Cloud Computing

Definitions :

Parallel Computer: is a collection of processing elements that communicate and cooperate to solve large problems fast (HPC), Processors are closely connected, all processors share the same memory, and the processors communicate by accessing this shared memory

Distributed System: A collection of **independent computers** that appears to its users as a **single coherent system** . This collection has **heterogeneous nodes** connected by one or more interconnection networks which provides access to system-wide shared **resources and services**. In addition that **hardware or software components** located at networked computers **communicate and coordinate** their actions only by **"message passing'** this message passing can be **synchronous or asynchronous** done by coordination between two or more **independent computers** , independent computers linked by a network with software designed to produce an integrated computing facility.

Middleware: Software that manages and supports the different components of a distributed system. In essence, it sits in the middle of the system

Centralized systems : System shared by users all the time which **All resources accessible** and Software runs in a single process , has single (physical location& point of control & point of failure**)**

Decentralizes systems : system has multiple autonomous components which some resources may not be accessible components shared by users ,has multiple(physical location& point of control & point of failure**)**

Service: manage a collection of related resources and present their functionalities to users and applications. the requesting process called Client.

Server: a process on networked computer that accepts requests from Client processes on other computers to perform a service and responds appropriately.

Remote invocation: A complete interaction between client and server, from the point when the client sends its request to when it receives the server's response.

Transparency	Description
Access	Hide differences in data representation and how a resource is accessed
Location	Hide where a resource is located
Migration	Hide that a resource may move to another location (i.e., Cold Migration)
Relocation	Hide that a resource may be moved to another location while in use (i.e., Hot Migration)
Replication	Hide that a resource may be redundant and shared by several competitive users
Concurrency	Hide that a resource may be shared by several competitive users
Failure	Hide the failure and recovery of a resource
Persistence	Hide whether a (software) resource is in memory or on disk

Access Transparency: Using identical operations to access local and remote resources, e.g. a Graphical User Interface (GUI) with folders.

Location Transparency: Resources to be accessed without knowledge of their location, e.g. URL (Support Availability).

Concurrency Transparency: Several processes operate concurrently using shared resources without interference with between them.

Replication Transparency: Multiple instances of resources to be used to increase reliability and performance without knowledge of the replicas by users or application programmers.

Failure Transparency: Users and applications complete their tasks despite the failure of hardware and software components (Allow fail and recovery) (reliability) (e.g. email).

Mobility Transparency: Movement of resources and clients within a system without affecting the operation of users and programs (Migration and Relocation), e.g., mobile.

Performance Transparency: Allows the system to be reconfigured to improve performance as loads vary.

Scaling Transparency: Allows the system and applications to expand in scale without change to the system structure or the application algorithms

Openness: is concerned with extensions and improvement of distributed systems. Detailed interfaces of components need to be published. New components have to be integrated with existing components .It is determined by the degree to which new resource can be added and made available for using by a variety of client programs.

Heterogeneity: Differences in data representation of interface types on different processors (of different vendors) have to be resolved.

Scalability: easy to expand and manage. A system is described as scalable if will remain effective when there is a significant increase in the number of resources and the number of users.
e.g: Internet

Latency: network delay before any data is sent.

Bandwidth: maximum channel capacity (analogue communication, digital communication).

Granularity: relative size of units of processing required. Distributed Systems operate best with coarse grain granularity because of the slow communication compared to processing speed in general. (kol ma el communication y2l el computation yzed)

Processor speed: MIPS, FLOPS.

Reliability: ability to continue operating correctly for a given time.

Fault tolerance: flexibility to partial system failure.

Security: policy to deal with threats to the communication or processing of data in a system.

Administrative/management domains: issues concerning the ownership and access to distributed systems components.

P2P: is the sharing of computer resources and services by direct exchange between systems. These resources and services include: Information Exchange, Processing Cycles, Cache Storage, & Disk Storage for Files.

Cluster Computing: Grouping multiple standalone computers in a cluster by a network.

