# Virtual computers and virtual data storage

Conference Paper · January 2013

**2 authors:**

Alen Šimec
University of applied sciences Zagreb
**19** PUBLICATIONS   **21** CITATIONS

SEE PROFILE

Ognjen Stanicic
Polytechnic of Zagreb
**2** PUBLICATIONS   **2** CITATIONS

SEE PROFILE

# Virtual computers and virtual data storage

Alen Šimec, Ognjen Staničić
Tehnical Polytehnic in Zagreb/Vrbik 8, 10000 Zagreb, Croatia
alen@tvz.hr, ognjen.stanici@tvz.hr

**Abstract – Virtual data storage represents a new business model which includes various concepts such as virtualization, design of distributed applications and management which enables flexible data access. These methods include using networks of remote servers instead of local servers and personal computers for storing, managing and editing data. Locations containing servers which execute applications and store data are not strictly defined, hence terms "virtual data storage" or "storing data in the cloud" are used. As the need for storing data increased, managing that data became harder as well. Backing up data in large organizations is an inconvenient task. In spite of the increase in power and storage capacity of computers, prices of storing and maintaining data remains high. Various technologies and solutions have developed over time to overcome this problem and in the end they evolved into virtualization of the system for data storage.**

## I. INTRODUCTION

Virtual data storage represents a new business model which includes various concepts such as virtualization, design of distributed applications and IT management which enables a flexible approach in application scaling and placement. Virtual data storage represents a pay per use model which enables simple and accessible access to computer resource groups via the internet. On the basic level, virtual data storage is a simple means of providing information technology resources as services. Virtual data storage includes using networks of remote servers instead of local servers and personal computers for storing, managing and editing data. Locations containing servers which execute applications and store data are not strictly defined, hence the term "in the cloud" is used. End users access programs via their browsers, whereas the software is stored on remote servers. Virtual data storage enables efficient resource usage, when needed only the required amount of resources is accessed. There is a current trend in the computer industry of developing programs which can be accessed by multiple users without installing anything on their computers. Virtual data storage is also useful if we want to increase the business capacity without investing in new infrastructure, training new staff or licensing new software.

For the end user, virtual data storage brings lots of benefits. Costs of hardware purchasing are eliminated, software licenses have been made obsolete and there is no need for hiring new employees. For these reasons virtual data storage represents a big improvement in IT evolution because it changes the business model. Access to data is unlimited and easily available from any location worldwide.

## II. CONCEPT AND DEFINITION

In the 1960s IBM introduces the concept of Virtual machines as means of providing simultaneous, interactive access to their mainframe computers. Every VM was an instance of a physical machine and provided users with an illusion of direct access to the physical machine. It was an elegant and transparent way of sharing resources and processor time of very expensive hardware. Every VM was a completely secure and isolated instance of the core system. Users could execute, develop and test applications without having to worry about crashing the system. Virtualization was used for decreasing hardware supply costs and improving productivity by enabling multi user access on the same computers. With the decrease in hardware prices and the appearance of multi-processor operating systems in the 1970s and 1980s, virtual machines have almost become extinct. The idea of virtualization reemerged in the 1990s with the appearance of a wide spectrum of PC hardware and operating systems. The main task of virtual machines is enabling execution of applications, designed for different hardware and operating systems on a certain machine. This trend is present today.[1]

In the computer world, applications and users perceive a virtual environment the same way they do a real one, even though core mechanisms are formally different. Virtual environments often

---

[1]

http://www.kernelthread.com/publications/virtualization

represent false images of machines or resources which have better or worse abilities than physical machines or resources. A typical computer system uses lots of such technologies. One example is usage of virtual memory in modern operating systems which provides processes with more memory than the actual amount of physical memory the computer provides. It also enables physical memory sharing amongst hundreds of processes. Another example is multitasking where CPU time is divided which achieves assigning each task its own virtual processor.

Virtualization is a way of separating programs and their core components from the hardware which supports them and presenting a logical or virtual view of these resources. This logical view can differ greatly from the physical view. The goal of virtualization is usually one of the following: higher levels of efficiency, scalability, reliability, availability, agility and creating a unique domain of security and management.[2]

## III. VIRTUALIZATION USAGE MODELS

The change in business from paper to computers, automation of business processes and the need for communication with clients caused the need for large numbers of computers and servers. The possibility of serving a greater number of systems on one physical server allows organizations to become data centers, which eliminates the costs of building one. Computer architects have to take into consideration various issues while changing from the standard model of distributed applications to a cloud based model. There are three basic virtualization usage models: public, private and hybrid clouds.

