A Model for Comparing Free Cloud Platforms

Radu LIXĂNDROIU, Cătălin MAICAN
Transilvania University of Brașov
lixi.radu@unitbv.ro, maican@unitbv.ro

VMware, VirtualBox, Virtual PC and other popular desktop virtualization applications are used only by a few users of IT techniques. This article attempts to make a comparison model for choosing the best cloud platform. Many virtualization applications such as VMware (VMware Player), Oracle VirtualBox and Microsoft Virtual PC are free for home users. The main goal of the virtualization software is that it allows users to run multiple operating systems simultaneously on one virtual environment, using one computer desktop.
Keywords: Cloud Computing, Virtualization, Jaccard Index, Comparing Model

1 Introduction

The desktop virtualization software allows multiple operating systems to install on a computer or laptop in a virtual environment without disturbing the host operating system. Moreover, in many such programs, one can change to another operating system without rebooting the computer. The main advantage of these applications is that they can be used on technically weaker computers, taking advantage of the virtual machine features on the server. [16], [17]

The advantages of virtualization platforms are:

1) Multiple operating systems can be used at a time, without rebooting.

2) They represent a completely isolated medium from the host operating system, because it is a virtual environment. Therefore, the possibility of infection with viruses or other unwanted items while working decreases.

3) They allow the running of different applications or different operating systems on one type of computer. For example, if one wishes to run some applications that only work in Win XP on a system that has installed Win 8, one can install a virtual machine on the existing system (with Win 8 installed) that simulates running Win XP.

4) The restoration of a virtual machine is very fast. For example, one can open the virtual machine selecting an older configuration. This function is usually called Snapshot. The Snapshot feature is usually used to create tests on the system.

Benefits obtained from the use of cloud computing services [15]

- Globalization for access to resources - people around the world can access cloud resources using the Internet.
- They make savings by increasing the volume or the productivity. Thus, lower costs per unit, project or product obtain.
- They allow running Streamline processes, which leads to obtaining a high efficiency by comparing the results obtained to the corresponding time and the people involved.
- They increase accessibility. Access can be made from anywhere, anytime.
- The monitoring of the projects is more effective.
- The levels of training of technical staff are lower. It requires fewer persons to configure and maintain hardware and software platforms.
- Savings on software licensing. There is no need to purchase expensive software licenses or software.
- They reduce the cost of capital. Companies can make great savings on hardware costs or software.
- Mobility - Employees with access to the internet can work anywhere. This flexibility positively affects the work-life balance for employees leading to increased labour productivity.
- Increasing flexibility. Changes can be made quickly without involving additional costs.

- They reduce spending money on technology infrastructure. Some technologies may be used only on demand when needed in certain projects.
- Flexibility - quickly meet business demands
- Payment for use - cloud services are usually paid only as long as they are used. Because cloud computing is quickly implemented in any company, start-up costs and running costs are minimal.
- Increased Collaboration - Cloud computing increases the level of collaboration, while enabling collaborative work
- Document Management - Cloud computing keeps all files in a central location and everyone is working on one version of the document. Employees can discuss with each other in real time to make changes together. The collaborative process increases efficiency and improves company communication.
- Security – the loss or destruction of the personal computer may have some serious financial implications, but when everything is stored in the cloud, data can still be accessed, regardless of what happens to a computer.
- Opportunities for back-up - companies that rely on cloud no longer need complex plans for disaster recovery. Cloud computing providers offer multiple services for data retention and backup.
- Competitiveness - Access through the cloud offers SMEs access to enterprise-class technology. Cloud computing allows small businesses to use superior resources.

