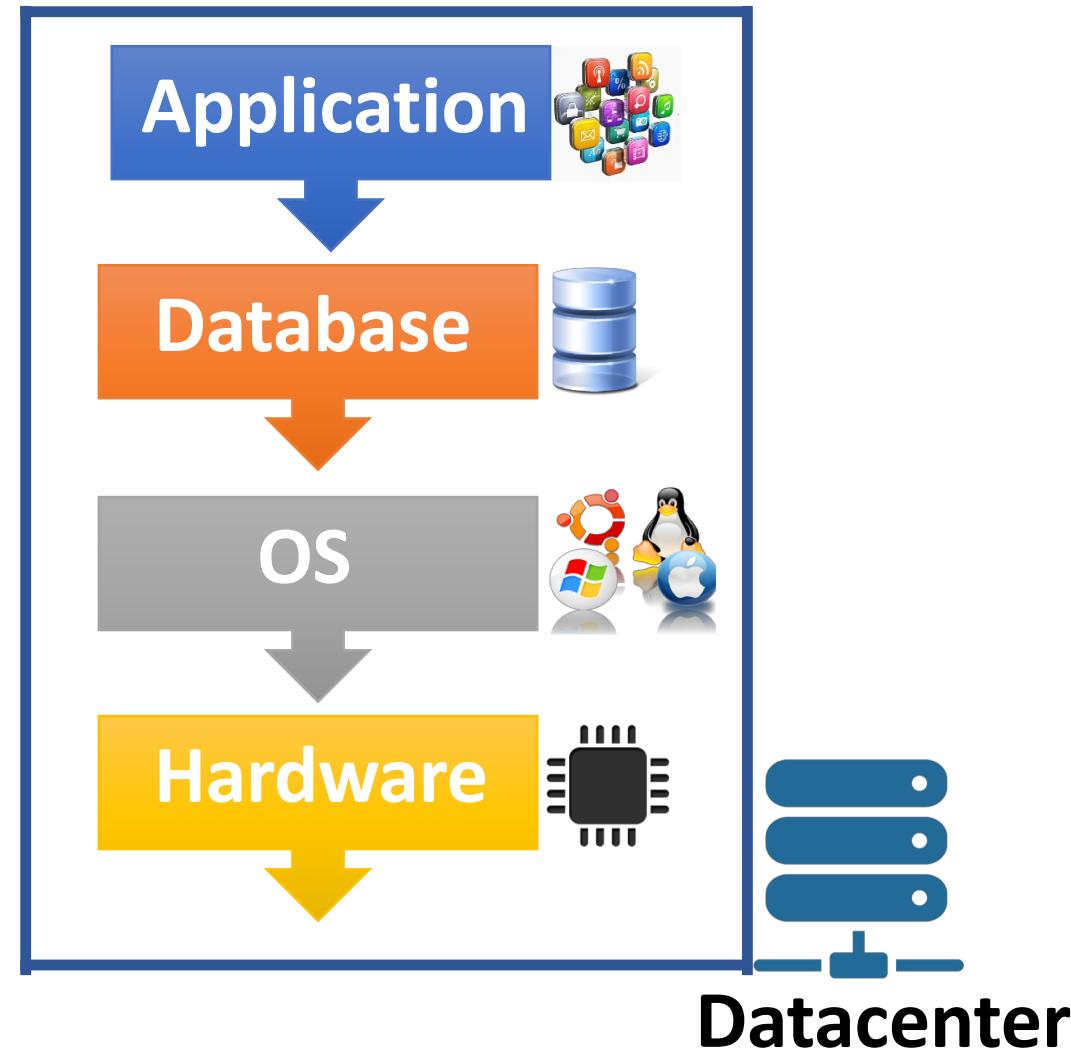
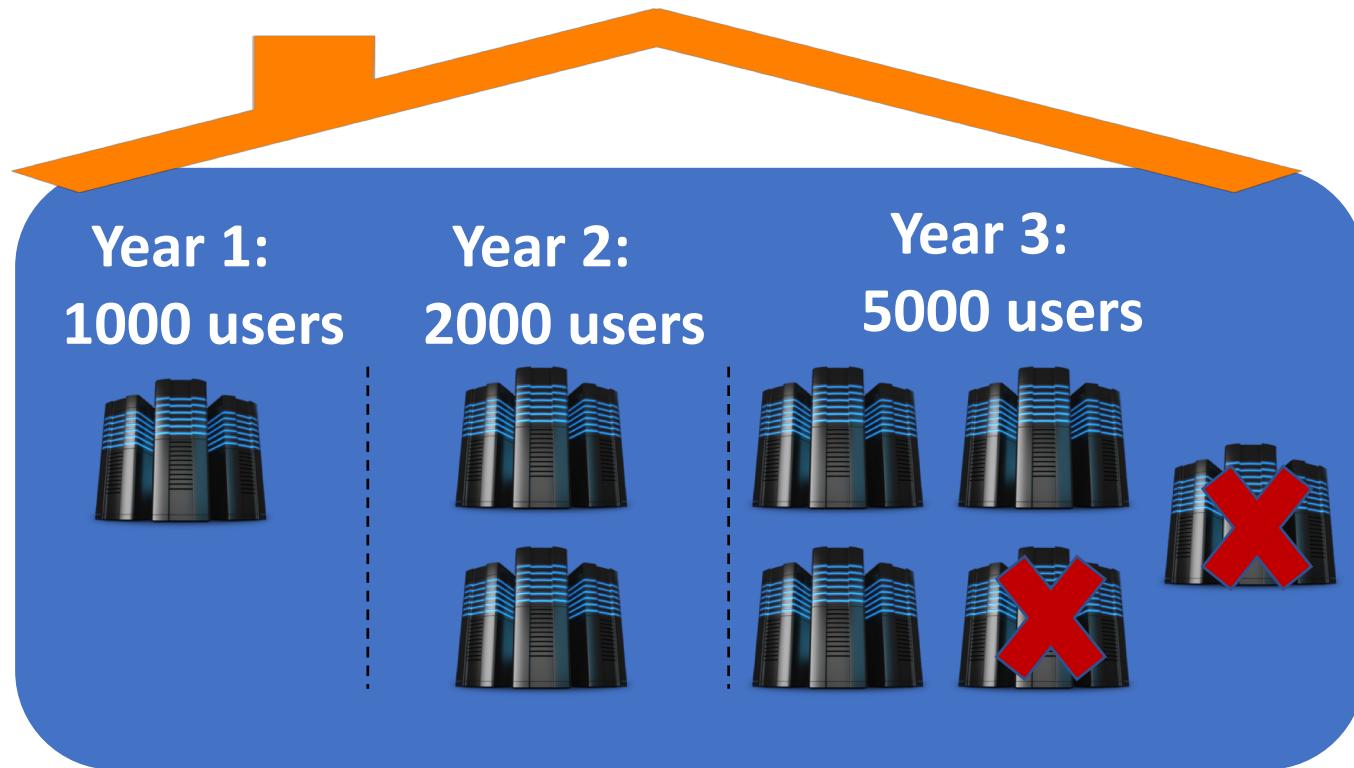




