Hawai‘i Cloud Innovation Summit 2023

Edge & Hybrid Computing
Enable AWS Cloud Computing Anywhere

Walt Whitecotton

AWS Worldwide Public Sector

Gov Cloud Business Development Mgr

TRUE TRANSFORM HAWAI'I GOVERNMENT DataHouse aws

Agenda

- Introduction
- AWS Edge - Hybrid portfolio Review
- Hybrid computing with AWS Outposts
- Edge computing with the AWS Snow Family
- AWS Modular Data Center
- Resiliency

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.

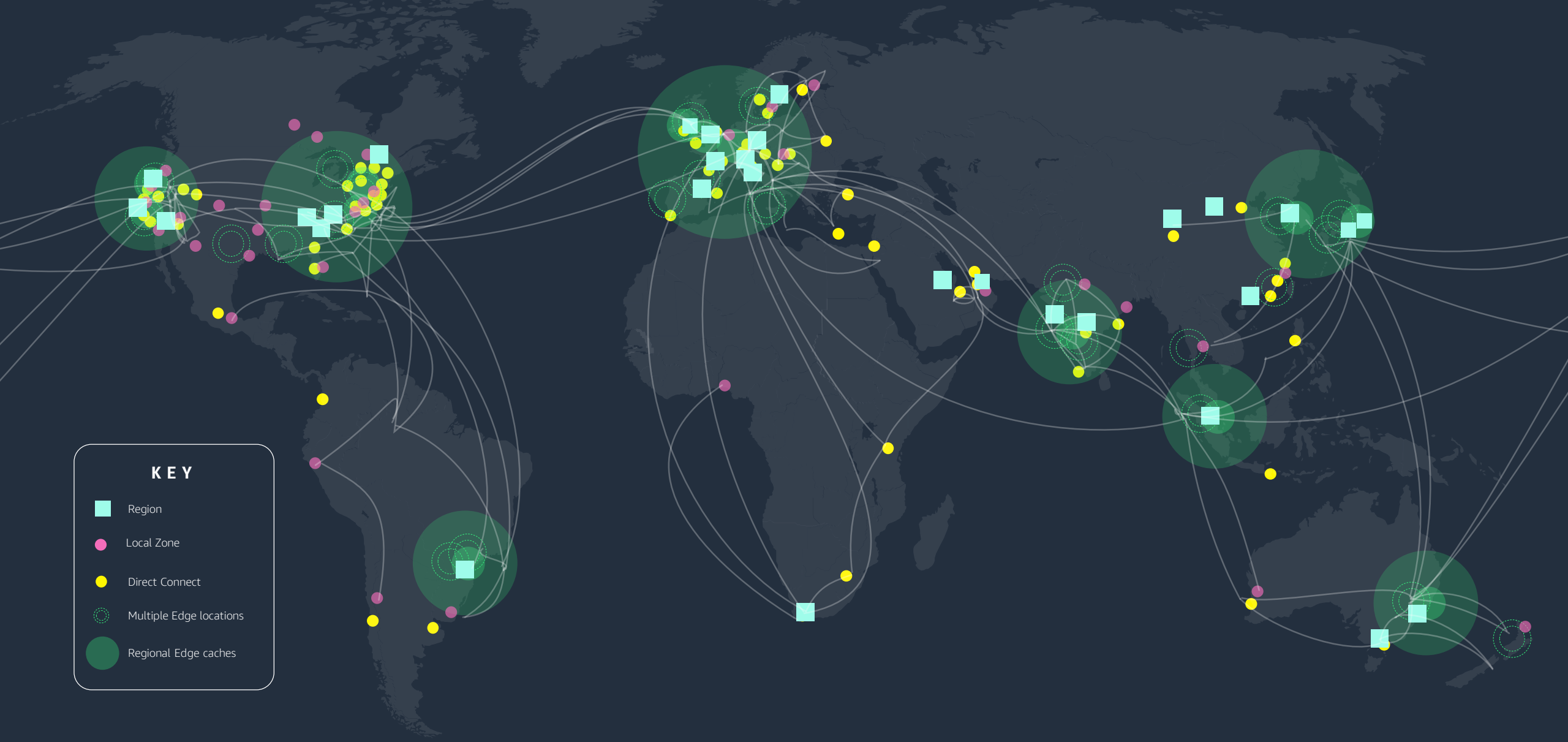
DIRECT CONNECT LOCATIONS

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

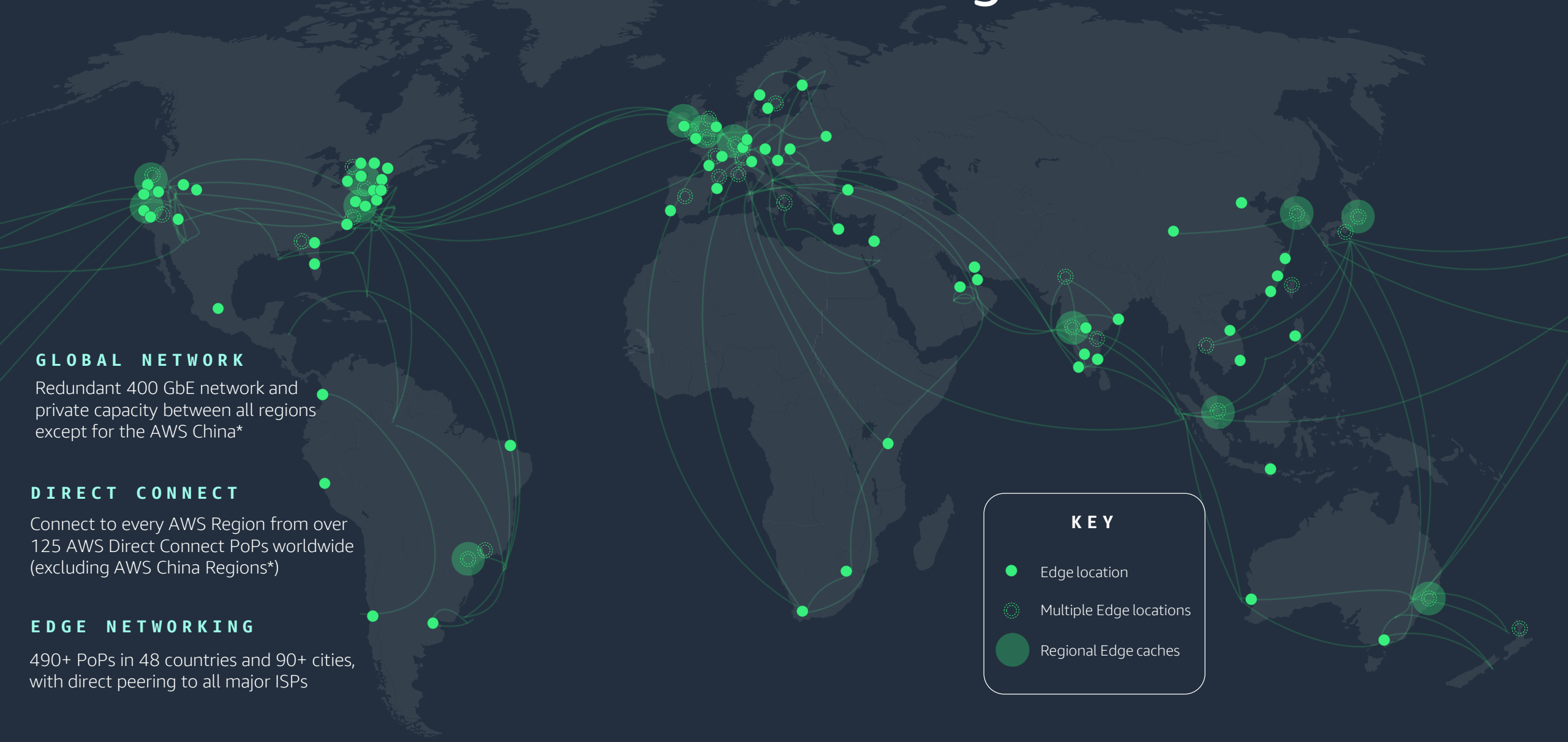
OUTPOSTS

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.

