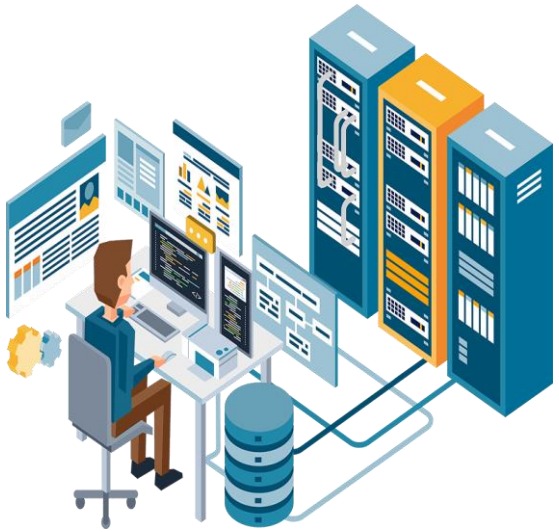
Welcome
to
Cloud Computing &
AWS Overview

Presenter - Babu

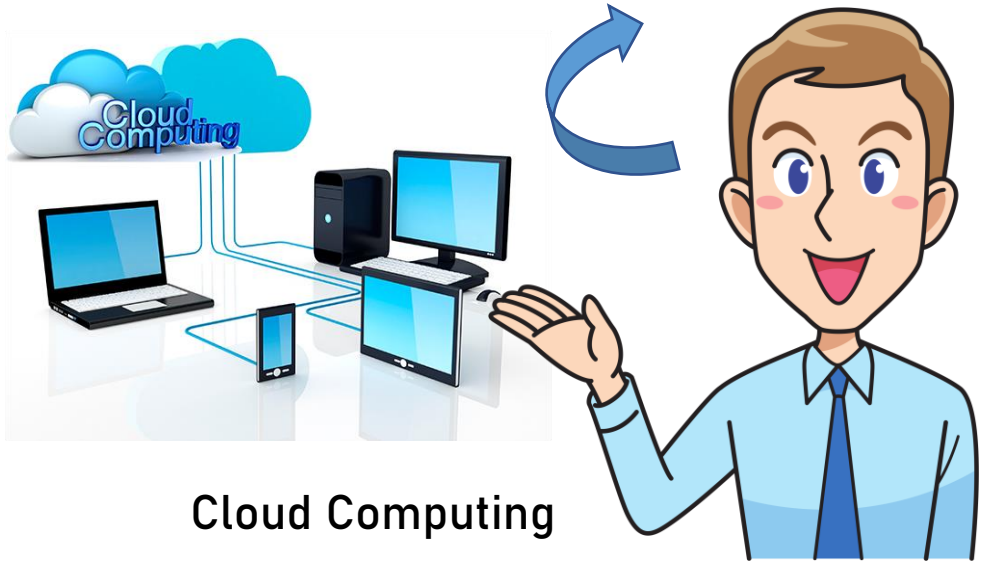
Why Cloud Computing ?

Hi, I'm about to start a company.
Can you list down the resources I will need to setup on-premise infrastructure?



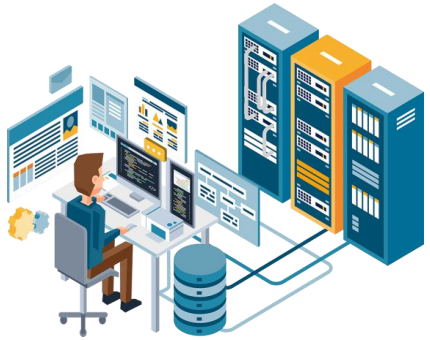
On-Premise

Why not you setup things on a cloud?



Cloud Computing

On-Premise Vs Cloud Computing ?



On-Premise

- Higher pay, less scalability
- Allow huge space for servers
- Less chance of data recovery
- Lack of flexibility
- Less Collaboration
- Longer Implementation time

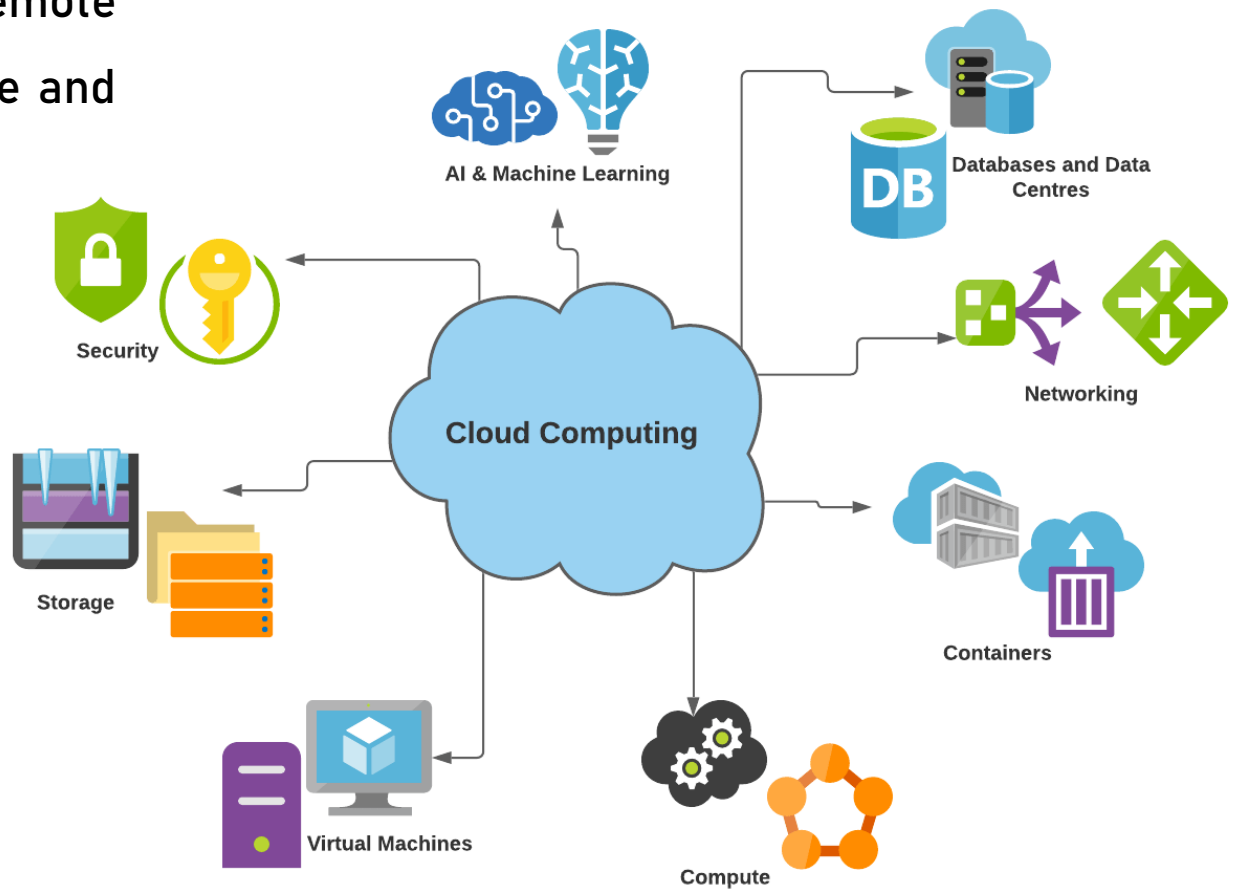
VS

Cloud Computing

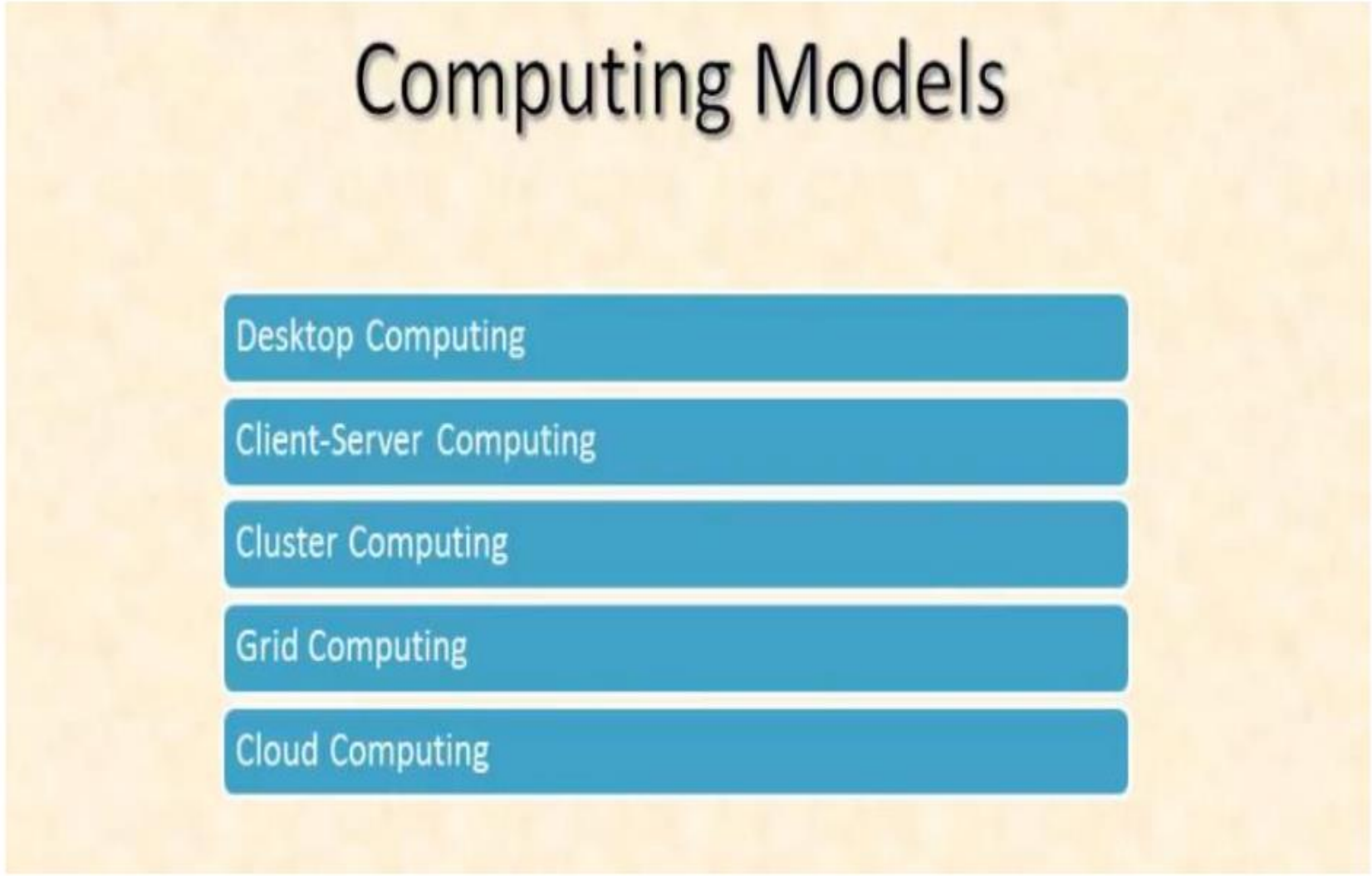
- Pay for what you use
  - Scale up = Pay more
  - Scale down = Pay less
- No server space required
- Disaster recovery
- High flexibility
- Collaboration from widespread location
- Rapid Implementation

What is Cloud Computing ?

