

AWS Academy Cloud Foundations

Module 3: AWS Global Infrastructure Overview



Topics

- AWS Global Infrastructure
- AWS service and service category overview

Demo

- AWS Global Infrastructure

Activities

- AWS Management Console clickthrough



Knowledge check

After completing this module, you should be able to:

- Identify the difference between AWS Regions, Availability Zones, and edge locations
- Identify AWS service and service categories

Module 3: AWS Global Infrastructure Overview

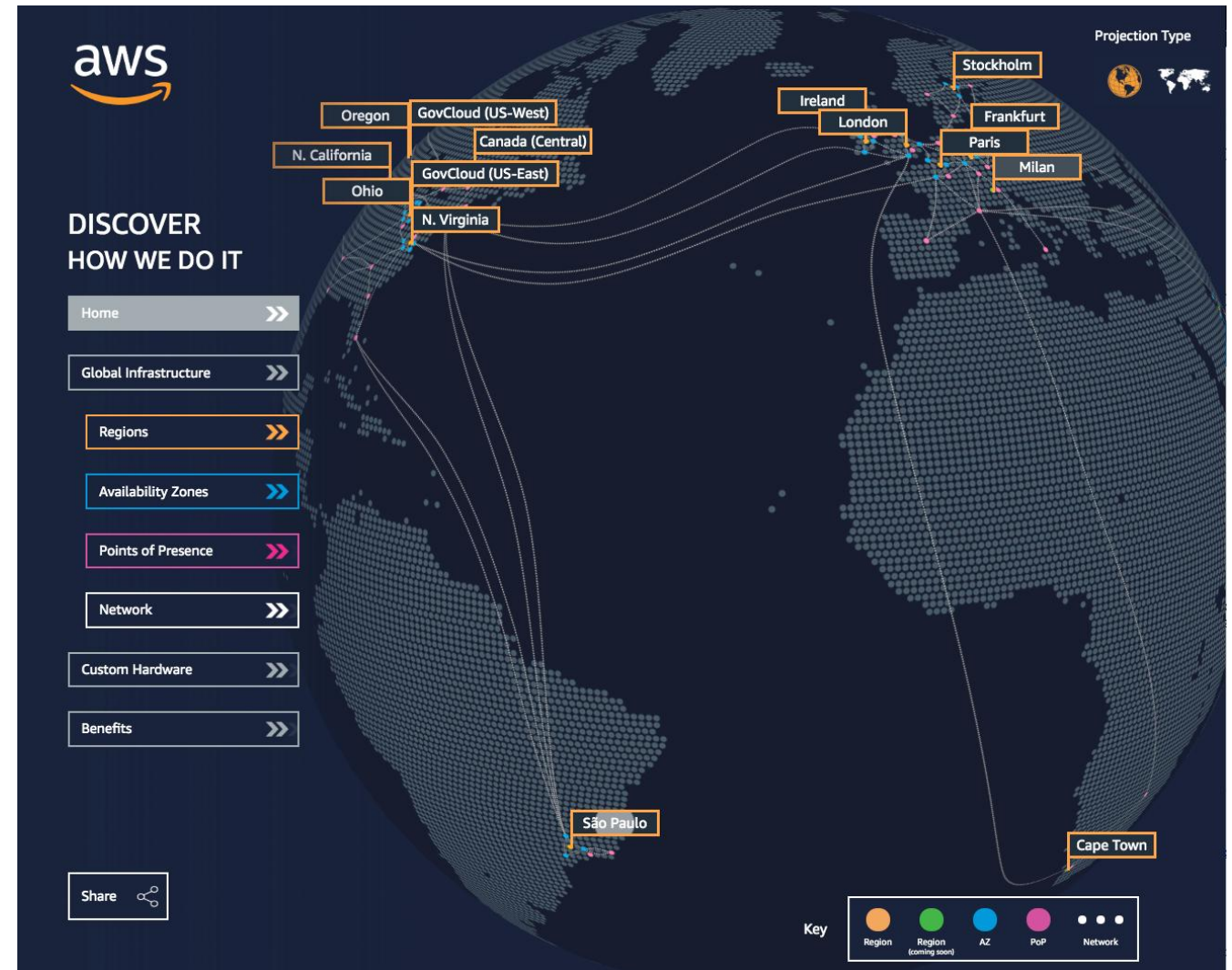
Section 1: AWS Global Infrastructure

AWS Global Infrastructure

- The **AWS Global Infrastructure** is designed and built to deliver a **flexible**, **reliable**, **scalable**, and **secure** cloud computing environment with high-quality **global network performance**.
- This map from <https://infrastructure.aws> shows the current **AWS Regions** and more that are coming soon.



Educator-Led Demo: AWS Global Infrastructure Details



AWS Regions

- An **AWS Region** is a geographical area.
 - **Data replication** across Regions is controlled by you.
 - **Communication** between Regions uses AWS backbone network infrastructure.
- Each Region provides full redundancy and connectivity to the network.
- A Region typically consists of two or more **Availability Zones**.



Example: London Region

Selecting a Region

Determine the right Region for your services, applications, and data based on these factors



Data governance, legal requirements



Proximity to customers (latency)



