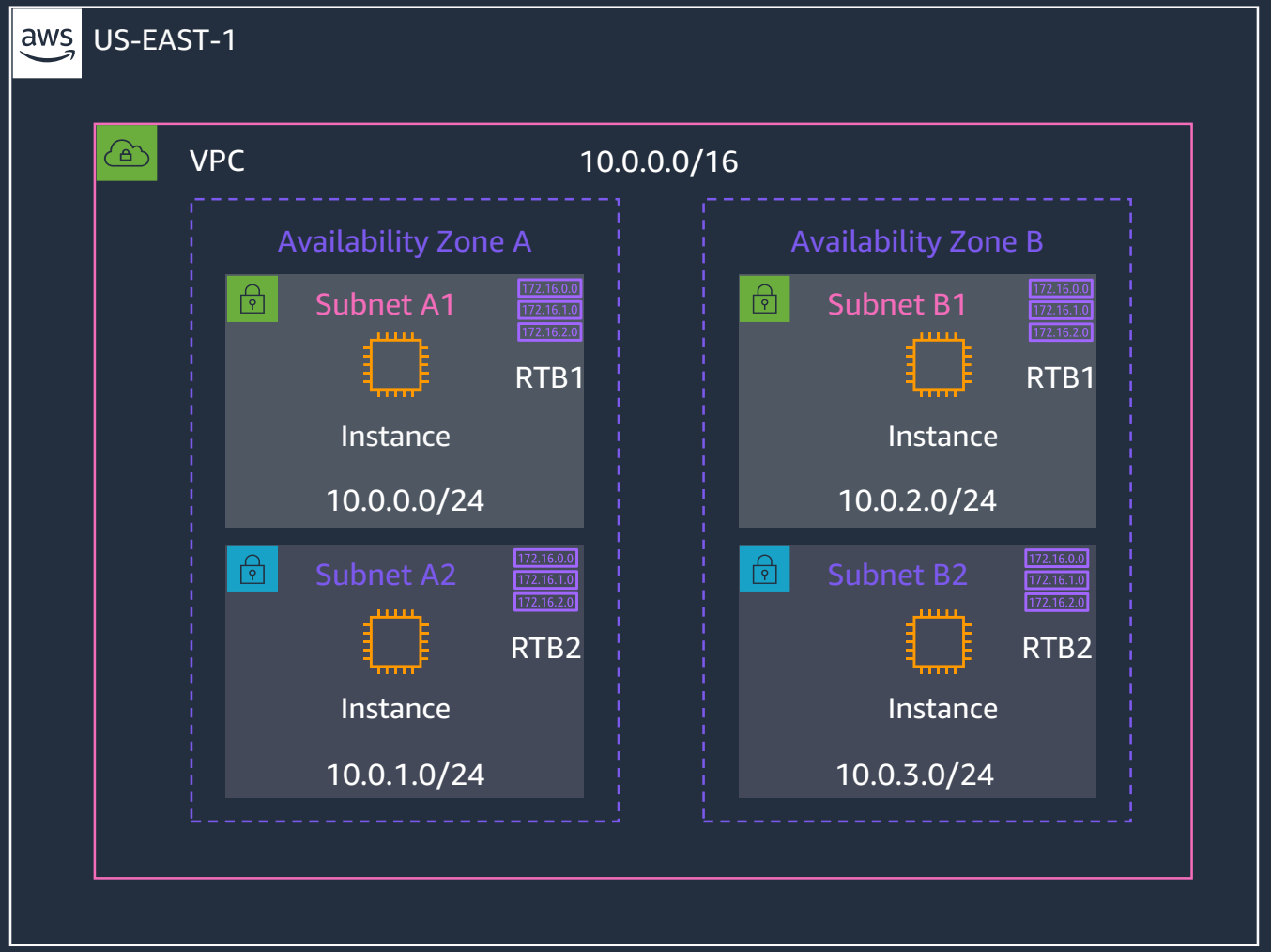
Subnets

- VPCs span a region
- Subnets are allocated as a subset of the VPC CIDR range and span a specific AZ
- You can have up to 200 subnets per VPC.
- Implicit route between all subnets within a VPC



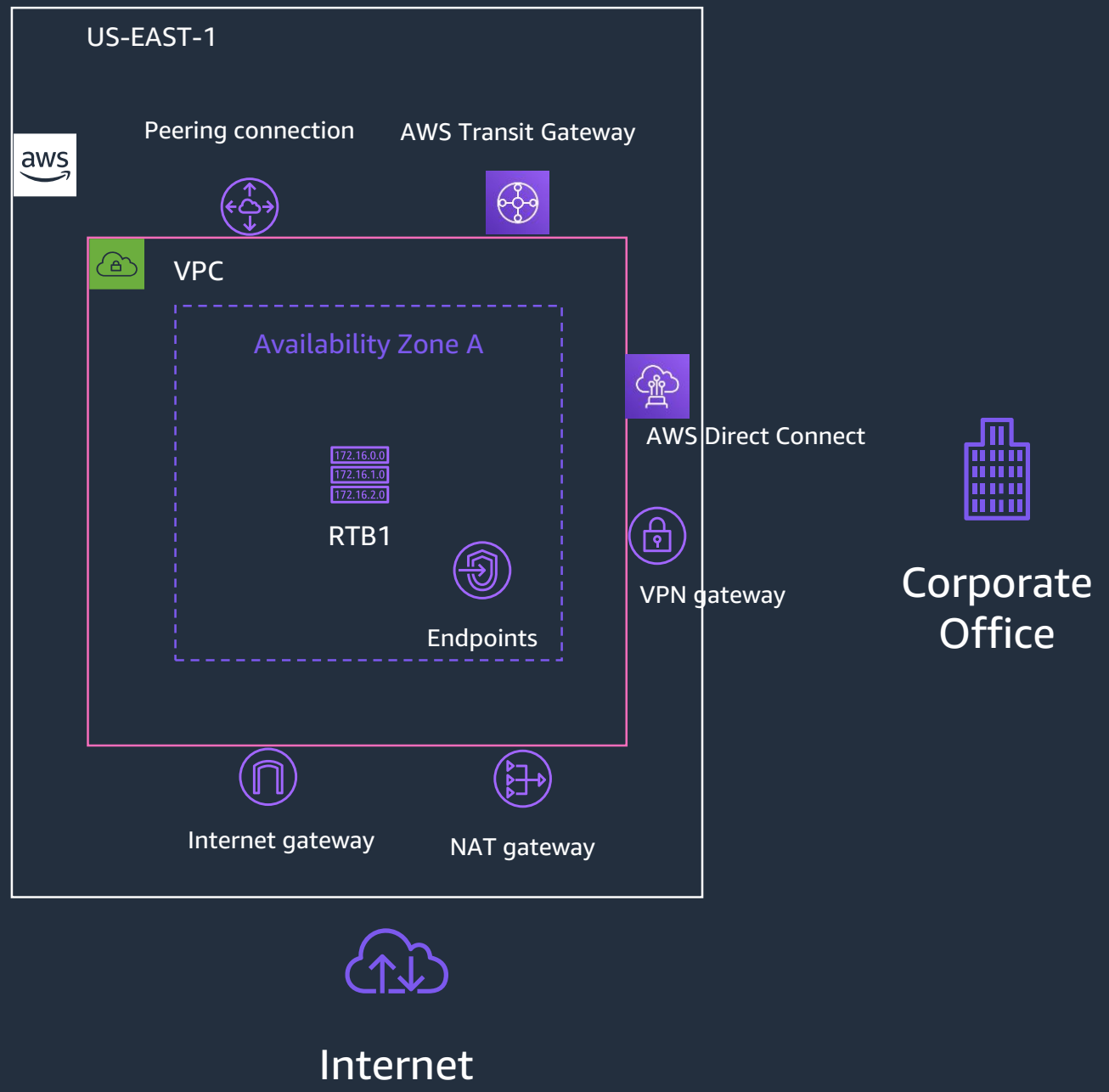
Routing tables

- Each subnet has associated routing table
- Routing tables can be associated with multiple subnets
- You can have 50 routes per route table.



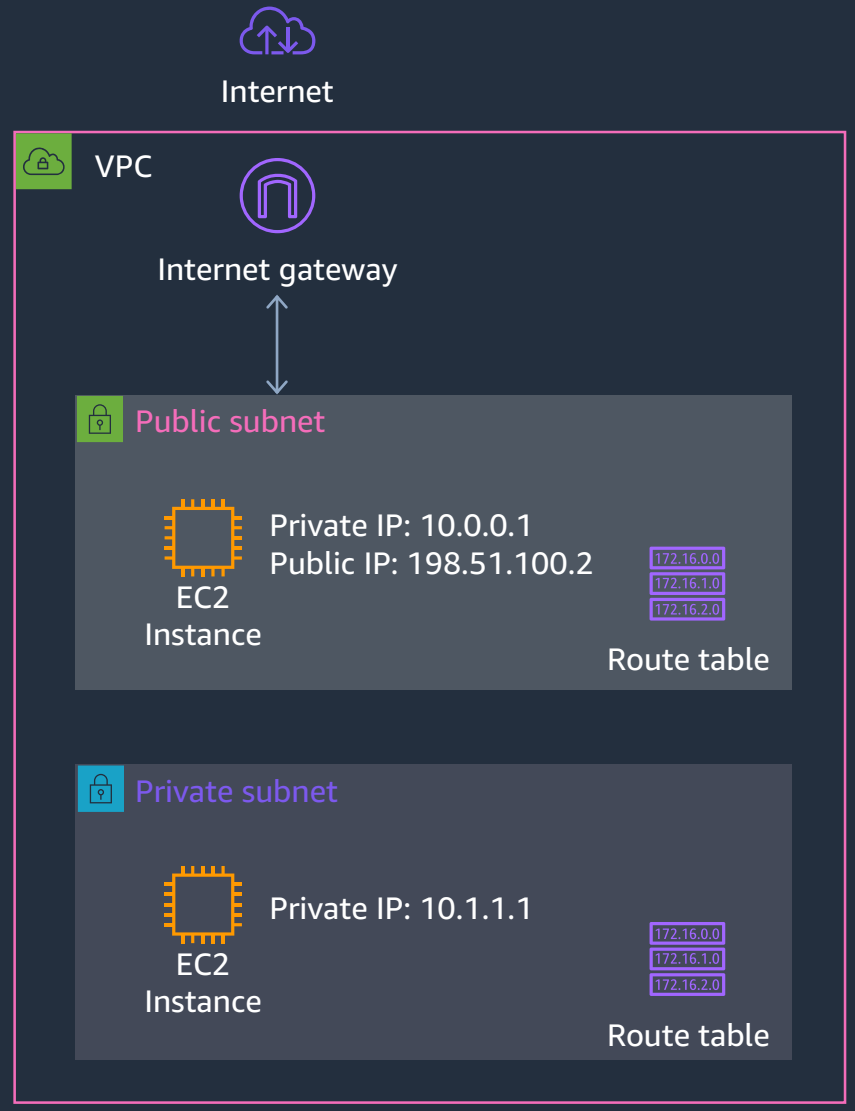
Routing

- Route Tables direct traffic towards:
  - Internet / NAT Gateway
  - Gateway Endpoint
  - VPC Peering / AWS Transit Gateway
  - VPN Gateway / Direct Connect
- Subnets are referred to as “Public Subnets” when there is a route to an Internet Gateway



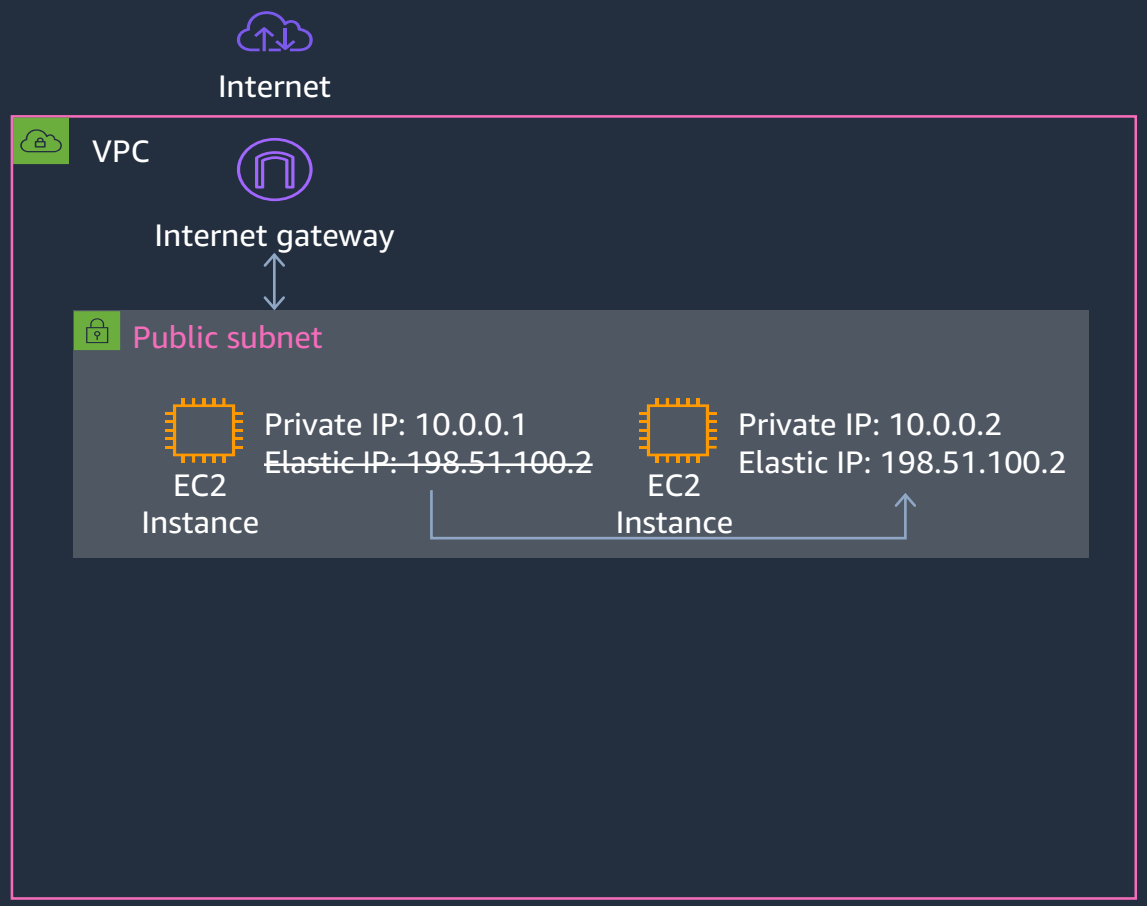
VPC to Internet: Internet Gateway

- Horizontally scaled, redundant, highly available VPC component
- Connect your VPC Subnets to the Internet
- Must be referenced on the Route Table
- Performs 1:1 NAT between Public and Private IP Addresses



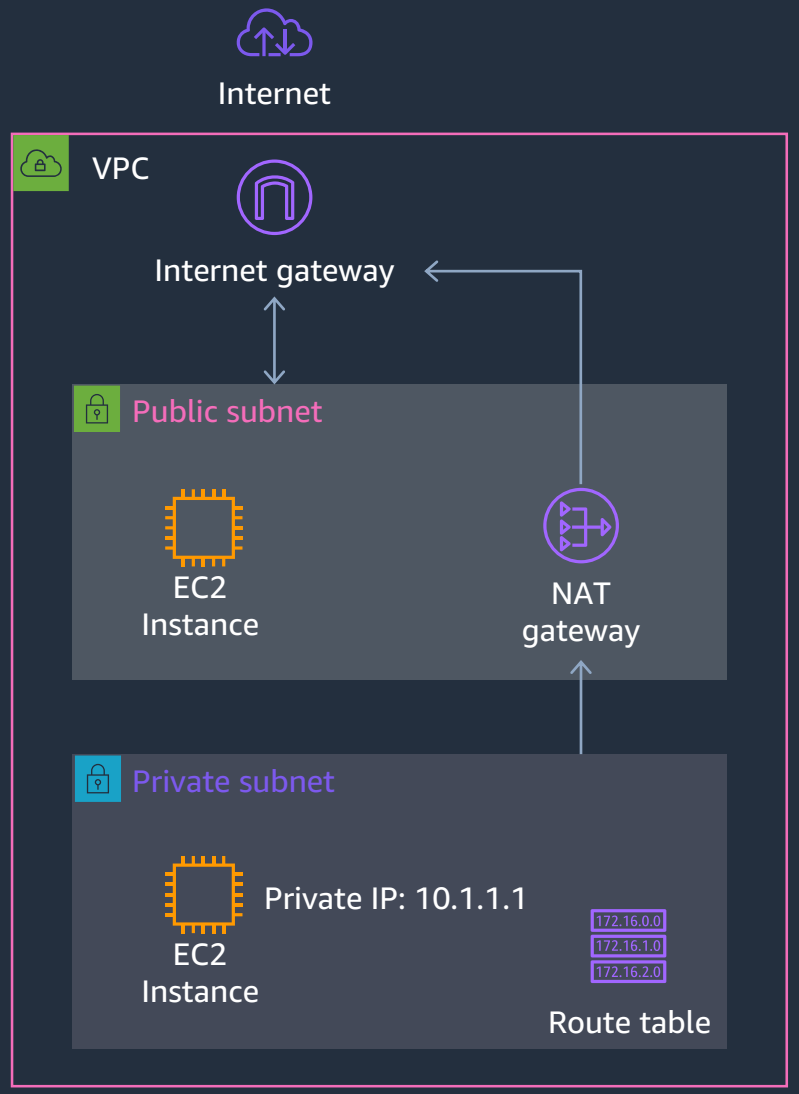
Public IP addressing: Elastic IP Address

- Static, Public IPv4 address, associated with your AWS account
- Dynamically assigned
- Specific to a region
- Can be associated with an instance or network interface
- Can be remapped to another instance in your account
- Useful for redundancy when Load Balancers are not an option



