Cloud Introduction and Concepts

Cloud Computing
Cloud Computing and Virtualization
Data Center
AWS Region and AZs
Networking Basics
Understanding CIDR
Shared Responsibility in AWS
Cloud Terminology
AWS Free Tier and Global Infrastructure

Cloud Computing

The practice of using a network of remote servers hosted on the Internet to store, manage, and process data, rather than a local server or a personal computer.

OR

Using someone else's computer to do your job.

Data Center

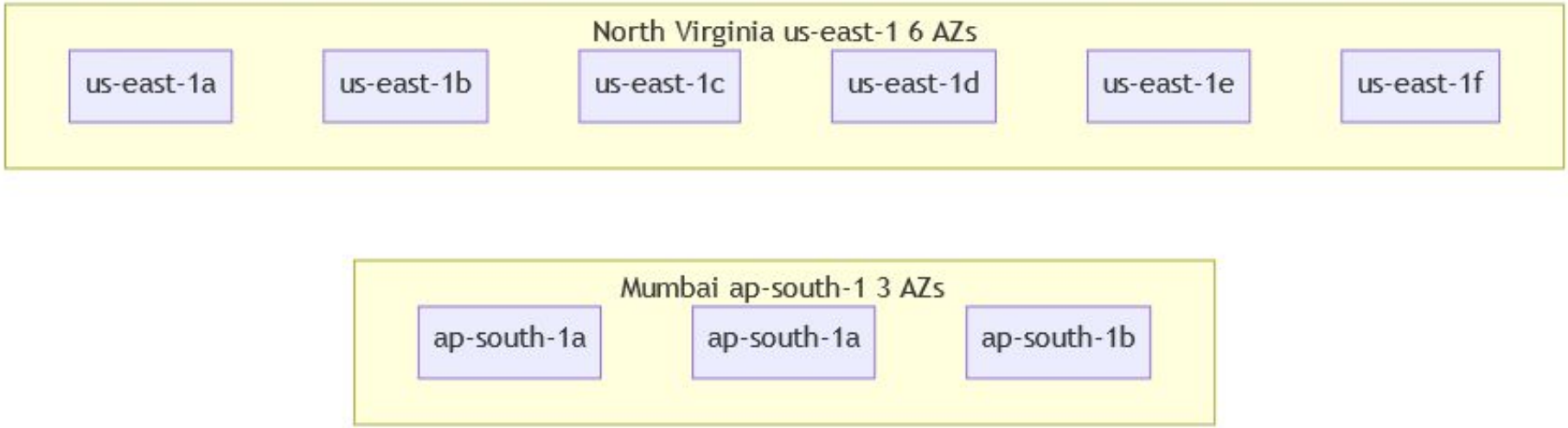
- A data center is a **physical facility** that organizations use to house their applications and data.
- Its design includes Computers/Machine/Servers, Routers, Switches, Storage systems etc.

AWS Region and AZs

Regions : AWS has the concept of a Region, which is a **physical location around the world** where we cluster data centers.

Availability Zone : We call each group of logical data centers an Availability Zone. Each **AWS Region** consists of *multiple, isolated, and physically separate AZs within a geographic area*. Each AZ has independent power, cooling, and physical security and is connected via redundant, ultra-low-latency networks.

