

Week 2: AWS Global Infrastructure

Understanding Regions, Availability Zones, and Edge Locations

■ Study Guide

These notes are designed to be comprehensive yet concise. Read through each section carefully, paying attention to key concepts, exam tips, and practical examples. Use these notes for quick review before assessments and as a reference during your AWS certification preparation.

■ AWS Global Infrastructure Overview

AWS operates the largest global cloud infrastructure, designed to deliver high availability, fault tolerance, and low latency to customers worldwide.

Key Components: The AWS infrastructure consists of Regions, Availability Zones (AZs), and Edge Locations, each serving specific purposes in delivering cloud services.

■ ■ AWS Regions

Definition: A Region is a physical location around the world where AWS clusters data centers. Each Region is completely independent.

Key Characteristics:

- Geographic areas with multiple isolated locations
- Full independence - no data replication between Regions without explicit action
- Named with geographic identifiers (e.g., us-east-1, eu-west-1, ap-southeast-1)

■ How to Choose a Region:

1. **Compliance & Data Residency:** Legal requirements for data location
2. **Latency:** Proximity to your users reduces response times
3. **Service Availability:** Not all services are available in all Regions
4. **Pricing:** Costs vary by Region based on local factors

■ **Exam Tip:** Always consider compliance first when choosing a Region!

■ Availability Zones (AZs)

Definition: An Availability Zone is one or more discrete data centers with redundant power, networking, and connectivity within a Region.

Key Features:

- Multiple AZs per Region (minimum 3 for most Regions)
- Physical separation protects against localized failures
- High-speed, low-latency connections between AZs

- Independent power, cooling, and networking

Multi-AZ Architecture Benefits:

- **High Availability:** Continue operating even if one AZ fails
- **Fault Tolerance:** Isolate and contain failures
- **Disaster Recovery:** Automatic failover capabilities

■ **Exam Tip:** Multi-AZ deployments provide high availability, while multi-Region provides disaster recovery

■ Edge Locations

Definition: Edge Locations are data centers used by AWS CloudFront (CDN) to cache content closer to end users.

Key Services Using Edge Locations:

- **CloudFront (CDN):** Delivers content with low latency
- **Route 53:** Global DNS service with health checking
- **AWS Global Accelerator:** Improves application performance

Differences: Edge vs Region

- More Edge Locations than Regions (200+ Edge Locations worldwide)
- Edge Locations cache content, Regions host full services
- Edge Locations reduce latency for content delivery

■ **Exam Tip:** Edge Locations are for content delivery, not for running EC2 instances!

■ Key Takeaways

- **Regions** provide geographic separation and comply with data residency requirements
- **Availability Zones** enable high availability within a Region
- **Edge Locations** reduce latency for content delivery
- Design for failure by using multiple AZs

- Consider compliance, latency, service availability, and pricing when choosing Regions

■ Study Tips & Review Strategies

Active Recall: After reading, close these notes and try to explain key concepts out loud.

Spaced Repetition: Review these notes on Day 1, Day 3, Day 7, and Day 14 for best retention.

Practice Questions: Test yourself on exam-style questions related to these topics.

Hands-On Practice: Apply concepts in the AWS Console or Free Tier account when possible.

Study Groups: Discuss concepts with classmates to identify gaps in understanding.

Remember: Understanding beats memorization. Focus on grasping WHY things work, not just WHAT they are.
Good luck with your studies! ■