

AWS Global Infrastructure

Cheat sheets, Practice Exams and Flash cards www.exampro.co/clf-c01

What is the AWS Global Infrastructure?

The AWS Global Infrastructure is **globally distributed hardware and datacenters** that **are physically networked together** to act as one large resource for the end customer.

The AWS Global Infrastructure is made up of the following resources:

- **25** Launched Regions
- **81** Availability Zones
- **108** Direct Connection Locations
- **275+** Points of Presence
- **11** Local Zone
- **17** Wavelength Zones



AWS has **millions** of active customers and **tens of thousands** of partners globally

Global Infrastructure - Regions

Cheat sheets, Practice Exams and Flash cards  www.exampro.co/clf-c01

Regions are **geographically distinct locations** consisting of one or more Availability Zones.

Every region is **physically isolated** from and independent of every other region in terms of **location, power, water supply**

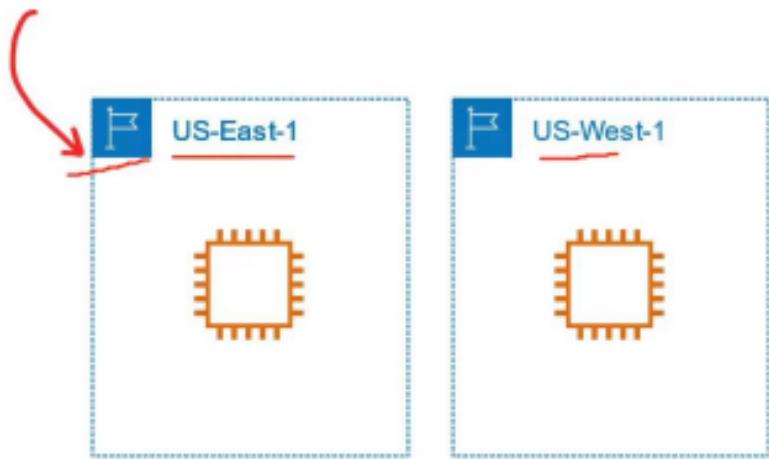
- Launch Regions
- Coming Soon



Global Infrastructure - Regions

Cheat sheets, Practice Exams and Flash cards  www.exampro.co/clf-c01

This is what a **region** will look like
represented in an architectural diagram.



Global Infrastructure - Regions

Cheat sheets, Practice Exams and Flash cards www.exampro.co/clf-c01

Each region generally has three Availability Zones

- Some new users are limited to two eg. US-West

New services almost always become available first in **US-EAST**

Not all AWS Services are available in all regions

All your billing information appears in **US-EAST-1** (North Virginia)

The cost of AWS services vary per region



When you choose a region there are four factors you need to consider:

1. What Regulatory Compliance does this region meet?
2. What is the cost of AWS services in this region?
3. What AWS services are available in this region?
4. What is the distance or latency to my end-users?

Global Infrastructure – Regional vs Global Services

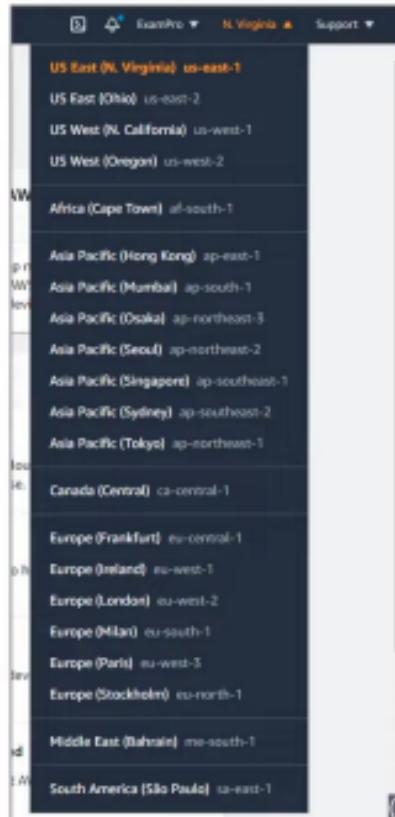
Cheat sheets, Practice Exams and Flash cards www.exampro.co/clf-c01

Regional Services

AWS **scopes** their AWS Management Console on a selected Region.