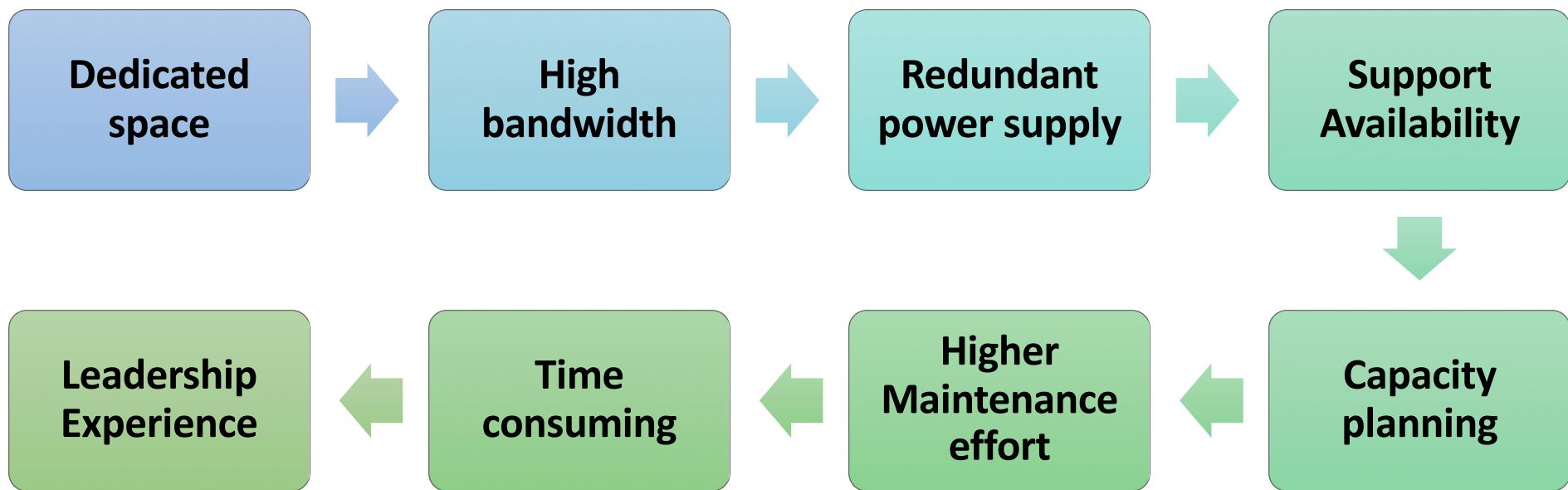
# AWS



# On-premise Infrastructure



# Datacenter Infrastructure Management



## **Business Requirements**

### **High Availability**

Creating the architecture in such a way that your system is always available

HA ensure that we can always access our data in the cloud

### **Fault Tolerant**

The ability of our system to withstand failures in one/more of its components & still remain available

FT ensure that if one of our web server failed, the backup server immediately took over

### **Scalability**

Scalability handles the changing needs of an application within the confines of the infrastructure via statically adding or removing resources to meet applications demands if needed.

### **Elasticity**

Elasticity is the ability to grow or shrink infrastructure resources dynamically as needed to adapt to workload changes in an autonomic manner, maximizing the use of resources. This can result in savings in infrastructure costs overall.