Private clouds are implemented for single user usage and they provide the highest data control, security and quality. Large companies with their own data centers are interested in private clouds (also called internal or corporative clouds) which provide additional optimization using the principle of virtualization or distributed computing. That way IT sectors become service providers for other business sections of the company. The company owns the infrastructure and has control over the execution of applications. The cloud can be implemented inside the company data center or in a remote location. This model enables companies control over resource usage, while a third party takes care of the required proficiency for establishing and managing the environment.[3]

Public clouds use internet access with hired resources. The complexity of implemented information and communication solutions is hidden from the user. The majority of management and system surveillance is done using browsers. Public clouds are maintained by independent service providers. Applications from different users are often mixed together on cloud servers, data storage systems and networks. A part of a public cloud can be separated for usage by a single user creating a virtual private data center. The virtual private data center is not limited to distributing images of virtual computers on the public cloud. It also provides users with a better insight in its infrastructure. This way, users can manipulate not only with images of virtual computers, but also with servers, data storage systems, network devices, and network topology. Creating a virtual private data center with all components placed at the same location helps with reducing data locality problems because network bandwidth is large and frequently free when connecting elements within the same location.

Hybrid clouds are combinations of public and private virtualization models. The hybrid model helps with enabling a system available on request, with the service being provided by a third party. Only a part of the programs, services and data are located in the cloud, while the rest of the IT system is located on the company infrastructure. The system structure allows end users undisturbed access to required data, regardless of the physical location on which they are located and the location of the business application which processes them.

## IV. VIRTUAL DATA STORAGE SYSTEM

Virtual data storage systems are used to join physical storage from various devices in order to create an appearance of a single data storage source. Storage servers are computer systems which allow general purpose systems safe access via user accounts. The operating system of the storage server is responsible in deciding which general purpose server is allowed access to which data on storage devices. If storage servers are connected to general purpose systems via special purpose networks, this configuration is often called Storage Area Network (SAN). Storage accessed via the network is called Network attached storage (NAS), regardless whether the general purpose system uses SAN or Local Area Network (LAN).[4]

[2] Kusnetzky, D. *Virtualization: A Manager's Guide,* O'Reilly Media, Inc., Sebastopol, 2011., pp 1

[3] Wolf, C. Virtualization: From the Desktop to the

Enterprise, Verlag New York, 2005., pp. 12

[4] Judith, R.; Davis, Eve, R.; Promotions, w. Data Virtualization: Going Beyond Traditional Data

Storage devices are based on rotating media such as traditional disks, solid state technology such as solid-state disks (SSD) or dynamic memory with direct access (DRAM). Storage can have several forms: direct attached storage (DAS), network attached storage (NAS) or storage area networks (SANs). It can be connected via several protocols: Fibre Channel, Internet SCSI (iSCSI), Fibre Channel over Ethernet or Network File System (NFS). Even though storage virtualization is not a prerequisite for server virtualization, one of the key consequences of storage virtualization is the ability to rely on "Thin provisioning".

Storage Area Networks are usually used in large data centers for enterprise resource planning applications (ERP) and they use expensive "fiber channel" links and switches.

Internet Small Computer System Interface (iSCSI) is a standard used for connecting data storage devices based on the IP protocol. It is a SAN protocol which allows organizations data storage consolidation into data storage arrays while providing hosts with the illusion of locally attached disks. This protocol is usually used while performing virtualization for starting and initializing virtual machines. ISCSI SAN can have efficiency problems which will prevent its usage in the virtualization architecture. Companies should balance Fiber Channel SAN and NAS with the managing complexity of the iSCSI protocol while implementing medium or large data storage systems.

Network File System is the preferred storage virtualization protocol because of its simplicity, scalability and inexpensive implementation and management. From the hardware aspect NFS is a "plug and play" system which uses network transport elements which are present in modern data centers.

