Module 02: AWS Overview

What is AWS (Amazon Web Services)?

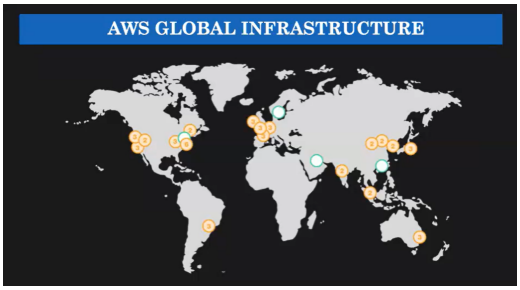
AWS is one of the cloud service providers. To deliver your application or service, you need infrastructure. This infra can be obtained from cloud service providers.

> 💡 Amazon Web Services (AWS) is a secure cloud services platform, offering compute power, database storage, content delivery and other functionality to help businesses scale and grow.

Cloud service providers: AWS, Azure, GCP, Alibaba Cloud, TenCent Cloud, IBM Cloud, Rackspace, etc.,

AWS Global Infrastructure

Components of global infrastructures - Region, Availability Zones, and Edge Locations.



Cloud means we are using AWS data centers.

AWS has data centers across the globe.

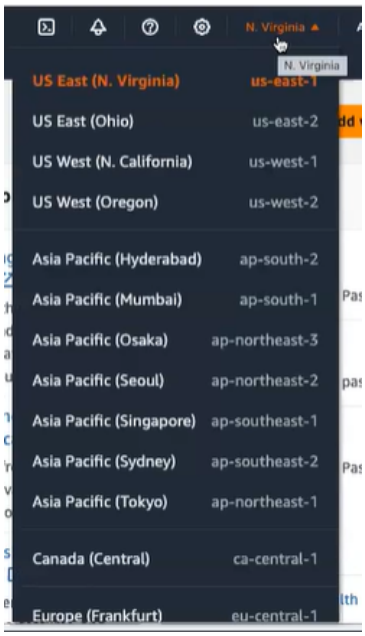
Global Infrastructure - AWS

The most secure, extensive, and reliable Global Cloud Infrastructure, for all your

https://aws.amazon.com/about-aws/global-infrastructure/

1. **Region:** AWS segregated entire world into different regions. There is no specific boundaries. Its just a physical location or geographic location.

   You can sit anywhere and launch a server in any region.



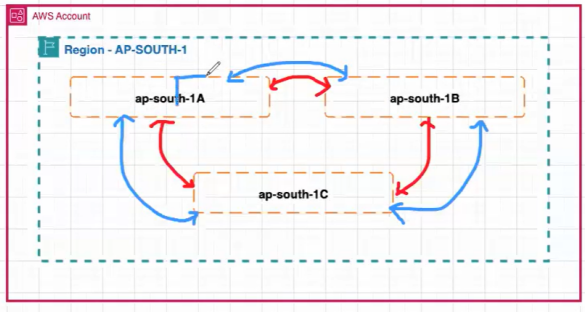
   34 regions are launched as of Sept 2023.

   Definition

   A region is **a large geographical area that contains multiple data centers, or Availability Zones (AZs), that are physically separated and isolated.**

2. **Data Center (or) Availability Zone:** AZ is nothing but a data center or combination of multiple data centers. Our actual workload (servers) runs in this AZs.

   Every region contains at least 3 AZs for high availability or fault tolerance purpose. Within in a region, all the running AZs will be connected with every other AZ with multiple connectivity lines to avoid downtime.



> 💡 As of sept 5, 2024 , there are 108 AZs

> 💡 **Definition**
>
> An Availability Zone (AZ) is **a collection of data centers within an AWS Region that are isolated from other AZs and have redundant power, networking, and connectivity**

3. **Edge location:** It is a CDN endpoint. CDN stands for Content Delivery Network. Cache locations, it stores data temporarily.

