kriativ.tech

Edition Nº 11, Vol. I – 4-12-2025
ISSN Print: 1646-9976 | ISSN Online: 2184-223X |
http://www.kriativ-tech.com
DOI: 10.31112/kriativ-tech-2023-06-101

Computing Service Models: General Concepts and Differences Among IaaS (Infrastructure as a Service), PaaS (Platform as a Service), and SaaS (Software as a Service)

Diogo Correia
ISTEC Lisboa – diogoribeiro.correia@my.istec.pt

Andreia Teles Vieira
Assistant Professor at ISTEC Lisboa, CIAC-PLDIS –andreia.vieira@my.istec.pt

Abstract: *This study focuses on cloud-computing service models, specifically the differences among IaaS (Infrastructure as a Service), PaaS (Platform as a Service) and SaaS (Software as a Service). Cloud computing is a fundamental technology in today's business environment, offering advantages such as scalability, cost-efficiency and flexibility. The project seeks to demystify each service model by highlighting its unique characteristics and ideal use cases. IaaS provides virtualised infrastructure, PaaS supplies development platforms, whereas SaaS delivers software applications as a service. The aim is to present a detailed analysis that helps readers identify which model best suits their needs, thereby optimising cloud adoption to achieve greater efficiency and innovation in business.*

Keywords: *Cloud Computing, IaaS (Infrastructure as a Service), PaaS (Platform as a Service), SaaS (Software as a Service), Scalability, Cost Efficiency, On-demand Self-service, Elasticity, and Scaling*

I. Introduction

The evolution of information technology has driven profound changes in the way companies operate, and the adoption of cloud computing has become an essential pillar for innovation, operational efficiency, and competitiveness in today's global business landscape. Thanks to its flexibility, scalability, and cost-efficiency, cloud computing has emerged as a pivotal solution to the dynamic challenges faced by contemporary enterprises. In this context, the three cloud-service models—Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS)—play fundamental roles by delivering configurable, on-demand computing resources that support a broad spectrum of business operations. This article therefore explores the general concepts of cloud computing and differentiates among these service models, clarifying their distinctions, unique characteristics, and ideal use cases. The goal is to show how each model can be strategically adopted to optimise business operations and drive innovation.

II. Cloud Computing

According to Braga et al. (2012), drawing on the definition adopted by the United States National Institute of Standards and Technology (NIST), cloud computing is a model that enables ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources—such as networks, servers, storage, applications, and services—that can be rapidly provisioned and released with minimal management effort or service-provider interaction.

This core concept of cloud computing is by no means new. The idea that computing capabilities could be delivered to the public as a utility service had already been envisioned in the 1960s. However, it was only after Eric Schmidt, then CEO of Google, used the term "cloud" in 2006 to describe delivering services over the Internet that the expression "cloud computing" gained the popularity and acceptance it enjoys today.

Unlike many technical terms, "cloud computing" does not denote a brand-new technology per se but rather a new operational model that brings together several pre-existing technologies to enable business to be conducted in innovative ways. This is why multiple interpretations of what truly constitutes cloud computing exist, making the task of establishing a single, standardised definition particularly challenging.

Moreover, the move to cloud computing represents a paradigm shift in how organisations and individuals connect to and use computing resources. The model offers unprecedented flexibility, allowing users to scale resources on demand while paying only for what they consume. This approach contrasts sharply with the traditional IT model, which required substantial capital investment in physical infrastructure and its maintenance.

The growth of cloud computing also raises important issues concerning information security, data privacy and dependence on service providers. These challenges are currently in the spotlight of the technology community, which is actively seeking solutions and best practices. As cloud computing continues to evolve, so too does our understanding of the optimal practices, security standards and management strategies needed to realise its benefits while minimising the inherent risks.

III. Essential Characteristics

The following attributes are fundamental and embody the benefits that cloud computing provides. When combined, these attributes uniquely define cloud computing and differentiate it from other computing models. The ability to provision resources rapidly, broad network access, and service metering are all indispensable elements in designing a cloud-based solution.