This will determine where an AWS service will be launched and what will be seen within an AWS Service's console.

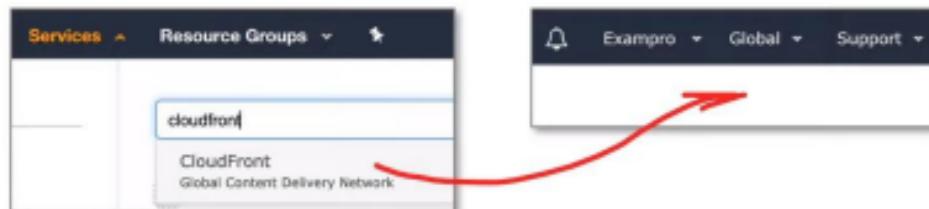
You generally don't explicitly set the Region for a service at the time of creation.



Global Services

Some AWS Services operate across multiple regions and the region will be fixed to "Global"

E.g. Amazon S3, CloudFront, Route53, IAM



For these global services at the time of creation:

- There is no concept of region. eg. IAM User
- A single region must be explicitly chosen eg. S3 Bucket
- A group of regions are chosen eg. CloudFront Distribution

Global Infrastructure – Availability Zones

Cheat sheets, Practice Exams and Flash cards  www.exampro.co/clf-c01

An **Availability Zone** (AZ) is physical location made up of one or more datacenter.

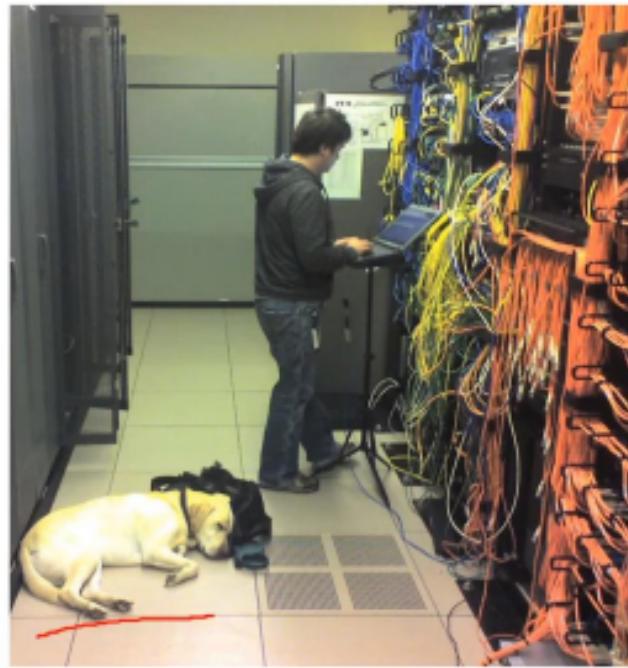
A datacenter is a secured building that contains hundreds of thousands of computers.

A region will ***generally contain 3 Availability Zones**

Datacenters within a region will be isolate from each other (different buildings). But they will be close enough to provide low-latency (< 10ms).

Its common practice to run workloads in at least 3 AZs to ensure services remain available in case one or two datacenters fail. (High Availability)

AZs are represented by a Region Code, followed by a letter identifier eg. **us-east-1a**



Global Infrastructure – Availability Zones

Cheat sheets, Practice Exams and Flash cards www.exampro.co/clf-c01

A Subnet is associated with an Availability Zone.

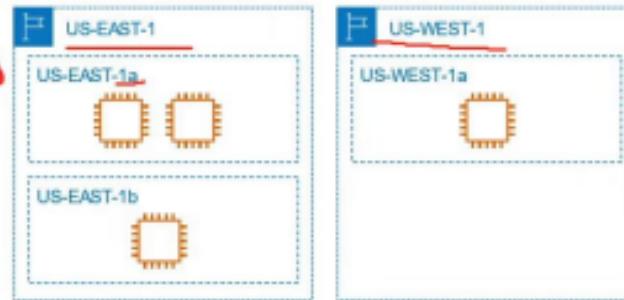
You never choose the AZ when launching resources.