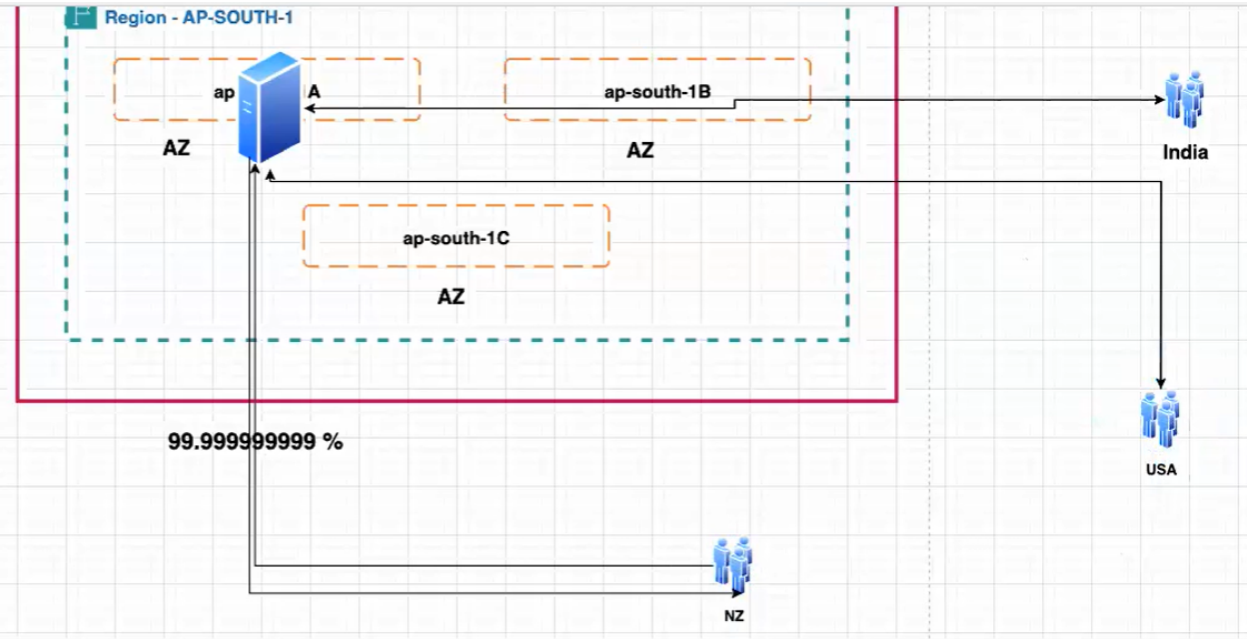
   Caching: Caching saves subsets of data, making it available so that when someone requests the data, it can be delivered faster from the nearest edge location.

Benefits: Edge locations help to reduce the distance that data has to travel to reach the user, which can lead to faster responses.

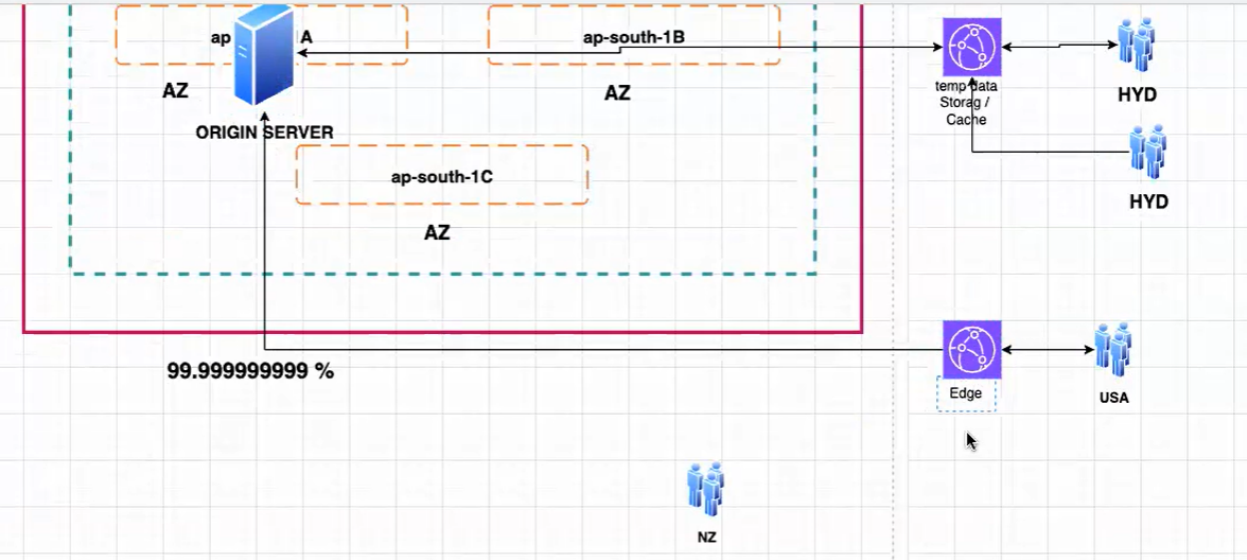
Function: Edge locations work with services like Amazon CloudFront, AWS's content delivery network (CDN), to cache content and deliver it to users faster.