AWS Region and AZs



AWS Free Tier and Global Infrastructure

https://aws.amazon.com/free/

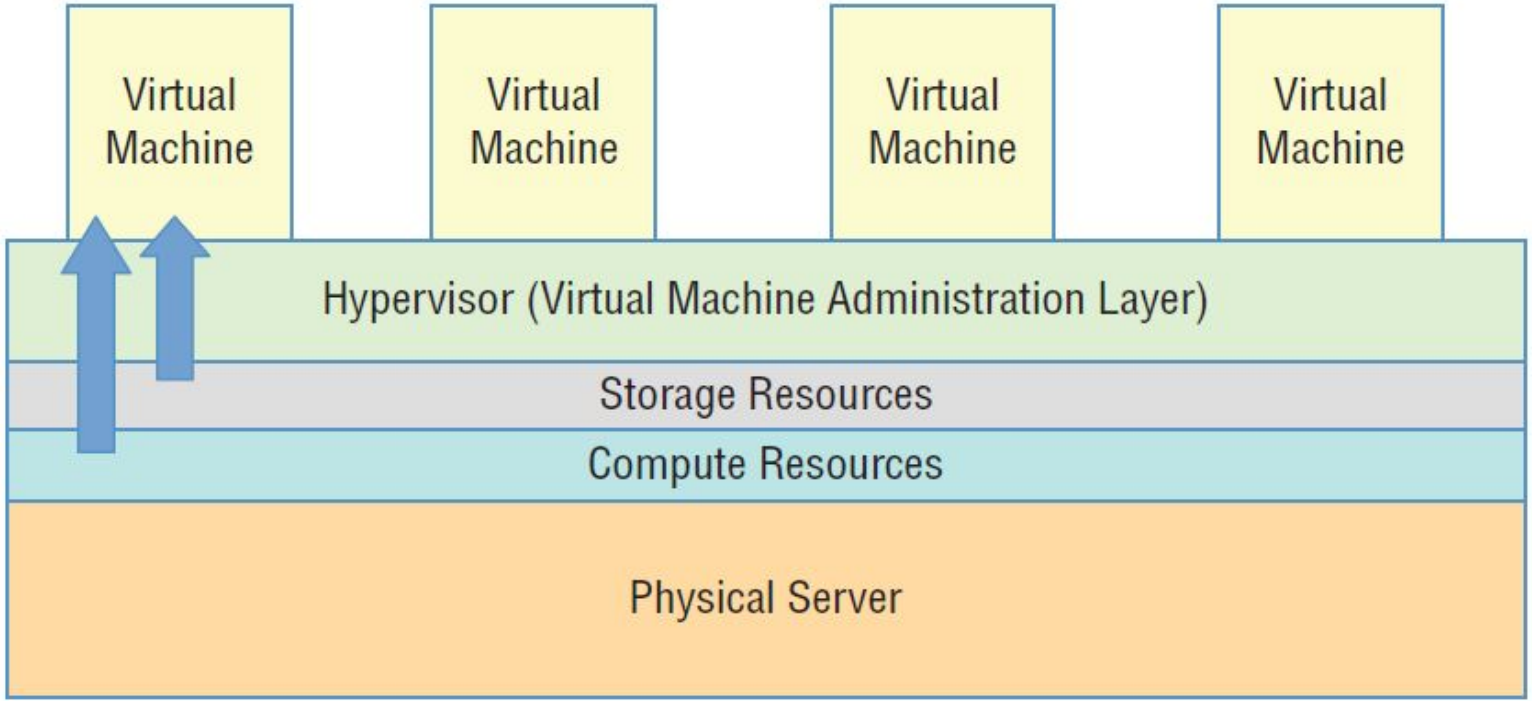
https://www.infrastructure.aws/

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

https://aws.amazon.com/economics/

Cloud Computing and Virtualization

The technology that lies at the core of all cloud operations is virtualization.



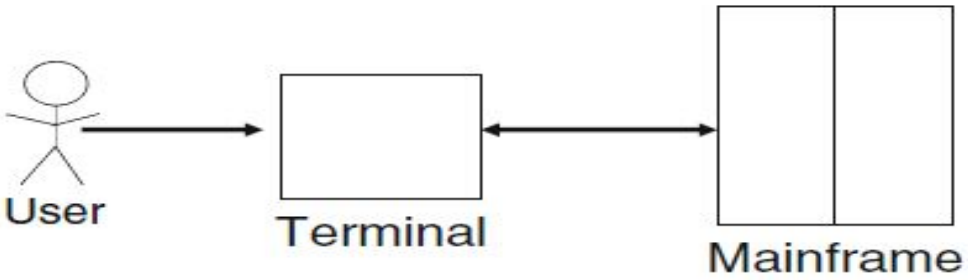
Cloud Computing and Virtualization

- Major cloud providers have **enormous server farms** where hundreds of **thousands of servers and data drives** are maintained along with the network cabling necessary to connect them.