# AWS Goal: Flip This Equation



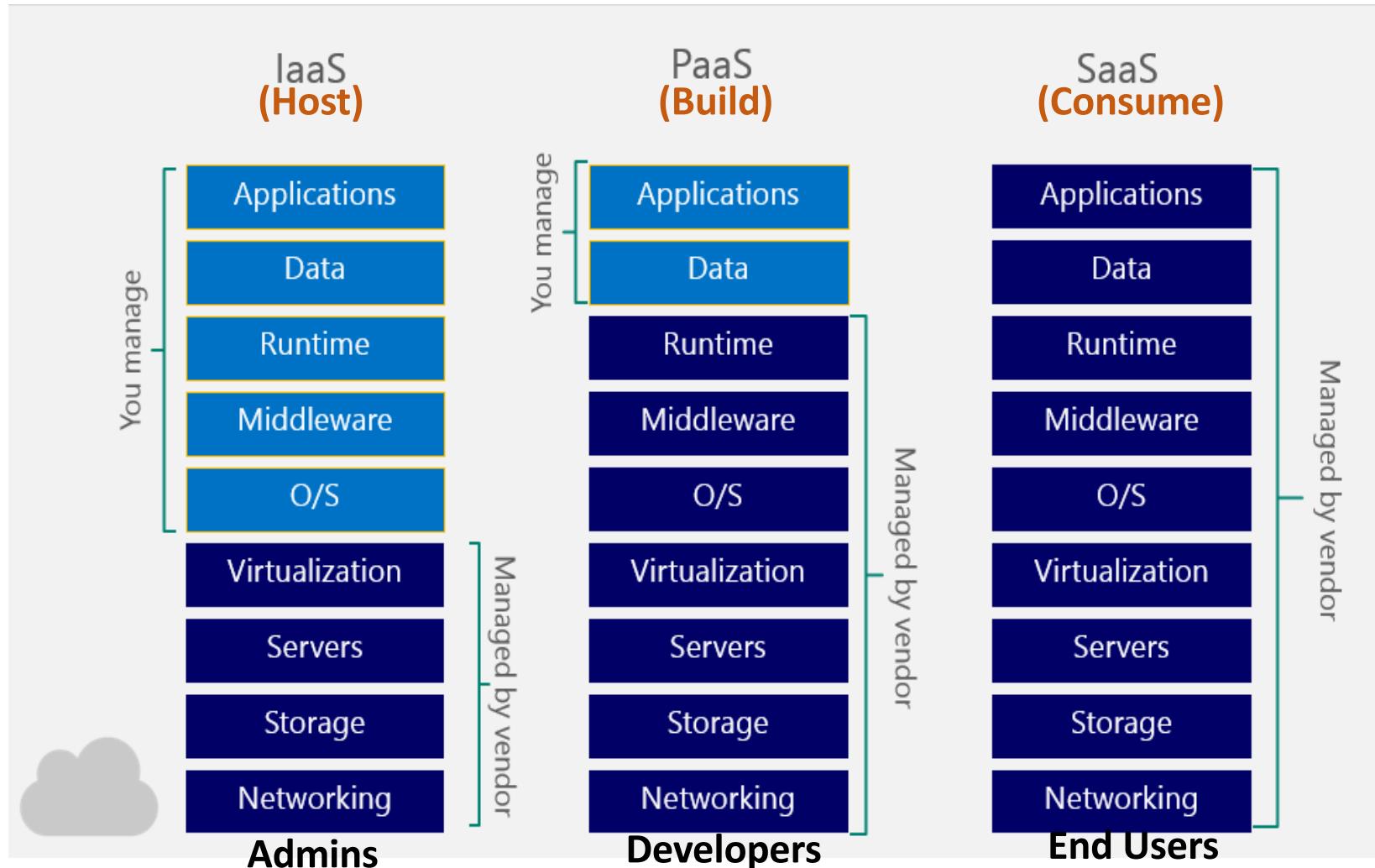
## **What is Cloud Service ?**

A **cloud service** is any **service** made available to users on demand via the Internet from a **cloud computing** provider's servers