Figure 1 - Five Essential Characteristics

a) Elasticity and Scaling

Elasticity refers to the provider's ability to allocate or release resources instantaneously and without interruption as the client's requirements change. This re-allocation must occur rapidly and "invisibly" to the user, creating the impression of an unlimited supply of services. Scalability, by contrast, gives users the perception of inexhaustible resources by automatically provisioning additional capacity as demand grows and withdrawing it as demand subsides. Whereas scalability facilitates the rapid adjustment of computing resources to a user's varying needs, elasticity is distinguished by its focus on enabling the infrastructure to adapt swiftly to demand fluctuations an agility-centred property derived from, yet extending beyond, mere scalability (Freitas et al., 2023).

b) On-demand Self-service

Users can unilaterally obtain computing resources—such as server processing time or network storage capacity—whenever they need them and without any human interaction with the service provider. The hardware and software within a cloud environment can be reconfigured and adapted automatically, and these changes occur transparently to the users. With different user profiles, they can therefore customise their computing environments, including installing software and configuring networking to grant specific privileges. This flexibility enables rapid adaptation to shifts in demand or project-specific requirements, leading to more efficient resource management and performance optimisation. Moreover, this approach minimises the need for manual interventions, simplifies the management of complex technological infrastructures, and allows users to concentrate on their core tasks (Sousa et al., 2010).

c) Broad Network Access

Resources are delivered over the network and reached through standardised mechanisms that allow low-capability devices and thin clients—such as mobile phones, laptops, and PDAs—to use them easily. The cloud-access interface is designed so that users do not need to change their usual configurations and working environments (for example, programming languages or operating systems). Client-side software installed locally to reach the cloud is intentionally lightweight and typically runs through a web browser (Sousa et al., 2010).

d) Resource Pooling

Cloud-service providers must support multiple clients simultaneously by flexibly configuring both physical and virtual resources that are allocated and reallocated in line with users' needs. This model operates under a principle of geographic independence: although users normally cannot determine or control the precise location of the resources they consume, they may be able to select a higher-level location—such as a country, state, or specific data centre (Oliveira et al., 2020).

e) Measured Service

Cloud-based infrastructures autonomously manage and monitor the specific resource needs of various services, including storage, processing capacity, and bandwidth. This management and monitoring must be carried out transparently, ensuring that both the service provider and the end user have a clear, intelligible view of resource utilisation and availability. It is essential that this process be performed in a way that enables both parties to trust the accuracy and fairness of the measurement of the services consumed (Vandresen et al., 2013).

f) Location Independence

Cloud resources must be reachable via the Internet and other networks, allowing access from conventional computing devices. This enables their use across a variety of platforms—such as mobile phones, laptops, and more—while ensuring constant availability without geographic constraints. Adopting this technology therefore requires a careful assessment of specific organisational needs (Vandresen et al., 2013).

g) Resource Virtualisation

Borges, Sousa, Schulze, and Mury, among others (2011), highlighted the capability to decouple infrastructure services from physical components such as hardware and networks. At a lower level, the details of resource location can thus be managed while remaining hidden from the upper layers of the architecture. Thanks to this separation, it becomes possible to provide resources in a utility-like manner, eliminating the need for direct interaction with the underlying hardware (Vandresen et al., 2013).

IV. Service Models

Cloud-computing services are organised into three main categories, each reflecting a different degree of abstraction and delivery structure. This layering shows how the higher tiers build on the foundational ones, creating a hierarchy of resources. The categories are: Infrastructure as a Service (IaaS) at the base; Platform as a Service (PaaS) at the intermediate level; and Software as a Service (SaaS) at the top. These divisions not only clarify the nature of the services offered but also make it easier to integrate them so that they can precisely meet an organisation's operational requirements. A typical scenario might involve using IaaS for virtual storage of applications and data, PaaS for developing and deploying bespoke applications, and SaaS to provide employees with enterprise tools such as email and project-management solutions.