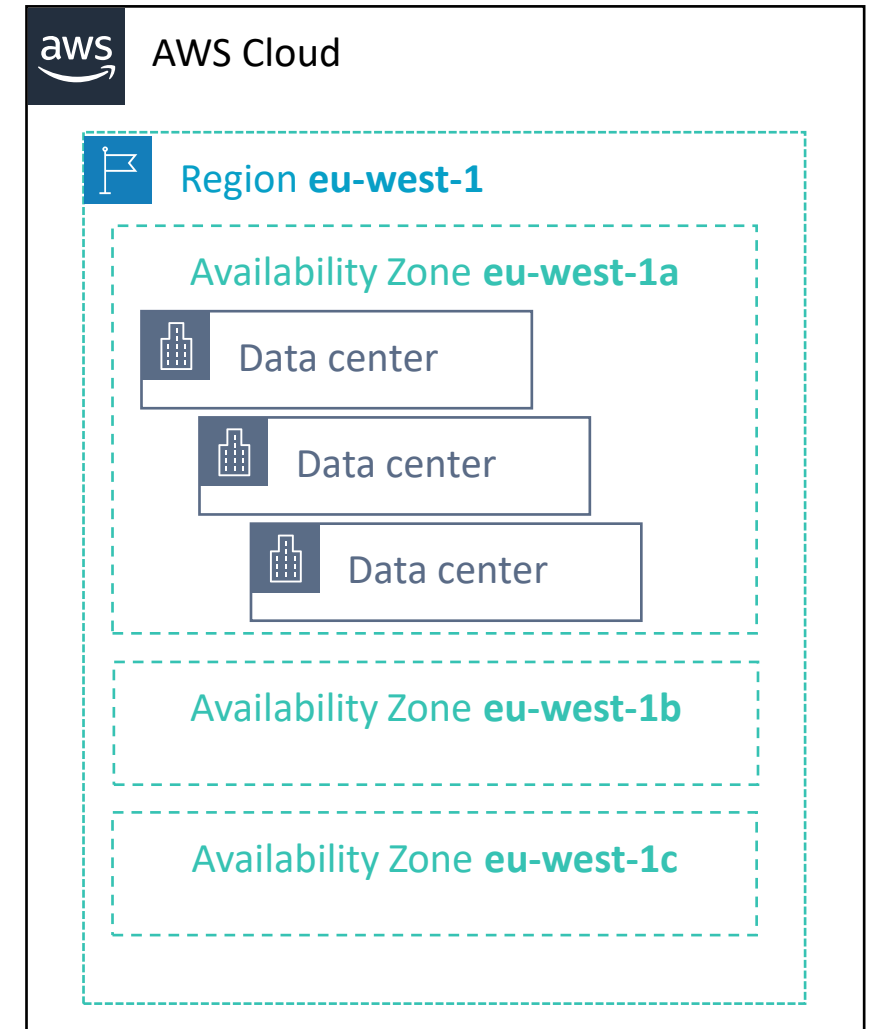
Services available within the Region



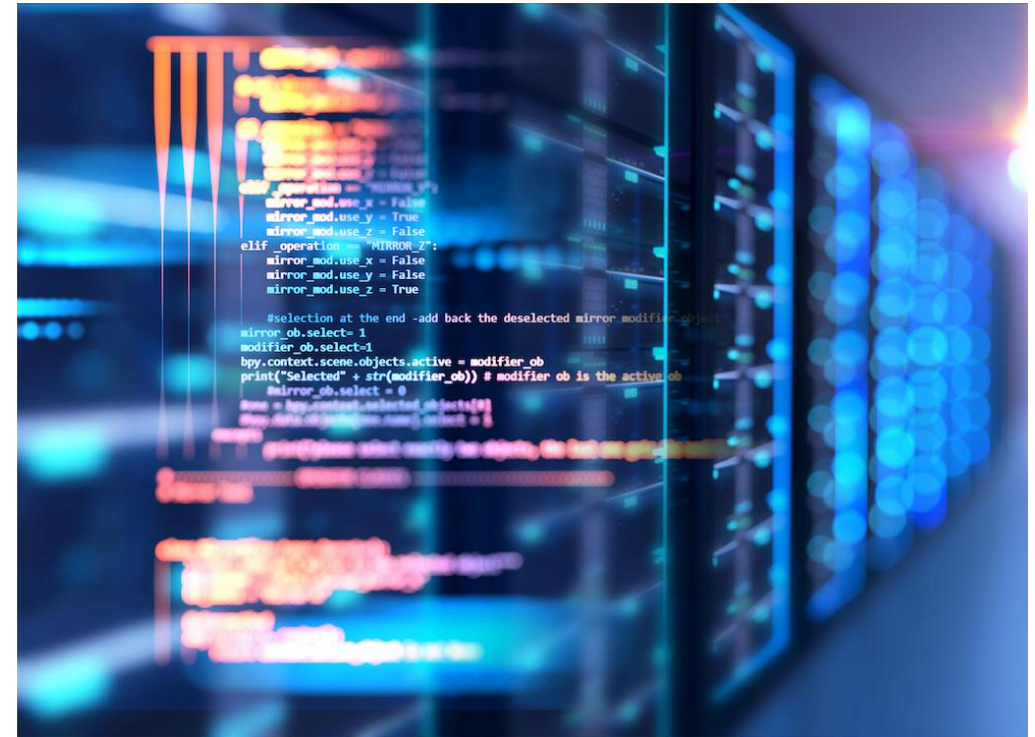
Costs (vary by Region)

Availability Zones

- Each **Region** has multiple Availability Zones.
- Each **Availability Zone** is a fully isolated partition of the AWS infrastructure.
 - There are currently 84 Availability Zones worldwide
 - Availability Zones consist of discrete **data centers**
 - They are designed for fault isolation
 - They are interconnected with other Availability Zones by using high-speed private networking
 - You choose your Availability Zones.
 - **AWS recommends replicating data and resources across Availability Zones** for resiliency.

