Getting started with AWS

AWS Cloud History

2002

AWS was launched internally within Amazon.

2003

Amazon recognized that its internal infrastructure was a core strength and began exploring the idea of offering it as a marketable product.

2004

AWS was publicly launched with its first service: Amazon SQS (Simple Queue Service).

2006

AWS was re-launched publicly with additional core services: SQS, S3 (Simple Storage Service), and EC2 (Elastic Compute Cloud).

2007

AWS expanded globally and was launched in Europe.

Early Adopters and Major Customers

Several major organizations started using AWS, including:

- **Netflix**
- **NASA**
- **Dropbox**
- **Airbnb**

AWS Cloud Number Facts

- In 2023, AWS generated **$90 billion** in annual revenue.
- As of Q1 2024, **AWS holds 31%** of the cloud market share.
    - Microsoft is second with **25%**.
- AWS has been the **pioneer and leader** of the cloud market for **13 consecutive years**.
- AWS serves **over 1,000,000 active users**.

Gartner Magic Quadrant (as of October 2023)

- AWS is positioned as a **Leader** in the Gartner Magic Quadrant for Strategic Cloud Platform Services.
- AWS leads in both **completeness of vision** and **ability to execute**.
- Other major players in the "Leaders" quadrant include:
    - **Microsoft**
    - **Google**
- Competitors in other quadrants:
    - "Challengers": Oracle
    - "Visionaries": IBM, Alibaba Cloud

    - "Niche Players": Huawei Cloud, Tencent Cloud

AWS Cloud Use Cases

- AWS enables the development of **sophisticated and scalable applications**.
- It is **applicable across a wide range of industries**.

Common Use Cases

- **Enterprise IT**, **Backup & Storage**, **Big Data Analytics**
- **Website Hosting**, **Mobile & Social Applications**
- **Gaming**

Example Companies Using AWS

- **McDonald's**
- **21st Century Fox**
- **Activision**
- **Netflix**

AWS Global Infrastructure

AWS has a robust and expansive global infrastructure that includes:

- **AWS Regions**: Geographically isolated areas that contain multiple Availability Zones.
- **AWS Availability Zones**: Multiple, isolated locations within each Region, designed for fault tolerance and high availability.
- **AWS Data Centers**: Physical infrastructure that powers Regions and Availability Zones.
- **AWS Edge Locations / Points of Presence**: Locations used to cache content closer to users to reduce latency, commonly used with services like CloudFront.

Global Presence

The map shows key AWS infrastructure locations across North America and Europe, including:

- **United States**: Oregon, N. California, Ohio, N. Virginia, GovCloud (US-West & US-East)
- **Canada**: Central
- **Europe**: Ireland, London, Paris, Frankfurt, Milan, Spain, Stockholm

For an interactive map and up-to-date details, visit:
https://infrastructure.aws/

AWS Regions

- AWS has **Regions** distributed all around the world.
- Region names follow a format such as `us-east-1`, `eu-west-3`, etc.
- A **Region** is a **cluster of data centers** that are geographically isolated.
- Most AWS services are **region-scoped**, meaning they operate and are accessible within a specific region only.

How to Choose an AWS Region?

When launching a new application, choosing the right AWS Region depends on several key factors:

- **Compliance** with data governance and legal requirements:
  Data will never leave a Region without your explicit permission, ensuring regulatory compliance.

- **Proximity** to customers:
  Selecting a region closer to your users helps reduce latency and improve performance.
- **Available services** within a Region:
  Not all AWS services and features are available in every Region. Some services may launch in specific Regions first.
- **Pricing**:
  Costs can vary between Regions. Pricing is transparent and can be reviewed on the AWS service pricing page.

AWS Availability Zones

- Each AWS Region contains multiple **Availability Zones (AZs)**.
  - The typical number is **3**, with a **minimum of 3** and **maximum of 6**.
  - Example AZs:
    - `ap-southeast-2a`
    - `ap-southeast-2b`
    - `ap-southeast-2c`
- An **Availability Zone** is one or more **discrete data centers** with **redundant power, networking, and connectivity**.
- AZs are **physically separated** within a region to ensure **isolation from failures and disasters**.
- Despite being isolated, they are **interconnected with high-bandwidth, ultra-low-latency networking**, allowing for seamless failover and replication.

AWS Points of Presence (Edge Locations)

- AWS operates **400+ Points of Presence**, which include:
  - **400+ Edge Locations**
  - **10+ Regional Edge Caches**
- These are distributed across **90+ cities** in **40+ countries**.

Purpose and Benefit

- Points of Presence are used primarily by **Amazon CloudFront** and other edge services.
- They enable **low-latency content delivery** to end users by caching content closer to their physical location.

Additional Resource

Learn more at: https://aws.amazon.com/cloudfront/features/

Tour of the AWS Console

AWS has Global Services

These services are **not tied to a specific region** and are managed globally:

- **Identity and Access Management (IAM)**
- **Route 53** (DNS service)
- **CloudFront** (Content Delivery Network)

- **WAF** (Web Application Firewall)

Most AWS Services are Region-Scoped

These services operate within a specific region:

- **Amazon EC2** (Infrastructure as a Service)
- **Elastic Beanstalk** (Platform as a Service)
- **Lambda** (Function as a Service)
- **Rekognition** (Software as a Service)

Region Table

To see which services are available in each region, visit:
https://aws.amazon.com/about-aws/global-infrastructure/regional-product-services