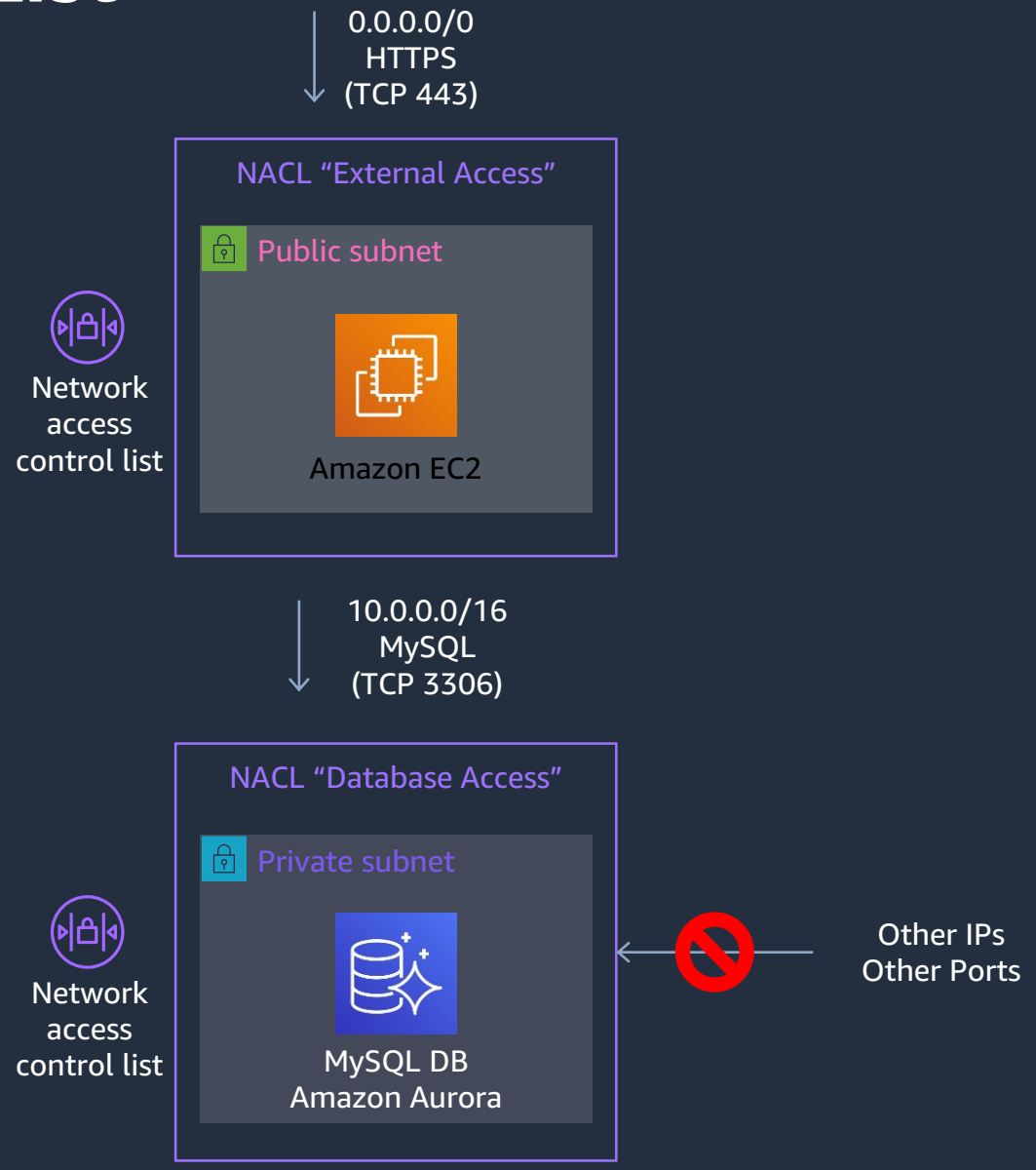
Outbound only traffic: NAT Gateway

- Enable outbound connection to the internet
- No incoming connection - useful for OS/packages updates, public web services access
- Fully managed by AWS
- Highly available
- Up to 45Gbps aggregate bandwidth
- Supports TCP, UDP, and ICMP protocols
- Network ACLs apply to NAT gateway traffic



IP FW: Network Access Control List

- Inbound and Outbound
- Subnet level inspection
- Optional level of security
- By default, allow all traffic
- Stateless
- IP and TCP/UDP port based
- Supports allow and deny rules
- Deny all at the end



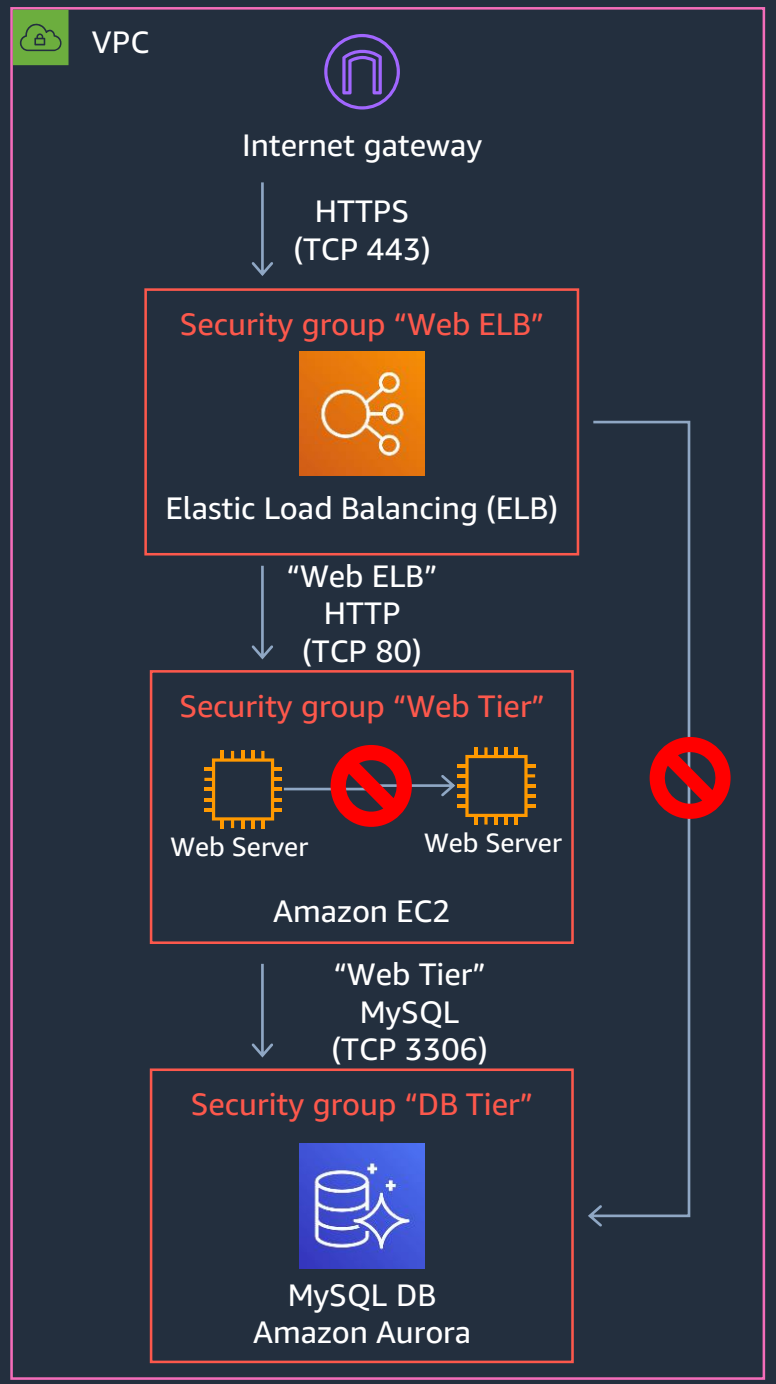
Resource FW: Security Groups

- Stateful firewall
- Inbound and Outbound customer defined rules
- Instance/Interface level inspection

  Micro segmentation

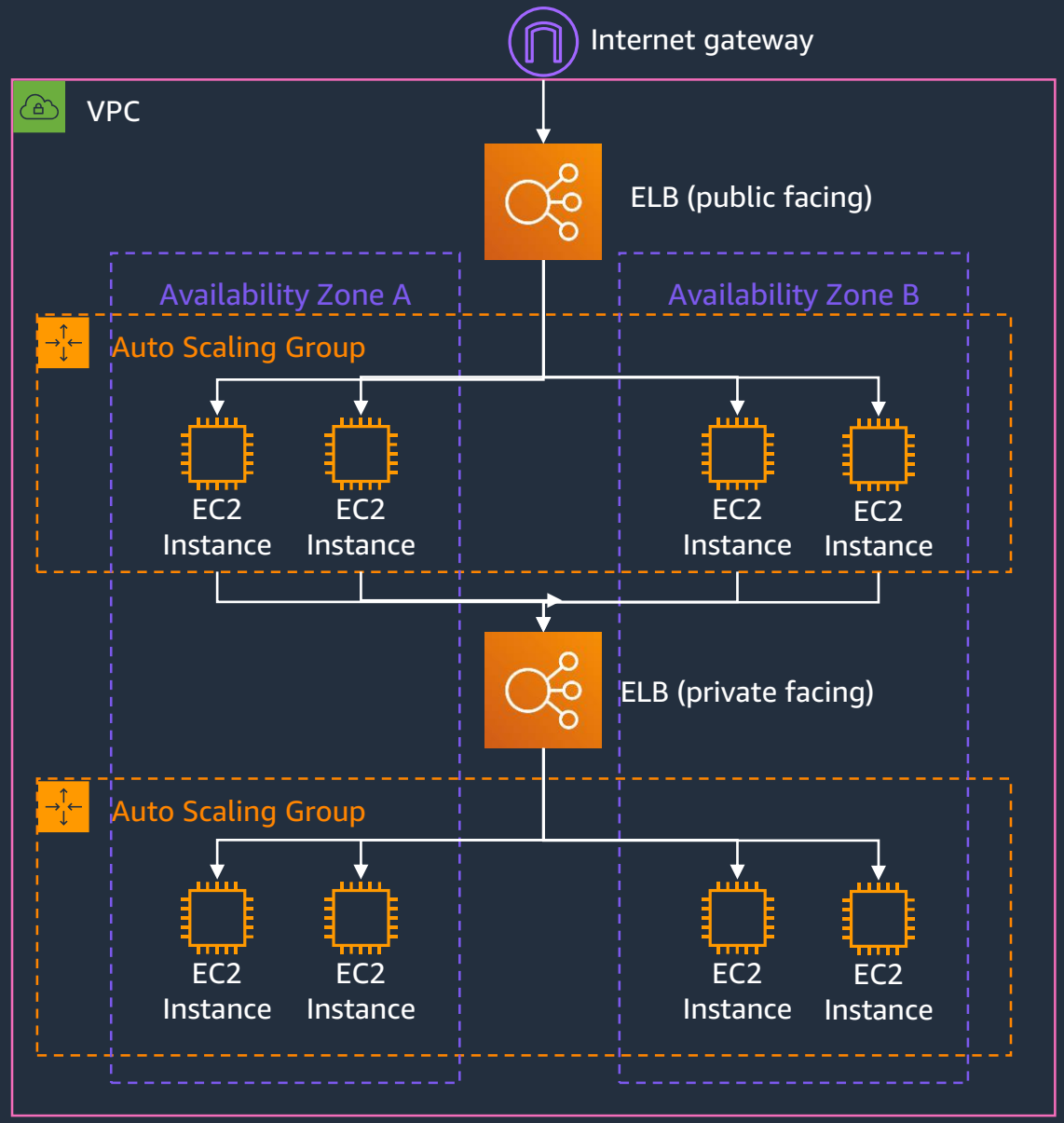
  Mandatory, all instances have an associated Security Group
- Can be cross referenced

  Works across VPC Peering
- Only supports allow rules
- Implicit deny all if not allowed
- AWS Verified Access adds a new logging functionality to improve troubleshooting



Horizontal scaling: Elastic Load Balancing

- Distribute traffic to multiple targets
  - EC2 instances
  - Containers
  - IP addresses
  - Lambda
- Multiple Availability Zones
- ELB Scales automatically
- Support Auto Scaling Groups
  - Automatically (de)register instances to the ELB based on health checks.



Types of ELB: NLB / ALB

Network Load Balancer (NLB)

- Layer 4 Load Balancing
- Connection-based Load Balancing
- High Throughput
- Low Latency
- Preserve source IP address
- Static IPs
- Long-lived TCP Connections

Application Load Balancer (ALB)

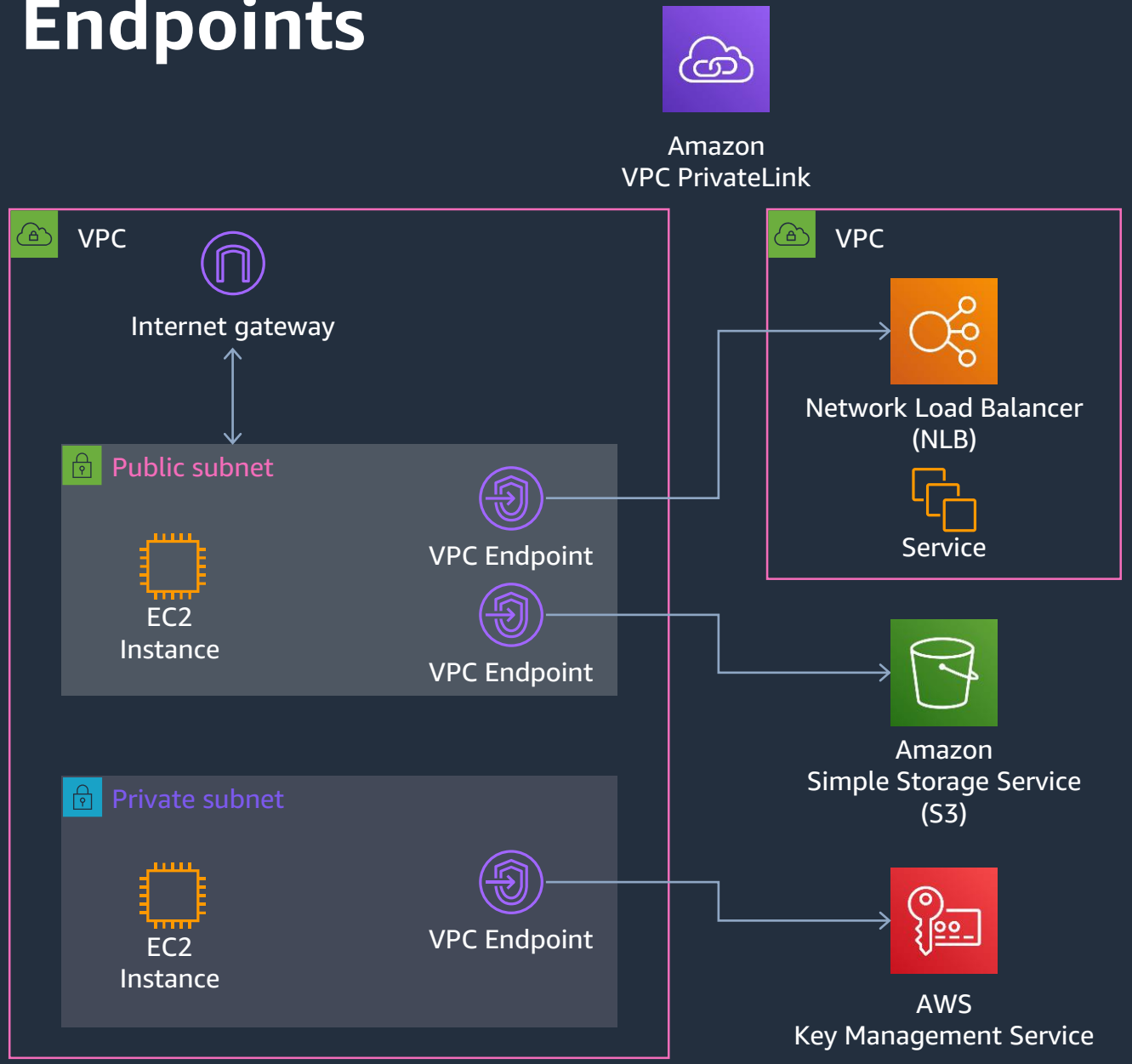
- Layer 7 Load Balancing
- Content-Based Routing (host and path based)
- HTTP/2 Support
- Request Tracing
- Web Application Firewall (WAF) integration
- Application Load Balancer now supports TLS 1.3

Supported by both

- Cross-zone load balancing
- WebSocket Support
- Deletion Protection

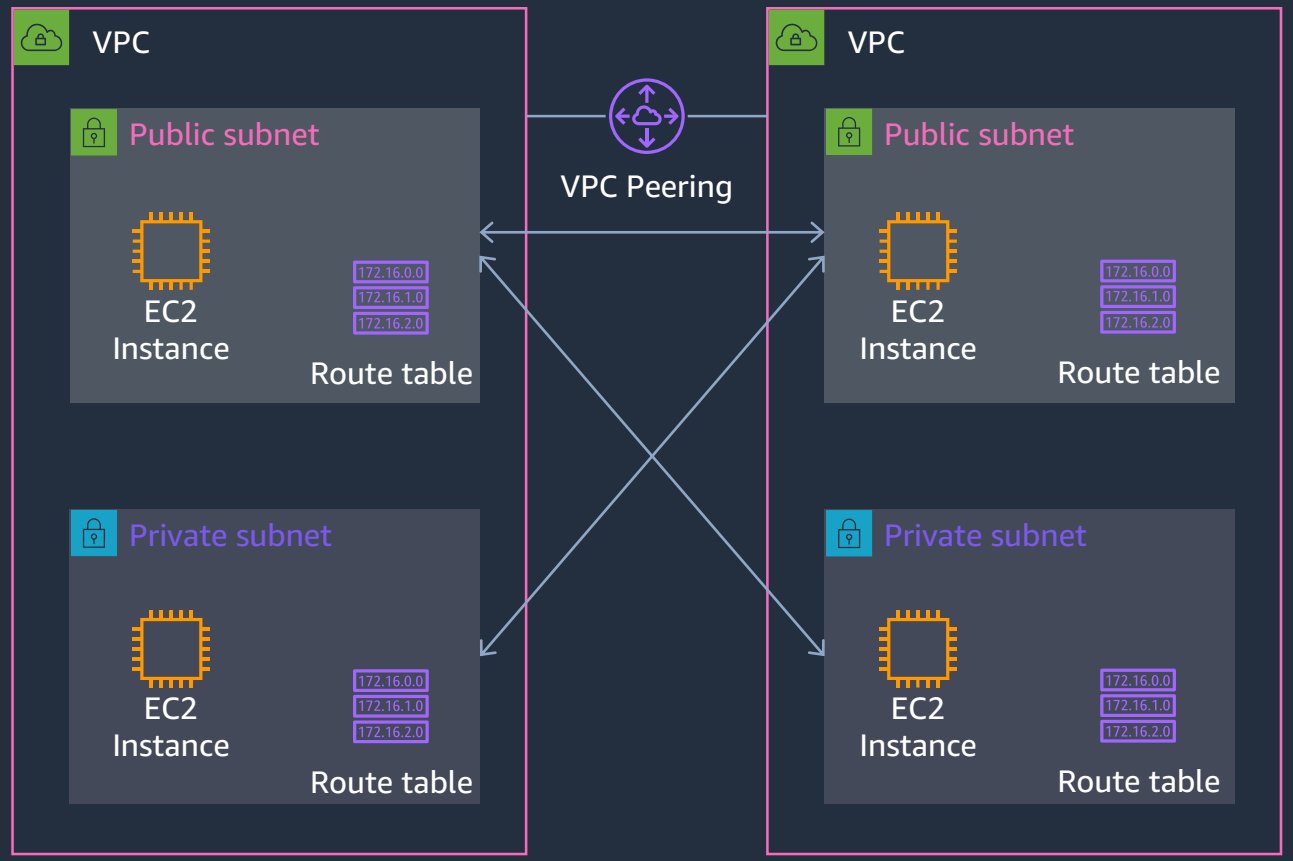
Stay on AWS network: VPC Endpoints

- Connect your VPC to:
  - Supported AWS services
  - VPC endpoint services powered by PrivateLink
- Doesn’t require public IPs or Internet connectivity
- Horizontally scaled, redundant, and highly available
- Robust access control
- Metrics for traffic visibility



