**Yucel TUREL**
Polish-Japanese Institute of Information Technology

# Cloud Computing and Virtualization: A Comprehensive Survey

***Abstract**. A short review of cloud computing and virtualization mechanisms is given. The benefits and risks of virtualization are presented. The differences between the grid and distributed computing are discussed. The role of the virtual machine manager is described.*

***Streszczenie.** Zaprezentowano zwarty przegląd obliczeń w chmurze i mechanizmów wirtualizacji. Przedstawiono zalety i ryzyka wynikające z procesu wirtualizacji. Przedyskutowano różnice pomiędzy obliczeniami gridowymi i rozproszonymi. Opisano rolę, jaką odgrywa zarządca maszyny wirtualnej. **Przegląd obliczeń w chmurze i mechanizmów wirtualizacji***

**Keywords:** cloud computing, virtualization, virtual machine.
**Słowa kluczowe:** obliczenia w chmurze, wirtualizacja, maszyna wirtualna.

## Introduction

Cloud computing is nowadays the most popular and flexible online technology using the virtual machines based on the pay per use method. It can be implemented by large, medium and small enterprises.

It is a convenient on-demand network access to the shared pool of the configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction [1].

Cloud computing is structurally quite similar to grid and distributed computer model. One of the main difference of cloud computing may be the scalability of data storing and retrieving.

Cloud computing offers facilities ranging from the simple word processing to complex data analysis. It can increase the performance of the IT infrastructure of an enterprise without investing many funds.

Companies are turning to the cloud computing for their datacenters to cut maintenance, labor and hardware costs. They can also improve scalability and elasticity. However, all risks should be projected and legal aspects should be considered carefully.

Virtualization is the mechanism that cloud computing works on. Virtualization of servers gives mobility and reduces the overall costs. Virtualized cloud computing provides virtual machines to the end-users to execute their software and applications on remote sites. There are different types of virtualization techniques which are described in the later sections of this article.

Hypervisor is the virtual machine operating system which creates virtual instances for each virtual machine. Each instance can be considered as a virtual machine.

The benefits of using cloud computing include lower operating costs, physical space savings, energy savings and increased availability. On the other hand, it brings some security related risks. It can be difficult to treat hypervisor as a real time operating system. Thus, extra configurations should be tuned on the virtual operating system such as authorization of users who wish to log on to the system.

There are other problems; for example, some virtual machines can be dormant (not running) at some point in time [2]. These virtual machines cannot be easily kept up-to-date, because typically this would require the virtual machines to be started, updated and shut down again, which is not only time-consuming but also is an irregular task to do on a permanent schedule.

Furthermore, machines used in an IaaS environment are subject to external attacks, i.e., they might be a selected or random target. Software updated and patching should be done periodically and virus scan should be run for known security vulnerabilities.

In this article, essential features of cloud computing and virtual machine mechanism are examined briefly.

## Cloud computing

Cloud computing may provide services such as IaaS, Paas and SaaS which will be described in this section. Those are the infrastructure, platform and software services. To be able to understand cloud computing, services and deployment should be explained first. Thus, data and application of customers are stored and distributed using on-line technology.

**IaaS** (Infrastructure as a Service): it is the delivery of computer infrastructure (typically a platform virtualization environment) as a service. Rather than purchasing servers, software, data center space or network equipment, clients buy those resources as a fully outsourced service. It focuses on managing virtual machines, and the risks are little different than with other cloud types, the main risk is malicious user or forgery of services. IaaS requires governance and usage monitoring.

**PaaS** (Platform as a Service): it is the delivery of a computing platform and solution stack as a service. It facilitates the deployment of applications without the cost and complexity of buying and managing the underlying hardware and software layers. At this level data encryption takes place and PaaS can be inherently secure, but the risk is the slow system performance. Still, any solution implemented should manage the connection to the cloud service and automatically encrypt 'confidential user' data such as home addresses, social security numbers or even medical records.

**SaaS** (Software as a Service): it is a model of software deployment whereby a provider licenses an application to customers for use as a service on demand. It delivers applications to the end users. At this level, authentication and password management take place. The main risk is likely to stem from multiple passwords accessing applications [3].

## Cloud Deployment Models

There are four deployment models for cloud services that address specific requirements. They all share the same basic features except authorization access to different cloud. These models are defined by CSA (Cloud security alliance) as listed below [3]:

**Public Clouds:** in a public cloud a service provider makes IT resources, such as CRM or payroll applications, storage capacity, or server compute cycles, available to any customer via the Internet.