Cloud computing is the use of a network of remote servers hosted on the internet to store, manage and process data rather than a local server



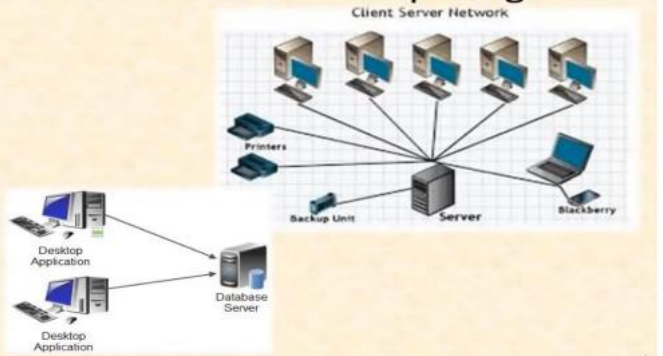
Computing Models



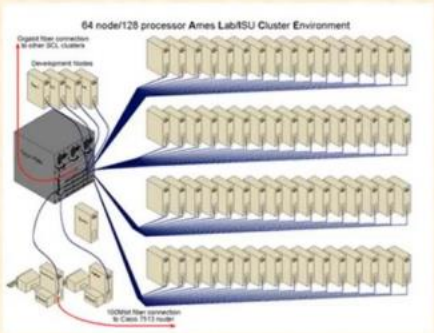
Computing Models

Desktop Computing

Client-Server Computing



Cluster Computing



Desktop Computing

1. Personal
2. Professional
   - Engineers
   - Artist
   - Authors
   - Doctors
   - Programmers
3. Office
4. Desktop Publishing

Software

1. MS-Office
2. AutoCAD
3. Photoshop
4. Illustrator
5. 3DStudio Max
6. Dreamweaver
7. Net Beans
8. Visual Studio
9. CMS

Client-Server Computing

- Banks
- Retail Stores
- Marketing & Sales
- Distribution
- Aviation Companies
- Automobile Companies
- Oil Companies

Software

1. Accounting/Finance
2. Sales ERP
3. CRM
4. Distribution ERP
5. Manufacturing ERP

- SAP
- Oracle Apps
- Microsoft Dynamics

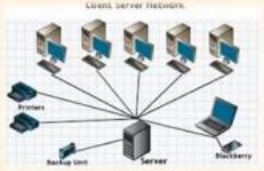
Cluster Computing

Computing Models

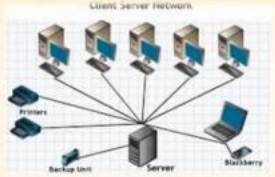
Grid Computing

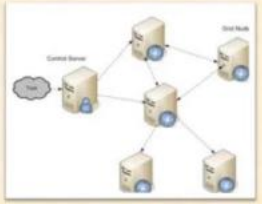
- **Different Locations**

Delhi Branch



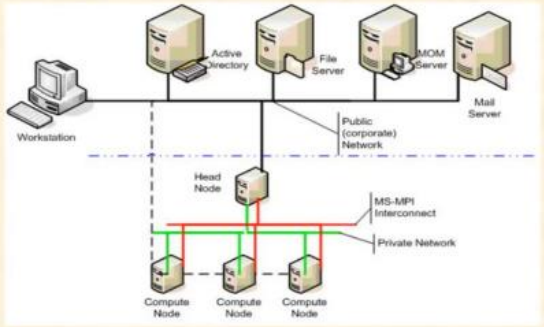
Bangaluru Branch



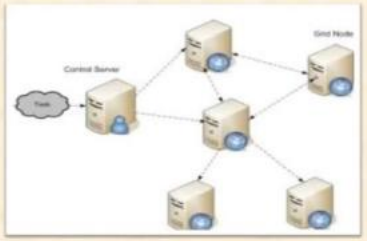


Grid Computing

- **Different Services**



Grid Computing



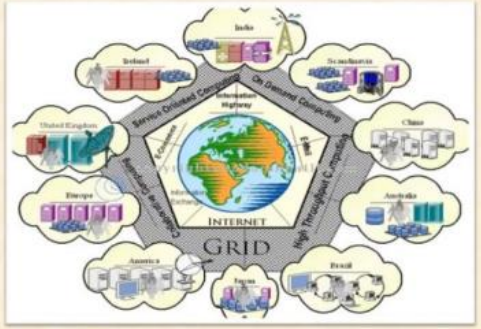
Heterogeneous Servers

- Different Operating Systems
- Different Application Servers

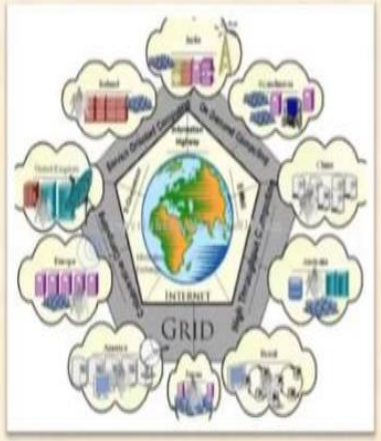
1. Database Server
2. Mail Server
3. Web Server
4. File Server

Cloud Computing

- Combination of Cluster and Grid Computing



Cloud Computing



Companies who Own Cloud

- Toyota
- Wal-Mart
- TATA Group
- Citi Bank

Infrastructure Maintenance

- IBM
- Dell
- Intel
- HP
- Cisco
- Sun

Software Companies

- SAP
- Oracle Apps
- Microsoft Dynamics

Objectives Cloud Computing ?

Elasticity

Ability to scale virtual machines resources up or down

On-demand usage

Ability to add or delete computing power (CPU, memory), and storage according to demand

Pay-per-use

Pay only for what you use

Multitenancy

Ability to have multiple customers access their servers in the data center in an isolated manner

Benefits of Cloud Computing ?

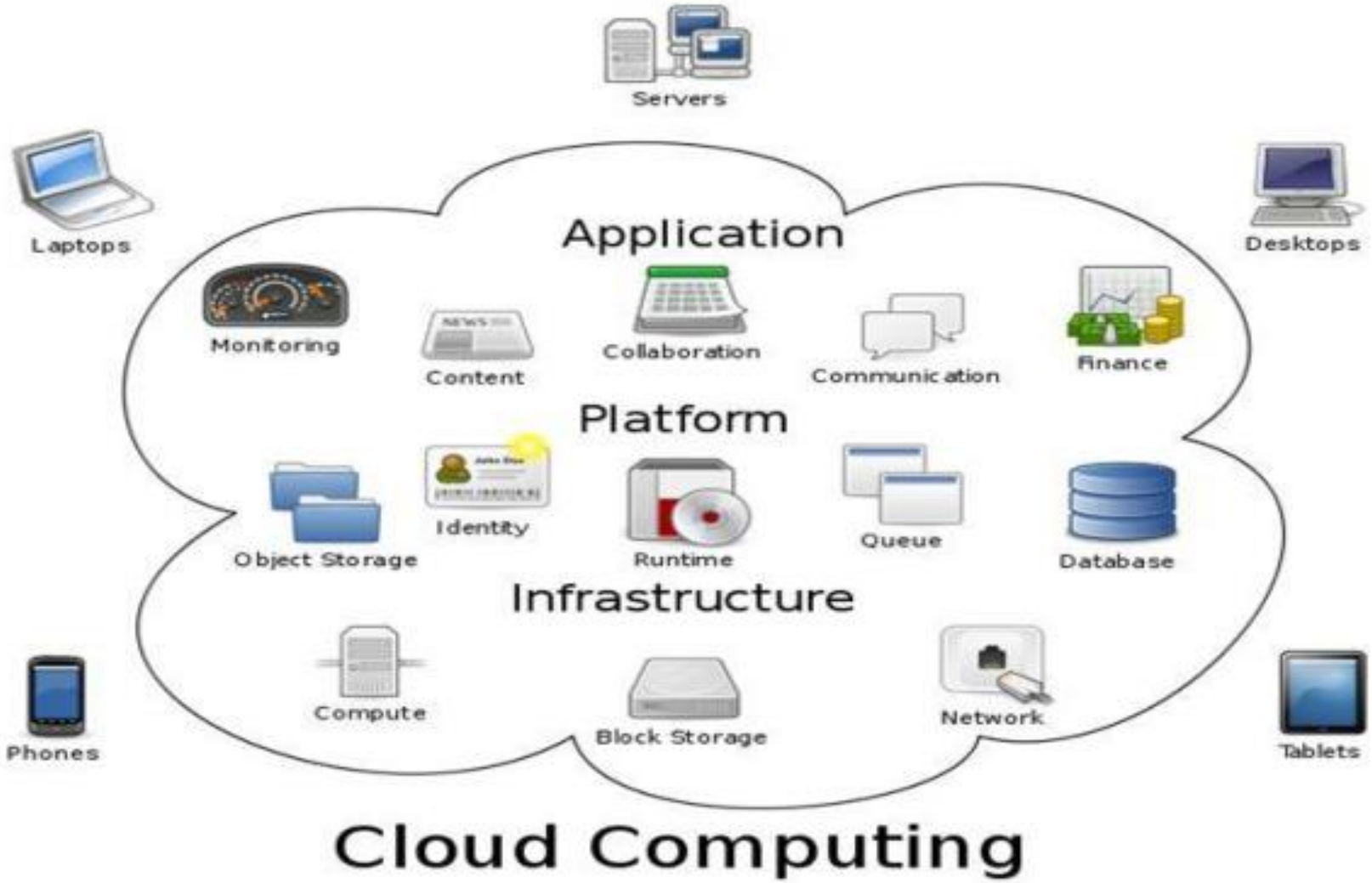
- COST – It eliminates the expense of buying computer hardware & software.
- SPEED – Vast amount of computing resources can be provisioned in minutes.
- SCALABILITY – Easy to scale up your capacity.
- ACCESSIBILITY – Easy to access data anywhere.
- BETTER SECURITY – With cloud, your data is stored in a centralized secure location.

Service Providers of Cloud Computing ?