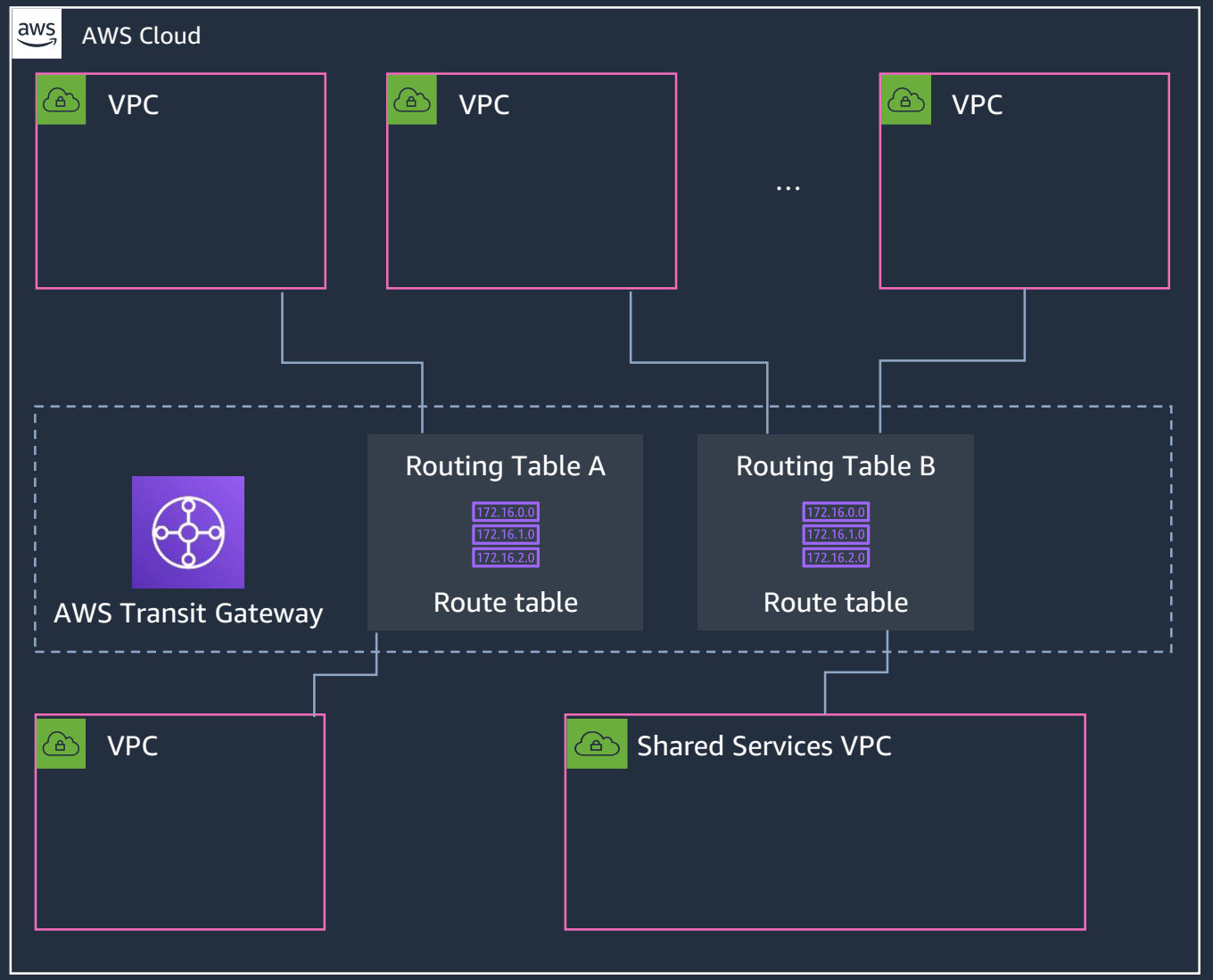
Connect multiple VPCs: VPC Peering

- Scalable and high available
- Supported between AWS accounts
- Supported across AWS Regions
- Bi-directional traffic
- Remote Security groups can be referenced
- Routing policy with Route Tables
  - Not all subnets need to connect to each other
- No overlapping IP addresses
- No transitive routing



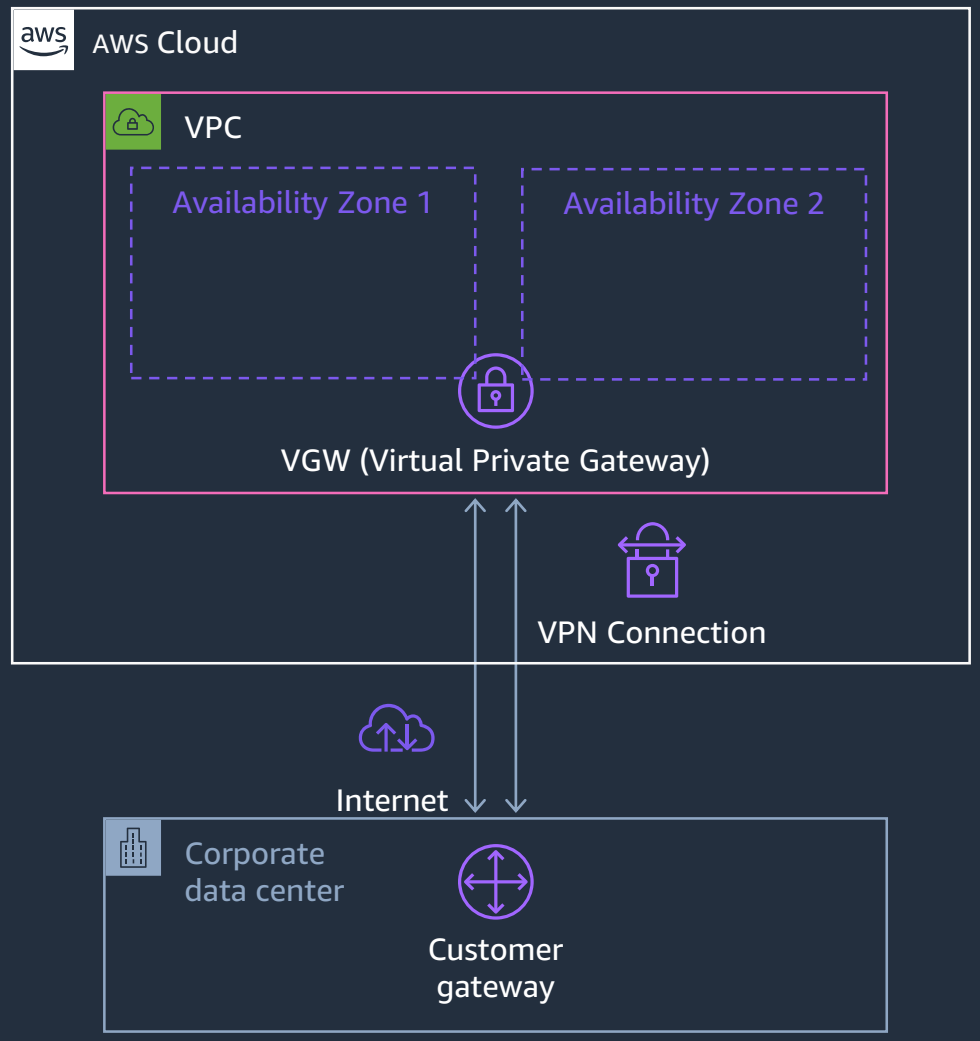
Connect multiple VPCs: Transit Gateway

- Connect thousands of VPC across accounts within a region
- Connect your VPCs and on-premises through a single transit gateway
- Centralize VPN and AWS Direct Connect connections
- Control segmentation and data flow with Route Tables
- Hub and Spoke design
- Up to 50 Gbps per attachment (burst)



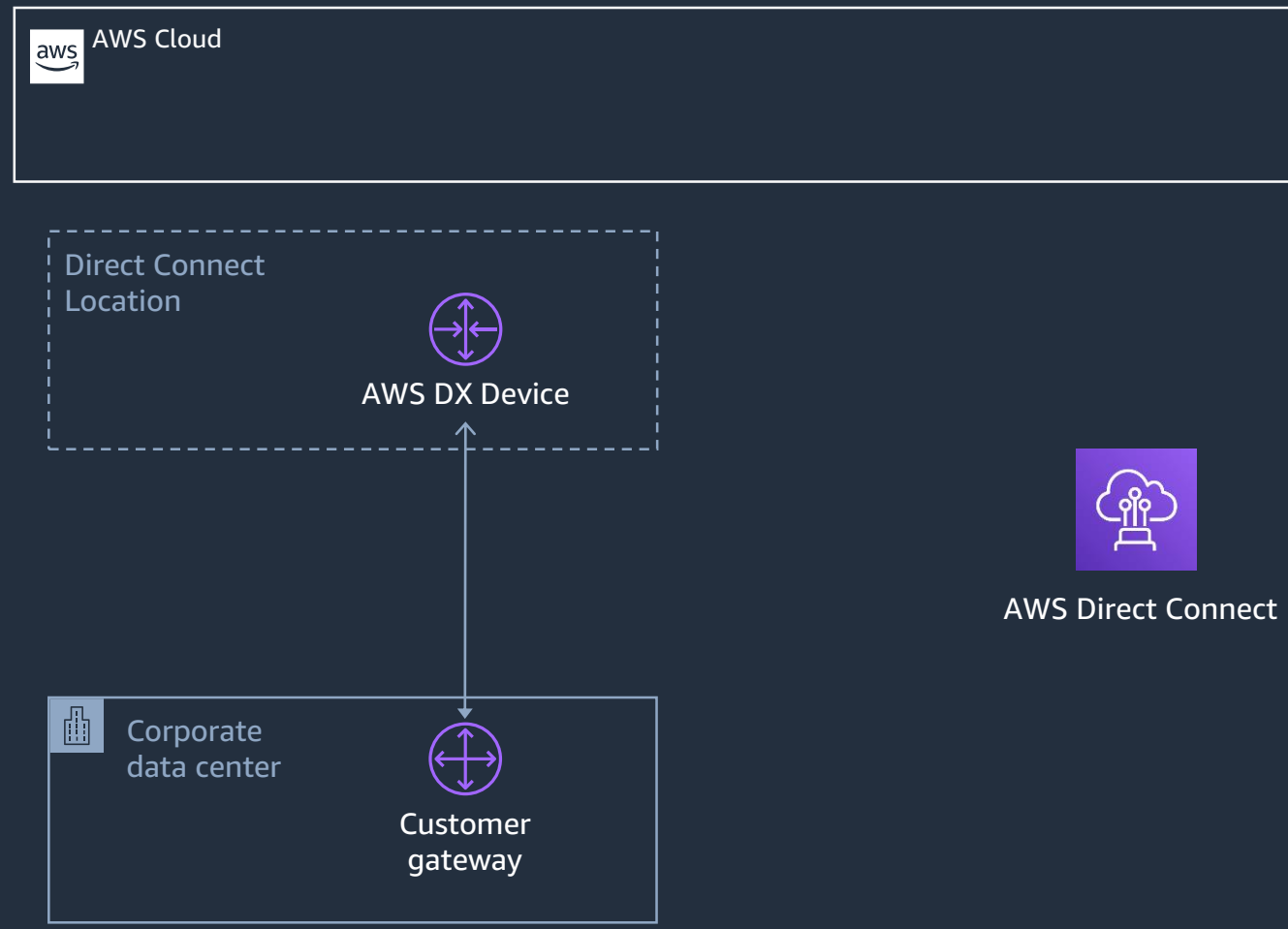
VPN to AWS: Virtual Private Gateway

- Fully managed VPN endpoint device
- One Virtual Private Gateway per VPC
- Redundant IPSec VPN Tunnels Terminating in different AZs
- IPSec AES 256-bit encryption SHA-2 hashing
- Scalable
- Dynamic (BGP) or Static Routing
- Default 10 Site-to-Site VPN connections per VGW – can increase limit



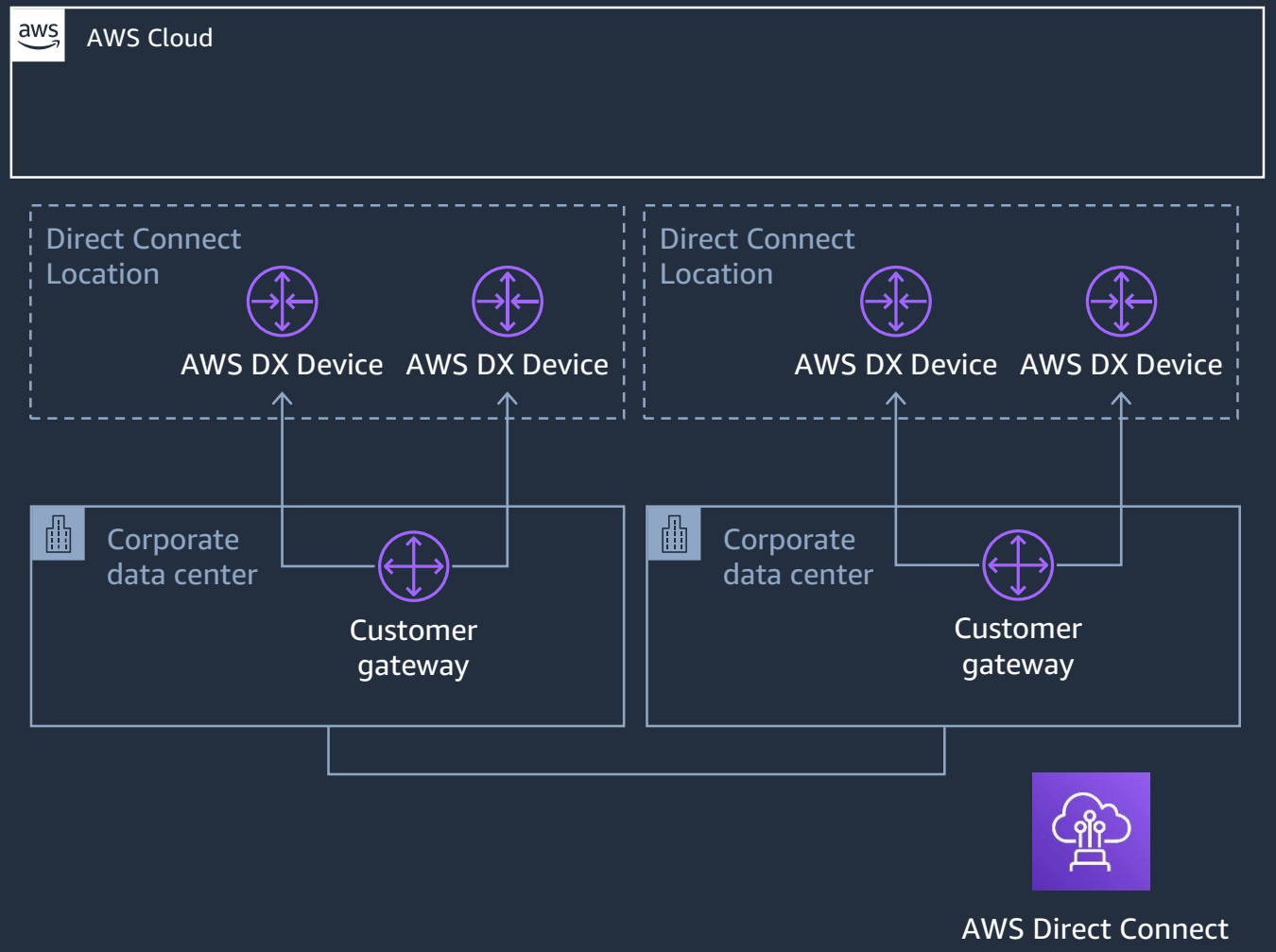
Dedicated link to AWS: AWS Direct Connect

- Dedicated network connection from your premises to AWS
- Dedicated Connection (1,10 or 100 Gbps, Supports multiple VIFs)
- AWS Partner Hosted Connection (50 Mbps to 10 Gbps, Single VIF)
- Consistent Network Performance
  - Dedicated bandwidth
  - Low latency
- Reduced egress data charges
- Connect to 108 Direct Connection Locations across the globe



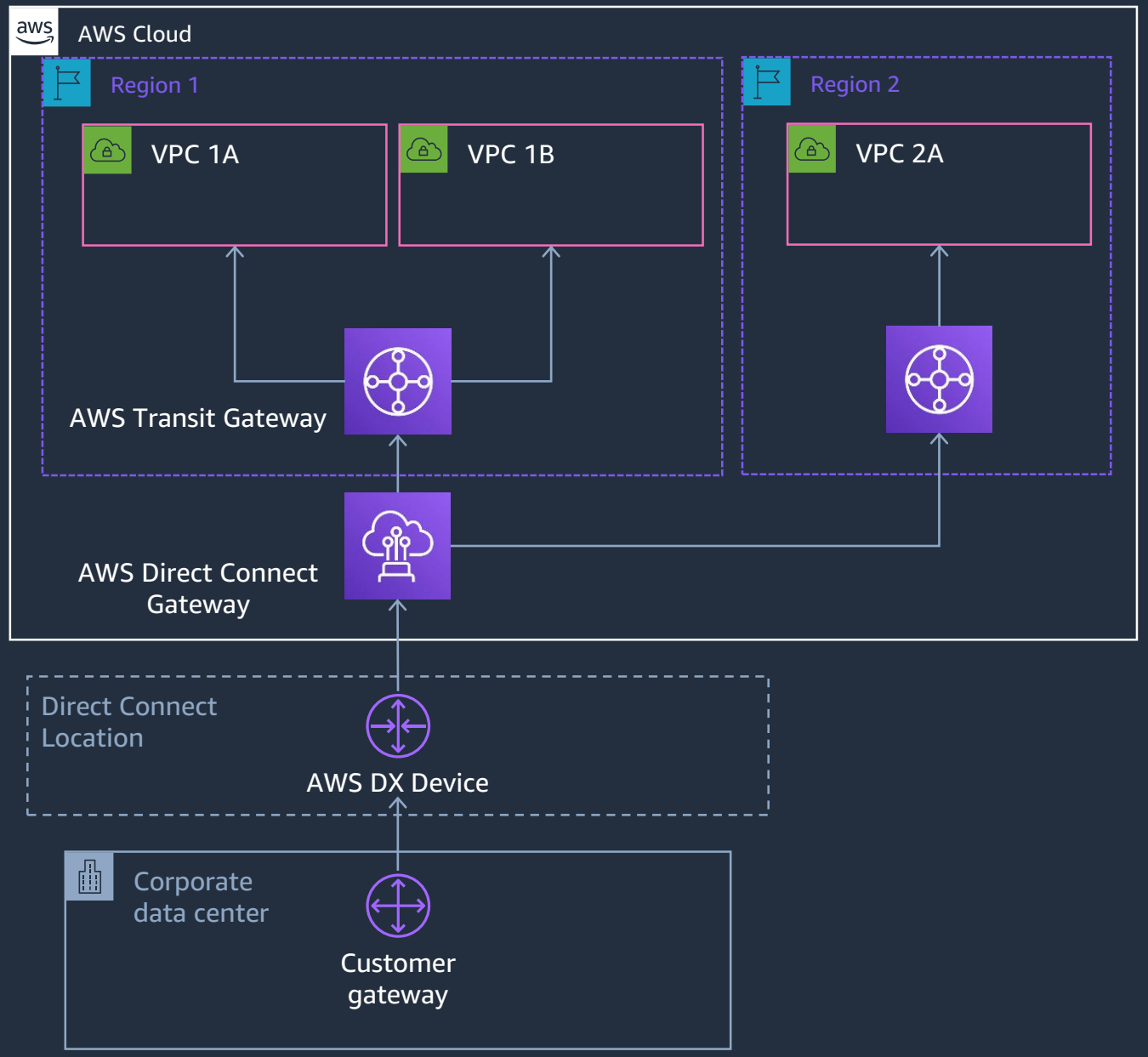
Dedicated link to AWS: AWS Direct Connect