2 Comparison model using Jaccard index

The Jaccard index, also known as the Jaccard similarity coefficient is a statistical indicator used to compare the similarity of several sets. The Jaccard index measures the similarity between finite two ore more sets of sentences and is defined as the intersection of sets divided by the sample sets meeting:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

If A and B are empty, define *J(A, B) = 1*

$$0 \le J(A,B) \le 1$$

The Jaccard distance is complementary to the Jaccard coefficient and is:

$$d_J(A,B) = 1 - J(A,B)$$
$$d_J(A,B) = \frac{|A \cup B| - |A \cap B|}{|A \cup B|}$$

2.1 Jaccard index for general information

Table 1. General information

Software	Initial release date	License(s)	Written in	As a service	Local installations	Jaccard index
Cloud Foundry	4/12/2011	Apache License	Ruby, Go	1	1	1
Cloud.com /CloudStack	5/4/2010	Apache license	Java, C	1	1	1
Eucalyptus[3]	5/29/2008	Proprietary, GPL v3	Java, C	1	1	1
OpenNebula[5]	03/2008	Apache License	C++, C, Ruby, Java, Shell	1	1	1

			script, lex, yacc			
OpenQRM[6]	03/2008	GPL License	C++, PHP, Shell script	1	1	1
OpenShift[7]	5/4/2011	Apache License	Ruby	1	1	1
OpenStack[8]	10/21/2010	Apache License	Python	1	1	1
oVirt	8/9/2012	Apache License	Java, Python	?	1	0.5
PetiteCloud	1/1/2014	BSD License	Java, C	0	1	0.5

Software as a Service (SaaS) [5] is a software distribution model where the applications are hosted by a provider and are available to customers over the Internet.

SaaS [9] is becoming more widespread as the technology develops Web, service-oriented architecture (SOA). Broadband companies have become increasingly available in order to support user access from more areas around the world.

The benefits of SaaS include:
• Easy Management
• Automatic Updates
• High compatibility: users access a single version of the software.
• Easier collaboration possibilities offered by working collaboratively.
• High mobility - can be accessed from any location with an Internet connection.

Software as a product (Local Installation) [6] - Software as a product (local installation) involves the purchase of a license. These software solutions are expensive purchases, but there are no monthly usage fees. In addition, there are extensive maintenance and upgrading costs.

Software as a product generally does not require Internet access to run the software. However, whenever a software update is needed, more money is needed. Some upgrades are not strictly necessary, but their purchase is recommended to increase their system reliability and to get new features. Some companies stop maintaining old software versions once a new version is released, leaving users with an expensive solution that is not usable.

2.2 Jaccard index for supported hosts

Table 2. Supported hosts

Software	Linux	FreeBSD	Windows	Bare Metal	Jaccard index
Cloud Foundry	1	0	0	1	0.5
Cloud.com / CloudStack	1	0	0	1	0.5
Eucalyptus	1	0	0	1	0.5
OpenNebula	1	0	?	0	0.25
OpenQRM	1	0	0	0	0.25
OpenShift	1	0	0	1	0.5
OpenStack	1	0	1	1	0.75
oVirt	1	0	0	1	0.5

PetiteCloud	1	1	1	0	0.75

FreeBSD - is a free operating system descendent from Unix. It is a complete, unitary developed operating system. The kernel, device drivers and user space utilities such as command interpreter or shell, are held in the same tree tracking source code revisions (CVS). This method of development is in contrast to the one used for Linux (an operating system similar to but better known), where the kernel is developed by a team of developers, and user-space utilities and applications are developed by other teams (such as GNU project). Then, all these are packaged together and released by other teams in the form of Linux distributions.

As an operating system, FreeBSD is generally regarded as very reliable and robust, and out of the operating systems that report the operating time since last boot (uptime) remotely, FreeBSD is the most popular choice of the free operating systems displayed in the Netcraft list of the top 50 web servers with the longest uptime.

BareMetal is an exokernel-based operating system created by Return Infinity. It is written in Assembly to achieve high-performance computing with minimal footprint. This operating system is not based on C, C++ or Unix-like kernels.

Linux: Designed for the Cloud – [8] Linux provides a perfect technology for cloud computing. This is modular, efficient, powerful, open source and omnipresent. The largest cloud infrastructures in the world are built on Linux.

Architecture compatible - Linux has a broad application and ISV ecosystem. With many users, Linux provides applications for the needs of all users.