AWS Global Infrastructure



AWS CloudFront 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*)

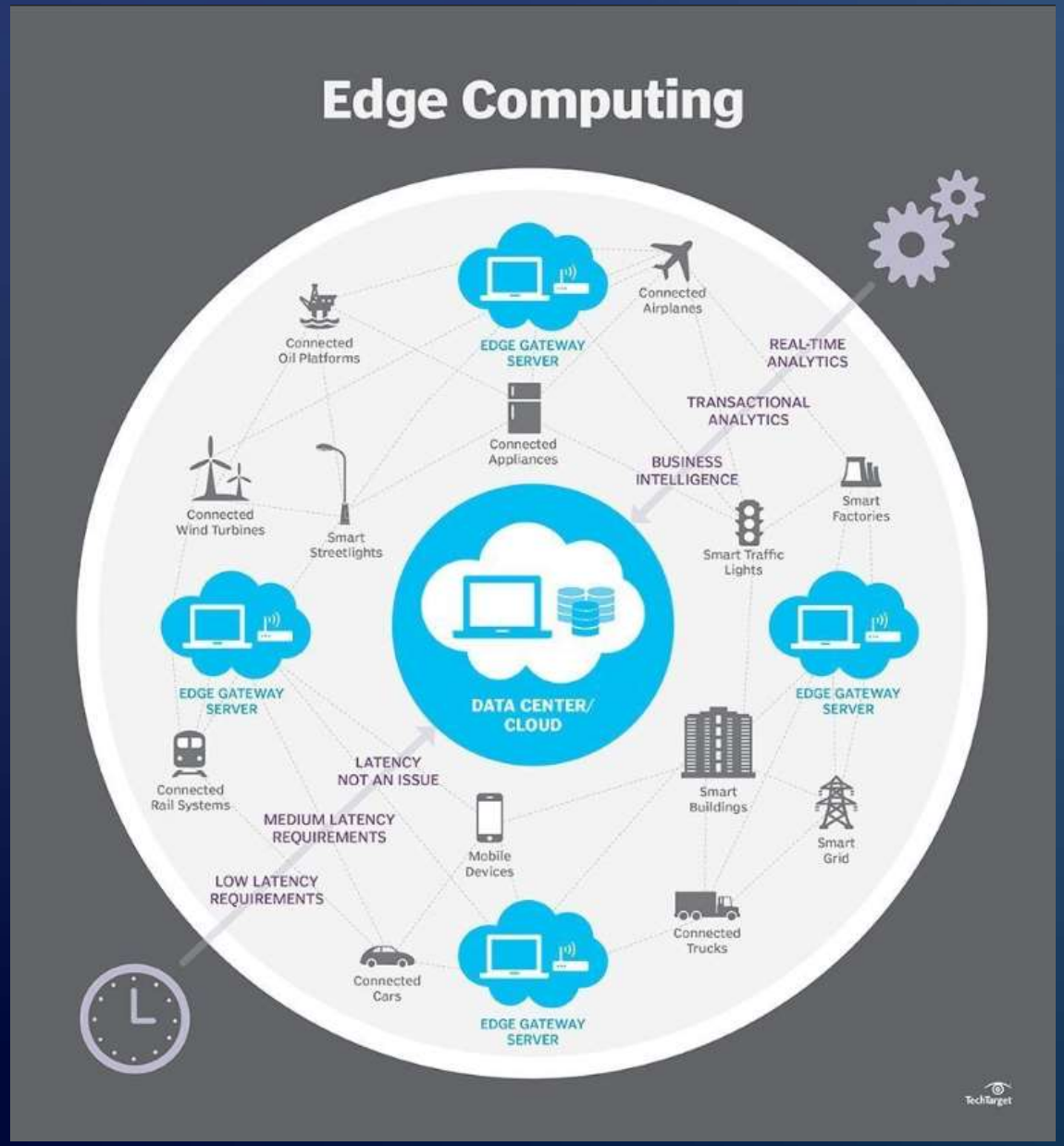
EDGE NETWORKING

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

KEY

- Edge location
- Multiple Edge locations
- Regional Edge caches

What is the edge?



Why is it important?

*"Historically around **10%** of enterprise-generated data is created and processed outside a traditional centralized data center or cloud."*

By 2025, Gartner predicts this figure will reach **75%**"

Gartner, Oct. 2018

Edge & Hybrid Cloud use cases

Various use cases across industries require cloud computing capabilities, but also need local processing for low latency & local storage to reduce data transfer cost. These use cases are often in environments where:

- data generation is decentralized
- data volumes are significant
- network connectivity is either consistent, intermittent or inaccessible
- application performance, suffer from impacts of latency
- data sovereignty requirements for certain types of information exist

  US Gov't FedRamp & EU GDPR, etc.

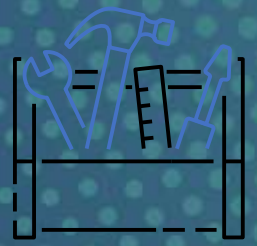



Customers want the same cloud experience at the edge

Same reliable, secure, and high-performance infrastructure

Same operational consistency

Same services and APIs



Same tools for automation, deployments, and security controls

Same pace of innovation as in the cloud

Customers' requirements for edge & hybrid applications

1/ Ability to securely connect and manage a broad range of device types and sizes at scale.

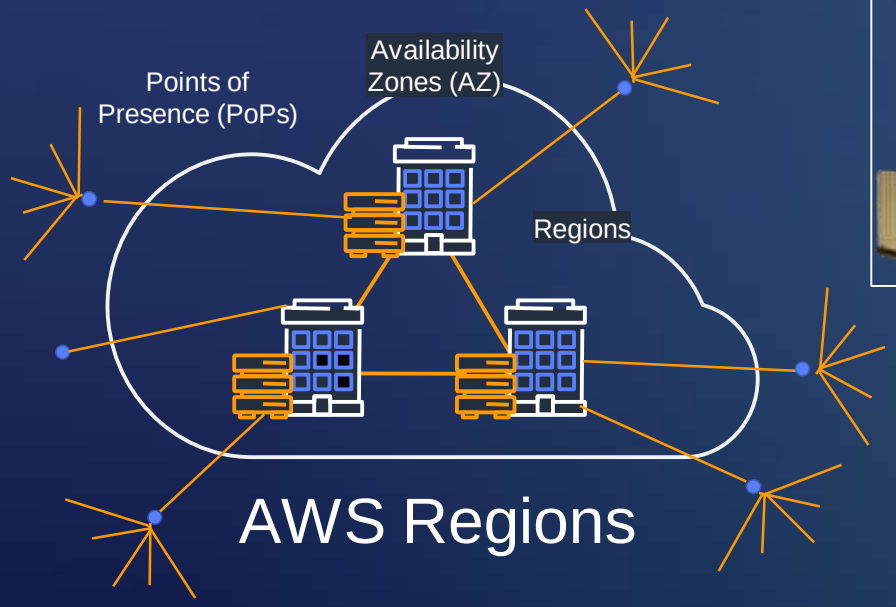
2/ Global cloud infrastructure that provides data processing and analysis capability as close to the end user or device as necessary.

3/ A broad range of integrated cloud and device services

4/ A single programming model for the cloud and local devices

AWS Edge/Hybrid Portfolio Overview

Edge to cloud continuum



AWS Regions

On premises

AWS Outposts

AWS Snowblade

AWS Modular Data Center

Metro centers

AWS Local Zones

5G networks

AWS Wavelength

IoT / robotics

AWS IoT Greengrass

AWS RoboMaker

Rugged, Mobile edge

AWS Snowball Edge

AWS Snowcone

AWS Snowblade

Cloud continuum

For most use cases

For low-latency, local data processing and data residency

…aligning to key IT initiatives

Migration and modernization

Migration of VMware-based apps that require low-latency access to on-premises compute and storage systems

Data center consolidation

Hardware refresh

VMware vSphere upgrades

Disaster recovery

Failover to fully managed VMware-based DR sites to meet low-latency or data residency requirements

Flexibility to failover to AWS Regions for global scalability and cost efficiency using same tools and processes

Test and development

Training and development environments for modern applications

Short-term projects

On-premises infrastructure that is ready to scale to AWS Regions

More information on VMware Cloud on AWS Outposts is available in the VMware Cloud on AWS Outposts (https://bit.ly/3Thk0y7)