- For redundancy, DX can deployed with single or multiples:
  - Circuits
  - Providers
  - Customer Gateways
  - Direct Connect Locations
  - Customer data centers
- BGP Routing for redundancy
  - AS Path Prepend
  - Scope BGP Communities
  - Local Preference BGP Communities



Connect at global scale: DX Gateway + Transit Gateway

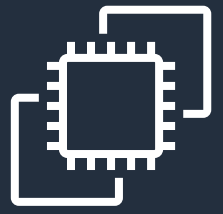
- Transit VIF
- Connects to a AWS Transit Gateway
- Simplify your network architecture and management overhead
- Create a hub-and-spoke model that spans multiple
  - VPCs
  - Regions
  - AWS accounts



Compute

Choices for Compute

World-class performance, security, and innovation



AMAZON EC2

Virtual server instances in the cloud

AMAZON ECS, EKS, and FARGATE*

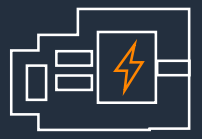
Container management service for running Docker on a managed cluster of EC2

AWS LAMBDA

Serverless compute for stateless code execution in response to triggers

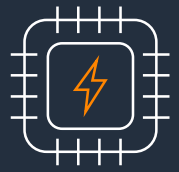
Nitro System is the foundation of AWS

Nitro Card



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

Nitro Security Chip



Integrated into motherboard
Protects hardware resources

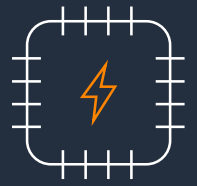
Nitro Hypervisor

Lightweight hypervisor
Memory and CPU allocation
Bare metal-like performance

Nitro Enclaves

Isolated environments for highly sensitive data processing
Utilizes EC2's Isolation Technology

Nitro SSD



60% lower I/O latency
Firmware Upgrades w/o Interruption
Encryption at rest

Nitro TPM

TPM 2.0 specification
Cryptographic attestation of instances integrity

Security is job Zero

Independently validated!: https://aws.amazon.com/blogs/compute/aws-nitro-system-gets-independent-affirmation-of-its-confidential-compute-capabilities/

Engineered with a **hardware-based root of trust** using the Nitro Security Chip, allowing for the system to be continuously measured and validated

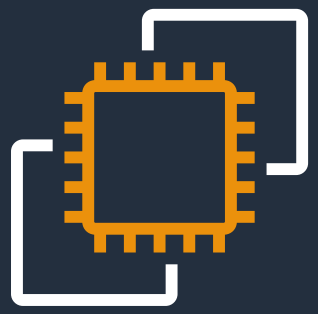
Dedicated **hardware-based virtualization** minimized attack surfaces, reducing risk of vulnerabilities

No operator or root access, including Amazon employees

VM tenancy protections. Instances are isolated from other tenants on the same EC2 host via VM, never containers.

Memory encryption enabled by default on Graviton2/3, Intel Ice Lake (TME) and AMD Milan (SME) instances

Amazon Elastic Compute Cloud (Amazon EC2)



AMAZON EC2

Linux | Windows | Mac

Arm and x86 architectures

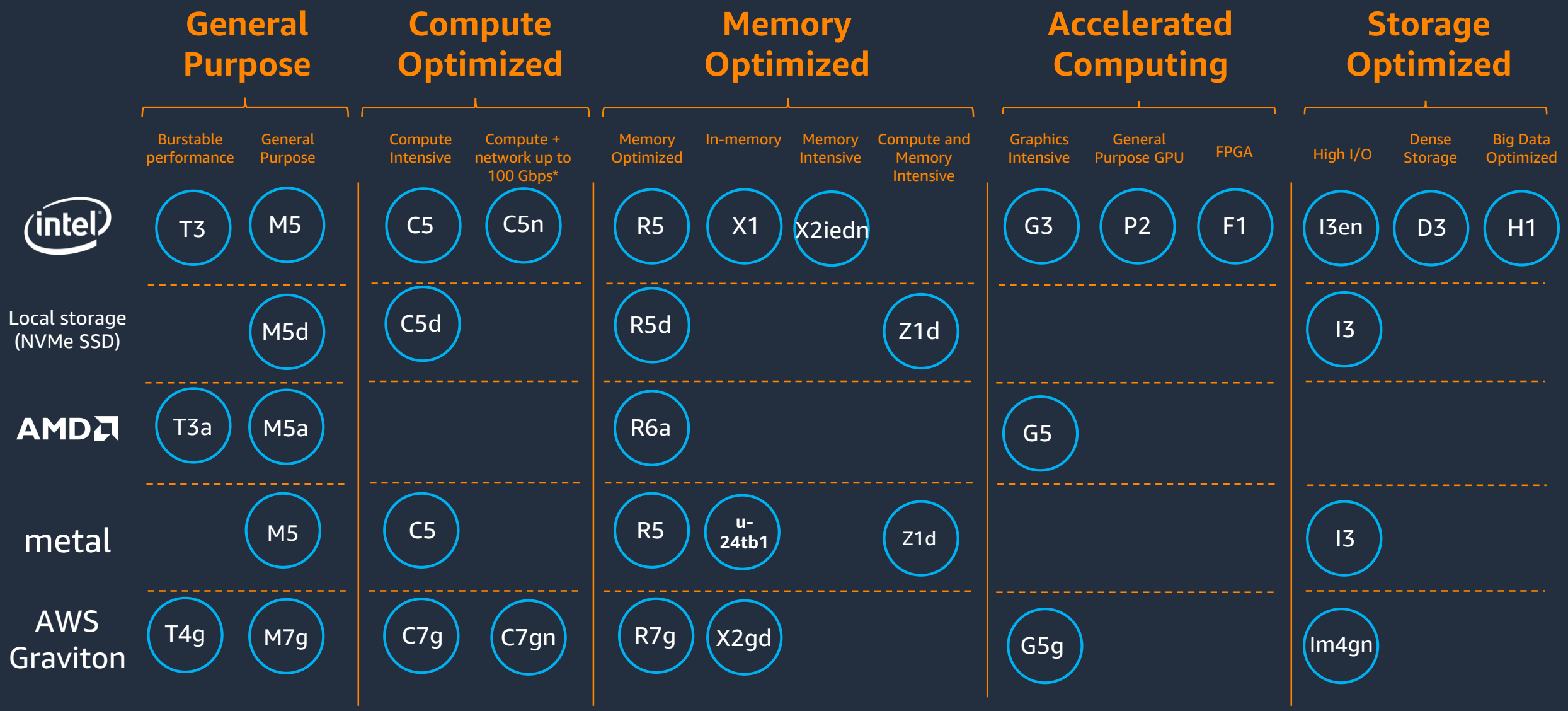
General purpose and workload optimized

Bare metal, disk, networking capabilities

Packaged | Custom | Community AMIs

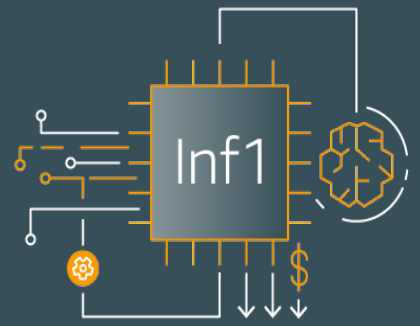
Multiple purchase options: On-Demand, Spot instances, Reserved Instances, Savings Plans, Dedicated Hosts

Instance Types



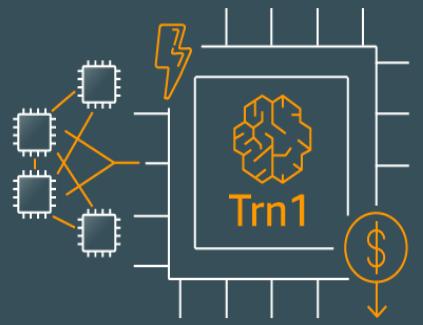
AWS chips optimized for deep learning

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

Instance Sizing

c7gn.8xlarge

2 – c7gn.4xlarge

4 – c7gn.2xlarge

8 – c7gn.xlarge

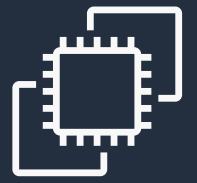
Choose your processor and architecture

Intel® Xeon® Scalable (Skylake) processor

NVIDIA V100 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

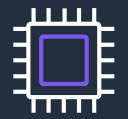
7x performance, 4x compute cores, and 5x faster memory

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

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

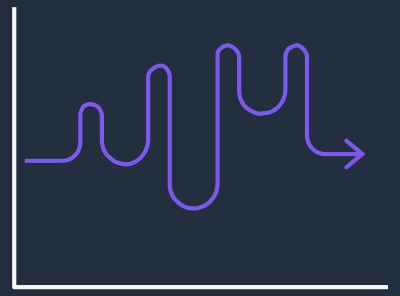
What is an Amazon Machine Image (AMI)?

- 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

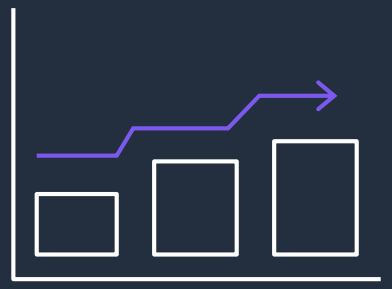
Pay for compute capacity by **the second** with no long-term commitments



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

Storage

You need the right storage for any data

OBJECT

Amazon S3

BLOCK

Amazon EBS

FILE

Amazon EFS

Amazon FSx for Windows File Server

Amazon FSx for Lustre

Amazon FSx for NetApp ONTAP

Amazon FSx for OpenZFS

BACKUP

AWS Backup

DATA TRANSFER AND EDGE PROCESSING

AWS Storage Gateway

AWS DataSync

AWS Transfer Family

AWS Snowball Edge

AWS Snowcone

Your choice of Amazon S3 storage classes

S3 Intelligent-Tiering

S3 Standard

S3 Standard-IA

S3 Glacier Instant Retrieval

S3 Glacier Flexible Retrieval

S3 Glacier Deep Archive

S3 One Zone-IA

S3 Outposts

AWS Region ≥ 3 Availability Zones

Changing access patterns

- Milliseconds access
- No retrieval charge
- **Archive Instant Access tier**

Frequently accessed data

- Milliseconds access
- No retrieval charge

Infrequently accessed data

- Milliseconds access
- Per-GB retrieval charge

Rarely accessed data

- Milliseconds access
- Per-GB retrieval charge

Archive data

- Retrieval options from minutes to hours
- **Free bulk retrievals**

Long term archive data

- Retrieval in hours

AWS AZ

Re-creatable, infrequently accessed data

- Milliseconds access
- Per-GB retrieval charge

AWS Outposts

On-premises data

- Milliseconds access

S3 Block Public Access

- Applies blanket protection against accidental public access
- Can be applied to ACL access, bucket policy access, or both
- Can be applied to new data or to all data
- Set at the bucket level or at the account level

S3 Block Public Access Settings

Amazon Elastic Block Store

Easy to use, high performance block storage at virtually any scale

Performance for any workload

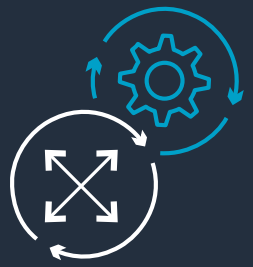
Up to 256,000 IOPS / volume
260,000 IOPS / server
As low as sub-millisecond latency

Easy to use

Easily add/remove capacity, or change volume types with Elastic Volumes

High reliability

99.999% availability
Up to 99.999% durability
Crash consistent Snapshots



Virtually unlimited scale

Use a single gigabyte or less, or scale up to petabytes of data

Secure

Encrypt all new volumes and data for a region by default with a single setting

Cost-effective

Pay as low as $0.015/GB-month for highly cost-effective block storage

EBS Snapshots

What are EBS snapshots

Point-in-time copy of EBS volume

Incremental copies
Only changed blocks are copied

Stored in S3
11x9's durability

Crash consistent

Contains all data necessary to recover a volume

Snapshot use-cases

Backup data on EBS volumes

Meet RPO and RTO SLA objectives

Copy volumes to another region for Disaster Recovery

Re-deploy production data for test/dev

Amazon Elastic File System (Amazon EFS)

Simple, serverless, set-and-forget, CLOUD Native, elastic file system for AWS compute

CLOUD NATIVE

Built from ground up in the cloud to be simple at cloud-scale and be fully elastic and serverless

Secured and protected using cloud constructs like AWS IAM, EFS Replication & AWS Backup

Consumable from any AWS computing model, supports tens of thousands of concurrent connections

Amazon EFS

4 Classes of Storage

1Zone | 1Zone IA | MAZ | MAZ IA

Highly durable – 11 9's

Highly available – 4 9's

PERFORMANT

Latencies as low as 250 micro-seconds

Up to 500 K IOPS, 10+ GB/s of thruput

2 Performance modes GP | MaxIO

Provision Throughput

COST OPTIMIZED

Lifecycle-based cost optimization

Intelligent Tiering (INT)

Blended storage costs as low as:
$0.043 GB/month

Amazon FSx for Windows File Server: fully featured, secure, reliable, and scalable

Accessibility
✓ Full SMB protocol support
✓ Windows Server 2008+, Windows 7+, Linux, and MacOS
✓ EC2, WorkSpaces and AppStream 2.0
✓ VMware Cloud on AWS
✓ ECS and EKS containers
✓ Cross-VPC/Account/Region access
✓ On-premises access (DirectConnect/VPN)

Administration
✓ Active directory integration
✓ Managing file shares
✓ Monitoring user sessions and open files
✓ Restoring locked files
✓ User storage quotas
✓ Monitoring actions via CloudTrail

Availability and durability
✓ High availability: automatic recovery
✓ High durability: automatic replication
✓ Multi-AZ deployment option
✓ SMB continuous availability (CA)

Performance and scale
✓ Consistent, sub-millisecond latencies
✓ PB-scale storage scalability
✓ Tens of GB/s throughput scalability
✓ Millions of IOPS scalability
✓ Select throughput and storage independently
✓ Server-side and client-side caching
✓ SMB Multichannel
✓ Perf. monitoring via CloudWatch
✓ Live scaling of throughput capacity

Cost optimization
✓ Storage type flex. (SSD/HDD)
✓ Deployment type (single-AZ/multi-AZ)
✓ Live scaling of storage and throughput capacity
✓ Data deduplication and compression

Data protection
✓ Snapshots (with end-user file restore)
✓ Backups

Security and compliances
✓ Encryption at rest and in transit
✓ Kerberos authentication
✓ Access controls via NTFS ACLs, share ACLs, VPC, and IAM
✓ File Access Auditing
✓ PCI DSS, ISO, SOC, GDPR, IRAP, and HIPAA compliances

What is Amazon FSx for NetApp ONTAP?

Complete NetApp ONTAP file systems

With the simplicity, agility, & scalability of an AWS service

- Native AWS service
- Built with NetApp

- Runs on AWS infrastructure
- Supported by AWS

Databases

Fully Managed Services on AWS

Automate undifferentiated heavy lifting

Self Managed

You

- Schema design
- Query construction
- Schema design
- Automatic fail-over
- Backup & recovery
- Isolation & security
- Industry compliance
- Push-button scaling
- Automated patching
- Advanced monitoring
- Routine maintenance
- Built-in best practices

Fully Managed

You

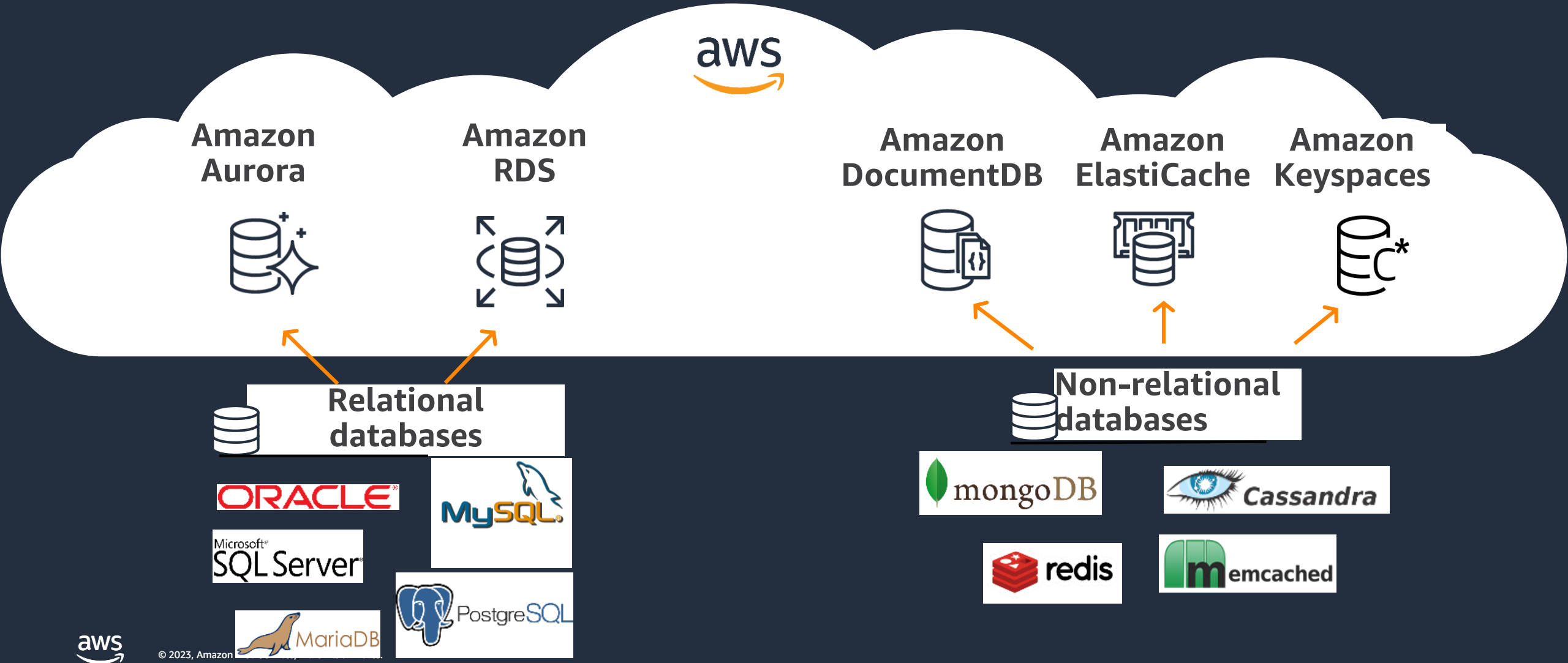
- Schema design
- Query construction
- Query optimization