Cost-licensing is not necessarily 0. There are many implementations for certain commercial costs to be paid. Linux is a more cost effective platform for implementing such solutions providers. Linux benefits arise from the speed of implementation and flexibility.

Virtualization - Linux can form a local or cloud environment. Linux is, therefore, an optimal platform for cloud, being a host operating system through technologies like KVM or Xen.

Windows Cloud Server Hosting Benefits - High availability: redundant architecture ensures maximum application / website uptime possible.

Particularly large scalability: cloud solutions are designed for each and every business.

Easy to use – the loading of the data on the server is performed using the most comfortable software used by the customer.

Possibility to manage monthly costs.

A history of over 16 years in hosting.

2.3 Jaccard index for supported clients

Table 3. Supported clients

Software	Linux	Windows	VMware	Xen	KVM	VirtualBox	Jaccard index
Cloud Foundry	1	0	1	1	1	1	0.833
Cloud.com / CloudStack	1	1	1	1	1	1	1.000
Eucalyptus	1	1	1	1	1	?	0.833
OpenNebula	1	1	1	1	1	1	1.000
OpenQRM	1	1	1	1	1	1	1.000
OpenShift	1	0	1	1	1	1	0.833
OpenStack	1	1	1	1	1	0	0.833

| oVirt | 1 | 1 | 0 | 0 | 1 | 0 | 0.500 |
| PetiteCloud | 1 | 1 | 1 | 0 | 1 | 1 | 0.833 |

VMware desktop software runs multiple operating systems such as Microsoft Windows, Linux and Mac OS X. VMware ESX servers and VMware ESXi servers are able to run directly on server hardware without requiring an additional operating system host.

VMware Workstation, Server and ESX use an optimized path to run target operating systems on the host compared to emulators, which simulate the function of each CPU instruction on the target machine one-by-one, or which offer a dynamic recompilation and a reuse of code, compiling machine-instruction blocks first time they run.

This makes the performance increase. At the same time, however, problems may arise when customers of virtual machines migrate between hardware hosts using different instruction or between hardware hosts with a different number of processors.

Xen Cloud Platform (XCP or) [12] is a virtualization solution that provides turnkey open source virtualization and out-of-the box cloud computing. XCP includes Xen hypervisor, the enterprise ready Xen API tool stack and integration of cloud solutions, storage and network. Additional functions are available in XCP: VM lifecycle: snapshots, checkpoints, migration, more resources for flexible storage and networking opportunity, opportunity for tracking events, upgrade and patching capabilities, performance monitoring in real time using templates based for Windows and Linux clients.

Kernel Virtual Machine (KVM) [13] is an open source hypervisor that provides enterprise-class performance. KVM advantages are: high scalability, high security, the possibility of operating on Windows and Linux on x86 systems.

Compared with other solutions, KVM provides a cost-effective alternative allowing cost reduction, high scalability.

KVM is a solution proposed by IBM and is one of the best solutions in its category.

Virtual machines that are based on KVM technology are the latest step in the evolution of open source virtualization technologies. KVM can use Linux for many of the things that a hypervisor has to do, like for example, planning tasks, memory management and interactions with hardware devices.

KVM creates virtual machines that can be opened by guest clients of operating systems like Linux or Windows, and it uses a modified version of another open source element - QEMU - to provide I / O device for emulation inside the virtual machine.

In conclusion, KVM is able to run efficiently for Windows and Linux virtual machines.

VirtualBox

VirtualBox is part of the most popular cross-platform virtualization software in the world, allowing one to run multiple operating systems.

VirtualBox (Sun) offers a free version and a subscription enterprise version. The free version allows only the use of a 32-bit operating system (as guest OS), while the paid edition allows a 64-bit guest operating system. It also has a free version of open source, but that is difficult to install in Windows. Virtualbox is available for all operating system platforms. A virtual machine created in a single operating system can be used in others as well. It is possible to transform a Virtualbox virtual machine created in VMW and vice versa.

2.4 Disclosure index computation – The simple cardinal valuation method

The decision maker establishes relationships between criteria ordered according to their importance, which progressively increase the associated share value. The evaluation algorithm is:

Criteria are sorted in ascending order based on their increasing importance. Let this order be C1, C2, ..., Cn (C1 being the least important criterion).