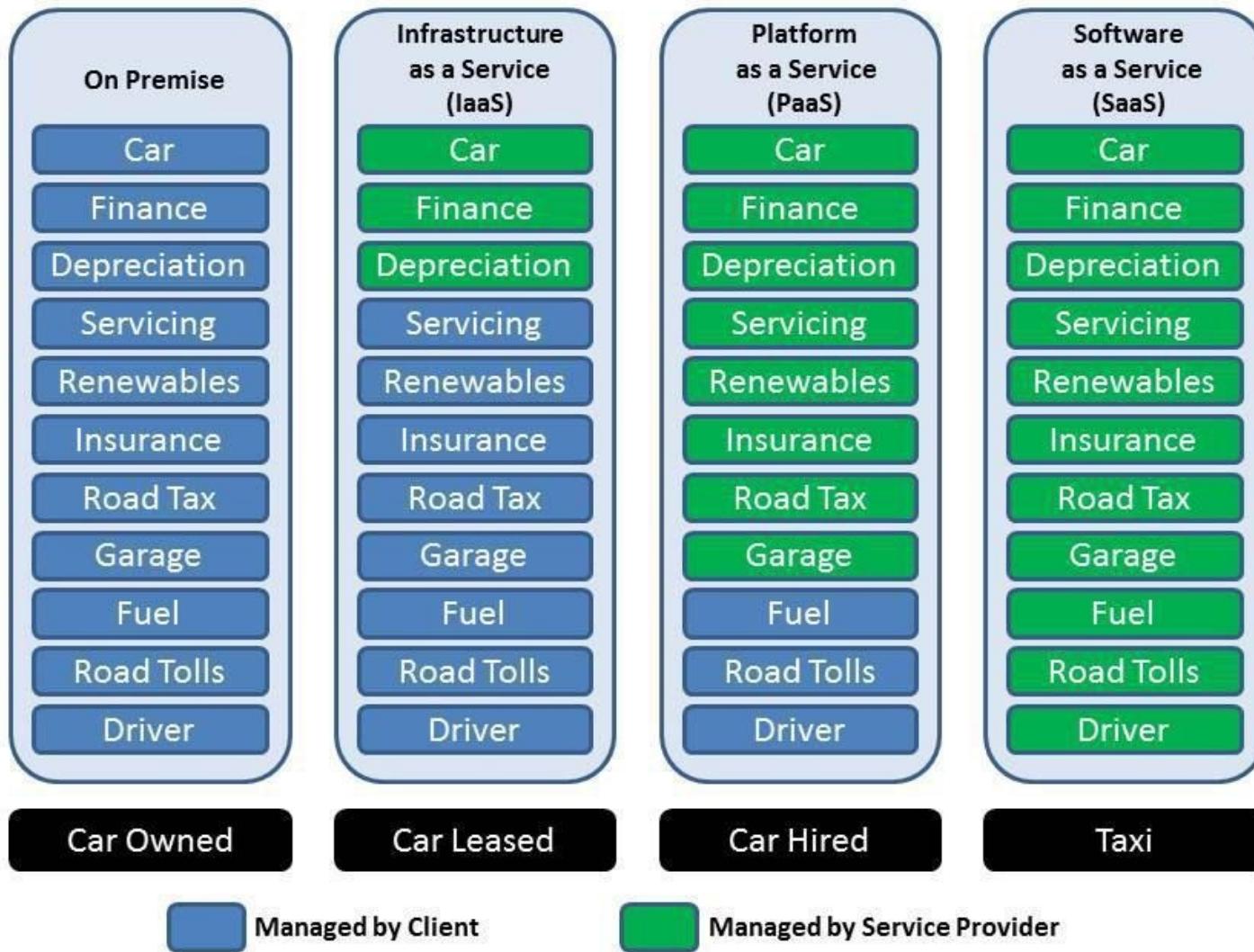
## **What is Cloud Computing ?**

**cloud computing** is the delivery of computing services-servers, storage, databases, networking, software, analytics and more-over the Internet

# Cloud Service Models

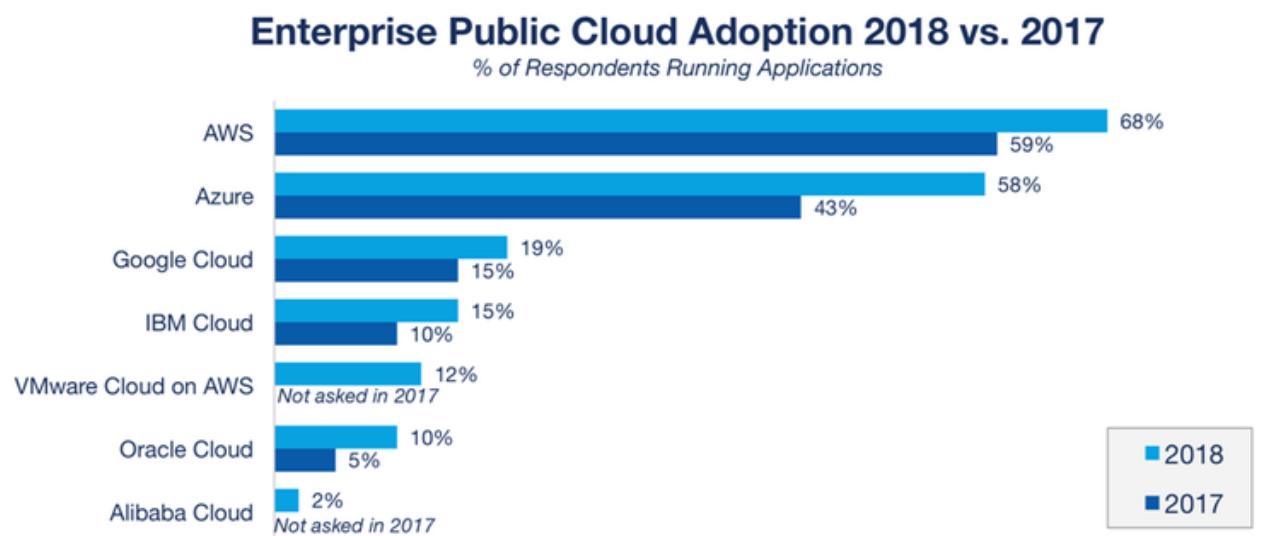


# Car as a Service



## Popular Cloud Providers

- Amazon Web Services
- Microsoft Azure
- Google Cloud Platform
- VMware
- IBM Cloud
- Oracle Cloud
- Rackspace
- Red Hat
- Salesforce