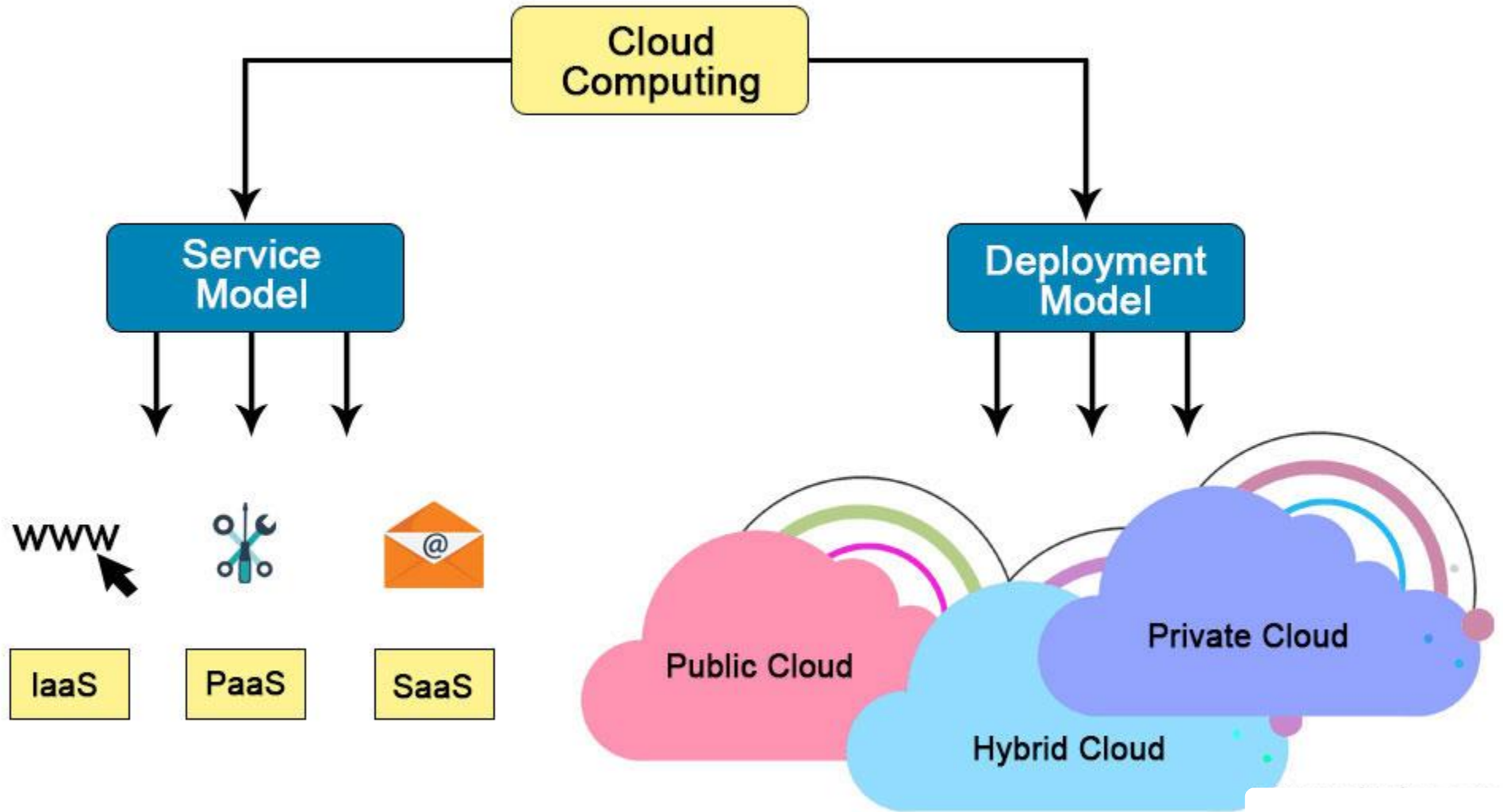
IBM **Cloud**

Note : Companies offering these computing services are called cloud providers

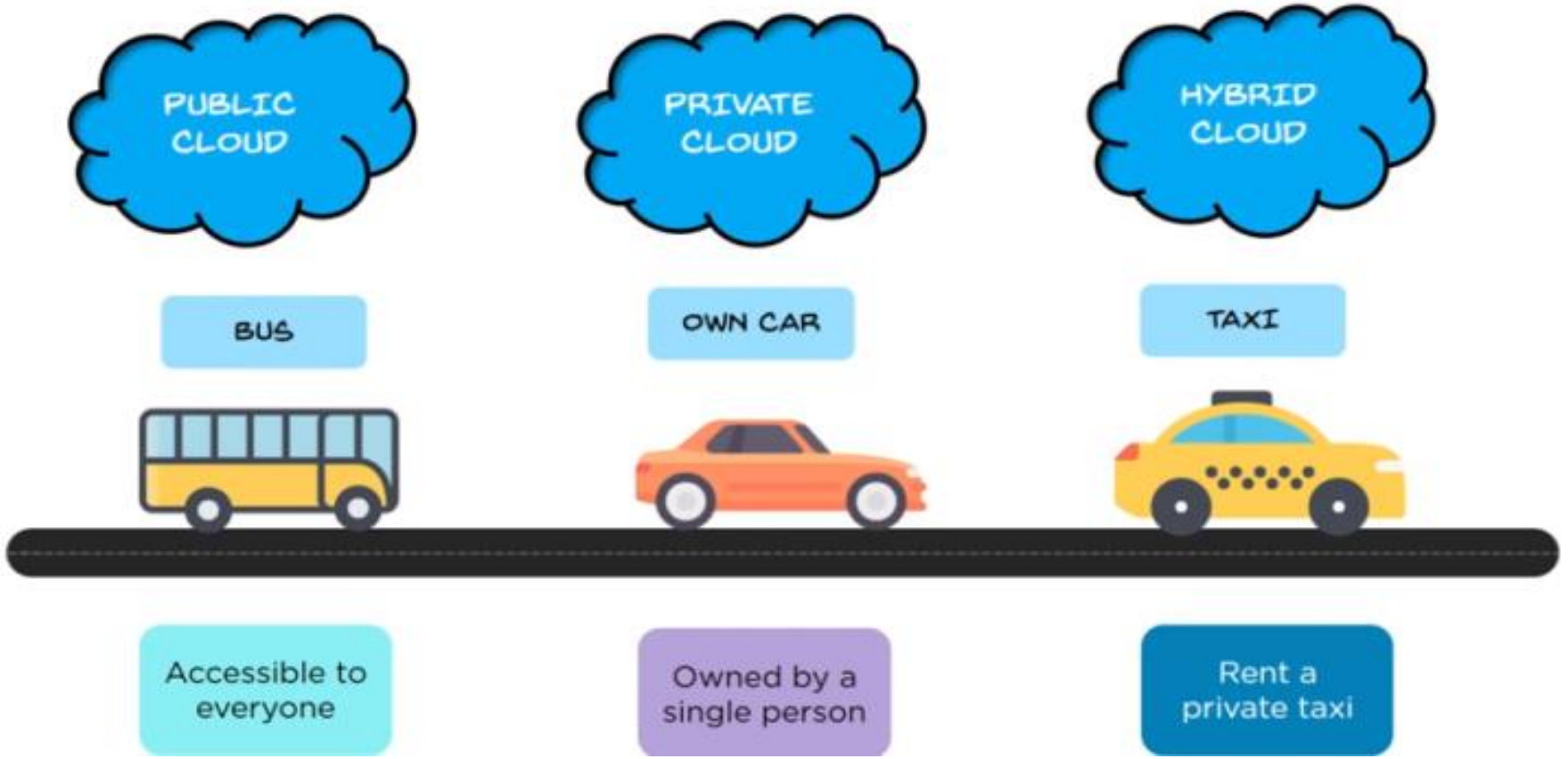
Cloud Architecture ?



Types of Cloud Computing ?



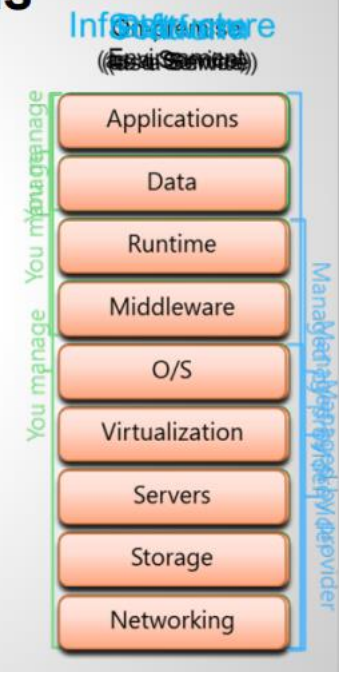
Types of Cloud Computing ?



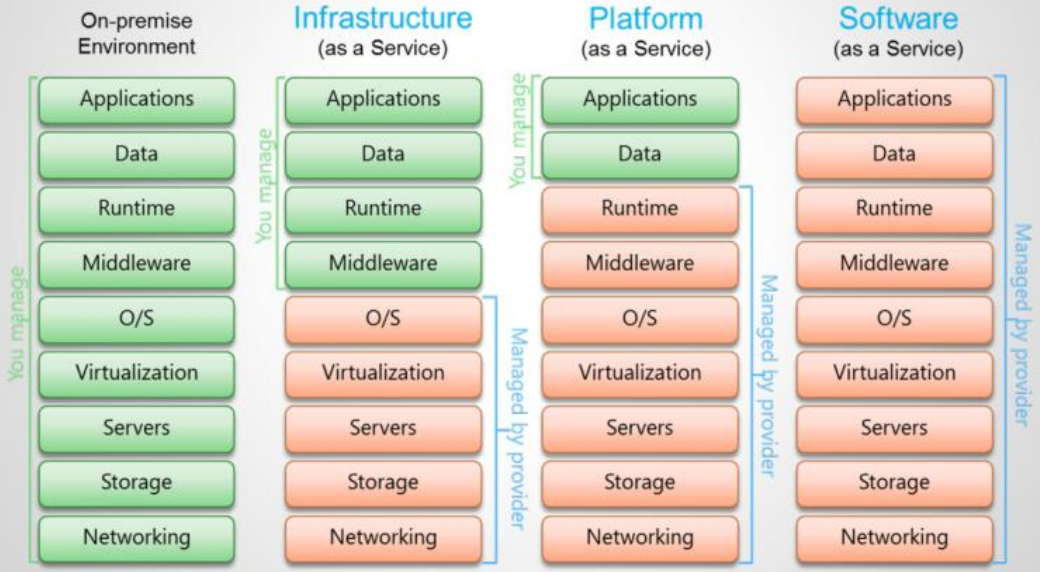
Service Model

Flavors and Service Models

- Infrastructure as a Service (IaaS)
  - Infrastructure components
  - Generally: $/component usage
  - e.g. Amazon Web Services, Rackspace
- Platform as a Service (PaaS)
  - Application backend
  - Generally: $/data
  - e.g. Google App Engine, MS Azure databases
- Software as a Service (Saas)
  - Application based
  - Generally: $/user
  - e.g. Google GMail, ServiceNow



Flavors and Service Models



Using the Cloud – Practical Examples of IaaS, PaaS and SaaS

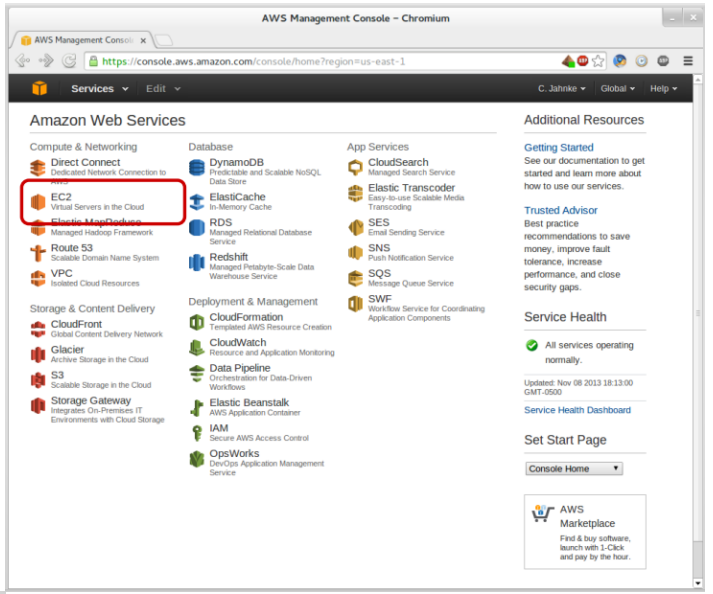
Infrastructure as a Service (IaaS)

Infrastructure components

- Compute, storage, network, etc.
- Pay for usage of each component
- Building blocks provide flexibility
- Often lower cost

Example:

amazon web services™



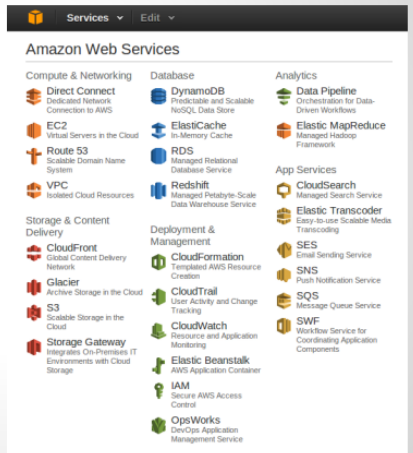
IaaS - Amazon Web Services

Amazon Provides

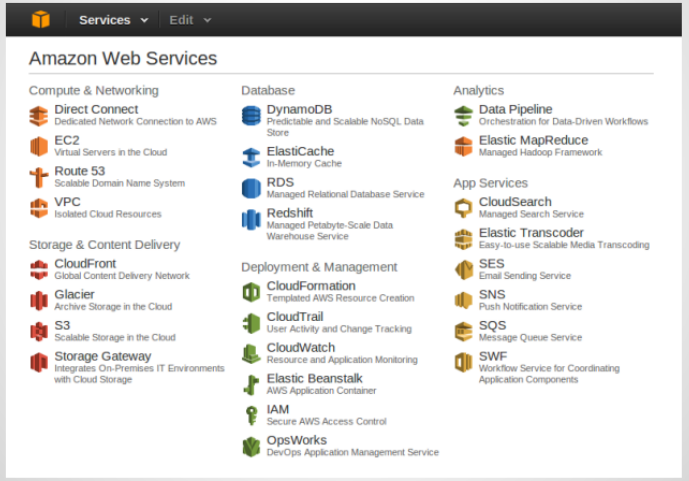
- Components

You Provide

- Component Config.
- Middleware
- Runtimes
- Data
- Applications
- **Time and Experience**



But that's just one component...