You choose the Subnet which is associated to the AZ.



Subnet	ⓘ	✓ No preference (default subnet in any Availability Zone) subnet-d9de91f7 Default in us-east-1c subnet-d0c28f8c Default in us-east-1a subnet-349fdf53 Default in us-east-1b subnet-a8c2f8a7 Default in us-east-1f subnet-b9db4c87 Default in us-east-1e subnet-13869659 Default in us-east-1d
Public IP	ⓘ	
Ent group	ⓘ	
ervation	ⓘ	

Example of an architectural diagram, representing two AZs, the Subnets associated with those AZs, and EC2 instances (Virtual Machines) launched in those subnets



The US-EAST-1 region has 6 AZs
(the most Availability Zones of any region)

Global Infrastructure – Availability Zones

Cheat sheets, Practice Exams and Flash cards www.exampro.co/clf-c01

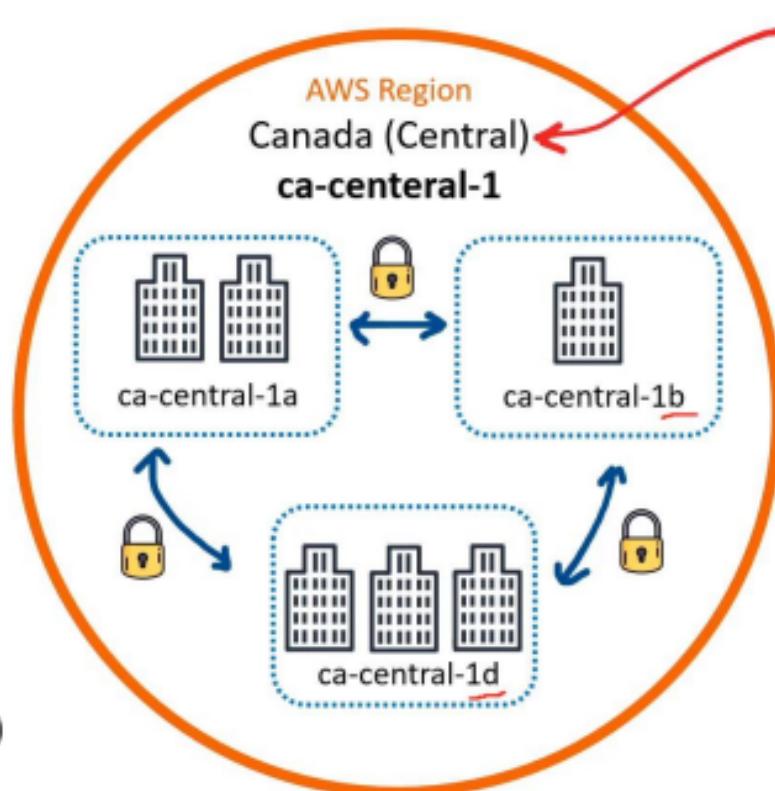
A region has multiple Availability Zones

An Availability Zone is made up of **one or more** datacenters

All AZs in an AWS Region are interconnected with high-bandwidth, low-latency networking, over fully redundant, dedicated metro fiber providing high-throughput, low-latency networking between

All traffic between AZs is encrypted

AZs are within 100 km (60 miles) of each other.



Montreal



@stevenwright Upsplash



Global Infrastructure – Fault Tolerance

Cheat sheets, Practice Exams and Flash cards www.exampro.co/clf-c01



Fault Domain

What is a fault domain?

A fault domain is a section of a network that is vulnerable to damage if a critical device or system fails. The purpose of a fault domain is that if a failure occurs **it will not cascade outside that domain**, limiting the damage possible.

You can have fault domains nested inside fault domains.

What is a fault level?

A fault level is a collection of fault domains.