Computing Phases

Phases

1. Mainframe Computing

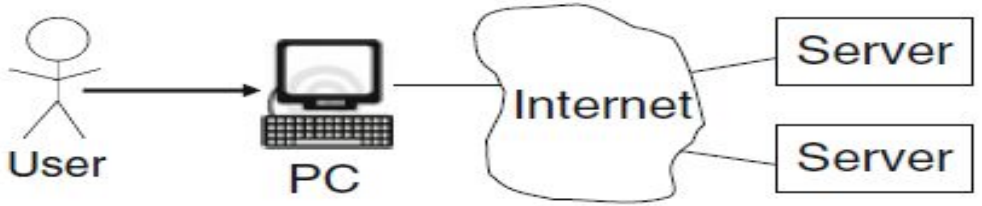


2. PC Computing

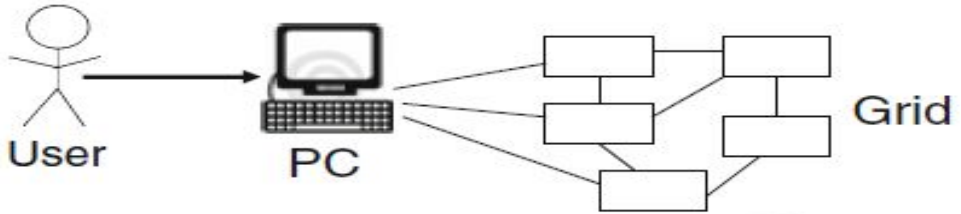
3. Network Computing

Computing Phases

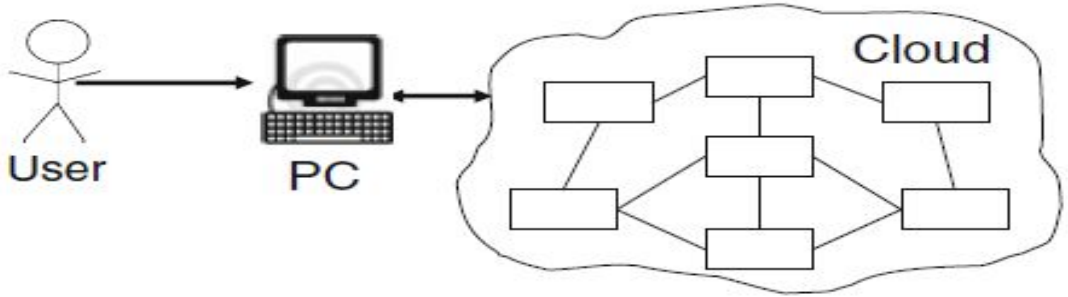
4. Internet Computing



5. Grid Computing



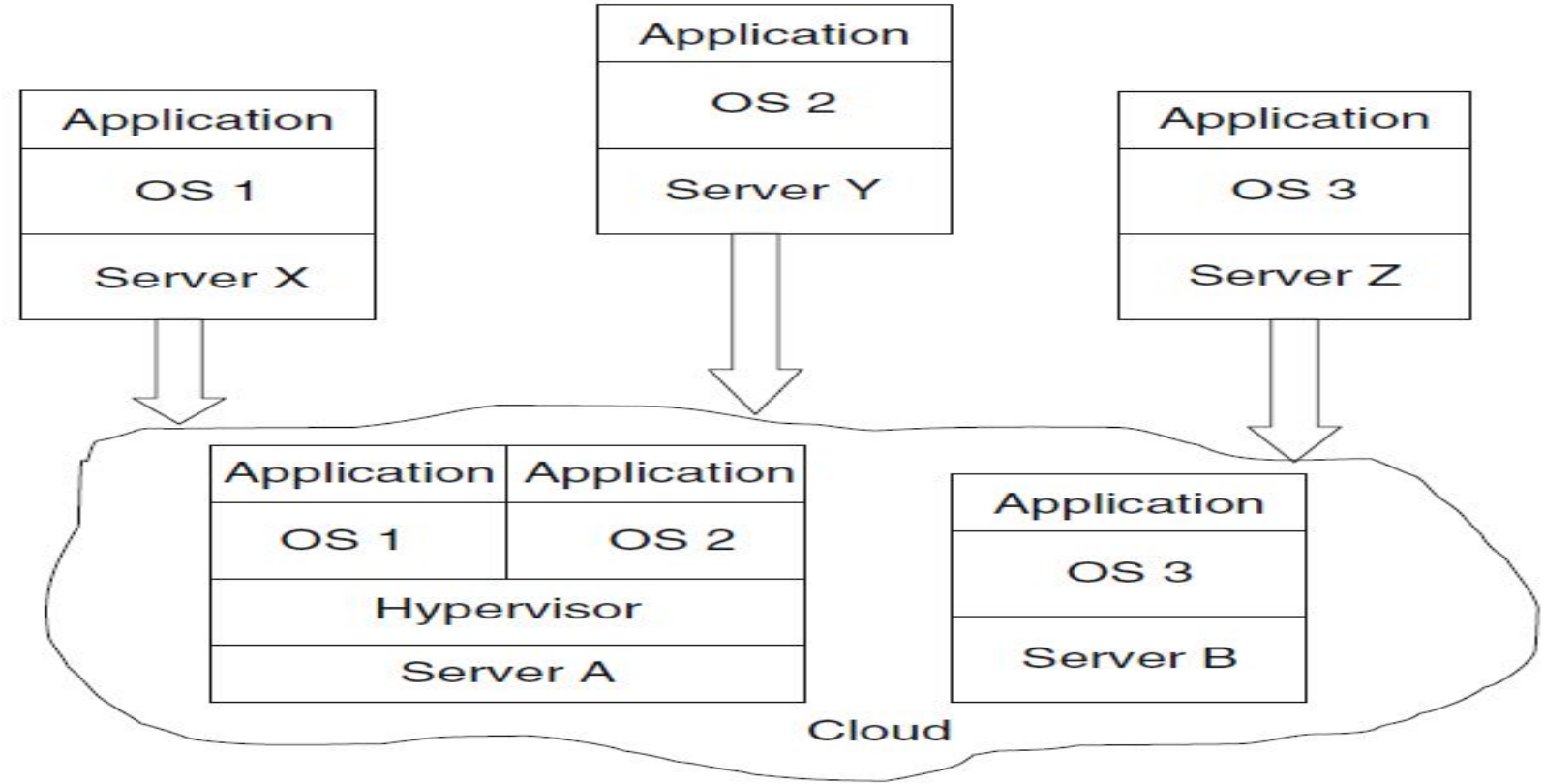
6. Cloud Computing



Virtualization

- The advantage of cloud computing is the ability to virtualize and **share resources** among different applications with the objective for **better server utilization.**
- A well-built virtualized environment could provide a virtual server using **storage, memory, compute cycles, and network bandwidth** collected from the most efficient mix of available sources it can find.
- In non-cloud computing three independent platforms exist for three different applications running on its own server. In the cloud, servers can be shared, or virtualized, for operating systems and applications resulting in fewer servers.

Virtualization



Networking Basics – IP Address

- **IP Address:**
  - An **IP address** is a number identifying of a computer or another device on the Internet.
  - IP addresses uniquely identify the source and destination of data transmitted with the Internet Protocol.

IPv4 and IPv6

- IPv4 addresses are **32 bits long** i.e 4 Bytes. (8bits is 1 Byte)
- An example of an IPv4 address is **216.58.216.164**.
- The maximum value of a 32-bit number is 2^{32}, or 4,294,967,296 i.e **4.3 Billion.**
- A major advantage of IPv6 is that it uses **128 bits** of data to store an address, permitting 2^{128} unique addresses or 340,282,366,920,938,463,463,374,607,431,768,211,456
- The size of IPv6's address space — 340 duodecillion this is much, much larger than IPv4.
- An example of an IPv6 address is **2402:8100:3000:378:b551:61bb:c1ff:977a**

IPv4 Overview

- Converting a dotted decimal IP address to binary is simple. Given that there are 4 numbers, each decimal number is represented in binary as an octet. The high order bits in binary are always the leftmost ones.
- The largest single value an IP address may have in decimal is 255 since this is the sum when all of the bits are added together i.e. **2^7 + 2^6 + 2^5 + 2^4 + 2^3 + 2^2 + 2^1 + 2^0**.
- When all of the bits are turned off, the number is 0. The range of an IP address is therefore 0 - 255, 256 values. (**216.58.216.164)**

Counting in Binary								
Binary	1	1	1	1	1	1	1	1
Powers of 2	2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0