Using the Cloud – Practical Examples of IaaS, PaaS and SaaS

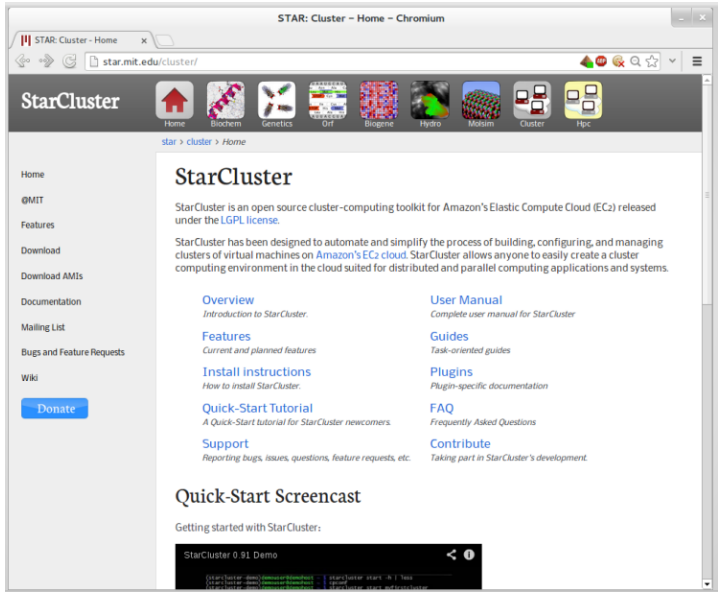
Platform as a Service (PaaS)

"Platform" or Application Backend

- Often a database or management toolset
- Pay for usage (time) of platform
- Consumer builds on top of platform
- Examples:
  - Google App Engine, Microsoft Azure Databases

Example:

StarCluster

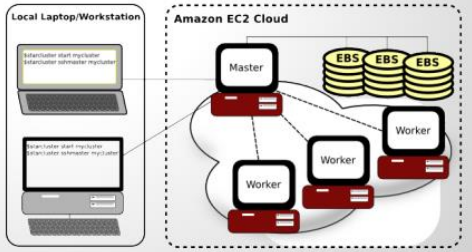


PaaS - MIT StarCluster

- StarCluster Provides
  - Builds on AWS "infrastructure" services
  - Cluster configuration
  - Middleware and runtime environment
  - Tools for management

Platform

- You provide
  - Management
  - Applications
  - Data



Using the Cloud – Practical Examples of IaaS, PaaS and SaaS

Software as a Service (SaaS)

Application or Software Suite

- Graphical Web Interface
- Often pay for each user
- Common Examples:
  - Google GMail
  - ServiceNow

Example:

DNAnexus

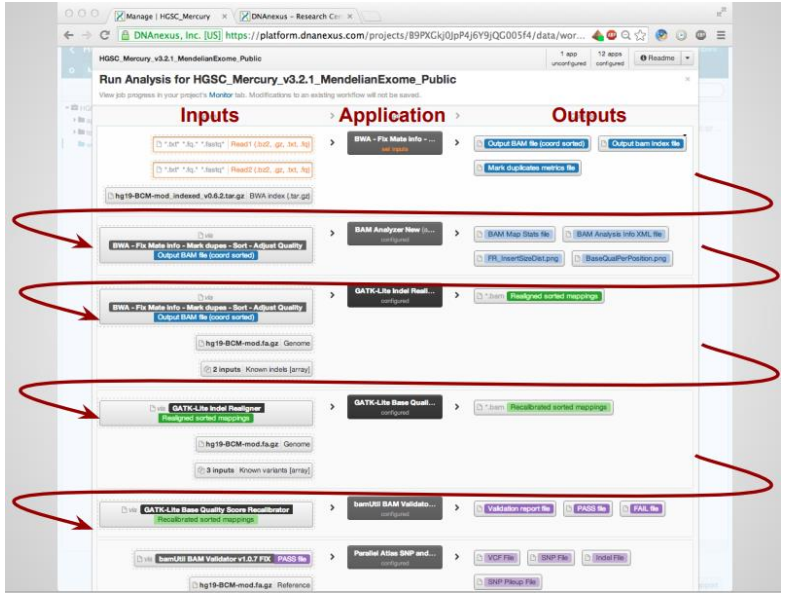
SaaS - DNA Nexus Graphic Interface

- GUI Available
- Drag and Drop
- Data Mgmt.
- Application Mgmt.
- Workspace Sharing

(This will get larger)

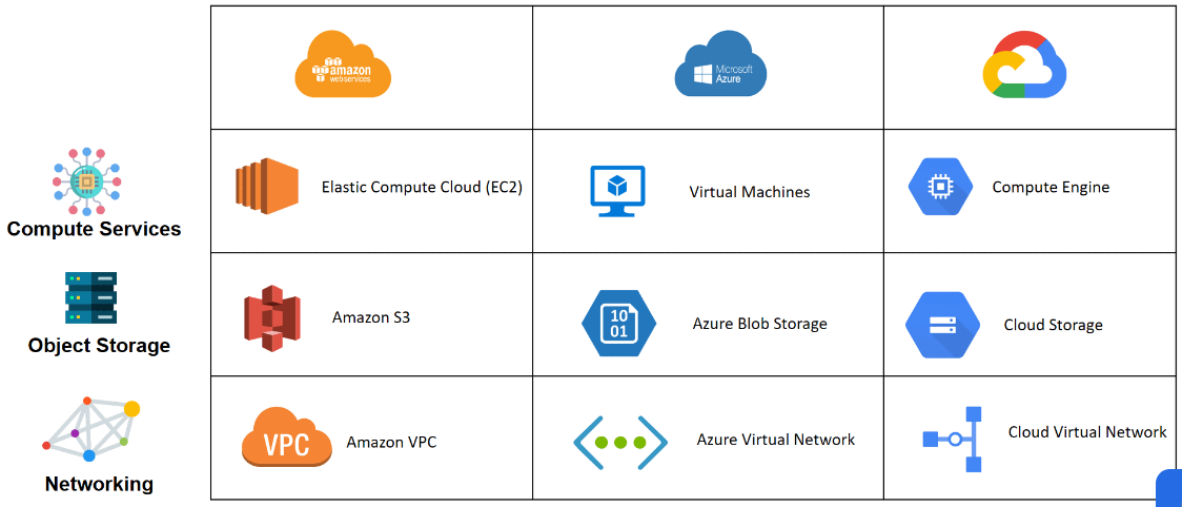
SaaS - DNA Nexus

- DNA Nexus Provides
  - Compute infrastructure
  - Virtualization, OS and cluster management
  - Standard and Public datasets
    - Study datasets if use their sequencing core.
  - Applications, tools and software
  - Graphical interface
- You provide
  - Study data
  - Application options and order



AWS, Azure, GCP Tools and Services

AWS, Azure GCP tools and services are divided into categories like Compute, Storage, Networking etc., in the table mentioned below.

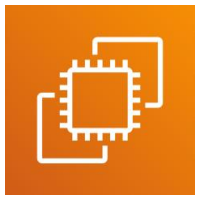


aws	Azure	Google Cloud Platform
AWS came into the market in the Year **2006**	Azure Started services in Year **2010**	GCP Launched in the year **2008**
AWS is Friendly with the open-source model from the beginning.	Azure has not so good a relationship with the open-source community.	GCP offers managed open source services that are tightly integrated into Google Cloud.
Large and Complex scale offerings of services that can potentially manipulate.	Low-Quality Support	Quite costly support fee of about $150 per month for the silver class, which is the most basic of services

What is AWS ?

- AWS (Amazon Web Services) is a Cloud Provider
- They provide you with servers and services that you can use on demand and scale easily
- AWS has revolutionized IT over time
- AWS powers some of the biggest websites in the world
- Amazon.com
- Netflix

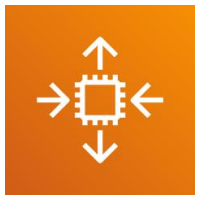
What we'll learn in this course ?



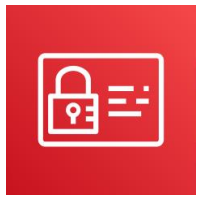
Amazon EC2

Amazon S3

Elastic Load Balancing



Auto Scaling



IAM

Amazon ECR

AWS CodeCommit

AWS CodeBuild

AWS CodeDeploy

AWS CodePipeline

Amazon Route 53

Amazon RDS

AWS Fundamentals

- Regions
- IAM
- EC2

US East (N. Virginia)	us-east-1
US East (Ohio)	us-east-2
US West (N. California)	us-west-1
US West (Oregon)	us-west-2
Africa (Cape Town)	af-south-1
Asia Pacific (Hong Kong)	ap-east-1
Asia Pacific (Mumbai)	ap-south-1
Asia Pacific (Seoul)	ap-northeast-2
Asia Pacific (Singapore)	ap-southeast-1
Asia Pacific (Sydney)	ap-southeast-2
Asia Pacific (Tokyo)	ap-northeast-1
Canada (Central)	ca-central-1
Europe (Frankfurt)	eu-central-1
Europe (Ireland)	eu-west-1
Europe (London)	eu-west-2
Europe (Paris)	eu-west-3
Europe (Stockholm)	eu-north-1
Middle East (Bahrain)	me-south-1
South America (São Paulo)	sa-east-1

26 Launched Regions

Each with multiple Availability Zones (AZ's)

84 Availability Zones

17 Local Zones

25 Wavelength Zones

For ultralow latency applications

8 Announced Regions

32 Announced Local Zones

2x More Regions

With multiple AZ's than the next largest cloud provider

245 Countries and Territories Served

108 Direct Connect Locations

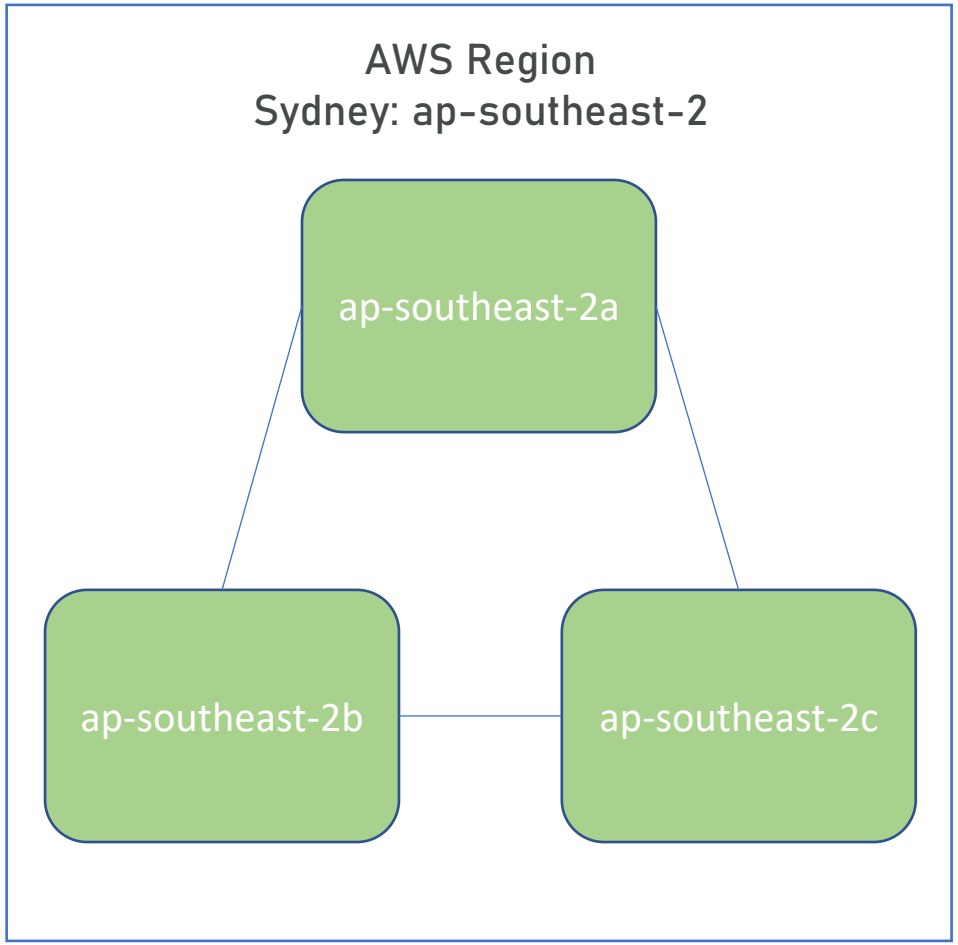
310+ Points of Presence

300+ Edge Locations and 13 Regional Edge Caches

AWS Availability Zones

Each Region has multiple, isolated locations known as Availability Zones. The code for Availability Zone is its Region code followed by a letter identifier. For example, ap-southeast-2.

Regions and Zones - Amazon Elastic Compute Cloud



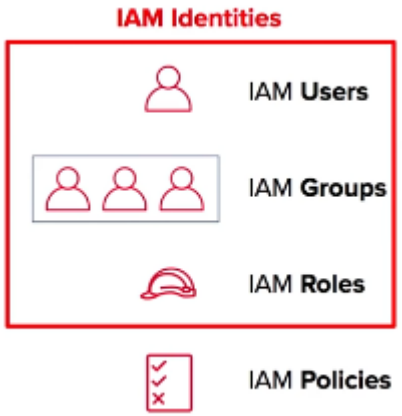
AWS Global – Quiz (Test your Skills)

1. True or False? Each AWS region is designed to be completely isolated from the other AWS regions – Ans (True)
2. True or False? Each region has a minimum number of 1 availability zones and the maximum is 4 -
   Ans - **False. The minimum is 2 while the maximum is 6.**
3. What considerations to take when choosing an AWS region for running a new application? -

   Ans –
   Services Availability: not all service (and all their features) are available in every region
   Reduced latency: deploy application in a region that is close to customers
   Compliance: some countries have more strict rules and requirements such as making sure the data stays within the borders of the country or the region. In that case, only specific region can be used for running the application
   Pricing: the pricing might not be consistent across regions so, the price for the same service in different regions might be different.

