



ORIGINAL ARTICLE

OPEN ACCESS

THE USE OF CLOUD COMPUTING AS AN AID TO SCALABILITY AND AVAILABILITY OF ONLINE SERVICES

A UTILIZAÇÃO DA COMPUTAÇÃO EM NUVEM COMO AUXÍLIO À ESCALABILIDADE E DISPONIBILIDADE DE SERVIÇOS ONLINE

EL USO DE LA COMPUTACIÓN EN NUBE COMO AYUDA A LA ESCALABILIDAD Y DISPONIBILIDAD DE LOS SERVICIOS EN LÍNEA

Marlon Tavares França^{1*}, Audrey Teles dos Santos², Igor Dias Costa de Jesus³, Samuel Molendolf Teixeira⁴, Wagner Azis Garcia de Araújo⁵, & Luan Diego de Lima Pereira⁶

¹²³⁴⁵ Instituto Federal do Norte de Minas Gerais. ⁶ Instituto Federal da Bahia

^{1*} marlontf@gmail.com ² audreytelesdossantos@gmail.com ³ diascostaigor@gmail.com ⁴ samuelmolendolf@hotmail.com
⁵ wagner.araujo@ifnmg.edu.br ⁶ luan.diego@ifba.edu.br

ARTICLE INFO.

Received: 14.03.2023

Approved: 10.04.2023

Available: 17.05.2023

KEYWORDS: Technology; services; scalability; availability; containerization; cloud computing.

PALAVRAS-CHAVE: Tecnologia; serviços; escalabilidade; disponibilidade; containerização; computação em nuvem.

PALABRAS CLAVE: Tecnología; servicios; escalabilidad; disponibilidad; contenedorización; computación en nube.

*Corresponding Author: França, M. T.

ABSTRACT

Cloud computing is a computational model that has stood out in recent years due to its ability to offer an online service in a flexible and scalable way. In light of this, this article aims to conceptualize cloud computing and explore its benefits, including scalability and availability, in order to keep the infrastructure working properly, regardless of the number of simultaneous accesses. Its limitations will also be addressed, such as the problem of data security and the complexity of integration into production systems. Finally, some cloud services will be analyzed and their respective results presented with respect to the research proposal, in order to contribute to the debate about cloud computing, as well as its respective characteristics, availability and scalability.

RESUMO

A computação em nuvem é um modelo computacional que tem se destacado nos últimos anos em virtude de sua capacidade de oferecer um serviço online de forma flexível e escalável. Diante disso, o artigo tem como objetivo conceituar a computação em nuvem e explorar seus benefícios, incluindo a escalabilidade e disponibilidade, de forma a manter a infraestrutura funcionando adequadamente, independente da quantidade de acessos

simultâneos. Também serão abordadas suas limitações, como a problematização da segurança dos dados e a complexidade de integração em sistemas em produção. Por fim, serão analisados alguns serviços em nuvem e apresentados os respectivos resultados no que tange à proposta da pesquisa, a fim de contribuir no debate acerca da computação em nuvem, bem como das suas respectivas características, disponibilidade e escalabilidade.

RESUMEN

La computación en nube es un modelo computacional que ha destacado en los últimos años por su capacidad de ofrecer un servicio online de forma flexible y escalable. A la luz de esto, el artículo pretende conceptualizar la computación en nube y explorar sus beneficios, incluyendo la escalabilidad y la disponibilidad, con el fin de mantener la infraestructura funcionando correctamente, independientemente del número de accesos simultáneos. También abordará sus limitaciones, como el problema de la seguridad de los datos y la complejidad de la integración en los sistemas de producción. Por último, se analizarán algunos servicios en nube y se presentarán los respectivos resultados en lo que respecta a la propuesta de investigación, con el fin de contribuir al debate sobre la computación en nube y sus respectivas características, disponibilidad y escalabilidad.