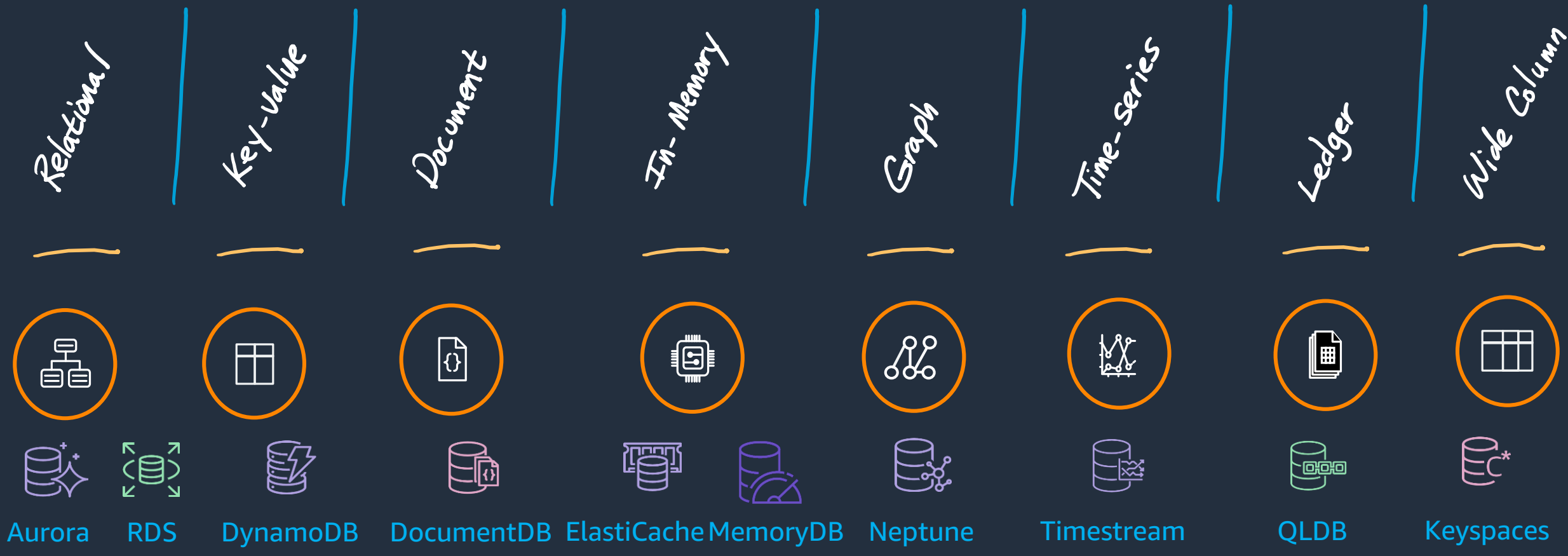
aws

- Automatic fail-over
- Backup & recovery
- Isolation & security
- Industry compliance
- Push-button scaling
- Automated patching
- Advanced monitoring
- Routine maintenance
- Built-in best practices

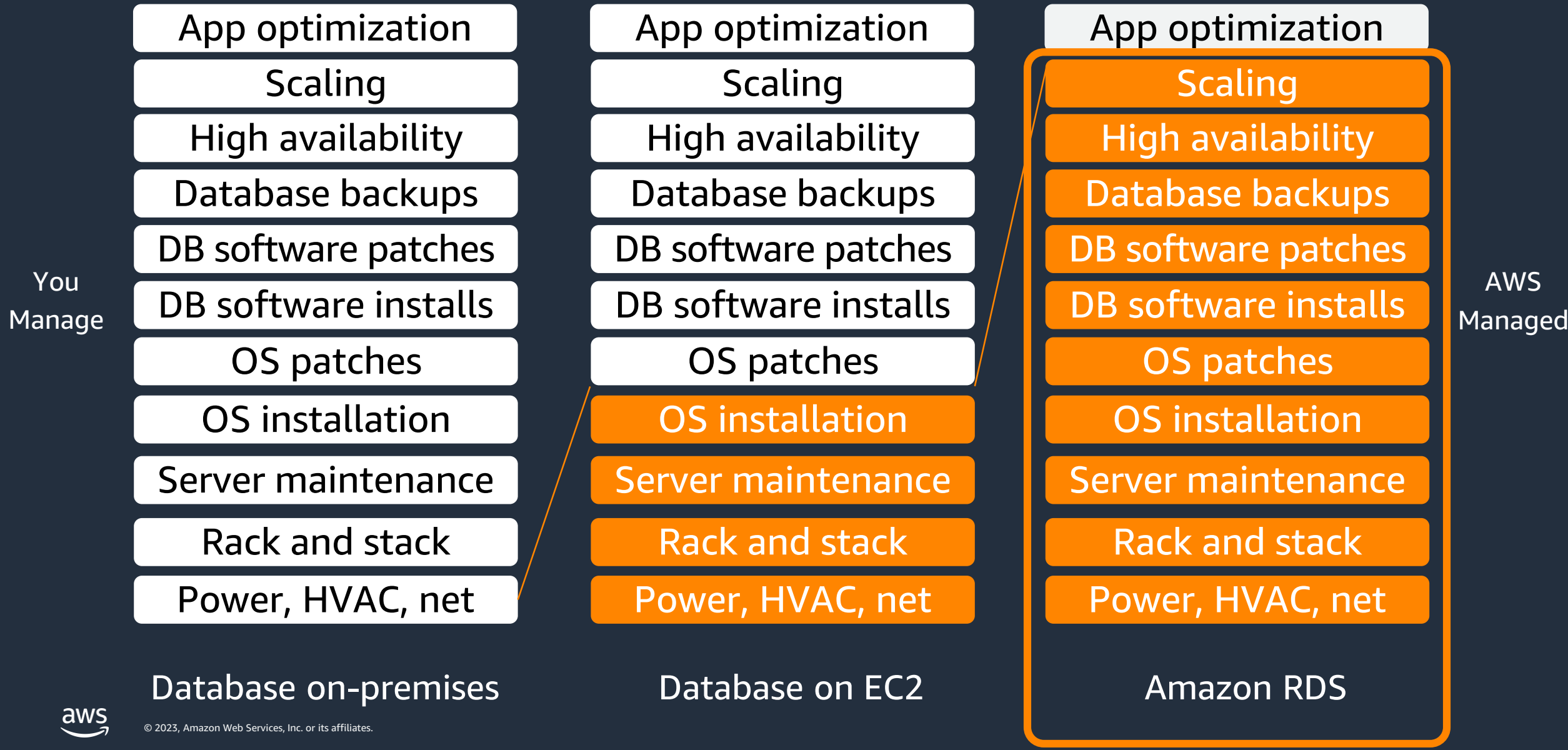
Move to Fully Managed Databases

Migrate on-premises or self-managed databases to fully managed services





Everything you get from Amazon RDS...



Amazon Aurora

MySQL and PostgreSQL-compatible relational database built for the cloud

Performance and availability of commercial-grade databases at 1/10th the cost

Performance and scalability

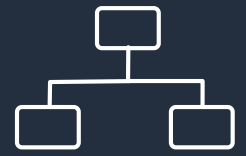
5x throughput of standard MySQL and 3x of standard PostgreSQL; scale-out up to 15 read replicas

Availability and durability

Fault-tolerant, self-healing storage; six copies of data across three Availability Zones; continuous backup to Amazon S3

Highly secure

Network isolation, encryption at rest and in transit



Fully managed

Managed by RDS: no server provisioning, software patching, setup, configuration, or backups

DynamoDB

Fast and flexible key-value database service for any scale

Performance at scale

Consistent, single-digit-millisecond response times at any scale; build applications with virtually unlimited throughput

Serverless architecture

No hardware provisioning, software patching, or upgrades; scales up or down automatically; continuously backs up your data

Enterprise security

Encrypts all data by default and fully integrates with AWS Identity and Access Management for robust security

Global replication

Build global applications with fast access to local data by easily replicating tables across multiple AWS Regions

Amazon ElastiCache

Managed Redis or Memcached-compatible in-memory caching service

µs is the new *ms*

Unlimited scale

Read scaling with replicas. Write and memory scaling with sharding. Nondisruptive scaling.

Consistent high performance

In-memory data store and cache for submillisecond response times

Fully managed

AWS manages all hardware and software setup, configuration, and monitoring

Amazon DocumentDB

Fast, scalable, highly available MongoDB-compatible database service

Millions of requests per second, millisecond latency

Same code, drivers, and tools you use with MongoDB

Simple and fully managed

Secure and compliant

2x throughput of managed MongoDB services

Deeply integrated with AWS services

Security

AWS Security, Identity, and Compliance Solutions

Identity and access management

AWS Identity and Access Management (IAM)

AWS IAM Identity Center (successor to AWS SSO)

AWS Organizations

AWS Directory Service

Amazon Cognito

AWS Resource Access Manager

Detective controls

AWS Security Hub

Amazon GuardDuty

Amazon Inspector

Amazon CloudWatch

AWS Config

AWS CloudTrail

VPC Flow Logs

AWS IoT Device Defender

Infrastructure protection

AWS Firewall Manager

AWS Network Firewall

AWS Shield

AWS WAF

Amazon VPC

AWS PrivateLink

AWS Systems Manager

Data protection

Amazon Macie

AWS Key Management Service (KMS)

AWS CloudHSM

AWS Certificate Manager

AWS Secrets Manager

AWS VPN

Server-Side Encryption

Incident response

Amazon Detective

Amazon EventBridge

AWS Backup

AWS Security Hub

AWS Elastic Disaster Recovery

Privacy and Compliance

AWS Artifact

AWS Audit Manager

Amazon CloudWatch

AWS CloudTrail

AWS Config

AWS Security Hub

AWS Systems Manager

AWS Identity and Access Management (IAM)

SECURELY CONTROL ACCESS TO AWS SERVICES AND RESOURCES FOR YOUR USERS.

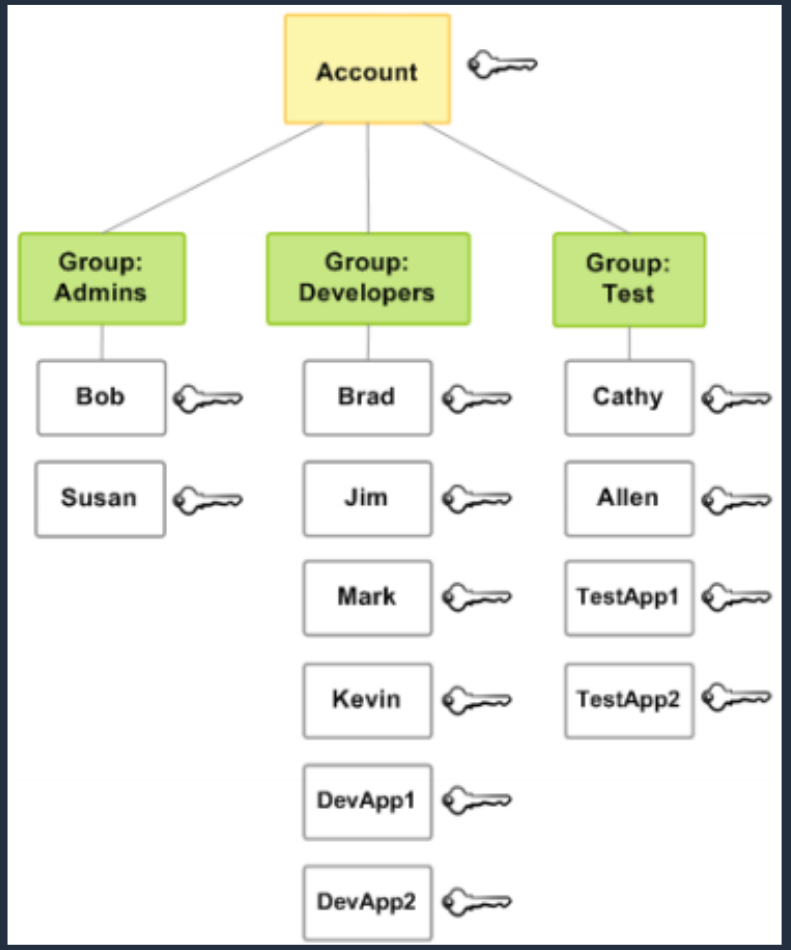
Username/User

Manage groups of users

Centralized Access Control

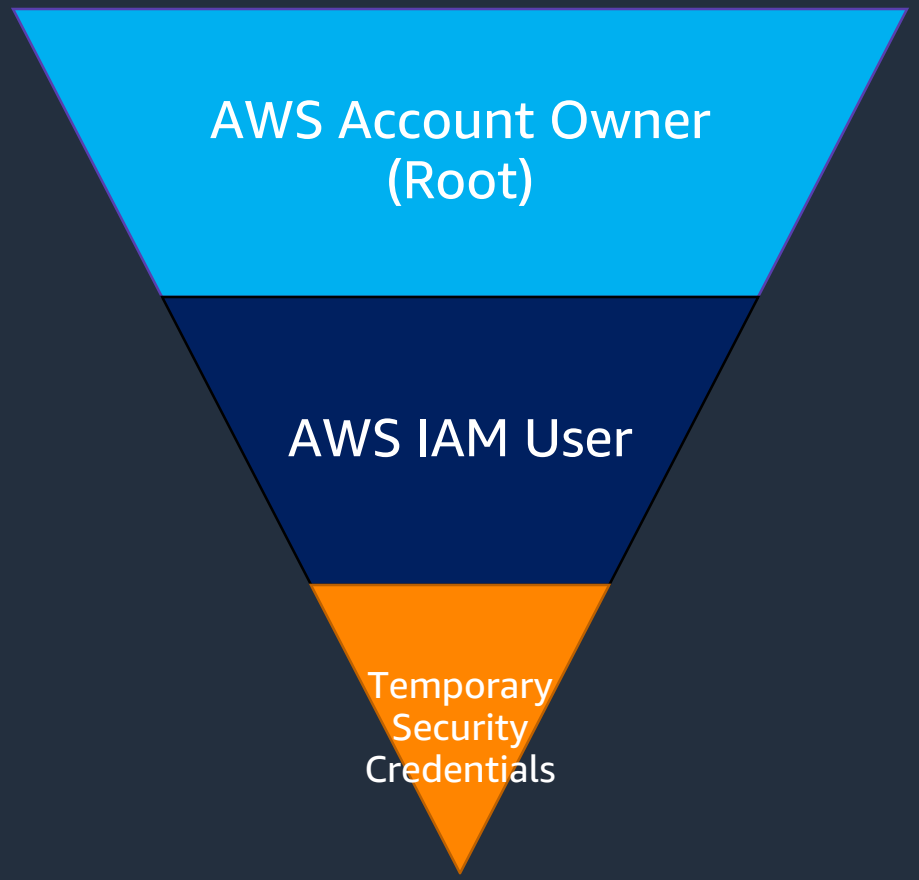
Optional Configurations:

- Password for console access.
- Policies for controlling access AWS APIs.
- Two methods to sign API calls:
  - X.509 certificate
  - Access/Secret Keys
- Multi-factor Authentication (MFA)



AWS IAM Hierarchy of Privileges

ENFORCE PRINCIPLE OF LEAST PRIVILEGE WITH IDENTITY AND ACCESS MANAGEMENT (IAM) USERS, GROUPS, AND POLICIES AND TEMPORARY CREDENTIALS.



Permissions	Example

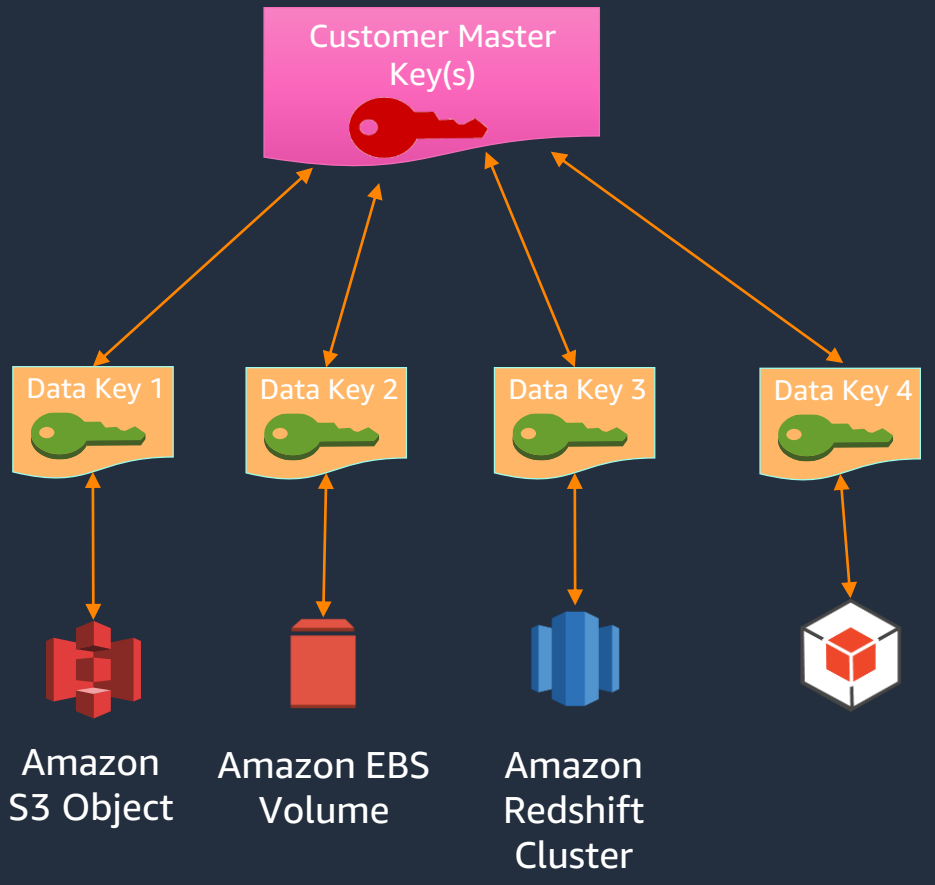
Unrestricted access to all enabled services and resources.	`Action: *`
	`Effect: Allow`
	`Resource: *`
	`(implicit)`

Access restricted by Group and User policies	`Action:`
	`[‘s3:*’,’sts:Get*’]`
	`Effect: Allow`
	`Resource: *`

Access restricted by generating identity and further by policies used to generate token	`Action: [ ‘s3:Get*’ ]`
	`Effect: Allow`
	`Resource:`
	`‘arn:aws:s3:::mybucket/*’`

AWS Key Management Service

MANAGED SERVICE TO CREATE AND MANAGE CRYPTOGRAPHIC KEYS AND CONTROL THEIR USE ACROSS A WIDE RANGE OF AWS SERVICES AND IN YOUR APPLICATIONS.