> 💡 **Definition**
>
> AWS Edge Locations are **strategically positioned points in the AWS network optimized for low-latency content delivery**, ensuring that data reaches users swiftly.

Illustration: When distance increases between the server and the user, there will be higher latency. A server placed in `ap-south-1a` region which is in Mumbai location, will provide faster response to users from India compared to users from USA or New Zealand.



To reduce this latency, we can depend on a component called *Edge Location*. Edge Location works with a mechanism called CDN. AWS CloudFront provides CDN mechanism. Data is cached temporarily in edge locations. TTL (time-to-live) is by default 86_400 seconds ~ 24hrs. If there is no data in Edge Location (cache), then request goes to original server.



Data Invalidation: When data is updated on the ORIGINAL SERVER but outdated information remains in the cache or temporary storage (due to the time-to-live (TTL) setting; data updates automatically after TTL expires), it becomes necessary to perform data invalidation to ensure consistency for immediate updating.

AWS Account Creation & Free tier limitations overview

Step 1: Log in to AWS Free Tier

1. Click **Create an AWS Account** or **Sign In**.
2. Enter your **Root User Email Address**: chaithra.kc22@email.com.
3. Set the **Account Name**: At the enterprise level, multiple accounts (sandbox, dev, SQA/testing, UAT, production) are used for different environments. The account name should clearly indicate the environment (e.g., prod-account or production account etc.,).

Step 2: Provide Personal Information

1. Choose **Account Type**: **Personal**, **Professional**, or Organization (select **Personal** for this tutorial).
2. Enter your **Name** and **Address**.

Step 3: Enter Payment Information

AWS accepts Visa, Mastercard, Amex, Discover, and Rupay. A temporary $1 USD charge will be applied for verification, refunded within 48-72 hours.

Step 4: Identity Verification

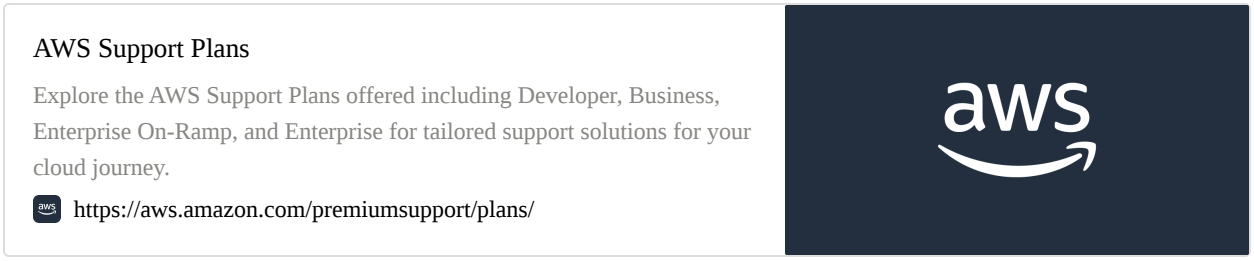
You can verify via phone or email:

- **Phone Verification**: Enter your phone number and select "Call me now" to receive a one-time password (OTP).
- **Email Verification**: Enter your email address, and retrieve the OTP from your inbox.

Step 5: Choose a Support Plan

- **Account and Billing Issues**: Free support across all plans, available 24/7 via phone, email, and chat.
- **Technical Issues**: Support varies depending on the plan selected.