The scope of a fault domain could be:

- specific servers in a rack
- an entire rack in a datacenter
- an entire room in a datacenter
- the entire data center building

It's up to the Cloud Service Provider (CSPs) to define the boundaries of a domain

An AWS Region would be a **Fault Level** →

Fault Level
us-east-1 (Region)

A Availability Zone would be a **Fault Domain** →

Fault Domain
us-east-1a (AZ)

Fault Domain
us-east-1b (AZ)

Global Infrastructure – Fault Tolerance

Cheat sheets, Practice Exams and Flash cards  www.exampro.co/clf-c01

Each Amazon Region is designed to be completely **isolated** from the other Amazon Regions.

- This achieves the greatest possible fault tolerance and stability

Each Availability Zone is **isolated**, but the Availability Zones in a Region are connected through low-latency links

Each Availability Zone is designed as an **independent failure zone**

- A "Failure Zone" is AWS describing a Fault Domain.

Failure Zone

- Availability Zones are physically separated within a typical metropolitan region and are located in lower risk flood plains
- discrete uninterruptible power supply (UPS) and onsite backup generation facilities
- data centers located in different Availability Zones are designed to be supplied by independent substations to reduce the risk of an event on the power grid impacting more than one Availability Zone.
- Availability Zones are all redundantly connected to multiple tier-1 transit providers



Multi-AZ for High Availability

If an application is partitioned across AZs, companies are better isolated and protected from issues such as **power outages, lightning strikes, tornadoes, earthquakes**, and more.

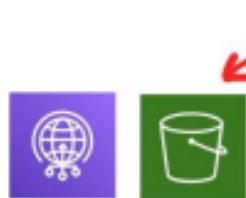
AWS Global Network

Cheat sheets, Practice Exams and Flash cards  www.exampro.co/clf-c01

The AWS Global Network represent the **interconnections between AWS Global Infrastructure**.

Commonly referred to as the “The Backbone of AWS”.

Think of it as private expressway, where things can move very fast between datacenters.



AWS Global Accelerator

AWS S3 Transfer Acceleration

Uses Edge Locations as an on-ramp to quickly reach AWS resources in other regions by traversing the fast AWS Global Network



Edge Locations can act as **on and off ramps** to the AWS Global Network



Amazon CloudFront (CDN)

Uses Edge Locations as an off-ramp, to provide at the Edge storage and compute near the end user.



VPC Endpoints

Ensuring your resources stay within the AWS Network and do no traverse over the public Internet.

Global Infrastructure – Point of Presence (PoP)

Cheat sheets, Practice Exams and Flash cards www.exampro.co/clf-c01

Points of Presence (PoP) is an intermediate location between an AWS Region and the end user, and this location could be a datacenter or collection of hardware.

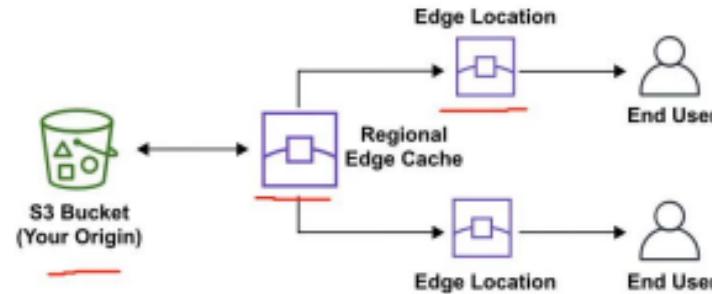
For AWS a Point of Presence is a data center **owned by AWS or a trusted partner** that is utilized by AWS Services related **for content delivery or expedited upload**.

PoP resources are:

- Edge Locations
- Regional Edge Caches

Edge Locations are datacenters that hold cached (copy) on the most popular files (eg. web pages, images and videos) so that the delivery of distance to the end users are reduced.