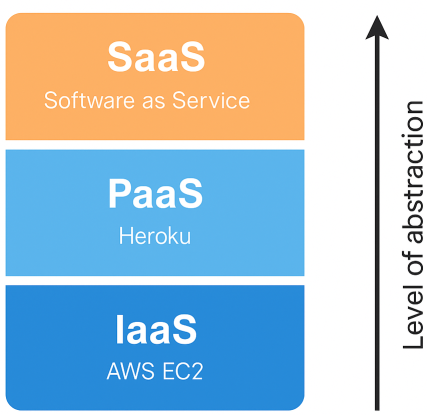

Figure 2 - Layered Diagram

a) Infrastructure as a Service (IaaS)

In an IaaS environment, services are available on demand, providing virtualised *hardware* resources such as compute, storage, and networking. This offering includes servers that can run custom software on a variety of operating systems. A single management-interface application streamlines interaction with hosts, switches, and routers and allows new devices to be added. As the foundational cloud layer, IaaS supplies the resources required by the intermediate and upper layers. It is viewed as the most fundamental model not because of its simplicity but because of the type of product it delivers—indispensable even to the other cloud-service tiers. By adopting IaaS, organisations gain a robust infrastructure for applications and data storage without having to build that base on premises or in branch offices, thereby significantly reducing up-front investment and required floor space.

Amazon is a leading player in this space: through Amazon Web Services it offers a broad portfolio that includes Amazon Elastic Compute Cloud, which provides a scalable architecture and fine-grained control over resource usage. Notably, Amazon did not begin as an IaaS provider but as an e-commerce company, later expanding into data-centre services first to meet its own needs and subsequently those of the market

Other prominent IaaS vendors include Microsoft Azure, Google Cloud Platform, and IBM Cloud, offering services such as virtual servers, storage, load balancing, security, and backup. The advantages of choosing IaaS include lower hardware expenditure and depreciation, elimination of security and maintenance costs, improved performance, freed-up physical space, and the flexibility to adjust compute and storage capacity as needed (Freitas et al., 2023).

b) Platform as a Service (PaaS)

In a PaaS environment, the main beneficiaries are not end-users but developers who want to build their own applications using third-party platforms. Instead of managing the underlying infrastructure, developers can focus on creating and refining their products. They have access to programming languages and a range of both native and third-party tools that are compatible with the platform.

PaaS simplifies software development by eliminating the need for significant investments in complex, expensive hardware and software. It provides an integrated tool-set that supports the entire lifecycle—from creation and testing to deployment and hosting—and incorporates features such as security, database management, and support for diverse web integrations. All of this is accessible over the Internet, offering developers a comprehensive solution.

A major advantage of PaaS is its scalability: users can easily adjust the resources their applications need as demand fluctuates. Leading providers in this segment include Google App Engine, Microsoft Azure, and Heroku, which offer not only complete development environments but also support a wide array of programming tools to streamline the creation of new applications (Freitas et al., 2023).

c) Software as a Service (SaaS)

The SaaS layer leverages the lower Platform-as-a-Service (PaaS) and Infrastructure-as-a-Service (IaaS) layers to deliver a seamless user experience. SaaS stands out for its simplicity and its deep impact on users' day-to-day activities, eliminating the need for advanced technical expertise. Applications are designed to be intuitive, allowing even less-skilled users to navigate them easily. No resource management or in-house software development is required; a device with a web browser and an Internet connection is enough.

Well-known SaaS offerings include email services (e.g., Gmail, Outlook.com, Yahoo Mail), data-storage platforms (Dropbox, OneDrive, Google Drive) and social-media networks (Facebook, Instagram, X/Twitter). Google's ecosystem is particularly valuable because a single account grants access to multiple services, from Gmail to Google Drive, enabling real-time document collaboration among multiple users. Cloud computing has likewise empowered innovative ventures—such as Foursquare, LinkedIn, Peixe Urbano, Zynga and Netflix—without the barrier of heavy up-front IT infrastructure investment. Netflix, for instance, lets subscribers pay a monthly fee to stream a vast catalogue of films, series and documentaries.

