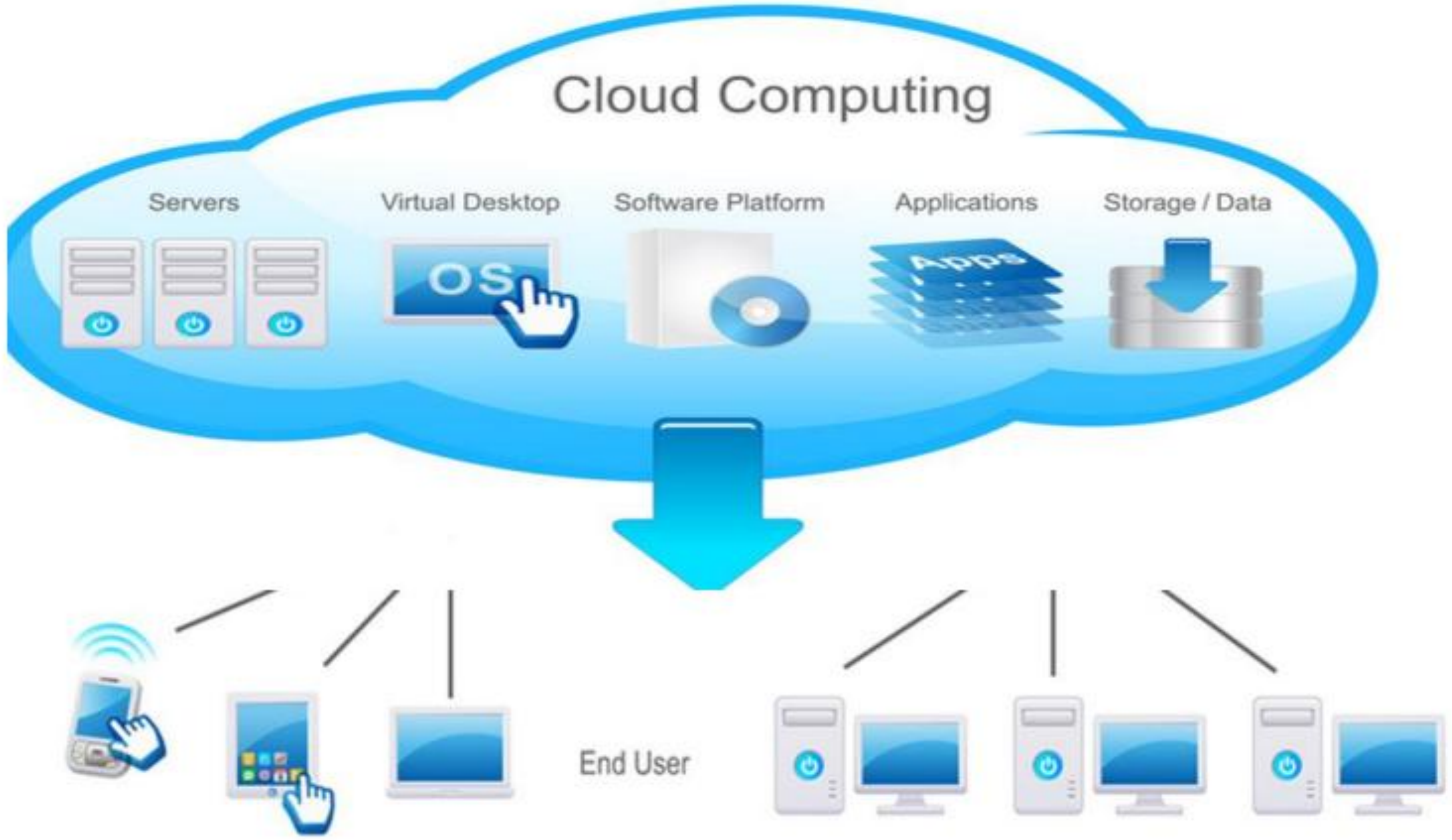
howest
hogeschool

Cloud concepts

Chris Roets
Guy Van Eeckhout

What is Cloud

Cloud computing is a service provisioning technique where computing resources like hardware such as servers and storage devices, software's and complete platform for developing applications are provided as a service by the cloud providers to the customers.