IAM Introduction-

IAM stands for Identity Access Management



- In Simple terms your whole AWS security is here

IAM Entities

- Users
- Groups
- Roles
- Policies

Priority Levels in IAM

- Explicit Deny
- Explicit Allow
- Default Deny (or Implicit Deny)

IAM Key Details

- Root account should never be used (and shared)
- New users have no permissions when their accounts are first created.
- Users must be created with proper permissions
- IAM is at the center of AWS
- Policies are written in JSON (JavaScript Object Notation)

IAM Do's and Don'ts-

- One IAM User per PHYSICAL PERSON
- One IAM Role per Application
- IAM credentials should NEVER BE SHARED
- Never, ever, ever, ever, write IAM credentials in code. EVER.
- And even less, NEVER EVER EVER COMMIT YOUR IAM credentials
- Never use the ROOT account except for initial setup.
- Never use ROOT IAM Credentials

IAM – Quiz (Test your Skills)

1. What is IAM? What are some of its features? – Ans - In short, it's used for managing users, groups, access policies & roles
2. True or False? IAM configuration is defined globally and not per region - Ans - True
3. True or False? When creating an AWS account, root account is created by default. This is the recommended account to use and share in your organization -Ans – False. Instead of using the root account, you should be creating users and use them.
4. True or False? Groups in AWS IAM, can contain only users and not other groups - Ans - True
5. True or False? Users in AWS IAM, can belong only to a single group -Ans -False. Users can belong to multiple groups
6. What are some best practices regarding IAM in AWS? - Ans –
   - Delete root account access keys and don't use root account regularly
   - Create IAM user for any physical user. Don't share users.
   - Apply "least privilege principle": give users only the permissions they need, nothing more than that.
   - Set up MFA and consider enforcing using it
   - Make use of groups to assign permissions (user -> group -> permissions)
7. What permissions does a new user have? - Ans - Only a login access.
8. True or False? If a user in AWS is using password for authenticating, he doesn't need to enable MFA – Ans - False(!). MFA is a great additional security layer to use for authentication.
9. What ways are there to access AWS? - Ans - AWS Management Console, AWS CLI & AWS SDK.
10. A user is unable to access an s3 bucket. What might be the problem?- Ans -There can be several reasons for that. One of them is lack of policy. To solve that, the admin has to attach the user with a policy what allows him to access the s3 bucket..
11. What should you use to:
    Grant access between two services/resources? Ans – Role
    Grant user access to resources/services? - Policy

IAM – Quiz (Test your Skills)

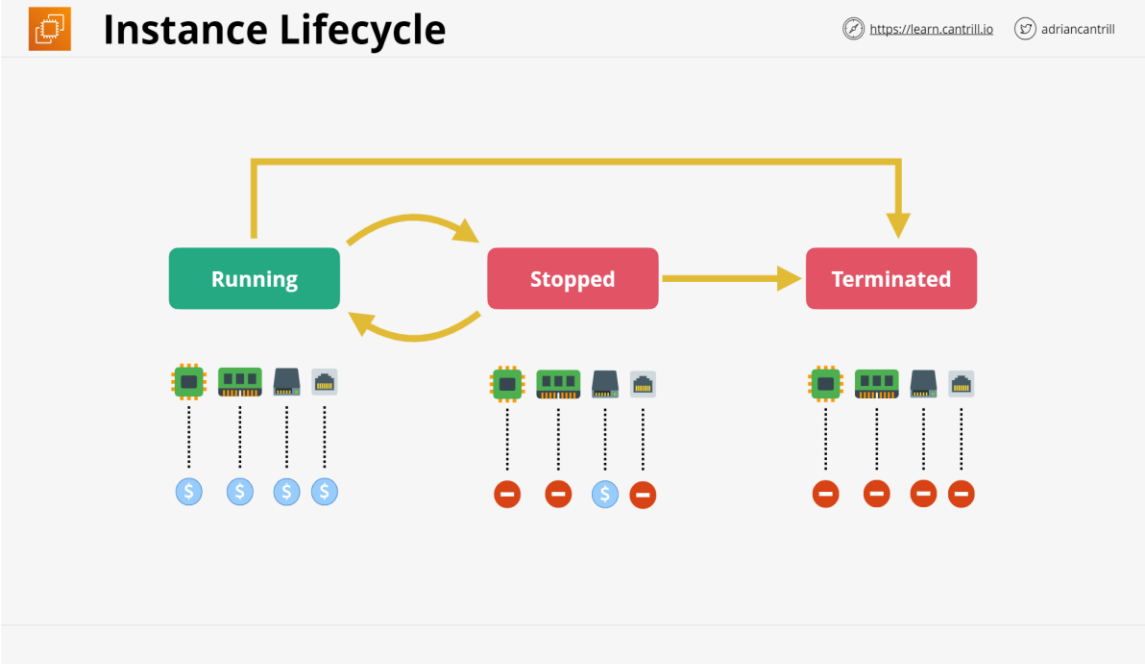
1. What statements AWS IAM policies are consisting of? – Ans –
Sid: identifier of the statement (optional)
Effect: allow or deny access
Action: list of actions (to deny or allow)
Resource: a list of resources to which the actions are applied
Principal: role or account or user to which to apply the policy
Condition: conditions to determine when the policy is applied (optional)

2. Explain the following policy:

```
{
   "Version": "2012-10-17",
   "Statement": [
      {
         "Effect:": "Allow",
         "Action": "*",
         "Resources": "*"
      }
   ]
}
```

Ans – This policy permits to perform any action on any resource. It happens to be the "AdministratorAccess" policy.

EC2- Elastic Cloud Computing



EC2 is one of most popular of AWS offering

- It mainly consists in the capability of :
- Renting virtual machines (EC2)
- Storing data on virtual drives (EBS)
- Distributing load across machines (ELB)
- Scaling the services using an auto-scaling group (ASG)
- Knowing EC2 is fundamental to understand how the Cloud works

You can :

1. Launch EC2
2. Reboot/Restart
3. Stop EC2 (like Shutdown)
4. Hibernate EC2 instance
5. Terminate EC2 (like release forever)
6. Start EC2 (if stopped)

When an EC2 instance is stopped -

- Private IP remains with the EC2 Instance
- Public IP may change when you restart a stopped EC2 instance
- The underlying lost (physical machine) may change
- EBS attached it EC2 will be charged.

EC2 Instance Pricing-

- On-Demand instances – Rent it for few seconds/hours/days etc. pay as you go $$$
- Reserved instances – Pay up front for 1-3 years and save a lot (up to 75%) $$
- Spot instances – You bid on them IF available. May terminate with 2 minutes notice. Up to 90% cheaper than On-Demand instances $
- Dedicated Instances – no other customers will share your hardware
- Dedicated hosts – book an entire physical server, control instance placement

- EC2 Pricing
- Example
- t2.small in US-EAST-1 (VIRGINIA), cost $0.023 per Hour
- If used for:
- 6 seconds, it costs $0.023/60 = $0.000383 (minimum of 60 seconds)
- 60 seconds, it costs $0.023/60 = $0.000383 (minimum of 60 seconds)
- 30 minutes, it costs $0.023/2 = $0.0115
- 1 month, it costs $0.023 * 24 * 30 = $16.56 (assuming a month is 30 days)
- X seconds (X > 60), it costs $0.023 * X / 3600
- The best way to know the pricing is to consult the pricing page: https://aws.amazon.com/ec2/pricing/on-demand/

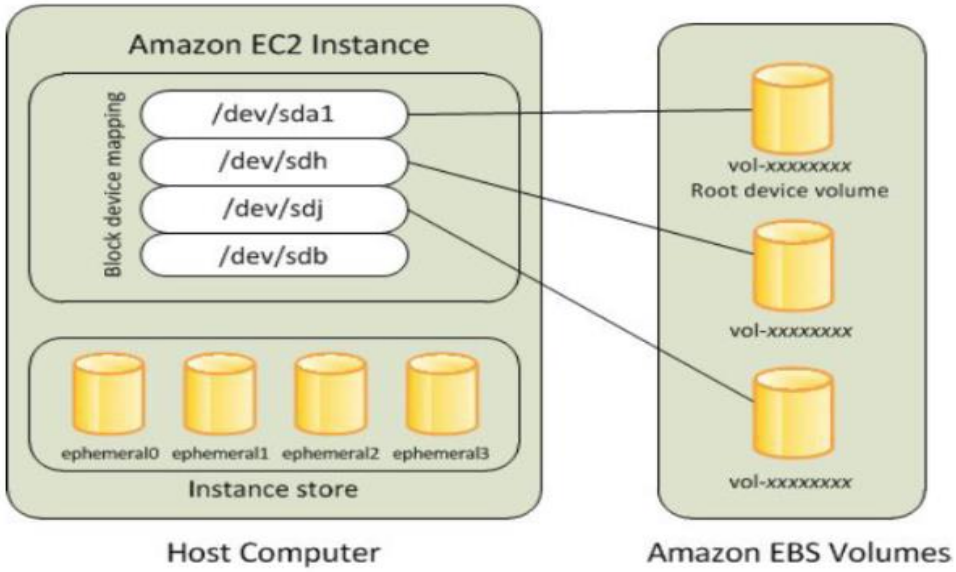
EC2 Instance

You need certain things/components to successfully launch an EC2 and log into it

- VPC, Subnet (there are default subnets)
- Security Group
- Keypair
- Storage (Usually EBS)
- AMI
- Instance Type