We attribute value x, $w_1 := x$ to the weight of criterion C1

The decision maker appreciates to what extent criterion C2 is more important than C1 by

subjectively determining the value of the ratio $\Delta w_2 = \dfrac{w_2}{w_1}$ ($\Delta w_2 \geq 1$, and the equality occurs if one considers that C1 and C2 are of equal importance).

The method is applied for each criterion by comparing it to the previous one. Thus, Δw_j is determined, for $j = 2,3,...,n$

$$\left(1 + \Delta w_2 + \Delta w_2 \cdot \Delta w_3 + \Delta w_2 \cdot \Delta w_3 \cdot \Delta w_4 + ... + \Delta w_2 \cdot \Delta w_3 \cdot ... \cdot \Delta w_{n-1} \cdot \Delta w_n\right) \cdot x = 1$$

The weights of the n criteria are calculated:

$w_j = w_{j-1} \cdot \Delta w_j$, with $j = 2,3,...,n$ and

$$w_1 = x$$

We note that $\displaystyle\sum_{j=1}^{3} w_j = 1$.

Table 4. The weight of each criteria

Criteria	Importance	w
Supported Hosts	x	0.25
Installation	1.2	0.3
Supported Clients	1.5	0.45

$$Final \quad score = \sum_{j=1}^{3} w_j \cdot Jaccard \quad index_j$$

Table 5. Final score

	Supported Clients	Installation	Supported Hosts	Final score
Cloud Foundry	1	0.5	0.833	0.808333
Cloud.com / CloudStack	1	0.5	1.000	0.85
Eucalyptus	1	0.5	0.833	0.808333
OpenNebula	1	0.25	1.000	0.775
OpenQRM	1	0.25	1.000	0.775
OpenShift	1	0.5	0.833	0.808333
OpenStack	1	0.75	0.833	0.883333
oVirt	0.5	0.5	0.500	0.5
PetiteCloud	0.5	0.75	0.833	0.658333
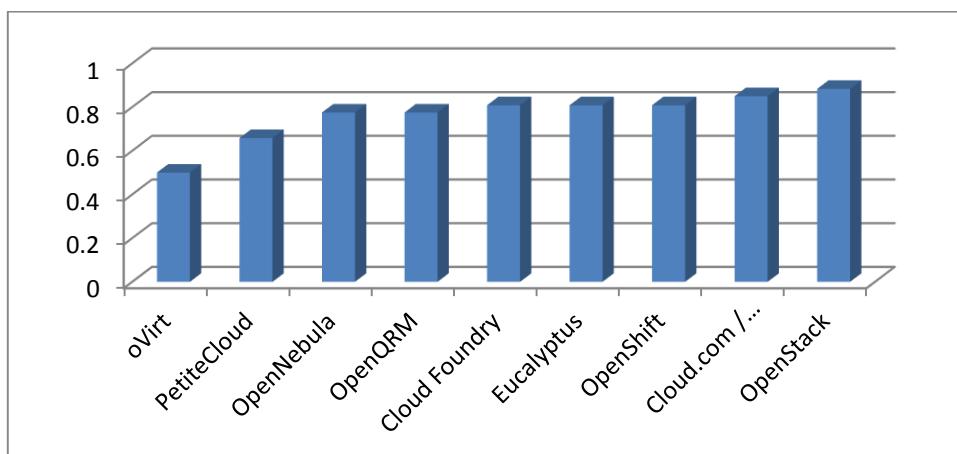
Fig. 1. Final score

3 Case Study - the use of OpenStack

In the following case study, we present a company from Brasov, Romania, which offers cloud services. At present, the company only uses VMware for all cloud solutions.

Physical servers are distributed across multiple sites (two in France and one in Romania). On the whole, there are about 13 "blade" servers. The servers are dual core CPU 8 (i.e. 16 physical cores per server, 32 virtual enabled

Hyperthreading) and they each have 128 GB of RAM.