SaaS offers numerous advantages, notably the flexibility to access software from anywhere, at any time, on any Internet-connected device. SaaS providers take responsibility for updates and security, freeing users from maintenance concerns. Typical SaaS applications include collaboration suites, project-management tools, customer-relationship-management (CRM) systems, human-resources administration, accounting and marketing platforms. Leading products in this category include Microsoft 365, Google Workspace, Salesforce, Dropbox and Slack (Freitas et al., 2023).

V. Methodology

This study adopts a documentary-analysis methodology focused exclusively on reviewing published articles to examine cloud-computing service models—Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS). It began with the careful identification and selection of relevant articles from academic databases, using specific keywords related to cloud computing and its respective service models. Selection followed inclusion criteria that prioritised publications offering detailed analyses, case studies or comprehensive theoretical reviews of the models in question.

The next phase involved a meticulous analysis of the chosen articles, from which essential information was extracted on each model's characteristics, advantages and disadvantages, as well as its most suitable use cases. The aim was to identify patterns, similarities and discrepancies in the ways different authors address cloud-computing service models.

Compiling and synthesising the extracted data enabled the construction of a comparative framework and an in-depth discussion of IaaS, PaaS and SaaS, reflecting the current state of the literature and providing an integrated perspective on these service models. The study was intentionally limited to published articles to ensure the reliability and authenticity of the information analysed, thereby contributing to a robust, up-to-date understanding of cloud-computing service models and their practical implications in business and technology

VI. Conclusion

The in-depth analysis of cloud-computing service models— IaaS, PaaS and SaaS—reveals a spectrum of solutions that can be tailored to diverse business needs and objectives, ranging from basic infrastructure to complex software applications.

Choosing among IaaS, PaaS and SaaS depends on an organisation's IT strategy, the specific requirements of each project, and the desired level of control over infrastructure, application development and software services. This paper underscores the importance of understanding the particularities of each model to enable effective adoption aligned with corporate goals, thereby ensuring scalability, cost efficiency and innovation. As the technology landscape continues to evolve, organisations that navigate the complex cloud-computing ecosystem efficiently will be better positioned to respond to market changes, capitalise on new opportunities and maintain a competitive advantage in the digital real.

VII. References

[1] Braga, A. S., Silva, G. M., & Barros, M. C. (2012). Cloud computing. Instituto de

Computação-Universidade Estadual de Campinas, 18.

[2] Sousa, F. R., Moreira, L. O., & Machado, J. C. (2009). Computação em nuvem: Conceitos, tecnologias, aplicações e desafios. II Escola Regional de Computação Ceará, Maranhão e Piauí (ERCEMAPI), 150-175.

[3] Oliveira Freitas[1], L., & Neto, A. P. F. COMPUTAÇÃO EM NUVEM: UMA BREVE REVISÃO BIBLIOGRAFICA. energia, 3, 12.

[4] Cogo, G. S. (2013). Análise da intenção de adoção da computação em nuvem por profissionais da área de TI.

[5] Mansur, A. F. U., Gomes, S. S., Lopes, A. D. A., & Biazus, M. C. B. (2010, August). Novos rumos para a Informática na Educação pelo uso da Computação em Nuvem (Cloud Education): Um estudo de Caso do Google Apps. In Foz do Iguaçu: Anais do XVI Congresso Internacional ABED de Educação a Distância (p. 35). Sn

[6] McCoy K. (2017) "Target to pay $18.5M for 2013 data breach that affected 41 million consumers". USAToday, May 23rd 2017 edition, Money section

[7] Oliveira, C., Fambrini, F., de Oliveira, M. G. C., &

Moreira, W. (2020). Computação em Nuvem