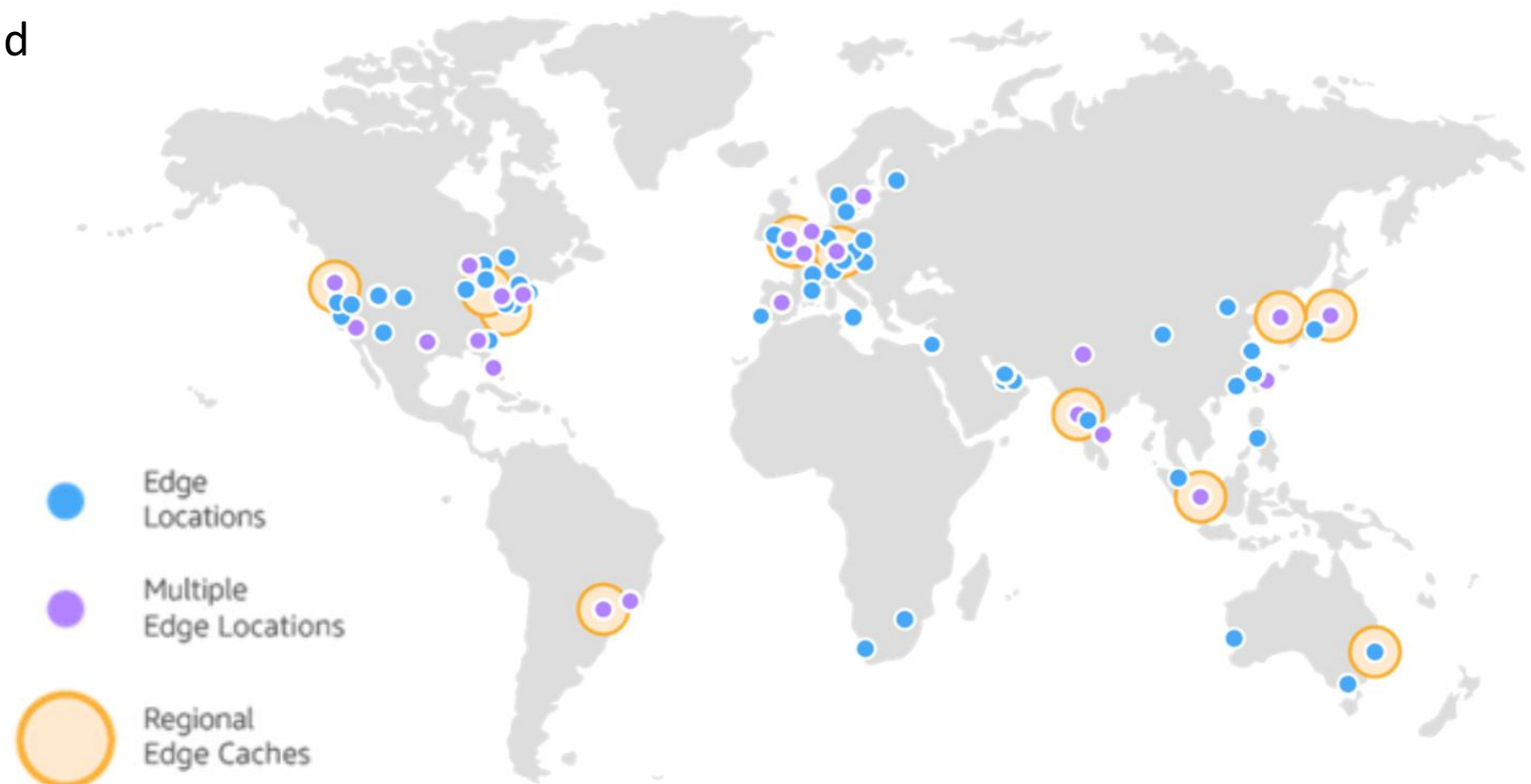


- AWS data centers are **designed for security**.
- Data centers are where the data resides and data processing occurs.
- Each data center has redundant power, networking, and connectivity, and is housed in a separate facility.
- A data center typically has 50,000 to 80,000 physical servers.