High Performance Computing (HPC) Cluster: HPC clusters used when Time to solution a problem is important & problem is too big and can't fit on one single computer.

Utility Computing: The design of Utility Computing based on a service providing (business) model when the users (consumers) need computing resources, they pay providers for using it.

Jungle Computing : refers to the use of diverse, distributed and highly non-uniform performance computer systems to achieve peak performance.

Cloud computing: It is a computational environment that provides transparent access to a shared pool of computing resources matching the user's needs

Cloud computing NIST : It's a Computing model for enabling everywhere, convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction

On-demand Self-Service: A consumer can provision computing capabilities (e.g., server time, network, and storage) as needed automatically without requiring human interaction with each service provider.

Broad Network Access: Capabilities available over network and accessed through standard mechanisms using heterogeneous client platforms (e.g., mobile phones, tablets, laptops, and workstations).

Resource Pooling: Computing resources are pooled to serve multiple consumers according to their demand.

Rapid Elasticity: Capabilities can elastically provision and released, in some cases automatically, to scale up and down according to consuming rate.

Measured Service: Resource usage can monitor, control, and report, providing transparency for both provider and consumer of the utilized service.

Infrastructure as a Service (IaaS): The capability provided to the consumer **(organizations)** to provision processing, storage, networks, and other fundamental computing resources where the consumer able to **deploy** and run arbitrary software.

Platform as a Service (PaaS): The capability provided to the consumer **(developer)** by needed platform environment to **develop** his application using programming languages, libraries, services, and tools supported by the provider.

Software as a Service (SaaS): The computing capability Provided to consumer **(External user)** with provider's applications running on a cloud infrastructure. The applications are accessible from various client devices through web browser

Human as a service (HuaaS) : extraction of information from crowds of people

Private Cloud: Cloud infrastructure provisioned for exclusive use by a single organization comprising multiple consumers (e.g., business units).it may be owned, managed, and operated by the organization, a third party, or some combination of them, and it may exist on or off premises.

Public Cloud : Cloud infrastructure is provisioned for open use by the general public. It may be owned, managed, and operated by a business, academic, or government organization, or some combination of them.

Community Cloud: It's provisioned for exclusive use by a specific community of consumers from organizations that have shared concerns (e.g., mission, security requirements, policy, and compliance considerations).It may be owned, managed, and operated by one or more of the organizations in the community, a third party, or some combination of them, and it may exist on or off premises.

Hybrid Cloud: it's a composition of two or more distinct cloud infrastructures (private, community, or public) bound together by standardized or proprietary technology that enables data and application portability. A cloud computing environment uses a mix of on-premises, private cloud and third-party, public cloud services with orchestration between their platforms.

Multi Cloud: It refers to the presence of more than one cloud deployment of the same type (Public or Private or Hybrid), sourced from different vendors.

Elasticity: Degree to which a system can adapt to workload changes by provisioning and deprovisioning resources in an autonomic manner, such that at each point in time the available resources match the current demand as closely as possible to avoid Over- Provisioning or Under-Provisioning.

Over-Provisioning: Allocating more resources than required.

Under-Provisioning: Allocating fewer resources than required.

Resource Scheduling : providing Virtual Machines with specific

configuration.

service Level Agreement (SLA): A contract between the user and the Provider. Service SLA contains terms and conditions to ensure the rights of the users, as well as the providers. Its role is identifying the user's needs and creates a relationship between the user and the service provider.

Virtualization: It is a technology to run multiple same or different isolated OSs on a single physical system by abstracting and partitioning it's physical resources into multiple virtual machines with different workloads to improve IT throughput and cost.

Dual System: A computer system in which two operating systems are installed on the same hard drive, allowing either operating system to be loaded and given control.

Emulation System: A system that pretends to be another system.

Virtualization System: A system that pretends to be two or more of the same system.

Host: underlying hardware system

Virtual Machine Manager (VMM) or hypervisor :creates and runs virtual machines by providing interface that is identical to the host

Guest: process provided with virtual copy of the Host

Hypervisor: A software that allows multiple OSs (guest) to share a single hardware host.