1. INTRODUCTION

At certain times of the year, some companies have a significant increase in the number of visits to their websites and applications, which often leaves them unprepared, causing slowdowns in operations and even a complete outage of services. A common example is the drop-in sales sites during Black Friday and the slowdown of sites when there is a large flow of simultaneous accesses, as is the case of the National High School Examination (Exame Nacional Do Ensino Médio – ENEM)

According to Monteiro (2011), cloud computing is nothing more than the combination of two previously well-known terms: computing, which alludes not only to the processing and storage of data, but also to the use of services and software for handling data themselves, and cloud, related to a large worldwide computer network, namely the internet.

Armbrust et al. (2010) states that there are two types of cloud: public and private. The public cloud is made available by companies providing cloud services and is available to anyone who is willing to pay for its use. Some examples of public cloud applications include online file storage, such as Dropbox, and the use of software-as-a-service platforms, such as Google Docs.

On the other hand, the private cloud is installed directly within a company or organization and can only be used by that company. Some advantages of the private cloud include greater security and data control, as well as greater flexibility for customizing services according to the company's needs. However, the private cloud can also have higher costs for implementation and management, as the company has to pay for the necessary hardware, software and maintenance costs.

Among the private cloud applications, the storage of confidential data, such as financial information, and the use of internal applications stand out. As an example, one can cite project management platforms, which need access to internal data to organize the tasks necessary for the execution of a company project.

When a company decides to create an online service, information technology (IT) analysts must assess the feasibility of acquiring a private structure or using a public one. The choice between a private and a public structure must consider the system's usage pattern and equipment requirements. If the system remains stable for a good part of the year and only at specific times receives an overload of use, not having much interference in its use and not requiring very robust equipment, a private structure may be the most suitable option. However, if there are frequent fluctuations in overhead, cloud computing may be the best choice, precisely because of its ability to provide resources dynamically, paying only when they are used.

According to Buyya (2011), there are reserved resources and on-demand resources. In the first case, in a business scenario, resources are purchased in advance and are available to the company, occasionally generating a waste of money if resources are not fully used. In the



second case, the resources are dynamically made available according to a set of computational resources (processor, memory, storage) obtained on demand, also known as pools, for a period of time and without long-term commitments, with payment being made according to resource consumption, known as pay-as-you-go.

The way services are scheduled and maintained has a direct impact on their availability. However, the hardware and services used can also have a significant impact. In view of this, the objective of the study is to investigate solutions that can avoid unforeseen events and instabilities in the systems, using technologies that allow to act dynamically in sustaining its operation and maintaining its scalability and high availability, such as micro services and containers.

2. CLOUD COMPUTING

According to Armbrust et al. (2010), the term cloud computing refers to hardware and software systems that provide applications as services over the Internet. These services are known as SaaS (Software as a Service), translated as “software as a service”, although some companies present them as IaaS (Infrastructure as a Service), translated as “infrastructure as a service”, and PaaS (Platform as a Service), translated as “platform as a service”.

For Carissimi (2015), cloud computing has two main concepts: the implementation of computing as a public service and the virtualization of hardware, development environments and service execution. In addition, the author points out that cloud computing is a business model where the user only pays for what he consumes of resources and the service provider maintains a physical infrastructure that is shared between different users.

According to Badger, Grance, Patt-Corner and Voas (2012), cloud computing is a model that allows access, on demand, to a shared network of configurable computing resources quickly and with minimal management effort. These authors emphasize that, contrary to what is suggested by other authors, cloud computing is not always related only to financial return, but also to availability performance issues.

According to IDG – International Data Group (2022), 69% of organizations accelerated their migration to the cloud in the last 12 months, and the percentage of companies with most or all of their IT infrastructure in the cloud is expected to jump from 41% today to 63% over the next 18 months. In this context, according to Veras (2015), the use of this technology proves to be efficient in making resources more financially economical, where the user only pays for what he uses. With this, cloud computing can be seen as a provider of applications as services over the Internet, and these services can have various architectures and types, aiming at saving financial resources, as well as hardware for a company.