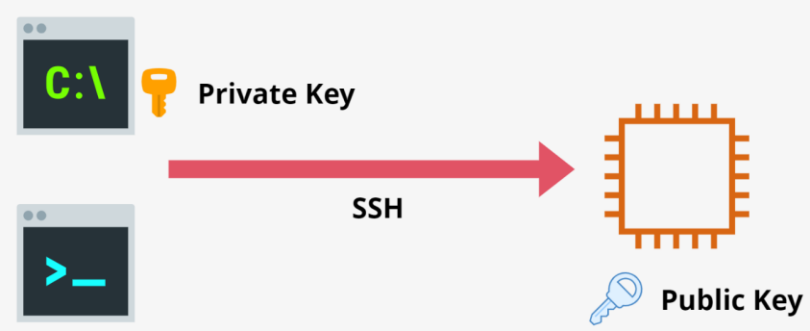
There are 2 kinds of EC2 storage

- Instance Store
- EBS Store – Elastic Block Storage



SSH Summary Table

How to SSH into your EC2 Instance - Linux / Mac OS X



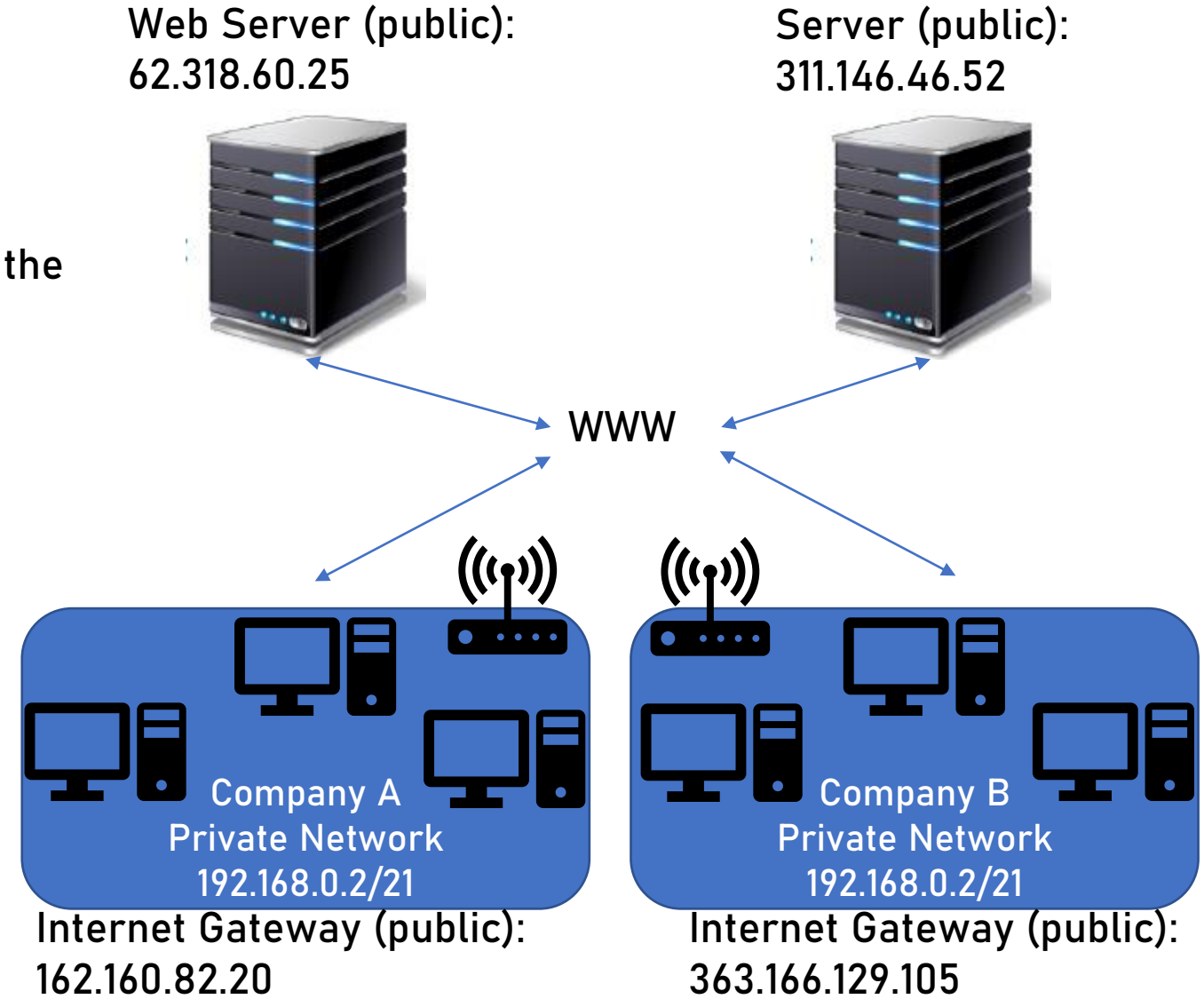
	SSH	Putty	EC2 Instance Connect
Mac	✓		✓
Linux	✓		✓
Windows < 10		✓	✓
Windows >= 10	✓	✓	✓

Private vs Public IP (IPv4)

- Networking has two sorts of IPs. IPv4 and IPv6:
- IPv4: 1.120.20.320
- IPv6: 5eef:2700:3434:1:400:e8ee:ef43:56be

- IPv4 allows for 3.7 billion different addresses in the public space
- IPv4: [0-255].[0-255].[0-255].[0-255].

Example-



Elastic IP

- When you stop and then start an EC2 instance, it can change its public IP.
- If you need to have a fixed public IP for your instance, you need an Elastic IP
- An Elastic IP is a public IPv4 IP you own as long as you don't delete it
- You can attach it to one instance at a time

Private vs Public IP (IPv4) In AWS EC2

- By default, your EC2 machine comes with:
  - A private IP for the internal AWS Network
  - A public IP, for the WWW.
- When we are doing SSH into our EC2 machines:
  - We can't use a private IP, because we are not in the same network
  - We can only use the public IP.
- If your machine is stopped and then started, the public IP can change

EC2 – Quiz (Test your Skills)

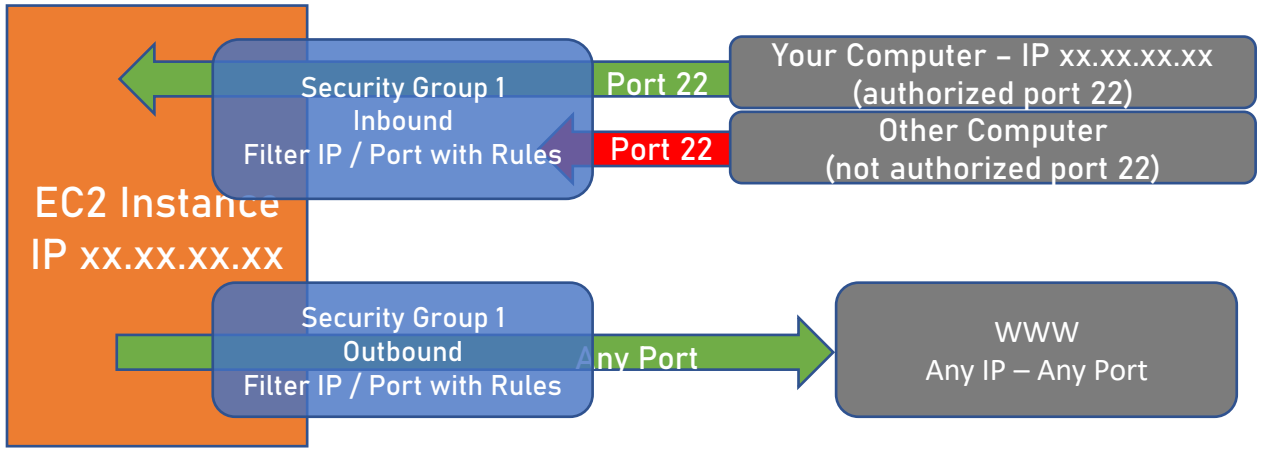
1. What is EC2– Ans - "a web service that provides secure, resizable compute capacity in the cloud"
2. True or False? EC2 is a Regional service - Ans – True. As opposed to IAM for example, which is a global service, EC2 is a regional service.
3. What are some of the properties/configuration options of EC2 instances that can be set or modified? Ans –OS (Linux, Windows) , RAM and CPU ,Networking - IP, Card properties like speed, Storage Space - (EBS, EFS, EC2 Instance Store), EC2 User Data, Security groups
4. What would you use for customizing EC2 instances? As in software installation, OS configuration, etc.-
   Ans - AMI. With AMI (Amazon Machine Image) you can customize EC2 instances by specifying which software to install, what OS changes should be applied, etc.
5. True or False? When stopping and starting an EC2 instance, its public IP changes - Ans - True
6. Why would you use an Elastic IP address - Ans - Let's say you have an instance that you need to shutdown or perform some maintenance on. In that case, what you would want to do is to move the Elastic IP address to another instance that is operational, until you finish to perform the maintenance and then you can move it back to the original instance (or keep it assigned to the second one).
7. What are the best practices around Elastic IP? Ans - The best practice is actually not using them in the first place. It's more common to use a load balancer without a public IP or use a random public IP and register a DNS record to it

Security Groups-

- Security Groups are the fundamental of network security in AWS
- They control how traffic is allowed into or out of our EC2 Machines.

Type	Protocol	Port Range	Source	Description
HTTP	TCP	80	0.0.0.0/0	test http page
SSH	TCP	22	122.149.196.85/32	
Custom TCP Rule	TCP	4567	0.0.0.0/0	java app

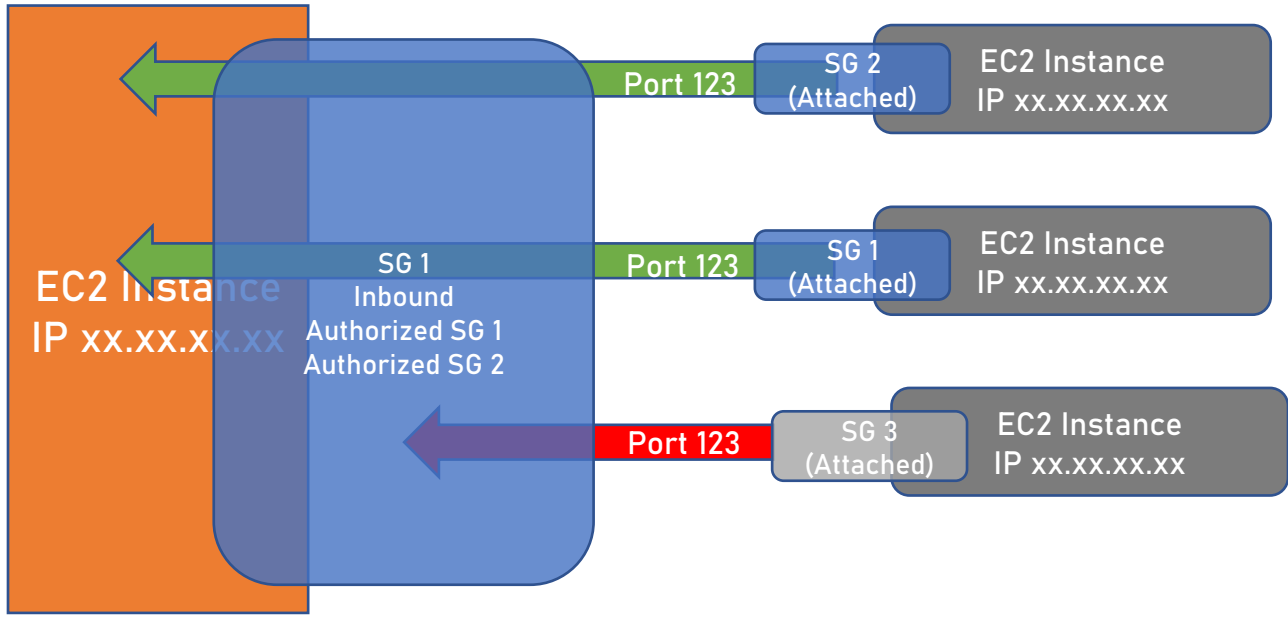
Security Groups Diagram-



Security Groups Good to Know-

- Can be attached to multiple instances
- Locked down to a region / VPC combination
- Does live “outside” the EC2 – if traffic is blocked the EC2 instance won’t see it
- It’s good to maintain one separate security group for SSH access
- If your application is not accessible (time out), then it’s a security group issue
- If your application gives a “connection refused“ error, then it’s an application error or it’s not launched
- All inbound traffic is blocked by default
- All outbound traffic is authorised by default.

Ref other Security Group Diagram



Security Groups – Quiz (Test your Skills)

1. What are security Groups– Ans - A security group acts as a virtual firewall that controls the traffic for one or more instances
2. True or False? Security groups only contain deny rules- Ans – False. Security groups only contain allow rules.
3. True or False? One security group can be attached to multiple instances ? Ans – True
4. True or False? Security groups are not locked down to a region and VPC (meaning you don't have to create a new one when switching regions)
   Ans - False. They are locked down to regions and VPC.
5. True or False? By default, when using security groups, all inbound traffic to an EC2 instance is blocked and all outbound traffic is allowed
   Ans - True.
6. Explain Security Groups and Network ACLs
   Ans - NACL - security layer on the subnet level.
   Security Group - security layer on the instance level.

S3- Simple Storage Service

- Amazon S3 is one of the main building blocks of AWS
- It's advertised as "infinitely scaling" storage
- It's widely popular and deserves its own section
- Amazon S3 allows people to store objects (files) in "buckets" (directories)
- Many websites use Amazon S3 as a backbone
- Many AWS services uses Amazon S3 as an integration as well
- Buckets are defined at the region level
- Naming convention
  - No uppercase
  - No underscore
  - 3-63 characters long
  - Not an IP
  - Must start with lowercase letter or number
- Objects (files) have a Key
- The key is the FULL path:
  - s3://my-bucket/my_file.txt
  - s3://my-bucket/my_folder1/another_folder/my_file.txt
- The key is composed of prefix + object name
  - s3://my-bucket/my_folder1/another_folder/my_file.txt
- There's no concept of "directories" within buckets
- Just keys with very long names that contain slashes

S3- Simple Storage Service

- S3 comes with the following main features:

1. Storage and pricing
2. Lifecycle management
3. Versioning
4. Encryption
5. MFA
6. Access control & policies to secure the data

- S3 charges by-

1. Storage size
2. Number of requests
3. Storage management pricing (known as tiers)
4. Data transfer pricing (objects leaving/entering AWS via the internet)
5. Transfer acceleration (an optional speed increase for moving objects via Cloudfront)
6. Cross region replication (more HA than offered by default

It can be used for hosting Website
The website URL will be:
▪ <bucket-name>.s3-website-<AWS-region>.amazonaws.com
OR
▪ <bucket-name>.s3-website.<AWS-region>.amazonaws.com

Amazon S3 – Versioning

- You can version your files in Amazon S3
- It is enabled at the bucket level
- Same key overwrite will increment the “version”: 1, 2, 3….
- It is best practice to version your buckets Protect against unintended deletes (ability to restore a version)
- Easy roll back to previous version
- Notes:
  - Any file that is not versioned prior to enabling versioning will have version “null”
  - Suspending versioning does not delete the previous versions

- Objects (files) have a Key
- The key is the FULL path:
  - s3://my-bucket/my_file.txt
  - s3://my-bucket/my_folder1/another_folder/my_file.txt
- The key is composed of prefix + object name
  - s3://my-bucket/my_folder1/another_folder/my_file.txt
- There’s no concept of “directories” within buckets
- Just keys with very long names that contain slashes