Responsible:

Controlling the host processor and resources,

Allocating what is needed to each operating system in turn, and

Making sure that the guest operating systems (called Virtual Machines (VMs)) cannot disrupt each other.

Small Scale Consolidation: Operate different OS's and applications on one single server

Production Consolidation: A company can achieve greater efficiency and increase profitability by selling all or part of its manufacturing operations.

Comparison

Parallel Verses Distributed:

	Parallel	Distributed
Hardware	identical processors regular interconnection	different types of processors Networks
Memory	Shared memory	Distributed memory
Control	synchronized (global clock)	A synchronized execution of tasks (no global clock)
Main focus	Performance	Information / resource sharing Reliability / availability, …
Homogeneity	tasks perform similar functions	Inhomogeneity, tasks perform different functions

Centralized and Decentralized Systems

Centralized Systems	Decentralized Systems
System shared by users all the time	Multiple autonomous components
All resources accessible	Some resources may not be accessible
Software runs in a single process	Components shared by users
Single physical location	Multiple physical locations
Single point of control	Multiple points of control
Single point of failure	Multiple points of failure
	No Global time & No Shared memory
	Software can run in concurrent processes on different processors
Example: Airplane booked, Banks	Example: Gird, Cloud

Comparison	Grid Computing	Cloud Computing
Means of utilization	Allocation of multiple servers onto a single task or job	Virtualization of servers; one server to compute several tasks concurrently
Typical usage pattern	Typically used for job execution, i.e., the execution of a program for a limited time	More frequently used to support long-running services
Level of abstraction	Expose high level of detail	Provide higher-level abstractions

Mainframe Verses Cloud Computing

	Mainframe	Cloud Computing
Differences	Offers finite computing power	Provides almost infinite power and capacity
	Dummy terminals as user interface devices	PCs can provide local computing and cashing
Similarity	Cloud is a return to mainframe computing	

Multi-Clouds Verses Hybrid Cloud

Compression	Hybrid Cloud	Multi-Cloud
Differences	An amalgamation of a private cloud with one or more public cloud.	An amalgamation of two or more public/hybrid cloud under centralized management
	It can be any combination of SaaS, IaaS, PaaS and any other as-a-service	A multi-cloud strategy offers the ability to select different cloud services from different providers
	It is a singular entity (ie., the cloud components are integrated to form singular entity)	It isn't single entity "individual clouds may not be integrated together"
	a hybrid cloud could be part of a multi-cloud deployment.	Enables organizations to locate IT resources closely to end users to achieve optimal performance and minimal latency
Similarity	Multi-cloud and hybrid cloud computing are similar, but different IT infrastructure model.	

Cloud Computing Services

	Who Uses it	What Services are available	Why use it?
SaaS	Business Users	Email, Office, CRM, Blogs.	To complete business tasks.
PaaS	Developers	Services and Applications test, development, integrations and deployment environment.	Create and deploy service for users
IaaS	System Manager	Virtual Machines, Operations Systems, Massaging Queue, Networks, Storage, CPU, Memory backup. Services.	Create platforms for services and applications test, development, integration and deployment

Server without virtualization	Server with virtualization
Single OS can run at time	Can run multiple OSs
Software & hardware tightly coupled	Hardware independence of OS & applications
Running multiple apps on same machine creates conflict	Save electricity, initial cost to buy server
Hardware changes require manual effort	

Full Virtualization	Para Virtualization
all software (including all OS's) capable of executing on the raw (bare) hardware	These hypervisors run on a conventional operating system (Host OS) just as other computer programs do.
It directly sitting on top of the bare hardware device	Para hypervisors abstract Guest OSs from the Host OS.
Hypervisors Enable to run multi-unmodified guest operating system	Improves performance & Lower overhead.
Guest OS is not aware that it is being virtualized.	A Guest OS runs as a process on the Host OS.