3. AVAILABILITY AND SCALABILITY

The availability of resources, according to Pereira et al. (2011), adds several advantages both for the organization and for the users, as complex concerns regarding infrastructure installations and maintenance would no longer exist, with this scalability being the exclusive responsibility of service providers, enabling a more comprehensive attention to needs. business rules relevant to the project.

For IBM Cloud – International Business Machines Corporation (2021), many events can affect server availability, such as hardware failures, software errors and even maintenance that may not have been performed as planned. Therefore, concerns with issues like these cease to be the customer's to be the provider's, hence the importance of these companies being reliable to keep the infrastructure and software in operation and, consequently, good availability.

According to Oracle (2020), the high availability of an architecture requires the implementation of redundancy, monitoring and fail over, these being the main elements. Redundancy involves creating multiple components capable of performing the same task, so that one component can take over the role of another if the latter fails. Monitoring, in turn, is responsible for assessing the proper performance of each component individually. Finally, fail over occurs when a minor component takes over as the primary component after a failure. Therefore, the focus on the infrastructure part is evident, based on the analysis of hardware equipment as the main point of attention.

Because providers need, most of the time, to consume resources at all times from providers, availability becomes a major concern. As the concept of cloud computing starts from the idea of delivering services over the internet, it is expected that they will always be working (Arruda, 2016).

According to Schwartz et al. (2012), scalability is an acquisition of resources through investments as they are needed to deal with the load and capacity required by the service, guaranteeing a quality service at the lowest possible cost.

Scalability is the ability of the system to expand without losing performance. In other words, the system can adjust itself according to demand, being able to increase or decrease its computational capacity if the infrastructure cannot perform normally (De Paula & Dian, 2021).

Availability and scalability are two important factors to consider when choosing a cloud service to deliver data and applications. It's important to choose a reliable provider and implement fail safe measures, as well as take advantage of the flexibility and cost savings offered by cloud scalability.

One way to keep a service stable, reliable and to avoid incompatibility between different hardware is to encapsulate it in a container. A container is a grouping of code and all the dependencies necessary for its execution, being light, autonomous and executable (Docker, 2022).



4. AVAILABLE TECHNOLOGIES AND TOOLS

Cloud computing is a technology that has been widely used in recent years due to the advantages it offers, such as flexibility, scalability and cost savings. The work by Endo, Gonçalves, Kelner, & Sadok (2010) is an example of how these researches are important to better understand cloud computing tools. The authors, Thomé, Hentges, & Griebler (2013) did a comparative analysis of seven different platforms, including Xen, Nimbus, OpenNebula, Eucalyptus, TPlatform, Apache and Enomaly. By analyzing Table 1, it is possible to identify the main characteristics of each cloud service platform. Based on this, it is concluded that there is a need for standardization between these platforms, since they have different service models.

Table 1. Key characteristics of cloud services.