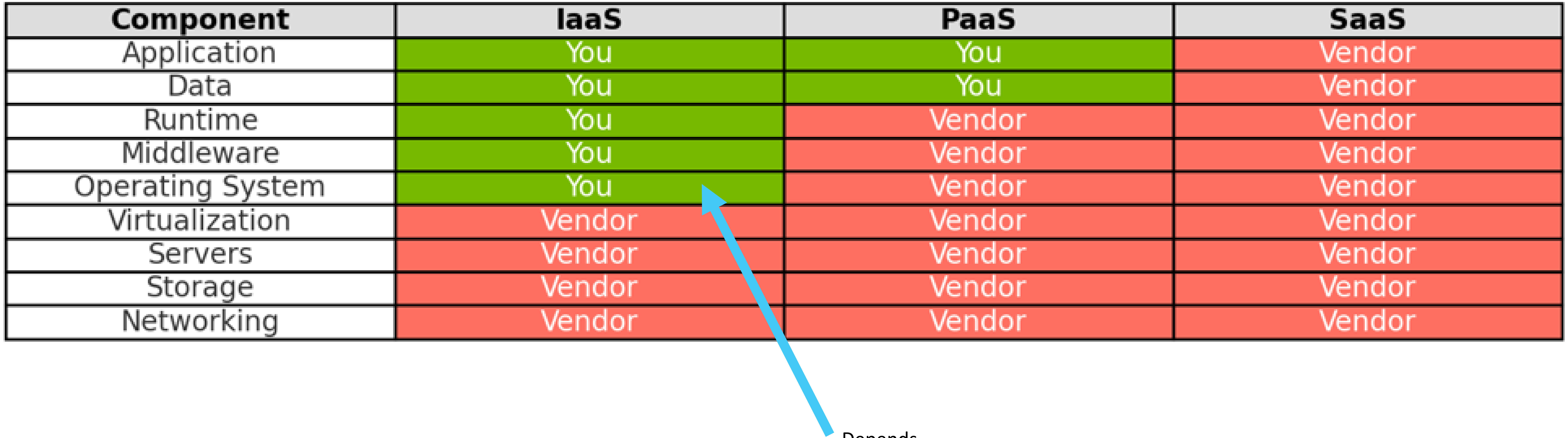


Advantages of Cloud computing

- Customers can use these resources as and when needed, can increase or decrease resource capacities dynamically according to their requirements and pay only for how much the resource were used.
- Customers do not need to invest money to purchase, manage and scale infrastructures, software upgradation and software licensing.

Different types

- The services that are provided by the cloud providers are broadly classified into three categories:
  - IAAS
  - PAAS
  - SAAS
- From deployment point of view, they can be seen as
  - Public Cloud
  - Private Cloud
  - Hybrid Cloud
  - Multi Cloud



Component	IaaS	PaaS	SaaS
Application	You	You	Vendor
Data	You	You	Vendor
Runtime	You	Vendor	Vendor
Middleware	You	Vendor	Vendor
Operating System	You	Vendor	Vendor
Virtualization	Vendor	Vendor	Vendor
Servers	Vendor	Vendor	Vendor
Storage	Vendor	Vendor	Vendor
Networking	Vendor	Vendor	Vendor