Source: RightScale 2018 State of the Cloud Report

RIGHTSCALE



**Amazon Web Services** is a subsidiary of Amazon.com that provides on-demand cloud computing platforms

- **Cost-Effectiveness – Pay as you Go**
- **Elasticity and Agility**
- **Flexibility and Openness**
- **Security**
- **Reliable & High Performance**

## AWS “Free Tier”

- AWS Free Tier refers to the limited free usage of AWS services
- AWS offers the Free Tier as means for a user to learn, experiment and get hands-on experience with AWS services
- Almost all AWS services offer some kind of Free Tier usage
- Free Tier is available for 12 months for an AWS account
- Some services extend past 12 months
- Free Tier is only available for new accounts

# Create an AWS account

URL: <https://aws.amazon.com>

- 1) Click on "Create a New AWS Account"
- 2) Enter your email address
- 3) Make sure "I am a new user" is selected
- 4) Complete the "Login Credentials" form
- 5) Select "Company" or "Personal" Account
- 6) Complete the rest of the "Contact Information" form
- 7) Complete the "Payment Information" form  
**You will need a validate credit/debit card**
- 8) Complete "Identity Verification" form & follow instructions to input the verification PIN number
- 9) Select a "Support Plan"  
**Select "Basic" for no fee/free tier use**
- 10) Sign in to the AWS Console with your newly created credentials

## AWS Global Infrastructure

The AWS Cloud spans 55 Availability Zones within 18 geographic Regions and 1 Local Region around the world, with announced plans for 15 more Availability Zones and five more AWS Regions in Bahrain, Cape Town, Hong Kong SAR, Stockholm, and a second GovCloud Region in the US.



### US East

N. Virginia (6),  
Ohio (3)

### US West

N. California (3),  
Oregon (3)

### Asia Pacific

Mumbai (2),  
Seoul (2),  
Singapore (3),  
Sydney (3),  
Tokyo (4),  
Osaka-Local (1)<sup>1</sup>



### Canada

Central (2)

### China

Beijing (2),  
Ningxia (3)

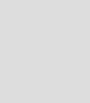
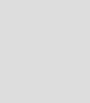
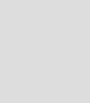
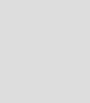
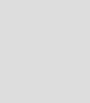
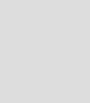
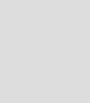
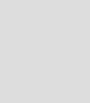
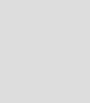
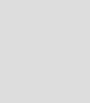
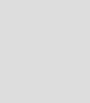
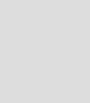
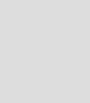
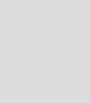
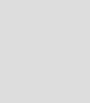
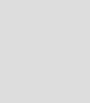
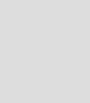
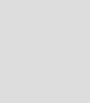
### Europe

Frankfurt (3),  
Ireland (3),  
London (3),  
Paris (3)

### South America

São Paulo (3)

### GovCloud (US-West) (3)



### New Region (coming soon)

Bahrain  
Cape Town  
Hong Kong

Stockholm  
GovCloud  
(US-East)



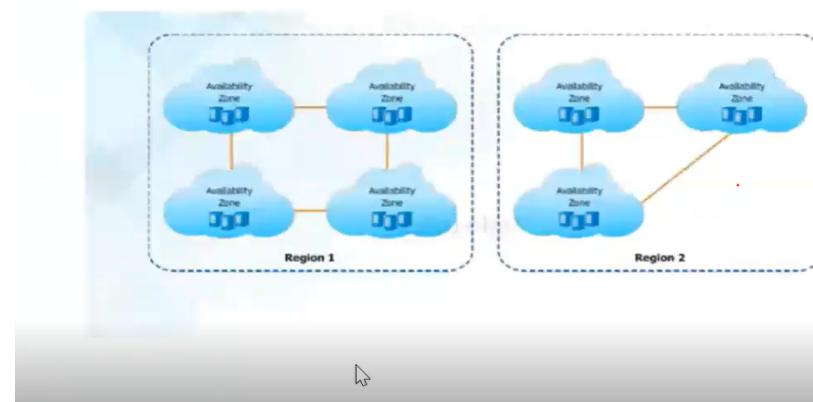
# Regions

- Amazon services are hosted in multiple locations world-wide.
- These locations are composed of regions(Logical Name) and Availability Zones(Physical).
- Each *region* is a separate geographic area.
- Each region has multiple, isolated locations known as *Availability Zones*.