Physical machines are standardized (they have the same characteristics and are produced by the same manufacturer).

One site in Romania and one in France have the same server model. The second site in France uses other equipment, having completely different servers, storage, network etc. The data centers are interconnected by high speed leased lines (MPLS L2).
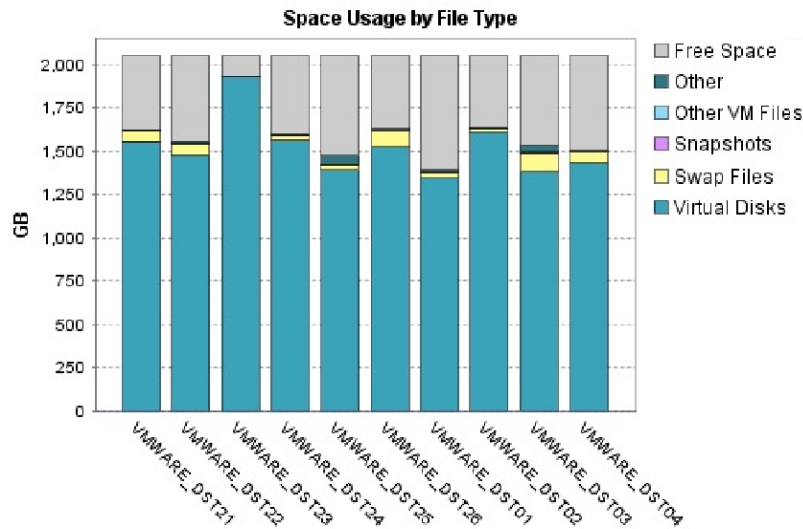

Fig. 2. Space usage by file type

If one considers a fair allocation of resources, from our perspective they are not all optimized. There are some virtual machines that have an over-allocation of resources, but some others have a sub-allocation of resources (CPU and RAM rule). Sub-allocation creates problems with the performance on individual virtual machines. Over-allocation can be achieved at the virtual platform (cluster) level, when more virtual machines are installed than physical resources, which makes virtual machines start "fighting" for resources...

A virtual machine can be allocated at one time on just one physical server. But, it can migrate from one server to another, live, at any time. There is indeed the possibility to put a virtual machine on two physical servers using the VMware Fault Tolerance function. This function is used when we want 100% availability for a virtual machine or a service on a virtual machine

The best virtual machine that this company has in production at present has 8 CPU, 16GB RAM and 1TB HDD. However, the company made tests on machines with more resources.

For "Self-service" for the company's users, we use a product developed by VMware, called vCloud Director. It is a tool devoted to final users, which they can use to manage the resources of the cloud. It can create virtual machines based on templates provided by a catalogue, can modify resources on existing machines, can make start / restart / stop operations, has access to the local console of the machine using a simple browser. It can also manage the security area for its resources; in vCloud, there is a virtual firewall that protects virtual machines
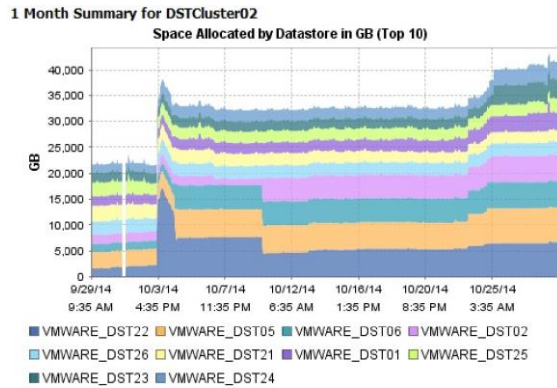
Fig. 3. Space allocated by data store in GB