**Private Clouds:** in a private cloud IT resources are consolidated and pooled so users across the company can have self-service access and increased scalability. Like a public cloud, a private cloud also makes provisioning an automated service request rather than a manual task processed by IT.

**Community Clouds:** the cloud infrastructure is shared by several organizations and supports a specific community that has shared concerns (e.g., mission, security requirements, policy, or compliance considerations).

**Hybrid Clouds**: like the private model let an organization continue to use their existing data center equipment and keep sensitive data secured on the organization's own network. Like the public cloud, a hybrid model lets an organization take advantage of a cloud's almost unlimited scalability.

Thus, mainly cloud computing covers IT infrastructures, software, operating system, security, backup and storage. Cloud computing are provided by Amazon, MS Azure, Google and VMware and others.

### Differences between grid and distributed computing

Pooling resources is an effective way of resource allocation and without it, managing online operations is almost impossible. Cloud technology is not a new way of sharing resources online. Hitherto, grid computing and distributed computing were well known network systems. Grid computing is an old terminology but spreading resources elsewhere and reaching them when demanded is the main idea.

As it is clear from the definition, traditional distributed computing can be characterized as a subset of grid computing. Some of the differences between these two are [4]:

1. Distributed computing normally refers to managing or pooling the hundreds or thousands of computer systems which individually are more limited in their memory and processing power. On the other hand, grid computing has some extra characteristics. It is concerned to efficient utilization of a pool of heterogeneous systems with optimal work load management utilizing an enterprise's entire compu-tational resources( servers, networks, storage, and information) acting together to create one or more large pools of computing resources. There is no limitation of users, departments or originations in grid computing.
2. Grid computing is focused on the ability to support computation across multiple administrative domains that sets it apart from traditional distributed computing. Grids offer a way of using the information technology resources optimally inside an organization involving virtualization of computing resources. Its concept of support for multiple administrative policies and security authentication and authorization mechanisms enables it to be distributed over a local, metropolitan, or wide-area network.

### Virtualization

Virtualization is the key word when one talks about the cloud computing. The cloud computing terminology implies virtualisation, virtual server and virtual machine, etc.

Virtual technology assigns a logical name for a physical resource and then provides a pointer to that physical resource when request is made. It increases the limits and capacity of hardware resources such as servers, workstations, networks or storage devices. Virtualization and cloud computing allow computer users access to powerful computers and software applications hosted by the remote group of servers but security as related to data privacy limits the public confidence and slows the adoption of new technologies [5].

Many organizations are already practising the cost savings from implementing the virtualized servers and system administrators benefit from the ease of deployment and management for the virtualized systems.

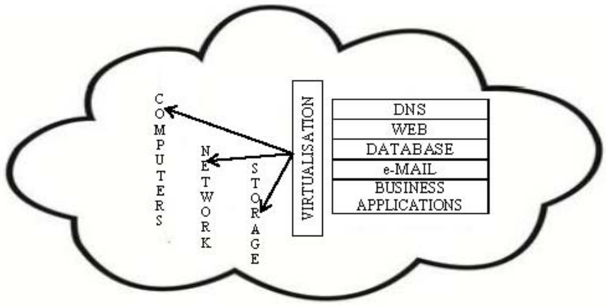


Fig. 1. Cloud computing virtualization

Virtualization is an old concept, first introduced in the 1960s with the appearance of mainframe computers. It was re-introduced to personal computers in the 1990s, and currently major products available are: Microsoft Virtual PC (Microsoft Virtual PC 2007), VMware software tools range (VMware, 2007), an open source software QEMU (Bellard, 2007), and few others [6].

Building on the foundation of vSphere virtualization, VMware vCloud Suite abstracts, pools, and automates all IT services to deliver an elastic, easy-to-manage Software-defined Datacenter. The combination of these advantages can direct resources previously spent on maintaining existing systems to work in investments aimed at creating innovative services that generate new revenue, streamline operations and meet business objectives [7].

- Memory partitioning – hardware-enforced partitioning of physical memory ensures that each VM can only access the assigned physical memory and in a fair manner only.
- Dedicated virtual I/O devices – I/O device modifications to support virtualization enables each VM to be given direct access to a dedicated (virtual) I/O device. The memory management facilities along with chipset support ensure that only the authorized VM can access the memory-mapped I/O and only at a given rate.