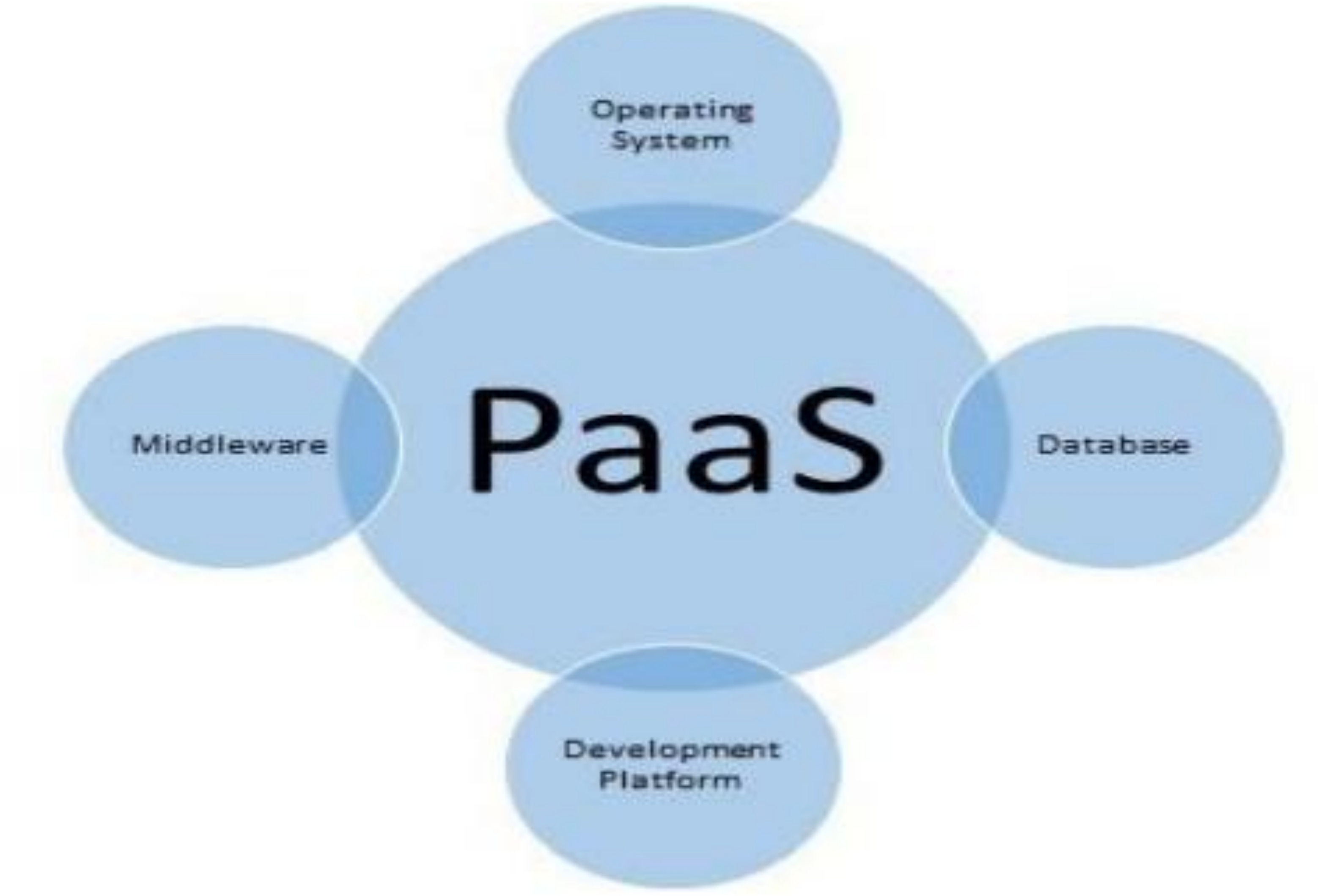
Depends

Infrastructure as a service

- the service provider owns the hardware equipment's such as Servers, Storage, Network and is provided as services to the clients. The client uses these equipment's and pays on per-use basis.
- E.g. Amazon Elastic Compute (EC2) and Simple Storage Service (S3)