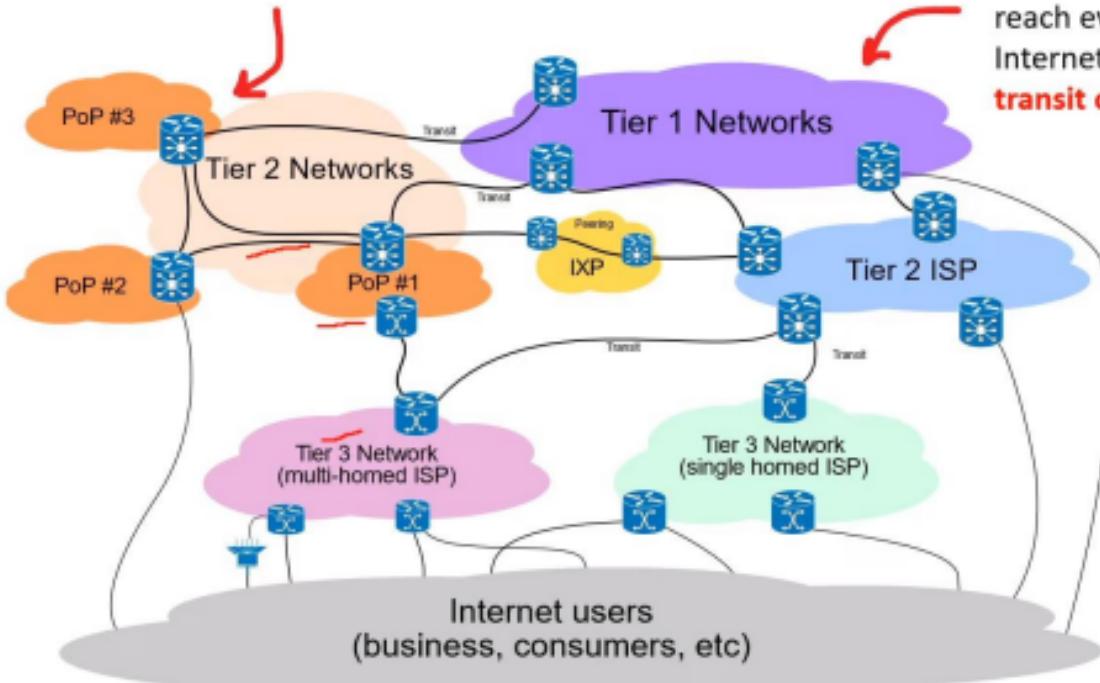
Regional Edge Locations are datacenters that hold much larger caches of less-popular files to reduce a full round trip and also to reduce the cost of transfer fees.



Global Infrastructure – Point of Presence (PoP)

Cheat sheets, Practice Exams and Flash cards  www.exampro.co/clf-c01

PoPs live at the edge/**intersection** of two networks



Tier 1 network is a network that can reach every other network on the Internet **without purchasing IP transit or paying for peering**.

AWS Availability Zones are all redundantly connected to multiple **tier-1 transit providers**

Global Infrastructure – Point of Presence (PoP)

Cheat sheets, Practice Exams and Flash cards  www.exampro.co/clf-c01

The following AWS Services use PoPs **for content delivery or expedited upload.**



Amazon CloudFront is a **Content Delivery Network (CDN) service** that:

- You point your website to CloudFront so that it will route requests to nearest Edge Location cache
- allows you to choose an **origin** (such as a web-server or storage) that will be source of cached
- caches the contents of what origin would returned to various Edge Locations around the world



Amazon S3 Transfer Acceleration allows you to generate a special URL that can be used by end users to upload files to a nearby Edge Location. Once a file is uploaded to an Edge Location, it can move much faster within the AWS Network to reach S3.



AWS Global Accelerator can find the optimal path from the end user to your web-servers. Global Accelerator are deployed within Edge Locations so you send user traffic to an Edge Location instead of directly to your web-application.