### Hypervisor

A hypervisor, also called a virtual machine manager, is a program that allows multiple operating systems to share a single hardware host. Each operating system appears to have the host's processor, memory, and other resources all to itself. However, the hypervisor is actually controlling the host processor and resources, allocating what are needed to each operating system in turn and making sure that the guest operating systems (called virtual machines) cannot disrupt each other.

A computer on which a hypervisor is running one or more virtual machines is defined as a *host machine*. Each virtual machine is called *a guest machine*. The hypervisor presents the guest operating systems with a virtual operating platform and manages the execution of the guest operating systems. Multiple instances of a variety of operating systems may share the virtualized hardware resources [8].

The hypervisor lets us show the same application on lots of systems without having physically copy that application onto each system. Because of the hypervisor architecture, it can load any (or many) different operating system as though it were just another application. Therefore, the hypervisor is a very practical way of getting things virtualized quickly and efficiently [8].

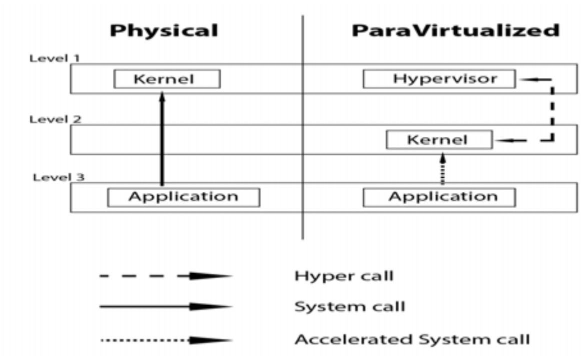


Fig 2. Hypervisor architecture

**Load balancing and virtualization**

One characteristics of a cloud computing is virtualized network access to a service. No matter where user can access the service, they are directed to available resource. Technology used to distribute service request to resources is referred to as load balancing. Load balancing can be implemented in hardware or in software. It is an optimization technique, it can be used to increase utilization and throughput, lower latency, reduces response time and avoid system overload. Load balancing systems can use different mechanisms to assign service direction [6].

Well, for example data in a specific database can be split into more than one hard drive on different workstations. Inserting data and retrieving data will be from the central database but the data could be divided into different hardware resources. As mentioned above, the main advantage is to balance hardware space and prevent system overload.

**Different types of cloud virtualization**

There are several types of the cloud virtualization listed below:

- **Network virtualization:** it creates virtual networks whereby each application sees its own logical network independent of the physical network. A VLAN is an example of hardware virtualization that provides an easy, flexible and less expensive way to manage network.
- **Server virtualization:** it enables multiple operating system and applications to run simultaneously on different virtual machines created on the same physical server. It provides a layer of abstraction between OS and underlying hardware.
- **Storage virtualization:** with storage virtualization, the disk/data storage for the user data is consolidated to and managed by virtual storage system. The servers connected to the storage system are not aware of where the data really is.
- **Application virtualization:** an application runs on another host from where it is installed in a variety of ways. It could be done by application streaming, desktop virtualization or a VDI or a VM packages (like VMware ACE created by a player). Microsoft Soft grid is an example of Application virtualization.
- **Para virtualization:** it requires that the host operating system provides a virtual machines interface for the Guest OS and that the Guest access network through that host virtual machine. An OS running as a guest on a paravirtualization system must be ported to work with the host interface.
- **Full virtualization:** all OS in full virtualization communicate directly with the VM hypervisors, so guest OS do not require any modification. Guest OS do not require any modification. Guest OS full virtualization systems are generally faster than other virtualization schemes [9].

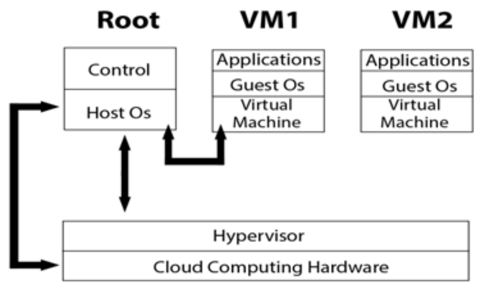


Fig 3. Cloud computing virtualization schema

**IP virtualization and Google cloud**

Google is the most frequently visited search engine on the internet. It has a giant infrastructure and its network is one of the largest in the world. It is estimated Google runs over million servers worldwide, maintains a pool of hundreds thousands of IP addresses, process a billion search requests and generates twenty peta bytes of data per day. Google never gives data centre tours to journalists and does not disclose where its data centre is located.
When ~~a~~ user initiates a Google search, his query is sent to a DNS server, which then queries Google's DNS server. The Google DNS server examines the pool, determines the geographical origin and uses a round robin policy to assign an IP address to that request. The request usually goes to the nearest data centre and that IP address is for a cluster of Google servers. This acts as a first level of IP utilisation [9].