Amazon S3 - Encryption

- There are 4 methods of encrypting objects in S3
- SSE-S3: encrypts S3 objects using keys handled & managed by AWS
- SSE-KMS: leverage AWS Key Management Service to manage encryption keys
- SSE-C: when you want to manage your own encryption keys
- Client Side Encryption

Amazon S3 – Storage Classes

- Amazon S3 Standard - General Purpose
- Amazon S3 Standard-Infrequent Access (IA)
- Amazon S3 One Zone-Infrequent Access
- Amazon S3 Intelligent Tiering
- Amazon Glacier
- Amazon Glacier Deep Archive
- Amazon S3 Reduced Redundancy Storage (deprecated - omitted)

Amazon S3 - Security

- User based
  - IAM policies - which API calls should be allowed for a specific user from IAM console
- Resource Based
  - Bucket Policies - bucket wide rules from the S3 console - allows cross account
  - Object Access Control List (ACL) – finer grain
  - Bucket Access Control List (ACL) – less common

Note: an IAM principal can access an S3 object if

- the user IAM permissions allow it OR the resource policy ALLOWS it
- AND there's no explicit DENY

Amazon S3 – Bucket Policies

- JSON based policies
  - Resources: buckets and objects
  - Actions: Set of API to Allow or Deny
  - Effect: Allow / Deny
  - Principal: The account or user to apply the policy to
- Use S3 bucket for policy to:
  - Grant public access to the bucket
  - Force objects to be encrypted at upload
  - Grant access to another account (Cross
  - Account)

```
{
    "Version": "2012-10-17",
    "Statement": [
        {
            "Sid": "PublicRead",
            "Effect": "Allow",
            "Principal": "*",
            "Action": [
                "s3:GetObject"
            ],
            "Resource": [
                "arn:aws:s3:::examplebucket/*"
            ]
        }
    ]
}
```

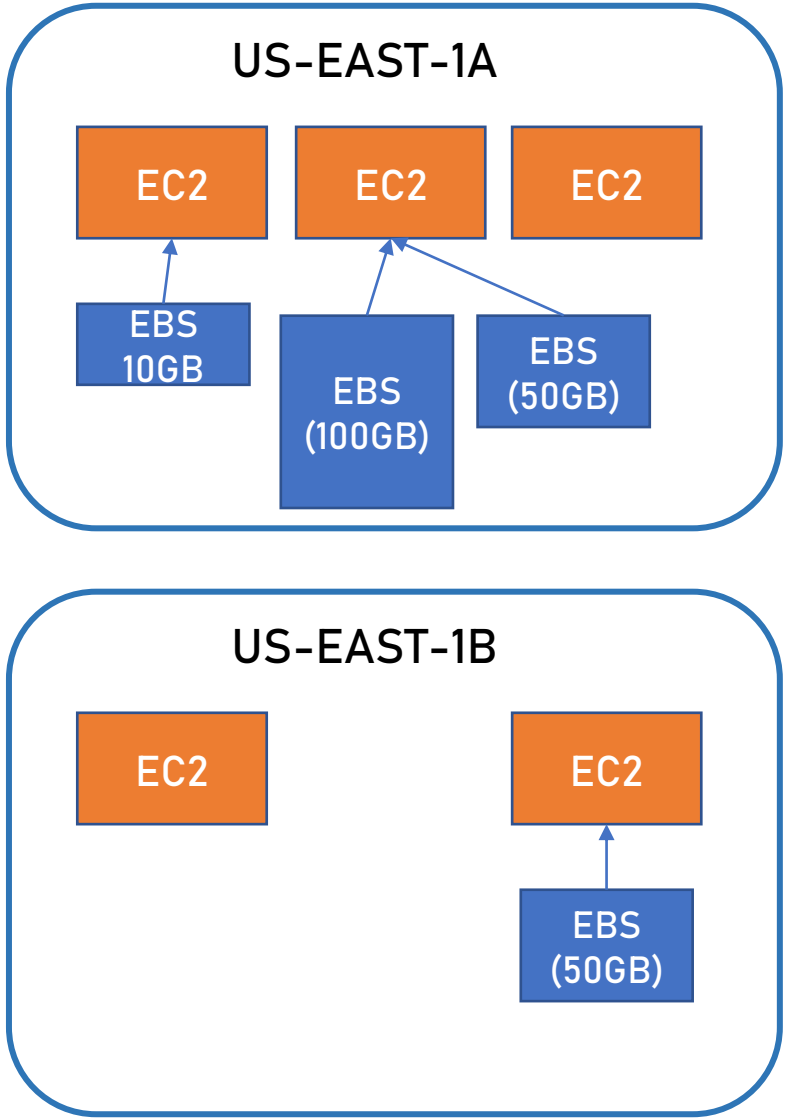
Amazon S3– Quiz (Test your Skills)

1. Explain what is AWS S3– Ans - S3 stands for: Simple Storage Service.
S3 is a object storage service which is fast, scalable and durable. S3 enables customers to upload, download or store any file or object that is up to 5 TB in size.
2. What is a bucket Ans – An S3 bucket is a resource which is similar to folders in a file system and allows storing objects, which consist of data
3. True or False? A bucket name must be globally unique - Ans – True
4. Explain folders and objects regarding buckets - Ans - Folder - any sub folder in an s3 bucket
Object - The files which are stored in a bucket
5. Explain Object Durability and Object Availability - Ans - The percent over a one-year time period that a file will not be lost
Object Availability: The percent over a one-year time period that a file will be accessible
6. What security measures have you taken in context of S3 - Ans - * Enable versioning. * Don't make bucket public.
* Enable encryption if it's disabled.
7. What is "Amazon S3 Transfer Acceleration“ - Ans - Amazon S3 Transfer Acceleration enables fast, easy, and secure transfers of files over long distances between your client and an S3 bucket
8. A user is unable to access an s3 bucket. What might be the problem? -

Ans - There can be several reasons for that. One of them is lack of policy. To solve that, the admin has to attach the user with a policy what allows him to access the s3 bucket.

EBS- Elastic Block Storage

- An EC2 machine loses its root volume (main drive) when it is manually terminated.
- Unexpected terminations might happen from time to time (AWS would email you)
- Sometimes, you need a way to store your instance data somewhere
- An EBS (Elastic Block Store) Volume is a network drive you can attach to your instances while they run
- It allows your instances to persist data



EBS Volume Example

EBS- Volume Types

- EBS Volumes come in 4 types
  - GP2 (SSD): General purpose SSD volume that balances price and performance for a wide variety of workloads
  - IO1 (SSD): Highest-performance SSD volume for mission-critical low-latency or high-throughput workloads
  - ST1 (HDD): Low cost HDD volume designed for frequently accessed, throughput-intensive workloads
  - SC1 (HDD): Lowest cost HDD volume designed for less frequently accessed workloads

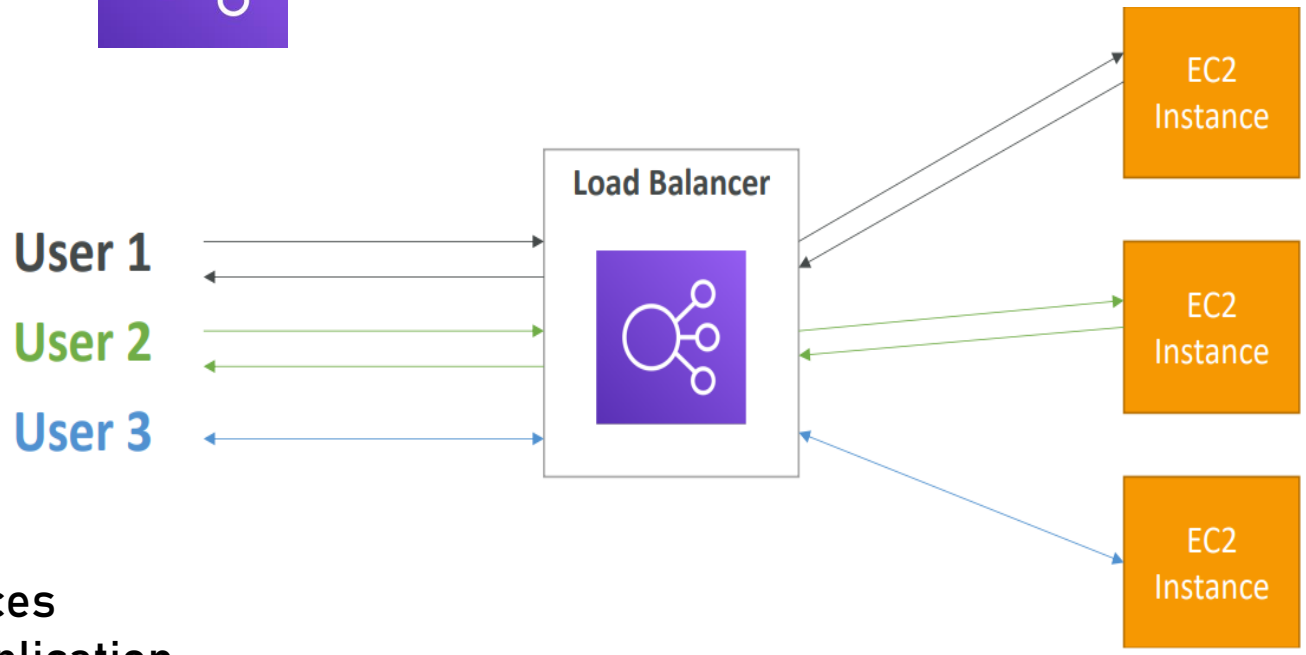
- EBS Volumes are characterized in Size | Throughput | IOPS (I/O Ops Per Sec)
- When in doubt always consult the AWS documentation – it's good!
- Only GP2 and IO1 can be used as boot volumes

Amazon EBS– Quiz (Test your Skills)

1. What happens to EBS volumes when the instance is terminated?– Ans - By default, the root volume is marked for deletion, while other volumes will still remain. You can control what will happen to every volume upon termination.
2. What happens to the EC2 disk (EBS) when the instance is stopped? - Ans – Disk is intact and can be used when the instance starts.
3. True or False? EBS volumes are locked to a specific availability zone– Ans – True
4. Explain EBS snapshot - Ans - EBS snapshots used for making a backup of the EBS volume at point of time.
5. What are the use cases for using EBS snapshots?- Ans - Backups of the data & Moving the data between AZs
6. Is it possible to attach the same EBS volume to multiple EC2 instances?- Ans - Yes, with multi-attach it's possible to attach a single EBS volume to multiple instances.
7. True or False? EBS is a network drive hence, it requires network connectivity - Ans - True
8. What EBS volume types are there? - Ans - HDD (ST 1, SC 1): Low-cost HDD volumes
   SSD -
   io1, io2: Highest performance SSD
   gp2, gp3: General purpose SSD
9. If you need an EBS volume for low latency workloads, which volume type would you use? - Ans - SSD - io1, io2
10. If you need an EBS volume for workloads that require good performance but the cost is also an important aspect for you, which volume type would you use? - Ans - SSD - gp2, gp3
11. If you need an EBS volume for high-throughput, which volume type would you use? - Ans - SSD - io1, io2
12. If you need an EBS volume for infrequently data access, which volume type would you use? - Ans - HDD - sc1
13. Which EBS volume types can be used as boot volumes for EC2 instances?? - Ans - SSD: gp2, gp3, io1, io2

Elastic Load Balancing(ELB)

- Load balancers are servers that forward internet traffic to multiple servers (EC2 Instances) downstream



Why use a load balancer?