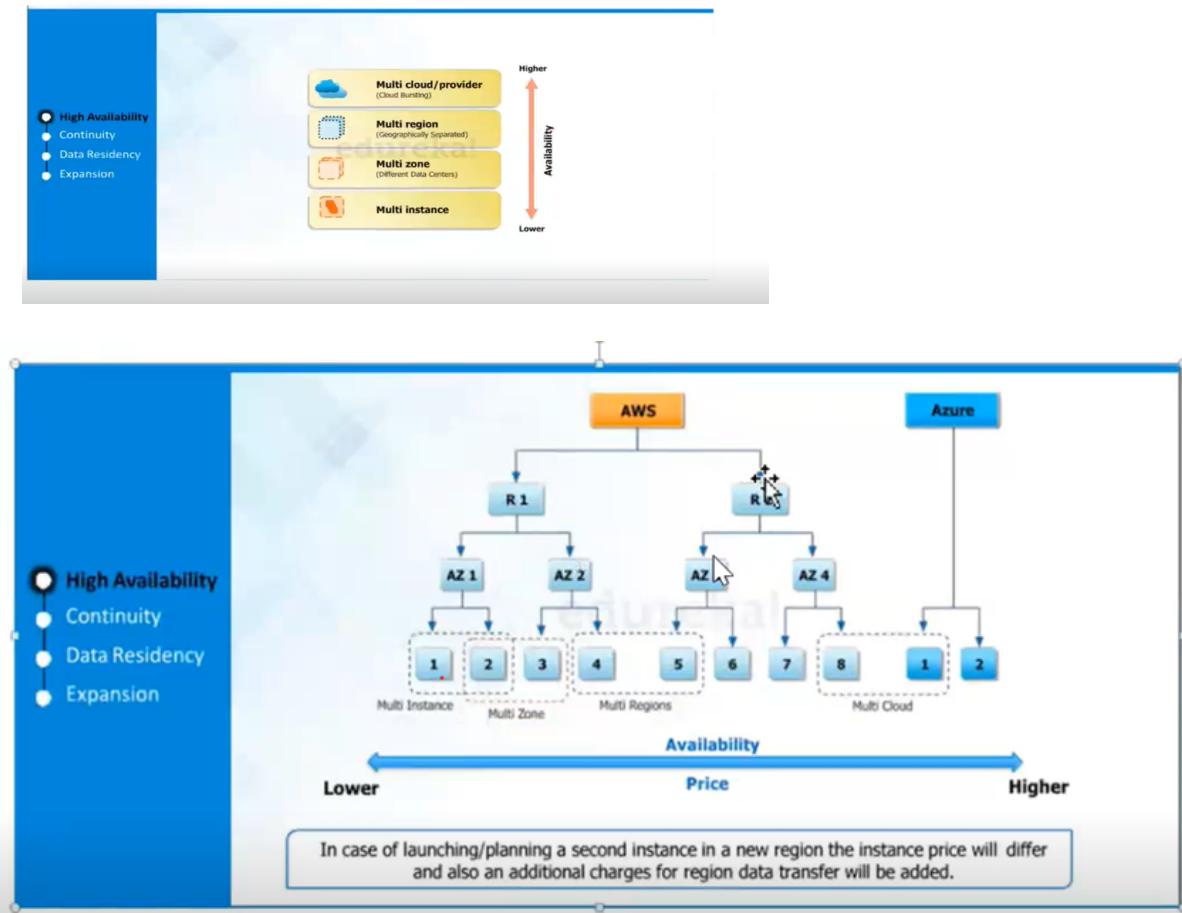
# Availability Zone

- Availability Zone are distinct locations that are engineered to be isolated from failures in other AZ
- By launching instances in separate Availability zones, we can protect our application from the failure of a single location.
- AZ is a local data center in a region & it can be a collection more than one data center.

Graphical view of Regions and AZ's



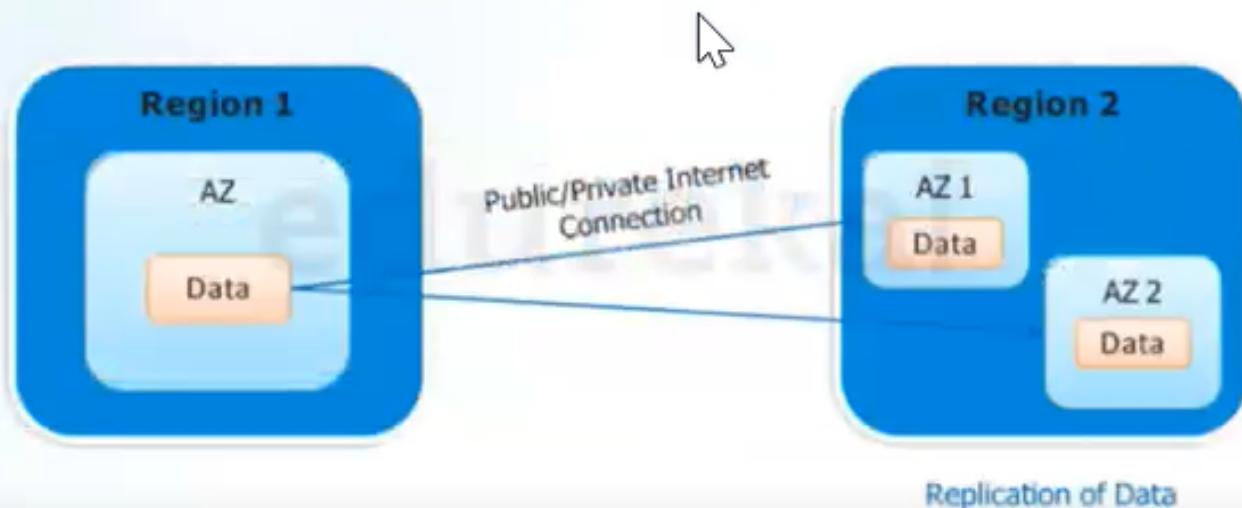
## 1. High Availability (HA)



