Cloud Computing and Services

Briefly Introduction

Cloud computing has indeed arrived, or rather returned, in the last decade to revolutionize the world of computing. Chronologically, there was the appearance in the 1980 virtualization, outsourcing and outsourcing; also the democratization of computing in the 1990 and especially during the last decade with the generalization of the Internet, the development of high-speed networks, application rental, payment to the use and quest for mobility, all this gave birth to the new concept infrastructure and IT solutions named cloud computing. The cloud consists of the interconnection and cooperation of IT resources, located within the same entity or in various internal, external or mixed structures. And whose fashions access is based on Internet protocols and standards. Cloud solutions are based on virtualization and automation technologies. Three key characteristics of the Cloud differentiate it from traditional solutions: first, “services instead of technological products with continuous and automatic updating”; second, “self-service and pay-per-use (depending on what you consume)”, and the last one, “pooling and dynamic allocation of capacity (elastic adaptation to peak loads). In the world of information technology, companies of all sizes and in all industries are less careful of the different online solutions on the market. Some of these companies have already migrated their data and services, such as emails, CRM and ERP applications or even online storage solutions to cloud services like Google Apps, Apple Apps, Dropbox, Microsoft OneDrive, Amazon AWS, and many more services. Also, several new applications are based on infrastructures or platforms offered in the form of the cloud such as, for example, Amazon AWS, Google Cloud or Microsoft Azure. The business case for these types of solutions is very satisfactory as far as savings and investment are concerned because cloud service providers will benefit from savings of scale in order to offer prices barely matched by traditional solutions. However, currently, cloud service customers have no way of verifying the complete confidentiality and integrity of their data and processing.

Cloud computing is not a new technology, but rather a new way of using current technological resources. Indeed, the Cloud is the provision of technological services of all kinds, such as e-mail, storage, office tools and so on, immediately and on-demand. This type of service allows the use of technological means unseen until now because of its flexibility. The main characteristics of cloud computing are as follows: high level of abstraction, on request, cost reduction resulting from economies of scale, flexibility and scalability, pay as you go, almost instantaneous disposition [4].

According to the main functionalities of the service " in the cloud", there are three models:

1. *Infrastructure as a Service (IaaS):* concerns servers, equipment and solutions of storage, network. The IaaS model consists of having an IT infrastructure available through a cloud computing deployment model. Access to the resource is complete and unrestricted, equivalent to the provision of an actual physical infrastructure. Thus a company can for example rent Linux-based OSs, Windows or Mac systems, which will run in a virtual machine at the IaaS vendor. The user of this model does not have control over the underlying technology infrastructure, but it does have control on network configuration, operating system and applications. In this model, the responsibility for safety is almost entirely transferred to the customer [1]. Example of IaaS services: Amazon Web Services (AWS), Linode, Google Compute Engine, Microsoft Azure.

2. *Software as a Service (SaaS):* concerns business applications: CRM, collaborative tools, messaging, BI, ERP. The SaaS model allows an application to be outsourced to a third party. This model is suitable for certain categories of applications that must be globally standard for everyone, standardization is one of the principles of the Cloud. The term SaaS does evoke a service in the sense that the supplier sells an operational function, and not technical components requiring computer skills. This is the model in which the user transfers almost the entire responsibility for IT security to the provider of services [1]. Example of SaaS services: Dropbox, Google Apps, Apple Apps, Apache Stratos, ZenDesk, Salesforce, BigCommerce.

3. *Platform as a Service (PaaS):* concerns middleware, development, test. The PaaS model consists of providing a ready-to-use environment, the infrastructure is hidden. For example, a PaaS platform makes it possible to have an immediately available development environment. However, changes to the configuration operating system, network configuration or storage system are not allowed [1]. Examples of PaaS services: Microsoft Azure, AWS Elastic, Open Shift, Heroku.

If the cloud service customer contracts just a few levels of the stack at the bottom, the provider de cloud will not be responsible for the security of the systems used by the customer. Conversely, if the customer rents multiple layers covering many levels of the stack, the risk will be transferred to the service provider. The above is the key point of security management in cloud computing models. The main disadvantage of transferring IT security risks to suppliers is that it will limit flexibility and functionality in contracted Cloud services. Another aspect of cloud computing service levels is the different interfaces that appear at the top of each model.