Hybrid Computing with AWS Outposts

AWS Outposts Family

Outposts rack

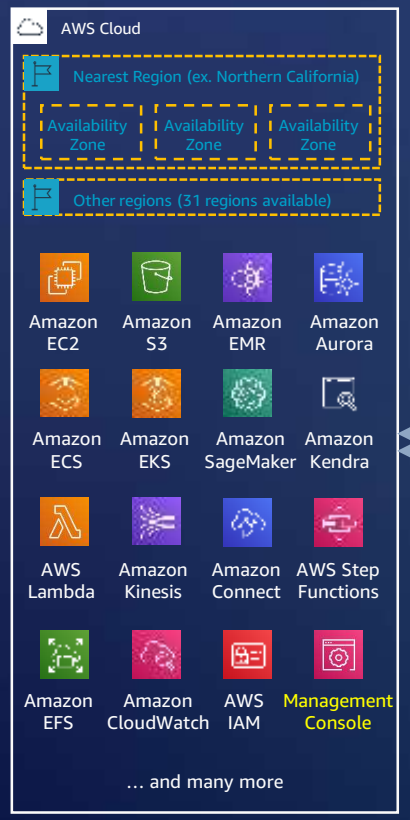
Bring the same AWS APIs, services, and features to virtually any datacenter or co-location space

Outposts servers

Run Outposts in locations with limited space or smaller capacity requirements

Deploy Applications Built in Region, Locally

AWS REGION



Leverage AWS commercial backbone or satellite failover

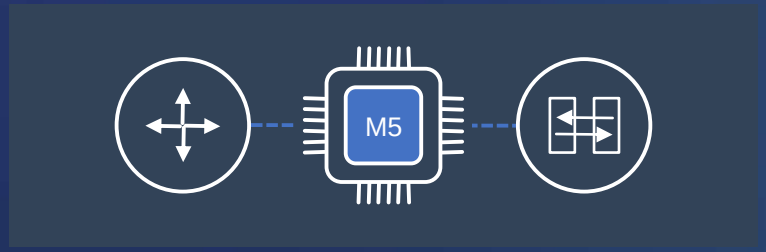
Direct Connect recommended, VPN available, min 1 Gbps
(backup connectivity also recommended)

OUTPOSTS

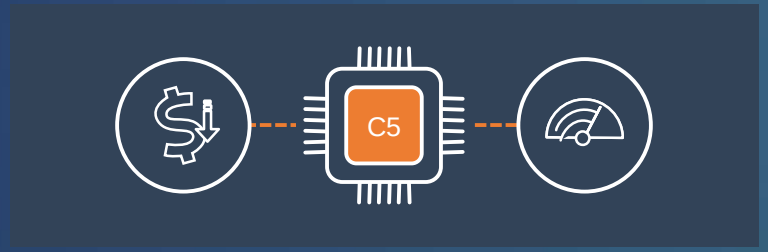
Latency sensitive, local pre-processing, data residency requirements
Ex: Data center, Embassy, major military installations

- Lowest latency for on premises users!
- Fully managed service, not just hardware
- Not intended for disconnected environments but will keep running in steady state
- Run managed services locally and easily access in-region services for extended workflows

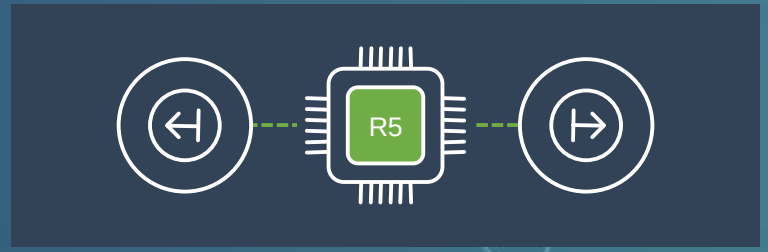
Build on the same EC2 Instances & EBS Volumes



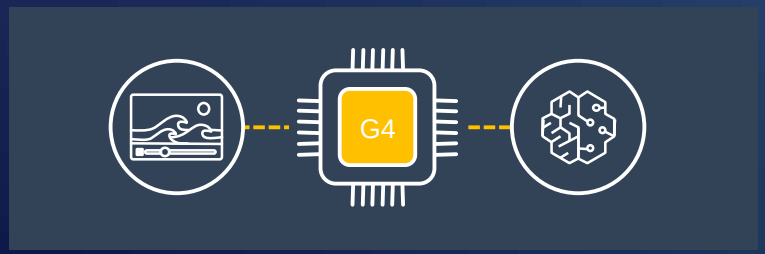
For general purpose applications



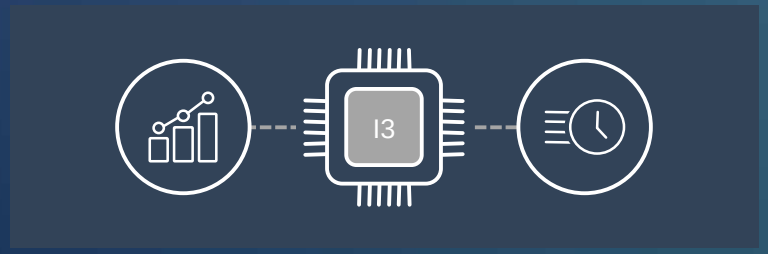
For compute intensive applications (media transcoding, gaming servers, machine learning inference)



For memory intensive applications (databases, in-memory caches, real time data analytics)



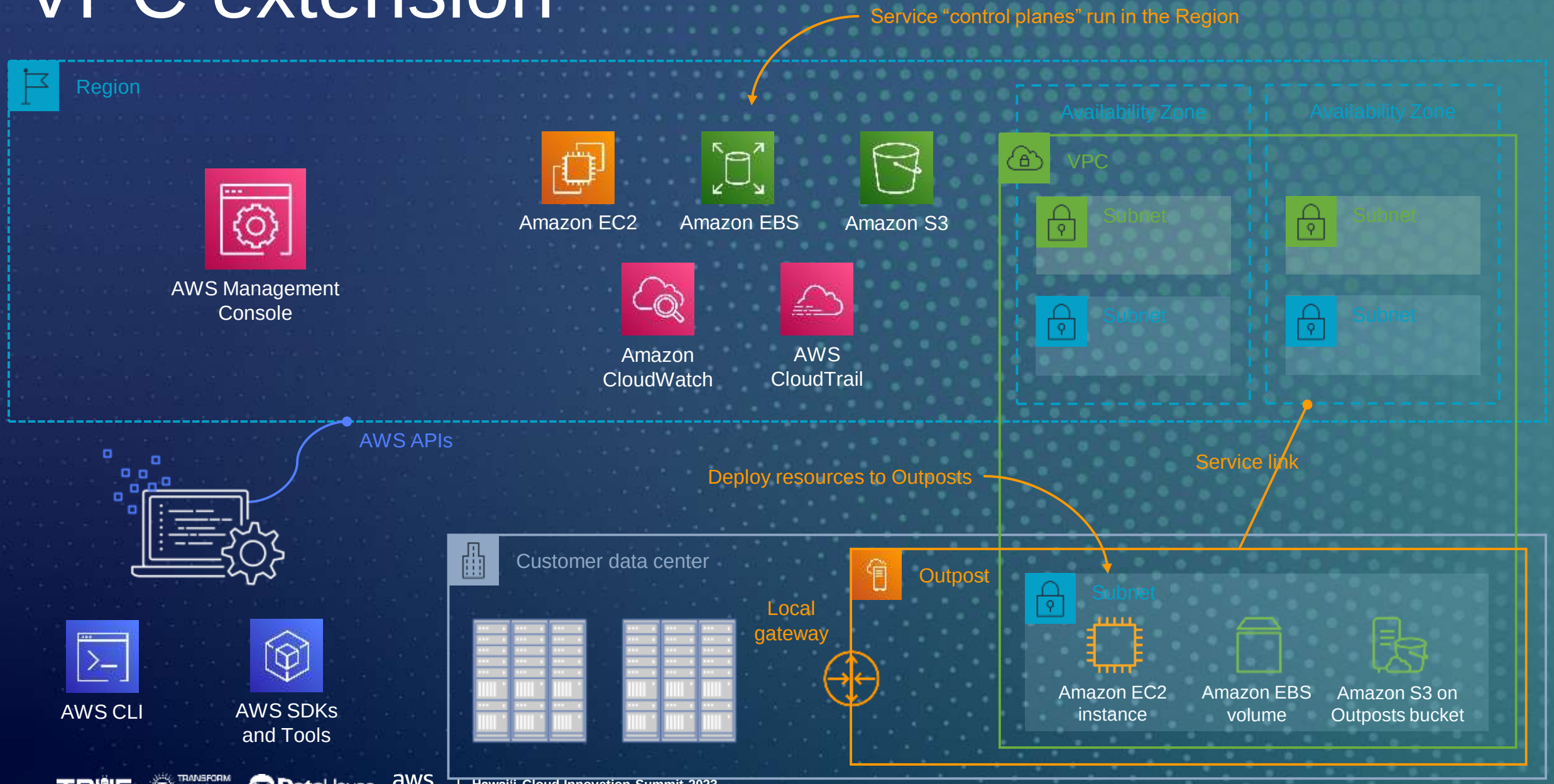
For machine learning inference and graphics workstations