Security is a complex topic and is achieved on several levels. There, one can distinguish:
- Physical security: access to servers, storage system (the data in the cloud)
- Virtual machine security: physical network firewall, virtual firewall in a "virtual data centre", IPS and SSL VPN connections etc.
- Data security on virtual machines: backup solutions, antivirus solutions, antispam solutions etc.
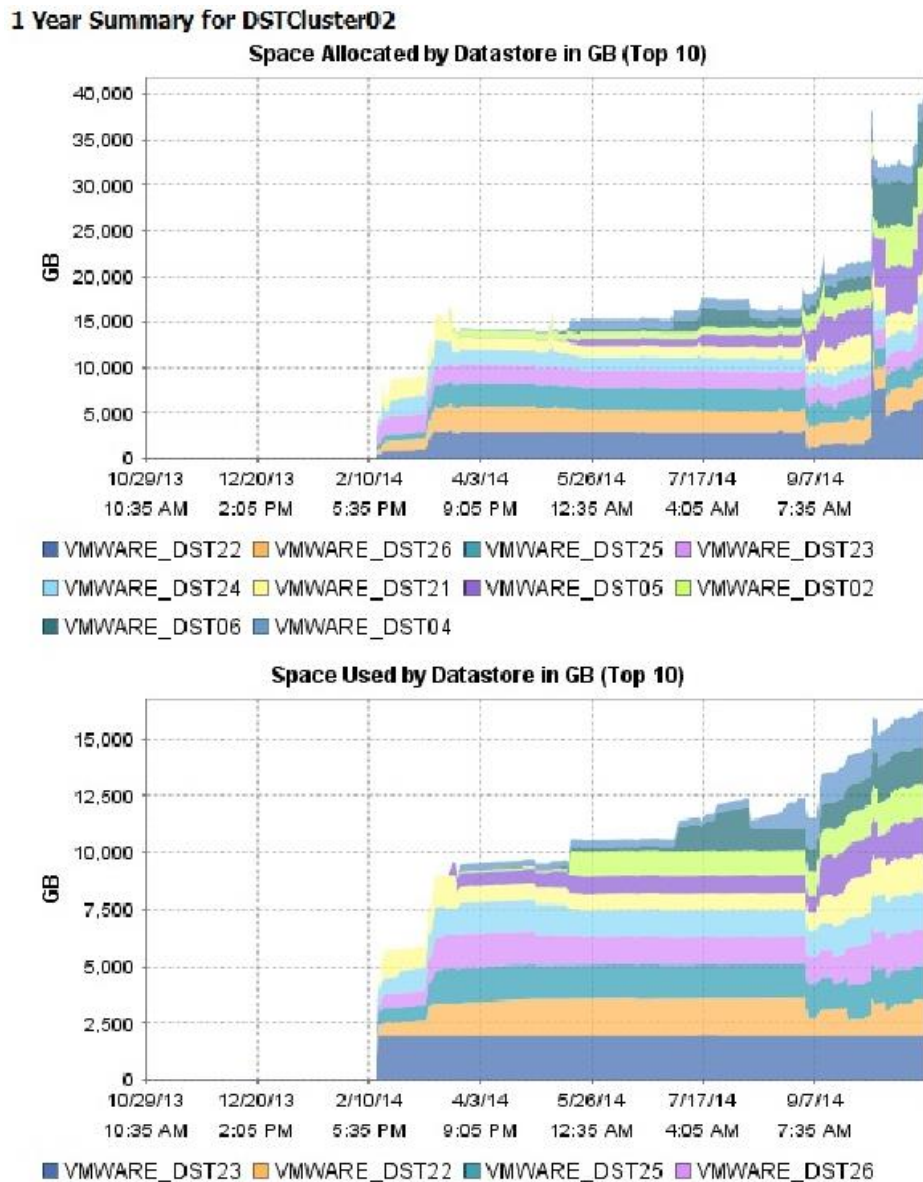

Fig. 4. Space allocated & used by data store in GB

As regards sales, this Romanian company mostly focuses on SaaS technologies. They are preferred by customers, especially by non-technical ones, because they do not need technical knowledge to manage their business platform in the cloud.
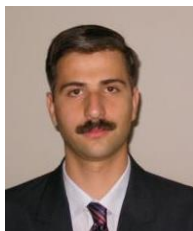
Prices vary significantly, depending on the quantity purchased and on the time the contract covers.