APIs (in IaaS), Integration & Middleware (in PaaS) and modality & Presentation Platform (in SaaS) are levels that act as an intermediary diary between the user's client (web browser) and the cloud computing service provider. However, Cloud Computing represents a major challenge for service providers, for its customers and, of course, for external attackers [1]. Thousands of gigabytes of information from different clients are stored in one place and exposed to the Internet. This represents feels like an ideal target for hundreds of malicious users, for example, industrial spies: imagine how attractive are the customer data of thousands of companies exposed to the Internet.

Security is an aspect that must be taken into account when deploying a cloud. The adoption of cloud-based technologies are inevitable, but care must be taken into account to prevent potential gains for adopting from being tarnished by a security threat. Computer security plays an important role in the deployment of cloud computing. In fact, security is the biggest challenge for IT managers, recent research finds who wish to adopt solutions and services hosted in the cloud. In a short phrase cloud computing is a model of providing infrastructure, platform and software services on the internet [5]. Lack of standardization of cloud service descriptions heterogeneous makes the discovery and selection of services very complex for users from the cloud. To alleviate this complexity, it is essential to describe the different information relevant to cloud service in a seamless model. And it is necessary to automate the mechanisms for discovering and selecting the appropriate cloud services to respond to the functional and non-functional needs of the user. These two problems constitute the issues studied in our work. To solve the first problem, namely the standardization of the description of heterogeneous cloud services, we built an ontology, which plays the role of a vocabulary operated by the main elements of the cloud system. In the event that the service desired by the user does not exist, we have proposed two methods for discovering services, the first is the search based on the name of the service. The second is based on selection functional properties, the latter offers a set of services with the same functional properties of the desired service. The latter represents the second contribution. And finally, we presented our third contribution, which consists of generating the different composition plans, taking into consideration the different relationships and dependencies between business services. The proposed approach has been evaluated according to several scenarios and the results obtained are encouraging and promising [2].

Cloud systems automatically control and optimize the use of resources by exploiting a measurement capacity corresponding to an appropriate level of abstraction of the type of service (for example, storage, processing, bandwidth, and active users)[4]. Resource use can be monitored, controlled and reported, offering transparency to both the supplier and the consumer of the service used. Cloud computing allows the customer to pay per use, only for what has been consumed depending on the type of services like storage, processing, bandwidth and active user accounts, and service usage time [2]. Virtualization is used to generate a simulated physical system on an actual physical system. It allows the use of a virtual IT resource from a machine real physical. In most cases, there are multiple simulated physical

systems. It is in this sense that virtualization is used to create a density of systems. We can have multiple virtual systems, called virtual machines, running on the one only physical system. These virtual systems share the use of physical resources such as a processor, a network interface or a hard disk, these are allocated to a virtual machine so that it functions as a physical machine [5]. When a virtual system does not use the resources of a physical system, those resources can be used by another virtual system. In a non-virtualized environment, system resources may be inactive for a while. Exist two types of virtualization: full virtualization and paravirtualization [3]. The difference between them is the fact that full virtualization allows multiple guest operating systems to execute a task on a host operating system independently of each other, while paravirtualization allows multiple operating systems to run on a host simultaneously communicating with the hypervisor for performance improvement.

Clusters and Cloud Computing

The goal of cloud computing is to build a cloud of clusters to interconnect a set of machines on a defined network. Users can then deploy virtual machines in this cloud, which allows them to use a number of resources (disk space, RAM, or CPU) [6]. This infrastructure, going into more detail, is made up of clusters and nodes. The clusters are used to manage the interface between nodes and the user. Thus, when deploying a virtual machine on a cluster, the cluster will create an instance, which will materialize through the useful resources in nodes [5]. Data protection is difficult for cloud computing service customer. It is very difficult to secure data that is spread across multiple locations. Make sure that the data is processed correctly is also complicated because the control over transfers data is beyond the reach of its owner [5].