Tool	Interface	Energy management	load balancing	Network	Storage	Monitoring	Integration	Virtualization	Security
Eucalyptus	SSH e WEB	It has	Elastic Load Balancer	Bridge e VLAN	AoE, iSCSI e NFS	Nagios	EC2, EBS, AMI, S3, IAM	Xen, KVM and VMware ESXi	Authentication, CUG and Active Directory
OpenNebula	SSH and WEB (Sustone GLUT)	CLUES	It has	Bridge, VLAN e Open Vswitch	NFS, iSCSI, LVM	OpenNebula Sunstone	EC2	Xen, KVM and VMware	Authentication and CUG
OpenStack	SSH and WEB (Horizon)	Power Management	Quantum Network Load Balancing	VLAN e Open Vswitch	AoE, iSCSI e NFS	OpenStack Clavi	EC2 and S3	XenServer, KVM e Hyper V	Keystone, LDAP, ... external methods
Cloud Stack	SSH and WEB	It has	Citrix NetScaler	VLAN	iSCSI e NFS	Traffic Sentinel	CloudBridge e EC2	Xen, KVM and VMware ESXi	Authentication and CUG
OpenQRM	SSH and WEB	It has	It has	Bridge e VLAN	NFS, iSCSI, AoE e LVM	openqrm-monitor	UEC, EC2 and Eucalyptus	Vmware ESX, Xen, KVM e XenServer	Authentication, CUG and LDAP
UEC	SSH and WEB	UEC Power Management	Does not have Has	Bridge e VLAN	iSCSI e AoE	UEC Monitor	EC2	KVM	Authentication and CUG
Abiquo	SSH and WEB	Does not have Has	It has	VLAN	NFS, iSCSI, LVM	Abiquo Monitor	Cisco UCS	Vmware ESXi, Hyper-V, XenServer, Xen, KVM	Authentication, CUG and LDAP
Convirt	SSH and WEB	Does not have Has	Does not have Has	VLAN	NFS, iSCSI e LVM	Convirt Monitor	EC2	Xen and KVM	Does not have
nimbus	SSH e WEB	It has	Does not have	VLAN	AOE, iSCSI e NFS	Nagios	EC2, S3, Cumulus	Xen and KVM	Authentication and CUG
Apache VCL	SSH and WEB	It has	Does not have	VLAN	iSCSI	Does not have	Does not have	Vmware, KVM	LDAP

Source: Adapted from (Thomé, Hentges, & Griebler, 2013)

Another relevant research is carried out by Khurshid, Al-nayeem and Gupta (2009). In this work, the authors evaluated the performance of Open Cirrus in relation to PlanetLab and EmuLab systems. The results showed that distances have a negative impact on cloud computing performance, which is important to consider when using this technology.

Machado (2011) carried out comparative tests between the OpenQRM and Eucalyptus tools in a controlled scenario, using six computers with the Ubuntu-10.04 operating system. The results pointed out the superiority of OpenQRM for this specific situation.

These works demonstrated the importance of research on cloud computing tools and the need to evaluate the performance and characteristics of these tools to meet market demands.



According to Veritas (2022), a multi-cloud data management company, the term containerization has become very popular in the cloud computing area. The use of VMs (Virtual Machines), the most common process at the beginning of migrations to the cloud, can generate environment inconsistency, dependence on the OS (Operating System), inability to provide isolation above the operating system level, among other problems. Containerization, on the other hand, allows a system to run in isolation and in user spaces, encapsulated with everything it needs to run (libraries, dependencies, etc.), allowing it to migrate between infrastructures without the need for reprogramming. In this context, being more efficient than virtualization, containerization has become a natural evolution.

According to the IBM Cloud Team (2022), when we talk about technologies encapsulated in containers, there are two names that emerge as leaders: Docker and Kubernetes. The most popular tool currently for creating containers is Docker, a toolkit that facilitates the deployment of containers, ensuring their speed and security. Kubernetes is a robust container orchestration tool, allowing you to schedule, automate deployment, and scale a large number of containerized applications.

When it comes to handling IaC (Infrastructure as Code) resources, Terraform does the job well. Hashicorp (s.d), the company that created Terraform, explains that the tool allows you to create versions of configuration files that can be shared and reused. These configurations will manage the infrastructure as a whole, throughout its life cycle, both at a low level (storage of network resources, computing) and at a high level (Domain Name System and Software as a Service resources). It works by generating an execution plan that outlines the changes needed to achieve the desired state of the infrastructure, and then applying those changes in an orderly and consistent manner. This allows infrastructure to be treated as code, with the same version control and collaboration advantages as software development.

5. PROS AND CONS OF CLOUD COMPUTING

For Google Cloud (n.d.), some of the benefits of cloud computing include, scalability and flexibility, advanced security and data loss prevention. In this sense, customers who use the cloud, most often using reliable platforms, have their data protected and with as little possibility of loss as possible, in addition to being able to adapt to the demands of use.