- Spread load across multiple downstream instances
- Expose a single point of access (DNS) to your application
- Seamlessly handle failures of downstream instances
- Do regular health checks to your instances
- Provide SSL termination (HTTPS) for your websites
- Enforce stickiness with cookies
- High availability across zones
- Separate public traffic from private traffic

Types of load balancer on AWS

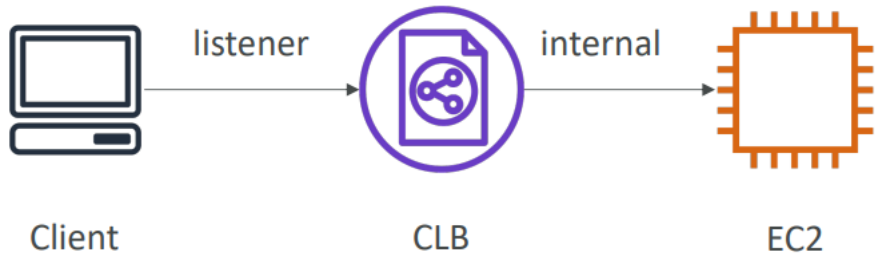
- AWS has 3 kinds of managed Load Balancers

- In AWS, there are three types of LBs:
  - Classic LBs - (v1 - old generation) – 2009 ▪ HTTP, HTTPS, TCP
  - Application LBs – (v2 - new generation) – 2016 ▪ HTTP, HTTPS, WebSocket
  - Network LBs - r (v2 - new generation) – 2017 ▪ TCP, TLS (secure TCP) & UDP

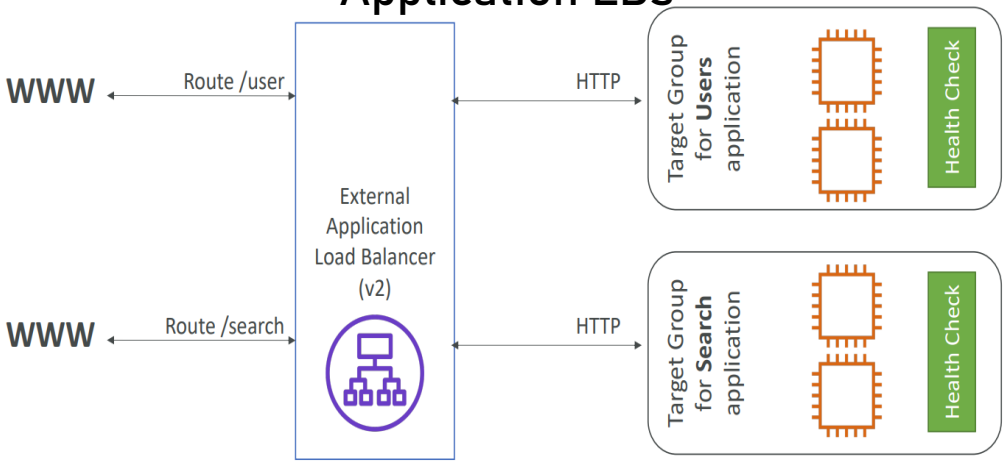
- Overall, it is recommended to use the newer / v2 generation load balancers as they provide more features

- You can setup internal (private) or external (public) ELBs

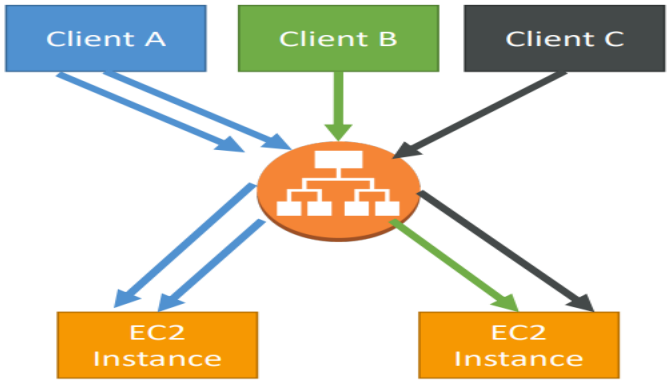
Classic LBs



Application LBs



Network LBs



Elastic Load Balancing(ELB)

- The lifecycle of a request to view a website behind an ELB:
- The browser requests the IP address for the load balancer from DNS.
- DNS provides the IP.
- With the IP at hand, your browser then makes an HTTP request for an HTML page from the Load Balancer.
- AWS perimeter devices checks and verifies your request before passing it onto the LB.
- The LB finds an active webserver to pass on the HTTP request.
- The webserver returns the requested HTML file.
- The browser receives the HTML file it requested and renders the graphical representation of it on the screen.

- Load balancers are a regional service. They do not balance load across different regions. You must provision a new ELB in each region that you operate out of.

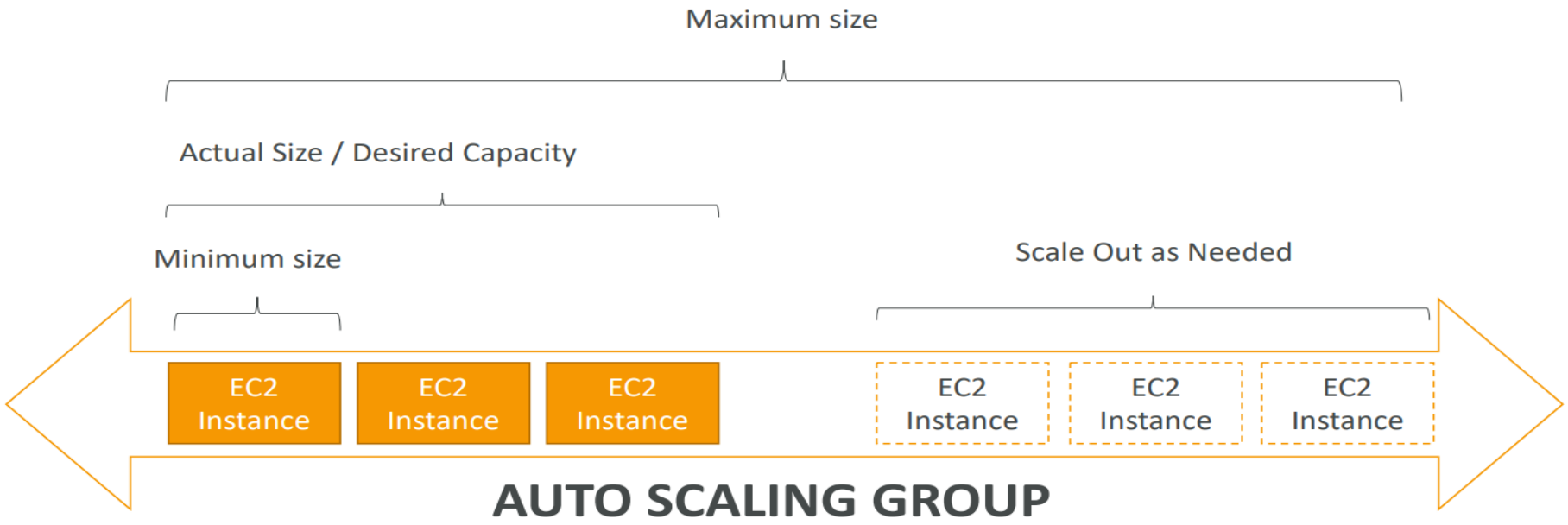
Amazon ELB– Quiz (Test your Skills)

1. What is ELB (Elastic Load Balancing)?– Ans - "Elastic Load Balancing automatically distributes incoming application traffic across multiple targets, such as Amazon EC2 instances, containers, IP addresses, and Lambda functions."
2. True or False? Elastic Load Balancer is a managed resource (= AWS takes care of it)? Ans – True. AWS responsible for making sure ELB is operational and takes care of lifecycle operations like upgrades, maintenance and high availability.
3. Which load balancer would you use for services which use HTTP or HTTPS traffic? Ans – Application Load Balancer (ALB).
4. True or False? For application load balancer, cross zone load balancing is always on and can't be disabled– Ans - False. It's disabled by default

Auto Scaling

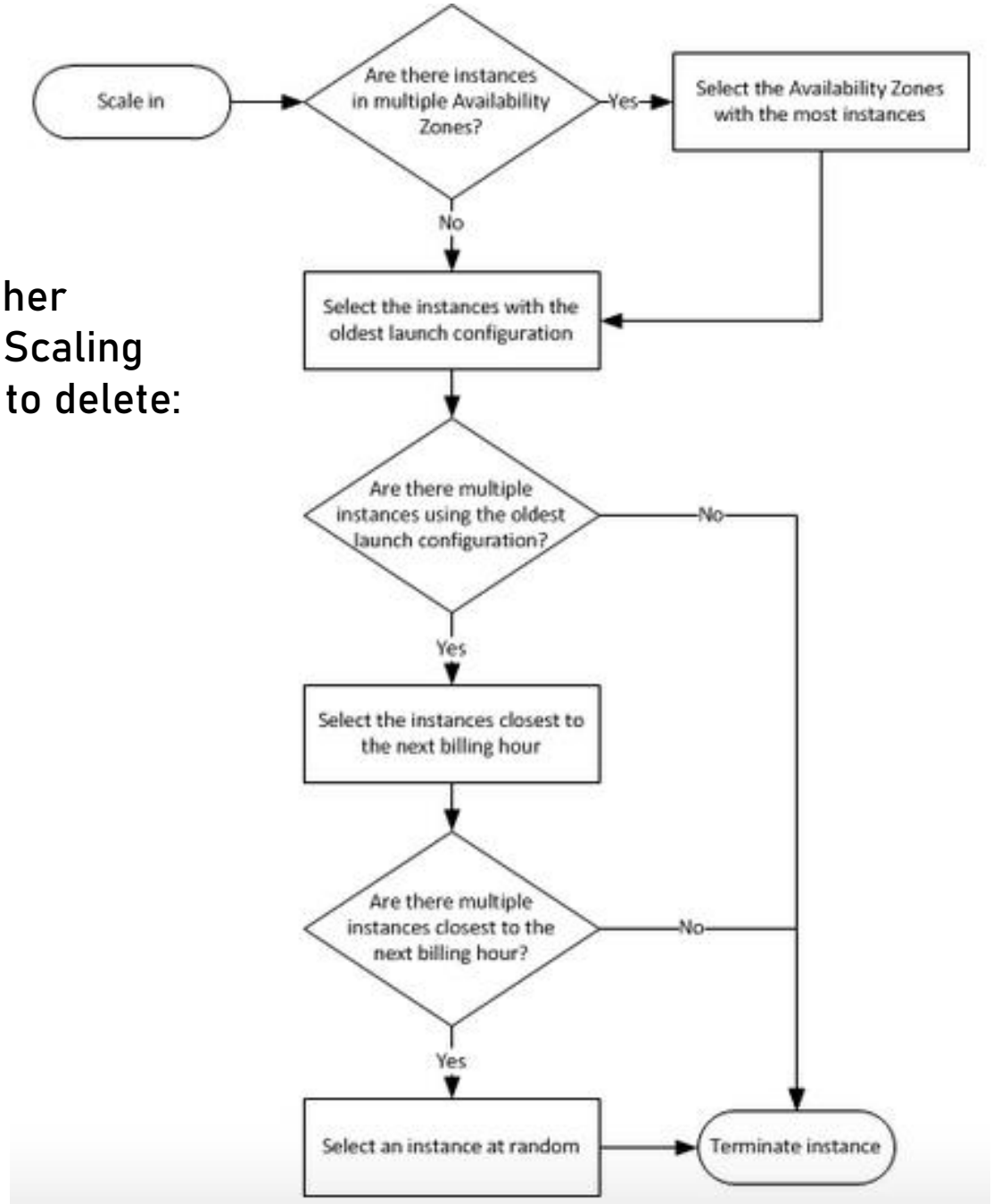
- In real-life, the load on your websites and application can change
- In the cloud, you can create and get rid of servers very quickly

- The goal of an Auto Scaling Group (ASG) is to:
- Scale out (add EC2 instances) to match an increased load
- Scale in (remove EC2 instances) to match a decreased load
- Ensure we have a minimum and a maximum number of machines running



Auto Scaling

This flow chart can provide further clarity on how the default Auto Scaling policy decides which instances to delete:



Amazon ASG– Quiz (Test your Skills)

1. Explain Auto Scaling Group?–

Ans - "An Auto Scaling group contains a collection of Amazon EC2 instances that are treated as a logical grouping for the purposes of automatic scaling and management. An Auto Scaling group also enables you to use Amazon EC2 Auto Scaling features such as health check replacements and scaling policies"

2. You have two instance running as part of ASG. You change the desired capacity to 1. What will be the outcome of this change?

Ans - One of the instances will be terminated.

3. How can you customize the trigger for the scaling in/out of an auto scaling group?

Ans - One way is to use CloudWatch alarms where an alarm will monitor a metric and based on a certain value (or range) you can choose to scale-in or scale-out the ASG.

4. How can you customize the trigger for the scaling in/out of an auto scaling group?

Ans - True, this is why when it terminates instances, it chooses the AZ with the most instances.

AWS RDS

- RDS stands for Relational Database Service
- It's a managed DB service for DB use SQL as a query language.
- It allows you to create databases in the cloud that are managed by AWS
  - Postgres
  - MySQL
  - MariaDB
  - Oracle
  - Microsoft SQL Server
  - Aurora (AWS Proprietary database)

Q & A

Q & A

?