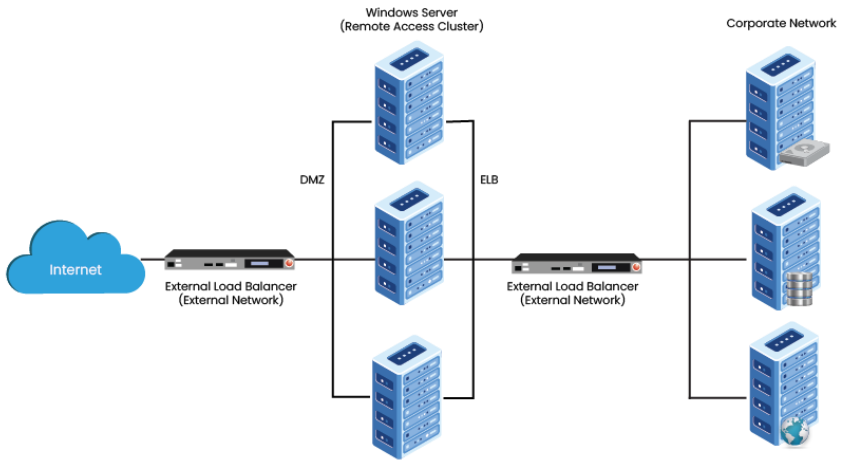

Figure 1: Clustering

Although basic cloud services have been offered since some time ago, the reality is that it is now expected to experience significant growth in both the number and the quality of services, providing organizations with the possibility not only of increasing their efficiency but also to develop new businesses for the that the existence of the cloud is essential.
In addition to the already existing agents and gradually in the function of the evolution levels of cloud computing is foreseeable that new agents associated with the needs of customers, who will demand the presence of experts in certain subjects that did not exist before the emergence of cloud services. For example, the cloud services broker will act as an intermediary between cloud service providers and end-users to speed up and facilitate the transition to the cloud. In the same way, the cloud presents itself as an opportunity for all those companies that have traditionally dedicated to information technology and what they see in the generalization of the use of cloud services a way of development to increase your sales volume through the provision of services in the cloud, doing what they have always been dedicated[6].
Faced with this change in the way of providing IT services, traditional operators are faced with the challenge of improving and restructure their infrastructures, taking decisions that allow them to maximize profits taking advantage of the increase in bandwidth that requires these services. Additionally, they are presented with the opportunity to increase the volume of business in new generation mobile communications (very demanded as the use of cloud services increases generalize) and increase their range of activity, beginning to offer services based on its own infrastructure by what they already know (network design and/or service provision) and innovating and accessing new businesses that are emerging every day.

There are logically limitations in the development of the business from the cloud.
The leap of companies to this type of services has high legal implications and associated risks

such as may be changes of the country conditions established in the service provider agreements and the data protection. The latter represents a theme of extreme sensitivity for organizations that hire cloud services, which will certainly require guarantee the security of your confidential data from the moment have your record moved to the physical location of the servers from the service provider [7]. The above question represents a critical point in those countries that have legislation especially sensitive to data protection. In this sense, services have been developed that allow the client to select the physical location to store your data and thus adhere to the current legislation of the country in which found registered.
In the same way, it is possible that certain agreements of provision of cloud services can generate situations in which the infrastructure, the situation of the servers or the own organizations involved belong to different countries that, in principle, may have nothing in common. You also have to take into account the requirements of investment in the development of network infrastructures such as another of the main limitations when it comes to achieving a 100% cloud scenario. The adoption of the cloud will generate bandwidth demands much higher than current, being able to saturate the access networks in case that the relevant investments were not made [7].

Data encryption: applications and limitations

One of the peculiarities of data security and processing in the cloud is the differentiation between static data and so-called moving data. Indeed, the practice of encrypting static data is different from using cryptography to protect data by transmission or in motion. Indeed, the particularity is that the encryption keys must be ephemeral, while for the static data, keys can be kept as long as the stored data is kept encrypted. Most of the data stored on the Internet are intended for use by users, it is primarily intended for use by other computers. Encryption keys cannot be stored by people. If data is stored on the same computer, or at least, the network where the data resides, represent a security risk [5]. Data security risks in the cloud do not fall outside the scope of IT risks globally, but they present vulnerabilities and should be considered when considering data security. These risks include phishing, privileged access in the cloud, and the source or origin of the data itself.