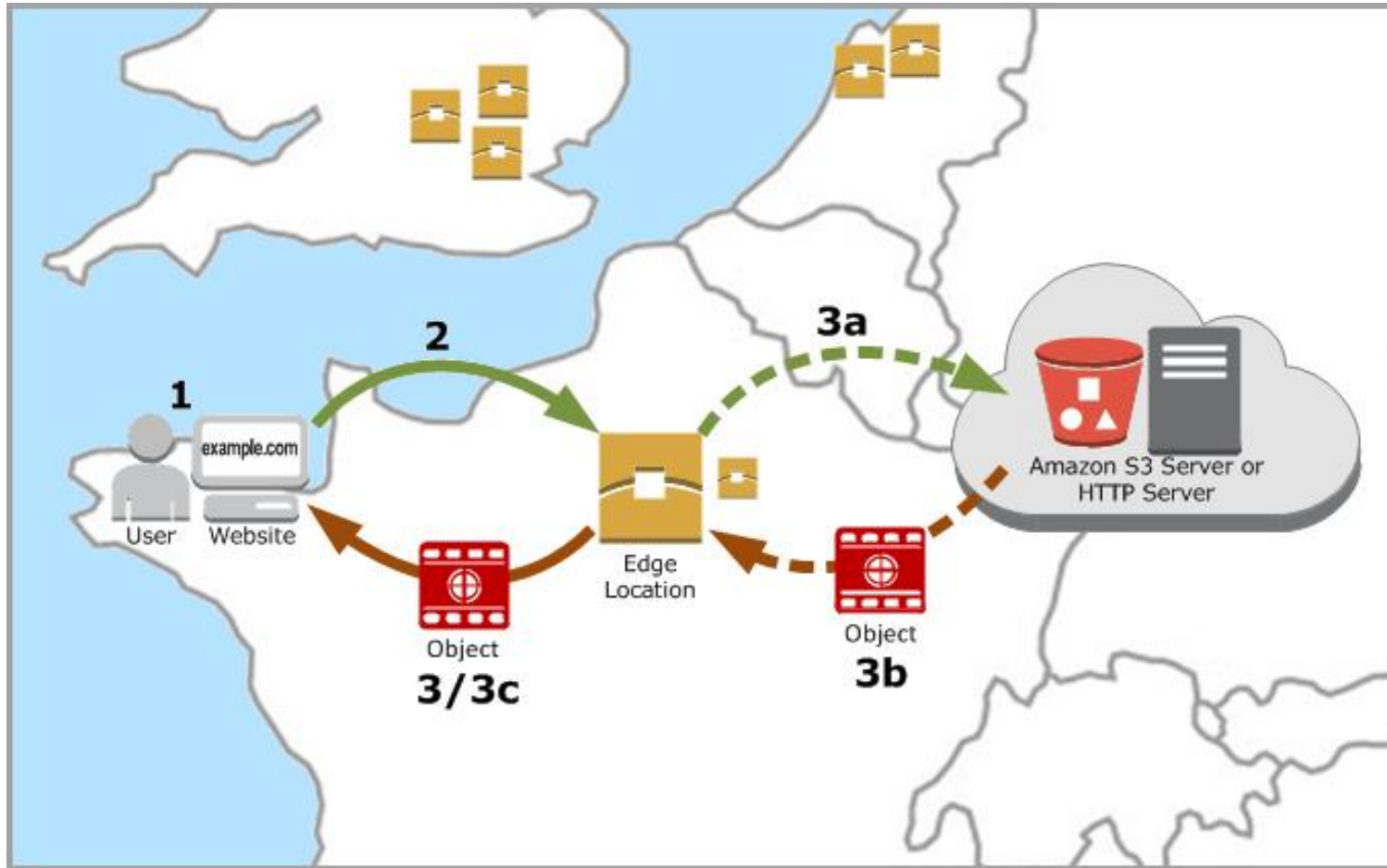


Points of Presence

- AWS provides a global network of 187 **Points of Presence** locations
- Consists of 176 **edge locations** and 11 **Regional edge caches**
- Used with Amazon CloudFront
 - A global Content Delivery Network (CDN), that delivers content to end users with **reduced latency**
- Regional edge caches used for content with infrequent access.



How CloudFront delivers content to your users



How CloudFront delivers content to your users



1. User ทำการเข้าถึงเว็บไซต์หรือแอปพลิเคชันของคุณและขอไฟล์อย่างน้อยหนึ่งไฟล์ เช่น image file และ HTML file
2. DNS ทำการกำหนดเส้นทางคำขอไปยัง CloudFront POP (edge location) ที่สามารถส่ง content ได้ดีที่สุด—โดยทั่วไปคือ CloudFront POP ที่ตำแหน่งใกล้ที่สุดโดยมองในแง่ของเวลา หรือ Latency —และทำการกำหนดเส้นทางเพื่อส่งคำขอไปยัง Edge location นั้น
3. ในตัว POP, CloudFront จะตรวจสอบแคชสำหรับไฟล์ที่ผู้ใช้ร้องขอ หากไฟล์อยู่ในแคช CloudFront จะส่งไฟล์ให้กับผู้ใช้ หากไฟล์ไม่อยู่ในแคช ให้ทำดังต่อไปนี้:
 - a. CloudFront เปรียบเทียบคำขอกับข้อกำหนดต่าง ๆ ในการแจกจ่ายข้อมูลของระบบและส่งต่อคำขอสำหรับไฟล์นี้ไปยังเซิร์ฟเวอร์ต้นทาง ตามประเภทของไฟล์ที่เกี่ยวข้อง—เช่น ค้นหา image file ที่บันทึกอยู่ใน Amazon S3 และทำการค้นหา HTML Files ที่อยู่ใน HTTP server
 - b. Origin servers ทำการส่งไฟล์กลับไป edge location
 - c. ทันทีที่ไบนารีแรกส่งมาจาก origin ตัว CloudFront จะเริ่มส่งต่อไฟล์ไปยังผู้ใช้ CloudFront ยังทำการเพิ่มไฟล์ไปยังแคชที่อยู่ภายใน Edge location สำหรับผู้ใช้รายอื่นที่จะเข้ามาทำการร้องขอไฟล์ในครั้งต่อไป

- CloudFront points of presence (POPs) (edge locations) ทำการตรวจสอบให้แน่ใจว่าเนื้อหาที่มักถูกเรียกใช้บ่อยสามารถแสดงต่อผู้ใช้งานได้อย่างรวดเร็ว CloudFront ยังมี **regional edge caches** เพื่อนำเนื้อหาที่ต้องการมาอยู่ใกล้กับผู้ร้องขอ (Users) มากขึ้น แม้ว่าเนื้อหาจะไม่ได้มีความถี่ที่ต้องการเพียงพอที่จะอยู่ที่ POP แต่เพื่อช่วยปรับปรุงประสิทธิภาพสำหรับ content นั้น
- **Regional edge caches** ช่วยในเรื่องของเนื้อหาทุกประเภทโดยเฉพาะเนื้อหาที่มีแนวโน้มที่จะได้รับความนิยมน้อยลงเมื่อเวลาผ่านไป ตัวอย่างเช่นเนื้อหาที่ผู้ใช้สร้างขึ้น เช่น วิดีโอ ภาพถ่าย หรืองานศิลปะ ส่วนเนื้อหาลักษณะของอีคอมเมิร์ซ เช่น ภาพถ่ายสินค้าและวิดีโอ และข่าวและเนื้อหาที่เกี่ยวข้องกับเหตุการณ์ที่อาจพบความนิยมใหม่