AWS CloudTrail

CLOUDTRAIL PROVIDES EVENT HISTORY OF YOUR AWS ACCOUNT ACTIVITY, INCLUDING ACTIONS TAKEN THROUGH THE AWS MANAGEMENT CONSOLE, AWS SDKS, COMMAND LINE TOOLS, AND OTHER AWS SERVICES.

Who?	When?	What?	Where to?	Where from?
Bill	3:27pm	Launch Instance	us-west-2	72.21.198.64
Alice	8:19am	Added Bob to admin group	us-east-1	127.0.0.1
Steve	2:22pm	Deleted DynamoDB table	eu-west-1	205.251.233.176


```
{
"Records": [
        {
        "eventVersion": "1.0",
        "userIdentity": {
            "type": "IAMUser",
            "principalId": "EX_PRINCIPAL_ID",
            "arn": "arn:aws:iam::123456789012:user/Alice",
            "accountId": "123456789012",
            "accessKeyId": "EXAMPLE_KEY_ID",
            "userName": "Alice",
            "sessionContext": {
                "attributes": {
                    "mfaAuthenticated": "false",
                    "creationDate": "2014-03-25T18:45:11Z"
                }
            }
        },
        "eventTime": "2014-03-25T21:08:14Z",
        "eventSource": "iam.amazonaws.com",
        "eventName": "AddUserToGroup",
        "awsRegion": "us-east-1",
        "sourceIPAddress": "127.0.0.1",
        "userAgent": "AWSConsole",
        "requestParameters": {
            "userName": "Bob",
            "groupName": "admin"
        },
        "responseElements": null
    },
    ...additional entries
  ]
}
```

AWS CloudWatch

MONITORING SERVICES FOR AWS RESOURCES AND AWS-BASED APPLICATIONS.

What does it do?

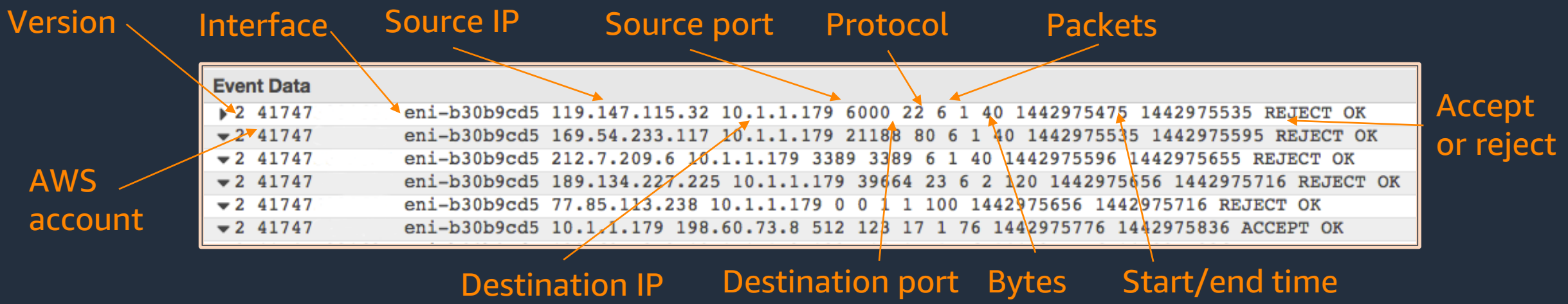
- Collect and Track Metrics
- Monitor and Store Logs
- Set Alarms (react to changes)
- View Graphs and Statistics

How can you use it?

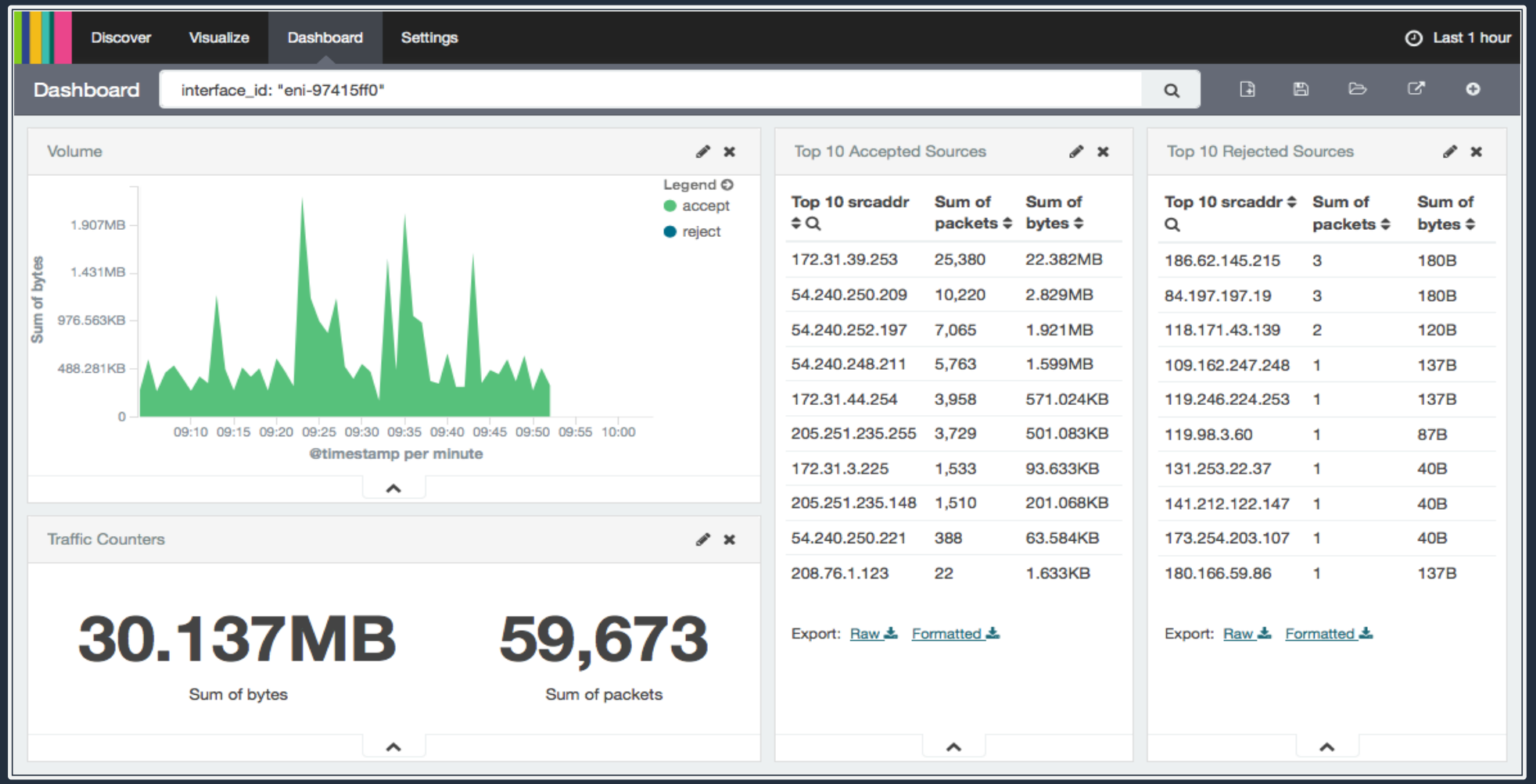
CloudWatch Metrics	Monitor CPU, Memory, Disk I/O, Network, etc.
CloudWatch Logs / CloudWatch Events	React to application log events and availability
CloudWatch Alarms	Automatically scale EC2 instance fleet
CloudWatch Dashboards	View Operational Status and Identify Issues

VPC Flow Logs

- Agentless
- Enable per ENI, per subnet, or per VPC
- Logged to AWS CloudWatch Logs
- Create CloudWatch metrics from log data
- Alarm on those metrics



VPC Flow Logs



What is Amazon GuardDuty?

Amazon GuardDuty is a threat detection service that uses **machine learning**, anomaly detection, and **integrated threat intelligence** to identify and prioritize potential threats.

One-step activation

Continuous monitoring of AWS accounts and resources

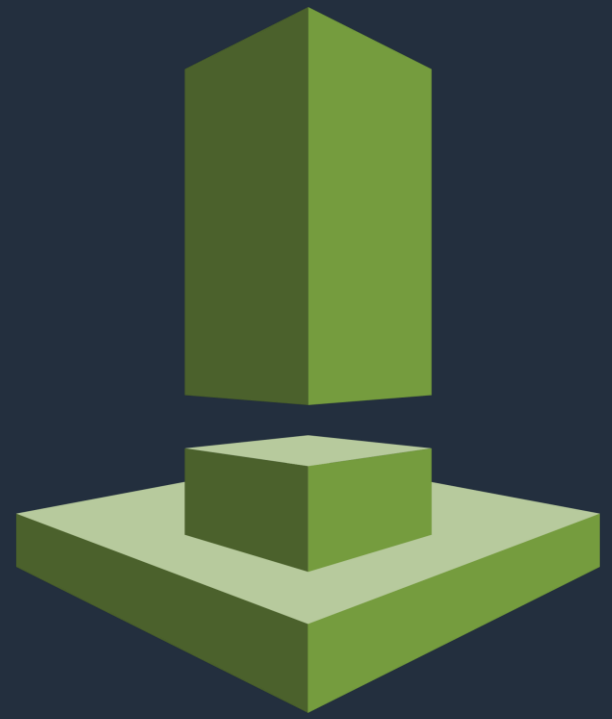
Global coverage with regional results

Detect known and unknown threats

Enterprise-wide consolidation & management

Amazon Inspector

- Vulnerability Assessment Service
  - Built from the ground up to support DevSecOps
  - Automatable via APIs
  - Integrates with CI/CD tools
  - On-Demand Pricing model
  - Static & Dynamic Rules Packages
  - Generates Findings



AWS WAF (Web Access Firewall)

Web Traffic Filtering with Custom Rules

Create custom rules that can block, allow or monitor requests based on IP address, HTTP headers, or a combination of both.

Malicious Request Blocking

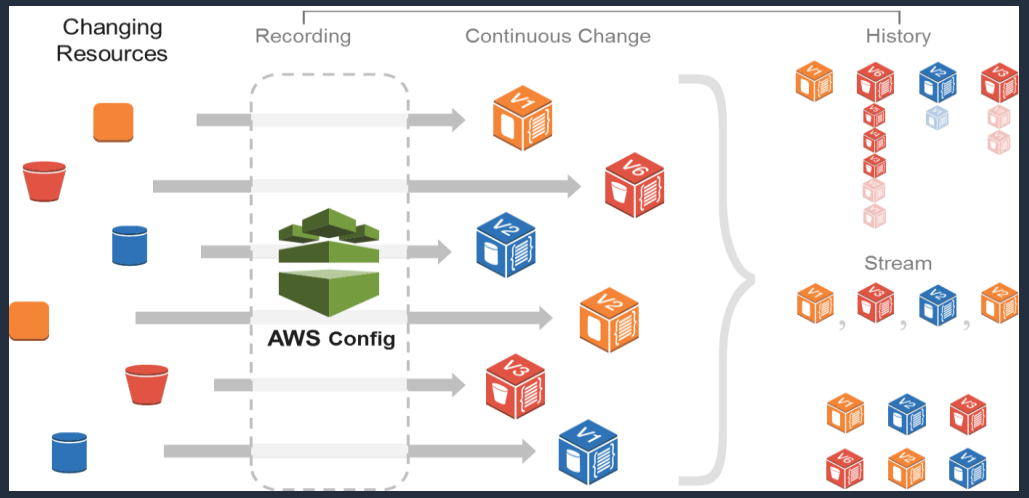
AWS WAF can recognize and block common web application security risks like SQL injection (SQLi) and cross-site scripting (XSS).

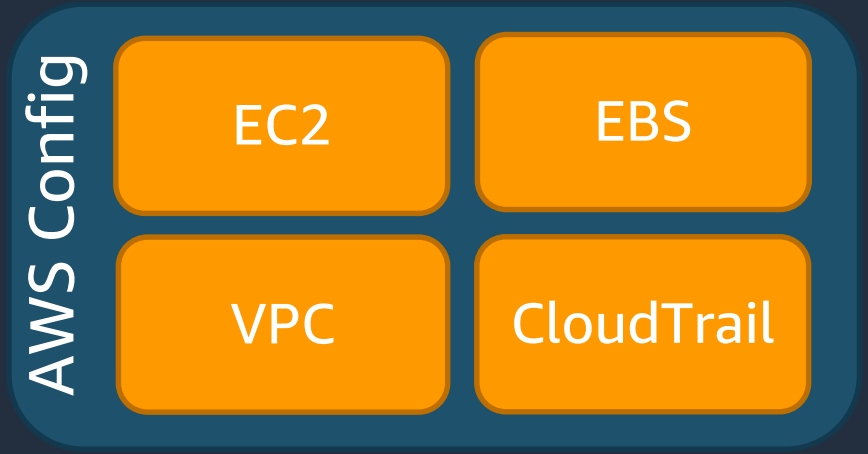
Active monitoring & tuning

Monitor and configure the requests that are being blocked and allowed by the Web ACL rules.

AWS Config

MANAGED SERVICE FOR TRACKING AWS INVENTORY AND CONFIGURATION, AND CONFIGURATION CHANGE NOTIFICATION.





Security Analysis

Audit Compliance

Change Management

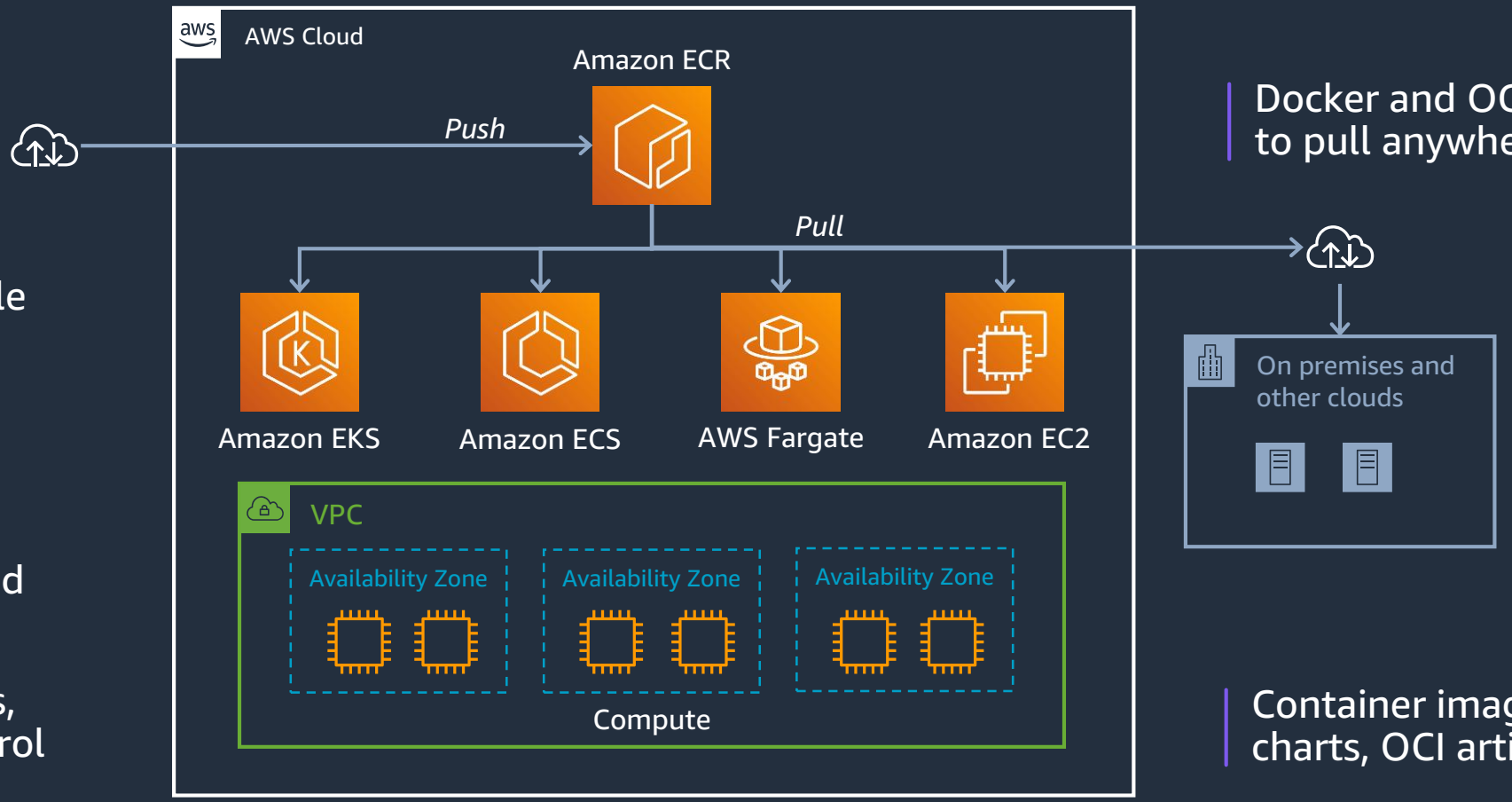
Troubleshooting

Discovery

Containers

Amazon ECR

Fully-managed container artifact registry



- Managed and scalable infrastructure
- Highly available, high performance
- Security with encrypted images and vulnerability scans
- Authenticated access, centralized IAM control

- Docker and OCI compliant to pull anywhere
- Container images, Helm charts, OCI artifacts

- Native integration to AWS orchestrators and compute

Which Orchestration Platform Should I Choose?