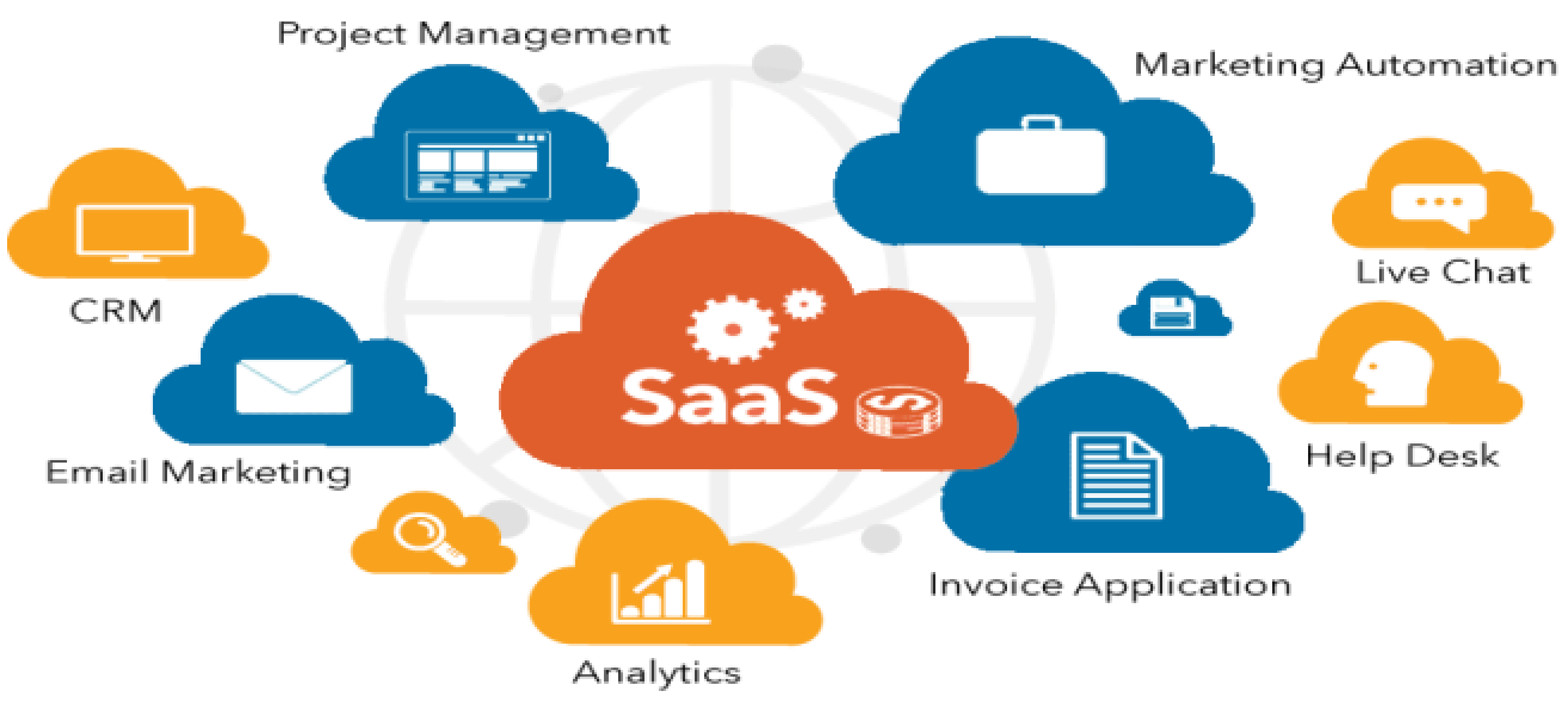
Platform as a service

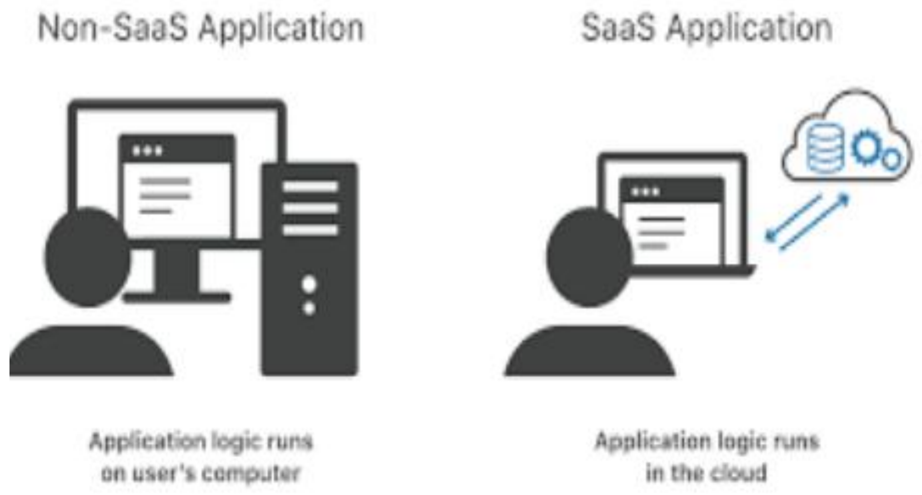
- complete resources needed to Design, Develop, Testing, Deploy and Hosting an application are provided as services without spending money for purchasing and maintaining the servers, storage and software.
- PaaS is an extension of IaaS. In addition to the fundamental computing resource supplied by the hardware in an IaaS offering, PaaS models also include the software and configuration required to create an applications.
- E.g. Google App Engine.