Edge Location vs Regional Edge caches

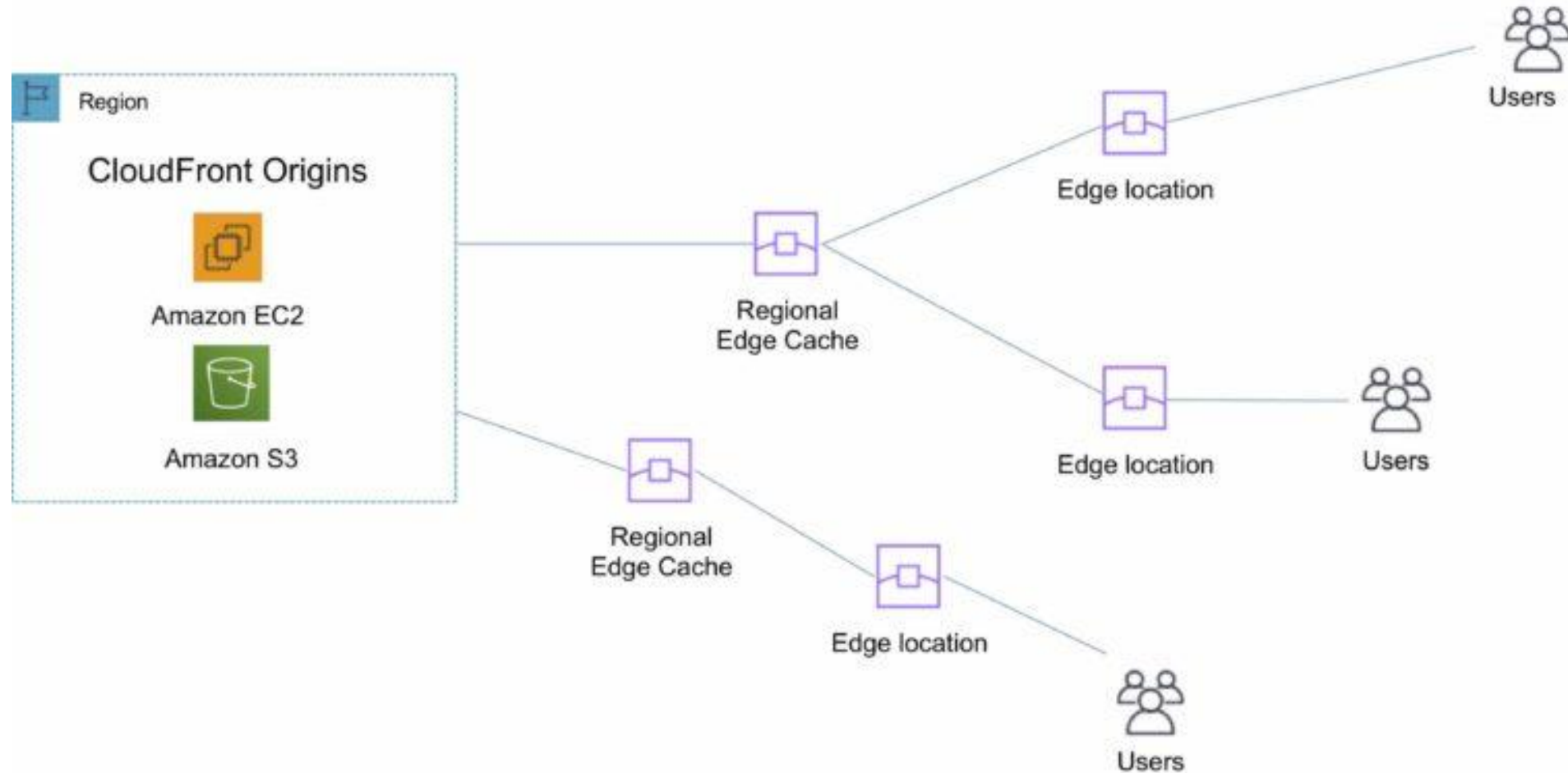
Edge Location

- Edge Location คือ การระบุตำแหน่งของ **Cached** ซึ่งแยกจาก **AWS regions/AZ**
- หากรับคำขอข้อมูลจากผู้ใช้ระบบจะทำการส่งข้อมูลที่อยู่ใน **edge location** ที่ใกล้ที่สุดโดยอัตโนมัติ
- Edge Location ไม่ถูกเชื่อมโยงกับ **Availability Zones** หรือ **Regions**
- ข้อมูลที่เป็นแบบ **Dynamic content** จะถูกส่งตรงไปยัง **origin** เลย โดยไม่ผ่าน **Regional Edge caches**
- Edge location เราสามารถอ่าน เขียน ได้

Regional Edge caches

- วัตถุประสงค์ของการสร้าง **Regional Edge caches** เพื่อให้ผู้ใช้งานเข้าถึง **content** มากขึ้นและรวดเร็วขึ้น
- ตำแหน่งของ **Regional Edge Caches** อยู่ระหว่าง **origin web servers** กับ **global edge locations** และขนาดของ **regional edge caches** มีขนาดของ **caches** ใหญ่กว่า
- **Regional Edge caches** ที่มีขนาดใหญ่กว่า **edge location** สามารถเก็บข้อมูล หรือ **Object data** ได้นานกว่า
- หากมีการใช้ **proxy methods** เช่น **PUT/POST/PATCH/OPTIONS/DELETE** สามารถส่งไปที่ต้นทางโดยตรงจาก **Edge location** โดยคำสั่งจะไม่ผ่านจาก **Regional Edge caches**
- ข้อมูลที่เป็นแบบ **Dynamic content** จะถูกส่งตรงไปยัง **origin** เลย โดยไม่ผ่าน **Regional Edge caches**