AWS Direct Connect

Cheat sheets, Practice Exams and Flash cards www.exampro.co/clf-c01



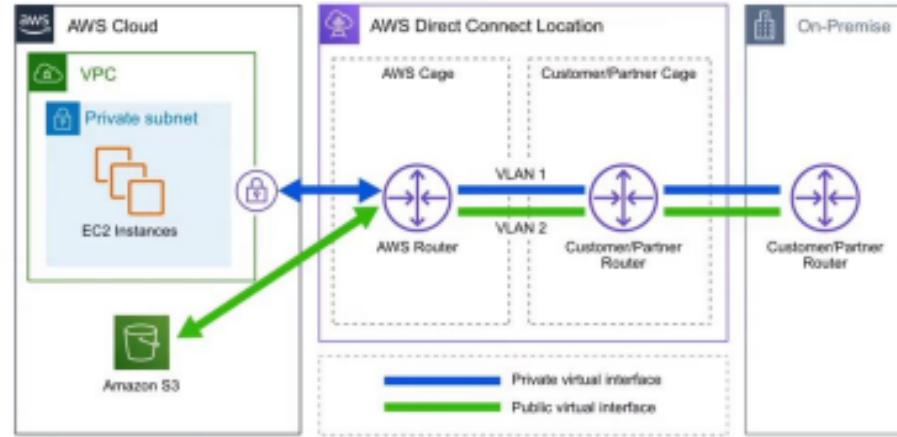
AWS Direct Connect is a **private/dedicated connection between your datacenter, office, co-location and AWS.**

Direct Connect has two **very-fast network** connection options:

1. Lower Bandwidth **50Mbps-500Mbps**
2. Higher Bandwidth **1GBps or 10GBps**



A co-location (aka carrier-hotel) is a data center where equipment, space, and bandwidth are available for rental to retail customers



Helps **reduce network costs** and **increase bandwidth throughput**. (great for high traffic networks)



Provides a **more consistent network experience** than a typical internet-based connection. (reliable and secure)

Global Infrastructure – Direct Connect Locations

Cheat sheets, Practice Exams and Flash cards  www.exampro.co/clf-c01

Direct Connect Locations are **trusted partnered datacenters** that you can establish a dedicated high speed, low-latency connection from your on-premise to AWS.



Allied Data Centres
250 Front Street West Toronto



You would use the **AWS Direct Connect** service to order and establish a connection

Global Infrastructure – Local Zones

Cheat sheets, Practice Exams and Flash cards www.exampro.co/clf-c01



Local Zones are datacenters located very close to a densely populated area to provide single-digit millisecond low latency performance (eg. 7ms) for that area.



- **Los Angeles, California** was the first Local Zone to be deployed
 - It is a logical extension of the US-West Region
 - The Identifier looks like the following: **us-west-2-lax-1a**
- Only specific AWS Services have been made available
 - EC2 Instance Types (T3, C5, R5, R5d, I3en, G4)
 - EBS (io1 and gp2)
 - Amazon FSx
 - Application Load Balancer
 - Amazon VPC

The purpose of Local Zone is the support highly-demanding applications sensitive to latencies:

- Media & Entertainment
- Electronic Design Automation
- Ad-Tech
- Machine Learning

To use Local Zones you need to Opt-In

Global Infrastructure – Wavelength Zones

Cheat sheets, Practice Exams and Flash cards www.exampro.co/clf-c01



AWS Wavelength Zones allows for **edge-computing on 5G Networks**.
Applications will have **ultra-low latency** being as close as possible to the users

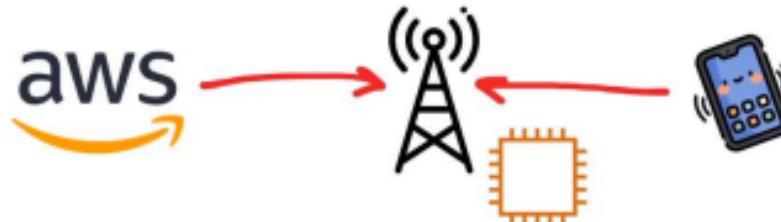
AWS has partnered with various Telecom companies to utilize their 5G networks

verizon ✓

KDDI

vodafone

SK telecom



You create a Subnet tied to a Wavelength Zone and then you can launch Virtual Machines (VMs) to the edge of the targeted 5G Networks.

Global Infrastructure – Data Residency

Cheat sheets, Practice Exams and Flash cards www.exampro.co/clf-c01

What is Data Residency?

The physical or geographic location of where an organization or cloud resources reside.

What is Compliance Boundaries?

A regulatory compliance (legal requirement) by a government or organization that describes where data and cloud resources are allowed to reside

What is Data Sovereignty?