According to Silva (2019), scalability is considered one of the main benefits of cloud computing, being a decisive factor in choosing this technology as an environment for provisioning data and applications. However, offering scalable services can be challenging, as it requires a flexible infrastructure and well-designed software architecture to support growing user demand. In this sense, it is important for companies to assess their needs and objectives before adopting cloud computing, in order to ensure that the chosen solutions are aligned with their business strategies and offer the best possible cost-benefit ratio. As mentioned earlier, Silva (2019) also highlights that the use of the cloud can bring other



benefits, such as data recovery, depending on the business model, implementation and services used.

According to Microsoft Azure (n.d.), migrating to the cloud can still be a challenge, including the need for planning, which must be extensive and well-prepared in order to avoid any unforeseen events and the cost, which, no matter how much the use of computing in the cloud reduce spending, an investment is still required for migration.

According to Google Cloud (n.d.), some obstacles can also be encountered when using cloud systems, including the complexity of integration with existing systems, as well as security concerns such as privacy and online threats. With this, although the cloud platform may be safe, it is necessary to be attentive as the applications are implemented so that there are no failures in relation to the data made available, after all, it is enough for the user not to make mistakes in the implementation of their services that, possibly, the data will be safe.

6. RESULTS AND DISCUSSION

With the use of physical servers, there is usually not a uniform load of utilization, most of the time causing an underutilization or overload of the hardware. Thus, in situations where there are a large number of simultaneous accesses to the physical server, it will have a certain amount of computational power, which will be limited to this value, since if exceeded there will be instabilities.

However, with the use of cloud services, the amount of resources to be used can be defined by demand, considering the use of the Terraform tool to build, change and create infrastructure versions, requesting only the necessary resources during a certain period of use and avoiding spending on a high amount of resources at times of low use. In the end, you only pay for what was used, without extra costs and without worrying about electricity costs to maintain these resources.

From the well-structured combination of Kubernetes and Terraform, resource waste will be minimized, since resources will be used on demand and, therefore, the system becomes self-scaling, allocating resources only when necessary: Kubernetes orchestrating the system nodes and Terraform managing resource consumption, instantiating new containers as needed.

Due to this characteristic, the system becomes more reliable in high demand scenarios, because even when the accesses increase a lot, the system will continue working, in the same way, if there is a problem and it falls, another one will automatically go up in its place and will function again without human intervention.

It becomes evident, therefore, that the adoption of cloud computing is a viable solution for users' data storage and processing needs. Since the service provider provides all the infrastructure resources, allowing the user to focus on strategic and operational issues, without worrying about issues of resources and physical equipment.



However, it is essential to emphasize that data security in the cloud is a shared responsibility between the service provider and the customer, and it varies according to the model adopted when provisioning resources in the cloud. Thus, it is necessary to adopt appropriate security measures to ensure the protection of data stored in the cloud. Top recommended security measures include implementing encryption to protect sensitive data and using firewalls and other security solutions to protect infrastructure.

Importantly, data security is a critical aspect when it comes to cloud computing and lack of proper care can lead to cyber security risks such as leakage of confidential information. In this sense, it is crucial that the service owner reinforces security and is aware of local legislation to avoid violating laws and guarantee the security of the server and end users.

Based on the results discussed, we can say that the combination of cloud services with container management brings numerous advantages for resource management and system scalability. In addition, updating services in real time is facilitated by the simultaneous execution of several containers from each pod, allowing updates to be made gradually, keeping the system online throughout the process.

In summary, the combination of cloud computing with container management using Kubernetes and Terraform is an efficient and scalable solution for managing infrastructure resources, provided that adequate security measures are adopted to ensure data protection.

