

AWS Global Infrastructure Overview

REGIONS · AVAILABILITY ZONES · DATA CENTERS · EDGE LOCATIONS · SERVICE CATEGORIES

STUDY & REVIEW

01 MUST KNOW

if you remember five things, remember these

- 01 The hierarchy:** a **Region** is a geographical area containing two or more **Availability Zones**; each AZ is one or more discrete **data centers**; **edge locations** sit outside this hierarchy, caching content near users. Customer choice stops at the AZ — never a specific data center.
- 02 Regions are isolated:** data stored in a Region is **never automatically replicated** to another Region — cross-Region replication is your decision and your responsibility. Communication between Regions rides the **AWS backbone network**.
- 03 Pick a Region on four factors:** **data governance / legal** requirements · **proximity to customers** (latency) · **services available** in the Region · **cost**, which varies by Region.
- 04 AZs are built for fault isolation:** own power infrastructure, physically separated by many kilometers (all within 100 km), interconnected by high-bandwidth, low-latency, fully redundant dedicated fiber — fast enough for **synchronous replication**. Replicate across AZs for resiliency.
- 05 Edge = low latency:** CloudFront (CDN) and Route 53 (DNS) automatically answer from the **nearest edge location**; **Regional edge caches** hold content too infrequently accessed to stay at the edge. “CloudFront + low latency” on the exam ⇒ *edge locations*.

02 KEY TERMS

TERM	DEFINITION
AWS Region	Geographical area with two or more Availability Zones; fully redundant and isolated from other Regions
Availability Zone	Fully isolated partition of AWS infrastructure: one or more discrete data centers designed for fault isolation
Data center	Facility where data resides and processing occurs; typically 50,000–80,000 physical servers
Edge location	Point of Presence site that caches content close to end users — the workhorse of CloudFront
Regional edge cache	Larger cache holding content accessed too infrequently to remain at an edge location
Points of Presence	AWS's global network of edge locations plus a much smaller number of Regional edge caches
AWS backbone network	Private AWS network infrastructure that carries communication between Regions
Amazon CloudFront	AWS's global content delivery network (CDN) — delivers content to users with reduced latency
Amazon Route 53	AWS's DNS service; requests are routed to the nearest edge location automatically
Elasticity	Capacity dynamically adapts to increases or decreases in requirements
Fault tolerance	Continues operating properly despite a component failure — redundancy is built in
High availability	High operational performance and minimized downtime, with no human intervention

03 THE INFRASTRUCTURE HIERARCHY

THE comparison this module exists for

LAYER	WHAT IT IS	REMEMBER
Region	Geographical area holding two or more Availability Zones	Isolated from other Regions; no automatic cross-Region replication; inter-Region traffic uses the AWS backbone; some Regions disabled by default or restricted

LAYER	WHAT IT IS	REMEMBER
Availability Zone	Fully isolated partition of AWS infrastructure — one or more discrete data centers (typically three)	Own power; many km apart yet all within 100 km; linked by redundant private fiber (synchronous replication); you choose your AZs
Data center	Where data resides and data processing occurs	50,000–80,000 servers; redundant power, networking, connectivity; undisclosed locations; cannot be selected — AZ is the most granular choice
Edge location	Point of Presence that caches content close to users	Used by CloudFront, Route 53, Shield, and WAF; located in most major cities; backed by Regional edge caches

04 CHOOSING A REGION — FOUR FACTORS

compliance first, then latency, services, price

FACTOR	ASK	DECK DETAIL
Data governance & legal	Must data stay inside a boundary?	Local laws (e.g., EU Data Protection Directive) can restrict where content and services may live
Proximity to customers	Which Region minimizes latency?	Run as close to users as possible; CloudPing tests latency to every Region
Services available	Does this Region offer what I need?	Not all services are available in all Regions
Cost	What does the same workload cost here?	On-Demand t3.medium Linux: \$0.0416/hr in US East (Ohio) vs \$0.0544/hr in Asia Pacific (Tokyo)

05 GLOBAL VS REGIONAL

straight from the console activity

GLOBAL <i>one instance spans the whole account</i> — IAM — users, groups, policies apply everywhere — Amazon Route 53 (DNS)	REGIONAL <i>the console Region selector matters</i> — Amazon EC2 — AWS Lambda — VPC — exists at the Region level — Subnet — exists at the Availability Zone level
---	---