Data Sovereignty is the jurisdictional control or legal authority that can be asserted over data because its physical location is within jurisdictional boundaries

For workloads that need to meet compliance boundaries strictly defining the data residency of data and cloud resources in AWS you can use:

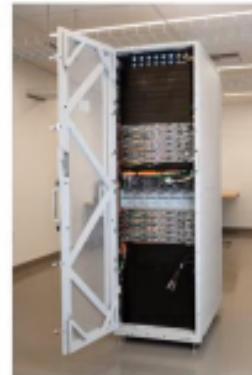


AWS Config is a Policy as Code service.

You can create rules to continuously check AWS resources configuration. If they deviate from your expectations you are alerted or AWS Config can in some cases auto-remediate.



IAM Policies can be written explicitly deny access to specific AWS Regions. A **Service Control Policy (SCP)** applies permissions organization wide.

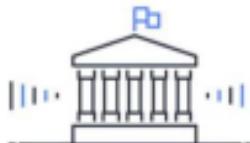


AWS Outposts is **physical rack of servers** that you can put in your data center. Your data will reside whenever the Outpost physically resides



Global Infrastructure – AWS for government

Cheat sheets, Practice Exams and Flash cards  www.exampro.co/clf-c01



What is Public Sector?

Public sectors include public goods and governmental services such as:

- military
- law enforcement
- infrastructure
- public transit
- public education
- health care
- the government itself

AWS can be utilized by public sector or organizations developing cloud workloads for the public sector.

AWS achieves this by meeting **regulatory compliance programs** along with specific governance and security controls



AWS has special regions for US regulation called **GovCloud**

Global Infrastructure – GovCloud (US)

Cheat sheets, Practice Exams and Flash cards  www.exampro.co/clf-c01



Federal Risk and Authorization Management Program (FedRAMP)

a US government-wide program that provides a standardized approach to security assessment, authorization, and continuous monitoring for cloud products and services.

What is GovCloud?

A Cloud Service Provider (CSP) generally will offer an **isolated region** to run FedRAMP workloads.



AWS GovCloud Regions allow customers to host sensitive **Controlled Unclassified Information** and other types of regulated workloads.

- GovCloud Regions are only operated by employees who are U.S. citizens, on U.S. soil.
- They are **only** accessible to U.S. entities and root account holders who pass a screening process

Customers can architect secure cloud solutions that comply with:

- FedRAMP High baseline
- DOJ's Criminal Justice Information Systems (CJIS) Security Policy
- U.S. International Traffic in Arms Regulations (ITAR)
- Export Administration Regulations (EAR)
- Department of Defense (DoD) Cloud Computing Security Requirements Guide

Global Infrastructure – AWS in China

Cheat sheets, Practice Exams and Flash cards  www.exampro.co/clf-c01



AWS China is the AWS cloud offerings in Mainland China.

AWS China is completely isolate *intentionally* from AWS Global to meet regulatory compliance for Mainland China.

AWS China is on its own domain at: amazonaws.cn

In order to operate in a AWS China Region you need have a Chinese Business License (ICP license)

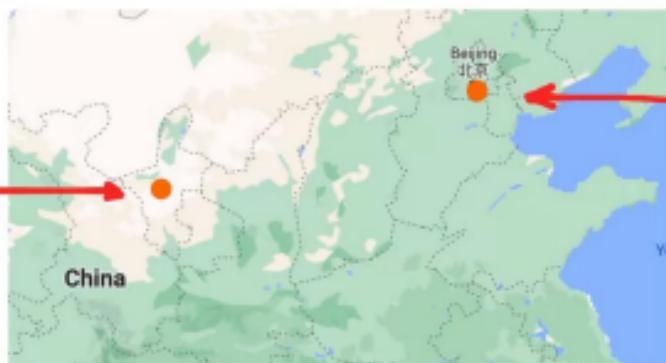
Not all services are available in china eg. Route53

Running in Mainland China (instead of Singapore) means you would not need to traverse the The Great Firewall.