One of the indirect risks to data outsourced to a cloud is phishing. Although generally considered impossible, today, to break the PKI or public key infrastructure and therefore break the authentication and encryption, it is possible to trick end users by the misappropriation of their credentials in the cloud. Also, even if phishing does not a new risk in the world of security, it represents an additional threat to cloud security [8]. As a short example, many Google services such as Apps / Docs / Sheets randomly prompt users to re-enter their passwords, especially when a suspicious event has been observed. By the way, many Google services display the IP address from the previous login session with automatic notification of a

suspicious event, such as the login from an Asian server shortly after an IP address from Europe has been logged in for the same account. Another potential risk to data security in the cloud relates to inappropriate access to sensitive customer data by cloud personnel [8]. Clearly stated, outsourced services, whether cloud-based or not, can bypass the typical controls that IT organizations generally enforce through physical and logical controls. This risk of two main factors: first, there is unencrypted data and second, there is the personnel of the privileged access cloud service provider. The assessment of this risk fully involves modifying the practices and security standards of the cloud service provider so that this provider's staff with privileged access cannot access customer data. The High-Availability and Integrity Layer is a distributed cryptographic system which allows a set of servers for proving to a client that a stored file is intact and fully recoverable. It strengthens, formally unifies, and rationalizes distinct approaches to the crypto community and distributed systems. The evidence in HAIL is efficiently processable by servers and very compact, typically a few tens or hundreds of bytes, regardless of file size. HAIL cryptographically verifies and reassigns file shares again. It is robust against an active, mobile adversary which can progressively corrupt all files of the servers. It offers a strong and official contradictory model for HAIL, and rigorous analysis and choice of parameters [8].

Classification of parallel applications

A parallel application is either composed only of a single task or of a set of tasks performed using multiple resources. The model used to represent a parallel application depends on the tasks that compose it and the interactions possible between them. Generally, two types of tasks can be distinguished: first, sequential tasks: a task is said to be sequential if and only if it does not can be performed by only one resource; and the second, parallel tasks: a task is said to be parallel if and only if it can be executed by multiple resources. Two main categories of parallel applications can be distinguished according to the tasks that compose them and the interactions between them, namely:

1. The first type of parallel applications refers to the case where the tasks defining a given application can run independently of each other. These types of applications include those that are made up of parallel tasks and those that are defined by sequential tasks. There is a third type of applications called parametric applications (parameter sweep application). In the latter case, the same application is executed several times with different values of its parameters (the input parameters of this application can be varied) [9].
2. The second type of parallel applications refers to the case where the tasks constituting the application in question cannot be run independently and are linked by precedence constraints representing time constraints or data exchange. For this type of application, there are two main categories:

(a) The first category refers to applications defined by tasks that interact by exchanging data during their executions. These applications can be modelled using undirected graphs called task interaction graphs (TIGs). The vertices of these graphs represent the tasks and stops represent communications between resources [9].

(b) The second category refers to applications defined by linked tasks by precedence constraints. In other words, a task of these applications can only be executed if all of the preceding tasks have been completed. These applications are generally represented by directed graphs cycle-free (Directed Acyclic Graphs (DAGs)). The vertices of these graphs represent the tasks (parallel or sequential) of the application to be executed and the arcs define the precedence relations between these tasks (dependencies time or data)[9].

The problem of resource allocation and scheduling of applications composed of independent tasks is the most studied compared to the other two types of applications. Indeed, many studies have been proposed in the literature for the scheduling of independent tasks[10]. However, most of the work uses homogeneous platforms. Independent task scheduling algorithms generally aim to minimize the overall execution time or the makespan of the application to be executed or to maximize the throughput (throughput) of the platform used if the tasks do not arrive. One of the fundamental problems that can be decreasing the performance of these applications is the mismanagement of resource use. It is therefore essential to develop methods and tools allowing the use of efficient use of available resources, almost all cloud services provide various plans with vast resources for developing our project.