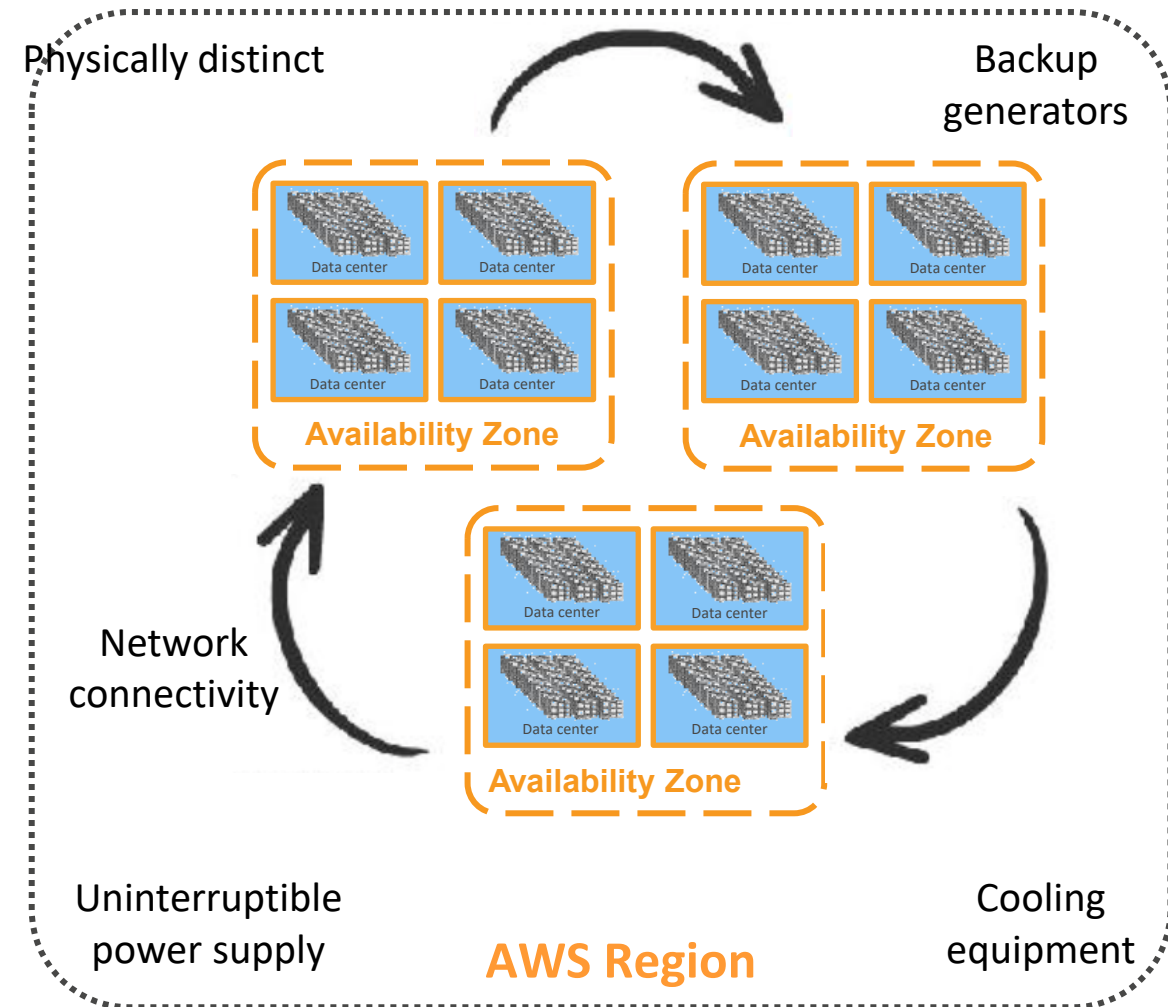
Example : Regional Edge Location and Edge location



Reference : <https://digitalcloud.training/certification-training/aws-solutions-architect-associate/networking-and-content-delivery/amazon-cloudfront/>

AWS infrastructure features

- Elasticity and scalability
 - Elastic infrastructure; dynamic adaption of capacity
 - Scalable infrastructure; adapts to accommodate growth
- Fault-tolerance
 - Continues operating properly in the presence of a failure
 - Built-in redundancy of components
- High availability
 - High level of operational performance
 - Minimized downtime
 - No human intervention



Key takeaways



- The **AWS Global Infrastructure** consists of **Regions** and **Availability Zones**.
- Your choice of a **Region** is typically based on **compliance requirements** or to **reduce latency**.
- Each **Availability Zone** is physically separate from other Availability Zones and has redundant power, networking, and connectivity.
- **Edge locations**, and **Regional edge caches** improve performance by **caching** content closer to users.

Module 3: AWS Global Infrastructure Overview

Section 2: AWS services and service category overview

AWS foundational services

Applications



Virtual desktops



Collaboration and sharing

Platform Services

Databases

Relational

NoSQL

Caching

Analytics

Cluster computing

Real-time

Data warehouse

Data workflows

Application services

Queuing

Orchestration

App Streaming

Transcoding

Email

Search

Deployment and management

Containers

DevOps tools

Resource templates

Usage tracking

Monitoring and logs

Mobile Services

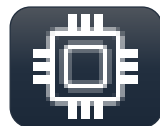
Identity

Sync

Mobile Analytics

Notifications

Foundation Services



Compute (virtual, automatic scaling, and load balancing)



Networking



Storage (object, block, and archive)

Infrastructure

Regions

Availability Zones



Edge locations

AWS categories of services



Analytics



Application
Integration



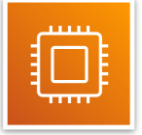
AR and VR



Blockchain



Business
Applications



Compute



Cost
Management



Customer
Engagement



Database



Developer Tools



End User
Computing



Game Tech



Internet
of Things



Machine
Learning



Management and
Governance



Media Services



Migration and
Transfer



Mobile



Networking and
Content Delivery



Robotics



Satellite



Security, Identity, and
Compliance



Storage

Storage service category



Photo from <https://www.pexels.com/photo/black-and-grey-device-159282/>



AWS storage services



Amazon Simple Storage Service (Amazon S3)



Amazon Elastic Block Store (Amazon EBS)



Amazon Elastic File System (Amazon EFS)



Amazon Simple Storage Service Glacier

Compute service category

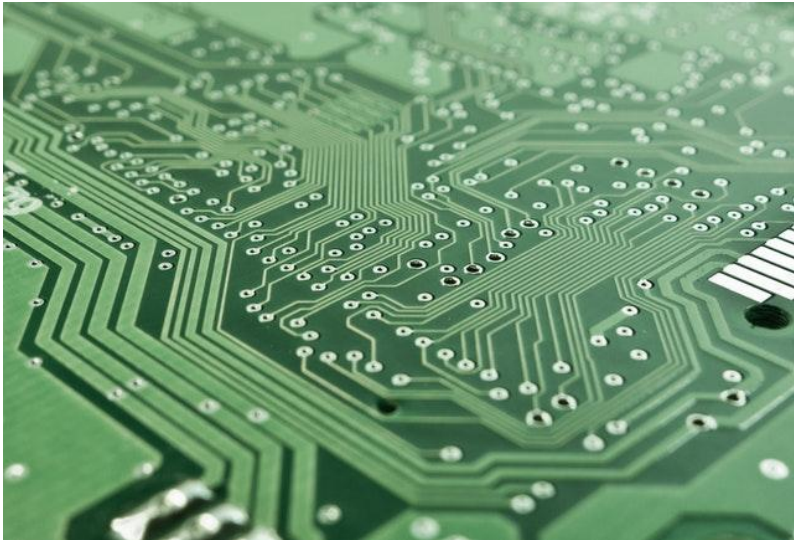
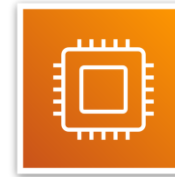