Decimal	128	64	32	16	8	4	2	1

IPv4 Overview

- The following is how the IP address **68.125.16.250** would be converted and represented in binary form.

	Decimal	Broken Down	Binary Addition	Binary
First Octet	68	64 + 4	$2^6 + 2^2$	01000100
Second Octet	125	64 + 32 + 16 + 8 + 4 + 1	$2^6 + 2^5 + 2^4 + 2^3 + 2^2 + 2^0$	01111101
Third Octet	16	16	2^4	00010000
Fourth Octet	250	128 + 64 + 32 + 16 + 8 + 2	$2^7 + 2^6 + 2^5 + 2^4 + 2^3 + 2^1$	11111010

- The end result looks like **01000100.***01111101*.**00010000.***11111010*.
- IP Address **68.125.16.250** is represented in 32-bits Binary Values as **01000100.***01111101*.**00010000.***11111010*

IP Address Classes

Class	Address range	Supports
Class A	**1**.0.0.1 to **126**.255.255.254	Supports 16 million hosts on each of 127 networks.
Class B	**128**.1.0.1 to **191**.255.255.254	Supports 65,000 hosts on each of 16,000 networks.
Class C	**192**.0.1.1 to **223**.255.254.254	Supports 254 hosts on each of 2 million networks.
Class D	**224**.0.0.0 to **239**.255.255.255	Reserved for multicast groups.
Class E	**240**.0.0.0 to **254**.255.255.254	Reserved for future use, or research and development purposes.

Ranges 127.x.x.x are reserved for the loopback or localhost

Private IP - IPv4

- The **Internet Assigned Numbers Authority** (IANA) established certain blocks of IPV4 addresses for the use of private (LAN) and public (Internet) addresses.
- Private IP can only allow certain values

10.0.0.0 – 10.255.255.255 (10.0.0.0/8) <= **in big networks (Class A)**

172.16.0.0 – **172.31**.255.255 (172.16.0.0/12) <= **default AWS one (Class B)**

192.168.0.0 – **192.168**.255.255 (192.168.0.0/16) <= **example: home networks (Class C)**

Public IP - IPv4

- All the rest of the IP on the internet are **public IP**

Connectivity

- If you want to access any machine in the **same network**, use **Private IP**.
- If you want to access any machine in some **other network**, use **Public IP**

Understanding CIDR - Classless Inter-Domain Routing

A CIDR has two components

- The base IP (XX.XX.XX.XX)
- The Subnet Mask (/26)

The base IP represents an IP contained in the range.
The subnet masks defines how many bits can change in the IP

The subnet mask can take two forms. Examples:

255.255.255.0 this is less common

/24 this is more common

Understanding CIDRs Subnet Masks

The subnet masks basically allows part of the underlying IP to get additional next values from the base IP.