Individual AWS Support Plans:



1. **Basic Support Plan (Free):**
    - No technical support from AWS.
    - Access to AWS knowledge base articles via AWS re:post for assistance.
    - Includes **Trusted Advisor** for core checks. AWS Trusted advisor will give you advices to secure your account, to reduce the cost, to optimize the performance. It will only give advices; not fix anything.
    - **Recommendation**: Choose this when creating your personal AWS account.
2. **Developer Support Plan (Starting at $29/month):**
    - Response time within 12-24 local business hours.

- Support via email from an AWS Associate.
- Case severity determines response time:
  - General guidance: < 24 hours
  - System impaired: < 12 hours
- 1 user can raise unlimited tickets, even if managing multiple accounts

3. **Business Support Plan (Starting at $100/month):**
   - 24/7 assistance via email, phone, and chat from an AWS Engineer.
   - Faster response times based on case severity:
     - General guidance: < 24 hours
     - System impaired: < 12 hours
     - Production system impaired: < 4 hours
     - Production system down: < 1 hour
   - Any user can raise unlimited tickets.
   - Full **Trusted Advisor** checks.

- **Note**: For multiple AWS accounts (e.g., 100 accounts), this can get costly ($100 x 100 = $10,000/month), so consider the **Enterprise On-Ramp** plan.

Advanced Support Plans (for Multiple Accounts):

1. **Enterprise On-Ramp Support Plan (Starting at $5,500/month):**
   - Response time within 30 minutes from an AWS Senior Engineer.
   - Unlimited tickets across unlimited accounts.
   - Annual architectural and operational reviews.
   - Full **Trusted Advisor** checks.
   - Case severity response times:
     - General guidance: < 24 hours

    - System impaired: < 12 hours
    - Production system impaired: < 4 hours
    - Production system down: < 1 hour
    - Business/mission-critical system down: < 30 minutes

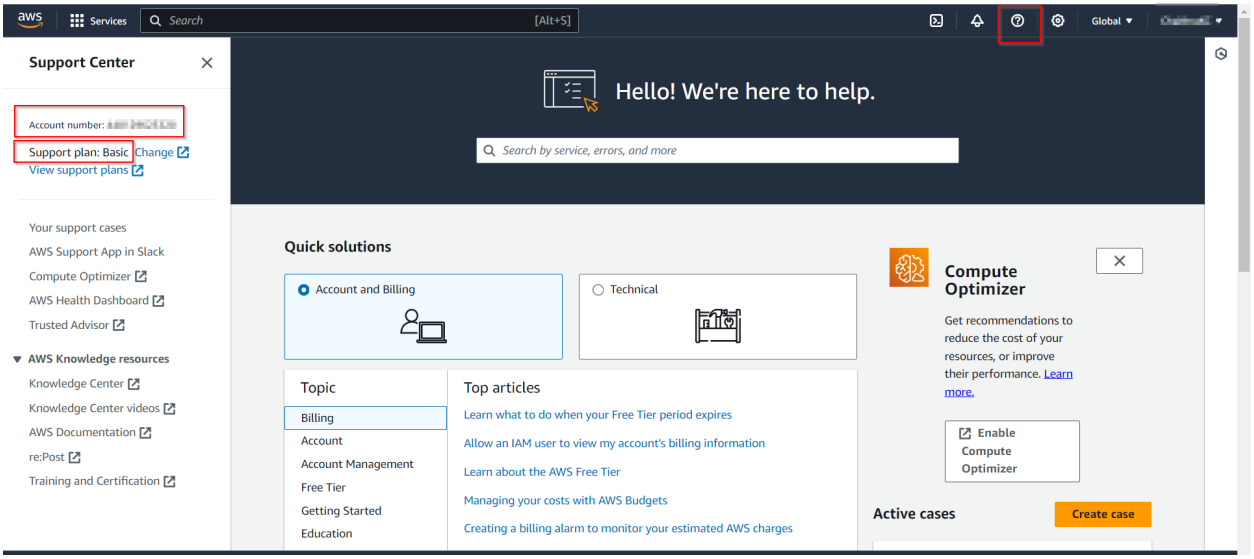
2. **Enterprise Support Plan (Starting at $15,000/month):**
    - Response time within 15 minutes from an AWS Senior Engineer.
    - Unlimited tickets across any account.
    - Dedicated Technical Account Manager (TAM) provided by AWS.
    - Annual architectural and operational reviews, with additional training on AWS services.
    - Full **Trusted Advisor** checks.
    - Case severity response times:
        - General guidance: < 24 hours
        - System impaired: < 12 hours
        - Production system impaired: < 4 hours
        - Production system down: < 1 hour
        - Business/mission-critical system down: < 15 minutes

The **Enterprise On-Ramp** and **Enterprise Support** plans can cover multiple AWS accounts. There's no need to purchase separate support plans for every account; one plan is sufficient for 100+ accounts.

