



**Cloud computing** is a type of Internet-based computing that provides shared computer processing resources and data to computers and other devices on demand. It is a model for enabling ubiquitous, on-demand access to a shared pool of configurable computing resources (e.g., computer networks, servers, storage, applications and services),[1][2] which can be rapidly provisioned and released with minimal management effort. Cloud computing and storage solutions provide users and enterprises with various capabilities to store and process their data in either privately owned, or third-party data centers[3] that may be located far from the user–ranging in distance from across a city to across the world. Cloud computing relies on sharing of resources to achieve coherence and economy of scale, similar to a utility (like the electricity grid) over an electricity network.

Advocates claim that cloud computing allows companies to avoid up-front infrastructure costs (e.g., purchasing servers). As well, it enables organizations to focus on their core businesses instead of spending time and money on computer infrastructure.[4] Proponents also claim that cloud computing allows enterprises to get their applications up and running faster, with improved manageability and less maintenance, and enables Information technology (IT) teams to more rapidly adjust resources to meet fluctuating and unpredictable business demand.[4][5][6] Cloud providers typically use a "pay as you go" model. This will lead to unexpectedly high charges if administrators do not adapt to the cloud pricing model.[7]

In 2009, the availability of high-capacity networks, low-cost computers and storage devices as well as the widespread adoption of hardware virtualization, service-oriented architecture, and autonomic and utility computing led to a growth in cloud computing.[8][9][10] Companies can scale up as computing needs increase and then scale down again as demands decrease.[11] In 2013, it was reported that cloud computing had become a highly demanded service or utility due to the advantages of high computing power, cheap cost of services, high performance, scalability, accessibility as well as availability. Some cloud vendors are experiencing growth rates of 50% per year,[12] but being still in a stage of infancy, it has pitfalls that need to be addressed to make cloud computing services more reliable and user friendly

**Infrastructure as a service (IaaS)**

See also: Category:Cloud infrastructure

According to the Internet Engineering Task Force (IETF), the most basic cloud-service model is that of providers offering computing infrastructure – virtual machines and other resources – as a service to subscribers. Infrastructure as a service (IaaS) refers to online services that abstract the user from the details of infrastructure like physical computing resources, location, data partitioning, scaling, security, backup etc. A hypervisor, such as Xen, Oracle VirtualBox, Oracle VM, KVM, VMware ESX/ESXi, or Hyper-V, runs the virtual machines as guests. Pools of hypervisors within the cloud operational system can support large numbers of virtual machines and the ability to scale services up and down according to customers' varying requirements. Linux containers run in isolated partitions of a single Linux kernel running directly on the physical hardware. Linux cgroups and namespaces are the underlying Linux kernel technologies used to isolate, secure and manage the containers. Containerisation offers higher performance than virtualization, because there is no hypervisor overhead. Also, container capacity auto-scales dynamically with computing load, which eliminates the problem of over-provisioning and enables usage-based billing.[63] IaaS clouds often offer additional resources such as a virtual-machine disk-image library, raw block storage, file or object storage, firewalls, load balancers, IP addresses, virtual local area networks (VLANs), and software bundles.[64]

IaaS-cloud providers supply these resources on-demand from their large pools of equipment installed in data centers. For wide-area connectivity, customers can use either the Internet or carrier clouds (dedicated virtual private networks). To deploy their applications, cloud users install operating-system images and their application software on the cloud infrastructure.[65][unreliable source?] In this model, the cloud user patches and maintains the operating systems and the application software. Cloud providers typically bill IaaS services on a utility computing basis: cost reflects the amount of resources allocated and consumed.[66][67][68][69]

EXAMPLE: storage

**Platform as a service (PaaS)**

Main article: Platform as a service

See also: Category:Cloud platforms

PaaS vendors offer a development environment to application developers. The provider typically develops toolkit and standards for development and channels for distribution and payment. In the PaaS models, cloud providers deliver a computing platform, typically including operating system, programming-language execution environment, database, and web server. Application developers can develop and run their software solutions on a cloud platform without the cost and complexity of buying and managing the underlying hardware and software layers. With some PaaS offers like Microsoft Azure and Google App Engine, the underlying computer and storage resources scale automatically to match application demand so that the cloud user does not have to allocate resources manually. The latter has also been proposed by an architecture aiming to facilitate real-time in cloud environments.[70][need quotation to verify] Even more specific application types can be provided via PaaS, such as media encoding as provided by services like bitcodin.com[71] or media.io.[72]

Some integration and data management providers have also embraced specialized applications of PaaS as delivery models for data solutions. Examples include iPaaS (Integration Platform as a Service) and dPaaS (Data Platform as a Service). iPaaS enables customers to develop, execute and govern integration flows.[73] Under the iPaaS integration model, customers drive the development and deployment of integrations without installing or managing any hardware or middleware.[74] dPaaS delivers integration—and data-management—products as a fully managed service.[75] Under the dPaaS model, the PaaS provider, not the customer, manages the development and execution of data solutions by building tailored data applications for the customer. dPaaS users retain transparency and control over data through data-visualization tools.[76] Platform as a Service (PaaS) consumers do not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage, but have control over the deployed applications and possibly configuration settings for the application-hosting environment.