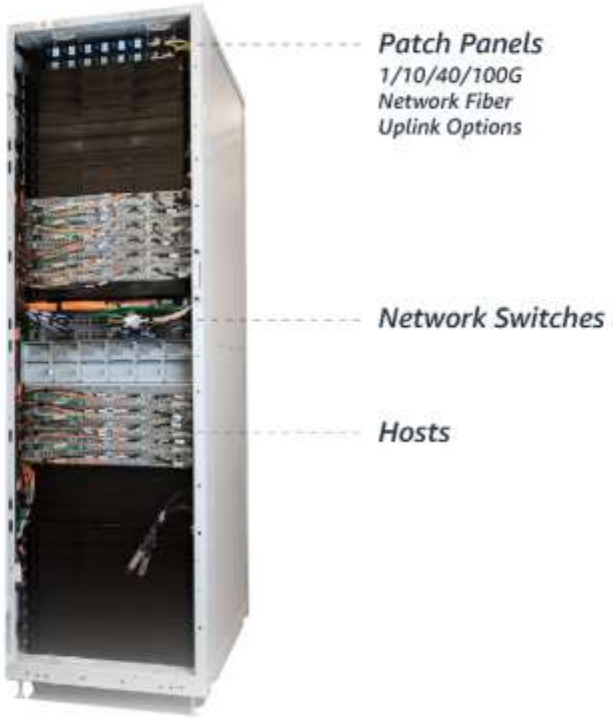
For I/O intensive applications (NoSQL databases, in-memory or transactional databases, distributed file systems)

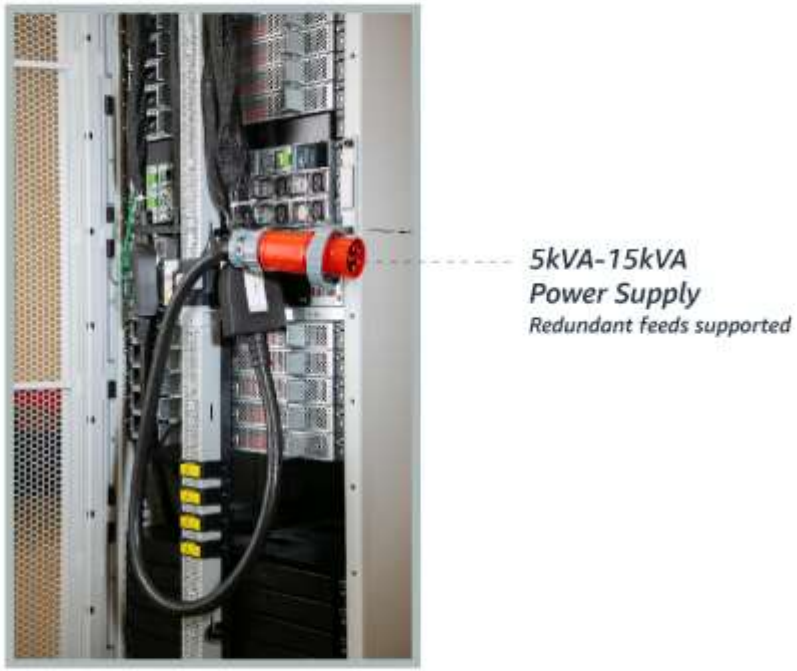
Local Instance Storage and EBS gp2 volumes for temporary and persistent storage