06 SERVICE CATEGORIES & CORE SERVICES

23 categories exist; the course + CCP exam lean on these 7

CATEGORY	CORE SERVICES IN THIS COURSE
Storage	Amazon S3 · Amazon EBS · Amazon EFS · Amazon S3 Glacier
Compute	Amazon EC2 · EC2 Auto Scaling · Amazon ECS · Amazon ECR · AWS Elastic Beanstalk · AWS Lambda · Amazon EKS · AWS Fargate
Database	Amazon RDS · Amazon Aurora · Amazon DynamoDB · Amazon Redshift
Networking & Content Delivery	Amazon VPC · Elastic Load Balancing · Amazon CloudFront · AWS Transit Gateway · Amazon Route 53 · AWS Direct Connect · AWS VPN
Security, Identity & Compliance	IAM · AWS Organizations · Amazon Cognito · AWS Artifact · AWS KMS · AWS Shield
Cost Management	AWS Cost and Usage Report · AWS Budgets · AWS Cost Explorer
Management & Governance	Management Console · AWS Config · Amazon CloudWatch · AWS Auto Scaling · AWS CLI · Trusted Advisor · Well-Architected Tool · AWS CloudTrail

- ✗ “Data stored in one Region is automatically replicated to other Regions.” — **False**. Regions are isolated for fault tolerance and stability; cross-Region replication is **your** responsibility if the business requires it.
- ✗ “An Availability Zone is a single data center.” — An AZ is **one or more discrete data centers** (typically three), together forming a fully isolated partition of the infrastructure.
- ✗ “You can deploy resources into a specific data center.” — The **Availability Zone is the most granular level** a customer can specify; data-center locations aren't even disclosed.
- ✗ “CloudFront achieves low-latency delivery by using Availability Zones.” — It uses **edge locations** (with Regional edge caches behind them). This is exactly the module's sample exam question.
- ✗ “Every Region is ready to use out of the box.” — Regions introduced **after March 20, 2019** (e.g., Hong Kong, Bahrain) are **disabled by default**; China's Regions require a separate account and GovCloud (US) is restricted to US government workloads.
- ✗ “Every service is available in every Region, at the same price.” — **Both false**. Service availability varies by Region, and so do prices (t3.medium: \$0.0416/hr Ohio vs \$0.0544/hr Tokyo).
- ✗ “EC2 and Lambda are global services.” — They're **Regional**. Of the activity's four, only **IAM and Route 53** are global.
- ✗ “Regional edge caches hold the most popular content.” — Backwards: they absorb content accessed **too infrequently** to remain at an edge location, sparing a round trip to the origin server.

08 SELF-QUIZ

answers are upside-down below

- Q1** Order these from largest to smallest: data center, Region, Availability Zone.
- Q2** Name the four factors for selecting a Region.
- Q3** What is the most granular infrastructure level a customer can specify when deploying resources?
- Q4** Describe the network that interconnects the AZs within a Region.
- Q5** Which two location types make up AWS Points of Presence?
- Q6** What kind of content ends up in a Regional edge cache?
- Q7** Of EC2, IAM, Lambda, and Route 53 — which are global?
- Q8** At which infrastructure level does a VPC exist? A subnet?
- Q9** Which service category holds IAM? Which holds Amazon VPC?
- Q10** Sample CCP exam: which component of the AWS global infrastructure does CloudFront use to ensure low-latency delivery — Regions, edge locations, Availability Zones, or Amazon VPC?

ANSWER KEY (rotate the page)

Q1 Region → Availability Zone → data center **Q2** data governance/legal, proximity to customers (latency), services available, cost **Q3** the Availability Zone → never a specific data center **Q4** high-bandwidth, low-latency, fully redundant dedicated fiber, supports synchronous replication **Q5** edge locations and Regional edge caches **Q6** content accessed too infrequently to remain at an edge location **Q7** IAM and Route 53 (EC2 and Lambda are Regional) **Q8** VPC: Region level · subnet: Availability Zone level **Q9** IAM: Security, Identity, & Compliance · VPC: Networking & Content Delivery **Q10** edge locations