A recent specialized PaaS is the Blockchain as a Service (BaaS), that some vendors such as Microsoft Azure have already included in their PaaS offering.[77]

EXAMPLE: DATABASE DEVELOPING TOOLS

**Software as a service (SaaS)**

Main article: Software as a service

In the software as a service (SaaS) model, users gain access to application software and databases. Cloud providers manage the infrastructure and platforms that run the applications. SaaS is sometimes referred to as "on-demand software" and is usually priced on a pay-per-use basis or using a subscription fee.[78] In the SaaS model, cloud providers install and operate application software in the cloud and cloud users access the software from cloud clients. Cloud users do not manage the cloud infrastructure and platform where the application runs. This eliminates the need to install and run the application on the cloud user's own computers, which simplifies maintenance and support. Cloud applications differ from other applications in their scalability—which can be achieved by cloning tasks onto multiple virtual machines at run-time to meet changing work demand.[79] Load balancers distribute the work over the set of virtual machines. This process is transparent to the cloud user, who sees only a single access-point. To accommodate a large number of cloud users, cloud applications can be multitenant, meaning that any machine may serve more than one cloud-user organization. PX email, Virtual

The pricing model for SaaS applications is typically a monthly or yearly flat fee per user,[80] so prices become scalable and adjustable if users are added or removed at any point.[81] Proponents claim that SaaS gives a business the potential to reduce IT operational costs by outsourcing hardware and software maintenance and support to the cloud provider. This enables the business to reallocate IT operations costs away from hardware/software spending and from personnel expenses, towards meeting other goals. In addition, with applications hosted centrally, updates can be released without the need for users to install new software. One drawback of SaaS comes with storing the users' data on the cloud provider's server. As a result,[citation needed] there could be unauthorized access to the data. For this reason, users are increasingly[quantify] adopting intelligent third-party key-management systems to help secure their data.[citation needed

**Grid computing** is the collection of computer resources from multiple locations to reach a common goal. The grid can be thought of as a distributed system with non-interactive workloads that involve a large number of files. Grid computing is distinguished from conventional high performance computing systems such as cluster computing in that grid computers have each node set to perform a different task/application. Grid computers also tend to be more heterogeneous and geographically dispersed (thus not physically coupled) than cluster computers.[1] Although a single grid can be dedicated to a particular application, commonly a grid is used for a variety of purposes. Grids are often constructed with general-purpose grid middleware software libraries. Grid sizes can be quite large.[2]

Grids are a form of distributed computing whereby a "super virtual computer" is composed of many networked loosely coupled computers acting together to perform large tasks. For certain applications, distributed or grid computing can be seen as a special type of parallel computing that relies on complete computers (with onboard CPUs, storage, power supplies, network interfaces, etc.) connected to a computer network (private or public) by a conventional network interface, such as Ethernet. This is in contrast to the traditional notion of a supercomputer, which has many processors connected by a local high-speed computer bus.

**BENEFITS OF CLUSTERING**

System availability (HA) : Cluster offers inherent high system availability due to the redundancy of hardware, operating systems, and applications.

Hardware Fault Tolerance: Cluster has some degree of redundancy in most system components including both hardware and software modules.

OS and application reliability : Run multiple copies of the OS and applications, and through this redundancy

Scalability : Adding servers to a cluster or adding more clusters to a network as the application need arises.

High Performance : Running cluster enabled programs to yield higher throughput.

MPP/DSM:

Compute across multiple systems: parallelism.

Network RAM:

Idle memory in other nodes. Page across other nodes’ idle memory

Software RAID:

file system supporting parallel I/O and reliability, mass-storage.

Multi-path Communication:

Communicate across multiple networks: Ethernet, ATM, Myrinet

**Operational Benefits of Clustering**

System availability (HA) : Cluster offers inherent high system availability due to the redundancy of hardware, operating systems, and applications.

Hardware Fault Tolerance: Cluster has some degree of redundancy in most system components including both hardware and software modules.

OS and application reliability : Run multiple copies of the OS and applications, and through this redundancy

Scalability : Adding servers to a cluster or adding more clusters to a network as the application need arises.

High Performance : Running cluster enabled programs to yield higher throughput.