## 2. Continuity (Latency)

- Can increase redundancy and fault tolerance by replicating data between geographic regions
- Provides low latency access(continuity) across the globe

- High Availability
- Continuity
- Data Residency
- Expansion



### 3. Data Residency

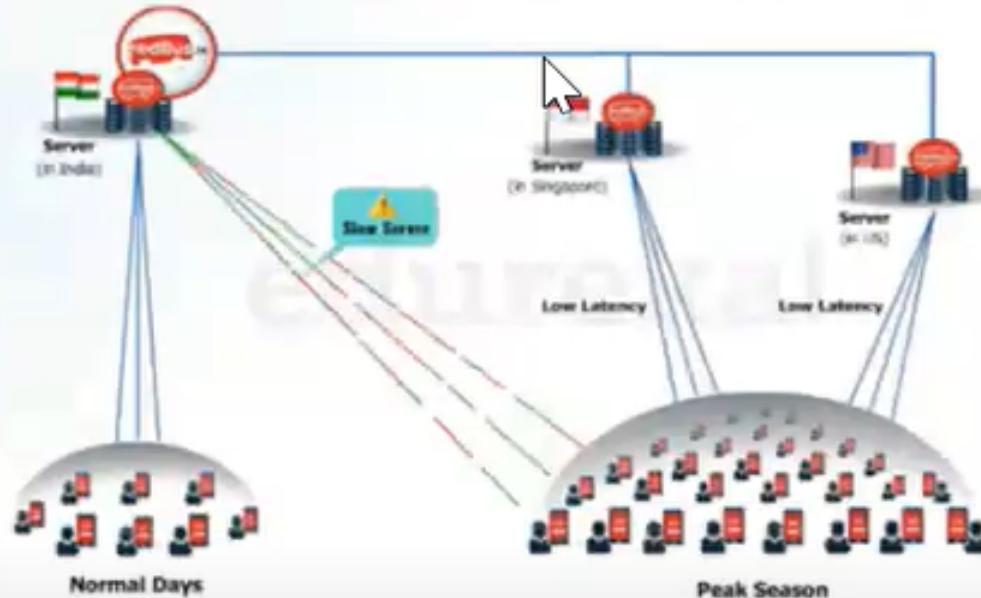
→ Complete control and ownership over the region, where user data is physically located



- High Availability
- Continuity
- **Data Residency**
- Expansion

## 4. Expansion (Elasticity)

- To provide flexible and low latency services to all the users, Amazon provides the flexibility of expanding the server globally
- Example: To provide better service to outside users during peak seasons, redBus started expanding globally to new geographic locations such as Singapore and USA using AWS.



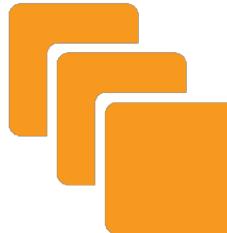
- High Availability
- Continuity
- Data Residency
- **Expansion**

# How to choose the right region...

---



## EC2 Basics



- **Elastic Compute Cloud is your computer in the cloud**
- **EC2 provides scalable computing capacity in the AWS cloud**
- **Use EC2 to launch as many or as few virtual servers as you need, configure security, networking & manage storage.**
- **EC2 enables you to scale up or down to handle changes in requirements**

# Conceptually understanding EC2

<b>Basic Computer components:</b>
Operating System
CPU
Hard Drive
Network Card
Firewall
RAM

<b>EC2 Instance components:</b>
AMI's
Instance type
EBS(local storage)
IP Addressing
Security Groups
RAM

# EC2 Instance Purchasing Options

## On-Demand:

- **Most expensive purchasing option**
- **Most flexible purchasing option**
- **You are charged only when instance is Running(billed by hour)**
- **You can provision/terminate an instance anytime**

## Reserved:

- **Allows us to purchase an instance for a set time period (1/3 yrs)**
- **Significant price discount**
- **Once you buy a reserved instance, we are responsible for the entire price - regardless of how often we use it**

## Spot:

- **Amazon sells the unused instances, for short amount of time at lower price**
- **We can Bid on an instance type & only use when the spot price is equal to or below your bid price**
- **Charged by hour**
- **Spot price fluctuate based on supply & demand in market**

# **How are we charged for using EC2 ?**

## **1) Purchasing Options:**

- on-Demand**
- Reserved**
- Spot**

## **2) Instance Type:**

- General Purpose**
- Compute Optimized**
- GPU Optimized**
- Memory optimized**
- Storage Optimizes**