ECS

Powerful simplicity

EKS

Open flexibility

Powerful simplicity

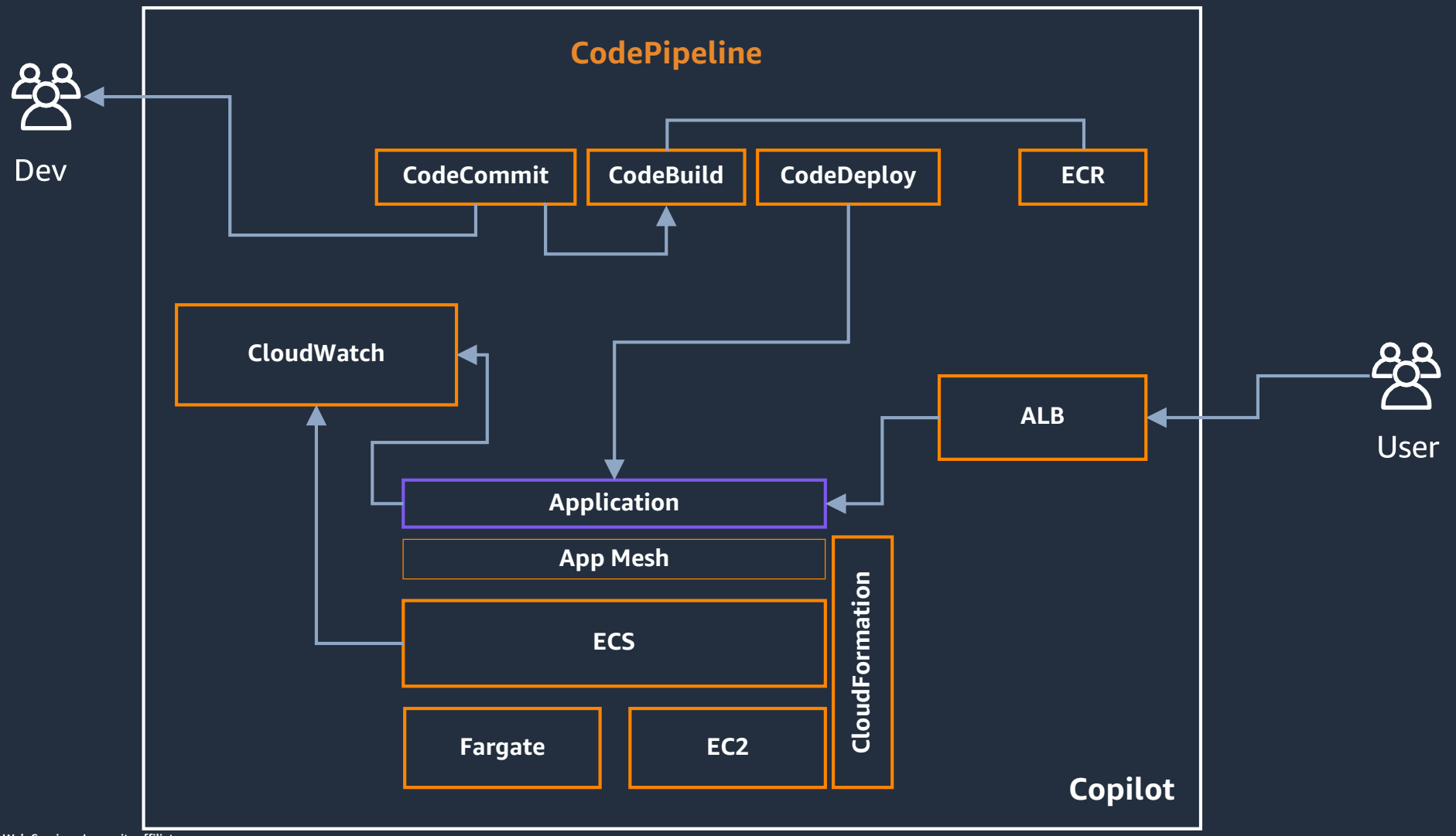
ECS

AWS-opinionated way to run containers at scale

Reduce decisions without sacrificing scale or features

Reduce time to build, deploy, and migrate applications

Powerful simplicity



Open flexibility

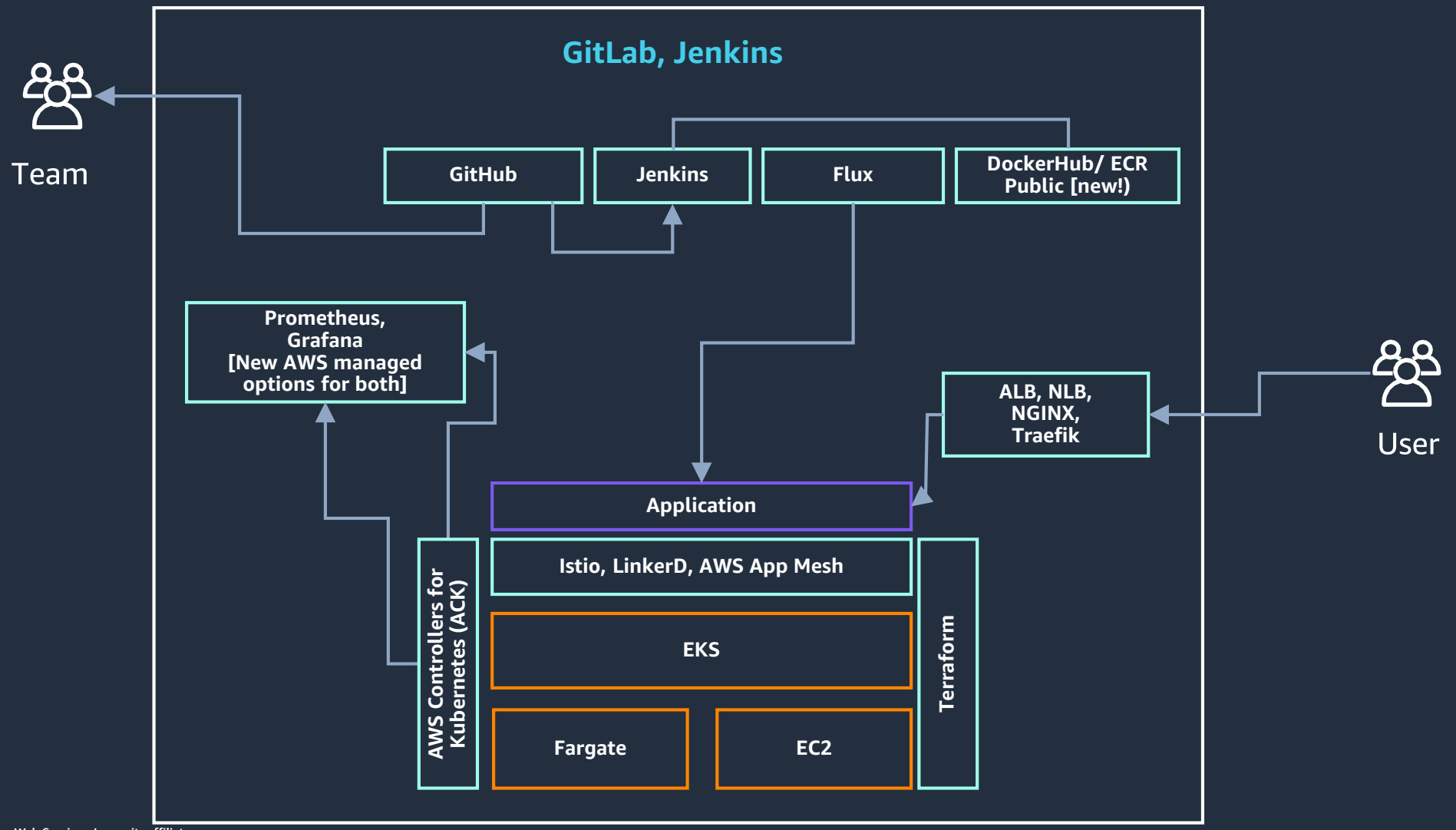
EKS

Gain agility and efficiency with AWS-optimized Kubernetes, and standardize operations everywhere

Secure, highly available, with observability across all Kubernetes deployments

Build with choice of solutions from the broader community around Kubernetes

Open flexibility



Run your containers anywhere based on your workload needs

Serverless

AWS Fargate

EC2 options

Amazon EC2

Edge and 5G

AWS Local Zones

AWS Wavelength

On-premises

AWS Outposts

EKS Anywhere
ECS Anywhere

Running Kubernetes clusters outside of AWS

EKS Distro

EKS Distro provides customers **version, patching, and security alignment** with Amazon EKS

EKS Anywhere

EKS Anywhere provides cluster creation and **lifecycle management** with a set of tools for EKS Distro and management through the EKS console

AI and ML

The AWS AI/ML stack

BROADEST AND MOST COMPLETE SET OF MACHINE LEARNING CAPABILITIES

AI SERVICES

SPECIALIZED

BUSINESS PROCESSES
- Amazon Personalize
- Amazon Forecast
- Amazon Fraud Detector
- Amazon Lookout for Metrics

SEARCH
- Amazon Kendra

CONVERSATION
- Amazon Lex
- Amazon Transcribe Call Analytics
- Contact Lens
- Voice ID

CODE + DEVOPS
- Amazon CodeGuru
- Amazon CodeWhisperer
- Amazon DevOps Guru

INDUSTRIAL
- Amazon Monitron
- Amazon Lookout for Equipment
- Amazon Lookout for Vision

HEALTH
- Amazon HealthLake
- Amazon Comprehend Medical
- Amazon Transcribe Medical
- Amazon Omics

CORE

TEXT
- Amazon Translate
- Amazon Comprehend

SPEECH
- Amazon Polly
- Amazon Transcribe

VISION
- Amazon Textract
- Amazon Rekognition
- AWS Panorama

AMAZON SAGEMAKER

CANVAS
No-code ML for business analysts

STUDIO LAB
Learn ML

GROUND TRUTH
Label data

STUDIO IDE
- Prepare data
- Geospatial ML
- Build with notebooks
- Train models
- Deploy in production
- Store features
- Tune parameters
- Manage and monitor

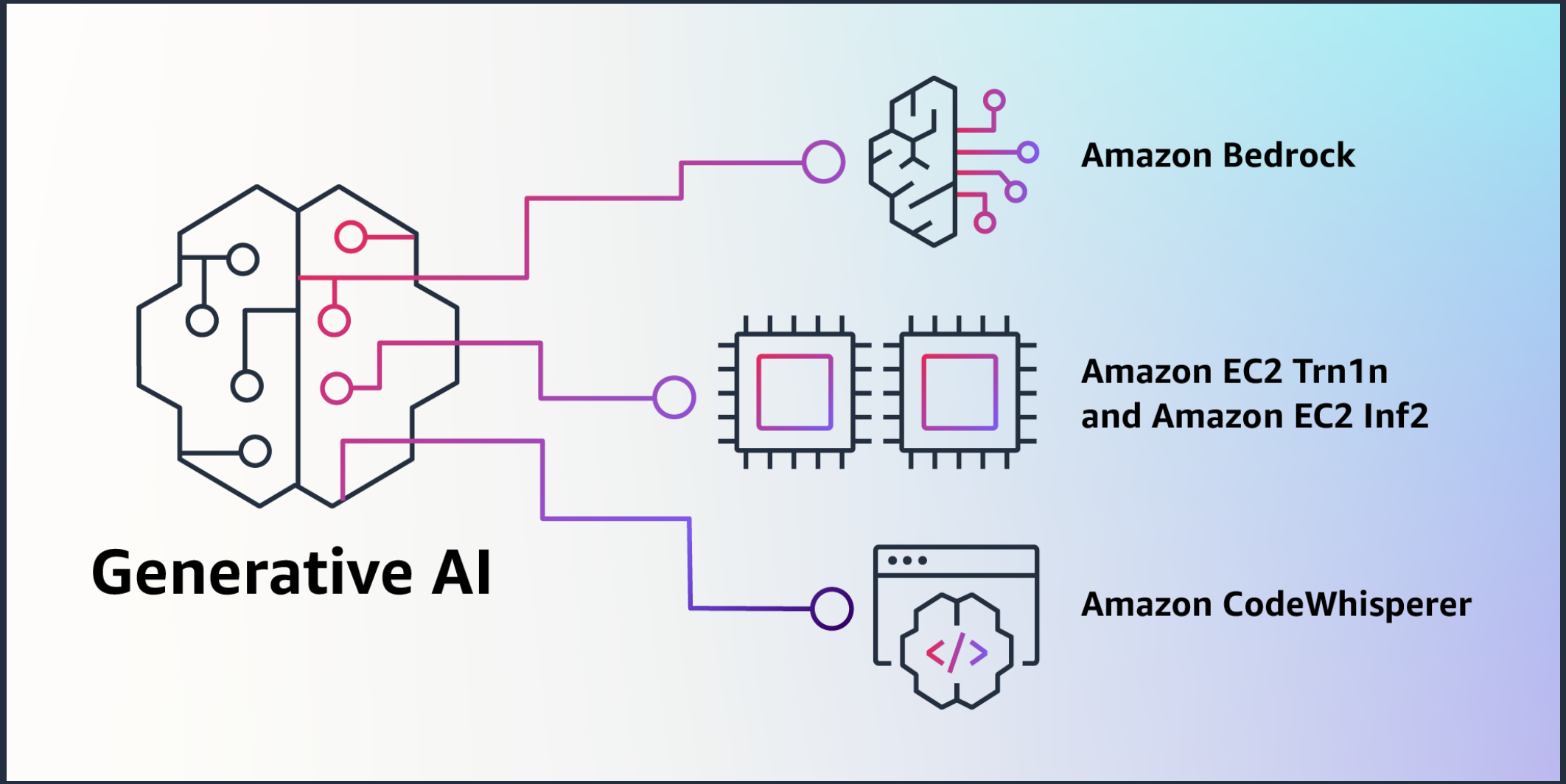
CI/CD | GOVERNANCE | RESPONSIBLE ML

EDGE MANAGER
Manage edge devices

ML FRAMEWORKS & INFRASTRUCTURE

PyTorch, Apache MXNet, TensorFlow | Amazon EC2 | CPUs | GPUs | AWS Inferentia | AWS Trainium | Habana Gaudi | FPGA

Building with generative AI on AWS



Build with publicly available foundation models

AVAILABLE ON SAGEMAKER JUMPSTART

AI21labs

Models

Jurassic-2 Ultra, Mid

Contextual answers

Summarize

Paraphrase

Grammatical error correction

Tasks

Text generation

Long-form generation

Summarization

Paraphrasing

Chat

Information extraction

Meta AI

Models

Llama 2 7B, 13B, 70B

Tasks

Question answering

Chat

Summarization

Paraphrasing

Sentiment analysis

Text generation

cohere

Models

Cohere Command XL

Tasks

Text generation

Information extraction

Question answering

Summarization

Hugging Face

Models

Falcon-7B, 40B

Open LlaMA

RedPajama

MPT-7B

BloomZ 176B

Flan T-5 models (8 variants)

DistilGPT2

GPT NeoXT

Bloom models (3 variants)

Tasks

Machine translation

Question answering

Summarization

stability.ai

Models

Stable Diffusion XL 1.0

2.1 base

Upscaling

Inpainting

Tasks

Generate photo-realistic images from text input

Improve quality of generated images

Features

Fine-tuning on Stable Diffusion 2.1 base model

LightOn

Models

Lyra-Fr 10B, Mini

Tasks

Text generation

Keyword extraction

Information extraction

Question answering

Summarization

Sentiment analysis

Classification

databricks

Models

Dolly

Tasks

Question answering

Chat

Summarization

Paraphrasing

Sentiment analysis

Text generation

alexa

Models

AlexaTM 20B

Tasks

Machine translation

Question answering

Summarization

Annotation

Data generation

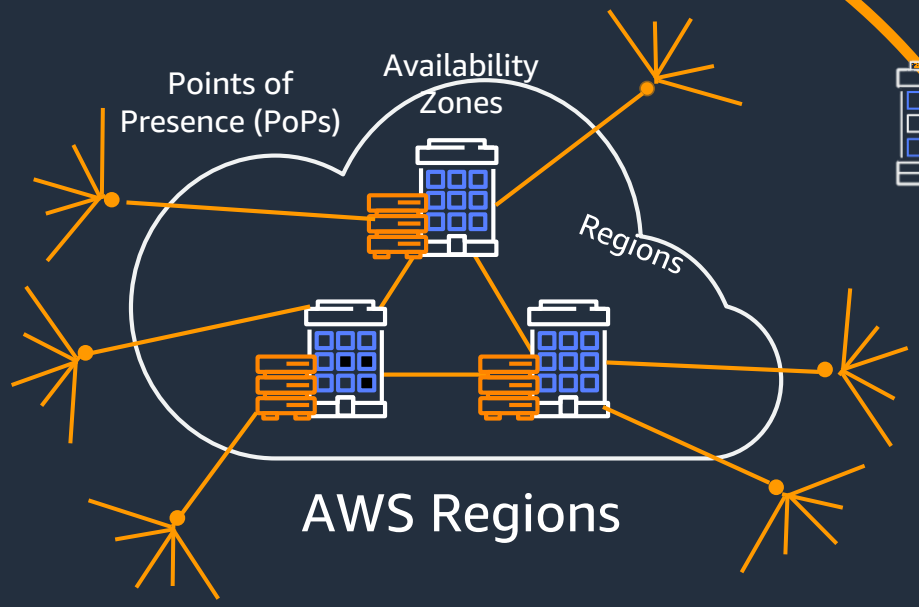
aws

Hybrid and Edge

AWS cloud-to-edge continuum

For most use cases

For low latency, local data processing, and data residency



On premises

AWS Outposts
AWS Storage Gateway

Metro centers

AWS Local Zones

5G networks

AWS Wavelength

IoT / robotics

AWS IoT Greengrass
AWS IoT SiteWise
FreeRTOS
AWS RoboMaker

Rugged, Mobile edge

AWS Snowball Edge
AWS Snowcone
AWS Snowblade
AWS Modular Data Center (MDC)

Cloud continuum

AWS Outposts Family

Outposts rack

Bring the same AWS APIs, services, and features to virtually any datacenter or co-location space

VMware Cloud™ on AWS Outposts

Deliver fully managed VMware Cloud on Outposts rack

Outposts servers

Run Outposts in locations with limited space or smaller capacity requirements

AWS Outposts rack

Industry standard 42U rack

Fully assembled, ready to be rolled into final position

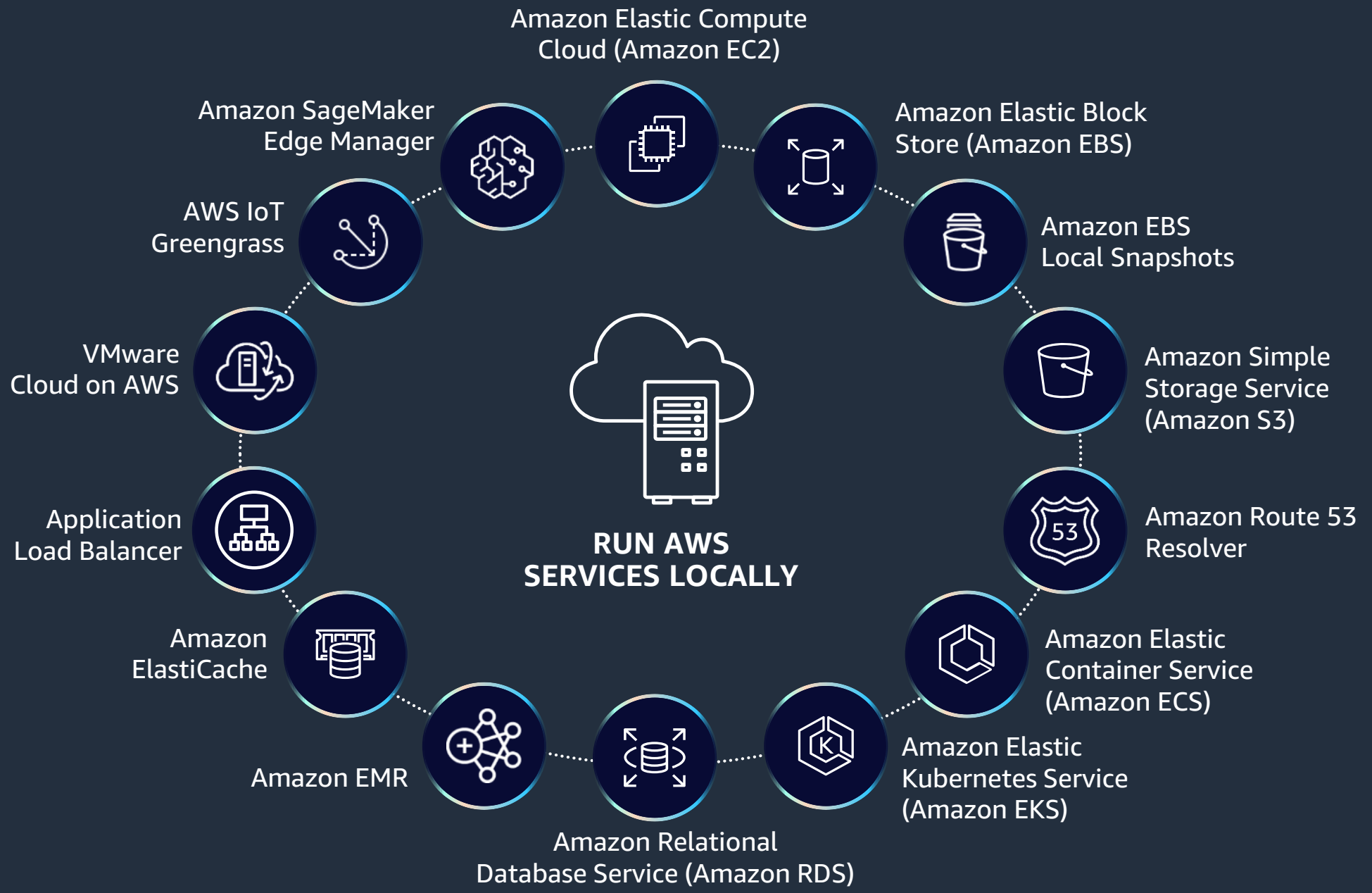
Installed by AWS, simply plugged into power and network

Centralized redundant power conversion unit and DC distribution system for higher reliability, energy efficiency, easier serviceability

Redundant active components including top of rack switches

© 2023, Amazon Web Services, Inc. or its affiliates.