- */32 allows for 1 IP = 2^0 -> **35.65.85.98**/32*
- */31 allows for 2 IP = 2^1*
- */30 allows for 4 IP = 2^2*
- */29 allows for 8 IP = 2^3*
- */28 allows for 16 IP = 2^4*
- */27 allows for 32 IP = 2^5*
- */26 allows for 64 IP = 2^6*
- */25 allows for 128 IP = 2^7*
- */24 allows for 256 IP = 2^8*
- */20 allows for 4096 IP = 2^12*
- */16 allows for 65,536 IP = 2^16*
- */0 allows for all IPs = 2^32 -> 0.0.0.0/0*

CIDR - Calculation

- 0.0.0.0/0 == all IPs
- But we can define for ex:
- 192.168.0.0/26: 192.168.0.0 – 192.168.0.63 (64 IP)
- https://www.ipaddressguide.com/cidr

CIDR - Overview

- CIDR are used for Security Groups rules, or AWS networking in general.
- They help to define an IP address range
- WW.XX.YY.ZZ/**32** == one IP
- 0.0.0.0/0 == all IPs
- But we can define for ex:
- 192.168.0.0/26: 192.168.0.0 – 192.168.0.63 (64 IP)
- https://www.ipaddressguide.com/cidr

How to remember this

- **/32** – no IP number can change
- **/24** - last IP number can change
- **/16** – last IP two numbers can change
- **/8** – last IP three numbers can change
- **/0** – all IP numbers can change

Lets Do the Math…

What will be the range and the total number of IP address available.

192.168.0.0/24 = … ?

192.168.0.0/16 = … ?

134.56.78.123/32 = … ?

0.0.0.0/0 = … ?

Lets Do the Math…

What will be the range and the total number of IP address available.

192.168.0.0/24 = … ?

- ***192.168.0.0 – 192.168.0.255 (256 IP)***

192.168.0.0/16 = … ?

- ***192.168.0.0 – 192.168.255.255 (65,536 IP)***

134.56.78.123/32 = … ?