## V. VIRTUAL DATA STORAGE FUNCTIONALITY

Data storage system virtualization creates an artificial view on the network which hides the physical network from the clients and the server. It manifests as a "layer" between physical and logical data storage processes. The layer is used for standardizing. Any physical solution for storage can be used in a logical solution, without the need for adjustment. Functionalities provided by virtual data storage are:

- Allowing distributed file systems: Remote storage devices are made to look as though

they are connected directly to the system. The local system is unaware of their location or the type of storage device being used.

- Creating artificial storage blocks: Multiple storage devices can be connected to create an illusion of a single, much larger storage device

- Creating block arrays for storage: Applications and data can be spread out over a large number of devices and storage servers in order to improve storage efficiency. This functionality can also be used for improving storage reliability. Same data can be stored on various devices or servers. If one of them fails, data can be reconstructed.

- Allowing better storage space management: Storage devices can be segmented into multiple file systems, allowing the storage device better usability.

- Allowing incompatible systems storage device sharing: Mainframe, Windows, Linux and UNIX all use different mechanisms for storing and accessing applications and data. Storage system virtualization allows all of them to share the same storage devices and files contained in them.[5]

## VI. STORING DATA "IN THE CLOUD"

Cloud storage is a network storage model in which data is stored into virtual data storage pools. Host companies control large data centers, and people needing to store their data somewhere buy or rent space from them. In the background, data center operators virtualize resources with regards to client's needs. Physical resources can extend over multiple servers. Cloud storage services can be accessed through an API or through a user interface.

The term "cloud" simply refers to going from local services to internet services, from local data storage to a safe and configurable environment, from using applications with limited space (in gigabytes) to applications without an upper limit, from using Microsoft Office to using web-based office. Between 2005 and 2008 network data storage has become cheaper and safer than local storage.

[5] Smoot, S.R.; Tan, N.K. Private Cloud Computing: Consolidation, Virtualization, and Service-Oriented Infrastructure. 2011.

Clouds offer access to cheap hardware and resources for storage using simple APIs and are based on pay-per-use models, hence renting these resources is much cheaper than looking for other solutions. Users are used to storing data remotely in clouds, so they gain popularity with engineers, small and medium sized companies and casual users.

## VII.EXAMPLES OF COMMERCIAL CLOUDS

The most famous commercial cloud service providers are Google, Amazon and Microsoft. With Google App Engine users develop their own web applications which are then deployed on the Google infrastructure. Applications can be written for execution in Java, Go or Python environment. Users write applications and deploy them onto the App Engine service using their Google App Engine accounts. The service is free with a limited amount of resources, whereas getting extra resources requires payment. Users can apply for the Google domain, or have their own domain. After deployment Google looks after the application. Their architecture grants resources to applications, looks after forwarding requests to them and gives them time for putting together responses. Google also offers additional services such as Google Cloud Storage which user applications can use.  It is estimated that Google App Engine stores 1 million active applications and has around 250.000 active users.

Infrastructure as a service (IaaS[6]) has improved with the development of virtual technologies and the appearance of server renting. The most known IaaS provider is Amazon, and the service is called Elastic Compute Cloud (EC2). Amazon EC2 is a real virtual computer environment which enables using web-service interfaces for running instances with different operating systems, loading them with an adjusted application environment and managing network access. In the App Engine users deploy their own web applications in the cloud, whereas in EC2 users pay for access to virtual computers which are run in the cloud. A user is granted a virtual computer which he can access through the ssh command line in case of Linux operating systems or through the Remote Desktop console in case of Windows operating systems. Prices depend on the strength of the computer and start from a couple of cents per hour using Linux and go to a dollar per hour using Windows with preinstalled additional software such as for example a database management system. Amazon Web Services (which

include Amazon EC2) hit revenue of around $500 Million in 2010.

Microsoft Azure is a typical example of a Platform-as-a-service (PaaS[7]) model, because it allows users application development on a remote computer (in the cloud). Windows Azure is a service which allows companies and individuals application deployment in Microsoft data centers, with payment for resources used. This approach absolves companies of infrastructure maintenance and allows them to focus on software development. Windows Azure enables savings in expenses for computer resources since users only pay for the resources they actually used. The platform is extremely easy to use and it is only necessary for a company to choose how much resources they want to reserve for a certain application. Windows Azure is completely scalable and allows simple increase in resource usage if the need arises as with an increased number of users of a web application. Visual Studio 2010[8] is used for developing Windows Azure applications. This way companies and individuals can develop applications on their own computers and deploy them on Windows Azure via integrated tools. The application appears instantly on the platform and users can start using it. Beside Visual Studio 2010, applications can also be developed in the Java Eclipse environment. Other programming languages are also supported such as PHP, Python and Ruby which allows interoperability of the platform. It is estimated that there are around 50 million active user accounts.