**Issues in Cluster Design**

Size Scalability (physical & application)

Enhanced Availability (failure management)

Single System Image (Middleware, OS extensions)

Fast Communication (networks & protocols)

Load Balancing (CPU, Net, Memory, Disk)

Security and Encryption (clusters and Grids)

Distributed Environment (User friendly)

Manageability (Jobs and resources )

Programmability (simple API required)

Applicability (cluster- and grid-awareness)

**SSI**

A single system image is the illusion, created by software or hardware, that presents a collection of resources as an integrated powerful resource.

SSI makes the cluster appear like a single machine to the user, applications, and network.

A cluster with multiple system images is nothing but a collection of independent computers (Distributed systems in general)

**GRIDS vs CLUSTERS vs CLOUD**

Cluster differs from Cloud and Grid in that a cluster is a group of computers connected by a local area network (LAN), whereas cloud and grid are more wide scale and can be geographically distributed. Another way to put it is to say that a cluster is tightly coupled, whereas a Grid or a cloud is loosely coupled. Also, clusters are made up of machines with similar hardware, whereas clouds and grids are made up of machines with possibly very different hardware configurations.

Cloud Computing refers to both the applications delivered as services over the Internet and the hardware and systems software in the datacenters that provide those services. The services themselves have long been referred to as Software as a Service (SaaS). The datacenter hardware and software is what we call a Cloud. When a Cloud is made available in a pay-as-you-go manner to the general public, we call it a Public Cloud; the service being sold is Utility Computing. We use the term Private Cloud to refer to internal datacenters of a business or other organization, not made available to the general public. Thus, Cloud Computing is the sum of SaaS and Utility Computing, but does not include Private Clouds. People can be users or providers of SaaS, or users or providers of Utility Computing.

The difference between a cloud and a grid can be expressed as below:

1. Resource distribution: Cloud computing is a centralized model whereas grid computing is a decentralized model where the computation could occur over many administrative domains.
2. Ownership: A grid is a collection of computers which is owned by multiple parties in multiple locations and connected together so that users can share the combined power of resources. Whereas a cloud is a collection of computers usually owned by a single party.

Examples of Clouds: Amazon Web Services (AWS), Google App Engine

Examples of Grids: FutureGrid

Dropbox, Gmail, Facebook, Youtube, Rapidshare, etc are all examples of cloud computing services.

**GPU PROCESSING**

General-purpose computing on graphics processing units (GPGPU, rarely GPGP or GP²U) is the use of a graphics processing unit (GPU), which typically handles computation only for computer graphics, to perform computation in applications traditionally handled by the central processing unit (CPU).[1][2][3] The use of multiple video cards in one computer, or large numbers of graphics chips, further parallelizes the already parallel nature of graphics processing.[4] In addition, even a single GPU-CPU framework provides advantages that multiple CPUs on their own do not offer due to the specialization in each chip.[5]

Essentially, a GPGPU pipeline is a kind of parallel processing between one or more GPUs and CPUs that analyzes data as if it were in image or other graphic form. While GPUs operate at lower frequencies, they typically have many times the number of cores. Thus, GPUs can operate on pictures and graphical data effectively far faster than a traditional CPU. Migrating data into graphical form and then using the GPU to scan and analyze it can result in profound speedup.

GPGPU pipelines were first developed for better, more general graphics processing (e.g., for better shaders). These pipelines were found to fit scientific computing needs well, and have since been developed in this direction.

**DATA CENTERS**

A data center is a facility used to house computer systems and associated components, such as telecommunications and storage systems. It generally includes redundant or backup power supplies, redundant data communications connections, environmental controls (e.g., air conditioning, fire suppression) and various security devices. Large data centers are industrial scale operations using as much electricity as a small town.[1][2]

**IOT**

The Internet of things (IoT) is the inter-networking of physical devices, vehicles (also referred to as "connected devices" and "smart devices"), buildings, and other items embedded with electronics, software, sensors, actuators, and network connectivity which enable these objects to collect and exchange data.[1][2][3] In 2013 the Global Standards Initiative on Internet of Things (IoT-GSI) defined the IoT as "a global infrastructure for the information society, enabling advanced services by interconnecting (physical and virtual) things based on existing and evolving interoperable information and communication technologies"[3] and for these purposes a "thing" is "an object of the physical world (physical things) or the information world (virtual things), which is capable of being identified and integrated into communication networks".[4] The IoT allows objects to be sensed or controlled remotely across existing network infrastructure,[5] creating opportunities for more direct integration of the physical world into computer-based systems, and resulting in improved efficiency, accuracy and economic benefit in addition to reduced human intervention.[6][7][8][9][10][11] When IoT is augmented with sensors and actuators, the technology becomes an instance of the more general class of cyber-physical systems, which also encompasses technologies such as smart grids, virtual power plants, smart homes, intelligent transportation and smart cities. Each thing is uniquely identifiable through its embedded computing system but is able to interoperate within the existing Internet infrastructure. Experts estimate that the IoT will consist of about 30 billion objects by 2020.[12]