Software as a service

- the service provider provides software's as a service over the Internet, eliminating the need to buy, install, maintain, upgradation and licensing on their local machine.
- In stead of running the app on you own pc, it runs "on the internet"
- E.g. Accounting, CRM, Google Docs Onedrive, office.com, etc…




Non-SaaS Application
SaaS Application
Application logic runs on user's computer
Application logic runs in the cloud

Public Cloud

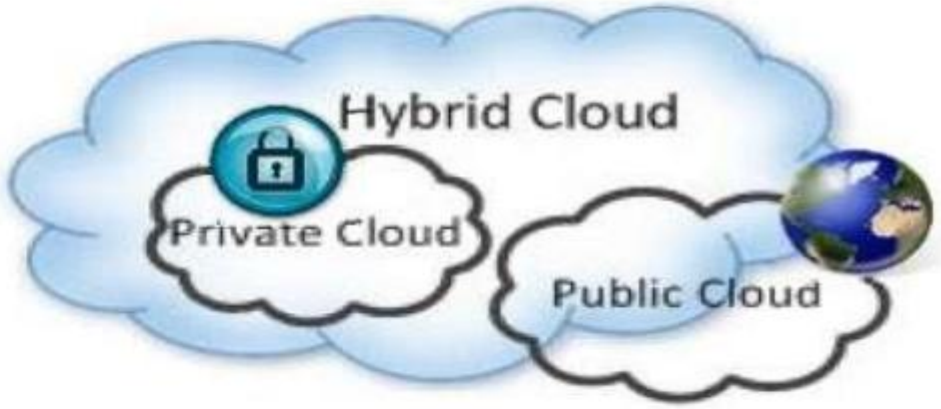
- A public cloud is a cloud in which services and infrastructure are hosted off-site by a cloud provider (owned by an organization selling cloud services) and easily accessible to general public via internet.

Private Cloud

- Private Cloud is a cloud where services and infrastructure are operated for a single operation accessible via private network, managed internally or by a third party. It is greater level of security