How to View Your Support Plan:

1. After logging in, search for **"Support"** in the AWS search bar or click the **question mark icon** in the top navigation bar.

2. Every AWS account you create is assigned a unique 12-digit account number.



AWS Free Tier Limitations

For the first 12 months after creating a new AWS account, certain services are available for free under the AWS Free Tier. However, not all services are entirely free, and limitations apply.

To explore the free tier services:

- Visit: AWS Free Tier
- Click on "12 months free services" for a detailed list: Free Tier Services.

Example: EC2 Instance

- The free tier includes **750 hours per month** of usage for **Linux, RHEL, or SLES t2.micro** or **t3.micro** instances (region-dependent).

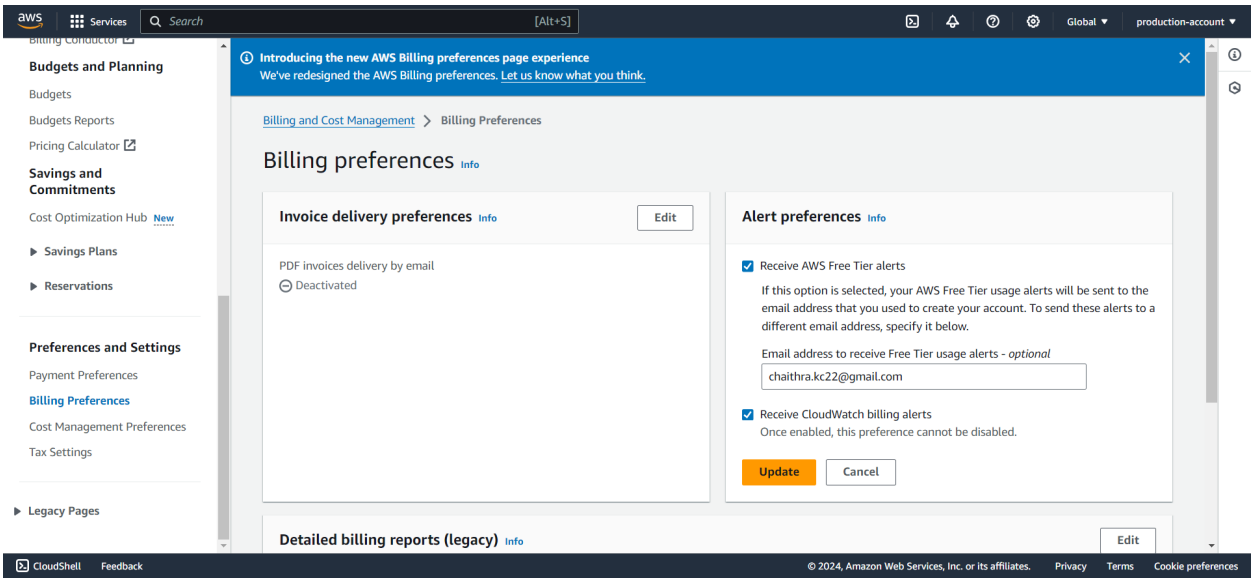
Usage Scenarios:

- **1 server** running 24/7 for 31 days equals **744 hours**, which is within the free tier limit. You can use 1 server for entire month for free.
- **2 servers** running for 16 days will consume **768 hours** (384 hours per server), exceeding the free limit. AWS will charge for the additional 18 hours and rest of the 14 days in a month.

Enabling Free Tier Alerts

To avoid unexpected charges, set up free tier usage alerts:

1. Log into your AWS account.
2. Navigate to **Billing and Cost Management**.
3. Go to **Billing Preferences**.
4. Enable **Alert Preferences** and enter your email to receive notifications.



Interview Question

What is the support plan your organization uses? Ans: Enterprise Support Plan