AWS has two Regions in Mainland China:



Ningxia CN-NorthWest-1
Operated by NSWCF



Beijing CN-North-1
operated by SINNET



Global Infrastructure – Sustainability

Cheat sheets, Practice Exams and Flash cards  www.exampro.co/clf-c01



Amazon co-founded the Climate Pledge to achieve Net-Zero Carbon Emissions by 2040 across all of Amazon's business (this includes AWS)
sustainability.aboutamazon.com

AWS Cloud's Sustainability goals are composed of three parts:

1. Renewable Energy

AWS is working towards having their AWS Global Infrastructure powered by 100% renewable energy by 2025.

AWS purchases and retires environmental attributes to cover the non-renewable energy for AWS Global Infrastructure:

- Renewable Energy Credits (RECs)
- Guarantees of Origin (GOs)

2. Cloud Efficiency

AWS's infrastructure is 3.6 times more energy efficient than the median of U.S. enterprise data centers surveyed.

3. Water Stewardship

Direct evaporative technology to cool our data center

Use of non-potable water for cooling purposes (recycled water)

On-site water treatment allows us to remove scale-forming minerals and reuse water for more cycles

Water efficiency metrics to determine and monitor optimal water use for each AWS Region

Global Infrastructure – AWS Ground Station

Cheat sheets, Practice Exams and Flash cards  www.exampro.co/clf-c01



AWS Ground Station is a fully managed service that **lets you control satellite communications**, process data, and scale your operations without having to worry about building or managing your own ground station infrastructure

Use cases for Ground Station: To use Ground Station:

- weather forecasting
 - You schedule a Contact (select satellite, start and end time, and the ground location)
 - use the AWS Ground Station EC2 AMI to launch EC2 instances that will uplink and downlink data during the contact or receive downlinked data in an Amazon S3 bucket.
- surface imaging
- communications
- video broadcasts

Use Case:

A company reaches an agreement with a Satellite Imagery Provider to take satellite photos of a specific region. They use AWS Ground Station to communicate that company's Satellite and download the S3 image data.



@isidurumm on Unsplash



Global Infrastructure – AWS Outposts

Cheat sheets, Practice Exams and Flash cards  www.exampro.co/clf-c01



AWS Outposts is a fully managed service that offers the same AWS infrastructure, AWS services, APIs, and tools to virtually any datacenter, co-location space, or on-premises facility for a truly consistent hybrid experience.

AWS Outposts is rack of servers running AWS Infrastructure on your physical location

42U Rack



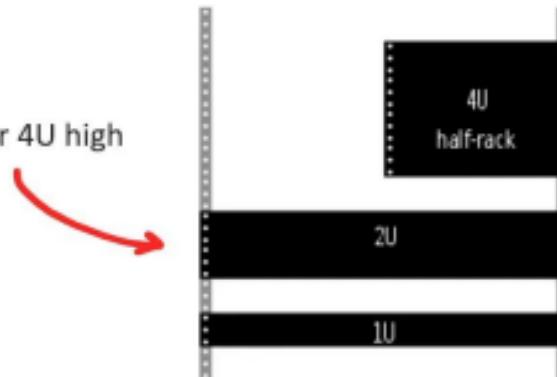
What is a Server Rack?

A frame design to hold and organize IT equipment.

Rack Heights

U stands for “rack units” or “U spaces” with is equal to 1.75 inches. The industry standard rack size is 48U (7 Foot Rack)

- full-size rack cage is 42U high
 - equipment is typically 1U, 2U, 3U, or 4U high



Global Infrastructure – AWS Outposts

Cheat sheets, Practice Exams and Flash cards  www.exampro.co/clf-c01

AWS Outposts comes in 3 form factors: 42U, 1U and 2U

This a full rack of servers provided by AWS

42U



These are servers that you can place into your existing racks:

1U

suitable for 19-inch wide
24-inch deep cabinets
AWS Graviton2 (up to 64 vCPUs)
128 GiB memory
4 TB of local NVMe storage

2U

suitable for 19-inch wide
36-inch deep cabinets,
Intel processor (up to 128 vCPUs)
256 GiB memory
8TB of local NVMe storage

AWS delivers it to your preferred physical site fully assembled and ready to be rolled into final position. It is installed by AWS and the rack needs to be simply plugged into power and network.