7. REFERENCES

- Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., ... & Zaharia, M. (2010). A view of cloud computing. *Communications of the ACM*, 53(4), 50-58. Recuperado de <https://doi.org/10.1145/1721654.1721672>
- Badger, L., Grance, T., Patt-Corner, R., & Voas, J. (2012). Cloud computing synopsis and recommendations. *National Institute of Standards and Technology Special Publication*, 800(146). Recuperado de <https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistpecialpublication800-146.pdf>
- Buyya, R., Broberg, J., & Goscinski, A. M. (Eds.). (2010). *Cloud computing: Principles and paradigms* (Vol. 87). John Wiley & Sons.
- Carissimi, A. (2015). *Desmistificando a computação em nuvem*. Instituto de Informática, Universidade Federal do Rio Grande do Sul. Recuperado de https://www.researchgate.net/publication/301298378_Desmistificando_a_Computacao_em_Nuvem
- De Paula, L., & Dian, M. de O. (2021). Computação em nuvem: os desafios das empresas ao migrar para a nuvem. *Interface Tecnológica*, 18(2), 304-315. Recuperado de <https://revista.fatectq.edu.br/interfacetecnologica/article/view/1304/689>
- Endo, P. T., Gonçalves, G. E., Kelner, J., & Sadok, D. (2010). A survey on open-source cloud computing solutions. In *VIII Workshop em Clouds, Grids e Aplicações*. Recife, PE, Brasil. Recuperado de https://www.researchgate.net/publication/228719828_A_Survey_on_Open-source_Cloud_Computing_Solutions
- Thomé, B., Hentges, E., & Griebler, D. (2013). *Computação em Nuvem: Análise Comparativa de Ferramentas Open Source para IaaS*. In 11th Escola Regional de Redes de Computadores (ERRC), page 4, Porto Alegre, RS, Brasil. *Sociedade Brasileira de Computação*. Recuperado de https://gmap.pucrs.br/dalvan/papers/2013/CR_ER_RC_2013.pdf
- Google Cloud. (n.d.). *Advantages of cloud computing*. Recuperado de <https://cloud.google.com/learn/advantages-of-cloud-computing>
- HashiCorp. (n.d.). *What is Terraform?* Recuperado de <https://developer.hashicorp.com/terraform/intro>
- IBM Cloud Docs. (2021). Recuperado de <https://cloud.ibm.com/docs/bare-metal?topic=bare-metal-ha-dr locale=pt-BR>
- IBM. *Kubernetes vs. Docker: Why not both?* (2022). *IBM Cloud Blog*. Recuperado de



Citation (APA): França, M. T., Santos, A. T., Jesus, I. D. C., Teixeira, S. M., Araújo, W. A. G., & Pereira, L. D. L. (2023). The use of cloud computing as an aid to scalability and availability of online services. *Brazilian Journal of Production Engineering*, 9(2), 79-87.

<https://www.ibm.com/cloud/blog/kubernetes-vs-docker>

IDG. (2022). 2022. *Cloud Computing Executive Summary*. Foundry. Recuperado de <https://resources.foundryco.com/download/cloud-computing-executive-summary>

Microsoft Azure. (n.d.). *Benefícios da Migração na Nuvem*. Recuperado de <https://azure.microsoft.com/pt-br/resources/cloud-computing-dictionary/benefits-of-cloud-migration/>

Monteiro, T. S. (2011). *Computação em nuvem, uma análise das abordagens arquiteturas atuais* (Master's thesis). Faculdade Natalense para o Desenvolvimento do Rio Grande do Norte.

Oracle. Learn about high availability in the cloud. (n.d.). Recuperado de <https://docs.oracle.com/pt-br/solutions/design-ha/index.html>

Schwartz, B., Zaitsev, P., & Tkachenko, V. (2012). *High performance MySQL: Optimization, backups, and replication* (3rd ed.). O'Reilly Media.

Veras, M. (2015). *Cloud Computing: nova Arquitetura da TI*. Rio de Janeiro: Brasport.

Veritas Technologies LLC. (n.d.). *O que é containerização? Quais são os benefícios?* Recuperado de <https://www.veritas.com/pt/br/information-center/containerization>