- ***Only 134.56.78.123***

0.0.0.0/0

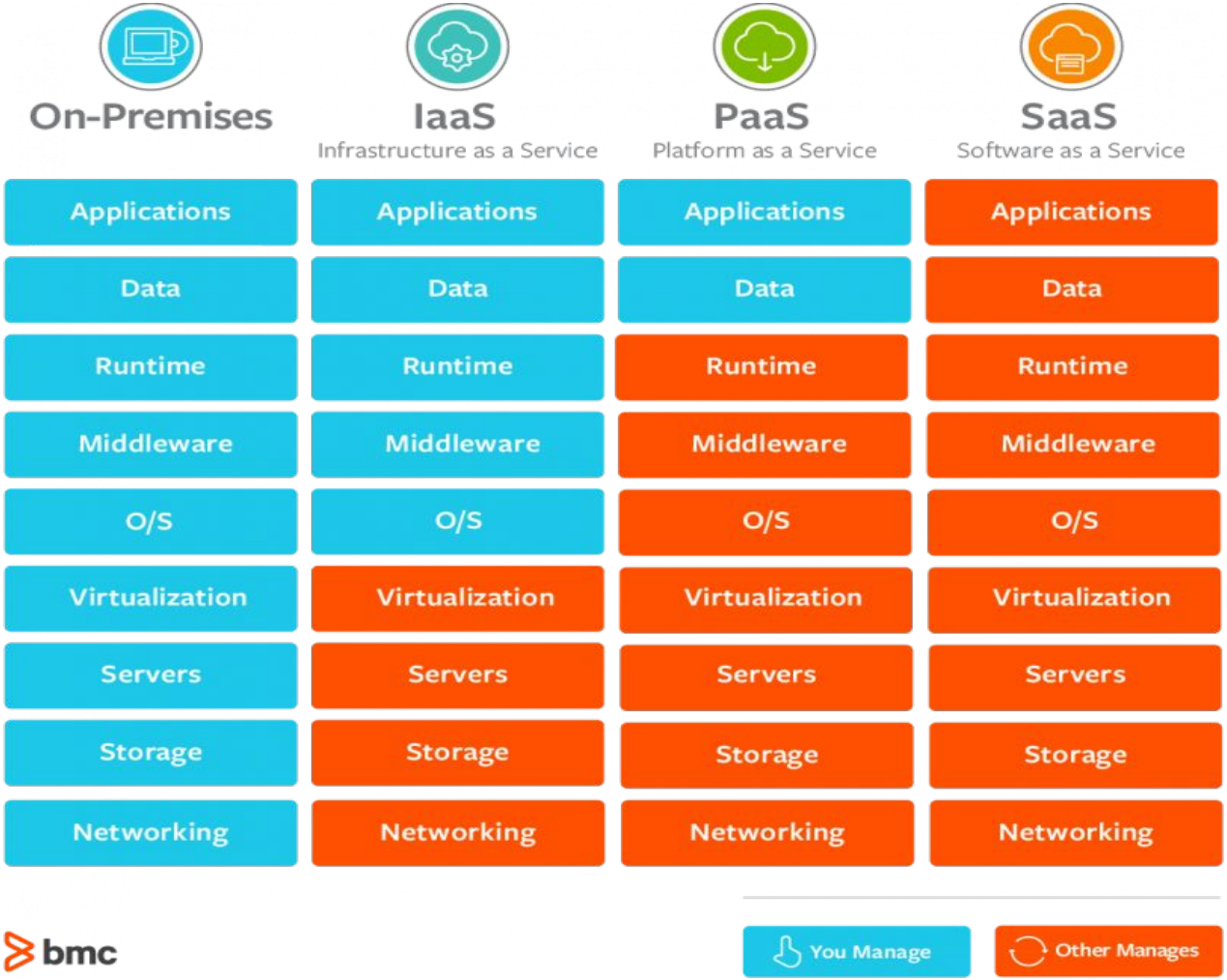
- ***All IP!***

- For above calculation : https://www.ipaddressguide.com/cidr

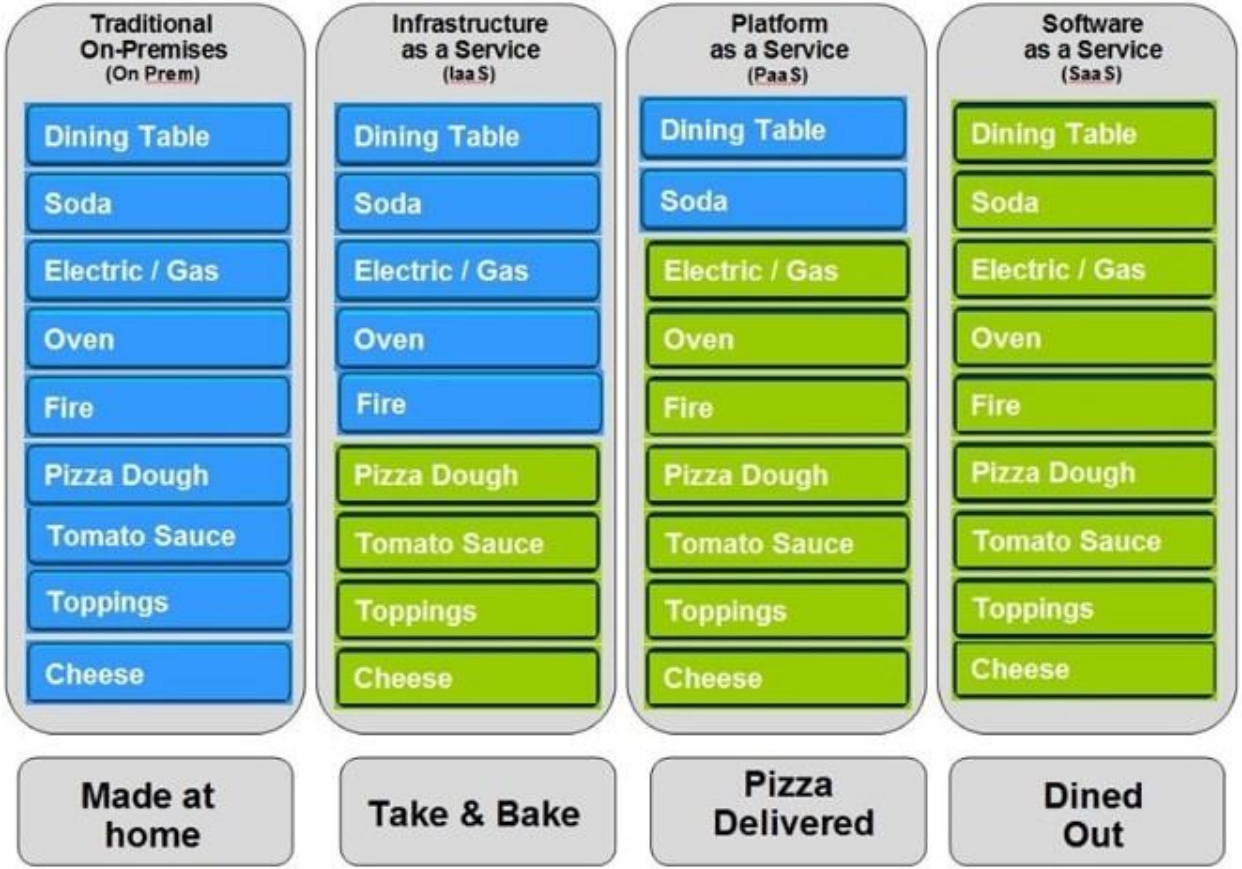
Service Models or Cloud Stack

SaaS, PaaS, and IaaS are simply three ways to describe how you can use the cloud for your business.