AWS Outposts services today



AWS Outposts rack vs. VMware Cloud on AWS Outposts

AWS Outposts rack

Same AWS APIs, services, and features as in the AWS Regions

EC2 and EBS with support for services including S3, RDS, ECS, EKS, EMR, ALB, and others

VMware Cloud on AWS Outposts

VMware APIs and services to leverage existing skills, automation, and governance policies

For customers running VMware SDDC on premises

More information on VMware Cloud on AWS Outposts is available in the VMware Cloud on AWS Outposts First Call Deck

AWS Outposts servers

RUN AWS OUTPOSTS IN LOCATIONS WITH LIMITED SPACE OR CAPACITY REQUIREMENTS

AWS OUTPOSTS FAMILY

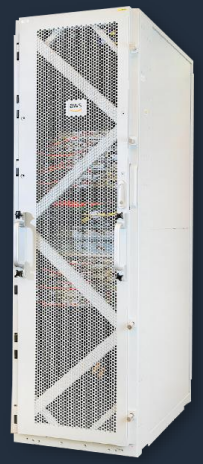
1U SINGLE SERVER

1¾ inches tall

2U SINGLE SERVER

3½ inches tall

42U FULL RACK



80 inches tall

Offers the same AWS infrastructure, services, APIs, and tools on premises, now with a smaller form factor

Choose between a 1U form factor with an AWS Graviton2 processor or a 2U form factor with an Intel processor

Run AWS services locally, including Amazon EC2, Amazon ECS, and edge services like AWS IoT Greengrass

Ideal for workloads that require low latency and local processing needs

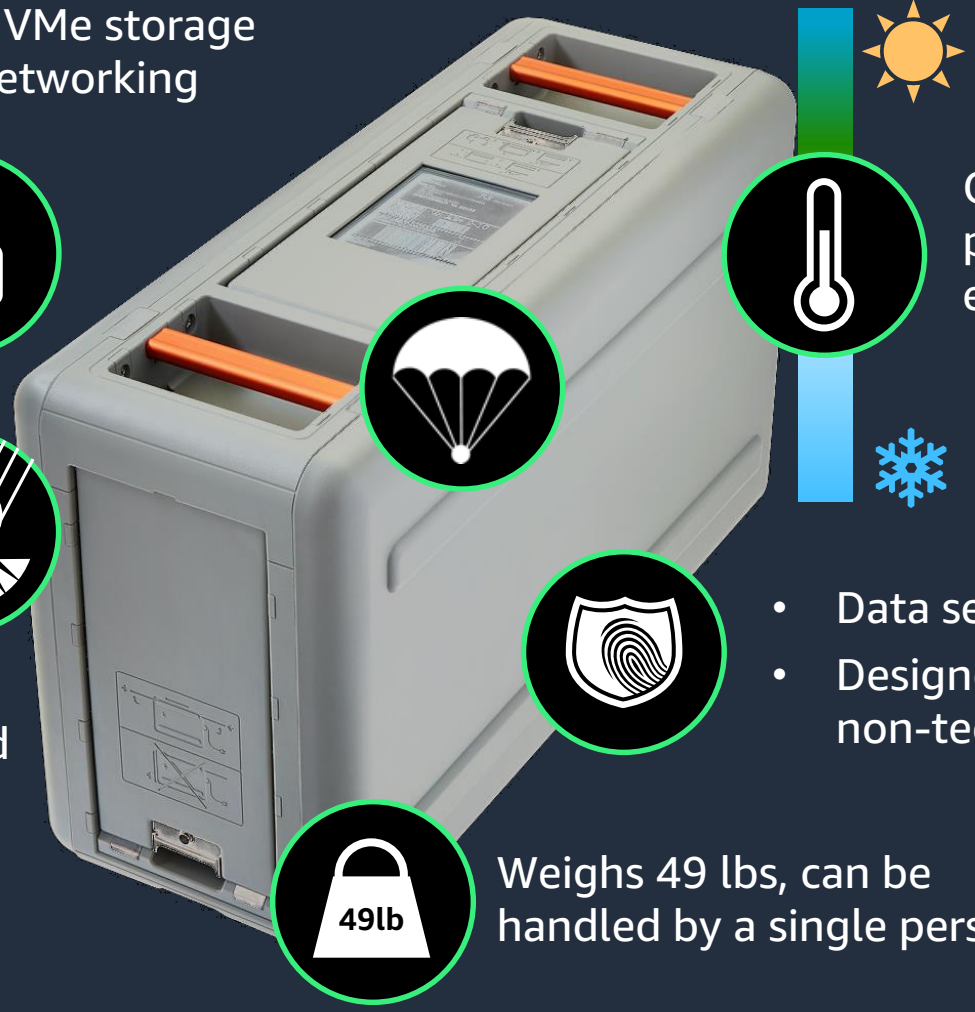
Simple device installation by either your own on-premises personnel or a third-party vendor

AWS Snowball Edge for Rugged and Tactical Edge

BRINGS AWS TO RUGGED EDGE AND LOCATIONS WITH NO OR LIMITED NETWORK CONNECTIVTY

Edge Compute and Edge Storage use cases:
32 - 104 vCPU, 208 - 416 GB RAM
28 – 210 TB NVMe storage
Up to 100G Networking

Tamper-resistant
Tamper-evident



Operates on unconditioned power and in non-datacenter environments (0° to 45°C)

- Handles up to **200G impact**
- Airdrop-able
- MIL-S-901D
- Meets FISMA High, FedRAMP ITAR, CJIS, and DoD SRG Impact Level 6 requirements

- Data secured with 256-bit encryption
- Designed to be operated by non-technical personnel

Weighs 49 lbs, can be handled by a single person

Snowball Edge Models

	Storage Optimized (data transfer)	Storage Optimized 210TB	Storage Optimized (with EC2)	Compute Optimized	Compute Optimized SSD
vCPU	40	104	40	52	104
Memory	80 GB	416 GB	80 GB	208 GB	416 GB
HDD	80 TB	N/A	80 TB	39.5 TB	N/A
SSD	1 TB SATA	210 TB NVMe	1 TB SATA	7.68 TB NVMe	28 TB NVMe
GPU	N/A	N/A	N/A	optional NVIDIA V100	N/A
NIC	2 x 10 Gbit – RJ45 1 x 25 Gbit – SFP28 1 x 100 Gbit – QSFP28	2 x 10 Gbit – RJ45 1 x 25 Gbit – SFP28 1 x 100 Gbit – QSFP28	2 x 10 Gbit – RJ45 1 x 25 Gbit – SFP28 1 x 100 Gbit – QSFP28	2 x 10 Gbit – RJ45 1 x 25 Gbit – SFP28 1 x 100 Gbit – QSFP28	2 x 10 Gbit – RJ45 1 x 25 Gbit – SFP28 1 x 100 Gbit – QSFP28
AWS Services*	EC2, EBS, S3, IAM, EKS-A, Tape Gateway, IoT Greengrass, Datasync, Lambda, Sitewise Edge, Sagemaker Edge				
Capabilities	Remote monitoring and management, multi-node, disconnected operations, Direct Network Interface				

AWS Snowcone for mobile workloads

SMALL, ULTRA-PORTABLE, RUGGED, AND SECURE EDGE COMPUTING AND DATA TRANSFER DEVICE

Capabilities:
2vCPU, 4 GB RAM
8 TB HDD or 14 TB SSD
1 or 10G Networking

Tamper-resistant
Tamper-evident

Operates on unconditioned power and in non-datacenter environments (0° to 45°C)

Handles vibration up to 50G RMS

Data secured with 256-bit encryption

Designed to be operated by non-technical personnel

Weighs 4 lbs and can be carried in a backpack

4 lb

AWS Snowblade

Brings AWS to the operational edge across Denied, Disrupted, Intermittent, Limited (DDIL) environments through the JWCC contract

Operates on unconditioned power and in non-datacenter environments (0° to 55°C)

Data secured with 256-bit encryption
FIPS compliance, NEBS-3 certification
2-layer encryption at Transport
FedRamp-Moderate and FedRamp-High
IL5 and IL6 certification

Tamper-resistant
Tamper-evident

50lb

Weighs 50 lbs (without enclosure), can be handled by a single person

208 vCPU, 832 GB RAM, 64 TB NVMe
4 server blades for redundant workloads
2 power-supplies
4 dual-port network switches
16TB/blade
Redundant interfaces (2 USB, 2 Ethernet)

DoD Operational Edge: [Snow Warrior] in optional TEMPEST enclosure

DoD Operational Edge: [Snow Warrior] in optional MIL-STD-810H enclosure

AWS Module Data Center

Brings AWS to the operational edge across Denied, Disrupted, Intermittent, Limited (DDIL) environments through the JWCC contract

- Ships as two 20’ ISO-standard containers
- Supports five 15kVA Snow or Outpost racks plus a sixth networking rack
- Redundant cooling
- Power distribution, takes up to two 400/480V, 50/60 Hz, 3-phase
- Intrusion detection, access control, fire suppression, monitoring
- Available in Unclassified and Classified variants (accreditable up to TS/SCI)

DevOps

Bringing DevOps practices to infrastructure

Teams with higher evolved DevOps practices:

Change management effectiveness

3x more effective

Restore service after incident less than a day

77% of teams

Remediate vulnerability in less than a day

60% of teams

What is “GitOps”?

GitOps: is an approach where infrastructure is defined as code (IaC) alongside application code, versioned in a git repository, and then provisioned following the same merge request process as application software code

Infrastructure as code
hosted just like software code

+

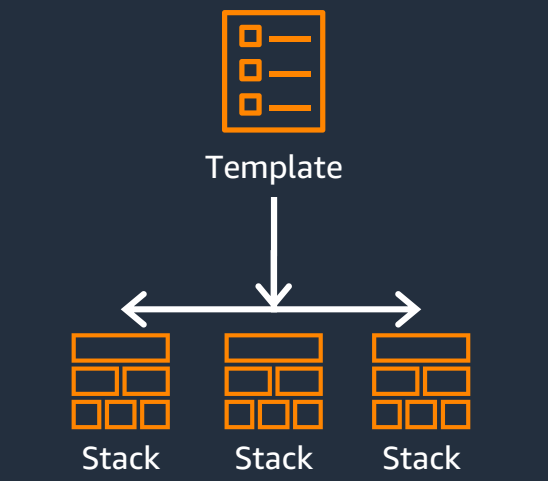
Git
versioning and pull requests workflows

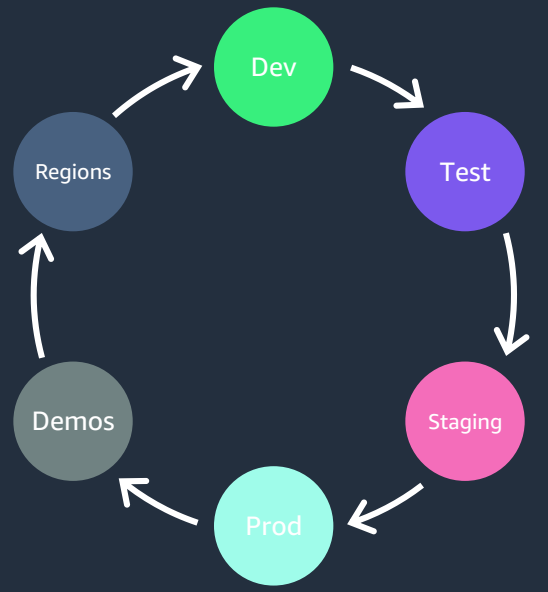
+

CI/CD
for automated build, test, and deploy

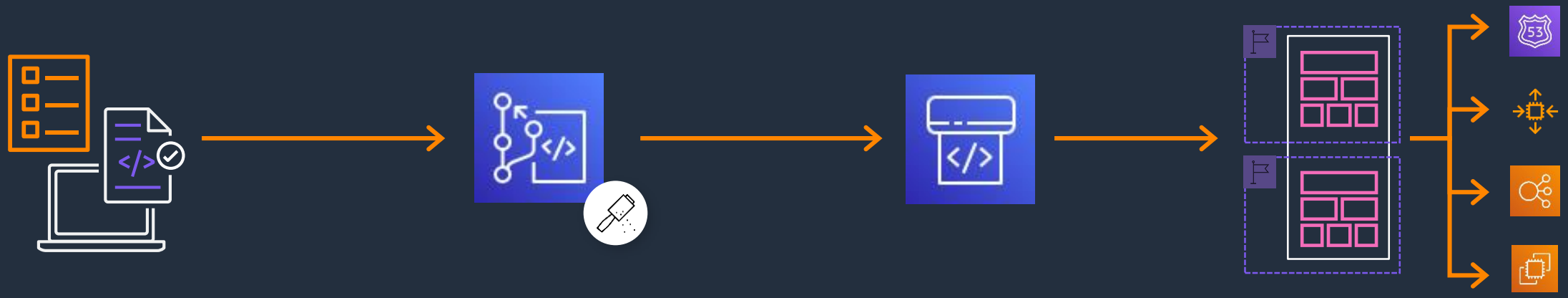
GitOps

- Single source of truth to deploy the whole stack
- A single workflow to control versioning on your infrastructure
- Service rolls back to the last good state on failures
- Infrastructure is defined, built, and run through CI/CD
- Drift detection





Provisioning and managing IaC



1. Code
2. Commit
   Git push/PR
3. Execute CI/CD
   test and
   deployment
4. Resources
   provisioned
   through
   CloudFormation

The AWS GitOps stack

Git solutions

- AWS CodeCommit

IaC solutions

- AWS CloudFormation
- AWS Cloud Development Kit (CDK)
- AWS Serverless Application Model (SAM)

CI/CD solutions

- AWS CodeBuild
- AWS CodePipeline
- AWS CodeDeploy

AWS Partner solutions

AWS CloudFormation powerful tool for IaC in the enterprise

AWS CloudFormation

Declarative JSON/YAML templates to describe infrastructure

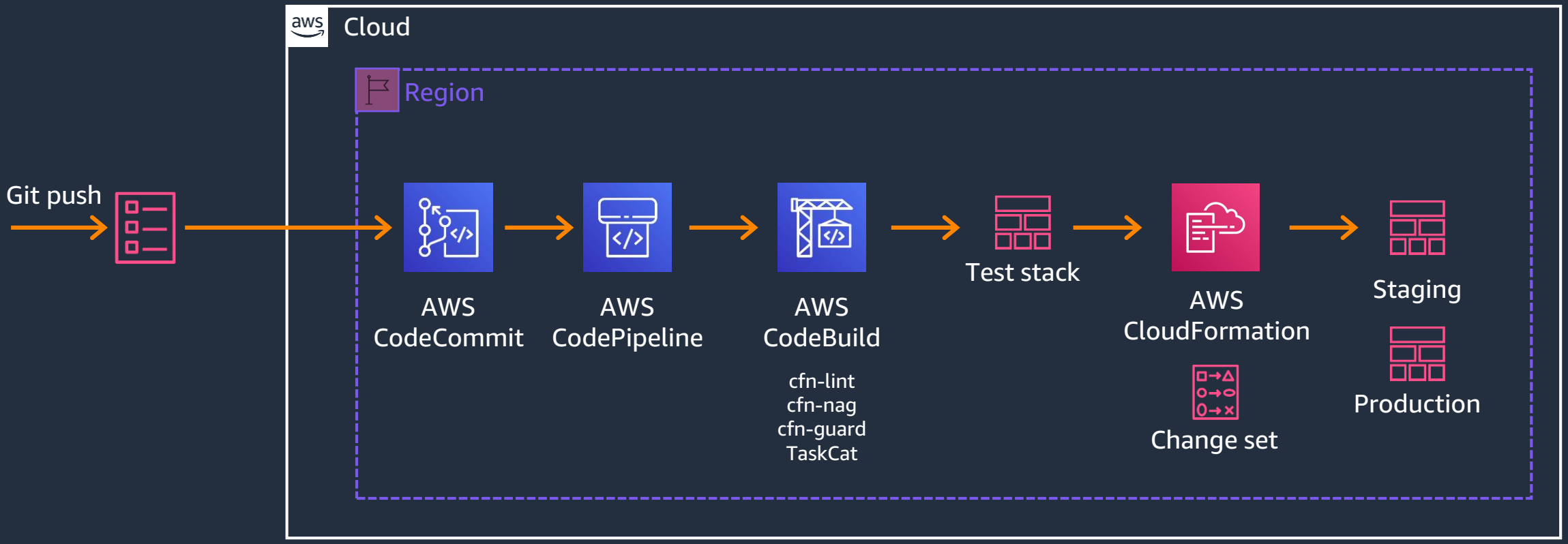
Automates provisions infrastructure as described in the template

Makes it easy to scale infrastructure

Tooling and integrations for enterprise requirements around compliance, governance, automation, and security

Manage AWS and third-party resources together

CI/CD on AWS



Thank you