VPC extension



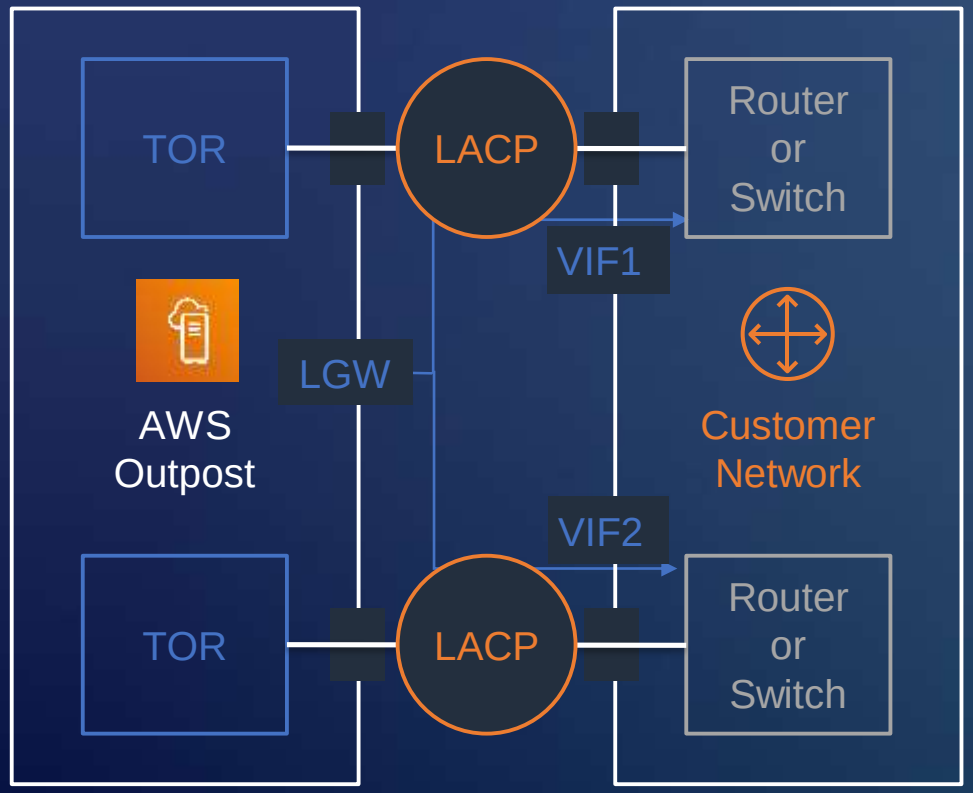
AWS Outposts Rack





AWS Outposts network connectivity

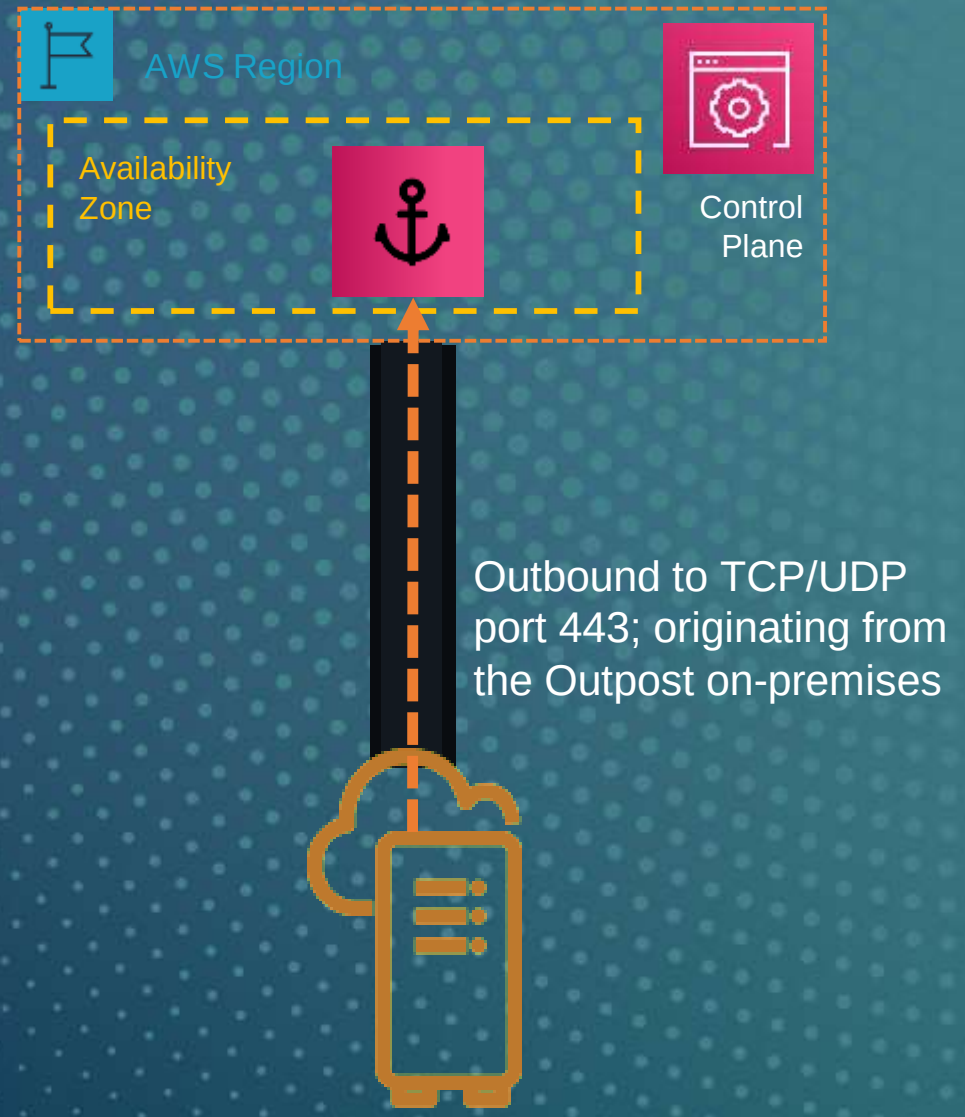
Connect to your local network



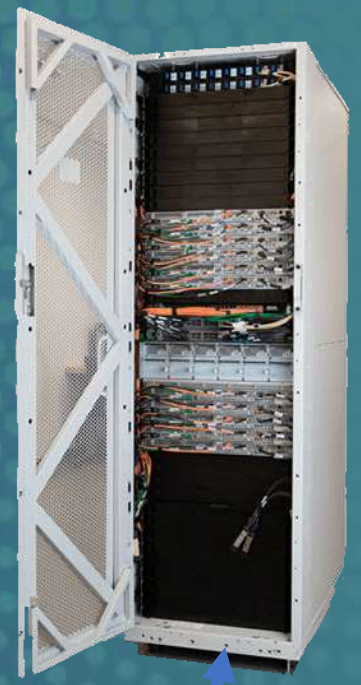
- Connect to local network equipment via ports provided in the Outpost's top of rack (TOR) switches
- Configure Virtual Interfaces (VIFs) mapping to your VLANs using Link Aggregation Control Protocol (LACP)
- Configure the new local gateway (LGW) on the Outpost to route traffic to and from your local network using these VIFs

AWS Outposts logical security - Service Link specifics

- Control plane stays within the region
- VPN tunnels to anchor points within a single Availability Zone
- TCP/UDP port 443 required
- Each Outposts server makes a service link connection to the region
- Service link is established outbound from the Outpost
- Consists of Data plane and Management Plane



Management and control plane in the AWS Region



Telemetry including health, state, hardware metrics

RDS Snapshots and Backups

Cloudwatch

EBS snapshots
(when the OutpostArn is not specified)

APIs calls

Systems Management

DNS

Amazon Machine Images

EKS, ECS, EMR Control Plane

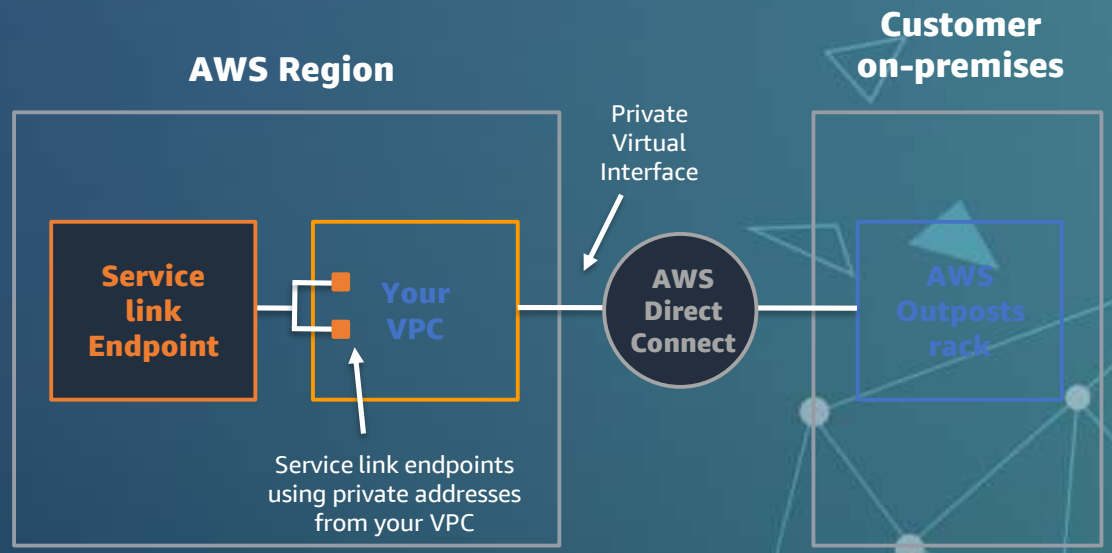
Region

Identity and Access Management

Connect to your AWS Region

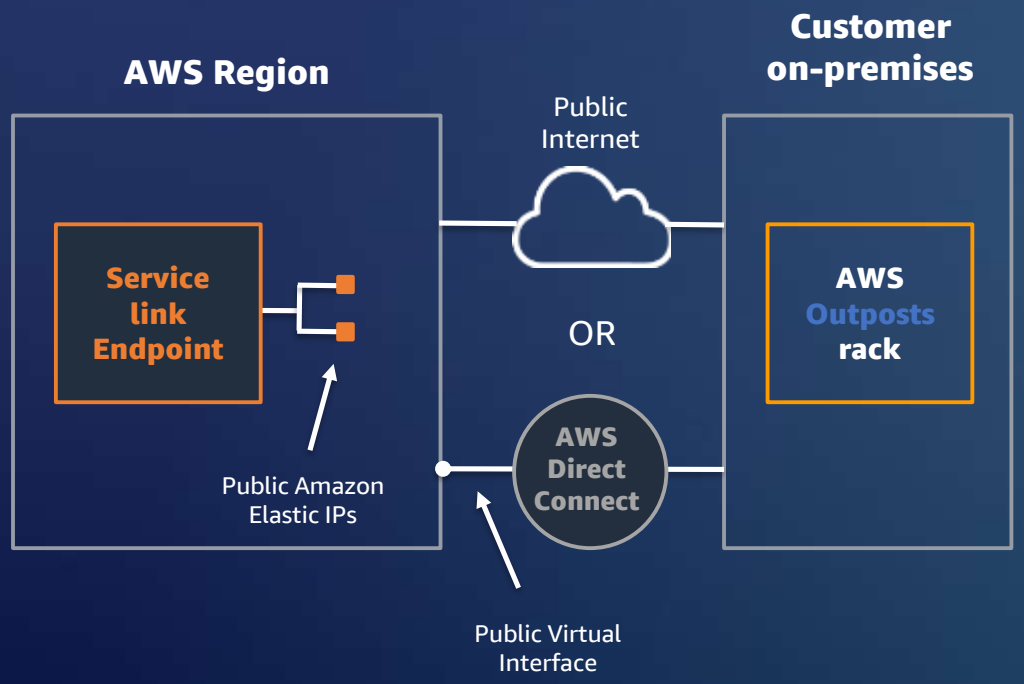
Private WAN access

- **AWS Service link access:** Connects to a VPC that you own, in your AWS Outposts account
- **Access from your VPC to your on-premises:** Uses an AWS Direct Connect private virtual interface, or other private means such as VPN
- **Service link Endpoints use private addressing:** Using private addresses in your VPC range, service link endpoints are reachable via VPC routing, no public IP's required
- No public IPs required: Through your AWS Direct Connect, all IP addressing can be private