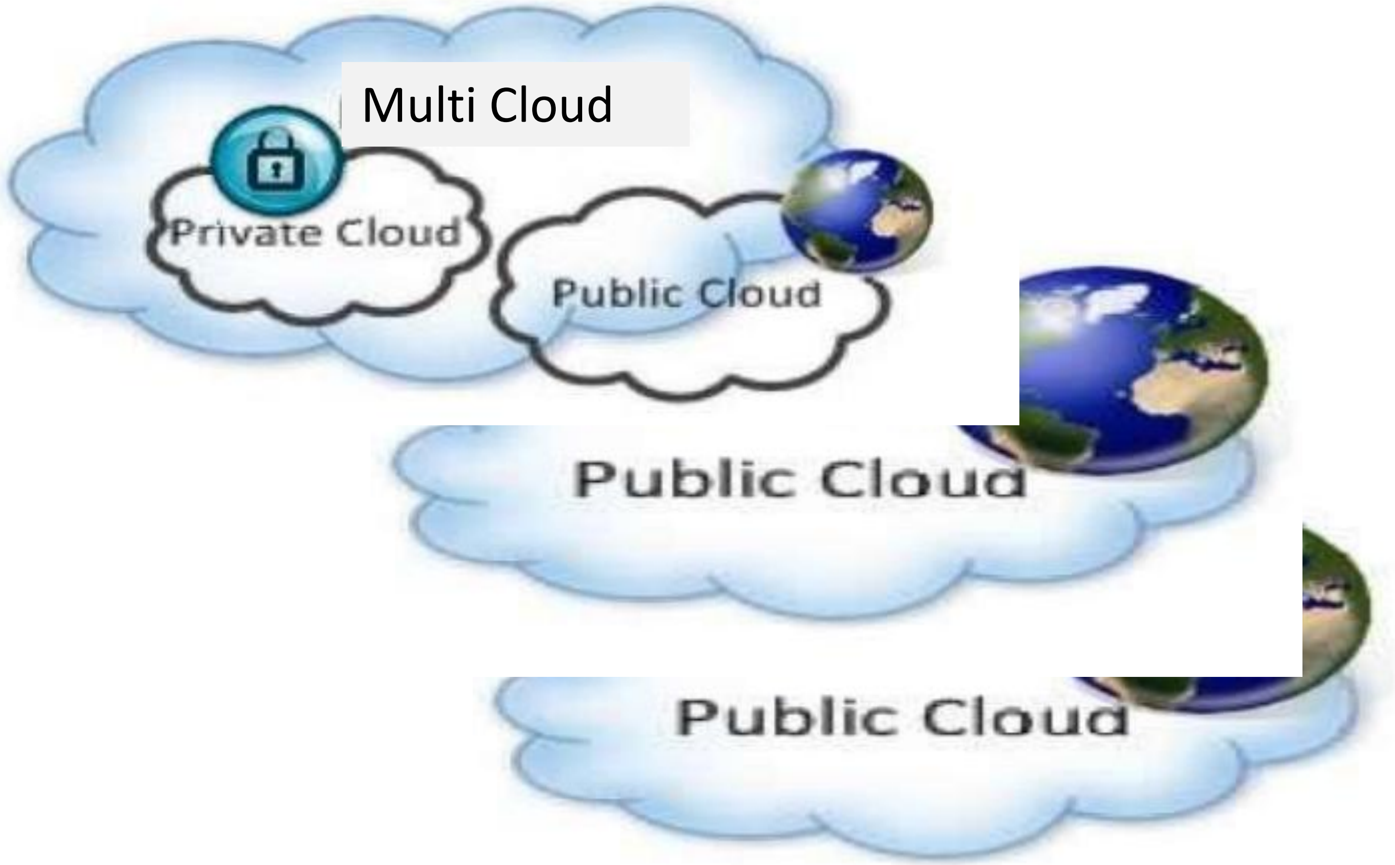
Hybrid Cloud

- Hybrid Cloud is a cloud which is a mixture of private and public cloud.
- In this type of cloud all critical and sensitive applications and data are stored in private cloud and non critical and non sensitive applications and data are stored in public cloud



Multi Cloud

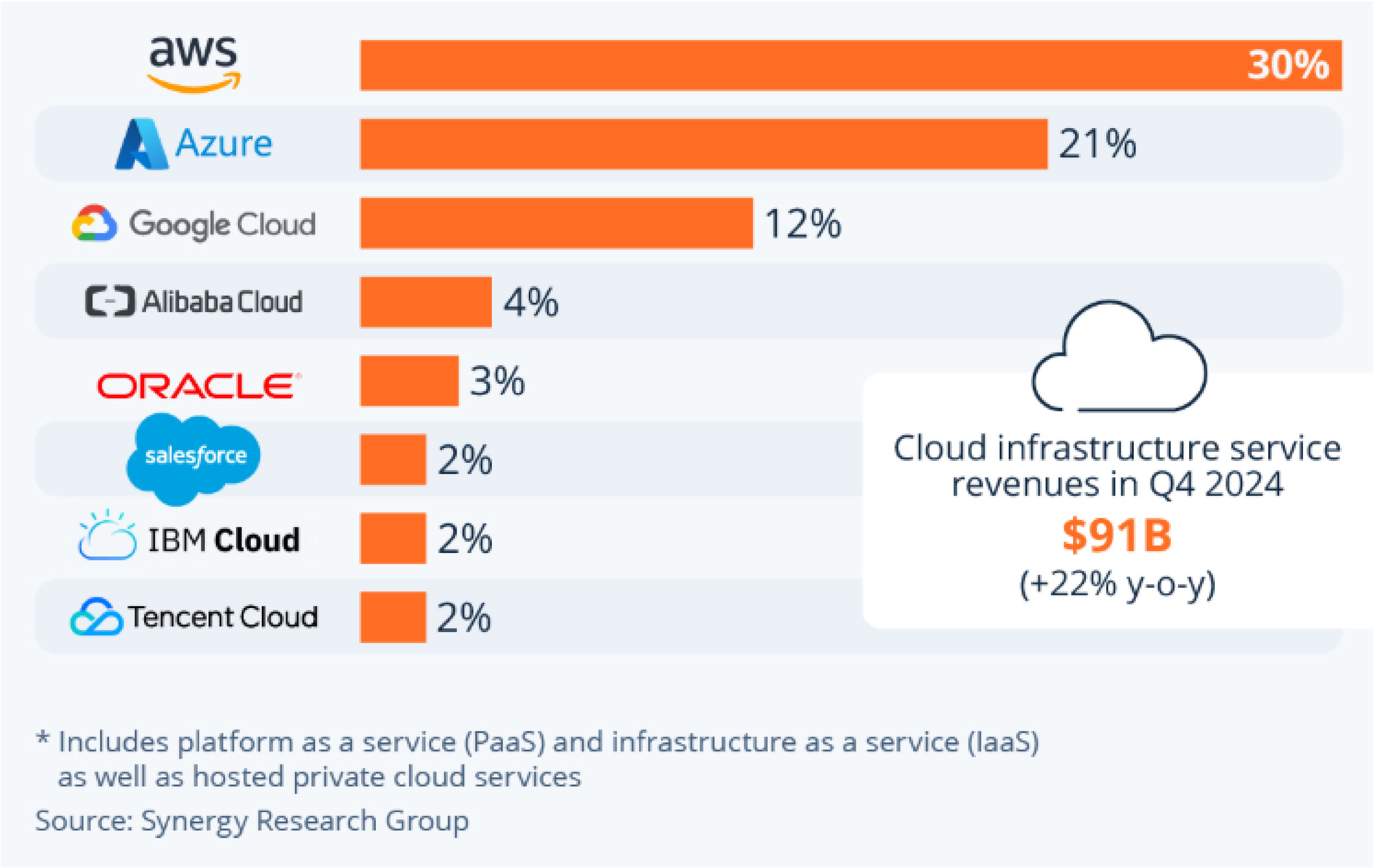
- Multi Cloud is a cloud which is a mixture of multiple private and public cloud.
- This is the most commonly used nowadays for SAAS.