- **IaaS**: cloud-based services, pay-as-you-go for services such as storage, networking, and virtualization
- **PaaS**: hardware and software tools available over the internet.
- **SaaS**: software that's available via a third-party over the internet.
- **On-premise**: software that's installed in the same building as your business.


On-Premises
IaaS
Infrastructure as a Service
PaaS
Platform as a Service
SaaS
Software as a Service
Applications
Data
Runtime
Middleware
O/S
Virtualization
Servers
Storage
Networking
bmc
You Manage
Other Manages

Pizza as a Service



Common Examples of IaaS, PaaS, SaaS

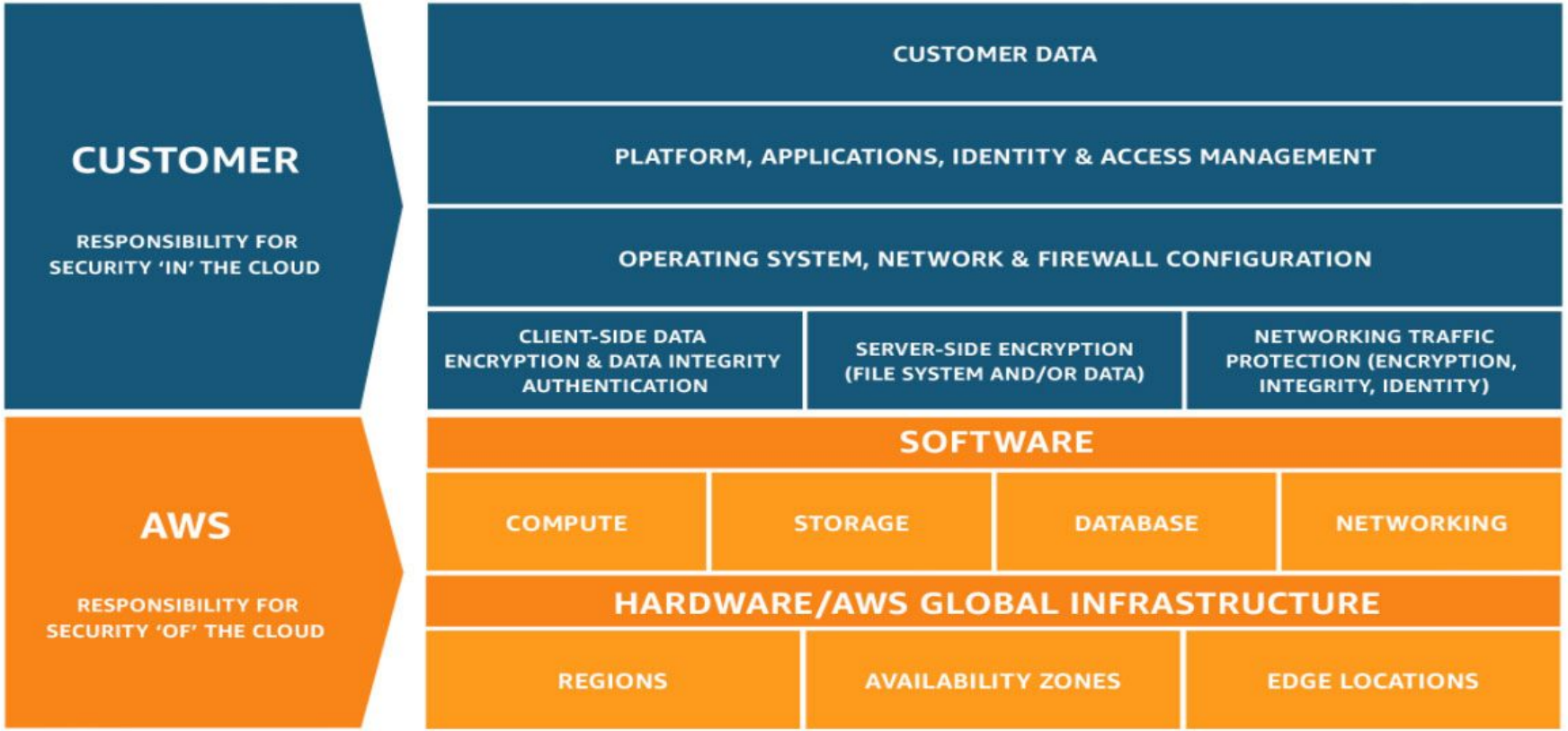
IaaS examples: AWS EC2, Google Compute Engine (GCE), Azure VM, Digital Ocean.