**Benefits of virtualization**

Virtualizing servers and machines is the main ad-vantage of cloud computing. From one hardware re-source as many virtual devices as possible can be derived. They can be independently offered to the users regardless of their physical location. In this section some of the benefits of cloud computing virtualization are outlined.

- ***Save money:*** x86 servers are running at an average of only 20 to 25 percent of total capacity. With virtualization, we can turn a single purpose server into a multi-tasking one, and turn multiple servers into a computing pool that can adapt more flexibly to changing workloads.
- ***Save energy:*** Businesses spend a lot of money powering unused server capacities. Virtualization reduces the number of physical servers, reducing the energy required to power and cool them.
- ***Save time***: With fewer servers, we can spend less time on the tasks required for server maintenance. Pooling many storage devices into a single virtual storage device, you can perform tasks such as backup, and recovery more easily and more quickly. It's also much faster to deploy a virtual machine than it is to deploy a new physical server [10].

**Risks of virtualization**

Cloud virtualization helps to save hardware resources but on the other hand brings some risk factors which are defined below.

- A holistic view of risk and vulnerability is essential. Security for the cloud, particularly when protecting high-security systems, requires extensive security monitoring and granular identity and access management controls.
- Virtualization technology is still maturing. Additional tools and controls are necessary and
- hypervisors must be treated like OSs and secured as such.
- Granular administrator controls are required. Defining custom roles versus built-in roles are needed to ensure

permission is restricted for tasks specific personnel are not expected to perform.

- More granular logging and monitoring is required. Because virtual infrastructure is holistic and has many parts, it's important to monitor the various management agents. For a complete picture, it is necessary to collect logs from network switches, storage subsystems, and management servers [11].

**Conclusions**

A brief description of cloud computing usage, benefits and risks which are today the most important virtualized data sharing techniques are presented.

***Acknowledgements***

*Author thanks to Prof. R. Kotowski for fruitful discussions and support.*

REFERENCES

[1] The NIST, National Institue of Standarts and Technology, UK, Cloud Computing Project, http://csrc.nist.gov/nice/states/maryland/posters/cloud-computing.pdf

[2] J. Leach, *The Rise of Service Oriented IT and the Birth of Infrastructure as a Service*,March 20, 2008; http://advice.ciocom/jim_leach/the_rise_of_service_ oriented_it_and_the_birth_of_infrastructure_as_a_service

[3] Roland Schwarzkopf, Matthias Schmidt, Christian Strack, Simon Martin and Bernd Freisleben, Increasing Virtual Machine Security in Cloud Environments, Journal of Cloud Computing: Advance Systems and Application, 2012, doi: 10.1186/2192 - 113X-1-12

[4] Cloud Security Alliance (CSA) Security guidance for critical areas of focus in cloud computing, V2.1, 2009

[5] http://www.jatit.org/distributed-computing/grid-vs-distributed.htm

[6] Kumar Surai, Pal Singh, Ashwanee Kumar Singh Jahangir Alidad, Virtualization, The Great Thing and Issues in Cloud Computing, International Journal of Current Engineering and Technology, Vol.3, No. 2, 2013

[7] Derek Bem, Ewa Huebner, Computer Forensic Analysis in a Virtual Environment, International Journal of Digital Evidence Fall 2007, V.6, I.2, www.ijde.org, University of Western Sydney, Australia www.vmware.com

[8] https://en.wikipedia.org/wiki/Hypervisor

[9] Erway Kupcu, C. Papamanthou, and R. Tamassia (2009), Dynamic provable data Possession, Proc. 16th ACM conference, 213-222.

[10] Eric Keller, Jakub Szefer, Jennifer Rex Ford, Ruby B. Lee, Virtualized Cloud Infrastructure without the Virtualization, Princeton University, Princeton, 2010.

[11] Intel IT, IT Best Practices: Cloud Computing and Information Security, Intel White Paper, January 2012.

---

***Author**: Mr Yucel Turel, Polish-Japanese Institute of Information Technology, Koszykowa 86, 02-008 Warsaw, Poland*
*E-mail: tyucel1@hotmail.com*