Skydrive is a personal cloud which provides users with immediate and protected access to private data from all devices with the possibility of sharing files and folders with others. Instead of using multiple services, SkyDrive users get a single service which allows them to access their data without the need to copy files from cloud to cloud in order to share or to search multiple data locations in order to find their files. Skydrive is used by 17 million active users which share their photographs privately or

---

[6] IaaS (*Infrastructure as a service*) servers deliver resources on demand, fetching them from large pools built in data centers.

[7] PaaS (*Platform-as-a-service*) facilitates the deployment of applications without the cost and complexity of buying and managing the underlying hardware and software. It provides everything neccessary for the application development nad delivery system to be available via the Internet. PaaS offers tools for application design, testing, implementation and server functionalities. It can also offer application services such as team collaboration, integrated database, security, scalability, warehousing, etc.

[8] A development tool completely adjusted for programming applications and services for the Windows Azure platform.

collaborate on Office documents. Currently Skydrive stores around a million gigabytes of user data and a big growth is expected in the future. These numbers are important because they show there is a tendency for Skydrive to become a service considered valuable by users and used daily.[9]

Cloud computing is definitely on the rise. A study done by Gartner estimates the cloud computing market is worth $150 Billion in 2013. Research also shows that 60% of server workloads will be virtualized by 2014, which compared to 12% in 2008 is a staggering increase. An independent research firm TNS surveyed more than 3500 cloud computing users in eight countries around the world. 93% of respondents said cloud computing had improved their data center efficiency and 80% saw these improvements within six months of moving to the cloud. 82% of respondents said they had saved money on their most recent cloud project. Another study surveyed 3258 companies showing areas of cloud usage, the two largest were accounting (20%) and payroll (20%). The rest were collaboration (17%), file/data storage and backup (15%), business class e-mail (14%) and CRM (14%).[10]

## VIII. CONCLUSION

All recent cloud computing forecasts show rapid adoption of cloud computing in enterprises. The cloud computing marketplace will reach $16.7 Billion in revenue by 2013, according to a new report from the 451 Market Monitor, which is a growth from $8.7 Billion in 2010, a compound annual growth rate of 24%. Forrester forecasts that the global market for cloud computing will grow from $40.7 Billion in 2011 to more than $241 Billion in 2020. Cisco predicts that Global cloud IP traffic will increase twelvefold over the next 5 years, accounting for 34% of total data center traffic by 2015. The growth of cloud computing will also have positive consequences on other parts of the IT industry such as mobile services and platforms.

---

[9] http://cloudcomputingtopics.com/2011/11/5-cloud-computing-statistics-you-may-find-surprising/

[10]

https://community.dynamics.com/product/crm/crm nontechnical/b/crmsoftwareblog/archive/2012/01/2 5/smb-cloud-adoption-study-statistics-show-that-cloud-usage-is-on-the-rise.aspx

## IX. REFERENCES

Columbus, L. Forecasting Public Cloud Adoption in the Enterprise, http://www.forbes.com/sites/louiscolumbus/2012/0 7/02/forecasting-public-cloud-adoption-in-the-enterprise-2. 2013.

Judith, R.; Davis, Eve, R.; Promotions, w. Data Virtualization: Going Beyond Traditional Data Integration to Achieve Business Agility, 2011

Kusnetzky, D. *Virtualization: A Manager's Guide.* O'Reilly Media, Inc., Sebastopol, 2011.

Singh, A.. An Introduction to Virtualization. http://www.kernelthread.com/publications/virtualiz ation. 2013

Smoot, S.R.; Tan, N.K. Private Cloud Computing: Consolidation, Virtualization, and Service-Oriented Infrastructure. 2011.

Toomey, L. 5 Cloud Computing Statistics. http://cloudcomputingtopics.com/2011/11/5-cloud-computing-statistics-you-may-find-surprising. 2013.

Wolf, C. Virtualization: From the Desktop to the Enterprise, Verlag New York, 2005.