Connect to your AWS Region

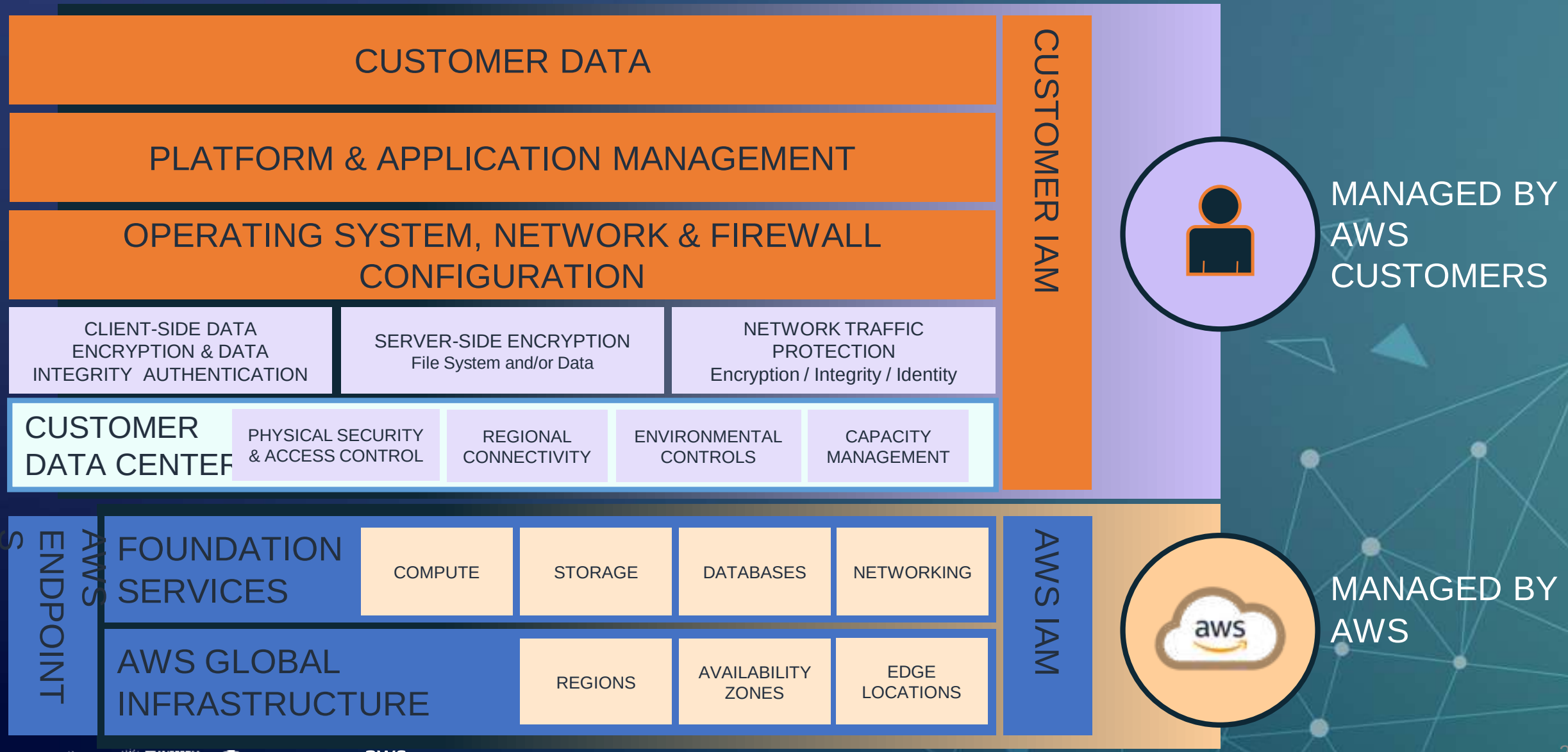
Public WAN access



- **AWS Service link access:** Uses public Amazon IP's in the public AWS realm, of the home region for your AWS Outposts rack
- **Reaching your service link from your on-premises:** You can either use the public internet or an AWS Direct Connect public virtual interface
- **AWS Direct Connect is not required:** With the public WAN model of access for your Outposts and service link anchors, public internet can be used without Direct Connect (DX), as DX is not a specific requirement for reaching the anchor endpoints

Security, compliance and responsibility

AWS Outposts shared responsibility model



AWS Outposts Rack security

- Built-in **tamper detection**
- Enclosed rack with a lockable door
- Data on Outposts rack is **encrypted**
- Removable and destroyable **hardware security key** on each server
- Encrypted network connection to the AWS Region
- **Physical security** of the Outposts rack location is the **customers** responsibility

Edge Computing with the AWS Snow Family

AWS Snow Family overview

AWS Snowcone

Compact and portable device, purpose-built for use outside of a traditional data center

AWS Snowball

Compute-optimized or storage-optimized device for extreme conditions

AWS Snowblade

Rack form-factor compute, storage, networking. MIL-STD-810H enclosure

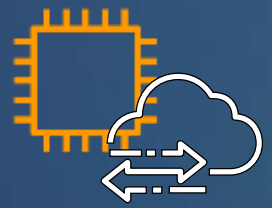
-JWCC Customers Only-

Available for clustering, scaling, and connectivity

AWS Snow Family key value propositions

AWS services and capabilities at the edge

Snow brings the core AWS services like EC2, EKS, EBS, S3, and IAM and AWS AI/ML capabilities like Greengrass, Lambda and more to the edge



Edge computing and data movement

Snow devices provide both edge computing and data movement capabilities

Cloud compatibility

Use the same AWS service APIs that you use in the cloud

Scalable

Add or remove Snow devices based on your needs

Secure

Snowball Edge devices use tamper-evident enclosures, 256-bit encryption, and industry-standard Trusted Platform Modules (TPM)