Features of Cloud

- It is elastic: Cloud computing is flexible in nature, where users can scale up and scale down the resources as needed.
- Pay per use: Usage is metered and user pays only for how much the resources were used.
- Operation: The services are completely handled by the provider.
- Reduce capital cost: No need to invest money on purchasing and maintaining of hardware and software, software licensing, training required for IT staff.
- Remote accessibility: Users can access applications and data stored on cloud from anywhere any time worldwide through a device with internet connection.
- Better use of IT staff: Staff with in enterprise need not worry on purchasing and maintaining of servers, softwares, up gradation of servers and softwares, software licensing etc., instead they can concentrate more on work

Most popular cloud providers



* Includes platform as a service (PaaS) and infrastructure as a service (IaaS) as well as hosted private cloud services

Source: Synergy Research Group

Cloud concepts and technology

- Virtualization
- Service portal
- Load Balancing
- Scalability and Elasticity
- Deployment
- Replication and backup
- Monitoring
- Software Defined Networking
- Identity and Access Management
- Service Level Agreements
- Billing
- Security

Virtualization

- Virtualization refers to the partitioning the resources of a physical system (such as computing, Storage, Network and Memory) into multiple virtual resources.
- In cloud computing, resources are pooled to serve multiple users using Multi-Tenancy.
- Multi-Tenant aspects of the cloud allow multiple users to be served by the same physical hardware.
- The physical resources such as computing, storage, memory and network resources are virtualized.
- The virtualization layer partitions the physical resources into multiple virtual machines.

Service portal

- A **cloud service portal** is a web-based interface that allows users to access, manage, and interact with cloud services provided by a cloud service provider. It's essentially the **dashboard or control panel** through which users can
  - Provision Resources
  - Monitor Usage & Performance
  - Manage Accounts & Permissions
  - Automate Tasks

Load Balancing

- One of the important features of cloud computing is scalability.
- Cloud resources can be scaled up on demand to meet the performance requirements of applications.
- Load balancing distributes workload across multiple servers to meet the application workloads.
- The goal of load balancing techniques are to achieve maximum utilization of resources, minimize the response times, maximizing throughput.
- Since multiple resources under a load balancer are used to serve the user requests, in the event of failure of one or more of the resources, the load balancer can automatically reroute the user traffic to the healthy resource

