# Cloud computing;

Computing resources are made available as a service

**Important Chars for Cloud;**

-On Demand and Self-service {Any time launch without manual Intervention}

-Elasticity (Scalability) {can scale up/down any time provided when it’s stopped}

-Measured Service {Pay for what you used}

Cloud computing models;

Software as service (SaaS)

Platform as service (PaaS)

Infrastructure as service (Iaas)

**High Level Architecture of Cloud computing;**

Virtual Servers | On Demand and Self service

-------- |

virtualization | -------------------->>>>>>> Elasticity

-------- |

Data Center | Pay per use

**\*\*Scalability\*\*;**

**--Horizontal Scalability**

If you need to upgrade 2 GB RAM to 8GB capacity, include 3 more server with 2GB RAM

**Cons;**

Can't be used in Database applications. Vertical scalability is preferred.

--**Vertical Scalability**

If you need to upgrade 2 GB RAM to 8GB capacity, it is done in a single server.

**Cons;**

If server goes down, it will stop promotions (business) Horizontal scalability is refereed.

**AUTO SCALING;**

-Auto scaling functionality for horizontal scalability is one of the AWS which can scale up/down based on the CPU requirement with threshold values.

-Max 5 servers/instances

\*\*\*CANNOT SCALE THE SERVERS WITHOUT SHUTTING DOWN SERVER IF DOING VERTICALLY\*\*

**Characteristics of CSP (Cloud service provider);**

Automated creating instance using GENKIN's job.

**On Demand Challenges;**

Limited resource for a Data center. To avoid 'insufficient capacity' we need to have reserved servers suggested by CSP.