Run partially or fully disconnected

Snow devices and capabilities are fully operable without network connection

AWS services supported on Snow Family

Amazon SageMaker Neo

AWS IAM Compatible API

AWS DataSync

AWS IoT SiteWise Edge

AWS IoT GreenGrass

Run AWS Services on Snow at the rugged, mobile Edge

Amazon EKS-D

AWS Lambda

Tape Gateway on Snow

Amazon S3 Compatible API

Amazon EBS Compatible API

Amazon EC2 Compatible API

AWS Snowcone for mobile workl[illegible]s

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

Tamper-resistant
Tamper-evident

Handles vibration u
to 50G RMS

Weighs 4 pounds and
can be carried in a backpack

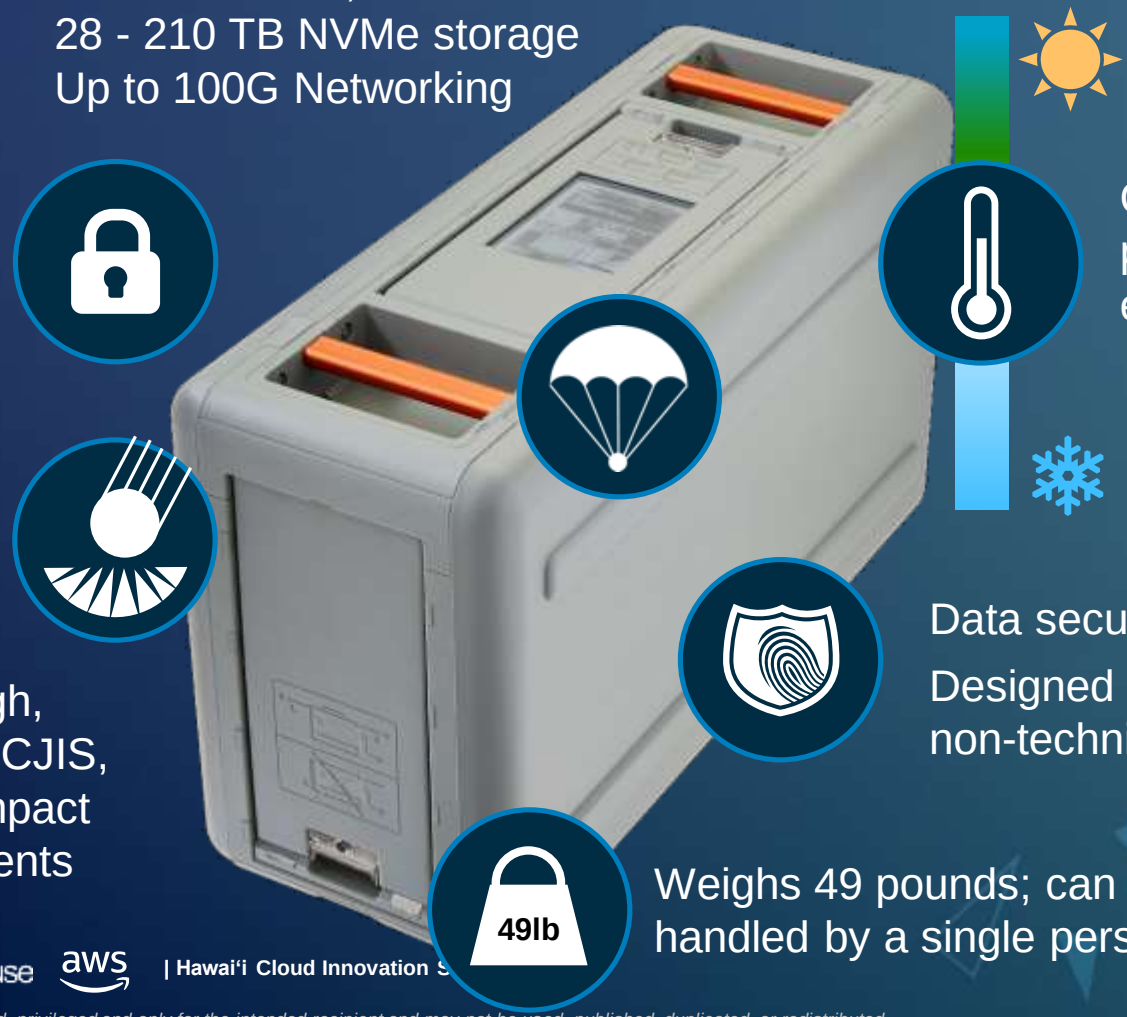
Operates on unconditioned power and in non-datacenter environments (0° to 45°C)

Data secured with 256-bit encryption

Designed to be operated by non-technical personnel

AWS Snowball Edge for rugged and tactical edge

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

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

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

Data secured with 256-bit encryption

Designed to be operated by non-technical personnel

Weighs 49 pounds; can be handled by a single person

TRUE TRANSFORM HAWAII GOVERNMENT DataHouse aws | Hawai'i Cloud Innovation S

AWS Snowblade ruggedized

Available for Joint Warfighting Cloud Capability (JWCC) customers

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

50lbs

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

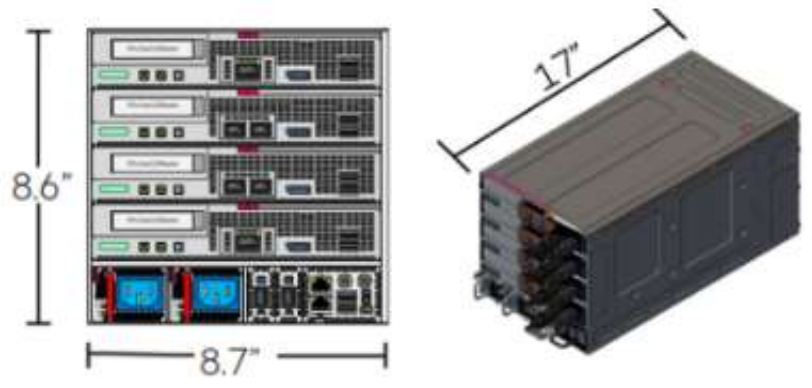
208 vCPU, 832 GB RAM, 64 TB NVMe
4 server blades for redundant workloads
2 power-supplies
4 dual-port network switches
16TB/blade
Redundant interfaces (2 USB, 2 Ethernet)

DoD Tactical Edge: Snowblade in optional TEMPEST enclosure

DoD Tactical Edge: Snowblade in optional MIL-STD-810H enclosure

AWS Snowblade rack form-factor

Available for Joint Warfighting Cloud Capability (JWCC) customers



- Rack-mounted, NEBS-3 compliant, size, weight, power (SWaP) optimized form-factor with dense storage, compute, integrated 100G NIC
- Each device has 4 x 1U compute blades and 4 x 100G NIC
- Usable capacity per device: 208 vCPU, 832GB RAM, 4 TB NVMe storage, 4x 100G NIC
- Redundant power supplies and interfaces
- Dimensions: 5U high, 8.7” wide (22 cm), 16.9” deep (43 cm)
- Up to 10 Snowblade devices per server rack (with ToR switch)

AWS Modular Data Center

AWS Modular Data Center

Available for Joint Warfighting Cloud Capability (JWCC) customers

- Supports up to six racks – any mix of AWS Outposts or a purpose-built rack solution for the AWS Snow Family
- Supports dual power feeds and includes redundant HVAC
- You provide power and networking
- Use across Unclassified, secret, and top secret workloads

37

AWS Modular Data Center benefits

Run applications wherever you need them

AWS Modular Data Center lets you bring compute and storage to virtually any location where data is generated

Deploy edge infrastructure rapidly

With AWS Modular Data Center, you have a fully built, self-contained data center ready to ship in weeks

Support your operations with trusted AWS data center expertise

AWS helps you confidently manage and maintain your data center by providing access to 24/7 support, a global maintenance team, and a secure monitoring system

Create a consistent experience across your environments

AWS Modular Data Center is designed to hold racks of AWS Outposts and AWS Snow Family devices, allowing you to build, deploy, and manage your applications on a common platform

Using AWS Snow Family devices in AWS MDC

Purpose-built racks for AWS Snow Family

42U Rack (24in width, 48in depth)

Top-of-rack switch

AWS Snowball Edge x6

Power shelf with uninterruptible power supply

AWS MDC

Deployment testing:

39

AWS Snowblade equipped MDC

Available for Joint Warfighting Cloud Capability (JWCC) customers

Self-contained MDC: using AWS Snowblade

- Up to 50 Snowblade devices in a AWS Modular Data Center
  - 10400 vCPU
  - 12.8 TB RAM
  - 3.2 PB storage
  - 20 Tbps Networking
- AWS Modular Data Center provides you with a self-contained data center, built into two 20’ shipping containers
- AWS MDC comes with a management system to proactively monitor and operate the MDC’s subsystems

40

Resiliency

Resiliency

Can mean different things to different stakeholders, what does it mean to you?

Resiliency is the ability of a system to respond, adapt, and eventually recover from unexpected conditions

Three foundational resiliency pillars: infrastructure, architecture, practices/tools

AWS solutions are designed to enable compliance with flexibility to balance resiliency and cost for each workload

Why is it important?

Embracing Failure as a Natural Occurrence

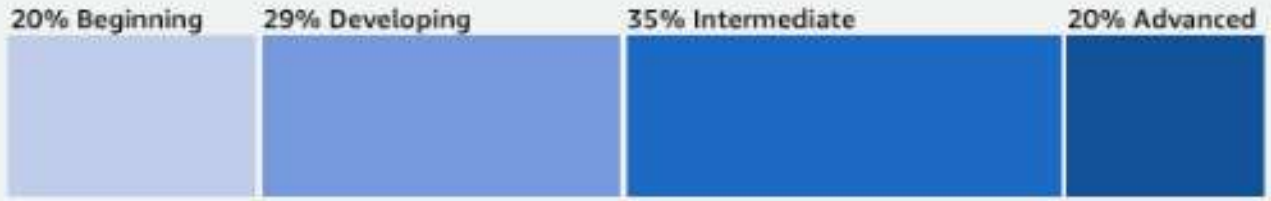
The cost and frequency of outages have soared in 2022:

- >60% of outages cost more than $100k, up from 39% in 2019
- 15% of outages cost more than $1 million, up from 11% in 2019

Successful organizations consider failure a natural part of the business

Not all organizations are equally resilient:

Survey across 2,100 industry leaders across 7 Key Industries shows that only 20% of these organizations have an advanced resiliency posture.