4. Conclusion

In this article we tried to make a comparison model of free licensed cloud platforms because the choice of a good platform leads to: great flexibility – it is able to optimize the IT infrastructure costs of a company; security - data in the cloud are more secure than those managed on traditional systems; high management capacity - with cloud computing, the management of a company can be accomplished by a smaller number of employees working more efficiently; small costs - cloud technology reduces maintenance fees. There are no longer multiple servers, software and upgrade fees. Many of the hidden costs typically associated with software implementation, customization, hardware, maintenance, and training are included in a transparent subscription fee; open projects - Internet and web services standards allow the connection to various services. This means information can be decentralized and can be accessed from anywhere in the world, from any computer or mobile device, at any time.

References

[1] S. Niwattanakul, J. Singthongchai, E. Naenudorn, S. Wanapu, "Using of Jaccard Coefficient for Keywords Similarity" in *Proceedings of the International MultiConference of Engineers and Computer Scientists 2013* Vol I

[2] R. Real, J.M. Vargas, "The Probabilistic Basis of Jaccard's Index of Similarity", http://sysbio.oxfordjournals.org/content/45/3/380.full.pdf+html

[3] A. N. Albatineh, M. Niewiadomska-Bugaj "Correcting Jaccard and other similarity indices for chance agreement in cluster analysis", Advances in Data Analysis and Classification, 2011, Volume 5, Issue 3, pp 179-200

[4] http://searchcloudcomputing.techtarget.com/definition/Software-as-a-Service

[5] http://www.pragmaticmarketing.com/resources/the-economics-of-software-as-a-service-saas-vs-software-as-a-product

[6] http://www.get-bynder.com/en/knowledge/our-blog/software-as-a-product-vs-software-as-a-service/

[7] https://www.freebsd.org/

[8] http://www.anhosting.com/blog/2012/10/top-five-benefits-of-using-linux-web-hosting/

[9] http://www.orcsweb.com/blog/windows-cloud-server-hosting-benefits-saas-hosting/

[10] http://www.ubuntu.com/cloud

[11] https://www.linux.com/learn/whitepapers/doc/8/raw

[12] http://www-archive.xenproject.org//cloudxen.html

[13] http://www-03.ibm.com/systems/kvm/

[14] http://www.sysprobs.com/why-to-use-vmware-virtualbox-virtual-pc-other-desktop-virtualization-software-advantages

[15] http://www.verio.com/resource-center/articles/cloud-computing-benefits/

[16] http://www.salesforce.com/uk/socialsuccess/cloud-computing/why-move-to-cloud-10-benefits-cloud-computing.jsp

[17] http://cloud4good.com/why-cloud/

[18] http://www.networkworld.com/article/2231739/tech-debates/who-has-the-better-virtualization-platform---vmware-or-microsoft-.html

[19] https://cloud.oracle.com/home

[20] http://cloudhostingmagazine.co.uk/Articles/index.php?mag=Cloud&page=compDetails&link=4097&cat=Cloud

[21] http://www.asugnews.com/article/sap-hana-cloud-platform-vs.-hana-enterprise-cloud-faq-definition

Radu LIXĂNDROIU Assoc. Prof., Economist, Ph.D. in Cybernetics, Statistics and Economic Informatics at the Academy of Economic Studies Bucharest. Post-graduate studies at the Faculty of Economic Sciences, Transilvania University of Brasov, specialized in Marketing Systems. Teaching and research activity in the fields of Databases, eCommerce, Accounting Software. Author or co-author of more than 40 papers in prestigious Romanian and foreign specialized publications. Author or co-author of 3 specialized books; team member in 7 scientific research contracts.

Cătălin MAICAN Assoc. Prof., Economist, Ph.D. in Cybernetics, Statistics and Economic Informatics. Teaching and research activity in the fields of Databases, Windows and Web Programming. Author or co-author of more than 50 papers in prestigious Romanian and foreign specialized publications. Author or co-author of 6 specialized books; coordinator of a CNCSIS (National Council for Scientific Research in Higher Education) grant and team member in 8 scientific research contracts.