PaaS examples: AWS Elastic Beanstalk, Windows Azure (mostly used as PaaS),

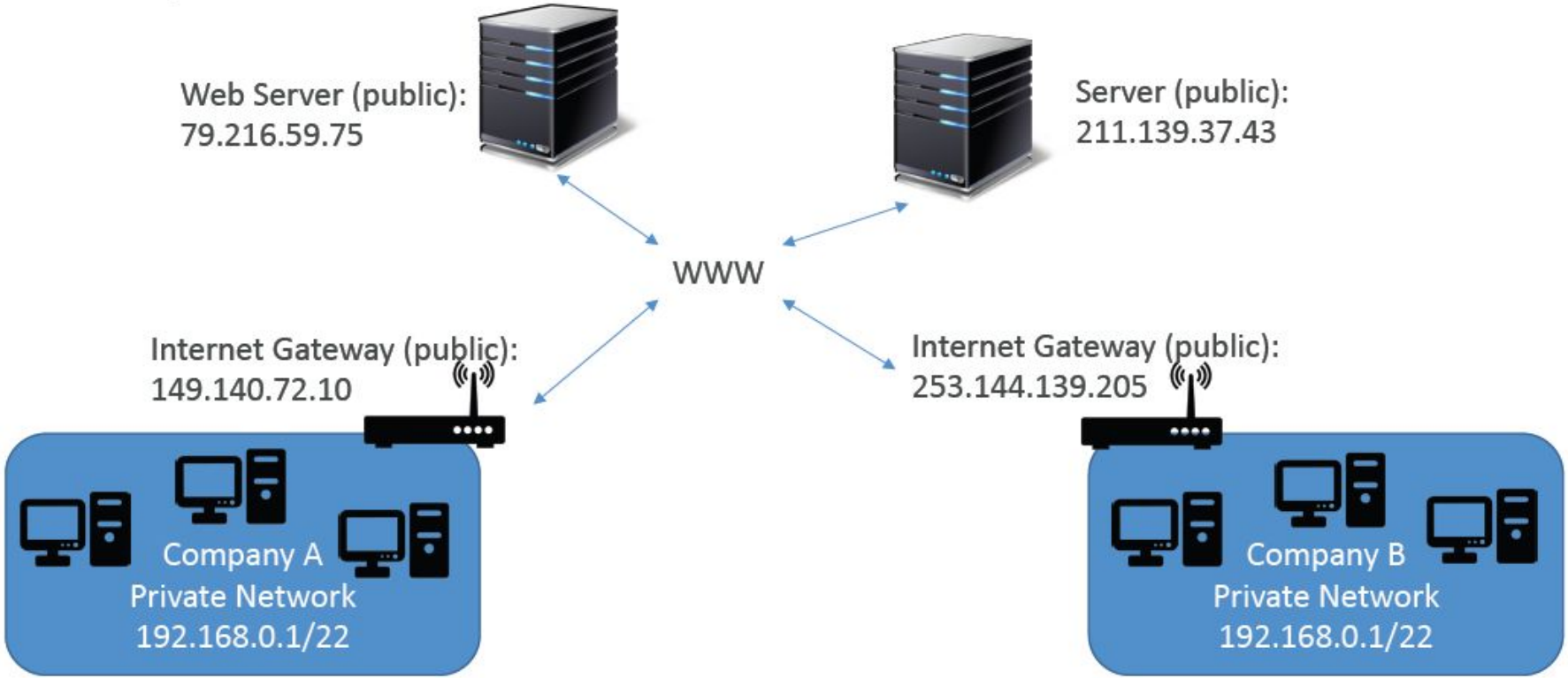
Force.com, OpenShift, Apache Stratos, Magento Commerce Cloud.

SaaS examples: Slack, DropBox, DocuSign, Salesforce, Google Apps

Shared Responsibility in AWS



Private vs Public IP (IPv4)



Cloud Terminology

Scalability & High Availability

- Scalability means that an application / system can handle greater loads by adapting.
- There are two kinds of scalability:
  - Vertical Scalability
  - Horizontal Scalability (= elasticity)

Cloud Terminology

Vertical Scalability

- It means increasing the size of the instance
- For example, your application runs on a t2.micro
- Scaling that application vertically means running it on a t2.large
- Vertical scalability is very common for non distributed systems, such as a database.

Horizontal Scalability

- It means increasing the number of instances / systems for your Application.
- Horizontal scaling implies distributed systems.
- common for web applications / modern applications

Cloud Terminology

High Available

- High Availability usually goes hand in hand with horizontal scaling
- High availability means running your application / system in at least 2 data centers (== Availability Zones)
- The goal of high availability is to survive a data center loss
- The high availability can be passive (for RDS Multi AZ for example)

QA Slide

- If we want to connect to a remote server, can it be connected using IP Address range 127.x.x.x ?