> **"We needed to build systems that embrace failure as a natural occurrence."**
>
> - Werner Vogels, VP and CTO, AWS

Major service provider outages… … and inevitably more to come

Google G-Suite	Texas Power Grid	Fastly	Azure Global	Facebook	AWS US-EAST-1	Spotify	Cloudflare
August 2020	February 2021	June 2021	October 2021	October 2021	December 2021	March 2022	June 2022

Plan for failure…

Resiliency is Critical, Complex, and is a key Cost driver (3C's)

Resiliency in the cloud need not be the same as traditional IT

*Greater use of a trio of cloud services: autoscaling, cloud watch, and load balancing provide significant automation & ease of use

Cloud to Edge, similar historical concepts & patterns tied to defining RTO & RPOs

Design w/ no single points of failure (balance cost vs risk)

Customers often ask for best practices and advice

*Caveat: The following Outpost based review of resiliency – patterns & anti-patterns are generalized. Ea application requires specific plng & review

Resiliency – Outpost Network review

Outpost deployment depends on a resilient connection to its anchor Availability Zone (AZs in Region) for management, monitoring, and service operations to function properly.

Provision your on-premises network to provide redundant network connections for each Outpost rack and reliable connectivity back to the anchor points in the AWS cloud.

Also, consider network paths between the application workloads running on the Outpost and the other on-premises and cloud systems they communicate with – how will you route this traffic in your network

Resiliency – Outpost Compute review

EC2 capacity on Outposts is **finite**. You are responsible for planning & managing the compute capacity of your Outposts deployments

Order sufficient compute capacity to support an N+M availability model, where N is the required number of servers & M is the number of spare servers provisioned to accommodate server failures. N+1 and N+2 are the most common availability levels.

Monitor and manage your Outpost capacity to ensure sufficient spare capacity is always available to accommodate server failures. use the same instance auto recovery & EC2 Auto Scaling mechanisms to recover or replace instances impacted by server failures and maintenance events.

Use Amazon EC2 placement groups on Outposts to control placement of instances across racks within a single Outpost.

Resiliency – Outpost Storage review

Use EBS snapshots to create point-in-time backups of block storage volumes to Amazon S3 in the Region or S3 on Outposts.

Use S3 on Outposts object versioning to maintain multiple versions and history of your objects.

Use S3 Replication on Outposts to automatically replicate your object data to another Outpost/region.

For non-data residency use cases, use AWS DataSync to back up objects stored in S3 on Outpost to Amazon S3 in the Region.

Use Elastic Disaster Recovery (legacy CEDR) to replicate instances between on-premises systems, logical Outposts, and the Region.

Hawaii Based Organization

On Premises Location

Edge Router

Edge Router

Direct Connect Location 1

Customer/ISP Router

AWS Router

Direct Connect Gateway

Direct Connect Location 2

Customer/ISP Router

AWS Router

Direct Connect Gateway

AWS Region

Transit Gateway

Internet Gateway

Internet Gateway

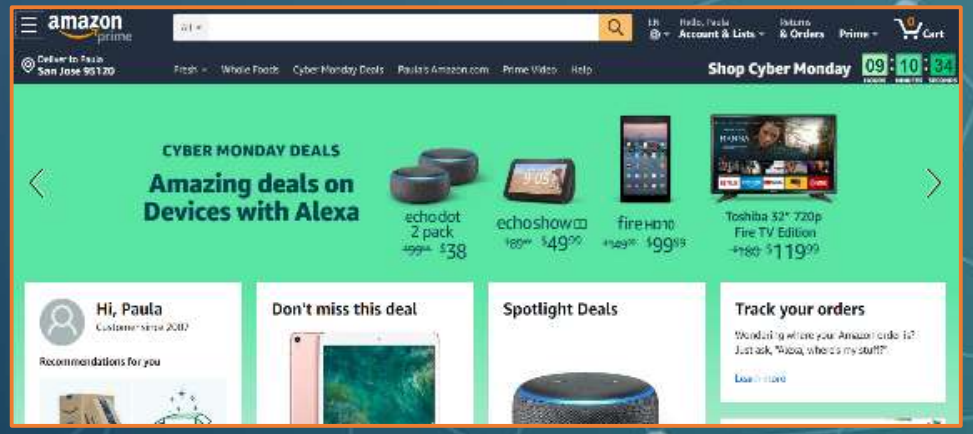
Internet Gateway

See link - Bldg Resilient & Hi Performing Cloud based Apps in HI diagram

Summary

Vast experience with edge computing for decades

- Globally, we have more than **175 fulfillment centers** and more than **40 sort centers**
- We have opened more than **50 robotic fulfillment** centers around the world
- Amazon currently uses the help of more than **200,000 robotic drive units** around the world
- Amazon has over **20,000 trailers** on the road
- Customers have purchased more than **one hundred million Alexa**-enabled devices



Please submit your **feedback**. Thank you!

https://www.pulse.aws/survey/DH53FSCQ

Feel free to reach out:
whitecot@amazon.com

Thank you!

Back-Up E-H Cloud Information

Edge/Hybrid Cloud links

Building a Hybrid Cloud with AWS (amazon.com)

Hybrid Cloud Use Cases | Amazon Web Services

Navigating common use cases spanning AWS GovCloud (US) and standard AWS | AWS Public Sector Blog (amazon.com)

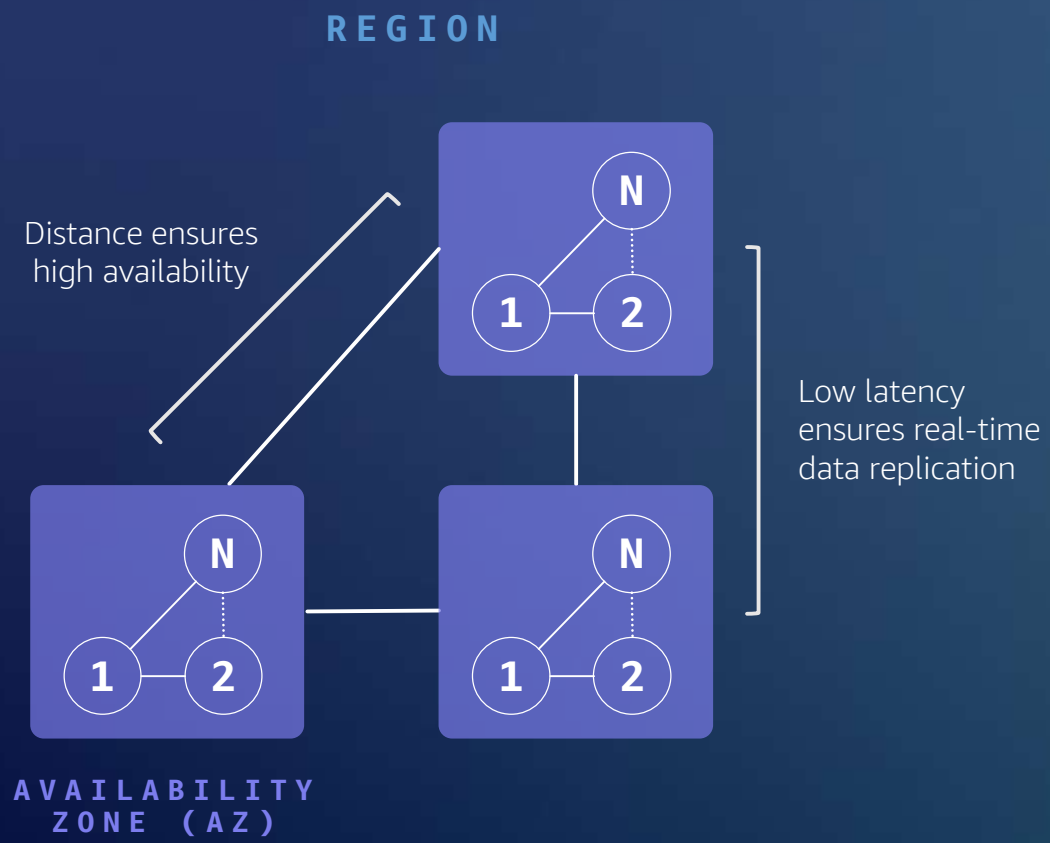
Resiliency Links

Building Resilient and High Performing Cloud-based Applications in Hawaii | AWS Architecture Blog (amazon.com)

AWS Outposts High Availability Design and Architecture Considerations - AWS Whitepaper (amazon.com)

Understand resiliency patterns and trade-offs to architect efficiently in the cloud | AWS Architecture Blog (amazon.com)

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 Outposts Customer Experience