Load Balancing Techniques

- In **round robin** load balancing, the servers are selected one by one in a circular fashion to server the incoming requests from the user.
- In **Weighted round robin** load balancing, servers are assigned some weights. The incoming requests are proportionally routed to a server based on its weight.
- In **low latency load balancing**, the load balancer monitors the latency of each server and request is routed to a server which has lowest latency.
- In **least connection load balancing**, the incoming requests are routed to the server with least number of connections.
- In **priority load balancing**, each server is assigned a priority. The incoming traffic is routed to the highest priority servers as long as the server is available. When the highest priority server fails, the incoming traffic is routed to a server with a lower priority.
- **Overflow load balancing** is similar to priority load balancing. When the incoming request to high priority servers overflow, the requests are routed to a lower priority server

Scalability and Elasticity

- Muti-tier applications such as e-Commerce, Social networking, business-to-business etc. can experience rapid changes in their traffic.
- Each website has a different traffic pattern which is determined by a number of factors that are generally hard to predict beforehand.
- Capacity planning involves determining the right sizing of each tier of the deployment of an application in terms of number of resources and capacity of each resource.
- Capacity planning may be for computing, storage, memory or network resources.
- Traditional approaches for capacity planning are based on predicted demands for applications and account for worst case peak load as of applications

Deployment

See service Architecture and deployment

Replication

- Replication is used to create and maintain multiple copies of the data in the cloud.
- Replication of data is important for practical reasons such as business continuity and disaster recovery.
- In the event of data loss at the primary location, organizations can continue to operate their applications from secondary data sources.
- Traditional business time objective (RTO).
- continuity and disaster recovery approaches don’t provide efficient, cost effective and automated recovery of data.
- Cloud based data replication approaches provide replication of data in multiple locations, automated recovery, low recovery point objective (RPO)and low recovery time.
- With cloud based data replication, organizations can plan for disaster recovery without making any capital expenditures on purchasing, configuring and managing secondary site locations

Backup

- Replication provides data redundancy
- But what if the data gets corrupted ?
  - Rm –r /
  - Drop table
  - Ransomware
  - ….
- Backup allows you to go back in time
- Backup is not standard in the offering and can be costly
  - See business case

Monitoring

Cloud providers provides monitoring service that allows cloud users can monitor their cloud resource usage.

A monitoring service at the cloud collects data on various system and application metrics from cloud computing instances.

Users can define various actions based on the monitoring data. For eg: Auto scaling when the CPU usage of the monitored resources becomes high.

Monitoring services also provides various statistics based on the monitoring data collected

Software Defined Networking

- SDN decouples the network and control forwarding functions. It separates the control plane (making traffic decision) from the data plane (packet forwarding)
  - Packet forwarding : layer 1 and 2
  - Control plane : layer 3 and 4

Identity and Access Management

Identity and Access Management (IDAM) for cloud describes the authentication and authorization of users to provide secure access to cloud resources.

Organization with multiple users can use IDAM services provided by the cloud service provider for management of user identifiers and user permissions.

IDAM services allow organizations to centrally manage users, access permission, security credentials and access keys.

IDAM services allow creation of user groups where all the users in a group have the same access permissions.

Identity and Management is enabled by a number of technologies such as OpenAuth, Role-based Access Control(RBAC), Digital Identities, Security Tokens, Identity Providers etc.

Service Level Agreement

- A Service Level Agreement (SLA) for cloud specifies the level of service that is formally defined as a part of the service contract with the cloud service provider.
- SLAs provide a level of service for each service which is specified in the form of minimum level of service guaranteed and a target level.
- SLAs contain a number of performance metrics and the corresponding service objectives.

Service Level Agreement

- Common criteria cloud SLAs
  - Availability
  - Performance
  - Disaster recovery
  - Incident response and resolution
  - Security and privacy of data
  - …..

Billing

- Cloud Service providers offer a number of billing models described as follows
  - pay as you use
  - Fixed pricing
  - Spot pricing
- More on this in the business case

Cloud security

- **Cloud Provider's Responsibility:** Understand that the cloud provider is responsible for securing the underlying infrastructure (physical servers, network, etc.).
- C**ustomer's Responsibility:** The customer is responsible for securing their data, applications, and access to their cloud resources.

reference

CLOUD COMPUTING

https://bmsce.ac.in/Content/CS/21CS7PECCT-CC-PPT-Oct-Jan_2022.pdf