Photo from <https://www.pexels.com/photo/technology-computer-lines-board-50711/>



AWS Compute services



Amazon EC2



Amazon EC2
Auto Scaling



Amazon Elastic
Container Service
(Amazon ECS)



Amazon EC2
Container Registry



AWS Elastic
Beanstalk



AWS Lambda



Amazon Elastic
Kubernetes Service
(Amazon EKS)



AWS Fargate

Database service category



Photo from <https://aws.amazon.com/compliance/data-center/data-centers/>



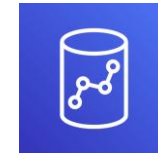
AWS Database services



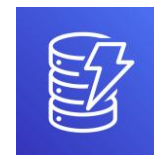
Amazon Relational
Database Service



Amazon Aurora

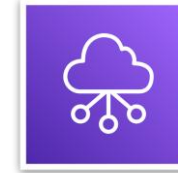


Amazon Redshift

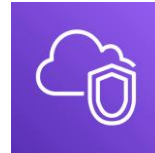


Amazon
DynamoDB

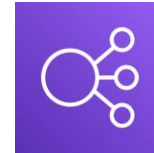
Networking and content delivery service category



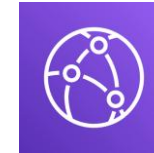
AWS networking and content delivery services



Amazon VPC



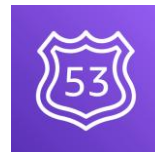
Elastic Load
Balancing



Amazon
CloudFront



AWS Transit
Gateway



Amazon
Route 53



AWS Direct
Connect



AWS VPN



Photo by Umberto on Unsplash

Security, identity, and compliance service category



Photo by Paweł Czerwiński on Unsplash



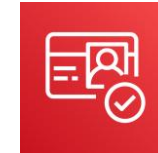
AWS security, identity, and compliance services



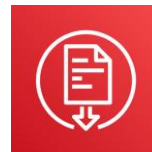
AWS Identity and Access Management (IAM)



AWS Organizations



Amazon Cognito



AWS Artifact



AWS Key Management Service



AWS Shield

AWS cost management service category



Photo by Alexander Mills on Unsplash



AWS cost management services



AWS Cost and
Usage Report



AWS Budgets



AWS Cost
Explorer

Management and governance service category

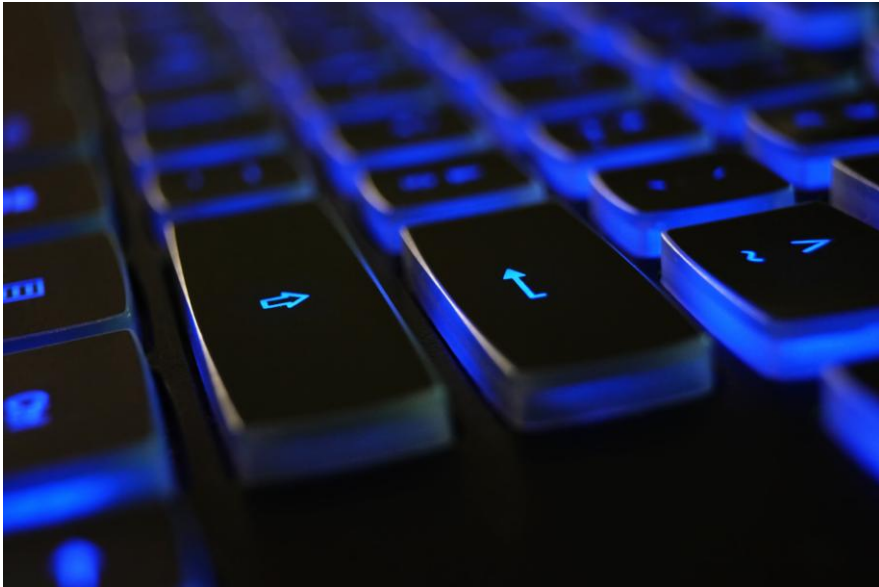


Photo by Marta Branco from Pexels



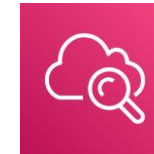
AWS management and governance services



AWS Management Console



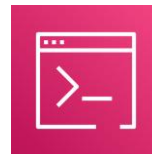
AWS Config



Amazon CloudWatch



AWS Auto Scaling



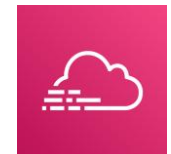
AWS Command Line Interface



AWS Trusted Advisor



AWS Well-Architected Tool



AWS CloudTrail

Activity: AWS Management Console clickthrough



Photo by Pixabay from Pexels.

Hands-on activity: AWS Management Console clickthrough



1. Launch the [Sandbox](#) hands-on environment and connect to the [AWS Management Console](#).
2. Explore the AWS Management Console.
 - A. Click the **Services** menu.
 - B. Notice how services are grouped into service categories. For example, the **EC2** service appears in the **Compute** service category.

Question #1: Under which service category does the **IAM** service appear?

Question #2: Under which service category does the **Amazon VPC** service appear?
 - C. Click the **Amazon VPC** service. Notice that the dropdown menu in the top-right corner displays an AWS Region (for example, it might display *N. Virginia*).
 - D. Click the Region menu and switch to a different Region. For example, choose **EU (London)**.
 - E. Click **Subnets** (on the left side of the screen). The Region has three subnets in it. Click the box next to one of the subnets. Notice that the bottom half of the screen now displays details about this subnet.

Question #3: Does the subnet you selected exist at the level of the Region or at the level of the Availability Zone?
 - F. Click **Your VPCs**. An existing VPC is already selected.