Dedicated Clusters for high-traffic applications and large datasets

• Additional hardware configurations available for specialized workloads

Tier	RAM	Storage	vCPU	Base Price
M30	8 GB	40 GB	2 vCPUs	from **$0.54**/hr
M40•	16 GB	80 GB	4 vCPUs	from **$1.04**/hr
M50•	32 GB	160 GB	8 vCPUs	from **$2.00**/hr
M60•	64 GB	320 GB	16 vCPUs	from **$3.95**/hr
M80•	128 GB	750 GB	32 vCPUs	from **$7.30**/hr
M140	192 GB	1000 GB	48 vCPUs	from **$10.99**/hr
M200•	256 GB	1500 GB	64 vCPUs	from **$14.59**/hr
M300•	384 GB	2000 GB	96 vCPUs	from **$21.85**/hr
M400	512 GB	3000 GB	64 vCPUs	from **$22.40**/hr
M700	768 GB	4096 GB	96 vCPUs	from **$33.26**/hr

Figure 2: Example of MongoDB cloud database various cluster selection

The optimization criterion taken into account by the algorithms aimed at scheduling a set of applications sharing the same resources is the makespan or the throughput of the platform of resources used. This is a difficult problem because you have to ensure fairness between different applications running in parallel. Scheduling no fairness can lead to starvation at worst [10]. In other words, a sub all the applications concerned will never have access to the available resources, some tasks will be indefinitely delayed. Automated tasks are performed by virtual machines [10]. Given one of the main characteristics of cloud computing, namely "illusion of infinite resources", the assignment of tasks at the level of virtual machines does not arise. Indeed, each time a task is assigned to a virtual machine, an instance of the latter is created. In what concerns the number of virtual machines, made available to users, they are supposed to be in infinite number. However, it should be pointed out that the number of types of virtual machines is finite. An interesting idea would be to separate these two parallels. In addition, the tasks constituting a process given are assumed to be indivisible in the sense that a task is executed by one and only one machine the fact that a given task can be divided and therefore the parts, as well as results, can be processed by separate resources.

References

"Architecting the Cloud: Design Decisions for Cloud Computing Service Models", Chapter 2, De Michael J. Kavis [1]

“Takwa Mohsni, Zaki Brahmi, and Mohamed Mohsen Gammoudi. Data-intensive service composition in cloud computing: State-of-the-art In *2016” IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA)*, pages 2–6. IEEE, 2016 [2]

Tim Abels, Puneet Dhawan, and Balasubramanian Chandrasekaran, “An overview of Xen virtualization. Dell Power Solutions”, 8:109–111, 2005 [3]

“Cloud Computing: Principles, Systems and Applications”, Nick Antonopoulos, Lee Gillam, 29-30, 2017 [4]

“An Analysis of the Cloud Computing Security Problem”, Mohamed Almorsy, John Grundy, Ingo Müller,5, 2016 [5]

S. T. Maguluri, R. Srikant and L. Ying, "Stochastic models of load balancing and scheduling in cloud computing clusters," 2012 Proceedings IEEE INFOCOM, 2012, pp. 702-710, doi: 10.1109/INFCOM.2012.6195815. IV-V [6]

S. Nurcan and F. Daoudi, “A benchmarking framework for methods to design flexible business processes, Improvement and Practice Journal on Business Process Management, Development and Support”, 2007 [7]

K. Gai, M. Qiu, H. Zhao and J. Xiong, "Privacy-Aware Adaptive Data Encryption Strategy of Big Data in Cloud Computing," 2016 IEEE 3rd International Conference on Cyber Security and Cloud Computing (CSCloud), 2016, pp. 273-278, doi: 10.1109/CSCloud.2016.52 [8]

A. Radulescu, C. Nicolescu, A. J. C. van Gemund, and P. P. Jonker, CPR : Mixed Task and Data Parallel Scheduling for Distributed Systems, In Proceedings of the 15th International Parallel and Distributed Processing Symposium (IPDPS’01), San Francisco, CA, USA, April 2001. IEEE Computer Society. [9]

P. Shroff, D. W. Watson, N. Flann, and R. Freund, Genetic Simulated Annealing for Scheduling Data dependent Tasks in Heterogeneous Environments, In Proceedings of the 5th IEEE Heterogeneous Computing Workshop (HCW’96), pages 98-117, Honolulu, Hawaii, April 1996 [10]