## **3) EBS Optimized (Option for higher IOPS performance)**

## **4) AMI Type (price varies on distribution/software packages)**

- Linux**
- Windows**

## **5) Data Transfer (in/out of the instance)**

## **6) Regions**

# AMI - Amazon Machine Image



- A preconfigured package required to launch an EC2 Instance; includes an Operating system, software packages & other settings
- AMI provides the information required to launch an instance, which is a virtual server in the cloud
- We can launch as many instances from the AMI as you need

## **Step1: Selecting an AMI**

- When you launch an EC2 Instance, the first thing you do is select an AMI
- AMIs come in 2 main categories:
  - 1) Community AMIs:
    - Free to use
    - Generally it contains only the OS
  - 2) AWS Marketplace AMIs:
    - pay to use
    - generally comes packaged with additional licensed software
  - 3) My AMIs:
    - AMIs that you can create yourself

## Step2: Instance Type?

- **Instance type is the CPU (compute power) of your instance**
- **When you launch an instance, the instance type determines the hardware of the host**
- **Each instance type offers different compute, memory & storage capabilities**
- **Select an instance type based on the requirement of the software that you plan to run on your instance**

Instance Type Components:

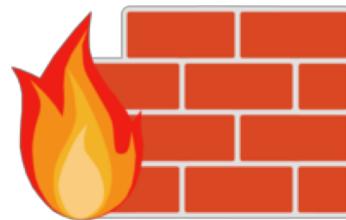
- **Family:** Categorizing instance types based on what they are optimized for
- **Type:** subcategory for each family type
- **vCPUs:** number of virtual CPUs the instance type uses
- **Memory:** Amount of RAM the instance type uses
- **Instance Storage(GB):** local instance storage volume(hard drive)
- **EBS-Optimized Available:** Indicates if EBS-optimization is an option for the instance type
- **Network Performance:** Rating based on its data transfer rate(bandwidth)

Families	Description	Example Use Cases
t2, m4, m3	<b>General Purpose</b> Balanced Performance	Websites, web applications, Dev, code repos, micro services, business apps
c3, c4, cc2	<b>Compute Optimized</b> High CPU Performance	Front-end fleets, web-servers, batch processing, distributed analytics, science and engineering apps, ad serving, MMO gaming, video-encoding
g2, p2	<b>GPU Optimized</b> High-end GPU	Amazon AppStream 2.0, video encoding, machine learning, high perf databases, science
r3, r4, x1, cr1	<b>Memory Optimized</b> Large RAM footprint	In-memory databases, data mining
d2, i2, i3, hi1, hs1	<b>Storage Optimized</b> High I/O, High density	NAS, data warehousing, NoSQL

EC2 Instance Type	Compute Optimized		General Purpose			Memory Optimized			Storage Optimized		
	C5	C4	M5	M4	T2	X1	X1e	R4	H1	I3	D2
Intel Processor	Xeon Platinum 8175M	Xeon E5 2666 v3	Xeon Platinum 8175M	Xeon E5 2686 v4 2676 v3	Xeon Family	Xeon E7 8880 v3	Xeon E7 8880 v3	Xeon E5 2686 v4	Xeon E5 2686 v4	Xeon E5 2686 v4	Xeon E5 2676 v3
Intel Processor Technology	Skylake	Haswell	Skylake	Broadwell Haswell	Yes	Haswell	Haswell	Broadwell	Broadwell	Broadwell	Haswell
Intel AVX	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Intel AVX2	Yes	Yes	Yes	Yes	-	Yes	Yes	Yes	Yes	Yes	Yes
Intel AVX-512	Yes	-	Yes	-	-	-	-	-	-	-	-
Intel Turbo Boost	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Storage	EBS-only	EBS-only	EBS-only	EBS-only	EBS-only	SSD EBS-Opt	SSD EBS-Opt	-	HDD	SSD	HDD

# Security Groups ?

- Firewall is a network security system designed to prevent unauthorized access to/from a private network
- Security groups acts as a virtual firewall that controls the traffic for one or more instances
- We add rules to each SG that allow/deny traffic from its associated instances
- Best practice is to allow only traffic that is required



Firewall



Security

Security Groups

# IP Addressing ?

- Similar to having home street address to send mail
- IP address is the instances address on the network

## Private IP:

- By default every EC2 instance will be provided with a private IP address
- Private IP addresses allow instances to communicate as long as they are located in the same VPC

## Public IP:

- EC2 instances can be launched with/without public IP address
- Public IP address is required for the instance to communicate with the network

## Elastic IP:

- Static public IP address for the instance.
- Chargeable for each elastic IP.

## **Launching an EC2 Instance:**

- **Select an AMI**
- **Select an Instance Type**
- **Configure Instance Details:**
- **Add Storage**
- **Add a Tag ( give the instance a name )**
- **Configure/Assign a Security Group**
  - Create a new security group**
- **Review & Launch**
- **Create a new Key pair & Download it.**

## **Connecting to an EC2 Instance(Linux/SSH)**

- **Select the instance**
- **Under "Actions", choose "connect"**
- **Follow these in order**
  - ✓ **Open a terminal to access the cmd line**
  - ✓ **Navigate into the dir that contains the Key pair we downloaded**
  - ✓ **Run the chmod on Key pair**
  - ✓ **Run commands**

Questions ?