Question #4: Does the VPC exist at the level of the Region or the level of the Availability Zone?

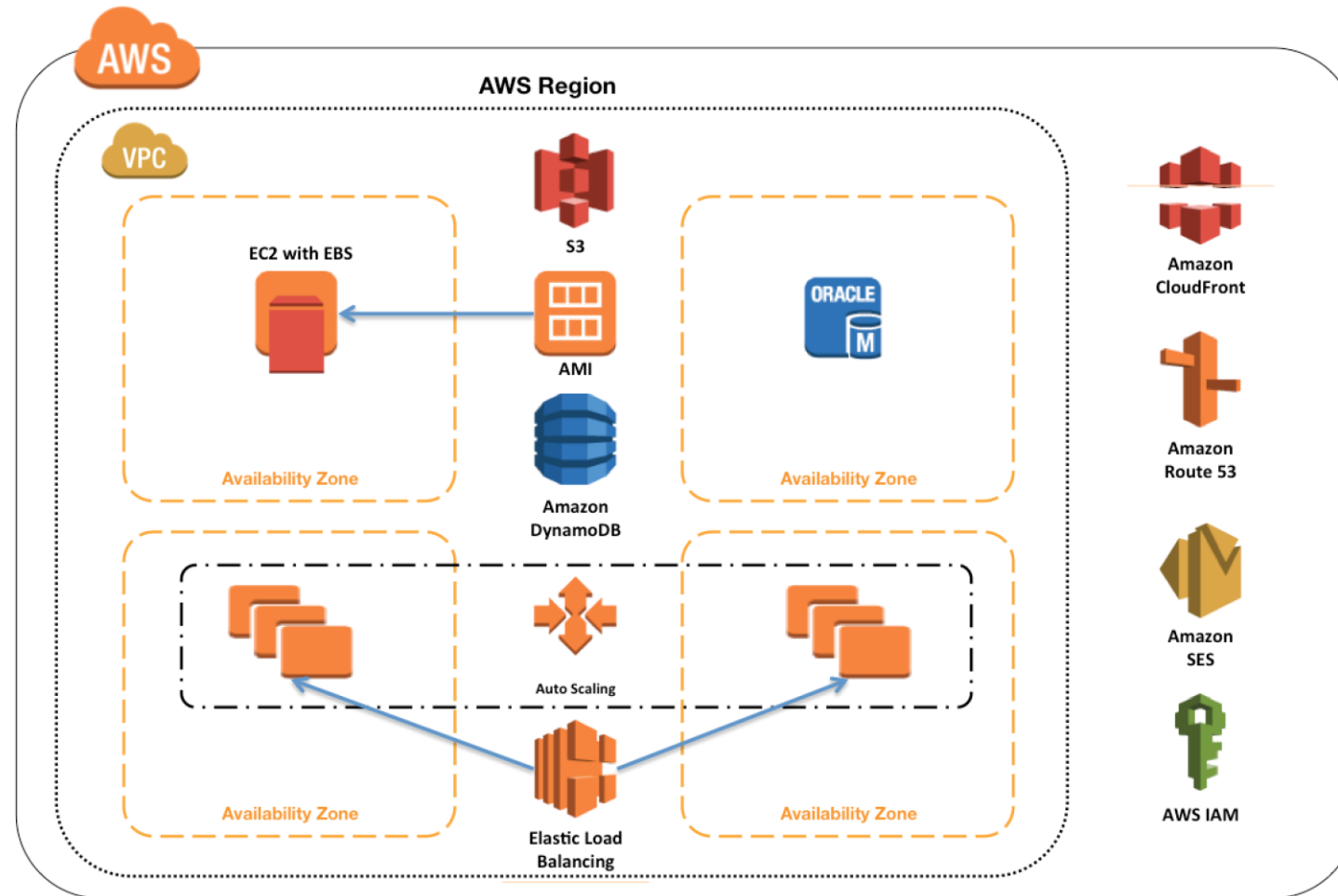
Question #5: Which services are global instead of Regional? Check Amazon EC2, IAM, Lambda, and Route 53.

Activity answer key

- **Question #1:** Under which service category does the **IAM** service appear?
 - **Answer:** **Security, Identity, & Compliance.**
- **Question #2:** Under which service category does the **Amazon VPC** service appear?
 - **Answer:** **Networking & Content Delivery**
- **Question #3:** Does the subnet that you selected exist at the level of the Region or the level of the Availability Zone?
 - **Answer:** Subnets exist at the **level of the Availability Zone.**
- **Question #4:** Does the VPC exist at the level of the Region or the level of the Availability Zone?
 - **Answer:** VPCs exist at the **Region level.**
- **Question #5:** Which of the following services are global instead of Regional? Check Amazon EC2, IAM, Lambda, and Route 53.
 - **Answer:** **IAM and Route 53 are global.** Amazon EC2 and Lambda are Regional.

Example : Regional vs Availability Zone

AWS Regional vs Availability Zone Services



Module 3: AWS Global Infrastructure Overview

Module wrap-up

In summary, in this module you learned how to:

- Identify the difference between AWS Regions, Availability Zones, and edge locations
- Identify AWS service and service categories

Complete the knowledge check



Sample exam question

Which component of AWS global infrastructure does Amazon CloudFront use to ensure low-latency delivery?

- A. AWS Regions
- B. AWS edge locations
- C. AWS Availability Zones
- D. Amazon Virtual Private Cloud (Amazon VPC)

- [AWS Global Infrastructure](#)
- [AWS Global Infrastructure Region Table](#)
- [AWS Cloud Products](#)

Thank you

© 2019 Amazon Web Services, Inc. or its affiliates. All rights reserved. This work may not be reproduced or redistributed, in whole or in part, without prior written permission from Amazon Web Services, Inc. Commercial copying, lending, or selling is prohibited. Corrections or feedback on the course, please email us at: aws-course-feedback@amazon.com. For all other questions, contact us at: <https://aws.amazon.com/contact-us/aws-training/>. All trademarks are the property of their owners.

